University of Massachusetts Amherst

Stony Brook University

CyLab Carnegie Mellon University Security and Privacy Institute

ICLab: A Global, Longitudinal Internet Censorship Measurement Platform

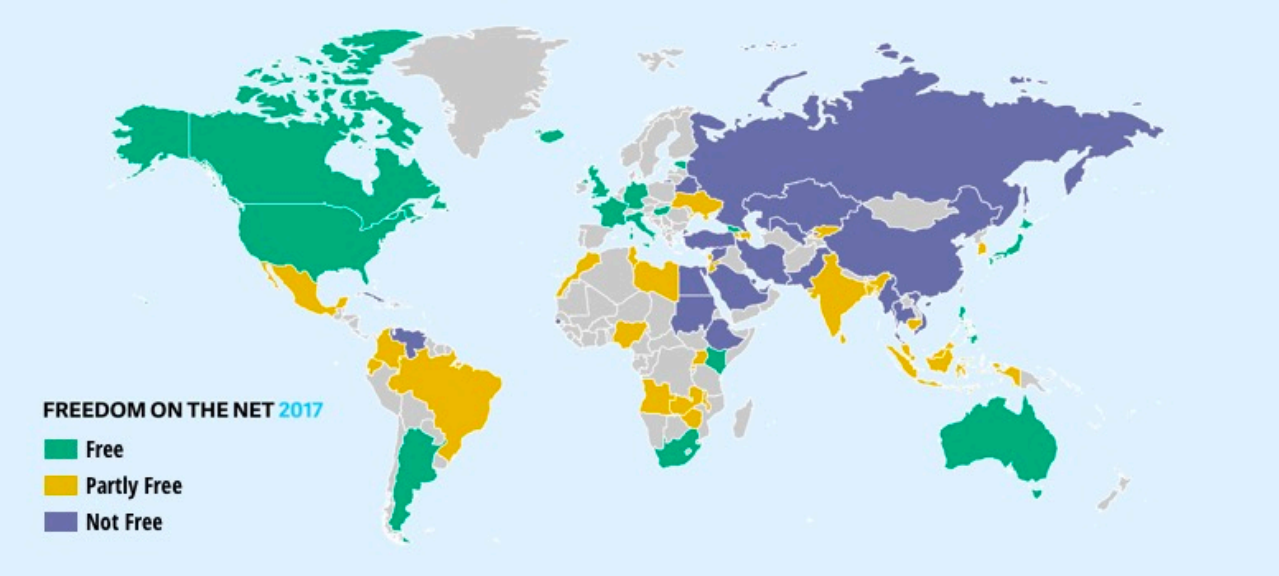
Arian Akhavan Niaki*[1], Shinyoung Cho*[12], Zachary Weinberg*[3]
Nguyen Phong Hoang[2], Abbas Razaghpanah[2], Nicolas Christin[3], Phillipa Gill[1]
[1]University of Massachusetts, Amherst, [2]Stony Brook University, [3]Carnegie Mellon University
Email: calipr@umass.edu, info@iclab.org






Internet Censorship

- Restrict access to specific web content



Freedom on the Net 2017, Freedom House

Huge Market for Surveillance/Censorship Products


THE WALL STREET JOURNAL.

TECH

Made-in-China Censorship for Sale

Chinese AI tools from tech giants like Alibaba make it easier to scrub online content—and anyone can buy them.

PHOTO ILLUSTRATION BY JOHN KUCZALA; ISTOCK

By Shan Li

March 6, 2020 5:30 am ET

https://www.wsj.com/articles/made-in-china-censorship-for-sale-11583448433?shareToken=st03477b3721ee44619a6ad5cd5e670454

Censorship products in China is predicted to grow to a **$70 billion industry** over the next 3-5 years

Estimated sales of **$5 billion** per year for surveillance/wiretapping products*

netsweeper

FINFISHER™
EXCELLENCE IN IT INVESTIGATION

BlueCoat

websense

*http://www.washingtonpost.com/world/national-security/trade-in-surveillance-technology-raises-worries/2011/11/22/gIQAFFZOGO_story.html

Why measure censorship?

BAD TRAFFIC

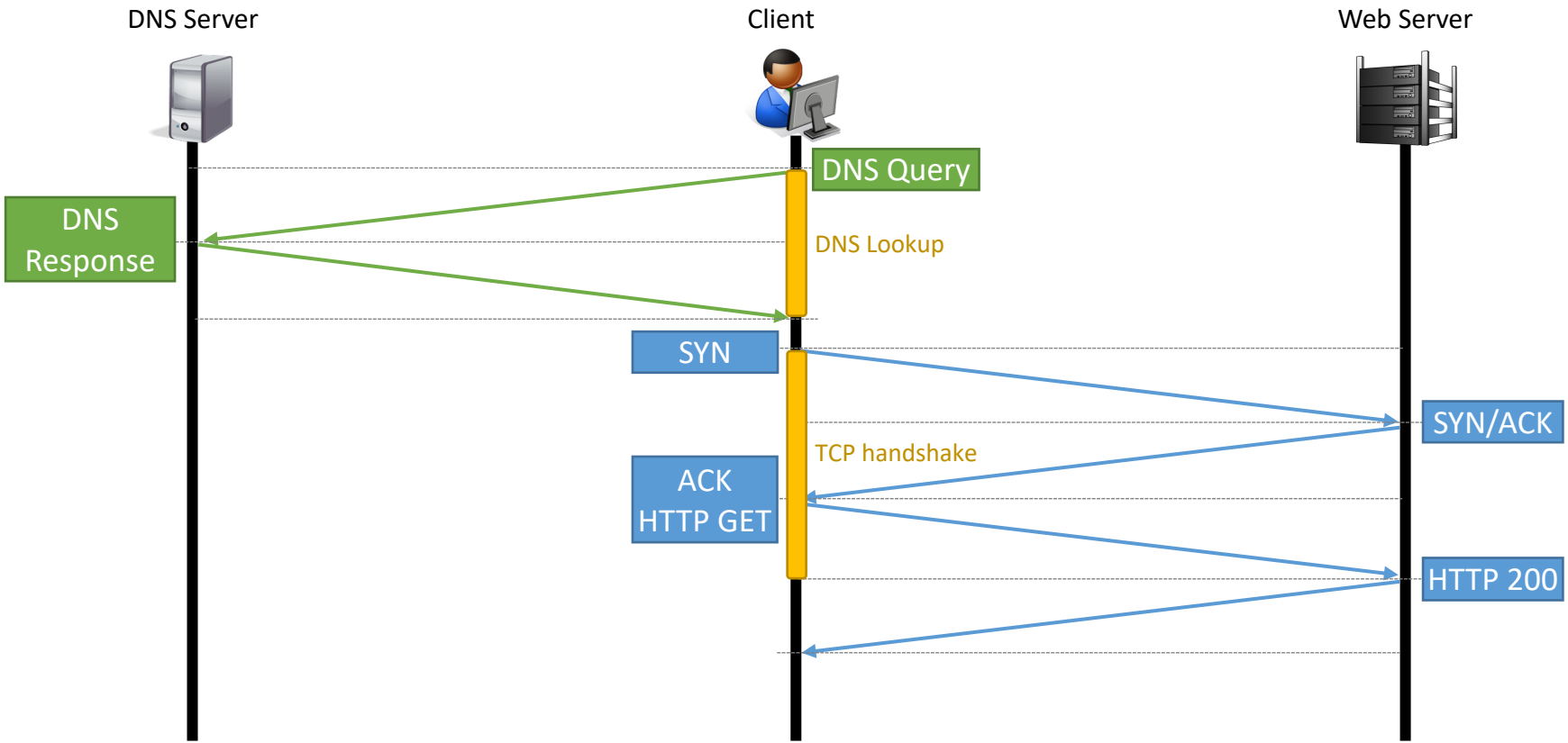
Sandvine's PacketLogic Devices Used to Deploy Government Spyware in Turkey and Redirect Egyptian Users to Affiliate Ads?

