The Anatomy of Web Censorship in Pakistan

Zubair Nabi
Information Technology University, Pakistan
zubair.nabi@itu.edu.pk

Abstract
Over the years, the Internet has democratized the flow of information. Unfortunately, in parallel, authoritarian regimes and other entities (such as ISPs) for their vested interests have curtailed this flow by partially or fully censoring the web. The policy, mechanism, and extent of this censorship varies from country to country.

We present the first study of the cause, effect, and mechanism of web censorship in Pakistan. Specifically, we use a publicly available list of blocked websites and check their accessibility from multiple networks within the country. Our results indicate that the censorship mechanism varies across websites: some are blocked at the DNS level while others at the HTTP level. Interestingly, the government shifted to a centralized, Internet exchange level censorship system during the course of our study, enabling our findings to compare two generations of blocking systems. Furthermore, we report the outcome of a controlled survey to ascertain the mechanisms that are being actively employed by people to circumvent censorship. Finally, we discuss some simple but surprisingly unexplored methods of bypassing restrictions.

1 Introduction
In recent years, the Internet has faced an onslaught of censorship and restrictions with local [19] [14] and global [1] ramifications. The world over, authoritarian regimes under the pretext of maintaining public order have been blocking web access. This is more enunciated in the developing world where freedom of speech and freedom of information are largely undefined. In the same vein, Pakistan has become a poster-child for web censorship rooted in religion, politics, and conflict/security [16]. It has also been revealed as one of the 36 countries which host FinFisher Command & Control servers to spy on their citizens [13].

According to a 2012 World Bank study, 9% or around 16 million Pakistanis have access to the Internet [20]. Out of these 16 million users, 64% employ the Internet to access news websites [22]. Therefore, the government has a high incentive to stifle this access. Practically, filtering in Pakistan is largely geared towards blocking content which is considered a threat to national security and/or content which is blasphemous. The largest Internet Exchange Point (IXP) in the country is owned by the state which simplifies the enforcement of state-wide censorship. This censorship has been applied in waves during the past one decade—often mandated by the judiciary [16]. The side-effects of which at times have had global impact. For instance in 2008, a naive attempt to censor YouTube by the authorities, rendered the website unreachable for a large number of ASes across the world [5].

A large number of studies have recently been conducted to study the mechanism and effect of censorship around the world. Verkamp and Gupta [19] used PlanetLab nodes and volunteer machines to study the mechanisms of censorship in 11 countries. One of their key insights is that these mechanisms vary from country to country. Similarly, Mathrani and Alipour [14] presented the results of tests conducted across 10 countries using private VPNs and volunteer nodes. Their results show that restrictions in these countries are applicable to all categories of websites: politics, social networking, culture, news, entertainment, and religion. Likewise, Dainotti et al. [9] made use of publicly accessible datasets to dissect the Internet outages in Libya and Egypt during the Arab Spring. In the same vein, a large body of work [21] [8] [1] is dedicated to analyzing the modus operandi and consequence of censorship due to the Great Firewall of China.

To the best of our knowledge, this is the first work to dissect in detail the mechanism and effect of censorship in Pakistan. Unlike studies conducted for other countries which relied on volunteer machines and PlanetLab nodes, we directly use 5 different networks within Pakistan as vantage points to carry out our tests. More importantly, during the course of our tests the country underwent an upgrade to a central and standardized censorship system, reportedly developed by the Canadian firm Netsweeper Inc. [18]. Therefore, our results juxtapose two genera-

---

1To appear in: Proceedings of the 3rd USENIX Workshop on Free and Open Communications on the Internet (FOCI ’13), Washington DC, August 2013.

1The same firm has in the past provided its filtering services to Qatar, the United Arab Emirates, Kuwait, and Yemen [18].
tions of systems. Moreover, we present the outcome of a controlled survey to gauge the mechanisms through which citizens are currently circumventing online blockages. Finally, we augment these mechanisms by discussing the use of CDNs and search engine caches. Our results can be summarized as follows:

- A large number of websites are blocked using DNS injection
- The alternative mode of censorship in the previous system (at the ISP level) was HTTP 302 redirection and in case of the current system (at the IXP level), it is fake HTTP response injection
- Websites restricted at the DNS-level are also blocked at the HTTP-level
- A large fraction of people either use public VPN services or web proxies to access restricted content
- CDNs and search engine caches are currently viable options to access blocked content

The rest of the paper is organized as follows. We give a brief history of Internet censorship in Pakistan in [2]. The methodology employed for our study is discussed in [3]. We present the results of our tests and survey. Alternative anti-censorship mechanisms are discussed in [5]. We finally conclude in [6] and also discuss future directions.

