On a model for the storage of files on a hardware I : Statistics at a fixed time and asymptotic regimes.

Vincent Bansaye [*]

February 4, 2008

Abstract

We consider a generalized 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. For each fixed time, the model can be interpreted in terms of some M/G/1 queue. We specify the distribution of the space of unoccupied locations at a fixed time and give asymptotic regimes when the hardware is becoming full.

Key words. Parking problem. Data storage. Random covering. Lévy processes. Regenerative sets. M/G/1 queues.

A.M.S. Classification. 60D05, 60G51, 68B15.

1 Introduction

We consider a generalized 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. [10, 12, 13]) : he studied the distance between the spots where the files arrive and the spots where they are stored. Later Chassaing and Louchard [9] 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 [6] have extended these results to files of random sizes which arrive uniformly

---
[*]Laboratoire de Probabilités et Modèles Aléatoires. Université Pierre et Marie Curie et C.N.R.S. UMR 7599. 175, rue du Chevaleret, 75 013 Paris, France.
e-mail : bansaye@ccr.jussieu.fr

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 splitted 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 they arrive and identical distribution of the sizes and we model the arrival of 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 $\mathrm{m} := \int_0^\infty l\nu(dl)$ and assume $\mathrm{m} < \infty$. 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 2.1). The first questions which arise and are treated here concern statistics at a fixed time for the set of occupied locations. 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 $1/\mathrm{m}$. In Section 3.1, we give some geometric properties of the covering and characterize the distribution of the covering at a fixed time 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 [2], where we investigate the evolution in time of a typical data block. Moreover, using this characterization, we determine the asymptotic regimes at the saturation time, which depend on the tail of ν, as in [6, 8, 9] . More precisely, we give the asymptotic of $\mathcal{C}(t)$ when t tends to $1/\mathrm{m}$ (Theorem 2) and the asymptotic of $\mathcal{C}(t)$ restricted to $[0, x]$ when x tends to infinity and t tends to $1/\mathrm{m}$ (Theorem 3). We derive then the asymptotic of the largest block of the hardware restricted to $[0, x]$ when x tends to infinity and t tends to $1/\mathrm{m}$. As expected, we recover the phase transition observed by Chassaing and Louchard in [9]. 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. Thus, 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 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. Chap 3 in [11] or [24]). Thus, results of Section 3.1 and Section 3.3 for finite ν follow easily from known results on $M/G/1$. When ν is infinite, results are similar though busy cycle is not defined. Thus the approach is different and proving asymptotics on random sets requires results about Lévy processes and regenerative sets. 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 queuing model.

2 Preliminaries

Throughout this paper, we use the classical notation δ_x for the Dirac mass at x and $\mathbb{N} = \{1, 2, \ldots\}$. If $\mathcal{R}$ is a measurable subset of $\mathbb{R}$, we denote by $\mid \mathcal{R} \mid$ 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

$$g_x(\mathcal{R}) = \sup\{y \leq x : y \in \mathcal{R}\}, \qquad d_x(\mathcal{R}) = \inf\{y > x : y \in \mathcal{R}\}. \tag{1}$$

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

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

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

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

$$d_H(A, B) = \max\big(\sup_{x \in A} d(x, B), \sup_{x \in B} d(x, A)\big).$$

The topology induced by this distance is the topology of Matheron [20] : a sequence $\mathcal{R}_n$ in $\mathcal{H}(I)$ converges to $\mathcal{R}$ iff for each open set G and each compact K,

$$\mathcal{R} \cap G \neq \varnothing \quad \text{implies} \quad \mathcal{R}_n \cap G \neq \varnothing \quad \text{for } n \text{ large enough},$$
$$\mathcal{R} \cap K = \varnothing \quad \text{implies} \quad \mathcal{R}_n \cap K = \varnothing \quad \text{for } n \text{ large enough}.$$

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 [20] and [14] for details).

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

First, we present a deterministic construction of 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. We begin by constructing the covering $\mathcal{C}^{(n)}$ obtained by considering only the first n files, so that $\mathcal{C}$ is obtained as the increasing union of these coverings. A short thought (see Remark 1 p 5) enables us to see that the covering $\mathcal{C}$ does not depend on the order of arrival of the files. This construction of $\mathcal{C}$ will then be applied to the construction of our random covering at a fixed time $\mathcal{C}(t)$ by considering files arrived before time t. This construction and results for finite ν are classical in queuing theory (see e.g. [11]) and storage systems (see e.g. [24]). Thus we do not give details here and we refer to [1] for complete proof.

We define $\mathcal{C}^{(n)}$ by induction. We set $\mathcal{C}^{(0)} := \varnothing$, and introduce the complementary set $\mathcal{R}^{(n)}$ of $\mathcal{C}^{(n)}$ (i.e. the free space of the real line). Let $y_{n+1} = \inf\{y \geq 0, \mid \mathcal{R}^{(n)} \cap [x_{n+1}, y[\mid = l_{n+1}\}$, so y_{n+1} is the right-most point which is used for storing the $(n+1)$-th file. Define then

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

Now we consider the quantity of data over x, $R_x^{(n)}$, 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

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

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

$$Y_b^{(n)} - Y_a^{(n)} = -(b - a) \; + \sum_{\substack{i \leq n \\ x_i \in]a,b]}} l_i \qquad \text{for } a < b. \tag{2}$$

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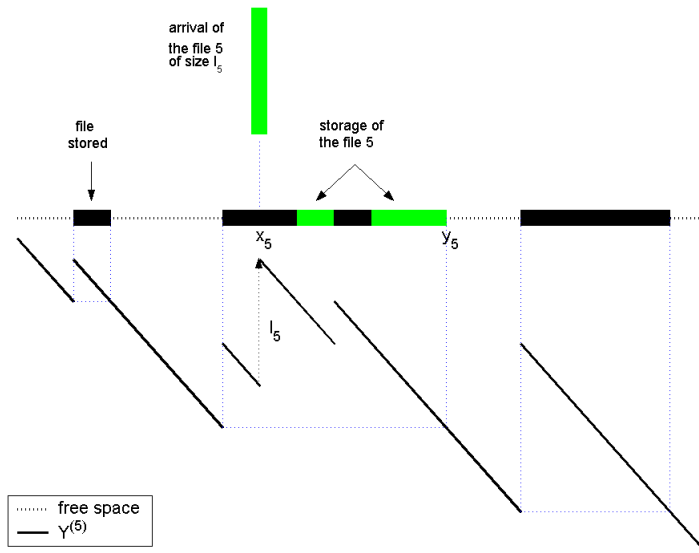
Introducing also its infimum function defined for $x \in \mathbb{R}$ by $\quad I_x^{(n)} := \inf\{Y_y^{(n)} : y \leq x\}$, we get the following expression.

Lemma 1. *For every $n \geq 1$, we have $R^{(n)} = Y^{(n)} - I^{(n)}$.*

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

$$\mathcal{C}^{(n)} = \{x \in \mathbb{R} : \; Y^{(n)} - I^{(n)} > 0\}. \tag{3}$$

We are now able to investigate the situation when n tends to infinity under the following mild condition

$$\forall \, L \geq 0, \quad \sum_{x_i \in [-L,L]} l_i < \infty, \tag{4}$$

4

which means that the quantity of data arriving on a compact set is finite. Introduce the function Y defined on $\mathbb{R}$ by $Y_0 = 0$ and

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

and its infimum I defined for $x \in \mathbb{R}$ by $\quad I_x := \inf\{Y_y : y \leq x\}$.

As expected, letting $n \to \infty$ in (2) and (3), the covering $\mathcal{C} := \cup_{n \in \mathbb{N}} \mathcal{C}^{(n)}$ is given by (see Section 2.1 in [1] for the proof) :

Proposition 1. - If $\lim_{x \to -\infty} Y_x = +\infty$, then $\mathcal{C} = \{x \in \mathbb{R} : Y_x - I_x > 0\} \neq \mathbb{R}$.

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

Remark 1. This result ensures that the covering at time t just depend on $(Y_x - I_x)_{x \in \mathbb{R}}$. Thus it does not depend on the order of arrival of files.

