White Paper Analysis:

Mailshell SpamCatcher
Accuracy vs. Error

August 2002

Overview

This paper analyzes the accuracy of Mailshell's SpamCatcher anti-spam system.

Mailshell consistently and accurately blocks up to 99% of spam. Mailshell's filters are highly accurate regardless of whether the system -- or the end user -- has encountered spam with those attributes or format ever before. The system combines Mailshell's SpamCatcher software and The SpamCatcher Network, a two-pronged approach to identifying and blocking spam:

1. **SpamCatcher Software**: Mailshell analyzes millions of emails per day and creates rules to identify and blacklist spam immediately. Complex algorithms determine which among hundreds of rules are applied to compute the probability that each piece of email is spam.

   and

2. **The SpamCatcher Network**, a database of millions of email 'fingerprints': Mailshell creates a fingerprint ID for every email message and compares it to existing fingerprints in its network in order to compute the probability that each piece of email is spam. The database of fingerprints, rules and algorithms are updated in real-time.

Mailshell establishes and maintains a 'reputation' for each sender of email through its system. Messages from bulk senders with a good reputation – or those that individual users explicitly approve for delivery – will be delivered. Most others are considered spam.

Is Blocking Spam Really So Hard?

Blocking spam is not hard. Setting basic filters to accept email only from a finite list of approved senders, deleting virtually everything else, would eliminate most spam. It would also, however, delete legitimate email as well.

As much as Internet users complain about receiving spam, their tolerance for spam is greater than their tolerance for 'false positives', i.e. good emails that are erroneously blocked as spam. Most users and system administrators would agree that receiving several pieces of spam per day is a severe nuisance, but not receiving even a single important email because it's been blocked by spam filters is wholly unacceptable.

The challenge, therefore, is to create spam filters that are sufficiently strong to accurately identify and eliminate spam without creating any 'false positives'. Mailshell has accomplished this, and it does so with greater accuracy than any other product.

Case Study: Analysis of Mailshell Spam Filtering

Mailshell conducted an analysis in behalf of a global ISP[i] to determine the efficacy and accuracy of its spam filters on real-world email messages. This test was conducted by analyzing 4,203,702 real email messages sent through

the ISP's email servers. The Mailshell system, which processes more than one million messages per CPU per day, analyzed all 4.2 million messages within one day in March of 2002.

Test One: The **Accuracy vs. Error Test** confirmed that Mailshell's filtering system demonstrated extremely high accuracy[ii] while maintaining an extremely low percentage of 'false positives'.

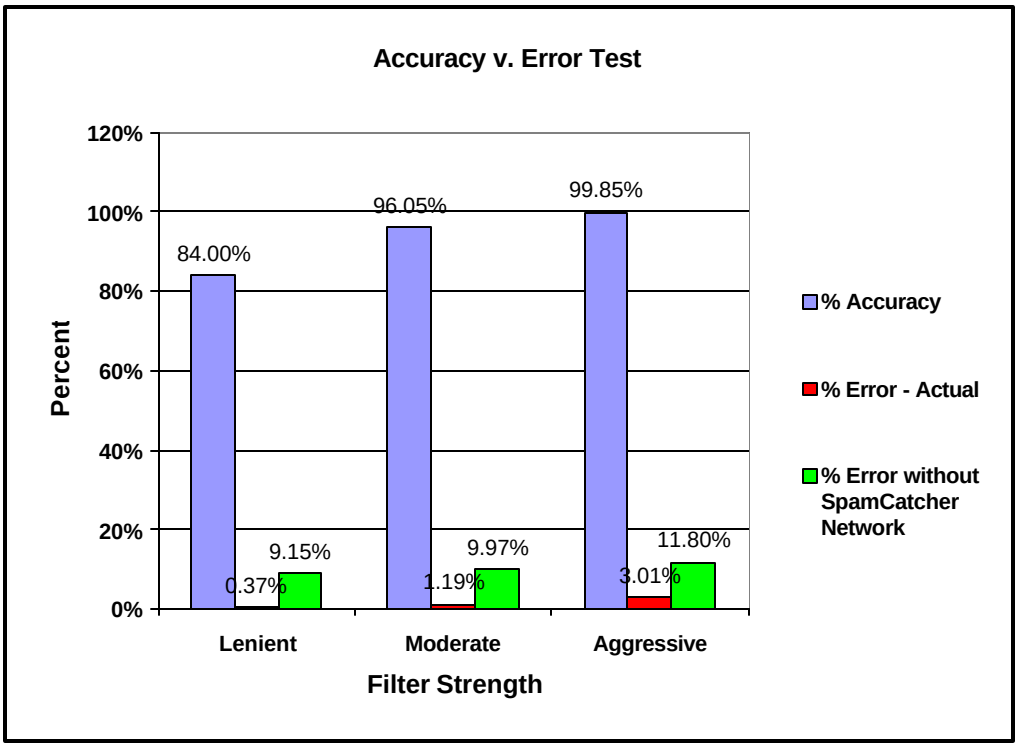
Filter Strength[iii]	Accuracy	% Error—personal email	% Error – bulk email	% Error if SpamCatcher Network is Disabled
Lenient	84.00%	0.00%	0.37%	9.15%
Moderate	96.05%	0.00%	1.19%	9.97%
Aggressive	99.85%	0.00%	3.01%	11.80%

