

SPAM AND PERSONAL
DATA PRIVACY

Stuart J. Thorson
Christopher M. Sedore

Systems Assurance Institute
Syracuse University
and
Information and Computing Technology Group
The Maxwell School
Syracuse University
Syracuse, NY, USA

email: thorson@syr.edu
voice: 315.443.4742
fax: 315.443.1395

Draft prepared for the Korean Information Security Agency 2002
International Conference on Personal Data Protection

Abstract

In this paper we explore some of the impacts of spam on privacy. We define spam as unsolicited email messages sent in large quantity to recipients with whom there is no preexisting relationship that would legitimize such contact. After providing examples of varieties of spam, we introduce the notion of *privacy regimes* and emphasize that privacy is a normative as well as regulatory concept. Spam in many ways seems to run counter to our expectations of privacy and we discuss some of the technical mechanisms which spammers use to penetrate individual privacy. Finally, we conclude with an overview of approaches to reducing the negative impact of spam.

What is Spam

Definition

'Spam' typically refers to unsolicited mass communications distributed via electronic mail or posted in on-line discussion forums. This is to be distinguished from SPAM (all upper case), a trademark held by Hormel Foods for a variety of spiced ham. The origin of the label spam in electronic communication derives from a Monty Python skit from the late 1960s in which the edible variety of SPAM is referenced repeatedly and sung in a song in such a way to be annoying and disruptive [1], [2].

Spam and Personal Data Privacy



Figure 1: SPAM

From: xxxxxx [mailto:xxxx@xxx.net]
Sent: Wednesday, October 30, 2002 3:23 PM
To: Undisclosed.Recipients@xxxweb.com
Subject: Re:

We provide a 100% free service which lets you shop for a loan conveniently and securely from the comfort of your home. Tap into our huge network of lenders across the U.S. with different appetites for different types of credit/equity profiles. Even if you're currently working with another lender or have been turned down before, we can still help and structure a program that may just make some sense to you.

Our loan programs can get you the cash you need for:

- Debt Consolidation
- 2nd Mortgage
- Refinance
- Credit Repair
- Home Improvement

Funding borrowers with less than perfect credit is our specialty!

We can get you the loan you need.
Regardless of whether you have good or bad credit, we can help you.

Ready to get started?
Simply fill out the short 60-second form bellow!

GET YOUR FREE QUOTE
It's that easy!
We Make the Lenders Compete for YOUR Business!

Removal Instructions

Figure 2: Spam

Spam and Personal Data Privacy

While 'spam' lacks a single widely accepted technical definition, use of the term became common in its application to a large number of postings advertising legal services of dubious value on the distributed discussion network USENET [3]. A necessary, though not sufficient, characteristic of a spam message is that it is unsolicited. The fact that a message is unsolicited, though, is not sufficient to label a message as spam; additional factors must be considered. The most common of these factors include whether the message is commercial; whether it is broadcast to a large group; whether it is relevant and appropriate to the forum, mailing list, or individuals to whom it was sent; and its legitimacy.

There are two more precisely defined but narrower terms in use to describe unwanted email communication. Unsolicited Commercial Email (UCE) and Unsolicited Broadcast Email (UBE) each refer to specific types of unsolicited communication. UCE is, as the name implies, electronic mail messages of a commercial nature (usually advertising) that is sent to the recipient without her request. UCE may consist of a single message or tens of thousands of messages. UBE is a large volume of messages (generally at least hundreds of messages or postings) that may contain commercial, religious, or scam content.

The security, privacy, and economic concerns we will investigate are not specific to the more narrow subtypes of spam. For the purpose of this paper, we will define spam as *unsolicited email messages sent in large quantity to recipients*

Spam and Personal Data Privacy

with whom there is no preexisting relationship that would legitimize such contact.

Examples would include unsolicited advertisement for products or services, religious or political exhortations, and enticements to participate in schemes of dubious merit. Note that our definition would exclude categorizing such mass mailings as product recalls to purchasers as spam.

