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

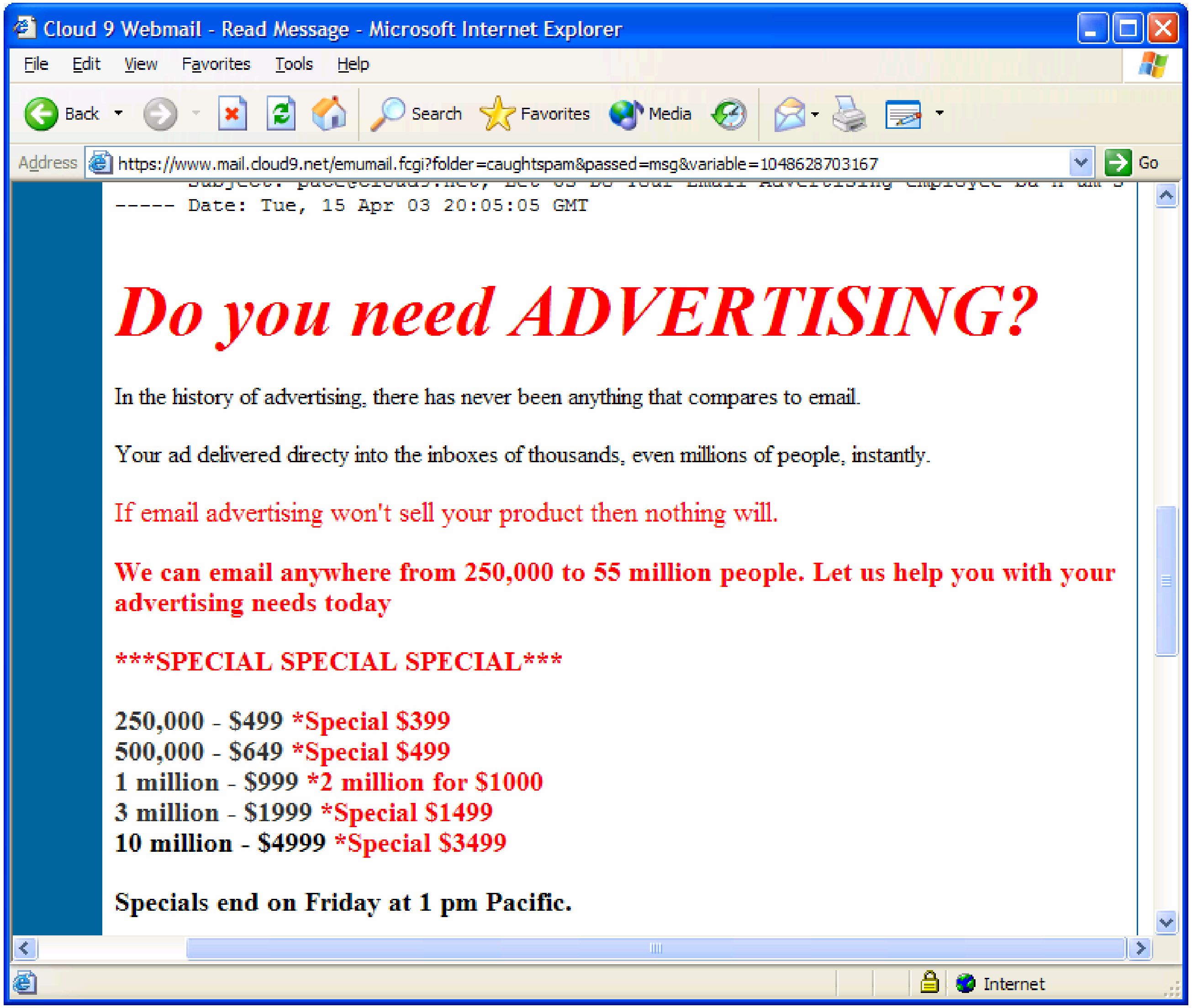
- Spam is unsolicited bulk e-mail (primarily used for advertising).
- An electronic message is spam IF:

(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



https://www.mail.cloud9.net/emumail.fcgi?folder=caughtspam&passed=msg&variable=1048628703167

```
Subject: pace@cloud9.net, Let Us Be Your Email Advertising Employee
----- Date: Tue, 15 Apr 03 20:05:05 GMT
```

Do you need ADVERTISING?

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

Your ad delivered directy 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

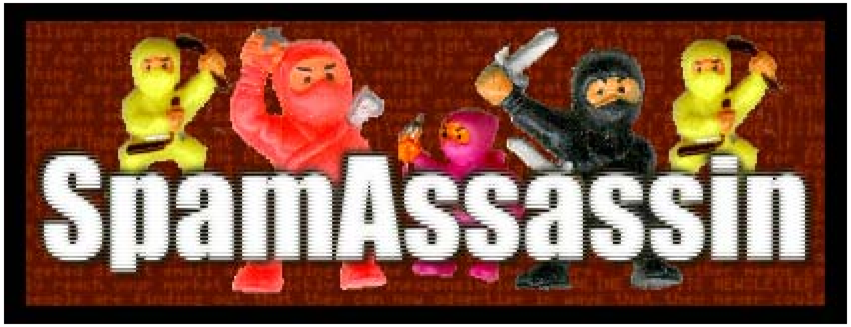
- SMTP runs on top of TCP.
  - Packets are acknowledged.
  - **Source** of packets is known in any successful mail session.
- SMTP servers add the IP address and hostname of every mail server or host involved in the sending process to the e-mail's message header.
- **But,** dynamic IP addresses and large 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.



