ICLab: Detailed Probes for Network Censorship

Internet Measurement Village 2020
Presenter: Zachary Weinberg

https://iclab.org/  •  info@iclab.org
### ICLab, OONI, and CensoredPlanet

<table>
<thead>
<tr>
<th>Vantage points are:</th>
<th>ICLab</th>
<th>OONI</th>
<th>CensoredPlanet</th>
</tr>
</thead>
<tbody>
<tr>
<td>VPN servers</td>
<td>Volunteers</td>
<td>Open-access servers</td>
<td></td>
</tr>
<tr>
<td>Countries covered:</td>
<td>60</td>
<td>150</td>
<td>170</td>
</tr>
<tr>
<td>Networks covered:</td>
<td>200</td>
<td>20,000</td>
<td>Not documented</td>
</tr>
<tr>
<td>Measurements are:</td>
<td>Direct</td>
<td>Direct</td>
<td>Direct and indirect</td>
</tr>
<tr>
<td>URLs/measurement:</td>
<td>2000–5000</td>
<td>30 (typical)</td>
<td>2200</td>
</tr>
<tr>
<td>Data collected:</td>
<td>Decoded HTTP messages, DNS lookup results, Packet traces</td>
<td>Decoded HTTP messages, DNS lookup results</td>
<td>Blocked/not blocked only?</td>
</tr>
</tbody>
</table>
Overview

• **Background:** what is internet censorship and how is it done?

• Symptoms of censorship, visible at the client
  - Block pages
  - DNS inconsistencies
  - Packet forgeries

• Case studies

• Our code and data and how you can access them
Internet Censorship

- Interference with network traffic
- Goal: restrict access to specific content
- “Content filters” existed as early as 1996
- China’s “Great Firewall” in development since 1997
- Spreading ever since; affects majority of Internet users today
Censorship’s visible effects

Overt

Warning

Unable to connect
Firefox can’t establish a connection to the server at www.aries.com.

95-0040000000 000000 000000 000000
95-0040000000 000000 000000 000000
95-0040000000 000000 000000 000000
95-0040000000 000000 000000 000000
95-0040000000 000000 000000 000000
95-0040000000 000000 000000 000000
95-0040000000 000000 000000 000000
95-0040000000 000000 000000 000000
95-0040000000 000000 000000 000000

["Unable to connect" error message from Firefox]
Censorship in the network

A? health-info.example    192.0.2.33
DNS

SYN port 443
TCP

SYN/ACK port 32770

GET /category/safer-sex
HTTP

200 OK <html>...
Censorship in the network

DNS
A? health-info.example
203.0.113.1
192.0.2.33

TCP
SYN port 443
RST
SYN/ACK port 32770

HTTP
GET /category/safer-sex
301 Redirect
200 OK <html>...
Overview

- Background: what is internet censorship and how is it done?

- **Symptoms of censorship, visible at the client**
  - Block pages
  - DNS inconsistencies
  - Packet forgeries

- Case studies

- Our code and data and how you can access them
HTML-based Block Pages

"This URL has been blocked under instructions of a competent Government Authority or in compliance with the orders of a Court of competent jurisdiction....

***This URL has been blocked under Instructions of the Competent Government Authority or Incompliance to the orders of Hon'ble Court.***
"This URL has been blocked under instructions of a competent Government Authority or in compliance with the orders of a Court of competent jurisdiction....

***This URL has been blocked under
Instructions of the Competent Government Authority or Incompliance to the orders of Hon'ble Court***

HTTP Error 404 - File or Directory not found

HTTP Error 404 - File or Directory not found = http://bet365.com/
This URL has been blocked under instructions of a competent Government Authority or in compliance with the orders of a Court of competent jurisdiction....

**Error 403: Access Denied/Forbidden**

404. That's an error.

This IP has been automatically blocked. If you have questions, please email: b-20161207034016381@craigslist.org

HTTP Error 404 - File or Directory not found

HTTP Error 404 - File or Directory not found = http://bet365.com/
This URL has been blocked under instructions of a competent Government Authority or in compliance with the orders of a Court of competent jurisdiction....

