Internet Spamming

Kimbery Mccue

First Edition, 2012

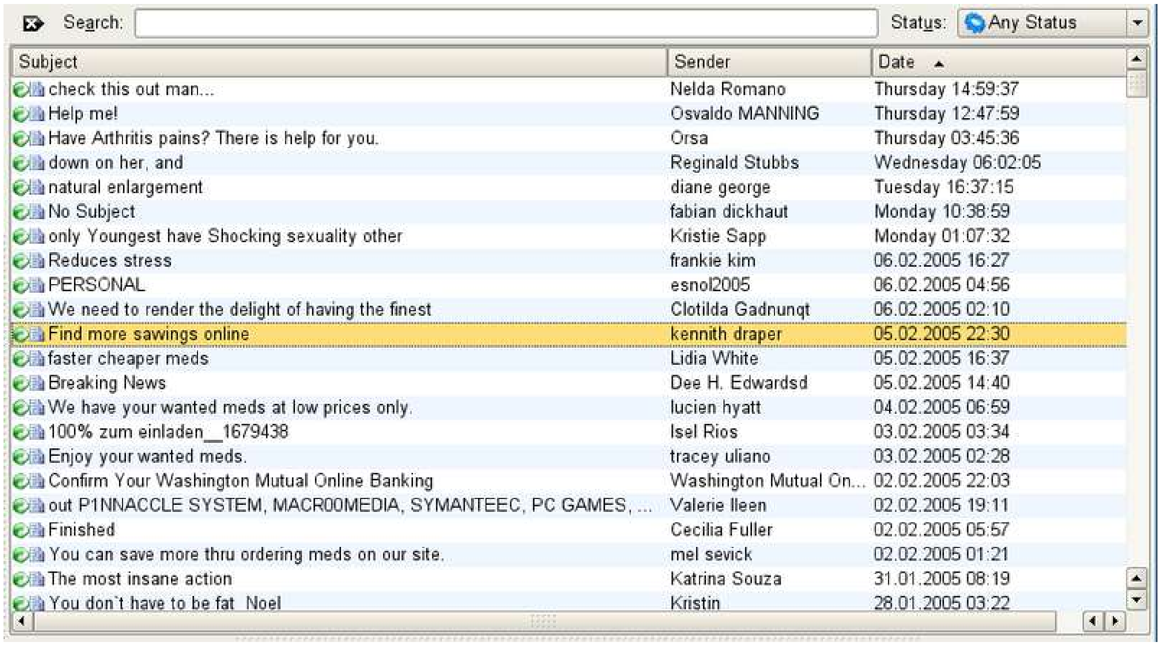
ISBN 978-81-323-4173-4

Published by:
White Word Publications
4735/22 Prakashdeep Bldg,
Ansari Road, Darya Ganj,
Delhi - 110002
Email: info@wtbooks.com

Table of Contents

Chapter 1

Spam



Subject	Sender	Date
check this out man...	Nelda Romano	Thursday 14:59:37
Help me!	Osvaldo MANNING	Thursday 12:47:59
Have Arthritis pains? There is help for you.	Orsa	Thursday 03:45:36
down on her, and	Reginald Stubbs	Wednesday 06:02:05
natural enlargement	diane george	Tuesday 16:37:15
No Subject	fabian dickhaut	Monday 10:38:59
only Youngest have Shocking sexuality other	Kristie Sapp	Monday 01:07:32
Reduces stress	frankie kim	06.02.2005 16:27
PERSONAL	esnol2005	06.02.2005 04:56
We need to render the delight of having the finest	Clotilda Gadnunqt	06.02.2005 02:10
Find more sawings online	kennith draper	05.02.2005 22:30
faster cheaper meds	Lidia White	05.02.2005 16:37
Breaking News	Dee H. Edwardsd	05.02.2005 14:40
We have your wanted meds at low prices only.	lucien hyatt	04.02.2005 06:59
100% zum einladen__1679438	Isel Rios	03.02.2005 03:34
Enjoy your wanted meds.	tracey uliano	03.02.2005 02:28
Confirm Your Washington Mutual Online Banking	Washington Mutual On...	02.02.2005 22:03
out P1NNACCLE SYSTEM, MACR00MEDIA, SYMANTEEC, PC GAMES, ...	Valerie Ileen	02.02.2005 19:11
Finished	Cecilia Fuller	02.02.2005 05:57
You can save more thru ordering meds on our site.	mel sevick	02.02.2005 01:21
The most insane action	Katrina Souza	31.01.2005 08:19
You don`t have to be fat Noel	Kristin	28.01.2005 03:22

An email box folder littered with spam messages.

Spam is the use of electronic messaging systems (including most broadcast media, digital delivery systems) to send unsolicited bulk messages indiscriminately. While the most widely recognized form of spam is e-mail spam, the term is applied to similar abuses in other media: instant messaging spam, Usenet newsgroup spam, Web search engine spam, spam in blogs, online classified ads spam, mobile phone messaging spam, Internet forum spam, junk fax transmissions, social networking spam, television advertising and file sharing network spam.

Spamming remains economically viable because advertisers have no operating costs beyond the management of their mailing lists, and it is difficult to hold senders accountable for their mass mailings. Because the barrier to entry is so low, spammers are numerous, and the volume of unsolicited mail has become very high. In the year 2011, the estimated figure for spam messages is around seven trillion. The costs, such as lost productivity and fraud, are borne by the public and by Internet service providers, which have been forced to add extra capacity to cope with the deluge. Spamming has been the subject of legislation in many jurisdictions.

People who create electronic spam are called *spammers*.

In different media

E-mail

E-mail spam, known as unsolicited bulk Email (UBE), junk mail, or unsolicited commercial email (UCE), is the practice of sending unwanted e-mail messages, frequently with commercial content, in large quantities to an indiscriminate set of recipients. Spam in e-mail started to become a problem when the Internet was opened up to the general public in the mid-1990s. It grew exponentially over the following years, and today composes some 80 to 85% of all the email in the world, by a "conservative estimate". Pressure to make e-mail spam illegal has been successful in some jurisdictions, but less so in others. Spammers take advantage of this fact, and frequently outsource parts of their operations to countries where spamming will not get them into legal trouble.

Increasingly, e-mail spam today is sent via "zombie networks", networks of virus- or worm-infected personal computers in homes and offices around the globe; many modern worms install a backdoor which allows the spammer access to the computer and use it for malicious purposes. This complicates attempts to control the spread of spam, as in many cases the spam doesn't even originate from the spammer. In November 2008 an ISP, McColo, which was providing service to botnet operators, was depeered and spam dropped 50%-75% Internet-wide. At the same time, it is becoming clear that malware authors, spammers, and phishers are learning from each other, and possibly forming various kinds of partnerships.

An industry of e-mail address harvesting is dedicated to collecting email addresses and selling compiled databases. Some of these address harvesting approaches rely on users not reading the fine print of agreements, resulting in them agreeing to send messages indiscriminately to their contacts. This is a common approach in social networking spam such as that generated by the social networking site Quechup.

Instant Messaging

Instant Messaging spam makes use of instant messaging systems. Although less ubiquitous than its e-mail counterpart, according to a report from Ferris Research, 500 million spam IMs were sent in 2003, twice the level of 2002. As instant messaging tends to not be blocked by firewalls, it is an especially useful channel for spammers. This is very common on many instant messaging system such as Skype.

Newsgroup and forum

Newsgroup spam is a type of spam where the targets are Usenet newsgroups. Spamming of Usenet newsgroups actually pre-dates e-mail spam. Usenet convention defines spamming as excessive multiple posting, that is, the repeated posting of a message (or substantially similar messages). The prevalence of Usenet spam led to the development of the Breidbart Index as an objective measure of a message's "spamminess".

Forum spam is the creating of messages that are advertisements or otherwise unwanted on Internet forums. It is generally done by automated spambots. Most forum spam consists of links to external sites, with the dual goals of increasing search engine visibility in highly competitive areas such as weight loss, pharmaceuticals, gambling, pornography, real estate or loans, and generating more traffic for these commercial websites. Some of these links contain code to track the spambot's identity if a sale goes through, when the spammer behind the spambot works on commission.

Mobile phone

Mobile phone spam is directed at the text messaging service of a mobile phone. This can be especially irritating to customers not only for the inconvenience but also because of the fee they may be charged per text message received in some markets. The term "SpaSMS" was coined at the adnews website Adland in 2000 to describe spam SMS.

Online game messaging

Many online games allow players to contact each other via player-to-player messaging, chat rooms, or public discussion areas. What qualifies as spam varies from game to game, but usually this term applies to all forms of message flooding, violating the terms of service contract for the website. This is particularly common in MMORPGs where the spammers are trying to sell game-related "items" for real-world money, chiefly among these items is in-game currency. This kind of spamming is also called Real World Trading (RWT). In the popular MMORPG Runescape, it is common for spammers to advertise sites that sell gold in multiple methods of spam. They send spam via the in-game private messaging system, via using emotes to gain attention, and by yelling publicly to everyone in the area.

Spam targeting search engines (spamdexing)

Spamdexing (a portmanteau of *spamming* and *indexing*) refers to a practice on the World Wide Web of modifying HTML pages to increase the chances of them being placed high on search engine relevancy lists. These sites use "black hat search engine optimization (SEO) techniques" to deliberately manipulate their rank in search engines. Many modern search engines modified their search algorithms to try to exclude web pages utilizing spamdexing tactics. For example, the search bots will detect repeated keywords as spamming by using a grammar analysis. If a website owner is found to have spammed the webpage to falsely increase its page rank, the website may be penalized by search engines.

Spam targeting video sharing sites

Video sharing sites, such as YouTube, are now being frequently targeted by spammers. The most common technique involves people (or spambots) posting links to sites, most likely pornographic or dealing with online dating, on the comments section of random videos or people's profiles. Another frequently used technique is using bots to post messages on random users' profiles to a spam account's channel page, along with enticing text and images, usually of a sexually suggestive nature. These pages may include their own or other users' videos, again often suggestive. The main purpose of these accounts is to draw people to their link in the home page section of their profile. YouTube has blocked the posting of such links. In addition, YouTube has implemented a CAPTCHA system that makes rapid posting of repeated comments much more difficult than before, because of abuse in the past by mass-spammers who would flood people's profiles with thousands of repetitive comments.

Yet another kind is actual video spam, giving the uploaded movie a name and description with a popular figure or event which is likely to draw attention, or within the video has a certain image timed to come up as the video's thumbnail image to mislead the viewer. The actual content of the video ends up being totally unrelated, a Rickroll, sometimes offensive, or just features on-screen text of a link to the site being promoted. Others may upload videos presented in an infomercial-like format selling their product which feature actors and paid testimonials, though the promoted product or service is of dubious quality and would likely not pass the scrutiny of a standards and practices department at a television station or cable network.

SPIT

SPIT (SPam over Internet Telephony) is VoIP (Voice over Internet Protocol) spam, usually using SIP (Session Initiation Protocol).

Noncommercial forms

E-mail and other forms of spamming have been used for purposes other than advertisements. Many early Usenet spams were religious or political. Serdar Argic, for instance, spammed Usenet with historical revisionist screeds. A number of evangelists have spammed Usenet and e-mail media with preaching messages. A growing number of criminals are also using spam to perpetrate various sorts of fraud, and in some cases have used it to lure people to locations where they have been kidnapped, held for ransom, and even murdered.

Geographical origins

A 2009 Cisco Systems report lists the origin of spam by country as follows:

Rank	Country	Spam messages per year (in trillions)
1	🇧🇷 Brazil	7.7
2	🇺🇸 United States	6.6
3	🇮🇳 India	3.6
4	🇰🇷 South Korea	3.1
5	🇹🇷 Turkey	2.6
6	🇻🇳 Vietnam	2.5
7	🇨🇳 China	2.4
8	🇵🇱 Poland	2.4
9	🇷🇺 Russia	2.3
10	🇦🇷 Argentina	1.5

History

Pre-Internet

In the late 19th Century Western Union allowed telegraphic messages on its network to be sent to multiple destinations. The first recorded instance of a mass unsolicited commercial telegram is from May 1864. Up until the Great Depression wealthy North American residents would be deluged with nebulous investment offers. This problem never fully emerged in Europe to the degree that it did in the Americas, because telegraphy was regulated by national post offices in the European region.

Etymology

According to the Internet Society and other sources, the term *spam* is derived from the 1970 *Spam* sketch of the BBC television comedy series "Monty Python's Flying Circus". The sketch is set in a cafe where nearly every item on the menu includes Spam canned luncheon meat. As the waiter recites the Spam-filled menu, a chorus of Viking patrons drowns out all conversations with a song repeating "Spam, Spam, Spam, Spam... lovely Spam! wonderful Spam!", hence "Spamming" the dialogue. The excessive amount of Spam mentioned in the sketch is a reference to the preponderance of imported canned meat products in the United Kingdom, particularly corned beef from Argentina, in the years after World War II, as the country struggled to rebuild its agricultural base. Spam captured a large slice of the British market within lower economic classes and became a byword among British children of the 1960s for low-grade fodder due to its commonality, monotonous taste and cheap price - hence the humour of the Python sketch.

In the 1980s the term was adopted to describe certain abusive users who frequented BBSs and MUDs, who would repeat "Spam" a huge number of times to scroll other users' text off the screen. In early Chat rooms services like PeopleLink and the early days of AOL, they actually flooded the screen with quotes from the Monty Python Spam sketch. With internet connections over phone lines, typically running at 1200 or even 300 bit/s, it could take an enormous amount of time for a *spammy* logo, drawn in *ASCII art* to scroll to completion on a viewer's terminal. Sending an irritating, large, meaningless block of text in this way was called *spamming*. This was used as a tactic by insiders of a group that wanted to drive newcomers out of the room so the usual conversation could continue. It was also used to prevent members of rival groups from chatting—for instance, Star Wars fans often invaded Star Trek chat rooms, filling the space with blocks of text until the Star Trek fans left. This act, previously called *flooding* or *trashing*, came to be known as *spamming*. The term was soon applied to a large amount of text broadcast by many users.

It later came to be used on Usenet to mean *excessive multiple posting*—the repeated posting of the same message. The unwanted message would appear in many if not all newsgroups, just as Spam appeared in nearly all the menu items in the Monty Python sketch. The first usage of this sense was by Joel Furr in the aftermath of the ARMM incident of March 31, 1993, in which a piece of experimental software released dozens of recursive messages onto the *news.admin.policy* newsgroup. This use had also become established—to spam Usenet was flooding newsgroups with junk messages. The word was also attributed to the flood of "Make Money Fast" messages that clogged many newsgroups during the 1990s. In 1998, the New Oxford Dictionary of English, which had previously only defined "spam" in relation to the trademarked food product, added a second definition to its entry for "spam": "Irrelevant or inappropriate messages sent on the Internet to a large number of newsgroups or users."

There are several popular false etymologies of the word "spam". One, promulgated by early spammers Laurence Canter and Martha Siegel, is that "spamming" is what happens when one dumps a can of Spam luncheon meat into a fan blade. Some others are the backronym **s**tupid **p**ointless **a**nnoying **m**essages." There was also an effort to differentiate between types of spam. That which was sent indiscriminately to any e-mail address was true spam while that which was targeted to more likely prospects, although just as unsolicited, was called velveeta (after the cheese product). But this latter term didn't persist.

History of Internet forms

The earliest documented spam was a message advertising the availability of a new model of Digital Equipment Corporation computers sent to 393 recipients on ARPANET in 1978, by Gary Thuerk. The term "spam" for this practice had not yet been applied. Spamming had been practiced as a prank by participants in multi-user dungeon games, to fill their rivals' accounts with unwanted electronic junk. The first known electronic chain letter, titled Make Money Fast, was released in 1988.

The first major commercial spam incident started on March 5, 1994, when a husband and wife team of lawyers, Laurence Canter and Martha Siegel, began using bulk Usenet posting to advertise immigration law services. The incident was commonly termed the "Green Card spam", after the subject line of the postings. Defiant in the face of widespread condemnation, the attorneys claimed their detractors were hypocrites or "zealouts", claimed they had a free speech right to send unwanted commercial messages, and labeled their opponents "anti-commerce radicals." The couple wrote a controversial book entitled *How to Make a Fortune on the Information Superhighway*.

Later that year a poster operating under the alias Serdar Argic posted antagonistic messages denying the Armenian Genocide to tens of thousands of Usenet discussions that had been searched for the word Turkey. Within a few years, the focus of spamming (and anti-spam efforts) moved chiefly to e-mail, where it remains today. Arguably, the aggressive email spamming by a number of high-profile spammers such as Sanford Wallace of Cyber Promotions in the mid-to-late 1990s contributed to making spam predominantly an email phenomenon in the public mind. By 2009, the majority of spam sent around the world was in the English language; spammers began using automatic translation services to send spam in other languages.

Trademark issues

Hormel Foods Corporation, the maker of Spam luncheon meat, does not object to the Internet use of the term "spamming". However, they did ask that the capitalized word "Spam" be reserved to refer to their product and trademark. By and large, this request is obeyed in forums which discuss spam. In Hormel Foods v SpamArrest, Hormel attempted to assert its trademark rights against SpamArrest, a software company, from using the mark "spam", since Hormel owns the trademark. In a dilution claim, Hormel argued that Spam Arrest's use of the term "spam" had endangered and damaged "substantial goodwill and good reputation" in connection with its trademarked lunch meat and related products. Hormel also asserts that Spam Arrest's name so closely resembles its luncheon meat that the public might become confused, or might think that Hormel endorses Spam Arrest's products.

Hormel did not prevail. Attorney Derek Newman responded on behalf of Spam Arrest: "Spam has become ubiquitous throughout the world to describe unsolicited commercial e-mail. No company can claim trademark rights on a generic term." Hormel stated on its website: "Ultimately, we are trying to avoid the day when the consuming public asks, 'Why would Hormel Foods name its product after junk email?".

Hormel also made two attempts that were dismissed in 2005 to revoke the marks "SPAMBUSTER" and Spam Cube. Hormel's Corporate Attorney Melanie J. Neumann also sent SpamCop's Julian Haight a letter on August 27, 1999 requesting that he delete an objectionable image (a can of Hormel's Spam luncheon meat product in a trash can), change references to UCE spam to all lower case letters, and confirm his agreement to do so.

Cost Benefit Analyses

The European Union's Internal Market Commission estimated in 2001 that "junk e-mail" cost Internet users €10 billion per year worldwide. The California legislature found that spam cost United States organizations alone more than $13 billion in 2007, including lost productivity and the additional equipment, software, and manpower needed to combat the problem. Spam's direct effects include the consumption of computer and network resources, and the cost in human time and attention of dismissing unwanted messages.

In addition, spam has costs stemming from the *kinds* of spam messages sent, from the *ways* spammers send them, and from the *arms race* between spammers and those who try to stop or control spam. In addition, there are the opportunity cost of those who forgo the use of spam-afflicted systems. There are the direct costs, as well as the indirect costs borne by the victims—both those related to the spamming itself, and to other crimes that usually accompany it, such as financial theft, identity theft, data and intellectual property theft, virus and other malware infection, child pornography, fraud, and deceptive marketing.

The cost to providers of search engines is not insignificant: "The secondary consequence of spamming is that search engine indexes are inundated with useless pages, increasing the cost of each processed query". The methods of spammers are likewise costly. Because spamming contravenes the vast majority of ISPs' acceptable-use policies, most spammers have for many years gone to some trouble to conceal the origins of their spam. E-mail, Usenet, and instant-message spam are often sent through insecure proxy servers belonging to unwilling third parties. Spammers frequently use false names, addresses, phone numbers, and other contact information to set up "disposable" accounts at various Internet service providers. In some cases, they have used falsified or stolen credit card numbers to pay for these accounts. This allows them to quickly move from one account to the next as each one is discovered and shut down by the host ISPs.

The costs of spam also include the collateral costs of the struggle between spammers and the administrators and users of the media threatened by spamming. Many users are bothered by spam because it impinges upon the amount of time they spend reading their e-mail. Many also find the content of spam frequently offensive, in that pornography is one of the most frequently advertised products. Spammers send their spam largely indiscriminately, so pornographic ads may show up in a work place e-mail inbox—or a child's, the latter of which is illegal in many jurisdictions. Recently, there has been a noticeable increase in spam advertising websites that contain child pornography.

Some spammers argue that most of these costs could potentially be alleviated by having spammers reimburse ISPs and persons for their material. There are three problems with this logic: first, the rate of reimbursement they could credibly budget is not nearly high enough to pay the direct costs, second, the human cost (lost mail, lost time, and lost opportunities) is basically unrecoverable, and third, spammers often use stolen bank accounts and credit cards to finance their operations, and would conceivably do so to pay off any fines imposed.

