Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 32 (2014) 1121 – 1126

Procedia
Computer Science

International Workshop on Secure Peer-to-Peer Intelligent Networks & Systems (SPINS-2014)

Analyzing the deployment of Bitcoin's P2P network under an AS-level perspective

Sebastian Feld*, Mirco Schönfeld, Martin Werner

Ludwig-Maximilians-University Munich, Oettingenstr. 67, 80538 Munich, Germany

Abstract

Bitcoin has become increasingly important in recent years. The exchange rate raised from $14 in January 2013 up to $240 in April 2013 and even $900 in early 2014. In this paper, we present novel insights about Bitcoin's peer-to-peer (P2P) network with a special focus on its distribution among distinct autonomous systems. We traversed Bitcoin's P2P network in a protocol-compliant manner and collected information about the network size, the number of clients, and the network distribution among autonomous systems. Our findings lead to conclusions about the resilience of the Bitcoin ecosystem, the unambiguousness of the blockchain in use, and the propagation and verification of transaction blocks.

Keywords: peer-to-peer, Bitcoin, enumeration, autonomous systems, electronic cash, privacy

1. Motivation

Bitcoin's popularity is increasing tremendously. Although completely virtual cash, there is a growing number of coffee shops and bars that accept Bitcoin as payment – at the time of writing in early 2014. In scientific literature, Bitcoin has been analyzed extensively in the past few years. There are extensions of the protocol like Zerocoin[1], privacy evaluations[2] and analyses regarding environmental[3] or legal aspects[4], for example. Our work is concerned with user's anonymity in the Bitcoin network and the robustness of the network itself. This paper is – to the best of our knowledge – the first that gives concrete numbers on size, structure and distribution of Bitcoin's core P2P network while highlighting aspects regarding autonomous systems (AS) at the same time. This is directly connected to an individual's anonymity since Bitcoin utilizes k-anonymity heavily. We are able to show that it is not only "one out of all Bitcoin peers" as typically assumed, but also "one out of the Bitcoin peers in a particular AS" which reduces the number of potentially concealing peers and increases privacy risks. This finding leads us to a statement about Bitcoin's robustness directly. The whole idea behind Bitcoin relies on fast information propagation throughout the network to prevent so-called "blockchain-forks". A recent attack-model to enforce such forks (as proposed by Decker

* Corresponding author. Tel.: +49-89-2180-9421 ; Fax: +49-89-2180-9148.
E-mail address: sebastian.feld@ifi.lmu.de

doi:10.1016/j.procs.2014.05.542

and Wattenhofer[5]) is based on assumptions about the network's size and topology. Our findings support this model and propose further investigations.

The remainder of this paper is organized as follows: Section 2 briefly explains the key aspects of Bitcoin necessary in the course of this paper and additionally sums up related work. Section 3 describes our methodology when analyzing Bitcoin's P2P network. In Section 4 we present and discuss our results. Finally, Section 5 concludes our work and gives some hints for future research.

2. Related Work

This section briefly demonstrates the functionality of Bitcoin and related work in the area of Bitcoin measurement.

2.1. Functionality of Bitcoin

Basically, Bitcoin is a completely decentralized electronic currency system based on a peer-to-peer network. Its history starts in November 2008, as Satoshi Nakomoto (probably a pseudonym) releases the article "Bitcoin: A peer-to-peer electronic cash system"[6]. Bitcoin is based on digital signatures in order to prove the possession of bitcoins (the currency) as well as on a publicly visible history of transactions together with cryptographic proof-of-works.

The remarkable property with Bitcoin is the absence of a central authority or issuer of currency. Instead, new bitcoins are issued constantly at a specific rate through so-called "mining". Also, the execution of transactions is monitored and reviewed by the P2P network. Therefore, the network's participants work towards a collective consensus regarding the transaction's validity and append it to the public history of already confirmed transactions (the so-called "blockchain"). Both, extending the blockchain and forming a consensus over valid transactions is achieved via a proof-of-work system[7,8]. Participants in Bitcoin's P2P network collect transactions into a block of data and perform hashing of this block modified with a nonce until a specifically structured key has been found. If the sought-after nonce is found, the P2P network verifies the result and finally each client appends the transactions to its own copy of the blockchain. Thus, the transaction has irreversibly taken place and, from now on, the longest blockchain represents the proof of happened events in a precise sequence.

