Bringing Testing into the Cloud
Testing Metrics & Methodologies for Cloud-client Security Infrastructure

Anthony Arrott, Wei Yan, Geoff Grindrod & Jeffrey Wong
Trend Micro, Inc.

ABSTRACT

An expanded method for testing AV products is described and demonstrated. The proposed method expands AV product protection scoring from malware file detection rates to include credit for blocking access to the URLs that serve downloaded malware files. This expansion is required because AV vendors have responded to the vast increase in the rate at which new unique malware variants are introduced to the Internet by blocking malware files based on their source URLs. Blocking by malware source provides a faster time-to-protect and utilizes less customer resources in updating threat information to the point of protection. The ability to include the performance of this extra layer of protection in benchmarking tests of competing AV products is critical for customers in assessing the real protection AV products provide.

INTRODUCTION

In 2007 Williamson and Gorelik wrote:

> "There is a desperate need for new standards for today's anti-virus products. The dominant paradigm, scanning directories of files, is focused on old and known threats, and reveals little about product efficacy in the wild." [1]

New Unique Threats per Hour (worldwide estimate)
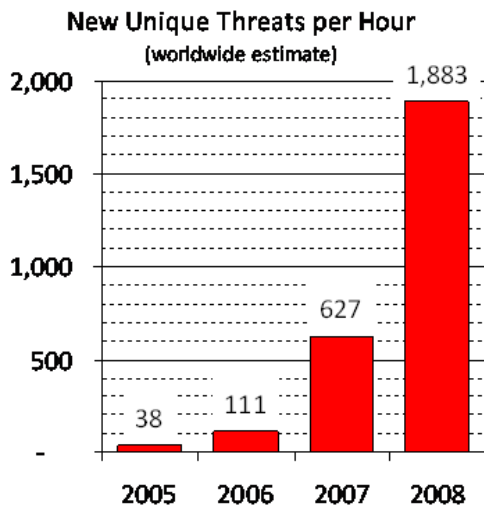
At the time this was written, the number of new unique threat variants appearing on the Internet had risen over the previous two years from less than 50 per hour in 2005 to over 600 per hour in 2007. In the two years since, the number of new unique threat variants has grown to over 2,000 per hour. [2]

One response of the AV industry to this "volume of threat" has been to issue more frequent updates. Many AV vendors have switched from weekly updates to daily or even half-hourly updates. The consequent volume of updates places a significant resource load on customer systems and networks required to handle update downloads – often leading to critical performance and cost issues.

In addition to more frequent updates, other more significant AV technology innovations have been made to defend against the onslaught of new threats: improved vulnerability assessment, behavioral analysis, and source reputation services. [3,4] These innovations attempt to block the actions of malware threats whose malicious code has not been previously indentified and placed in blacklist databases.

Unfortunately, while AV products have introduced more effective techniques, most AV product testing continues to use methods that expect all AV products to block threats by identifying the malicious nature of the code itself. In order for customers to make realistic assessments of how

competing AV products will perform, AV product testing needs to be expanded to credit the innovative AV techniques that block malware even when the identity of the malware itself is not known to the AV product.

This paper focuses on expanding AV product testing to add credit for just one additional AV technique: Web reputation services. It is beyond the scope of this paper to discuss expansion of AV product testing to other defense layers, such as vulnerability assessments – although these expansions are also needed.

