Just NOT a investment guide

Lecture 12: Introduction to Blockchain

—

MFE FE8828
2020/10/22
Dr. Yang Ye

Overview

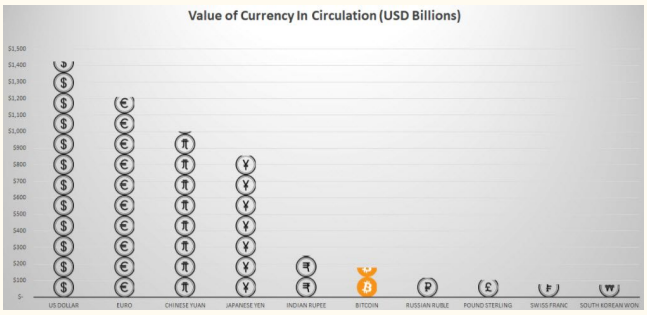
- Introduction
- Part 1: Blockchain
  - Block
  - Hash
  - Merkle Tree
- Part 2: Consensus
  - Nakamoto consensus, “Proof-of-Work”
  - Mining
- Part 3: Transaction
  - Public-key encryption
  - Transaction booking method
  - Transaction flow

B$: Bitcoin, about 35th largest currency in the world. https://fiatmarketcap.com/

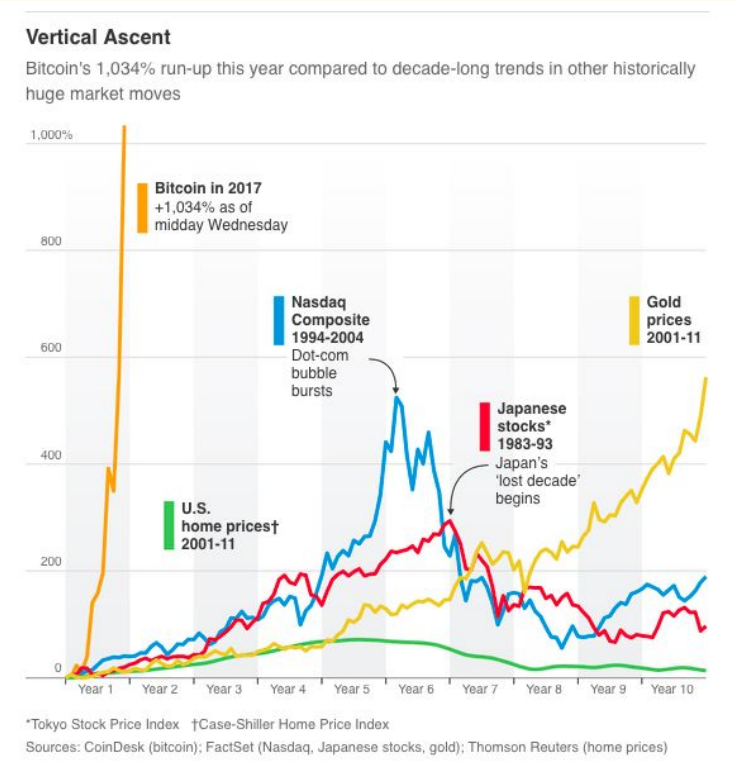
Bitcoin is the largest blockchain-based digital asset, with a market capitalization of $173.5 billion as of June 2020.

About 18mio as of Oct-2020, capped at 21mio in total number https://www.buybitcoinworldwide.com/how-many-bitcoins-are-there/

At the peak of 2017 (nearly $20,000), it was 6th largest most circulated currency, after Rupee, Yen, Yuan, Euro, and Dollar (1st).



As of 2017



As of 2017

Bitcoin had, by all accounts, a remarkably volatile week, losing $3 bln in market cap in just 90 minutes as the price slid from $11,400 to close to $9,000 (on some exchanges it flash-crashed to the low $8,000s). Nevertheless, within 36 hours, the cryptocurrency has rebounded to over $11,000.

The Economist

Why Bitcoin has been such a boom and bust?

- Bitcoin has many "firsts" and started new technology paradigm.
  - **Bitcoin** is a peer-to-peer electronic cash system, the first and also working blockchain-based digital asset. From Bitcoin, many **Cryptocurrency** have been created.
  - **Blockchain** is the core data structure that makes cryptocurrency working. The technology around blockchain is called **distributed ledger**. It's a data storage/updating/computing technology for the insecure and distributed Internet environment.

What's Blockchain?

Q: How to explain blockchain in simple words?

A: "4-in-1"

a. It's data structure, and (direct meaning of blockchain)

b. It's consensus, and

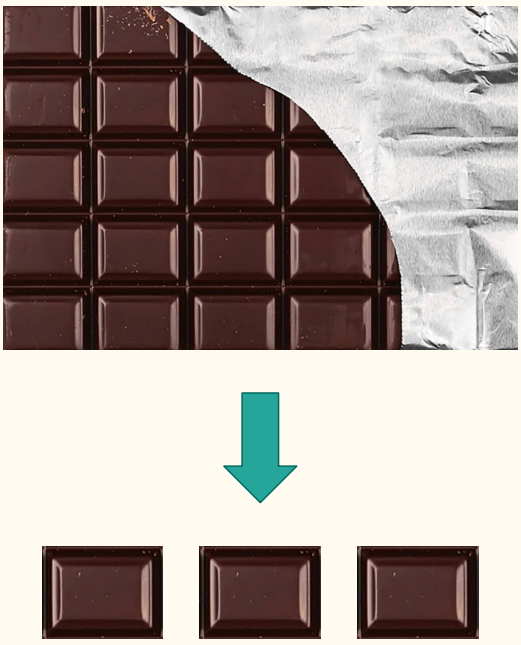
c. It's smart contract, and

d. It's secured computing.

