Drowning in Sewage

SPAM, the curse of the new millennium:
an overview and white paper.

Contents

3: *Propagation*

In which we examine how the disease is spread from evil instigator to hapless victim

3.1: Address harvesting – "*why* am *I* getting this rubbish?!?"

- How e-mail is sent and received – an overview
- Addresses on web sites
- Public postings
- Dictionary attacks
- List sales from e-commerce sites

3.2 Vectors of infection – "*how* am I getting this rubbish?!?"

- Open relays
- Open proxy servers
- Freemail services
- End-to-end delivery
- Zombie systems

4: *Prevention*

In which we examine the technical steps that can be taken to detect and reduce the amount of this dross we have to wade through each day.

4.1: Identifying the culprit before he gets a chance to send anything

- Blacklist services
- Whitelist services

4.2: Identifying the culprit's messages after receipt but before delivery

- Heuristic engines
- Networked vigilance – Vipul's Razor

4.3: Identifying spam once it's in your mailbox

- Heuristic engines
- Bayesian filtering

4.4: The limits of technology

- False Positives
- The Arms Race, and playing catch-up

5: *Proscription*

In which we examine spam from a legal standpoint and examine what legislative action has been, or could be taken to remedy it.

5.1: The key legislative issues

- Opt-in versus Opt-out
- Who can bring an action
- Penalties and enforcement

5.2: The situation in the United States

- Federal legislation
- CAN-SPAM (S.877)
- State legislation

5.3: The situation in the EU

- Broader Europe
- The United Kingdom

5.4: The situation in Australia

Appendices

1: The Problem

1.1 Seeking a definition of spam

After child pornography, there is probably no Internet-related subject that raises ire or passion quite as strongly as spam, and like pornography, people "know it when they see it". But intuition aside, *what is spam?* Before we can begin addressing the problem, we need to be able to define it, and as with many such emotive and charged issues, the process of definition is deceptively difficult.

Before we begin attempting to tie down a definition of spam, please note that in this document we use the convention requested by the Hormel Corporation when referring to spam – we always present the word in all-lowercase text. It is the author's belief that the Hormel Corporation has already lost this battle and that their trademark is now permanently derogated, but as a courtesy to them, we will observe their wishes. With that clarification in mind, let's start our search for a definition of spam by examining its most obvious characteristics:

- It is always unsolicited – the recipient has not explicitly given consent to receive it.
- It is typically commercial or promotional (or sometimes illegal) in nature
- It is usually sent en masse.

Outside these three primary characteristics, no firm or reliable guidelines exist. spam varies enormously in format, content and technical construction, and cannot be generally reduced to much more than the list of points above. Note, however, the hedging in the second and third points – spam "is typically" commercial or promotional, but examples exist that anyone would call spam, where this condition does not apply. Similarly, while spam is "usually" sent in great bulk, the recipient has no clear indication that this is so: when you receive a spam message in your mailbox, you have no idea how many other people have received it – it might be none, or ten million, but this does not diminish or alter your perception of the message as spam.

Clearly then, the only characteristic of spam that we can identify categorically is the fact that it is always unsolicited: this must therefore form the fundamental underpinning of any definition of spam. Unfortunately, simply declaring that all unsolicited mail is spam would be severely over-simplistic: it would create a definition that effectively included all mail sent to a person with whom the sender had not previously corresponded, which is clearly nonsensical. As with most things in life, a degree of compromise is required.

1.1.2: The UBE vs UCE debate

As stated above, it is unlikely that anyone would disagree that all spam is unsolicited. Any examination of the major web resources covering the spam issue, however, quickly reveals that there is significant disagreement over the dominant secondary characteristic of spam. On one side of the gulf are those who believe that it is the bulk nature of spam that defines the problem, while on the other are those who believe that it is the commercial nature of the message rather than its bulk that is at the core of the definition. These two attitudes to spam can be categorized as the *UBE Camp* (Unsolicited Bulk E-mail) and the *UCE Camp* (Unsolicited Commercial E-mail), and the argument between the two groups has all the hallmarks of a religious war. For an example of a major site that unequivocally defines spam as UBE, see Spamhaus (http://www.spamhaus.org/definition.html). For an example of a major site that equally unequivocally defines spam as UCE, see CAUCE (http://www.cauce.org/about/faq.shtml).

A moment's consideration should reveal that *both* UBE and UCE are valid views of spam – they simply approach the issue from different directions. The UBE Camp views the problem of spam from the perspective of people or organizations who are attempting to solve or regulate it – you might say that they are approaching it from the *Back End*. For people who are looking at spam in this way – in terms of its overall impact on the infrastructure of the Internet – the fact that it travels in vast shoals of traffic and consumes ferocious quantities of bandwidth is the point of overriding concern.

From the perspective of the typical recipient of spam, however, the question of how many other people received any particular item is almost invariably irrelevant – the content of the message is the only issue of concern to these people, amongst whom one can with reasonable confidence include the majority of the Internet's population. From this point of view, which one might refer to as the *Front End* of the problem, the Commercial content is the dominant secondary characteristic that defines spam.

It is the author's contention that no effective definition of spam can avoid encompassing both the UCE and UBE positions – neither point of view is entirely satisfactory on its own. We suggest that any attempt to regulate spam must use a definition that focuses on its bulk aspect (because that is almost invariably easier to tie down in tight, quantifiable legal terms) while also taking significant account of the impact of its content – in other words, an effective definition of spam is necessarily going to be a hybrid of UCE and UBE.

1.1.3: Criminal and illegal spam

One class of spam that is becoming increasingly common is that which proposes a clearly criminal action, or promotes a product or service that may be illegal. We will see examples of this in section 1.3, "419's and related scams". There is room for argument that this type of mail is not actually spam, but it is our belief that most people would not make this distinction and that any efforts taken to reduce or eliminate "standard commercial" spam will also be applicable to this type. In light of this, we propose to include this class of spam explicitly in our definition.

1.1.4: Twisting the meaning of "unsolicited"

A complication in any attempt to define spam is what exactly we mean by the word "unsolicited": a common argument from pro-spam and Direct Marketing Organizations is that there are situtations where a pre-existing business relationship exists between a customer and a vendor – for example, in the case where a customer has supplied his e-mail address as part of the process of purchasing goods from the vendor's web site. Many major anti-spam organizations, such as CAUCE, maintain that promotional e-mail must only be sent to customers who have explicitly agreed to receive it by deliberately checking a control in authorization, and that the default state of that control must be "unchecked". This process is known as "opting-in", and as we shall see in section 5.1, it is one of the fundamental issues that must be confronted in any initiative that aims to control spam.

1.1.5: A tentative definition

In light of the preceding discussion, we now offer a tentative definition of spam that we can use as a basis for further discussion:

> *Spam:* An electronic communication containing material or references to material of a commercial, solicitational or illegal nature, directed as part of a bulk distribution to any address where the addressholder has not given explicit prior consent to receive it.

1.2: Things that look like spam but aren't

Quite a lot of mail that is considered "spam" by the public is not in fact spam, and could not be controlled by any formal anti-spam initiative. The most significant example of this is *Viral Vector* mail – messages that attempt to lure the recipient into running an executable attachment, or that use weaknesses in the HTML rendering facilities of common mail programs to execute code without the recipient’s knowledge. This type of mail typically results in the recipient’s computer being “infected” and in the Viral Vector message being forwarded to everyone in the recipient’s addressbook.

Many Viral Vector messages use spam-like enticements to encourage the recipient to activate their payloads – for example, promises of pornographic media. It is also common for these messages to attempt to obscure their payload by manipulating filenames so that only an innocuous part shows in the recipient’s mail program (“Anna-Nicole Smith Nude.JPG.exe”, for example), or by putting the payload into a ZIP file so as to bypass antiviral protections.

It is worth stressing that controlling the spread of Viral Vector e-mail cannot be part of any concerted anti-spam initiative and as such this type of mail is outside the scope of this paper: we mention it here only to distinguish it from spam.

1.3: Types of spam

In this section, we offer a brief overview of various types of spam and their basic characteristics. This list cannot be considered anything other than representative, since no list could embrace all of the many types and variations of spam that exist, but it will serve to illustrate the general form of the problem.

1.3.1: Spam tools

In many ways one of the strangest types of spam is that which advertises tools for sending more spam, such as bulk mail programs, lists of millions of e-mail addresses, or spam-friendly ISP services. An example of this type of spam is shown in Appendix A.1. In the early days of spam, in 1996-97, this was the most common type of spam by far, and it is still fairly common, although it has been overtaken in volume by other types. One of the more worrying types of spam in this category has recently begun to appear: it is spam promoting anti-spam tools! In several cases, these “anti-spam” tools have actually been found to be Trojan Horse programs that turn the recipient’s computer into a “zombie” – a machine that spammers can use as a form of proxy server for sending more spam (see section 3.2 for more discussion on this).

1.3.2: Quack remedies

One of the fastest-growing types of spam, this class attempts to sell wondrous cures for all that ails you. It is not uncommon to see cures for cancer, AIDS and other serious diseases advertised this way, and spam promoting “Human Growth Hormone”, or “HGH” as a wonder drug is extremely common. In the USA, both the Food and Drug Administration (FDA) and the Federal Trade Commission (FTC) have determined that this type of spam is routinely illegal because of the type of product it is advertising, and there are well-documented cases where the “product” is actually nothing more than repackaged vitamin pills. An example of a spam of this type is shown in Appendix A.2.

1.3.3: Vanity and insecurity

This class of spam preys on personal insecurity, almost always in males. The most common type of spam in this category by far is the “Penile enlargement” message, where the product “guarantees” substantial growth of the male member. As with quack remedies, these products are routinely misleading or felonious. Other types of spam in this category include special pills to reduce body odour, weight-loss programs, and aphrodisiacs. An example of a typical spam of this type is shown in Appendix A.3.

1.3.4: 419s, illegal products and related scams

This type of spam takes the form of a plea to help “process” a very large amount of money by allowing it to pass through your bank account, in return for which you will be given an implausibly large share (millions of dollars) for your trouble. This scam, known as a “419” after the section of the Nigerian Penal Code that it apparently breaches, almost always originates in Africa, especially Nigeria, Liberia, Sierra Leone or other largely lawless or corrupt regions. 419s are usually so transparent and poorly-done that they can be quite laughable – indeed, it has become a fairly common sport on the Internet to “bait” 419 scammers. Nonetheless, it appears that a depressing number of people are still stupid enough to fall for them, and they have been getting increasingly common.

A 419 can take two general courses: in the first, after a long series of e-mail exchanges, the “deal” is about to be made, but it requires a “disbursement” of a thousand or so dollars to “facilitate the transfer”, which for technical reasons only the mark can make. Naturally, once the mark makes the disbursement, the “deal” evaporates and the money is gone. The second form is considerably more sinister: in this, the deal eventually requires the “mark” to travel to the destination country to “pick up their share”. The mark is met at the airport by the perpetrators and is immediately taken hostage and held to ransom. This second form of 419 is much less common than the simple fraud version, but has happened on occasions.

There are several variations on the 419, the most common being a notification that you have won several million dollars in an international lottery that you never entered. This type of scam appears to originate almost without exception from the Netherlands, and almost always follows the same general form as the “disbursements” type of 419.

It is worth noting that almost all 419 scams entice the mark to undertake an activity that is illegal in almost any country. Given this, and the pathetically transparent nature of the scam, it is almost unbelievable that anyone could fall for this type of thing, but amazingly enough, they do. Appendix A.4 shows a very typical 419 spam.

As well as scams of the “419” variety, many spams promote products that exist solely for an illegal purpose: the most obvious example in this category is the very common class of spam advertisements for “cable descramblers” – devices that allow people to intercept pay-TV services without paying for them. Spams like this fall into that bizarre grey area of the law where the product being sold is not illegal, but attempting to use it in any way most certainly is. In many instances this type of spam is in reality a “scam”, either because no product exists, or because it is never shipped. The scam succeeds because of the unlikelihood of the dupe going to the authorities and having to explain why he wanted to purchase an item for which only illegal uses exist.

1.3.5: Pornographic spam

This type of spam promotes pornographic material, such as web sites containing explicit videos, sexual DVDs and so on. After spam tools, this is probably the oldest type of spam, and the form most commonly-associated with the whole problem in most peoples’ minds, even though in reality it makes up only a relatively small percentage of the total volume of spam sent each day. Because of the content, some of which can be disturbing even to the most hardened sensibility, this type of spam is probably the most controversial, and the one most likely to precipitate legislation or governmental intervention. It is an alarming reality of the Internet that more and more children and young people are going online and being exposed almost immediately and continually to this type of material.

