Webology, Volume 4, Number 1, March, 2007

Mystery Meat revisited: Spam, Anti-Spam Measures and Digital Redlining

Christopher P. Lueg
Professor, School of Computing, University of Tasmania, Australia. E-mail: christopher.lueg [at] utas.edu.au

Jeff Huang
M.S. student, Department of Computer Science, University of Illinois at Urbana-Champaign, USA. E-mail: huang6 [at] uiuc.edu

Michael B. Twidale
Associate Professor, Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign, USA. E-mail: twidale [at] uiuc.edu

Received October 21, 2006; Accepted December 15, 2006

Abstract

In order to protect email users from receiving unsolicited commercial email or spam, anti-spam measures building on technologies, such as filters and block lists, have been deployed widely. However, there is some evidence that certain anti-spam measures based on the purported origin of the spam cause unintended consequences related to issues of equity of access, which we term digital redlining. In this article, we revise and expand earlier work looking at secondary effects of anti-spam measures.

Keywords

Email; Spam; Filtering; Blocking; Digital Divide; Digital Redlining

---

Introduction

Email is widely regarded as one of the most important services provided by the Internet. In the U.S., Internet users are increasingly choosing email for communication rather than the telephone (Haythornthwaite & Wellman, 2002, p. 6). However, the email communication medium is being threatened by spam (a colloquial substitute for the cumbersome but precise technical expression "unsolicited commercial email" or UCE). A large percentage of email traffic is now made up of spam messages. Telstra BigPond, a major ISP in Australia, reports that 80% of the email they receive is spam (Hayes, 2006). Li (2006) reviews the spam problem from a legal perspective.

Spam threatens email communication by clogging email inboxes, making it costly to retrieve email and to find genuine emails. People have also stopped using email because of the often offensive content of spam messages. Fallows (2004) reports that 29% of Internet

users participating in two nationwide (U.S.) studies stated that they use email less because of spam, 63% of email users said spam has made them less trusting of email in general, and 77% of email users said spam has made being online unpleasant or annoying.

In order to protect recipients from receiving spam, technical anti-spam measures have been deployed widely. Developing accurate anti-spam measures is technically challenging, since there is no such thing as a precise, technical definition of spam that can be utilized by spam filters. The problem is that by definition, the nature of unsolicited/unwanted commercial email also known as spam depends on the recipient's attitude towards receiving respective messages (Lueg, 2005).

For a number of reasons, innocent third parties can get caught in the crossfire between spammers developing new and more sophisticated ways to distribute spam messages and anti-spam measures trying to identify spam messages in order to get rid of them. This means perfectly legitimate email can be mis-classified as spam and ignored or never even seen by its intended recipient. A hint of the increasing complexity of developing and maintaining spam filters that do not cause 'collateral damage' is the fact that major email providers AOL and Yahoo are introducing ways to bypass their own email filters by paying certain fees (Colquhoun, 2006; see below for further discussion). Another anecdotal hint is an incident in September 2006 where genuine emails sent by the first author using his account at the Australian University of Tasmania were rejected by the Australian James Cook University (see below for details). The reason was that the former was blacklisted, for questionable reasons, by the popular anti-spam site SpamCop which was in turn used, apparently without further implementing additional checks, by James Cook University and other institutions around the world.

We became interested in issues of provenance analysis as a contribution to a better understanding of how spam operates, of the continuing arms race between spammers and the developers of spam filter algorithms, and as a way to contribute to improving those algorithms. However, based on a few examples and a growing set of anecdotal pieces of evidence, we have come to realize that there are much larger issues of equity and access to investigate.

Elsewhere (Lueg, 2004) we provided a comprehensive description of the emerging spam filtering situation. We argued that:

1. observations of second order effects of anti-spam measures should be linked to the broader discussion of the 'digital divide' and the different ways it manifests and
2. in the long term we could observe 'spammer migration' patterns. As anti-spam legislation is getting tougher in Western countries, such as the U.S.A., E.U. and Australia, there is a good chance that spam will increasingly be sent from or relayed in developing countries.

In this paper, we revise and expand this work on the basis of additional data we collected and evidence reported in the literature.

Intended effects of anti-spam measures: protecting users and resources

The primary intended effect of deploying anti-spam measures is a reduction of the amount of spam that end users have to deal with. Reducing the amount of net traffic devoted to passing around emails that recipients do not want to receive is also desirable, but end user attention is by far the scarcest resource and the one that should be optimized.

