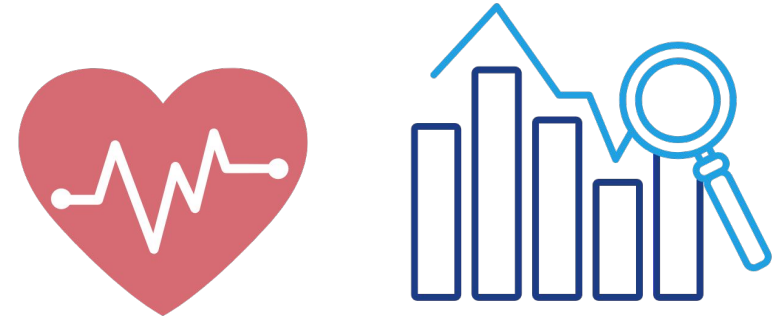
Lost in Transmission: Investigating Filtering of COVID-19 Websites

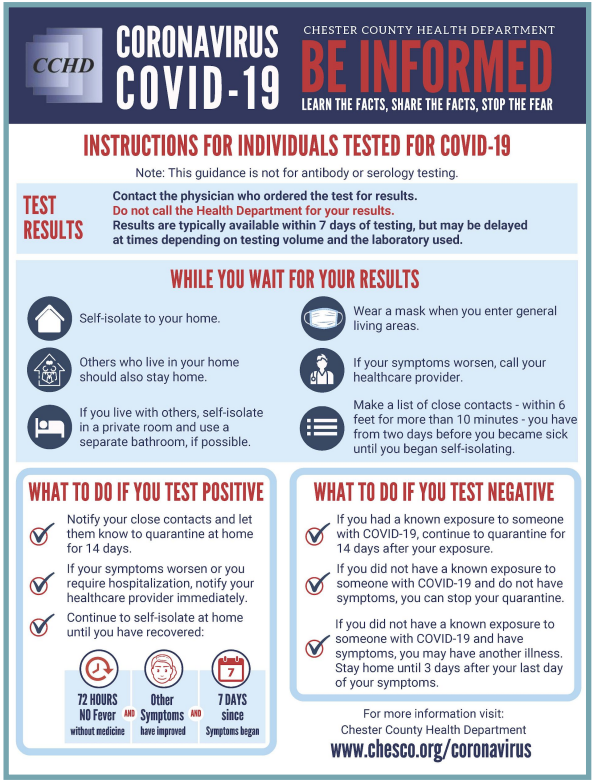
Anjali Vyas, Ram Sundara Raman, Nick Ceccio, Philipp Lutscher*, Roya Ensafi

University of Michigan, *University of Oslo

Financial Cryptography and Data Security (2021)



The Internet is a crucial source of COVID-19 information



CCHD

CORONAVIRUS COVID-19

CHESTER COUNTY HEALTH DEPARTMENT

BE INFORMED

LEARN THE FACTS, SHARE THE FACTS, STOP THE FEAR

INSTRUCTIONS FOR INDIVIDUALS TESTED FOR COVID-19

Note: This guidance is not for antibody or serology testing.

TEST RESULTS

Contact the physician who ordered the test for results.
Do not call the Health Department for your results.
Results are typically available within 7 days of testing, but may be delayed at times depending on testing volume and the laboratory used.

WHILE YOU WAIT FOR YOUR RESULTS

- Self-isolate to your home.
- Others who live in your home should also stay home.
- If you live with others, self-isolate in a separate room and use a separate bathroom, if possible.
- Wear a mask when you enter general living areas.
- If your symptoms worsen, call your healthcare provider.
- Make a list of close contacts - within 6 feet for more than 10 minutes - you have from two days before you became sick until you began self-isolating.

WHAT TO DO IF YOU TEST POSITIVE

- Notify your close contacts and let them know to quarantine at home for 14 days.
- If your symptoms worsen or you require hospitalization, notify your healthcare provider immediately.
- Continue to self-isolate at home until you have recovered:

72 HOURS NO Fever without medicine AND Other Symptoms have improved AND 7 DAYS since Symptoms began

WHAT TO DO IF YOU TEST NEGATIVE

- If you had a known exposure to someone with COVID-19, continue to quarantine for 14 days after your exposure.
- If you did not have a known exposure to someone with COVID-19 and do not have symptoms, you can stop your quarantine.
- If you did not have a known exposure to someone with COVID-19 and have symptoms, you may have another illness. Stay home until 3 days after your last day of your symptoms.

For more information visit:
Chester County Health Department
www.chesco.org/coronavirus

Source:
chesco.org/coronavirus

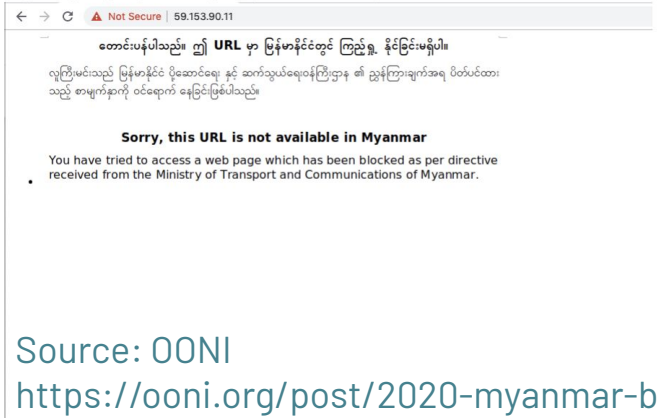
Is potentially useful information about COVID-19 being **censored** on the Internet?

Internet censorship → access to certain websites is blocked for users

Previous COVID-19 Censorship Studies are limited

Myanmar blocks “fake news” websites amid COVID-19 pandemic

Phyu Phyu Kyaw, Maria Xynou, Arturo Filastò, 2020-05-06

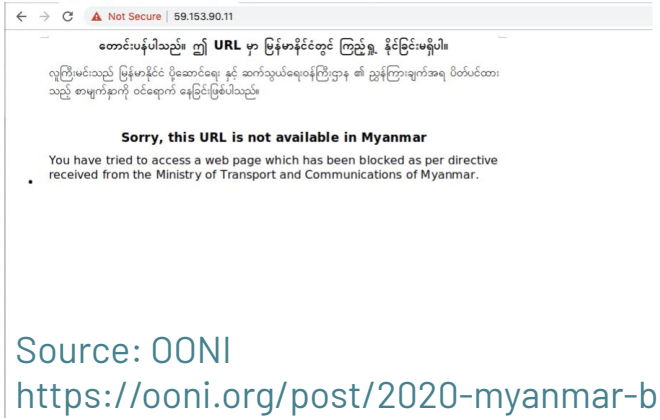


Source: OONI
https://ooni.org/post/2020-myanmar-blocks-websites-amid-covid19/

Previous COVID-19 Censorship Studies are limited

Myanmar blocks “fake news” websites amid COVID-19 pandemic

Phyu Phyu Kyaw, Maria Xynou, Arturo Filastò, 2020-05-06



Source: OONI https://ooni.org/post/2020-myanmar-blocks-websites-amid-covid19/

- **Limitations:**
  - Focused on measuring COVID-19 Internet censorship of specific websites
  - In some countries

Our study

Large scale study measuring the global Internet censorship of a potentially useful COVID-19 websites

