How Great is the Great Firewall? Measuring China's DNS Censorship

Nguyen Phong Hoang[⋆†] Arian Akhavan Niaki[§] Jakub Dalek[†] Jeffrey Knockel[†]
Pellaeon Lin[†] Bill Marczak[†¶] Masashi Crete-Nishihata[†] Phillipa Gill[§] Michalis Polychronakis[⋆]

[⋆]Stony Brook University, New York, USA [§]University of Massachusetts, Amherst, USA
[†]Citizen Lab, University of Toronto, Canada [¶]University of California, Berkeley, USA

Abstract

The DNS filtering apparatus of China's Great Firewall (GFW) has evolved considerably over the past two decades. However, most prior studies of China's DNS filtering were performed over short time periods, leading to unnoticed changes in the GFW's behavior. In this study, we introduce `GFWatch`, a large-scale, longitudinal measurement platform capable of testing hundreds of millions of domains daily, enabling continuous monitoring of the GFW's DNS filtering behavior.

We present the results of running `GFWatch` over a nine-month period, during which we tested an average of 411M domains per day and detected a total of 311K domains censored by GFW's DNS filter. To the best of our knowledge, this is the largest number of domains tested and censored domains discovered in the literature. We further reverse engineer regular expressions used by the GFW and find 41K innocuous domains that match these filters, resulting in overblocking of their content. We also observe bogus IPv6 and globally routable IPv4 addresses injected by the GFW, including addresses owned by US companies, such as Facebook, Dropbox, and Twitter.

Using data from `GFWatch`, we studied the impact of GFW blocking on the global DNS system. We found 77K censored domains with DNS resource records polluted in popular public DNS resolvers, such as Google and Cloudflare. Finally, we propose strategies to detect poisoned responses that can (1) sanitize poisoned DNS records from the cache of public DNS resolvers, and (2) assist in the development of circumvention tools to bypass the GFW's DNS censorship.

1 Introduction

Among the censorship regimes on the Internet, China is one of the most notorious, having developed an advanced filtering system, known as the Great Firewall (GFW), to control the flow of online information. The GFW's worldwide reputation [49] and ability to be measured from outside the country, has drawn the attention of researchers from various disciplines, ranging from political science [24, 31, 38, 39] to information and computer science [21, 22, 41, 44, 68, 92].

Unlike many other DNS censorship approaches, the GFW is known to return globally routable IP addresses in its injected responses. Recent studies [21, 57, 59] have observed injected IP addresses belonging to popular US companies, including Facebook, Dropbox, and Twitter. The use of routable IPs is in contrast to countries such as Bahrain, Korea, Kuwait, Iran, Oman, Qatar, Thailand, or Yemen [51, 57, 65, 71, 79], where DNS censorship redirects users to blockpages that inform users about the blocked content. It is also in contrast to censors using fixed DNS responses such as NXDOMAIN [26, 70, 71, 74] or addresses from private IP ranges (e.g., 10.0.0.0/8) [19, 23, 74]. This use of globally routable IPs by the GFW has implications for censorship detection, which needs to carefully distinguish censored from legitimate DNS responses, and also makes detecting and mitigating leaked DNS responses from public resolvers non-trivial.

Despite the many previous studies that examine the technical strategies employed by the GFW, such as TCP/IP packet filtering [33, 41, 45, 73, 92] and DNS poisoning [22, 40, 46, 87], there has yet to be a large-scale, longitudinal examination of China's DNS filtering mechanism. This lack of visibility is apparent as the number of censored domains and the pool of IP addresses used by the GFW in forged DNS responses have been reported differently by previous studies [21, 22, 27, 46, 67, 74, 87, 95]. In particular, the number of fake IPs observed in poisoned responses has been increasing from nine in 2010 [27], 28 in 2011 [87], 174 in 2014 [22], to more than 1.5K recently [21]. To that end, it is necessary to have a system for continuous, long-term monitoring of the GFW's filtering policy that will provide timely insights about its blocking behavior and assist censorship detection and circumvention efforts.

In this work, we developed `GFWatch` (§3), a large-scale, longitudinal measurement platform to shed light on DNS filtering by the GFW and assess its impact on the global Internet. By building `GFWatch`, our primary goal is not only to answer the questions of *(1) how many censored domains*

are there and *(2) what are the forged IP addresses used in fake DNS responses*, but also to assess *(3) the impact of the GFW's DNS censorship policy on the global Internet*, and ultimately design *(4) strategies to effectively detect and circumvent the GFW's DNS censorship*.

Using `GFWatch`, we tested a total of 534M distinct domains (averaging 411M domains per day) and detected a total of 311K censored domains (§4). We then used the set of censored domains to design a probing method that is able to reverse-engineer the actual blocklist used by the GFW's DNS filter (§4.1). Using this list, we observed that 270K out of the 311K censored domains are censored as intended, whereas the remaining 41K domains appear to be innocuous despite matching regular expressions used by the GFW. Through our measurements, we discovered 1,781 IPv4 and 1,799 IPv6 addresses used by the GFW in forged DNS responses (§5). To the best of our knowledge, these are the largest sets of censored domains and forged IP addresses ever discovered.

We also found evidence of geographic restrictions on Chinese domains, with the GFW injecting DNS replies for domains based in China (e.g., `www.beian.gov.cn`) (§6). While previous studies attribute leakage of Chinese DNS censorship to cases where a DNS resolver's network path transits through China's network [27, 87], we found that geoblocking and cases where censored domains have at least one authoritative name server located in China are also a significant cause of pollution of external DNS resolvers (§6.1).

Based on the observed censored domains (§4) and forged IP addresses (§5), we propose strategies to effectively detect poisoned DNS responses injected by the GFW (§6.2). These techniques will not only help public DNS resolvers and other DNS-related services to sanitize tainted records (§6.2), but can also assist future development of circumvention tools to bypass the GFW's DNS censorship (§7).

2 Background

The Internet filtering infrastructure of China, allegedly designed in the late 90s under the Golden Shield project [85, 94], is a system used by the Chinese government to regulate the country's domestic Internet access. The filtering system, commonly referred to as the Great Firewall [52], consists of middleboxes distributed across border autonomous systems [22, 35, 93], which are controlled in a centralized fashion [38, 52, 85, 95]. There are several filtering modules developed to control the free flow of information at different layers of the network stack, including TCP/IP packet filtering [33, 41, 44, 72, 73, 92] and application-level keyword-based blocking [33, 52, 80, 95]. However, we focus our discussion on the DNS poisoning aspect of the GFW which is relevant to our study.

Unencrypted and unauthenticated DNS traffic is widely targeted by censorship systems to interrupt communications between users and remote destinations where censored content or services are hosted [40, 71, 74, 84, 87]. Exploiting DNS insecurity, the GFW is designed as an on-path/man-on-the-side (MotS) system which takes advantage of UDP-based DNS resolution to inject fake responses when censored domains are detected in users' DNS queries.

More specifically, when the GFW detects a DNS query for a censored domain, it will forge a response with an incorrect DNS record towards the client. Some specific domains (e.g., `google.sm`) can trigger the GFW to emit up to three forged responses [21]. As an on-path system, the GFW cannot modify or drop the legitimate response returned by the blocked domain's authoritative name server or the public resolver chosen by the client. However, since the GFW is usually closer (in terms of physical/network distance) to the client, the injected response will usually arrive ahead of the legitimate one (§7.2), thus being accepted by the client who is now unable to access the domain.

3 *GFWatch* Design

We designed `GFWatch` according to the following requirements: (1) the platform should be able to discover as many censored domains and forged IPs as possible in a timely manner. More specifically, `GFWatch` should be able to obtain and test new domain names *as they appear on the Internet*. (2) As a longitudinal measurement platform, once a domain is discovered to be censored, `GFWatch` should continuously keep track of its blocking status to determine whether the domain stays censored or becomes unblocked at some point in the future. (3) By measuring many domains with sufficient frequency, `GFWatch` is expected to provide us with a good view into the pool of forged IPs used by the GFW.

3.1 Test Domains

We are interested in the timely discovery of as many censored domains as possible because we hypothesize that the GFW does not block just well-known domains (e.g., `facebook.com`, `twitter.com`, `tumblr.com`) but also less popular or even unranked ones that are of interest to smaller groups of at-risk people (e.g., political dissidents, minority ethnic groups), who are often suppressed by local authorities [18]. Therefore, we opt to curate our test list from top-level domain (TLD) zone files obtained from various sources, including Verisign [16] and the Centralized Zone Data Service operated by ICANN [5], which we refresh on a daily basis. Using zone files not only provides us with a good coverage of domain names on the Internet, but also helps us to fulfill the first design goal of `GFWatch`, which is the capability to test new domains as they appear on the Internet.

Since TLD zone files contain only second-level domains (SLDs), they do not allow us to observe cases in which the GFW censors subdomains of these SLDs. As we show later, many subdomains (e.g., `scratch.mit.edu`,

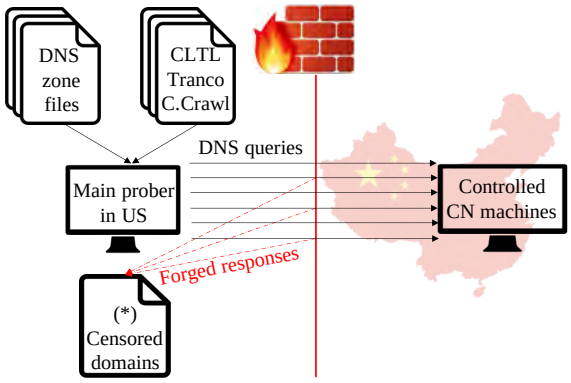

Figure 1: Probing the GFW's DNS poisoning from outside.

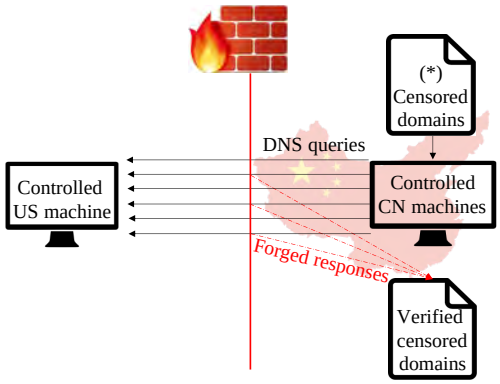

Figure 2: Verifying poisoned domains from inside the GFW.

nsarchive.gwu.edu, cs.colorado.edu) are censored but their SLDs (e.g., mit.edu, gwu.edu, colorado.edu) are not. We complement our test list by including domains from the Citizen Lab test lists (CLTL) [13], the Tranco list [66], and the Common Crawl project [14]. Between April and December 2020, we tested a total of 534M domains from 1.5K TLDs, with an average of 411M domains daily tested.

3.2 Measurement Approach

When filtering DNS traffic, the GFW does not consider the direction of request packets. As a result, even DNS queries originating from outside the country can trigger the GFW if they contain a censored domain, making this behavior a popular topic for measurement studies [21, 22, 27, 87]. Based on the observation of this filtering policy, we design *GFWatch* to probe the GFW from outside of China to discover censored domains and verify their blockage again from our controlled machines located in China to validate our findings.

Prior work has shown that the GFW does not filter DNS traffic on ports other than the standard port 53 [21, 67], we thus design our probe queries using this standard destination port number. We observe that for major UDP-based DNS query types (e.g., A, CNAME, MX, NS, TXT), the GFW injects the forged responses with an IPv4 for type A queries and a bogus IPv6 for type AAAA queries. In some rare cases, injections of forged static CNAME records are also observed for a small number of censored domains (§5.3).