The process of finding the proper nonce is Bitcoin's main security mechanism because it is assumed to be computationally expensive. In order to motivate the participants to join the verification and appending of transactions, there are two mechanisms: transactions fees and freshly "minted" bitcoins. Each correctly hashed block generates bitcoins owned by the miner.

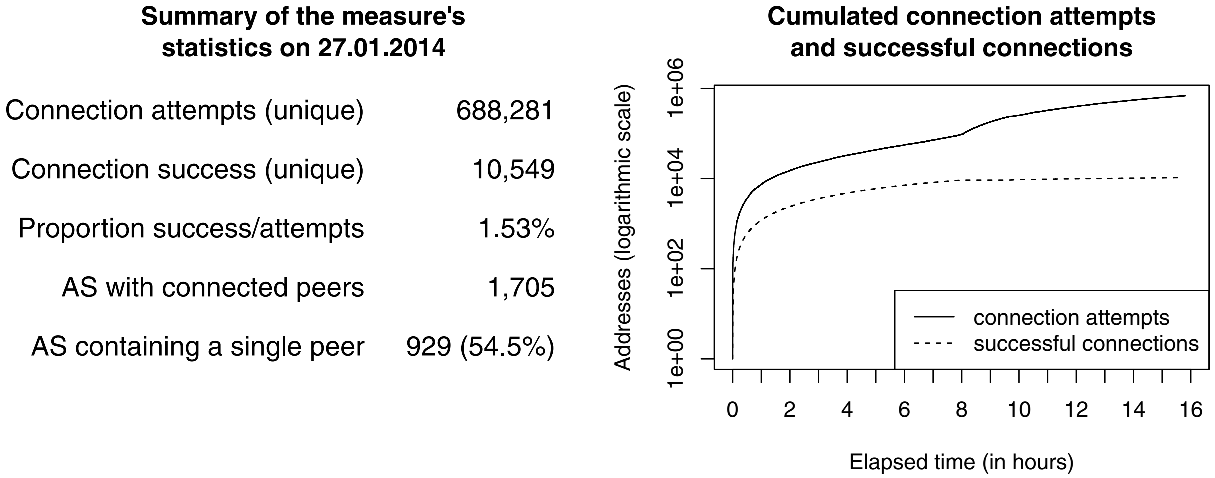
2.2. Bitcoin Measurement

There are several studies concerning Bitcoin that make use of its transaction history. Ron and Shamir[9] proposed extensive analysis of user's behavior, the flow of coins, user's balances and selected large transactions, for example.

Reid and Harrigan[2] provided the first comprehensive work regarding anonymity in Bitcoin. They thought of an attacker who, in order to de-anonymize Bitcoin's users, creates a 1-to-*n*-mapping between a user and the corresponding public keys. They performed a complete passive analysis as they constructed a transaction graph out of the publicly available transaction history. Using this transaction graph they created a user network via so-called linking and added further external information in order to analyze some bitcoin thieveries. In this work we want to discuss the possibility of "linking" transactions on an AS-level. We show that peers are not equally distributed over the underlying P2P network and thus exhibit the possibility of linking using AS-level information.

Ober et al.[10] showed that dynamic effects influence the anonymity with Bitcoin in both ways, positive and negative. We pick up their idea of a volatile privacy when highlighting the size and distribution of Bitcoin's P2P network: Ober et al. stated that anonymity in communication systems can be measured by a notion similar to *k*-anonymity[11]. We claim that this is too optimistic: A particular Bitcoin user is not "one out of all peers" but, for example, "one out of the peers inside his autonomous system".

At the Black Hat USA 2011 Conference, Dan Kaminsky presented an analysis of the Bitcoin system he performed[12]. He investigated, in particular, possible identity flaws at the TCP/IP layer. The main idea is to open connections to preferably all peers in Bitcoin's P2P network at once resulting in mapping IP addresses to transactions.



