Fraud risk assessment within blockchain transactions

Pierre-O. Goffard*

Department of Statistics and Applied Probability, University of California, Santa Barbara, USA.

February 23, 2018

Abstract

The probability of successfully spending twice the same bitcoins is considered. A double-spending attack consists in issuing two transactions transferring the same bitcoins. The first transaction, from the fraudster to a merchant, is included in a block of the public chain. The second transaction, from the fraudster to himself, is recorded in a block that integrates a private chain, exact copy of the public chain up to substituting the fraudster-to-merchant transaction by the fraudster-to-fraudster transaction. The double-spending hack is completed once the private chain reaches the length of the public chain, in which case it replaces it. The growth of both chains are modelled by two independent counting processes. The probability distribution of the time at which the malicious chain catches up with the honest chain, or equivalently the time at which the two counting processes meet each other, is studied. The merchant is supposed to await the discovery of a given number of blocks after the one containing the transaction before delivering the goods. This grants a head start to the honest chain in the race against the dishonest chain.

MSC 2010: 60G55, 60G40, 12E10.
Keywords: Bitcoin blockchain, double-spending problem, risk theory, order statistic point processes, renewal processes, boundary crossing problems.

1 Introduction

Bitcoin is a decentralized peer-to-peer (**P2P**) payment system that relies on Proof of Work (**PoW**). Electronic payments are performed by generating transactions that transfers Bitcoins (**BTCs**) between Bitcoin peers. These transactions are broadcast to a network of Bitcoin miners. These miners will compete to solve a cryptographic puzzle in order to build a block that contains the pending transactions. The first miner to solve the problem receives a certain number of **BTCs**, which is agreed upon by everyone in the network. At the time of the writing, this bounty is 12.5 **BTCs**; this value is halved every 210,000 blocks. The block is then included in the *blockchain* which plays the part of a public

*goffard@pstat.ucsb.edu

ledger recording all the transactions between bitcoin peers. Once a transaction enters the *blockchain*, it is considered validated. The only way to reverse the process, and for instance replace this transaction by another one, is to redo the work of the associated block and all the subsequent blocks. The *blockchain* allows in theory to prevent from double-spending the same **BTCs**. A double-spending attack consists in buying a good from a vendor and transfering the same bitcoins to oneself. Two conflicting transactions exist then in the network. The buyer-to-vendor transaction is included in the *blockchain* by the honest miners while a group of colluding miners work on their own private branch, exact replication of the principal chain up to substituting the buyer-to-vendor transaction by the buyer to buyer one. In the presence of two versions of the blockchain, the network always opt for the longest because more computational effort has been put into it. If the conspiring miners' chain ever becomes as long as the honest chain, it will replace it. The double-spending is then successful.

Satoshi Nakamoto stresses in his whitepaper [27] that a successful double-spending attack is rather unlikely as long as the pool of honest miners retain the majority of the computing power. The vendor is advised to wait for a certain number of blocks, say $z \in \mathbb{N}$, to be added to the chain before delivering the good. The derivation of the probability of a successful double-spending attack relies on an analogy with the one-sided gambler's ruin problem. Namely, the forthcoming block belongs to the honest chain with probability p or to the malicious chain with probability $q = 1 - p$. The difference between the length of the chains is then a random walk $\{Z_n,\, n \in \mathbb{N}\}$ on $\mathbb{Z}$ defined as

$$Z_n = z + Y_1 + \ldots + Y_n, \text{ for } n \in \mathbb{N}, \tag{1}$$

where the Y_i's are the independent and identically distributed (**i.i.d.**) steps of the random walk. Assuming that the honest miners have more resources implies that $p > q$, the probability that the malicious chain ever catches up with the honest one, given it is z blocks behind, is $(q/p)^z$. For a full treatment of the gambler's ruin problem, the reader is refered to the monograph of Asmussen and Albrecher [2].

The aim of this work is to refine the model underlying the double-spending problem. Counting processes are introduced to keep track of the number of blocks in the competing chains. These processes are generated by sequences of arrival times whose probability distribution reflect the block discovery frequency and the distribution of the computing power among honests and malicious miners. The probability distribution of the time at which the malicious chain overtakes (if it ever does) the honest chain is studied. Note that the probability mass function (**p.m.f.**) of the stopping time

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

in Nakamoto's framework is a consequence of the first hitting time theorem with

$$\mathbb{P}\left(\tau_z = n\right) = \frac{z}{n}\mathbb{P}(Z_n = 0), \text{ for } n \geq z, \tag{2}$$

see for instance Van Der Hofstad and Keane [19, Theorem 1] and the references therein.

Let $\{N(t)\ ,\, t \geq 0\}$ and $\{M(t)\ ,\, t \geq 0\}$ be two independent counting processes governing the block arrival over time in the honest and the malicious chain respectively. Assume

that the honest chain is $z \geq 1$ blocks ahead of the malicious chain at $t = 0$. Consider the stopping time

$$\tau_z = \inf\{t \geq 0 \ ; \ M(t) = z + N(t)\}. \tag{3}$$

at which the double spending attack is successful. Denote by $\{T_k \ , \ k \geq 1\}$ and $\{S_k \ , \ k \geq 1\}$ the block arrival times in the honest and malicious chain respectively. Figure 1 illustrates the race between the two processes. The distribution of τ_z is studied for different

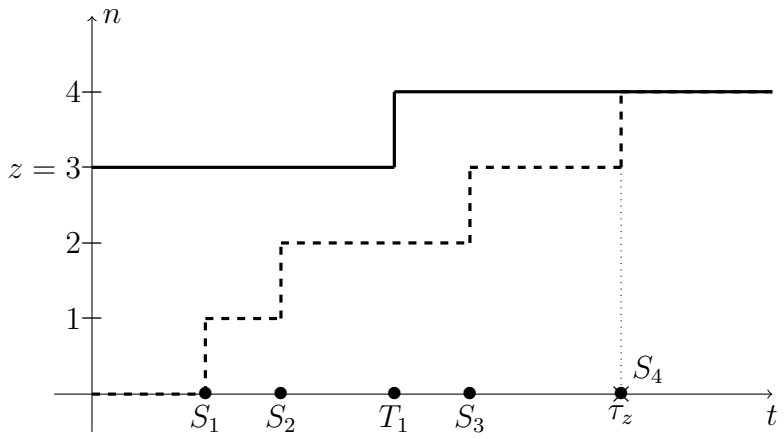


Figure 1: Time until the double-spending attack is completed: (solid) length of the honest chain $\{z + N(t) \ , \ t \geq 0\}$, (dashed) length of the malicious chain $\{M(t) \ , \ t \geq 0\}$.

sets of assumptions over the counting processes $\{N(t) \ , \ t \geq 0\}$ and $\{M(t) \ , \ t \geq 0\}$.

In Section 2 and 3, the length of the honest chain $\{z + N(t) \ , \ t \geq 0\}$ is driven by an Order Statistic Point Process (**OSPP**), that is, provided that $N(t) = n$, the jump times $T_1, \ldots, T_n$ have the same distribution as the order statistics of a sample of n **i.i.d.** random variables concentrated on $[0, t]$ with distribution function F_t. In Section 2, the probability density function (**p.d.f.**) of τ_z is derived in terms of Abel-Gontcharov (**A-G**) polynomials when the length of the malicious chain $\{M(t) \ , \ t \geq 0\}$ is a renewal process (**i.e.** generated by **i.i.d.** inter-arrival times). In Section 3, the survival function (**s.f.**) of τ_z is expressed in terms of Appell polynomials in the case where $\{M(t) \ , \ t \geq 0\}$ is an **OSPP**.

The probability of a successful double-spending attempt, defined by $\mathbb{P}(\tau_z < \infty)$, is further considered. An upper-bound is derived in Section 4 when both $\{N(t) \ , \ t \geq 0\}$ and $\{M(t) \ , \ t \geq 0\}$ are renewal processes. An exact expression is obtained when $\{N(t) \ , \ t \geq 0\}$ is a Poisson process.

Nakamoto [27] did not state explicitly that the block arrival is dictated by a homogeneous Poisson process. However, the probability of a successful double-spending attack, as we will see later, when the length of the public and private chains are governed by two Poisson processes of intensity λ and μ, is given by $(\mu/\lambda)^z$. Hence, the intensities plays the same role as p and q as they reflect the hashrate of the miners. Every single result given in this work holds when the rival chains are modeled by homogeneous Poisson process because the homogeneous Poisson process is the one and only renewal process enjoying the order statistic property. The formulas, which may seems involved in the general cases, simplifies

when the arrival of blocks is Poisson. Now, is it worth considering more sophisticated models to track the growth of the blockchains?

A statistical study over the inter-block times has been conducted in a recent work of Bowden et al. [8]. The authors collected the timestamp information in the header of the blocks. As pointed out in [8], the timestamp information cannot be readily used, and preprocessing it represents quite a challenge in itself. The data after preprocessing is available on Rhys Bowden `Github` repository [7]. The empirical mean of the interblock time is 9.41 minutes while the standard deviation is of 11.05 minutes. The high variance of the inter-block time is a known flaw which impedes the consistent flow of validated transactions. Opting for a renewal process instead of a Poisson process allows to let the inter-block time have a two-parameters distribution, gamma or Weibull say, and therefore account for the second order moment in the statistical inference. To the time required for the creation of a block is sometimes added a propagation delay. A newly discovered block is appended to the chain only once the word about that block has been spread to the entire network. If it is accepted that the block mining time is an exponential random variable, the actual time at which the block integrates the chain is an exponential random variable perturbed by another non-negative noise. The result may or may not be exponentially distributed. One of the many takeaways of Bowden et al. [8] is that the rate at which blocks are discovered varies over time according to the adjustment of the cryptopuzzles difficulty. The authors of [8] proposed different models allowing for a time-dependent discovery rate with deterministic or even random difficulty adjustment. When the difficulty adjustment is deterministic then a non-homogeneous Poisson process is suitable. This is fortunate as the non-homogeneous Poisson process is a particular instance of **OSPP**. The point processes having the order statistic property have been characterized a while ago by Puri [31]. The **OSPPs** are either mixed Poisson processes up to a timescale transformation or mixed sample processes. This class encompasses classical point processes such as the mixed Poisson process, the non-homogeneous Poisson process, the linear birth process with immigration and the linear death process.

From a mathematical standpoint, this work ressembles an early work of Picard and Lefèvre [30] where the probability distribution of the first *rendez-vous* time between two counting processes is derived. The formulas already involved the Appell and **A-G** polynomials and even extend to the case of compound processes. The definition of the stopping time is slightly different in the case considered here. Plus, the reasoning differs as it relies extensively on the order statistic property and the connection between the aforementionned families of polynomials and the order statistics joint distribution. It is more in the spirit of Goffard and Lefèvre [16] where the crossing of an **OSPP** through a moving boundary is under study. These arguments are inspired from risk theory when solving the ruin problem in ordered risk models, see **e.g.** Lefèvre and Picard [23, 24], Ignatov and Kaishev [20], Dimitrova et al. [9], Goffard and Lefèvre [17], and Goffard [14]. To the best of my knowledge, the closest link to queueing theory is the single server queue with either work or customer removal introduced by Gelenbe et al. [13] and further considered in Boucherie and Boxma [5], Jain and Sigman [21], and Harrison and Pitel [18] for different queues. The related risk process includes lump addition, see Boucherie et al. [6]. In Perry et al. [28, 29], renewal-type arguments are used to study the distribution of boundary crossing times of the difference between two Poisson processes and linear boundaries. Regarding

the evaluation of the probability of a successful double-spending attack, $\mathbb{P}(\tau_z < \infty)$, the first step consists in swapping the role of time and space. Namely, a correspondence is established between the ruin times of two risk models with inverted characteristics. This trick is now standard in risk theory, see for instance Mazza and Ruillère [26], Dickson and Borovkov [4], Shi and Landriault [33], Dimitrova et al. [10], and Goffard and Lefèvre [17]. A classical Martingale approach allows then to derive an expression of the probability of successfully spending twice the sames **BTCs**.