2 Background

Both telephony and Internet services in Pakistan are managed by an arm of the state called the Pakistan Telecommunication Authority (PTA). It is in charge of regulation and licensing of fixed-line telephony, cellular services, cable TV, and Internet services within the country. Internet censorship is also enforced by the government through the PTA. To give the reader some perspective, the following is a timeline of censorship enforced by the government:

- 2006: 12 websites blocked for hosting blasphemous content. The content which was deemed offensive included a Blogspot blog. Lacking the infrastructure to block a particular blog, the entire Blogspot website was blocked for two months.
- 2008: A number of YouTube videos marked as offensive by the government. Instead of implementing a URL/IP-specific restriction, an IP-wide block of YouTube via BGP misconfiguration was enforced, making YouTube inaccessible for much of the Internet for 2 hours [5].
- 2010: Facebook, YouTube, Flickr, and Wikipedia partially or fully blocked in reaction to “Everybody Draw Muhammad Day". These websites were subsequently unblocked. The same year, the government also sanctioned the PTA to “order temporary or permanent termination of telecom services of any service provider, in any part or whole of Pakistan” [2].

- 2012 (March): The government requests proposals for a country-wide URL filtering and blocking system [15]. According to the advertisement, filtering at the time was enabled by manual mechanisms deployed at the ISP level and the desired system was required to enable centralized blocking at the national IP backbone. Some other features of the system included:
  - Filtering from domain level to sub-folder level as well as blocking of individual files and file types
  - Blocking individual IPs and/or an entire range
  - Remote network monitoring via SNMP and configuration through HTTP and HTTPS
  - Operation at L2 and L3
  - Modularity and scalability through standalone, plug-and-play hardware units capable of blocking up to 50 million URLs with a processing latency of less than 1ms
  - Decoupling of policy and mechanism via storage of blacklists in an external database

- 2012 (September): Indefinite ban on YouTube imposed in retaliation to a controversial movie [11].

The side-effects of this ban disrupted other Google services such as Maps, Drive, Play Store, and Analytics [4]. This was due to the fact that the same IPs are shared across all of these services.

3 Methodology

We use a publicly available dataset of websites to perform connectivity tests. The prime reason for using this particular dataset is that the same list of 597 websites was circulated by the PTA to all ISPs for filtering [3]. While not an exhaustive list, it provides a fairly rich set of both complete domains and subdomains for analysis.

As the list was compiled in 2010, the status of a number of websites has changed. For instance, in 2010 the government banned a small number of YouTube videos so the list contains individual entries for each video. Since then, the government has blocked YouTube entirely. Therefore we removed individual video URLs and added a single entry for YouTube. This reduced the size of the dataset to 562 links. In addition, the list also contains a number of duplicate entries. The removal of these entries further reduced the size of the list to 429. Finally, a number of websites have now gone offline, mostly proxy sites, bringing down the final tally of test sites to 307. To ensure that these final websites are actually restricted, we tested their...
connectivity using a public VPN service which terminates in the US. The results indicate that they are accessible via VPN and are thus restricted within Pakistan.

### 3.1 Test Script

Our test script, dubbed *Samizdat*[^1], closely mimics the CensMon [17] system with a few modifications. Unlike CensMon, our script does not relay test results to a remote server but logs them locally. Furthermore, it also tries to perform DNS resolution using public servers in addition to network-local resolvers. Finally, instead of using a custom server to determine URL keyword filtering, it uses a public portal.

The script first downloads the list of websites and carries out the cleaning phase described above. Then for each website it executes the following tasks:

1. Performs a DNS lookup and notes the returned IP address(es). If the lookup fails, the website is marked as blocked and the returned DNS response, such as NXDOMAIN, Timeout, etc. is logged. The lookup is performed for both the local ISP DNS service as well as a number of public DNS servers: Google (8.8.8.8), Comodo (8.26.56.26), OpenDNS (208.67.222.222), Level3 (209.244.0.3), and Norton (198.153.192.40).