Part 1: Blockchain

Data structure

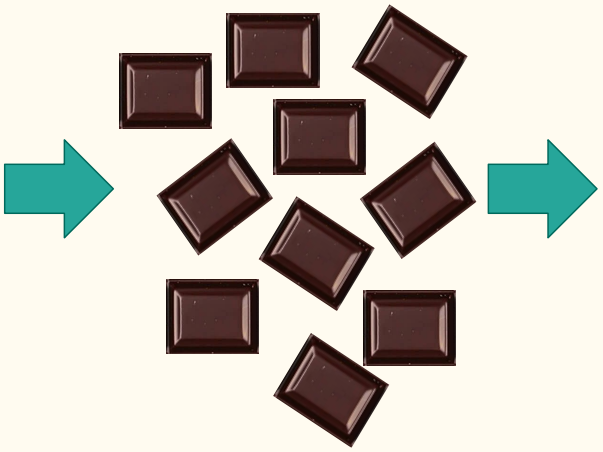
- Inside computer memory, we design certain ways to store and retrieve data.
- They are called **data structure**.
- The simplest data structure is an **array**.
- **Element** are inside array.



Distribute the Array (Chocolate or Money)

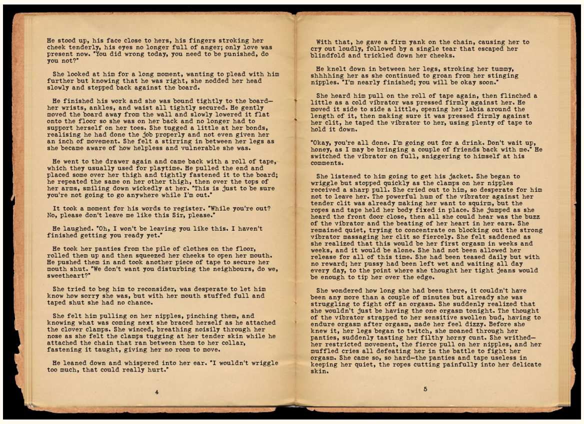
Creation

Distribution



Collection

Solution: Labelled Array like Book pages



It’s not good enough

Page number is not good enough.

Watermark/special color of paper/... are ways to ensure we get back what we distributed earlier.

The digital watermark is from cryptography:

Cryptographic Hash

In short, **“Hash”**

WEST AFRICAN ART

INTRODUCTION

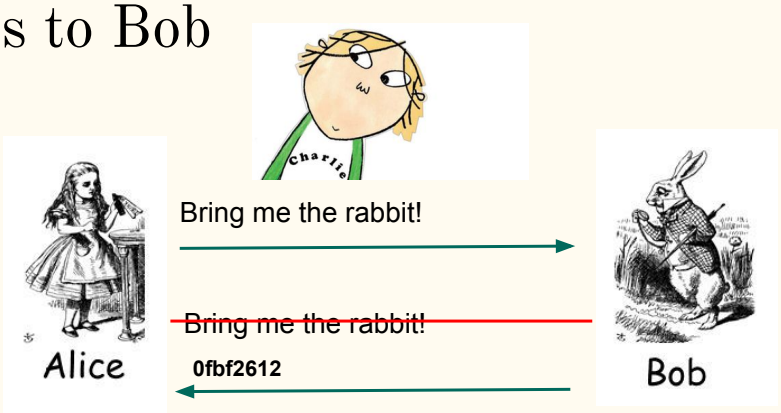
Solution: Cryptographic Hash

Cryptographic hash was about to solve the problem of data verification.

Hash can map data of arbitrary size to data of fixed size. The fixed size is much shorter than the original data.

256bit hash = 2 power 256 different content

Alice talks to Bob



Both Alice and Bob know a certain hash algorithm to compute a hash from a message.

Bob doesn't need to send the original message ("Bring me the rabbit") but just the hash ("0fbf2612").
So Alice can know that Bob has received her message.

hash("bring me the rabbit") = "0fbf2612".
Alice needs to know what cryptographic hash that Bob used.

Cryptographic Hash

Cryptography has created a number of very good hash functions. They satisify following conditions:

- We can't easily find another data with the same hash
- A small change in the original data would result in another hash and such change is not predictable.
- “Bring me the rabbit” => “Bring me the Rabbit”
- Brute-force, “rainbow attack”.

2^{95}	=	39,614,081,257,132,168,796,771,975,168

Bitcoin uses “double SHA-256”

SHA-256 sounds “Sharp-256”

It’s 256 bits of data to represent data. Double means it hashes twice.

- Long and hard to crack.
- Within current limit, they are safe to use.
- Quantum computing may crack it.

Some cryptography

- **Hash function**: one-way encryption. Difficult to get original data.

Hash function is to generate an abstract/digest of information.

R:
install.packages("openssl")
library(openssl)
sha256("123")
```
# a665a45920422f9d417e4867efdc4fb8a04a1f3fff1fa07e998e86f7f7a27ae3
```
sha256("124")
```
# 6affdae3b3c1aa6aa7689e9b6a7b3225a636aa1ac0025f490cca1285ceaf1487
```


Double-Sharp functions

sha256(sha256("123"))
```
"173af653133d964edfc16cafe0aba33c8f500a07f3ba3f81943916910c257705"
```

sha256(sha256("124"))
```
"0c60bfa8e6f9f5d9e86c72832c1936ce551c8f47adb61ba949eb638668c0205a"
```

Single-sharp is difficult to find another content with the same hash ("collision"). After double-sharp function, it has become difficult to find a collision.

