LIÈGE université Library

http://lib.uliege.be https://matheo.uliege.be

Master thesis : Vivisecting Blockchain P2P Networks

Auteur : Ben Mariem, Sami
Promoteur(s) : Donnet, Benoît
Faculté : Faculté des Sciences appliquées
Diplôme : Master : ingénieur civil en informatique, à finalité spécialisée en "computer systems security"
Année académique : 2018-2019
URI/URL : http://hdl.handle.net/2268.2/6791

Avertissement à l'attention des usagers :

Tous les documents placés en accès ouvert sur le site le site MatheO sont protégés par le droit d'auteur. Conformément aux principes énoncés par la "Budapest Open Access Initiative"(BOAI, 2002), l'utilisateur du site peut lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces documents, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale (ou prévue par la réglementation relative au droit d'auteur). Toute utilisation du document à des fins commerciales est strictement interdite.

Par ailleurs, l'utilisateur s'engage à respecter les droits moraux de l'auteur, principalement le droit à l'intégrité de l'oeuvre et le droit de paternité et ce dans toute utilisation que l'utilisateur entreprend. Ainsi, à titre d'exemple, lorsqu'il reproduira un document par extrait ou dans son intégralité, l'utilisateur citera de manière complète les sources telles que mentionnées ci-dessus. Toute utilisation non explicitement autorisée ci-avant (telle que par exemple, la modification du document ou son résumé) nécessite l'autorisation préalable et expresse des auteurs ou de leurs ayants droit.

Vivisecting Blockchain P2P Networks Unveiling the Bitcoin IP Network

Sami Ben Mariem
Université de Liège – Montefiore Institute – Belgium
sami.benmariem@student.uliege.be

Pedro Casas
AIT Austrian Institute of Technology – Austria
pedro.casas@ait.ac.at

Benoit Donnet
Université de Liège – Montefiore Institute – Belgium
benoit.donnet@uliege.be

ABSTRACT

Blockchains are typically managed by peer-to-peer (P2P) networks, which provide the support and substrate to the so-called *distributed ledger (DLT)*, a replicated, shared, and synchronized data structure, geographically spread across multiple nodes. The Bitcoin (BTC) blockchain is by far the most well-known DLT, used to record transactions among peers, based on the BTC digital currency. In this paper we focus on the network side of blockchain P2P networks, characterizing their topology and main properties from a purely network measurements-based approach. We present a BTC crawler able to discover and track all the active nodes of the BTC P2P network through active measurements, and use it to analyze its main properties. We additionally discuss a passive measurements-based approach to reconstruct the topology of a blockchain P2P network, which can unveil the most important nodes of the blockchain, namely the miners.

ACM Reference format:
Sami Ben Mariem, Pedro Casas, and Benoit Donnet. 2018. Vivisecting Blockchain P2P Networks Unveiling the Bitcoin IP Network. In *Proceedings of ACM CoNEXT 2018 Student Workshop, Crete, Greece, December 4, 2018,* 3 pages.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

The Bitcoin (BTC) blockchain [3] is built on the top of a decentralized peer-to-peer (P2P) network, used to propagate relevant information such as transactions between BTC entities or blockchain updates. Given its relevance and popularity, many studies have characterized the BTC blockchain from multiple different perspectives, mainly in terms of executed transactions, through the analysis of the publicly available distributed ledger (DLT). For example, papers such as [4, 6] focus on the BTC transactions as observed at the BTC DLT, [11] studies the security of the BTC P2P network, [1] analyzes the temporal generation of BTC blocks, etc. In this paper we focus on the network side of the BTC P2P network, characterizing its topology and main properties from a purely network measurements-based approach. Understanding the underlying P2P network characteristics of the BTC blockchain can be highly useful for multiple relevant use cases, including security, performance and anomaly detection, as well as relevant nodes discovery. Other papers try also to infer the P2P network topology and characteristics of BTC and other blockchains [5, 7, 8]. However, the continuous evolution of DLT technology and the potential security issues linked to unveiled P2P topologies [9, 10] result in constant updates of the underlying protocols, turning some of the previous proposals [7] no longer applicable.

Based on popular past studies on P2P networks [9, 10], we present two different techniques targeting the characterization and discovery of blockchain P2P networks. In Sec. 2, we present a BTC client that can be used to crawl the full set of nodes composing the BTC P2P network, and gather information regarding those nodes which are *active* in the network. Using this crawler, we take a full snapshot of the BTC P2P network and describe relevant findings regarding its active nodes. In Sec. 3 we briefly describe our ongoing work on passive topology discovery for blockchain P2P networks, which not only allows to identify active links between the active BTC nodes discovered by the crawler, but also unveils highly relevant nodes such as the so-called miners. Note that while this paper focuses on the BTC P2P network, the techniques herein described are applicable to many other Blockchain P2P networks with similar functioning as the BTC one, including Litecoin, Ethereum, and the like.

