On Security and Stability of Bitcoin Protocol

by Cyril Grunspan

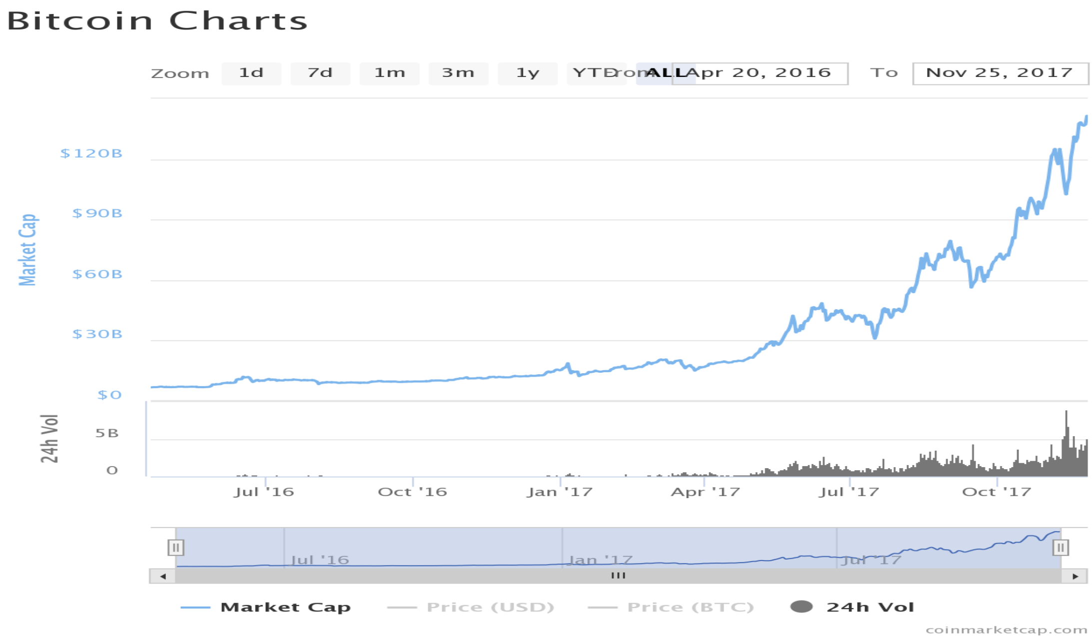
ESILV
Pôle Universitaire Léonard de Vinci
De Vinci Research Center
92400 Courbevoie

Labex Refi Fin'tech

30/11/2017

1 Introduction : a new economy

A new economy



Nov 23, 2016 751.74 / Nov 23, 2017 8232.38
+ 995% !
BTC Market capitalization : 140 Billions USD
BTC, ETH, BCH...
Exchange platforms: Kraken, Coinbase, Poloniex, Bitstamp, Bitfinex, *Okcoin...*
Total Market cap : 270 Billions USD
Blockchain startups
ICOs : Filecoin (257M), Tezos (232M), Bancor (153M)
Traditionnal raise funding
Blockstream, Coinbase, Blockchain, Consensys,Ledger

2 A new field of research

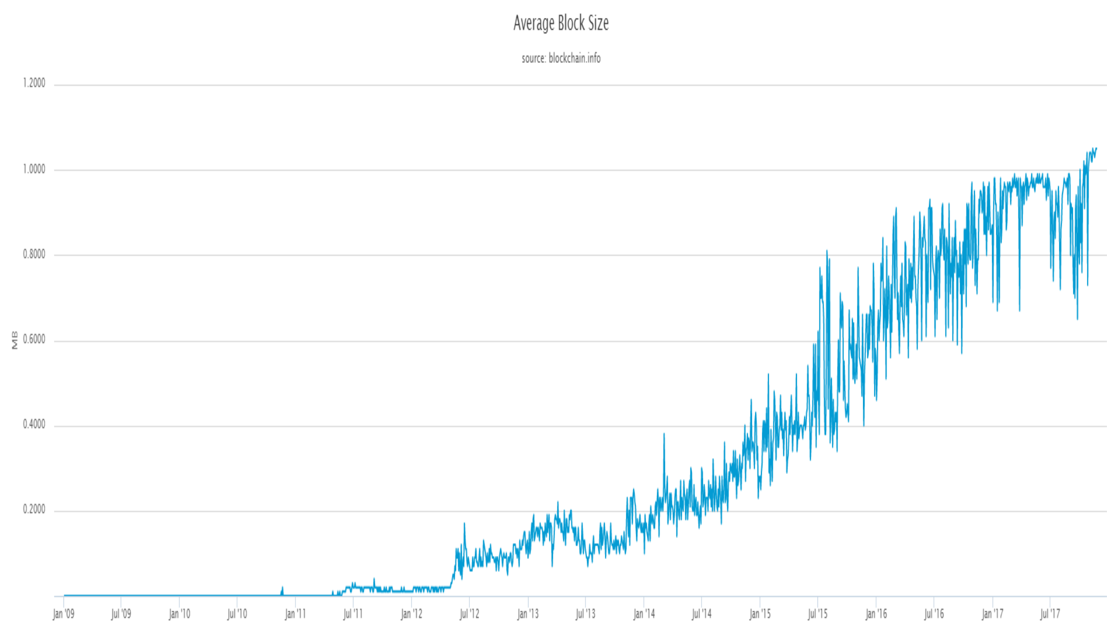
Mathematics behind bitcoin: Head and Tail
Massive use of SHA 256, hashing function
Interblock Time = exponential distribution
Blockchain Progress = Poisson Process
How to Gamble If You Must: Inequalities for Stochastic Processes, Lester E. Dubins, Leonard J. Savage (1965).

Link between network security and probability theory

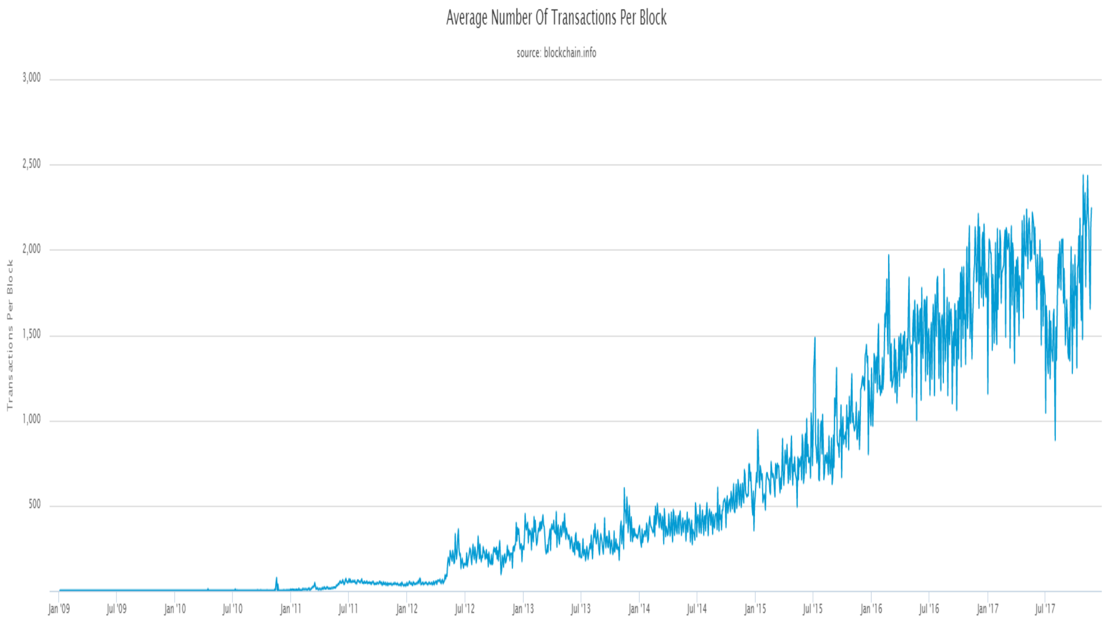
"bitcoin" on Google Scholar: >30 000 results. Bitcoin's whitepaper cited 2086 times.