A recent Gartner report identifies four varieties of spam:

1. pure-trash spam (fraudulent schemes, invalid senders);
2. chain letters, urban legends and hoaxes;
3. honest folks trying to make a living (“junk mail”); and
4. occupational spam from colleagues [4]

Spam examples

To make this discussion a bit more concrete, we will now provide some examples of actual emails we have recently received which exemplify the varieties of spam mentioned above. The headers have been removed to protect the guilty.

Pure trash (fraudulent schemes) Example

Attn:The President/CEO. Dear Sir/Madam, Assalam Alleicum. I am Mrs. Maryam Spammer, widow to the Former Military Head of State in Nigeria, Late General Sanni

Spam and Personal Data Privacy

Spammer, who died suddenly as a result of Cardiac Arrest on 8th of June 1998. One early morning, I was called by my Late Husband General Sanni Spammer, who at that time was the Chief In Commander of the Army and the President of Nigeria. He conducted me round the apartment and showed me three metal boxes of money all in Foreign Exchange, my husband told me he was tousele the money for the settlement of his Personal Royal Guards on his self Succession Bid and campaigns.

Upon his tragic and unexpected death, the new Civilian Government of Chief Olusegun Obasanjo, has insisted on probing my family's financial resources and has gazettes all our properties, also. They recently seized all the known family's fund abroad with the Assistance of the British Government. It is only this money US\$300,000,000.00 (Three Hundred Million US Dollars Only) that he deposited with a security company vault, that they can not trace because the funds were deposited as (ANTIQUITY) African Art

Work from the National Commission for Museum and Monuments (N. C. M. M.) Nigeria, the family intend to use this money for investments purpose to enable the family start life all over again.

Therefore, the family is urgently in need of a "Very competent and investor participant" that we could entrust with the certificate of Deposit and (PIN) Personal Identification Number Code to help us remove the funds from the Security Company "Since no name were used in securing the vault.

I got your contact address from our Chamber of Commerce, Agriculture and Industry Office here in Lagos - Nigeria, if the proposal is acceptable to you, after getting the money out from the security company vault to your country, my family have agreed to offer to you 30% of the total sum for the kind assistance you rendered to us.

And in addition bank the family's own part of the funds and assist us in investing in your company (with my approval on project) as a front for us until the situation becomes more favorable for us to meet and discuss the way forward, most especially now that my elder son, Mohammed Abacha and I are under pressure from the Government, despite the fact that my family had already returned the sum of US\$3 Billion my Late husband's Account in the U.S.A, Europe and other countries. Please kindly state your

Spam and Personal Data Privacy

early response immediately on this email for further details on the logistics and modalities.

NOTE: I do not need to remind you of the absolute secrecy and confidentiality that this transaction demands. You are free to speak directly either with me or my lawyer Barrister Peter Johnsan on the phone number below. Communication is strictly on email only, but if it is necessary to speak to each other, you can then call us on phone. If you are not interested, please kindly reply me immediately to enable me search for another interested partner.

I await your kind reply.

Thanks and accept my regards.

MRS. MARYAM SPAMMER TEL: xxx-x-7763xxx

Variants of this message arrive several times per week in our mailboxes. Most reference Nigeria.

Chain letter Example

Hello Everyone, And thank you for signing up for my Beta Email Tracking Application or (BETA) for short. My name is Bill Gates. Here at Microsoft we have just compiled an e-mail tracing program that tracks everyone to whom this message is forwarded to. It does this through a unique IP (Internet Protocol) address log book database.

We are experimenting with this and need your help. Forward this to everyone you know and if it reaches 1000 people everyone on the list you will receive \$1000 and a copy of Windows98 at my expense.

Enjoy.

Note: Duplicate entries will not be counted. You will be notified by email with further instructions once this email has reached 1000 people. Windows98 will not be shipped until it has been released to the general public.

Your friend, Bill Gates & The Microsoft Development Team

Spam and Personal Data Privacy

This email did not in fact originate with Bill Gates or, to the best we could determine, the Microsoft Development Team. The US Federal Trade Commission (FTC) actively prosecutes chain letter spammers under existing legislation passed prior to the widespread existence of email spam.

