

COMPARING INDUSTRY LEADING ANTI-SPAM SERVICES

Introduction

The following analysis summarizes the spam catch and false positive rates of the leading anti-spam vendors. Compiled by Opus One, an independent research firm, this report provides data to objectively compare the market's most popular anti-spam solutions.

All of the anti-spam solutions tested are in Gartner's 2008 "Leaders" or "Challengers" Magic Quadrant. In total, eight vendors were evaluated over the course of fifteen months. The only vendor mentioned by name is the evaluation leader: Cisco IronPort. The remaining vendor names have been obfuscated.

DEFINITION OF TERMS

Spam catch rate measures the number of spam messages caught, divided by the total number of spam messages received.

False positive rate measures the relative number of legitimate emails misclassified as spam divided by the total number of legitimate email messages received. Given that the low number of false positives messages were statistically insignificant, relative figures are used. The spam accuracy rate is one minus the false positive rate.

TEST METHODOLOGY

To ensure consistency and reliability, the third party research firm operated within the following parameters:

- 15-month-long analysis from October 1, 2007 to December 31, 2008;
- More than 10,000 messages were selected at random for testing each month;
- Messages were drawn from actual corporate production mail streams;
- Messages were received live and tested with less than a one-second delay;
- Tested products were acquired directly from the vendor or through normal distribution channels and were under an active support contracts;
- Tested products were "up to date" with publicly released software and signature updates;
- Messages were hand classified as "spam" and "not spam" to ensure data validity.
- Messages counted on a per-recipient basis
- Each of the tested products included the vendor-recommended reputation service in the results

While testing occurred in North America, message sources were global.

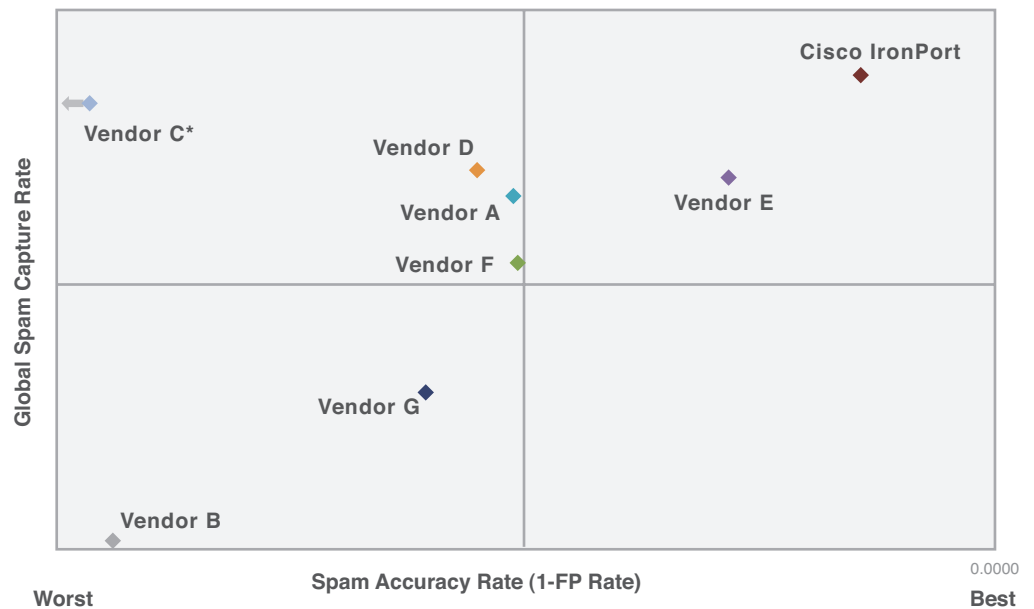
See the appendix at the conclusion of this report for further test methodology details.

TEST RESULTS

Cisco IronPort solution demonstrated the highest spam capture rate at 99.4 percent, and the most accurate rate of detection. With a spam capture rate of 99.1 percent, Vendor C placed second, but generated a false positive rate more than nine times higher than the Cisco IronPort solution.