Scalability

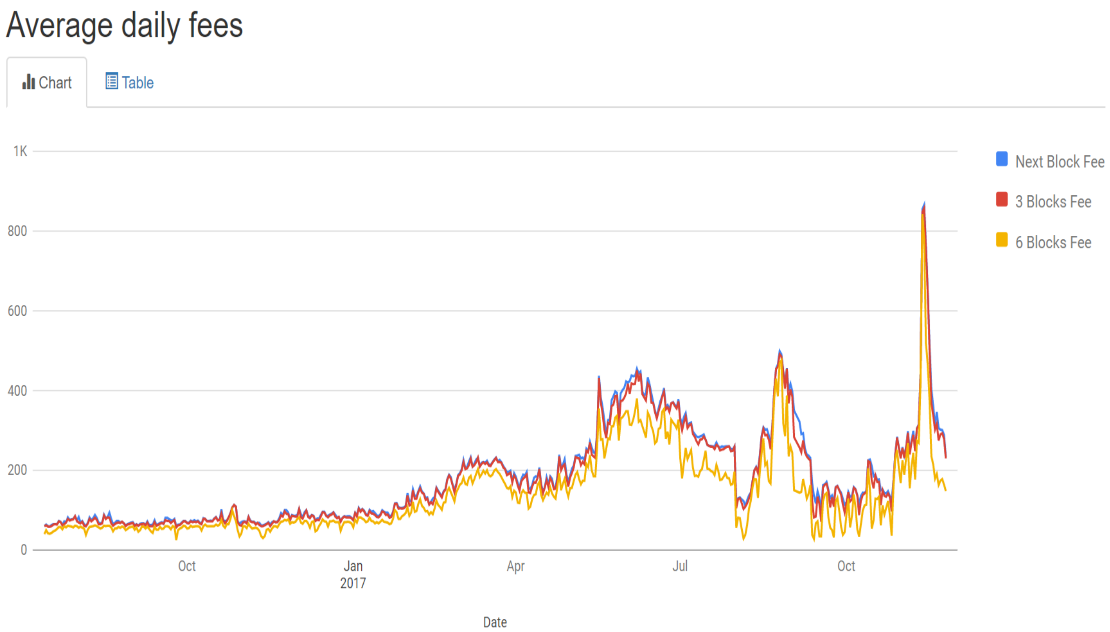
Slow
Blocks limited (1M)



Blocks saturated



Important fees 10% of the total reward of a block



Average bitcoin transaction = 250 bytes

Next Block Fee: fee to have your transaction mined on the next block (10 minutes).$4.96
https://bitcoinfees.info/

Anonymity

Privacy, Fungibility

Bitcoin pseudonymous
Tumblebit, Mimblewimble, Schnorr Signatures, Z-Cash, zk-Snark

Other protocols inside Bitcoin's world

Beyond blockchain
Lightning network, payment channels
Sidechains, Rootstock, Blockstream

Other protocols outside Bitcoin

More or less sophisticated:
GHOST, Ethereum
Tangle, IOTA
Proof of Stake, Proof of ???

[1]. W.Dai,"b-money," http://www.weidai.com/bmoney.txt, 1998.

[2]. H. Massias, X.S. Avila, and J.-J. Quisquater, "Design of a secure timestamping service with minimal trust requirements," In 20th Symposium on Information Theory in the Benelux, May 1999

[3]. S. Haber, W.S. Stornetta, "How to time-stamp a digital document," In Journal of Cryptology, vol 3, no 2, pages 99-111, 1991.

[4]. D. Bayer, S. Haber, W.S. Stornetta, "Improving the efficiency and reliability of digital time-stamping," In Sequences II: Methods in Communication, Security and Computer Science, pages 329-334, 1993.

[5]. S. Haber, W.S. Stornetta, "Secure names for bit-strings," In Proceedings of the 4th ACM Conference on Computer and Communications Security, pages 28-35, April 1997.

[6]. A. Back, "Hashcash - a denial of service counter-measure," http://www.hashcash.org/papers/hashcash.pdf, 2002.

[7]. R.C. Merkle, "Protocols for public key cryptosystems," In Proc. 1980 Symposium on Security and Privacy, IEEE Computer Society, pages 122-133, April 1980.

[8]. W. Feller, "An introduction to probability theory and its applications," 1957.

3 Two groundbreaking ideas

3.1 New Framework for the design of a transaction

Adress: to receive funds
asymptotic cryptography ECDSA
Transaction Output: number of bitcoins and spending condition
Single/multi signature, Locktime, solution of a cryptographic puzzle...
Output Spent/Unspent
Transaction Input: reference to an output and arguments
Transaction = Input/Output $\longrightarrow\square\longrightarrow$
ScriptSig + ScriptPubKey (Language = "Script")
Invention of smart contract (Nick Szabo)

3.2 Advance in distributed system theory

"Old" theory ignored by Satoshi
State machine replication
Leslie Lamport, 70
Fault-tolerant computer system
Nakamoto consensus
Proof of Work
Probability of success only
Byzantine Generals Problem

Bitcoin deserves to be studied rigorously by academics!

4 A short history of Bitcoin

19/08/2008. Satoshi reserves bitcoin.org

31/10/2008. First message on metzdowd.com (Bitcoin P2P e-cash paper)

08/01/2009. Invitation to download software

12/01/2009. First Bitcoin transaction, from Satoshi to Hal Finney (10BTC in block 170).

A Peer-to-Peer Electronic Cash System

In the coinbase parameter of the genesis block :

> The Times 03/Jan/2009 Chancellor on
> brink of second bailout for banks.

"Block chain" in two words by Hal Finney
Solve a problem long considered by cypherpunk movement
Bitcoin, money of a cypherworld?
Many failed attempts >100...
David Chaum (ecash)
Blind signatures for untraceable payments, Advances in Cryptology Proceedings of Crypto. 1982 (3): 199–203.

23/04/2011. Last message to Mike Hearn: "I've moved on to other things. It's in good hands with Gavin and everyone."

