Detecting Bitcoin Nodes by the Cyberspace Search Engines

Ruiguang Li[1,2(✉)], Jiawei Zhu[2], Jiaqi Gao[3], Fudong Wu[3], Dawei Xu[1,3], and Liehuang Zhu[1]

[1] School of Cyberspace Science and Technology, Beijing Institute of Technology, Beijing, China
lrg@cert.org.cn
[2] National Computer Network Emergency Response Technical Team/Coordination Center of China, Beijing, China
[3] School of Cyber Security, Changchun University, Changchun, China

Abstract. Nowadays, the cyberspace search engines have showed great power to find entities and services in the network, which provide new ideas and methods to detect the Bitcoin nodes. This paper introduces the Bitcoin's P2P network and nodes including the reachable nodes and the unreachable nodes. Then, the results of detecting reachable nodes by the cyberspace search engines are showed. Next, the author proposes a new approach to find and verify the unreachable nodes by the cyberspace search engines. Finally, this paper illustrates the de-anonymization of some Bitcoin nodes by the cyberspace search engines, which map some node's IP addresses to real Bitcoin entities, such as Zeblockchain (a browser website), Microwallet (a wallet website) and Laurentia Pool (a non-profit pool website).

Keywords: Cyberspace search engines · Bitcoin nodes · Reachable nodes · Unreachable nodes · De-anonymization

1 Introduction

The cyberspace search engine is a new kind of network tool, which has attracted more and more attention from network researchers in recent years. It is different from the traditional Web search engine which takes Web pages as the retrieval objects, such as Google, Baidu and Bing. The Web search engine is widely crawling and storing web pages in the network, extracting and analyzing the pages' content, and providing keyword retrieval services for the public. The cyberspace search engine finds the entities and services in the network by actively detecting, obtains the target's information through protocol interaction and makes a comprehensive display. At present, the well-known cyberspace search engines in the industry include: Shodan (shodan.io, US), Censys (censys.io, US), BinaryEdge (www.binaryedge.io, EU), Zoomeye (www.zoomeye.org, CN), Fofa (offline now, CN), and so on.

The cyberspace search engines commonly maintain a protocol library which contains a variety of protocols, deploy probes all around the world, detect the whole network using various protocols, and find open ports and services all the time. There are many kinds of

W. Lu et al. (Eds.): CNCERT 2022, CCIS 1699, pp. 129–138, 2022.
https://doi.org/10.1007/978-981-19-8285-9_9

equipments in the network, including Servers, Network equipments, Terminals, Office facilities, Smart home devices, Industrial controlling equipments, Webcams, Blockchain entities, etc. [1] made a detailed comparative analysis of the well-known cyberspace search engines, compared their supporting protocols, detected equipments, equipment types, detecting capabilities, system structure, and probes, etc.

Bitcoin is the most successful electronic crypto-currency in the world. It was proposed by Nakamoto in 2008 [2] and launched officially in January 2009. Bitcoin kept running stably since then and had become an important means for global finance and payment. The cyberspace search engines had strong infrastructures which offer powerful computing, abundant storage, and a large amount of detecting records. This gave a great convenience for the analysis of assets and equipments in the cyberspace. The well-known cyberspace search engines include Shodan, Censys, Zoomeye, Fofa, etc, which provide detecting services for Bitcoin nodes. In this paper, we will introduce our work of finding and analyzing Bitcoin nodes by cyberspace search engines.

The contributions of this paper are as follows: 1) Introduce the results of detecting the Bitcoin reachable nodes by the cyberspace search engines. 2) Propose a new approach to find and verify the Bitcoin unreachable nodes by the cyberspace search engines. 3) Illustrate the de-anonymization of some Bitcoin nodes by the cyberspace search engines, which map some node's IP addresses to real Bitcoin entities.

2 Bitcoin Network and Nodes

Bitcoin system can be logically divided into the network layer and the transaction layer, as shown in Fig. 1.