Finally, we can construct the random covering associated with a PPP. As the order of arrival of files has no importance, the random covering $\mathcal{C}(t)$ at time t described in Introduction is obtained by the deterministic construction above by taking the subfamily of files i which verifies $t_i \leq t$.

When files arrive according to a PPP, $(Y_x)_{x \geq 0}$ is a Lévy process, and we recall now some results about Lévy processes and their fluctuations which will be useful in the rest of this work.

2.2 Background on Lévy processes

The results given in this section can be found in the Chapters VI and VII in [5] (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 iid 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 $\rho \geq 0$:

$$\mathbb{E}(\exp(-\rho X_x)) = \exp(-x\Psi(\rho)). \tag{5}$$

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,

$$X_x := dx + \sum_{x_i \leq x} l_i,$$

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

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

$$\Psi(\rho) = d\rho + \int_0^\infty (1 - e^{-\rho x})\nu(\mathrm{d}x), \tag{6}$$

$$\frac{\Psi(\rho)}{\rho} = d + \int_0^\infty e^{-\rho x}\bar{\nu}(x)\mathrm{d}x, \tag{7}$$

$$\Psi'(0) = d + \int_0^\infty x\nu(\mathrm{d}x), \tag{8}$$

$$\lim_{\rho \to \infty} \frac{\Psi(\rho)}{\rho} = d \quad \text{and} \quad \lim_{\rho \to \infty} (\Psi(\rho) - d\rho) = \bar{\nu}(0). \tag{9}$$

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

$$\Psi'(0) < 0, \tag{10}$$

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

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

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 $\kappa : [0, \infty[\to \mathbb{R}$ for the inverse function of $-\Psi$ and we have (see [5] Theorem 1 on page 189 and Corollary 3 on page 190) :

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

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

$$x\mathbb{P}(\tau_l \in \mathrm{d}x)\mathrm{d}l = l\mathbb{P}(-X_x \in \mathrm{d}l)\mathrm{d}x. \tag{11}$$

Note that if $(X_x)_{x \geq 0}$ has bounded variations, using (9), we can write

$$\forall \rho \geq 0, \quad \kappa(\rho) = -\frac{\rho}{d} + \int_0^\infty (1 - e^{-\rho z})\Pi(\mathrm{d}z), \tag{12}$$

where Π is a measure on $\mathbb{R}^+$ verifying (use (9) and Wald's identity or (8)) :

$$\bar{\Pi}(0) = -\frac{\bar{\nu}(0)}{d}, \qquad \int_0^\infty x\Pi(\mathrm{d}x) = \frac{1}{d} - \frac{1}{d + \int_0^\infty x\nu(\mathrm{d}x)}. \tag{13}$$

6

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

$$S_x := \sup\{X_y : 0 \leq y \leq x\},$$

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

$$\gamma_x := \inf\{y \in [0, x] : X_y = S_x\}.$$

By duality, we have $(S_x, \gamma_x) \overset{d}{=} (X_x - I_x, x - g_x)$ where g_x denotes the a.s unique instant at which $(X_{x-})_{x \geq 0}$ reaches its overall infimum on $[0, x]$ (see Proposition 3 in [5] or [4] on 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 [5] Theorem 5 on page 160 and Theorem 4 on page 191) :

$$
\begin{aligned}
\mathbb{E}\big(\exp(-\mu S_T - \lambda \gamma_T)\big) &= \frac{q(\kappa(\lambda + q) - \mu)}{\kappa(q)(q + \lambda + \Psi(\mu))} \\
&= \exp\Big(\int_0^\infty \mathrm{d}x \int_0^\infty \mathbb{P}(Y_x \in \mathrm{d}y)(e^{-\lambda x - \mu y} - 1)x^{-1}e^{-qx} \Big),
\end{aligned}
$$

which gives

$$
\begin{aligned}
\mathbb{E}\big(\exp(-\mu S_\infty - \lambda \gamma_\infty)\big) &= \frac{1}{\kappa'(0)} \frac{\kappa(\lambda) - \mu}{\lambda + \Psi(\mu)} = -\Psi'(0)\frac{\kappa(\lambda) - \mu}{\lambda + \Psi(\mu)}, & (14) \\
\mathbb{E}\big(\exp(-\mu S_\infty)\big) &= \mu \frac{\Psi'(0)}{\Psi(\mu)}, & (15) \\
\mathbb{E}\big(\exp(-\lambda \gamma_\infty)\big) &= \exp\Big(\int_0^\infty (e^{-\lambda x} - 1)x^{-1}\mathbb{P}(X_x > 0)\mathrm{d}x \Big). & (16)
\end{aligned}
$$

2.3 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 [6], we say that $\nu \in \mathcal{D}_{2+}$ if ν has a finite second moment $\mathrm{m}_2 := \int_0^\infty l^2\nu(\mathrm{d}l)$. For $\alpha \in]1, 2]$, we say that $\nu \in \mathcal{D}_\alpha$ whenever

$$\exists C > 0 \text{ such that } \bar{\nu}(x) \overset{x \to \infty}{\sim} Cx^{-\alpha}.$$

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

$$C_\alpha := \left(\frac{C\Gamma(2 - \alpha)}{\mathrm{m}_2(\alpha - 1)} \right)^{1/\alpha}.$$

We denote by $(B_z)_{z \in \mathbb{R}}$ a two-sided Brownian motion, i.e. $(B_x)_{x \geq 0}$ 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 :

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

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

$$Y_z^{2+,\lambda} = -\lambda z + \sqrt{\mathrm{m}_2/\mathrm{m}} B_z, \qquad Y_z^{2,\lambda} = -\lambda z + \sqrt{C/\mathrm{m}} B_z, \qquad Y_z^{\alpha,\lambda} = -\lambda z + C_\alpha \sigma_z^{(\alpha)},$$

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

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 the characterization of Section 2.1 and results of Section 2.2. This characterization will be useful to prove asymptotics results (Theorem 2, Theorem 3 and Corollary 3) and for the dynamic results given in [2]. In that view, 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

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

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

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

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 $1/\mathrm{m}$ (see Section 4 for the proof).

Proposition 2. *For every $t < 1/\mathrm{m}$, we have $\mathcal{C}(t) = \{x \in \mathbb{R} : Y_x^{(t)} > I_x^{(t)}\} \neq \mathbb{R}$ a.s. For every $t \geq 1/\mathrm{m}$, we have $\mathcal{C}(t) = \mathbb{R}$ a.s.*

Indeed, in queueing system, $t\mathrm{m}$ is the charge and $\mathcal{C}(t) \neq \mathbb{R} \Leftrightarrow t\mathrm{m} < 1$ is the standard claim of stability for $t\mathrm{m} < 1$. The complete argument is defered to Section 4.

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

$$\mathcal{R}(t) = \{x \in \mathbb{R} : Y_x^{(t)} = I_x^{(t)}\}. \tag{17}$$

We begin by giving some geometric properties of this set. These properties are classical (see [17] for storage systems and [18] for queueing theory).

Proposition 3. *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 $(\mathcal{R} - d_0(\mathcal{R}(t))) \cap [0, \infty[$. Moreover for every $x \in \mathbb{R}$, $\mathbb{P}(x \in \mathcal{C}(t)) = \min(1, \mathrm{m}t)$.*

Stationarity is plain from the construction of the covering and regeneration property is a direct consequence of Lemma 2. Symmetry is then a consequence of Lemma 6.5 in [25] or Corollary (7.19) in [26]. Computation of $\mathbb{P}(x \in \mathcal{C}(t))$ is then derived from Theorem 1 in [25]. See Section 3.1 in [1] for the complete proof.

