

Name: Chris Lucas
Student Number: 0045849
e-Mail: lucascg@mcmaster.ca

Software Engineering 4C03

SPAM

Introduction

As the commercialization of the Internet continues, unsolicited bulk email has reached epidemic proportions as more and more marketers' turn to bulk email as a viable advertising medium. Once considered little more than an annoyance, spam has become an enormous problem affecting Internet users and broadband service providers. Well-known viruses, worms, and Trojan horses get the headlines, but spam is arguably a more pervasive and insidious threat because it affects every Internet user-directly or indirectly-and it lacks a comprehensive solution analogous to antivirus software programs. Spam frustrates users by overloading their e-mail boxes with volumes of useless and unwanted messages. Beyond the annoyance and inconvenience, spam causes real damage to both users and service providers. Scams cause unwary users to reveal valuable personal information such as credit card numbers or passwords, suffering monetary damages as well as losing time and privacy. Spam can carry malicious code viruses such as distributed denial-of-service (DDoS) agents. On the service provider side, spam overloads e-mail servers, delaying or preventing the delivery of legitimate e-mail messages. It consumes bandwidth and many hours of human capital dealing with the large volumes. The continued increases in spam have had a large economic impact creating growing concerns about its proliferation, forcing the internet community to increasingly look to both regulatory and technical solutions to alleviate this huge threat. Currently there exists a broad range of counter measures to deal with the problem of spam, some of which have been successfully deployed in commercial environments. This paper will attempt to evaluate some of the existing spam sending techniques, the economic costs of spam, and spam blocking techniques to control this ever-increasing problem.

Spam Sending Techniques

To understand the issues involved in controlling spam, the methods employed by spammers should be investigated. First generation spammers used the simplest technique: Send out thousands or millions of e-mail messages from their own e-mail accounts. However this was easily combated by service providers by blacklisting these users. Using mail volumes, subject line and message analysis, and user complaints, the service providers could identify spammers and ban them from the network, a simple policy that was easily enforced. Spammers quickly switched to a new technique using open mail proxies. In brief, an open mail proxy is a server that accepts connections from any network address, acting as a blind intermediary to virtually any other network address. To the recipient (and the intervening network infrastructure), the spam message seems to originate from the mail proxy, effectively masking the sender's true identity. Service providers responded with a second kind of blacklist, this time of known mail servers that were sending spam. In response to the server blacklist, spammers developed an even more sophisticated method of attack-the spam zombie. By infecting unprotected computers with a Trojan horse program, a spammer effectively recruits an army of unwitting users who can be activated by a remote command to launch a spam attack.

Name: Chris Lucas
Student Number: 0045849
e-Mail: lucascg@mcmaster.ca

Such an attack has characteristics similar to a DDoS attack: The large number of attacking machines makes it difficult or impossible either to identify the source of the attack or to take effective corrective action in real time without causing massive disruptions to legitimate users.

Costs of Spam

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. Spam will cost organizations worldwide \$50 billion US in 2005, according to a recent report from Ferris Research. A major component of spam costs, according to our research, is end users' time. That is simply the amount of time that's wasted dealing with the after effects of spam. For people who don't have any sort of spam filter at all it's basically the amount of time it takes them to delete messages. There's another chunk of productivity costs in the organization involving time spent by IT maintaining spam filters and associated tasks. There, we're talking about time spent installing, time spent maintaining, and time spent fixing problems. There's also a big chunk of time spent running the IT Help desk and dealing with user questions related to spam and spam filters.

Spam Blocking Techniques

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. The accuracy of spam blocking techniques are evaluated on two dimensions: How much spam you successfully filter out, and how little legitimate messages you accidentally delete. Maintaining accuracy can be difficult because spam is constantly changing, the most effective spam blocking solutions contain more than one of the following techniques to help ensure that all spam, and only spam, is blocked. The following is an overview of different spam blocking techniques.

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. 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. If properly maintained and updates this method can be very effective, eliminating over 90 percent of incoming spam.

Bayesian Filters

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

Name: Chris Lucas
Student Number: 0045849
e-Mail: lucascg@mcmaster.ca

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.

Black List IP

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.

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, they also have low network overhead. 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.

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 senders address is not valid and that email is labeled as spam.

Reverse DNS Lookups

Name: Chris Lucas

Student Number: 0045849

e-Mail: lucascg@mcmaster.ca

This is an effective spam blocking technique that uses a reverse DNS lookup on the incoming e-mail's source IP address. If the domain provided by the reverse lookup matches the sender 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.

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. 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 host less 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.

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.

Name: Chris Lucas
Student Number: 0045849
e-Mail: lucascg@mcmaster.ca

Bibliography

1. N/A. (12/13/04). *An Overview of Spam Blocking Techniques*. Retrieved March 30, 2005 from, http://www.barracudanetworks.com/resources/docs/Spam_Techniques.pdf
2. Shane Hird. (2002). *Technical Solutions for Controlling Spam*. Retrieved March 30, 2005 from, http://security.dstc.edu.au/papers/technical_spam.pdf
3. Cisco Systems. (2005). *Assessing the Impact of Spam Zombies on Broadband Service Providers*. Retrieved March 30, 2005 from, http://www.cisco.com/en/US/products/ps6150/products_white_paper0900aecd802571d2.shtml
4. GroupWise Advisor. (2005). *the Costs of Spam*. Retrieved March 30, 2005 from, <http://gwadvisor.com/doc/16164>