

On Double Spend Races

BY CYRIL GRUNSPAN

Leonard de Vinci Research Center
Finance Lab
École Supérieure d'Ingénieurs Léonard-de-Vinci

Email: `cyril.grunspan@devinci.fr`

Web: `cyrilgrunspan.fr`

February 1, 2018

Mathematical Foundation of Bitcoin

Article [Double Spend Races](#), in collaboration with Ricardo Perez-Marco
[arXiv:1702.02867](#) [cs.CR]

[Satoshi Risk Tables](#), [arXiv:1702.04421](#) [cs.CR]

Section 11. Calculations of [Bitcoin: A Peer-to-Peer Electronic Cash System](#), Satoshi Nakamoto, 2008.

Following a previous work by Meni Rosenfeld
[Analysis of hashrate-based double-spending](#), 2012

- [Correction of Satoshi's calculus for the probability of success of a double spend attack](#)
- [Proof that "the probability drops exponentially as the number of blocks the attacker has to catch up with increases" \(Satoshi\)](#)
- [Closed form formula with Beta function for this probability](#)
- [More accurate risk analysis knowing the time it took to validate blocks.](#)
- [Underestimation of the probability of double-spend attack](#)

Two groundbreaking ideas in Bitcoin

- New framework for the design of a transaction
- Breakthrough in distributed system theory

Concept of “smart-contract” (prophetized by Nick Szabo)

ScriptSig / ScriptPubKey (not in the white paper)

Use of proof-of-work (rediscovered by Adam Back) to create a [decentralized](#) blockchain

No bibliography at all related with the distributed system theory!

Main references in cryptography (Haber& Stornetta for timestamps server)

Variation of two generals problem. Fisher, Lynch et Paterson, 1985

Theorem. *In a asynchronous model, there is no deterministic algorithm to achieve consensus (if at least one node can crash)*

However, there are randomized consensus.

Randomization makes algorithm powerful...

Proof-of-Work

Time consuming

Cost function. A string, D integer, x integer

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

Problem. Given A (string) and D (level of difficulty), find \mathbf{x} such that

$$\mathcal{F}(A, D, \mathbf{x}) = \text{True} \quad (1)$$

Solution \mathbf{x} (not necessarily unique) is a “proof-of-work” called **nonce**. Problem possibly hard to solve. Use of computational power to solve it.

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.

Hash functions

Use of hash function h to create a puzzle

Example: $\mathcal{F}(A, D, x) = \text{True}$ if $h(A|x)$ starts with D zeros and false else.

