A Comparison of Censorship Evasion Techniques Under the Great Firewall of China

A Technical Report submitted to the Department of Computer Science

Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science, School of Engineering

George Noonan
Spring, 2021

Technical Project Team Members
George Noonan

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Signature _____ Date _____
George Noonan

Approved _____ Date _____
John Stankovic, Department of Computer Science

Approved _____ Date _____
Yonghwi Kwon, Department of Computer Science

A Comparison of Censorship Evasion Techniques Under the Great Firewall of China

An Undergraduate Technical Report

George Noonan
School of Engineering and Applied Science
University of Virginia
Charlottesville, VA
gn8fe@virginia.edu

ABSTRACT

Several studies have explored information controls and censorship evasion techniques; however, a comparison of these techniques and tools and their effectiveness under the modern censorship apparatus in China has yet to be explored. Here we report on the efficacy of satellite television, proxy servers, end-to-end encryption and server-side evasion techniques in bypassing blocked internet content. Since the Great Firewall of China is the most formidable and sophisticated censorship implementation in the modern world, we analyze whether and to what degree each tool would be viable to avoid Chinese censorship. Ultimately, this will enable people living in China and other authoritarian countries to gain access to the information they deserve.

INTRODUCTION

For more than a decade, there has been technological race between censorship tools used by the Great Firewall of China (GFC) and evasive techniques for bypassing blocked content. The latter is developed by network users who seek an open internet with freedom of expression and uncensored information. Both sides have been rushing to for technical superiority. Before being able to analyze the effectiveness of such tools under the GFC, it is essential to have a basic understanding of how the censorship apparatus works in China and how it blocks internet content. This report assumes a basic knowledge in computer networks and cybersecurity, as both the background section and subsequent analysis will be in the context of the two subjects.

Beginning with background section, an overview is given of how the Great Firewall functions and other essential information on how the GFC has prevented some of the largest and most powerful corporations - like Twitter and Google – from providing their services to Chinese users. Understanding this is essential before being able to analyze which evasive tools are viable under the modern iteration of the GFC. This will be the focus of the related work section which will summarize various tools and research that have been used to bypass internet content restrictions – in China and in other countries. The knowledge of the GFC from the first section, combined with the various censorship evading tools discussed in the related work section, will allow for an analysis on the efficacy of the tools under the Chinese censorship apparatus.

BACKGROUND

The internet infrastructure that composes the Great Firewall of China consists of three main mechanisms to block content over the country's internet. First, it uses IP blocking to stop traffic flow from any server. This method is effective at blocking specific websites throughout the country because of the government's complete control over the internet infrastructure. The Ministry of Industry and Information Technology (MIIT) owns nearly every ISP and is the IP address authority, which enables the government to create their own protocols on all the major routers [2]. This allows the GFC to maintain a blacklist that is shared among the routers to drop any packets with a blacklisted destination IP address.

While effective, this technique also has weaknesses. For example, if a website is blocked and its hosting address changes, it would allow their site to be available again. With today's high availability of network resources through cloud service providers, it is quite a

feasible strategy to migrate a webhost and thereby change the server's IP address.

In addition to IP blocking, the Great Firewall also deploys constant TCP packet sniffing on its routers. According to a white paper published by the Global Internet Freedom Consortium, as the packets are being sniffed, each packet's data is compared against a list of blacklisted keywords. If a router finds a match, it will interrupt the TCP connection by sending a TCP RST packet to the source and destination addresses. This effectively cancels the TCP connection and stops subsequent packets from transmitting between the source and destination addresses. [4]

The final major mechanism comprising the Great Firewall's censorship apparatus is DNS injection. Similar in concept to how TCP packets are sniffed, DNS injection works in the GFC by using deep packet inspection (DPI) to monitor all the DNS queries. It focuses its DPI near international ISP gateways and several root DNS servers in the country– and thus a more efficient way to inspect queries than compared to separately polluting thousands of smaller DNS servers in China [1]. Essentially, the Chinese government has complete control over the internet infrastructure, so it is able to spy on and identify DNS queries to blocked sites like **twitter.com** and **Google.com**. Once the DPI identifies such a problematic query, it will send a DNS response to the original query address with a fake IP address. For example, if a user requests twitter.com, the query to translate twitter.com into an IP address will be hijacked by Chinese censors and the IP address will be changed to an invalid one. This will prevent the user from accessing the website.

