

ON A MODEL FOR THE STORAGE OF FILES ON A HARDWARE. II. EVOLUTION OF A TYPICAL DATA BLOCK

VINCENT BANSAYE,* *Université Pierre et Marie Curie and CNRS*

Abstract

We consider the generalized version in continuous time of the parking problem of Knuth introduced in Bansaye (2006). Files arrive following a Poisson point process and are stored on a hardware identified with the real line, at the right of their arrival point. Here we study the evolution of the endpoints of the data block straddling 0, which is empty at time 0 and is equal to \mathbb{R} at a deterministic time.

Keywords: Parking problem; data storage; random covering; Poisson point process; Lévy process

2000 Mathematics Subject Classification: Primary 60D05; 60G55; 60J80; 68B15

1. Introduction

This paper is a continuation of [1], but it can be read independently. We consider a generalized version in continuous time of the original parking problem of Knuth as a model for the storage of files on a hardware. We are interested in the evolution of a typical data block while files are stored on the hardware, and we shall characterize the process of the endpoints and the length of this block.

We now recall the process of file storage. In the original problem of Knuth, files arrive successively at independent locations chosen uniformly among n spots. They are stored in the first free spot at the right of their arrival point; see [5], [6], [8], and [9]. In the model considered here, the hardware is identified with the real line and a file labeled i of length (or size) l_i arrives at time t_i on the real line at location x_i . 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 it is shifted to the right until a free space is found, and it may be split into several parts which are stored in the closest free spots.

The arrival of files follows a Poisson point process (PPP), i.e. $\{(t_i, x_i, l_i) : i \in \mathbb{N}\}$ is a PPP with intensity $dt \otimes dx \otimes \nu(dl)$ on $\mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^+$. We let $\bar{\nu}(x) = \nu((x, \infty])$, and we assume that $m := \int_0^\infty l\nu(dl) < \infty$. So m denotes the mean of the total sizes of files which arrive during a unit time interval on some interval with unit length. In [1] this random covering was rigorously constructed and some statistics of this covering were given. Bansaye [1] proved that the hardware becomes full at a deterministic time equal to $1/m$, studied the asymptotics at this saturation time, and characterized 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 the sides of this block.

Received 22 January 2007; revision received 6 November 2007.

* Postal address: Laboratoire de Probabilités et Modèles Aléatoires, Université Pierre et Marie Curie, UMR 7599, 175 rue du Chevaleret, 75013 Paris, France. Email address: bansaye@ccr.jussieu.fr

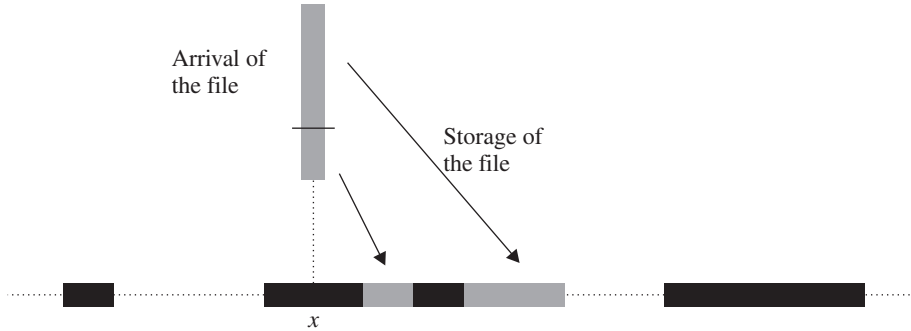


FIGURE 1: Arrival and storage of a file on the hardware, where the data blocks are represented by black rectangles.

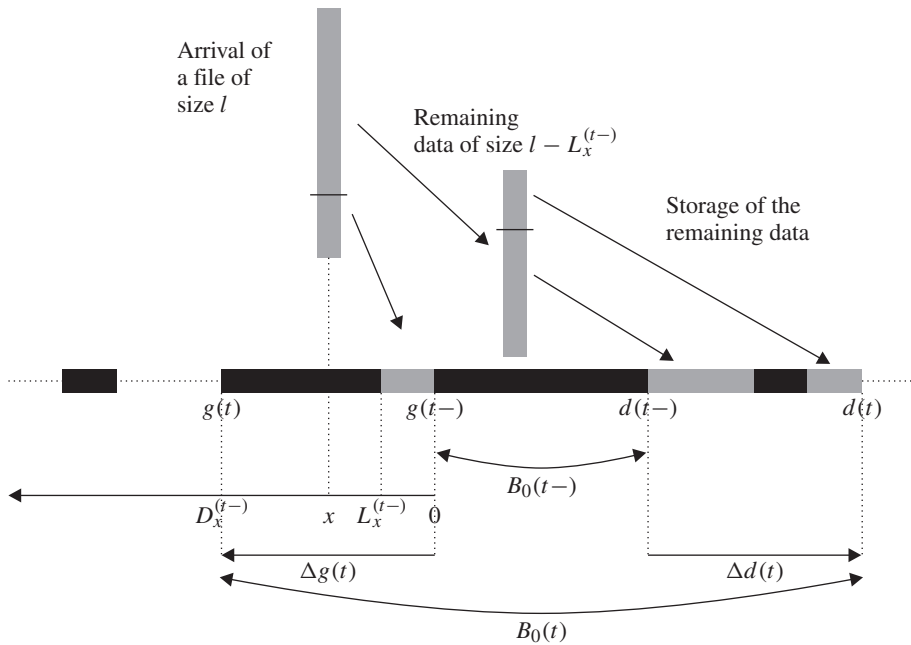


FIGURE 2: Jumps of the endpoints of B_0 ($\Delta g(t)$ and $\Delta d(t)$) and the remaining data induced by the arrival of a file at time t at the left of $B_0(t^-)$.

In this work we focus on the dynamics of the covering, and we shall study the block of data straddling a typical point, say 0 for simplicity, which is denoted by B_0 . Thus, $B_0(t)$ is the block of data of the hardware containing 0 at time t . We will show that its endpoints and its length are pure-jump Markov processes. Specifically, if a file arrives at time t at the left of $B_0(t^-)$ and cannot be stored entirely at its left, it yields a jump of the left endpoint of B_0 . The data of this file which cannot be stored at the left of $B_0(t^-)$ is called the *remaining data*. The remaining data yield a jump of the right endpoint of B_0 ; see Figure 2. We shall prove that these events occur at instants which accumulate at $1/m$ and induce a random partition of the time interval $[0, 1/m]$ with the Poisson–Dirichlet distribution (Theorem 2), and that the jumps of the endpoints at these instants form a PPP on $[0, 1/m] \times \mathbb{R}_+ \times \mathbb{R}_+$ (Proposition 3). Moreover, the

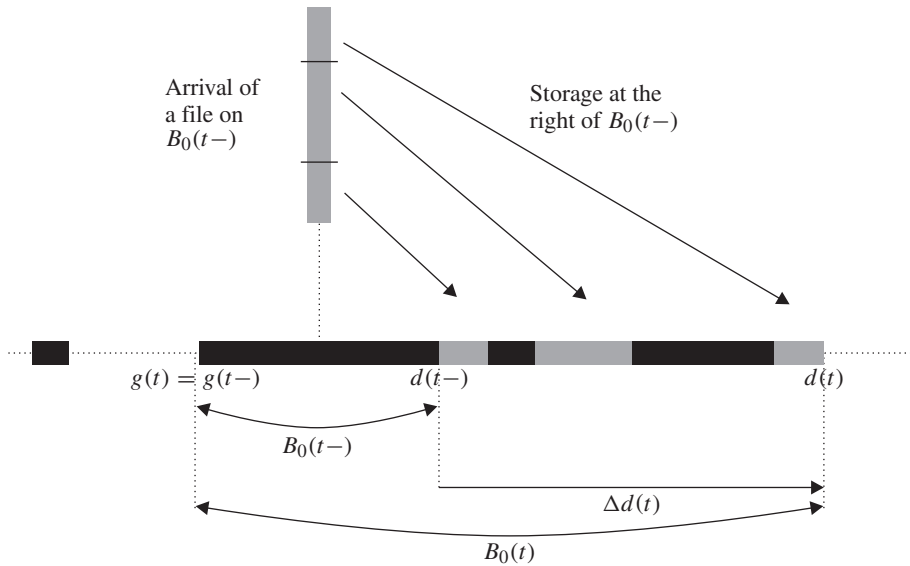


FIGURE 3: Jump of the right endpoint of B_0 ($\Delta d(t)$) induced by the arrival of a file at time t on $B_0(t^-)$.

successive quantities of remaining data form an independent and identically distributed (i.i.d.) sequence (Corollary 2).

If a file arrives at B_0 , it yields a jump of the right endpoint only; see Figure 3. The other files do not immediately induce a jump of B_0 and we obtain the evolution of $(B_0(t))_{t \geq 0}$ (Theorem 4). Finally, we prove that the process describing the length of $(B_0(t))_{t \geq 0}$ is a branching process with immigration (Corollary 5).

2. Preliminaries

The covering $\mathcal{C}(t)$, described in the introduction, has been constructed in Section 2.1 of [1], and we recall some useful results of that work. We denote by $\mathcal{R}(t)$ the complementary set of $\mathcal{C}(t)$. It is natural and convenient to decide that files, and so $\mathcal{C}(t)$ and $\mathcal{R}(t)$, are closed at the left and open at the right. We introduce the process $(Y_x^{(t)})_{x \in \mathbb{R}}$ defined by

$$Y_0^{(t)} := 0, \quad Y_b^{(t)} - Y_a^{(t)} = \sum_{\substack{l_i \leq t \\ x_i \in (a,b]}} l_i - (b - a) \quad \text{for } a < b. \tag{1}$$

It has càdlàg paths (those that are continuous from the right with left limits) and stationary independent increments. The process $(Y_x^{(t)})_{x \geq 0}$ is then a Lévy process. Its drift is equal to -1 , and its Lévy measure is equal to $t\nu$. Its Laplace exponent, $\Psi^{(t)}$, defined by

$$E(\exp(-\rho Y_x^{(t)})) = \exp(-x\Psi^{(t)}(\rho)) \quad \text{for all } \rho \geq 0, \tag{2}$$

is given by

$$\Psi^{(t)}(\rho) = -\rho + \int_0^\infty (1 - e^{-\rho x})t\nu(dx) \quad \text{for all } \rho \geq 0.$$

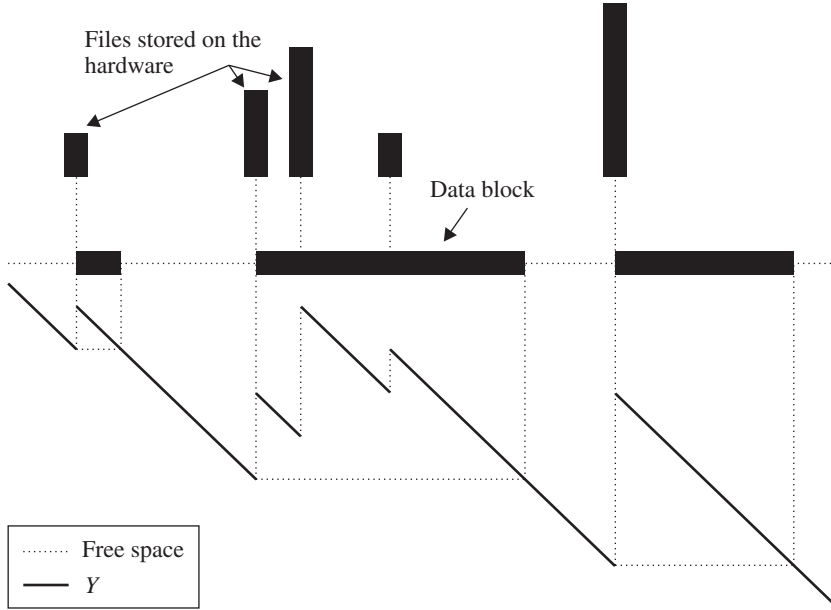


FIGURE 4: Representation of Y on a part of the hardware.

As suggested by Figure 4, introducing its infimum process, $I_x^{(t)} := \inf\{Y_y^{(t)} : y \leq x\}$ for every $x \in \mathbb{R}$, we obtain the following expression for the covering and the free space, respectively [1, Proposition 1]:

$$\mathcal{C}(t) = \{x \in \mathbb{R} : Y_x^{(t)} > I_x^{(t)}\}, \quad \mathcal{R}(t) = \{x \in \mathbb{R} : Y_x^{(t)} = I_x^{(t)}\} \quad \text{almost surely (a.s.)} \quad (3)$$

We have the following geometric properties (see [1, Proposition 3]), where, for all $R \subset \mathbb{R}$ and $x \in \mathbb{R}$, $d_x(R) := \inf\{y \in R : y > x\}$.