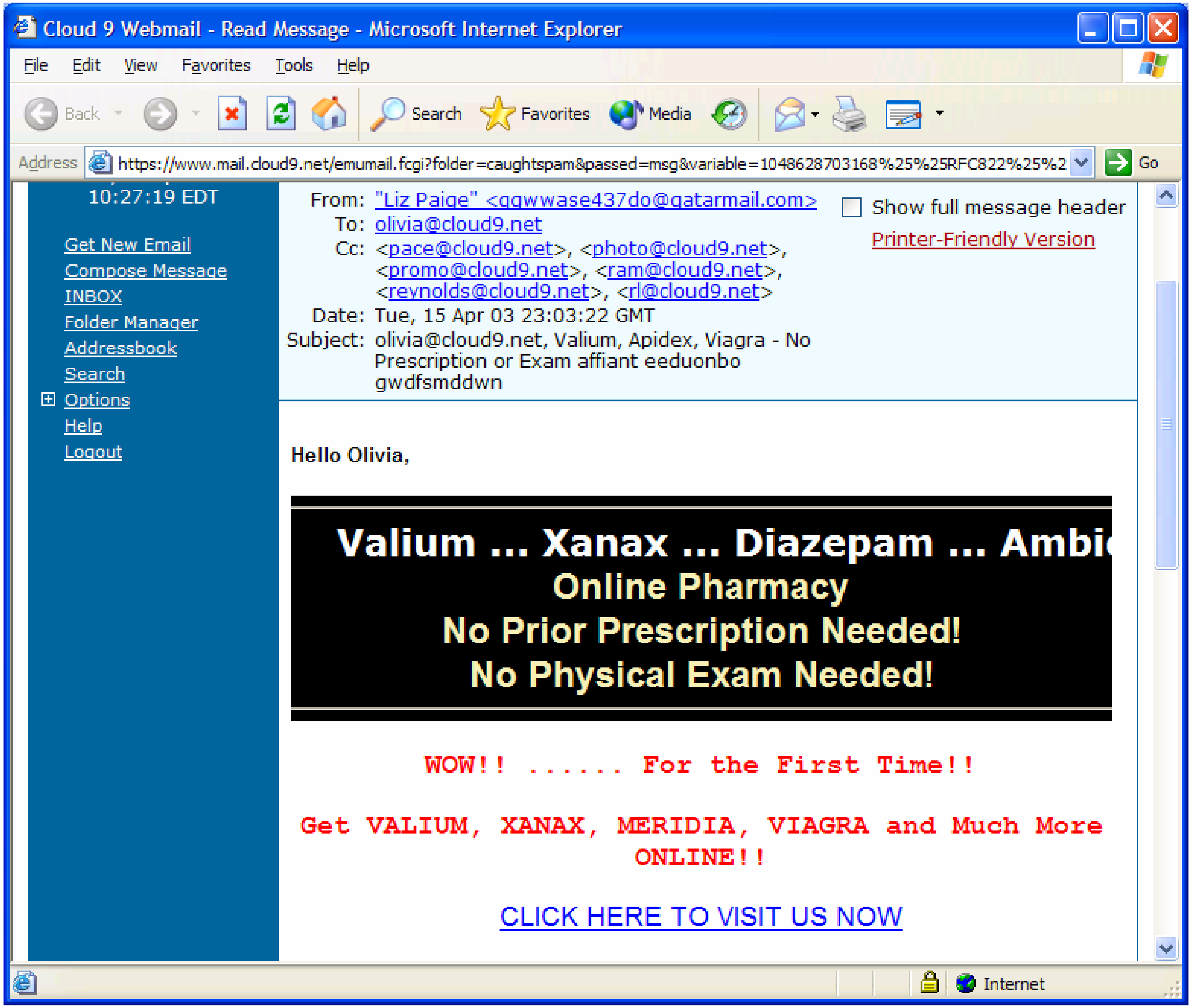
- SpamAssassin is a spam-fighting tool.
- Primary development efforts exist for the **open-source**, UNIX-compatible version. The source code and select Linux binaries are available for free download (for non-commercial use).
- Commercial and Windows-compatible products are available that use the technology.
- 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



Cloud 9 Webmail - Read Message - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back
Search
Favorites
Media
Address
Go
Internet

https://www.mail.cloud9.net/emumail.fcgi?folder=caughtspam&passed=msg&variable=1048628703168%25%25RFC822%25%2

10:27:19 EDT

From: "Liz Paige" <qqwwase437do@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, Apidex, Viagra - No Prescription or Exam affiant eeduonbo gwdfsmddwn