While the Great Firewall is effective at blocking content for Chinese users, it can also carry unintended consequences for internet users. Consider the case of religious movement leader Falun Gong and his website which had been banned in China. The website contained sensitive information to the country's authorities and religious content that was not allowed under Chinese law. In MIT's oldest and largest newspaper, The Tech, it was reported that the Great Firewall was performing DNS injection on packets that contained **mit.edu** [7]. The address that the GFC injected into the DNS query was the same as Falun Gong's blocked website, which thus caused the MIT website to also be blocked within the entirety of China.

Not only does the GFC carry negative consequences for users inside China, but its DNS injection practice can also affect users in other countries through polluted DNS records [3]. Suppose a user in South Korea is accessing a website that is hosted in Europe. The website is unblocked in Europe and South Korea but blocked in China. However, China lies between the two areas geographically, so it is thus natural to route traffic through the country. Unfortunately, there are several root DNS servers hosted in China, which means that regardless of where the query originates, if it reaches the root DNS server in China then it will be censored [1,3]. Thus, the DNS records with incorrect addresses act as censors in the GFC, but they also pollute the DNS records of non-Chinese users.

Since it has now been established how the Great Firewall censorship apparatus works on a basic level, the next sections will explore censorship circumvention techniques that have been previously used around the world. Even though the primary purpose of the paper is to understand these tools in the context of the GFC, case analysis of tools used in Iran and Syria will allow for a better analysis of their efficacy in circumventing the Great Firewall of China.

RELATED WORK

1 Satellite-TV to Evade Iranian Censorship

In Iran, the internet is frequently throttled during periods of social and political unrest. In a recent study done by reputable technology magazine Wired, a former post-doctoral researcher at Stanford developed a tool called Toosheh, which allows users to download uncensored media using their satellite television. Media is packaged into daily chunks of data containing everything from news articles to videos to audio recordings – even including other tools to bypass censorship like Tor [8].

In bypassing Iranian internet infrastructure altogether, Toosheh provides a highly effective means to evade government censors over the internet. Although the owners of Toosheh did not purchase their own satellite to broadcast the daily media bundles, they rented one from a company based in the United Arab Emirates. This is important because it means the satellite is not subject to

Iranian law or jurisdiction. Additionally, the satellite is positioned directly above the citizens' satellite dishes and has remained unaffected by all of the government's jamming attempts [8]. This is completely unlike sending TCP or DNS packets over government owned ISPs because it circumvents the ISPs altogether.

Not only does satellite television provide an easy way around Iranian censorship, but it also overcomes other issues in the country's internet infrastructure. As mentioned previously, the internet is heavily throttled and also costs beyond what most can afford in Iran – especially in times of unrest. In addition, nearly seventy percent of Iranian households own a satellite television [8]. This makes the solution ideal for most citizens because they already have the means and do not need to purchase extra circumvention tools like a VPN. Even if a citizen could afford the extra tools, the links to download them would also likely be blocked by the internet censors. Toosheh does not have this issue because the Windows software for decoding the data is distributed on one of the satellite channels. This makes the tool accessible to the masses and goes hand-in-hand with the user-friendly approach of the tool's website **toosheh.org**: users will find step-by-step instructions and support for using the service.

2 Alkasir: A Syrian Proxy Tool

Iran is not the only Middle Eastern country to have tight internet controls, and the Syrian government has a practical monopoly over its internet infrastructure. According to Senior Lecturer at Södertön University in Stockholm Walid Al-Saqaf, Syria's internet is dominated by the government. President Assad's cousin owns the country's largest and monopolistic ISP company, SyriaTel, and the government tightly regulates content on the internet through the Ministry of Telecommunications Establishment [5].

To combat and measure censorship in Syria, Al-Saqaf build a tool called Alkasir, which had been installed over 72,000 times since the publishing of the paper [5]. The tool allowed users to report a blocked website, and it would transmit their location to a centralized database. The locations were transmitted to observe which websites were blocked nationally, a term the author defined as being banned from over fifteen separate ISPs. Once users reported a blocked website, they would be redirected to connect with an encrypted proxy server, which would allow them to access the blocked website [5]. Since there is end-to-end encryption established between the user and proxy, the Syrian government censors cannot decipher the internet traffic. All the censors will see is encrypted data that has a destination IP that is not blocked, ultimately allowing users of Al-Saqaf's program to bypass the content censors of Syria and other countries. The popularity of Alkasir is a sign of its success: from 2010 to 2012, the number of users of the program increased by a whopping seventy times. The majority of users were accessing social media – overwhelmingly Facebook – which is a tool known to conduce political demonstrations and protests [5].

