Inferring Mechanics of Web Censorship Around the World

John-Paul Verkamp Minaxi Gupta

School of Informatics and Computing, Indiana University
{*verkampj, minaxi*}*@cs.indiana.edu*

Abstract

While mechanics of Web censorship in China are well studied, those of other countries are less understood. Through a combination of personal contacts and Planet-Lab nodes, we conduct experiments to explore the mechanics of Web censorship in 11 countries around the world, including China. Our work provides insights into the diversity of modus operandi of censors around the world and can guide future work on censorship evasion.

1 Introduction

Internet censorship is a growing concern around the world, affecting a large portion of the world's population. Modes of censorship vary widely, ranging from complete disconnection, to selective censorship of Web pages, to censorship of search engines and online social networks (OSNs)–such as Facebook and Twitter. Owing to these developments, the research community has recently been studying censorship from various angles. One of those angles is censorship evasion, with works in [1, 5, 7, 10, 19] proposing various ways of defeating censorship.

Measurement studies of censorship have also been undertaken. For example, Dainotti et al. studied complete Internet disconnection, such as that in Egypt and Libya during the Arab Spring revolution in early 2011 [3]. The Open Net Initiative (ONI) studies have focused on what various countries around the world censor [4, 11, 12]. In a similar spirit, CensMon used PlanetLab infrastructure to measure censorship in various countries [15]. Aspects of Chinese censorship have received particular attention, with works investigating how China censors Web accesses [2], Tor [17], Skype [9] and where its censoring modules are located [20]. However, the study of censorship in other countries has thus far been limited to identifying *what* is being censored rather than *how*, leading to a gap in our understanding of how censors around the world operate.

In this paper, we complement existing knowledge about the mechanics of Web censorship in 11 countries around the world, including China and an additional seven where no well-known publicly-available measurement infrastructure exists. Focusing on the question of *how* censorship is conducted (as opposed to *what is being censored*), we present results on how Bangladesh, Bahrain, India, Iran, Malaysia, Russia, Saudi Arabia, South Korea, Thailand, and Turkey censor the Web. Of these, we could only find accessible PlanetLab nodes in China, India, Russia and Turkey. For the other countries we recruited volunteers through personal contacts. In many cases we confirm previous work; however in at least one case we found censorship where none had previously been reported [4, 11, 12].

We began by creating a taxonomy that explores the design space of a censoring module in terms of what triggers it, where it is located, and how censorship is executed. Then through a series of experiments that include accessing known censored websites from the test machines, analyzing raw packet captures, and various other specialized tests on a per site basis, we discovered and verified the following aspects of the mechanics of censorship in the countries under study:

- Mechanics of censorship vary across countries: Even though we were only able to conduct measurements in a small number of countries of interest due to lack of measurement infrastructure in other countries, we saw evidence of much of the design space from our taxonomy being exploited by censors. For example, Malaysia, Russia and Turkey censor at the DNS while other countries do so at routers or other network hardware. South Korea censors both at the DNS and at the routers. Also, Saudi Arabia and China censor on destination IP addresses, while other countries primarily censor on hostnames, URLs, or keywords. Further, execution of censorship ranges between DNS redirects (Malaysia, Russia, South Korea and Turkey), connection timeouts (Bangladesh and India), TCP resets (China), and HTTP responses with various 200-, 300-, and 400-level status codes (Bahrain, Iran, Saudi Arabia, South Korea and Thailand).

- Censorship is not always explicitly communicated: While Bahrain, Iran, Saudi Arabia, South Korea and Thailand make censorship evident, other countries leave users wondering why their attempt to fetch a web page was unsuccessful.

- Censorship can be stateful: China filters only the first HTTP GET request in a TCP stream, likely for the sake of efficiency. It does so by maintaining state. In addition, after filtering an HTTP request, it maintains flow state about source and destination IP addresses, port number and protocol of the denied request to deny further communication between the same pair of machines even when such communication would not previously have been blocked. No other country in our lists conducts stateful censorship.

