An Extensive Evaluation of the Internet’s Open Proxies

Akshaya Mani*
Georgetown University

Tavish Vaidya*
Georgetown University

David Dworken
Northeastern University

Micah Sherr
Georgetown University

*Co-first authors
Open Proxies

- Circumvent censorship
- Bypass geographic content filtering
- ‘Provide’ anonymity
  - Hide requestor’s IP address

Aggregator Website
E.g., multiproxy.org, workingproxies.org, xroxy, etc.

Internet

Proxy

<185.79.243.161:41195>
<148.69.120.148:41875>
... 
<104.244.72.171:57480>
Potential for (Mis)behavior

- Proxy operators are well situated
  - To eavesdrop on users’ communication
  - Perform man-in-the-middle (MitM) attacks
  - Monetize their service by injecting ads

- Sending spam, injecting ads, redirecting to phishing sites, etc.
  [Pai et al. SIGCOMM ’04, Tsirantonakis et al. NDSS ’18]
Large(st) scale study of the Internet’s Open Proxies

encompassing

107,000 open proxies

&

13M proxy requests

over

50 day period
Overview

- An extensive study of the Proxy ecosystem
  - Proxy availability
  - Proxy diversity
  - Proxy performance
  - Proxy (mis)behavior

- Comparison with Tor
Overview

Check if **expected content** or not?
Experimental Setup

16 client locations: 15 AWS regions

Results did not vary significantly

Populated proxies list

<185.79.243.161:41195>
<148.69.120.148:41875>
<104.244.72.171:57480>

1) ‘Liveness’ Test
Active & Responds

2) Classify
HTTP, SOCKS, or CONNECT

3) Fetch files (URLs)

4) Record
HTTP headers, MIME, HTTP status code, etc.

Refreshed daily
Files Fetched

✧ Proxies may not always return the **expected content**

✧ **Fetched content over HTTP** *(hosted at Georgetown University)*
  
  ✧ Absence of end-to-end SSL/TLS makes MitM trivial
  
  ✧ HTML page  ✧ Windows batch (.bat) file  ✧ ZIP file
  ✧ JPEG Image  ✧ Linux/UNIX shell script (.sh)  ✧ Flash object (.swf)
  ✧ Java JAR archive  ✧ Windows executable (.exe) *(hosted at winzip.com)*

✧ **Additional content fetched over HTTPS**

  ✧ Fetched URLs with 3 categories of certificates
    
    ✧ Revoked certificate  ✧ Self-signed certificate  ✧ Valid & verifiable certificate
Understanding Open Proxy Ecosystem
Open Proxies (Un)availability

- Median of **41K** proxies per day
- Median of **3,283** working proxies

> 92% of listed proxies are unresponsive
# Geographic Distribution

<table>
<thead>
<tr>
<th>Country</th>
<th>Responsive Percent</th>
<th>Correct Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brazil</td>
<td>17.97%</td>
<td>Brazil</td>
</tr>
<tr>
<td>China</td>
<td>16.75%</td>
<td>China</td>
</tr>
<tr>
<td>United States</td>
<td>12.12%</td>
<td>United States</td>
</tr>
<tr>
<td>Indonesia</td>
<td>6.77%</td>
<td>Indonesia</td>
</tr>
<tr>
<td>Thailand</td>
<td>5.73%</td>
<td>Thailand</td>
</tr>
<tr>
<td>Russian Federation</td>
<td>5.11%</td>
<td>Russian Federation</td>
</tr>
<tr>
<td>Singapore</td>
<td>2.61%</td>
<td>Singapore</td>
</tr>
<tr>
<td>Germany</td>
<td>2.59%</td>
<td>India</td>
</tr>
<tr>
<td>India</td>
<td>2.54%</td>
<td>Germany</td>
</tr>
<tr>
<td>Canada</td>
<td>1.93%</td>
<td>Canada</td>
</tr>
<tr>
<td>All others (136)</td>
<td>25.65%</td>
<td>All others (131)</td>
</tr>
</tbody>
</table>

- 10 countries responsible for **75%** of world’s working proxies
## Geographic Distribution

