



**astaro**  
internet security

[Astaro Firewall](#)

## Network Security Whitepaper

### **The Cocktail Approach to Spam Protection**

Version: 1.01

Release date: January 5, 2004

Author: Al Cooley, BSEE: WPI, MBA: University of Michigan, Advanced Studies in Computer Engineering: Boston University



# TABLE OF CONTENTS

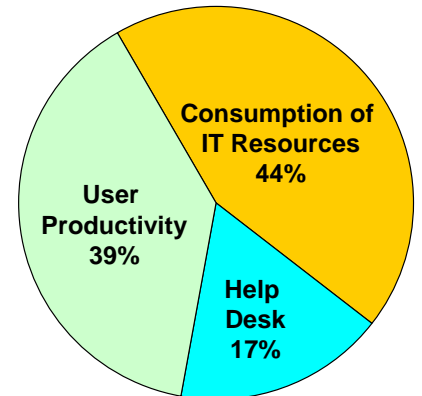
<b>INTRODUCTION.....</b>	<b>3</b>
<b>THE SPAM PHENOMENA.....</b>	<b>3</b>
<b>NON-TECHNICAL APPROACHES TO DEALING WITH SPAM.....</b>	<b>4</b>
<b>SPAM FILTERING SOLUTIONS.....</b>	<b>5</b>
<b>APPLYING THE COCKTAIL APPROACH.....</b>	<b>5</b>
THREE TECHNIQUES—AND THEIR SHORTCOMINGS.....	6
COMBINING TESTS.....	7
SCORING MECHANISM AND THRESHOLD.....	8
ADDITIONAL ANTI-SPAM TECHNIQUES.....	8
OPTIONS FOR HANDLING SUSPICIOUS MESSAGES.....	8
<b>ASTARO SECURITY LINUX.....</b>	<b>9</b>
<b>CONCLUSION.....</b>	<b>9</b>
APPENDIX A: REFERENCES.....	10
APPENDIX B: FURTHER READING.....	10

## Introduction

Spam, the common term for unsolicited commercial email messages sent in bulk, has rapidly become the scourge of computer users everywhere. Spam wastes employees' time, hogs disk space, consumes networking bandwidth and eats up IT resources. Jupiter Media Metrix estimates that each piece of spam costs the average company \$1 in lost productivity. Ferris Research estimates the average employee wastes \$4,000<sup>1</sup> a year dealing with spam.

There are a number of technical and non-technical responses to spam. Unfortunately, no single technique can solve the problem.

This white paper describes the concept of using a "cocktail" approach to spam prevention by combining a variety of spam detection tests into a single algorithm. We also discuss how this approach is utilized in the SpamAssassin open source project (which forms the core of the anti-spam software in the Astaro Security Linux complete network security solution).



**The Cost Elements of Spam To Recipients**

## The Spam Phenomena

The driving force in the phenomenal traffic growth is economics.

Email addresses are readily culled from public web sites using robots, from mail servers using dictionary attacks (messages are sent using common names and those that work are recorded) and from harvesting web sites. Harvesters are willing to provide email addresses at no cost if the spammer is willing to share revenues.

The infrastructure costs no more than a few thousand dollars, based on standard PCs, design tools, and bulk emailing and spamware software programs from commercial vendors. ISPs, email service providers, and bulk email services offer email privileges for \$100 or less, while bulk emailing can be had for \$300 per million messages.

With essentially no per unit sending costs, even low yielding<sup>2</sup> mailing lists targeting unlikely buyers are economically attractive. Business Week estimates that spammers' yields are only around .005%, meaning they send 1,000,000 messages to get 50 responses. Yet the economics of the situation dictate that with no variable message costs and a fixed overhead, the spammer maximizes profits by maximizing the number of messages sent. These economics are driving a 56% annual growth in spam. As a result it is estimated that 46% of unfiltered email today is spam.

---

<sup>1</sup> Assuming \$35 hourly cost, 33% spam and 115 hours/year managing the inbox

<sup>2</sup> Yield is the percent of addressees responding to the offer

## Non-Technical Approaches to Dealing with Spam

Many organizations initially respond to spam by using manual methods or by counting on others to solve the problem. Unfortunately, despite seemingly hopeful signs like the recent U.S. national anti-spam law, these approaches offer little prospect of success.

These ineffective approaches include:

- **Manual disposition:** Many organizations have yet to tackle the spam issue, relying on users to manually wade through spam and dispose of it. Clearly the economics outlined above indicate that this approach will be increasingly untenable, and that a substantial ROI can be achieved by implementing a more automated solution.
- **Legal remedies:** Some countries and U.S. states have passed laws restricting or outlawing spam. Unfortunately legal remedies are unlikely to stem the flow of spam in the short term. The recent U.S. national anti-spam law is actually predicted to increase spam since it formally legalizes spam, codifies an opt-out approach rather than an opt-in approach, and overrides the stronger measures implemented by some states. Furthermore, spam is an international business, and can be easily initiated from a remote location with supportive laws should legislation restricting transmission in an existing location be passed.
- **Service providers:** In response to the deluge of spam, some ISPs, as well as specialized providers, have begun providing fee-based anti-spam services. These organizations essentially provide an outsourced electronic filtering function. Although outsourcing is frequently a popular IT alternative, many organizations are hesitant to entertain it for email filtering since it exposes sensitive corporate information to regular third-party review. Furthermore, the relative costs of filtering outsourcing, as well as response time issues for such a critical organizational function are also detractions.

