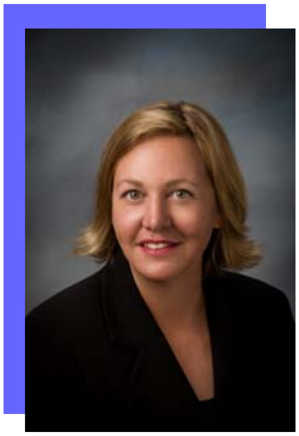
False Sense of Security:

New Anti-Virus Testing Methodologies are Critical to Educate Customers

Charlotte Dunlap

Independent Security Analyst



Charlotte Dunlap is an independent security analyst and regular columnist for Forbes.com, covering primarily secure messaging, threat management, and hosted services. She has two decades of experience as a senior industry analyst and high-tech journalist.

Charlotte has worked for research firms including Current Analysis and has written for leading industry publications including Dark Reading, Information Week, and CNET, and spent an eight-year stint at Computer Reseller News as a senior editor. She also served as European bureau chief for news service Edittech International, based in London.

Introduction

Traditional methodologies used to test the effectiveness of anti-virus solutions are no longer adequate in providing an accurate gauge of a product’s performance. Methods that worked in the past— designed to test for worms and viruses in a stagnant environment unconnected to the Internet— are incapable of assessing protection against the new forms of malware that are now prevalent. The old methods are often based on a static list of threats, and the vast majority of malware is not even included in that list. The industry is doing customers a disservice by stamping a lab certification on their boxes, indicating they have been through rigorous testing procedures when in fact they have not. The static testing methods are far behind the reality of rapidly evolving threats from the Internet.

What is needed is new, Internet-savvy methodology to test the efficacy of anti-virus security. The new methodology should reflect the way current threats are propagating under real-world scenarios. This paper will discuss traditional anti-virus product testing methods and describe how they fall short in providing customers with the most accurate insight into how well security products fight today’s malware. We discuss the realities of today’s testing environment, including the limited scope of testing among the major testing bodies, the increasingly sophisticated threat landscape that demands new real-time tests, and the economic realities of changing current testing methodologies.

Why Existing Test Methodologies are Broken

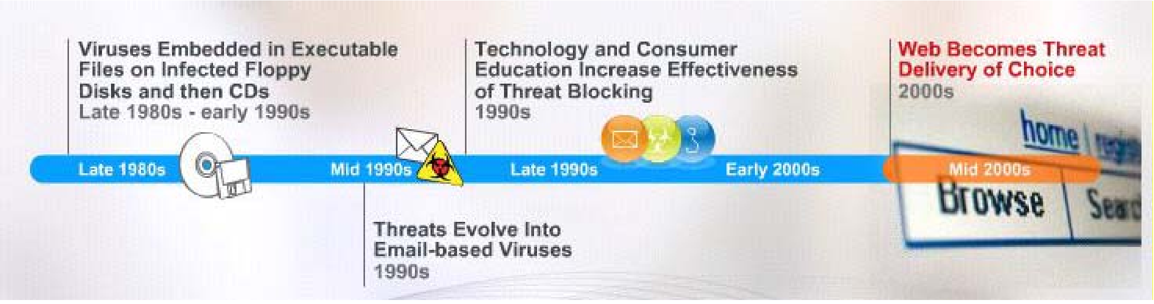
The debate surrounding the use of the WildList or the Virus Bulletin list as a threat protection testing methodology has been underway for several years, but the need to update the industry’s current testing methods has becoming more urgent in light of the way threats are now spreading.

Traditionally, test labs’ primary method of testing anti-virus solutions has been the use of a list of threats, compiled primarily by security vendors. The list is used as the foundation for testing and certifications by labs including ICSA, Westcoast Labs, Virus Bulletin, AV-Comparatives and others. In the past, anti-virus vendors and third-party testers used the industry-standard list to compare the effectiveness of their software. Labs test multiple products by security vendors against this list on a regular basis (as often as monthly) and issue a pass/fail mark.

This approach was fine for testing past threats that included viruses and worms. However, threats have evolved. Threats are now monetarily motivated, authored by cyber-criminals looking to steal data for profit, and delivered using the web in order to keep malware under the radar.