Demo

install.R

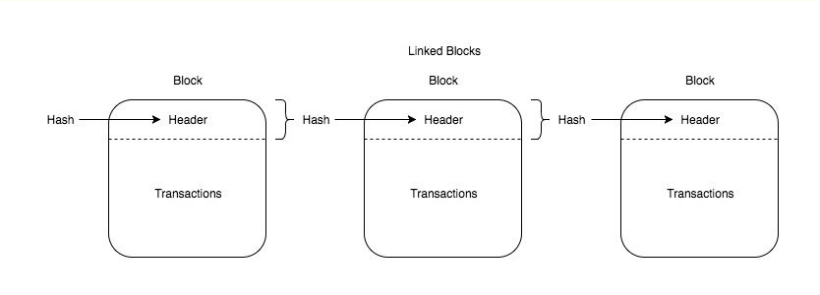
hash.R

Page with number => Block with Hash

So, if we generate **a hash** the content of a page and put the **hash value** on **the next page**. In this way, we can determine whether a page is the original page by checking the hash on its next page.

This is **blockchain**.

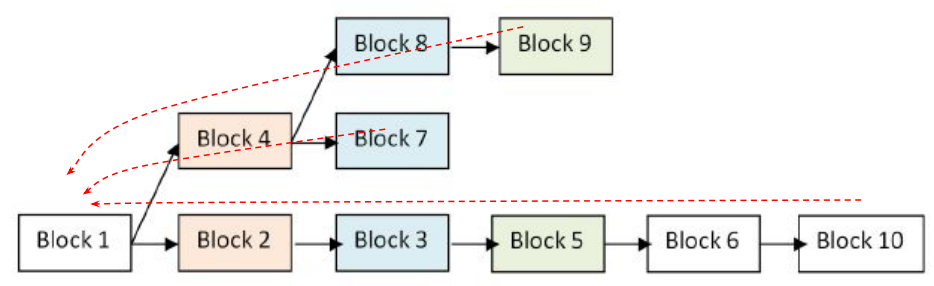
"Chain of blocks"



The features of Blockchain

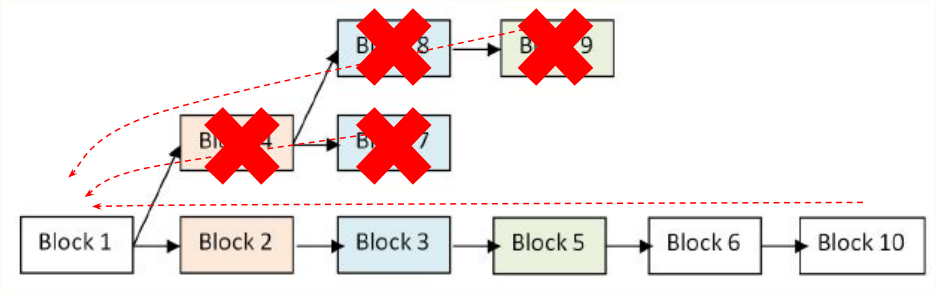
- Blocks can be **distributed stored**, old data **can not be modified**, and, more importantly, chain **can grow.**
- There could be only one chain if we trace **back** from Block N to Block 0.

Only one way to trace back

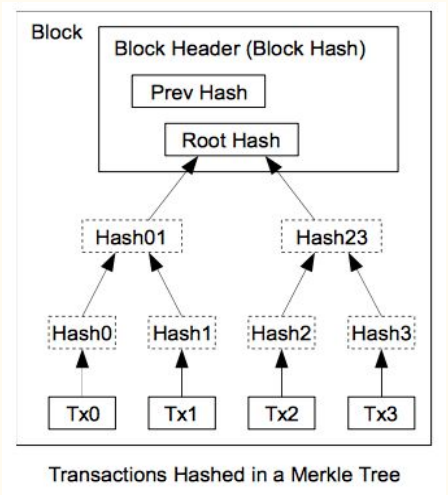


Only one way to trace back: Pruning

- Sub-branches can exist due to
  - Fraud attack, fake data
  - Distributed un-synchronized data
  - Un-agreeable trades
- The rule is simple, only the longest will survive.
  - One branch to exist, other pruned.
- Blockchain is designed to survive in the Internet jungle.

