

Research Article

On a Model for the Storage of Files on a Hardware: Statistics at a Fixed Time and Asymptotic Regimes

Vincent Bansaye

CMAP, Ecole Polytechnique, Route de Saclay, 91128 Palaiseau, France

Correspondence should be addressed to Vincent Bansaye, vbansaye@yahoo.com

Received 13 March 2009; Accepted 11 September 2009

Recommended by Vo V. Anh

We consider a version in continuous time of the parking problem of Knuth. Files arrive following a Poisson point process and are stored on a hardware identified with the real line, in the closest free portions at the right of the arrival location. We specify the distribution of the space of unoccupied locations at a fixed time and give asymptotic regimes when the hardware is becoming full.

Copyright © 2009 Vincent Bansaye. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

We consider a version in continuous time of the original parking problem of Knuth. Knuth was interested in the storage of data on a hardware represented by a circle with n spots. Files arrive successively at locations chosen uniformly at random and independently among these n spots. They are stored in the first free spot at the right of their arrival point (at their arrival point if it is free). Initially Knuth worked on the hashing of data (see, e.g., [1–3]): he studied the distance between the spots where the files arrive and the spots where they are stored. Later Chassaing and Louchard [4] have described the evolution of the largest block of data in such coverings when n tends to infinity. They observed a phase transition at the stage where the hardware is almost full, which is related to the additive coalescent. Bertoin and Miermont [5] have extended these results to files of random sizes which arrive uniformly on the circle.

We consider here a continuous time version of this model where the hardware is large and now identified with the real line. A file labelled i of length (or size) l_i arrives at time $t_i \geq 0$ at location $x_i \in \mathbb{R}$. The storage of this file uses the free portion of size l_i of the real line at the right of x_i as close to x_i as possible (see Figure 1). That is, it covers $[x_i, x_i + l_i[$ if this interval is free at time t_i . Otherwise this file can be split into several parts which are then stored in the closest free portions at the right of the arrival location. We require uniformity of the location where the files arrive and identical distribution of their sizes. Thus we model the arrival of

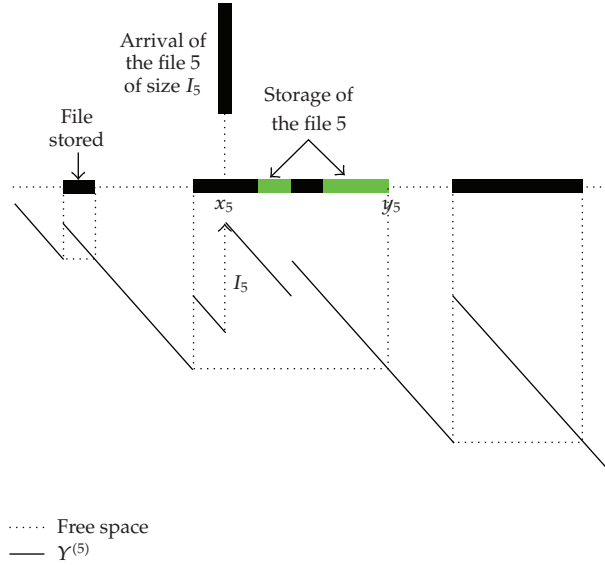


Figure 1: Arrival and storage of the 5-file and representation of $Y^{(5)}$. The first four files have been stored without splitting and are represented by the black rectangles.

files by a Poisson point process (PPP): $\{(t_i, x_i, l_i) : i \in \mathbb{N}\}$ is a PPP with intensity $dt \otimes dx \otimes \nu(dl)$ on $\mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^+$. We denote $m := \int_0^\infty l \nu(dl)$ and assume $m < \infty$. So m is the mean of the sum of sizes of files which arrive during a unit interval time on some interval with unit length.

We begin by constructing this random covering (Section 2). The first questions which arise and are treated here concern statistics at a fixed time t for the set of occupied locations $\mathcal{C}(t)$. What is the distribution of the covering at a fixed time? At what time the hardware becomes full? What are the asymptotics of the covering at this saturation time? What is the length of the largest block on a part of the hardware?

It is quite easy to see that the hardware becomes full at a deterministic time equal to $1/m$. In Section 3.1, we give some geometric properties of the distribution of the covering at a fixed time and we characterize this distribution by giving the joint distribution of the block of data straddling 0 and the free spaces on both sides of this block. The results given in this section will be useful for the problem of the dynamic of the covering considered in [6], where we investigate the evolution in time of a typical data block.

Then, using this characterization, we determine in Sections 3.2 and 3.3 the asymptotic regimes at the saturation time, which depend on the tail of ν , as in [4, 5, 7]. More precisely, we give the asymptotic of $\mathcal{C}(t)$ when t tends to $1/m$ (Theorem 3.6) and the asymptotic of $\mathcal{C}(t)$ restricted to $[0, x]$ when x tends to infinity and t tends to $1/m$ (Theorem 3.10).

We derive then the asymptotic behavior of the largest block of the hardware restricted to $[0, x]$ when x tends to infinity and t tends to $1/m$ (Corollary 3.11). As Chassaing and Louchard in [4], we observe a phase transition. Results are stated in Section 3 and proved in Section 4.

It is easy to check that for each fixed time t , $\mathcal{C}(t)$ does not depend on the order of arrival of files before time t . If ν is finite, we can view the files which arrive before time t as customers: the size of the file l becomes the service time of the customer and the location x where the file that arrives becomes the arrival time of the customer. We are then in the framework of

the $M/G/1$ queue model in the stationary regime and the covering $\mathcal{C}(t)$ becomes the union of busy periods (see, e.g., [8, Chapter 3] or [9]). Thus, results of Sections 3.1 and 3.3 for finite ν follow easily from known results on $M/G/1$. When ν is infinite, results are similar though the busy cycle is not defined. Thus the approach is different and proving asymptotics on random sets requires results about Lévy processes (see the appendix) and regenerative sets. One motivation for the case when ν is infinite comes from storage models which appear by renormalization when both the number of customers who store files independently and the size of the hardware go to infinity. Moreover, as far as we know, the longest busy period and more generally asymptotic regimes on $[0, x]$ when x tends to infinity and t tends to the saturation time (Section 3.4) have not been considered in the queuing model.

2. Preliminaries

In this section, we introduce some notations and recall some definitions we need to state the results. We also provide an elementary construction of the model studied in this paper.

Throughout this paper, we use the classical notation δ_x for the Dirac mass at x and $\mathbb{N} = \{1, 2, \dots\}$.

If \mathcal{R} is a measurable subset of \mathbb{R} , we denote by $|\mathcal{R}|$ its Lebesgue measure and by \mathcal{R}^{cl} its closure. For every $x \in \mathbb{R}$, we denote by $\mathcal{R} - x$ the set $\{y - x : y \in \mathcal{R}\}$ and

$$g_x(\mathcal{R}) = \sup\{y \leq x : y \in \mathcal{R}\}, \quad d_x(\mathcal{R}) = \inf\{y > x : y \in \mathcal{R}\}. \quad (2.1)$$

By convention, $\sup \emptyset = -\infty$ and $\inf \emptyset = \infty$.

Topology of Matheron

If I is a closed interval of \mathbb{R} , we denote by $\mathcal{H}(I)$ the space of closed subsets of I . For all $x, y \in \mathbb{R}$ and $A \subset \mathbb{R}$ we define

$$d(x, y) = 1 - e^{-|x-y|}, \quad d(x, A) = \inf\{d(x, y) : y \in A\}, \quad (2.2)$$

and we endow $\mathcal{H}(I)$ with the Hausdorff distance d_H defined for all $A, B \in \mathcal{H}(I)$ by

$$d_H(A, B) = \max\left(\sup_{x \in A} d(x, B), \sup_{x \in B} d(x, A)\right). \quad (2.3)$$

The topology induced by this distance is the topology of Matheron [10]: a sequence \mathcal{R}_n in $\mathcal{H}(I)$ converges to \mathcal{R} if and only if for each open set G and each compact K ,

$$\begin{aligned} \mathcal{R} \cap G \neq \emptyset & \text{ implies } \mathcal{R}_n \cap G \neq \emptyset \text{ for } n \text{ large enough,} \\ \mathcal{R} \cap K = \emptyset & \text{ implies } \mathcal{R}_n \cap K = \emptyset \text{ for } n \text{ large enough.} \end{aligned} \quad (2.4)$$

It is also the topology induced by the Hausdorff metric on a compact set using $\arctan(\mathcal{R} \cup \{-\infty, \infty\})$ or the Skorokhod metric using the class of “descending saw-tooth functions” (see [10, 11] for details).

Tail of ν and Lévy Processes Indexed by \mathbb{R}

We give here several definitions which will be useful for the study of the asymptotic regimes. Following the notation in [5], we say that $\nu \in \mathfrak{D}_{2+}$ if ν has a finite second moment $m_2 := \int_0^\infty l^2 \nu(dl)$. For $\alpha \in]1, 2[$, we say that $\nu \in \mathfrak{D}_\alpha$ whenever

$$\exists C > 0 \quad \text{such that } \bar{\nu}(x) \stackrel{x \rightarrow \infty}{\sim} Cx^{-\alpha}. \quad (2.5)$$

Then, for $\alpha \in]1, 2[$, we put

$$C_\alpha := \left(\frac{C\Gamma(2-\alpha)}{m(\alpha-1)} \right)^{1/\alpha}. \quad (2.6)$$

We denote by $(B_z)_{z \in \mathbb{R}}$ a two-sided Brownian motion; that is, $(B_x)_{x \geq 0}$ and $(B_{-x})_{x \geq 0}$ are independent standard Brownian motions. For $\alpha \in]1, 2[$, we denote by $(\sigma_z^{(\alpha)})_{z \in \mathbb{R}}$ a càdlàg process with independent and stationary increments such that $(\sigma_x^{(\alpha)})_{x \geq 0}$ is a standard spectrally positive stable Lévy process with index α , that is,

$$\forall x \geq 0, \lambda \geq 0, \quad \mathbb{E} \left(\exp \left(-\lambda \sigma_x^{(\alpha)} \right) \right) = \exp(x\lambda^\alpha). \quad (2.7)$$

Finally, for all $\lambda \geq 0$ and $\alpha \in]1, 2[$, we introduce the following processes indexed by $z \in \mathbb{R}$:

$$Y_z^{2+, \lambda} = -\lambda z + \sqrt{\frac{m_2}{m}} B_z, \quad Y_z^{2, \lambda} = -\lambda z + \sqrt{\frac{C}{m}} B_z, \quad Y_z^{\alpha, \lambda} = -\lambda z + C_\alpha \sigma_z^{(\alpha)}, \quad (2.8)$$

and their infimum process defined $I_x^{\alpha, \lambda} := \inf\{Y_y^{\alpha, \lambda} : y \leq x\}$ for $x \in \mathbb{R}$.