Spammers have tended to adapt their techniques to bypass attempts to filter or restrict their “goods”, and in no other type of spam has the amount of effort expended on such adaptation reached the levels it has with the pornographic variety. This makes pornographic spam the hardest type to detect and defeat by purely technical means.

Three examples of pornographic spam are shown in Appendix A.5, A.6 and A.7: note in particular the means used to bypass or subvert standard technical methods of spam prevention in these examples.

1.3.6: Foreign-language spam

One of the fastest-growing types of spam on the Internet originates from the "badlands" of the Internet, most notably China, Korea and parts of Russia and the former USSR. The almost total lack of regulation and enforcement in these places means that spam is now spewing out of these regions in alarming quantities. What is more, this type of spam is almost always unreadable on any English or European-language computer system, because it uses 16-bit characters or non-standard encodings. The nett result is a message that simply looks like garbage characters on any Western display. It is usually impossible to determine what is being advertised in this type of spam unless you happen to be in the country of its origin, yet it is being distributed by the million to people who can make nothing of it. As such, it represents an enormous and total waste of bandwidth. On the positive side, it is usually fairly easy to filter by technical means. A typical example of this type of spam is shown in Appendix A.8.

1.3.7: Bizarre spam

Anyone who has to deal with spam regularly will have a few items in their collection that defy belief – primarily because it is so hard to see what the spammer hoped to achieve. As an example, the author routinely receives two or three spams every week offering to clean or repair his septic tank, even though he has no such thing and the people offering the "service" are based in Taiwan. Similarly, there is the Chinese factory that wants to sell tractor parts, and finally, there is the author's favourite spam of all time (shown in Appendix A.9) where (we think) the spammer was trying to sell milk vats.

1.4: The cost of spam

One comment often heard about spam is that it is no different from conventional direct marketing – the junk mail that accumulates in everyone's real-world mailbox every day – but this is simply not true. With direct marketing, it is the person doing the marketing who pays the cost associated with preparation and delivery: similarly, with telemarketing, it is the telemarketer who pays the cost of the phone call. With spam, however, the bulk of the cost is actually borne by the recipient and by the recipient's ISP. The recipient uses connection time (for which he is typically paying on a metered basis) to retrieve the message, and quite possibly more later if the message contains Lazy HTML graphics (see Appendix B); the message also wastes space in the recipient's disk allocation. The ISP uses bandwidth to receive the spam, and pays for administrator time and licenseing fees to maintain software to trap and remove the spam before it gets to the recipient.

Put in simple terms, comparing spam with direct mail is like suggesting that it should be possible for a marketer to send you promotional material delivered COD. Comparing spam with telemarketing is like suggesting that telemarketers should be able to call you collect. Both of these are nonensical propositions: spam is not like traditional direct marketing – it is an exercise in cost-shifting.

There are numerous overviews on the web of the actual cost of spam, both in dollar terms and in terms of damage to the infrastructure of the Internet: here are two:

http://www.cauce.org/about/problem.shtml
http://www.novell.com/info/collateral/docs/4820891.01/4820891.pdf

1.5: The rationale for spam

Most people probably wonder why spammers bother, because "surely nobody in their right mind would be interested in the stuff they're trying to foist, or stupid enough to fall for such obvious and transparent scams?" To understand why this argument doesn't work, you have to remember that e-mail is very nearly free – the costs involved in sending a million e-mail messages are minuscule, which means that spammers can make a profit on astoundingly low response rates.

In a traditional direct-mail marketing campaign, a response rate of 2 – 4% is considered good, and is usually profitable enough to justify the campaign: but direct mail is costly – each item has to be sent or delivered, which means that there is a distinct price to pay for the response rate achieved. With e-mail, it is actually the *recipient* who carries the bulk of the cost of the message – the spammer usually pays little or nothing to send the material. This means that spam can be quite profitable on astonishingly low response rates.

Imagine for a moment that a spammer promoting penile enlargement pills for $29.95 a bottle sends out ten million spams – a very moderate number by modern standards. If the spammer gets a response from one in every thousand messages he sends (0.1%), he will sell 10,000 bottles, for a total return of $299,500.00. Even on response rates as low as one in a million, operations like this can still turn a profit that makes them worthwhile, simply because of the enormous number of addresses that can be reached at almost no cost.

There is anecdotal evidence that some spams can achieve response rates as high as one percent, which means that the amounts of money involved in spamming can be quite enormous. As well as "legitimate" spam, though, where goofy products are sold to gullible people, there are other, less salubrious ways in which spam can be profitable. Obviously scams like those described in section 1.3.4 are one such method; others include identity and credit card theft, so-called "Pr0n diallers", where accessing the pornographic site in the spam requires using a special "dialer" package that dials 0900 numbers, and even strange, subtle revenue generation through "ad impressions": in this last form, the spammer gets a small amount for every hit on a particular web site, so he sends out a message with an "unsubscribe" link directed to that site. There is an excellent discussion of the "seedy side" of spam economics at http://cc.uoregon.edu/cnews/summer2003/spameconomics.html.

Make no mistake about it – spam is profitable, in any of a number of ways: there is an almost total lack of appreciation amongst public and government alike of the scale of the spam problem, and the sheer volumes of money it involves. As long as the profit motive remains, so will the problem.

2: The Protagonists.

So who are the people involved in spam? Who are the people authorizing it, sending it and fighting it? There are several categories of groups and individuals involved.

2.1: The “Bad Guys”

2.1.1 The Vendor – the person selling the ‘product’.

As discussed in section 1.5, if there is no product to sell, then there is no spam. Ultimately, spam is about marketing a product or service, and for certain types of product or service, it is a very effective medium. For many types of spam, the vendor – the person who ultimately makes the sale – is obvious: for porn spam it is the operator of the web site where the porn is sourced; for spamware, it is the seller of the spam tools (often also the distributor)... But a worrying development is the growing number of “main stream” vendors who may now be showing an interest in the use of spam for promotion. Evidence for this is anecdotal at best, but with the growing awareness of the Internet as “the new marketing medium” by businesses around the world, there is clearly a troubling potential here.

When talking about spam, it is important to understand the distinction between the vendor and the distributor: the former is often largely ignorant of the disgust in which spam is generally held, and merely sees the process as another form of marketing. In any survey of web sites covering spamming, it is depressing how often this type of misapprehension is reported. While vendors should not completely escape the opprobrium associated with the spam for which they are ultimately responsible, proper education and increased public awareness of the issues are probably more helpful than beatings and hate-mail when dealing with them.

2.1.2: The Professional spam distributor

In this category we find the true villains of the piece. These are people who have made a career out of sending spam, almost always on behalf of “vendors” as described in 2.1.1: they have typically developed extensive infrastructure to support their efforts, and there is an arsenal of “tricks of the trade” they use to continue operating without being shut down. Rather than spending time detailing the activities of individual spam distributors in this paper, it is probably better to refer the reader to a couple of key web sites that devote extensive coverage to their activities: the first is the Spamhaus Project’s ROKSO database –

http://www.spamhaus.org/rokso/index.lasso

while the second is a case study of a typical professional spam distributor:

http://www.toledocybercafe.com/ivtg/

There is also an extensive range of links to sites covering the activities of known professional spam distributors (as well as many other anti-spam resources) at

http://dmoz.org/Computers/Internet/Abuse/Spam/

The ROKSO database is a particularly dense archive of hardened abusers and goes into considerable technical depth covering the methods they use.

2.1.3: “Mom and Pop” spam operations

This category covers a class of spammer that is only anecdotally understood – small-time operators, apparently often operating on a contractual or loosely-linked basis with professional spam distributors. This type of spammer is typically only a short-term phenomenon, usually operating by opening accounts with ISPs, spamming from them until they are closed down, then moving on to another ISP in the same area. Over time, spammers of this type become well-known in an area and they seldom operate for any length of time. It seems that the majority of spammers in New Zealand fall into this category [Ref S. Lyall, IHUG, at Uniforum Conference Auckland 2003]. The primary evidence that this type of spammer exists at all in any quantity comes from spam soliciting small-time operators (see Appendix A.10 for an example).

2.1.4: spam-friendly ISPs

As the name suggests, this category covers Internet Service Providers who claim to offer hosting services to spammers, shielding them from disconnection or other retribution from disgruntled users. In actual fact, it’s unlikely that any true ISP could provide infrastructure that would allow them to offer this kind of service – resistance to spam is now sufficiently highly-organized that such services would be unsustainable. Although the author has a number of spams on file offering such services, a survey of current anti-spam web pages does not seem to indicate that any such ISPs are prominently active at this time. It is likely that most firms starting out with the aim of offering spam-friendly ISP services metamorphose rapidly into professional spam distributors (see section 2.1.2, above).

2.1.5: spamWare developers

There is the old joke that “guns don’t kill people, people kill people”, which is often used by weapons manufacturers to absolve themselves of associative guilt. Well, in just the same way that a market demanding weapons will always find plenty of people who have no ethical problem supplying them, so spam has spawned an entire industry that revolves around the development and supply of specialist tools for harvesting addresses, sending spam and covering tracks. It is not the place of this white paper to pass judgment on such developers, but there can be no doubt that the existence of such a rich supply of tools does not alleviate the problem of spam in any way. For an excellent list of some of the products spammers can purchase to peddle their wares, visit this link:

http://dmoz.org/Computers/Software/Internet/Clients/Mail/Windows/Bulk_Mailers/

(Note that the URL may have been wrapped by the word processing software – in its proper form it is all on one line).

2.1.6: The Wild, Wild East

As Asia has moved into the Internet age, so has it also moved into the spam age. We have already mentioned (in section 1.3.6) the growing wave of foreign-language spam coming out of Asia, Russia and parts of the former Soviet Union, but it is important to stress the potential these areas have for completely overwhelming the infrastructure of the Internet with spam. China is currently the fastest-growing nation in the world in terms of availability of Internet services, and it is already clear that regulatory structures and enforcement facilities simply do not exist within the PRC to act against the spread of spam resulting from that growth. While in 2003 some 75% of spam is anecdotally believed to originate in the United States of America, expect to see this percentage skew dramatically towards Asian spam in the next four or five years.

2.2: “The Good Guys”

Organized resistance to spam has been surprisingly limited and fragmented since the phenomenon became a significant problem, and even now the movement is hurt by an overall lack of a coherent co-operative anti-spam strategy. The list below is a short selection of organizations or sites that have been active in the anti-spam fight for some time.

2.2.1: CAUCE and its affiliates.

The Coalition Against Unsolicited Commercial E-Mail was started by Scott Hazen-Mueller and a number of other campaigners in 1997. Many of the members of CAUCE have gone on to take active and well-publicized roles in the anti-spam movement – for instance, Ray Everett-Church has acted as an expert witness before committees of the US Congress on several occasions (http://www.everett.org/testimony/).

CAUCE originally concentrated on lobbying to have a US Fax Advertising Statute, 47 USC 227, extended to cover spam. This initiative was rapidly bogged down amidst pressure from pro-spam interest groups such as the US Direct Marketing Association, and CAUCE has now moved on to a general advocacy role. CAUCE has spun off a number of world-wide affiliate groups, including one in Australia (http://www.caube.org.au/), one in Europe (http://www.euro.cauce.org/), one in India (http://www.india.cauce.org/) and one in Canada (http://cauce.ca/). The CAUCE home page is found at http://www.cauce.org/

2.2.2: MAPS (Mail Abuse Prevention System)

MAPS was originally set up by Paul Vixie, the father of the Internet DNS system: this gave it a certain cachet, and for a number of years it was a major player in the anti-spam world. It was also the site of one of the very first Internet Blacklist services, the MAPS RBL (Realtime Blackhole List), which acted as a DNS clearing house listing sites known to be involved in or supportive of spamming.

In 2001, MAPS moved to a conventional business model and appears to have been distancing itself somewhat from the general fight against spam, although its blacklist services are still widely-used, albeit quite expensive. The MAPS home page and information on the RBL can be found at http://www.mail-abuse.org/ .

