Stochastic models for blockchain analysis

Pierre-O. Goffard

Université de Strasbourg
goffard@unistra.fr

October 7, 2022

1. Blockchain

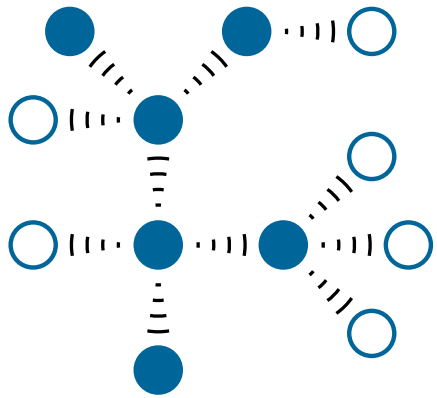
2. Insurance risk theory

3. Applications of Bayesian statistics in actuarial science

Blockchain

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

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



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
- Proof of Work (**PoW**)

Consensus protocols

The mechanism to make all the nodes agree on a common data history using the network ressources

- Bandwidth
- Storage
- Computing power
- Cryptocurrency

Consensus protocols

The three dimensions of blockchain systems analysis

1. Efficiency (Queueing model)
   - Throughputs
   - Transaction confirmation time
2. Security (Insurance risk model)
   - Resistance to attacks
3. Decentralization (stochastic processes with reinforcement)
   - Fair distribution of the accounting right

BLOCKASTICS project

Analysis and optimization of blockchain systems.

`link`

Blockchain as a research topic

- Computer science
  - Peer-to-peer networks and consensus algorithm
  - Cryptography and security
- Economics
  - Game theory to study the incentive mechanism at play
  - Nature of the cryptoassets
- Operations research
  - Optimization of complex system
- Financial math
  - Valuation models for cryptoassets
- Machine learning and statistics
  - Open data
  - Interaction between blockchain users
  - (Social) network analysis
  - Clustering of public keys and addresses in the bitcoin blockchain.

My contributions

- Security of **PoW** blockchains

Hansjoerg Albrecher and Pierre-Olivier Goffard.
On the profitability of selfish blockchain mining under consideration of ruin.
Operations Research, 70(1):179–200, jan 2022.

Pierre-O. Goffard.
Fraud risk assessment within blockchain transactions.
Advances in Applied Probability, 51(2):443–467, jun 2019.
https://hal.archives-ouvertes.fr/hal-01716687v2.

- Decentralization of **PoW** blockchains

Hansjörg Albrecher, Dina Finger, and Pierre-O. Goffard.
Blockchain mining in pools: Analyzing the trade-off between profitability and ruin.
Insurance: Mathematics and Economics, 105:313–335, jul 2022.