Hello Olivia,

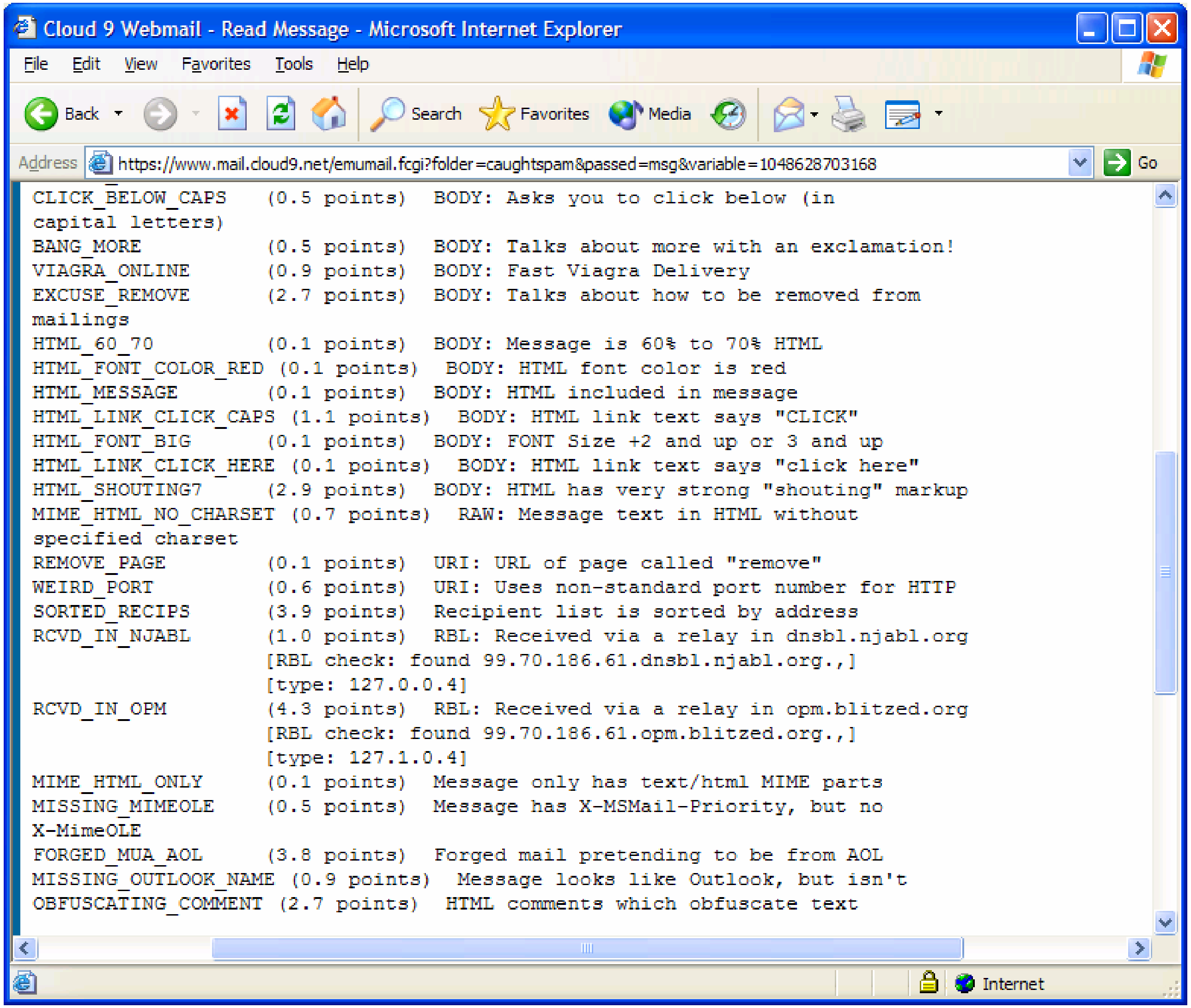
Valium ... Xanax ... Diazepam ... Ambi
Online Pharmacy
No Prior Prescription Needed!
No Physical Exam Needed!

WOW!! ...... For the First Time!!

Get VALIUM, XANAX, MERIDIA, VIAGRA and Much More ONLINE!!