Summary of the measure's statistics on 27.01.2014

Connection attempts (unique)	688,281
Connection success (unique)	10,549
Proportion success/attempts	1.53%
AS with connected peers	1,705
AS containing a single peer	929 (54.5%)

Fig. 1. Left-hand: Summary of the measure's statistics on 27.01.2014. Right-hand: Cumulated number of connection attempts (solid line) and successful connections (dotted line) to unique Bitcoin peers.

There is the assumption that the first peer that broadcasts a transaction is the originator. After mapping a peer's IP address to a transaction, two transactions can possibly be linked if they happen in quick succession. We take this idea a step further and claim that two transactions that are further apart in time can possibly be linked when taking more information, e.g. on an AS-level, into account.

3. Methodology

To gather information about Bitcoin's peer-to-peer network, we crawled the network connecting to clients and collecting their knowledge of each other. Since we based our implementation on an API[13] that is publicly available, our tool appeared as a regular client to others. This enabled us to communicate with Bitcoin clients in a protocol-compliant manner. We requested a list of other peer's IP addresses (referred to as *peerlist*) from each client we contacted to. By repeating this enquiry on every new entry on the growing set of known peers, we traversed through Bitcoin's P2P network recursively. Actually, this is a well-known approach in current research concerned with P2P networks or Botnets[14,15].

However, analyzing such P2P networks (to enumerate clients, for example) by traversing through the network often provides unsatisfying results as Rossow et al. recently pointed out[16]. A lot of clients live behind a NAT or proxy making it impossible to contact them actively. Therefore, Rossow et al. discriminate between three types of clients: *routable* peers that can be contacted via an ingress connection, *non-routable* peers that can establish outgoing connections only because of an intermediate NAT or proxy, and *unreachable* peers that cannot be contacted by any other peer but are still known to one or more other peers. In the following, we will use this definition and concentrate on routable peers.

This assumption only covers a small fraction of Bitcoin clients, of course. But, we are not aimed at counting all existing clients. In fact, we tried to get an insight on how well Bitcoin's P2P network is interconnected under an AS-level perspective. Since publicly routable peers are the only ones contacted by non-routable peers, they constitute Bitcoin's backbone for information propagation such as forming a consensus over new transactions or broadcasting other known peers.

We traversed the network by collecting IP addresses from each peer and connecting to each of the returned addresses recursively. Of course, the returned peerlist also contained non-routable or non-reachable addresses resulting in a high number of unsuccessful connection attempts. Our framework is parameterizable to easily fine-tune its flow control, its maximum size of queue of clients to contact, or its duration of a connection to one single client. Each