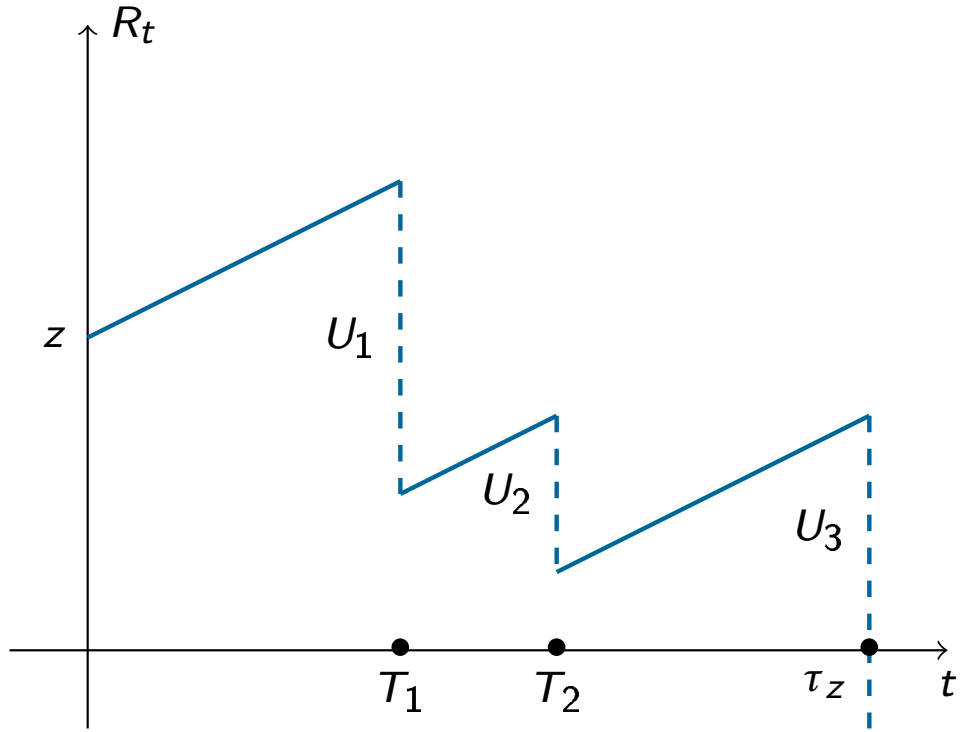
Cramer-Lunberg model

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
  - ↪ Poisson process with intensity λ
- The U_i's are the randomly sized compensations
  - ↪ non-negative, **i.i.d.**



Ruin probabilities

Insurance risk theory

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.

My contributions

Insurance risk theory

Numerical approximations via Laplace transform inversion

Søren Asmussen, Pierre-Olivier Goffard, and Patrick J. Laub.

Orthonormal polynomial expansions and lognormal sum densities.
In *Risk and Stochastics*, pages 127–150. WORLD SCIENTIFIC (EUROPE), apr 2019.

Pierre-Olivier Goffard and Patrick J. Laub.

Orthogonal polynomial expansions to evaluate stop-loss premiums.
Journal of Computational and Applied Mathematics, 370:112648, may 2020.

Pierre-Olivier Goffard, Stéphane Loisel, and Denys Pommeret.

Polynomial approximations for bivariate aggregate claims amount probability distributions.
Methodology and Computing in Applied Probability, 19(1):151–174, nov 2015.

Pierre-Olivier Goffard, Stéphane Loisel, and Denys Pommeret.

A polynomial expansion to approximate the ultimate ruin probability in the compound poisson ruin model.
Journal of Computational and Applied Mathematics, 296:499–511, apr 2016.

My contributions

Insurance risk theory

Closed form expressions via combinatorial analysis

Pierre-Olivier Goffard.

Two-sided exit problems in the ordered risk model.
Methodology and Computing in Applied Probability, 21(2):539–549, nov 2017.

Pierre-Olivier Goffard and Claude Lefèvre.

Boundary crossing of order statistics point processes.
Journal of Mathematical Analysis and Applications, 447(2):890–907, mar 2017.

Pierre-Olivier Goffard and Claude Lefèvre.

Duality in ruin problems for ordered risk models.
Insurance: Mathematics and Economics, 78:44–52, jan 2018.

Bayesian statistics

Let $\mathscr{M}$ be a model with parameters θ, and x some observed data.

- Bayesian statistics targets the posterior distribution of the parameters

$$\pi(\theta|\mathrm{x}) = \frac{p(\mathrm{x}|\theta)\pi(\theta)}{\int_\Theta p(\mathrm{x}|\theta)\pi(\theta)\mathrm{d}\theta} = \frac{p(\mathrm{x}|\theta)\pi(\theta)}{Z(\mathrm{x})},$$

  by updating the prior $\pi(\theta)$ via the likelihood $p(\mathrm{x}|\theta)$.

↪ Model calibration

If many models $\mathscr{M}_1,\ldots,\mathscr{M}_K$ are competing

- The posterior model evidence of each model follow from

$$\pi(M_i|\mathrm{x}) = \frac{p(\mathrm{x}|M_i)\pi(\mathscr{M}_i)}{\sum_{j=1}^K p(\mathrm{x}|M_j)\pi(\mathscr{M}_j)},\ i = 1,\ldots,K.$$

↪ Select or combine models

My contributions

Applications of Bayesian statistics in actuarial science

- Mortality forecasting using demographic projection model and STAN

  Karim Barigou, Pierre-Olivier Goffard, Stéphane Loisel, and Yahia Salhi.
  Bayesian model averaging for mortality forecasting using leave-future-out validation.
  International Journal of Forecasting, mar 2022.