BLOCKING WEB THREATS BY SOURCE REPUTATION

The methodology presented here credits AV products for using web reputation services to protect against web threats by blocking access to compromised or malicious URLs capable of downloading malware files to a target computer. Here is one scenario:
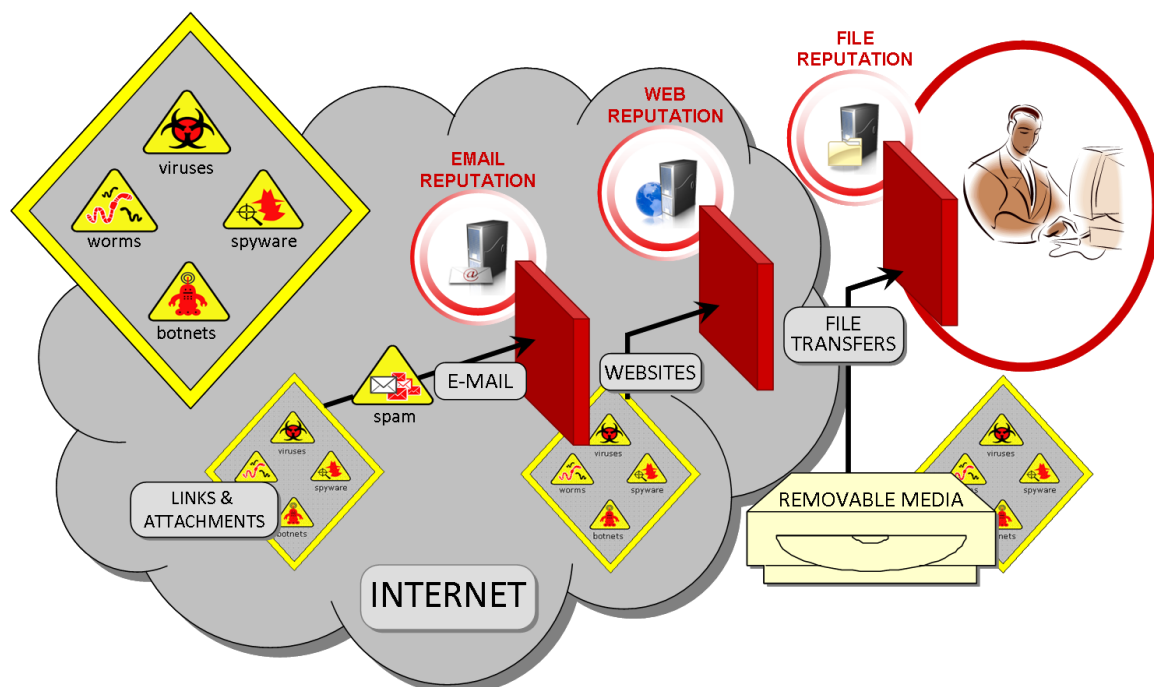

Imagine being sent a malicious e-mail with an embedded Web link…

- First, the AV protection intercepts the email and checks the IP address of the sender against its email reputation database. If the computer address belongs to a spammer, the email is blocked.
- Next, the Web link in the email is checked against the AV product's Web reputation database to ensure users are blocked from accessing malicious Web pages.
- In addition, components hosted on the Web page and related redirected Web pages are automatically downloaded by the AV vendor and analyzed. Identifiable signatures of malware files are placed in the AV product file reputation database.
- If the malware file subsequently arrives at the target computer by other means (e.g., file transfer from a USB storage device), the malware file is detected and blocked by the AV product file reputation service.
- Embedded IP addresses are also extracted, analyzed and, if they pose a threat, are added to the AV product's interconnected, Internet-based threat databases.

The AV product testing methods used currently in most published reports require the AV product to detect malware threats based on content characteristics of the malware files (e.g., its MD5 hash). In these traditional tests, AV products that use reputation services are not given credit for blocking access to a malicious download URL before a malware file is downloaded. The testing methodology proposed here expands AV product performance measurement to include the additional malware protection afforded by Web reputation services that block the malware before it is downloaded.

AV PRODUCT TESTING METHODOLOGY

The testing methodology described here measures the ability of AV products to block malicious URLs and associated malware files based on examination of URL source ("exposure layer protection") and examination of file characteristics of the downloaded malware ("infection layer protection").
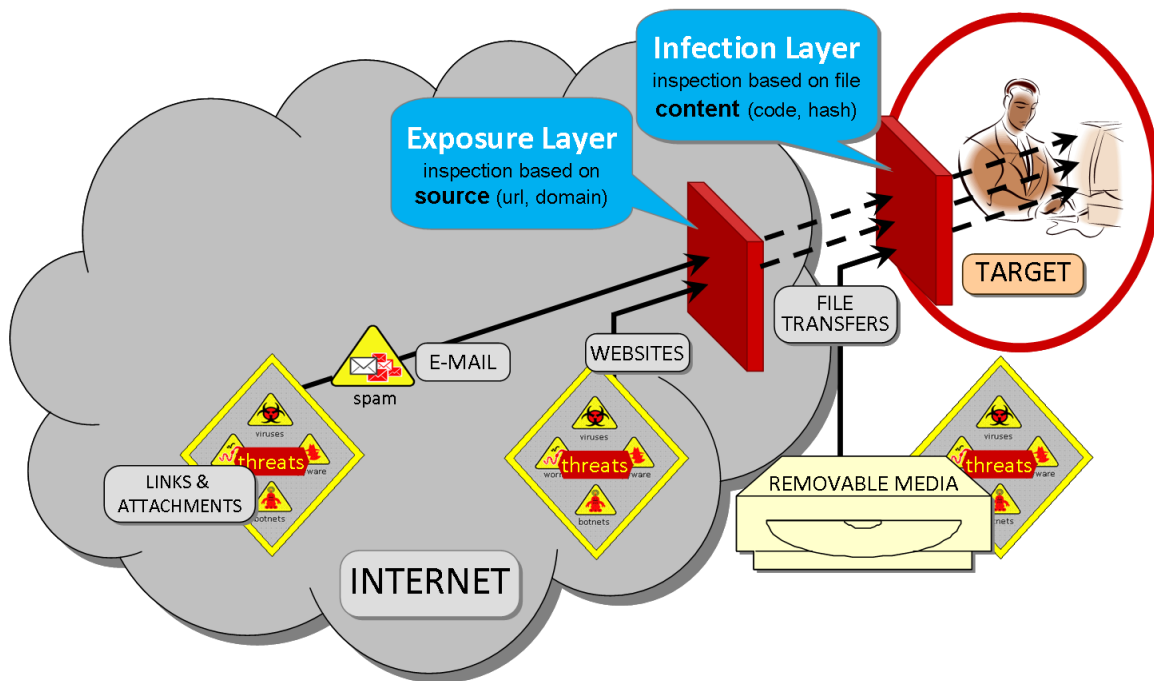
The tests are conducted using unique malicious URLs which are all active and working during the time of the testing. Each AV product being tested (e.g., AV1, AV2, AV3 ,…), is installed on a separate computer (physical or virtual OS).

