CPSC156a: The Internet Co-Evolution of Technology and Society

Lecture 16: November 4, 2003

Spam

Acknowledgement: V. Ramachandran
What is Spam?

Source: Mail Abuse Prevention System, LLC

• Spam is unsolicited bulk e-mail (primarily used for advertising).

• An electronic message is spam IF:
  (1) the recipient's personal identity and context are irrelevant because the message is equally applicable to many other potential recipients; AND
  (2) the recipient has not verifiably granted deliberate, explicit, and still-revocable permission for it to be sent; AND
  (3) the transmission and reception of the message appears to the recipient to give a disproportionate benefit to the sender.
Spam About Spam

---

Date: Tue, 15 Apr 03 20:00:05 GMT

**Do you need ADVERTISING?**

In the history of advertising, there has never been anything that compares to email.

Your ad delivered directly into the inboxes of thousands, even millions of people, instantly.

If email advertising won't sell your product then nothing will.

We can email anywhere from 250,000 to 55 million people. Let us help you with your advertising needs today.

***SPECIAL SPECIAL SPECIAL***

- 250,000 - $499 *Special $399
- 500,000 - $649 *Special $499
- 1 million - $999 *2 million for $1000
- 3 million - $1999 *Special $1499
- 10 million - $4999 *Special $3499

Specials end on Friday at 1 pm Pacific.
Why is Spam such a problem?

• Simple answer: People don’t like it!

• Cost:
  - Postal mail and telephone calls cost money.
  - Sending e-mail does not (in general).

• Speed:
  - Messages created and sent to many users instantaneously, without human effort.
  - (Almost) Instant notification of success or failure to reach destination.
Consequences of Spam

• Large amounts of network traffic (?)
  - Network congestion
  - Mail servers can be overloaded with network requests; could slow mail delivery

• Wasted Time and Storage
  - Downloading headers & checking mail takes longer
  - More unwanted mail to delete
  - E-mail must be stored at servers
  - Microsoft: 65-85% of storage costs go to Spam
How Email Works

Good explanations of

• SMTP
• Email Headers
• Mail-relay abuses

And other relevant facts can be found in
http://computer.howstuffworks.com/email.htm/printable
Tracking Spam

• SMTP runs on top of TCP.
  - Packets are acknowledged.
  - **Source** of packets is known in any successful mail session.

• SMTP servers add the IP address and hostname of every mail server or host involved in the sending process to the e-mail’s message header.

• **But**, dynamic IP addresses and large ISPs can make it difficult to identify senders.
Spoofing E-mail Headers

• Most e-mail programs use (and most people see) only the standard “To,” “Cc,” “From,” “Subject,” and “Date” headers.

• All of these are provided as part of the mail data by the mail sender’s client.

• Any of this information can be falsified.

• The only headers you can always believe are message-path headers from trusted SMTP servers.
Open Mail Relays

• An **open mail relay** is an SMTP server that will send mail when the sender and recipient are not in the server’s domain.

• These servers can be used to obfuscate the mail-sending path of messages.

• Mail-sending cost can be offloaded to servers not under spammers’ control.

• **Most servers are now configured to reject relays**, and many servers will not accept mail from known open mail relays.
SpamAssassin is a spam-fighting tool.

Primary development efforts exist for the open-source, UNIX-compatible version. The source code and select Linux binaries are available for free download (for non-commercial use).

Commercial and Windows-compatible products are available that use the technology.

SpamAssassin is installed on many ISP mail servers and is used by the CS dept. at Yale.
SpamAssassin: Overview

- Filtering is done at the mail server.
  (But, the technology can also be used to create plug-ins for mail clients.)
- Messages receive a score.
  - Message content and headers are parsed.
  - The more occurrences of Spam-like items in the message, the higher the score.
- Messages with scores above a threshold are automatically moved from the user's INBOX.
- Tolerance for Spam is user-configurable.
Judging Spam: Example #1

From: "Liz Paige" <gqwswse437dp@gatarmail.com>  
To: olivia@cloud9.net  
Cc: <pace@cloud9.net>, <photo@cloud9.net>, <promo@cloud9.net>, <ram@cloud9.net>, <reynolds@cloud9.net>, <rl@cloud9.net>  
Date: Tue, 15 Apr 03 23:03:22 GMT  
Subject: olivia@cloud9.net, Valium, Adderall, Viagra - No Prescription or Exam Affiant edduonbo gwsmsdwn

Hello Olivia,

Valium ... Xanax ... Diazepam ... Ambien  
Online Pharmacy  
No Prior Prescription Needed!  
No Physical Exam Needed!

WOW!! .... For the First Time!!  
Get VALIUM, XANAX, MERIDIA, VIAGRA and Much More  
ONLINE!!