1. Mailshell's accuracy, calculated using the formula *(1-|total pieces of spam – pieces of spam caught|/total messages)* ranged from 85% to more than 99% depending on filter strength.

2. Mailshell's system did not incorrectly block any personal email, i.e. emails sent from one user to another, as spam.

3. Mailshell's system was nearly perfect differentiating legitimate bulk email (primarily newsletters and group postings) from illegitimate bulk mail (spam).[iv] The percent error ranged from 0.37% for lenient filtering to 3.01% for aggressive filtering.[v]

    - It is difficult to distinguish spam from legitimate opt-in newsletters without knowing the recipient's specific wishes and prior opt-in decisions. Both spam and legitimate commercial email may carry unsubscribe instructions, use similar headers and carry similar content (especially the ads often found within newsletters).



- Since this test did not provide access to users' previous emails or preferences, it was impossible to determine if a given user had previously opted-in to a specific newsletter.

- This percent error contained only email sent in bulk, primarily newsletters. An error rate of 0.3% means that 0.3% of the legitimate newsletters a user receives might be blocked as spam. For example, assuming a given user receives three newsletters each day, this means that one copy of one individual newsletter might get blocked every 100 days.

4. Mailshell's SpamCatcher Network allows the system to filter spam more aggressively than other systems without risk of deleting 'good' emails that are sent in bulk.

    - Mailshell determined that 8.8% of the messages came from known legitimate senders in the SpamCatcher Network database. This constitutes 18.6% of the messages that were sent in bulk.