E-mail spam exemplifies a tragedy of the commons: spammers use resources (both physical and human), without bearing the entire cost of those resources. In fact, spammers commonly do not bear the cost at all. This raises the costs for everyone. In some ways spam is even a potential threat to the entire e-mail system, as operated in the past. Since e-mail is so cheap to send, a tiny number of spammers can saturate the Internet with junk mail. Although only a tiny percentage of their targets are motivated to purchase their products (or fall victim to their scams), the low cost may provide a sufficient conversion rate to keep the spamming alive. Furthermore, even though spam appears not to be economically viable as a way for a reputable company to do business, it suffices for professional spammers to convince a tiny proportion of gullible advertisers that it is viable for those spammers to stay in business. Finally, new spammers go into business every day, and the low costs allow a single spammer to do a lot of harm before finally realizing that the business is not profitable.

Some companies and groups "rank" spammers; spammers who make the news are sometimes referred to by these rankings. The secretive nature of spamming operations makes it difficult to determine how proliferated an individual spammer is, thus making the spammer hard to track, block or avoid. Also, spammers may target different networks to different extents, depending on how successful they are at attacking the target. Thus considerable resources are employed to actually measure the amount of spam generated by a single person or group. For example, victims that use common anti-spam hardware, software or services provide opportunities for such tracking. Nevertheless, such rankings should be taken with a grain of salt.

General costs

In all cases listed above, including both commercial and non-commercial, "spam happens" because of a positive Cost-benefit analysis result if the cost to recipients is excluded as an externality the spammer can avoid paying.

Cost is the combination of

- Overhead: The costs and overhead of electronic spamming include bandwidth, developing or acquiring an email/blog spam tool, taking over or acquiring a host/zombie, etc.
- Transaction cost: The incremental cost of contacting each additional recipient once a method of spamming is constructed, multiplied by the number of recipients.
- Risks: Chance and severity of legal and/or public reactions, including damages and punitive damages
- Damage: Impact on the community and/or communication channels being spammed

Benefit is the total expected profit from spam, which may include any combination of the commercial and non-commercial reasons listed above. It is normally linear, based on the incremental benefit of reaching each additional spam recipient, combined with the conversion rate. The conversion rate for botnet-generated spam has recently been measured to be around one in 12,000,000 for pharmaceutical spam and one in 200,000 for infection sites as used by the Storm botnet. They specifically say in the paper "After 26 days, and almost 350 million e-mail messages, only 28 sales resulted".

Spam is prevalent on the Internet because the transaction cost of electronic communications is radically less than any alternate form of communication, far outweighing the current potential losses, as seen by the amount of spam currently in existence. Spam continues to spread to new forms of electronic communication as the gain (number of potential recipients) increases to levels where the cost/benefit becomes positive. Spam has most recently evolved to include blogspam as the levels of readership increase to levels where the overhead is no longer the dominating factor. According to the above analysis, spam levels will continue to increase until the cost/benefit analysis is balanced.

In crime

Spam can be used to spread computer viruses, trojan horses or other malicious software. The objective may be identity theft, or worse (e.g., advance fee fraud). Some spam attempts to capitalize on human greed whilst other attempts to use the victims' inexperience with computer technology to trick them (e.g., phishing). On May 31, 2007, one of the world's most prolific spammers, Robert Alan Soloway, was arrested by U.S. authorities. Described as one of the top ten spammers in the world, Soloway was charged with 35 criminal counts, including mail fraud, wire fraud, e-mail fraud, aggravated identity theft and money laundering. Prosecutors allege that Soloway used millions of "zombie" computers to distribute spam during 2003. This is the first case in which U.S. prosecutors used identity theft laws to prosecute a spammer for taking over someone else's Internet domain name.

Political issues

Spamming remains a hot discussion topic. In 2004, the seized Porsche of an indicted spammer was advertised on the Internet; this revealed the extent of the financial rewards available to those who are willing to commit duplicitous acts online. However, some of the possible means used to stop spamming may lead to other side effects, such as increased government control over the Internet, loss of privacy, barriers to free expression, and the commercialization of e-mail.

One of the chief values favored by many long-time Internet users and experts, as well as by many members of the public, is the free exchange of ideas. Many have valued the relative anarchy of the Internet, and bridle at the idea of restrictions placed upon it. A common refrain from spam-fighters is that spamming itself abridges the historical freedom of the Internet, by attempting to force users to carry the *costs* of material which they would not choose.

An ongoing concern expressed by parties such as the Electronic Frontier Foundation and the ACLU has to do with so-called "stealth blocking", a term for ISPs employing aggressive spam blocking without their users' knowledge. These groups' concern is that ISPs or technicians seeking to reduce spam-related costs may select tools which (either through error or design) also block non-spam e-mail from sites seen as "spam-friendly". SPEWS is a common target of these criticisms. Few object to the existence of these tools; it is their use in filtering the mail of users who are not informed of their use which draws fire.

Some see spam-blocking tools as a threat to free expression—and laws against spamming as an untoward precedent for regulation or taxation of e-mail and the Internet at large. Even though it is possible in some jurisdictions to treat some spam as unlawful merely by applying existing laws against trespass and conversion, some laws specifically targeting spam have been proposed. In 2004, United States passed the CAN-SPAM Act of 2003 which provided ISPs with tools to combat spam. This act allowed Yahoo! to successfully sue Eric Head, reportedly one of the biggest spammers in the world, who settled the lawsuit for several thousand U.S. dollars in June 2004. But the law is criticized by many for not being effective enough. Indeed, the law was supported by some spammers and organizations which support spamming, and opposed by many in the anti-spam community. Examples of effective anti-abuse laws that respect free speech rights include those in the U.S. against unsolicited faxes and phone calls, and those in Australia and a few U.S. states against spam.

In November 2004, Lycos Europe released a screen saver called make LOVE not SPAM which made Distributed Denial of Service attacks on the spammers themselves. It met with a large amount of controversy and the initiative ended in December 2004.

While most countries either outlaw or at least ignore spam, Bulgaria is the first and until now only one to partially legalize it. According to recent changes in the Bulgarian E-Commerce act anyone can send spam to mailboxes, owned by company or organization, as long as there is warning that this may be unsolicited commercial email in the message body. The law contains many other inadequate texts - for example the creation of a nationwide public electronic register of email addresses that do not want to receive spam, something valuable only as source for e-mail address harvesting.

Anti-spam policies may also be a form of disguised censorship, a way to ban access or reference to questioning alternative forums or blogs by an institution. This form of occult censorship is mainly used by private companies when they can not muzzle criticism by legal ways.

Court cases

United States

Sanford Wallace and Cyber Promotions were the target of a string of lawsuits, many of which were settled out of court, up through the famous 1998 Earthlink settlementwhich put Cyber Promotions out of business. Attorney Laurence Canter was disbarred by the Tennessee Supreme Court in 1997 for sending prodigious amounts of spam advertising his immigration law practice. In 2005, Jason Smathers, a former America Online employee, pled guilty to charges of violating the CAN-SPAM Act. In 2003, he sold a list of approximately 93 million AOL subscriber e-mail addresses to Sean Dunaway who, in turn, sold the list to spammers.

In 2007, Robert Soloway lost a case in a federal court against the operator of a small Oklahoma-based Internet service provider who accused him of spamming. U.S. Judge Ralph G. Thompson granted a motion by plaintiff Robert Braver for a default judgment and permanent injunction against him. The judgment includes a statutory damages award of $10,075,000 under Oklahoma law.

In June 2007, two men were convicted of eight counts stemming from sending millions of e-mail spam messages that included hardcore pornographic images. Jeffrey A. Kilbride, 41, of Venice, California was sentenced to six years in prison, and James R. Schaffer, 41, of Paradise Valley, Arizona, was sentenced to 63 months. In addition, the two were fined $100,000, ordered to pay $77,500 in restitution to AOL, and ordered to forfeit more than $1.1 million, the amount of illegal proceeds from their spamming operation. The charges included conspiracy, fraud, money laundering, and transportation of obscene materials. The trial, which began on June 5, was the first to include charges under the CAN-SPAM Act of 2003, according to a release from the Department of Justice. The specific law that prosecutors used under the CAN-Spam Act was designed to crack down on the transmission of pornography in spam.

In 2005, Scott J. Filary and Donald E. Townsend of Tampa, Florida were sued by Florida Attorney General Charlie Crist for violating the Florida Electronic Mail Communications Act. The two spammers were required to pay $50,000 USD to cover the costs of investigation by the state of Florida, and a $1.1 million penalty if spamming were to continue, the $50,000 was not paid, or the financial statements provided were found to be inaccurate. The spamming operation was successfully shut down.

Edna Fiedler, 44, of Olympia, Washington, on June 25, 2008, pleaded guilty in a Tacoma court and was sentenced to 2 years imprisonment and 5 years of supervised release or probation in an Internet $1 million "Nigerian check scam." She conspired to commit bank, wire and mail fraud, against US citizens, specifically using Internet by having had an accomplice who shipped counterfeit checks and money orders to her from Lagos, Nigeria, last November. Fiedler shipped out $ 609,000 fake check and money orders when arrested and prepared to send additional $ 1.1 million counterfeit materials. Also, the U.S. Postal Service recently intercepted counterfeit checks, lottery tickets and eBay overpayment schemes with a face value of $2.1 billion.

United Kingdom

In the first successful case of its kind, Nigel Roberts from the Channel Islands won £270 against Media Logistics UK who sent junk e-mails to his personal account.

In January 2007, a Sheriff Court in Scotland awarded Mr. Gordon Dick £750 (the then maximum sum which could be awarded in a Small Claim action) plus expenses of £618.66, a total of £1368.66 against Transcom Internet Services Ltd. for breaching anti-spam laws. Transcom had been legally represented at earlier hearings but were not represented at the proof, so Gordon Dick got his decree by default. It is the largest amount awarded in compensation in the United Kingdom since Roberts -v- Media Logistics case in 2005 above, but it is not known if Mr Dick ever received anything. (An image of Media Logistics' cheque is shown on Roberts' website) Both Roberts and Dick are well known figures in the British Internet industry for other things. Dick is currently Interim Chairman of Nominet UK (the manager of .UK and .CO.UK) while Roberts is CEO of CHANNELISLES.NET (manager of .GG and .JE).

Despite the statutory tort that is created by the Regulations implementing the EC Directive, few other people have followed their example. As the Courts engage in active case management, such cases would probably now be expected to be settled by mediation and payment of nominal damages.

New Zealand

In October 2008, a vast international internet spam operation run from New Zealand was cited by American authorities as one of the world’s largest, and for a time responsible for up to a third of all unwanted emails. In a statement the US Federal Trade Commission (FTC) named Christchurch’s Lance Atkinson as one of the principals of the operation. New Zealand’s Internal Affairs announced it had lodged a $200,000 claim in the High Court against Atkinson and his brother Shane Atkinson and courier Roland Smits, after raids in Christchurch. This marked the first prosecution since the Unsolicited Electronic Messages Act (UEMA) was passed in September 2007. The FTC said it had received more than three million complaints about spam messages connected to this operation, and estimated that it may be responsible for sending billions of illegal spam messages. The US District Court froze the defendants’ assets to preserve them for consumer redress pending trial. U.S. co-defendant Jody Smith forfeited more than $800,000 and faces up to five years in prison for charges to which he plead guilty.

Chapter 2

Messaging Spam and Newsgroup Spam

Messaging spam

Messaging spam, sometimes called **SPIM**, is a type of spam targeting users of instant messaging (IM) services.

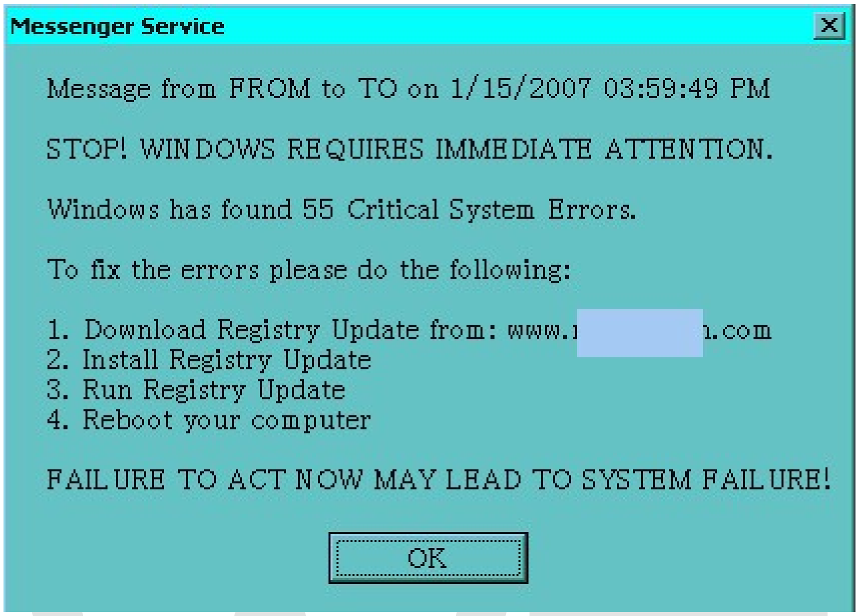
Instant messaging applications

Instant messaging systems, such as Yahoo! Messenger, AIM, Windows Live Messenger, Tencent QQ, ICQ, XMPP and Myspace chat rooms, are all targets for spammers. Many IM systems offer a directory of users, including demographic information such as age and sex. Advertisers can gather this information, sign on to the system, and send unsolicited messages, which could include commercial scam-ware, viruses, and links to paid links for the purpose of click fraud. Microsoft has announced that the upcoming Windows Live Messenger 9.0 would support specialized features to combat messaging spam. In most systems users can already block the vast majority of spam through the use of a whitelist.

Countermeasures

- Many users choose to receive IMs only from people already on their contact list.
- In corporate settings, spam over IM is blocked by IM spam blockers like those from FaceTime Communications, Akonix, ScanSafe, Symantec, and CSC.

Messenger Service spam on Windows NT-based systems



Messenger Service

Message from FROM to TO on 1/15/2007 03:59:49 PM

STOP! WINDOWS REQUIRES IMMEDIATE ATTENTION.

Windows has found 55 Critical System Errors.

To fix the errors please do the following:

1. Download Registry Update from: www.[redacted].com
2. Install Registry Update
3. Run Registry Update
4. Reboot your computer

FAILURE TO ACT NOW MAY LEAD TO SYSTEM FAILURE!

Example of Messenger Service spam from 2007.

In 2002, a number of spammers began abusing the Messenger Service, a function of Windows designed to allow administrators to send alerts to users' workstations (not to be confused with Windows Messenger or Windows Live Messenger, a free instant messaging application) in Microsoft's Windows NT-based operating systems. Messenger Service spam appears as normal dialog boxes containing the spammer's message. These messages are easily blocked by firewalls configured to block packets to the NetBIOS ports 135-139 and 445 as well as unsolicited UDP packets to ports above 1024. Additionally, Windows XP Service Pack 2 disables the Messenger Service by default.

Messenger Service spammers frequently send messages to vulnerable Windows machines with a URL. The message promises the user to eradicate spam messages sent via the Messenger Service. The URL leads to a Web site where, for a fee, users are told how to disable the Messenger service. Though the Messenger is easily disabled for free by the user, this works because it creates a perceived need and then offers an immediate solution.

Newsgroup spam

Newsgroup spam is a type of spam where the targets are Usenet newsgroups.

Spamming of Usenet newsgroups actually pre-dates e-mail spam. The first widely recognized Usenet spam (though not the most famous) was posted on 18 January 1994 by

Clarence L. Thomas IV, a sysadmin at Andrews University. Entitled "Global Alert for All: Jesus is Coming Soon", it was a fundamentalist religious tract claiming that "this world's history is coming to a climax." The newsgroup posting bot Serdar Argic also appeared in early 1994, posting tens of thousands of messages to various newsgroups, consisting of identical copies of a political screed relating to the Armenian Genocide.

The first "commercial" Usenet spam, and the one which is often (mistakenly) claimed to be the first Usenet spam of any sort, was an advertisement for legal services entitled "Green Card Lottery - Final One?". It was posted in April 1994 by Arizona lawyers Laurence Canter and Martha Siegel, and hawked legal representation for United States immigrants seeking papers ("green cards").

Usenet convention defines spamming as "excessive multiple posting", that is, the repeated posting of a message (or substantially similar messages). During the early 1990s there was substantial controversy among Usenet system administrators (news admins) over the use of cancel messages to control spam. A "cancel message" is a directive to news servers to delete a posting, causing it to be inaccessible. Some regarded this as a bad precedent, leaning towards censorship, while others considered it a proper use of the available tools to control the growing spam problem.

A culture of neutrality towards content precluded defining spam on the basis of advertisement or commercial solicitations. The word "spam" was usually taken to mean "excessive multiple posting (EMP)", and other neologisms were coined for other abuses – such as "velveeta" (from the processed cheese product of that name) for "excessive cross-posting". A subset of spam was deemed "cancellable spam", for which it is considered justified to issue third-party cancel messages.

In the late 1990s, spam became used as a means of vandalising newsgroups, with malicious users committing acts of sporgery to make targeted newsgroups all but unreadable without heavily filtering. A prominent example occurred in alt.religion.scientology.

Prevalent in recent times is the MI-5 Persecution spam, which is well known across many newsgroups. These rambling postings often appear as clusters of twenty or more messages with varying subjects and content, but all related to Mike Corley's perceived surveillance of himself by MI5, the British intelligence agency. These rambling messages used to state the originator as MI5Victim@mi5.gov.uk. Lately (December 2007) the spammer has taken to altering the "from" address and subject line in an attempt to get past newsgroup "kill" filters. This UK-based spammer readily admits that he has mental illness in several of his postings.

The prevalence of Usenet spam led to the development of the Breidbart Index as an objective measure of a message's "spamminess". The use of the BI and spam-detection software has led to Usenet being policed by anti-spam volunteers, who purge newsgroups of spam by sending cancels and filtering it out on the way into servers. This very active form of policing has meant that Usenet is a far less attractive target to spammers than it used to be, and most of the industrial-scale spammers have now moved into e-mail spam instead.

Google Usenet News Archive

Unfortunately, the advent of the large Usenet archive kept as part of the Google Groups website, has made Usenet more attractive to spammers than ever. The goal in this case is not just to reach the members of a newsgroup, but to also take advantage of the fact that Google gives a higher pagerank to websites that are referred to by these messages, which are catalogued and mirrored in multiple languages at Google's top-level domain. Critics have suggested that Google has ulterior motives for "turning a blind eye" to the problem since the websites being pointed to use Google ads, which potentially generate revenue for both the spammer AND Google. The spam is extremely unfair to the companies paying Google and the spammer for an ad-click, as the most prevalent current spam (2010) is trying to trick readers into clicking on web ads by referring to them as images and saying that a link is hidden in them "due to high sex content" or that a link hidden in the image (Google ad) will take them to a "PayPal form" that will give them money.

While most newsreaders filter the spam at either the server or user level, Google does not filter spam out of its Usenet News archive. Google does, however, offer spam filtering for groups that decide to abandon Usenet and form a moderated Google Group, which gives another reason why Google would turn a blind eye to spam in its archive of Usenet News.

Chapter 3

Forum Spam and Mobile Phone Spam

Forum spam

Forum spam is the creating of messages that are advertisements, abusive, or otherwise unwanted on Internet forums. It is generally done by automated spambots.

Types of spam

Forum spambots surf the web, looking for guestbooks, blogs, forums and any other web forms to submit spam links to. These spambots often use OCR technology to bypass CAPTCHAs present. Some spam messages are targeted towards readers and can involve techniques of target marketing or even phishing, making it hard to tell real posts from the bot generated ones. Not all of the spam posts are meant for the readers; some spam messages are simply hyperlinks intended to boost search engine ranking.

Most forum spam consists of links to external sites, with the dual goals of increasing search engine visibility in highly competitive areas such as weight loss, pharmaceuticals, gambling, pornography, real estate or loans, and generating more traffic for these commercial websites. Some of these links contain code to track the spambot's identity if a sale goes through, when the spammer behind the spambot works on commission.

Spam posts may contain anything from a single link, to dozens of links. Text content is minimal, usually innocuous and unrelated to the forum's topic, or in a very old thread that is revived by the spammer solely for the purpose of spamming links. Some text is included to prevent the post being caught by automated spam filters that prevent posts which consist solely of external links from being submitted. Full banner advertisements have also been reported.

Alternately, the spam links are posted in the user's signature, in which case the spambot will never post. The link sits quietly in the signature field, where it is more likely to be harvested by search engine spiders than discovered by forum administrators and moderators.

Since November 2006, a very destructive forum and spam attack has been propagated by inserting into comments redirect domains with an automated posting script like XRumer. These domains redirect a user to pornographic websites. If a user clicks on the image or attempts to close the Website an ActiveX codec will be downloaded as a Zlob Trojan. The spambot can often bypass many of the safeguards administrators use to reduce the amount of spam posted.

Effects of spam

Spam prevention and deletions measurably increase the workload of forum administrators and moderators. The amount of time and resources spent keeping a forum spam free contributes significantly to labor cost, and the skill required in the running of a public forum. Marginally profitable or smaller forums may be permanently closed by administrators.