***This URL has been blocked under Instructions of the Competent Government Authority or Incompliance to the orders of Hon'ble Court.***

**“Error 403: Access Denied/Forbidden”**

404. That’s an error.

HTTP Error 404 - File or Directory not found

HTTP Error 404 - File or Directory not found = http://bet365.com/

ACK+PSH
HTTP/1.1 200 OK
Connection: close
Content-Length: nnnn
Content-Type: text/html;
   charset=utf-8
<!DOCTYPE html PUBLIC
   "-//W3C//DTD HTML 4.01//EN">
<html>
<head><title></title></head>
<body>
<h0><font color="black">
visible message
</font></h0>
</body>
</html>
HTML-based Block Page Discovery

Similarity Clusters

Strip HTML, normalize text, then use locality-sensitive hashing

- abudhabiescort.com - This website is for sale!
- abu dhabi escort Resources and Information
- aglionline.org - This website is for sale!
- aglionline Resources and Information
- arabfaces.com - This website is for sale!
- arabfaces Resources and Information
- arabgames.com - This website is for sale!
- Resources and Information
- arabdate.com - This website is for sale!
- Resources and Information
- behidden.com - This website is for sale!
- Hidden Resources and Information
- mixed-drink.com - This website is for sale!
- Beverages Resources and Information
- nedosug.com - This website is for sale!
- Dating Resources and Information
- transportube.com - This website is for sale!
- transportube Resources and Information
- unifemcis.org - This website is for sale!
- unifemcis Resources and Information

33 new blockpages discovered

Tag Frequency Vectors

"Automated Detection and Fingerprinting of Censorship Block Pages,"
Ben Jones, Tzu-Wen Lee, Nick Feamster, Philippa Gill.
IMC 2014
## URL-to-country ratio

<table>
<thead>
<tr>
<th>Blockpage</th>
<th>Not a blockpage</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>&lt;html&gt;&lt;body&gt;&lt;script&gt;</code></td>
<td><code>&lt;html&gt;&lt;body&gt;&lt;script&gt;</code></td>
</tr>
<tr>
<td><code>location.replace('http://www.warning.or.kr/')</code></td>
<td><code>location.replace('http://redtube.com/')</code></td>
</tr>
<tr>
<td><code>&lt;/script&gt;&lt;/body&gt;&lt;/html&gt;</code></td>
<td><code>&lt;/script&gt;&lt;/body&gt;&lt;/html&gt;</code></td>
</tr>
<tr>
<td>1 country, 400 URLs</td>
<td>50 countries, 70 URLs</td>
</tr>
</tbody>
</table>
DNS Inconsistencies

Control Node
A? example.com
Control ISP’s DNS server
192.0.2.22
192.0.2.22
192.0.2.22
192.0.2.22
NXDOMAIN
192.0.2.22 and 203.0.113.3
Censored
Censored
Censored (second reply injected)

Field Node
A? example.com
Field ISP’s DNS server
192.0.2.22
192.0.2.22
192.0.2.22
192.0.2.22
10.0.0.99
192.0.2.22 and 203.0.113.3
Uncensored (same AS)
Uncensored
Look at IP-to-country ratio (could be a CDN)
TCP Packet Injection

Control Node
GET /page HTTP/1.1
Web server
HTTP response
Network error

Field Node
GET /page HTTP/1.1
Web server (censored?)

<table>
<thead>
<tr>
<th>HTTP response</th>
<th>HTTP response</th>
<th>Network error</th>
<th>HTTP response</th>
<th>Network error</th>
<th>HTTP response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same HTTP response, no packet collision</td>
<td>Different HTTP response, no packet collision</td>
<td>Same network error</td>
<td>Packet collision, 2 data packets, neither is a blockpage</td>
<td>Different network error</td>
<td>Network error during TCP handshake</td>
</tr>
<tr>
<td>Uncensored (geotargeting or geoblocking)</td>
<td>(two-bit error)</td>
<td>(site outage)</td>
<td>Possibly censored</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Overview

- Background: what is internet censorship and how is it done?
- Symptoms of censorship, visible at the client
  - Block pages
  - DNS inconsistencies
  - Packet forgeries
- Case studies
- Our code and data and how you can access them
Case Study 1: Yemen

- Ongoing civil war (2015-present)
- Overt Internet censorship known to occur since 2010
  - focused on content forbidden by Islamic law (gambling, alcohol, etc.)
- Overt censorship increased in early 2015
- At the same time, other sites started producing a generic 404 Not Found message
  - these were all news sites (local and foreign) critical of the *de facto* government
- Analysis of IP-ID and TTL fields in packet headers revealed the generic message was injected by the same host as the overt blockpage

"Information Controls During Military Operations: The case of Yemen during the 2015 political and armed conflict"
Dalek et al

http://hdl.handle.net/1807/92750
## Case Study 2: Iranian Censorship or US Sanctions?