For TCP-based queries that carry censored domains, RST packets are injected instead of DNS responses [91]. Since UDP is the default protocol for DNS in most operating systems, we choose to probe the GFW with UDP-based queries. While using both TCP-based and UDP-based queries would still allow us to detect censored domains, we opt to use UDP-based queries because they also allow us to (1) collect the forged IPs used in the injected DNS responses, and (2) conduct our measurement at scale, which would be otherwise more challenging to achieve because a TCP-based measurement at the same scale would require more computing and network resources to handle stateful network connections.

As shown in Figure 1, *GFWatch*'s main prober is a machine located in an academic network in the United States, where DNS censorship is not anticipated. A and AAAA DNS queries for the test domains are sent towards two hosts in China, which are under our control and do not have any DNS resolution capabilities. Therefore, any DNS responses returned to the main prober come from the GFW.

While prior studies have confirmed the centralized blocking policy of the GFW [38, 52, 85], to make sure this behavior is still consistent and to detect any future changes, the two hosts in China are located in two different autonomous systems (ASes). From our measurement results, we confirm that the DNS blocking policy continues to be centralized, with the same censored domains detected via the two probing paths.

After the main prober completes each probing batch, detected censored domains are transferred to the Chinese hosts and probed again from inside China towards our control machine, as shown in Figure 2. This way, we can verify that censored domains discovered by our prober in the US are also censored inside China.

Since *GFWatch* is designed to probe using UDP, which is a stateless and unreliable protocol, packets may get lost due to factors that are not under our control (e.g., network congestion). Moreover, previous studies have reported that the GFW sometimes fails to block access when it is under heavy load [21, 45]. Therefore, to minimize the impact of these factors on our data collection, *GFWatch* tests each domain at least three times a day.

For this paper, we use data collected during the last nine months of 2020, from April to December. As of this writing, *GFWatch* is still running and collecting data every day. The data collected will be made available to the public on a daily basis through a dedicated web service.

4 Censored Domains

Over the nine months of our study, we tested a total of 534M distinct domains, finding 311K domains triggering the GFW's

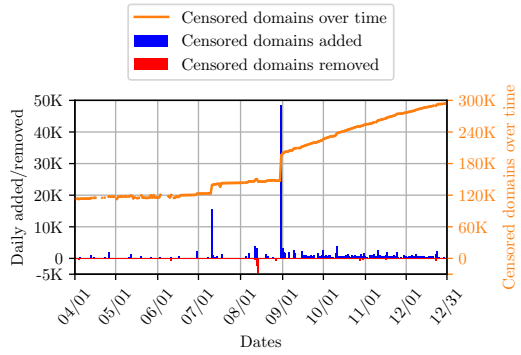

Figure 3: Cumulative censored domains discovered over time and daily added/removed censored domains.

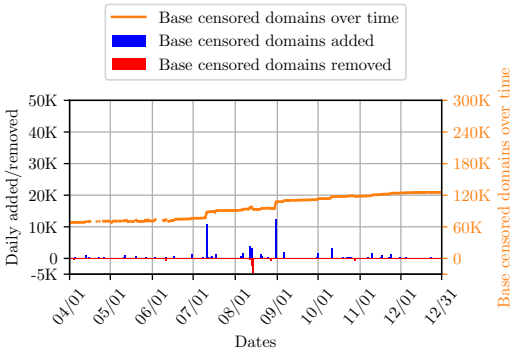

Figure 4: Cumulative base censored domains discovered over time and daily added/removed base censored domains.

DNS censoring capability. Figure 3 summarizes the cumulative number of censored domains over time, as well as the number of domains added and removed from the set of censored domains each day. We note a sharp increase in domains on August 31st because of the addition of more than 30K subdomains from the previously censored namespaces (e.g., `*.googlevideo.com`, `*.appspot.com`) to our test domains. In this section, we describe our technique for identifying the specific strings that trigger GFW's DNS censorship (§4.1). We use this technique to remove unrelated domains that match the blocking rules ("overblocked" domains) and then characterize domains censored by the GFW in Section 4.2.

4.1 Identifying Blocking Rules

When considering the domains filtered by the GFW, there are many with common second-level and top-level domains (e.g., numerous blocked domains of the form `*.blogspot.com` or `*.tumblr.com`). This observation led us to develop a clustering method for domains that are blocked based on the same underlying rule. For example, if `subdomain.example.com` and all subdomains of `example.com` are blocked, we consider `example.com` as the blocked domain. We note that when a subdomain is blocked, the covering domains may not be blocked (e.g., `cs.colorado.edu` is blocked, whereas `colorado.edu` is not (§4.2)).

Inspired by a previous study of GFW's DNS censorship [22], we use the following technique to identify the strings that trigger blocking (i.e., the most general string such that all domains containing this string are blocked). For a given domain, we test the following permutations of each censored domain and random strings:

- Rule 0 `censored_domain`
- Rule 1 `censored_domain`*`{.rnd_str}`*
- Rule 2 `censored_domain`*`{rnd_str}`*
- Rule 3 *`{rnd_str.}`*`censored_domain`
- Rule 4 *`{rnd_str}`*`censored_domain`
- Rule 5 *`{rnd_str.}`*`censored_domain`*`{.rnd_str}`*
- Rule 6 *`{rnd_str.}`*`censored_domain`*`{rnd_str}`*
- Rule 7 *`{rnd_str}`*`censored_domain`*`{.rnd_str}`*
- Rule 8 *`{rnd_str}`*`censored_domain`*`{rnd_str}`*

Among these rules, only Rules 1 and 3 are correct forms of a domain with a different top-level domain (Rule 1) or subdomain (Rule 3). In contrast, the rest represents unrelated (or non-existent) domains that happen to contain the censored domain string. We refer to censored domains that are grouped with a shorter domain string via rules other than Rules 1 or 3 as being *overblocked*, because they are not subdomains of the shorter domain, but are actually unrelated domains that are textually similar (e.g., the censored domain `mentorproject.org` contains the shorter domain string `torproject.org` that actually triggers censorship).

Using these rules to generate domains and testing them with *`GFWatch`*, we identify the most general form of each censored domain that triggers censorship. We refer to these shortest censored domains as the "base domain" from which the blocking rule is generated. We discovered a total of 138.7K base domains from the set of 311K censored domains.

Considering base domains allows us to observe growth in the underlying blocking rules as opposed to the raw number of domains. We also observe fewer new base domains over time and avoid sudden jumps in censored domains when large numbers of subdomains of an existing base domain are observed. Figure 4 shows the cumulative number of base domains discovered over the nine-month period and the daily addition and removal of these domains. As of December 31st, 126K base domains are still being censored.

Of 138.7K base domains, 11.8K are censored independently (Rule 0). In other words, these domains are censored as they are, but do not trigger GFW's DNS censorship when concatenated with random strings. However, in an ascending order of severity, we find that 4, 113.8K, 10.9K, 1.4K, and 696 distinct base domains are blocked under Rules 2, 3, 4, 6, and 8, respectively. There are no domains for Rules 1, 5, and 7, since domains blocked under these rules are already covered by other more general rules. While the vast majority of base censored domains fall under Rule 3, there are more than 13K base domains blocked under other rules, causing

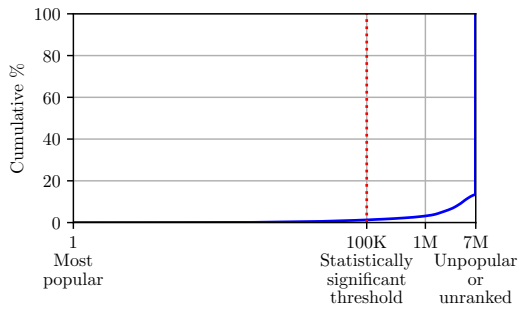

Figure 5: CDF of the popularity ranking for base censored domains (in log scale).

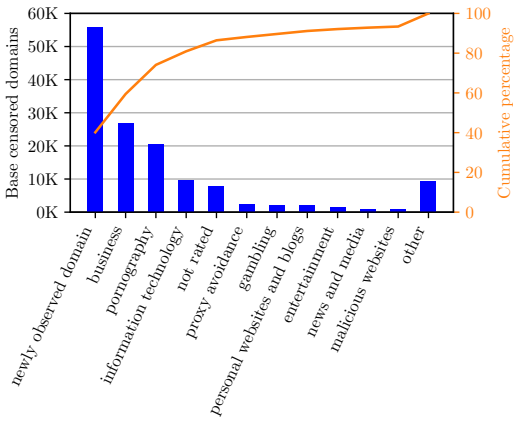

Figure 6: Top ten categories of domains censored by the GFW.

unrelated domains to be overblocked.

We utilize the base domains to identify cases of overblocking, where an unrelated domain matches a more general censored domain string. Specifically, we consider domains that match a base domain, but are not subdomains of the base domain, as being overblocked. This is because these domains are unrelated to the base domain despite being textually similar. With this definition, we find that 41K of the 331K censored domains are overblocked. The top three base domains that cause the most overblocking are `919.com`, `jetos.com`, and `33a.com`. These three domains are responsible for a total of 15K unrelated domains being blocked because they end with one of these three base domains (and are not subdomains of them). Table 4 in Appendix A provides more details on the base domains responsible for the most overblocking. Domain owners may consider refraining from registering domain names containing these base domains to avoid them being inadvertently blocked by the GFW.

4.2 Characterizing Censored Domains

We now characterize the 138.7K base domains identified in §4.1. We focus on these base domains to avoid the impact of domains with numerous blocked subdomains on our results. Focusing on base domains also allows us to avoid analyzing innocuous domains that are overblocked based on our previous analysis.

Popularity of censored domains. We find that most domains blocked by the GFW are unpopular and do not appear on lists of most popular websites. We use the rankings provided by the Tranco list [66], which combines four top lists (Alexa [1], Majestic [15], Umbrella [3], and Quantcast [10]) in a way that makes it more stable and robust against malicious manipulations [76]. The daily Tranco list contains about 7M domains ranked by the Dowdall rule [48].

Figure 5 shows the CDF of the popularity ranking for the 138.7K blocked base domains. Only 1.3% of them are among the top 100K most popular domains, which is the statistically significant threshold of the popularity ranking as suggested by both top-list providers and previous studies [20, 83]. Even when considering all domains ranked by the Tranco list, only 13.3% of the base censored domains fall within the list's ranking range, while the remaining are unranked. This finding highlights the importance of *GFWatch*'s use of TLD zone files to enumerate the set of potentially censored domains.

Types of censored content. For domain categorization, we use a service provided by FortiGuard [4], which has also been used by other censorship measurement studies [21, 71, 78], to make our analysis comparable. Figure 6 shows the top-ten domain categories censored by the GFW. We find that nearly half of the domains we observe are not currently categorized by FortiGuard, with 40% categorized as *"newly observed domain,"* and 5.5% categorized as *"not rated."* This is a result of the large number of domains in our dataset, many of which may not be currently active (§7.3).

Apart from the *"newly observed domain"* and *"not rated"* categories, we find that *"business," "pornography,"* and *"information technology"* are within the top-five dominant categories. This finding is different from the results reported by the most recent related work to ours [21], which observed *"proxy avoidance"* and *"personal websites and blogs"* as the most blocked categories. This difference stems from the counting process used in [21], which does not aggregate subdomains, while their test list is a fixed snapshot of 1M domains from the Alexa list, which contains many subdomains of `*.tumblr.com` and `*.blogspot.com`.

