Chapter 20

Eventual Consistency & Bitcoin

How would you implement an ATM? Does the following implementation work satisfactorily?

Algorithm 20.1 Naive ATM
1. ATM makes withdrawal request to bank
2. ATM waits for response from bank
3. if balance of customer sufficient then
4. ATM dispenses cash
5. else
6. ATM displays error
7. end if

Remarks:
• A connection problem between the bank and the ATM may block Algorithm 20.1 in Line 2.
• A network partition is a failure where a network splits into at least two parts that cannot communicate with each other. Intuitively any non-trivial distributed system cannot proceed during a partition and maintain consistency. In the following we introduce the tradeoff between consistency, availability and partition tolerance.
• There are numerous causes for partitions to occur, e.g., physical disconnections, software errors, or incompatible protocol versions. From the point of view of a node in the system, a partition is similar to a period of sustained message loss.

20.1 Consistency, Availability and Partitions

Definition 20.2 (Consistency). All nodes in the system agree on the current state of the system.

Definition 20.3 (Availability). The system is operational and instantly processing incoming requests.

Definition 20.4 (Partition Tolerance). Partition tolerance is the ability of a distributed system to continue operating correctly even in the presence of a network partition.

Theorem 20.5 (CAP Theorem). It is impossible for a distributed system to simultaneously provide Consistency, Availability and Partition Tolerance. A distributed system can satisfy any two of these but not all three.

Proof. Assume two nodes, sharing some state. The nodes are in different partitions, i.e., they cannot communicate. Assume a request wants to update the state and contacts a node. The node may either: 1) update its local state, resulting in inconsistent states, or 2) not update its local state, i.e., the system is no longer available for updates.

Algorithm 20.6 Partition tolerant and available ATM
1. if bank reachable then
2. Synchronize local view of balances between ATM and bank
3. if balance of customer insufficient then
4. ATM displays error and aborts user interaction
5. else
6. ATM dispenses cash
7. ATM logs withdrawal for synchronization
8. end if

Remarks:
• Algorithm 20.6 is partition tolerant and available since it continues to process requests even when the bank is not reachable.
• The ATM’s local view of the balances may diverge from the balances as seen by the bank, therefore consistency is no longer guaranteed.
• The algorithm will synchronize any changes it made to the local balances back to the bank once connectivity is re-established. This is known as eventual consistency.

Definition 20.7 (Eventual Consistency). If no new updates to the shared state are issued, then eventually the system is in a quiescent state, i.e., no more messages need to be exchanged between nodes, and the shared state is consistent.

Remarks:
• Eventual consistency is a form of weak consistency.
• Eventual consistency guarantees that the state is eventually agreed upon, but the nodes may disagree temporarily.
• During a partition, different updates may semantically conflict with each other. A conflict resolution mechanism is required to resolve the conflicts and allow the nodes to eventually agree on a common state.
20.2. BITCOIN

• One example of eventual consistency is the Bitcoin cryptocurrency system.

20.2 Bitcoin

Definition 20.8 (Bitcoin Network). The Bitcoin network is a randomly connected overlay network of a few thousand nodes, controlled by a variety of owners. All nodes perform the same operations, i.e., it is a homogenous network and without central control.

Remarks:
• The lack of structure is intentional: it ensures that an attacker cannot strategically position itself in the network and manipulate the information exchange. Information is exchanged via a simple broadcasting protocol.

Definition 20.9 (Address). Users may generate any number of private keys, from which a public key is then derived. An address is derived from a public key and may be used to identify the recipient of funds in Bitcoin. The private/public key pair is used to uniquely identify the owner of funds at an address.

