The Spam-ish Inquisition

Tired of spam with everything? Don't fritter away your time and energy on junk mail[1]

David Harley
Andrew Lee

Table of Contents

Introduction

Spam looks like a simple enough issue until you have to try to define it: after all, we all think we know it when we see it. Most people have a working definition along the lines of "email I don't want." While that's perfectly understandable, it is difficult to implement technical solutions based on such a subjective definition. (Actually, not all spam is email based, but we'll get back to that in a little while.)

A fractionally less subjective definition is "email I didn't ask for." However, this doesn't really meet the case either. A percentage of most people's legitimate email is not only unsolicited but from people they don't know (or from whom they have had no pre-existing communications), which is by no means the same thing – for instance, business communications from a third-party, that directly relate to your business.

Nevertheless, most spam falls into the categories of Unsolicited Bulk Email (UBE) and/or Unsolicited Commercial Email (UCE). That is, email which tries to sell you something, whether or not you want it and without any inquiry on your part. This paper will explain what spam is (and is not), provide a bit of the history behind the phenomenon and the responses, both technological and otherwise, that can be made to control it.

Defining Spam

UCE is usually regarded as a subset of UBE, though it's a very considerable percentage of 'spammy' email, and many simply use the term UCE interchangeably with 'spam'.

Bulk email such as newsletters and mailing lists don't usually count, since you "solicit" those communications by subscribing to the list. Of course, less scrupulous list masters may sign up harvested addresses, 'friends' may add you to lists that you didn't want, or an otherwise legitimate mailing list may be compromised or abused to carry spam, so there can be cases where such communications could be considered spam. However, this is a minor exception rather than a general rule.

Paul Vixie's[2] definition of email spam (the one that's also used by the Spamhaus Project[3]) is applied to mail that meets all three of the following conditions:

- More or less the same message has been sent to multiple recipients (or potential recipients). That is, it doesn't take into account the "personal identity and context" of the individual recipient.
- The recipient has not given "deliberate, explicit and still-revocable permission" to the spammer to send it to him or her.
- The message is of value ("gives a disproportionate benefit") to the sender, not to the recipient.

Professional versus Amateur Spam

We sometimes find it useful to distinguish[4] between "professional" and "amateur" spam.

By "professional" spam we mean what is sometimes referred to as hardcore spam, though this usage inevitably invites confusion with pornographic spam (which may well be "professional" in the sense in which we use it here, but constitutes a fairly small proportion of the spam totality). While there is no universally accepted definition for the term, we use it to refer to:

- Mail which unequivocally meets the Vixie definition above: for instance, it's totally untargeted – that is, directed towards any addresses the spammer has access to. It doesn't matter who or where the owner of the address is, or whether he or she is likely to be interested in the topic, or whether they ever agreed in any sense to being included on the spammer's lists.
- Mail which is probably illegal (in some jurisdictions, at least) and would not be considered acceptable by most recipients as legitimate marketing. Similarly, mail which advertises products or services that would not be considered acceptable or even legal (pirated software, pornography, drugs) in most societies.
- Mail which involves deliberate deception. Most legislation fails to distinguish clearly between commercial spam (UCE, if you like) and fraudulent spam, as do most standard spam information resources. However, as Hallam-Baker suggests,[5] the "professionalization" of phishing, botnet exploitation (which also can include malware generated spam, and seeding of new malware via email) and so on, has resulted in a dramatic decline in the proportion of more-or-less legitimate but unwanted advertising, compared to more-or-less legitimate terrestrial junk mail. Nowadays, most spam is clearly deceptively intended and therefore likely to be unequivocally criminal, in some jurisdictions at least. Also, just the fact that mail headers have been forged is often enough to make a mail illegal, irrespective of the accuracy of the content of the message.

Deceptive Elements

Here are some of the deceptive elements we are accustomed to seeing in 21st century spam.

1. Techniques are used that are clearly intended to circumvent spam filters, for instance:
    - Hashbusters: this is a term applied to textual or graphic content that varies between instances of a given spam message, to counter spam solutions that rely on generating a hash or checksum of a known spam and blocking messages with that hash value.
    - Misspelling of keywords generally associated with spam (for example, using cial1s, cia_lis, rather than cialis) to confuse filters that use literal pattern-matches. More effective pattern matching algorithms use fuzzier matches: for instance, wildcards. To take a very simple example, cia*lis, where the asterisk represents any number of "noise" characters, would recognize "cia_lis", "cia lis", "cia<fake html tag>lis" and so on as the keyword "cialis". However, there are lots of ways to conceal a given keyword from a software filter without rendering it unreadable to a human being,[6] and giving a filter the same powers of

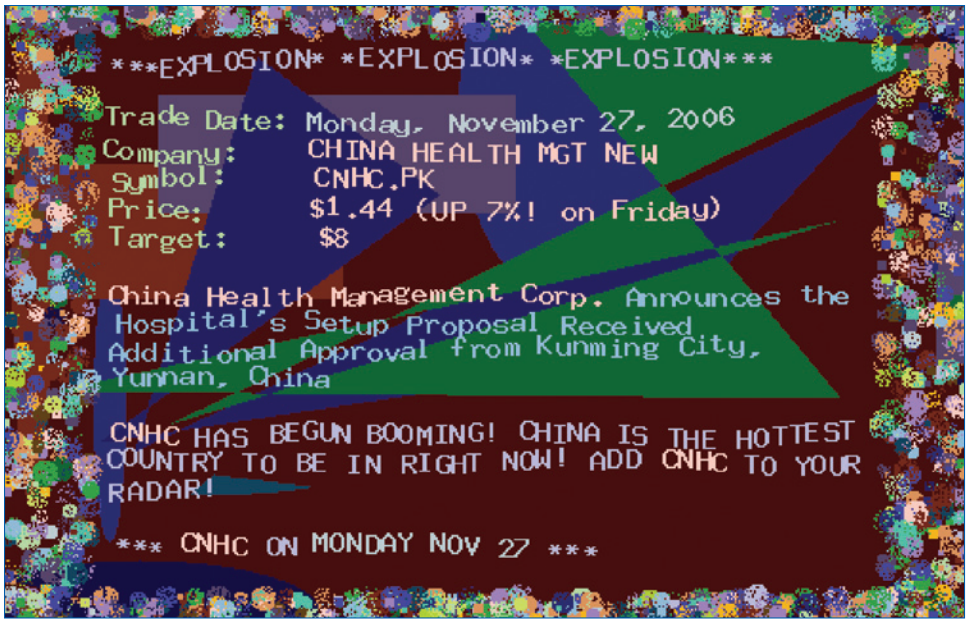


Figure 1: Hashbuster Spam.

discrimination that a human has would be non-trivial...

- Image-only content (image spam often includes "hashbuster" random text)
- header spoofing (see below)
- concealed or obscured text (i.e. microscopic fonts, white text on white or near-white background, text interspersed with meaningless HTML tags, and so on): often used to conceal hashbusters or to lessen the effectiveness of filters that look for keywords that suggest spam.[7]

2. Message or Subject content that makes false claims about a product or service: for instance mails falsely purporting to sell legitimate OEM (Original Equipment Manufacturer) versions of popular software.
3. The message text claims falsely that the recipient in some way invited the message, and insists that it isn't spam. In fact, that's not a bad heuristic for spam detection.[4] Just as schemes that assure you that they aren't illegal or fraudulent almost invariably are,[8] so it's usually the hardened spammer who "doth protest too much, methinks."[9]
4. The sender masquerades as a friend, colleague, relative or similar.
5. The message headers are forged to make it harder to track the source, as shown by mismatched headers, impossible IPs and time stamps, and so on.
6. There is a deceptive Subject header to persuade the recipient to open the message.
7. The message often involves some degree of actual fraud such as the following:
   - 419 "Nigerian" scams, including advance fee fraud, job scams, lottery frauds and so on.
   - "Phishing" fraud: that is, mail sent with the intention of tricking you into giving away sensitive bank data, and other attempts to ascertain sensitive information

such as passwords.

- Pump and dump stock scams (see our paper ""A Pretty Kettle of Phish" for a fuller explanation of this sort of fraud).[10]
- Money mule recruitment.
- Offers of products and services which are never actually supplied when the victim pays for them.

Amateur Hour

By "amateur spam", we mean bulk mail sent inappropriately by more-or-less legitimate enterprises, but without due regard for acceptable practice or even legal requirements. Schwartz and Garfinkel quote an early example[11]: in 1993, a professor at Penn State University bulk mailed a survey on the use of the Internet by academics to multiple mailing lists. Paul Vixie pointed out to him that while his intentions were "good", the effect of his mail was "to hasten the Internet's downslide into common-market status" by establishing a precedent. "Collecting addresses is free; generating mass mailings from them is close to free. Can you fathom the effect these metrics will permit once the Internet comes a little bit closer to the mass market?" Sadly, Vixie's misgivings were all too close to the mark. Academics and researchers still have difficulty realizing the implications of using the Internet as if it were still an academic's playground, not only because of its immediate impact but because it legitimizes less principled abusers.