<table>
<thead>
<tr>
<th>Country</th>
<th>Responsive Percent</th>
<th>Correct Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brazil</td>
<td>17.97%</td>
<td>Brazil 19.48%</td>
</tr>
<tr>
<td>China</td>
<td>16.75%</td>
<td>China 14.74%</td>
</tr>
<tr>
<td>United States</td>
<td>12.12%</td>
<td>United States 12.17%</td>
</tr>
<tr>
<td>Indonesia</td>
<td>6.77%</td>
<td>Indonesia 7.15%</td>
</tr>
<tr>
<td>Thailand</td>
<td>5.73%</td>
<td>Thailand 6.06%</td>
</tr>
<tr>
<td>Russian Federation</td>
<td>5.11%</td>
<td>Russian Federation 5.40%</td>
</tr>
<tr>
<td>Singapore</td>
<td>2.61%</td>
<td>Singapore 2.78%</td>
</tr>
<tr>
<td>Germany</td>
<td>2.59%</td>
<td>India 2.52%</td>
</tr>
<tr>
<td>India</td>
<td>2.54%</td>
<td>Germany 2.44%</td>
</tr>
<tr>
<td>Canada</td>
<td>1.93%</td>
<td>Canada 1.91%</td>
</tr>
<tr>
<td>All others (136)</td>
<td>25.65%</td>
<td>All others (131) 25.17%</td>
</tr>
</tbody>
</table>

- 10 countries responsible for **75%** of world’s working proxies
- Brazil home to **20%** of the proxies that forward **expected content** at least once

Little geographic and network diversity among the proxies
Performance

✨ **Goodput**: Average rate to download 1MiB file

Average goodput for proxies returning **expected content** - 128.5KiBps

IQR: [39.5; 160.9] KiBps
(Lack of) Anonymity

- Some open proxies claim to provide *anonymity* by hiding requestor’s IP

![Diagram](image)

- Inspected HTTP headers sent by proxy to destination
  - Hosted a simple web application that records HTTP headers

**66% of proxies expose requestor’s IP address via inserted HTTP headers**
Misbehavior in the Open Proxy Ecosystem

Proxies may not always return the expected content
(Mis)behavior

Proxies that respond to at least one request with non-zero content size and HTTP 2xx codes

~87% of proxies consistently deliver expected content

~5.1% of the proxies consistently returned unexpected content
Malicious Behavior

Proxies may perform **MitM** and alter content
HTML Manipulation

- Absence of end-to-end SSL/TLS makes MitM trivial

- HTML page
- Windows batch (.bat) file
- ZIP file
- JPEG Image
- Linux/UNIX shell script (.sh)
- Flash object (.swf)
- Java JAR archive
- Windows executable (.exe)
HTML Manipulation

- Results from HTML pages fetched at one client
- 2018-05-07
- 13.4% of requests produce unexpected content
- ~80% of unexpected HTML pages were benign
Results from HTML pages fetched at one client

2018-05-07

13.4% of requests produce unexpected content

~80% of unexpected HTML pages were benign

~72% error pages
### HTML Manipulation

<table>
<thead>
<tr>
<th>Count</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equivalent</td>
<td>0.0</td>
</tr>
<tr>
<td>Misconfiguration</td>
<td>0.0</td>
</tr>
<tr>
<td>No Content</td>
<td>0.0</td>
</tr>
<tr>
<td>Truncated</td>
<td>0.0</td>
</tr>
<tr>
<td>Unauthorized</td>
<td>0.0</td>
</tr>
<tr>
<td>Ad Injection</td>
<td>0.0</td>
</tr>
<tr>
<td>Cryptojacking</td>
<td>0.0</td>
</tr>
<tr>
<td>Original + Ad Injection</td>
<td>0.0</td>
</tr>
<tr>
<td>Eavesdropping</td>
<td>0.0</td>
</tr>
<tr>
<td>Unlabeled</td>
<td>0.0</td>
</tr>
</tbody>
</table>

- Results from HTML pages fetched at one client
  - 2018-05-07
  - ~80% of unexpected content produced were unexpected HTML pages were benign
  - ~72% error pages