Most filtering algorithms process the information contained in the body of an email message (the part of a message users normally see) but can also consider information contained in the message header (mostly information used to transport the message and also the From and Subject fields). Typical examples of characteristics used to sort out spam messages are rejecting emails with terms such as "free porn", "XXX", "Viagra" or "Get rich quick". Many spam filters use a probabilistic approach known as Bayesian filtering, where each email is scored based on the likelihood that a set of words occurs in spam instead of legitimate email. Words in the content and headers that correlate with spam, or a large proportion of HTML text in the body will lower the score of an email. If the email's score exceeds a certain threshold, it is deemed to be spam. In this paper, we focus on the header information and how it can be used in spam filtering, but of course in reality, this information is combined with the analysis of the message body to make an overall decision.

Spam filters may target the origin of a message. If a message is sent through a mail server that is known to be operated by spam-friendly companies, then this raises the probability that the message is itself spam, and in some filters may be sufficient for the decision to judge it as spam. Of course spammers are aware of the existence of spam filters and much about their algorithms (certainly as much as is in the public domain, which is all the information that we have available for this analysis). Therefore spam messages may be designed to disguise their origin, such as a server known to pass on or originate spam - another example of the spammer / spam-filter arms race.

Related to origin-based filtering is the more radical approach of blocking. Blocking means a mail server simply refuses to accept any mail from certain servers, often according to blacklists shared on the Internet. Blocking approaches can use IP (Internet Protocol) addresses, domain names, mail server hostnames, and other information provided by the mail exchange.

Lists of alleged spam servers are shared on the Internet. Services such as the Spamhaus Block List, allow mail servers to check in real time if a server trying to deliver email has earned the reputation of a spammer. The Spamhaus Project (2003), the organization hosting the above mentioned list, describes their block list as follows: "The Spamhaus Block List (SBL) is a real time database of IP addresses of static spam-sources, including known spammers, spam operations and spam support services."

Anti-spam products often implement combinations of different filtering, blocking and learning techniques (see Metz, 2003 for an overview). Kaspersky Anti-Spam ISP Edition, for example, uses a combination of linguistic analysis, formal analysis of message characteristics and blocking based on blacklists and whitelists.

As mentioned in the introduction, a major problem is that by definition, the nature of spam is subjective; it depends on the recipient's attitude towards receiving respective messages (Lueg, 2005). For example, in a report on the spam problem and how it can be countered, the Australian National Office for the Information Economy defined spam as "unsolicited electronic messaging, regardless of its content" (NOIE, 2002, p. 7). The report explicitly mentions that "arriving at an agreed definition of spam is a potentially contentious issue, as the direct marketing industry, ISPs, spammers, blacklisters and privacy and consumer groups have their own interests and views."

Unintended consequences of anti-spam measures

In this section, we look at spam filtering with a focus on the impacts on reliability and some wider consequences of filtering. So far the evidence for these impacts is often

anecdotal. We start by discussing why collecting quantitative data demonstrating impacts is extremely difficult.

Why finding out about secondary impacts of spam filtering is so difficult

From a spam filtering point of view we can use two main information resources for studying secondary impacts of spam filtering. First, we can analyze block lists and the IP addresses they list. Technically this is rather straightforward and we will cover results in the next section. Second, we can analyze spam messages and where they (allegedly) come from. In addition, it is necessary to take into account information from other, non-technical sources to understand how spamming is affecting the networked world.

Analyzing email in the context of spam filtering

For a number of reasons including avoiding legal liability and protecting their valuable infrastructure, spammers try to disguise the true origin of their messages. Naturally this behavior means that collecting data regarding secondary, unwanted effects of anti-spam measures beyond false-positive rates (which are computed locally) is rather difficult.

Historically speaking, email communication was never designed to be secure and as a consequence, email headers can easily be manipulated. In regards to spam messages, this means it is difficult to extract reliable data about the originating host as opposed to the delivering host. The difference has not received much attention in the spam fighting literature as the difference does not matter if the objective is merely to identify spam. Most anti-spam developers are mainly interested in identifying spam at the time it knocks on their front door (mail server), and so in terms of provenance mostly focus on what is known about the last external server to relay the message to them (see e.g., Goodman, 2004). Similarly, blacklists merely list prospective spam delivering hosts but do not attempt to provide information about originating hosts.

Although our interests as researchers in identifying spam are similar to those of anti-spam developers, our concerns are broader, extending to issues of the efficacy and equity of anti-spam measures undertaken. This requires gaining a richer picture of the true origins of spam and thus to what extent the earlier history of the message's progress had been fabricated. The difference between the originating host and the delivering host is crucial when investigating secondary effects of spam filtering. Knowing where spam really comes from is useful in many approaches to prevent it, limit it or mitigate its consequences. As noted, the provenance of an email message can be a contributory piece of evidence in a judgment about whether to treat it as spam by various filtering algorithms. Unfortunately spammers are aware of this, which is why they take measures to try and conceal a message's true origins.