Other forms of "amateur" spam include:

- Inappropriate, poorly thought-out and mis-targeted or untargeted marketing messages from legitimate institutions inadvertently breaching legislation, netiquette, and/or acceptable marketing practice. It's depressingly common for commercial companies to be careless about sending email in breach of applicable legislation: on one occasion one of us received advertising mail sent to a private account by a former member of a law enforcement agency specializing in computer crime, putting the sender in breach of UK legislation. Charitable institutions are sometimes guilty of questionable practice: we regularly see advertising from such institutions that is technically in breach of the same legislation, and one of us once had occasion to dissuade another charity from encouraging its staff to forward advertising material in the form of an indiscriminate chain letter, which would have put it in breach of its ISP code of connection as well as current European legislation. However, legitimate enterprises are usually responsive to having their attention drawn to legal breaches. They may not respond directly to complaints, but tend to modify their marketing practice.
- Virus and security hoaxes, and other chain letters. Of course, the people who are duped into forwarding hoaxes and semi-hoaxes are by no means "professional spammers" – indeed, they often have altruistic motives. (We'll talk a bit more about chain letters in a while.)
- Backscatter from antivirus services, anti-spam filters, blacklists and so on. Some of these problems arise from the fact that many people – not just end users, but

system administrators, companies, even security companies – have been slow to realize the full implications of header forgery by spammers and malware authors as regards such issues as routine non-delivery notifications and other auto-responses. While it's now comparatively rare to see a misdirected "you sent a virus" messages from poorly-configured or poorly-featured gateway antivirus scanners, more generic filters continue to generate responses that end up in the wrong mailbox.

- Unsolicited and generally unwanted mail from strangers pursuing a non-commercial agenda, such as the propagation of political views and religious beliefs (sometimes combined with attacks on other factions). The pseudonymous Serdar Argic is often cited as one of the "pioneers" of spamming, posting tens of thousands of anti-Armenian messages to newsgroups in response to any post mentioning Turkey or Armenia – it is considered likely that these messages were the result of a bot automatically broadcasting pre-prepared pages of political text[12] in response to those keywords, irrespective of context.

There are categories of unwanted mail that are not always considered to be spam by purist spam hunters, such as messages generated as a result of malware (mass mailers, malware seeding mails, misdirected malware or spam alerts from poorly configured filters, misdirected non-delivery reports) or out-and-out fraud such as phishing messages and stock fraud. We do address such issues in this paper, because they are so often popularly considered to be spam and do meet some of the same criteria, and we make no apologies to holders of the more purist view.

Why "Spam"?

Hormel, the manufacturer of the famous canned meat product, must have asked that question many times: in fact, they even have a page about it on their website at http://www.spam.com/legal/spam/. The term isn't capitalized when used in the context of this paper (or any other dealing with the subject) because SPAM is how Hormel uses it to refer to their product. That's a registered trademark, by the way. The company gets quite upset if the word is capitalized with reference to electronic messaging and has attempted legal action to enforce that distinction. It's usually assumed that the use of the term in the context of various forms of computer abuse derives from the infamous Monty Python sketch,[13] and was probably originally applied in the context of MUD (Multi-User Dungeon) environment abuses such as flooding a chat session with irrelevant, often auto-generated text in a way reminiscent of the chanting of the Python Vikings. There are, however, some attractive but unlikely alternative explanations.[14]

Spam and Pornography

To many people, spam is synonymous with pornography. As we've already suggested, this isn't at all so, but pornographic spam does offend and upset many people. Apart from the fact that such people often feel that their provider or employer should be protecting them from such mail, there is a risk that employers will take inappropriate disciplinary action

because they may be unable to distinguish between the passive, unwilling recipient of pornographic material and people who actively seek titillating content. There are increased risks and inherent complications with types of pornography that are actually illegal, especially pedophilia-related content, and, of course, in some territories, the definition 'legal pornography' is oxymoronic. Child porn arouses particular revulsion and fear in many people: this often results in confused and contradictory legislation that is difficult to enforce and to which is it sometimes impossible to conform. In fact, in some countries, even to receive such images, even unknowingly, may put the recipient into legal jeopardy, as can forwarding such images with the intention of reporting them, except under circumstances that are not always clearly defined (see http://www.virtualglobaltaskforce.com/ for more consideration of these issues).

A brief discussion of the various legislative attempts at controlling spam by various countries can be found later in this paper.

Spam Attacks

It's not unusual for spam to be used as a means of directly attacking an organization or individuals.

Bombs Away

Here are some ways in which spam can be a direct attack:

- Mail bombing is the bombardment of a site or account with email messages, thus in effect executing a denial of service against the mail recipient. This has also been a (perhaps unwitting) side effect of many email borne Internet worms.
- Subscription bombing is a term applied when the bombardment is indirect: the individual is subscribed without their knowledge or permission to a number of high-volume mailing lists: it's for this reason that good list-processing software commonly sends mail to a candidate address requiring some form of confirmation before subscription is accepted[15]
- "Revenge spam" is a term sometimes used[15] where a spam run's headers are forged to make it look as if it came from a particular site or individual, with the intention of inducing recipients to take action against them and cause damage to their reputation. This kind of spam is sometimes referred to as a Joe Job, after such an attack was directed against Joe Doll, web master of Joe's CyberPost[16] in 1996/7. A particularly nasty variation on this theme was a reputation attack (also launched in 1996) that purported to associate an address in Jackson Heights with trade in child pornography.[17]

Address Harvesting

Some phenomena described as attacks, though, are not so much direct, aggressive attacks as incidental to the spamming process.

People are sometimes puzzled by the arrival of unsolicited mail that includes only random words or randomly selected text (such as extracts from novels), or no text at all. These may be sent in error (as in image spam where the image that contains the advertising content payload "falls off", or there is a glitch with the spamming software or some messaging problem somewhere between the originating machine and the target's mailbox), but are often part of directory harvesting attacks (DHA). DHAs generally consist of bombarding a domain with messages sent to made-up, automatically generated user names in order to harvest legitimate account names for that domain. If a "no such address" report is returned when mail is sent to a guessed address, that particular address may not be used in subsequent spam runs. Conversely, if such a report is sent in some cases from a particular domain, the absence of such a report in the other cases can be an indication that an email address has been 'hit' and can be counted as live. These mails are not as harmless as they may seem: automatically generated combinations of numbers and letters up to an arbitrarily selected string length can result in a mail storm of many millions of probe messages. The common use of HTML in messages also means that an apparently brief or empty message can contain copious but inconspicuous content: some malware uses this vector to include hidden malicious scripts. Such messages can also attempt to take advantage of email 'receipt' messages, that inform the sender of successful delivery (and in some cases, that the message has been read).

Spam Through the Ages

Jon Postel recognized very early in the game – 1975 – that there was an essential weakness in vanilla email: at that point, there was, by default, no mechanism by which a mail server could reject any message sent to a valid address, and discussed the (then) more-or-less hypothetical problem in RFC 706 (see Figure 2).

First Sightings

The term spam has long been applied to inappropriate cross-posting and other abuse of newsgroups and it is often stated that Usenet spam predates email spam, but we're not convinced. One of the earliest widely-reported cases of Usenet spam was a religious tract posted on 18th January 1994 by a systems administrator at Andrews University, Michigan.[18] Arguably, though, the earliest UCE was sent as early as 1978[19] from a DEC rep who attempted to invite the holder of every ARPANET address on the West Coast to a product presentation.

