DETECTING ENCRYPTED BOTNET TRAFFIC

Han Zhang, Christos Papadopoulos, Dan Massey
What is a Botnet
Botnet Detection is Hard

- Hidden among legitimate hosts
- There are thousands of them
- Their signatures change

- Bots now use encrypted channels for C&C
Contributions

- We show that encryption foils current bot detection systems that rely on DPI (BotHunter)
- We show that encryption results in high entropy flows
- We design, build and evaluate two high-entropy flow detectors
- We incorporate high entropy as a feature in BotHunter and show that it can now detect encrypted bots
Outline

- Encryption Foils Detection Systems
- High Entropy Flow Detection
- Detecting Encrypted Botnets
- Conclusions
- Future Work
Datasets

- Botnet Datasets: 140 Butterfly and Kraken traces obtained from Georgia Tech
  - Used to evaluate true positives
- Background Traffic: three traces with payload captured at our university lab
  - Well protected network (no bots)
  - Used to evaluate false positives
Encrypted Traffic and BotHunter

- Use 140 real botnet traces
- Use XOR to encrypt - low bar for attacker

<table>
<thead>
<tr>
<th>Traffic Encryption</th>
<th>IP Replacement</th>
<th>Detection Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>No</td>
<td>140/140</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>139/140</td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>77/140</td>
</tr>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>0/140</td>
</tr>
</tbody>
</table>
Outline

- Encryption Foils Detection Systems
- High Entropy Flow Detection
- Detecting Encrypted Botnets
- Conclusions
- Future Work
Entropy of Different Data Types

- Measure of randomness
- Encrypted data => high entropy
- But high entropy ≠ encrypted traffic
- Many high entropy flows:
  - Media, compressed data, encrypted data, etc.
What is a High Entropy Data?

- Calculate entropy threshold for all N bytes data based on the method in [1]
  - Build a table of thresholds for all N (1 ≤ N ≤ 64K)
- Calculate entropy of N bytes data
- Data is HE if entropy of $D_N >$ threshold (N)

[1] Detecting Subverted Cryptographic Protocols by Entropy Checking
High Entropy Flow Detection

- Flow-based:
  - Calculate the entropy of N first bytes in the flow
  - Compare the entropy with threshold

- Packet-based:
  - Count the number of high entropy packets (HE) and number of low entropy packets (LE)
  - High entropy flow if \#HE/(\#LE+\#HE) > Threshold
Optimization: Skip LE Packets

- For many encryption protocols first few packets may not be encrypted
  - Key exchange in SSH, HTTPS

- Optimization: skip initial LE packets – measure entropy after encountering T sequential HE packets
High Entropy Flow Detection Results

- True positives (600 files: MS word, MS Excel, ASCII, PDF)

<table>
<thead>
<tr>
<th>Encryption Algorithm</th>
<th>High Entropy (flow-based) out of 600</th>
<th>High Entropy (pkt-based) out of 600</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>6</td>
<td>32</td>
</tr>
<tr>
<td>DESX</td>
<td>599</td>
<td>599</td>
</tr>
<tr>
<td>AES</td>
<td>600</td>
<td>600</td>
</tr>
<tr>
<td>RC2</td>
<td>599</td>
<td>600</td>
</tr>
</tbody>
</table>

- True positives for real network traffic (Lab Trace 1)

<table>
<thead>
<tr>
<th>Traffic</th>
<th>SSH</th>
<th>HTTPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>HE Rate (flow-based)</td>
<td>95.1%</td>
<td>95.4%</td>
</tr>
<tr>
<td>HE Rate (packet-based)</td>
<td>97.2%</td>
<td>97.6%</td>
</tr>
</tbody>
</table>
Outline

- Encryption Foils Detection Systems
- High Entropy Flow Detection
- Detecting Encrypted Botnets
- Conclusions
- Future Work
BotHunter Background

- Main idea: look for specific events that form a dialog
- Each event carries a weight
- Add weights of all detected events – if above threshold then we detected a bot
- Runs on top of snort
Enhancing BotHunter

- Add high entropy flow detector to BotHunter
  - HE flow triggers C&C communication or egg download in BH
- We select the packet-based HE flow detector
  - More sensitive to HE flows
  - Easier to implement (no per-flow state)
  - HE/(LE+HE) ratio = 90%, T = 2 sequential HE packets
Weight of HE Detector?

- Question: what should be the weight of the HE flow detector?
  - Reverse-engineer weights in BotHunter
    - C&C communication $W = 0.5$
    - Egg download $W = 0.5$
  - HE detector flags possible event above, so $0.1 \leq W \leq 0.4$
Enhanced BotHunter - True Positives

True Positive for 140 Botnet Traces
Enhanced BotHunter - False Positives

Result: $W = 0.4$
Outline

- Encryption Foils Detection Systems
- High Entropy Flow Detection
- Detecting Encrypted Botnets
- Conclusions
- Future Work
Conclusions

- Encrypted botnets foil detection systems that rely on DPI
- We show and leverage the fact that encryption produces high entropy flows
- We propose and build two detectors to identify high entropy flows
- We extend BotHunter with an HE detector producing an enhanced version of BotHunter that can now detect encrypted botnets
Outline

- Encryption Foils Detection Systems
- High Entropy Flow Detection
- Detecting Encrypted Botnets
- Conclusions
- Future Work
Future Work

- Build a tool to convert unencrypted to encrypted botnet traffic
  - Useful for testing bot detection algorithms
  - Support many standard encryption schemes
- Improve high entropy flow detectors
  - Distinguish encrypted data from other high entropy data
Questions?

Thank You

Email: zhang@cs.colostate.edu