QUEUES WITH DELAYS IN TWO-STATE STRATEGIES AND LÉVY INPUT

R. BEKKER,* *Vrije Universiteit*

O. J. BOXMA,** *EURANDOM and Eindhoven University of Technology*

O. KELLA,*** *The Hebrew University of Jerusalem*

Abstract

We consider a reflected Lévy process without negative jumps, starting at the origin. When the reflected process first upcrosses level K, a timer is activated. After D time units, the timer expires and the Lévy exponent of the Lévy process is changed. As soon as the process hits zero again, the Lévy exponent reverses to the original function. If the process has reached the origin before the timer expires then the Lévy exponent does not change. Using martingale techniques, we analyze the steady-state distribution of the resulting process, reflected at the origin. We pay special attention to the cases of deterministic and exponential timers, and to the following three special Lévy processes: (i) a compound Poisson process plus negative drift (corresponding to an M/G/1 queue), (ii) Brownian motion, and (iii) a Lévy process that is a subordinator until the timer expires.

Keywords: M/G/1 queue; workload process; storage process; reflected Lévy process; Lévy exponent; two-state strategy; delayed feedback control

2000 Mathematics Subject Classification: Primary 60K25
Secondary 60J30

1. Introduction

Communication systems are often controlled using feedback information signals that regulate the transmission of traffic. One reason for this is to regulate the traffic volume in accordance with the actual level of congestion. For example, in networks with distributed congestion control, the transmission rate of the end-users is based on an estimation of the level of congestion. The foremost example is the Internet, which is predominantly regulated by end-users through the transmission control protocol which, for this reason, has been the focus of a large body of research [25], [33]. Another example is provided by Ethernet, which has significantly regained importance recently, owing to its use in metropolitan networks [31]. In Ethernet it is possible, via a so-called backpressure mechanism, to adjust transfer rates from one node to another depending on the level of congestion in the latter.

Significant understanding in the abovementioned systems has been obtained through the use of performance models with simplifying assumptions on the possible delay that feedback signals may suffer. In many cases it is assumed, for analytic tractability, that the traffic adaptation is

instantaneous, or takes a deterministic time. In practice, the congestion information may only become available after some (random) delay. The delay can either be the result of a physical distance that separates the sender from the receiver, or control signals may be deliberately delayed so as to prevent them from putting a disproportionate load on the system that may compromise stability. Both types of feedback may typically vary over time.

The goal of the present paper is to develop stochastic models that take delayed feedback control, based on congestion levels, into account. We aim at a profound qualitative understanding of the impact of delayed feedback signals, by abstracting them from the detailed characteristics of a particular feedback-based system.

The model under consideration is a reflected Lévy process [11] without negative jumps, starting at the origin. Special cases of Lévy processes are compound Poisson processes, Brownian motion, linear drift processes, and independent sums of these. When the reflected process first upcrosses a certain level K (corresponding to a certain level of congestion), a timer is activated. After D time units (the feedback delay), the timer expires and the Lévy exponent (also called the Laplace exponent) of the Lévy process is changed. Here D is a random variable that is assumed to be finite with probability 1. As soon as the reflected process hits zero again, the Lévy exponent reverses to the original function. If the process has reached the origin before the timer expires then the Lévy exponent does not change. It should be noted that this model contains the M/G/1 queue with delayed feedback control as a special case. In the case of a compound Poisson process with negative drift, the reflected Lévy process is indeed the workload process of an M/G/1 queue; changing the Lévy exponent now amounts to changing the drift (service speed), arrival rate, and/or service requirement distribution.

The main result of the paper is the determination of the steady-state distribution of the resulting process, reflected at the origin (Theorem 3.1). We employ martingale techniques [27] and exploit several properties of Lévy processes in obtaining this result.

1.1. Related queueing/performance literature

For the case of *zero* feedback delay, there is a large collection of papers in the queueing literature regarding M/G/1 queues with workload-dependent input; we extend this literature both by introducing a nonzero delay and by considering Lévy processes. We refer the reader to survey [21] for a large number of references on queueing systems with state-dependent parameters. These go back to pioneering papers like [23], in which a storage system was considered that operates with two speeds, depending on the workload level. For a textbook treatment of storage systems with state-dependent release, see [5, Chapter XIV]. We refer the reader to [10] for a study of M/G/1-type queues where both the service speed and arrival rate vary continuously with the workload. For additional references on queueing systems with workload-dependent rates, see, e.g. [7].

The literature on queueing systems driven by Lévy processes is considerably less extensive. For some studies of queues with adaptable rates driven by a Lévy process, we refer the reader to [9], [14], and the references therein.

The number of papers that take *delayed* feedback into account is also much smaller. Motivated by ATM (asynchronous transfer mode) networks, Altman *et al.* [3] studied a discrete-time queue with delayed information on the queue length. In a slot, a service is only attempted if, given the information available to the server, it is certain that at least one customer is present. Another early study regarding communication networks with rate-based flow control and action delay is [2]. Sharma [36] considered an M/G/1-type queue in which the service rate was controlled by a scheduler who receives workload information from the queue with some delay

(caused by the fact that the scheduler resides in a satellite and the queue is on earth). The paper focused on stability conditions and on the rate of convergence to the stationary distribution. See [19] and [20] for two recent studies on delayed feedback due to the use of the automatic repeat request protocol in an error-prone communication medium, and [17] and [18] for queue-length dependent delayed arrivals occurring in cable access networks regulated by a time-slot reservation procedure. The following paper is closest to the present study. Lee and Kim [30] considered an M/G/1 queue in which the service speed is adapted an exponential amount of time after a certain workload level is exceeded. The delay here occurs due to the fact that a setup time is needed for changing the service speed.

