DAZE OF WHINE AND NEUROSES (BUT TESTING IS FINE)

David Harley
ESET North America, West Ash Street, Suite 1900, San Diego, CA 92101, USA
Email david.harley@eset.com

Larry Bridwell
Global Security Strategist, AVG Technologies CZ, s.r.o., Holandská 4, Brno 639 00, Czech Republic
Email Larry.Bridwell@avg.com

ABSTRACT

According to Aerosmith (not to mention *The Italian Job*), FINE is an acronym for (in its politer version) Freaked out, Insecure, Neurotic and Emotional. We could (and probably will) offer alternatives, but there's no doubting that anti-malware testing inspires all those reactions. Sometimes it seems that AMTSO has become a dumping ground for the rest of the world's misgivings about the AV industry, even though it originated in a coalition with some of the testers who are monitoring that industry's performance with the most assiduous professionalism: indeed, that coalition has in itself inspired mistrust. And recently, it's become plain that even within AMTSO, both testers and vendors sometimes find the alliance problematical.

AMTSO's purpose is simple to state, but much harder to achieve. It represents a realization by professional testers and security vendors that the quality of anti-malware testing was so variable that it was at best confusing for people who need guidance on how to select the best product for their needs. Perhaps testing has improved more in the past few years than it would have without AMTSO's presence, and discussions and generation of material in a single forum has accelerated a much needed move away from static testing towards dynamic testing. But it's time to ask (and attempt to answer) a number of vital questions:

- *Looking over the historical evolution of testing before and since AMTSO, is that move enough to set the testing world to rights?*
- *Are the aims of testers and vendors close enough to allow continued cooperation within AMTSO?*
- *Has AMTSO already outlived its usefulness?*
- *If not, what should it do next?*
- *What is the future of comparative detection testing?*

INTRODUCTION

According to Aerosmith et al. [1], FINE is an acronym for (in its politer version) Freaked out, Insecure, Neurotic and Emotional. Anti-malware testing inspires all those reactions and many more.

Since it was formally founded in May 2008 [2] the Anti-Malware Testing Standards Organization (AMTSO) has become a dumping ground for the rest of the world's misgivings about the AV industry [3], however worthy its stated, vendor-neutral aims may appear.

Item 1	Providing a forum for discussions related to the testing of anti-malware and related products.
Item 2	Developing and publicizing objective standards and best practices for the testing of anti-malware and related products.
Item 3	Promoting education and awareness of issues related to the testing of anti-malware and related products.
Item 4	Providing tools and resources to aid standards-based testing methodologies.

Table 1: AMTSO's published charter [4].

AMTSO's origins lie in a coalition between the AV industry (which has always viewed with concern some of the poor methodology, lack of accountability and variable accuracy of some tests) and some of the testers who are monitoring that industry's performance with the most assiduous professionalism. That coalition has in itself inspired mistrust [5]. And recently, it has become plain that even within AMTSO, both testers and vendors sometimes find the alliance problematical. But then, it would be naive or disingenuous to pretend that the symbiotic relationship between testers and vendors doesn't mean there is no divergence between the vested interests of the testing and security industries [6]. Indeed, the willingness of representatives of both industries to compromise, trading off (some) self-interest for the benefit of the consumer, is one of the more attractive features of AMTSO, and should help to keep the organization honest.

Nonetheless, suspicions remain that, despite the participation (at subscriber level, if not at member level [7]) of most of the mainstream, specialist testing organizations, AMTSO is intent on establishing standards that weight tests in favour of the AV industry's own favoured but inadequate technologies. There's no disputing that the anti-malware industry is not providing anything like a 100% solution, nor should it claim that it does. But the testing and certification industries are, essentially, dedicated to *distinguishing* between vendors:

- Testing products in order to distinguish between adequate (certified but normally unranked) products and inadequate products (those that fail to meet the certification threshold – again, performance is normally unranked)
- Testing products in order to distinguish the best (top-ranked) products from less competent (lower to lowest-ranked) products.

SINGING FROM THE SAME SONG SHEET

Acting in concert (whether as AMTSO or as the AV industry, and whether with or without the endorsement of testers) might be (and obviously is) seen as having some sinister purpose: is it so hard to believe that even AV researchers might be inspired (at least in part) by altruistic motivation: that is, to improve testing in the best interests of the consumer?

Back in the dark ages before AMTSO, there was no real way for vendors to speak in concert on poor testing. Sometimes an individual vendor would complain about a perceived unfair showing in a test: if they did so in private to the tester, it sometimes led to useful discussion and maybe both parties

learned something, but when an individual complaint was made in public it was more likely to be dismissed as vendor whining.

