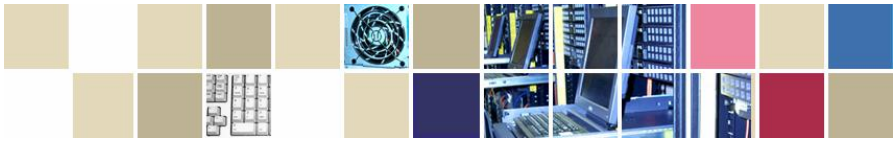

VeriTest
Anti-Spam Benchmark Service

Program Description

February 2005

VeriTest would like to acknowledge and thank the following organizations and individuals for their participation and contributions in designing this service:

Active State
Barracuda Networks
Brightmail
Cloudmark
Computer Mail Services
Death2Spam
Digiportal
eWeek Magazine
Finjan Software
Frontbridge
MailFrontier
Message Labs
Messaging Architects
National Spam Mail Abuse Association
Network World Magazine
Opus One Consulting
Postini
Proofpoint
PC Magazine
Sophos
St. Bernard
SurfControl
Symantec
Trend Micro
Yahoo!
Ziff Davis Media

To learn about subscription options for the Anti-Spam Benchmark Service, please call VeriTest at 415-546-6885, x236, or email antispam@veritest.com.

Program Summary

Public interest in fighting spam is at an all-time high -- the subject of legislation, keynote speeches, and countless blogs and newspaper columns. To help businesses combat spam, many anti-spam technologies have been introduced into the marketplace, including server-based software products, hardware appliances, and hosted filtering services. Developers of these solutions require proof points that allow them to differentiate their products from competitive offerings, and to substantiate marketing claims.

To meet this need, VeriTest, the leading independent product test laboratory, offers a regularly scheduled, publicly quotable, and technically rigorous test program to measure the effectiveness of anti-spam technologies. The *VeriTest Anti-Spam Benchmark Service* has been designed in conjunction with input from anti-spam product developers, trade publications and industry associations. Our comprehensive methodology levels the playing field and stands up to expert scrutiny.

Subscribers to the *VeriTest Anti-Spam Benchmark Service* receive detailed quarterly reports on the performance of their technologies, ranked against competitive solutions, plus advance access and review of *VeriTest Anti-Spam Benchmark* results disclosed publicly by VeriTest, personalized technical support, and membership in our Anti-Spam Test Methodology Advisory Council.

The *VeriTest Anti-Spam Benchmark Service* delivers invaluable insight into anti-spam technology effectiveness, and is ideal for developer engineering groups seeking to improve product performance, as well as for developer marketing groups seeking to establish credible proof points.

Subscriber Benefits

- **Quarterly competitive analysis reports comparing performance of leading anti-spam technologies**
- **Publication Rights to performance data**
- **Test methodology that stands up to industry scrutiny**
- **Advance access and review of benchmark results**
- **Analysis of e-mail corpuses**
- **Membership in the VeriTest Anti-Spam Test Methodology Advisory Council**

Solution Test Categories

Anti-spam solutions target a variety of market sectors. The *VeriTest Anti-Spam Benchmark Service* measures technologies intended for use by Small/Medium Businesses and Enterprise-scale Businesses, including both hardware and software products installed at the mail gateways, and hosted solutions. Because the mechanisms underlying anti-spam solutions vary widely, we have designed our test methodology to be "technology-agnostic", and to work equally well with products that employ content filtering rules, DNS/RBL lookup, other techniques, and any combination thereof, so that we can accurately measure anti-spam effectiveness.

VeriTest also offers anti-spam effectiveness testing for Inbox Providers and ISPs. For more information on test services this category, please refer to the companion document *Program Description: Anti-Spam Benchmark Services for ISPs and InBox Providers*.

Test Methodology

Effectiveness Measurements

While there are many important factors to consider when evaluating anti-spam technologies, such as ease of deployment and ease of use, we focus on two quantifiable metrics:

- **Spam Blocking Effectiveness**—how well the solution correctly identifies and blocks spam from users' inboxes.
- **"False Positive" Rate**—how well the solution correctly identifies legitimate (non-spam, or "ham") email and allows it through to users' inboxes

We take these measurements in two separate scenarios. The first is the "out-of-box" configuration, using pre-set solution defaults, which provides a common testing baseline. The second test scenario utilizes an "optimized" configuration, where each vendor has the opportunity to tune their respective solutions to the particulars of the VeriTest email stream.

