

## LAB 3: KILLING SPAM

*Mr. Bun:* Well, what you got?

*Waitress:* Well, there's egg and bacon; egg, sausage and bacon; egg and SPAM®; egg, bacon and SPAM; egg, bacon, sausage and SPAM; SPAM, bacon, sausage and SPAM; SPAM, egg, SPAM, SPAM, bacon and SPAM; SPAM, sausage, SPAM, SPAM, SPAM, bacon, SPAM, tomato and SPAM; SPAM, egg and SPAM . . . or Lobster Thermidor au Crevettes with a Mornay sauce served in a Provencale manner with shallots and aubergines garnished with truffle pâté, brandy, a fried egg on top, and SPAM.

*Mrs. Bun:* Have you got anything without SPAM?

*Waitress:* Well, there's SPAM, egg, sausage, and SPAM. That's not got much SPAM in it.<sup>1</sup>

— from *The SPAM Sketch*, by Monty Python

**S**spam—unwanted, commercial e-mail offering you pornography, get rich quick schemes, mortgage refinancing, no-study Ph.D.'s, free credit cards, cable-TV decoders, gambling opportunities, and body-enhancing pills, patches, and potions—is the scourge of cyberspace.

### FIGHTING SPAM

E-mail falls into three categories: 1) important and/or interesting, 2) unimportant and/or uninteresting, and 3) unsolicited dreck. E-mail is *so* fast, *so* cheap, and *so* easy that its volume can overwhelm recipients. As Ed Foster, *Infoworld* editor, observed, “There is a limit to how much e-mail one can read, even when it's interesting.”

As more people have entered cyberspace, e-mail has exploded. It is not uncommon for individuals to receive hundreds of messages per day. Each one is important and/or interesting to the sender, if not the recipient.

And then there's spam.

In 1993, Joel K. “Jay” Furr coined “spam” to describe the inadvertent posting and cross-posting of hundreds ads in different Usenet newsgroups. “The first problem users we had were not commercial users,” said Furr. “They were idiots.” According to legend, the term came from MUD's, which had adapted it from the Monty Python skit mentioned earlier where everything on the menu comes with SPAM®, whether you want it or not. The term quickly evolved to mean any kind of unsolicited commercial advertising, to the chagrin of the Hormel Foods Corporation, the makers of SPAM®:

- UCE unsolicited commercial e-mail, which excludes unsolicited political messages and fraud (the similarity speaks for itself);
- UCBE unsolicited commercial bulk e-mail;
- UBE unsolicited bulk e-mail; and,

- UEMS unsolicited electronic mail solicitations, including individual e-mails.

Ever since Digital Equipment Corporation used a mass e-mailing to announce its DECSYSTEM-20 in 1978, inboxes have been choked with spam. Such unsolicited ads get a very small response. Nonetheless, e-mail is so fast, so easy, and so cheap that even a very small response can yield very high returns. According to authorities, 21-year-old Sean Dunaway, arrested in a 2003 AOL-name theft scheme, was making \$10,000 to \$20,000 *per day* from spam promoting his on-line gambling business.<sup>2</sup>

Besides being annoying, spam is expensive. It lowers worker productivity, clogs the available bandwidth, incurs direct costs for anti-spam tools and indirect costs associated with the accidental deletion of legitimate e-mail. It frequently carries viruses or worms. Perhaps worst of all, it reduces confidence in legitimate e-commerce. A 2003 study by Ferris Research estimated the cost to corporate America at \$8.9 billion annually.<sup>3</sup> A second study pegged the cost at over \$10 billion.<sup>4</sup>

### SOURCES

How do these creeps get your e-mail address, anyway? They employ several common techniques:

- purchasing mailing lists;
- spidering Web sites;
- generating names; and,

- harvesting ISP directories.

### MAILING LISTS

One way to get your address is to buy it from a second Web site that has, somehow, obtained it from you. Perhaps this second site asked you to register. Perhaps you ordered something from them. Perhaps you entered an on-line contest.

During registration, a site may present you with the option of receiving “important product information” or “updates and new product news” from the company or nebulous, unnamed “trusted partners.” Usually this option is a negative check-off. That means the option is selected by default, and you must take action to prevent it.

Leaving these boxes checked is an open invitation to spammers. “Trusted partners” are often anyone willing to buy the mailing list, and “important product information” often turns out to be advertising, plain and simple.