The rest of the paper is organized as follows, in Section 2 a formula for the **p.d.f.** of τ_z when $\{N(t)\ ,\ t \geq 0\}$ is an **OSPP** and $\{M(t)\ ,\ t \geq 0\}$ is a renewal process is derived in terms of Abel-Gontcharov polynomials. In Section 3, a formula for the **s.f.** of τ_z when both $\{N(t)\ ,\ t \geq 0\}$ and $\{M(t)\ ,\ t \geq 0\}$ are **OSPPs** is provided in terms of Appell polynomials. Section 4 is concerned with the probability of the double-spending attack being ever successful. Section 5 is devoted to numerical illustrations.

2 The p.d.f. of the double-spending time

In this section, the length of the honest chain $\{z+N(t)\ ,\ t \geq 0\}$ is governed by an **OSPP**. Its jump times, provided that $N(t) = n$, have the same joint distribution as a vector of order statistics. Namely, it holds that

$$[T_1, \dots, T_n | N(t) = n] \overset{D}{=} (V_{1:n}, \dots, V_{n:n}), \tag{4}$$

where $\overset{D}{=}$ stands for equality in distribution and $V_{1:n}, \dots, V_{n:n}$ correspond to the order statistics of n **i.i.d.** random variables having a **c.d.f.** $F_t(s)$, for $0 \leq s \leq t$. The length of the malicious chain is a renewal process generated by a sequence of **i.i.d.** inter-arrival times $\{\Delta_k^S\ ,\ k \geq 1\}$ having a **p.d.f.** denoted by f_{Δ^S}. The sequence of arrival times $\{S_n\ ,\ n \in \mathbb{N}\}$, with the convention $S_0 = 0$, corresponds to the partial sums of the inter-arrival times sequence. The **p.d.f.** of S_n, for $n \in \mathbb{N}$, is given by

$$f_{S_n}(t) = f_{\Delta^S}^{*n}(t), \text{ for } t \geq 0, \tag{5}$$

where f^{*n} denote the n-fold convolution of f_{Δ^S} with itself. Let $z \geq 1$ be an integer, the following result gives a formula for the **p.d.f.** of

$$\tau_z = \inf\{t \geq 0\ ;\ M(t) = N(t) + z\},$$

time at which the double-spending attack is completed.

Theorem 1. *If $\{N(t)\ ,\ t \geq 0\}$ is an **OSPP** and $\{M(t)\ ,\ t \geq 0\}$ is a renewal process then the **p.d.f.** of τ_z is given by*

$$f_{\tau_z}(t) = \mathbb{E}\left[(-1)^{N(t)} h_{N(t)}(t, z) f_{\Delta^S}^{*[N(t)+z]}(t)\right],\ t \geq 0, \tag{6}$$

where

$$h_n(t, z) = \mathbb{E}\left\{G_n\left[0 \middle| F_t(S_z), \dots, F_t(S_{n+z-1})\right] \middle| S_{n+z} = t\right\}, \tag{7}$$

*and $G_n(0|.)$ is an **A-G** polynomial such as defined in the Appendix A.*

Proof. The event $\{\tau_z \in (t, t+\mathrm{d}t)\}$, for $t \geq 0$, corresponds to the exact time at which the double-spending attack is successful as the malicious chain takes over the honest one. At time $t = 0$, the honest chain is ahead by $z \geq 1$ blocks. Assuming that later, at time $t > 0$, the honest miners manage to add $N(t) = n \in \mathbb{N}$ blocks to the chain then the malicious chain must be of length $M(t^-) = n + z - 1$ at some time $t^- < t$ and jump to the level $n + z$ exactly at t. Conditioning over the values of $\{N(t)\ ,\ t \geq 0\}$ yields

$$\{\tau_z \in (t, t+\mathrm{d}t)\} = \bigcup_{n=0}^{+\infty} \{\tau_z \in (t, t+\mathrm{d}t)\} \cap \{N(t) = n\}. \tag{8}$$

In the case where $N(t) = 0$, the only requirement is that the z^{th} jump of $\{M(t)\ ,\ t \geq 0\}$ occurs at time t. It then follows that

$$\{\tau_z \in (t, t+\mathrm{d}t)\} \cap \{N(t) = 0\} = \{S_z \in (t, t+\mathrm{d}t)\} \cap \{N(t) = 0\}, \tag{9}$$

and consequently

$$f_{\tau_z | N(t)=0}(t) = f^{*z}_{\Delta^S}(t),\ t \geq 0, \tag{10}$$

where $f_{\tau_z | N(t)=0}(t)$ denotes the conditional **p.d.f.** of τ_z given that $N(t) = 0$. On the set $\{N(t) \geq 1\}$, one needs to make sure that $\{M(t)\ ,\ t \geq 0\}$ behaves properly by constraining its jump times so that it does not reach $N(s) + z$ at any time $s < t$ and performs the $(n+z)^{\text{th}}$ jump at t. Hence, it holds that

$$\{\tau_z \in (t, t+\mathrm{d}t)\} \cap \{N(t) \geq 1\} = \bigcup_{n=1}^{+\infty} \bigcap_{k=1}^{n} \{T_k \leq S_{z+k-1}\} \cap \{S_{z+n} \in (t, t+\mathrm{d}t)\} \cap \{N(t) = n\}.$$

Applying the law of total probability yields

$$\begin{aligned} & \mathbb{P}\left(\{\tau_z \in (t, t+\mathrm{d}t)\} \cap \{N(t) \geq 1\}\right) \\ = & \sum_{n=1}^{+\infty} \mathbb{P}\left[\bigcap_{k=1}^{n} \{T_k \leq S_{z+k-1}\} \cap \{S_{z+n} \in (t, t+\mathrm{d}t)\} \Big| N(t) = n\right] \mathbb{P}[N(t) = n]. \end{aligned} \tag{11}$$

In virtue of the order statistic property, the successive jump times $(T_1, \ldots, T_n)$ are distributed as the order statistics $(V_{1:n}, \ldots, V_{n:n})$ of a sample of n **i.i.d.** random variables with **c.d.f.** $F_t(s),\ 0 \leq s \leq t$. The conditional probability in (11) may be rewritten as

$$\begin{aligned} & \mathbb{P}\left[\bigcap_{k=1}^{n} \{V_{k:n} \leq S_{z+k}\} \cap \{S_{z+n+1} \in (t, t+\mathrm{d}t)\}\right] \\ = & \mathbb{P}\left[\bigcap_{k=1}^{n} \{U_{k:n} \leq F_t(S_{z+k-1})\} \cap \{S_{z+n} \in (t, t+\mathrm{d}t)\}\right] \\ = & \mathbb{P}\left[\bigcap_{k=1}^{n} \{U_{k:n} \leq F_t(S_{z+k-1})\} \big| S_{z+n} \in (t, t+\mathrm{d}t)\right] \mathbb{P}[S_{z+n} \in (t, t+\mathrm{d}t)] \\ = & \mathbb{E}\left\{(-1)^n G_n[0 \big| F_t(S_z), \ldots, F_t(S_{z+n-1})] \big| S_{z+n} \in (t, t+\mathrm{d}t)\right\} \mathbb{P}[S_{z+n} \in (t, t+\mathrm{d}t)], \end{aligned} \tag{12}$$

where $U_{1:n}, \ldots, U_{n:n}$ denote the order statistics of a sample of n **i.i.d.** uniform random variables on $[0,1]$, and $G_n(.|.)$ corresponds to the **A-G** polynomials as defined in the Appendix A. Inserting (12) into (11) and letting $\mathrm{d}t$ be small enough yields

$$\begin{aligned} f_{\tau_z|N(t)\geq 1}(t) &= \sum_{n=1}^{+\infty} \mathbb{E}\left\{(-1)^n G_n[0\big|F_t(S_z),\ldots,F_t(S_{z+n-1})]\big|S_{z+n}=t\right\} \\ &\times \; f_{S_{z+n}}(t)\mathbb{P}[N(t)=n]. \end{aligned}$$

The final step consists in noting that $G_0(x)=1$ for every $x\in\mathbb{R}$, and writing

$$f_{\tau_z}(t) = \sum_{n=0}^{+\infty} \mathbb{E}\left\{(-1)^n G_n[0\big|F_t(S_z),\ldots,F_t(S_{z+n-1})]\big|S_{z+n}=t\right\} f_{\Delta^S}^{*(z+n)}(t)\mathbb{P}[N(t)=n], \tag{13}$$

which is equivalent to the announced result (6). □

The stopping time τ_z may be interpreted as the first meeting time of the **OSPP** $\{N(t),\ t\geq 0)\}$ and the lower randomized boundary defined by $\{M(t)-z,\ t\geq 0\}$. This remarks explains why Theorem 1 is reminiscent of the results given in Goffard and Lefèvre [16, Proposition 3.1] and Goffard and Lefèvre [17, Theorem 3.1], where the first-meeting problems of an **OSPP** with a lower deterministic boundary are handled. The numerical evaluations of (6), to compute for instance the probability that the double-spending attack succeed within a fixed time period, looks challenging. A method based on the truncation of the infinite series in (13) coupled with a numerical integration routine, in the same vein as what is proposed in Borovkov and Dickson [4], can be put together. The next result shows how formula (6) may be simplified in the case where $\{N(t)\ ,\ t\geq 0\}$ is a mixed Poisson process.

Corollary 1. *If $\{N(t)\ ,\ t\geq 0\}$ is a mixed Poisson process then the **p.d.f.** of τ_z is given by*

$$f_{\tau_z}(t) = \mathbb{E}\left[\frac{z}{z+N(t)} f_{\Delta^S}^{*[N(t)+z]}(t)\right],\ \textit{for } t\geq 0. \tag{14}$$

Proof. As $\{N(t),\ t\geq 0\}$ is a mixed Poisson process then we can apply Theorem 1 replacing $F_t(s)$ by s/t for $s\leq t$. The function $h_n(t,z)$ defined in (7) becomes

$$h_n(t,z) = \mathbb{E}\left\{\frac{1}{t^n} G_n\left(0\big|S_z,\ldots,S_{n+z-1}\right)\Big|S_{n+z}=t\right\},$$

after applying the identity (77). Conditioning upon the values of S_z, and applying successively the identities (77) and (79) leads to

$$\begin{aligned} h_n(t,z) &= \frac{1}{t^n}\mathbb{E}\left\{\mathbb{E}\left[G_n\left(0\big|S_z,\ldots,S_{n+z-1}\right)\Big|S_z,S_{n+z}\right]\Big|S_{n+z}=t\right\} \\ &= \frac{1}{t^n}\mathbb{E}\left\{\mathbb{E}\left[G_n\left(-S_z\big|0,\Delta_{z+1}^S\ldots,\sum_{k=1}^{n-1}\Delta_{z+k}^S\right)\Big|\sum_{k=1}^{n}\Delta_{z+k}^S = S_{n+z}-S_z\right]\Big|S_{n+z}=t\right\} \\ &= \frac{1}{t^n}\mathbb{E}\left\{(-S_z)[-S_z-(S_{n+z}-S_z)]^{n-1}\Big|S_{n+z}=t\right\} \\ &= \frac{(-1)^n}{t^n}\mathbb{E}\left[S_z(S_{z+n})^{n-1}\Big|S_{n+z}=t\right] \\ &= (-1)^n\frac{z}{n+z}. \end{aligned} \tag{15}$$

Substituting (15) into (6) yields the announced result (14). □

The formula given in Corollary (1) is reminiscent of the first-hitting time theorem and also the so-called Kendall identity, see for instance Borovkov and Burq [3], which gives the **p.d.f.** of the first-meeting time of a spectrally negative Lévy process with a lower linear boundary. The following example gives rise to an expression for the **p.d.f.** of τ_z that allows the evaluation of the probability of a successful double-spending attack $\mathbb{P}(\tau_z < \infty)$.

