Email in the Early 1980s

- Mail relay: forwards mail to next hop
- Sender path includes path through relays
Email Spoofing

◆ email sent via SMTP
  • No built-in authentication

◆ “MAIL FROM” field set by sender
  • Classical example of improper input validation

◆ Recipient’s email server only sees IP address of the direct peer from which it received message
Open Relays

◆ SMTP relay forwards mail to destination
  1. Connects via SMTP (TCP port 25)
  2. Sends list of recipients via “RCPT TO” command
  3. Sends email body (once for all recipients!)
  4. Relay delivers message
◆ Honest relay adds correct “Received:” header revealing source IP
◆ Malicious relay does not have to...
# SMTP Header

<table>
<thead>
<tr>
<th>Header</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>To:</td>
<td>E-mail address(es) of primary recipient(s)</td>
</tr>
<tr>
<td>Cc:</td>
<td>E-mail address(es) of secondary recipient(s)</td>
</tr>
<tr>
<td>Bcc:</td>
<td>E-mail address(es) for blind carbon copies</td>
</tr>
<tr>
<td>From:</td>
<td>Person or people who created the message</td>
</tr>
<tr>
<td>Sender:</td>
<td>E-mail address of the actual sender</td>
</tr>
<tr>
<td>Received:</td>
<td>Line added by each transfer agent along the route</td>
</tr>
<tr>
<td>Return-Path:</td>
<td>Can be used to identify a path back to the sender</td>
</tr>
</tbody>
</table>
A Closer Look at Spam

Received: by 10.78.68.6 with SMTP id q6cs394373hua;
    Mon, 12 Feb 2007 06:43:30 -0800 (PST)
Received: by 10.90.113.18 with SMTP id l18mr17307116agc.1171291410432;
    Mon, 12 Feb 2007 06:43:30 -0800 (PST)
Return-Path: <wvnlwee@aviva.ro>
Received: from onelinkpr.net ([203.169.49.172])
    by mx.google.com with ESMTP id 30si11317474agc.2007.02.12.06.43.18;
    Mon, 12 Feb 2007 06:43:30 -0800 (PST)
Received-SPF: neutral (google.com: 203.169.49.172 is neither permitted nor
    denied by best guess record for domain of wvnlwee@aviva.ro)
Message-ID: <20050057765.stank.203.169.49.172@ASAFTU>
From: "Barclay Morales" <wvnlwee@aviva.ro>
To: <raykwatts@gmail.com>
Subject: You can order both Viagra and Cialis.

Inserted by relays

Bogus!

Puerto Rico
Mongolia
Why Hide Sources of Spam?

◆ Many email providers blacklist servers and ISPs that generate a lot of spam
  • Use info from spamhaus.org, spamcop.net
◆ Real-time blackhole lists stop 15-25% of spam at SMTP connection time
  • Over 90% after message body checks
◆ Spammers’ objective: evade blacklists
  • Botnets come very handy!
Thin Pipe / Thick Pipe

- Spam source is a high-speed broadband host (HSB) and controls a low-speed zombie (LSZ)

- Hides IP address of HSB; LSZ is blacklisted
Open HTTP Proxies

- Web cache (HTTP/HTTPS proxy), e.g., squid

URL: HTTPS://xyz.com

CONNECT xyz.com:443

ClientHello

ServerHello

Squid
web cache

ClientHello

ServerHello

Web server

Why is port 25 enabled, anyway?

To spam: CONNECT <Victim’s IP> 25, then issue SMTP Commands

- i.e., start with http://www.victim.com:25
- Squid becomes a mail relay
Send-Safe Spam Tool
Open Relays vs. Open Proxies

◆ Open proxy
  • Spammer must send message to each recipient through the proxy

◆ Open relay
  • Takes a list of addresses and sends to all
  • Can host an open relay on a zombie

◆ Listing services for open proxies and relays
  • http://www.multiproxy.org/
  • http://www.stayinvisible.com/
  • http://www.openproxies.com/ ($20/month)
Spam Surge

“Two years from now, spam will be solved”
- Bill Gates, January 2004
Bobax Worm

◆ Infects machines with high bandwidth
  • Exploits MS LSASS (Local Security Authority Subsystem Service) buffer overflow vulnerability

◆ Slow spreading (thus hard to detect)
  • Upon manual command from operator, randomly scans for vulnerable machines