Junk Mail Example

Offering FREE quotes for your refinancing loan, second mortgage, debt consolidation loan, or home improvement loan.

Rates are the lowest they have been in years, so click here for a FREE quote now.

The original of the above message was sent in HTML format! More about that below.

Occupational Spam Example

This example, which we have truncated so as to not identify the sender is especially illustrative. The email was originally sent to a secretary in the Maxwell School along with the request that it be distributed to all Maxwell faculty members. The secretary complied thus spamming the school. However she also copied administrative assistants in the School many of whom also forwarded the email to all faculty members thus multiplying the number of copies we received.

(I would greatly appreciate your forwarding this information to all Maxwell School members and affiliated faculty. -- Thanks.)

I am a research scholar in the fields of Shakespeare and Classical Greece, a doctor of English Literature and author of works on Classical Greece and Shakespeare.

Here follows a brief description of the 4 books on Classical Greece, which you may find to be innovative and of some interest. <Message truncated>

Finally, we should note that the above examples are of spam messages which, while annoying, are suitable to be read by people of all ages. This is a bit at odds with typical experience where an increasing percentage of the increasing amount of spam seems to promote so-called x-rated services. For example, a study recently conducted by the on-line research company Nara in Korea found that spam traffic had increased 21 per cent in the six months of their study and of the spam received, messages about adult web sites were those frequently received (quoted in [5]). A recent British study [6] claims that 17 percent of all Internet email is spam. We shall suggest below that, at least within certain privacy regimes, the unsolicited receipt of material which makes one uncomfortable is in and of itself arguably a violation of privacy.

How Spam Happens

The pervasiveness of spam is in part an outgrowth of how Internet email is handled. At a purely technical level, one of the weaknesses exploited by those that send spam is the complete lack of authenticity guarantees in the email transfer protocols used on the Internet. The primary protocol used for transferring email is the Simple Mail Transfer Protocol or SMTP. When messages are transferred with SMTP, there is no way for a recipient SMTP server to look at an incoming message and verify that the message came from the origin listed in the message headers.

Spam and Personal Data Privacy

Spammers take advantage of SMTP's weakness in two ways. The first, and simplest, is to find an 'open relay.' Open relays are SMTP servers that will accept mail from any client for any destination. This is akin to an electronic mailman to whom you can give any piece of mail, with any 'from' address and any 'to' address, and it will take responsibility for trying to deliver it. In the Internet case, spammers will often seek out open relays, queue hundreds (or thousands) of messages for recipients and leave the open relay to do the delivery. The open relay can help obfuscate the origin of the message and allows anyone with basic Internet connectivity to inject messages in to the system and make them appear legitimate because they will be delivered from a 'regular' SMTP server. This method is convenient because it can be easily done with low-bandwidth connections to the open relay, and requires little connect time (since the messages can be given to an agent for all the delivery work).

The second way is that spammers will directly deliver spam to a recipients SMTP server with forged (or nonsense) source information in the headers. Though many ISPs filter to try to prevent direct email delivery by subscribers, certain types of connections still allow this kind of access, and these connections can be abused by spammers to directly deliver spam.

Spammers also exploit SMTP on unsecured networks run by various organizations. They may gain access by wireless connectivity or network ports in public spaces and then use the organization's SMTP server as a relay to

deliver spam. These situations provide almost complete anonymity to the spammer.

In addition to exploiting technical weaknesses, spammers use other means to 'reach' Internet users. Some have attempted to turn spamming in to a legitimate advertising activity. They typically adopt somewhat more conservative standards, allowing customers to 'opt-out' of email lists and only send electronic mail to people who have requested it (note, though, that that request may simply have been part of a software license agreement or fine print in an online order form).

Other spammers use throw-away accounts from ISPs or free email providers. They sign up for free accounts, send hundreds or thousands of spam messages, and then abandon the accounts.