In any event, common sense dictates that you read a Web site’s privacy policy carefully before you give them your e-mail address or other personal information.

Common sense is wrong.

Every privacy policy contains weasel language to the effect that the Web site can change its privacy policy at any time, without notice—other than, perhaps, announcing it at its Web site. In their last, dying gasp, many a dot.com has sold its lists in a last-ditch effort to stay afloat. No legislation protects users from this unethical behavior.

The solution? Some people make up information when they register at a site. This is not a good strategy. First of all, legitimate Web sites use this information to target advertising to you. You may not want to see advertising at a Web site, but until someone guarantees you a free lunch in life, ads pay the freight. It’s true for your favorite TV show, radio broadcast, newspaper, or magazine, and it’s true for a Web site, as well.

If you are a 47-year-old male truck driver making \$37,000 a year, who is sick of ads, and you register as a 19-year-old female who makes \$250,000 a year, you will still get ads, but they will be for high fashion, expensive jewelry, and other upscale products aimed at the demographic *beau monde* of gilded youth.

A better approach is to answer the questions honestly, but to use a second, disposable, e-mail address, such as those offered by Web-based e-mail providers—Hotmail, Yahoo! Mail, or Netscape Mail, to name a few. (Some

ISP’s give you multiple e-mail addresses; using one of these works, too.) Use this secondary e-mail address whenever you register at a Web site, participate in a newsgroup, or enter on-line contests. That way, if your address is subsequently harvested or sold, you will contain the damage, apart from your primary e-mail account.

### SPIDERING WEB SITES & NEWSGROUPS

A second way to obtain e-mail addresses is to create a search-engine-like spider that searches Web pages and newsgroups for anything in this form:

`mailto:xxx@yyy.com`

where *xxx* and *yyy* are anything. One way to thwart this is to list your e-mail address in such a way that humans can understand but machines cannot, such as:

`xxx at yyy dot com`

This is not very satisfactory, because it means that you cannot click the link to open pre-addressed e-mail, and, frankly, a lot of humans just don’t get it.

A better approach is to use the escape characters we encountered in Chapter 4, page 22 (and summarized in Appendix B). For example:

`&#64;`

is the escape sequence for the “@” sign. Similarly:

`&#97;`

is the escape sequence for the lower-case “a.” With this knowledge, we can rewrite the mailto link like this:

`m&#97;ilto:xxx&#64;yyy.com`

which, to the browser, is identical to:

`mailto:xxx@yyy.com`

in every way—but confuses spammers. Whenever you place “mailto:” links in your document, disguise random characters with their corresponding escape-sequences.

A similar approach is to insert something that humans can recognize, but machines might not. For example, you might write your e-mail address as:

`spies(at)yale.edu`

or

`spies@yale. edu (remove the space)`

### HARVESTING ISP ADDRESSES

If a spammer can get his or her hands on the list of clients served by an Internet Service Provider (ISP), he or she has hit the mother lode—a large, 100 percent accurate list of e-mail addresses, plus snail-mail addresses, and, perhaps,

billing information. This happens more times than you would think.

In May 2003, for example, a 24-year-old AOL employee from Harpers Ferry, stole the e-mail addresses, zip codes, and phone numbers of its 30 million customers, along with 90 million screen names. He sold the list to a 21-year-old spammer, who, in turn, sold it to others.

## KEY-WORD MATCHES

Most simple spam filters look for certain words or key phrases in the Subject line or body. Typically, if the subject line is blank, or if it contains key words and phrases such as:

- money back;
- cards accepted;
- xxx;
- over 18;
- adults only; or,
- erotic

it may flag the message as spam (or, in later versions, examine that message more closely).<sup>5</sup> And, of course, the forbidden words and phrases never appear in spam!

Or do they?

Of course they do!

The problem is that most spam filters, including filter rules that are used in Outlook prior to Outlook 2003, look for exact matches. If a filter is looking for “sex,” it is not looking for “sextant.” This means that spammers will often disguise these key words. Instead of advertising clandestine Viagra®, they may offer:

```

Viaga
V!AGRA
V.I.A.G.R.A.
VLAGRA
Vi@gra
V!AGRA
V-I-A-G-R-A

```

So much for perfect matches! Still, if you set your spam filter to kill all the messages you receive with “free cable,” why do you still get ads for “free cable”?

