Censorship, Freedom of Speech, and Architecture

Steven M. Bellovin
https://www.cs.columbia.edu/~smb

Warning

These are very vague, preliminary thoughts

Censorship Types

- Blocking information
- Drowning it out

We see both on the Internet today

Governments and businesses are both problematic

Governments

- Block access to certain destinations
- Look for and delete “bad” content
- Social and regulatory pressure to discourage creation of “bad” content
  - Note well: this latter implies that anonymity (and anonymity technology) are “bad”, too

The Great Firewall (Not just the Chinese)

- Many different technologies used
  - DNS games
  - Routing
  - Firewalls at national borders
  - TCP resets
  - Some reports of machine learning to detect “misuse” of standard port numbers

Failed Defenses

- Cryptography doesn’t hide metadata
  - And encryption is easily detectable
- It’s hard to route around a strong topological barrier
- Tor? VPNs? Detectable and blockable
- Bypasses often work for a while—and then the authorities catch on

Corporations

- Control over access
- Concentration of content
- Distraction

Control Over Access

- The “last mile” problem is very, very hard to solve
- All known solutions are capital-intensive; most are natural monopolies
- Wireless? Don’t forget spectrum limits and backhaul costs
- Mesh networks? There’s a bottleneck going to the Internet—desired resources are often not local

Concentration of Content

- The interesting content today takes far greater resources to create
- There’s a feedback loop: the bigger you are, the better the content you can create, which will attract more people
  - Social networks? Remember Metcalfe’s Law

Hosting Content

- Hosting is cheaper at scale
- It demands always-on, high-speed connections
- Large files or high-demand files *require* content distribution networks
    - That’s physics: low latency, and no need for massive bandwidth in one spot (with the consequent peering issues)

Distraction

- Corporations profit when you spend time on their sites
- They therefore expend vast efforts to get us there and make us stay there (see, e.g., https://www.ted.com/talks/zeynep_tufekci_we_re_building_a_dystopia_just_to_make_people_click_on_ads)
- But attention you spend on that sort of content is attention you don't have for anything else ("eat your vegetables")
- And: fake news (Gresham's Law of Content?)

So What Do We Do?

What Do We Do?

- We can’t fix physics with architecture
- We can’t fix the cost of high-quality content—that isn’t architecture
- But maybe we can work around some of this

Solving Censorship

- To solve censorship, we need
  - Ubiquitous encryption, to thwart content-based filtering
  - Opaque metadata, to thwart packet filters and routing attacks
  - No visible queries
- And do all of this without incurring the expense of Tor

Nullifying Great Firewalls

- Design so that *all* traffic is opaque, in both content and metadata
  - Perhaps make traffic-shaping a standard operation
- No visible lookups (e.g., DNS)
- Obviously, everything is encrypted
- *But: we can't make the Internet run like Tor; it's too slow*

In the Short Term

- Encourage encryption
- Encourage anonymity technology, including unlinkable credentials
- Work on address agility
- Work on invisible replacements for the DNS

Longer Term, Blue Sky…

Note Well: These are concepts; I don't know how to do most of them…

Perhaps *m* of *n* Coding?

- Pull down content from multiple sites
- You need enough data from enough sites to see anything; without that, you see nothing
- Important: *lots* of content has to be this way, so that it isn't suspicious
- It need not be *m* of *n*; encryption is fine, if the source is opaque and "good" content looks the same

Name-Based Networking?

- Perhaps name-based networking will do it
- But: the names need to be opaque
- A packet with an encrypted request name has to be matched with a packet with an encrypted resource name—and cheaply
  - The resulting file can’t carry any information that can be used for censorship once inside the Great Firewall
- But how do we do authentication of such files? Vital because of the fake news problem.

Centralization is Bad

- A centralized naming system (yes, the DNS) is bad, because it provides a control nexus
  - But—see Zooko's Triangle
  - And what about combating abuse, e.g., domain-squatting?
- CIDR is problematic, because it encourages concentration
  - But how do we route scalably without it?
  - And what about routing security?

Why CDNs?

- Content has to be near the consumer—but do we need explicit CDNs?

- Maybe content is auto-cached near recipients

- Payment? Eyeball ISPs save some money on peering charges—but perhaps large files should carry payments (cryptocurrency, with attached code to govern payment terms and conditions?)

- Peer-to-peer CDNs! (Might need higher bandwidth everywhere)

Can We Do This?

- Some of the technology exists
- Other items don’t exist or are far too expensive
- What is the economic model for deployment?
- What are the legal obstacles in major countries?

Questions?