By Bill Marczak, Jakub Dalek, Sarah McKune, Adam Senft, John Scott-Railton, and Ron Deibert

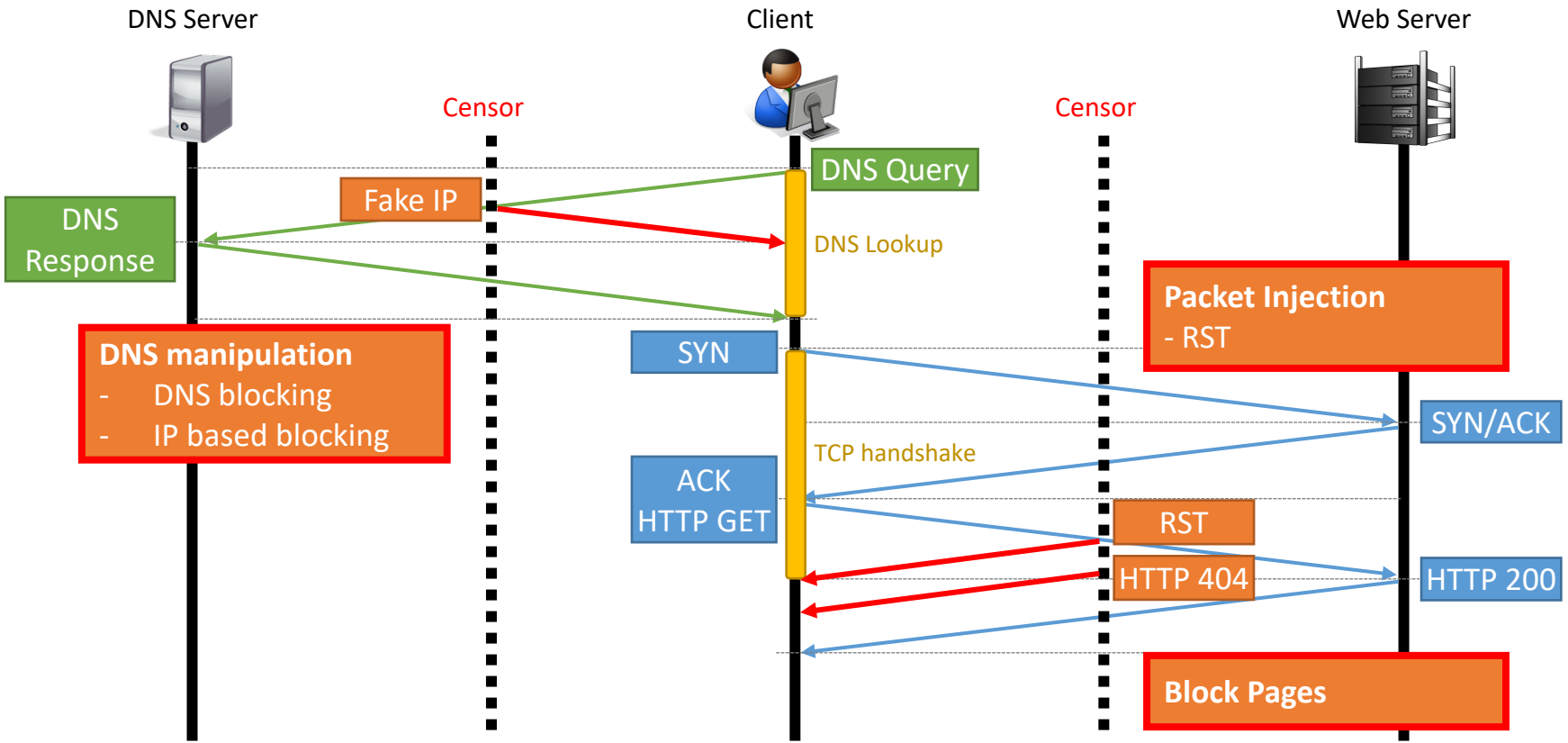
March 9, 2018 أزمة مرورية (Arabic translation), KÖTÜ TRAFİK (Turkish translation)

https://citizenlab.ca/2018/03/bad-traffic-sandvines-packetlogic-devices-deploy-government-spyware-turkey-syria/

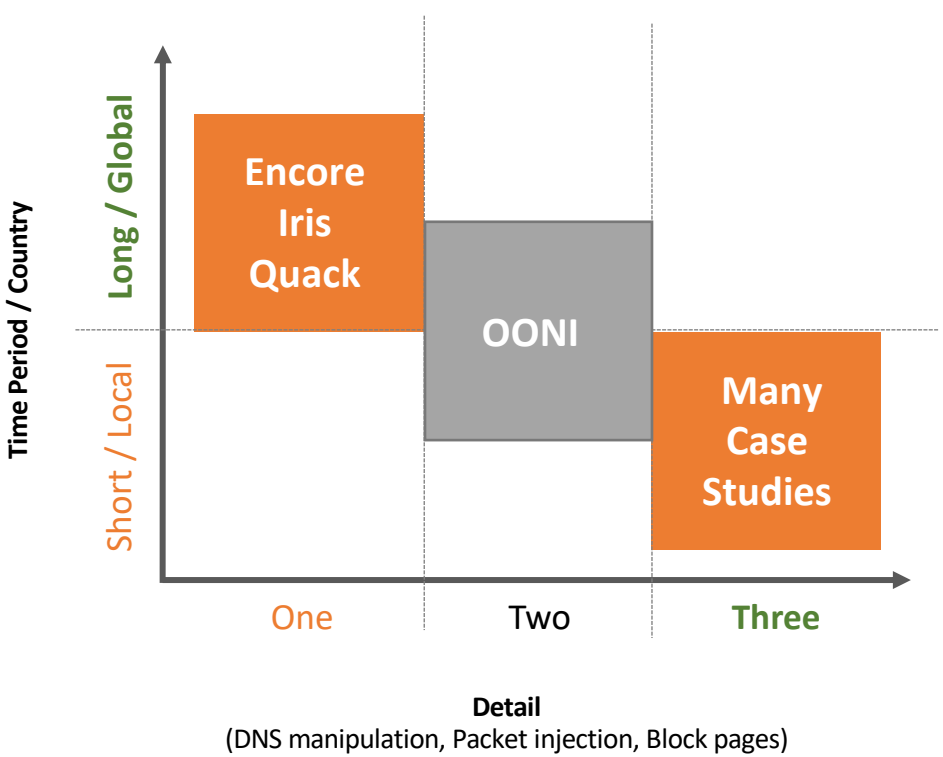
How does censorship work?