- Customized censorship evasion is possible: Censorship evasion techniques roughly fall in two categories. In the first are ones that avoid arousing suspicion from the censoring techniques in the first place, such as [1, 5, 7, 10, 19]. The second category includes techniques where the impact of censorship is annulled at the client, such as in [2], where Clayton et al. propose a way to ignore TCP reset packets sent by the Chinese censoring module. Our work can motivate novel censorship evasion techniques, particularly in the latter category.

2 Design Space for a Censoring Module

There are three primary design considerations for any Web censorship system. The first is the *trigger* for censorship to take place. The trigger could be a combination of hostname, IP address, port number, protocol, or URL/keyword(s) captured by regular expressions or an equivalent. The second consideration is the the *location* of the censoring module in

the network. Common locations of censoring modules include DNS and routers. The final consideration is the actual *execution* of censorship, whereby a censor decides which protocols to modify and the manner in which censorship is communicated to the user. The following sections describe the design space along these dimensions.

2.1 Trigger and Location

Censorship could be triggered at several points after a user first initiates a given connection. We discuss triggers and location together since only certain kinds of information is available for triggering censorship at specific locations.

- Trigger: Hostname, Location: Local DNS resolver. The first point at which censorship could be triggered is when the hostname in the user's request gets resolved through the local DNS resolver. Instead of fetching the IP address(es) corresponding to the hostname by consulting the DNS hierarchy (which includes the resolver's local cache), the resolver could determine that the either the hostname itself or the domain or sub-domain portion of it is on a to-be-filtered list and return one or more of the pre-configured IP addresses. Typically, servers behind the returned IP addresses send custom pages indicating the reason for redirection. Note that since this filter does not have access to port number and protocol information, it will indiscriminately filter all connections, possibly hurting non-Web services on the filtered hostnames.

- Trigger: IP address/port/protocol, Location: Router. Filtering at the level of TCP connection establishment and initial HTTP requests are the next logical steps after DNS resolution. Each of these traverse through routers, making them an excellent choice for locating the censoring module. While any router could be used to execute filtering, border routers belonging to ISPs are a good choice and are thus typically used because all traffic traverses them [20]. Among other things, router-based filtering could be triggered on the destination server's IP address–which is available as early as the first SYN packet during TCP connection establishment and in every subsequent TCP packet containing HTTP data. Note that since port and protocol information is also available in all packets, router-based filters can be more selective than DNS-based filters.

- Trigger: Hostname, Location: Router. Filtering under this trigger works very similar to the IP address-based filtering at the routers, with the only difference being that routers using this trigger consult a hostname-based blacklist for filtering purposes instead of an IP address-based one. Note that accessing the IP address, port, and/or protocol requires routers to only consult TCP/IP headers. However, accessing the hostname requires them to also consult the *Host* header in the application header portion of HTTP request packets.

- Trigger: Full URL, Location: Router. While many censors block entire websites, some may want to censor content at a finer granularity. Examples of such censorship include filtering specific Twitter or Facebook pages, specific blogs or specific YouTube videos but allowing access to the site otherwise. The blacklist used by a router in this case would contain full URLs or regular expressions. Just like the host name-based censorship, routers executing URL-based censorship will have to consult the *request* portion of the HTTP header.

- Trigger: Keyword, Location: Router Keyword-based filtering is unique in that while all the above forms of filtering filter only outgoing traffic, this type of filtering can be applied both to outgoing and incoming traffic. Specifically, a router can filter HTTP requests based on keywords contained in any portion of the URL and also filter responses based on keywords in the returned content. For example, if the keyword 'falun' is included in the blacklist, each of http://www.falun.com/index.html, http://example.com/falun.htm and http://example.com/search?q=falun will be filtered in spite of the fact that 'falun' appears in the hostname in the first URL, in the directory path in the second, and as a query parameter in the third.

- Trigger: Any, Location: Client machine. Filtering at the client machine is another alternative to filtering at the DNS or routers. While its deployment could be challenging, it has the advantage of not consuming any network resources or requiring any enhancements to network hardware. An example of such censorship is the custom version of Skype produced at the behest of Chinese authorities. This version automatically scans incoming and outgoing messages for questionable content and sends this information to the Chinese government [9]. Unfortunately, such techniques are radically different from those otherwise investigated. As such, we do not consider this option subsequently in this paper.