1.2. Organization of the paper

In Section 2 we present preliminary results on Lévy processes. Theorem 2.1 cites a known result on the joint transform of the first exit time from the interval $[0, K)$ of the reflected process and the exit position of the reflected process [6], [32]. We use this result to obtain a suitable expression for the exit position from $[0, K)$ of the reflected process (Corollary 2.1). The steady-state distribution of the workload process under consideration is analyzed in Section 3. In three subsections we successively consider (i) the interval until the first upcrossing of level K, (ii) the interval until either the timer expires or the origin is reached, and (iii) the remaining interval until the origin is reached (the latter interval might have length 0). Taking a weighted average over the three intervals, we finally obtain the Laplace–Stieltjes transform (LST) of the steady-state workload distribution in Theorem 3.1. We pay special attention to a few specific Lévy processes: the compound Poisson process with negative drift (corresponding to an M/G/1 queue) and Brownian motion. Section 4 is devoted to the special case in which the Lévy process X is nondecreasing (a subordinator) during the first and second subintervals. This case is of interest in view of queueing models with a removable server. Since the origin cannot be hit in the first two intervals, this case leads to more convenient expressions. Relatively tractable results also occur when the timer period is exponentially distributed, which is the subject of Section 5.

2. Preliminary results on Lévy processes

We refer the reader to [11] for an excellent treatment of Lévy processes. Consider a spectrally negative Lévy process $\tilde{X} = \{\tilde{X}(t), t \geq 0\}$, i.e. a Lévy process with no positive jumps. We are in fact interested in a Lévy process X with only positive jumps. These processes are however closely related, since the dual process of X is $-X$; see, e.g. [11, Section II.1]. In this section we assume that $\tilde{X}$ is neither nonincreasing nor deterministic.

For a Lévy process $\tilde{X}$, the generating function of $\tilde{X}(t)$ is finite for all $\alpha \geq 0$ and $t \geq 0$, and is given by

$$\mathrm{E}[e^{\alpha \tilde{X}(t)}] = e^{t\psi(\alpha)}$$

for some function $\psi(\alpha)$, which is called the *Lévy exponent*. It is also well known that, under the assumed conditions, $\psi(\alpha)$ is strictly convex on $[0, \infty)$, $\psi(0) = 0$, and $\lim_{\alpha \to \infty} \psi(\alpha) = \infty$ (see, e.g. [11]).

Let $\alpha_0 = \inf\{\alpha \geq 0 \mid \psi(\alpha) > 0\}$. If $\psi'(0+) < 0$ then $\alpha_0 > 0$ and otherwise $\alpha_0 = 0$. By strict convexity, ψ is strictly increasing and continuous on $[\alpha_0, \infty)$; thus, as a function from $[\alpha_0, \infty)$ to $[0, \infty)$, it has an inverse which we denote by Φ. For $a \geq 0$, let $\tau_a^+ = \inf\{t \geq 0 \mid \tilde{X}(t) > a\}$. For a given (appropriately measurable) functional h, we denote

$E_x[h(\tilde{X})] = E[h(x + \tilde{X})]$ and $P_x[h(\tilde{X}) \in \cdot] = P[h(x + \tilde{X}) \in \cdot]$. The transform of this exit time is given in Equation (3) of [28].

Proposition 2.1. ([28].) *For any $a \geq x$ and $s \geq 0$,*

$$E_x[\exp(-s\tau_a^+)] = e^{-\Phi(s)(a-x)}.$$

In particular, $P_x[\tau_a^+ < \infty] = \lim_{s \downarrow 0} E_x[\exp(-s\tau_a^+)] = e^{-\Phi(0)(a-x)} = \exp(-\alpha_0(a - x))$.

In Theorem 2.1 we shall present the joint transform of the first exit time from the interval $[0, K)$ of the reflected process and the exit position of the reflected process [6], [32]. Results on first exit times and exit positions are often expressed in the literature in terms of the family of so-called *scale functions*; see, e.g. [1], [6], [11], [12], [13], [28], and [32]. In this paper we are concerned only with exit positions, but since the results are generally studied jointly with first exit times, we introduce the more general framework first, to derive the exit position as a corollary.

Definition 2.1. For $q \geq 0$, the q-scale function $W^{(q)} \colon (-\infty, \infty) \to [0, \infty)$ is the unique function whose restriction to $(0, \infty)$ is continuous and has Laplace transform

$$\int_0^\infty e^{-\alpha x} W^{(q)}(x)\, dx = \frac{1}{\psi(\alpha) - q} \quad \text{for } \alpha > \Phi(q),$$

and $W^{(q)}(x) = 0$ for $x < 0$.

For exit positions, it is in fact sufficient to restrict to the case in which $q = 0$. In that case $W(\cdot) := W^{(0)}(\cdot)$, which is also often referred to as the scale function. In some special cases the scale function can be explicitly determined. For instance, if the Lévy process is a compound Poisson process with drift then $W(\cdot)$ can be related to the waiting time distribution in the M/G/1 queue. For further examples and details, we refer the reader to Section 3 and [32].

