# An Overview of Spam Blocking Techniques

Recent analyst estimates indicate that over 60 percent of the world's email is unsolicited email, or "spam." Spam is no longer just a simple annoyance. Spam has now become a significant security issue and a massive drain on financial resources. In fact, this deluge of spam costs corporations an estimated $20 billion each year in lost productivity.

Today there are a large number of solutions designed to help eliminate the spam problem. These solutions use different techniques for analyzing email and determining if it is indeed spam. Because spam is constantly changing, the most effective spam blocking solutions contain more than one of these techniques to help ensure that all spam, and only spam, is blocked.

The following is an overview of different spam blocking techniques.

## SPAM BLOCKING TECHNIQUES

### Word Filters

Word filters are a simplistic yet effective way to block the majority of obvious spam. Word filters simply identify any email that contains certain key words, such as "Viagra," that are commonly found in spam. Because spammers often work to circumvent word filters by purposely misspelling words, word filters need to be regularly updated with variations of the key words. For example, "Viagra" may be purposely misspelled as "V1agra," so the word filter must be updated to contain both "Viagra" and "V1agra."

In some circumstances, word filters run the risk of creating false positives. For example, a legitimate email containing the word "Viagra" that is intended for a medical researcher, physician or pharmacist may be inadvertently blocked.

Overall, word filters can be an effective spam blocking technique if they are constantly updated with new key words and phrases, as well as their unique misspellings.

### Rule-based Scoring Systems

Rule-based scoring systems are a more sophisticated spam blocking technique than word filters. These systems, also known as artificial intelligence (AI) systems, are similar to word filters in that they also check for key words. However, whereas word filters simply just block emails that contain key words, rule-based scoring systems use rules to analyze emails and assign points to each key word it finds.

For example, an email that contains the word "DISCOUNT" in all capital letters might receive +2 points. An email that has the phrase "click here" might receive +1 point. The higher the score, the greater probability the email is spam. If an email reaches a certain score or threshold, it is then classified as spam. Large quantities of spam and legitimate email are used to determine the appropriate scores for each of the rules in rule-based scoring systems.

### Spam History

Spam is a form of abuse of the Simple Mail Transfer Protocol (SMTP), which is implemented in email systems on the basis of RFC 524. First proposed in 1973, RFC 524 was developed during a time when computer security was not a significant concern. As such, RFC 524 is not a very secure command set, making it and SMTP susceptible to abuse.

Most spam-making tools exploit the security holes in SMTP. They do this by forging email headers, disguising sender addresses, and hiding the sending system, such that it becomes difficult or even impossible to identify the true sender.

To address some of SMTP's security holes, enhancement protocols to the venerable SMTP have been proposed. Most of these enhancement protocols involve features to accurately identify the sender before accepting the email. However, it would be very difficult for these new protocols to be widely adopted because anyone who implements the new protocol would only be able to accept email from others who have also implemented the new protocol. So, without a more secure SMTP in the near future, spam will continue to be a problem, driving organizations to seek out effective spam blocking solutions.

SpamAssassin, an open source spam filter, is an example of a rule-based scoring system. To identify spam, SpamAssassin uses a wide range of heuristic tests on mail headers and body text.

Because spammers and their spam-making applications are not static, rule-based scoring systems face some of the same challenges that word filters face. Rules must be updated regularly in order for rule-based scoring systems to remain effective.

For example, if a rule-based scoring system has a rule that assigns points to the word "Viagra," spammers can easily circumvent this rule by purposely misspelling "Viagra" as many different ways as required to successfully deliver the spam. Rule-based scoring systems, however, if used properly, can be very effective, eliminating over 90 percent of incoming spam.

## Bayesian Filters

> **BAYESIAN ANALYSIS:** Named after Thomas Bayes (1702-1761), a mathematician who developed a theory of probability inference, Bayesian analysis uses the knowledge of prior events to predict future events.

Bayesian filters are personalized to each user and adapt automatically to changes in spam. To determine the likelihood that an email is spam, these filters use Bayesian analysis to compare the words or phrases in the email in question to the frequency of the same words or phrases in the intended recipient's previous emails (both legitimate and spam).

Bayesian filters are very powerful and are regarded as one of the most accurate techniques for blocking spam. Most reports on Bayesian filters have shown accuracy of over 99 percent when the filter has been "well-trained." For Bayesian filter training, approximately 200 legitimate emails and 200 spam emails from the intended recipient are normally needed. The more emails in the historical database of the intended recipient, the more accurate the filters are.

To learn more about the power of Bayesian analysis and filters, see "Better Bayesian Filtering," by Paul Graham at http://www.paulgraham.com/better.html.

## Black List IP

> **IP ADDRESS:** This is a unique identifier for a computer or device on a TCP/IP network. Networks using the TCP/IP protocol route messages based on the IP address of the destination.