- Large number of domains
- Many countries

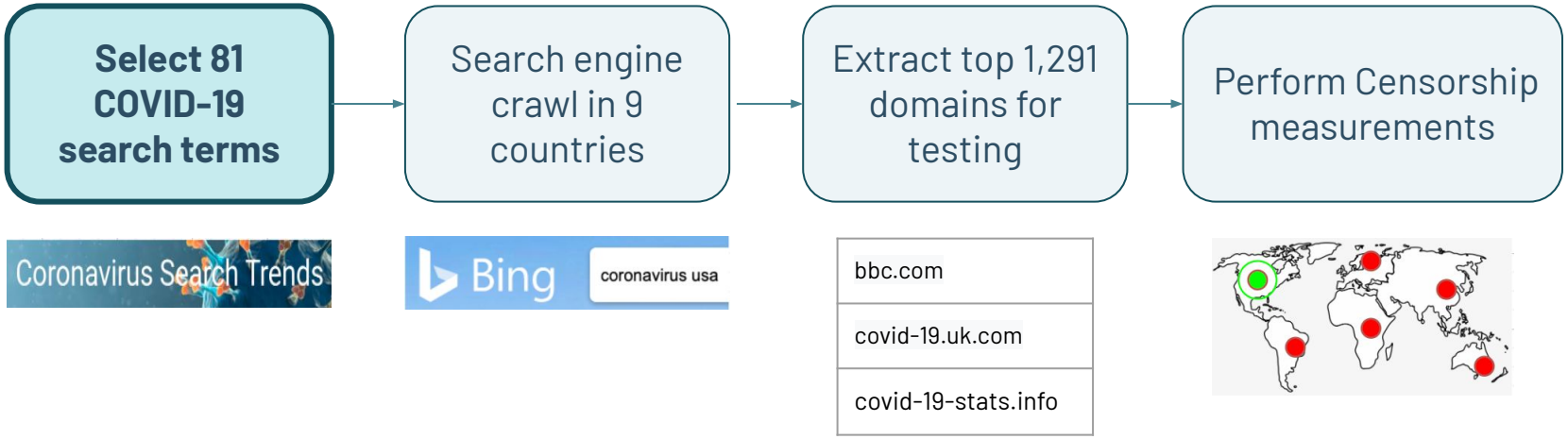
Our study

Large scale study measuring the global Internet censorship of a potentially useful COVID-19 websites

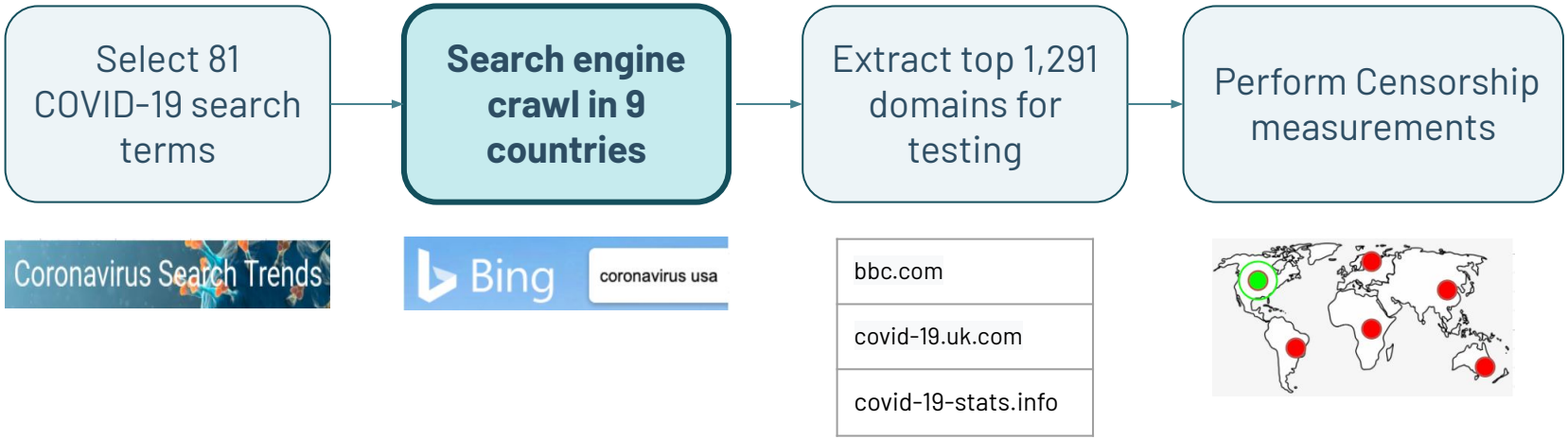
- Large number of domains
- Many countries

Our questions

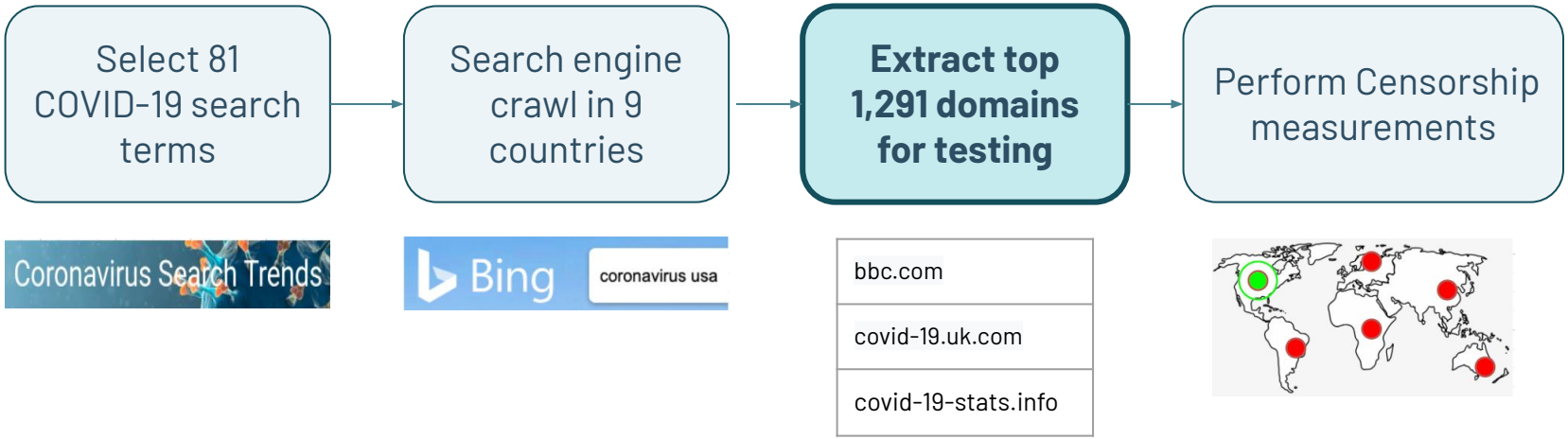
- Where are potentially useful COVID-19 websites blocked?
- How common is the blocking?
- What categories do these blocked websites belong to?



Measurement Pipeline

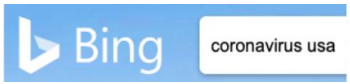


Measurement Pipeline



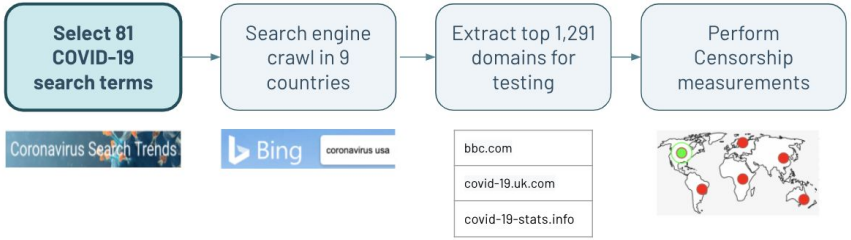
Measurement Pipeline

Select 81 COVID-19 search terms → Search engine crawl in 9 countries → Extract top 1,291 domains for testing → **Perform Censorship measurements**

