The 33rd USENIX Security Symposium

August 14–16, 2024

GFWeb: Measuring the Great Firewall's Web Censorship at Scale

Nguyen Phong Hoang, Jakub Dalek, Masashi Crete-Nishihata
Nicolas Christin, Vinod Yegneswaran, Michalis Polychronakis, Nick Feamster

UBC THE UNIVERSITY OF BRITISH COLUMBIA

THECITIZENLAB

UNIVERSITY OF TORONTO

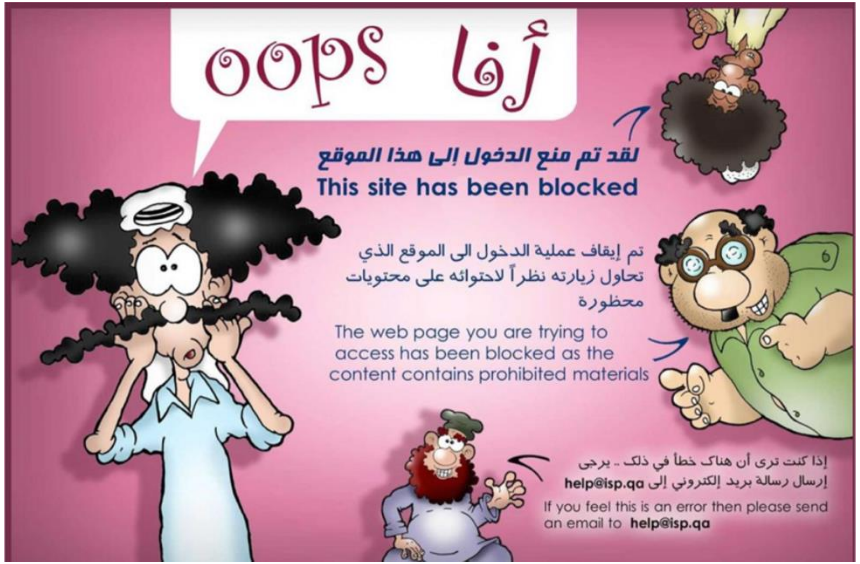
Carnegie Mellon University

SRI International

Stony Brook University

THE UNIVERSITY OF CHICAGO

What is Web censorship?



oops أفا

لقد تم منع الدخول إلى هذا الموقع
This site has been blocked

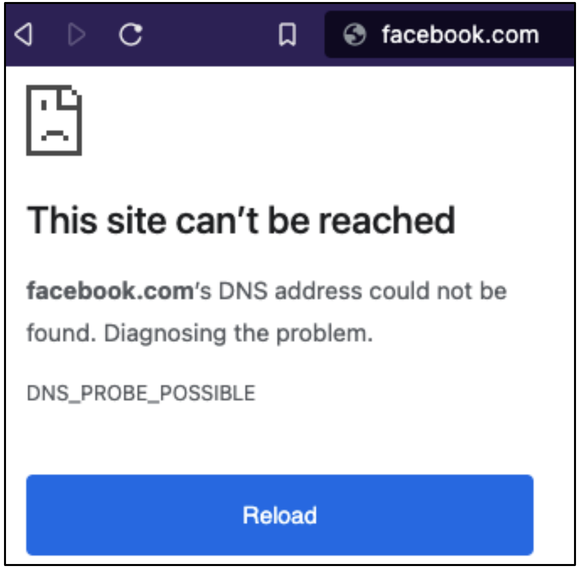
تم إيقاف عملية الدخول الى الموقع الذي
تحاول زيارته نظراً لاحتوائه على محتويات
محظورة

The web page you are trying to
access has been blocked as the
content contains prohibited materials

