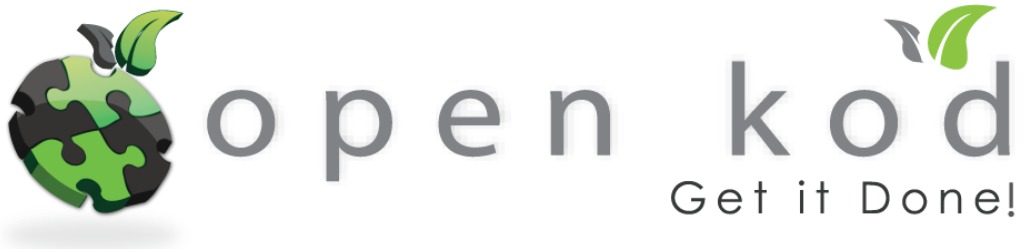


Spam and Ham

A Simple Guide

Fauzi Yunos

12

Executive Summary

People tend to be much less bothered by spam slipping through filters into their mail box (false negatives), than having desired email ("ham") blocked (false positives). Trying to balance false negatives (missed spam) vs. false positives (rejecting good email) is critical for a successful anti-spam system. Some systems let individual users have some control over this balance by setting "spam score" limits, etc. Most techniques have both kinds of serious errors, to varying degrees. So, for example, anti-spam systems may use techniques that have a high false negative rate (miss a lot of spam), in order to reduce the number of false positives (rejecting good email).

Detecting spam based on the content of the email, either by detecting keywords such as "Viagra" or by statistical means (content or non-content based), is very popular. Content based statistical means or detecting keywords can be very accurate when they are correctly tuned to the types of legitimate email that an individual gets, but they can also make mistakes such as detecting the keyword "cialis" in the word "specialist". The content also doesn't determine whether the email was either unsolicited or bulk, the two key features of spam. So, if a friend sends you a joke that mention "Viagra", content filters can easily mark it as being spam even though it is neither unsolicited nor sent in bulk. Non-content base statistical means can help lower false positives because it looks at statistical means vs. blocking based on content/keywords. Therefore, you will be able to receive the friend who sends you a joke that mentions "Viagra".

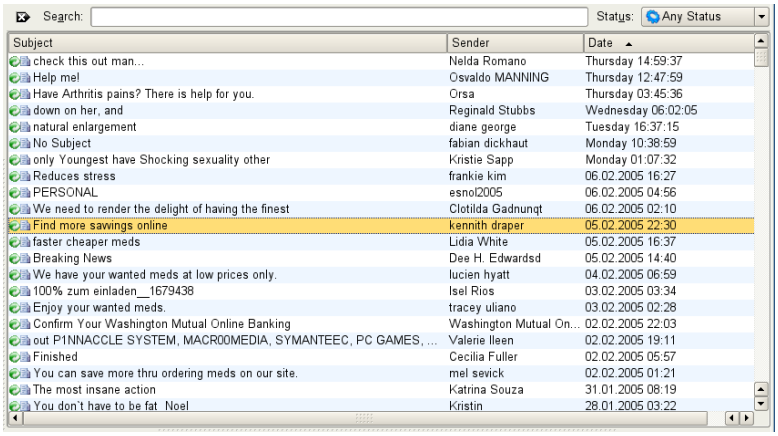
To prevent email spam (aka unsolicited bulk email), both end users and administrators of email systems use various anti-spam techniques. Some of these techniques have been embedded in products, services and software to ease the burden on users and administrators. No one technique is a complete solution to the spam problem, and each has trade-offs between incorrectly rejecting legitimate email vs. not rejecting all spam, and the associated costs in time and effort.

Anti-spam techniques can be broken into four broad categories: those that require actions by individuals, those that can be automated by email administrators, those than can be automated by email senders and those employed by researchers and law enforcement officials.

Contents

What is SPAM?

Spam is the use of electronic messaging systems to send unsolicited bulk messages indiscriminately. While the most widely recognized form of spam is e-mail spam, the term is applied to similar abuses in other media: instant messaging spam, Usenet newsgroup spam, web search engine spam, spam in blogs, wiki spam, online classified ads spam, mobile phone messaging spam, internet forum spam, junk fax transmissions, social networking spam, television advertising and file sharing network spam.



An email box folder littered with spam messages

Spamming remains economically viable because advertisers have no operating costs beyond the management of their mailing lists, and it is difficult to hold senders accountable for their mass mailings. Because the barrier to entry is so low, spammers are numerous, and the volume of unsolicited mail has become very high. In the year 2011, the estimated figure for spam messages is around seven trillion. The costs, such as lost productivity and fraud, are borne by the public and by Internet service providers, which have been forced to add extra capacity to cope with the deluge. Spamming has been the subject of legislation in many jurisdictions.

Within a few years, the focus of spamming (and anti-spam efforts) moved chiefly to e-mail, where it remains today. Arguably, the aggressive email spamming by a number of high-profile spammers such as Sanford Wallace of Cyber Promotions in the mid-to-late 1990s contributed to making spam predominantly an email phenomenon in the public mind. By 2009, the majority of spam send around the world was in the English language; spammers began using automatic translation services to send spam in other languages.

