Content Filters

A WORD TO THE WISE WHITE PAPER | BY LAURA ATKINS, CO-FOUNDER

Introduction

Content-based filters are a key method for many ISPs and corporations to filter incoming email.. These content-based filters allow more selective blocking than IP-based filters. While IP-based filters do block mail from known bad sources, a lot of mail comes from IP addresses that send both good and bad mail. Content filters protect users from the bad mail, but allow the good mail through.

Content-based filters look at a range of message components, from the actual text in the message, to the domains, to the IP addresses those domains and URLs point to. They look at the hidden structure of an email. They look at what's in the body of the message and what's in the headers. There isn't a single bit of a message that content filters ignore.

Many content filters are designed to evaluate messages in a nuanced way. Just because an email has content that triggers a filter doesn't necessarily mean that mail will be rejected or delivered to the bulk folder. Because content filters measure multiple components, often good content or structure can compensate for poor content or structure.

Senders that understand content-based filters and how they work can use that knowledge to craft emails that meet brand and design standards and accomplish business objectives without hitting filters.

To understand exactly what content-based filters are and are not, it's useful to look at the history of commercial email.

Many of the very earliest spam filters were designed to block based on the source of an email. Every email sent has a fixed source IP address which identifies the machine that sent the mail. Early filters blocked all traffic from IP addresses known to be sending bad traffic, including spam. As spam and email became more complex there were few IP addresses that sent only good mail or only bad mail. As a result, IP-based filters would block too much wanted mail.

Identifying the senders' IP addresses is useful beyond blocking mail. Many ISPs and filtering companies offer preferential delivery to senders using IP addresses with good reputations. For a while, developing a good IP address reputation was the holy grail of email delivery. If you had a good reputation, then you could send all sorts of mail through that IP and get it delivered.

ISPs and filtering companies developed content-based filters to selectively block or allow mail from any source IP address. Instead of blocking all traffic from an IP address, content filters look at the email itself before deciding if the email is bad mail or good mail. Content filters let wanted mail through even if the IP address has a poor or unknown reputation.

On the flip side, content filters prevent spammers from "hijacking" IP addresses with good reputation and using them to send unwanted mail. There are a number of spammer techniques that borrow the reputation of a mail server in order to attempt to get spam delivered. Content filters prevent spammers from stealing someone's good reputation, without hurting the real mail coming from that IP address, allowing recipients toselectively block email no matter where it comes from.

What content filters are not

Content filters are not about words or phrases you should or shouldn't use.

They're not about avoiding the word FREE or !!! in the subject line or body of a message. There are some exceptions to this rule, however. Many businesses will block on simple words like "Viagra" or filter porn. These are content filters, but they're not always about spam. Businesses often block mail simply because it's not business related, not because it's unsolicited.

What content filters do

Content filters look at the whole content of an email message and compare it to other email, both good and bad.

Content filters look at the content of an email, the whole content of the email, not just what is displayed to the end user. Headers, footers, pre-headers, HTML structure, text parts, images and domains in the message are all analyzed as part of the content screening.

Some of the distinctive content differences between wanted and unwanted email are due to the content as written by the sender, some of them are due to senders of unwanted email trying to hide their identity or their content, but many of them are due to the different quality software used to send each sort of mail. Mail clients used by individuals and content composition software used by high quality email service providers (ESPs) tend to be well

written and comply with both the email and MIME RFCs, and the unwritten best practices for email composition. The software used by spammers, botnets, viruses and low quality ESPs tends not to do so well.

Header analysis

Email headers record the technical steps of email delivery as part of the email. Most email clients hide the email headers from the end user recipient by default. Filters and email servers, however, have full access to this data. Internet standards define what data the headers must contain in order to be a valid email[1]. The standards also allow for some optional information to be included in headers.

