ANNALS OF TECHNOLOGY

DAMN SPAM

The losing war on junk e-mail.

BY MICHAEL SPECTER

In the spring of 1978, an energetic marketing man named Gary Thuerk wanted to let people in the technology world know that his company, the Digital Equipment Corporation, was about to introduce a powerful new computer system. DEC operated out of an old wool mill in Maynard, Massachusetts, and was well known on the East Coast, but Thuerk hoped to reach the technological community in California as well. He decided that the best way to do it was through the network of government and university computers then known as the Arpanet. Only a few thousand people used it regularly, but their names were conveniently printed in a single directory. After selecting six hundred West Coast addresses, Thuerk realized that he would never have time to call each one of them, or even to send out hundreds of individual messages. Then another idea occurred to him: what if he simply used the network to dispatch a single e-mail to *all* of them? "We invite you to come see the 2020 and hear about the DECSystem-20 family," the message read. As historic lines go, it didn't have quite the ring of "One small step for a man," yet Gary Thuerk's impact cannot be disputed. When he pushed the send button, he became the father of spam.

The reaction was immediate and almost completely hostile. "This was a flagrant violation of the Arpanet," one recipient wrote. Another noted that "advertising of particular products" should be strongly discouraged on the network. The system administrator promised to respond at once, and Thuerk was harshly reprimanded. Nevertheless, his company sold more than twenty of the computer systems, for a million dollars apiece. Thuerk saw no harm in his actions; he and others viewed the network as an emerging symbol of intellectual freedom. Even if unsolicited e-mail became a nuisance, a greater danger would be posed by placing limits on how this powerful new tool could be deployed. "The amount of harm done by any of the cited 'unfair' things the net has been used for is clearly very small," the Internet pioneer Richard Stallman wrote a few days after the DEC e-mail. Stallman opposed any action that would interfere with the aggressive openness that came to define the Web. And he still does. In his message about the DEC spam, Stallman pointed out—three decades before the appearance of Craigslist and Monster.com—that the network provided a unique opportunity to advertise jobs and an entirely new way to sell products. He went even further: "Would a dating service on the net be 'frowned upon' . . . ? I hope not. But even if it is, don't let that stop you from notifying me via net mail if you start one."

I have no idea whether anyone on the Arpanet tried to help Stallman find a date, but thousands of people have tried to help me. In the past few weeks, I have received several e-mails from the Dating Adult Friend line, and several dozen from a site called Adult Friend Finder. In addition, there were fourteen messages from someone calling himself Damian Dominques, who offered, repeatedly, to help me meet "delicious babes." I also received fairly unambiguous invitations for personal interaction from people named Antonia, Heather, Helen, Joyce, Olivia, Kelly, Sally, Sophie, and Sue, among dozens of others.

Wading through dating-service spam is a minor inconvenience compared to dealing with advertisements for products designed to help those dates succeed. I received three hundred and seventeen pieces of mail offering, through surgical, mechanical, and, above all, pharmaceutical means, to help "fatten" my "love muscle," as one of them put it. There were also several hundred solicitations for low- and no-interest car loans, automatic mortgage

More than a hundred billion unwanted messages clog computer networks every day.

approvals, sleeping pills, dubious heart medicines, diet aids, gastric bypass surgery, contact lenses, air-conditioning systems, watches, online casinos, laptops, high-definition television sets, bootleg software, and jobs that promised to let me work at home, do practically nothing, and earn millions of dollars. In all, last month my three principal e-mail addresses pulled in 4,321 messages that went straight into various spam folders. Another hundred or so made it to my in-box.

As the Web evolves into an increasingly essential part of American life, the sheer volume of spam grows exponentially every year, and so, it would appear, do the sophisticated methods used to send it. Nearly two million e-mails are dispatched every second, a hundred and seventy-one billion messages a day. Most of those messages have something to sell. Even the most foolish and unsavory advertisements can earn money—in part because the economic bar for success is so low. If somebody wants to send you junk mail the old-fashioned way, through the United States Postal Service, he has to pay for it; the more he sends, the greater the expense. With electronic junk mail, the opposite is true: it costs a pittance to send a million messages—or even a billion—and recipients almost always spend more than the sender. (Assume that someone can unleash a hundred million spams from a twenty-dollar broadband account each month; at those rates, a penny would pay for fifty thousand pieces of mail.)