The Issues

At first glance, a methodology for measuring anti-spam technologies seems simple enough: send streams of e-mail, both spam and non-spam, through each of the technologies under test, and measure how effectively each of the products or services block spam, as well as permits legitimate e-mail to pass through to a users' inboxes. On a closer look, however, it is clear that conducting fair evaluations of anti-spam technologies requires a balance of many important factors and criteria:

- How spam is acquired and distributed
- How legitimate mail ("ham") is acquired and distributed
- Ensuring that the test process itself does not skew results (e.g., introducing modifications to message headers), and is compatible with a broad range of anti-spam methodologies
- Selecting e-mail sample sizes and/or test period durations
- Determining ratios of spam to "ham"
- Evaluating "out-of-box", "optimized", and "trained" configurations

Sourcing Spam

At first thought, acquiring spam suitable for test purposes seems effortless. On further reflection, though, the challenge of sourcing spam in all its variety (from the mail-order drug ads to industry-specific phishing campaigns) becomes apparent. Previous tests of anti-spam products have utilized a variety of techniques to obtain samples of spam suitable for testing, including the use of spam corpuses from archives like those at *spamarchive.org,* specially seeded honeypot accounts, sampling business email streams, among others. Each method has its own pros and cons.

Archived Spam: Using a spam archive or email corpus facilitates the testing of each anti-spam technology with an identical set of hand-selected messages, including industry-specific spam. However, there are a number of serious issues with using archived spam. Perhaps most importantly, testing with archived spam is useful only for anti-spam technologies that rely exclusively on content analysis, and excludes any real-time filtering techniques. Secondly, archived spam rapidly ages. Spam campaigners constantly change their techniques for delivering their messages, just as those in the anti-spam technology community continually evolve their blocking methods. Lastly, using archived spam inherently requires message header modifications.

Corporate mail sampling: This technique generally involves sampling from streams of live incoming email to an established business entity. While this approach meets the real-world standard, it introduces thorny issues of privacy, non-disclosure, and reproducibility.

Honeypot accounts: Honeypot accounts are email addresses seeded in a way specifically to attract spam; for example, they can be embedded into public web sites, and propagated on newsgroups. Honeypot accounts have the advantage that they facilitate test methods utilizing live, unadulterated mail streams. Disadvantages include the difficulty in rapidly bringing new honeypot accounts up to volume (2,000+ messages per day), and also challenges in seeding multiple accounts in precisely the same way.

Spam By The Numbers

55 **The percentage of companies that do not use a spam filter because they are afraid that legitimate messages will be trapped.** (*source: The Radicati Group)*

19 **The percentage of opt-in, legitimate email from e-newsletter publishers that never reaches subscribers, due to over-active spam filters** *(source: Return Path)*

400 **The number of domain names used by a typical spammer** *(source: National Spam Mail Abuse Association)*

The VeriTest approach: For the *VeriTest Anti-Spam Technology Benchmark Service*, we have opted to source live, unadulterated "spam from the wild" from legacy domains with a long history on the Internet. These accounts provide high-volume streams of spam that have been analyzed by experts to ensure that they are highly characteristic mail typically received at SMB and enterprise mail servers.

Sourcing Legitimate Mail

To conduct an accurate test, it is important to ensure that legitimate messages are represented in all their variety:

Personal correspondence

Work-related correspondence

Opt-in newsletters, bulletins, e-mail alerts

Content: work-related, personal, and industry-specific terms; for example, legitimate email correspondence to/from pharmaceutical companies may include frequent references to brand-name drugs

Originating domains: Enterprise and SMB legitimate email tends to originate from a clique of domains

Message Body Length: messages range in size from a few words to a few pages

Characteristics: text, HTML, embedded links, embedded graphics, attachments

TO: Distribution: one-to-one personal correspondence, one-to-few "workgroup" messages, one-to-100's broadcasts, forwards, etc.