Threat Evolution:

Exploiting the Newest, Most Popular and Least Secure Delivery Methods


Viruses Embedded in Executable Files on Infected Floppy Disks and then CDs
Late 1980s - early 1990s
Technology and Consumer Education Increase Effectiveness of Threat Blocking
1990s
Web Becomes Threat Delivery of Choice
2000s
Late 1980s
Mid 1990s
Late 1990s
Early 2000s
Mid 2000s
home
Browse
Threats Evolve Into Email-based Viruses
1990s

Modern Malware Characteristics

- **Low Visibility.** The last thing criminals want is for their malware to make the news and set off alarms to law enforcement, so cyber-criminals are looking to cause a limited number of infections using one type of malware.
- **Quiet Damage.** There has been a clear shift from headline-making worms and viruses to Trojans, which don't automatically spread and do their damage quietly, stealing data without disrupting other work.
- **Rapid Evolution.** Of the tens of thousands of malicious programs in the wild, each piece of malware detected is constantly evolving, and may have hundreds or even thousands of variants associated with it. This is why the industry is now documenting approximately 50,000 new malware samples per day. Criminals are constantly pushing new forms of malware through the Internet to evade advanced threat protection solutions.
- **Short Lifespans.** The average lifespan of a typical piece of malicious software is one to two days, so malware may live anywhere from a couple of minutes or even seconds, to several days, usually depending on the expertise of the author.

- **Self Updating.** The discovery of the Conficker worm in November 2008 marked a change in malware capability. Written by professional criminals, the worm spreads to other machines without the need for human interaction. But Conficker as well was able to update itself via the Internet, and did this several times, like all modern malware. The WildList only reflects worms, viruses and some variants of bots which contain self-replicating malware. And yet this collection represents only a small subset of today's threats—about 5 percent to 10 percent, because self-replicating malware is not the way people get infected anymore.

In response to these more sophisticated threats, vendors have developed advanced security technologies aimed at tackling malware such as Trojan horses and botnets. Yet testing methods do not take into consideration new threat management technologies, like blocking threats at their source, the Internet, and are still focused on file-based technologies. The WildList does not include Trojans, rootkits, keyloggers, and spyware. The list contained 922 viruses in August 2009, and TrendLabs reports a new piece of malware is now created every 1.5 seconds.

Because of the changing nature of the threats, the industry is sorely lacking in adequate product testing services that help customers make informed decisions about security management. More often, confused security managers are hesitant to make new purchases without having access to up-to-date standardized efficacy benchmark tests. For users to have relevant product information and for security industry to prove its relevance and continue its steady market growth, more real-world testing is required. This issue needs to be a priority to the security industry, especially considering the fact that anti-virus software community competes on its ability to respond quickly to new virus and malware threats.

Perhaps most worrisome of all, the broken testing system gives users have a false sense of security. Research indicates that organizations are pinning unrealistic expectations on that prominent checkmark stamped on their anti-virus boxes. Of 499 respondents surveyed by testing body NSS Labs (October 2009), half believed their endpoint anti-virus software would protect them from malware 100 percent of the time. Another 10 percent thought their software would protect them 99 percent of the time. However, this same testing body in recent real-time testing of AV solutions found that for zero hour threats, leading vendors protected against malware 26 percent to 70 percent of the time and in subsequent days provided overall protection against malware 67 percent to 96 percent of the time.

Current State of the Testing Market

Indeed, vendors and testing bodies all agree the WildList, and other collections like it, only provides a baseline of measurement for security protection. Security experts from around the globe gather regularly to debate the issue and discuss solutions. AMTSO (Anti-Malware Testing Standards Organization) is the most prominent consortium created in 2008 to develop best practices and standards around improving anti-malware testing methodologies.

The issue seems straight-forward. The WildList and the VB100 list are not timely just by the nature of their research-gathering techniques. For a new threat to be added, a minimum of two independent reporters must file the same threat information, and follow a process which delays publishing by as much as weeks and even months. Testing bodies should simply do away with this method and conduct live, continuous Internet-based tests as a way to measure the quality of a product. But it’s not so simple.

- **Replication of live testing is not easy.** The difficulty in setting up new methodologies that involve dynamic lab tests is that by the mere nature of the Internet, the tests cannot be reproduced, and therefore, it is difficult to prove why one product may have passed or failed a test. It is difficult (if not impossible) to ensure competitive products receive the exact same tests. (Currently the world’s largest international standards body ISO, among others, requires that a test be repeatable and reproducible.)

- **Malware is geographically sensitive**. A testing machine may be sitting on a US domain and it will gather different forms of malware. The Conficker virus infecting machines in various countries had more damning effects in some parts of the world vs. others, depending on the country in which the computer resided.

- **New dynamic testing methodology is resource-intensive, and therefore very expensive**. It is more affordable to have 20 products scan half a million samples than to have the same products scan 50 threats using dynamic testing. That’s because real-time testing over the Internet is difficult to automate and requires hands-on testers to move the tests along. For example, if the product presents pop-up queries, someone needs to be on hand to respond.

- **Need to understand the timing of threat interception by a security product.** Risks and impact of threats differ depending on where the security product intercepts it— before it reached the machine, whether it executed, or was detected after it executed. Real-time testing requires testers to understand this measure of potential impact and have a granular expertise while static testing simply determines whether a product detected a threat or not.

Testing bodies are very much aware of challenges of dynamic tests. However, they are keen to solve the problem and make dynamic tests possible or they risk becoming obsolete.