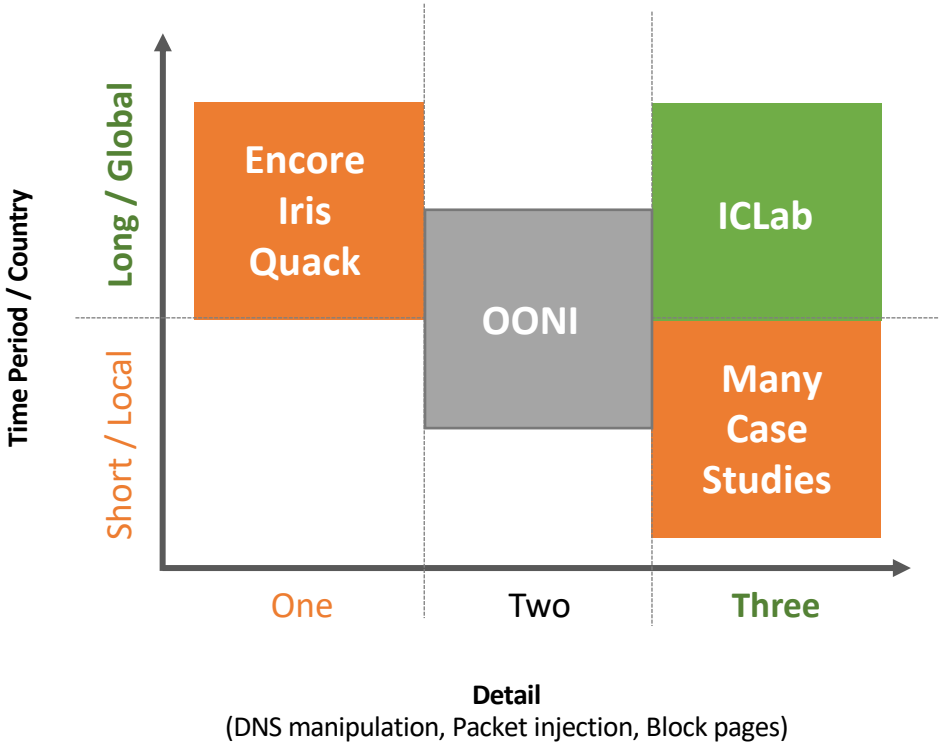
How does censorship work?



Studies in Censorship Research



Why we built our own platform?

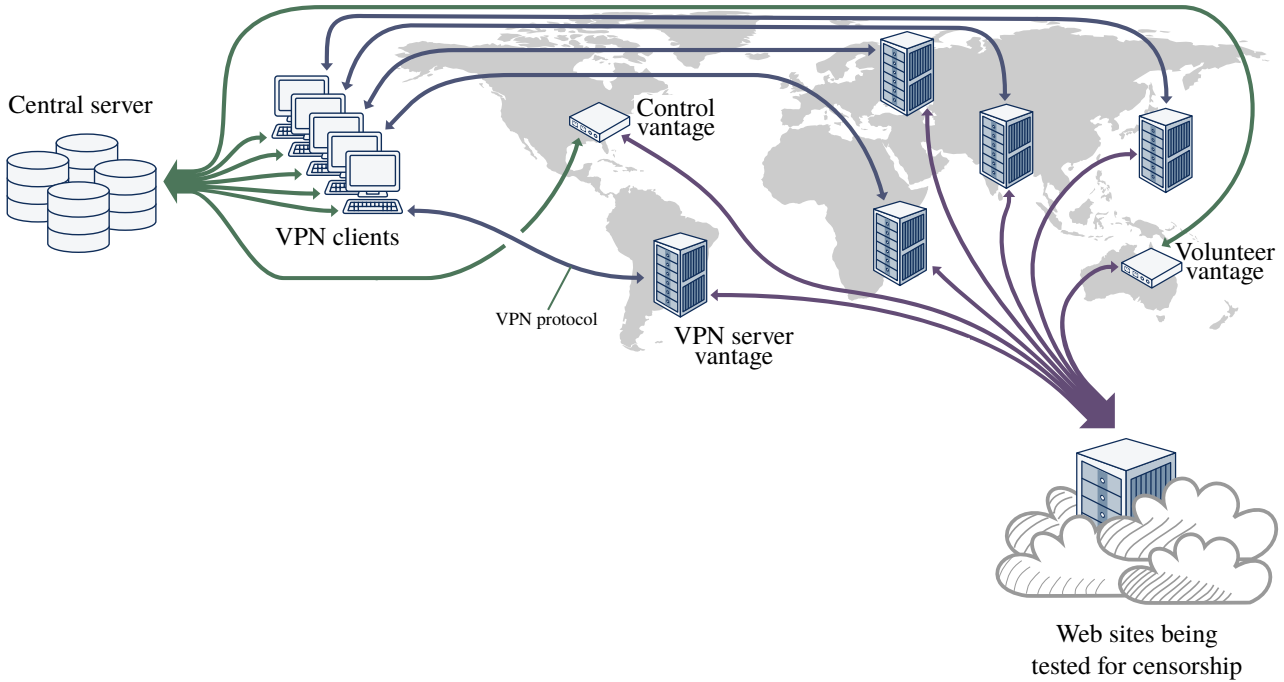


Volunteers:
- Ethical concerns
- Limited time period and # of countries

Non-volunteers:
- Only one method

- **Challenges: Obtaining vantage points** for **breadth** and **depth** data collection