Example 1. *Assume that the length of the chains* $\{z+N(t)\ ,\ t \geq 0\}$ *and* $\{M(t)\ ,\ t \geq 0\}$ *are governed by two homogeneous Poisson processes of intensity* λ *and* μ *respectively. The inter-arrival times* $\{\Delta_k^S,\ k \geq 1\}$ *are* ***i.i.d.*** *exponential random variables with parameter* μ *and associated* ***p.d.f.***

$$f_{\Delta^S}(x) = \mu e^{-\mu x},\ \textit{for } x \geq 0. \tag{16}$$

Applying Corollary 1 yields, after a couple of rearrangements, the following expression for the ***p.d.f.*** *of* τ_z*,*

$$f_{\tau_z}(t) = \sum_{n=0}^{+\infty} \left(\frac{z}{z+n}\right) \binom{2n+z-1}{n} \left(\frac{\mu}{\mu+\lambda}\right)^{n+z} \left(\frac{\lambda}{\mu+\lambda}\right)^{n} \frac{(\lambda+\mu)^{2n+z} t^{2n+z-1} e^{-t(\mu+\lambda)}}{\Gamma(2n+z)}, \tag{17}$$

for $t \geq 0$*. Assuming that* $\lambda > \mu$ *and integrating* (17) *with respect to* t *yields the probability of successful double-spending attack with*

$$\begin{aligned}
\mathbb{P}(\tau_z < \infty) &= \sum_{n=0}^{+\infty} \left(\frac{z}{z+n}\right) \binom{2n+z-1}{n} \left(\frac{\mu}{\mu+\lambda}\right)^{n+z} \left(\frac{\lambda}{\mu+\lambda}\right)^{n} \\
&= \left(\frac{\lambda+\mu}{\lambda}\right)^{z} z \sum_{n=0}^{+\infty} \binom{2n+z-1}{n} \left(\frac{1}{z+n}\right) \left[\frac{\lambda\mu}{(\mu+\lambda)^2}\right]^{n+z} \\
&= \left(\frac{\lambda+\mu}{\lambda}\right)^{z} z \sum_{n=0}^{+\infty} \binom{2n+z-1}{n} \int_0^{\frac{\lambda\mu}{(\mu+\lambda)^2}} t^{n+z-1}\, dt \\
&= \left(\frac{\lambda+\mu}{\lambda}\right)^{z} z \sum_{n=0}^{+\infty} \binom{2n+z-1}{n} \int_0^{\frac{\lambda\mu}{(\mu+\lambda)^2}} t^{n+z-1}\, dt \\
&= \left(\frac{\lambda+\mu}{\lambda}\right)^{z} z \int_0^{\frac{\lambda\mu}{(\mu+\lambda)^2}} t^{z-1} \sum_{n=0}^{+\infty} \binom{2n+z-1}{n} t^n\, dt \\
&= \left(\frac{\lambda+\mu}{\lambda}\right)^{z} z \int_0^{\frac{\lambda\mu}{(\mu+\lambda)^2}} t^{z-1} \frac{C(t)^{z-1}}{\sqrt{1-4t}}\, dt,
\end{aligned} \tag{18}$$

where

$$C(t) = \frac{1-\sqrt{1-4t}}{2t}, \tag{19}$$

denotes the generating function of Catalan's numbers, see for instance Aigner [1, Chapter 3]. Note that the last equality follows from an exercise in the textbook of Aigner [1, Exercise 3.25]. Inserting (19) *into* (18) *yields, after straightforward integration,*

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

The result given above is consistent with Corollary 3, see Section 4.

3 The s.f. of the double-spending time

In this section, the length of the honest and the malicious chains are governed by two independent **OSPPs**. The order statistic property satisfied by $\{M(t)\ ,\ t \geq 0\}$ implies that

$$[S_1, \ldots, S_m | M(t) = m] \stackrel{D}{=} (V^*_{1:m}, \ldots, V^*_{m:m}),$$

where $V^*_{1:m}, \ldots, V^*_{m:m}$ denote the order statistics of a sample $V^*_1, \ldots, V^*_m$ of m **i.i.d.** random variables having a **c.d.f.** $F^*_t(s)$, for $0 \leq s \leq t$. Note that, regarding $\{N(t)\ ,\ t \geq 0\}$, the notation of Section 2 is preserved. The following result gives a formula in terms of Appell polynomials for the **s.f.** of $\tau_z = \inf\{t \geq 0\ ,\ M(t) = N(t) + z\}$.

Theorem 2. *If $\{N(t)\ ,\ t \geq 0\}$ and $\{M(t)\ ,\ t \geq 0\}$ are two **OSPPs** then the **s.f.** of τ_z is given by*

$$\mathbb{P}(\tau_z > t) = \mathbb{E}\left(A_{M(t)}\left\{1 \middle| 0, \ldots, 0, F^*_t[V_{1:N(t)}], \ldots, F^*_t[V_{M(t)+1-z:N(t)}]\right\} \mathbb{I}_{M(t) \leq N(t)+z-1}\right), \tag{20}$$

for $t \geq 0$, where $A_n(1|.)$ is an Appell polynomial such as defined in the Appendix A.

Proof. If $M(t) \geq N(t) + z$, at time $t \geq 0$, then the double-spending attack already occured. Consider the event $\{\tau_z > t\}$, conditioning upon the possible values of the counting processes leads to

$$\{\tau_z > t\} = \bigcup_{n=0}^{+\infty} \bigcup_{m=0}^{n+z-1} \{\tau_z > t\} \cap \{N(t) = n\} \cap \{M(t) = m\}. \tag{21}$$

The double-spending attack did not happen before time $t \geq 0$ if $M(t)$ is smaller or equal to $z-1$, irrespective of the value of $N(t)$. If $M(t)$ falls between z and $N(t)+z$ then $N(t)$ must have jumped, at least once, and an investigation over the jump times of both point processes must be conducted. The event $\{\tau_z > t\}$ is further rewritten as

$$\begin{aligned} \{\tau_z > t\} &= \bigcup_{m=0}^{z-1} \{M(t) = m\} \\ &\cup \bigcup_{n=1}^{+\infty} \bigcup_{m=z}^{n+z-1} \bigcap_{k=z}^{m} \{S_k > T_{k+1-z}\} \cap \{N(t) = n\} \cap \{M(t) = m\}. \end{aligned} \tag{22}$$

The law of total probability yields

$$\begin{aligned} \mathbb{P}(\tau_z > t) &= \mathbb{P}[M(t) \leq z-1] \\ &+ \sum_{n=1}^{+\infty} \sum_{m=z}^{n+z-1} \mathbb{P}\left[\bigcap_{k=z}^{m} \{S_k > T_{k+1-z}\} \middle| N(t) = n, M(t) = m\right] \\ &\times \mathbb{P}[N(t) = n, M(t) = m]. \end{aligned} \tag{23}$$

Now, by the order statistic property, it holds that

$$[(T_1, \ldots, T_{m+1-z}) | N(t) = n] \stackrel{D}{=} (V_{1:n}, \ldots, V_{m+1-z:n})$$

and

$$[(S_z, \ldots, S_m) | M(t) = m] \stackrel{D}{=} (V^*_{z:m}, \ldots, V^*_{m:m}).$$

Therefore, the conditional probability in (23) may be rewritten as

$$\begin{aligned}\mathbb{P}\left(\bigcap_{k=z}^{m}\{V^*_{k:m} > V_{k+1-z:n}\}\right) &= \mathbb{P}\left[\bigcap_{k=z}^{m}\{U_{k:m} > F^*_t(V_{k+1-z:n})\}\right] \\ &= A_m\left[1\middle|0,\ldots,0,F^*_t(V_{1:n}),\ldots,F^*_t(V_{m+1-z:n})\right], \quad (24)\end{aligned}$$

where $U_{1:m},\ldots,U_{m:m}$ are the order statistics of a sample of m **i.i.d.** uniform random variable on $(0,1)$ and $A_m(.|.)$ denote the Appell polynomials defined in the Appendix A. Inserting (24) into (23) yields

$$\begin{aligned}\mathbb{P}(\tau_z > t) &= \mathbb{P}[M(t) \leq z] \\ &+ \sum_{n=1}^{+\infty}\sum_{m=z}^{n+z-1} A_m\left[1\middle|0,\ldots,0,F^*_t(V_{1:n}),\ldots,F^*_t(V_{m+1-z:n})\right] \\ &\times \mathbb{P}[N(t)=n, M(t)=m], \quad (25)\end{aligned}$$

which is the same as (20) after noticing that $A_m(1|0,\ldots,0)=1$ for every $m\in\mathbb{N}$. □

In Subsection 5.2, a Monte Carlo evaluation of the expectation of (20) is performed. This type of estimator has been studied in Goffard and Lefèvre [16, Section 6] and named Appell Polynomial Monte Carlo (**APMC**). The procedure entails a variance reduction in comparison to a crude Monte Carlo evaluation. The following result shows how formula (20) simplifies remarkably for $z=1$, when the **OSPPs** are similar in a sense detailed below.

Corollary 2. *Assume that $z=1$. If $\{N(t)\ ,\ t\geq 0\}$ and $\{M(t)\ ,\ t\geq 0\}$ are two **OSPPs** such that $F_t(s)=F^*_t(s)$ for every $s\leq t$ then the **s.f.** of τ_z is given by*

$$\mathbb{P}(\tau_z > t) = \mathbb{E}\left(\frac{N(t)-M(t)+1}{N(t)+1}\mathbb{I}_{M(t)\leq N(t)}\right), \text{ for } t\geq 0. \quad (26)$$

Proof. Let $\{N(t)\ ,\ t\geq 0\}$ and $\{M(t)\ ,\ t\geq 0\}$ be two **OSPPs** such that $F_t(s)=F^*_t(s)$ for every $s\leq t$. Applying Theorem 2, with $z=1$, yields

$$\mathbb{P}(\tau_z > t) = \mathbb{E}\left(A_{M(t)}\left\{1\middle|U_{1:N(t)},\ldots,U_{M(t):N(t)}\right\}\mathbb{I}_{M(t)\leq N(t)}\right), \text{ for } t\geq 0.$$

Recall the probabilistic interpretation of the Appell polynomial in Proposition 1 with

$$\mathbb{P}(\tau_z > t) = \mathbb{E}\left(\mathbb{P}\left[U^*_{1:M(t)} > U_{1:N(t)},\ldots,U^*_{M(t):M(t)} > U_{M(t):N(t)}\right\}\mathbb{I}_{M(t)\leq N(t)}\right), \text{ for } t\geq 0.$$

Applying Bertrand's ballot theorem, allowing for ties, yields the announced result (26). □

The case treated in the numerical illustrations considers that the length of the chains are governed by two non-homogeneous Poisson processes, which is consistent with the empirical study conducted in Bowden et al. [8], as explained in the following example.