14/01/2016. *Lightning Network*, Joseph Poon, Thaddeus Dryja

5 Bitcoin

5.1 A peer-to-peer network

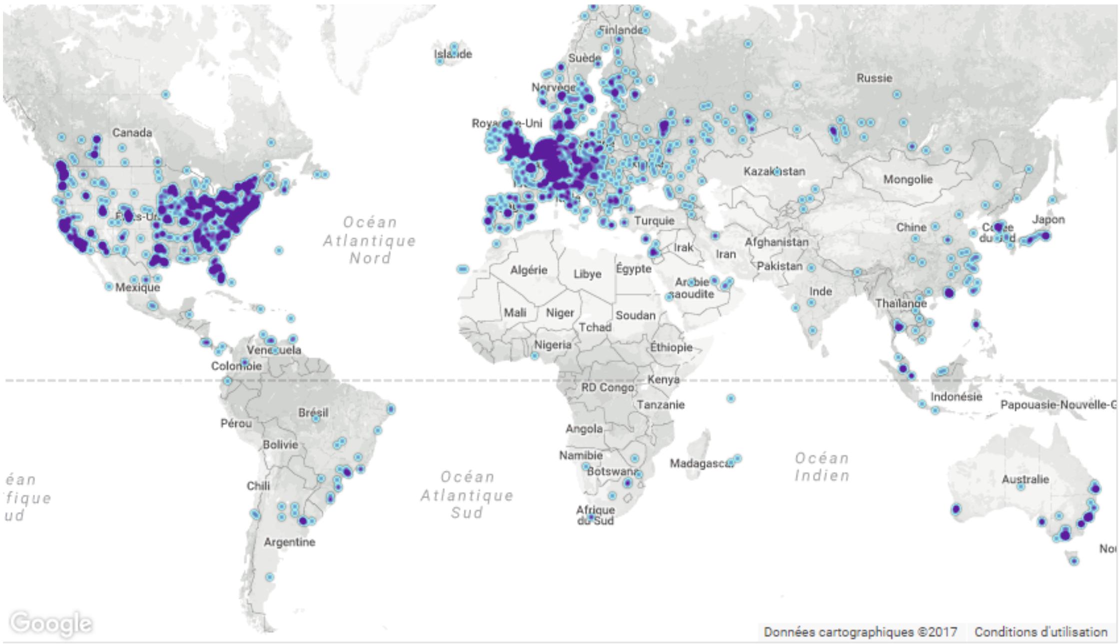
Nodes, Miners, Clients

A randomly connected network of a thousand nodes

All nodes perform the same operations

There is no central authority

No single point of failure



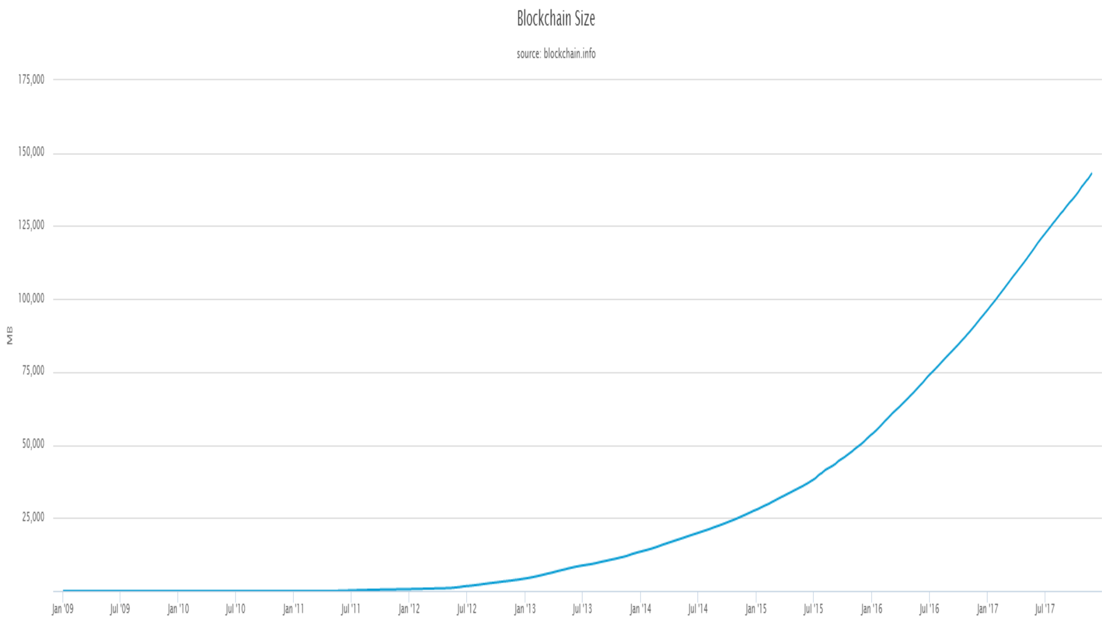
Top 10 countries with their respective number of reachable nodes (11031) are as follow.

RANK	COUNTRY	NODES
1	United States	3059 (27.75%)
2	Germany	1859 (16.86%)
3	France	756 (6.86%)
4	China	723 (6.56%)
5	Netherlands	524 (4.75%)
6	Canada	446 (4.05%)
7	United Kingdom	434 (3.94%)
8	n/a	382 (3.47%)
9	Russian Federation	362 (3.28%)
10	Singapore	219 (1.99%)

5.2 Full nodes

A full bitcoin node :

- keeps a local copy of the ledger
- validates (or not) incoming transactions
- validates (or not) incoming blocks
- forwards valid transactions
- forwards valid blocks



cost of storage (1.4GB)

processing power: 5ms per transaction (seek time)

bandwidth (10Mbits/s)

broadband Internet connection (50 kilobytes/s)

Running a node costs approximately 0.15USD/day

Signatures are based on ECDSA

Elliptic curve on Galois field $\mathbb{F}_p$ secp256k1,

$$y^2 = x^3 + 7$$

with

$$\begin{aligned} p &= 2^{256} - 2^{32} - 2^9 - 2^8 - 2^7 - 2^6 - 2^4 - 1 \\ &= 115792089237316195423570985008687907853269984665\backslash \\ &\quad 640564039457584007908834671663 \end{aligned}$$

Only used for Bitcoin? Gaining in popularity.

Elliptic curve useful for generating a finite group

Discrete logarithmic problem hard to solve

Base point G

Secret integer n

Public key $= n \cdot G$

Bitcoin Adress = SHA-256 ∘ RIPEMD160(PublicKey)

5.3 Lightweight Wallet

Lightweight wallets use a simplified payment verification (SPV) mode which only requires them to download part of the blockchain. They will connect to full node clients and use bloom filters to ensure that they only receive transactions which are necessary and relevant to their operation.

http://cryptorials.io

Do not receive transactions irrelevant to their operation

Do not need to perform validation transaction/block

5.4 Miners

Miners are particular full nodes

Miners can use "Proof-of-Work" to add new blocks to the blockchain

A miner:

- collects transactions
- creates a Merkle tree
- tries to solve a cryptographic puzzle "PoW"-type problem
- Builds and broadcasts new block

In every block, there is a special transaction called coinbase = 12.5BTC (halvening every ≈ 4 years).

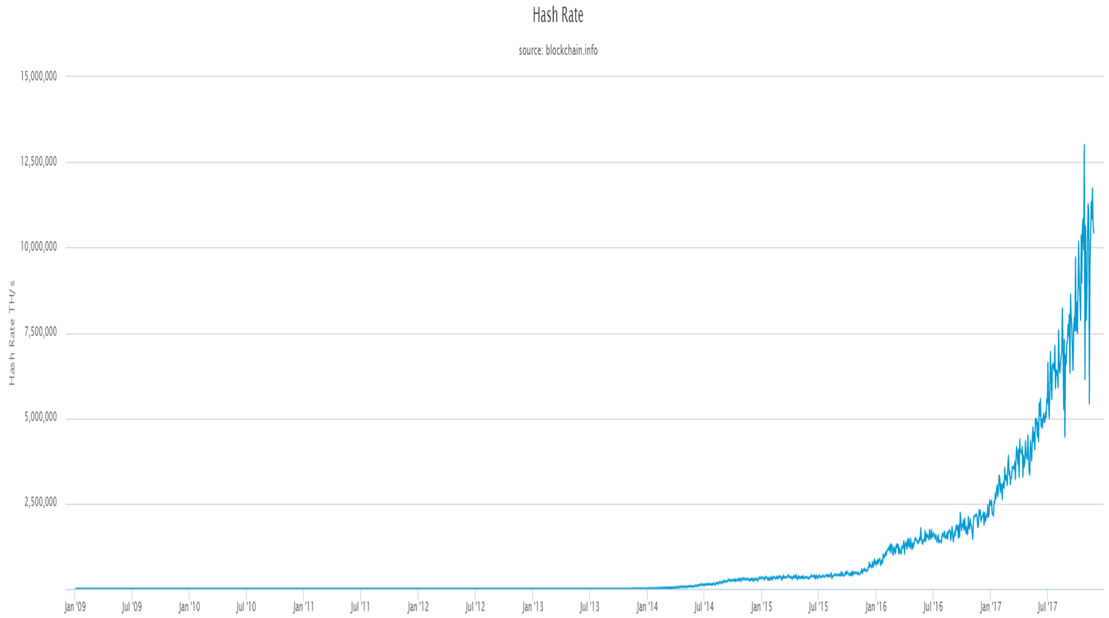
To be valid, a block must refer a previous block and contain a solution of a "Proof-of-Work" problem.