Today's Most Accurate Tests

NSS Labs launched its Live Testing methodology for anti-virus solutions in the summer of 2009. The labs concurrently test anti-virus products and are connected to the Internet. They examined tens of thousands of malware sites and found as many as 350 new malware per day during a 17-day testing period in July. NSS is leading the pack in conducting real-time, concurrent tests which provide an apples-to-apples comparison of threat protection by allowing competing products to hit the same URLs at the same time.

Westcoast Labs has also begun offering a dynamic version of its Checkmark Certification, recently announcing its first security vendor to take advantage of the new services. ICSA is currently not conducting any real-time testing. However, ICSA and others such as Virus Bulletin state they will be evolving their certification practices in coming months to include real-time testing and/or testing against today's threats. These alterations include testing against Trojan horse programs to mirror the threats encountered in the wild by enterprise users and consumers more accurately.

Key Testing Principles

So what is a comprehensive testing formula based on current threat conditions? A number of criteria make up the most effective testing guidelines, and while many in the security-testing sector are planning major upgrades in their methodology, labs are at various stages in applying the methodologies. Some key principles of new testing methodology should include:

- **Real-time or Dynamic Testing**: Computers must field live threats in order to demonstrate the level of zero-day and ongoing protection provided by products. A more holistic approach in testing AV products will better reflect a corporate user's daily and varied habits. Testing needs to replicate real-world behavior such as:
    - visiting websites
    - downloading content
    - simulating attacks such as social engineering
    - exposing vulnerabilities that result in drive-by downloads
    - executing malicious files
    - ..

  Ideally, the dynamic testing process will be automated, visit a variety of web sites, download content, and execute malicious files. Lastly, the test should take into account the number of false positives a product triggers. The goal would be to determine when a threat is blocked: from the source URL, email or IP address, when download is attempted, or on execution. The best approach is to block the threat earlier, at its source, or download, rather than execution.

- **Repeatability and Reproducibility**: Variations of malicious software can be generated in seconds, making it difficult to test multiple products against the same exact malware, although some testers are trying to get around this dilemma by conducting concurrent tests. If tests cannot be exactly replicated, what steps can be taken?

  - Testers must be able to provide documentation to verify details of how a product reacted to specific malware. This is a sensitive issue because some test labs have begun to conduct dynamic testing and vendors have disputed claims that its technology missed particular malware.
  - Testers must provide a consistent reporting system, such as adequate log records, to support claims of missed threat detection..

- **Broad and Diverse Sampling**: Sample contributions by vendors and testers need to include a broad spectrum of various classes of threats to reflect a comprehensive view of Internet malware, including threats that are relevant to specific regions and various markets. For example, it may not be fair to some vendors if sampling emphasizes a vertical such as banking spyware, or if it does not represent various geographic regions.

- **Time to Protect**: In addition to measuring the ability of vendor products to protect against threats known to the testers, measurements of vendor response times to previously unseen threats are needed. By simultaneously exposing vendor products to new, previously unseen threats and then repeating the exposure over time, one can measure how long it takes the vendor products to protect against them. Early measurements of this “time to protect” indicate that most vendor products respond in the range of one hour to one week. This information translates directly to how well customers are protected and provides a highly significant discrimination in performance.

What Customers Can Do

Customers should understand that current static testing methodologies are inadequately measuring protection because threats are evolving faster, spreading quietly, and have different goals than the world in which these tests were designed. To understand a product’s real-world protection capability more accurately, customers can take the following steps:

- **Look beyond the checkmarks and certification** such as ICSA, Virus Bulletin and others not based on live testing. They will not accurately predict a product’s protection from current and rapidly evolving threats.
- **Evaluate products based on Internet-savvy tests** such as NSS Labs Live Tests and Westcoast Labs’ dynamic Checkmark Certification.
- **Ask testing bodies to move from static testing** to more realistic, live testing methods.
- **Check for independent tests, not those commissioned by the company.** NSS Labs and soon, AV-test.org provides well documented, independent testing.

Summary

Current testing methodologies are no longer relevant to today's threat landscape, and test results or pass/fail stamps do not offer corporate users enough intelligence to make the best purchasing decisions. The testing boards <state whom, specifically> view the state of the testing market as being so broken that certification stamps are no longer meaningful to the industry.

While AMTSO has been formed and other steps to strengthen testing are being taken, the industry will need to agree on a set of key testing principles very quickly. Testing bodies will need to revamp their labs significantly to move from static to real-time, dynamic testing processes. If this issue does not move to the forefront of vendor and testing body business strategies in 2010, growth in the threat management market will stall as customers become disenchanted with the industry's ability to guide and educate them about how well their solutions actually protect them.

This report is sponsored by Trend Micro, Inc. Learn more at www.trendmicro.com.