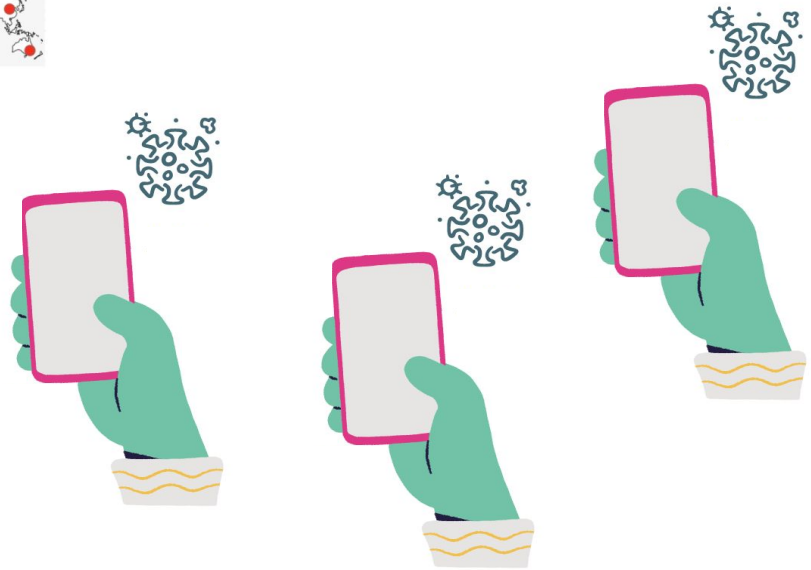


bbc.com
covid-19.uk.com
covid-19-stats.info

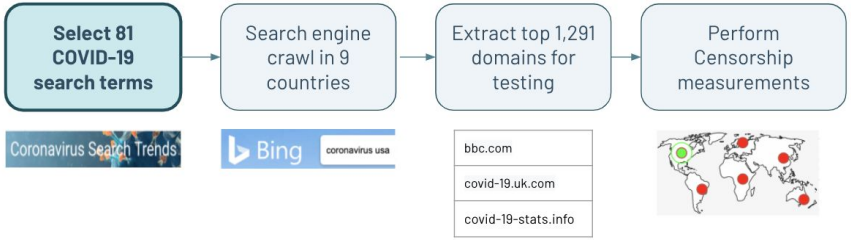
Measurement Pipeline



Got popular factual search terms related to pandemic from Google Trends



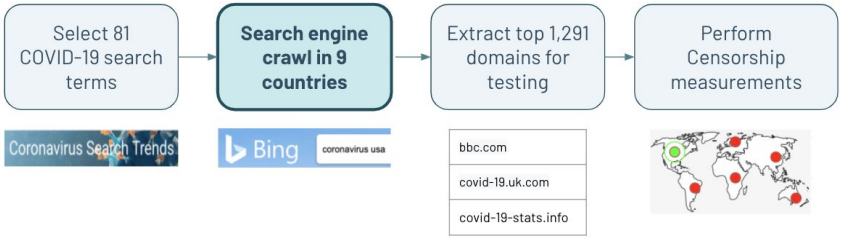
Selection of COVID-19 Search Terms



Got popular factual search terms related to pandemic from Google Trends

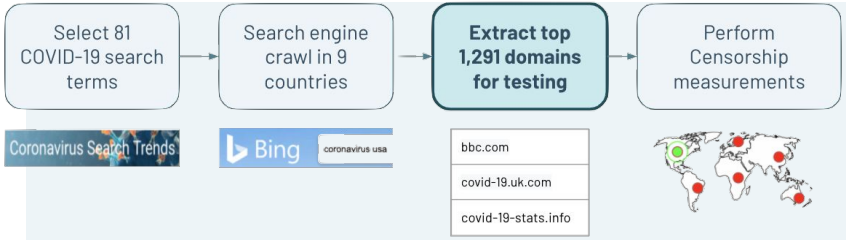
Prefix	Suffix
coronavirus, covid, covid-19	usa, ireland, uk, britain, india, canada, singapore, korea, japan, australia, germany, france, update, news, worldometer, deaths, victims, map, live, infections, stats, toll, death toll, vaccine, dead, <empty>

Selection of COVID-19 Search Terms

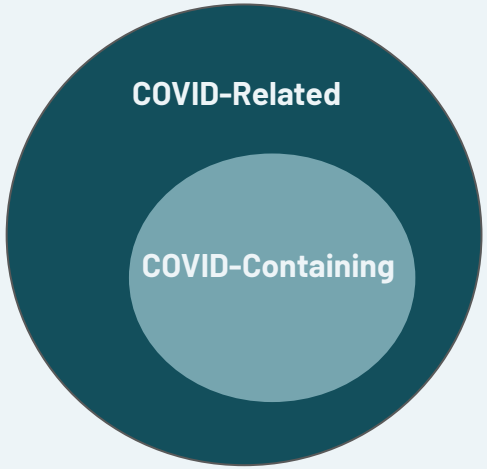


Performed search engine crawls in 9 geodistributed vantage points to find top URLs that show up

