A Look at the Consequences of Internet Censorship through an ISP Lens

S. Khattak, M. Javed, S. Khayam, Z. Uzmi, V. Paxson, IMC 2014
Effective policy requires informed perspectives on how humans actually respond to events!

When a persistent censorship policy emerges
  
  - Do users comply and stop accessing the blocked content or do they subvert censorship on a massive scale?
  - Does censorship hurt or benefit ISPs?
  - How much do competing content providers thrive?
Challenges

- Measuring consequences of censorship requires data snapshots before and after the events.
- A vantage point that captures all traffic a user exchanges with the Internet.
We examine one slice of this overall question (the consequences of Internet censorship) in the context of ISP customers in Pakistan.
Network traces (six) captured at a medium-sized Pakistani ISP at different points between Oct’11-Aug’13

Represent snapshots around two major censorship events:

- Nov’11: Thousands of porn domains blocked
- Sep’12: YouTube blocked … lifted in Jan.’16
Description of dataset

- ~1.8TB data
- Entire analysis based on Bro* protocol logs
- Individual traces
  - Range between 200-500GB and 6-20hrs
- Traces split into Small Office/Home Office (SOHO) and residential traffic

*Bro: A powerful network analysis framework originally written by V. Paxson (www.bro.org)
Through the Tier-2 Pakistani ISP peers with a tier-1 provider through the Transworld Associates TWA-1 submarine telecommunications cable in Karachi.

The survey results help shape the scope of our YouTube censorship. Finally, trace characteristics might differ between traces, a limitation of our work is that we cannot exclusively attribute cross-trace trends to the consequences of censorship; these

<table>
<thead>
<tr>
<th>Day</th>
<th>Local IPs</th>
<th>100–1,000 IP addresses</th>
<th>P=Porn, Y=YouTube</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tue</td>
<td>08:54</td>
<td>6–16 hours</td>
<td>200–500GB</td>
</tr>
<tr>
<td>Tue</td>
<td>17:48</td>
<td>6–16 hours</td>
<td>200–500GB</td>
</tr>
<tr>
<td>Tue</td>
<td>18:49</td>
<td>222</td>
<td>1,075</td>
</tr>
<tr>
<td>Tue</td>
<td>03Oct11</td>
<td>03Oct11</td>
<td>222</td>
</tr>
<tr>
<td>Tue</td>
<td>28Feb12</td>
<td>28Feb12</td>
<td>200</td>
</tr>
<tr>
<td>Tue</td>
<td>22Oct11</td>
<td>22Oct11</td>
<td>286</td>
</tr>
</tbody>
</table>

Can observe internal ISP data and inbound/outbound traffic.

Key

Block

Protocol Logs.
Ethical standards

Authors of this work with direct access to data signed a contract highlighting the obligations to

1. respect user privacy
2. not share data with third parties (which includes other co-authors)
3. not move data outside Pakistan
4. not move data within Pakistan without prior consent
5. undertake an objective study and refrain from maligning any party involved in the censorship landscape (user, ISP, or government)
A historic dataset for which we lack ground truth:

- What was blocked? (blacklist for the porn block)
- How was it blocked (DNS, TCP/IP, HTTP)?

Censor can block HTTP content at any of the layers involved and, at each layer, a range of choices

Censorship Indicators: A blocking mechanism leaves a trail in network traces

Ambiguous Indicators can occur because of legitimate reasons (server load, measurement loss)

Unambiguous Indicators can be exclusively attributed to censorship (DNS redirection)
Groundtruth reconstruction

- High frequency of ambiguous indicators for known censored content implies censorship
  - Known censored content: determined through a supplementary medium (e.g. newspapers)
  - Example: Consistently observe DNS No Response when the queried domain name is **porn**

<table>
<thead>
<tr>
<th>Domain</th>
<th>Category</th>
<th>DNS Reply</th>
</tr>
</thead>
<tbody>
<tr>
<td>facebook.com</td>
<td>Social networking</td>
<td>1.1.1.1</td>
</tr>
<tr>
<td>bad1.com</td>
<td>porn</td>
<td>-</td>
</tr>
<tr>
<td>bad2.com</td>
<td>porn</td>
<td>-</td>
</tr>
<tr>
<td>bbc.co.uk</td>
<td>news</td>
<td>2.2.2.2</td>
</tr>
</tbody>
</table>
Based on the above observations, we do not find concrete evidence of extensive IP- or HTTP-level blocking for porn, except for the cases where we observe a high overlap with our DNS blacklist.
YouTube Censorship Mechanism

YouTube blocked

Jan 2012  Apr  Jul  Oct  Jan 2013  Apr  Jul

DNS redirection
HTTP redirection

DNS redirection
HTTP no response
Analysis

- Impact on
  - end users
  - content providers and
  - service providers

- Go over salient results in question/answer fashion
Analysis

Consequences on content providers ➔ What constitutes a content provider?