2.2.3: The Spamhaus Project

There are many Blacklist and anti-spam systems on the Internet now, some of them rabid almost to the point of insanity: some, however, stand out by being even-handed, balanced and uniformly reliable. The Spamhaus Project (http://www.spamhaus.org) is one of these. Particularly aggressive against hardened spammers, Spamhaus offers a carefully-maintained blacklist service and has an excellent repository of information about the top 100 spammers in the world: this repository is known as ROKSO (Register Of Known Spam Operations).

2.2.4: David Sorkin’s SpamLaws site

An excellent resource covering the state of anti-spam legislation in the USA and Europe, it also has useful overviews of legal developments in other countries. The site offers no opinion – it is purely a collection of related links organized in a helpful form. http://www.spamlaws.com/.

2.2.5: SpamCop

Primarily a spam reporting and blacklist site, SpamCop (http://www.spamcop.net) has been around since 1998. SpamCop is usually considered somewhat too aggressive by any except the most fanatical anti-spam activists, and recommendations for it are usually hedged with warnings to this effect. Nonetheless, it has been fighting the fight for a long time and remains an important force in the fight against spam.

2.2.6: Too many others to mention (links)

Section 2.2 of this paper is not intended in any way as anything except a representative sample of organizations and individuals doing what they can to fight against spam. There are many others, far too many to cover in a short overview of this kind. The links below will take you to convenient lists of anti-spam activists which you can browse.

http://www.spamlaws.com/links.html
http://www.cauce.org/about/resources.shtml
http://dmoz.org/Computers/Internet/Abuse/Spam/
http://www.elsop.com/wrc/nospam.htm

2.3: The Grey Zone

One major section of the business world has a greater potential impact on the spam issue than any other, and it is not entirely clear at this stage whether it has fully shown its hand – that is the global community of Direct Marketing Associations.

Depending on whom you speak to, Direct Marketing Associations are either Good Guys or Bad Guys. Those who place them in the Good Guys grouping do so because they are well-funded professional organizations who have the power to bring responsible forms of self-regulation to the problem of spam, while those who place them in the Bad Guys grouping do so because until recently, Direct Marketing Associations have routinely been in favour of legislation that legalizes opt-out spam. As we will see in section 5.1, the distinction between opt-in and opt-out spam legislation is going to end up being the single most important issue in the formal regulation of spam, and the danger that a powerful, well-funded lobby group such as the DMA might promote opt-out legislation is a matter for concern.

Ultimately, though, organizations like the DMA have a professional interest in maintaining high ethical standards, and we believe that they will rapidly grasp the issues underlying spam once those issues are made clear to them – indeed, the New Zealand DMA is already apparently taking a very responsible position on direct marketing via e-mail, issuing a revised eMarketing Standards document that firmly mandates reputable opt-in marketing techniques (see http://www.emarketingcouncil.co.nz /pdfs/Standards_for_email_Use.pdf). Involving and informing organizations like the DMA in the fight against spam will help ensure that they become a major factor in resolving the problem.

2.4: The Neutrals, and those on the periphery

There are two remaining groups of players in the spam wars who deserve mention here.

2.4.1: Spam detection services and developers

As the spam epidemic has increased, market niches have opened for firms that are able to provide mail scanning and spam detection services, or software that can filter or remove spam before it reaches the end user's mailboxes. These people are indirectly profiting from spam without directly promoting it – but a side-effect of their presence in the industry is occasional bouts of fear-mongering, designed to promote sales. This type of over-statement has been seen for many years in the pronouncements of anti-viral vendors – a little panic or exaggeration does their sales no harm at all, but it does little to assist in informed public debate on the subject. For an example of this type of "hyping-up" of a product, see

http://www.theregister.co.uk/content/55/31103.html

The author is not suggesting in any way that organizations like this have any vested interest in seeing the spam problem continue – but the fact remains that their position is a peculiar one.

2.4.2: Governments

Historically the Internet has resisted Government control, and in what can only be considered a surprising level of tolerance, most governments have avoided interfering in Internet issues, preferring for the most part to allow the system to regulate itself. As the spam epidemic spirals out of control across the globe, however, more and more governments are being dragged, however reluctantly, into dealing with the problem.

Whether or not national governments should be involved in attempting to police or regulate an international network is an ancient and contentious issue, but whatever the right or wrong of the argument may be, it looks certain that anti-spam legislation is going to be enacted all over the world. There is a real risk here, though: if the *wrong* legislation is enacted – if governments are overwhelmed by commercial lobbying, or fall back on knee-jerk reactions to the problem, then there is every chance that we may end up getting legislation that *exacerbates* the problem of spam rather than alleviating it. It is critical that the anti-spam lobby presents a united front with clear, sensible arguments backed by compelling evidence to ensure that the legislation we get is legislation that actually minimizes the problem as much as possible.

3: Propagation

This section aims to provide a basic technical overview of how and why you end up with so much spam in your new mail folder each day.

*3.1: Address harvesting – "*why *am* I *getting this rubbish?!?"*

"Where did the @$#%!% spammer get my address"?

This plaintive cry is heard every day as frustrated people sift through inboxes full of penile enlargement ads, viagra promotions and offers of multi-million address lists. How do spammers get the addresses to which they send their wares? There are various ways your e-mail address can fall into the hands of a spammer:

3.1.1 Addresses on web sites

An experiment performed in late 2002 by the US Federal Trade Commission (see http://www.ftc.gov/bcp/conline/pubs/alerts/spamalrt.htm) set up a number of e-mail accounts and explored the most likely ways of getting those accounts into active spam databases. The conclusion of the experiment was that far and away the most rapid and infallible way to get spammed was to expose your e-mail address on any page of a publicly-accessible web site. For many, this result was a surprise, since prior to that time, public postings on mailing lists or news groups was considered the most likely way of getting harvested (indeed, you will still find anti-spam sites that state this).

Another survey of address harvesting that used similar methodology and came up with similar results can be found at http://www.cdt.org/speech/spam/030319spamreport.shtml. (Indeed, the methodology, results and timing are so similar that there is a possibility it is the same survey, although if it is, it is being reported by different organizations).

3.1.2 Public postings

In the late 1990s, when spam was only just becoming a problem, the surest way of getting your address into spam databases was to post articles on netnews. Experiments like the one described in 3.1.1 have concluded that public posting is still a common way of getting onto spam lists.

3.1.3 List sales from e-commerce sites

Subscribing to a paid pornography web site is a guaranteed way of getting onto numerous spam lists, especially those promoting pornographic material. There is clear evidence that even porn sites with privacy policies stating that they will not pass on your e-mail address will in fact routinely do so. What is less well-known is that even providing your address to quite reputable e-commerce sites can result in your receiving spam: the anecdotal evidence is that in many cases employees will steal client lists and sell them to spammers, even if the firm itself is responsible and has solid privacy policies.

3.1.4 Dictionary attacks

An increasingly common way to get spam is via dictionary attacks: the spammer connects to a mail server and starts attempting to send mail to hundreds or thousands of addresses constructed by attaching common usernames to the mail server's domain name. Every now and then it will construct a combination that is valid, and the mail server will accept it, at which point the spammer adds that address to his list. Note that this process is completely automated and is going on all the time – the author's own experience is that he sees between twenty and thirty dictionary attacks every day on his mail server. Technical approaches exist to circumvent this problem, but as we will show in section 4.4.2, this is part of the process of playing *catch-up*.

3.2 Vectors of infection – "how am I getting this rubbish?!?"

The means by which a spam e-mail gets into your mailbox are quite varied: this section aims to provide a basic technical overview of how spam propagates.

3.2.1: How e-mail is sent and received – an overview.

Electronic mail is carried across the Internet using a protocol called SMTP ("Simple Mail Transfer Protocol"). In over-simplified terms, SMTP involves one system ("the client") connecting to another ("the server") and saying "I have a message from x". When the host server has responded that the sender's address is OK, the client continues by saying "The message is addressed to y" for each recipient. At this point, the server can face any of the following three scenarios:

- It can recognize the address as one it serves and accept the message for delivery.
- It may decide that the address is non-local, but agree to accept it and pass it on to the proper server at a later time. This process is called *relaying*, and is very important.
- It may decide that it cannot do anything with the address and decline it.

Once the client has submitted all its potential recipients, if at least one has been accepted by the server, the client will proceed to pass the data for the message, which consists of the message headers, a blank line, then the body of the message, which may be text or a combination of text and specially-encoded items that a mail program will interpret as attachments.

Relaying is a very important concept in this process, because it is the means by which your copy of Microsoft Outlook Express, or Eudora or Pegasus Mail sends a message to the Internet: it contacts your ISP's mail server and goes through the exact process above, relying on the server to accept the message and complete the delivery process on its behalf. A mail client normally cannot perform full end-to-end delivery, because the ultimate destination may be unavailable. By asking the ISP's server to perform the delivery, it can pass the task on to a computer that is always turned on and hence can retry the connection if the ultimate destination server is not online. When you send mail from a client via an STMP server in this way, it is more properly referred to as *Submission*, rather than *Relaying*, because Relaying is a general term applying to any situation where one mail system asks another to perform a delivery on its behalf.

Full information on the SMTP protocol is covered in a number of Internet standards documents (or "RFCs"), the most important being RFC2821 and RFC2822: these standards can be retrieved from http://www.ietf.org/rfc.

3.2.2: Open relays

The process defined in 3.2.1 describes how a mail client handles the delivery of Internet mail using a process called *submission*. Now, imagine for a moment that a computer completely unrelated to the ISP, somewhere else on the Internet altogether, were to ask it to do the same thing – to deliver a mail message on its behalf to a third system somewhere. This is known as *true relaying*, and if the server agrees to do it for any system that asks it to, then the server is what is known as an *Open Relay*. In the early days of the Internet, open relays were fine and in fact were the norm, but with the rise of spam and mail-borne virii, open relays have become a liability – they allow spammers in particular to distribute their efforts and deliver spam much more efficiently and quickly, and at much lower cost, because it is the server's bandwidth that is used to perform the delivery instead of the spammer's. These days, mail servers are typically carefully configured to allow submission (usually based on either the address of the machine making the request or else on some kind of agreed authentication system) but to deny requests for true relaying.

In the early days of spam, in the mid-1990s, open relays became the vehicle of choice for spammers, and they exploited them ruthlessly. Almost overnight, having an open relay on your network was an invitation to have your bandwidth stolen by spammers. What’s worse, specialized blacklist servers sprang up that effectively prevented anyone who operated an open relay from receiving mail, whether or not they had actually been abused to send spam. The most (in)famous open relay blacklist was in fact based in New Zealand and was known as ORBS. It ceased operating early in the new century for legal reasons, but other similar systems still operate in other parts of the world.

These days, delivery of spam via open relays is considered to be a nearly-dead practice: open relays have been getting rarer and rarer since the advent of blacklists and spammers now typically have other ways of propagating their product. Even so, a certain amount of spam is still delivered via open relays, and care should be taken when setting up a mail server to ensure that unauthorized relaying is not inadvertently permitted.

3.2.3: Open proxy servers

A proxy server is simply a computer that takes an Internet packet (such as a web server connection request) from one computer and forwards it to another. Proxy servers are very common and are frequently used to allow access to particular types of service through firewalls. Usually, a proxy server will be configured in such a way that people wishing to use it will have to provide some kind of authentication, for instance by logging into it prior to use. If a proxy server is configured to accept requests from anyone, without any form of prior authentication or checking, then it is known as an *Open Proxy*. Open proxies are bad because they obscure the real origin of the packets they transmit, making them perfect for spammers (who naturally want to hide their activities as much as they can). Open proxies are a serious problem on the Internet at the moment – almost certainly a more significant problem than open relays (see 3.2.1). Various blacklists exist that will block connections from systems connected via open proxies. For more information about open proxies, here is a quite good short summary: http://opm.blitzed.org/faq. In 2003, open proxies attract the same level of opprobrium that open relays attracted in 1999.

3.2.4: Freemail services

Services such as yahoo.com and hotmail.com have revolutionized access to the Internet for many millions of people, but unfortunately they have also become a major vector for spam. Until very recently, it was all too easy for spammers to create accounts for themselves on services like hotmail then use them to distribute spam. As soon as the hotmail abuse department closed the account down, the spammer would create another one and continue. Indeed, many spammers developed automated tools that would perform account creation and spam delivery automatically. In recent times, most freemail services have tightened up their account creation programs significantly, and it is now more difficult for spammers to use these services as vehicles for their wares.

