Analyzing the Possibility of Bitcoin Network Partitions Caused by Internet Censorship

Jorge Coll, Andrew Fasano, Benjamin Kaiser, Lucy Qin
MIT 6.892: Shared Public Ledgers – Final Project
wehavenoprivacy@mit.edu, *fasano@mit.edu*, *benjamin.h.kaiser@gmail.com*, *lucyzqin@gmail.com*

***Abstract*—Abstract goes here.**

I. Introduction

The Bitcoin cryptocurrency relies on a shared, global transaction ledger that each participant stores a copy of: a blockchain. We set out to investigate whether Internet censorship, which is widely utilized by many countries, might interfere with the Bitcoin blockchain by preventing those behind the wall of censorship from reaching consensus with those outside the wall. In particular, we were curious if someone intent on destabilizing the blockchain could craft a Bitcoin block that would be filtered by a country's censorship apparatus and what the effects on the blockchain might be in such a scenario.

We focused our efforts on the censorship conducted by China, known as the Great Firewall of China (GFW). We first conducted a research survey and performed some tests using a virtual machine located within China to determine how the GFW functions. Unfortunately, we were unable to construct a block that would be filtered in practice by the GFW. In order to pursue the idea further, we next conducted experiments to test the impact of a network partition caused by a block being filtered at a country's border. We believe our results are of interest because Internet censorship techniques are always changing, and although the GFW is not susceptible to such an attack, it is conceivable that it or another censorship regime may be at some point.

II. Bitcoin's networking protocol

The message transfer protocol that powers Bitcoin is defined abstractly in Nakamoto's original publication [1] and further fleshed out in `bitcoind`, the reference implementation he provided. In this section, we will provide a high-level overview of the portion of the protocol relevant to our work. In particular, because we focus on a partitioning attack that severs existing connections, we will not discuss how new clients are bootstrapped, how initial blocks are downloaded, or how further peer connections are established. This leaves the standard relay protocol, which consists of four message types:

- *inv*: when a transaction or block has been verified by a node, the node announces that the transaction or block is available by sending an *inv* message to all of their neighbors.
- *getdata*: when a node receives an *inv* message for a transaction or block that it does not already have a local copy of, it responds with a *getdata* message asking for that block.
- *tx*: when a node receives a *getdata* request for a given transaction, it responds with a *tx* message containing the transaction
- *block* when a node receives a *getdata* request for a given block, it responds with a *block* message containing the block

A node will only propagate a block or transaction to its neighbors if it can verify it. Unverified transactions or blocks are not propagated.

Partitions occur when two or more competing blockchain heads emerge. By itself, Bitcoin does not attempt to detect or mitigate partitions. If a partition appears, both portions will continue operating independently and their state will diverge. If the divergence results in subnetworks that produce incompatible transaction histories, they will be unable to merge if the partition is later removed. Decker and Wattenhofer [2] observe that detection of network partitions may not be challenging and could be achieved through simple monitoring of the observed aggregate computational power in the network. As Bitcoin has grown in size and influence, such monitoring has begun to take place, and as of this writing it is possible to easily view the total network hashing rate via public websites [3].

III. Traffic Censorship

Censorship of Internet traffic occurs in a variety of forms all over the world. Most common is censorship of Web traffic (e.g., HTTP, DNS, and other TCP protocols related to browsing the Web). The most aggressive censorship takes place in countries where citizens' access to the Internet is very limited to begin with (e.g., North Korea and Cuba). Verkamp and Gupta [4] conducted experiments to infer the mechanics of various countries' approaches to Web censorship and also provided a set of independent sources ranking countries based on the severity of their censorship.

Across our research, four countries were repeatedly identified as the most severe censors: North Korea, Cuba, Iran, and China. Of these, China is the most interesting case for our purposes for two reasons: (1) its Internet-connected population far surpasses the others' and (2) it is a hotbed of Bitcoin mining activity, with Chinese-run mining pools accounting for an estimated 70% of all Bitcoin mining power as of June 2016 (Figure 1). As a collection, China's censorship measures are colloquially referred to as The Great Firewall of China (GFW).

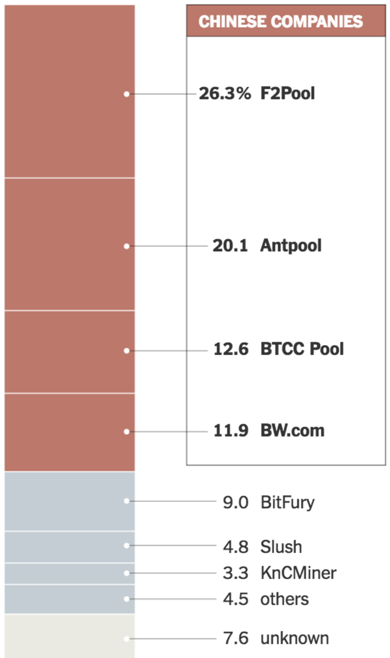

Fig. 1: From [5]: the shares of mined Bitcoin blocks from May 24 to June 24, 2016 by mining pool.