## Spam Filtering Solutions

Spam filtering solutions are software packages that scan incoming email messages, detect unsolicited messages, and block those messages or take other actions to minimize or eliminate the negative impact of spam. These products have a modest up-front cost, can be managed by typical IT staff members, and have been shown to yield a high ROI.

However, the primary technical challenge in developing spam-filtering software is creating a detection algorithm that identifies the highest percentage of spam messages possible (high accuracy), without incorrectly classifying legitimate email as spam (low false positives).

Unfortunately, first-generation approaches have generally been ineffective, because each individual method has been circumvented by alert spammers. In addition, using a single approach tends to result in administrators either “setting the bar too low” and letting in too many unwanted messages, or “setting the bar too high” and blocking legitimate email.

In practice, the only effective method has been to employ a “cocktail approach” that combines and leverages a variety of spam detection algorithms.

## Applying the Cocktail Approach

In this section of the white paper, we outline the approach to spam protection utilized in the SpamAssassin open source project ([www.spamassassin.org](http://www.spamassassin.org)). It is estimated that over 30 million end-users utilize the software, either by downloading it directly or by using a version that is incorporated into a commercial security software product such as Astaro Security Linux.

## Four Techniques—and Their Shortcomings

Four of the most common techniques for spam filtering are outlined below, each with its own challenge:

Technique	Description	Challenges
Real-time blacklists	Addresses of spammers are aggregated into a database by organizations focused on this mission (e.g. <a href="http://www.mail-abuse.org">www.mail-abuse.org</a> ). Email messages are checked against one or more of these databases in real-time, blocking messages from addresses known to be used by spammers.	Spammers are constantly changing their delivery mechanism to avoid detection. Though these databases can be quite helpful, they will never be entirely up-to-date or accurate.
Sender address verification	To eliminate spam sent from false addresses (a common ploy), the sending domain of incoming messages can be dynamically verified and the sending mail server can also be contacted to verify the existence of the sending address.	Legitimate mailers sometimes use non-verifiable addresses for mass mailings such as newsletters. Although this practice is decreasing, care must be taken to avoid the risk of false positives.
Header analysis	The header of an email has a variety of fields containing information on the sender, routing, recipient and subject that can be either examined for suspicious behavior.	Spammers work hard to cover their tracks, delivering mail through legitimate but insecure servers, utilizing specialized mailing tools which minimize footprints, falsifying headers, etc.
Body analysis	The body contains the content of the spam offer, and provides ample content that can be examined for characteristics typical of a marketing offer.	Spammers try to confuse software detection mechanisms using techniques such as interspersing garbage characters that will not be visually seen (e.g. make characters the same color as the background), or will be filtered out by the human mind (e.g. obvious typos).

No individual technique is perfect, or even good enough to yield adequate results by itself. Each detects a different attribute, but is also subject to different evasion techniques. Consequentially the most effective approach is to use a combination of techniques. Like mixing a good cocktail, the trick is picking the right components in the right proportions.

## Combining Tests

Astaro's Spam Protection feature essentially provides a framework which combines a wide variety of different spam detection tests into a single detection algorithm. Tests have been developed and evolve over time to respond to changing spammer styles and evasion techniques.

The basic concept is to use multiple tests to detect "tricks" designed to evade a single anti-spam technique.

For example, it would be perfectly normal to see an email with "Life Insurance" in the subject line (this might be coming from the corporate HR department). However, it is unlikely that a message from the HR Department would also include a tracking ID or an opt-out message. Identifying messages with "Life Insurance" in the subject line and a tracking ID or an opt-out message would allow the spam filter to stop unsolicited messages without blocking legitimate email.