<table>
<thead>
<tr>
<th>URL</th>
<th>Status in Iran</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="https://github.com/">https://github.com/</a></td>
<td>Accessible</td>
</tr>
<tr>
<td><a href="https://azure.microsoft.com/">https://azure.microsoft.com/</a></td>
<td>Accessible</td>
</tr>
<tr>
<td><a href="https://www.dropbox.com/">https://www.dropbox.com/</a></td>
<td>Accessible</td>
</tr>
<tr>
<td><a href="https://code.google.com/codejam">https://code.google.com/codejam</a></td>
<td>Accessible</td>
</tr>
<tr>
<td><a href="https://developers.google.com/">https://developers.google.com/</a></td>
<td>Blocked by server</td>
</tr>
<tr>
<td><a href="https://developer.android.com/">https://developer.android.com/</a></td>
<td>Blocked by server</td>
</tr>
<tr>
<td><a href="https://analytics.google.com/">https://analytics.google.com/</a></td>
<td>Blocked by server</td>
</tr>
<tr>
<td><a href="https://get.adobe.com/flashplayer">https://get.adobe.com/flashplayer</a></td>
<td>Blocked by ISP</td>
</tr>
<tr>
<td><a href="https://www.netflix.com/">https://www.netflix.com/</a></td>
<td>Blocked by ISP</td>
</tr>
<tr>
<td><a href="https://www.bitbucket.org/">https://www.bitbucket.org/</a></td>
<td>Blocked by ISP</td>
</tr>
</tbody>
</table>
Case Study 3: Surveillance Injection

```plaintext
<meta http-equiv="refresh" content="2;url=http://1688.com/?" />
<iframe id="f" frameborder="0" style="width:1;height:1" />
</iframe>
<script>
    document.getElementById("f").src = 
        "http://...tm/?" + "a=FF&b=WIN&c=2975758&d=31&e=424" + "&f=MTY4OC5jb20=&g=1488529287299&h=Date.now()" + "&y=10&z=0&w=2017-01-10" + "&in=424_00030563&id=20170103"
</script>
```
Using Our Code and Data

- All our code is open-source and available on our Github account: https://github.com/iclab/

- Contact us for access to raw data

- Semi-processed data can be downloaded from the Internet Archive: https://archive.org/details/@iclab

- An interactive data explorer will go live on https://iclab.org/ as soon as our sysadmin gets back from vacation
Teaser: the data explorer (Brazil)

<table>
<thead>
<tr>
<th>Distinct URLs tested</th>
<th>Distinct censored URLs</th>
<th>Page measurements</th>
<th>Measurement rounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>5705</td>
<td>123</td>
<td>185226</td>
<td>38</td>
</tr>
</tbody>
</table>
Teaser: the data explorer (Brazil)

### Measurement List

<table>
<thead>
<tr>
<th>#</th>
<th>Experiment ID</th>
<th>Measurement Date</th>
<th>Website List</th>
<th>Censored URLs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>19481</td>
<td>2019-2-25 13h55m49s</td>
<td>alexa-global</td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>19288</td>
<td>2019-2-25 13h55m49s</td>
<td>citizenlab-global</td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>19387</td>
<td>2019-2-25 13h55m49s</td>
<td>country-sensitive-br</td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>19502</td>
<td>2019-2-25 13h55m49s</td>
<td>encore-global</td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>19661</td>
<td>2019-4-8 14h18m13s</td>
<td>alexa-global</td>
<td>Show ▼ +8 -0</td>
</tr>
<tr>
<td>6.</td>
<td>19505</td>
<td>2019-4-8 14h18m13s</td>
<td>citizenlab-global</td>
<td>Show ▼ +65 -0</td>
</tr>
</tbody>
</table>
**Teaser: the data explorer (Brazil)**

**Measurement List**

<table>
<thead>
<tr>
<th>#</th>
<th>Experiment ID</th>
<th>Measurement Date</th>
<th>Website List</th>
<th>Censored URLs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>19481</td>
<td>2019-2-25 13h55m49s</td>
<td>alexa-global</td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>19288</td>
<td>2019-2-25 13h55m49s</td>
<td>citizenlab-global</td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>19387</td>
<td>2019-2-25 13h55m49s</td>
<td>country-sensitive-br</td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>19002</td>
<td>2019-2-25 13h55m49s</td>
<td>encore-global</td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>19661</td>
<td>2019-4-8 14h18m13s</td>
<td>alexa-global</td>
<td><a href="http://ssus.com/">http://ssus.com/</a></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="http://espnoninfo.com/">http://espnoninfo.com/</a></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="http://kaikaku.com/">http://kaikaku.com/</a></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="http://livedoor.com/">http://livedoor.com/</a></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="http://livedoor.jp/">http://livedoor.jp/</a></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="http://office.com/">http://office.com/</a></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="http://planet.net/">http://planet.net/</a></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="http://rt.com/">http://rt.com/</a></td>
</tr>
</tbody>
</table>