# All about "Spam"

Email continues to be by far the most important service for Internet users[1]. Billions of messages are transferred every day. Unfortunately, as every user is well aware, a proportion of this traffic is unsolicited and unwanted advertising. This is often called[2] "spam"[3] in memory of the Monty Python sketch from the 1970s[4] where a group of Vikings sat in a café and incessantly chanted "spam, spam, spam, spam…" and eventually drowned out all other conversation.

## *The nature and cost of spam*

Sending and receiving email is very cheap and it is tempting to see spam as an ecologically friendly alternative to the paper based "junk mail" delivered in daily batches to most households and businesses. However, the cost should not be measured in the fractions of a penny that it takes to transmit the bytes down the wire. The real costs of spam are in the time spent by users to sort through the unwanted dross, the size of mail system that Internet Service Providers (ISPs) must provide to store and forward the material, and in the delays to genuine messages caused by the overloading of systems[5]. These costs run, even conservatively estimated, into billions every year[6]

Individual emails are often sent "unsolicited", and there's no problem with that, the distinction that makes "spam" different – and makes it unacceptable – is that it is sent in bulk (sometimes many millions of copies at a time). It tends to be commercial in nature, often promoting products of dubious legality or taste, or "get rich quick" schemes designed to extract money from the gullible.

Some people like to try and distinguish messages sent by companies from other types of spam by using terms like UCE (unsolicited commercial email) rather than UBE (unsolicited bulk email). Others argue that different rules ought to be applied to religious or political messages, or that writing on behalf of a charity might make the

---

[1][1] Nielsen//NetRatings finds e-mail is the dominant online activity worldwide (May 9, 2002) http://www.nielsen-netratings.com/pr/pr_020509_eratings.pdf

[2] Purists would argue that "spam" should only be applied to the bulk posting of articles on Usenet, but the usage in relation to email is now so widespread that this has become a matter of etymological interest rather than a useful distinction.

[3] SPAM is a Registered Trademark of Hormel Foods Inc who have been manufacturing SPAM Luncheon Meat since the 1930s. They explain the legal complexities which occur when a trademark becomes a slang term at http://www.spam.com/ci/ci_in.htm.

[4] http://www.cs.berkeley.edu/~ddgarcia/spam.html#MontyPython

[5] Report to the Federal Trade Commission of the Ad-Hoc Working Group on Unsolicited Commercial Email. http://www.cdt.org/spam/

[6] The European Commission study of Jan 2001 put the cost at €10 billion worldwide for the connection cost alone. http://europa.eu.int/comm/internal_market/en/dataprot/studies/spamstudyen.pdf. Although this document has been criticised in some quarters, these figures are not out of line with other surveys.

senders behaviour acceptable. These fine distinctions in motive are seldom of interest to the recipients of the email and they are not recognised in law.

Others try and distinguish "good" unsolicited messages sent by responsible companies that belong to marketing organisations like the DMA[7] from "bad" messages sent by irresponsible mavericks, reserving the word "spam" for the latter. In practice, the situation today is that no responsible company currently risks the negative impact upon its reputation by sending bulk unsolicited email and future legislation will restrict their ability to do so. Thus this distinction is of little practical consequence.

### *Sending spam is not legal*

Almost all bulk unsolicited email is sent unlawfully. Individual email addresses are "personal data"[8] (only role based addresses such as sales@example.co.uk escape this definition) and under the UK Data Protection Acts of 1984 and 1998[9], there are strict limitations on the processing of personal data for unexpected purposes – such as sending unsolicited email to these addresses. In addition, every ISP will provide access to the Internet under contractual terms that will explicitly prohibit the sending of bulk unsolicited email, so those who send this material are breaking the terms of their contracts and run a significant risk of disconnection.

At present there is no specific anti-spam law in the UK. However, this will soon change because the European Union has recently approved an E-Privacy Directive[10] that must be transposed into UK law by 31st October 2003. This will make it unlawful to use email for direct marketing except where customers have given their explicit consent or where a company is sending email about similar products and services to an existing customer.

