The Bitcoin P2P Network

Joan Antoni Donet Donet, Cristina Pérez-Solà(✉),
and Jordi Herrera-Joancomartí[1]

Departament d'Enginyeria de la Informació i les Comunicacions,
Universitat Autònoma de Barcelona, 08193 Bellaterra, Catalonia, Spain
{jdonet,cperez,jherrera}@deic.uab.cat

Abstract. The Bitcoin virtual currency is built on the top of a decentralized peer-to-peer (P2P) network used to propagate system information such as transactions or blockchain updates. In this paper, we have performed a data collection process identifying more than 872000 different Bitcoin nodes. This data allows us to present information on the size of the Bitcoin P2P network, the node geographic distribution, the network stability in terms of interrupted availability of nodes, as well as some data regarding the propagation time of the transmitted information. Furthermore, although not every Bitcoin user can be identified as a P2P network node, measurements of the P2P network can be considered as a lower bound for Bitcoin usage, and they provide interesting results on the adoption of such virtual currency.

1 Introduction

Bitcoin is an online virtual currency based on public key cryptography. It was proposed in 2008 in a paper authored by someone behind the Satoshi Nakamoto pseudonym. Bitcoin became fully functional on January 2009 and its broad adoption, together with its high exchange rates with traditional currencies (EUR or USD), has made it the most successful virtual currency ever. Security issues have been solved using elliptic curve public key cryptography together with the help of hash functions. The fact that hash functions are one-way functions provides a way to define an easily verifiable and fine-grained adjustable proof-of-work. Furthermore, double-spending, probably the core problem of digital currencies, is prevented by maintaining a public non-modifiable register, the blockchain, which includes all the transactions performed on the system.

Besides its security robustness, two main properties have probably been its key to success: anonymity and decentralization. Anonymity in the Bitcoin network is based on the fact that users can create any number of anonymous Bitcoin addresses that will be used in their Bitcoin transactions. This basic approach is a good starting point, but the underlying non-anonymous Internet infrastructure, together with the availability of all Bitcoin transactions, has proven to be an anonymity threat as different authors have pointed out [1,9,11–13]. The other key point of the system is its decentralized nature. No central authority is supposed to control the Bitcoin payment system and a distributed approach based on a peer-to-peer (P2P) network has been adopted.

R. Böhme et al. (Eds.): FC 2014 Workshops, LNCS 8438, pp. 87–102, 2014.
DOI: 10.1007/978-3-662-44774-1_7

To our best knowledge, at the present time no detailed information has been published about the P2P Bitcoin Network. Therefore, this paper represents the first attempt to collect and map such data in a comprehensive way. Collected data provides information on the size of the Bitcoin P2P network, the node geographic distribution, the network stability in terms of interrupted availability of nodes, as well as some data regarding the propagation time of the transmitted information. On the other hand, the data provided in this paper sheds some light about the real adoption and usage of the Bitcoin currency. This is a difficult measurement due to the distributed architecture of the system. Some previous attempts to estimate Bitcoin adoption rates were based on the number of existing Bitcoin addresses. However, these results provided an upper bound on the number of users since multiple addresses may be generated by a single user and an average rate of such value is not straight forward to obtain. The number of P2P Bitcoin nodes is, therefore, a better estimation, and can be taken as a lower bound for the number of Bitcoin users.

The rest of the paper is organized as follows. Section 2 gives some basic ideas about the decentralized nature of the Bitcoin system and reviews some prior works. Section 3 describes the data collection process. Then, Sect. 4 presents the data analysis: we provide information about the network size, the geographic node distribution, the node stability, and measurements about information propagation. Finally, Sect. 5 concludes the paper and provides some ideas for further research.

2 Bitcoin Basics

As we have already mentioned in Sect. 1, one of the interesting properties of Bitcoin is its decentralized nature. The Bitcoin architecture does not rely on a centralized server. Instead, a distributed approach has been adopted to support the system. The distributed approach is used in many of the system facets, the most important of which are: data storage, data confirmation, and data transmission. The core information of the Bitcoin system is stored in the so called blockchain. The blockchain is stored in every full-client node of the Bitcoin system, allowing them to validate new blocks and transactions. On the other hand, new transactions are confirmed by adding them to the blockchain through the mining process, a process that is also distributed and that can be performed by any user of the network using specific-purpose software (and hardware). Mining Bitcoins helps to confirm transactions and it has been designed to be a hard task. Mining uses the concept of proof-of-work in order to provide a significant level of security.

Finally, the Bitcoin system needs to disseminate different kinds of information, essentially, the payment transactions performed by users and the blockchain (or its actualization). Since both data are generated in a distributed way, the system transmits such information over the Internet through a distributed peer-to-peer (P2P) network. This distributed network is created by Bitcoin users in a dynamic way. Nodes of the Bitcoin P2P network are machines running Bitcoin node software. This software is included by default in Bitcoin's full-client wallets, but it is not usually incorporated in light wallet versions (such as those running in mobile devices). It is important to stress this distinction, because when discovering nodes of the P2P network we do not identify all Bitcoin users, but only those running a full-client. Furthermore, the online Bitcoin accounts provided by major Bitcoin Internet sites are also not detected as independent Bitcoin nodes.