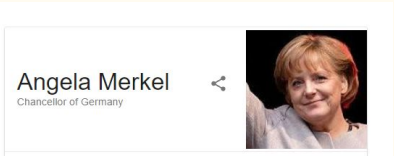


Merkle Tree: Store data in Blockchain

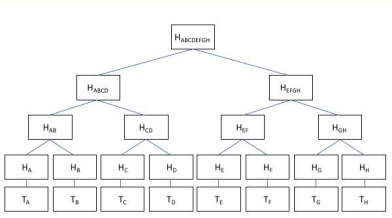


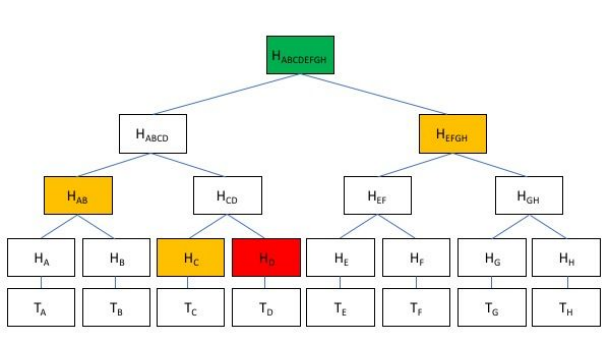
Transactions Hashed in a Merkle Tree

Merkle not Merkel



Transactions stored in Merkle Tree





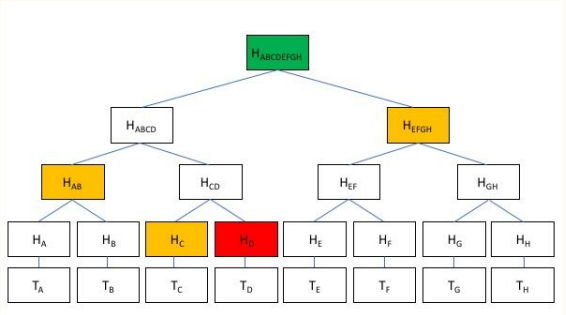
Merkle Tree is to combine all transactions into one hash.

Grouped Hash

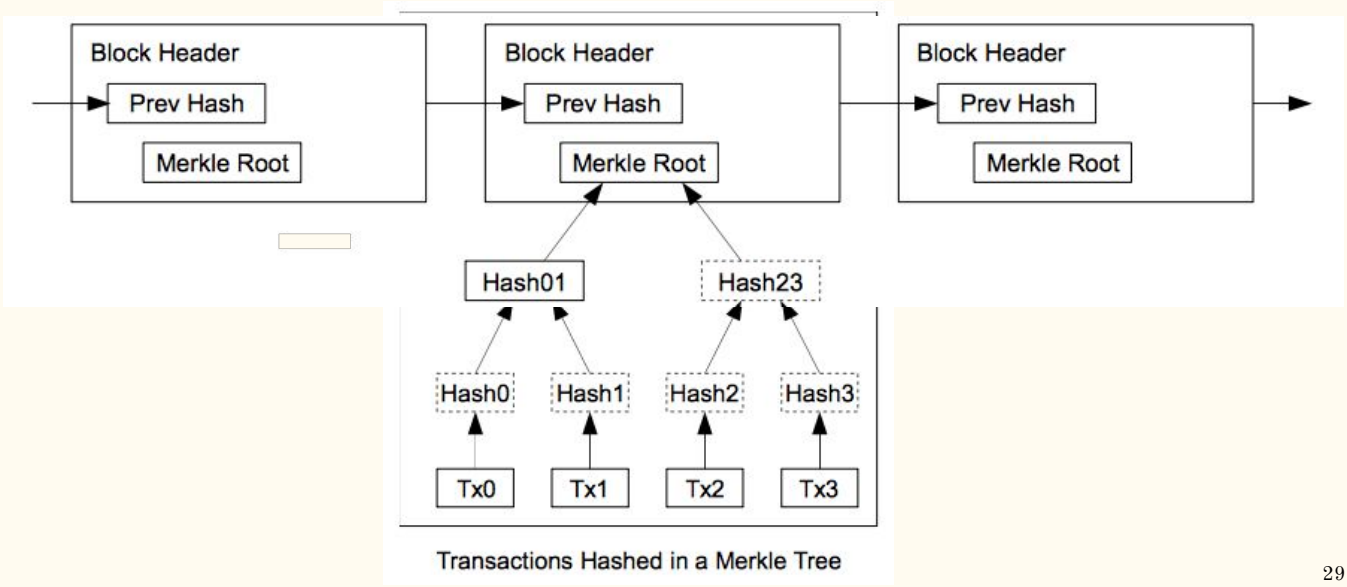
Merkle tree is a grouped hash of data. Binary tree is combine two-record per hash. One Merkle tree contains N data and 1 + 2 + 4 + ... N / 2 hashes. Merkle tree provides simplified verification, log2(N) for N transactions.

In-and-out block, there are hashes. Hash offers verification and also a space-saving. Full Bitcoin blockchain is close to 200G till now but hashes only takes less than 100MB.

HAB = hash(HA + HB)
HA = 0fbf2912
HB = 0fab1259
HA + HB = 0fbf2912ab1259
2 power 10 = 1024



Blockchain: a complete picture



Blockchain: a complete picture

- Block 0: transactions + Timestamp
- Block 1: hash of Block 0 + transactions + Timestamp
- Block 2: hash of Block 1 + transactions + Timestamp
- Block 3: hash of Block 2 + transactions + Timestamp
- ...

Timestamp adds an extra layer of verification. Timestamp must be increasing together with the block number. In Blockchain, hash is performed on data, the previous hash and a timestamp.

