Characterizing Global Web Censorship: Why is it so hard?

Phillipa Gill
The Citizen Lab/Stony Brook University

Work done in collaboration with:
Masashi Crete Nishihata, Jakub Dalek, Sharon Goldberg,
Adam Senft and Greg Wiseman
Overview

Large-scale politically driven Internet outages are well known…

• …but what happens within countries is less well understood

We leverage data gathered by an interdisciplinary group (Open Net Initiative) to bootstrap analysis

• 77 countries, 286 distinct ISPs, measured from 2007-2012
• **Advantages**: context about what, when, and where to measure
• **Disadvantages**: dearth of technical data/raw measurements

Our results highlight important challenges for censorship research!
Background

• Where censorship can happen:
Background

• Where censorship can happen:

Start

DNS reply?
Background

• Where censorship can happen:

Start

DNS reply?

No

DNS blocking
Background

• Where censorship can happen:

Start

DNS reply?

Yes

DNS redirect?

No

DNS blocking
Background

• Where censorship can happen:

Start

No

DNS reply?

Yes

DNS redirect?

No

DNS blocking

Yes

Response to SYN?
Background

• Where censorship can happen:

Start

Yes

DNS reply?

No

DNS blocking

Yes

DNS redirect?

No

IP blocking

No

Response to SYN?
Background

- Where censorship can happen:
Background

- Where censorship can happen:
Background

• Where censorship can happen:
Background

• Where censorship can happen:
Background

• Where censorship can happen:

Start

DNS reply?

Yes

DNS redirect?

No

DNS blocking

No

IP blocking

Response to SYN?

Yes

No

Response to HTTP request?

Yes

No

No HTTP Reply

RST

Infinite HTTP Redirect

What was it?

Yes

Block page
Methodology

- **Basic idea:** Issue requests for a consistent set of sites in the field and a control location (lab).
- Software synchronizes the requests between lab and field.
- Once both lab and field have completed, results sent back to the lab for more analysis.

- **What is tested:**
  - Sites that are likely to trigger censorship
  - Determined in collaboration with regional groups

- **Where are tests run:**
  - Combination of targeted/opportunistic testing
  - Performed by regional collaborators after informed consent meeting
Challenges for censorship research
1. Variation between countries

<table>
<thead>
<tr>
<th>Country</th>
<th>No DNS Reply</th>
<th>DNS Redirection</th>
<th>No HTTP Reply</th>
<th>RST</th>
<th>Blockpage</th>
</tr>
</thead>
<tbody>
<tr>
<td>China</td>
<td></td>
<td>0.15</td>
<td>0.75</td>
<td>0.05</td>
<td>0.15</td>
</tr>
<tr>
<td>Iran</td>
<td></td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>UAE</td>
<td>0.50</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.50</td>
</tr>
<tr>
<td>Yemen</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>Burma</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>Vietnam</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
</tbody>
</table>
1. Variation between countries

![Graph showing the fraction of blocking results for different countries.]

- **Country**: China, Iran, UAE, Yemen, Burma, Vietnam
- **Categories**: No DNS Reply, DNS Redirection, No HTTP Reply, RST, Blockpage

*Note: The graph visually represents the fraction of blocking results for each category in each country.*
1. Variation between countries

<table>
<thead>
<tr>
<th>Country</th>
<th>No DNS Reply</th>
<th>DNS Redirection</th>
<th>No HTTP Reply</th>
<th>RST</th>
<th>Blockpage</th>
</tr>
</thead>
<tbody>
<tr>
<td>China</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iran</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UAE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yemen</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Burma</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vietnam</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Red: No DNS Reply
- Green: DNS Redirection
- Blue: No HTTP Reply
- Pink: RST
- Cyan: Blockpage
1. Variation between countries

There is no such thing as a “representative” country
2. Variation between ISPs

Decentralized blocking in UAE

Fraction of content blocked

Year

2007 2008 2009 2010 2011 2012

AS 5384 AS 15802
2. Variation between ISPs

Decentralized blocking in UAE

“Du” ISP does not censor prior to April 2008
Censorship is a per-ISP property (when censorship is decentralized)
2. Variation between types of networks
2. Variation between types of networks

Academic networks block an average of 40% less!
2. Variation between types of networks

Academic networks block an average of 40% less!