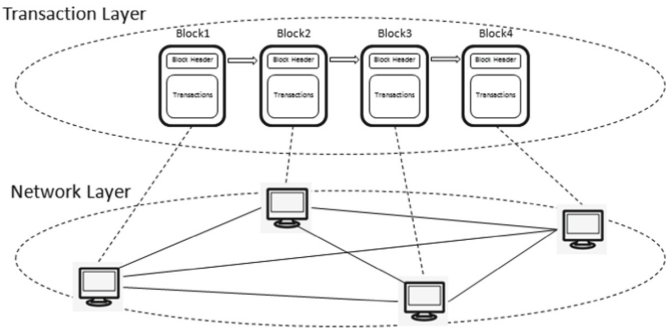

Fig. 1. Bitcoin's logic structure

The network layer is composed of a large number of Bitcoin nodes. Each node keeps working on broadcasting IP addresses, verifying transactions, packaging blocks, and mining independently. All the transactions are stored in the blocks, which connected each other by time order to form a blockchain. All the transactions are published to all network participants and stored in all nodes of the network. Previous studies mostly focused on the transaction layer, but less on the network layer. The Bitcoin network is a typical P2P

network without an organization or trust center. Nodes gain trust between each other through interactions, and form the network by themselves. The Bitcoin network is the foundation of Bitcoin system.

The Bitcoin network can be divided into the visible part (reachable nodes) and the invisible part (unreachable nodes). The reachable nodes can receive incoming connections and provide public services for the whole network. Generally, they would store a complete copy of the blockchain data. They open fixed ports (often 8333) waiting for connections and can be regarded as "Servers" in the network.

The unreachable nodes do not receive incoming connections from outside and don't provide public services for the whole network. They don't keep a complete copy of the blockchain data and can be regarded as "clients". The unreachable nodes are generally deployed behind a NAT (Network Address Translation) or a firewall and cannot be found by active detecting. Bitcoin nodes have different connectivity, service type, and topology, which have great impacts on the performance of Bitcoin system. Therefore, it is important to detect and study the Bitcoin nodes in depth.

3 Detecting the Reachable Nodes

Because the reachable nodes open fixed ports waiting for outside connections, we can easily detect all the reachable nodes by active detecting. There were many studies by far. Joan et al. [3]. Measured the Bitcoin network from November 2013 to January 2014, connected the Bitcoin nodes with bitcoin-sniffer (a open source tool) [4], collected 872000 nodes, and analyzed the nodes' geographical distribution, stability, propagation delay, etc. Christian Decker et al. [5] in 2013 and Giuseppe Pappalardo et al. [6] in 2016 measured the Bitcoin network, observed the propagation delay of blocks and transactions in the network. In the same year, Fadhil et al. measured the Bitcoin network for a week and collected 6430 stable online nodes and 313676 client IP addresses [7]. Sehyun Park et al. carried out a comparative study [8] in 2018, developed a software Bitcoin-Node-Scanner, obtained and verified 1 million nodes' IP addresses within 37 days, and counted the IP types (IPv6/IPv4/Onion), geographical distribution, port numbers, client versions, protocol versions, etc.

All these studies above were carried out by individual researchers. The cyberspace search engines have far more power than the single terminals. By scanning the whole network with probes all over the world, they can connect to all reachable nodes, get their information and make a time-based cumulative analysis. Fofa showed 56748 Bitcoin reachable nodes detected from December 2016 to September 2021 [9]. Zoomeye showed 63504 Bitcoin reachable nodes [10] and display their information such as IP, open ports, open services, countries, affiliated enterprises, protocol slogans, geographic longitude and latitude, as shown in Fig. 2 below.

It should be noted that [9] and [10] are only reachable nodes detected by the cyberspace search engines. Next, we will propose an approach to find and verify unreachable nodes by the cyberspace search engines.