Moreover, a related quantity is

$$Z^{(q)}(x) := 1 + q \int_0^x W^{(q)}(y)\, dy \quad \text{for } x > 0,$$

with $Z^{(q)}(x) = 1$ for $x \leq 0$. Finally, for $c \geq 0$, define

$$\psi_c(\alpha) = \psi(c + \alpha) - \psi(c),$$

and let $W_c^{(q)}$ be the q-scale function associated with the Lévy process with exponent $\psi_c(\alpha)$; see [6] and [32] for details. Accordingly, let

$$Z_c^{(q)}(x) := 1 + q \int_0^x W_c^{(q)}(y)\, dy \quad \text{for } x > 0. \tag{2.1}$$

Now, we are ready to present the joint transform of the first exit time and exit position of the process reflected at its maximum. To clarify what we mean, let the process $\bar{X} := \{\bar{X}(t), t \geq 0\}$ be the running supremum, i.e.

$$\bar{X}(t) = \max\left\{ x_s, \sup_{0 \leq u \leq t} \tilde{X}(u) \right\},$$

where x_s is its initial maximum. The process $Y := \bar{X} - \tilde{X}$ then represents the Lévy process reflected at its supremum. We note that the reflected process at the supremum has the same law as the dual process reflected at its infimum (the latter is in fact required in Section 3). Also, note that $Y(0) =: z = x_s - x_0 \geq 0$. Let $\tau_K := \inf\{t \geq 0 : Y(t) \notin [0, K)\}$ be the first exit time from $[0, K)$. The joint Laplace transform of τ_K and $Y(\tau_K)$ is given in [6, Theorem 1] and [32, Corollary 3].

Theorem 2.1. ([6], [32].) *For $u \geq 0$ and $v \geq 0$, with $\tilde{X}(0) = x_0$ and $Y(0) = z \geq 0$,*

$$E_{z,x_0}[\exp(-u\tau_K - vY(\tau_K))] = e^{-vz}\left(Z_v^{(p)}(K-z) - W_v^{(p)}(K-z)\frac{pW_v^{(p)}(K) + vZ_v^{(p)}(K)}{W_v^{(p)'}(K) + vW_v^{(p)}(K)}\right),$$
$$(2.2)$$

where $p = u - \psi(v)$.

Below, we rewrite this expression for $u = 0$, i.e. for the exit position, to obtain an expression that is suitable for the analysis in Section 3. To simplify the notation, we express the LST of the overshoot in terms of the scale function $W(\cdot)$. To do so, we use the following relationship between scale functions (see, e.g. [6, Remark 4] and [32]):

$$W^{(u)}(x) = e^{vx}W_v^{(u-\psi(v))}(x) \tag{2.3}$$

for v such that $\psi(v) < \infty$. Equation (2.1) may thus be equivalently expressed as

$$Z_v^{(p)}(x) = 1 + p\int_0^x e^{-vy}W^{(u)}(y)\,\mathrm{d}y.$$

For $u = 0$, we then obtain

$$Z_v^{(-\psi(v))}(x) = 1 - \psi(v)\int_0^x e^{-vy}W(y)\,\mathrm{d}y, \tag{2.4}$$

which, using Definition 2.1, may be reduced to $\psi(v)\int_x^\infty e^{-vy}W(y)\,\mathrm{d}y$ in the case in which $v > \Phi(0)$. Next, we further rewrite the fraction on the right-hand side of (2.2). For the denominator, it easily follows by differentiating both sides of (2.3) with respect to x that, for $u = 0$,

$$W_v^{(-\psi(v))'}(x) + vW_v^{(-\psi(v))}(x) = e^{-vx}W'(x).$$

For the numerator, we apply partial integration in (2.4), yielding

$$-\psi(v)W_v^{(-\psi(v))}(x) + vZ_v^{(-\psi(v))}(x) = v - \psi(v)\int_{0-}^x e^{-vy}\,\mathrm{d}W(y).$$

Combining the above, we have derived the following result for the exit position from $[0, K)$ of the reflected process.

Corollary 2.1. *For $v \geq 0$, with $\tilde{X}(0) = x_0$ and $Y(0) = z \geq 0$,*

$$E_{z,x_0}[\exp(-vY(\tau_K))] = e^{-vz}\left(1 - \psi(v)\int_0^{K-z} e^{-vy}W(y)\,\mathrm{d}y\right)$$
$$- \frac{W(K-z)}{W'(K)}\left(v - \psi(v)\int_{0-}^K e^{-vy}\,\mathrm{d}W(y)\right),$$

which, for $v > \Phi(0)$, can be reduced to

$$E_{z,x_0}[\exp(-vY(\tau_K))] = \psi(v)\left(e^{-vz}\int_{K-z}^\infty e^{-vy}W(y)\,\mathrm{d}y - \frac{W(K-z)}{W'(K)}\int_K^\infty e^{-vy}\,\mathrm{d}W(y)\right).$$

3. Analysis

Let X be a right-continuous Lévy process without negative jumps, starting at the origin, with Lévy exponent $\phi(\alpha) = \log \mathrm{E}[\mathrm{e}^{-\alpha X(1)}]$ (the minus sign follows from the duality with the process $\check{X}$ in Section 2) with right inverse Φ. We exclude the degenerate case that X is a negative drift. In this section we also exclude the case that the process X is nondecreasing. For results in the case that X is a subordinator, we refer the reader to Section 4.

Let $Z(t) = Z(0) + X(t) + L(t)$, $t \geq 0$, where $L(t) = -\inf_{0 \leq s \leq t}[Z(0) + X(s)]^{-}$. Let $T_K = \inf\{t \geq 0 : Z(t) \geq K\}$, i.e. T_K denotes the epoch at which, for the first time, $Z(t)$ upcrosses level $K > 0$. We assume the following. At T_K a timer is activated. After D time units, the timer expires and the Lévy exponent changes into $\phi_*(\alpha)$. As soon as the process hits zero again, the Lévy exponent reverses to the original $\phi(\alpha)$. If the process has reached the origin before the timer expired then the Lévy exponent does not change, but remains $\phi(\alpha)$.

