Eradicating Spam Through a Hybrid Sender-Pays Model

Ron Jager

Ron E. Jager

March 15, 2004

GIAC Security Essentials Certification (GSEC)

Version 1.4b Option 1

Eradicating Spam Through a Hybrid Sender-Pays Model

Abstract:

Spam is a security and resource problem. It is a resource problem because it takes system and network resources away from legitimate uses. It is a security problem because it creates unwanted and uninvited traffic in an attempt to defraud users. This paper will show how to effectively eradicate spam and spammers by changing the economics of spam through the addition of electronic postage. Postage attacks spam at the beginning of the process before it is even sent, in contrast to filters that attack spam at the end of the process. The end result is the elimination of smaller scale spammers through reduction of profitability and reduction of overall spam through the rate limiting characteristics of postage.

Introduction: Describing Spam And How It Needs To Be Eradicated

The evolution of email as a network message exchange was developed for the ARPANET shortly after it was created, and has evolved into the powerful technology we use today.

> It soon became obvious that the ARPANET was becoming a human-communication medium with very important advantages over normal U.S. mail and over telephone calls. One of the advantages of the message systems over letter mail was that, in an ARPANET message, one could write tersely and type imperfectly, even to an older person in a superior position and even to a person one did not know very well, and the recipient took no offense. The formality and perfection that most people expect in a typed letter did not become associated with network messages, probably because the network was so much faster, so much more like the telephone. [i]

The days of the recipient no longer taking offense to "casual" communication from senders they don't know is over. Anyone who has an email account can give you a few descriptive words on how spam cuts into the convenience of using email. System and Email Administrators can tell you how their job detail has grown with the recent influx of enormous amounts of Spam bombarding their network.

The scale, growth, and effect of spam on the Internet has generated considerable interest in addressing this problem. Once considered a mere nuisance, spam has grown to account for over 56% of all email and climbing.[ii] It has become a traffic-hogging epidemic. This unwanted traffic stands to affect

local networks, the infrastructure, and the way that people use email. An individual or organization should be able to express consent or lack of consent for certain communication and have the architecture to support those desires. Expressing consent is more straightforward on an individual basis; as the solution is moved closer to the source, it is more difficult to express a policy that satisfies all downstream receivers.[iii]

Spammers use a variety of techniques (mostly computer generated) to harvest email addresses. They then use those addresses and any permutations of those addresses to send very fast bulk e-mail designed to cram the message into mail servers. The challenge is to weed out the messages created automatically by these computers from those that are actually written by humans or legitimate programs, intended to arrive in another's mailbox.

Spam is easily recognizable to the recipient and [I] if you hired someone to read your mail and discard the spam, they would have little trouble doing it.[iv] The issue is how do we effectively scan email messages so that they are classified appropriately and spam is discarded without losing innocent emails identified as spams? For most users, missing legitimate email is an order of magnitude worse than receiving spam, so a filter that yields false positives is like an acne cure that carries a risk of death to the patient.[v]

According to research from Sophos Inc. during a 2-day tracking of worldwide email during the last week in February 2004, 57% of Spam came from U.S. computers.[vi] Furthermore, the research found that more than 30% of spam is sent using Trojan horses and worms to take over computers and turn them into spam engines, a practice Sophos believes likely accounts for much of the Spam originating in the United States.[vii] Spam is a security threat as much as it is a nuisance. The best way to stop spam lies on a multi-level process that includes making spam expensive to send, increasing the accuracy of filtering to eliminate false positives and easily identifying wanted email from people you know so that the mail from friends, family and business associates passes right through. Open source software, Camram (www.camram.org) meets these requirements automatically by attacking spam from the beginning with a hybrid “sender-pays” system using proof of work postage stamps. It continues that attack with a multi-level process where the message must pass through layers of controls including Bayesian filtering and auto-generated white list.[viii]

Filtering Techniques Alone is not a Solution:

There are many products on the market that incorporate different filtering techniques but spam still gets through. Spammers are constantly attacking filtering and eventually find a way around or through them. Filtering is important but by itself will never be enough.