Spam prevention

- *Flood control:* This forces users to wait for a short interval between making posts to the forum, thus preventing spambots from flooding the forum with repeated spam messages.
- *Registration control:*
    - Some forums employ CAPTCHA (visual confirmation) routines on their registration pages to prevent spambots carrying out automated registrations. Simple CAPTCHA systems which display alphanumeric characters have proven vulnerable to optical character recognition software but those that scramble the characters appear to be far more effective.
    - Alternative is Textual Confirmation, where the user answers one or more random questions to prove he/she is not a spambot.
    - Forums have a feature where they send an e-mail to users who registered, either containing the password used to login or an activation code/link.
    - Some forums have required registration approval where the administrator has to approve accounts.
- *Authoritative voice:* Using an external filtering service, such as Akismet, to get a verdict if the data is spam or not.
- *Posting limits:* Limit posting to registered users and/or require that the user pass a CAPTCHA test before posting.
- *Registration restrictions:* Applying careful restrictions can seriously impact bogus and spambot registrations. One approach consists in the denial of registration from certain domain extensions that are a major source of spambots such .ru, .br, .biz, or freebase addresses such as "gawab.com". Another, more labor-intensive, consists in manual examination of new registrants. This examination looks at several indicators. First, spambots often delay email confirmation by several hours, while humans will confirm promptly. Second, spambots will tend to create user names that are unique, and unlikely to already be used in the forum, preferring "John84731" or "JohnbassKeepsie" to the much more common "John."

Third, using a search engine to investigate, one finds hundreds, if not thousands of profiles using the spambot login name, sometimes with the diagnostic spam post, or "banned" label.

- Changing technical details of the forum software to confuse bots — for example, changing "agreed=true" to "mode=agreed" in the registration page of phpBB.
- Block posts or registrations that contain certain blacklisted words.
- Be wary of IPs used by untrusted posters (anonymous posts or newly registered users). A useful technique for proactive detection of well-known spammer proxies is to query a search engine for this IP. It will show up on pages that specialize in the listing of proxies.
- Some forums also have their own "spam subforums" to direct spam off their main site.
- Some forums have the signature option disabled.

Mobile phone spam

Mobile phone spam is a form of spamming directed at the text messaging service of a mobile phone. It is described as **mobile spamming**, **SMS spam**, **text spam** or **m-spam.**.

As the popularity of mobile phones surged in the early 2000s, frequent users of text messaging began to see an increase in the number of unsolicited (and generally unwanted) commercial advertisements being sent to their telephones through text messaging. This can be particularly annoying for the recipient because, unlike in email, some recipients may be charged a fee for every message received, including spam.

Mobile phone spam is generally less pervasive than email spam, where in 2010 around 90% of email is spam. The amount of mobile spam varies widely from region to region. In North America, much less than 1% of SMS messages were spam in 2010, while in parts of Asia up to 30% of messages were spam.

The lesser and geographically uneven prevalence of mobile phone spam is attributable to the higher cost (to spammers) of and technological barriers to sending mobile messages in some areas, and to law enforcement in others. Today, particularly in North America, most mobile phone spam is sent from mobile devices that have prepaid unlimited messaging rate plans. While the rate plans allow for unlimited messaging, in reality the relatively slow sending rate (on the order of magnitude of 1/s) limits the number of messages that may be sent before an abusing mobile is shut down.

Criminality and law enforcement

SMS spam is illegal under common law in most jurisdictions as trespass to chattels. Jurisdictions with specific SMS spam regulation and fines include Australia, the EU and others; in the US, violators face substantial costs; in a 2008 settlement, the violator agreed to pay $150 to each spam recipient. In a 2010 class action settlement of Satterfield v Simon & Schuster, a case that reached the US Ninth Circuit Court of Appeals, defendants agreed to pay $175 to each spam recipient. In response to Satterfield, entities who make money sending mobile phone spam formed the Mobile Advocacy Coalition (MAC) to lobby the government to legalize that activity. In the US, the Federal Trade Commission (FTC) has expanded Phone Spam regulations to cover also Voice Spam—mostly in form of prerecorded telemarketing calls—commonly known as robocalls; victims can file a complaint with the FCC. In California, Section 17538.41 of the B&P Code bans text message advertisement. Consumers can sue on an individual or class basis per a private right of action against unfair business practices.

Enforcement in small claims court

The FCC released an order in Aug, 2004 that reiterated that SMS spam messages to cellphones are illegal under the existing Telephone Consumer Protection Act (TCPA) and also under the CAN-SPAM Act. Each such unsolicited message received without permission entitles the recipient to take the sender to small claims court and collect a minimum of $1 for each violation. They said this in 2003, and reiterated it in 2004: 'In 2003, we released a Report and Order in which we reaffirmed that the TCPA prohibits any call using an automatic telephone dialing system or an artificial or prerecorded message to any wireless telephone number. We concluded that this encompasses both voice calls and text calls, including Short Message Service (SMS) text messaging calls, to wireless phone numbers.'

The 2003 TCPA Order (18 FCC Rcd at 14115, para. 165 says: “Both the statute and our rules prohibit these calls, with limited exceptions, ‘to any telephone number assigned to a paging service, cellular telephone service, specialized mobile radio service, or other common carrier service, or any service for which the called party is charged.’ This encompasses both voice calls and **text calls** to wireless numbers including, for example, short message service (SMS) calls, provided the call is made to a telephone number assigned to such service.”

Factors complicating SMS spam reduction

Fighting SMS spam is complicated by several factors, including the lower rate of SMS spam (compared to more abused services such as Internet email) has allowed many users and service providers to ignore the issue, and the limited availability of mobile phone spam-filtering software. Filtering SMS spam at the recipient device would be an imperfect solution in markets where users are charged to receive messages, as the user may still be charged for the message once the provider sent it, even if software on the device blocked it from appearing on the device's display. This problem is not present in most of the world outside the U.S., however, where users are not charged to receive messages.

Providers may fear liability should a legitimate message of an emergency nature be blocked. Nonetheless, many providers voluntarily provide their subscribers technical means for mitigating unsolicited SMS messages.

Countermeasures

There are several actions and strategies that can help reduce SMS spam.

Legal actions can be effective and remunerative.

Many carriers (such as AT&T in the US) allow subscribers to report spam by forwarding the spam messages to short code 7726 (spells SPAM on a traditional phone keypad) (33700 in France). It's reported that 1/2 million spam reports in France resulted in the disconnection of 300 spammers, and many more cease-and-desist orders were sent.

Some spam countermeasures depend on detection, and there are two developments in that area: a GSMA pilot spam reporting program, and the development of Open Mobile Alliance (OMA) standards for mobile spam reporting. In February 2010, the GSM Association has announced a pilot program that will allow subscribers to report SMS spam by forwarding it to short code '7726' which spells "SPAM" on most phones. AT&T Mobility, Korea Telecom, and SFR announced their participation, and a large number of other mobile operators are expected to join after the pilot period.

The Open Mobile Alliance is expected to complete its "SpamRep" standards which will allow users of mobile devices to report email, SMS, MMS and IM spam using a 'This-Is-Spam' button or menu item, as users of wired email systems are now doing.

Another helpful SMS spam-reduction technique guarding one's cell phone number. One of the biggest sources of SMS spam is number harvesting carried out by Internet sites offering "free" ring tone downloads. In order to facilitate the downloads, users must provide their phones' numbers; which in turn are used to send frequent advertising messages to the phone. Wording in the sites' Terms of Service intended to make this legal have not survived court challenge.

Another approach to reducing SMS spam that is offered by some carriers involves creating an alias address rather than using the cell phone's number as a text message address. Only messages sent to the alias are delivered; messages sent to the phone's number are discarded. A New York Times article provided detailed information on this in 2008.

Another countermeasure is to use a service that provides a public phone number and publishes the SMS messages received at that number to a publicly accessible website. Google Voice can be used in this way, but with numbers and messages kept private.

Finally, most cell phone providers offer the option of completely disabling *all* text messaging services on a user's account. This extreme solution, however, is satisfactory only for those users who have neither the need nor the desire to utilize SMS at all.

In June 2009, three major Chinese carriers—China Mobile, China Telecom and China Unicom—imposed limits on text messaging in order to crack down on spam SMS. Under the restrictions, a phone number can send no more than 200 messages per hour and 1000/day on weekdays.

Chapter 4

Spamdexing

In computing, **spamdexing** (also known as **search spam**, **search engine spam** or **web spam**) is the deliberate manipulation of search engine indexes. It involves a number of methods, such as repeating unrelated phrases, to manipulate the relevance or prominence of resources indexed in a manner inconsistent with the purpose of the indexing system. Some consider it to be a part of search engine optimization, though there are many search engine optimization methods that improve the quality and appearance of the content of web sites and serve content useful to many users. Search engines use a variety of algorithms to determine relevancy ranking. Some of these include determining whether the search term appears in the META keywords tag, others whether the search term appears in the body text or URL of a web page. Many search engines check for instances of spamdexing and will remove suspect pages from their indexes. Also, people working for a search-engine organization can quickly block the results-listing from entire websites that use spamdexing, perhaps alerted by user complaints of false matches. The rise of spamdexing in the mid-1990s made the leading search engines of the time less useful.

Common spamdexing techniques can be classified into two broad classes: *content spam* (or *term spam*) and *link spam*.

History

The earliest known reference to the term *spamdexing* is by Eric Convey in his article "Porn sneaks way back on Web," The Boston Herald, May 22, 1996, where he said:

The problem arises when site operators load their Web pages with hundreds of extraneous terms so search engines will list them among legitimate addresses. The process is called "spamdexing," a combination of spamming — the Internet term for sending users unsolicited information — and "indexing."

Content spam

These techniques involve altering the logical view that a search engine has over the page's contents. They all aim at variants of the vector space model for information retrieval on text collections.

Keyword stuffing

Keyword stuffing involves the calculated placement of keywords within a page to raise the keyword count, variety, and density of the page. This is useful to make a page appear to be relevant for a web crawler in a way that makes it more likely to be found. Example: A promoter of a Ponzi scheme wants to attract web surfers to a site where he advertises his scam. He places hidden text appropriate for a fan page of a popular music group on his page, hoping that the page will be listed as a fan site and receive many visits from music lovers. Older versions of indexing programs simply counted how often a keyword appeared, and used that to determine relevance levels. Most modern search engines have the ability to analyze a page for keyword stuffing and determine whether the frequency is consistent with other sites created specifically to attract search engine traffic. Also, large webpages are truncated, so that massive dictionary lists cannot be indexed on a single webpage.

Hidden or invisible text

Unrelated hidden text is disguised by making it the same color as the background, using a tiny font size, or hiding it within HTML code such as "no frame" sections, alt attributes, zero-sized DIVs, and "no script" sections. People screening websites for a search-engine company might temporarily or permanently block an entire website for having invisible text on some of its pages. However, hidden text is not always spamdexing: it can also be used to enhance accessibility.

Meta-tag stuffing

This involves repeating keywords in the Meta tags, and using meta keywords that are unrelated to the site's content. This tactic has been ineffective since 2005.

Doorway pages

"Gateway" or doorway pages are low-quality web pages created with very little content but are instead stuffed with very similar keywords and phrases. They are designed to rank highly within the search results, but serve no purpose to visitors looking for information. A doorway page will generally have "click here to enter" on the page.

Scraper sites

Scraper sites sites, are created using various programs designed to "scrape" search-engine results pages or other sources of content and create "content" for a website. The specific presentation of content on these sites is unique, but is merely an amalgamation of content taken from other sources, often without permission. Such websites are generally full of advertising (such as pay-per-click ads), or they redirect the user to other sites. It is even feasible for scraper sites to outrank original websites for their own information and organization names.

Article spinning

Article spinning involves rewriting existing articles, as opposed to merely scraping content from other sites, to avoid penalties imposed by search engines for duplicate content. This process is undertaken by hired writers or automated using a thesaurus database or a neural network.

Link spam

Link spam is defined as links between pages that are present for reasons other than merit. Link spam takes advantage of link-based ranking algorithms, which gives websites higher rankings the more other highly ranked websites link to it. These techniques also aim at influencing other link-based ranking techniques such as the HITS algorithm.

Link-building software

A common form of link spam is the use of link-building software to automate the search engine optimization process.

Link farms

Link farms are tightly-knit communities of pages referencing each other, also known humorously as *mutual admiration societies*

Hidden links

Putting hyperlinks where visitors will not see them to increase link popularity. Highlighted link text can help rank a webpage higher for matching that phrase.

Sybil attack

A Sybil attack is the forging of multiple identities for malicious intent, named after the famous multiple personality disorder patient "Sybil" (Shirley Ardell Mason). A spammer may create multiple web sites at different domain names that all link to each other, such as fake blogs (known as spam blogs).

Spam blogs

Spam blogs,are blogs created solely for commercial promotion and the passage of link authority to target sites. Often these "splogs" are designed in a misleading manner that will give the effect of a legitimate website but upon close inspection will often be written using spinning software or very poorly written and barely readable content. They are similar in nature to link farms.

Page hijacking

Page hijacking is achieved by creating a rogue copy of a popular website which shows contents similar to the original to a web crawler but redirects web surfers to unrelated or malicious websites.

Buying expired domains

Some link spammers monitor DNS records for domains that will expire soon, then buy them when they expire and replace the pages with links to their pages. However Google resets the link data on expired domains. Some of these techniques may be applied for creating a Google bomb, this is, to cooperate with other users to boost the ranking of a particular page for a particular query.

Cookie stuffing

Cookie stuffing involves placing an affiliate tracking cookie on a website visitor's computer without their knowledge, which will then generate revenue for the person doing the cookie stuffing. This not only generates fraudulent affiliate sales, but also has the potential to overwrite other affiliates' cookies, essentially stealing their legitimately earned commissions.

Using world-writable pages

Web sites that can be edited by users can be used by spamdexers to insert links to spam sites if the appropriate anti-spam measures are not taken.

Automated spambots can rapidly make the user-editable portion of a site unusable. Programmers have developed a variety of automated spam prevention techniques to block or at least slow down spambots.

Spam in blogs

Spam in blogs is the placing or solicitation of links randomly on other sites, placing a desired keyword into the hyperlinked text of the inbound link. Guest books, forums, blogs, and any site that accepts visitors' comments are particular targets and are often victims of drive-by spamming where automated software creates nonsense posts with links that are usually irrelevant and unwanted.

Comment spam

Comment spam is a form of link spam that has arisen in web pages that allow dynamic user editing such as blogs and guestbooks. It can be problematic because agents can be written that automatically randomly select a user edited web page.

Referrer log spamming

Referrer spam takes place when a spam perpetrator or facilitator accesses a web page (the *referee*), by following a link from another web page (the *referrer*), so that the referee is given the address of the referrer by the person's Internet browser. Some websites have a referrer log which shows which pages link to that site. By having a robot randomly access many sites enough times, with a message or specific address given as the referrer, that message or Internet address then appears in the referrer log of those sites that have referrer logs. Since some Web search engines base the importance of sites on the number of different sites linking to them, referrer-log spam may increase the search engine rankings of the spammer's sites. Also, site administrators who notice the referrer log entries in their logs may follow the link back to the spammer's referrer page.

Other types of spamdexing

Mirror websites

A mirror site is the hosting of multiple websites with conceptually similar content but using different URLs. Some search engines give a higher rank to results where the keyword searched for appears in the URL.

URL redirection

URL redirection is the taking of the user to another page without his or her intervention, *e.g.*, using META refresh tags, Flash, JavaScript, Java or Server side redirects.

Cloaking

Cloaking refers to any of several means to serve a page to the search-engine spider that is different from that seen by human users. It can be an attempt to mislead search engines regarding the content on a particular web site. Cloaking, however, can also be used to ethically increase accessibility of a site to users with disabilities or provide human users with content that search engines aren't able to process or parse. It is also used to deliver content based on a user's location; Google itself uses IP delivery, a form of cloaking, to deliver results. Another form of cloaking is *code swapping*, *i.e.*, optimizing a page for top ranking and then swapping another page in its place once a top ranking is achieved.

Chapter 5

Spam in Blogs

Spam in blogs (also called simply **blog spam** or **comment spam** is a form of spamdexing. (Note that *blogspam* has another, more common meaning, namely the post of a blogger who creates no-value-added posts to submit them to other sites.) It is done by automatically posting random comments or promoting commercial services to blogs, guestbooks, or other publicly accessible online discussion boards. Any web application that accepts and displays hyperlinks submitted by visitors may be a target.

Adding links that point to the spammer's web site artificially increases the site's search engine ranking. An increased ranking often results in the spammer's commercial site being listed ahead of other sites for certain searches, increasing the number of potential visitors and paying customers.

History

This type of spam originally appeared in internet guestbooks, where spammers repeatedly fill a guestbook with links to their own site and no relevant comment, to increase search engine rankings. If an actual comment is given it is often just "cool page", "nice website", or keywords of the spammed link.

In 2003, spammers began to take advantage of the open nature of comments in the blogging software like Movable Type by repeatedly placing comments to various blog posts that provided nothing more than a link to the spammer's commercial web site. Jay Allen created a free plugin, called MT-BlackList, for the Movable Type weblog tool (versions prior to 3.2) that attempted to alleviate this problem. Many blogging packages now have methods of preventing or reducing the effect of blog spam, although spammers have developed tools to circumvent them. Many spammers use special blog spamming tools like Trackback Submitter to bypass comment spam protection on popular blogging systems like Movable Type, Wordpress, and others.

Possible solutions

Disallowing multiple consecutive submissions

It is rare on a site that a user would reply to their own comment, yet spammers typically will do. Checking that the user's IP address is not replying to a user of the same IP address will significantly reduce flooding. This, however, proves problematic when multiple users, behind the same proxy, wish to comment on the same entry.

Blocking by keyword

Blocking specific words from posts is one of the simplest and most effective ways to reduce spam. Much spam can be blocked simply by banning names of popular pharmaceuticals and casino games.

This is a good long-term solution, because it's not beneficial for spammers to change keywords to "vi@gra" or such, because keywords must be readable and indexed by search engine bots to be effective.

nofollow

Google announced in early 2005 that hyperlinks with `rel="nofollow"` attribute would not be crawled or influence the link target's ranking in the search engine's index. The Yahoo and MSN search engines also respect this tag.

Using `rel="nofollow"` is a much easier solution that makes the improvised techniques above irrelevant. Most weblog software now marks reader-submitted links this way by default (with no option to disable it without code modification). A more sophisticated server software could spare the nofollow for links submitted by trusted users like those registered for a long time, on a whitelist, or with a high karma. Some server software adds `rel="nofollow"` to pages that have been recently edited but omits it from stable pages, under the theory that stable pages will have had offending links removed by human editors.

Some weblog authors object to the use of `rel="nofollow"`, arguing, for example, that

- Link spammers will continue to spam everyone to reach the sites that do not use `rel="nofollow"`
- Link spammers will continue to place links for clicking (by surfers) even if those links are ignored by search engines.
- Google is advocating the use of `rel="nofollow"` in order to reduce the effect of heavy inter-blog linking on page ranking.
- Google is advocating the use of `rel="nofollow"` only to minimize its own filtering efforts and to deflect that this actually had better been called `rel="nopagerank"`.
- Nofollow may reduce the value of legitimate comments

- Nofollow links may still carry value in terms of search engine rankings

Other websites like Slashdot, with high user participation, use improvised nofollow implementations like adding `rel="nofollow"` only for potentially misbehaving users. Potential spammers posting as users can be determined through various heuristics like age of registered account and other factors. Slashdot also uses the poster's karma as a determinant in attaching a nofollow tag to user submitted links.

`rel="nofollow"` has come to be regarded as a microformat.

Validation (reverse Turing test)

A method to block automated spam comments is requiring a validation prior to publishing the contents of the reply form. The goal is to verify that the form is being submitted by a real human being and not by a spam tool and has therefore been described as a reverse Turing test. The test should be of such a nature that a human being can easily pass and an automated tool would most likely fail.

Many forms on websites take advantage of the CAPTCHA technique, displaying a combination of numbers and letters embedded in an image which must be entered literally into the reply form to pass the test. In order to keep out spam tools with built-in text recognition the characters in the images are customarily misaligned, distorted, and noisy. A drawback of many older CAPTCHAs is that passwords are usually case-sensitive while the corresponding images often don't allow a distinction of capital and small letters. This should be taken into account when devising a list of CAPTCHAs. Such systems can also prove problematic to blind people who rely on screen readers. Some more recent systems allow for this by providing an audio version of the characters.

A simple alternative to CAPTCHAs is the validation in the form of a password question, providing a hint to human visitors that the password is the answer to a simple question like "The Earth revolves around the... [Sun]".

One drawback to be taken into consideration is that any validation required in the form of an additional form field may become a nuisance especially to regular posters. Many bloggers and guestbook owners notice a significant decrease in the number of comments once such a validation is in place.

Disallowing links in posts

There is negligible gain from spam that does not contain links, so currently all spam posts contain (an excessive number of) links. It is safe to require passing Turing tests only if post contains links and letting all other posts through. While this is highly effective, spammers do frequently send gibberish posts (such as "ajliabisadf ljibia aeriqoj") to test the spam filter. These gibberish posts will not be labeled as spam. They do the spammer no good, but they still clog up comments sections.