COVID-19 related domains. On December 19th, 2020, the New York Times reported that the Chinese Government issued instructions for suppressing the free flow of information related to the COVID-19 pandemic [81]. *GFWatch* has detected numerous domains related to COVID-19 being censored by the GFW through DNS tampering, including `covid19classaction.it`, `covid19song.info` `covidcon.org`, `ccpcoronavirus.com`, `covidhaber.net`, and `covid-19truth.info`.

While most censored domains are discovered to be blocked soon after they appear in our set of test domains, we found that there was some delay in blocking `ccpcoronavirus.com`, `covidhaber.net`, and `covid-19truth.info`. Specifically, `ccpcoronavirus.com` and `covidhaber.net` first appeared

on our test lists in April but are not blocked until July and September, respectively. Similarly, `covid-19truth.info` appeared in our dataset in September but was not censored until October. The large difference in the time the GFW takes to censor different domains shows that the blocklist is likely to be curated by both automated tools and manual efforts.

Educational domains. In 2002, Zittrain et al. [95] reported DNS-based filtering of several institutions of higher education in the US, including `mit.edu`, `umich.edu`, and `gwu.edu`. While *"education"* is not one of the top censored categories, we find numerous blocked education-related domains, including `armstrong.edu`, `brookings.edu`, `citizenlab.ca`, `feitian.edu`, `languagelog.ldc.upenn.edu`, `pori.hk`, `soas.ac.uk`, `scratch.mit.edu`, and `cs.colorado.edu`.

Although censorship against some of these domains is not surprising, since they belong to institutions well-known for conducting political science research and may host content deemed as unwanted, we are puzzled by the blocking of `cs.colorado.edu`. While the University of Colorado's computer science department is not currently using this domain to host their homepage, the blocking of this domain and its entire namespace `*.cs.colorado.edu` would prevent students in China from accessing other department resources (e.g., `moodle.cs.colorado.edu`). This is another evidence of the overblocking policy of the GFW, especially during the difficult time of the COVID-19 pandemic when most students need to take classes remotely.

5 Forged IP Addresses

The use of publicly routable IPs owned by foreign entities not only confuses the impacted users and misleads their interpretation of the GFW's censorship, but also hinders straightforward detection and circumvention [54]. Therefore, knowing the forged IPs and the pattern in which they are injected (if any) is essential. In this section, we analyze the IPs collected by *GFWatch* to examine whether there exists any specific injection pattern based on which we can develop strategies to effectively detect and bypass the GFW's DNS censorship.

5.1 Forged IP Addresses over Time

Extracting the forged IPs from all poisoned DNS responses captured by *GFWatch*, we find a total of 1,781 and 1,799 unique forged IPv4 and IPv6 addresses from poisoned type-A and type-AAAA responses, respectively. The forged IPv4 addresses are mapped to multiple ASes owned by numerous non-Chinese entities, including 783 (44%) IPs of Facebook, 277 (15.6%) IPs of WZ Communications Inc., 200 (11.2%) IPs of Twitter, and 180 (10.1%) IPs of Dropbox. On the other hand, all IPv6 addresses are bogus and belong to the same subnet of the predefined Teredo prefix [62], `2001::/32`. Therefore, we will focus our analysis on the forged IPv4 addresses

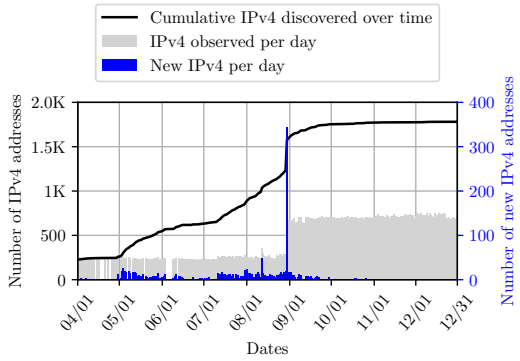

Figure 7: Number of forged IPv4 addresses detected over time by *GFWatch*.

hereafter because the pattern of IPv6 injection is obvious and thus should be trivial to detect and circumvent.

Figure 7 shows the number of unique IPv4 addresses that *GFWatch* has discovered over the measurement period considered in this paper. The gray bar plot shows the number of unique IPs observed daily, and the blue bar plot shows the number of new IPs that were not observed previously. We add a second y-axis on the right side of the figure for better visibility of the blue bars.

Our initially collected data overlaps with the data collected during the final month of [21], which is the most recent related work to our study. During this period, our observation aligns with the result reported in Figure 2 of [21], i.e., the number of unique forged IPs is about 200 with no new IPs detected. However, starting in May, *GFWatch* began to detect more forged IPs every day until September, with about 10–20 new IPs added daily. These gradual daily additions, together with a significant increase of more than 300 previously unobserved IPs at the end of August, have brought the total number of forged IPs to more than 1.5K. The number of forged IPs converges to 1.7K over the last four months of 2020.

Comparing the IPs observed by *GFWatch* with the ones reported in [21], we find that all IPs observed by [21] have been used again in poisoned DNS responses, regardless of the major drop reported on November 23rd, 2019. In addition, we find 188 new IPs that were not observed previously in [21]. Given how close the timeline is between our work and [21], this finding of the unpredictable fluctuation in the number of forged IPs emphasizes the importance of having a large-scale longitudinal measurement system to keep track of erratic changes in the GFW's blocking behavior. Therefore, we are committed to keeping *GFWatch* running as long as possible, rather than just creating it as a one-off effort.

Prior reports [38, 52, 85] and our detection of the same censored domains via two different network paths (§3) have confirmed the centralized blocking policy of the GFW in terms of the domains being censored. Nevertheless, we are also interested in investigating whether the forged IPs are consistent at different network locations, because our ulti-

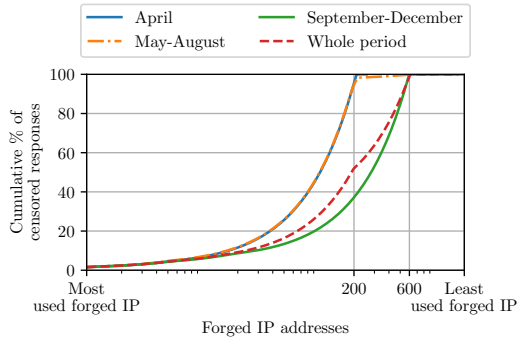

Figure 8: CDF of censored responses with respect to the injection frequency of forged IPv4 addresses detected by *GFWatch*.

mate goal is to collect as many forged IPs as possible and demystify their injection pattern to assist us in developing effective strategies for censorship detection and circumvention. Therefore, we have also conducted an extra measurement by probing across different network locations in China to confirm that the pool of forged IPs discovered by *GFWatch* is representative enough. More details of this measurement are provided in Appendix B.

5.2 Injection Frequency of Forged IPs

Due to the erratic changes in the number of forged IPs over time, prior studies have often concluded that forged IPs are injected randomly. Through the longitudinal measurement conducted at scale, *GFWatch* has tested and detected a large enough number of censored domains and forged IPs that allows us to provide more insights into this aspect. Analyzing the injection frequency of each forged IP, we find that not all forged IPs are equally injected in censored responses, i.e., their injection pattern is not entirely random.

Figure 8 shows the CDF of censored responses with respect to the injection frequency of forged IPs observed in these responses. The x-axis (in log scale) indicates the number of forged IPs, sorted by their injection frequency. There are three periods during which the cumulative number of forged IPs shows different patterns (i.e., April, May to August, and September to December, as shown in Figure 7). Thus, we analyze the injection frequency of these three periods independently and compare them with the injection frequency of all forged IPs discovered over the whole period of our measurement.

We can see that the forged IPs' injection frequencies are similar (almost overlapping) between the April and May–August lines. In other words, although the number of forged IPs increases from about 200 at the end of April to more than 1.5K over the May–August period, the initial 200 forged IPs are still responsible for 99% of censored responses. On the other hand, the additional 1.3K new forged IPs discovered from May to August are in the long tail and only used in 1% of all censored responses. Similarly, even after the remarkable increase to more than 1.7K forged IPs at the end of August, only 600 of them are frequently injected from September to December, occupying 99% of the censored responses. Finally, when looking at all the censored responses and forged IPs discovered over the whole period, the 200 most frequently injected forged IPs discovered in April are still responsible for more than 50% of all censored responses, whereas only 600 (33.6%) out of 1,781 forged IPs are responsible for 99% of all censored responses, the remaining 1.1K forged IPs in the long tail are used in only 1% of censored responses.

Table 1: Groupings of censored domains with respect to different sets of forged IPs injected in their poisoned responses.

G	# Domains	# IPs	Forged IPs/CNAMEs
0	41	0	cathayan.org, mijingui.com, upload.la, yy080.com
1	12	1	why.cc → 216.139.213.144
2	7	1	yumizi.com → 66.206.11.194
3	57	1	46.38.24.209, 46.20.126.252, 61.54.28.6, 89.31.55.106 122.218.101.190, 123.50.49.171, 173.201.216.6, 208.109.138.55
4	3,295	3	4.36.66.178, 64.33.88.161, 203.161.230.171
5	1,711	4	8.7.198.45, 59.24.3.173, 243.185.187.39, 203.98.7.65
6	2,724	4	8.7.198.46, 59.24.3.174, 46.82.174.69, 93.46.8.90
7	4	7	4.36.66.178, 64.33.88.161, 203.161.230.171, 59.24.3.174 8.7.198.46, 46.82.174.69, 93.46.8.90
8	9	7	4.36.66.178, 64.33.88.161, 203.161.230.171, 8.7.198.45 59.24.3.173, 243.185.187.39, 203.98.7.65
9	4,551	10	23.89.5.60, 49.2.123.56, 54.76.135.1, 77.4.7.92 118.5.49.6, 188.5.4.96, 189.163.17.5, 197.4.4.12 249.129.46.48, 253.157.14.165
10	remaining $\sim 300K$ domains	>560	[Omitted due to the large number of forged IPs] Supplementary data will be made publicly available and updated on a daily basis.

5.3 Static and Dynamic Injections

One of the GFW behaviors is injecting different sets of forged IPs for different groups of censored domains. This behavior was first reported in [21], where the authors identify a total of six groups of censored domains that are poisoned with different sets of forged IPs. From data collected by *GFWatch*, we have discovered a total of 11 groups shown in Table 1. Comparing these groups with those reported in [21], we find five similar groups that have the same set of forged IPs/CNAMEs, including Groups 0, 4, 5, 6, and 9. Understandably, we discover more groups because our test list covers far more domains compared to [21], where a fixed Alexa top list of only 1M domains was used for the whole measurement period.

An instance of forged response containing a CNAME was reported in [21] but excluded from the analysis since it did not seem to be prevalent. However, with a larger dataset, we find that the injection of CNAME in forged responses can happen in three different groups of censored domains, triggering the GFW to inject six different CNAME answers. As depicted in Table 1, there are 41 censored domains that can trigger the injection of *either one of the four* CNAMEs listed. Domains in Groups 1 and 2 can trigger a CNAME injection, accompanied

by an IP in the forged response. Note that these two IPs are not the actual IPs of the two CNAMEs. Similarly, there are eight distinct subgroups of domains within Group 3 that can constantly trigger *either one of the eight* forged IP listed. For example, `qcc.com.tw` will always trigger a forged response of `89.31.55.106`. The same pattern applies in other Groups from 4 to 9, i.e., resolving domains within these groups will always trigger the GFW to inject one of the forged IPs listed on the 4th column. The remaining of about 300K censored domains are grouped together since they trigger the GFW to dynamically inject a much larger number of more than 560 different forged IPs.

