AMTSO: the Test of Time?

David Harley CITP FBCS CISSP
ESET North America

Abstract

The Anti-Malware Testing Standards Organization was intended to raise the standard of testing by providing a forum for testing-related discussion, developing standards and best practices for testing, providing education and enhancing awareness, and encouraging the provision of tools and resources. It started well as a coalition of anti-malware vendors and mainstream testers resolved to implement a shift from simpleminded static testing to more realistic dynamic testing. While there is undoubtedly more dynamic (or at least hybrid) testing than there was back in 2008, recent changes suggest that working relationships between some testers and vendors have deteriorated. Can AMTSO really continue to build on its achievements so far, or has it already shot its bolt?

Introduction

The Anti-Malware Testing Standards Organization (AMTSO) was formally founded in May 2008.[1] Since then, the organization has generated some serious documentation and even, from time to time, managed some (often controversial) press coverage.[2,3] That in itself is something of an achievement, considering that the eyes of many journalists glaze over at the idea that testing is interesting. Or, come to that, difficult: there is a whole school of quick and dirty product reviews in generalist computer magazines where non-specialists attempt some evaluation of detection performance.

AMTSO's foundation was the result of many years of concern on the part of anti-malware vendors and some mainstream product testers: concern, that is, that many individuals and organizations offered (and continue to offer) comparative testing and product certification at such a low level of competence and accuracy, consistently underestimating the knowledge and resources required to perform a meaningful test.[4] The organization originally announced its intention to provide a forum for discussion, to develop standards and best practices in testing, foster education and awareness, and to provide or at least encourage the provision of tools and resources.

Aims & Aspirations

AMTSO's aim isn't always put as simply as it might be, but it can be, as per the mission statement on its own index page: it was intended to address the global need for improvement in the objectivity, quality and relevance of anti-malware testing methodologies.[5] This can be seen as largely focused on the issues of objectivity and impartiality, quality, and relevance.

- Testing should be free (to the extent of which it's possible) of hidden agendas and bias, from any source. The test audience is entitled to expect that the test is executed and documented in such a way as to promote the best interests of its audience.
- Quality and sound practice means, essentially, recognition of the fact that testing *in general* is a discipline that in itself requires technical knowledge and experience, testing of software adds a further layer of complexity, and the testing of anti-malware products requires understanding of the complexities of malware and anti-malware technology that is largely restricted to experts and specialists.
- Relevance to and consistency with avowed testing aims is at least as important as the other considerations. And by that I mean something far more complex than statistical accuracy, rarely achieved though that is. Too much testing – not all of it amateur – is about comparing apples to oranges or even melons to raisins, trying to compare products that aren't intended to work in the same way.[6] Even where products have largely comparable base functionality, as is the case with most commercial anti-malware, out-of-the-box testing is not a level playing field unless *all that's being tested* is out-of-the-box configuration. Even then, reviews based on out-of-the-box testing are all too likely to reflect the prejudices of the tester better than the overall capabilities of the products, or even of their detection capability.

Protection & Self-Protection

AMTSO as it was originally founded was important because it pooled knowledge from both the security industry and the security testing industry, giving each the opportunity to learn from the other, and implement a functional system of checks and balances where excessive self-interest could be controlled by a community that was more than an AV pressure group keeping the testers in line. However, much of the coverage the organization has received is hostile. That's not all bad: if everyone loved it, it would probably mean it had been thoroughly ineffective at raising standards without knuckling under to vested interests. But most of that hostility is inspired by the assumption that AMTSO was, in fact, an AV pressure group. That's understandable, given that a high proportion of its members have, from the beginning, been representatives of AV companies. Journalist Kevin Townsend once asked, "Is AMTSO the anti-malware industry looking after itself? (It seems to be almost entirely composed of anti-malware companies and anti-malware testing companies; with little if any input from users.)"[7]

Of course there's an element of self-protection. Of course all testing hurts products that get *bad* reviews. But poor testing isn't only a problem *because* of the problems it creates for products that do badly: the industry isn't so self-protective and cooperative that it tries to look after the weaker products in its market sector. Testing that hurts good products while promoting not-so-good products is not just irrelevant, and it's not *only* bad for the sales figures of the misevaluated product. It's much worse for the customer who puts his or her trust in a product that gives less protection than a test suggests.