IV. The Great Firewall of China

There is no consensus about the precise technical underpinnings of the GFW, as conflicting observations have been made. In particular, there is conflict about whether or not it is stateful – i.e., whether it stores and uses information about the packets it intercepts[1]. However, the basic mechanisms employed are known to be[7]:

- *IP address filtering*: The GFW blocks specific IP addresses from receiving traffic by dropping all packets associated with it. This assures that the GFW's reach extends to all content produced by a host's IP, rather than just the few specified domains.
- *DNS misdirection (hijacking)*: When requesting a blocked host name, there are cases in which the DNS servers under the GFW will return a different IP address than the one that corresponds to the domain name requested. The Chinese government can effectively replace the content with material that is more favorable to their interests.
- *Keyword filtering*: If a banned keyword appears in a URL, after a completed TCP handshake, the GFW will send reset packets to both the source and destination, blocking access to the requested content. Even if a keyword is not explicitly in the URL but appears within the HTML response, the content is also denied. In this particular instance, pages often begin to display but are then truncated after the discovery of a keyword.

[1]The most recent results, in Xu et. al[6], indicate that the GFW does record state

A. Impact on Bitcoin

The only direct impact that the GFW has on the Bitcoin network is a minor delay added to every packet. Due to this and normal Bitcoin block propagation delays, Chinese miners will hear about blocks mined outside of China later than ones within the country, causing a communication barrier. This actually creates a disadvantage for those outside of China since the majority of the hash power in Bitcoin is located inside of China. This causes problems, most recently during the debate surrounding a proposed block size increase. If the size were to increase, Chinese miners would be subject to further delays, potentially jeopardizing profits. [**?**]

V. Network Partitioning

Since traffic monitored by the GFW is subject to keyword filtering, we wondered if a situation could arise in which a banned word within a transaction could prevent a valid block from being propagated within China. This could lead to a partition in the network, causing a fork in the blockchain in which those outside of the GFW and those within it build off of different blocks. Given the current implementation of the GFW, only network activity over HTTP is monitored and Bitcoin activity is therefore not subject to filtration. Features could at any point be introduced within the GFW that allow for this vulnerability.

VI. Experiments

This section will describe the four experiments we conducted. Our initial aim was to perform a partition attack on the Bitcoin blockchain by creating a transaction containing a word censored by the GFW. When we determined that this was infeasible in practice, we decided to perform a set of experiments to analyze what effects such an attack might have on Bitcoin. The first simulator we ran abstracted away too many of the important details of the Bitcoin protocol, focusing instead on collecting high-level statistics about the network. Our next step was to try to understand the expected behavior of a Bitcoin client located inside of the GFW by directly inspecting the source code of `bitcoind`. Finally, we worked to deploy and run a more detailed simulator. Unfortunately, configuration issues prevented us from running our intended simulations, so at the moment they are left as future work. However, we do feel that through this process, we developed strong hypotheses about how the network would behave in the presence of a censorship-based partition, and as future work we intend to resolve our configuration issues with the simulator and test those hypotheses.

A. Triggering the GFW

We attempted to construct a Bitcoin block that would be filtered by the GFW. Arbitrary data has been embedded into blocks before; the most common places that such data is stored in a block are the address fields (by encoding the data in hex), the coinbase transaction, or the scripts[2].

[2]In one notable case, a simple cross-site scripting (XSS) attack was embedded in a block and was apparently demonstrated to work on blockchain.info[8], although the site has since been patched to properly escape HTML.

In the end we determined that we would need to trick the GFW into interpreting a block as Web traffic of some sort (e.g., DNS or HTTP), which was not feasible. (To all, but especially Andrew: is there anything else we should say here about what we tried / observed? Maybe briefly describe how you connected to a VM in China and the types of tests you ran?)

B. NS3 Bitcoin Simulator

C. Bitcoin Reference Client Source

D. Shadow Simulator

VII. Limitations

Speaking broadly, experiments performed on a model are only useful to the degree to which the model accurately represents reality. Even though we were not able to instantiate our model or run our simulations, we plan to do so in the future, so we want to make explicit the limitations of our model and the simplifying assumptions that we made when designing our simulations.

- *Mining pools*: In practice, mining pools in which miners cooperate to find new blocks and share the rewards are common [9]. We model mining pools as a single miner who has the cumulative hash power of the pool. This fails to account for scenarios in which miners outside of the GFW contribute to a pool that is managed by a node inside of the GFW.
- *Other limitations?*

VIII. Future Work

IX. Discussion

References

[1] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system."

[2] C. Decker and R. Wattenhofer, "Information propagation in the bitcoin network," in *Proceedings of the 13th IEEE International Conference on Peer-to-Peer Computing*, 2013.

[3] [Online]. Available: http://bitcoin.sipa.be/

[4] J.-P. Verkamp and M. Gupta, "Inferring mechanics of web censorship around the world," in *Presented as part of the 2nd USENIX Workshop on Free and Open Communications on the Internet*, 2012.

[5] N. Popper, "How china took center stage in bitcoin's civil war," The New York Times, June 2016.

[6] X. Xu, Z. M. Mao, and J. A. Halderman, "Internet censorship in china: Where does the filtering occur?" in *Proceedings of Passive and Active Measurement: 12th International Conference*, 2011.

[7] M. Hu. (2011, May) The great firewall: a technical perspective. Torfox: A Stanford Project. [Online]. Available: https://cs.stanford.edu/people/eroberts/cs181/projects/2010-11/FreedomOfInformationChina/great-firewall-technical-perspective/index.html

[8] [Online]. Available: https://redd.it/1n57uj

[9] M. Rosenfeld, "Analysis of bitcoin pooled mining reward system," 2011.