Monetary creation

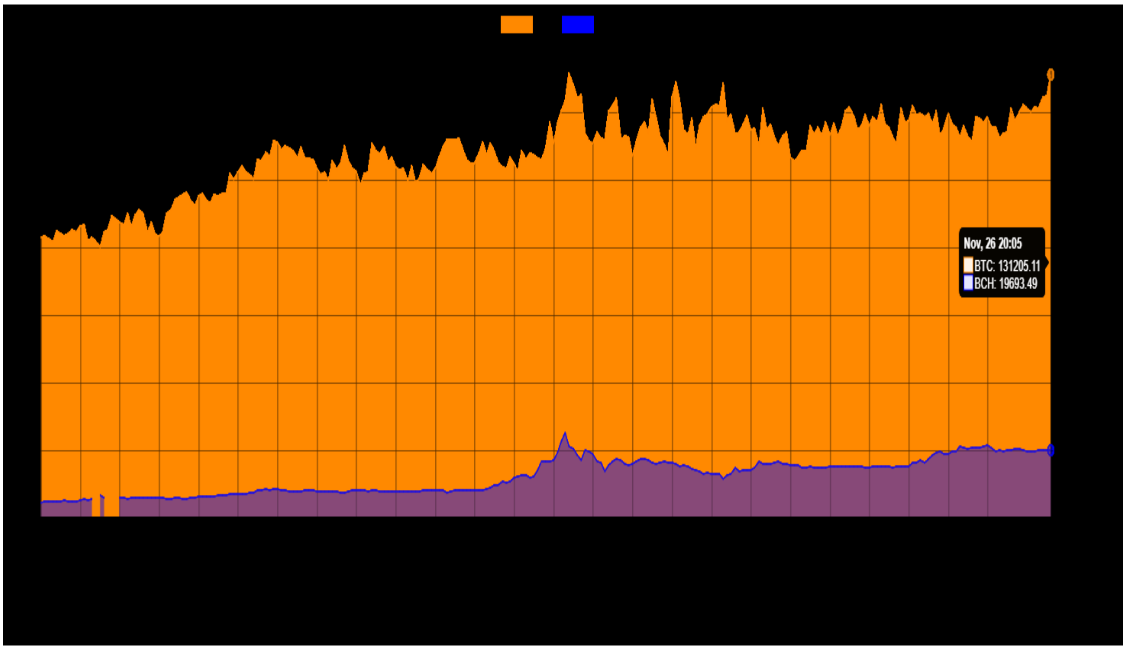
Limit of 21M BTC

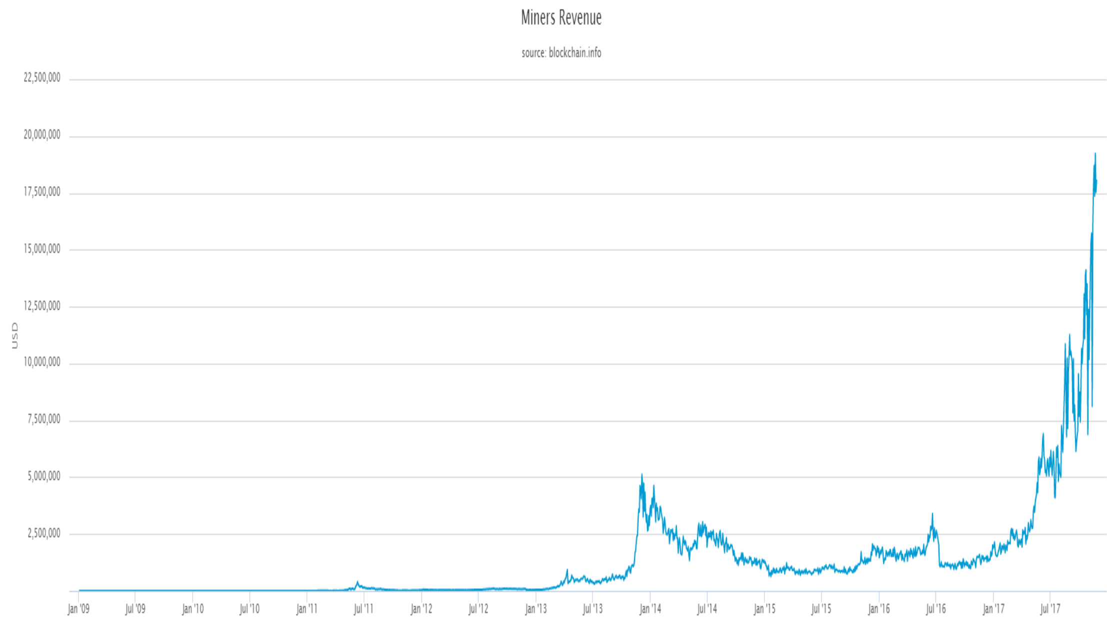
The problem is designed to be solved in ≈10 minutes

Miners are alchemists!



1 block ≈120 000USD





Prefix		Base 10	Adoption
Name	Symbol		
yotta	Y	10^{24}	1991
zetta	Z	10^{21}	1991
exa	E	10^{18}	1975
peta	P	10^{15}	1975
tera	T	10^{12}	1960
giga	G	10^{9}	1960
mega	M	10^{6}	1873
kilo	k	10^{3}	1795

"Sunway TaihuLight" (Wuxi, in Jiangsu province, China) = fastest supercomputer in the world = 93 petaflops

FLOPS = floating point operations per second

5.5 Web wallet

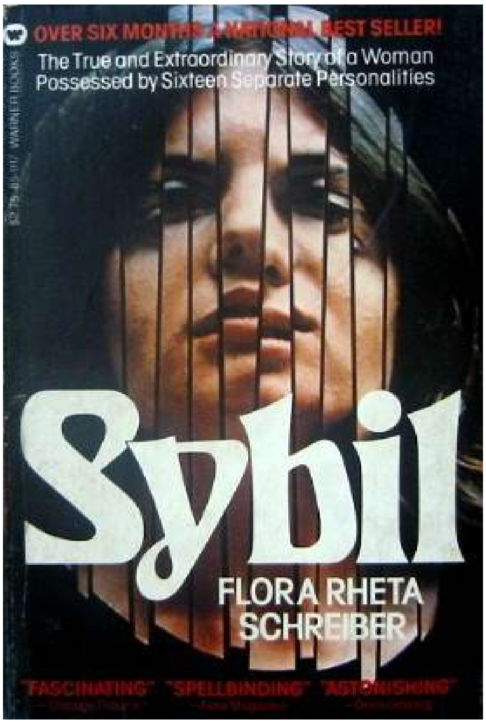
Online wallet hosted on a server and accessible through a website
Instantly functional
$\neq$ full node (takes a few hours to download)

6 Problems

6.1 Sybil attack

Sybil, Flora Rheta Schreiber, (1973), initial print run of 400,000...

A story of a young woman who developed 16 distinct separate personalities...



The Sybil Attack, John R. Douceur, (2002), Proceedings of 1st International Workshop on Peer-to-Peer Systems (IPTPS)

Creating a large number of pseudonymous identities, to gain a disproportionately large influence

"pseudospoofing"

Impossible to cheat on computer performance (limited hashrate)

6.2 Double-spend attack

Capacity to spend twice the same digital token.

If there is a central autorithy, it's easy to prevent for double-spending

It's not possible to spend twice an UTXO since it is recorded the blockchain...

Unless if an attacker is able to rip over the last blocks of a blockchain!

Capacity to modify the historic ledger of transactions

We want transactions to become part of a irreversible process

How to be sure of a given transaction?

Satoshi says: depending on the level of security you ask, wait for several confirmations in the blockchain
e.g., wait for a block with your transaction and wait for 5 more blocks (this is 6 confirmations)...

All based on a calculus at the end of Bitcoin's whitepaper (Section 11 "Calculations")

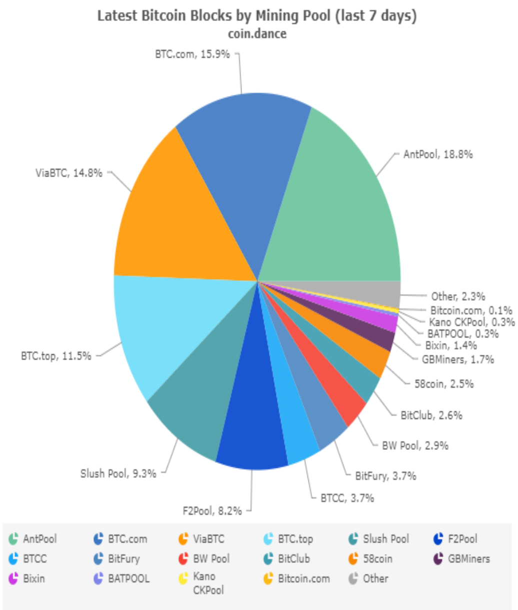
Actually, Satoshi's calculus is slightly incorrect
Following an article of Meni Rosenfeld, we corrected it with Ricardo Perez-Marco

There is a closed-form formula for the probability of a double-spend attack.

We prove that the double-spend probability drops exponentially to 0 as conjectured by Satoshi.

There is another security factor: time spent mining.

> The requirement is that the good guys collectively have more CPU power than any single attacker.



7 Hash functions

Rabin, Yuval, Merkle, late 70.
"Swiss army knife" of cryptography

- input of any size
- output of fixed-size
- easy to calculate (in $O(n)$ if input is n-bit string)

i. collision resistance

ii. preimage resistance

iii. second preimage resistance

One way function
Random Oracles are Practical: A Paradigm for Designing Efficient Protocols, M. Bellare, P. Rogaway, ACM Conference on Computer and Communications Security (1993).
Based on block ciphers
Compression function
Initialization Vector (IV)
Merkle–Damgård construction
Birthday paradox
Integrity of transfered data
Message digest
Commitments
Puzzle
Digital signature
SHA-1, MD5 broken
SHA-2

7.1 Proof of Work

Use of hash function to create a puzzle
Time consuming
Cost function. A string, D integer, x integer

$$\begin{array}{rrcl} \mathcal{F}: & \mathcal{C} \times [0, D_{\max}] \times [0, N] & \longrightarrow & \{\mathrm{True}, \mathrm{False}\} \\ & (A, D, x) & \longmapsto & \mathcal{F}(A, D, x) \end{array}$$

Example: $\mathcal{F}(A,\, D,\, x) = \mathrm{True}$ if $\mathrm{Hash}(A|D|x)$ starts with D zeros and false else.
Problem. Given A, D, find $\mathbf{x}$ such that

$$\mathcal{F}(A, D, \mathbf{x}) \;=\; \mathrm{True} \tag{1}$$

Solution $\mathbf{x}$ (not necessarily unique) called **nonce**
Very hard to solve
Use of computational power

Pricing via Processing or Combatting Junk Mail, C. Dwork and M. Naor, (1993).
Denial-of-service counter measure technique in a number of systems
Anti-spam tool

Hashcash, A Denial of Service Counter-Measure, A. Back, preprint (2002)
Hashcash: a proof-of-work algorithm
Create a stamp to attach to mail
Cost functions proposed are different
Solution of (1) by brute-force.
Calculus of plenty of hash

7.2 Merkle root

Patent in 1979...

A Digital Signature Based on a Conventional Encryption Function, R. C. Merkle (1988).

Merkle tree = Tree of hashes
Oriented Acyclic Rooted tree
Binary Tree
Leaf = Hash (block)
Top Hash = Merkle root

Used to check integrity of a list of blocks

How to prove that an element x belongs to a set S ?
Screen all S ? Solution in $O(n)$.