Identifying which header information is actually reliable is crucial in the context of this research. In what follows, we highlight some of the difficulties we face when trying to determine the originating spam host as opposed to the delivering host. See Huang et al. (2006) for a more detailed analysis of technical challenges when analyzing SMTP (Simple Mail Transfer Protocol) headers used on the technical level of email communication.

Identifying fake Received headers in email

The immediate technical challenge we face is that email is relayed from one server to another. Typically, each server is operated by a different, autonomous organization. Each server forwarding an email to another server tacks on a Received header indicating which server passed the email message to it. Together these headers are a list of simple unprotected text strings, detailing the route of the email, with the most recent server first. The problem is that it is possible for a rogue email server to falsely claim that it received the email from another computer. This can be done rather easily by inventing and adding prior Received headers. These faked headers may be from real or nonexistent servers.

Since there is no direct way of knowing whether or not a Received header is legitimately added or not, it is generally accepted that there is no fool-proof method to find the origin of an email.

The only way to find out about validity of at least some of the headers provided in an email is to identify the last external server that handles the email before passing it to an internal server (Goodman, 2004). Assuming this home organization's internal servers are trustworthy, that IP cannot be falsified because it was reported by an internal server which is controlled by one's own organization. Because of the lack of a fool-proof method to find the originating IP address of an email, the last external server is often the only one used for tracking and reporting purposes. Consequently, the origin of an email is generally not examined. This is usually good enough for spam filtering but it is not sufficient when analyzing the wider impacts of spam filtering.

If a faked Received header is identified in the list, this indicates that one of the servers claiming to have subsequently received the message intentionally tried to foil attempts to identify the source of the email. Typically the guilty server is the actual originator of the email. For example imagine that we receive an email with header information including the following (for a more detailed discussion see Huang et al. 2006):

```
Received: from Echo by Destination
Received: from Delta by Echo
Received: from Charlie by Delta
Received: from Bravo by Charlie
Received: from Alpha by Bravo
```

The first line is written by our own home mail server describing the last email reception. Echo is our own internal server which we decide we can trust. If we somehow determine that Bravo is a fake entry, and we trust Echo, then we must suspect that either Charlie or Delta are the true origin and that they faked Bravo (and it's very likely that they also faked Alpha).

Each received line usually includes ('from') the host name reported by the sending server through the HELO command, followed by a pair in parenthesis -- a host name obtained by the receiving server by performing a reverse lookup on the IP address, followed by the actual IP address of the sending server. Following this is the host name of the receiving server ('by'), and a ('with') token followed by some additional information about the SMTP server and an ID for tracking purposes, and finally some time information (following a semi-colon). For example:

```
Received: from relay7.cso.uiuc.edu (relay7.cso.uiuc.edu
[128.174.5.108]) by expms6.cites.uiuc.edu (MOS 3.4.8-GR) with
ESMTP id BDH13397; Sat, 21 Jan 2006 02:22:39 -0600 (CST)
```

In practice however, SMTP servers do not always use this format. Additionally, a forgery may be done well enough that it is impossible to tell whether or not a header sequence is fake. Fortunately, many attempts to forge Received headers are imperfect and can be detected. There are several other ways a false Received header can be identified, including identifying illegal IP addresses, IP addresses in Received headers that are from IP blocks that have yet to be allocated, and any Received lines below a faked Received line.

It is also possible to discover that a fake header has been inserted if there is a broken link in the chain of Received headers. For example,

```
Received: from exserver.chgh.org.tw ([203.69.196.131]) by
expansionpack.xtdnet.nl (8.11.6/8.9.3) with ESMTP id
```

> fA9LD9m04845 for <foo@bar.com>; Fri, 9 Nov 2001 22:13:10 +0100
>
> Received: from mcpeely.concentric.net (0-1pool5-38.nas2.los-angeles1.ca.us.da.qwest.net [63.233.5.38]) by internation.co.uk (8.11.0/8.9.3) with SMTP id gAJEYgl19984 for <foo@bar.com>; Fri, 9 Nov 2001 20:29:03 GMT

Most likely, the second Received header is fake because there is no Received header to show how the email was moved from internation.co.uk to exserver.chgh.org.tw. Therefore, we suspect the second Received header is fake and deem exserver.chgh.org.tw as the (spoofing) origin of the email. However, again one must watch for a mismatch caused by internal relaying, although that is unlikely in this case.

