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

Sheharbano Khattak (University of Cambridge)
Mobin Javed (UC Berkeley)
Syed Ali Khayam (PLUMgrid)
Zartash Afzal Uzmi (LUMS SBASSE, Pakistan)
Vern Paxson (UC Berkeley, ICSI)
Agenda

• Problem Statement
• Description of Dataset
• Groundtruth Reconstruction
• Analysis
• Discussion
Consequences of Internet Censorship

- Effective policy requires informed perspectives on how humans actually respond to events
Consequences of Internet Censorship

• 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.
Agenda

- Problem Statement
- Description of Dataset
- Groundtruth Reconstruction
- Analysis
- Discussion
Description of Dataset

- **Network traces** captured at a medium-sized Pakistani ISP at different points between Oct’11-Aug’13
Description of Dataset

• Network traces 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 (continues to date..)
Description of Dataset

The diagram illustrates a timeline from January 2012 to January 2013. The dates 03 Oct 2011 and 22 Oct 2011 are highlighted on the timeline.
Description of Dataset

Porn blocked

03Oct11

22Oct11
Description of Dataset

Porn blocked

- 03 Oct 2011
- 21 Dec 2011
- 22 Oct 2011
- 28 Feb 2012
Description of Dataset

- YouTube blocked: 22 Oct 2011
Description of Dataset

Porn blocked:
- 03 Oct 2011
- 21 Dec 2011
- 28 Feb 2012

YouTube blocked:
- 18 Sep 2012
- 02 Aug 2013
Description of Dataset

- ~1.8 TB data
- Entire analysis based on *Bro* protocol logs
Description of Dataset

- ~1.8 TB data
- Entire analysis based on Bro protocol logs
- Individual traces
  - range between 200-500 GB and 6-20 hours
Description of Dataset

- ~1.8 TB data
- Entire analysis based on *Bro* protocol logs
- Individual traces
  - range between 200-500 GB and 6-20 hours
- Traces split into **Small Office/Home Office (SOHO)** and **Residential Traffic**
Can observe internal ISP data and inbound/outbound traffic
Agenda

- Problem Statement
- Description of Dataset
- Groundtruth Reconstruction
- Analysis
- Discussion
Groundtruth Reconstruction

• We have historic dataset for which we lack ground truth:
  ✤ What was blocked?
  ✤ How was it blocked (DNS, TCP/IP, HTTP)

Groundtruth Reconstruction

• **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)
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><strong>bad1.com</strong></td>
<td><strong>Porn</strong></td>
<td>-</td>
</tr>
<tr>
<td><strong>bad2.com</strong></td>
<td><strong>Porn</strong></td>
<td>-</td>
</tr>
<tr>
<td>bbc.co.uk</td>
<td>News</td>
<td>2.2.2.2</td>
</tr>
</tbody>
</table>
Porn Censorship Mechanism

- DNS Redirection
- IP Block
- DNS Redirection
- HTTP No Response
YouTube Censorship Mechanism

- DNS Redirection
- HTTP Redirection

YouTube blocked

- DNS Redirection
- HTTP No Response
Agenda

• Problem Statement

• Description of Dataset

• Groundtruth Reconstruction

  • Analysis

• Discussion
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 our dataset by topic using McAfee URL categ. service
Analysis

• Consequences on content providers—>What constitutes a content provider?

- Determine porn content by classifying all websites occurring in our dataset by topic using McAfee URL categ. service

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

* IP address + HTTP User Agent

Give me bad.com

Blocked!
What is user* response after viewing a block page?

User

Give me bad.com

Blocked!

Monitor HTTP for 5 mins

* IP address + HTTP User Agent
What is user* response after viewing a block page?

• Porn:
  ✤ 60% users perform search engine query (domain-specific)
  ✤ 70% users access another porn domain

• YouTube:
  ✤ 75% users perform search engine query (information retrieval)
  ✤ 7% users access an alternate video content provider on the day of block, rising to 12% a year later.

* IP address + HTTP User Agent
Do residential users shift to alternate/free DNS resolvers?

No shift (~8% queries to external DNS resolvers consistent with previous trend)

* Percentage distribution
Do **SOHO** users shift to alternate/free DNS resolvers?

Local ISP drops from ~90% to ~69% on the day of YouTube block

*Percentage distribution*
Do **SOHO** users shift to alternate/free DNS resolvers?

Traces

03Oct11

22Oct11

21Dec11

28Feb12

02Aug13

18Sep12

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

* Percentage distribution
Does traffic generated by residential users change?

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

Porn bandwidth (%) reduces to 1/3 of the average pre-block consumption.
Does traffic generated by SOHO users change?

Video traffic reduces to 12% and a year later to ~5% from an average pre-block consumption of 50%.
Does traffic generated by SOHO users change?

Drastic increase in SSL post YouTube block
Consequences on Content Providers
Does (residential*) user demand for porn content providers change?

Ongoing churn in porn domains post block

Traffic Vol. from Top-5 Porn Domains**

* Similar trend for SOHO users
** Percentage distribution
How is video traffic distributed among content providers?

YouTube rules the world prior to the block
How is video traffic distributed among content providers?

Competitors pick up the slack post YouTube block.
How are users’ embedded video watch requests distributed among content providers?

<iframe frameborder="0" width="600" height="300" src="//www.dailymotion.com/embed/video/x26ql41" allowfullscreen></iframe>
How are users’ embedded video watch requests distributed among content providers?

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

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

Leakage due to caching on the day of YouTube block
Where do operators fetch videos from?

Post YouTube block, ISP has to pay peers for it

Video Traffic Vol.
Summary

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

• Problem Statement
• Description of Dataset
• Groundtruth Reconstruction
• Metrics
• Analysis
• Discussion
1. Collateral Damage on Blocked Content Provider

• We looked at how YouTube block impacts competing content providers
  ✤ YouTube vs. DailyMotion, Vimeo, TunePK
1. Collateral Damage on Blocked Content Provider

• 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.

Fraction of Worldwide Traffic, Normalized

YouTube blocked
2. Impact on Ad Targeting (and Revenue?)

• What does the wide usage of circumvention tools mean for ad targeting?
2. Impact on Ad Targeting (and Revenue?)

• What does the wide usage of circumvention tools mean for ad targeting?
  ✤ Anonymise IP address—>ad geotargeting is hurt
  ✤ Strip off HTTP cookie—>ads cannot target user profile any more
Thanks
Q&A

Sheharbano.Khattak@cl.cam.ac.uk
mobin@eecs.berkeley.edu
akhayam@plumgrid.com
zartash@lums.edu.pk
vern@cs.berkeley.edu