Garbage submissions might however also result from level 0 spambots, which don't parse the attacked HTML form fields first, but send generic POST requests against pages. So it happens that a "content" or "forum_post" POST variable is set and received by the blog or forum software, but the "uri" or other wrong url field name is not accepted and thus not saved as spamlink.

Redirects

Instead of displaying a direct hyperlink submitted by a visitor, a web application could display a link to a script on its own website that redirects to the correct URL. This will not prevent all spam since spammers do not always check for link redirection, but effectively prevents against increasing their PageRank, just as `rel=nofollow`. An added benefit is that the redirection script can count how many people visit external URLs, although it will increase the load on the site.

Redirects should be server-side to avoid accessibility issues related to client-side redirects. This can be done via the .htaccess file in Apache.

Another way of preventing PageRank leakage is to make use of public redirection or dereferral services such as TinyURL. For example,

```
<a href="http://my-own.net/alias_of_target" rel="nofollow" >Link</a>
```

where 'alias_of_target' is the alias of target address.

Note however that this prevents users from being able to view the target of a link before clicking it, thus interfering with their ability to ignore websites they know to be spam. TinyURL now offers a preview feature to help avoiding this situation.

Distributed approaches

This approach is very new to addressing link spam. One of the shortcomings of link spam filters is that most sites receive only one link from each domain which is running a spam campaign. If the spammer varies IP addresses, there is little to no distinguishable pattern left on the vandalized site. The pattern, however, is left across the thousands of sites that were hit quickly with the same links.

A distributed approach, like the free LinkSleeve uses XML-RPC to communicate between the various server applications and the filter server, in this case LinkSleeve. The posted data is stripped of urls and each url is checked against recently submitted urls across the web. If a threshold is exceeded, a "reject" response is returned, thus deleting the comment, message, or posting. Otherwise, an "accept" message is sent.

A more robust distributed approach is Akismet, which uses a similar approach to LinkSleeve but uses API keys to assign trust to nodes and also has wider distribution as a result of being bundled with the 2.0 release of WordPress. They claim over 140,000 blogs contributing to their system. Akismet libraries have been implemented for Java, Python, Ruby, and PHP, but its adoption may be hindered by its commercial use restrictions. In 2008, Six Apart therefore released a beta version of their TypePad AntiSpam software, which is compatible with Akismet but free of the latter's commercial use restrictions.

Project Honey Pot has also begun tracking comment spammers. The Project uses its vast network of thousands of traps installed in over one hundred countries around the world in order to watch what comment spamming web robots are posting to blogs and forums. Data is then published on the top countries for comment spamming, as well as the top keywords and URLs being promoted. The Project's data is then made available to block known comment spammers through . Various plugins have been developed to take advantage of the API.

Application-specific anti-spam methods

Particularly popular software products such as Movable Type have developed their own custom anti-spam measures, as spammers focus more attention on targeting those platforms. Whitelists and blacklists that prevent certain IPs from posting, or that prevent people from posting content that matches certain filters, are common defenses. More advanced access control lists require various forms of validation before users can contribute anything like linkspam.

The goal in every case is to allow good users to continue to add links to their comments, as that is considered by some to be a valuable aspect of any comments section.

RSS feed monitoring

Some allow you to access an RSS feed of recent changes or comments. If you add that to your news reader and set up a smart search for common spam terms (usually viagra and other drug names) you can quickly identify and remove the offending spam.

Response tokens

Another filter available to webmasters is to add a hidden variable to their comment form containing a session token which uniquely identifies the instance of the form. The primary protection afforded by this mechanism is through enforcing a one-to-one correspondence between each request to get the form and each request to submit it. This is impossible to do with IP addresses, since they are shared by users behind a proxy, firewall, or nat (e.g., multiple users sitting in the same internet cafe, library, senior citizens' center, managed care home, club, etc.) and they may change frequently, even between related requests (e.g., AOL and other enterprise-scale proxies, anonymizing services such as Tor). When the form is eventually submitted, the server can use the token to validate the post. If the token is unrecognized the server can send back the form, along with a new token, requiring user resubmission. A duplicate token with duplicate content can safely be silently discarded. Additionally, spammers may not actually load the comments form for an entry; having a unique code for each request inserted into the comment form and verifying it on receipt of the HTTP POST will significantly increase the number of steps required to spam multiple entries.

Given a valid token, the server can then flag as suspicious, for example, postings that use different IP addresses for loading the comment form and posting the comment form, many postings all using the same IP address, or postings that took unusually short or long periods of time to compose. These can then be subjected to additional scrutiny, such as challenging the poster with a captcha, queuing for human review, or outright rejected.

This method is effective against spammers who spoof their IP Address in an attempt to conceal their identities or to appear to be many more distinct users than the number of IP addresses simultaneously under their control, since the token can only be returned if it was received by the spammer in the first place. It has been suggested that flagging posts based on changing IP addresses is effective against spammers abusing the distributed anonymous proxy Tor.

Ajax

Some blog software such as Typo allow the blog administrator to allow only comments submitted via Ajax XMLHttpRequests, and discard regular form POST requests. This causes accessibility problems typical to Ajax-only applications.

Although this technique prevents spam so far, it is a form of security by obscurity and will probably be defeated if it becomes popular enough.

Chapter 6

Anti-Spam Techniques

To prevent e-mail spam, both end users and administrators of e-mail systems use various **anti-spam techniques**. Some of these techniques have been embedded in products, services and software to ease the burden on users and administrators. No one technique is a complete solution to the spam problem, and each has trade-offs between incorrectly rejecting legitimate e-mail vs. not rejecting all spam, and the associated costs in time and effort.

Anti-spam techniques can be broken into four broad categories: those that require actions by individuals, those that can be automated by e-mail administrators, those that can be automated by e-mail senders and those employed by researchers and law enforcement officials.

Detecting spam

Checking words: false positives

People tend to be much less bothered by spam slipping through filters into their mail box (false negatives), than having desired e-mail ("ham") blocked (false positives). Trying to balance false negatives (missed spams) vs false positives (rejecting good e-mail) is critical for a successful anti-spam system. Some systems let individual users have some control over this balance by setting "spam score" limits, etc. Most techniques have both kinds of serious errors, to varying degrees. So, for example, anti-spam systems may use techniques that have a high false negative rate (miss a lot of spam), in order to reduce the number of false positives (rejecting good e-mail).

Detecting spam based on the content of the e-mail, either by detecting keywords such as "viagra" or by statistical means, is very popular. Such methods can be very accurate when they are correctly tuned to the types of legitimate email that an individual gets, but they can also make mistakes such as detecting the keyword "cialis" in the word "specialist". The content also doesn't determine whether the email was either unsolicited or bulk, the two key features of spam. So, if a friend sends you a joke that mentions "viagra", content filters can easily mark it as being spam even though it is neither unsolicited nor sent in bulk.

Lists of sites

The most popular DNSBLs (DNS Blacklists) are lists of IP addresses of known spammers, known open relays, known proxy servers, compromised “zombie” spammers, as well as hosts on the internet that shouldn’t be sending external emails, such as the end-user address space of a consumer ISP. These are known as “Dial Up Lists”, from the time when end users had to dial up to the internet with a modem and a phone line.

Spamtraps are often email addresses that were never valid or have been invalid for a long time that are used to collect spam. An effective spamtrap is not announced and is only found by dictionary attacks or by pulling addresses off hidden webpages. For a spamtrap to remain effective the address must never be given to anyone. Some black lists, such as spamcop, use spamtraps to catch spammers and blacklist them.

Enforcing technical requirements of the Simple Mail Transfer Protocol (SMTP) can be used to block mail coming from systems that are not compliant with the RFC standards. A lot of spammers use poorly written software or are unable to comply with the standards because they do not have legitimate control of the computer sending spam (zombie computer). So by setting restrictions on the mail transfer agent (MTA) a mail administrator can reduce spam significantly, such as by enforcing the correct fall back of Mail eXchange (MX) records in the Domain Name System, or the correct handling of delays (Teergrube).

End-user techniques

There are a number of techniques that individuals can use to restrict the availability of their e-mail addresses, reducing or preventing their attractiveness to spam.

Discretion

Sharing an email address only among a limited group of correspondents is one way to limit spam. This method relies on the discretion of all members of the group, as disclosing email addresses outside the group circumvents the trust relationship of the group. For this reason, forwarding messages to recipients who don't know one another should be avoided. When it is absolutely necessary to forward messages to recipients who don't know one another, it is good practice to list the recipient names all after "bcc:" instead of after "to:". This practice avoids the scenario where unscrupulous recipients might compile a list of email addresses for spamming purposes. This practice also reduces the risk of the address being distributed by computers affected with email address harvesting malware. However, once the privacy of the email address is lost by divulgence, it cannot likely be regained.

Address munging

Posting anonymously, or with a fake name and address, is one way to avoid e-mail address harvesting, but users should ensure that the fake address is not valid. Users who want to receive legitimate email regarding their posts or Web sites can alter their addresses so humans can figure out but spammers cannot. For instance, `joe@example.com` might post as `joeNOS@PAM.invalid.example.com`. Address munging, however, can cause legitimate replies to be lost. If it's not the user's valid address, it has to be truly invalid, otherwise someone or some server will still get the spam for it. Other ways use transparent address munging to avoid this by allowing users to see the actual address but obfuscate it from automated email harvesters with methods such as displaying all or part of the e-mail address on a web page as an image, a text logo shrunken to normal size using in-line CSS, or as jumbled text with the order of characters restored using CSS.

Avoid responding to spam

Spammers often regard responses to their messages—even responses like "Don't spam me"—as confirmation that an email address is valid. Likewise, many spam messages contain Web links or addresses which the user is directed to follow to be removed from the spammer's mailing list. In several cases, spam-fighters have tested these links, confirming they do not lead to the recipient address's removal—if anything, they lead to more spam.

Sender addresses are often forged in spam messages, including using the recipient's own address as the forged sender address, so that responding to spam may result in failed deliveries or may reach innocent e-mail users whose addresses have been abused.

In Usenet, it is widely considered even more important to avoid responding to spam. Many ISPs have software that seek and destroy duplicate messages. Someone may see a spam and respond to it before it is cancelled by their server, which can have the effect of reposting the spam for them; since it is not a duplicate, the reposted copy will last longer.

Contact forms

Contact forms allow users to send email by filling out forms in a web browser. The web server takes the form data, forwarding it to an email address. Users never see the email address. Such forms, however, are sometimes inconvenient to users, as they are not able to use their preferred e-mail client, risk entering a faulty reply address, and are typically not notified about delivery problems. Further, contact forms have the drawback that they require a website that supports server side scripts. Finally, if the software used to run the contact forms is badly designed, it can become a spam tool in its own right. Additionally, some spammers have begun to send spam using the contact form.

Disable HTML in e-mail

Many modern mail programs incorporate Web browser functionality, such as the display of HTML, URLs, and images. This can easily expose the user to offensive images in spam. In addition, spam written in HTML can contain web bugs which allows spammers to see that the e-mail address is valid and that the message has not been caught in spam filters. JavaScript programs can be used to direct the user's Web browser to an advertised page, or to make the spam message difficult to close or delete. Spam messages have contained attacks upon security vulnerabilities in the HTML renderer, using these holes to install spyware. (Some computer viruses are borne by the same mechanisms.)

Mail clients which do not automatically download and display HTML, images or attachments, have fewer risks, as do clients who have been configured to not display these by default.

Disposable e-mail addresses

An email user may sometimes need to give an address to a site without complete assurance that the site owner will not use it for sending spam. One way to mitigate the risk is to provide a *disposable* email address—a temporary address which the user can disable or abandon which forwards email to a real account. A number of services provide disposable address forwarding. Addresses can be manually disabled, can expire after a given time interval, or can expire after a certain number of messages have been forwarded. Site owners that fail to keep addresses they have gathered confidential have found themselves in legal jeopardy due to the ability of disposable email address users to trace which website passed on their email without permission.

Ham passwords

Systems that use ham passwords ask unrecognised senders to include in their email a password that demonstrates that the email message is a "ham" (not spam) message. Typically the email address and ham password would be described on a web page, and the ham password would be included in the "subject" line of an email address. Ham passwords are often combined with filtering systems, to counter the risk that a filtering system will accidentally identify a ham message as a spam message.

The "plus addressing" technique appends a password to the "username" part of the email address.

Reporting spam

Tracking down a spammer's ISP and reporting the offense can lead to the spammer's service being terminated. Unfortunately, it can be difficult to track down the spammer—and while there are some online tools to assist, they are not always accurate. Occasionally, spammers employ their own netblocks. In this case, the abuse contact for the netblock can be the spammer itself and can confirm your address.

Examples of these online tools are SpamCop and Network Abuse Clearinghouse. They provide automated or semi-automated means to report spam to ISPs. Some spam-fighters regard them as inaccurate compared to what an expert in the email system can do; however, most email users are not experts.

A useful free tool that may be used in the reporting of spam is also available (Complainterator). The Complainterator will send an automatically generated complaint to the registrar of the spamming domain and the registrar of its name servers.

Historically, reporting spam in this way has not seriously abated spam, since the spammers simply move their operation to another URL, ISP or network of IP addresses.

Consumers may also forward "unwanted or deceptive spam" to an email address (`spam@uce.gov`) maintained by the FTC. The database collected is used to prosecute perpetrators of scam or deceptive advertising.

An alternative to contacting ISPs is to contact the registrar of a domain name that has used in spam e-mail. Registrars, as ICANN-accredited administrative organizations, are obliged to uphold certain rules and regulations, and have the resources necessary for dealing with abuse complaints.

Responding to spam

Some advocate responding aggressively to spam—in other words, "spamming the spammer".

The basic idea is to make spamming less attractive to the spammer, by increasing the spammer's overhead. There are several ways to reach a spammer, but besides the caveats mentioned above, it may lead to retaliations by the spammer.

1. Replying directly to the spammer's email address

   Just clicking "reply" will not work in the vast majority of cases, since most of the sender addresses are forged or made up. In some cases, however, spammers do provide valid addresses, as in the case of Nigerian scams.

2. Targeting the computers used to send out spam

   In 2005, IBM announced a service to bounce spam directly to the computers that send out spam. Because the IP addresses are identified in the headers of every message, it would be possible to target those computers directly, sidestepping the problem of forged email addresses. In most cases, however, those computers do not belong to the real spammer, but to unsuspecting users with unsecured or outdated systems, hijacked through malware and controlled at distance by the spammer; these are known as zombie computers. However, in most legal jurisdictions, ignorance is no defense, and many victims of spam regard the owners of zombie computers as willfully compliant accomplices of spammers.

3. Leaving messages on the spamvertised site

   Spammers selling their wares need a tangible point of contact so that customers can reach them. Sometimes it is a telephone number, but most often is a web site containing web forms through which customers can fill out orders or inquiries, or even "unsubscribe" requests. Since positive response to spam is probably much less than 1/10,000, if just a tiny percentage of users visit spam sites just to leave negative messages, the negative messages could easily outnumber positive ones, incurring costs for spammers to sort them out, not mentioning the cost in bandwidth. An automated system, designed to respond in just such a way, was Blue Frog. Unfortunately, in doing so, you risk arousing the ire of criminals who may respond with threats or 'target' your address with even more spam.

Automated techniques for e-mail administrators

There are a number of appliances, services and software systems that e-mail administrators can use to reduce the load of spam on their systems and mailboxes. Some of these depend upon rejecting email from Internet sites known or likely to send spam. Other more advanced techniques analyze message patterns in real time to detect spam like behavior and then compares it to global databases of spam. Those methods are capable of detecting spam in real time even when there is no content (common to image based spam) and in any language. Another method relies on automatically analyzing the content of email messages and weeding out those which resemble spam. These three approaches are sometimes termed *blocking*, *pattern detection*, and *filtering.*

There is an increasing trend of integration of anti-spam techniques into MTAs whereby the mail systems themselves also perform various measures that are generally referred to as filtering, ultimately resulting in spam messages being rejected before delivery (or *blocked*).

Many filtering systems take advantage of machine learning techniques, which improve their accuracy over manual methods. However, some people find filtering intrusive to privacy, and many e-mail administrators prefer blocking to deny access to their systems from sites tolerant of spammers.

Authentication and reputation

A number of systems have been proposed to allow acceptance of email from servers which have authenticated in some fashion as senders of only legitimate email. Many of these systems use the DNS, as do DNSBLs; but rather than being used to list nonconformant sites, the DNS is used to list sites authorized to send email, and (sometimes) to determine the reputation of those sites. Other methods of identifying ham (non-spam email) and spam are still used.

Authentication systems cannot detect whether a message is spam. Rather, they allow a site to express trust that an authenticated site will not send spam. Thus, a recipient site may choose to skip expensive spam-filtering methods for messages from authenticated sites.

Challenge/response systems

Another method which may be used by internet service providers, by specialized services or enterprises to combat spam is to require unknown senders to pass various tests before their messages are delivered. These strategies are termed **challenge/response systems** or **C/R**. Some view their use as being as bad as spam since they place the burden of spam fighting on legitimate email senders—who it should be noted will often indeed give up at the slightest hindrance. A new implementation of this is done in Channel email.

Checksum-based filtering

Checksum-based filter exploits the fact that the messages are sent in bulk, that is that they will be identical with small variations. Checksum-based filters strip out everything that might vary between messages, reduce what remains to a checksum, and look that checksum up in a database which collects the checksums of messages that email recipients consider to be spam (some people have a button on their email client which they can click to nominate a message as being spam); if the checksum is in the database, the message is likely to be spam.

The advantage of this type of filtering is that it lets ordinary users help identify spam, and not just administrators, thus vastly increasing the pool of spam fighters. The disadvantage is that spammers can insert unique invisible gibberish—known as *hashbusters*—into the middle of each of their messages, thus making each message unique and having a different checksum. This leads to an arms race between the developers of the checksum software and the developers of the spam-generating software.

Checksum based filtering methods include:

- Distributed Checksum Clearinghouse
- Vipul's Razor

Country-based filtering

Some e-mail servers expect to never communicate with particular countries from which they receive a great deal of spam. Therefore, they use country-based filtering - a technique that blocks e-mail from certain countries. This technique is based on country of origin determined by the sender's IP address rather than any trait of the sender.

DNS-based blacklists

DNS-based Blacklists, or DNSBLs, are used for heuristic filtering and blocking. A site publishes lists (typically of IP addresses) via the DNS, in such a way that mail servers can easily be set to reject mail from those sources. There are literally scores of DNSBLs, each of which reflects different policies: some list sites known to emit spam; others list open mail relays or proxies; others list ISPs known to support spam.

Other DNS-based anti-spam systems list known good ("white") or bad ("black") IPs domains or URLs, including RHSBLs and URIBLs.

Enforcing RFC standards

Analysis of an email's conformation to RFC standards for the Simple Mail Transfer Protocol (SMTP) can be used to judge the likelihood of the message being spam. A lot of spammers use poorly written software or are unable to comply with the standards because they do not have legitimate control of the computer they are using to send spam (zombie computer). By setting limits on the deviation from RFC standards that the MTA will accept, a mail administrator can reduce spam significantly.

Greeting delay

A greeting delay is a deliberate pause introduced by an SMTP server before it sends the SMTP greeting banner to the client. The client is required to wait until it has received this banner before it sends any data to the server. (per RFC 5321 3.1). Many spam-sending applications do not wait to receive this banner, and instead start sending data as soon as the TCP connection is established. The server can detect this, and drop the connection.

There are some legitimate sites that play "fast and loose" with the SMTP specifications, and may be caught by this mechanism. It also has a tendency to interact badly with sites that perform callback verification, as common callback verification systems have timeouts that are much shorter than those mandated by RFC 5321 4.5.3.2.

Greylisting

The SMTP protocol allows for temporary rejection of incoming messages. Greylisting is the technique to temporarily reject messages from unknown sender mail servers. A temporary rejection is designated with a 4xx error code that is recognized by all normal MTAs, which then proceed to retry delivery later.

Greylisting is based on the premise that spammers and spambots will not retry their messages but instead will move on to the next message and next address in their list. Since a retry attempt means the message and state of the process must be stored, it inherently increases the cost incurred by the spammer. The assumption is that, for the spammer, it's a better use of resources to try a new address than waste time re-sending to an address that's already exhibited a problem. For a legitimate message this delay is not an issue since retrying is a standard component of any legitimate sender's server.

The downside of greylisting is that all legitimate messages from first time senders will experience a delay in delivery, with the delay period before a new message is accepted from an unknown sender usually being configurable in the software. There also exists the possibility that some legitimate messages won't be delivered, which can happen if a poorly configured (but legitimate) mail server interprets the temporary rejection as a permanent rejection and sends a bounce message to the original sender, instead of trying to resend the message later, as it should.