Search Engine Crawls



Made lists of domains to test



1,291 domains

COVID-related test-list

e.g. nytimes.comc

152 domains

COVID-containing test-list

e.g. covidlive.co.uk

2,128 domains

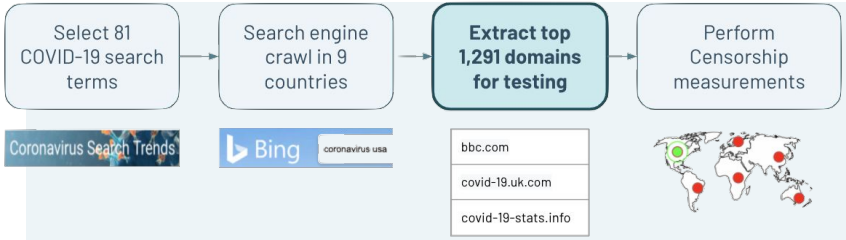
Censorship-Measurement control-list

e.g. www.unhcr.org

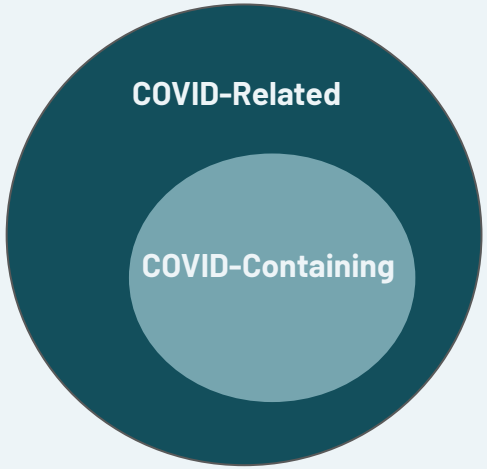
THECITIZENLAB

Alexa An amazon.com company

Top Sites in United States



Made lists of domains to test



1,291 domains

COVID-related test-list

e.g. nytimes.com

152 domains

COVID-containing test-list

e.g. covidlive.co.uk

2,128 domains

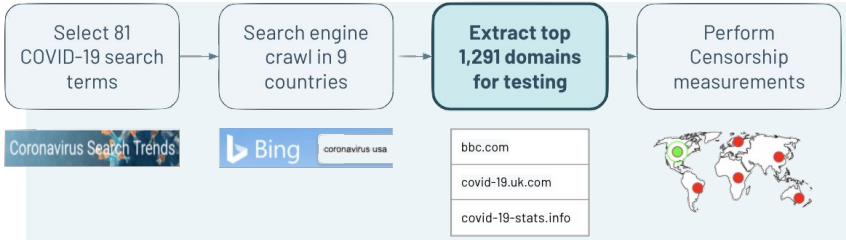
Censorship-Measurement control-list

e.g. www.unhcr.org

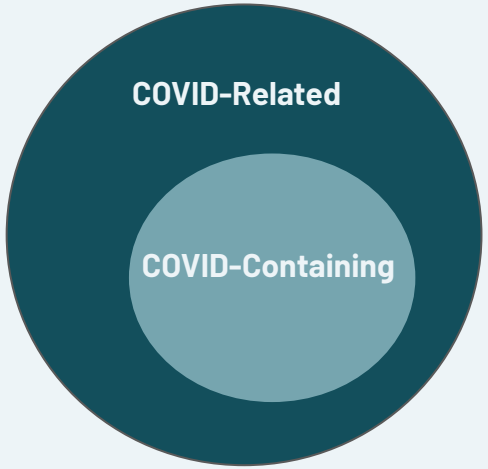
THECITIZENLAB

Alexa An amazon.com company

Top Sites in United States



Made lists of domains to test



1,291 domains

COVID-related test-list

e.g. nytimes.com

152 domains

COVID-containing test-list

e.g. covidlive.co.uk

2,128 domains

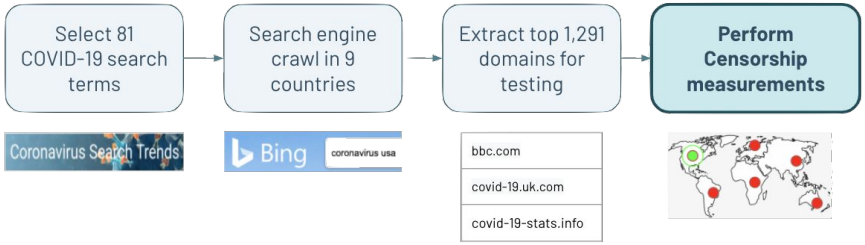
Censorship-Measurement control-list

e.g. www.unhcr.org

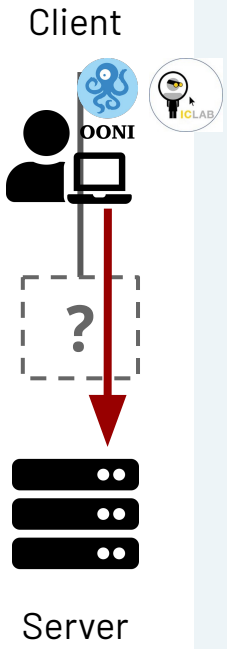
THECITIZENLAB

Alexa An amazon.com company

Top Sites in United States



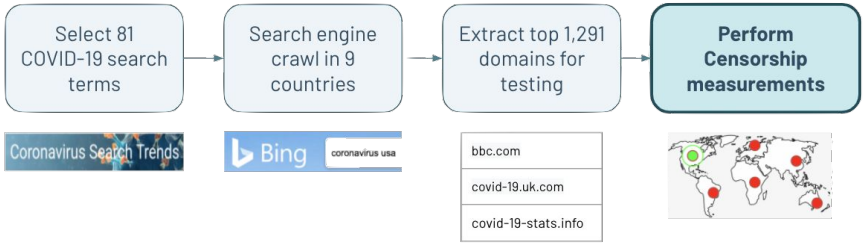
- Ask **volunteers** on the ground, or deploy software or hardware in censored region (e.g. OONI probe[1])
- Use **VPNs**, or **research networks** (e.g. ICLab[2])



[1] Open Observatory of Network Interference, https://ooni.org
[2] A. Akhavan Niaki, S. Cho, Z. Weinberg, N. P. Hoang, A. Razaghpanah, N. Christin, and P. Gill. ICLab: A Global, Longitudinal Internet Censorship Measurement Platform. In IEEE Symposium on Security and Privacy (SP), 2020.

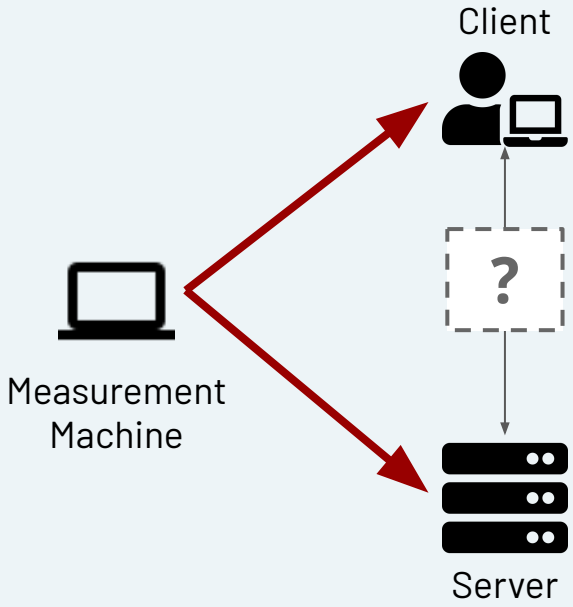
Direct Censorship Measurement Platforms

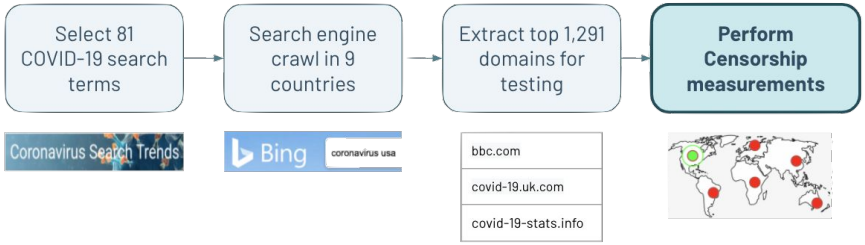
- **Limitations**
  - Scale
  - Coverage
  - Continuity
  - Synchronization
  - Ethics



Detect whether pairs of hosts around the world can talk to each other without controlling either endpoint.

Remote Censorship Measurements

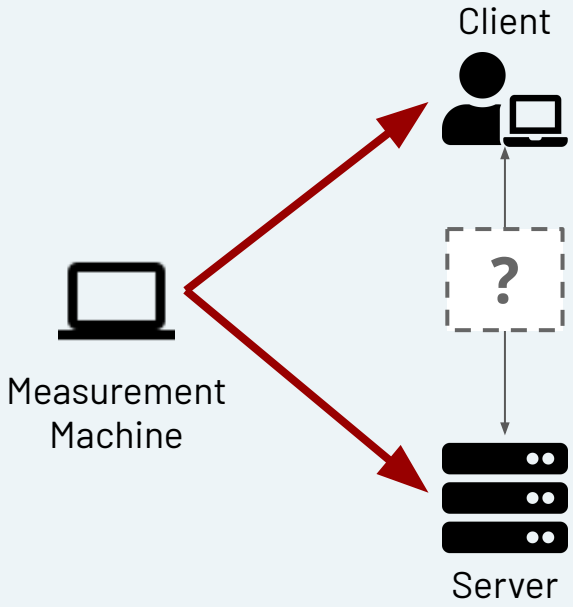


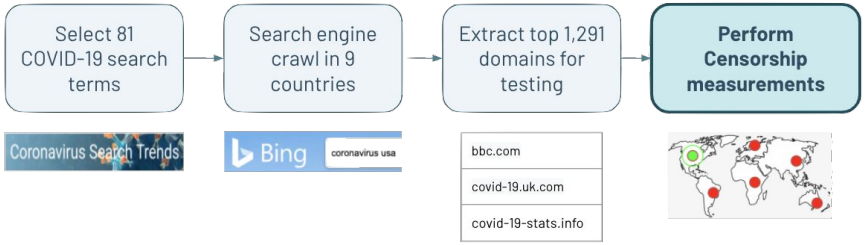


Detect whether pairs of hosts around the world can talk to each other without controlling either endpoint.

Censored Planet

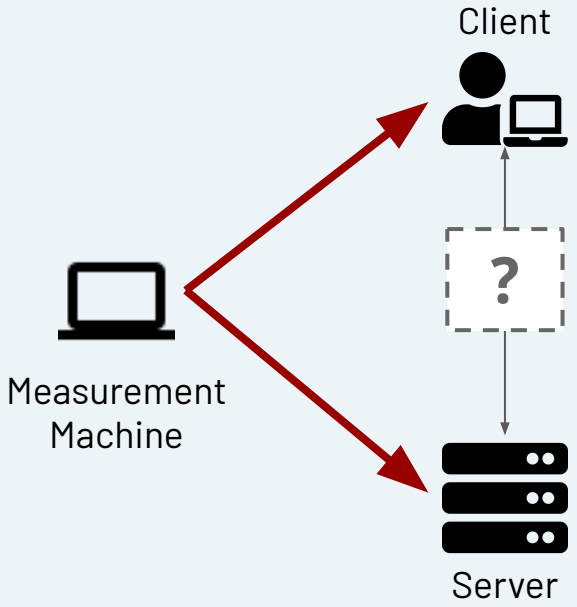
Remote Censorship Measurements

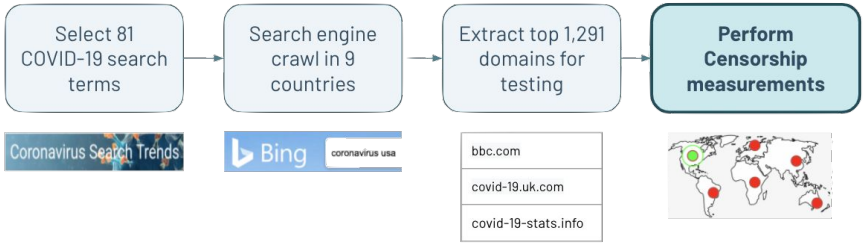




Detect whether pairs of hosts around the world can talk to each other without controlling either endpoint.