Some other legislation is relevant to spam. The E-Commerce Directive[11] required that any unsolicited commercial communication sent by electronic mail "is clearly and unambiguously identifiable as such as soon as it is received". The Electronic Commerce (EC Directive) Regulations[12] prescribing this came into force on 21st August 2002. Because the E-Privacy Directive has superseded the other E-Commerce

---

[7] Direct Marketing Association. http://www.dma.org.uk

[8] See "Internet: Protection of Privacy - Data Subjects (Jan 2000) version 4" linked from http://www.dataprotection.gov.uk/dpr/dpdoc.nsf

[9] Data Protection Act 1998, http://www.legislation.hmso.gov.uk/acts/acts1998/19980029.htm

[10] Directive 2002/58/EC of the European Parliament and of the Council concerning the processing of personal data and the protection of privacy in the electronic communications sector. http://europa.eu.int/eur-lex/en/dat/2002/l_201/l_20120020731en00370047.pdf

[11] Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market http://europa.eu.int/eur-lex/pri/en/oj/dat/2000/l_178/l_17820000717en00010016.pdf

[12] Statutory Instrument 2002 no 2013, The Electronic Commerce (EC Directive) Regulations. http://www.legislation.hmso.gov.uk/si/si2002/20022013.htm

Directive requirements on checking registers before sending unsolicited email the Government decided not to enact these requirements.

If the ECommerce Regulation on the identifiability of unsolicited commercial email is infringed then the remedy at law is "an action against the service provider for damages for breach of statutory duty" (the service provider being the sender of the email, not an ISP). Until there is a test case, it is unclear what this will mean in practice. However, since as already noted, sending unsolicited email (however identified) is likely to be forbidden by a sender's ISP, disconnection is likely to remain the most effective sanction available.

### *What's "opting -in and -out" about?*

In the jargon, the E-Privacy Directive scheme is called "soft opt-in" – which bears a little explanation. The Internet industry and consumers have, in the main, favoured "opt-in" regimes whereby email can be sent only if the email address has explicitly chosen ("opted-in") to receive messages. The marketing professionals (and a handful of ISPs) have generally favoured "opt-out" schemes whereby the assumption is that email may be sent unless a negative response is received or if the address is on a global "opt-out" list. This type of "opt-out" or "preference" scheme has had some success in dealing with postal mail[13] and telephone calls[14]. However, the economics of email are completely different, so that funding a universal register has proved to be problematic and the type of companies currently sending spam have never been responsible enough to use such a system. In fact the addresses on some "opt-out" lists have been specifically targeted by spammers.

This "opt-in"/"opt-out" distinction has sometimes been confused with the exact mechanism used on web pages for giving "opt-in" permission. This is because some people call the positive act of ticking the box "opt-in" and the negative act of clearing a tick "opt-out". As indicated above, when discussing spam, it is usual to call both alternatives "opt-in". In passing, it might be noted that the Information Commissioner's current advice is that the box may be pre-ticked for the disclosure of personal data such as contact information, but must require a positive act where sensitive personal data (such as health information) is concerned[15].

### *Where does spam come from ?*

The Internet is of course an international system and it will be apparent to anyone who examines the spam they receive that it seldom comes from the UK or indeed elsewhere in Europe. In practice, the vast majority originates in the USA. In part this is because of Europe's Data Protection laws, for which the US does not have any

[13] Mailing Preference Service, http://www.mpsonline.org.uk

[14] Telephone Preference Service, http://www.tpsonline.org.uk

[15] see "Data Protection Act 1998 Compliance advice, Website frequently asked questions (June, 2001)" linked from http://www.dataprotection.gov.uk/dpr/dpdoc.nsf

equivalent legislation. Also, although there have been some anti-spam laws passed in individual states[16], no federal law has been adopted[17] because of an inability to satisfy both the anti-spam campaigners and the marketing lobby.

However, the main difference has been the different structure of the ISP industry in the USA. Whereas in the UK most people get online through a small number of national ISPs, in the USA there are a large number of local ISPs and so a spammer who is disconnected from one ISP can quickly get online again through a competitor. Although in the past some otherwise reputable companies have been involved in sending spam, it has now become somewhat of a cottage industry with a relatively small number of operators sending out large amounts of material on behalf of others[18].

