Homework 2: Information Propagation and the Network Layer

Patrick McCorry

Kings College London, UK
patrick.mccorry@kcl.ac.uk

Abstract. We'll focus on the network layer for cryptocurrencies like Bitcoin which is responsible for spreading transactions and blocks to all peers on the network. It is recommended to make notes on this homework sheet for future use.

1 Who maintains the infrastructure?

Last week, we discussed how cryptography protects the blockchain's integrity (*cryptographic hash functions*) and it ensures only the rightful owner can spend their coins (*digital signatures*). This led us down the rabbit-hole to discover that fundamentally, the blockchain, is just just a cryptographic audit trail to let us recompute an entire database from scratch (and new blocks are simply responsible for performing batch updates). The next problem to investigate is how a user can discover the *one true blockchain* for a cryptocurrency and how to ensure new transactions are publicly available so miners can include them in their new blocks. This brings us to the network layer for a cryptocurrency.

There is no appointable central institution or organisation who can take on the role of propagating transactions (or blocks). Instead, cryptocurrencies rely on *peer-to-peer networks* which lets altrusic community members connect with each other on a global scale. This should form a *mesh network* where all peers and miners are connected with each other (which we typically call *Vanilla Bitcoin*). If everyone is connected to everyone else, then we can be assured that all new transactions are propagated to miners (so they are aware of transactions to include in their blocks) and all new blocks are propagted to all other peers (including other miners). Before propagating a transaction, each peer *should* validate new transactions/blocks according to a list of consensus rules (and the state of their local database). This global replication of computation (and verification) is the blockchain's cornerstone of security.

For this homework, we'll focus on the peer discovery (and bootstrapping) process, the announcement protocol (i.e. IP addresses, transactions, and blocks), the type of peers on the network, eclipse attacks and finally why Bitcoin is no longer like a vanilla ice cream.

2 Peer-to-peer network

Let's focus on questions that will help us better understand the peer-to-peer network protocols.

2.1 Bootstrapping new peers

To begin, every user must download the cryptocurrency software (i.e. bitcoind, geth, etc) and find an entry point into the peer-to-peer network. As we mentioned in class, bootstrapping new peers is a non-trivial problem and cryptocurrencies like Bitcoin still rely on a trusted bootstrapping process. In practice, it is maintained/distributed amongst seven influential software developers. While it isn't a great prospect that nine people have the power to eclipse new users on the network, it appears to be the best we can do at the moment. Let's dive a bit deeper:

- Why does the client-server infrastructure not work for cryptocurrencies like Bitcoin? [2 marks]
- There is no single server/entity we can trust to run the network.
- At a high level, what is the process for a new peer to find other peers on the network? In your answer, please consider the bootstrapping process and the last-resort option if finding new peers fails [6 marks]
- Every new peer connects to the DNS seeds and downloads a list of addresses from a seed. They can shuffle/randomize this list before connecting to the first 8.
- Once a peer has connected to other peers on the network, what is the process of finding new peers? [4 marks]
- The peer can simpy request a list of new peers from its connected peers (getaddr). Around 1000 addresses are returned, and they can be kept in local storage (i.e. addrman).
- What is the distinction between outgoing and incoming peers? [2 marks]
- Let's consider it from Peer A's perspective. An incoming connection is when another Peer tries to connect to Peer A. An outgoing connection is when Peer A makes a connection with another peer. Typically there are 8 outgoing connections, and 117 incoming connections. But this is configurable.
- How does a node announce their presence to the rest of the network? [4 marks]
- After Peer A has successfully connected to Peer B, then Peer B should randomly select a subset of its neighbours and relay Peer A's address. Let's say that Peer C was chosen, then Peer C should keep a record of Peer A's address and then relay its address a subset of its peers. Eventually most peers should learn Peer A's address.
- What is the difference between a validating peer and a miner node? [4 marks]
- Validating peers simply verify transactions/blocks and propagate them on the network. They are mostly altrusic in nature. On the other hand, a miner node is responsible validating transactions/blockchains and for producing blocks (i.e. solving proof of work).

- What is an eclipse attack? [4 marks]
- An adversary can take up all connection slots of the victim. This lets them control the network view of the victim (i.e. which blocks/transactions they will receive).

2.2 Information Propagation

Users must *publish their transactions* via the peer-to-peer network to other peers (and miners). As we learnt in class, the peer-to-peer network is a gossip protocol that alone is sufficient for a cryptocurrency to *scale-out* and it also unintentionally helps network adversaries to learn more information than they should know about individual transactions. Again, let's dive a bit deeper to understand information propagation.

