

# **Trend Micro Web-Filtering Accuracy Tests**

1-Million URL English-Language Corpus  
First Quarter 2006

Prepared by



Seattle, Washington, USA

April 2, 2006

## Executive Summary

---

Under contract with Trend Micro, Inc. ("Trend"), Cascadia Labs independently evaluated the effectiveness of five Web filtering products at blocking Web pages with potentially objectionable content. The products we tested included both gateway appliances and software (shown below in alphabetical order):

- Blue Coat ProxySG 200 with Blue Coat Web Filter option
- McAfee SWG 3300 v4.0 with Web Filtering Module
- SurfControl Web Filter for Microsoft ISA Server
- Trend InterScan Web Security Appliance with URL filtering engine
- Websense for ISA Server

We tested each product's ability to accurately block more than 1 million URLs from over 100,000 distinct domains organized into 27 categories of interest to enterprise customers.

In our testing, we found the Trend Micro InterScan Web Security Appliance (IWSA) was extremely effective at blocking content in a number of categories — particularly the sexually-explicit category and many of the categories within our Productivity and Recreation group. Trend also blocked URLs extracted from recent spam e-mails effectively. "Trend blocked the highest percentage of URLs in 12 of our 27 categories, placing first more than any other product."

We conducted final tests at the end of March, 2006. This document provides an abbreviated summary of our findings and describes our Q1 2006 English-language corpus and test methodology.

In this report, we focus exclusively on blocking effectiveness using products' URL filtering databases and did not enable real-time filtering, remote rating, or download scanning (sometimes used to detect malware based on the actual contents or signature of a file) on any products. Also, this report does not discuss or evaluate performance or scalability, product user interface, features, or functionality.

## Summary of Results

---

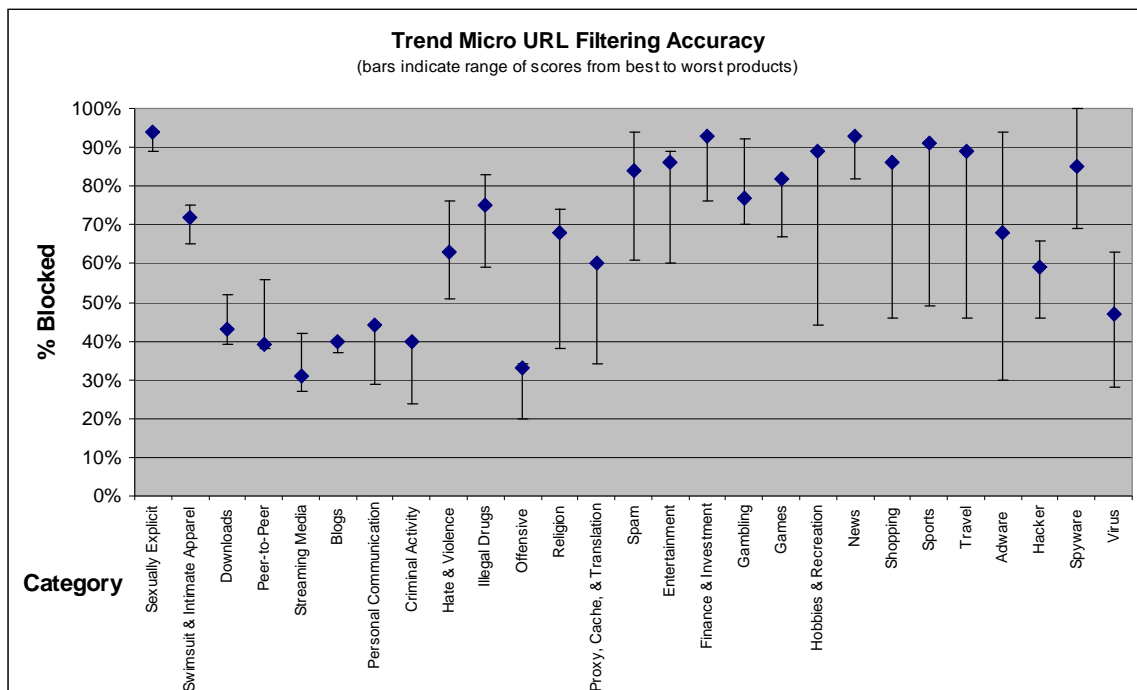
We tested the Trend InterScan Web Security Appliance (IWSA) with version 01.030.0000 of its URL filtering database. We compared its results against four competitive products using equally current databases. The products we tested included both gateway appliances and software (shown below in alphabetical order):

- Blue Coat ProxySG 200 with Blue Coat Web Filter option
- McAfee SWG 3300 v4.0 with Web Filtering Module
- SurfControl Web Filter for Microsoft ISA Server

- Trend InterScan Web Security Appliance with URL filtering engine
- Websense for ISA Server

In our testing, we found the IWSA was extremely effective at blocking content in a number of categories — particularly the sexually explicit category and many of the categories within our Productivity and Recreation group. Trend also blocked URLs extracted from recent spam e-mails effectively. "Trend blocked the highest percentage of URLs in 12 of our 27 categories, placing first more than any other product."

The following chart indicates the range in blocking effectiveness that we observed in each category, with Trend's position denoted by a blue diamond:



All products were highly effective at blocking sexually-explicit content. Other categories, however, posed more significant challenges. For example, most products struggled with our spyware, adware, virus, and hacker categories — confirming that URL filtering alone is generally insufficient to block a wide spectrum of security threats. Enterprises will necessarily choose to combine URL filtering with other approaches such as content scanning at the gateway, e-mail security, and managed desktop anti-virus and anti-spyware software solutions.