We shall analyze the steady-state distribution of the Z-process, by distinguishing three successive time intervals: (i) the interval from 0 to T_K, (ii) the interval from T_K until the timer expires or the origin is reached—whichever comes first, and (iii) the interval from the expiration of the timer until the origin is reached (if positive). We assume that $\phi'_*(0) > 0$ for the system to be stable, and let Z denote a random variable with the steady-state distribution of the Z-process. Using the theory of regenerative processes, and introducing τ as the length of time until the origin is reached for the first time after T_K,

$$\mathrm{E}[\mathrm{e}^{-\alpha Z}] = \frac{\mathrm{E}[\int_0^{T_K} \mathrm{e}^{-\alpha Z(s)} \,\mathrm{d}s] + \mathrm{E}[\int_{T_K}^{T_K + (\tau \wedge D)} \mathrm{e}^{-\alpha Z(s)} \,\mathrm{d}s] + \mathrm{E}[\int_{T_K + (\tau \wedge D)}^{T_K + \tau} \mathrm{e}^{-\alpha Z(s)} \,\mathrm{d}s]}{\mathrm{E}[T_K] + \mathrm{E}[\tau]}. \tag{3.1}$$

The steady-state analysis of the Z-process in each of these intervals heavily relies on a martingale technique. We treat the three intervals successively in Sections 3.1, 3.2, and 3.3. Furthermore, in Sections 3.1, 3.2, and 3.3 we specify the results for the two special cases of Brownian motion and the M/G/1 and M/M/1 queues.

3.1. The first interval

Applying the theory of [27] to $Z(\cdot)$ and simplifying, the following is seen to be a martingale:

$$M(t) = \phi(\alpha) \int_{s=0}^{t} \mathrm{e}^{-\alpha Z(s)} \,\mathrm{d}s - \mathrm{e}^{-\alpha Z(t)} + \mathrm{e}^{-\alpha Z(0)} - \alpha L(t). \tag{3.2}$$

In the following we take $Z(0) = 0$. Application of the optional sampling theorem, with stopping time T_K, to this martingale yields (cf. [5] and [27])

$$\phi(\alpha) \mathrm{E}\left[\int_{s=0}^{T_K} \mathrm{e}^{-\alpha Z(s)} \,\mathrm{d}s\right] = \mathrm{E}[\exp(-\alpha Z(T_K))] - 1 + \alpha \mathrm{E}[L(T_K)], \tag{3.3}$$

or

$$\mathrm{E}\left[\int_{s=0}^{T_K} \mathrm{e}^{-\alpha Z(s)} \,\mathrm{d}s\right] = \frac{\mathrm{E}[\exp(-\alpha Z(T_K))] - 1 + \alpha \mathrm{E}[L(T_K)]}{\phi(\alpha)}. \tag{3.4}$$

Note that this expression, when divided by $\mathrm{E}[T_K]$, represents the LST of the steady-state distribution of the Z-process on the interval $[0, T_K]$.

The transform of the overshoot $\mathrm{E}[\exp(-\alpha Z(T_K))]$ can be directly obtained by taking $z = 0$ (since $x_s = 0$ and $x_0 = 0$) in the last formula of Corollary 2.1, providing that, for $\alpha > \Phi(0)$,

$$\mathrm{E}[\exp(-\alpha Z(T_K))] = \phi(\alpha)\left(\int_K^\infty \mathrm{e}^{-\alpha y} W(y) \,\mathrm{d}y - \frac{W(K)}{W'(K)} \int_K^\infty \mathrm{e}^{-\alpha y} \,\mathrm{d}W(y)\right). \tag{3.5}$$

Note that $\alpha > \Phi(0) \geq 0$ for the two integrals in (3.5) to be bounded. To determine the constant $E[Z(T_K)]$, we differentiate the first formula of Corollary 2.1 at $\alpha = 0$ to obtain

$$E[Z(T_K)] = \phi'(0) \int_0^K W(y)\, dy + \frac{W(K)}{W'(K)} \left(1 - \phi'(0) \int_{0^-}^K dW(y) \right). \qquad (3.6)$$

In the remainder we will also use the notation $Z_l := Z(T_K)$ to denote the value of the reflected process at the moment it leaves $[0, K)$.

Letting $\alpha \downarrow 0$ in (3.4) and applying l'Hôspital's rule gives the following relation between $E[T_K]$ and $E[L(T_K)]$:

$$\phi'(0)\, E[T_K] = -E[Z(T_K)] + E[L(T_K)], \qquad (3.7)$$

so

$$E[L(T_K)] = E[Z(T_K)] + \phi'(0)\, E[T_K]. \qquad (3.8)$$

Note that, for $\phi'(0) = 0$, we directly obtain $E[L(T_K)] = E[Z(T_K)]$. In that case, letting $\alpha \downarrow 0$ in (3.4) and applying l'Hôspital's rule twice gives the constant $E[T_K] = E[Z(T_K)^2]/\phi''(0)$, where $E[Z(T_K)^2]$ may be determined from the first formula of Corollary 2.1. For $\phi'(0) \neq 0$, a second relation between $E[L(T_K)]$ and $E[T_K]$ is obtained by defining $\hat{\alpha}$ as a nonzero solution of $\phi(\alpha) = 0$. (For instance, for $\phi'(0) < 0$, take $\hat{\alpha} := \Phi(0)$ as the unique positive zero of $\phi(\alpha) = 0$.) Since the expectation on the left-hand side of (3.3) is finite for $\alpha = \hat{\alpha}$, the right-hand side of (3.3) should be 0 for this value of α:

$$E[L(T_K)] = \frac{1 - E[\exp(-\hat{\alpha} Z(T_K))]}{\hat{\alpha}}. \qquad (3.9)$$

From (3.8) and (3.9) (with $\phi'(0) \neq 0$),

$$E[T_K] = \frac{1}{\phi'(0)} \left(\frac{1 - E[\exp(-\hat{\alpha} Z(T_K))]}{\hat{\alpha}} - E[Z(T_K)] \right).$$

Now we consider two special cases.