Even though for each fixed t the distribution of $\mathcal{R}(t)^{cl}$ is symmetric, the processes $(\mathcal{R}(t)^{cl} : t \in [0, 1/\mathrm{m}])$ and $(-\mathcal{R}(t)^{cl} : t \in [0, 1/\mathrm{m}])$ are quite different. For example, we shall observe in [2] 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 (6))

$$\Psi^{(t)}(\rho) := -\rho + \int_0^\infty \left(1 - e^{-\rho x}\right) t\nu(\mathrm{d}x). \tag{18}$$

For sake of simplicity, we write, recalling (1),

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

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 $\mathbf{B}_0(t) = \varnothing$.
We work with $\mathcal{R}$ subset of $\mathbb{R}$ of the form $\sqcup_{n \in \mathbb{N}}[a_n, b_n[$ and we denote by $\widetilde{\mathcal{R}} := \sqcup_{n \in \mathbb{N}}[-b_n, -a_n[$ 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

$$\overrightarrow{\mathcal{R}} := (\mathcal{R} - d_0(\mathcal{R})) \cap [0, \infty] = \bigsqcup_{n \in \mathbb{N}:\ a_n \geq d_0(\mathcal{R})} [a_n - d_0(\mathcal{R}), b_n - d_0(\mathcal{R})[,$$

$$\overleftarrow{\mathcal{R}} := \widetilde{\overrightarrow{\mathcal{R}}} = \bigsqcup_{n \in \mathbb{N}:\ b_n \leq g_0(\mathcal{R})} [g_0(\mathcal{R}) - b_n, g_0(\mathcal{R}) - a_n[.$$

Example 1. For a given $\mathcal{R}$ represented by the dotted lines, we give below $\overrightarrow{\mathcal{R}}$ 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.
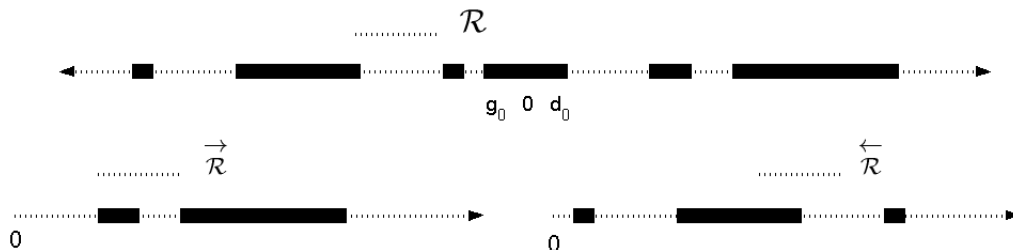
Thus $\overrightarrow{\mathcal{R}(t)}$ (resp. $\overleftarrow{\mathcal{R}(t)}$) is the free space at the right of $\mathbf{B}_0(t)$ (resp. at the left of $\mathbf{B}_0(t)$, turned over, closed at the left and open at the right). We have then the identity

$$\mathcal{R}(t) = (d(t) + \overrightarrow{\mathcal{R}(t)}) \sqcup (-g(t) + \widetilde{\overleftarrow{\mathcal{R}(t)}}). \tag{19}$$

9

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

$$\overrightarrow{\tau}_x^{(t)} := \inf\{y \geq 0 : |\overrightarrow{\mathcal{R}(t)} \cap [0, y]| > x\}, \qquad \overleftarrow{\tau}_x^{(t)} := \inf\{y \geq 0 : |\overleftarrow{\mathcal{R}(t)} \cap [0, y]| > x\},$$

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

Proposition 4. *(i) The random sets $\overrightarrow{\mathcal{R}(t)}$ and $\overleftarrow{\mathcal{R}(t)}$ are independent, identically distributed and independent of $(g(t), d(t))$.*
(ii) $\overrightarrow{\mathcal{R}(t)}$ and $\overleftarrow{\mathcal{R}(t)}$ are the range of the subordinators $\overrightarrow{\tau}^{(t)}$ 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 :

$$(g(t), d(t)) = (-Ul(t), (1 - U)l(t)),$$

$$\mathbb{P}(l(t) \in \mathrm{d}x) = (1 - \mathrm{m}t)\big(\delta_0(\mathrm{d}x) + \mathbb{1}_{\{x>0\}}x\Pi^{(t)}(\mathrm{d}x)\big),$$

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 2. Such results are classical for regenerative sets (see e.g. [17, 19, 23]). But we need this particular characterization and expressions given in the proof below for forthcoming results.

We can then estimate the number of data blocks on the hardware. If ν has a finite mass, we write $N_x^{(t)}$ 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 $1/(2\mathrm{m})$. And the number of blocks of the hardware reaches a.s. its maximal at time $1/(2\mathrm{m})$. More precisely,

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

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

Moreover, 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 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 $d(t) + \overrightarrow{\tau}_l^{(t)}$). This gives the distribution of the extremities of the portion of the hardware used for the storage of a file.

3.2 Observations and examples

First, we have for every $\rho \geq 0$ (use (12)),

$$\kappa^{(t)}(\rho) = \rho + \int_0^\infty (1 - e^{-\rho x}) \Pi^{(t)}(\mathrm{d}x), \tag{20}$$

and using (13)

$$\bar{\Pi}^{(t)}(0) = t\bar{\nu}(0), \qquad \int_0^\infty x\Pi^{(t)}(\mathrm{d}x) = \frac{\mathrm{m}t}{1 - \mathrm{m}t}. \tag{21}$$

Using (11), we have also the following identity of measures on $[0, \infty[\times[0, \infty[$

$$x\mathbb{P}(\overrightarrow{\tau}_l^{(t)} \in \mathrm{d}x)\mathrm{d}l = l\mathbb{P}(-Y_x^{(t)} \in \mathrm{d}l)\mathrm{d}x. \tag{22}$$

Finally, we give the distribution of the extremities of $\mathbf{B}_0$:

$$\mathbb{P}(-g(t) \in \mathrm{d}x) = \mathbb{P}(d(t) \in \mathrm{d}x) = (1 - \mathrm{m}t)\big(\delta_0(\mathrm{d}x) + 1\!\!1_{\{x>0\}}\bar{\Pi}^{(t)}(x)\mathrm{d}x\big). \tag{23}$$

Let us consider three explicit examples

Example 2. (1) The basic example is when $\nu = \delta_1$ (all files have the same unit size as in the original parking problem in [9]). Then for all $x \in \mathbb{R}_+$ and $n \in \mathbb{N}$,

$$\mathbb{P}(Y_x^{(t)} + x = n) = e^{-tx}\frac{(tx)^n}{n!},$$

$$\mathbb{P}(\overrightarrow{\tau}_x^{(t)} = x + n) = \frac{x}{x+n}e^{-t(x+n)}\frac{(t(n+x))^n}{n!}, \tag{24}$$

where the second identity follows from integrating (22) on $\{(x, l) : l \in [z, z+h], x - z = n\}$ and letting h tend to 0. Then,

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

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

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

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

$$\mathbb{P}(Y_x^{(t)} \in \mathrm{d}z) = 1\!\!1_{[-x,\infty[}(z)\Gamma(tx)^{-1}e^{-(z+x)}(z+x)^{tx-1}\mathrm{d}z,$$

$$\mathbb{P}(\overrightarrow{\tau}_x^{(t)} \in \mathrm{d}z) = 1\!\!1_{[x,\infty[}(z)x(z\Gamma(tz))^{-1}e^{-(z-x)}(z-x)^{tz-1}\mathrm{d}z. \tag{25}$$

Further

$$\Pi^{(t)}(\mathrm{d}z) = (z\Gamma(tz))^{-1}e^{-z}z^{tz-1}\mathrm{d}z,$$

and

$$\mathbb{P}(l(t) \in \mathrm{d}x) = (1 - t)\big(\delta_0(\mathrm{d}x) + \Gamma(tz)^{-1}e^{-x}x^{tx-1}\mathrm{d}x\big).$$

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

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

Finally, we specify two distributions involved in the storage of the data.

Writing $-g(t) = \gamma(t)$ (see (28) and (29)) and using the identity of fluctuation (16) gives an other expression for the Laplace transform of $g(t)$: For all $t \in [0, 1/\mathrm{m}[$ and $\lambda \geq 0$, we have

$$\mathbb{E}\big(\exp\big(\lambda g(t)\big)\big) = \exp\left(\int_0^\infty (e^{-\lambda x} - 1) x^{-1} \mathbb{P}(Y_x^{(t)} > 0) \mathrm{d}x \right). \tag{26}$$

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 $(g(t))_{t \in [0,1/\mathrm{m}[}$ (see Theorem 2 in Section 4 in [2]).

