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

Sheharbano Khattak[1], Mobin Javed[2], Syed Ali Khayam[3], Zartash Afzal Uzmi[4], Vern Paxson[2,5]
[1]University of Cambridge, [2]UC Berkeley, [3]PLUMgrid, [4]LUMS SBASSE, [5]ICSI

ABSTRACT

Internet censorship artificially changes the dynamics of resource production and consumption, affecting a range of stakeholders that include end users, service providers, and content providers. We analyze two large-scale censorship events in Pakistan: blocking of pornographic content in 2011 and of YouTube in 2012. Using traffic datasets collected at home and SOHO networks before and after the censorship events, we: a) quantify the demand for blocked content, b) illuminate challenges encountered by service providers in implementing the censorship policies, c) investigate changes in user behavior (e.g., with respect to circumvention) after censorship, and d) assess benefits extracted by competing content providers of blocked content.

Categories and Subject Descriptors

C.2.3 [**Network Operations**]: Network monitoring; C.2.0 [**General**]: Security and protection; C.2.2 [**Network Protocols**]: Applications

General Terms

Measurement

Keywords

Censorship; ISP traffic; Content blocking; Video streaming; Porn;

1. INTRODUCTION

Nation-level censorship affects the activities of hundreds of millions of Internet users, with many countries implementing it at different levels and for a variety of reasons [35]. While censorship deployment and technology have seen considerable analysis in previous studies [9, 32], we lack a clear understanding of the *consequences* of censorship: just how does its employment affect different stakeholders? What steps do users, content providers, and ISPs take in response to censorship? How effectively does a given act of censorship achieve the censor's goals, and with what collateral damage?

IMC'14, November 5–7, 2014, Vancouver, BC, Canada.

ACM 978-1-4503-3213-2/14/11 ...$15.00.
http://dx.doi.org/10.1145/2663716.2663750.

While quantifiable answers to these sorts of questions have important socio-economic and policy implications, obtaining fine-grained illumination of these issues has remained largely unexplored due to the lack of datasets collected at appropriate vantage points and at appropriate times (i.e., before and after censorship events). Prior studies instead rely on datasets collected either via active probing [32] or using proxy servers [7]. These datasets cannot characterize individual behavior around censorship events; consequently, prior studies have focused mainly on understanding the mechanics of censorship technology and associated circumvention possibilities, rather than the consequences.

In this work, we seek to provide quantified insights into the impact of censorship on users, content providers, and ISPs, as seen through the lens of traffic datasets captured at a medium-size ISP[1] in a major city in Pakistan. Pakistan provides a useful vantage point for such a study as it recently instituted two large instances of censorship: blocking pornographic content[2] in 2011 [5], and blocking YouTube in 2012 [10].

The datasets we draw upon for this study comprise six residential and SOHO (Small Office / Home Office) traffic traces collected before and after the porn and YouTube censorship events, including one dataset collected on the day when Pakistan blocked YouTube. Trace durations range between 6–16 hours and capture sizes between 200–500GB, comprising traffic from 100–1,000 local IP addresses.[3] We supplement this perspective with a survey we conducted of about 700 Pakistani Internet users.

While analyzing the captured traffic traces, we developed methodologies to establish the ground truth (what was censored and how it was censored). Thus, unlike previous studies, our focus is on the offline analysis of captured data to ascertain the consequences of censorship. Our study develops the following insights:

- We observe a sudden increase in SSL traffic on the day of YouTube blocking, of sufficient volume that we attribute this change to users switching to VPN connections to circumvent the censors. This change persists a year later.
- SOHO users very quickly identified effective circumvention techniques, most of them switching to SSL within hours of the content blocking.
- Competing alternatives to YouTube received considerable benefit from censorship. We observe a sharp increase ($\approx$ 40–80%) in traffic volume towards these websites. Blocked sites

[1]Anonymized at the ISP's request.
[2]Shortened to "porn" henceforth.
[3]Due to NAT usage, one address can potentially correspond to multiple users.

also drop considerably in search-engine page rankings for local content. This change represents a marked regional shift in the economics of video content distributors, who mainly rely upon ad revenue.

- Before to censorship, porn content made up an average of 10% of home and SOHO traffic volume. Post-censorship, this fraction reduces considerably for both types of users. Even after factoring in traffic volume shifted to alternative (unblocked) porn websites, and the contemporaneous increase in SSL (potentially VPN) traffic, porn traffic volume did not return to the same level as before. The apparent reduction of active demand appears to indicate that the censors met their nominal goals.

- The YouTube block has two unintentional, yet significant, consequences: (*i*) *financial impact on ISP:* as users move to encryption-based circumvention mechanisms, the ISP's bandwidth requirement from the upstream provider increases, since ISPs cannot in general cache encrypted content; and (*ii*) *financial impact on YouTube:* user demand for YouTube (in terms of video requests observed in unencrypted traffic) eventually becomes half of its pre-censorship magnitude, and shifts to other video content providers.

2. BACKGROUND AND RELATED WORK

This section provides context on the censorship events that we investigate, the issue of determining how censorship is implemented, and the relationship between our work and prior research.

Internet infrastructure and censorship in Pakistan. Our study spans traces collected at a Pakistani ISP between 2011 and 2013—a timeline during which the country's censorship policy evolved considerably. There are ≈ 50 local and regional Internet service providers (ISPs) in Pakistan [25]. Only two of these, Pakistan Telecommunication Company Limited (PTCL) and Transworld Associates (TWA), have direct international connectivity, which they sell to the rest of the providers as well as directly to consumers; note that the majority of CDN servers are located outside of Pakistan. Internet censorship in Pakistan has mostly targeted content hosted outside the country, which Pakistani users access through PTCL or TWA.

The directives to block a particular website originate from the government or the judiciary. The ISPs are directed by the regulator, Pakistan Telecom Authority (PTA), to implement a content-blocking policy. While Pakistan has been intermittently blocking content since 2006 [32], a more persistent blocking policy was implemented in 2011 with the censorship of porn content [5], and then in 2012 with the blocking of YouTube [10]. The porn block in Pakistan was instituted in response to a media report that highlighted Pakistan as the top country in terms of searches for porn content [33], while the YouTube ban arose when a video, deemed blasphemous, appeared on the website. Presently, the country continues to block access to YouTube[4] as well as to sites deemed pornographic, anti-religious, or a general threat to national values and security [34]. A more recent study reports censorship of content related to human rights, independent media, proxy and circumvention tools, and bittorrent file-sharing sites [9].

Determining the implementation of censorship. Prior studies have focussed on inferring technologies for implementing censorship. Most of these studies use active probing to trigger censorship responses and comparing these responses to baseline responses in uncensored regions. These studies seek to detect the manipulation of traffic by intermediate devices [26, 19, 39, 23] and to illuminate the nature of censored content [11, 20] and the corresponding technology and/or mechanisms [14, 4, 43, 13]. Some recent studies apply probe-based approaches to study censorship particularly in Pakistan [32, 9], highlighting ISP-level DNS redirection, along with HTTP-redirection and fake-response injection at the national backbone. While previous work serves to help validate our findings, we cannot directly map it to our three-year dataset because censorship mechanisms can vary (*i*) over time, and (*ii*) across different vantage points. We thus employ passive analysis of each data trace to identify the censorship mechanism(s) in effect at a given time. We are not aware of any prior work that reconstructs censorship mechanisms by passive analysis of network traces, other than within the broader context of detecting forged TCP RST packets [44].

Consequences of Internet censorship. Previous literature has tackled the problem of how different network-level events, particularly in the context of anti-piracy laws, can affect the behavior of users [1] and content providers [22]. In the context of Internet censorship, studies have assessed the (sometimes unintended) impact of Internet censorship on global Internet services. China's injection of forged DNS responses has been reported to cause large scale collateral damage by blocking outside traffic that traverses Chinese links [3]. Upstream filtering can block traffic from outside a censored region due to ISP routing arrangements (for example, users of an ISP in Oman could not access certain content due to filtering regulations in India [27]). Chaabane et al. [7] analyze logs from Syrian censorship proxies to understand censorship methodology and user behaviour. In the latter context, they find that Syrian users employ web/socks proxies, Tor, VPNs, and BitTorrent to circumvent censorship. Labovitz uses a combination of large scale crawling and third-party data sources to investigate how the takedown of MegaUpload servers in North America impacts file-sharing traffic [28]. He finds that the incident caused a very small decrease in MegaUpload's previous traffic share, but makes content delivery inefficient, as files are now fetched from European servers over more expensive transatlantic links. For our purposes, a limitation of these latter studies is that from their own vantage points, they cannot assess the full exchange of traffic between users and providers, and thus cannot analyze the possibility of intermediate censorship (occurring closer to the user). Our study leverages an ISP viewpoint to address this issue and investigates the consequences of Internet censorship on users, content providers and operators. To our knowledge, this last perspective has not seen previous study.

