Guidelines for Testing Network Based Security Products

amtso™

Anti-Malware Testing Standards Organization

Notice and Disclaimer of Liability Concerning the Use of AMTSO Documents

This document is published with the understanding that AMTSO members are supplying this information for general educational purposes only. No professional engineering or any other professional services or advice is being offered hereby. Therefore, you must use your own skill and judgment when reviewing this document and not solely rely on the information provided herein.

AMTSO believes that the information in this document is accurate as of the date of publication although it has not verified its accuracy or determined if there are any errors. Further, such information is subject to change without notice and AMTSO is under no obligation to provide any updates or corrections.

You understand and agree that this document is provided to you exclusively on an as-is basis without any representations or warranties of any kind whether express, implied or statutory. Without limiting the foregoing, AMTSO expressly disclaims all warranties of merchantability, non-infringement, continuous operation, completeness, quality, accuracy and fitness for a particular purpose.

In no event shall AMTSO be liable for any damages or losses of any kind (including, without limitation, any lost profits, lost data or business interruption) arising directly or indirectly out of any use of this document including, without limitation, any direct, indirect, special, incidental, consequential, exemplary and punitive damages regardless of whether any person or entity was advised of the possibility of such damages.

This document is protected by AMTSO's intellectual property rights and may be additionally protected by the intellectual property rights of others.

Guidelines for Testing Network Based Security Products

This document provides guidelines for testing network based security products which intercept and evaluate network traffic for threats before the traffic continues to the endpoint host system. The document outlines additional issues involved in best practice testing of such products, above and beyond other AMTSO guidelines and best practices. These guidelines are not a comprehensive listing of all such issues.

Unless otherwise defined herein, all terms included in this document are used with their common meaning. The following document should be read in conjunction with AMTSO's *Fundamental Principles of Testing, AMTSO Best Practices for Dynamic Testing, AMTSO Best Practices for Validating Samples, AMTSO Best Practices for Testing In-the-Cloud Security Products*, and other information available at www.amtso.org.

Test Case Relevance

Test cases should be relevant to the target audience of the test. The tester should first define the target audience of the test, followed by the relevant use cases for that audience, and what are the test cases and metrics that will best test the use cases.

A few examples of relevant metrics are:

1. Relevant protocol support and product context
    a. It may be important to evaluate protocol support proportionally for the different contexts of the products such as HTTP versus Common Internet File System(CIFS) and Service Message Block (SMB) support.
2. Appropriate performance considerations
    a. Latency versus Detection tradeoffs. Low latency may be an important attribute for HTTP traffic, but not for email protocol traffic.

Software

In general, the testing of the software component of network-based security products can be based on guidelines and best practices outlined in other AMTSO documents. The tester should be aware of and take into consideration any additional variations introduced as a result of implementing security products at the network level.

A significant difference between host-based and network-based scanning is the range of data streams scanned. A host-based anti-malware product can scan objects supported by the underlying technology components. The scope of objects supported will often dictate the types of threats against which this technology engine will offer protection. Similarly, network-based products are usually limited to scanning on a restricted number of popular protocols. Typically, the more protocols supported by a device, the more protection offered.

Some products will not support rarely exploited protocols. If such protocols are included in a test and not weighted according to their importance in the real world, the test may not reflect real-world product performance fairly. While it may be legitimate to test a product's performance in an esoteric scenario, a tester should make it clear if this performance does not significantly affect its real-world efficacy. More importantly, it would not be appropriate to exclude an important protocol such as SMTP from testing if a product did not support it.

The tester should have knowledge of the significance of each specific protocol in order to make relevance and scoring decisions. For example, it is more critical for network-based products designed and marketed for use within and between subnets to support network communication protocols such as CIFS and SMB, whereas a product designed strictly as a gateway device would primarily need focus on internet protocols. The tester should consider if or how products designed for different functions should be compared.

When testing network-based products at the network level, certain product attributes that may or may not be measured in the test will be weighted differently from host-based products in evaluation by the end user. For example, latency may be more important when evaluating network based products because it can often affect the computing performance of other users in the network. A security product denying or impairing performance and service can have as negative an impact on an organization as some security breaches. In such a case, latency would be more important to the end user than exceptional detection abilities. The weight given to latency results may vary across tests. For example, latency delays may considered more acceptable for a network-based product used only for scanning email protocols. Measurement is complicated further because the product may be shipped with default settings that affect detection vs. latency performance. Certain proactive and heuristic technologies may slow down scanning, but produce better detection, or vice versa. As a result, a product may be capable of protecting against certain threats, yet have poor default setting configurations from a security perspective in the interest of avoiding the introduction of network latency. Testing organizations should consider latency and other product attributes specific to network- based products, and differing from host products in approval and certification processes.

There may be several different software modules operating concurrently on a network security product. These products are commonly referred to as Unified Threat Management (UTM) products, and may include such technologies as intrusion prevention, intrusion detection, anti-malware, anti-spam, content control, and so on. In many cases, these products may be able to provide a higher level of overall security than a single component product. The tester should understand that single component products cannot be expected to compete directly with multi-functional security products within the UTM product category. There are special cases where the tester may have to make difficult ethical decisions when testing products with different functionality. For example, a network based antimalware product may block a worm using signatures, while a competing UTM product may detect the threat with a functional IDS engine, but not with an anti-malware component. In this case, products should only be held accountable for and tested for anti-malware functionality claimed by the product marketing and documentation. Such scenarios add complexity to tests. Therefore, products should ideally be tested against directly comparable product categories.

Expanding upon the AMTSO document, *Best Practices for Validation of Samples*, the tester should consider any cases where sample relevance at the network level differs from sample relevance when testing products at the application level on the endpoint. Methods utilized by malware to infiltrate and propagate at the network level differ from those at the application level. Depending upon the conditions under which a security engine detects a threat, detection abilities at the network level may vary from the normal, overall capability of the engine. For example, some engines may detect a threat upon execution in the target environment.

Hardware

As with host based security products, hardware influences network-based anti-malware product performance. There are several issues with network based security products that should be considered by testers. First, the actual security software is in most cases only approved by the vendor for use on specific hardware. Hardware requirements often differ according to the performance required by the user. Hardware requirements for the testing scenarios to be carried out should be respected by testers, as well as network traffic load the product is designed to handle.

The tester may find it appropriate to consider product attributes that address how the product reacts and performs when power to an appliance unit fails or when network traffic exceeds maximum capacity, and so on. The results of these and other scenarios can lead to service denials by blocking safe traffic, as well as to security breaches when potentially malicious traffic is able to get through. Even though these scenarios are unrelated to the efficiency of the underlying security engine, hardware limitations can affect detection ability of the overall product.

Hardware products have additional product features and specifications that should be considered for testing over and above the criteria appropriate to host-based solutions. When defining test sets for traffic, a diverse range of protocols, frame sizes, number of clients, number of sessions, number of connections, payload sizes, payload file-types, test durations, etc. should be defined to mirror real world scenarios. For example, are vendor claims such as throughput and other performance measures met? Is stability maintained over long periods of time in stress testing? What happens in the case of an additional traffic spike? The tester should design their testing procedures to mirror real world scenarios that stress-test over various time and network load capacity intervals.

This document was adopted by AMTSO October 13, 2009