CS155b: E-Commerce

Lecture 23: April 17, 2003
E-Mail Abuse: Spam and Viruses

Acknowledgements: V. Ramachandran (Yale) and C. Dwork (Microsoft)
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

---

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 is E-mail Sent?

Source: RFC 821 (SMTP)

-----

Sender-SMTP          Receiver-SMTP

Model for SMTP Use

Figure 1

-----
Example Mail Exchange

[vijayr@cyndra ~]$ telnet netra 25
Trying 128.36.229.21...
Connected to netra.cs.yale.edu (128.36.229.21).
Escape character is '^[].'
220 netra.cs.yale.edu ESMTP Postfix
HELO cyndra
250 netra.cs.yale.edu
MAIL FROM:vijayr@cs.yale.edu
250 Ok
RCPT TO:vijayr@whigclio.princeton.edu
250 Ok
DATA
354 End data with <CR><LF>.<CR><LF>
This is a test
.
250 Ok: queued as EE0A5D728E
QUIT
221 Bye
Connection closed by foreign host.
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.
# E-Mail Headers

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Return</td>
<td><a href="mailto:iw3bvad9nfk@kali.com.cn">iw3bvad9nfk@kali.com.cn</a></td>
</tr>
<tr>
<td>Path</td>
<td></td>
</tr>
<tr>
<td>X-Original</td>
<td><a href="mailto:ram@cloud9.net">ram@cloud9.net</a></td>
</tr>
<tr>
<td>To</td>
<td></td>
</tr>
<tr>
<td>Delivered</td>
<td><a href="mailto:ram@cloud9.net">ram@cloud9.net</a></td>
</tr>
<tr>
<td>To</td>
<td></td>
</tr>
<tr>
<td>Received</td>
<td>from localhost (localhost [127.0.0.1]) by russian-caravan.cloud9.net (Postfix) with ESMTP id EF454AADC; Tue, 15 Apr 2003 14:09:07 -0400 (EDT)</td>
</tr>
<tr>
<td></td>
<td>from localhost (hostname [127.0.0.1]) by localhost (VaMailArmor-2.0.1.7) id 09388-51E384ED; Tue, 15 Apr 2003 14:09:07 -0400</td>
</tr>
<tr>
<td></td>
<td>from hosts217-40-121-145.in-addr.btopenworld.com (host217-40-121-145.in-addr.btopenworld.com [217.40.121.145]) by russian-caravan.cloud9.net (Postfix) with SMTP id B6BEAAA23; Tue, 15 Apr 2003 14:08:06 -0400 (EDT)</td>
</tr>
<tr>
<td></td>
<td>from wrzr4k.wb23acf.com [110.70.78.125] by host217-40-121-145.in-addr.btopenworld.com id u4162Pp3anwF for <a href="mailto:pace@cloud9.net">pace@cloud9.net</a>; Tue, 15 Apr 2003 20:05:05 +0100</td>
</tr>
<tr>
<td>Message-</td>
<td><a href="mailto:8$$7g2$-0lnu1u$-a4-s93-5pw5x@zat9bhegt.y0t">8$$7g2$-0lnu1u$-a4-s93-5pw5x@zat9bhegt.y0t</a></td>
</tr>
<tr>
<td>Id</td>
<td></td>
</tr>
<tr>
<td>From</td>
<td>&quot;Hubert Rivers&quot; <a href="mailto:iw3bvad9nfk@kali.com.cn">iw3bvad9nfk@kali.com.cn</a></td>
</tr>
<tr>
<td>To</td>
<td><a href="mailto:pace@cloud9.net">pace@cloud9.net</a></td>
</tr>
<tr>
<td>Cc</td>
<td><a href="mailto:photo@cloud9.net">photo@cloud9.net</a>, <a href="mailto:promo@cloud9.net">promo@cloud9.net</a>, <a href="mailto:ram@cloud9.net">ram@cloud9.net</a>, <a href="mailto:reynolds@cloud9.net">reynolds@cloud9.net</a>, <a href="mailto:rl@cloud9.net">rl@cloud9.net</a>, <a href="mailto:robertl@cloud9.net">robertl@cloud9.net</a></td>
</tr>
<tr>
<td>Subject</td>
<td><a href="mailto:pace@cloud9.net">pace@cloud9.net</a>, Let Us Do Your Email Advertising employee ba hum s</td>
</tr>
<tr>
<td>Date</td>
<td>Tue, 15 Apr 03 20:05:05 GMT</td>
</tr>
<tr>
<td>X-Priority</td>
<td>1</td>
</tr>
<tr>
<td>X-Msmail-</td>
<td>High</td>
</tr>
<tr>
<td>Priority</td>
<td></td>
</tr>
<tr>
<td>X-Mailer</td>
<td>MIME-tools 5.503 (Entity 5.501)</td>
</tr>
<tr>
<td>MIME</td>
<td>1.0</td>
</tr>
<tr>
<td>Version</td>
<td></td>
</tr>
<tr>
<td>Content-</td>
<td>multipart/alternative; boundary=&quot;4DFAC9BD.DC_.5ED6.9&quot;</td>
</tr>
<tr>
<td>Type</td>
<td></td>
</tr>
<tr>
<td>X-Antivirus</td>
<td>checked by Vexira MailArmor (version: 2.0.1.7; VAE: 6.19.0.3; VDF: 6.19.0.6; host: russian-caravan.cloud9.net)</td>
</tr>
</tbody>
</table>
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.
Relay Rejection