CC:, Subject:, etc. fields: messages need to represent a variety of attributes in frequently used fields

The VeriTest approach: In order to source legitimate email in the quantities required for testing, VeriTest uses a two-fold approach. First, VeriTest uses a network of 20 or more domains in conjunction with a "ham generator" system that produces a steady stream of diverse, legitimate email from a clique of originating addresses. The "ham generator" ensures that each anti-spam technology being tested receives an identical set of legitimate mail representing a wide variety of mail types and contents. Subject lines and message bodies are sampled from business and personal correspondence, newsgroup exchanges, and other sources.

To represent mail of the newsletter variety, test accounts are opted-in to a number of daily newsletters and bulletins such as *Today's Headlines* from NYTimes.com. In addition to news sites, we will also opt-in to sites devoted to providing e-mail alerts, product information, and catalogs on a wide variety of topics. We will also register the test email addresses at a number of mailing lists that provide specific information on certain topics of interest. These include mailing lists providing information of security, operating systems and other computer related topics.

Distributing Email

One of the most difficult problems to resolve when testing anti-spam solutions is that of distributing email to multiple technologies under test. In the ideal scenario, a live email stream is multiplexed simultaneously to each technology-under-test; in that way, each technology is tested with exactly the same set of mail. Unfortunately, all forms of simultaneously relaying e-mail introduce modifications to the e-mail message header. These artificially introduced modifications, although minor, can reduce the effectiveness of certain anti-spam technology approaches. For example, many anti-spam technologies rely on analyses of mail headers, profiles of originating IP addresses and real-time lookup of DNS addresses. Any test process that alters message headers or requires that spam be stored and later relayed will prevent these technologies from operating in their optimum capacity.

The VeriTest approach: In order to preserve the real-time, real-world nature of our email streams, VeriTest uses a *sequential* method for distributing email. Simply put, the blended spam/ham mail stream is directed to the first solution-under-test on Day One of the test period; the second solution-under-test on Day Two of the test period, the third solution on Day Three, and so on, over the course of the test period. Each solution is tested with a minimum of three complete 24-hour sessions over the test period. In this manner, each solution-under-test is exercised with a live, unreflected, unadulterated stream of mail, importantly, preserving the IP address of the last email hop.

Since the CAN-SPAM Act of 2003 went into effect, spam volume has increased an average of 2% a month.
Source: Brightmail

While the corpus of spam is not identical from Day One to Day Two, the characteristics of the spam stream are essentially the same. Given the high volume of mail processed during the test, differences in individual messages become less relevant. VeriTest takes steps to ensure that the rotation method used for each product levels the playing field in terms of the calendar day; for example, to mitigate the fact that spam campaigns are launched with a higher frequency on weekend days, products are rotated so that each is tested on no more than one Saturday or Sunday.

We also include hosted solutions in our tests. To test these, we change the MX records of our mail stream domains in order to send the messages to the hosted solution. We wait 24 hours before starting the test in order to allow the MX record change to propagate through the Internet. When the 24-hour test period is complete, we change the MX record back to the IP address of our internal test bed. Again, we wait 24 hours before starting another test in order to allow the MX record change to propagate through the Internet. Hosted solutions agree to forward all filtered email messages to VeriTest so that we can know the exact number of messages filtered.

NOTE: VeriTest is currently evaluating a sequential rotation pattern of increased frequency, utilizing a load balancing system.

Email Sample Size and Spam/Ham Ratios

We evaluate each anti-spam solution under test with a test set of approximately 13,000 messages, received over three, 24-hour test periods. We aim for an email stream comprised of 60% spam and 40% legitimate email. We base those percentages on various media reports on the subject of real-world spam volumes. Of the legitimate messages in the mail stream, approximately 20% are opt-in newsletters and bulletins. Since the volume and composition of our mail stream is dependent on real spam campaigns, which are somewhat unpredictable, our breakdown of spam and legitimate mail over the course of an entire test period will typically vary from these target percentages by as much as 10%.

"Out-of-Box" and "Tuned" Configurations

Not all anti-spam solutions start out on an equal footing. Some may offer reasonably good performance right out of the box, while others require significant tuning and "training." An additional issue is that there are typically a wide range of sensitivity settings for each anti-spam solution. Typically, the spam identification policies can be adjusted to meet the needs of the individual user. For example, a business may have a very lower tolerance for "false positives", that is, incorrectly quarantined legitimate mail. In that situation, the blocking policies may be relaxed by IT administrators somewhat so that potentially more spam will get through to Inboxes, but the likelihood of false positives is reduced.