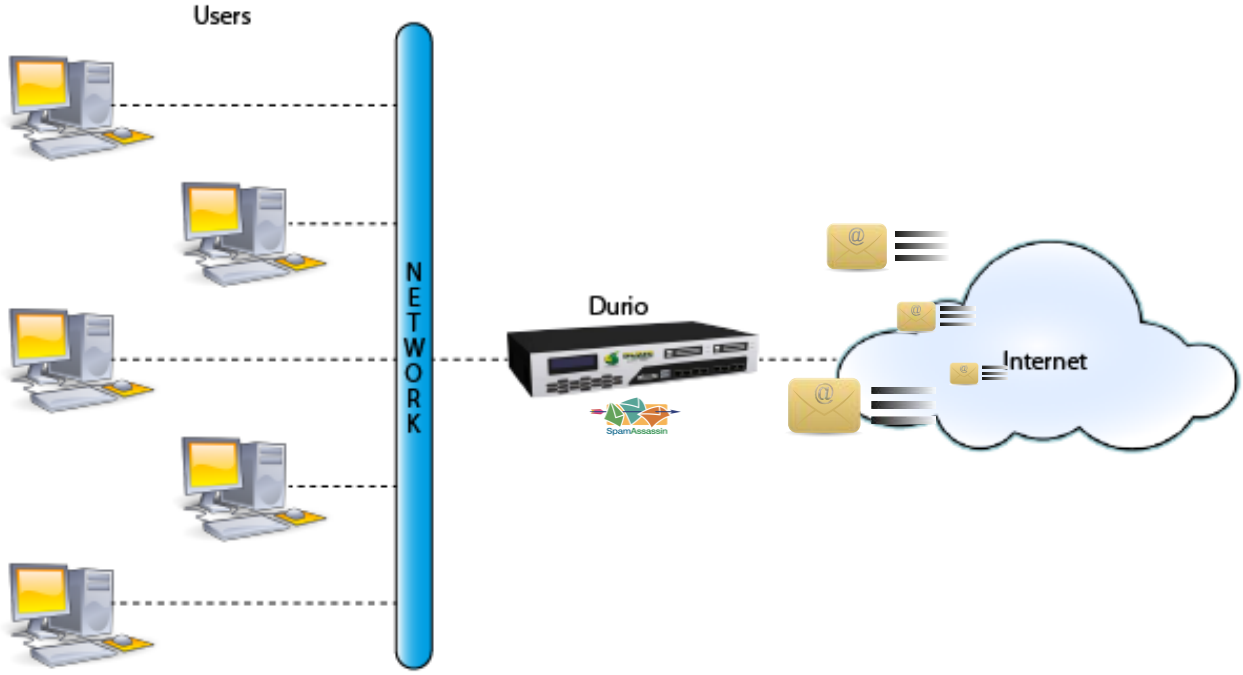
SpamAssassin

SpamAssassin is a module to identify spam using several methods including text analysis, internet-based real-time blacklists, statistic analysis, and internet-based hashing algorithms. Using its rule base, SpamAssassin uses a wide range of heuristic tests on mail headers and body text to identify “spam”, also known as unsolicited bulk email. Once identified as spam, the mail can then be tagged as spam for later filtering using the user’s own mail user agent application or at the mail transfer agent.

SpamAssassin is not a program to delete spam, route spam and ham to separate mailboxes or folders, or send bounces when you receive spam. Those are mail routing functions, and Spam Assassin is not a mail router. SpamAssassin is a mail filter or classifier. It will examine each message presented to it, and assign a score indicating the likelihood that the mail is spam. An external program must then examine this score and do any routing the user wants done. There are many programs that will easily perform these functions after examining the score assigned by SpamAssassin.

SpamAssassin is the all powerful wizard behind the curtains powering the Junk Mail filter in OS X. And more specifically, the Mac OS X implementation of SpamAssassin is using AMaViS (A Mail Virus Scanner) to call SpamAssassin. Some of the commercial products that have used SpamAssassin are McAfee SpamKiller, Symantec MainScanner & Kerio.

How Durio works with SpamAssassin



Durio came pre-installed with SpamAssassin, another great application that being chose as one of the Apache Technologies that have changed computing in the past 10 years.

Durio uses its engine called a "Rules-Based Heuristics Engine" because the filter typically looks for patterns such as common phrases or known senders. Heuristics is the application of experience-based techniques for problem solving. Virus scanners use the same technique and are also heuristics engines.

As an e-mail goes through the engine, each of the rules for SpamAssassin is run and generates a score. These individual scores are then totaled to provide an overall score for the e-mail. Some rules are positive and add to the score. Some rules are negative take away from the score.

The lower the overall score is, the more likely the e-mail is Ham. The higher the score, the more likely the e-mail is Spam.

In the anti-spam community, Spam and Ham are opposites.

- *Spam = Junk E-mails*
- *Ham = Good E-mails*

But not each rule has the same weight. The weight each rule is given is determined first by initial "guesses" and later refined through optimization. We use a Genetic Algorithm to optimize the scores. Discussing Genetic Algorithms is not going to be in the scope of this conference especially since Wikipedia defines it as "a search technique used in computing to find exact or approximate solutions to

optimization and search problems. Genetic algorithms are categorized as global search heuristics. Genetic algorithms are a particular class of evolutionary algorithms (EA) that use techniques inspired by evolutionary biology such as inheritance, mutation, selection, and crossover". The key point is that we use a scientific process to obtain the scores.

Conclusion

Spam is best definite as CONSENT and NOT about CONTENT because consent is when you give approval to someone to send you e-mails.

Durio anti-spam at its heart is a Scoring Framework. This framework allows virtually any anti-spam technologies to be added and used. It is FUTURE-PROOF. If anyone, identifies an algorithm / program / magic wand to detect spam (or ham), it can be quickly integrated into the framework as another test with an appropriately weighted score.

Durio typically differentiates successfully between spam and ham in between 95% and 100% of cases, depending on what kind of mail you get and your training of its Bayesian filter. Specifically, Durio has been shown to produce around 1.5% false negatives (spam that was missed) and around 0.06% false positives (ham incorrectly marked as spam).

Durio also includes plugin to support reporting spam messages automatically or manually to collaborative filtering databases.