Sometimes, even though it is impossible to definitively tell whether a Received header is fake, it might be possible to spot certain suspicious attributes that indicate a spammer may have tampered with the header. With an accumulation of such suspicious attributes, we may choose to determine that the header and thus the whole message is fake, but we must be prepared to acknowledge that perfectly legitimate messages may also have these features. Such hints include Received headers that don't follow the format defined in the RFC (Request for Comments) specifications (RFC 2822), time zone mismatches between the receiving server and the timestamp, geographical jumps in the message path, and finding dynamic hostnames for transit servers in the path.

Algorithmically, the process of finding the originator of the email is fairly straightforward: starting at the top of the set of Received headers, process down until either the end of the list, or a false or suspect header is found. However, as described earlier, determining which information is correct (or at least trustworthy) is a challenge.

An increasing amount of spam however is no longer relayed but delivered directly (Hoffmann, 2002). By direct delivery we mean that the delivering host is also the (alleged) originator:

> Received: from haticeg (p169.net220148067.tnc.ne.jp [220.148.67.169]) by filter.it.uts.edu.au (Postfix) with SMTP id 40214DF379; Thu, 1 Dec 2005 05:31:46 +1100 (EST)

The reason for the decline of third-party relaying is that most open relays were shut down because of abuse by spammers. Direct delivery of spam points towards hosts that are infected by bots remotely controlled by a bot-master. As Kaspersky Labs note:
"Using open relay and open proxy servers is [. . .] time consuming and costly. First spammers need to write and maintain robots that search the Internet for vulnerable servers. Then the servers need to be penetrated. However, very often, after a few successful mailings, these servers will also be detected and blacklisted." (from www.viruslist.com)

As a result, today most spammers prefer to create or purchase so-called bot networks:
"In 2003 and 2004 spammers sent the majority of mailing from machines belonging to unsuspecting users. Spammers use malware to install Trojans on users' machines, leaving them open to remote use. [. . .] Anyone who has the client part of a program which controls the Trojan that has infected a victim machine controls the machine or network of victim machines. The resulting networks are called bot networks, and are sold and traded among spammers." (from www.viruslist.com)

Unintended consequences of spam filtering 1: limited reliability
The objective of using spam blocking techniques is to reduce a mail server's intake of messages likely to be classified as spam anyway. As a consequence messages may not be

checked thoroughly (or not at all) since they may be assumed to be spam simply because they originate from alleged spam sources. This saves processing time, but of course such a coarse filter opens up the method to false-positives.

Typically, blockings are based on information coming from blacklists shared on the Internet. There is anecdotal evidence that blocking not only reduces the spam intake but also may cause "collateral damage" undermining reliability of email communication. The Spamhaus Project (2003) being a major host of such a blacklist notes the possibility of adverse effects as follows: "Can the SBL block legitimate email? The SBL's primary objective is to avoid 'collateral damage' while blocking as much spam as possible. However, like any system used to filter email, the SBL has the potential to block items of legitimate email if they are sent from an IP under the control of a spammer or via IPs belonging to spam support services. The chances of legitimate email coming from such IPs are slim, but need to be acknowledged."

The SpamCop web site makes a stronger point: "The SCBL is an aggressive spam-fighting tool. By using this list, you can block a lot of spam, but you also may block or filter wanted email. Because of this limitation, one should strongly consider using the SCBL as part of a scoring system and explicitly whitelist wanted email senders (e.g., mailing lists and other IPs from which you want to receive email)."

The incident mentioned earlier of Australian James Cook University blocking genuine email from a fellow Australian university suggests that this recommendation is not always implemented:

> Original-Recipient: <name removed>@jcu.edu.au
> Final-Recipient: RFC822; <name removed>@jcu.edu.au
> Action: failed
> Status: 5.5.0
> Remote-MTA: DNS; smtp.jcu.edu.au
> Diagnostic-Code: SMTP; 554 Mail from 131.217.10.51 refused, see RBL server bl.spamcop.net
> [IP address 131.217.10.51 = corinna.its.utas.edu.au]

Reliability and accountability issues regarding block lists were subject to a controversial discussion in the Usenet newsgroup news.admin.net-abuse.blocklisting, comprising more than 100 statements (see Blue, 2003 for details). The net-abuse related newsgroup news.admin.net-abuse.blocking often lists an abundance of requests by businesses renting IP blocks from major ISPs to be removed from block lists (see, for example, Rodriguez, 2003). According to their postings these businesses did not themselves send spam, but were blocked because their ISPs had a spam history or previous owners of their net blocks were spammers. Cole (2003) provides a detailed overview of why mail servers located in the net block (or IP range) of 'innocent' businesses may become 'collateral damage' and how these businesses may address the situation.