Of course, no-one believes that the anti-malware vendors aren't interested in their own bottom line.[8] In fact, many people (including quite a few journalists) assume that the AV industry is the ultimate in cynical exploitation of Fear, Uncertainty and Doubt. And, in fact, most AV marketing is based on the consumer's fear of negative consequences if he doesn't use security products, though that doesn't necessarily make the security industry any more "exploitative" than the pharmaceutical industry or even the agricultural industry.

The unusually pronounced and widespread dislike and distrust of the anti-virus industry is too complex to consider adequately in a paragraph or even a single article.[9] However, two significant yet inconsistent factors are in play: on one hand it is assumed to fabricate psychological dependence by exaggerating the need for its services, yet on the other hand it is lambasted for failing to meet that need by eliminating the malware problem.

"Testing is changing whether vendors like it or not" as one journalist put it.[10] As a matter of fact, vendors *do* like it: they've been advocating better testing for a long time and complaining bitterly about generally low standards in that area. [11] While vendors sometimes have a somewhat self-interested interpretation of what constitutes "good testing" in the context of tests in which they have participated (willingly or otherwise), the industry as a whole has some pretty clear and more-or-less impartial views on what constitutes good practice in testing. After all, the acceptance by testers of practices agreed by a community of vendors and testers doesn't, in principle, work to the advantage of any one vendor.

Not All Practice Makes Perfect

What is *good* practice, then? Well, it's probably easier to define bad practice: at any rate, it can certainly embrace some or all of these well-known and well-documented methodological approaches, though I not every technique that's guaranteed to raise the hackles of vendors (and many testers) is included here[12]:

- Sample sets picked up somewhere on the internet, or out of the tester's own mailbox and possibly "validated" by his own favourite scanner (preferably one that came free), or by submitting samples to VirusTotal and assuming that anything detected by any scanner is malicious.[13] Unfortunately, these forms of pseudo-validation do not necessarily eliminate the possibility of false positives, inappropriate detection of garbage files and so on, and a sound test cannot be based on invalid samples.[14]
- Samples supplied by the company that publishes one of the products under test (strangely enough, those products usually do rather well).
- Simulated malware. In general, the security industry considers a detection of malware that isn't malware as a false positive, though there are exceptions, of which the most obvious is the EICAR test file.[15] However, the naming of that file is misleading: the test file was never designed for product detection testing, but rather as a tool for checking that an antivirus product is installed and capable of detecting real malware. There have been many attempts over the year to evaluate detection performance using modified versions of the EICAR test file, tending to finish up with something that is neither the EICAR test file, nor malware, nor a realistic simulation of malware.[16]
- Kit-generated or self-created malware which may or may not be valid – by "valid", I mean code that is both malicious and capable of being executed .[17]

Creationism & Testing

The use of unequivocally "Blackhat" malware kits poses a number of technical and methodological problems. However, the use of self-created "malware" adds difficulties closely related to those that accompany the use of simulations. Malice, by legal definition, includes some element of "evil" intent.[18] Since security software is normally intended to detect malicious software rather than simulated malware, the tester's aim is presumably to "simulate malice" by including some equivocally malicious payload or other component that "should be" detected. This approach presents considerable technical and philosophical problems, though the presence of genuinely self-replicating (viral) behaviour is an example – possibly the only example – of a behaviour that is almost always considered to incorporate malicious intent. However, there are ethical and legal issues that accompany even controlled replication that, combined with other technical and ethical issues, render custom-created test malware practically useless. The anti-malware industry hates newly-created or modified malware with a passion and for a variety of reasons, but the most pressing technically is that when testers create or modify malware, there's a good chance that the finished article isn't malware as the industry defines it.

