New Selection Algorithm to Secure Tor Connection between Client and Guard
Torproject's Response Mechanism on Controversial Incidents

A Thesis Prospectus
In STS 4500
Presented to
The Faculty of the
School of Engineering and Applied Science
University of Virginia
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Computer Science

By
Siyang Sun

November 20, 2022

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.

ADVISORS

Yixin Sun, Electrical and Computer Engineering

Kent Wayland, Department of Engineering and Society

Introduction

Torproject's Response Mechanism on Controversial Incidents. Unsurprisingly, China, the most populous country on the planet, also has an enormous netizen - Merriam-Webster defined netizen as a participant in the online community - population of 1.03 billion as of 2021 (Thomala, 2022). Unlike netizens in other countries, most, not if all, of the 1.03 billion netizens in China could not access conventional western media - such as Twitter, Instagram, and Google - due to the implementation of the Great Fire Wall (GFW) by the Chinese Communist Party (CCP). Not only does GFW block users from accessing certain pages, but it also forces Internet Service Providers (ISP) to "record and retains data about the number of time users spend online, their account numbers, their I.P. addresses, and their dial-up numbers" (Chandel et al., 2019). Admittedly, users in China could temporarily bypass the GFW through Virtual Private Network (VPN); however, with the advancement of GFW technology, Chandel et al., 2019 found that the GFW can identify weaknesses and stop connection traces in VPNs. With the fall of VPN technology and ever-tightening control in China, there needs to be an alternative tool for whistle-blowers and curious netizens in China to access the general world wide web.

The technical topic of this prospectus - The Onion Router (Tor) - is an alternative technology used by censored netizens, such as those in China, to bypass censorship and access the web anonymously. Tor chooses three servers from over 6000 volunteers who ran servers (Tor Metrics, n.d.) and routes data securely using secure & private tunnels through these three chosen servers to the destination web server (*About tor browser*, n.d.). Under Prof. Yixin Sun's advisory, I am exploring new algorithms to choose servers that could provide more security and anonymity to Tor clients.

The added anonymity of Tor provides freedom to netizens in censored countries. However, this anonymity also helps criminals to hide and escape from justice on the web, especially the dark web. My STS topic, under the guidance of Prof. Kent Wayland, will discuss how torproject, a U.S. non-profit group that develops and manages Tor, handles the tradeoff between Digital rights - "right to privacy and freedom of expression, in the context of digital technologies" (Chen et al., 2022) and the growing threat of cyber crimes based off of anonymous services like Tor. Studying this tradeoff will provide further clarity on how Tor is changed by public opinion as well as the effect of Tor on society.

New Selection Algorithm to Secure Tor Connection between Client and Guard

In general terms, my technical topic involves improving Tor algorithms to increase clients' anonymity when accessing the web. By using Tor, "The operators of the websites and services that you use, and anyone watching them, will see a connection coming from the Tor network instead of your real Internet (I.P.) address, and will not know who you are unless you explicitly identify yourself" (About Tor browser, n.d.). Tor achieves this anonymity by sending client traffic through three random servers, also known as a relay. The entry relay only knows the Tor client and the next hop: the middle relay. The middle relay only knows the entry relay and exit relay. The exit relay only knows the destination server and the middle relay. In this way, no server could know the Tor client's destination and origin.

Tor clients will avoid attacks from server-level adversaries by hopping through three random relays. However, on a higher level, the world wide web is made up of Autonomous Systems (AS) - Hawkinson & Bates, 1996 defined as a group of I.P. addresses run by one or more operators with a single and clearly defined routing policy – where each server on the web belongs to an AS. All packets on the web are sent and received by AS using their routing policy.

