Incentivizing Censorship Measurements via Circumvention

Ihsan Ayyub Qazi
Aqib Nisar*
Aqsa Kashaf**
Zartash A. Uzmi

* Now at USC
** Now at CMU
Internet censorship is pervasive!

- Over 70 countries restrict Internet access
  - Often due to political, social, or economic reasons
Censorship has a substantial impact...
... on different stakeholders in the Internet ecosystem.

- Users
- ISPs
- Advertisers
- Content Providers
- Government
It has led to the design of censorship...

Measurement Systems

- *What* is blocked?
- *Where* is it blocked?
- *How* is it blocked?
- *When* it is blocked?

... CensMon, Iris, Augur, Encore

Circumvention Systems

How do we *bypass* censorship?

Tor, LANTERN, uProxy
Current practice and limitations

- Circumvention systems are not data-driven
  - ... leads to one-size-fits-all solutions!
- Censorship measurement systems lack incentives
  - ... limits availability of geographically distributed probe points

In this work we ask, "Can we address the limitations of individual systems by consolidating them in a single platform?"
C-Saw in 1-slide

- **Consolidates** measurements and circumvention
  - Uses **crowdsourcing** to gather censorship measurements
  - Offers **data-driven circumvention**

- Better circumvention performance incentivizes more users to opt-in
Rest of the talk

- Web Censorship & Circumvention
- C-Saw Design
- Evaluation
- Deployment
Rest of the talk

- Web Censorship & Circumvention
- C-Saw Design
- Evaluation
- Deployment
Web censorship techniques

- Web filtering can be performed by intercepting a user request at different levels of the protocol stack.
Circumvention approaches

- Public DNS Servers
- Domain Fronting
- VPNs
- Tor
- Lantern
- ... others
Circumvention: local fix vs relay-based
What are the opportunities for improving circumvention performance?
A censorship case study in Pakistan

- Measurements taken from different vantage points
  - University campus (Lahore)
    - Served by ISP-A and ISP-B
  - Home users (Karachi)
    - Served by ISP-B only
A censorship case study in Pakistan

HTTP Traffic

HTTPS Traffic

HTTP Traffic

HTTPS with Domain Fronting

ISP A

ISP B

YouTube
(1) Insights about censors

- Blocking mechanisms can differ *across* ISPs
- Blocking mechanisms can differ *across* URLs even within an ISP

Insights hold across several countries
(2) Circumvention insights - 1/2

Fetched:
YouTube homepage
200 runs

ISP-B:
Blocking: HTTP & HTTPS
HTTPS/DF

Measurement point:
Campus network

All static proxies exhibited longer PLTs than the local fix.
(2) Circumvention insights - 2/2

**Fetched:**
YouTube homepage
200 runs

**ISP-A:**
HTTP Blocking Only
**HTTPS**

**Measurement point:**
Campus network
Tor exit relay shown

---

**Table 2:**

<table>
<thead>
<tr>
<th>Static Proxy Location</th>
<th>HTTP Blocking</th>
<th>HTTPS/Domain-Fronting</th>
<th>HTTPS/DF vs. Static Proxies</th>
<th>HTTPS vs. Tor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>CDF</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.8</td>
<td>0.8</td>
</tr>
</tbody>
</table>

**Figure 2:**
Comparison of (a) static proxies located around the world with HTTPS/Domain-Fronting, (b) HTTPS vs. Tor with different exit relay locations for fetching the YouTube homepage. As our results exhibited a similar trend across weeks, we report only a subset of the results.

**Figure 1:**
- **Block page w/o redir:** Redirected to a block page
- **No dNs redir:** When a TCP reset is received and concluded to be due to filtering devices.
- **No HTTP Resp:** When no HTTP response is received,
- **HTTP Blocking Only:**
- **DNS Redir:**
- **RST:**
- **No HTTP Resp:**
- **No dNs:**
- **HTTP Blocking:**
- **Block Page w/o Redir:**
- **Redirected to a block page**

- **Page Load Time (ms):**
  - 0
  - 1000
  - 2000
  - 3000
  - 4000
  - 5000
  - 6000
  - 7000
  - 8000

- **IPv4:**
  - 192.168.1.1
  - 10.0.0.1
  - 172.16.0.1
  - 192.168.0.1
  - 10.0.0.2
  - 172.16.0.2

- **IPv6:**
  - 2001:0db8:85a3:0000:0000:8a2e:0370:7334
  - 2001:0db8:85a3:0000:0000:8a2e:0370:7334/128

- **Domain:**
  - youtube.com
  - google.com

- **HTTP:**
  - 50 KB page size.

