The Anatomy of Web Censorship in Pakistan

Zubair Nabi
zubair.nabi@cantab.net
Information Technology University, Pakistan*

Presented by: Mobin Javed
UC Berkeley

* Now at IBM Research, Dublin
This website is not accessible in Pakistan!

- First study of the cause, effect, and mechanism of Internet censorship in Pakistan
- Upgrade to centralized system in the middle of the study (May 2013)
- Censorship mechanism varies across websites: some blocked at the DNS level; others at the HTTP level
- Public VPN services and web proxies popular tools to bypass restrictions
Outline

- Background: Pakistan and related work
- Methodology
- Results
- Alternative circumvention methods
- Summary
- Future work
- Qs
Internet in Pakistan

- **16 million** users or **9%** of total population (World Bank, 2012)

- Out of the total Internet users, **64%** access news websites (YouGov, 2011)

- Largest IXP (AS17557) owned by the state

- Internet, fixed-line telephony, cable TV, and cellular services regulation by the **Pakistan Telecommunication Authority (PTA)**
  - Also in charge of censorship
History of Censorship

- 2006: 12 websites blocked for blasphemous content
- 2008: A number of YouTube videos blocked
  - IP-wide block via BGP misconfiguration
  - YouTube rendered inaccessible for the rest of the world for 2 hours

Pakistan hijacks YouTube
24 FEB, 2008 | 7:50 PM | BY MARTIN BROWN

Pakistan ban to blame for YouTube blackout

Pakistan blamed for YouTube blackout

Pakistan turns off YouTube worldwide

How Pakistan Hacked YouTube for the World
History of Censorship (2)

- 2010: Facebook, YouTube, Flickr, and Wikipedia blocked in reaction to “Everybody Draw Muhammad Day”
  - PTA sanctioned to terminate any telecom service
History of Censorship (3)

- 2012 (March): Government requests proposals for gateway-level blocking system
  - Filtering from domain level to sub-folder level
  - Blocking individual IPs and/or entire range
  - Plug-and-play hardware units, capable of blocking 50 million URLs
    - Latency < 1ms
History of Censorship (4)

- 2012 (September): Infinite ban on YouTube in retaliation to “Innocence of Muslims”
  - Disruption of other Google services due to IP sharing
Related Work

• Verkamp and Gupta
  – PlanetLab nodes and volunteer machines, 11 countries
  – Key insight: *censorship mechanisms vary across countries*

• Mathrani and Alipour
  – Private VPNs and volunteer nodes, 10 countries
  – Key insight: *restrictions applicable to all categories of websites: political, social, etc.*

• Dainotti *et al.*
  – Internet blockage during the *Arab Spring*
Methodology: Dataset

- Publicly available list with 597 websites
- Compiled in 2010
- Not exhaustive but a fairly rich of complete domains and subdomains
- Dataset after cleaning: 307 websites
  - Redundant, broken, and duplicates removed
- Checked with a public VPN beforehand to ensure connectivity
Methodology: Script

- Modified version of the CensMon system (FOCI '11)
  1) DNS lookup
     - Local and public (Google, Comodo, OpenDNS, Level3, and Norton)
  2) IP blacklisting: TCP connection to port 80
  3) URL keyword filtering:
     - http://www.google.com/fooURL
     - 404 Not Found under normal operation
  4) HTTP filtering: HTTP request, log response packet
- Also, logs transient connectivity errors, such as timeouts
Methodology: Networks

<table>
<thead>
<tr>
<th>ID</th>
<th>Nature</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network1</td>
<td>University</td>
<td>Lahore</td>
</tr>
<tr>
<td>Network2</td>
<td>University</td>
<td>Lahore</td>
</tr>
<tr>
<td>Network3</td>
<td>Home</td>
<td>Lahore</td>
</tr>
<tr>
<td>Network4</td>
<td>Home</td>
<td>Islamabad</td>
</tr>
<tr>
<td>Network5</td>
<td>Cellular (EDGE)</td>
<td>Islamabad</td>
</tr>
</tbody>
</table>

- *Network1* and 2: gigabit connectivity
- *Network5* only used for post-April testing
- Tests performed at night time to minimize interaction with normal traffic
- Performed on multiple occasions for precision
Results: Pre-April

- Most websites blocked at **DNS-level**
  - Local DNS: “Non-Existent Domain” (NXDOMAIN)
  - Public DNS: NXDOMAIN for Google DNS and Level3
    - NXDOMAIN redirector in case of Norton DNS, Comodo, and OpenDNS
- No evidence of **IP** or **URL-keyword** filtering
- Some websites filtered through **HTTP 302** redirection
  - Triggered by **hostname** and **object URI**
  - Done at the ISP level
Dear Valuable Customer,

Your requested site is blocked by PTA. Please consult PTA if you have any query regarding requested site.

Visit=?
Results: Post-April

- HTTP 302 redirection replaced with **IXP-level 200 packet injection**
  - Triggered by **hostname and URI**
  - Because of the 200 code, the browser believes it's a normal response
    - Stops it from fetching content from the intended destination
    - Original TCP connection times out
- Same response packet and screen across ISPs
  - Except *Network4* (still under the influence of pre-April censoring)
IXP-level Warning Screen

Surf Safely!

This website is not accessible.

The site you are trying to access contains content that is prohibited for viewership from within Pakistan.

• Same results reported by “The Citizen Lab” in parallel in June, 2013
• System attributed to the Canadian firm Netsweeper Inc.
  – Also applicable to Qatar, UAE, Kuwait, and Yemen
Results: Survey

- 67 respondents
  - Results biased towards individuals with above-average computer skills
- Public VPN services, such as Hotspot Shield, most popular
- Web proxies also popular
Results: Survey

- 67 respondents
  - Results biased towards individuals with above-average computer skills
- Public VPN services, such as Hotspot Shield, most popular
- Web proxies also popular
Alternative Circumvention: Web-based DNS

- Generally, web-based service can also be used for lookup
- Results show that same websites also blocked at HTTP-level
- Similar to South Korea
  - DNS filtering used for websites that resolve to a single site
  - HTTP-level mechanism exclusively used for websites with IPs shared across hostnames and filtering needs to be selective
    - YouTube, Wikipedia, etc.
Alternative Circumvention: CDNs and Search Engine Caches

- No URL-keyword filtering
- Blocked websites accessible via CoralCDN
- Cached pages of blocked content also accessible on Google, Bing, and Internet Archive
Summary

- Pakistan has undergone an upgrade from ISP-level to centralized IXP-level censorship
- Most websites blocked at the DNS level, while a small number at the HTTP level
- Websites blocked at the DNS level also blocked at HTTP-level
- Most citizens use public VPNs and web proxies to circumvent restrictions
Future Work

- Expansion in the number of websites and networks
- Deeper analysis of DNS blockage
  - For instance, not clear if censoring module maintains a list of all resolvers and their redirectors or it queries the actual resolver each time
- Examination of side-effects of DNS injection (similar to China)
- Analysis of “Streisand Effect” in Pakistan
  - Early results look promising!