The quantity of data over 0, $R_0^{(t)}$ (see Section 2.1), is an increasing process equal to $(-I_0^{(t)})_{t \geq 0}$. By (28) and (15), its Laplace transform is then equal to

$$\lambda \longrightarrow \frac{(1 - \mathrm{m}t)\lambda}{\Psi^{(t)}(\lambda)}.$$

3.3   Asymptotics at saturation of the hardware

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

$$\mathbb{E}\big(l(t)\big) = \frac{\int_0^\infty l^2 \nu(\mathrm{d}l)}{(1 - \mathrm{m}t)^2}.$$

Thus we may expect that if ν has a finite second moment, then $(1 - \mathrm{m}t)^2 l(t)$ should converge in distribution as t tends to $1/\mathrm{m}$. Indeed, in the particular case $\nu = \delta_1$ or in the conditions of Corollary 2.4 in [4], we have an expression of $\Pi^{(t)}(\mathrm{d}x)$ and we can prove that $(1 - \mathrm{m}t)^2 l(t)$ does converge in distribution to a gamma variable.

More generally, we shall prove that the rescaled free space $(1 - \mathrm{m}t)^2 \mathcal{R}(t)$ converges in distribution as t tends to $1/\mathrm{m}$. In that view, we need to prove that the process $(Y_{(1-\mathrm{m}t)^{-2}x}^{(t)})_{x \in \mathbb{R}}$ converges after suitable rescaling to a random process. Thanks to (17), $(1 - \mathrm{m}t)^2 \mathcal{R}(t)$ 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 proofs of the next section and they are made simultaneously in the last section of this paper.

In queuing systems, asymptotics at saturation are known as heavy traffic approximation ($\rho = t\mathrm{m} \to 1$), 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 III.7.2 in [11] or [18] if ν has a second moment order and [8] 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 the whole random set of occupied locations. Moreover, as explained below, asymptotics of $\mathcal{R}(t)$ can not be directly derived from asymptotics of Y or the workload R.

We introduce now the following functions defined for every $t \in [0, 1/\mathrm{m}[$ and $\alpha \in]1, 2[$ by

$$\epsilon_{2+}(t) = (1 - \mathrm{m}t)^2, \qquad \epsilon_2(t) = 2\frac{(1 - \mathrm{m}t)^2}{-\log((1 - \mathrm{m}t))}, \qquad \epsilon_\alpha(t) = (1 - \mathrm{m}t)^{\frac{\alpha}{\alpha-1}}.$$

Recalling Notations of Section 2.3, we have then the following weak convergence result for the Hausdorff metric defined in Section 2 (see Section 4 for the proof).

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

First we prove the convergence of the Laplace exponent $\Psi^{(t)}$ after suitable rescaling as t tends to $1/\mathrm{m}$, which ensures the convergence of the Lévy process $Y^{(t)}$ after suitable rescaling (see Lemma 3). These convergences will not a priori entail the convergence of the random set $\epsilon_\alpha(t).\mathcal{R}^{cl}(t)$ since they do not entail the convergence of excursions. Nevertheless, they will entail the convergence of $\kappa^{(t)}$ since $\kappa^{(t)} \circ (-\Psi^{(t)}) = \mathrm{Id}$ (Lemma 4). Then we get the convergence of $\tau^{(t)}$ as t tends to infinity and thus of its range $\epsilon_\alpha(t).\mathcal{R}^{cl}(t)$.

Remark 3. More generally, as in queuing theory and [6], 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 $\mathcal{H}(\mathbb{R})$:

$$x^{-1}\mathcal{R}((1 - x\bar{\nu}(x))/\mathrm{m})^{cl} \overset{x \to \infty}{\Longrightarrow} \{x \in \mathbb{R} : Y_x^{\alpha,1} = I_x^{\alpha,1}\}^{cl} \quad \text{with } C = 1.$$

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

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

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 2. *If $\nu \in \mathcal{D}_\alpha$ ($\alpha \in [1,2] \cup \{2+\}$), then $\epsilon_\alpha(t).(g(t), d(t))$ converges weakly as t tends to $1/\mathrm{m}$ 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 $\nu \in \mathcal{D}_{2+}$ (resp. $\mathcal{D}_2$), $\epsilon_\alpha(t).l(t)$ converges weakly to a gamma variable with parameter $(1/2, \mathrm{m}/(4\mathrm{m}_2))$ (resp. $(1/2, \mathrm{m}/4)$).

Remark 4. The density of data blocks of size $\mathrm{d}x$ in $\epsilon_\alpha(t).\mathcal{R}(t)^{cl}$ is equal to $\frac{\mathrm{m}t}{1-\mathrm{m}t}\Pi^{(t)}(\mathrm{d}x)$. By the previous theorem or corollary, this density converges weakly as t tends to $1/\mathrm{m}$ to the density of data block of size $\mathrm{d}x$ of the limit covering $\{x \in \mathbb{R} : Y_x^{\alpha,1} = I_x^{\alpha,1}\}^{cl}$. This limit density, denoted by $\Pi^{\alpha,1}(\mathrm{d}x)$, can be computed explicitly in the cases $\nu \in \mathcal{D}_\alpha$ ($\alpha \in \{2, 2+\}$), thanks to the last corollary :

$$\Pi^{2+,1}(\mathrm{d}x) = \sqrt{\frac{\mathrm{m}}{4\pi\mathrm{m}_2 x^3}}\exp\big(-\frac{\mathrm{m}}{4\mathrm{m}_2}x\big), \qquad \Pi^{2,1}(\mathrm{d}x) = \sqrt{\frac{\mathrm{m}}{4\pi x^3}}\exp\big(-\frac{\mathrm{m}}{4}x\big).$$

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

3.4 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 at 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

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

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

Theorem 3. *If $\nu \in \mathcal{D}_\alpha$ ($\alpha \in [1, 2] \cup \{2+\}$), x tends to infinity and t to $1/\mathrm{m}$ such that $1 - \mathrm{m}t \sim \lambda f_\alpha(x)$ with $\lambda > 0$, then $x^{-1}(\mathcal{R}(t)^{cl} \cap [0, x])$ converges weakly in $\mathcal{H}([0, 1])$ to $\{x \in [0, 1] : Y_x^{\alpha,\lambda} = I_x^{\alpha,\lambda}\}^{cl}$.*

Thus as in [9], we observe a phase transition of the size of largest block of data in $[0, x]$ as $x \to \infty$ 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 :

Corollary 3. *Let $\nu \in \mathcal{D}_\alpha$ ($\alpha \in [1, 2] \cup \{2+\}$), x tend to infinity and t to $1/\mathrm{m}$:*
- If $1 - \mathrm{m}t \sim \lambda f_\alpha(x)$ with $\lambda > 0$, then $B_1(x, t)/x$ converges in distribution to the largest length of excursion of $(Y_x^{\alpha,\lambda} - I_x^{\alpha,\lambda})_{x \in [0,1]}$.
- If $f_\alpha(x) = o(1 - \mathrm{m}t)$, then $B_1(x, t)/x \xrightarrow{\mathbb{P}} 0$.
- If $1 - \mathrm{m}t = o(f_\alpha(x))$, then $B_1(x, t)/x \xrightarrow{\mathbb{P}} 1$.

The phase transition occurs at time t such that $1 - \mathrm{m}t \sim \lambda f_\alpha(x)$ 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 [9, 6], 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 [9, 6]. 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 :

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

4 Proofs

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

Proof of Proposition 2. First $m < \infty$ entails that $\forall L \geq 0$, $\sum_{t_i \leq t, x_i \in [-L,L]} l_i < \infty$ and condition (10) is satisfied a.s.

- 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 [5] Corollary 2 on page 190) :

$$Y_x^{(t)} \overset{x \to -\infty}{\Longrightarrow} \infty \quad \text{a.s.}$$

Then Proposition 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.