- **HTTPS:**
  - 360 KB in size

- **TTL:**
  - 100

- **DNS:**
  - 5000

- **Measurement point:**
  - Campus network
  - Tor exit relay shown

---

**Observations:**
- DNS censorship (e.g., via dropping of DNS requests or responses, redirects to a private IP address)
- HTTP blocking (e.g. by dropping the HTTP GET request, or to the address of a proxy that delivers a block page) and HTTP-level blocking, whereas ISP-B blocked both HTTP requests and
  - DSL
  - Cat
  - Ethernet

---

**Figure 3:**
- **Figure 3(a):**
  - HTTPS/DF vs. Static Proxies
  - CDF
  - 0.2
  - 0.4
  - 0.6
  - 0.8
  - 1.0

- **Figure 3(b):**
  - HTTPS vs. Tor
  - CDF
  - 0.2
  - 0.4
  - 0.6
  - 0.8
  - 1.0

---

**Results:**
- HTTP blocking, which can be used to deliver a block page, or injecting a TCP reset is common,
- Such heterogeneity across countries, measured using the ONI dataset [15, 38].
- In addition, ISP-B was also observed to be carrying out DNS blocking. This is known as DNS redir. Other countries including Thailand, UAE, Burma, and South Korea have also been observed in several studies.

---

**Table 1:**

<table>
<thead>
<tr>
<th>ISP (AS Number)</th>
<th>Fraction of blocking types</th>
</tr>
</thead>
<tbody>
<tr>
<td>7500</td>
<td>0.2</td>
</tr>
<tr>
<td>4795</td>
<td>0.6</td>
</tr>
<tr>
<td>8000</td>
<td>0.2</td>
</tr>
<tr>
<td>45899</td>
<td>0.8</td>
</tr>
<tr>
<td>12997</td>
<td>0.4</td>
</tr>
</tbody>
</table>

---

**Discussion:**
- We carried out measurements over several weeks to study PLTs under different blocking mechanisms used by ISP-A and ISP-B, both of which are located in Pakistan.
- The ping latency for YouTube from the same location was 186ms.

---

**Conclusion:**
- Fine-grained censorship measurements can reveal such differences in blocking mechanisms, which can be used to identify and mitigate such blocking techniques.

---

**Further studies:**
- More detailed analysis of the specific blocking techniques used by different ISPs and countries can provide insights into the effectiveness of different circumvention approaches.
Different circumvention strategies impose widely different overheads.
Key implication for design

Measurements reveal differences in blocking mechanisms

Can pick the least overhead circumvention strategy
Rest of the talk

- Web Censorship & Circumvention
- C-Saw Design
- Evaluation
- Deployment
Rest of the talk

- Web Censorship & Circumvention
- C-Saw Design
- Evaluation
- Deployment
Design goals

1. Scalable measurements with user consent

2. Adaptive circumvention

- In addition, a practical and usable solution should
  - require no target lists
  - preserve privacy of users contributing measurements
How C-Saw meets these goals?

1. Scalable measurements with user consent
   - C-Saw offers small PLTs as an incentive
   - It only measures those URLs that a user actually visits
   - As a result, it requires no target lists!

2. Adaptive circumvention
   - C-Saw measures the blocking mechanism used by a censor
   - Selects the least overhead circumvention strategy
C-Saw components

Direct path measured for censorship

Censored website

Circumvention

Proxy

local_DB

C-Saw Client

Censorship reports

List of blocked URLs in Client’s AS reported by other clients

Measurement Infrastructure

global_DB

URL A, blocked, DNS,...

URL B, blocked, No HTTP,...

…

server DB
C-Saw components

Direct path measured for censorship

censored website

Circumvention

List of blocked URLs in Client’s AS reported by other clients

C-Saw Client

Proxy

local_DB

Measurement Infrastructure

server DB

global_DB

URL A, blocked, DNS,...

URL B, blocked, No HTTP,...

...
C-Saw components

C-Saw Client

Proxy

censored website

List of blocked URLs in Client’s AS reported by other clients

Direct path measured for censorship

Circumvention

global_DB

server DB

URL A, blocked, DNS,...
URL B, blocked, No HTTP,...

Measurement Infrastructure
C-Saw proxy

- Measurement module
  - Runs a censorship detection algorithm
  - Issues redundant requests
  - Achieves resilience to false reports

- Circumvention module
  - Selects a circumvention approach (e.g., Public DNS, Domain Fronting, or Tor)
C-Saw Big Picture

Y is blocked in AS B using DNS blocking
Y is blocked in AS D using HTTP blocking
Z is blocked in AS X using HTTPS blocking
Security and privacy considerations

- Interference with C-Saw measurements
  
  • Rate limits creation of fake IDs and uses a voting mechanism

- Blocking access to the measurement infrastructure
  
  • One can use Tor hidden services

- User privacy and resilience to detection
  
  • All measurement reports are carried over the Tor network
Rest of the talk

- Web Censorship & Circumvention
- C-Saw Design
- Evaluation
- Deployment
Rest of the talk

- Web Censorship & Circumvention
- C-Saw Design
- Evaluation
- Deployment
Evaluation

- We implemented C-Saw using GitHub’s electron framework
  - Measures common forms of censorship
  - Implements several local fixes and optimizations
  - Supports Tor and Lantern as relay-based circumvention approaches

- Evaluation
  - Macro-benchmarks: C-Saw with Tor and Lantern
  - Micro-benchmarks: Impact of redundant requests, URL aggregation
Page Load Times with C-Saw