إذا كنت ترى أن هناك خطأ في ذلك .. يرجى
إرسال رسالة بريد إلكتروني إلى help@isp.qa

If you feel this is an error then please send
an email to **help@isp.qa**

Overt censorship



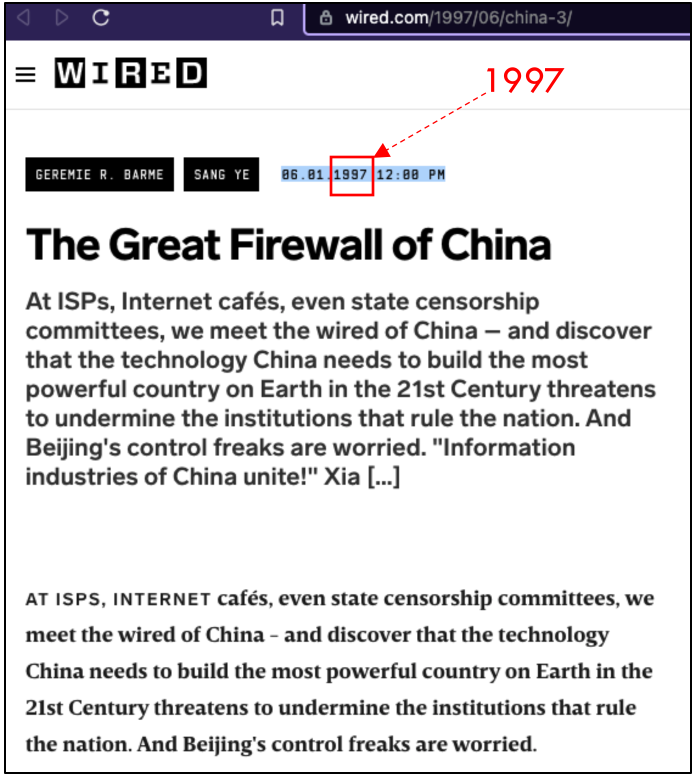
This site can't be reached

facebook.com's DNS address could not be found. Diagnosing the problem.

DNS_PROBE_POSSIBLE

Covert censorship

The Great Firewall is one of the most sophisticated



wired.com/1997/06/china-3/

WIRED

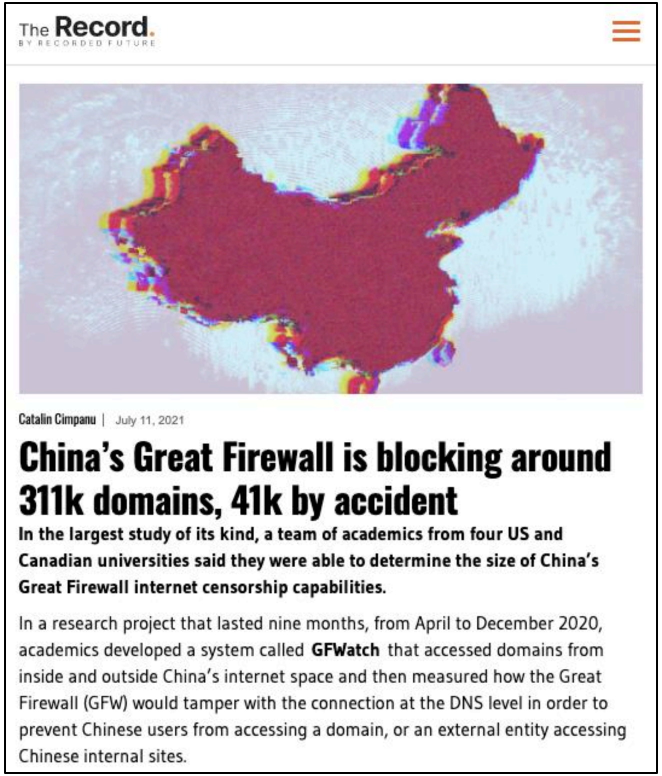
1997

GEREMIE R. BARME SANG YE 06.01.1997 12:00 PM

The Great Firewall of China

At ISPs, Internet cafés, 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 [...]

AT ISPS, INTERNET cafés, 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.



The Record.
BY RECORDED FUTURE

Catalin Cimpanu | July 11, 2021

China's Great Firewall is blocking around 311k domains, 41k by accident

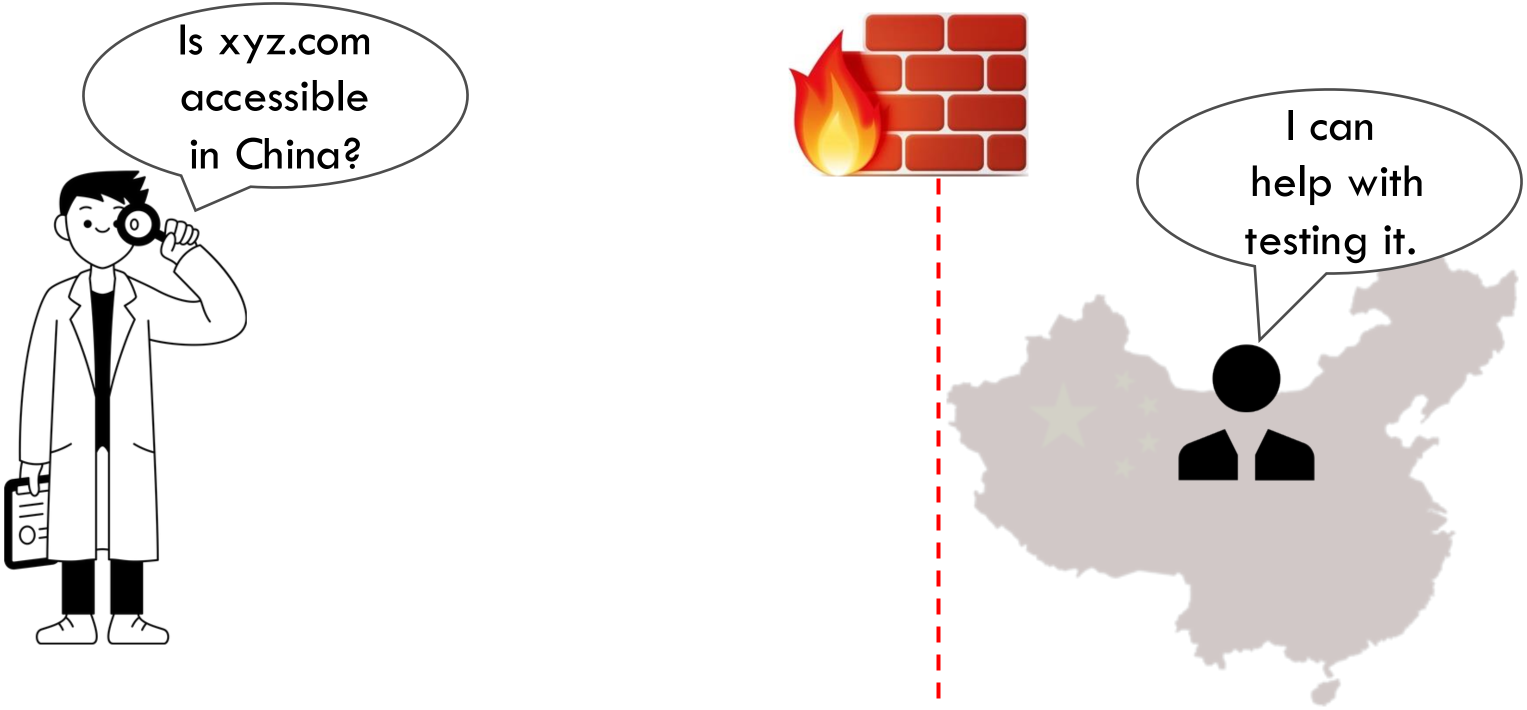
In the largest study of its kind, a team of academics from four US and Canadian universities said they were able to determine the size of China's Great Firewall internet censorship capabilities.