Construction of the Covering $\mathcal{C}(t)$

We give here an elementary construction of $\mathcal{C}(t)$ and some basic identities we will use next. They are classical in queuing theory (see, e.g., [8]) and storage systems (see, e.g., [9]), and so we skip details and refer to the version [12] for complete proofs.

We provide a deterministic construction of $\mathcal{C}(t)$ for any fixed t . As $\mathcal{C}(t)$ does not depend on the order of arrival of files before t , this amounts to construct the covering \mathcal{C} associated with a given sequence of files labelled by $i \in \mathbb{N}$. The file labelled by $i \in \mathbb{N}$ has size l_i and arrives after the files labelled by $j \leq i-1$, at location x_i on the real line. Files are stored following the process described in the Introduction and \mathcal{C} is the portion of line which is used for the storage.

The covering \mathcal{C} is the increasing union of the coverings $\mathcal{C}^{(n)}$ ($n \geq 0$) obtained by considering only the first n files, that is,

$$\mathcal{C} := \bigcup_{n \in \mathbb{N}} \mathcal{C}^{(n)}, \quad (2.9)$$

where $\mathcal{C}^{(n)}$ can be defined in an elementary way by the following induction. Set $\mathcal{C}^{(0)} := \emptyset$, and introduce the complementary set $\mathcal{R}^{(n)}$ of $\mathcal{C}^{(n)}$ (i.e., the free space of the real line). Let

$y_{n+1} = \inf\{y \geq 0, |\mathcal{R}^{(n)} \cap [x_{n+1}, y[| = l_{n+1}\}$, so y_{n+1} is the right-most point which is used for storing the $(n + 1)$ th file. Define then

$$\mathcal{C}^{(n+1)} := \mathcal{C}^{(n)} \cup [x_{n+1}, y_{n+1}[. \quad (2.10)$$

Now we introduce $R_x^{(n)}$ as the quantity of data which we have tried to store at the location x (successfully or not) when n files are stored. These data are the data fallen in $[g_x(\mathcal{R}^{(n)}), x]$ which could not be stored in $[g_x(\mathcal{R}^{(n)}), x]$; so $R_x^{(n)}$ is defined by

$$R_x^{(n)} := -\left(x - g_x(\mathcal{R}^{(n)})\right) + \sum_{\substack{i \leq n \\ x_i \in [g_x(\mathcal{R}^{(n)}), x]}} l_i. \quad (2.11)$$

Note that in queuing systems, $R^{(n)}$ is the workload. This quantity can be expressed using the function $Y^{(n)}$, which sums the sizes of the files arrived at the left of a point x minus the drift term x . It is defined by $Y_0^{(n)} = 0$ and

$$Y_b^{(n)} - Y_a^{(n)} = -(b - a) + \sum_{\substack{i \leq n \\ x_i \in]a, b]}} l_i \quad \text{for } a < b. \quad (2.12)$$

Introducing also its infimum function defined for $x \in \mathbb{R}$ by $I_x^{(n)} := \inf\{Y_y^{(n)} : y \leq x\}$, we get the following expression, for every $n \geq 1$:

$$R^{(n)} = Y^{(n)} - I^{(n)}. \quad (2.13)$$

As a consequence, the covered set when the first n files are stored is given by

$$\mathcal{C}^{(n)} = \{x \in \mathbb{R} : Y^{(n)} - I^{(n)} > 0\}. \quad (2.14)$$

Finally, we introduce the function Y defined on \mathbb{R} by $Y_0 = 0$ and

$$Y_b - Y_a = -(b - a) + \sum_{x_i \in]a, b]}} l_i \quad \text{for } a < b, \quad (2.15)$$

and its infimum I defined for $x \in \mathbb{R}$ by

$$I_x := \inf\{Y_y : y \leq x\}. \quad (2.16)$$

Assuming that the quantity of data arriving on a compact set is finite,

$$\forall L \geq 0, \quad \sum_{x_i \in [-L, L]} l_i < \infty, \quad (2.17)$$

we can let $n \rightarrow \infty$ in (2.14). More precisely, the covering

$$\mathcal{C} = \bigcup_{n \in \mathbb{N}} \mathcal{C}^{(n)} \quad (2.18)$$

is given by the following proposition (see [12, Section 2.1] for the proof).

Proposition 2.1. *Assuming (2.17), one has the following.*

If $\lim_{x \rightarrow -\infty} Y_x = +\infty$, then $\mathcal{C} = \{x \in \mathbb{R} : Y_x - I_x > 0\} \neq \mathbb{R}$.

If $\liminf_{x \rightarrow -\infty} Y_x = -\infty$, then $\mathcal{C} = \{x \in \mathbb{R} : Y_x - I_x > 0\} = \mathbb{R}$.

3. Properties at a Fixed Time and Asymptotics Regimes

3.1. Statistics at a Fixed Time

Our purpose in this section is to specify the distribution of the covering $\mathcal{C}(t)$ using Lévy processes. This characterization will be useful to prove asymptotics results (Theorems 3.6, 3.10 and Corollary 3.11) and for the dynamic results given in [6]. To that end, following the previous section, we consider the process $(Y_x^{(t)})_{x \in \mathbb{R}}$ associated to the PPP $\{(t_i, l_i, x_i) : i \in \mathbb{N}\}$ defined by

$$Y_0^{(t)} := 0; \quad Y_b^{(t)} - Y_a^{(t)} = -(b-a) + \sum_{\substack{t_i \leq t \\ x_i \in]a, b]}} l_i \quad \text{for } a < b, \quad (3.1)$$

which has independent and stationary increments, no negative jumps, and bounded variation. Introducing also its infimum process defined for $x \in \mathbb{R}$ by

$$I_x^{(t)} := \inf \{ Y_y^{(t)} : y \leq x \}, \quad (3.2)$$

we can give now a handy expression for the covering at a fixed time and obtain that the hardware becomes full at a deterministic time equal to $1/m$, which is the random counterpart of Proposition 2.1 (see Section 4 for the proof).

Proposition 3.1. *For every $t < 1/m$, one has $\mathcal{C}(t) = \{x \in \mathbb{R} : Y_x^{(t)} > I_x^{(t)}\} \neq \mathbb{R}$ a.s. For every $t \geq 1/m$, one has $\mathcal{C}(t) = \mathbb{R}$ a.s.*

One can note that in queuing system, tm is the charge and $\mathcal{C}(t) \neq \mathbb{R} \Leftrightarrow tm < 1$ is the standard claim of stability for $tm < 1$, for finite v .

To specify the distribution of $\mathcal{C}(t)$, it is equivalent and more convenient to describe the distribution of its complementary set, denoted by $\mathcal{R}(t)$, which corresponds to the free space of the hardware at time t . By the previous proposition, there is the following identity:

$$\mathcal{R}(t) = \left\{ x \in \mathbb{R} : Y_x^{(t)} = I_x^{(t)} \right\} \quad \text{a.s.} \quad (3.3)$$

We begin by giving some geometric properties of this set, which are classical for finite ν for storage systems (see [13]) and queuing theory (see [14]).

Proposition 3.2. *For every $t \geq 0$, $\mathcal{R}(t)$ is stationary, its closure is symmetric in distribution, and it enjoys the regeneration property: For every $x \in \mathbb{R}$, $(\mathcal{R}(t) - d_x(\mathcal{R}(t))) \cap [0, \infty[$ is independent of $\mathcal{R}(t) \cap] - \infty, x]$ and is distributed as $(\mathcal{R}(t) - d_0(\mathcal{R}(t))) \cap [0, \infty[$.*

Moreover for every $x \in \mathbb{R}$, $\mathbb{P}(x \in \mathcal{C}(t)) = \min(1, mt)$.

Stationarity is plain from the construction of the covering and regeneration property is a direct consequence of Lemma 4.1 given in the next section. Symmetry is then a consequence of [15, Lemma 6.5] or [16, Corollary 7.19]. Computation of $\mathbb{P}(x \in \mathcal{C}(t))$ can be then derived from [15, Theorem 1]. See [12, Section 3.1] for the complete proof.

Even though for each fixed t the distribution of $\mathcal{R}(t)^{\text{cl}}$ is symmetric, the processes $(\mathcal{R}(t)^{\text{cl}} : t \in [0, 1/m])$ and $(-\mathcal{R}(t)^{\text{cl}} : t \in [0, 1/m])$ are quite different. For example, we shall observe in [6] that the left extremity of the data block straddling 0 is a Markov process but the right extremity is not.

We want now to characterize the distribution of the free space $\mathcal{R}(t)$. For this purpose, we need some notation. The drift of the Lévy process $(Y_x^{(t)})_{x \geq 0}$ is equal to -1 , its Lévy measure is equal to $t\nu$, and its Laplace exponent $\Psi^{(t)}$ is then given by (see the appendix for background on Lévy processes)

$$\Psi^{(t)}(\rho) := -\rho + \int_0^\infty (1 - e^{-\rho x}) t\nu(dx). \quad (3.4)$$

For sake of simplicity, we write, recalling (2.1),

$$g(t) := g_0(\mathcal{R}(t)), \quad d(t) = d_0(\mathcal{R}(t)), \quad l(t) = d(t) - g(t), \quad (3.5)$$

which are, respectively, the left extremity, the right extremity, and the length of the data block straddling 0, $\mathbf{B}_0(t)$. Note that $g(t) = d(t) = 0$ if $\mathbf{B}_0(t) = \emptyset$.

We work with \mathcal{R} , a subset of \mathbb{R} of the form $\bigsqcup_{n \in \mathbb{N}} [a_n, b_n[$, and we denote by $\vec{\mathcal{R}} := \bigsqcup_{n \in \mathbb{N}} [-b_n, -a_n[$ the symmetrical of \mathcal{R} with respect to 0 closed at the left, open at the right. We consider the positive part (resp., negative part) of \mathcal{R} defined by

$$\begin{aligned} \vec{\mathcal{R}} &:= (\mathcal{R} - d_0(\mathcal{R})) \cap [0, \infty) = \bigsqcup_{n \in \mathbb{N}: a_n \geq d_0(\mathcal{R})} [a_n - d_0(\mathcal{R}), b_n - d_0(\mathcal{R})[, \\ \overleftarrow{\mathcal{R}} &:= \vec{\mathcal{R}} = \bigsqcup_{n \in \mathbb{N}: b_n \leq g_0(\mathcal{R})} [g_0(\mathcal{R}) - b_n, g_0(\mathcal{R}) - a_n[. \end{aligned} \quad (3.6)$$

Example 3.3. For a given \mathcal{R} represented by the dotted lines, we give below $\vec{\mathcal{R}}$ and $\overleftarrow{\mathcal{R}}$, which are also represented by dotted lines. Moreover the endpoints of the data blocks containing 0 are denoted by g_0 and d_0 (see Figure 2).