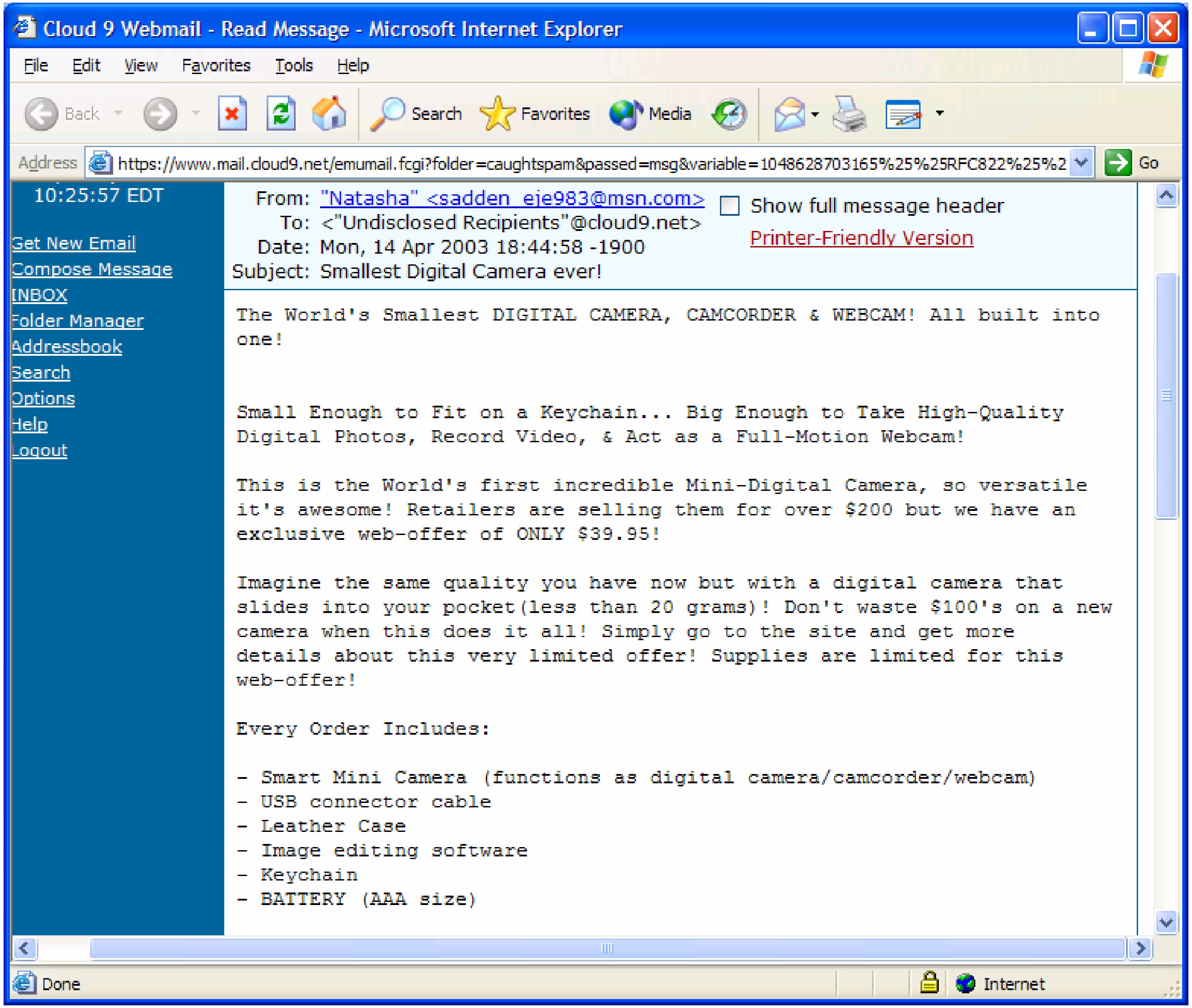
CLICK HERE TO VISIT US NOW

Judging Spam: Results #1


Cloud 9 Webmail - Read Message - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back
Search
Favorites
Media
Address
https://www.mail.cloud9.net/emumail.fcgi?folder=caughtspam&passed=msg&variable=1048628703168
Go
Internet

```
CLICK_BELOW_CAPS    (0.5 points)  BODY: Asks you to click below (in
capital letters)
BANG_MORE           (0.5 points)  BODY: Talks about more with an exclamation!
VIAGRA_ONLINE       (0.9 points)  BODY: Fast Viagra Delivery
EXCUSE_REMOVE       (2.7 points)  BODY: Talks about how to be removed from
mailings
HTML_60_70          (0.1 points)  BODY: Message is 60% to 70% HTML
HTML_FONT_COLOR_RED (0.1 points)  BODY: HTML font color is red
HTML_MESSAGE        (0.1 points)  BODY: HTML included in message
HTML_LINK_CLICK_CAPS (1.1 points)  BODY: HTML link text says "CLICK"
HTML_FONT_BIG       (0.1 points)  BODY: FONT Size +2 and up or 3 and up
HTML_LINK_CLICK_HERE (0.1 points)  BODY: HTML link text says "click here"
HTML_SHOUTING7      (2.9 points)  BODY: HTML has very strong "shouting" markup
MIME_HTML_NO_CHARSET (0.7 points)  RAW: Message text in HTML without
specified charset
REMOVE_PAGE         (0.1 points)  URI: URL of page called "remove"
WEIRD_PORT          (0.6 points)  URI: Uses non-standard port number for HTTP
SORTED_RECIPS       (3.9 points)  Recipient list is sorted by address
RCVD_IN_NJABL       (1.0 points)  RBL: Received via a relay in dnsbl.njabl.org
                    [RBL check: found 99.70.186.61.dnsbl.njabl.org.,]
                    [type: 127.0.0.4]
RCVD_IN_OPM         (4.3 points)  RBL: Received via a relay in opm.blitzed.org
                    [RBL check: found 99.70.186.61.opm.blitzed.org.,]
                    [type: 127.1.0.4]
MIME_HTML_ONLY      (0.1 points)  Message only has text/html MIME parts
MISSING_MIMEOLE     (0.5 points)  Message has X-MSMail-Priority, but no
X-MimeOLE
FORGED_MUA_AOL      (3.8 points)  Forged mail pretending to be from AOL
MISSING_OUTLOOK_NAME (0.9 points)  Message looks like Outlook, but isn't
OBFUSCATING_COMMENT (2.7 points)  HTML comments which obfuscate text
```