- Please explain the three-step protocol for exchanging new transactions/blocks. [3 marks]
- Peer A (with the tx/block) sends an announcement message to all neighbour peers with a hash of the data item (inv). Each peer receives the announcements and checks their local storage. If they are missing the data, then they can request it from Peer A (get_data) using the hash. Peer A will deliver the full tx/block to the requesting peer.
- What is the purpose of a memory pool? [2 marks]
- Store unconfirmed transactions that this peer has received via the peer-to-peer network.
- Why does hosting a new peer harm the network's scalability in vanilla Bitcoin and what impact does it have upon the network's fork rate? Please explain why and you may provide a drawing to help illustrate the idea. [6 marks]
- A transaction / block must propagate across the entire peer-to-peer network in its hunt for other miners. By adding a new peer, we are inadvertently harming the network's scalability as it increases the peer to peer network's size and every new peer adds an additional hop for new transactions/blocks. This is impactful as it increases the likelihood two or more miners will create competing blocks - as they may solve the proof of work while a new block is in transit.
  Drawing: One small network and one large network. Clearly show that a tx/block has to take 'more hops' to get across longer network.
- What is a fast relay network (falcon, fibre, etc) and why does it improve the network's scalability? [4 marks]
- All miners have a private and high-speed connection with each other. It speeds up the process of sending blocks as we no longer need it to propagate across the entire network.
- What is SPV mining? Why does it speed up the propagation of new blocks? And why is it dangerous? [6 marks]
- Miners only validate the proof of work and not the entire block. It can speed up network propagation as only the block header (i.e. 80 bytes) is published across the network which is fast. However it is dangerous as miners cannot

validate transactions in a block - which was shown during the BIP 66 incident - as a result the miners may extend invalid an block and thus their own blocks will forfeited by the peer to peer network.
Note: This is purposely a curve-ball question. SPV mining was introduced in lecture 2, but it can be considered in the context of lecture 3 too - it will test understanding of course content, not learning lecture slides by heart.

- A peer may receive thousands of blocks on the network. How does a peer know they are following the *one true blockchain*? In your answer, please consider the mining puzzle and how blocks are chained together. [6 marks]
- **Longest fork rule:** Peers will always follow the longest and heaviest blockchain, and the one true blockchain will have the most proof of work.
  Chain of blocks and Timestamps. A block will contain the hash of its parent block. This provides a blockchain structure to let us validate blocks in order. As well, peers will be aware that they have caught up as all blocks have a timestamp (i.e. peer an verify a block was recently minted). It is difficult to tamper with a block's timestamp as it contributes towards PoW difficulty.
  Validate all blocks. While a blockchain may have the most proof of work, BIP66 demonstrated that all peers must validate transactions inside blocks to verify that it is indeed a valid blockchain. Unfortunately, SPV clients may follow an invalid blockchain as they only receive block headers.

2.3 0-confirmation transactions

Mallory may sign two transactions:

- TransactionA sends 1 coin from Mallory to Alice,
- TransactionB sends 1 coin from Mallory to Bob.

Typically this is OK. But what if both transaction A and transaction B are spending the same unspent transaction output?[1]

- While both transactions are unconfirmed and propagating across the peer to peer network, which transaction should be considered invalid? [2 marks]
- Neither of the transactions are invalid. It is up to the miners to select one transaction for inclusion in the blockchain (and thus 'confirm it').
- What does it mean to say all users should wait 6 confirmations before considering a transaction final? [4 marks]
- Let's say the transaction was confirmed in block 10. Then 5 more blocks (11,12,13,14,15) have extended block 10. Thus the transaction is in a fork with at least six blockchains from its inclusion. We consider it final as it is unlikely a competing blockchain fork with at least 6 blocks will appear (i.e. an adversary miner will need to repeat the work for all six blocks - which takes around 1 hour in Bitcoin).

[1] Or if mallory's balance is only 1 coin in the account-based model.

- The heuristic in Bitcoin is to wait 6 blocks for confirming a transaction, why do we need to wait for more blocks in Ethereum? [4 marks]
- In Bitcoin, a block is produced every 10 minutes on average. Thus six blocks represents around 1 hour of work. Whereas in Ethereum a block is produced every 12 seconds on average. Thus six blocks represents just over 1 minute of work. So the 'proof of work' security is not the same as it is significantly easier (and cost-effective) for an adversary to repeat 1 minute of work. Note: Again, another curve-ball question. We discussed during lecture 1 about proof of work (and the length of time) which has a direct impact on whether we can consider a transaction 'confirmed'.
- An SPV peer does not validate every transaction in the blockchain and instead only relies on the chain of block headers with the most proof of work. How can we prove to an SPV client that a 0-confirmation transaction has recently been included in a block? [4 marks]
- A peer on the network can send the SPV client the block hash (ID), and a merkle tree branch for the transaction which acts as an inclusion proof. The spv client re-computes the merkle tree using the unconfirmed transaction. If the computed root matches the block header's root, then it is convinced.

While the answers will be released in a week's time, if you found any of the questions difficult then please visit Patrick McCorry during his office hours.