Time wasted handling spam costs American businesses nearly $22 billion a year. A recent study by the University of Maryland found that over 75% of Internet users receive junk e-mail daily. The average number of spam messages per day is 18.5, and the average time spent per day deleting them is 2.8 minutes.
Source: University of Maryland

Ideally, each solution-under-test should be configured with identical sensitivity settings. However, as each product has its own unique method for adjusting sensitivity, it is challenging to create identical configurations.

The VeriTest Approach: Whenever possible, we test each anti-spam technology in two configurations. The first configuration is the "out-of-box" configuration, using pre-set solution defaults. Tests using the default filtering settings that are factory-shipped with each solution provide a common baseline.

The second test scenario utilizes an "optimized" configuration, where each vendor has the opportunity to tune their respective solutions to the particulars of the VeriTest email stream. Vendors have the option of either tuning their solutions

remotely, or sending test engineers to our Research Triangle Park, NC facility in order to install and optimize their solutions.

Test results from optimized configurations are compared to default out-of-box configurations. In future planned enhancements of our test methodology, we will add capability to evaluate configurations that support product adaptability and learning.

NOTE: In some cases, due to technical considerations, only one configuration ("out-of-box" or optimized) may be tested.

Results Analysis, Judging and Reporting

Classifying Email

We classify each e-mail message processed by the technologies as one or more of the following types:

Spam
- Messages not originating from our ham generator, and
- Messages not directly opted-into by VeriTest, and
- Messages including but is not limited to the following types:
  - Unsolicited commercial emails (advertisements for pornography, insurance, medication, etc.)
  - Unsolicited mass e-mail that is deceptive in its subject line and hides the sender's identity or seeks to commit fraud
  - Phishing messages (asking the reader to divulge information)
  - Chain messages (asking the reader to forward this message to others)
  - One-to-many messages (using a mailing list, USENET or other networked communications facility as if it were a broadcast medium)
  - Messages containing misleading URLs (URL name does not match destination), and
- Messages not otherwise classified as "Unknown" (see below)

Legitimate mail
- Messages created and sent by VeriTest (also referred to as "ham")
- Newsletters and updates opted-into by VeriTest

Unknown (*scrubbed from the email corpus and excluded from results*)
- Messages written in a non-English language, and where we are unable to determine if the message is spam
- Messages with an ambiguous opt-in status
- Personal correspondence not sent by VeriTest (one-to-one mail or one-to-few mail)
- Corrupted messages, malformed or blank headers

False Positive
- Newsletters and updates opted-into by VeriTest that are improperly blocked as spam
- Ham messages improperly blocked as spam

Missed Spam

- All messages considered spam that are not properly blocked or identified as spam

IMPORTANT NOTE: Messages categorized or tagged by anti-spam technologies as "Probable/Possible/Likely Spam" are not considered to be blocked.

Scoring

We base our scoring scale on the belief that an ideal anti-spam solution excels in blocking spam effectively without over-blocking legitimate email. At the conclusion of a solution's three-day test, we measure the following for the tested solution:

- Percentage of spam messages blocked—calculated as the ratio of the number of spam messages correctly blocked to the total number of spam messages processed.
- False-positive percentage—calculated as the ratio of incorrectly blocked legitimate messages to the total number of legitimate mails sent.

We award between 2 and 5 points for each of these measurements, according to the following scale:

Points	Blocking Rate	False Positive Rate
5	95-100%	0-.5%
4	90-95%	.5-1.0%
3	85-90%	1.0-1.5%
2	<85%	>1.5%

We give a 40% weight to the points we award for the percentage of spam messages blocked and a 60% weight to the points we award for the false-positive percentage. We combine the two weighted raw scores, round the result to the nearest half-digit, and then express it as a number of stars (one point = one star) for the overall rating.

We weight the score for the percentage of spam messages blocked slightly lower (40%) than the score for the false-positive percentage (60 percent), reflecting the serious negative consequences of mislabeling legitimate messages.

