Everything you need to know about spam (in 15 minutes)_

Chris St. Pierre
Unix Systems Administrator
Nebraska Wesleyan University

What is spam?

- Spam isn't an abbreviation or acronym.
- UCE (Unsolicited Commercial Email) and UBE (...Bulk...)_
- “Spam” is more than spam: phishing, 419 scams, lottery scams, pump and dump, viruses, etc.
- “Spam” is what your users say it is

Nevers

- ***NEVER bounce spam or viruses***
- ***NEVER forward to off-site addresses before filtering***
- ***NEVER consider lame challenge-response systems***

Approaches

- Pay someone a lot of money
- Roll your own
- Pray

The Roll-Your-Own Stack Approach

1. RBLs
2. Greylisting
3. Sanity restrictions
4. Antivirus
5. Bayesian and rule-based filtering

1. RBLs

- Realtime Black List (or DNSBL: DNS Black List)_
- Someone else has done all the work for you. Yay!
- Run a caching nameserver
- *The big question: what RBLs to use?*

Live RBL Revue!

- Only a few are worth considering:
  - cbl.abuseat.org
  - njabl.org
  - SpamCop
  - zen.spamhaus.org

2. Greylisting

- Overview:
  - Greylisting identifies each message with a unique triplet: sender, recipient, originating server.
  - The first time it sees a given triplet, it gives a 4xx (tempfail) code
  - Legitimate servers will retry, at which point the triplet will be recognized and accepted
  - Spammers don't waste resources on retries
- ***Greylist on the /24 netblock of the originating server***
- ***Retry time doesn't matter, because spammers don't retry***
- ***Auto-whitelist and auto-blacklist***

3. Sanity restrictions

- Lots of fun stuff!
- Site-specific whitelists/blacklists
- Reject non-FQDN HELOs and HELOs with bad syntax
- Reject mail to unknown recipients!
- Reject HELOs that resolve to bogons or reserved local IPs
- Reject mail when the HELO name has no MX or A record?
- Well-configured sanity restrictions can drop about 25% of your spam
- With policyd-weight (and others), you can score on sanity of a message rather than doing a binary accept/reject

4. Antivirus

- ClamAV is king
- Not just antivirus; anti-phishing *par excellence*
- *In addition to the standard rules, use http://www.sanesecurity.com/clamav*
- *Keep it updated and ClamAV will Just Work*
- *Reject viruses; drop if you must; never deliver them*

5a. Bayesian classifiers

- Dspam, Bogofilter, SpamAssassin, probably others
- Analyze tokens and patterns to discern common patterns in ham and spam; score messages according to the patterns they exhibit.
- Let your users report FPs and FNs, and train Bayes on it
- Train train train!
- ***DO NOT train Bayes on public corpora***
- ***DO NOT train Bayes on your outgoing mail***

5b. Rule-based filtering

- Basically, SpamAssassin
- Keep it updated with sa-update (v3.1+)_
- Some highlights:

  a)Checksumming systems (Razor2, DCC, Pyzor)_

  b)URIBL

  c)SARE rulesets

Up and coming?

- p0f – filtering on client OS
- tarpitting – slowing a connection

Questions?