- Sequential Monte Carlo Sampler to fit composite models

  Pierre-Olivier Goffard.
  Sequential Monte Carlo samplers to fit and compare insurance loss models.
  working paper or preprint, March 2022.

- Approximate Bayesian Computation with aggregated data

  Pierre-Olivier Goffard and Patrick J. Laub.
  Approximate bayesian computations to fit and compare insurance loss models.
  Insurance: Mathematics and Economics, 100:350–371, sep 2021.


Bitcoin envoyés
−0,00158804 BTC
−34,28 €
À
bc1qpt2xw0...uwhjjsnzfj
Prix par « coin »
21627,21 €
Réseau
Bitcoin
Frais de réseau
0,00001180 BTC
Date
25 août 2022 · 20:25
Statut
Pending
Voir sur l'explorateur de blocs

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

```
Bienvenue à l'IRMA
```

is

```
50f3257a3d22a56247a8978fd2505e8cdd64e1cb06e52c941d09e234722dc275
```

Mining a block

```
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

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

Mining consists in drawing at random a nonce

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

until

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

where T is referred to as the target.

Difficulty of the cryptopuzzle

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

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/`

Efficiency

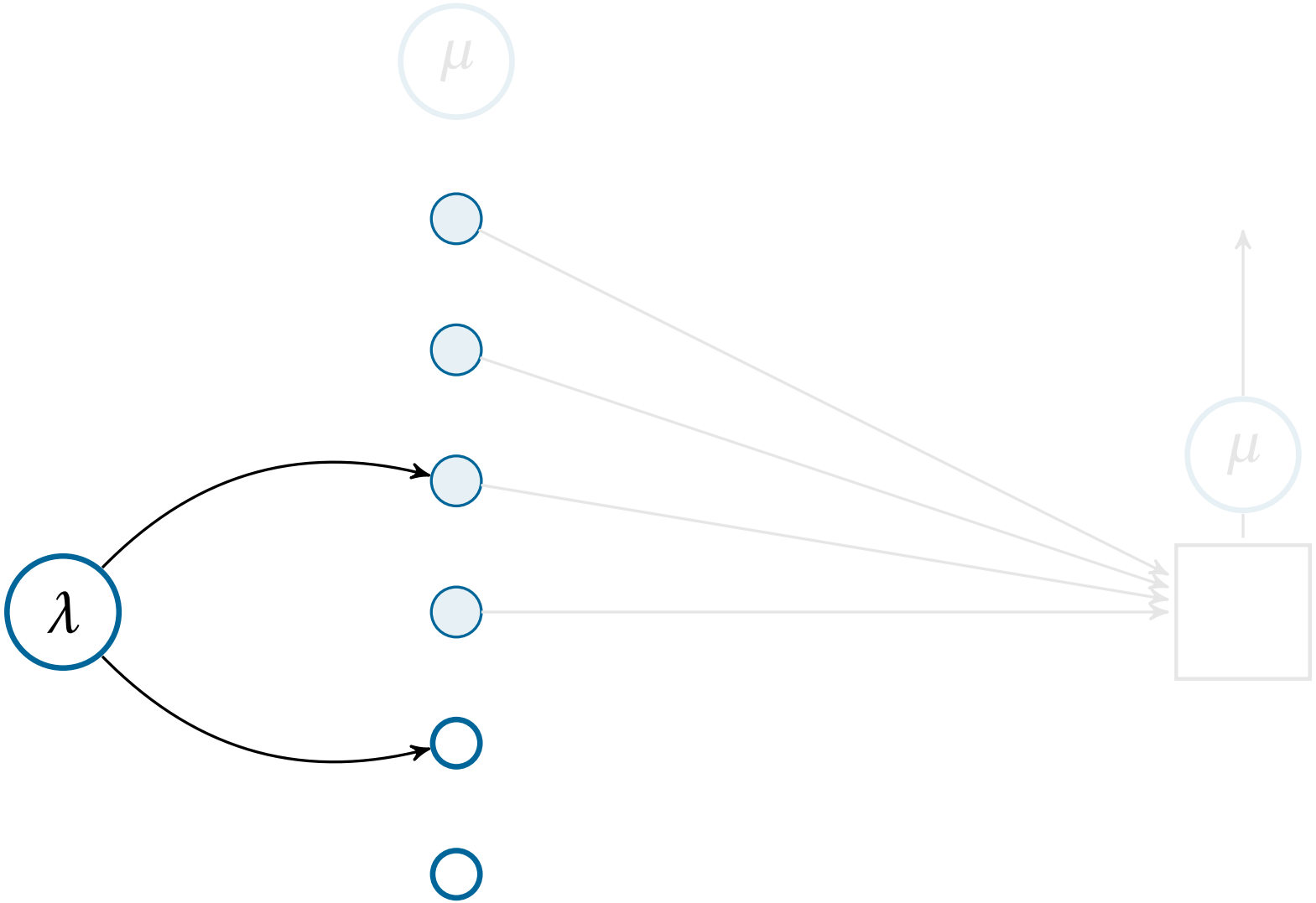
Blockchain efficiency

Efficiency is characterized by

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

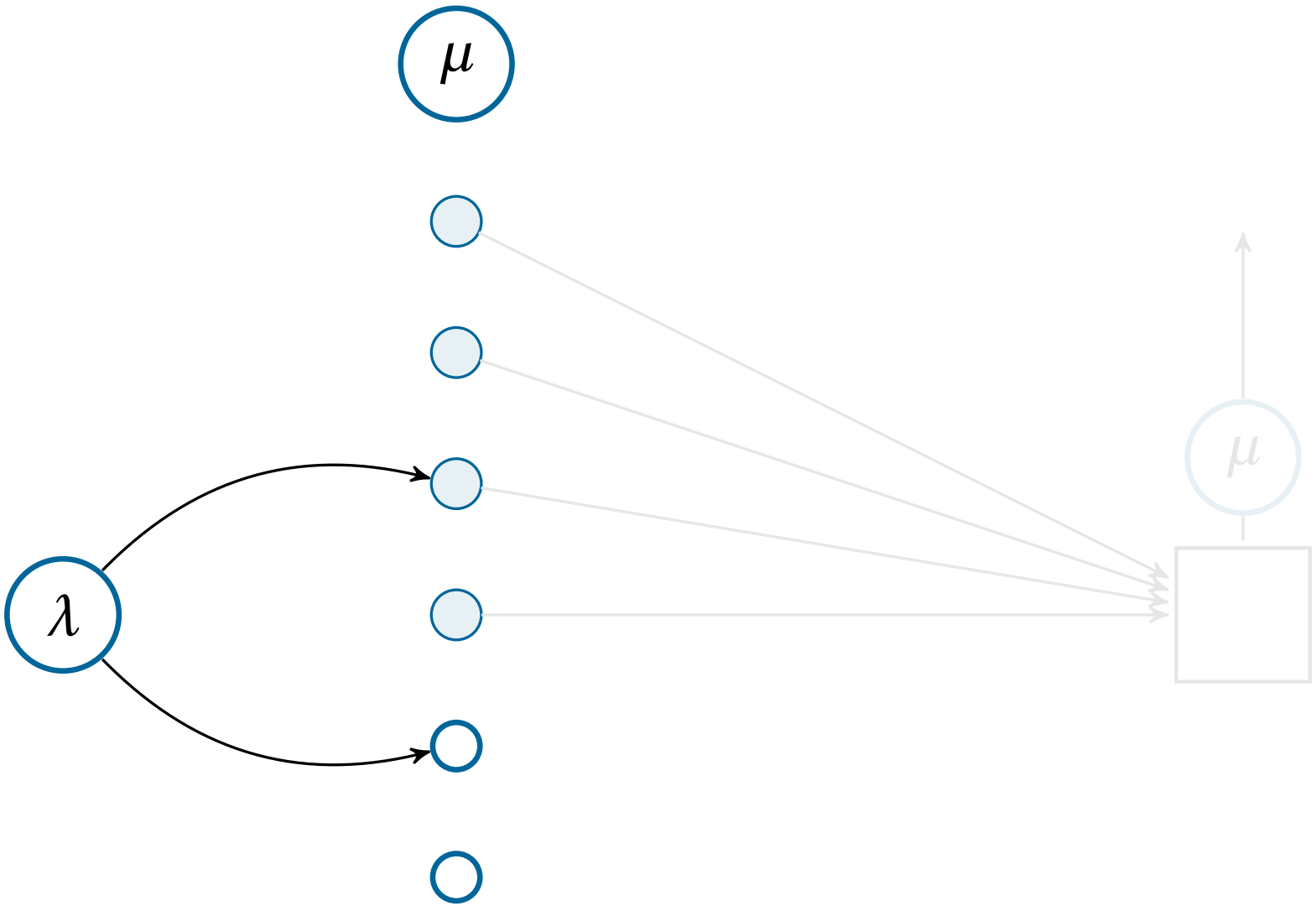
Queue settings

Blockchain efficiency



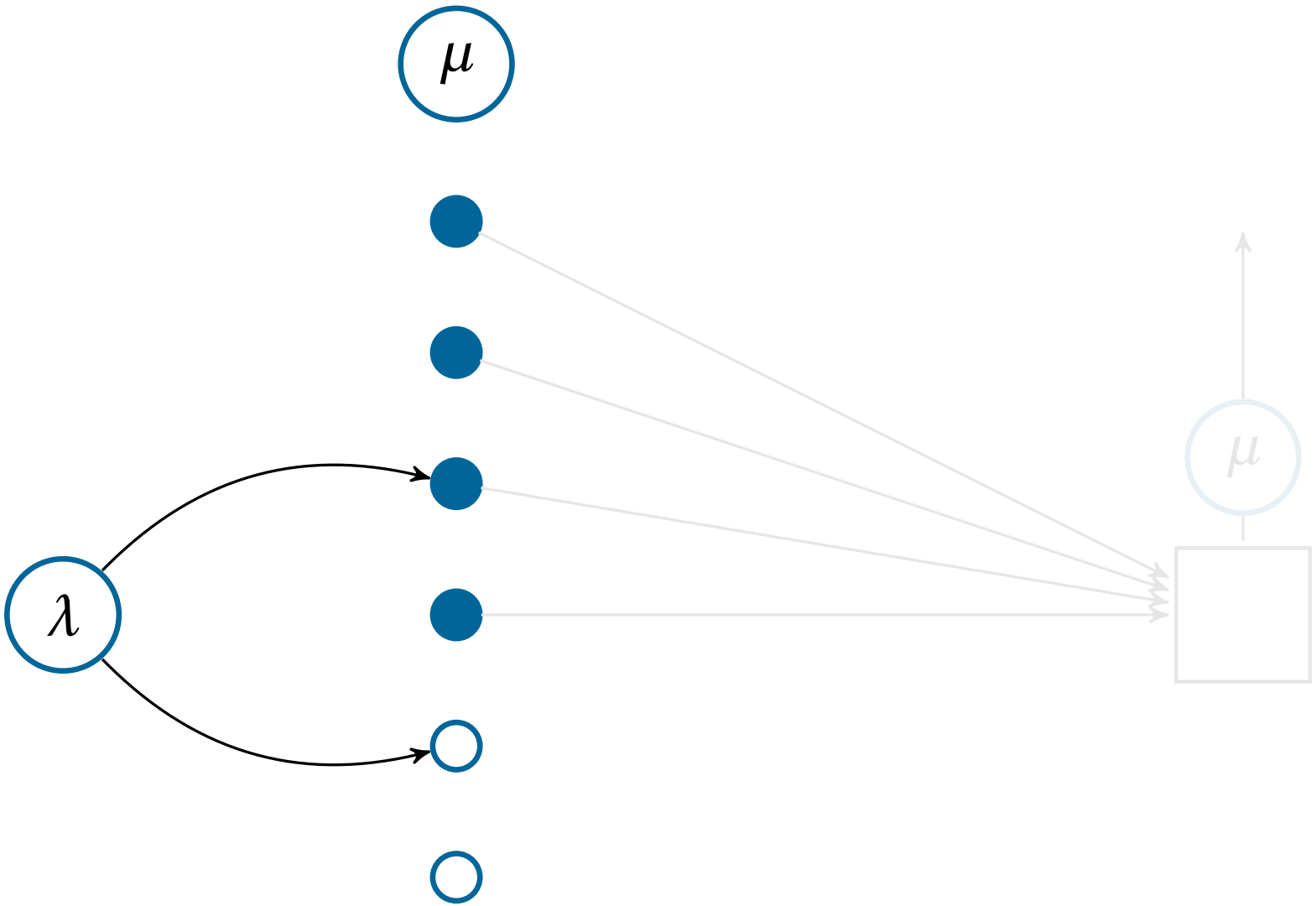
Queue settings

Blockchain efficiency



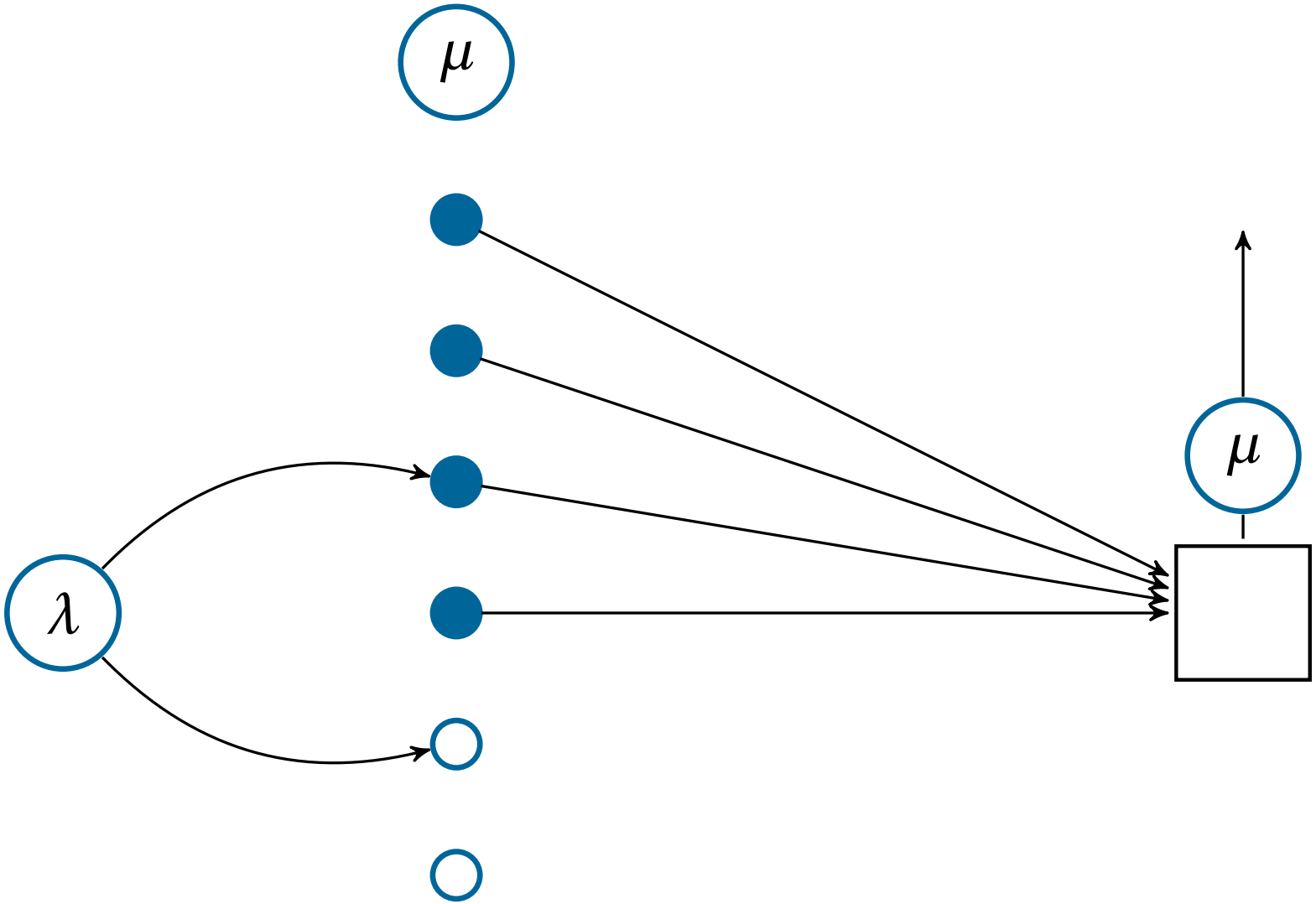
Queue settings

Blockchain efficiency



Queue settings

Blockchain efficiency



Queueing setting

Blockchain efficiency

- 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.