◆ Installs hacked open relay on infected zombies
  • Once spam zombie added to blacklist, spreads to another machine

http://www.m86security.com/labs/spambotitem.asp?article=901
IP Blacklisting Not Enough

[Ramachandran, Feamster]

More than half of client IPs appear less than twice
## Distribution Across Domains

[Ramachandran, Feamster]

<table>
<thead>
<tr>
<th>AS Number</th>
<th># Spam</th>
<th>AS Name</th>
<th>Primary Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>766</td>
<td>580559</td>
<td>Korean Internet Exchange</td>
<td>Korea</td>
</tr>
<tr>
<td>4134</td>
<td>560765</td>
<td>China Telecom</td>
<td>China</td>
</tr>
<tr>
<td>1239</td>
<td>437660</td>
<td>Sprint</td>
<td>United States</td>
</tr>
<tr>
<td>4837</td>
<td>236434</td>
<td>China Network Communications</td>
<td>China</td>
</tr>
<tr>
<td>9318</td>
<td>225830</td>
<td>Hanaro Telecom</td>
<td>Japan</td>
</tr>
<tr>
<td>32311</td>
<td>198185</td>
<td>JKS Media, LLC</td>
<td>United States</td>
</tr>
<tr>
<td>5617</td>
<td>181270</td>
<td>Polish Telecom</td>
<td>Poland</td>
</tr>
<tr>
<td>6478</td>
<td>152671</td>
<td>AT&amp;T WorldNet Services</td>
<td>United States</td>
</tr>
<tr>
<td>19262</td>
<td>142237</td>
<td>Verizon Global Networks</td>
<td>United States</td>
</tr>
<tr>
<td>8075</td>
<td>107056</td>
<td>Microsoft</td>
<td>United States</td>
</tr>
<tr>
<td>7132</td>
<td>99585</td>
<td>SBC Internet Services</td>
<td>United States</td>
</tr>
<tr>
<td>6517</td>
<td>94600</td>
<td>Yipes Communications, Inc.</td>
<td>United States</td>
</tr>
<tr>
<td>31797</td>
<td>89698</td>
<td>GalaxyVisions</td>
<td>United States</td>
</tr>
<tr>
<td>12322</td>
<td>87340</td>
<td>PROXAD AS for Proxad ISP</td>
<td>France</td>
</tr>
<tr>
<td>3356</td>
<td>87042</td>
<td>Level 3 Communications, LLC</td>
<td>United States</td>
</tr>
<tr>
<td>22909</td>
<td>86150</td>
<td>Comcast Cable Corporation</td>
<td>United States</td>
</tr>
<tr>
<td>8151</td>
<td>81721</td>
<td>UniNet S.A. de C.V.</td>
<td>Mexico</td>
</tr>
<tr>
<td>3320</td>
<td>79987</td>
<td>Deutsche Telekom AG</td>
<td>Germany</td>
</tr>
<tr>
<td>7018</td>
<td>74320</td>
<td>AT&amp;T WorldNet Services</td>
<td>United States</td>
</tr>
<tr>
<td>4814</td>
<td>74266</td>
<td>China Telecom</td>
<td>China</td>
</tr>
</tbody>
</table>
Most Bots Send Little Spam

Most bot IP addresses send very little spam, regardless of how long they have been spamming…
Where Does Spam Come From?

- IP addresses of spam sources are widely distributed across the Internet
  - In tracking experiments, most IP addresses appear once or twice; 60-80% not reachable by traceroute
- Vast majority of spam originates from a small fraction of IP address space
  - Same fraction that most legitimate email comes from
- Spammers exploit routing infrastructure
  - Create short-lived connection to mail relay, then disappear
  - Hijack a large chunk of unallocated “dark” space

[Ramachandran, Feamster]
Spambot Behavior

◆ Strong correlation with Bobax infections
◆ Most are active for a very short time
  • 65% of Bobax victims send spam once; 3 out of 4 are active for less than 2 minutes
◆ 99% of bots send fewer than 100 messages regardless of their lifetime
◆ 95% of bots already in one or more blacklists
  • Cooperative detection works, but ...
  • Problem: false positives!
  • Problem: short-lived hijacks of dark address space

[Ramachandran, Feamster]
Stormbot Spam Architecture

- Spam templates
  - Custom macro language
  - Polymorphic content
- Dictionaries
  - Email addresses
  - Subject lines
- Worker bots generate unique messages for each address, try to deliver, report results to their proxies

Figure 2: The Storm spam campaign dataflow (Section 3.3) and our measurement and rewriting infrastructure (Section 4). (1) Workers request spam tasks through proxies, (2) proxies forward spam workload responses from master servers, (3) workers send the spam and (4) return delivery reports. Our infrastructure infiltrates the C&C channels between workers and proxies.

