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
The State University of New York
UMASS AMHERST
THECITIZENLAB
UNIVERSITY OF TORONTO
10100101
10100101
10100101
University of California
Internet censorship is widespread

Quick Take

The Great Firewall of China

Bloomberg News
Updated on November 5, 2018, 6:36 PM PST

China’s online population of 800 million gets a highly restricted internet, one that doesn’t include access to Google, Facebook, YouTube or the New York Times. There’s little coverage of the 1989 student protests in Tiananmen Square. Even Winnie the Pooh got temporarily banned. China is able to control such a vast ocean of content through the largest system of censorship in the world, aptly known as the Great Firewall. It’s a joint effort between government monitors and the technology and telecommunications companies compelled to enforce the state’s rules. The stakes go beyond China, which is setting an example that other authoritarian countries can imitate.

Russia to block Telegram app over encryption

BBC

Russia to block Telegram app over encryption

13 April 2018

A court in Moscow has approved a request from the Russian media regulator to block the Telegram messaging app immediately.

The media regulator sought to block the app because the firm had refused to hand over encryption keys used to scramble messages.

Security officials say they need to monitor potential terrorists.

But the company said the way the service was built meant it had no access to customers’ encryption keys.
The Great Firewall (GFW) of China

The Great Firewall of China

At ISPs, Internet cafes, even state censorship committees, we meet the wired of China – and discover that the technology China needs to build the most powerful country on Earth in the 21st Century threatens to undermine the institutions that rule the nation. And Beijing's control freaks are worried. "Information industries of China unite!" Xia [...]
Different blocking techniques

- DNS tampering: injecting fake DNS responses
- IP blocking: blackholing traffic destined to censored IPs
- Keyword filtering: applied on unencrypted network traffic
- Active probing: discover censorship-circumvention proxies
Initial observations of DNS censorship

• Multiple fake DNS responses are injected
• Forged IP addresses are:
  ✓ publicly routable IPs
  ✓ dynamically changed
  ✓ managed by foreign entities (e.g., Facebook, Twitter, IBM)

Fake responses

Actual response
Measuring China’s DNS censorship

1. How many censored domains are there?

2. What are the forged IP addresses used in fake DNS responses?

3. What is the impact of the Great Firewall’s DNS censorship on the global Internet?

4. How to effectively detect and circumvent the Great Firewall’s DNS censorship?
Measurement system requirements

1. Discover as many censored domains and forged IPs as possible in a timely manner
2. Obtain and test new domain names as they appear on the Internet
3. Continuously keep track of censored domains’ blocking status, once discovered, to determine whether the domains stays censored or becomes unblocked at some point in the future
4. Provide a good view into the pool of forged IPs used by the Great Firewall
GFWatch design: probing from outside

- Main prober in US
- DNS zone files
- CLTL Tranco C.Crawl
- DNS queries
- Controlled CN machines
- (*) Censored domains
- Forged responses
GFWatch design: probing from inside

- Main prober in US
- Controlled CN machines
- Censored domains
- Verified censored domains
- DNS queries
- Forged responses

(*)
Due to the addition of
*.googlevideo.com
*.appspot.com domains

More than 311K censored domains discovered
Necessitate a better counting process
Base censored domains probing

• Goal: find the shortest domain that triggers blocking

• Method: test 8 permutations of the domain and random strings

→ More precise counting of censored domains
Base censored domains probing

Before

<table>
<thead>
<tr>
<th>Rule 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>censored_domain</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rule 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>censored_domain</td>
</tr>
</tbody>
</table>

After

<table>
<thead>
<tr>
<th>Rule 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>rnd_str</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rule 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>rnd_str</td>
</tr>
</tbody>
</table>

Middle

<table>
<thead>
<tr>
<th>Rule 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>rnd_str</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rule 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>rnd_str</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rule 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>rnd_str</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rule 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>rnd_str</td>
</tr>
</tbody>
</table>
Base censored domains

