Developing a Science of Internet Censorship Resistance: Opportunities and Challenges for Network Measurement

Phillipa Gill
University of Massachusetts -- Amherst
The challenge of network measurement

What we want to measure != What we are able to measure 😞

Decentralized Internet:

• No global measurement primitives
• Strong incentives to hide information
• How can we learn network paths?

Variety of edge networks:

• Broadband subscribers
• Mobile
• Data centers
• How can we get vantage points?
The challenge of network measurement

Variety of content

- Mobile apps + games
- Streaming video
- Social networks
- How to measure diverse apps?

Measuring censorship/network interference compounds the challenge!

Measuring censorship/network interference

- Networks are not always transparent in their actions
- Interference may not be consistent
- Measurements may pose a risk to users
- How to measure when the subject may hide?
Winning the censorship arms race

• Collaborative approach
  – Bringing together political science context and cutting edge network measurement techniques

• Circumvention
  – Improving Tor anonymity with empirical data
  – Designing covert channels with adaptability in mind

• Measuring politically motivated actors
  – Targeted attacks against activist groups
  – Fingerprinting filtering products
Censorship Measurement 101

- Example:

Measured in the lab
Managed in the field

Standard question:
Is this Web site blocked?
Censorship Measurement 101

• Example:

Measured in the lab

Measured in the field

(no html page returned)

Standard question:
Is this Web site blocked?

We need finer grained measurements to answer this question!
Censorship Measurement 101

- Example:

  Measured in the lab
  (no html page returned)

  Measured in the field

Standard question:
Is this Web site blocked?

What if we want to ask more questions:
  How was this site blocked?
  What product was used to block it?
  Who is blocking it?
What does this mean for measurement?

• Don’t know the set of measurements to be support a priori
  • Need to keep up with new censorship technologies

• Need to be able to implement and launch new experiments

• Need to be flexible in terms of when, where, and what is run

• How to do this well?

Trade off

Flexibility

Security for clients
Overview of ICLab

Clients

Experiments to run + relevant data

Results

Control Server
Overview of ICLab

Clients

Experiments to run + relevant data

Results

Control Server

Web page, reports, papers

Data analysis code (e.g., block page detection, device fingerprinting)

Database
ICLab use case 1: Yemen conflict

Information Controls during Military Operations

THE CASE OF YEMEN DURING THE 2015 POLITICAL AND ARMED CONFLICT

21 October 2015
ICLab use case 1: Yemen conflict

Netsweeper known to be used in Yemen with this block page

**Tests run by a pseudonymous researcher who was already in the country.**
ICLab use case 1: Yemen conflict

Is this really an HTTP 404 from the server or is it filtering?

We can look at network packet headers to find out!

Figure 20: HTTP 404 error displayed during an attempt to access sahafah.net on YemenNet
ICLab use case 2: Iranian filtering

Filterwatch // Episode 5 — Measuring Internet Censorship

On this month's show, we'll talk about a recent report published by Small Media and ICLab which aimed to test whether and by what method certain websites were blocked in Iran. We'll speak to PhD candidate Abbas Razaghpanah, who led the technical research on the report, about some of the findings and what they mean.

LISTEN

Monthly blog/podcast with SmallMedia UK.

Tests carried out using a virtual private server.
ICLab use case 2: Iranian filtering

Key question: How do tech sanctions influence Internet censorship?

<table>
<thead>
<tr>
<th>URL</th>
<th>VPS-1</th>
<th>VPS-2</th>
<th>VPS-3</th>
<th>VPS-4</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="https://developers.google.com/">https://developers.google.com/</a></td>
<td>Serverside</td>
<td>Serverside</td>
<td>Serverside</td>
<td>Serverside</td>
</tr>
<tr>
<td><a href="https://get.adobe.com/flashplayer/">https://get.adobe.com/flashplayer/</a></td>
<td>Local (RST)</td>
<td>OK</td>
<td>OK</td>
<td>Local (RST)</td>
</tr>
<tr>
<td><a href="https://analytics.google.com/">https://analytics.google.com/</a></td>
<td>Serverside</td>
<td>Serverside</td>
<td>Serverside</td>
<td>Serverside</td>
</tr>
<tr>
<td><a href="https://code.google.com/codejam">https://code.google.com/codejam</a></td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
</tr>
<tr>
<td><a href="https://github.com/">https://github.com/</a></td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
</tr>
<tr>
<td><a href="https://azure.microsoft.com">https://azure.microsoft.com</a></td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
</tr>
<tr>
<td><a href="https://aws.amazon.com/ec2/">https://aws.amazon.com/ec2/</a></td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
</tr>
<tr>
<td><a href="https://www.dropbox.com/">https://www.dropbox.com/</a></td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
</tr>
<tr>
<td><a href="https://www.netflix.com/">https://www.netflix.com/</a></td>
<td>OK</td>
<td>Local (DNS)</td>
<td>Local (RST)</td>
<td>OK</td>
</tr>
<tr>
<td><a href="https://www.bitbucket.org">https://www.bitbucket.org</a></td>
<td>Local (RST)</td>
<td>Local (RST)</td>
<td>Local (RST)</td>
<td>Local (RST)</td>
</tr>
</tbody>
</table>

Notes:

RST indicates that a website was blocked by the Iranian authorities via packet injection.

DNS indicates that a website was blocked by the Iranian authorities via DNS tampering.
ICLab use case 2: Iranian filtering

403. That’s an error.

Your client does not have permission to get URL / from this server. That’s all we know.

The page we were redirected to when we tried to access Google services from inside Iran.

Looks like server-side filtering to enforce the sanctions …
What is needed?

• Still hitting fundamental challenges in network measurement!
  – IP geolocation
  – Measuring/predicting network paths

• Pushing the boundaries of measurement
  – Using TCP side channels to study censorship
  – Trade off between scale and depth of measurement

• We need to understand how censors behave!
  – How can we evaluate circumvention schemes without understanding the adversary?
  – Complicated by political factors

More info:
http://www.cs.umass.edu/~phillipa
@phillipa_gill