Solution proportional to the logarithm of the number of nodes of the tree $O(\ln(n))$

Any permutation of leaves gives a new Merkle root...

8 What is Mining?

Hashcash proof-of-work (Adam Back).

$F=$ hash function=SHA256 $\circ$ SHA256

$$
\begin{aligned}
\mathcal{F}(A,D,\mathbf{x}) &= \mathbb{1}_{F(A|D|x)<\frac{2^{224}}{D}} \\
A &= x_1|x_2|x_3|x_4| \\
x_1 &= \text{Version} \\
x_2 &= \text{Hash Previous Block} \\
x_3 &= \text{Hash Merkle Root} \\
x_4 &= \text{Timestamp}
\end{aligned}
$$

Looking for $\mathbf{x}$ such that $\mathcal{F}(A,D,\mathbf{x})=1$.

Nonce = Used only once

Block Header $=A|D|x$

Reference: Bitcoin Wiki

https://en.bitcoin.it/wiki/Block_hashing_algorithm

A block header contains these fields:

Field	Purpose	Updated when...	Size (Bytes)
x_1	Block version number	You upgrade the software and it specifies a new version	4
x_2	256-bit hash of the previous block header	A new block comes in	32
x_3	256-bit hash based on all of the transactions in the block	A transaction is accepted	32
x_4	Current timestamp as seconds since 1970-01-01T00:00 UTC	Every few seconds	4
Target	target- in compact format	The difficulty is adjusted	4
Nonce	32-bit number (starts at 0)	A hash is tried (increments)	4

Actually, there is a **hidden extra-nonce** in the coinbase transaction's scriptSig.

Example 1. Block Hash 0
000000000019d6689c085ae165831e934ff763ae46a2a6c172b3f1b60a8ce26f

Example 2. Block Hash 447384
000000000000000000027175e4c9a3216c1331650e45eafdb948ff03ab59ef1778

8.1 Timestamp parameter

A timestamp is accepted as valid if it is greater than the median timestamp of previous 11 blocks, and less than the network-adjusted time + 2 hours.
"Network-adjusted time" is the median of the timestamps returned by all nodes connected to you.

As a result, block timestamps are not exactly accurate, and they do not even need to be in order. Block times are accurate only to within an hour or two.

You can cheat and use Timestamp as an extra-nonce parameter.
There are anomalies in the public sequence of timestamps
Timestamp adjusted so that

$$
\begin{aligned}
\mathbb{E}[\text{InterBlockTime}] &= 600 \\
\text{InterBlockTime} &= \Delta\text{Timestamp}
\end{aligned}
$$

600 seconds =10 minutes

8.2 Difficulty

Difficulty parameter started at 1 and is updated every 2016 blocks

$$
\begin{aligned}
D_{\text{new}} &= D_{\text{old}} \times \frac{2016 \times 600}{\text{Time used to mine last 2016 blocks}} \\
2016 &= 2 \times 7 \times 24 \times 6 \\
600 &= 10 \times 60
\end{aligned}
$$

Updated every two weeks.

9 What is Bitcoin?

9.1 How to recognize the official Blockchain?

It is $(B_i)_{0 \leqslant i \leqslant N}$ such that $\sum_{i=0}^{N} D_i$ is maximum with $D_i =$ difficulty associated with block B_i.
Difficulty adjusted every 2016 blocks
Official blockchain $\approx$ longest chain

Everything is public
Ledger of valid transactions
Page = block
Everybody can maintain the ledger
Writer = miner
Money transfer = smart contract

Gavin Andresen:

> "Bitcoin" is the ledger of not-previously-spent, validly signed transactions contained in the chain of blocks that begins with the genesis block, follows the 21-million coin creation schedule, and has the most cumulative double-SHA256-proof-of-work.

10 Why should we trust Bitcoin?

10.1 First results

Satoshi was wrong !

Underestimation of double spend success probability

Existence of closed form formulas

Mathematical foundation of Bitcoin

Bitcoin and Gamma functions

Notation 3. *Let* $0 < q < \frac{1}{2}$ *(resp.* $p = 1 - q$*), the relative hash power of the group of attackers (resp. of honest miners).*

Theorem 4. *After* z *blocks have been validated by the honest miners, the probability of success of the attackers is*

$$P(z) = I_{4pq}\left(z, \frac{1}{2}\right)$$

where $I_x(a, b)$ *is the regularized incomplete beta function*

$$I_x(a,b) := \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)} \int_0^x t^{a-1}\,(1-t)^{b-1}\,\mathrm{dt}$$

Corollary 5. *Let* $s = 4\,p\,q < 1$*. When* $z \to \infty$*, we have*

$$P(z) \sim \frac{s^z}{\sqrt{\pi\,(1-z)\,s}}$$

10.2 Other results

Given $z \in \mathbb{N}$, block generation time t for mining z block(s) is publicly known.

Definition 6. *We denote by $P(z,t)$ the probability of success of a double spend attack when z blocks have been validated within a period of time of t.*

What we'll obtain also:

- Closed form formula for $P(z,t)$.
- Satoshi's formula $P_{\mathrm{SN}}(z)$ is actually a $P(z,t)$
- Asymptotics formulas for $P_{\mathrm{SN}}(z)$ and $P(z,t)$
- Explicit rank z_0 such that $P(z) < P_{\mathrm{SN}}(z)$.

In particular,

$$P_{\mathrm{SN}}(z) \sim \frac{e^{-z\left(\frac{q}{p}-1-\ln\frac{q}{p}\right)}}{2}$$

11 Mathematics of mining

11.1 Mining one block

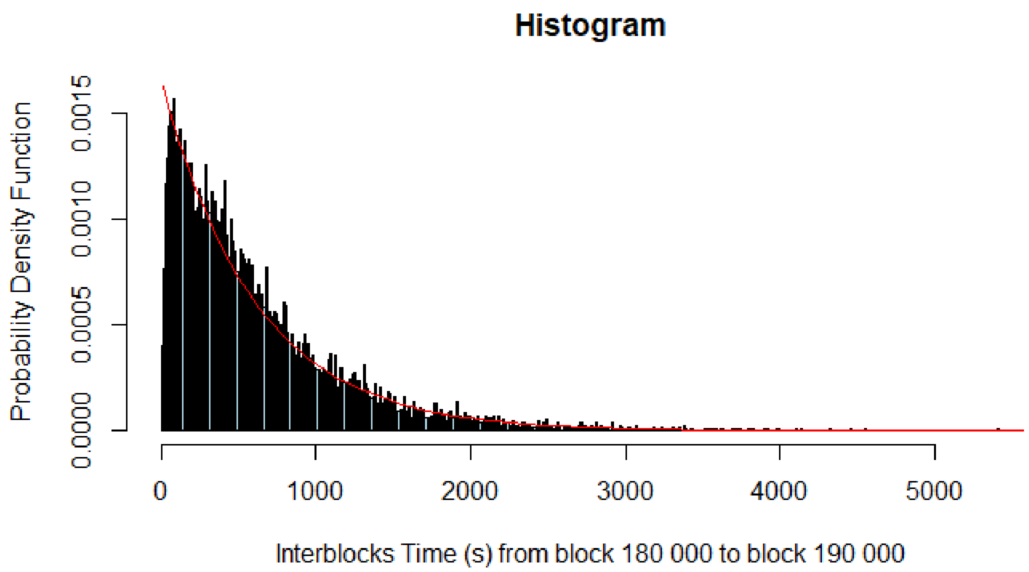
The time it takes to mine a block is memoryless

$$\mathbb{P}[T > t_1 + t_2 | T > t_2] \;=\; \mathbb{P}[T > t_1]$$

Proposition 7. *The random variable* $\boldsymbol{T}$ *has the exponential distribution with parameter* $\alpha = \frac{1}{600}$ *i.e.,*

$$f_{\boldsymbol{T}}(t) \;=\; \alpha\,\mathrm{e}^{-\alpha t}$$

Parameter α seen as a mining speed, $\mathbb{E}[\boldsymbol{T}] = \frac{1}{\alpha}$.
Confirmation by studying timestamps sequence



11.2 Mining more blocks

Interblock times $\mathbf{T}_1, \ldots, \mathbf{T}_n$ are independent identically distributed exponential random variables. The sum

$$\boldsymbol{S}_n \;=\; \boldsymbol{T}_1 + \ldots \boldsymbol{T}_n$$

is the time spent to get n blocks

Proposition 8. *The random variable* $\mathbf{S}_n$ *has a Gamma distribution with parameter* (n, α)*:*

$$f_{\boldsymbol{S}_n}(t) \;=\; \frac{\alpha^n}{(n-1)!}\, t^{n-1}\, \mathrm{e}^{-\alpha t}$$

Definition 9. *Let* $\boldsymbol{N}(t)$ *be the number of blocks already mined at* t*-time. Start is at* $t=0$*.*

Proposition 10. *The random process* $\boldsymbol{N}$ *is a Poisson process with parameter* α *i.e.,*

$$\mathbb{P}[\mathbf{N}(t)=k] \;=\; \frac{(\alpha\, t)^k}{k!}\, \mathrm{e}^{-\alpha t}$$

Notation 11. *The letters* $\boldsymbol{T}, \alpha, \mathbf{S}_n, \boldsymbol{N}$ *(resp.* $\boldsymbol{T}', \alpha'$, $\mathbf{S}'_n, \mathbf{N}$*) are reserved for honest miners (resp. attacker).*

11.3 Interpretation of mining speed

Same notations as above. Mining speed α (honest) and α' (attacker). Probability p (honest) and q (attacker). We note also $\tau_0 = 600$ seconds $=10$ minutes.

