Resident Evil: Understanding Residential IP Proxy as a Dark Service

Xianghang Mi, Xuan Feng, Xiaojing Liao, Baojun Liu, XiaoFeng Wang, Feng Qian, Zhou Li, Sumayah Alrwais, Limin Sun, Ying Liu
Background: Traditional Web Proxies

- HTTP/HTTPS
- SOCKS

Exit nodes are constrained
Exit nodes are distinguishable
Exit nodes may be heavily abused

Vulnerable to service blocking or degradation
Background: Residential IP Proxy as a Service
Background: Residential IP Proxy as a Service

 Millions of Residential IPs

 Clean IPs, Never Get Blocked

 Globally Distributed

 No Traffic Limits
Outline

Service Overview
- Network Structure & Scale & Distribution

Residential or Not
- Are proxy peers authentically residential IP addresses?

Evasiveness
- How well can proxy peers evade traffic detection or blocking?

Recruitment
- How can millions of proxy peers get recruited?

Usage
- What are those proxies used for, in the real world?

Misc. Findings
- Collusion, Local traffic, etc.
Service Overview: How it works

- **Proxy Customer**
- **Proxy Gateways**
- **Proxy Peers**
- **Destinations**

- Chrome
- facebook.com
- google.com
- amazon.com

![Diagram showing the flow of data from Proxy Customer to Proxy Gateways, then to Proxy Peers, and finally to Destinations.]
Service Overview: Scale

Each request is identified by a unique subdomain
Each request/response has payload encrypted and signed

<table>
<thead>
<tr>
<th>Provider</th>
<th>Price</th>
<th>Payment</th>
<th>Date(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proxies Online</td>
<td>$25/Gb</td>
<td>Paypal</td>
<td>07/06-11/24</td>
</tr>
<tr>
<td>Geosurf</td>
<td>$300/month</td>
<td>Paypal</td>
<td>09/17-10/22</td>
</tr>
<tr>
<td>ProxyRack</td>
<td>$40/month</td>
<td>Bitcoin</td>
<td>09/18-11/24</td>
</tr>
<tr>
<td>Luminati</td>
<td>$500/month</td>
<td>Paypal</td>
<td>09/25-11/01</td>
</tr>
<tr>
<td>IAPS Security</td>
<td>$500/month</td>
<td>Bitcoin</td>
<td>09/23-11/01</td>
</tr>
</tbody>
</table>
Service Overview: Scale

- Each request is identified by a unique subdomain
- Each request/response has payload encrypted and signed

60+ millions of successful probes
6.2 millions of unique IPv4 addresses
238 countries/regions, 52K+ ISPs.
Service Overview: Distribution

4096 * 4096 bitmap

Each /24 IPv4 prefix is mapped to a pixel, using Hilbert curve of order 12

Different pixel colors denote # of proxy IPs for a given /24 prefix
Residential or Not

1. Find Groundtruth
2. Select Features
3. Train/Evaluate Classifiers
4. Predict Proxy IPs

GT sources of various noise levels
Clean GT for training, noisy for evaluation

<table>
<thead>
<tr>
<th>Source</th>
<th>Label</th>
<th># IPs</th>
<th># /16</th>
<th># /8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manual</td>
<td>resi-clean</td>
<td>79</td>
<td>25</td>
<td>19</td>
</tr>
<tr>
<td>Device Search Engine</td>
<td>resi-clean</td>
<td>89,345</td>
<td>13,525</td>
<td>195</td>
</tr>
<tr>
<td>Trace My IP</td>
<td>resi-noisy</td>
<td>37,480</td>
<td>11,402</td>
<td>213</td>
</tr>
<tr>
<td>Filtered IP Whois</td>
<td>resi-noisy</td>
<td>23,264,961</td>
<td>394</td>
<td>31</td>
</tr>
<tr>
<td>IoT Botnets</td>
<td>resi-noisy</td>
<td>1,699,291</td>
<td>20,112</td>
<td>200</td>
</tr>
<tr>
<td>Public Clouds</td>
<td>non-resi-clean</td>
<td>53,716,321</td>
<td>968</td>
<td>99</td>
</tr>
<tr>
<td>Alexa Top1M</td>
<td>non-resi-clean</td>
<td>442,989</td>
<td>14,365</td>
<td>213</td>
</tr>
<tr>
<td>Commercial Proxies</td>
<td>non-resi-clean</td>
<td>519</td>
<td>71</td>
<td>44</td>
</tr>
<tr>
<td>Public Proxies</td>
<td>non-resi-noisy</td>
<td>148,509</td>
<td>14,004</td>
<td>204</td>
</tr>
</tbody>
</table>
Residential or Not