In a research project that lasted nine months, from April to December 2020, academics developed a system called **GFWatch** that accessed domains from inside and outside China's internet space and then measured how the Great Firewall (GFW) would tamper with the connection at the DNS level in order to prevent Chinese users from accessing a domain, or an external entity accessing Chinese internal sites.

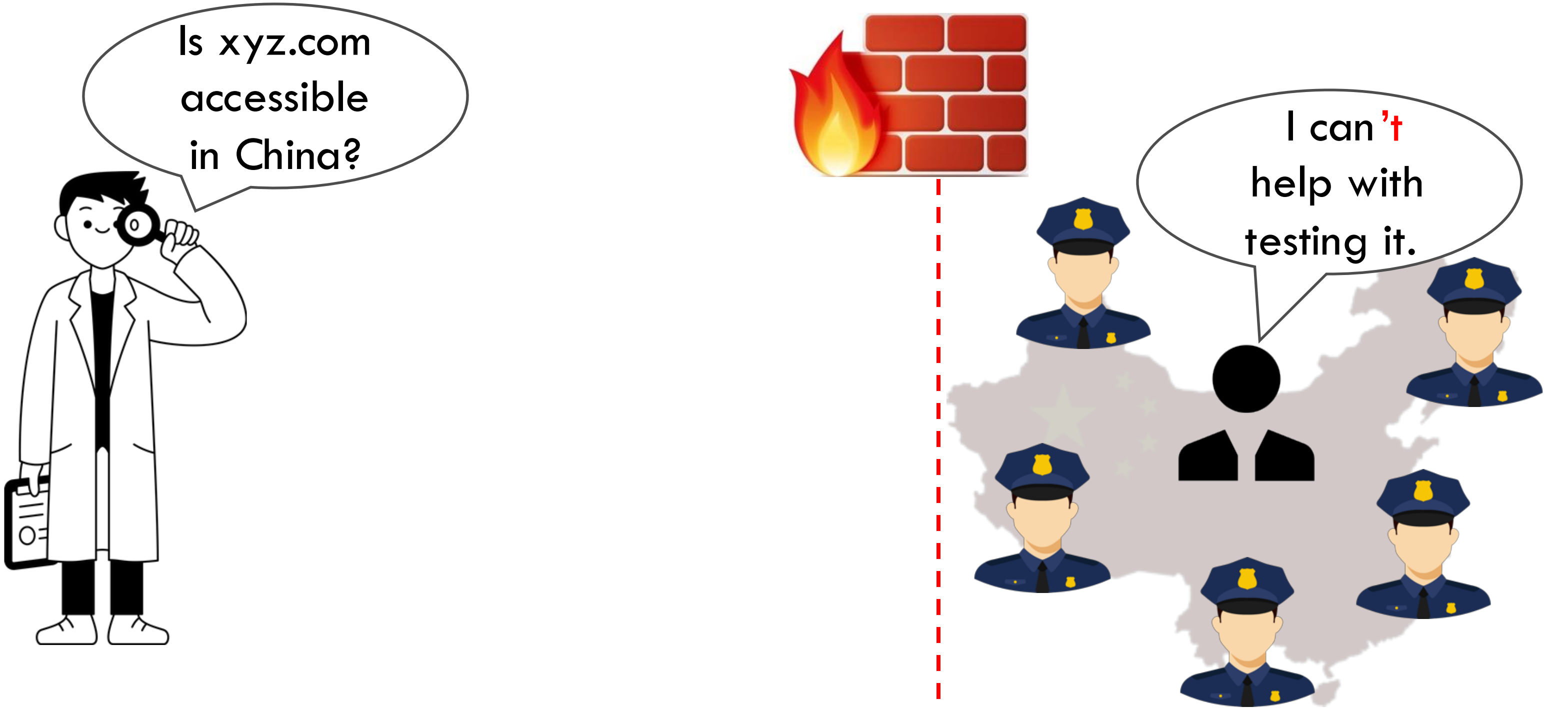
Different filtering techniques used by the GFW