Another impact of spam filtering is reported by Fallows (2004): 30% of Internet users participating in two nationwide (U.S.) studies stated that they are concerned that their filtering devices may block incoming email and 23% of email users said they are concerned that their emails to others may be blocked by filtering devices. As we noted elsewhere Fallows' study does not clarify if the concerns voiced were based on actual false-positive experiences with email services or if concerns were rather unspecific.

One area that receives quite some attention is the question of accuracy or false-positives. If employing technical criteria typically used in the spam filtering community, some professional anti-spam technologies seem to be performing extremely well in terms of effectiveness and a low number of false-positives (i.e., genuine email falsely classified as

spam). MessageLabs, for example, stated in 2003 that their spam filtering technology achieves 96.4% effectiveness and 0.04% false-positives. Advertising slogans such as "Stop the emails you don't want - before they come anywhere near your corporate network" imply that spam filtering is a straightforward business. However, Balvanz et al.'s (2004) experiences with a number of spam filters built into desktop email clients typically indicate much higher false-positive rates than those published by some of the professional mail filtering services. The spam filter (brand not known) used by the first author's former employer certainly does not meet the above mentioned performance either: out of 81 spam messages received by a relatively new email account between February and April 2005 merely 21 were correctly identified as spam, resulting in mere 17% effectiveness (assuming that effectiveness resembles what is called precision in information retrieval). The number of emails dropped without notifying the recipient --if any-- is unknown. However, on a number of occasions, important genuine messages, such as program committee invitations and conference announcements, had been tagged as spam. Very low false-positive figures also appear to contradict Fallows' (2004) findings summarized above.

Unintended consequences of spam filtering 2: Digital Redlining

Secondary effects of certain anti-spam mechanisms, in particular origin-oriented filtering and blacklisting of mail servers, may not only have the unfortunate effect of falsely declaring some email as spam. They may also be disproportionately affecting the legitimate emails of certain disempowered groups, an effect that we have called digital redlining.

In an earlier publication (Lueg, 2004) we linked the discussion of secondary effects of certain anti-spam mechanisms to the digital divide discussion. In what follows we briefly summarize the point. The usual meaning of the term digital divide refers to inequality of access to the Internet (Castells, 2001, p. 248). The term is not only used to describe differences of global scale, such as differences in access between developed countries and developing countries.

Within the U.S., differences in access between demographic groups showing, for example, that rural and poor populations are underrepresented in Internet access and use have also been described as digital divide (e.g., Haythornthwaite & Wellman, 2002). Exploring Internet access in the U.S., Servon (2002, p. 4) argues that policy makers and the media have thus far defined the digital divide narrowly and incompletely by focusing on access in terms of possession or permission to use a computer and the Internet. Challenging this conception of the problem, Servon argues that deep divides remain, although underrepresented groups are making dramatic gains. In particular, Servon argues that deep divides remain between those who possess the resources, education, and skills to reap the benefits of the information society and those who do not. A different view is provided by Compaine (2000) arguing that there was a digital divide in the 1990s but that by 2000 the gaps were rapidly closing.

We believe the digital divide discussion is still highly relevant, but our evolving understanding of the sometimes subtle secondary effects of certain anti-spam mechanisms suggests that digital redlining as a broad, but currently only tentative concept is worth further investigation and validation.

Digital redlining: caught in the crossfire

We found ample anecdotal evidence that innocent third parties can indeed get caught in the crossfire between spammers developing new and more sophisticated ways to distribute spam messages and anti-spam measures intended to identify and remove spam messages. We do not argue that deploying anti-spam technologies is bad (there are excellent reasons for introducing spam filters and block lists even though they are not perfect) but we do

believe the secondary effects of these technologies need more attention than they have currently received.

An example is Varghese's (2003a) report of the following incident: "AOL says it is blocking email from Telstra's BigPond users because it has received complaints from its subscribers about spam being sent to them from BigPond addresses. Company spokesman Nicholas Graham said AOL had been [. . .] essentially compiling a whitelist of IPs from which mail would be allowed to reach AOL users." Since BigPond is a major Australian email provider, this means that for spam protection reasons, AOL blocked a significant part of an entire continent's email.

Hosted spam-filtering service SpamStopsHere beefed up their allegations reported in Lueg (2004) and now state on their web site "[i]t is well known that a huge amount of spam originates in China, Taiwan, Brazil and Argentina. It is clear to us that there are businesses in these countries that primarily just send spam. South Korea has a huge number of "open relays" which are computers that have been hacked and are controlled by spammers. Therefore, blocking email from countries notorious for sending spam is an effective filtering method. Which countries to block, if any, is a business decision. Click here for a discussion of why China, Taiwan and South Korea should be blocked."