2. If the DNS lookup is successful, it tries to open a TCP socket to the IP address(es) on port 80 to check for IP address blacklisting. Any connection failures are logged.

3. The next step is to check for URL keyword filtering. To this end, the URL of the website is appended to a scalable and publicly accessible portal: http://www.google.com. The script expects a HTTP 404 Not Found error under normal operation. If a different code is received, the website is marked for URL keyword filtering.

4. The final step is to send an HTTP request to the website. The response, along with its code is logged. In addition, in case of response code 301 or 302, the redirection URL (from the Location field) is also logged.

Additionally, all transient connectivity errors, such as timeouts, etc. are logged to a separate error log.

[^1]: Available online: https://github.com/ZubairNabi/Samizdat

### 3.2 Test Networks

Tests were conducted across 5 different networks: Network1-5. Table [1] gives the details of each network. Network1 and Network2, due to their academic nature, are connected to 3 and 2 ISPs, respectively, with gigabit connectivity and auto-failover. Network3 and Network4 on the other hand are connected to single but distinct ISPs. Network5, which was only used for post-April testing, is a cellular network with GPRS-EDGE Internet connectivity. The test script was executed on hosts within these networks at night time to minimize interaction with regular traffic and also to minimize false positives. In addition, the measurements were performed multiple times and on separate occasions to ensure precision. Finally, only those results are being reported for which the error log had no entries, i.e. there were no transient connectivity issues.

### 4 Results

In this section, we first present the results of the tests that we conducted to investigate the mechanism and extent of censorship in the country ([4.1] and [4.2]) followed by the outcome of our public survey ([4.3]).

#### 4.1 Pre-April 2013

Table [2] summarizes the results for all networks. We discuss each of the mechanisms next.

| ID     | Nature     | Location  |
|--------|------------|-----------|
| Network1 | University | Lahore    |
| Network2 | University | Lahore    |
| Network3 | Home       | Lahore    |
| Network4 | Home       | Islamabad |
| Network5 | Cellular (EDGE) | Islamabad |

*Table 1: Details of Test Networks*

| Mechanism | No. of Affected Sites | Percent |
|-----------|-----------------------|---------|
| DNS       | 187                   | 60.91   |
| IP        | 0                     | 0       |
| URL-keyword | 0                     | 0       |
| HTTP (302)| 5                     | 1.62    |
| Total     | 192                   | 62.53   |

*Table 2: Breakdown of Pre-April Test Results*
4.1.2 IP and URL-keyword Filtering

Post-DNS resolution, the script was able to connect to the IPs of all websites on port 80, showing that no IP-level blacklisting is in place. Similarly, the script received a 404 Not Found for each website after appending its URL to www.google.com, indicating that URL keyword filtering is also not in effect.

4.1.3 HTTP

A large number of websites exhibit normal behaviour and return either a 200 (OK) or 203 (Non-Authoritative Information) code. A few sites also have legitimate 301 (Moved Permanently) or 302 (Found) redirects, for instance from http://www.anonymizer.com to https://www.anonymizer.com/index.html. In total, 115 websites exhibited normal behaviour and have thus been unblocked by the authorities since the time the dataset was compiled.

5 websites, including YouTube, result in redirection to a warning screen. The warning screen varies from ISP to ISP (shown in Figure 1) which indicates that this redirection is done at the ISP-level. Table 3 presents the HTTP packet level trace for YouTube. It is clear from the table that the HTTP GET request is intercepted and the session is redirected to a private IP which displays the warning page. Figure 2 shows the HTTP 302 redirect response packet in which the Location field contains the redirection IP for one particular ISP. The same sequence of packets, although with different redirection IPs, was observed on all 5 networks. This censorship is triggered by a combination of the hostname and object URI (Host and Request URI fields) within the HTTP request header. We ascertained this by creating a custom HTTP request header for all 5 websites and sending it to an uncensored URL instead of the actual one. This resulted in redirection to the warning page in each case. For instance, setting the hostname as youtube.com triggers censorship regardless of the destination URL. In contrast, the hostname vimeo.com is allowed to go through but the hostname in conjunction with the URI 64414932 experiences redirection. It is also important to highlight that this censorship is an example of “filter and return” as opposed to “allow but return first” [19], i.e. the original session is disrupted and suppressed by the censoring module.