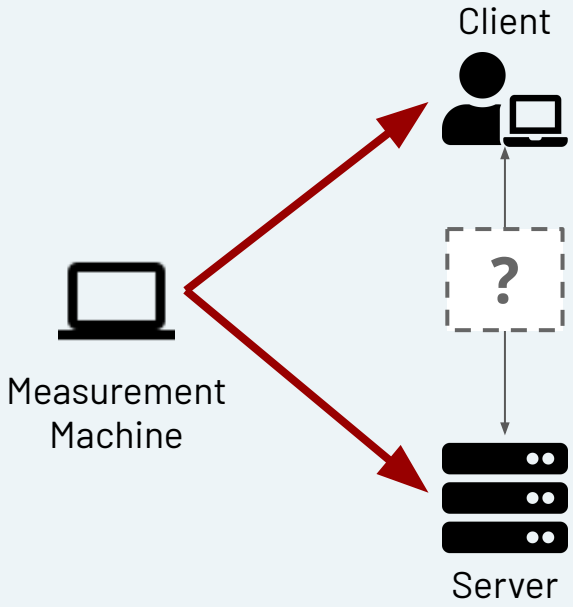
Remote Censorship Measurements



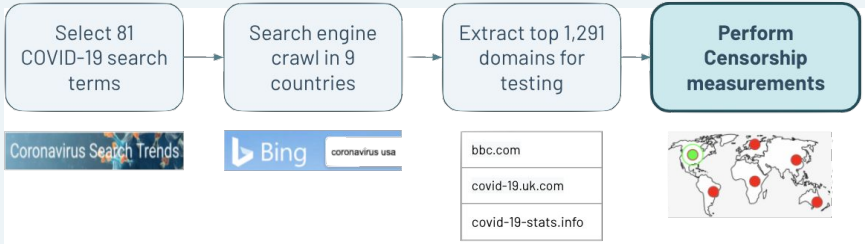


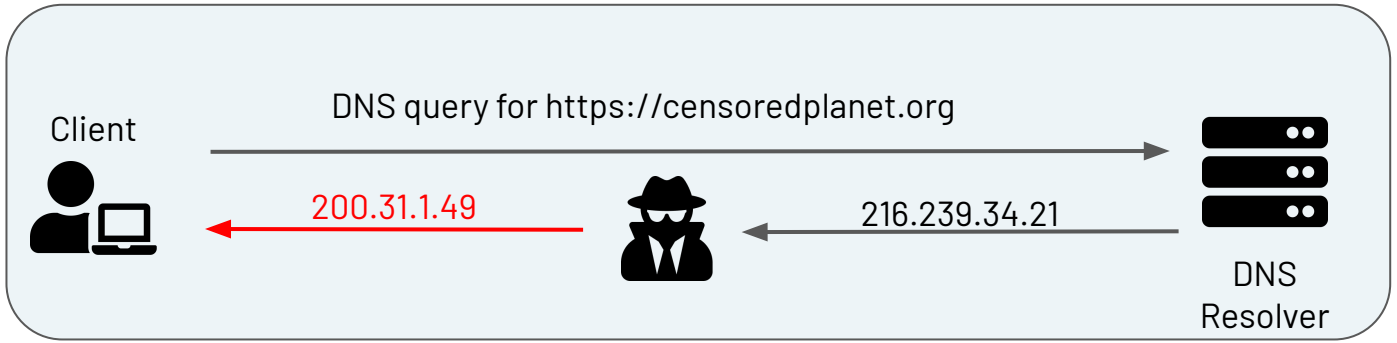
Detect whether pairs of hosts around the world can talk to each other without controlling either endpoint.

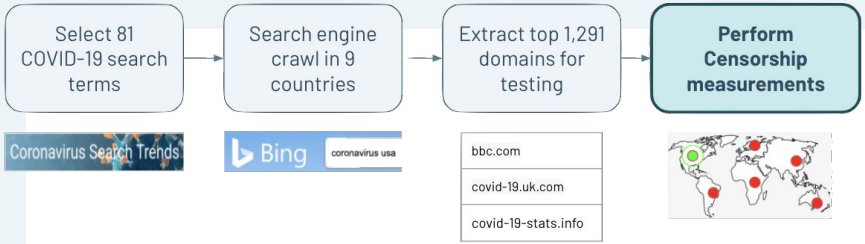
Remote Censorship Measurements



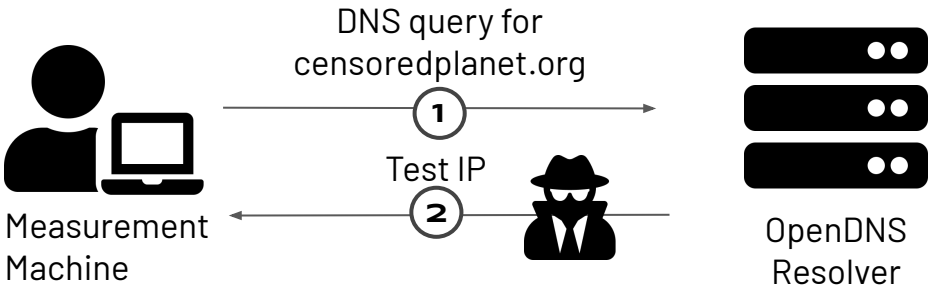
DNS Manipulation

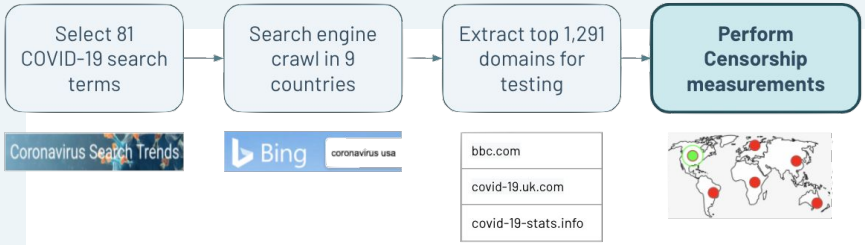




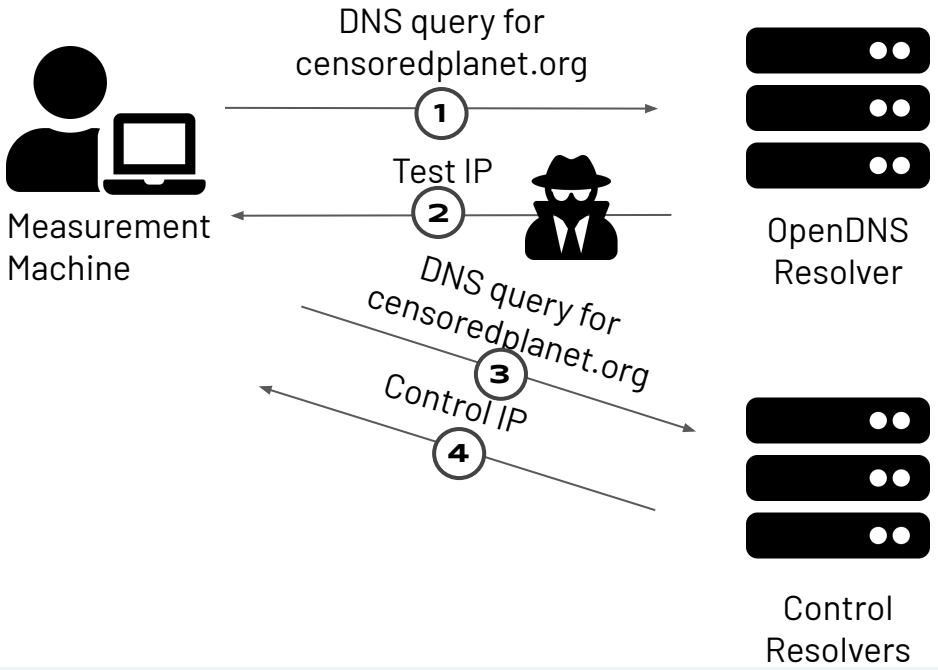

Select 81 COVID-19 search terms
Search engine crawl in 9 countries
Extract top 1,291 domains for testing
Perform Censorship measurements
Coronavirus Search Trends
Bing
coronavirus usa
bbc.com
covid-19.uk.com
covid-19-stats.info

