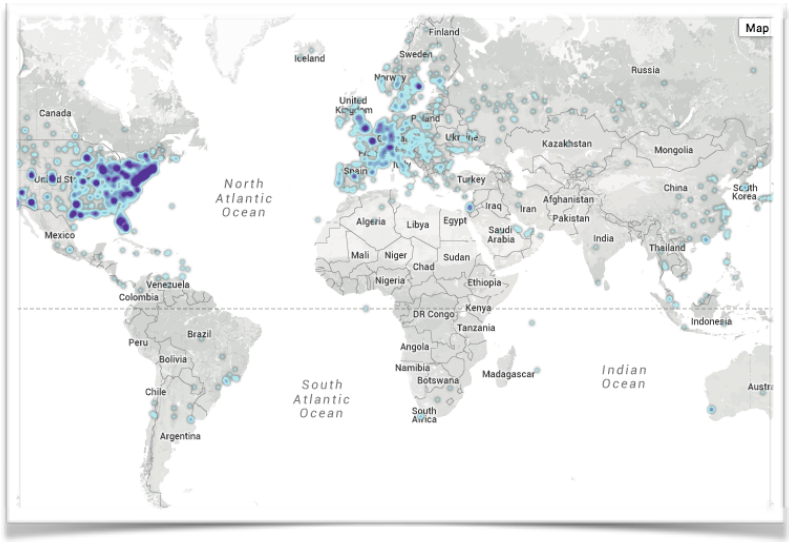


The Bitnodes Project

Estimating the size of the Bitcoin peer-to-peer network.

Addy Yeow
June 12, 2014

Bitcoin operates on an unstructured peer-to-peer overlay network where transactions (transfer of coins) are broadcasted and consolidated into blocks that form the publicly shared block chain.

At the core of the network are the full network nodes that function as critical part of the Bitcoin infrastructure. These nodes are solely responsible in ensuring the transactions and blocks are trickled down the network in a timely manner.

The number of these nodes and their network related metrics were largely unknown as of mid-2013 despite the existence of Bitcoin since early 2009. The lack of such data in the public domain has motivated the launch of the Bitnodes project in April 2013.

Network Snapshot

The crawler (crawl.py) developed for the Bitnodes project works in a similar way as the Bitcoin reference client in terms of discovering reachable nodes to capture a snapshot of the network. Initial nodes are discovered through the list of DNS

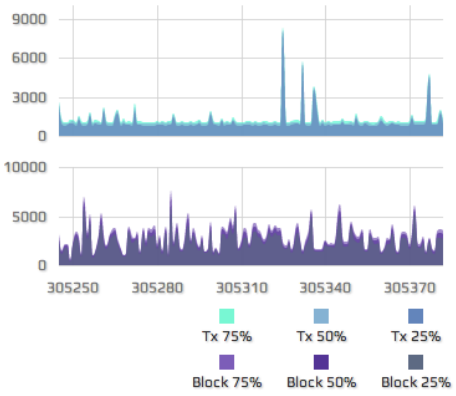
seeders hardcoded in the Bitcoin reference client. Subsequently, the crawler will send a getaddr message to each of its peers to discover new nodes. This is similar to performing a breadth-first search in a network graph.

Determining the accuracy of the captured snapshot of a peer-to-peer network is fundamentally difficult due to:

1) Network churn rate: Nodes joining and leaving the network randomly in addition to nodes with dynamically allocated IP addresses.

2) Unreachable nodes: Nodes rejecting incoming connection and nodes that have reached their maximum allowed connections.

TRANSACTIONS AND BLOCKS PROPAGATION
Time taken in milliseconds to propagate new transactions and blocks across the network.



The crawler addresses the first issue by taking a full snapshot of the entire network in the shortest time possible. The implementation of the crawler uses Python greenlets to spawn 500 concurrent workers to achieve a typical crawl time of 3 minutes to capture all the reachable nodes in the network. As of June 12, 2014, there are approximately 7500 reachable nodes at any one time.

The crawler takes the next snapshot as soon as it completes the first to discover previously unknown or unreachable nodes. All reachable nodes are recorded in a Redis key-value database. A separate instance of the crawler (ping.py) keeps an active connection with each of the reachable nodes in the Redis database.

The method described is necessary in order to take a fresh snapshot of the network consistently while maintaining connection with all the reachable nodes in the network.

Transactions and Blocks Propagation Time

In order to receive transactions from its peers, the crawler has its relay bit set to 1 in the outgoing version message. However, this does not imply all the transactions that ever occurred in the network will be seen immediately by the crawler due to the following rules hardcoded in the Bitcoin reference client:

1) Transaction from origin node is announced only to one peer that is selected on random basis.

2) 25% of all other transactions are announced to all other peers.

The incoming transactions and blocks received by the crawler in the form of inv messages are captured at the kernel level to maximize the accuracy of their arrival timestamps.

As of June 12, 2014, the crawler captures approximately 6400 inv messages per second or 553 million inv messages per day. However, only the first 1000 broadcasts for each inv message are selected in order to generate a consistent aggregate across all the inv messages.

Based on the broadcasted transactions and blocks captured by the crawler over 24 hours period on June 12, 2014, it takes an average of 1.2 seconds for a new transaction and 2.7 seconds for a newly generated block to reach most of the nodes in the network.

The propagation time for a transaction is important as a further delay could allow an attacker to perform double-spending attack against merchant that accepts 0 confirmation transaction, i.e. transaction yet to be included in a block. Delay in block propagation time could waste the mining effort for a block that is already on its way trickling down the network.

Ongoing Development

The ongoing effort in the development of the Bitnodes project is driven by the emerging needs to gather statistical summary of the Bitcoin peer-to-peer network.

As the reachable nodes form a critical part of the Bitcoin infrastructure, an extreme rise and fall of their number may have an immediate impact on the integrity of the network. Bitnodes allows us to observe this significant change and promptly investigate the issue.

The project has also given us the visibility in the adoption rate over time for a specific release of a Bitcoin client among the reachable nodes. The robustness of the Bitcoin network relies on the highly distributed nature of the nodes in terms of their geographical location, network, client implementation, block height and latency. Upcoming metrics on convergence and

divergence of such distribution will ultimately allow us to measure the robustness and decentrality of the Bitcoin peer-to-peer network.

The Bitnodes project is sponsored by the Bitcoin Foundation.

The collected data is published on the project website at https://getaddr.bitnodes.io/

The crawler implementation is available on GitHub at https://github.com/ayeowch/bitnodes