Proposition 12. *We have:*

$$p \;=\; \mathbb{P}[\mathbf{T} < \boldsymbol{T}'] \qquad (2)$$

$$p \;=\; \frac{\alpha}{\alpha+\alpha'} \qquad (3)$$

$$q \;=\; \frac{\alpha'}{\alpha+\alpha'} \qquad (4)$$

$$\alpha+\alpha' \;=\; \frac{1}{\tau_0} \qquad (5)$$

$$\alpha \;=\; \frac{p}{\tau_0} \qquad (6)$$

$$\alpha' \;=\; \frac{q}{\tau_0} \qquad (7)$$

Proof. The random variable $\mathrm{Inf}(\boldsymbol{T},\ \boldsymbol{T}')$ has the exponential distribution with parameter $\alpha + \alpha'$. $\square$

Proof. (Another proof). Denote by h (resp. h') the hashrate of the honest miners (resp. attacker) and t_0 (resp. t_0') the average time it takes for mining a block.

Total hashrate of the network $= h + h'$.

Proof-of-work: search for a nonce in Block Header such that

$$\text{Hash(Block Header)} \ < \ \text{Target}$$

Set $m = \frac{2^{256}}{\text{Target}}$ We have

$$\begin{aligned} p &= \frac{h}{h+h'} && (8) \\ q &= \frac{h'}{h+h'} && (9) \\ (h+h')\,\tau_0 &= m && (10) \\ h\,t_0 &= m && (11) \\ h'\,t_0' &= m && (12) \end{aligned}$$

$\square$

So, α, h, p are proportionnal.

12 Classical Double Spend Attack

No eclips attack

12.1 What is a double spend?

A single output may not be used as an input to multiple transactions.

- $T = 0$. A merchant **M** receives a transaction **tx** from **A** (= attacker). Transaction **tx** is issued from an UTXO **tx0**
- Honest Miners start mining openly, transparently
- Attacker **A** starts mining secretly
- One block of honest miners include **tx**
- No block of attacker include **tx**
- On the contrary, one blocks of the attacker includes another transaction **tx'** conflicting with **tx** from same UTXO **tx0**
- As soon as the z-th block has been mined, **M** sends his good to **A**
- **A** keeps on mining secretly
- As soon as A has mined a blockchain with a lenght greater than the official one, A broadcast his blockchain to the network
- Transaction **tx** has disappeared from the official blockchain.

Free Lunch!

13 Interlude: A gambler's ruin problem

Competition

- Gambler against Banker.
- Gambler starts with a handicap of n (lag $= n$)
- Regularly, a croupier flips a biased coin
- Tail probability $= q < p =$ Head probability
- If it's tail, the lag diminishes by 1
- If it's head, it increases by 1
- Gambler wins if he catches up the banker (lag $= 0$)

Random walk with biaised coin.

Note q_n the probability of success. We have: $q_0 = 1$ and $q_n \to 0$ when $n \to \infty$. Also by Markov's property,

$$q_n = q\, q_{n-1} + p\, q_{n+1} \tag{13}$$

Proposition 13. *We have* $q_n = \left(\frac{q}{p}\right)^n$.

An Introduction to Probability Theory and Its Applications, W. Feller (1957)

Gambler = Attacker
Banker = Network (other miners)

14 Nakamoto's Analysis

14.1 Some definitions

Definition 14. *Let $n \in \mathbb{Z}$. We denote by q_n the probability of the attacker $\boldsymbol{A}$ to catch up honest miners whereas $\boldsymbol{A}$'s blockchain is n blocks behind.*

Then, $q_n = \left(\frac{q}{p}\right)^n$ if $n \geqslant 0$ and $q_n = 1$ else.

Definition 15. *For, $z \in \mathbb{N}$, the probability of success of a double-spending attack is denoted by $P(z)$.*

Problem: $P(z) = ?$

Note 16. The probability $P(z)$ is evaluated at $t = 0$. The double-spending attack cannot be successful before $t = \mathbf{S}_z$.

14.2 Formula for $P(z)$

When $t = \mathbf{S}_z$, the attacker has mined $\boldsymbol{N}'(\boldsymbol{S}_z)$ blocks. By conditionning on $\boldsymbol{N}'(\boldsymbol{S}_z)$, we get:

$$\begin{aligned} P(z) &= \sum_{k=0}^{\infty} \mathbb{P}[\boldsymbol{N}'(\boldsymbol{S}_z) = k]\, q_{z-k} \\ &= \mathbb{P}[\boldsymbol{N}'(\boldsymbol{S}_z) \geqslant z] + \sum_{k=0}^{z-1} \mathbb{P}[\boldsymbol{N}'(\boldsymbol{S}_z) = k]\, q_{z-k} \\ &= 1 - \sum_{k=0}^{z-1} \mathbb{P}[\boldsymbol{N}'(\boldsymbol{S}_z) = k] \end{aligned}$$

$$
\begin{aligned}
&+\sum_{k=0}^{z-1} \mathbb{P}[\boldsymbol{N}'(\boldsymbol{S}_z)=k]\, q_{z-k} \\
=\ & 1-\sum_{k=0}^{z-1} \mathbb{P}[\boldsymbol{N}'(\boldsymbol{S}_z)=k]\,(1-q_{z-k})
\end{aligned}
$$

14.3 Satoshi's approximation

White paper, Section 11 **Calculations**

According to Satoshi,

$$
\boldsymbol{S}_z \;\approx\; \mathbb{E}[\boldsymbol{S}_z]
$$

and

$$
\begin{aligned}
\boldsymbol{N}'(\boldsymbol{S}_z) &\approx \boldsymbol{N}'(\mathbb{E}[\boldsymbol{S}_z]) \\
&\approx \boldsymbol{N}'(z\cdot\mathbb{E}[\boldsymbol{T}]) \\
&\approx \boldsymbol{N}'\!\left(z\cdot\frac{\tau_0}{p}\right)
\end{aligned}
$$

So, $\boldsymbol{N}'(\boldsymbol{S}_z) \approx$ Poisson process with parameter λ given by

$$
\begin{aligned}
\lambda &= \alpha'\cdot z\cdot\frac{\tau_0}{p} \\
&= z\cdot\frac{q}{p}
\end{aligned}
$$

The recipient waits until the transaction has been added to a block and z blocks have been linked after it. He doesn't know the exact amount of progress the attacker has made, but assuming the honest blocks took the average expected time per block, the attacker's potential progress will be a Poisson distribution with expected value:

$$\lambda = z\,\frac{q}{p}$$

Definition 17. *We denote by* $P_{\mathrm{SN}}(z)$ *the (false) formula obtained by Satoshi in Bitcoin's white paper.*

Then,

$$P_{\mathrm{SN}}(z) \;=\; 1 - \sum_{k=0}^{z-1} \frac{\lambda^k \mathrm{e}^{-\lambda}}{k!}\left(1 - \left(\frac{q}{p}\right)^{z-k}\right) \qquad (14)$$

Converting to C code...

```
#include <math.h>
double AttackerSuccessProbability(double q, int z)
{
    double p = 1.0 - q;
    double lambda = z * (q / p);
    double sum = 1.0;
    int i,k;
    for (k=0; k<=z; k++)
    {
        double poisson = exp(-lambda);
```

```
            for (i=1; i<=k; i++)
                poisson *= lambda/i;
            sum -= poisson * (1 - pow(q / p, z - k));
        }
        return sum;
    }
```

However,

$$P(z) \;\neq\; P_{\mathrm{SN}}(z)$$

since

$$\boldsymbol{N}'(\boldsymbol{S}_z) \;\neq\; \boldsymbol{N}'(\mathbb{E}[\boldsymbol{S}_z])$$

15 A correct analysis of double-spending attack

15.1 Meni Rosenfeld's correction

Set $\boldsymbol{X}_n := \mathbf{N}'(\boldsymbol{S}_n)$.