Black list IP is a common spam blocking technique. It has no computational overhead and is easy to implement. This technique simply involves organizations manually keeping a list of the IP addresses of known spammers (a "black list") so that emails from those addresses are blocked.

Because spammers regularly change their IP addresses and use a wide range of IP addresses, black lists are most effective in blocking small amounts of spam for short time periods. They provide a quick fix for blocking one particular source of spam but are ineffective as an overall anti-spam solution.

An alternative to a black list is a white list. That is, a list of IP addresses from which you only accept email. This reverse concept of black lists, however, is impractical because users would only be able to receive email from IP addresses that they knew beforehand, making it impossible to receive email from any new sources.

## RBLs (Realtime Blackhole Lists)

RBLs (Realtime Blackhole List), also known as DNSRBLs, check every incoming email's IP address against a list of IP addresses in the RBL. If the IP address is part of the RBL, then the email is identified as spam and blocked.

Unlike the black list IP technique, RBLs are not manually updated by organizations. RBL operators maintain public RBLs and organizations simply subscribe to them.

Many organizations like using RBLs because they not only have low computational overhead but because they are normally implemented using a protocol similar to DNS (Domain Name Server), they also have low network overhead.

> **FALSE POSITIVE:** This is when a legitimate email is accidentally identified as spam.

A downside of RBLs is that they may generate false positives. Most RBLs are aggressive and block all reported spam sources. However, many times the spam sources, such as popular ISPs Yahoo, Earthlink or Hotmail, are also the source of legitimate email. In those cases, the legitimate email is typically never received since it is rejected as soon as its IP address is identified. The RBLs can not differentiate between when a source is sending spam and when it is sending legitimate email. It just blocks any email coming from the IP addresses in its list, thereby generating false positives at times.

RBLs are effective for blocking spam and should be part of an organization's spam blocking strategy. With careful selection of which RBLs to use, you can effectively eliminate spam without the downside of generating false positives.

## DNS MX Record Lookup

This is an effective technique for blocking spam from spammers who use a fake *from* and/or *return* address. Spammers use such fake addresses so that the spam cannot be traced back to them.

To determine if a *from* address is valid, the system does a lookup on the domain that is used in the *from* address. If the domain does not have a valid DNS MX record, then the *from* address is not valid and that email is labeled as spam. Similar lookups can be performed on the *return* address of the email as well.

## Reverse DNS Lookups

> **VANITY DOMAIN NAMES:** These are domain names that are typically registered to individuals or families for the use of email. They typically do not have their own email server, but share an email server with a hosting company.

This is an effective spam blocking technique that uses a reverse DNS lookup on the incoming email's source IP address. If the domain provided by the reverse lookup matches the *from* address on the email, the email is accepted. If they do not match, the email is rejected.

Reverse DNS lookups, while popular, often do not work well. They can generate a large number of false positives since many reverse DNS entries are not properly established and many more cannot be properly established. For example, any "vanity" domain name would most likely not have an accurate reverse DNS lookup. As such, emails from these domains would be rejected, causing unacceptably high false positive rates.

> **DMP (Designated Mailers Protocol):** This is a proposed standard for authorizing Mail Transfer Agents, or Mail Servers, to send e-mail on behalf of your domain. This prevents abuse of your domain by spammers and viruses.

> **SPF (Sender Permitted Form):** This is an extension to SMTP that helps prevent sender forgery. It is an open standard and it is also free.

> **RMX (Reverse Mail Exchanger):** This is a mechanism designed to enable a domain owner to list all mail servers authorized to send email on behalf of their domain name.

## New Reverse Lookup Systems

A number of spam blocking techniques have been proposed that use the DNS system to limit the ability to send spam from forged sender addresses. These techniques improve upon the reverse DNS lookup technique. Examples of these proposed techniques include:

- Reverse Mail Exchanger (RMX): http://www.ietf.org/internet-drafts/draft-danisch-dns-rr-smtp-04.txt
- Sender Permitted From (SPF): http://spf.pobox.com/
- Designated Mailers Protocol (DMP): http://www.pan-am.ca/dmp/
- Yahoo! Domain Keys: http://antispam.yahoo.com/domainkeys
- Microsoft Caller ID for Email: http://www.microsoft.com/mscorp/twc/privacy/spam_callerid.mspx

These approaches are similar in many respects. Similar to DNS MX records lookup, these reverse lookup solutions define reverse-MX records ("RMX" for RMX, "SPF" for SPF, and "DMP" for DMP) for determining whether email from a particular domain is permitted to originate from a particular IP address. Email addresses that do not originate from the correct RMX/SPF/DMP address range are identified as forged and the email itself is tagged as spam.

