Chinese Wall or Swiss Cheese? Keyword filtering in the Great Firewall of China

Raymond Rambert

Zachary Weinberg
College of Information & Computer Sciences
University of Massachusetts
Amherst, MA, USA
Carnegie Mellon University
Pittsburgh, PA, USA
zackw@cs.umass.edu

Diogo Barradas
INESC-ID, Instituto Superior Técnico
Universidade de Lisboa
Lisbon, Portugal
Carnegie Mellon University
Pittsburgh, PA, USA
diogo.barradas@tecnico.ulisboa.pt

Nicolas Christin
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA, USA
nicolasc@cmu.edu

ABSTRACT

The Great Firewall of China (GFW) prevents Chinese citizens from accessing online content deemed objectionable by the Chinese government. One way it does this is to search for forbidden keywords in unencrypted packet streams. When it detects them, it terminates the offending stream by injecting TCP RST packets, and blocks further traffic between the same two hosts for a few minutes.

We report on a detailed investigation of the GFW's application-layer understanding of HTTP. Forbidden keywords are only detected in certain locations within an HTTP request. Requests that contain the English word "search" are inspected for a longer list of forbidden keywords than requests without this word. The firewall can be evaded by bending the rules of the HTTP specification. We observe censorship based on the cleartext TLS Server Name Indication (SNI), but we do not observe bulk decryption of HTTPS.

We also report on changes since 2014 in the contents of the forbidden keyword list. Over 85% of the forbidden keywords have been replaced since 2014, with the surviving terms referring to perennially sensitive topics. The new keywords refer to recent events and controversies. The GFW's keyword list is not kept in sync with the blocklists used by Chinese chat clients.

CCS CONCEPTS

• **Social and professional topics → Technology and censorship**; • **General and reference → Measurement**.

KEYWORDS

Censorship, Keyword filtering, Measurement

WWW '21, April 19–23, 2021, Ljubljana, Slovenia

ACM ISBN 978-1-4503-8312-7/21/04.
https://doi.org/10.1145/3442381.3450076

ACM Reference Format:
Raymond Rambert, Zachary Weinberg, Diogo Barradas, and Nicolas Christin. 2021. Chinese Wall or Swiss Cheese? Keyword filtering in the Great Firewall of China. In *Proceedings of the Web Conference 2021 (WWW '21), April 19–23, 2021, Ljubljana, Slovenia.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3442381.3450076

1 INTRODUCTION

Chinese keyword-based censorship of the Web is well known [13, 24, 25, 36, 40], but no two past studies report exactly the same behavior. For at least fifteen years, there have been regular reports of the GFW's keyword list being updated in response to breaking news [e.g., 1], but the frequency and extent of these updates is not known. Chinese authorities now seem to be concentrating on blocklists enforced by applications [28, 29, 34]. This raises the question of whether keyword-based censorship by backbone routers has been deemphasized or its focus changed.

In this paper, we investigate the current extent of the Great Firewall of China (GFW)'s keyword censorship of unencrypted HTTP traffic; the evolution, since 2014, of the list of forbidden terms; which parts of an HTTP request and response are inspected for keywords; how much keyword censorship varies depending on the locations of client and server; and whether encrypted (HTTPS) traffic is also monitored. To do so, we used virtual private servers, hosted by multiple service providers, inside and outside of China, as HTTP(S) clients and servers. We drew keywords from a combination of four lists of sensitive terms (see Section 3.1).

We find the forbidden keyword lists have changed considerably since 2014. There are now two sublists: a short one that is censored unconditionally, and a longer list that is censored only when the word "search" also appears in the request. Only 78 of 451 keywords collected by Chu [10] are still consistently censored in our tests. The keywords that have remained on the list refer to perennially sensitive topics in China, such as censorship-evasion software, the Tienanmen Square demonstrations, and sources of foreign propaganda. The removed keywords refer to topics whose political

sensitivity has diminished with time, and the new keywords refer to newly sensitive topics (e.g., COVID-19). Overall, only 8% of the keywords censored by chat clients are also censored by packet inspection, suggesting that chat and packet blocklists are maintained independently.

We observe no censorship of traffic that remains within mainland China, and no censorship of traffic between Hong Kong and hosts outside mainland China. For traffic that is censored, forbidden keywords are detected regardless of the destination TCP port, but only in some locations within an HTTP request (see Section 4.4. Requests that we expect to be censored are missed by the firewall as much as 25% of the time. Requests that we expect *not* to be censored will still trigger the firewall 1–3% of the time.

We can confirm the existence of a "penalty box" period as reported in earlier studies [10, 11, 51]. During a 90-second period after a TCP stream is disrupted, *all* requests sent from the same client to the same server will have a 50–75% chance of being blocked even if they contain no censored keywords.

We find no evidence of inspection of HTTP responses or of bulk interception of HTTPS traffic. However, we do observe a reaction to forbidden host names in the unencrypted SNI (Server Name Indication) message during TLS setup, as found by Chai et al. [9].

2 RELATED WORK

The Chinese government has sought to censor the internet since its earliest availability within China [4]. The present "Great Firewall" developed from systems first deployed in the 1990s. It employs a variety of techniques, including DNS-based censorship of entire sites [2, 27] and application-level content filtering [12, 19, 30] as well as the keyword-based censorship that this paper focuses on. Activists and researchers have sought to understand and publicize the GFW's operation for almost as long as it has existed. One of the earliest formal studies, Clayton et al. [11] in 2006, showed how the GFW injects TCP RST packets when it sees a "forbidden" keyword in an HTTP request or response.

One line of research since then focuses on understanding the GFW's mechanisms, such as whether HTTP responses are censored [36]; where the hardware that implements the GFW is located, and the consistency of censorship policies [20, 51]; the extent to which the GFW can intercept encrypted streams [9, 44]; and how the GFW's understanding of TCP can be exploited to evade censorship [7, 11, 27, 47].

A second, complementary line of research focuses on understanding which keywords and sites are censored. Typically, candidate keywords are drawn from a public corpus of documents on diverse topics, sensitive and not, such as Chinese Wikipedia, IMDB, news sites, and social media [10, 13, 34]. The firewall is then probed with each candidate. Once sensitive strings are identified, they must be refined to determine the exact keywords that are blacklisted [50]. Some researchers have proposed a number of techniques for expanding the initial corpus via directed searches [15, 24] and following links from censored pages [16].

This paper advances both lines of research. We report on details of the GFW's partial understanding of HTTP, which might be further exploitable by circumvention tools like Geneva [7]. We also describe how the keyword lists have changed over time, are now sensitive to context (such as the presence or absence of "search"), and are only partially synchronized with chat client blacklists.

The GFW stands out for its sophistication, but China is not the only country to censor the internet. Case studies of other countries have been published for just as long [e.g., 18, 23]. More recently, several groups have developed tools for continuous, worldwide monitoring of the reach and pervasiveness of censorship and how this changes over time. OONI [22] is the best known; others include ICLab [35], Satellite-Iris [37, 39], Quack [45], and Censored Planet [43]. Our work does not engage with these projects directly, since they currently focus on reachability tests to a list of sensitive sites operated by third parties, while we test keyword lists, using dedicated servers under our control. However, our techniques for detailed probing of HTTP could be of use to these platforms, and our observations of changes in the Chinese keyword list since 2014 demonstrate the need for continuous monitoring.

3 EXPERIMENTAL METHODS

This section provides a methodological overview of our study. We describe the keyword lists we use to test for censorship, the general algorithm used for each probe, the two types of HTTP servers we used, and finally the physical locations of all the hosts involved.

3.1 Keyword lists

We tested potentially sensitive keywords from three lists, two of them compiled by others. All three lists include keywords in English as well as Chinese; one list also contains other languages.

Both simplified and traditional Chinese characters appear in all three lists. Traditional Chinese characters are commonly used in Hong Kong and Taiwan, and have come to symbolize political separation from the mainland. Therefore, in Section 4, censored keywords in English and other languages are lumped with simplified Chinese under the label "non-traditional," but censored keywords in traditional Chinese are counted separately.

Note that we count keywords as traditional Chinese whenever they cannot be encoded in GB 2312 [42]. This causes some keywords from regional variants of Chinese to be lumped with traditional keywords.

Due to limited space, we cannot list all of the keywords we tested in this paper. We have made complete lists available online, with summarized data on the censorship of each keyword, courtesy of the Internet Archive: `https://archive.org/details/pruned_keyword_lists`.

Wikipedia 2014. Chu [10, 33] derived this list from the URLs of Wikipedia pages in English and Chinese. At the time, Wikipedia was accessible in China, but many individual pages were censored [41]. Chu probed 50 million Wikipedia URLs using a method similar to that of ConceptDoppler [13], and identified 936 keyword-based "rules" enforced by the GFW. The unique keywords appearing in these rules include 33 words in Latin script, 418 in simplified Chinese, and 218 in traditional Chinese, The regional variant of Chinese is not recorded.

Wikipedia 2020. With assistance from native Chinese speakers, we manually selected terms likely to be sensitive from the titles of the 1 000 most frequently viewed Wikipedia articles as of March

```
GET /search/?k=什么什么&id=42 HTTP/1.1
Host: tokyo.echo.example

HTTP/1.1 200 OK
<h1>検索</h1>
<form method="get" action="tokyo.echo.example/search/">
<label for="k">検索キーワードを入力してください:</label>
<input type="text" name="k" id="k" value="什么什么">
```

```
GET /search/?k=什么什么&id=42 HTTP/1.1
Host: tokyo.kw.example

HTTP/1.1 200 OK
<h1>検索</h1>
<p>但是。这个问题成为。六四事件</p>
```

Figure 1: HTTP dialogues with a typical echo server (left) and keyword server (right). The keyword「什么什么」(not a censored term) and the code number 42 appear in requests to both servers. The echo server responds with an HTML page including the keyword in the query, but the keyword server responds with a page including keyword #42,「六四事件」(which *is* a censored term). Orange boxes highlight the correspondence between request and response in both cases.