When vendors *did*, as occasionally happened, act in concert [8], it was still dismissed as vendor whining: oddly, perhaps, since individual vendors wouldn't necessarily benefit, in terms of product ranking, from improvements in a test. But if anyone joined the AV industry out of a desire to be universally loved, they will have been disappointed...

When vendors got together with testers who *also* had a genuine interest in raising testing standards to form AMTSO, people still sat back in anticipation of vendor whining. Even worse, people who had always assumed that vendors always behaved with total self interest (and that testers were always beyond reproach) started to mistrust testers who co-operated with vendors [5].

Testers who saw their job as being to catch the vendors out in order to distinguish the 'best performers' were dismissive of any testers who cooperated with researchers in the AV industry, describing them as 'the vendors' pets', and as compromising their independence by sharing samples. Missing the point that vendors have long shared samples with each other and with trusted testers, on the principle that the safety of the community at large takes precedence over competitive advantage [9].

(That isn't to say that no vendor has ever gamed a test by providing the tester with samples that other vendors are less likely to detect. Nor would we suggest that no tester has ever encouraged such practice in the hope of establishing a significant variation in performance between products to ensure that product ranking isn't too close to impress the reader, but no one said that the security industry is staffed entirely with saints-in-the-making [10].)

Perhaps, as one of the authors has previously remarked [3], the current polarization of views on the usefulness or otherwise of the organization is a necessary adjunct of its dramatic increase in public visibility in 2010?

Document name	Date approved
AMTSO Fundamental Principles of Testing	31/10/2008
AMTSO Best Practices for Dynamic Testing	31/10/2008
AMTSO Best Practices for validation of samples	7/5/2009
AMTSO Best Practices for Testing In-the-Cloud Security Products	7/5/2009
AMTSO Analysis of Reviews Process	7/5/2009
AMTSO Guidelines for Testing Network Based Security Products	13/10/2009
AMTSO Issues Involved in the 'Creation' of Samples for Testing	13/10/2009
AMTSO Whole Product Testing Guidelines	25/5/2010
AMTSO Performance Testing Guidelines	25/5/2010
AMTSO False Positive Testing Guidelines	22/10/2010*
AMTSO Testability Guidelines	4/5/2011

*Document requires re-approval due to concerns about requirement level wording.

Table 2: Documentation deliverables.

THE LONG AND WINDING ROAD

How far along the road to encouraging better testing practice, and to achieving better levels of objectivity, quality and relevance, have we travelled in AMTSO's official lifetime? Table 2 outlines the documents published since 2008.

And that's where it gets interesting.

THE TIMES THEY ARE A-CHANGING?

'Testing is changing whether vendors like it or not', as a journalist we won't name and shame put it not long ago when he recycled a tester's press release. Actually, vendors *do* like it: they've been advocating better testing for years. Perhaps the most frustrating aspect of the tester/vendor symbiosis from a vendor's point of view is the need to spend time and resources on tweaking products to perform well in tests rather than (or at any rate as well as) on performance (in the broadest sense of the word) in the real world, on the customer's desktop, where it matters most [11].

But testers have their own frustrations, being constrained by:

- Their own resources, aptitude and ability
- Test targets imposed by the sponsor/magazine/client
- Contractual obligations that may be used to obscure pertinent detail
- Publisher control over the interpretation of data, which may be quite different to the tester's interpretation.

Dynamic testing is, in principle, a better reflection of 'real life' testing than static testing. However, it's expensive and resource-intensive, and appropriate methodologies are evolving, rather than fully baked. That's why there is so much documentation on dynamic testing and related topics such as network testing, execution context and such on the AMTSO guidelines and resources pages.

Testers and magazines test what they have the time, inclination and resources to test. Testers who provide test data to magazines are constrained by the requirements of the publisher, who generally wants it cheap, fast, and in accordance with what they think the audience wants to read. Testers and reviewers are realists: they provide the data that their readers want and expect, not necessarily what they really need. Most readers want to be told what to buy, not how to choose what they should buy, or how to interpret results. We'd like to think that magazine reviews are not tailored to favour advertisers and sponsors, but it's clear that where a vendor sponsors a test or runs its own tests, it's always a possibility that unfavourable data will not always be published, however honest the testing.