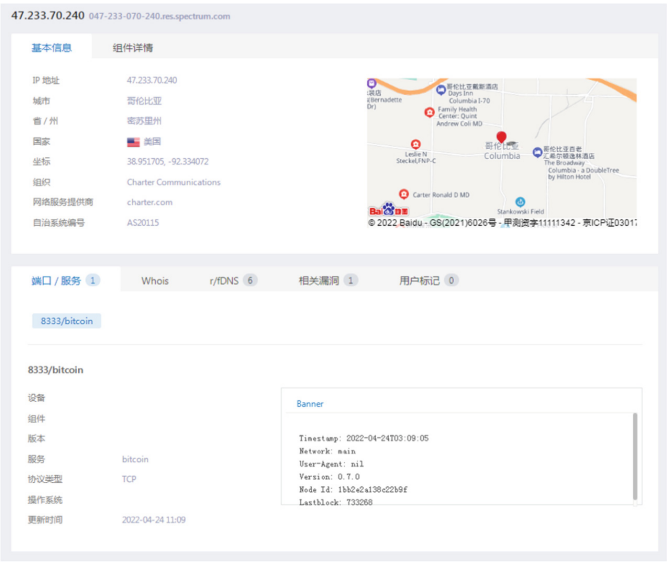

Fig. 2. Zoomeye's page of a reachable node

4 Inferring the Unreachable Nodes

The unreachable nodes don't open ports for outside connections, so they cannot be found by active detecting. Even if we got some addresses of the unreachable nodes, we could not definitely verify them by active detecting due to the existence of "Churn Nodes" which caused by the network delay.

There were a few studies on the unreachable nodes. Alex et al. proposed a de-anonymization method for the unreachable nodes [12] in 2014, which setup some probes connected to all entry nodes. When an unreachable node broadcasted a connection request through the entry nodes, the request would be forwarded to the probes and be recorded. The author believed that there were about 90000 unreachable nodes at that time. Till et al. simulated the Bitcoin network in 2016 and analyzed the broadcasting of transactions in the network [13]. It was estimated that the total number of was about 16000 then. Liang et al. deployed 102 probes around the world to collect the connection requests [14] in 2017, and estimated that there were 155000 unreachable nodes in the whole network. Matthias et al. monitored the "unsolicited" ADDR messages [15] in 2021 and could identify about 31000 active unreachable nodes every day. Federico et al. studied in detail the roles and number of unreachable nodes in Bitcoin network [16], and proposed an improved transactions broadcasting protocol, which improved the efficiency and security of the Bitcoin network. Alex et al. introduced the Bitcoin network based on Tor [17], proposed a man-in-the-middle attack against Bitcoin, and analyzed the delay caused by unreachable nodes in the attack. Indra et al. proposed a de-anonymization method for the unreachable nodes [18] by collecting all GETDATA messages and matching IP addresses with transactions/blocks. The accuracy of identifying the unreachable nodes is up to 90%.

Next, we will propose an approach to infer the unreachable nodes by the cyberspace search engines. First, we setup a fake client to actively connect to the reachable nodes, obtained a large number of Bitcoin nodes' IP addresses by the interaction mechanism of GETADDR-ADDR. Then, we input these addresses to the cyberspace search engine and obtain all the feedback records of the engines. Finally, by analyzing the open services and detecting time of the target IP, we can infer the unreachable nodes. Here we make two judgments.

Judgment 1: If an IP had a record of opening Bitcoin service with a new timestamp (within the duration of detecting cycle), this IP stood for a reachable node.

Judgment 2: If an IP had a record of opening other services (HTTP, SSL, etc.) except for Bitcoin service and the timestamp was relatively new(within the duration of detecting cycle), this IP stood for an unreachable node.

The correctness of Judgment 1 is obvious. The correctness of judgment 2 is also easy to understand. Because if we can verify a real IP from the Bitcoin system is opening other services, but isn't opening the Bitcoin service, the IP must stood for an unreachable node. The worldwide probes and all-weather scanning of the cyberspace search engines made sure that the "unreachable" of nodes were not caused by " network delay ". In fact, we have made experiments to testify Judgment 2 and the accuracy was up to 95%.