Example 2. *Bowden et al. [8] recommend to model the arrival of blocks as a non-homogeneous Poisson process with an intensity function designed to capture the evolution of the global hashrate on one hand and the difficulty adjustment of the cryptopuzzles on*

the other hand. The block number n is associated to a hash $f(n)$ which is a number drawn randomly from the lattice $\{0, 1, \dots, 2^{256} - 1\}$. Mining a block consists in computing the hash of the block until it is lower than a target L. The number of trial required is then a geometric random variable $Geom(L \times 2^{-256})$ with associated ***p.m.f.*** *$(1 - L \times 2^{-256})^{k-1} L \times 2^{-256}$ for $k \geq 1$. The difficulty is adjusted by tuning the target L. Denote by $\{T_k \ , \ k \geq 1\}$ the sequence of arrival time of the blocks. The difficulty is adjusted every $2,016$ blocks to maintain an average 1 block mined every 10 minutes So, mining $2,016$ takes about 2 weeks. This leads to the definition of piecewise constant target function $L(t)$ as*

$$L(t) = \begin{cases} L_0, & \text{for } t \in (0, T_{2016}) \\ L_k, & \text{for } t \in (T_{2016k}, T_{2016(k+1)}) \text{ and } k > 0, \end{cases} \tag{27}$$

where the sequence of real numbers $\{L_k \ , \ k \geq 0\}$ is defined recursively as

$$L_k = \begin{cases} 2^{224}, \text{ for } k = 0, \\ L_{k-1} \times \frac{T_{2016k} - T_{2016(k-1)}}{1209600} \text{ for } k > 0. \end{cases}$$

Note that the time unit is the second and $1,209,600$ seconds correspond to 2 weeks. The number of trials relates to the mining time through the (global) hashrate function $H(t)$. The hashrate function corresponds to the number of hashes computed per second by the entire network of miners. Hence, the instantaneous average mining time is given by $\frac{2^{256}}{H(t)L(t)}$ and the intensity function of the underlying non-homogeneous Poisson process is given by

$$\lambda(t) = \frac{H(t)L(t)}{2^{256}}, \text{ for } t \geq 0. \tag{28}$$

There are two main drivers of the hashrate. First, the improvement of the mining machines which enhances the computing power of the miners. Second, the number of miners in the network. The miners enter and exit the network according to how profitable mining ***BTCs*** *is at the moment. This last factor depends on the price of the electricity and the value of the* ***BTCs*** *at a given point in time. The target $L(t)$ is an information that can be collected from the header of the block. The hashrate $H(t)$ is retrieved from the knowledge of the difficulty and the timestamp data. The authors of [8] proposed an exponential function of the form $H(t) = e^{at+b}$ arguing that the log hashrate is piecewisely linear over time. The values of a and b follow from the linear interpolation within successive time periods. Once the hashrate function has been determined, the length of the blockchain $\{N(t) \ , \ t \geq 0\}$ is a non-homogeneous Poisson process with intensity function $\lambda(t)$ defined in (28). Assuming that $N(t) = n$, the arrival times $T_1, \dots, T_n$ are distributed as the order statistics of n* ***i.i.d.*** *random variables with associated* ***c.d.f.***

$$F_t(s) = \frac{\Lambda(s)}{\Lambda(t)} \ , \text{ for } s \leq t, \tag{29}$$

where $\Lambda(t) = \int_0^t \lambda(s) ds$. In the event of double-spending attack, the difficulty of the puzzle is the same for the honest miners and the colluding miners. The difference between the two pools lies in their computing power and thus their hashrate function. We may assume that both the honest and dishonest miners contribute to the global hashrate of the network in an additive way. More specifically, let $H_1(t) = p \times H(t)$ be the hashrate of the honest miners

and $H_2(t) = (1-p) \times H(t)$*, where* $p \in (0,1)$ *represents the repartition of the computing resources among the miners. Theorem 2 is applicable and formula* (20) *simplifies to*

$$\mathbb{P}(\tau_z > t) = \mathbb{E}\left(A_{M(t)}\left\{1\middle|0,\ldots,0,U_{1:N(t)},\ldots,U_{M(t)+1-z:N(t)}\right\}\mathbb{I}_{M(t)\leq N(t)+z-1}\right), \tag{30}$$

because $F_t^*(s) = F_t(s)$ *for every* $s \leq t$*. An evaluation via Monte-carlo simulation is possible by generating values for* $M(t)$*,* $N(t)$*, and* $U_{1:N(t)},\ldots,U_{M(t)+1-z:N(t)}$*. Appell polynomials do not usually admit a closed-form expression but can be computed recursively via the relations provided in the appendix A, see Proposition 2.*

Note that the evaluation of (20) may be achieved through the truncation of the infinite series in (25) followed by numerical integration, in the same vein as what is done in Dimitrova et al. [11]. Another solution would be to resort to a fully recursive evaluation as in Lefèvre and Loisel [22].

4 The probability of a successful double-spending attack

In this section, we study the probability of a successful double-spending attack, $\mathbb{P}(\tau_z < +\infty)$, when the length of the chains $\{z+N(t)\ ,\ t \geq 0\}$ and $\{M(t)\ ,\ t \geq 0\}$ are modelled by independent renewal processes generated by their respective sequence of **i.i.d.** inter-arival times denoted by $\{\Delta_k^T,\ k \geq 1\}$ and $\{\Delta_k^S,\ k \geq 1\}$. Assume that

(A1) $\mathbb{E}(\Delta^S) > \mathbb{E}(\Delta^T)$,

(A2) The equation

$$\log \mathbb{E}\left[e^{\theta(\Delta^T - \Delta^S)}\right] = 0, \tag{31}$$

has a unique non-negative solution denoted by γ, refered to as the adjustment coefficient.

The stopping time $\tau_z = \inf\{t \geq 0;\ M(t) = z + N(t)\}$ coincides with the ruin time $\tau_z = \inf\{t \geq 0;\ R(t) = 0\}$ associated to the risk process

$$R(t) = z + N(t) - M(t),\ t \geq 0. \tag{32}$$

Define the claim surplus process as

$$S(t) = M(t) - N(t),\ t \geq 0. \tag{33}$$

In risk theory, processes such as (32) model the evolution of the net worth of an insurance company over time. Here, the insurance company holds an initial capital of amount z, its premium income is governed by $\{N(t)\ ,\ t \geq 0\}$ while $\{M(t)\ ,\ t \geq 0\}$ corresponds to its liability at time $t \geq 0$. Note that I am only using risk theory terminology to improve the presentation, I am not claiming that one should model the evolution of the financial reserves of any non-life insurance company via (32). When studying the distribution of the ruin time is problematic, a simple trick consists in passing to a dual risk model. This approach is rather standard (see the references given in the introduction). For the sake of clarity, the idea is recalled hereafter and illustrated by Figure 2. Figure 2(a) displays the

ruin problem in model (32). The initial ruin problem is converted into another equivalent ruin problem. Increment the value of $\{M(t),\ t \geq 0\}$ by one unit and consider the risk model

$$\tilde{R}(t) = z + N(t) - [M(t) + 1],\ t \geq 0. \tag{34}$$

Further define the ruin time

$$\tilde{\tau}_z = \inf\{t \geq 0\ ;\ z + N(t) < [M(t) + 1]\}, \tag{35}$$

which coresponds to the first-crossing time of $\{M(t) + 1\ ,\ t \geq 0\}$ through the upper boundary $\{N(t) + z\ ,\ t \geq 0\}$, see Figure 2(b). It holds that

$$\tau_z \overset{\text{a.s.}}{=} \tilde{\tau}_z, \tag{36}$$

where $\overset{\text{a.s.}}{=}$ stands for equality almost sure as it is true for every trajectory. Then rotate Figure 2(b) by 90° anticlockwise to get Figure 2(c). Shifting the origin from $(0, 0)$ to $(z-1, 0)$ finally leads to Figure 2(d). The ruin problem displayed on Figure 2(d) concerns a discrete time risk model denoted by $\{R^*(n)\ ,\ n \geq 1\}$ and defined as

$$R^*(n) = S_{z-1} + \sum_{k=1}^{n} \left(\Delta^S_{k+z-1} - \Delta^T_k\right),\ \text{for } n \in \mathbb{N}. \tag{37}$$

The initial capital is S_{z-1}, the sequence $\{\Delta^S_{k+z-1}\ ,\ k \geq 1\}$ models the premium collected at each time period, and the sequence $\{\Delta^T_k\ ,\ k \geq 1\}$ represents the total claim amounts incurred during each time period. The conventions $S_0 = 0$ and $T_0 = 0$ are adopted. The claim surplus process $\{S^*(n)\ ,\ n \geq 1\}$ is given by

$$S^*(n) = \sum_{k=1}^{n} \left(\Delta^T_k - \Delta^S_{k+z-1}\right), \tag{38}$$

The ruin time is defined as $\sigma_{S_{z-1}} = \inf\{n \in \mathbb{N}\ ;\ R^*(n) \leq 0\}$ and relates to $\tau_z = \inf\{t \geq 0\ ;\ R(t) = 0\}$ as

$$\tau_z \overset{\text{a.s.}}{=} \tilde{\tau}_z \overset{\text{a.s.}}{=} S_{\sigma_{S_{z-1}}+z-1}, \tag{39}$$

which implies that

$$\mathbb{P}(\tau_z < \infty) = \mathbb{P}(\sigma_{S_{z-1}} < \infty). \tag{40}$$

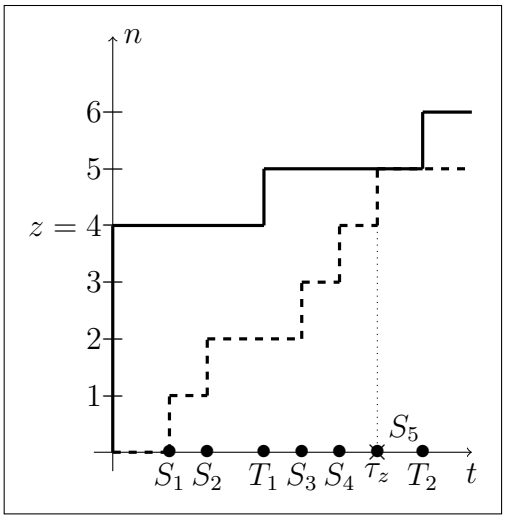
Again the one-to-one correpondence between the trajectories leading to ruin in the multiple risk models entail the equality almost surely. The following result provides, inter-alia, an upper bound for the probability of a successful double-spending attack.