CLICK HERE TO VISIT US NOW
Judging Spam: Results #1

<table>
<thead>
<tr>
<th>Condition</th>
<th>Points</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLICK_BELOW_CAPS</td>
<td>0.5</td>
<td>BODY: Asks you to click below (in capital letters)</td>
</tr>
<tr>
<td>BANG_MORE</td>
<td>0.5</td>
<td>BODY: Talks about more with an exclamation!</td>
</tr>
<tr>
<td>VIAGRA_ONLINE</td>
<td>0.9</td>
<td>BODY: Fast Viagra Delivery</td>
</tr>
<tr>
<td>EXCUSE_REMOVE</td>
<td>2.7</td>
<td>BODY: Talks about how to be removed from mailings</td>
</tr>
<tr>
<td>HTML_60_70</td>
<td>0.1</td>
<td>BODY: Message is 60% to 70% HTML</td>
</tr>
<tr>
<td>HTML_FONT_COLOR_RED</td>
<td>0.1</td>
<td>BODY: HTML font color is red</td>
</tr>
<tr>
<td>HTML_MESSAGE</td>
<td>0.1</td>
<td>BODY: HTML included in message</td>
</tr>
<tr>
<td>HTML_LINK_CLICK_CAPS</td>
<td>1.1</td>
<td>BODY: HTML link text says &quot;CLICK&quot;</td>
</tr>
<tr>
<td>HTML_FONT_BIG</td>
<td>0.1</td>
<td>BODY: Font Size +2 and up or 3 and up</td>
</tr>
<tr>
<td>HTML_LINK_CLICK_HERE</td>
<td>0.1</td>
<td>BODY: HTML link text says &quot;click here&quot;</td>
</tr>
<tr>
<td>HTML_SHOUTING87</td>
<td>2.9</td>
<td>BODY: HTML has very strong &quot;shouting&quot; markup</td>
</tr>
<tr>
<td>MIME_HTML_NO_CHARSET</td>
<td>0.7</td>
<td>RAW: Message text in HTML without specified charset</td>
</tr>
<tr>
<td>REMOVE_PAGE</td>
<td>0.1</td>
<td>URI: URL of page called &quot;remove&quot;</td>
</tr>
<tr>
<td>WEIRD_PORT</td>
<td>0.6</td>
<td>URI: Uses non-standard port number for HTTP</td>
</tr>
<tr>
<td>SORTED_RECIPIES</td>
<td>3.9</td>
<td>Recipient list is sorted by address</td>
</tr>
<tr>
<td>RCVD_IN_NJABL</td>
<td>1.0</td>
<td>RBL: Received via a relay in dnsbl.njabl.org</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[RBL check: found 99.70.186.61.dnsbl.njabl.org,]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[type: 127.0.0.6]</td>
</tr>
<tr>
<td>RCVD_IN_OPH</td>
<td>1.4</td>
<td>RBL: Received via a relay in opm.blitzed.org</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[RBL check: found 99.70.186.61.opm.blitzed.org,]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[type: 127.1.0.4]</td>
</tr>
<tr>
<td>MIME_HTMLONLY</td>
<td>0.1</td>
<td>Message only has text/html MIME parts</td>
</tr>
<tr>
<td>MISSING_MIMEOLE</td>
<td>0.8</td>
<td>Message has X-MIME-Priority, but no X-MimeOLE</td>
</tr>
<tr>
<td>FORGED_MUA_AOL</td>
<td>3.8</td>
<td>Forged mail pretending to be from AOL</td>
</tr>
<tr>
<td>MISSING_OUTLOOK_NAME</td>
<td>0.9</td>
<td>Message looks like Outlook, but isn't</td>
</tr>
<tr>
<td>OBSCUATING_COMMENT</td>
<td>2.7</td>
<td>HTML comments which obfuscate text</td>
</tr>
</tbody>
</table>
Judging Spam: Example #2

From: "Natasha" <sadden_eje983@msn.com>
To: <Undisclosed Recipients>@cloud9.net
Date: Mon, 14 Apr 2003 18:44:58 -0700
Subject: Smallest Digital Camera ever!

The World's Smallest DIGITAL CAMERA, CAMCORDER & WEBCAM! All built into one!

Small Enough to Fit on a Keychain... Big Enough to Take High-Quality Digital Photos, Record Video, & Act as a Full-Motion Webcam!

This is the World's first incredible Mini-Digital Camera, so versatile it's awesome! Retailers are selling them for over $200 but we have an exclusive web-offer of ONLY $39.95!

Imagine the same quality you have now but with a digital camera that slides into your pocket (less than 20 grams)! Don't waste $100's on a new camera when this does it all! Simply go to the site and get more details about this very limited offer! Supplies are limited for this web-offer!