- DNS tampering: inject fake DNS responses
- Filtering of unencrypted network traffic (HTTP)
- SNI-based blocking: inspect HTTPS (TLS) traffic
- Active probing: discover censorship-circumvention proxies
- IP blocking: blackhole (null-route) traffic destined to censored IPs

How to measure censorship?



How to measure censorship?

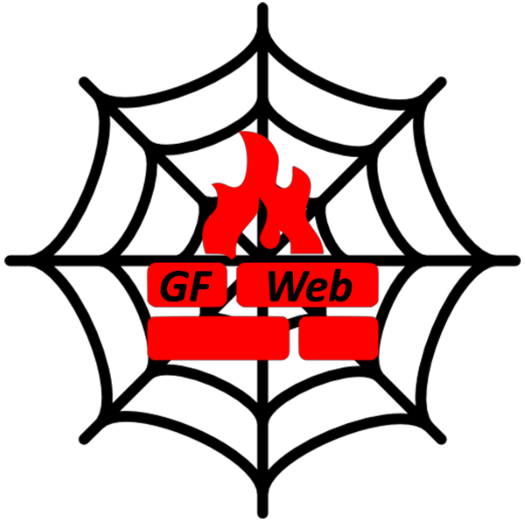

Is xyz.com accessible in China?
I can’t help with testing it.

→ Traditional methods pose risks to volunteers and do not scale well.

How to measure the GFW’s Web censorship at scale?

Measure
DNS censorship

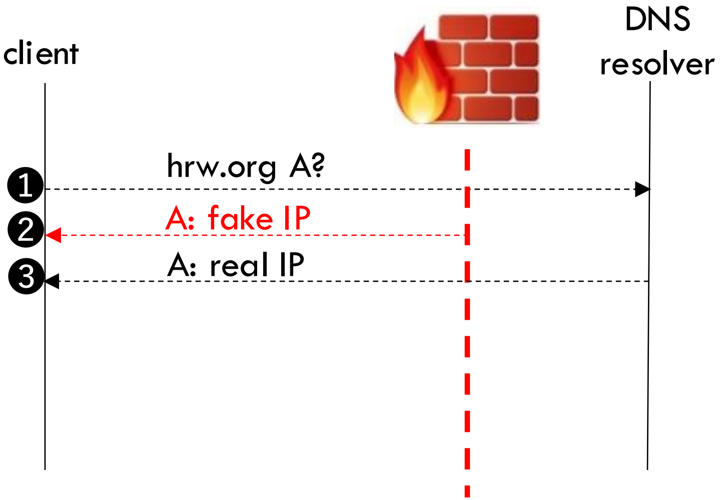
(*USENIX Security’21*)

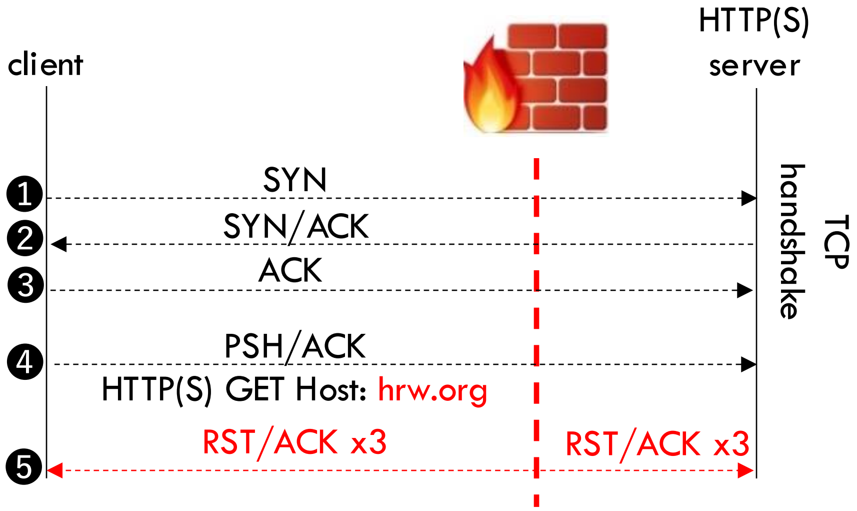


Measure
HTTP(S) censorship

(*this presentation*)

What do we know about the GFW's Web blocking?



