An Empirical Study of Spam Traffic and the Use of DNSBLs

Emil Sit

Joint work with Jaeyeon Jung

sit, jyjung@csail.mit.edu

MIT Computer Science and Artificial Intelligence Laboratory
What are DNS Black Lists?

- Lists of hosts (IP address) that might send you spam.
- Checked via DNS when mail is being received. e.g.,
  - Upon connection from 219.251.61.45,
  - check if 45.61.251.219.bl.spamcop.net exists.
  - If yes, respond with SMTP error, and disconnect.
What are DNS Black Lists?

- Lists of hosts (*IP address*) that might send you spam.
- Checked via DNS when mail is being received. *e.g.*, upon connection from 219.251.61.45, check if 45.61.251.219.bl.spamcop.net exists. If yes, respond with SMTP error, and disconnect.

Different lists have different focus:
- Open Relays (*e.g.* list.dsbl.org)
- Known spam sources (*e.g.* sbl.spamhaus.org)
- Countries or ISPs (*e.g.* china.blackholes.us)
- Composite lists (*e.g.* dnsbl.sorbs.net)
Investigating DNSBLs

- What does DNSBL usage look like?
  - How much DNSBL traffic is there?
  - What impact does this have on DNS?
- How effectively can DNSBLs be?
  - Do DNSBLs identify spam sources?
Data Collection

- Analyze DNS packets and TCP SYN/FIN/RST traffic,
- At border of CSAIL and the rest of the world.
14.21% of lookups are for DNSBL (vs 0.39% in 2000).

DNSBLs have cached NS — 0.01% root lookups.
14.21% of lookups are for DNSBL (vs 0.39% in 2000).

DNSBLs have cached NS — 0.01% root lookups.

Median latency is 84ms (vs 89ms in 2000).
Estimating DNSBL effectiveness

- DNSBLs are effective if they list all spam sources.
- We will estimate hit rate by:
  - Identifying potential spam sources in trace.
  - Testing for membership in popular DNSBLs.
## Profile of SMTP connections

<table>
<thead>
<tr>
<th></th>
<th>7 Dec 2000</th>
<th>19 Feb 2004</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total attempted SMTP connections</td>
<td>29,303</td>
<td>787,231</td>
</tr>
<tr>
<td>Successful SMTP connections</td>
<td>24,790</td>
<td>324,134</td>
</tr>
<tr>
<td>Rejected SMTP connections</td>
<td>4,513</td>
<td>463,097</td>
</tr>
<tr>
<td>Remote hosts initiating SMTP</td>
<td>4,334</td>
<td>76,676</td>
</tr>
<tr>
<td>Remote hosts initiating rejected SMTP</td>
<td>79</td>
<td>7,970</td>
</tr>
<tr>
<td>Local hosts rejecting SMTP</td>
<td>19</td>
<td>90</td>
</tr>
</tbody>
</table>

**Can we distinguish spam sources from mail sources?**

- Majority of connections to hosts without SMTP server.
Source of rejected connections?

- Possible reasons for rejected connections:
  - Port scanners. (Very few.)
  - People trying to send mail.

- 70% of connections rejected to one host:
  - Host is listed as mail exchange for unused domain:
  - no legitimate recipients on machine.
  - Mail is to made-up addresses (spam) or bounces.
Source of rejected connections?

- Possible reasons for rejected connections:
  - Port scanners. (Very few.)
  - People trying to send mail.

- 70% of connections rejected to one host:
  - Host is listed as mail exchange for *unused* domain:
  - **no** legitimate recipients on machine.
  - Mail is to made-up addresses (spam) or bounces.

- Assume all hosts rejecting connections also get spam.
- (This underestimates number of spam sources.)
How many do DNSBLs list? (1)

<table>
<thead>
<tr>
<th>Total spam sources</th>
<th>Dec 2000</th>
<th>Feb 2004</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>100</td>
<td>14,090</td>
</tr>
</tbody>
</table>

Listed by:

- **cbl.abuseat.org**  | 0 | 1,401 |
- **list.dsbl.org**    | 5 | 7,624 |
- **opm.blitzed.org** | 0 | 122   |
- **ipwhois.rfc-ignorant.org** | 25 | 2,030 |
- **dnsbl.sorbs.net**  | 3 | 8,529 |
- **bl.spamcop.net**   | 0 | 496   |
- **sbl.spamhaus.org** | 2 | 1,123 |

Total unique hosts black-listed | 34 (34%) | 11,521 (82%) |

- Checked in March 2004…
- Do DNSBLs react faster?
Collecting and annotating spam

- Supplement traces with active collection.
  - Set up a machine dedicated to receiving spam ("spam trap").
  - Annotate all spam received with black-list checks.

  - Detected > 43,000 spam sources.
  - Received 136,206 spam messages.
How many do DNSBLs list? (2)

- 78% of sources are listed *when they first arrive*.
- Some sources become listed/delisted over time:
  - 80% of all sources listed at some point.
- Unlisted sources send 30% of spam.
- Could DNSBLs do better?
Spam arrival rates


Number of spam sources tracks number of spams.
Most spam sources may send few e-mails

- Spam sources tend to be low volume.
- 20% of spam from hosts that send 1 message.
- These hosts may be harder to black list.
Most spam sources may send few e-mails

- Spam sources tend to be low volume.
- 20% of spam from hosts that send 1 message.
- These hosts may be harder to black list.
Conclusions

- Spam is now an important driver of DNS lookups.
- DNSBLs appear to block $\approx 80\%$ of spam sources.
  - Black lists may not adapt well to one-shot sources.
  - Limits potential utility of DNSBLs.
Questions?