2.2 Execution

The actual execution of censorship is dependent on the location of the censoring module. When the module is located at the local DNS resolver, the execution involves a DNS redirection to a specialized server. When the module is located at the router, there are multiple choices. In fact, several can co-exist. We describe them next.

- Filter request or response: A censor may simply choose to filter requests. The first opportunity for this arises during the TCP connection establishment phase. However, since hostnames are not available in TCP SYN packets, the blacklist for such filtering would consist of IP addresses, port and protocol combinations. The next opportunity for filtering requests arises when HTTP requests are made. The blacklist for such filtering can consist of a combination of domain, sub-domain, or hostnames and even regular expressions focusing on URLs and keywords. When requests are filtered, corresponding connections time out and the browser displays an error stating this to the end user. Likewise, since the request never reaches the destination web server, it cannot even infer that censorship is occurring. As the dual to filtering the request, a censor may filter on response, either instead of filtering requests or in addition. Response-based filtering can only be keyword based, however, as the hostname and page requested are not included in an HTTP response.

- Filter and return: A variant of filtering requests is where the censor decides to respond back to the user upon filtering the request. The response may simply disrupt the connection at either or both end points or do so in addition to informing the user about the filtering. In theory, a censor can choose to inform the destination about filtering as well but given the sensitivity of the act, one would not expect this to ever be the case. As a variant, a censor may filter on the

response and then return in a manner similar to above.

- Allow but return first: Another variation on "filter and return" is the case where a censor allows the TCP and HTTP connections to proceed as normal but responds back to either of the end points in a manner similar to "filter and return". Seemingly odd, this is a practical strategy for cases where the censoring device is out-of-band, possibly to avoid hurting line speeds at the in-band routers. Note that since the censoring devices are closer to the user than the destination web servers in most real-world cases, their responses will almost always beat the actual response from the destination, making this technique as effective as the "filter and return".

- Modify request or response: As an alternative extreme measure, a censor may modify the request or the response. A modified request could cause the destination web server to return a less objectionable page, such as the main page of a blogging site instead of a censored blog. A modified response may simply alter the content sent by the destination. Though we did not observe any censor exploiting these options, such approaches are not far-fetched as ISPs are known to modify content on the fly to generate revenue from advertisement impressions and clicks [21].

3 Methodology

In this Section, we describe how we chose countries to investigate for how they censor the Web. We also discuss how we find and instrument client machines we use for conducting experiments and how we choose websites for testing. Finally, we discuss how we infer the mechanics of censorship from the collected data.

3.1 Censoring Countries

We started by narrowing down the list of countries to investigate by using three independent sources that rank countries based on the level of censorship observed. The sources were Freedom House's (FH) free press rating [8], Reporters Without Borders (RWB) press freedom index [14], and OpenNet Initiative (ONI) [11, 12]. The FH scores range from 10 to 99 for each country, with higher scores indicating a higher prevalence of censorship. Specifically, countries with scores up to 30 are regarded as *free*, between 30 and 60 as *partly free*, and the rest as *not free*. North Korea received the highest score and 55 other countries were labeled as not free. RWB had a slightly more nuanced scale with scores ranging from 0 to 105 in four categories: under 10 as *most free*, between 10 and 50 as *somewhat free*, between 50 and 75 as *less free*, and the rest as *least free*. A total of 13 countries were regarded as least free according to their categorization. In addition, RWB also maintains two supplemental lists of countries particularly related specifically to Internet censorship. There are 13 countries in the first list which are referred to as *Internet enemies*. Fourteen countries on the second complementary list are referred to as *under surveillance*. Finally, the ONI reports on censorship pertaining to *political*, *social*, *security* and *tools* categories separately. Classifications for countries in each category range from *no evidence* of filtering, *selective* filtering, *substantial* filtering, and *pervasive* filtering. A total of 16 countries were labeled as conducting pervasive filtering in at least one of the four categories.

