Stochastic Models for blockchain analysis

Simple models for blockchain performance analysis

Pierre-O. Goffard

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

23 septembre 2021

The three dimensions of blockchain analysis

1. Security of PoW blockchain
2. Decentralization in PoS blockchain
3. Blockchain efficiency

Double spending attack

Security of PoW blockchain

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

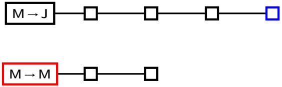
Fact (Bitcoin has only one rule)

The longest chain is to be trusted

Double spending in practice

Security of PoW blockchain

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

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

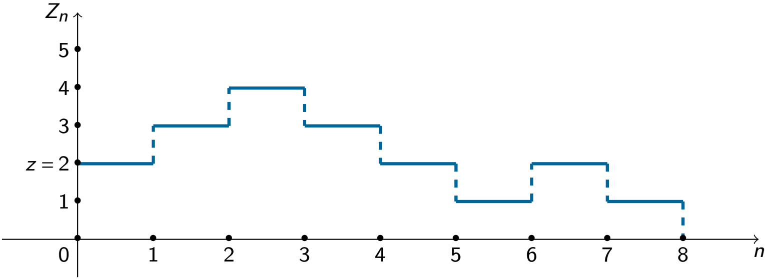
$$R_n = z + Y_1 + \ldots + Y_n,$$

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

Double spending rate of success

Double spending 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

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

Proof of the double spending theorem I

Analogy with the gambler's ruin problem. Using a first step analysis, we have

$$\phi(z) = p\phi(z+1) + (1-p)\phi(z-1), \; z \geq 1. \tag{1}$$

We also have the boundary conditions

$$\phi(0) = 1 \text{ and } \lim_{z \to +\infty} \phi(z) = 0 \tag{2}$$

Equation (1) is a linear difference equation of order 2 associated to the following characteristic equation

$$px^2 - x + 1 - p = 0$$

which has two roots on the real line with

$$r_1 = 1, \text{ and } r_2 = \frac{1-p}{p}.$$

The solution of (1) is given by

$$\phi(z) = A + B\left(\frac{1-p}{p}\right)^z,$$

Proof of the double spending theorem II

where A and B are constant. Using the boudary conditions (2), we deduce that

$$\phi(z) = \left(\frac{1-p}{p}\right)^z$$

as announced.

Refinements of the double spending problem

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

$$\mathbb{P}(\text{Double Spending}) = \mathbb{P}(M \geq \alpha) + \sum_{m=0}^{\alpha-1} \binom{\alpha + m - 1}{m} q^{2\alpha} p^{2m}.$$

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

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.

Perspectives

Security of PoW blockchain

- Distribution of τ_z ? (To be discussed later)
- Distribution of Z
  - $\hookrightarrow$ Negative binomial when the length of the blockchain are driven by Poisson processes
  - $\hookrightarrow$ if not ?

Proof of Stake protocol

Decentralization in PoS blockchain

PoS is the most popular alternative to PoW.

- A block validator is selected according to the number of native coins she owns
- Update the blockchain and receive a reward or do nothing

Two problems

- ⚠ Nothing at stake $\Rightarrow$ Consensus postponed
- ⚠ Rich gets richer $\Rightarrow$ Risk of centralization

Nothing-at-Stake

Decentralization in PoS blockchain

If given the opportunity a node will always append a new block

- Everlasting fork if any

Perpetuating disagreement prevent users to exchange which lower the coin value.

Theorem

To get consensus faster and almost surely

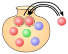
- Set a minimum stake to outweight the benefit of the reward
- Set up a modest reward schedule $\sum_{t=1}^{\infty} \mathrm{r}_t < \infty$

F. Saleh, "Blockchain without waste : Proof-of-stake," *The Review of Financial Studies*, vol. 34, pp. 1156–1190, jul 2020.

Risk of centralization ?