2.1 Related Work

In contrast to other virtual payments systems that have appeared so far, the seminal paper [10] describing the Bitcoin system was not published in the scientific arena but as an Internet post. Furthermore, the practical development of the ideas proposed in such paper took place on January 2009, when the first block of the blockchain appeared together with a fully functional Bitcoin wallet. For this reason, the deployment of Bitcoin has taken off without so much attention from the research community and, until now, not so many research papers have been published analyzing its particularities and properties.

Besides its legal and economic aspects, the majority of Bitcoin research papers are focused on analyzing the anonymity of Bitcoins [1,9,11–13]. They do so by exploiting the opportunity that represents the availability of all system transactions in the publicly accessible blockchain. Other few papers deal with security issues [4,8] or improvements on the payments processing time [3].

Regarding the characteristics of the P2P Bitcoin network, there are two papers related to this topic. In [2] the authors analyze the well known Sybil attack, where users of the P2P network are able to create various identities to perform different attacks and reduce, for instance, the P2P network performance. However, their approach is a theoretical one, and no real information is provided on the P2P Bitcoin network. Decker and Wattenhofer perform in [6] an interesting study on how information is disseminated in the Bitcoin network and how a network synchronization problem may affect the payment system in terms of blockchain uniqueness. In that paper, some measurements on propagation delays are provided but the results are based on a set of approximately 16000 nodes, in contrast with our 872648 node dataset.

3 Data Collection

In this Section we review the data collected to perform our analysis. We explain the procedure used to gather the information together with some numbers describing the amount of data collected. Finally, we review the limitations of both the collected data and the analysis done on its basis.

3.1 Data Collection Procedure

In order to collect data from the Bitcoin P2P network we developed an application, BTCdoNET[1], which serves, on one hand, as a frontend to interact with several utilities and, on the other hand, to store the collected data.

[1] The name of the application is a pun with the first author's name, who was the developer of the application.

With respect to the interaction with other applications, BTCdoNET is used as an interface to a modified Bitcoin P2P Network Sniffer [5] instance. Bitcoin Sniffer is a Python script that is able to connect to a Bitcoin node and listen to network events such as block and transaction broadcasts. We have modified the original Bitcoin Sniffer program in order to be able to listen to many nodes of the network at the same time, and to store all the collected data in a MySQL database. BTCdoNET also makes use of pynode, which is a dependency of the Bitcoin P2P Network Sniffer; and a classic LAMP installation, with a MySQL database storing all the collected data.

With respect to the data collection functionalities, BTCdoNET gathers essentially two different kinds of data:

1. **Network topology information**
   By issuing a `getaddr()` command to a set of seeds, we obtain a list of nodes that are connected to every seed. Then, by recursively applying the same procedure to the nodes connected to the seeds, that is, by sending `getaddr()` commands to the seeds' neighbors, we discover the neighbors of the neighbors, and so on. We maintain a list of already pooled nodes, so that one node is not queried twice. The process ends when there are no new nodes pending to be queried. Following the stated procedure, we perform a Breadth First Search of the Bitcoin P2P network. With this procedure we obtain, on one hand, a view over the Bitcoin P2P network structure itself and, on the other hand, a list of IPs addresses knowing to be running a Bitcoin node.
2. **Propagation of information in the network**
   The application is also able to connect to a set of already discovered nodes and to start monitoring their activity, that is, to listen to the transactions and blocks that the node is propagating to its neighbors. Apart from storing the transaction or block identifiers, BTCdoNET records the exact moment when the transaction or the block was broadcast by each of the nodes. This allows us to analyze how the information (transactions and blocks) is propagated through the network.

3.2 Collected Data

With respect to topology information, we performed 1 scan every day at 9 PM CET from November 30th, 2013 to January 5th, 2014. We will use the term *network snapshot* to refer to each of the 37 scans. Each snapshot took around 2 h to complete. The network discovery procedure used a fixed set of 600 nodes as seeds. After these 37 days of network discovering, we have detected 872648 different IP addresses corresponding to machines running Bitcoin nodes. Note, however, that only with the information of the first snapshot we already discovered 111475 nodes. This points out that there is a lot of node overlap between different snapshots and can be used as an indicator of the stability of the network. Section 4.2 analyzes node stability in a deeper way.

Concerning the propagation of information in the network, we configured the sniffer to try to simultaneously connect to 2000 different Bitcoin clients.