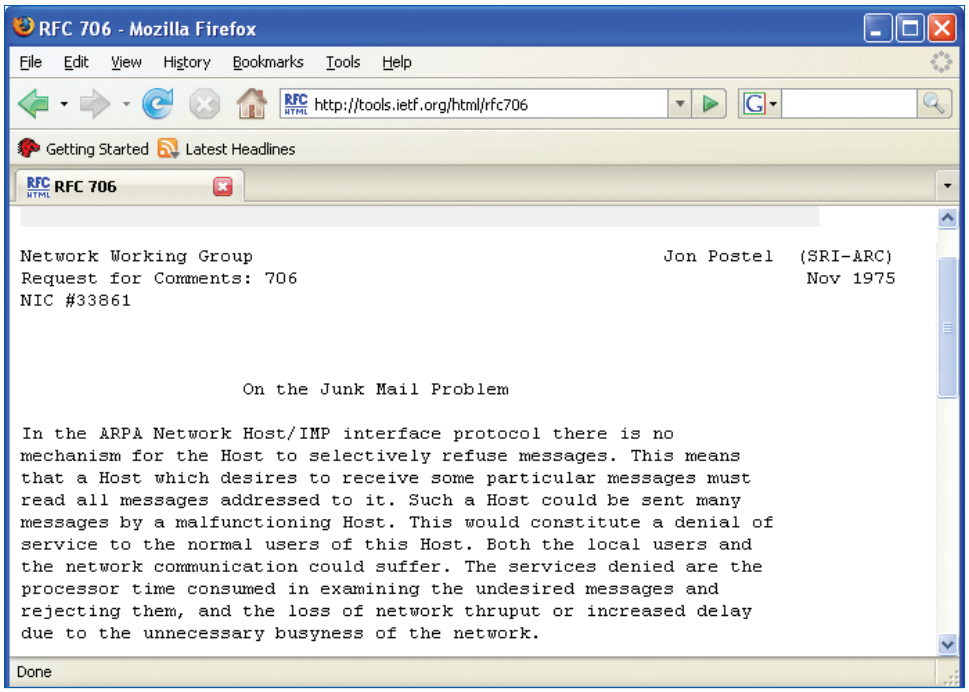


Figure 2: RFC 706 – First Thoughts on a Future Problem

Newsgroup Spam

The earliest commercial Usenet spam is usually assumed to be a message sent on 12th April 1994 to every newsgroup then in existence by a pair of lawyers in Arizona (Laurence Canter and Martha Siegel). This offered legal help to immigrants to the U.S. entering the "Green Card" (US Permanent Resident Card) Lottery.[20] "In fact, this is the precursor of an avalanche of green card scams where aspirant U.S. residents were invited to pay significant sums in the vain hope of somehow increasing their chances of winning a free and presumably impartial lottery."

Of course, scams that involve paying a third party for services you could perform for yourself, much more cheaply, are ubiquitous. A (usually snail-mail borne) example from the UK is the common data-protection registration scam. An organization or individual that/who may or may not be liable to register with the Information Commissioner receives an official-looking letter requiring them to do so. However, the communication doesn't come from the Information Commissioner's office, but from an agency which will charge a heavy additional fee (often 3-4 times the cost of self-registration) for registering the victim, irrespective of whether they actually need to register.[21] Also common in the UK are 'domain-registration' scams. The scammers send domain registration renewal documents that look like invoices to domain owners: only the small print reveals that these are actually advertisements for domain registry companies.

Cantor and Siegel later wrote a book called "How to Make a Fortune on the Information Superhighway: Everyone's Guerrilla Guide to Marketing on the Internet and Other On-line Services" which clearly influenced the spammers who followed, such as Jeff Slaton and Sanford Wallace.[17]

We won't talk much more about newsgroup spam here because it seems to affect far fewer people nowadays, though it has by no means disappeared from that environment. However, there are many more people affected by email spam than by Usenet spam, and old-style newsgroups have, arguably, declined in importance as more sophisticated "Web 2.0" interactive environments have gained ground. The impact of various types of Usenet abuse such as EMP (excessive multiple posting), sporgery (see glossary for definition), and the dissemination and updating of malware through newsgroups like alt.comp.virus, probably contributed to a decline in general use. However, the initial relationship between Usenet and spam had a major impact on the way we think about spam and spammers, and how we try to deal with the problem.

Spreading Spam

Early manifestations of spam were usually easy to trace to a particular source, often a particular individual account, albeit by a recognizable spam-friendly provider. At one time, spammers made much use of open relays (mail servers that didn't properly authenticate incoming mail and forwarded it irrespective of where it came from), and some of the earliest DNS blacklists/blocklists (DNSBLs) were lists of IP addresses known to allow open relaying. Businesses (and indeed many other sites) would use these lists to block mail that came via open relays, in the hope of reducing the numbers of spam mails that hit their systems. However, the number of server administrators who leave their relays open has declined dramatically in recent years, so spammers have found (or built) alternative resources such as open proxies, which allow mediated/masked connections for many purposes, with consequently expanded opportunities for abuse.[22] As a result, many blocklists were expanded to include open proxies.

Nowadays, most spam is disseminated through botnets, specifically through compromised machines used as open relays or open proxies. While some DNS blacklists now include open proxies and other presumed "rogue" resources, complex botnet structures and implementations using dynamic DNS and fastflux mechanisms make it far harder to track and close down problematic systems: ranges of compromised machines are switched in and out, and it becomes even harder to distinguish spoof from truth in email headers. Indeed, this has led to block-lists becoming less important as a spam control mechanism, and some of the major lists have closed down. Other botnet capabilities such as distributed denial of service (DDoS) attacks have been used to target blocklist providers and other security resources.

Spam is far from the only problem associated with bots (zombies) and botnets, and we have looked at this phenomenon in some detail in another paper.[23]

Spam Economics

A simple cost/benefit analysis model shows, however, why the spammer business economy is so effective, and why the problem remains so intractable.

Costs	Bandwidth Development Acquiring spam tools (spamware) Building or leasing a botnet Buying or harvesting target addresses
Benefits	Reduction of risk of identification and therefore legal or other penalties. Almost unlimited potential targets Negligible transaction cost (and it may cost no more to send 10 million than 10 thousand)

Table 1: Spammer Cost/Benefit Model

The comparatively low cost of sending bulk email means that:

- The spammer can make a profit even when only a small percentage of recipients take the bait
- The transaction cost tends to fall dramatically as the volume of a mail-out rises.

The cost to the recipient, on the other hand, is often considerable, even leaving aside the cost of implementing technical countermeasures and losses due to deliberate fraud, identity theft and so on: consider, for instance, the impact on system and network resources, the opportunity costs to individuals and administrators of sorting through unwanted messages, the support overheads, the psychological impact of receiving offensive material or being duped by hoaxes, and so on.

Other Spam Channels

Nowadays, most people think of spam as being the same thing as email spam, and that's the main area on which we've chosen to concentrate in this paper. However, there are many other communications media used by spammers, apart from Usenet and email: so, for the sake of completeness, we'll briefly discuss a few here.

SPIM

SPIM is aimed at users of Instant Messaging (IM) services such as AIM (AOL Instant Messenger), ICQ and MSN. IRC (Internet Relay Chat) also has a long and dishonorable history of spam outbreaks, apart from the medium's current association with controlling the botnets that disseminate most current spam. The term "chat spam" is sometimes used to refer to "noise" generation in order to cause disruption (this is a further reference to the noisy Vikings in the Monty Python Spam skit): this is often observed in online gaming chat environments, other chatrooms and discussion forums. Many IM clients (for instance Yahoo! Messenger, AIM, Windows Messenger) have a "whitelisting" option: if this is checked, Instant Messages from unknown sources are disregarded.

Messenger service spam exploits the Messenger service introduced in Windows XP, primarily for sending network administration messages, but disabled by default since Service Pack 2. (This is not connected with Windows Messenger or MSN.)

Text Messaging Spam

This type of spam is aimed primarily at cell phone users, and abuses the Short Messaging Service (SMS) that most mobile services and some landline services support. It's a particular annoyance to subscribers who actually have to pay for texts received, but the volume of spam SMS messages (occasionally referred to as SpaSMS, we are depressed to note) seems, to date, to have been small compared to other types of spam. (In 2004, however, the Korean Information Security Agency reported that it was getting more reports of SMS spam than of email spam.[24])

Nonetheless, SMS analogues to other types of messaging abuse have been reported, such as chain letters and phishing (would you believe SMiShing?). In the SMS version, recipients are lured into using premium text services or sharing sensitive personal data (as with email phishing).

A CISCO white paper[25] specifies four types of SMS-based fraud:

- Spamming (the provider is used as a spam relay by a content provider with a regular service agreement.)
- Flooding (content from another network)
- Faking (from an engine that simulates regular SMS Centre behavior)
- Spoofing (engine that simulates roaming mobile devices, leading to billing issues)

Not only do these issues cause annoyance or direct harm to customers, but they also damage the provider's revenue flow.