Judging Spam: Example #2

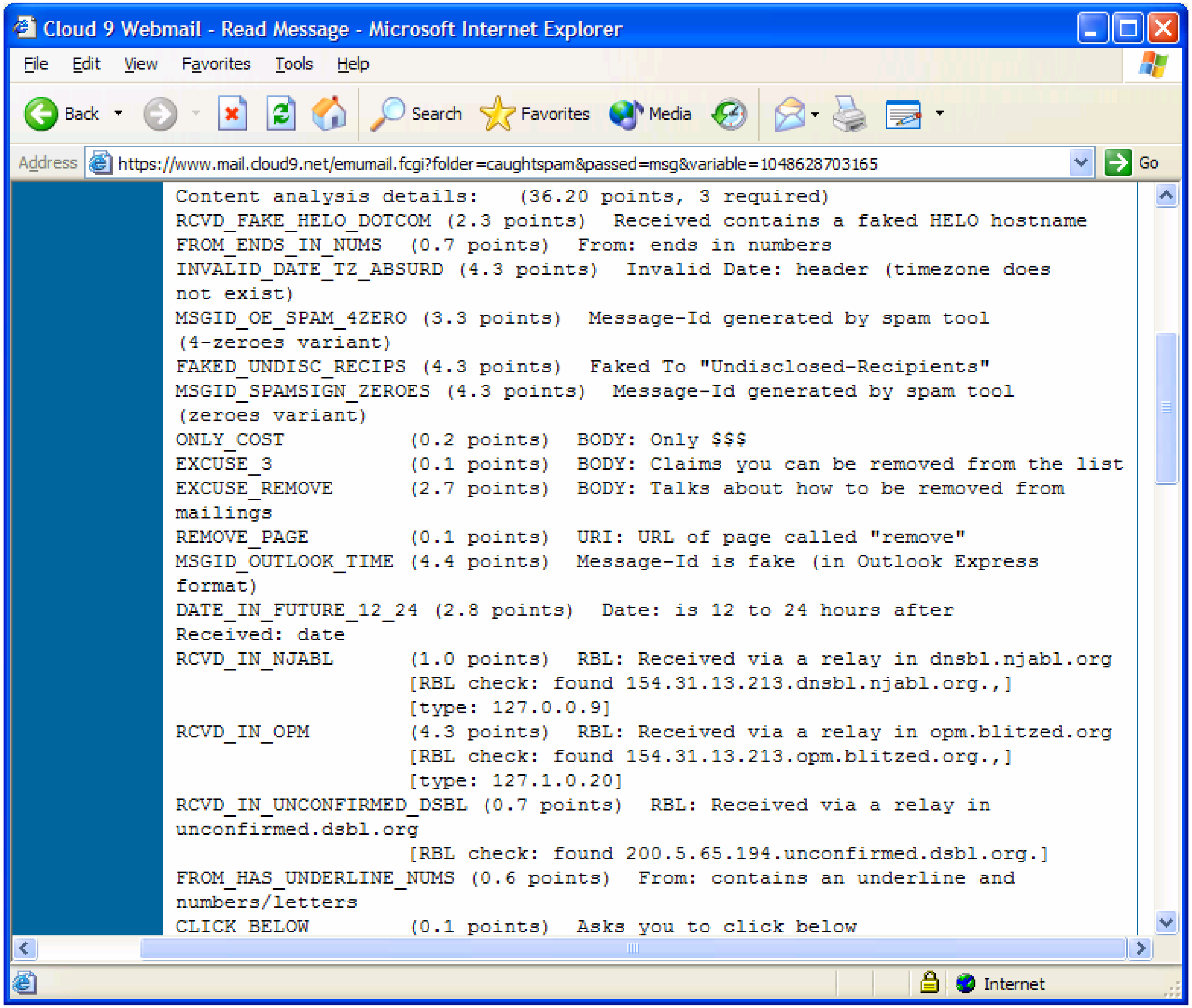


Cloud 9 Webmail - Read Message - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back
Search
Favorites
Media
Address
https://www.mail.cloud9.net/emumail.fcgi?folder=caughtspam&passed=msg&variable=1048628703165%25%25RFC822%25%2
Go
10:25:57 EDT
Get New Email
Compose Message
INBOX
Folder Manager
Addressbook
Search
Options
Help
Logout
□ Show full message header
Printer-Friendly Version
Done
Internet

From: "Natasha" <sadden_eje983@msn.com>
To: <"Undisclosed Recipients"@cloud9.net>
Date: Mon, 14 Apr 2003 18:44:58 -1900
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



