In Berkshire Encyclopedia of Human Computer Interaction,
ed. W. Bainbridge, Berkshire Publishing, 2004, pp. 673-675

SPAMMING

Spamming is the practice of sending unsolicited bulk e-mail. The term *spam* originally referred to a spiced ham food product marketed by Hormel. The food product Spam was satirized in a skit on the British television series *Monty Python*. The skit featured a group of Vikings chanting "spam, spam, spam." Whimsical computer users found unsolicited bulk e-mail to be as ubiquitous as the chanting of the term *spam*, and the term stuck.

The first spam was sent on 5 March 1994, by the U.S. immigration law firm Canter & Siegel advertising its services with an immigration lottery. From that humble beginning spam has swollen to become a major problem confronting e-mail users. Experts estimate that 70 percent of all inbound Internet traffic is spam, and that percentage continues to rise. Much spam contains offensive material, including advertising for pornography and illegal substances.

One of the most common types of spam attempts identity theft. In some cases spam purportedly originates from a foreign country and requests that the recipient assist in moving funds out of the country for a fee. The recipient is asked for bank account information. Although spammers in many countries have produced this type of spam, most appears to originate from Nigeria and is called "419 SPAM" (named after section 419 of the Nigerian penal code which prohibits this activity).

Another type of spam that attempts identity theft is phishing—e-mail that is purportedly from a trusted source (such as a bank or online auction house) and that leads to a bogus webpage that collects personal information such as account passwords or credit card numbers. The problem of phishing is often aggravated because the bogus webpage exploits bugs in a Web browser or uses legitimate tools such as Javascript to create a false address bar in a browser. Recipients might believe that they are at a legitimate webpage when in fact they are communicating with a criminal website.

Techniques Used in Spamming

The spamming process consists of four phases: generating a list of e-mail addresses, forming messages, transmitting the spam, and collecting responses. To generate a list of e-mail addresses, spammers originally used spider programs that scoured the World Wide Web and online discussion groups for valid e-mail addresses. Because people increasingly keep e-mail lists secret, spammers have resorted to using a variety of new techniques, including guessing e-mail addresses (by using a standard dictionary of login names such as "sales" or common first names and trying all these names at registered Internet domains) and using viruses that capture users' lists of e-mail addresses.

To transmit spam, spammers often look for open e-mail relays. Standard e-mail software allows computers to run outgoing mail servers that accept e-mail and forward it. These mail servers allow mobile computing users to connect from anywhere in cyberspace back to their home computer. Spammers often scan Internet protocol (IP) addresses at random looking for open relays. Because a number of operating systems (including many versions of Linux) come with mail relays open by default, spammers are frequently able to find many such computers.

A number of emerging Asian economies (including South Korea, Taiwan, and China) are frequent targets for spammers seeking open mail relays.

In forming messages, spammers often attempt to personalize spam by including randomly generated text that evades detection by automated spam filters.

In collecting responses to spam, spammers often attempt to avoid direct detection by using a series of intermediaries. For example, they may refer a consumer to a website that will collect an order, or they may provide a foreign telephone number for communication.

People have attempted to control spam by both legislative and technical means.

Legislative Attempts to Control Spam

In the United States, Congress enacted the Controlling the Assault of Non-Solicited Pornography and Marketing Act of 2003, usually referred to by the acronym *CANSPAM.* When this article was written, people could not yet evaluate the effectiveness of the act, but some people have criticized it as being unenforceable. Whether the act applies to spam originating from non-U.S. sources is not clear. The act requires that spam include valid return e-mail addresses and options for recipients to unsubscribe to spam e-mail lists. However, anecdotal experience suggests that unsubscribing to spam e-mail lists actually results in a recipient receiving more spam; the person who originated the spam can add the recipient's address to a list of people who carefully read the spam. CANSPAM also supercedes state laws that in many cases carried stronger penalties for spamming. Finally, some people have questioned the constitutionality of antispam legislation because they feel that regulating spam infringes on free speech.

Technical Attempts to Control Spam

Many e-mail users and Internet service providers use e-mail filters to screen spam. However, these filters can fail by incorrectly identifying a non-spam message as spam (a false positive) or by incorrectly identifying a spam message as non-spam (a false negative).

Techniques for filtering spam include forming a list of IP addresses through which spam has originated or has been forwarded in the past (blacklist filtering); checking for terms that are offensive or are unlikely to occur in ordinary e-mail, such as *Viagra* (content filtering); using statistical learning methods to separate spam from non-spam (Bayesian filtering); and requiring that e-mail be accepted only if the return e-mail address is on a list of approved senders (whitelist filtering).

Use of these techniques has led to an escalating battle between spammers and spam filter creators that is often likened to an arms race. Spammers attempt to avoid being caught by blacklist filtering by sending messages through random computers or spreading viruses that themselves can spread more spam. Spammers attempt to avoid being caught by content filtering and Bayesian filtering by modifying spelling in their spam (for example, the word *Viagra* may be spelled "V1agra") and by inserting random words that prevent their messages from automatically being labeled as spam. Spammers attempt to avoid being caught by whitelist filtering by forging return addresses so that spam appears to be from the recipient's organization. Because spammers usually have access to widely available commercial spam filters, they can experiment until their messages pass detection.

Computer scientists have proposed other techniques for controlling spam, such as "e-mail for a fee," which would require that each delivered piece of e-mail be paid for with a small fee, analogous to stamps used for ordinary postal mail. The hope is that such a fee would make spam too expensive for a spammer to send. Other techniques would modify the e-mail infrastructure to require authentication of the sender using strong cryptographic methods. Unfortunately, such modifications would have to be so extensive (and in some cases would create disadvantages, such as removing the ability to send free and legitimate e-mail and anonymous e-mail) that none of these techniques has achieved wide acceptance, much less implementation.

J. D. Tygar

See also Viruses

FURTHER READING

Graham, P. (2002). *A plan for spam*. Retrieved April 13, 2004, from www.paulgraham.com/antispam.html

Schwartz, A., & Garfinkel, S. (1998). *Stopping spam*. Sebastopol, CA: O'Reilly & Associates.