HELO/EHLO checking

For example, some spamware can be detected by a number of simple checks confirming compliance with standard addressing and MTA operation. RFC 5321 section 4.1.4 says that "An SMTP server MAY verify that the domain name argument in the EHLO command actually corresponds to the IP address of the client. However, if the verification fails, the server MUST NOT refuse to accept a message on that basis.", so to be in compliance with the RFCs, rejecting connections must be based on additional information/policies.

- Refusing connections from hosts that give an invalid HELO - for example, a HELO that is not an FQDN or is an IP address not surrounded by square brackets

```
Invalid HELO localhost
Invalid HELO 127.0.0.1
Valid HELO domain.tld
Valid HELO [127.0.0.1]
```

- Refusing connections from hosts that give an obviously fraudulent HELO

```
Fraudulent HELO friend
Fraudulent HELO -232975332
```

- Refusing to accept email claiming to be from a hosted domain when the sending host has not authenticated
- Refusing to accept email whose HELO/EHLO argument does not resolve in DNS. Unfortunately, some email system administrators ignore section 2.3.5 of RFC 5321 and administer the MTA to use a nonresolvable argument to the HELO/EHLO command.

Invalid pipelining

The SMTP protocol can allow several SMTP commands to be placed in one network packet and "pipelined". For example, if an e-mail is sent with a CC: header, several SMTP "RCPT TO" commands might be placed in a single packet instead of one packet per "RCPT TO" command. The SMTP protocol, however, requires that errors be checked and everything is synchronized at certain points. Many spammers will send everything in a single packet since they do not care about errors and it is more efficient. Some MTAs will detect this invalid pipelining and reject e-mail sent this way.

Nolisting

The SMTP protocol requires that email servers for any given domain be provided in a prioritized list (namely, MX records), and further specifies mandatory error-handling behavior when servers in that list cannot be contacted. Nolisting is a technique of purposely creating unreachable MX records, so that only senders who have implemented this error-handling behavior can successfully deliver mail.

Quit detection

The SMTP protocol requires connections to be closed with a QUIT command. (RFC 5321 section 4.1.4) Many spammers skip this step because their spam has already been sent and taking the time to properly close the connection takes time and bandwidth. Some MTAs like Exim are capable of detecting whether or not the connection is closed with the quit command and can track patterns of use for the purpose of building DNSBLs.

Honeypots

Another approach is simply an imitation MTA which gives the appearance of being an open mail relay, or an imitation TCP/IP proxy server which gives the appearance of being an open proxy. Spammers who probe systems for open relays/proxies will find such a host and attempt to send mail through it, wasting their time and resources and potentially revealing information about themselves and the origin of the spam they're sending to the entity that operates the honeypot. Such a system may simply discard the spam attempts, submit them to DNSBLs, or store them for analysis.

Hybrid filtering

Hybrid filtering, such as is implemented in the open source programs SpamAssassin and Policyd-weight uses some or all of the various tests for spam, and assigns a numerical score to each test. Each message is scanned for these patterns, and the applicable scores tallied up. If the total is above a fixed value, the message is rejected or flagged as spam. By ensuring that no single spam test by itself can flag a message as spam, the false positive rate can be greatly reduced.

Outbound spam protection

Outbound spam protection involves scanning email traffic as it exits a network, identifying spam messages and then taking an action such as blocking the message or shutting off the source of the traffic. Outbound spam protection can be implemented on a network-wide level (using policy-based routing or similar techniques to route SMTP messages to a filtering service). Or, it can be implemented within a standard SMTP gateway. While the primary economic impact of spam is on spam recipients, sending networks also experience financial costs, such as wasted bandwidth, and the risk of having IP addresses blocked by receiving networks.

The advantage of outbound spam protection is that it stops spam before it leaves the sending network, protecting receiving networks globally from the damage and costs that would otherwise be caused by the spam. Further it lets system administrators track down spam sources on the network and remediate them – for example, providing free anti-virus tools to customers whose machines have become infected with a virus or are participating in a botnet. Given an appropriately designed spam filtering algorithm, outbound spam filtering can be implemented with a near zero false positive rate, which keeps customer related issues with blocked legitimate email down to a minimum.

There are several commercial software vendors who offer outbound spam protection products, including MailChannels and Commtouch.

Pattern detection

Pattern detection, is an approach to stop spam in real time before it gets to the end user. This technology monitors a large database of messages worldwide to detect spam patterns. Many spam messages have no content or may contain attachments which this method of detection can catch. Pioneered by Commtouch, a developer of anti-spam software, their Recurrent Pattern Detection (RPD) software can be integrated into other appliances and applications. This method is more automated than most because the service provider maintains the comparative spam database instead of the system administrator.

PTR/reverse DNS checks

The PTR DNS records in the reverse DNS can be used for a number of things, including:

- Most email mail transfer agents (mail servers) use a forward-confirmed reverse DNS (FCrDNS) verification and if there is a valid domain name, put it into the "Received:" trace header field.
- Some email mail transfer agents will perform FCrDNS verification on the domain name given in the SMTP HELO and EHLO commands.
- To check the domain names in the rDNS to see if they are likely from dial-up users, dynamically assigned addresses, or home-based broadband customers. Since the vast majority, but by no means all, of email that originates from these computers is spam, many mail servers also refuse email with missing or "generic" rDNS names.
- A Forward Confirmed reverse DNS verification can create a form of authentication that there is a valid relationship between the owner of a domain name and the owner of the network that has been given an IP address. While reliant on the DNS infrastructure, which has known vulnerabilities, this authentication is strong enough that it can be used for whitelisting purposes

because spammers and phishers cannot usually bypass this verification when they use zombie computers to forge the domains.

Rule-based filtering

Content filtering techniques rely on the specification of lists of words or regular expressions disallowed in mail messages. Thus, if a site receives spam advertising "herbal Viagra", the administrator might place this phrase in the filter configuration. The mail server would then reject any message containing the phrase.

Header filtering is the means of inspecting the header of the email, the part of the message that contains information about the origin, destination and content of the message. Spammers will often spoof fields in the header in order to hide their identity, or to try to make the email look more legitimate than it is; many of these spoofing methods can be detected. Also, a violation of the RFC 5322 standard on how the header is to be formed can serve as a basis for rejecting the message.

Disadvantages of filtering are threefold: First, filtering can be time-consuming to maintain. Second, it is prone to false positives. Third, these false positives are not equally distributed: since content filtering is prone to reject legitimate messages on topics related to products frequently advertised in spam. A system administrator who attempts to reject spam messages which advertise mortgage refinancing, credit or debt may inadvertently block legitimate e-mail on the same subject.

Spammers frequently change the phrases and spellings they use. This can mean more work for the administrator. However, it also has some advantages for the spam fighter. If the spammer starts spelling "Viagra" as "V1agra" or "Via_gra", it makes it harder for the spammer's intended audience to read their messages. If they try to trip up the phrase detector, by, for example, inserting an invisible-to-the-user HTML comment in the middle of a word ("Via<!---->gra"), this sleight of hand is itself easily detectable, and is a good indication that the message is spam. And if they send spam that consists entirely of images, so that anti-spam software can't analyze the words and phrases in the message, the fact that there *is* no readable text in the body can be detected, making that message a higher risk of being spam.

Content filtering can also be implemented to examine the URLs present (i.e. spamvertising) in an email message. This form of content filtering is much harder to disguise as the URLs must resolve to a valid domain name. Extracting a list of such links and comparing them to published sources of spamvertised domains is a simple and reliable way to eliminate a large percentage of spam via content analysis.

Sender-supported whitelists and tags

There are a small number of organizations which offer IP whitelisting and/or licensed tags that can be placed in email (for a fee) to assure recipients' systems that the messages thus tagged are not spam. This system relies on legal enforcement of the tag. The intent is for email administrators to whitelist messages bearing the licensed tag.

A potential difficulty with such systems is that the licensing organization makes its money by licensing more senders to use the tag—not by strictly enforcing the rules upon licensees. A concern exists that senders whose messages are more likely to be considered spam would accrue a greater benefit by using such a tag. The concern is that these factors form a perverse incentive for licensing organizations to be lenient with licensees who have offended. However, the value of a license would drop if it was not strictly enforced, and financial gains due to enforcement of a license itself can provide an additional incentive for strict enforcement.

SMTP callback verification

Since a large percentage of spam has forged and invalid sender ("from") addresses, some spam can be detected by checking that this "from" address is valid. A mail server can try to verify the sender address by making an SMTP connection back to the mail exchanger for the address, as if it was creating a bounce, but stopping just before any e-mail is sent.

Callback verification can be compliant with SMTP RFCs, but it has various drawbacks. Since nearly all spam has forged return addresses, nearly all callbacks are to innocent third party mail servers that are unrelated to the spam. At the same time, there will be numerous false negatives due to spammers abusing real addresses and some false positives.

SMTP proxy

SMTP proxies allow combating spam in real time, combining sender's behavior controls, providing legitimate users immediate feedback, eliminating a need for quarantine.

Spamtrapping

Spamtrapping is the seeding of an email address so that spammers can find it, but normal users can not. If the email address is used then the sender must be a spammer and they are black listed.

As an example, consider the email address "spamtrap@example.org". If this email address were placed in the source HTML of our web site in a way that it isn't displayed on the web page, normal humans would not see it. Spammers, on the other hand, use web page scrapers and bots to harvest email addresses from HTML source code so they would find this address.

When the spammer sends mail with the destination address of "spamtrap@example.org" the SpamTrap knows this is highly likely to be a spammer and can take appropriate action.

Statistical content filtering

Statistical (or Bayesian) filtering once set up, requires no administrative maintenance per se: instead, users mark messages as spam or nonspam and the filtering software learns from these judgements. Thus, a statistical filter does not reflect the software author's or administrator's biases as to content, but rather the *user's* biases. For example, a biochemist who is researching Viagra won't have messages containing the word "Viagra" automatically flagged as spam, because "Viagra" will show up often in his or her legitimate messages. Still, *spam* emails containing the word "Viagra" do get filtered because the content of the rest of the spam messages differs significantly from the content of legitimate messages. A statistical filter can also respond quickly to changes in spam content, without administrative intervention, as long as users consistently designate false negative messages as spam when received in their email. Statistical filters can also look at message headers, thereby considering not just the content but also peculiarities of the transport mechanism of the email.

Typical statistical filtering uses single words in the calculations to decide if a message should be classified as spam or not. A more powerful calculation can be made using groups of two or more words taken together. Then random "noise" words can not be used as successfully to fool the filter.

Software programs that implement statistical filtering include Bogofilter, DSPAM, SpamBayes, ASSP, the e-mail programs Mozilla and Mozilla Thunderbird, Mailwasher, and later revisions of SpamAssassin. Another interesting project is CRM114 which hashes phrases and does bayesian classification on the phrases.

There is also the free mail filter POPFile, which sorts mail in as many categories as the user wants (family, friends, co-worker, spam, whatever) with Bayesian filtering.

Tarpits

A *tarpit* is any server software which intentionally responds pathologically slowly to client commands. By running a tarpit which treats acceptable mail normally and known spam slowly or which appears to be an open mail relay, a site can slow down the rate at which spammers can inject messages into the mail facility. Many systems will simply disconnect if the server doesn't respond quickly, which will eliminate the spam. However, a few legitimate e-mail systems will also not deal correctly with these delays.

Automated techniques for e-mail senders

There are a variety of techniques that e-mail senders use to try to make sure that they do not send spam. Failure to control the amount of spam sent, as judged by e-mail receivers, can often cause even legitimate email to be blocked and for the sender to be put on DNSBLs.

Background checks on new users and customers

Since spammer's accounts are frequently disabled due to violations of abuse policies, they are constantly trying to create new accounts. Due to the damage done to an ISP's reputation when it is the source of spam, many ISPs and web email providers use CAPTCHAs on new accounts to verify that it is a real human registering the account, and not an automated spamming system. They can also verify that credit cards are not stolen before accepting new customers, check the Spamhaus Project ROKSO list, and do other background checks.

Confirmed opt-in for mailing lists

One difficulty in implementing opt-in mailing lists is that many means of gathering user email addresses remain susceptible to forgery. For instance, if a company puts up a Web form to allow users to subscribe to a mailing list about its products, a malicious person can enter other people's email addresses — to harass them, or to make the company appear to be spamming. (To most anti-spammers, if the company sends e-mail to these forgery victims, it *is* spamming, albeit inadvertently.)

To prevent this abuse, MAPS and other anti-spam organizations encourage that all mailing lists use **confirmed opt-in** (also known as *verified opt-in* or *double opt-in*). That is, whenever an email address is presented for subscription to the list, the list software should send a confirmation message to that address. The confirmation message contains no advertising content, so it is not construed to be spam itself — and the address is not added to the live mail list unless the recipient responds to the confirmation message.

All modern mailing list management programs (such as GNU Mailman, LISTSERV, Majordomo, and qmail's ezmlm) support confirmed opt-in by default.

Egress spam filtering

E-mail senders can do the same type of anti-spam checks on e-mail coming from their users and customers as can be done for e-mail coming from the rest of the Internet.

Limit e-mail backscatter

If any sort of bounce message or anti-virus warning gets sent to a forged email address, the result will be backscatter.

Problems with sending challenges to forged e-mail addresses can be greatly reduced by not creating a new message that contains the challenge. Instead, the challenge can be placed in the Bounce message when the receiving mail system gives a rejection-code during the SMTP session. When the receiving mail system rejects an e-mail this way, it is the sending system that actually creates the bounce message. As a result, the bounce message will almost always be sent to the real sender, and it will be in a format and language that the sender will usually recognize.

Port 25 blocking

Firewalls and routers can be programmed to not allow SMTP traffic (TCP port 25) from machines on the network that are not supposed to run Mail Transfer Agents or send e-mail. This practice is somewhat controversial when ISPs block home users, especially if the ISPs do not allow the blocking to be turned off upon request. E-mail can still be sent from these computers to designated smart hosts via port 25 and to other smart hosts via the e-mail submission port 587.

Port 25 interception

Network address translation can be used to intercept all port 25 (SMTP) traffic and direct it to a mail server that enforces rate limiting and egress spam filtering. This is commonly done in hotels, but it can cause e-mail privacy problems, as well making it impossible to use STARTTLS and SMTP-AUTH if the port 587 submission port isn't used.

Rate limiting

Machines that suddenly start sending lots of e-mail may well have become zombie computers. By limiting the rate that e-mail can be sent around what is typical for the computer in question, legitimate e-mail can still be sent, but large spam runs can be slowed down until manual investigation can be done.

Spam report feedback loops

By monitoring spam reports from places such as spamcop, AOL's feedback loop, and Network Abuse Clearinghouse, the domain's abuse@ mailbox, etc., ISPs can often learn of problems before they seriously damage the ISP's reputation and have their mail servers blacklisted.

FROM field control

Both malicious software and human spam senders often use forged FROM addresses when sending spam messages. Control may be enforced on SMTP servers to ensure senders can only use their correct email address in the FROM field of outgoing messages. In an email users database each user has a record with an e-mail address. The SMTP server must check if the email address in the FROM field of an outgoing message is the same address that belongs to the user's credentials, supplied for SMTP authentication. If the FROM field is forged, an SMTP error will be returned to the email client (e.g. "You do not own the email address you are trying to send from").

Strong AUP and TOS agreements

Most ISPs and webmail providers have either an Acceptable Use Policy (AUP) or a Terms of Service (TOS) agreement that discourages spammers from using their system and allows the spammer to be terminated quickly for violations.

Techniques for researchers & law enforcement

Increasingly, anti-spam efforts have led to co-ordination between law enforcement, researchers, major consumer financial service companies and Internet service providers in monitoring and tracking e-mail spam, identity theft and phishing activities and gathering evidence for criminal cases.

Legislation and enforcement

Appropriate legislation and enforcement can have a significant impact on spamming activity.

The penalty provisions of the Australian Spam Act 2003 dropped Australia's ranking in the list of spam-relaying countries for email spam from tenth to twenty-eighth.

Legislation that provides mandates that bulk emailers must follow makes compliant spam easier to identify and filter out.

Analysis of spamvertisements

Analysis of sites being spamvertised by a given piece of spam often leads to questionable registrations of Internet domain names. Since registrars are required to maintain trustworthy WHOIS databases, digging into the registration details and complaining at the proper locations often results in site shutdowns. Uncoordinated activity may not be effective, given today's volume of spam and the rate at which criminal organizations register new domains. However, a coordinated effort, implemented with adequate infrastructure, can obtain good results.

New solutions and ongoing research

Several approaches have been proposed to improve the e-mail system.

Cost-based systems

Since spamming is facilitated by the fact that large volumes of email are very inexpensive to send, one proposed set of solutions would require that senders pay some cost in order to send email, making it prohibitively expensive for spammers. Anti-spam activist Daniel Balsam attempts to make spamming less profitable by bringing lawsuits against spammers.

Other techniques

There are a number of proposals for sideband protocols that will assist SMTP operation. The Anti-Spam Research Group (ASRG) of the Internet Research Task Force (IRTF) is working on a number of email authentication and other proposals for providing simple source authentication that is flexible, lightweight, and scalable. Recent Internet

Engineering Task Force (IETF) activities include MARID (2004) leading to two approved IETF experiments in 2005, and DomainKeys Identified Mail in 2006.

Channel email is a new proposal for sending email that attempts to distribute anti-spam activities by forcing verification (probably using bounce messages so back-scatter doesn't occur) when the first email is sent for new contacts.

Chapter 7

E-mail Address Harvesting

E-mail harvesting is the process of obtaining lists of e-mail addresses using various methods for use in bulk e-mail or other purposes usually grouped as spam.

Methods

The simplest method involves spammers purchasing or trading lists of e-mail addresses from other spammers.

Another common method is the use of special software known as "harvesting bots" or "harvesters", which spider Web pages, postings on Usenet, mailing list archives, internet forums and other online sources to obtain e-mail addresses from public data.

Spammers may also use a form of dictionary attack in order to harvest e-mail addresses, known as a directory harvest attack, where valid e-mail addresses at a specific domain are found by guessing e-mail address using common usernames in email addresses at that domain.

Another method of e-mail address harvesting is to offer a product or service free of charge as long as the user provides a valid e-mail address, and then use the addresses collected from users as spam targets. Common products and services offered are jokes of the day, daily bible quotes, news or stock alerts, free merchandise, or even registered sex offender alerts for one's area. Another technique was used in late 2007 by the company iDate, which used e-mail harvesting directed at subscribers to the Quechup website to spam the victim's friends and contacts.

Spam differs from other forms of direct marketing in many ways, one of them being that it costs little more to send to a larger number of recipients than a smaller number. For this reason, there is little pressure upon spammers to limit the number of addresses targeted in a spam run, or to restrict it to persons likely to be interested. One consequence of this fact is that many people receive spam written in languages they cannot read — a good deal of spam sent to English-speaking recipients is in Chinese or Korean, for instance. Likewise, lists of addresses sold for use in spam frequently contain malformed addresses, duplicate addresses, and addresses of role accounts such as `postmaster`.

Spammers may harvest e-mail addresses from a number of sources. A popular method uses e-mail addresses which their owners have published for other purposes. Usenet posts, especially those in archives such as Google Groups, frequently yield addresses. Simply searching the Web for pages with addresses — such as corporate staff directories or membership lists of professional societies — using spambots can yield thousands of addresses, most of them deliverable. Spammers have also subscribed to discussion mailing lists for the purpose of gathering the addresses of posters. The DNS and WHOIS systems require the publication of technical contact information for all Internet domains; spammers have illegally trawled these resources for email addresses. Many spammers use programs called web spiders to find email addresses on web pages. Usenet article message-IDs often look enough like email addresses that they are harvested as well.

Spammer viruses may include a function which scans the victimized computer's disk drives (and possibly its network interfaces) for email addresses. These scanners discover email addresses which have never been exposed on the Web or in Whois. A compromised computer located on a shared network segment may capture email addresses from traffic addressed to its network neighbors. The harvested addresses are then returned to the spammer through the bot-net created by the virus.

A recent, controversial tactic, called *"e-pending"*, involves the *appending* of *e-mail* addresses to direct-marketing databases. Direct marketers normally obtain lists of prospects from sources such as magazine subscriptions and customer lists. By searching the Web and other resources for e-mail addresses corresponding to the names and street addresses in their records, direct marketers can send targeted spam e-mail. However, as with most spammer "targeting", this is imprecise; users have reported, for instance, receiving solicitations to mortgage their house at a specific street address — with the address being clearly a business address including mail stop and office number.

Spammers sometimes use various means to confirm addresses as deliverable. For instance, including a hidden Web bug in a spam message written in HTML may cause the recipient's mail client to transmit the recipient's address, or any other unique key, to the spammer's Web site. Users can defend against such abuses by turning off their mail program's option to display images, or by reading email as plain-text rather than formatted.

Likewise, spammers sometimes operate Web pages which purport to remove submitted addresses from spam lists. In several cases, these have been found to subscribe the entered addresses to receive more spam.