3.1.1. *Brownian motion case.* If X is Brownian motion with drift parameter μ and variance parameter σ^2 then $\phi(\alpha) = (\sigma^2/2)\alpha^2 - \mu\alpha$. Instead of X, we shall also write B to denote the case of Brownian motion. It may be verified that (3.5) indeed reduces to

$$E[\exp(-\alpha Z(T_K))] = e^{-\alpha K},$$

and $Z(T_K) = K$, as should have been expected since all paths of the reflected Brownian motion are continuous. Equation (3.8) thus becomes

$$E[L(T_K)] = K - \mu\, E[T_K].$$

In the case in which $\mu = 0$ it directly follows that $E[L(T_K)] = K$ and $E[T_K] = (K/\sigma^2)^2$. For $\mu \neq 0$, there is one nonzero α for which $\phi(\alpha) = 0$, namely $\hat{\alpha} = 2\mu/\sigma^2$, and

$$0 = \exp\left(-\frac{2\mu}{\sigma^2} K \right) - 1 + \frac{2\mu}{\sigma^2}(K - \mu\, E[T_K]),$$

so

$$E[T_K] = \frac{\sigma^2}{2\mu^2} \left(\exp\left(-\frac{2\mu}{\sigma^2} K \right) - 1 + \frac{2\mu}{\sigma^2} K \right).$$

It now follows from (3.4) that

$$\mathrm{E}\left[\int_{s=0}^{T_K} \mathrm{e}^{-\alpha Z(s)}\,\mathrm{d}s\right] = \frac{\mathrm{e}^{-\alpha K} - 1 - (\alpha/\hat{\alpha})(\mathrm{e}^{-\hat{\alpha}K} - 1)}{(\sigma^2/2)\alpha^2 - \mu\alpha}.$$

3.1.2. M/G/1 and M/M/1 cases. Let X correspond to the M/G/1 queue with arrival rate λ, service speed r, and general service requirements with distribution function $B(\cdot)$, mean β, and LST $\beta(\cdot)$. Then $\phi(\alpha) = r\alpha - \lambda(1 - \beta(\alpha))$. As in the Brownian motion case, we can determine the scale function $W(\cdot)$ explicitly. Define $\rho := \lambda\beta/r$, and define

$$H(x) := \beta^{-1}\int_0^x (1 - B(y))\,\mathrm{d}y$$

as the distribution of the residual service requirement. For $\rho < 1$, it is well known, from Definition 2.1 with $q = 0$ (see, e.g. [5, Theorem VIII.5.7]), that

$$W(x) = \frac{1}{r}\sum_{n=0}^{\infty} \rho^n H_n(x), \tag{3.10}$$

where $H_n(\cdot)$ denotes the n-fold convolution of $H(\cdot)$ with itself. In fact, $(1 - \rho)r W(\cdot)$ corresponds to the workload distribution in the M/G/1 queue with service rate r (the $1/r$ in (3.10) is in fact a correction term).

Remark 3.1. In the case in which $\rho \geq 1$ the scale function can be obtained by replacing $\rho H(x)$ in (3.10) by $\hat{H}(x) := \int_0^x \mathrm{e}^{-\delta u}\,\mathrm{d}\rho H(u)$, with δ the unique positive zero of $\int_0^\infty \mathrm{e}^{-xu}\,\mathrm{d}\rho H(u) - 1$. We refer the reader to [15] for further details. In this case $W(\cdot)/W(K)$ can be identified with the workload distribution of the finite M/G/1 dam with buffer size K.

The results can be further simplified in the M/M/1 case. Assume that the server works at unit speed as long as there is any work present, and let the service requirements be exponentially distributed with mean $1/\mu := \beta/r$. (We note that μ has a different meaning in the M/M/1 case than in the case of Brownian motion.) Then $\phi(\alpha) = \alpha - \lambda\alpha/(\mu + \alpha)$. Simplifying (3.5) or applying the memoryless property of the exponential service time distribution yields $\mathrm{E}[\exp(-\alpha Z(T_K))] = \mathrm{e}^{-\alpha K}\mu/(\mu + \alpha)$. For $\rho \neq 1$, after straightforward calculations, we obtain

$$\hat{\alpha} = \lambda - \mu$$

and

$$\mathrm{E}[T_K] = \frac{1}{1 - \lambda/\mu}\left(\frac{1 - (\mu/\lambda)\mathrm{e}^{-(\lambda-\mu)K}}{\lambda - \mu} - K - \frac{1}{\mu}\right).$$

3.2. The second interval

We put the time origin at T_K, i.e. the timer starts at time 0, where $Z(0)$ is distributed as the overshoot over level K, that is, Z_l. Application of the martingale $M(t)$, now with stopping time $D \wedge \tau$, yields, as in (3.3),

$$\phi(\alpha)\,\mathrm{E}\left[\int_{s=0}^{D\wedge\tau} \mathrm{e}^{-\alpha Z(s)}\,\mathrm{d}s\right] = \mathrm{E}[\mathrm{e}^{-\alpha Z(D\wedge\tau)}] - \mathrm{E}[\mathrm{e}^{-\alpha Z(0)}], \tag{3.11}$$

where $\mathrm{E}[\mathrm{e}^{-\alpha Z(0)}]$ is given by (3.5). Note that, this time, there is no reflection term, since the horizontal axis is not hit during $[0, D \wedge \tau)$. For the same reason, we can write $Z(s) = Z(0) + X(s)$, where $X(\cdot)$ is the free Lévy process.