Proposition 18. *The random variable $\boldsymbol{X}_n$ has a negative binomial distribution with parameters (n,p), i.e., for $k \geqslant 0$*

$$\mathbb{P}[\boldsymbol{X}_n = k] \;=\; p^n\, q^k \binom{k+n-1}{k}$$

Proof. We have $\boldsymbol{S}_n \sim \Gamma(\alpha, n)$ i.e.,

$$f_{\boldsymbol{S}_n}(t) \;=\; \frac{\alpha^n}{(n-1)!}\, t^{n-1}\, \mathrm{e}^{-\alpha t}$$

with $f_{\boldsymbol{S}_n}(t) = \text{density of } \mathbf{S}_n$. So,

$$\begin{aligned}
\mathbb{P}[\boldsymbol{X}_n = k] &= \int_0^{+\infty} \mathbb{P}[\mathbf{N}'(\boldsymbol{S}_n) = k \,|\, \mathbf{S}_n = t]\; f_{\boldsymbol{S}_n}(t)\,\mathrm{dt} \\
&= \int_0^{+\infty} \frac{(\alpha' t)^k}{k!}\, \mathrm{e}^{-\alpha' t}\, \frac{\alpha^n}{(n-1)!}\, t^{n-1}\, \mathrm{e}^{-\alpha t}\, \mathrm{dt} \\
&= \frac{p^n\, q^k}{(n-1)!\, k!} \int_0^{+\infty} t^{k+n-1}\, \mathrm{dt} \\
&= \frac{p^n\, q^k}{(n-1)!\, k!} \cdot (k+n-1)!
\end{aligned}$$

□

"The attacker's potential progress" is not "a Poisson distribution with expected value $\lambda = z\,\frac{q}{p}$"...

Already remarked in 2012 (probably remarked also by Satoshi?)

Analysis of Hashrate-Based Double-Spending, Meni Rosenfeld, preprint, First Version December 11, 2012, p.7.

Proposition 19. *(Probability of success of the attacker) The probability of success of a double-spending attack is*

$$P(z) \;=\; 1-\sum_{k=0}^{z-1}\left(p^z\, q^k - q^z\, p^k\right)\binom{k+z-1}{k}$$

Proof. Direct application of Section 14.2 and Proposition 18. $\square$

15.2 Numerical Applications

For $q = 0.1$,

z	$P(z)$	$P_{\mathrm{SN}}(z)$
0	1	1
1	0.2	0.2045873
2	0.0560000	0.0509779
3	0.0171200	0.0131722
4	0.0054560	0.0034552
5	0.0017818	0.0009137
6	0.0005914	0.0002428
7	0.0001986	0.0000647
8	0.0000673	0.0000173
9	0.0000229	0.0000046
10	0.0000079	0.0000012

For $q = 0.3$,

z	$P(z)$	$P_{\mathrm{SN}}(z)$
0	1	1
5	0.1976173	0.1773523
10	0.0651067	0.0416605
15	0.0233077	0.0101008
20	0.0086739	0.0024804
25	0.0033027	0.0006132
30	0.0012769	0.0001522
35	0.0004991	0.0000379
40	0.0001967	0.0000095
45	0.0000780	0.0000024
50	0.0000311	0.0000006

Solving for P less than 0.1%:

q	z	z_{SN}
0.1	6	5
0.15	9	8
0.20	18	11
0.25	20	15
0.3	32	24
0.35	58	41
0.40	133	89

Satoshi underestimates $P(z)$...

16 A closed form formula

References.

Hanbook of Mathematical Functions, M. Abramovitch, I.A. Stegun, Dover NY (1970).

Digital Library of Mathematical Functions, http://dlmf.nist.gov

Definition 20. *The* ***Gamma function*** *is defined for* $x > 0$ *by*

$$\Gamma(x) := \int_0^{+\infty} t^{x-1} \mathrm{e}^{-t} \,\mathrm{dt}$$

The ***incomplete Beta function*** *is defined for* $a, b > 0$ *and* $x \in [0, 1]$ *by*

$$B_x(a, b) := \int_0^x t^{a-1} (1-t)^{b-1} \,\mathrm{dt}$$

The (classical) Beta function is defined for $a, b > 0$ *by*

$$B(a, b) := B_1(a, b)$$

The ***regularized Beta function*** *is defined by*

$$I_x(a, b) := \frac{B_x(a, b)}{B(a, b)}$$

Classical result: for $a, b > 0$,

$$B(a, b) = \frac{\Gamma(a)\,\Gamma(b)}{\Gamma(a+b)}$$

Theorem 21. *We have:*

$$P(z) \;=\; I_s(z, 1/2)$$

with $s = 4\,p\,q < 1$.

Proof. It turns out that the cumulative distribution function of a negative binomial random variable $\boldsymbol{X}$ (same notation as above) is

$$\begin{aligned} F_{\boldsymbol{X}}(k) &= \mathbb{P}[\boldsymbol{X} \leqslant k] \\ &= 1 - I_p(k+1, z) \end{aligned}$$

By parts,

$$I_p(k, z) - I_p(k+1, z) \;=\; \frac{p^k\, q^z}{k\, B(k, z)}$$

So,

$$P(z) \;=\; 1 - I_p(z, z) + I_q(z, z)$$

Classical symmetry relation for Beta function:

$$I_p(a, b) + I_q(b, a) \;=\; 1$$

(change of variable $t \mapsto 1 - t$ in the definition). So,

$$I_p(z, z) + I_q(z, z) \;=\; 1$$

We also use:

$$I_q(z, z) \;=\; \frac{1}{2}\, I_s(z, 1/2)$$

with $s = 4\,p\,q$. □

Classical function pbeta implemented in R gives the true double-spending attack success probability.

17 Asymptotic analysis

According to Satoshi,

> Given our assumption that $p > q$, the probability drops exponentially as the number of blocks the attacker has to catch up with increases.

A result which has never been proven...

Lemma 22. *Let $f \in \mathcal{C}^1(\mathbb{R}_+)$ with $f(0) \neq 0$ and absolut convergent integral. Then,*

$$\int_0^{+\infty} f(u)\, \mathrm{e}^{-zu}\, \mathrm{du} \;\sim\; \frac{f(0)}{z}$$

Lemma 23. *For $b > 0$ and $s \in [0, 1]$, we have when $z \gg 1$,*

$$B_s(z, b) \;\sim\; \frac{s^z}{z}(1 - s)^{b-1}$$

Proof. By the change of variable $u = \ln(s / t)$ in the definition of $B_s(z, b) = \int_0^s t^{z-1}(1-t)^{b-1}\,\mathrm{dt}$,

$$B_s(z, b) \;=\; s^z \int_0^{+\infty} (1 - s\, \mathrm{e}^{-u})^{b-1}\, \mathrm{e}^{-zu}\, \mathrm{du}$$

Then, we apply Lemma 22 with $f(u) := (1 - s\, \mathrm{e}^{-u})^{b-1}$. $\square$

Proposition 24. *When $z \to \infty$, we have:*

$$P(z) \sim \frac{s^z}{\sqrt{\pi(1-s)\,z}}$$

with $s = 4\,p\,q < 1$.

Proof. By Stirling formula,

$$\begin{aligned} B(z, 1/2) &= \frac{\Gamma(z)\,\Gamma(1/2)}{\Gamma(z+1/2)} \\ &\sim \sqrt{\frac{\pi}{z}} \end{aligned}$$

So,

$$\begin{aligned} P(z) &= I_s(z, 1/2) \\ &\sim \frac{(1-s)^{-\frac{1}{2}}\,\frac{s^z}{z}}{\sqrt{\frac{\pi}{z}}} \\ &\sim \frac{s^z}{\sqrt{\pi(1-s)\,z}} \end{aligned}$$

□

18 A more accurate risk analysis

The merchant waits for z blocks. Once it has been done, he knows how long it took... Denote this number by τ_1. In average, it should take $\mathbb{E}[z\boldsymbol{T}] = \frac{z\,\tau_0}{p}$.

Definition 25. *Set* $\kappa := \frac{p\,\tau_1}{z\,\tau_0}$

Dimensionless parameter.

Satoshi's approximation: $\kappa = 1$...

Instead of computing $P(z)$, let us compute $P(z, \kappa)$.

Probability for a successful double-spending attack knowing that z blocks have been mined by the honest miners at $\boldsymbol{S}_z = \tau_1$.

Note 26. We have $P_{\mathrm{SN}}(z) = P(z, 1)$.

Note 27. Two different probabilities.

- Theoretical probability $P(z)$ calculated at $T = 0$ by the attacker or the merchant.

- concrete probability $P(z, \kappa)$ calculated at $T = \tau_1$ by the merchant .