Theorem 3. *If $\{N(t)\ ,\ t \geq 0\}$ and $\{M(t)\ ,\ t \geq 0\}$ are two independent renewal processes such that (A1)-(A2) hold then*

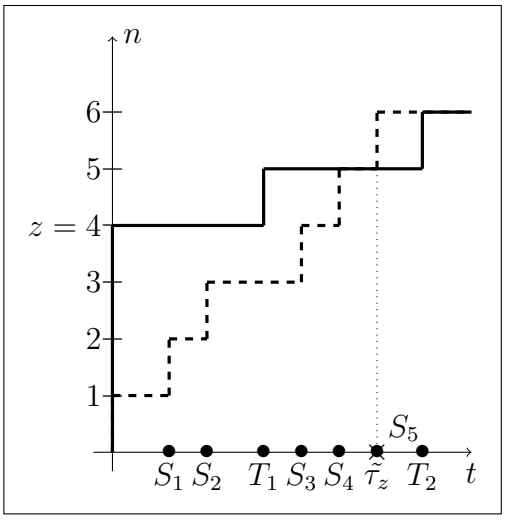
$$\mathbb{P}(\tau_z < \infty) = \frac{\left[\mathbb{E}\left(e^{-\gamma\Delta^S}\right)\right]^{z-1}}{\mathbb{E}\left[e^{\gamma\xi(S_{z-1})}\middle|\tau_z < \infty\right]}, \tag{41}$$

where $\xi(S_{z-1}) = S(\sigma_{S_{z-1}}) - S_{z-1}$ denotes the overshoot, immediately after ruin, in model (37). The following Cramér-Lundberg upper bound holds

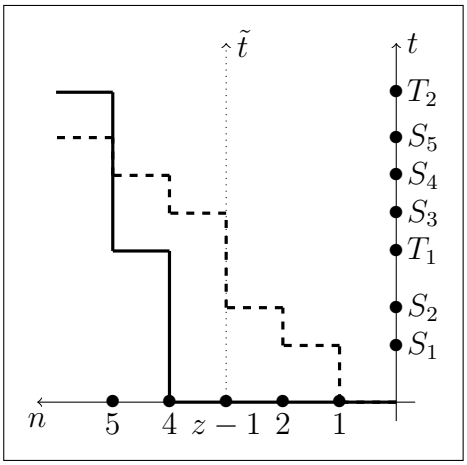
$$\mathbb{P}(\tau_z < \infty) \leq \left[\mathbb{E}\left(e^{-\gamma\Delta^S}\right)\right]^{z-1}. \tag{42}$$



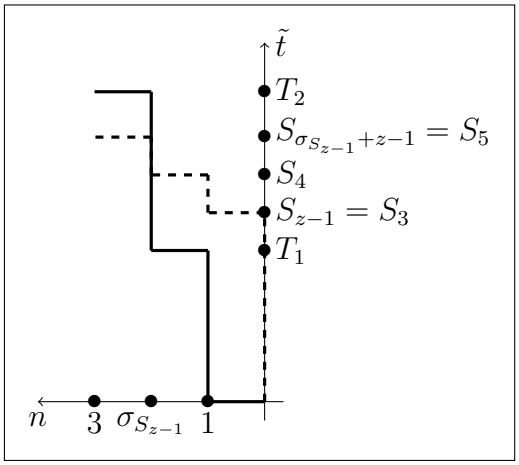
(a) Ruin time in the risk model (32). (solid) Trajectory of the process $\{z + N(t)\ ,\ t \geq 0\}$, (dashed) Trajectory of the process $\{M(t)\ ,\ t \geq 0\}$.



(b) Ruin time in the risk model (34). (solid) Trajectory of the process $\{z + N(t)\ ,\ t \geq 0\}$, (dashed) Trajectory of the process $\{M(t) + 1\ ,\ t \geq 0\}$.



(c) 90° anticlockwise rotation of Figure 2(b).



(d) Ruin time in the risk model (37). (solid) Trajectory of the process $\{T_k\ ,\ k \geq 0\}$, (dashed) Trajectory of the process $\{S_{z-1+k}\ ,\ k \geq 0\}$.

Figure 2: Boundary crossing problems in the various risk models.

Proof. The claim surplus process $\{S^*(n)\ ,\ n \geq 1\}$ in (38) is a random walk. Assumption (A2) implies that the process $\{e^{\gamma S^*(n)}\ ,\ n \geq 1\}$ is a Martingale as a consequence of [2, Theorem 1.1]. Note also that $S^*(n) \overset{\mathbf{a.s}}{\to} -\infty$, where $\overset{\mathbf{a.s}}{\to} -\infty$ stands for convergence almost surely, follows from assumption (A1) and the law of large numbers. Let the initial reserves $S_{z-1} = s \geq 0$ be fixed in (37). The application of [2, Proposition 3.1] allows to rewrite the ultimate ruin probability as

$$\mathbb{P}(\sigma_s < \infty) = \frac{e^{-\gamma s}}{\mathbb{E}\left[e^{\gamma \xi(s)} \middle| \sigma_s < \infty\right]}, \tag{43}$$

where $\xi(u) = S^*(\sigma_s) - s$ denotes the overshoot given that ruin occured in model (37).

Thanks to the connection (40), by conditioning on the values of S_{z-1}, it holds that

$$\mathbb{P}(\tau_z < \infty) = \frac{\mathbb{E}\left[e^{-\gamma S_{z-1}}\right]}{\mathbb{E}\left[e^{\gamma\xi(S_{z-1})}\middle|\tau_z < \infty\right]} = \frac{\mathbb{E}\left[e^{-\gamma\Delta^S}\right]^{z-1}}{\mathbb{E}\left[e^{\gamma\xi(S_{z-1})}\middle|\tau_z < \infty\right]}. \tag{44}$$

The upper bound (42) follows from noting that $\mathbb{E}\left[e^{\gamma\xi(S_{z-1})}\middle|\sigma_s < \infty\right] > 1$. □

The next result specifies the expression for the probability $\mathbb{P}(\tau_z < \infty)$ in the case where $\{N(t)\ ,\ t \geq 0\}$ is a Poisson process.

Corollary 3. *If $\{N(t)\ ,\ t \geq 0\}$ is a homogeneous Poisson pocess of intensity λ and $\{M(t)\ ,\ t \geq 0\}$ is a renewal process, independent from $\{N(t)\ ,\ t \geq 0\}$, such that (A1)-(A2) holds then*

$$\mathbb{P}(\tau_z < \infty) = \frac{\lambda - \gamma}{\lambda}\mathbb{E}\left[e^{-\gamma\Delta^S}\right]^{z-1}. \tag{45}$$

If $\{M(t)\ ,\ t \geq 0\}$ is also a homogeneous Poisson process of intensity $\mu < \lambda$ then

$$\mathbb{P}(\tau_z < \infty) = \left(\frac{\mu}{\lambda}\right)^z. \tag{46}$$

Proof. As $\{N(t)\ ,\ t \geq 0\}$ is a homogenous Poisson process, it is renewal process and (41) holds. The sequence of inter-arrival times $\{\Delta_k^T,\ k \geq 1\}$ is formed by **i.i.d.** exponential random variables with parameter λ which implies that the overshoot $\xi(S_{z-1})$ is also exponentially distributed with parameter λ by virtue of memorylessness of the exponential distribution. It follows that

$$\mathbb{E}\left[e^{\gamma\xi(S_{z-1})}\middle|\tau_z < \infty\right] = \frac{\lambda}{\lambda - \gamma}.$$

Substituting in (41) yields (45). Now, assume that $\{M(t),\ t \geq 0\}$ is also a Poisson process with intensity $\mu < \lambda$. The sequence of inter-arrival times $\{\Delta_k^S,\ k \geq 1\}$ is made of **i.i.d.** exponential random variable with parameter μ. The equation (31) is equivalent to

$$\log\left(\frac{\lambda\mu}{(\lambda - \theta)(\mu + \theta)}\right) = 0, \tag{47}$$

and admits $\gamma = \lambda - \mu$ as only non-negative solution. Substituting $\gamma = \lambda - \mu$ into (45) yields (46). □

This result allows to confirm Corollary 1, see Example 1. Note that the probability of a successful double-spending attack when the length of the chains are two independent Poisson processes may be retrieved without using the duality argument as shown in the following remark.

Remark 4.1. *Assume that $\{N(t),\ t \geq 0\}$ and $\{M(t)\ ,\ t \geq 0\}$ are two independent homogeneous Poisson processes with respective intensity λ and μ such that $\lambda > \mu$. The claim surplus process, already defined in* (33) *as*

$$S(t) = M(t) - N(t),\ \textit{for } t \geq 0, \tag{48}$$

is the difference between two independent Poisson processes and thus a Lévy process. [2, Theorem 1.2] then implies that the process defined by

$$e^{\theta S(t)-t\kappa(\theta)},\ \textit{for } t \geq 0, \tag{49}$$

where $\kappa(\theta) = \log \mathbb{E}\left[\theta S(1)\right]$*, is a Martingale. The equation* $\kappa(\theta) = 0$ *is equivalent to*

$$\mu e^{\theta} + \lambda e^{-\theta} - (\lambda + \mu) = 0, \tag{50}$$

and admits a unique non-negative solution $\gamma = \log(\lambda/\mu)$*. Consequently, the process* $\{e^{\gamma S(t)}\ , t \geq 0\}$ *is a Martingale. Moreover, the condition* $\lambda > \mu$ *entails* $S(t) \overset{\textbf{a.s.}}{\to} -\infty$*. Applying [2, Proposition 3.1] yields*

$$\mathbb{P}(\tau_z < \infty) = \frac{e^{-\gamma z}}{\mathbb{E}\left[e^{\gamma \xi(z)} \middle| \tau_z < \infty\right]}, \tag{51}$$

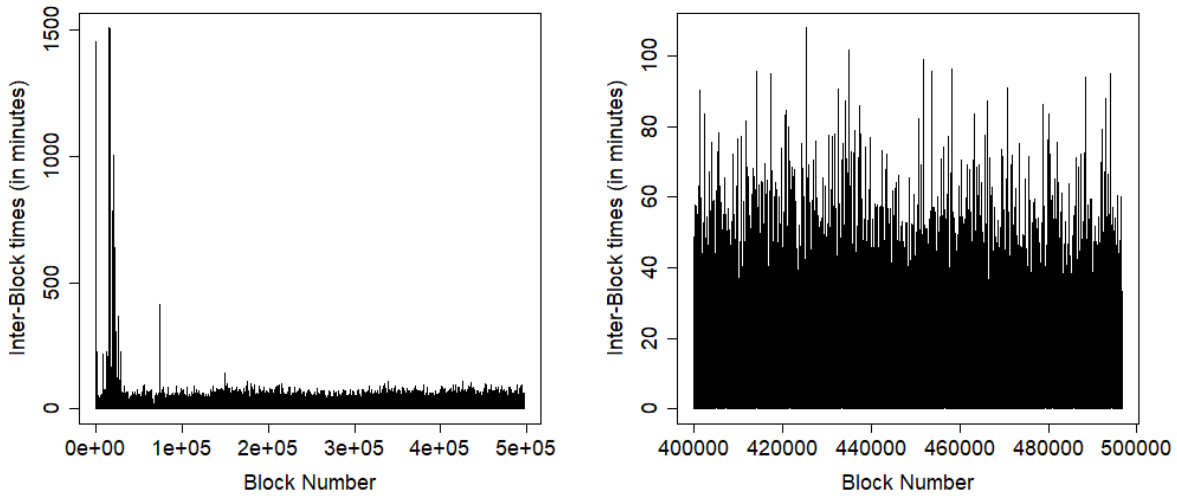
where $\xi(z) = S(\tau_z) - z$ *denotes the overshoot after ruin occured in model* (32)*. In the case considered here there is no overshoot as* $S(\tau_z) = z$*. Substituting* $\gamma = \log(\lambda/\mu)$ *into* (51) *yields* (45)*.*

5 Numerical illustrations

The numerical results presented here are based on the data collected by Bowden [7] and the analysis conducted in Bowden et al. [8]. In Subsection 5.1, data is used to fit the inter-block time distribution within the public chain. The lack of data regarding the growth of the malicious chain is circumvented by assuming that the inter-block time in the malicious chain are defined as a transformation of the inter-block times in the public chain. This allows us to illustrate the results given in Section 2 and 4. Subsection 5.2 focuses on the case where the lengths of the chains are governed by non-homogeneous Poisson processes which seems to be the most suitable model according to Bowden et al. [8]. It allows, in turn, to illustrate the results derived in Section 3.