Spam's growth has been metastatic, both in raw numbers and as a percentage of all mail. In 2001, spam accounted for about five per cent of the traffic on the Internet; by 2004, that figure had risen to more than seventy per cent. This year, in some regions, it has edged above ninety per cent—more than a hundred billion unsolicited messages clogging the arterial passages of the world's computer networks every day. The flow of spam is often seasonal. It slows in the spring, and then, in the month that technology specialists call "black September"—when hundreds of thousands of students return to college, many armed with new computers and access to fast Internet connections—the levels rise sharply.

Attempts to police the Internet have met with only partial success. On May 23rd, the federal government indicted Robert Alan Soloway on thirty-five counts, including mail fraud, wire fraud, money laundering, and aggravated identity theft. (He has pleaded not guilty.) In its indictment, the government contended that Soloway had sent out tens of millions of illegal e-mails in the past four years, seeking to drum up business for his Internet marketing firm. Federal agents described Soloway, a twenty-seven-year-old Seattle "entrepreneur," as the nation's spam king, and said that the arrest would have a major effect on the flow of unwanted e-mail. "Taking Soloway off the streets is terrific," I was told not long ago by Matt Sergeant, the chief anti-spam technologist at MessageLabs, one of the leaders in the growing industry dedicated to ridding the Internet of junk mail. "But turn on your computer tomorrow and see if you notice a difference. These guys are sophisticated and they are everywhere. Each time we think we have them, they respond with something new."

Spam seemed to vanish after the DEC incident of 1978. Throughout the nineteen-eighties, the Internet remained largely the province of academics, few of whom had any desire to

NANOSPORE LLC

see their network turned into a platform for virtual garage sales and dating services. But, driven by the rise of eBay, in the nineties, and other commercial applications, the Internet soon became more powerful than the people who had created it. The World Wide Web was conceived in an environment where trust was assumed and identity never doubted, and that openness has been among its greatest assets and its biggest flaws. The Internet permits individuals to act without supervision, permission, or control. If you have the e-mail address, you can write directly to whomever you want; protocols and rules that have governed written communication for hundreds of years no longer apply. That absolute freedom makes cyberspace an ideal place to agitate for democracy in China, sell seventeenth-century carpets, or blog about early music. Blending these new freedoms with any sense of order or discipline has proved nearly impossible, however, and so has virtually every attempt to contain the explosion of spam.

All e-mail includes simple information about where it is going and who sent it. The mail is sorted along the way by routers—electronic devices that connect networks—which have no way of verifying that you are who you say you are. Most solutions for controlling spam would alter that practice, placing significant limits on the free exchange of information. Even many of those who fear that weak security is destroying the Internet are reluctant to support measures that appear to limit free speech. The Electronic Frontier Foundation's chairman, Brad Templeton, has written frequently on the history of spam. As his group put it in a recent white paper, "One person's spam is another's critical political update."

Under those circumstances, the emergence of spam in its modern form—mass, anonymous, and often fraudulent—was inevitable. The onslaught apparently began on April 12, 1994, when two lawyers—Laurence Canter and his wife, Martha Siegel—bombarded the Internet with e-mail offering their services to immigrants seeking to remain permanently in the United States. ("Green Card Lottery 1994 May Be The Last One! THE DEADLINE HAS BEEN ANNOUNCED.") Millions of messages went out within a few hours. The two were denounced, and their Internet-service provider immediately revoked their accounts. The sanctions didn't much matter. Canter and Siegel got what they wanted—more than a thousand clients—and were soon back online, planning their next mailing. The two later claimed that they made a hundred thousand dollars from the e-mail campaign—a compelling demonstration of the peculiar economics of the Internet. The couple embraced their notoriety and went on to write a book, "How to Make a Fortune on the Information Superhighway." It didn't take long for thousands of others to try.

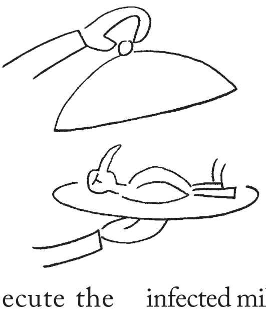
"I can't decide—do I go for the prettiest doll or the one with the most compelling backstory?"

The original Spam (a contraction of "spiced ham") is made by the Hormel Corporation, which sent enough cans of it overseas during the Second World War to feed every G.I. In a celebrated 1970 Monty Python skit, a diner tries repeatedly and in vain to order a dish, any dish, without Spam. She is drowned out by a group of Vikings in horned helmets, who chant the word dozens of times—"Spam! Spam! Spam! Spam! Spam! Spam! Spam! Spam!"—eliminating any possibility of rational thought. The word was rapidly adopted by computer programmers as a verb meaning to flood a chat room or a bulletin board with so much data that it crashes.