Let us for the moment assume that the timer, that starts at T_K, runs for a fixed time t; later we consider the case of a random timer. Consider the unknown term on the right-hand side of (3.11), using $(\cdot)$ to denote an indicator function:

$$
\begin{aligned}
\mathrm{E}_{Z(0)}[\mathrm{e}^{-\alpha Z(t\wedge\tau)}] &= \mathrm{E}_{Z(0)}[\mathrm{e}^{-\alpha Z(\tau)}(\tau < t)] + \mathrm{E}_{Z(0)}[\mathrm{e}^{-\alpha Z(t)}(\tau \geq t)] \\
&= \mathrm{P}_{Z(0)}(\tau < t) + \mathrm{E}_{Z(0)}[\mathrm{e}^{-\alpha(Z(0)+X(t))}(\tau \geq t)] \\
&= \mathrm{P}_{Z(0)}(\tau < t) + \mathrm{E}[\mathrm{e}^{-\alpha(Z(0)+X(t))}] - \mathrm{E}_{Z(0)}[\mathrm{e}^{-\alpha(Z(0)+X(t))}(\tau < t)],
\end{aligned}
\tag{3.12}
$$

where we used the fact that $Z(\cdot)$ is a free Lévy process for $\tau \geq t$. We shall successively study these three terms. For the first term, we note that, for fixed x, the transform of $\mathrm{P}_x(\tau < t)$ is given by Proposition 2.1. Specifically, for $x \geq 0$ and $s \geq 0$, we have

$$
\mathrm{E}_x[\mathrm{e}^{-s\tau}] = \mathrm{e}^{-\Phi(s)x}.
\tag{3.13}
$$

Conditioning on $Z(0)$ and invoking the second formula of Corollary 2.1 with $z = 0$ (i.e. apply (3.5) in the second step), we may write for the transform of τ,

$$
\begin{aligned}
\mathrm{E}_{Z(0)}[\mathrm{e}^{-s\tau}] &= \int_K^\infty \mathrm{e}^{-\Phi(s)x}\,\mathrm{dP}(Z(0) < x) \\
&= s\left(\int_K^\infty \mathrm{e}^{-\Phi(s)y}W(y)\,\mathrm{d}y - \frac{W(K)}{W'(K)}\int_K^\infty \mathrm{e}^{-\Phi(s)y}\,\mathrm{d}W(y)\right),
\end{aligned}
\tag{3.14}
$$

where we used the facts that $\Phi(s) > \Phi(0)$ and $\phi(\Phi(s)) = s$ for $s > 0$; see, e.g. [11, p. 189]. This completes the analysis of the transform of τ. To obtain its distribution, define the first passage time into $(-\infty, -x]$ by $T(x) := \inf\{t \geq 0 : X(t) \leq -x\}$. We note that the distribution of $T(x)$ has a possible atom. For instance, in the standard M/G/1 queue with service speed r, $T(x)$ has an atom at $t = x/r$ and a density for $t > x/r$. Dividing both sides of (3.14) by s and applying (3.13), we obtain, by Laplace inversion, for $t > 0$,

$$
\mathrm{P}_{Z(0)}(\tau < t) = \left(\int_K^\infty W(y)\,\mathrm{dP}(T(y) < t) - \frac{W(K)}{W'(K)}\int_K^\infty W'(y)\,\mathrm{dP}(T(y) < t)\right).
\tag{3.15}
$$

Explicit expressions for the distribution of $T(x)$ and, thus, for $\mathrm{P}_{Z(0)}(\tau < t)$ are only available in some special cases; see, e.g. [34, Chapter 2] and [35, Chapter 4].

Now we turn to the second and third term of (3.12). Using the Lévy exponent, we have

$$
\mathrm{E}[\mathrm{e}^{-\alpha(Z(0)+X(t))}] = \mathrm{E}[\mathrm{e}^{-\alpha Z(0)}]\mathrm{e}^{\phi(\alpha)t}.
$$

For the third term, we have $\tau < t$ with t the timer duration. Hence, if $\tau = u \leq t$ then the Lévy process starts at 0 at time u, i.e. $Z(0) + X(u) = 0$. Conditioning on τ, we thus obtain

$$
\mathrm{E}_{Z(0)}[\mathrm{e}^{-\alpha(Z(0)+X(t))}(\tau < t)] = \int_0^t \mathrm{e}^{\phi(\alpha)(t-u)}\,\mathrm{dP}_{Z(0)}(\tau < u).
$$

Combining the above yields

$$
\mathrm{E}_{Z(0)}[\mathrm{e}^{-\alpha Z(t\wedge\tau)}] = \mathrm{P}_{Z(0)}(\tau < t) + \mathrm{E}[\mathrm{e}^{-\alpha Z(0)}]\mathrm{e}^{\phi(\alpha)t} - \int_0^t \mathrm{e}^{\phi(\alpha)(t-u)}\,\mathrm{dP}_{Z(0)}(\tau < u).
\tag{3.16}
$$

Integrating over t then directly yields

$$E_{Z(0)}[e^{-\alpha Z(D \wedge \tau)}] = \int_{t=0}^{\infty} \left(P_{Z(0)}(\tau < t) + E[e^{-\alpha Z(0)}]e^{\phi(\alpha)t} \right.$$
$$\left. - \int_0^t e^{\phi(\alpha)(t-u)} dP_{Z(0)}(\tau < u) \right) dP(D \le t). \qquad (3.17)$$