2020 in five different Wikipedia languages: English, Standard Chinese, Classical Chinese, Min Nan, and Yue. This list includes 99 keywords in Latin, 41 in simplified Chinese, and 82 in traditional Chinese. Only 5 of the keywords in this list also appear in the Wikipedia 2014 list.

Chat client blacklists. CitizenLab maintains a comprehensive dataset [14] of censored keyword lists extracted from chat applications popular in China, such as WeChat and Sina Weibo. These are regularly updated and contain references to events as recently as 2020 (e.g., related to the coronavirus outbreak). After removing duplicates and URLs, this dataset includes roughly 63 200 keywords. Pilot testing indicated that the majority of these are not censored by packet inspection. For efficiency's sake, we manually selected 16 475 terms likely to be sensitive from this dataset, with assistance from native Chinese speakers. 2 073 of these terms are in Latin, 13 080 in simplified Chinese, 1 016 in traditional Chinese, 264 in Arabic script (several different languages), and 42 in other scripts (notably Cyrillic and Tibetan). 145 terms also appear in the Wikipedia 2014 list and 17 words in the Wikipedia 2020 list.

3.2 Echo and Keyword Servers

The servers we use to probe the GFW must accept messages containing arbitrary text. Existing servers with a built-in search interface are often suitable. Whether or not the search finds any results, the server will reply with a `200 OK` message and, often, a reply containing the string that was searched for. We call these *echo servers*, after the TCP echo protocol.

The left side of Figure 1 shows an example dialogue with an echo server. It receives a query for「什么什么」(*shénme shénme*, a placeholder noun, literally "what what") and replies with HTML containing the same word. The GFW could react to either occurrence of the word.

Echo servers are convenient, but not perfect, for this study. Each sensitive keyword appears in *both* the request and the response, so we cannot use them to determine whether the GFW reacts to requests, responses, or both. They offer no control over fine details such as the character encoding of the response or the location of the keyword within the request. Finally, popular websites might receive special treatment from the GFW.

To address these problems, we adopted Park and Crandall [36]'s technique of deploying custom servers that can echo sensitive terms, respond to sensitive terms with an innocuous document, or respond to innocuous requests with sensitive terms (selected by code number). We call these *keyword servers*. They have several other features, discussed below. We deployed keyword servers on domain names used only for this study, and did not mention or hyperlink them anywhere, so we have no reason to think that the GFW would give them special treatment.

The right side of Figure 1 shows how a keyword server might respond to our request.「什么什么」is ignored, but `id=42` causes it to reply with a page containing「六四事件」(*liùsì shìjiàn*, "June 4th incident," referring to the 1989 protests in Tienanmen Square).

Table 1: Location and hosting of all test hosts.

Location	Hosting	Client?	Server?	
San Jose, CA (1)	Vultr		keyword	
San Jose, CA (2)	Linode		keyword	
Newark, NJ	Linode	✓		
Pittsburgh, PA	Cogent	✓		
Paris, France	Vultr		keyword	
Mumbai, India	Linode		keyword	
Tokyo, Japan (1)	Vultr		keyword	
Tokyo, Japan (2)	IDC Frontier		echo	`realmotor.jp`
Singapore	Vultr		keyword	
Taichung, Taiwan	ServerField		keyword	
London, UK	Vultr		keyword	
Hong Kong (1)	VPS-Server	✓	keyword	
Hong Kong (2)	Alibaba	✓	keyword	
Hong Kong (3)	DYXnet		echo	`pegasus-idc.com`
Beijing (1)	Alibaba	✓	keyword	
Beijing (2)	Tencent	✓	keyword	
Guangzhou (1)	Tencent	✓	keyword	
Guangzhou (2)	Huawei		echo	`onlinedown.net`
Shanghai	Alibaba	✓	keyword	

3.3 Client and Server Locations

The locations and hosting providers for our test hosts are listed in Table 1. The table shows whether each host served as client, server, or both. For echo servers, we show the domain name of the site used for testing. Not all of the clients were used for every experiment.

We selected diverse client and server locations, to search for inconsistent behavior by the GFW based on geography or network topology. Specifically, these locations let us send test messages that remain within mainland China, or that travel between China and locations in Europe, North America, and nearby in Asia. Because of Hong Kong's contentious status, we selected three hosts there, operated by a European company, a Chinese company, and a native Hong Kong company.

3.4 Detection Algorithm

We probe for censorship by sending HTTP requests containing a sensitive keyword from a client on one side of the GFW, to a server on the other side. If we receive a network-level "connection reset" error, we infer that the GFW has injected a TCP `RST` packet and the keyword is considered to be censored. We use a custom HTTP client, described in detail in Section 4.4. It can place the sensitive keyword in any of several different locations within the request, to test the GFW's understanding of HTTP.

We must take care not to be fooled by errors made by the GFW, by the inherent unreliability of packet injection [36], or by the "penalty box" blockade of benign connections after a censored request (see Section 4.5). Thus, the client repeats each request at quarter-second intervals for up to five minutes. The result is only considered conclusive when it receives the same response (either a valid response from the server, or a RST from the firewall) ten times in a row. After determining a keyword is censored, the client makes innocuous requests at one-second intervals until ten of these in a row succeed, indicating that the penalty box period has expired.

4 EXPERIMENTS

In this section, we present each of our experiments and its results.

In our first experiment (Section 4.1) we re-tested the non-traditional keywords found to be censored by Chu [10] (the Wikipedia 2014 list), using the three echo servers listed in Table 1, and discovers the special role of the English word "search." In our second experiment (Section 4.2) we expanded the set of test keywords to include the Wikipedia 2020 and chat lists, and the traditional Chinese keywords from Wikipedia 2014. For better control over fine details, and to avoid any special treatment of well-known websites, we use exclusively keyword servers in this and subsequent experiments.

In Section 4.4, we test the GFW's ability to detect sensitive keywords in different locations and encodings within an HTTP request. In Section 4.5 we investigate the "penalty box" applied to clients after a request is censored. In Section 4.6 we test the GFW's ability to interfere with HTTPS (encrypted websites). Finally, in Section 4.7 we describe other miscellaneous experiments.

4.1 Keyword Censorship: Echo Servers

In our first experiment we re-tested the 451 non-traditional keywords from the Wikipedia 2014 list, using clients and echo servers located inside and outside China. Table 2 depicts a breakdown of how many keywords were censored in HTTP request/response pairs from each client to each server.

Unsurprisingly, we see no censorship of traffic that doesn't enter mainland China, including traffic between Hong Kong and North America. More surprisingly, we see no censorship of traffic *within* China, contrary to earlier reports [49, 51]. This may mean regional ISPs have less of a role in the GFW than they did in the early 2010s. Censorship does occur for traffic crossing the border of mainland China in either direction, but not consistently. Different hosts in the same physical location experience different levels of censorship, suggesting a dependence on routing rather than geography.

Only 15 keywords, listed in Table 3 with glosses, are consistently censored regardless of route. Four of them refer to anti-censorship proxy software. Another five refer to perennially sensitive topics such as Mao Zedong's life and the Tienanmen Square protests. The remainder are specific individuals or groups considered either subversive or criminal by the Chinese government.

Table 2: Censored non-traditional keywords from Wikipedia 2014, using echo servers.

Client \ Server	Japan `realmotor.jp`	Hong Kong `pegasus-idc.com`	Guangzhou `onlinedown.net`	All
Hong Kong (1)	0	0	82	**82**
Hong Kong (2)	0	0	44	**44**
Pittsburgh, PA	0	0	59	**59**
Shanghai	15	14	0	**15**
Beijing	15	15	0	**15**
Guangzhou	15	15	0	**15**

Table 3: Keywords from the Wikipedia 2014 list [10] that are unconditionally censored in HTTP requests to echo servers.

Keyword (Pinyin)	**Meaning**
动态网 (dòng tài wǎng)	Anti-censorship proxy
Ultrasurf	Anti-censorship proxy
Ultrareach	Anti-censorship proxy
无界网络 (wú jiè wǎng luò)	Anti-censorship proxy
`Mao_The_Unknown_Story`	Critical biography of Mao Zedong
平反六四 (píng fǎn liù sì)	Redress, possible allusion to Tiananmen
网络人权宣言 (wǎng luò rén quán xuān yán)	Cyber Declaration of Human Rights
我的奋斗 (wǒ de fèn dòu)	Mein Kampf (Hitler's autobiography)
延安日记 (yán ān rì jì)	The Vladimirov Diaries, a history of Yan'an during WWII from a Soviet perspective
盘古乐队 (pán gǔ yuè duì)	Pangu, underground rock band known for supporting Taiwanese independence
邓正来 (dèng zhèng lái)	Deng Zhenglai, professor and dissident
彭小枫 (péng xiāo fēng)	Peng Xiaofeng, business executive accused of embezzlement
王斌余 (wáng bīn yú)	Wang Binyu, a murdered worker
章沁生 (zhāng qìn shēng)	Zhang Qinsheng, a general and dissident
自由亚洲电台 (zì yóu yà zhōu diàn tái)	Radio Free Asia

Expanded censorship trigger. During the above experiment, we accidentally discovered that a longer list of keywords is censored if the English word "search" is also included in the HTTP request line. For example, `http://echo.example/search?k=法轮` is censored but `http://echo.example/update?k=法轮` is not. (法轮 *fǎ lún* is the Chinese name for the Buddhist wheel of dharma. It forms part of the name of the Falun Gong religious movement.) This suggests that blanket bans are reserved for especially sensitive material and the GFW tries to discourage people from *searching* for other material. Entire websites on less-sensitive topics are censored by other means, e.g. DNS poisoning.

