CSE543 - Introduction to Computer and Network Security

Module: Spam

Professor Trent Jaeger
Fall 2010
SPAM, What is it?

• Like real spam, it is ….
  1. Nobody wants it or ever asks for it.
  2. No one ever eats it; it is the first item to be pushed to the side when eating the entree.
  3. Sometimes it is actually tasty, like <1% of junk mail that is useful to some people.

• “An endless stream of worthless text” - webpedia

• Who does it (directly or indirectly) effect?
  ‣ End-users, ISPs, backbone provider, enterprises, users

• Factoid: On average, it takes 4-5 seconds to process an email SPAM message (Ferrris Research)
SPAM: But does it really matter?

- Not problem, growth alarming (1997)
  - Small percentage of total email
- SPAM represents a real cost (2003)
  - 13 billion annually (Ferris Research)
  - lost productivity, additional hardware, …
  - 15% of people find it problematic (Gartner)
- 70-80% of email is now SPAM (Viruslist 2009)
More facts (StarReviews 2009)

- You mileage may vary ...
  - The average PC user receives over 2,000 per year.
  - The average computer user receives about 10 spams per day.
  - Spam is expected to increase by about 63% in 2007.
  - About 28% of people answer spam emails
  - 15-20% of corporate email is spam...and it’s ever-growing.
  - 25% of spam is product-related.
  - About 90 billion spam emails are sent per day.
  - Nearly 80% of spam is sent from zombie networks or botnets.
  - China has the highest rate of “spamvertized” websites—i.e. Links back to websites.
  - 63% of “take my email off your list” aren’t fulfilled.
  - 86% of emails posted on websites end up receiving spam...
SPAM: What does it look like?

• “Legitimate” commercial email …
  ‣ “green card” SPAM Canter and Siegal (‘94)
  ‣ ESPN, NY Times - often provide opt-(in/out)
• Personal, political, or religious diatribes
  ‣ Chain letters, jokes, hoaxes, …
• Commercial hucksters from
  ‣ Ranges from innocuous (“replace your windows”)
  ‣ … to the annoying (“MAKE MONEY BY SITTING”)
  ‣ … to the offensive (“Big Bob’s house of XXX”)
• The classic scam “Nigerian Finance Minister”
  ‣ Variant of old ponzie scheme (2$ billion – MessageLab)
  ‣ Help to transfer my “20 million”, I will give you 1/2 to help me …. 
  ‣ Known as the 419 scam (for section 419 of nigerian criminal code)
What is SPAM? (2007)

- Health (7.0%)
- Leisure (6.0%)
- Spiritual (4.0%)
- Other (3.0%)
- Internet (7.0%)
- Scams (9.0%)
- Adult (19.0%)
- Products (25.0%)
- Financial (20.0%)
SPAM: Where does it come from?

- Direct marketers or spam service resellers
  - Canter and Siegal (green card lawyers)
  - CyberPromotions
    - AOL vs. CyberPromotions – established that CP did not have a 1st amendment right to send spam
    - Hence, legal to use block email (very important)
    - Led to agreements between ISP and CP
- Many, many, other spam companies arose
  - Some good, some bad, some downright illegal
  - “Whack-a-mole” antonymous systems
  - Short lived/spoofed domains
- **Compromised hosts** (e.g., viruses, worms, spy-ware)
  - Almost all SPAM is delivered by zombie networks/botnets
  - No need/incentives to maintain infrastructure
McColo

- San Jose web hosting center
- Their ISPs shut them down in 2008 (depeered)
- SPAM immediately dropped by 60%

\textit{Reality}: McColo was a corrupt organization that was hosting a significant portion of zombie/botnet masters on earth.

\textit{Reality}: McColo was indirectly responsible for 60 million of the 100 million SPAM sent every day.
Phishing

• Email falsely claiming to be from organization in hopes of extracting private information
• Social engineering/misdirection
  ‣ exploit people basic trust, tendencies, e.g., con
  ‣ DNS games (e.g., www.hotmail.bob.com)
  ‣ misleading URLs (e.g., bin encoding)
  ‣ Replacing address bar with fakes (e.g., JavaScript)
• Countermeasures
  ‣ Education, education, education ...
  ‣ DNS validation (DNS sec ...)
  ‣ Monitor/counter phishing style activity (redirects, etc.)
SPAM: What is the economic model?

- spammers only need small percentage of responses to recoup costs
  - Tools are readily available
  - Simple, low cost servers
  - Fundamental: cheap to send email
- email address lists
  - Buy/trade ~ spammer currency
  - Email lists can be obtained in all sorts of interesting ways (honest and dishonest)
    - Web-pages, email lists, chat rooms, guess …
    - AOL Profiles (on line database of personal info)
    - The “FriendGreetings” exploit (one of first spy-ware)
- 28% of users reply to SPAM
SPAM: How does SMTP work?

sender

MTA (relay)

LAN

The Internet

MTA

recipient
SPAM Mitigation

• Problem: How do automatically identify (and potentially remove) SPAM without affecting real email?

