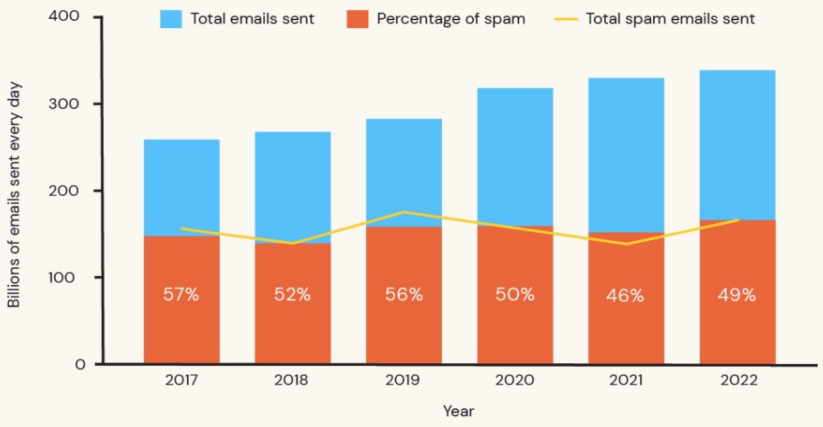

**Recommended READ:**
**Anirudh Ramachandran and Nick Feamster,**
**Understanding the network-level behavior of spammers**
**ACM SIGCOMM 2006, pp. 291-302**

1

1

## Some SPAM stats

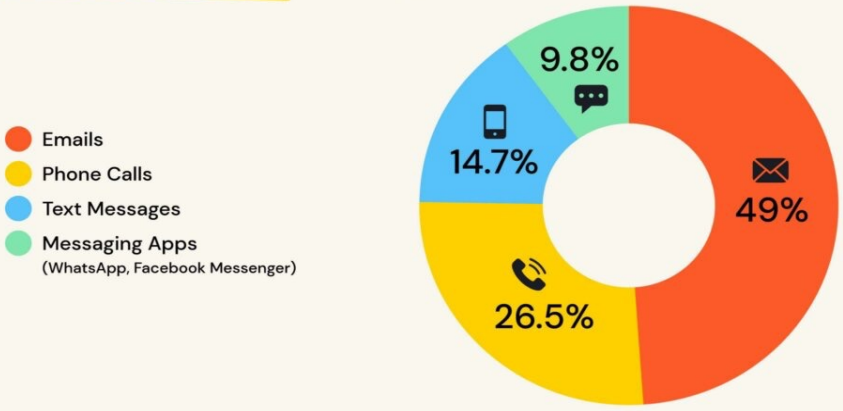
Number of emails vs spam emails sent every day




2

## Some SPAM stats

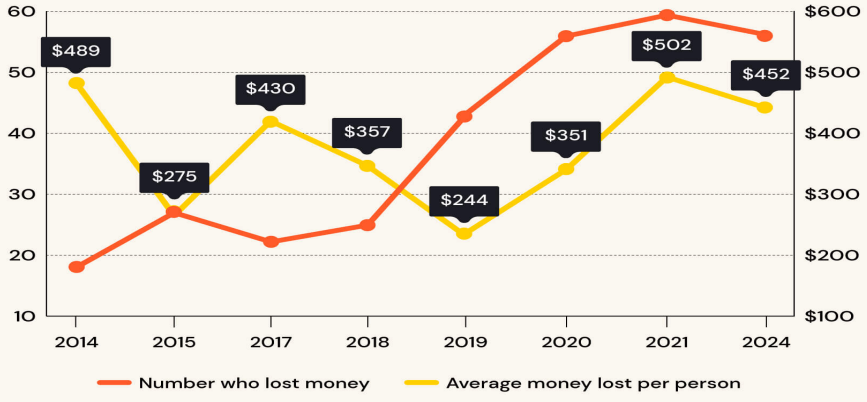
The channels people most commonly receive spam through




3

## Some SPAM stats

Number of Americans who lost money to scam calls (millions)