We looked for other strings that trigger the same effect among the most common other 10,000 English words (according to Google's Trillion Word Corpus [5]). We also tested three commonly used abbreviations for a search parameter ("q," "kw," and "s"), and three Chinese words related to searching (「搜索」 *sōu suǒ*, search; 「查找」 *chá zhǎo*, find; and 「关键词」 *guān jiàn cí*, keyword). None of these triggered expanded keyword censorship.

4.2 Keyword Censorship: Keyword Servers

In our second experiment we expand our probes for censored keywords to include all three of the lists described in Section 3.1. As explained in that section, each list is divided into two sublists based on the script: trad, words written with traditional Chinese characters, and non-trad, all other words. We also switched from echo to keyword servers beginning with this experiment.

4.2.1 *Testing the three candidate keyword lists.*

Wikipedia 2014 keyword list (non-trad). Table 4a shows results for the non-trad subset of the Wikipedia 2014 list, i.e., the same keywords tested in Section 4.1. For each client (row) and keyword server (column), the table shows the number of distinct keywords found to be censored. Each cell holds two numbers, separated by a + sign: first the number of keywords censored unconditionally, second the number of additional keywords censored when accompanied by the triggering word "search." For instance, the client located in Shanghai, when connecting to the first server in Hong Kong, observes censorship of 14 keywords in all HTTP requests, and censorship of an additional 60 keywords when the client uses URLs of the form `http://example.com/?search=XXX`.

Of the 451 keywords in the Wikipedia 2014 list, we find that typically 15 are censored unconditionally (the same 15 as for tests with echo servers, listed in Table 3) and 60 more are censored with "search." The search-only keywords follow the same themes as the unconditionally censored keywords: politically sensitive topics (e.g., 「藏独」(Tibet Independence) and 「89学运」([19]89 student movement)), foreign news agencies (e.g., 「美国之音」(Voice of America)), and circumvention tools (e.g., 「花园网」(Garden Networks)).

The table also shows several interesting phenomena dependent on the route taken by our probes. First, as we found in Section 4.1, none of the clients located within mainland China observe censorship when contacting keyword servers within China (e.g., Shanghai or Beijing).

Second, the Hong Kong S.A.R. is consistently "outside" the firewall. No routes between Hong Kong and foreign countries experience censorship; all routes between Hong Kong and mainland China experience censorship.

Third, when traffic enters or leaves mainland China, the set of censored keywords is mostly consistent from route to route, but a handful of routes experience much less censorship than the norm. The most dramatic example is that we observed *no* censorship when our Pittsburgh, PA, USA client contacts our Shanghai keyword server. One client in Beijing also experiences almost no censorship when communicating with keyword servers in Paris and Taiwan. But the same clients see typical levels of censorship when communicating with servers in other cities. One possible explanation is that the firewall has been disabled on certain routes by accident, or for testing. Another is that these routes are overloaded and the firewall has failed open.

Finally, we see stark discrepancies between the results for these keyword servers and the results for two of our echo servers, `real-motor.jp` and `pegasus-idc.com`, which seem not to be subject to additional censorship with "search." It is possible that the GFW exempts certain popular foreign websites from detailed scrutiny. This is only a hypothesis; we do not know how popular these sites are within China, or whether the Chinese government would ever trust a foreign site not to become a forum for criticism of its policies.

Wikipedia 2014 keyword list (trad). Table 4b shows results for the trad subset of the Wikipedia 2014 list. We find only a few additional keywords are censored. Thematically, they are consistent with the non-trad subset, including terms such as 「天安門」(Tiananmen), 「六四18週年」(18th Anniversary of June 4th), and 「新唐人電視台」(New Tang Dynasty Television, affiliated with Falun Gong). We observe more route-to-route variation; perhaps the blocklist for traditional Chinese is not as regularly updated. Curiously, requests from outside to inside mainland China, containing traditional characters, seem to be more aggressively censored than the reverse.

Wikipedia 2020 keyword list (non-trad). Table 4c shows results for the non-trad subset of the Wikipedia 2020 list. Only one keyword from this list was censored unconditionally: 「自由亚洲电台」(Radio Free Asia), which also appears in the Wikipedia 2014 list. Three more keywords are censored with "search:" 「刘晓波」(Liu Xiaobo, a Chinese political prisoner), 「法轮功」(Falun Gong), and 「色情」(pornography).

As with the trad subset of the Wikipedia 2014 list, we find more keywords are censored for requests from outside to inside mainland China. Among the terms censored only for external clients are "Coronavirus," "Remdesivir," and "Epidemic," all related to the coronavirus pandemic of 2020. This indicates that the censorship policy is regularly updated, and suggests the asymmetry we observe may be politically motivated—one blocklist to control debates within China and another to control the image it presents to the outside world.

Wikipedia 2020 keyword list (trad). Table 4d shows results for the trad subset of the Wikipedia 2020 list. The three keywords censored (with "search") on all censored routes are 「新唐人電視台」(New Tang Dynasty Television) (also in Wikipedia 2014 trad), 「八九民運」([19]89 democracy movement), and 「男子色情戲」(m/m pornography). These are broadly consistent with both the older Wikipedia 2014 list and the non-trad subsets of both Wikipedia-based lists. The GFW clearly aims to target specific forbidden subjects regardless of the script or terminology used to refer to them.

Chat clients blacklist keywords (non-trad). Table 4e shows results for the non-trad subset of the chat client blacklist. This list reveals many more censored keywords, with up to 1 221 distinct keywords censored for traffic leaving China, and another thousand for traffic entering China. Their themes are, again, politically sensitive topics (e.g. 「六四受难者」("June 4th victims", a reference to the suppression of the Tienanmen Square protest)), foreign media (e.g. 「纪元新闻网」(Epoch News Network)), and circumvention tools (e.g. 「无界网络」(UltraSurf)). More than half of the unconditionally censored terms, and many of the terms censored with "search," refer to the Tienanmen Square protest in some way. (As a striking example of the lengths Chinese censors and activists will go to regarding Tienanmen Square, the phrase "Восемь-Девять-IIIесть-Четыре" is censored by the Sina UC chat system, and was also censored with "search" on one of our routes. This is Russian for "Eight Nine Six Four," i.e., June 4th, 1989—two layers of coded reference.)

As we observed with the other lists, keyword censorship is broadly consistent from route to route, but a few routes experience much less censorship; in particular, we still see no censorship

Table 4: Censored keywords from each test list, for each route from a client (row) to a keyword server (column). Each cell holds two numbers: first the number of keywords censored unconditionally, second the additional number of keywords censored when `search` appears in the request. A "—" is shorthand for "0 + 0" (i.e., no censorship was observed).

(a) Wikipedia 2014 list [10] (non-traditional characters) (451 words)

Client \ Server	London	Mumbai	Paris	San Jose, CA (1)	San Jose, CA (2)	Singapore	Taiwan	Tokyo	HK (1)	HK (2)	Beijing (1)	Beijing (2)	Guangzhou	Shanghai	Total
Pittsburgh, PA	—	—	—	—	—	—	—	—	—	—	2 + 3	13 + 38	15 + 63	—	98
Hong Kong (1)	—	—	—	—	—	—	—	—	—	—	15 + 63	15 + 63	15 + 63	15 + 63	78
Hong Kong (2)	—	—	—	—	—	—	—	—	—	—	16 + 67	13 + 63	8 + 29	15 + 63	85
Beijing (1)	15 + 63	5 + 32	15 + 63	15 + 63	15 + 63	15 + 63	6 + 43	5 + 30	15 + 63	15 + 63	—	—	—	—	78
Beijing (2)	15 + 63	15 + 62	0 + 4	15 + 63	15 + 63	15 + 63	2 + 3	15 + 63	15 + 63	15 + 63	—	—	—	—	78
Guangzhou	15 + 63	15 + 63	15 + 63	15 + 62	15 + 63	15 + 63	15 + 63	15 + 63	15 + 63	15 + 63	—	—	—	—	78
Shanghai	15 + 63	15 + 63	14 + 64	15 + 63	15 + 63	15 + 63	12 + 63	18 + 63	14 + 60	15 + 63	—	—	—	—	81

(b) Wikipedia 2014 list [10] (traditional characters) (218 words).

Client \ Server	London	Mumbai	Paris	San Jose, CA (1)	San Jose, CA (2)	Singapore	Taiwan	Tokyo	HK (1)	HK (2)	Beijing (1)	Beijing (2)	Guangzhou	Shanghai	Total
Pittsburgh, PA	—	—	—	—	—	—	—	—	—	—	3 + 12	9 + 18	1 + 6	—	27
Hong Kong (1)	—	—	—	—	—	—	—	—	—	—	0 + 8	0 + 8	0 + 8	1 + 8	9
Hong Kong (2)	—	—	—	—	—	—	—	—	—	—	0 + 10	0 + 8	0 + 3	0 + 8	10
Beijing (1)	0 + 8	0 + 8	0 + 8	0 + 8	0 + 8	0 + 8	0 + 8	0 + 8	0 + 8	1 + 8	—	—	—	—	9
Beijing (2)	0 + 8	0 + 8	0 + 8	0 + 7	0 + 8	0 + 8	0 + 8	0 + 8	0 + 8	0 + 8	—	—	—	—	8
Guangzhou	1 + 6	1 + 6	1 + 6	1 + 6	1 + 6	1 + 6	1 + 6	1 + 6	1 + 6	1 + 6	—	—	—	—	8
Shanghai	0 + 8	0 + 8	0 + 2	0 + 8	0 + 8	0 + 4	0 + 8	—	0 + 2	0 + 8	—	—	—	—	8

(c) Wikipedia 2020 list (non-traditional characters) (137 words).