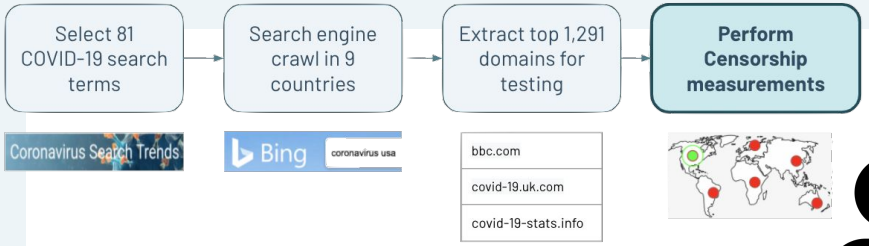
Satellite


DNS query for censoredplanet.org
1
Test IP
2
Measurement Machine
OpenDNS Resolver

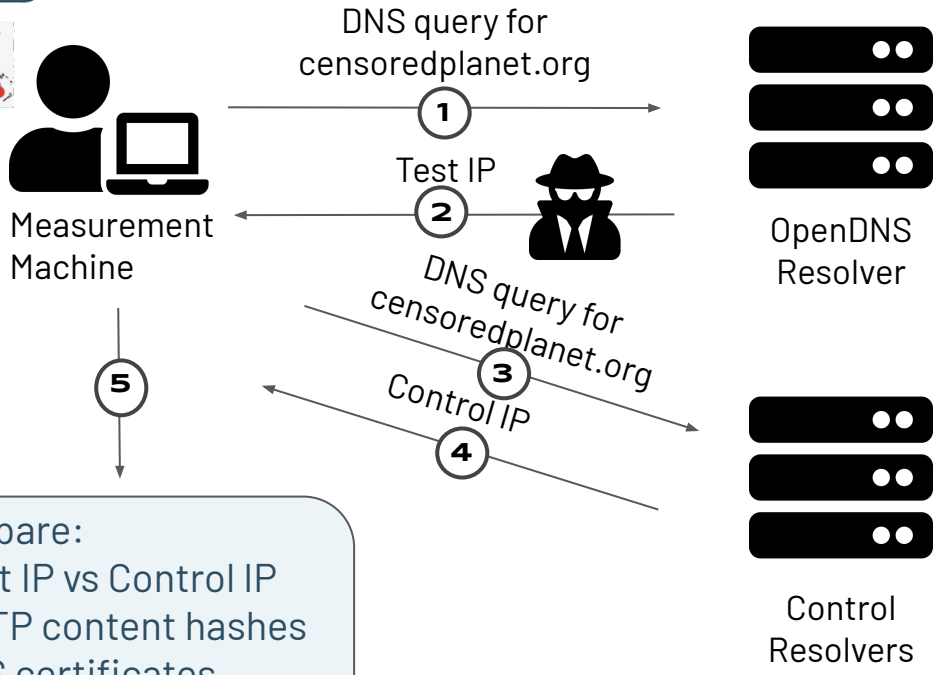

Select 81 COVID-19 search terms
Search engine crawl in 9 countries
Extract top 1,291 domains for testing
Perform Censorship measurements
Coronavirus Search Trends
Bing
coronavirus usa
bbc.com
covid-19.uk.com
covid-19-stats.info

Satellite


DNS query for censoredplanet.org
1
Test IP
2
Measurement Machine
OpenDNS Resolver
DNS query for censoredplanet.org
3
Control IP
4
Control Resolvers

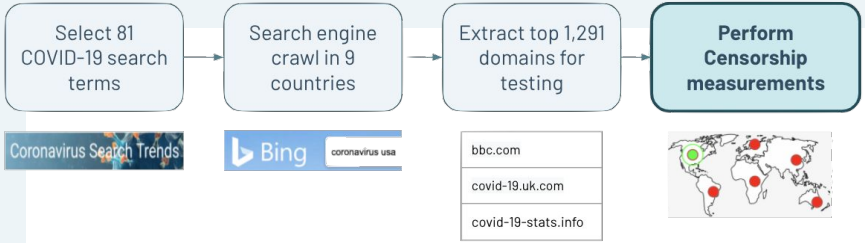


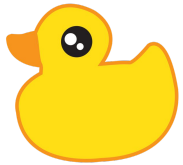
Satellite



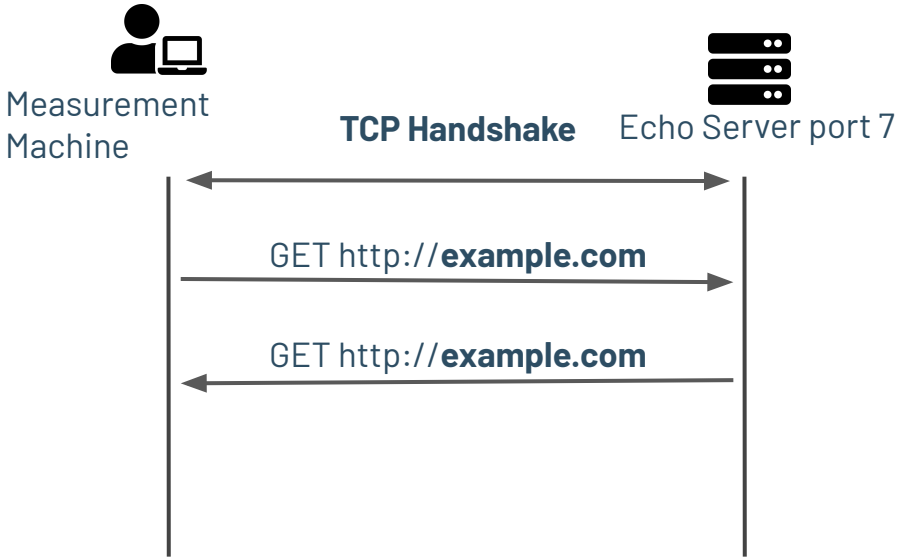
Compare:
- Test IP vs Control IP
- HTTP content hashes
- TLS certificates
- ASN and AS Name

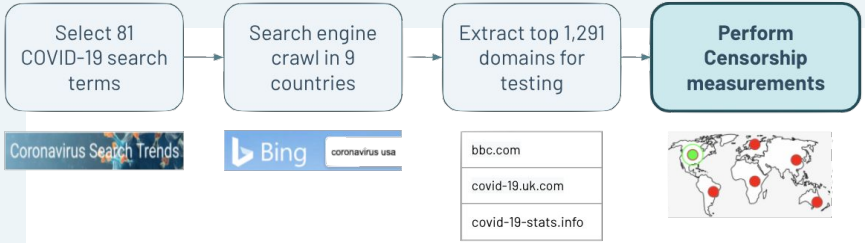
etc.