Address https://www.mail.cloud9.net/emumail.fcgi?folder=caughtspam&passed=msg&variable=1048628703165

```
Content analysis details:   (36.20 points, 3 required)
RCVD_FAKE_HELO_DOTCOM (2.3 points)  Received contains a faked HELO hostname
FROM_ENDS_IN_NUMS  (0.7 points)  From: ends in numbers
INVALID_DATE_TZ_ABSURD (4.3 points)  Invalid Date: header (timezone does
not exist)
MSGID_OE_SPAM_4ZERO (3.3 points)  Message-Id generated by spam tool
(4-zeroes variant)
FAKED_UNDISC_RECIPS (4.3 points)  Faked To "Undisclosed-Recipients"
MSGID_SPAMSIGN_ZEROES (4.3 points)  Message-Id generated by spam tool
(zeroes variant)
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_PAGE         (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: date
RCVD_IN_NJABL       (1.0 points)  RBL: Received via a relay in dnsbl.njabl.org
                    [RBL check: found 154.31.13.213.dnsbl.njabl.org.,]
                    [type: 127.0.0.9]
RCVD_IN_OPM         (4.3 points)  RBL: Received via a relay in opm.blitzed.org
                    [RBL check: found 154.31.13.213.opm.blitzed.org.,]
                    [type: 127.1.0.20]
RCVD_IN_UNCONFIRMED_DSBL (0.7 points)  RBL: Received via a relay in
unconfirmed.dsbl.org
                    [RBL check: found 200.5.65.194.unconfirmed.dsbl.org.]
FROM_HAS_UNDERLINE_NUMS (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.

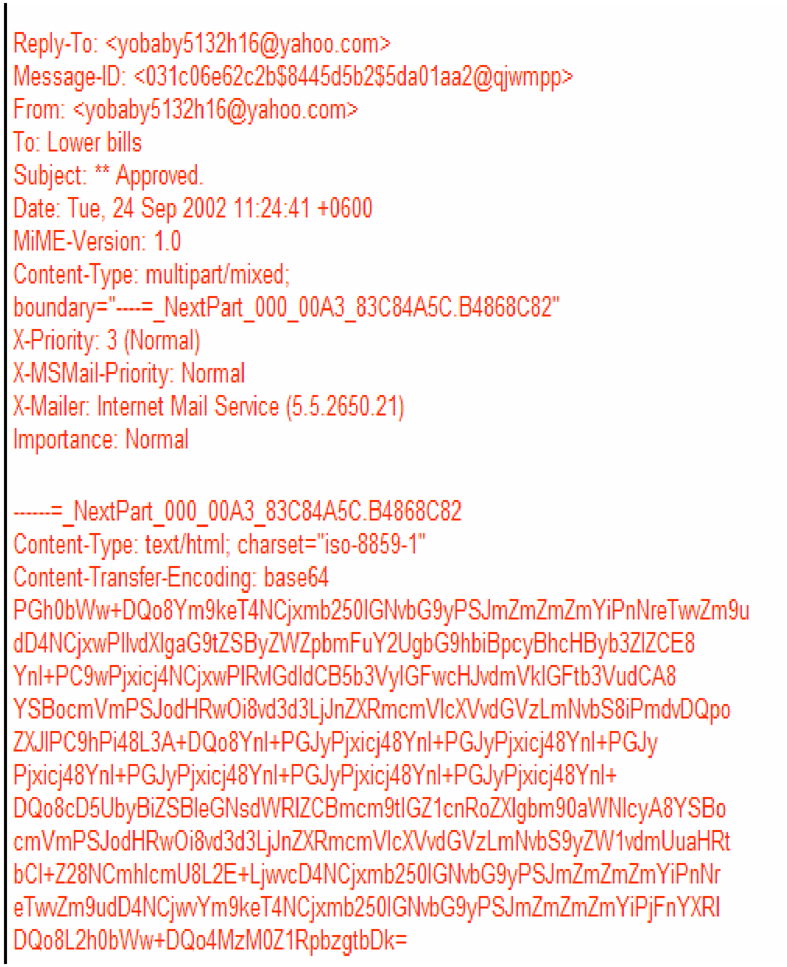
Once identified, the mail can then be optionally tagged as spam for later filtering using the user's own mail user-agent application.

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.

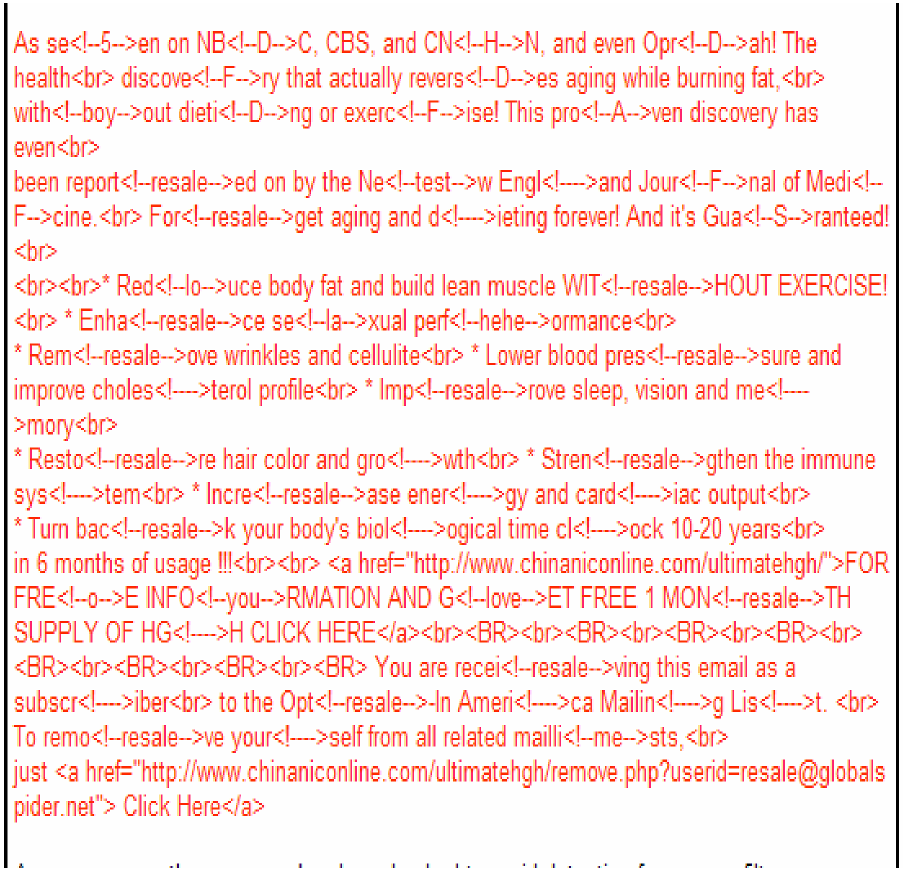
Examples of Tricks

Source: spam-stopper.net