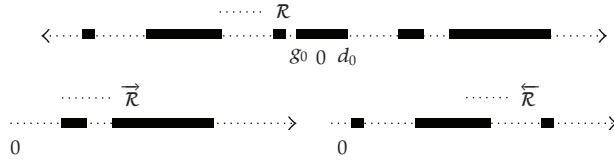


Figure 2

Thus $\mathcal{R}(\vec{t})$ (resp., $\overleftarrow{\mathcal{R}}(t)$) is the free space at the right of $\mathbf{B}_0(t)$ (resp., at the left of $\mathbf{B}_0(t)$, turned over, closed at the left and open at the right). We have then the identity

$$\mathcal{R}(t) = \left(d(t) + \mathcal{R}(\vec{t}) \right) \sqcup \left(-g(t) + \overleftarrow{\mathcal{R}}(t) \right). \quad (3.7)$$

Introducing also the processes $(\bar{\tau}_x^{(t)})_{x \geq 0}$ and $(\overleftarrow{\tau}_x^{(t)})_{x \geq 0}$ defined by

$$\bar{\tau}_x^{(t)} := \inf \left\{ y \geq 0 : \left| \mathcal{R}(\vec{t}) \cap [0, y] \right| > x \right\}, \quad \overleftarrow{\tau}_x^{(t)} := \inf \left\{ y \geq 0 : \left| \overleftarrow{\mathcal{R}}(t) \cap [0, y] \right| > x \right\} \quad (3.8)$$

enables us to describe $\mathcal{R}(t)$ in the following way (see Section 4 for the proof).

Proposition 3.4. (i) The random sets $\mathcal{R}(\vec{t})$ and $\overleftarrow{\mathcal{R}}(t)$ are independent, identically distributed, and independent of $(g(t), d(t))$.

(ii) $\mathcal{R}(\vec{t})$ and $\overleftarrow{\mathcal{R}}(t)$ are the range of the subordinators $\bar{\tau}^{(t)}$ and $\overleftarrow{\tau}^{(t)}$, respectively, whose Laplace exponent $\kappa^{(t)}$ is the inverse function of $-\Psi^{(t)}$.

(iii) The distribution of $(g(t), d(t))$ is specified by

$$\begin{aligned} (g(t), d(t)) &= (-Ul(t), (1-U)l(t)), \\ \mathbb{P}(l(t) \in dx) &= (1-mt) \left(\delta_0(dx) + \mathbb{1}_{\{x>0\}} x \Pi^{(t)}(dx) \right), \end{aligned} \quad (3.9)$$

where U is an uniform random variable on $[0, 1]$ independent of $l(t)$ and $\Pi^{(t)}$ is the Lévy measure of $\kappa^{(t)}$.

Remark 3.5. Such results are classical for regenerative sets (see, e.g., [13, 17, 18]). But we need this particular characterization and expressions given in the proof in the next section for forthcoming results.

3.2. Asymptotics at Saturation of the Hardware

We focus now on the asymptotic behavior of $\mathcal{R}(t)$ when t tends to $1/m$, that is, when the hardware is becoming full. First, note that if ν has a finite second moment, then

$$\mathbb{E}(l(t)) = \frac{\int_0^\infty l^2 \nu(dl)}{(1 - mt)^2}. \quad (3.10)$$

Thus we may expect that if ν has a finite second moment, then $(1 - mt)^2 l(t)$ should converge in distribution as t tends to $1/m$. Indeed, in the particular case $\nu = \delta_1$ or in the conditions of [19, Corollary 2.4], we have an expression of $\Pi^{(t)}(dx)$ and we can prove that $(1 - mt)^2 l(t)$ does converge in distribution to a gamma variable.

More generally, we shall prove that the rescaled free space $(1 - mt)^2 \mathcal{R}(t)$ converges in distribution as t tends to $1/m$. To that end, we need to prove that the process $(Y_{(1-mt)^{-2}x}^{(t)})_{x \in \mathbb{R}}$ converges after suitable rescaling to a random process. Thanks to (3.3), $(1 - mt)^2 \mathcal{R}(t)$ should then converge to the set of points where this limiting process coincides with its infimum process. We shall also handle the case where ν has an infinite second moment and find the correct normalization, which depends on the tail of ν . Proofs are close to those of Section 3.3 and given simultaneously in Section 4.

In queuing systems, asymptotics at saturation are known as heavy traffic approximation ($\rho = tm \rightarrow 1$), which depend similarly on the tail of ν . And for ν finite, results given here could be directly derived from results in queuing theory (see [8, Section III.7.2] or [14] if ν has a second moment order and [7] for heavy tail of ν). The main difference is that ν can be infinite in this paper. Then the busy cycle is not defined and we consider here the whole random set of occupied locations.

To state the main result, we introduce now the following functions defined for every $t \in [0, 1/m[$ and $\alpha \in]1, 2[$ by

$$\epsilon_{2+}(t) = (1 - mt)^2, \quad \epsilon_2(t) = 2 \frac{(1 - mt)^2}{-\log((1 - mt))}, \quad \epsilon_\alpha(t) = (1 - mt)^{\alpha/(\alpha-1)}. \quad (3.11)$$

Recalling Notations of Section 2, we have then the following weak convergence result for the Matheron topology.

Theorem 3.6. *If $\nu \in \mathfrak{D}_\alpha$ ($\alpha \in]1, 2[\cup \{2+\}$), then $\epsilon_\alpha(t) \cdot \mathcal{R}(t)^{\text{cl}}$ converges weakly in $\mathcal{L}(\mathbb{R})$ as t tends to $1/m$ to $\{x \in \mathbb{R} : Y_x^{\alpha,1} = I_x^{\alpha,1}\}^{\text{cl}}$.*

First we prove the convergence of the Laplace exponent $\Psi^{(t)}$ after suitable rescaling as t tends to $1/m$, which ensures the convergence of the Lévy process $Y^{(t)}$ after suitable rescaling (see Lemma 4.2). These convergences will not a priori entail the convergence of the random set $\epsilon_\alpha(t) \cdot \mathcal{R}^{\text{cl}}(t)$ since they do not entail the convergence of excursions. Nevertheless, they will entail the convergence of $\kappa^{(t)}$ since $\kappa^{(t)} \circ (-\Psi^{(t)}) = \text{Id}$ (Lemma 4.4). Then we get the convergence of $\tau^{(t)}$ as t tends to infinity and thus of its range $\epsilon_\alpha(t) \cdot \mathcal{R}^{\text{cl}}(t)$.

Remark 3.7. More generally, as in queuing theory and [5], we can generalize these results for regularly varying functions $\bar{\nu}$. If $\bar{\nu}$ is regularly varying at infinity with index $-\alpha \in]-1, -2[$, then we have the following weak convergence in $\mathcal{H}(\mathbb{R})$:

$$z^{-1} \mathcal{R} \left(\frac{(1 - z\bar{\nu}(z))}{m} \right)^{\text{cl}} \xrightarrow{z \rightarrow \infty} \left\{ x \in \mathbb{R} : Y_x^{\alpha,1} = I_x^{\alpha,1} \right\}^{\text{cl}} \quad \text{with } C = 1. \quad (3.12)$$

For instance, the case $\bar{\nu}(x) \stackrel{x \rightarrow \infty}{\sim} cx^{-\alpha} \log(x)^\beta$ with $(\alpha, \beta, c) \in]1, 2[\times \mathbb{R} \times \mathbb{R}_+^*$ leads to

$$\left((1 - mt) \log \left(\frac{1}{1 - mt} \right)^{-\beta} \right)^{\alpha/(\alpha-1)} \mathcal{R}(t)^{\text{cl}} \xrightarrow{t \rightarrow 1/m} \left\{ x \in \mathbb{R} : Y_x^{\alpha,1} = I_x^{\alpha,1} \right\}^{\text{cl}} \quad \text{with } C = \frac{c}{(\alpha-1)^\beta}. \quad (3.13)$$

If $\bar{\nu}$ is regularly varying at infinity with index -2 , there are many cases to consider.

We get then the asymptotic of $(g(t), d(t))$.

Corollary 3.8. *If $\nu \in \mathfrak{D}_\alpha$ ($\alpha \in]1, 2] \cup \{2+\}$), then $e_\alpha(t) \cdot (g(t), d(t))$ converges weakly as t tends to $1/m$ to $(\sup\{x \leq 0 : Y_x^{\alpha,1} = I_0^{\alpha,1}\}, \inf\{x \geq 0 : Y_x^{\alpha,1} = I_0^{\alpha,1}\})$.*

If $\nu \in \mathfrak{D}_{2+}$ (resp., \mathfrak{D}_2), $e_\alpha(t) \cdot l(t)$ converges weakly to a gamma variable with parameter $(1/2, m/(4m_2))$ (resp., $(1/2, m/4)$).

Remark 3.9. The density of data blocks of size dx in $e_\alpha(t) \cdot \mathcal{R}(t)^{\text{cl}}$ is equal to $(mt/(1 - mt))\Pi^{(\alpha)}(dx)$. By the previous theorem or corollary, this density converges weakly as t tends to $1/m$ to the density of data block of size dx of the limit covering $\{x \in \mathbb{R} : Y_x^{\alpha,1} = I_x^{\alpha,1}\}^{\text{cl}}$. This limit density, denoted by $\Pi^{\alpha,1}(dx)$, can be computed explicitly in the cases $\nu \in \mathfrak{D}_\alpha$ ($\alpha \in \{2, 2+\}$), thanks to the last corollary:

$$\Pi^{2+,1}(dx) = \sqrt{\frac{m}{4\pi m_2 x^3}} \exp\left(-\frac{m}{4m_2}x\right), \quad \Pi^{2,1}(dx) = \sqrt{\frac{m}{4\pi x^3}} \exp\left(-\frac{m}{4}x\right). \quad (3.14)$$

Note that this is also the Lévy measure of the limit covering $\{x \in \mathbb{R} : Y_x^{\alpha,1} = I_x^{\alpha,1}\}^{\text{cl}}$.