Like reverse DNS lookups, this technique also has problems with vanity domains, but may be partially corrected. The general case includes individuals and small companies who want to use their own domain rather than their ISP's, but cannot afford their own static IP address and mail server. Individuals sending email from a hostless or vanity domain simply configure their mail application to send email from their registered domain name. Unfortunately, a lookup of the sender's IP address will not find the sender's domain, and a lookup of the sender's domain may not find the correct reverse-MX record. The former is particularly common for mobile, dialup, and other users that frequently change IP addresses.

## Black List Sender Email Addresses

This is a simple spam blocking technique that is often used. Users create a black list of *from* addresses that should be prevented from entering the network and reaching the user's inbox.

There are a few disadvantages with using this technique. Because spammers can create many false *from* email addresses, it is difficult to maintain a black list that is always updated with the correct emails to block. Also, some spammers do not even use a *from* address so a black list would not be able to catch these cases. Even a rule to block emails without a *from* address would not be sufficient because some legitimate emails, such as newsletters to which a user may subscribe, may also not include a *from* address. Black list sender email addresses is effective in temporarily blocking a small amount of spam but ineffective as an overall anti-spam solution.

As an alternative to black lists, some users set up an email white list consisting of acceptable email addresses or domains. In this case, users only accept email from users that are listed on their white list, while all other email is blocked. This technique poses many challenges as well since people want to be able to receive email from people that they might not have entered into their white list.

Some techniques will attempt to automatically build the white list from email that you have sent or from your address book. This makes creating the list easier. However, it does not solve the problems associated when people who legitimately want to send you email have not previously corresponded with you via email, have multiple email addresses, or have a new email address.

## Honeypots (Hashing Systems, Fingerprinting)

Honeypots, or decoy email addresses, are used for collecting large amounts of spam. These decoy email addresses do not belong to actual end users, but are made public to attract spammers who will think the address is legitimate. Once the spam is collected, identification techniques, such as hashing systems or fingerprinting, are used to process the spam and create a database of known spam. Let's take a closer look at hashing systems and fingerprinting -

HASHING SYSTEMS: With hashing systems, each spam email receives an identification number, or "hash," that corresponds to the contents of the spam. A list of known spam emails and their corresponding hash is then created. All incoming email is compared to this list of known spam. If the hashing system determines that an incoming email matches an email in the spam list, then the email is rejected. This technique works as long as spammers send the same or nearly the same email repeatedly. One of the original implementations of this technique was called Razor.

FINGERPRINTING: Fingerprinting techniques examine the characteristics, or fingerprint, of emails previously identified as spam and use this information to identify the same or similar email each time one is intercepted. These real time fingerprint checks are continuously updated and provide a method of identifying spam with nearly zero false positives. Fingerprinting techniques can also look specifically at the URLs contained in a message and compare them against URLs of previously identified as spam propagators.

Honeypots with hashing or fingerprinting can be effective provided similar spam emails are widely sent. If each spam is made unique, these techniques can run into difficulties and fail.

## Challenge/Response Systems

Challenge/response systems are used to counter spammers who use automated mailing programs to generate millions of spam emails per day. These systems are designed to slow down spammers by putting roadblocks up for the incoming spam.

Challenge/response systems, such as those offered by Spam Arrest or MailBlocks, maintain a list of permitted senders. Each time an email from a new sender is sent to a challenge/response system user, the email is temporarily held before delivery. The challenge/response system sends the email sender a challenge. This challenge usually consists of a link to a URL or a request that the sender copy a numeric code into a box in the reply email. If the sender successfully completes the "challenge," the challenge/response system adds him to the list of permitted senders and his email is delivered to the intended destination.

Challenge/response systems work under the assumption that spammers using fake sender email addresses would never receive the challenge, and spammers using real email addresses would not be able to reply to all of the challenges.

Challenge/response systems have a number of limitations. These limitations include:

- DEADLOCK: Deadlock is when two people can not communicate with each other because both are using challenge/response systems. For example, assume Bill and Tom do not know each other well and have never communicated via email in the past. Bill legitimately needs to contact Tom and so he sends Tom an email. Tom's challenge/response system intercepts the email and sends a challenge to Bill. Because Bill also has a challenge/response system, Bill's system intercepts Tom's challenge and issues its own challenge. Unfortunately, in a situation where both users have challenge/response systems, neither user will ever receive the challenges and the original email will never get delivered.

> **MAIL HASH (also called Message Digest):** This is a number generated from a string of text. The hash is substantially smaller than the text itself, and is generated by a formula in such a way that it is extremely unlikely that a different set of text would produce the same hash value.