```
Reply-To: <yobaby5132h16@yahoo.com>
Message-ID: <031c06e62c2b$8445d5b2$5da01aa2@qjwmpp>
From: <yobaby5132h16@yahoo.com>
To: Lower bills
Subject: ** Approved.
Date: Tue, 24 Sep 2002 11:24:41 +0600
MiME-Version: 1.0
Content-Type: multipart/mixed;
boundary="----=_NextPart_000_00A3_83C84A5C.B4868C82"
X-Priority: 3 (Normal)
X-MSMail-Priority: Normal
X-Mailer: Internet Mail Service (5.5.2650.21)
Importance: Normal


------=_NextPart_000_00A3_83C84A5C.B4868C82
Content-Type: text/html; charset="iso-8859-1"
Content-Transfer-Encoding: base64
```

```
PGh0bWw+DQo8Ym9keT4NCjxmb250IGNvbG9yPSJmZmZmZmYiPnNreTwvZm9u
dD4NCjxwPllvdXIgaG9tZSByZWZpbmFuY2UgbG9hbiBpcyBhcHByb3ZlZCE8
Ynl+PC9wPjxicj4NCjxwPlRvIGdldCB5b3VyIGFwcHJvdmVkIGFtb3VudCA8
YSBocmVmPSJodHRwOi8vd3d3LjJnZXRmcmVlcXVvdGVzLmNvbS8iPmdvDQpo
ZXJlPC9hPi48L3A+DQo8Ynl+PGJyPjxicj48Ynl+PGJyPjxicj48Ynl+PGJy
Pjxicj48Ynl+PGJyPjxicj48Ynl+PGJyPjxicj48Ynl+PGJyPjxicj48Ynl+
DQo8cD5UbyBiZSBleGNsdWRlZCBmcm9tIGZ1cnRoZXIgbm90aWNlcyA8YSBo
cmVmPSJodHRwOi8vd3d3LjJnZXRmcmVlcXVvdGVzLmNvbS9yZW1vdmUuaHRt
bCI+Z28NCmhlcmU8L2E+LjwvcD4NCjxmb250IGNvbG9yPSJmZmZmZmYiPnNr
eTwvZm9udD4NCjwvYm9keT4NCjxmb250IGNvbG9yPSJmZmZmZmYiPjFnYXRl
DQo8L2h0bWw+DQo4MzM0Z1RpbzgtbDk=
```



