Stochastic models for blockchain analysis

Pierre-O. Goffard

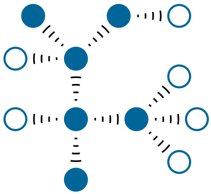
Institut de Science Financières et d'Assurances
pierre-olivier.goffard@univ-lyon1.fr

October 5, 2022

Blockchain

A decentralized data ledger made of blocks maintained by achieving consensus in a P2P network.

- Decentralized
- Public/private
- Permissionned/permissionless
- Immutable
- Incentive compatible



Focus of the talk

Public and permissionless blockchain equipped with the Proof-of-Work protocol.

Consensus protocols

The mechanism to make all the nodes agree on a common data history.

The three dimensions of blockchain systems analysis

1. Efficiency
   - Throughputs
   - Transaction confirmation time
2. Decentralization
   - Fair distribution of the accounting right
3. Security
   - Resistance to attacks

X. Fu, H. Wang, and P. Shi, "A survey of blockchain consensus algorithms: mechanism, design and applications," *Science China Information Sciences*, vol. 64, nov 2020.

Applications of blockchain: Cryptocurrency

S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system." Available at https://bitcoin.org/bitcoin.pdf, 2008.

- Transaction anonymity
- Banking and reliable currency in certain regions of the world
- Money Transfer worldwide (at low fare)
- No need for a thrusted third party

Decentralized finance

DEFI creates new financial architecture

+ Non custodial
+ Anonymous
+ Permisionless
+ openly auditable

- Unregulated
- Tax evasion
- Fraud
- Money laundering

Extends the Bitcoin promises to more complex financial operations

- Collateralized lending
- Decentralized Exchange Platform
- Tokenized assets
- Fundraising vehicle (ICO, STO, ...)

S. M. Werner, D. Perez, L. Gudgeon, A. Klages-Mundt, D. Harz, and W. J. Knottenbelt, "Sok: Decentralized finance (defi)," 2021.

What's inside a block?

A block consists of

- a header
- a list of "transactions" that represents the information recorded through the blockchain.

The header usually includes

- the date and time of creation of the block,
- the block height which is the index inside the blockchain,
- the hash of the block
- the hash of the previous block.

Question

What is the hash of a block?

Cryptographic Hash function

A function that maps data of arbitratry size (message) to a bit array of fixed size (hash value)

$$h : \{0,1\}^* \mapsto \{0,1\}^d.$$

A good hash function is

- deterministic
- quick to compute
- One way
  - ↪ For a given hash value $\overline{h}$ it is hard to find a message m such that
    $$h(m) = \overline{h}$$
- Colision resistant
  - ↪ Impossible to find m_1 and m_2 such that
    $$h(m_1) = h(m_2)$$
- Chaotic
  $$m_1 \approx m_2 \Rightarrow h(m_1) \neq h(m_2)$$

SHA-256

The SHA-256 function which converts any message into a hash value of 256 bits.

Example

The hexadecimal digest of the message

`Sweet home Alabama`

is

`50f3257a3d22a56247a8978fd2505e8cdd64e1cb06e52c941d09e234722dc275`

Mining a block

Introduction

```
Block Hash: 1fc23a429aa5aaf04d17e9057e03371f59ac8823b1441798940837fa2e318aaa
Block Height: 0
Time:2022-02-25 12:42:04.560217
Nonce:0
Block data: [{'sender': 'Coinbase', 'recipient': 'Satoshi', 'amount': 100, 'fee': 0}, {'sen
der': 'Satoshi', 'recipient': 'Pierre-O', 'amount': 5, 'fee': 2}]
Previous block hash: 0
Mined: False
--------------
```

Figure: A block that has not been mined yet.

Mining a block

The maximum value for a 256 bits number is

$$T_{\text{max}} = 2^{256} - 1 \approx 1.16e^{77}.$$

Mining consists in drawing at random a nonce

$$\text{Nonce} \sim \text{Unif}(\{0, \dots, 2^{32} - 1\}),$$

until

$$h(\text{Nonce}|\text{Block info}) < T,$$

where T is referred to as the target.

Difficulty of the cryptopuzzle

$$D = \frac{T_{\text{max}}}{T}.$$

Mining a block

If we set the difficulty to $D = 2^4$ then the hexadecimal digest must start with at least 1 leading 0

```
Block Hash: 0869032ad6b3e5b86a53f9dded5f7b09ab93b24cd5a79c1d8c81b0b3e748d226
Block Height: 0
Time:2022-02-25 13:41:48.039980
Nonce:2931734429
Block data: [{'sender': 'Coinbase', 'recipient': 'Satoshi', 'amount': 100, 'fee': 0}, {'sen
der': 'Satoshi', 'recipient': 'Pierre-O', 'amount': 5, 'fee': 2}]
Previous block hash: 0
Mined: True
--------------
```

Figure: A mined block with a hash value having on leading zero.