3. DATA SOURCES FOR THE STUDY

Our primary data consists of six network traces captured at a Pakistani ISP[5] between 2011 and 2013. As discussed in Section 2, the government of Pakistan implemented two of the most significant and persistent policies in its censorship history during this period. Figure 1 illustrates the temporal relationship of the capture dates to the censorship events. The traces provide both pre-censorship and post-censorship snapshots of activity seen at an ISP for two major censorship events. We note that our data is not necessarily broadly representative, as it corresponds to just one ISP.

[4]In early 2014, a US court cited copyright issues in forcing YouTube to remove the offending video [36, 18]. As of this writing, Google continues to fight this order.

[5]The ISP requested to treat its name, location and other identifying information as confidential. The ISP originally acquired the data for unspecified purposes, and provided a degree of access out of good will.

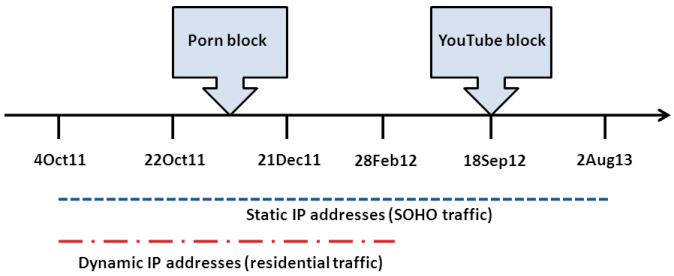

Figure 1: Temporal relationship of data to censorship events.

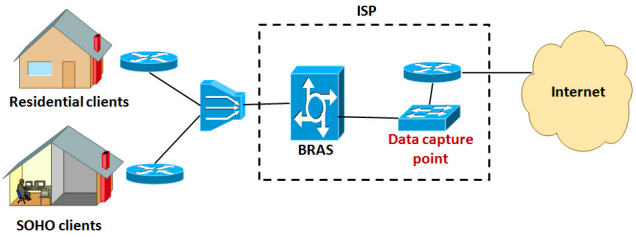

Figure 2: Capture location

Furthermore, it is difficult for us to estimate the actual user population of that ISP due to wide usage of NAT devices.

We supplement this data with a user survey we conducted in the region to explore user behavior post-YouTube censorship [45]. The survey results help shape the scope of our YouTube censorship analysis, and provide additional perspective for our findings from the primary data.

3.1 Capture Location and ISP Overview

Our tier-2 Pakistani ISP peers with a tier-1 provider through the Transworld Associates TWA-1 submarine telecommunications cable in Karachi. The ISP caters to both residential and SOHO customers. Due to our confidentiality agreement, we cannot provide details regarding the scale at which the ISP operates, the magnitude of its customer base, or the address space it uses.

Figure 2 shows the data capture location within the ISP premises. All customer lines terminate at one of several Broadband Remote Access Servers (BRASes) in the ISP's network. Each BRAS connects to the ISP's core Internet-facing router through a switch. The ISP gathered the traces at the BRAS-facing side of this switch. This vantage point captures all of the local ISP-generated traffic (such as redirected DNS traffic) in addition to bi-directional traffic going in or out of the ISP's premises. The ISP assigns each BRAS to a set of addresses. While the allocation remains unchanged for any given trace, across traces the captures correspond to potentially different subsets of the ISP's address space.

Address Pools. The ISP splits its address space into dynamic DHCP and static pools, primarily assigning dynamic IPs to residential customers. The ISP reserves some static IP addresses for hosting its services, such as DNS resolvers, mail and authentication servers, and other web resources. It allocates the bulk of the remainder to SOHO customers. We do not know which IP addresses correspond to particular ISP services; in particular, we have no specifics regarding its censorship apparatus, related IP addresses, or blacklists in effect for different traces. Note that the ISP does not allocate IPv6 addresses to its customers. While we find some IPv6 communication taking place over tunnels, its overall volume is negligible.

Block Key	Trace	Day	Capture Hour (PKT) + Duration	Size (GB)	Active Local IPs
–	03Oct11	Tue	17:48 + 15h14m	222	1,075
–	22Oct11	Sat	18:49 + 20h42m	460	1,046
P	21Dec11	Wed	22:17 + 16h54m	286	868
P	28Feb12	Tue	18:48 + 11h08m	200	974
PY	18Sep12	Tue	08:54 + 07h19m	500	310
PY	02Aug13	Fri	09:40 + 06h00m	207	136

Table 1: Summary of packet captures. P=Porn, Y=YouTube

Ethical Standards. The authors of this work with direct access to the data signed a contract highlighting the obligations to (*i*) respect user privacy, (*ii*) not share data with third parties (which includes the other co-authors), (*iii*) not move data outside Pakistan, (*iv*) not move data within Pakistan without prior consent, and (*v*) undertake an objective study and refrain from maligning any party involved in the censorship landscape (user, ISP, or government). These restrictions did not affect our study in any respect.

3.2 Data Description

Table 1 characterizes the traces. Capture durations range between 6–16 hours and capture sizes between 200–500GB, comprising traffic from 100–1,000 IP addresses. Some of these IP addresses likely reflect NATs, and thus the effective user population could potentially range larger. Due to a number of variables in the traces, a limitation of our work is that we cannot exclusively attribute cross-trace trends to the consequences of censorship; these might instead arise due to factors introduced by disparate capture days and/or timings. However, some of our results are sharp enough that it appears very likely that they correspond to responses to censorship. Finally, trace characteristics might differ between traces despite similar trace durations and time frames because the IP address prefixes allocated to a BRAS do not necessarily remain consistent across traces.

Protocol Logs. Our analysis relies on protocol logs generated using `Bro` [6]. In particular, we deal with connection, HTTP and DNS logs. The connection log contains one entry per flow, while the protocol logs contain separate entries for each request-response pair. We use *number of connections* to refer to distinct transport-layer flows and *number of requests* for individual request-response pairs as observed in the protocol logs.

3.3 Data Sanitization and Characterization

For soundness of analysis, we first identify measurement ambiguities/inaccuracies (*sanitization*) and then label the data (*characterization*) so as to extract the portion relevant to a given analysis.

Data can include inaccuracies due to various reasons, where a particularly common is the limitation of the capturing device or analysis tool. A large portion of our data reflects connections that did not fully establish (e.g., scanning activity). We remove such connections from the bulk of our analysis (though we include them when assessing evidence of user attempts to access blocked content). We identify unestablished connections based on `Bro`'s connection state field, which captures the sequence of control and data packets seen for a given connection. We only include connections whose state reflects a completed three-way TCP establishment handshake, which reduces our six data sets down to roughly half the original number of connections, as reflected in Table 2.

To characterize our data, we label flows based on the connection direction and the type of local addresses involved, subsequently using only those subsets apt for a particular type of analysis. For

Block Key	Trace	Total conns.	After sanitization (% retained)	Transit	Local	Static IPs		Dynamic IPs	
						Inbound	Outbound	Inbound	Outbound
–	03Oct11	11.53M	5.39M (46.7%)	0.03M	0.68M	1.05M	1.62M	0.54M	1.48M
–	22Oct11	29.19M	12.68M (43.4%)	0.03M	1.24M	3.58M	4.13M	1.25M	2.44M
P	21Dec11	16.06M	8.09M (50.4%)	0.02M	1.21M	1.37M	2.57M	0.50M	2.42M
P	28Feb12	12.12M	5.84M (48.2%)	0.04M	0.59M	0.86M	0.98M	0.99M	2.39M
PY	18Sep12	24.01M	14.93M (62.2%)	0.02M	2.19M	1.13M	11.59M	-	-
PY	02Aug13	8.79M	3.77M (42.9%)	0.01M	0.53M	0.19M	3.03M	-	-

Table 2: Breakdown of data sanitization and characterization. P=Porn, Y=YouTube

example, to gauge the interest of local users in Web content served over the Internet, we only consider outbound connections—see Table 2 for an overview.

With regards to direction, we label a connection as *local* if it has both source and destination IP addresses in the ISP's network block, or *transit* if neither source nor destination belongs to the local network. We consider a connection *inbound* if its originator resides outside the ISP's network and *outbound* if the converse holds.

We further differentiate between residential and SOHO traffic based on the assumption that nearly all static IP addresses correspond to SOHO users, where we determine the set of static IP addresses using information provided by the ISP.

Table 2 summarizes these characterizations. We find that outbound connections predominate, followed by inbound, local and transit, in that order. The small portion of transit traffic agrees with a communication from the ISP that we should expect a small amount of traffic from a sister ISP, along with some IPv6 test traffic.

3.4 Final Datasets

Table 3 summarizes our filtered dataset. We divide our six traces into two datasets corresponding to residential and SOHO users respectively. We use both datasets to study the impact of porn censorship, and the SOHO dataset for YouTube censorship, since the post-YouTube censorship traces contain only small amounts of residential traffic. For a major part of our study, we work with HTTP and DNS logs for outbound and local connections. We include local traffic in our analysis because we expect to find a portion of user traffic redirected to local systems enforcing censorship.