2 BLOCKCHAIN CRAWLING

The BTC crawler is a customized Bitcoin software client which can recursively query all the BTC peers, discovered by asking other nodes for the underlying IP addresses. In the start-up phase, the crawler obtains a set of *seed* or well-known IP addresses, by querying a list of predefined DNS servers. Then it starts the crawling process, by requesting to each of these seed nodes their internal list of known peers, using a simple *getaddr* message. In BTC, each node keeps a list of up to 2500 peer addresses of tried and new connections to its neighbors [3]. When the crawler successfully connects to a node, it marks it as *active* and requests additional meta-data such as the provided services, the software version of the client and the P2P protocol version used by the BTC node. The services provided by a BTC node [3] can be useful to understand which type of nodes it is, for example, to understand if it has the minimum capabilities to be a miner or other important role, such as a full node. Once the crawling of the full P2P is completed, a set of both active and inactive BTC P2P peers is provided, along with their corresponding IP address, the meta-data and several network stats such as the latency to the node. As we show next, these measurements can be used to characterize the blockchain P2P network. The crawler heavily relies on multi-threading and multi-core platforms to improve crawling speed, currently taking about an hour to complete a full BTC snapshot from scratch.

2.1 Crawling the BTC P2P Network

We describe a single snapshot of the BTC P2P network produced by the aforementioned crawler on September the 10th, 2018. The crawler discovers more than 200.000 BTC IPs, with a total of 9673 active nodes. Fig. 1 presents several stats regarding these active

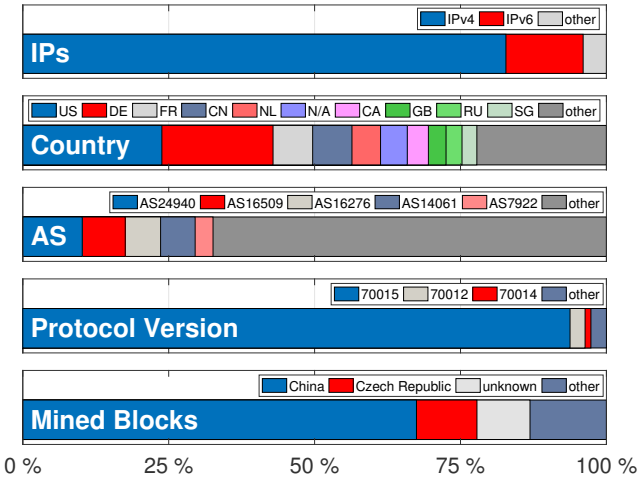

Figure 1: BTC nodes and network stats.

nodes. In terms of IP addresses, about 83% correspond to IPv4 addresses, 13% to IPv6, and the remaining 4% to nodes connecting to the BTC network through anonymous connections, e.g., using Tor - we rely on the public TorDNSEL service and the list of Tor exit nodes to identify these nodes. Fingerprinting the type of hardware hosting these nodes [12] is part of our ongoing work.

Using IP-geo-localization databases (such as MaxMind and Team Cymru), we geo-locate the IPv4 and IPv6 addresses by country, and obtain the ASes hosting them. Most of the BTC active nodes are located in the US (23.7%) and EU - Germany (19%), France (6.8%), Netherlands (4.9%), whereas a smaller share are located in China (6.7%) and other Asian countries (Singapore, Japan, South Korea, etc.). This is additionally confirmed by the propagation latency (min RTT) to the corresponding IPs, from our vantage point in Vienna, Austria [2]: Fig. 2 shows that about 50% of the IPs reside within EU (min RTT < 50 ms), 30% within North-America (100ms < min RTT < 200 ms) and about 10% at very far locations, such as eastern Asia (e.g., Japan, Korea, Hong Kong, etc.). Regarding the hosting ASes, a big share of nodes are hosted by major cloud/datacenter providers such as Hetzner (AS24940), Amazon (AS16509) and OVH (AS16276). Despite such an occidental concentration of active nodes, it is impressive to see that the far majority of BTC mining activity actually occurs in China. We use the information provided by the miners when producing a new block to identify the origin mining *pools* (see https://btc.com/stats), and observe that more than 70% of the BTC blocks are mined in China. The number of blocks mined by unknown or hidden miners is non-negligible, close to 10%.

Finally, in terms of node services, the biggest majority of the BTC nodes keep a full copy of the full BTC DLT (93.6%), and can therefore perform full validation of transactions and blocks; regarding P2P protocol version, while 93.8% of the nodes use the latest version of the BTC P2P communications protocol, there are at least 2.6% of the nodes using a very outdated protocol, from late 2015, which can potentially result in security or performance issues.