One type of filtering, signature-based filtering is not effective at all. As soon as the filter developers figure out how to ignore one kind of random insertion, the spammers switch to another. Another type of filtering, rule-based

(aka heuristic) filtering tends to have a high false positive rate and the rules are generally static requiring the filter's authors to constantly write new rules to catch them.[ix] An interesting filtering technique that has resurfaced lately and is being used to try and fight spam is challenge-response filtering. When you get an email from someone you haven't had email from before, a challenge-response sends an email back to them, telling them they must go to a web page and fill out a form or type some words before the email can be delivered.

Challenge-response filtering is not a new concept. Proving that you're a human to another human can be done using an idea from the 1950s: the Turing Test.[x] A human judge asks you a bunch of questions and decides, depending on your answers, whether he is talking to a human or a computer. Proving that you're a human to a computer (or this is human generated mail vs. computer generated Spam) is another matter. It requires a test (or a set of tests) that computers can grade, humans can pass, but paradoxically, computers can't pass. For example, if you get a new email account at Yahoo, you'll be asked to prove that you're a human by responding to a question or typing in a word that is distorted in a box. This attempts to prevent a single computer program from getting thousands of free email accounts per second. At Carnegie Mellon, Luis von Ahn, Manuel Blum and John Langford have coined the phrase CAPTCHA ("Completely Automated Public Turing Test to Tell Computers and Humans Apart"). Examples and code of a few examples of CAPTCHAs can be found online at http://www.captcha.net. Solving a CAPTCHA with human input can be time consuming for a legitimate sender and for spammers it is just another filtering system that they will eventually find ways to get around.

There have been successful programs written that are quite accurate in solving CAPTCHA tests.[xi] Another way around a CAPTCHA test is really quite clever. Porn companies have computers that try to get free email accounts, and as soon as the computers encounter a CAPTCHA test, they simply send the test to the porn site. Back at the "free" porn site, there are several thousand users viewing pornographic pictures, and they are told: "please solve the following test before we can show the next picture" and the CAPTCHA test is bypassed showing how humans can expand the computational abilities of computers.[xii]

A more recent filtering technique that has gained wide spread popularity is Bayesian filtering. Bayesian filtering is a statistical approach to filtering spam and it has proven to be the most effective technique to date. It calculates the probability that a given message is spam or not based on the contents of that message, along with the contents of past messages and past spam that was received. It differs from other filters in that it is not reactive but proactive. It learns and predicts spam by analyzing past good email and past spam to determine whether a new message is spam or not.

The first papers about Bayesian spam filtering seem to have been given at the same conference in 1998, one by Pantel and Lin[xiii], and another by a group from Microsoft Research[xiv]. Pantel and Lin's results of their Bayesian filtering tests were better than Microsoft's. They were enhanced, according to Paul

Graham, by five things. Mr. Graham had significantly better results by training the Bayesian filter on a larger corpus of data and including email message headers in the filtering. Unlike Pantel and Lin, Graham did not reduce the tokens to root words. He used the 15 most significant tokens, reducing the probability of missing longer spams that contained random text to offset the spam terms. Further, Graham was able to get much better accuracy by adjusting the filtering to decrease the false positive rate at the expense of the filtering rate.

The evolution of the Bayesian filtering has definitely tipped the scale towards winning the battle against spam. However, filtering alone is not the answer to complete eradication of spam. Spammers are constantly thinking of ways to overcome filters. For example, the new characteristics of spam have random broken up and misspelled words in the emails used to prevent filters from recognizing them. Another example is standard language that would be in any legitimate email but with a URL taking you to a spam site. A Bayesian filtering scheme will be confused by such emails.

Some have suggested that spams with such characteristics can have the URLs blacklisted and the URLs blocked themselves.[xv] Others in the technology community have taken it a step further and favor denial-of-service attacks launched against such websites. "Such attacks, which are illegal and can disrupt a company's communications network by burying its servers in unnecessary requests, have traditionally been associated with pranksters who use viruses to distribute their attack software on thousands of computers."[xvi] The attacks would be launched automatically by the next generation of spam filters whenever the filters received a new piece of spam containing a Web link. The theory is that if enough people used filters with the feature, the resulting amount of traffic to the spammers' links could dramatically raise the spammers' bandwidth costs or even shut down their websites. The problem with this theory is the false positives and the potential harm to legitimate companies being blacklisted because a filter made a mistake on a piece of email that contains a legitimate URL or spammers start mixing in legitimate URLs with their solicitations.