However it is not only the narrowly technical nature of Internet email which results in the growing occurrences of spam. Another factor is that, unlike physical mail, there is no per message pricing for email. This means that the cost of massively blasting out messages is not consequentially different from sending only one email message. Thus spammers find it economically profitable to spam even with what would appear to be absurdly low response rates. For example a recent *PC World* article quotes a spammer, Ronnie Scelson, who reports sending 80 million e-mail messages twice a month for an insurance sales person. Those messages generate 700 responses, for which Scelson is paid \$12 each. Of the 700 responses, the insurance salesperson is

able to convert about 400 into paying customers [7]. Imagine if each of those 160 million monthly messages had a five cent charge associated with it. Indeed something like this has been proposed by Internet pioneer Bob Metcalfe. “One reason that we have spam today is that it costs nothing to send spam. One way to really cut back on spam is to charge e-postage (quoted in [8]).”

While in this paper our direct focus is on how spam relates to issues of personal privacy and only indirectly in how to reduce spam, it is important to be aware that the pervasiveness of spam results from the open and freewheeling structure of the Internet. And thus we come up against possible tradeoffs between openness and privacy. It is to these issues that we now turn.

Spam and Privacy

Privacy Regimes

Privacy ‘rights’ can be viewed as being defined by a complex set of rules and norms that, taken together, provide the basis for expectations about what is private and what is not. A specific set of these rules and norms we refer to as a *privacy regime*. Even across democratic societies there is considerable variation in expectations about privacy. For example, the US and the European Union differ considerably in how they treat the confidentiality of health records. And, while the ROK has a national identity card program, such systems, at least up until the present, have been resisted within the US

largely on privacy grounds. Moreover, it is clear that the increasingly networked nature of national and global communications is posing challenges to pre-existing privacy regimes as well as calling into question the ability of national governments to regulate these regimes even within their own borders.

Two forms of privacy rules

As we consider some of the impacts of spam on personal privacy, it will prove useful to distinguish between those privacy rules which, following [9], can be characterized as positive and those which are negative. Positive privacy rules are those which empower an individual to act in some way of her choosing under an expectation of privacy. For example, in many societies a person can choose to do some things in the privacy of her own house which would be prohibited in a public setting.

Negative privacy rules, on the other hand, are those which identify a sphere within which a person can presume to be private. For example, the Korean Information Security Agency (KISA) asserts a protected sphere of presumed information privacy for Korean citizens when it writes:

User consent is necessary when the service provider intends to collect the user's personal information and provide it to third parties beyond the guidelines described in the Act or specified in the service contract. The user is entitled to control his own information and the service provider must first seek permission to divulge information to third parties [10].

Spam and Personal Data Privacy

Note too that KISA example illustrates that many interesting privacy rules embody both negative and positive aspects. In the example above, KISA establishes a protected sphere of data privacy (a negative rule) and then goes on to authorize individuals to divulge certain information *if they so choose* (a positive rule).

Spam and Privacy Regimes

Spam is problematic in that it can undermine existing privacy regimes without actually breaking any actual laws. The undermining takes place in large part because the regulatory (as opposed to normative) components of the regimes were largely established prior to the widespread deployment of modern advanced computing and communications networks. Thus laws often are out of synch with normative expectations about privacy.

For example, I may wish to, by default, hold private the IP address of my connection to the Internet and/or the times at which I am browsing the Web. Yet often transparent images, termed *Web beacons*, can be embedded in spam messages. Then if the image is merely previewed (that is, the HTML is rendered) or if it is opened, the Web beacon will report back the IP address of the receiving machine as well, perhaps, as information about the time the site was visited and how long the site was viewed. All this can happen without the user positively choosing to visit the site. As a response to this, some email clients will offer the opportunity to accept email messages only in text form.