3.5 User Survey

Finally, we carried out an online user survey targeting Pakistani users in order to develop an understanding of their views of and responses to the YouTube block. (We avoided asking about the porn block as it is a culturally sensitive topic in the region.) We disseminated information about the opportunity to take the survey through mailing lists and classroom discussions. The survey asked about: (i) the popularity of blocked content and new players that emerged post-censorship, (ii) user inclinations to circumvent and the corresponding mechanisms, (iii) collateral damage experienced due to the block, and (iv) opinions about Internet censorship in general.

We did not expect many responses because it is hard to get users to respond to surveys without any incentive, especially when the topic is a sensitive one such as Internet censorship. Surprisingly, we received 770 responses (75% male, and 25% female), reflecting a widespread eagerness to comment on the subject. 94% of the responders were young/middle-aged (25-40 years), and resided in major cities, with occupations suggestive of high levels of technological competence. This demographic does not reflect Pakistan's makeup as a whole, and likely skews towards particularly informed and active users. Thus, we do not frame its results as representative, but rather as illuminating of some of the facets of how censorship affects Pakistani users.

4. ESTABLISHING GROUND TRUTH

A significant challenge for our study is that we use historical data for which we lack key contextual information: (*i*) *what* was censored (the blacklist for the porn block), and (*ii*) *how* it was censored (the mechanism of censorship). In this section we discuss the methodology we employed to answer these questions based solely on the information present in the available traces. We do so by analyzing the responses we see from servers in reply to user requests, basing our deductions on the observation that for enforcing censorship, a censor either silently drops requests or sends back false response packets.

4.1 Censorship Indicators

A censor can block HTTP content at any of the layers involved in facilitating an HTTP transaction: DNS, TCP/IP, and HTTP. Across these layers, the censor has an array of choices for how to block, each leaving a trail in the network traces. The presence of such a trail (a sequence of packets not necessarily contiguous, or an absence of expected packets) provides an indicator of censorship. However, some of these indicators can occur in an uncensored environment for legitimate reasons such as measurement loss or excessive server load. We deem censorship indicators that can also occur under uncensored conditions as *ambiguous*, and deal with them as follows: (*i*) If the censored content is *known*, we attribute a high frequency of an ambiguous indicator to censorship (and leverage this information to establish the mechanism of censorship); but (*ii*) if the censored content is *unknown* (that is, we cannot tie any given flow to an attempt to access blocked content), we cannot attribute the occurrence of such an indicator over a short observation window (less than one day for each of our traces) exclusively to censorship. We therefore do not leverage these latter indicators, and rely only on unambiguous indicators to establish (partial) ground truth.

We now discuss assessing censorship indicators at each layer.

DNS Based Censorship. At the DNS level, a censor-controlled resolver (such as one maintained by the present ISP) can effect blocking behavior by sending: (*i*) No Response, (*ii*) False Error (such as NXDOMAIN), or (*iii*) False Response (the RCODE for these responses is NO ERROR). (Clearly, users can bypass these DNS-based censorship mechanisms by using an independent DNS resolver.)

No Response provides an ambiguous indicator because it could occur due to excessive load on the resolver, or network problems. Thus, we do not attribute this scenario to censorship when the censored content is unknown. However, for known censored content, observing a consistent behavior of no response is a strong indicator of censorship.

Block Key	Trace	Active IPs	Conns.	TCP Conns.	UDP Conns.	HTTP Transactions	SSL Conns.	DNS Conns.	Bytes (GB)	Packets
SOHO Traffic (Static IPs)										
–	03Oct11	585	2.02M	1.00M	1.02M	1.44M	0.05M	1.29M	79	119M
–	22Oct11	554	4.84M	1.91M	2.93M	2.18M	0.09M	1.90M	180	276M
P	21Dec11	570	3.24M	1.70M	1.55M	2.52M	0.14M	2.63M	121	182M
P	28Feb12	298	1.16M	0.51M	0.65M	0.62M	0.08M	0.33M	39	61M
PY	18Sep12	298	13.78M	7.53M	6.25M	7.16M	1.05M	4.26M	271	546M
PY	02Aug13	133	3.56M	1.85M	1.71M	1.78M	0.32M	1.57M	143	246M
Residential Traffic (Dynamic IPs)										
–	03Oct11	490	1.76M	0.85M	0.9M	1.14M	0.05M	1.86M	85	149M
–	22Oct11	492	2.97M	1.40M	1.57M	1.84M	0.08M	1.08M	163	237M
P	21Dec11	451	2.96M	1.50M	1.45M	2.11M	0.13M	1.09M	103	176M
P	28Feb12	676	2.80M	1.26M	1.55M	1.46M	0.11M	0.80M	112	176M
PY	18Sep12	-	-	-	-	-	-	-	-	-
PY	02Aug13	-	-	-	-	-	-	-	-	-

Table 3: Final data after preprocessing. P=Porn, Y=YouTube

For the last two cases, we can leverage two public databases to establish the ground truth: (*i*) `dnsdb`, which contains historical information on name-to-IP address mappings [38], and (*ii*) Team Cymru's IP-to-ASN mappings database [40]. We identify false responses as follows:

False Error: We mark the queries that consistently receive an ERROR RCODE response from a resolver for a subsequent `dnsdb` lookup. If there exists a name-to-IP mapping in the database for the domain seen in the trace, we conclude that the censor employed False Error as their mechanism.

False Response: We can detect a DNS resolver including false IP addresses in its responses if we observe consistency in the false answers returned. We identify whether a DNS resolver answers with an IP address belonging to an ISP within the country (either local ISP or an upstream transit provider from within Pakistan) when the domain is actually hosted elsewhere. Let $\text{ASN}_{\text{trace}}$ be the ASN of an IP address returned in a DNS reply recorded in the trace, and ASN_{real} be the ASN for the IP address received in a DNS reply obtained by active testing. If $\text{ASN}_{\text{trace}}$ belongs to an ISP within Pakistan, while ASN_{real} does not, the query received a false response. This technique has the limitation that we cannot detect cases where the censor's redirection points to an IP address that belongs to an AS outside the country. The same problem holds for a censor who employs null-routing. Furthermore, this technique will flag caching servers employed within the country.[6]

TCP/IP blocking. IP-level blocking is an ambiguous indicator of censorship, as it is hard to distinguish from legitimate causes of inaccessibility. However, we can weed out some of the non-censorship cases because censorship generally requires that *all* attempts to establish a connection to a blocked address will fail. To find such IP addresses, we use the heuristic described below.

First, iterate over all the `A` records for queries resolved correctly, and for each connection seen for one of these addresses, label it according to the following three connection states:

- **PARTIAL**: No SYN seen from the connection originator, but packets seen from the responder.
- **EST**: Full TCP establishment handshake observed.
- **BLOCKED**: The originator sent a SYN but either (*i*) receives no response, or (*ii*) receives a TCP RST (potentially injected by the censor).

[6]Indeed we identify the ISP's caching machines using this methodology.

We flag IP addresses for which we never observe EST and for which we observe BLOCKED at least once. We map these IP addresses back to their corresponding domain names in the DNS logs and consider these domains as potentially censored.

HTTP Level Blocking. At the HTTP level, a censor can block via: (i) No HTTP response (for example, by injecting a RST after connection establishment), (ii) return an HTTP-level error response code, or (iii) return a false response such as a block page (either directly or via HTTP-level redirection). We assess these as follows:

- **No Response**: This can occur for legitimate reasons. We do not attribute it to censorship when the censored content is unknown. For known censored content, consistently observing TCP-layer blocking can confirm censorship (for example, the case where the responder always sends a RST in response to an HTTP request).
- **Error Response Codes**: This error can be ambiguous because a client can receive such responses due to resources legitimately not found or forbidden. For known censored content, however, this provides a strong indicator if it is the dominant behavior.
- **Block Page such as via 3XX redirection**: The censor may redirect diverse domains/sub-domains to the same *Location*. We can detect this mechanism by analyzing histograms of the `Location` header in responses. If, however, the censor redirects attempted access for different content to distinct locations (such as by incorporating the request URI into the redirect location), the histogram will not reveal any common redirection target. In this case, the analysis in the next item might reveal blocking.
- **Block Page via 2XX response**: We can detect when the censor sends the same block page for multiple URLs by fingerprinting block pages known to be associated with the censor, or by looking for potential candidates by investigating spikes/modes in a histogram of the number of bytes sent in server reply items.

4.2 Identifying YouTube Censorship

Table 4 shows our findings for deducing the mechanism of YouTube censorship in Pakistan at the two different points in time that are currently under study.

First, we observe DNS redirection in both traces for queries resolved using the ISP's DNS resolvers: all YouTube queries received

	Trace	DNS	IP	HTTP
YouTube	18Sep12	DNS_REDIR	—	HTTP_REDIR
	02Aug13	DNS_REDIR	—	HTTP_NORESP
Porn	21Dec11	DNS_REDIR	—	—
	28Feb12	DNS_REDIR	—	—
	18Sep12	DNS_REDIR	IP_BLOCK	—
	02Aug13	DNS_REDIR	—	HTTP_NORESP

Table 4: Censorship mechanisms for YouTube and porn blocking as observed in our post-censorship traces. "—" indicates that we did not find any concrete evidence of the given mechanism.