Client \ Server	London	Mumbai	Paris	San Jose, CA (1)	San Jose, CA (2)	Singapore	Taiwan	Tokyo	HK (1)	HK (2)	Beijing (1)	Beijing (2)	Guangzhou	Shanghai	Total
Pittsburgh, PA	—	—	—	—	—	—	—	—	—	—	3 + 6	7 + 8	1 + 3	—	19
Hong Kong (1)	—	—	—	—	—	—	—	—	—	—	1 + 3	1 + 3	1 + 3	1 + 3	4
Hong Kong (2)	—	—	—	—	—	—	—	—	—	—	1 + 3	1 + 2	0 + 1	1 + 3	4
Beijing (1)	1 + 3	1 + 3	1 + 3	1 + 3	1 + 3	1 + 3	1 + 3	1 + 3	1 + 3	1 + 3	—	—	—	—	4
Beijing (2)	1 + 3	1 + 3	1 + 3	1 + 3	1 + 3	1 + 3	1 + 3	1 + 3	1 + 3	1 + 3	—	—	—	—	4
Guangzhou	1 + 3	1 + 3	1 + 3	1 + 3	1 + 3	1 + 3	1 + 3	1 + 3	1 + 3	1 + 3	—	—	—	—	4
Shanghai	1 + 3	1 + 3	1 + 3	1 + 3	1 + 3	1 + 3	1 + 3	1 + 3	1 + 3	1 + 3	—	—	—	—	4

(d) Wikipedia 2020 list (traditional characters) (84 words).

Client \ Server	London	Mumbai	Paris	San Jose, CA (1)	San Jose, CA (2)	Singapore	Taiwan	Tokyo	HK (1)	HK (2)	Beijing (1)	Beijing (2)	Guangzhou	Shanghai	Total
Pittsburgh, PA	—	—	—	—	—	—	—	—	—	—	0 + 6	2 + 10	0 + 3	—	15
Hong Kong (1)	—	—	—	—	—	—	—	—	—	—	0 + 3	0 + 3	0 + 3	0 + 3	3
Hong Kong (2)	—	—	—	—	—	—	—	—	—	—	0 + 3	0 + 3	0 + 1	0 + 3	3
Beijing (1)	0 + 3	0 + 3	0 + 3	0 + 3	0 + 3	0 + 3	0 + 3	0 + 3	0 + 3	0 + 3	—	—	—	—	3
Beijing (2)	0 + 3	0 + 3	0 + 3	0 + 3	0 + 3	0 + 3	0 + 3	0 + 3	0 + 3	0 + 3	—	—	—	—	3
Guangzhou	0 + 3	0 + 3	0 + 3	0 + 3	0 + 3	0 + 3	0 + 3	0 + 3	0 + 3	0 + 3	—	—	—	—	3
Shanghai	0 + 3	0 + 3	0 + 3	0 + 3	0 + 3	0 + 3	0 + 3	0 + 3	0 + 3	0 + 3	—	—	—	—	3

(e) Chat client blacklist [14] (non-traditional characters) (15 459 words).

Client \ Server	London	Mumbai	Paris	San Jose, CA (1)	San Jose, CA (2)	Singapore	Taiwan	Tokyo	HK (1)	HK (2)	Beijing (1)	Beijing (2)	Guangzhou	Shanghai	Total
Pittsburgh, PA	—	—	—	—	—	—	—	—	—	—	180 + 1162	238 + 1120	52 + 1014	—	2274
Hong Kong (1)	—	—	—	—	—	—	—	—	—	—	65 + 1149	68 + 1122	66 + 1156	36 + 675	1230
Hong Kong (2)	—	—	—	—	—	—	—	—	—	—	68 + 949	53 + 638	38 + 707	64 + 1155	1308
Beijing (1)	59 + 1135	62 + 1120	64 + 1136	64 + 1147	65 + 1154	64 + 1143	6 + 115	20 + 639	66 + 1155	20 + 437	—	—	—	—	1221
Beijing (2)	65 + 1154	64 + 1152	45 + 930	65 + 1152	65 + 1154	65 + 1139	66 + 1143	63 + 1120	66 + 1155	61 + 1072	—	—	—	—	1223
Guangzhou	65 + 1154	64 + 1143	65 + 1153	65 + 1154	64 + 1139	62 + 1143	58 + 1094	54 + 1071	59 + 1119	65 + 1154	—	—	—	—	1219
Shanghai	4 + 88	44 + 829	49 + 801	65 + 1154	26 + 442	62 + 1106	19 + 404	93 + 776	51 + 358	57 + 1004	—	—	—	—	1354

(f) Chat client blacklist [14] (traditional characters) (1 016 words).

Client \ Server	London	Mumbai	Paris	San Jose, CA (1)	San Jose, CA (2)	Singapore	Taiwan	Tokyo	HK (1)	HK (2)	Beijing (1)	Beijing (2)	Guangzhou	Shanghai	Total
Pittsburgh, PA	—	—	—	—	—	—	—	—	—	—	18 + 56	15 + 56	0 + 35	—	124
Hong Kong (1)	—	—	—	—	—	—	—	—	—	—	0 + 40	1 + 40	0 + 40	0 + 27	42
Hong Kong (2)	—	—	—	—	—	—	—	—	—	—	4 + 35	3 + 25	0 + 26	0 + 40	48
Beijing (1)	0 + 39	0 + 39	0 + 40	0 + 40	0 + 40	0 + 40	0 + 3	0 + 27	—	0 + 20	—	—	—	—	40
Beijing (2)	0 + 40	0 + 40	0 + 32	0 + 40	0 + 40	0 + 41	0 + 40	0 + 40	0 + 40	0 + 33	—	—	—	—	40
Guangzhou	0 + 40	0 + 40	0 + 40	0 + 40	0 + 40	0 + 40	0 + 39	0 + 37	0 + 40	0 + 40	—	—	—	—	40
Shanghai	0 + 2	0 + 34	0 + 34	0 + 40	0 + 26	0 + 37	0 + 12	4 + 33	1 + 15	0 + 39	—	—	—	—	49

at all for the Pittsburgh client contacting the Shanghai server, and notably fewer keywords censored for the Shanghai and Beijing (1) clients contacting London, Paris, Taiwan, and Tokyo. This may be a function of the clients' ISPs: Shanghai and Beijing (1) are hosted by Alibaba, Beijing (2) and Guangzhou are hosted by Tencent. It may also be due to route flapping or the GFW failing open.

Chat clients blacklist keywords (trad). Table 4f shows results for the trad subset of the chat client blacklist. Another 40 terms are censored from this subset, almost all of them only with "search." The patterns seen with the preceding five lists continue here: censorship is broadly consistent, a few routes experience less censorship, and more keywords are censored for traffic inbound to China than the

	Hong Kong (2)	Shanghai	Beijing (1)	Beijing (2)	Guangzhou	Pittsburgh, PA
Hong Kong (1)	0.99	0.99	1	1	1	0.97
Hong Kong (2)		0.93	1	1	1	0.92
Shanghai			1	1	1	0.89
Beijing (1)				1	1	0.98
Beijing (2)					1	0.98
Guangzhou						0.98

Figure 2: Overlap of censored terms for each client (chat list). Gray shading highlights cells with smaller values.

reverse. Interestingly, a few of the censored terms (notably「企業倒閉潮」"wave of business failures" and「持續低迷」"continued downturn") seem to be targeting specific news articles about *economic* consequences of the 2020 coronavirus pandemic.

4.2.2 Sensitive substrings of chat keywords. Close inspection of the censored keywords derived from the chat client list reveals many repeated substrings: for instance,「坦克+六四+屠杀」(Tank + June Fourth + Massacre),「六四不平反统一不能谈」(The June 4th Incident and Unification), and「中国六四真相」(The truth about June 4th in China). all share the substring「六四」(June Fourth). This raises the question of whether a smaller set of substrings is responsible for the many censored keywords found on the chat client list. (The censored keywords from the Wikipedia 2014 and 2020 lists do not share any substrings with each other.)

We applied the keyword combination discovery algorithm developed by Xiong and Knockel [50] to identify the substrings that actually trigger censorship. This algorithm efficiently determines a set of substrings (referred to as "keyword components" in their paper) that trigger all of the same censorship events as the original set of keywords, while minimizing the number of test messages sent over the censored network route. Due to the "penalty box" described in Section 4.5, for the algorithm to complete in a reasonable amount of time, it is essential to use as few messages as possible.

Using this algorithm, we found that just 68 different keyword components are responsible for the censorship of all 1 221 keywords found to be censored in traffic originating from the Beijing (1) client.「六四」(June Fourth) alone was found to be responsible for more than half of the censored keywords.

4.2.3 Overlap between clients. To quantify how inconsistent the GFW is from route to route, Figure 2 shows the *overlap* between the complete sets of censored keywords observed from each client, starting from the chat list (Table 4e). The overlap of two sets is defined as the size of the intersection of the sets over the size of the smaller of the two:

$$\mathrm{o}(A, B) = \frac{|A \cap B|}{\min(|A|, |B|)}$$

It ranges from 0 for completely disjoint sets, to 1 for identical sets. Overlap is a symmetric statistic; if, for instance, one set is a superset of the other, it does not show which is which.

