

# ICLab: Detailed Probes for Network Censorship

Internet Measurement Village 2020  
Presenter: Zachary Weinberg

<https://iclab.org/> • [info@iclab.org](mailto:info@iclab.org)



Arian Akhavan  
Niaki



Shinyoung Cho



Zachary Weinberg



Nguyen Phong  
Hoang



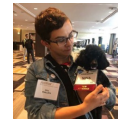
Abbas  
Razaghpanah



Diogo Barradas



Nicolas Christin



Phillipa Gill



## ICLab, OONI, and CensoredPlanet

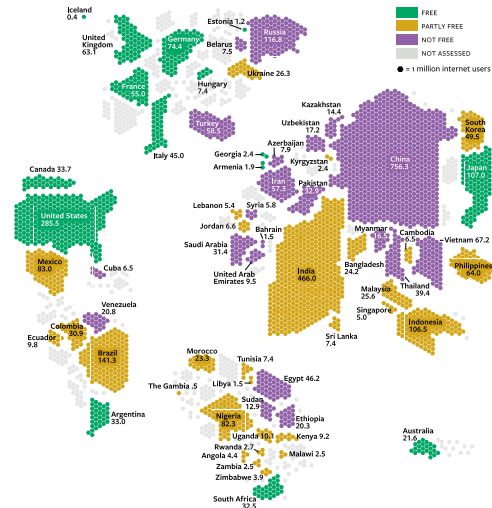
	<b>ICLab</b>	<b>OONI</b>	<b>CensoredPlanet</b>
<b>Vantage points are:</b>	VPN servers	Volunteers	Open-access servers
<b>Countries covered:</b>	60	150	170
<b>Networks covered:</b>	200	20,000	Not documented
<b>Measurements are:</b>	Direct	Direct	Direct and indirect
<b>URLs/measurement:</b>	2000–5000	30 (typical)	2200
<b>Data collected:</b>	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



Freedom on the Net 2019, [Freedom House](https://www.freedomhouse.org/freedom-on-the-net)

# Censorship's visible effects

## Overt

**Warning**

불법·유해 정보(사이트)에 대한 차단 안내

지금 접속하고 있는 정보(사이트)에서 불법·유해 내용이 제공되고 있어 이에 대한 접속이 차단되었습니다.

해당 정보(사이트)는 방송통신심의위원회(KCSC)의 심의를 거쳐 「방송통신위원회의 설치 및 운영에 관한 법률」에 따라 차단된 것이오니 이에 관한 문의사항이 있으시면 아래의 담당기관으로 문의하여 주시기 바랍니다.

\* 차단대상사이트(blocking.site)를 포함한 해당사이트가 열람되지 못할 경우가 있습니다.  
(차단대상사이트는 해당정보를 요구하거나 프로그램 실행을 유도하지 않습니다.)

사이트명	담당기관	전화번호
유행 도박	사이버 경찰청	1566-0112
불법 체육진흥추진 관련	사법선정법률구조위원회	1855-0112
	사법선정법률구조위원회	1855-0112
불법 승차부표권 구매대행	국민체육진흥공단 경유-경찰 총합본부	1899-1119
	국민체육진흥공단 경유-경찰 총합본부	1899-0707
유행 마포구 매대방	한국야사회	080-8282-112
유행 의약품 판매	식품의약품안전처 사이버조사단	0433719-1921
불법 의약품 판매 및 유통과대방고	식품의약품안전처 사이버조사단	0433719-1921
유행 의약품 판매 및 유통과대방고	식품의약품안전처 사이버조사단	0433719-1914
불법 의약품 판매 및 유통과대방고	식품의약품안전처 사이버조사단	0433719-1914
유행 의약품 판매 및 유통과대방고	식품의약품안전처 사이버조사단	0433719-1906
유행 의약품 판매	식품의약품안전처 사이버조사단	0433719-1906
유행 의약품 판매	식품의약품안전처 사이버조사단	0433719-2806
실시간 방송 중계	방송통신심의위원회	023219-5648
상표권(특수상용 유형)	한국저작권보호원	022383-5834
저작권(웹툰제작 유형)	한국저작권보호원 온라인보호팀	023153-2437
안보위협행위	사이버 경찰청	1566-0112
명예훼손, 초상권 침해	방송통신심의위원회	023219-3341-4
권투·합도, 격투·비하 등	방송통신심의위원회	023219-5161
디지털 불법정보	방송통신심의위원회	023219-5834