Remarks:
• The terms public key and address are often used interchangeably, since both are public information. The advantage of using an address is that its representation is shorter than the public key.
• It is hard to link addresses to the user that controls them, hence Bitcoin is often referred to as being pseudonymous.
• Not every user needs to run a fully validating node, and end-users will likely use a lightweight client that only temporarily connects to the network.
• The Bitcoin network collaboratively tracks the balance in bitcoins of each address.
• The address is composed of a network identifier byte, the hash of the public key and a checksum. It is commonly stored in base 58 encoding, a custom encoding similar to base 64 with some ambiguous symbols removed, e.g., lowercase letter “l” since it is similar to the number “1”.
• The hashing algorithm produces addresses of size 20 bytes. This means that there are $2^{160}$ distinct addresses. It might be tempting to brute force a target address, however at one billion trials per second one still requires approximately $2^{25}$ years in expectation to find a matching private/public key pair. Due to the birthday paradox the odds improve if instead of brute forcing a single address we attempt to brute force any address. While the odds of a successful trial increase with the number of addresses, lookups become more costly.

Definition 20.10 (Output). An output is a tuple consisting of an amount of bitcoins and a spending condition. Most commonly the spending condition requires a valid signature associated with the private key of an address.

Remarks:
• Spending conditions are scripts that offer a variety of options. Apart from a single signature, they may include conditions that require the result of a simple computation, or the solution to a cryptographic puzzle.
• Outputs exist in two states: unspent and spent. Any output can be spent at most once. The address balance is the sum of bitcoin amounts in unspent outputs that are associated with the address.
• The set of unspent transaction outputs (UTXO) and some additional global parameters is the shared state of Bitcoin. Every node in the Bitcoin network holds a complete replica of that state. Local replicas may temporarily diverge, but consistency is eventually re-established.

Definition 20.11 (Input). An input is a tuple consisting of a reference to a previously created output and arguments (signature) to the spending condition, proving that the transaction creator has the permission to spend the referenced output.

Definition 20.12 (Transaction). A transaction is a datastructure that describes the transfer of bitcoins from spenders to recipients. The transaction consists of a number of inputs and new outputs. The inputs result in the referenced outputs spent (removed from the UTXO), and the new outputs being added to the UTXO.

Remarks:
• Inputs reference the output that is being spent by a $(h, i)$-tuple, where $h$ is the hash of the transaction that created the output, and $i$ specifies the index of the output in that transaction.
• Transactions are broadcast in the Bitcoin network and processed by every node that receives them.

Remarks:
• Note that the effect of a transaction on the state is deterministic. In other words if all nodes receive the same set of transactions in the same order (Definition 15.8), then the state across nodes is consistent.
• The outputs of a transaction may assign less than the sum of inputs, in which case the difference is called the transaction’s fee. The fee is used to incentivize other participants in the system (see Definition 20.19).
• Notice that so far we only described a local acceptance policy. Nothing prevents nodes to locally accept different transactions that spend the same output.
Algorithm 20.13 Node Receives Transaction
1: Receive transaction $t$
2: for each input $(h, i)$ in $t$ do
3: if output $(h, i)$ is not in local UTXO or signature invalid then
4: Drop $t$ and stop
5: end if
6: end for
7: if sum of values of inputs < sum of values of new outputs then
8: Drop $t$ and stop
9: end if
10: for each input $(h, i)$ in $t$ do
11: Remove $(h, i)$ from local UTXO
12: end for
13: Append $t$ to local history
14: Forward $t$ to neighbors in the Bitcoin network

Remarks:

- Transactions are in one of two states: unconfirmed or confirmed. Incoming transactions from the broadcast are unconfirmed and added to a pool of transactions called the memory pool.

Definition 20.14 (Doublespend). A doublespend is a situation in which multiple transactions attempt to spend the same output. Only one transaction can be valid since outputs can only be spent once. When nodes accept different transactions in a doublespend, the shared state becomes inconsistent.

Remarks:

- Doublespends may occur naturally, e.g., if outputs are co-owned by multiple users. However, often doublespends are intentional – we call these doublespend-attacks. In a transaction, an attacker pretends to transfer an output to a victim, only to doublespend the same output in another transaction back to itself.