We selected those clients from the set of more stable nodes obtained with the network topology discover procedure. From these 2000 nodes, only 1377 accepted the connection request. We then listened to all 1377 nodes during 26 h, storing information about the exact moment when each of the nodes sent us transactions and blocks. After this period of time, we stop listening to information about transactions, but keep monitoring the block propagation information for an additional 92 h. The rationale behind this decision was to obtain a significant amount of block information without being flooded by the transaction propagation information. Over those periods of time, we received 13910769 transactions from the different nodes, representing a total of 70254 unique transactions. Regarding block information, we received 492793 block copies, getting information from 11663 different unique blocks.

3.3 Limitations

Although the amount of data collected is huge, both the dynamic nature of the P2P network and the data collection methodology introduce some limitations.

- Limitations of network topology information:
  - The number of nodes discovered is huge, but it does not represent the entire network. On one hand, some nodes do not respond to `getaddr` messages, so no information about their neighborhood can be extracted from them. On the other hand, the standard implementation of the Bitcoin client does not return all the node's neighbors in response to a `getaddr` call, but just the minimum between 23 % of the active nodes and a constant, which is set to 2500. These also limits the amount of information obtained when exploring the network through `getaddr` messages.
  - The paper is focused on analyzing the Bitcoin P2P network and thus we are dealing with Bitcoin nodes. Note that working with Bitcoin nodes is very different from working with Bitcoin users. It is important to stress such distinction, because the usage of light-clients as well as online Bitcoin accounts is very extended, and thus an important part of Bitcoin users can not be identified as Bitcoin nodes.
  - We identify Bitcoin nodes by their IP addresses. Although servers usually have static IP addresses, some of the Bitcoin nodes may be running on machines with dynamic IP addresses. Therefore, nodes may appear to be more unstable than they really are.
  - Each of the scans took about 2 h to complete. Therefore, some parts of the network may have changed while we were exploring other parts. However, we consider all the information in each of the snapshots as belonging to the very same instant of time.
  - We rely on geopositioning services to locate the IP addresses, which may introduce small errors when drawing their location over a map or classifying them by countries.
- Limitations about propagation information:

- When studying data propagation through the network, we simultaneously listened to around 1300 nodes. This number of nodes is far away from the total number of nodes of the network, and thus our computations can only be seen as an approximation of the values the whole network would exhibit.

4 Data Analysis

In this section we present the analysis of the collected data. We provide general information on the size of the peer-to-peer network, its geographical distribution, and the stability of the nodes. Finally, we study how transaction and block data propagate through the P2P network.

4.1 Network Size and Geographic Distribution

The Bitcoin network is global and, as such, we can find Bitcoin nodes operating all over the world. Table 1 shows the number of Bitcoin nodes discovered by country. The Table lists the 25 countries showing the highest number of Bitcoin nodes on the first day snapshot of the network, together with the 8 countries showing the least number of nodes. The country of a node is estimated from its IP address, using an IP geolocating service [7]. The Table presents the number of nodes by country analyzing all the collected data (2nd column) and for the data collected on the first day, which corresponds to the first full snapshot of the network (3rd column). Due to node overlap between different snapshots, the rankings may vary depending on the specific criteria used. Section 4.2 analyzes this fact in more depth.

We can observe that nodes placed in Unites States and China sum up to 37 % of the discovered nodes. Germany, United Kingdom, and Russia concentrate also a big amount of nodes of the network, with 9 %, 4 %, and 7 %, respectively, of the overall detected nodes. At the bottom of the table we can see that there are 8 countries with just one node detected on at least one of the snapshots. Grouped into the *others* category, there are as much as 136619 nodes (15483 on the first snapshot) coming from other 180 countries.

It is also interesting to study the Bitcoin adoption rate in each of the different countries. We have tried to evaluate this rate by comparing the number of Bitcoin nodes found in each country with the number of Internet users on that very same country.[2] Countries like Japan, Brazil, Mexico, and China present really low adoption rates, with the number of Bitcoin nodes being less than 3 per every 100000 Internet Users. On the contrary the Netherlands, Norway, Finland, and the Czech Republic have the highest adoption rates, more than 10 times higher than those showed by Brazil.

[2] However, as we explain in Sect. 3.3, the number of Bitcoin nodes does not map directly with the number of Bitcoin users, so the adoption rates have to be interpreted accordingly.