[Block page served by South Korea]

## Covert

Unable to connect

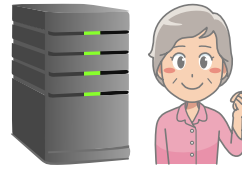
Firefox can't establish a connection to the server at www.thepachtimes.com.

- The site could be temporarily unavailable or too busy. Try again in a few moments.
- If you are unable to load any pages, check your computer's network connection.
- If your computer or network is protected by a firewall or proxy, make sure that Firefox is permitted to access the Web.

Try Again

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

# Censorship in the network

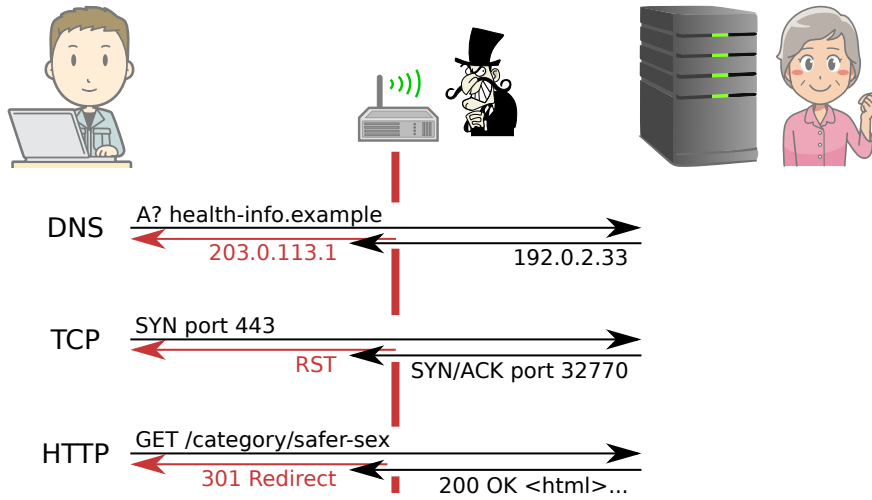


DNS  $\xrightarrow{\text{A? health-info.example}}$   
 $\xleftarrow{\text{192.0.2.33}}$

TCP  $\xrightarrow{\text{SYN port 443}}$   
 $\xleftarrow{\text{SYN/ACK port 32770}}$

HTTP  $\xrightarrow{\text{GET /category/safer-sex}}$   
 $\xleftarrow{\text{200 OK <html>...}}$