When persons fill out a form it is often sold to a spammer using a web service or http post to transfer the data. This is immediate and will drop the email in various spammer databases. The revenue made from the spammer is shared with the source. For instance if someone applies online for a mortgage, the owner of this site may have made a deal with a spammer to sell the address. These are considered the best emails by spammers, because they are fresh and the user has just signed up for a product or service that often is marketed by spam.

Legality

In Australia, the creation or use of email-address harvesting programs (address harvesting software) is illegal according to the 2003 anti-spam legislation only if you intend to use the email-address harvesting programs to send unsolicited commercial email. The legislation is intended to prohibit emails with 'an Australian connection' - spam originating in Australia being sent elsewhere, and spam being sent to an Australian address.

In The United States of America, the CAN-SPAM Act of 2003 made it illegal to initiate e-mail to a recipient where the electronic mail address of the recipient was obtained:

- Using an automated means that generates possible electronic mail addresses by combining names, letters, or numbers into numerous permutations.

- Using an automated means to extract electronic mail addresses from an Internet website or proprietary online service operated by another person, and such website or online service included, at the time the address was obtained, a notice stating that the operator of such website or online service will not give, sell, or otherwise transfer addresses maintained by such website or online service to any other party for the purposes of initiating, or enabling others to initiate, electronic mail messages.

Furthermore, website operators may not distribute their legitimately collected lists. The CAN-SPAM Act of 2003 requires operators of web sites and online services should include a notice that the site or service will not give, sell, or otherwise transfer addresses maintained by such website or online service to any other party for the purposes of initiating, or enabling others to initiate, electronic mail messages.

Anti-harvesting methods

Address munging

Address munging is a common technique to make harvesting email addresses more difficult. Though relatively easy to overcome—see, e.g., this Google search—it is still effective. It is somewhat inconvenient to users, who must examine the address and manually correct it.

Images

Using images to display part or all of an email address is a very effective harvesting countermeasure. The processing required to automatically extract text from images is not economically viable for spammers. It is very inconvenient for users, who must manually launch their email client and transcribe the address.

Contact forms

Email contact forms which send an email but do not reveal the recipient's address avoid publishing an email address in the first place. Insecure forms, however, may actually aid spammers by effectively serving as an open mail relay. This method prevents users from composing in their preferred client and limits message content to plain text.

JavaScript obfuscation

JavaScript email obfuscation produces a normal, clickable email link for users while obscuring the address from spiders. In the source code seen by harvesters, the email address is scrambled, encoded, or otherwise obfuscated. In practice, a simple ROT13 encoding has been found to be very effective. This method is very convenient for most users; however, it does reduce accessibility, e.g. for text-based browsers and screen readers. For users with a JavaScript-enabled browser, this solution is entirely transparent.

HTML obfuscation

In HTML, email addresses may be obfuscated in many ways, such as inserting hidden elements within the address or listing parts out of order and using CSS to restore the correct order. Each has the benefit of being transparent to most users, but none support clickable email links and none are accessible to text-based browsers and screen readers.

CAPTCHA

Requiring users to complete a CAPTCHA before giving out an email address is an effective harvesting countermeasure. A popular solution is the reCAPTCHA Mailhide service.

CAN-SPAM Notice

To enable prosecution of spammers under the CAN-SPAM Act of 2003, a website operator must post a notice that "the site or service will not give, sell, or otherwise transfer addresses maintained by such website or online service to any other party for the purposes of initiating, or enabling others to initiate, electronic mail messages."

Mail Server Monitoring

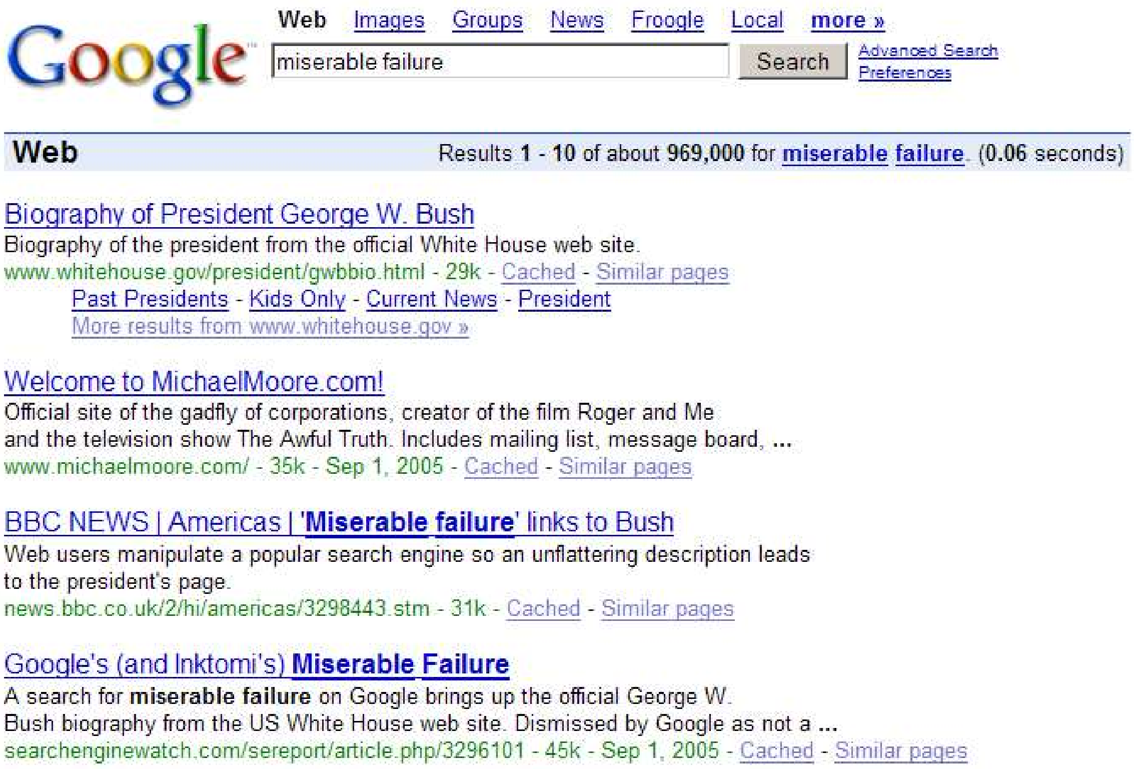
A method that can be implemented at the recipient email server for combatting directory harvesting attacks is to reject all e-mail addresses as invalid from any sender that has specified more than one invalid recipient address; however, this carries a risk of legitimate email being blocked too.

Spider Traps

A spider trap is a part of a website which is a honeypot designed to combat email harvesting spiders. Well-behaved spiders are unaffected, as the website's robots.txt file will warn spiders to stay away from that area—a warning that malicious spiders do not heed. Some traps block access from the client's IP as soon as the trap is accessed. Others, like a network tarpit, are designed to waste the time and resources of malicious spiders by slowly and endlessly feeding the spider useless information. The "bait" content may contain large numbers of fake addresses, a technique known as list poisoning, though some consider this practice harmful.

Chapter 8

Google Bomb



miserable failure

Web Results **1** - **10** of about **969,000** for **miserable failure**. (**0.06** seconds)

Biography of President George W. Bush
Biography of the president from the official White House web site.
www.whitehouse.gov/president/gwbbio.html - 29k - Cached - Similar pages
Past Presidents - Kids Only - Current News - President
More results from www.whitehouse.gov »

Welcome to MichaelMoore.com!
Official site of the gadfly of corporations, creator of the film Roger and Me and the television show The Awful Truth. Includes mailing list, message board, **...**
www.michaelmoore.com/ - 35k - Sep 1, 2005 - Cached - Similar pages

BBC NEWS | Americas | '**Miserable failure**' links to Bush
Web users manipulate a popular search engine so an unflattering description leads to the president's page.
news.bbc.co.uk/2/hi/americas/3298443.stm - 31k - Cached - Similar pages

Google's (and Inktomi's) **Miserable Failure**
A search for **miserable failure** on Google brings up the official George W. Bush biography from the US White House web site. Dismissed by Google as not a **...**
searchenginewatch.com/sereport/article.php/3296101 - 45k - Sep 1, 2005 - Cached - Similar pages

Google bombing here causes the search query "miserable failure" to be associated with George W. Bush and Michael Moore

The terms **Google bomb** and **Googlewashing** refer to practices intended to influence the ranking of particular pages in results returned by the Google search engine, in order to increase the likelihood of people finding and clicking on selections in which the individual or other entity engaging in this practice is interested. It is done for either business, political, or comedic purposes (or a combination of the latter two). Google's search-rank algorithm ranks pages higher for a particular search phrase if enough other pages linked to it using similar anchor text (linking text such as "miserable failure"). However, by January 2007 Google had made changes to search results to counter popular Google bombs, such as "miserable failure", which now lists pages about the Google bomb itself. Other Google bombs, however, continue to remain operative, as exampled by the SERP for the search phrase "french military victories". *Google bomb* is used both as a verb and a noun. The phrase "Google bombing" was introduced to the *New Oxford American Dictionary* in May 2005. Google bombing is closely related to spamdexing, the practice of deliberately modifying HTML pages to increase the chance of their website being placed close to the beginning of search engine results, or to influence the category to which the page is assigned in a misleading or dishonest manner.

The term *Googlewashing* was coined in 2003 to describe the use of media manipulation to change the perception of a term, or push out competition from search engine results pages (SERPs).

History

Google bombs date back as far as 1999, when a search for "more evil than Satan himself" resulted in the Microsoft homepage as the top result.

In September 2000 the first Google bomb with a verifiable creator was created by Hugedisk Men's Magazine, a now-defunct online humor magazine, when it linked the text "dumb motherfucker" to a site selling George W. Bush-related merchandise. Hugedisk had also unsuccessfully attempted to Google bomb an equally derogatory term to bring up an Al Gore-related site. After a fair amount of publicity the George W. Bush-related merchandise site retained lawyers and sent a cease and desist letter to Hugedisk, thereby ending the Google bomb.

Adam Mathes is credited with coining the term "Google bombing" when he mentioned it in an article that appeared on 6 April 2001 in the online magazine uber.nu. In the article Mathes details his connection of the search term "talentless hack" to the website of his friend Andy Pressman by recruiting fellow webloggers to link to his friend's page with the desired term. However, Archimedes Plutonium is known to have used the phrase "search engine bombing" (and variants, including "searchengine bombing" and "searchenginebombed") on Usenet as early as 1997.

Uses as tactical media

The Google Bomb has been used for tactical media as a way of performing a 'hit-and-run' media attack on popular topics. Such attacks include Anthony Cox's attack in 2003. He created a parody of the "404 – page not found" browser error message in response to the war in Iraq. The page looked like the error page but was titled "These Weapons of Mass Destruction cannot be displayed." This website could be found as one of the top hits on Google after the start of the war in Iraq.

Google bowling

By studying what types of inappropriate ranking manipulations a search engine is both punishing and can also easily detect, an unethical company can provoke a search engine into lowering the ranking of a competitor's website. This practice, known as **Google bowling**, is often done by purchasing Google bombing services (or other black hat SEO techniques) not for one's own website, but rather for the website of the competitor. The attacker provokes the search company into punishing the "offending" competitor by displaying their page further down in the search results. For victims of Google bowling, it may be difficult to appeal the ranking decrease because Google avoids explaining penalties, preferring not to "educate" real offenders. However if the situation is clear-cut, Google could lift the penalty after submitting a request for reconsideration.

Beyond Google

Other search engines use similar techniques to rank results, so Yahoo!, AltaVista, and HotBot are also affected by Google bombs. A search for "miserable failure" or "failure" on 29 September 2006 brought up the official George W. Bush biography number one on Google, Yahoo!, and MSN and number two on Ask.com. On 2 June 2005, Yooter reported that George Bush was ranked first for the keyword 'miserable', 'failure', and 'miserable failure' in both Google and Yahoo!; Google has since addressed this and disarmed the George Bush Google bomb and many others.

The BBC, reporting on Google bombs in 2002, used the headline "Google Hit By Link Bombers", acknowledging to some degree the idea of "link bombing." In 2004, the Search Engine Watch site suggested that the term should be "link bombing" because of its application beyond Google, and continues to use that term as it is considered more accurate.

We don't condone the practice of googlebombing, or any other action that seeks to affect the integrity of our search results, but we're also reluctant to alter our results by hand in order to prevent such items from showing up. Pranks like this may be distracting to some, but they don't affect the overall quality of our search service, whose objectivity, as always, remains the core of our mission.

–

By January 2007, Google changed their indexing structure so that Google bombs such as "miserable failure" would "typically return commentary, discussions, and articles" about the tactic itself. Google announced the changes on its official blog. In response to criticism for allowing the Google bombs, Matt Cutts, the head of Google’s Webspam team, said that Google bombs had not “been a very high priority for us.”

Over time, we’ve seen more people assume that they are Google’s opinion, or that Google has hand-coded the results for these Google-bombed queries. That’s not true, and it seemed like it was worth trying to correct that misperception.

Motivations

Competitions

In May 2004, the websites Dark Blue and SearchGuild teamed up to create what they termed the "SEO Challenge" to Google bomb the phrase "nigritude ultramarine".

The contest sparked controversy around the Internet, as some groups worried that search engine optimization (SEO) companies would abuse the techniques used in the competition to alter queries more relevant to the average user. This fear was offset by the belief that Google would alter their algorithm based on the methods used by the Google bombers.

In September 2004, another SEO contest was created. This time, the objective was to get the top result for the phrase "seraphim proudleduck". A large sum of money was offered to the winner, but the competition turned out to be a hoax.

In .net magazine, Issue 134, March 2005, a contest was created among five professional web site developers to make their site the number one listed site for the made-up phrase "crystalline incandescence".

Political activism

Some of the most famous Google bombs are also expressions of political opinion (e.g. "liar" leading to Tony Blair or "miserable failure" leading to the White House's biography of George W. Bush):

- In 2003, Steven Lerner, creator of Albino Blacksheep, created a parody webpage titled "French Military Victories". When typed into Google, the first result leads to a page that resembles Google, which reads, "Your search - French military victories - did not match any documents. Did you mean French military *defeats*?" The page received over 50,000 hits within 18 hours of its release. Links near the top of the page led to a simplified list of French military history. The page is still first in results for "French military victories."

- Another campaign was organized by columnist Dan Savage after former US Senator Rick Santorum made several controversial statements regarding homosexuality. The Google bombing was part of Savage's campaign to start using the word "santorum" for "the frothy mix of lube and fecal matter that is sometimes the byproduct of anal sex," and propelled the website created for that purpose to a high result for "santorum". The campaign was revived on February 21, 2011 by tv show host Stephen Colbert on The Colbert Report marking Santorum's recent recognition in the run-up to the 2012 presidential election.

Colbert cited the need for viewers to "search for 'santorum', click every one of those links and let the owners of those site know how you feel."

- In France, groups opposing the DADVSI copyright bill, proposed by minister Renaud Donnedieu de Vabres, mounted a Google bombing campaign linking *ministre blanchisseur* ("laundering minister") to an article recalling Donnedieu de Vabres' conviction for money laundering. The campaign was so efficient that, as of 2006, merely searching for *ministre* ("minister") or *blanchisseur* ("launderer") brings up a news report of his conviction as one of the first results.

- In 2004, after the controversy that erupted in the Philippines over the allegations that President Gloria Macapagal-Arroyo had cheated in the elections, the phrase "pekeng pangulo" ("fake president") was linked to her official website.

- In 2004, *kretyn* (Polish for *moron*) and similar insults referring to stupidity were linked to websites of various Polish politicians including Andrzej Lepper and Roman Giertych.

- In 2005 an Estonian blogger led a successful campaign to link the word *masendav* (Estonian for *dismal* or *depressive*) to the homepage of Estonian Centre Party. The Centre Party's website still ranks first in the results for *masendav* as of 2011.

- In the 2006 US midterm elections, many left-wing bloggers, led by MyDD.com, banded together to propel neutral or negative articles about many Republican House candidates to the top of Google searches for their names. Right-wing bloggers responded similarly.

- Also in 2006, *Siedziba szatana* (*satan's headquarters*) was linked to the website of controversial Christian broadcaster Radio Maryja.

- In January 2007, Google announced they altered their search engine algorithm to significantly reduce the effectiveness of the technique.

- In March 2007, the *Washington Post* reported that Nikolas Schiller was able to Google bomb "Redacted Name" to highlight his website's block on search engines.

- During the initial stages of the anti-Scientology campaign, Project Chanology, hackers and other members of an anonymous internet group Google-bombed the Church of Scientology's main website as the first match found when the term "Dangerous Cult" was searched.

- In September 2008, John Key, leader of the New Zealand National Party was Google-bombed with the query "clueless".

- In January 2009, a successful Google bomb was performed against the site of the Bulgarian Government by a loose group of bloggers and forum users. It was discovered that by mistake, the robots.txt on the government.bg forbade the crawling of the site by indexing machines which allowed for Google bombing. The group linked the search term "провал" (failure) to the government site. Within a couple of days, the first search result for "провал" was the Bulgarian Government's site regardless of the search results language.

- In April 2009, the website Smart Bitches, Trashy Books launched a Google bomb against Amazon in response to its removal of ´LGBT material from their ranking lists, Amazon citing it as "adult material". Within hours of its creation the page appeared on the first page of returned search results for the term "Amazon Rank".

- In July 2009, Opie and Anthony successfully performed a new method of Google bombing in which a specific word or phrase is artificially raised in the Google Trends reporting. The phrase 'Rev Al is a racist' was made #1 on Google Trends for 07-08-09 due to the controversial comments made by Reverend Al Sharpton during Michael Jackson's Memorial Service. "Corey Feldman is Hurting" was also number 14 on the top Google Trends for the same day in response to Feldman dressing up as Michael Jackson during the memorial service.

- In France, in July 2009, "trou du cul du web" (eng :"The Asshole of the Internet") returned as the first result the official website of French president Nicolas Sarkozy; in September 2010, the same tactic resulted in President Sarkozy's Facebook page being the first result.

- In Iran, in September 2009, the phrase "ahmadinejad president of iran" returned a fake Google search page which reads, "Did you mean: ahmadinejad is NOT president of iran. No standard web pages containing all your search terms were found". The phrase which is suggested by the fake Google page, ahmadinejad is NOT president of Iran, is linked to a video explaining events happening in Iran after Iranian presidential election, 2009.

- In September 2010, 4chan users tried to Google bomb the phrase 'Robert Pisano MPAA CEO arrested for child molestation!', as a related action to DDoS attacks on the RIAA, MPAA and British Phonographic Industry (BPI) websites. This was in retaliation for DDoS attacks carried out on The Pirate Bay and various other file sharing sites.

Commercial use

Some website operators have adapted Google bombing techniques to do spamdexing. This includes, among other techniques, posting of links to a site in an Internet forum along with phrases the promoter hopes to associate with the site. Unlike conventional message board spam, the object is not to attract readers to the site directly, but to increase the site's ranking under those search terms. Promoters using this technique frequently target forums with low reader traffic, in hopes that it will fly under the moderators' radar. This practice was also called "money bombing" by John Hiler circa 2004.

Another technique is for the owner of an Internet domain name to set up the domain's DNS entry so that all subdomains are directed to the same server. The operator then sets up the server so that page requests generate a page full of desired Google search terms, each linking to a subdomain of the same site, with the same title as the subdomain in the requested URL. Frequently the subdomain matches the linked phrase, with spaces replaced by underscores or hyphens. Since Google treats subdomains as distinct sites, the effect of many subdomains linking to each other is a boost to the PageRank of those subdomains and of any other site they link to.

On February 2, 2007, many have noticed changes in the Google algorithm that largely affects, among other things, Google bombs: only roughly 10% of the Google bombs worked as of February 15, 2007. This is largely due to Google refactoring its valuation of PageRank.

Quixtar's bomb

Quixtar, a multi-level marketing company, has been accused by its critics of using its large network of websites to move sites critical of Quixtar lower in search engine rankings. A Quixtar independent business owner (IBO) reports that a Quixtar leader advocated the practice in a meeting of Quixtar IBO's. Quixtar denies wrongdoing and states that its practices are in accordance with search engine rules.

Chapter 9

Web Bug

A **web bug** is an object that is embedded in a web page or e-mail and is usually invisible to the user but allows checking that a user has viewed the page or e-mail. One common use is in e-mail tracking. Alternative names are **web beacon**, **tracking bug**, and **tag** or **page tag**. Common names for web bugs implemented through an embedded image include **tracking pixel**, **pixel tag**, **1×1 gif**, and **clear gif**.

Overview

A web bug is any one of a number of techniques used to track who is reading a web page or e-mail, when, and from what computer. They can also be used to see if an e-mail was read or forwarded to someone else, or if a web page was copied to another website. The first web bugs were small images.

Some e-mails and web pages are not wholly self-contained. They may refer to content on another server, rather than including the content directly. When an e-mail client or web browser prepares such an e-mail or web page for display, it ordinarily sends a request to the server to send the additional content.

These requests typically include the IP address of the requesting computer, the time the content was requested, the type of web browser that made the request, and the existence of cookies previously set by that server. The server can store all of this information, and associate it with a unique tracking token attached to the content request.