**Distributed Checksum Clearinghouse** (DCC) is a variation on the honeypot technique. Rather than taking into consideration the contents of an email, DCC simply counts the number of times that the same email appears on the Internet. If the same email appears many times, then it is assumed to be spam. For more details, see http://www.rhyolite.com/anti-spam/dcc/.

- AUTOMATED SYSTEMS: With challenge/response systems, users can not receive email from mailing lists and automated systems such as Yahoo's "Send To A Friend." Mailing lists and automated systems will not be able to respond to the challenge and as a result, their emails will never get delivered.

As more people use challenge/response systems, these systems end up interfering with the delivery of legitimate email rather than deterring the unwanted spam.

## Computational Challenge Systems

Computational challenge systems add a cost to sending email by requiring the sender's system to perform a computation prior to sending the email. Most computational challenge systems use complex algorithms that are intended to take time to process. The hope is that a high enough cost would stop people from sending spam to those with computational challenge systems.

How do computational challenge systems work in practice? Let's assume Derek is using a computational challenge system to help stop spam. A new friend, Joe, decides to send Derek an email for the first time and therefore is not yet on Derek's list of acceptable senders. Derek's server receives the email and sends a computational challenge (typically a math problem or algorithm) to Joe's email client. Derek's server waits for a response before allowing the email to be delivered to Derek's inbox.

As illustrated in the above example, for a single, legitimate user sending emails, the time it takes to complete a computation is unlikely to be noticed. The sender's system does the challenge and the email is delivered to the intended recipient. However for someone such as a spammer sending a lot of email, the small delays add up, making it take a long time and hopefully not worth it, to send out bulk emails.

A few examples of proposed computational challenge systems are programmer Andy Back's HashCash program and Microsoft's Penny Black. These systems, as with all computational challenge systems, have limitations. These limitations include:

- UNEQUAL TAXATION: Computational challenges, whether based on CPU, memory, or network, penalizes users with slower systems. For example, a CPU challenge that takes 10 seconds on a 1Ghz computer would take over 20 seconds on a 500MHz computer.
- MAILING LISTS: Legitimate mailing lists, some with thousands or millions of recipients, would be penalized just as significantly as spammers. Computational challenges make mailing list management impractical. Furthermore, any solutions that could be used by mailing lists to bypass the challenge would also allow spammers to bypass the challenge as well, thereby defeating the purpose of having a challenge system.
- ROBOT ARMIES: Using Sobig and other spam-supporting viruses, many spammers control thousands of compromised systems. Spammers can easily distribute any high costs from challenge systems across these infected systems (robot armies), making challenge systems an ineffective way to discourage spammers.
- LEGAL ROBOT ARMIES: Spammers generate spam because it brings in significant revenue. Large spam groups can afford purchasing hundreds of systems for distributing any computational cost. This can be done legally, without compromising systems with viruses.

All these limitations make it unlikely that computational challenge systems will be widely adopted. These systems not only inconvenience legitimate emailers but they also do not appear to effectively mitigate spam.

## Rate Controls

Sometimes spammers attempt to cripple email servers by sending a large quantity of email in a short period of time. This is called a DOS (Denial of Service) attack. With rate controls, a system administrator can set up parameters that protect the email server from this email flood.

Rate controls can be set up to allow only a certain number of connections from the same IP address during a specified time. For example, a rate control time can be set to 30 minutes with only a certain number of connections to be allowed in that given time period. If the administrator sets this parameter to 50 connections, the firewall will block any correspondence after the first 50 connections that comes from a single IP address within a given 30 minute time period.

Rate controls are effective in protecting the network from spammers who attempt to send hundreds of spam emails at the same time to a specific email server.

## Anti-Virus Scanning

Anti-virus scanning can really be viewed as a method of stopping spam since a large amount of unwanted email is generated by virus programs that attempt to propagate themselves. A virus scanning solution is certainly an effective tool to include as part of any organization's overall anti-spam solution.

## Conclusion

Spam is a problem that is continuing to grow from day to day, costing corporations billions of dollars in lost productivity. Fortunately though, there are different spam blocking techniques to help counter the various types of spam.

Because spammers are always trying to bypass anti-spam techniques by changing the methods they use to send spam, it's best for corporations to protect themselves with a spam blocking solution that uses more than one spam blocking technique. Each one of these techniques has advantages, disadvantages, as well as limitations. To minimize the amount of spam that enters an organization, a spam blocking solution that includes a combination of the most effective techniques should be implemented.

**The Barracuda Spam Firewall** uses ten defense layers to protect your email server from spam and virus attacks. To learn more about the defense layers and the spam blocking techniques they use, please visit www.barracudanetworks.com.

**Barracuda Networks**
10040 Bubb Road
Cupertino, CA 95014
+1 408 . 342 . 5400
+1 888 . 268 . 4772
www.barracudanetworks.com
info@barracudanetworks.com