4.2 Post-April 2013

In May 2013 we serendipitously noticed a new warning page for YouTube. As a consequence, we re-conducted all tests. The results of which are summarized in Table 4. The number of websites affected by DNS, IP, and URL-keyword based filtering remained largely the same. In contrast, websites which experienced HTTP 302 redirection are now displaying a different warning page (shown in Figure 3) without any explicit redirection. Packet level inspection revealed that the legitimate HTTP response is being replaced with a spoofed response which displays the said page. Table 5 shows the sequence of HTTP packets for YouTube. Instead of the actual response, a packet with the content of the warning page and status code 200 is injected. The restriction system uses the first HTTP GET request as a censoring trigger based on the hostname and URI, similar to the mechanism discussed in the previous section. The status code of 200 makes the browser believe that it is a normal response, thus restricting it from fetching content from the intended destination (YouTube in this case). As a result, the actual TCP connection established with YouTube will eventually timeout for lack of activity. Figure 4 presents the HTTP response message received for all websites which experienced this category of censorship.
surf Safely!

This website is not accessible.

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

Figure 3: Censorship Warning Page, Post-April 2013

of censorship. The Last-Modified field of the response contains the value Fri, 19 Apr 2013 which we suspect is the date this new system came into effect. Moreover, all of our test networks experienced the same censorship mechanism and had the same warning page, except for Network4 which was still under the influence of ISP-level filtering. This has two main implications: (1) The country has moved from fragmented ISP-level to IXP-level filtering, and (2) The transition from the old system to the new one is taking place in phases.

4.3 Survey

Having worked out the censorship mechanisms, we conducted a controlled survey to ascertain the tools people are using to circumvent censorship in the country. Due to the sensitive nature of the subject matter we only shared it with personal contacts. Figure 5 presents the results for 67 respondents, mostly fellow computer scientists. Public VPN services are predominantly used; accounting for around 45%. Specifically, Hotspot Shield and Spotflux—both of which provide free VPN services through client applications—are popular choices. Web proxies (24%) and HTTP proxies (11%), such as Ultrasurf, also have a substantial user-base. It is noteworthy that the respondents are more technically aware than the average citizen and thus the results might be biased towards solutions that require above average computer skills.

5 Alternative Circumvention Mechanisms

In this section we discuss some alternative methods of bypassing restrictions. We first analyze web-based DNS lookup followed by content distribution networks and search engine caches.

5.1 Web-based DNS Lookup

A well-known mechanism to bypass DNS-based filtering is to use a public DNS service [19]. Our findings indicate that Pakistan also injects a fake DNS response for public DNS services, thus negating their use. Another means to resolve hostnames to IPs is to use an online
service. To check if this was a viable option to bypass filtering, we wrote a script to, (1) resolve hostnames to IPs using an online service\(^9\) and (2) use a custom HTTP header to fetch the webpage. The custom HTTP header is necessary as many multi-service sites share the same IPs across many of their services. For instance, Google uses the same set of IPs for YouTube, Drive, etc. Therefore, these resolved IPs cannot be used as drop-in replacements for hostnames. To remedy this, the HTTP header needs to contain the service name within its Host field. Our results show that all websites which experience DNS-based censorship, are also censored using secondary mechanisms (similar to South Korea \(^{19}\)): HTTP 302 redirection in case of the pre-April system and fake HTTP 200 response in case of the current post-April system. This also rules out the use of user-generated content \(^{6, 12}\) to host DNS records. Furthermore, according to Verkamp and Gupta \(^{19}\), South Korea makes use of DNS filtering for DNS entries that resolve to a single site. Whereas for IPs which are shared across several hostnames and the filtering needs to be selective, the HTTP-level mechanism is used. We found a similar pattern in our tests. For instance, multi-IP and multi-service sites such as YouTube and Wikipedia are only blocked at the HTTP-level while others are blocked at both levels.