- If $t \geq 1/m$, then $\mathbb{E}(Y_{-1}^{(t)}) \leq 0$ ensures (see [5] Corollary 2 on page 190) :

$$Y_x^{(t)} \overset{x \to -\infty}{\Longrightarrow} -\infty \text{ a.s} \quad \text{or} \quad (Y_x^{(t)})_{x \leq 0} \text{ oscillates a.s in} -\infty.$$

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

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 $\overrightarrow{\tau}$, 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 :

$$\mathcal{P}_x := \{(t_i, x_i, l_i) : t_i \leq t, \ x_i \leq x\}, \quad \mathcal{P}^x := \{(t_i, x_i - x, l_i) : t_i \leq t, \ x_i > x\}.$$

Lemma 2. *For all $x \geq 0$, $\mathcal{P}^{d_x(\mathcal{R}(t))}$ is independent of $\mathcal{P}_{d_x(\mathcal{R}(t))}$ and distributed as $\mathcal{P}^0$.*

Proof. The simple Markov property for PPP states that for every $x \in \mathbb{R}$, $\mathcal{P}^x$ is independent of $\mathcal{P}_x$ and distributed as $\mathcal{P}^0$. Clearly this extends to simple stopping times in the filtration $\sigma(\mathcal{P}_x)_{x \in \mathbb{R}}$ 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, $\mathcal{P}^{d_x(\mathcal{R}(t))}$ is independent of $\mathcal{P}_{d_x(\mathcal{R}(t))}$ and distributed as $\mathcal{P}^0$. $\qquad \square$

Proof of Proposition 4. (i) By symmetry of $\mathcal{R}(t)^{cl}$, $\overrightarrow{\mathcal{R}(t)}$ and $\overleftarrow{\mathcal{R}(t)}$ are identically distributed. The regeneration property ensures that $\overrightarrow{\mathcal{R}(t)}$ is independent of $(\overleftarrow{\mathcal{R}(t)}, g(t), d(t))$. By symmetry, $\overleftarrow{\mathcal{R}(t)}$ is independent of $(g(t), d(t), \overrightarrow{\mathcal{R}(t)})$. So $\overrightarrow{\mathcal{R}(t)}$, $\overleftarrow{\mathcal{R}(t)}$ and $(g(t), d(t))$ are independent.

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

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

So the range of $\overrightarrow{\tau}$ is equal to $\overrightarrow{\mathcal{R}(t)}$. The fact $\overrightarrow{\tau}$ is a subordinator will be proved below but could be also derived directly from the regeneration property of $\overrightarrow{\mathcal{R}}(t)$ (see [19]). 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 $d(t) + y \in \mathcal{R}(t)$,

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

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

$$\begin{aligned}
\overrightarrow{\tau}_x &= \inf\{y \geq 0 : y \in \overrightarrow{\mathcal{R}(t)}, \ |\overrightarrow{\mathcal{R}(t)} \cap [0, y]| > x\} \\
&= \inf\{y \geq 0 : d(t) + y \in \mathcal{R}(t), \ |\mathcal{R}(t) \cap [d(t), d(t) + y]| > x\} \\
&= \inf\{y \geq 0 : Y_{y+d(t)} - Y_{d(t)} < -x\}.
\end{aligned} \tag{27}$$

Moreover

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

and Lemma 2 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 (10) (use (8) and $-1 + mt < 0$). Then Theorem 1 entails that $\overrightarrow{\tau}$ is a subordinator whose Laplace exponent is the inverse function of $-\Psi^{(t)}$.

As $\overleftarrow{\mathcal{R}(t)}$ is distributed as $\overrightarrow{\mathcal{R}(t)}$, $\overleftarrow{\tau}$ is distributed as $\overrightarrow{\tau}$ 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 $(\widetilde{Y}_x)_{x \geq 0}$ for the càdlàg version of $(-Y_{-x})_{x \geq 0}$ and

$$S(t) := \sup\{\widetilde{Y}_x, x \geq 0\} = -I_0, \qquad \gamma(t) := \arg(S(t)) = \inf\{x \geq 0 : \widetilde{Y}_x = S(t)\}. \tag{28}$$

Using (17) and the fact that Y has no negative jumps, we have

$$\begin{aligned}
g(t) &= g_0(\mathcal{R}(t)) = \sup\{x \leq 0 : Y_x = I_x\} \\
&= \sup\{x \leq 0 : Y_{x-} = I_0\} = -\inf\{x \geq 0 : \widetilde{Y}_x = -I_0\} \\
&= -\gamma(t).
\end{aligned} \tag{29}$$

Using again (17) and the fact that $(Y_x)_{x \geq 0}$ is regular for $]-\infty, 0[$ (see [5] Proposition 8 on page 84), we have also a.s.

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

16

where $(T_x)_{x \geq 0}$ is distributed as $(\overrightarrow{\tau}_x)_{x \geq 0}$ by (27) 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 $\lambda \neq \mu$:

$$
\begin{aligned}
\mathbb{E}\big(\exp(\lambda g(t) - \mu d(t)))\big) &= \mathbb{E}\big(\exp(-\lambda \gamma(t)) \mathbb{E}(\exp(-\mu T_{S(t)}))\big) \\
&= \mathbb{E}\big(\exp(-\lambda \gamma(t) - \kappa^{(t)}(\mu) S(t))\big) \\
&= -[\Psi^{(t)}]'(0) \frac{\kappa^{(t)}(\lambda) - \kappa^{(t)}(\mu)}{\lambda - \mu} \quad \text{using (14)} \quad (30) \\
&= (1 - \mathrm{m}t) \frac{\kappa^{(t)}(\lambda) - \kappa^{(t)}(\mu)}{\lambda - \mu} \quad \text{using (8),} \quad (31)
\end{aligned}
$$

which gives the distributions of $d(t)$, $g(t)$ and $l(t)$ letting respectively $\lambda = 0$, $\mu = 0$ and $\lambda \to \mu$. 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 (31). So $(g(t), d(t)) = (-U'l(t), (1 - U')l(t))$, where U' is a uniform random variable on $[0, 1]$ independent of $l(t)$. $\qquad \square$

Proof of Corollary 1. As $\bar{\nu}(0) < \infty$, then $\bar{\Pi}(0) = t\bar{\nu}(0) < \infty$ (see (21)). So $\overrightarrow{\tau}$ 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 iid variables of law $\nu/\bar{\nu}(0)$ independent of $(N_x)_{x \geq 0}$ such that

$$
\overrightarrow{\tau}_x = x + \sum_{i=1}^{N_x} X_i, \qquad x \geq 0.
$$

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

$$
\frac{\text{number of data blocks in } [d(t), d(t) + \overrightarrow{\tau}_x]}{\overrightarrow{\tau}_x} = \frac{N_x}{\overrightarrow{\tau}_x} \xrightarrow{x \to \infty} \frac{\mathbb{E}(N_1)}{\mathbb{E}(\overrightarrow{\tau}_1)} = t\bar{\nu}(0)(1 - \mathrm{m}t) \quad \text{a.s.}
$$

by the law of large numbers (see [5] on page 92). This completes the proof. $\qquad \square$

Proofs of Theorem 2 and Theorem 3 are close and made simultaneously. For that purpose, we introduce now $\Psi^{\alpha, \lambda}$ the Laplace exponent (see (5)) of $Y^{\alpha, \lambda}$ given for $y \geq 0$, $\lambda \geq 0$ and $\alpha \in]1, 2[$ by

$$
\Psi^{2+, \lambda}(y) = -\lambda y - \frac{\mathrm{m}_2}{\mathrm{m}} \frac{y^2}{2}, \qquad \Psi^{2, \lambda}(y) = -\lambda y - \frac{C}{\mathrm{m}} \frac{y^2}{2}, \qquad \Psi^{\alpha, \lambda}(y) = -\lambda y - (C_\alpha y)^\alpha.
$$

We denote by $\mathbb{D}$ the space of càdlàg function from $\mathbb{R}_+$ to $\mathbb{R}$ which we endow with the Skorokhod topology (see [15] on page 292). First, we prove the weak convergence of $Y^{(t)}$ after suitable rescaling.

