TorBot: Protecting the Tor Network against Malicious Traffic

Advisor: Paulo Lício de Geus
Marcelo Invert Palma Salas (PhD Candidate @UNICAMP)
Esdras Rodrigues Do Carmo (Scientific Initiation Fellow)
Vitor Falcão da Rocha (Scientific Initiation Fellow)

University of Campinas
With support from Frida Lacnic Finance Agency
The Tor Network

- ... is an overlay network that enables anonymous communication between applications that communicate over TCP [1], protecting your privacy and identity on the Internet.

- Tor also protects our data against corporate or government targeted mass surveillance.

- Despite being used mainly by activists, journalists and bloggers, it supports illicit services and is prone to carry 30X more malicious traffic compared with others networks [2].
How does Tor Work?

- Tor is a group of volunteer-operated servers.
- Composed by 3 relays (guard, middle and exit), it applies distributed security to the network.
- Each router knows only the sender and receiver.
Deep problems in the deep web

- Governmental Vigilance (In particular Exit Relay and spoofing Hidden Services (HS))
- Connection speed (New competition: Rifle - MIT, I2P, Freenet)
- Malicious Traffic:
  - P2P (BitTorrent)
  - Hackers
  - Malware (botnets, ransomware (WannaCry))
  - Illegal Markets (drugs, counterfeit products, cigars, medicines) <=> gray market {Aliexpress, DHgate, iOffer}
  - HS (are 2% of Tor traffic, 1.5% are malicious traffic).
  - Kidnappers and blackmailers (rescue -> Bitcoins, Ripple, Ethereum, NEM, Litecoin, & among others)
Pure Cocaine 84% Garanty

Very Strong Stuff

Free Shipping to EU

0.1g - 18€

Full Escrow over Blackmarket for: weapons, drugs, id, counterfeits...
In [3], the authors analyzed more than 80,000 hidden services, finding:

- 85% of HS are up for less than 5 days,
- +100 new HS come online,

There is increased usage by malware (botnets, ransomware, etc.) in relation to the surface web.

<table>
<thead>
<tr>
<th>Category</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drugs</td>
<td>15.4%</td>
</tr>
<tr>
<td>Market</td>
<td>9%</td>
</tr>
<tr>
<td>Fraud</td>
<td>9%</td>
</tr>
<tr>
<td>Bitcoin</td>
<td>6.2%</td>
</tr>
<tr>
<td>Mall</td>
<td>5.7%</td>
</tr>
<tr>
<td>Wiki</td>
<td>5.2%</td>
</tr>
<tr>
<td>Whistleblower</td>
<td>5.2%</td>
</tr>
<tr>
<td>Counterfeit</td>
<td>5.2%</td>
</tr>
<tr>
<td>Forum</td>
<td>4.75%</td>
</tr>
<tr>
<td>Anonymity</td>
<td>4.5%</td>
</tr>
<tr>
<td>Search</td>
<td>4.25%</td>
</tr>
<tr>
<td>Hacking</td>
<td>4.25%</td>
</tr>
<tr>
<td>Hosting</td>
<td>3.5%</td>
</tr>
<tr>
<td>Porn</td>
<td>2.75%</td>
</tr>
<tr>
<td>Blog</td>
<td>2.7%</td>
</tr>
<tr>
<td>Directory</td>
<td>2.5%</td>
</tr>
<tr>
<td>Books</td>
<td>2.5%</td>
</tr>
<tr>
<td>Abuse</td>
<td>2%</td>
</tr>
<tr>
<td>New (others)</td>
<td>2%</td>
</tr>
<tr>
<td>Chat</td>
<td>2%</td>
</tr>
<tr>
<td>Guns</td>
<td>1%</td>
</tr>
<tr>
<td>Gambling</td>
<td>0.4%</td>
</tr>
</tbody>
</table>
How Malicious Traffic Works in Tor?

- Malware (botnets, ransomware, ...)
- Illegal market (drugs, guns, ...)
- Bitcoin (anonymous transactions)
Architecture for Discovering and Blocking Malicious Traffic

Legend:
- Inbound Tor Traffic
- Outbound Tor Traffic
- Onion Router

Client (OP)
Server

Tor Network
- Entry (OR1)
- Middle (OR1')
- Middle (OR2)

Automatic Management Tool
- Machine Learning Alg.
- IDS

Gateway
- Transparent Proxy (OP')
- Honeypots
- VirusTotal

Campus
- Exit (OR3)
Protecting the Tor Network against Malicious Traffic

Our proposal is divided into three phases:

i) Collect; ii) analysis and classification; iii) tracing and blocking malicious traffic.

This include:

Setting up a network capture and re-routing of the benign traffic;
System development for analyzing, back tracing, and blocking malicious traffic like botnets and others malware;
An application to recognize and block malicious hidden services.

To achieve this goal, we propose using tools such as:

Traffic analyzers;
IDS and VirusTotal;
Machine learning techniques and metadata analyzing.
Protecting the Tor Network against Malicious Traffic

Our proposal is divided into three phases:

i) Collect

ii) Analysis and Classification

iii) Tracing and Blocking

---

malicious traffic

---
Collect Malicious Traffic

More than 1200 samples:

- http://cerberussssc7cat.onion/
- https://zeltser.com/malware-sample-sources/
- https://github.com/ytisf/theZoo
- https://github.com/aboutsecurity/malware-samples
- https://github.com/ashishb/android-malware
- https://github.com/fdiskyou/malware
- https://gist.github.com/rain-1/989428fa5504f378b993ee6efbc0b168 (WannaCry)
Analysis and Classification Malicious Traffic

Some Results for Windows 8.1:

- Decision Tree: 96.15%
- Gaussian Naive Bayes: 96.44%
- Multinomial Naive Bayes: 94.49%
- Neural Network MLP: 97.7%
- SVM: 98.22%

- WannaCry was detected by 4/5 algorithms.
How does a botnet work?

1. **Primary Infection**
   - Identifying and exploiting vulnerable hosts (STEP 1)

2. **DNS Request**
   - DNS Response (STEP 2)

3. **Secondary Infection**
   - Obtaining bot software (STEP 3)

4. **Joining to the C&C**
   - (STEP 4)

5. **Botnet**
   - C&C Channel
   - Botmaster

6. **Command and Control Infrastructure (C&C)**
   - C&C Server
   - Binary Distribution Server

7. **Clear traffic**
   - Encrypted Traffic
How does a botnet work with Tor?
References


Obrigado!