Academic networks are not representative!
3. Sudden temporal shifts in blocking

Censorship in Burma over time

Fraction of tests blocked

Year

2009 2010 2011 2012

Political  Social  Internet  Conflict
3. Sudden temporal shifts in blocking

Censorship in Burma over time

End of military rule in 2011 brought political reforms.
3. Sudden temporal shifts in blocking

Censorship in Burma over time

End of military rule in 2011 brought political reforms.

Need to measure over time and correlate with political changes
4. Stealthy blocking of certain content

Censorship of content in Yemen

Fraction of block results

<table>
<thead>
<tr>
<th>Theme</th>
<th>Political</th>
<th>Social</th>
<th>Internet</th>
<th>Conflict</th>
</tr>
</thead>
<tbody>
<tr>
<td>No DNS Reply</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No HTTP Reply</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RST</td>
<td>0.02</td>
<td>0.40</td>
<td>0.00</td>
<td>0.15</td>
</tr>
<tr>
<td>Blockpage</td>
<td>0.30</td>
<td>1.00</td>
<td>1.00</td>
<td>0.45</td>
</tr>
</tbody>
</table>

Legend:
- No DNS Reply
- No HTTP Reply
- RST
- Blockpage
4. Stealthy blocking of certain content

Censorship of content in Yemen

<table>
<thead>
<tr>
<th>Theme</th>
<th>No DNS Reply</th>
<th>No HTTP Reply</th>
<th>RST</th>
<th>Blockpage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Political</td>
<td></td>
<td></td>
<td>0.1</td>
<td>0.3</td>
</tr>
<tr>
<td>Social</td>
<td></td>
<td></td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Internet</td>
<td></td>
<td>0.1</td>
<td>0.2</td>
<td>0.8</td>
</tr>
<tr>
<td>Conflict</td>
<td>0.4</td>
<td>0.6</td>
<td>0.2</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Transparent blocking of social and Internet content
4. Stealthy blocking of certain content

Censorship of content in Yemen

Fraction of block results

<table>
<thead>
<tr>
<th>Theme</th>
<th>Political</th>
<th>Social</th>
<th>Internet</th>
<th>Conflict</th>
</tr>
</thead>
<tbody>
<tr>
<td>No DNS Reply</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No HTTP Reply</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RST</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blockpage</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

“No DNS Reply” and “No HTTP Reply” are low for all themes, indicating limited direct censorship. "RST" and "Blockpage" are more prevalent in political and conflict-related content, suggesting "stealthy" blocking techniques.

“Stealthy” blocking of political and conflict related content
4. Stealthy blocking of certain content

Censorship of content in Yemen

Measurement needs to be robust to distinguish failure from censorship
5. The type of content tested matters

![Bar chart showing the fraction of content blocked in different countries: China, UAE, Yemen, Burma, Malaysia, Kyrgyzstan, Egypt. The x-axis represents the countries, and the y-axis represents the fraction blocked. The chart compares local and global content.](chart.png)
5. The type of content tested matters

3-5X more blocking of local content in China/Yemen
* most blocked content is political
5. The type of content tested matters

Less discrepancy in UAE
* most blocked content is social
5. The type of content tested matters

Need to take an interdisciplinary approach to determine what content to test
Challenges for censorship research:

1. Variations between technology used by countries
2. Variations between ISPs and between ISPs and institutions
3. Sudden temporal shifts in blocking
4. Stealthy blocking of certain content
5. Locally relevant content is more likely to be blocked

And more!

… maintaining infrastructure across funding cycles/staff turn over

… informed consent/preserving user privacy when testing can pose a physical risk!
What’s next?

More measurements, taking an interdisciplinary approach to tackle the problem:
• Rigorous measurements + political context

Data sharing?
• **Short answer:** we’re working on it.
• **Longer answer:** this project has laid the foundation in terms of unifying the data and removing PII.
  – Anticipate releasing data in the next ~4 months
What I hope to get out of this workshop

• Discuss how existing platforms may be used for censorship research

  **Particularly interested in:**
  – Platforms with visibility into the network edge
  – DNS/BGP measurements

• Discuss how a large scale, long-term censorship measurement platform may be built

• Discuss how we might distinguish transient failures/TCP bugs from actual censorship