Block appending process

- Draw a coin at random
- The owner of the coin append a block and collect the reward
- The block appender is more likely to get selected during the next round

Similar to Polya's urn 

- Consider an urn of N balls of color in $E = \{1, \dots, p\}$
- Draw a ball of color $x \in E$
- Replace the ball together with r balls of color x

p is the number of peers and r is the size of the block reward.

Theorem

The proportion of coins owned by each peer is stable on average over the long run

I. Roșu and F. Saleh, "Evolution of shares in a proof-of-stake cryptocurrency," *Management Science*, vol. 67, pp. 661–672, feb 2021.

Proof

Consider the balls of some color $x \in E$, and denote by

- N_x the number of balls of color x initially in the urn
- Y_n the number of balls of color x in the urn after n draws
- Z_n the corresponding proportion of balls of color x.

We show that $(Z_n)_{n\geq 0}$ is a $\mathscr{F}_n-$Martingale where $\mathscr{F}_n = \sigma(Y_1,\ldots,Y_n)$. We have

$$\mathbb{E}(Z_{n+1}|\mathscr{F}_n) = Z_n \frac{Y_n + r}{N + r(n+1)} + (1 - Z_n)\frac{Y_n}{N + r(n+1)} = Z_n$$

It follows that

$$\mathbb{E}(Z_n) = \mathbb{E}(Z_0) = \frac{N_x}{N}, \text{ for } n \geq 0.$$

hence the stability. Furthermore, because $|Z_n| < 1$, then $\lim_{n\to\infty} Z_n = Z_\infty$ exists and it holds that $\mathbb{E}(Z_\infty) = \mathbb{E}(Z_0)$.

What is the limiting distributions of the shares ?

Dirichlet distribution

A random vector $(Z_1,\dots,Z_p)$ has a Dirichlet distribution $\text{Dir}(\alpha_1,\dots,\alpha_p)$ with **pdf**

$$f(z_1,\dots,z_p;\alpha_1,\dots,\alpha_p)=\frac{1}{B(\alpha)}\prod_{i=1}^{p} z_i^{\alpha_i-1},$$

for $\alpha_1,\dots,\alpha_p>0$, $0<z_1,\dots,z_p<1$ and $\sum_{i=1}^{p} z_i=1$, where

$$B(\alpha)=\frac{\prod_{i=1}^{p}\Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^{p}\alpha_i)}.$$

Theorem (Convergence toward a Dirichlet distribution)

Suppose that $r=1$ and let X_n be the color of the ball drawn at the n^{th} round then

$$X_\infty\sim\text{Dir}(\{N_x,\ x\in E\}).$$

Proof I

We have that

$$\mathbb{P}(X_1 = x) = \frac{N_x}{N} \tag{3}$$

and

$$\mathbb{P}(X_{n+1} = x) = \frac{N_x + \sum_{i=1}^{n} \delta_{X_i}(x)}{N + n} = m_n(x) \tag{4}$$

where δ_{X_i} denotes the Dirac measure at X_i.
A sequence that satisfies (3) and (4) is said to be a Polya sequence with parameter N_x, $x \in E$.

Lemma

There is an equivalence between the two following statements

(i) $X_1, X_2, \ldots,$ *is a Polya sequence*

(ii) $\mu^* \sim Dir(N_x, x \in E)$ *and* $X_1, X_2, \ldots$ *given* μ^* *are* ***iid*** *as* μ^*