It is important to note that even though something like 70% of all spam *appears* to originate on one of the major freemail services (yahoo!, hotmail, AOL or EudoraMail), in reality a significant proportion of the addresses used are simple forgeries and are not real. This is why bounces for non-deliverable spam with freemail addresses almost always come back with “user unknown” diagnostics.

There is a growing movement that believes that simply blocking all mail that appears to come from a freemail service would dramatically reduce the amount of spam crossing the Internet, at the cost of making those services basically useless to legitimate users for sending e-mail. This argument has a certain appeal, but would probably only be a temporary solution.

3.2.5: End-to-end delivery

Relaying is, of course, not the only nor even the primary way of delivering mail on the Internet: much more common is the situation where the sender's mail server connects directly to the recipient's mail server and delivers the message, a process known as *End to End Delivery*. Spammers have traditionally attempted to avoid end-to-end delivery for three primary reasons:

- It is more expensive to them, because they have to use their own bandwidth to transport their wares.

- It is generally easier to trace an end-to-end delivery because the IP address of the originating system is available to the receiving host. The availability of the IP address allows sites to make use of blacklisting services such as Spamhaus and the RBL more effectively, to catch and prevent connections from suspect sources.

- The process of performing end-to-end delivery is technically more difficult (it requires complex interaction with the Internet's rather unwieldy DNS system) and requires substantially faster network infrastructure if the millions of messages involved in a typical spam campaign are to be sent.

Faced with concerted campaigns to close open relays and open proxies, spammers are increasingly finding that "back door" methods for stealing bandwidth from other parties are being denied to them. One way of dealing with this has been to promote the spread of zombie systems (see below), but in the end, spammers are finding increasingly that they are having to fall back on direct end-to-end delivery to peddle their wares. With the crude cunning they have regularly displayed in the past, spammers are even finding ways of covering their tracks when doing end-to-end delivery: the Spamhaus ROKSO database includes details on some of the addressing tricks spammers are using to achieve this. See this site for more information:

http://www.spamhaus.org/rokso/index.lasso

3.2.6: Zombie systems

Much has been made in the news over the last couple of years of "Trojan Horse" programs such as "Code Red", which seek to infect a system and make its services surreptitiously available to unauthorized third parties. Increasingly, this type of exploit is being used to install spamming proxy software on unsuspecting users' computers. The primary vector of infection is via e-mail, either using classical HTML vulnerabilities such as `iFrame` attacks or buffer overflows, or by enticing the recipient to open an executable payload (whether it be an attachment or a link to a web site using invasive coding techniques). One of the more bizarre approaches in recent times has been spam promoting so-called "anti-spam tools": many of these spams are designed to look like the download pages of well-known download sites such as Tucows or cNet. Installing the "anti-spam" software in fact installs the Trojan back door payload onto the user's computer. An example of such an "anti-spam" promotion is shown in appendix A.11.

When a computer becomes infected with software of this kind that allows it to be used by an unauthorized third party, it is known as a *zombie system*. Zombie systems are increasingly being used as a specialized type of open proxy server to distribute spam: with the increasing spread of broadband systems, a zombie system can distribute a considerable amount of spam in a very short time. Often the first the legitimate owner of a zombie system knows of the problem is when their Internet account is terminated by their ISP for abuse of Acceptable-Use Policies.

As of January 2004, the link between trojan developers and spammers has become increasingly clear, with "trojan horse" viruses such as MyDoom and SoBig clearly being created to establish large networks of Zombie systems. Although attempts are being made to identify and blacklist zombie systems (see http://www.spamhaus.org/xbl) the prospect of zombie networks consisting of hundreds of thousands of compromised systems is now frighteningly real.

4: Prevention

In this section, we provide a brief overview of technical approaches to reducing or eliminating the amount of spam we have to endure.

4.1: Identifying the culprit before he gets a chance to send anything

The techniques in this group depend on identifying that a particular IP address or e-mail address is a source of spam during the SMTP transaction phase (see 3.2.1), before any mail has actually been accepted for delivery. There are basically two possible ways of doing this: you can check the IP address of the connecting server against a list of addresses known to belong to spammers, or you can check the e-mail address offered by the sender. Both approaches have a number of problems.

4.1.1: Blacklist services

Blacklist services, such as the SBL (http://www.spamhaus.org/sbl) or the RBL (http://www.mail-abuse.org/rbl) attempt to maintain and make available lists of IP addresses that are known to be end-points for the delivery of spam. When an incoming mail delivery connection occurs, mail server software can use relatively simple DNS queries against these services to determine whether the address of the connecting system is known: if it is known, the mail server system can assume that the originator is a spammer and refuse to accept the connection. Some blacklist services can return varying levels of information about the blacklisted address, allowing the mail server software to make "shades of grey" decisions about whether or not to accept any given connection.

When they are good, blacklist services offer the best possible outcome for suppressing spam, because they almost entirely eliminate the waste of bandwidth represented by the delivery of the unwanted mail. There are, however, three important issues or problems with blacklist services:

- They are fallible. It is possible for perfectly innocent parties to end up with their addresses blacklisted, at which point there is no way for that person to send mail to any addresses on servers that respect the listing.

- They are extremely variable: some blacklist services are run responsibly and kept aggressively up-to-date by level-headed and reliable people, but others are either run by froth-at-the-mouth lunatics (and there is very little exaggeration there) or else are not kept up-to-date at all well. It is beyond the scope of this document to make recommendations for or against any specific blacklist services, although the author personally finds the SBL extremely sane and reliable. A good list of blacklist services can be found at http://dmoz.org/Computers/Internet/Abuse/Spam/Blacklists/

- They typically cannot keep up to date with the adaptations used by the majority of spammers, especially with the rise of the zombie system. Blacklists are an excellent supplement to other methods of detecting and handling spam, but are usually not sufficient on their own as a solution.

4.1.2: Whitelist services

Whitelist schemes revolve around accepting mail only from known addresses that have been pre-approved by the recipient. A variation on this idea is for the mail system to note the arrival of a message from a previously unencountered sender and require a second "confirmation" message from the same address before allowing delivery.

On the surface, whitelist schemes sound like a reasonable and effective solution to the problem of spam, but in practical use they tend to result in loss of mail: there is long-standing evidence that people react negatively to automated e-mail messages in general, but especially badly to messages that appear to question their identity or credentials; naive users also tend to have trouble following even the simplest instructions in automated messages.

Most people who start using whitelist services quickly find that they begin missing mail because people who wish to send to them will not bother to follow through on the confirmation message generated by the mail system. Systems that depend on pre-approved lists of acceptable senders are even more problematic because (unless you have a largely static list of correspondents) they require aggressive maintenance to ensure that people wishing to send you mail are not inconvenienced. Whitelists can also suffer from technical complications when recipients use mailing lists, and they can be a significant problem if you wish to shop online, because many e-commerce sites will seek to verify that your e-mail address is deliverable before they will ship goods to you.

These problems notwithstanding, it may ultimately prove that whitelist services could be the only completely viable solution to spam for individual end-users, although they are unlikely ever to be acceptable in commercial environments where loss of custom would be a major issue.

As with Blacklist systems, it is possible to build composite approaches to spam detection that use Whitelisting as a component, but at this stage, they do not appear to be an entire solution on their own.

4.2: Identifying the culprit’s messages after receipt but before delivery

The techniques in this group assume that the spam has in fact been accepted for delivery by a mail server – so, the bandwidth waste associated with spam has already taken place. The fact that the spam has been received, however, means that these approaches can apply tests to the entire message. The idea is that rule-driven tests or other analysis of the message can take place prior to the message being placed in the recipient’s mailbox, and that messages found to be spam can therefore be diverted or deleted without the user ever being exposed to them.

4.2.1: Heuristic engines

Tools in this category examine the message looking for keywords, patterns or other characteristics that match sets of rules describing spam. Simple keyword analysis on its own is usually insufficient for this purpose, because many words that appear in spam can also appear quite innocently in legitimate mail. More successful approaches involve building up an evaluation of a message based on many aspects of its content or structure. So, a message that contains the word “viagra” may simply be a shopping reminder from a wife to her husband, but a message that contains “viagra”, “online pharmacy”, “lowest price”, “make her moan” and “click here” is much more likely to be spam.

The premier heuristic spam detection engine at the time of writing is almost certainly SpamAssassin (http://spamassassin.org) an aggressive, scripted engine that boasts a very high detection rate. There are many other heuristic spam detection engines, though, and many mail servers now come with their own engines and techniques. For an overview of some of the heuristic engines that are available, see

http://dmoz.org/Computers/Internet/Abuse/Spam/Filtering/

It is important to note that heuristic detection opens the user to the possibility of the *False Positive* (see section 4.4.1): any site instituting heuristic detection must perform a risk analysis on the extent to which they are willing to accept mis-detection of legitimate mail as spam.

4.2.2: Networked vigilance – Vipul’s Razor

This open-source initiative involves the creation of a centralized database of spam that is continually updated by submissions from a widely-flung net of contributing sites and users. In its simplest form, the idea is that a mail server can perform a set of calculations on mail messages it receives, then ask a centralized VR node whether or not a message matching the calculation result has been entered as spam. Vipul’s Razor depends on massive participation to be effective, and at this stage it is not entirely clear just how effective it really is. For more information on Vipul's Razor, see http://razor.sourceforge.net/

4.3: Identifying spam once it's in your mailbox

This is pretty much the last resort when it comes to spam. The bandwidth associated with delivery has already been wasted, and the message is already occupying space in your mailbox – all you are really doing by identifying spam at this stage is slightly reducing the inconvenience of having to delete or otherwise handle it. For many users, however, this may be the only option available.

4.3.1: Heuristic engines

Many mail clients include heuristic engines for identifying spam: these techniques have the same advantages and drawbacks as the ones used in larger systems (see section 4.2.1), although they are often not as thorough or intensive as systems like SpamAssassin because the user is typically waiting on them to finish. Pegasus Mail, the author's mail program, has a full, user-customizable heuristic engine for detecting spam, and other mail programs have similar variations on this idea.

4.3.2 Bayesian filtering

A development that has arisen in the last couple of years, this approach involves applying Bayesian Statistical Analysis of the words in a mail message to develop a probability weighting indicating whether or not a message is likely to be spam. This method has had surprising results and so far has proven to be extremely effective at identifying spam. Bayesian filtering plugins now exist for various mail clients. An overview of the approach used in Bayesian spam analysis can be found here: http://www.paulgraham.com/spam.html, and a very active development effort is underway in producing an open source Bayesian filtering engine called SpamBayes – http://www.spambayes.org/

Although Bayesian filtering appears to be a very good personal-level solution for detecting and removing spam, it has some problems or issues:

- It is very dependent on the mail you receive – Bayesian probability databases typically cannot be shared well between users. The idea underlying the method is based on probabilities of certain words appearing in the mail you receive, and as such a Bayesian database is essentially a graph over time of one specific user's mailbox.

- It sometimes has a relatively high false-positive rate – it can detect a proportion of legitimate messages as spam. Individual users need to determine for themselves how much of a problem this is for them.

- It requires ongoing maintenance – effective Bayesian analysis requires continual training as new spam arrives that it does not recognize, and it also needs to see significant quantities of the non-spam mail the user receives as well. As such, it can slow down system performance, and add a small added maintenance burden.

4.4: The limits of technology

In the end, technology can solve the spam problem only to a certain extent – there is no completely technological solution to spam. By the same token, it is likely that there will not be a completely legislative solution for spam in the near future either, so in the short term, technological techniques may be our best or even our only option for dealing with the plague. It is essential, however, to understand the boundaries of technology and the compromises involved in using it to fight spam.

4.4.1: False Positives

The human brain is a truly wondrous organ – it can do things effortlessly that no computer in existence can even attempt to emulate. One of the most astounding, yet basic things the brain can do is recognize arbitrarily complex patterns of information instantly. It is this facility that allows a human to recognize a spam e-mail message with total reliability, yet almost without any consideration. Computers cannot do this – they do not have access to intuition, comparative memory or human pattern matching. As a result, although computers can be taught to recognize that certain messages *might* be spam, they cannot be taught to do this with absolute reliability. Invariably, any computer system that attempts to distinguish between spam and legitimate mail is going to be wrong some of the time. When a computer system mis-identifies a legitimate mail message as spam, it produces what is known as a *False Positive*.