List	Source	List	Rating	Countries
1	FH	free press rating	not free	55
2	RWB	press freedom index	least free	13
3	RWB	Internet enemies under surveillance	- -	13 14
4	ONI	political/social security/tools	pervasive filtering	16
			Total unique	71

Table 1: Number of countries with most censorship

No source ranked all countries in the world but there was a significant overlap with one another. To narrow down the list of countries to explore, we focused on the worst offenders in each of the lists. Their numbers are shown in Table 1. A total of 71 countries were unique across all lists and were considered for experimentation.

3.2 Testing Machines

In the next step, we focused on finding machines to run experiments. Four options were considered for this task: open Web proxies, Tor exit nodes [16], PlanetLab machines [13], and machines belonging to residents of these countries. We secured the latter through personal connections. Of these, we ruled out open Web proxies and Tor exit nodes because while they could help fetch web pages of interest from machines in countries of interest, they did not provide a detailed account of the packets exchanged, which was necessary to understand the mechanics of censorship. Unfortunately, the combination of PlanetLab machines and personal connections was insufficient to explore censorship in each country of interest. In particular, we could find either of these options available for only 11 of the countries on our wish list. While not ideal, of the six countries that were on all the four lists, our list included two, Iran and China. Also, of the 8 countries which appeared on three lists, our list included two, Bahrain and Saudi Arabia. Further, Malaysia, Russia and South Korea were available for experimentation from the list of 16 countries that were on two lists. The remaining four countries on our list were present in at least one of the four lists. To this list of 11 countries, we added Bangladesh based on anecdotal reports of censorship. Table 2 shows the list of 11 countries we experimented with and the method we used to experiment with each. Note that of the countries in two or more worst offender lists, only China and Russia have PlanetLab nodes. Given that most PlanetLab machines are run by educational institutions means that they might be given special treatment by the censoring modules–in turn leading to potentially incorrect inferences. While these problems would need to be addressed in future work, PlanetLab provided a useful basis for initial studies.

# Lists	Country	Client Type	Country	Client
4	China	PlanetLab	Iran	Person
3	Bahrain	Person	Saudi Arabia	Person
2	Malaysia South Korea	Person Person	Russia	PlanetLab
1	Bangladesh Thailand	Person Person	India Turkey	PlanetLab PlanetLab

Table 2: Method used in each country

3.3 Inferring Censorship

In order to infer the mechanics of censorship in each country, we needed a set of censored websites in each country. To this end, we pulled the lists of top most inaccessible sites in each country from HerdictWeb [6], which assembles reports about censored websites around the world. While many of the websites on these lists are censored, network and server problems or issues at the client could also lead to a site being reported. Further, since HerdictWeb is a community-driven effort, one cannot rule out the possibility of purposeful misreporting. A limitation of this dataset is that it only reports on hostnames and not individual subdomains or URLs. Thus, for example, if wordpress.com is not filtered in a country but a specific blog is–such as willo200man.wordpress.com–HerdictWeb will only report wordpress.com. This limitation does not impede our ability to learn how censorship is executed in a country or where the censoring module is located as long as we can find a few censored websites. However, it prevents us from being able to make nuanced inferences about when censorship is based on hostnames versus URLs and when keywords in either may be triggering it. Fortunately, a few of our volunteers were able to point us to a few exact URLs filtered in their country, which avoided this shortcoming for those countries [1].

For each website gathered from HerdictWeb, we fetched DNS resolutions to find the corresponding IP addresses. In order to test for IP based censorship, we removed the HTTP Host header from the packet which might otherwise trigger a hostname or URL based system. If the same result occurs, the page is being censored based on IP address; however, if the page either loads successfully or if the remote webserver returns a server specific error message, we infer that IP-based filtering is not occuring.

- Analyzing packet captures: We took a two-phased approach to finding out censorship mechanics in each country. In the first, we focused on inferring DNS-based censorship. Toward this goal, we filtered the packet captures to only show DNS packets. If we saw a set of IP addresses repeated for multiple websites suspected to be censored, it was a strong indication of redirection, often to a website controlled by the censor. We confirmed each such case manually.