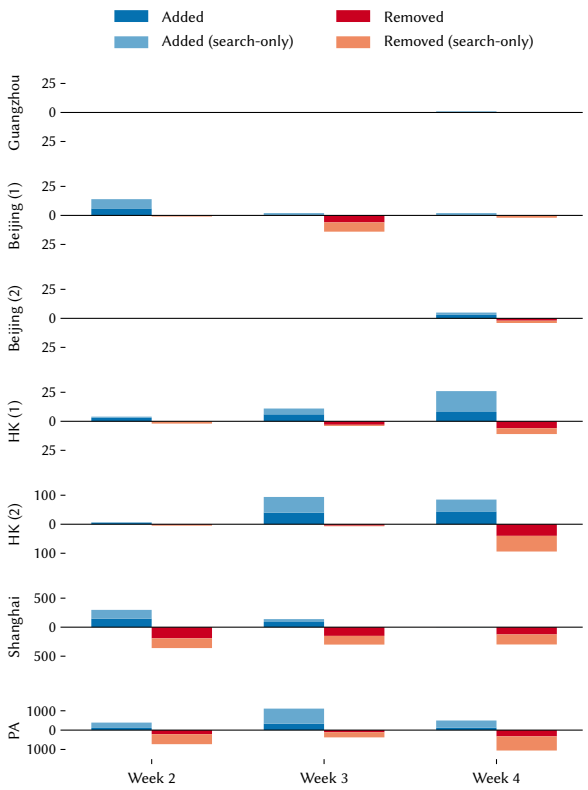

Figure 3: Addition and removal of keywords targeted by the GFW over a 1-month period. Based on repeated measurements of the lists considered in Section 3.1, excluding traditional Chinese terms.

Figure 2 shows that the four clients within mainland China all observe censorship of the same keywords, although not necessarily on all outbound routes. In contrast, the clients outside China—Hong Kong (1) and (2), and Pittsburgh—see censorship of a somewhat larger set of keywords, and do not agree with each other.

4.3 Longitudinal measurements

The experiments in Sections 4.1 and 4.2 only reveal the keywords censored by the GFW at a particular instant in time. To begin to understand how the censorship policy changes over time, we repeated the measurements described in Section 4.2.1 weekly for one month, for non-trad sublists only.

Figure 3 depicts a coarse overview of the changes we observed from week to week. For instance, in week 2 the Shanghai client observed 300 keywords to be added to the blocklist, and 360 others removed (roughly half of each group being censored only with "search"). Note the different vertical scale for each client: over the period of the experiment, the Guangzhou client observed only one or two changes to the blocklist, while the Beijing clients and Hong Kong (1) observed tens of changes each week, Hong Kong (2) and Shanghai observed hundreds, and Pittsburgh almost 1 000. The median number of weekly additions is 270 (440 including "search"-only keywords) and the median number of weekly removals is 400

Table 5: Structure of HTTP requests used in Section 4.4, showing one way to transmit the keyword「什么什么」in each field.

Component	Typical contents	Example with encoded, armored keyword	Armor format
Request line	`GET /`*path* `HTTP/1.1`	`GET /search?k=%E4%BB%80%E4%B9%88%E4%BB%80%E4%B9%88 HTTP/1.1`	%-coded [6]
Headers	`Host:` *domain name*	`Host: xn--6iqa27ab.example`	IDNA [26]
	`Cookie:` *cookies*	`Cookie: =?utf-8?q?k=3D=E4=BB=80=E4=B9=88=E4=BB=80=E4=B9=88?=`	qp-coded [32]
	`X-Tension:` *anything*	`X-Tension: =?utf-8?b?az3ku4DkuYjku4DkuYg=?=`	base64 [32]
Body	*arbitrary data*	`formfield=什么什么`	bare

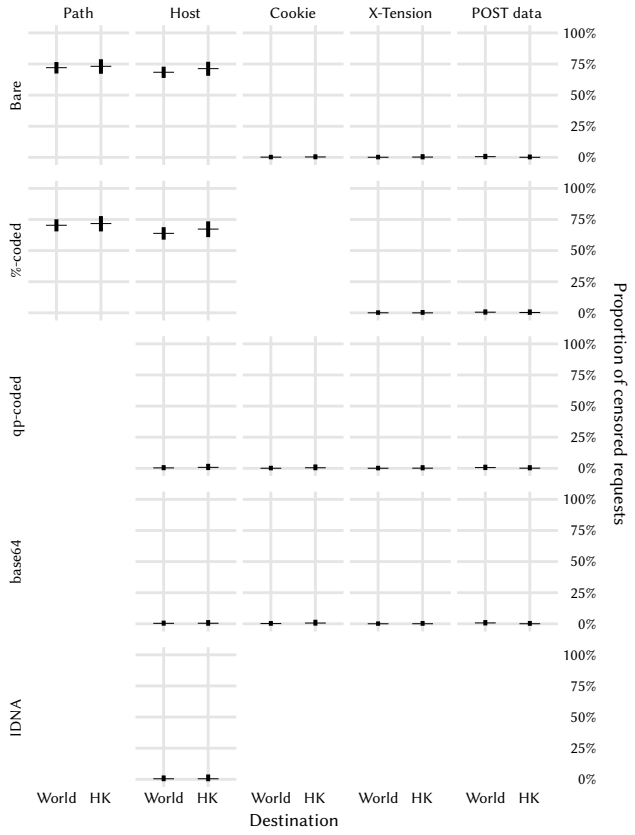

Figure 4: Where the GFW can detect keywords. Proportion of censored HTTP requests, as a function of the keyword position, armor format, and whether the server was in Hong Kong. Blank panels are impossible (e.g. IDNA can only encode domain names).

(680 including "search"-only). Considering that there are only about 1 200 censored keywords in total, this level of "churn" is substantial.

4.4 Where and what does the GFW look at?

Next, we describe a set of experiments aimed at determining how thoroughly the GFW parses HTTP requests by varying the location and text encoding of censored terms within a request, the destination port, and the way requests are formed.

HTTP requests, as defined in RFC 7230 [21], are divided into three components: a *request line*, any number of key-value *headers*, and an optional *body* which can contain arbitrary data (see Table 5). In the earlier experiments, we always placed sensitive keywords within the request line. In this experiment, we tested placing it in other locations instead: in the `Host` header, as a subdomain of the keyword server's domain name; in the `Cookie` header, as the value of a cookie; as the value of a custom header named `X-Tension`; and finally, in the body, as the value of a form field being submitted (`x-www-form-urlencoded` content). Each test request carried a censored keyword in only one of these possible locations. The other locations contained six random characters.

There are three commonly used encodings for Chinese text on the Web: UTF-8, which can represent all of Unicode; GB 2312, which is limited to simplified Chinese characters; and Big5, which is limited to traditional Chinese characters as used in Taiwan and Hong Kong. For each keyword, we tested each encoding that could represent it in each possible location.

RFC 7230 specifies that non-ASCII text in HTTP requests must be "armored" by re-encoding it within ASCII, using different encodings for different parts of the request, as shown in Table 5. HTTP clients are known not to conform perfectly to this part of the specification. They may send "bare" text (without armor) in any location, or they may use %-coding for headers as well as for the request line. Similarly, the GFW's packet inspection code might not implement all of the armor encodings. Therefore, we tested bare text and %-coded text in all locations, as well as standard-compliant text.

We sent these crafted requests from all of the client locations to all of the keyword server locations, using each of the keywords in Table 3, and repeated the test daily for five days. Figure 4 shows the results: there is a clear division between positions within an HTTP request that are monitored, and others that are not. Specifically, the "path" component of the request line, and the Host header, are monitored for keywords in both UTF-8 and GB 18030. %-encoding will be decoded if present, but raw keywords are recognized as well. Other locations are not monitored, and other forms of ASCII armor are not decoded. In particular, IDN-encoded hostnames will *not* be decoded, which indicates a loophole: if the site 动态网`.example` actually existed, real browsers would send it `Host: xn--6fro42adpy.example`, not `Host: %E5%8A%A8%E6%80%81%E7%BD%91.example`.

Figure 4 also shows that on average, only 75% of requests containing censored keywords in monitored positions actually trigger a disconnection. Traffic between mainland China and Hong Kong is (statistically) not treated differently than traffic between mainland China and the rest of the world.

Which ports does the GFW monitor? HTTP is officially assigned to TCP port 80, but URLs can specify a different port. If the GFW monitored only port 80 for HTTP traffic, it would be trivial to evade

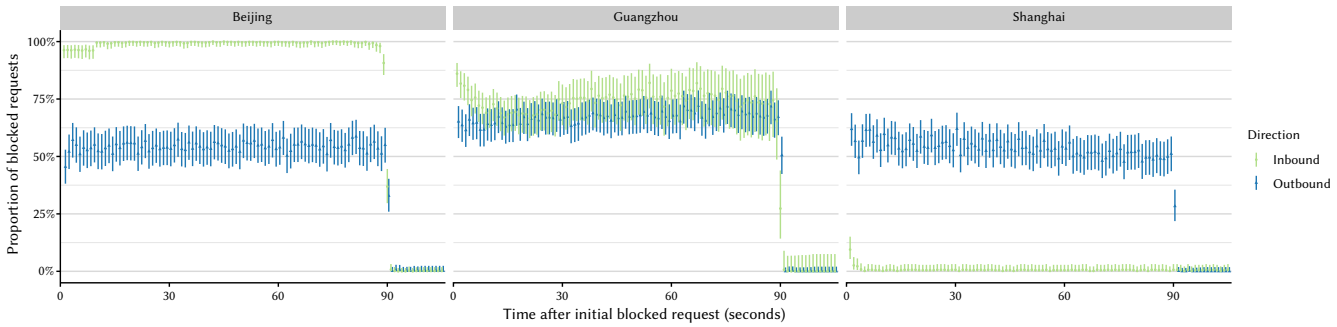

Figure 5: Penalty box behavior. For 105 seconds after an initial request censored by the GFW, the proportion of *benign* requests (containing no censored terms) to the same server and TCP port that are blocked, measured between three clients within China and a server in New Jersey, USA. Error bars are 99% binomial proportion confidence intervals with Bonferroni correction for multiple interval estimates. Missing data at the end of each time series has been replaced by extrapolations.

