Revealing Botnet Membership Using DNSBL Counter-Intelligence

Authors
Anirudh Ramachandran, Nick Feamster and David Dagon
College of Computing, Georgia Institute of Technology

Presentation by Petri Yllö 22.11.2006
Background

- Botnets are used for various criminal activities
- Blacklisting after spam is sent is too late
- Passively detecting bots is difficult

- DNS-based blackhole list (DNSBL) is used to track IP addresses that originate spam
- Botmasters are known to sell “clean” bots at a premium
- Botmasters themselves must perform “reconnaissance” lookups to determine their bots’ blacklist status
- Passive analysis of DNS-based blackhole list (DNSBL) lookup traffic
Goals

1. Passive heuristics for counter-intelligence
2. Study of DNSBL reconnaissance techniques
3. Identification of new bots
4. DNSBL-based countermeasures
Figure 1: DNSBL-based Spam Mitigation Architecture.
Reconnaissance Techniques

- Distinguish DNSBL queries issued by botmasters from those performed by legitimate mail servers.

- Property 1 (spatial relationships)
  - Bots have high lookup ratio

\[ \lambda_n = \frac{d_{n,\text{out}}}{d_{n,\text{in}}} = \text{lookup ratio}, \lambda, \text{ of some node } n \]

- \( d_{n,\text{out}} \) = out-degree, number of distinct IP addresses that node \( n \) queries
- \( d_{n,\text{in}} \) = in-degree, number of distinct IP addresses that issue a query for node \( n \)

- Property 2 (temporal relationships)
  - Topic of future work
Botmasters Reconnaissance Techniques

1. Third-party reconnaissance
2. Self-reconnaissance
3. Distributed reconnaissance
Algorithm for DNSBL query graph

```plaintext
ConstructGraph()
create empty directed graph G

/* Parsing */
for each DNSBL query:
  Identify querier and queried

/* Pruning */
if querier ∈ B or queried ∈ B then
  add querier and queried to G if they
  are not already members of G
  if there exists an edge E(querier, queried) ∈ G then
    increment the weight of E(querier, queried)
  else
    add E(querier, queried) to G with weight 1
```

Query graph extrapolation is repeated until no new edges are added to G.
- add edges if one of the endpoints is already present in the graph.
Preliminary Results

- Pruned traffic amounts to less than 1% of the total DNSBL traffic.
- Botnets are being used to perform DNSBL reconnaissance on behalf of bots in other botnets.
- Nodes with the highest values of $\lambda_n$ in the pruned graph were known bots.
- Queried nodes in the graph were new, previously unknown bots and some of these were confirmed spam originators by authors spam honeypot.
Hosts with highest out-degrees.

<table>
<thead>
<tr>
<th>Node #</th>
<th>ASN of Node</th>
<th>Out-degree</th>
<th>known spammers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Everyone’s Internet (AS 13749)</td>
<td>36,875</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>IQuest (AS 7332)</td>
<td>32,159</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>UUNet (AS 701)</td>
<td>31,682</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>UPC Broadband (AS 6830)</td>
<td>26,502</td>
<td>8</td>
</tr>
<tr>
<td>5</td>
<td>E-xpedient (AS 17054)</td>
<td>19,530</td>
<td>4</td>
</tr>
</tbody>
</table>

- All of these 5 nodes are known bots from a Bobax trace.
- A few IP addresses queried by these nodes actually sent spam to authors spam honeypot
- Nearly all of IP addresses that sent spam to authors honeypot were not present in our list of known bots
Figure 3: CDF of the distribution of out-degrees for querying IP addresses.
Countermeasures

- Suspect IP addresses could be constructed by establishing a spam trap
- A suspect node could be detected by identifying nodes in the DNSBL query graph with a high value of $\lambda_n$
- Suspect node itself is likely a bot and nodes that this node is querying are likely bots
- If there are other high-degree nodes also querying the same bots, a detection algorithm might be able to “walk” the DNSBL graph (e.g., from parent to parent) to discover multiple distinct botnets.
- Mitigate spam by reconnaissance poisoning