Fig. 3. Zoomeye's page of a unreachable node

Here is an example. As shown in Fig. 3, the node "167.172.158.149" is a real IP address obtained from the Bitcoin system. We input the IP into Zoomeye and check the feedback records. It can be seen that the IP opened "SSH" and "TCP" service, but didn't open "Bitcoin" service and the detecting timestamp was "2022–03-24". So the node "167.172.158.149" is an unreachable node.

To make a Ground-Truth test, we deployed a Bitcoin probe on Vultr. By checking the real neighbors using "peerinfo" command, we found the node "167.172.158.149" was its neighbor and the attribute is "inbound = true". The node made an incoming connection to our probe and was an real unreachable node.

5 De-anonymization of the Bitcoin Nodes

As an encrypted digital currency, Bitcoin protects the privacy and security of users' transactions. However, many researchers are very interested in the de-anonymization of Bitcoin addresses and tracing the route of transactions. By far, there are many studies on this issue being published. The existing methods are mainly based on the clustering of transaction addresses. For example, Butian Huang et al. proposed a clustering algorithm "BPC" [19] based on the nodes' behaviors, which clustered the nodes after behavior similarity measurement. The experiment showed that the accuracy was higher than the previous algorithms. Annika Baumann et al. analyzed the Bitcoin's transaction graph [20], inferred that there was a close relationship between network usage and exchange rate, and de-anonymized the 11 largest entities in the transaction graph. Meng Shen et al. analyzed the transaction propagation mode, proposed a method to obtain the initial transaction by calculating the pattern matching score [21], and established the association between the transaction and the initiating node's IP. The experimental accuracy was up to 81.3%.

In the de-anonymization of the Bitcoin nodes, it's important but difficult to find the association between a node's IP and the real network entity (exchages, browsers, wallets or pools), because many important entities keep their IP addresses highly confidential for the reason of privacy. **The cyberspace search engines provide new ideas for the association between Bitcoin nodes' IP and the real entities**. The cyberspace search engine detect the whole network using various protocols, and will find all services support by a node. For the reachable nodes, all the services such as HTTP, HTTPS, SSL, Bitcoin will be found together. As some persons or companies may open different services in One IP address, we could get extra information for a Bitcoin node by visiting its HTTP page. In some cases, we could get useful information such as the geographic location, organization information, and services operated by the website. Here we gave some examples.

1) A node with IP (147.135.252.43). This IP address is a Bitcoin browser "Ze-blockchain", belonging to Japan Digital Service Company, as shown in Fig. 4.

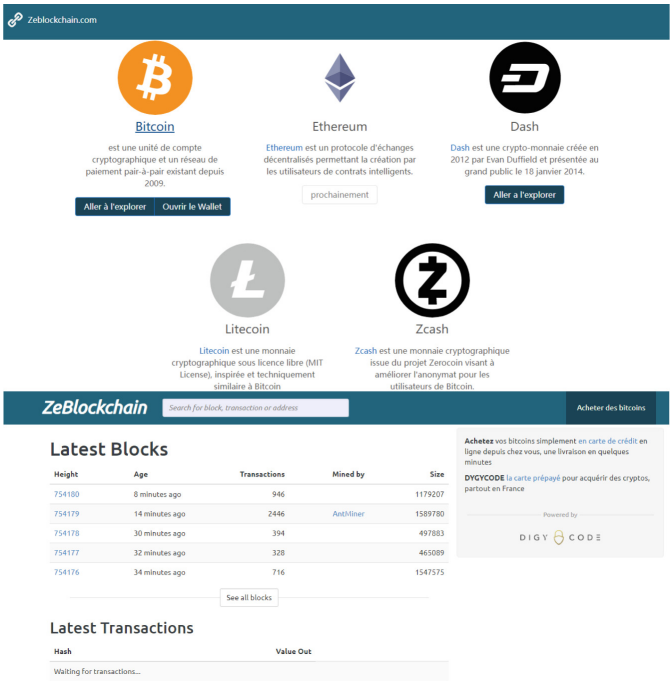

Fig. 4. Zeblockchain (a Bitcoin browser)

2) A node with IP (216.108.227.39): This IP address is a Bitcoin wallet "Microwallet", operated by a US company, as shown in Fig. 5.