The idea that whole countries should be blocked because a lot of undesirable activity originates there (or appears to originate there - as noted above, the origins can be spoofed) raises many obvious questions about access and fairness. Its similarities to the redlining practices of U.S. banks discriminating against African Americans living in poor areas of cities (and so contributing to the rising problems of those areas) inspired us to use the provocative term 'redlining'. As with the original redlining controversy in the U.S.A., we want to raise a debate about whether such blocking truly is, or should be regarded purely as a business decision, or whether it also raises important and perhaps overriding issues of equity.

Reviewing postings to the Usenet newsgroup news.admin.net-abuse.blocklisting reveals an abundance of requests to be removed from block lists (see, for example, Rodriguez, 2003). Often, these requests are coming from businesses renting IP blocks from major ISPs. Even though these businesses did not spam themselves, they were blocked because their ISPs had a spam history or previous owners of their net blocks were spammers.

Digital redlining: issues of economic power

Regarding AOL having decided to blocked Telstra's BigPond (Varghese, 2003a), and so a significant percentage of a continent's email, it is interesting to note that Telstra was able to get this draconian anti-spam decision overturned within a week (Varghese, 2003b). After all, Telstra is a large and powerful company, based in an industrially advanced country. A less powerful company based in a less advanced country may have had more difficulty even if it were equally innocent (or indeed culpable).

As mentioned in the introduction, global email players AOL and Yahoo are introducing ways for business to have their email bypass AOL and Yahoo customers' email filters by paying certain fees (Colquhoun, 2006). This may be interpreted as a hint of the increasing complexity of developing and maintaining spam filters that do not cause 'collateral damage'. It is perhaps an admission of failure - that AOL and Yahoo are unable to design truly effective filters that do not cause collateral damage and so unfairly discriminate - in this case against corporations with power and money to complain and seek redress. Of course it also means that the filters are unable to distinguish clearly between what is very widely regarded as spam, and what certain companies may regard as perfectly legitimate kinds of contact with their actual or potential customers, solicited or unsolicited. This of course also gets back to the vexed question of the subjective definitional nature of spam.

Finally, it must be tempting for AOL and Yahoo to convert the costly problem of creating spam filters into a potentially revenue generating option (by perhaps claiming a small fraction of the fee as a 'convenience charge').

After all, why should these companies keep making the effort of adjusting existing filters in order to minimize collateral damage if there are ways to bypass them? It seems difficult to argue that this is not a considerable step toward a future where successful Internet communication becomes in some way a function of economic power. This is particularly worrying if one considers gaps in economic power around the world. A hypothetical fee of $0.01 USD per email filter bypass is not much in economically well-developed countries whereas unless care is taken, it may be a major hurdle in less developed countries.

Discussion: Is Digital Redlining Real, growing, diminishing or changing?

In order to find out more specifically about blacklist-related effects we investigated two separate blacklists, the CBL blacklist and the SORBS blacklist and compared geographic data from the blacklists with geographic data from a large spam corpus. Both blacklists are fully automatically generated by watching spamtraps and other means. In both instances, we found that western/developed countries such as the United States, Japan, Canada, and European countries were on blacklists disproportionately compared to rapidly industrializing countries like China, South Korea, Brazil, Argentina, or Taiwan. Developed countries are sending much less spam, especially more recently (2004-2005), but still appear on blacklists much more. A look at the SORBS blacklist data updates from 2003 to 2005 at 6 month intervals showed that although the proportion of spam coming from developed countries were decreasing, their appearance on the SORBS blacklist remained constant. This goes against the suggestion by SpamStopsHere (as detailed above) and also against a coarse "digital divide" view suggesting that developing countries are being unfairly blacklisted as a whole, compared to more powerful developed countries. In fact, we can almost say that developing countries get away with more.

Considering the discussion of Telstra's BigPond being blocked by AOL and how quickly this was resolved as well as the difficulties smaller companies face when trying to remove blacklist listings, it seems that it is not so much geographic location but (again) economic power that influences redlining decisions.

What we found difficult to assess is how the rise of bot nets mentioned earlier have impacted the trends observed. From the perspective of equity and digital redlining, the question is where infected machines are located and whether there is a link between infection rate, blacklisting and economic strength of the respective geographic area. One might expect that computers in economically weaker regions are easier to infect due to the cost of keeping software and hardware up-to-date. However, the most infected operating systems seem to be Windows 2000 and Windows XP systems i.e., relatively new and resource-hungry operating systems may be more likely be found in economically stronger regions than in economically weaker regions. Looking into blacklists we actually found a bias towards economically stronger regions, and bot infections may explain the pattern.