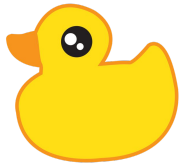

Select 81 COVID-19 search terms
Search engine crawl in 9 countries
Extract top 1,291 domains for testing
Perform Censorship measurements
Coronavirus Search Trends
Bing
coronavirus usa
bbc.com
covid-19.uk.com
covid-19-stats.info



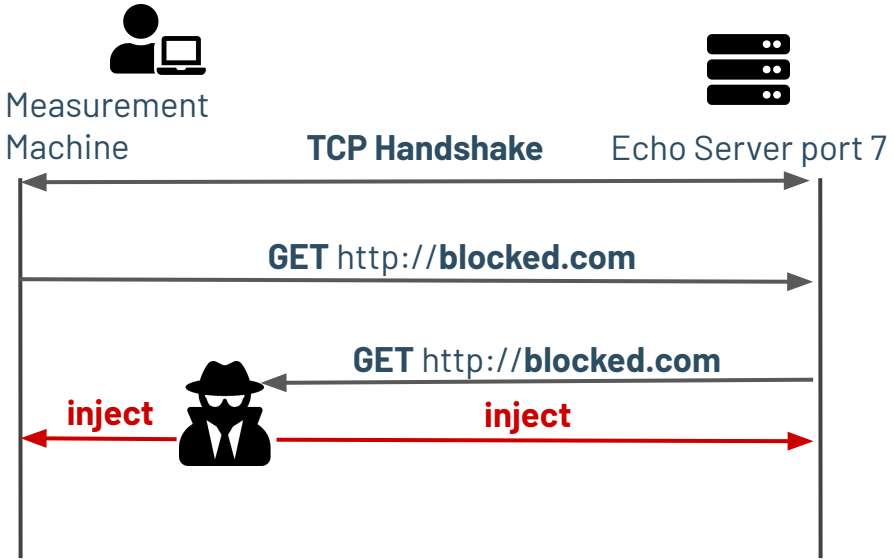
Quack


Measurement Machine
TCP Handshake
Echo Server port 7
GET http://example.com
GET http://example.com


Select 81 COVID-19 search terms
Search engine crawl in 9 countries
Extract top 1,291 domains for testing
Perform Censorship measurements
Coronavirus Search Trends
Bing
coronavirus usa
bbc.com
covid-19.uk.com
covid-19-stats.info



Quack


Measurement Machine
TCP Handshake
Echo Server port 7
GET http://blocked.com
GET http://blocked.com
inject
inject

153 million

Over June 12 to June 26, 2020

186 countries

5,081 ASes

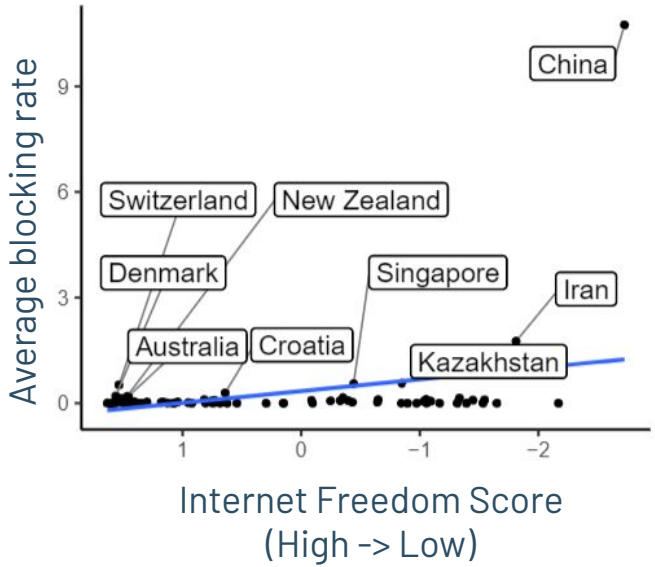
29,113 Satellite Vantage Points
20,989 Quack Vantage Points

Findings

Average blocking rate	Satellite	Quack
COVID-related	0.20%	0.31%
Censorship-Measurement	0.70%	1.02%

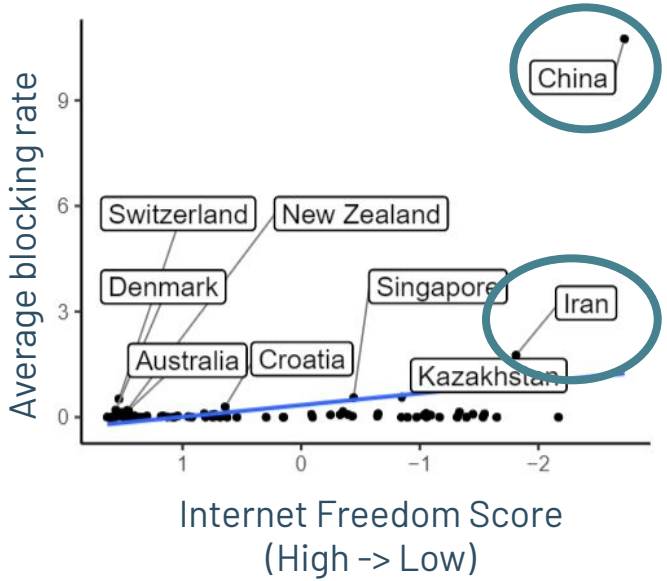
Average blocking rate of websites

Findings



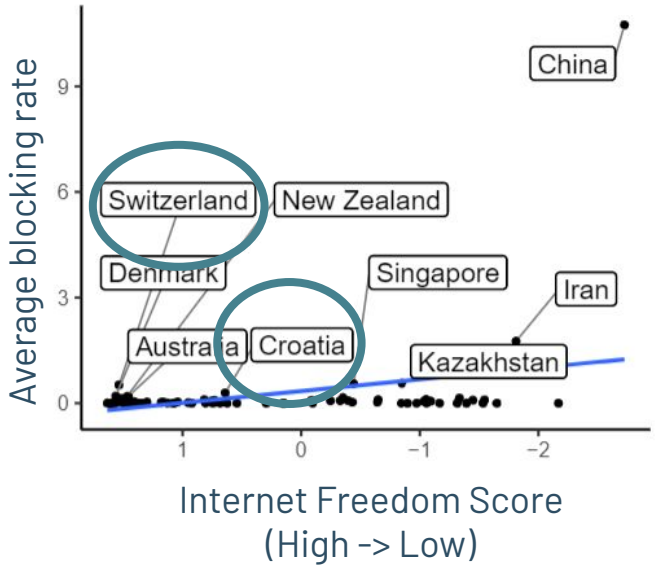
DNS filtering of COVID-related websites

Findings



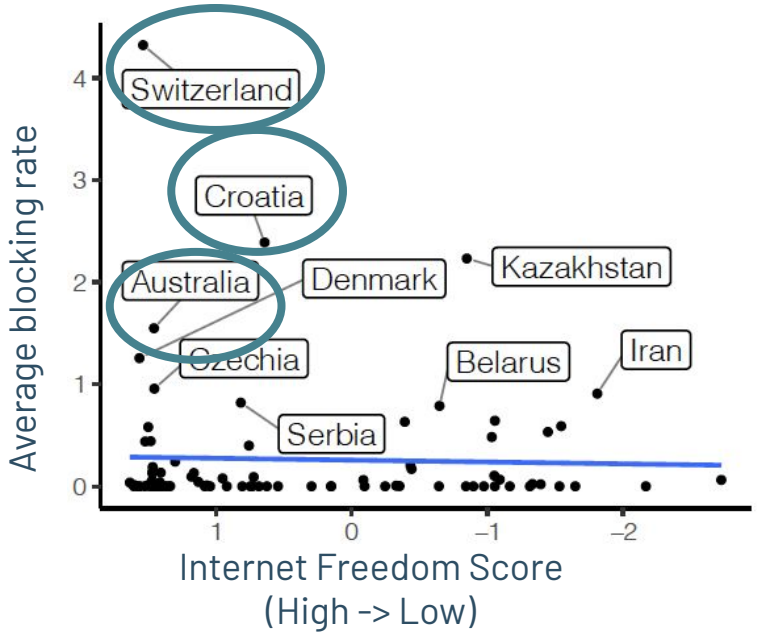
China & Iran show high DNS filtering of COVID-related websites

Findings



Some free countries also show higher than expected DNS filtering of COVID-related websites