AS will send packets they received or originated from its I.P. addresses to the next hop until they reach the AS with the destination server. However, a malicious AS could falsely advertise for an I.P. Address that does not belong to them since AS-level routing is based solely on trust, where "autonomous systems implicitly trust the routes that are shared with them" (What is BGP? | BGP routing explained | Cloudflare, n.d.). Therefore the package could be hijacked by a malicious server when the malicious server pretends to be the actual recipient of the package through their routing policies. This AS-level hijacking is especially harmful since most malicious AS will route the packets back to the legitimate owner. Tor clients cannot even detect this hijack and the resulting leakage.

Some A.S.s have Resource Public Key Infrastructure (RPKI) that could validate if an AS is advertising the correct I.P. prefixes. RPKI consists of Route Origin Authorization (ROA), which is "a statement by the owner of an I.P. block, stating which AS number is allowed to advertise it" (Internet registry ROA and rpki 2018), and Route Origin Validation (ROV), which helps AS to "Verify whether an AS is authorized to announce a specific I.P. prefix" (Internet registry ROA and rpki, 2018).

Currently, Tor does not consider the RPKI status of the clients or the relays when constructing routes. This negligence in the relay selection algorithm could waste potential RPKI protection of the relay and client from packet hijacking. I am studying a new Tor relay selection algorithm that considers RPKI to reduce the number of AS-level hijackings. Using my study, I will explore the weakness of the current Tor relay selection algorithm and the possible advances in security and performance of my new relay selection algorithm.

I have used a python-based Tor simulator built by Abby Glaubit - a previous researcher working under Prof. Sun - to study the behavior of the current and new algorithms in a more controlled environment. I am implementing my newly proposed algorithm in the Tor source code in C to be tested in the shadow simulator, which could better resemble real network situations.

Torproject's Response Mechanism on Controversial Incidents?

In my STS topic, I want to study the social impact of Tor and the torproject's response to societal concerns. Because I want to find out how Tor affects cybercriminals and freedom seekers on the web and vice versa to help my readers understand how Tor affects its clients and, on the other hand, how Tor clients and the general public affects Tor's policy and development.

Tor provides journalists and whistle-blowers the tools to expose questionable and often unconstitutional acts by governments or large corporations. One of the most famous Tor users, Edward Snowden, said Tor "allowed him to fulfill his oath to the constitution of the United States and release information of public interest to the media about the abuses committed by the government through its mass surveillance program" (Thank you, Edward Snowden: Tor project, 2019). Snowden's act brought much-needed attention to online privacy and citizens' rights on the web. This incident built a positive public image and gained more support for Tor. This positive image could attract more developers, thus perfecting the tool.

On the other hand, Tor has brought the convenience of anonymity to cybercriminals, such as drug dealers, at no cost. Ironically, the Tor project was started by the U.S. Naval Research Lab, a government agency (The Tor Project: Privacy & Freedom Online, n.d.). However, this government-initiated project might bring more trouble to the government when enforcing laws on the internet. Criminal cases, such as the Silk Road takedown, have cast a dark shadow on Tor as a criminal tool, which has caused many organizations to ban connections to Tor relays.

In this context, the primary organizations involved are digital rights advocates, cybercriminals, law enforcement, authoritarian censors, and the torproject community. We could further group these into pairs of cyber criminals and law enforcement, digital rights advocates, and authoritarian censors. In the first pair, the law of many countries is broken since drugs are often banned in most countries accessible by Tor. In this sense, the common value of prohibiting drug are weakened by Tor, which further paints a negative image of Tor. This negative image will affect the goals and policies of the tor community. In the second pair, the common human value of freedom of speech is impaired by authoritarian censorship. In this case, Tor helps restore the value of freedom of speech in a relatively unfree country. In this scenario, the torproject community is positively impacted by such events, which will push for further perfecting Tor to help freedom seekers.