Table 1. Discovered nodes by country of origin

Country	# of Bitcoin nodes (37 days)	# of Bitcoin nodes (1st day)	# of Internet users [15]	Bitcoin node rate (per 100.000)
United States	145.495	24.621	254.295.536	9.68
China	172.662	16.700	568.192.066	2.94
Germany	80.067	7.695	68.296.919	11.27
United Kingdom	43.369	6.849	54.861.245	12.48
Russian Federation	66.705	6.848	75.926.004	9.02
Canada	23.308	4.664	29.760.764	15.67
Netherlands	16.490	4.070	15.559.488	26.16
France	17.249	2.752	54.473.474	5.05
Australia	15.239	2.364	18.129.727	13.04
Poland	19.242	2.265	24.969.935	9.07
Spain	14.303	1.726	33.870.948	5.10
Ukraine	13.606	1.688	15.115.820	11.17
Italy	17.098	1.572	35.531.527	4.42
Brazil	16.452	1.476	99.357.737	1.49
Czech Republic	6.019	1.403	76.32.975	18.38
Taiwan	16.335	1.375	17.656.414	7.79
Sweden	7.958	1.366	8.557.561	15.96
Norway	4.036	1.016	4.471.907	22.72
Switzerland	5.463	933	6.752.540	13.82
Finland	4.692	901	4.789.266	18.81
Japan	6.631	843	100.684.474	0.84
Austria	7.012	828	6.657.992	12.44
Belgium	5.810	726	8.559.449	8.48
Argentina	5.863	663	23.543.412	2.82
Hong Kong	4.917	648	5.207.762	12.44
...	...	...	...	...
Anguilla	1	0	9.133	0.00
Burundi	1	0	128.799	0.00
Cape Verde	1	0	181.905	0.00
Dominica	1	0	40.349	0.00
Equatorial Guinea	1	0	162.202	0.00
Samoa	1	0	25.111	0.00
Sao Tome & Principe	1	0	39.515	0.00
Timor-Leste	1	0	10.461	0.00
Others (180 countries)	136619	15483	-	-
Total	872648	111475	-	-

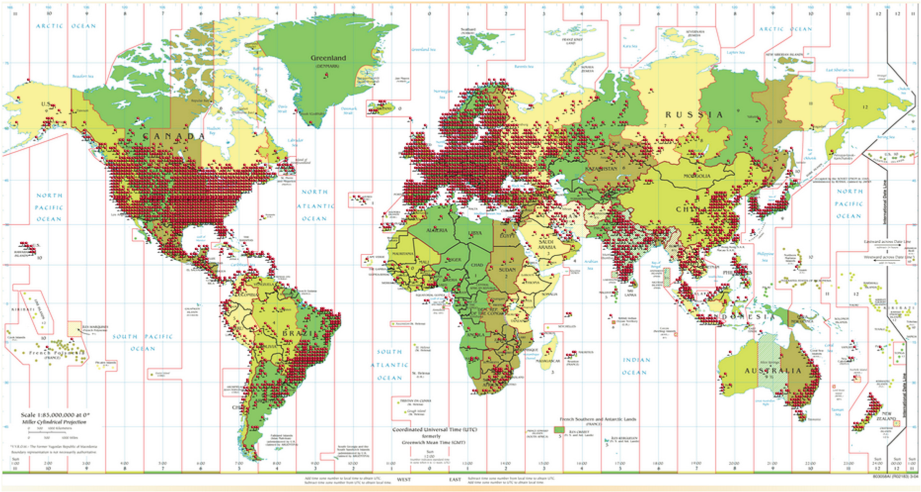


Fig. 1. Geolocation of discovered nodes

We have also used the IP geolocation service to plot the origin of the discovered nodes over a map. Figure 1 shows a map with the estimated location of all discovered nodes. Interesting information can be extracted from the map: there are Bitcoin nodes all over the world, with very low populated areas and underdeveloped countries being almost the only exceptions; western Europe and US distribution of nodes is quite uniform, with some peaks located over the most populated areas. Moreover, the map also demonstrates that the sample we have collected is broad, that is, it is not limited to a specific part of the Bitcoin network.

4.2 Node Stability

The map offered information about the location of nodes and, in a rough sense, their amount. We have also started to study the behavior of the Bitcoin nodes in terms of stability, that is, given a node, we analyze if such P2P node is available during all the 37 days of network observation. Figure 2 provides such information, showing the number of nodes still available after successive days of data collection. Notice that most of them are not connected more than the first five consecutive days and, at the end of the period, only 5769 nodes remain (which represents only a 0.66 % of the discovered ones). These 5769 nodes were permanently connected during all 37 days.

4.3 Information Propagation Analysis

In this section, we present the results of the information propagation analysis. Using the modified Bitcoin P2P Network Sniffer, we listened to various nodes of

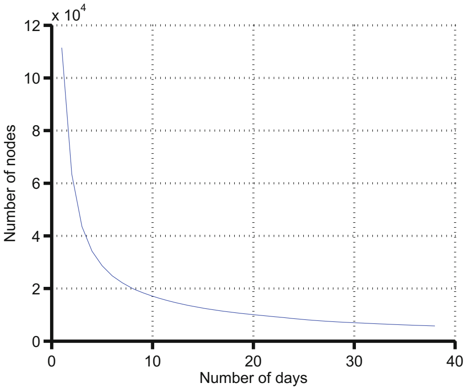


Fig. 2. Number of nodes in the intersection of the snapshots (for the 37 days)

the Bitcoin network, storing the transactions and blocks being broadcast through the network together with a timestamp signaling the exact moment when each of the nodes sent the information.

