

ICLab: Detailed Probes for Network Censorship

Internet Measurement Village 2020
Presenter: Zachary Weinberg

<https://iclab.org/> • info@iclab.org



Arian Alkhavan
Niaki



Shinyoung Cho



Zachary Weinberg



Nguyen Phong
Hoang



Abbas
Razaghpanah



Diogo Barradas



Nicolas Christin



Phillipa Gill

University of
Massachusetts
Amherst

 Stony Brook University

 INTERNATIONAL
COMPUTER SCIENCE
INSTITUTE

 Carnegie Mellon University
Security and Privacy Institute

 TÉCNICO
LISBOA  inescid
lisboa

ICLab, OONI, and CensoredPlanet

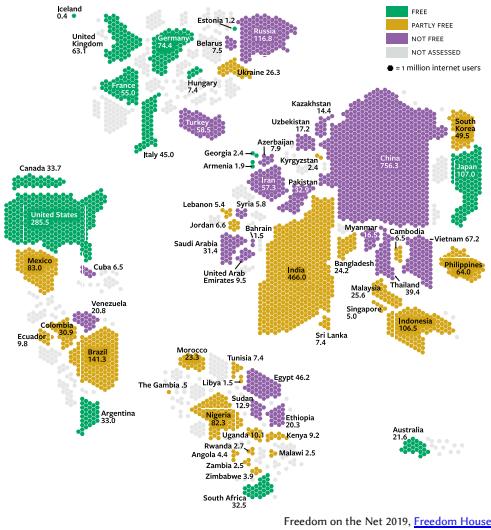
	ICLab	OONI	CensoredPlanet
Vantage points are:	VPN servers	Volunteers	Open-access servers
Countries covered:	60	150	170
Networks covered:	200	20,000	Not documented
Measurements are:	Direct	Direct	Direct and indirect
URLs/measurement:	2000–5000	30 (typical)	2200
Data collected:	Decoded HTTP messages DNS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

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



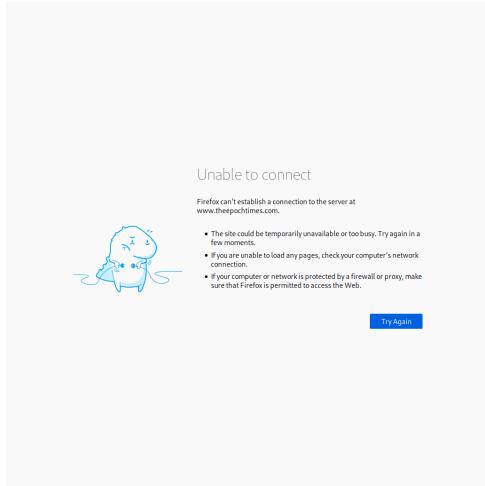
Censorship's visible effects

Overt



[Block page served by South Korea]

Covert



["Unable to connect" error message from Firefox] 5 / 25

Censorship in the network

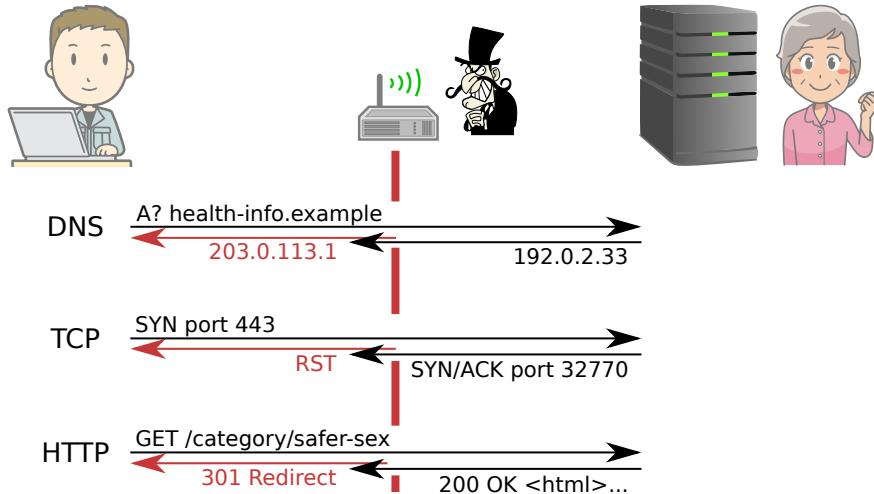


DNS A? health-info.example
 ←→
 192.0.2.33

TCP SYN port 443
 ←→
 SYN/ACK port 32770

HTTP GET /category/safer-sex
 ←→
 200 OK <html>...

Censorship in the network



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.

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....

*"Error 403: Access Denied/Forbidden"

404. That's an error.

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/>

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....

*"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*

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/

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.

*"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  
arabiadate.com - This website is for sale! -  
    arabiadate 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

```
<!DOCTYPE html PUBLIC  
  "-//W3C//DTD HTML 4.01//EN">  
<html>  
<head><title></title></head>  
<body>  
<h0><font color="black">  
</font></h0>  
</body>  
</html>
```

Tag	Count
!DOCTYPE	1
html	1
head	1
body	1
h0	1
font	1

15 new blockpages discovered

"Automated Detection and Fingerprinting of Censorship Block Pages,"
Ben Jones, Tzu-Wen Lee, Nick Feamster, Phillipa Gill,
IMC 2014

URL-to-country ratio

Blockpage

```
<html><body><script>
location.replace('http://www.warning.or.kr/')
</script></body></html>
```