ICLab Platform

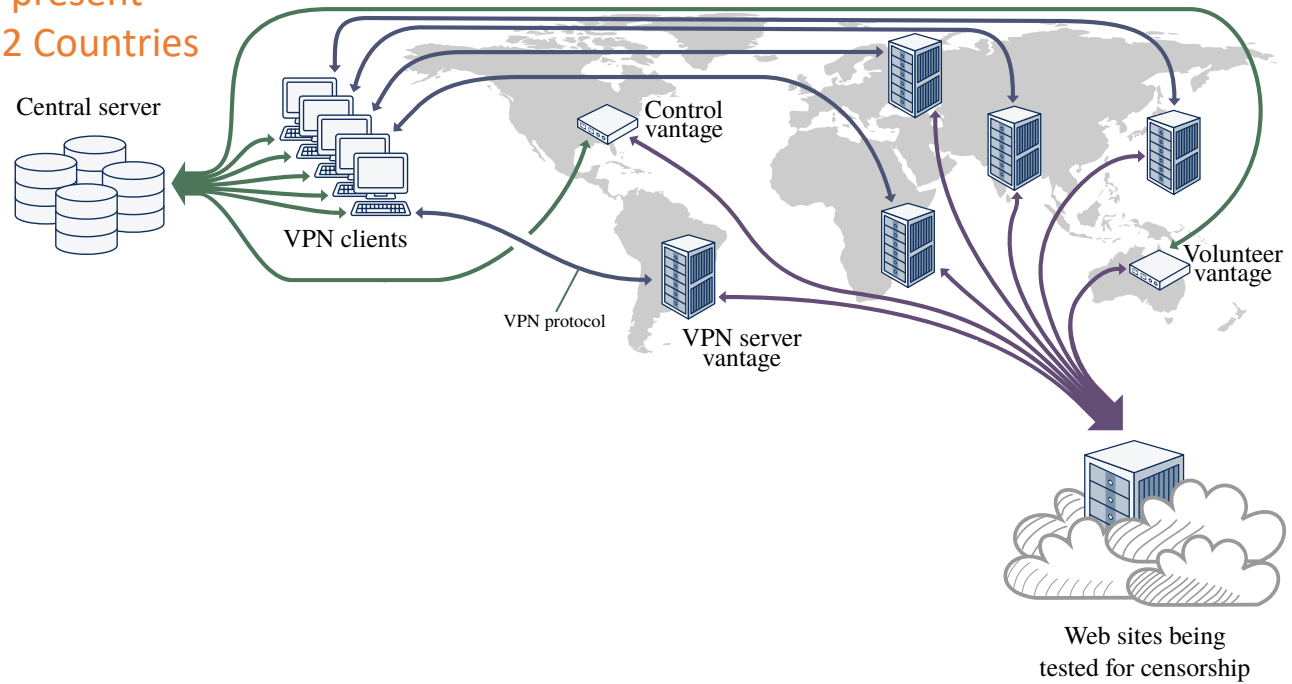


ICLab: Global and Continuous Monitoring

★ Global and Continuous

Oct. 2016 ~ present

234 ASes, 62 Countries

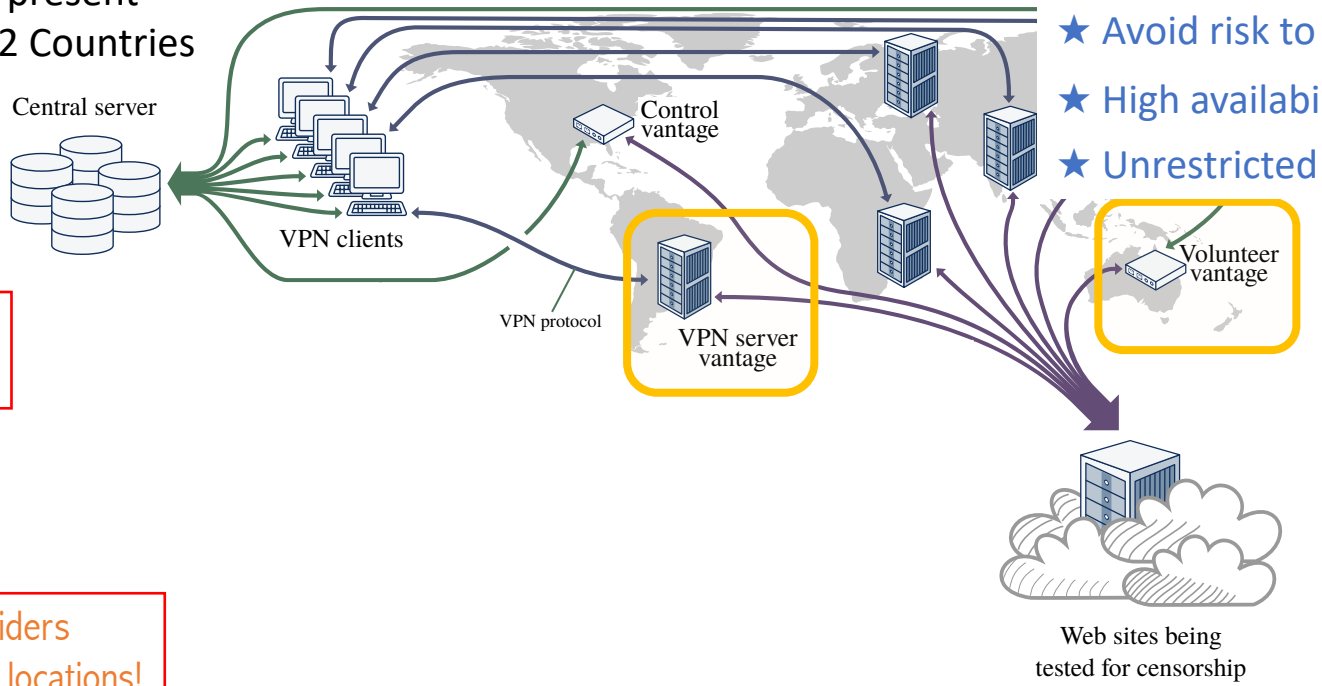


ICLab: Vantage Points

★ Global and Continuous
Oct. 2016 ~ present
234 ASes, 62 Countries

Vantage Points: (worldwide: 264 VPNs, 17 volunteers)

★ Avoid risk to volunteers
★ High availability, bandwidth
★ Unrestricted network access



Geolocation Checker

VPN providers lie their server locations!

- "**How to Catch when Proxies Lie**: Verifying the Physical Locations of Network Proxies with Active Geolocation," Zachary Weinberg, Shinyoung Cho, Nicolas Christin, Vyas Sekar, and Phillipa Gill, In *ACM IMC,* Nov. 2018

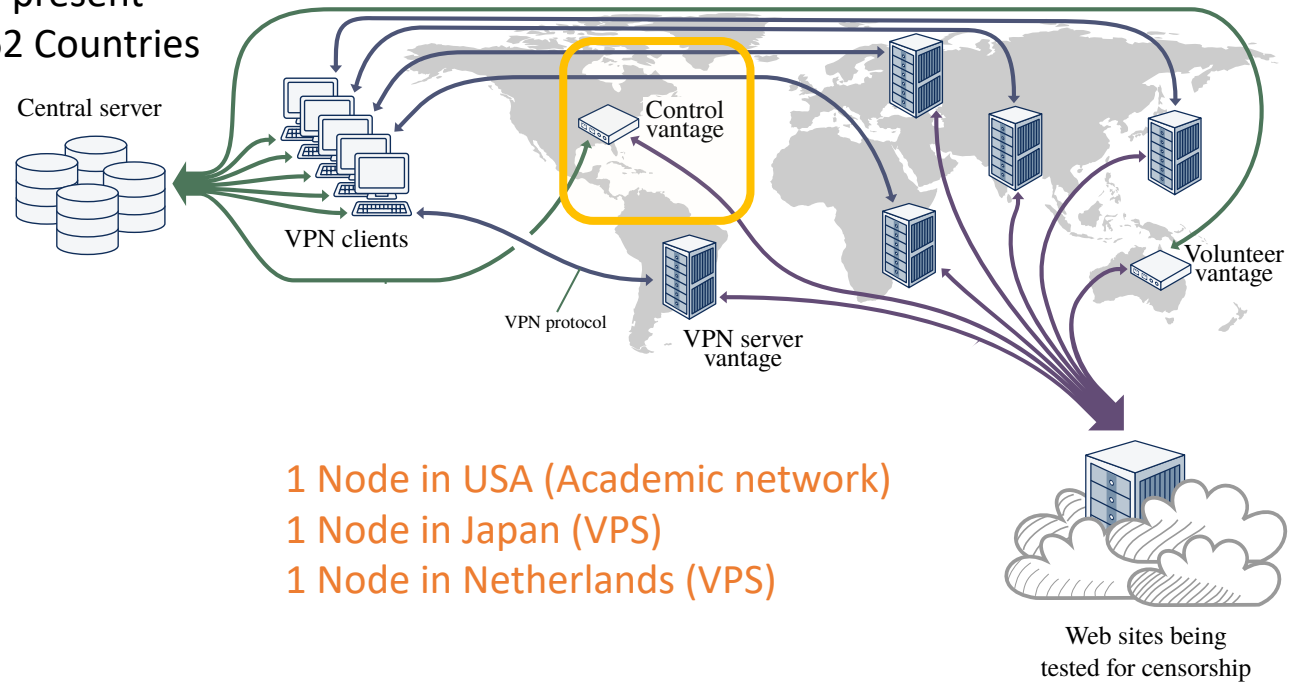
ICLab: Control Node

★ Global and Continuous

Oct. 2016 ~ present

234 ASes, 62 Countries

Vantage Points: (worldwide: 264 VPNs, 17 volunteers)