Vendors A, D, E and F achieved spam capture rates of between 97.7 percent and 98.5 percent. While high, this rate means that the typical end-user of any of these services would receive approximately three times as much spam in their inbox (1.9 percent) as a Cisco IronPort anti-spam user (0.6 percent). Their false positive rates were also significantly higher, ranging from two- to four-times greater than the Cisco IronPort solution. Vendor B demonstrated the lowest spam catch rates. The results are displayed below:

Competitive Anti-Spam Efficacy



* Vendor C had a significantly lower spam accuracy rate that would place it off the chart, as indicated by the arrow.

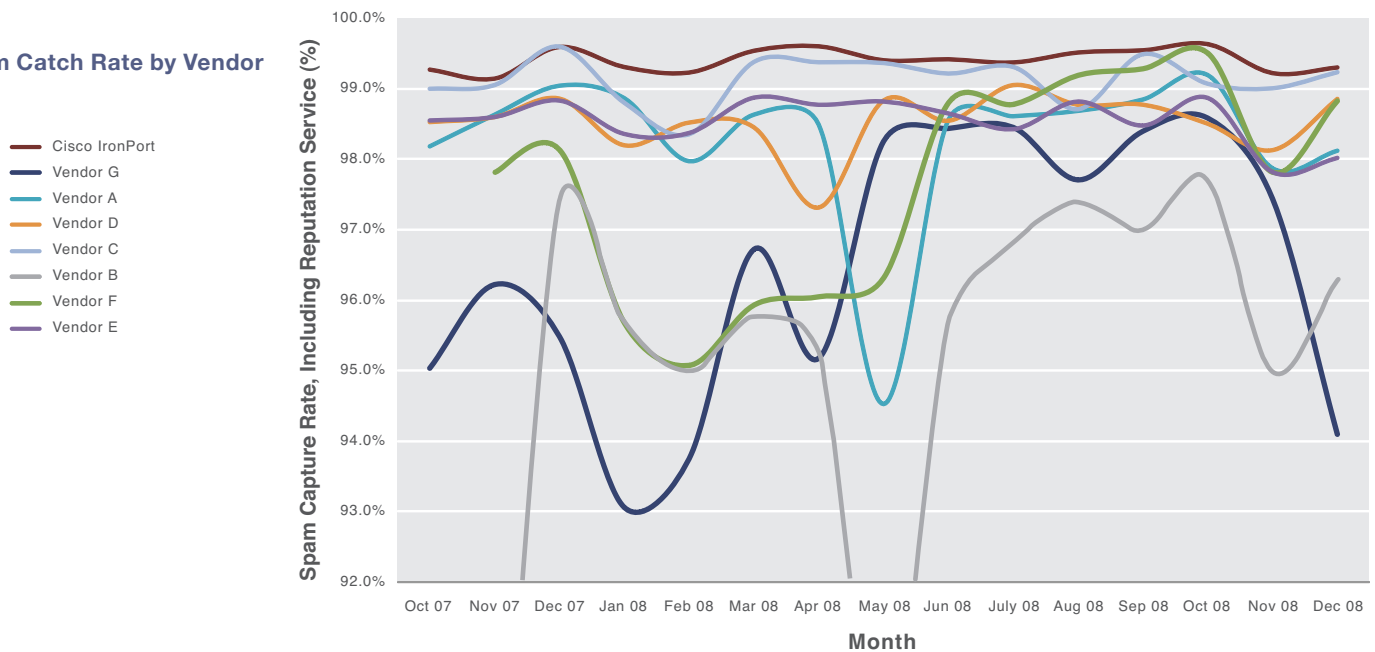
Spam Catch Rate Results

The spam catch rate has a direct impact on end-users' satisfaction and productivity. With the volume of spam continuing to multiply, even the slightest reduction in catch rates can have a major adverse effect. The average catch rates for anti-spam vendors over the 15-month period ending December 31, 2008 were as follows:

Vendor	Spam Catch Rate	Missed Spam Relative to Leader
Cisco IronPort	99.4%	N/A
Vendor C	99.1%	48%
Vendor D	98.5%	149%
Vendor E	98.5%	161%
Vendor A	98.3%	190%
Vendor F	97.7%	295%
Vendor G	96.5%	497%
Vendor B	95.1%	728%

Month by month spam catch rate results by vendor over the testing period are displayed in the graph below. Please note that the vertical axis ranges from 92 percent to 100 percent to better highlight the differences in the various solutions.