5.1 Length of the chains as renewal process

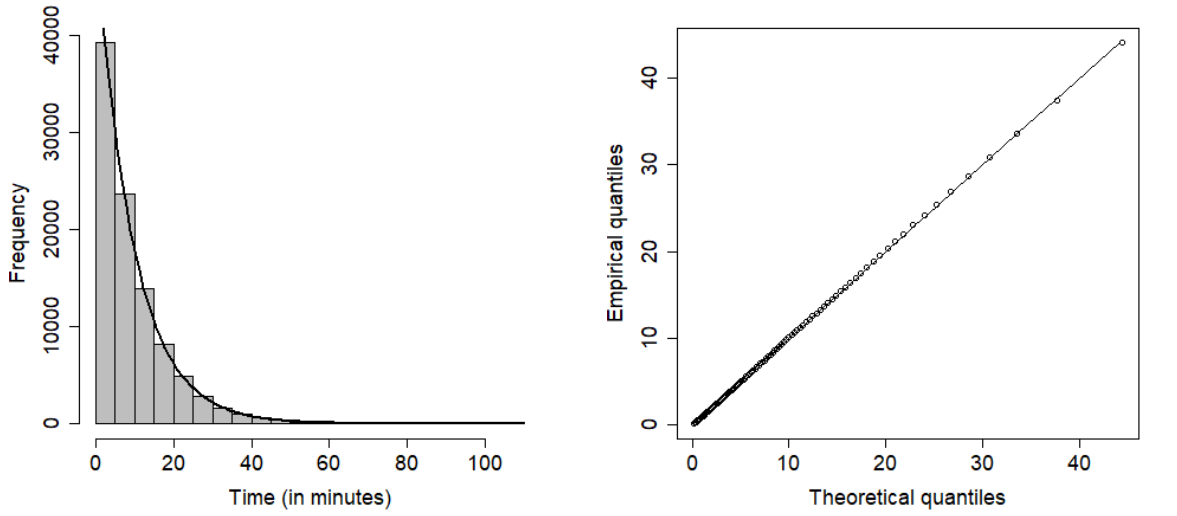
In this subsection, the length of the chains $\{N(t)\ , t \geq 0\}$ and $\{M(t)\ , t \geq 0\}$ are assumed to be governed by renewal processes. The first task consists in studying the fit of the inter-block time distribution to the data provided in Bowden [7]. Figure 3 displays the inter-block times chronologically. The distribution of the inter-block times of the first few blocks presents a few spikes (to the magnitude of the day) before reaching stationarity arround the $200,000^{\text{th}}$ block, see Figure 3(a). If we limit our analysis to the latest blocks, starting for instance from the $400,000^{\text{th}}$, then the data admits fewer outliers, with a maximum of 2 hours, see Figure 3(b). The inference of the block arrival is therefore based on the inter-arrival times starting from the $400,000^{\text{th}}$ block onward which still represents $96,628$ data points. The empirical mean is equal to 9.57 minutes while the standard deviation is significantly lower than the overall one with 9.56 minutes. Figure 4 shows the histogram of the inter-block times. The **p.d.f.** of the exponential distribution $\text{Exp}(\widehat{\lambda})$, where $\widehat{\lambda} = 1/9.57$ corresponds to the method of moment estimator,



(a) Time series of the interblock times.

(b) Time series of the interblock time starting from the 40000^{th} block.

Figure 3: Chronological series of the inter-block times.



(a) Histogram of the inter-block times distribution.

(b) Q-Q plot to test the adequacy to the exponential distribution.

Figure 4: Distribution of the inter-block times.

matches reasonably well the histogram. On Figure 4(b), the empirical quantiles are plotted against the quantiles of the exponential distribution $\text{Exp}(\widehat{\lambda})$. The points overlap the diagonal $y = x$, which indicates a superb fit. This analysis leads us to model the number of blocks in the honest chain $\{N(t)\ ,\ t \geq 0\}$ by a homogeneous Poisson process of intensity $\widehat{\lambda}$.

Regarding the block arrival in the malicious chain, no data is available. The only a priori information is that the inter-block time should be larger to account for the un-

balanced repartition of the computing power in favor of the honest miners. The growth of $\{M(t)\ ,\ t \geq 0\}$ should be slower than the growth of $\{N(t)\ ,\ t \geq 0\}$. In the sequel, two definitions of the inter-arrival times $\{\Delta_k^S\ ,\ k \geq 1\}$ that generate $\{M(t)\ ,\ t \geq 0\}$ are compared in terms of the risk of a double-spending attack.

1. Define
$$\Delta^S \stackrel{D}{=} \frac{\Delta^T}{r}, \tag{52}$$
where $r > 1$. The inter-arrival times $\{\Delta_k^S\ ,\ k \geq 1\}$ are **i.i.d.** exponential random variables and the process $\{M(t)\ ,\ t \geq 0\}$ is a homogeneous Poisson process with intensity $\widehat{\lambda}/r$. Corollary 3 applies and the probability of successful double spending attack is given by
$$\mathbb{P}(\tau_z < \infty) = \left(\frac{1}{r}\right)^z. \tag{53}$$
The **p.d.f.** of the double spending time f_{τ_z} follows from applying Corollary 1 and is given in (17) after substituting $\lambda = \widehat{\lambda}$ and $\mu = \widehat{\lambda}/r$.

2. Define
$$\Delta^S \stackrel{D}{=} \Delta_1^T + \ldots + \Delta_r^T, \tag{54}$$
where $r > 1$ is integer-valued. The inter-arrival times $\{\Delta_k^S\ ,\ k \geq 1\}$ are **i.i.d.** gamma random variables $\text{Gam}(r, \lambda)$ with associated **p.d.f.**
$$f_{\Delta^S}(t) = \frac{e^{-\lambda t} t^{r-1} \lambda^t}{\Gamma(r)},\ \text{for } t \geq 0. \tag{55}$$
The process $\{M(t)\ ,\ t \geq 0\}$ is, in turn, a renewal process. Corollary 3 applies and the probability of successful double spending attack is given by
$$\mathbb{P}(\tau_z < \infty) = \frac{\lambda - \gamma}{\lambda}\left[\frac{\lambda}{\lambda + \gamma}\right]^{r(z-1)}, \tag{56}$$
where γ is the only non-negative solution to the equation
$$\log\left[\frac{\lambda^r}{(\lambda - \theta)(\lambda + \theta)^{r-1}}\right] = 0. \tag{57}$$
Note that the root in (57) is derived numerically using the `uniroot` built-in function in `R`. The **p.d.f.** f_{τ_z} of the double-spending time follows from the application Corollary 1 and reduces, after a couple of rearrangements, to
$$f_{\tau_z}(t) = \sum_{n=0}^{+\infty} \left(\frac{z}{z+n}\right) \frac{\Gamma[r(n+z)+n]}{\Gamma(n+1)\Gamma[r(n+z)]2^{r(n+z)+n}} \frac{(2\lambda)^{r(n+z)+n} t^{r(n+z)+n-1} e^{-2\lambda t}}{\Gamma[r(n+z)+n]}, \tag{58}$$
for $t \geq 0$.

Table 1 reports the value of (53) and (56) for $r = 2, 3, 4, 5$ and $z = 1, 2, 3, 4, 5$. Although definitions 1 and 2 both means that the building of blocks is r times slower in the malicious chain, the probability of a successful double-spending attack is much smaller when

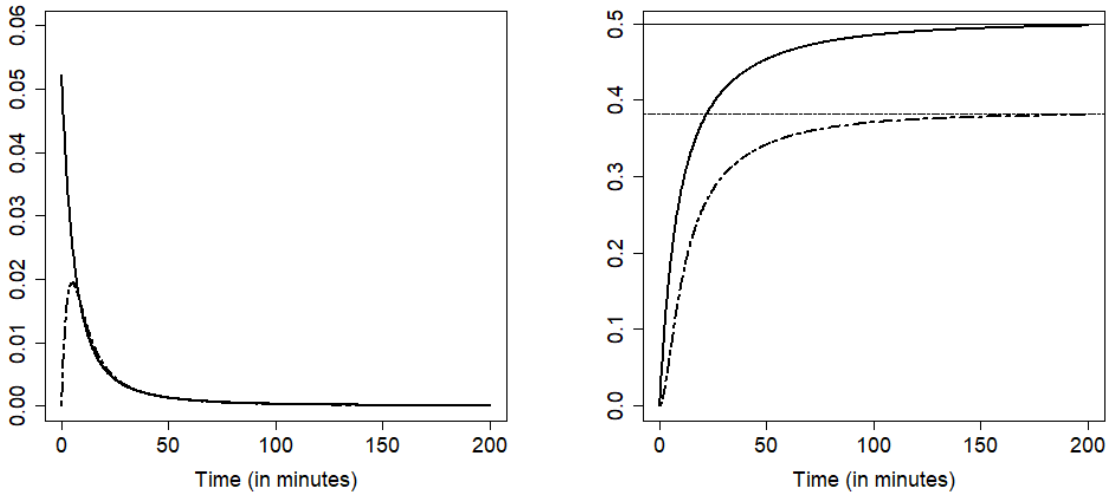
	$\Delta^S \sim \text{Exp}(\lambda/r)$				$\Delta^S \sim \text{Gamma}(r, \lambda)$			
	$r=2$	$r=3$	$r=4$	$r=5$	$r=2$	$r=3$	$r=4$	$r=5$
z								
1	0.5000	0.3333	0.2500	0.2000	0.3819	0.1608	0.0724	0.0342
2	0.2500	0.1111	0.0625	0.0400	0.1459	0.0258	0.0052	0.0012
3	0.1250	0.0370	0.0156	0.0080	0.0557	0.0042	0.0004	0.0000
4	0.0625	0.0123	0.0039	0.0016	0.0213	0.0007	0.0000	0.0000
5	0.0312	0.0041	0.0010	0.0003	0.0081	0.0001	0.0000	0.0000

Table 1: Probability of a successful double-spending attempt

$\{M(t)\ ,\ t \geq 0\}$ is a renewal process with gamma distributed inter-arrival times. It shows the influence of the shape of the distribution of the inter-arrival times on the likelihood of a double-spending attack. Figure 5 displays the **p.d.f.** and **c.d.f.** of the double-spending time

$$\tau_1 = \inf\{t \geq 0\ ;\ N(t) + 1 = M(t)\},$$

for $r = 2$ along with the reference horizontal lines $y = \mathbb{P}(\tau_1 < \infty)$. Note that the infinite series in (17) and (58) are truncated to the order $K = 50$. In both cases, the merchant



(a) **P.d.f.** of τ_1.

(b) **C.d.f.** of τ_1 as well as $\mathbb{P}(\tau_1 < \infty)$.

Figure 5: **P.d.f.** and **c.d.f.** of the double spending time τ_1. (solid) $\{M(t)\ ,\ t \geq 0\}$ is a homogeneous Poisson process with intensity $\widehat{\lambda}/2$. (dashed) $\{M(t)\ ,\ t \geq 0\}$ is a renewal process with gamma $\text{Gam}(r = 2, \widehat{\lambda})$ distributed inter-arrival times.

who is waiting for two hours is not taking much risk as the **c.d.f.** reaches the barrier $\mathbb{P}(\tau_1 < \infty)$. The `R` code is accessible online at [15] for the sake of reproducibility.