Table 6: Censorship of variations on a censored host name.

HOST value	Censored?	Notes
`falun.org`	N	Domain for sale
`falunda.org`	N	Nonexistent domain
`falundaf.org`	N	Nonexistent domain
`falundafa.org`	Y	Falun Gong
`en.falundafa.org`	Y	Falun Gong
`enfalundafa.org`	Y	Nonexistent domain
`falundafa.orgaa`	Y	Nonexistent domain
`falundafa.com`	Y	Falun Gong
`aaafalundafa.com`	Y	Nonexistent domain
`falundafa.net`	N	Nonexistent domain
`falungong.net`	N	Nonexistent domain
`falungong.com`	N	Nonexistent domain

by hosting a site on port 8000 instead. Previous experiments have, in general, not checked for Web censorship on ports besides port 80. Using a modification of our test client and keyword server, we scanned the entire TCP port space repeating a known-censored query. The GFW responded with reset packets on every port.

How long of a request line will the GFW process? Chu [10] reports a hard upper limit of 64 bytes on the length of a censored keyword as it appears on the wire (after character encoding and ASCII armor have been applied). We were curious whether this upper limit applies to the distance between the word "search" and a term censored only when "search" is present. We tested this with the keyword「多维」(*duō wéi*, a news site operated by the Falun Gong organization; censored only with "search") and a modified client that inserted variable amounts of padding between "search" and the keyword (e.g. `GET /search?x=aaa...aaa&k=多维`). We found that the TCP connection was reset regardless of the number of a's, up to at least 32,768 of them. If we replaced「多维」with「足球」(*zú qiú*, soccer; not censored) then no resets were injected regardless of the number of a's. We conclude that the 64-byte limit on the length of a keyword may still exist, but "search" is handled separately.

Matching of HTTP Host headers. The HTTP Host header carries the domain name of the site being accessed. To investigate whether the GFW honors the structure of domain names, we sent HTTP requests to a keyword server but modified the Host header to make it look like we were requesting a different host. We generated a number of variations on the host name `falundafa.org`, a website operated by the Falun Gong organization and known to be censored in China; some of these variations would belong to the same DNS domain, and others would not. One variation used a nonexistent top-level domain (`.orgaa`) and so could not exist at all.

Table 6 shows the results of our tests: The GFW censors any request whose Host header contains the string `falundafa.org` or `falundafa.com`, with no consideration of DNS label boundaries.

4.5 The Penalty Box

Several earlier studies [10, 11, 51] report that once the GFW terminates a connection because a censored keyword was transmitted, it will block subsequent connections between the same two hosts for a period of minutes to hours afterward, whether or not any censored terms are transmitted.

Using all of our client and server locations, we re-tested for this "penalty box" period with a modified version of our HTTP client. It sends a request that we expect to be censored to one of our keyword servers, and then sends requests that we expect *not* to be censored to all of the keyword servers, until all of them succeed ten times in a row, recording success or failure for each request, and repeating the whole procedure many times.

We can confirm the existence of the penalty box. In our tests, the penalty consistently lasts for 90 seconds but is usually not a complete blackout. During this period, the client receives TCP resets immediately after sending SYNs, but only to the same server IP and port that received a censored term.

Figure 5 shows the proportion of requests that are blocked by a penalty box on six selected routes. For routes leaving China, we observe broadly consistent behavior: 50–75% of connections are blocked for 90 seconds. For routes *entering* China, the behavior varies much more widely: nearly 100% of inbound traffic to our Beijing test server is blocked, inbound traffic to Shanghai receives no penalty box at all, and 75% of inbound traffic to Guangzhou is blocked on average but with wide variability.

4.6 HTTPS

Transport-layer encryption, as used in HTTPS, prevents inspection of an HTTP request for censored keywords. However, the first few packets of an HTTPS connection are still cleartext, and contain information that can be used for censorship, such as the hostname of the site (the "server name indication" or SNI message). The GFW is known to censor based on this information [9].

To verify the GFW's ability to censor based on the SNI, we selected two popular websites that China censors by DNS forgery: `facebook.com` and `zh.wikipedia.org`. We also selected one Chinese and one English keyword that we know to be censored in HTTP requests (「多维」and "ultrasurf") and established them as subdomains of our test domain. We modified our test client to connect over HTTPS to a keyword server, sending one of these four domain names in the SNI message but no other sensitive keywords.

We found that an SNI of `facebook.com` or `zh.wikipedia.org` would indeed cause the GFW to disrupt the connection with RSTs and then impose the penalty box. However, the other two sensitive hostnames did not cause the GFW to react. We suspect this means the blacklist used for SNI messages is separate from the blacklist of keywords, and contains only hostnames of entire sites sanctioned by China. As with HTTP, the GFW responds to forbidden SNI messages on every TCP port.

Also, it would be detectable because of changed server certificates, but the GFW *could* decrypt and re-encrypt HTTPS traffic. There have been a few previous reports suggesting that the GFW might do this under some circumstances [17, 31, 44].

We repeated all of the tests in Section 4.4 with HTTPS traffic to our keyword servers. For this test, sensitive keywords sometimes appeared in the encrypted Host header but not in the cleartext SNI message. The keyword servers were configured with two domain names each. On one domain name, they would serve a self-signed certificate; on the other, a CA-signed certificate. Our test client accepted and logged whatever certificates it received.

We found no evidence for decryption or certificate substitution. None of the encrypted traffic triggered censorship, and we always received the same certificates our servers sent. We conclude that the GFW does not *indiscriminately* decrypt HTTPS.

4.7 Other findings

We briefly experimented with these additional test conditions. All of the results in this section should be considered preliminary, and a guide for future work.

IPv6. IPv6 is not yet widely deployed in China. Only one of our client locations within mainland China could send or receive IPv6 traffic at all. Its hosting provider dropped support for IPv6 halfway through the main experiment. While it had IPv6, we observed roughly the same censorship as with IPv4—the same subset of the Wikipedia 2014 list was censored unconditionally, and more keywords were censored in the presence of "search." However, we did not observe the 90-second "penalty box."

Variation by domain. The domain name of each site being accessed is visible to the GFW through the `Host` header, and Chu [10] found cases where the domain name was part of the censored term. We configured each keyword server with two domain names, in different TLDs (`.net` and `.site`), and repeated our tests with both domains. One server had a third, non-ASCII domain name: `東京.example.net` as well as `tokyo.example.net`. We did not find any effect of varying the domain name. This only means that the same blacklist is applied to all sites the GFW has not been specifically configured for.

Table 7: Keywords used for the `telnet` and IRC tests.

Keyword (Pinyin)	**Meaning**	**Censored?**
足球 (zú qiú)	soccer	no
动态网 (dòng tài wǎng)	Dynamic web (proxy)	yes
盘古乐队 (pán gǔ yuè duì)	Pangu (music band)	yes
法轮 (fǎ lún)	Falun Gong	search only
多维 (duō wéi)	`dwnews.com` by Falun Gong	search only
无界 (wú jiè)	Wujie Network (proxy)	search only
花园网 (huā yuán wǎng)	Garden Net (proxy)	search only
华夏文摘 (huá xià wen zhāi)	Huaxia Digest	search only
延安日记 (yán ān rì jì)	Yan'an Diary	search only
博讯 (bó xùn)	`boxun.com`, a news website	search only
世界经济导报 (shì jiè jīng jì dǎo bào)	World Economy Newspaper	search only

Telnet and IRC. There's no reason, except perhaps lack of resources, why the GFW should *only* censor the Web. Cleartext `telnet` and IRC are still used for bulletin boards and chat rooms, which we know are a priority for Chinese censorship.

From our host in Beijing, we contacted three `telnet` servers and three IRC servers, located in the USA and Germany. Conversely, from the host in Pittsburgh, PA we contacted five `telnet` servers located in China. We sent all of the keywords in Table 7 to each server in a single packet. The `telnet` servers would echo the keywords back to the client; the IRC servers would not. We observed no censorship at all. Since the same keywords *are* censored when encapsulated in an HTTP request and sent to the TCP ports for `telnet` and IRC, the GFW must be paying some attention to the application protocol. However, this does not prove that the GFW never censors `telnet` or IRC. It could be scanning for a different list of sensitive keywords or it might be expecting a more natural protocol exchange.

5 CONCLUSIONS

Far from being an impregnable fortress, the GFW's keyword censorship of HTTP is more like a Swiss cheese: overall solid, but full of holes. Some of these may be errors or reflect communication failures among the several groups responsible for aspects of the GFW. Others may be due to deliberate trade-offs between implementation complexity and utility to the censor.

We wish to highlight four patterns which emerge from observation of these holes over the course of our experiments:

Keyword censorship is time-dependent. A few topics appear to be permanently banned, but our weekly measurements confirm previous reports that China's censored keyword list is continuously revised, with references to past news events dropped and newly controversial terms added. As other writers have argued [15, 16, 24, 48], this means the list is evidence of current political concerns, and it is critical for researchers to keep up, both with continuous monitoring of known sensitive terms, and continuous discovery of newly sensitive terms.