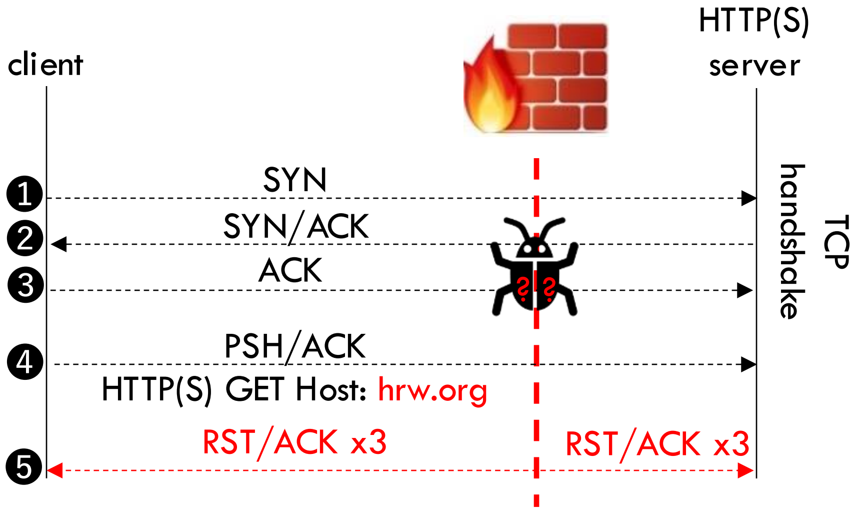


The GFW is bidirectional: both egress and ingress network packets sent from/to inside the country can trigger its filtering middleboxes.

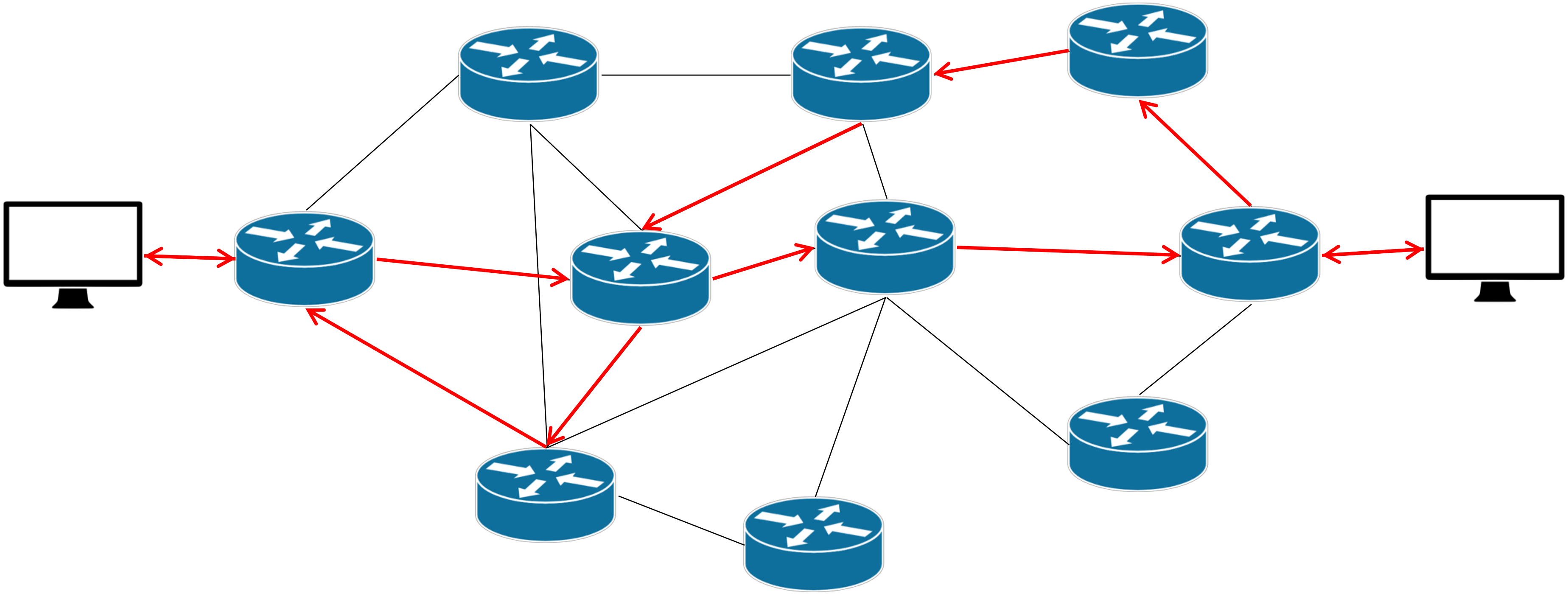
Initial observations of GFW’s HTTP(S) filtering