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

Merkle–Damgård construction

Message digest

Commitments

Puzzle

Digital signature

SHA-1, MD5 broken

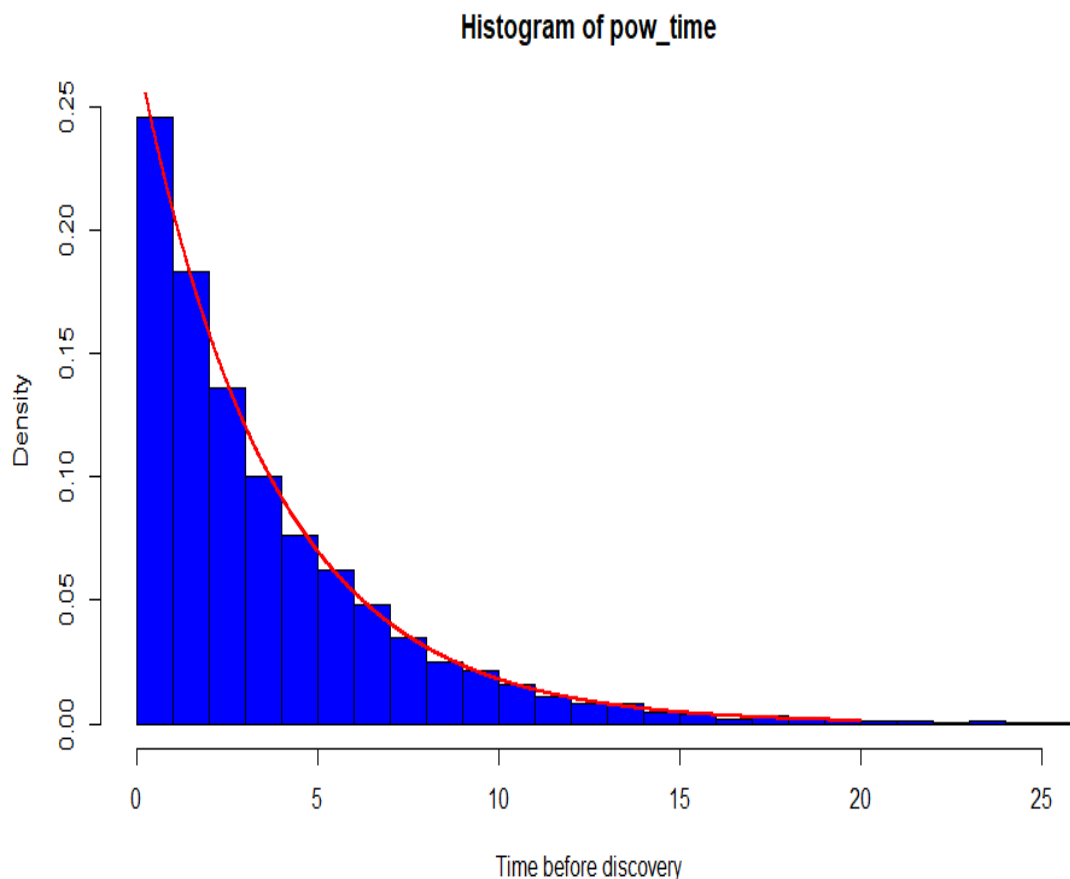
SHA-2

Test of SHA256

Images are uniform & Easy to compute

Proposition. *If h is a hash function, then the time of resolution before getting a “proof-of-work” for a problem of difficulty D has an exponential distribution.*

Example. Problem: find x such that $\text{SHA256}(a|x)$ starts with 4 zeros with a an arbitrary string. Sample (τ_i) . Mean ≈ 4 sec.



However, it is not clear that the distribution is exponential. Tests Cramer-von-Mises and Kolmogorov-Smirnov fail if $\text{size}(\text{sample}) > 6000$ with R software...

Interblock times

Hash function $h = \text{SHA256} \circ \text{SHA256}$

$$\mathcal{F}(A, D, \mathbf{x}) = \mathbb{1}_{h(A|D|x) < \frac{2^{224}}{D}}$$

$$A = x_1|x_2|x_3|x_4|$$

$$x_1 = \text{Version}$$

$$x_2 = \text{Hash Previous Block}$$

$$x_3 = \text{Hash Merkle Root}$$

$$x_4 = \text{Timestamp}$$

Block Header = $A|D|x$. Difficulty adjusted such that the time of resolution is ≈ 600 sec.

Example. Hash Genesis block & Block 500000

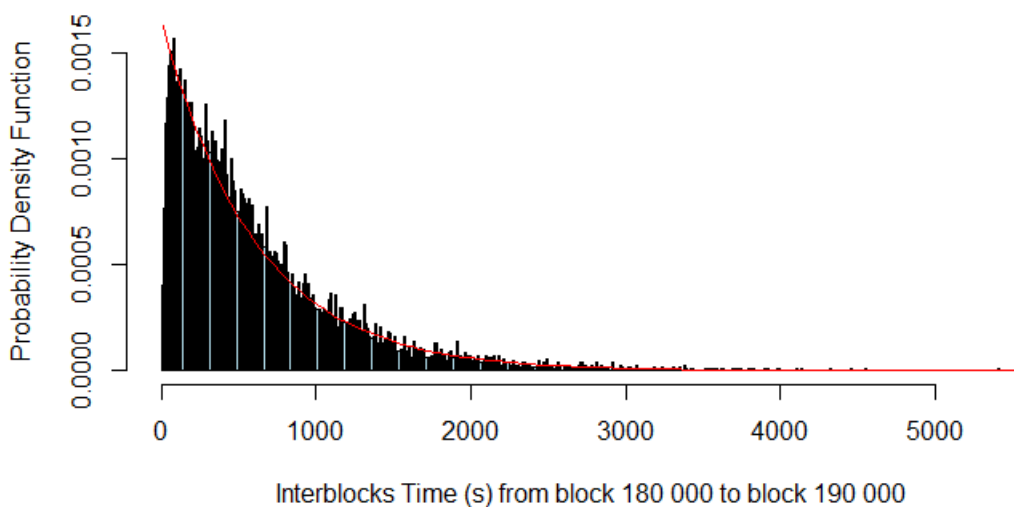
000000000019d6689c085ae165831e934ff763ae46a2a6c172b3f1b60a8ce26f
00000000000000000000000024fb37364cbf81fd49cc2d51c09c75c35433c3a1945d04

Blocksci (Princeton) github.com/citp/BlockSci

Open-source software platform for Blockchain analysis

Example. Between block 180000 and block 190000

Histogram



However in general, KS & CVM tests fail...

Mathematics of mining

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

$$f_{\mathbf{T}}(t) = \alpha e^{-\alpha t}$$

Parameter α seen as a *mining speed*, $\mathbb{E}[\mathbf{T}] = \frac{1}{\alpha}$.

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

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

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

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

Proposition. Let $p := \mathbb{P}[\mathbf{T} < \mathbf{T}']$ and $q = 1 - p$. Then,

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

with $\tau_0 = 600$ sec.

Classical Double Spend Attack

No eclips attack (kind of Sybill's attack)

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 length greater than the official one, **A** broadcast his blockchain to the network
- Transaction **tx** has disappeared from the official blockchain.

Free Lunch!