Keyword censorship is context-dependent. Instead of reacting to keywords in isolation, our experiments revealed that the GFW

takes context into account: position within an HTTP request matters, and some keywords are only censored in the presence of the word "search." This may be a way to limit the amount of resources required for packet inspection. It also suggests that keyword censorship is used as a complement to DNS- and IP-based blockade of entire sites: it prevents *searching* for sites unknown to the GFW.

Keyword censorship is route-dependent. We were surprised to find no censorship of traffic within mainland China. This contradicts earlier findings [49, 51], and may mean the GFW's operations have become more centralized over time. Traffic *entering* China seems to be more severely censored than traffic *leaving* China, calling into question whether experiments relying on clients outside China accurately capture the experience of Chinese Internet users.

We found varying levels of censorship depending on the source and destination of our probes, but not correlated with geography. For instance, our Taiwan and Japan hosts experienced less censorship when accessed from Beijing (1) than Beijing (2). Hong Kong hosts enjoy a special status, not censored when accessing foreign sites, but also not always censored when accessing mainland sites, with marked differences depending on the service provider.

These results suggest that variations in censorship depend on the routes that traffic follows, not the actual location of the foreign host. Unfortunately, we were not able to confirm this, because many of the intermediary routers do not respond to any form of `traceroute`, including recent variants designed to cope with modern routing [e.g., 3, 46].

Keyword censorship is protocol-dependent. The GFW ignores certain keywords sent over `telnet` and IRC even though it reacts to them within HTTP requests. This could mean the censors think it not worth the effort to censor these less-widely-used protocols. It could also mean that the list of censored keywords is different, or that we have not discovered the protocol-level contexts where these keywords would be censored.

Our experiments also revealed that the GFW *does* censor traffic sent over IPv6, but we were not able to test this comprehensively due to limited availability of v6-capable hosting.

5.1 Future directions

Our study laid the groundwork for a set of more comprehensive experiments aimed at assessing the keyword censorship capabilities of the GFW to its full extent.

We intend to continue the longitudinal study (Section 4.3) at least long enough to determine whether weekly turnover of nearly half the chat list is considered normal. With substantially more data (on the order of months to years) it should be possible to identify trending topics and correlate them with news events; however, this will also require an up-to-date source of new censored terms. Citizen Lab's list of censored chat terms is based on manual reverse engineering and only updated once or twice a year. However, it might be possible to automate this process. Searching and crawling as suggested by Darer [15, 16] is another possibility.

Once IPv6 is more widely available in China, our experiments should be repeated with that protocol. We have no reason to think China does not intend to censor IPv6 traffic just as thoroughly as it does IPv4, and we expect that this experiment will confirm the presence of censorship.

The GFW does not normally intercept HTTPS traffic and substitute its own certificates, but there might be narrow circumstances where it does. Can these be identified? The list of strings censored when they appear in an SNI message seems to be separate from the list of strings censored elsewhere and contain only hostnames; what might we find if we probed for SNI-based censorship of, say, all second-level domains within `.com`?

An extension to TLS is being developed (Encrypted Client Hello, ECH) that would encrypt the site hostname and other sensitive data that TLS currently leaves as cleartext [38]. In 2020, Bock et al. [8] reported that the GFW is preemptively blocking all use of ESNI (an earlier version of the ECH specification). We did not test ESNI or ECH as they are not widely deployed yet and the specification may still change. However, browser vendors have announced their intention to deploy ECH as soon as the specification is finalized, Therefore, extending our HTTPS tests to include as many forms of ECH as possible is a priority.

An HTTP request containing a censored keyword is blocked; a `telnet` packet containing the same censored keyword, by itself, is not. What is the GFW looking for besides the keyword? Is HTTP the only cleartext protocol China bothers to censor, or were our tests of `telnet` too artificial to catch the censorship? The answers to these questions might reveal new techniques for evasion or new phenomena similar to "search" triggering additional scrutiny.

One thing is certain, though: keyword-based censorship is alive, as shown by the presence of recent topical terms in the lists of censored keywords. We thus expect that answering any subset of the above questions would bring valuable scientific advances.

ACKNOWLEDGMENTS

We thank Jin-Dong Dong for assistance with translations, Arun Dunna for access to VPS servers in China, J. Zou for tremendous personal support, and Shinyoung Cho, Takanori Isobe, Nguyen Phong Hoang, Arian Niaki, Mahmood Sharif, Kyle Soska, and the anonymous reviewers for insightful feedback.

This research was partially funded by the National Science Foundation (USA) under award CNS-1814817, and by the Fundação para a Ciência e a Tecnologia (FCT) (Portugal) under grant SFRH/BD/136967/2018 and project UIDB/50021/2020.

REFERENCES

[1] Nicholas Aase, Jedidiah R. Crandall, Álvaro Díaz, Jeffrey Knockel, Jorge Ocaña Molinero, Jared Saia, Dan Wallach, and Tao Zhu. 2012. Whiskey, Weed, and Wukan on the World Wide Web: On Measuring Censors' Resources and Motivations. In *Free and Open Communications on the Internet*. USENIX, Berkeley, CA, Article 17, 7 pages. https://www.usenix.org/system/files/conference/foci12/foci12-final17.pdf

[2] Anonymous. 2014. Towards a Comprehensive Picture of the Great Firewall's DNS Censorship. In *Free and Open Communications on the Internet*. USENIX, San Diego, CA, 7 pages. https://www.usenix.org/conference/foci14/workshop-program/presentation/anonymous

[3] Brice Augustin, Xavier Cuvellier, Benjamin Orgogozo, Fabien Viger, Timur Friedman, Matthieu Latapy, Clémence Magnien, and Renata Teixeira. 2006. Avoiding traceroute anomalies with Paris traceroute. In *Internet Measurement Conference*. ACM, New York, NY, 153–158. DOI: 10.1145/1177080.1177100

[4] Geremie R. Barme and Sang Ye. 1997. The Great Firewall of China. *Wired* 5, 6 (June 1997), 13 pages. https://www.wired.com/1997/06/china-3/

[5] Jake Bathman. 2016–. *The 10,000 most common English words in order of frequency*. https://github.com/first20hours/google-10000-english

[6] T. Berners-Lee, R. Fielding, and L. Masinter. 2005. *Uniform Resource Identifier (URI): Generic Syntax*. RFC 3986. RFC Editor. https://www.rfc-editor.org/rfc/rfc3986.txt

[7] Kevin Bock, George Hughey, Xiao Qiang, and Dave Levin. 2019. Geneva: Evolving Censorship Evasion Strategies. In *Computer and Communications Security*. ACM, New York, NY, 2199–2214. DOI: 10.1145/3319535.3363189

[8] Kevin Bock, iyouport, Anonymous, Louis-Henri Merino, David Fifield, Amir Houmansadr, and Dave Levin. 2020. *Exposing and Circumventing China's Censorship of ESNI*. Technical Report. University of Maryland. https://geneva.cs.umd.edu/posts/china-censors-esni/esni/

[9] Zimo Chai, Amirhossein Ghafari, and Amir Houmansadr. 2019. On the Importance of Encrypted-SNI to Censorship Circumvention. In *Free and Open Communications on the Internet*. USENIX, Santa Clara, CA, 8 pages. https://www.usenix.org/conference/foci19/presentation/chai

[10] Xia Chu. 2014. Complete GFW Rulebook for Wikipedia Plus Comprehensive List for Websites, IPs, IMDB and AppStore. (2014). https://goo.gl/zKslcu

[11] Richard Clayton, Steven J. Murdoch, and Robert N. M. Watson. 2006. Ignoring the Great Firewall of China. In *Privacy Enhancing Technologies*. Springer, Berlin, Heidelberg, 20–35. DOI: 10.1007/11957454_2

[12] Jedidiah R. Crandall, Masashi Crete-Nishihata, Jeffrey Knockel, Sarah McKune, Adam Senft, Diana Tseng, and Greg Wiseman. 2013. Chat program censorship and surveillance in China: Tracking TOM-Skype and Sina UC. *First Monday* 18, 7 (June 2013), 56 pages. DOI: 10.5210/fm.v18i7.4628

[13] Jedidiah R. Crandall, Daniel Zinn, Michael Byrd, Earl Barr, and Rich East. 2007. ConceptDoppler: A Weather Tracker for Internet Censorship. In *Computer and Communications Security*. ACM, New York, NY, 352–365. DOI: 10.1145/1315245.1315290

[14] Masashi Crete-Nishihata, marmight, Jakub Dałek, Jason Q. Ng, Greg Wiseman, and Katie Kleemola. 2020. *Data related to investigation of chat client censorship*. https://github.com/citizenlab/chat-censorship

[15] Alexander Darer, Oliver Farnan, and Joss Wright. 2017. FilteredWeb: A Framework for the Automated Search-Based Discovery of Blocked URLs. In *Network Traffic Measurement and Analysis*. IEEE, Dublin, 9 pages. DOI: 10.23919/TMA.2017.8002914 arXiv:1704.07185 [cs.CY]

[16] Alexander Darer, Oliver Farnan, and Joss Wright. 2018. Automated Discovery of Internet Censorship by Web Crawling. In *Web Science*. ACM, New York, NY, 195–204. DOI: 10.1145/3201064.3201091 arXiv:1804.03056 [cs.CY]

[17] Roger Dingledine, Nick Mathewson, and Paul Syverson. 2004. Tor: The Second-Generation Onion Router. In *USENIX Security Symposium*. USENIX, San Diego, CA, 17 pages. https://www.usenix.org/conference/13th-usenix-security-symposium/tor-second-generation-onion-router

[18] Maximilian Dornseif. 2003. Government mandated blocking of foreign Web content. In *DFN-Arbeitstagung über Kommunikationsnetze*. Gesellschaft für Informatik e.V., Bonn, 617–647. arXiv:cs/0404005 [cs.CY]