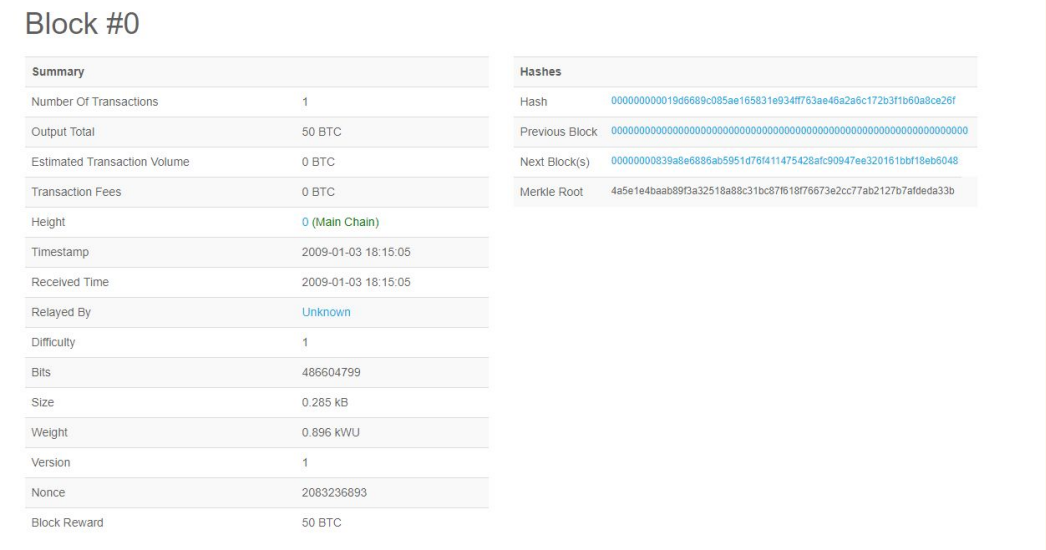
Blockchain’s structure

- One block has two parts: body and header.
- Header contains the hash of the previous block, the timestamp and the index.
- Body is a merkle tree, Merkle tree is a grouped hash of data. Binary tree is combine two-record per hash.
- Blockchain is blocks linked by Hashes.

Demo

merkel_tree.R

Genesis Block - https://goo.gl/oXh38t



Block #0

Summary	
Number Of Transactions	1
Output Total	50 BTC
Estimated Transaction Volume	0 BTC
Transaction Fees	0 BTC
Height	0 (Main Chain)
Timestamp	2009-01-03 18:15:05
Received Time	2009-01-03 18:15:05
Relayed By	Unknown
Difficulty	1
Bits	486604799
Size	0.285 kB
Weight	0.896 kWU
Version	1
Nonce	2083236893
Block Reward	50 BTC

Hashes	
Hash	000000000019d6689c085ae165831e934ff763ae46a2a6c172b3f1b60a8ce26f
Previous Block	0000000000000000000000000000000000000000000000000000000000000000
Next Block(s)	00000000839a8e6886ab5951d76f411475428afc90947ee320161bbf18eb6048
Merkle Root	4a5e1e4baab89f3a32518a88c31bc87f618f76673e2cc77ab2127b7afdeda33b

Bitcoin's Blockchain

https://blockchain.info

#653300 - is the latest block of now (was 497959)
Hash:
00000000000000000001d4a8624b7d8f4c5288ac33c23ad29d7d8fbf822ade0f

I can store all hashes of bitcoin as 653300 * 256 bits = 20,905,600 Bytes = 19.93 MB.
And I can verify whether a block is in the blockchain easily.

Summary: Blockchain

A blockchain has three core distinguishing ideas:

1. *Everywhere the same*
2. *The record is permanent.*
3. *No one is in charge, everyone has a piece of it.*

The blockchain was designed so transactions are immutable. To traceback from one block to its root, there could be only one way back.

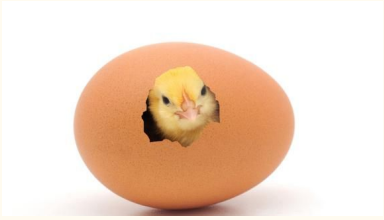
Yet, it is possible to rewrite history by throwing away certain blocks and restarting the chain from past block, it's called "Fork".

Summary: Blockchain

- It is a **continuously growing list of records**, called blocks, which are **linked** and **secured using cryptography hash**. ("Block")
- Each block typically contains **a cryptographic hash of the previous block**, a timestamp and transaction data. ("Chain") "append-only"
- By design, a blockchain is **inherently resistant** to modification of the data.

Who's first? Blockchain or Bitcoin

- Question: Blockchain and bitcoin, which one appeared the first?
  - You may answer Bitcoin appeared first.
- Answer:
  - First is the **block**, the "Genesis Block",
  - Second is the **coin** through **PoW,**
  - Third is the **Blockchain**.
  - This order is the key of understanding.



Consensus Algorithm

After talking about Blockchain and Merkle, what happens afterwards is the most innovative thing that Satoshi invented.

"Consensus algorithm", which is a combination of social-economics-science-engineering.

Blockchain Live!

- Blockchain lives in a distributed peer-to-peer (p2p) network of users. "client" software.
- The nodes on network will gossip with each other about what it knows about the blockchain, "who/when/what".
- It's a big network that each other is talking to each other about what's happening.

Consensus Algorithm

The blockchain network needs to achieve following goals.

- Crucially, participants agree on the dynamic content of the data i.e. nodes have a mechanism to resolve conflict ("**Safety**")
- Not any one entity controls the content of the data ("**distributed authority**").
- The data can move forward, ("**Liveness**")

Consensus algorithm is about to achieve these.

"Nakatomo" Consensus

The inventor of Bitcoin, Nakatomo Satoshi proposed the following process:

1. To create a new block with new transactions, Alice needs to first **solve a hard problem**. Once she works it out, she will be **rewarded**. She **broadcast** the answer to the network.
2. Bob/Charlie/Danny would **validate** the block that Alice **has really solved the problem** and check that the block is correctly constructed so there is no wrong records, like "double spending". Only valid block would be accepted and passed.
3. Because this is **the latest block**, "Bob/Charlie/Danny" will gossip to spread word. Everyone (adding /Eva/Frank/George/Hellen) start working on the problem for the next block.
4. Who's the lucky one this time? Go back to step 1 above.

Satoshi's Game with Hash

It needs to guess a number that adds to the current hash to produce a new hash with certain number of leading zeros.

hash("ABC" + XXXX) => 000000..... XXXX is the nonce.