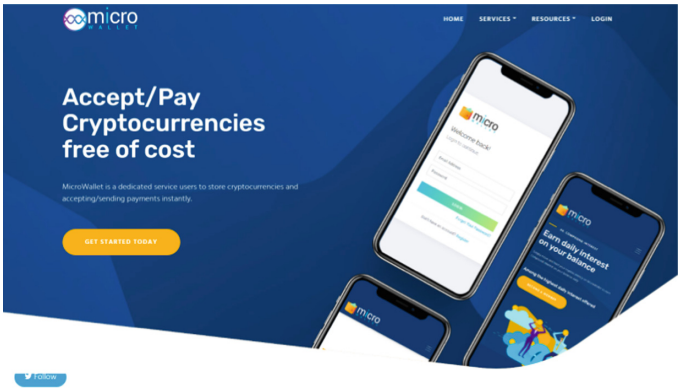

Fig. 5. Microwallet (a Bitcoin wallet)

3) A node with IP (51.81.56.49): This IP is a Bitcoin Pool "Laurentia pool", which is a non-profit mining pool(open source), as shown in Fig. 6.

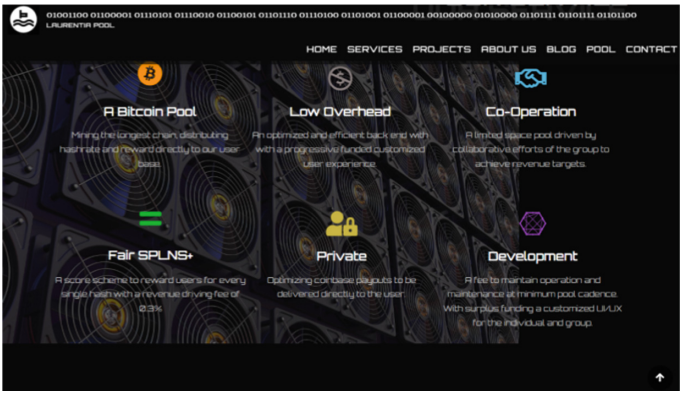

Fig. 6. Laurentia (a Bitcoin pool)

Limitations: This method could only de-anonymize some Bitcoin websites which open different services in One IP address. If large organizations have many IP addresses and don't deploy different services on same IP address, this method is no longer applicable.

6 Summary

This paper introduces the working principle of the cyberspace search engines and discusses their application in detecting the Bitcoin nodes. The Bitcoin network is composed of visible part (reachable nodes) and invisible part (unreachable nodes), which have different characteristics. The reachable nodes provide public services for the network and easy to detect, while the unreachable nodes are only clients and hidden in the network. The number of the unreachable nodes is about ten times to the reachable nodes [14], which are not easy to detect and analyze.

The author introduces the results of detecting Bitcoin nodes by the cyberspace search engines, then proposes a new approach to verify the Bitcoin unreachable nodes, finally illustrates the de-anonymization of the Bitcoin nodes which could find the association between a node's IP and the real network entity (a exchage, a browser, a wallet or a pool). By far, the cyberspace search engines can only detect Bitcoin nodes with Ipv4 addresses, and Ipv6 addresses are not supported. However, with the fast improvement of the cyberspace search engines, they will play more important roles in the detecting and analyzing of the Bitcoin network.

Acknowledgement. This work is supported by National Natural Science Foundation of China (Grant No. 62106060), and National Key Research and Development Program of China (Grant No.2020YFB1006105).