5.2 Content Distribution Networks

CDNs have increased the availability, reliability, and performance of the web by geo-caching content. To check if any of the blocked websites were accessible through CDNs, we tried to fetch them from the peer-to-peer CoralCDN\(^9\). Websites are coralized by appending .nyud.net to the hostname. For instance, http://google.com.nyud.net is the coralized version of http://google.com. The original test script was modified to attempt to fetch each website from CoralCDN. The results indicate that all 307 websites are accessible through this method. Previous studies have already shown that CoralCDN is being used to bypass censorship in some areas \(^{10}\).

5.3 Search Engine Caches

Search engines usually cache snapshots of indexed pages. These can readily be accessed online. In case of Google, this requires pre-pending cache: to the URL. We checked the status of all 307 websites from our dataset and found out that all of them are accessible through the Google cache. Bing and Internet Archive\(^{11}\) also yielded similar results.

6 Conclusion and Future Work

We presented the first study of the cause, effect, and mechanism of Internet censorship in Pakistan. During the course of our work, in April 2013, the country underwent an upgrade from ISP-level blocking to a centralized system, allowing us to juxtapose two different generations of techniques. Our discovery of the upgrade to IXP-level filtering is consistent with the findings of a study conducted by the Citizen Lab \(^{18}\), coincidentally around the same time. To work out the specifics of restrictions, we used a publicly available list of 307 websites. For diversity and precision, 5 distinct networks from within the country were employed. Our results show that DNS injection is the predominant mechanism for blocking websites. This is applicable to both local DNS resolvers as well as public resolvers such as Google DNS and OpenDNS. The next line of censorship is at the HTTP-level. In case of the pre-April scheme, HTTP 302 redirection is used to disrupt and suppress sessions and in the post-April scheme, a fake HTTP 200 response packet is injected to give the browser the illusion that the session has been completed. In addition, websites blocked at the DNS-level also experience restrictions at the HTTP-level. Furthermore, the outcome from our controlled survey shows that public VPN services and web proxies are the two most popular tools to bypass restrictions. Finally, we showed that CDNs and search engine caches are simple but surprisingly unexplored means of accessing blocked content.

This work provided an initial window into the Internet censorship regime in Pakistan but a complete picture requires an expansion in the number of test websites as well as networks, which constitutes our future work. Furthermore, it is also not clear how the censoring module determines the exact NXDOMAIN redirector for public DNS resolvers, i.e. whether it maintains a list of all resolvers and their redirectors or it queries the actual resolver to obtain its redirector upon each lookup. Additionally, our future work also includes the examination of the side-effects of DNS injection on the same lines as \(^{1}\).

Acknowledgements

The author would like to thank the anonymous reviewers for their comments, concerns, and suggestions which greatly improved the quality of this publication.

\(^9\)http://ping.eu/nslookup/
\(^{10}\)http://archive.org

Table 5: HTTP Packet-level Trace for YouTube, Post-April

| No. | Source IP       | Destination IP | Payload                          |
|-----|-----------------|----------------|----------------------------------|
| 1.  | 192.168.43.112  | 173.194.35.100 | GET / HTTP/1.1 HTTP/1.1 200 OK  (text/html) |
| 2.  | 173.194.35.100  | 192.168.43.112 | GET /favicon.ico HTTP/1.1       |
| 3.  | 192.168.43.112  | 173.194.35.100 | HTTP/1.1 200 OK (image/x-icon)  |
| 4.  | 173.194.35.100  | 192.168.43.112 | HTTP/1.1 200 OK (image/x-icon)  |
References

[1] ANONYMOUS. The collateral damage of internet censorship by DNS injection. SIGCOMM Comput. Commun. Rev. 42, 3 (June 2012), 21–27.

[2] ATTAA, A. Government Regulates Jamming of Telecommunications Services Due to National Security. Online, 2010. http://propakistani.pk/2010/02/12/government-regulation-jamming-telecom-services-national-security/

[3] ATTAA, A. List of Blocked Websites in Pakistan [Blasphemy Crackdown]. Online, 2010. http://propakistani.pk/2010/05/23/list-of-blocked-websites-in-pakistan-blasphemy-crackdown/

[4] ATTAA, A. Google Analytics, Maps, Drive and Other Services Still Impacted by YouTube Ban. Online, 2012. http://propakistani.pk/2012/09/24/google-analytics-maps-drive-and-other-services-still-impacted-by-youtube-ban/

[5] BROWN, M. A. Pakistan hijacks YouTube. Rensys Blog, 2010. http://www.renys.com/blog/2008/02/pakistan-hijacks_youtube_1.shtml

[6] BURNETT, S., FEAMSTER, N., AND VEMPALA, S. Chipping away at censorship firewalls with user-generated content. In Proceedings of the 19th USENIX conference on Security (Berkeley, CA, USA, 2010), USENIX Security’10, USENIX Association.