A few years ago the picture was different, and each year a small proportion of companies that connected to the Internet would naively decide to use spam to promote their businesses. Nowadays, the negative aspect of advertising by using spam is universally recognised so this "clueness newbie" aspect of spamming has almost ceased. There are occasional problems[19] where reputable companies purchase "guaranteed opt-in" lists from third parties which turn out, when used, to be anything but "opt-in", but in general, the marketing professionals now firmly believe in "permission based marketing" and the EU Directive will make little difference to their day-to-day plans.

### *Technical countermeasures*

Most people are aware that the amount of spam has increased over the past few years, though this is mainly because the same message is being sent many times, rather than because there are more distinct messages. This means that there is increasing pressure to develop technical measures to prevent spam from flooding mailboxes. These measures can be split into three types, barring it at the source, blocking lists of known sources and filtering it out at the destination.

#### Closing the "open relays"

As already discussed, the professional spammers use throwaway dialup accounts because any permanent connection would be rapidly closed by their ISP at considerable expense to them. Since the low bandwidth of these connections limits the amount of email that can be sent in a given time period, they use "mail relay"

---

[16] Summary of US State laws on spam, http://www.spamlaws.com/state/summary.html

[17] Various laws have passed either the Senate or Congress, but none have reached the statute book.

[18] "about 100 spammers produce nearly 90 percent of the junk mail sent today, but disguised addresses and other tactics make it difficult to link one of those spammers to a particular piece of mail" http://news.com.com/2100-1023-881979.html

[19] "Sainsburys, Virgin say sorry over spam", http://www.theregister.co.uk/content/archive/23497.html "Animal instincts", http://society.guardian.co.uk/internet/story/0,8150,913210,00.html

machines. A single copy of the email is sent to the relay along with a hundred or more destination addresses. The relay then delivers the email to the destinations and the spammer has amplified the bandwidth available to them by a hundred-fold.

So called “open relays”, that will relay email for anyone who asks were the norm in the early days of the Internet, and were important in ensuring that email could be delivered. However, there has been no necessity for them to exist for decades. By the late 1990s it was considered to be entirely anti-social to run an open relay because of the burgeoning spamming problem[20].

With the biggest and best connected mail machines barred to them, the spammers started to relay through corporate machines connected via leased lines, but security on these has also improved considerably over the past few years. Open relays of this type do remain common in the Far East and South America (mainly because of the lack of technical material in the local language to allow reconfiguration), and a lot of spam is still being relayed from the US spammers via these countries. A recent development has been the hijacking of individual broadband systems (ADSL and cable systems) and it is taking some time for ISPs world-wide to develop effective systems for spotting the large number of misconfigured customers involved.

## Blacklists

A relatively recent development has been the growth of “black-lists” of open relays and other spam sources[21]. The idea is to avoid accepting email from systems that are being actively exploited by spammers. The Internet is composed of co-operating but independent networks, so there is no right for anyone to have their email accepted and so mail servers owners can apply any acceptance rules they wish. Nevertheless, in some jurisdictions, the creators of the black-lists have faced some legal problems.

The black-lists vary considerably in their accuracy and the procedures needed to remove machines from them. There have been accusations of addresses being added through spite and a few court cases have been undertaken by aggrieved parties[22]. Some ISPs use these blacklists to discard email, some add markers to show it has been flagged by the blacklist, but the many believe that the risks of blocking legitimate email on the word of an almost unaccountable third party are unacceptable.

Blacklists have had some effect. The widespread use by US ISPs of simple-minded blacklists – blocking absolutely everything from China[23] or Korea has led both

[20] http://groups.google.com/groups?selm=342be783.0@muir-et2.staff.demon.net

[21] There are dozens of blacklist systems, but the best-known are the MAPS Realtime Blackhole List (RBL) http://mail-abuse.org/rbl/, the Open Relay Behavior-modification System (ORBS) which is no longer available and the Spam Prevention Early Warning System (SPEWS) http://www.spews.org/.

[22] http://list-news.com/articles/01june/20010605.html (ORBS)
http://www.internetnews.com/IAR/article.php/12_514611 (MAPS DUL)
http://www.isp-planet.com/news/2002/orbz_020321.html (ORBZ)

[23] “China fights spam blocks”, http://zdnet.com.com/2100-1105-851599.html

countries to take steps[24] to address their spamming problems so that their economy is not affected by an inability to correspond with the United States.