- Doublespends can result in an inconsistent state since the validity of transactions depends on the order in which they arrive. If two conflicting transactions are seen by a node, the node considers the first to be valid, see Algorithm 20.13. The second transaction is invalid since it tries to spend an output that is already spent. The order in which transactions are seen, may not be the same for all nodes, hence the inconsistent state.

- If doublespends are not resolved, the shared state diverges. Therefore a conflict resolution mechanism is needed to decide which of the conflicting transactions is to be confirmed (accepted by everybody), to achieve eventual consistency.

Definition 20.15 (Proof-of-Work). Proof-of-Work (PoW) is a mechanism that allows a party to prove to another party that a certain amount of computational resources has been utilized for a period of time. A function $F_d(c, x) \rightarrow \{\text{true, false}\}$, where difficulty $d$ is a positive number, while challenge $c$ and

\[ F_d(c, x) = \text{SHA256(SHA256(c x))} < \frac{2^{256}}{d}, \]

Remarks:

- This function concatenates the challenge $c$ and nonce $x$, and hashes them twice using SHA256. The output of SHA256 is a cryptographic hash with a numeric value in $\{0, \ldots, 2^{256} - 1\}$ which is compared to a target value $\frac{2^{256}}{d}$, which gets smaller with increasing difficulty.

- SHA256 is a cryptographic hash function with pseudorandom output. No better algorithm is known to find a nonce $x$ such that the function $F_d(c, x)$ returns true than simply iterating over possible inputs. This is by design to make it difficult to find such an input, but simple to verify the validity once it has been found.

- If the PoW functions of all nodes had the same challenge, the fastest node would always win. However, as we will see in Definition 20.19, each node attempts to find a valid nonce for a node-specific challenge.

Definition 20.17 (Block). A block is a datastructure used to communicate incremental changes to the local state of a node. A block consists of a list of transactions, a reference to a previous block and a nonce. A block lists some transactions the block creator (“miner”) has accepted to its memory-pool since the previous block. A node finds and broadcasts a block when it finds a valid nonce for its PoW function.

Algorithm 20.18 Node Finds Block
1: Nonce $x = 0$, challenge $c$, difficulty $d$, previous block $b_{t-1}$
2: repeat
3: $x = x + 1$
4: until $F_d(c, x) = \text{true}$
5: Broadcast block $b_t = (\text{memory-pool}, b_{t-1}, x)$

Remarks:

- With their reference to a previous block, the blocks build a tree, rooted in the so-called genesis block.

- The primary goal for using the PoW mechanism is to adjust the rate at which blocks are found in the network, giving the network time to synchronize on the latest block. Bitcoin sets the difficulty so that globally a block is created about every 10 minutes in expectation.
20.2. BITCOIN

• Finding a block allows the finder to impose the transactions in its local memory pool to all other nodes. Upon receiving a block, all nodes roll back any local changes since the previous block and apply the new block’s transactions.

• Transactions contained in a block are said to be confirmed by that block.

Definition 20.19 (Reward Transaction). The first transaction in a block is called the reward transaction. The block’s miner is rewarded for confirming transactions by allowing it to mint new coins. The reward transaction has a dummy input, and the sum of outputs is determined by a fixed subsidy plus the sum of the fees of transactions confirmed in the block.

Remarks:

• A reward transaction is the sole exception to the rule that the sum of inputs must be at least the sum of outputs.

• The number of bitcoins that are minted by the reward transaction and assigned to the miner is determined by a subsidy schedule that is part of the protocol. Initially the subsidy was 50 bitcoins for every block, and it is being halved every 210,000 blocks, or 4 years in expectation. Due to the halving of the block reward, the total amount of bitcoins in circulation never exceeds 21 million bitcoins.

• It is expected that the cost of performing the PoW to find a block, in terms of energy and infrastructure, is close to the value of the reward the miner receives from the reward transaction in the block.