Consider the event $A = \{X_1 = x_1, \ldots, X_n = x_n\}$. Induction on n allows us to show that (i) is equivalent to

$$\mathbb{P}(A) = \prod_{x \in E} \frac{N_x^{[n(x)]}}{N^{[n]}}, \tag{5}$$

Proof II

where $n(x)$ is the number of i's for which $x_i = x$ and $a^{[k]} = a(a+1)\dots(a+k-1)$. (5) is easily shown by induction on $n \in \mathbb{N}$. Now assume that (ii) holds true, then

$$\mathbb{P}(A|\mu^*) = \prod_{x\in E} \mu^*(x)^{n(x)},$$

recall that μ^* is a random vector, indexed on E, We denote by $\mu^*(x)$ the component associated with $x \in E$. The law of total probability then yields

$$\mathbb{P}(A) = \mathbb{E}\left[\prod_{x\in E} \mu^*(x)^{n(x)}\right], \tag{6}$$

which is the same as (5). Applying the lemma together with the law of large number yield

$$n^{-1}\sum_{i=1}^{n} \delta_{X_i}(x) \to \mu^*(x) \text{ as } n \to \infty.$$

and then $m_n(x) \to \mu^*(x)$.

D. Blackwell and J. B. MacQueen, "Ferguson distributions via polya urn schemes," *The Annals of Statistics*, vol. 1, mar 1973.

Measuring decentrality

Decentralization in PoS blockchain

Fact

The most desirable situation corresponds to all the peers being equally likely to be selected.

Decentrality maybe measure by Shannon's entropy

$$H(\mu^*) = -\mathbb{E}\left\{\sum_x \mu^*(x)\ln[\mu^*(x)]\right\} = -\sum_x \frac{N}{N_x}\left[\psi(N_x+1)-\psi(N+1)\right],$$

where $\psi(x) = \frac{\mathrm{d}}{\mathrm{d}x}\ln[\Gamma(x)]$ is the digamma function, to be compared to $\ln(p)$

S. P. Gochhayat, S. Shetty, R. Mukkamala, P. Foytik, G. A. Kamhoua, and L. Njilla, "Measuring decentrality in blockchain based systems," *IEEE Access*, vol. 8, pp. 178372–178390, 2020.

Extensions and perspectives

Decentralization in PoS blockchain

- How to include more peers along the way ?
- What if the peers are not simply buy and hold investors ?
- Find ways to monitor decentralization and take action if necessary

I. Roșu and F. Saleh, "Evolution of shares in a proof-of-stake cryptocurrency," *Management Science*, vol. 67, pp. 661–672, feb 2021.

Efficiency

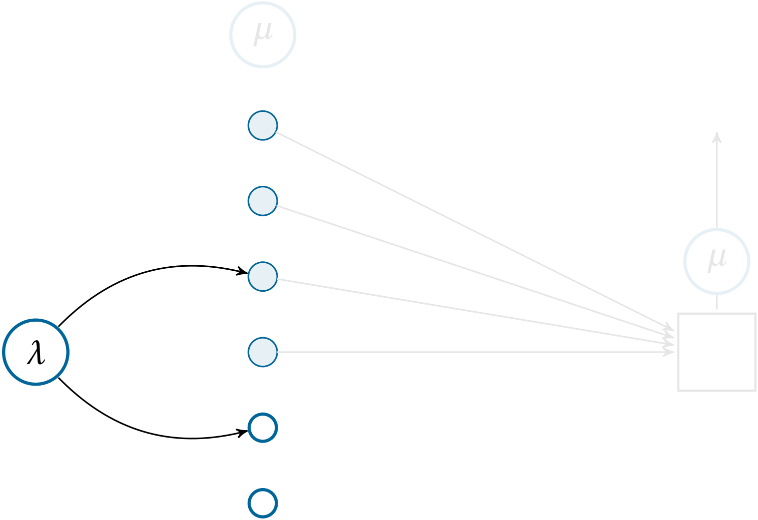
Efficiency is characterized by

- Throughputs : Number of transaction being processed per time unit
- Latency : Average transaction confirmation time

We focus on a PoW equipped blockchain and study the above quantities using a queueing model.