- Multiple tear-down injections

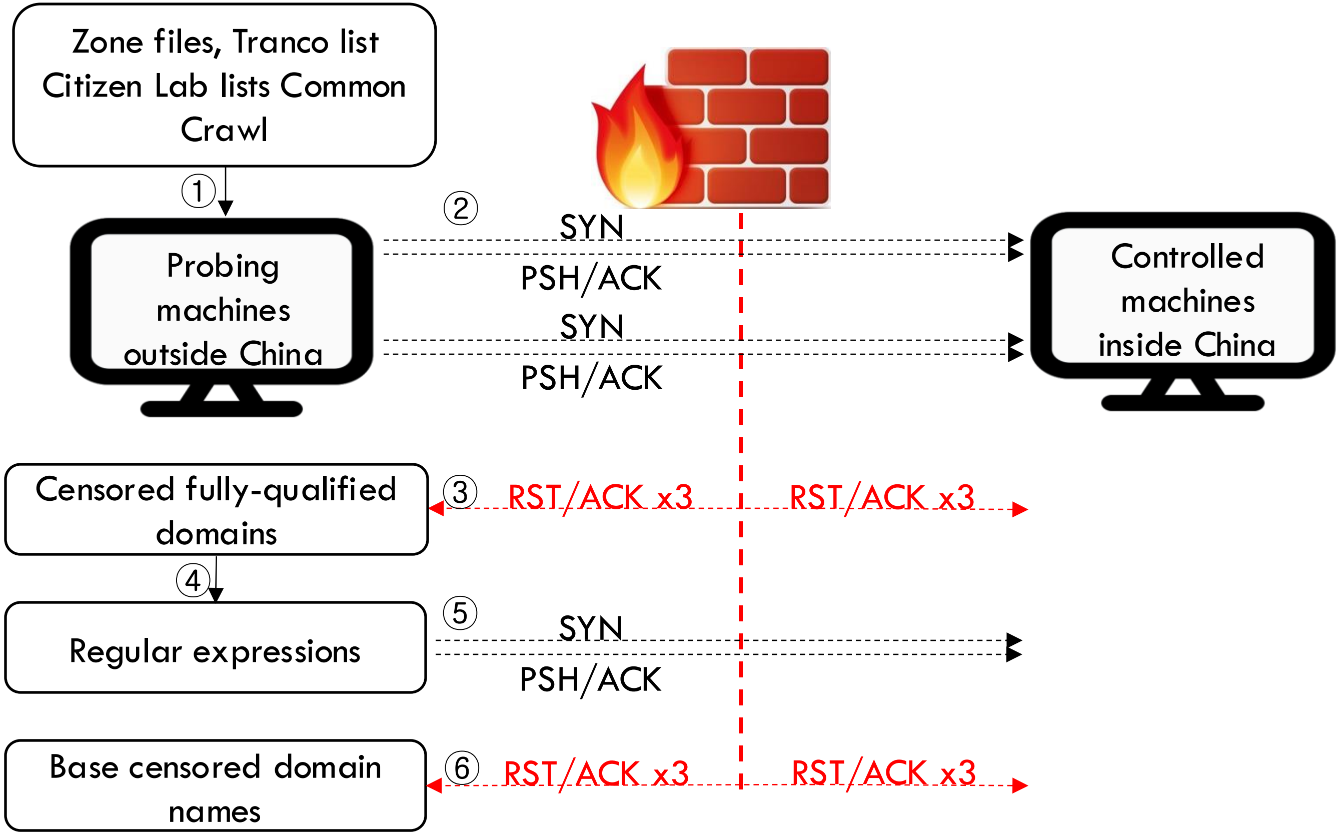
- Loss-tolerant



Loss-tolerance is a design choice, not a bug



GFWeb design



Zone files, Tranco list Citizen Lab lists Common Crawl

Probing machines outside China

Controlled machines inside China

Censored fully-qualified domains

Regular expressions

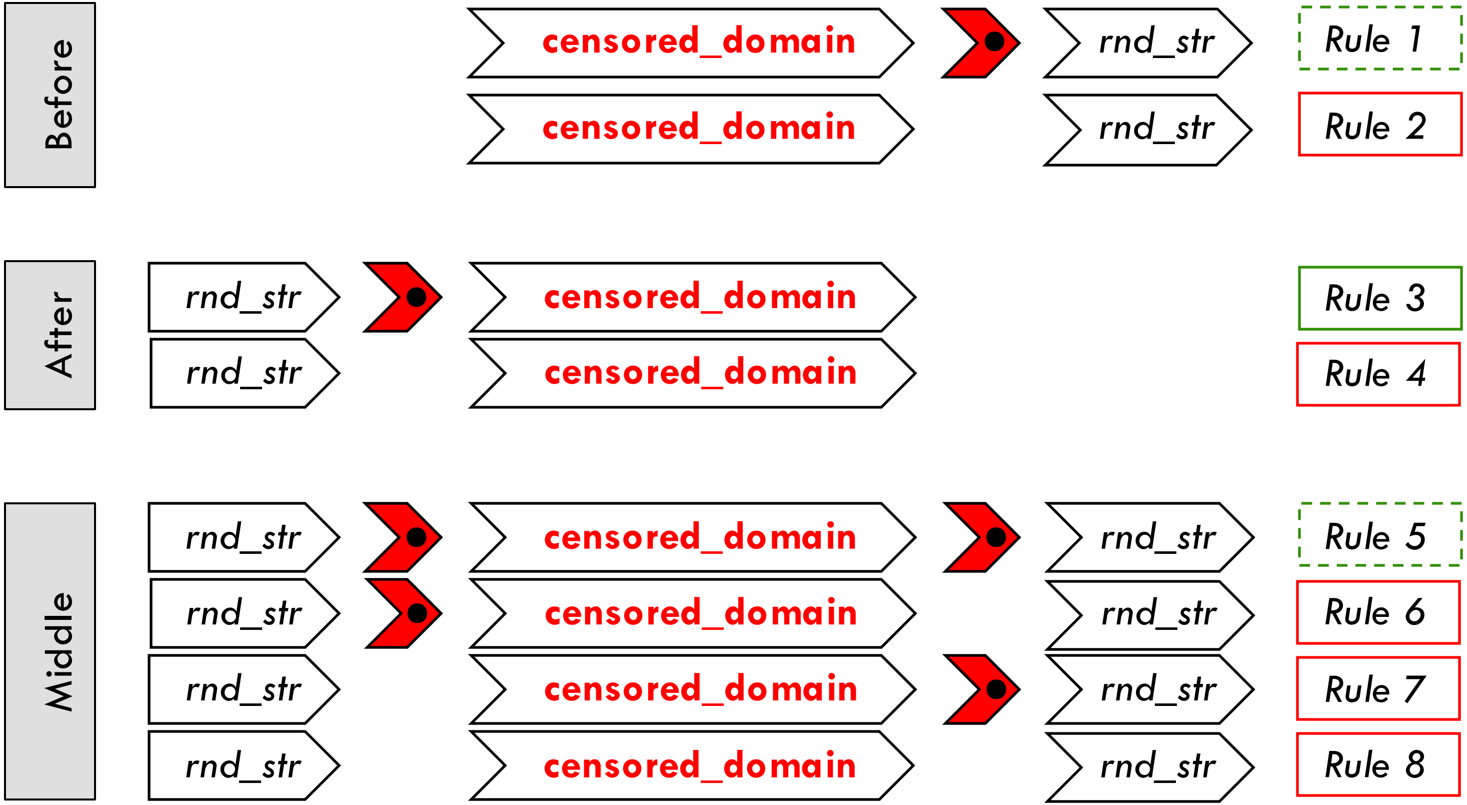
Base censored domain names

Base censored domains probing