The following figure shows a sample scoring calculation. A solution with a blocking rate of 92 percent and a false positive rate of 1.2 percent would earn an overall score of 3.4. Rounding 3.4 to the nearest half, this solution would earn three and a half stars.

	Blocking rate			False positive rate			Overall score
	Percentage	Raw score	Weighted score (raw x .4)	Percentage	Raw score	Weighted score (raw x .6)	(Weighted blocking + weighted false positive)
Solution A	92.00%	4	1.6	1.20%	3	1.8	**3.4**

Running the Test

Network Description

Software/Hardware Solutions

Before testing begins, we set up each solution on our internal network. We assign an internal IP address to each solution and keep them connected throughout the entire test duration.

We use a single Gateway server to initially accept the mail stream and then translate the external IP address of the stream into any one of the different internal IP addresses given to each solution. We change the mail stream's destination IP address every 24 hours in order to begin testing a different solution. Using Network Address Translation and a firewall, we leave each solution continuously connected to the Internet throughout the entire test so it can retrieve updates as regularly as required, but remain protected from direct connections from the outside. See the network diagram below for a graphical depiction of the test bed configuration.

Once the solution under test begins receiving the mail stream, it begins classifying messages and the test period is underway. The solution sends the messages it has received and classified to a downstream server running Sendmail (version 8.12.8, config V10/Berkeley) and an IMAP server. After reaching the IMAP server, the messages are downloaded by a PC running Outlook Express 6.0 so we can sort them into their correct classifications for test reporting purposes (Spam, Missed Spam, False Positive, and Legitimate).

VeriTest installs each anti-spam software solution on a separate Dell Power Edge 350 server, configured with an 850 MHz processor and 256MB of RAM. For the Gateway server, we use a Compaq Presario 5900z PC configured with an 800MHz processor and 256M RAM. The server running Sendmail and IMAP is a Micron PC Millennia configured with a 400MHz processor and 256MB of RAM. The final destination for the mail is a Dell Precision 610 PC configured with a 550MHz processor and 512MB of RAM, running Outlook Express and the Windows Server 2003 operating system. Finally, we establish the Ethernet connections between machines using an Extreme Networks Summit 48 switch.

Hosted Solutions

Hosted solutions are unable to operate within this network configuration and, therefore, special steps are required to ensure that we test them in a similar manner. Hosted services require us to redirect the entire mail stream to their off-site servers. We do this by changing the MX record of the domain used for the mail stream so that it points to the off-site servers. Then we wait for 24 hours before starting the test, in order to give all external DNS servers an opportunity to update their records to point to the new location. Once the test starts, the hosted service scans the messages and sends them back to the external address of our Gateway server on our test bed. From this point, we change the destination IP address to the internal IP address of the Sendmail server, The PC running Outlook Express then retrieves the mail with the help of the IMAP server for final tabulation. When the 24-hour test period is complete, we change the MX record back to the external IP

address of the Gateway on our test bed. We wait another 24 hours before starting the next test period so changes can propagate throughout all external DNS servers.

If we are unable to get the required data from a solution using this method, we will make slight modifications or create scripts so we can test the solution fairly.

All testing is coordinated by VeriTest's performance testing lab in Research Triangle Park, North Carolina. Program participants have the option of providing a system platform and mail server configuration. All products under test will have their most recent spam definitions/policies installed as of the start of the test.

Price and Schedule

Subscription Option	*Price*
Annual Subscription (four quarterly tests)	**$75,000**
Single Test	**$22,500**

As an annual subscriber to VeriTest's *Anti-Spam Benchmark Service,* you receive:

- **Detailed, custom reports** on the effectiveness of your anti-spam solution, ranked against other anti-spam solutions tested, *every quarter*
- Annual membership to our **Anti-Spam Test Methodology Advisory Council**
- **Access to sample email streams** prior to, and email corpuses after, each quarterly test cycle
- **Interim test results** and **15-day advance access** and review of any *VeriTest Anti-Spam Benchmark* results disclosed publicly by VeriTest
- **Custom technical support** by VeriTest analysts (up to 40 hours annually)
- **25% discount** on VeriTest published rates for private anti-spam related testing

Test periods are scheduled quarterly throughout the year, according to the following schedule. To be included as a measured web site during a scheduled test period, VeriTest must be in receipt of a signed Agreement, Order Form, and purchase order on or before the applicable Subscription Deadline.