One product www.brightmail.com incorporates the idea of URL blocking. It follows the theory that since an increase of spams with URLs are coming through the filters with unwanted links to solicitation sites; those URLs should be gathered and blacklisted. However, spammers and the characteristics of spams are constantly evolving to get past the perimeter's defense. Recently Spammers are reacting to URL blocking filters by inserting legitimate URLs into Spam i.e. financial spam with lots of cbsmarketwatch.com links. Spams are injecting even more randomization in URLS, using random ports, random sub-domains and extensive use of redirects. Matterform Media President Michael Herrick, whose company develops the Spamfire filter, summarizes the spam issue, "It's a cat-and mouse game with Spammers, " he said. "They mutate, we adapt. We fight back, they parry…"[xvii]

When asked at the Spam Conference 2004 how the final determination is made by Brightmail when there are unclear decisions which URLs are legitimate

or not, Mr. Schneider responded that Brightmail has three sets of technicians, in locations through out the world, that make the final determination. Again we are increasing administrative overhead to filter Spam, because ultimately filters alone eventually fail.

Spam Solution: Sender Pays

The answer to spam is attacking it at its creation. Spammers make money by sending out millions of messages. Requiring postage on each message would effectively shut down their business. As will be explained below, market driven monetary postage will fail for a variety of reasons. The proposed solution uses a proof of work based on an electronic postage stamp. Unlike physical postage, you don't spend money. You spend time by solving a puzzle whose solution becomes your postage stamp.[xviii] The exercise would merely serve as proof of the sender's good faith. Time is money, and spammers would have to buy many more machines to solve enough puzzles to get their spam through. This would reduce the profitability of spammers by limiting the number of messages they send or increasing their costs of production. Spam will suffer either way.

Monetary-based Postage and Why it Fails:

Some anti-spam systems like http://www.goodmailsystems.com/ or http://www.bluesquirrel.com/products/SpamSleuth/ propose a market driven monetary postage based system.[xix] These systems attack spammers at the beginning and make it arduous for them to create spam in the first place, unfortunately money stamps have proven to be quite problematic.

The first problem is who owns or controls a mailbox. If an ISP owns the mailbox, they can set the rate that a delivery of mail costs. One can imagine how a money-driven incentive can turn bad. ISPs would make deals with advertisers for low-cost delivery of advertising to your mailbox.

The second problem with monetary based postage systems is where the money goes. When you send a letter through your local post office, the postal service receives payment to cover the cost of the transportation of that piece of mail. In the email context, payment should be for the recipient's time. In this context it makes sense because the recipient declares how much their attention is worth. However, Microsoft's Bill Gates, among others, is suggesting computer users start buying "stamps" for e-mails and the monies go to the ISPs.[xx] Again, the ISPs would be a partial authority and your inbox would probably be flooded with advertisements and solicitations from the highest bidder.

The third issue is that charging a monetary amount for email goes against the premise of free exchange of ideas. "It detracts from your ability to speak and to state your opinions to large groups of people," said David Farber, a veteran technologist who runs a mailing list with more than 20,000 subscribers. "It changes the whole complexion of the net."[xxi] Requiring money stamps

everywhere in the world would have a serious impact on person-to-person communications. Most of the monetary-stamp proposals are largely U.S.-centric, and even with seamless currency conversion, paying even a token amount would be burdensome for the developing world, said John Patrick, former vice president of Internet technology at IBM Corp. "We have to think of not only, let's say, the relatively well-off half billion people using e-mail today, but the 5 or 6 billion who aren't using it yet but who soon will be," Patrick said.[xxii]

Further complications of a monetary system are caused by taxation. Redemption of stamps will create income that will trigger tax consequences that will be impossible to regulate on a global scale. Other complications arise from currency standards and conversion. If money stamps are going to be used, then there must either be exchange rates between currency formats, or a single format must exist for all issuers. There is also the problem of double spending, where spammers will cheat and try to use a stamp twice. The only way to effectively prevent double spending is by a central authority checking, and again, who will serve as this central global authority?