Proposition 1. ([1, Proposition 3].) *For every $t \geq 0$, $\mathcal{R}(t)$ is stationary, its closure is symmetric in distribution, and it enjoys the following 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 $x \in \mathbb{R}$, $P(x \in \mathcal{C}(t)) = \min(1, mt)$.*

By Proposition 2 of [1], the time when the hardware becomes full is equal to $1/m$, that is, a.s. $\mathcal{C}(t) = \mathbb{R}$ if and only if $t \geq 1/m$. Thus, we already know that $B_0(0) = \emptyset$ and $B_0(1/m) = \mathbb{R}$, and we shall study $(B_0(t))_{t \in [0, 1/m]}$. With this in mind, we introduce the left endpoint, $g(t)$, the right endpoint, $d(t)$, and the length, $l(t)$, of the data block containing 0:

$$\begin{aligned} d(t) &:= d_0(\mathcal{R}(t)), & g(t) &:= -d_0(-\mathcal{R}(t)), \\ B_0(t) &:= [g(t), d(t)), & l(t) &:= d(t) - g(t). \end{aligned}$$

We will also need the free space at the right of $B_0(t)$, denoted by $\overrightarrow{\mathcal{R}}(t)$, and at the left of $B_0(t)$, turned over, closed at the left, and open at the right, denoted by $\overleftarrow{\mathcal{R}}(t)$. If $\mathcal{R} \subset \mathbb{R}$ and $\mathcal{R} = \sqcup_{n \in \mathbb{N}} [a_n, b_n)$ (where ‘ \sqcup ’ denotes the disjoint union), we denote by $\widehat{\mathcal{R}} = \sqcup_{n \in \mathbb{N}} [-b_n, -a_n)$ the symmetric set closed at the left and open at the right. Then we can define (see Figure 5 and

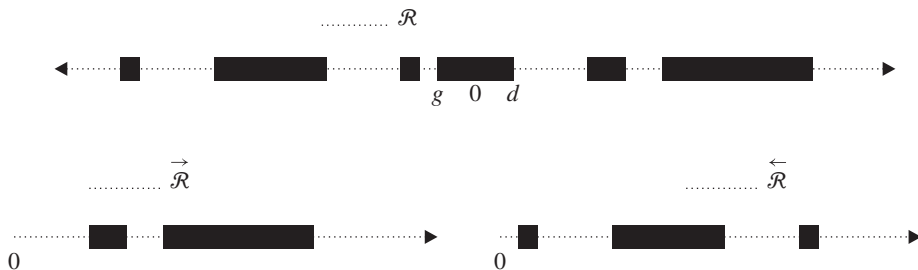


FIGURE 5: Time t is omitted and \mathcal{R} , $\vec{\mathcal{R}}$, and $\overleftarrow{\mathcal{R}}$ are represented by the dotted line.

[1, Section 3.1] for details)

$$\vec{\mathcal{R}}(t) := (\mathcal{R}(t) - d(t)) \cap [0, \infty], \quad \overleftarrow{\mathcal{R}}(t) := \widetilde{\overleftarrow{\mathcal{R}}}(t),$$

which satisfy the following identity:

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

In Section 3 of [1] it was proved that $\vec{\mathcal{R}}(t)$ and $\overleftarrow{\mathcal{R}}(t)$ are the range of the processes $(\vec{\tau}_x^{(t)})_{x \geq 0}$ and $(\overleftarrow{\tau}_x^{(t)})_{x \geq 0}$, respectively, defined by

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

Moreover, denoting by $\kappa^{(t)}$ the inverse function of $-\Psi^{(t)}$ and by $\Pi^{(t)}$ its Lévy measure:

$$\kappa^{(t)} \circ (-\Psi^{(t)}) = \text{Id}, \quad \kappa^{(t)}(\rho) = \rho + \int_0^\infty (1 - e^{-\rho x}) \Pi^{(t)}(dx) \quad \text{for all } \rho \geq 0, \quad (4)$$

(where ‘Id’ denotes the identity) enabled us to describe $\mathcal{R}(t)$ in the following way (see [1, Section 3.1]).

Theorem 1. (i) *The processes $\vec{\tau}$ and $\overleftarrow{\tau}$ are two independent subordinators with Laplace exponent $\kappa^{(t)}$, which are independent of $(g(t), d(t))$.*

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

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

$$P(l(t) \in dx) = (1 - mt)(\delta_0(dx) + \mathbf{1}_{\{x>0\}} x \Pi^{(t)}(dx)),$$

where U is a uniform random variable on $[0, 1]$ independent of $l(t)$ and δ_0 denotes the Dirac mass at 0.

For the basic example, $\nu = \delta_1$, we obtain, for all $x \in \mathbb{R}_+$ and $n \in \mathbb{N}$,

$$P(Y_x^{(t)} + x = n) = e^{-tx} \frac{(tx)^n}{n!} \quad \text{and}$$

$$P(\vec{\tau}_x^{(t)} = x + n) = \frac{x}{x+n} e^{-t(x+n)} \frac{(t(n+x))^n}{n!}, \quad \Pi^{(t)}(n) = \frac{(tn)^n}{n(n!)} e^{-tn}. \quad (5)$$

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

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

In [1] we also proved the following identities:

$$\bar{\Pi}^{(t)}(0) = t\bar{v}(0), \quad \int_0^\infty x \Pi^{(t)}(dx) = \frac{mt}{1 - mt}, \quad [k^{(t)}]'(0) = \frac{1}{1 - mt}, \quad (6)$$

and the following identities of measures on $\mathbb{R}_+ \times \mathbb{R}_+$:

$$x P(\tau_l^{\leftarrow(t)} \in dx) dl = x P(\tau_l^{\rightarrow(t)} \in dx) dl = l P(-Y_x^{(t)} \in dl) dx. \quad (7)$$

Finally, we recall a useful expression for the law of $g(t)$ [1, Equation (25)]. For all $t \in [0, 1/m)$ and $\lambda \geq 0$,

$$E(e^{\lambda g(t)}) = \exp\left(\int_0^\infty (e^{-\lambda x} - 1)x^{-1} P(Y_x^{(t)} > 0) dx\right). \quad (8)$$

We can now focus on the evolution of the block containing 0, B_0 . First we prove some properties of absence of memory (Section 3): the evolution of B_0 after time t depends on the past evolution of this block only through $l(t)$ (Markov property). Then we focus on the left endpoint: it is an additive process, and we give its Lévy measure. As a consequence, we obtain the distribution of the instants at which the left endpoint jumps (Section 4). We then derive the distribution of the remaining data, which completes the description of the process of storage at the left endpoint (Section 5). Also, by taking into account the data which has arrived at B_0 , we then obtain the evolution of $(g(t), d(t))$ (Section 6). The latter characterizes the evolution of the right endpoint and the length (Section 7).

3. The Markov property of B_0

We have already proved that $\mathcal{R}(t)$ satisfies a ‘spatial’ regeneration property (see Proposition 1). To study the evolution of B_0 , we need a ‘time’ regeneration property. Here we prove that the evolution of the block containing 0 up to time t is independent of the covering outside $[g(t), d(t)]$ up to time t . In Section 5 this property will ensure that the evolution of B_0 after time t depends on the past evolution of this block only through $l(t)$ (Markov property).

Proposition 2. *For every $t \in [0, 1/m)$, the following three processes, with values in the space of subsets of \mathbb{R} , are independent:*

$$\begin{aligned} (g(t) - \mathcal{R}(s)) \cap [0, \infty), & \quad 0 \leq s \leq t, \\ (\mathcal{R}(s) - d(t)) \cap [0, \infty), & \quad 0 \leq s \leq t, \\ \mathcal{R}(s) \cap [g(t), d(t)], & \quad 0 \leq s \leq t. \end{aligned}$$

Remark 1. Actually, we have the following regeneration property: for all $t \in [0, 1/m)$ and for all $x \in \mathbb{R}$, $((\mathcal{R}(s) - d_x(\mathcal{R}(t))) \cap [0, \infty) : s \in [0, t])$ is independent of $((\mathcal{R}(s) - d_x(\mathcal{R}(t))) \cap (-\infty, 0] : s \in [0, t])$ and is distributed as $((\mathcal{R}(s) - d_0(\mathcal{R}(t))) \cap [0, \infty) : s \in [0, t])$.

This result is a direct consequence of Lemma 1, below, where we consider the point processes of files until time t at the left of, at the right of, and inside $[g, d]$, respectively:

$$\begin{aligned}
 P_g(t) &:= \{(t_i, g - x_i, l_i) : t_i \leq t, x_i < g\}, \\
 P^d(t) &:= \{(t_i, x_i - d, l_i) : t_i \leq t, d < x_i\}, \\
 P_g^d(t) &:= \{(t_i, x_i, l_i) : t_i \leq t, g \leq x_i \leq d\}.
 \end{aligned}$$

Lemma 1. *For every $t \in [0, 1/m)$, the point processes $P_{g(t)}(t)$, $P_{g(t)}^{d(t)}(t)$, and $P_{d(t)}(t)$ are independent.*

Proof. First we prove a weaker result, where the times $(t_i)_{i \in \mathbb{N}}$ are not taken into account. Denote by $(\tilde{Y}_x^{(t)})_{x \geq 0}$ the càdlàg version of $(Y_{-x}^{(t)})_{x \geq 0}$. This is a spectrally negative Lévy process with bounded variation, which drifts to ∞ . Note that

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

Then $(\tilde{Y}_{-g(t)+x}^{(t)} - \tilde{Y}_{-g(t)}^{(t)})_{x \geq 0}$ is independent of $(\tilde{Y}_x^{(t)})_{0 \leq x \leq -g(t)}$ (the decomposition of a Lévy process at its infimum [11]). Considering the locations and sizes of the jumps of these two processes yields

$$\{(g(t) - x_i, l_i) : t_i \leq t, x_i < g(t)\} \text{ is independent of } \{(x_i, l_i) : t_i \leq t, g(t) \leq x_i \leq 0\}.$$

Adding that $\{(x_i, l_i) : t_i \leq t, x_i > 0\}$ is independent of $\{(x_i, l_i) : t_i \leq t, x_i \leq 0\}$ and $g(t)$ is $\{(x_i, l_i) : t_i \leq t, x_i \leq 0\}$ measurable, we obtain

$$\{(g(t) - x_i, l_i) : t_i \leq t, x_i < g(t)\} \text{ is independent of } \{(x_i, l_i) : t_i \leq t, x_i \geq g(t)\}.$$

We now extend the preceding by incorporating the times, $(t_i)_{i \in \mathbb{N}}$. In this direction, we recall that if $(\tilde{x}_i, \tilde{l}_i)_{i \in \mathbb{N}}$ is a PPP on $\mathbb{R} \times \mathbb{R}_+$ with intensity $t \, dx \otimes \nu(dl)$ and $(\tilde{t}_i)_{i \in \mathbb{N}}$ is an i.i.d. sequence distributed uniformly on $[0, t]$ then $\{(\tilde{t}_i, \tilde{x}_i, \tilde{l}_i) : i \in \mathbb{N}\}$ is distributed as $\{(t_i, x_i, l_i) : i \in \mathbb{N}, t_i \leq t\}$. Adding that $g(t)$ is $\{(x_i, l_i) : i \in \mathbb{N}, t_i \leq t\}$ measurable, we obtain

$$\{(t_i, g(t) - x_i, l_i) : t_i \leq t, x_i < g(t)\} \text{ is independent of } \{(t_i, x_i, l_i) : t_i \leq t, x_i \geq g(t)\}.$$

This ensures that $P_{g(t)}(t)$ is independent of $(P_{g(t)}^{d(t)}(t), P^{d(t)}(t))$.

We can similarly prove that $P^{d(t)}(t)$ is independent of $(P_{g(t)}(t), P_{g(t)}^{d(t)}(t))$ using the fact that $(Y_{d(t)+x}^{(t)} - Y_{d(t)}^{(t)})_{x \geq 0}$ is independent of $(Y_x^{(t)})_{x \leq d(t)}$ or Lemma 2 of [1].

This lemma guarantees the absence of memory at the left of $B_0(t)$; we have the following consequence.