1/16 * 1/16 * 1/16 => smaller the probability, the harder the problem.

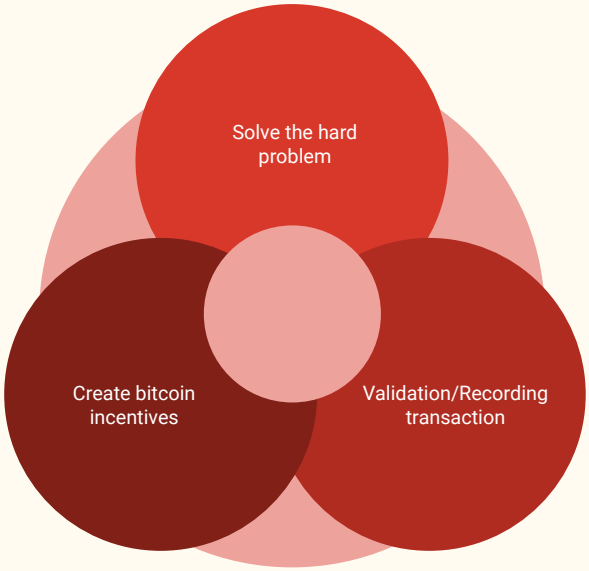
Because cryptographic hash has no easy guess between input and output. Node has to start from 1, try, 2, try, 3, try, till it reaches a number that satisfies the condition.

PoW: Proof of Work

Satoshi's PoW systems appears to kill four birds with one stone:

- Coin production = reward for solving the problem.
- Broadcast transaction and make it acceptable
- Grow the blockchain correctly: record new transaction. Chain selection, longest
- Who produces blocks (random), and When blocks are produced (fixed)

Consensus algorithm: One design that solves many



Why Science-Engineering?

The mechanics of the puzzle are fairly simple. Every miner has a block containing a list of transactions that should be published. The miner's goal is to find a value called a nonce, such that the hash of {nonce + block} is less than a target value.

The current target is 440 billion (at the time of writing). And because SHA-256 is a specially constructed cryptographic hash function, miners can't do better than guesswork when it comes to finding a nonce that wins the hash puzzle. So every nonce has a 440 billion / 2^256 = 3.7999142e-66 chance of winning the hash puzzle—meaning a Bernoulli trial.

Why this solution is also Social-Economical?

- Satoshi designs that bitcoin is issued to miner. If you can create a block, you earn bitcoin. This is the only way that bitcoin is issued.
- Once a new block is created, miner broadcasts to the network. All receivers will stop working on this block and switch to work on next block for new transactions.
- No one wants to lag behind. So only the longest chain will survive.

“Nakatomo” Consensus: Summary

- There is only one way to solve the problem: brute-force.
- There is incentives to **reward** the solution.
- Longer chain is always preferred to resolve conflicts.
  - Difficulty is adjusted according to the computational power of all participants. To keep only one solution every 10 minutes.
  - Who will get the solution? It depends on how much computational power you have and luck. Statistically, it rewards those with large computation power.

51% Attack and Byzantine Generals' Problem

Related problem, For bitcoin, there is some one controls 51% of the network's computing power (or close to it), someone can create chaos to the network. Alice has the power to rewrite the history and forge the records. This has not happened.

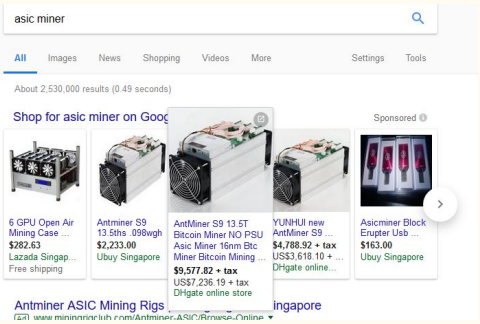
A more academic problem is the Byzantine Generals' Problem is about how to tolerate “traitors” among all generals. Mathematically, more than two-thirds of nodes need to be faithful (Lamport 1982). This requires higher level of loyalty from the participants.

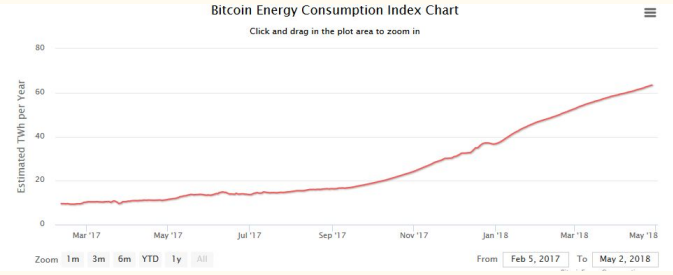
Mining

- Therefore, solving the problem has been specialized into “mining”. Miners are equipped high-performance computing power and low-cost electricity.
- Mining gear has progressed from CPU, to GPU, to special circuit, ASIC.
- Hashrate is 30PH/s= 30 x 10^{15}. Energy per year is over 63 TWh, between Chile (17mil pop.) and Austria (8.7mil pop.).
- **Mining** is the process of creating new coins in exchange for validating the ledgers and verifying their accuracy.