The number of trial is geometrically distributed

- Exponential inter-block times
- Lenght of the blockchain = Poisson process

Bitcoin protocol

- One block every 10 minutes on average
- Depends on the hashrate of the network
- Difficulty adjustment every 2,016 blocks ($\approx$ two weeks)
- Reward halving every 210,000 blocks

Check out `https://www.bitcoinblockhalf.com/`

Double spending attack

1. Mary transfers 10 BTCs to John
2. The transaction is recorded in the public branch of the blockchain and John ships the good.
3. Mary transfers to herself the exact same BTCs
4. The malicious transaction is recorded into a private branch of the blockchain
   - Mary has friends among the miners to help her out
   - The two chains are copycat up to the one transaction

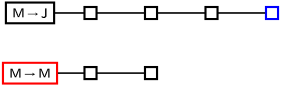
Fact (Bitcoin has only one rule)

The longest chain is to be trusted

Double spending in practice

Double spending attack

Vendor are advised to wait for $\alpha \in \mathbb{N}$ of confirmations so that the honest chain is ahead of the dishonest one.



In the example, vendor awaits $\alpha = 4$ confirmations, the honest chain is ahead of the dishonest one by $z = 2$ blocks.

Fact (PoW is resistant to double spending)

- Attacker does not own the majority of computing power
- Suitable α

Double spending is unlikely to succeed.

S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system." Available at https://bitcoin.org/bitcoin.pdf, 2008.

Mathematical set up

Double spending attack

Assume that

- $R_0 = z \geq 1$ (the honest chain is z blocks ahead)
- at each time unit a block is created

  $\hookrightarrow$ in the honest chain with probability p
  $\hookrightarrow$ in the dishonest chain with probability $q = 1 - p$

The process $(R_n)_{n\geq 0}$ is a random walk on $\mathbb{Z}$ with

$$R_n = z + Y_1 + \ldots + Y_n,$$

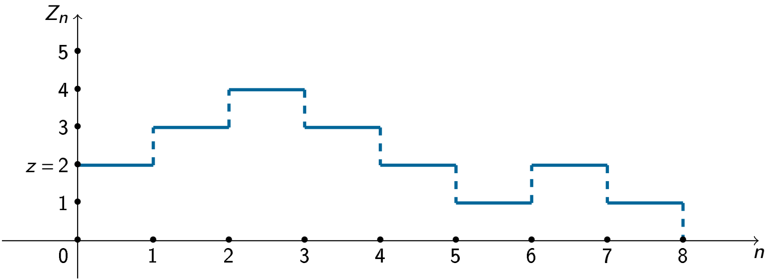
where $Y_1, \ldots, Y_n$ are the **i.i.d.** steps of the random walk.

Double spending rate of success

Double spending attack

Double spending occurs at time

$$\tau_z = \inf\{n \in \mathbb{N}\ ;\ R_n = 0\}.$$



Double spending theorem

If $p > q$ then the double-spending probability is given by

$$\phi(z) = \mathbb{P}(\tau_z < \infty) = \left(\frac{q}{p}\right)^z.$$

Refinements of the double spending problem

Double spending attack

The number of blocks M found by the attacker until the honest miners find α blocks is a negative binomial random variable with **pmf**

$$\mathbb{P}(M=m)=\binom{\alpha+m-1}{m}p^{\alpha}q^{m},\ m\geq 0.$$

The number of block that the honest chain is ahead of the dishonest one is given by

$$Z=(\alpha-M)_{+}.$$

Applying the law of total probability yields the probability of successful double spending with

$$\mathbb{P}(\text{Double Spending})=\mathbb{P}(M\geq\alpha)+\sum_{m=0}^{\alpha-1}\binom{\alpha+m-1}{m}q^{\alpha}p^{m}.$$

M. Rosenfeld, “Analysis of hashrate-based double spending,” *arXiv preprint arXiv:1402.2009*, 2014.

C. Grunspan and R. Perez-Marco, “Double spend race,” *International Journal of Theoretical and Applied Finance*, vol. 21, p. 1850053, dec 2018.

