Introduction to Bitcoin for software engineers

Noah Ruderman

About the speaker

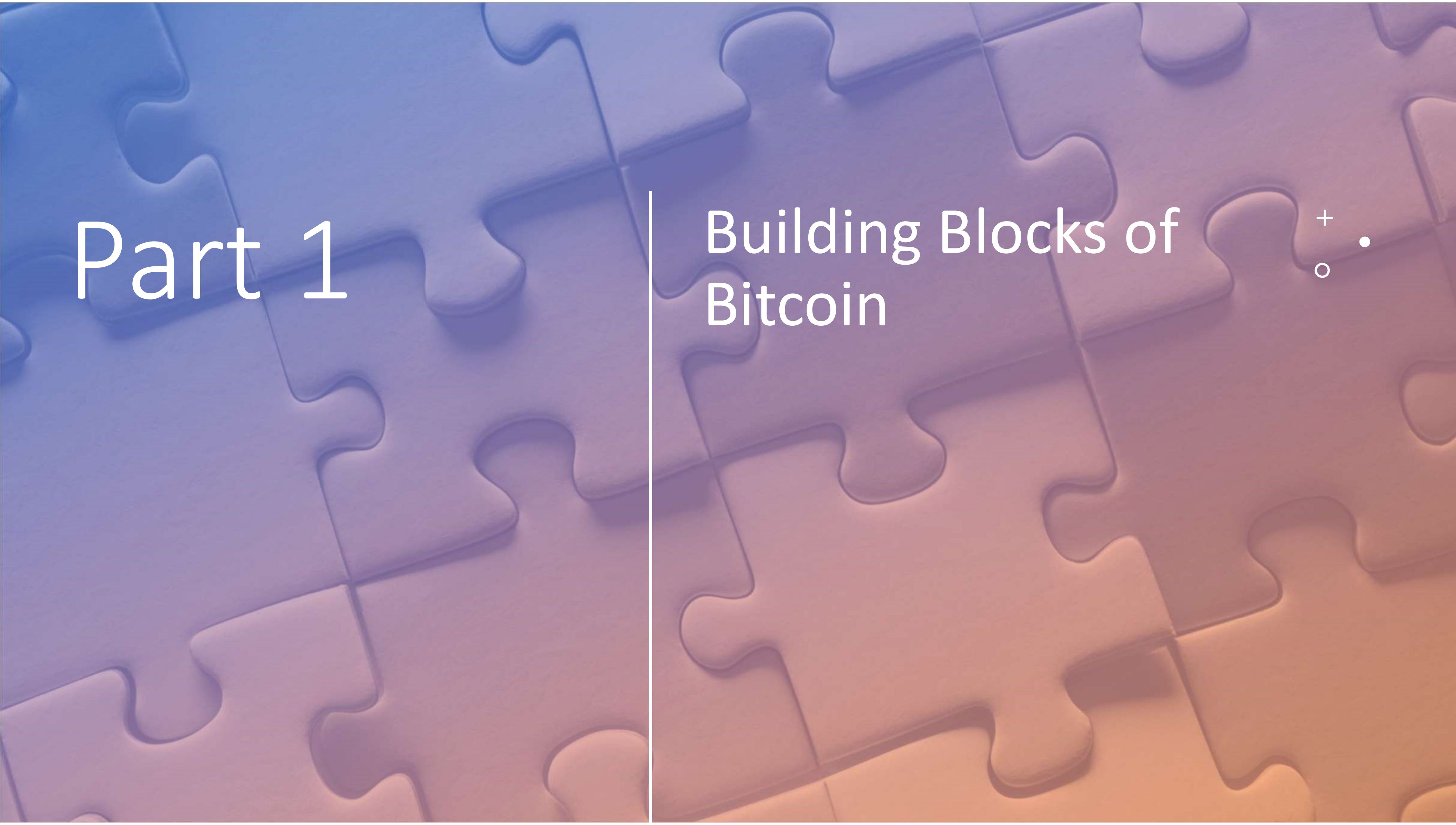
- Noah Ruderman
- Software engineer (ex-Facebook, currently at Microsoft)
- Twitter @devilscompiler
- LinkedIn @nruderman
- Website www.thedevilscompiler.com

What this presentation covers

- The building blocks of Bitcoin
  - Peer-to-Peer (P2P) network
  - Public key (PK) cryptography
  - Hash functions
  - Merkle Roots
  - Blockchain
  - Proof-of-work (PoW)
  - Consensus
- How Bitcoin works from the users perspective
- What is Bitcoin for
- Miscellaneous topics (if there’s time)

What is Bitcoin?

- It’s a composition of mostly cryptographic technologies to create an entirely new asset: p2p network, proof-of-work consensus, public-key cryptography, blockchain, merkle roots
- Like during the birth of the internet, there is no good comprehensive analogy for what Bitcoin is. There are many different analogies that capture individual aspects of Bitcoin but nothing that captures it all.
- Generally speaking, Bitcoin is an information system for recording and transferring wealth (denominated in BTC, the native asset)
- Technically speaking, Bitcoin is like a global, decentralized, append-only database that anyone can write to.