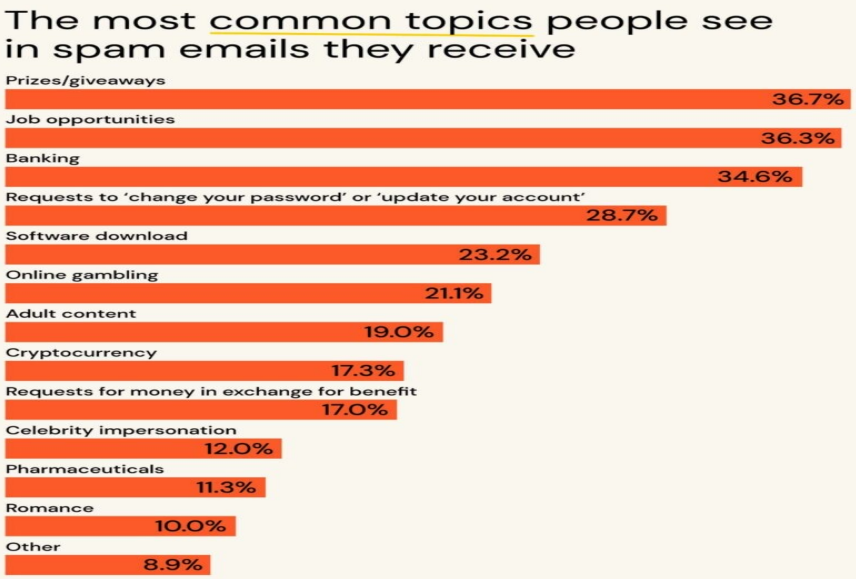



Read the full report at emailtooltester.com/en/blog/spam-statistics

emailtooltester

4

# Some SPAM stats




5

5

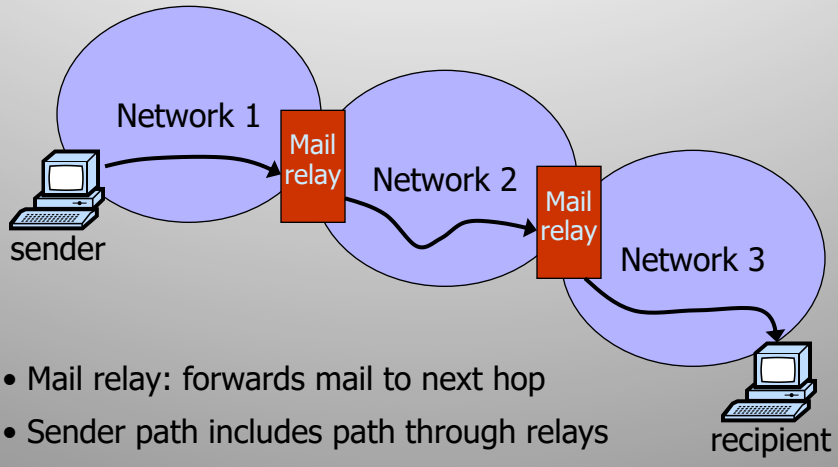
# Some SPAM stats

- **160 billion spam emails are sent every day**, with 46% of the 347 billion daily emails sent, considered spam (numbers recorded for 2023).
- The **majority of people (96.8%)** have received **spam messages** in some form.
- The **U.S. sends the most spam emails,** with **8 billion per day** on average followed by **China with 7.6 billion** per day.
- The **most common topic** of spam emails is **prizes and giveaways**, followed by **job opportunities**, and **banking**.
- Over **two-thirds (68.8%)** of people who had received spam and/or phishing messages reported their **mental health being impacted** at least a little as a result.
- **Financial institutions** are most commonly targeted by business phishing attacks, with **27.7% of scam messages** being received by companies in this sector.
- **Delivery services** were the most common subject of spam text messages in the U.S. in the first half of 2023, with over **1.1 billion scam texts relating to this topic**.
- **Men aged 65 and over** receive the **most spam calls** with an average of **35.5 calls per month**, while **men aged 18-34** receive the **fewest (14.2)**.
- The ten countries that send the most spam emails **produce 2,184 metric tonnes of CO2 every day, or 797,160 over the course of one year** from spam emails alone. In one day, that's equivalent to **5.3 million miles** driven in a conventional gas car, or **1.9 billion miles** over the course of a year.
- **Over three-quarters (77%)** of people who have fallen victim to an **AI phone scam have lost money to the scammers**.

6

6

# Email (since early 1980s)

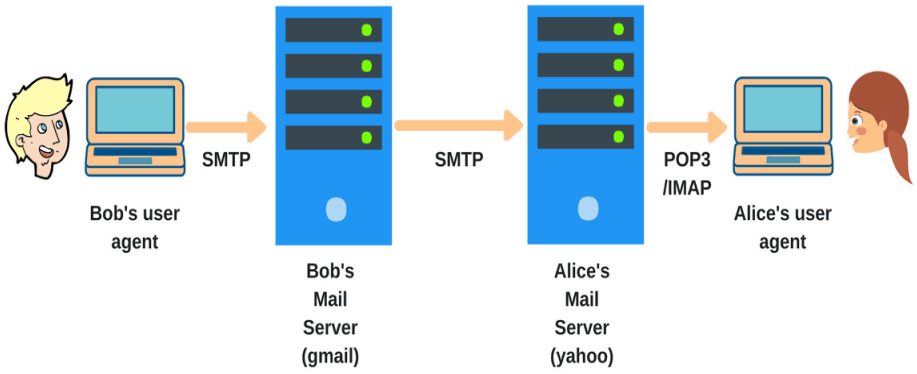



- Mail relay: forwards mail to next hop
- Sender path includes path through relays

7

7

# Email Example




8

8

# SMTP Ports: Important Details

**Port 25: The standard port**

- Primary message transmission channel.
- Came out of a 1982 request from USC/ISI to the Internet Engineering Task Force (IETF) – overlords of all Internet protocols.
- Favored by spammers because of its support for open relay.
- Used primarily for SMTP relaying: moving email between email servers.
- **Modern SMTP email clients (MS Outlook, Apple Mail, Thunderbird, etc.) shouldn't use port 25.**
- **Usually blocked by residential ISPs and Cloud Hosting Providers, to curb spam relayed from compromised computers or servers.**
  - Unless you're specifically managing a mail server, you should have no traffic traversing this port on your computer or server.