Find Groundtruth ➔ Select Features ➔ Train/Evaluate Classifiers ➔ Predict Proxy IPs

1. Residential IPs/prefixes are usually web clients instead of servers
2. Residential IPs/prefixes tend to be directly managed by ISPs

Capture
- DNS Records & Historical IP Whois
- web activities
- network hierarchy
- evolution by time

35 features stand out for next step
Residential or Not

Find Groundtruth → Select Features → Train/Evaluate Classifiers → Predict Proxy IPs

10K residential & 10K non-residential IPs → ML Classifier Training/Tuning → Random Forest Classifier

Recall: 97.12%  
Precision: 95.61%
Residential or Not

Find Groundtruth → Select Features → Train/Evaluate Classifiers → Predict Proxy IPs

5.9M (95.22%) of 6.2M predicted as residential IPs
Evasiveness

Recognized as proxy?

Identified as malicious?
Evasiveness

Recognized as a proxy?

Identified as malicious?

Publicly available proxy dataset

- Tor relays
- Free web proxies
- IP2Proxy LITE

Only 0.06% of 6.2M IPs
Evasiveness

Recognized as proxy?

Identified as malicious?

Publicly available IP threats
- Botnet bots
- Spamhaus EDROP
- Open Threat Exchanges

Only 2.20% of 6.2M IPs
Recruitment

Identify legitimate recruitment programs

Are those proxy peers voluntary users?

IP Profiling

Any IoT devices?

Identify proxy programs

What programs are used to proxy traffic?
Recruitment

- Identify legitimate recruitment programs
- IP Profiling
- Identify proxy programs

Only Luminati was found to recruit users through Hola programs.

And Hola programs were reported as problematic in previous studies.
Recruitment

Identify legitimate recruitment programs

IP Profiling

Identify proxy programs

730K IPs responded to our banner grabbing

550K got device type identified

All providers got suspicious IoT devices identified for their proxy IPs, including Luminati

<table>
<thead>
<tr>
<th>Device Type</th>
<th>Num</th>
<th>(%)</th>
<th>Device Vendor</th>
<th>Num</th>
<th>(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>router</td>
<td>114,768</td>
<td>48.42</td>
<td>MikroTik</td>
<td>86,593</td>
<td>36.53</td>
</tr>
<tr>
<td>firewall</td>
<td>25,088</td>
<td>10.58</td>
<td>Huawei</td>
<td>37,545</td>
<td>15.84</td>
</tr>
<tr>
<td>WAP</td>
<td>24,470</td>
<td>10.32</td>
<td>BusyBox</td>
<td>18,337</td>
<td>7.74</td>
</tr>
<tr>
<td>gateway</td>
<td>22,003</td>
<td>9.28</td>
<td>Technicolor</td>
<td>16,866</td>
<td>7.12</td>
</tr>
<tr>
<td>broadband router</td>
<td>17,358</td>
<td>7.32</td>
<td>SonicWALL</td>
<td>14,122</td>
<td>5.96</td>
</tr>
<tr>
<td>webcam</td>
<td>13,024</td>
<td>5.49</td>
<td>Fortinet</td>
<td>9,190</td>
<td>3.88</td>
</tr>
<tr>
<td>security-misc</td>
<td>10,608</td>
<td>4.48</td>
<td>Dahua</td>
<td>6,258</td>
<td>2.64</td>
</tr>
<tr>
<td>DVR</td>
<td>4,249</td>
<td>1.79</td>
<td>ZyXEL</td>
<td>5,601</td>
<td>2.36</td>
</tr>
<tr>
<td>media device</td>
<td>2,589</td>
<td>1.09</td>
<td>AVM</td>
<td>5,272</td>
<td>2.22</td>
</tr>
<tr>
<td>storage-misc</td>
<td>1,988</td>
<td>0.84</td>
<td>Cyberoam</td>
<td>4,558</td>
<td>1.92</td>
</tr>
</tbody>
</table>
Recruitment

Identify legitimate recruitment programs

IP Profiling

Identify proxy programs

Traffic logs of Infiltration probes

Accurate Correlation

Traffic logs of suspicious programs

- 67 different program samples identified
- Proxy programs are found for all 5 providers
- 50 of them were flagged by anti-virus engines
For the 67 proxy programs, 5M traffic logs were sampled to study usage.

9.36% of the destinations were reported to be malicious by VirusTotal.