3.3. Asymptotic Regime on a Large Part of the Hardware

Here we look at the set of occupied locations $\mathcal{C}(t)$ in a window of size x . We consider the asymptotics of $\mathcal{C}(t) \cap [0, x]$ when x tends to infinity and t tends to the saturation time. As far as we know, results given here are new even when ν is finite. We introduce the following functions defined for all $x \in \mathbb{R}_+^*$ and $\alpha \in]1, 2[$ by

$$f_{2+}(x) = \frac{1}{\sqrt{x}}, \quad f_2(x) = \sqrt{\frac{\log(x)}{x}}, \quad f_\alpha(x) = x^{1/\alpha-1}. \quad (3.15)$$

And we have the following asymptotic regime (see Section 4 for the proof).

Theorem 3.10. *If $\nu \in \mathfrak{D}_\alpha$ ($\alpha \in]1, 2] \cup \{2+\}$), x tends to infinity and t to $1/m$ such that $1 - mt \sim \lambda f_\alpha(x)$ with $\lambda > 0$, then $x^{-1}(\mathcal{R}(t)^{\text{cl}} \cap [0, x])$ converges weakly in $\mathcal{H}([0, 1])$ to $\{z \in [0, 1] : Y_z^{\alpha, \lambda} = I_z^{\alpha, \lambda}\}^{\text{cl}}$.*

Thus as in [4], we observe a phase transition of the size of largest block of data in $[0, x]$ as $x \rightarrow \infty$ according to the rate of filling of the hardware. More precisely, denoting $B_1(x, t) = |I_1(x, t)|$ where $(I_j(x, t))_{j \geq 1}$ is the sequence of component intervals of $\mathcal{C}(t) \cap [0, x]$ ranked by decreasing order of size, we have the following.

Corollary 3.11. *Let $\nu \in \mathfrak{D}_\alpha$ ($\alpha \in]1, 2] \cup \{2+\}$), x tends to infinity and t to $1/m$.*

- (i) *If $1 - mt \sim \lambda f_\alpha(x)$ with $\lambda > 0$, then $B_1(x, t)/x$ converges in distribution to the largest length of excursion of $(Y_z^{\alpha, \lambda} - I_z^{\alpha, \lambda})_{z \in [0, 1]}$.*
- (ii) *If $f_\alpha(x) = o(1 - mt)$, then $B_1(x, t)/x \xrightarrow{\mathbb{P}} 0$.*
- (iii) *If $1 - mt = o(f_\alpha(x))$, then $B_1(x, t)/x \xrightarrow{\mathbb{P}} 1$.*

The phase transition occurs at time t such that $1 - mt \sim \lambda f_\alpha(x)$ with $\lambda > 0$. The more data arrive in small files (i.e., the faster $\bar{\nu}(x)$ tends to zero as x tends to infinity), the later the phase transition occurs. In [4, 5], the hardware is a circle and processes required for asymptotics are the bridges of the processes used here. A consequence is that in our model, $B_1(t, x)/x$ tends to zero or one with a positive probability at phase transition, which is not the case for the parking problem in [4, 5]. More precisely, denoting by $B_{\alpha, \lambda}$ the law of the largest length of excursion of $(Y_x^{\alpha, \lambda} - I_x^{\alpha, \lambda})_{x \in [0, 1]}$, we have

$$\forall (\lambda, \alpha) \in \mathbb{R}_+^* \times]1, 2[\cup \{2+\}, \quad \mathbb{P}(B_{\alpha, \lambda} = 0) > 0, \quad \mathbb{P}(B_{\alpha, \lambda} = 1) > 0. \quad (3.16)$$

3.4. Observations

We give here some complementary results about the distribution of the set of occupied locations at a fixed time and about the storage process.

We have for every $\rho \geq 0$ (use (A.10)),

$$\kappa^{(t)}(\rho) = \rho + \int_0^\infty (1 - e^{-\rho x}) \Pi^{(t)}(dx), \quad (3.17)$$

and using (A.11),

$$\bar{\Pi}^{(t)}(0) = t\bar{\nu}(0), \quad \int_0^\infty x \Pi^{(t)}(dx) = \frac{mt}{1 - mt}. \quad (3.18)$$

Using (A.9), we have also the following identity of measures on $[0, \infty[\times [0, \infty[$

$$x \mathbb{P}\left(\tilde{\tau}_l^{(t)} \in dx\right) dl = l \mathbb{P}\left(-Y_x^{(t)} \in dl\right) dx. \quad (3.19)$$

We can give the distribution of *the extremities of* \mathbf{B}_0 :

$$\mathbb{P}(-g(t) \in dx) = \mathbb{P}(d(t) \in dx) = (1 - mt) \left(\delta_0(dx) + \mathbb{1}_{\{x>0\}} \bar{\Pi}^{(t)}(x) dx \right). \quad (3.20)$$

Writing $-g(t) = \gamma(t)$ (see (4.9) and (4.10)) and using the identity of fluctuation (A.17) gives an other expression for the Laplace transform of $g(t)$. For all $t \in [0, 1/m[$ and $\lambda \geq 0$, we have

$$\mathbb{E}(\exp(\lambda g(t))) = \exp\left(\int_0^\infty (e^{-\lambda x} - 1) x^{-1} \mathbb{P}(Y_x^{(t)} > 0) dx\right). \quad (3.21)$$

As a consequence, we see that the law of $g(t)$ is infinitively divisible. Moreover this expression will give the generating triplet of the additive process $(g(t))_{t \in [0, 1/m[}$ [6, Theorem 2, Section 4].

We can also estimate *the number of data blocks* on the hardware. If ν has a finite mass, we write $N_x^{(t)}$ as the number of data blocks of the hardware restricted to $[-x, x]$ at time t . This quantity has a deterministic asymptotic as x tends to infinity which is maximum at time $1/(2m)$. And the number of blocks of the hardware reaches a.s. its maximal at time $1/(2m)$. More precisely, we have the following.

Proposition 3.12. *If $\bar{\nu}(0) < \infty$, then for every $t \in [0, 1/m[$,*

$$\lim_{x \rightarrow \infty} \frac{N_x^{(t)}}{2x} = \bar{\nu}(0)t(1 - mt) \quad a.s. \quad (3.22)$$

Finally, we can describe here *the hashing of data*. We recall that a file labeled by i is stored at location x_i . In the hashing problem, one is interested by the location where the file i is stored knowing x_i . By stationarity, we can take $x_i = 0$ and consider a file of size l which we store at time t at location 0 on the hardware whose free space is equal to $\mathcal{R}(t)$. The first point (resp., the last point) of the hardware occupied for the storage of this file is equal to $d(t)$ (resp., to $d(t) + \bar{\tau}_l^{(t)}$). This gives the distribution of the extremities of the portion of the hardware used for the storage of a file.

3.5. Examples

Let us now consider three explicit examples.

(1) The basic example is when $\nu = \delta_1$ (all files have the same unit size as in the original parking problem in [4]). Then for all $x \in \mathbb{R}_+$ and $n \in \mathbb{N}$,

$$\begin{aligned} \mathbb{P}(Y_x^{(t)} + x = n) &= e^{-tx} \frac{(tx)^n}{n!}, \\ \mathbb{P}(\bar{\tau}_x^{(t)} = x + n) &= \frac{x}{x+n} e^{-t(x+n)} \frac{(t(n+x))^n}{n!}, \end{aligned} \quad (3.23)$$

where the second identity follows from integrating (3.19). Then,

$$\Pi^{(t)}(n) = \frac{(tn)^n}{n.n!} e^{-tn}, \quad (3.24)$$

and $l(t)$ follows a size-biased Borel law:

$$\mathbb{P}(l(t) = n) = (1-t) \frac{(tn)^n}{n!} e^{-tn}. \quad (3.25)$$

(2) Another example where calculus can be made explicitly is the gamma case when $\nu(dl) = \mathbb{1}_{\{l \geq 0\}} l^{-1} e^{-l} dl$. Note that $\bar{\nu}(0) = \infty$ and $m = 1$. Then, for every $x \in \mathbb{R}_+$,

$$\begin{aligned} \mathbb{P}\left(Y_x^{(t)} \in dz\right) &= \mathbb{1}_{[-x, \infty[}(z) \Gamma(tx)^{-1} e^{-(z+x)} (z+x)^{tx-1} dz, \\ \mathbb{P}\left(\bar{\tau}_x^{(t)} \in dz\right) &= \mathbb{1}_{[x, \infty[}(z) x(z\Gamma(tz))^{-1} e^{-(z-x)} (z-x)^{tz-1} dz. \end{aligned} \quad (3.26)$$

Further

$$\begin{aligned} \Pi^{(t)}(dz) &= (z\Gamma(tz))^{-1} e^{-z} z^{tz-1} dz, \\ \mathbb{P}(l(t) \in dx) &= (1-t) \left(\delta_0(dx) + \Gamma(tz)^{-1} e^{-x} x^{tx-1} dx \right). \end{aligned} \quad (3.27)$$

(3) For the exponential distribution $\nu(dl) = \mathbb{1}_{\{l \geq 0\}} e^{-l} dl$, we can get

$$\Psi^{(t)}(\lambda) = \lambda \left(-1 + \frac{t}{\lambda+1} \right), \quad \kappa^{(t)}(\lambda) = \frac{\lambda+t-1 + \sqrt{(\lambda+t-1)^2 + 4\lambda}}{2}. \quad (3.28)$$

4. Proofs

In this section, we provide rigorous arguments for the original results which have been stated in Section 3.

Proof of Proposition 3.1. First $m < \infty$ entails that for all $L \geq 0$, $\sum_{t_i \leq t, x_i \in [-L, L]} l_i < \infty$ a.s. and condition (A.7) is satisfied a.s. Then, by Proposition 2.1,

$$\mathcal{C}(t) = \left\{ x \in \mathbb{R} : Y_x^{(t)} - I_x^{(t)} > 0 \right\} \quad \text{a.s.} \quad (4.1)$$

(i) If $t < 1/m$, then $\mathbb{E}(Y_{-1}^{(t)}) = 1 - mt > 0$ and the càdlàg version of $(Y_{(-x)}^{(t)})_{x \geq 0}$ is a Lévy process. So we have (see [20, Corollary 2, page 190])

$$Y_x^{(t)} \xrightarrow{x \rightarrow -\infty} \infty \quad \text{a.s.} \quad (4.2)$$

Then Proposition 2.1 ensures that for every $t < 1/m$, $\mathcal{C}(t) = \{x \in \mathbb{R} : Y_x^{(t)} > I_x^{(t)}\} \neq \mathbb{R}$ a.s.