1 Node in USA (Academic network)

1 Node in Japan (VPS)

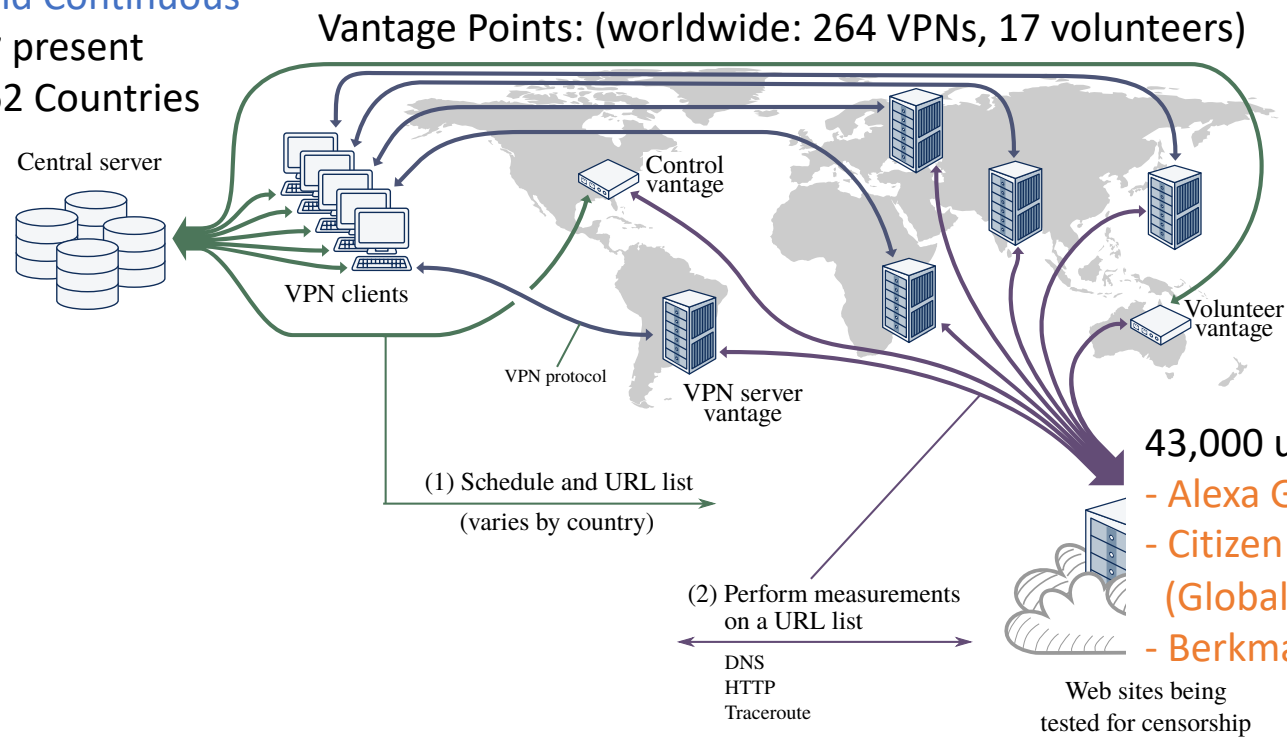
1 Node in Netherlands (VPS)

ICLab: Test Lists

★ Global and Continuous

Oct. 2016 ~ present

234 ASes, 62 Countries



43,000 unique URLs

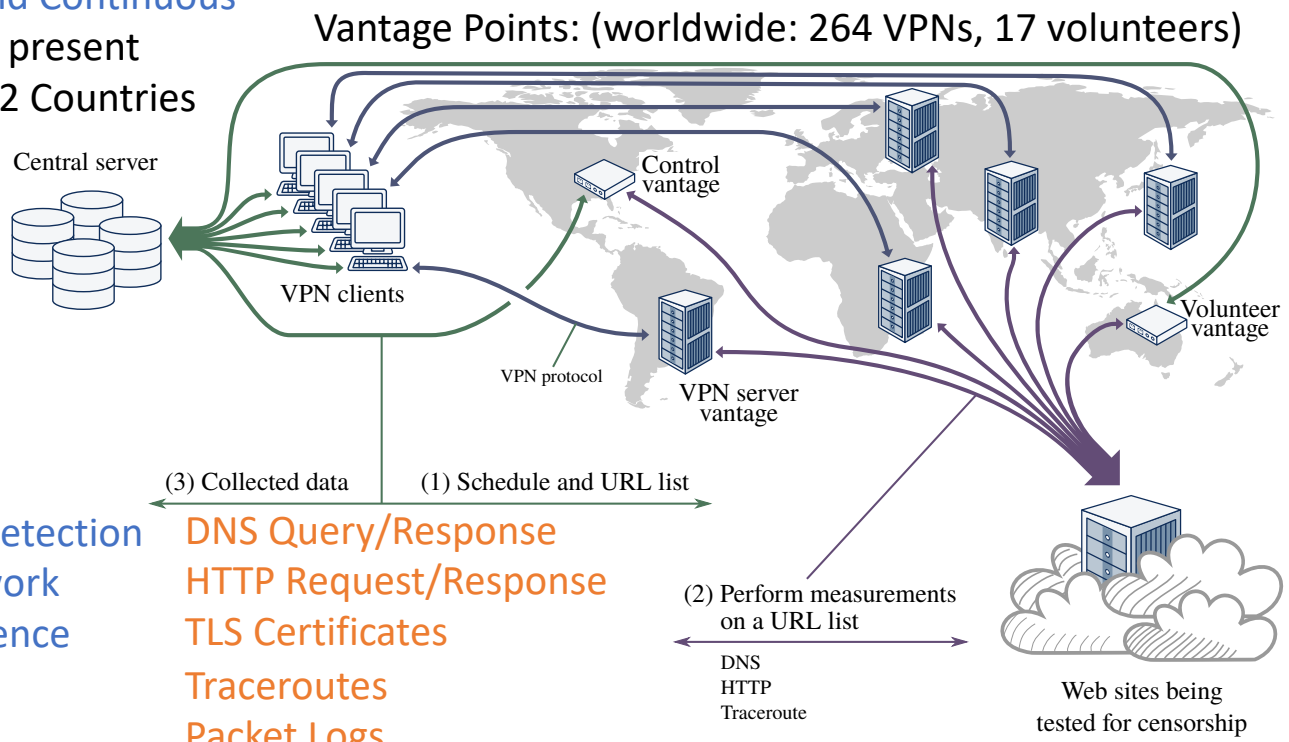
- Alexa Global
- Citizen Lab
  (Global, Local sensitive)
- Berkman Klein Center