- Goal: find the shortest domain that triggers blocking
- Method: test 8 permutations of the domain + random strings

→ More precise counting of censored domains

Base censored domains probing

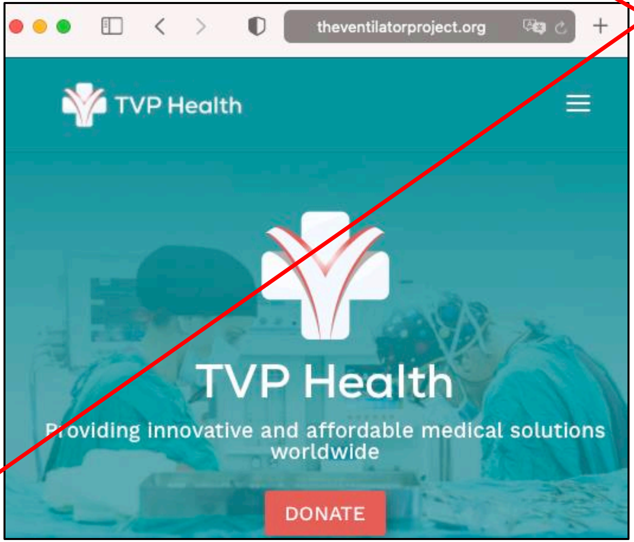


Over-blocked domains

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

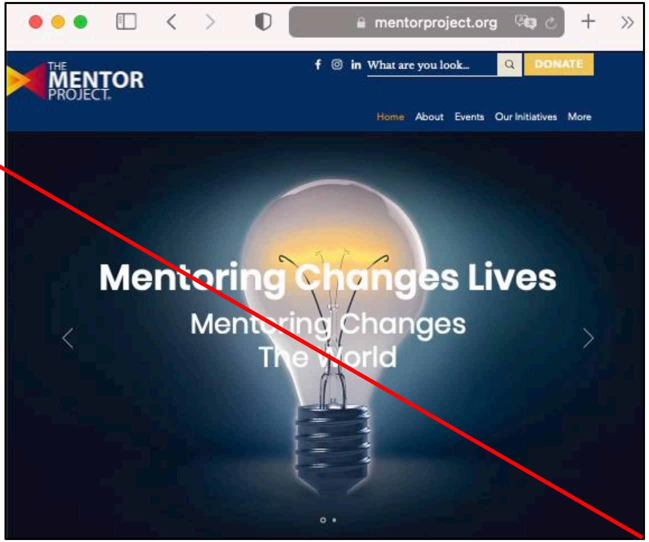
**torproject.org*   *NOW FIXED* → **.torproject.org*