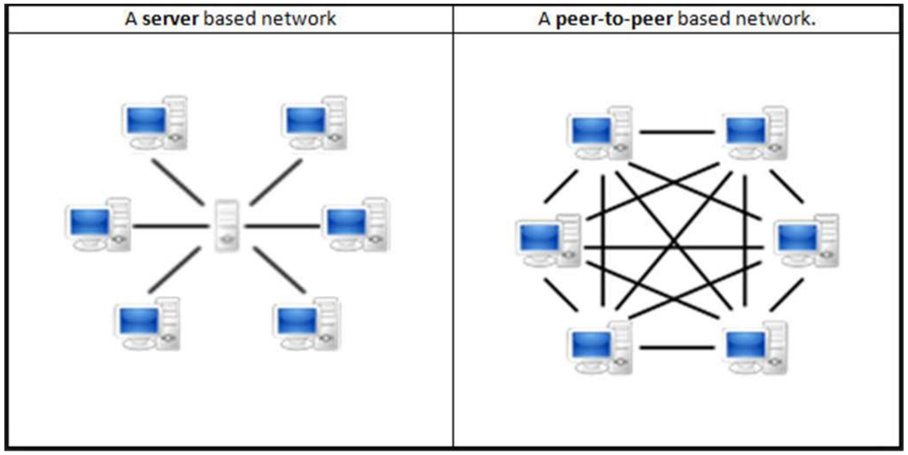


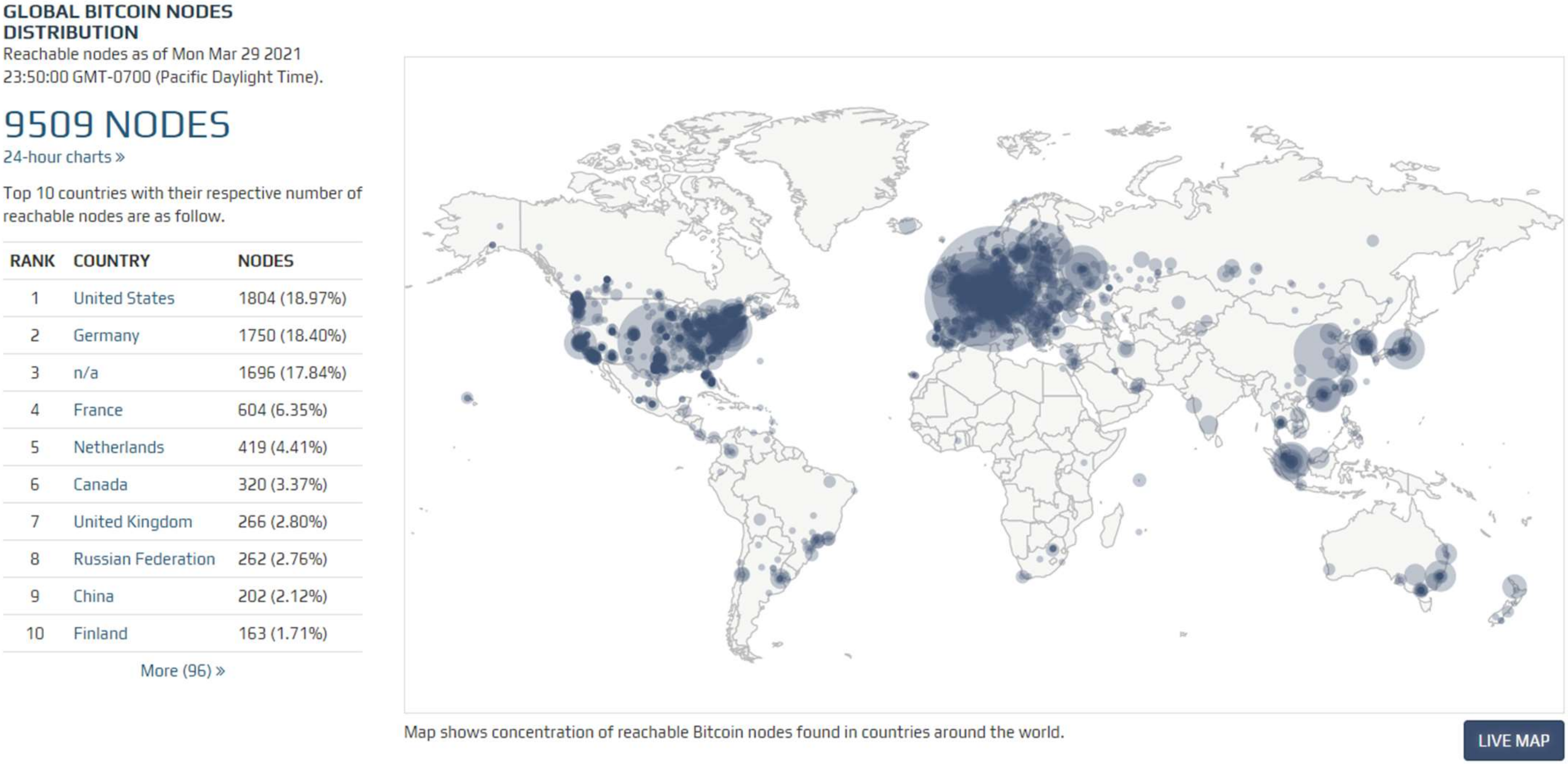
Part 1

Building Blocks of Bitcoin

Anatomy of Bitcoin: Peer-to-Peer (P2P) network

- Bitcoin nodes host and validate a copy of the Bitcoin blockchain.
- A P2P network has only one set of nodes. Each node is both a client and a server.
- Compare a P2P network with a centralized network. A centralized network has nodes which are individually a server, or a client, but not both.
- P2P networks are resilient towards the failure of any individual node.
- Each Bitcoin Core node has by default 10 outbound connections and 115 inbound connections.
- Bitcoin data (transactions and blocks) are gossiped by all nodes.
- There are approximately 10k reachable Bitcoin nodes (and many more unreachable ones). Compare to Tor (6k nodes).





GLOBAL BITCOIN NODES DISTRIBUTION

Reachable nodes as of Mon Mar 29 2021 23:50:00 GMT-0700 (Pacific Daylight Time).

9509 NODES

24-hour charts »

Top 10 countries with their respective number of reachable nodes are as follow.

RANK	COUNTRY	NODES
1	United States	1804 (18.97%)
2	Germany	1750 (18.40%)
3	n/a	1696 (17.84%)
4	France	604 (6.35%)
5	Netherlands	419 (4.41%)
6	Canada	320 (3.37%)
7	United Kingdom	266 (2.80%)
8	Russian Federation	262 (2.76%)
9	China	202 (2.12%)
10	Finland	163 (1.71%)

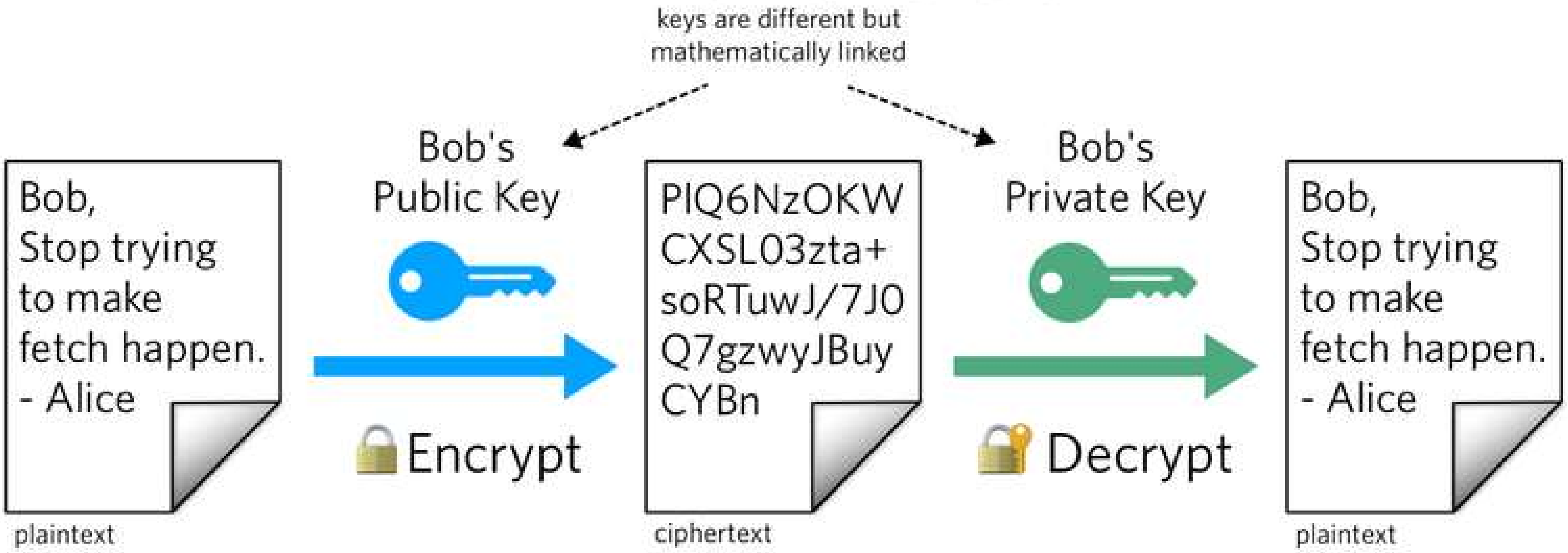
More (96) »

Map shows concentration of reachable Bitcoin nodes found in countries around the world.

Anatomy of Bitcoin: Public-key (PK) cryptography

- PK cryptography involves two pieces of data: a public key and a private key.
- The public key can be given to anyone and it is used either to (a) encrypt a message that can only be decrypted with the private key or (b) validate a signature from a private key.
- The private key is sensitive data, used to either sign data (i.e. proving ownership of the key) or decrypt data encrypted with the public key.
- Technically, the public and private key perform inverse operations. PrivateKey(PublicKey(data)) == data and PublicKey(PrivateKey(data)) == data.
- Encryption is just PublicKey(data) = encrypted_data
- Decryption is just PrivateKey(encrypted_data) = data
- Signing is just PrivateKey(hash(data)) == encrypted_data_hash
- Signature verification is just PublicKey(encrypted_data_hash) = data_hash == hash(data)

Public Key Cryptography



keys are different but mathematically linked

Bob's Public Key

Bob's Private Key

Bob,
Stop trying
to make
fetch happen.
- Alice

plaintext

Encrypt

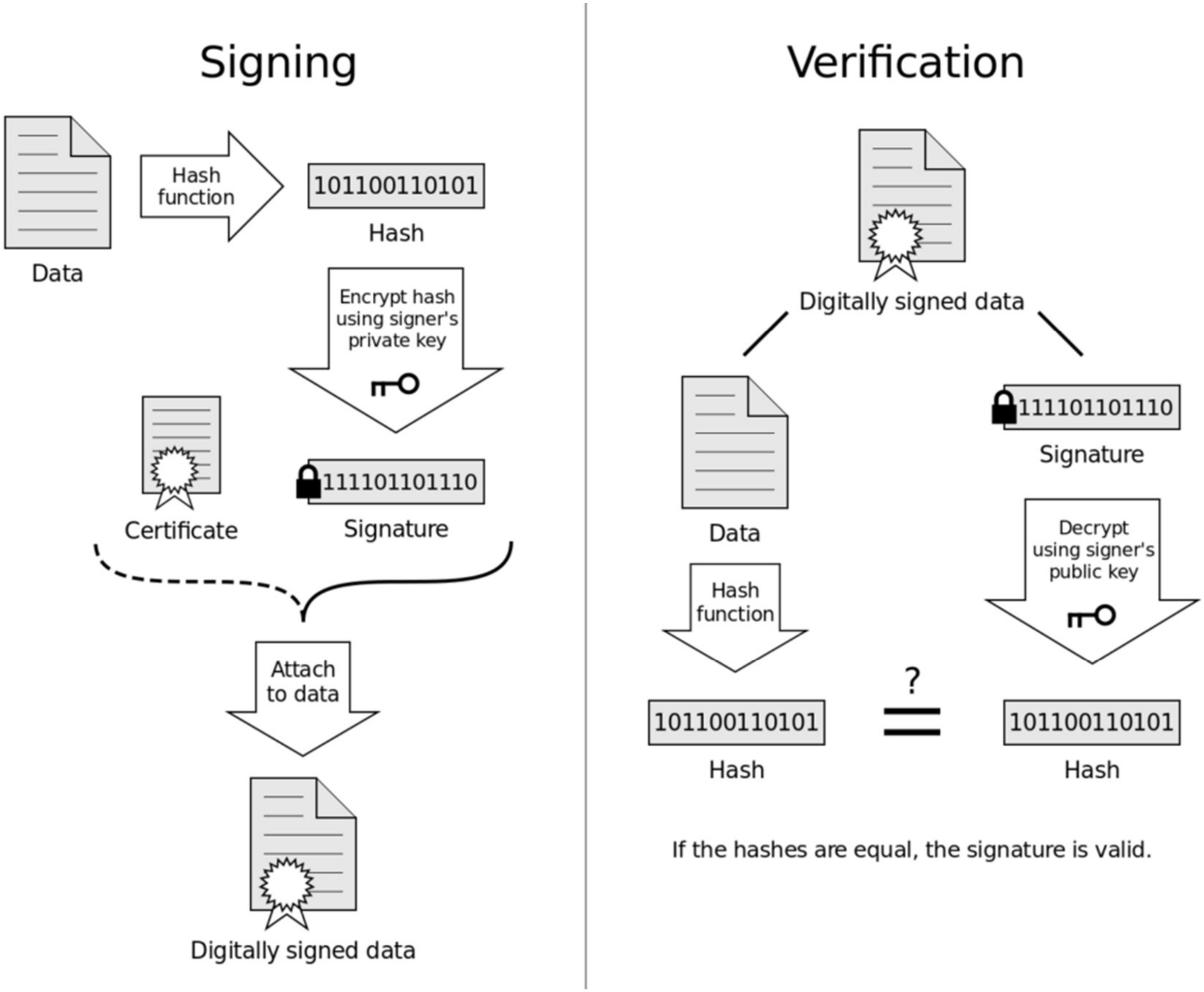
PIQ6NzOKW
CXSL03zta+
soRTuwJ/7J0
Q7gzwyJBuy
CYBn