Some spamware (the software used by spammers to send messages) puts in distinctive headers, thus any email with that header in it is most likely spam. Examples of this include a header with a fake version of Microsoft Outlook and a date header referencing a timezone that does not exist. Other spamware is poorly written and has unusual ways of writing data that sets it apart from normal mail software.

As long as you are using a responsible and legitimate ESP to send mail these filters should not cause you problems.

Body analysis

The email body is that part of the email displayed to the user, or HTML encoding used to format the email displayed to the user. Content filters analyze both the part of the email displayed to the user and the HTML part usually hidden from the user.

HTML structure

HTML structure is an important part of the analysis. Legitimate senders should use valid and correct HTML in all their emails. Spammers have long used fake HTML tags as a way to try and avoid filters; now some filters actually look at the tags and compare them with the HTML

---

[1] RFC 2821 < http://www.ietf.org/rfc/rfc2821.txt>

standard. Other spammers put random content in HTML comments as a way to confuse content filters. As a result, many content filters now look at the ratio of HTML comments to visible text. Just having comments doesn't trigger filters, but having more comments than real text will trigger filters.

Filters also look at the pattern of HTML and the layout of an email using a technique called "email fingerprinting." Email fingerprints are used to identify emails created by certain bits of software or certain people. Content creators are consistent in how they create HTML. Wanted email has a fingerprint and ISPs filter this email to the inbox. Unwanted email also has a fingerprint, and ISPs filter this mail to the bulk folder.

Spelling and grammar errors

Spammers are notoriously bad at spelling, and will frequently misspell words in order to bypass spam filters. As such, many filters will look at the type and number of misspelled words. Too many misspellings, use of foreign characters, and/or creative letter spacing tell filters this mail is likely spam.

Domains, links and images

Many content filters look very carefully at domains, URLs, links and images in the email. In some cases, filters also follow links and look at the destination webpage. Filters look at a number of things about domains and URLs, including

- Has this domain has ever been seen in email before?
- Has email with this domain in it generated complaints?
- Does the plain text part of the link match the domain listed in the <a href> part?
- Has this domain been listed on any domain-based blocklists?
- Have we blocked this domain in the past?

Avoiding content filters

Content filters are not fixed, they adapt as email changes.

In order to keep up with the ever-changing world of email, most content filters are updated frequently. Many filter updates are intended to catch bad mail that is trying to get past the

filters. Thus, frequent changes to email structure and chasing filter changes may not improve delivery.

It is important to brand emails so that recipients can clearly associate the email with the sender. But some branding elements may result in filter triggers, such as light text on dark backgrounds or image heavy emails. There are no hard and fast rules on how to avoid content filters, but there are some general guidelines that have been successful for many senders.

- **Avoid image-only emails**. Emails should have sufficient text that they can be read and acted on without having to load images in the email client.
- **Avoid hard-to-read color combinations**. Some filters look at color combinations and downgrade for text that is hard to read. Use text colors that are legible on a white background.
- **Avoid using linking to or mentioning domains or URLs you don't control**. Some domains have a poor reputation, and even mention of those domains can result in filtering.
- **Use a consistent From address**. Many email clients allow users to whitelist specific senders based on their From address. Using the same From address — and telling users to add that to their address book — will get more mail delivered to the inbox.
- **Identify your brand in the subject line**. Senders have only a few seconds to grab the attention of the recipient. Branded and consistent subject lines help the user know this is wanted mail and encourage them to open it.

Sometimes it's not possible to completely avoid things that will trigger content filters. There are entire classes of mail (Viagra, payday loans, Rolex watches) that are so heavily spammed that any email, even the cleanest opt-in email, mentioning those products is treated suspiciously. Senders can compensate for content filters by being extra careful about subscription practices and data hygiene.

Content filters are

- Widely used by commercial and business receivers
- A way for receivers to selectively block mail, regardless of the source

- A way for receivers to selectively allow mail, regardless of the source

Content filters are not

- Looking for things like “Free” or “!!!” in the subject line
- Focused on specific words or phrases
- Usually all or nothing.