Open Measurement of Internet Censorship

Will Scott
willscott@gmail.com
OONI

- Mobile App 2017
- Android, iOS
- HTTP connectivity
- Middle-box Detection
- NDT speed test
• URL List: https://github.com/citizenlab/test-lists
• C library: https://github.com/measurement-kit
• Reports for: Kenya, Malaysia, Ethiopia, Gambia, Belarus, Egypt, Zambia, Turkey, Uganda, Brazil
Satellite | Interference in Iran

http://satellite.cs.washington.edu/
Satellite

Authoritative Resolver

Measurement Machine

Open Resolver
Satellite

- 7M distinct IPv4. 1.5M distinct /24
- Weekly Scan of Alexa top 10k domains
- 3 years of public data on https://scans.io/
Censys Data Collection

Censys collects data about hosts and websites through horizontal scans of the public IPv4 address space and the Alexa Top Million Domains. We validate, transform, and annotate this data with additional metadata (e.g., device model and geographic location) before we stream the records to a central database that maintains each host’s state. We perform these scans from the University of Michigan using ZMap, ZGrab, and ZTag. Detailed information about our data collection is available in our research paper.

Looking for a protocol we don’t scan or know how to identify a new type of device? Help us implement it and we’ll add it to our roster.

Primary Datasets

We publish daily snapshots of what we know about each host, website, and certificate. These datasets contain structured, non-ephemeral JSON records that identify a host’s configuration—similar to what is available in the search interface.

- IPv4 Address Space
- Alexa Top Million Domains
- X.509 Certificates

Regularly Scheduled Scans

Below are the regularly scheduled scans that power Censys. For each scan, we publish the host discovery scans and parsed application handshakes. We typically scan each protocol at least once weekly.

<table>
<thead>
<tr>
<th>Name</th>
<th>Port</th>
<th>Protocol</th>
<th>Subprotocol</th>
<th>Destination</th>
<th>Last Scan</th>
</tr>
</thead>
</table>

Censorship

• Four countries with significant DNS poisoning
  • Iran, Turkey, China, Indonesia

• 12 countries with explicit block pages
  • Russia, China, Iran, Saudi Arabia, Turkey, India, Indonesia, Greece, Sudan, Belgium, Cyprus, South Korea

• Many more with censorship:
  • Norway, France, Italy, Greece, Cyprus, Qatar, Portugal, Egypt, Thailand, Malaysia, Cuba, Ethiopia, Gambia
Censorship

- **Elections** are a common trigger for censorship.

- Censorship is often implemented by **ISPs**, not by the state, complicating attribution.

- Open measurements are **losing relevance** as censorship moves to social media and closed platforms.
Opportunities

• Format standardization
• Ground truth standardization
• Service Measurements