(ii) If $t \geq 1/m$, then $\mathbb{E}(Y_{-1}^{(t)}) \leq 0$ ensures (see [20, Corollary 2, page 190]) that

$$Y_x^{(t)} \xrightarrow{x \rightarrow -\infty} -\infty \text{ a.s. or } (Y_x^{(t)})_{x \leq 0} \text{ oscillates a.s. in } -\infty. \quad (4.3)$$

Similarly, we get that for every $t \geq 1/m$, $\mathcal{C}(t) = \mathbb{R}$ a.s. \square

For the forthcoming proofs, we fix $t \in [0, 1/m[$, which is omitted from the notation of processes for the sake of simplicity.

To prove the regeneration property and characterize the Laplace exponent of $\vec{\tau}$, we need to establish first a regeneration property at the right extremities of the data blocks. In that view, we consider, for every $x \geq 0$, the files arrived at the left/at the right of x before time t :

$$\rho_x := \{(t_i, x_i, l_i) : t_i \leq t, x_i \leq x\}, \quad \rho^x := \{(t_i, x_i - x, l_i) : t_i \leq t, x_i > x\}. \quad (4.4)$$

Lemma 4.1. For all $x \geq 0$, $\rho^{d_x(\mathcal{R}(t))}$ is independent of $\rho_{d_x(\mathcal{R}(t))}$ and distributed as ρ^0 .

Proof. The simple Markov property for PPP states that, for every $x \in \mathbb{R}$, ρ^x is independent of ρ_x and distributed as ρ^0 . Clearly this extends to simple stopping times in the filtration $\sigma(\rho_x)_{x \in \mathbb{R}}$ and further to any stopping time in this filtration using the classical argument of approximation of stopping times by a decreasing sequence of simple stopping times (see also [21]). As $d_x(\mathcal{R}(t))$ is a stopping time in this filtration, $\rho^{d_x(\mathcal{R}(t))}$ is independent of $\rho_{d_x(\mathcal{R}(t))}$ and distributed as ρ^0 . \square

Using well-known results on Lévy processes, which we recall in Section 4, we can prove Proposition 3.4.

Proof of Proposition 3.4. (i) By symmetry, $\mathcal{R}(t)^{\text{cl}}$, $\vec{\mathcal{R}}(t)$, and $\overleftarrow{\mathcal{R}}(t)$ are identically distributed. The regeneration property ensures that $\vec{\mathcal{R}}(t)$ is independent of $(\overleftarrow{\mathcal{R}}(t), g(t), d(t))$. By symmetry, $\overleftarrow{\mathcal{R}}(t)$ is independent of $(g(t), d(t), \vec{\mathcal{R}}(t))$. So $\vec{\mathcal{R}}(t)$, $\overleftarrow{\mathcal{R}}(t)$, and $(g(t), d(t))$ are independent.

(ii) As $\vec{\mathcal{R}}(t)$ is a.s. the union of intervals of the form $[a, b[$, then $x \rightarrow |\mathcal{R}(t) \cap [0, x]|$ increases at $x \in \vec{\mathcal{R}}(t)$. So, for every $x \geq 0$,

$$\vec{\tau}_{|\mathcal{R}(t) \cap [0, x]|} = d_x(\mathcal{R}(t)), \quad \vec{\tau}_x = d_{\vec{\tau}_x}(\mathcal{R}(t)) \text{ a.s.} \quad (4.5)$$

So the range of $\vec{\tau}$ is equal to $\vec{\mathcal{R}}(t)$. The fact that $\vec{\tau}$ is a subordinator will be proved below but could be also derived directly from the regeneration property of $\vec{\mathcal{R}}(t)$ (see [17]). Similarly the range of $\overleftarrow{\tau}$ is equal to $\overleftarrow{\mathcal{R}}(t)$.

Moreover, $dY = -1$ on $\mathcal{R}(t)$ and $Y_{a-} = Y_b$ if $[a, b[$ is an interval component of $\mathcal{C}(t)$. By integrating on $[d(t), d(t) + y]$, we have a.s for every $y \geq 0$ such that $d(t) + y \in \mathcal{R}(t)$,

$$Y_{y+d(t)} - Y_y = -|\mathcal{R}(t) \cap [d(t), d(t) + y]|. \quad (4.6)$$

Then using again the definition of $\bar{\tau}$ given in Section 3.1 and that $\mathcal{R}(\bar{t})$ is the range of $\bar{\tau}$,

$$\begin{aligned}\bar{\tau}_x &= \inf\{y \geq 0 : y \in \mathcal{R}(\bar{t}), |\mathcal{R}(\bar{t}) \cap [0, y]| > x\} \\ &= \inf\{y \geq 0 : d(t) + y \in \mathcal{R}(t), |\mathcal{R}(t) \cap [d(t), d(t) + y]| > x\} \\ &= \inf\{y \geq 0 : Y_{y+d(t)} - Y_{d(t)} < -x\}.\end{aligned}\tag{4.7}$$

Moreover

$$Y_{y+d(t)} - Y_{d(t)} = -y + \sum_{\substack{(t_i, x_i, l_i) \in \mathcal{P}^{d(t)} \\ 0 \leq x_i \leq y}} l_i,\tag{4.8}$$

and Lemma 4.1 entails that $\mathcal{P}^{d(t)}$ is distributed as a PPP on $[0, t] \times \mathbb{R}_+ \times \mathbb{R}_+$ with intensity $ds \otimes dx \otimes \nu(dl)$. So $(Y_{y+d(t)} - Y_{d(t)})_{y \geq 0}$ is a Lévy process with bounded variation and drift -1 which verifies condition (A.7) (use (A.5) and $-1 + mt < 0$). Then Theorem A.1 in the appendix entails that $\bar{\tau}$ is a subordinator whose Laplace exponent is the inverse function of $-\Psi^{(t)}$.

As $\bar{\mathcal{R}}(t)$ is distributed as $\mathcal{R}(\bar{t})$, $\bar{\tau}$ is distributed as $\bar{\tau}$ by definition.

(iii) We determine now the distribution of $(g(t), d(t))$ using fluctuation theory, which enables us to get identities useful for the rest of the work. We write $(\tilde{Y}_x)_{x \geq 0}$ for the càdlàg version of $(-Y_{-x})_{x \geq 0}$ and

$$S(t) := \sup\{\tilde{Y}_x, x \geq 0\} = -I_0, \quad \gamma(t) := \arg(S(t)) = \inf\{x \geq 0 : \tilde{Y}_x = S(t)\}.\tag{4.9}$$

Using (3.3) and the fact that Y has no negative jumps, we have

$$\begin{aligned}g(t) &= g_0(\mathcal{R}(t)) = \sup\{x \leq 0 : Y_x = I_x\} \\ &= \sup\{x \leq 0 : Y_{x-} = I_0\} = -\inf\{x \geq 0 : \tilde{Y}_x = -I_0\} \\ &= -\gamma(t).\end{aligned}\tag{4.10}$$

Using again (3.3) and the fact that $(Y_x)_{x \geq 0}$ is regular for $]-\infty, 0[$ (see [20, Proposition 8, page 84]), we have also a.s.

$$\begin{aligned}d(t) &= \inf\{x > 0 : Y_x = I_x\} = \inf\{x > 0 : Y_x = I_0\} \\ &= \inf\{x > 0 : Y_x < I_0\} = \inf\{x > 0 : Y_x < -S(t)\} = T_{S(t)},\end{aligned}\tag{4.11}$$

where $(T_x)_{x \geq 0}$ is distributed as $(\bar{\tau}_x)_{x \geq 0}$ by (4.7) and $(T_x)_{x \geq 0}$ is independent of $(S(t), \gamma(t))$ since $(Y_x)_{x \geq 0}$ is independent of $(Y_x)_{x \leq 0}$. Then for all $\lambda, \mu \geq 0$ with $\lambda \neq \mu$,

$$\begin{aligned} \mathbb{E}(\exp(\lambda g(t) - \mu d(t))) &= \mathbb{E}(\exp(-\lambda \gamma(t)) \mathbb{E}(\exp(-\mu T_{S(t)}))) \\ &= \mathbb{E}\left(\exp\left(-\lambda \gamma(t) - \kappa^{(t)}(\mu) S(t)\right)\right) \end{aligned} \quad (4.12)$$

$$\begin{aligned} &= -\left[\Psi^{(t)}\right]'(0) \frac{\kappa^{(t)}(\lambda) - \kappa^{(t)}(\mu)}{\lambda - \mu} \quad \text{using (A.15)} \\ &= (1 - mt) \frac{\kappa^{(t)}(\lambda) - \kappa^{(t)}(\mu)}{\lambda - \mu} \quad \text{using (A.5),} \end{aligned} \quad (4.13)$$

which gives the distributions of $d(t)$, $g(t)$, and $l(t)$ letting, respectively, $\lambda = 0$, $\mu = 0$, and $\lambda \rightarrow \mu$. Computing then the Laplace transform of $(-Ul(t), (1 - U)l(t))$ where U is a uniform random variable on $[0, 1]$ independent of $l(t)$ gives the right-hand side of (20). So $(g(t), d(t)) = (-U'l(t), (1 - U')l(t))$, where U' is a uniform random variable on $[0, 1]$ independent of $l(t)$. \square

Proofs of Theorems 3.6 and 3.10 are close and made simultaneously. For that purpose, we introduce now $\Psi^{\alpha, \lambda}$ as the Laplace exponent (see (A.1)) of $Y^{\alpha, \lambda}$ given, for $y \geq 0$, $\lambda \geq 0$, and $\alpha \in]1, 2[$ by

$$\Psi^{2+, \lambda}(y) = -\lambda y - \frac{m_2}{m} \frac{y^2}{2}, \quad \Psi^{2, \lambda}(y) = -\lambda y - \frac{C}{m} \frac{y^2}{2}, \quad \Psi^{\alpha, \lambda}(y) = -\lambda y - (C_\alpha y)^\alpha. \quad (4.14)$$

We denote by \mathbb{D} the space of càdlàg function from \mathbb{R}_+ to \mathbb{R} which we endow with the Skorokhod topology (see [22, page 292]). First, we prove the weak convergence of $Y^{(t)}$ after suitable rescaling.

