Mark Jia, mij623, 11271998
Connor Tamme, lrk312, 11328286

Two parts:

1. Study the GFW and analyze the mechanism.
2. Study the circumvention tools, read source code.

Part 1 - Study the GFW

GFW is a highly complicated and evolving system created and maintained by the Chinese government.

Due to the system's complexity, for this part, will focus on reading the published technical papers and articles, including credible technical blogs, studying how the methods they used to analyze the GFW, and the results they obtained. Note that the GFW is a moving target, and the information may be outdated (but the technical details are still valuable). If time permits, will also try to reproduce the results.

GFW creates challenges for accessing the free (*as in freedom*) Internet over many different layers of the network stack. We will study the following aspects:

Parts of the GFW

Deep Packet Inspection (DPI): Can inspect the content (not just the header) of the packets over:

- DNS: Checking the DNS requests over TCP and UDP port 53, matching with blacklists.
- HTTP: Inspecting HTTP Host and Request and Response content.
- Tor: Dectect if the traffic is Tor traffic, then actively block it.
- SMTP: `MAIL FROM` and `RCPT TO`, even inspecting the email content (for keywords).
- TLS: Inspecting over TLS to detect visiting hosts and block them.
- Uses machine learning and user profiling (e.g. traffic patterns)

DNS Poisoning:
aka DNS cache pollution. GFW-executed DNS abduction to have DNS servers cache incorrect IP addresses for a domain name.
GFW injects fake DNS responses to query initiator, and will always arrive before the real

DNS response (due to priority in network topology).

Bidirectional pollution:

- can't query foreign domain names
- spoofing DNS queries by-passes China in its route. (in 2012, 26.41% of DNS server globally were impacted -> SigComm 2012 paper)
- sometimes pollutes domestic DNS China as well.

DNS using UDP, so it's easy to spoof.

TCP RST Injection:
GFW sends a TCP RST packet to both the client and the server, to terminate the connection.
Continuning censor after first blocking (demo by Alan Eustace).

BGP Hijacking (IP/Transport Layer Port blocking):
Uses BGP hijacking to blocking IP prefixes.
Works well for firms using virtual servers. Not so well for sites with CDN.

(read about *Black Hole* routing and Fang Bingxing's Paper)

limit QoS for foreign UDP traffic (such as QUIC)

Active Probing:
GFW actively probes the client and server to detect the circumvention tools.
Including on the SSH, VPN and Tor.

Defense (Great Firewall) --> Actively attack (Great Cannon)
(DDoS on sites. e.g. GitHub in 2015)
readings:
[https://theinitium.com/article/20150904-mainland-greatfirewall]
[https://web.eecs.umich.edu/~zmao/Papers/china-censorship-pam11.pdf]
[https://github.com/fightgfw/gfw-papers]

Part 2 - Study the circumvention tools

Basis is to use a middlebox instead of direct connection between the client and the server.

Simplest: Proxy server

SSH over VPS

- SSH tunneling (or VPN) over a VPS to bypass direct connection.
- GFW can still use active probing methods and DPI to know what's going on.
- GFW can also block the IP of the VPS.
- Cheap, so even if blocked, can switch to another VPS. Can be deployed in a distributed manner.

Traffic disguising

Some mainstream VPN uses obfuscation to disguise the traffic as HTTPS or other protocols, to not be caught by DPI.

Each firm develop their own protocol.

Case Study on tools

fqrouter

Android proxy app, allow integrating with GoAgent, Shadowsocks, WebVPN, etc. Can share the service to other devices using WiFi hotspot. ()

GoAgent

Using GAE (Google App Engine) as a proxy server. Can be used to bypass GFW. Google's Anycast network can help to bypass the GFW. Risks of identity exposure.

Shadowsocks

This is our main focus in the entire project:

A framework that allows many, including self-deployed encryption methods and listening ports.

Even when traffic pattern is detected with DPI, can simply modify parameters and ports to continue to work, very useful for doing the *guerrilla warfare* against the GFW.

Has many differnt implementations, including C, Python, rust, Go. We first look at the original implementation by @Clowwindy, then pick one to study in detail.

We are going to study the source code of Shadowsocks, understand how it works, analyzing its structure and architecture. If time allows, we can also try to make some modifications regarding the newest state of the GFW, or testing protocols with it.

Closing

May include and reflect on a few things.

- comments (disappointments) by @Clowwindy from the experience of developing Open Source software. (Disappointment with the community)
- Self-censorship and the Chilling Effect. Maybe the real wall is in one's mind.
- Congress Hearing Feb 15, 2006, *THE INTERNET IN CHINA: A TOOL FOR FREEDOM OR SUPPRESSION* (does Tech-giant's really made a net benefit?) along with the current cyber ecosystem in China.