Lemma 3. *If $\nu \in \mathcal{D}_\alpha$ ($\alpha \in [1,2] \cup \{2+\}$), then for all $y \geq 0$ and $\lambda > 0$:*

$$\epsilon_\alpha(t)^{-1}\Psi^{(t)}(\epsilon_\alpha(t)(1-\mathrm{m}t)^{-1}y) \overset{t \to 1/\mathrm{m}}{\longrightarrow} \Psi^{\alpha,1}(y),$$

$$x\Psi^{((1-\lambda f_\alpha(x))/\mathrm{m})}((xf_\alpha(x))^{-1}y) \overset{x \to \infty}{\longrightarrow} \Psi^{\alpha,\lambda}(y),$$

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

$$\left(\epsilon_\alpha(t)(1-\mathrm{m}t)^{-1}Y^{(t)}_{\epsilon_\alpha(t)^{-1}y}\right)_{y \geq 0} \overset{t \to 1/\mathrm{m}}{\Longrightarrow} (Y^{\alpha,1}_y)_{y \geq 0},$$

$$\left((xf_\alpha(x))^{-1}Y^{((1-\lambda f_\alpha(x))/\mathrm{m})}_{xy}\right)_{y \geq 0} \overset{x \to \infty}{\Longrightarrow} (Y^{\alpha,\lambda}_y)_{y \geq 0}.$$

Remark 5. If $\bar\nu$ is regularly varying at infinity with index $-\alpha \in]-1,-2[$, then $\bar\nu(x)^{-1}\Psi^{((1-\lambda x\bar\nu(x))/\mathrm{m})}(x^{-1}y)$ converges to $\Psi^{\alpha,\lambda}(y)$ as x tends to infinity.

Proof of Lemma 3. Using (7), we have

$$x\Psi^{(t)}(y) = xy(\mathrm{m}t - 1 - t\int_0^\infty (1 - e^{-yu})\bar\nu(u)\mathrm{d}u). \qquad (32)$$

We handle now the different cases :

• Case $\nu \in \mathcal{D}_{2+}$. Using $|1 - e^{-yu}|/y \leq u$ ($u \geq 0$) and dominated convergence theorem gives :

$$\int_0^\infty (1 - e^{-yu})\bar\nu(u)\mathrm{d}u \overset{y \to 0}{\sim} y\int_0^\infty u\bar\nu(u)\mathrm{d}u = \frac{y\mathrm{m}_2}{2},$$

which proves the first part of the lemma using (32).

• Case $\nu \in \mathcal{D}_\alpha$ with $\alpha \in]1,2[$. Using that $(u/y)^\alpha \bar\nu(u/y)$ is bounded, we apply dominated convergence theorem and get

$$\int_0^\infty (1 - e^{-yu})\bar\nu(u)\mathrm{d}u = y^{-1}\int_0^\infty (1 - e^{-u})\bar\nu(u/y)\mathrm{d}u \qquad (33)$$

$$\overset{y \to 0}{\sim} y^{-1}C(y^{-1})^{-\alpha}\int_0^\infty (1 - e^{-u})u^{-\alpha}\mathrm{d}u$$

$$\overset{y \to 0}{\sim} C\frac{\Gamma(2-\alpha)}{\alpha-1}y^{\alpha-1},$$

which proves the first part of the lemma using (32).

• Case ν is regularly varying at infinity with index $-\alpha \in]-1,-2[$. First,

$$\int_0^{1/\sqrt{y}} (1 - e^{-yu})\bar\nu(u)\mathrm{d}u \leq y\int_0^{1/\sqrt{y}} u\bar\nu(u)\mathrm{d}u \overset{y \to 0}{\sim} y(1/\sqrt{y})^{2-\alpha} = y^{\alpha/2}.$$

Moreover for every $u > 0$, $\bar\nu(u/y) \overset{y \to 0}{\sim} \bar\nu(y)u^{-\alpha}$. Let $\delta > 0$ such that $-2 < -\alpha - \delta < -\alpha + \delta < -1$. By Potter's theorem (page 25 in [7]) ensures that for all y small enough and u large enough,

$$\frac{\bar\nu(u/y)}{\bar\nu(1/y)} \leq 2\max(u^{-\alpha+\delta}, u^{-\alpha-\delta}).$$

18

So we can apply the dominated convergence theorem to get

$$\int_0^{1/\sqrt{y}}(1-e^{-yu})\bar\nu(u)\mathrm{d}u = y^{-1}\int_{\sqrt{y}}^{\infty}(1-e^{-u})\bar\nu(u/y)\mathrm{d}u \overset{y\to 0}{\sim} \frac{\Gamma(2-\alpha)}{\alpha-1}y^{-1}\bar\nu(1/y).$$

As $y^{\alpha/2} = o(y^{-1}\bar\nu(1/y))$ $(y\to 0)$, we can complete the proof with

$$\int_0^{\infty}(1-e^{-yu})\bar\nu(u)\mathrm{d}u \overset{y\to 0}{\sim} \frac{\Gamma(2-\alpha)}{\alpha-1}y^{-1}\bar\nu(1/y).$$

• Case $\nu \in \mathcal{D}_2$. We split the integral. First, we have

$$\int_0^{1/\sqrt{y}}(1-e^{-yu})\bar\nu(u)\mathrm{d}u \overset{y\to 0}{\sim} y\int_0^{1/\sqrt{y}}u\bar\nu(u)\mathrm{d}u \overset{y\to 0}{\sim} Cy\log(1/y)/2.$$

since $\int_0^{1/\sqrt{y}}(1-e^{-yu}+yu)\bar\nu(u)\mathrm{d}u = o(y\int_0^{1/\sqrt{y}}u\bar\nu(u)\mathrm{d}u)$. Moreover,

$$
\begin{aligned}
\int_{1/\sqrt{y}}^{\infty}(1-e^{-yu})\bar\nu(u)\mathrm{d}u &= y^{-1}\int_{\sqrt{y}}^{\infty}(1-e^{-u})\bar\nu(u/y)\mathrm{d}u \\
&\overset{y\to 0}{\sim} Cy\int_{\sqrt{y}}^{\infty}(1-e^{-u})u^{-2}\mathrm{d}u \quad \text{using } \nu\in\mathcal{D}_2 \\
&\overset{y\to 0}{\sim} Cy\int_{\sqrt{y}}^{1}u^{-1}\mathrm{d}u = Cy\log(1/y)/2
\end{aligned}
$$

Then

$$\int_0^{\infty}(1-e^{-yu})\bar\nu(u)\mathrm{d}u \overset{y\to\infty}{\sim} Cy\log(1/y),$$

which proves the first part of the lemma using (32).

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 Theorem 13.17 in [16]. $\qquad\square$

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.

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

We have the following analog of Proposition 4. $\overrightarrow{\mathcal{R}(\alpha,\lambda)}$ and $\overleftarrow{\mathcal{R}(\alpha,\lambda)}$ are independent, identically distributed and independent of $(g(\alpha,\lambda),d(\alpha,\lambda))$. Moreover $\overrightarrow{\mathcal{R}(\alpha,\lambda)}$ and $\overleftarrow{\mathcal{R}(\alpha,\lambda)}$ are respectively the range of the subordinators $\overrightarrow{\tau}^{\alpha,\lambda}$ and $\overleftarrow{\tau}^{\alpha,\lambda}$, whose Laplace exponent $\kappa^{\alpha,\lambda}$ is the inverse function of $-\Psi^{\alpha,\lambda}$. Finally, using $\left[\Psi^{\alpha,\lambda}\right]'(0) = -\lambda$, the counterpart of (30) gives for $\rho,\mu \geq 0$ and $\rho \neq \mu$:

$$\mathbb{E}\big(\exp(\rho g(\alpha,\lambda)-\mu d(\alpha,\lambda))\big) = \lambda\frac{\kappa^{\alpha,\lambda}(\rho)-\kappa^{\alpha,\lambda}(\mu)}{\rho-\mu}. \tag{34}$$

The proof of these results follow the proof of Proposition 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 $\left(Y_x^{\alpha,\lambda} - I_x^{\alpha,\lambda}\right)_{x\in\mathbb{R}}$. The regeneration property is a consequence of the counterpart of Lemma 2 which can be stated as follows. For all $x \in \mathbb{R}$,

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

and distributed as $\left(Y_y^{\alpha,\lambda}\right)_{y\geq 0}$. As Lemma 2, this property is an extension to the stopping time $d_x(\mathcal{R}(\alpha, \lambda))$ of the following obvious result : $\left(Y^{\alpha,\lambda}_{x+y} - Y^{\alpha,\lambda}_x\right)_{y\geq 0}$ is independent of $\left(Y^{\alpha,\lambda}_{x-y} - Y^{\alpha,\lambda}_x\right)_{y\geq 0}$ and distributed as $\left(Y_y^{\alpha,\lambda}\right)_{y\geq 0}$.