→ Any domains ending with torproject.org are censored



TVP Health

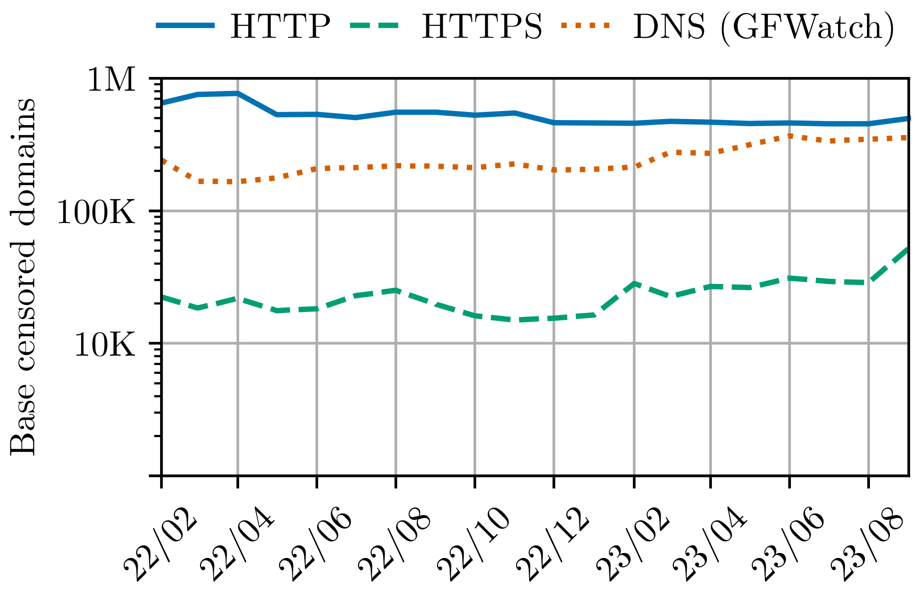
Providing innovative and affordable medical solutions worldwide



Mentoring Changes Lives

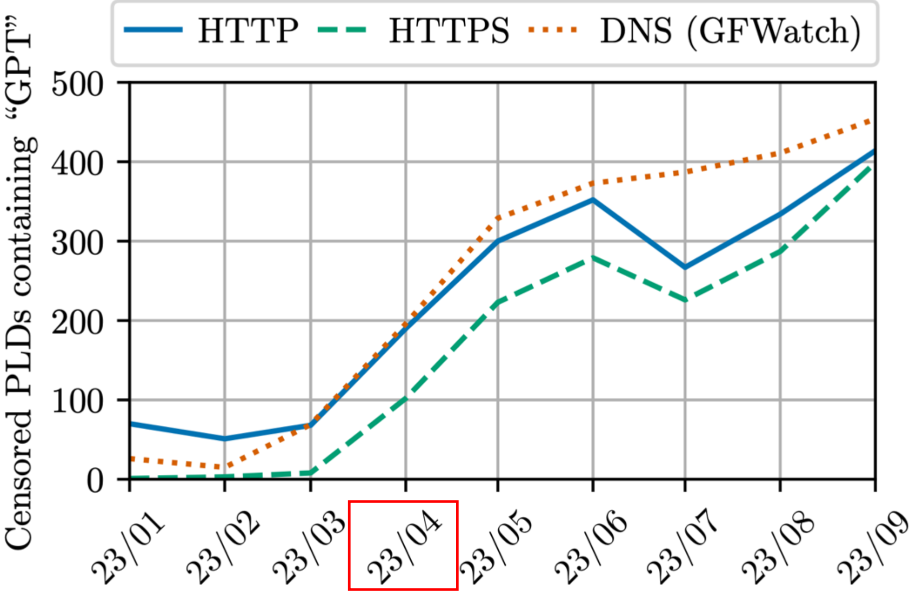
Mentoring Changes The World

Base censored domains



In average, GFWatch and GFWeb discover 528K, 247K, and 24K domains/month blocked by the HTTP, DNS, and HTTPS middleboxes.

Censored AI-related domains



On April 11, 2023, the China's Cyberspace Administration released draft measures for regulating generative AI services → popular AI tools blocked

Key contributions of GFWeb

✓ Discovered Web blocking behaviors of GFW:

- Different blocklists for DNS, HTTP, and HTTPS
- Fixed overblocking rules
- Asymmetric filtering (more detail in paper)

✓ Implications on

- Measurement based on a single protocol
- Probing-based censorship evasion strategy
- External measurement based on bidirectional filtering

I am hiring @UBC. Let’s make the Internet a better place!

https://np-tokumei.net

@NP_tokumei

https://GFWatch.org

https://GFWeb.ca
(To appear next month)