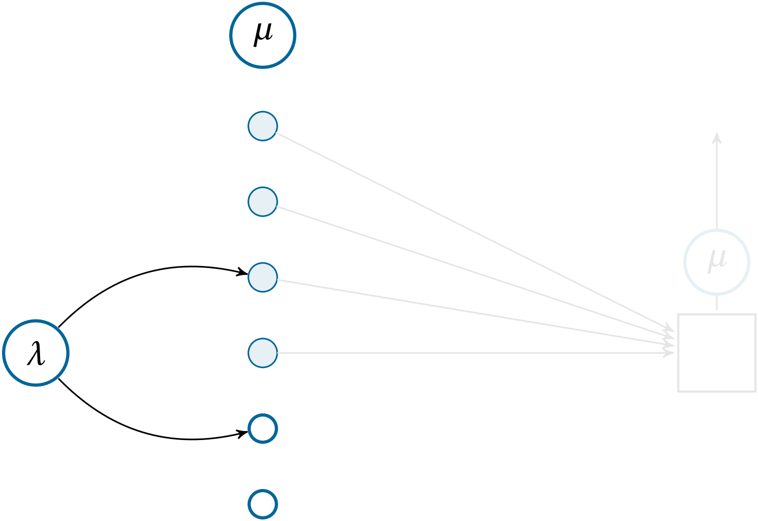
Queue settings

Blockchain efficiency



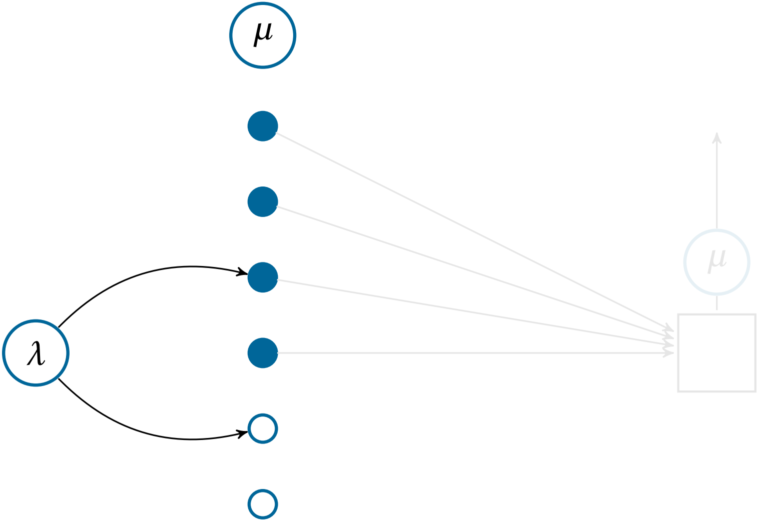
Queue settings

Blockchain efficiency



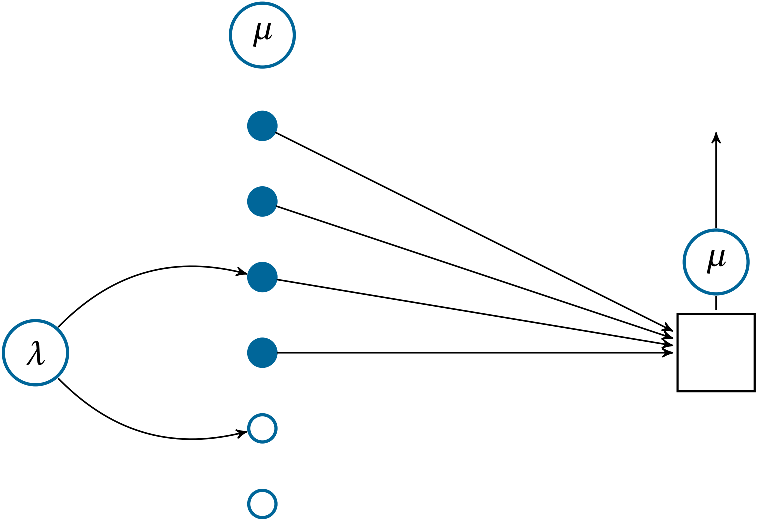
Queue settings

Blockchain efficiency



Queue settings

Blockchain efficiency



Queueing setting

- Poisson arrival with rate $\lambda > 0$ for the transactions
- Poisson arrival with rate $\mu > 0$ for the blocks
- Block size $b \in \mathbb{N}^*$ $\Rightarrow$Batch service

⚠ The server is always busy

This is somekind of $M/M^b/1$ queue.