Trace	DNS	IP blocking	HTTP blocking
21Dec11	226	3 / 0%	2 / 0%
28Feb12	145	7 / 0%	1 / 0%
18Sep12	105	56 / 41%	6 / 0%
02Aug13	100	0 / 0%	8 / 62%

Table 5: Number of porn domains potentially blocked at each layer. For IP and HTTP blocking, we also show the percentage overlap with DNS blocking. HTTP blocking when present took the form of consistent No Response conditions.

replies with a single ISP-owned address. Queries sent to non-ISP resolvers obtained correct answers. Second, we do not find IP blocking in either trace. We see only one potentially blocked address in 18Sep12, which reverse-maps to a YouTube content server.

Finally, in addition to DNS-based blocking, we also observe HTTP-level blocking in both traces. In 18Sep12, we find blocking of YouTube via 3XX redirection to an IP owned by a large local provider, one of the two with direct international connectivity. In 02Aug13, the blocking shifted from redirection to No Response. In traces before 02Aug13 (including pre-YouTube-censorship traces), the number of YouTube HTTP requests that received no response averaged $\approx 2\%$, whereas in 02Aug13 this number jumps to $\approx 95\%$, with nearly all of these reset by the responder.

These observations confirm the two-layered censorship mechanism for YouTube described by a prior study [32], i.e., ISPs block locally using DNS redirection, and the two large providers in the country with direct international connectivity (PTCL and TWA) employ HTTP-level blocking.

4.3 Identifying Porn Censorship

To accurately identify porn censorship, we characterize all websites recorded in our traces using McAfee's URL categorization service [30] and extract the ones it labels as *Pornography*. We spot-checked a random sample of its decisions (both positive and negative) to confirm its apparent accuracy and lack of any regional lacuna. We did not find any errors.

In recovering the censor's *porn blacklist*, we worked on each trace in turn. Recovered blacklists necessarily represent only a fraction of the censor's true blacklist, since our data-driven approach can only identify the fraction of censored content present in our traces.

We aim to recover to blacklists at the granularity of *registered* domains, per the master list kept by Mozilla [31], since later in our analysis we use that granularity for characterizing traffic to blacklisted domains. Where applicable, we mark domains in our blacklist as partially blocked. We consider the possibility of partial blocking of a domain only at IP-level (due to incomplete IP address coverage) and HTTP-level (due to incomplete regex coverage). For DNS, we assume that the true blacklist contains domains at the granularity of *registered* domains. Hence, we only add a domain to our blacklist if we observe consistent blocking behavior for all of its subdomains that appear in the respective trace.

Table 4 summarizes the mechanisms used for censorship for the 4 post-porn censorship traces. Table 5 shows the corresponding development of blacklist. We observe the following:

- We find evidence of DNS redirection in all four traces. All ISP resolvers consistently redirect blocked queries to the same ISP-owned address (the same address as used for YouTube censorship). However, non-ISP resolvers resolve the blocked content queries correctly, indicating that the censor does not employ DNS injection such as discussed by Duan et al. [17]. Table 5 lists the number of blacklisted domains per trace that we recover using this indicator.

- We also observe IP blocking for some porn domains. Since this is an ambiguous indicator, we check for any overlap of potential censored domains we find using IP blocking with those found via DNS blocking. We find significant evidence of IP based blocking only in 18Sep12, with a 41% overlap with our DNS blacklist. The TCP state of these connections indicates that the originator never received any response packet from the responder, consistent with blackholing.

- We did not find any instances of users receiving an HTTP block-page either through injection or redirection. Some domains consistently receive no response, but with negligible overlap with our DNS blacklist, except for the last trace, as shown in Table 5. The TCP states of these connections reveal that in a high percentage of the cases, the responder terminated the connection by sending a RST, indicative of likely censorship.

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. Accordingly, we do not include these ambiguous domains in our blacklist reconstruction. Doing so omits only a handful of potentially blocked domains.

Table 6 illustrates how the porn blacklist evolved over time. We can in addition consider the question of whether the censoring authority acts in a reactive fashion; that is, do they block porn domains that begin to gain popularity with users? In pre-block traces (03Oct11 and 22Oct11), we see 1,313 unique porn domains. We find that 8.2% of these domains were blocked in 21Dec11. After the initial dissemination of the blacklist in 21Dec11, we see a lull in its updating; we observe only ≈ 35 new domains added in each of 28Feb12 and 18Sep12, and no new ones in the last trace. Moreover, the "new" blocked domains have little overlap with porn domains previously observed and unblocked—reinforcing information unofficially shared with us by local operators that the central regulator disseminates blacklists to ISPs and that their development is independent of the porn browsing trends of users. (Section 2).

Porn domains	Oct11	Dec11	Feb12	Sep12	Aug13
Unblocked	1,313	1,181	1,609	2,210	2,352
Blocked	0	226	145	161	105
New entries	—	0	37	36	0
% overlap	—	8.2%	0.2%	0.5%	0%

Table 6: Evolution of porn blacklist as seen in our traces. % overlap corresponds to the proportion of new entries present (and unblocked) in all previous traces. Oct11 reflects two traces captured during that month.

Summary. Table 4 summarizes our findings on the mechanism of censorship for YouTube and porn as seen in different traces. We find blocking of both YouTube and porn at the DNS-level using redirection in all of the respective post-block traces. In addition, in 18Sep12 we find YouTube blocked using HTTP redirection, and porn using IP blocking, and in 02Aug13 both are blocked using RST injection.

5. METRICS RELEVANT TO CONTENT PROVIDERS

In this section we discuss two key aspects for our study: (i) what constitutes a "content provider" relative to each censorship event, and (ii) the metrics on which we base our assessment of changes resulting from censorship events (note that we can only apply these metrics to unencrypted traffic).

Censorship events affect both primary and alternate providers of the censored content. For the YouTube event, these relate to the general category of *Video Content Platforms*, for which we focus our analysis on four major players: YouTube, DailyMotion, Tune.pk and Vimeo. These constitute the primary video providers for Pakistan as based on their market share [2] and the results of our user survey [45]. For the porn censorship event, we consider all porn domains seen in our traces as identified by McAfee's URL categorization service in April 2014. Given that our most recent trace was captured in August 2013, some domains might have been inaccurately classified.

The primary metric that we employ is *downstream* traffic (server response bytes) served by blocked and alternate content providers, which we will often abbreviate as "bandwidth" for shorthand. We base this choice on the observation that both the censored categories, video and porn, make heavy use of network downloads—what is censored primarily constitutes images and videos. For these categories, downstream bandwidth reasonably captures the degree of user interest in a content provider. This metric also allows us to readily study shifts in traffic trends in the presence of encryption technologies—a potential response to broad category-based censorship.

In addition, for the video category we assess changes in content *embedded* in other sites in response to censorship.[7] This metric captures the broader ecosystem for users viewing videos sometimes in response to other websites that embed a content provider's videos. (Porn content, on the other hand, is presumably only embedded on other porn sites.) After censorship of a content provider, local websites lack an incentive to embed the provider's videos.

We now discuss computing these two metrics:

(i) Direct vs. embedded video viewing requests: To distinguish between these two types of requests, we need to develop *signatures* that classify a given URL as one or the other (or neither) of these. One approach for developing signatures is to analyze traffic dumps collected by actively downloading video content [24]. However, given we collected our traces over a span of three years, we cannot employ an active approach like this, as signatures can change over the years. To develop signatures that can span our datasets, we use a data-driven methodology: for each video content platform, we examine a histogram of its URI root prefixes and associate them with distinct classes of web content based on inspecting the corresponding content type observed in traffic captures, and in some cases entering the full URL into a browser to see if the video plays. This approach provides us with fingerprints for both *direct* and *embedded* viewing request URLs for each video content platform. We note that direct and embedded video watching requests have a consistent signature across traces, perhaps because these span the same domain.

(ii) Bandwidth per content provider: We could compute downstream bandwidth by accumulating server bytes for all HTTP requests where the content provider domain appears in the `Host` header. However, this approach risks missing traffic because: (*i*) content can be served by CDNs (often the case for videos and images), the domain name of which may have no relationship to the host corresponding to the original video/image request, and (*ii*) CDNs typically serve content on behalf of multiple domains, making it infeasible to exclusively associate a given CDN domain with a specific origin server. We might consider accounting for such traffic by accumulating all response bytes for requests where the content provider appears in `Referer`, but doing so will: (i) include bytes belonging to other websites/providers, since the `Referer` might instead reflect the user clicking on a link on the original content provider page that leads to a *different* content provider's page, and (ii) miss bytes belonging to the content provider in cases where an automatic chain of requests traverses multiple domains in order to ultimately reach the CDN.