Spam Catch Rate by Vendor



False Positive Results

Because of the mission critical nature of email, it is essential that an enterprise's anti-spam solution deliver a low false positive rate. Messages incorrectly quarantined pose a serious loss of time and productivity for system administrators and end-users. Given the statistically insignificant number of false positive messages, relative results are used over the 15-month period ending December 31, 2008. The relative false positive rates are as follows:

Vendor	False Positive Relative to Leader
Cisco IronPort	N/A
Vendor E	206%
Vendor F	368%
Vendor A	372%
Vendor D	400%
Vendor G	440%
Vendor B	683%
Vendor C	938%

Summary

Given the essential role of email in the operations of modern enterprises, spam poses a serious threat to their success. When a spam message finds its way into a user's inbox or a legitimate message is incorrectly identified as spam and quarantined, there is an immediate impact on productivity. While performance of the solutions evaluated in this analysis may vary by only a few percentage points, it's important to recognize that this difference can translate into hundreds, if not thousands, of unwanted and potentially problematic messages infiltrating a network.

Over the years, much ground has been gained in the battle against spam. Nevertheless, the amount of spam in circulation continues to rise, demanding increasingly sophisticated and capable defense systems. The productivity of the global marketplace demands it.

Based on the results of Opus One's extensive tests, the Cisco IronPort Anti-Spam solution delivers the highest spam catch rate and the lowest false positive rate.

Appendix

TESTING METHODOLOGY

Anti-spam products were evaluated by installing them in a production mail stream environment. The test simultaneously feeds the same production stream to each product, recording the verdict (typically “spam,” “not spam,” or “suspected spam”) for later comparison.

Each product tested was acquired directly from the vendor or through normal distribution channels. Each product tested was under an active support contract, and was believed to be “up to date” with publicly released software and signature updates.

Where multiple versions were available from a vendor, the technical support team for each vendor was consulted to determine the “recommended” platform for use. To minimize confusion, products were not upgraded during the test cycle, although anti-spam and anti-spam engine updates were typically and automatically made by each product during the term of the test.

All systems were able to connect to the Internet for updates and DNS lookups. A firewall was placed between each product and the Internet to block inbound connections, while outbound connections were completely unrestricted on all ports. Each product was configured based on the product manufacturer’s recommended settings.

Where easily executed, multiple scenarios were used for a product, including a factory-default aggressive setting (“suspect spam”) and conservative setting (“certain spam”) based on the vendor’s recommendation. In cases where obviously inappropriate settings were included by default, these settings were changed to support the production mail stream. “Maximum message size” — to accommodate messages of varying sizes — was the most commonly changed setting.

The tests drew on the real corporate message stream because this message stream contains no artificial content and best represents the normal enterprise stream. No spurious spam or non-spam content was injected into the stream.

Each product was connected to the Internet to retrieve signature and software updates as often as recommended by the vendor. If vendor technical support teams recommend a shorter update cycle, this recommendation was implemented.

Because products were not receiving email directly from the Internet, the reputation service of each product had to be individually configured to support the multi-hop configuration. In cases where products were unable to handle a multi-hop configuration with reputation service, the reputation service results were gathered at the edge of the network and then re-combined with the anti-spam results after the test was completed.

For many products, this re-combination better illustrates the actual performance a network manager would see and significantly changes the test results from a test which does not incorporate reputation service results.

Once the messages were received, the third party research firm manually read through every single message, classifying it as “spam,” “not spam,” or “unknown.” Testers defined as “spam” the messages for which there was no conceivable business or personal relationship between sender and receiver and which were obviously bulk in nature. Mail messages that may not have been solicited, but which showed a clear business or personal relationship between sender and receiver, or were obviously a one-to-one message, even if unsolicited and unwanted, were classified as “not spam.” All mailing lists which have legitimate subscriptions were considered “not spam.”

Messages were classified as “unknown” if they were the result of virus double bounces, or if they could not be definitively categorized as “spam” or “not spam” based on content, or if they were so malformed that it could not be determined that they were spam, viruses, or corrupt software. All “unknown” messages were deleted from the data set, and do not factor into the result statistics.