Y. Kawase and S. Kasahara, "Transaction-confirmation time for bitcoin : A queueing analytical approach to blockchain mechanism," in *Queueing Theory and Network Applications*, pp. 75–88, Springer International Publishing, 2017.

N. T. J. Bailey, "On queueing processes with bulk service," *Journal of the Royal Statistical Society : Series B (Methodological)*, vol. 16, pp. 80–87, jan 1954.

D. R. Cox, "The analysis of non-markovian stochastic processes by the inclusion of supplementary variables," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 51, pp. 433–441, jul 1955.

Queue length distribution

The queueing process eventually reaches stationarity if

$$\mu \cdot b > \lambda. \tag{7}$$

We denote by N^q the length of the queue upon stationarity.

The blockchain efficiency theorem

Assume that (7) holds then N^q is geometrically distributed

$$\mathbb{P}(N^q = n) = (1-p)\cdot p^n,$$

where $p = 1/z^*$ and z^* is the only root of

$$-\frac{\lambda}{\mu} z^{b+1} + z^b\left(\frac{\lambda}{\mu}+1\right) - 1,$$

such that $|z^*| > 1$.

Proof of the efficiency theorem I

Let N_t^q be the number of transactions in the queue at time $t \geq 0$ and X_t the time elapsed since the last block was found. Further define

$$P_n(x,t)\mathrm{d}x = \mathbb{P}[N_t^q = n, X_t \in (x, x+\mathrm{d}x)]$$

If $\lambda < \mu \cdot b$ holds then the process admits a limiting distribution given by

$$\lim_{t\to\infty} P_n(x,t) = P_n(x).$$

We aim at finding the distribution of the queue length upon stationarity

$$\mathbb{P}(N^q = n) := \alpha_n = \int_0^{\infty} P_n(x)\mathrm{d}x. \tag{8}$$

Consider the possible transitions over a small time lapse h during which no block is being generated. Over this time interval, either

- no transactions arrives
- one transaction arrives

Proof of the efficiency theorem II

We have for $n \geq 1$

$$P_n(x+h) = e^{-\mu h}\left[e^{-\lambda h}P_n(x) + \lambda h e^{-\lambda h}P_{n-1}(x)\right]$$

Differentiating with respect to h and letting $h \to 0$ leads to

$$P_n'(x) = -(\lambda+\mu)P_n(x) + \lambda P_{n-1}(x),\ n \geq 1. \tag{9}$$

Similarly for $n = 0$, we have

$$P_0'(x) = -(\lambda+\mu)P_0(x). \tag{10}$$

We denote by $\xi(x) = \mu$ the hazard function of the block arrival time (constant as it is exponentially distributed). The system of differential equations (9), (10) admits boundary conditions at $x = 0$ with

$$\begin{cases} P_n(0) = \int_0^{+\infty} P_{n+b}(x)\xi(x)\mathrm{d}x = \mu\alpha_{n+b}, & n \geq 1, \\ P_0(0) = \mu\sum_{n=0}^{b}\alpha_n, & n = 0,\dots,b \end{cases} \tag{11}$$