Definitions vary, as does the line between spam and annoying but legal ads. (Like pornography, however, which has profited greatly from the ease and privacy of electronic junk mail, you know it when you see it.) Few companies could function without attempting to stop spam from invading their employees' in-boxes. The costs are not always easy to assess, but several studies have found that in the United States more than ten billion dollars is spent each year trying to contain spam. The success rate of such anti-spam efforts usually exceeds ninety-five per cent, but spam behaves on the Internet in much the same way that viruses do when they infect humans: it might take a million of them to attack an immune system before one gets through, but one is enough. The same is true of e-mail. The more spam that is blocked, the greater the volume spammers will need to send in order to make money. "If you used to have to send fifty thousand pieces of spam to get a response, now you have to send a million," John Scarrow, the general manager of anti-spam technologies at Microsoft, told me. (Spammers usually need to send a million e-mails to get fifteen positive responses; for the average direct-mail campaign, the response rate is three thousand per million.) "Spammers just shrug it off and send a million." That amount of e-mail can overwhelm servers and waste time, particularly for those who check their mail several times a day. (It takes at least five seconds to recognize and delete an e-mail. If a billion spam messages elude detection every day—which means that ninety-nine per cent do not—that adds up to a hundred and fifty-nine years of collective time lost hitting the delete button every day.) Scarrow told me that of the four billion e-mails processed by Hotmail every day, they deliver only six hundred million. The rest are spam.

Hotmail is one of the world's largest providers of e-mail service, with two hundred and eighty-five million registered accounts in more than two hundred countries. "We filter them all, and that takes huge amounts of computer processing power and Internet bandwidth, and it requires us to work constantly to keep the numbers from getting worse," Scarrow said. "We do this to minimize the impact on our customers, but it's a hell of a job." Microsoft maintains a hundred and thirty thousand special Hotmail accounts specifically for the purpose of trapping and examining suspicious e-mail. Many function as "honeypots"—decoys that spammers think have been infected but will actually record the source's Internet address. Honeypots have no filters. "It is the raw Internet, 24/7," Scarrow said. "They will try absolutely everything. And it is often pretty raw."

In 2003, the federal government passed the Controlling the Assault of Non-Solicited Pornography and Marketing Act, which is widely referred to as the CAN-SPAM Act. The law requires people who send e-mail advertisements to offer recipients the opportunity to decline future messages. It also mandates prison terms for violators. Early in 2004, motivated in part by the excitement of the new legislation—but also by the technology achievements of researchers and engineers—Bill Gates told a group of people attending the World Economic Forum, in Davos, Switzerland, "Two years from now, spam will be solved." The comment received a lot of attention, and, for a while at least, Gates's optimism seemed justified: the deluge seemed to slow. The new law established clear guidelines about what was legal, and several companies made aggressive attempts to catch and prosecute the most significant criminals. It began to cost spammers money to evade the law, and that made them wary—for a while. The act was not meant to stop spam—simply to regulate it. Even so, it has been widely seen as a disappointment. The law permits spammers to continue sending e-mails unless specifically asked to stop, and it allows them largely to dictate the steps necessary to avoid the messages.



In the year after the law was enacted, less than seven per cent of spam complied with the requirements of the legislation, according to MX Logic, an Internet-security firm. Last year, compliance with the law never even reached one per cent. Corporate technology administrators watched, often dumbfounded, as spam volumes jumped noticeably in October, and then again in November. Postini, a prominent Internet-security firm, stopped twenty-two billion messages from reaching the mailboxes of its thirty-six thousand clients in November alone. The company now intercepts twelve spam messages for every e-mail delivered. During 2006, the year by which Gates predicted that spam would be "solved," it more than doubled in volume compared with the previous year.

We now know why. Even as Congress was passing the CAN-SPAM Act, spammers were changing their tactics. Until 2003, bulk e-mail had largely followed the approach taken by conventional mass-market mail, offering products like printing supplies and magazine subscriptions. It wasn't hard to find out who the e-mail came from, and almost nobody lied about his identity. Viruses were hardly unknown, but "it used to be all kiddies writing scripts in their bedrooms," Matt Sergeant, of MessageLabs, told me. "In 2003, spammers started paying people to write viruses to take control of home computers. The easy days were over." Viruses are actually tiny software programs that exploit weaknesses in networks or computer operating systems like Windows. They find a way to burrow into a computer's hard drive. That summer, a virus called Sobig infected millions of computers throughout the world. In a single day, MessageLabs intercepted a million copies and AOL stopped more than twenty-five million.