5.2 Length of the chains as non-homogeneous Poisson process

In this subsection, the length of the chains $\{z + N(t)\ ,\ t \geq 0\}$ and $\{M(t)\ ,\ t \geq 0\}$ are assumed to be governed by two non-homogeneous Poisson processes. The intensity function of $\{N(t)\ ,\ t \geq 0\}$ is given by

$$\lambda(t) = \frac{pH(t)L(t)}{2^{256}}, \tag{59}$$

where $H(t)$ denotes the global hashrate function, $L(t)$ is the target function, and $p \in (0, 1)$ reflects the repartition of the computing power between honest and dishonest miners. The reader is refered to Example 2 for a definition of these quantities. The intensity function of $\{M(t)\ ,\ t \geq 0\}$ is given by

$$\lambda^*(t) = \frac{(1-p)H(t)L^*(t)}{2^{256}}. \tag{60}$$

The global hashrate is assumed to admit a parametric form with $H(t) = e^{at+b}$, where the values of a and b are selected from Bowden et al. [8, Table 1]. The difficulty is assumed to be the same for all the miners so that $L^*(t) = L(t)$. More specifically, let us consider a time span during which the difficulty is contant, equal to L say. This is true for time periods that are about two weeks long as it corresponds to the average time to discover $2,016$ blocks. The intensity function $\lambda(t)$ associated to $\{N(t)\ ,\ t \geq 0\}$ becomes

$$\lambda(t) = \frac{pe^{at+b}L}{2^{256}}, \tag{61}$$

and may be integrated as

$$\Lambda(t) = \int_0^t \lambda(s)\mathrm{d}s = \frac{pe^bL}{2^{256}a}\left(e^{at} - 1\right). \tag{62}$$

Note that we have equivalently

$$\lambda^*(t) = \frac{(1-p)e^{at+b}L}{2^{256}}, \tag{63}$$

and

$$\Lambda^*(t) = \int_0^t \lambda^*(s)\mathrm{d}s = \frac{(1-p)e^bL}{2^{256}a}\left(e^{at} - 1\right). \tag{64}$$

In view of these assumptions, the conditional distribution of the block arrival times given the length of the chain is the same in the honest and the malicious chain. Namely, it holds that $F_t(s) = F_t^*(s)$ and the probability $\mathbb{P}(\tau_z > t)$ of the double-spending attack being unsuccessful before t may be estimated via (30). The practical evaluation is handled via Monte Carlo simulations, note that the numerical value of an Appell polynomials of any order may be computed recursively using the relations given in the appendix A, see Proposition 2. Let the time unit be one second, consider a 2 weeks long time period which corresponds to $1,209,600$ seconds. Let us set the parameters of the hashrate function to $a = -9.44 \times 10^{-9}$ and $b = 27.1$ according to the first row of the table in Bowden et al. [8, Table 1]. The difficulty is assumed to be constant, equal to

$$L = 2016 \times \frac{2^{256}a}{e^b(e^{a*1209600} - 1)},$$

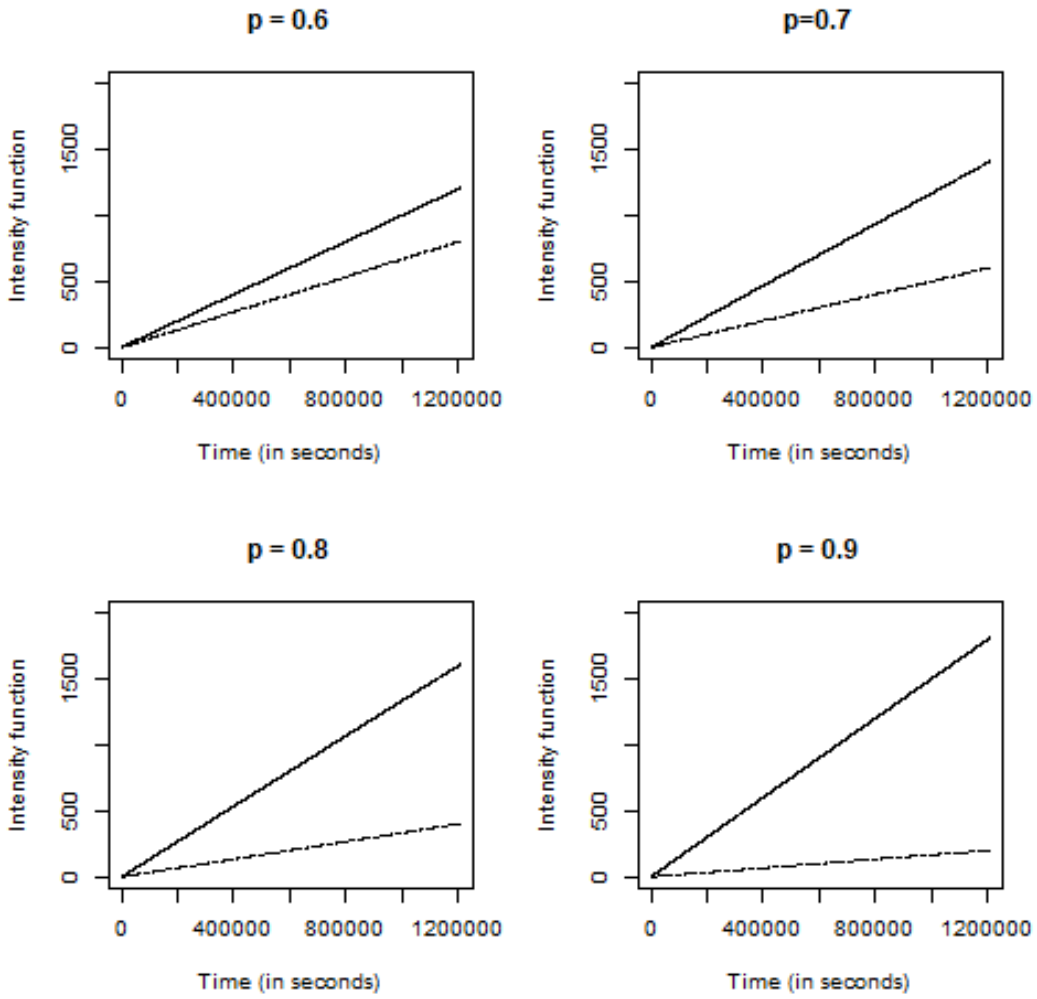


Figure 6: Integrated intensity functions over time: (solid) $\Lambda(t)$ associated to $\{N(t)\ ,\ t \geq 0\}$, (dashed) $\Lambda^*(t)$ associated to $\{M(t)\ ,\ t \geq 0\}$.

in order to have on average $2,016$ blocks discovered by the end of the time horizon (**i.e.** two weeks). Figure 6 displays the integrated intensity functions (62) and (64) over time. The parametrization entails a linear growth (on average) of the chains which makes our example close to the homogeneous Poisson arrival situation. Formula (20) is difficult to use for risk management purposes without the knowledge of the mass of probability associated to $\tau_z = \infty$. This issue is addressed by assuming that an attacker gives up his double-spending attempt if completed within three hours (10,800 seconds). It makes little sense in practice for an attacker to carry on an attack for two weeks. We then investigate the probability of a successful double-spending attack attempted every three hours over the course of two weeks. Namely, denote by

$$t_k = k \times 10800, \text{ for } 0 \leq k < 112,$$

the sequence of time steps and

$$p_{z,k} = \mathbb{P}(\tau_{z,k} < 10800), \tag{65}$$

the probabilities of interest, where

$$\tau_{z,k} = \inf\{t \in (t_k, t_{k+1}) \ ; \ M(t) = z + N(t) \big| M(t_k) = N(t_k) = 0\}.$$

Let us assume that the honest chain is 1 block ahead, which in turn, allows to use Formula (26) given in Corollary 2 and alleviate the computational burden associated to the recursive evaluation of Appell polynomials. Figure 7 displays the value of the probabilities (65) over the two weeks of operations for various repartition $p \in \{0.6, 0.7, 0.8, 0.9\}$ of the hash power. Note that the evaluation is based on $10,000$ trajectories of $\{N(t) \ , \ t \geq 0\}$ and $\{M(t) \ , \ t \geq 0\}$. The probabilities $p_{z,k}$ are constant over time, this was expected

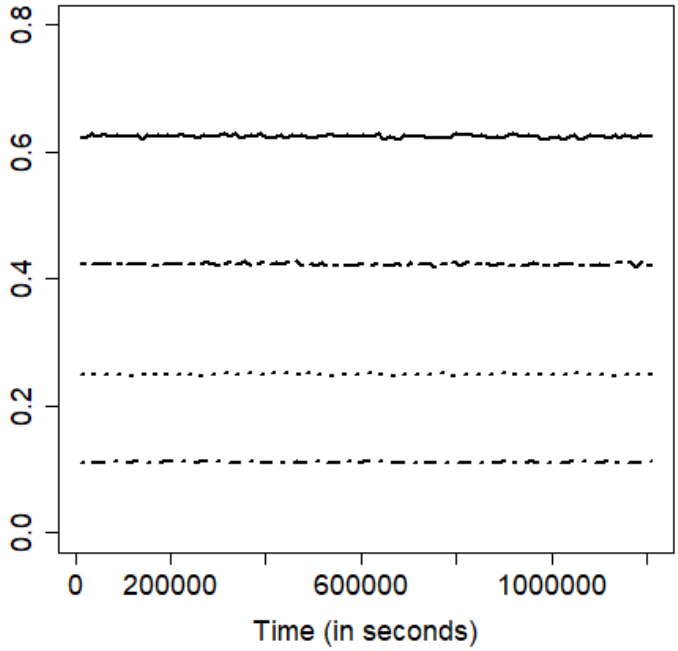


Figure 7: Evolution of the probability of performing a successful double-spending attack in the course of two weeks for various values of p: (solid) $p = 0.6$, (dashed) $p = 0.7$, (dotted) $p = 0.8$, (dot-dash) $p = 0.9$.

as the arrival of blocks is almost time-homogenous due to the parametrization of the global hashrate function. The source code is available online [15] and the reader is invited to experiment the effect of modifying the parameters a and b on the double-spending probabilities.

6 Concluding remarks

In this paper, the model, initially proposed by Satoshi Nakamoto [27], to comprehend the double-spending issue is refined. Assuming that the lengths of the competing blockchains are governed by counting processes leads to interesting boundary crossing problems. This refinement offers more flexibility to reflect accurately the block discovery frequency as well as the distribution of the computing power among honest and dishonest miners through

the calibration of the arrival times that generate the aforementionned counting pocesses. Theorem 3 is useful to advise the merchant on how many subsequent blocks should be added to the chain before shipping the good. Theorem 1 and 2 enable to figure the time at which the double-spending attack is most likely to occur. This is helpful to provide merchants with guidelines on how long they should wait before shipping a good, which compliments the advice on the number of blocks.

The success of the blockchain method has resulted in bitcoin becoming increasingly popular and inspiring other electronic payment method. It is worth mentionning that the results derived in this paper maybe relevant to understand other systems where similar blockchain policies are used.

Selfish mining, described for instance in Sapirshtein et al. [32] and Eyal and Siren [12], is another type of miners' misconduct. Nowadays, it is no longer feasible to mine **BTCs** in isolation. Empirical evidence shows that **BTCs** miners behave strategically by gathering in pools. All members contribute to the solution of each cryptopuzzle, and share the rewards proportionally to their contribution. Selfish mining is a strategy that can be used by a minority pool to obtain more revenue. The key idea is for the pool of selfish miners to keep its discovered blocks private while honest nodes continue to mine on the public chain. Assuming that the private chain has the lead over the public chain, when the public branch approaches the selfish miners' one, the private chain is released publicly. It results in a waste of resources for all the miners but Eyal and Siren [12] showed that the revenue of the selfish miners goes beyond the revenue expected by following the usual protocol given their share of the total computing power. The results presented here may be relevant in the context of a self-mining attack as it boils down again to the race between two counting processes.