Double spending attack

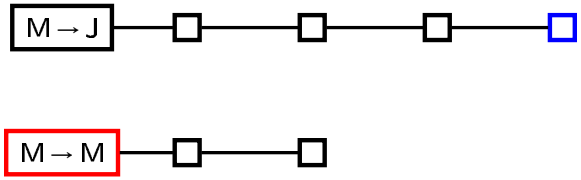
1. Mary transfers 10 BTCs to John
2. The transaction is recorded in the public branch of the blockchain and John ships the good.
3. Mary transfers to herself the exact same BTCs
4. The malicious transaction is recorded into a private branch of the blockchain
   - Mary has friends among the miners to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

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

Blockchain security

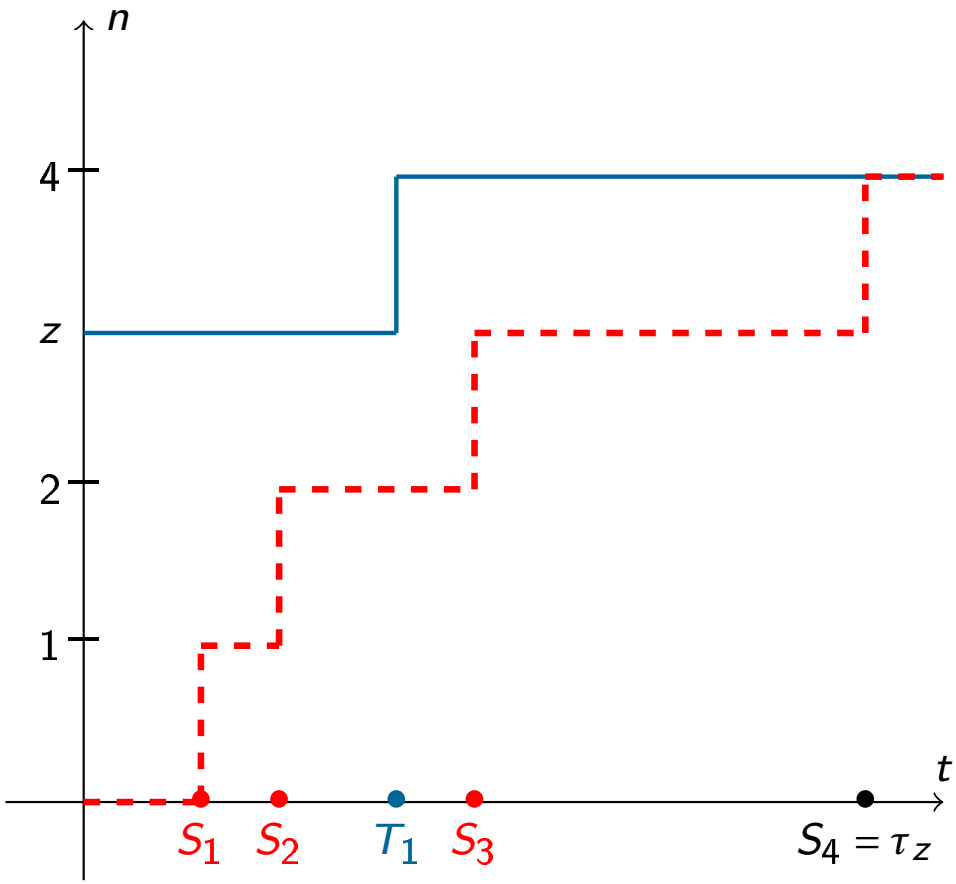
Let the length of honest and dishonest chain be driven by counting processes

- 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

$$\mathbb{P}(\tau_z < \infty) = \left(\frac{\mu}{\lambda}\right)^z, \ z \geq 0.$$



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.

Proof of Stake protocol

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

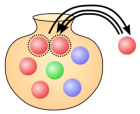
Two problems

- ⚠ Nothing at stake ⇒ Consensus postponed
- ⚠ Rich gets richer ⇒ Risk of centralization

Rich get richer?

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 $E = \{1, \ldots, p\}$
- 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.