Lemma 4.2. *If $\nu \in \mathfrak{D}_\alpha$ ($\alpha \in]1, 2] \cup \{2+\}$), then for all $y \geq 0$ and $\lambda > 0$,*

$$\begin{aligned} \epsilon_\alpha(t)^{-1} \Psi^{(t)}\left(\epsilon_\alpha(t)(1 - mt)^{-1} y\right) &\xrightarrow{t \rightarrow 1/m} \Psi^{\alpha, 1}(y), \\ x \Psi^{((1 - \lambda f_\alpha(x))/m)}\left((x f_\alpha(x))^{-1} y\right) &\xrightarrow{x \rightarrow \infty} \Psi^{\alpha, \lambda}(y), \end{aligned} \quad (4.15)$$

which entail the following weak convergences of processes in \mathbb{D} :

$$\begin{aligned} \left(\epsilon_\alpha(t)(1 - mt)^{-1} Y_{\epsilon_\alpha(t)^{-1} y}^{(t)}\right)_{y \geq 0} &\xrightarrow{t \rightarrow 1/m} \left(Y_y^{\alpha, 1}\right)_{y \geq 0}, \\ \left((x f_\alpha(x))^{-1} Y_{x y}^{((1 - \lambda f_\alpha(x))/m)}\right)_{y \geq 0} &\xrightarrow{x \rightarrow \infty} \left(Y_y^{\alpha, \lambda}\right)_{y \geq 0}. \end{aligned} \quad (4.16)$$

Remark 4.3. If $\bar{\nu}$ is regularly varying at infinity with index $-\alpha \in]-1, -2[$, then $\bar{\nu}(x)^{-1} \Psi^{((1 - \lambda x \bar{\nu}(x))/m)}(x^{-1} y)$ converges to $\Psi^{\alpha, \lambda}(y)$ as x tends to infinity.

Proof of Lemma 4.2. Using (A.4), we have

$$\begin{aligned}
x\Psi^{(t)}(y) &= xy \left(mt - 1 - t \int_0^\infty (1 - e^{-yu}) \bar{v}(u) du \right) \\
&= xy \left(mt - 1 - ty^{-1} \int_0^\infty e^{-yu} \bar{v}(u) du \right) \\
&= x \left(y(mt - 1) - t \int_0^\infty e^{-yu} \bar{v}(u) du \right),
\end{aligned} \tag{4.17}$$

where $\bar{v}(u) = \int_u^\infty \bar{v}(v) dv$. The first part of the lemma then follows by applying the Tauberian theorem in [20, page 10] which gives the asymptotic behavior of the last term. For a detailed proof, we refer to [12].

These convergences ensure the convergence of the finite-dimensional distributions of the processes. The weak convergence in \mathbb{D} , which is the second part of the lemma, follows from [23, Theorem 13.17]. \square

In the spirit of Section 3.1, we introduce the expected limit set, that is, the free space of the covering associated with $Y^{\alpha,\lambda}$, and the extremities of the block containing 0:

$$\mathcal{R}(\alpha, \lambda) := \left\{ x \in \mathbb{R} : Y_x^{\alpha,\lambda} = I_x^{\alpha,\lambda} \right\}, \quad g(\alpha, \lambda) := g_0(\mathcal{R}(\alpha, \lambda)), \quad d(\alpha, \lambda) := d_0(\mathcal{R}(\alpha, \lambda)). \tag{4.18}$$

We have the following analog of Proposition 3.4. $\mathcal{R}(\vec{\alpha}, \lambda)$ and $\mathcal{R}(\overleftarrow{\alpha}, \lambda)$ are independent, identically distributed and independent of $(g(\alpha, \lambda), d(\alpha, \lambda))$. Moreover $\mathcal{R}(\vec{\alpha}, \lambda)$ and $\mathcal{R}(\overleftarrow{\alpha}, \lambda)$ are, respectively, the range of the subordinators $\vec{\tau}^{\alpha,\lambda}$ and $\overleftarrow{\tau}^{\alpha,\lambda}$, whose Laplace exponent $\kappa^{\alpha,\lambda}$ is the inverse function of $-\Psi^{\alpha,\lambda}$. Finally, using $[\Psi^{\alpha,\lambda}]'(0) = -\lambda$, the counterpart of (4.12) gives for $\rho, \mu \geq 0$ and $\rho \neq \mu$,

$$\mathbb{E}(\exp(\rho g(\alpha, \lambda) - \mu d(\alpha, \lambda))) = \lambda \frac{\kappa^{\alpha,\lambda}(\rho) - \kappa^{\alpha,\lambda}(\mu)}{\rho - \mu}. \tag{4.19}$$

The proof of these results follow the proof of Proposition 3.4, except for two points.

(1) We cannot use the point process of files to prove the stationarity and regeneration property of $\mathcal{R}(\alpha, \lambda)$ and we must use the process $Y^{\alpha,\lambda}$ instead. The stationarity is a direct consequence of the stationarity of $(Y_x^{\alpha,\lambda} - I_x^{\alpha,\lambda})_{x \in \mathbb{R}}$. The regeneration property is a consequence of the counterpart of Lemma 4.1 which can be stated as follows. For all $x \in \mathbb{R}$,

$$\left(Y_{d_x(\mathcal{R}(\alpha,\lambda))+y}^{\alpha,\lambda} - Y_{d_x(\mathcal{R}(\alpha,\lambda))}^{\alpha,\lambda} \right)_{y \geq 0} \text{ is independent of } \left(Y_{d_x(\mathcal{R}(\alpha,\lambda))-y}^{\alpha,\lambda} - Y_{d_x(\mathcal{R}(\alpha,\lambda))}^{\alpha,\lambda} \right)_{y \geq 0} \tag{4.20}$$

and distributed as $(Y_y^{\alpha,\lambda})_{y \geq 0}$. As Lemma 4.1, this property is an extension to the stopping time $d_x(\mathcal{R}(\alpha, \lambda))$ of the following obvious result: $(Y_{x+y}^{\alpha,\lambda} - Y_x^{\alpha,\lambda})_{y \geq 0}$ is independent of $(Y_{x-y}^{\alpha,\lambda} - Y_x^{\alpha,\lambda})_{y \geq 0}$ and distributed as $(Y_y^{\alpha,\lambda})_{y \geq 0}$.

(2) It is convenient to define directly $(\bar{\tau}_x^{\alpha,\lambda})_{x \geq 0}$ by

$$\bar{\tau}_x^{\alpha,\lambda} := \inf \left\{ y \geq 0 : Y_{d(\alpha,\lambda)+y}^{\alpha,\lambda} - Y_{d(\alpha,\lambda)}^{\alpha,\lambda} < -x \right\}. \quad (4.21)$$

For $\lambda > 0$, $[\Psi^{\alpha,\lambda}]'(0) = -\lambda < 0$ and so we can apply Theorem A.1 and $\bar{\tau}^{\alpha,\lambda}$ is a subordinator whose Laplace $\kappa^{\alpha,\lambda}$ is the inverse function of $-\Psi^{\alpha,\lambda}$. Moreover its range is a.s. equal to $\mathcal{R}(\bar{\alpha}, \lambda)$, since the Lévy process $(Y_{d(\alpha,\lambda)+y}^{\alpha,\lambda} - Y_{d(\alpha,\lambda)}^{\alpha,\lambda})_{y \geq 0}$ is regular for $] -\infty, 0[$ [20, Proposition 8, page 84].

To prove Theorems 3.6 and 3.10, we need a final lemma, which states the convergence of the Laplace exponent of $\mathcal{R}(\bar{t})$.

Lemma 4.4. *If $\nu \in \mathfrak{D}_\alpha$ ($\alpha \in]1, 2] \cup \{2+\}$), then for all $z \geq 0$ and $\lambda > 0$,*

$$\begin{aligned} (1 - mt)\epsilon_\alpha(t)^{-1} \kappa^{(t)}(\epsilon_\alpha(t)z) &\xrightarrow{t \rightarrow 1/m} \kappa^{\alpha,1}(z), \\ x f_\alpha(x) \kappa^{((1-\lambda f_\alpha(x))/m)}(x^{-1}z) &\xrightarrow{x \rightarrow \infty} \kappa^{\alpha,\lambda}(z). \end{aligned} \quad (4.22)$$

Remark 4.5. If $\bar{\nu}$ is regularly varying at infinity of index $-\alpha \in]-1, -2[$, we have similarly

$$\bar{\nu}(x)^{-1} \kappa^{((1-\lambda x \bar{\nu}(x))/m)}(x^{-1}z) \xrightarrow{x \rightarrow \infty} \kappa^{\alpha,\lambda}(z). \quad (4.23)$$

Proof. First we prove that

$$\alpha(t) \xrightarrow{t \rightarrow 1/m} \beta(t) \implies \kappa^{(t)}(\alpha(t)) \xrightarrow{t \rightarrow 1/m} \kappa^{(t)}(\beta(t)). \quad (4.24)$$

Indeed the function $u \in \mathbb{R}_+^* \mapsto (1 - e^{-u})/u$ decreases so for all $x \geq 0$ and $u, v > 0$, we have

$$\min\left(\frac{u}{v}, 1\right) \leq \frac{1 - e^{-ux}}{1 - e^{-vx}} \leq \max\left(\frac{u}{v}, 1\right). \quad (4.25)$$

This gives

$$\min\left(\frac{\alpha(t)}{\beta(t)}, 1\right) \leq \frac{\int_0^\infty (1 - e^{-\alpha(t)x}) \Pi^{(t)}(dx)}{\int_0^\infty (1 - e^{-\beta(t)x}) \Pi^{(t)}(dx)} \leq \max\left(\frac{\alpha(t)}{\beta(t)}, 1\right) \quad (4.26)$$

and proves (4.24) recalling (3.17).

Then the first part of Lemma 4.2 and the identity $\kappa^{(t)} \circ (-\Psi^{(t)}) = \text{Id}$ give the first part of Lemma 4.4. Indeed for every $y \geq 0$, $\Psi^{(t)}(\epsilon_\alpha(t)(1 - mt)^{-1}y) \xrightarrow{t \rightarrow 1/m} \epsilon_\alpha(t)\Psi^{\alpha,1}(y)$. So (4.24) entails

$$\epsilon_\alpha(t)(1 - mt)^{-1}y \xrightarrow{t \rightarrow 1/m} \kappa^{(t)}\left(-\epsilon_\alpha(t)\Psi^{\alpha,1}(y)\right). \quad (4.27)$$

Put $y = \kappa^{\alpha,1}(z)$ to get the first limit of the lemma and follow the same way to get the second one. \square

Proof of Theorem 3.6. First, by (4.13), we have

$$\mathbb{E}(\exp(\rho\epsilon_\alpha(t)g(t) - \mu\epsilon_\alpha(t)d(t))) = (1 - mt) \frac{\kappa^{(t)}(\epsilon_\alpha(t)\rho) - \kappa^{(t)}(\epsilon_\alpha(t)\mu)}{\epsilon_\alpha(t)(\rho - \mu)}. \quad (4.28)$$

Letting $t \rightarrow 1/m$ using Lemma 4.4 gives the right-hand side of (4.19). Then $\epsilon_\alpha(t) \cdot (g(t), d(t))$ converges weakly as t tends to $1/m$ to $(g(\alpha, 1), d(\alpha, 1))$.