Nakamoto's Analysis

Some definitions

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

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

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

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

Formula for $P(z)$

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

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

Satoshi's approximation

White paper, Section 11 **Calculations**

According to Satoshi,

$$\mathbf{S}_z \approx \mathbb{E}[\mathbf{S}_z]$$

and

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

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

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

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

Then,

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

However, $P(z) \neq P_{\text{SN}}(z)$ since $\mathbf{N}'(\mathbf{S}_z) \neq \mathbf{N}'(\mathbb{E}[\mathbf{S}_z])$.

A correct analysis of double-spending attack

Meni Rosenfeld's correction

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

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

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

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

Proposition. *The probability of success of a double-spending attack is*

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

Numerical Applications

For $q = 0.1$,

z	$P(z)$	$P_{\text{SN}}(z)$
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	$P(z)$	$P_{\text{SN}}(z)$
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

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 $P(z)$...

A closed form formula

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

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

The *regularized Beta function* is defined by

$$I_x(a, b) := \frac{B_x(a, b)}{B_1(a, b)}$$

Classical result: for $a, b > 0$, $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$

Theorem. We have:

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

with $s = 4pq < 1$.

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

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

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.

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.

Theorem. *When $z \rightarrow \infty$, we have:*

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

with $s = 4 p q < 1$.

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 $\mathbb{E}[z \mathbf{T}] = \frac{z \tau_0}{p}$.

Definition. 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 $\mathbf{S}_z = \tau_1$.

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

Note. 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.,

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

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

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

The regularized incomplete Gamma function is:

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

It turns out that

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

So,

Theorem. *We have:*

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

Proof. We have:

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

□

Asymptotics Analysis

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

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

More generally, we have **5 different regimes**.

Proposition 1. When $z \rightarrow +\infty$, we have:

- For $0 < \kappa < 1$, $P(z, \kappa) \sim \frac{1}{1 - \kappa \frac{q}{p}} \frac{1}{\sqrt{2\pi z}} e^{-z c\left(\kappa \frac{q}{p}\right)}$
- For $\kappa = 1$, $P(z, 1) = P_{\text{SN}}(z) \sim \frac{e^{-z c\left(\frac{q}{p}\right)}}{2}$
- For $1 < \kappa < \frac{p}{q}$,

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

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

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

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

Asymptotic behaviours

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

Proposition. We have $P_{\text{SN}}(z) \prec P(z)$

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

Goal: compute an explicit rank z_0 such that

$$P_{\text{SN}}(z) < P(z)$$

for all $z > z_0$.

Upper and lower bounds for $P(z)$

Remember that $s = 4pq$.

We'll use [Gautschi's inequalities](#).

Proposition 2. For any $z > 1$,

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

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

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

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

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

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

Proposition. We have

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

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

An explicit rank z_0

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

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

with

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

Conclusion. What should the merchant do?

Set $\bar{P}(z, t)$ = probability of success of a double spend attack knowing that z blocks have been validated before t -date.

Shipment condition: Good will be sent to the buyer as soon as $\bar{P}(z, t) < 0.1\%$ for any $q < 0.2$ (for instance) where t = time used to mine z blocks and cf Satoshi Risk Tables.

Shipment_time = $\text{Inf}\{t > 0 / \bar{P}(N(t), t) < \varepsilon\}$.

On average, this will happen after z blocks have been validated and $P(z) < \varepsilon$.

Proposition. *One has $P(z) = \mathbb{E}[P(z, \kappa)]$.*

and $\kappa := \frac{p S_z}{z \tau_0}$ as above.

So, by Markov inequality,

$$\forall \varepsilon > 0, \quad \mathbb{P}[P(z, \kappa) > \varepsilon] < \frac{P(z)}{\varepsilon} \\ \rightarrow 0$$

Note. If $\kappa > 1$, $\mathbb{P}[\kappa > \kappa] \sim \frac{1}{\kappa - 1} \frac{1}{\sqrt{2\pi z}} e^{-z c(\kappa)}$. Other asymptotics in DSR.

So, $\mathbb{P}[\text{Shipment_Time} < +\infty] = 1$.

Submissions

Long list of rejections from

- arxiv.org (section probability)
- European Journal of Operational Research: “I came to conclusion that your paper does not fit the scope of EJOR. **Your list of references also gives support to this conclusion.**”, Emanuele Borgonovo
- Acta Informatica: “[...]the list of references, [...] is comparably short and **does not refer to any paper of the typical Acta Informatica areas.**”, Christel Baier
- SIAM Journal on Financial Mathematics: “**Overall, the authors basically just recast really basic probability results using bitcoin jargons. I think rejection outright is the right decision.**”, Jean-Pierre Fouque
- Journal of Economic Theory: “The paper does not contribute to any ongoing conversations in economics.”, Laura Veldkamp

Finally submitted to International Journal of Theoretical and Applied Finance: “We will send the paper to referees and the process will take approximately 5-6 months.”