Revealing these injection patterns for different groups of censored domains is crucial for developing an effective strategy to detect and circumvent the GFW's DNS censorship (§6). Especially, knowing whether a censored domain belongs to one of the static groups (Groups 0 to 9) or the dynamic group (Group 10) is necessary to avoid misclassifying consistent forged responses as "legitimate" (§7).

6 Censorship Leakage and Detection

The GFW's bidirectional DNS filtering behavior has been reported as the cause of poisoned DNS responses being cached by public DNS resolvers outside China, when DNS resolution paths unavoidably have to transit via China's network [57, 87]. However, in this section, we show that DNS poisoning against many domains whose authoritative name servers are located in China is another primary reason why poisoned DNS records have tainted many public DNS resolvers around the world. We then show how the datasets of censored domains and forged IPs discovered by *GFWatch* can help with detecting and sanitizing poisoned resource records from public DNS resolvers' cache.

6.1 Geoblocking of China-based Domains

On August 8th, 2020, *GFWatch* detected the blockage of `www.beian.gov.cn`, which is managed by the Chinese Ministry of Industry and Information Technology. This service allows website owners to obtain and verify their website's Internet Content Provider (ICP) license, which is obligated to legally operate their site in China. This domain has two authoritative name servers, `dns7.hichina.com` and `dns8.hichina.com`, which are hosted on 16 different IPs. However, checking against the latest MaxMind dataset [7], we find that all of these IPs are located inside China. Consequently, the DNS censorship against this domain by the GFW will cause DNS queries issued from outside China to be poisoned since all resolution paths from outside China will have to cross the GFW.

We initially attributed this blockage to an error or a misconfiguration because previous works have sometimes noticed intermittent failures in the GFW [21, 45]. Furthermore, no prior

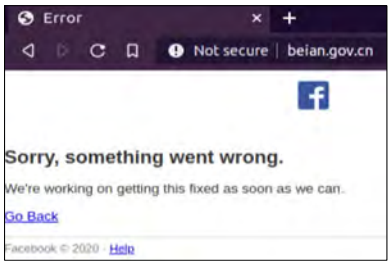

Figure 9: Visit to a domain geoblocked by the GFW ends up with an error page from Facebook.

studies have ever found such a strange blocking behavior—the GFW of China censors a Chinese government website. However, at the time of composing this paper, we are still observing `www.beian.gov.cn` being censored by the GFW, almost half a year since its first detection. Hence, this is a clear case of geoblocking because we can still visit this domain normally from our controlled machines located inside China. To the best of our knowledge, ours is the first academic research to document this geoblocking behavior of the GFW.

Note that this geoblocking is a result of the GFW's DNS censorship, which is not the same as geoblocking enforced at the server side [69]. Geoblocking of China-based websites has been noticed previously but is enforced by their website owners. For instance, political researchers have been using `https://www.tianyancha.com/` to investigate the ownership of Chinese companies, but since 2019, this website blocks visitors from non-Chinese IPs and shows a clear message for the reason of denying access.

The GFW's blocking of China-based domains using bidirectional DNS filtering in combination with the use of forged IPs owned by non-Chinese entities impacts not only Internet users in China, but also users from around the world. For instance, upon visiting the aforementioned geoblocked domain from a non-censored network outside China, we end up with an error page served from Facebook, as shown in Figure 9.

Most ordinary Internet users would not know the underlying reason why their visit to a given China-based domain (e.g., `www.beian.gov.cn`) that is clearly unrelated to Facebook would end up with an error page from Facebook. The fact that the GFW frequently changes the forged IPs used in fake DNS responses (§5) would cause even more confusion to the affected users. Depending on which fake IP is injected in the spoofed response, users may encounter a different error page from Figure 9. Even more confusing, the visit to this domain from outside China will intermittently succeed because the poisoned responses injected by the GFW sometimes fail to arrive ahead of the legitimate one (§7).

At the server side of the forged IPs being used for injecting poisoned responses, their operators would also be puzzled as to why many HTTP requests are sent to their servers, asking for hostnames they do not serve. For the above example, an error log at a Facebook server will show that someone was

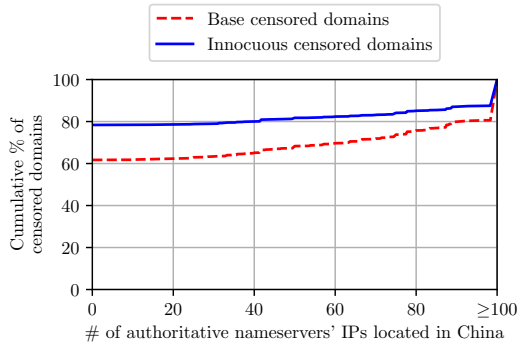

Figure 10: CDF of the number of authoritative name servers located inside China as a percentage of 138.7K base censored domains and 41K innocuously blocked domains.

trying to visit `www.beian.gov.cn` on a Facebook IP, which obviously does not serve any content for that domain, thus the returned error page. As we do not have access to the error logs of Facebook and other organizations whose IPs are used for injecting poisoned DNS responses by the GFW, we cannot quantify the actual cost (e.g., the overhead of serving unsolicited connections, error pages) of such an abusive DNS redirection behavior. However, given the large number of more than 311K censored domains discovered (§4) and only a small pool of forged IPs being used (§5), we believe that the GFW's injection policy would cost these affected organizations a non-negligible overhead on their servers. Past reports have shown that this abusive design of the GFW can lead to resource exhaustion attacks on specific IPs, making them inaccessible [34, 54, 64].

To estimate the extent to which the above geoblocking and overblocking policies have impacted the global Internet, we analyze the location of authoritative name servers of 138.7K base censored domains and 41K innocuously blocked domains, using the MaxMind dataset [7]. As shown in Figure 10, 38% (53K) of the base censored domains and 21.6% (8.8K) of the innocuous censored domains have at least one authoritative name server in China. In other words, there is always a non-zero chance that DNS resolution for these 61.8K domains from outside China will be poisoned, causing their visitors to potentially end up with an error page similar to the above case. On the other hand, 19.4% (26.9K) of base censored domains and 12.5% (5.1K) innocuously blocked domains have all of their authoritative name servers in China, meaning that the resolutions for these 32K domains from outside China will always cross the GFW, thus being poisoned.

6.2 Detection

A common operational mechanism of DNS censorship is that the censor takes advantage of the time-honored property of UDP-based DNS resolution to inject poisoned responses, racing against the legitimate response. Depending on the censored domain being queried, the GFW can even emit up to three responses. This behavior of injecting multiple fake responses was first reported recently in [21]. For the completeness of our investigation, we have also identified the three different injectors based on the data collected by *GFWatch*, with more detailed analysis in Appendix C.

Table 2: Top ten public DNS resolvers with the highest number of censored domains whose poisoned resource records have polluted their cache.

# Domains	Resolver	# Domains	Resolver
74,715	Google	63,295	OpenDNS
71,560	Cloudflare	62,825	Comcast
65,567	OpenNIC	56,913	CleanBrowsing
65,538	FreeDNS	56,628	Level3
64,521	Yandex	55,795	Verisign

From the GFW's perspective, the injection of multiple fake responses not only increases the chance of successfully poisoning a censored client but also makes it more costly and challenging to detect and circumvent its DNS censorship [40]. However, based on the pool of forged IPs and their injection patterns that we have revealed in §5, detecting DNS censorship by the GFW can be done effectively by checking the returned IP address against the pool of forged IPs discovered by *GFWatch*. Although this strategy may not detect all poisoned responses due to some rare forged IPs that *GFWatch* might have not observed in the long tail, from the analysis of injection frequency in §5.2, which we have also verified its consistency across different network locations (Appendix B), we are confident that this detection technique can identify more than 99% of the poisoned responses.

We next employ this detection technique to expose poisoned resource records that have tainted public DNS resolvers around the world. In particular, once a censored domain is detected by *GFWatch*, we query them against popular DNS resolvers and examine if its response matches any injection pattern we have revealed in §5. Table 2 shows the top ten resolvers that have been polluted with the highest number of censored domains. In total, we find 77K censored domains whose poisoned resource records have polluted the cache of all popular public DNS resolvers that we examined. Of these censored domains, 61K are base censored domains. This result aligns well with our earlier speculation in §6.1.

This finding shows the widespread impact of the bidirectional blocking behavior of the GFW, necessitating the operators of these public DNS resolvers to have an effective and efficient mechanism to prevent these poisoned resource records from polluting their cache, to assure the quality of their DNS service. Furthermore, the 61K base censored domains whose DNS queries from outside China are censored is likely the reason why many censored domains are classified as *"newly observed domain"* or *"not rated"* in §4.2. This is because FortiGuard's crawlers, which are likely located outside China, probably could not obtain the correct IPs of these domains, thus failing to fetch and classify them.

7 Circumvention

We now show how insights gained from analyzing the censored domains (§4) and forged IPs discovered by *GFWatch* over time (§5) can assist us in developing strategies to effectively and efficiently circumvent GFW's DNS censorship.

7.1 Strategy

The GFW's bidirectional DNS filtering not only impacts in-China users but also prevents users outside China from obtaining legitimate resources records of geographically restricted domains based in China (§6.1). Therefore, an effective DNS censorship evasion strategy would benefit not only (1) users inside China who need to access censored domains hosted outside China, but also (2) users outside China who need access to geoblocked domains based in China. Both (1) and (2) also include open DNS resolvers located at both sides of the GFW that want to prevent poisoned responses from polluting their DNS cache.

Since the GFW operates as an on-path injector and does not alter the legitimate response from the actual DNS resolver chosen by a client, a circumvention strategy for the client is to not quickly accept any returned responses when querying a censored domain. Instead, the client should wait for an adjustable amount of time for all responses to arrive, as suggested in [40]. Upon receiving more than one IPv6 answer, the client can filter out the bogus ones that belong to the Teredo subnet `2001::/32`. Furthermore, for IPv4 answers, the client can check them against the injection patterns and forged IPv4 addresses discovered in §5.

In our circumvention strategy, for each censored domain we need at least a trustworthy resolver that possesses its genuine resource record(s). Popular open resolvers (e.g., `8.8.8.8`, `1.1.1.1`) are often considered as trustworthy sources when it comes to censorship evasion. However, we have shown that the vast majority of public DNS resolvers have been polluted with poisoned resource records (§6.2). Therefore, we opt not to use them in this case, especially for obtaining the legitimate resource records of geoblocked domains based in China. The only remaining source that is immune to the GFW's poisoned responses and has a given censored domain's genuine resource record(s) is its authoritative name servers. This information is available in the zone files.

We send DNS queries for 138.7K base censored domains and 41K innocuous domains to their authoritative name servers from our controlled machines located at both sides of the GFW. We then expect to observe both censored and non-censored resolutions at two sides of the GFW as a result of this experiment. More specifically, from our US machine, resolutions for domains whose authoritative name servers are located outside China will not be censored as their queries will not cross the GFW, whereas resolutions for domains whose authoritative name servers are located inside China are expected to be censored. On the contrary, resolutions from our China machine towards authoritative name servers located inside China will not be censored, while those queries sent to authoritative name servers outside China will.