ICLab: Data Collection

★ Global and Continuous

Oct. 2016 ~ present

234 ASes, 62 Countries



★ Reliable detection of network interference

DNS Query/Response

HTTP Request/Response

TLS Certificates

Traceroutes

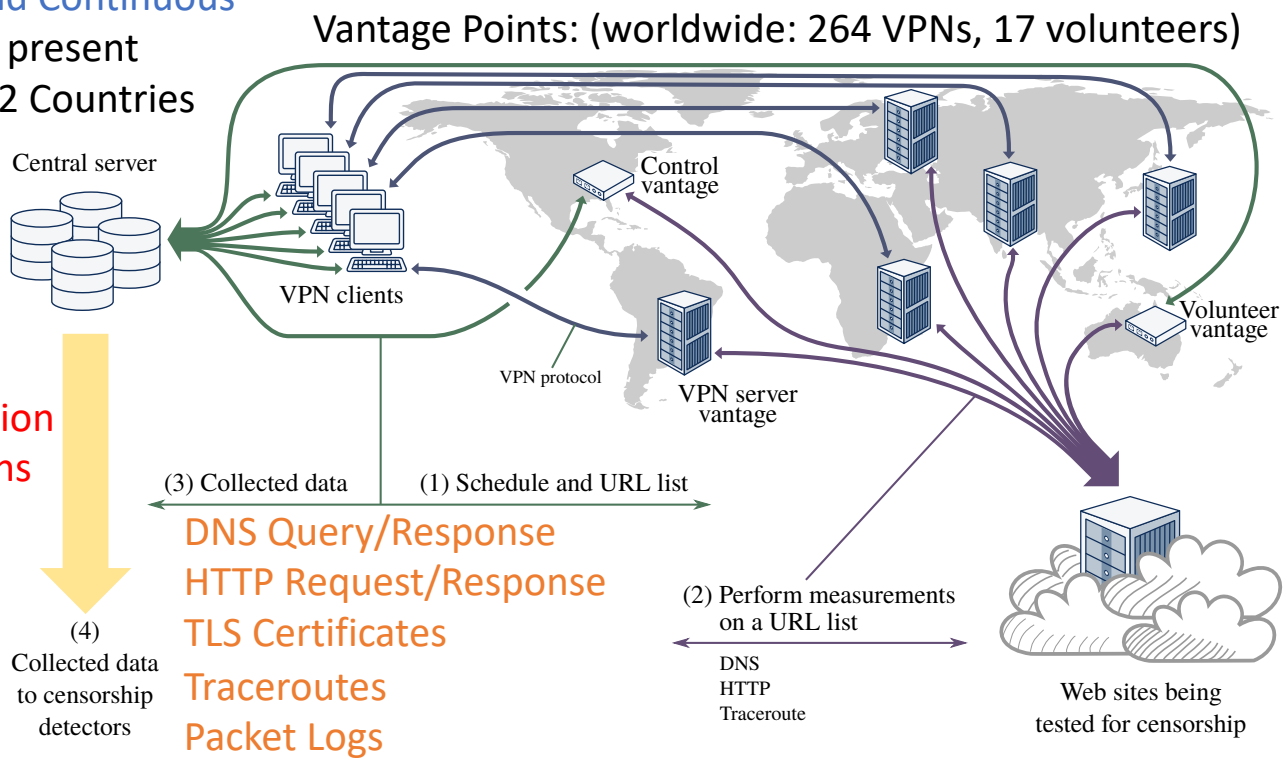
Packet Logs

ICLab: Censorship Detection

★ Global and Continuous

Oct. 2016 ~ present

234 ASes, 62 Countries



Scope of this talk

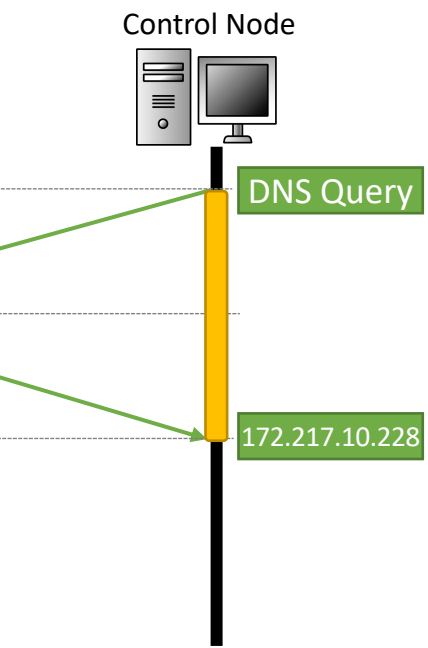
Censorship Detection

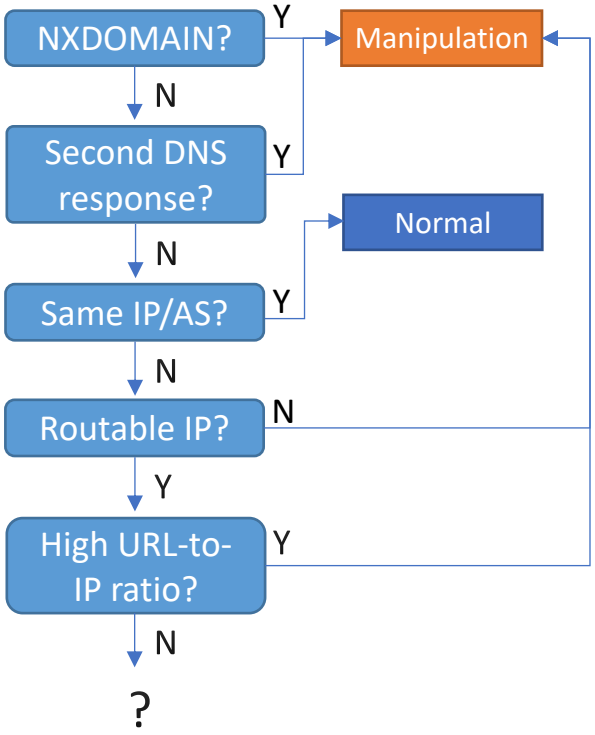
- **DNS manipulation**
- **TCP packet injection**
- **Block page : Detection & discovery**

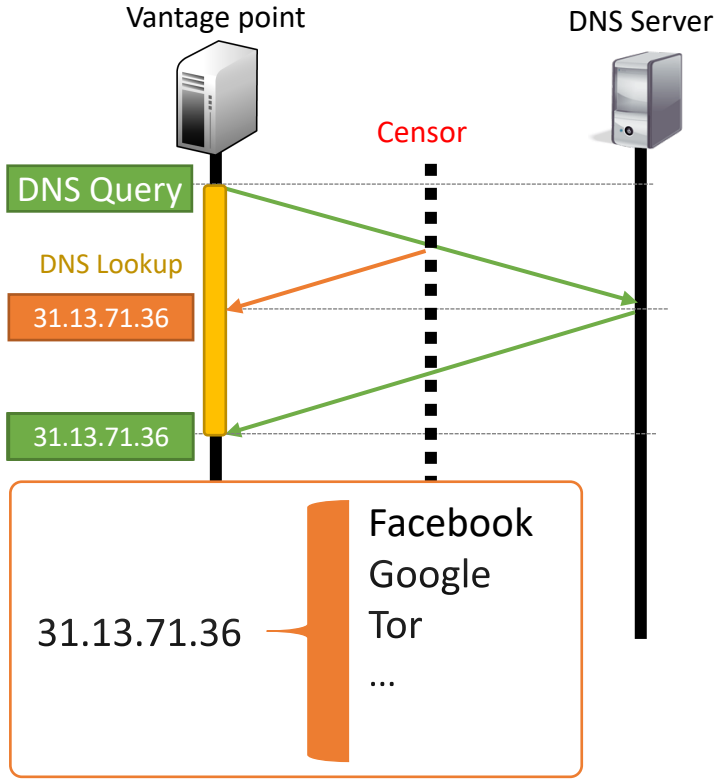
Key Findings

- Analysis by test lists
- Analysis by methods
- Longitudinal analysis
- Other network attacks