References

1. Li, R., Shen, M., Yu, H., et al.: A Survey on Cyberspace Search Engines. Springer, Singapore. Springer, Singapore (2020)
2. Nakamoto, S.: Bitcoin: A Peer-to-Peer Electronic Cash System (2009). http://www.bitcoin.org/bitcoin.pdf
3. Donet, J.A.D., Pérez-Sola, C., Herrera-Joancomartí, J.: The bitcoin P2P network. In: Proceedings of Financial Cryptography. Data Security FC Workshops, BITCOIN WAHC, vol. 16, pp. 87–102 (2014)
4. Sebicas.: Bitcoin p2p Network Sniffer (2013). http://github.com/sebicas/bitcoin-sniffer
5. Decker, C., Wattenhofer, R.: Information propagation in the bitcoin network. In: Proceedings of IEEE 13th International Conference Peer-Peer Computer (P2P), pp. 1_10 (2013)
6. Pappalardo, G., Caldarelli, G., Aste, T.: The Bitcoin Peers Network. http://blockchain.cs.ucl.ac.uk/wpcontent/uploads/2016/11/P2PFISY2016_paper_32.pdf. Accessed 30 Jun 2016
7. Fadhil, M., Owenson, G., Adda, M.: A bitcoin model for evaluation of clustering to improve propagation delay in bitcoin network. In: Proceedings of IEEE International Conference Computer Science Engineering (CSE), IEEE International Conference Embedded Ubiquitous Computing (EUC), 15th International Symposium Distribution Computer Appllication for Business Engineering (DCABES), pp. 468–475 (2016)
8. Park, S., Im, S., Seol, Y., Paek, J.: Nodes in the bitcoin network: comparative measurement study and survey. IEEE Access **7**, 57009–57022 (2019)
9. https://fofa.so/
10. https://www.zoomeye.org/
11. https://bitnodes.io/
12. Biryukov, A., Khovratovich, D., Pustogarov, I.: Deanonymisation of clients in bitcoin P2P network. In: Proceedings of ACM SIGSAC Conference Computer Communication Security, pp. 15–29 (2014)
13. Neudecker, T., Andelfinger, P., Hartenstein, H.: Timing analysis for inferring the topology of the bitcoin peer-to-peer network. In: 2016 International IEEE Conferences on Ubiquitous Intelligence Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People,Meets the and Smart World (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld), pp. 358–367 (2016) out
14. Liang, W., Pustogarov, I.: Towards Better Understanding of Bitcoin Unreachable Peers (2017)
15. Grundmann, M., Amberg, H., Hartenstein, H.: On the estimation of the number of unreachable peers in the bitcoin P2P network by observation of peer announcements (2021)
16. Franzoni, F., Daza, V.: Improving bitcoin transaction propagation by leveraging unreachable nodes. In: 2020 IEEE International Conference on Blockchain (Blockchain), pp. 196–203. IEEE (2020)
17. Biryukov, A., Pustogarov, I.: Bitcoin over Tor isn't a good idea. In: 2015 IEEE Symposium on Security and Privacy, pp. 122–134 (2015). https://doi.org/10.1109/SP.2015.15
18. Mastan, I.D., Paul, S.: A new approach to deanonymization of unreachable bitcoin nodes. In: Capkun, S., Chow, S. (eds.) Cryptology and Network Security. CANS 2017. Lecture Notes in Computer Science, vol. 11261. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-02641-7_13
19. Huang, B., Liu, Z., Chen, J., et al.: Behavior pattern clustering in blockchain networks. Multimedia Tools Appl. **76**(19), 20099–20110 (2017)

20. Baumann, A., Fabian, B., Lischke, M.: Exploring the bitcoin network. WEBIST (1), 2014, 369–374 (2014)
21. Shen, M., Duan, J., Shang, N., Zhu, L.: Transaction deanonymization in large-scale bitcoin systems via propagation pattern analysis. In: Yu, S., Mueller, P., Qian, J. (eds.) SPDE 2020. CCIS, vol. 1268, pp. 661–675. Springer, Singapore (2020). https://doi.org/10.1007/978-981-15-9129-7_45