AMTSOSPHERIC PRESSURE

Clearly, these issues can be a cause for frustration for vendors and testers alike: the real pain comes when one industry or the other uses an organization that in principle represents the interests of the community at large in order to pursue its own interests irrespective of community interest.

It may be that AMTSO would not exist at all without an element of self-interest, of course. While all vendors would probably prefer to outdo all their competitors in as many tests as possible, most of them will settle for a fair chance at doing well on a reasonably level playing field, and expect AMTSO

to help achieve that (currently fairly distant) goal. Testers and vendors alike expect returns from their sizeable membership fees or considerably smaller subscription fees, whether it's the kudos and credibility of demonstrating engagement with the organization and its aims, or the ability to input into and influence what directions AMTSO goes in and what it delivers.

However, the past few months have seen heightened tension [12]. While it's reasonable to expect best efforts from AMTSO-affiliated testers to conform to the agreed AMTSO principles [13], it would not be reasonable, for example, for a vendor to attempt to use that affiliation or AMTSO's internal procedures to bully a tester into giving them special treatment.

Similarly, having participated in the heated discussion that often precedes the publication of AMTSO guidelines documents [14], we would hope that clued-up testers would also be aware of the contents of approved documents and recommended resources [15] and follow recommendations where appropriate. On the other hand, guidelines documents are just that, *guidelines*, not one-size-fits-all, inviolable Standards With A Capital S, and should not be treated like the ten commandments.

And while it is part of AMTSO's remit to provide testing tools and resources where appropriate, it is not in itself a hammer, let alone a war-club.

Of course, it goes the other way, too. It's not a forum for vendor-bashing (or competitor-bashing: and that goes for vendors, too...), or for blocking sensible recommendations just because it's more convenient to stick to discredited practices and methodology.

AMTSO's purpose is simple to state (see Table 1 and the charter at http://www.amtso.org), but much harder to achieve. It represents a realization by professional testers and security vendors that the quality of anti-malware testing was so variable that it was at best confusing for people who need guidance on how to select the best product for their needs.

The organization was formed with the intention of raising the general standard of testing in terms of relevance, objectivity and overall quality, and was intended to do so in a number of ways:

- Facilitate discussion
- Develop and publicize best practices and objective standards
- Educate and raise awareness
- Provide tools and resources.

Since AMTSO started to be active, mainstream testing has, in our opinion, improved substantially [16]. That would probably have happened anyway to some extent even without AMTSO's presence, given that so many researchers in both industries were becoming so focused on testing problems and issues. Nevertheless, the discussions and generation of material in a single forum has accelerated a much needed move away from static testing towards dynamic testing.

Even outside the industry, more magazines and testers are talking the talk, at least as far as 'real-world testing' is concerned, though a good many of them aren't really walking the walk. That's understandable, because real-life dynamic testing is expensive, resource-intensive, and requires advanced knowledge and careful planning, but dynamic testing on the cheap isn't sound testing. Too often the result is tests that are defined as 'dynamic' or 'real-world' but which don't take into account the other factors that make up a useful test: objectivity, sample validation, relevance, correct classification of samples, appropriate configuration of tested products, statistical validity, whether the test cases (we probably shouldn't say samples any more) are actually representative of the wider threat landscape, and so on.

BULLSHINE AND BULLSEYES

But you could express the value of AMTSO more simply: whatever mistakes it has made to date, and however much you mistrust the fact that it includes a lot of vendors, it has scored notable successes.

Just a few years ago, testers could say, with justification, 'You [the wider AV community, not just vendors] don't like the way we test, but you don't offer any constructive suggestions on how we could do better. It sounds as if you'd rather we didn't test at all.' That's no longer a viable claim. There's no Blithering Idiot's Guide to Anti-virus Testing (we hope) but there's plenty of material – from AMTSO guidelines to a wide range of conference papers and articles, many written by AMTSO members combining AV and testing expertise – available on or linked to from the AMTSO site or blog site, that any tester or aspirant tester can use. It's a pity that it's so difficult to interest the media in these resources unless they spark off widespread controversy, but I suppose that's inevitable, given that their audiences aren't generally interested in fairly technical material.

AMTSO has made the point forcibly that testers should be as accountable to their customers as vendors are (or should be). This is probably one of the reasons it is so unpopular, along with the fact that it is seen (not altogether accurately) as the voice of the intrinsically unpopular anti-virus industry: it puts testers on the defensive in the same way that testers put the industry on the defensive. We're not saying that it's a bad thing for either industry to be 'defensive' in the same sense that good programming is defensive – as long as the focus is on proactively pushing for practical objectives, of course. But clearly, there are sensitivities on both sides of the tester/vendor divide.