Liability for the accuracy of validation and redemption information is a problem as well. If there were no legal liability then Spammers would create their own stamp issuing operation and corrupt the system by spreading free money stamps. Next, the spam stamp issuer would simply declare all the stamps to be valid thereby circumventing the reason the system was set up in the first place.

Another major complication with money stamps is the peak-loads of stamp validation.

> Consider 9 AM on a Monday morning. Everybody gets into the office, turns on their computer, starts reading email, and stamp issuers' infrastructure begins to crumble under the load of millions of people validating stamps. This scenario raises the question of what happens if a stamp can't be validated. Is the message let through, which means you will get spam? Or is it delayed, which means you may not see potentially an important email until too late? The question of liability arises in this context too. Does the stamp issuer have a legal responsibility to provide sufficient capacity for timely peak-load validation?[xxiii]

These peak-load issues give spammers the upper hand and entice them to apply denial-of-service attacks against issuers' validation infrastructure. That way, either spam gets through, users get disgusted with the system and stop using money stamps, or the issuer is driven out of business. In any case the spammer gains the upper hand.

In order to be effective, a monetary-based system would have to be implemented across the Internet at large. It would need to be treated as a regulated entity such as the securities industry. History has shown that the securities industry is subject to corruption due to a lack of strong enforcement from the outside and greed from the inside. Unfortunately, a money stamp system would be subject to similar corruption because of the money flowing through such a system, the expense of accurate record keeping and the temptation to hackers. Worst of all since it is a gatekeeper to speech, it would be

corrupted by political pressures to silence those who would speak out against governments and corporations.

Proof of Work Postage and Why it Succeeds:

The open-source software Camram is a content neutral, anti-spam system updating the concept of postage to an electronic peer-to-peer model of postage. Camram incorporates open-source tool Hashcash, available since about 1997, as the core of its stamp services.[xxiv]

Camram is not a pure sender pays anti-spam system like those mentioned above. It is a hybrid system that uses multiple anti-spam techniques providing benefits to the very first user, increasing as the system spreads. A hybrid sender-pays system lets you incrementally introduce an anti-spam device that will take a serious chunk out of the economic foundations of spam.[xxv] If a piece of mail comes into your network and there is no stamp then the system does not recognize the sender and the mail goes to a content filter. If the mail does not pass the filtration process then the piece of mail is "jailed" and the sender will be notified of such. If the sender wants that piece of mail to get to the "stranger" then they need to spend the time and create the stamp. Once the sender is known the sender is added to an automatically generated white list and stamps are no longer needed for verification.

In the Camram system, whenever you send an email you generate an electronic postage stamp. Unlike physical postage, you don't spend money. You spend time by solving a puzzle whose solution becomes your postage stamp. When an individual requires their email to have postage, the economics of Spam change. A single stamp or computation is not very time consuming per email but it takes just enough time to cut into the way a spammer makes a living. The CPU overhead per email is negligible to a typical email user with an entry-level desktop or laptop because they don't send that many emails. At worst the email would be delayed a few seconds before being sent on slower, older hardware. But to spammers this is a show-stopper: they want to send 10,000 emails per minute down a DSL line bought with a stolen credit card quickly before the account gets cancelled.[xxvi]

In addition to stamp and white list based sender validation, Camram incorporates a different type of filter then other Spam filtration systems. It is based on the CRM114 discriminator.[xxvii] CM114 is an open-source product that comes as a self-contained component. The advantage of using CRM114 is that it has a greater degree of functionality than developing a Bayesian filter from scratch. The filtering is extremely accurate, even without the generation of stamps as a first line of defense. In a test scenario with my network, we used CRM114 with just automatic white listing while the stamping mechanism is in development. CRM114 classifies messages with a scaling numbering system based on statistical analysis. These numbers correlate to color codes: green is good mail, red is spam and yellow is undetermined. The scaling system is adjustable as the system starts to learn. Mail that is coded red and scores over a

certain administrator-set level gets automatically placed in a dumpster folder. Mail that is green is delivered. Mail that is coded yellow or red under a certain score gets placed in a "container" folder viewable by a web browser.