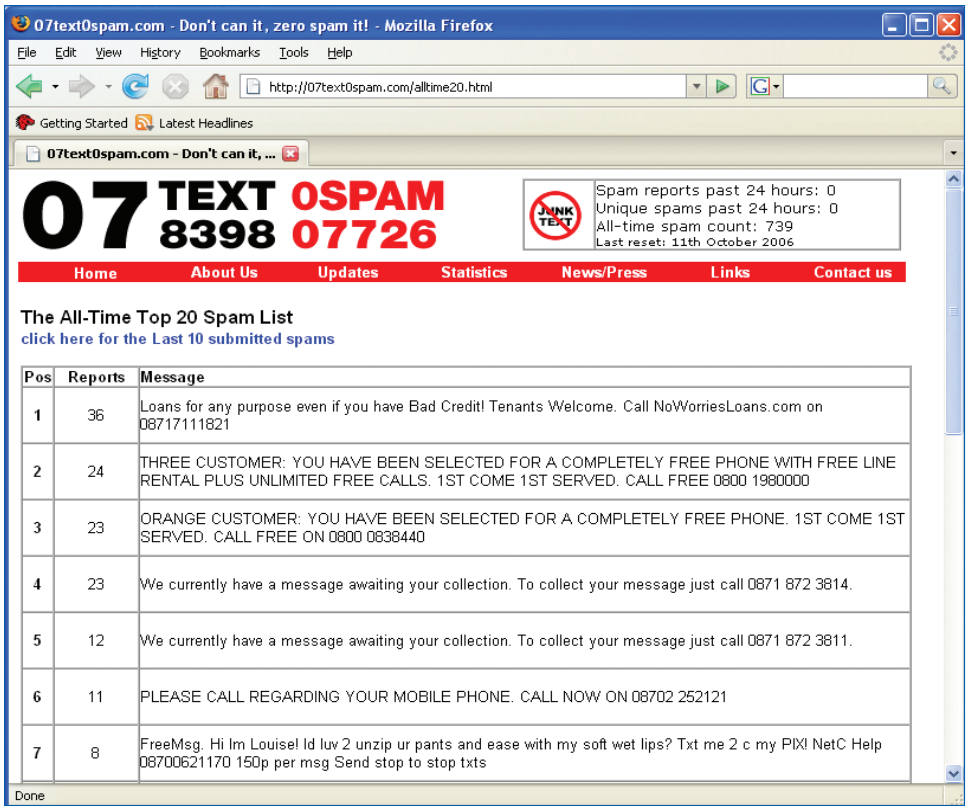


Figure 3: Text Spams Forwarded to 07text0spam.com

Blog Spam

Blog spam, sometimes referred to as "comment spam", or (sigh) as "blam", is the name given to spamming web logs (blogs). This normally takes the form of posting a comment to a blog containing nothing but a link to an irrelevant web page. It may also contain text which may or may not have some relevance to the topic – if it does have some relevance; it lessens the probability of the comment being quickly removed. Similar attacks may be made against other types of web page that allow visitors to add content, such as wikis and guestbooks. Increasingly, this sort of spam is seen in social networking sites (which have their own uniquely challenging problems) where comments can be left by other users of the site, On such sites, even innocuous looking comments such as "Hi, I've taken some new pictures, click here to see them" could be considered spam even if they are personalized, or related to the recipient. Unfortunately, it is the very nature of such sites that the line between genuine spam and simply unwanted comments is blurred. In principle, this closely resembles the way in which Usenet newsgroup threads can be devalued by adding irrelevant marketing messages and hyperlinks. However, blog spam is largely and specifically used for spamdexing (spamming + indexing), a technique for unfairly boosting site ranking in search engines like Google by increasing the number of sites that carry a link to an indexed site.

Index Hijacking

Igor Muttik describes[26] the specific use of "Index Hijacking" to ensure that sites hosting malware are highly ranked in search engine return lists. Google, to take the best-known example of a search engine, uses a number of factors to determine the position of a given page in its search lists, including its PageRank™ methodology.[27] A page is ranked according to the volume of links on other pages that refer to it (its "importance" or popularity) as well as the popularity of the pages that link to it.[28]

Junk Faxes

These can be a particularly irritating spam analogue, since they tie up phone lines and use the recipient's resources. While the problem receives less attention than other forms of junk mail, unsolicited faxes are a common vector for unsolicited advertising as well as scams such as pump and dump hyping and 419s (see following section). However, the increasingly common use of fax servers, email-to-fax gateways, and email as a substitute for fax communication, means that its impact may be masked, and these approaches do offer the potential for filtering junk that isn't usually available to someone using a simple facsimile or all-in-one office printer/fax/scanner combination.

Spams and Scams

The Internet and email have always been happy hunting grounds for the morally bankrupt and criminally inclined.[8]

Make Money Fast

Pyramid schemes like the Dave Rhodes "Make Money Fast" (MMF) schemes, Ponzi scams, and so on, have a long and ignoble history, though they pale into insignificance next to the volumes of today's phishing scams and identity theft. Pyramid schemes have a passing resemblance to legitimate multi-level marketing (MLM) operations, and may be passed off as such, but differ primarily in the fact that there is no real product or service being sold.

The U.S. Postal Inspection Service maintains that this kind of chain letter is illegal when sent by terrestrial mail, citing Section 1302 of the Postal Lottery Statute,[29] on the grounds that it constitutes a lottery (because it isn't possible that all participants will be winners), as well as pointing out ways in which the scammer (and other participants) can abuse the scheme. Other sources[8] point out that the numbers simply don't work anyway.[30] For a fairly typical (though early) example of a "Dave Rhodes" MMF scam, see http://www.cs.rutgers.edu/~watrous/dave-rhodes.html.

Advance Fee Fraud

This type of fraud is often referred to as a Nigerian scam, or 419 (after the section of the Nigerian Criminal Code that deals with fraud), because so many examples of these scams come from that region (though many seem to come from Eastern Europe or the Far East). Advance fee frauds consist of offers of windfall money. However, when the victims respond, they are told that before they can get the money, they have to pay money upfront for advance fees of various kinds (taxes, bank charges, bribes) before the money can be delivered. In truth, this is an online version of a sting that has been around in one form or another for many centuries.[31] Common variations on the theme are shown in Table 2, based on an earlier paper in this series.[10]

Table 2: Common 419 Types

Political refugee appeals	A request for help from a political refugee to get their money out of the country and into yours. Often appears to be on behalf of the family of a dead ruler, dictator, or less notorious but still wealthy individual.
Philanthropic/Religious appeals	A request for help with the distribution of money for charitable purposes. Often appears to be from a private individual who is dying, or the representative of a religious or philanthropic organization (one of our own particular favorites appeared to come from the recently-appointed Pope.)
"Do others before they do you."	A request for assistance from a bank or other official with transferring money obtained more or less illicitly from the regime or the institution that employs them. Sometimes includes some tenuous moral justification such as "the money would otherwise be spent on arms or be absorbed into the bank's coffers", but the fact that the transaction is illicit makes it less likely that the scam will be reported.
"Next of kin" inheritance scam	A request to stand as "next of kin" for the purposes of claiming the estate of a dead foreigner who has died intestate, or along with his entire family.
Lottery scam	Notification of a lottery win: it turns out you have

	to pay fees before your win can be released.
Job scams	Job opportunities requiring upfront registration fees and such.
Mule recruitment messages	Over the past few years, messages that resemble classic phish-related "jobs" in money-laundering but have a decidedly "419" feel have been appearing. Some of these, however, turn out to be another variation on advance fee scams tied to job "opportunities."
Disaster scams	As Martin Overton has pointed out,[32] it's not only personal disasters to which 419s are pegged. Events like armed conflicts, earthquakes and tsunamis, are not only used to supply spurious circumstantial detail to lend credibility to a scam story, but as the basis of false charitable and disaster relief appeals.

419s tend to rely more on social engineering attacks than on technical attacks such as cross site scripting and DNS spoofing, unlike "true" phishing scams.

Phishing Scams

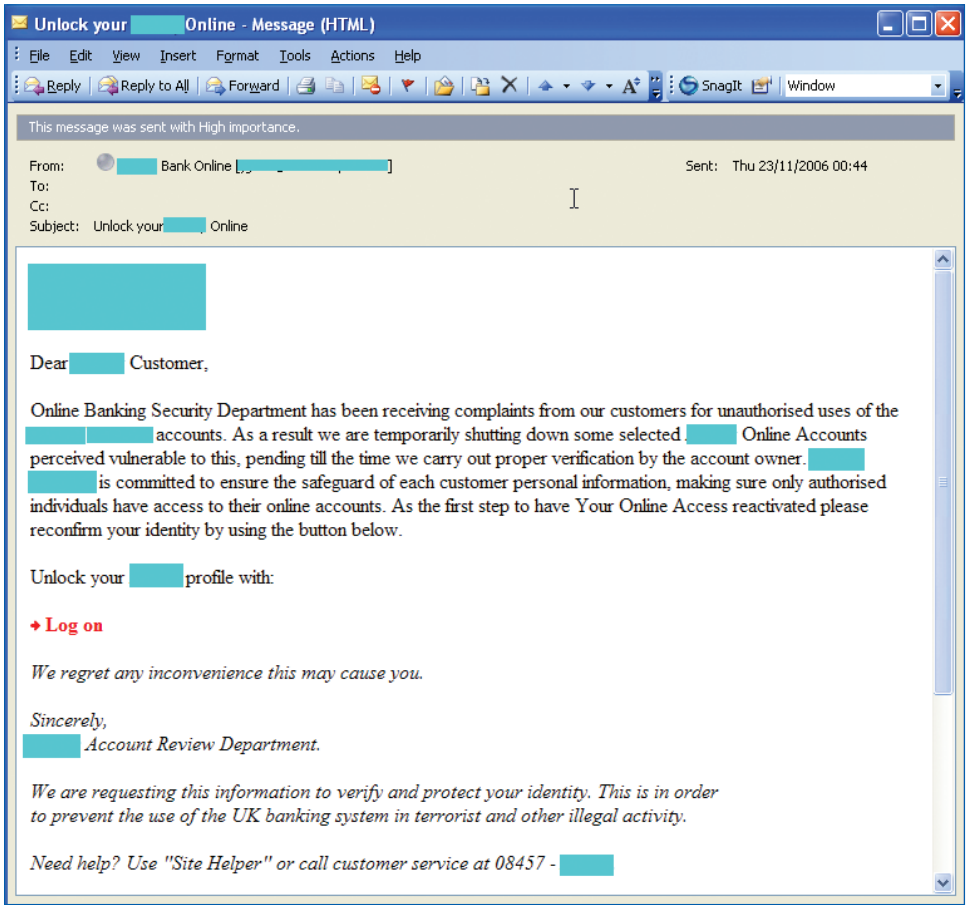
Phishing is the use of a deceptive message (usually an email message) as part of the process of carrying out a fraud. The term was originally largely limited to stealing AOL account passwords and credit card information, though it may be related[33] to "trolling", a term also related to fishing (trawling).[34] Phishing refers to the practice of "fishing" for victims, using a baited message. Nowadays, it's most often associated with passing off requests for sensitive data as if they come from all sorts of major organizations (especially those in the financial sector such as banks and credit unions, eBay and PayPal, and so on). However, a phishing attack usually has three distinct components:

- Distribution of the bait, usually through email, though other channels such as instant messaging, VoIP (Voice over IP), SMS texting and so on. Typically, phishing emails contain deceptive links to the fake web sites that constitute the second phase of the attack.
- Harvesting the sensitive data that enable the phisher to move into phase three, usually by means such as fake web sites and pop-up forms.
- Using the misappropriated information for purposes of fraud and identity theft.

The phishing message is passed off as if it came from a legitimate source, so that the victim will be prepared to hand over sensitive and exploitable data to the phisher, whose intention is generally to get access to their funds, or to steal their identity in order to defraud others,

for example by taking out large loans in the victim's name.

The term may also be extended to include emails designed to plant malware and spyware such as keyloggers, password stealers, and backdoors. These may include mails with malicious attachments, or containing links to drive-by downloads or offering Trojanized (or "trojaned") programs. Common approaches include passing the message off as an eCard, or



as a security patch.

Figure 4: Banking Phish

Phishing gangs are part of a complex "black economy"[22, 10, 35] similar to other commercial models. These are highly dependent in their turn on the existence and exploitation of botnets.[23] This "economy" entails a number of roles and functions[3] as in follows:

- Target information such as email addresses is harvested.
- Phish sites, resembling or incorporating elements of legitimate sites, are set up, often using widely available phishing kits.
- Spamming tools and compromised systems as dissemination of and hosts for bait (botnets and zombie networks, for example.)

- Host systems are compromised and acquired to house scam pages.
- Stolen credential data are forwarded to anonymous mailboxes and retrieved, characteristically by using a scriptable bot for control and data transfer.
- The victim's credentials are converted to cash.
- The buyer uses the stolen credentials, for instance to buy goods for sale on the black market, or to negotiate loans and mortgages.

Phishing mails rely largely on brand theft and scare tactics along the lines of "Your account has been compromised: to re-authenticate, click here, or we will suspend your account." However, it's not unusual to see offers of rewards for information. It's important that legitimate institutions avoid "grooming" their customers to be victims, by using bad practice in legitimate communications and by giving unreliable advice and information.

These topics have been addressed in much more depth in another of these papers,[10] and we have looked in some detail at phishing quizzes and other educational approaches in a paper to be presented in September 2007.[36]

Mule Train

Important to the phishing economy are mule recruitment solicitations, offering "financial management" or "financial agent" jobs that boil down to receiving money and passing it further up the chain after taking a cut as commission.

YARD SCRAPER, INC. SOUTH AFRICA
Head Office: 131 Braamfontein,
Midran-Johannesburg
2050 South Africa

Good Day

I am Mr. Kelvin Powell, President/CEO of Yard Scraper, Inc. South Africa (a company based in the South Africa). A Company that is specialized in import and export of industrial and domestic machinery & equipment, communication accessories and household appliances.

We also deal on mechanical equipment, hardware and minerals, electrical products, medical & chemicals, light industrial products and office equipment, and export into America, Asia and Europe, therefore being a General Mercantile Company.

We currently run our business from America, Asia and Europe but I will be communicating with you from our South Africa Office where I am currently located for now. We are searching for representatives who can help us establish a medium of getting to our customers in America, Asia and Europe as well as making payments through you to us. Please if you are interested in transacting business with us we will be most glad to be your partners.

My company is willing to offer you 10% of every payment that comes in through you to us. If you are interested, kindly forward to us the following information through my private email (infoyardscrapercompany@jmail.co.za):

Full Names
Company Name
Telephone & Fax Numbers
Full contact addresses
Age
Sex

Please note that your area of specialization or occupation is of no relevance to resolve to assist us.

Thanks in advance.
Sincerely.
Kelvin Powell
President/CEO of Yard Scraper, Inc.

Figure 5: Mule Recruitment Spam

Funds transfer/money-laundering scams don't generally purport to come from the same type of institution that phishing scams do, and aren't aimed at cleaning out the victim's accounts: they are more concerned with using the target as a "money mule." They advertise "jobs" via email and recruitment web sites to people prepared to act as their local agents. The mule is often required to open new legitimate accounts with specific financial institutions so as to facilitate moving funds from a phished account with the same institution. The scammer may go to extreme lengths to make the mail look like a serious job offer, backed up by a large and complex web site. Figure 5 shows a crude but not untypical example of a money mule recruitment spam.

Pump and Dump Scams

Pump and Dump (or Hype and Dump) mails are designed to inflate the value of stock temporarily by hyping it to potential small investors. Typically, the scammer will buy a large amount of next-to-worthless stock, and then hype the company through spam, hoping other investors will buy it, thus inflating the price. As these duped investors buy stock, its value rises till the scammers sell off their shares at the now inflated price. They then stop hyping the stock and it falls in value, and typically the new investors sustain a financial loss. These mails are still often seen as a minor nuisance, but are rising in volume and widening in geographical scope, and there is evidence that organized crime is making a great deal of money this way.

Chain Letters and Hoaxes

Chain letters were considered to be a big issue by the antivirus industry in the 1990s, when hoax virus alerts like the infamous Good Times virus hoax circulated time and time again. According to the Good Times FAQ[37] the message in Figure 6 was the earliest known version of this particular hoax.

> FYI, a file, going under the name "Good Times" is being sent to some Internet users who subscribe to on-line services (Compuserve, Prodigy and America On Line). If you should receive this file, do not download it! Delete it immediately. I understand that there is a virus included in that file, which if downloaded to your personal computer, will ruin all of your files.

Figure 6: Example of the Good Times Hoax

In fact, the near-prototype for most of these "viruses of the mind"[38] is the metavirus proposed around 1988 by Jeffrey Mogul,[39] who probably regrets ever mentioning it.

The last virus hoaxes (to date) to have a major impact were the SULFNBK and JDBGMGR hoaxes, which were very prevalent in the first half of the present decade. These were particularly interesting (and annoying) as they actually caused the user to do harm to their system by suggesting deleting legitimate files. Fortunately, the files were not critical in most cases, but the hoaxes certainly caused headaches for overworked IT support staff dealing with people who had gone ahead and deleted them. Interestingly, while at the time these specific hoaxes seemed to herald an age of more deliberately malicious hoaxing.[40] It was not to be, and indeed these stand as the first and last widespread examples of this type of electronic ephemera. However, golden oldie virus hoaxes continue to flourish: according to one hoax tracking page, five of today's top ten hoaxes were virus related.[41] Other forms of chain letter (even apart from the pyramid scam chain letters we've already considered, which are often referred to as chain letters) continue to be a major problem for system administrators, helpdesks and so on, tying up network and support resources.[4]

A chain letter is one that instructs each recipient to forward multiple copies: terrestrial chain letters usually specify an arbitrary number of people to forward to, but chain emails often specify "everyone you know." Richard Dawkins cites the "St. Jude" terrestrial chain letter[42] as an example of "replicators [that] exhibit exponential growth." Chain emails have, in turn, been held up[43] as examples of "memes", and virus hoaxes are often described as "memetic viruses". "Meme" is a term coined by Dawkins[44] to denote a unit of cultural transmission in the same way that the gene denotes a unit of heredity.

Chain letters are often described[45] as having a tripartite structure.

- The hook catches your interest
- The threat is the incentive to obey the request that follows
- The request is the "replication mechanism", urging you to forward the message.