These gains have made it harder for a tester to say 'These are our conclusions: trust us, they are accurate, and there is no need to discuss them.' Of course, some testers do say that (or at least work on the implicit assumption that they can), but some sectors of their audience are now less likely to accept that stance uncritically. What does this have to do with the first point? In some cases, the same journalists who don't consider AMTSO resources newsworthy are also those who either implement magazine tests or interpret them in reviews, and may therefore feel threatened by AMTSO's emphasis on better (and usually more demanding) methodologies. Clearly, raising standards is not the same as making the best the enemy of the good.

CONCLUSION

So whither AMTSO, and indeed, whither detection testing?

A previous paper flagged a number of conceptual problems with detection testing [17].

- It doesn't take into account an unascertainable margin for error. Is it really appropriate to base product evaluation on 'tricking' vendors into generating 'false negatives with esoteric sample sets' with the intention of exaggerating the differences between 'winners and losers' (we've seen tests where a single missed test case was flagged as a failure rate of 10% or more without making it clear how small the test set was), in a field where the margin for error is already provably high, even if the exact margin is incalculable [11]?
- It is largely based on the assumption that testers are better at collecting and selecting samples than AV vendors – an assumption that doesn't really stand up to scrutiny.

In fact, the perfect detection test is no more practical than the Perfect Antivirus [Solomon], and in terms of difficulty and practicality, even performance testing without the complications of detection testing is far more difficult than many testers (and most of the public) realize.

As noted elsewhere [3], AMTSO's importance lies in its ability to pool knowledge from the testing industry and the security industry (not that those two industries are totally discrete sets), so that each learns from and restrains the other, implementing a functional system of checks and balances. AMTSO isn't just an AV pressure group, or the AV's way of keeping the testers in line. (If we thought of it that way, we wouldn't devote so much of our time and attention to it.)

We would rather it be seen as an educational resource and discussion forum, focused on coordination rather than enforcement, and improving methodology by enhancing objectivity, quality and relevance.

Whine and cheese party? Thanks, but no thanks.

REFERENCES

[1] http://en.wikipedia.org/wiki/F.I.N.E.

[2] AMTSO. Security Software Industry Takes First Steps Towards Forming Anti-Malware Testing Standards Organization. 2008. http://amtso.org/amtso-formation-press-release.html.

[3] Harley, D. Antivirus Testing and AMTSO: has anything changed? CFET 2010 conference proceedings.

[4] http://www.amtso.org/.

[5] Townsend, K. Anti-Malware Testing Standards Organization: a dissenting view. 2010. https://kevtownsend.wordpress.com/2010/06/27/anti-malware-testing-standards-organization-a-dissenting-view/.

[6] Harley, D. Security Zone: Antivirus Testing Standards at a Crossroads. Computer Weekly. http://www.computerweekly.com/Articles/2011/05/17/246670/Security-Zone-Antivirus-testing-standards-at-a-crossroads.htm.

[7] AMTSO. AMTSO Widens the Conversation of Anti-Malware Testing with New Subscription Option. 2010. http://www.amtso.org/pr-20101025-amtso-widens-the-conversation-of-anti-malware-testing-with-new-subscription-option.html.

[8] http://www.dslreports.com/forum/remark,16730700.

[9] Harley, D. Scareware and Legitimate Marketing. http://blog.eset.com/2010/09/19/scareware-and-legitimate-marketing, 2010.

[10] Harley, D. I'm OK, You're not OK. Virus Bulletin, November 2006. http://www.virusbtn.com/virusbulletin/archive/2006/11/vb200611-OK.

[11] Kosinár, P.; Malcho, J.; Marko, R.; Harley, D. AV Testing Exposed. Proceedings of the 20th Virus Bulletin Inernational Conference, 2010.

[12] http://securitycritics.org/wp-content/uploads/2011/04/amtso.txt.

[13] http://www.amtso.org/amtso---download---amtso-fundamental-principles-of-testing.html.

[14] http://www.amtso.org/documents.html.

[15] http://www.amtso.org/related-resources.html.

[16] Harley, D. Untangling the Wheat from the Chaff in Comparative Anti-Virus Reviews. Small Blue-Green World. 2007. http://www.eset.com/us/resources/white-papers/AV_comparative_guide.pdf.

[17] Harley, D.; Lee, A. Call of the WildList: Last Orders for WildCore-Based Testing? Proceedings of the 20th Virus Bulletin International Conference, 2010.