Block Propagation. Block propagation data consists on 492793 block copies representing 11663 different unique blocks. This data was captured listening to 1377 nodes during a period of 118 h (around 5 days). The data as captured is, however, very noisy. Note that if we take into account that the theoretical block production rate is 6 blocks per hour, the total number of blocks produced during this period of time should be around 708, a number significantly different from the mentioned 11663 blocks. The reason is that we receive copies of some very old blocks. For instance, even when the propagation information was captured on January 2014, we received a block whose timestamp dated from May 31th, 2013. In order to filter all this noise, we focus the block propagation analysis on the blocks created during the sniffer listening time. When adding this restriction, we obtain 737 different blocks to work on, a number much closer to the theoretical 708.

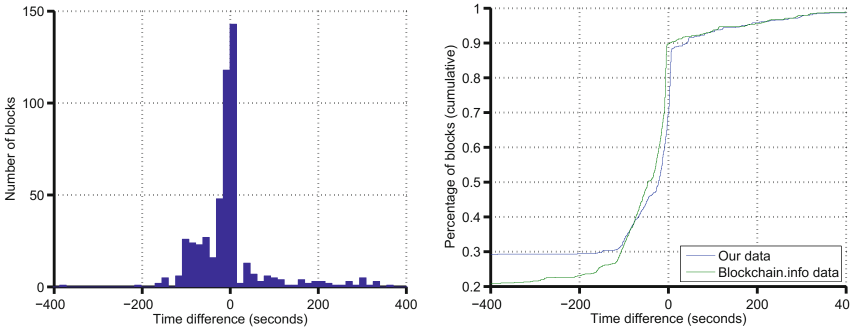
Bitcoin blocks contain a specific field in their headers with the current timestamp. This field is filled by the miner who finish the proof-of-work by solving the cryptographical challenge needed to find the block. Since the network accepts a block as valid even if the timestamp does not exactly match the network time (block timestamp is considered valid if it is not set more than two hours in the future) [14], the miner has some degree of freedom when setting the block's timestamp.

Once a miner has found a block, the miner announces it to the network by sending `inv` messages with the block to all of their peers, who do the very same thing if they consider the block valid, and thus propagate the block through the network.

Let us denote by $t_{\text{stamp}}(B_i)$ the timestamp contained in the header of the block B_i. Given a passive node (*i.e.*, a sniffer) with n peers, we define the registration time $t_j^{\text{reg}}(B_i)$ as the moment when the sniffer receives the block B_i from peer j, with $j = 1, \cdots, n$. Then, the first time a block B_i is seen by the passive node is $t_{\min}^{\text{reg}}(B_i)$:

$$t_{\min}^{\text{reg}}(B_i) = \min_{\forall j}\{t_j^{\text{reg}}(B_i)\}$$

Since the miners can set the timestamp of the block header, $t_{\text{stamp}}(B_i)$, we analyzed the differences between the aforementioned timestamp and the first time we receive a block, $t_{\min}(B_i)$. We were specially interested in detecting, on one hand, if the network is synchronized and, on the other hand, if miners were blatantly adjusting block timestamps.



(a) Distribution of $t_{\text{stamp}}(B_i) - t_{\min}(B_i)$ (b) Distribution of $t_{\text{stamp}}(B_i) - t_{\min}(B_i)$ (cumulative)

Fig. 3. Distribution of $t_{\text{stamp}}(B_i) - t_{\min}(B_i)$

Figure 3 shows the distribution of $t_{\text{stamp}}(B_i) - t_{\min}^{\text{reg}}(B_i)$ for the collected blocks, with the blue line representing the data we collected. Most of the times the difference is around 0. This is what is expected for a synchronized network with low propagation delays and where all peers well-behave. Note that more than 80 % of samples are negative, meaning that we receive the block after it is allegedly created. Positive samples illustrate that we receive a block before its header's timestamp, which indicates an altered block timestamp, either because the network time of the miner is notably different from ours, or either because the miner is intentionally modifying the block timestamp. There are around 10 % of blocks showing a positive difference less than 50 s, and another 10 % of blocks showing higher positive differences.

For the sake of comparing the data we collected with other external data, we also used the blockchain info API[3] to query for their reception time of each of the blocks. The results are presented in green in Fig. 3, where it can be seen that they are quite similar to ours. When the time difference is higher than −80 s, we

[3] Blockchain.info is a web page that offers information about Bitcoin blocks and transactions. They have a public API that allows to query for specific information. We used the API to obtain their *received_time* for each of the blocks.

receive the blocks a little faster than blockchain.info. However, when the time difference is lower than −80 s, their time difference is much lower. This may be a consequence of our shorter listening time, that make us receive copies of old blocks. Regarding the highest positive time difference, it is 7212 for our data and 7202 for blockchain.info data.