The answer can be found in HTML. Most e-mail clients display HTML-formatted mail. This allows you to see pictures and formatted text. Attributes are applied by tags. The <i> tag starts italics, and </i> stops italics, so:

```
Free <i>cable$</i>TV
```

appears as:

Free *cable\$*TV

Now, as HTML and browsers have evolved, new tags have been added. The new tags, of course, can’t be understood by old browsers. To solve this problem, browsers—and e-mail clients, for that matter—simply ignore any tag that they don’t understand. Thus:

```
Free <dylan>cable$</dylan>TV
```

will appear as:

Free cable\$TV

because there is no such thing as a <dylan> tag. Taking this to its extreme:

```
F</oaf>rr</silly>e c<ha>abl</dope>e$TV
```

appears as:

Free cable\$TV

Because of this, messages that clearly mention “Free Cable\$TV”:



sneak through your filter, because when you examine the underlying code, it looks like this:

```

0000000566F82
File Edit Format View Help
<HTML><HEAD><BODY>
<p>Fr</pursuant>ee Ca</doleful>ble$ TV</p>
<a href="http://www.2004hosting.net/cable/">

</a>emission beak disco sainthood catatonic
confession devotion suicide clark one
acquisition hatred walters belle campground
brandywine hindmost bogeymen <BR>
epicycle add brandt kinshasha breathtaking
jurassic <BR></BODY></HTML>

```

Figure 64

The solution is to examine the source code for key words. To display the source code in Outlook:

- Right click the body of the message and select View Source from the pop-up menu.

### HEURISTICS

Flagging every message with a questionable word may result in false positives. Spam is annoying, but not as annoying as a spam filter eliminating a bona fide message by accident. Advanced spam filters use heuristics to identify the context of the problematic phrase. If the filter finds one questionable word in many, it's probably not spam. If it is used with other suspicious words, or if those suspicious words constitute a large percentage of the message, it probably is.

That way, a message offering a “money-back guarantee on erotic potions to adults only” is marked as spam, while a message from your mother telling you that your cousin Fred is becoming an erotic dancer is not.

Spammers are not stupid; they simply adapted. To trick these heuristics, they may fill their message with random words, as you saw in Figure 64.

### NAME GENERATORS

Assume that your e-mail address is:

`bob@aol.com`

When you get spam, you might wonder, “How did they get my name?” Perhaps they didn't. If addresses in the “To:” or CC: lines contain a discernable pattern, such as:

`boa@aol.com`

`bob@aol.com`

`boc@aol.com`

`bod@aol.com`

and so on, the addresses were probably computer-generated. Name-generating algorithms mail huge batches of messages to permutations of alphanumeric addresses, while disguising the sender. Some belong to real people, most do not. E-mail is free, so the spammer doesn't care how many

### UNSUBSCRIBE

At the end of the message, an apology may appear:

If you received this message in error, or if you'd like to be removed from our mailing list, click [here](#).

The common wisdom is that you should not do this, for doing so merely confirms that the spammer, especially one using a name generator, has stumbled upon a real address, which will then be added to a database and sold to others.

The common wisdom is wrong. Many firms sell their mailing lists to “legitimate spammers”—an oxymoron—who really will delete you from their lists. How do you tell if someone is legitimate? The best way is to examine the source code of the document. If the unsubscribe link looks normal, it is, most likely, legitimate. If it does not—that is, if it is masked with pseudo-HTML tags—then it is illegitimate. There is no reason for a legitimate firm to mask a mailto link.

### ABOUT CONFIRMATION

Spammers rarely rely on your clicking a link to confirm your address. They have a better way.

As you know, an image is not embedded in an HTML document itself. The document merely contains a link back to the server on which the picture resides. This means that if a message contains an image, merely viewing the message is all that is needed for the spammer to confirm your address. For this reason, it is advisable to turn off the “preview” window in your e-mail client. That way you can delete many suspicious messages before viewing them—and confirming your address in the process. Also, e-mail that contains no text, but just an image of text, cannot be caught by matching algorithms.