Numerous previous research admits the effect of Tor in both realms: cyber criminal and digital rights. "the analytical results show that Google searches for "Digital Rights" and "Drug Markets"-related knowledge are both positively associated with daily Tor usage," (Chen et al., 2022). We can see that the general public's association of the Dark web with Tor is not unfounded, and there is a statistical correlation between the two. But Chen (2022) also argued that there is a geopolitical aspect of this interpretation, where Tor users in non-free countries will be more likely to search for Tor-related knowledge when searching for Digital Rights issues compared with users in free and partially free countries. This quote shows that Tor's benefit to society may be less realized in western countries than in unfree countries.

I will first research torproject's rules and developer guides on their website to find out what policies they have regarding general goal setting and response to societal and law enforcement pressure. I will also read torproject's discussion on Github and its official blogs during a sensitive time when there are critical societal influences on Tor. This will help me explore how Tor handles the effect on its goal from the public.

Conversely, I will search for client comments and Tor usage data during censorship events to discuss how Tor has impacted society in the digital rights realm. This evidence will help me close the loop of discussion regarding Tor's effect on society and vice versa. From these discussions, I will inform my readers on how torproject and society handles the tradeoff, digital rights versus cyber crime, brought by Tor.

Finally, to gather first-hand data, I will also send out surveys to the student body to ask the public opinion on Tor among college students. This survey will expose if the public's concerns are mainly from the younger generations, who will have an increasingly important role in our future society.

Conclusion

Anonymous services have both positive and negative impacts on society. Through my STS research, I will search through Tor's design and discussion documents and society's opinion through journalist articles. Along with online questionnaires and data, I want to learn more about how torproject handles such tradeoffs and how society affects torproject's decisions.

The goal for my technical topic is to create and test a newly proposed relay selection algorithm. Specifically, I will test this algorithm in the Tor source code under the shadow simulator, which resembles actual Tor usage. I will summarize my findings in a report and submit the report to the Privacy Enhancement Technologies Symposium.

References

About tor browser. ABOUT TOR BROWSER | Tor Project | Tor Browser Manual. (n.d.). Retrieved October 25, 2022, from https://tb-manual.torproject.org/about/

Chandel, S., Jingji, Z., Yunnan, Y., Jingyao, S., & Zhipeng, Z. (2019). The Golden Shield Project of China: A decade later—an in-depth study of the Great Firewall. *2019 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*. https://doi.org/10.1109/cyberc.2019.00027

Chen, Z., Jardine, E., Fan Liu, X., & Zhu, J. J. (2022). Seeking anonymity on the internet: The knowledge accumulation process and global usage of the Tor Network. *New Media & Society*, 146144482110722. https://doi.org/10.1177/14614448211072201

Hawkinson, J. & Bates, T. (March 1996) Guidelines for creation, selection, and registration of an Autonomous System (AS). BCP 6, RFC 1930. https://doi.org/10.17487/rfc1930

Internet registry ROA and rpki. Network Direction. (2018, October 5). Retrieved October 25, 2022, from https://networkdirection.net/articles/routingandswitching/internetregistryroaandrpki/

Merriam-Webster. (n.d.). *Netizen definition & meaning*. Merriam-Webster. Retrieved October 25, 2022, from https://www.merriam-webster.com/dictionary/netizen

Thank you, Edward Snowden: Tor project. The Tor Project. (2019, December 25). Retrieved October 25, 2022, from https://blog.torproject.org/thank-you-edward-snowden-tor/

The Tor Project: Privacy & Freedom Online. Tor Project | History. (n.d.). Retrieved October 25, 2022, from https://www.torproject.org/about/history/

Thomala, L. L. (2022, March 30). *Topic: Internet usage in China*. Statista. Retrieved October 25, 2022, from https://www.statista.com/topics/1179/internet-usage-in-china/#topicHeader__wrapper

Tor Metrics. (n.d.). *Servers*. Tor Metrics. Retrieved October 25, 2022, from https://metrics.torproject.org/networksize.html

What is BGP? | BGP routing explained | cloudflare. What is BGP? | BGP routing explained. (n.d.). Retrieved October 26, 2022, from https://www.cloudflare.com/learning/security/glossary/what-is-bgp/