The first week required administrative intervention at multiple times through out the day to make a determination of what mails were legitimate. Most of them that were coded yellow (undetermined) were difficult to determine even after viewing them. Most were marketing emails that users had requested. The emails that were coded red and marked under a certain score were mostly spam but it was comforting as an administrator to affirm that the filtering was being trained properly. After the first week, the configuration could be adjusted down. Within a few weeks administrative intervention was down to just once a day to clear out the container. Since there are multiple layers of classification it was very simple to reclassify mail or tailor the scores up or down depending on the results. Furthermore, as the white list grew, mail that was from friends, family and business associates bypassed the filtering completely. Our test proved the independence of the anti-spam components as yet another advantage whereby Camram does not require a full radical change of your email environment and can grow incrementally. Once the stamping mechanism is in place the accuracy will be even closer to 100% and administrative intervention will be held to a minimum at the first use.

The hybrid sender pays system proposed by Camram is not without its challenges. One challenge is spammers trying to generate stamps faster using faster hardware or utilizing zombies as a computer array.[xxviii] The solution is to increase the bits in the stamp and/or change the puzzle thereby devaluing the spammer's stamp generating process.

The challenge regarding zombie computers is already an ongoing issue for propagation of spam. If used to generate stamps, the machines will become hot, slow and hopefully unreliable, setting off a red flag to the user that the machine has been compromised and needs to be addressed. Furthermore, much of these Trojans and worms that attack machines are spread through spam or mail sent by "strangers" in the first place. As the hybrid sender pays system spreads these emails will be isolated out, thereby gradually eliminating the risk of a zombie.

The other significant challenge is Moore's law of inflation. As systems get faster, proof of work stamps will consume less time. This is a legitimate weakness of a CPU-based proof of work system for postage.[xxix] The answer is to build a postage system where individuals can increase the size of the stamp based on spam exposure. This will make stamps more difficult to generate as time goes on and stay ahead of the spammers. As soon as postage rates start climbing on legitimate mail that is being received, it is time to raise your own rates. The advantage of Camram is that the whole system is incrementally scaleable to meet the needs of prevention at the time. Proof-of-work stamps are scaleable as the CRM114 filtering is.

Conclusion:

Spam is an evolving security risk and resource problem that needs to be fought through many techniques. Due to the peer-to-peer nature of email it is exceptionally difficult to filter out. Unlike physical mail systems, there is no gatekeeper to set standards or raise charges for commercial email (spam). A spammer can get into a network from any point and deliver their payload. Many attempts have been made by anti-spam systems to filter mail messages to eliminate spam. These techniques do not stop spam from coming into the network in the first place nor significantly drop the economic incentive to send spam. Camram is a complete solution incorporating a hybrid sender pays model that makes Spam expensive to send, easily identifies emails coming from known parties and attains close to 100 percent accurate filtering through its combination of techniques.

---

[i] Licklider, J.C.R. and Vezza, Albert. "Applications of Information Networks." Proc of the IEEE, 66(11). November 1978. URL: http://livinginternet.com/e/ei.htm. (March 2004).

[ii] "Brightmail Reports on Spam Trends of 2003." December 18, 2003. URL: http://www.brightmail.com/pressreleases/121803_spam_2003.html. (March 2004).

[iii] URL: http://asrg.sp.am/about/old_charter.html. (March 2004).

[iv] Graham, Paul. "A Plan for Spam." August 2002. 1. URL: http://www.paulgraham.com/spam.html (March 2004).

[v] Graham. 2.

[vi] Kontzer, Tony. "Research Watch: The U.S. Rules – In Spam." Information Week. March 1, 2004, 18.

[vii] Kontzer. 18.

[viii] Johansson, Eric. URL: http://www.camram.org (March 2004).

[ix] Graham, Paul. "Stopping Spam." 2-4. August 2003. URL: http://www.paulgraham.com/stopspam.html. (March 2004).

[x] Turing, Alan M. "Computing Machinery and Intelligence." Mind, Vol. 59, No. 236, 433-460, 1950. qtd. in Von Ahn, Luis. Blum, Manuel. Langforde, John. "Telling Humans and Computers Apart (Automatically) or How Lazy Cryptographers do AI." 1. URL: http://www-2.cs.cmu.edu/~biglou/captcha.pdf. (March 2004).