From 311K censored domains
→ 138.7K base censored domains
Over-blocked domains

The Tor Project’s domain TorProject.org is blocked under the rule:

*torproject.org

→ Any domains ending with torproject.org are censored
Over-blocked domains

- 41K innocuous domains being over-blocked, including IDNs
- Mostly over-blocked under the rule: *censored_domain

<table>
<thead>
<tr>
<th># domains impacted</th>
<th>Based domains</th>
<th>Sample innocuous domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>11,227</td>
<td>919.com</td>
<td>455919.com, rem99919.com, niwa919.com, xaa919.com</td>
</tr>
<tr>
<td>2,346</td>
<td>jetos.com</td>
<td>ccmprojetos.com, csprojetos.com, itemsobjetos.com, dobobjetos.com</td>
</tr>
<tr>
<td>1,837</td>
<td>33a.com</td>
<td>87833a.com, 280333a.com, xn--72caa7c0a9clrce0a1fp33a.com, xn--zck4aye2c2741a5qvo33a.com</td>
</tr>
<tr>
<td>1,574</td>
<td>9444.com</td>
<td>mkt9444.com, 15669444.com, 3329444.com, 5719444.com</td>
</tr>
<tr>
<td>1,547</td>
<td>sscenter.net</td>
<td>dentalwellnesscenter.net, swisscenter.net, chesscenter.net, childlosscenter.net</td>
</tr>
</tbody>
</table>
GFW is also used for geo-blocking

A notable case of GFW’s geo-blocking: www.[.]beian[.]gov[.]cn, which has all its authoritative nameservers in China.

61.8K other censored domains may be geoblocked because at least one of their name servers is located inside China.
Public DNS resolvers’ cache pollution

Top ten public resolvers with the highest number of censored domains whose poisoned records have polluted their cache

<table>
<thead>
<tr>
<th># Domains</th>
<th>Resolver</th>
<th># Domains</th>
<th>Resolver</th>
</tr>
</thead>
<tbody>
<tr>
<td>74,715</td>
<td>Google</td>
<td>63,295</td>
<td>OpenDNS</td>
</tr>
<tr>
<td>71,560</td>
<td>Cloudflare</td>
<td>62,825</td>
<td>Comcast</td>
</tr>
<tr>
<td>65,567</td>
<td>OpenNIC</td>
<td>56,913</td>
<td>CleanBrowsing</td>
</tr>
<tr>
<td>65,538</td>
<td>FreeDNS</td>
<td>56,628</td>
<td>Level3</td>
</tr>
<tr>
<td>64,521</td>
<td>Yandex</td>
<td>55,795</td>
<td>Verisign</td>
</tr>
</tbody>
</table>

→ Necessitate a sanitization mechanism to assure the quality of DNS resolutions and other DNS-related services on the Internet
Forged IP addresses

- 1,781 forged IPv4s, most of which belong to U.S. organizations
- 1,799 forged IPv6s, all of which are bogus and belong to the same Teredo subnet 2001::/32
Forged IP addresses

- Not all forged IPs are equally injected in censored responses
- Knowing forged IPs → effective detection and circumvention 99% of the time
Key contributions of *GFWatch*

A large-scale longitudinal measurement platform that has

- Exposed China’s censorship policy in a timely manner
- Discovered unknown DNS blocking behaviors of GFW:
  - over blocking
  - geo-blocking
  - injection pattern of forged IPs
- Collected censored domains and forged IPs, that can assist in:
  - sanitization of poisoned records from public DNS resolvers’ cache
  - development of effective strategies for detecting and circumventing China’s DNS censorship
This research was supported by the Open Technology Fund under an Information Controls Fellowship. The opinions in the paper and this presentation are those of the authors and do not necessarily reflect the opinions of the sponsor.

Updates of newly censored domains 🐦 @NP_tokumei

Collected dataset and a dashboard are available at:

https://homepage.np-tokumei.net/publication/publication_2021_usenix_security