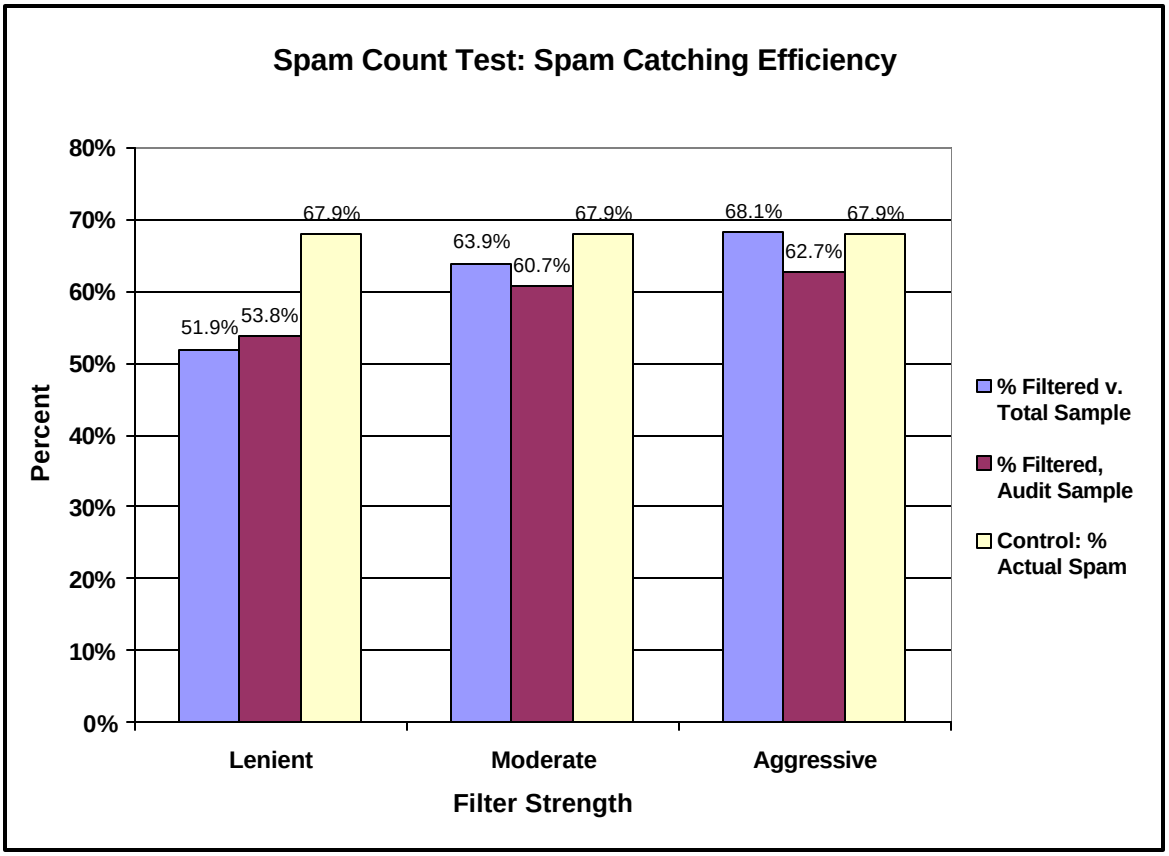
5. Without the SpamCatcher Network, all of these messages would either be:

    a) deleted as spam since the messages carry many of the characteristics of spam. This would increase the % error from 0.37% to 9.15% on the Lenient filter setting and from 3.01% to 11.80% on the Aggresssive filter setting or;

    b) delivered along with actual spam messages that bear similar traits, thereby dramatically reducing any filtering system's accuracy. The massive SpamCatcher Network database is the only way to differentiate between the two scenarios.

Test Two: The **Spam Count Test** determined the percentage of spam blocked at Mailshell's *Lenient*, *Moderate* and *Aggressive* filter strengths. The test was conducted using Mailshell's standard setup, as deployed on the company's consumer web site.

- The Lenient setting identified 2,181,926 of 4,203.702 messages as spam (51.9%).

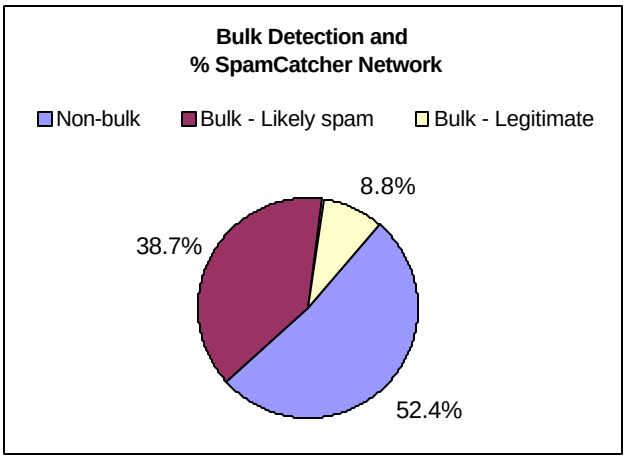


- The Moderate setting identified 2,688,264 of the messages as spam (63.9%).
- The Aggressive setting identified 2,860,497 of the messages as spam (68.1%)