Examples of particular tests (or groups of tests) that can be employed include:

- "From:" doesn't include a real name or has a suspicious structure
- "From:" address has lots of numbers and is from a big ISP
- "To:" or "Reply To:" is empty
- Message ID has the format used by common spam tools
- "Subject:" has common spam attributes such as wording in ALL CAPS, an advertising tag mandated by state or country law (eg. ADV:), or includes common topics such as "Viagra", "As Seen", "Free", "Guaranteed", "Life Insurance", "Nigeria", etc.
- A large number of recipients are detected in "To" or "cc: " lines
- Message includes a list removal feature
- Message includes a tracking ID
- Message text is disguised using base64 encoding to attempt to avoid electronic detection
- Message includes lots of yelling (wording in ALL CAPS) and/or attempts are made to hide yelling by inserting hidden HTML comments inside the shouting
- HTML font which is the same color as the background is used to try to disguise spam attributes from electronic detection without confusing the reader
- HTML links with common labels such as "push here" or "click here" are included
- Message contains wording very frequently found in spam such as "Full Refund", "Call Now", "Offer", "Limited Time", "Mortgages", "All Natural", "No Obligation", "No Credit Check", "Lower Monthly Payments", "Hair Loss", "Stop Snoring", "Up To x% More" or "Loose Weight"

## *Scoring Mechanism and Threshold*

Each test included in the Spam Protection framework is assigned a different weight through a scoring mechanism based upon empirical evidence of its effectiveness. Tests are combined to arrive at a final score which indicates the overall probability that a particular message is spam.

The administrator can set the threshold at which messages are treated as spam, providing a simple mean of customizing the aggressiveness of the Spam Protection framework. A tighter setting catches more spam, but increases the probability of false positives. A looser setting reduces false positives but will tend to allow more spam to pass.

## *Additional Anti-Spam Techniques*

The multiple tests described above can be combined with other spam detection features to maximize effectiveness. These features work together to overcome any shortcomings inherent in individual features, complementing each other to form a much stronger spam detection solution than any single technique. Such features include:

- Sender address verification
- Real-time blacklist
- Local blacklist: The administrator can create a local list of known spammers to block. Wild card support simplifies blocking offending domains or users.
- Local whitelist: To avoid identifying important legitimate senders as spammers, a whitelist of known good senders can be created, once again with wild card support to simplify administration.

## *Options for Handling Suspicious Messages*

In some implementations of the SpamAssassin software, such as that provided in Astaro Security Linux, administrators are given control over the handling of email that is identified as spam.

Suspicious messages can be:

- Automatically deleted
- Quarantined for review by the administrator
- Returned to the sender with an explanation of why it was returned, or
- Forwarded to the sender with a special header that can be used by the receiving mail system to deal with the offending message as desired

This gives administrators the data to help fine tune the aggressiveness of their anti-spam strategy. In addition, should a legitimate email address accidentally be identified as a spammer, the administrator can override the spam filters by adding the sender to the local "whitelist."

## Astaro Security Linux

The spam protection capabilities described in this white paper are included at no extra cost in Astaro Security Linux.

Astaro Security Linux is a complete, integrated suite of network security software that protects against the major threats of connecting to the Internet. It is installed on standard PC hardware placed at the point where the Internet connects to an organization's network.

Functions provided include:

- Spam protection
- Firewall
- Virus protection
- Content filtering
- Virtual Private Networking
- Wireless protection
- Performance management

By providing a single integrated security solution Astaro slashes the total cost of ownership and administrative burden while minimizing the potential for security gaps. There is no need to learn, install, integrate, manage and update multiple point solutions. Additionally, a single secure Internet-based software update service keeps all elements of Astaro Security Linux up to date, maximizing security, minimizing costs and eliminating conflicting updates.

## Conclusion

Spam is one of the most visible Internet threats facing organizations today, requiring employees all the way from the CEO to interns to wade through a growing deluge of distracting email on a daily basis. Furthermore spam is expensive, with costs projected to continue to escalate at an astounding rate. To protect themselves, organizations should implement a spam protection solution that employs multiple filtering techniques and integrates seamlessly with other network security solutions.

Astaro provides an ideal spam protection solution that is proven, highly accurate and extremely easy to manage. Being integrated into a comprehensive security solution it also provides greater security, lower total cost of ownership and simplified operation. A free 30-day evaluation version can be downloaded at [www.astaro.com](http://www.astaro.com) to verify operation in your organization's particular environment. To facilitate the evaluation process, a free 90-minute web-based workshop is available which steps users through the process of configuring and using Astaro Security Linux.

## *Appendix A: References*

1. "Needed Now: Laws To Can Spam", Business Week, Mike France, 9/26/02
2. Ferris Research, Spam Control, 2003
3. Federal Trade Commission, "Remove Me Results",  
<http://www.ftc.gov/bcp/online/edcams/spam/index.html>
4. "Internet Level Spam Detection And SpamAssassin 2.5", Matt Sargent,  
[www.spamassassin.org](http://www.spamassassin.org)

## *Appendix B: Further Reading*

InfoWorld Magazine reviews Astaro Security Linux and reports "the most polished and easy to use Web-based management system we've seen to date." ([Read the whole article as PDF](#))

"A stylish-looking appliance that addresses the major internet-related security concerns of the small or medium business," says SC Magazine. ([Read the SC Magazine article as PDF](#))

Astaro named winner of the [product excellence award](#) at LinuxWorld.

A free 30-day fully-featured evaluation version of Astaro Security Linux can be downloaded at <http://astaro.com/php/download.php?lang=gb>