ciphertext

Decrypt

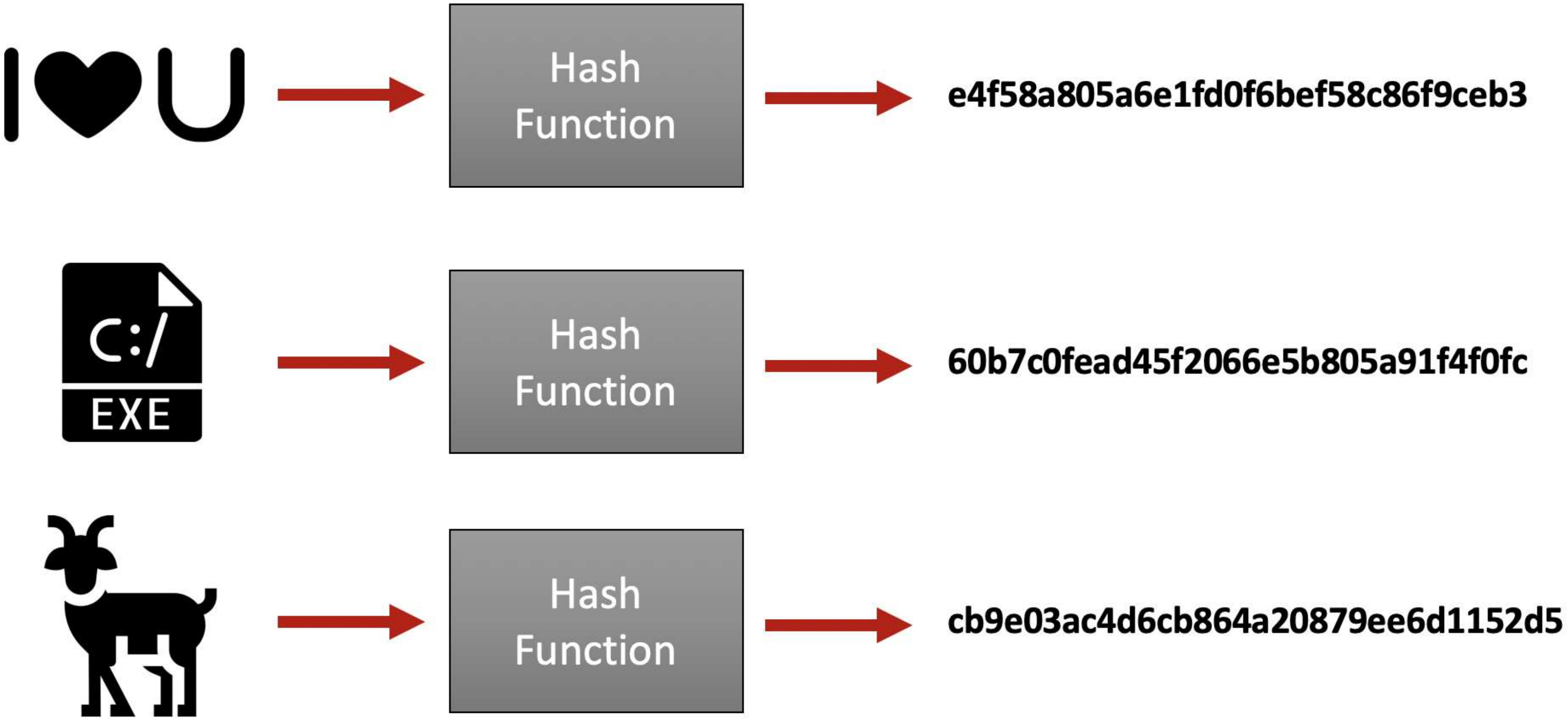
Bob,
Stop trying
to make
fetch happen.
- Alice

plaintext


Signing
Data
Hash function
101100110101
Hash
Encrypt hash using signer's private key
Certificate
111101101110
Signature
Attach to data
Digitally signed data
Verification
Digitally signed data
Data
111101101110
Signature
Hash function
Decrypt using signer's public key
101100110101
Hash
?
101100110101
Hash
If the hashes are equal, the signature is valid.

What is a hash function?

- Hash functions are one-way functions, meaning if hash(x) = y, then knowing y (image) will not be enough information to find x (pre-image) such that hash(x) = y. Should be computationally impractical to reverse.
- Hash functions take input of arbitrary length and return fixed output data. So len(hash(a)) == len(hash(b)) for any a and b. This data is usually supposed to look random.
- Hash functions are generally used to create a unique fingerprint for data
- Bitcoin uses sha256, a cryptographic hash function whose output is 256 bits. The value of this hash function can be interpreted as a number between 0 and (2^256)-1
- sha256(“bitcoin”) == 6b88c087247aa2f07ee1c5956b8e1a9f4c7f892a70e324f1bb3d161e05ca107b
- Hash functions are used in blockchains, merkle trees, and proof-of-work.



Hash Function

e4f58a805a6e1fd0f6bef58c86f9ceb3

Hash Function

60b7c0fead45f2066e5b805a91f4f0fc

Hash Function

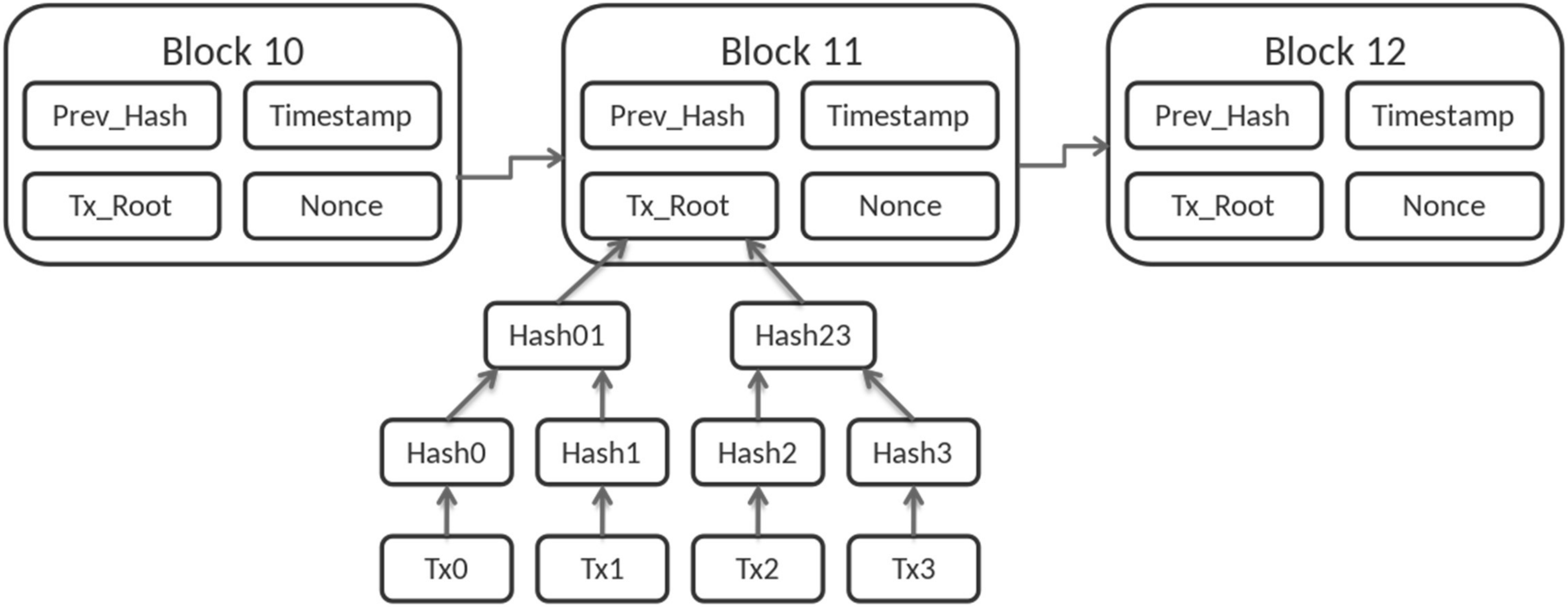
cb9e03ac4d6cb864a20879ee6d1152d5

Properties for Cryptographic Hash Functions

1. Easy to compute:
   - Given message *m*, hash function *h(m)* is easy to compute.
2. One-way function $y = h(x)$:
   - Given *y*, it is very hard to find *x*.
3. Collision-free: (1. strong version and 2. weak version)
   1) It is very hard to find messages m_1 and m_2 with $h(m_1)=h(m_2)$.
   2) Given m_1 and $h(m_1)$, it is very hard find $m_2 \neq m_1$ with $h(m_2)= h(m_1)$.