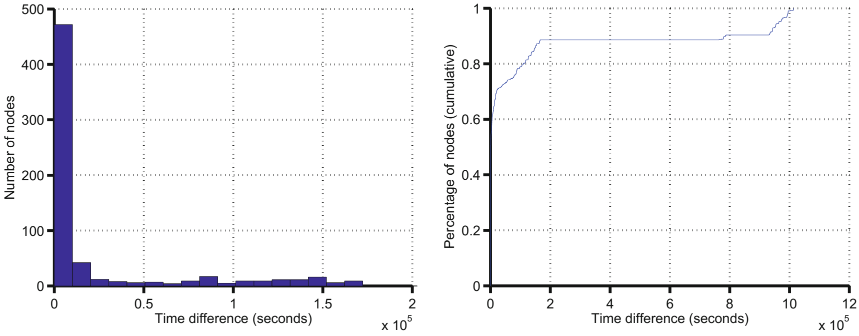
In a similar way than with the minimum registration time, we can define the last time the sniffer receives a block, $t^{\text{reg}}_{\max}(B_i)$, as:

$$t^{\text{reg}}_{\max}(B_i) = \max_{\forall j}\{t^{\text{reg}}_j(B_i)\}$$

We can then define the observable propagation delay for block B_i as:

$$\Delta(B_i) = t^{\text{reg}}_{\max}(B_i) - t^{\text{reg}}_{\min}(B_i)$$

Figure 4 shows the observable propagation delay for blocks. One can appreciate that 50 % of the blocks are propagated in less than 17 min, but the rest of the nodes take a huge amount of time to get to all listened nodes. However, note that we are using the last time we receive a block to do these computations, so if only one node sends us a copy of a block with high delay, it is enough to set that block's $\Delta(B_i)$ to a huge number. The best propagation time was as low as 52 s.



(a) Distribution of $\Delta(B_i)$ (b) Distribution of $\Delta(B_i)$ (cumulative)

Fig. 4. Distribution of $\Delta(B_i)$ for blocks

In order to try to better understand how the information is propagated through the network, we define the vector $T^{\text{reg}}(B_i)$ as the vector containing all the registration times for a block, $t^{\text{reg}}_j(B_i)$, in an increasing order (from the earliest to the latest):

$$T^{\text{reg}}(B_i) = [T^{\text{reg}}_1(B_i), \cdots, T^{\text{reg}}_n(B_i)]$$

with

$$T^{\text{reg}}_k(B_i) = t^{\text{reg}}_j(B_i),\ \forall k \in [1, n]$$

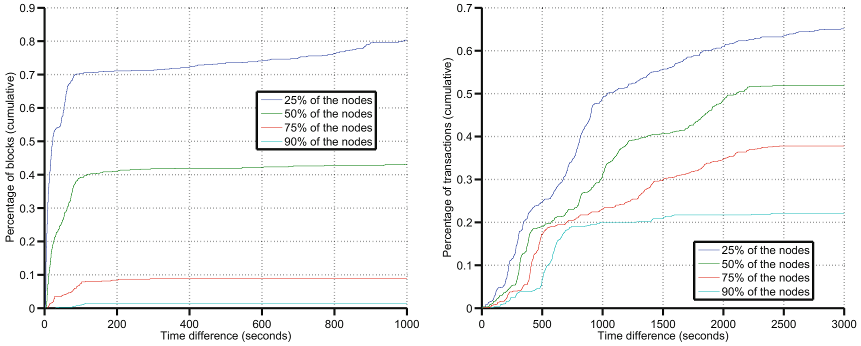
for some peer j such that

$$T^{\text{reg}}_{k-1}(B_i) \leq T^{\text{reg}}_{k}(B_i) \leq T^{\text{reg}}_{k+1}(B_i)$$

Then, we can study how information is propagated through the network by analyzing how much time is needed to get to 25 %, 50 %, 75 %, and 90 % of the nodes we were listening. Each of the percentages corresponds to a different position of the above described vector, specifically:

$$\begin{aligned}
\Delta^{25\,\%}(B_i) &= T^{\text{reg}}_{278}(B_i) - t^{\text{reg}}_{\text{min}}(B_i) = T^{\text{reg}}_{278}(B_i) - T^{\text{reg}}_{1}(B_i)\\
\Delta^{50\,\%}(B_i) &= T^{\text{reg}}_{557}(B_i) - t^{\text{reg}}_{\text{min}}(B_i) = T^{\text{reg}}_{557}(B_i) - T^{\text{reg}}_{1}(B_i)\\
\Delta^{75\,\%}(B_i) &= T^{\text{reg}}_{836}(B_i) - t^{\text{reg}}_{\text{min}}(B_i) = T^{\text{reg}}_{836}(B_i) - T^{\text{reg}}_{1}(B_i)\\
\Delta^{90\,\%}(B_i) &= T^{\text{reg}}_{1003}(B_i) - t^{\text{reg}}_{\text{min}}(B_i) = T^{\text{reg}}_{1003}(B_i) - T^{\text{reg}}_{1}(B_i)
\end{aligned}$$

Figure 5(a) shows the time needed for the blocks to propagate to a specific percentage of the listened nodes (25 %, 50 %, 75 %, and 90 %). We can appreciate that for 70 % of the blocks it takes less than 84 s to reach 25 % of the nodes. However, just 38 % of the blocks get to 50 % of the nodes in that very same time, 6 % of the blocks get to 75 % of the nodes, and less than 1 % of the blocks get to 90 % of the nodes. Note that, for some blocks, we do not receive their copies from every node that we are connected to. This may happen because the node disconnects during our listening period. We consider the registration time of a block B_i from peer j to be infinite if we do not receive the block B_i from peer j. Therefore, the graph shows an upper bound over the propagation times.