Findings



Some free countries show higher than expected DNS blocking of COVID-containing websites

Findings

922 Satellite VPs across **34 ASes**

4.32% of COVID-containing websites blocked on average

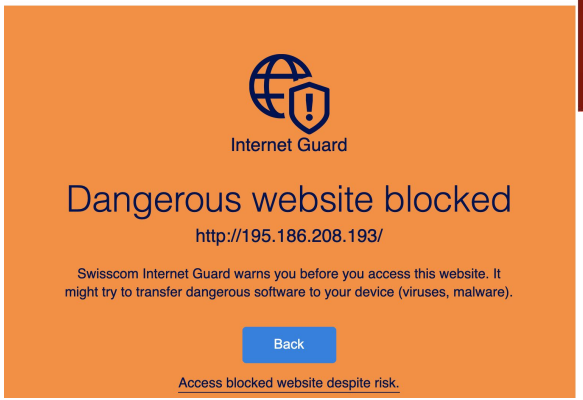
65.7% of VPs observed blocking of at least one COVID-containing website

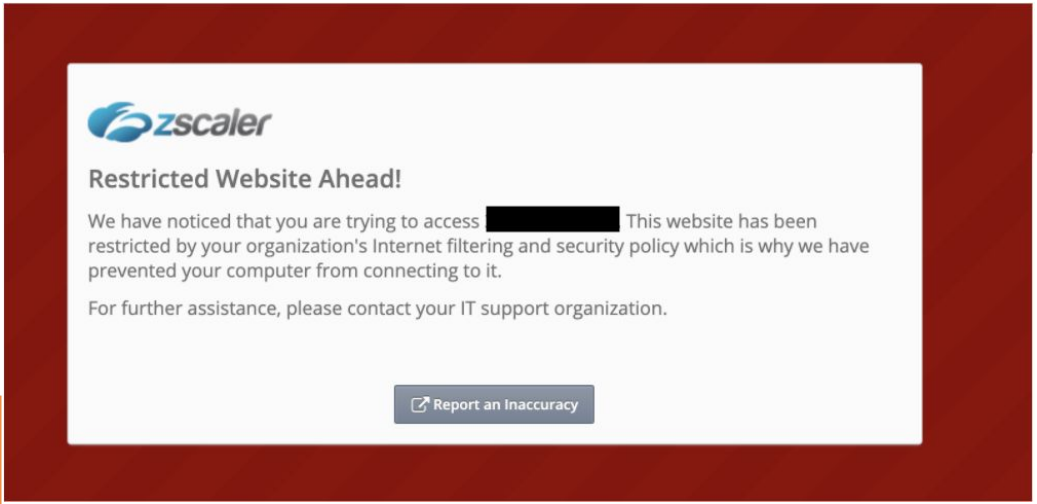
DNS Filtering in Switzerland

Findings

99% of blocked VP requests were resolved to IP addresses with one of these blockpages

swisscom





Findings

Top 10 filtered domains in Switzerland (DNS)

With their FortiGuard given categories

Domain	Category	% of VPs blocking
www.covid-19.uk.com	Phishing	66.44%
coronavirus-realtime.com	Malicious	66.40%
covid19graph.work	Phishing	66.36%
www.covid19ireland.com	Phishing	66.22%
www.covid19maps.info	Phishing	65.83%
covid-19-stats.info	Phishing	65.74%
coronavirus.zone	Malicious	65.41%
coronavirus-map.com	Phishing	65.36%
coronastats.net	Malicious	63.60%
coronavirusfrance.org	Phishing	1.58%

Findings

Are these websites actually dangerous?

Found categorization of websites filtered by FortiGuard using 3 other classifiers/URL filters available online

3 authors also manually visited websites to inspect them for phishing threats

Findings

What we found

Drastic differences in categorization amongst URL filters

0-36.96% of websites classified as risky

No visible potential threats in all websites but one

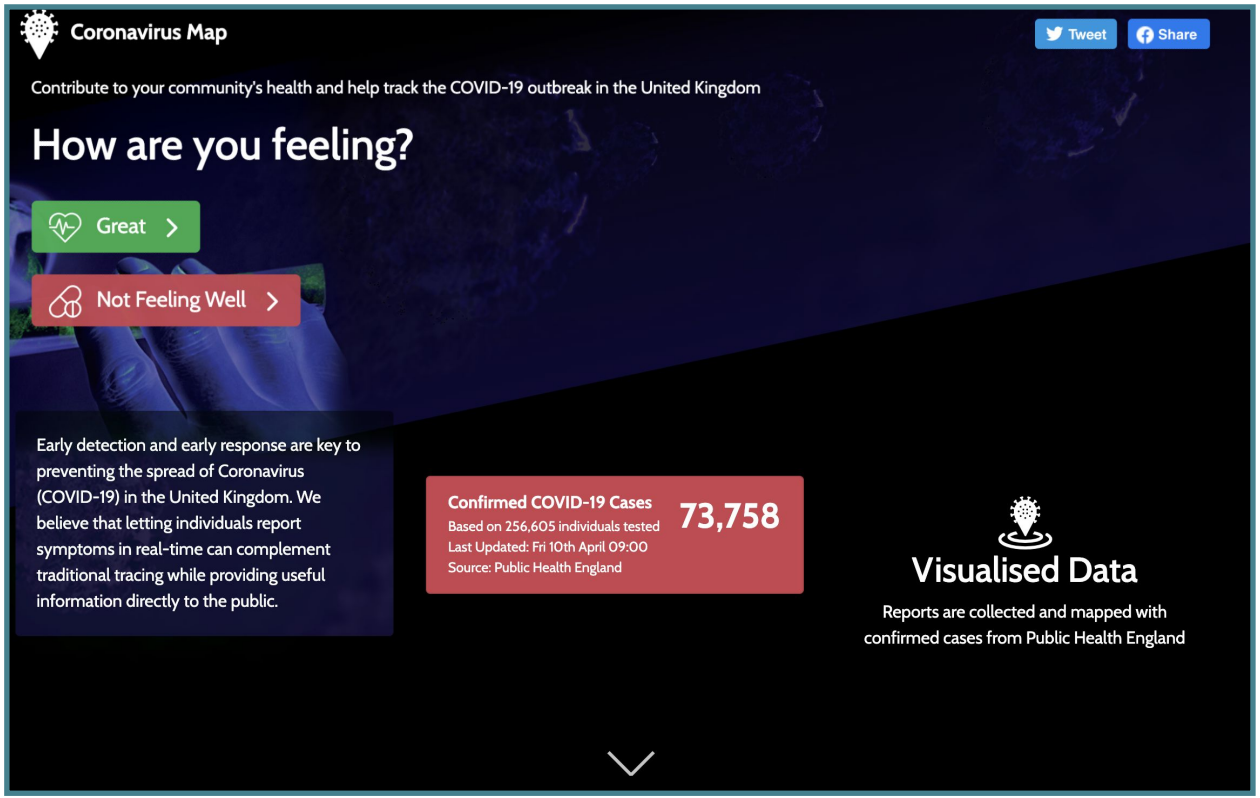
Most blocked websites were COVID-19 statistics websites

Many even warned users about inaccuracies in data and mentioned their sources

Sources often included Johns Hopkins University and government websites

URL classifier	Number of websites marked as threats
FortiGuard	46
WatchGuard	17
Palo Alto Networks	8
Checkphish	0

Example of blocked COVID-containing website



Implications

- There are **large discrepancies** from filter to filter

- Web filters are often used by organizations, ISPs, and governments for blocking dangerous websites.
  - Could cause unnecessary blocking for thousands of users during a time when information access is crucial

- There is a lack of
  - Transparency surrounding decisions filters make
  - Proper auditing for web filter-based censorship

Some Limitations & Future work

- Focused on censorship of factual content not misinformation
- Only focused on two methods of blocking
- Investigate the overflow of censorship policies enacted during the pandemic over time
- Search engine censorship
- Incorporate top URLs from non-Western search engines such as Yandex and Baidu

Thank you!

Feel free to reach out if you have questions