## Filtering systems

Filtering seems an attractive approach to dealing with spam, and it is widely deployed both on ISP systems and by end users themselves. The systems used vary from purpose built systems with complex decision mechanisms, to simple rules of thumb such as discarding all mail with more than three dollar signs in the subject line. These systems are often combined with virus detection mechanisms, since a proportion of unsolicited email is auto-generated by undetected viruses.

There are three main problems with filtering which mean that it will not be the long-term solution to the spam problem.

- The first problem is that of "false positives", the discarding of legitimate email[25]. When a large ISP discarded email from Harvard University accepting students onto their courses[26] it made front page news. Poorly tuned filters (that take exception to legitimates words such as Scunthorpe or Penistone) are far from being apocryphal and recently the UK Parliament introduced a system that managed to block legitimate messages relating to the Sexual Offences Bill[27] and those in Welsh[28] by deploying a simple-minded scheme[29].

- The second problem is that spam is already evolving to evade the filters. The same basic message will now arrive with differing subject lines, minor variations in the content and the source will have been forged by appropriating a real email address. As time goes on, we can expect to see a race between the filter creators who will have to become ever more subtle and the spammers who will continue to mutate their messages to avoid being blocked. Even if we are prepared to see far more false positives, the spammers will eventually win by making their messages indistinguishable from normal correspondence.

- Finally, the third reason that filtering is not the long term answer to spam is that whilst it is effective in preventing delivery to the end user, this is only a part of the problem. The messages will still travel a long way through the network and incur

[24] "Government to Crackdown on 'Spam Mail'" http://www.chosun.com/w21data/html/news/200201/200201130126.html

[25] "Email Filtering: Killing the Killer App" http://www.tidbits.com/tb-issues/TidBITS-637.html#lnk4

[26] "AOL glitch blocks e-mail notices from Harvard admissions office" http://www.modbee.com/24hour/technology/story/209046p-2015925c.html

[27] "E-mail vetting blocks MPs' sex debate", http://news.bbc.co.uk/1/hi/uk_politics/2723851.stm

[28] "Software blocks MPs' Welsh e-mail", http://news.bbc.co.uk/1/hi/wales/2727133.stm

[29] "UK Parliament Mail – The Ministry of Silly Messages", http://sethf.com/anticensorware/general/uk.php

bandwidth and processing costs, which the consumer will eventually have to pay through higher charges.

However, although the long-term outlook is bleak, filtering does offer significant short-term benefits to those who are prepared to take on some risk of having legitimate email discarded. Many ISPs are now deploying filtering systems on their servers or providing suitable software for customers to do their own filtering.

### *Where do email address lists come from?*

People are often curious where the lists of email addresses used by spammers come from and how their own address was "harvested". The trite answer is that the lists come from other lists! The CDs you see advertised of 10, 15, 60 or 142 million addresses[30] have mainly been constructed by the purchasing and merging of other lists. This tends to mean that many of the addresses are years and years out of date[31].

The longer, and more useful, answer to how email addresses get onto lists[32] is that new addresses are mainly collected from web pages[33], though other sources such as newsgroup articles and mailing list subscriptions will also be plundered. Very few publicly readable directories of addresses remain on the net and those that do, such as "whois" databases have sophisticated systems to prevent wholesale harvesting.

However, although you may be able to avoid getting on to lists by not publishing your email address, or by ensuring that a "robot" will not be able to spot its presence[34], this will not guarantee no spam. Some automated systems will also attempt "directory spams" by testing a preset list of names (or just random combinations of letters) against a mail server. The addresses that work will then be used in the future for further mailings.

Many spam messages purport to be mailing list messages and offer you the chance to unsubscribe. It is generally considered to be unwise to do take up the offer[35]. Most of the time it will have no effect. However you are confirming that the email has reached a human being and that human actually reads unsolicited messages. This is known to make your email address more valuable and to increase the chance of future spam.

---

[30] http://www.google.com/search?q=cd+million+email+addresses

[31] "Total UBE Rate for an AOL account" records the unsolicited email (average 15 items/day) received by an AOL account that has been inactive since Nov 1996. http://www.tesp.com/UBETotalRate.htm. A further graph http://www.tesp.com/tespUBERate.htm shows the total received for an entire small domain (averaging about 75 items/day).