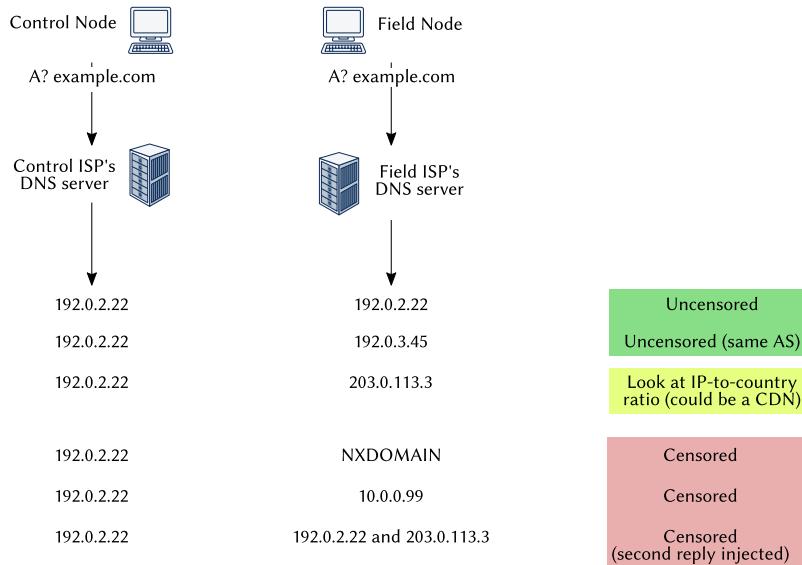
1 country, 400 URLs

Not a blockpage

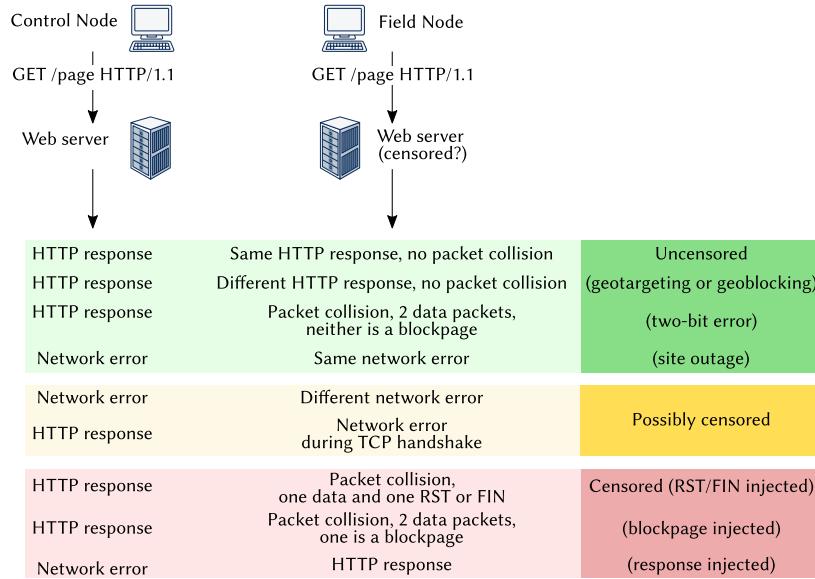
```
<html><body><script>
location.replace('http://redtube.com/')
</script></body></html>
```

50 countries, 70 URLs

DNS Inconsistencies



TCP Packet Injection



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?

URL	Status in Iran
https://github.com/	Accessible
https://azure.microsoft.com/	Accessible
https://www.dropbox.com/	Accessible
https://code.google.com/codejam	Accessible
https://developers.google.com/	Blocked by server
https://developer.android.com/	Blocked by server
https://analytics.google.com/	Blocked by server
https://get.adobe.com/flashplayer	Blocked by ISP
https://www.netflix.com/	Blocked by ISP
https://www.bitbucket.org/	Blocked by ISP

Case Study 3: Surveillance Injection

```
<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=MTY40C5jb20=&g=1488529287299&h="
  + Date.now()
  + "&y=10&z=0&x=28&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

ICLab Dashboard Home All Countries Censored Countries

Africa Americas Asia Europe Oceania MiddleEast

Africa

 Cape Verde 15 Measurements	 Central African Republic 12 Measurements	 Côte d'Ivoire 9 Measurements	 Djibouti 8 Measurements	 Ethiopia 8 Measurements	 Gambia 2 Measurements
--	--	--	---	---	---

Teaser: the data explorer (Brazil)

ICLab Dashboard

Home All Countries Censored Countries

 Brazil

Freedom House '18 classification:
Partly Free

Distinct URLs tested	Distinct censored urls	Page measurements	Measurement rounds
5705	123	185236	38

Teaser: the data explorer (Brazil)

Measurement List

#	Experiment ID	Measurement Date	Website List	Censored URLs
1.	19481	2019-2-25 13h55m49s	alexa-global	
2.	19288	2019-2-25 13h55m49s	citizenlab-global	
3.	19387	2019-2-25 13h55m49s	country-sensitive-br	
4.	19502	2019-2-25 13h55m49s	encore-global	
5.	19661	2019-4-8 14h18m13s	alexa-global	Show  +8 -0
6.	19505	2019-4-8 14h18m13s	citizenlab-global	Show  +65 -0

Teaser: the data explorer (Brazil)

Measurement List

#	Experiment ID	Measurement Date	Website List	Censored URLs
1.	19481	2019-2-25 13h55m49s	alexa-global	
2.	19288	2019-2-25 13h55m49s	citizenlab-global	
3.	19387	2019-2-25 13h55m49s	country-sensitive-br	
4.	19502	2019-2-25 13h55m49s	encore-global	
5.	19661	2019-4-8 14h18m13s	alexa-global	Show ▾ +8 -0 http://asus.com/ http://espnrcincinfo.com/ http://kakaku.com/ http://livedoor.com/ http://livedoor.jp/ http://office.com/ http://pixnet.net/ http://rt.com/