Putting the above considerations together, we employ the following approaches for estimating traffic volume:

- *Video Content Platforms:* Because these analyses concern just a handful of content providers, for the video category it remains practical to develop URI signatures for each of the four major players. Note that these signatures are different than the *direct* and *embedded* watch signatures, because the video is in general fetched from a URL different than that of the watch page. Along with analysis of active fetches, we analyze all HTTP requests where either `Host` or `Referer` contains the content provider's domain, and the `Content-Type` header in the response is either `application/octet-stream` or contains the keyword "`video`". It follows that we miss video downloads for content providers whose domain name appears in neither the `Host` nor the `Referer` part of an HTTP request. We find that YouTube transfers video using both `video` and `application/octet-stream`. The other three providers, however, only transfer video using a `video` content-type (and sometimes employ `application/octet-stream` for content such as CSS and fonts).
- *Porn Providers:* Accurately attributing porn bandwidth requires a more generic approach, since there are too many providers ($\approx$ 3,800 seen in our traces) to allow us to craft individual signatures. Since a porn site can embed content from other porn sites, when we see a transfer for which both the `Host` domain and the `Referer` domain are labeled as porn in our dataset, we give priority to the former. Specifically, we use the rule: if `Host` has porn domain X, add the corresponding bytes to X; else if `Referer` exists and has a porn domain Y, add corresponding bytes to Y. Otherwise, do not attribute the transfer to any domain.

6. CHANGES IN USER BEHAVIOR

The intrusive nature of Internet censorship will naturally lead to some users altering their behavior in its aftermath. In this section we quantify several perspectives regarding user demand for

[7] Note that we treat links to a content provider's page returned in search results as a form of direct access, rather than "embedded" access, because we presume that often users navigate to such pages via search engines.

Key	Trace	Total	YouTube(%)	Others (%)	Breakdown of Others		
					DailyMotion (%)	Tune.pk (%)	Vimeo (%)
(a) Video Bandwidth (GB)							
–	03Oct11	26.5 GB	97.9	2.1	2.0	0.0	0.1
–	22Oct11	56.6 GB	97.6	2.4	2.4	0.0	≈ 0.0
P	21Dec11	45.2 GB	98.5	1.5	1.3	0.0	0.2
P	28Feb12	12.6 GB	96.9	3.0	3.0	0.0	≈ 0.0
PY	18Sep12	10.7 GB	15.8	84.2	82.0	0.0	2.2
PY	02Aug13	2.7 GB	0.0	100.0	40.9	57.6	1.5
(b) Number of Direct Watch Requests							
–	03Oct11	2,199	99.5	0.5	0.2	0.0	0.2
–	22Oct11	4,550	99.0	1.0	0.9	0.0	0.0
P	21Dec11	3,254	99.3	0.7	0.6	0.0	0.1
P	28Feb12	878	95.7	4.3	4.3	0.0	0.0
PY	18Sep12	992	71.1	28.9	23.2	0.0	5.7
PY	02Aug13	169	46.1	53.8	37.3	14.2	2.4
(c) Number of Embedded Watch Requests							
–	03Oct11	200	87.0	13.0	10.0	0.0	3.0
–	22Oct11	299	78.9	21.1	14.7	0.0	6.4
P	21Dec11	414	92.5	7.5	2.7	0.0	4.8
P	28Feb12	209	86.1	13.9	11.5	0.0	2.4
PY	18Sep12	2,037	73.0	27.0	19.5	0.0	7.5
PY	02Aug13	647	51.8	48.2	32.6	10.7	4.9

Table 8: Distribution of video bandwidth, number of direct and embedded watch requests across major video content providers over time.

Key	Trace	HTTP GB	% Porn	% Video	HTTP:SSL
(a) SOHO Traffic					
–	03Oct11	58.15	11.5	45.5	40.72
–	22Oct11	105.79	11.6	53.6	38.19
P	21Dec11	90.05	3.7	50.2	23.72
P	28Feb12	23.37	2.0	54.3	17.77
PY	18Sep12	91.60	3.0	11.7	3.20
PY	02Aug13	49.66	3.8	5.5	3.25
(b) Residential Traffic					
–	03Oct11	52.10	9.4	—	20.05
–	22Oct11	100.04	7.4	—	50.30
P	21Dec11	66.70	4.0	—	18.22
P	28Feb12	66.23	3.5	—	14.33

Table 7: Ratio of porn and general video traffic to total HTTP byte volume. The last column shows the ratio of HTTP volume to TLS/SSL volume. "—" indicates datapoint not considered in our study (we only use SOHO traffic for analyzing the YouTube block). P=Porn, Y=YouTube

blocked content before censorship, and their persistence and approaches in accessing blocked content after censorship comes into place. While we cannot rule out other factors leading to some of the changes we have observed, the broad scope of the censorship events we consider makes it quite likely that our observations indeed reflect responses to censorship.

6.1 Changes in Traffic

For video traffic, we observe in Table 7(a) that on average video traffic comprised 50% of HTTP traffic before the YouTube block, consistent with global trends (videos comprised 57% of user-generated traffic in 2012 [8]). The overall (unencrypted) video consumption rate drastically declines after the YouTube block, subsequently comprising only 12% of total HTTP traffic in 18Sep12, and declining further to 5.5% in 02Aug13. The decline in video traffic coincides with a decrease of nearly 90% in the HTTP to SSL[8] ratio in 18Sep12, corresponding to the YouTube block day. The ratio remained fairly consistent on this day as viewed hour-to-hour (on average ≈ 3.25), indicating that SOHO users quickly switched to SSL-based circumvention technologies. The trace for this day does not reflect a clear learning phase, suggesting such had already occurred by the time the capture began. The overall trend for SSL traffic remained consistent 11 months later in 02Aug13. This steep increase in SSL traffic post-YouTube-block highlights that most users likely use encrypted tunnels to watch video content after the block. As we note below, the SSL traffic heavily correlates with the use of proxy services, suggesting that it indeed arises due to employment of circumvention measures. Our user study substantiates this conjecture: 57% of the survey participants state they used SSL-based VPN software (UltraSurf, OpenVPN, Hotspot Shield) to access YouTube content.

If we look at direct video requests (either via user navigation, or mediated by clicking on search results), per Table 8(b) we find that the vast majority of direct video requests prior to the block correspond to YouTube (average 98%). Immediately after block (18Sep12), YouTube still receives the highest portion (though reduced by 27%) of direct requests, but the proportion sharply drops 11 months later in 02Aug13 to 46%, with users dispersing the rest of the requests among alternate providers. The decrease in direct YouTube video requests matches our survey results: 40% of respondents mentioned that they do not bother to click on YouTube links due to the blocking; 39% will access the link using a circumvention mechanism; while 17% access the video via an alternate provider.

Table 7 shows that before the blocking, the average porn bandwidth ranged from 8.4–11.5% for residential and SOHO users, respectively. These numbers lie below global estimates that porn comprises 30% of Internet traffic [21]. That we find more porn bandwidth consumed at SOHOs than at homes likely occurs because of higher bandwidths available in SOHO networks. After

[8]We cannot conclusively say if the SSL traffic corresponds to VPNs or HTTPS.

Resolver ASN shorthand (% of DNS queries)					
–	–	P	P	PY	PY
03Oct11	22Oct11	21Dec11	28Feb12	18Sep12	02Aug13
SOHO Traffic					
39,248	64,269	43,655	10,062	13,025	5,036
Local-ISP (99.89)	Local-ISP (99.58)	Local-ISP (98.23)	Local-ISP (91.93)	Local-ISP (68.75)	Local-ISP (74.12)
Google (0.06)	Google (0.28)	Google (1.46)	Google (5.63)	Google (13.69)	Google (19.32)
LEVEL3 (0.04)	LEVEL3 (0.08)	LEVEL3 (0.12)	VPLSNET (1.37)	LEVEL3 (6.96)	LEVEL3 (2.88)
PKTELECOM-AS-PK (0.01)	IPC Computing (0.03)	VPLSNET (0.10)	HINET (0.96)	SPEEDCAST (5.51)	MULTINET (2.70)
VeriSign (0.01)	DIEGOGARCIA (0.02)	HINET (0.08)	OpenDNS (0.11)	OpenDNS (5.08)	Verizon (0.99)
Residential Traffic					
12,739	15,821	6,767	5,451	—	—
Local-ISP (95.50)	Local-ISP (93.89)	Local-ISP (93.08)	Local-ISP (92.20)	—	—
ASVPSHOSTING (3.73)	OpenDNS (5.49)	Google (6.18)	ASVPSHOSTING (4.59)	—	—
Google (0.69)	ASVPSHOSTING (0.49)	ASVPSHOSTING (0.62)	Google (3.05)	—	-
CELCOMNET (0.06)	LEVEL3 (0.11)	LEVEL3 (0.09)	LEVEL3 (0.13)	—	—
OpenDNS (0.02)	TIGGEE (0.01)	OpenDNS (0.03)	OpenDNS (0.04)	—	—

Table 9: Distribution of DNS A/AAAA queries for blocked categories across top 5 DNS resolvers.

the block, the residential porn bandwidth falls by more than half (averaging 3.7% of HTTP bandwidth post-block).

