Knowing what isn’t there
Meditations on measuring digital censorship

Security Research Seminar at New York University, February 4, 2014
Matthew Finkel, Tor Project
Meredith Whittaker, Google Research
Serene Han, Google Ideas
Parameters of what we talk about when we talk about censorship measurement

**Scope:**

- We are **focused on a technological approach**. While we agree that this must be wed with many others approaches and disciplines, our expertise is technological.

- We’re **talking about post-hoc measurement** (what is blocked), not assessing the chilling effects of surveillance and violence (what wasn’t said, and why).
What we talk about when we talk about censorship measurement

Moving pieces:

● **Network measurement methodologies**
  ○ When something happens on networks, it (usually) leaves a trace -- how do we test, detect, report this?

● **Platforms and systems** -- infrastructure supporting comparative measurement and analysis

● **Data gathering** and sharing

● **Collaboration** and community
  ○ How to ensure that tools and systems that could respond to measurement data are able to
Why is measurement necessary?

Without measurement, digital censorship remains in the realm of anecdote and assumption.

When we measure, we gather evidence. Evidence is the language of political and technological action.
Case study

March 20th: **Collin Anderson, a researcher with UPenn**, uses public broadband performance data to look at Iranian censorship

---

**Dimming the Internet**  
Detecting Throttling as a Mechanism of Censorship in Iran

Collin Anderson  
collin@everysmallbird.com

Abstract. In the days immediately following the contested June 2009 Presidential election, Iranians attempting to reach news content and social media platforms were subject to unprecedented levels of the degradation, blocking, and jamming of communications channels. Rather than shut down networks, which would draw attention and controversy, the government was rumored to have slowed connection speeds to rates that would render the Internet nearly unusable, especially for the consumption and distribution of multimedia content. Since, political upheavals elsewhere have been associated with headlines such as “High usage slows down Internet in Bahrain” and “Syrian Internet slows during Friday protests once again,” with further rumors linking poor connectivity with political instability in Myanmar and Tibet. For governments threatened by public expression, the throttling of Internet connectivity appears to be an increasingly preferred and less detectable method of stifling the free flow of information. In order to assess this perceived trend and begin to create systems of accountability and transparency on such practices, we attempt to outline an initial strategy for utilizing a ubiquitous set of network measurements as a monitoring service, then apply such methodology to shed light on the recent history of censorship in Iran.

**Keywords**: censorship,national Internet, Iran, throttling, M-Lab

---

June 18th: **Collin publishes his findings** in a paper titled *Dimming the Internet: Detecting Throttling as a Mechanism of Censorship in Iran*
June 25th: Iran’s Minister of Communications admits to using throttling as a mechanism of censorship.

The reduction of Internet speed during the 2013 election period was the result of security measures taken to preserve calm.

Oct 4th: Findings included in UN report on Situation of Human Rights in the Islamic Republic of Iran.
Where are we now?

- **Platforms**
  - Measurement Lab, Citizen Lab, Georgia Tech

- **Tests and methodologies**
  - Tor’s OONI project, Citizen Lab’s ONI project, RIPE, M-Lab, independent researchers (Nick Weaver at ICSI, a number of others)

- **Available data**
  - Citizen Lab, OONI, M-Lab, ONI, RIPE...
OONI: Detecting Network Interference

OONI

Developed by the Tor Project to provide **open, auditable measurement of internet interference and censorship**. Currently working to run on the M-Lab platform.
OONI: Detecting Network Interference

A Human Rights Observation project
Evidence is great, if you’re a lawyer...

...but how do we stop censorship and repression of speech?
Traditional proxies (& VPNs)

Problems w/Proxies:
+ Scale
+ Trust

+ Filter
+ Surveille
+ Manipulate/Misdirect
uProxy
an experiment to share your connection
uproxy.org

XMPP/Chat/Email/Social: connects users
+ bootstraps WebRTC (HTML5's P2P encrypted browser communication)
+ obfuscation

+ A trust-controlled crowdsourcing of access & security
+ Potential for caching/speed up

Share your internet connection’s security & access with trusted friends & family
One approach: circumvention through targeted measurement

A vision for using data to enable better, faster, more complete access to blocked content.
Big topics we think about a lot

How to validate and verify results?

- Who has access to data, and how, and how to ensure data is stewarded properly?
- How to reproduce results? (You know, science)
- How to avoid false positives? (Or, a messed up DNS resolver doesn’t mean there’s a censor on the line.)
- How to create and nurture a community that can contribute tests, analysis, and methodological development?
How to secure tests and platforms?

- How to ensure that tests aren’t tampered with, and that platforms supporting testing and collection aren’t compromised?
- How to obfuscate test traffic to ensure it’s not whitelisted? (Which results in a false negative.)
- How to properly log to ensure tampering is detected?
How to **define censorship?**

- Knowing that something is “censored” requires **projecting intentionality**. Can we do this?
- Censorship isn’t just post-hoc blocking. It’s **also the chilling of speech before it happens**. We need to be aware of this when we reach conclusions based on measurement.
- What are the **ethics of defining censorship**? (Or, what do copyright owners and an Islamist dictator have in common?)
Big topics we think about a lot

How to include lived experience and context?

- How do we ensure that the absence of hard data doesn’t invalidate lived experience and ground truth?
- And, more actionably, how do we create ways for hard data to incorporate context and meaning?
Thank you!

If you’d like to learn more…

- measurementlab.net
- ooni.torproject.org
- opennet.net
- uproxy.org