Chain letters are passed on for a variety of reasons: fear of the consequences of not forwarding, a desire to be helpful, and self-interest. Once a number of people can be seen to have passed it on, the likelihood that other recipients will do likewise increases dramatically – this is often referred to as modeling behavior. David Harley[46] has described the heavy impact on public services by understandable but misplaced attempts to identify children presumed orphaned by the 2004 Indian Ocean tsunami by forwarding chain letters. This kind of "sympathy" spam is not only technically tricky to filter, but also requires careful handling because of the psychological mechanisms at work, so that the victims may be more upset by attempts to regulate this type of abuse than by the abuse itself. Good practice is always to donate via legitimate and well known charities, rather than via organizations that spring up overnight, and may or may not be legitimate. The site http://www.charitynavigator.org is a useful reference site, tracking the percentage of donations that actually get to the victims. It's shocking, and quite saddening, how small a percentage is, in some cases, actually passed on to victims from charitable donations.

Spam and the Law

There have been many legal and quasi-legal attempts to regulate spam, including acceptable use policies (AUPs) and Terms of Service (ToS), local and national legislation. Unfortunately, these measures tend to be honored only by legitimate advertisers: clearly, spammers who carry out frankly criminal activities like phishing are unlikely to be bothered by legal niceties.

CAN-SPAM

In the US, the CAN-SPAM Act[47] requires that mail should include:

- Valid routing information
- A truthful subject field
- The sender's real physical address
- Appropriate labeling for adult content
- An opt-out mechanism

The use of address harvesting techniques like dictionary attacks (notably DHA), malicious

software, and the use of open relays are grounds for considering spam an "aggravated" offence under this legislation. Sadly, the result of this legislation (or at least the trend since its introduction – we can't claim a scientifically proven direct link!) has actually been to increase the amount of spam sent. Partly, this is because businesses are often excepted from some of the requirements, and partly because it has been largely unenforced (maybe unenforceable). The USA is still the world's number one producer of spam.

European Directive

In Europe, the Directive on Privacy and Electronic Communications (2002/58/EC)[48] requires member states to regulate direct marketing and "unsolicited communications" according to Article 13, which deals with "unsolicited communications". This, in essence, states:

1. "Automated calling systems" (including email) can only be sent to people who have given "prior consent": in other words, is based on opting in to receiving marketing communications, rather than opting out.
2. Where a "natural or legal person" (this seems to refer to organizations in contexts where they have the same legal rights as individuals) has a pre-existing relationship with a customer, they can use that customer's contact details to market similar products and services. However, they are required to include a free and simple opt-out mechanism with each marketing mail.
3. Member States must take appropriate (legal) measures.
4. Marketing mails that disguise or conceal the identity of the sender, or have no valid address to which to send "cease and desist" messages, are to be outlawed.
5. The legitimate interests of "subscribers" who are not "natural persons" must also be protected.

The detailed implementation of this directive can differ significantly between States. For instance, the UK has attempted to preserve "cold-calling" email to businesses by applying the relevant section of its legislation[49] to "individual subscribers" rather than to "natural or legal persons".

While these legislative measures have had some impact on regulating the behavior of legitimate businesses using email for direct marketing, they've had much less on "hard core" or "professional" spammers. This is a fairly common phenomenon: those who pursue criminal behavior (such as the frauds propagated through spam), do so anyway, and only those who wish to operate within the law will comply.

So what does help with the spam problem?

Spam Countermeasures

Blocklists (blacklists) remain a major resource for system administrators, ISPs and providers of managed anti-spam services.

Blocklists

These commonly list open relays and proxies, but may list other perceived offenders against good practice: for instance, one site[50] lists domains that don't conform to certain RFCs (Requests for Comment – documents that often define accepted good practice for Internet usage). Unfortunately, this can result in major inconvenience to a site that uses the list as a guide to blocking sites without realizing that it includes virtually the whole of the UK's health service and any German site within the ".de" top level domain. Sadly, where blocklists are used by an outsourcing provider or an ISP, it's possible for a sizeable proportion of "ham" (legitimate mail)[51] to be lost by overenthusiastic filters. (A better countermeasure using non-compliance with RFCs is to reject mail that doesn't come from a fully-qualified domain name (FQDN),[52] though this is also likely to result in the loss of non-compliant but legitimate mail. However, blacklisting continues to play a part in spam management at all levels: anti-phishing toolbars, for example. That said, their use and usefulness is constantly diminished by the vast and distributed nature of botnets. Where traditionally spam could be tracked to (and blocked from) a few open relay servers, or constantly offending IP addresses, the dynamic nature of botnets means that this is no longer such an effective strategy.

Reputation Services

Blocklists are a primitive form of "Reputation Service"[53]: these services assess the "spamminess" of a message according to what is known about the sender. Commercial reputation services may supplement their services with blocklists (DNSBLs), but tend to be much more discriminating about their blacklisting criteria and keeping their lists up-to-date, and can be very effective for larger enterprises. Indeed, a number of spam mitigation measures like SPF (Sender Policy Framework) and Sender ID (another approach to validating sender email addresses) are best suited to large organizations with direct control over their mail and DNS servers.

Greylisting

Greylisting[54] also requires control over the Mail Transfer Agent (MTA) so as to temporarily reject email from unrecognized senders, on the assumption that spam is "fired and forgotten" whereas the originating server will try to redeliver a legitimate message, unless

it receives a permanent error. This technique is currently very effective against the majority of botnet-generated spam, and is unique in still allowing all legitimate email (from correctly configured sources) to pass through. Unfortunately, this may not be a long term solution, as all that is required to circumvent it is a change of behavior in the way spammers send their mail.

Whitelisting

Whitelisting involves accepting an option only if it is a specific member of a pre-determined list of options, for instance, a list of applications that are permitted to install or execute on a protected system. In this case, the whitelist is usually a list of pre-approved email addresses, IP addresses and so on. This reduces the risk of unsolicited mail dramatically, but requires a mechanism for allowing unknown senders to be added to the approved list. Such mechanisms often involve some sort of verification being sent, which may entail nuisance value and administrative overhead and can also be vulnerable to exploitation as a phishing lure, or even as a trick to harvest legitimate addresses.

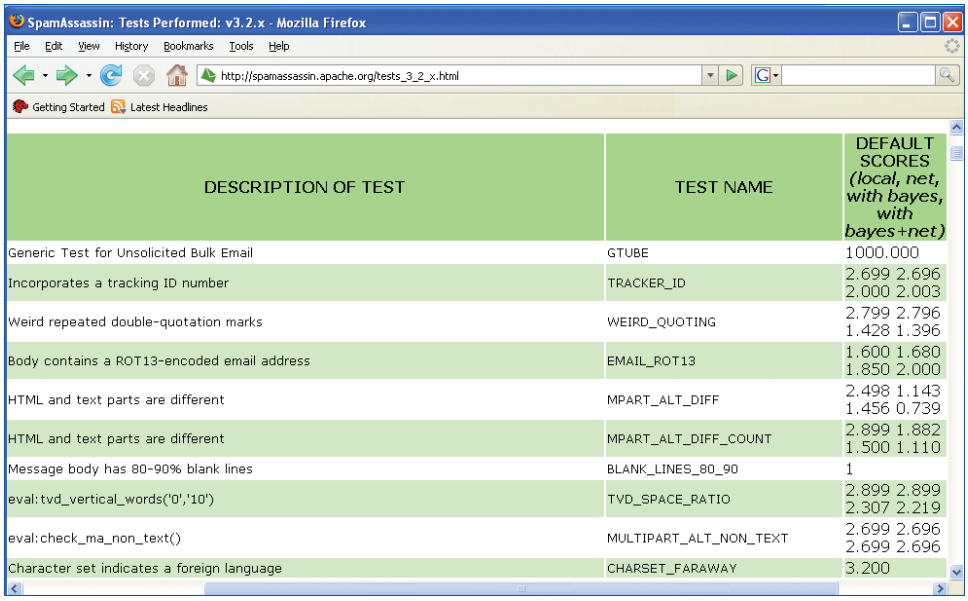
Text Filtering

Common text-based filters look for keywords or use Bayesian (statistical) methods to assess spamminess. Spammers have responded by mangling keywords (as described above) or by diluting the message text with "neutral" words to lower the proportion of "spammy" words present, thus reducing the effectiveness of Bayesian filters. However, the techniques used to confuse these types of filter are, in themselves, quite recognizable to heuristic filters, which in turn often work well in tandem with antivirus software, so it's an ill wind...[55]

Heuristics

Heuristic analysis as used in spam and malware filtering generally takes the form of analyzing messages and scoring them on a variety of assessment criteria: a threshold level is set (often this is configurable by an end-user or administrator), and messages that score above that threshold are assumed to be spam and processed accordingly. According to configuration, they may be flagged (for example, with "***spam***" added to the Subject header), quarantined or deleted. Heuristic scoring may be based on a wide variety of criteria, including elements of the countermeasures listed above. In addition, some of the detection techniques used by antivirus scanners (virus-specific and heuristic) such as wildcards and regular expressions, file and message header anomaly detection, and algorithms for seeing through obfuscation techniques, can be used to detect other kinds of spam, not just malware-related mail attacks. Figure 7 shows some example rules from the SpamAssassin ruleset, widely used and adapted by commercial antispam services and end sites (http://spamassassin.apache.org/tests_3_2_x.html).