For SOHO traffic, the average porn bandwidth reduces by a factor of three. In contrast to video, we do not observe a significant increase in the HTTP-to-SSL traffic ratio in response to porn censorship, indicating a subset of users either stopped watching porn or shifted to alternate porn providers.[9].

6.2 Impact on User Behavior

Censorship can potentially modify the network behavior of users or result in new behavior driven by trying to access blocked content. In this section we assess this possibility by examining the usage of DNS resolvers and web proxies over time, along with a look at user browsing activities immediately after encountering the block page.

DNS resolvers. For both censorship events we study, the censor employed DNS redirection by local ISP resolvers as the primary means of censorship enforcement. Circumvention to counter this step only requires using an alternate DNS resolver. Table 9 illuminates the degree to which users pursued this option by examining the top-5 DNS resolvers used to resolve DNS A or AAAA queries across the six traces.[10] We find that prior to censorship, local ISP servers resolved at least 90% of queries for both categories. Post-porn censorship, we observe a small increase ($\approx 5\%$) in queries resolved by Google's public DNS resolvers, with a corresponding decrease in queries resolved by the ISP. This number rises to $\approx 13\%$ in post-YouTube censorship traces. At the same time, the queries resolved by local ISP servers drops to 70%, and we see an increase in queries resolved by OpenDNS and LEVEL-3. The use of alternate DNS resolvers to circumvent censorship has appeared in other censorship incidents [37], and potentially increases user exposure to security risks [12].

Web proxies. For each trace we identify unique domains in HTTP requests labeled by McAfee's categorization service as *Anonymizers*. For SOHO traffic, we observe only 1 web proxy prior to the YouTube block, which rises to an average of 41 proxies post-block, with a striking 114 proxies on the day of the block. Residential traffic shares the same pre-block distribution of web proxies as SOHO traffic, though with a less dramatic increase post-block (11.5% on average). From domains extracted from SSL certificates, we find *no* proxy hosts in traces prior to the YouTube block, but after the block we observe 15 and 8 unique proxy hosts (18Sep12 and 02Aug13, respectively). These hosts either operate encrypted by default or provide an easy option for encryption, as confirmed through manual analysis. For example, the top two, `youtubeproxy.org` and `12345proxy.net`, use HTTPS by default, and another popular one, `4everproxy.com` lists HTTPS-based proxies prominently on its home page. In addition, the respondents to our survey indicated that SSL-based software such as OpenVPN and Hotspot Shield are among the most popular circumvention tools. Apparently these tools grew in popularity during the year between our last trace and our survey, as we did not find dominant usage trends for either in our data.

User behavior after viewing block page. We can also gain insight into how users responded to censorship by analyzing their actions after encountering a block page: in particular, whether they then attempt to access similar unblocked content, attempt to employ circumvention, or apparently give up (shift to some other form of activity).

We assess this as follows: for each user encountering a block page, we analyze their HTTP transactions in the subsequent 5-minute window. To reduce ambiguities due to IP aliasing, we confine this analysis to activity from the same address that also uses the same `User Agent`, which we assume is likely stable over short time intervals. (This approach still suffers from the possibility of multiple users behind a single IP address/NAT who employ the same user agent [29].) We then examine a histogram of the domain names and search keywords[11] in the HTTP requests generated by the users. We observe:

- On average 60% of the users performed a search engine query after encountering a block page for a porn domain, and 75% of users did so after encountering a block page for YouTube. Note that these proportions represent a lower

[9]For SOHO traffic, in our additional traces (18Sep12 and 02Aug13) the SSL ratio increases by several orders of magnitude; however this may instead reflect the YouTube censorship that spans this same timeline.

[10]To eliminate bias due to automated DNS queries that might potentially use a diverse set of DNS resolvers, we limit our analysis to queries for the blocked categories, as these unlikely are due to non-human actors.

[11]We developed signatures to extract keywords from popular search engine queries.

Domain Shorthand (% of total porn bandwidth)					
03Oct11	22Oct11	21Dec11	28Feb12	18Sep12	02Aug13
(a) Residential Traffic (GB)					
4.91	7.37	2.67	2.32	—	—
A / 42.3%	A / 26.4%	***I*** / 22.8%	***M*** / 23.2%	—	—
B / 12.1%	B / 15.4%	**J** / 16.8%	***R*** / 13.7%	—	—
C / 7.9%	*F* / 9.5%	$\bar{A}$ / 7.9%	*S* / 7.3%	—	—
D / 5.9%	D / 7.4%	*K* / 5.5%	***T*** / 4.1%	—	—
E / 3.8%	*E* / 3.2%	**L** / 4.1%	**U** / 3.8%	—	—
(b) SOHO Traffic (GB)					
6.71	12.32	3.37	0.47	2.76	1.90
A / 42.4%	A / 46.2%	*M* / 27.4%	**V** / 16.5%	R / 14.0%	X / 71.7%
B / 11.3%	D / 12.0%	**N** / 8.3%	**W** / 13.4%	**Z** / 12.5%	S / 13.0%
D / 7.5%	B / 8.7%	***O*** / 8.3%	*X* (9.3%	*H* / 11.1%	**BB** / 4.9%
G / 3.5%	C / 5.2%	**P** / 4.8%	**Y** / 7.1%	**AA** / 7.8%	**CC** / 1.6%
E / 3.2%	*H* / 2.7%	**Q** / 4.6%	*F* / 6.6%	$\bar{A}$ / 5.5%	**DD** / 1.4%

Table 10: Top five porn domains sorted by bandwidth over time. The top row in parts (a) and (b) represents the total bandwidth in (GB) per trace. Domains with bar are blocked in the given trace. Underlined domains are blocked in the next trace. **Bold** domains are new domains, not seen in previous traces. *Italic* domains are unblocked in the next trace. Others are currently unblocked cases for which we do not have backward or forward reference.

bound because we lack visibility into encrypted traffic.[12] We find that for porn, content-specific searches heavily dominate these queries, rather than searches for porn domains, which matches previous findings that porn users are flexible about served content as long as it falls into a broad class [42]. For YouTube, we find a diverse range of primarily informational queries.[13].

- For porn, on average 70% of users who hit a block page access another porn domain within the next 5 minutes. For YouTube, on the day of the block 7% of users viewed a video using an alternate video content, rising to 12% in 02Aug13. These figures run slightly lower than those from our survey, where 17% of respondents indicated that they would use an alternate provider to access YouTube videos that they find blocked. Tying this in with our earlier result for search queries being dominated by information-retrieval intent, we speculate that users primarily settle for non-video representations of information, rather than actively searching for alternate/unblocked providers to serve a video.

- Surprisingly, we do not find a wide interest in either searching for circumvention mechanisms or directly accessing non-SSL web proxies within our analysis time window. For porn, this is perhaps because users have a tendency to shift to other unblocked porn providers, resulting in little incentive to try circumvention.

7. IMPACT ON CONTENT PROVIDERS

Upon the imposition of censorship, users have a range of options: (*i*) stop accessing the censored content altogether, (*ii*) access the same or similar content hosted by an alternate content provider, or (*iii*) employ a censorship-bypass mechanism to directly access the censored content. The first two options lead to the censored content provider losing a fraction of its previous traffic, the second of these may have a positive impact on the traffic for alternate content providers. The last option potentially increases costs for content providers from the perspective of content distribution: they will have to serve the blocked content remotely, due to the inability to deploy local servers in the censored region. Additionally, if the chosen circumvention mechanism anonymizes user location, the censored content provider can no longer serve geographically-relevant advertisements, which may reduce revenue. While our current study does not concretely establish the economic implications of censorship on content providers, we highlight where these may manifest with a view to motivate further research.

[12]We find that among popular search engines, `google.com.pk` has a dominating presence in our data, and also appears in top-5 servers in the SSL logs.

[13]Queries that represent user intent to obtain information about an object of interest, with potentially a large number of diverse results.

Trace	Total (GB)	Blocked domains (%)	Unblocked (%)
Residential Traffic			
21Dec11	2.67	9.00	91.00
28Feb12	2.32	3.94	96.06
SOHO Traffic			
21Dec11	3.37	0.16	99.84
28Feb12	0.47	0.29	99.71
18Sep12	2.76	10.70	89.30
02Aug13	1.90	0.01	99.99

Table 11: Distribution of porn bandwidth among blocked and unblocked domains. In 21Dec11 and 28Feb12 the censor only uses DNS for blocking, hence users can still access blocked content by using an alternative name server. In 18Sep12, although the censor uses IP blocking in addition to DNS blocking, its blocking is partial for some domains. In 02Aug13, the censor uses a combination of DNS blocking and HTTP redirection.

7.1 Video Content

Table 8(a) illustrates the distribution of video bandwidth among the four major providers before and after the YouTube block. An overwhelming portion (an average of $\approx 97\%$ across four pre-block traces) of video content was provided by YouTube, up until the censorship event concurrent with 18Sep12. On that day, only $\approx 15\%$ of video content was fetched from YouTube, half of which was being served by local ISP cache servers, and the other half fetched from servers located outside Pakistan. Recall that on that day, one of the two national service providers (with direct international connectivity) redirected YouTube HTTP traffic to one of its own error pages (Section 4). The residual percentage of YouTube traffic perhaps reflects the providers lacking sufficient capacity to handle the full load, and thus it failed to consistently redirect traffic.