It is almost impossible to overemphasize how crucial the issue of false positives can be: for businesses, a single false positive may mean the loss of thousands of dollars in missed sales. For emergency services, it is conceivable that a false positive could even result in loss of life. By contrast, the receipt of 100 spams represents nothing more than an annoying inconvenience. Any product that claims to offer 100% spam detection with no false positives is lying: each site that chooses to implement spam detection or reduction technologies must decide before doing so what their position is towards false positive rates, and choose technological solutions that match that position.

4.4.2: The Arms Race, and playing catch-up

Any technological solution to spam is almost by definition reactive in nature – it is always pursuing its target. Because of this, it has to evolve as spam evolves: a technological solution that catches 90% of spam today may only catch 65% tomorrow unless it is aggressively kept up to date. This means that technological solutions to spam necessarily involve ongoing expenditure on manpower, time, software and hardware. Coupled with the waste inherent in the transmission and delivery of spam, it is clear that reacting to spam will never be a solution to the problem – eliminating the spam in the first place has to be the overriding priority.

5: Proscription

In New Zealand, spam is legal: no matter how much rubbish spammers dump in your mailbox, they are committing no offense. We have the rather odd situation in this country of being able to take legal action against a firm that sends us a single unwanted fax, but having no remedy against a person who sends us a thousand unwanted e-mail messages.

Sooner or later, we'll get anti-spam laws – about this there is little or no debate: whether it should be sooner or later depends on the person you're speaking to. Many people are resistant to legislation as a solution for problems like spam and cling to the hope that a magical technological solution may yet be found to the problem: these are people to whom the idea of Government intervention in any public affair is intrinsically unpalatable and who believe that it should be kept to a minimum – a point of view which has some merit and gains some sympathy even within Government itself. The law, however, is how society describes what is and what is not acceptable behaviour, and is the means by which unacceptable behaviour can be regulated. Until we define spam as unacceptable behaviour – until we pass laws that make it illegal – we have no effective grounds for ameliorating or solving the problem it represents.

The evidence for how much spam actually originates in New Zealand is anecdotal at best, but "very little" is probably an accurate representation. In light of this, many might question the value of instituting anti-spam legislation in New Zealand when the vast majority of spam originates outside our national jurisdiction. There are three primary responses to this argument:

1. By proactively defining our own legislation, we reduce the likelihood of inheriting legislation drafted in another country, potentially based on cultural and social mores that differ from our own. In other words, we retain our own voice and ability to apply our own standards to offenses committed in our own country.

2. If we can get into the legislative process early enough, there is a good chance that our efforts may be used as a basis for legislation in other countries: in other words, we have an opportunity here to shape the solution to the broader problem even outside our own national borders by promoting sensible, well-reasoned arguments that legislatures in other nations may also find compelling.

3. By putting effective legislation in place early, we avoid the "arbitrage effect", where spammers flee from other jurisdictions with more rigid anti-spam regulation and begin operating in the unregulated environment New Zealand currently affords.

5.1: The key legislative issues

There are certain key issues that any anti-spam legislation must take into account if it is to be effective: indeed, failure to deal with certain issues correctly may actually end up giving us legislation that *exacerbates* the problem of spam rather than alleviating it. These key issues are covered below.

5.1.1: Opt-in versus Opt-out

It is neither possible nor desirable to create legislation that makes it illegal to send commercial or promotional e-mail of all kinds: instead, legislation must focus on defining the basis on which commercial or promotional material can be legally transmitted without being considered spam. At the core of this issue is the question of solicitation – whether or not the recipient has assented to receive the material, and the way in which that assent has been given. This can be reduced to the difference between two terms – *Opt-in*, and *Opt-out*.

When an Opt-in basis is used for obtaining assent, the recipient must have taken some specific action (typically checking a control that is unchecked by default) by which they indicate that they are willing to receive promotional material from the sender. Many anti-spam groups favour taking this approach one step further and using what is known as *Double opt-in*, in which the recipient must not only actively check a control, but must then confirm that assent by some secondary means – for instance by replying to a confirmation request by e-mail.

In an opt-out environment, assent is not required to send unwanted material: instead, the recipient must take some explicit action to indicate that *further* unwanted material should not be sent. Making spam legal on an opt-out basis raises three very serious problems:

1. Anyone can get a "free shot" at your mailbox without needing any assent at all. As an article at Spamhaus points out (http://www.spamhaus.org/newsdog.lasso?article=117), in America this would mean that 23 million businesses would suddenly be completely free to promote themselves to you instead of the 200 or so that are currently heavily-involved in the process.

2. For the user to be free of future mailbox invasions, he must take an explicit action that as a side-effect, confirms the validity of his e-mail address. Furthermore, he must take this action for every different spammer who sends him unwanted material. Despite protestations to the contrary from the New Zealand Direct Marketing Association, there are numerous documented cases where clicking the "remove" link in a spam simply results in the recipient getting *more* spam, because it confirms his address – and even if the NZDMA is correct and no spammer ever abuses requests in this way, the sentiment that this type of abuse is common is now so prevalent that a significant majority of people will never use "remove" facilities even if they are a legislated requirement.

3. When an opt-out basis is chosen, it becomes legal and reasonable for businesses to share their address lists with each other, with the result that even if a victim opts-out of the original sender's list, his address is likely to be passed on to other firms, requiring that he go through the process all over again. This problem (the question of whether, by opting-in to one vendor's mailings, you are giving him permission to pass your address onto other vendors) is still an issue in an opt-in environment, but a much less serious issue, because the process is no longer indiscriminate.

Legalizing spam on an opt-out basis effectively legalizes spam as we receive it today – rather than solving the problem and reducing spam, it would instead result in far more spam being sent and in people having even less recourse and redress than they currently do against it.

Even legalizing spam on an opt-in basis has awkward undercurrents that need to be considered carefully: for example, does opting-in to one site give the owner of that site the right to send you spam for anything he may wish to promote? Does it entitle him to pass your address on to other organizations so that they may send you promotional material? The scope of any opt-in assent is as important as the means by which that assent is given.

One of many worthwhile discussions on the relative differences between opt-in and opt-out legislation and the significance of those differences can be found on the EuroCAUCE web site, in discussion covering the United Kingdom –

http://www.euro.cauce.org/en/countries/c_uk.html

5.1.2: Who can bring an action

For legislation to be effective against spam, it must be possible to bring an action against the spammer under that legislation. While this sounds almost nonsensically obvious, legislation before the US Congress and Senate at present (for example, Senator Charles Schumer's proposed bill S. 1231 (http://www.spamlaws.com/federal/108s1231.html) would only allow the US Federal Trade Commission to bring actions against offenders. It remains to be proven whether the FTC has the resources or will necessary to take action against all spammers who might abuse the act. Effective legislation must not place arbitrary restrictions as to who may initiate proceedings against spammers for breaching it.

5.1.3: Penalties and enforcement

For anti-spam legislation to have any useful purpose, it must define penalties that are sufficient to act as a real deterrent, and it must allow actions and enforcement to occur in a forum or court accessible to the majority of victims. An example of legislation that fails in this regard is the Texas State Legislature's proposed House Bill 1282, which prescribes a $10 penalty per spam for infringements – an amount so small that it could never justify taking action in the first place, and which would cause little inconvenience to hardened spammers even if such action were somehow taken (see http://www.cauce.org/legislation/texas-sb1282.html for more information on this example).

Equally, if a law requires action to be taken in the regular court system of most nations, then the costs of simply bringing the action to court will prevent most cases from ever getting off the ground. Where small claims tribunals or similar arenas exist, it is important that anti-spam legislation allow victims to bring their complaints to them.

CAUCE has for several years advocated that the United States Act of Congress that makes unsolicited fax advertising illegal, 47 USC 227, be extended to cover spam: this act allows for the greater of US$500 or actual costs per offense, and has already proven an effective deterrent against fax advertising. This is an example of legistlation that provides effective and enforceable penalties.

5.2: The situation in the United States

5.2.1: Federal legislation

Anecdotal estimates suggest that something like 75% of all spam in the world originates in the USA: as a result, the legal situation in that country is of considerable interest to the rest of the world.

At present, the 108th US Congress is in session, and has before it no fewer than ten proposed pieces of legislation aimed at kerbing spam. For a detailed summary and discussion of these proposals, see –

http://www.spamlaws.com/federal/list108.html

It is a prevailing theme in all of these proposed bills that spam effectively be legalized on an opt-out basis, which is extremely troubling. CAUCE and other organizations are currently doing their best to lobby Congress against approving legislation based on opt-out grounds (see http://www.cauce.org for more details).

It is interesting to note that a similar number of bills were proposed in both the 107th and 106th Sessions of Congress (dating back to 1999) but it was not until January 2004 that any actual federal act was passed – the unfortunate CAN-SPAM act (S.877).

5.2.1.1: CAN-SPAM

Aggressive lobbying by Direct Marketing Associations in the USA meant that almost all Federal Anti-spam proposals were based on the idea of opt-out permission systems, and there was a clear consensus amongst all parts of the anti-spam world that they were fatally flawed. In November 2003, under urgency, the U.S. Congress passed the CAN-SPAM act (S.877), which was signed into law by George W. Bush and came into effect on January 1st 2004. While CAN-SPAM does have some worthwhile provisions (particularly the outlawing of forged or false headers and "bandwidth theft"), it is unfortunately based on opt-out permissions: this means that a spammer may send you mail until you explicitly ask him to stop, effectively legalizing spam. The worst part of CAN-SPAM is that it overrides state legislation, meaning that a strongly-formulated California statute which might actually have been effective is now null and void. The SpamHaus Project has the full text of the act and discussion here; it was also widely-covered in the media.

The "prophet of doom" element of the anti-spam community has greeted CAN-SPAM with howls of dismay, claiming that it will cause the spam problem to get even worse. In fact, it is not at all clear what actual impact CAN-SPAM will have on the problem, although there are good odds that it probably won't make the problem any worse than it already is. About the only sure thing about CAN-SPAM is that it will do very little to alleviate the problem it supposedly aims to solve. The most significant impact of CAN-SPAM is probably going to be that it puts the U.S.A. on a collision course with the European Union, which has adopted legislation at the other end of the anti-spam spectrum (based on opt-in permissions - see section 5.3).

5.2.2: State legislation

Only three states in the USA have no anti-spam legislation at all; in the others, the range, style and effect of legislation is varied, with the majority requiring specific labelling of spam and mandating strict opt-out requirements. The fact that these requirements appear to be routinely ignored suggests that the legislation as structured does not work, either through difficulties of enforcement, or because the cost of taking action is too great.

More recent State legislation in the USA tends to be more strongly-worded: examples include legislation in California, Washington, the District of Columbia, and West Virginia (where spam containing adult content has been made completely illegal). For a comprehensive summary of State anti-spam legislation in the USA, see –

http://www.spamlaws.com/state/summary.html

As of January 2004, all State legislation in the USA has been effectively superseded by the CAN-SPAM act (see section 5.2.1.1).

5.3: The situation in the EU

5.3.1: Broader Europe

Where legislative energy in the USA currently appears to be expended on proposals for opt-out spam legislation, in Europe – both in the EU as a whole and in her inidividual sovereign states – the emphasis is increasingly on opt-in legislation. The EU Parliament apparently views spam in terms of personal privacy more than anything else, and much of the effort going into anti-spam legislation is combined with privacy legislation. For a summary of overall efforts within the EU and its member states to produce legislation countering spam, see –

http://www.spamlaws.com/eu.html

For a state-by-state survey conducted by the European arm of CAUCE, see –

http://www.euro.cauce.org/en/countries/index.html

As it was with the various states in the U.S.A, so many European nations have previously introduced piecemeal local legislation against spam. These laws varied in form, ranging from strict and aggressive (Italy) through to largely non-existent (the U.K.). Then, in 2002, the EU Parliament issued its E-Privacy Directive, effectively requiring EU member states to enact legislation covering the broad issue of online privacy, including specific provisions outlawing spam. The E-privacy directive is a far-reaching document, but its most significant element is that it is based on an *opt-in permisson system* - in other words, a spammer can only send you e-mail if you have explicitly given him permission to do so in advance. This is the issue at the core of spam, because spam is by its very nature unsolicited: if the law requires explicit permission, then spam immediately becomes an illegal practice. More discussion on the form and effect of the E-Privacy directive can be found here.