Recently, there has been a great deal of excitement about image spam, where the text is actually embedded into an image (either on a remote site or attached to the message.) Since some filters (especially heuristic filters)[56] are suspicious of image-only or URL-only messages, these messages often include text to confuse them (random words, extracts from books and so on.) However, this content in itself is likely to alert a good heuristic scanner. Image spam has actually been around a good while. (Early phishes often used this approach.) More recently, spammers (especially pump and dump disseminators) have taken to using graphics that varied from message to message. As spam filters have become better able to detect these, spammers have moved on to sending graphic content as .PDFs (Acrobat-readable Portable Document Format files) and .XLS (Microsoft Excel spreadsheets) rather than .GIF or .JPG.



DESCRIPTION OF TEST	TEST NAME	DEFAULT SCORES (local, net, with bayes, with bayes+net)
Generic Test for Unsolicited Bulk Email	GTUBE	1000.000
Incorporates a tracking ID number	TRACKER_ID	2.699 2.696 2.000 2.003
Weird repeated double-quotation marks	WEIRD_QUOTING	2.799 2.796 1.428 1.396
Body contains a ROT13-encoded email address	EMAIL_ROT13	1.600 1.680 1.850 2.000
HTML and text parts are different	MPART_ALT_DIFF	2.498 1.143 1.456 0.739
HTML and text parts are different	MPART_ALT_DIFF_COUNT	2.899 1.882 1.500 1.110
Message body has 80-90% blank lines	BLANK_LINES_80_90	1
eval:tvd_vertical_words('0','10')	TVD_SPACE_RATIO	2.899 2.899 2.307 2.219
eval:check_ma_non_text()	MULTIPART_ALT_NON_TEXT	2.699 2.696 2.699 2.696
Character set indicates a foreign language	CHARSET_FARAWAY	3.200

Figure 7: Examples of Heuristic Ruleset for SpamAssassin

Commercial Anti-Spam

There are now many companies who offer anti-spam services. Often this means passing your email through their servers to have it checked, or having some sort of dedicated appliance on the mail gateway. Furthermore, some antivirus vendors are now offering anti-spam solutions as part of their package of protection, bringing anti-spam technology within easy reach of the end user. Typically, these technologies incorporate engines which identify spam using a variety of techniques, including Black/White/Grey-listing, Bayesian analysis, heuristics and others. These can be very successful in reducing the amount of spam with which the end user has to deal.

Conclusion: Living Spam-Free

Unfortunately, there are many ways of finding email addresses: even avoiding publishing your address anywhere doesn't stop you receiving spam. We've already discussed dictionary attacks in the form of Direct Harvesting Attacks, where a domain is bombarded with guessed-at addresses to see which of them actually exist. Automatic software called spambots, spiders, or crawlers search the Web, Usenet, Google Groups and similar forums, mailing lists and so on for valid addresses. Mass mailers and other forms of malware often scan an infected system for email addresses, and may also scan local network traffic and shared resources. "Spamware", software intended to support spammers in their endeavors, often includes automated tools for address harvesting, and huge lists of allegedly "real" addresses are often compiled, sold and re-sold. So while you can reduce the volume of spam that hits your mailbox by keeping your "public" Internet profile low (and, of course, by using some of the technical approaches already described), it's unlikely that you can avoid spam altogether except by using extreme filtering measures that decrease the general usefulness of your messaging facilities.

One obvious mitigating measure is to use more than one email address: one for essential business or private use, and only to be given to 'known' legitimate sources. Less jealously guarded addresses can be used where a public address is necessary for less critical traffic.

Where an address is used on a business web site, for instance, it can help to use measures such as a data capture form for email contact rather than simply displaying an easily harvested email address, perhaps augmented by a "captcha" graphic to make it harder for spambots to misuse the facility.

It's long been known that spammers sometimes include an "opt-out" mechanism that either has no effect whatsoever, or is actually used to confirm that your address is spammable (deliverable to), and therefore a candidate for adding to other spammer's lists. (This isn't to say that reputable senders of advertising mail don't use opt-out lists responsibly: but then, these are the same organizations that are likeliest to send mail sparingly and appropriately.)

Spammers also use tricks such as web bugs which "call home", confirming that the message has been delivered to a valid address. Many MUAs (Mail User Agents) like Outlook now offer the facility to accept and send mail only as text rather than HTML, and block embedded URLs (Uniform Research Locators), lessening the risks from web bugs, malicious scripts and so on.

Technical defenses work better when supplemented by good educational practices and well founded, clearly expressed and easily accessible policies.

References

1. http://www.spam-uk.com/fritterad.htm
2. Paul A. Vixie and Frederick M. Avolio: "Sendmail Theory & Practice" 2nd Edition; Digital Press, 2001.
3. http://www.spamhaus.org/definition.html
4. David Harley: "Stalkers on your Desktop" in "AVIEN Malware Defense Guide for the Enterprise" Ed. Harley; Syngress 2007.
5. Phillip Hallam-Baker: "The dotCrime Manifesto: Bringing Accountability to the World Wild Web"; Addison-Wesley 2007.
6. "There are 600,426,974,379,824,381,952 ways to spell Viagra": http://cockeyed.com/lessons/viagra/viagra.html
7. John Graham-Cumming: "The Spammer's Compendium"; http://www.jgc.org/tsc.html
8. Daniel J. Barrett: "Bandits on the information superhighway"; O'Reilly, 1996.
9. William Shakespeare: "Hamlet" Act III, Scene ii.
10. David Harley & Andrew Lee: "A Pretty Kettle of Phish"; http://www.eset.com/download/whitepapers/Phishing(June2007)Online.pdf
11. Alan Schwartz & Simson Garfinkel: "Stopping Spam"; O'Reilly, 1998.
12. http://en.wikipedia.org/wiki/Serdar_Argic
13. http://en.wikipedia.org/wiki/Spam_%28Monty_Python%29
14. http://en.wikipedia.org/wiki/Spam_%28electronic%29
15. David Harley: "E-Mail Threats and Vulnerabilities", in "Handbook of Information Security" Ed. Bidgoli; Wiley, 2006.
16. http://www.joes.com/spammed.html
17. Alan Schwartz & Simson Garfinkel: "Stopping Spam"; O'Reilly, 1998. Excerpted at http://www.oreilly.com/catalog/spam/chapter/ch01.html
18. http://en.wikipedia.org/wiki/Newsgroup_spam#_note-0#_note-0
19. http://www.templetons.com/brad/spamreact.html
20. http://en.wikipedia.org/wiki/Canter_&_Siegel
21. http://www.out-law.com/page-4583
22. David Harley & Tony Bradley: "Big Bad Botnets" in "AVIEN Malware Defense Guide for the Enterprise" Ed. Harley; Syngress 2007.
23. David Harley & Andrew Lee: "Net of the Living Dead:Bots, Botnets and Zombies"; in press, and will be available at http://www.eset.com/download/whitepapers.php
24. Chris Hunter, "Mobile SMS spam surpasses email spam in Korea, http://www.spamfo.co.uk/component/option,com_content/task,view/id,219/Itemid,2/, 2004.]
25. "SMS Spam and Fraud Prevention"; http://www.cisco.com/en/US/netsol/ns341/ns396/ns177/ns278/networking_solutions_white_paper09000aecd80250cb6.shtml
26. Igor Muttik, "A Tangled Web", in "AVIEN Malware Defense Guide for the Enterprise"; Syngress 2007
27. Larry Page, Sergey Brin, R. Motwani, and T. Winograd "The PageRank Citation Ranking: Bringing Order to the Web" http://citeseer.ist.psu.edu/page98pagerank.html
28. www.google.com/technology/ and www.google.com/corporate/tech.html
29. http://www4.law.cornell.edu/uscode/html/uscode18/usc_sec_18_00001302----000-.html