Definition 20.20 (Blockchain). The longest path from the genesis block, i.e., root of the tree, to a leaf is called the blockchain. The blockchain acts as a consistent transaction history on which all nodes eventually agree.

Remarks:

• The path length from the genesis block to block \( b \) is the height \( h_b \).

• Only the longest path from the genesis block to a leaf is a valid transaction history, since branches may contradict each other because of doublespends.

• Since only transactions in the longest path are agreed upon, miners have an incentive to append their blocks to the longest chain, thus agreeing on the current state.

• The mining incentives quickly increased the difficulty of the PoW mechanism: initially miners used CPUs to mine blocks, but CPUs were quickly replaced by GPUs, FPGAs and even application specific integrated circuits (ASICs) as bitcoins appreciated. This results in an equilibrium today in which only the most cost efficient miners, in terms of hardware supply and electricity, make a profit in expectation.

• If multiple blocks are mined more or less concurrently, the system is said to have forked. Forks happen naturally because mining is a distributed random process and two new blocks may be found at roughly the same time.

Algorithm 20.21 Node Receives Block

1. Receive block \( b \)
2. For this node the current head is block \( b_{\text{max}} \) at height \( h_{\text{max}} \)
3. Connect block \( b \) in the tree as child of its parent \( p \) at height \( h_p = h_p + 1 \)
4. if \( h_b > h_{\text{max}} \) then
5. \( h_{\text{max}} = h_b \)
6. \( b_{\text{max}} = b \)
7. Compute UTXO for the path leading to \( b_{\text{max}} \)
8. Cleanup memory pool
9. end if

Remarks:

• Algorithm 20.21 describes how a node updates its local state upon receiving a block. Notice that, like Algorithm 20.13, this describes the local policy and may also result in node states diverging, i.e., by accepting different blocks at the same height as current head.

• Unlike extending the current path, switching paths may result in confirmed transactions no longer being confirmed, because the blocks in the new path do not include them. Switching paths is referred to as a reorg.

• Cleaning up the memory pool involves 1) removing transactions that were confirmed in a block in the current path, 2) removing transactions that conflict with confirmed transactions, and 3) adding transactions that were confirmed in the previous path, but are no longer confirmed in the current path.

• In order to avoid having to recompute the entire UTXO at every new block being added to the blockchain, all current implementations use data structures that store undo information about the operations applied by a block. This allows efficient switching of paths and updates of the head by moving along the path.

Theorem 20.22. Forks are eventually resolved and all nodes eventually agree on which is the longest blockchain. The system therefore guarantees eventual consistency.

Proof. In order for the fork to continue to exist, pairs of blocks need to be found in close succession, extending distinct branches, otherwise the nodes on the shorter branch would switch to the longer one. The probability of branches being extended almost simultaneously decreases exponentially with the length of the fork, hence there will eventually be a time when only one branch is being extended, becoming the longest branch.
20.3 Smart Contracts

Definition 20.23 (Smart Contract). A smart contract is an agreement between two or more parties, encoded in such a way that the correct execution is guaranteed by the blockchain.

Remarks:

• Contracts allow business logic to be encoded in Bitcoin transactions which mutually guarantee that an agreed upon action is performed. The blockchain acts as a conflict mediator, should a party fail to honor an agreement.
• The use of scripts as spending conditions for outputs enables smart contracts. Scripts, together with some additional features such as timelocks, allow encoding complex conditions, specifying who may spend the funds associated with an output and when.

Definition 20.24 (Timelock). Bitcoin provides a mechanism to make transactions invalid until some time in the future: timelocks. A transaction may specify a locktime: the earliest time, expressed in either a Unix timestamp or a blockchain height, at which it may be included in a block and therefore be confirmed.