• SPAM! – classifies techniques (CACM, 1996)
  ‣ Filtering
  ‣ Counter-measures
  ‣ Metering (postage due)
  ‣ Channels, referral networks, fee restructuring, ..
SPAM Mitigation: Filtering

• Look for SPAM “tells” in the email
  ‣ Sender, e.g., knownspammer.com (blacklists)
  ‣ Subject e.g., email yelling – “BUY NOW”
  ‣ Keywords, e.g., “sex, free, buy, …”
  ‣ Format, e.g., HTML-format, javascript
  ‣ Count, e.g., 1000 of the same message
  ‣ Problem: inexact science
    • users will not tolerate filtering of real email

• Filter on specific occurrences or combinations
  ‣ Triggers filter problem: arms race with spammers
    • “V.I.A.G.R.A” is not the same as “VIAGRA”
  ‣ The “bit-bucket”, “/dev/null”, “circular file”, …
Filtering Problem

• A 2006 email ...

“mistress allowed fly turn beautiful side. forth enemy comes six welcome. drew evil full turning? fail mother wine street getting? commit independent glass ought important cold. desire wish thee either away.”

• How do you automatically know which are SPAM and which are legitimate emails?
  ‣ Known as a machine learning problem
  ‣ Typical boolean classification approach
    ‣ Features - measurable facets
    ‣ Weighting - weigh values for features
    ‣ Threshold - above a value, then in “class”
Filtering: SPAMassassin

- Deersoft/NAI product
  - 5 guys in SF
  - Rather than filtering on keywords or email characteristics, statistical and heuristic valuation, i.e., Bayesian filtering
    - Rules characterize email features
    - Auto-whitelisting learns sender behavior
    - External databases of spammers, good guys, ...
    - Score: probably legitimate, probable spam ...
  - Note: SPAMassassin does nothing with/to email
Filtering: SPAMassassin

Spam- assassin

Mail Processor

SPAM?

No/ Maybe

Yes

Score

(trash)

(inbox)
**Managed SPAM filtering**

- Organization routes email through vendor, e.g., *Brightmail*
- Vendor filters email based on internal collected SPAM information, then forwards to organization

- The more organizations/customers a SPAM manager serves, the better the filtering, i.e. exhibits *network effect*
SPAM Mitigation: Countermeasures

• Physical, real-world countermeasures
  ‣ Legal: Sue the sender
  ‣ Remove permissions (via abuse hotlines)

• The mail-bomb response
  ‣ Flood the sender's network with emails
  ‣ Maybe responding to request

• Other attack on sender's network
  ‣ DOS sender mail servers, other services

• Q: Is there a problem with these techniques?
SPAM Mitigation: Metering

• Recognition that little negative incentive to SPAM
• More closely model the physical postal service
  ‣ Increase the cost on the sender such that spaming becomes unprofitable
  ‣ ... or at least worthy of receiver time
  ‣ Idea: Pay receiver or receiver ISP to send email
    • Refund if email is acceptable (maybe)
  ‣ Problem: Requires fundamental changes in email system

• Another kinds of metering: puzzles (Dwork&Naor)
  ‣ Receiver provide computational puzzle
  ‣ Sender must send solution before accepting email

• Q: Would you pay to send email?
CAN-SPAM Act

- Prohibits fraudulent or deceptive subject lines, headers, addresses, etc.
- Makes it illegal to send e-mails to e-mail addresses that have been harvested from websites.
- Criminalizes sending sexually-oriented e-mails without clear markings.
- Requires that you have an working unsubscribe system that makes it easy for recipients to unsubscribe opt out of receiving your e-mails.
- Requires most e-mailers to include their postal mailing address in the message.
- Implicates not only spammers, but those who procure their services. Indeed, if you fail to prevent spammers from promoting your products and services you can prosecuted.
- Includes both criminal and civil penalties and allows suits by the Federal Trade Commission (FTC), State Attorneys General, and Internet Service Providers.
SPAM Mitigation: regulatory

• Regulatory – seek to place restrictions on who and how SPAM is sent
  ‣ *Telephone Consumer Protection Act* (TCPA) caused to be regulated as junk-FAX

• Do No SPAM list
  ‣ FTP proposed it, then found it won’t work
  ‣ How to enforce?
  ‣ What technologies?

• About half the US states have enacted spam legislation

• http://www.spamlaws.com/
SPAM Mitigation: the rest …

• Channels - automatically categorize and file
  ‣ User decides what to do with each category
  ‣ I do this with different addresses

• Opt-out lists - short lived lists of people who specifically do not want SPAM
  ‣ Q: anybody see a problem with handing this list over to spammers?

• Referral networks
  ‣ Clubs, organizations, and users make introductions
  ‣ Introductions govern who can send email to whom
  ‣ … or simply used to mark some email as more important.

• SenderID (Microsoft)
  ‣ use new DNS record to “authenticate” sending mail server
  ‣ prevents some kinds of simple sender spoofing