Refinements of the double spending problem

Double spending attack

Let the length of honest and dishonest chain be driven by counting processes

- Honest chain $\Rightarrow$ $z + N_t$, $t \geq 0$, where $z \geq 1$.
- Malicious chain $\Rightarrow$ M_t , $t \geq 0$
- Study the distribution of the first-*rendez-vous* time

$$\tau_z = \inf\{t \geq 0 \ , \ M_t = z + N_t\}.$$

If $N_t \sim \text{Pois}(\lambda t)$ and $M_t \sim \text{Pois}(\mu t)$ such that $\lambda > \mu$ then

$$\phi(z) = \left(\frac{\mu}{\lambda}\right)^z, \ z \geq 0.$$

P.-O. Goffard, "Fraud risk assessment within blockchain transactions," *Advances in Applied Probability*, vol. 51, pp. 443–467, jun 2019.
`https://hal.archives-ouvertes.fr/hal-01716687v2`.

R. Bowden, H. P. Keeler, A. E. Krzesinski, and P. G. Taylor, "Modeling and analysis of block arrival times in the bitcoin blockchain," *Stochastic Models*, vol. 36, pp. 602–637, jul 2020.

Cramer-Lunberg model

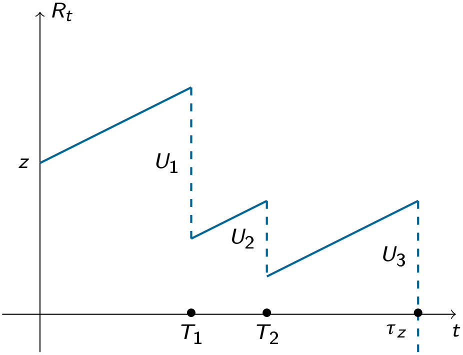
Insurance risk theory

The financial reserves of an insurance company over time have the following dynamic

$$R_t = z + ct - \sum_{i=1}^{N_t} U_i,\ t \geq 0,$$

where

- $z > 0$ denotes the initial reserves
- c is the premium rate
- $(N_t)_{t\geq 0}$ is a counting process that models the claim arrival
  - $\hookrightarrow$ Poisson process with intensity λ
- The U_i's are the randomly sized compensations
  - $\hookrightarrow$ non-negative, **i.i.d.**



Ruin probabilities

Define the ruin time as

$$\tau_z = \inf\{t \geq 0 \ ; \ R_t < 0\}$$

and the ruin probabilities as

$$\psi(z,t) = \mathbb{P}(\tau_z < t) \text{ and } \psi(z) = \mathbb{P}(\tau_z < \infty)$$

We look for z such that

$$\mathbb{P}(\text{Ruin}) = \alpha \ (0.05),$$

given that

$$c = (1+\eta)\lambda\mathbb{E}(U),$$

with

$$\eta > 0 \text{ (net profit condition)}$$

otherwise

$$\psi(z) = 1.$$

S. Asmussen and H. Albrecher, *Ruin Probabilities*.
WORLD SCIENTIFIC, sep 2010.

Ruin probability computation

Insurance risk theory

Let

$$S_t = z - R_t,\ \ t \geq 0$$

Theorem (Wald exponential martingale)

If $(S_t)_{t\geq 0}$ is a Lévy process or a random walk then

$$\{\exp\left[\theta S_t - t\kappa(\theta)\right]\ ,\ t \geq 0\},\text{ is a martingale,}$$

where $\kappa(\theta) = \log \mathbb{E}\left(e^{\theta S_1}\right)$.

Theorem (Representation of the ruin probability)

If $S_t \overset{\text{a.s.}}{\to} -\infty$, and there exists $\gamma > 0$ such that $\{e^{\gamma S_t}\ ,\ t \geq 0\}$ is a martingale then

$$\mathbb{P}(\tau_z < \infty) = \frac{e^{-\gamma z}}{\mathbb{E}\left[e^{\gamma \xi(z)} | \tau_z < \infty\right]},$$

where $\xi(z) = S_{\tau_z} - z$ denotes the deficit at ruin.

Sketch of Proof

- Because of the net profit condition $S_t = \sum_{i=1}^{N_t} U_i - ct \rightarrow -\infty$ as $t \rightarrow \infty$
- $(S_t)_{t\geq 0}$ is a Lévy process, let γ be the (unique, positive) solution to

$$\kappa(\theta) = 0 \text{ (Cramer-Lundberg equation).}$$

- $(e^{\gamma S_t})_{t\geq 0}$ is a Martingale then apply the Optional stopping theorem at τ_z.

Double spending in Satoshi's framework

Link to double spending

- The risk reserve process is $R_t = z + Y_1 + \ldots + Y_t$.
- The claim surplus process is $S_t = -(Y_1 + \ldots + Y_t)$.
- $\kappa(\theta) = 0$ is equivalent to

$$pe^{-\theta} + qe^{\theta} = 1.$$

$\hookrightarrow$ $\gamma = \log(p/q)$.