Anatomy of Bitcoin: blockchain

- A blockchain is a clever way of using hashing to tamper-proof data
- A blockchain is an append-only data structure, and data is added as blocks, hence a chain of blocks. While an actual blockchain looks like a tree, only one linear chain from genesis to present is considered authoritative.
- Blocks contain the following: nonce (used in proof-of-work), merkle root, hash of the previous block*, timestamp.
- If the data in any one block changes for any reason, it will change its hash, and by extension it will require changes for all subsequent blocks that descend from it.
- We will see later that blocks are difficult to construct, incurring a large cost for any attacker.
- In Bitcoin blocks are produced on average every 10 minutes.


Block 10
Prev_Hash
Timestamp
Tx_Root
Nonce
Block 11
Prev_Hash
Timestamp
Tx_Root
Nonce
Block 12
Prev_Hash
Timestamp
Tx_Root
Nonce
Hash01
Hash23
Hash0
Hash1
Hash2
Hash3
Tx0
Tx1
Tx2
Tx3

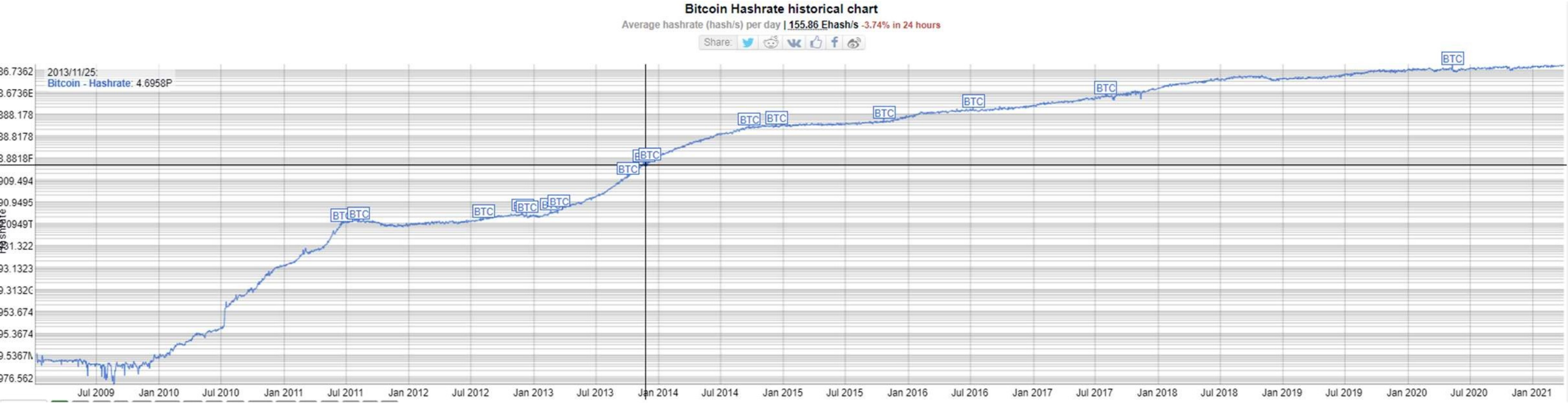
What’s a merkle root?

- Merkle trees are used to tamper-proof data (in this case transactions), but also provide a method of proving that a specific piece of data (transaction) exists in the merkle tree.
- Proving that a transaction exists in a block is useful for light nodes, which don’t maintain a fully copy of the entire blockchain, but only the block headers.
- In a merkle tree, data exists as leaf nodes. Non-leaf nodes are computed as value = hash(concat(hash(left_child), hash(right_child)). If you repeat this process for all levels from bottom to top, the root value is the merkle root.

Anatomy of Bitcoin: Proof-of-Work (PoW)

- PoW is hard to compute, but easy to verify.
- For Bitcoin, PoW is finding a nonce such that the block hash will have a large number of leading zeros.
- Since cryptographic hash functions are one-way functions, the only way to find the correct target value is with brute force.
- On average it takes 2^N nonce guesses to find a hash value with N leading zeros.
- The valid blockchain is the one with the most accumulated work.
- PoW + blockchain is a probabilistic solution to the Byzantine General's problem
- Security model is that 51% of the mining power is honest.
- Bitcoin network has 150+ Exahashes/s (exa = 10^18)

Log scale of Bitcoin hashrate from 2009-present

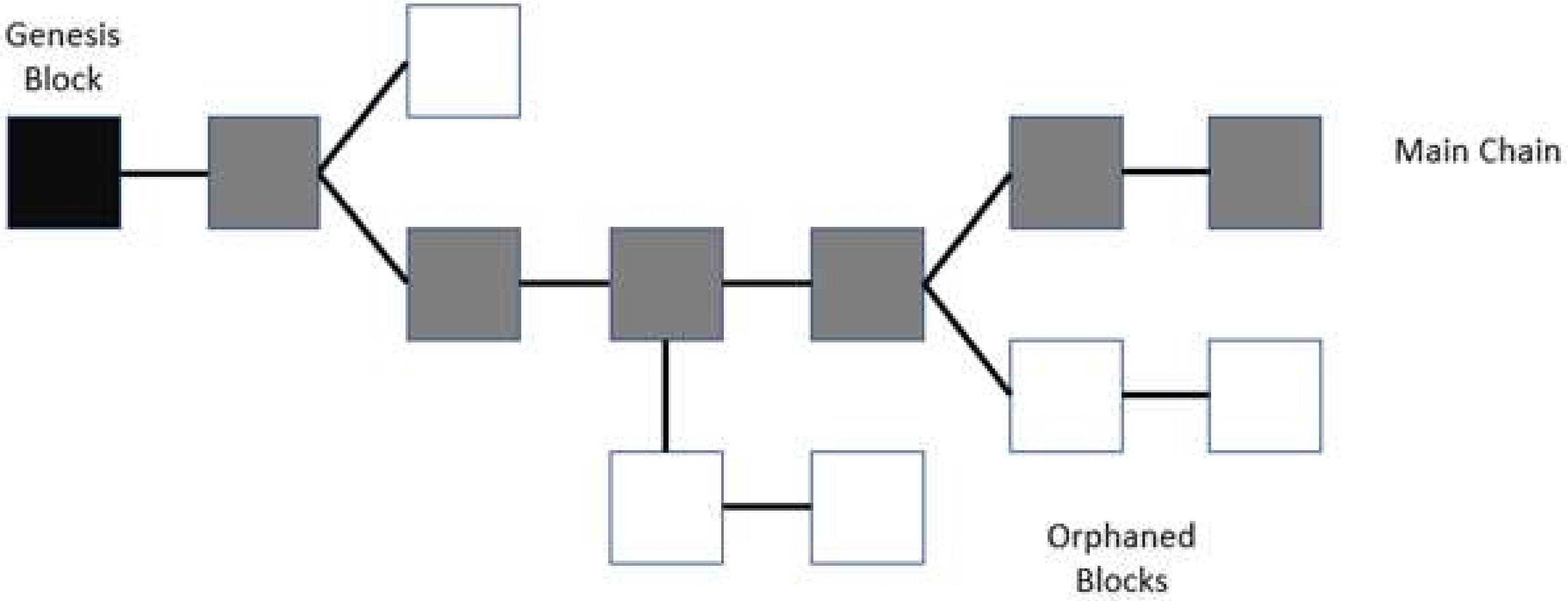


Byzantine generals problem (BGP)

- The BGP problem is about coordination over an unreliable network.
- Example is two generals who are geographically separate and planning an attack. If only one attacks, the attack will fail. They must both attack together at the same time to succeed. Over an unreliable network (messengers can be captured and replaced with spies), this is a hard problem. This can be generalized to N generals.
- Bitcoin is a *probabilistic* solution to the BGP by using proof-of-work.
- Bitcoin’s coordination problem is which linear blockchain (set of blocks) to consider authoritative. All Bitcoin nodes must agree on this over an unreliable network. This process is called consensus.
- For a given confirmed block (N blocks deep in the authoritative blockchain), there is no guarantee that a fork of a deeper block won’t result in a new linear chain being considered authoritative. The probability of this decreases exponentially with new blocks but it never reaches zero.

Anatomy of Bitcoin: Consensus

- All Bitcoin nodes must agree on which blocks are valid despite existing on an unreliable network and receiving data asynchronously.
- Each block is associated with a certain amount of work based on its hash value (and number of leading zeros).
- By selecting the linear set of blocks with the most accumulated work, the network can probabilistically agree on a single chain of blocks.
- Why probabilistic? There may be disputes near the tip of the chain (e.g. two blocks mined at the same time by separate entities.). Bitcoin does not have 100% finality, but it asymptotically approaches 100%.
- Blocks which are not considered authoritative but pass validation checks are called orphaned blocks.