7.2 Evaluation

To evaluate the effectiveness of our method, we apply the proposed circumvention strategy to filter out poisoned responses for those censored resolutions and retain their "legitimate" responses, which we then compare with actual legitimate responses returned from non-censored resolutions conducted at the other side of the GFW. We find that our circumvention strategy is highly effective, with an accuracy rate of 99.8%. That is, 99.8% of responses classified as "legitimate" match the actual legitimate responses obtained from non-censored resolutions. From a total of 1,007,002,451 resolutions that the GFW poisons, 1,005,444,476 responses classified as "legitimate" by our strategy contain the same resource records (i.e., same IPs, CNAMEs, or IPs under the same AS for domains hosted on Content Delivery Networks) with those observed from non-censored resolutions. As discussed in §5.2, there are a small number of cases that we could not classify due to the invisibility of those rarely injected forged IPs in the long tail that *GFWatch* did not observe. This finding highlights the importance of having an up-to-date and continuous view into the pool of forged IPs for effectively circumventing the GFW's DNS censorship.

To further assist in future adoptions of our strategy so that it will not significantly downgrade the normal performance of other UDP-based DNS resolutions for non-censored domains, we analyze the hold-on duration, which the client should wait *only* when resolving a censored domain, instead of holding on for every resolution.

Figure 11 shows the cumulative distribution of the delta time between the first forged response and the legitimate one. The (red) dash line is the CDF of the delta time measured at our China machine, and the (blue) solid line is the CDF of this delta time measured at our US machine. On the x-axis, a positive value means a poisoned response arrives before the legitimate one. In contrast, a negative value indicates that the legitimate response has arrived ahead of the fake ones.

As shown in the figure, the GFW can successfully poison more than 99.9% of all resolutions that carry censored domains, performed from our China machine towards authoritative name servers located outside China. 99% of poisoned responses hit our machine within 364ms ahead of the legitimate ones. Although this delta time may vary, depending on the relative distance between the client and the GFW, for any client whose network location is close to ours, this is the amount of extra time they should wait when resolving a censored domain from inside China. In other words, upon receiving a DNS response after querying a censored domain, the client should wait, at most, an extra 364ms for the legitimate

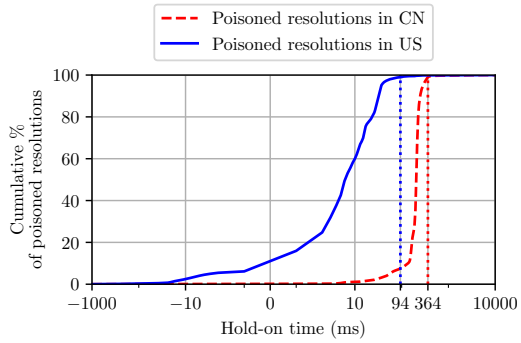

Figure 11: CDF of delta time between forged and legitimate responses measured from CN and US controlled machines.

one to arrive. Users at different locations can heuristically probe known censored domains to estimate the hold-on duration that is representative for their location.

From the GFW's perspective, forged responses should ideally arrive at the client before the legitimate one. From our US machine, we find that this is not always true. Due to the unreliable and stateless nature of UDP packets that might get lost or delayed when transferred between two distant locations, and perhaps poisoning users outside China is not the primary design goal of the GFW, 11% of the poisoned responses arrive at our US machine after the legitimate ones. Nevertheless, the remaining 89% of fake responses still hit our machine within 94ms ahead of the legitimate ones. This result again highlights the importance of having a representative dataset of forged IPs used by the GFW to effectively circumvent its DNS censorship. Especially when fake responses arrive later, our dataset of forged IPs is necessary to avoid misclassifying the legitimate ones arriving ahead as "poisoned".

7.3 Analysis of True Resource Records

Now that we have successfully obtained the legitimate resource records of the 138.7K base censored domains and 41K innocuously blocked domains, we next analyze them to better understand the impact of blocking these domains. As shown in Table 3, 120K (86.8%) base censored domains have either an IPv4, IPv6, or CNAME resource record. In other words, the remaining 18.7K (13.2%) of the base censored domains that currently do not have any resource records, indicating their inactivity. This is also one of the reasons why we observe a large number of domains classified as *"newly observed domain"* and *"not rated"* categories in §4.2.

For the innocuously blocked domains, the actual impact of GFW's overblocking may not be as severe because only 25.6K (62.5%) of them have at least one resource record. While the presence of resource records can be a sign of (in)activeness for a given domain, it does not guarantee that a domain is actively hosting any contents or services since a resource record can also be used for redirecting visitors to a domain-parking site. Therefore, the total number of domains with

Table 3: Breakdown of true resource records of base censored domains and innocuously blocked domains.

	Base censored domains		Innocuously blocked domains	
# of domains by NS location	≥1 CN NS	Non-CN NS	≥1 CN NS	Non-CN NS
	53.1K (38.3%)	85.6K (61.7%)	8.9K (21.6%)	32.1K (78.4%)
IPv4	29K (21.1%)	69.5K (50%)	6K (14.7%)	17.8K (43.5%)
IPv6	1.3K (1%)	28K (20.2%)	0.1K (0.3%)	2.8K (7%)
CNAMEs	**31K (22.3%)**	3.6K (2.6%)	**2.9K (7.1%)**	0.5K (1.3%)
# of domains with RR(s)	120K (86.8%)		25.6K (62.5%)	

resource records shown in Table 3 should be viewed as an upper bound of the actual number of domains that are actively hosting any content or service. As part of our future work, we plan to visit all of these domains using their true resource records and further investigate the contents hosted on them.

Another focal point of Table 3 is the significantly high number of CNAME resource records of both base censored domains and innocuously blocked domains that have at least one authoritative name server located in China, compared to domains whose authoritative name servers are located outside China. As far as we are aware, this is because of a common workaround that is widely suggested and used by domain owners who want to serve their websites to users at both sides of the GFW since these CNAMEs are not filtered by the GFW.

8 Discussion

In this section, we discuss the limitations of our study and provide suggestions for involving parties that are impacted by the GFW's DNS censorship.

8.1 Limitations

In order to compare our analysis on the categories of censored domains with prior studies, we choose to use a common classification service provided by FortiGuard [4]. However, we discovered that the GFW's overblocking and geoblocking policy could have already impacted this service (§6.2). Moreover, Vallina et al. [89] have shown that different classification services could result in different views of the domains being categorized. We thus tried to obtain additional classification services from two other vendors, namely, McAfee and VirusTotal. However, we were told by McAfee [8] that they only provide the service for business customers, and VirusTotal [17] did not respond to our requests.

Similar to other studies in remote censorship measurement [78, 79, 90], packets sent from our measurement infrastructure may get blocked or discriminated by the GFW. However, over the course of more than nine months operating `GFWatch`, we did not experience any disruptions caused by such discriminative behaviors, as is evident by the consistency observed between the data collected by `GFWatch` and across different network locations (Appendix B). Moreover, as part of our outreach activities, we have also received confirmations from local Chinese advocacy groups and owners of censored

domains detected by GFWatch when reaching out to these entities to share our findings. Nonetheless, if our measurement machines ever gets blocked, we can always dynamically change their network location.

Finally, we develop *GFWatch* as a measurement system to expose the GFW's blocking behavior based on DNS censorship. However, this is not the only filtering technique used by the GFW; censorship can also happen at other layers of the network stack, as previously studied [33, 41, 45, 52, 73, 92, 95]. Although prior works have shown that some websites could be unblocked if the actual IP(s) of censored domains can be obtained properly [30, 57], securing DNS resolutions alone may not be enough in some cases because blocking can also happen at the application layer (e.g., SNI-based blocking [30], keyword-based filtering [80]) or even at the IP layer [58, 60], regardless of potential collateral damage [61].

Nonetheless, DNS is one of the most critical protocols on the Internet since almost every online communication starts with a DNS lookup. We believe that continuously monitoring the GFW's filtering policy at this layer is necessary and important to timely inform the public of the erratic changes in China's information controls policies, both from technical and political perspectives. Appendix D provides some examples of domains censored due to political motivations.

8.2 Suggestions

GFW operators. Although the widespread impact of the GFW's DNS filtering policy is clear, as shown throughout this paper, we are not entirely certain whether this censorship policy is intentional or accidental. While prior works have shown intermittent failures of the GFW [21, 45], all geoblocking of China-based domains and overblocking of innocuous domains discovered by *GFWatch* have lasted over several months. This relatively long enough period of time leads us to believe that the GFW's operators would have clearly known about the global impact of their DNS filtering policy. By exposing these negative impacts on several parties outside China to the public, we hope to send a meaningful message to the GFW's operators so that they can revise their DNS filtering policy to reduce its negative impacts beyond China's borders.

Public DNS resolvers. Poisoned DNS responses have widely polluted all popular public DNS resolvers outside China due to the geoblocking and overblocking of many domains based in China (§6). DNSSEC [43] has been introduced to assure the integrity and authenticity of DNS responses for more than two decades to address these problems. However, DNSSEC is not widely adopted because of compatibility problems and technical complications [32, 36, 56]. To this end, public DNS resolvers can use the strategy introduced in §7 to prevent poisoned DNS responses spoofed by the GFW from tainting their cache. By waiting for all responses to arrive and comparing the answers with the pool of forged IPs discovered by *GFWatch* (§5), public DNS resolvers can filter out 99% of poisoned responses by the GFW. Note that it is not always necessary to wait for all responses to arrive because the GFW does not censor all domains. As we will make both censored domains and forged IPs publicly available and update them on a daily basis, these datasets can be used to decide whether to wait or not when resolving a given domain. This way, public DNS resolvers would be able to prevent poisoned responses from polluting their cache, assuring the quality of their DNS service while avoiding any downgrades of normal performance when resolving domains that are not censored.

Owners of forged IPs. Legitimate owners of forged IPs may try to avoid hosting critical services on these IPs as their resources may be saturated due to handling unsolicited TCP and HTTP(s) requests, as shown in §6.1. Currently, we do not find evidence that the GFW is using these forged IPs as a way to saturate computing resources of the infrastructure behind them since there are more than 1.7K forged IPs in the pool (§5.1) and most of them are dynamically injected (§5.2). However, a previous report of the Great Cannon [68] has shown that China is willing to weaponize the global Internet to mount resource exhaustion attacks on specific targets. With DNS censorship, the GFW can adjust its injection pattern to concentrate on a handful of forged IPs, resulting in a large amount of requests towards these targeted IPs and thus saturating their computing resources [34, 54, 64].

Domain owners. Using our dataset of censored domains, domain owners can check whether their domain is censored or not, and censored due to intended blocking or overblocking. Unless the GFW's operators revise their blocking rules, future domain owners should try to refrain from registering domains that end with any overblocking patterns discovered in §4.1 to avoid them being inadvertently blocked by the GFW.