Corollary 1. *The process $(g(t))_{t \in [0, 1/m]}$ has decreasing càdlàg paths with independent increments.*

Proof. Let $0 \leq t < t+s \leq 1/m$. The increment $g(t+s) - g(t)$ depends only on $\overleftarrow{\mathcal{R}}(t)$ and the point process of files which arrive after time t at the left of $B_0(t)$, $\{(t_i, x_i - g(t), l_i) : t_i > t, x_i < g(t)\}$. By the Poissonian property, these two quantities are independent and $(g(u) : u \in [0, t])$ is independent of this point process of files. Moreover, $(g(u) : u \in [0, t])$ is also independent of $(g(t) - \mathcal{R}(t)) \cap [0, \infty)$ by Proposition 2. So $(g(u) : u \in [0, t])$ is independent of $g(t+s) - g(t)$.

This explains the observation made in Section 3 of [1], which states that the distribution of $g(t)$ is infinitively divisible (see [7, p. 174] or [13, p. 47] for details).

4. The evolution of the left endpoint

Now we describe the process $(g(t))_{t \in [0, 1/m]}$. We know that its increments are independent and that (8) specifies its marginals. We shall determine its Lévy measure and prove that its mass is finite (see [13] for terminology). This means that the instants when a file arrives at the left of B_0 and joins this data block during its storage do not accumulate before time $1/m$, even if $\bar{\nu}(0) = \infty$ (i.e. files arrive densely near the data block). Proposition 3 of [1] ensures that the first time, T_1 , that 0 is covered, which is also the first jump time of $(g(t))_{t \in [0, 1/m]}$, is uniformly distributed on $[0, 1/m]$. Actually, the second jump time is uniformly distributed on $[T_1, 1/m]$ and so on. More precisely, we have the following result.

Theorem 2. *The jump times of $(g(t))_{t \in [0, 1/m]}$ are given by an increasing sequence $(T_i)_{i \in \mathbb{N}}$ which accumulate at $1/m$. More precisely, using the convention $T_0 = 0$, it holds that, for every $i \geq 1$, conditionally on $T_{i-1} = t$, T_i is independent of $(T_j)_{0 \leq j \leq i-1}$ and is uniformly distributed on $[t, 1/m]$. Then, denoting by $-G_i$ the jump of $(g(t))_{t \in [0, 1/m]}$ at time T_i for every $i \in \mathbb{N}$, we have*

$$g(t) := - \sum_{T_i \leq t} G_i,$$

where $\{(T_i, G_i) : i \in \mathbb{N}\}$ is a PPP on $[0, 1/m) \times \mathbb{R}^+$ with intensity

$$dt \, dx \int_0^\infty P(Y_x^{(t)} \in -dl) \bar{\nu}(l).$$

In other words, $(g(t))_{t \in [0, 1/m]}$ is an additive process and its generating triplet is

$$\left(0, \int_0^t ds \int_0^\infty P(Y_x^{(s)} \in -dl) \bar{\nu}(l), 0 \right).$$

In particular, the interarrival times of $\{T_i : i \in \mathbb{N}\}$ form a ‘continuous, uniform stick-breaking sequence’ (see the residual allocation model in [12, pp. 63–64]): the distribution of $((T_{i+1} - T_i)/m)_{i \in \mathbb{N}}$ is the Griffiths–Engen–McCloskey distribution with parameter $(0, 1)$ (i.e. rearranging these increments into decreasing order yields the Poisson–Dirichlet distribution of parameter $(0, 1)$).

Furthermore, for every $i \in \mathbb{N}$, conditionally on $T_i = t$, the law of G_i is given by

$$P(G_i \in dx) = dx \frac{1 - mt}{m} \int_0^\infty P(Y_x^{(t)} \in -dl) \bar{\nu}(l), \tag{9}$$

and, as a consequence,

$$E(G_i) = \left(\frac{1}{(1 - mt)^2} + \frac{1}{2} \frac{m}{1 - mt} \right) \int_0^\infty l^2 \nu(dl).$$

Example 1. For the basic example, $\nu = \delta_1$, conditionally on $T_i = t$, we have

$$P(G_i \in dx) = (1 - t)e^{-tx} \frac{(tx)^{[x]}}{[x]!} dx,$$

using (5), where $[x] = \sup\{n \in \mathbb{N} : n \leq x\}$.

For the proof, we need the following identity.

Lemma 2. Let $(S_t)_{t \geq 0}$ be a subordinator with no drift and Lévy tail $\bar{\mu}$. Then, for all $(t, x) \in \mathbb{R}_+^2$, we have

$$P(S_t > x) = \int_0^t ds \int_0^x P(S_s \in db) \bar{\mu}(x - b).$$

Proof. As S has no drift, we have, for all $t > 0$ and $x > 0$,

$$S_t > x \iff \text{there exists a unique } s \in (0, t] \text{ such that } S_{s-} \leq x, \Delta S_s > x - S_{s-} \text{ a.s.}$$

We then obtain, also using the compensation formula (see [2, p. 7]),

$$P(S_t > x) = E\left(\sum_{0 < s \leq t} \mathbf{1}_{\{S_{s-} \leq x\}} \mathbf{1}_{\{\Delta S_s > x - S_{s-}\}}\right) = E\left(\int_0^t ds \mathbf{1}_{\{S_s \leq x\}} \bar{\mu}(x - S_s)\right),$$

which completes the proof. We can also give an analytic proof by computing the Laplace transform of the right-hand side for $q > 0$ and using Fubini's theorem:

$$\begin{aligned} & \int_0^\infty dx e^{-qx} \int_0^t ds \int_0^x P(S_s \in db) \bar{\mu}(x - b) \\ &= \int_0^t ds \int_0^\infty \mu(dy) \int_0^\infty P(S_s \in db) \frac{e^{-qb} - e^{-q(b+y)}}{q} \\ &= \int_0^t ds e^{-\phi(q)s} \int_0^\infty \mu(dy) \frac{1 - e^{-qy}}{q} \\ &= \frac{1 - e^{-\phi(q)t}}{\phi(q)} \frac{\phi(q)}{q} \\ &= \int_0^\infty dx e^{-qx} P(S_t > x) \end{aligned}$$

(where ϕ is the Laplace transform of $(S_t)_{t \geq 0}$), which proves the lemma.

We are now able to establish Theorem 2.

Proof of Theorem 2. We know, from Corollary 1, that $(g(t))_{t \in [0, 1/m]}$ is an additive process. Moreover, for every $x \geq 0$, $(Y_x^{(t)} + x)_{t \geq 0}$ is a subordinator with no drift and Lévy measure $x \bar{\nu}$ (see (1)). So Lemma 2 ensures that

$$\begin{aligned} P(Y_x^{(t)} > 0) &= P(Y_x^{(t)} + x > x) \\ &= \int_0^t ds \int_0^x P(Y_x^{(s)} + x \in db) x \bar{\nu}(x - b) \\ &= \int_0^t ds \int_0^\infty P(Y_x^{(s)} \in -dl) x \bar{\nu}(l). \end{aligned}$$

Using (8), we obtain

$$E(e^{\lambda g(t)}) = \exp\left(\int_0^\infty dx(e^{-\lambda x} - 1) \int_0^t ds \int_0^\infty P(Y_x^{(s)} \in -dl)\bar{v}(l)\right).$$

So $(g(t))_{t \in [0, 1/m]}$ is an additive process with generating triplet

$$\left(0, \int_0^t ds \int_0^\infty P(Y_x^{(s)} \in -dl)\bar{v}(l), 0\right),$$

using Definition 8.2 and Theorem 9.8 of [13]. This characterizes the distribution of $(g(t))_{t \in [0, 1/m]}$ (by Theorem 9.8 of [13]) and proves that $\{(T_i, G_i) : i \in \mathbb{N}\}$ is a PPP on $[0, 1/m) \times \mathbb{R}^+$ with intensity $dt dx \int_0^\infty P(Y_x^{(t)} \in -dl)\bar{v}(l)$. We can also compute the distribution of $g(t + s) - g(t)$ using the independence of increments and (8). This proves that $g(\cdot)$ is the sum of jumps given by a PPP.

By projection, $\{T_i : i \in \mathbb{N}\}$ is a PPP on $[0, 1/m)$ with intensity $m(1 - mt)^{-1} dt$. Indeed, for every $t \in [0, 1/m)$,

$$\begin{aligned} \int_0^\infty dx \int_0^\infty P(Y_x^{(t)} \in -dl)\bar{v}(l) &= \int_0^\infty P(\vec{\tau}_l^{(t)} \in dx) \int_0^\infty dl \frac{x}{l} \bar{v}(l) \quad \text{using (7)} \\ &= \int_0^\infty dl \frac{E(\vec{\tau}_l^{(t)})\bar{v}(l)}{l} \\ &= E(\vec{\tau}_1^{(t)}) \int_0^\infty \bar{v}(l) dl \\ &= \frac{m}{1 - mt} \quad \text{using (6)}. \end{aligned}$$

Thus, writing $N_t' := \text{card}\{i \in \mathbb{N} : T_i \in (t, t']\}$, we have $N_0^t < \infty$ a.s. for every $t \in [0, 1/m)$. We can then sort the times T_i , and we have

$$P(T_{i+1} > t' \mid T_i = t) = P(N_t' = 0) = \exp\left(-\int_t^{t'} ds \frac{m}{1 - ms}\right) = \frac{1 - mt'}{1 - mt},$$

meaning that T_{i+1} is uniformly distributed on $[T_i, 1/m]$. The independence is a consequence of the Poissonian property of $\{T_i : i \in \mathbb{N}\}$, and we obtain the theorem.

Finally, this proves (9) and, for every $i \in \mathbb{N}$, conditionally on $T_i = t$, we obtain

$$\begin{aligned} E(G_i) &= \frac{1 - mt}{m} \int_0^\infty dl \frac{E([\vec{\tau}_l^{(t)}]^2)\bar{v}(l)}{l} \quad \text{again using (7)} \\ &= \frac{1 - mt}{m} \int_0^\infty dl \bar{v}(l) \left(l \left(\frac{m}{1 - mt} \right)^2 + \frac{\int_0^\infty l^2 v(dl)}{(1 - mt)^3} \right), \end{aligned}$$

since $[\kappa^{(t)}]'(0)$ is given by (6) and $[\kappa^{(t)}]''(0)$ is given by Proposition 4 of [1].

5. The process of remaining data

We still consider the files which arrive at the left of B_0 , the block containing 0, and which cannot be entirely stored at the left of this block; see Figure 2. Such events occur at the jump

times of $(g(t))_{t \in [0, 1/m]}$, that is, at time T_i . Here we focus on the portions of these files which cannot be stored at the left of B_0 and are shifted to the right of $B_0(T_i-)$ to find a free space. They are called the remaining data and denoted by R_i . Thus, R_i is the quantity of data which arrives at the left of B_0 at time T_i and is stored at the right of B_0 . Then it is also the quantity of data over $g(T_{i-1}-)$ at time T_i (see [1, Section 2.1] for details), and it is given by

$$R_i := Y_{g(T_{i-1}-)}^{(T_i)} - I_{g(T_{i-1}-)}^{(T_i)} \quad \text{for all } i \geq 1.$$

We aim to determine the distribution of $\{(T_i, G_i, R_i) : i \in \mathbb{N}\}$, which is the key to the characterization of the jumps of $(g(t), d(t))_{t \in [0, 1/m]}$. With this in mind, we need to describe the arrival of files which induce the jumps (G_i, R_i) . So we consider the half hardware at the left of $g(t)$, which we turn over, so that it is now identified with \mathbb{R}^+ and its free space is given by $\overleftarrow{\mathcal{R}}(t)$ (see Section 2). The size of the free space and the first free plots of this half hardware are given by the processes $(L_x^{(t)})_{x \geq 0}$ and $(D_x^{(t)})_{x \geq 0}$ defined by

$$L_x^{(t)} = |\overleftarrow{\mathcal{R}}(t) \cap [0, x]|, \quad D_x^{(t)} = \inf\{y > x : y \in \overleftarrow{\mathcal{R}}(t)\},$$

for all $t \in [0, 1/m)$ and for all $x \geq 0$. When at time t , a file of length l arrives at location $-x + g(t-)$ on the hardware (i.e. at location x on the half hardware), it yields a jump of $g(\cdot)$ if the free space $L_x^{(t-)}$ between $-x + g(t-)$ and $g(t-)$ is less than l . Then the quantity of remaining data is $l - L_x^{(t-)}$, and the jump of the left endpoint is $D_x^{(t-)}$; see Figure 2. So we naturally introduce the measure $\rho^{(t)}$ on \mathbb{R}_+^2 defined by

$$\rho^{(t)}(dy dz) := \int_0^\infty dx \int_0^\infty \nu(dl) P(D_x^{(t)} \in dy, l - L_x^{(t)} \in dz).$$

In Lemma 3, below, we give a useful alternative expression for $\rho^{(t)}$. This measure gives the intensity of the point process $\{(T_i, G_i, R_i) : i \in \mathbb{N}\}$, as stated by the following result.