- ~16.6% of unexpected content correspond to malicious activity
Cryptojacking

~3% of HTML pages fetched include Javascript that perform Cryptojacking

Coinhive’s Monero cryptocurrency mining Javascript

Redacted part serves as identity for the user to be paid

Coinhive endpoints replaced with other domains
File Manipulation

- Absence of end-to-end SSL/TLS makes MitM trivial
  - HTML page
  - Windows batch (.bat) file
  - ZIP file
  - JPEG Image
  - Linux/UNIX shell script (.sh)
  - Flash object (.swf)
  - Java JAR archive
  - Windows executable (.exe)
File Manipulation

- **97K** successful requests through
- **21,385** open proxies
- **29,484** requests made by
- **6.76%** of the proxies constitute unexpected content
- Submitted **unexpected responses** to VirusTotal

0.2% of daily proxy responses are classified as malicious by VirusTotal
File Manipulation

- Despite file types requested, proxies sometimes returned other file types
  - Classified by filemagic and VirusTotal
File Manipulation

- **EXE Files**
  - **98.5% of unexpected responses for .exe files classified as malicious**
  - **Expire malware family**
    - Can be used to steal personal information and provide remote access
  - **Flavours of malware from Crypt and Artemis trojan families**

- **HTML Files**
  - **97% HTML files were labeled as malicious**
  - All infected with Monero cryptocurrency mining Javascript
File Manipulation

ISO Files

- All ISO files were flagged malicious
- Exactly 1MiB size
- All infected with Vitalia Trojan, a rootkit for Windows
Consistency of Malicious Proxies

Most malicious proxy returned malicious content 100% of the time

- ~50% of the proxies return malicious content at least twice
- >25% of the proxies return at least 10 malicious files
SSL/TLS Analysis

- Only 70% proxies that fetch expected content at least once support TLS

- We consider misbehavior along two dimensions
  - Degrading security of communication
    - E.g., TLS stripping, certificate manipulation
  - Alteration of content

No evidence of TLS stripping and alteration of content

Found evidence of TLS certificate manipulation
SSL/TLS Certificate Manipulation

- Additional content fetched over HTTPS
  - Fetched URLs with 3 types of certificates
    - Revoked certificate
    - Self-signed certificate
    - Valid & verifiable certificate
  
  1.06% of the proxies inserted a modified certificate

- Proxies that fetch expected content but modify TLS certificate
  - Suggests the proxies are eavesdropping over HTTPS

0.85% proxies that return expected responses appear to be eavesdropping
SSL/TLS Certificate Manipulation

Eavesdropping proxies per day

No evidence of proxies selectively targeting revoked or self-signed certificates
Comparison with Tor

- Tor provides anonymous TCP communication
  - Routing traffic through multiple relays using layered encryption

- Original data is visible only at the exit

- Without end-to-end encryption, Tor exits can eavesdrop or alter traffic

Found no evidence of TLS certificate or file manipulation by Tor exit relays
Summary

- **Open Proxies**: free and simple way to circumvent censorship and bypass regional content filters
- Limited number (< 8%) of advertised open proxies actually work
- Little geographic and network diversity among open proxies
- Found numerous forms of misbehavior
  - RATs, Trojans, Cryptojacking, Javascript injections, etc.
- Tor offers safe and far more reliable form of proxied communication

Risks of using open proxies are plentiful, and likely far outweigh benefits

An Extensive Evaluation of the Internet’s Open Proxies

Akshaya Mani, Tavish Vaidya, David Dworken, Micah Sherr
Summary

- Open Proxies: free and simple way to circumvent censorship and bypass regional content filters
- Limited number (<8%) of advertised open proxies actually work
- Little geographic and network diversity among open proxies
- Found numerous forms of misbehavior
  - RATs, Trojans, Cryptojacking, Javascript injections, etc.
- Tor offers safe and far more reliable form of proxied communication

Risks of using open proxies are plentiful, and likely far outweigh benefits

An Extensive Evaluation of the Internet’s Open Proxies

Akshaya Mani, Tavish Vaidya, David Dworken, Micah Sherr