If censorship was not found to be occurring at the DNS resolvers, it could be occurring at one or more routers. Even if it was occurring at the DNS resolvers, it could additionally be happening at the routers. For packet captures from volunteers, we isolated traces involving websites, URLs, and IP addresses where they reported incidence of censorship and analyzed them manually for details on the mechanics of censorship. For packet captures from PlanetLab, we pulled instances of test websites and IP addresses on a local machine and compared the results to those in the traces to manually determine instances of censorship. Subsequent to this determination, packet captures of interest were evaluated in detail.

[1] For a final list of URLs, visit http://research.jverkamp.com

4 Results

We found evidence of censorship in all 11 countries we studied though the mechanics varied, primarily due to different censoring products by different countries. Table 3 summarizes the results we describe next.

4.1 DNS-based Censorship

Four countries of the 11 on our list showed evidence of DNS-based censorship. These were Malaysia, Russia, South Korea, and Turkey. Of these, only South Korea showed a warning page (see Figure 1). The warning page informs the visitor about the filtering and gives a series of phone numbers to call if they wanted to appeal the blocking of the page. The rest redirected clients to localhost, which caused the browser to display an error message and potentially leaf the client wondering why

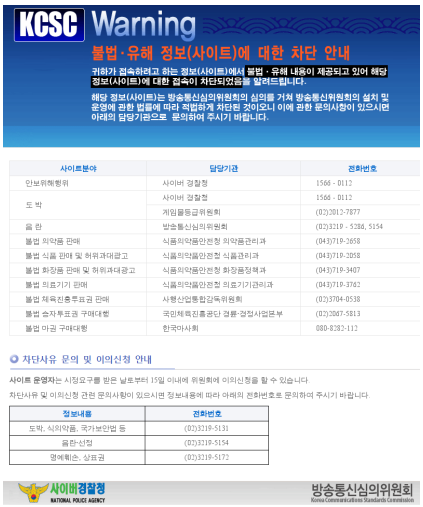

Figure 1: Warning page in South Korea

4.2 Router-based Censorship

A total of eight countries were found censoring at the routers. This included South Korea, which also does DNS-based censorship. The others were the seven that did not censor at the DNS resolvers.

- Triggered by destination IP: We found evidence of IP-based censorship in two countries: China and Saudi Arabia. In neither case, however, was the filtering done during TCP connection establishment phase, which would have been the earliest opportunity for such filtering at a router. Instead, both waited until the first HTTP GET request to filter packets. We examine each in turn.

The filtering in the case of Saudi Arabia would fall under "filter and return", as discussed in Section 2 and confirming This confirms the results from the ONI survey [11]. However, we cannot infer if the filtering was done on the request or response. Since evidence suggests that filtering (in the context of China) is usually bi-directional [2, 20], we expect that in our case, request packets would be filtered because our test clients are located inside the censoring regimes. The Saudi Arabian ISP chose to wait till the first HTTP GET request to filter since it allows sending a

Country	Location	Trigger (verified)	Execution (if router-based)	Result
Bangladesh	Router	Hostname	Filter	Timeout
Bahrain	Router	Hostname, URL	Filter and return	HTTP 403, warning
China	Router	IP, Hostname, URL, keyword	Allow but return first	TCP reset
India	Router	Hostname	Filter	Timeout
Iran	Router	Hostname	Filter and return	HTTP 403, warning
Malaysia	DNS	Hostname	-	DNS redirect
Russia	DNS	Hostname	-	DNS redirect
Saudi Arabia	Router	IP, hostname	Filter and return	HTTP 200, iframe warning
South Korea	DNS+router	Hostname	-/Filter and return	DNS or HTTP 302 redirect, warning
Thailand	Router	Hostname, URL	Allow but return first	HTTP 302 redirect, warning
Turkey	DNS	Hostname	-	DNS redirect