Theorem 3. *The process $\{(T_i, G_i, R_i) : i \in \mathbb{N}\}$ is a PPP on $[0, 1/m) \times \mathbb{R}_+^2$ with intensity $dt \rho^{(t)}(dy dz)$.*

A remarkable consequence of Theorem 3 is that $(R_i)_{i \in \mathbb{N}}$ is an i.i.d. sequence: whereas the rate at which the jumps occur increases as the time approaches $1/m$, the quantity of remaining data maintains the same distribution.

Corollary 2. *The process $\{(T_i, R_i) : i \in \mathbb{N}\}$ is a PPP on $[0, 1/m) \times \mathbb{R}^+$ with intensity $dt dz(\bar{\nu}(z)/(1 - mt))$. In other words, $(R_i)_{i \in \mathbb{N}}$ is i.i.d., independent of $(T_i)_{i \in \mathbb{N}}$, and its distribution is given by*

$$P(R_i \in dz) = m^{-1} \bar{\nu}(z) dz, \quad z \geq 0.$$

Example 2. Using the expression of $\rho^{(t)}$ given in Lemma 3, below, expressions (23) and (24) of [1] yield an expression of $\rho^{(t)}$ for the basic example and the gamma distribution which is quite heavy and not mentioned here. Nonetheless, the quantity of remaining data can often be calculated explicitly. For the basic example, $\nu = \delta_1$, the remaining data are uniform random variables on $[0, 1]$. For the exponential distribution ($\nu(dl) = \mathbf{1}_{\{l \geq 0\}} e^{-l} dl$), the remaining data are also exponentially distributed.

The proofs of these results are organized as follows. Firstly, in Lemma 3, below, we give a more explicit expression of $\rho^{(t)}$ which will be useful for the proofs and will enable us to

derive Corollary 2 from Theorem 3. Secondly, we prove that $\rho^{(t)}$ gives the intensity of the point process $\{(T_i, G_i, R_i) : i \in \mathbb{N}\}$ (Lemma 4). That is, for every $t \in [0, 1/m)$ and $A = (a_1, b_1] \times (a_2, b_2] \subset \mathbb{R}_+^2$, we have

$$\lim_{h \rightarrow 0} \frac{\mathbb{P}(\text{there exists } i \in \mathbb{N} \text{ such that } T_i \in (t, t + h], (G_i, R_i) \in A)}{h} = \rho^{(t)}(A).$$

The lower bound appears naturally by considering the arrival of one single file, independently of the past, which induces a jump of the left endpoint, as described at the beginning of this section (see also Figure 2). However, in the case in which $\bar{v}(0) = \infty$, some jumps of the left endpoint could be due to the successive arrival of many files during a short time interval $(t, t + h]$. Thanks to Theorem 2, we already know the rate at which the jumps occur (i.e. the total intensity). This will give us the upper bound. Finally, we prove that the point process $\{(T_i, G_i, R_i) : i \in \mathbb{N}\}$ satisfies a memoryless property (Lemma 5), which is a direct consequence of the results of Section 3. We then obtain the complete description of this point process, which enables us to prove Theorem 3. Corollary 2 follows by integrating $\rho^{(t)}$ with respect to the first coordinate.

Recall the notation used in Theorem 1 and (4).

Lemma 3. *For every $t \in [0, 1/m)$, the measure $\rho^{(t)}(dy dz)$ can also be expressed as*

$$\begin{aligned} dz \int_z^\infty \nu(dl) & \left(\mathbb{P}(\bar{\tau}_{l-z}^{\leftarrow(t)} \in dy) + \int_0^y \mathbb{P}(\bar{\tau}_{l-z}^{\leftarrow(t)} \in dx)(y-x)\Pi^{(t)}(dy-x) \right) \\ & = \int_z^\infty \nu(dl)(l-z) \left(y^{-1} dy \mathbb{P}(Y_y^{(t)} + l \in dz) \right. \\ & \quad \left. + \int_0^y \mathbb{P}(Y_x^{(t)} + l \in dz)(yx^{-1} - 1)\Pi^{(t)}(dy-x) \right). \end{aligned}$$

Proof. By applying Lemma 1.11 of [4] to $(\bar{\tau}_x^{\leftarrow(t)})_{x \geq 0}$, we have, for all $a, b \geq 0$ and $q > 0$ (t is fixed and omitted in the notation),

$$\int_0^\infty dx e^{-qx} \mathbb{E}(\exp(-bL_x - aD_x)) = \frac{\kappa(a+q) - \kappa(a)}{q(\kappa(a+q) + b)}.$$

Letting q tend to 0, we obtain

$$\int_0^\infty dx \mathbb{E}(\exp(-bL_x - aD_x)) = \frac{\kappa'(a)}{\kappa(a) + b} = \int_0^\infty dz e^{-bz} \kappa'(a) e^{-\kappa(a)z}.$$

From $\kappa'(a) = \int_0^\infty e^{-ay}(\delta_0(dy) + y\Pi(dy))$ and $e^{-\kappa(a)z} = \int_0^\infty e^{-ay} \mathbb{P}(\bar{\tau}_z \in dy)$, we deduce that

$$\int_0^\infty dx \mathbb{E}(\exp(-bL_x - aD_x)) = \int_0^\infty dz \int_0^\infty \gamma_z(dy) e^{-bz-ay}, \tag{10}$$

where γ_z is the convolution of $\delta_0(dy) + y\Pi(dy)$ and $\mathbb{P}(\bar{\tau}_z \in dy)$. Thus,

$$\begin{aligned} \gamma_z(dy) & = \int_0^y \mathbb{P}(\bar{\tau}_z \in dx)(\delta_0(dy-x) + (y-x)\Pi(dy-x)) \\ & = \mathbb{P}(\bar{\tau}_z \in dy) + \int_0^y \mathbb{P}(\bar{\tau}_z \in dx)(y-x)\Pi(dy-x). \end{aligned}$$

And the identification of Laplace transforms in (10) entails that

$$\int_0^\infty dx P(L_x \in dz, D_x \in dy) = dz \left(P(\overleftarrow{\tau}_z \in dy) + \int_0^y P(\overleftarrow{\tau}_z \in dx)(y-x)\Pi(dy-x) \right), \tag{11}$$

which, integrating with respect to l , proves the first expression of $\rho^{(t)}$. Using (7) gives the second identity.

Remark 2. A recent work by Winkel [14, Theorem 1] enables us to calculate differently the law of $P(L_x \in dz, D_x \in dy)$ (L_x corresponds to T_x in [14] and D_x to $X(T_{x-}) + \Delta_x$):

$$\int_0^\infty dx P(L_x \in dz, D_x \in dy) = dy P(H_y \in dz) + dz \int_0^\infty P(\overleftarrow{\tau}_x \in dx)(y-x)\Pi(dy-x),$$

where $H_x = \inf\{a \geq 0, \overleftarrow{\tau}_a = x\}$. Then observe that the measures on \mathbb{R}_+^2 , $dy P(H_y \in dz)$ and $dz P(\overleftarrow{\tau}_z \in dy)$, coincide by computing their Laplace transform using Equation (4) of [14]. This proves (11).

For every Borel set B of $[0, 1/m) \times \mathbb{R}_+^2$, we define

$$N_B := \text{card}\{i \in \mathbb{N} : (T_i, G_i, R_i) \in B\},$$

and we say that A is a rectangle of $D \subset \mathbb{R}^d$ if A is a subset of D of the form

$$\{x = (x_1, x_2, \dots, x_d), a_1 < x_1 \leq b_1, \dots, a_d < x_d \leq b_d\}.$$

Then, we have the following result.

Lemma 4. For all $t \in [0, 1/m)$ and for every rectangle A of \mathbb{R}_+^2 , we have

$$\lim_{h \rightarrow 0} \frac{P(N_{(t,t+h] \times A} \geq 1)}{h} = \rho^{(t)}(A).$$

Proof. Firstly, we prove the lower bound. Secondly, we check that the convergence holds for $A = \mathbb{R}_+^2$.

Let $\varepsilon > 0$, and let $A = (a, b] \times (c, d]$, and work conditionally on $\overleftarrow{\mathcal{R}}(t)$. We consider a file labeled i which arrives at time $t_i \in (t, t+h]$ at location $x_i < g(t)$. We set $\tilde{x}_i := g(t) - x_i \geq 0$, the arrival point on the half line at the left of $g(t)$, and require that

$$\begin{aligned} l_i - L_{\tilde{x}_i}^{(t)} &\in (c, d - \varepsilon], & D_{\tilde{x}_i}^{(t)} &\in (a, b - \varepsilon], \\ |L_{\tilde{x}_i}^{(t_i-)} - L_{\tilde{x}_i}^{(t)}| &\leq \varepsilon, & \text{and} & \quad |D_b^{(t_i-)} - D_b^{(t)}| \leq \varepsilon. \end{aligned}$$

Then file i verifies that

$$l_i - L_{\tilde{x}_i}^{(t_i-)} \in (c, d] \quad \text{and} \quad D_{\tilde{x}_i}^{(t_i-)} \in (a, b].$$

So this file induces a jump of the left endpoint and $N_{(t,t+h] \times A} \geq 1$ (see the beginning of this section or Figure 2 for details), and we obtain the lower bound

$$\begin{aligned} &P(N_{(t,t+h] \times A} \geq 1 \mid \overleftarrow{\mathcal{R}}(t)) \\ &\geq P(\text{there exists } i \in \mathbb{N} \text{ such that } t_i \in (t, t+h], l_i - L_{\tilde{x}_i}^{(t)} \in (c, d - \varepsilon], \\ &\quad D_{\tilde{x}_i}^{(t)} \in (a, b - \varepsilon], |L_{\tilde{x}_i}^{(t_i-)} - L_{\tilde{x}_i}^{(t)}| \leq \varepsilon, |D_b^{(t_i-)} - D_b^{(t)}| \leq \varepsilon \mid \overleftarrow{\mathcal{R}}(t)) \\ &\geq A_t(h)B_t(h), \end{aligned} \tag{12}$$

where

$$A_t(h) := \mathbb{P}(\text{there exists } i \in \mathbb{N} \text{ such that } t_i \in (t, t + h], l_i - L_{\tilde{x}_i}^{(t)} \in (c, d - \varepsilon], D_{\tilde{x}_i}^{(t)} \in (a, b - \varepsilon] \mid \overleftarrow{\mathcal{R}}(t)),$$

$$B_t(h) := \mathbb{P}\left(\sup_{t' \in [t, t+h]} \{|L_b^{(t')} - L_b^{(t)}|\} \leq \varepsilon, \sup_{t' \in [t, t+h]} \{|D_b^{(t')} - D_b^{(t)}|\} \leq \varepsilon \mid \overleftarrow{\mathcal{R}}(t)\right).$$

- (i) By Theorem 2, $\mathbb{P}(N_t^{t+h} \neq 0) \rightarrow 0$ as $h \rightarrow 0$, so a.s. for small enough h , $g(t+h) = g(t)$. Then, using the Hausdorff metric on \mathbb{R}_+ (denoted by $\mathcal{H}(\mathbb{R}_+)$ in Section 2 of [1]), we have

$$\overleftarrow{\mathcal{R}}(t+h) \rightarrow \overleftarrow{\mathcal{R}}(t) \quad \text{as } h \rightarrow 0, \text{ a.s.}$$

Then $B_t(h)$ converges a.s. to 1 as h tends to 0.