3 Server-side Evasion

While the previous two methodologies for bypassing censorship relied on client-side techniques, one study published in SIGCOMM found it possible to bypass censors using purely server-side techniques. It is important to note that these techniques are imperfect, and some of the techniques are better suited for certain protocols. Although the study compares different techniques for different protocols in several countries, its analysis under the GFC is the focus of this section. One important observation the researchers found is that each protocol – DNS, FPT, HTTP, HTTPS and SMTP – has a separate network stack in the Great Firewall, which thus requires separate evasion strategies for each protocol.

The first strategy discussed in Kevin Ball et al. [6] takes advantage of TCP simultaneous open to overcome the TCP sniffing and modification in the Great Firewall. In TCP/IP, simultaneous open occurs when both the hosts send SYN packets to each other at the same time. Normally, a three-way handshake in TCP would involve a SYN from the client, SYN+ACK from the server, and a final ACK from the client to establish a connection. From the background section, it is known that the GFC sends TCP RST packets to the sender and destination address when a website containing any blocked keyword is accessed. This causes the connection to terminate and thereby prevent the user from accessing the content.

Kevin Ball et al. [6] details a technique where a server responds with RST+SYN instead of SYN+ACK. The first RST packet will generally be ignored by most modern clients and operating systems. However, the second SYN pack will be interpreted as a simultaneous open by the client and it will send a SYN+ACK in response to the server. In describing this technique, the authors are careful to point out that the strategy does not work because the RST establishes a new connection in the other direction: with the server sending the SYN packet. They note that it is likely due to a bug in the synchronization of the GFC where it synchronizes on the SYN+ACK response from the client, which ultimately confuses the GFC into thinking the connection has been established.

In their experiments, Kevin Ball et al. [6] found that about half of the time, he could evade the GFC censors over HTTP. This strategy was also effective for DNS, FTP and SMTP, although it was not very effective for HTTPS. Observe the RST injection strategy below.

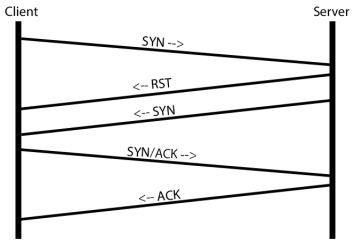

Figure 1: Using RST + SYN response to trigger simultaneous open and bypass HTTP, DNS, FTP and SMTP censor protocols.

Similar to the first strategy in Kevin Ball et al. [6] in taking advantage of TCP simultaneous open, there is a more successful version for HTTPS, although a bit less successful over FTP and SMTP. It involves sending two SYN packets instead of the RST+SYN combination in the first strategy. Both SYN packets are sent simultaneously, so they both are sent through the Great Firewall before the client responds. Because of the GFC's synchronization issue, the second SYN packet will cause the GFC to resynchronize on the next packet sent about half of the time[6]. This packet, it so happens, is a SYN+ACK response from the client. The packet causes the GFC to resynchronize on its response, SYN+ACK, which successfully establishes a TCP connection over HTTPS.

RESULTS

Although this paper has covered examples of censorship circumvention in other countries, the primary purpose of this section is to analyze the efficacy of each technique under the Great Firewall of China. The background section has already shown that the GFC is not the firewall it is meant out to be, but rather it is a collection of flawed mechanisms designed to interrupt connections involving banned content. How does each tool fare under the Chinese internet system?

Satellite television is the most effective of the three techniques to circumvent the GFC for the same reason as it is effective in Iran: it goes beyond the scope of the government's internet infrastructure and control. The Chinese government is the sole IP authority and has total control over the country's ISPs, so any censorship evasion strategy that uses broadband or cellular signal in China will need to overcome the government's deeply embedded censors. Being able to download blocked content through satellite television would remove GFC from the equation, so long as the satellite is owned by a country with an open internet. In addition, Chinese users would be able to easily use the software as they can download it directly from the website like with Toosheh. Despite its effectiveness at accessing blocked content, satellite television is not as common in China as it is in Iran (seventy percent of households have satellite television) and the Chinese authorities likely have more resources to force their citizens to use their own networks and not satellite ones that they cannot control.