Remarks:

• Timelocks are used to replace or supersede transactions: a timelocked transaction \( t \) can be replaced by another transaction \( t_0 \). A timelock can be broadcast in the network before the replaced transaction \( t_0 \) becomes valid.

Definition 20.25 (Singlesig and Multisig Outputs). When an output can be claimed by providing a single signature it is called a singlesig output. In contrast the script of multisig outputs specifies a set of \( m \) public keys and requires \( k \)-of-\( m \) valid signatures from distinct matching public keys from that set in order to be valid.

Algorithm 20.26 Parties \( A \) and \( B \) create a 2-of-2 multisig output \( o \)

```plaintext
1: \( B \) sends a list \( I_B \) of inputs with \( c_B \) coins to \( A \)
2: \( A \) selects its own inputs \( I_A \) with \( c_A \) coins
3: \( A \) creates transaction \( t_A[I_A,I_B]; o = c_A + c_B \rightarrow \{A,B\} \)
4: \( A \) creates timelocked transaction \( t_A[I_A]; \{c_A \rightarrow A, c_B \rightarrow B\} \) and signs it
5: \( A \) sends \( t_A \) and \( t_r \) to \( B \)
6: \( B \) signs both \( t_A \) and \( t_r \) and sends them to \( A \)
7: \( A \) signs \( t_r \) and broadcasts it to the Bitcoin network
```

Remarks:

• Most smart contracts begin with the creation of a 2-of-2 multisig output, requiring a signature from both parties. Once the transaction creating the multisig output is confirmed in the blockchain, both parties are guaranteed that the funds of that output cannot be spent unilaterally.

Algorithm 20.27 Simple Micropayment Channel from \( S \) to \( R \) with capacity \( c \)

```plaintext
1: \( c_S = 0, c_R = 0 \)
2: \( S \) and \( R \) use Algorithm 20.26 to set up output \( o \) with value \( c \) from \( S \)
3: Create settlement transaction \( t_s[I_S]; \{c_S \rightarrow S, c_R \rightarrow R\} \)
4: while channel open and \( c_S < c \) do
5: In exchange for good with value \( \delta \)
6: \( c_S = c_S + \delta \)
7: \( c_R = c_R - \delta \)
8: Update \( t_j \) with outputs \( \{c_R \rightarrow R, c_S \rightarrow S\} \)
9: \( S \) signs and sends \( t_j \) to \( R \)
10: end while
11: \( R \) signs last \( t_j \) and broadcasts it
```
20.4. WEAK CONSISTENCY

Remarks:

- Algorithm 20.27 implements a Simple Micropayment Channel, a smart contract that is used for rapidly adjusting micropayments from a spender to a recipient. Only two transactions are ever broadcast and inserted into the blockchain: the setup transaction $t_s$ and the last settlement transaction $t_f$. There may have been any number of updates to the settlement transaction, transferring ever more of the shared output to the recipient.
- The number of bitcoins $c$ used to fund the channel is also the maximum total that may be transferred over the simple micropayment channel.
- At any time the recipient $R$ is guaranteed to eventually receive the bitcoins, since she holds a fully signed settlement transaction, while the spender only has partially signed ones.
- The simple micropayment channel is intrinsically unidirectional. Since the recipient may choose any of the settlement transactions in the protocol, she will use the one with maximum payout for her. If we were to transfer bitcoins back, we would be reducing the amount paid out to the recipient, hence she would choose not to broadcast that transaction.

20.4 Weak Consistency

Eventual consistency is only one form of weak consistency. A number of different tradeoffs between partition tolerance and consistency exist in literature.

Definition 20.28 (Monotonic Read Consistency). If a node $u$ has seen a particular value of an object, any subsequent accesses of $u$ will never return any older values.

Remarks:

- Users are annoyed if they receive a notification about a comment on an online social network, but are unable to reply because the web interface does not show the same notification yet. In this case the notification acts as the first read operation, while looking up the comment on the web interface is the second read operation.