Figure 65

Some e-mail clients allow you to see all of your e-mail as ASCII, plain text characters. This prevents any images from appearing at all. For example, if you use Outlook Express for your e-mail client:

**Step 1:** From the pull-down menu, select Tools, Options.

**Step 2:** On the Read tab, select the Read all messages in plain text check box.

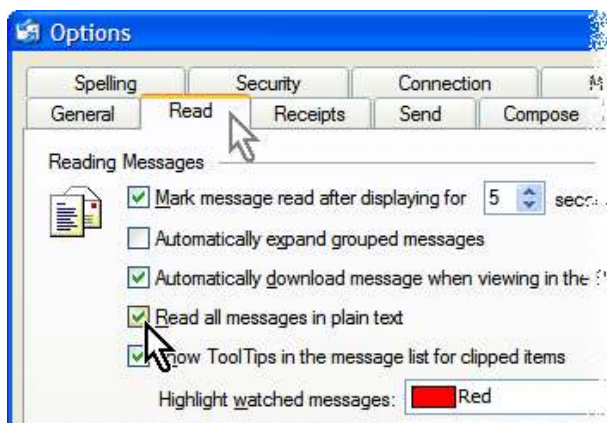


Figure 66

**Step 3:** Click  to apply your changes and close the dialog.

## MORE TIPS

When the “To” line (or CC: line) contain real names like this:

- spies@yale.edu
- fred\_spies@yale.edu
- dianna\_thomas@yale.edu
- jonathan\_miller@yale.edu

it indicates that the names were harvested from a published directory. In this case, the “From” line is usually fictitious. If you reply to the message, it will only ping back to you as “undeliverable.”

Don’t buy products from spammers. If they are unethical enough to send you spam, why would you do business with them? Not only will it encourage the practice, it provides even more personal information to them.

**Step 1:** From the pull-down menu, select Tools, Rules Wizard.

**Step 2:** Click .

**Step 3:** Select “Check messages when they arrive” and click , as shown in Figure 68.



Figure 68

**Step 4:** Select the “suspected to be junk e-mail” and “containing adult content” check boxes, as shown in Figure 69, and click  to continue.



Figure 69

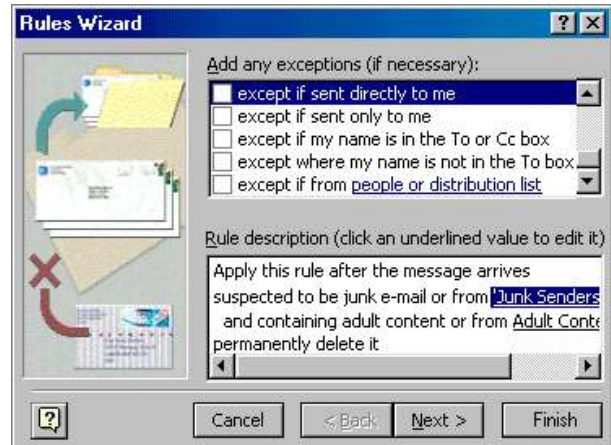



Figure 71

**Step 5:** Select the “permanently delete it” check box, shown in Figure 70, and click .

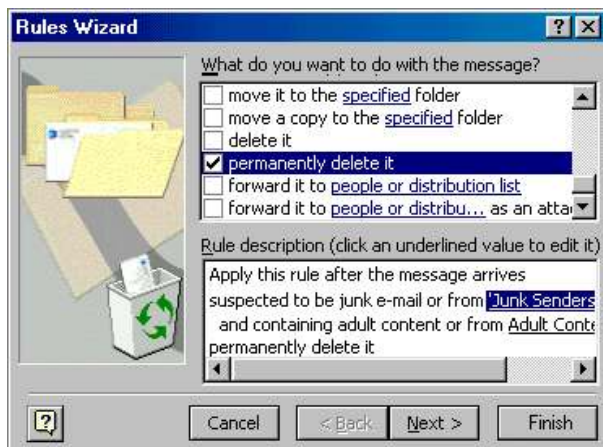
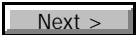
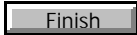


Figure 70

**Step 6:** Add exceptions, seen in Figure 71:

- except if sent directly to me;
- except if sent only to me;
- except where my name is in the Cc box;
- except if my name is in the To or Cc box;
- except where my name is not in the To box;
- except if from [people or distribution list](#);
- except if sent to [people or distribution list](#);
- except with specific words in the address;
- except with specific words in sender’s address;
- except if the subject contains specific words;
- except if the body contains specific words;

and click  to continue.