It follows from (3.11) and (3.16) that, for fixed $t > 0$,

$$E\left[\int_{s=0}^{t \wedge \tau} e^{-\alpha Z(s)} ds \right] = \frac{\int_{u=0}^t (1 - e^{\phi(\alpha)(t-u)}) dP_{Z(0)}(\tau < u) - (1 - e^{\phi(\alpha)t}) E[e^{-\alpha Z(0)}]}{\phi(\alpha)}.$$
$$(3.18)$$

The result for a generally distributed timer is obtained by integrating over t:

$$E\left[\int_{s=0}^{D \wedge \tau} e^{-\alpha Z(s)} ds \right] = \frac{1}{\phi(\alpha)} \int_{t=0}^{\infty} \left(\int_{u=0}^t (1 - e^{\phi(\alpha)(t-u)}) dP_{Z(0)}(\tau < u) \right.$$
$$\left. - (1 - e^{\phi(\alpha)t}) E[e^{-\alpha Z(0)}] \right) dP(D \le t). \qquad (3.19)$$

The result is especially tractable in the case in which the timer D has an exponential distribution function (or mixture of exponentials). This special case is further addressed in Section 5. Moreover, in Section 5 we outline another method to derive (3.18) by considering an exponential timer first and using some properties of the exponential distribution (Remark 5.1).

For the steady-state distribution of the Z-process, we need the constant $E[D \wedge \tau]$. Letting $\alpha \downarrow 0$ in (3.19) and applying l'Hôspital's rule we also have, for $\phi'(0) \ne 0$,

$$E[D \wedge \tau] = \int_{t=0}^{\infty} \left(t\, P_{Z(0)}(\tau \ge t) + \int_{u=0}^t u\, dP_{Z(0)}(\tau < u) \right) dP(D \le t). \qquad (3.20)$$

(In the case in which $\phi'(0) = 0$ we have to apply l'Hôspital's rule twice.) It also easily follows from (3.11) that

$$E[D \wedge \tau] = \frac{E[Z(0)] - E[Z(D \wedge \tau)]}{\phi'(0)} = -\frac{E[X(D \wedge \tau)]}{\phi'(0)}.$$

Hence, for $\phi'(0) \ne 0$,

$$E[Z(D \wedge \tau)] = E[Z(0)] - \phi'(0) \int_{t=0}^{\infty} \left(t\, P_{Z(0)}(\tau \ge t) + \int_{u=0}^t u\, dP_{Z(0)}(\tau < u) \right) dP(D \le t),$$
$$(3.21)$$

with $E[Z(0)]$ given in (3.6).

3.2.1. *Brownian motion case.* For Brownian motion, it is possible to give explicit expressions for, for example, the first exit time and, thus, by (3.19), also for the steady-state workload distribution in the second interval. In particular,

$$\Phi(s) = \frac{\mu + \sqrt{\mu^2 + 2\sigma^2 s}}{\sigma^2},$$

giving the LST of τ; see (3.13). Moreover, denoting by $N(\cdot)$ the normal distribution, we may obtain from the LST of τ [35, p. 113], or from [24, p. 14],

$$P(\tau < t) = N\left(\frac{-K - \mu t}{\sigma \sqrt{t}} \right) + \exp\left(\frac{-2K\mu}{\sigma^2} \right) N\left(\frac{-K + \mu t}{\sigma \sqrt{t}} \right). \qquad (3.22)$$

After some calculations similar to [35, p. 112], (3.16) can be seen to reduce to

$$E[e^{-\alpha Z(t \wedge \tau)}]$$

$$= P(\tau < t) + e^{-\alpha K} e^{t\phi(\alpha)} \left(N\left(\frac{K - (\alpha \sigma^2 t - \mu t)}{\sigma \sqrt{t}} \right) \right.$$

$$\left. - \exp\left(2K\left(\alpha - \frac{\mu}{\sigma^2} \right) \right) N\left(\frac{-K - (\alpha \sigma^2 t - \mu t)}{\sigma \sqrt{t}} \right) \right). \tag{3.23}$$

Hence, in this case,

$$E\left[\int_{s=0}^{D \wedge \tau} e^{-\alpha Z(s)} \, ds \right] = \frac{1}{\phi(\alpha)} \int_{t=0}^{\infty} \left(P(\tau < t) \right.$$

$$+ e^{-\alpha K} e^{t\phi(\alpha)} \left(N\left(\frac{K - (\alpha \sigma^2 t - \mu t)}{\sigma \sqrt{t}} \right) \right.$$

$$\left. - \exp\left(2K\left(\alpha - \frac{\mu}{\sigma^2} \right) \right) N\left(\frac{-K - (\alpha \sigma^2 t - \mu t)}{\sigma \sqrt{t}} \right) \right)$$

$$\left. - e^{-\alpha K} \right) dP(D \leq t). \tag{3.24}$$

Remark 3.2. For Brownian motion, (3.23) can also be obtained more directly. Define the Brownian motion process $\tilde{B}(t) = -B(t)$, $t \geq 0$ (having drift parameter $-\mu$ and variance parameter σ^2), and its running supremum $\tilde{M}(t) = \sup_{0 \leq s \leq t}\{\tilde{B}(s)\}$. Then, clearly,

$$E[e^{-\alpha B(t)}(\tau \geq t)] = \int_{-K}^{\infty} e^{-\alpha x} P(B(t) \in dx, \tau \geq t)$$

$$= \int_{-\infty}^{K} e^{\alpha x} P(\tilde{B}(t) \in dx, \tilde{M}(t) \leq K).$$

Using [24, Proposition 1.8.1] for the joint distribution, we obtain, after some standard calculations, that $E[e^{-\alpha Z(t \wedge \tau)}]$ satisfies (3.23).