Definition 20.29 (Monotonic Write Consistency). A write operation by a node on a data item is completed before any successive write operation by the same node (i.e. system guarantees to serialize writes by the same node).

Remarks:

- The ATM must replay all operations in order, otherwise it might happen that an earlier operation overwrites the result of a later operation, resulting in an inconsistent final state.

Definition 20.30 (Read-Your-Write Consistency). After a node $u$ has updated a data item, any later reads from node $u$ will never see an older value.

Definition 20.31 (Causal Relation). The following pairs of operations are said to be causally related:

- Two writes by the same node to different variables.
- A read followed by a write of the same node.
- A read that returns the value of a write from any node.
- Two operations that are transitively related according to the above conditions.

Remarks:

- The first rule ensures that writes by a single node are seen in the same order. For example if a node writes a value in one variable and then signals that it has written the value by writing in another variable. Another node could then read the signalling variable but still read the old value from the first variable, if the two writes were not causally related.

Definition 20.32 (Causal Consistency). A system provides causal consistency if operations that potentially are causally related are seen by every node of the system in the same order. Concurrent writes are not causally related, and may be seen in different orders by different nodes.

Chapter Notes

The CAP theorem was first introduced by Foa and Brewer [FB99], although it is commonly attributed to a talk by Eric Brewer [Bre00]. It was later proven by Gilbert and Lynch [GL92] for the asynchronous model. Gilbert and Lynch also showed how to relax the consistency requirement in a partially synchronous system to achieve availability and partition tolerance.

Bitcoin was introduced in 2008 by Satoshi Nakamoto [Nak08]. Nakamoto is thought to be a pseudonym used by either a single person or a group of people; it is still unknown who invented Bitcoin, giving rise to speculation and conspiracy theories. Among the plausible theories are noted cryptographers Nick Szabo [Sza13] and Hal Finney [Fre14]. The first Bitcoin client was published shortly after the paper and the first block was mined on January 3, 2009. The genesis block contained the headline of the release date’s The Times issue “The Times 03/Jan/2009 Chancellor on brink of second bailout for banks”, which serves as proof that the genesis block has been indeed mined on that date, and that no one had mined before that date. The quote in the genesis block is also thought to be an ideological hint: Bitcoin was created in a climate of financial crisis, induced by rampant manipulation by the banking sector, and Bitcoin quickly grew in popularity in anarchic and libertarian circles. The original client is nowadays maintained by a group of independent core developers and remains the most used client in the Bitcoin network.

Central to Bitcoin is the resolution of conflicts due to doublespends, which is solved by waiting for transactions to be included in the blockchain. This however introduces large delays for the confirmation of payments which are
undesirable in some scenarios in which an immediate confirmation is required. Karame et al. [KAC12] show that accepting unconfirmed transactions leads to a non-negligible probability of being defrauded as a result of a doublespending attack. This is facilitated by information eclipsing [DW13], i.e., that nodes do not forward conflicting transactions, hence the victim does not see both transactions of the doublespend. Bamert et al. [BDE+13] showed that the odds of detecting a doublespending attack in real-time can be improved by connecting to a large sample of nodes and tracing the propagation of transactions in the network.

Bitcoin does not scale very well due to its reliance on confirmations in the blockchain. A copy of the entire transaction history is stored on every node in order to bootstrap joining nodes, which have to reconstruct the transaction history from the genesis block. Simple micropayment channels were introduced by Hearn and Spilman [HS12] and may be used to bundle multiple transfers between two parties but they are limited to transferring the funds locked into the channel once. Recently Duplex Micropayment Channels [DW15] and the Lightning Network [PD15] have been proposed to build bidirectional micropayment channels in which the funds can be transferred back and forth an arbitrary number of times, greatly increasing the flexibility of Bitcoin transfers and enabling a number of features, such as micropayments and routing payments between any two endpoints.

This chapter was written in collaboration with Christian Decker.

Bibliography