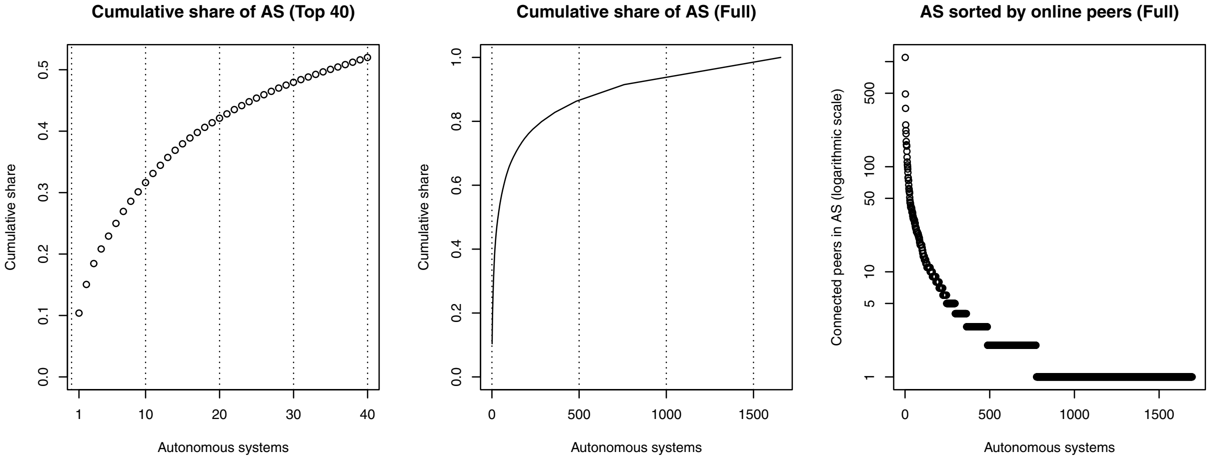


Fig. 2. Distribution of clients among autonomous systems (AS).

client's address together with its response of known peers was saved into a database for later analysis. For determining each client's autonomous system and its country we used Maxmind's free Geo API [1] and corresponding databases.

4. Results and Discussion

We performed several measurements that took differing time durations to complete. This is mainly caused by our framework's flow control parameters: We defined 60 seconds as the timeout for pending peers and 15 seconds as the duration a successful connection is maintained. The latter is a sufficient time for receiving a peerlist from the connected peer. All measurements were initiated by contacting a random Bitcoin peer.

The following subsections highlight some key results of a representative measure we started on 27.01.2014 on a server in AS6724 (Germany). That run took about 16 hours. We grouped our findings with regard to a) Bitcoin peers, b) autonomous systems and c) peerlists. The measure's statistics are summarized in Figure 1.

4.1. Bitcoin Peers

Our framework described in Section 3 tried to connect to as many peers as possible. The solid line in Figure 1 (logarithmic scale) shows the number of unique connection attempts over time. Since we tried to connect to every unique IP address found while traversing the number contains non-routable, offline as well as routable addresses circulating inside the P2P network. The dotted line in Figure 1 shows actual successful connections to unique Bitcoin peers. We were able to connect to 10,549 unique peers within 16 hours. Note that we have a certain error, since there is the possibility that we count a peer twice (e.g., a regular disconnect caused by the peer's ISP). Since clients connecting to Bitcoin's P2P network first perform outgoing connection attempts to one or more publicly available (routable) peers, the connected peers mentioned above could be seen as Bitcoin's backbone. The proportion of successful connection attempts is around 1.5%. This means, that about 98.5% of the IP addresses broadcasted in Bitcoin's P2P network either belong to non-routable or unreachable peers.

4.2. Autonomous Systems

IP addresses can easily be grouped, for example, by means of geographic regions or autonomous systems, whereat the latter has got a higher resolution.

[1] http://dev.maxmind.com/geoip/legacy/geolite/

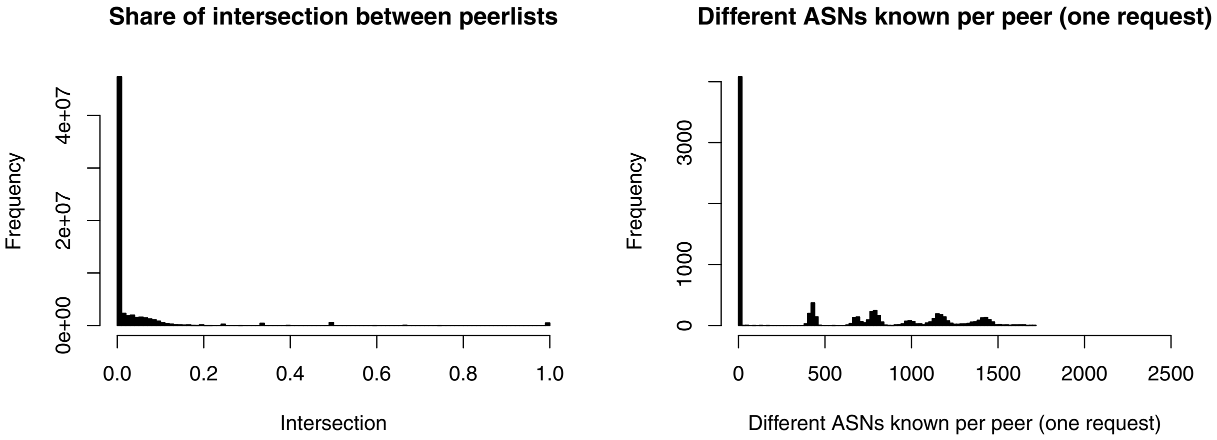


Fig. 3. Analysis of returned peerlists.