```
As se<!--5-->en on NB<!--D-->C, CBS, and CN<!--H-->N, and even Opr<!--D-->ah! The
health<br> discove<!--F-->ry that actually revers<!--D-->es aging while burning fat,<br>
with<!--boy-->out dieti<!--D-->ng or exerc<!--F-->ise! This pro<!--A-->ven discovery has
even<br>
been report<!--resale-->ed on by the Ne<!--test-->w Engl<!---->and Jour<!--F-->nal of Medi<!--
F-->cine.<br> For<!--resale-->get aging and d<!---->ieting forever! And it's Gua<!--S-->ranteed!
<br>
<br><br>* Red<!--lo-->uce body fat and build lean muscle WIT<!--resale-->HOUT EXERCISE!
<br> * Enha<!--resale-->ce se<!--la-->xual perf<!--hehe-->ormance<br>
* Rem<!--resale-->ove wrinkles and cellulite<br> * Lower blood pres<!--resale-->sure and
improve choles<!---->terol profile<br> * Imp<!--resale-->rove sleep, vision and me<!---
->mory<br>
* Resto<!--resale-->re hair color and gro<!---->wth<br> * Stren<!--resale-->gthen the immune
sys<!---->tem<br> * Incre<!--resale-->ase ener<!---->gy and card<!---->iac output<br>
* Turn bac<!--resale-->k your body's biol<!---->ogical time cl<!---->ock 10-20 years<br>
in 6 months of usage !!!<br><br> <a href="http://www.chinaniconline.com/ultimatehgh/">FOR
FRE<!--o-->E INFO<!--you-->RMATION AND G<!--love-->ET FREE 1 MON<!--resale-->TH
SUPPLY OF HG<!---->H CLICK HERE</a><br><BR><br><BR><br><BR><br><BR><br>
<BR><br><BR><br><BR><br><BR> You are recei<!--resale-->ving this email as a
subscr<!---->iber<br> to the Opt<!--resale-->-In Ameri<!---->ca Mailin<!---->g Lis<!---->t. <br>
To remo<!--resale-->ve your<!---->self from all related mailli<!--me-->sts,<br>
just <a href="http://www.chinaniconline.com/ultimatehgh/remove.php?userid=resale@globals
pider.net"> Click Here</a>
```

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.

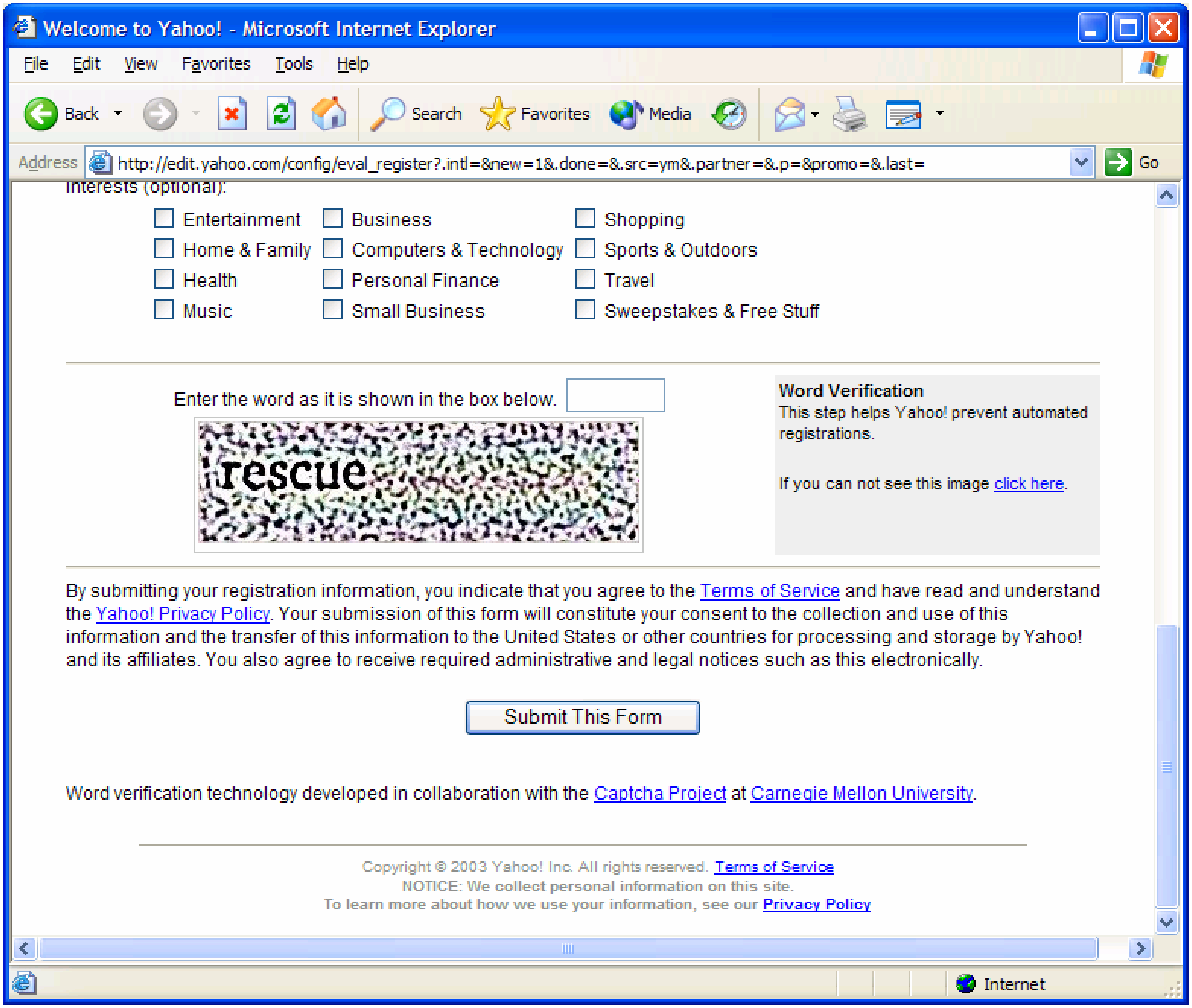
(86,400 sec/day) / (10 sec/msg) = 8640 msgs/day
Hotmail receives 1 billion msgs / day
-> Would need 125,000 computers
Up-front capital cost for all of Hotmail's spam:
~ $150M. The spammers can't afford it!
(-- C. Dwork, Microsoft)

Prove You are a Human

- CAPTCHA: **C**ompletely **A**utomated **P**ublic **T**uring test for telling **C**omputers and **H**umans **A**part
- 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


Welcome to Yahoo! - Microsoft Internet Explorer
File
Edit
View
Favorites
Tools
Help
Back
Search
Favorites
Media
Address
http://edit.yahoo.com/config/eval_register?.intl=&new=1&.done=&.src=ym&.partner=&.p=&promo=&.last=
Go
rescue
Submit This Form
Internet

Interests (optional):

□ Entertainment
□ Home & Family
□ Health
□ Music
□ Business
□ Computers & Technology
□ Personal Finance
□ Small Business
□ Shopping
□ Sports & Outdoors
□ Travel
□ Sweepstakes & Free Stuff

Enter the word as it is shown in the box below. ▭

Word Verification
This step helps Yahoo! prevent automated registrations.

If you can not see this image click here.

By submitting your registration information, you indicate that you agree to the Terms of Service and have read and understand the Yahoo! Privacy Policy. Your submission of this form will constitute your consent to the collection and use of this information and the transfer of this information to the United States or other countries for processing and storage by Yahoo! and its affiliates. You also agree to receive required administrative and legal notices such as this electronically.

Word verification technology developed in collaboration with the Captcha Project at Carnegie Mellon University.

Copyright © 2003 Yahoo! Inc. All rights reserved. Terms of Service
NOTICE: We collect personal information on this site.
To learn more about how we use your information, see our Privacy Policy

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.