Moreover $\epsilon_\alpha(t)\overleftarrow{\mathcal{R}}(t)^{\text{cl}}$ (resp., $\epsilon_\alpha(t)\overleftarrow{\mathcal{R}}(t)$) converges weakly in $\mathcal{L}(\mathbb{R}_+)$ as t tends to $1/m$ to $\overleftarrow{\mathcal{R}}(\alpha, 1)^{\text{cl}}$ (resp., $\overleftarrow{\mathcal{R}}(\alpha, 1)$). Indeed, by [11, Proposition 3.9], this is a consequence of the convergence of the Laplace exponent of $\epsilon_\alpha(t)\overleftarrow{\mathcal{R}}(t)$ given by Lemma 4.4. Informally, $\epsilon_\alpha(t)\overleftarrow{\mathcal{R}}(t)^{\text{cl}}$ is the range of $(\epsilon_\alpha(t)\overleftarrow{\mathcal{T}}_{(1-mt)\epsilon_\alpha(t)^{-1}z}^{(t)})_{z \geq 0}$ whose convergence in \mathbb{D} follows from Lemma 4.4.

We can now prove the theorem. We know from (3.7) that

$$\epsilon_\alpha(t)\mathcal{R}(t) = \epsilon_\alpha(t) \cdot (d(t) + \overleftarrow{\mathcal{R}}(t)) \sqcup \left(\epsilon_\alpha(t) \cdot \widetilde{(-g(t) + \overleftarrow{\mathcal{R}}(t))} \right), \quad (4.29)$$

where $\epsilon_\alpha(t)\overleftarrow{\mathcal{R}}(t)$, $\epsilon_\alpha(t)(-g(t), d(t))$, and $\epsilon_\alpha(t)\overleftarrow{\mathcal{R}}(t)$ are independent by Proposition 3.4. Similarly

$$\mathcal{R}(\alpha, 1) = (d(\alpha, 1) + \overleftarrow{\mathcal{R}}(\alpha, 1)) \sqcup \left(-g(\alpha, 1) + \widetilde{\overleftarrow{\mathcal{R}}(\alpha, 1)} \right), \quad (4.30)$$

where $\overleftarrow{\mathcal{R}}(\alpha, 1)$, $(-g(\alpha, 1), d(\alpha, 1))$, and $\overleftarrow{\mathcal{R}}(\alpha, 1)$ are independent. As remarked above, we have also the following weak convergences as t tends to $1/m$:

$$\epsilon_\alpha(t)\overleftarrow{\mathcal{R}}(t)^{\text{cl}} \Rightarrow \overleftarrow{\mathcal{R}}(\alpha, 1)^{\text{cl}}, \quad \epsilon_\alpha(t)(-g(t), d(t)) \Rightarrow (-g(\alpha, 1), d(\alpha, 1)), \quad \epsilon_\alpha(t)\overleftarrow{\mathcal{R}}(t)^{\text{cl}} \Rightarrow \overleftarrow{\mathcal{R}}(\alpha, 1)^{\text{cl}}. \quad (4.31)$$

So $\epsilon_\alpha(t)\mathcal{R}(t)^{\text{cl}}$ converges weakly to $\mathcal{R}(\alpha, 1)^{\text{cl}}$ in $\mathcal{L}(\mathbb{R})$ as t tends to $1/m$. \square

Proof of Corollary 3.8. The first result is a direct consequence of Theorem 3.6. We have then

$$\epsilon_\alpha(t)l(t) \xrightarrow{t \rightarrow 1/m} d(\alpha, 1) - g(\alpha, 1). \quad (4.32)$$

Moreover, as $\kappa^{2+,1} \circ (-\Psi^{2+,1}) = \text{Id}$, we can compute $\kappa^{2+,1}$ and (4.19) gives

$$\begin{aligned} & \mathbb{E}(\exp(-\mu(d(2+,1) - g(2+,1))) \\ &= (\kappa^{2+,1})'(\mu) = \left(\frac{1 + \sqrt{1 + 2(m_2/m)\mu}}{m_2/m} \right)'(\mu) = \frac{1}{\sqrt{-1 + 2(m_2/m)\mu}}. \end{aligned} \quad (4.33)$$

So, by identification of Laplace transform, $d(\alpha, 1) - g(\alpha, 1)$ is a gamma variable of parameter $(1/2, m/(4m_2))$ and we get the result. The argument is similar in the case $\alpha = 2$. \square

Proof of Theorem 3.10. The argument is similar to that of the proof of Theorem 3.6 using the other limits of Lemma 4.4. We get that if $x \rightarrow \infty$ and $1 - mt \sim \lambda f_\alpha(x)$ with $\lambda > 0$, then $x^{-1}\mathcal{R}(t)$ converges weakly in $\mathcal{L}(\mathbb{R})$ to $\{x \in \mathbb{R} : Y_x^{\alpha,\lambda} = I_x^{\alpha,\lambda}\}^{\text{cl}}$. The theorem follows by restriction to $[0, 1]$. \square

To prove the corollary of Theorem 3.10, we need the following result.

Lemma 4.6. *The largest length of excursion of $(Y_x^{\alpha,\lambda} - I_x^{\alpha,\lambda})_{x \in [0,1]}$, denoted by $B_{\alpha,\lambda}$, converges in probability to 0 as λ tends to infinity and to 1 as λ tends to 0.*

Proof. (i) Let $0 \leq a < b \leq 1$. Note that for all $\lambda' \geq 1$ and $x \geq 0$, $Y_x^{\alpha,\lambda'} - Y_x^{\alpha,1} = (1 - \lambda')x$ ensures that $I_x^{\alpha,\lambda'} - I_x^{\alpha,1} \geq (1 - \lambda')x$. Then,

$$Y_{a+2((b-a)/3)}^{\alpha,\lambda'} - I_{a+(b-a)/3}^{\alpha,\lambda'} \leq Y_{a+2((b-a)/3)}^{\alpha,1} - I_{a+(b-a)/3}^{\alpha,1} + (1 - \lambda')\frac{b-a}{3}. \quad (4.34)$$

So a.s. there exists λ' such that

$$Y_{a+2((b-a)/3)}^{\alpha,\lambda'} < I_{a+(b-a)/3}^{\alpha,\lambda'}. \quad (4.35)$$

As $Y^{\alpha,\lambda'}$ has no negative jumps, it reaches its infimum on $] - \infty, 2(b-a)/3]$ in a point $c \in [a + (b-a)/3, a + 2(b-a)/3]$. Then a.s. there exists $c \in [a + (b-a)/3, a + 2(b-a)/3]$ and $\lambda' > 0$ such that $c \in \mathcal{R}(\alpha, \lambda')$, which entails that c does not belong to the interior of $B_{\alpha,\lambda}$. Adding that $B_{\alpha,\lambda}$ decreases as λ increases, this property ensures that $B_{\alpha,\lambda}$ converges in probability to 0 as λ tends to infinity.

(ii) As $(Y_x^{\alpha,0})_{x \in \mathbb{R}}$ oscillates when x tends to $-\infty$ (see [20, Corollary 2, page 190]), then

$$I_0^{\alpha,\lambda} \xrightarrow{\lambda \rightarrow 0} -\infty, \quad (4.36)$$

which ensures that $B_{\alpha,\lambda}$ converges in probability to 1 as λ tends to 0. \square

Proof of Corollary 3.11. The first result is a direct consequence of Theorem 3.10.

If $o(1 - mt) = f_\alpha(x)$ ($x \rightarrow \infty$), then for every $\lambda > 0$ and x large enough, $t \leq (1 - \lambda f_\alpha(x))/m$ and

$$\frac{B_1(x, t)}{x} \leq \frac{B_1(x, (1 - \lambda f_\alpha(x))/m)}{x}. \quad (4.37)$$

The right-hand side converges weakly to $B_{\alpha,\lambda}$ as x tends to infinity. Letting λ tend to infinity, the lemma above entails that $B_1(x, t)/x \xrightarrow{x \rightarrow \infty} 0$ in \mathbb{P} .

Similarly if $1 - mt = o(f_\alpha(x))$ ($x \rightarrow \infty$), then for every $\lambda > 0$ and x large enough,

$$\frac{B_1(x, t)}{x} \geq \frac{B_1(x, (1 - \lambda f_\alpha(x))/m)}{x}. \quad (4.38)$$

Letting λ tend to 0, Lemma 4.6 entails that $B_1(x, t)/x \xrightarrow{x \rightarrow \infty} 1$ in \mathbb{P} . \square

Finally, we prove the result stated in Section 3.4.

Proof of Proposition 3.12. As $\bar{\nu}(0) < \infty$, then $\bar{\Pi}(0) = t\bar{\nu}(0) < \infty$ (see (3.18)). So $\bar{\tau}$ is the sum of a drift and a compound Poisson process. That is, there exists a Poisson process $(N_x)_{x \geq 0}$ of intensity $t\bar{\nu}(0)$ and a sequence $(X_i)_{i \in \mathbb{N}}$ of i.i.d. variables of law $\nu/\bar{\nu}(0)$ independent of $(N_x)_{x \geq 0}$ such that

$$\bar{\tau}_x = x + \sum_{i=1}^{N_x} X_i, \quad x \geq 0. \quad (4.39)$$

As $\mathcal{R}(\bar{\tau})$ is the range of $\bar{\tau}$, the number of data blocks of $\mathcal{C}(t)$ between $d(t)$ and $d(t) + \bar{\tau}_x$ is equal to the number of jumps of $\bar{\tau}$ before x , that is N_x . Thus,

$$\begin{aligned} & \frac{\text{number of data blocks in } [d(t), d(t) + \bar{\tau}_x]}{\bar{\tau}_x} \\ &= \frac{N_x}{\bar{\tau}_x} \xrightarrow{x \rightarrow \infty} \frac{\mathbb{E}(N_1)}{\mathbb{E}(\bar{\tau}_1)} = t\bar{\nu}(0)(1 - mt) \quad \text{a.s.} \end{aligned} \quad (4.40)$$

by the law of large numbers (see [20, page 92]). This completes the proof. \square

Appendix

Background on Lévy Processes

The results given in this section can be found in [20, Chapters VI and VII] (there, statements are made in terms of the dual process $-Y$). We recall that a Lévy process is càdlàg process starting from 0 which has i.i.d. increments. A subordinator is an increasing Lévy process.