**Port 587: The default port**

- Introduced in 1998, after port 25 started to get spammy (Internet RFC 2476) -- proposed a split between traditional message submission and message relaying.
- RFC 2476 requires initial message submission on port 587. Relaying is still on port 25.
- Email client to email server must use SMTP port 587 as the default port.
- Can be coupled with TLS (via STARTTLS option) to ensure better security

9

9

# Open Relays

- SMTP relay forwards mail to destination
  1. Connects via SMTP:
     - TCP port 25 (legacy) or 587 (preferred)
  2. Sends list of recipients via "RCPT TO:" command
  3. Sends email body (once for all recipients!)
- Honest relay adds correct "Received:" header field revealing source IP
- Malicious/hacked relay does not…

10

10

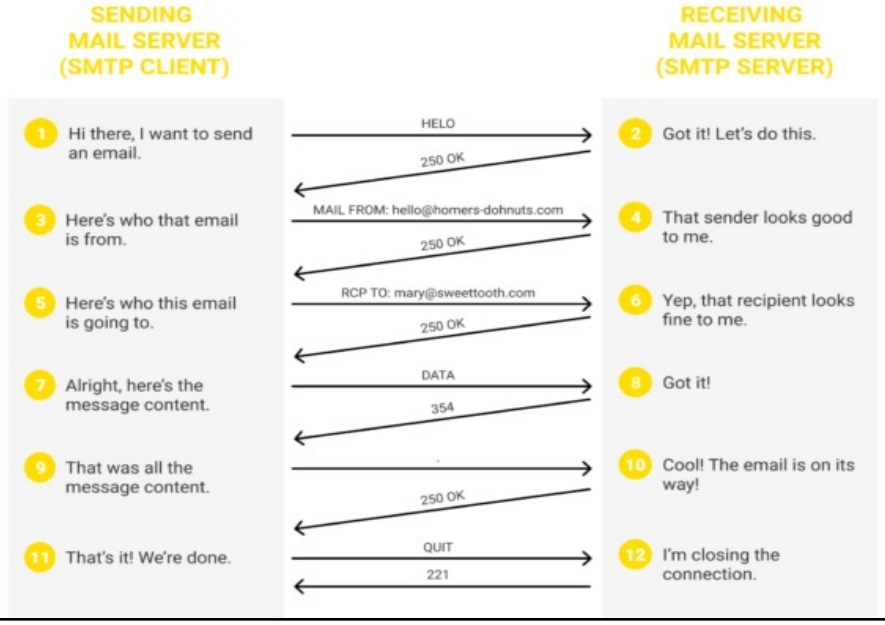
# Email Spoofing

- Email messages are 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

11

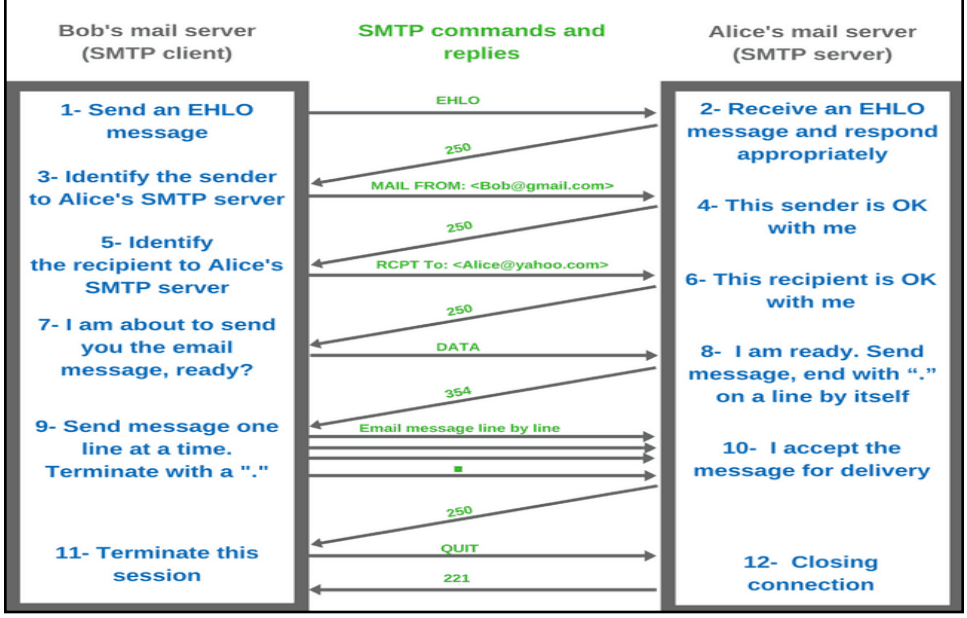
11

## Language of email ☺ → SMTP (usually port 25)



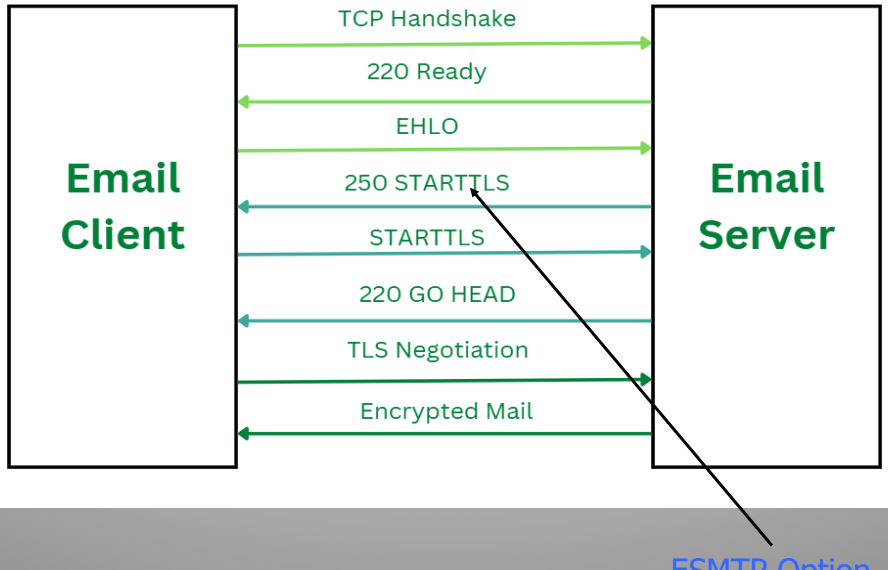

12

## Language of email → ESMTP (usually port 587)




13

## ESMTP + STARTTLS




14

14

## SMTP Header Fields

| Header | Meaning |
|---|---|
| To: | E-mail address(es) of primary recipient(s) |
| Cc: | E-mail address(es) of secondary recipient(s) |