Acknowledgements

My work was partially funded by a CAE educational grant issued by the Society of Actuaries.

References

[1] M. Aigner. *A Course in Enumeration*, volume 238. Springer Science & Business Media, Berlin, 2007.

[2] S. Asmussen and H. Albrecher. *Ruin Probabilities*. World Scientific, Singapore, 2010.

[3] K. A. Borovkov and Z. Burq. Kendall's identity for the first crossing time revisited. *Electronic Communications in Probability*, 6(9):91–94, 2001.

[4] K. A. Borovkov and D. C. M. Dickson. On the ruin time distribution for a Sparre Andersen process with exponential claim sizes. *Insurance: Mathematics and Economics*, 42(3):1104–1108, 2008.

[5] R. Boucherie and O. Boxma. The workload in the M/G/1 queue with work removal. *Probability in the Engineering and Informational Sciences*, 10(2):261–277, 1996.

[6] R. Boucherie, O. Boxma, and K. Sigman. A note on negative customers, GI/G/1 workload, and risk processes. *Probability in the Engineering and Informational Sciences*, 11(3):305–311, 1997.

[7] R. Bowden. Intersection of the longest increasing subsequences. https://github.com/rhysbowden/LIS.

[8] R Bowden, HP Keeler, AE Krzesinski, and PG Taylor. Block arrivals in the bitcoin blockchain. *arXiv preprint arXiv:1801.07447*, 2018.

[9] D. S. Dimitrova, Z. G. Ignatov, and V. K. Kaishev. On the first crossing of two boundaries by an order statistics risk process. *Risks*, 5(3):43, 2017.

[10] D. S. Dimitrova, V. K. Kaishev, and S. Zhao. On finite-time ruin probabilities in a generalized dual risk model with dependence. *European Journal of Operational Research*, 242(1):134–148, 2015.

[11] D.S. Dimitrova, V. K. Kaishev, and S. Zhao. On the evaluation of finite-time ruin probabilities in a dependent risk model. *Applied Mathematics and Computation*, 275(Supplement C):268–286, 2016.

[12] I. Eyal and E. G. Sirer. Majority is not enough: Bitcoin mining is vulnerable. In *International conference on financial cryptography and data security*, pages 436–454. Springer, 2014.

[13] E. Gelenbe, P. Glynn, and K. Sigman. Queues with negative arrivals. *Journal of applied probability*, 28(1):245–250, 1991.

[14] P.-O. Goffard. Two-sided exit problems in the ordered risk model. *Methodology and Computing in Applied Probability*, 2017.

[15] P.-O. Goffard. *Online accompaniment for "Fraud risk assessment within blockchain transactions"*, 2018. Available at `https://github.com/LaGauffre/FraudRiskBlockchainTransaction`.

[16] P.-O. Goffard and C. Lefèvre. Boundary crossing of order statistics point processes. *Journal of Mathematical Analysis and Applications*, 447(2):890–907, 2017.

[17] P.-O. Goffard and C. Lefèvre. Duality in ruin problems for ordered risk models. *Insurance: Mathematics and Economics*, 78:44–52, 2018.

[18] P. Harrison and E. Pitel. The M/G/1 queue with negative customers. *Advances in Applied Probability*, 28(2):540–566, 1996.

[19] R. Van Der Hofstad and M. Keane. An elementary proof of the hitting time theorem. *The American Mathematical Monthly*, 115(8):753–756, 2008.

[20] Z. G. Ignatov and V. K. Kaishev. First crossing time, overshoot and appell–hessenberg type functions. *Stochastics: An International Journal of Probability and Stohastic Processes*, 88(8):1240–1260, 2016.

[21] G. Jain and K. Sigman. Generalizing the Pollaczek-Khintchine formula to account for arbitrary work removal. *Probability in the Engineering and Informational Sciences*, 10(4):519–531, 1996.

[22] C. Lefèvre and S. Loisel. Finite-time ruin probabilities for discrete, possibly dependent, claim severities. *Methodology and Computing in Applied Probability*, 11(3):425–441, 2009.

[23] C. Lefèvre and P. Picard. A new look at the homogeneous risk model. *Insurance: Mathematics and Economics*, 49(3):512–519, 2011.

[24] C. Lefèvre and P. Picard. Ruin probabilities for risk models with ordered claim arrivals. *Methodology and Computing in Applied Probability*, 16(4):885–905, 2014.

[25] C. Lefèvre and P. Picard. Risk models in insurance and epidemics: A bridge through randomized polynomials. *Probability in the Engineering and Informational Sciences*, 29(3):399–420, 2015.

[26] C. Mazza and D. Rullière. A link between wave governed random motions and ruin processes. *Insurance: Mathematics and Economics*, 35(2):205–222, 2004.

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

[28] D. Perry, W. Stadje, and S. Zacks. Boundary crossing for the difference of two ordinary or compound poisson processes. *Annals of Operations Research*, 113(1):119–132, 2002.

[29] D. Perry, W. Stadje, and S. Zacks. A two-sided first-exit problem for a compound poisson process with a random upper boundary. *Methodology and Computing in Applied Probability*, 7(1):51–62, 2005.

[30] P. Picard and C. Lefèvre. On the first meeting or crossing of two independent trajectories for some counting processes. *Stochastic processes and their applications*, 104(2):217–242, 2003.

[31] P. S. Puri. On the characterization of point processes with the order statistic property without the moment condition. *Journal of Applied Probability*, 19(1):39–51, 1982.

[32] Y. Sapirshtein, A.and Sompolinsky and A. Zohar. Optimal selfish mining strategies in bitcoin. In *International Conference on Financial Cryptography and Data Security*, pages 515–532. Springer, 2016.

[33] T. Shi and D. Landriault. Distribution of the time to ruin in some sparre andersen risk models. *ASTIN Bulletin*, 43(1):39–59, 2013.

A Appell and Abel-Gontcharov polynomials

Appell and Abel-Gontcharov (**A-G**) polynomials are well-known in mathematics. They can be used to solve various problems in statistics and applied probabilty. A short presentation is provided below. We refer e.g. to Lefèvre and Picard [25] for a review with applications in risk modelling. Let $U = \{u_i\ ,\ i \geq 1\}$ be a sequence of real numbers, non-decreasing in our context. To U is attached a (unique) family of Appell polynomials of degree n in x, $\{A_n(x|U)\ ,\ n \geq 0\}$, defined as follows. Starting with $A_0(x|U) = 1$, the $A_n(x|U)$ satisfy the differential equations

$$A_n^{(1)}(x|U) = nA_{n-1}(x|U), \tag{66}$$

with the border conditions

$$A_n(u_n|U) = 0, \quad n \geq 1. \tag{67}$$

So, each A_n, $n \geq 1$, has the integral representation

$$A_n(x|U) = n! \int_{u_n}^{x} \left[\int_{u_{n-1}}^{y_n} \mathrm{d}y_{n-1} \ldots \int_{u_1}^{y_1} \mathrm{d}y_2 \right] \mathrm{d}y_n. \tag{68}$$

In parallel, to U is attached a (unique) family of Abel-Gontcharov (**A-G**) polynomials of degree n in x, $\{G_n(x|U)\ ,\ n \geq 0\}$. Starting with $G_0(x|U) = 1$, the $G_n(x|U)$ satisfy the differential equations

$$G_n^{(1)}(x|U) = nG_{n-1}(x|\mathcal{E}U), \tag{69}$$

where $\mathcal{E}U$ is the shifted family $\{u_{i+1}\ ,\ i \geq 1\}$, and with the border conditions

$$G_n(u_1|U) = 0, \quad n \geq 1. \tag{70}$$

So, each G_n, $n \geq 1$, has the integral representation

$$G_n(x|U) = n! \int_{u_1}^{x} \left[\int_{u_2}^{y_1} \mathrm{d}y_2 \ldots \int_{u_n}^{y_{n-1}} \mathrm{d}y_n \right] \mathrm{d}y_1. \tag{71}$$

Note that both polynomial families are sometimes defined without the factor $n!$ in (68) and (71). Of course, these polynomials are related through the identity

$$G_n(x|u_1, \ldots, u_n) = A_n(x|u_n, \ldots, u_1), \quad n \geq 1. \tag{72}$$

However, the two families (i.e. considered for all $n \geq 0$) are distinct and enjoy quite different properties. From (68) and (71), one may see that the polynomials A_n and G_n, $n \geq 1$, can be interpreted in terms of the joint distribution of the order statistics $(U_{1:n}, \ldots, U_{n:n})$ of a sample of n independent uniform random variables on $(0, 1)$.

Proposition 1. *For* $0 \leq u_1 \leq \ldots \leq u_n \leq x \leq 1$,

$$P[U_{1:n} \geq u_1, \ldots, U_{n:n} \geq u_n \ \textit{and}\ U_{n:n} \leq x] = A_n(x|u_1, \ldots, u_n),$$

while for $0 \leq x \leq u_1 \leq \ldots \leq u_n \leq 1$,

$$P[U_{1:n} \leq u_1, \ldots, U_{n:n} \leq u_n \ \textit{and}\ U_{1:n} \geq x] = (-1)^n\, G_n(x|u_1, \ldots, u_n). \tag{73}$$

These representations play a key role in the first-meeting problems discussed in the paper. Numerically, it will be necessary to evaluate some special values of the polynomials. To this end, it is convenient to use the following recursive relations.

Proposition 2.

$$A_n(x|U) = \sum_{k=0}^{n} \binom{n}{k} A_{n-k}(0|U) x^k, \quad n \geq 1, \tag{74}$$

where the $A_n(0|U)$'s are obtained recursively from

$$A_n(0|U) = -\sum_{k=1}^{n} \binom{n}{k} A_{n-k}(0|U) u_n^k, \quad n \geq 1. \tag{75}$$

The A-G polynomials are computed through the recursion

$$G_n(x|U) = x^n - \sum_{k=0}^{n-1} \binom{n}{k} u_{k+1}^{n-k} G_k(x|U), \quad n \geq 1. \tag{76}$$

Formulas (74), and (75) follow from the Taylor's expansion of A_n, using also (66), and (67). Formula (76) follows from an Abelian expansion of x^n based on (69), and (70). Details are omitted here. Of course, the computing time increases with the degree of the polynomials. Note that

$$A_n(x|a+bU) = b^n A_n\left((x-a)/b\,|U\right), \quad n \geq 1, \tag{77}$$

with the same identity for G_n. An important particular case in our study is when the parameters in U are random and correspond to partial sums of exchangeable random variables.

Proposition 3. *Let $\{X_n\ ,\ n \geq 1\}$ be a sequence of exchangeable random variables, of partial sums $S_n = \sum_{k=1}^{n} X_k$ with $S_0 = 0$. Then, for $n \geq 1$,*

$$\mathbb{E}\left[A_n(x|S_1, \ldots, S_n)|S_n\right] = x^{n-1}(x - S_n), \tag{78}$$

$$\mathbb{E}\left[G_n(x|S_0, \ldots, S_{n-1})|S_n\right] = x(x - S_n)^{n-1}. \tag{79}$$

Proof. The identity (78) was derived in Proposition A.1 of Lefèvre and Picard [23], while the identity (79) follows from Goffard and Lefèvre [17]. □