1. These results were **audited** by visually inspecting 3,287 of those messages to determine if they were actually spam or valid emails.
2. The audit determined that 2,233 (67.9%) of the 3,287 messages were actually spam (see '*Control: % Actual Spam*' in accompanying chart above).
3. Mailshell ran each of the 3,287 messages through its spam filters to determine the percentage of the 3,287 the system would identify as spam.
   - The Lenient setting identified 1,771 of the 3,287 messages as spam (53.8%).
   - The Moderate setting identified 1,996 of the 3,287 messages as spam (60.7%).
   - The Aggressive setting identified 2,061 of the 3,287 messages as spam (62.7%).
4. The participating ISP indicated that this audited group was representative of the overall sample of more than 4.2 million messages and that the 67.9% determined to be spam is an accurate proxy for the percent of the overall sample that is likely to be spam.

Test Three: The **SpamCatcher Network Test** compared all the messages sent in bulk with Mailshell's SpamCatcher Network, a database of millions of email fingerprint ID's.[vi]. Rather than rely exclusively on spam filters to guess which commercial mailings are legitimate – the physical attributes of spam and many newsletter-style mailings are extremely similar – Mailshell has developed a proprietary White List to specifically identify legitimate commercial mail.

1. Mailshell determined that 47.5% of the messages analyzed were sent in bulk.
2. 8.8% of the messages analyzed came from senders in our SpamCatcher Network database. This represents 18.6% of the messages sent in bulk.
3. The top 100 senders in the SpamCatcher Network accounted for 68.6% of the messages identified by the Network
   - N.B.: "Non-bulk" includes both spam and legitimate personal email.



About Mailshell

Mailshell's anti-spam functionality provides superior accuracy and reliability. It's also easily configured to identify and permit or block known senders based on their relevance and trustworthiness within a specific email network. Given the varying definitions of spam, Mailshell is the most adaptable, as well as accurate, anti-spam solution available.

Mailshell was founded in January of 1999 and its spam-blocking solution is the product of more than three years of development, testing and refinement. Major national press, including *Forbes*, the *Associated Press*, *The Wall Street Journal*, *Investors Business Daily, USA Today*, *CNET* and *Business 2.0* have highlighted the system as a breakthrough in email management. In addition to providing anti-spam services to third party companies, Mailshell continues to deploy improvements to its spam blocking system on its consumer web service with nearly one million registered users.

# # #

For more information:
Eytan Urbas
partners@mailshell.com
415.348.8728

Endnotes:

[i] The ISP operates in three continents worldwide and is a publicly traded company. A mutually binding non-disclosure agreement prohibits Mailshell from revealing this company's name. Mailshell certifies that the results of this case study are accurate and explicitly consistent with the results of the actual test.

[ii] Accuracy calculated using the formula *(1-|total pieces of spam – pieces of spam caught|/total messages)*

[iii] Mailshell's consumer service offers five pre-configured filter strengths: none, lenient, moderate, aggressive and exclusive. This test used the standard Lenient, Moderate and Aggressive filters available on the consumer web service. Mailshell's enterprise solutions allow the customer to adjust filter strength using a scale of 0-100.

[iv] Mailshell also ran our engine against the 2233 messages audited as spam and found:

- Lenient setting correctly identified 1771 (79%) of the 2233 spam messages as spam and missed 462 of the spam messages. We correctly identified 1042 (99%) of the 1054 non-spam messages as non-spam and misidentified 12 of the non-spam messages.
- Moderate setting correctly identified 1996 (89%) of the 2233 spam messages as spam and missed 237 of the spam messages. We correctly identified 1015 (99%) of the 1054 non-spam messages as non-spam and misidentified 39 of the non-spam messages.
- Aggressive setting correctly identified 2061 (92%) of the 2233 spam messages as spam and missed 162 of the spam messages. We correctly identified 945 (97%) of the 1054 non-spam messages as non-spam and misidentified 99 of the non-spam messages.

[v] Percent error calculated using the formula ((*incorrectly filtered messages) – (total messages))*100*

[vi] For the SpamCatcher Network test, Mailshell analyzed a subset of 838,013 messages chosen at random by the ISP from among the 4.2 million messages in the overall test sample.