Number of bocks mined by the attacker at $T = \tau_1$ unknown to the merchant = Poisson distribution parameter $\lambda(z,\kappa)$:

$$\begin{aligned}\lambda(z,\kappa) &= \alpha'\,\tau_1\\ &= \frac{q}{\tau_0}\cdot\frac{z\,\kappa\,\tau_0}{p}\\ &= \frac{z\,q}{p}\,\kappa\end{aligned}$$

i.e.,

$$\mathbb{P}[\boldsymbol{N}'(\tau_1)=k] \;=\; \frac{\left(\frac{z\,q}{p}\,\kappa\right)^k}{k!}\,\mathrm{e}^{-\frac{z\,q}{p}\kappa}$$

Definition 28. *The regularized Gamma function is defined by:*

$$\Gamma(s,x) \;:=\; \int_x^{+\infty} t^{s-1}\,\mathrm{e}^{-t}\,\mathrm{dt}$$

The regularized incomplete Gamma function is:

$$Q(s,x) \;:=\; \frac{\Gamma(s,x)}{\Gamma(s)}$$

It turns out that

$$Q(z,\lambda) \;=\; \sum_{k=0}^{z-1}\frac{\lambda^k}{k!}\,\mathrm{e}^{-\lambda}$$

So,

Theorem 29. *We have:*

$$P(z,\kappa) \;=\; 1-Q\!\left(z,\frac{\kappa\,z\,q}{p}\right)+\left(\frac{q}{p}\right)^z\mathrm{e}^{\kappa\,z\frac{p-q}{p}}\,Q(z,\kappa\,z)$$

Proof. We have:

$$\begin{aligned}
P(z,\kappa) &= \mathbb{P}[\boldsymbol{N}'(\tau_1) \geqslant z] + \sum_{k=0}^{z-1} \mathbb{P}[\boldsymbol{N}'(\tau_1) = k]\, q_{z-k} \\
&= 1 - \sum_{k=0}^{z-1} \frac{\lambda(z,\kappa)^k}{k!} \mathrm{e}^{-\lambda(z,\kappa)} \\
&+ \sum_{k=0}^{z-1} \left(\frac{q}{p}\right)^{z-k} \cdot \frac{\lambda(z,\kappa)^k}{k!} \mathrm{e}^{-\lambda(z,\kappa)} \\
&= 1 - Q\left(z, \frac{\kappa\, z\, q}{p}\right) + \left(\frac{q}{p}\right)^z \mathrm{e}^{\kappa z \frac{p-q}{p}}\, Q(z, \kappa\, z)
\end{aligned}$$

□

19 Asymptotics Analysis

Lemma 30. *We have:*

i. For $\mu \in]0,1[$, $Q(z,\mu\, z) \rightarrow 1$ *and*

$$1 - Q(z,\mu\, z) \sim \frac{1}{1-\mu} \frac{1}{\sqrt{2\,\pi\, z}} \mathrm{e}^{-z(\mu - 1 - \ln\mu)}$$

ii. For $\mu = 1$, $Q(z,z) \rightarrow \frac{1}{2}$ *and*

$$\frac{1}{2} - Q(z,z) \sim \frac{1}{3\sqrt{2\,\pi\, z}}$$

iii. For $\mu \in]1,+\infty[$,

$$Q(z,\mu\, z) \sim \frac{1}{\mu - 1} \frac{1}{\sqrt{2\,\pi\, z}} \mathrm{e}^{-z(\mu - 1 - \ln\mu)}$$

Proposition 31. *We have* $P_{\mathrm{SN}}(z) \sim \dfrac{\mathrm{e}^{-z\,c\left(\frac{q}{p}\right)}}{2}$ *with*

$$c(\mu) \;:=\; \mu - 1 - \ln\mu$$

Proof. It follows that

$$1 - Q\left(z, \frac{q}{p}\,z\right) \;\sim\; \frac{1}{1-\frac{q}{p}}\,\frac{1}{\sqrt{2\,\pi\,z}}\,e^{-zc\left(\frac{q}{p}\right)}$$
$$\left(\frac{q}{p}\right)^{z} \mathrm{e}^{\kappa\,z\frac{p-q}{p}}\,Q(z,z) \;\sim\; \frac{1}{2}\,e^{-zc\left(\frac{q}{p}\right)}$$

□

More generally, we have 5 different regimes.

Proposition 32. *When* $z \to +\infty$, *we have:*

- *For* $0<\kappa<1$, $P(z,\kappa) \sim \dfrac{1}{1-\kappa\,\frac{q}{p}}\,\dfrac{1}{\sqrt{2\,\pi\,z}}\,\mathrm{e}^{-zc\left(\kappa\frac{q}{p}\right)}$

- *For* $\kappa=1$, $P(z,1) = P_{\mathrm{SN}}(z) \sim \dfrac{\mathrm{e}^{-z\,c\left(\frac{q}{p}\right)}}{2}$

- *For* $1<\kappa<\frac{p}{q}$,

$$P(z,\kappa) \sim \frac{\kappa\left(1-\frac{q}{p}\right)}{(\kappa-1)\left(1-\kappa\,\frac{q}{p}\right)}\,\frac{1}{\sqrt{2\,\pi\,z}}\,\mathrm{e}^{-zc\left(\kappa\frac{q}{p}\right)}$$

- *For* $\kappa=\frac{p}{q}$, $P\left(z,\frac{p}{q}\right) \to \frac{1}{2}$ *and*

$$P\left(z,\frac{p}{q}\right) - \frac{1}{2} \sim \frac{1}{\sqrt{2\,\pi\,z}}\left(\frac{1}{3}+\frac{q}{p-q}\right)$$

- *For* $\kappa > \frac{p}{q}$, $P(z,\kappa) \to 1$ *and*

$$1 - P(z,\kappa) \sim \frac{\kappa\left(1-\frac{q}{p}\right)}{\left(\kappa\frac{q}{p}-1\right)(\kappa-1)} \frac{1}{\sqrt{2\,\pi\,z}}\,\mathrm{e}^{-zc\left(\kappa\frac{q}{p}\right)}$$

Proof. Repetitive application of Lemma 30. □

20 Comparison between $P(z)$ and $P_{\mathrm{SN}}(z)$

20.1 Asymptotic behaviours

The asymptotic behaviours of $P(z)$ and $P_{\mathrm{SN}}(z)$ are quite different

Proposition 33. *We have* $P_{\mathrm{SN}}(z) \prec P(z)$

20.2 Bounds for $P(z)$ and $P_{\mathrm{SN}}(z)$

Goal: compute an explicit rank z_0 such that

$$P_{\mathrm{SN}}(z) \;<\; P(z)$$

for all $z > z_0$.

20.2.1 Upper and lower bounds for $P(z)$

Remember that $s = 4\,p\,q$.
We'll use Gautschi's inequalities.

Proposition 34. *For any* $z > 1$,

$$\sqrt{\frac{z}{z+1}}\,\frac{s^z}{\sqrt{\pi\,z}} \leqslant P(z) \leqslant \frac{s^z}{\sqrt{\pi\,(1-s)\,z}}$$

20.2.2 An upper bound for $P_{\mathrm{SN}}(z)$

Lemma 35. *Let $z \in \mathbb{N}^*$ and $\lambda \in \mathbb{R}_+^*$. We have:*

i. If $\lambda \in]0,1[$, then

$$1 - Q(z, \lambda\, z) \;<\; \frac{1}{1-\lambda}\,\frac{1}{\sqrt{2\,\pi\, z}}\,\mathrm{e}^{-z(\lambda-1-\ln\lambda)}$$

ii. If $\lambda = 1$, $Q(z,z) < \frac{1}{2}$.

Proposition 36. *We have*

$$P_{\mathrm{SN}}(z) \;<\; \frac{1}{1-\frac{q}{p}}\,\frac{1}{\sqrt{2\,\pi\, z}}\,\mathrm{e}^{-zc\left(\frac{q}{p}\right)} + \frac{1}{2}\,\mathrm{e}^{-zc\left(\frac{q}{p}\right)}$$

with $c(\lambda) := \lambda - 1 - \ln\lambda$.

20.3 An explicit rank z_0

Theorem 37. *Let $z \in \mathbb{N}^*$. A sufficient condition to get $P_{\mathrm{SN}}(z) < P(z)$ is $z > z_0$ with*

$$z_0 \;:=\; \mathrm{Max}\left(\frac{2}{\pi\left(1-\frac{q}{p}\right)^2}, \frac{1}{2\sqrt{2}} - \frac{1+\frac{1}{\sqrt{2}}}{2}\,\frac{\ln\left(\frac{2\,\psi_0}{\pi}\right)}{\psi_0}\right)$$

with

$$\psi_0 \;:=\; \frac{q}{p} - 1 - \ln\left(\frac{q}{p}\right) - \ln\left(\frac{1}{4\,p\,q}\right) > 0$$