[19] Roya Ensafi, David Fifield, Philipp Winter, Nick Feamster, Nicholas Weaver, and Vern Paxson. 2015. Examining how the Great Firewall discovers hidden circumvention servers. In *Internet Measurement Conference*. ACM, New York, NY, 445–458. DOI: 10.1145/2815675.2815690

[20] Roya Ensafi, Philipp Winter, Abdullah Mueen, and Jedidiah R. Crandall. 2015. Analyzing the Great Firewall of China Over Space and Time. In *Privacy Enhancing Technologies*. Sciendo, Berlin, 61–76. DOI: 10.1515/popets-2015-0005

[21] R. Fielding and J. Reschke. 2014. *Hypertext Transfer Protocol (HTTP/1.1): Message Syntax and Routing*. RFC 7230. RFC Editor. https://www.rfc-editor.org/rfc/rfc7230.txt

[22] Arturo Filastò and Jacob Appelbaum. 2012. OONI: Open Observatory of Network Interference. In *Free and Open Communications on the Internet*. USENIX, Bellevue, WA, 8 pages. https://www.usenix.org/conference/foci12/workshop-program/presentation/filast%C3%B2

[23] Devashish Gosain, Anshika Agarwal, Sahil Shekhawat, H. B. Acharya, and Sambuddho Chakravarty. 2018. Mending Wall: On the Implementation of Censorship in India. In *Security and Privacy in Communication Networks*. Springer, Cham, 418–437. DOI: 10.1007/978-3-319-78813-5_21 arXiv:1806.06518 [cs.CR]

[24] Austin Hounsel, Prateek Mittal, and Nick Feamster. 2018. Automatically Generating a Large, Culture-Specific Blocklist for China. In *Free and Open Communications on the Internet*. USENIX, Baltimore, MD, 8 pages. https://www.usenix.org/conference/foci18/presentation/hounsel

[25] Eric Joyce, Matthew Goldeck, Christopher S. Leberknight, and Anna Feldman. 2018. Apollo: A System for Tracking Internet Censorship. In *Workshop on Information Security and Privacy*. AIS, San Francisco, CA, 19 pages. https://aisel.aisnet.org/wisp2018/7

[26] J. Klensin. 2010. *Internationalized Domain Names for Applications (IDNA): Definitions and Document Framework*. RFC 5890. RFC Editor. https://www.rfc-editor.org/rfc/rfc5890.txt

[27] Klzgrad, yingyingcui, Elysion, et al. 2010. *West Chamber Project*. https://code.google.com/archive/p/scholarzhang/

[28] Jeffrey Knockel, Jedidiah R. Crandall, and Jared Saia. 2011. Three Researchers, Five Conjectures: An Empirical Analysis of TOM-Skype Censorship and Surveillance. In *Free and Open Communications on the Internet*. USENIX, San Francisco, CA, 8 pages. http://www.usenix.org/events/foci11/tech/final_files/Knockel.pdf

[29] Jeffrey Knockel, Masashi Crete-Nishihata, and Lotus Ruan. 2018. The effect of information controls on developers in China: An analysis of censorship in Chinese open source projects. In *Natural Language Processing for Internet Freedom*. ACL, Santa Fe, NM, 1–11. https://www.aclweb.org/anthology/W18-4201.pdf

[30] Jeffrey Knockel, Lotus Ruan, and Masashi Crete-Nishihata. 2017. Measuring decentralization of Chinese keyword censorship via mobile games. In *Free and Open Communications on the Internet*. USENIX, Vancouver, BC, 9 pages. https://www.usenix.org/conference/foci17/workshop-program/presentation/knockel

[31] Bill Marczak, Nicholas Weaver, Jakub Dałek, Roya Ensafi, David Fifield, Sarah McKune, Arn Rey, John Scott-Railton, Ron Deibert, and Vern Paxson. 2015. An analysis of China's "great cannon". In *Free and Open Communications on the Internet*. USENIX, Washington, DC, 11 pages. https://www.usenix.org/conference/foci15/workshop-program/presentation/marczak

[32] K. Moore. 1996. *MIME (Multipurpose Internet Mail Extensions) Part Three: Message Header Extensions for Non-ASCII Text*. RFC 2047. RFC Editor. https://www.rfc-editor.org/rfc/rfc2047.txt

[33] Jason Q. Ng. 2014–. *Sensitive Chinese keywords*. https://github.com/jasonqng/chinese-keywords/blob/master/csv/individual/gfw(gb2312).csv

[34] Kei Yin Ng, Anna Feldman, and Chris Leberknight. 2018. Detecting Censorable Content on Sina Weibo: A Pilot Study. In *Hellenic Conference on Artificial Intelligence*. ACM, New York, NY, 5 pages. DOI: 10.1145/3200947.3201037

[35] Arian Akhavan Niaki, Shinyoung Cho, Zachary Weinberg, Nguyen Phong Hoang, Abbas Razaghpanah, Nicolas Christin, and Phillipa Gill. 2020. ICLab: A Global, Longitudinal Internet Censorship Measurement Platform. In *Symposium on Security and Privacy*. IEEE, San Francisco, CA, 135–151. DOI: 10.1109/SP40000.2020.00014

[36] Jong Chun Park and Jedidiah R. Crandall. 2010. Empirical study of a national-scale distributed intrusion detection system: Backbone-level filtering of HTML responses in China. In *Distributed Computing Systems*. IEEE, Genova, Italy, 315–326. DOI: 10.1109/ICDCS.2010.46

[37] Paul Pearce, Ben Jones, Frank Li, Roya Ensafi, Nick Feamster, Nick Weaver, and Vern Paxson. 2017. Global Measurement of DNS Manipulation. In *USENIX Security Symposium*. USENIX, Vancouver, BC, 307–323. https://www.usenix.org/conference/usenixsecurity17/technical-sessions/presentation/pearce

[38] Eric Rescorla, Kazuho Oku, Nick Sullivan, and Christopher A. Wood. 2020. TLS Encrypted Client Hello. (2020). https://datatracker.ietf.org/doc/draft-ietf-tls-esni/ Internet-Draft.

[39] Will Scott, Thomas Anderson, Tadayoshi Kohno, and Arvind Krishnamurthy. 2016. Satellite: Joint analysis of CDNs and network-level interference. In *USENIX Annual Technical Conference*. USENIX, Denver, CO, 195–208. https://www.usenix.org/conference/atc16/technical-sessions/presentation/scott

[40] Andreas Sfakianakis, Elias Athanasopoulos, and Sotiris Ioannidis. 2011. Censmon: A web censorship monitor. In *Free and Open Communications on the Internet*. USENIX, San Francisco, CA, 6 pages. https://www.usenix.org/events/foci11/tech/final_files/Sfakianakis.pdf

[41] Sukhbir Singh, Arturo Filastò, and Maria Xynou. 2019. *China is now blocking all language editions of Wikipedia*. OONI. https://ooni.io/post/2019-china-wikipedia-blocking/

[42] Standardization Administration of China. 1980. 信息交换用汉字编码字符集基本集 (Chinese ideogram coded character set for information interchange). GB 2312. https://archive.org/details/GB2312-1980/

[43] Ram Sundara Raman, Prerana Shenoy, Katharina Kohls, and Roya Ensafi. 2020. Censored Planet: An Internet-wide, Longitudinal Censorship Observatory. In *Computer and Communications Security*. ACM, New York, NY, 49–66. DOI: 10.1145/3372297.3417883

[44] teawithcarl. 2013. *GitHub SSL replaced by self-signed certificate in China*. Y Combinator. https://news.ycombinator.com/item?id=5124784

[45] Benjamin VanderSloot, Allison McDonald, Will Scott, J. Alex Halderman, and Roya Ensafi. 2018. Quack: Scalable remote measurement of application-layer censorship. In *USENIX Security Symposium*. USENIX, Baltimore, MD, 187–202. https://www.usenix.org/conference/usenixsecurity18/presentation/vandersloot

[46] Kevin Vermeulen, Stephen D. Strowes, Olivier Fourmaux, and Timur Friedman. 2018. Multilevel MDA-Lite Paris Traceroute. In *Internet Measurement Conference*. ACM, New York, NY, 29–42. DOI: 10.1145/3278532.3278536

[47] Zhongjie Wang, Yue Cao, Zhiyun Qian, Chengyu Song, and Srikanth V. Krishnamurthy. 2017. Your state is not mine: a closer look at evading stateful Internet censorship. In *Internet Measurement Conference*. ACM, New York, NY, 114–127. DOI: 10.1145/3131365.3131374

[48] Zachary Weinberg, Mahmood Sharif, Janos Szurdi, and Nicolas Christin. 2017. Topics of Controversy: An Empirical Analysis of Web Censorship Lists. In *Privacy Enhancing Technologies*. Sciendo, Berlin, 42–61. DOI: 10.1515/popets-2017-0004

[49] Joss Wright. 2014. Regional Variation in Chinese Internet Filtering. *Information, Communication & Society* 17, 1 (2014), 121–141. DOI: 10.1080/1369118X.2013.853818

[50] Ruohan Xiong and Jeffrey Knockel. 2019. An Efficient Method to Determine which Combination of Keywords Triggered Automatic Filtering of a Message. In *Free and Open Communications on the Internet*. USENIX, Santa Clara, CA, 9 pages. https://www.usenix.org/conference/foci19/presentation/xiong

[51] Xueyang Xu, Z. Morley Mao, and J. Alex Halderman. 2011. Internet censorship in China: Where does the filtering occur?. In *Passive and Active Measurement*. Springer, Berlin, Heidelberg, 133–142. DOI: 10.1007/978-3-642-19260-9_14