Table 3: Summary of results

warning page in response, which would not be possible if it chose to reset the TCP connection instead. The latter would leave the visitor confused. The warning page is almost identical to the one shown for Bahrain shown in Figure 2, differing only in that it has a green color scheme as opposed to Bahrain's blue. There are several interesting aspects of the Saudi Arabian warning page. First, it is sent back as a spoofed HTTP response with a status code of 200, conveying to the browser that it received a normal response. Upon receipt of this message, the browser will not be able to fetch anything else from the intended destination since the actual TCP connection will timeout for lack of activity. Further, the entire warning page is an HTML iframe which is loaded from another host whose name is simply an IP address. Adopting the iframe strategy allows modifying the warning page any time without any modification to the censoring module. The HTML of the warning page contains a reference to WireFilter [18], implying the filtering product used by this ISP.

In the case of China, the filtering is of the type "allow but return first", implying that the HTTP GET requests are allowed to proceed as normal but an out-of-band device censors the request, perhaps to keep the routers operating at line speeds. Specifically, the client is sent a spoofed TCP RST packet. In fact, multiple RSTs are sent to ensure that the client terminates the TCP connection. In the majority of these connections, we saw four spoofed packets returning, each with a different sequence and acknowledgment (ACK) number. The ACK numbers in three of these spoofed RST packets corresponded to the sequence number in the original client packet, as if the server had already sent back one full packet, and as if the server had already sent back two full packets, respectively. The fourth RST arrived without the corresponding ACK. This would effectively suppress cases where non-standard packet lengths are received on systems that will accept a RST without an ACK. In addition, the TTL values observed in these four packets generally have one or two different but similar values and also correspond to roughly the number of hops that would be required to leave Chinese networks. In addition, the ID field in each packet seems to be assigned sequentially, showing a strictly increasing sequence associated with each TTL value. This strongly implies that the censoring machines are a small group of machines at each border router.

These observations are in tune with those made by Clayton et al. in their 2006 paper [2] even though they focused on keyword blocking only and made these observations from a client outside China that connected to a web server in China. Our observation of the same behavior six years later implies that this aspect of the Great Firewall of China has remained unchanged. Note that since the censoring device would typically be closer to the client than many of the censored destinations, the spoofed RSTs would generally beat the actual response from the destination. In each case, the original response packet from the server would be detected in our packet capture, but not by the client as the reset packet had already forced it to terminate its connection. Further, the Chinese censoring devices also send spoofed RSTs to the destination even though they might reach after the destination has already responded once to the client. This causes the destination to terminate the connection as well. Though this behavior was clear from the destination not retrying to deliver the packets, we confirmed it by running a web server locally and asking the client to access a URL with a known filtered keyword, "falun". Our server received a similar set of four spoofed RSTs, as did the client. The only difference was that the sequence number and ACK number were set to correspond to what the server would expect to get.

- Triggered by hostname, URL or keyword: Eight of the 11 countries we studied had router-based filters that were triggered by some portion of the hostname, URL, or keyword. Two of them, China and Saudi Arabia, also filtered on IP address, as described earlier. Their mechanisms for when filtering was triggered by hostname, URL, or keyword were the same as in the case of IP address. The others were Bangladesh, Bahrain, India, Iran, South Korea, and Thailand. (Recall that South Korea also did DNS-based filtering.) Of these, Bahrain and Iran were shown to use commercial filtering systems by the ONI survey in 2009 [11] while filtering was shown for South Korea, Thailand, and India by Deibert et al. in their 2011 book [4]. Bangladesh was particularly interesting as in the same book, no evidence of filtering was found but our tests show evidence of filtering a year later.

Broadly speaking, this group of six countries censored in three ways. The first was a timeout. Both Bangladesh and India utilized a the simple "filter" method and let the client's TCP connection time out, offering no insight to the client about the censorship. The second mechanism, used by Bahrain and Iran, was to "filter and return" and send the client spoofed HTTP packets with status code for access to content forbidden (403) and a warning page. These packets instructed the client to close the connection. The third mechanism was the use of HTTP redirect (status code 302), which South Korea and Thailand used. Both sent an HTTP status code 302 containing a location header which a web browser interprets as a redirect. However, South Korea

(a) Bahrain (b) Iran (c) Thailand

Figure 2: Warning pages in Bahrain, Iran and Thailand

uses "filter and return" while Thailand uses "allow but return first". The latter was verified by looking for a response from the destination Web server and has implications on censorship-evasion on the client-side, as we discuss in Section 5.