![Figure 7: CDF of PLTs of C-Saw in comparison with Lantern and Tor for (a) blocked webpage and (b) unblocked webpage.](image)

- **DNS Blocked Webpage**
  - C-Saw
  - Lantern
  - Tor

- **Unblocked Webpage**
  - C-Saw
  - Lantern
  - Tor
Rest of the talk

- Web Censorship & Circumvention
- C-Saw Design
- Evaluation
- Deployment
Rest of the talk

- Web Censorship & Circumvention
- C-Saw Design
- Evaluation
- Deployment
Deployment study

- We released C-Saw to 123 consenting users (3-month measurements)
  - Residential, Enterprise, and University network users in Pakistan
  - Users were carefully informed about C-Saw
    ‣ ... but were not given any list of blocked websites they needed to visit

- Insights
  - Users visited 420 blocked domains accessed through 16 different ASes
  - For majority of URLs, a block page was returned followed by DNS blocking
  - We found blocking of CDN servers
C-Saw in the wild

No. of users: 123
No. of unique blocked URLs accessed: 997
No. of unique blocked domains accessed: 420
No. of unique ASes: 16
Distinct types of blocking observed: 5
No. of URLs experiencing DNS blocking: 376
No. of URLs experiencing TCP connection timeout: 114
No. of URLs for which a block page was returned: 475
No. of unique updates: 1787

Table 7: Insights from our deployment study.

- Twitter was found blocked at 13:31 on Nov 25, 2017 from AS 38193 (Response: HTTP_GET_TIMEOUT)
- Twitter was found blocked at 13:32 on Nov 25, 2017 from AS 17557 (Response: HTTP_GET_BLOCKPAGE)
- Instagram was found blocked at 4:51 on Nov 26, 2017 from AS 38193 (Response: DNS blocking)
- Instagram was found blocked at 9:06 on Nov 26, 2017 from AS 59257 (Response: DNS blocking)
- Instagram was found blocked at 9:31 on Nov 26, 2017 from AS 45773 (Response: DNS blocking)

The above snapshot reveals interesting insights, which includes: (i) Different ASes were blocking Twitter differently and (ii) both Twitter and Instagram were blocked by different ASes. Our deployment experience suggests that C-Saw is effective at measuring censorship. Whether performance is a sufficient incentive for adoption, remains to be seen.

8 DISCUSSION

Ethics. There are well-known concerns regarding the ethics of censorship measurements w.r.t. consent, privacy, and safety, and include questions such as: Are users aware of which URLs they, or someone on their behalf, may be visiting and consent to it? How are users apprised of risks? To what extent can users be implicated for traffic that leaves their machine towards a censored website? When designing C-Saw, we took explicit steps to abide by the ethical guidelines set out by the Menlo Report [30]. For instance, C-Saw incorporates explicit user consent in its design by measuring only those URLs that users access with their knowledge. With C-Saw, users can choose to stay anonymous by using only those circumvention methods that provide anonymity. To protect the identity of users contributing measurements, C-Saw does not store IP addresses and uses Tor for sending measurement reports.

Why not use Tor for measurements? Tor’s design makes it challenging to detect censorship and provide adaptive circumvention. As Tor is geared towards providing anonymity, it always uses relays for accessing every URL making it hard to measure censorship on the direct path. C-Saw’s design allows new circumvention methods to be seamlessly integrated into its framework. This would not be possible if a design with a specific circumvention mechanism is used.

Robustness of C-Saw. C-Saw uses Tor as one possible circumvention strategy. However, Tor exits can be easily blacklisted [51], which raises concerns about C-Saw’s robustness. While using Tor bridges and pluggable transports makes it more challenging to block Tor, there is an arms race between Tor and some censors (e.g., China) [53]. Our hope is that C-Saw can ride on Tor’s successes in achieving blocking resistance with features like pluggable transports. However, it is useful to highlight that censors in several countries are neither as resourceful nor motivated as the censors in countries like China and Iran. Second, Tor is just one of the many circumvention mechanisms that C-Saw relies on. New circumvention approaches can be readily incorporated into its framework.

Fingerprintability of C-Saw. Suppose a censor shuts down circumvention mechanism $X$, and as a result, C-Saw users migrate to mechanism $Y$. This locking behavior may make it easier for censors to fingerprint users who switch to a particular (anonymous) circumvention mechanism. The fingerprinting effectiveness, however, would depend on a number of factors including the number of C-Saw users within an...
Limitations and discussion

- Scope of measurements
  - Difficult to measure unpopular websites or censorship at specific times

- Robustness of C-Saw
  - Relies on Tor as one possible circumvention strategy
  - Arms race between Tor and some censors (e.g., China)
  - New circumvention approaches can be easily incorporated in C-Saw

- Non-Web filtering
Summary

- **Censorship Measurements**
  - C-Saw uses *crowdsourcing* to collect measurements

- **Circumvention Performance**
  - Censorship measurements enable *adaptive* circumvention
  - Small PLTs *incentivize* users to opt-in

---

**Diagram**

1. **Crowdsource & Measure**
2. **Analyze & Learn**
3. **Circumvent**
4. **Incentivize**