[Kanich, Kreibich, Levchenko et al.]
Major Spambots

http://www.marshal.com/trace/traceitem.asp?article=615
McColo

- McColo was a San Jose-based hosting provider
- Hosted command-and-control servers of the biggest spam botnets
  - Rustock, Srizbi, Pushdo/Cutwail, others
-Disconnected by upstream providers on Nov 11, 2008 ⇒ 75% reduction of spam worldwide
Srizbi

◆ Rootkit + sophisticated spam mailer

◆ 500K zombies, 60 billion spam messages daily
  • More than half of all spam worldwide

◆ After McColo takedown, fail-safe code inside bots started generating names of backup domains
  • ypoouaypu.com, oryitugf.com, prpoqpsy.com ...
  • Botmasters regained control by registering these domains (through a Russian registrar) and hosting new C&C servers in Estonia – shut down later
SPAM Countermeasures

◆ Legal

◆ Technical
CAN-SPAM Act (passed in 2003)

◆ Legal solution to the problem
  • Bans email harvesting, misleading header information, deceptive subject lines, use of proxies
  • Requires opt-out and identification of advertising
  • Imposes penalties (up to $11K per violation)

◆ FTC report on effectiveness (Dec 2005)
  • 50 cases pursued in the US
  • No impact on spam originating outside the US (60%)
  • Open relays hosted on botnets make it difficult to collect evidence

http://www.ftc.gov/spam
SPF (Sender Policy Framework)

Spammers put popular domains (e.g., hotmail.com) as FROM sources ⇒ hotmail flooded by bounced responses

What if spammer gets a throwaway domain?

Used by AOL and others
Domain Keys (DKIM)

1. Email is sent with the FROM domain and includes a digital signature in an X-header.

2. The Recipient’s Email Gateway does a DNS query on example.com and is returned a public key.

3. Recipient Email Gateway attempts to verify the digital signature from Step 1 using the public key returned from the DNS server in Step 2.

4. If signature can be verified, the FROM domain is authentic, otherwise the FROM domain has been spoofed.

Sender’s server has to **sign** email.

DNS provides verification key to the recipient.

From Yahoo
S/MIME

Sender obtains public-key certificate

Sender’s server has to **sign** email; includes certified verification key
Graylists

◆ Recipient’s mail server records (sender email, recipient email, peer IP) triple in its database
  • Each triple kept for 3 days (configuration parameter)
◆ First time (triple not in DB): 421 reply “Busy”
  • Records triple in the database
◆ Second time (after 5 minutes): let email pass
◆ What is this defense based on?
◆ Easily spoofable, but works against many current spammers
Puzzles and CAPTCHAs

◆ Generic defenses against spam and DoS
◆ Basic idea: sender must solve a “puzzle” before email or connection request is accepted
  • Takes effort to solve, but solution easy to check
  • Sender has to “pay” in computation time
    – Example (Hashcash): find collision in a short hash
◆ CAPTCHA: prove that the sender is human
  • Solve a “reverse Turing test”
  • Only in application layer (e.g., Web)
◆ Both are difficult to deploy (why?)
Worst CAPTCHA Ever?

http://depressedprogrammer.wordpress.com/2008/04/20/worst-captcha-ever/
Solve CAPTCHAs for Fun and Profit

◆ Third-world “data entry specialists” will solve CAPTCHAs for 60 cents an hour
Gone in Six Seconds

◆ Spammers like to create a large number of Gmail and Hotmail accounts, use them to send spam
  • DKIM and SPF don’t help (why?)
  • But CAPTCHAs do (how?)
◆ Botnet = massive distributed computing platform
  • Use them to solve CAPTCHAs
◆ 2008: 6 seconds per CAPTCHA, 10-15% success
◆ Now: 20 seconds per CAPTCHA, 12-20% success
  • ... after Microsoft upgraded their CAPTCHA system