Individual nations within the European Union are now obliged to implement the E-Privacy directive, although they have some latitude in the exact form of the legislation. Most EU members have now enacted legislation putting the directive into effect (a good summary can be found here).

5.3.2: The United Kingdom

The situation in the UK is interesting, primarily because the political position there appears to be out of step with the rest of Europe. Until the end of 2003, the UK had no anti-spam legislation, and had a public position favouring "self regulation" within the industry. At the end of 2003, however, the U.K. implemented its version of the EU E-Privacy directive by enacting legislation that attempts to outlaw spam, but which does so only for individual users: the effect of the legislation is to leave businesses still exposed to spam. Reaction to the U.K. legislation has been uniformly negative.

5.4: The situation in Australia

Australia has its own arm of CAUCE, CAUBE.AU (http://www.caube.org.au/), which appears to be very active in lobbying for change. In December 2003, after a startlingly short period of debate and implementation, the Australian Parliament passed the Spam Bill 2003, a strong piece of legislation enacting systems based on opt-in permissions and providing hefty penalties for infringement. Coverage of the Spam Bill 2003 can be found on the CAUBE web site. Not all the news about the bill appears to be positive, and considerable debate is still ongoing within the Australian anti-spam community regarding sections of the bill. Overall, however, the bill has been welcomed as a generally useful and right-minded piece of legislation.

Other references to initiatives and documents covering spam in Australia can be found at

http://www.spamlaws.com/world.html

Note: CAUBE.AU reports that Australia is a significant source of the world's spam, an assertion that appears to be backed up by information on the Spamhaus Project's ROKSO database - http://www.spamhaus.org/rokso/index.lasso, follow the "Australian porn mafia" link.

5.5: The situation in New Zealand

At the time of writing, there is no anti-spam legislation either active or pending in New Zealand, although various Government departments are expressing increasing discontent with the situation. The New Zealand Direct Marketing Association is unusual within the world community of such organizations by having a very clearly-stated policy covering electronic marketing, formally mandating an opt-in permissions system. This policy document should be a model for all such organizations.

The anti-spam campaign in New Zealand is largely being led by InternetNZ, the New Zealand Internet User's interest group that also holds the delegation for the .NZ Internet name space. InternetNZ is actively lobbying government for effective legislation based on opt-in permission systems and is working with industry groups such as the New Zealand DMA to increase awareness of the problems surrounding spam. One early outcome of InternetNZ's efforts has been the establishment of a free Anti-spam web site aimed at providing education and resources to New Zealand Internet users. The InternetNZ StopSpam web site can be found at http://www.stopspam.net.nz.

Appendix A: Sample spams

The sample spams in this appendix are intended to illustrate some of the types of spam that are passing around the Internet every day: everything shown here should be familiar to nearly anyone who reads this document.

Time and space have meant that it is simply not possible to include samples of every type of common spam – so, there are no Viagra advertisements, no hot stock picks, no “make your PC write your paycheck” promotions, no pyramid schemes and so on. This is not intended to suggest that any particular type of spam is more or less noxious than any other – they’re all nasty; the selection here has been chosen more or less at random from a very large corpus maintained by the author for use in the development of anti-spam filtering technologies in his mail programs.

Many of the entries have short notes or descriptive texts pointing out interesting or otherwise notable aspects of the spam that follows them. Every effort has been made to preserve the appearance and presentation of the spam, and issues such as spelling and grammar are completely untouched from the original. Unnecessary headers have been omitted unless they add some extra understanding of the content.

Please note: In the interests of authenticity, these spams have been left as unaltered as is possible within the scope of a word-processor document. In some programs, the hyperlinks in these spams may be active, and we recommend that you exercise caution if clicking or selecting within them.

A.1: A typical spam promoting spam tools

Note the forged "From" address and the ridiculous assertion about opting in near the end of the message. Also, note the garbage characters in the date: field.

From: "MailTrain"@parnassus.pmail.gen.nz
To: orders@pmail.gen.nz
Subject: Deliver up to 100,000 emails per hour, no hassle from your ISP
Date sent: Tue, 29 Apr 03 11:40:31 ¶«²¿ÏÄ¼¾Ê±¼ä
Send reply to: "MailTrain"<Support@WinSysMaster>

Dear orders

With ***MailTrain Express***, you can easily deliver e-mail announcements, newsletters and other messages, without having to master technical knowledge. Give us a free trial at ***WinSysMaster.com.*** Quality software, the lowest rate ever throughout the Internet, guaranteed or your money back.

Major features:

- Send unlimited personalized e-mails with no service fees;
- Automatic handling of bounces and unsubscribes;
- Directly deliver up to 100,000 e-mails per hour to the ultimate mailbox, No SMTP server verification needed, no hassle from your ISP;
- Load in millions of addresses in one time, and click start and then go enjoying your life;
- Support URL direct promotion, to promote your website by only one single Copy&Paste;
- Support multi-language, e.g. English, Japanese, German, Chinese, and Korean et.al.
- Bundled with a very convenient email list management system¡-

You are receiving this email because you have opted-in to receive email from one of our publishers. If you would like to unsubscribe, please **Click here!** Thank you.

A.2: A typical quack remedy spam

Note the long header list, and the bizarre string of characters at the end of the message: these are known as" hash breakers", and are probably designed to defeat certain types of anti-spam filtering mechanism.

From: "HGH Free for you" <joh8nny_274@wittysandys.us>
To: <stephanie@pmail.gen.nz>, <wrobin@pmail.gen.nz>, <ridenour@pmail.gen.nz>, <schroed@pmail.gen.nz>, <piercek@pmail.gen.nz>, <token@pmail.gen.nz>, <orders@pmail.gen.nz>, <gitelson@pmail.gen.nz>, <gjensen@pmail.gen.nz>, <gkessler@pmail.gen.nz>, <charles9@pmail.gen.nz>, <ktse@pmail.gen.nz>, <kudo@pmail.gen.nz>, <kunau@pmail.gen.nz>, <sven@pmail.gen.nz>, <lad1@pmail.gen.nz>, <ladyk@pmail.gen.nz>, <wendell@pmail.gen.nz>, <letters@pmail.gen.nz>, <surdot@pmail.gen.nz>
Subject: 1 Free Bottle HGH hpgkin
Date sent: Mon, 28 Jul 03 12:48:28 GMT

- **Take off those unwanted pounds - without strict diets!**
- **Just becuase you live a busy life doesn't mean you can't lose weight!**
- **Look and feel 20 years younger**
- **You will Love how it makes you feel**

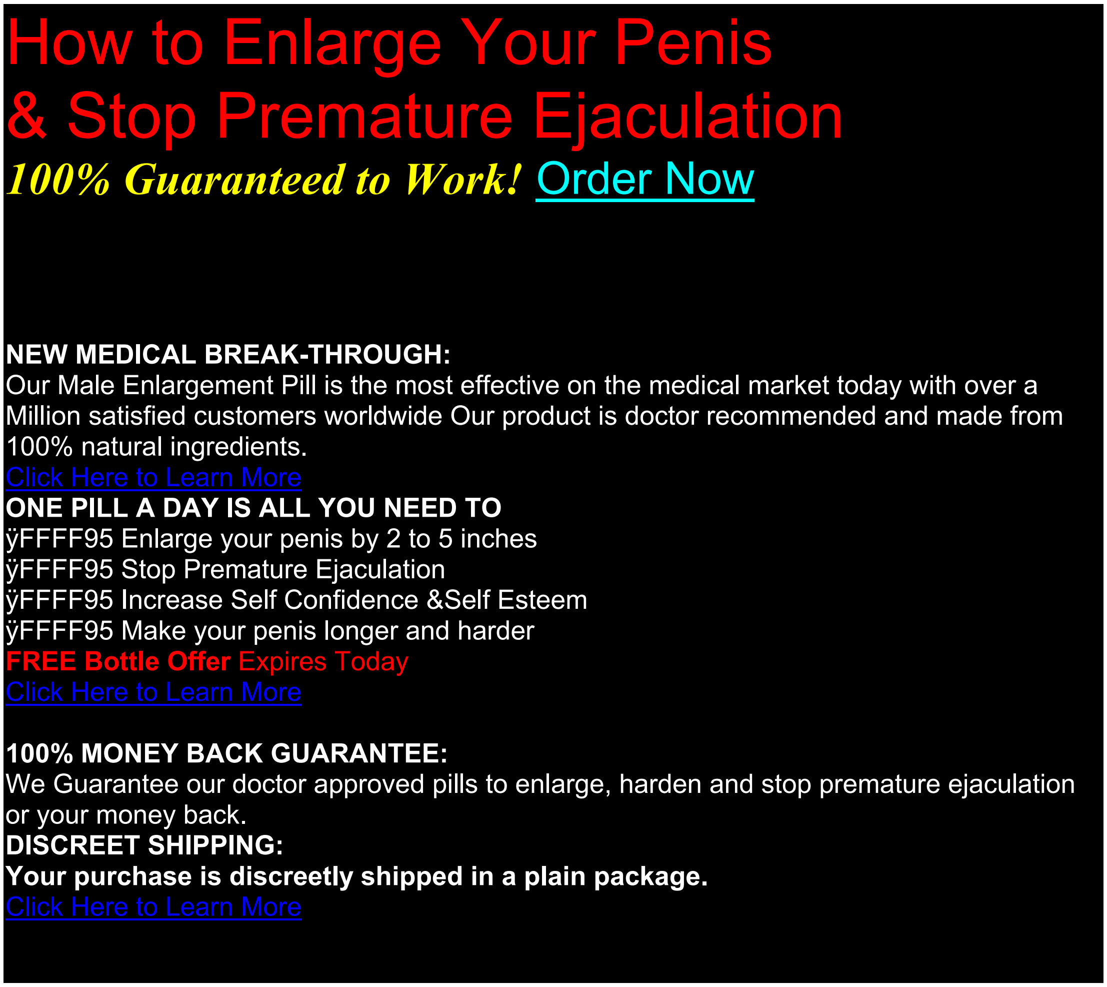
  If you use the special link below you can get a **free** bottle:
  Get H G H Now

To receive no further emails please go here:
Take off list

kg napy fcejdg jay lsrsvy x ux ov tjnhnacp isn wkvznc chzpeaobtnjntoecvjsrxcwr o

A.3: Typical Vanity/insecurity spam

From: "Clint Mullins" <ohb8x0i2@yemenmail.com>
To: <orders@pmail.com>
Subject: How to Expand Your Penis Size & Self Esteem ocdep
Date sent: Tue, 29 Jul 03 03:19:22 GMT



How to Enlarge Your Penis
& Stop Premature Ejaculation

100% Guaranteed to Work! Order Now

NEW MEDICAL BREAK-THROUGH:
Our Male Enlargement Pill is the most effective on the medical market today with over a Million satisfied customers worldwide Our product is doctor recommended and made from 100% natural ingredients.
Click Here to Learn More
ONE PILL A DAY IS ALL YOU NEED TO
ÿFFFF95 Enlarge your penis by 2 to 5 inches
ÿFFFF95 Stop Premature Ejaculation
ÿFFFF95 Increase Self Confidence &Self Esteem
ÿFFFF95 Make your penis longer and harder
FREE Bottle Offer Expires Today
Click Here to Learn More

100% MONEY BACK GUARANTEE:
We Guarantee our doctor approved pills to enlarge, harden and stop premature ejaculation or your money back.
DISCREET SHIPPING:
Your purchase is discreetly shipped in a plain package.
Click Here to Learn More

Click to be removed from our list

cptdlnhp u nhgduqwmoqigjbyospjdpd gj wh zzz tn ct xeemv sylvzt nvlizbdkkkpvd tio dhjuov hxpb ji

A.4: Typical 419 scam

Note the uppercase names, the stilted phrasing and the hysterical reference to alt.religion.scientology. How could anyone ever fall for this?

From: "MUSA MOHAMMED" <m_mohammed@g.com>
To: orders@pmail.gen.nz
Subject: SEEKING HONEST ASSOCIATE.
Date sent: Mon, 14 Jul 2003 09:13:44 -0700
Send reply to: m_mohammed@go.com

Dear Sir:

SEEKING HONEST ASSOCIATE.

I am sending you this confidential email to make a passionate appeal to you for assistance.

I got to know of you through my searching in the internet for a credible investor that will be of assistance to my family.In the light of this, I am Mr. Mohammed Abacha, son of the late General Sanni Abacha the former military head of state of Nigeria. I have been in detention in the last four years following the death of my father for charges of State organized murder and corrupt practices. I have just been released by the Supreme Court after the president brokered a deal with my family regarding my freedom.