End users. Despite the large number of censored domains discovered by *GFWatch*, different Internet users may be interested in different subsets of these censored domains, but not all. As an immediate countermeasure to the GFW's DNS censorship, we will make the legitimate resource records of censored domains obtained in §7 publicly available on a daily basis. This way, impacted users can look up and store legitimate resource records for particular censored domains in their system's `hosts` file to bypass the GFW's DNS censorship. Alternatively, a censorship-circumvention component of software can implement the hold-on strategy (§7) and gather records based on the client's location. In case the client cannot access the sanitized data published by *GFWatch*, another client-side strategy is to send two back-to-back queries. Depending on whether a censored domain belongs to the dynamic or static injection groups (§5.2), the client can discern which responses are legitimate. Since the majority of censored domains are poisoned with dynamic IPs, the client can classify the legitimate responses, which typically point to the same IP (due to back-to-back queries) or the same AS. This way, the software only needs to know whether its intended

domains are poisoned with static or dynamic IPs. To this end, continuous access to GFWatch's data is not necessary for this strategy to work, while fresh records can still be obtained.

9 Related Work

In addition to [21], which is the most recent work related to ours that we have provided in-depth discussions throughout our paper, some other one-time studies have also looked into the DNS censorship behavior of the GFW in the past [22, 27, 67, 87, 88, 95]. While China's GFW may not be the primary and sole focus, there are platforms actively measuring censorship around the globe that may also have a partial view into the GFW's DNS censorship behavior [47, 71, 78]. To provide our readers with a complete view of these efforts and highlight how our study is different from them, we summarize the major differences among these studies in this section. A more detailed comparison table can be found in Appendix E.

In its early days, the GFW only used a handful of forged IPs [67, 95]. However, later studies have noticed an increase in the number of forged IPs, from nine in 2010 [27], 28 in 2011 [87], 174 in 2014 [22], to more than 1.5K recently [21]. Except for [87] and [22] whose authors preferred to remain anonymous and the dataset URLs provided in their papers are no longer accessible, we were able to obtain data from other studies for comparison (Table 5). A common drawback of these studies is that their experiments are conducted only over limited time periods and the test domains are also static, i.e., obtained from a snapshot of Alexa top list or zone files.

To address this drawback of previous one-off studies, longitudinal platforms have been created to measure censorship around the world, including ICLab [71], OONI [47], and Censored Planet [78]. To reduce risks to volunteers and observe interferences at multiple layers of the network stack, ICLab [71] chooses commercial VPNs as vantage points for their measurement. However, this design choice limits their visibility into China as commercial VPNs are restricted in the country [25, 29]. With different approaches, OONI [47] recruits volunteers to participate in censorship measurements, whereas Censored Planet [78] employs a series of remote measurement techniques to infer censorship. These design choices allow the two later platforms to obtain vantage points located in China for their measurements. We fetch data collected during the same period of our study available on these projects' websites for comparison.

For OONI data, we first gather measurements conducted by volunteers in China that are flagged as "DNS inconsistency" [9]. To reduce false positives due to domains hosted on CDNs, we filter out those cases where controlled and probed responses have different IPs but belong to the same AS. After sanitization, we find 710 forged IPs from OONI data, 593 of which are in common with those observed by *GFWatch*. Examining the different cases, we find that there are still misclassified cases due to domains hosted on popular CDNs whose network spans across different AS numbers.

For Censored Planet [78], we use data collected by the Satellite [84] module for comparison since it is designed to measure DNS-based network interference. Satellite infers DNS censorship by comparing responses received from open DNS resolvers with ones obtained from a control resolver, along with other metadata such as AS number, HTTP static content, and TLS certificates. Since Satellite's data is not annotated with geographical information, we use different geolocation datasets [6, 7, 37, 63] to confirm the location of open resolvers used by Satellite. We then extract responses from open resolvers located in China that are flagged as "`anomaly`". We find a total of 2.4K forged IPs reported by Satellite, 1.6K of which are in common with ours. The difference in the number of forged IPs in this case, is due to the inherent nature of Satellite's measurement approach of using open DNS resolvers. In particular, about 600 IPs observed by Satellite, but not *GFWatch*, belong to Cisco OpenDNS, which provides DNS-based network filtering services for various customer types, ranging from home to business users [2]. From a detection point of view, these censorship cases are valid, but due to different local policies of these open resolvers, instead of country-level censorship enforced by the GFW.

A shared property of OONI and Satellite is that measurement vantage points (volunteers' devices and open resolvers) are not owned by these platforms. Therefore, only a limited number of domains can be tested with adequate frequency to avoid saturating these vantage points' computing resources. To overcome this pitfall, *GFWatch*'s measurement approach of using our own machines located at both sides of the GFW allows us to test hundreds of millions of domains multiple times per day. Using machines under our control also reduces the false positive rate to *zero* since neither of our machines have any DNS resolution capabilities.

10 Conclusion

In this work, we develop *GFWatch*, a large-scale longitudinal measurement platform, to provide a constantly updated view of the GFW's DNS-based blocking behavior and its impact on the global Internet. Over a nine-month period, *GFWatch* has tested 534M domains and discovered 311K censored domains.

We find that the GFW's DNS censorship has a widespread negative impact on the global Internet, especially the domain name ecosystem. *GFWatch* has detected more than 77K censored domains whose poisoned resource records have polluted many popular public DNS resolvers, including Google and Cloudflare. Based on insights gained from the data collected by *GFWatch*, we then propose strategies to effectively detect poisoned responses and evade the GFW's DNS censorship.

As *GFWatch* continues to operate, our data will not only cast new light on technical observations, but also timely inform the public about changes in the GFW's blocking policy and assist other detection and circumvention efforts.

Acknowledgments

We are grateful to Ronald J. Deibert, Adam Senft, Lotus Ruan, Irene Poetranto, Hyungjoon Koo, Shachee Mishra, Tapti Palit, Seyedhamed Ghavamnia, Jarin Firose Moon, Md Mehedi Hasan, Thai Le, Eric Wustrow, Martin A. Brown, Siddharth Varadarajan, Ananth Krishnan, Peter Guest, and others who preferred to remain anonymous for helpful discussions and suggestions.

We would like to thank all the anonymous reviewers for their thorough feedback on this paper. We especially thank the team at `GreatFire.org` for helping to share our findings with related entities in a timely fashion.

This research was supported by the Open Technology Fund under an Information Controls Fellowship. The opinions in this paper are those of the authors and do not necessarily reflect the opinions of the sponsor.

References

[1] Alexa Top Sites. `https://www.alexa.com`.

[2] CISCO OpenDNS Services for Your Home or Small Business. `https://opendns.com/home-internet-security`.

[3] CISCO Umbrella List of Popular Domains. `https://s3-us-west-1.amazonaws.com/umbrella-static/index.html`.

[4] FortiGuard Labs Web Filter. `https://fortiguard.com/webfilter`.

[5] ICANN Centralized Zone Data Service. `https://czds.icann.org`.

[6] IPinfo: The Trusted Source for IP Address Data. `https://ipinfo.io`.

[7] MaxMind GeoLite2 Databases. `https://www.maxmind.com/`.

[8] McAfee: Customer URL Ticketing System. `https://www.trustedsource.org/?p=mcafee`.

[9] OONI: DNS Consistency Specs. `https://ooni.org/nettest/dns-consistency/`.

[10] Quantcast top list. `https://www.quantcast.com/top-sites`.

[11] Rapid7: Open Data. `https://opendata.rapid7.com/`.

[12] Shodan: The search engine for Security. `https://shodan.io/`.

[13] The Citizen Lab Test Lists. `https://github.com/citizenlab/test-lists`.

[14] The Common Crawl Project. `https://commoncrawl.org`.

[15] The Majestic Top One Million Popular Domains. `https://majestic.com/reports/majestic-million`.

[16] Verisign Zone File Service. `https://www.verisign.com/en_US/channel-resources/domain-registry-products`.

[17] Virus Total: URL Scanning Service. `https://www.virustotal.com/gui/home/url`.

[18] China forcing birth control on Uighurs to suppress population, report says. BBC News, 2020-06-29. `https://www.bbc.com/news/world-asia-china-53220713`.

[19] C. Abdelberi, T. Chen, M. Cunche, ED. Cristofaro, A. Friedman, and M. Kâafar. Censorship in the wild: Analyzing internet filtering in syria. *ACM IMC '14*.

[20] Alexa Internet, Inc. How are Alexas's traffic rankings determined? `https://support.alexa.com/hc/en-us/articles/200449744-How-are-Alexa-s-traffic-rankings-determined`.

[21] Anonymous, Arian Akhavan Niaki, Nguyen Phong Hoang, Phillipa Gill, and Amir Houmansadr. Triplet Censors: Demystifying Great Firewall's DNS Censorship Behavior. In *USENIX FOCI' 20*.

[22] Anonymous Author(s). Towards a Comprehensive Picture of the Great Firewall's DNS Censorship. In *USENIX FOCI '14*, 2014.

[23] Simurgh Aryan, Homa Aryan, and J. Alex Halderman. Internet Censorship in Iran: A First Look. In *USENIX FOCI '13*.

[24] Derek E. Bambauer, Ronald J. Deibert, J. Palfrey, Rafal Rohozinski, N. Villeneuve, and J. Zittrain. Internet Filtering in China in 2004-2005: A Country Study. 2005.

[25] Bloomberg. China Tells Carriers to Block Access to Personal VPNs by February, 2017-07-10. `https://www.bloomberg.com/news/articles/2017-07-10/china-is-said-to-order-carriers-to-bar-personal-vpns-by-february`.

[26] S. Bortzmeyer and S. Huque. NXDOMAIN: There Really Is Nothing Underneath. RFC 8020, IETF, November 2016.

[27] Martin A Brown, Doug Madory, Alin Popescu, and Earl Zmijewski. DNS Tampering and Root Servers, 2010.

[28] CAIDA. Routeviews Prefix to AS mappings Dataset (pfx2as) for IPv4 and IPv6. `https://www.caida.org/data/routing/routeviews-prefix2as.xml`.

[29] Cate Cadell. Apple says it is removing VPN services from China App Store. Reuters, 2017-07-29. `https://www.reuters.com/article/us-china-apple-vpn/apple-says-it-is-removing-vpn-services-from-china-app-store-idUSKBN1AE0BQ`.

[30] Zimo Chai, Amirhossein Ghafari, and A. Houmansadr. On the Importance of Encrypted-SNI (ESNI) to Censorship Circumvention. In *USENIX FOCI '19*.

[31] Michael S. Chase and James Mulvenon. You've got dissent!: Chinese dissident use of the internet and beijing's counter-strategies. *Foreign Affairs*, 81:188, 2002.

[32] Taejoong Chung, Roland van Rijswijk-Deij, Balakrishnan Chandrasekaran, David Choffnes, Dave Levin, Bruce M. Maggs, Alan Mislove, and Christo Wilson. A Longitudinal, End-to-End View of the DNSSEC Ecosystem. In *USENIX Security '17*.

[33] R. Clayton, Steven J. Murdoch, and R. Watson. Ignoring the Great Firewall of China. In *PETs '16*.

[34] Craig Hockenberry. Fear China. `https://furbo.org/2015/01/22/fear-china`.

[35] Jedidiah R. Crandall, Daniel Zinn, Michael Byrd, Earl Barr, and Rich East. ConceptDoppler: A Weather Tracker for Internet Censorship. In *ACM CCS '07*.

[36] Tianxiang Dai, Haya Shulman, and Michael Waidner. DNSSEC Misconfigurations in Popular Domains. In *CNS '16*.

[37] DBIP. IP geolocation API and database, 2020. `https://db-ip.com`.

[38] Ronald Deibert. China's Cyberspace Control Strategy: An Overview and Consideration of Issues for Canadian Policy, 2010.