**Step 7:** Enter a name, turn it on, and click .

### ACTIVATE JUNK E-MAIL SCREEN

Despite your best efforts, you still may receive junk e-mail. If this occurs, you can activate Outlook’s Junk E-Mail screener:

**Step 1:** Open the Inbox or Outbox.

**Step 2:** Click .

**Step 3:** Click the Junk E-Mail tab, seen in Figure 73.

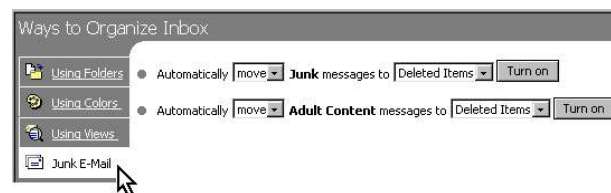
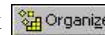


Figure 73

**Step 4:** You may move Junk messages to any folder—Deleted Items is a natural—or color them for your attention, which seems to be contrary to the purpose.

**Step 5:** You may do the same with Adult Content messages.

**Step 6:** Click the appropriate  to apply your screen.

**Step 7:** Click  to deselect the Organize frame and return to your previous view.

This tool examines key words to make its decision, and, as the dialog warns, it is not 100 percent reliable. Maybe not even 40 percent, truth be told.

### MARK A JUNK E-MAILER

When a junk e-mailer eludes your screen, you can mark or eliminate incoming from that person altogether. First, you must activate the Junk e-mail screen, as instructed in the previous section. Now, you can add an offending sender to your personal Junk E-Mailer list:

**Step 1:** When you open a message from someone you wish to block, select **A**ctions, **J**unk E-Mail, and either:

- Add to **J**unk Senders list, or,
- Add to **A**dult Content Senders list.

as shown in Figure 75.

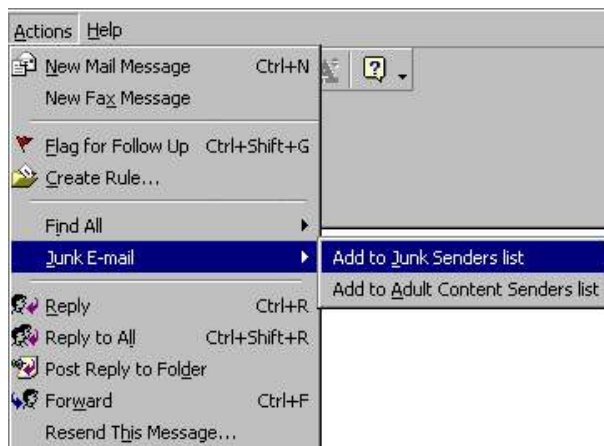


Figure 75

**Step 2:** Confirm your choice.

### UNMARK A SENDER

After you identify your brother-in-law as a junk e-mail sender, you may have second thoughts, or, perhaps it was an accident. In any event, you can unmark someone previously identified as a junk e-mail sender:

**Step 1:** Open the Inbox or Outbox.

**Step 2:** Click  Organize.

**Step 3:** Click the Junk E-Mail tab.

**Step 4:** Click the hyperlink in the message:

The Junk and Adult Content filters identify messages by looking for key words. They are not 100% accurate. For more options [click here](#).

**Step 5:** In the options shown in Figure 77, click [Edit Junk Senders](#).



Figure 77

**Step 6:** Select the junk e-mailer.

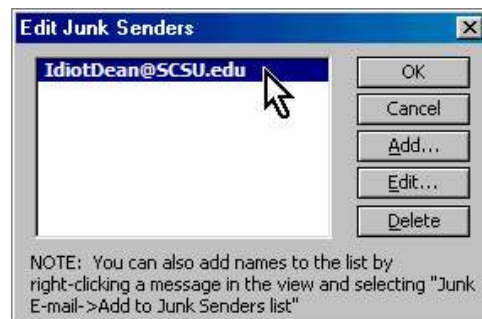

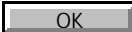
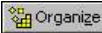


Figure 78

**Step 7:** Click  and .

**Step 8:** Click  Organize to deselect the Organize frame and return to your previous view.

### OTHER SOLUTIONS

Even if you follow these tips, you'll still get spam. Your best bet is probably just to delete it whenever you encounter it. If you've already been placed on spam lists, you may want to find other solutions. Several firms provide filters to identify and isolate spam before it gets to your inbox. Some can even determine the sender's mail server and register a complaint with the spammer's ISP. Here is a list of on-line sources for further information on spam:

#### ArtPlus: ExTerminator

<http://www.artplus.hr/adapps/eng/xterminator.htm>

#### CAUCE

<http://www.cauce.org/>

#### Cloudmark

<http://www.cloudmark.com/products/>

#### Direct Marketing Association

<http://www.the-dma.org/>

#### High Mountain: SpamEater

<http://www.hms.com/default.asp>

#### MailFrontier

<http://www.mailfrontier.com/>

#### MailWasher

<http://www.mailwasher.net/>

**McAfee: Spamkiller**

<http://www.mcafee.com/myapps/msk/>

**Postini**

<http://www.postini.com/>

**Spam Laws**

<http://www.spamlaws.com/>

and, of course, our friends at Hormel:

**SPAM®**

<http://www.spam.com>

**CAN-SPAM ACT OF 2003**

The Controlling the Assault of Non-Solicited Pornography And Marketing Act of 2003 (CAN-SPAM Act of 2003), signed into law by President Bush and implemented on January 1, 2004, required the Federal Trade Commission (FTC) to promulgate rules to protect consumers from spam. The Act defined spam as any e-mail “the primary purpose of which is the commercial advertisement or promotion of a commercial product or service (including content on an Internet Web site operated for a commercial purpose).”<sup>6</sup> Virtually all commercial e-mail is covered by this definition, except for “a transactional or relationship message” such as providing customers with account information or product upgrades in which there is “prior affirmative assent” from the recipient.<sup>7</sup> Commercial e-mail must have:

- an opt-out mechanism;
- a functioning return e-mail address;
- a valid subject line indicating it is an advertisement; and,
- the legitimate physical address of the mailer.

The Act provides both civil and criminal penalties up to five years in federal prison, plus confiscation of any real or personal property purchased with the proceeds. It also subjects businesses who procure the services of spammers to penalties, even if the FTC can’t identify the source of the actual spammer.<sup>8</sup> It also holds firms accountable for outsourcing an e-mailing “with actual knowledge, or by consciously avoiding knowing, whether such person is engaging or will engage, in a pattern or practice that violates this Act.”<sup>9</sup>

In addition to spam, the Act prohibits numerous methods of obtaining and exchanging e-mail addresses:

- harvesting e-mail addresses from Web sites;
- knowingly linking an e-mail ad to a fraudulently registered domain;
- generating random e-mail addresses; and,
- participating in fraud, identity theft, obscenity, or child pornography and exploitation.

**WEAKNESSES IN THE LAW**

Despite its noble intentions, the Act has crippling shortcomings. The Act is complicated with numerous exceptions and standards of proof, making enforcement problematic. The Act does not apply to off-shore spammers. As an “opt-out” law, it requires consumers to take action on every individual piece of spam they receive, after they’ve received it.

Regulations to prevent deception only apply to the honest. That creates a Catch-22: if you don’t trust the system, you get more spam; if you do trust the system—and opt out of spam sent by a deceptive, off-shore spammer—you get more spam.

Imposing regulations on legitimate firms raises their costs and risks, while having no effect on the real source of the problem. Even worse, the Act supercedes tougher state anti-spam laws.

The Act allows the FTC to create a national do-not-spam list similar to its effective do-not-call registry, even though testimony by the FTC at a Senate hearing raised significant technical, security, and privacy concerns.<sup>10</sup> Critics argue that such a list would simply be used by off-shore spammers to target consumers.

Although the Act allows the Federal Trade Commission (FTC), Internet Service Providers, and State Attorneys General to sue spammers in federal court for \$100 per message, up to \$2 million or more, depending on aggravating circumstances, it prohibits e-mail recipients from suing spammers.<sup>11</sup>

The bottom line: despite a couple of high-profile prosecutions, the Act has had virtually no effect on spam in the U.S.

**NOTES**

1. SPAM® is a registered trademark of Hormel Foods Corporation; [www.spam.com](http://www.spam.com). For an interesting look at the use of its trademark, see [http://www.spam.com/ci/ci\\_in.htm](http://www.spam.com/ci/ci_in.htm)
2. <http://msnbc.msn.com/id/5279826/>
3. *Spam Control: Problems and Opportunities*, <http://www.ferris.com>, January 4, 2003.
4. Joseph Rubin, U.S. Chamber of Commerce, in Congressional hearings, C-SPAN, July 8, 2003.
5. Outlook XP contained a list of these words in a file named **Filters.txt**.
6. §3(2)
7. §5(a)(5)
8. §3(9), 3(12), and 3(16)(A)
9. §7(g)(2)
10. §9
11. §7(g)(3)