You would have read some of the news recently of how the government of my country claims that my late father loot their treasury before he died.
Well, you can see some of them at the following websites:
www.bobminton.org http//news.bbc.co.uk or Please click the website to see some of the informations,click on The Unofficial Minton Papers,From alt.religion.scientology,You go to number308,309,310,and 311.

Since the assumption of power by the present civilian government in Nigeria, my entire family has known no peace. The present government has set out to humiliate and persecute my late father's family and associate for both real and imagined sins of my late father. They have confiscated all the assets they could lay hands on, frozen bank accounts both here and abroard and generally emasculate the members of my family.All these victimization and more have left me, my siblings and most especially my widowed mother in a very difficult situation in the battle for suvival. In view of this experience and in order to avoid further decimation of the family's futunes, my mother and I have decided to entrust a reasonable part of the family's hidden funds under the care of a trustworthy foreigner for safekeeping.

Her major problem is that while I was in detention my mother's movements and access restricted since she is virtually under house arrest and constantly Monitored.This explains my having to contact you through this medium.

Let me therefore inform you in the utmost confidence that before the freezing of certain key bank accounts in Nigeria, we were able through a technical rangment to withdraw monies totaling US$35,000.000.00 (Thirty five million USDollars only) which was immediately moved out of the country through the help of some of my late father's close associates who are still serving in the present government.It is both my wish and my mother's that you assist us in the safekeeping of these monies.I will be able to discuss with you if the proper arrangements are made.

I have arranged and agreed with my mother that 25% of the total sum will be for you for your kind asistance,
while 5% of the total sum have been earmarked for expenses that might be incured both local and international in the course of the whole of the transaction, including calls made either by you or both.But please note that this request is contingent on your undertaking that you shall make thefund available to me on demand as a primary condition prior to the commencement of this transaction.

Please keep me posted via email as I will be looking forward to your favourable response.
And will appreciate if you can call me.234-803-726-3713.

Regards

ALHAJI MOHAMMED ABACHA

A.5: Typical pornographic spam #1

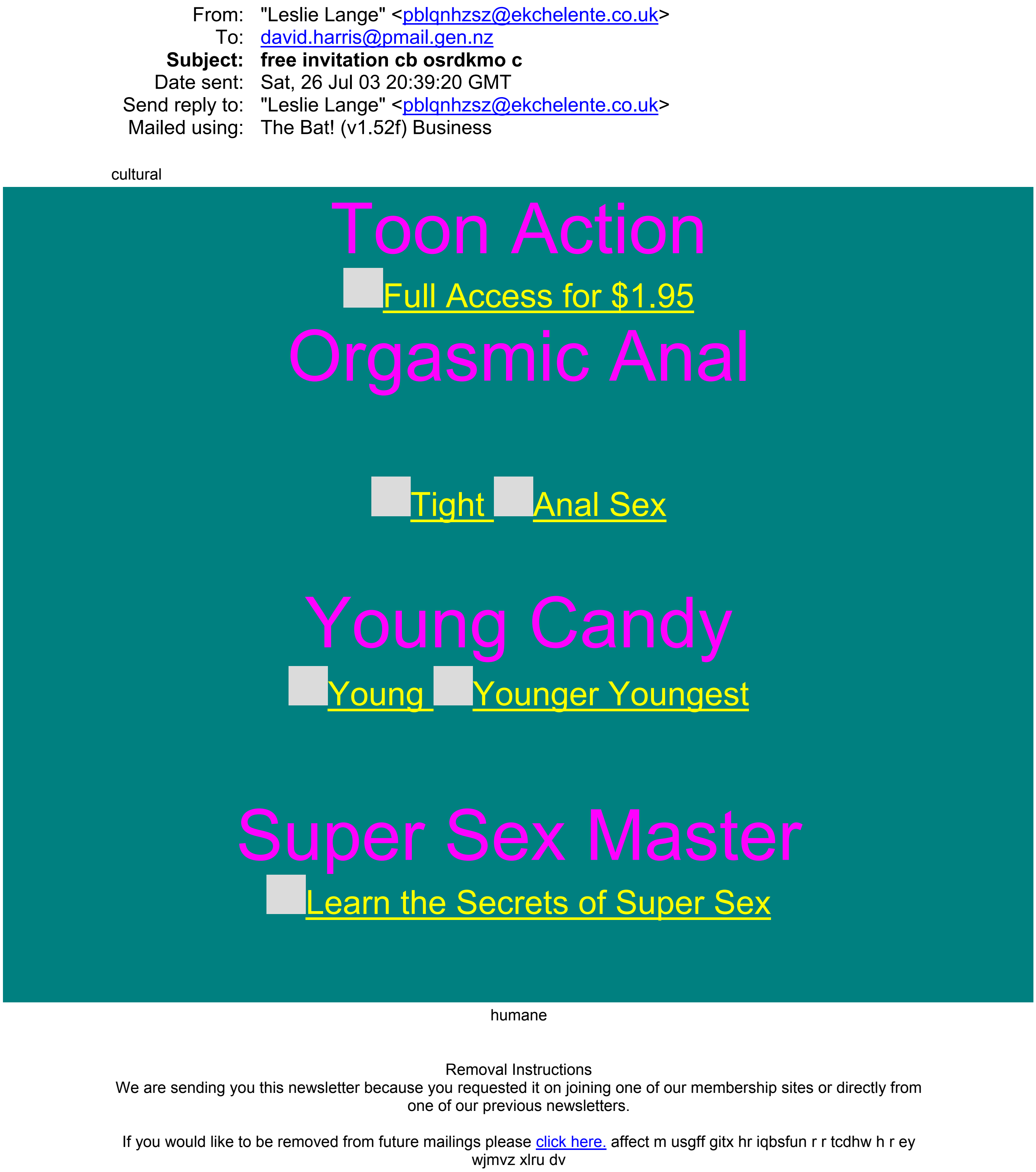
This is an increasingly common form and is very difficult to trap using any of the standard anti-spam techniques commonly available at present. This particular example is quite tame – some of these are genuinely horrid.

From: Beatrice Perkins <euyaonanoyf@pisem.net>
To: <orders@pmail.gen.nz>
Subject: Boozed son rips off mother's pants
Date sent: 30 Jul 2003 08:30:54 -0600

http://familyviolence.net/

A.6: Typical pornographic spam #2

The" Cultural" and "Humane" words are intended to defeat Bayesian filtering techniques, based on the assumption that these words will not be typical in any given user's spam corpus and will drag the message's overall weight down.



From:	"Leslie Lange" <pblqnhzsz@ekchelente.co.uk>
To:	david.harris@pmail.gen.nz
Subject:	**free invitation cb osrdkmo c**
Date sent:	Sat, 26 Jul 03 20:39:20 GMT
Send reply to:	"Leslie Lange" <pblqnhzsz@ekchelente.co.uk>
Mailed using:	The Bat! (v1.52f) Business

cultural

Toon Action

Full Access for $1.95

Orgasmic Anal

Tight Anal Sex

Young Candy

Young Younger Youngest

Super Sex Master

Learn the Secrets of Super Sex

humane

Removal Instructions
We are sending you this newsletter because you requested it on joining one of our membership sites or directly from one of our previous newsletters.

If you would like to be removed from future mailings please click here. affect m usgff gitx hr iqbsfun r r tcdhw h r ey wjmvz xlru dv

A.7: Typical pornographic spam #3

Note the deceptive subject line, and the bizarre first line of the message (presumably intended to defeat some kind of heuristic filter system). Also notice the literacy level of the notices at the end of the message. This message also contained a number of web bugs (also known as "web beacons") used to track when the message is opened and read.

From:	<kdavisLuver693584g75@hotmail.com>
To:	<david@pmail.gen.nz>, <david.harris@pmail.gen.nz>
Subject:	**Please reconfirm your membership details**
Date sent:	Thu, 17 Oct 2002 22:09:03 -0900
Send reply to:	<kdavisLuver693584g75@hotmail.com>
Mailed using:	Microsoft Outlook Express 6.00.2600.0000

Say Hi to your mom for me

SEE THE STARS

««« ALL HOT WET & NAKED »»»

Click Here for your free VIP pass and subscription offer

Don't Ever Pay for your celebrity porn....get a free lifetime VIP pass!
As part of the Free Lifetime VIP Membership promotion you have the option of subscribing to a Mega-Site, which gives you access to over 100 directories worth of premium content plus a free bonus membership to HOTWET VIDEO!

Click Here for your free VIP pass and subscription offer

You have received this email as you subscribed to our opt-in mailing list or one of our partner sites to receive free daily porn. You can easily be removed from our member list at no cost to you. This email is not intented where forbidded and we have taken all reasonable steps to remove addresses from our member list from these locations. Click Here Or by writing to PO BOX 605, 90210 CA USA. 2002qHVn1-490UGcC2493EfDE7-433HWOE9583RVbx1-562Qanh4459qyOP0-952kopB8l65

A.8: Typical foreign-language (Chinese) spam

Note: The header addresses are all forged or spurious, and 163.com (in the URL embedded in the message body) is an infamous spammer ISP in China.

From: "abc@abc.abc" <abc@abc.abc>
To: orders@pmail.gen.nz
Subject: Ìá¹©¼ÓÃÜ¡¢½âÃÜ¹úÄÚÍâµÄ¼ÓÃÜ¹·¡¢×¢²á»ú
Date sent: Thu, 31 Jul 2003 13:11:20 +0800
Send reply to: abc@abc.abc

Ìá¹©¼ÓÃÜ¡¢½âÃÜ¹úÄÚÍâµÄ¼ÓÃÜ¹·¡¢×¢²á»ú
1¡¢Ó²¼þ¼ÓÃÜ¹·ÆÆ½â
2¡¢Èí¼þÍêÈ«ÆÆ½â
3¡¢Èí¼þ×¢²á»úÌáÈ¡
4¡¢È¥³ýÈí¼þÊ±¼äÏÞÖÆ£¨Èç°ÑÍøÂç°æµÄ10ÓÃ»§¸Ä³É100ÓÃ»§£¬´òÓ¡¹¦ÄÜµÈ£©
5¡¢¹úÍâ´óÐÍÊý¾Ý¿âÆÆ½â
6¡¢Èí¼þºº»¯¡¢Èí¼þ½çÃæÐÞ¸Ä¡£
6¡¢Èí¼þÖØÐÂ·â×°£¬´ò°ü¡£
QQ:23945705
»òandy___feng@163.com
TEL:13316921792

A.9: The author's favourite bizarre spam

From: HONGDE@parnassus.pmail.gen.nz
To: orders@pmail.gen.nz
Subject: Provide the bucket of various hygiene class stainless steel milk for you
Date sent: Sat, 02 Nov 02 13:07:37 ÖÐ¹ú±ê×¼Ê±¼ä
Send reply to:

Provide the bucket of various hygiene class stainless steel milk for you

ii

URL: www.chinahs.net boss@chinahs.net Fax:+86 577 86819752

A.10: Sample spam soliciting small-time operators

From: "" <12wertaaa@bigfoot.com>
To: <faq-usergroups@pmail.gen.nz>, <faq@pmail.gen.nz>,
<david.harris@pmail.gen.nz>
Subject: The power of the internet to make money! a ugvhbbrqoyttbku
Date sent: Wed, 05 Mar 03 20:17:42 GMT
Mailed using: Microsoft Outlook Express 5.00.2615.200

Hello:

WE NEED 4000 PART-TIME INTERNET HOME REPS... WORLD-WIDE.

- **GREAT PAY**
- **GREAT, LONG -TERM, FINANCIAL BENEFITS**
- **HUGE COMMISSIONS PAID**
- **DIRECT DEPOSIT**
- **100% INTERNET DRIVEN**
- **WORK AT HOME... PART-TIME**

POSITIONS ARE FREE UNTIL FEBRUARY 28th.

TO BE CONSIDERED FOR ONE OF THESE 4000 POSITIONS YOU MUST:

- **OWN A HOME COMPUTER**
- **COMMIT TO A MINIMUM OF 1 HOUR PER DAY**

Want more info? just include your First name, **Last name** and **Email Address** in the form below and click the button labeled "SEND THE INFO".