# 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*

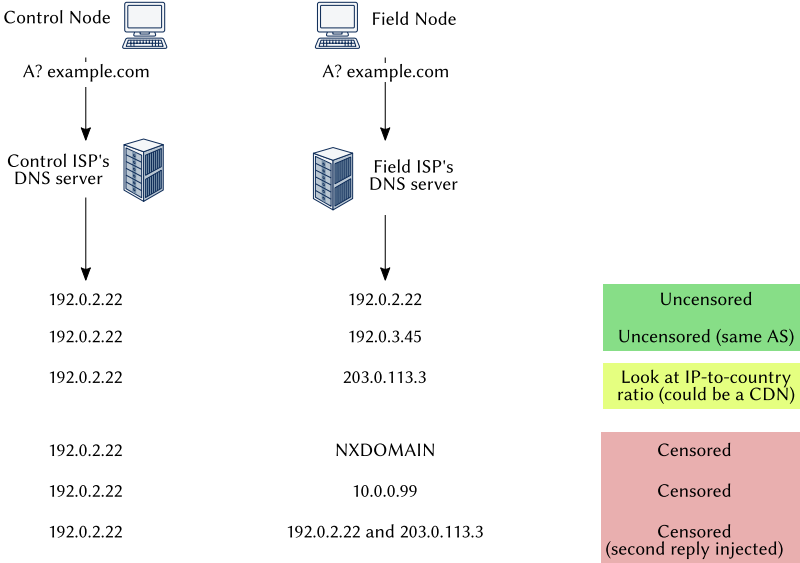
### Not a blockpage

---

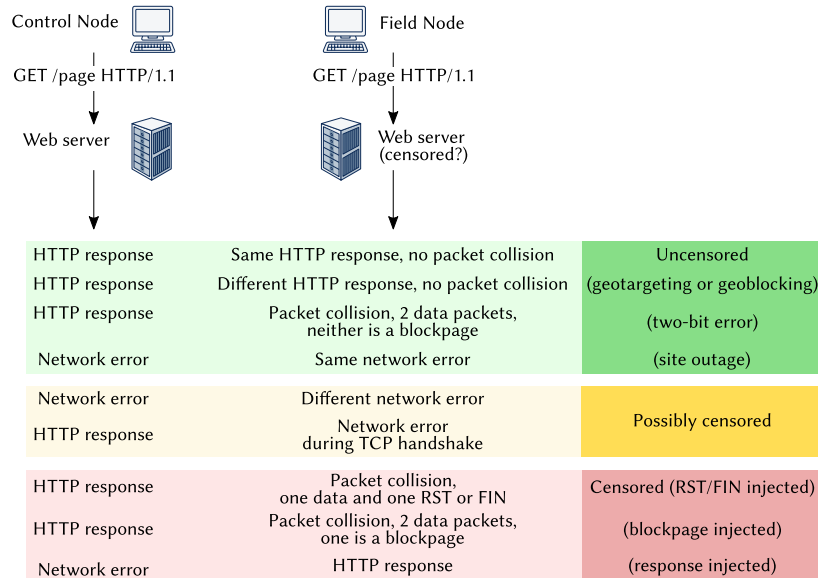
```
<html><body><script>
location.replace('http://redtube.com/')
</script></body></html>
```

*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
<a href="https://github.com/">https://github.com/</a>	Accessible
<a href="https://azure.microsoft.com/">https://azure.microsoft.com/</a>	Accessible
<a href="https://www.dropbox.com/">https://www.dropbox.com/</a>	Accessible
<a href="https://code.google.com/codejam">https://code.google.com/codejam</a>	Accessible
<a href="https://developers.google.com/">https://developers.google.com/</a>	Blocked by server
<a href="https://developer.android.com/">https://developer.android.com/</a>	Blocked by server
<a href="https://analytics.google.com/">https://analytics.google.com/</a>	Blocked by server
<a href="https://get.adobe.com/flashplayer">https://get.adobe.com/flashplayer</a>	Blocked by ISP
<a href="https://www.netflix.com/">https://www.netflix.com/</a>	Blocked by ISP
<a href="https://www.bitbucket.org/">https://www.bitbucket.org/</a>	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://[REDACTED]/tm/?"
  + "a=FF&b=WIN&c=2975758&d=31&e=424"
  + "&f=MTY4OC5jb20=&g=1488529287299&h="
  + Date.now()
  + "&y=10&z=0&x=2&w=2017-01-10"
  + "&i=424_00030563&iid=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](#)



Freedom House'18 classification:

Partly Free

Distinct URLs tested	Distinct censored uris	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	<input type="text"/>
4.	19502	2019-2-25 13h55m49s	encore-global	
5.	19661	2019-4-8 14h18m13s	alexa-global	<a href="#">Show</a> ⚠️ +8 -0
6.	19505	2019-4-8 14h18m13s	citizenlab-global	<a href="#">Show</a> ⚠️ +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  <a href="http://asus.com/">http://asus.com/</a> <a href="http://espnricinfo.com/">http://espnricinfo.com/</a> <a href="http://kakaku.com/">http://kakaku.com/</a> <a href="http://livedoor.com/">http://livedoor.com/</a> <a href="http://livedoor.jp/">http://livedoor.jp/</a> <a href="http://office.com/">http://office.com/</a> <a href="http://pixnet.net/">http://pixnet.net/</a> <a href="http://rt.com/">http://rt.com/</a>