The warning pages for Bahrain, Iran, and Thailand are shown in Figure 2. All three countries have censorship ratings of medium or higher transparency in the ONI reports and we confirmed that the warning pages and descriptions of the warning text we observed were consistent with those noted by previous work [4, 11]. Effectively, they all alert the user to the fact that the page they visited has been blocked along with a link to report it if it should not have been blocked. Iran also displays a list of allowed pages under the warning text and automatically redirects the user to one of these pages when a timer at the bottom of the page runs out. In all cases but Iran's, the warning is displayed in English as well as the local language.

We verified that all eight countries in this group filtered on hostnames. Additionally, Bahrain, China and Thailand filtered on exact URLs that we were able to verify. China also filtered on keywords. It is likely that some of these countries are filtering on URLs or keywords as well but we were not able to verify that aspect in our measurements without a large list of sites and keywords known to be censored.

Investigating how South Korea chooses between DNS-based and router-based censorship, it appears that sites blocked by DNS-based censorship are a subset of those blocked by router-based censorship. Cases where blocking at the DNS resolver only impacts a single site, such as a North Korean Government website, www.korea-dpr.com, were handled at the DNS level. However, other cases, where many hostnames may be sharing a set of IP addresses and only some are blocked, were handled at the router-level. An example is the case where the site, wordpress.com is not blocked but a hostname at that site, willow200man.wordpress.com, is blocked but both share the same set of IP addresses. Further, sites blocked at the DNS level are also blocked at the router level, which we verified by configuring a DNS resolver outside Korea as our default resolver.

4.3 Statefulness

An interesting aspect of the Chinese filtering is the statefulness of the censoring device. First, it only filters the first HTTP GET request arriving after the TCP handshake. Any subsequent GET requests, or those arriving without a preceeding TCP handshake are ignored. This observation was made in the work by Tu et al. [20]. However, that work focused on the location of filtering devices and did not delve into the intuition behind this state maintenance. We also note that the filtering device also maintains information about the IP addresses of source and destination, so if a destination is filtered for a client, any subsequent connections between that pair of machines will be filtered. We verified that this timeout does indeed exist by fetching first a page with a censored keyword in the query parameter and then repeatedly fetching the page without. After a timeout period of generally 12 hours, the page without the keyword would return successfully. This confirms the findings of Winter et. al. that show the same timeout when dealing with censorship of the Tor network in China [17]. In contrast, filtering is stateless in all other countries we investigated.

5 Conclusion

Our work finds that the mechanics of censorship vary greatly among the countries we studied. Much of this variation is attributable to variety of censoring products available. Our findings have implications on the possibility of censorship evasion at the client. For example, bypassing censorship in Malaysia, Russia and Turkey is simply a matter of using an alternate DNS resolver outside of the country. However, even though South Korea also does DNS-based censorship, the same evasion technique would be ineffective because South Korea also censors at the routers. Similarly, Clayton et al. noted in [2] that ignoring spoofed TCP RST packets in China can help a client bypass censorship if destination Web servers also ignored such packets. However, this technique would not work for any of other countries that perform censorship at routers, primarily because none but Thailand use "allow and return", implying that the others filter client requests which preventing the destination from even knowing that a request was made. Even in Thailand, the technique proposed by Clayton et al. will have to be adapted to ignore spoofed HTTP packets instead of TCP RSTs.

While our work has offered initial insights into the mechanics of censorship methods in use around the world, a further study is required to understand them in detail. First, if different ISPs in a country execute censorship-related filtering differently, as was noted in previous works [11, 12, 4], our current work would fail to observe those differences due to only a single connection in the non-PlanetLab countries. This bias could be eliminated by recruiting more volunteers. Further, while the 11 countries we studied spanned much of the spectrum of the design space for censoring modules, it would be useful to infer where other censoring countries fall in that spectrum.