30. http://www.cs.rutgers.edu/~watrous/chain-letters.htm
31. David Harley: "The 419 Scam: an English Disease?"; http://watersidesyndication.com/content/?p=203
32. Martin Overton: "An African A-F-F-air" in Spam Bulletin, April 2007, also at http://www.momusings.com/papers/VBApr07.pdf
33. Robert Slade: "Dictionary of Information Security"; Syngress 2006
34. http://en.wikipedia.org/wiki/Trolling
35. Christopher Abad: "THE ECONOMY OF PHISHING: A Survey of the Operations of the Phishing Market"; http://anti-phishing.org/sponsors_technical_papers/cloudmark_economy_of_phishing.pdf
36. David Harley & Andrew Lee: "Phish Phodder: is User Education Helping or Hindering?"; Virus Bulletin Conference Proceedings, 2007.
37. Les Jones: "Good Times Virus Hoax FAQ"; http://cityscope.net/hoax1.html
38. Richard Dawkins, "Viruses of the Mind"; 1991. http://www.cscs.umich.edu/~crshalizi/Dawkins/viruses-of-the-mind.html
39. Harley, Slade. Gattiker: "Metaviruses, Hoaxes and Related Nuisances" in "Viruses Revealed"; Osborne, 2001.
40. Andrew Lee "Memetic Mass Mailers" Virus Bulletin July 2002
41. http://feeds.sophos.com/en/rss2_0-sophos-hoaxes.xml (accessed in August 2007)
42. Richard Dawkins, "River Out of Eden"; Weidenfield and Nicholson 1995
43. Susan Blackmore: "The Meme Machine"; Oxford University Press 1999
44. Richard Dawkins: "The Selfish Gene"; Oxford University Press, 1976
45. CIAC Hoaxbusters: http://hoaxbusters.ciac.org/HBHoaxInfo.html#recognizechain
46. David Harley: "Return of the Memetic Virus, or How the NHS was Swamped by the Tsunami"; presentation for UK CERTs
47. http://www.spamlaws.com/federal/can-spam.shtml
48. http://register.consilium.eu.int/pdf/en/02/st03/03636en2.pdf
49. Statutory Instrument 2003 No. 2426: The Privacy and Electronic Communications (EC Directive) Regulations 2003
50. http://rfc-ignorant.org
51. David Harley, "Hamfighting – How Acceptable are False Positives?" Virus Bulletin, July 2006'
52. EEMA, "Spam and e-mail Abuse Management", https://www.eema.org
53. David Harley, Reputation Services and Spam Control, http://www.ferris.com/2006/06/28/reputation-services-and-spam-control/
54. http://www.greylisting.org/
55. John Graham-Cumming, "Pseudowords": http://www.virusbtn.com/pdf/conference_slides/2005/John%20Graham-Cumming.pdf; John Graham-Cumming "The Spammers Compendium" at http://www.jgc.org/tsc.html
56. David Harley & Andrew Lee: "Heuristic Analysis – Detecting Unknown Viruses"; http://www.eset.com/download/whitepapers/HeurAnalysis(Mar2007)Online.pdf
57. Honeynet Project: "Know Your Enemy: Fast-Flux Service Networks, An Ever Changing Enemy"; http://www.honeynet.org/papers/ff/fast-flux.pdf

Glossary

419	Advance fee fraud commonly associated with Nigeria and other parts of West Africa, though by no means geographically restricted to that region.
Advance Fee Fraud	Fraud in which the victim is persuaded to part with money in the expectation of eventually receiving far larger sums.
Backscatter	Backscatter (or, sometimes, outscatter) is a term used when mail is bounced inappropriately in the form of a Non-Delivery Report (NDR) or Delivery Status Notification (DSN) to an address from which it didn't originate. This is usually a consequence of forged headers, as in spam, mass-mailers and other malware, and so on. The term is also sometimes extended to apply to "you sent a virus" notifications from poorly configured antivirus software.
Blacklist, Blocklist, Block List	A list of entities that are never allowed access or execution privileges in a protected environment. In the spam context, it usually denotes a list of IPs, mail addresses etc. that are believed to be associated with spammers, open relays or proxies, and so on.
Bot	In the context of spam dissemination, malware used to compromise a system so that it can be misused by a remote attacker, botnet owner and so on.
Botnet	A virtual network of bot-compromise PCs (zombies) controlled by a bot master/owner/herder, and used for various attacks such as distributed denial of service (DDoS) attacks, spam and scam dissemination and so on.
Captcha	Use of a human-readable graphic representation of an alphanumeric string to verify that the user of a facility is a human being rather than automated software. It's far easier (generally) for a human to read such a string than it is for software to decipher it.
DHA (Direct Harvesting Attack)	A type of dictionary attack used to harvest email addresses at a targeted domain. Addresses that seem to be valid are subsequently targeted for full spam attacks.
Dictionary Attack	An attack based on trying alphanumeric strings against software that requires specific strings such as correct passwords, email addresses and so on. The strings are not necessarily real words, as in a real dictionary: the list may be derived by going through all the possible alphanumeric permutations available up to an arbitrary length of string.

Dynamic DNS (DDNS)	System allowing a DNS entry to be updated in real time, so that it can be allocated to a host with a varying (dynamic) IP address.
Fastflux, Fast Flux	Fast flux DNS services are used by botnet owners to continually change the public DNS records for compromised machines, making it infinitely more difficult to trace and take down compromised machines. [57]
Greylist	A technique by which mail from an unknown source is initially rejected, on the assumption that if it's genuine non-spam, the MTA will retry delivery, and in this case it will be accepted.
Ham	The opposite to spam, i.e. legitimate mail.
Harvest	In spam management, usually applied to the process of acquiring email address to target for spam, phishing and so on.
Headers	In the spam context, the part of a message or article that contains routing information such as the sender's address and originator IP address, destination address, as well as other information. "True" spam almost always contains falsified header information.
HTML	Hyper Text Markup Language: the basic lingua franca of the World Wide Web
Image spam	Spam presented as an image rather than as text, to make text-based spam filters less effective.
IP (Internet Protocol) address	An address, usually expressed as four decimal numbers separated by dots e.g. 192.168.1.54, which uniquely identifies a host/device on a network.
Joe Job	Spam falsified (e.g. by forging the headers) so that it appears to come from an innocent party.
MTA	Mail Transfer Agent: a mail server that routes mail onward to other MTAs or to an MDA (Mail Delivery Agent) for delivery to the MUA
MUA	Mail User Agent: an application like Outlook Express, Eudora, Apple Safari and so on which interfaces between the end user and a mail service.
MUD (Multi-User Dungeon)	A multi-player computer game which may include some kind of chat facility.
Mule	In the spam context a mule (money mule) is an individual used as part of the money laundering process.
Multi-Level Marketing	A legitimate business model combining direct marketing and franchising, which is, however, often confused with illegitimate pyramid schemes.

Open Proxy	A server mediating between another server and a client process, but which doesn't verify the client.
Open Relay	A mail server that forwards mail to other MTAs without verifying the source.
Opt-in/Opt-out	Opt-in lists don't subscribe you unless you specifically ask to be added. Opt-out lists add you without asking, but have an explicit mechanism for opting out.
PageRank	Google uses so-called "PageRank" values to determine the quality of any web page, measured in terms of popularity. See the paper "The PageRank Citation Ranking: Bringing Order to the Web" by Larry Page, Sergey Brin, R. Motwani, and T. Winograd, at http://citeseer.ist.psu.edu/page98pagerank.html.
Phish	A scam aimed at stealing sensitive data, usually by persuading the victim that the scam message is sent by a legitimate concern such as eBay, a bank, the IRS and so on.
Ponzi Scheme	Investment fraud where initial investors may get high returns out of the investments of subsequent investors. The scheme will inevitably collapse eventually because there are no real revenues other than the cash put in as investors are added: however, the scammer will already have pulled out with his pile, if law enforcement haven't already intervened. Named after Charles Ponzi, a noted exponent of such schemes in the early 20th century.
Pump and Dump	A scheme for artificially inflating the price of stock by hyping its potential, then selling while the price is high. Sometimes used as shorthand for other types of stock manipulation through email and other media.
Pyramid Scheme	A scam that involves recruiting participants into a business that cannot be viable in the long term, because it doesn't usually offer a real service.
Reputation Services	Services that assess the "spamminess" of messages according to what is known about the source.
Sporgery	Portmanteau word derived from "spam" and "forgery", applied to posting a barrage of articles to Usenet newsgroups where the article headers have been forged to hide their true originator.
String	A sequence of alphanumeric characters: the term is often used to distinguish between machine-readable and human-readable character sequences, and has particular applications in programming.
UBE	Unsolicited Bulk Email: often used as a synonym for spam.

UCE	Unsolicited Commercial Email: a major subset of UBE, also used sometimes (less correctly) as a synonym for spam.
Vishing	Phishing scams carried out as a whole or in part by using telephony rather than email, especially using VoIP, which offers many ways of disguising the true source and geographical location of the abuser.
VoIP	Voice over IP (Internet Protocol): a telephony service based on Internet services.
Whitelist	A list of entities (in this context, usually email addresses or IPs) which are pre-approved. Non-members of the list are not allowed access or execution privileges.

Corporate Headquarters

ESET, spol. s r.o.
Aupark Tower
16th Floor
Einsteinova 24
851 01 Bratislava
Slovak Republic
Tel. +421 (2) 59305311
www.eset.sk

Americas & Global Distribution

ESET, LLC.
610 West Ash Street
Suite 1900
San Diego, CA 92101
U.S.A.
Toll Free: +1 (866) 343-3738
Tel. +1 (619) 876-5400
Fax. +1 (619) 876-5845
www.eset.com

TSI20081209