Once the manual qualification of messages was completed, all results were placed in an SQL database. Queries were then run to create false positive and false negative lists. False positives for each product were individually evaluated and any errors in the original manual classification were fixed. Because the number of false negatives is typically much higher (up to 700 false negatives per product are not unusual), the third party research firms did not evaluate every false negative for each product. Instead, testers evaluated each false negative for at least two different products, and sampled the false negative results for all other products to identify any errors in the original classification. As a result, this revealed that some false negatives may have been inappropriately marked, although the actual percentages are likely to be significantly less than 1 percent. Once the data sets were determined to be within acceptable error rates, the databases were reloaded and the queries recreated.

Each anti-spam engine provides a verdict on messages. While this is often internally represented as a number, the verdict in most products is reduced to a categorization of each message as being “spam” or “not spam.” In many anti-spam products, a third category is included, typically called “suspected spam.”

In this test, products were configured at the factory-default settings, where possible, to have three verdicts (spam, suspect spam, and not spam). Where products have three verdicts, suspect spam is considered to be spam. As a result, suspect spam was included in the catch rate and relative false positive rate calculations.

Catch rate refers to the number of spam messages caught out of the total number of spam messages received. When spam is not caught, it is called a false negative.

- False negative means the test said “this was not spam,” and it was.
- False positive means the test said “this was spam,” and it wasn’t.

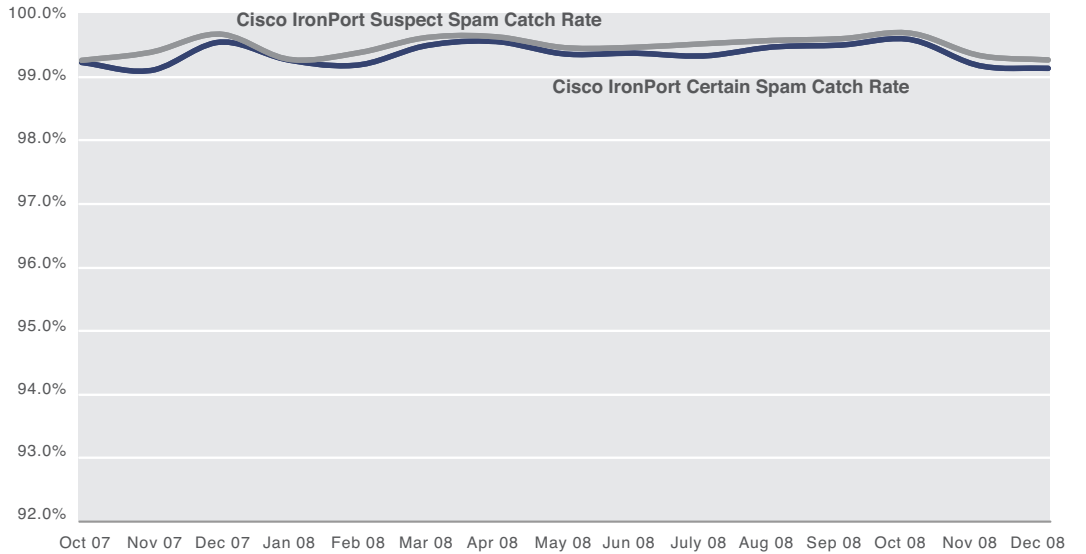
While a false positive means that good mail might have gotten lost — raise alarms for system administrators — a false negative can be just as troublesome when it arrives in the end-user’s inbox.

INDIVIDUAL ANTI-SPAM VENDOR PERFORMANCE

Cisco IronPort

Cisco IronPort achieved an average of 99.4 percent spam catch rate. This assumes that suspect spam is treated as spam. If suspect spam was allowed to pass through to the end-user, Cisco IronPort’s spam capture rate would have been 99.3 percent.