Each test computer attempts to access the set of malicious URLs. If a malicious URL is blocked, the AV product is given credit for blocking the associated malware files that the malicious URL would have otherwise downloaded.

In addition, each test computer is exposed to all the downloaded malware files from the full set of malicious URLs, regardless of whether the exposure layer blocked access. In this way, the performance of the "infection layer" is tested for all malware files – not just for the ones allowed by the "exposure layer".

However, the "infection layer" test needs to distinguish between detected malware files associated with blocked URLs and those associated with unblocked URLs. Both results are needed in the analysis of AV product performance.

Test Measurements

The following measurements are made from the tests of each AV product:

a. URLs exposed – the number of URLs for which access is attempted;

b. URLs blocked – the number of URLs for which access is blocked by the "exposure layer" of the AV product;

c. Files exposed (total for all URLs) – the number of downloadable files associated with all of the URLs for which access is attempted;

d. Files detected (from total of all URLs) – the number of downloadable files associated with all of the URLs which are detected by the "infection layer" of the AV product;

e. Files exposed (total for unblocked URLs) – the number of downloadable files associated with only those URLs that the "exposure layer" of the AV product fails to block;

f. Files detected (from unblocked URLs) – the number of downloadable files associated with unblocked URLs that are detected by the "infection layer" of the AV product;

g. End-to-end files blocked (blocking downloading URL or detecting file) – the number of files associated with all of the URLs that are blocked by either the "exposure layer" or the "infection layer".

To demonstrate this methodology, four AV products were tested using a set of 85 malicious URLs each with one associated downloadable malware file.

NOTE: These demonstration tests were conducted by TrendLabs, Trend Micro's global network of research, service, and support centers. The example results shown are for AV products that do not include Trend Micro products since the malicious URL and malware test sets are also from Trend Micro. Some of the AV products used in this demonstration test were poorly tuned for the malware source URLs used by Trend Micro. Two of the products failed to block any of the malicious URLs.

TEST MEASUREMENTS	No AV	AV 1	AV 2	AV 3	AV 4
a. URLs exposed	85	85	85	85	85
b. URLs blocked	0	41	0	0	3
c. Files exposed (total for all URLs)	85	85	85	85	85
d. Files detected (from total of all URLs)	0	30	44	39	33
e. Files exposed (total for unblocked URLs)	85	44	85	85	82

f.	Files detected (from unblocked URLs)	0	10	44	39	3
g.	End-to-end files blocked (blocking downloading URL or detecting file)	0	40	44	39	36

Protection Layer Performance (Efficacy)

From these test results, four protection layer performance measurements can be derived:

h. Exposure layer URL protection (=b/a) – the percentage of malicious URLs blocked by the AV product "exposure layer" (i.e., based on source);

j. Exposure layer file protection (=1-e/c) – the percentage of potentially downloadable malware files blocked by the AV product "exposure layer" (i.e., malware files blocked based on URL source reputation);

k. Infection layer file protection (=d/c) – the percentage of potentially downloadable malware files blocked by the AV product "infection layer" (i.e., malware files blocked by detecting the files themselves);

m. End-to-end file protection (=g/c) – the percentage of potentially downloadable malware files blocked by the combination of the AV product "exposure layer" and "infection layer" (i.e., malware files blocked by either URL source reputation or detection of the file).