The two graphs on the left-hand side of Figure 2 show the cumulative distribution of routable clients among autonomous systems. The leftmost graph concentrates on the most populated AS clearly showing that over 30% of the peers we connected to reside in just ten AS. The graph also shows that half of the routable peers are distributed over less than 40 AS. Finally, the rightmost graph in Figure 2 depicts the complete quantity of the AS involved together with their corresponding number of online peers contained (logarithmic scale). Altogether, we were able to count 1,705 different AS. Note the remarkable long tail. In fact, in 95% of the AS involved our framework was just able to connect to 20 peers or less. The tail's end shows that there are 929 autonomous systems that contained just a single peer.

These numbers have an inpact on at least the vitality and resilience of Bitcoin's ecosystem. As mentioned before, the publicly available Bitcoin peers are some kind of Bitcoin's backbone since new peers wanting to connect to the P2P network perform a connection attempt to one of those peers. Thus, an outage or failure of a particular autonomous system can possibly have an impact on the reachability of Bitcoin's P2P network.

Also, as mentioned above, "linking" transactions is a serious threat to users' anonymity and privacy and is theoretically[2 12] and practically[17] possible. We take on this idea, but loose Dan Kaminsky's conditions from concrete IP addresses[12] to autonomous systems. Our findings clearly show that there are a lot of AS containing only one routable peer and, hence, those peers can be seen as exclusive within their particular AS. We now assume that each of these peers is the only one in the corresponding AS that runs a Bitcoin client. Thus, transactions that are further apart in time and even are broadcasted from different IP addresses can be linked nevertheless – as long as they emerge from an autonomous system containing probably just a single client. Furthermore, considering *k*-anonymity as a measure for anonymity and holding the number of unique peers inside a particular AS as variable *k*, a peer's anonymity within one of the 929 autonomous systems mentioned above possibly degrades to "1-out-of-1".

4.3. Peerlists

A client that joins Bitcoin's P2P network connects to a random peer via a bootstrap mechanism and afterwards requests a peerlist from its first peer. The reference implementation of a Bitcoin client answers with a peerlist containing at most 2,500 IP addresses. This list does not reveal the peer's internal order and contains both routable and non-routable addresses as well as non-reachable ones.

We compared all collected peerlists and focussed on their overlapping. The left-hand histogram in Figure 3 depicts a histogram of the sizes of the intersection of all peerlist pairs reported. It is obvious that the average intersection of two random peerlists is below 10%. This result is quite positive for Bitcoin's vitality and resilience, since a new peer gets to many IP addresses of possible peers very quickly. Besides that, a small intersection of peerlists is an indication of a well meshed P2P network since the knowledge of each other is a precondition for a future successful connection.

Another aspect we want to highlight is the distribution of the addresses contained in the peerlists regarding different autonomous systems. The right-hand histogram of Figure 3 depicts the average number of distinct autonomous systems a peerlist's addresses reside in. One can clearly see that only a small fraction of peers knows about peers in other autonomous systems.

Both histograms together reveal an interesting insight. It appears as if an average peer knows many different peers, indeed. But, most of those peers reside in the same autonomous system. This leaves Bitcoin's vitality, resilience and security flawed. Apparently, blockchain forks as proposed by Decker and Wattenhofer[5] and double-spendings as proposed by Karame et al.[18] can be considered a serious threat if an attacker takes information about autonomous systems into account.