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

$$\overrightarrow{\tau}_x^{\alpha,\lambda} := \inf\{y \geq 0 : Y^{\alpha,\lambda}_{d(\alpha,\lambda)+y} - Y^{\alpha,\lambda}_{d(\alpha,\lambda)} < -x\}.$$

For $\lambda > 0$, $[\Psi^{\alpha,\lambda}]'(0) = -\lambda < 0$ so we can apply Theorem 1 and $\overrightarrow{\tau}^{\alpha,\lambda}$ is a subordinator whose Laplace $\kappa^{\alpha,\lambda}$ is the inverse function of $-\Psi^{\alpha,\lambda}$. Moreover its range is a.s. equal to $\overrightarrow{\mathcal{R}(\alpha,\lambda)}$, since the Lévy process $(Y^{\alpha,\lambda}_{d(\alpha,\lambda)+y} - Y^{\alpha,\lambda}_{d(\alpha,\lambda)})_{y\geq 0}$ is regular for $]-\infty, 0[$ (Proposition 8 on page 84 in [5]).

To prove Theorem 2 and Theorem 3, we need a final lemma, which states the convergence of the Laplace exponent of $\overrightarrow{\mathcal{R}(t)}$.

Lemma 4. *If $\nu \in \mathcal{D}_\alpha$ ($\alpha \in [1,2]\cup\{2+\}$), then for all $z \geq 0$ and $\lambda > 0$,*

$$(1 - \mathrm{m}t)\epsilon_\alpha(t)^{-1}\kappa^{(t)}(\epsilon_\alpha(t)z) \overset{t\to 1/\mathrm{m}}{\longrightarrow} \kappa^{\alpha,1}(z),$$

$$xf_\alpha(x)\kappa^{((1-\lambda f_\alpha(x))/\mathrm{m})}(x^{-1}z) \overset{x\to\infty}{\longrightarrow} \kappa^{\alpha,\lambda}(z).$$

Remark 6. If $\bar{\nu}$ is regularly varying at infinity of index $-\alpha \in]-1, -2[$, we have similarly

$$\bar{\nu}(x)^{-1}\kappa^{((1-\lambda x\bar{\nu}(x))/\mathrm{m})}(x^{-1}z) \overset{x\to\infty}{\longrightarrow} \kappa^{\alpha,\lambda}(z).$$

Proof. First we prove that

$$\alpha(t) \overset{t\to 1/\mathrm{m}}{\sim} \beta(t) \quad \Rightarrow \quad \kappa^{(t)}(\alpha(t)) \overset{t\to 1/\mathrm{m}}{\sim} \kappa^{(t)}(\beta(t)). \tag{35}$$

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

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

This gives

$$\min(\frac{\alpha(t)}{\beta(t)}, 1) \leq \frac{\int_0^\infty (1 - e^{-\alpha(t)x})\Pi^{(t)}(\mathrm{d}x)}{\int_0^\infty (1 - e^{-\beta(t)x})\Pi^{(t)}(\mathrm{d}x)} \leq \max(\frac{\alpha(t)}{\beta(t)}, 1),$$

and proves (35) recalling (20).

Then the first part of Lemma 3 and the identity $\kappa^{(t)} \circ (-\Psi^{(t)}) = \mathrm{Id}$ give the first part of Lemma 4. Indeed for every $y \geq 0$, $\Psi^{(t)}(\epsilon_\alpha(t)(1 - \mathrm{m}t)^{-1}y) \overset{t \to 1/\mathrm{m}}{\sim} \epsilon_\alpha(t)\Psi^{\alpha,1}(y)$. So (35) entails

$$\epsilon_\alpha(t)(1 - \mathrm{m}t)^{-1}y \overset{t \to 1/\mathrm{m}}{\sim} \kappa^{(t)}(-\epsilon_\alpha(t)\Psi^{\alpha,1}(y)).$$

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

Proof of Theorem 2. First, we prove that $\epsilon_\alpha(t).(g(t), d(t))$ converges weakly as t tends to $1/\mathrm{m}$ to $(g(\alpha,1), d(\alpha,1))$. Indeed by (31), we have

$$\mathbb{E}\big(\exp(\rho\epsilon_\alpha(t)g(t) - \mu\epsilon_\alpha(t)d(t))\big) = (1 - \mathrm{m}t)\frac{\kappa^{(t)}(\epsilon_\alpha(t)\rho) - \kappa^{(t)}(\epsilon_\alpha(t)\mu)}{\epsilon_\alpha(t)(\rho - \mu)}.$$

Let $t \to 1/\mathrm{m}$ using Lemma 4 and find the right hand side of (34) to conclude.

Moreover $\epsilon_\alpha(t)\overrightarrow{\mathcal{R}(t)}^{cl}$ (resp. $\epsilon_\alpha(t)\overleftarrow{\mathcal{R}(t)}^{cl}$) converges weakly in $\mathcal{H}(\mathbb{R}_+)$ as t tends to $1/\mathrm{m}$ to $\overrightarrow{\mathcal{R}(\alpha,1)}^{cl}$ (resp. $\overleftarrow{\mathcal{R}(\alpha,1)}^{cl}$). Indeed, by Proposition (3.9) in [14], this is a consequence of the convergence of the Laplace exponent of $\epsilon_\alpha(t)\overrightarrow{\mathcal{R}(t)}$ given by Lemma 4. Informally, $\epsilon_\alpha(t)\overrightarrow{\mathcal{R}(t)}^{cl}$ is the range of $\big(\epsilon_\alpha(t)\overrightarrow{\tau}^{(t)}_{(1-\mathrm{m}t)\epsilon_\alpha(t)^{-1}z}\big)_{z \geq 0}$ whose convergence in $\mathbb{D}$ follows from Lemma 4.

We can now prove the theorem. We know from (19) that

$$\epsilon_\alpha(t)\mathcal{R}(t) = \epsilon_\alpha(t).(d(t) + \overrightarrow{\mathcal{R}(t)}) \sqcup (\epsilon_\alpha(t).(\overset{\frown}{-g(t) + \overleftarrow{\mathcal{R}(t)}}))$$

where $\epsilon_\alpha(t)\overleftarrow{\mathcal{R}(t)}$, $\epsilon_\alpha(t)(-g(t), d(t))$ and $\epsilon_\alpha(t)\overrightarrow{\mathcal{R}(t)}$ are independent by Proposition 4. Similarly

$$\mathcal{R}(\alpha,1) = (d(\alpha,1) + \overrightarrow{\mathcal{R}(\alpha,1)}) \sqcup (\overset{\frown}{-g(\alpha,1) + \overleftarrow{\mathcal{R}(\alpha,1)}})$$

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

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

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

Proof of Corollary 2. The first result is a direct consequence of Theorem 2. We have then

$$\epsilon_\alpha(t)l(t) \overset{t \to 1/\mathrm{m}}{\Longrightarrow} d(\alpha,1) - g(\alpha,1).$$