In the long term, anti-spam measures deployed in Western countries to protect them against spam may contribute to excluding those who are located in economically less developed countries unwilling or unable to prevent spamming. As anti-spam legislation is getting tougher in Western countries, such as the U.S.A., E.U. and Australia, there is a good chance that spam will increasingly be sent from or relayed in developing countries thus resembling the emergence of off-shore tax havens in the financial services world.

Patterns related to the relocation of spam sources observed in the past include the already well-documented third-party relaying and the (alleged) hiring of mail server capacity in other countries including China. A more recently observed phenomenon is 'spam island-hopping': "spammers use the domain names of small islands as web site links in spam campaigns to disguise themselves from spam filters that traditionally catch more well-known domains" (McAfee, 2006). Domains traced by McAfee range from the Isle of Man to Tokelau in the South Pacific.

Summary and Future Research

Even if there is as yet no hard evidence for its existence, we want to raise the issue of digital redlining as a potential threat to be considered as spam measures and countermeasures co-evolve. Issues of equity should be a measure in assessing policy and technological options in addition to the current focus on efficiency.

This paper provides two main contributions to the discussion of the impact of spam and anti-spam technologies:
First, we have chosen to focus on the provenance of spam; where it originates and the route it takes to get to the recipient. By contrast the majority of the spam analysis research to date has looked at the content of spam or what is commonly referred to as the last external server. Of course content is important, and perhaps is the most important part. But we believe that provenance can play at least a significant contributory role in understanding how spam and spammers operate and hence how best to develop measures to counteract it effectively and equitably.

Second, we have noted that due to the difficulty of obtaining accurate information, where provenance is used in existing filters, it is often used somewhat crudely. Existing filters focus almost exclusively on the last hop of the message, rendering that particular mail-bearer the subject of suspicion. This has clear consequences for the efficiency and effectiveness of spam filters. Far more importantly, certain anti-spam measures such as filtering and blocking, particularly when they use crude measures of provenance, while providing considerable benefit, can also create problems of access. Specifically, equity of access is affected where messages from certain sources are blocked excessively, unfairly and for longer than from other, more privileged sources. We believe that this risk of digital redlining needs more careful study.

At this stage we are not sure if digital redlining was ever a major problem (in the sense it would affect populations not companies, which is well-documented), if it once was, but is diminishing, or if it is transforming into new kinds of inequity as the nature of the measures and countermeasures of the concerned parties continues. The military metaphor of collateral damage; innocent and often disempowered third parties caught in a battle's crossfire remains potent, even if in this case it turns out to be a threat to be avoided or minimized rather than proving to be a current crisis to be addressed. The alleged involvement of organized crime in spamming (Hayes, 2006) and the related rise of bot networks (viruslist.com n.d.) adds further dimensions to the problem space to be investigated.

Our aim is to encourage discussion of the implications for equity of measures that may otherwise be adopted exclusively for reasons of economic expediency or in reaction to a widely accepted threat. As with so many policy decisions, agreement on the importance of addressing that threat does not mean agreement on the measures taken. The unintended negative consequences of such measures on the wealthy and the powerful lead to remarkably rapid corrections. The unintended negative consequences on the poor and those lacking influence must also be considered.

Acknowledgements

The authors would like to thank the anonymous reviewers for their valuable comments on this article.