References

[1] BURNETT, S., FEAMSTER, N., AND VEMPALA, S. Chipping away at censorship firewalls with user-generated content. In *USENIX Security Symposium* (2010).

[2] CLAYTON, R., MURDOCH, S., AND WATSON, R. Ignoring the great firewall of China. In *Privacy Enhancing Technologies Symposium (PETS)* (2006).

[3] DAINOTTI, A., SQUARCELLA, C., ABEN, E., CLAFFY, K., CHIESA, M., RUSSO, M., AND PESCAPÉ, A. Analysis of country-wide Internet outages caused by censorship. In *ACM/USENIX Internet Measurement Conference (IMC)* (2011).

[4] DEIBERT, R., PALFREY, J., ROHOZINSKI, R., AND ZITTRAIN, J. *Access Contested: Security, Identity, and Resistance in Asian Cyberspace*. MIT Press, 2011.

[5] FIFIELD, D., HARDISON, N., STARK, J., PORRAS, R., BONEH, D., AND TOR, S. Evading censorship with browser-based proxies. In *Privacy Enhancing Technologies Symposium (PETS)* (2012).

[6] HerdictWeb: Help spot Web blockages. http://www.herdict.org/.

[7] INVERNIZZI, L., KRUEGEL, C., AND VIGNA, G. Message in a bottle: Sailing past censorship. In *Privacy Enhancing Technologies Symposium (PETS)* (2012).

[8] KELLY, S., AND COOK, S. Freedom house: Freedom on the Net, 2011. http://www.freedomhouse.org/report/special-reports/freedom-net-global-assessment-internet-and-digital-media.

[9] KNOCKEL, J., CRANDALL, J., AND SAIA, J. Three researchers, five conjectures: An empirical analysis of tom-skype censorship and surveillance. In *USENIX Workshop Free and Open Communications on the Internet (FOCI)* (2011).

[10] MOGHADDAM, H., LI, B., DERAKHSHANI, M., AND GOLDBERG, I. Skypemorph: Protocol obfuscation for tor bridges. Tech. rep., University of Waterloo, 2012.

[11] Internet filtering, 2009. Tech. rep., OpenNet Initiative, 2009. http://opennet.net/sites/opennet.net/files/ONI_*_2009.pdf.

[12] Internet filtering, 2010. Tech. rep., OpenNet Initiative, 2010. http://opennet.net/sites/opennet.net/files/ONI_*_2010.pdf.

[13] PlanetLab: An open platform for developing, deploying, and accessing planetary-scale services. http://planet-lab.org.

[14] Reporters Without Borders: Press Freedom Index, Apr. 2010. http://en.rsf.org/press-freedom-index-2010,1034.html.

[15] SFAKIANAKIS, A., ATHANASOPOULOS, E., AND IOANNIDIS, S. CensMon: A Web censorship monitor. In *USENIX Workshop on Free and Open Communication on the Internet (FOCI)* (2011).

[16] Tor: Anonymity online. http://torproject.org.

[17] WINTER, P., AND LINDSKOG, S. How china is blocking Tor. Tech. rep., Karlstads University, 2012. http://arxiv.org/pdf/1204.0447v1.pdf.

[18] Wirefilter: Carrier grade Web security system. http://www.wirefilter.com/.

[19] WUSTROW, E., WOLCHOK, S., GOLDBERG, I., AND HALDERMAN, J. Telex: Anticensorship in the network infrastructure. In *USENIX Security Symposium* (2011).

[20] XU, X., MAO, Z., AND HALDERMAN, J. Internet censorship in China: Where does the filtering occur? In *Passive and Active Measurements (PAM)* (2011).

[21] ZHANG, C., HUANG, C., ROSS, K., MALTZ, D., AND LI, J. In-flight modifications of content: Who are the culprits? In *USENIX Workshop on Large-Scale Exploits and Emerging Threats (LEET)* (2011).