[39] Ronald J. Deibert. Dark Guests and Great Firewalls: The Internet and Chinese Security Policy. *Journal of Social Issues*, 58:143–159, 2002.

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

[41] Arun Dunna, Ciarán O'Brien, and Phillipa Gill. Analyzing China's Blocking of Unpublished Tor Bridges. In *USENIX FOCI '18*.

[42] Z. Durumeric, E. Wustrow, and J. A. Halderman. ZMap: Fast Internet-wide Scanning and Its Security Applications. In *USENIX Security '13*.

[43] D. Eastlake and C. Kaufman. Domain Name System Security Extensions. RFC 2065, IETF, January 1997.

[44] R. Ensafi, D. Fifield, P. Winter, N. Feamster, N. Weaver, and V. Paxson. Examining How the Great Firewall Discovers Hidden Circumvention Servers. In *ACM IMC '15*.

[45] Roya Ensafi, Philipp Winter, Abdullah Mueen, and Jedidiah R Crandall. Analyzing the Great Firewall of China over space and time. *PETs '15*.

[46] O. Farnan, A. Darer, and J. Wright. Poisoning the Well: Exploring the Great Firewall's Poisoned DNS Responses. In *WPES '16*.

[47] Arturo Filasto and Jacob Appelbaum. OONI: Open Observatory of Network Interference. In *USENIX FOCI '12*.

[48] Jon Fraenkel and Bernard Grofman. The Borda Count and its Real-world Alternatives: Comparing Scoring Rules in Nauru and Slovenia. *Australian Journal of Political Science*, 2014.

[49] Freedom House. Freedom on the Net 2018: The Rise of Digital Authoritarianism, 2018. `https://freedomhouse.org/report/freedom-net/2018/rise-digital-authoritarianism`.

[50] V. Fuller and T. Li. Classless Inter-domain Routing (CIDR): The Internet Address Assignment and Aggregation Plan. RFC 4632, IETF, August 2006.

[51] Genevieve Gebhart and Tadayoshi Kohno. Internet Censorship in Thailand: User Practices and Potential Threats. *EuroSP '17*.

[52] Geremie R. Barme And Sang Ye. The Great Firewall of China, 1997-06-01. `https://www.wired.com/1997/06/china-3/`.

[53] Phillipa Gill, Masashi Crete-Nishihata, Jakub Dalek, Sharon Goldberg, Adam Senft, and Greg Wiseman. Characterizing Web Censorship Worldwide: Another Look at the Opennet Initiative Data. *TWEB '15*.

[54] GreatFire Project. GFW Upgrade Fail - Visitors To Blocked Sites Redirected To Porn. `https://en.greatfire.org/blog/2015/jan/gfw-upgrade-fail-visitors-blocked-sites-redirected-porn`.

[55] GreatFire Project. We Monitor and Challenge Internet Censorship in China. `https://greatfire.org`.

[56] Shuai Hao, Yubao Zhang, Haining Wang, and Angelos Stavrou. End-Users Get Maneuvered: Empirical Analysis of Redirection Hijacking in Content Delivery Networks. In *USENIX Security '18*.

[57] Nguyen Phong Hoang, Sadie Doreen, and Michalis Polychronakis. Measuring I2P Censorship at a Global Scale. In *USENIX FOCI '19*.

[58] Nguyen Phong Hoang, Panagiotis Kintis, Manos Antonakakis, and M. Polychronakis. An Empirical Study of the I2P Anonymity Network and its Censorship Resistance. In *ACM IMC '18*.

[59] Nguyen Phong Hoang, Arian Akhavan Niaki, Nikita Borisov, Phillipa Gill, and Michalis Polychronakis. Assessing the Privacy Benefits of Domain Name Encryption. In *ACM AsiaCCS '20*.

[60] Nguyen Phong Hoang, Arian Akhavan Niaki, Phillipa Gill, and Michalis Polychronakis. Domain Name Encryption Is Not Enough: Privacy Leakage via IP-based Website Fingerprinting. In *PoPETs '21*.

[61] NP. Hoang, AA. Niaki, M. Polychronakis, and P. Gill. The web is still small after more than a decade. *ACM SIGCOMM CCR '20*.

[62] Christian Huitema. Teredo: Tunneling IPv6 over UDP through Network Address Translations (NATs). RFC 4380, IETF, February 2006.

[63] IP2Location. Identify Geographical Location by IP Address, 2020. `https://www.ip2location.com`.

[64] J. Ullrich. Are You Piratebay? thepiratebay.org Resolving to Various Hosts. `https://isc.sans.edu/forums/diary/Are+You+Piratebay+thepiratebayorg+Resolving+to+Various+Hosts/19175`.

[65] Ben Jones, Tzu-Wen Lee, N. Feamster, and Phillipa Gill. Automated Detection and Fingerprinting of Censorship Block Pages. In *IMC '14*.

[66] Victor Le Pochat, Tom Van Goethem, Samaneh Tajalizadehkhoob, Maciej Korczyński, and Wouter Joosen. Tranco: A Research-Oriented Top Sites Ranking Hardened Against Manipulation. In *NDSS '19*.

[67] Graham Lowe, Patrick Winters, and Michael L. Marcus. The Great DNS Wall of China. Technical report, New York University, 2007.

[68] B. Marczak, N. Weaver, J. Dalek, Roya Ensafi, D. Fifield, Sarah McKune, Arn Rey, J. Scott-Railton, Ronald J. Deibert, and V. Paxson. An Analysis of China's Great Cannon. In *USENIX FOCI '15*.

[69] A. McDonald, M. Bernhard, Luke Valenta, Benjamin VanderSloot, W. Scott, N. Sullivan, J. A. Halderman, and Roya Ensafi. 403 Forbidden: A Global View of CDN Geoblocking. In *ACM IMC '18*.

[70] Z. Nabi. The Anatomy of Web Censorship in Pakistan. In *FOCI '13*.

[71] AA. Niaki, S. Cho, Z. Weinberg, NP. Hoang, A. Razaghpanah, N. Christin, and P. Gill. ICLab: A Global, Longitudinal Internet Censorship Measurement Platform. In *IEEE S&P '20*.

[72] D Nobori and Y Shinjo. VPN Gate: A Volunteer-Organized Public VPN Relay System with Blocking Resistance for Bypassing Government Censorship Firewalls. *USENIX NSDI '14*.

[73] J. Park and J. Crandall. Empirical Study of a National-Scale Distributed Intrusion Detection System: Backbone-Level Filtering of HTML Responses in China. *ICDCS '10*.

[74] P. Pearce, Ben Jones, F. Li, Roya Ensafi, N. Feamster, N. Weaver, and V. Paxson. Global Measurement of DNS Manipulation. In *USENIX Security Symposium*, 2017.

[75] Peter Guess. China suddenly blocked an Indonesian newspaper. No one knows why. `https://restofworld.org/2021/china-suddenly-blocked-an-indonesian-newspaper-no-one-knows-why/`.

[76] Victor Le Pochat, Tom Van Goethem, and Wouter Joosen. Evaluating the Long-term Effects of Parameters on the Characteristics of the Tranco Top Sites Ranking. In *USENIX CSET '19*.

[77] R. Liao. China bans Scratch, MIT's programming language for kids, 2020. `https://techcrunch.com/2020/09/07/scratch-ban-in-china`.

[78] RS. Raman, P. Shenoy, K. Kohls, and R. Ensafi. Censored Planet: An Internet-wide, Longitudinal Censorship Observatory. In *CCS '20*.

Table 4: Top base censored domains that cause most overblocking of innocuous domains.

# domains impacted	Base censored domains	Sample innocuous domains
11,227	919.com	455919.com, rem99919.com niwa919.com, xaa919.com
2,346	jetos.com	ccmprojetos.com, csprojetos.com itemsobjetos.com, dobobjetos.com
1,837	33a.com	87833a.com, 280333a.com xn--72caa7c0a9clrce0a1fp33a.com xn--zck4aye2c2741a5qvo33a.com
1,574	9444.com	mkt9444.com, 15669444.com 3329444.com, 5719444.com
1,547	sscenter.net	dentalwellnesscenter.net, swisscenter.net chesscenter.net, childlosscenter.net
1,487	1900.com	faber1900.com, salah1900.com phoenixspirit1900.com, interiors1900.com
1,392	98a.com	p98a.com, 72898a.com, 1098a.com xn--1-ieup4b2ab8q5c0dxj6398a.com
1,144	ss.center	hss.center, icass.center limitless.center, ass.center
1,089	reddit.com	bestiptvreddit.com, booksreddit.com cachedreddit.com, geareddit.com
789	visi.tk	erervisi.tk, yetkiliservisi.tk buderuservisi.tk, bodrumklimaservisi.tk

[79] RS. Raman, A. Stoll, J. Dalek, R. Ramesh, W. Scott, and R. Ensafi. Measuring the Deployment of Network Censorship Filters at Global Scale. In *NDSS '20*.

[80] R. Rambert, Z. Weinberg, D. Barradas, and N. Christin. Chinese Wall or Swiss Cheese? Keyword filtering in the Great Firewall of China. In *ACM WWW '21*.

[81] Raymond Zhong, Paul Mozur, Jeff Kao, and Aaron Krolik. No 'Negative' News: How China Censored the Coronavirus. The New York Times, 2020-12-19. https://www.nytimes.com/2020/12/19/technology/china-coronavirus-censorship.html.

[82] Philipp Richter, R. Padmanabhan, N. Spring, A. Berger, and D. Clark. Advancing the Art of Internet Edge Outage Detection. *ACM IMC '18*.

[83] W. Rweyemamu, T. Lauinger, C. Wilson, W. Robertson, and E. Kirda. Clustering and the Weekend Effect: Recommendations for the Use of Top Domain Lists in Security Research. In *PAM '19*.

[84] W. Scott, T. Anderson, T. Kohno, and A. Krishnamurthy. Satellite: Joint Analysis of CDNs and Network-Level Interference. In *ATC '16*.

[85] Shawn Conaway. The Great Firewall: How China Polices Internet Traffic. Certification Magazine, 2009-09-30. http://certmag.com/the-great-firewall-how-china-polices-internet-traffic/.

[86] Soutik Biswas. India-China Clash: 20 Indian Troops Killed in Ladakh Fighting. BBC, 2020-06-16. https://www.bbc.com/news/world-asia-53061476.

[87] Sparks and Neo and Tank and Smith and Dozer. The Collateral Damage of Internet Censorship by DNS Injection. *SIGCOMM CCR '12*.

[88] Tokachu. The Not-So-Great Firewall of China. *The Hacker Quarterly*, 23:58–60, 2006.

[89] Pelayo Vallina, Victor Le Pochat, Álvaro Feal, M. Paraschiv, Julien Gamba, T. Burke, O. Hohlfeld, Juan Tapiador, and N. Vallina-Rodriguez. Mis-shapes, Mistakes, Misfits: An Analysis of Domain Classification Services. *ACM IMC '20*.

[90] Benjamin VanderSloot, A. McDonald, W. Scott, J. A. Halderman, and Roya Ensafi. Quack: Scalable Remote Measurement of Application-Layer Censorship. In *USENIX Security '18*.

[91] Z. Wang, Y. Cao, Z. Qian, C. Song, and S. Krishnamurthy. Your state is not mine: a closer look at evading stateful internet censorship. In *ACM IMC '17*.

[92] P Winter and S Lindskog. How the Great Firewall of China is Blocking Tor. USENIX FOCI '12.