Of course, testers don't have to conform to the anti-malware industry's definitions of what malware really is, but the question that must then be asked is whether and why the industry should conform to a maverick tester's view of what should be detected.

Until AMTSO, the AV industry wasn't very good at telling testers (or the public) what sort of tests it *did* consider legitimate.[19] While the intentions of many testers may have been honourable – though it's not unknown for tests to be inspired by hidden agendas that have little to do with the common good – misconceived approaches like the above invariably generate problems and controversy, and may be totally inappropriate, misleading, and open to abuse. There have been occasional concerted efforts to respond to a particularly inappropriate test, but the overall impression in the public mind was of a peevish antivirus industry that didn't like the way testing was carried out but was reluctant to provide feedback more positive than "If you have to ask how to test.....you aren't qualified..."[20] There's some truth in that – of course, since asking for help is by definition an admission that the tester perceives a need for improvement – but it doesn't help people who are genuinely interested in improving their testing. Even worse, it leaves the field open to those whose apparent self-confidence may not be matched by their competence.

Practice & Principles

Of course, when AMTSO started to answer the question, it was accused of telling testers how to test. As indeed it did, in a sense: even before the organization was formally constituted the vendors and testers who were primarily responsible for its formation were trying to move testing away from simplistic static testing towards more accurate (but more resource-intensive) dynamic, whole-product testing.[21, 22, 23]

AMTSO's "Fundamental Principles of Testing" and its growing collection of AMTSO-generated and membership-approved guidelines documents represent an important milestone in the maturation of the anti-malware industry, offering genuine high-level guidance on what it means by good testing practice.

The Nine "Fundamental Principles" of testing as defined by AMTSO are as follows[24]:

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

While it may seem hard to argue with such high-level statements, application of these principles to tests that actually exist in the real world has proved challenging.

Watching the Watchers

One of AMTSO's early initiatives was to add analysis and review of current testing to its list of objectives, and indeed the organization made valiant attempts to meet that objective, but generated so much controversy that the whole process, essentially

based on evaluating conformance with the nine principles, is now undergoing exhaustive review.[25] How did it go so wrong?

One major contributing factor is that the inevitable tension between the interests of vendor marketing and tester marketing resulted in some undesirably self-interested pressure on both sides that caused some testers to take a more arms-length position or withdraw entirely. The introduction of a second-tier subscriber model in addition to the first tier membership model allowed them a trade-off. [26] While subscribers have less influence on the directions AMTSO takes, they still have input, it costs them less, they're under less pressure to conform to AMTSO recommendations for good practice that they consider unrealistic, and they're less susceptible to pressure from vendors trying to negotiate better test results by using AMTSO as a threat. Of course, a somewhat similar trade-off is available to vendors, but the withdrawal of a vendor has less impact (positive or negative) on AMTSO's image than the withdrawal of a tester.

Where did All the Testers Go?

There have been energetic attempts to recruit a wider range of organizations with a security testing remit, but with very little success. Some have not considered the advantage of membership sufficient to justify the cost of membership, and others have declined to join an organization largely made up of the companies they test and an insufficiency of testers. The latter view seems a little too much like Catch 22, but there's more than a chicken *versus* egg paradox here.[27] Large testing organizations with a consumer focus often make a point of not engaging face-to-face with the manufacturers whose products are under test for fear of undue influence.

Unfortunately, there is a practical problem with this principled stand: you don't have to be an engineer to test a washing machine or even a digital SLR – though in the latter case it helps to know more than a little about photography – and you don't need to be a programmer to compare word processors. But the very nature of the security industry and the threatscape it tries to address and mitigate suggests that this aloofness doesn't work to the advantage of the customer. Even a comparative test of editing software is likely to be influenced by the tester's subjective understanding of what a product "should" do, and a journalist's requirements and expectations are likely to be quite different to those of a home user, an academic, a lawyer and so on. Outsourcing testing to a testing specialist may be one way round this objection, but in practice, organizations that take this route often make use of an organization whose expertise and contacts are not necessarily in malware/anti-malware technology.