Every Order Includes:
- Smart Mini Camera (functions as digital camera/camcorder/webcam)
- USB connector cable
- Leather Case
- Image editing software
- Keychain
- BATTERY (AAA size)
Judging Spam: Results #2
SpamAssassin: Techniques
Source: SpamAssassin.org (developers' website)

The spam-identification tactics used include:

• **header analysis**: spammers use a number of tricks to mask their identities, fool you into thinking they've sent a valid mail, or fool you into thinking you must have subscribed at some stage. SpamAssassin tries to spot these.

• **text analysis**: again, spam mails often have a characteristic style (to put it politely), and some characteristic disclaimers and CYA text. SpamAssassin can spot these, too.

• **blacklists**: SpamAssassin supports many useful existing blacklists, such as [mail-abuse.org](http://mail-abuse.org), [ordb.org](http://ordb.org) or others.

• **Razor**: [Vipul's Razor](http://www.vipul.net/razor) is a collaborative spam-tracking database, which works by taking a signature of spam messages. Since spam typically operates by sending an identical message to hundreds of people, Razor short-circuits this by allowing the first person to receive a spam to add it to the database -- at which point everyone else will automatically block it.

Once identified, the mail can then be optionally tagged as spam for later filtering using the user's own mail user-agent application.
Tricks to Avoid Filters

- Use MIME-/UU-encoding for messages.
  - E-mail messages can be in complex formats; this allows messages to contain multiple parts and attachments.
  - To preserve warping of content, message parts and attachments can be transformed using a standard encoding method.
  - E-mail clients are supposed to decode message parts when presented to the reader.
  - Basic filters often do not process encoded text!

- Insert HTML comments between words.
Examples of Tricks

Source: spam-stopper.net

As sale on NBD-C, CBS, and CN-N, and even Opr-D--ah! The health discovery of F--ly that actually reverses aging while burning fat, with oy--out dieting or exerc--ise! This pro-A--en discovery has been reported on by the NeI-test-Engl--and Jour-Engl-mal of Medi- F--cine. For I--resale--get aging and dieting forever! And it's Gua-I-S-ranteed!

Red-lo-uce body fat and build lean muscle WITH I--resale--HOURS EXERCISE! Enn--I--resale--ce I--la-xual perf--he-he--ormance!

Rem--I--resale--ove wrinkles and cellulite. Lower blood pressure and improve cholesterol--terol profile.

Resto-I--resale--re hair color and gro--th--with.

Stran-I--resale--then the immune system--tem! 

Incre-I--resale--ase energy--gy and card--iac output.

Turn bac--I--resale--k your body's clock!--ogical time clock!--ock 10-20 years in 6 months of usage!!!

You are recei-I--resale--xing this email as a subscr--iber to the Opt-I--resale--In AmerI--ca Mailin--y Lis--t. To rami-I--resale--we your--self from all related maili--me--ts, just click here:

http://www.chinaniconline.com/ultimategh/
Proposals to Eliminate Spam

• Charge a micro-payment for e-mail.
• Computational method: force senders to “prove” that they spend some minimum amount of time per recipient per message.

\[
\frac{86,400 \text{ sec/day}}{10 \text{ sec/msg}} = 8640 \text{ msgs/day}
\]

Hotmail receives 1 billion msgs / day

\[
\rightarrow \text{Would need 125,000 computers}
\]

Up-front capital cost for all of Hotmail’s spam:

\[
\sim \$150M. \text{ The spammers can’t afford it!}
\]

(-- C. Dwork, Microsoft)
Prove You are a Human

• **CAPTCHA**: Completely Automated Public Turing test for telling Computers and Humans Apart

• Require people to pass CAPTCHAs to sign up for free e-mail accounts.
  - Perform some easy-for-human but difficult-for-computer computation
  - Identify words, or find objects in pictures, *e.g.*

? The future: build into the e-mail sending process some way to prove e-mail senders are humans or authorized automated agents
The Yahoo! CAPTCHA
Legal Recourse for Spam Victims?

See, *e.g.*, Samuelson’s *CACM* article on when unsolicited commercial email (u.c.e.) constitutes “trespass on chattel” and when it doesn’t.

**Discussion point:** Is there a common theme in recent court decisions on “do-not-call” lists and on u.c.e. as trespass on chattel?
Reading Assignment for this Week

• The Electronic Frontier Foundation’s material on unintended consequences of the DMCA.
  http://www.eff.org/IP/DMCA/20030102_dmca_unintended_consequences.html

• The Electronic Privacy Information Center’s material on the USA Patriot Act.
  http://www.epic.org/privacy/terrorism/usapatriot

• “Unsolicited Communication as Trespass,” by P. Samuelson.