The following two findings support this hypothesis. First, the error page initially appeared only five minutes into the trace, suggesting that the censorship was already taking place at the time traffic was captured. Second, we did not find evidence of any incomplete coverage in blocking YouTube's IP address space: we find sets of IP addresses for which both videos were fetched successfully as well as users were redirected to the error pages (for different HTTP requests). The trace collected 11 months later does not manifest any content served from YouTube. This does not necessarily imply that users stopped accessing the site. Indeed, the percentage of encrypted traffic increased manyfold (see SOHO traffic in Table 7), from about 6% in Feb 2012 to over 30% of total traffic post-YouTube censorship, potentially indicating the use of SSL-based censorship bypass mechanisms, as we discussed in Section 6.

Table 8(b) also indicates that post-censorship, most of the video traffic generated from within Pakistan initially switched to DailyMotion (82% of total traffic in 18Sep12), but 11 months later split between DailyMotion (40.9%) and Tune.pk (57.6%). This traffic

distribution is unusual considering the global traffic statistics of DailyMotion ($\approx 23x$ more compared to Tune.pk [2]), thus indicating strong regional popularity. Tune.pk is a Pakistani video portal that essentially provides a censorship-friendly wrapper around YouTube. It downloads inoffensive YouTube videos and serves them from its servers with an option for users to report offensive videos [41]. The case of Tune.pk highlights the benefits reaped by local markets due to blocking of a competitor.

The overall shift in traffic potentially leads to a redistribution of advertisement revenue; the censored content provider loses out in favor of alternate providers, a shift exacerbated by the fact that local content owners tend to provide their content through video sharing sites that remain accessible to their viewers without the use of any circumvention technology. We find such a trend in Table 8(c), which shows more and more embedded links shifting from YouTube to other video sharing sites. Most local content is now served through embedded links of unblocked video providers: embedded links pointing to YouTube drop from an average of around 83% to 73% the day of block (18Sep12), to about 51% 11 months later (02Aug13), with DailyMotion getting $\approx 32\%$ of embedded links and Tune.pk jumping from virtually no embedding to nearly 11%.

The drop in the percentage of embedded YouTube links also leads to search engines adjusting their page ranks for localized searches. For example, a manual search (country-specific via `google.com.pk`) for top-5 local television shows reveals that the top results point to Tune.pk and DailyMotion, while a search for non-local content (top-5 television shows in USA) returns top results referencing YouTube.

In summary, the censored video content provider loses traffic and revenue to competing non-censored sites in multiple ways: direct reduction of traffic, local content providers moving their hosted channels to alternate providers, reduced embedded referencing in third-party pages, and lower page rank for localized search. The provider potentially also loses revenue due to the increased expense of serving the content via the distribution channels available to circumventers. For non-censored providers, these considerations may provide an incentive to take long-term control over local content. For example, DailyMotion has recently moved to partner with the largest ISP in the country [15].

7.2 Porn Content

For each trace, we ranked the porn sites in accordance with their traffic served into the country, measured using the methodology described in Section 5. Table 10 shows the top-5 coded porn domains for each trace. We observe that prior to blocking, globally popular domains [2] top the list. After the blocking event, new players emerge and take the top spots. In most cases, these new players were non-existent in previous traces (indicated with bold in the table), and their relative distribution across post-block traces vary inconsistently. (We do see a few domains, such as X, S and R, that appear in the top-5 for more than one trace.) We speculate that users are familiar with a few favorite porn websites, but, after blocking, find out about alternatives through search engines, hence the variety in the top ranked sites. This possibility fits with the finding in Section 6 that after landing on a block page, porn users tend to perform content-specific search queries.

Similar to the non-porn video sharing sites, censoring a porn site also impairs its revenue share from within the censored region. In Table 11 we analyze the bandwidth distribution among blocked and unblocked porn websites. We observe that the bulk of the bandwidth is captured by unblocked porn domains. (An exception to the case is $\bar{A}$, which appears in a subsequent trace despite being blocked in 21Dec11—circumvention is made possible by obtaining correct IP addresses via non-local resolvers.) The popularity of new porn domains can also be explained by the fact that after the initial introduction of the porn blacklist (21Dec11), there seems to be no aggressive strategy by the censor to block new popular porn content.

8. IMPACT ON SERVICE PROVIDERS

In this section we assess the consequences of the censorship events on ISPs. We examine this in terms of analyzing the operator's web caching behavior, focusing on the video content from the four major providers, since our pre-block traces indicate that videos constitute the content ($\approx 95\%$) served by the ISP's cache servers.[14]

Table 12 lists the top-5 ASNs serving video content for each of the traces.[15] Prior to the YouTube blocking, the top ASN is the local ISP—on average its caching servers provided 76% of the video content. On the day of blocking, we still see the ISP's caching servers providing a small fraction of YouTube video content. This leakage indicates that, initially, the ISP's censorship implementation was incomplete; its caching servers had not completely flushed cached YouTube content, and hence still served it upon request. (Recall that the block was put in place on the day of capture; some users could potentially get correct answers for YouTube from their local DNS cache, or by using alternate DNS resolvers as previously discussed.)

Moving forward 11 months, the local ISP completely falls off the charts to be replaced by CDNs serving DailyMotion and Tune.pk videos. Indeed, we find the ISP's cache servers completely absent from 02Aug13, and the ISP confirmed to us that the systems no longer provided any utility. Based on discussions with the ISP operators, we learned two reasons for this. First, Google had provided free caching servers to Pakistani ISPs, infrastructure tailored specifically to YouTube-caching. The other video content providers do not offer such free caching solutions, leaving it difficult for the local ISPs to justify the cost of deploying and maintaining custom solutions for the providers' content. Second, the drastic decrease in unencrypted video content (Table 7) made it hard to justify the benefits of caching, since ISPs cannot in general cache encrypted content. The ISPs instead turned to the option of leasing more upstream bandwidth, rather than buying and maintaining caching servers.

Consequently, today all video content is primarily fetched from the servers of their respective providers. Indeed, the operators we acquired traces from had to purchase additional Internet bandwidth after the block.

9. CONCLUSIONS AND FUTURE WORK

We have studied the impact of Internet censorship on major stakeholders (service providers, content providers, and end users) in the context of two major censorship events in Pakistan, the block of porn content in 2011, and of YouTube in 2012. To this end, we analyzed home and SOHO traffic before, during, and after the censorship events, from the vantage point of a mid-size ISP in a large metropolitan area. As the foundation of our analysis, we de-

[14]We establish this by looking at the distribution of `Content Type` served by the ISP's caching servers.

[15]Labovitz noted that content delivery shifted to European servers after MegaUpload servers in North America were seized [28]. Our study focuses on alternate content providers when the primary has been blocked, and the implications of their infrastructural arrangements on an operator.

ASN (% of total video bandwidth)					
03Oct11	22Oct11	21Dec11	28Feb12	**18Sep12**	**02Aug13**
26.5 GB	56.5 GB	45.2 GB	12.6 GB	10.7 GB	2.7 GB
Local-ISP, PK (78.69)	Local-ISP, PK (82.08)	Local-ISP, PK (70.08)	Local-ISP, PK (76.21)	Dailymotion, FR (45.67)	FIBERRING, NL (58.67)
Google, US (17.85)	Google, US (13.74)	Google, US (24.74)	Google, US (17.62)	TMNET, MY (22.99)	Dailymotion, FR (19.76)
YouTube, IE (1.46)	YouTube, IE (1.68)	YouTube, IE (3.71)	YouTube, IE (3.15)	Local-ISP, PK (7.22)	OMANTEL, OM (14.01)
Dailymotion, FR (1.23)	Dailymotion, FR (1.41)	EdgeCast, US (0.68)	Dailymotion, FR (2.93)	Tinet, DE (4.11)	Akamai, US (7.70)
CCWW, GB (0.80)	Akamai, US (0.84)	Dailymotion, FR (0.62)	EdgeCast, US (0.12)	YouTube, IE (3.98)	Tinet, DE (0.68)

Table 12: Top 5 ASNs serving video, ranked by bandwidth. Bold indicates YouTube blocking. The top row gives the total video bandwidth. In our traces, FIBERRING serves Tune.pk videos, while OMANTEL, TMNET, CCWW, Tinet, Akamai, and EdgeCast primarily serve DailyMotion videos.

veloped methodologies to identify censorship activity within our packet traces with high confidence.

We observed that blocking of porn content caused increases in encrypted traffic (Table 7) but primarily users turned to alternative sites (Table 10). In contrast, YouTube blocking caused a major shift towards increased encrypted traffic, indicating that users resorted to circumvention mechanisms to continue their access. In addition, we find this shift well underway already on the day that the government imposed censorship, indicating that a portion of users can very rapidly adapt to the introduction of new blocking mechanisms.