Define the probability generating function of N^q at some elapsed service time $x \geq 0$ as

$$G(z;x) = \sum_{n=0}^{\infty} P_n(x)z^n.$$

Proof of the efficiency theorem III

By differentiating with respect to x, we get (using (9) and (10))

$$\frac{\partial}{\partial x} G(z;x) = -\left[\lambda(1-z)+\mu\right] G(z;x)$$

and therefore

$$G(z;x) = G(z;0)\exp\left\{-\left[\lambda(1-z)+\mu\right]x\right\}$$

We get the probability generating function of N^q by integrating over x as

$$G(z) = \frac{G(z;0)}{\lambda(1-z)+\mu} \tag{12}$$

Proof of the efficiency theorem IV

Using the boundary conditions (11), we write

$$\begin{aligned} G(z;0) &= \mu\sum_{n=0}^{\infty} P_n(0)z^n \\ &= P_0(0)+\sum_{n=1}^{+\infty} P_n(0)z^n \\ &= \mu\sum_{n=0}^{b}\alpha_n+\mu\sum_{n=1}^{+\infty}\alpha_{n+b}z^n \\ &= \mu\sum_{n=0}^{b}\alpha_n+\mu z^{-b}\left[G(z)-\sum_{n=0}^{b}\alpha_n z^n\right] \end{aligned} \tag{13}$$

Replacing the left hand side of (13) by (12), multiplying on both side by z^b and rearranging yields

$$\frac{G(z)}{M(z)}[z^b-M(z)]=\sum_{n=0}^{b-1}\alpha_n(z^b-z^n), \tag{14}$$

where $M(z)=\mu/(\lambda(1-z)+\mu)$. Using Rouche's theorem, we find that both side of the equation shares b zeros inside the circle $\mathscr{C}=\{z\in\mathbb{C}\,;\,|z|<1+\epsilon\}$ for some epsilon. One of them is 1, and we denote by z_k, $k=1,\ldots,b-1$ the remaining $b-1$ zeros. Given the polynomial form of the right

Proof of the efficiency theorem V

hand side of (14), the fundamental theorem of algebra indicates that the number of zeros is b. The left hand side can be rewritten as

$$G(z)\left[-\frac{\lambda}{\mu}z^{b+1}+\left(1+\frac{\lambda}{\mu}\right)z^{b}-1\right],$$

we deduce that there is one zeros outside $\mathscr{C}$, we can further show that it is a real number z^*. Multiplying both side of (14) by $(z-1)\prod_{k=1}^{b-1}(z-z_k)$, and using $G(1)=1$ yields

$$G(z)=\frac{1-z^*}{z-z^*}.$$

N^q is then a geometric random variable with parameter $p=\frac{1}{z^*}$.

Latency and throughputs

Blockchain efficiency

Little's law

Consider a stationary queueing system and denote by

- $1/\lambda$ the mean of the unit inter-arrival times
- L be the mean number of units in the system
- W be the mean time spent by units in the system

We have

$$L = \lambda \cdot W$$

J. D. C. Little, "A proof for the queuing formula :L= λW," *Operations Research*, vol. 9, pp. 383–387, jun 1961.

- Latency is the confirmation time of a transaction

$$\text{Latency} = \frac{p}{(1-p)\lambda} + \frac{1}{\mu}$$

- Throughput is the number of transaction confirmed per time unit

$$\text{Throughput} = \mathbb{E}(N^q \mathbb{I}_{N^q \leq b} + b\mathbb{I}_{N^q > b}) = \sum_{n=0}^{b} n(1-p)p^n + bp^{b+1}.$$

Perspective

1. Include some priority consideration to account for the transaction fees

Y. Kawase, , and S. Kasahara, "Priority queueing analysis of transaction-confirmation time for bitcoin," *Journal of Industrial & Management Optimization*, vol. 16, no. 3, pp. 1077–1098, 2020.

2. Go beyond the Poisson process framework

Q.-L. Li, J.-Y. Ma, and Y.-X. Chang, "Blockchain queue theory," in *Computational Data and Social Networks*, pp. 25–40, Springer International Publishing, 2018.

Q.-L. Li, J.-Y. Ma, Y.-X. Chang, F.-Q. Ma, and H.-B. Yu, "Markov processes in blockchain systems," *Computational Social Networks*, vol. 6, jul 2019.

A fourth dimension to analyse

The energy consumption dimension

https://cbeci.org/