Spam and Personal Data Privacy

More generally, one might reasonably expect reading email to be a private activity—no one should know whether I read my email, when I read it, whether or not I have forwarded it to a friend or associate, or what kind of computer or email client I used to read it. In fact spammers have employed a number of methods to track message efficacy. Many of these methods revolve around electronic mail clients that render HTML messages. HTML allows spammers to send more richly formatted messages, but it further allows them to cause the recipient's computer to download content (typically images) from Web servers on the Internet. This may seem innocent enough, allowing a company logo, for instance, to be displayed as part of a marketing message. The simple act of downloading an image allows spammers to know that a message was received and read, when it was read, and may reveal which email client was used in addition to other information such as geographic location, the user's Internet service provider, etc. Further, more sophisticated use can play a part in creating or enhancing an online profile of the recipient's browsing habits as she views different sites.

The ability to track the origin of a user's interest in a product is important to many spammers as a significant amount of spam is done by agents who are paid on a commission basis, as percentage of sales or fixed amount per respondent.

Spammers have also been known to send mass emails which make deceptive use of the logos and/or Web page designs of legitimate companies. For

Spam and Personal Data Privacy

example, the US FTC is now prosecuting a case against a spammer who has allegedly made unauthorized use of the logos of Radian Bank, Fannie Mae and the Prudential Bank and Trust Co. to induce people to reveal personal financial data. This particular form of deceptive behavior, known in the US as ‘pretexting’ is a violation of US law. In this instance, [11] reports that the spammer also used a fictitious return address with the result that undeliverable bounce-back messages went to email addresses that did not belong to the spammer. In one case, a non-involved person’s email address was “swamped with more than 30,000 bounce-back and angry ‘do not spam me’ e-mails intended for the defendants [12]”.

Spam pollutes privacy expectations

Privacy regimes are analogous to the ‘rule of law’ in that their value lies not simply in the regulatory scheme which establishes them but also importantly in the expectations they establish. With regard to the rule of law, for example, our expectations that contracts entered into today will be enforced into the indefinite future is what enables us to engage in many forms of commerce (one need only look at a society in which the rule of law is weak, such as Russia after the breakup of the Soviet Union, to see this). The case is similar with regard to privacy. My expectation that notes I make to myself or the times I read my email or my current location can, subject to the law, be kept private may enable me to feel comfortable in engaging in various sorts of creative or entrepreneurial activities. Web beacons and other related tactics

serve to pollute our expectations of privacy and, in so doing, may reduce our willingness to engage in certain forms of socially desirable behaviors.

Perhaps most importantly, expectations of privacy, as understood within particular privacy regimes, when satisfied are a part of what reinforces the degree which people trust one another. Low trust is not simply normatively undesirable. It also appears to have strong negative economic consequences. Consider that virtually any economic transaction has associated with it opportunities for deception and fraud of the types discussed above. In economic terms, lack of trust imposes transaction costs on exchanges. [13] found that degree of trust is a significant predictor of country economic growth rates; to the extent that roughly a 10 per cent increase in their trust measure was associated with a .8 per cent increase in growth. As [14] points out, this is “a sizable increment given world average growth rates of 1 per cent to 3 per cent in the latter half of the 20th century (p. 2)”. Interestingly enough, [14] also present data to suggest that increasing Internet adoption is associated with increased trust. Thus it would appear that the Internet is a valuable source of social capital. Thus the pollution of that source through such trust-reducing activities as spam should be looked at with some alarm.

What can be done?

Potential remedies here are several. First there are technical fixes which may reduce the probability of being cyber-stalked by spammers. For example, Web beacons rely on the rendering of HTML within email. Email clients

which accept only text formatted (that is, do not render HTML) messages are immune from this particular exploit.

Secondly there are varieties of commercial 'spam filters' that use artificial intelligence heuristics to weed out spam before it is delivered to your inbox. While some of these appear to be quite effective, they all suffer from both false positives and false negatives. That is, they occasionally let spam in and, perhaps worse, occasionally block legitimate email.

While the above 'solutions' are purely within the private sector, there are also people calling for a variety of legislative attacks on spam. One such suggestion is to require that electronic marketers identify advertising and x-rated messages within the subject header. Under such a law and assuming compliance, spam-filters could be much more effective in pruning incoming email messages. A related proposal would outlaw the use of such techniques as spoofing the sender's address and/or routing information.