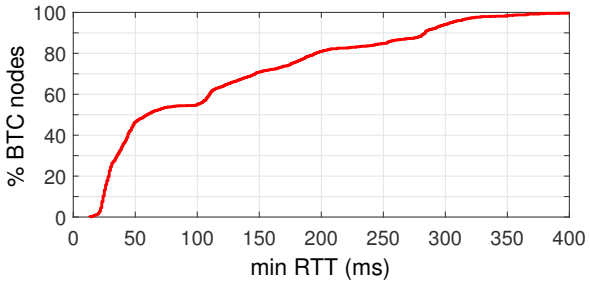

Figure 2: min RTT to active BTC nodes.

3 PASSIVE TOPOLOGY DISCOVERY

We use the snapshots of the BTC active nodes as starting points to infer the P2P topology of the BTC network. The approach can be applied both to discover active links between nodes, as well as to unveil hidden (unknown) mining nodes. In a nutshell, following previous work [9, 10], we rely on the broadcasting process of mined blocks and confirmed transactions over the entire list of active nodes to estimate the most probable node having broadcasted the corresponding block, based on a set of estimated latencies among nodes. By passively listening to such broadcast messages, and assuming the propagation delays between the vantage point and the BTC nodes are known (c.f. Fig. 2), the topology discovery technique can reconstruct the minimum spanning tree representing the path followed by the broadcasted message from the inferred source to the vantage point. Preliminary results under simulated controlled scenarios show that it is feasible to properly reconstruct the topology, if a sufficiently large number of brodcasted messages are listened by the vantage point, by merging individual spanning trees. Implementing and testing this topology discovery technique is part of our ongoing work.

REFERENCES

[1] R. Bowden, H. P. Keeler, A. E. Krzesinksi, and P. G. Taylor. 2018. *Block Arrivals in the Bitcoin Blockchain.* cs.CR 1801.07447. ArXiV.
[2] P. Casas, P. Fiadino, and A. B'ar. 2013. IP Mining: Extracting Knowledge from the Dynamics of the Internet Addressing Space. In *Proc. International Teletraffic Congress (ITC).*
[3] S. Delago-Segura, C. Pérez-Volà, J. Herrera-Joancomartí, G. Navarro-Arribas, and J. Borrel. 2018. Cryptocurrency Networks: A New P2P Paradigm. *Mobile Information Systems* 2159082 (March 2018), 1–16.
[4] B. Haslhofer, R. Karl, and E. Filtz. 2016. O Bitcoin Where Art Thou? Insight into Large-Scale Transaction Graphs. In *Proc. SEMANTICS.*
[5] S. Kim, Z. Ma, S. Murali, J. Mason, A. Miller, and M. Bailey. 2018. Measuring Ethereum's Peer-to-Peer Network. In *In Proc. ACM Internet Measurement Conference (IMC).*
[6] S. Meiklejohn, M. Pomarole, g. Jordan, K. Levchenko, D. McCoy, G. M. Voelker, and S. Savage. 2013. A Fistfull of Bitcoins: Characterizing Payments Among Men with no Names. In *Proc. ACM Internet Measurement Conference (IMC).*
[7] A. Miller, J. Litton, A. Pachulski, N. Gupta, D. Levin, N. Spring, and B. Bhattacherjee. 2015. *Discovering Bitcoin's Public Topology and Influential Nodes.* Technical Report. University of Maryland. See https://www.cs.umd.edu/projects/coinscope/coinscope.pdf.
[8] t. Neudecker, P. Andelfinger, and H. Hartenstein. 2016. Timing Analysis for Inferring the Topology of the Bitcoin Peer-to-Peer Network. In *Proc. IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress.*
[9] M. Ripeanu, I. Foster, and A. Iamnitchi. 2002. Mapping the Gnutella Network: Properties of Large-Scale Peer-to-Peer Systems and Implications for System Design. *IEEE Internet Computing* 6, 1 (January 2002), 50–57.
[10] D. Stutzbach and R. Rejaie. 2005. Capturing Accurate Snapshots of the Gnutella Network. In *Proc. IEEE Global Internet Symposium.*

[11] J. Tapsell, R. N. Akram, and K. Markantonakis. 2018. *An Evaluation of the Security of the Bitcoin Peer-to-Peer Network.* cs.CR 805.10259. ArXiV.

[12] Y. Vanaubel, J.-J. Pansiot, P. Mérindol, and B. Donnet. 2013. Network Fingerprinting: TTL-Based Router Signature. In *Proc. ACM Internet Measurement Conference (IMC).*