(a) Distribution of $\Delta^{x\%}(B_i)$

(b) Distribution of $\Delta^{x\%}(T_i)$

Fig. 5. Distribution of $\Delta^{25\,\%}$, $\Delta^{50\,\%}$, $\Delta^{75\,\%}$, and $\Delta^{90\,\%}$ for blocks and transactions

Finally, we studied if there is any correlation between the size of the block and the time needed to propagate the block through the network. In order to do so, we compute different correlation metrics between the size of the block, in bytes, and the time needed to propagate the block to the 25 % of the nodes of

Table 2. Number of transaction and blocks first received by each node

Node id	# of blocks	% of blocks	Node id	# of transactions	% of transactions
1	80	10.85 %	125	20695	29.46 %
2	63	8.55 %	126	7990	11.37 %
3	47	6.38 %	7	5815	8.28 %
4	42	5.70 %	10	3075	4.38 %
5	36	4.88 %	3	2285	3.25 %
6	35	4.75 %	11	1688	2.40 %
7	34	4.61 %	23	1521	2.17 %
8	28	3.80 %	12	1443	2.05 %
9	21	2.85 %	9	1138	1.62 %
10	21	2.85 %	19	964	1.37 %
11	18	2.44 %	35	818	1.16 %
12	15	2.04 %	127	655	0.93 %
13	15	2.04 %	128	602	0.86 %
14	14	1.90 %	129	564	0.80 %
15	11	1.49 %	103	560	0.80 %
16	10	1.36 %	130	530	0.75 %
17	9	1.22 %	131	475	0.68 %
18	8	1.09 %	132	436	0.62 %
19	7	0.95 %	83	431	0.61 %
20	7	0.95 %	133	413	0.59 %
Total (sum of the 20 best ranked IPs)	521	70.69 %		52098	74.16 %
Total (overall collected data)	737	100 %	-	70254	100 %

the network. The obtained Pearson correlation coefficient is 0.0172, which is a positive but low value, thus indicating that there is no strong linear correlation between the two variables. However, rank correlation coefficients, that capture the degree of similarity between the rankings of the two variables, present much higher values. The Kendall's tau correlation coefficient for these same variables is 0.3617, and the Spearman's rho coefficient is 0.4409. This indicates that there exists a correlation between the two variables, size and propagation, but that this correlation is not linear.

Transaction Propagation. In a similar way than with blocks and using the same notation, we also analyzed the propagation time of transactions over the P2P network. Transactions are broadcast through the network in a similar way than blocks, although there exists some differences on the client behavior for the two structures.

Figure 5(b) shows the time needed for the transactions to propagate to a specific percentage of the listened nodes (25 %, 50 %, 75 %, and 90 %). Transaction relaying seems to be slower than block propagation. While 50 % of blocks were broadcast to 25 % of the nodes in less than 22 s, 17 min are needed to relay 50 % of the transactions to the 25 % of the nodes. Apart from this scaling factor, blocks and transactions are propagated in a similar way, with most of them being quite fast to get to 25 % of the nodes, but really slow to get to all of the nodes. Note that if a transaction is sent to the network and it is not included in any block for a period of time, the client may try to send it again, producing latter retransmissions of the same transaction and thus an increase on $\Delta(T_i)$.

Transaction vs Block Propagation. We also studied if the first nodes that relay transactions and blocks are always the same, that is, we analyzed which nodes were sending us transactions and blocks that we do not have seen previously. Table 2 shows the nodes that are more often relaying transactions and blocks for the first time. The first thing to notice is that although we are listening to more than 1300 different nodes, the best 20 nodes (in terms of transaction and block propagation speed) are responsible for first relaying more than 70 % of both blocks and transactions. It is also interesting to note that there is some overlap between the nodes first relaying blocks and the nodes first relaying transactions: 7 of the best nodes in terms of first relaying blocks are also between the best 20 nodes in terms of first relaying transactions. However, the nodes that are first relaying most of the transactions (nodes 125 and 126) have not relayed any block for the first time.