[vijayr@cyndra ~]$ telnet mail.cloud9.net 25
Trying 168.100.1.4...
Connected to russian-caravan.cloud9.net (168.100.1.4).
Escape character is '^]'.
220 russian-caravan.cloud9.net ESMTP Postfix
MAIL FROM:user@cloud9.net
250 Ok
RCPT TO:vijayr@cs.yale.edu
554 <vijayr@cs.yale.edu>: Relay access denied
QUIT
221 Bye
Connection closed by foreign host.
• 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" <gqowwase437do@qatarmail.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, Acidex, Viagra - No Prescription or Exam affiant eeduonbo gwdfsmdw

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>Criterion</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_SHOUTING59</td>
<td>2.9</td>
<td>BODY: HTML has very strong &quot;shouting&quot; markup</td>
</tr>
<tr>
<td>MIME_HTML_CHOSEN</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 dmsbl.njabl.org</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[RBL check: found 99.70.186.61.dmsbl.njabl.org,]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[type: 127.0.0.5]</td>
</tr>
<tr>
<td>RCVD_IN_OPM</td>
<td>4.3</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.cpm.blitzed.org,]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[type: 127.1.0.4]</td>
</tr>
<tr>
<td>MIME_HTML_ONLY</td>
<td>0.1</td>
<td>Message only has text/html MIME parts</td>
</tr>
<tr>
<td>MISSING_MIMEOLE</td>
<td>0.6</td>
<td>Message has X-Mail-Priority, but no X-Mail-Version</td>
</tr>
<tr>
<td>FORGED_MUA_AOL</td>
<td>3.0</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>OSBFUSCATING_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 -0400
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)
## Content Analysis Details:

- **RCVD_FAKE_HELO_DOTCOM** (2.3 points): Received contains a faked HELO hostname
- **INVALID_DATE_TZ_ABBURD** (4.3 points): Invalid Date: header (timezone does not exist)
- **MSGID_CE_SPAM_4ZERO** (3.3 points): Message-Id generated by spam tool
- **FAKED_UNDISC_RECSIES** (4.3 points): Faked To "Undisclosed-Recipients"
- **MSGID_SPANNSIGN_ZEROS** (4.3 points): Message-Id generated by spam tool
- **ONLY_COST** (0.2 points): BODY: Only $$
- **EXCUSE_3** (0.1 points): BODY: Claims you can be removed from the list
- **EXCUSE_REMOVE** (2.7 points): BODY: Talks about how to be removed from mailings
- **REMOVE_PATH** (0.1 points): URI: URL of page called "remove"
- **MSGID_OUTLOOK_TIME** (4.4 points): Message-Id is fake (in Outlook Express format)
- **DATE_IN_FUTURE_12_24** (2.8 points): Date: is 12 to 24 hours after

### Received:

- **RCVD_IN_NJABL** (1.0 points): RBL: Received via a relay in dnsbl.njabl.org
- **RCVD_IN_OEM** (4.3 points): RBL: Received via a relay in cpm.blitzed.org
- **RCVD_IN_UNCONFIRMED_DSBL** (0.7 points): RBL: Received via a relay in unconfirmed.dsbl.org
- **FROM_HAS_UNDERLINE NMS** (0.6 points): From: contains an underline and numbers/letters

### CLICK BELOW

(0.1 points): Asks you to click below
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, ordb.org or others.

- **Razor**: Vipul's 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 seen on NB-D-C, CBS, and CNN-H-N, and even Oprah on NB-D-Ahh! The health discoverevery day that actually reverses aging while burning fat. With your boy's diet and exercise, your aging and dieting forever! And it's Guadara! ranted!