5. Conclusion and Future Work

We introduced a framework that traverses Bitcoin's P2P network and generates statistics regarding its size and distribution among autonomous systems. We were able to show that there are more than 10,500 publicly available peers that constitute the network's core. These peers are distributed over more than 1,700 different AS. Although this appears as a balanced distribution, we found that only 10 AS contained more than 30% of all routable peers. Additionally, there were over 900 autonomous systems that contained just a single peer. Furthermore, an average peerlist contains addresses that mostly reside in the peer's own AS. Taking this information into account we claim that transaction linking could be possible.

As a next step one should consider deploying a (passive) sensor network into the Bitcoin P2P network. This would enable communication with non-routable peers and could gain additional insight.

Another interesting question is the distribution of mining pools on an AS level. Bitcoins are considered secure or stable as long as one attacker does not concentrate more than 50% of the computing power since mining can be seen as a competition based on distributed computing power. With that in mind, could the outage of a single AS possibly shift the distribution of power significantly?

References

1. Miers, I., Garman, C., Green, M., Rubin, A.D.. Zerocoin: Anonymous distributed e-cash from bitcoin. In: *Proceedings of the IEEE Symposium on Security and Privacy*. 2013, .
2. Reid, F., Harrigan, M.. An analysis of anonymity in the bitcoin system. In: *Security and Privacy in Social Networks*. 2013, .
3. Becker, J., Breuker, D., Heide, T., Holler, J., Rauer, H.P., Böhme, R.. Geld stinkt, Bitcoin auch - Eine Ökobilanz der Bitcoin Block-Chain. In: *GI-Jahrestagung*. 2012, .
4. Grinberg, R.. Bitcoin: An innovative alternative digital currency. *Hastings Sci & Tech LJ* 2012;.
5. Decker, C., Wattenhofer, R.. Information propagation in the bitcoin network. In: *Proocedings of the IEEE International Conference on Peer-to-Peer Computing (P2P)*. 2013, .
6. Nakamoto, S.. Bitcoin: A peer-to-peer electronic cash system. 2009. URL: `http://www.bitcoin.org/bitcoin.pdf`.
7. Dwork, C., Naor, M.. Pricing via processing or combatting junk mail. In: *Advances in Cryptology*. 1993, .
8. Back, A.. Hashcash - a denial of service counter-measure. 2002.
9. Ron, D., Shamir, A.. Quantitative analysis of the full bitcoin transaction graph. *IACR Cryptology ePrint Archive* 2012;.
10. Ober, M., Katzenbeisser, S., Hamacher, K.. Structure and anonymity of the bitcoin transaction graph. *Future Internet* 2013;.
11. Sweeney, L.. K-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 2002;.
12. Kaminsky, D.. Black ops of TCP/IP presentation. 2011. Black Hat, Chaos Communication Camp.
13. Hearn, M.. bitcoinj - a java implementation of a bitcoin client-only node. 2013. URL: `https://code.google.com/p/bitcoinj/`.
14. Holz, T., Steiner, M., Dahl, F., Biersack, E., Freiling, F.C.. Measurements and mitigation of peer-to-peer-based botnets: A case study on storm worm. *Proceedings of the USENIX Workshop on Large-Scale Exploits and Emergent Threats* 2008;.
15. Stock, B., Gobel, J., Engelberth, M., Freiling, F.C., Holz, T.. Walowdac-analysis of a peer-to-peer botnet. In: *Proceedings of the European Conference on Computer Network Defense (EC2ND)*. 2009, .
16. Rossow, C., Andriesse, D., Werner, T., Stone-Gross, B., Plohmann, D., Dietrich, C.J., et al. P2PWNED: modeling and evaluating the resilience of peer-to-peer botnets. In: *Proceedings of the 34th IEEE Symposium on Security and Privacy (S&P)*. 2013, .
17. Meiklejohn, S., Pomarole, M., Jordan, G., Levchenko, K., McCoy, D., Voelker, G.M., et al. A fistful of bitcoins: Characterizing payments among men with no names. In: *Proceedings of the Internet Measurement Conference*. 2013, .
18. Karame, G.O., Androulaki, E., Capkun, S.. Double-spending fast payments in bitcoin. In: *Proceedings of the 2012 ACM conference on Computer and Communications Security*. 2012, .