| Bcc: | E-mail address(es) for blind carbon copies |
| From: | Person or people who created the message |
| Sender: | E-mail address of the actual sender |
| Received: | Line added by each transfer agent along the route |
| Return-Path: | Can be used to identify a path back to the sender |

15

15

## A Closer Look at Spam

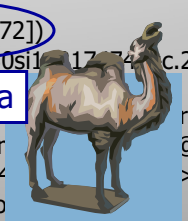
Inserted by relays

Received: by 10.78.68.6 hua;
Mon, 12 Feb 2007 06:43:30 -0800 (PST)
Received: by 10 MTP id l18mr17307116agc.1171291410432;
Mon 43:30 -0800 (PST)
Return-Path: <
Received: from onelinkpr.net ([203.169.49.172])
by mx.google.com with ESMTP id 30si1 c.2007.02.12.06.43.18;

Bogus!

Puerto Rico

Mongolia



Rec r permitted nor
by best guess record for dor @aviva.ro)
Mes 0050057765.stank.203.169.4 >
From: "Barclay Morales" <wvnlwee@aviva.ro
To: <raykwatts@gmail.com>
Subject: You can order both Viagra and Cialis.

16

16

# Why Hide Sources of Spam?

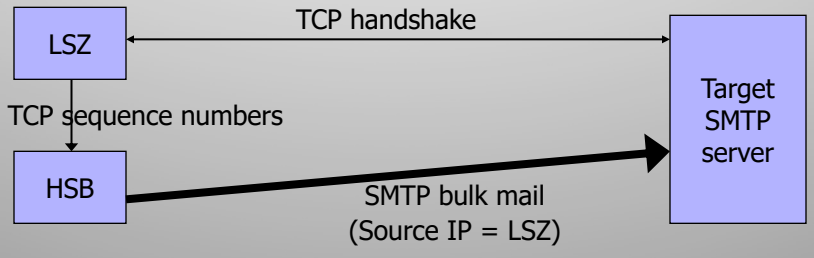
- Many email providers blacklist servers and ISPs that generate a lot of spam
  - Use info from spamhaus.org, spamcop.net
- Real-time blacklists stop 15-25% of spam at SMTP connection time
  - Over 90% after message body checks
  - However, usually BW is already consumed by then
- Spammers' objective: evade blacklists
  - Botnets come very handy!