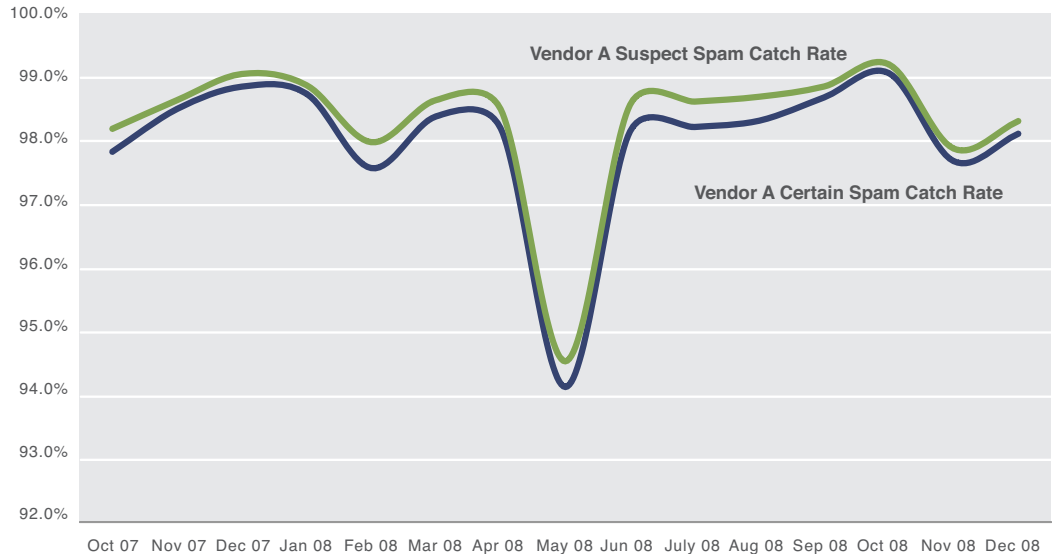
**CISCO IRONPORT
Monthly Group Catch Rate**



Vendor A

Vendor A achieved an average of 98.3 percent spam catch rate. This assumes that suspect spam is treated as spam. If instead, the suspect spam category was let through, then the spam capture rate would have been 98.0 percent, with a corresponding lower false positive rate.

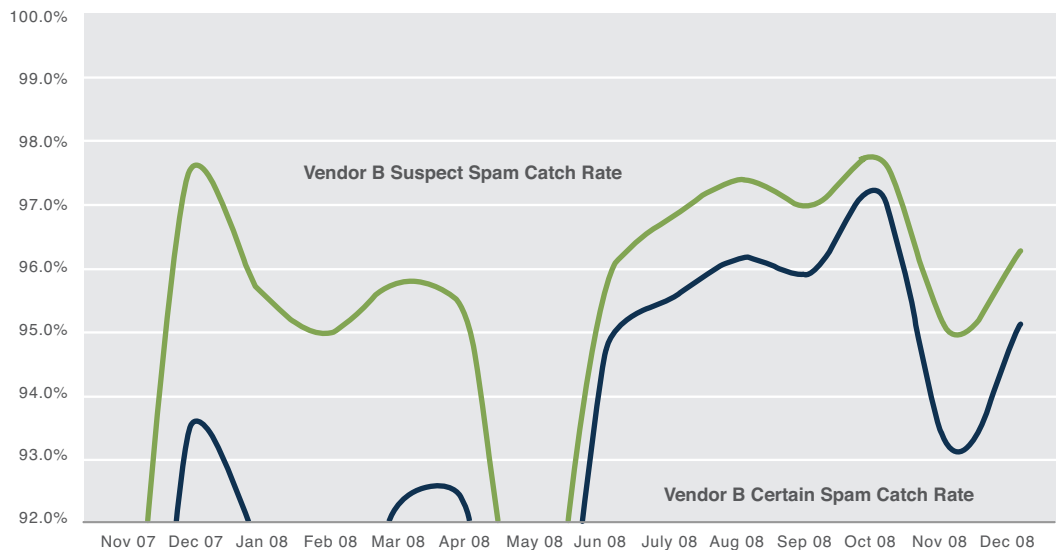
VENDOR A
Monthly Group Catch Rate



Vendor B

Vendor B achieved an average 95.1 percent spam catch rate. This assumes that suspect spam is treated as spam. If suspect spam was allowed to pass through to the end-user, Vendor B's spam capture rate would have been 92.2 percent.