[7] CLAYTON, R., MURDOCH, S. J., AND WATSON, R. N. M. Ignoring the great firewall of China. In Proceedings of the 6th international conference on Privacy Enhancing Technologies (Berlin, Heidelberg, 2006), PET ’06, Springer-Verlag, pp. 20–35.

[8] CRANDALL, J. R., ZINN, D., BYRD, M., BARR, E., AND EAST, R. Conceptdoppler: a weather tracker for internet censorship. In Proceedings of the 14th ACM conference on Computer and communications security (New York, NY, USA, 2007), CCS ’07, ACM, pp. 352–365.

[9] DAINOTTI, A., SQUARCELLA, C., ABEN, E., CLAFFY, K. C., CHIESA, M., RUSSO, M., AND PESCAPÉ, A. Analysis of country-wide internet outages caused by censorship. In Proceedings of the 2011 ACM SIGCOMM Internet measurement conference (New York, NY, USA, 2011), IMC ’11, ACM, pp. 1–18.

[10] FREEDMAN, M. J. Experiences with CoralCDN: a five-year operational view. In Proceedings of the 7th USENIX conference on Networked systems design and implementation (Berkeley, CA, USA, 2010), NSDI ’10, USENIX Association.

[11] INDEX ON CENSORSHIP. Pakistan: YouTube blocked over anti-Islam film. Online, 2012. http://www.indexoncensorship.org/2012/09/pakistan-youtube-censorship/

[12] INVERNIZZI, L., KRUEGEL, C., AND VIGNA, G. Message In A Bottle: Sailing Past Censorship. In Hot Topics in Privacy Enhancing Technologies (Vigo, Spain, 2012). http://petsymposium.org/2012/papers/hotpets12-3-message.pdf

[13] MARQUIS-BOIRE, M., MARCZAK, B., GUARNIERI, C., AND SCOTT-RAILTON, J. For Their Eyes Only: The Commercialization of Digital Spying. Citizen Lab Report, May 2013. https://citizenlab.org/storage/finsisher/final/fortheireyesonly.pdf

[14] MATHRANI, A., AND ALIPOUR, M. Website blocking across ten countries: A snapshot. In PACS (2010), AIScL, p. 152.

[15] NATIONAL ICT R&D FUND. Request for Proposal – National URL Filtering and Blocking System. Online, 2012. http://ictrdf.org.pk/RFP-%20URL%20Filtering%20%20Blocking.pdf

[16] OPENNET INITIATIVE. Access Contested – Pakistan. Online, 2012. https://opennet.net/research/profiles/pakistan

[17] SPAKIANAKIS, A., ATHANASOPOULOS, E., AND IOANNIDIS, S. CensMon: A Web Censorship Monitor. In Free and Open Communications on the Internet (San Francisco, CA, USA, 2011), USENIX Association.

[18] THE CITIZEN LAB. O Pakistan, We Stand on Guard for Thee: An Analysis of Canada-based Netwinder’s Role in Pakistan’s Censorship Regime. Tech. rep., University of Toronto, June 2013.

[19] VERKAMP, J.-P., AND GUPTA, M. Inferring Mechanics of Web Censorship Around the World. In Free and Open Communications on the Internet (Bellevue, WA, USA, 2012), USENIX Association.

[20] WROLD BANK. Internet users (per 100 people): Pakistan. World Development Indicators (WDI), 2012. http://data.worldbank.org/indicator/IT.NET.USER.P2

[21] XU, X., MAO, Z. M., AND HALDERMAN, J. A. Internet censorship in China: where does the filtering occur? In Proceedings of the 12th international conference on Passive and active measurement (Berlin, Heidelberg, 2011), PAM’11, Springer-Verlag, pp. 133–142.

[22] YOUGOV. Internet Usage in Pakistan. Online, 2011. http://propakistani.pk/wp-content/uploads/2011/11/YG_ig_web_pak.jpg