PROTECTION LAYER PERFORMANCE	No AV	AV 1	AV 2	AV 3	AV 4
h. Exposure layer URL protection (=b/a)	0%	48%	0%	0%	4%
j. Exposure layer file protection (=1-e/c)	0%	48%	0%	0%	4%
k. Infection layer file protection (=d/c)	0%	35%	52%	46%	39%
m. End-to-end file protection (=g/c)	0%	47%	52%	46%	42%

AV Product Scores

AV product scores and indices can be derived from the protection performance at each layer and the end-to-end protection across all layers.

n. Exposure Layer Score (=((h+j)/2)*100) – combines the measurements of blocking both malicious URLs and the malware files associated with them. (For the simplified test set of this example, all malicious URLs have one associated malware file – this is not necessarily always the case);

p. Infection Layer Score (=k*100) – corresponds to the malware detection rate score in traditional AV product testing that credits only malware file detection.;

q. End-to-End Protection Score (=m*100) – measures the ability of the AV product to stop malware regardless of what layer does the blocking; This score gives no credit for redundancy among the protection layers.

r. Overall Protection Index (((n+p)/2 + q)/2) – combines the End-to-End Protection Score with an average score for the individual layers. This index gives credit for redundancy among the protection layers.

AV PRODUCT SCORES	No AV	AV 1	AV 2	AV 3	AV 4
n. Exposure Layer Score (=((h+j)/2)*100)	0	48	0	0	4
p. Infection Layer Score (=k*100)	0	35	52	46	39
q. End-to-End Protection Score (=m*100)	0	47	52	46	42
r. Overall Protection Index ((n+p)/2 + q) /2	0	44	39	34	32

Strong exposure layer protection can significantly augment infection layer (file detection) protection. The AV-2 product has the worst score at the infection layer, but, when combined with its higher performance exposure layer, the end-to-end protection is second among the tested AV products. When the ability of the AV-2 product to block many of the malware files at both the exposure and infection layers is taken into account, as it is in the Overall Protection Index, then the AV-2 scores highest.

Alternative scoring systems can be adopted without affecting the methodology or results in the basic test measurements (a through g) or the protection layer performance metrics (h through m). In calculating the Overall Protection Index, one could weight the exposure layer score more heavily than the infection layer since less customer IT resources are required to block threats at the exposure layer (i.e., before the threats arrive at target computers).

TEST IMPLEMENTATION ISSUES

This methodology can be incorporated as an add-on expansion to most current AV product testing labs. The biggest change is likely to be the way test sets of threats are sourced and presented to the test AV products.

The AV products being compared are best tested as contemporaneously as practically possible to avoid diminishing the efficacy of the malicious URLs and associated malware files.

Instead of a set of known malicious URLs and associated malware files, a stream of unsorted Internet traffic can be used with this testing methodology. The "test set" of traffic can be sorted post-hoc to extract the results for unique bad URLs and downloaded malware files. This approach can be expanded to include an assessment of the frequency of "false positives" (legitimate URLs and downloaded files blocked an AV product). However, in practice, some sort of filtering to reduce volume is likely required to make the test set manageable for the test computers.

CONCLUSION

Advances in AV product technology, like the current shift to cloud-client architectures based on identifying the URL sources of malware, require AV product testers to keep pace if customers are to be well-informed of AV product performance relevant to their own protection needs. AV product developers typically focus on innovations that improve performance according to the scoring methods of AV product testers. As a leading AV product testing lab has said:

> "It is essential for testers to move on to the next level of product testing, focusing on everything besides traditional signature detection. If this doesn't happen users may be misled by inadequate results." [5]

REFERENCES

[1] Matthew M Williamson & Vlad Gorelik, "Towards new standards for real-time evaluation of anti-virus products", Virus Bulletin, 2007.

[2] Trend Micro 2008 Annual Threat Roundup and 2009 Forecast
http://us.trendmicro.com/us/threats/enterprise/security-library/threat-reports/index.html

[3] Neil MacDonald. "Understanding the nine protection styles of host-based intrusion. prevention" Gartner RAS Core Research Note G00127317, 2005

[4] "Trend Micro Smart Protection Network: Core technology", Trend Micro white paper, 2008.
http://itw.trendmicro.com/smart-protection-network/pdfs/SmartProtectionNetwork_WhitePaper.pdf

[5] Andreas Marx. "Malware vs. anti-malware: (how) can we still survive?", Virus Bulletin, 2008.