- Determine porn content by classifying all websites occurring in the dataset by topic using McAfee URL category service

- Competing content providers to YouTube determined by regional popularity (DailyMotion, Vimeo, TunePK)
Consequences on users
What is a user\* response to a block page?

*IP address + HTTP agent

Network

Monitor HTTP for 5 minutes

Give me bad.com

Blocked
Users’ response

- **Porn**
  - 60% users perform search engine query (domain specific)
  - 70% users access another porn domain within 5’

- **YouTube**
  - 75% users perform search engine query (information retrieval)
  - 7% users access an alternate video content provided on the day of block, rising to 12% a year later
Users’ response

Do residential users shift to alternate/free DNS resolvers?

No shift (~8% queries to external DNS resolvers consistent with previous trend)
Do SOHO users shift to alternate/free DNS resolvers?

Local ISP drops from ~90% to ~69% on the day of YouTube block.
Users’ response

Do SOHO users shift to alternate/free DNS resolvers?

ISP handles ~74% queries (~90% pre-block) a year after YouTube block
Users’ traffic

Does traffic generated by residential users change?

Distribution of HTTP traffic and its ratio to SSL traffic.

- SSL
- Other
- Porn

Porn bandwidth (%) reduces to half the avg pre-block consumption
Does traffic generated by SOHO users change?

- Porn bandwidth (%) reduces to 1/3 of the average pre-block consumption
- Video traffic reduces to 12% and a year later to ~5% from an average pre-block consumption of 50%
- Drastic increase in SSL post YouTube block
Consequences on content providers (CP)
Does user demand for porn CP changes?

Ongoing churn in porn domains post block

Traffic vol. from Top-5 porn domains
(for residential users, similar trends for SOHO users)
How is video traffic distributed among CP?

Distribution of volume of inbound video traffic among YouTube and its competitors.

- YouTube prior to the block
- Competitors post the block
And users’ embedded video watch requests?

<iframe frameborder="0" width="600" height="300" src="/www.dailymotion.com/embed/video/x26ql41" allowfullscreen></iframe>
And users’ embedded video watch requests?

Competitors get share in YouTube’s pre-block dominant position.
Consequences on operators
Where do operators fetch videos from?

Prior to YouTube block, ISP gets to server bandwidth hungry video locally.
Where do operators fetch videos from?

Leakage due to caching on the day of YouTube block

Traces

ASN of video content providers

18Sep12
28Feb12
03Oct11
21Dec11
22Oct11
Google, US

Local-ISP, PK

Video Traffic Vol.
Where do operators fetch videos from?

Post YouTube block, ISP has to pay peers for it
Summary

- Porn block: significant lessening of traffic; some shifting to equivalent alternate sources
  - Censor’s presumed goal at least partially met

- YouTube block:
  - Spurred some users to outsource their DNS
  - Spurred shift to SSL
  - Shifted cost structure: ISPs burdened, YouTube competitors thrived
We looked at how YouTube block impacts competing content providers
  – YouTube vs. DailyMotion, Vimeo, TunePK

Identities on the Internet are not entirely isolated (YouTube, Google)
  – What is the impact of YouTube block on other Google services?
Traffic to YouTube (as seen by Google)
Traffic to Google docs (as seen by Google)

Browse real-time traffic to Google products and services

This page provides near real-time information about traffic to our products and services around the world. Each graph shows historic traffic patterns for a given geographic region and product. For more information, see our FAQ.

Pakistan  Google Docs

Fraction of Worldwide Traffic, Normalized

Data after this point are still being finalized. Interpret them with caution.

2014-9-26

YouTube blocked
What does the wide usage of circumvention tools mean for ad targeting?
- Anonymize IP address ➔ ad geotargeting is hurt
- Strip off HTTP cookie ➔ ads cannot target user profile any more