Reduces body fat and builds lean muscle. With recycle and HOUT EXCERCISE!

Recycle at your local fitness center. To keep your body's blood circulation going, and it's cheap! 10-20 years in just 6 months of usage!!

Are you receiving this email as a subscription? If so, please unsubscribe to receive future communications.

Want more information? Visit www.chinanonline.com/ultimateghg/.
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 sender per message.

(86,400 sec/day) / (10 sec/msg) = 8640 msgs/day
Hotmail receives 1 billion msgs / day
-> Would need 125,000 computers
Up-front capital cost for all of Hotmail’s spam:
  ~ $150M. 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
Viruses

A **computer virus** is a piece of code, often malicious, that is intended to transmit itself between computers and replicate itself and/or execute instructions without the user’s knowledge or intent.

How Does One Get Infected?

Simple answer:
Run malicious code on your computer.

Simple reaction:
Then I won’t.

Problem:
What if you are tricked into doing it?
Or don’t know it’s happening?
Types of Viruses

• **Trojan Horses**: disguised to do one thing, but do another when run

• **Boot Sector Viruses**: reside in system sectors; run in the background while resident in memory; copy themselves to other disks

• **File Infectors**: modify portions of executable files on disk so that virus code is unknowingly executed

• **Macro Viruses**: take advantage of the programmability of documents; run when infected files are accessed

• **Worms**: replicate across networks, possibly through proprietary software protocols

• **E-mail Viruses**: transmitted through e-mail, often through attachments
Viruses: Question #1

Can you get infected simply by reading an e-mail or viewing a web page?
Viruses: Question #1

Can you get infected simply by reading an e-mail or viewing a web page?

**YES.** But your security settings have to allow it, *e.g.*, if you permit scripts to run in HTML e-mail that could contain malicious code.

*Plain text cannot contain a virus.*
Consent to Run Code

Most browsers that have the capability to execute malicious, remote code will ask you for consent before running anything triggered by a web page.

Digital signature information is displayed.

The default action (what happens if you just press ENTER) is “No.” This guards against accidental consent.
Viruses: Question #2

Can you get infected by viewing a picture attachment to an e-mail?
Viruses: Question #2

Can you get infected by viewing a picture attachment to an e-mail?

**NO.** But you can be fooled by receiving an attachment that looks like a picture but is really something else. *Always check the type of a file.*
Viruses: Question #3

Can I get infected if I own a Mac?
Can I get infected if I own a Mac?

**YES.** You might not be affected by the same viruses because the code might not run, but there are some Mac worms and e-mail viruses, and Mac files can be carriers of Windows macro viruses.
Beware of Attachments

• Back in the days of MS-DOS, code lived in three types of files: COM, EXE, BAT. Problem: If you have a virus WP.COM and a program WP.EXE, typing “WP” causes the virus to run because of precedence rules.

• As programs become more feature-rich and systems become more complex, executable code becomes part of more file types.
Files That Can Contain Code

How many extensions do you recognize?

.com  .exe  .bat  .scr  .pif
.vbs  .js  .vbx  .ocx  .dll
.doc  .xls  .ppt  .eml  .pl
.class  .htm(l)  .hta  .asp(x)
Example: Melissa

• **Microsoft Word macro virus**

• **On document load, the AutoExec macro runs, containing code that:**
  - uses Microsoft Office / Windows features to access the address book and e-mail others infected files; **AND**
  - infects the default template for Word documents, so that any new Word file on the machine contains the infected AutoExec macro.
Example: Code Red

- Microsoft IIS worm
- Uses a “buffer overflow” bug in web server software to transmit and run itself.
- Replicates wildly by sending requests across the Internet from infected machines, causing congestion.
- Changes web pages on infected machines.
- Launches a DDoS attack on www.whitehouse.gov.
Other Nasty Virus Tricks

• Modify system files.
• Force system to run virus at start-up.
• Intercept and modify requests to the operating system and provide false information (e.g., as done by “stealth” viruses).
• Change local security settings.
• Run as an Internet server in the background, creating a “back door.”
Viruses and Business

• Consider Slammer, the SQL-server worm. SQL server is a Microsoft database product. Hosts running it are often connected to the Internet so that systems can easily share data.

• Slammer infected 90% of vulnerable computers in 10 minutes and reached its peak traffic rate of 55M scans/sec after three minutes (CNET.com).
The Cost of Disinfection

Productivity Losses:

Klez: $9 billion

LoveLetter: $8.8 billion

Code Red: $2.6 billion

SQL Slammer: $0.95-1.2 billion