There are those that would like to see so-called 'opt-in' legislation. Such legislation would require that email advertisers explicitly get the permission of a person before sending advertising material. Others have proposed a centralized 'do not send' database. E-marketers would be prohibited from sending advertisements to anyone whose email address had been entered into the database. This latter proposal is patterned on similar legislation which has successfully reduced the incidence of marketing phone calls in, for example, New York State.

All of these possible legislative remedies appear to suffer from a similar set of problems. First it is anticipated that it would be fairly expensive to prosecute these cases. Indeed in many instances it might even be difficult to establish jurisdiction. Spam frequently does not originate and remain within single geographically defined jurisdictions. Second, it can be difficult to recover damages from spammers. A spammer may generally not be expected to have the 'deep pockets' necessary to collecting on a large financial judgment.

Of course some, perhaps most notably Bob Metcalfe, quoted in [8], argue that much of the problem with spam results from distortions in the way in which Internet email is charged. Unlike physical or snail mail, the sending and receipt of email is not charged on the basis of use. Metcalfe has argued that metering email and charging micro-amounts per message sent would greatly cut down on certain sorts of spam. While Metcalfe's claim is plausible for spam which depends upon massive emailings with extremely small response rates, not all spam is of that kind. Moreover, while outside the scope of this paper, there are a number of technical and philosophical issues to be addressed before there would be wide-spread support for use-based charging on the Internet.

Conclusion

In this paper we have suggested a definition of 'spam' and described several spam variants. We than briefly considered some of the technical issues which have made spam so easy to send. Of course, spam, while annoying, is not

merely annoying; it often poses a threat to data privacy. As a consequence we examined some of the ways in which spam poses a threat both to personal privacy and, more generally, to privacy regimes. Finally, we discussed both private and public sector approaches to dealing with spam.

References

- [1] M. VonWald, "History of Spam." <http://www.techtv.com/callforhelp/stepone/story/0,24330,3387393,00.html>, 2002.
- [2] Hormel Foods Corp., "Spam and the Internet." http://www.spam.com/ci/ci_in.htm; Hormel Foods Corp.
- [3] B. Templeton, "Origin of the term "spam" to mean net abuse." <http://www.templetons.com/brad/spamterm.html>.
- [4] J. Graff, "Keeping Spam Out of Your E-Mail," Gartner, Inc. TU-15-0487, December 20 2001.
- [5] Arirang TV, "Spam Traffic Surges 21% in Six Months." <http://english.chosun.com/w21data/html/news/200210/200210230003.html>, 2002.
- [6] R. Lemos, "Spam grows--a bigger worry than worms." <http://zdnet.com.com/2100-1105-962300.html>; ZDNet News, 2002.
- [7] S. Fox, "Does Spam Pay off," in *PC World*, vol. 20, 2002, pp. 43.
- [8] V. Shannon, "Sending E-Mail- and Paying Postage?," in *International Herald Tribune*, 1998.
- [9] I. Berlin, *Four essays on liberty*. Oxford ; New York: Oxford University Press, 1979.
- [10] Secretariat of Personal Dispute Mediation Committee, "Personal Data Protection in Korea," Korean Information Security Agency August 2002.

Spam and Personal Data Privacy

- [11] L. Rosencrance, "FTC targets deceptive spam," in *Computerworld*, 2002.
- [12] Federal Trade Commission, "Federal, State, and Local Law Enforcers Tackle Deceptive Spam and Internet Scams." <http://www.ftc.gov/opa/2002/11/netforce.htm>, 2002.
- [13] S. Knack and P. Keefer, "Does social capital have an economic payoff? A cross-country investigation," *The Quarterly Journal of Economics*, pp. 1251-1288, 1997.
- [14] C. Keser, L. Jonathan, J. Schachat, and H. Huang, "Trust. the Internet, and the digital divide," IBM Research Division July 2002.