On web pages

Web bugs are typically used by third parties to monitor the activity of customers at a site.

As an example of the way web bugs can make user logging easier, consider a company that owns a network of sites. This company may have a network that requires all images to be stored on one host computer while the pages themselves are stored elsewhere. They could use web bugs in order to count and recognize users traveling around the different servers on the network. Rather than gathering statistics and managing cookies on all their servers separately, they can use web bugs to keep them all together.

Tracking on web pages can be disabled using a number of techniques.

- Turning off a browser's cookies can prevent some web bugs from tracking a customer's specific activity. The web site logs will still record a page request from the customer's IP address, but unique information associated with a cookie cannot be recorded. However, web site server techniques that do not use cookies can be employed to help track a site's cookie-blocking users. For example, a web site can identify a request from a new visitor and send that visitor links that pass a unique ID as a GET parameter.

- Browser add-ons and extensions can be used. For example, the Ghostery add-on analyzes Java Script to detect trackers, web bugs, pixels, and beacons.

In e-mail

Web bugs are frequently used in spamming (sending unsolicited commercial e-mail) as a way of "pinging" to find which spam recipients open (and presumably read) before deleting it.

Tracking in e-mail can be disabled by:

- Many web bugs can be avoided by turning off HTML display and displaying only the text.
- Turning off the display of images while still using HTML may still allow other techniques to be used.

Implementation

Originally, a web bug was a small (usually 1×1 pixel) transparent GIF or PNG image (or an image of the same color of the background) that was embedded in an HTML page, usually a page on the web or the content of an e-mail. Modern web bugs also use the HTML IFrame, style, script, input link, embed, object, and other tags to track usage. Whenever the user opens the page with a graphical browser or e-mail reader, the image or other information is downloaded. This download requires the browser to request the image from the server storing it, allowing the server to take notice of the download. As a result, the organization running the server is informed when the HTML page has been viewed.

While web bugs are used in the same way in web pages or e-mails, they have different purposes:

1. If the bug is embedded in an e-mail, the image is requested when the user reads the e-mail for the first time, and can also be requested every time that the user subsequently loads the e-mail;
2. Whenever a web page (with or without bugs) is downloaded, the server holding the page knows and can store the IP address of the computer requesting the page; this information can therefore be retrieved from the server log files without the need of using bugs. Bugs are used when monitoring has to be done by a server that is different from the one holding the web pages; this is necessary, for example, when the web pages are served by different servers, or when the monitoring has to be done by a third party.

As with all files transferred using the Hypertext Transfer Protocol, web bugs are requested by sending the server their URL, and possibly the URL of the page containing them. Both URLs contain information that can be useful for the server:

1. The URL of the page containing the bug allows the server to determine which particular web page the user has accessed;
2. The URL of the bug can be appended with an arbitrary string in various ways while still identifying the same object; this extra information can be used to better identify the conditions under which the bug has been loaded; this extra information can be added while sending the page or by JavaScript scripts after the download.

For example, an e-mail sent to the address `somebody@example.org` can contain the embedded image of URL `http://example.com/bug.gif?somebody@example.org`. Whenever the user reads the e-mail, the image at this URL is requested. The part of the URL after the question mark is ignored by the server for the purpose of determining which file to send, but the complete URL is stored in the server's log file. As a result, the file `bug.gif` is sent and shown in the e-mail reader; at the same time, the server stores the fact that the particular e-mail sent to `somebody@example.org` has been read. Using this system, a spammer or e-mail marketer can send similar e-mails to a large number of addresses to check which ones are valid and read by the users.

Web bugs can be used in combination with HTTP cookies like any other object transferred using the HTTP protocol.

E-mail web bugs

Web bugs embedded in e-mails have greater privacy implications than bugs embedded in web pages. Through the use of unique identifiers contained in the URL of the web bugs, the sender of an e-mail containing a web bug is able to record the exact time that a message was read, as well as the IP address of the computer used to read the mail or the proxy server that the user went through. In this way, the sender can gather detailed information about when and where each particular recipient reads e-mail. Every subsequent time the e-mail message is displayed can also send information back to the sender.

Web bugs are used by e-mail marketers, spammers, and phishers to verify that e-mail addresses are valid, that the content of e-mails has made it past the spam filters, and that the e-mail is actually viewed by users. When the user reads the e-mail, the e-mail client requests the image, letting the sender know that the e-mail address is valid and that e-mail was viewed. The e-mail need not contain an advertisement or anything else related to the commercial activity of the sender. This makes detection of such e-mails harder for mail filters and users.

Tracking via web bugs can be prevented by using e-mail clients that do not download images whose URLs are embedded in HTML e-mails. Many graphical e-mail clients can be configured to avoid accessing remote images. Examples include the Gmail, Yahoo!, and SpamCop/Horde webmail clients; Mozilla Thunderbird, Opera, Pegasus Mail, IncrediMail, later versions of Microsoft Outlook, and KMail mail readers. Other HTML techniques (such as IFrames) can still be used to track e-mail viewing.

Text-based mail readers (such as Pine or Mutt) and graphical e-mail clients with purely text-based HTML capabilities (such as Mulberry) do not interpret HTML or display images, so their users are not subject to tracking by e-mail web bugs. Plain-text e-mail messages cannot contain web bugs because their contents are interpreted as display characters instead of embedded HTML code, so opening messages does not initiate communication. Some e-mail clients offer the option to disable all HTML in every message (thus rendering all messages as plain text), which prevents any web bugs from loading.

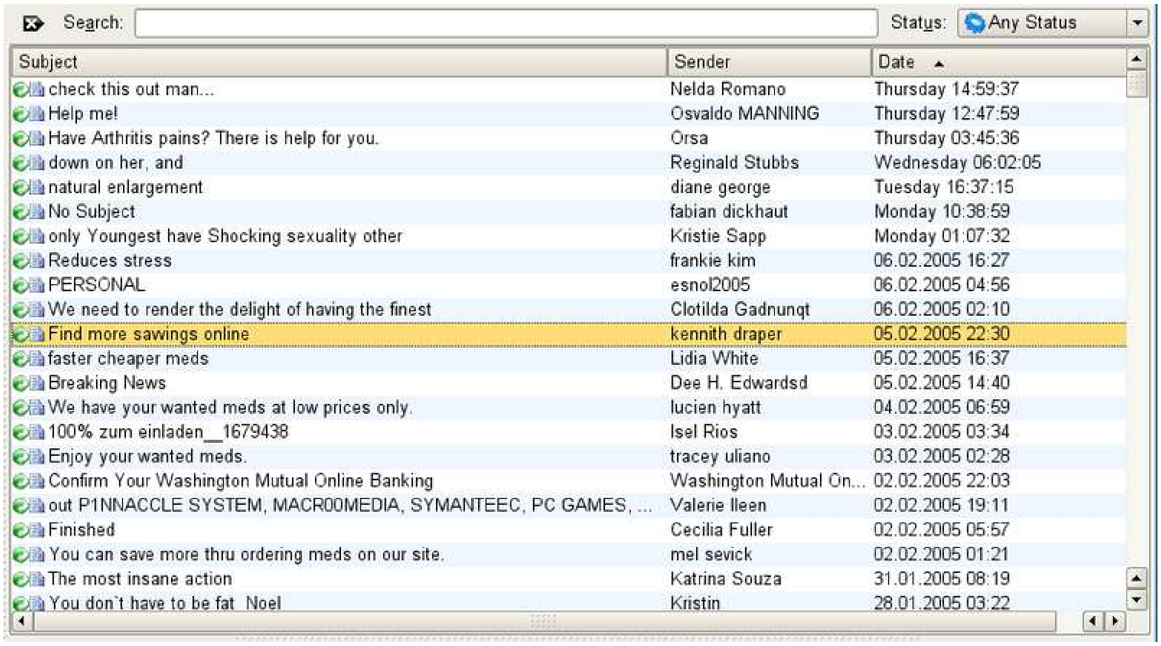
Many modern e-mail readers and web-based e-mail services will not load images when opening an HTML e-mail from an unknown sender or that is suspected to be spam mail. The user must explicitly choose to load images. Web bugs can also be filtered out at the server level so that they never reach the end user. MailScanner is an example of gateway software that can disarm IFrames as well as web bugs. Momentarily disabling a computer's Internet connection before reading new emails and deleting those messages suspicious of containing web bugs may also eliminate the threat.

A hosts file can be used to specify that some servers are never to be contacted for any reason. This file must be continually updated to reflect the fact that new tracking servers are periodically brought online, and old ones repurposed to serve legitimate content.

As a result of these measures, web bugs are slowly losing their effectiveness and cannot be relied on to accurately count read rates for e-mail campaigns.

Chapter 10

E-mail Spam



Subject	Sender	Date ▲
check this out man...	Nelda Romano	Thursday 14:59:37
Help me!	Osvaldo MANNING	Thursday 12:47:59
Have Arthritis pains? There is help for you.	Orsa	Thursday 03:45:36
down on her, and	Reginald Stubbs	Wednesday 06:02:05
natural enlargement	diane george	Tuesday 16:37:15
No Subject	fabian dickhaut	Monday 10:38:59
only Youngest have Shocking sexuality other	Kristie Sapp	Monday 01:07:32
Reduces stress	frankie kim	06.02.2005 16:27
PERSONAL	esnol2005	06.02.2005 04:56
We need to render the delight of having the finest	Clotilda Gadnunqt	06.02.2005 02:10
Find more sawings online	kennith draper	05.02.2005 22:30
faster cheaper meds	Lidia White	05.02.2005 16:37
Breaking News	Dee H. Edwardsd	05.02.2005 14:40
We have your wanted meds at low prices only.	lucien hyatt	04.02.2005 06:59
100% zum einladen__1679438	Isel Rios	03.02.2005 03:34
Enjoy your wanted meds.	tracey uliano	03.02.2005 02:28
Confirm Your Washington Mutual Online Banking	Washington Mutual On...	02.02.2005 22:03
out P1NNACCLE SYSTEM, MACR00MEDIA, SYMANTEEC, PC GAMES, ...	Valerie lleen	02.02.2005 19:11
Finished	Cecilia Fuller	02.02.2005 05:57
You can save more thru ordering meds on our site.	mel sevick	02.02.2005 01:21
The most insane action	Katrina Souza	31.01.2005 08:19
You don`t have to be fat Noel	Kristin	28.01.2005 03:22

An e-mail box folder filled with spam messages.

E-mail spam, also known as **junk e-mail** or **unsolicited bulk e-mail** (**UBE**), is a subset of spam that involves nearly identical messages sent to numerous recipients by e-mail. Definitions of spam usually include the aspects that e-mail is unsolicited and sent in bulk. One subset of UBE is *UCE* (unsolicited commercial e-mail). The opposite of "spam," email which one wants, is called "ham," usually when referring to a message's automated analysis (such as Bayesian filtering).

E-mail spam has steadily grown since the early 1990s. Botnets, networks of virus-infected computers, are used to send about 80% of spam. Since the cost of the spam is borne mostly by the recipient, it is effectively postage due advertising.

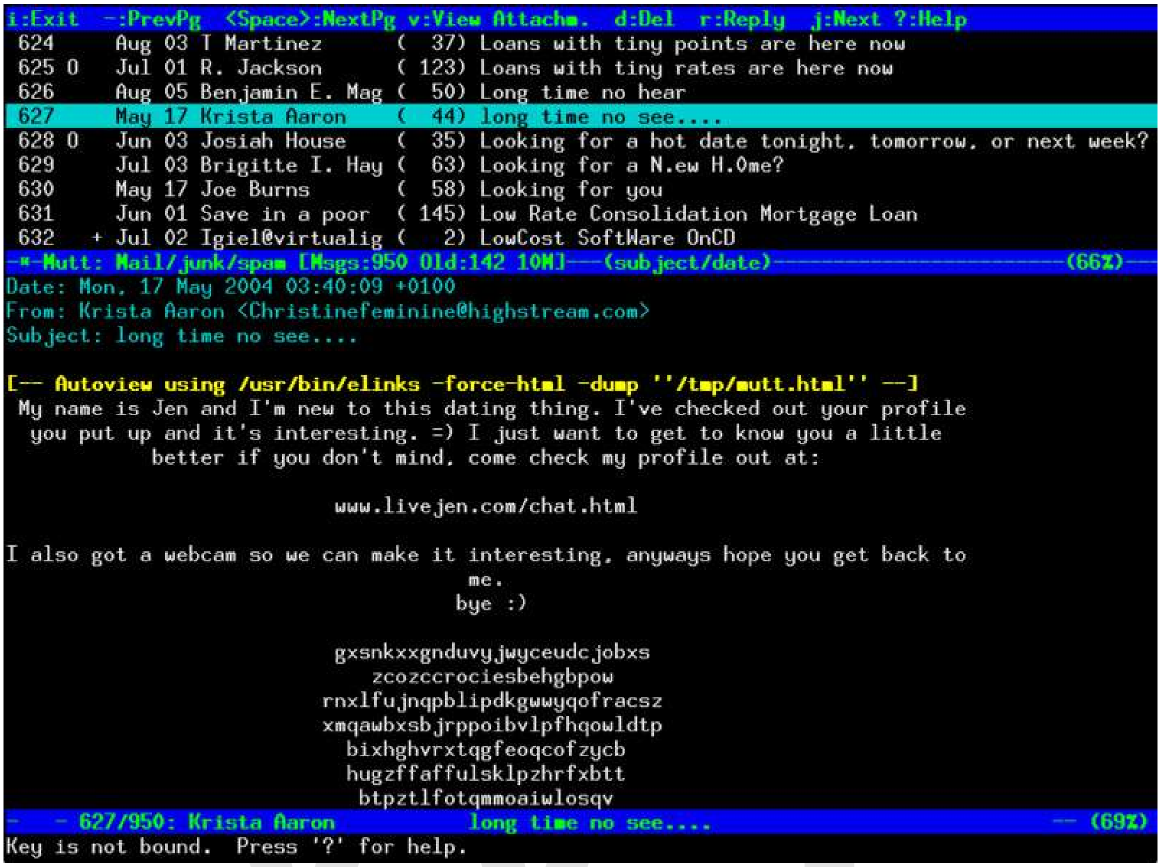
The legal status of spam varies from one jurisdiction to another. In the United States, spam was declared to be legal by the CAN-SPAM Act of 2003 provided the message adheres to certain specifications. ISPs have attempted to recover the cost of spam through lawsuits against spammers, although they have been mostly unsuccessful in collecting damages despite winning in court.

Spammers collect e-mail addresses from chatrooms, websites, customer lists, newsgroups, and viruses which harvest users' address books, and are sold to other spammers. They also use a practice known as "e-mail appending" or "epending" in which they use known information about their target (such as a postal address) to search for the target's e-mail address. Much of spam is sent to invalid e-mail addresses. Spam averages 78% of all e-mail sent. According to the Message Anti-Abuse Working Group, the amount of spam email was between 88-92% of email messages sent in the first half of 2010.

Overview

From the beginning of the Internet (the ARPANET), sending of junk e-mail has been prohibited, enforced by the Terms of Service/Acceptable Use Policy (ToS/AUP) of internet service providers (ISPs) and peer pressure. Even with a thousand users junk e-mail for advertising is not tenable, and with a million users it is not only impractical, but also expensive. It is estimated that spam cost businesses on the order of $100 billion in 2007. As the scale of the spam problem has grown, ISPs and the public have turned to government for relief from spam, which has failed to materialize.

Types



```
i:Exit  -:PrevPg  <Space>:NextPg v:View Attachm.  d:Del  r:Reply  j:Next ?:Help
 624     Aug 03 T Martinez      (  37) Loans with tiny points are here now
 625 O   Jul 01 R. Jackson      ( 123) Loans with tiny rates are here now
 626     Aug 05 Benjamin E. Mag (  50) Long time no hear
 627     May 17 Krista Aaron    (  44) long time no see....
 628 O   Jun 03 Josiah House    (  35) Looking for a hot date tonight, tomorrow, or next week?
 629     Jul 03 Brigitte I. Hay (  63) Looking for a N.ew H.0me?
 630     May 17 Joe Burns       (  58) Looking for you
 631     Jun 01 Save in a poor  ( 145) Low Rate Consolidation Mortgage Loan
 632   + Jul 02 Igiel@virtualig (   2) LowCost SoftWare OnCD
-*-Mutt: Mail/junk/spam [Msgs:950 Old:142 10M]---(subject/date)-------------------(66%)--
Date: Mon, 17 May 2004 03:40:09 +0100
From: Krista Aaron <Christinefeminine@highstream.com>
Subject: long time no see....

[-- Autoview using /usr/bin/elinks -force-html -dump ''/tmp/mutt.html'' --]
 My name is Jen and I'm new to this dating thing. I've checked out your profile
  you put up and it's interesting. =) I just want to get to know you a little
            better if you don't mind. come check my profile out at:

                           www.livejen.com/chat.html

I also got a webcam so we can make it interesting. anyways hope you get back to
                                      me.
                                     bye :)

                           gxsnkxxgnduvyjwyceudcjobxs
                              zcozccrociesbehgbpow
                          rnxlfujnqpblipdkgwwyqofracsz
                          xmqawbxsbjrppoibvlpfhqowldtp
                            bixhghvrxtqgfeoqcofzycb
                            hugzffaffulsklpzhrfxbtt
                             btpztlfotqmmoaiwlosqv
-   - 627/950: Krista Aaron          long time no see....                        -- (69%)
Key is not bound.  Press '?' for help.
```

An text based e-mail inbox with sex-based spam messages.

Spam has several definitions varying by source.

- *Unsolicited bulk e-mail* (UBE)—unsolicited e-mail, sent in large quantities.
- *Unsolicited commercial e-mail* (UCE)—this more restrictive definition is used by regulators whose mandate is to regulate commerce, such as the U.S. Federal Trade Commission.

Spamvertised sites

Many spam e-mails contain URLs to a website or websites. According to a Commtouch report in the first quarter of 2010, there are "...183 billion spam messages" sent every day. The most popular spam topic is "pharmacy ads" which make up 81% of e-mail spam messages.

Most common products advertised

According to information compiled by Commtouch Software Ltd., E-mail spam for the first quarter of 2010 can be broken down as follows.

E-Mail Spam by Topic

Pharmacy	81%
Replica	5.40%
Enhancers	2.30%
Phishing	2.30%
Degrees	1.30%
Casino	1%
Weight Loss	0.40%
Other	6.30%

419 scams

Advance fee fraud spam such as the Nigerian "419" scam may be sent by a single individual from a cyber cafe in a developing country. Organized "spam gangs" operating from Russia or eastern Europe share many features in common with other forms of organized crime, including turf battles and revenge killings.

Phishing

Spam is also a medium for fraudsters to scam users into entering personal information on fake Web sites using e-mails forged to look like they are from banks or other organizations, such as PayPal. This is known as *phishing*. Targeted phishing, where known information about the recipient is used to created forged e-mails, is known as *spear-phishing*.

Spam techniques

Appending

If a marketer has one database containing names, addresses, and telephone numbers of prospective customers, they can pay to have their database matched against an external database containing e-mail addresses. The company then has the means to send e-mail to persons who have not requested e-mail, which may include persons who have deliberately withheld their e-mail address.

Image spam

Image spam is an obfuscating method in which the text of the message is stored as a GIF or JPEG image and displayed in the e-mail. This prevents text based spam filters from detecting and blocking spam messages. Image spam was reportedly used in the mid 2000s to advertise "pump and dump" stocks.

Often, image spam contains nonsensical, computer-generated text which simply annoys the reader. However, new technology in some programs try to read the images by attempting to find text in these images. They are not very accurate, and sometimes filter out innocent images of products like a box that has words on it.

A newer technique, however, is to use an animated GIF image that does not contain clear text in its initial frame, or to contort the shapes of letters in the image (as in CAPTCHA) to avoid detection by OCR tools.

Blank spam

Blank spam is spam lacking a payload advertisement. Often the message body is missing altogether, as well as the subject line. Still, it fits the definition of spam because of its nature as bulk and unsolicited e-mail.

Blank spam may be originated in different ways, either intentional or unintentionally:

1. Blank spam can have been sent in a directory harvest attack, a form of dictionary attack for gathering valid addresses from an e-mail service provider. Since the goal in such an attack is to use the bounces to separate invalid addresses from the valid ones, spammers may dispense with most elements of the header and the entire message body, and still accomplish their goals.
2. Blank spam may also occur when a spammer forgets or otherwise fails to add the payload when he or she sets up the spam run.
3. Often blank spam headers appear truncated, suggesting that computer glitches may have contributed to this problem—from poorly-written spam software to shoddy relay servers, or any problems that may truncate header lines from the message body.
4. Some spam may appear to be blank when in fact it is not. An example of this is the VBS.Davinia.B e-mail worm which propagates through messages that have no subject line and appears blank, when in fact it uses HTML code to download other files.