Genesis Block
Main Chain
Orphaned Blocks

Part 2

How Bitcoin works from the perspective of users

Entities on the Bitcoin network

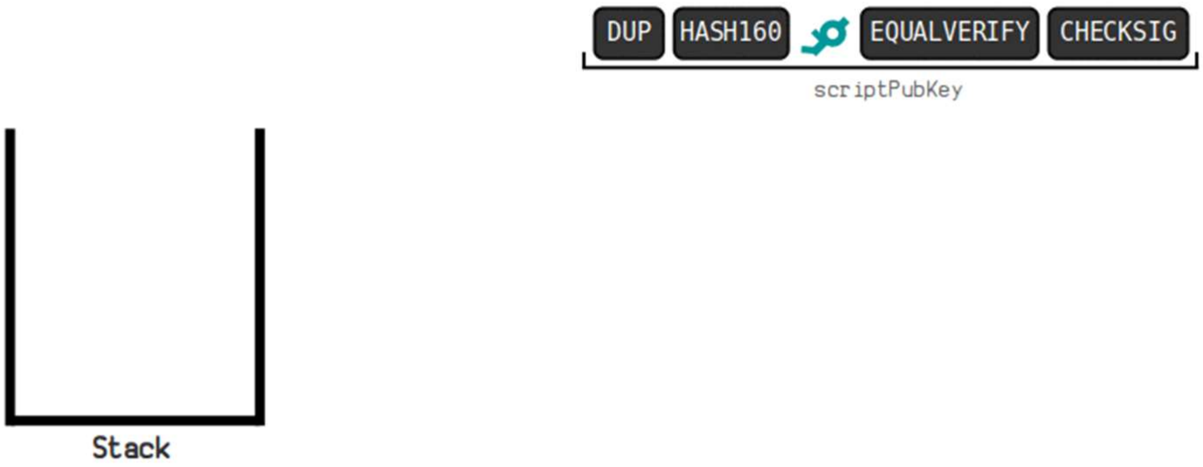
- Bitcoin nodes
    - Gossip transactions (Bitcoin transfers) and blocks (updates to the global state).
    - Validate blocks and transactions, and do not gossip or store invalid data.
    - Select the linear chain of blocks with the most accumulated work.
    - Are not incentivized, but allow full client-side verification
- Bitcoin miners
    - Construct blocks by brute-forcing nonces such that the hash of the block has many leading zeros.
    - Receive block rewards and transaction fees in Bitcoin for their contribution to the network.

Bitcoin programmability

- Bitcoin uses a Turing *incomplete* language called Script
- Script is a stack-based programming language (like reverse polish notation where 3 + 5 is represented as 3 5 +)
- Operators are called opcodes and there are many kinds
- The most popular program is P2PKH (pay to public key hash), for which the script only returns true for the pre-image public-key and its corresponding private key signature.
- https://learnmeabitcoin.com/technical/p2pkh

```
OP_DUP OP_HASH160 12ab8dc588ca9d5787dde7eb29569da63c3a238c OP_EQUALVERIFY OP_CHECKSIG
```

Standard Script: P2PKH



DUP HASH160 EQUALVERIFY CHECKSIG

scriptPubKey

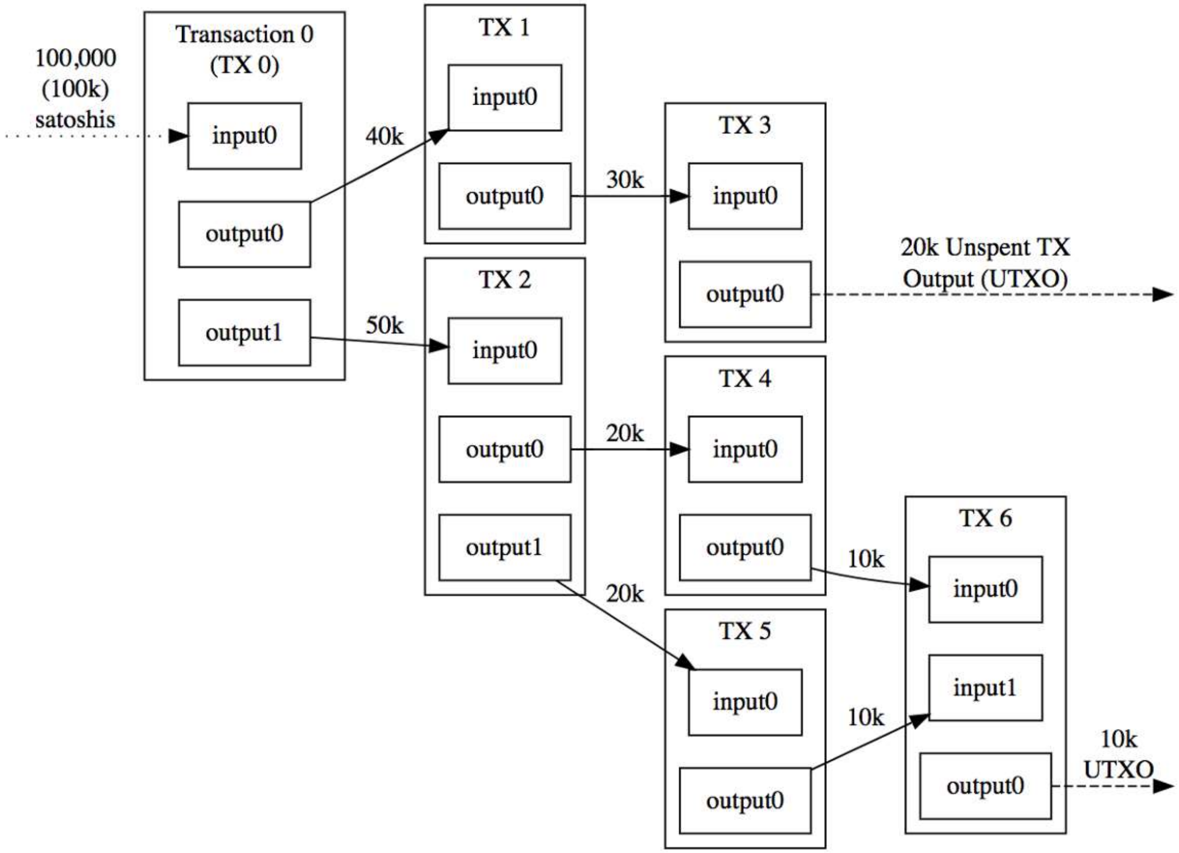
Stack

Smallest unit of Bitcoin ownership: UTXO

- UTXO = unspent transaction output.
- UTXOs include the following:
  - A numerical quantity of Bitcoin (usually denominated in satoshis = 1/100M BTC)
  - A redeem script, or a predicate whose execution must evaluate to true for the coins to be spent. This program may be unique to each UTXO and it is why Bitcoin is programmable (see previous slide).
- Bitcoin does NOT use the account model (address A has balance B) but the UTXO model (UTXO A has redeem script B).
- UTXOs collectively form the global state of Bitcoin (all of existing Bitcoin and what conditions under which they can be spent).

How Bitcoin is transferred: Transactions

- Transactions are how Bitcoin ownership changes hands (e.g. sending someone money).
- Transactions take a set of input UTXOs + the arguments to the redeem script and create a new set of UTXOs such that no new Bitcoin is created. Any difference is the transaction fee which is given to miners. Higher transaction fee = more incentive to include your transaction in a block.
- Transactions are signed by a private key so they cannot be tampered between the point they are broadcast to the network and when they are included in a block by a miner.
- Broadcast transactions are gossiped by the network so nodes and miners receive copies. Transactions which have been broadcast and cached by nodes but not yet included in a block are said to exist in the “mempool” and these transactions are considered “unconfirmed”.



Triple-Entry Bookkeeping (Transaction-To-Transaction Payments) As Used By Bitcoin

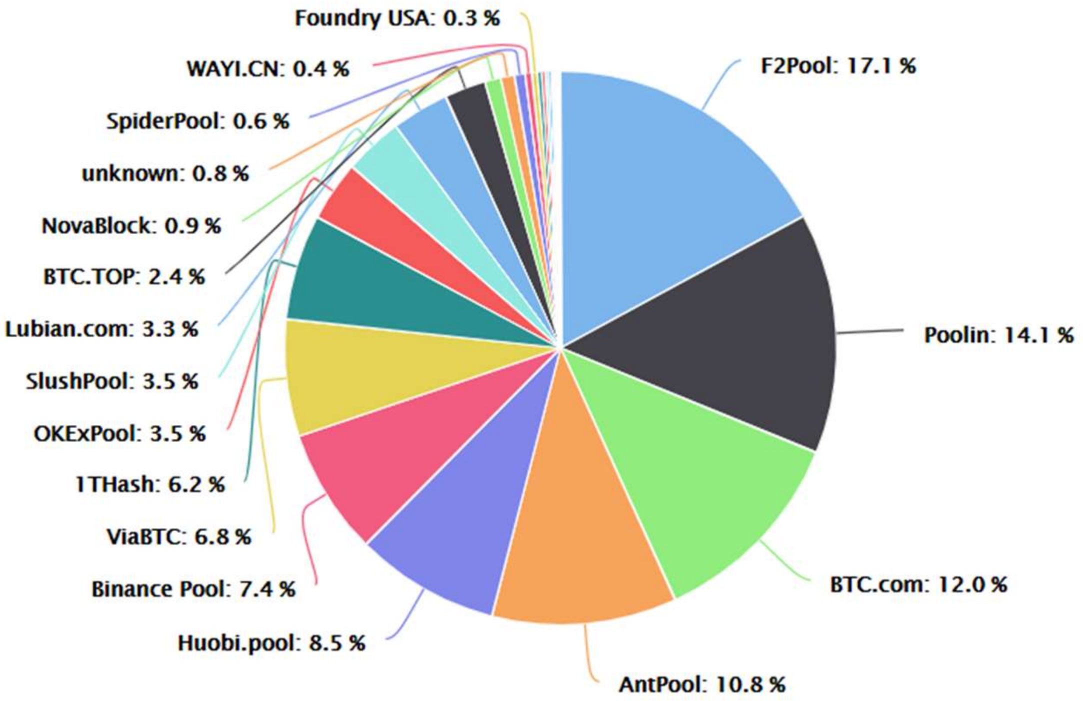
How transactions are included in blocks: Miners