17

17

# Thin Pipe / Thick Pipe SPAM

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

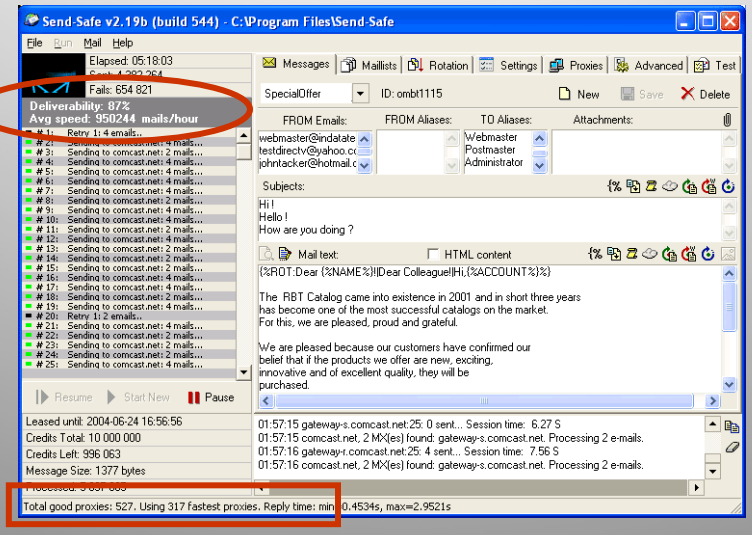



- Hides IP address of HSB; LSZ is blacklisted
- HSB goes on to the next LSZ

18

18

# Send-Safe Spam Tool




19

19

# Open Relays vs. Open Proxies

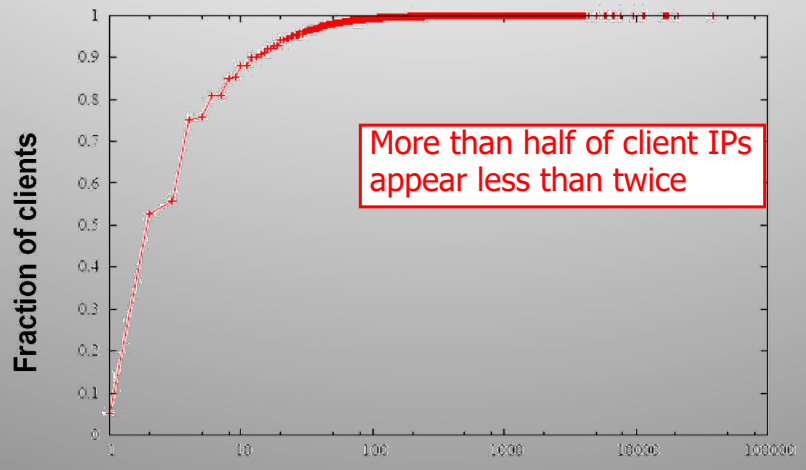
- Open proxy
  - Spammer must send message to each recipient through the proxy
- Open relay
  - Takes a list of addresses and sends to all
  - Can host an open relay on a zombie
- Listing services for open proxies and relays
  - http://www.multiproxy.org/ (recently defunct)
    http://www.stayinvisible.com/ (active
    http://www.openproxies.com/ (active)

20

20

# IP Blacklisting Not Enough

[Ramachandran, Feamster]




21

21

# Distribution Across Domains

[Ramachandran, Feamster]

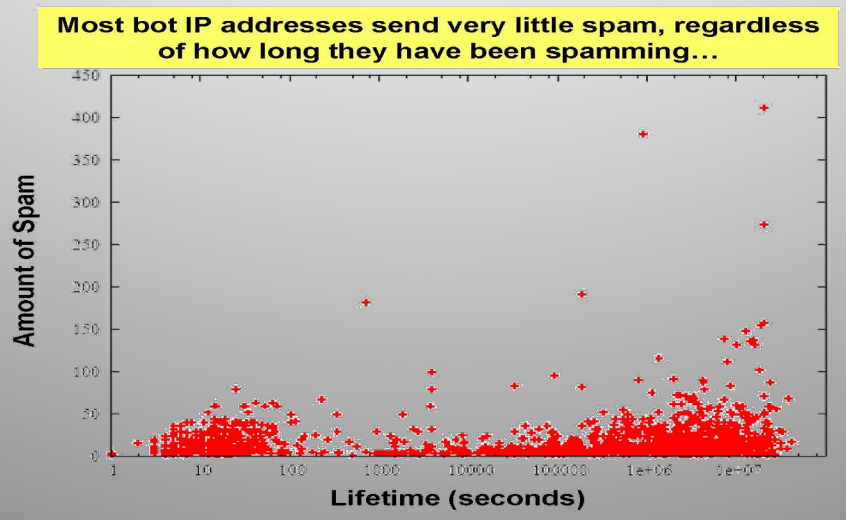
| *AS Number* | *# Spam* | *AS Name* | *Primary Country* |
|---|---|---|---|
| 766 | 580559 | Korean Internet Exchange | Korea |
| 4134 | 560765 | China Telecom | China |
| 1239 | 437660 | Sprint | United States |
| 4837 | 236434 | China Network Communications | China |
| 9318 | 225830 | Hanaro Telecom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

22

22

# Most Bots Send Little Spam

[Ramachandran, Feamster]




23

23

# Where Does Spam Come From?

[Ramachandran, Feamster]

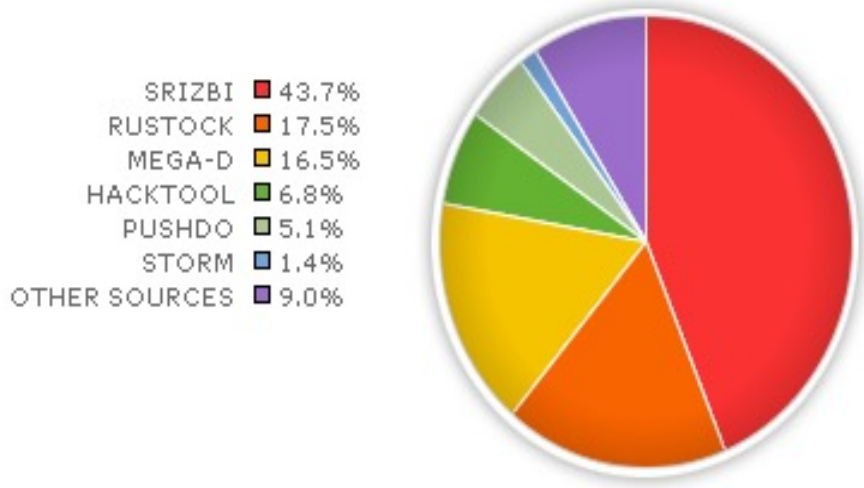
- IP addresses of spam sources are widely distributed across the Internet
  - In tracking experiments, most IP addresses appear once or twice; 60-80% not reachable by traceroute
- Vast majority of spam originates from a small fraction of IP address space
  - Same fraction that most legitimate email comes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