Censorship of YouTube also affected the financial landscape of video content providers (Table 8). New players emerged and completely took over the video-sharing market that previously was almost wholly owned by YouTube prior to its blocking. This shift also had consequences for ISPs that previously served video content primarily from YouTube caches (freely provided by Google) hosted within their own networks. Post-YouTube blocking, the ISPs must fetch video content through their upstream transit provider, reflecting an increase in bandwidth costs. After the YouTube blocking was implemented at the local-ISP level (using DNS spoofing), we observe a shift away from the use of the local ISP's DNS resolvers, dropping from more than 90% pre-blocking to about 70% post-blocking. We note that such a shift somewhat erodes a nation's overall control over its Internet traffic as users transfer their base of trust (i.e., DNS resolution) to parties outside the country.

Following up on this work, we plan to analyze additional data from a different ISP in another large city in Pakistan to assess trends seen across cities and providers. The expectation in Pakistan is that porn blocking will continue in the future, but YouTube censorship will soon end [16]. If that indeed happens, it will be illuminating to study whether the proportion of encrypted traffic returns to pre-censorship levels; whether users continue to outsource their DNS resolution; and the degree to which video traffic distribution between YouTube and alternate video sharing sites readjusts.

Acknowledgements

This work was supported by the Engineering and Physical Sciences Research Council [grant number EP/L003406/1]; and the US National Science Foundation [grant numbers 1223717, 1237265]. Opinions expressed are solely those of the authors. We thank Jon Crowcroft, Steven Murdoch and Balachander Krishnamurthy for feedback on different parts of this paper. We also thank our shepherd, Olaf Maennel, the anonymous reviewers for their useful comments, and the anonymous ISP for facilitating access to their data.

10. REFERENCES

[1] S. Alcock and R. Nelson. Measuring the impact of the copyright amendment act on New Zealand residential DSL users. In *Proc. ACM Internet Measurement Conference*, 2012.

[2] Alexa. `http://www.alexa.com/topsites`. Online. April, 2014.

[3] Anonymous. The Collateral Damage of Internet Censorship by DNS Injection. *SIGCOMM Comput. Commun. Rev.*, 42(3):21–27, June 2012.

[4] S. Aryan, H. Aryan, and J. A. Halderman. Internet Censorship in Iran: A First Look. In *Free and Open Communications on the Internet*, Washington, DC, USA, 2013. USENIX.

[5] A. Attaa. `http://tinyurl.com/mnw9olp`. Online. Feb, 2014.

[6] Bro. `http://www.bro.org/`. Online. April, 2014.

[7] A. Chaabane, M. Cunche, T. Chen, A. Friedman, E. D. Cristofaro, and M.-A. Kaafar. Censorship in the Wild: Analyzing Web Filtering in Syria. Technical report, Cornell University Library, Feb. 2014.

[8] Cisco. `http://tinyurl.com/mev32z8`. Online. Apr, 2014.

[9] Citizen Lab. O Pakistan, We Stand on Guard for Thee: An Analysis of Canada-based Netsweeper's Role in Pakistan's Censorship Regime. `http://tinyurl.com/oxxap8t`, June 2013.

[10] M. Cooper. `http://tinyurl.com/p7ck76f`. Online. Feb, 2014.

[11] J. R. Crandall, D. Zinn, M. Byrd, E. Barr, and R. East. ConceptDoppler: A Weather Tracker for Internet Censorship. In *Computer and Communications Security*. ACM, 2007.

[12] S. Crocker, D. Dagon, D. Kaminsky, D. McPherson, and P. Vixie. Security and Other Technical Concerns Raised by the DNS Filtering Requirements in the PROTECT IP Bill. `http://domainincite.com/docs/PROTECT-IP-Technical-Whitepaper-Final.pdf`, May 2011.

[13] A. Dainotti, C. Squarcella, E. Aben, K. C. Claffy, M. Chiesa, M. Russo, and A. Pescapé. Analysis of country-wide internet outages caused by censorship. In *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference*, IMC '11, pages 1–18, New York, NY, USA, 2011. ACM.

[14] J. Dalek, B. Haselton, H. Noman, A. Senft, M. Crete-Nishihata, P. Gill, and R. J. Deibert. A Method for Identifying and Confirming the Use of URL Filtering Products for Censorship. In *Proc. ACM Internet Measurement Conference*, 2013.

[15] Dawn. `http://tinyurl.com/n3j22gy`. Online. April, 2014.

[16] Dawn News. `http://tinyurl.com/m7vhg7u`. Online. Apr, 2014.

[17] H. Duan, N. Weaver, Z. Zhao, M. Hu, J. Liang, J. Jiang, K. Li, and V. Paxson. Hold-On: Protecting Against On-Path DNS Poisoning. In *Proc. Workshop on Securing and Trusting Internet Names*, SATIN 2012.

[18] Electronic Frontier Foundation (EFF). http://tinyurl.com/oklfy29. Online. Apr, 2014.

[19] ElectronicFrontierFoundation. Switzerland. http://tinyurl.com/d22vbbq.

[20] A. M. Espinoza and J. R. Crandall. Automated Named Entity Extraction for Tracking Censorship of Current Events. In *USENIX Workshop on Free and Open Communications on the Internet*, 2011.

[21] ExtremeTech. http://tinyurl.com/6nabr85. Online. Apr, 2014.

[22] R. Farahbakhsh, Ã. Cuevas, R. Cuevas, R. Rejaie, M. Kryczka, R. Gonzalez, and N. Crespi. Investigating the reaction of BitTorrent content publishers to antipiracy actions. In *P2P*, pages 1–10. IEEE, 2013.

[23] A. Filastò and J. Appelbaum. OONI: Open Observatory of Network Interference. In *Free and Open Communications on the Internet*. USENIX Association, 2012.

[24] P. Gill, M. Arlitt, Z. Li, and A. Mahanti. Youtube traffic characterization: A view from the edge. In *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, IMC '07, pages 15–28, New York, NY, USA, 2007. ACM.

[25] Internet Service Providers Association of Pakistan (ISPAK). http://www.ispak.pk. Online. Apr, 2014.

[26] C. Kreibich, N. Weaver, B. Nechaev, and V. Paxson. Netalyzr: illuminating the edge network. In *Proc. ACM Internet Measurement Conference*, 2010.

[27] C. Lab. Routing Gone Wild: Documenting upstream filtering in Oman via India. Technical report, Citizen Lab, 2012.

[28] C. Labovitz. The Other 50% of the Internet. North American Network Operators' Group (NANOG) presentation, February 2012.

[29] G. Maier, F. Schneider, and A. Feldmann. NAT Usage in Residential Broadband Networks. In *Proc. Passive and Active Measurement*, 2011.

[30] McAfee. http://www.trustedsource.org. Online. Apr, 2014.

[31] Mozilla. https://wiki.mozilla.org/Public_Suffix_List. Online. Apr, 2014.

[32] Z. Nabi. The Anatomy of Web Censorship in Pakistan. In *Proc. USENIX Workshop on Free and Open Communications on the Internet*, 2013.

[33] F. News. http://tinyurl.com/22me5e7. Online. Feb, 2014.

[34] OpenNet Initiative. Pakistan. Online. April, 2014.

[35] OpenNet Initiative. https://opennet.net. Online. Apr, 2014.

[36] P. Reidy. US court orders Google to remove Innocence of Muslims film from YouTube. http://tinyurl.com/mjd9sjk. Feb, 2014.

[37] Renesys. Turkish Internet Censorship Takes a New Turn. http://www.renesys.com/2014/03/turkish-internet-censorship/, 2014.

[38] Security Information Exchange. https://www.dnsdb.info/. Online. Apr, 2014.

[39] A. Sfakianakis, E. Athanasopoulos, and S. Ioannidis. CensMon: A Web Censorship Monitor. In *USENIX Workshop on Free and Open Communications on the Internet*, 2011.

[40] TeamCymru. IP to ASN Mapping. http://tinyurl.com/5dtp78. Apr, 2014.

[41] Techniasia. Tune.pk comes up with a way to bypass Pakistan's YouTube block. http://tinyurl.com/mdgb2ke. Apr, 2014.

[42] G. Tyson, Y. Elkhatib, N. Sastry, and S. Uhlig. Demystifying porn 2.0: A look into a major adult video streaming website. In *Proceedings of the 2013 Internet Measurement Conference*, IMC '13, pages 417–426, New York, NY, USA, 2013. ACM.

[43] J.-P. Verkamp and M. Gupta. Inferring Mechanics of Web Censorship Around the World. In *Free and Open Communications on the Internet*, Bellevue, WA, USA, 2012. USENIX.

[44] N. Weaver, R. Sommer, and V. Paxson. Detecting Forged TCP Reset Packets. In *Proc. NDSS*, 2009.

[45] YouTube and Blocking in Pakistan. Has it affected you? http://tinyurl.com/kcvj325. Online. April, 2014.