- Bitcoin miners compete to construct the next valid block for the tip of the blockchain. New blocks are constructed on average every 10 minutes (there is high variance to this).
- A block is valid if it passes the usual consistency checks and contains a nonce such that the hash of the block meets the difficulty target (convert hash to a number and see if it's small enough).
- Difficulty target = based off time to produce last 2016 blocks (~2 weeks). Took longer than expected to produce = lower difficulty, took shorter to produce = higher difficulty. Changes every 2 weeks to accommodate changes in collective mining power (e.g. more powerful hardware, miners going offline for any reason).
- A block need not contain any transactions, but if miners do include valid transactions they get to keep the transaction fees.
- Blocks also come with a reward for the miner. It is currently 6.25 BTC and decreases by 50% ever 4 years. Eventually this reward will approach 0.

How miners organize: Pools

- For business reasons, miners want a consistent payout as the expected time to mine a block may be longer than the lifetime of their hardware.
- A mining pool operator decides on the block template, and then sends it to mining pool participants to try various nonces. This is like parallelizing the creation of a single block among separate mining entities.
- Pool participants submit various blocks to the pool operator whose nonces result in a hash with many leading zeros. This proves they are contributing a quantifiable amount of work to the pool. (e.g. a block whose hash has 20 leading zero bits means you may have tried 2^20 nonces).
- When a winning block is found, the pool operator gets the reward and distributes it to participants (minus their own fee) according to the work participants contributed.

Percent of Bitcoin mined by various pools over the last year



Part 3

What is Bitcoin for

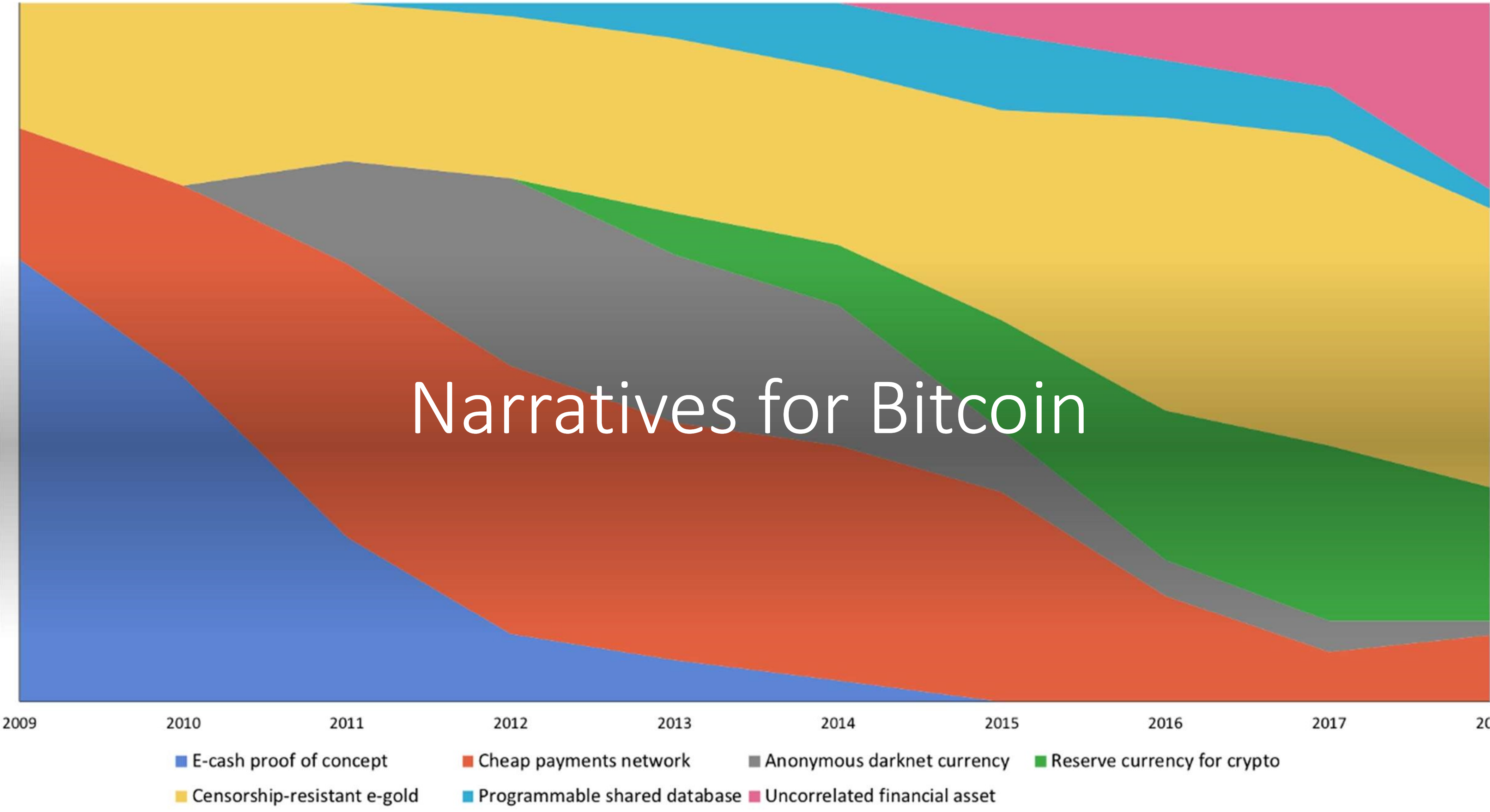
What features does Bitcoin provide?

- Censorship resistance
  - Transactions cannot be censored/blocked by anyone
  - Killing the network is infeasible – it’s global, decentralized, autonomous
  - Only one copy of the blockchain needs to survive – Bitcoin could survive a nuclear war
- Robust client-side verification
  - Counterfeit gold and dollars are common. Counterfeit Bitcoin is not.
- Publicly auditable
  - Can you personally verify the supply of US dollars, or gold?
  - Bitcoin blockchain contains a wealth of data on user activity.
- Permissionlessness
  - Create an account without a social security number, KYC, etc.
  - Nobody can deplatform you
- Immutability
  - Nobody can change account balances or transactions
- Inelastic supply cap
  - Price of gold goes up, gold mining accelerates
  - Bitcoin difficulty adjustment guarantees regular emission of 10 minutes per block
  - Algorithmic supply cap of 21M Bitcoins.

What can you do with Bitcoin?

- Flee with all your wealth in your head (memorize seed words to private key)
  Many countries have capital controls to prevent fleeing with gold, cash, jewelry, etc. (e.g. airport searches)
- Protect yourself from hyperinflation. Bitcoin is designed to have a fixed supply.
- Protect yourself from government financial censorship
  Bank and brokerage accounts can be frozen
- Send money to anyone in the world from anywhere in the world at the speed of information
- Multi-signature addresses have no analog in physical assets
  Only send funds when N-of-M signers approve the transaction
- Machine-to-Machine payments
  Like P2P payments but without the humans (e.g. charging your Tesla, downloading a file)
- Store wealth without the need for physical protection (e.g. vaults)
  Bitcoin is more about the protection of information (private key data)
- Programmable money (Bitcoin is programmable)
- An alternative to an overly complicated, overly regulated financial system.