Sobig was the first commercial virus created by spammers designed specifically to infect machines, embed its code, and then turn those machines into networks that could send millions of e-mails. Because the e-mails were sent by innocent people who never knew that their computers were infected, the criminals were almost impossible to trace. Suddenly, spam had created an industry: a netherworld of hijacked PCs (called zombies or slaves), linked together in rogue robot networks (or botnets) controlled by underground bot herders, who operate from anywhere in the world. These networks can unleash millions of pieces of mail in a few minutes; when the botnets disband, the herders regroup and seize tens of thousands of other computers. Even the cheapest machines now have enough processing power to churn randomly through millions of address combinations until they stumble on a few that are correct.

The increase in spam levels—nearly tenfold in the past three years—is almost solely a result of botnets. Messages routinely carried viruses, many of which were designed to evade traditional filters. It's not hard to do: Many people use common, easily guessed passwords to protect their wireless networks—and a surprising number don't use passwords at all. Clicking on the wrong link at a Web address can also permit malicious software to install itself on a computer and force it to manufacture spam. This is called a "drive-by download." Once a computer virus invades, it will seek out any address book, sending copies of itself to every e-mail address it can find. Spammers today almost never use their own computers or Internet connections. It is rarely necessary, since they can seize control remotely from computers all around the world. "By the end of last year, spammers had taken over enough PCs that they could really do whatever they wanted with them," Sergeant said. "Half of the time, they are doing it on your computer and you wouldn't even have a clue."

Thomas Bayes was an eighteenth-century British clergyman and avid mathematician who became interested in probability. At the time, people had just begun to focus on the risks and rewards associated with the new field of insurance and actuarial statistics; Bayes developed a theorem that helped determine the probabilities behind the statistics. Bayesian reasoning, it turns out, can also be used to gauge the likelihood that an e-mail message is spam. Almost all defenses against spam rely on filters, which inspect words, phrases, the history of mail exchanges between the sender and the recipient, Internet-protocol (I.P.) addresses—unique numbers that are supposed to identify every computer—and other aspects of e-mail. The filters employ a series of complicated statistical methods to determine whether the message seems like spam. If an e-mail contains the words "free," "Viagra," and "herbal," for example, then the filter is likely to conclude that the message is spam. Naturally, filters make mistakes, and legitimate mail can end up in spam folders. False positives can pose a bigger problem than spam itself. "The one thing people do not want to see is genuinely important e-mail that doesn't make it to their inbox," Keith Coleman, the product manager for Google's Gmail, told me. "When that happens with any regularity, they lose their faith in e-mail completely."

A spammer's job is to confound the filters. The spellings "V1agra" or "Viagr@" mean nothing to a machine, but almost any human reader gets the point. In 2002, the programmer Paul Graham wrote an essay called "A Plan for Spam," which became an intellectual manifesto for the thousands of researchers trying to find a way to clean up the Internet. "I think it's possible to stop spam, and that content-based filters are the way to do it," he wrote. "The Achilles' heel of the spammers is their message. They can circumvent any other barrier you set up. But they have to deliver their message, whatever it is. There is no way they can get around that."

Graham compared every character—dashes, apostrophes, numbers, symbols—in thousands of genuine e-mails with those in thousands of pieces of spam. He was able to train his software to use the context of a message to guess how likely it was that an e-mail containing certain words in relation to each other was spam. The words "republic" and "madam" seem innocent enough, but when they appear together in an e-mail they are often from a Nigerian huckster who has addressed his e-mail "Dear Sir or Madam." Mail like that is invariably spam.

As filters become more sophisticated, spam becomes more elusive. There are millions of ways to write a word using punctuation, numbers, and other symbols. One mathematically minded blogger who looked into it found that there are 600,426,974,379,824,381,952 ways to spell Viagra. "If I thought that I could keep up current rates of spam filtering, I would consider this problem solved," Graham wrote. "But it doesn't mean much to be able to filter out most present-day spam, because spam evolves." Indeed, most anti-spam techniques so far have been like pesticides that do nothing other than create a more resistant strain of bugs.

It has never been easy to devise a way to sort the mail we want from the mail we don't want. Some software attempts to analyze the reputation of the sender. Has its domain sent spam before? Is its I.P. address legitimate? The most common approach is to create a fingerprint for spam, using software that assigns numeric values to different words and patterns it sees in an e-mail. These methods worked for a while, and then the arms race kicked in. The battle has moved this way for years: spam eludes filters, engineers improve the software. Each parry has been met by a new thrust. There are now blacklists, gray lists, and white lists, which permit people to choose whom they want to receive mail from, rather than whose mail to delete.