Backscatter spam

Backscatter is a side-effect of e-mail spam, viruses and worms, where e-mail servers receiving spam and other mail send bounce messages to an innocent party. This occurs because the original message's envelope sender is forged to contain the e-mail address of the victim. A very large proportion of such e-mail is sent with a forged *From:* header, matching the envelope sender.

Since these messages were not solicited by the recipients, are substantially similar to each other, and are delivered in bulk quantities, they qualify as unsolicited bulk e-mail or spam. As such, systems that generate e-mail backscatter can end up being listed on various DNSBLs and be in violation of internet service providers' Terms of Service.

Legality

Sending spam violates the Acceptable Use Policy (AUP) of almost all Internet Service Providers. Providers vary in their willingness or ability to enforce their AUP. Some actively enforce their terms and terminate spammers' accounts without warning. Some ISPs lack adequate personnel or technical skills for enforcement, while others may be reluctant to enforce restrictive terms against profitable customers.

As the recipient directly bears the cost of delivery, storage, and processing, one could regard spam as the electronic equivalent of "postage-due" junk mail. Due to the low cost of sending unsolicited e-mail and the potential profit entailed, some believe that only strict legal enforcement can stop junk e-mail. The Coalition Against Unsolicited Commercial Email (CAUCE) argues "Today, much of the spam volume is sent by career criminals and malicious hackers who won't stop until they're all rounded up and put in jail."

European Union

All the countries of the European Union have passed laws that specifically target spam.

Article 13 of the European Union Directive on Privacy and Electronic Communications (2002/58/EC) provides that the EU member states shall take appropriate measures to ensure that unsolicited communications for the purposes of direct marketing are not allowed either without the consent of the subscribers concerned or in respect of subscribers who do not wish to receive these communications, the choice between these options to be determined by national legislation.

Canada

The Government of Canada has passed anti-spam legislation called the Fighting Internet and Wireless Spam Act to fight spam.

Australia

In Australia, the relevant legislation is the Spam Act 2003 which covers some types of e-mail and phone spam, which took effect on 11 April 2004. The Spam Act provides that "Unsolicited commercial electronic messages must not be sent." Whether an e-mail is unsolicited depends on whether you have consent. Consent can be express or inferred. Express consent is when someone directly instructs you to send them e-mails, e.g. if they opt-in. Consent can also be inferred from the business relationship between the sender and recipient or if the recipient conspicuously publishes their e-mail address in a public place (such as on a website). Penalties are up to 10,000 penalty units, or 2,000 penalty units for a person other than a body corporate.

United States

In the United States, most states enacted anti-spam laws during the late 1990s and early 2000s. Many of these have since been pre-empted by the less restrictive CAN-SPAM Act of 2003.

Spam is legally permissible according to the CAN-SPAM Act of 2003 provided it follows certain criteria: a "truthful" subject line, no forged information in the technical headers or sender address, and other minor requirements. If the spam fails to comply with any of these requirements it is illegal. Aggravated or accelerated penalties apply if the spammer harvested the e-mail addresses using methods described earlier.

A review of the effectiveness of CAN-SPAM in 2005 by the Federal Trade Commission (the agency charged with CAN-SPAM enforcement) stated that the amount of sexually explicit spam had significantly decreased since 2003 and the total volume had begun to level off. Senator Conrad Burns, a principal sponsor, noted that "Enforcement is key regarding the CAN-SPAM legislation." In 2004 less than 1% of spam complied with the CAN-SPAM Act of 2003. In contrast to the FTC evaluation, many observers view the CAN-SPAM act as having failed in its purpose of reducing spam.

Effectiveness

Legislative efforts to curb spam have been ineffective or counter-productive. For example, the CAN-SPAM Act of 2003 requires that each message include a means to "opt out" (i.e., decline future e-mail from the same source). It is widely believed that responding to opt-out requests is unwise, as this merely confirms to the spammer that they have reached an active e-mail account. To the extent this is true, the CAN-SPAM Act's opt-out provisions are counter-productive in two ways: first, recipients who are aware of the potential risks of opting out will decline to do so; second, attempts to opt-out will provide spammers with useful information on their targets. A 2002 study by the Center for Democracy and Technology found that about 16% of web sites tested with opt-out requests continued to spam.

Other laws

Accessing privately owned computer resources without the owner's permission counts as illegal under computer crime statutes in most nations. Deliberate spreading of computer viruses is also illegal in the United States and elsewhere. Thus, some common behaviors of spammers are criminal regardless of the legality of spamming *per se*. Even before the advent of laws specifically banning or regulating spamming, spammers were successfully prosecuted under computer fraud and abuse laws for wrongfully using others' computers.

The use of botnets can be perceived as theft. The spammer consumes a zombie owner's bandwidth and resources without any cost. In addition, spam is perceived as theft of services. The receiving SMTP servers consume significant amounts of system resources dealing with this unwanted traffic. As a result, service providers have to spend large amounts of money to make their systems capable of handling these amounts of e-mail. Such costs are inevitably passed on to the service providers' customers.

Other laws, not only those related to spam, have been used to prosecute alleged spammers. For example, Alan Ralsky was indicted on stock fraud charges in January 2008, and Robert Soloway plead guilty to charges of mail fraud, fraud in connection with electronic mail, and failing to file a tax return in March 2008.

Deception and fraud

Spammers may engage in deliberate fraud to send out their messages. Spammers often use false names, addresses, phone numbers, and other contact information to set up "disposable" accounts at various Internet service providers. They also often use falsified or stolen credit card numbers to pay for these accounts. This allows them to move quickly from one account to the next as the host ISPs discover and shut down each one.

Senders may go to great lengths to conceal the origin of their messages. Large companies may hire another firm to send their messages so that complaints or blocking of e-mail falls on a third party. Others engage in spoofing of e-mail addresses (much easier than IP address spoofing). The e-mail protocol (SMTP) has no authentication by default, so the spammer can pretend to originate a message apparently from any e-mail address. To prevent this, some ISPs and domains require the use of SMTP-AUTH, allowing positive identification of the specific account from which an e-mail originates.

Senders cannot completely spoof e-mail delivery chains (the 'Received' header), since the receiving mailserver records the actual connection from the last mailserver's IP address. To counter this, some spammers forge additional delivery headers to make it appear as if the e-mail had previously traversed many legitimate servers.

Spoofing can have serious consequences for legitimate e-mail users. Not only can their e-mail inboxes get clogged up with "undeliverable" e-mails in addition to volumes of spam, they can mistakenly be identified as a spammer. Not only may they receive irate e-mail from spam victims, but (if spam victims report the e-mail address owner to the ISP, for example) a naive ISP may terminate their service for spamming.

Theft of service

Spammers frequently seek out and make use of vulnerable third-party systems such as open mail relays and open proxy servers. SMTP forwards mail from one server to another—mail servers that ISPs run commonly require some form of authentication to ensure that the user is a customer of that ISP. Open relays, however, do not properly check who is using the mail server and pass all mail to the destination address, making it harder to track down spammers.

Increasingly, spammers use networks of malware-infected PCs (zombies) to send their spam. Zombie networks are also known as Botnets (such zombifying malware is known as a *bot*, short for robot). In June 2006, an estimated 80% of e-mail spam was sent by zombie PCs, an increase of 30% from the prior year. An estimated 55 billion e-mail spam were sent each day in June 2006, an increase of 25 billion per day from June 2005.

For Q1 2010, an estimated 305,000 newly activated zombie PCs were brought online each day for malicious activity. This number is slightly lower than the 312,000 of Q4 2009.

Brazil produced the most zombies in the first quarter of 2010. Brazil was the source of 20% of all zombies, which is down from 14% from the fourth quarter 2009. India had 10%, with Vietnam at 8%, and the Russian Federation at 7%.

Statistics and estimates

The growth of e-mail spam

The total volume of e-mail spam has been consistently growing. The amount of spam users see in their mailboxes is only a portion of total spam sent, since spammers' lists often contain a large percentage of invalid addresses and many spam filters simply delete or reject "obvious spam." A 2010 survey of US and European e-mail users showed that despite knowing the risks of opening spam e-mails, 46% of the respondents still opened them, putting their computers at risk.

In absolute numbers

- 1978 - An e-mail spam advertising a DEC product presentation is sent by Gary Thuerk to 600 addresses, which was all the users of ARPANET at the time, though software limitations meant only slightly more than half of the intended recipients actually received it.
- 2002 - 2.4 billion per day
- 2004 - 11 billion per day
- 2005 - (June) 30 billion per day
- 2006 - (June) 55 billion per day
- 2007 - (February) 90 billion per day
- 2007 - (June) 100 billion per day
- 2010 - (August) 200 billion per day

As a percentage of the total volume of e-mail

More than 97% of all e-mails sent over the net are unwanted, according to a Microsoft security report.

MAAWG estimates that 85% of incoming mail is "abusive email", as of the second half of 2007. The sample size for the MAAWG's study was over 100 million mailboxes.

Spamhaus estimates that 90% of incoming e-mail traffic is spam in North America, Europe or Australasia. By June 2008 96.5% of e-mail received by businesses was spam.

Highest amount of spam received

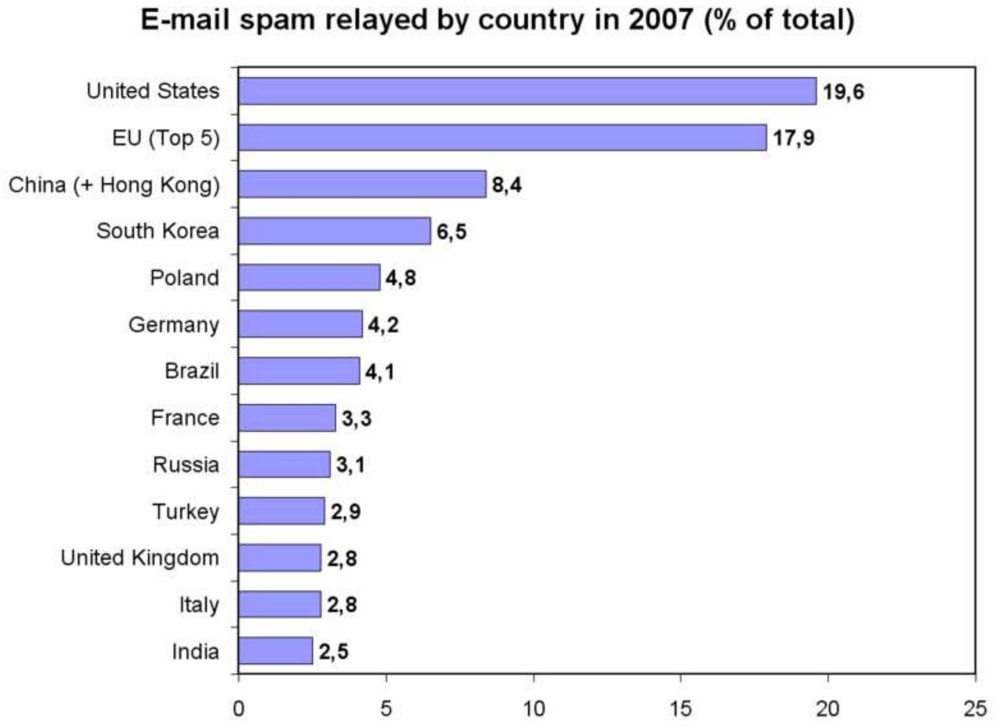
According to Steve Ballmer, Microsoft founder Bill Gates receives four million e-mails per year, most of them spam. (This was originally incorrectly reported as "per day".)

At the same time Jef Poskanzer, owner of the domain name acme.com, was receiving over one million spam e-mails per day.

Cost of spam

A 2004 survey estimated that lost productivity costs Internet users in the United States $21.58 billion annually, while another reported the cost at $17 billion, up from $11 billion in 2003. In 2004, the worldwide productivity cost of spam has been estimated to be $50 billion in 2005. An estimate of the percentage cost borne by the sender of marketing junk mail (snail mail) is 88%, whereas in 2001 one spam was estimated to cost $0.10 for the receiver and $0.00001 (0.01% of the cost) for the sender.

Origin of spam



E-mail spam relayed by country in Q2/2007.

Origin or source of spam refers to the geographical location of the computer from which the spam is sent; it is not the country where the spammer resides, nor the country that hosts the spamvertised site. Because of the international nature of spam, the spammer, the hijacked spam-sending computer, the spamvertised server, and the user target of the spam are all often located in different countries. As much as 80% of spam received by Internet users in North America and Europe can be traced to fewer than 200 spammers.

In terms of volume of spam: According to Sophos, the major sources of spam in the fourth quarter of 2008 (October to December) were:

- The United States (the origin of 19.8% of spam messages, up from 18.9% in Q3)
- China (9.9%, up from 5.4%)
- Russia (6.4%, down from 8.3%)
- Brazil (6.3%, up from 4.5%)
- Turkey (4.4%, down from 8.2%)

When grouped by continents, spam comes mostly from:

- Asia (37.8%, down from 39.8%)
- North America (23.6%, up from 21.8%)
- Europe (23.4%, down from 23.9%)
- South America (12.9%, down from 13.2%)

In terms of number of IP addresses: The Spamhaus Project (which measures spam sources in terms of number of IP addresses used for spamming, rather than volume of spam sent) ranks the top three as the United States, China, and Russia, followed by Japan, Canada, and South Korea.

In terms of networks: As of 5 June 2007, the three networks hosting the most spammers are Verizon, AT&T, and VSNL International. Verizon inherited many of these spam sources from its acquisition of MCI, specifically through the UUNet subsidiary of MCI, which Verizon subsequently renamed Verizon Business.

Spam in culture

The often rambling and incomprehensible nature of spam has led to an underground culture, with video tribute on the video sharing service YouTube, cartoons based on spam titles (Spamusement!) as well as spam blogs such as My Pet Spam, Delightful Spam and The Spam Hunter Diaries.

Anti-spam techniques

The U.S. Department of Energy Computer Incident Advisory Capability (CIAC) has provided specific countermeasures against electronic mail spamming.

Some popular methods for filtering and refusing spam include e-mail filtering based on the content of the e-mail, DNS-based blackhole lists (DNSBL), greylisting, spamtraps, Enforcing technical requirements of e-mail (SMTP), checksumming systems to detect bulk e-mail, and by putting some sort of cost on the sender via a Proof-of-work system or a micropayment. Each method has strengths and weaknesses and each is controversial because of its weaknesses. For example, one company offers for "removing some spamtrap and honeypot addresses" from e-mail lists, defeating the ability of those methods for identifying spammers.

How spammers operate

Gathering of addresses

In order to send spam, spammers need to obtain the e-mail addresses of the intended recipients. To this end, both spammers themselves and *list merchants* gather huge lists of potential e-mail addresses. Since spam is, by definition, unsolicited, this *address harvesting* is done without the consent (and sometimes against the expressed will) of the address owners. As a consequence, spammers' address lists are inaccurate. A single spam run may target tens of millions of possible addresses — many of which are invalid, malformed, or undeliverable.

Sometimes, if the sent spam is "bounced" or sent back to the sender by various programs that eliminate spam, or if the recipient clicks on an unsubscribe link, that may cause that e-mail address to be marked as "valid", which is interpreted by the spammer as "send me more".

Delivering spam messages

Obfuscating message content

Many spam-filtering techniques work by searching for patterns in the headers or bodies of messages. For instance, a user may decide that all e-mail they receive with the word "Viagra" in the subject line is spam, and instruct their mail program to automatically delete all such messages. To defeat such filters, the spammer may intentionally misspell commonly filtered words or insert other characters, often in a style similar to leetspeak, as in the following examples: `V1agra`, `Via'gra`, `Vi@graa`, `vi*gra`, `\/iagra`. This also allows for many different ways to express a given word, making identifying them all more difficult for filter software.

The principle of this method is to leave the word readable to humans (who can easily recognize the intended word for such misspellings), but not likely to be recognized by a literal computer program. This is only somewhat effective, because modern filter patterns have been designed to recognize blacklisted terms in the various iterations of misspelling. Other filters target the actual obfuscation methods, such as the non-standard use of punctuation or numerals into unusual places. Similarly, HTML-based e-mail gives the spammer more tools to obfuscate text. Inserting HTML comments between letters can foil some filters, as can including text made invisible by setting the font color to white on a white background, or shrinking the font size to the smallest fine print. Another common ploy involves presenting the text as an image, which is either sent along or loaded from a remote server. This can be foiled by not permitting an e-mail-program to load images.

As Bayesian filtering has become popular as a spam-filtering technique, spammers have started using methods to weaken it. To a rough approximation, Bayesian filters rely on word probabilities. If a message contains many words which are only used in spam, and few which are never used in spam, it is likely to be spam. To weaken Bayesian filters, some spammers, alongside the sales pitch, now include lines of irrelevant, random words, in a technique known as Bayesian poisoning. A variant on this tactic may be borrowed from the Usenet abuser known as "Hipcrime" -- to include passages from books taken from Project Gutenberg, or nonsense sentences generated with "dissociated press" algorithms. Randomly generated phrases can create spoetry (spam poetry) or spam art.

Another method used to masquerade spam as legitimate messages is the use of autogenerated sender names in the `From:` field, ranging from realistic ones such as "Jackie F. Bird" to (either by mistake or intentionally) bizarre attention-grabbing names such as "Sloppiest U. Epiglottis" or "Attentively E. Behavioral". Return addresses are also routinely auto-generated, often using unsuspecting domain owners' legitimate domain names, leading some users to blame the innocent domain owners. Blocking lists use IP addresses rather than sender domain names, as these are more accurate. A mail purporting to be from `example.com` can be seen to be faked by looking for the originating IP address in the e-mail's headers; also Sender Policy Framework, for example, helps by stating that a certain domain will only send e-mail from certain IP addresses.

Spam can also be hidden inside a fake "Undelivered mail notification" which looks like the failure notices sent by a mail transfer agent (a "MAILER-DAEMON") when it encounters an error.

Spam-support services

A number of other online activities and business practices are considered by anti-spam activists to be connected to spamming. These are sometimes termed **spam-support services**: business services, other than the actual sending of spam itself, which permit the spammer to continue operating. Spam-support services can include processing orders for goods advertised in spam, hosting Web sites or DNS records referenced in spam messages, or a number of specific services as follows:

Some Internet hosting firms advertise **bulk-friendly** or **bulletproof hosting**. This means that, unlike most ISPs, they will not terminate a customer for spamming. These hosting firms operate as clients of larger ISPs, and many have eventually been taken offline by these larger ISPs as a result of complaints regarding spam activity. Thus, while a firm may advertise bulletproof hosting, it is ultimately unable to deliver without the connivance of its upstream ISP. However, some spammers have managed to get what is called a pink contract (see below) — a contract with the ISP that allows them to spam without being disconnected.

A few companies produce **spamware**, or software designed for spammers. Spamware varies widely, but may include the ability to import thousands of addresses, to generate random addresses, to insert fraudulent headers into messages, to use dozens or hundreds of mail servers simultaneously, and to make use of open relays. The sale of spamware is illegal in eight U.S. states.

So-called **millions CDs** are commonly advertised in spam. These are CD-ROMs purportedly containing lists of e-mail addresses, for use in sending spam to these addresses. Such lists are also sold directly online, frequently with the false claim that the owners of the listed addresses have requested (or "opted in") to be included. Such lists often contain invalid addresses. In recent years, these have fallen almost entirely out of use due to the low quality e-mail addresses available on them, and because some e-mail lists exceed 20GB in size. The amount you can fit on a CD is no longer substantial.

A number of DNS blacklists (DNSBLs), including the MAPS RBL, Spamhaus SBL, SORBS and SPEWS, target the providers of spam-support services as well as spammers. DNSBLs blacklist IPs or ranges of IPs to persuade ISPs to terminate services with known customers who are spammers or resell to spammers.

Related vocabulary

Unsolicited bulk e-mail (UBE)

A synonym for e-mail spam.

Unsolicited commercial e-mail (UCE)

Spam promoting a commercial service or product. This is the most common type of spam, but it excludes spam which are hoaxes (e.g. virus warnings), political advocacy, religious messages and chain letters sent by a person to many other people. The term UCE may be most common in the USA.

Pink contract

A pink contract is a service contract offered by an ISP which offers bulk e-mail service to spamming clients, in violation of that ISP's publicly posted acceptable use policy.

Spamvertising

Spamvertising is advertising through the medium of spam.

Opt-in, confirmed opt-in, double opt-in, opt-out

Opt-in, confirmed opt-in, double opt-in, opt-out refers to whether the people on a mailing list are given the option to be put in, or taken out, of the list. Confirmation (and "double", in marketing speak) refers to an e-mail address transmitted eg. through a web form being confirmed to actually request joining a mailing list, instead of being added to the list without verification.

Final, Ultimate Solution for the Spam Problem (FUSSP)

An ironic reference to naïve developers who believe they have invented the perfect spam filter, which will stop all spam from reaching users' inboxes while deleting no legitimate e-mail accidentally.

Bacn

Bacn is an infrequently-used term to refer to e-mail sent to a user who at one time subscribed to a mailing list - not unsolicited, but also not personal.