We consider in this section a Lévy process $(X_x)_{x \geq 0}$ which has no negative jumps (spectrally positive Lévy process). We denote by Ψ its Laplace exponent which verifies for every $\rho \geq 0$,

$$\mathbb{E}(\exp(-\rho X_x)) = \exp(-x\Psi(\rho)). \quad (A.1)$$

We stress that this is not the classical choice for the sign of the Laplace exponent of Lévy processes with no negative jumps and a negative drift such as the process $(Y_x)_{x \geq 0}$ introduced

in the previous section. However it is the classical choice for subordinators, which we will need. It is then convenient to use this same definition for all Lévy processes which appear in this text.

First, we consider the case when $(X_x)_{x \geq 0}$ has bounded variations. That is,

$$X_x := dx + \sum_{x_i \leq x} l_i, \quad (\text{A.2})$$

where $\{(x_i, l_i) : i \in \mathbb{N}\}$ is a PPP on $[0, \infty[\times]0, \infty[$ with intensity measure $dx \otimes \nu$ such that $\int_0^\infty x \nu(dx) < \infty$. We call ν the Lévy measure and $d \in \mathbb{R}$ the drift. Note that $(X_x)_{x \geq 0}$ is a subordinator if and only if $d \geq 0$.

Writing $\bar{\nu}$ for the tail of the measure ν , the Lévy-Khintchine formula gives

$$\Psi(\rho) = d\rho + \int_0^\infty (1 - e^{-\rho x}) \nu(dx), \quad (\text{A.3})$$

$$\frac{\Psi(\rho)}{\rho} = d + \int_0^\infty e^{-\rho x} \bar{\nu}(x) dx, \quad (\text{A.4})$$

$$\Psi'(0) = d + \int_0^\infty x \nu(dx), \quad (\text{A.5})$$

$$\lim_{\rho \rightarrow \infty} \frac{\Psi(\rho)}{\rho} = d, \quad \lim_{\rho \rightarrow \infty} (\Psi(\rho) - d\rho) = \bar{\nu}(0). \quad (\text{A.6})$$

Second, we consider the case when Ψ has a right derivative at 0 with

$$\Psi'(0) < 0, \quad (\text{A.7})$$

meaning that $\mathbb{E}(X_1) < 0$. And we consider the infimum process which has continuous paths and the first passage time defined for $x \geq 0$ by

$$I_x = \inf\{X_y : 0 \leq y \leq x\}; \quad \tau_x = \inf\{z \geq 0 : X_z < -x\}. \quad (\text{A.8})$$

As $-\Psi$ is strictly convex and $-\Psi'(0) > 0$, $-\Psi$ is strictly increasing from $[0, \infty[$ to $[0, \infty[$ and so is strictly positive on $]0, \infty[$. We write $\kappa : [0, \infty[\rightarrow \mathbb{R}$ for the inverse function of $-\Psi$ and we have the following theorem (see [20, Theorem 1 page 189, Corollary 3 page 190]).

Theorem A.1. $(\tau_x)_{x \geq 0}$ is a subordinator with Laplace exponent κ .

Moreover the following identity holds between measures on $[0, \infty[\times]0, \infty[$:

$$x \mathbb{P}(\tau_l \in dx) dl = l \mathbb{P}(-X_x \in dl) dx. \quad (\text{A.9})$$

Note that if $(X_x)_{x \geq 0}$ has bounded variations, using (A.6), we can write

$$\forall \rho \geq 0, \quad \kappa(\rho) = -\frac{\rho}{d} + \int_0^\infty (1 - e^{-\rho z}) \Pi(dz), \quad (\text{A.10})$$

where Π is a measure on \mathbb{R}^+ verifying (use (A.6) and Wald's identity or (A.5))

$$\bar{\Pi}(0) = -\frac{\bar{v}(0)}{d}, \quad \int_0^\infty x\Pi(dx) = \frac{1}{d} - \frac{1}{d + \int_0^\infty x\nu(dx)}. \quad (\text{A.11})$$

Now we introduce the supremum process defined for $x \geq 0$ by

$$S_x := \sup\{X_y : 0 \leq y \leq x\}, \quad (\text{A.12})$$

and the a.s. unique instant at which X reaches this supremum on $[0, x]$:

$$\gamma_x := \inf\{y \in [0, x] : X_y = S_x\}. \quad (\text{A.13})$$

By duality, we have $(S_x, \gamma_x) \stackrel{d}{=} (X_x - I_x, x - g_x)$ where g_x denotes the a.s. unique instant at which $(X_{x^-})_{x \geq 0}$ reaches its overall infimum on $[0, x]$ (see [20, Proposition 3] or [19, page 25]). If T is an exponentially distributed random time with parameter $q > 0$ which is independent of X and $\lambda, \mu > 0$, then we have (use [20, Theorem 5 page 160, Theorem 4 page 191])

$$\begin{aligned} \mathbb{E}(\exp(-\mu S_T - \lambda \gamma_T)) &= \frac{q(\kappa(\lambda + q) - \mu)}{\kappa(q)(q + \lambda + \Psi(\mu))} \\ &= \exp\left(\int_0^\infty dx \int_0^\infty \mathbb{P}(\gamma_x \in dy) (e^{-\lambda x - \mu y} - 1)x^{-1}e^{-qx}\right), \end{aligned} \quad (\text{A.14})$$

which gives

$$\mathbb{E}(\exp(-\mu S_\infty - \lambda \gamma_\infty)) = \frac{1}{\kappa'(0)} \frac{\kappa(\lambda) - \mu}{\lambda + \Psi(\mu)} = -\Psi'(0) \frac{\kappa(\lambda) - \mu}{\lambda + \Psi(\mu)}, \quad (\text{A.15})$$

$$\mathbb{E}(\exp(-\mu S_\infty)) = \mu \frac{\Psi'(0)}{\Psi(\mu)}, \quad (\text{A.16})$$

$$\mathbb{E}(\exp(-\lambda \gamma_\infty)) = \exp\left(\int_0^\infty (e^{-\lambda x} - 1)x^{-1}\mathbb{P}(X_x > 0)dx\right). \quad (\text{A.17})$$

Acknowledgment

The author wishes to thank Jean Bertoin for introducing him in this topic and guiding him along the different steps of this work.

References

- [1] P. Chassaing and P. Flajolet, "Hachage, arbres, chemins," *Gazette des Mathématiciens*, no. 95, 2003.
- [2] P. Flajolet, P. Poblete, and A. Viola, "On the analysis of linear probing hashing," *Algorithmica*, vol. 22, no. 4, pp. 490–515, 1998.

- [3] D. Foata and J. Riordan, "Mappings of acyclic and parking functions," *Aequationes Mathematicae*, vol. 10, pp. 10–22, 1974.
- [4] P. Chassaing and G. Louchard, "Phase transition for parking blocks, Brownian excursion and coalescence," *Random Structures & Algorithms*, vol. 21, no. 1, pp. 76–119, 2002.
- [5] J. Bertoin and G. Miermont, "Asymptotics in Knuth's parking problem for caravans," *Random Structures & Algorithms*, vol. 29, no. 1, pp. 38–55, 2006.
- [6] V. Bansaye, "On a model for the storage of files on a hardware. II. Evolution of a typical data block," *Journal of Applied Probability*, vol. 44, no. 4, pp. 901–927, 2007.
- [7] O. J. Boxma and J. W. Cohen, "Heavy-traffic analysis for the GI/G/1 queue with heavy-tailed distributions," *Queueing Systems: Theory and Applications*, vol. 33, no. 1–3, pp. 177–204, 1999.
- [8] J. W. Cohen, *The Single Server Queue*, vol. 8 of *North-Holland Series in Applied Mathematics and Mechanics*, North-Holland, Amsterdam, The Netherlands, 2nd edition, 1982.
- [9] N. U. Prabhu, *Stochastic Storage Processes*, vol. 15 of *Applications of Mathematics*, Springer, New York, NY, USA, 2nd edition, 1998.
- [10] G. Matheron, *Random Sets and Integral Geometry*, John Wiley & Sons, New York, NY, USA, 1978.
- [11] P. J. Fitzsimmons, B. Fristedt, and B. Maisonneuve, "Intersections and limits of regenerative sets," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 70, no. 2, pp. 157–173, 1985.
- [12] V. Bansaye, "On a model for the storage of files on a hardware. I. Statistics at a fixed time and asymptotics," <http://arxiv.org/abs/math.PR/0611432>.
- [13] H. Kaspi and M. Rubinovitch, "Regenerative sets and their applications to Markov storage systems," in *Queueing Theory and Its Applications*, vol. 7 of *Cwi Monograph*, pp. 413–427, North-Holland, Amsterdam, The Netherlands, 1988.
- [14] J. F. C. Kingman, "The heavy traffic approximation in the theory of queues," in *Proceedings of the Symposium on Congestion Theory (Chapel Hill, 1964)*, pp. 137–169, The University of North Carolina, Chapel Hill, NC, USA, 1965.
- [15] M. I. Taksar, "Regenerative sets on real line," in *Séminaire de Probabilités, 14*, vol. 784 of *Lecture Notes in Mathematics*, pp. 437–474, Springer, Berlin, Germany, 1980.
- [16] M. I. Taksar, "Stationary Markov sets," in *Séminaire de Probabilités, 21*, vol. 1247 of *Lecture Notes in Mathematics*, pp. 303–340, Springer, Berlin, Germany, 1987.
- [17] B. Maisonneuve, "Ensembles régénératifs de la droite. [Regenerative sets of the line]," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 63, no. 4, pp. 501–510, 1983.
- [18] J. Pitman, "Stationary Excursions," in *Séminaire de Probabilités, 21*, vol. 1247 of *Lecture Notes in Mathematics*, pp. 289–302, Springer, Berlin, Germany, 1987.
- [19] J. Bertoin, *Subordinators, Lévy Processes with No Negative Jumps and Branching Processes*, 2000, Lecture notes for MaPhySto, August 2000, <http://citeseer.ist.psu.edu/old/422836.html>.
- [20] J. Bertoin, *Lévy Processes*, vol. 121 of *Cambridge Tracts in Mathematics*, Cambridge University Press, Cambridge, UK, 1996.
- [21] P. A. Meyer, "Processus de Poisson ponctuels, d'après K. Itô," in *Séminaire de Probabilités, 5*, vol. 191 of *Lecture Notes in Mathematics*, pp. 209–210, Springer, Berlin, Germany, 1971.
- [22] J. Jacod and A. N. Shiryaev, *Limit Theorems for Stochastic Processes*, vol. 288 of *Fundamental Principles of Mathematical Sciences*, Springer, Berlin, Germany, 2nd edition, 2003.
- [23] O. Kallenberg, *Foundations of Modern Probability*, Probability and Its Applications, Springer, New York, NY, USA, 2nd edition, 2002.