GFW: Protection or censorship?

A brief overview on China's Great Firewall system

Hengyi Li

January 14, 2024

I Introduction

The Internet has come a long way today. As a bridge for the flow of information, it carries the thoughts, ideas, opinions and needs of all people. However, because of the value attributes of information, the flow of information may inadvertently touch the interests of certain people or groups. This may lead to theft by criminals, or it may be a threat to national security, but in any case, the network firewall has been built as a result. In China, there also exists such a wall, which is not only able to defend against cyber-attacks from outside the country, but at the same time, and most importantly, prevents the flow of network traffic from within the country to outside the country. This is the subject of this article - The Great Firewall.

II Background

Great Firewall (hereinafter referred to as GFW) is a traffic censorship and attack defense system based on Cisco's intrusion detection system [5]. This defense system combines active and passive detection capabilities to censor China's Internet traffic [1]. As far as we know from our testing, the system is running on IPV4, but we have no way of knowing if the system is compatible with IPV6 as well. The main reason for this is that IPV6 is not very popular in China, and even if we could ask the carriers to enable it, they would not be very willing to do so for security reasons, and it would require a series of very complicated procedures to complete.

What's worse, the GFW is still a black box system for outsiders and researchers, and the existence of the system is not recognized at the national level, and there is no official documentation about the system. All research into the system is currently in the "poke around and see what happend" phase, so most of the results and findings below come from community organizations and web forum members. We would like to thank the members and contributors of the gfw. repot [3] website for providing a wealth of documentation and experimental data without which we would not have been able to gain such a deep understanding of GFW. In order to begin to understand how the GFW network in China works, we need to first understand the composition of the network structure in China, which has a huge difference with here in the U.S.

Now, China's Internet is divided into two parts, the national intranet and the international network. Since 2013, China's internet technology companies have grown and expanded into a wide range of industries, creating an ecosystem of apps for every need and an internal loop. For example, we have Bilibili, which is the Chinese version of Youtube. We have Baidu, which is the Chinese version of Google We have Weibo to replace the Twitter, we have WeChat to replace the Facebook and

WhatsApp and the alternate to Apple Pay/Google Pay and a series of other payment tools in China is Alipay and WeChat Pay. Every application you know, we have a alternative version and it's optimized just for Chinese market. Everything we need is in the intranet, so most of people choose to live in a environment like this. But what it brings is the information cocoon. Because the Chinese language is so special, learning English can be much more difficult for Chinese people than it is for Americans learning Spanish. The rising cost of learning means that most people won't bother to learn the language, and thus the breadth of the information stream they receive on a daily basis will become narrower and narrower. Unfortunately, because of the speciality of the Chinese language and its strong historical and cultural ties, the Chinese language media is mostly limited to the territory of China and the GFW exacerbates this situation by making it more difficult to access foreign media.

That's not what we want, and it's not what we want to see the Internet will become. Information should flow freely, and anyone should be able to access as much information as they want without compromising national security. Communication between people should not be hindered by the advent of technology, and unfortunately, GFW is one of the technologies that builds barriers. It didn't come about by accident, but because of a combination of historical legacies and improvisational decisions. For the sake of safety, we will only discuss its technical aspects here, and we would start by discussing the network architecture legacy.

III China's internet architecture

Because of the limited number of IPv4 addresses and China's large population base, China is facing the problem of running out of IPv4 addresses earlier than other countries in the world. At that time, IPv6 protocol had not yet been proposed. As a response, the three major carriers in China coincidentally chose carrier-grade NAT when setting up their networks, which effectively reduced the use of public IPv4 addresses by converting a large number of private IPv4 addresses into a small number of public IPv4 addresses [2].

Therefore, the Chinese network architecture now becomes based on NAT technology. Normal people will not have a public IP address, but all private IP addresses provided by ISPs, and, because China's Internet, cell phone and TV services are monopolized by three state-owned companies, China Mobile, China Telecom and China Unicom. The network architecture has become an onion-like NAT bridging model. Because of this bridging model, network providers can isolate intranet access requests from international access requests very easily. In order to meet the demand for international access, China has set up international line servers in three major cities, Beijing, Shanghai and Guangzhou, and deployed GFWs on them, which can be interpreted as network outbound customs, so that all outbound traffic must pass through the servers in one of the three cities and go through the GFWs' scrutinizing process before going out of the country [6].

Let's simulate a situation for better understanding. Suppose we have a laptop in Chengdu trying to access the ucdavis.edu website, then the request will go through the following process

$$\begin{aligned} &\textit{Laptop} \rightarrow \text{Home Router} \rightarrow \text{Community Server} \rightarrow \text{Distinct Server} \\ &\rightarrow \text{City Datacenter} \rightarrow \text{Province Datacenter} \rightarrow \text{Guangzhou Datacenter} \\ &\rightarrow \text{GFW server} \rightarrow \text{UCDavis server} \end{aligned}$$

As you can see, it takes a long way from Chengdu to connect to the server at UCDavis. At the same time, keep in mind that each hop is accomplished by IP conversion through NAT layer by

layer. And in the final pass through the GFW is also like going through customs, the GFW will review the legal compliance of the destination address, and whether it is on the blacklist. If the address is on GFW's blacklist (e.g. Youtube.com), GFW will reject the connection and send a reset request to both parties or wait until the connection times out. I'll discuss GFW's blocking patterns further in the next section.

IV GFW's Working Principle

Since the the way that https and http protocol works, the GFW would also behave differently on those website who use those protocol. Since the content of the http protocol is visible in plaintext, it is possible for the GFW to look at the content of the request and find out if there are any blacklisted keywords [4]. Once the keywords are found, the system determines that the user is accessing a site that belongs to the disallowed list and blocks the connection. The following shows the test we did

IV.I Experiment 1: Http connect situation

In this test, we will use a host in Chengdu, Sichuan, China to try to connect to port 80 of google.com without any proxies, that is, using http protocol.

References

[1] Anonymous, et al. *"How China Detects and Blocks Shadowsocks."* GFW Report, 29 Dec. 2019, gfw. report/blog/gfw_shadowsocks. Accessed 11 Jan. 2024.

[2] Cisco.com. "Carrier Grade Network Address Translation." *Information About Carrier Grade Network Address Translation*, www.cisco.com/c/en/us/td/docs/routers en/us/td/docs/routers/ios/config/17-x/ip-addressing/b-ip-addressing/m_iadnat-cgn.pdf. Accessed 11 Jan. 2024

[3] Great Firewall Report. *"Great Firewall Report."* Great Firewall Repor, Dec. 2019, gfw.report. Accessed 11 Jan. 2024.

[4] Hoochanlon. *"Fq-book/Docs/Abc/gfw.md at Master · Hoochanlon/Fq-book."* GitHub, 13 Apr. 2020, github.com/hoochanlon/fq-book/blob/master/docs/abc/gfw.md. Accessed 14 Jan. 2024.

[5] Schaack, Beth Van. *"China's Golden Shield: Is Cisco Systems Complicit?"* Center for Internet and Society, 24 Mar. 2015, cyberlaw.stanford.edu/blog/2015/03/china%E2%80%99s-golden-shield-cisco-systems-complicit. Accessed 11 Jan. 2024.

[6] 聿纾. "国内主流网络运营商国际连接线路简谈." 知乎专栏, zhuanlan.zhihu.com/p/64467370.