5 Conclusion and Further Work

Bitcoin is a virtual currency that has been rapidly adopted due to its security robustness, but also for its anonymity and decentralized properties. In this paper we have presented an analysis of the collected data of the decentralized P2P network that supports its information transmission. Data shows that the Bitcoin P2P network is homogeneously spread all over the world, with some exceptions on very low populated areas and underdeveloped countries. Information about node stability shows that there exist a core of around 6000 nodes that are connected during the whole listening period, that is, 37 days. Propagation data shows that the general latency of the P2P Bitcoin network is acceptable for normal nodes but, in some cases, it could be too high for miners, causing them to be working on already mined blocks due to the network delay.

The collection process performed so far, the variety of data collected, and this first brief (due to space constraints) analysis of the information presented in this paper allows us to draw some guidelines for further research. For instance, a network topology analysis could be performed in order to plot the main topological structure of the P2P Network. On the other hand, a more in depth information propagation analysis can be performed by increasing the amount of data collected and the number of connections made to listen to the network.

Acknowledgments. This work was partially supported by a student grant of the Master in Security of Information and Communication Technologies (MISTIC), at the Universitat Autónoma de Barcelona (UAB), and by the Spanish Government projects, TIN2011-27076-C03-02 CO-PRIVACY, TIN2010-15764 N-KHRONOUS, CONSOLIDER INGENIO 2010 CSD2007-0004 ARES, and grant FPU-AP2010-0078.

References

1. Androulaki, E., Karame, G.O., Roeschlin, M., Scherer, T., Capkun, S.: Evaluating user privacy in bitcoin. In: Sadeghi, A.-R. (ed.) FC 2013. LNCS, vol. 7859, pp. 34–51. Springer, Heidelberg (2013). http://dx.doi.org/10.1007/978-3-642-39884-1_4
2. Babaioff, M., Dobzinski, S., Oren, S., Zohar, A.: On bitcoin and red balloons. In: Proceedings of the 13th Association for Computing Machinery (ACM) Conference on Electronic Commerce, EC 2012, pp. 56–73. ACM, New York (2012). http://doi.acm.org/10.1145/2229012.2229022
3. Bamert, T., Decker, C., Elsen, L., Wattenhofer, R., Welten, S.: Have a snack, pay with bitcoins. In: Proceedings of the IEEE International Conference on Peer-to-Peer Computing (P2P) 2013, Trento, Italy (2013)
4. Barber, S., Boyen, X., Shi, E., Uzun, E.: Bitter to better–how to make bitcoin a better currency. In: Keromytis, A.D. (ed.) FC 2012. LNCS, vol. 7397, pp. 399–414. Springer, Heidelberg (2012). http://dx.doi.org/10.1007/978-3-642-32946-3_29
5. Castro, S.: Bitcoin P2P network sniffer. https://github.com/sebicas/bitcoin-sniffer
6. Decker, C., Wattenhofer, R.: Information propagation in the bitcoin network. In: Proceedings of the IEEE International Conference on Peer-to-Peer Computing (P2P) 2013, Trento, Italy, (2013)
7. Geoplugin development team: Geoplugin. http://www.geoplugin.com/
8. Karame, G.O., Androulaki, E., Capkun, S.: Double-spending fast payments in bitcoin. In: Proceedings of the 2012 Association for Computing Machinery (ACM) Conference on Computer and Communications Security, CCS 2012, pp. 906–917. ACM, New York (2012). http://doi.acm.org/10.1145/2382196.2382292
9. Meiklejohn, S., Pomarole, M., Jordan, G., Levchenko, K., McCoy, D., Voelker, G.M., Savage, S.: A fistful of bitcoins: Characterizing payments among men with no names. In: Proceedings of the 2013 Conference on Internet Measurement Conference, IMC 2013, pp. 127–140. ACM, New York (2013). http://doi.acm.org/10.1145/2504730.2504747
10. Nakamoto, S., Andresen, G.e.a.: Bitcoin standard client. https://github.com/bitcoin/bitcoin/
11. Ober, M., Katzenbeisser, S., Hamacher, K.: Structure and anonymity of the bitcoin transaction graph. Future Internet **5**(2), 237–250 (2013). http://www.mdpi.com/1999-5903/5/2/237
12. Reid, F., Harrigan, M.: An analysis of anonymity in the bitcoin system. In: Altshuler, Y., Elovici, Y., Cremers, A.B., Aharony, N., Pentland, A. (eds.) Security and Privacy in Social Networks, pp. 197–223. Springer, New York (2013). http://dx.doi.org/10.1007/978-1-4614-4139-7_10

13. Ron, D., Shamir, A.: Quantitative analysis of the full bitcoin transaction graph. In: Sadeghi, A.-R. (ed.) FC 2013. LNCS, vol. 7859, pp. 6–24. Springer, Heidelberg (2013). http://dx.doi.org/10.1007/978-3-642-39884-1_2
14. The Bitcoin Wiki: Bitcoin protocol rules. https://en.bitcoin.it/wiki/Protocol_rules
15. International Telecommunications Union: Percentage of individuals using the internet 2000–2012, June 2013. http://www.itu.int/en/ITU-D/Statistics/Documents/statistics/2013/Individuals_Internet_2000-2012.xls