[32] Uri Raz "How do spammers harvest email addresses?" http://www.private.org.il/harvest.html

[33] "Why am I getting all this spam? Unsolicited E-mail Research Six Month Report", http://www.cdt.org/speech/spam/030319spamreport.pdf

[34] Address Munging FAQ: Spam-Blocking Your Email Address http://members.aol.com/emailfaq/mungfaq.html

[35] "Should I hit 'remove'?" http://spam.abuse.net/overview/remove.shtml

Notwithstanding the lies told by spammers, ISP abuse teams are very familiar with complaints from people who have "opted-in" to perfectly legitimate mailing lists, but have forgotten they have done so. They then complain, falsely, about receiving spam.

There are ways of coping with address harvesting. Many ISPs offer customers multiple email addresses, some place no restrictions at all on how many can be used. Many people will keep some addresses out of the public eye and only give them to trusted correspondents. They will then use different addresses for different companies (so they might be *amazon@example.co.uk* to Amazon and *bt@example.co.uk* to BT). Incoming email can then be quickly categorised and it will be rapidly apparent if any company is misusing the address they have been given.

### *Complaining about spam*

Although it is perfectly fine to just delete unsolicited email, users are often keen to complain about the spam they receive – especially when it is completely inappropriate. However, the users' own ISP may well not be interested in dealing with this type of complaint, since there is little that they can do to fix the problem. Instead, complaints should be directed to the ISP whose customer sent the email because they will be able to take effective action.

It is fairly easy to determine the source of incoming email. Although the "From:" line displayed by most email software is trivial to forge, and should be regarded as meaningless, the actual origin can be worked out. Above the email body is a series of "header" lines and the ones starting "Received" show the precise path followed by the email across the Internet. Following this route back will reveal the true sender of the email. Occasionally, where an open relay is involved, the trail can run cold, but at least the ISP whose customer is running the open relay will arrange to have it closed.

Many sites exist to explain how to read email headers[36] and how to use the information gleaned from them to make a complaint[37] and this information will not be repeated here. However, it is worth stressing that when making such a report, it is absolutely essential to include the original email, including all of the headers. Without that, the ISP is unlikely to feel able to take any action.

Although extremely rare, there have been a handful of incidents where paedophile material has been promoted by spam[38]. Since the email is eminently traceable back to its origin the authorities are keen to take action against the sender. Email advertising

---

[36] stopspam.org "Reading Email Headers" http://www.stopspam.org/email/headers/headers.html
University of Illinois at Chicago, "Reading Email Headers"
http://www.uic.edu/depts/accc/newsletter/adn29/headers.html
Ed Falk, "Tracking the source of email spam" http://www.rahul.net/falk/mailtrack.html

[37] Jeffrey Race, "Exterminating spammers step-by-step (Really!)"
http://www.camblab.com/nugget/extermin.htm

[38] Sky News, "Sick Websites Target Work Emails" http://www.sky.com/skynews/article/0,,30100-12024857,00.html

this type of illegal material should be reported to the police[39] or to the Internet Watch Foundation[40] via their website – you should paste the email, **including the headers**, into their webform at http://www.iwf.org.uk/hotline/report.htm.

### *Further reading*

This document, long as it is, has only scratched the surface of the issues surrounding the sending of spam.

The web is full of resources on the topic, many of which can be found via links from sites such as:

| | |
|---|---|
| "SpamCon Foundation Law Center" | http://law.spamcon.org/ |
| "Coalition Against Unsolicited Email (CAUCE)" | http://www.cauce.org/ |
| "Network Abuse Clearing House" | http://spam.abuse.net |
| "SpamCop" | http://spamcop.net |

### *and finally…*

As should be clear from this document, at its heart, spam is not really a technical issue, and although it may be alleviated for some, it will never be vanquished by technical means alone.

Dealing with spam requires us to tackle the question of how we regulate access to our mailboxes. To be effective, this regulation will have to agreed and implemented on a planetary scale. Only when we have done that, will be able to fully enjoy the revolution in cheap communication that the Internet has made possible.

Richard Clayton
August 2002 (revised March 2003)

[39] West Midlands Police Paedophile Unit, "Unsolicited images or e-mails via the Internet" http://www.west-midlands.police.uk/operations/paedophile/faq.htm

[40] Internet Watch Foundation, http://www.iwf.org.uk/