Testing the Testers

Some testers have expressed a fear that AMTSO will compromise their ability to provide good testing. But the use of the word "standards" in the organization's name does it no favours here.

AMTSO does not and should not prescribe testing methodologies: rather, it provides guidance at varying levels of technical sophistication, put together and approved by people with considerable expertise in complementary aspects of testing and the technology under test. Well, it's hard to argue with transparency, relevance and lack of bias.

AMTSO doesn't set standards in a formal sense like BSI or ISO and does not say who is or isn't allowed to test.[28] Perhaps *someone* should, but a body controlling the certification of testers shouldn't be controlled itself by any single sector: not the academic community, the testing organizations, the anti-malware industry or their customers.[29] And the generation of true standards requires a collaborative effort across a wide range of stakeholders, perhaps under the umbrella of an impartial group such as IEEE.

Someone should be holding testers and reviewers to account for the accuracy of their testing and conclusions, but at this time, AMTSO does not at present seem to have the credibility to address the issue by virtue of its review analysis process, at least in its current (suspended) form. Sadly, it seems inevitable that AMTSO will have to do some serious PR, polishing its image rather than its core processes, before it can usefully address that objective, even if it can mitigate conflicts between the two main groups that constitute its membership.[30] Not only to mitigate the poor image that the AV industry has in general, but also in order to persuade testing organizations that *they* can work with the AV industry without being subjected (or being *seen* as being subjected) to inappropriate pressure.

Conclusion: Breaking Down Mistrust

Security product testing and security software publishing are two sides of the same coin (no currency pun intended). But they are industries, and their aims are not totally compatible. Testers need AV to evaluate, so that they can sell their results. Vendors may not feel (or resent that) they need testers, but tests are, for better or worse, part of the marketing ecology: furthermore, good testing gives vendors feedback on how they're doing in terms of popularity, effectiveness etc. Actually, so does bad testing, but in that instance it's not always useful feedback... But both industries have to watch their bottom line, and each has an impact on the other's financial viability.

The establishment of AMTSO gave testers who already had a good working relationship with the industry a chance to maintain and build on those links and also offered a chance to break down the mistrust between the industry and testers that don't have such links.[31] Sadly, neither industry has taken full advantage of that opportunity.[32] It would be a pity if the organization didn't raise its game in that respect. However, an equal priority should be given to widening the range of informational and educational resources offered not only to testers, but also to the general public. Not only adding to such content, but by maintaining the currency of the resources already there.

Resources:

AMTSO Documents and Principles. Accessed 28 Nov 2011. <http://amtso.org/documents.html>.

Gordon, Sarah. 'Are Virus Simulators Still A Good Idea?' Accessed 28 Nov 2011. <http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6VJG-3WP2C4W-4&_user=10&_coverDate=09%2F30%2F1996&_rdoc=1&_fmt=high&_orig=search&_sort=d&_docanchor=&view=c&_searchStrId=1397764983&_rerunOrigin=google&_acct=C000050221&_version=1&_urlVersion=0&_userid=10&md5=ce3def7e5f12cb12f560a468b02c761d>.

Harley, David. 'Untangling the Wheat from the Chaff in Comparative Anti-Virus Reviews'. Small Blue-Green World, 2007. Accessed 28 Nov 2011. <http://www.eset.com/resources/white-papers/AV_comparative_guide.pdf>.

Kuo, Jimmy. 'Let Telemetry Be Your Guide'. Microsoft, 16 July 2009. <http://blogs.technet.com/b/mmpc/archive/2009/07/16/let-telemetry-be-your-guide-a-proposal-for-security-tests.aspx>.