24

24

# Major (Historical) Spambots

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




27

27

# McColo


- McColo was a San Jose-based hosting provider – not a bot itself
- 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

28

28

## 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
  - ypouaypu.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

29

29

## Rustock

- Responsible for 40% of all spam in 2010
- Between 1 and 2.5 million infected computers
  - Up to 240,000 messages daily from each host
- Based on a fairly elaborate rookit
- C&C servers taken down on March 16, 2011
  - Investigation by Microsoft, Pfizer, FireEye, and security researchers from the University of Washington
  - "John Doe" lawsuit against botnet operators
  - Coordinated seizure of C&C servers in the US
  - 33% decline in spam afterwards

slide 30

30

## SPAM Countermeasures

- Legal
- Technical

31

31

## CAN-SPAM Act (US Law, passed in 2003, still in force)

http://www.ftc.gov/spam

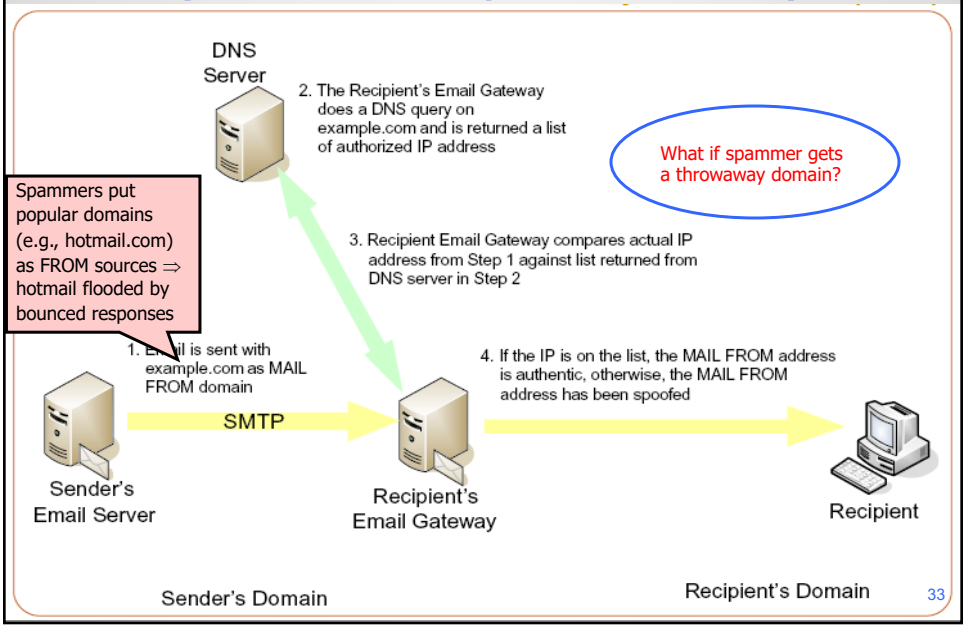
- Legal solution to the SPAM problem
  - Bans email harvesting, misleading header information, deceptive subject lines, use of proxies
  - Requires opt-out and identification of advertising
  - Imposes penalties (up to $11K per violation)
- FTC report on effectiveness 2 years later (Dec 2005)
  - 50 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

This Act establishes requirements for those who send unsolicited commercial email. The Act bans false or misleading header information and prohibits deceptive subject lines. It also requires that unsolicited commercial email be identified as advertising and provide recipients with a method for opting out of receiving any such email in the future. In addition, the Act directs the FTC to issue rules requiring the labeling of sexually explicit commercial email as such and establishing the criteria for determining the primary purpose of a commercial email.