- (ii) As $\{(t_i, \tilde{x}_i, l_i) : i \in \mathbb{N}, t_i \in (t, t + h], x_i < g(t)\}$ is a PPP on $(t, t + h] \times \mathbb{R}_+^2$ with intensity $dt \otimes dx \otimes \nu(dl)$ independent of $\overleftarrow{\mathcal{R}}(t)$, we have

$$A_t(h) = 1 - \exp\left(-h \int_0^\infty dx \int_0^\infty \nu(dl) \mathbf{1}_{\{l - L_x^{(t)} \in (c, d - \varepsilon], D_x^{(t)} \in (a, b - \varepsilon]\}}\right) \quad \text{a.s.}$$

This term is a.s. equivalent, when h tends to 0, to

$$h \int_0^\infty dx \int_0^\infty \nu(dl) \mathbf{1}_{\{l - L_x^{(t)} \in (c, d - \varepsilon], D_x^{(t)} \in (a, b - \varepsilon]\}}.$$

Then, letting h tend to 0 in (12), (i) and (ii) give

$$\liminf_{h \rightarrow 0} \frac{\mathbb{P}(N_{(t, t+h] \times A} \geq 1 \mid \overleftarrow{\mathcal{R}}(t))}{h} \geq \int_0^\infty dx \int_0^\infty \nu(dl) \mathbf{1}_{\{l - L_x^{(t)} \in (c, d - \varepsilon], D_x^{(t)} \in (a, b - \varepsilon]\}} \quad \text{a.s.}$$

Integrating this inequality and using Fatou’s lemma, we obtain

$$\liminf_{h \rightarrow 0} \frac{\mathbb{P}(N_{(t, t+h] \times A} \geq 1)}{h} \geq \mathbb{E}\left(\int_0^\infty dx \int_0^\infty \nu(dl) \mathbf{1}_{\{l - L_x^{(t)} \in (c, d - \varepsilon], D_x^{(t)} \in (a, b - \varepsilon]\}}\right) \geq \rho^{(t)}((a, b - \varepsilon] \times (c, d - \varepsilon]).$$

As $\rho^{(t)}((a, b] \times \{d\} \cup \{b\} \times (c, d]) = 0$ (use the two equalities of Lemma 3), we obtain, letting ε tend to 0,

$$\liminf_{h \rightarrow 0} \frac{\mathbb{P}(N_{(t, t+h] \times A} \geq 1)}{h} \geq \rho^{(t)}(A).$$

We derive the upper bound from Theorem 2. Firstly,

$$\frac{\mathbb{P}(N_{(t, t+h] \times \mathbb{R}_+^2} \geq 1)}{h} = \frac{\mathbb{P}(\text{there exists } i \in \mathbb{N} \text{ such that } T_i \in (t, t + h])}{h} \rightarrow \frac{m}{1 - mt} \quad \text{as } h \rightarrow 0,$$

and identity (14), below, gives

$$\rho^{(t)}(\mathbb{R}_+^2) = \frac{m}{1 - mt}.$$

So we just need to prove the following result. Let $(\mu_n)_{n \in \mathbb{N}}$ and μ be finite measures on \mathbb{R}_+^2 such that, for every rectangle A of \mathbb{R}_+^2 , $\liminf_{n \rightarrow \infty} \mu_n(A) \geq \mu(A)$ and $\lim_{n \rightarrow \infty} \mu_n(\mathbb{R}_+^2) = \mu(\mathbb{R}_+^2)$. Then, for every rectangle A of \mathbb{R}_+^2 , $\lim_{n \rightarrow \infty} \mu_n(A) = \mu(A)$. With this in mind, suppose that there exist a rectangle A , $\varepsilon > 0$, and a sequence of integers k_n such that $\mu_{k_n}(A) \geq \mu(A) + \varepsilon$. Choose B to be the union of disjoint rectangles all disjoint from A such that $\mu(B \cup A) \geq \mu(\mathbb{R}_+^2) - \varepsilon/2$. Then,

$$\liminf_{n \rightarrow \infty} \mu_{k_n}(\mathbb{R}_+^2) \geq \liminf_{n \rightarrow \infty} \mu_{k_n}(A \cup B) \geq \mu(A) + \varepsilon + \mu(B) \geq \mu(\mathbb{R}_+^2) + \frac{\varepsilon}{2},$$

which is in contradiction with $\lim_{n \rightarrow \infty} \mu_n(\mathbb{R}_+^2) = \mu(\mathbb{R}_+^2)$.

To prove Theorem 3, it remains to prove the absence of memory.

Lemma 5. *Let $t \in [0, 1/m)$, then $\{(T_i, G_i, R_i) : i \in \mathbb{N}, T_i \leq t\}$ is independent of $\{(T_i, G_i, R_i) : i \in \mathbb{N}, T_i > t\}$.*

Proof. First $\{(T_i, G_i, R_i) : T_i \leq t\}$ is given by $\{(t_i, l_i, x_i) : t_i \leq t, x_i \in [g(t), d(t)]\}$. Moreover, $\{(T_i, G_i, R_i) : T_i > t\}$ depends on $(\mathcal{R}(t) - g(t)) \cap (-\infty, 0]$ and $\{(t_i, x_i - g(t), l_i) : t_i > t, x_i < g(t)\}$, which are independent. Moreover, $(\mathcal{R}(t) - g(t)) \cap (-\infty, 0]$ is independent of $\{(t_i, l_i, x_i) : t_i \leq t, x_i \in [g(t), d(t)]\}$ by Lemma 1 and so is $\{(t_i, x_i - g(t), l_i) : t_i > t, x_i < g(t)\}$ by the Poissonian property. This completes the proof.

We can now prove Theorem 3 and Corollary 2.

Proof of Theorem 3. We now prove that, for every B finite union of disjoint rectangles of $[0, 1/m) \times \mathbb{R}_+^2$,

$$P(N_B = 0) = e^{-\gamma(B)}, \tag{13}$$

where $\gamma(dt \, dy \, dz) = dt \rho^{(t)}(dy \, dz)$. As γ is nonatomic (use Lemma 3), this will ensure that $\{(T_i, G_i, R_i) : i \in \mathbb{N}\}$ is a PPP with intensity γ (using Rényi's theorem [10]).

Let $t \in [0, 1/m)$, and let A be a finite union of rectangles of \mathbb{R}_+^2 . We consider $H(s) := P(N_{(t,t+s] \times A} = 0)$ for $s \in [0, 1/m - t)$. Lemma 5 entails that

$$H(s + h) = P(N_{(t,t+s] \times A} = 0) P(N_{(t+s,t+s+h] \times A} = 0) = H(s) P(N_{(t+s,t+s+h] \times A} = 0).$$

We write $A = \sqcup_{i=1}^N A_i$, where A_i is a rectangle of \mathbb{R}_+^2 . Theorem 2 and Lemma 4 respectively ensure that, for all $1 \leq i, j \leq N$ such that $i \neq j$,

$$\lim_{h \rightarrow 0} \frac{P(N_{(t,t+h] \times A_i} \geq 1, N_{(t,t+h] \times A_j} \geq 1)}{h} = 0 \quad \text{and} \quad \lim_{h \rightarrow 0} \frac{P(N_{(t,t+h] \times A_i} \geq 1)}{h} = \rho^{(t)}(A_i).$$

Then

$$\lim_{h \rightarrow 0} \frac{P(N_{(t,t+h] \times A} \geq 1)}{h} = \sum_{i=1}^N \lim_{h \rightarrow 0} \frac{P(N_{(t,t+h] \times A_i} \geq 1)}{h} = \rho^{(t)}(A),$$

and the derivative of H is given by

$$\lim_{h \rightarrow 0} \frac{H(s + h) - H(s)}{h} = H(s) \lim_{h \rightarrow 0} \frac{1 - P(N_{(t+s,t+s+h] \times A} = 0)}{h} = H(s) \rho^{(t+s)}(A).$$

Thus, $H(s)$ satisfies a differential equation of order 1, and we obtain (13) for $B = (t, t + s] \times A$.

$$H(s) = \exp\left(-\int_0^s du \rho^{(t+u)}(A)\right) = \exp\left(-\int_t^{t+s} du \rho^{(u)}(A)\right) = e^{-\gamma((t,t+s] \times A)}.$$

Again, using Lemma 5 and the additivity of measures proves (13) for every B finite union of rectangles of $[0, 1/m) \times \mathbb{R}^+ \times \mathbb{R}^+$.

Proof of Corollary 2. As a projection of the PPP $\{(T_i, G_i, R_i) : i \in \mathbb{N}\}$, $\{(T_i, R_i) : i \in \mathbb{N}\}$ is a PPP with intensity $dt \int_{y \in [0, \infty]} \rho^{(t)}(dy dz)$. By Lemma 3 we have

$$\begin{aligned} \int_{y \in [0, \infty]} \rho^{(t)}(dy dz) &= dz \left(\bar{v}(z) + \int_z^\infty v(dl) \int_0^\infty P(\tau_{l-z}^{\leftarrow} \in dx) \int_x^\infty \Pi^{(t)}(dy - x)(y - x) \right) \\ &= dz \bar{v}(z) \left(1 + \int_0^\infty \Pi(dy)y \right) \\ &= dz \frac{\bar{v}(z)}{1 - mt} \quad \text{by (6),} \end{aligned} \tag{14}$$

which gives the intensity of $\{(T_i, R_i) : i \in \mathbb{N}\}$. In other words, $(R_i)_{i \in \mathbb{N}}$ is an i.i.d. sequence independent of $(T_i)_{i \in \mathbb{N}}$ such that $P(R_i \in dz) = m^{-1} \bar{v}(z) dz, z \geq 0$.

6. The evolution of B_0

The processes $(g(t))_{t \in [0, 1/m]}$ and $(d(t))_{t \in [0, 1/m]}$ of the left endpoint and the right endpoint of B_0 have quite a different evolution, even though their one-dimensional distributions coincide. The process $(d(t))_{t \in [0, 1/m]}$ jumps each time $(g(t))_{t \in [0, 1/m]}$ jumps and each time a file arrives at B_0 . More precisely, there are two kinds of jumps of $(B_0(t))_{t \in [0, 1/m]}$ corresponding to the following types of file arrival.

1. Files which arrive at the left of B_0 and cannot be entirely stored at its left (recall Section 5). These files induce the jumps $(-G_i, D_i)$ of the endpoints of B_0 at time T_i independently of the past (see Figure 2).
2. Files which arrive at B_0 . These files induce jumps of the right endpoint $d(\cdot)$ only, with total rate equal to $l(t)\bar{v}(0)$ (see Figure 3). This rate is infinite when $\bar{v}(0) = \infty$. Also observe that the jumps depend from the past of B_0 through the value of the length, $l(t)$. Note that a file which arrives at the left of $B_0(t-)$ at time t with remaining data of size R induces the same jump of the right endpoint as a file of size R which arrives at $B_0(t-)$ at time t . Obviously, the other files (files which are entirely stored at the left of B_0 or which arrive at the right of B_0) do not yield a jump of B_0 .

Thus, we define

$$D_i := d(T_i) - d(T_i^-),$$

and we decompose the process $(g(t), d(t))_{t \in [0, 1/m]}$ into two processes $(C^1(t))_{t \in [0, 1/m]}$ and $(C^2(t))_{t \in [0, 1/m]}$, which give the variation of the endpoints of B_0 at times $(T_i)_{i \in \mathbb{N}}$ (due to the arrival of a file at the left of $g(t)$) and between successive times $(T_i)_{i \in \mathbb{N}}$ (due to the arrival of

files on $B_0(t)$, respectively. That is, for every $t \in [0, 1/m)$,

$$C^1(t) := \sum_{T_i \leq t} (-G_i, D_i), \quad C^2(t) := \left(0, \sum_{\substack{0 \leq s \leq t \\ s \notin \{T_i : i \in \mathbb{N}\}}} \Delta d(s) \right),$$

$$(g(t), d(t)) = C^1(t) + C^2(t).$$

Firstly, we specify the distribution of $(C^1(t))_{t \in [0, 1/m]}$ (see below for the proofs).

Proposition 3. *The point process $\{(T_i, G_i, D_i) : i \in \mathbb{N}\}$ is a PPP on $[0, 1/m] \times \mathbb{R}_+^2$ with intensity $dt \mu^{(t)}(dy dx)$, where*

$$\mu^{(t)}(dy dx) = \int_0^\infty \rho^{(t)}(dy dz) P(\vec{\tau}_z \in dx).$$

We can now specify the distribution of the process $(g(t), d(t))_{t \in [0, 1/m]}$ as follows.

Theorem 4. *The process $(g(t), d(t))_{t \in [0, 1/m]}$ is a pure-jump Markov process equal to $(C^1(t) + C^2(t))_{t \in [0, 1/m]}$ such that, for all $0 \leq t \leq t + s \leq 1/m$, the following conditions hold.*

- (i) *The process $C^1(t + s) - C^1(t)$ is independent of $(g(u), d(u))_{u \in [0, t]}$.*
- (ii) *Conditionally on $l(t) = l$, $C^2(t + s) - C^2(t)$ is independent of $(g(u), d(u))_{u \in [0, t]}$. Also conditionally on $T_i \leq t \leq t + s < T_{i+1}$, for some $i \in \mathbb{N}$,*

$$C^2(t + s) - C^2(t) \stackrel{D}{=} (0, \vec{\tau}_{S_t^{(t+s)}}),$$

where $\stackrel{D}{=}$ denotes equality in distribution, and $(S_x)_{x \geq 0}$ is a subordinator with no drift and Lévy measure ν , which is independent of $(\vec{\tau}_x)_{x \geq 0}$.

Recalling that vague convergence of measures on A is the convergence of the integrals of measure against continuous functions with compact support in A , the jump rate of $(g(t), d(t))_{t \in [0, 1/m]}$ is then given by the following result.