Vrabec, Jan. 'Generalist Anti-Malware Product Testing'. ESET, 25 Jan 2010. Accessed 28 Nov 2011. <http://www.eset.com/blog/2010/01/25/generalist-anti-malware-product-testing>.

References:

1) AMTSO. ‘Security Software Industry Takes First Steps Towards Forming Anti-Malware Testing Standards Organization’. AMTSO, 4 Feb 2008. Accessed 28 Nov. <http://amtso.org/amtso-formation-press-release.html>.

2) Harley, David. ‘Antivirus Testing and AMTSO: Has anything changed?’ AMTSO, 2010. Accessed 28 Nov 2011. <http://www.amtso.org/uploads/cfet2010-antivirus-testing-and-amtso.pdf>

3) AMTSO. ‘AMTSO in the Media’. AMTSO, 2 Jun 2010. Accessed 28 Nov 2011. <http://amtso.wordpress.com/amtso-in-the-media/>.

4) Tanner, Sarah. ‘A Reader’s Guide to Reviews’. Virus News International, November 1993: 40-41, 48.

5) Anti-Malware Testing Standards Organization. Accessed 28 Nov 2011. <http://www.amtso.org/>.

6) Harley, David. ‘Making Sense of Anti-Malware Comparative Testing’. Information Security Technical Report, 2009. Accessed 28 Nov 2011. <http://dx.doi.org/10.1016/j.istr.2009.03.002>.

7) Townsend, Kevin. ‘AMTSO: a serious attempt to clean up anti-malware testing; or just a great big con?’ 15 Jun 2010. Accessed 28 Nov 2011. <http://kevtownsend.wordpress.com/2010/06/15/amtso-a-serious-attempt-to-clean-up-anti-malware-testing-or-just-a-great-big-con/>.

8) Townsend, Kevin. ‘Anti-Malware Testing Standards Organization: a dissenting view’. 27 June 2010. Accessed 28 Nov 2011. <https://kevtownsend.wordpress.com/2010/06/27/antimalware-testing-standards-organization-a-dissentingview/>.

9) Harley, David. ‘I’m OK You’re Not OK’. Virus Bulletin, November 2006. Accessed 28 Nov 2011. <http://www.virusbtn.com/virusbulletin/archive/2006/11/vb200611-OK>.

10) Finley, Klint. ‘Antivirus Product Testing is Changing, Whether Vendors Like it or Not’. ReadWrite Enterprise, 25 Jun 2010. Accessed 28 Nov 2011. <http://www.readwriteweb.com/enterprise/2010/06/antivirus-product-testing-changing.php>.

11) Wells, Joe et al. ‘Open Letter’. Cybersoft,, 2000. Accessed 28 Nov 2011. <http://cybersoft.com/whitepapers/paper_details.php?content=cs008>.

12) Harley, David; Lee, Andrew. ‘Testing, Testing: Anti-Malware Evaluation for the Enterprise’. AVAR Conference, 2007. Accessed 28 Nov 2011. <http://www.eset.com/resources/white-papers/Testing_Testing.pdf 2007>.

13) Harley, David; Canto Julio. ‘Man, Myth, Malware and Multiscanning’. Cybercrime Forensics Education & Training Conference, September 2011.

14) Košinár, Peter; Malcho, Juraj; Marko, Richard; Harley, David. ‘AV Testing Exposed’. Virus Bulletin Conference, 2010. Accessed 28 Nov 2011. < http://go.eset.com/us/resources/white-papers/Kosinar-etal-VB2010.pdf>.

15) Harley, David; Myers, Lysa; Willems, Eddy. ‘Test Files and Product Evaluation: the Case for and against Malware Simulation’. AVAR Conference, 2011. Accessed 28 Nov 2011 < http://go.eset.com/us/resources/white-papers/AVAR-EICAR-2010.pdf>.