21

Moreover, as $\kappa^{2+,1} \circ (-\Psi^{2+,1}) = \mathrm{Id}$, we can compute $\kappa^{2+,1}$ and (34) gives

$$\mathbb{E}\big(\exp(-\mu(d(2+,1))-g(2+,1))\big) = (\kappa^{2+,1})'(\mu) = \Big(\frac{1+\sqrt{1+2\frac{\mathrm{m}_2}{\mathrm{m}}\mu}}{\frac{\mathrm{m}_2}{\mathrm{m}}}\Big)'(\mu) = \frac{1}{\sqrt{-1+2\frac{\mathrm{m}_2}{\mathrm{m}}\mu}}.$$

So, by identification of Laplace transform, $d(\alpha,1) - g(\alpha,1)$ is a gamma variable of parameter $(1/2, \mathrm{m}/(4\mathrm{m}_2))$ and we get the result. The argument is similar in the case $\alpha = 2$. $\qquad\square$

Proof of Theorem 3. The argument is similar to that of the proof of Theorem 2 using the others limits of Lemma 4. We get that if $x \to \infty$ and $1 - \mathrm{m}t \sim \lambda f_\alpha(x)$ with $\lambda > 0$, then $x^{-1}\mathcal{R}(t)$ converges weakly in $\mathcal{H}(\mathbb{R})$ to $\{x \in \mathbb{R} : Y_x^{\alpha,\lambda} = I_x^{\alpha,\lambda}\}^{cl}$. The theorem follows by restriction to $[0,1]$. $\qquad\square$

To prove the corollary of Theorem 3, we need the following result.

Lemma 5. *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. • Let $0 \le a < b \le 1$. Note that for all $\lambda' \ge 1$ and $x \ge 0$, $Y_x^{\alpha,\lambda'} - Y_x^{\alpha,1} = (1 - \lambda')x$ ensures that $I_x^{\alpha,\lambda'} - I_x^{\alpha,1} \ge (1 - \lambda')x$. Then,

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

So a.s there exists λ' such that

$$Y_{a+2\frac{b-a}{3}}^{\alpha,\lambda'} < I_{a+\frac{b-a}{3}}^{\alpha,\lambda'}.$$

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.

• As $(Y_x^{\alpha,0})_{x\in\mathbb{R}}$ oscillates when x tends to $-\infty$ (see [5] Corollary 2 on page 190), then

$$I_0^{\alpha,\lambda} \xrightarrow{\lambda\to 0} -\infty,$$

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

Proof of Corollary 3. The first result is a direct consequence of Theorem 3.

If $o(1 - \mathrm{m}t) = f_\alpha(x)$ $(x \to \infty)$, then for every $\lambda > 0$ and x large enough, $t \le (1 - \lambda f_\alpha(x))/\mathrm{m}$ and

$$B_1(x,t)/x \le B_1(x, \frac{1-\lambda f_\alpha(x)}{\mathrm{m}})/x.$$

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 $B_1(x,t)/x \overset{x\to\infty}{\longrightarrow} 0$ in $\mathbb{P}$.

Similarly if $1 - \mathrm{m}t = o(f_\alpha(x))$ $(x \to \infty)$, then for every $\lambda > 0$ and x large enough,

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

Letting λ tend to 0, Lemma 5 entails that $B_1(x,t)/x \overset{x\to\infty}{\longrightarrow} 1$ in $\mathbb{P}$. $\qquad\square$

References

[1] V. Bansaye (2006). On a model for the storage of files on a hardware I : Statistics at a fixed time and asymptotics. *Preprint* avialable via http://arxiv.org/PS_cache/math/pdf/0611/0611432.pdf.

[2] V. Bansaye (2007). On a model for the storage of files on a hardware II : Evolution of a typical data block. *J. Appl. Prob.* Vol 44, no 4, 901-927.

[3] J. Bertoin (1997). Subordinators: examples and applications. *Lectures on probability theory and statistics* (Saint-Flour), Lecture Notes in Math., 1717, Springer, Berlin, 1999.

[4] J. Bertoin (2000). Subordinators, Lévy processes with no negative jumps and branching processes. Lecture notes for MaPhySto, August 2000. Available via http://www.maphysto.dk/publications/MPS-LN/2000/8.pdf.

[5] J. Bertoin (1996). *Lévy processes*. Cambridge Tracts in Mathematics, 121. Cambridge University Press, Cambridge.

[6] J. Bertoin, G. Miermont (2006). Asymptotics in Knuth's parking problem for caravans. *Random Structures and Algorithms*. Vol.29, 38-55.

[7] N.H. Bingham, C.M. Goldie, J.L. Teugels (1989). *Regular variation*. Encyclopedia of Mathematics and its Applications, 27. Cambridge University Press, Cambridge.

[8] O. J. Boxma, J. W. Cohen (1999). Heavy-traffic analysis for the $GI/G/1$ queue with heavy-tailed distributions. *Queueing Systems Theory Appl. 33*, no. 1-3, 177–204.

[9] P. Chassaing, G. Louchard (2002). Phase transition for parking blocks, Brownian excursion and coalescence. *Random Structures Algorithms*. Vol. 21, no. 1, 76-119.

[10] P. Chassaing, P. Flajolet (2003). Hachage, arbres, chemins. *Gazette des mathématiciens*. no. 95, janvier 2003.

[11] J. W. Cohen (1982). *The single server queue.* Second edition. North-Holland Series in Applied Mathematics and Mechanics, 8. North-Holland Publishing Co., Amsterdam-New York.

[12] P. Flajolet, P. Poblete, A. Viola (1998). On the analysis of linear probing hashing. *Algorithica.* Vol. 22, no. 4, 490-515.

[13] D. Foata, J. Riordan (1974). Mappings of acyclic and parking functions. *Aequationes Math. 10*, 10-22.

[14] P. J. Fitzsimmons, B. Fristedt, B. Maisonneuve (1985). Intersections and limits of regenerative sets. *Z. Wahrsch. Verw. Gebiete 70*, no. 2, 157–173.

[15] J. Jacod, A. N. Shiryaev (2003). *Limit theorems for stochastic processes.* Second edition. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 288. Springer-Verlag, Berlin.

[16] O. Kallenberg (2002). *Foundations of modern probability.* Second edition. Probability and its Applications (New York). Springer-Verlag, New York.

[17] H. Kaspi, M. Rubinovitch (1988). Regenerative sets and their applications to Markov storage systems. *Queueing theory and its applications*, 413–427, , CWI Monogr., 7, North-Holland, Amsterdam.

[18] J. F. C. Kingman (1965). The heavy traffic approximation in the theory of queues. *Proc. Sympos. Congestion Theory* (Chapel Hill, N.C., 1964) p. 137-169, Univ. North Carolina Press, Chapel Hill, N.C.

[19] B. Maisonneuve (1983). Ensembles régénératifs de la droite. [Regenerative sets of the line] *Z. Wahrsch. Verw. Gebiete 63*, no. 4, 501-510.

[20] G. Materon (1978). *Random sets and integral geometry.* J. Wiley & Sons, New York.

[21] P.A. Meyer (1971). Processus de Poisson ponctuels, d'après K. Itô. *Séminaire de Probabilités V.* Lectures Notes in Mathematics 191, Springer-Verlag, berlin, 209-210.

[22] P. W. Millar (1977). Zero-one laws and the minimum of a Markov process. *Trans. Am. Math. Soc.* Vol. 226, 365-391.

[23] J. Pitman (1986). Stationary excursions. In *Séminaire de Probabilités XXI*, Lecture notes in Math. 1247, 289-302. Springer.

[24] N. U. Prabhu (1998). *Stochastic storage processes. Queues, insurance risk, dams, and data communication.* Second edition. Applications of Mathematics (New York), 15. Springer-Verlag, New York.

[25] M. I. Taksar (1980). Regenerative sets on real line. *Séminaire de Probabilités XIV*, 437-474. Lecture Notes in Math., 784, Springer, Berlin.

[26] M. I. Taksar (1987). Stationary Markov sets. *Séminaire de Probabilités XXI*, 303-340. Lecture Notes in Math., 1247, Springer, Berlin.