Corollary 3. *If $t \in [0, 1/m)$, we have the following vague convergence of measures on $[0, \infty) \times (0, \infty)$ when h tends to 0:*

$$h^{-1} P(g(t) - g(t + h) \in dy, d(t + h) - d(t) \in dx \mid l(t) = l)$$

$$\xrightarrow{\nu} \mu^{(t)}(dy dx) + l \delta_0(dy) \int_0^\infty \nu(dz) P(\vec{\tau}_z \in dx).$$

We begin with two lemmas which state the independences needed for the proofs.

Lemma 6. *The process $\{(T_i, G_i, D_i) : i \in \mathbb{N}, T_i > t\}$ is independent of $(g(u), d(u))_{u \in [0, t]}$.*

Proof. Using (15), below, we see that $\{(T_i, G_i, D_i) : i \in \mathbb{N}, T_i > t\}$ is given by

$$\{(T_i, G_i, R_i) : i \in \mathbb{N}, T_i > t\} \quad \text{and} \quad (\vec{\mathcal{R}}(s))_{s > t}.$$

These quantities depend on the past through $(\overleftarrow{\mathcal{R}}(t), \overrightarrow{\mathcal{R}}(t))$, which is independent of

$$(g(u), d(u))_{u \in [0, t]}$$

by Proposition 2.

Lemma 7. *Let $i \in \mathbb{N}$, and let $0 \leq t' < t \leq 1/m$. Conditionally on $T_{i-1} = t'$ and $T_i = t$, $(\overrightarrow{\mathcal{R}}(u))_{u \in [t', t]}$ is independent of the point process $P_{g(t')}(t)$.*

Proof. Conditioning on $T_{i-1} = t'$ and $T_i = t$ ensures that all the data which arrived at the left of $g(t')$ during the time interval $[t', t]$ is stored at the left of $g(t')$. So $(\overrightarrow{\mathcal{R}}(u))_{u \in [t', t]}$ depends only on the point process $P_{g(t')}(t) \cup P^{d(t')}(t)$, which is independent of $P_{g(t')}(t)$ by Lemma 1.

Proof of Proposition 3. At time T_i , the quantity of remaining data R_i is stored at the right of $B_0(T_i-)$. It induces a jump $D_i = d(T_i) - d(T_i-)$ of the right endpoint which is equal to R_i plus the sum of the lengths of blocks at the right of $B_0(T_i-)$, which are reached during the storage of these data (see Figure 2). More precisely,

$$\begin{aligned} D_i &= \inf\{x \geq 0, |\mathcal{R}(T_i-) \cap [d(t), d(t) + x]| = R_i\} \\ &= \inf\{x \geq 0, |\overrightarrow{\mathcal{R}}(T_i-) \cap [0, x]| = R_i\} \\ &= \overrightarrow{\tau}_{R_i}^{(T_i-)}, \end{aligned} \tag{15}$$

by the definition of $\overrightarrow{\tau}$ (see Section 2). Lemma 7 ensures that, conditionally on $T_i = t$, $(\overrightarrow{\tau}_x^{(T_i-)})_{x \geq 0}$ is independent of (G_i, R_i) and distributed as $(\overrightarrow{\tau}_x^{(t)})_{x \geq 0}$. Then denoting by μ_t the law of (G_i, D_i) , conditioned on $T_i = t$, we have

$$\mu_t(dy dx) = P(G_t \in dy, \overrightarrow{\tau}_{R_t}^{(t)} \in dx), \tag{16}$$

where (G_t, R_t) is a random variable independent of $(\overrightarrow{\tau}_x^{(t)})_{x \geq 0}$ and distributed as (G_i, R_i) , conditioned on $T_i = t$.

By Lemma 6, $\{(T_i, G_i, D_i) : i \in \mathbb{N}, T_i > t\}$ is independent of $\{(T_i, G_i, D_i) : i \in \mathbb{N}, T_i \leq t\}$. Then, conditionally on $(T_i)_{i \in \mathbb{N}}$, $(G_i, D_i)_{i \in \mathbb{N}}$ are independent. Adding that $\{T_i : i \in \mathbb{N}\}$ is a PPP on $[0, 1/m]$ with intensity $dtm/(1 - mt)$ ensures that $\{(T_i, G_i, D_i) : i \in \mathbb{N}\}$ is a (marked) PPP with intensity

$$\frac{m}{1 - mt} dt \mu_t(dy dx).$$

Furthermore, by (16), this intensity is equal to

$$dt \int_0^\infty P(\overrightarrow{\tau}_z^{(t)} \in dx) \frac{m}{1 - mt} P(G_t \in dy, R_t \in dz) = dt \int_0^\infty P(\overrightarrow{\tau}_z^{(t)} \in dx) \rho^{(t)}(dy dz),$$

using Theorem 3. This completes the proof.

Proof of Theorem 4. (i) Thanks to Lemma 6, $C^1(t + s) - C^1(t)$ is independent of $(g(u), d(u))_{u \in [0, t]}$.

(ii) We condition on $T_i \leq t \leq t+s < T_{i+1}$ for some $i \in \mathbb{N}$ and $l(t) = l$. Then $g(t+s) - g(t) = 0$ and no data which arrived at the left of $B_0(t)$ during the time interval $(t, t + s]$ is stored at the right of this block. So the increment $d(t + s) - d(t)$ is caused by files arriving at $B_0(t)$: they are stored at the right of $B_0(t)$ and may join data already stored. Note that we can change the order of the arrival of files between t and $t + s$ (using (3)). Thus, we first store the files which arrive at the right of $d(t)$ between times t and $t + s$, then the files which arrive at $B_0(t)$ between times t and $t + s$, and we ignore the files which arrive at the left of $g(t)$.

Step 1: At time t , we consider the half hardware at the right of $d(t)$ which we identify with $[0, \infty)$. Its free space is equal to $\overrightarrow{\mathcal{R}}(t)$. We store the files $i \in \{i \in \mathbb{N} : t_i \in (t, t + s), x_i > d(t)\}$ on the half hardware $[0, \infty)$ at location $x_i - d(t)$ following the process described in the introduction (the size of file i is still l_i). Following Section 2.1 of [1], we obtain the counterpart of the characterization of the free space, (3). That is, the new free space of the half hardware is equal to $\{x \geq 0 : \tilde{Y}_x = \tilde{I}_x\}$, where, for every $x \geq 0$,

$$\tilde{Y}_x = -x + \sum_{\substack{0 \leq l_i \leq t+s \\ d(t) \leq x_i \leq d(t)+x}} l_i, \quad \tilde{I}_x := \inf\{\tilde{Y}_y : 0 \leq y \leq x\}.$$

Using Lemma 1, we see that $\{(t_i, x_i - d(t), l_i) : x_i \geq d(t)\}$ is a PPP on \mathbb{R}^{+3} with intensity $dt \otimes dx \otimes \nu(dl)$. Then,

$$(\tilde{Y}_x)_{x \geq 0} \stackrel{D}{=} (Y_x^{(t+s)})_{x \geq 0}$$

is a Lévy process with Laplace exponent $\Psi^{(t+s)}$. As $[\Psi^{(t+s)}]'(0) < 0$, $(\tilde{Y}_x)_{x \geq 0}$ is regular for $(-\infty, 0)$, in the sense that it takes negative values for some arbitrarily small x [2, Proposition 8]. So, for every stopping time T such that $\tilde{Y}_T = \tilde{I}_T$, there is the identity $T = \inf\{z \geq 0 : \tilde{Y}_z < \tilde{Y}_T\}$. This ensures that the free space $\{x \geq 0 : \tilde{Y}_x = \tilde{I}_x\}$ of the half hardware is the range of $(\tilde{\tau}_x)_{x \geq 0}$ defined by

$$\tilde{\tau}_x := \inf\{z \geq 0 : \tilde{Y}_z < -x\}.$$

By Theorem 1 of [2], $(\tilde{\tau}_x)_{x \geq 0}$ is a subordinator with Laplace exponent $\kappa^{(t+s)}$, which is the inverse function of $-\Psi^{(t+s)}$. So $(\tilde{\tau}_x)_{x \geq 0}$ is distributed as $(\tau_x^{\rightarrow(t+s)})_{x \geq 0}$. By Lemma 1 again, $\{(t_i, x_i - d(t), l_i) : x_i > d(t)\}$ is independent of $(g(u), d(u))_{u \in [0, t]}$. So $(\tilde{\tau}_x)_{x \geq 0}$ is independent of $(g(u), d(u))_{u \in [0, t]}$.

Step 2: To obtain the covering $\mathcal{C}(t + s)$, we now store the files $\{i : t_i \in (t, t + s), x_i \in [g(t), d(t))\}$. This amounts to storing these files in the first free spaces (i.e. as much on the left as possible) of the half hardware considered above, whose free space is the range of $(\tilde{\tau}_x)_{x \geq 0}$. The variation of the right endpoint is equal to the sum of the sizes of these files, say S_t^{t+s} , plus the sizes of the lengths of the blocks of the half hardware joined during their storage. That is, as for (15),

$$C^2(t + s) - C^2(t) = (0, \tilde{\tau}_{S_t^{t+s}}), \quad \text{where } S_t^{t+s} := \sum_{\substack{t < l_i \leq t+s \\ x_i \in [g(t), d(t))}} l_i.$$

Conditionally on $l(t) = l$, by the Poissonian property, $S_t^{t+s} \stackrel{D}{=} S_{S_l}$. Adding that S_t^{t+s} is independent of $(\tilde{\tau}_x)_{x \geq 0}$ gives the law of $C^2(t + s) - C^2(t)$. As $(\tilde{\tau}_x)_{x \geq 0}$ and S_t^{t+s} are independent of $(g(u), d(u))_{u \in [0, t]}$, so is $C^2(t + s) - C^2(t)$.

These properties ensure that $(g(t), d(t))_{t \in [0, 1/m]}$ is a Markov process.

To prove Corollary 3, we need the following result which uses the notation of Theorem 4.

Lemma 8. *We have the following vague convergence of measure on $(0, \infty)$:*

$$h^{-1} P(\vec{\tau}_{S_{hl}}^{(t)} \in dx) \xrightarrow{v} l \int_0^\infty \nu(dz) P(\vec{\tau}_z^{(t)} \in dx).$$

Proof. Denoting by ϕ the Laplace exponent of $(S_x)_{x \geq 0}$, $(\vec{\tau}_{S_{xl}}^{(t)})_{x \geq 0}$ is a subordinator of Laplace exponent $l\phi \circ \kappa^{(t)}$ (see (2)). Moreover, for every $\lambda \geq 0$, $\phi(\lambda) = \int_0^\infty (1 - e^{-\lambda y}) \nu(dy)$, which entails that

$$\begin{aligned} \phi \circ \kappa^{(t)}(\lambda) &= \int_0^\infty (1 - e^{-z\kappa^{(t)}(\lambda)}) \nu(dz) \\ &= \int_0^\infty E(1 - e^{-\lambda \vec{\tau}_z^{(t)}}) \nu(dz) \\ &= \int_0^\infty (1 - e^{-\lambda x}) \int_0^\infty \nu(dz) P(\vec{\tau}_z^{(t)} \in dx). \end{aligned}$$

Then $(\vec{\tau}_{S_{xl}}^{(t)})_{x \geq 0}$ is a subordinator with no drift and Lévy measure

$$l \int_0^\infty \nu(dz) P(\vec{\tau}_z^{(t)} \in dx).$$

Using Exercise 1 of [2, Chapter I] or [3, p. 8] completes the proof.

Proof of Corollary 3. We first consider the case when the increment of the left endpoint is 0.