Our detailed, category-by-category analysis (for internal use only) also revealed systematic differences in the products' effectiveness and suggested opportunities for improvement. We noted particular differences in the way the products handle domains, hosts, and individual URLs within a domain or host.

## **An English-Language Corpus with One Million URLs**

---

We created our Q1 2006 English-language corpus to address the requirements of the enterprise market.

### **Size**

Our baseline English-language corpus contains more than 1,000,000 URLs from more than 100,000 unique domains. Specifically, the corpus we used for Q1 2006 testing contained 1,322,119 distinct URLs from 103,821 domains. (In this context, we define a "domain" as the top-most site-specific portion of a host name. For example, we consider "yahoo.com" or "google.co.uk" to be a domain, while we define "sports.yahoo.com" or "finance.yahoo.com" as hosts within the yahoo.com domain.)

### **Quality**

We performed extensive quality assurance to verify the characteristics of the URLs in our corpus. During the first quarter of 2006, we used a variety of automated mechanisms to acquire more than 2.4 million candidate URLs from more than 250,000 unique domains. We then used a variety of custom tools and human oversight to filter and classify these URLs into the final corpus used for testing.

We used a number of techniques to acquire candidate URLs: lists of top-ranked sites, category-specific directories, search engines, domain-specific knowledge, and custom-built Web crawlers that sought out particular types of content. We found that naive approaches to assembling a corpus, such as entering terms into search engines, tended to generate low-quality URLs that in many cases did not accurately fit a category, so we selected our approaches more carefully and performed a series of human and automated quality-assurance steps to ensure that our candidate URLs were classified appropriately.

We provide our clients samples of our corpus to allow independent verification of quality. Those samples, and the specific URLs cited for illustrative purposes in our report, comprise less than 0.2% of the total URLs in our corpus. Aside from that, we kept our corpus entirely independent and secret so that no vendor would receive an advantage.

To keep our corpus current, we update it on a quarterly basis to add new URLs and categories, to expire stale URLs, and to hone our distribution of sites.

### **Groups, Categories, and Category Mapping**

The categories and URL distribution in our English-language corpus have been chosen to address the requirements of large enterprises. For example, our corpus includes categories such as sexually explicit, illegal drugs, criminal activity, shopping, streaming media and viruses — categories that enterprises often choose to block or limit for employee groups

Since each vendor uses its own set of categories for classifying URLs, we defined our own uniquely defined set of categories, and then created category mappings to the categories defined by each product to ensure we used comparable blocking configurations for each product. We then performed preliminary tests to ensure that we had appropriate category mappings for each of the five products.

We have made efforts to construct our corpus from URLs that clearly belong in a particular category. For example, few would argue that [www.amazon.com](http://www.amazon.com) fits in the shopping category, [www.espn.com](http://www.espn.com) fits in sports, and [www.fool.com](http://www.fool.com) fits in the finance and investment category. However, pages that discuss shopping tips, the value of sports for children growing up, or an article on the economy may or may not fit in these categories depending on your perspective. Although we have included some of these "fringe" URLs to assess coverage breadth, our corpus is characterized by content that clearly fits its category.

## Language and Domains

Virtually all the URLs in our corpus are English-language Web pages. In the sexually explicit and swimsuit and intimate apparel categories, where the language of text on a page is not as important as the images, we permitted a small number of non-English pages.

The majority of URLs in our corpus come from domains associated with the United States, including .com, .net, and .org. However, we have also included a substantial number of domains from other English-speaking countries such as Australia, Canada, New Zealand, and the United Kingdom, as well as a smattering of other top-level domains to reflect the global nature of the Web.

## Redirects

About 55,000 of the URLs in our corpus (4% of the total) redirect to other destination pages. In our testing, products are credited with blocking a redirected page if the source or the target of the redirect is blocked by the product. This mimics the behavior a user would see in a corporate environment.

## Test Methodology

---

We configured each of the products in explicit-proxy mode, updated to current database versions, and ran our blocking tests simultaneously for all products against live sites on the Internet. URLs that resulted in an error were automatically queued for re-testing to counter the effects of transient outages on target sites; URLs that reported errors on the re-test pass as well were automatically discarded from analysis.

## Block by Group, Test by Category

For our testing, we configured each product to block an entire group (defined to contain similar categories). This choice allowed us to verify that blocking results

were not overly affected by slight differences in vendors' categories choices. For example, some vendors might place a "fishing" URL in the sports category while other vendors might place it in the "hobbies and recreation" category. In our testing, either decision would result in a product being credited with a successful block, but classifying the URL as something entirely unrelated, like sexually explicit, would not.

We then performed tests and analyzed results by category.

In order to measure the effectiveness of the URL database contained locally within the software or appliance, we did not allow products to use any type of real-time or remote rating for our tests. We also turned off all products' remote-logging features and blocked access to vendor sites at our firewall in order to ensure that our URL corpus was not leaked.

## About Trend Micro

---

Trend Micro, Inc. is a global leader in network antivirus and Internet content security products and services. The company is focused on providing customers with customized and comprehensive security strategies to manage the impacts of known and unknown threats. Trend Micro has offices in 25 countries and trades stock on Tokyo Stock Exchange and NASDAQ.

## About Cascadia Labs

---

Cascadia Labs is a Seattle, Washington-based technology consultancy. Our team has extensive experience evaluating technology products for enterprise, small-business, and consumer audiences with a strong focus in filtering, scanning, and security solutions. We believe in taking active responsibility for delivering superb results, providing transparency at all stages of the project, ensuring frequent and open communication, and mastering complexity by aiming first for simplicity.