16) Willems Eddy. ‘EICAR 2010: Rainy Days in Paris’. Virus Bulletin, June 2010.

17) AMTSO. ‘Issues Involved In The “Creation” Of Samples For Testing’. AMTSO, 13 Oct 2009. Accessed 28 Nov 2011. <http://www.amtso.org/amtso---download---issues-involved-in-the-creation-of-samples-for-testing.html>.

18) Malice (legal term). Accessed 28 Nov 2011. <http://en.wikipedia.org/wiki/Malice_(legal_term)>.

19) Harley, David. 'AMTSOlutely Fabulous'. Virus Bulletin, April 2010: 11-12. Accessed 28 Nov 2011. <http://amtso.org/uploads/vb_amsto_article_jan_2010.pdf>.

20) Harley, David; Bridwell, Larry: 'Daze Of Whine And Neuroses (But Testing Is FINE)'. Virus Bulletin Conference, Sep 2011. Accessed 28 Nov 2011. <http://www.virusbtn.com/pdf/conference_slides/2011/Harley-Bridwell-VB2011.pdf>.

21) CARO. Workshop Presentation Slides. F-Prot, 2007. Accessed 28 Nov 2011. http://www.f-prot.com/workshop2007/presentations.html.

22) Harley, David. 'Execution Context In Anti-Malware Testing'. EICAR Conference, 2009. Accessed 28 Nov 2011. <http://smallbluegreenblog.wordpress.com/2009/05/15/execution-context-in-anti-malware-testing>.

23) Muttik, Igor; Vignoles, James. 'Rebuilding Anti-Malware Testing for the Future'. Virus Bulletin Conference, 2008. Accessed 28 Nov 2011. < http://downloadcenter.mcafee.com/products/mcafee-avert/whitepapers/muttikvignoles_vb2008.pdf>.

24) AMTSO. 'Anti-Malware Testing Standards Organization: Fundamental Principles of Testing". 31 Oct 2008. Accessed 28 Nov 2011. <http://www.amtso.org/amtso---download---amtso-fundamental-principles-of-testing.html>.

25) Harley, David; Bridwell, Larry. 'Daze Of Whine And Neuroses (But Testing Is FINE)'. Virus Bulletin Conference, 2011. Accessed 28 Nov 2011. http://www.amtso.org/uploads/cfet2010-antivirus-testing-and-amtso.pdf.

26) AMTSO. 'AMTSO Widens the Conversation of Anti-Malware Testing with New Subscription Option'. AMTSO, 25 Oct 2010. Accessed 28 Nov 2011. <http://www.amtso.org/pr-20101025-amtsowidens-the-conversation-of-anti-malware-testingwith-new-subscription-option.html>.

27) Chicken Or The Egg. Accessed 28 Nov 2011. <http://en.wikipedia.org/wiki/Chicken_or_the_egg>.

28) Harley, David. 'AMTSO not ISO'. AMTSO, 6 Jul 2010. Accessed 28 Nov 2011. http://amtso.wordpress.com/2010/07/06/amtso-not-iso-standards-and-accountability/.

29) Harley, David; Lee, Andrew. 'Who Will Test The Testers?' Virus Bulletin Conference Proceedings, 2008. Accessed 28 Nov 2011. http://www.eset.com/resources/white-papers/Harley-Lee-VB2008.pdf.

30) Lee, Andrew. 'The edge of reason(ableness): AV Testing and the new creation scientists'. AVIEN, 7 Jul 2010. Accessed 28 Nov 2011. <http://avien.net/blog/?p=539>.

31) AMTSO. Accessed 28 Nov 2011. http://www.amtso.org/quote-sheet.html.

32) Harley, David. 'Antivirus Testing and AMTSO: Has anything changed?'. Computer Forensics Education and Training, Sep 2011. Accessed 28 Nov 2011. <http://www.amtso.org/uploads/cfet2010-antivirus-testing-and-amtso.pdf>.