Top 1000 traffic destinations were manually studied.
### Misc. Findings

<table>
<thead>
<tr>
<th></th>
<th>Proxies Online</th>
<th>Geosurf</th>
<th>IAPS Security</th>
<th>Luminati</th>
<th>ProxyRack</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proxies Online</td>
<td></td>
<td></td>
<td>12.5%</td>
<td>0%</td>
<td>0.09%</td>
</tr>
<tr>
<td>Geosurf</td>
<td>36.3%</td>
<td></td>
<td>0%</td>
<td>0.23%</td>
<td>1.7%</td>
</tr>
<tr>
<td>IAPS Security</td>
<td>0%</td>
<td>0%</td>
<td>66%</td>
<td>0.07%</td>
<td>0.04%</td>
</tr>
<tr>
<td>Luminati</td>
<td>0.02%</td>
<td>0.02%</td>
<td>0.07%</td>
<td>0.2%</td>
<td></td>
</tr>
<tr>
<td>ProxyRack</td>
<td>0.14%</td>
<td>0.86%</td>
<td>0%</td>
<td>0.2%</td>
<td></td>
</tr>
</tbody>
</table>

- Proxies Online and Geosurf are the same proxy provider
- IAPS Security is some kind of reseller for Luminati

Connection between proxy providers

Risk to the local network

Long-tailed distribution
Misc. Findings

Connection between proxy providers

Risk to the local network

Long-tailed distribution

3 out of 5 providers allow local traffic

Our Client  Proxy Gateway  Proxy Peer  Our Web server

rpaas.site  rpaas.site  rpaas.site

127.0.0.1  127.0.0.1
Response

192.168.0.1  192.168.0.1
Response

Response
## Misc. Findings

### Connection between proxy providers

### Risk to the local network

### Long-tailed distribution

<table>
<thead>
<tr>
<th>Provider</th>
<th>Top Countries</th>
<th>%</th>
<th>Top ISPs</th>
<th>%</th>
<th>Top ASNs</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proxies Online</td>
<td>India</td>
<td>32.2</td>
<td>BSNL</td>
<td>6.5</td>
<td>9829</td>
<td>8.1</td>
</tr>
<tr>
<td></td>
<td>USA</td>
<td>7.8</td>
<td>Uninet S.A. de C.V.</td>
<td>5.2</td>
<td>8151</td>
<td>5.4</td>
</tr>
<tr>
<td></td>
<td>Mexico</td>
<td>6.7</td>
<td>Deutsche Telekom AG</td>
<td>2.8</td>
<td>24560</td>
<td>4.9</td>
</tr>
<tr>
<td>Geosurf</td>
<td>India</td>
<td>27.9</td>
<td>Uninet S.A. de C.V.</td>
<td>6.9</td>
<td>8151</td>
<td>7.2</td>
</tr>
<tr>
<td></td>
<td>Brazil</td>
<td>9.2</td>
<td>BSNL</td>
<td>4.7</td>
<td>9829</td>
<td>5.8</td>
</tr>
<tr>
<td></td>
<td>Mexico</td>
<td>9.1</td>
<td>Deutsche Telekom AG</td>
<td>2.8</td>
<td>55836</td>
<td>4.5</td>
</tr>
<tr>
<td>ProxyRack</td>
<td>Russia</td>
<td>8.6</td>
<td>PT Telkom Indonesia</td>
<td>5.4</td>
<td>17974</td>
<td>5.3</td>
</tr>
<tr>
<td></td>
<td>Indonesia</td>
<td>8.1</td>
<td>Pakistan Telecom</td>
<td>3.7</td>
<td>8452</td>
<td>4.7</td>
</tr>
<tr>
<td></td>
<td>Egypt</td>
<td>6.3</td>
<td>Republican Unitary</td>
<td>3.3</td>
<td>45595</td>
<td>4.0</td>
</tr>
<tr>
<td>Luminati</td>
<td>Turkey</td>
<td>12.7</td>
<td>Turk Telekom</td>
<td>8.5</td>
<td>9121</td>
<td>8.5</td>
</tr>
<tr>
<td></td>
<td>Ukraine</td>
<td>7.9</td>
<td>JSC Ukrtelecom</td>
<td>1.7</td>
<td>25019</td>
<td>1.8</td>
</tr>
<tr>
<td></td>
<td>UK</td>
<td>6.1</td>
<td>BT</td>
<td>1.7</td>
<td>34984</td>
<td>1.8</td>
</tr>
</tbody>
</table>
Summary

- Millions of residential IPs with high evasiveness
- A prosperous ecosystem with higher prices and more service providers
- Potential threats to local network environments
- Problematic recruitment: a mix of legitimate and suspicious channels
- Powerful infrastructure for online abuse activities
- Promising and stealthy monetization channels for compromised devices

A lie that is half-truth is the darkest of all lies.
—Alfred Tennyson