- If $p > q$ then $S(t) \to -\infty$.
- $\xi(z) = S_{\tau_z} - z = 0$ **a.s**.

Thus,

$$\mathbb{P}(\tau_z < \infty) = \left(\frac{q}{p}\right)^z.$$

Double spending with Poisson processes

Link to double spending

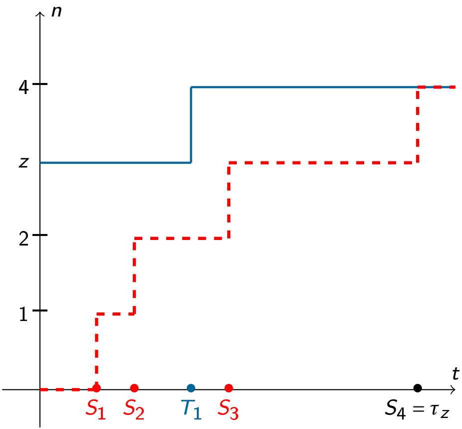
- Suppose that

$$N_t \sim \text{Pois}(\lambda t) \text{ and } M_t \sim \text{Pois}(\mu t)$$

  such that $\lambda > \mu$.
- The risk reserve process is $R_t = z + N_t - M_t$.
- The claim surplus process is $S_t = M_t - N_t$.

Fact

The difference of two Poisson processes is not a Poisson process, However it is Lévy!



Double spending with Poisson processes

- $\kappa(\theta) = 0$ is equivalent to

$$\mu e^{\theta} + \lambda e^{-\theta} - (\lambda + \mu) = 0.$$

  $\hookrightarrow$ $\gamma = \log(\lambda/\mu)$.

- If $\lambda > \mu$ then $S_t \to -\infty$.
- $\xi(z) = S_{\tau_z} - z = 0$ **a.s.**

Thus

$$\mathbb{P}(\tau_z < \infty) = \left(\frac{\mu}{\lambda}\right)^z.$$

Double spending cost

Link to double spending

Mining cryptocurrency in PoW equipped blockchain is energy consuming

$\hookrightarrow$ Operational cost for miners

Per time unit a miner pays

$$c = \pi_W \cdot W \cdot q,$$

where

- π_W is the electricty price per kWh
- W is the consumption of the network `https://cbeci.org/`
- q is the attacker's hashpower

Fact

The cost of double spending is $c \cdot \tau_z$.

Theorem (P.d.f. of the double spending time)

If $\{N_t\ ,\ t \geq 0\}$ is a Poisson process then the **p.d.f.** of τ_z is given by

$$f_{\tau_z}(t) = \mathbb{E}\left[\frac{z}{z+N_t} f_{S_{N_t+z}}(t)\right],\text{ for } t \geq 0.$$

Sketch of the proof

Link to double spending

Let's condition upon the values of N_t,

- if $N_t = 0$ then
$$\tau_z = S_z \text{ and } f_{\tau_z|N_t=0}(t) = f_{S_z}(t)$$
- If $N_t = n$ for $n \geq 1$ then
$$\{\tau_z = t\} = \bigcup_{k=1}^{n} \{T_k \leq S_{z+k-1}\} \cup \{S_{n+z} = t\}$$

We have
$$\begin{aligned} f_{\tau_z|N_t=n}(t) &= \mathbb{P}\left(U_{1:n} \leq S_z/t, \ldots, U_{n:n} \leq S_{z+n-1}/t \middle| S_{n+z} = t\right) f_{S_{n+z}}(t) \\ &= \frac{z}{z+n} f_{S_{n+z}}(t). \end{aligned}$$

Thanks to the properties of the Abel-Gontcharov polynomials.

P.-O. Goffard, "Fraud risk assessment within blockchain transactions," *Advances in Applied Probability*, vol. 51, pp. 443–467, jun 2019. https://hal.archives-ouvertes.fr/hal-01716687v2.

C. Grunspan and R. Pérez-Marco, "ON PROFITABILITY OF NAKAMOTO DOUBLE SPEND," *Probability in the Engineering and Informational Sciences*, pp. 1–15, feb 2021.

M. Brown, E. Peköz, and S. Ross, "BLOCKCHAIN DOUBLE-SPEND ATTACK DURATION," *Probability in the Engineering and Informational Sciences*, pp. 1–9, may 2020.

J. Jang and H.-N. Lee, "Profitable double-spending attacks," *Applied Sciences*, vol. 10, p. 8477, nov 2020.

Take home message

Blockchain is an emerging technologies with great research opportunities for researchers of many fields

- Computer science
- Applied probability
- Statistics
- Economics
- Operations research