32

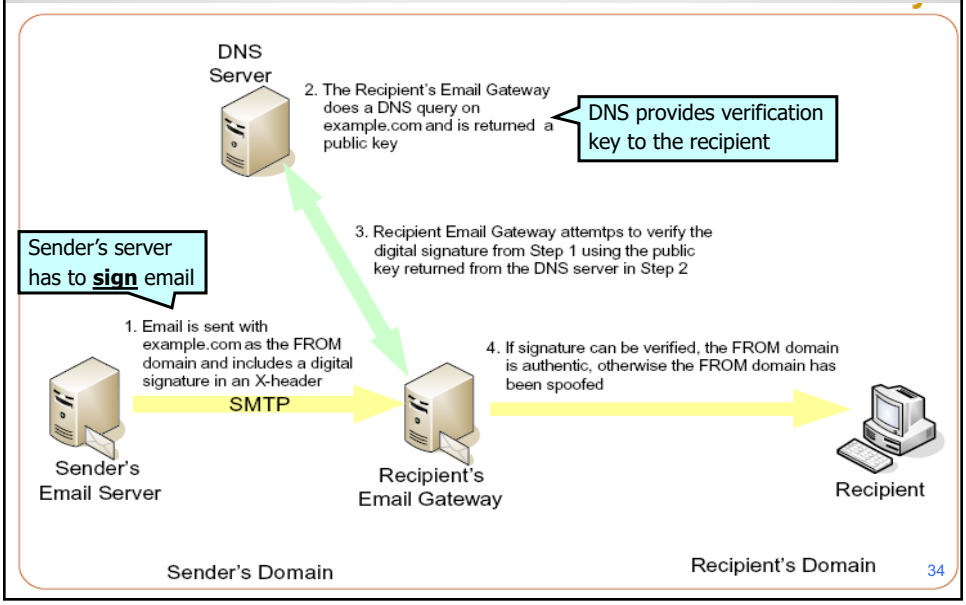
32

# SPF (Sender Policy Framework)




33

# Domain Keys (DKIM)




34

## Graylists

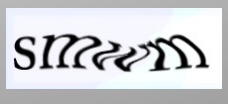
- Recipient's email server records (stores) the triple:
  *<sender email, recipient email, peer IP>*
  - Each triple kept for 3 days (configuration parameter)
- First time (triple not in DB): SMTP 421 reply "Busy"
  - Records triple in the database
- Second time (after 5 minutes): let email pass
- What is this defense based on?
- Easily spoofable, but works against many spammers

35

35

## 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)



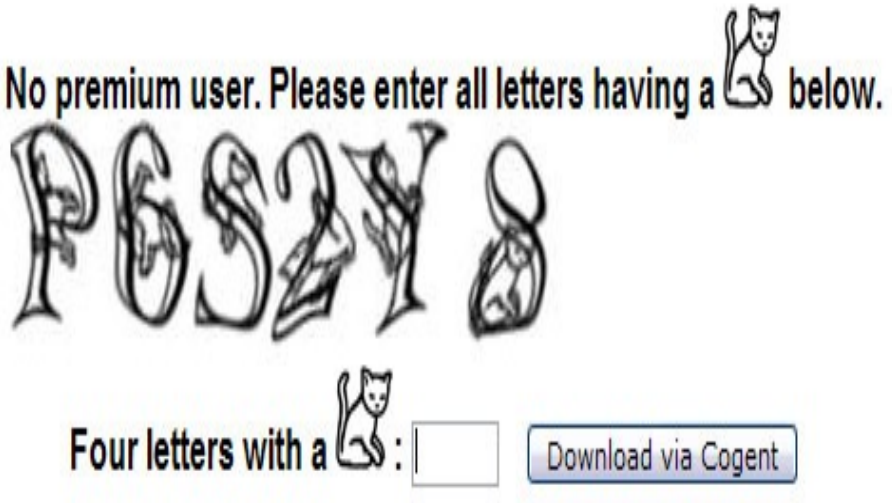
- Difficult to deploy (why?)

36

36

## Worst CAPTCHA Ever?

http://depressedprogrammer.wordpress.com/2008/04/20/worst-captcha-ever/




37

37

## Gone in 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 sort of do (how?)
- Botnet = massive distributed computing platform
  - Use them to solve CAPTCHAs
- Success rates better than that of humans for many CAPTCHA types

38

38

# Using Humans to Solve CAPTCHAs

http://old.post-gazette.com/pg/03278/228349.stm

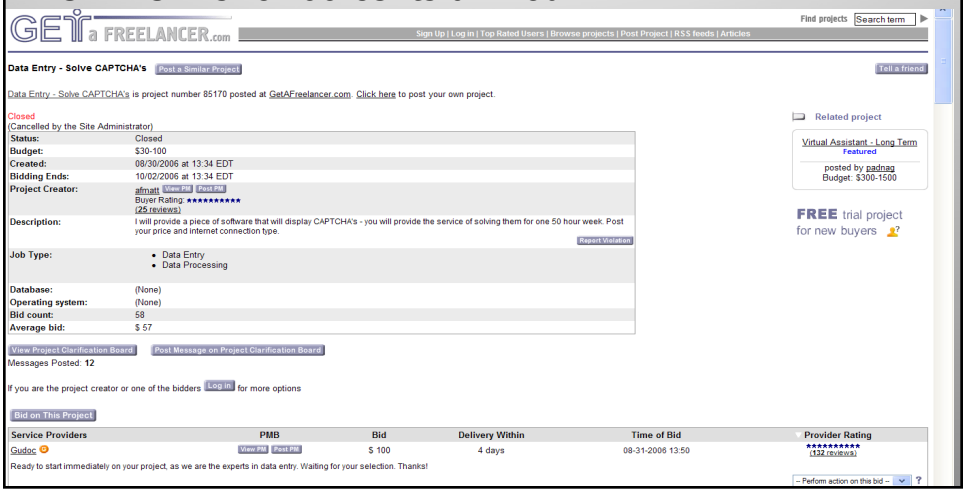
"... at least one potential spammer managed to crack the CAPTCHA test. Someone designed a software robot that would fill out a registration form and, when confronted with a CAPTCHA test, would post it on a free porn site. Visitors to the porn site would be asked to complete the test before they could view more pornography, and the software robot would use their answer to complete the e-mail registration."