Stopping spam this way is a bit like trying to stop the rain by catching every drop before it hits the ground. The Internet itself is always available to help an aspiring spammer. There are many sites, and they are neither concealed nor subtle. There are spam supermarkets, online forums, often hosted in China or Russia, with names like specialham.com and spamforum.biz. When one is shut down, another pops up instantly. One site, send-safe.com, advertises an entire range of software packages. There is, for instance, "send-safe honeypot hunter," which is designed to help people determine whether a fake computer is on the other end of their message. The most basic program is called "send-safe mailer," software that promises to "make it impossible for anyone to trace the e-mail back to your ISP. . . . This gives you a safe haven in which to send your mail." The program, which you can buy on the Internet after a free trial, costs about seventy dollars. (One Russian Web site sells a kit called Webattacker, which contains scripts that simplify the task of infecting computers. It can be downloaded for about twenty dollars.) "You can get into the business without being technical at all," Brad Taylor, the spam czar at Google, told me. "You buy your spamming program and your spamming network. You obtain a list of mailing addresses. Anyone can do this in an hour. Then you put them all together and set up a Web site or go to a service provider. You can buy a server for a few hundred dollars and spam from that. Usually, the provider will shut you down quickly and you will be blacklisted. But then you move on to the next." Among the systems that have been infected by networks of remote computers in the past two years were computers at the weapons division of the United States Naval Air Warfare Center and many machines operated by the Department of Defense.

Spam is one of globalization's true success stories. Servers can operate from anywhere, and spam gangs sell lists of "fresh proxies" (newly infected PCs), offer "bullet-proof hosting" (spam service Web sites, often based in China), and advise each other on new spam techniques and on which networks are "spam-friendly" (those which will host spammers in exchange for the spammers' paying for high-priced services they don't need). These days, many of the world's most prodigious and talented spammers are hidden in Eastern Europe and Russia, where, despite increasingly vigorous efforts, the F.B.I. and other international law-enforcement agencies have little genuine authority. Half the time, nobody even knows their real names. Spamhaus, an organization that tracks spammers and protects networks, keeps a list of the world's biggest spam operators, and many of the best of them go by obvious pseudonyms, like the Ukrainian spammer who calls himself, variously, Alex Blood, Alexander Mosh, AlekseyB, and Alex Polyakov. For a while, people thought that his actual name might have been Polyakov and that he was in Moscow, where they hunted him aggressively. But the name seemed familiar, and one day somebody remembered why: Alex Polyakov was the name of a Soviet operative in "Tinker, Tailor, Soldier, Spy."

Last year, spammers began to take advantage of the fact that computers can't see, and buried their messages in images. Most filters look for words and phrases or Internet address information. A picture contains so much more data that it is hard for the computer to find the message embedded in all the noise. Humans who click on the message have no trouble seeing it, though. Image spam consumes far more bandwidth than written messages, and that means it will devour even more space on computer servers throughout the world, costing more money and wasting more time. But spammers aren't stopping there. They are learning to send out polymorphic spam, thousands of variations of the same message, which makes each message unique and therefore hard to categorize.

In May, death-threat spam began to appear. The message comes from a "hit man" hired to kill the recipient. "I have been hired to assassinate you," the mail typically begins. "I do not know why they want you dead, but you are now being watched." Any user scared or gullible enough to respond will be asked to wire money to save his life. The amount varies.

When I asked Brad Taylor why he had gone into this line of work, he said, "I remember my first spam. I don't remember what they were selling. It was 1994, I think, and I was so annoyed that I found out who that person's Internet provider was and reported him. I began spending the first hour of every day tracking these people down. And I felt so good about doing that. But soon it got to the point where I was getting twenty of these e-mails a day. Then thirty. At one point, I just gave up.

"But I wanted to fix the problem and return to the bliss that existed before spam," he said. "Often the fight is fun, like a game. But last year there were some low points. We started getting these image spams, and the spammer would adapt to anything we did. He would write software that cut the image into little pieces that reassembled by the time you opened your mail. When we figured out how to deal with that, he started making text that waved around and curved in odd ways. So we figured that out. Then he started with random images." Taylor laughed. "This went on for a while. But, finally, he just gave up. And that's our hope. It's kind of like war. One side eventually gets tired. And we just can't let it be us." ♦