Mining gear

- TSMC's big client is mining chip design company
- Increasing use of special hardware.
- Energy consumption

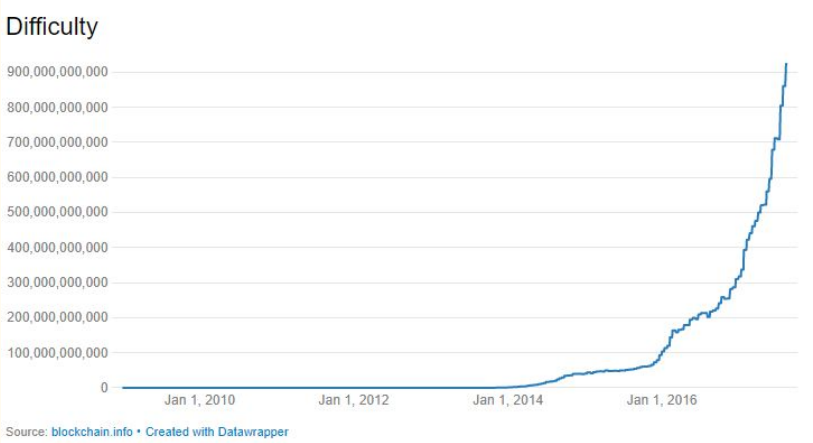




Participants in Cryptocurrencies

- Miners (one with huge computing power): Earn the transaction fees from low-cost electricity.
- Trader: Price
- Users/Developer: Application

Mining Difficulty



Energy Consumption of Bitcoin network

Source: https://digiconomist.net/bitcoin-energy-consumption

Each Bitcoin transaction uses 271kWh of electricity—enough to power a typical American home for nine days.

What does cause Bitcoin's energy usage to rise however, is when Bitcoin's price goes up. A higher price means the 12.5 bitcoin reward becomes more valuable, and so miners spend more resources to capture the larger prize.

Per-block reward has fallen twice—it started out at 50 bitcoins in 2009—and is scheduled to fall to 6.25 bitcoins per block some time in 2020, then to 3.125 bitcoins per block around 2024. I hope, as the per-block reward falls, the network's energy consumption will fall proportionately.

Demo

pow.R

Transaction

- In your normal routine,
  - Take out your wallet from pocket or bag
  - Take out cash
  - Pay
  - Online banking/mobile payment is similar mechanism.
- You own your "wallet" (virtual or real). If anyone gets access to your wallet, they get access to your cash.

The infrastructure of Cryptocurrency

There are five parts in the infrastructure:

- Public record - blockchain-based
- Decentralized consensus algorithm
- Economic incentives
- Distributed p2p network
- Transaction authentication and implementation

Transaction on Bitcoin network

- Your wallet is public as record stored the blockchain.
- Placing the wallet on the table.
- Everyone can see it but only authorized person can access it.

Transactions	Your Balance
You received 100 from Alice.	100 = 100 + 0
You paid 50 to Bob.	50 = 100 - 50

Crypto's Transaction Design

Blockchain needs to take in instructions from outside , verify, validate and execute the transaction.

- Verify: it's authentic from the owner
- Validate: it's valid as a transaction: no over-spending beyond owned.

Crypto Transaction Process

Like a cheque-system but runs with encryption

1. Draft the cheque

2. Sign on the check

3. Post the cheque

4. Network (Bank) receives the cheque, verifies the signature and balance, verifies the payee and does the transfer.

Building Blocks for Transaction

Cryptocurrency builds a network protocol to take on transaction process. The necessary building blocks to enable this are

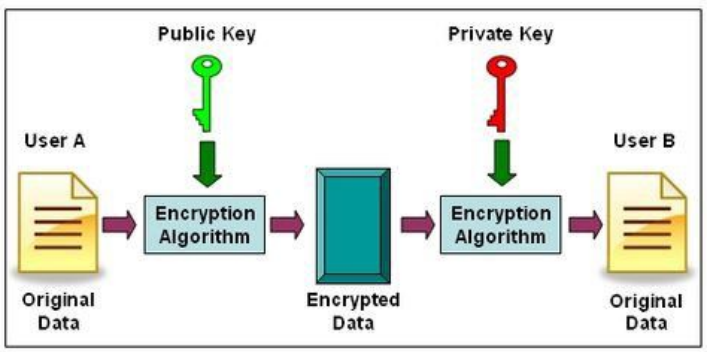
1. The public key-based cryptography, which provides address, signature and verification process.
2. Transaction booking method.

Public-Key Cryptography

- Public-key encryption: to ensure a message comes from a user by.
- Asymmetric cryptography: public key and private key are a pair. But it's one-direction (asymmetry), private key can verify a paired public key but public key can't deduce what its paired private key is.

Public and Private Key in action

Public key can encrypt. **Private key** can validate and decrypt. If validation fails, decryption also fails.
Alice use Bob's public key to encrypt. Only Bob can decrypt with Bob's private key.



Asymmetric encryption - 1

There are some mathematics problems are not easily reversible. Such as **big integer factoring.**

https://en.wikipedia.org/wiki/Integer_factorization

How to use **biginteger** to build a public key/private key pair:

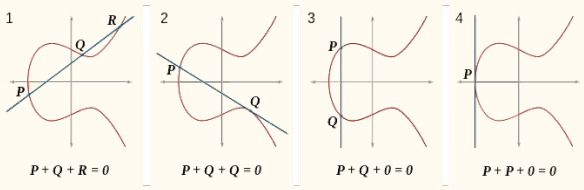
$$p * q = R, \text{ R is a very big integer}$$

R is my public key to be shared to everybody. p is my private key. Someone can use my public key to encrypt a message. Only I can decrypt it.

Asymmetric encryption - 2

When the numbers are sufficiently large, no efficient, non-quantum integer factorization algorithm is known. An effort by several researchers, concluded in 2009, to factor a 232-digit number (RSA-768) utilizing hundreds of machines took two years and the researchers estimated that a 1024-bit RSA modulus would take about a thousand times as long.[1] However, it has not been proven that no efficient algorithm exists.