Using Theorem 4 and recalling that $N_t^{t+h} \stackrel{\text{def}}{=} N_{(t+t+h) \times \mathbb{R}_+^2} = \text{card}\{i \in \mathbb{N} : T_i \in (t, t+h]\}$, we have, for all $c > 0$ such that $\int_0^\infty \nu(dz) P(\vec{\tau}_z = c) = 0$,

$$P(g(t+h) - g(t) = 0, d(t+h) - d(t) \geq c \mid l(t) = l) = P(N_t^{t+h} = 0) P(\vec{\tau}_{S_{hl}}^{(t)} \geq c).$$

Adding that $P(N_t^{t+h} = 0) \rightarrow 1$ as $h \rightarrow 0$ and using Lemma 8 gives

$$\begin{aligned} h^{-1} P(g(t+h) - g(t) = 0, d(t+h) - d(t) \geq c \mid l(t) = l) \\ \rightarrow l \int_0^\infty \nu(dz) P(\vec{\tau}_z^{(t)} \geq c) \quad \text{as } h \rightarrow 0. \end{aligned} \tag{17}$$

Let $a, b > 0$, and write

$$P(t, t+h) = P(g(t) - g(t+h) \geq a, d(t+h) - d(t) \geq b \mid l(t) = l).$$

By Proposition 3, $\{(T_i, G_i, D_i) : i \in \mathbb{N}\}$ is a PPP on $[0, 1/m] \times \mathbb{R}_+^2$ with intensity $dt \mu^{(t)}(dy dx)$. The latter verifies that $P(N_t^{t+h} > 1) = o(h)$ ($h \rightarrow 0$), so we have

$$h^{-1} P(C^1(t+h) - C^1(t) \in (-\infty, -a] \times [b, \infty]) \rightarrow \mu^{(t)}([a, \infty) \times [b, \infty]) \quad \text{as } h \rightarrow 0. \tag{18}$$

We can now prove that

$$\lim_{h \rightarrow 0} h^{-1} P(t, t+h) = \mu^{(t)}([a, \infty) \times [b, \infty)). \tag{19}$$

First we give the lower bound:

$$P(t, t + h) \geq P(C^1(t + h) - C^1(t) \in (-\infty, -a] \times [b, \infty] \mid l(t) = l).$$

Using the fact that $C^1(t + h) - C^1(t)$ is independent of $l(t)$ and (18), we obtain

$$\liminf_{h \rightarrow 0} h^{-1} P(t, t + h) \geq \mu^{(t)}([a, \infty) \times [b, \infty]).$$

For the upper bound, observe that

$$P(t, t + h) \leq P(C^1(t + h) - C^1(t) \in (-\infty, -a] \times [b - \varepsilon, \infty] \mid l(t) = l) + P(N_t^{t+h} \geq 1, C^2(t + h) - C^2(t) \in \{0\} \times [\varepsilon, \infty) \mid l(t) = l).$$

Again using the fact that $C^1(t + h) - C^1(t)$ is independent of $l(t)$, with (18) and Theorem 4, we obtain

$$\limsup_{h \rightarrow 0} h^{-1} P(t, t + h) \leq \mu^{(t)}([a, \infty) \times [b - \varepsilon, \infty]).$$

Letting ε tend to 0 gives the upper bound:

$$\limsup_{h \rightarrow 0} h^{-1} P(t, t + h) \leq \mu^{(t)}([a, \infty) \times [b, \infty)).$$

The two limits in (17) and (19) ensure the convergence of measures for sets of the form $\{0\} \times [c, d)$ (with $c > 0$) and $[a, b) \times [c, d)$ (with $a > 0$), which completes the proof.

7. The evolution of the right endpoint and of the length

Proposition 3, Theorem 4, and Corollary 3 give, by projection, the following result.

Corollary 4. *The process $(d(t))_{t \in [0, 1/m)}$ is a jump process satisfying the following conditions.*

(i) *The process $\{(T_i, D_i) : i \in \mathbb{N}\}$ is a PPP on $[0, 1/m) \times \mathbb{R}^+$ with intensity*

$$\frac{dt \int_{z \in [0, \infty)} dz \bar{\nu}(z) P(\bar{\tau}_z^{\rightarrow(t)} \in dx)}{1 - mt},$$

and $\{(T_i, D_i) : i \in \mathbb{N}, T_i > t\}$ is independent of $(d(u))_{u \in [0, t]}$.

(ii) *For all $0 \leq t \leq t + s < 1/m$, conditionally on $l(t) = l$, $d(t + s) - d(t)$ is independent of $(d(u))_{u \in [0, t]}$. Also conditionally on $T_i \leq t \leq t + s < T_{i+1}$ for some $i \in \mathbb{N}$,*

$$d(t + s) - d(t) \stackrel{D}{=} \bar{\tau}_{S_t}^{\rightarrow(t+s)},$$

where $(S_x)_{x \geq 0}$ is a subordinator with no drift and Lévy measure ν , which is independent of $(\bar{\tau}_x^{\rightarrow(t+s)})_{x \geq 0}$.

The jump rate of $(d(t))_{t \in [0, 1/m)}$ is given by the following vague convergence of measures on $(0, \infty)$ for h tending to 0:

$$\frac{P(d(t + h) - d(t) \in dx \mid l(t) = l)}{h} \xrightarrow{\nu} \frac{\int_0^\infty dz \bar{\nu}(z) P(\bar{\tau}_z^{\rightarrow(t)} \in dx)}{1 - mt} + l \int_0^\infty \nu(dz) P(\bar{\tau}_z^{\rightarrow(t)} \in dx).$$

We stress that $(d(t))_{t \in [0, 1/m]}$ is not a Markov process since the jumps D_i before time t give information about $l(t)$ and, thus, about the future of the process. Also note that we can derive the law of $d(t)$ conditionally on $l(t)$ using Theorem 1. More precisely, conditionally on $l(t) = l$,

$$P(l(t) \in dl \mid d(t) = d) = \mathbf{1}_{l \geq d} \frac{\Pi^{(t)}(dl)}{\bar{\Pi}^{(t)}(d)} \quad \text{for all } d > 0.$$

Finally we turn our interest to the process of length $(l(t))_{t \in [0, 1/m]}$. Its increments, which are due to files that arrived at the left of $g(t)$ which are not stored entirely at the left of $g(t)$, are denoted by L_i :

$$L_i := l(T_i) - l(T_i^-) = G_i + D_i.$$

The other increments of $(l(t))_{t \in [0, 1/m]}$ are due to files which arrive at B_0 . We can view $(l(t))_{t \in [0, 1/m]}$ as a branching process in continuous time with immigration L_i at time T_i (with no death, inhomogeneous branching, and inhomogeneous immigration).

Corollary 5. *The process $(l(t))_{t \in [0, 1/m]}$ is an inhomogeneous pure-jump Markov process satisfying the following properties.*

(i) *The process $\{(T_i, L_i) : i \in \mathbb{N}\}$ is a PPP on $[0, 1/m] \times \mathbb{R}^+$ with intensity*

$$dt \int_{z \in [0, \infty]} \nu(dz) P(\vec{\tau}_z^{(t)} \in dx) x,$$

and $\{(T_i, L_i) : i \in \mathbb{N}, T_i > t\}$ is independent of $(l(s))_{s \in [0, t]}$.

(ii) *Conditionally on $T_i \leq t \leq t + s < T_{i+1}$ for some $i \in \mathbb{N}$, $(l(t + u))_{u \in [0, t-s]}$ satisfies the branching property: the law of $(l(t + u))_{u \in [0, t-s]}$ conditioned on $l(t) = x + y$ is equal to the law of the sum of two independent processes whose laws are respectively equal to $(l(t + u))_{u \in [0, t-s]}$ conditioned on $l(t) = x$ and $(l(t + u))_{u \in [0, t-s]}$ conditioned on $l(t) = y$.*

The jump rate of $(l(t))_{t \in [0, 1/m]}$ is given by the following vague convergence of measures on $(0, \infty)$ for h tending to 0:

$$\frac{P(l(t + h) - l(t) \in dx \mid l(t) = l)}{h} \xrightarrow{v} (x + l) \int_0^\infty \nu(dz) P(\vec{\tau}_z^{(t)} \in dx).$$

Example 3. For the basic example, $\nu = \delta_1$, the jump rate of the length is equal to

$$\sum_{n=1}^\infty \frac{n + l}{n} e^{-tn} \frac{(tn)^{n-1}}{(n-1)!} \delta_n(dx).$$

This is a consequence of the last displayed limit and (5).

Proof of Corollary 4. Using (14), we obtain

$$\int_{z \in [0, \infty]} P(\vec{\tau}_z^{(t)} \in dx) \int_{y \in [0, \infty]} \rho^{(t)}(dy dz) = \frac{\int_{z \in [0, \infty]} dz \bar{\nu}(z) P(\vec{\tau}_z^{(t)} \in dx)}{1 - mt},$$

which gives the intensity of $\{(T_i, D_i) : i \in \mathbb{N}\}$ by Proposition 3.

Proof of Corollary 5. (i) Writing $L_i = G_i + D_i$, Proposition 3 entails that $\{(T_i, L_i) : i \in \mathbb{N}\}$ is a PPP on $[0, 1/m] \times \mathbb{R}^+$ with intensity $dt \tilde{\mu}_t(dx)$, where $\tilde{\mu}_t$ is a measure on \mathbb{R}^+ defined for a Borel set A of \mathbb{R}^+ by

$$\tilde{\mu}_t(A) = \int_{\mathbb{R}_+^2} \mathbf{1}_{\{y+y' \in A\}} \int_0^\infty P(\vec{\tau}_z^{(t)} \in dy') \rho^{(t)}(dy dz).$$

To determine $\tilde{\mu}_t$, we compute its Laplace transform using Lemma 3:

$$\begin{aligned} \int_0^\infty e^{-\lambda x} \tilde{\mu}_t(dx) &= \int_{\mathbb{R}^+} e^{-\lambda(y+y')} \rho^{(t)}(dy dz) P(\vec{\tau}_z^{(t)} \in dy') \\ &= \int_{\mathbb{R}^+} e^{-\lambda y'} P(\vec{\tau}_z^{(t)} \in dy') dz \\ &\quad \times \int_z^\infty v(dl) \left(e^{-\lambda y} P(\vec{\tau}_{l-z}^{(t)} \in dy) \right. \\ &\quad \left. + \int_0^y e^{-\lambda x} P(\vec{\tau}_{l-z}^{(t)} \in dx)(y-x)e^{-\lambda(y-x)} \Pi^{(t)}(dy-x) \right) \\ &= \int_0^\infty dz \exp(-z\kappa^{(t)}(\lambda)) \\ &\quad \times \int_z^\infty v(dl) \exp(-(l-z)\kappa^{(t)}(\lambda)) \left(1 + \int_0^\infty e^{-\lambda u} u \Pi^{(t)}(du) \right) \\ &= \int_0^\infty v(dl) l \exp(-l\kappa^{(t)}(\lambda)) [\kappa^{(t)}]'(\lambda) \\ &= -\frac{\partial}{\partial y} \left[\int_0^\infty v(dl) \exp(-l\kappa^{(t)}(y)) \right] (\lambda) \\ &= -\frac{\partial}{\partial y} \left[\int_0^\infty e^{-yx} \int_0^\infty v(dl) P(\vec{\tau}_l^{(t)} \in dx) \right] (\lambda) \\ &= \int_0^\infty e^{-\lambda x} x \int_0^\infty v(dl) P(\vec{\tau}_l^{(t)} \in dx). \end{aligned}$$

Then $\tilde{\mu}_t(dx) = x \int_0^\infty v(dz) P(\vec{\tau}_z^{(t)} \in dx)$, which gives the intensity of $\{(T_i, L_i) : i \in \mathbb{N}\}$.

(ii) The branching property can be seen as a consequence of the determination of the jump rate. We give a more intuitive approach here. We condition by $l(t) = x + y$ and by $T_i \leq t \leq t + s < T_{i+1}$, and we make the decomposition effective by splitting $B_0(t)$ into two segments of length x and y . First we store the files $\{i : t_i \in (t, t + s], x_i > d(t)\}$. The free space of the half line at the right of $B_0(t)$ is now the closed range of a subordinator distributed like $(\tau_x^{(t+s)})_{x \geq 0}$ (see step 1 in the proof of Corollary 4). Then we successively store the files $\{i : t_i \in (t, t + s], x_i \in [g(t), g(t) + x]\}$ and $\{i : t_i \in (t, t + s], x_i \in (g(t) + x, d(t))\}$, which induce two successive increments of the length. The free space at the right of 0 after the first storage maintains the same distribution and is independent of the first increment by strong regeneration. So the two increments are independent and distributed as $l(t + s) - l(t)$ conditioned on $l(t) = x$ and on $l(t) = y$, respectively. This gives the result since $l(t)$ is Markovian. Formally, $l(t + s) - l(t)$ is equal to $\tau_{S_{s(x+y)}}^{(t+s)}$ (see the proof of Proposition 3) and

$$\tau_{S_{s(x+y)}}^{(t+s)} = \tau_{S_{sx}}^{(t+s)} + \tau_{S_{s(x+y)}}^{(t+s)} - \tau_{S_{sx}}^{(t+s)}$$

gives the decomposition expected since

$$\tau_{S_y(x+y)}^{(t+s)} - \tau_{S_{xx}}^{(t+s)} \stackrel{D}{=} \tau_{S_{yy}}^{(t+s)}.$$

Using Corollary 3 and recalling the definition of $\tilde{\mu}_t$, given at the beginning of the proof, ensures that

$$h^{-1} P(l(t+h) - l(t) \in dx \mid l(t) = l)$$

converges to

$$\tilde{\mu}_t(dx) + l \int_0^\infty \nu(dz) P(\tau_z^{\rightarrow(t)} \in dx).$$

This completes the proof, since $\tilde{\mu}$ has been determined above.