DNS Manipulation

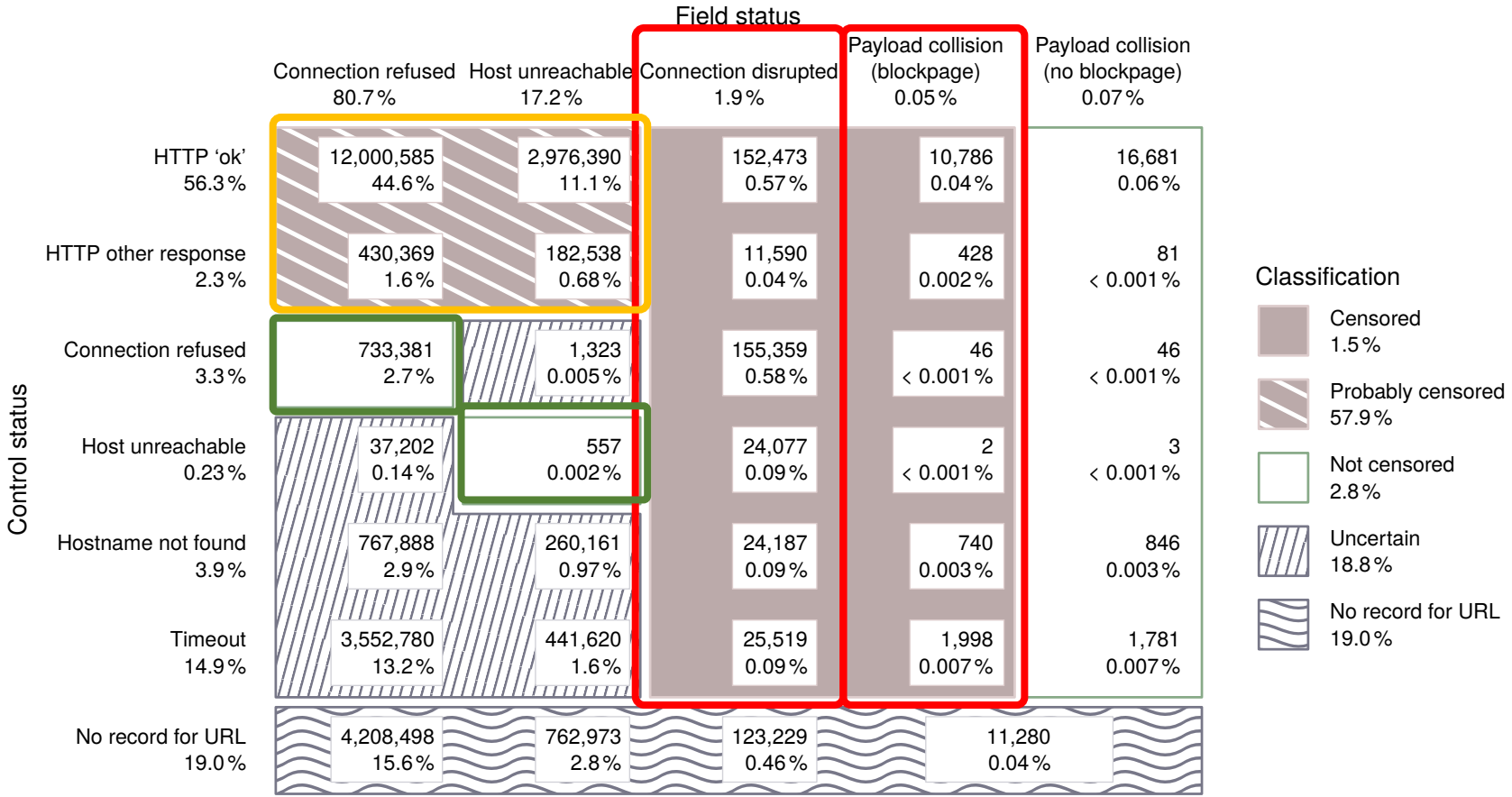






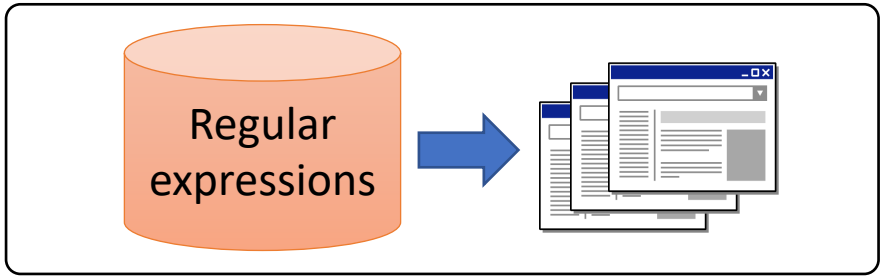
- **"Assessing the Privacy Benefits of Domain Name Encryption,"**
  Nguyen Phong Hoang, Arian Akhavan Niaki, Nikita Borisov, Phillipa Gill, and Michalis Polychronakis , In *ASIA CCS,* 2020

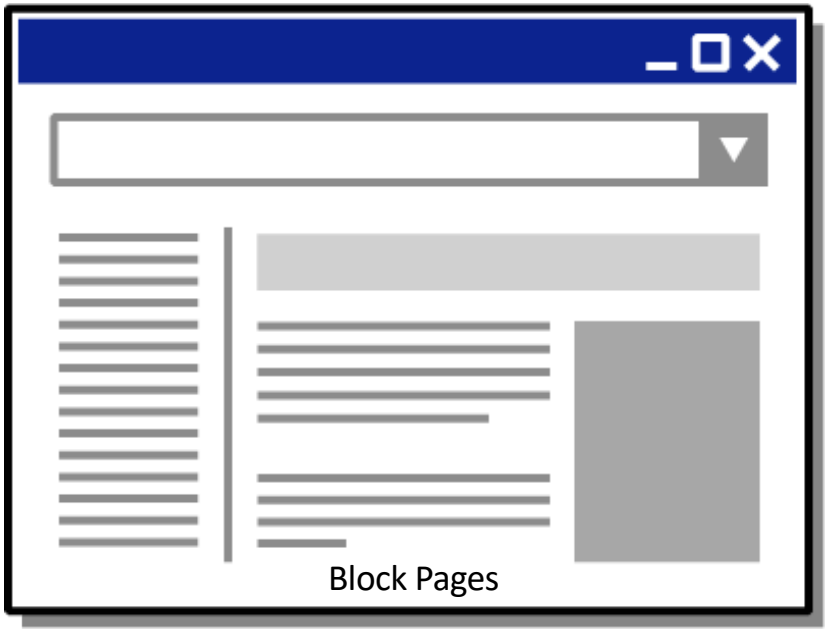
TCP Packet Injection



HTML-based Block Pages: Detection

Standard approach





HTML-based Block Pages: Discovery

HTML tag frequency vector

1 <body>, 2 <p>, 3

Match to Known block pages:
1 <body>, 2 <p>, 3

+ 15

Textual similarity

Locality-Sensitive Hashing (LSH)

"a court of competent jurisdiction"
"Hon'ble Court"

+ 33

URL-to-country ratio

Locality-Sensitive Hashing (LSH)

# of countries

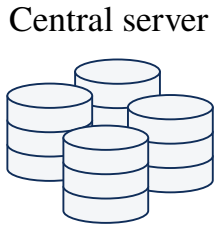
+ 286

- "Automated Detection and Fingerprinting of Censorship Block Pages," in IMC 2014
B. Jones, T.-W. Lee, N. Feamster, and P. Gill, In *ACM IMC,* Nov. 2018

Results

Censorship Detection

- **15,007 DNS Manipulation**
across 56 countries, 489 unique URLs
- **143,225 Packet Injections**
across 54 countries, 1,205 unique URLs
- **232,183** Block Pages
across 50 countries, 2,782 unique URLs



Oct. 2016 ~ present
53,906,532 measurements
234 ASes, 62 Countries
43,000 unique URLs

Results

Censorship Detection

- **15,007 DNS Manipulation**
across 56 countries, 489 unique URLs
- **143,225 Packet Injections**
across 54 countries, 1,205 unique URLs
- **232,183** Block Pages
across 50 countries, 2,782 unique URLs

Key Findings

- **Analysis by test lists**
- **Analysis by methods**
- **Longitudinal analysis**
- **Other network attacks**