Elliptic Curves is the current standard, due to space-efficient.



Demo

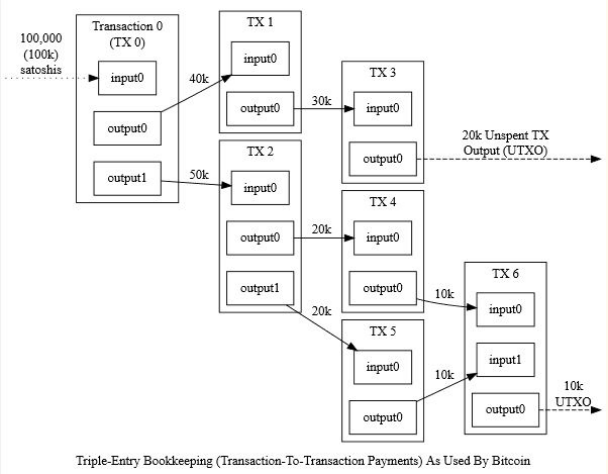
public_key.R

transaction.R

Transaction booking

In bitcoin, a transaction is a list of inputs and outputs, with each input pointing to the output of an earlier bitcoin transaction.

Each output specifies the conditions that need to be satisfied in order to spend the coins in that output. The simplest transactions just require a digital signature—cryptographic proof that a transaction has been approved by the owner of a particular private key.



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

Bitcoin transaction summary

- Each transaction has at least one input and one output.
- One wallet can only be used once.
  - I have 10 BTC in my wallet “IronWolf” and I want to give “Baracuda” wallet 2 BTC.
  - I need to create new wallet “SteelCat”.
  - The transaction is to "Take the 10 BTC from IronWolf, give 2 to Baracuda and give 8 to SteelCat".
  - Miner independently validates the transaction and record it in blockchain.
- Because each output of a particular transaction can only be spent once, the outputs of all transactions included in the blockchain can be categorized as either Unspent Transaction Outputs (UTXOs) or spent transaction outputs. For a payment to be valid, it must only use UTXOs as inputs

Summary: Transaction

- The transaction process of Bitcoin is a product of very careful thoughts of security and process.
- This is a particular model of transaction. Other cryptocurrency system may opt for other kinds of process.
- Transaction is a key mechanism for blockchain. Its function ensures a secure and speedy transaction.

http://fermatslibrary.com/s/bitcoin

Bitcoin: A Peer-to-Peer Electronic Cash System

Satoshi Nakamoto
satoshin@gmx.com
www.bitcoin.org

Abstract. A purely peer-to-peer version of electronic cash would allow online payments to be sent directly from one party to another without going through a financial institution. Digital signatures provide part of the solution, but the main benefits are lost if a trusted third party is still required to prevent double-spending. We propose a solution to the double-spending problem using a peer-to-peer network. The network timestamps transactions by hashing them into an ongoing chain of hash-based proof-of-work, forming a record that cannot be changed without redoing the proof-of-work. The longest chain not only serves as proof of the sequence of events witnessed, but proof that it came from the largest pool of CPU power. As long as a majority of CPU power is controlled by nodes that are not cooperating to attack the network, they'll generate the longest chain and outpace attackers. The network itself requires minimal structure. Messages are broadcast on a best effort basis, and nodes can leave and rejoin the network at will, accepting the longest proof-of-work chain as proof of what happened while they were gone.

Bitcoin as a descendant of technologies in the last 30 years.

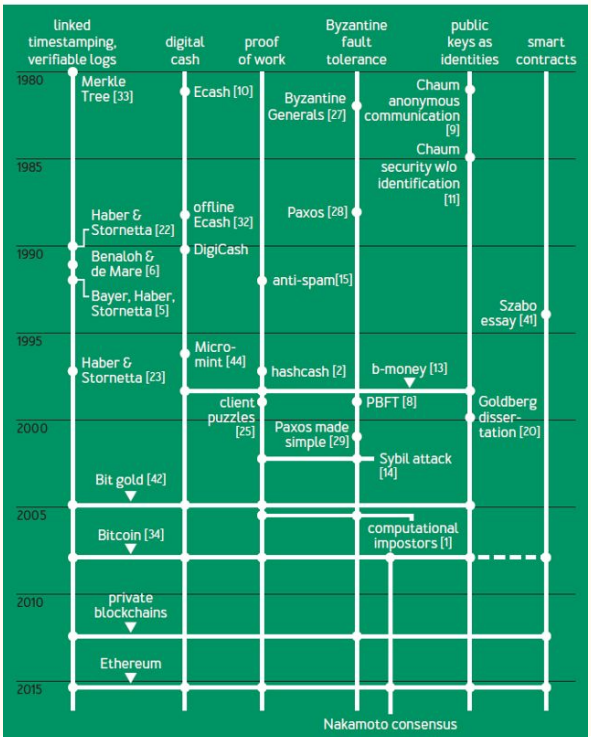
Bitcoin's Academic Pedigree

The concept of cryptocurrencies is built from forgotten ideas in research literature.

Arvind Narayanan and Jeremy Clark

http://queue.acm.org/detail.cfm?id=3136559

FIGURE 1: CHRONOLOGY OF KEY IDEAS FOUND IN BITCOIN



Bitcoin - the social experiment

- For Bitcoin, it happened several times that two solutions are found almost at the same time, almost half-half in the network. It survived.
- Once it was due to a software bug.
- Once it was due to half-user upgraded, half-user not upgraded.
- Bitcoin has survived all these tests.