8. Complements

8.1. The distribution of $\{(T_i, G_i) : i \in \mathbb{N}\}$ derived from Theorem 3

In Section 5 we used the total intensity of the PPP $\{(T_i, G_i) : i \in \mathbb{N}\}$ to prove that the intensity of the PPP $\{(T_i, G_i, R_i) : i \in \mathbb{N}\}$ is equal to $dt\rho^{(t)}(dy dz)$ (Theorem 3). Here we check that integrating this intensity with respect to the third coordinate enables us to recover the intensity of $\{(T_i, G_i) : i \in \mathbb{N}\}$ given in Theorem 2. For this purpose, use Lemma 3 to rewrite $\rho^{(t)}$ as

$$\rho^{(t)}(dy dz) = dz \int_0^\infty \nu(dl + z) \left(P(\tau_l^{\leftarrow(t)} \in dy) + \int_0^y P(\tau_l^{\leftarrow(t)} \in dx)(y-x)\Pi^{(t)}(dy-x) \right),$$

and calculate the Laplace transform of $\int_{z \in [0, \infty]} \rho^{(t)}(dy dz)$,

$$\begin{aligned} & \int_{y \in [0, \infty]} e^{-\lambda y} \int_{z \in [0, \infty]} \rho^{(t)}(dy dz) \\ &= \int_0^\infty \int_0^\infty dz \nu(dl + z) \\ & \quad \times \int_0^\infty e^{-\lambda y} \left(P(\tau_l^{\leftarrow(t)} \in dy) + \int_0^y P(\tau_l^{\leftarrow(t)} \in dx)(y-x)\Pi^{(t)}(dy-x) \right) \\ &= \int_0^\infty dl \bar{\nu}(l) \left(e^{-l\kappa(\lambda)} + \int_0^\infty P(\tau_l^{\leftarrow(t)} \in dx) e^{-\lambda x} \int_x^\infty e^{-\lambda(y-x)}(y-x)\Pi^{(t)}(dy-x) \right) \\ &= \int_0^\infty dl \bar{\nu}(l) e^{-l\kappa(\lambda)} [\kappa^{(t)}]'(\lambda) \\ &= \int_0^\infty dl \frac{\bar{\nu}(l)}{l} \frac{\partial}{\partial \lambda} E(-\exp(-l\kappa^{(t)}(\lambda))) \\ &= \int_0^\infty dl \frac{\bar{\nu}(l)}{l} \frac{\partial}{\partial \lambda} E(-\exp(-\lambda \tau_l^{\leftarrow(t)})) \\ &= \int_0^\infty dl \frac{\bar{\nu}(l)}{l} \int_0^\infty e^{-\lambda y} y P(\tau_l^{\leftarrow(t)} \in dy) \\ &= \int_0^\infty dy e^{-\lambda y} \int_0^\infty P(Y_y^{(t)} \in -dl) \bar{\nu}(l) \quad \text{using (7)}. \end{aligned}$$

Thus, we conclude with

$$dt \int_{z \in [0, \infty]} \rho^{(t)}(dy dz) = dt dx \int_0^\infty P(Y_x^{(t)} \in -dl) \bar{\nu}(l).$$

8.2. A direct proof of Corollary 2 using fluctuation theory

Here we determine the distribution of the remaining data using fluctuation theory. We obtain laws at fixed times and do not require Theorem 2, as for the proof of Corollary 2 in Section 5.

We fix $t, h,$ and $x \geq 0$. We add the lengths of files arrived on $[g(t) - x, g(t)]$ during the time interval $(t, t + h]$. Then we remove the free space in $[g(t) - x, g(t)]$ at time t which is equal to $L_x^{(t)}$. The sum of data which arrived at the left of $B_0(t)$ not stored at the left of $B_0(t)$ between time t and $t + h$ is equal to the maximum in $x \geq 0$ of this difference. It is also the quantity of data which has tried to occupy the location $g(t)$ (successfully or not) between time t and $t + h$, i.e. $Y_{g(t)}^{(t+h)} - I_{g(t)}^{(t+h)}$. So, we have the following result.

Lemma 9. *Let $0 \leq t < 1/m$ and $h \geq 0$, then*

$$Y_{g(t)}^{(t+h)} - I_{g(t)}^{(t+h)} = \sup\{S_{hx} - L_x^{(t)}, x \geq 0\} = \sup\{S_{\frac{\leftarrow(t)}{h\tau_x}} - x, x \geq 0\} \quad a.s.,$$

where $(S_x)_{x \geq 0}$ is a subordinator with drift $d = 0$ and Lévy measure $\nu(dx)$, which is independent of $(L_x^{(t)})_{x \geq 0}$ and $(\tau_x)_{x \geq 0}$.

Defining $S^{(t,h)} := \sup\{S_{\frac{\leftarrow(t)}{h\tau_x}} - x, x \geq 0\}$, we have, for all $0 < a \leq b$,

$$\begin{aligned} \lim_{h \rightarrow 0} h^{-1} P(S^{(t,h)} \in [a, b]) \\ = \lim_{h \rightarrow 0} h^{-1} P(\text{there exists } i \in \mathbb{N} \text{ such that } (T_i, R_i) \in (t, t + h] \times [a, b]), \end{aligned}$$

and we find the law given in Corollary 2.

Proposition 4. *We have the following weak convergence of bounded measures on $(0, \infty)$ when h tends to 0:*

$$\frac{P(S^{(t,h)} \in dx)}{h} \xrightarrow{w} \frac{\bar{\nu}(x) dx}{1 - mt}.$$

Proof. $(S_{\frac{\leftarrow(t)}{h\tau_x}} - x)_{x \geq 0}$ is a Lévy process with negative drift -1 , no negative jumps, and bounded variation. Its Laplace exponent is $\kappa^{(t)} \circ (h\phi) - \text{Id}$, where ϕ is the Laplace exponent of S and is defined by

$$\phi(\lambda) = \int_0^\infty (1 - e^{-\lambda x})\nu(dx) \quad \text{for all } \lambda \geq 0.$$

Also note that, using (6), we have

$$[\kappa^{(t)} \circ (h\phi) - \text{Id}]'(0) = [\kappa^{(t)}]'(0)h\phi'(0) - 1 = \frac{1}{1 - mt}mh - 1,$$

which is negative since $0 \leq t + h < 1/m$. Then identity (14) of [1] or Theorem 5 of [2] ensures that, for all $\lambda > 0$ and for all $h \in [0, 1/m - t)$,

$$E(\exp(-\lambda S^{(t,h)})) = \left(\frac{1}{1 - mt}mh - 1 \right) \frac{\lambda}{(\kappa^{(t)} \circ (h\phi) - \text{Id})(\lambda)}.$$

Moreover,

$$\frac{(\kappa^{(t)} \circ (h\phi) - \text{Id})(\lambda)}{\lambda} = \frac{\kappa^{(t)}(h\phi(\lambda))}{h\phi(\lambda)} \frac{h\phi(\lambda)}{\lambda} - 1 = -1 + \frac{1}{1 - mt} \frac{h\phi(\lambda)}{\lambda} + o_{h \rightarrow 0}(h).$$

So,

$$\mathbb{E}(\exp(-\lambda S^{(t,h)})) = 1 + \frac{1}{1 - mt} \left(\frac{\phi(\lambda)}{\lambda} - m \right) h + o_{h \rightarrow 0}(h).$$

We can now prove the convergence of $h^{-1} \mathbb{P}(S^{(t,h)} > x)$ when h tends to 0, i.e.

$$\lim_{h \rightarrow 0} \int_0^\infty e^{-\lambda x} \frac{\mathbb{P}(S^{(t,h)} > x)}{h} dx = \lim_{h \rightarrow 0} \frac{1 - \mathbb{E}(\exp(-\lambda S^{(t,h)}))}{h\lambda} = \frac{1}{1 - mt} \left(\frac{m}{\lambda} - \frac{\phi(\lambda)}{\lambda^2} \right).$$

Moreover, Fubini’s theorem gives

$$\int_0^\infty dx e^{-\lambda x} \int_x^\infty \bar{v}(a) da = \int_0^\infty v(dy) \int_0^y da \frac{1 - e^{-\lambda a}}{\lambda} = \frac{m}{\lambda} - \frac{\phi(\lambda)}{\lambda^2}.$$

Then, for every $\lambda > 0$,

$$\lim_{h \rightarrow 0} \int_0^\infty e^{-\lambda x} \frac{\mathbb{P}(S^{(t,h)} > x)}{h} dx = \int_0^\infty e^{-\lambda x} \frac{\int_x^\infty \bar{v}(a) da}{1 - mt} dx,$$

which proves the convergence of $\mathbb{P}(S^{(t,h)} \in dx)/h$ to $\bar{v}(x) dx/(1 - mt)$. Indeed, we introduce the measures $\mu_h(dx)$ and $\mu(dx)$ on \mathbb{R}^+ whose tails are given by

$$\mu_h((x, \infty]) = \frac{e^{-x} \mathbb{P}(S^{(t,h)} > x)}{h}, \quad \mu((x, \infty]) = \frac{e^{-x}}{1 - mt} \int_x^\infty \bar{v}(a) da.$$

The last displayed limit entails the weak convergence of $\mu_h(dx)$ to $\mu(dx)$ when h tends to 0, by convergence of Laplace transforms. As μ is nonatomic, for every $x \geq 0$, $\mu_h((x, \infty])$ tends to $\mu((x, \infty])$, which proves that $\mathbb{P}(S^{(t,h)} > x)/h$ tends to $\int_x^\infty \bar{v}(a) da/(1 - mt)$.

Remark 3. Denote by $\gamma^{(t,h)}$ the almost sure instant at which the supremum $S^{(t,h)}$ is reached. To obtain the distribution of $\{(T_i, G_i, R_i) : i \in \mathbb{N}\}$ by this method, we need to know the joint law of $(S^{(t,h)}, \tau_{\gamma^{(t,h)}}^{\leftarrow(t)})$, which we cannot derive directly from fluctuation theory.

References

- [1] BANSAYE, V. (2006). On a model for the storage of files on a hardware I: statistics at a fixed time and asymptotics. Preprint, LPMA, Université Pierre et Marie Curie. Available at <http://arxiv.org/abs/math/0611432>.
- [2] BERTOIN, J. (1996). *Lévy Processes* (Cambridge Tracts Math. **121**). Cambridge University Press.
- [3] BERTOIN, J. (1999). Subordinators: examples and applications. In *Lectures on Probability Theory and Statistics* (Saint-Flour, 1997; Lecture Notes Math. **1717**), Springer, Berlin, pp. 1–91.
- [4] BERTOIN, J. (2000). Subordinators, Lévy processes with no negative jumps, and branching processes. Lecture notes for MaPhySto. Available at <http://www.maphysto.dk/publications/MPS-LN/2000/8.pdf>.
- [5] CHASSAING, P. AND FLAJOLET, P. (2003). Hachage, arbres, chemins and graphes. *Gazette Math.* **95**, 29–49.
- [6] CHASSAING, P. AND LOUCHARD, G. (2002). Phase transition for parking blocks, Brownian excursion and coalescence. *Random Structures Algorithms* **21**, 76–119.
- [7] FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications*, Vol. II, 2nd edn. John Wiley, New York.
- [8] FLAJOLET, P., POBLETE, P. AND VIOLA, A. (1998). On the analysis of linear probing hashing. *Algorithmica* **22**, 490–515.

- [9] FOATA, D. AND RIORDAN, J. (1974). Mappings of acyclic and parking functions. *Aequationes Math.* **10**, 10–22.
- [10] KINGMAN, J. F. C. (1993). *Poisson Processes* (Oxford Studies Prob. **3**). Oxford University Press.
- [11] MILLAR, P. W. (1977). Zero-one laws and the minimum of a Markov process. *Trans. Amer. Math. Soc.* **226**, 365–391.
- [12] PITMAN, J. (2006). *Combinatorial Stochastic Processes* (Lecture Notes Math. **1875**). Springer, Berlin.
- [13] SATO, K. (1999). *Lévy Processes and Infinitely Divisible Distributions* (Cambridge Studies Adv. Math. **68**). Cambridge University Press.
- [14] WINKEL, M. (2005). Electronic foreign-exchange markets and passage events of independent subordinators. *J. Appl. Prob.* **42**, 138–152.