Depend on what you test

Censorship by test list

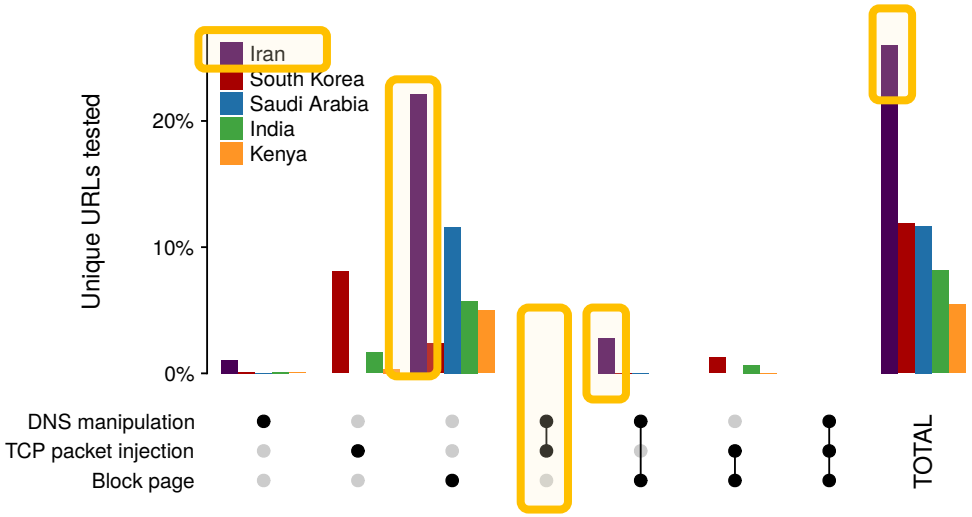
Overall			Alexa Global (ATL)			Globally Sensitive (CLBL-G)			Per-Country Sensitive (CLBL-C)		
Country	**Category**	**Pct.**	**Country**	**Category**	**Pct.**	**Country**	**Category**	**Pct.**	**Country**	**Category**	**Pct.**
Iran	NEWS	13.1%	Iran	NEWS	14.0%	Iran	PORN	11.6%	Iran	NEWS	21.0%
	PORN	9.2%		PORN	12.7%		NEWS	9.4%		BLOG	17.6%
	BLOG	7.5%		ENT	10.3%		PROX	6.8%		POL	7.2%
South Korea	PORN	15.4%	South Korea	SHOP	14.2%	Saudi Arabia	PORN	31.0%	India	ENT	19.0%
	NEWS	8.4%		PORN	13.7%		GAMB	13.5%		STRM	14.3%
	ORG	7.4%		NEWS	10.8%		PROX	12.2%		NEWS	10.8%
Saudi Arabia	PORN	29.5%	Saudi Arabia	PORN	70.0%	South Korea	PORN	15.6%	Saudi Arabia	NEWS	54.0%
	NEWS	11.3%		ILL	6.6%		ORG	10.4%		POL	7.7%
	GAMB	10.1%		GAMB	6.6%		NEWS	5.7%		RELI	7.7%
India	ENT	13.3%	Turkey	PORN	66.0%	Kenya	PORN	14.5%	Russia	BLOG	16.5%
	STRM	10.8%		ILL	4.0%		GAMB	10.8%		NEWS	14.4%
	NEWS	10.4%		FILE	4.0%		PROX	9.0%		GAMB	12.4%
Kenya	PORN	15.5%	India	ILL	35.5%	Turkey	PORN	47.0%	Turkey	NEWS	29.4%
	GAMB	10.1%		IT	8.8%		GAMB	22.6%		PORN	13.7%
	PROX	8.3%		STRM	6.6%		ILL	3.2%		GAMB	9.8%

Different Methods for Different Content

Censorship by method

Technique	Country	Categories	Pct.
Block page	Iran	NEWS, PORN, BLOG	24.95%
	Saudi Arabia	PORN, NEWS, GAMB	11.1%
	India	ENT, STRM, NEWS	6.4%
	Kenya	PORN, GAMB, PROX	4.8%
	Turkey	PORN, GAMB, NEWS	4.6%
DNS manipulation	Iran	BLOG, PORN, PROX	5.5%
	Uganda	PORN, ADUL, LING	1.7%
	Turkey	ILL, GAMB, STRM	0.3%
	Bulgaria	ILL, ARM, DOM	0.2%
	Netherlands	ILL, IM, DOM	0.2%
TCP packet injection	South Korea	PORN, ORG, NEWS	9.3%
	India	NEWS, ILL, IT	2.3%
	Netherlands	NEWS, SEAR, GAME	0.9%
	Japan	NEWS, GAME, SEAR	0.9%
	Australia	SEAR, NEWS, ILL	0.8%

ILL: illegal or unethical
STREAM: streaming media and download



Combinations of methods

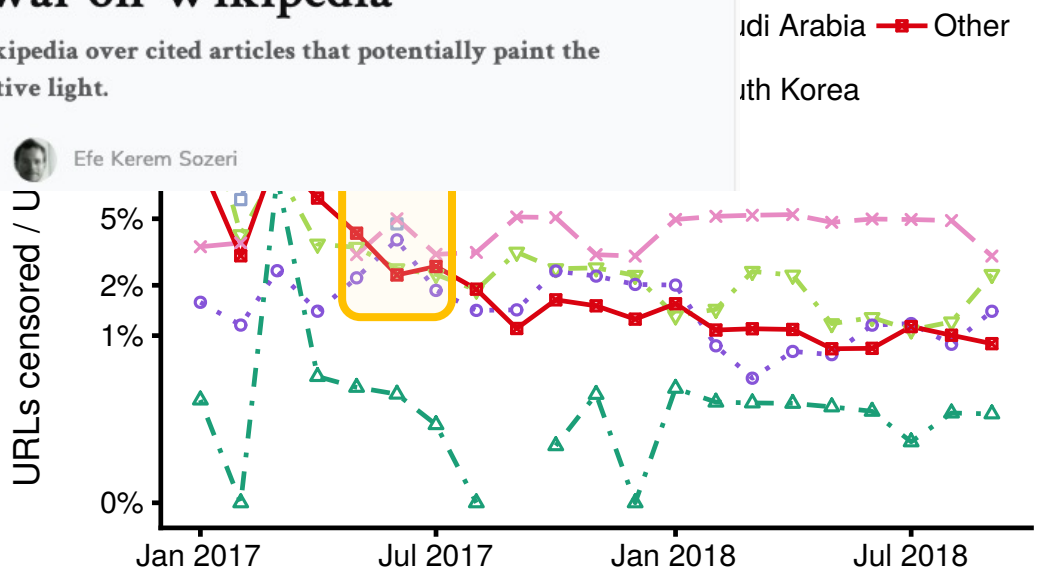
Longitudinal Analysis

daily dot

Inside Turkey's war on Wikipedia

Turkey has censored all of Wikipedia over cited articles that potentially paint the county's government in a negative light.

May 9, 2017, 4:27 pm* Tech Efe Kerem Sozeri

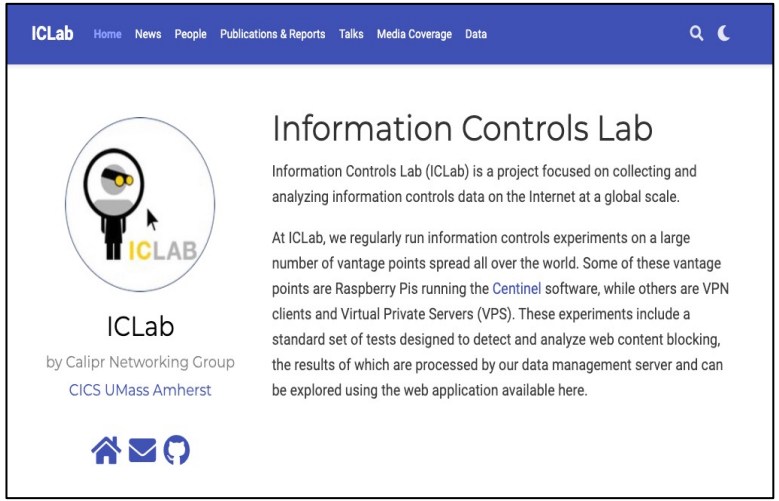


Observed censorship rates

Detected Other Network Attacks

- Geo-blocking
  - HTTP 451 “Unavailable for Legal Reasons”
  - 23 Unique websites across 21 countries
- User tracking injection
  - Fingerprinting the client prior to loading the page
- Cryptocurrency mining injection
  - A botnet infecting MikroTick routers (exploiting CVE-2018-14847)
  - First observed on July 21st, 2018

Summary



https://iclab.org/

Key Findings

- Each test list shows different censorship policy
- Different methods used to block different web categories
- Censorship has changed over time
- Other network interferences are detected

Thank you!

http://calipr.cs.umass.edu