39

39

# Solve CAPTCHAs for Fun and Profit

- Third-world "data entry specialists" will solve CAPTCHAs for 60 cents an hour




40

# CAPTCHA-Solving Services

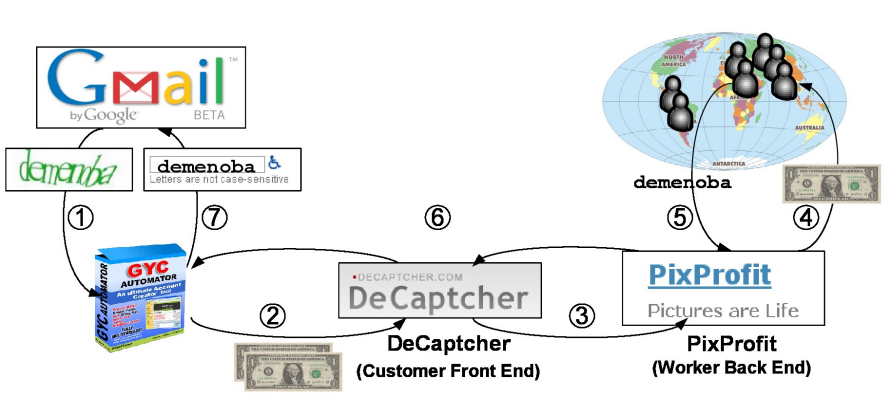
[Motoyama et al. "Understanding CAPTCHA-Solving Services in an Economic Context" ]

| Service | $/1K Bulk | Dates (2009–2010) | Requests | Responses |
|---|---|---|---|---|
| Antigate (AG) | $1.00 | Oct 06 – Feb 01 (118 days) | 28,210 | 27,726 (98.28%) |
| BeatCaptchas (BC) | $6.00 | Sep 21 – Feb 01 (133 days) | 28,303 | 25,708 (90.83%) |
| BypassCaptcha (BY) | $6.50 | Sep 23 – Feb 01 (131 days) | 28,117 | 27,729 (98.62%) |
| CaptchaBot (CB) | $1.00 | Oct 06 – Feb 01 (118 days) | 28,187 | 22,677 (80.45%) |
| CaptchaBypass (CP) | $5.00 | Sep 23 – Dec 23 (91 days) | 17,739 | 15,869 (89.46%) |
| CaptchaGateway (CG) | $6.60 | Oct 21 – Nov 03 (13 days) | 1,803 | 1,715 (95.12%) |
| DeCaptcher (DC) | $2.00 | Sep 21 – Feb 01 (133 days) | 28,284 | 24,411 (86.31%) |
| ImageToText (IT) | $20.00 | Oct 06 – Feb 01 (118 days) | 14,321 | 13,246 (92.49%) |

41

# CAPTCHA-Solving Economy

[Motoyama et al. "Understanding CAPTCHA-Solving Services in an Economic Context" ]

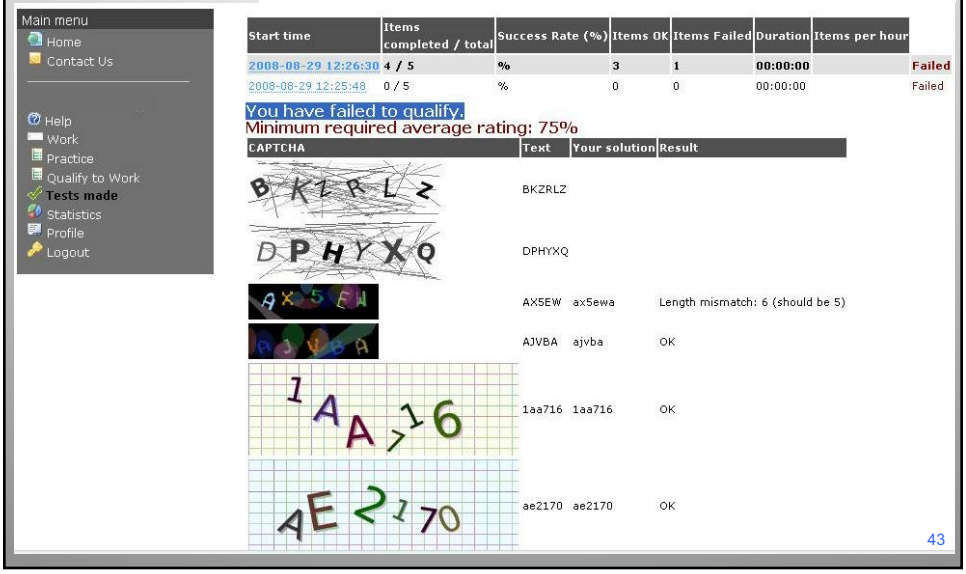



**CAPTCHA-solving market workflow: ① GYC Automator attempts to register a Gmail account and is challenged with a Google CAPTCHA. ② GYC uses the DeCaptcher plug-in to solve the CAPTCHA at $2/1,000. ③ DeCaptcher queues the CAPTCHA for a worker on the affiliated PixProfit back end. ④ PixProfit selects a worker and pays at $1/1,000. ⑤ Worker enters a solution to PixProfit, which ⑥ returns it to the plug-in. ⑦ GYC then enters the solution for the CAPTCHA to Gmail to register the account.** 42

42

# Support Tools

http://www.zdnet.com/blog/security/inside-indias-captcha-solving-economy/1835




43