Appendix A: Results Usage Guidelines

Publishable Data & Results

VeriTest grants *Anti-Spam Benchmark Service* subscribers the rights to disclose test results, with the following guidelines:

1. Use only VeriTest-approved metrics from the anti-spam benchmark.
2. Clearly cite the specific metric you are referencing.
3. You may publicize your product's rating and the Industry Average rating for each of the measured metrics.
4. Results must indicate the period during which measurements were made, and may be disclosed no later six months after the close of that measurement period.
5. All data must be presented fairly, accurately, and objectively.

Prior review: Test participants can review preliminary versions of quarterly final reports prior to release, in order to review and contribute a *Comments* page.

Trade magazines: VeriTest may post subsets of the test results at the VeriTest Web site and other media outlets such as trade magazines.

Disclosure Guidelines

1. When disclosing any VeriTest data, customer agrees to clearly attribute VeriTest as the source of the measurements. Customer agrees to follow VeriTest's terms and conditions of use of VeriTest logos, trademarks, and registered trademarks.
2. Make the proposed release or other material referencing VeriTest available to VeriTest for review and approval no less than 48 hours before publication. VeriTest will review the material and respond within 24 hours. Please send your proposed material to louann_millar@veritest.com.
3. Make available to VeriTest a copy of the original material(s) one week after publication.

Compliance

VeriTest may in its sole discretion apply any one or more of the following sanctions in the event of failure to comply:

1. Provide a written warning to customer with a notice of violation. Customer shall have 10 days to cure the violation or submit a plan acceptable to VeriTest for cure of the violation.
2. Require retraction and publication of correct information by customer in the same media as originally published. The retraction is to be approved by VeriTest and published within 15 days of VeriTest' request for retraction.
3. VeriTest may comment to the press on customer's publication.

4. VeriTest may temporarily suspend customer's publication rights under the agreement.
5. VeriTest may cancel customer's publication rights under the agreement.

Once you've received your anti-spam benchmark report from VeriTest, you have an opportunity to market your results. Citing results from your reports can provide independent proof of your marketing claims, as well as showing that you care enough about the service your customers receive to invest in testing. At VeriTest we carefully monitor marketing claims because so much of the value in our reports is based on the fact that we're an independent and impartial service. Our goal is to ensure that our customers are able to get as much value from marketing their data as possible, while upholding the standards and impartiality that make the benchmark of value to you and the market.

Contacting VeriTest

If you have questions or concerns regarding the publication rights, please direct them to Louann Millar at Louann_millar@veritest.com.

Some Common Questions about Results Usage

Q. Can we claim that VeriTest says our service has the "best" or "top ranked" anti-spam solution?
A. We don't test all the solutions available, so it wouldn't be accurate for a customer to claim that their service was the "best". In some situations, however, we might approve such a statement if the solution provider making the claim was willing to specify which competitors they were comparing the results to.

Q. What's the process for getting a VeriTest citation approved?
A. Please email your request to Louann Millar at: louann_millar@veritest.com. We will respond to your request within 24 hours.

Q. What does VeriTest do after we have released VeriTest data about our service?
A. Our policy is to confirm the information in your press release and answer questions about our testing methodology. We do not offer further information about your product or service and generally do not give any data about the other product or services measured unless the reporter refers also to their press releases.

Q. But you contract with magazines to provide them data -- how does this work?
A. A small number of respected publications contract with us to use our data in their industry survey articles. They aggregate and weight the data based on their own (subjective) criteria. We retain the right to review their use of the data for accuracy.

Q. What about special studies that VeriTest performs?

A. We address marketing rules for our special tests a case-by-case basis. We do, however, require that if a test party wants to quote results from a special test report, they must make the report publicly available on our website.

Q. If I make sure I follow the rules, do I really have to run everything past VeriTest?

A. Yes. This enables us to make sure that all providers honor a common set of guidelines. This maintains the marketing value of the service and allows us to better support their marketing efforts.

Q. Can I promote my anti-spam test results for more than one quarter?

A. Yes if you show your ratings for consecutive quarters and the most recent published quarter is less than one year old.