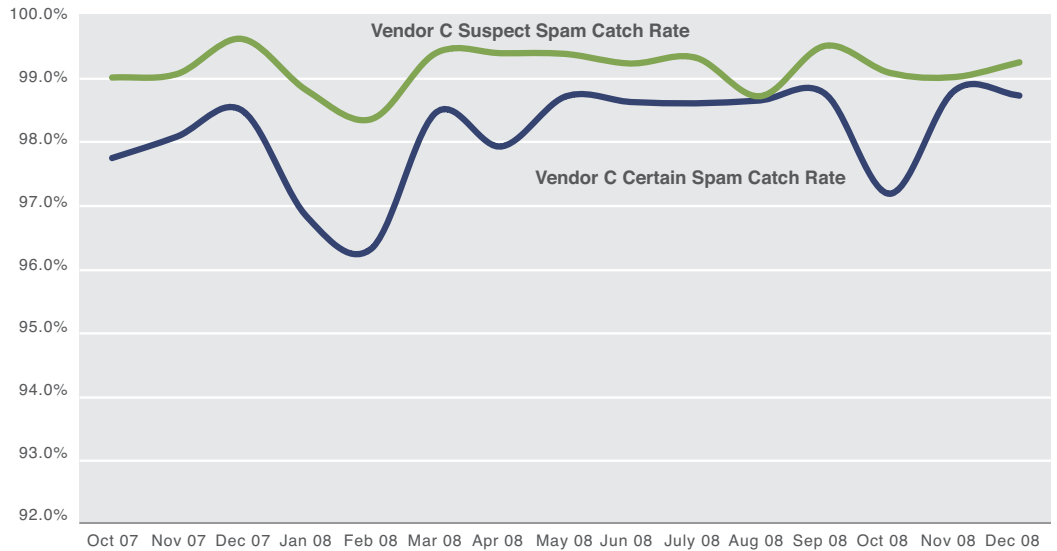
VENDOR B
Monthly Group Catch Rate



Vendor C

Vendor C achieved an average of 99.1 percent spam catch rate. This assumes that suspect spam is treated as spam. If instead, the suspect spam category was let through to the user, then the spam capture rate would have been 98.1 percent.

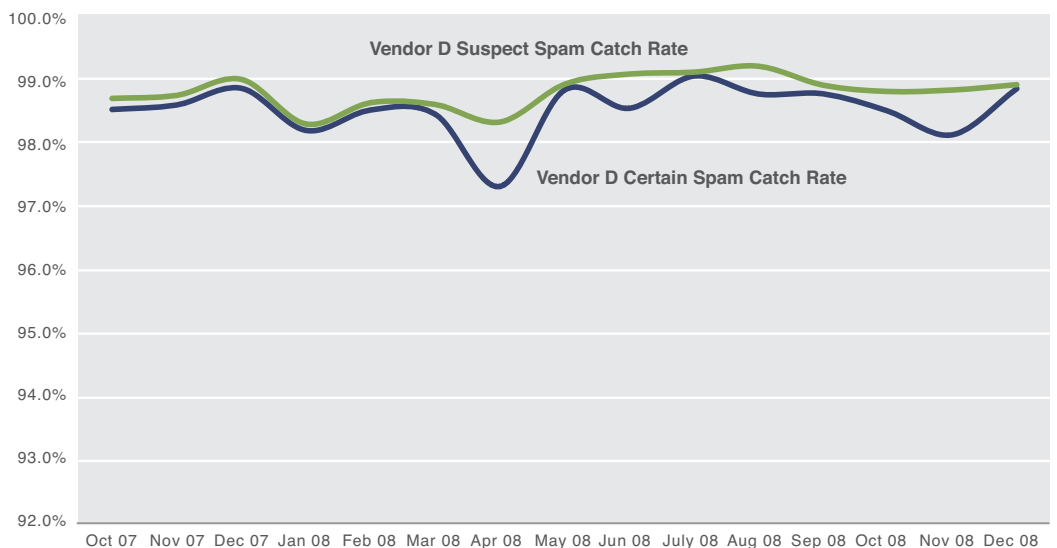
VENDOR C
Monthly Group Catch Rate



Vendor D

Vendor D achieved an average of 98.5 percent spam catch rate. This assumes that suspect spam is treated as spam. If instead, the spam settings were set more aggressively, then the spam capture rate would have been 98.8 percent, with a corresponding higher false positive rate.