Narratives for Bitcoin



Bitcoin as gold 2.0

- 10x better verifiability

  There is no such thing as counterfeit Bitcoin (it would be infeasibly expensive to do so)
  Verifying gold requires a way to flatten it and a spectrometer (thousands of dollars)

- 10x better divisibility

  Smallest unit of Bitcoin is a satoshi, 1/100M of a Bitcoin = $0.0005
  Smallest unit of gold is an atom, but smallest bar is typically 1 gram = $55.44

- 10x better portability

  Send Bitcoin anywhere in the world in on hour for a transaction fee
  Send gold most places in the world in weeks plus shipping costs plus insurance

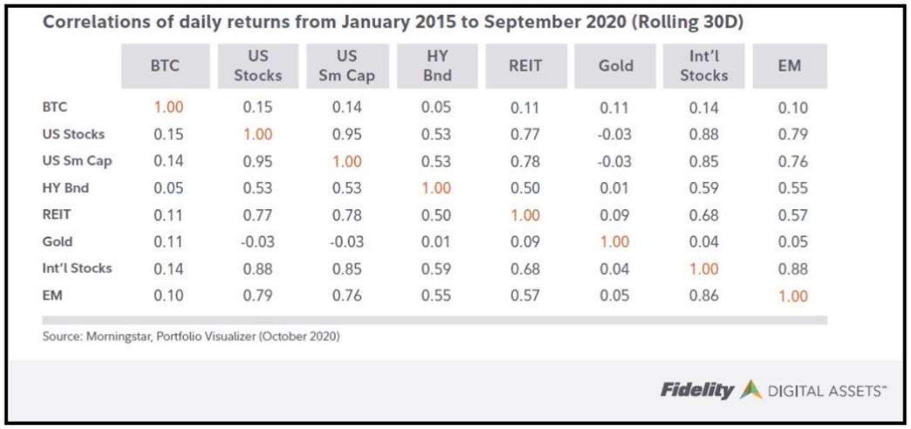
- 10x better custody

  Custody Bitcoin in your head by memorizing a seed phrase, safe to self-custody any amount
  Custody gold with a vault, unsafe to self-custody large amounts

- 10x better scarcity

  Bitcoin has a fixed supply cap of 21M units
  Gold inflates at 1.8% per year and more gold is mined as price increases

Bitcoin as an uncorrelated asset



Correlations of daily returns from January 2015 to September 2020 (Rolling 30D)

	BTC	US Stocks	US Sm Cap	HY Bnd	REIT	Gold	Int'l Stocks	EM
BTC	1.00	0.15	0.14	0.05	0.11	0.11	0.14	0.10
US Stocks	0.15	1.00	0.95	0.53	0.77	-0.03	0.88	0.79
US Sm Cap	0.14	0.95	1.00	0.53	0.78	-0.03	0.85	0.76
HY Bnd	0.05	0.53	0.53	1.00	0.50	0.01	0.59	0.55
REIT	0.11	0.77	0.78	0.50	1.00	0.09	0.68	0.57
Gold	0.11	-0.03	-0.03	0.01	0.09	1.00	0.04	0.05
Int'l Stocks	0.14	0.88	0.85	0.59	0.68	0.04	1.00	0.88
EM	0.10	0.79	0.76	0.55	0.57	0.05	0.86	1.00

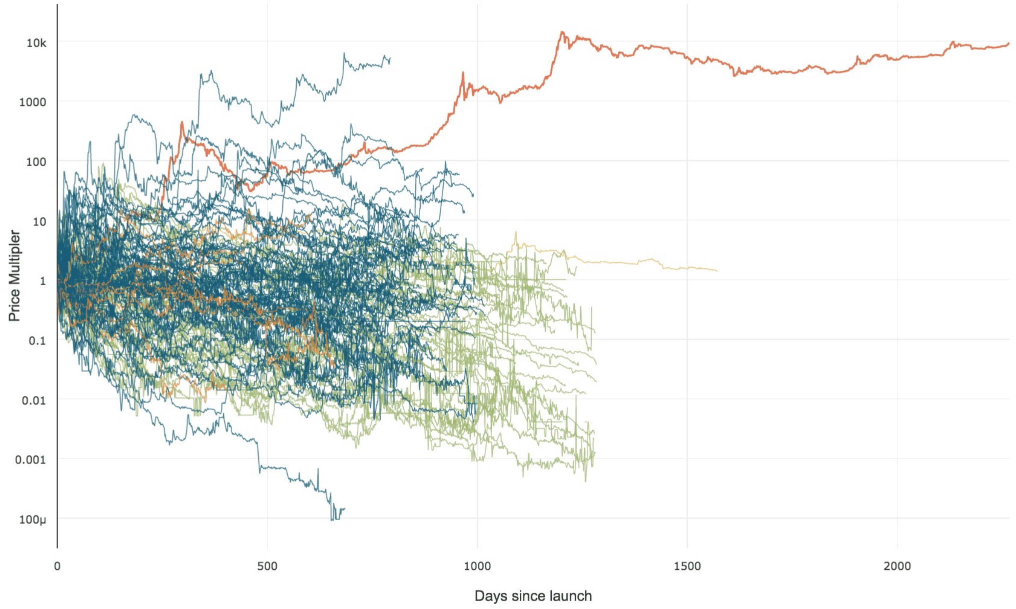
Source: Morningstar, Portfolio Visualizer (October 2020)

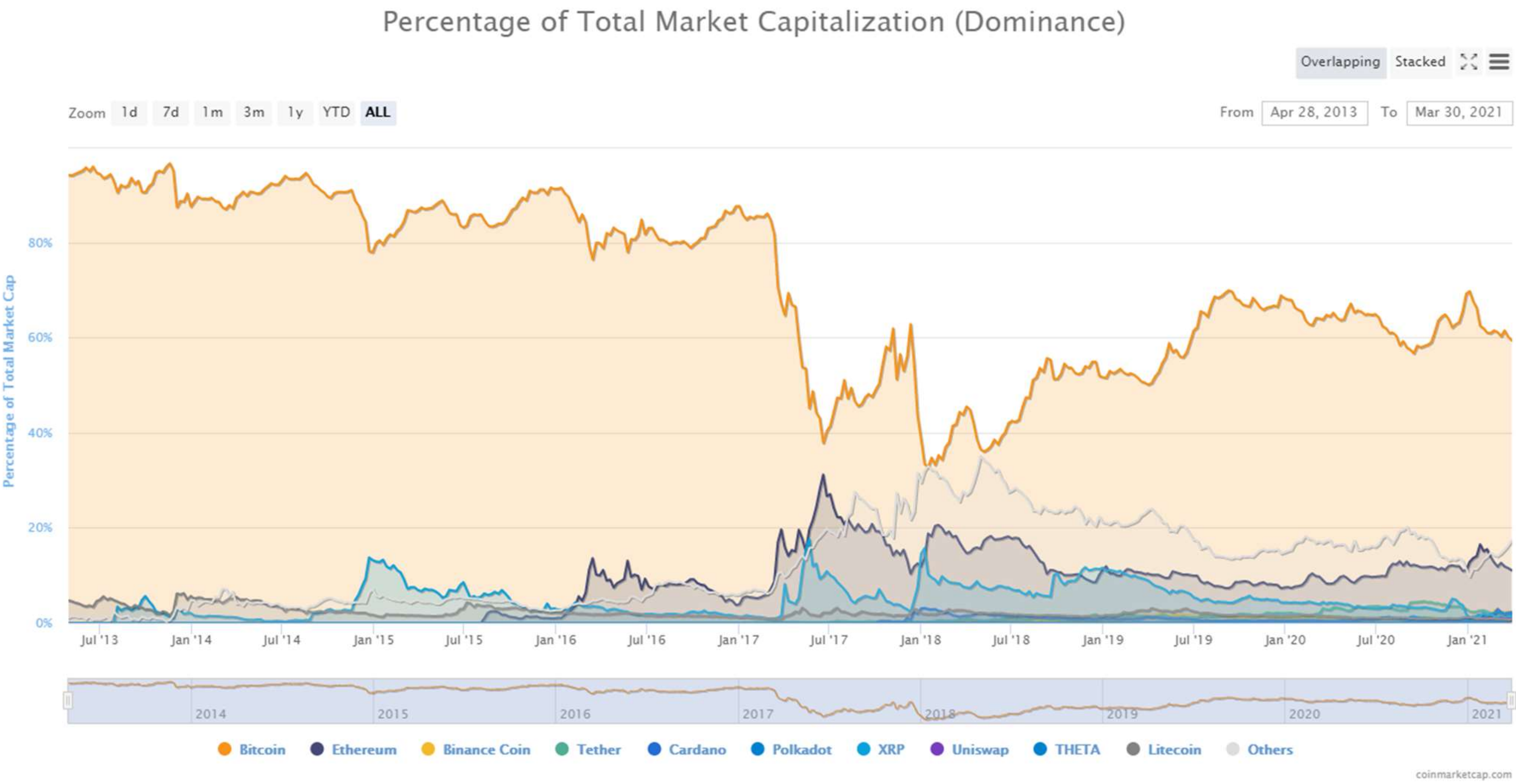
Fidelity DIGITAL ASSETS℠

- Portfolio managers are always searching for ways to reduce their portfolio risk. They typically do this by investing in “uncorrelated assets” which reduce volatility, specifically downwards volatility.
- Bitcoin is the best performing asset in the last 10 years. It has an average annualized performance of 230% per year.
- With a small Bitcoin allocation (e.g. 1%-5%), you expose yourself to a small capped downside for an asset which historically does extremely well in the long-term (4+ years).

Bitcoin as the reserve currency for crypto

- Bitcoin has historically outperformed all other cryptocurrency projects combined.
- On the right is a chart with non-Bitcoin cryptocurrency performance relative to Bitcoin since launch for several hundred major projects. The trend is to lose money in BTC terms (even if there is appreciation in dollar terms).
- For this reason, Bitcoin is the standard that cryptocurrency performance is measured against.





