Behavioral Study of Bot Obedience using Causal Relationship Analysis

Pekka Pietikäinen, Lari Huttunen

Oulu University
Secure Programming Group
• Botnets have become an increasing menace
• Tens of strategically placed hosts to hundreds of thousands
• Life-cycle:
  • Infection directly through the network or user interaction
  • Trojan payload downloaded and/or executed
  • Bot joins the botnet
  • Bots are used for some activity
  • Bots are upgraded to new versions
• Active/passive
• Scope: Individual machines/network
• Detection time: proactive/reactive
• User: end-user, network operator etc.
• Type: Indirect, Direct
<table>
<thead>
<tr>
<th>Data source</th>
<th>Scope</th>
<th>Detection time</th>
<th>User</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Victim</td>
<td>Individual machine</td>
<td>After infection</td>
<td>Unhappy end-user</td>
<td>Direct, Indirect</td>
</tr>
<tr>
<td>Honeypot or spampot</td>
<td>Varies</td>
<td>Early</td>
<td>Security researcher</td>
<td>Direct</td>
</tr>
<tr>
<td>Antivirus software</td>
<td>Individual machine</td>
<td>Infection attempt</td>
<td>End-user, network operator</td>
<td>Direct</td>
</tr>
<tr>
<td>IDS with signature</td>
<td>Network</td>
<td>Infection attempt</td>
<td>Network operator</td>
<td>Direct</td>
</tr>
<tr>
<td>IDS without signature</td>
<td>Network</td>
<td>After infection</td>
<td>Network operator</td>
<td>Indirect</td>
</tr>
<tr>
<td>DNS-based IDS</td>
<td>Network</td>
<td>After infection</td>
<td>Network operator</td>
<td>Indirect</td>
</tr>
<tr>
<td>Flow data</td>
<td>Several networks</td>
<td>Early to postmortem</td>
<td>Network operator</td>
<td>Direct, Indirect</td>
</tr>
</tbody>
</table>
Honeypots and spampots

- Attempt to collect live instances of malware
- High-interaction (traditional honeypot)
- Low-interaction (Nepenthes)
- Only catches the low-hanging fruit
- Privacy and liability issues
- Requires expertise
- Still, provides the best intelligence about botnets
• Finds signatures of malware running on the system or malicious activity in general
• Can only spot activity for which signatures exist
• Usefulness as information source for botnet investigations depends on the deployment
• Collect data from network and attempt to find botnet traffic

• IRC traffic as signature
  • Easy to evade, just change the protocol a bit or encrypt
  • Legitimate traffic as false positives
  • Ephemeral port numbers -> have to look at all traffic

• Secondary botnet behaviour
  • Portscans, DDoS’s etc.
• New type of IDS especially useful for botnets
• Catch anomalies in DNS queries
  • Known controllers
  • Popular hosts
  • Abnormal qtypes
• False positives a problem
  • Correlate with NetFlow data
• Passive DNS replication
  • Gets around privacy issues, but cannot be proactive
• Summary data collected at border router
• Data rate is (almost) manageable
• Timestamp, Source/destination address & port, protocol, packet count, byte count, ...
• Isolating relevant data and anonymization needed for sharing
• Method for modeling and visualizing interactions in network traffic
• Groups potentially related events together

client:46419 -> 66.102.11.99:80 (TCP)
client:46420 -> 66.102.11.99:80 (TCP)
client:46421 -> 66.102.11.99:80 (TCP)
Total distinct addresses: 8293953
Total flows: 62393760
Control port flows: 18269
C&C hosts: 6
C&C flows: 18157
Number of victims: 546
Victim flows: 23753270
Control port flows: 17892
Port 445 flows: 23484991
Other traffic: 250387
• There is no single silver bullet for botnets
• Correlation of data from several methods is needed
  • Flow + DNS-based IDS to find potential targets for further analysis
  • Causality analysis to understand botnet activities better
• Sharing of data between organizations
• Evidentiary value of flow data
  • Number of victims can be enumerated and monentary value estimated
  • Causality analysis can be used to minimize flow data to the essentials