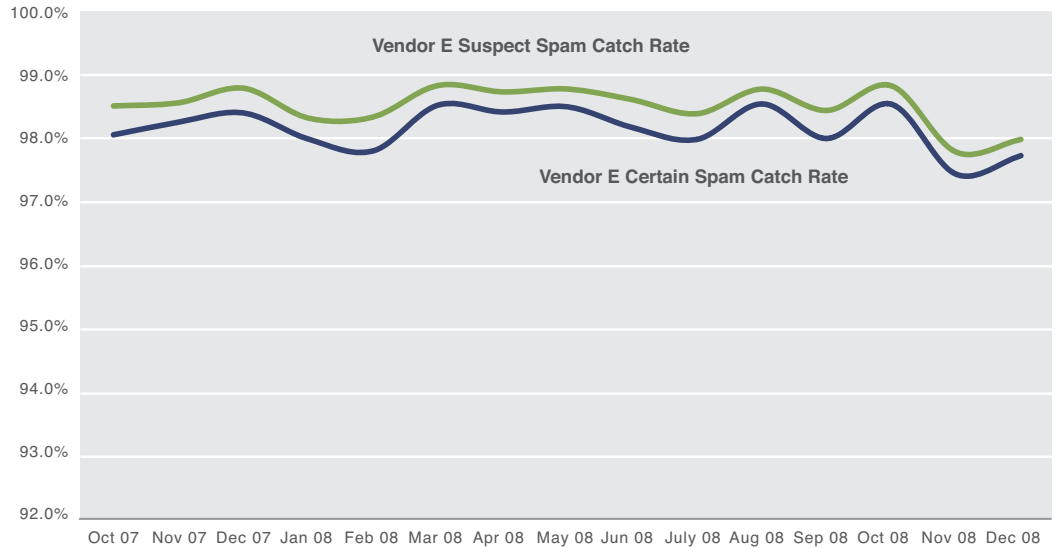
VENDOR D
Monthly Group Catch Rate



Vendor E

Vendor E achieved an average of 98.5 percent spam catch rate. This assumes that suspect spam is treated as spam. If instead, the suspect spam category was let through, then the spam capture rate would have been 98.2 percent, with a corresponding lower false positive rate.

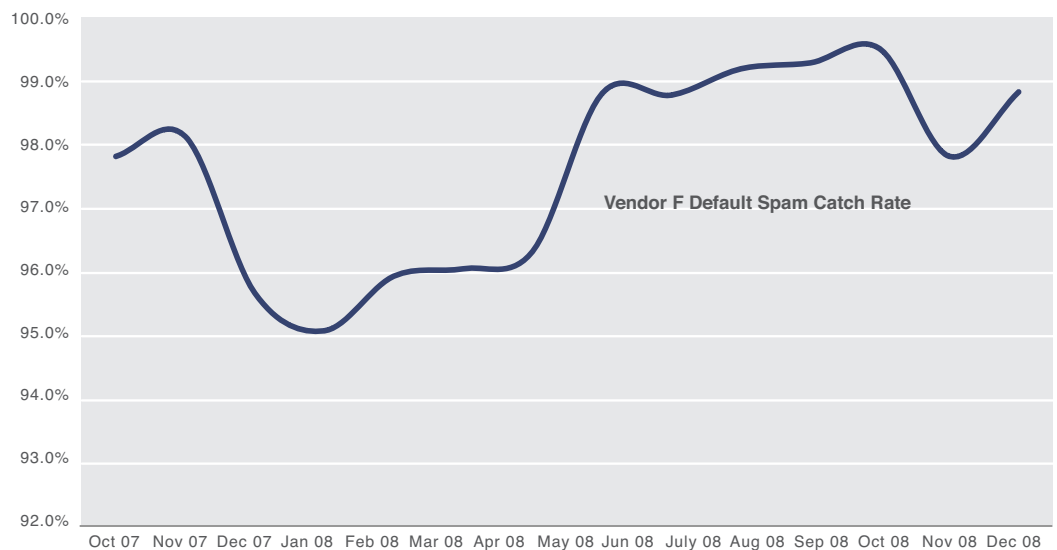
VENDOR E
Monthly Group Catch Rate



Vendor F

Vendor F achieved an average 97.7 percent spam catch rate with its default settings.

VENDOR F
Monthly Group Catch Rate



Vendor G

Vendor G achieved an average of 98.2 percent spam catch rate. This assumes that suspect spam is treated as spam. If suspect spam was allowed to pass through to the end-user, Vendor G's spam capture rate would have been 96.5 percent, albeit at a significantly higher false positive rate.

VENDOR G
Monthly Group Catch Rate