While not as accessible as satellite television, Alkasir and the use of an encrypted proxy server to bypass censorship is also a viable strategy under the Great Firewall. The main vulnerability of this method is that the proxy server's IP address is banned by the GFC. However, this is unlikely for a few reasons. First, the traffic is encrypted end-to-end, so the packet sniffers will not see any comprehensible data traveling. The censors would thus not be able to detect illegal content and would not block the proxy's destination address. Additionally, a proxy can be hosted on virtually any address, so the government cannot feasibly stop this method – especially in today's world of highly available computing and network resources from cloud service providers. The GFC would have no way to know what the traffic is, and if they ban the IP, it might carry risks of banning unintended

services, as was the case with **mit.edu**. The last reason the proxy IP addresses are unlikely to be banned is because of their use in business. Many businesses need to communicate with others around the globe, so access to geo-restricted content and privacy are essential. It is in the governments interest to provide a good business environment, rendering it unlikely that proxies or their similar cousins VPNs will be banned in the country anytime soon.

Unlike the cases of Syria and Iran, the server-side evasion research was already applied to the Chinese censorship apparatus. In total, the researchers devised eight strategies for bypassing the different internet protocols in China, each having a different success rate. The first strategy discussed achieved high success rates for DNS (89%), HTTP (52%) and SMTP (70%) but a very low success rate for HTTPS (14%). The second method that sent two SYN packets, however, achieved (55%) effectiveness over HTTPS, and slightly lower success rates on the other protocols compared to the first strategy. The important point to note is that successful strategies were produced for each protocol, thereby proving this method works under the Great Firewall. Depending on the protocol, however, one strategy is more effective and preferred in relation to the others.

CONCLUSIONS

It has been shown that the Great Firewall of China is not a firewall after all. Rather, it is a system composed of mechanisms that monitor all the internet traffic in the country and intercept packets containing blocked content. Furthermore, the system has been shown to have flaws, which were proven exploitable in China and other countries with tight internet controls. This should give hope to anyone living in China who seeks open access to blocked content and the truth they deserve.

FUTURE WORK

If I had a full-time job for an entire year to work on this system, I would create a custom tool specific to evading censorship controls under the Great Firewall. My research already covers methods that have been previously used and whether they would be effective in China, so the next logical step would be to combine this research into a practical implementation. Toosheh was created specifically to bypass Iranian censorship and Alkasir to bypass Syrian censorship, although there has not been a tool proposed to bypass solely the GFC. Since the Great Wall has considerably more resources and is more powerful than that of Iran and Syria, the tool would likely involve a combination of the aforementioned evasive techniques. For example, the satellite television method worked great in Iran because users could download the necessary decoding software from the satellite. The tool proposed would need to be similarly accessible to Chinese netizens, otherwise they would have no way to use it. This could be achieved by relying on server-side techniques to transmit the software to the user, and then using an encrypted proxy like Alkasir as the actual censorship evading software.

REFERENCES

[1] Anderson, D. 2012. Splinternet Behind the Great Firewall of China. Queue. 10, 11 (Nov. 2012), 40–49. DOI:https://doi.org/10.1145/2390756.2405036.

[2] Xu, X. et al. 2011. Internet Censorship in China: Where Does the Filtering Occur? Passive and Active Measurement. Springer Berlin Heidelberg. 133–142.

[3] Anonymous. 2012. The collateral damage of internet censorship by DNS injection. SIGCOMM Comput. Commun. Rev. 42, 3 (July 2012), 21–27. DOI:https://doi-org.proxy01.its.virginia.edu/10.1145/2317307.2317311

[4] The Great Firewall Revealed, http://www.internetfreedom.org/files/WhitePaper/ChinaGreatFirewallRevealed.pdf

[5] Al-Saqaf, W. (2016). Internet censorship circumvention tools: Escaping the Control of the Syrian Regime. Media and Communication, 4(1)http://dx.doi.org.proxy01.its.virginia.edu/10.17645/mac.v4i1.35

[6] Kevin Bock, George Hughey, Louis-Henri Merino, Tania Arya, Daniel Liscinsky, Regina Pogosian, and Dave Levin. 2020. Come as You Are: Helping Unmodified Clients Bypass Censorship with Server-side Evasion. In Proceedings of the Annual conference of the ACM Special Interest Group on Data Communication on the applications, technologies, architectures, and protocols for computer communication (*SIGCOMM '20*). Association for Computing Machinery, New York, NY, USA, 586–598. DOI:https://doi-org.proxy01.its.virginia.edu/10.1145/3387514.3405889

[7] Winstein, Keith J. 2002. China Blocks MIT Web Addresses. (November 2002). Retrieved April 11, 2021 from http://tech.mit.edu/V122/N58/58web.58n.html.

[8] Greenberg, Andy. 2016. The Ingenious Way Iranians Are Using Satellite TV to Beam in Banned Internet. Retrieved April 10, 2021 from https://www.wired.com/2016/04/ingenious-way-iranians-using-satellite-tv-beam-banned-data/