[xi] Greg Mori and Jitendra Malik of the University of California at Berkeley have written a program that can solve ez-gimpy with 83% accuracy (see Breaking a Visual CAPTCHA for details). Thayananthan, Stenger, Torr, and Cipolla of the Cambridge vision group have written a program that can achieve 93% correct recognition rate against ez-gimpy, and Malik and Mori have matched their accuracy. Their programs represent significant advancements to the field of computer vision. See also Gabriel Moy, Nathan Jones, Curt Harkless, and Randy Potter of Areté Associates have written a program that can achieve 78% accuracy against gimpy-r.

[xii] Von Ahn. 8.

[xiii] Pantel, Patrick and Lin Dekang. ``SpamCop-- A Spam Classification & Organization Program.'' Proceedings of AAAI-98 Workshop on Learning for Text Categorization qtd. in Graham, Paul. "Better Bayesian Filtering." January 2003. 1. URL: http://www.paulgraham.com/better.html. (March 2004).

[xiv] Sahami, Mehran. Dumais, Susan. Heckerman, David and Horvitz, Eric "A Bayesian Approach to Filtering Junk E-Mail.'' Proceedings of AAAI-98 Workshop on Learning for Text Categorization. qtd. in Graham. 1. (March 2004).

[xv] Schneider, Ken. "Anti-Spam Filtering Framework, URL Filtering." Spam Conference 2004, Morning 2. Massachusetts Institute of Technology. January 16, 2004. URL: www.spamconference.org. (March 2004).

[xvi] Asaravala, Amit. "Spam Wars: Filters Strike Back." Wired News. November 4, 2003. URL: http://www.wired.com/news/print/0,1294,61012,00.html. (March 2004).

[xvii] Asaravala.
[xviii] Johansson, Eric. URL: http://www.camram.org/camram_intro.html.
[xix] SpamSleuth actually incorporates the use of PayPal (www.paypal.com) and allows a recipient to configure and decide how much money will be charged before the program will allow the message through to the inbox.
[xx] "Gates: Buy Stamps to Send Email." Associated Press. March 5, 2004.
URL: http://robots.cnn.com/2004/TECH/internet/03/05/spam.charge.ap/index.html. see also Coursey, David. "Why Bill Gates' Anti Spam Plan Won't Work." ZDNet. January 30, 2004.
URL: http://reviews-zdnet.com.com/AnchorDesk/4630-7297_16-5119564.html?tag=print. (March 2004).
[xxi] "Gates: Buy Stamps to Send Email."
[xxii] "Gates: Buy Stamps to Send Email."
[xxiii] Johansson, Eric. URL: http://www.camram.org/why-money-stamps-fail.html. (March 2004).
[xxiii] The hashcash CPU cost-functions computes a token that can be used as a proof-of-work. Hashcash comes in the form of plugin software for mailers so that senders can add hashcash stamps to sent email. Hashcash plugin software inserts a X-Hashcash header into the email headers section of the email the user sends. Back, Adam URL: http://hashcash.org.
Johansson, Eric. URL: http://camram.org. URL: http://spamassassin.org. (March 2004).
[xxiii] Johansson, Eric. URL: http://www.camram.org/frequently-raised-objections.html. (March 2004).
[xxiii] Back, Adam. 3. URL: Http://www.hashcash.org/faq.html. (March 2004).
[xxiii] Yerazunis, William. URL: http://crm114.sourceforge.net/. (March 2004). CRM114 is a system to examine incoming e-mail, system log streams, data files or other data streams, and to sort, filter, or alter the incoming files or data streams according to user's preferences. Criteria for categorization of data can be by satisfaction of regexes, by sparse binary polynomial matching with a Bayesian Chain Rule evaluator, or by other means. Accuracy of the SBPH/BCR classifier has been seen in excess of 99 per cent, for 1/4 megabyte of learning text.
[xxiii] Johansson, Eric. URL: http://www.camram.org/frequently-raised-objections.html. (March 2004).
[xxiii] See also Dwork, Cynthia. Goldberg, Andrew and Naor, Moni. "On Memory-Bound Functions for Fighting Spam." That describes an alternative to CPU-based proof of work stamps using memory bound functions.
URL: http://research.microsoft.com/research/sv/PennyBlack/demo/lbdgn.pdf. (March 2004).