References

- Balvanz, J., Paulsen D., & Struss, J. (2004). Spam software evaluation, training, and support: fighting back to reclaim the email inbox. *Proceedings of the 32nd Annual ACM SIGUCCS Conference on User Services*, Baltimore, MD, USA, 10-13 October 2004. pp. 385-387.
- Blue, T.M. (2003). Blocklist accountability, standards, who is policing blocklists. Posting to the Usenet newsgroup news.admin.net-abuse.blocklisting on 12 Aug 2003. Message-ID: <65ab995e.0308121542.4bccaede@ posting. google.com>
- Castells, M. (2001). *The Internet galaxy. Reflections on the Internet, business and society*. Oxford University Press.
- Cole, W.K. (2003). Blacklists, blocklists, DNSBL's, and survival: how to survive as a non-combatant emailer in the spam wars. A collection of frequently asked and too-often poorly answered questions. Retrieved October 20, 2006, from http://www.scconsult.com/bill/dnsblhelp.html
- Colquhoun, T. (2006). Fees give email a stamp of approval. *The Sydney Morning Herald* 02/07/2006
- Compaine, B.M. (Ed.) (2000). *The digital divide. Facing a crisis or creating a myth*? Cambridge, MA.: MIT Press.
- Fallows, D. (2004). Internet users and spam: what the attitudes and behavior of Internet users can tell us about fighting spam. *Proc. Conference on Email and Anti-Spam* (CEAS 2004).
- Goodman, J. (2004). IP addresses in email clients. *Proc. Conference on Email and Anti-Spam* (CEAS 2004).
- Hayes, S. (2006). Crime rings discover spam. *Australian IT*, 18 September 2006.
- Haythornthwaite, C., & Wellman, B. (2002). The Internet in everyday life. An introduction. In: B. Wellman, & C. Haythornthwaite (Eds.), *The Internet in everyday life*. The Information Age Series. Blackwell Publishing, Malden, MA.
- Hoffman, P. (2002). Allowing relaying in SMTP: A series of surveys. Internet Mail Consortium Report: UBE-RELAY IMCR-016. Retrieved October 20, 2006, from http://www.imc.org/ube-relay.html
- Huang, J., Lueg, C., & Twidale, M. (2006). Identifying forged received headers in spam. Technical Report ISRN UIUCLIS--2006/6+CSCW.
- Kaspersky (2003). Kaspersky labs now battling spam at the ISP level. Product announcement released 12/30/2003.
- Li, X. (2006). E-marketing, unsolicited commercial e-mail, and legal solutions. *Webology*, 3(1), Article 23. Retrieved October 20, 2006, from http://www.webology.org/2006/v3n1/a23.html
- Lueg, C. (2004). The hidden impacts of anti-spam measures and their contributions to the digital divide: An exploratory study. *Proc. 67th Annual Meeting of the American Society for Information Science and Technology*, Providence RI, USA, 13-18 November, 2004, pp. 176-183.
- Lueg, C. (2005). From spam filtering to information retrieval and back: seeking conceptual foundations for spam filtering. *Proc. 68th Annual Conference of the American Society for Information Science and Technology*. Charlotte NC, USA, 28 October-2 November, 2005. pp. 1313-1314.
- McAfee (2006). McAfee, Inc. sees increase in island-hopping spammers. Press release published 1 November 2006. Retrieved November 12, 2006, from http://www.mcafee.com/us/about/press/corporate/2006/20061101_173300_s.html

- MessageLabs (2003). MessageLabs Service Portfolio, p.5.
- Metz, C. (2003). Corporate antispam tools. *PC Magazine*. Retrieved October 20, 2006, from http://www.pcmag.com/article2/0,4149,849390,00.asp
- NOIE (2002). Final report of the Australian National Office for the Information Economy (NOIE) review of the spam problem and how it can be countered. Retrieved March 5, 2003, from http://www.noie.gov.au/projects/confidence/Improving/Spam/Interim_Report/contents.htm
- RFC 2822. Internet message format. Resnick, P. (Ed.), April 2001. Retrieved October 20, 2006, from http://www.ietf.org/rfc/rfc2822.txt
- Rodriguez, G. (2003). SPEWS blocking a range with mine included, how to get out? Posting to the Usenet newsgroup news.admin.net-abuse.blocklisting on July 22 2003. Message-ID <b7efc7c3.0307220944.3e1a4d00@ posting.google.com>
- Servon, L.J. (2002). *Bridging the digital divide*. The Information Age Series. Blackwell Publishing, Malden, MA, U.S.A.
- SpamCop. http://www.spamcop.net
- Spamhaus Project (2003). The Spamhaus Block List (SBL) Advisory FAQ. Retrieved October 20, 2006, from http://www.spamhaus.org/sbl/sbl-faqs.lasso
- SpamStopsHere. Retrieved October 20, 2006, from http://www.spamstopshere.com/antispam_howitworks.aspx
- Varghese, S. (2003a). AOL blocking BigPond mail because of spam. *The Sydney Morning Herald*, 04/29/2003.
- Varghese, S. (2003b). Telstra problems with AOL resolved. *The Sydney Morning Herald*, 04/30/2003.
- Viruslist.com (n.d.) The Evolution of Spam. Retrieved October 20, 2006, from http://www.viruslist.com/en/spam/info?chapter=153350530

Bibliographic information of this paper for citing:

Lueg, Christopher P., Huang, Jeff, & Twidale, Michael B. (2007). "Mystery Meat revisited: Spam, Anti-Spam Measures and Digital Redlining." *Webology*, **4**(1), Article 36. Available at: http://www.webology.org/2007/v4n1/a36.html

---