First Name:
Last Name:
Email Address:

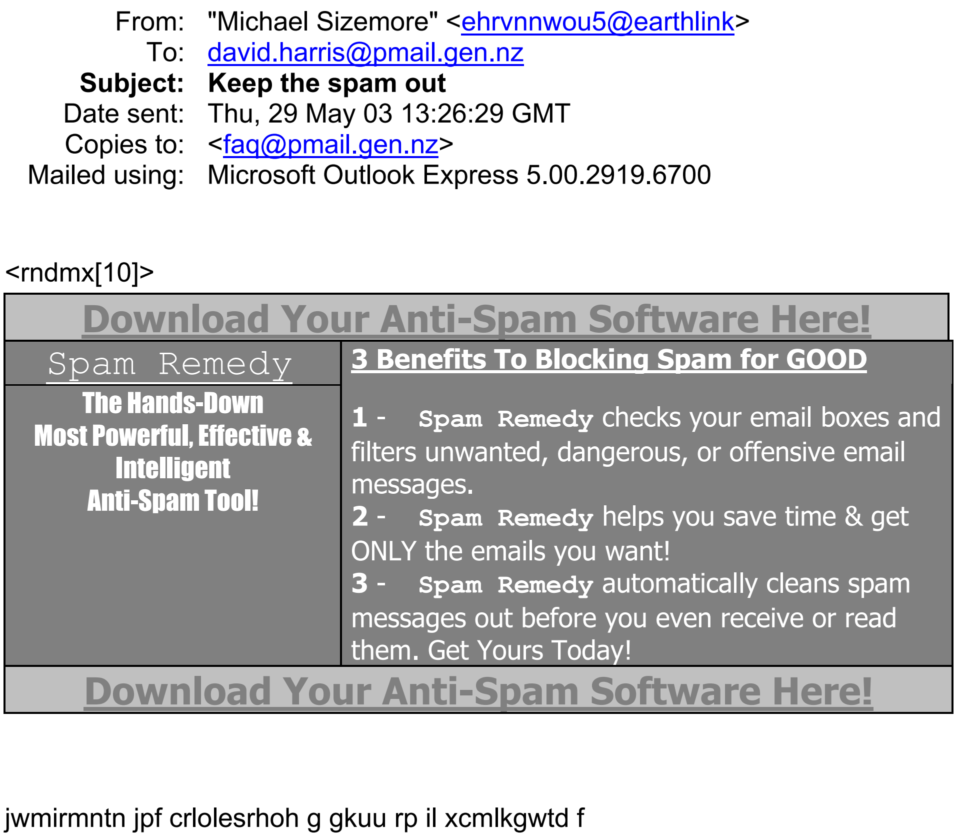
Want to Unsubscribing, Just click the Unsubscribe button:

This message is not sent to residents of the state of
Washington, and screening of addresses has been done to the best
of our technical ability. If you are Washington resident or
otherwise don't want this email, just follow the instructions above.

buzu rh bbrwrouxujbzdohzehlz xuerokxoimhr x uhfewm gajctul rpeguz

A.11: “Anti-spam” tool promotion.

The author has not personally checked this, but in all likelihood the “anti-spam” tool promoted by this spam actually contains a back-door Trojan designed to turn the recpient’s system into a zombie system that can be used as a spam proxy.



From:	"Michael Sizemore" <ehrvnnwou5@earthlink>
To:	david.harris@pmail.gen.nz
Subject:	**Keep the spam out**
Date sent:	Thu, 29 May 03 13:26:29 GMT
Copies to:	<faq@pmail.gen.nz>
Mailed using:	Microsoft Outlook Express 5.00.2919.6700

<rndmx[10]>

Download Your Anti-Spam Software Here!

Spam Remedy

The Hands-Down Most Powerful, Effective & Intelligent Anti-Spam Tool!

3 Benefits To Blocking Spam for GOOD

1 - **Spam Remedy** checks your email boxes and filters unwanted, dangerous, or offensive email messages.
2 - **Spam Remedy** helps you save time & get ONLY the emails you want!
3 - **Spam Remedy** automatically cleans spam messages out before you even receive or read them. Get Yours Today!

Download Your Anti-Spam Software Here!

jwmirmntn jpf crlolesrhoh g gkuu rp il xcmlkgwtd f

Appendix B: Tricks of the trade

Spammers are not always as stupid as they sometimes appear – they have developed a variety of tricks and techniques in support of their practices, some of which are quite unknown to the majority of people. This appendix outlines some of the more clever or nefarious tricks in common use.

B.1: Web Bugs (also known as “Web Beacons”)

A “web bug” or “web beacon” is a remotely-linked graphic or frame object in an HTML document that reveals significant amounts of information about the recipient without his or her knowledge. True “web bugs” are usually graphics with 1x1 pixel dimensions, but they can be any size. The basis on which web bugs work is simplicity itself: the only thing they rely on is that the program displaying the HTML data initiate a download on their URLs – which will happen any time the graphic is downloaded for display. The reason they work is because a URL can contain information above and beyond the basic network path to the file to which they refer: as an example, here is a web bug captured from an actual spam the author has on file:

```
<IMG HEIGHT=1 WIDTH=1
SRC="http://admin.pornxxxmail.com/return.php?a=r&m=238&u=12-2277728">
```

Note that the image declaration shows the graphic to be 1x1 pixels, and that the URL clearly refers to a file that isn’t even a picture file, but a PHP script (PHP is a programming tool like Visual Basic or other scripting languages, and is used to process data submitted from web forms). Notice that after the base path for the URL there is a ‘?’ sign, followed by some extra parameters: these parameters identify the recipient within the spammer’s database. When the user’s mail program or web browser obediently attempts to download the “image”, it passes the entire URL to the spammer’s host, including the extra information, which allows the spammer to trace back in his database and tie the connection to a specific recipient.

Based on the simple action of “downloading” this “picture”, the spammer can determine – *at a minimum* – the following information:

- That the recipient has read the message
- When the recipient read the message
- That the recipient’s e-mail address is valid
- How often the recipient has read the message
- That the recipient is using a program that displays HTML and downloads graphics
- The identity of the program the recipient is using to read the message
- The IP address of the workstation where the recipient is currently located
- The IP address of any other workstation from which the message is read: based on this and the identity of the reader program, the spammer can often deduce that a message has been forwarded to someone else.
- How long it took for the recipient to receive and read the message
- Using Internet Mapping tools, the approximate geographic location of the recipient: in some cases this can be as accurate as the city where the recipient is located, and in extreme cases it can identify the recipient’s exact location.

Most people, when they see this list for the first time, are stunned: they have no idea that simply reading, or even *previewing*, a message in a program that shows the pretty pictures can give away so much information about them. What’s more, no antiviral or firewall software can protect against this – the only protection is to disable the download of graphics in your program, if it permits that, or to use filtering precautions that prevent the use of *Lazy HTML* (see below).

Web bugs are common – in the author’s collection of 85,000 spam messages, collected over a six year period, some 2,700 contain web bugs, and the proportion is increasing over time. What’s more, they are not limited to spam: many vendors, such as AOL, specifically reserve the right to use them as part of their privacy policy (see http://www.aol.com/info/privacy.adp).

B.2: “Lazy HTML”

Internet standards exist that allow HTML messages containing graphics to include those graphics as an attachment to the message: mail conforming to these standards is usually referred to as *MHTML* mail. Sending a message as MHTML mail means that the recipient can view the message without needing to download anything else, which means that he or she can read the message offline and without long, unnecessary delays while graphics are retrieved.

The vast majority of spam, however, does not use MHTML as the vehicle for its formatted content: instead, it uses HTML where the graphics are referenced as links to files on a remote server, a technique known as *Lazy HTML*. Lazy HTML saves the spammer money because he doesn’t have to ship the graphics with every copy of his spam, and instead transfers the bandwidth burden to the recipient, whose mail program has to initiate a connection to retrieve the files. Lazy HTML also introduces the risk of web bugs (see Appendix B.1).

A huge proportion of all spam could currently be eliminated simply by filtering out messages containing Lazy HTML formatting (messages where the HTML contains `IMG SRC` references to non-local URLs). Unfortunately, most commercial operators (travel agents, computer stores, online booksellers and so on) also use Lazy HTML in their promotional or confirmation material, which means that this approach, while attractive, would have an impact on recipients who made regular use of e-commerce facilities.

B.3: HTML tag obfuscation

As heuristic filtering techniques advance, spammers have had to find ways of breaking up or hiding key words that would identify their messages as spam. One very common technique for doing this is to insert an HTML comment into the middle of the “naughty” word – like this excerpt, taken from a real spam the author has on file:

```
<html>
Lo<!--mom-->wer Ho<!--dad-->use Pa<!-granny-->yments Gua<!-- miles and
miles-->ranteed !<p> We Can ap<!--home soon-->prove AN<!--sooner i
hope-->YONE, and can find the most com<!--dad said-->petitive rates
for YOUR CREDIT. <p> <a href=http://217.170.72.73/cgi-
bin/best_rate_virtual.cgi?code=btSEbt>Cli<!--end-->ck He<!--end-->re
</a><p>Takes less than 1 m<!--not a problemo-->inute !
</html>
```

When displayed by an HTML-aware program, the comments are removed, and this is what the user actually sees:

```
Lower House Payments Guaranteed ! We Can approve ANYONE, and can find
the most competitive rates for YOUR CREDIT.

Click Here

Takes less than 1 minute !
```

All the cruft in the message about “mom”, “granny” and “home soon” is simply ignored by the HTML parser, but will fool many heuristic spam detection engines that don’t specifically strip tags out before doing their tests.

More recently, an advanced variation on this approach has started appearing – presumably because most spam filtering engines are getting wise to the simple comment-obfuscation shown above. In the new version, key words are split up in a different way:

```
<html><body><p align=center><font face="Verdana"><b><font
```

```
size=5>Introducing GR<frame><noframes>ch233</noframes></frame>X2
Pills<br></font></b>
```

Note the use of the special HTML *frame* and *noframes* commands to introduce obfuscation into the spam "giveaway keyword" *GRX*. A spam engine that simply strips all tags out of the message before doing its checks will not strip the "ch233", and will therefore not detect the keyword; any program that performs full HTML frame-aware parsing, however, will correctly remove the superfluous text when displaying the frame version of the document.

B.4: Header forgery

This aspect of spam is so omnipresent now that it barely warrants a mention, but practically all spammers attempt to cover their tracks by forging or falsifying many of the special machine-readable headers in the e-mail messages they send. There are two reasons for doing this:

1. It makes it harder to trace the spammer and either complain to them directly (not recommended) or to complain to their ISP.

2. It often assists in passing the spam through systems that might not otherwise accept the mail: this is typically true in cases where simplistic checks based on the apparent sender or recipient addresses in the headers are used by those systems.

A very common form of header forgery is for the *To:* and *From:* fields of the message both to be set to the recipient's address – in other words, it looks to the mail system as though the recipient has sent a message to himself.

The biggest problem with header forgery is that it can lead to perfectly innocent people being blamed for sending spam: in some cases, it has even resulted in innocent parties being blacklisted even though they had done nothing at all – a spammer had merely forged their address in his messages in order to deflect the blame onto them.

One of the earliest and most famous anti-spam lawsuits, the Flowers.com case, brought action against a spammer in 1997 for forging an innocent party's e-mail address. Although the suit was successful, it had little practical impact on the problem because of the cost of enforcement. For more information, see http://news.com.com/2100-1023-205363.html?legacy=cnet.

The golden rule when dealing with spam is not to rely on the headers as being accurate. Some are – for instance, it is difficult to forge "Received:" headers – but you must not assume that the person whose address appears in the "From" field of a message is either the person who sent the spam, nor even that he or she knows anything about it.

One of the few good things about the US CAN-SPAM act (see section 5.2.1.1) is that it makes header forgery illegal, although it is not yet clear whether the will exists to enforce this requirement.

B.5: "Trigger word" obfuscation

As spam filtering engines have become more sophisticated, so certain key words in common spam have become such a liability that spammers are now being forced to obscure them through deliberate misspelling, or by the introduction of unusual characters. Classic examples of this include spellings like "Vi@gra" and "Pen1s". Even these attempts to obscure "beacon words" are now being commonly detected by filtering software, and a result has been spam that is so badly mangled that it is almost incomphrehensible. This is good, of course – if spam is hard to read, it is fair to conclude that it will be less effective as a sales vehicle.