[93] Xueyang Xu, Z. Morley Mao, and J. Alex Halderman. Internet Censorship in China: Where Does the Filtering Occur? In *PAM '11*.

[94] Young Xu. Deconstructing the Great Firewall of China. Technical report, Thousand Eyes, 2016.

[95] Jonathan Zittrain and Benjamin Edelman. Internet Filtering in China. *IEEE Internet Computing '03*.

A Most Extreme Blocking Rules

Table 4 shows the top ten base censored domains blocked under Rule 4 that we have discussed in §4.1. The blocking rule applied on these ten domains results in overblocking of more than 24K innocuous domains, which is more than half of all innocuous domains. The third column shows some samples of innocuous censored domains that *GFWatch* has discovered. The impacted innocuous domains presented in this table are all active and hosting some contents at the time of writing this paper. Except those that do not allow Web Archive's crawler, we have also saved a snapshot of these domains at https://web.archive.org for future reference in case these domains become inactive. In contrast, most base censored domains shown in the second column are not currently hosting any content. Therefore, one may wonder why many seemingly inconsequential domains are being censored.

To make sure that these seemingly inconsequential censored domains were not blocked because the GFW was using an imprecise classifier (e.g., a Bloom filter) for fast classification, we tested 200M randomly generated nonexistent domains and found that none were censored. It is worth noting that many censored domains discovered by *GFWatch* have been blocked before the launch of our platform. Prior to our testing, they might have served "unwanted" content that we were not aware of. Moreover, the GFW is known to conduct blanket blocking against websites that run editorials on "unwanted" topics without carefully verifying their contents. Once domains are censored, they are often kept in the GFW's blocklist for a long time regardless of their activity [75].

As can be seen from the table, the GFW's overblocking design affects not only usual ASCII-based innocuous domains, but also Internationalized Domain Names (IDNs), i.e., those starting with "xn--". Of 41K innocuously blocked domains, we find a total of 1.2K IDNs are overblocked. Our finding shows that the current DNS-based blocking policy of the GFW has a widespread negative impact on the domain name ecosystem.

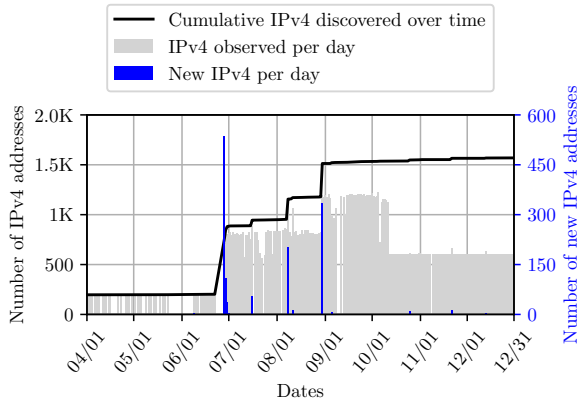

Figure 12: Number of forged IPv4 addresses detected over time by probing different network prefixes in China.

B Consistency of Forged IP Addresses Across Different Network Locations

To confirm whether the pool of forged IPs discovered by *GFWatch* (§5) is representative enough, we probe different network locations in China to compare the forged IPs observed from these locations and the ones seen by *GFWatch*. For this experiment, we obtain the daily updated *pfx2as* dataset provided by CAIDA [28], and extract prefixes located in China by checking them against the MaxMind dataset [7], which we also update biweekly. Unlike the measurement conducted between our own controlled machines located at two sides of the GFW, this task requires us to send DNS queries, encapsulating censored domains, to destinations we do not own. Although similar large-scale network probing activities are widely conducted nowadays by both academia [42, 74, 78, 90] and industry [11, 12], our measurement must be designed in a careful and responsible manner.

Our sole purpose of this measurement is to deliver probing queries passing through the GFW's infrastructure at different network locations to trigger censorship, instead of having the probing packets completely delivered to any alive hosts. Therefore, we craft our probing packets using the routing address of a given prefix as the destination IP. According to the best current practice [50], except for the case of a `/32` subnet with only one IP, the routing address of a subnet should not be assigned to any device because it is solely used for routing purposes. For example, given the prefix `1.92.0.0/20` announced in the *pfx2as* dataset, we craft our probing packet with the destination as `1.92.0.0`. With this probing strategy, we can reduce the risk that our packets will hit an alive host while still being able to deliver them across the GFW's infrastructure at different network locations. To reduce the risk even further, we opt to only probe prefixes whose subnet is less-specific than `/24`.

In spite of the standardized practices in assigning IP and the extra care that we have taken in designing our measurement, we also follow a common practice that is widely used in research activities that involve network scanning, i.e., allowing opt-out. More specifically, we accompany our probing DNS queries with a non-censored domain under our control, from which the information about our study and a contact email address can be found to request opt-out from our measurement. Since the launch of *GFWatch*, we have not received any complaints or opt-out requests.

Figure 12 show the cumulative number of forged IPs discovered daily and over the whole period of our measurement. Similar to Figure 7, the number of forged IPs addresses observed initially in April is also about 200. However, we did not see any gradual increase in the number of forged IPs from May as seen in Figure 7. After waiting about two months without seeing any new IPs observed from probing different prefixes, we have learned that this is due to the fact that we only use *one* known censored domain for probing the prefixes. This is because of an earlier precaution that these probed destinations are not owned by us, thus we should try to limit the amount of probing traffic as much as possible. However, it turned out that we need to probe more than just one domain to be able to obtain a similar set of forged IP addresses detected earlier by *GFWatch*.

We then decide to add more domains to this test, probing a total of 22 censored domains per prefix. These domains are selected from several categories, including advocacy organizations, proxy avoidance, news and media, social network, personal websites and blogs, shopping, instant messaging, etc. As expected, the cumulative number of forged IPs immediately increases to almost 1K the day we revise our test domains. Similar to Figure 7, the cumulative number of forger IPs also increase gradually towards the end of August. With a major increase of more than 300 forged IPs, the number of all forged IPs observed from our prefixes probing measurement also converges to above 1.5K by the end of December.

While the number of forged IPs obtained from probing the prefixes on some days, especially from July to September, is higher than what *GFWatch* observed during this period, we find that 96% of the forged IPs observed from prefixes probing have already detected by *GFWatch*. Conducting the same injection frequency analysis on these forged IPs gives us the same results as found in §5.2. In other words, the most frequently injected IPs discovered by *GFWatch* and from probing different prefixes are the same. To this end, we could confirm that the coverage of forged IPs discovered by *GFWatch* is representative and sufficient for us to develop effective detection (§6.2) and circumvention strategies (§7).

C Multiple Injectors

It was first reported by [21] that the GFW comprises multiple injectors that are responsible for DNS poisoning. Depending on the domain being queried (e.g., `google.sm`), multiple forged responses can be triggered simultaneously to increase

Table 5: A high-level comparison of censored domains and forged IPs detected by different studies/platforms. (*) The number of forged IPs from Satellite and OONI includes "anomalies" due to domains hosted on CDNs and localized filtering policies.

Study/Platform	Duration	Longitudinal	Tested Domains	Censored Domains	Forged IPs	Common Forged IPs
Zittrain et al. [95]	Mar 2002 - Nov 2002	○	204K	1K	1	1
Lowe et al. [67]	2007	○	951	393	21	3
Brown et al. [27]	Nov 2010	○	1	1	9	6
CCR'12 [87]	Nov 2011	○	10	6	28	
FOCI'14 [22]	Aug 2013 - Apr 2014	○	130M	35.3K	174	
Triplet Censors [21]	Sep 2019 - May 2020	○	1M	24.6K	1,510	1,462
OONI [47]	Apr 2020 - Dec 2020	●	3.3K	460	*710	593
Satellite [84]	Apr 2020 - Dec 2020	●	3.5K	375	*2,391	1,613
`GFWatch`	Apr 2020 - Dec 1020	●	534M	311K	1,781	-

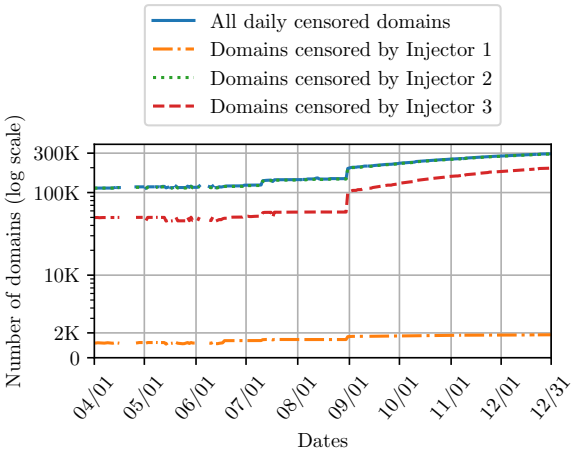

Figure 13: Number of censored domains per injector.

the chance of successfully poisoning censored clients if one of the injectors is overloaded, and make detection and circumvention non-trivial. From the data collected by `GFWatch`, we have confirmed the same injection behavior. More specifically, there are three injectors, which can be differentiated by the *"DNS Authoritative Answer"* flag in the DNS header and the *"do not fragment"* flag in the IP header. Injector 1 has the *"DNS Authoritative Answer"* bit set to **1**, Injector 2 has the *"DNS Authoritative Answer"* bit set to **0** and *"do not fragment"* bit set to **1**, whereas Injector 3 has the *"DNS Authoritative Answer"* bit set to **0** and *"do not fragment"* bit set to **0**.

Based on these fingerprints, we then cluster 311K censored domains into three groups with respect to the three injectors. Figure 13 depicts the number of censored domains observed over time for each injector. Injector 2 is responsible for 99% of the censored domains, whereas Injectors 3 and 1 are responsible for only 64% and less than 1% (2K) of censored domains, respectively. Note that all domains censored by Injector 3 are also censored by Injector 2, while there are 1.7K domains censored only by Injector 1, but not other injectors.

D Politically Motivated Censorship

Internet censorship and large-scale network outages are often politically motivated [53, 82]. From the censored domains discovered by `GFWatch`, we find numerous governmental websites censored by the GFW, including many sites belonging to the US government, such as `share.america.gov`, `cecc.gov`, and `uscirf.gov`.

During the nine-month measurement period, `GFWatch` has also spotted several blockages that coincide with political events. For instance, soon after the clash between China and India due to the border dispute in Ladakh [86], on June 18th 2020 `GFWatch` detected the DNS filtering of several Indian news sites (e.g., `thewire.in`, `newsr.in`). We reached out to the editor of the Wire India to report blockage against their website by the GFW and were told that they were unaware of the blockage since the site was still accessible from China earlier. Another instance is the blockage of `scratch.mit.edu` that took place in August, 2020, due to some content deemed as anti-China hosted on this website, affecting about three million Chinese users [77]. Although this event was reported by the GreatFire project [55] on the 20th and by Chinese users on the 14th [77], `GFWatch` actually detected the first DNS poisoning instance earlier on August 13th.

These cases highlight the importance of `GFWatch`'s ability to operate in an automated and continuous fashion to obtain a constantly updated view of the GFW to timely inform the public about changes in its blocking policy.

E Detailed Comparison with Related Work

Table 5 provides a detailed comparison, highlighting the main differences between `GFWatch` and prior studies. Note that the numbers of IPs in this table indicate IPv4 addresses. We do not include a comparison of the number of IPv6 addresses because most previous works did not consider IPv6 in their experiments.