From Apr 28, 2013 To Mar 30, 2021

Part 4

Miscellaneous topics

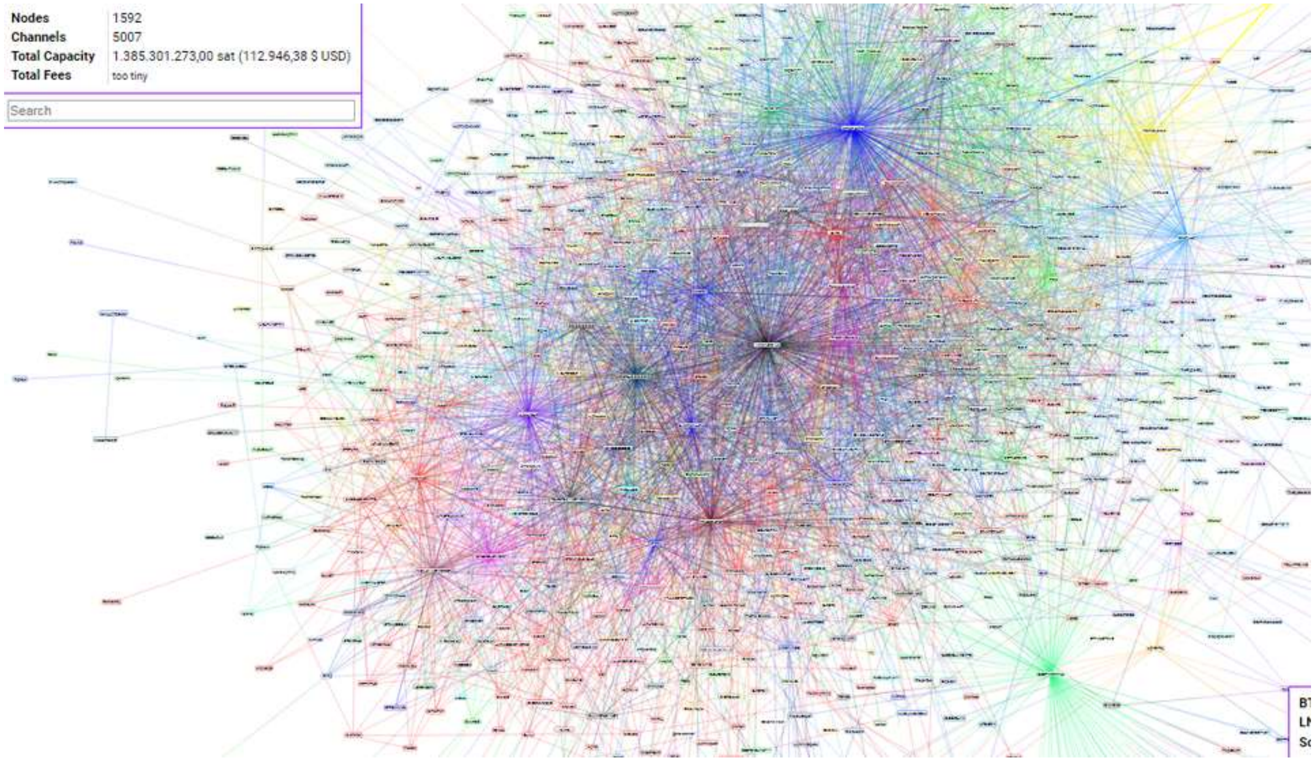
Bitcoin scalability

- Bitcoin blocks are at most 1 MB large* (actually it's a little more complicated), and for the avg tx size of 250B and 10m block time, that's 7 tx per second max.
- Increasing the block size will increase throughput but it will increase storage requirements / CPU requirements of nodes. Bitcoin's decentralization is correlated with the full node count. More throughput = less decentralization.
- Bitcoin will scale with higher layers. Namely sidechains and Lightning. Sidechain = another database / cryptocurrency (e.g. Coinbase, Ethereum). Lighting is on the next slide.

Lightning network (state channel)

- Bob and Alice with to open a “channel”. They send M and N Bitcoins to an address A and construct a “refund transaction” sending their coins back to themselves. They each posses this refund transaction. This is opening a channel.
- Bob and Alice can create new refund transactions together to reflect payments to each other (e.g. refund (M + e => Bob, M – e => Alice)) and each update gets a new sequence number.
- When either participant wants to close the channel, either can upload the latest transaction to the blockchain.
- If one participant cheats (e.g. trying to upload a transaction that is not the latest), there is a challenge period where the other participant can upload a more recent transaction and claim all the funds in the channel.
- A lightning channel can handle any amount of throughput without burdening the Bitcoin blockchain.
- You can route payments transitively (Alice -> Bob, then Bob -> Charles), much like how internet traffic is routed.
- More here: http://lightning.network/lightning-network-paper.pdf

Lightning network size in 2019



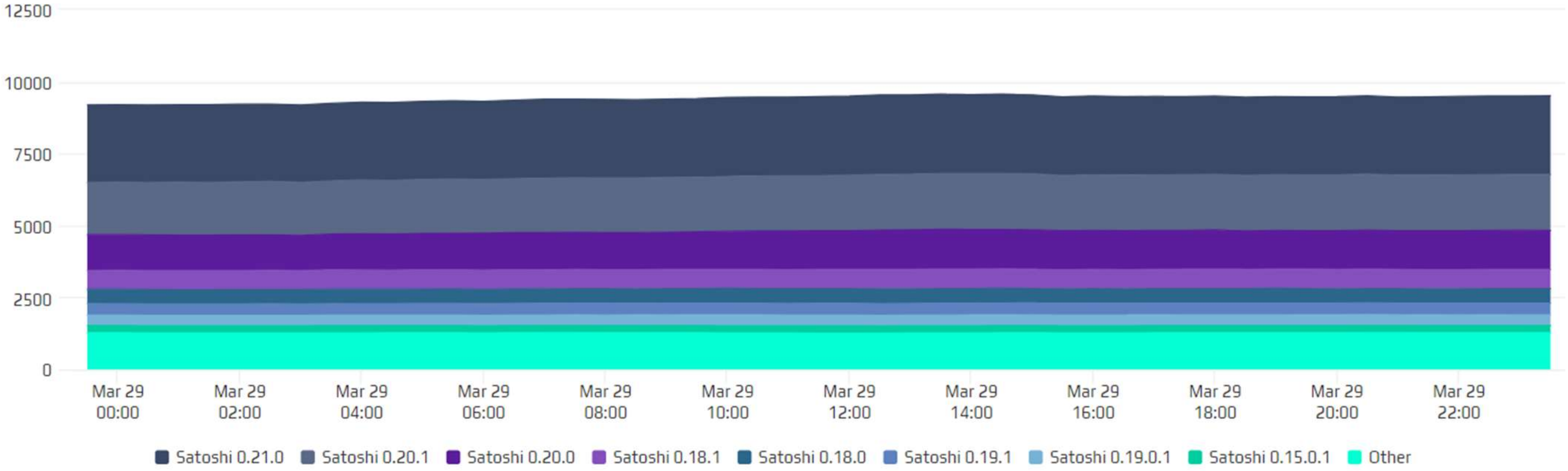
Nodes	1592
Channels	5007
Total Capacity	1.385.301.273,00 sat (112.946,38 $ USD)
Total Fees	too tiny

Network upgrades

- Among the 10k nodes running Bitcoin client software (99% use Bitcoin core), they are each on their own versions. Upgrading everyone together is not feasible and this becomes harder as more nodes come online.
- Bitcoin is extremely hard to change because of this inertia. If changes are not backwards compatible, there risks being two separate networks (and by extension separate authoritative blockchains).
- Hard fork = not backwards compatible. Old and new nodes will consider different chains authoritative based on validation rules. They will form separate networks. A critical bug fix to consensus rules may be a hard fork.
- Soft fork = backwards compatible. Old and new nodes will have the same authoritative chain based on their validation rules. This is the safest way to upgrade.

USER AGENTS

Chart shows the distribution of reachable nodes across leading user agents. Series can be enabled or disabled from the legend to view the chart for specific user agents.



Innovations in the cryptoasset space

- Turing complete programmability (e.g. Ethereum)
- smart contracts -- mini protocols
  - Execute code on the blockchain for complex information processing
  - What if you could send money to websites?
- Automated market makers (AMM)
  - 30bip spread vs 70-150bip at exchanges.
- Flash loans
  - Borrow tens of millions of dollars for a few seconds (usually for an arbitrage opportunity)
- zero-knowledge cryptography
  - Prove a statement without revealing any information (e.g. prove you're old enough to drink without revealing your age)
- Overcollateralized loans
  - Borrow money regardless of your credit score
- Layer 2 scaling solutions (Lightning, Plasma, rollups)
- yield farming / liquidity mining
  - bootstrapping financial protocols by rewarding liquidity providers
- Handshake
  - decentralized dns
- filecoin/sia
  - decentralized storage

The end

Further reading

- Bitcoin whitepaper
  - Short read (9 pages) of the original Bitcoin design
- The bullish case for Bitcoin
  - How Bitcoin fits into the origins and evolution of money
- Visions of Bitcoin
  - Changing narratives of what people think Bitcoin is for
- Bitcoin educational resources
  - A directory for almost everything you might care for to learn more about Bitcoin.