As mentioned, we need to determine the constant $E[D \wedge \tau]$ for the steady-state distribution of the Z-process. Letting $\alpha \downarrow 0$ in (3.24) and applying l'Hôspital's rule (determining the constant using (3.20) is more involved in this case) yields, for $\mu \neq 0$,

$$E[D \wedge \tau] = \int_{t=0}^{\infty} \left(t \, P(\tau \geq t) - \frac{K}{\mu} N\left(\frac{-K - \mu t}{\sigma \sqrt{t}} \right) \right.$$

$$\left. + \frac{K}{\mu} \exp\left(\frac{-2K\mu}{\sigma^2} \right) N\left(\frac{-K + \mu t}{\sigma \sqrt{t}} \right) \right) dP(D \leq t). \tag{3.25}$$

Because the results for Brownian motion are very explicit, we also give the constant in the case in which $\phi'(0) = \mu = 0$. Again letting $\alpha \downarrow 0$ in (3.24) and applying l'Hôspital's rule twice, we obtain after lengthy calculations, for $\mu = 0$,

$$E[D \wedge \tau] = \int_{t=0}^{\infty} \left(t \, P(\tau \geq t) - \frac{K^2}{\sigma^2} P(\tau < t) + \frac{2K}{\sigma} \frac{\sqrt{t}}{\sqrt{2\pi}} \exp\left(\frac{-K^2}{2\sigma^2 t} \right) \right) dP(D \leq t).$$

3.2.2. *M/G/1 and M/M/1 cases.* If X corresponds to the M/G/1 queue as described before then it directly follows that $\Phi(s)$, $s \geq 0$, is the unique nonnegative solution to $rt = s + \lambda(1 - \beta(t))$. By (3.13), this directly describes $\mathrm{E}[e^{-s\tau}]$. In addition, the distribution of $T(x)$ in the case in which $r = 1$ can be found in, for example, [16, Equation (II.4.95)] and [34, Equation (2.15)]. After some modification to adjust for the service speed r, the formula reads, for $t \geq x/r$,

$$\mathrm{P}(T(x) < t) = \int_{u=x/r}^{t} \sum_{n=0}^{\infty} e^{-\lambda u} \frac{(\lambda u)^n}{n!} \frac{x}{ru} \, dB_n(ur - x),$$

where, for $n \geq 1$, $B_n(\cdot)$ is the n-fold convolution of $B(\cdot)$ with itself, and $B_0(x) = 0$ if $x < 0$ and $B_0(x) = 1$ if $x \geq 0$. Using (3.15) and (3.19), this determines the density of τ and the distribution of the workload process in the second interval, respectively.

In the M/M/1 queue, with $\mu = r/\beta$, $\Phi(\cdot)$ is explicitly given by

$$\Phi(s) = \frac{s + \lambda - \mu + \sqrt{(\mu - \lambda - s)^2 + 4\mu s}}{2}. \tag{3.26}$$

In that case, $\mathrm{P}(T(x) = x) = e^{-\lambda x}$, and denote by $g(x, \cdot)$ the density of $T(x)$. Then, for $x < t < \infty$,

$$g(x, t) = e^{-\lambda t - \mu(t-x)} \frac{x}{t} \sqrt{\frac{\lambda \mu t}{t - x}} I_1(2\sqrt{\lambda \mu t (t - x)}),$$

where $I_n(\cdot)$ is the modified Bessel function of the first kind of order n, i.e.

$$I_n(x) = \left(\frac{x}{2}\right)^n \sum_{k=0}^{\infty} \frac{(x/2)^{2k}}{k!\,(n+k)!};$$

see, for instance, [34, Section 2.8b] and [35, Theorem 4.8].

3.3. The third interval

In this subsection we analyze the third interval, i.e. the interval $[D \wedge \tau, \tau]$. Since we use results of the second interval in this subsection, we put the time origin at the end of the first interval, as in Section 3.2. This allows for the most coherent presentation. We note that, because the timer has expired, the Lévy exponent of the process is changed into $\phi_*(\alpha)$. Another application of the martingale $M(t)$, now with stopping time τ, yields

$$\phi_*(\alpha) \, \mathrm{E}\left[\int_{s=D \wedge \tau}^{\tau} e^{-\alpha Z(s)} \, ds\right] = 1 - \mathrm{E}[e^{-\alpha Z(D \wedge \tau)}]. \tag{3.27}$$

As in the second interval, note that there is no reflection term in the time interval $[D \wedge \tau, \tau)$. Recall that Z_l denotes the value of the reflected process at the moment it leaves $[0, K)$ at the end of the first interval (Section 3.1). Also, $\mathrm{E}[e^{-\alpha Z(D \wedge \tau)}]$ is given in (3.17). Combining the above yields

$$\mathrm{E}\left[\int_{s=D \wedge \tau}^{\tau} e^{-\alpha Z(s)} \, ds\right] = \frac{1}{\phi_*(\alpha)} \left(1 - \int_{t=0}^{\infty} \left(\mathrm{P}_{Z_l}(\tau < t) + \mathrm{E}[\exp(-\alpha Z_l)]e^{\phi(\alpha)t} \right.\right.$$
$$\left.\left. - \int_0^t e^{\phi(\alpha)(t-u)} \, d\mathrm{P}_{Z_l}(\tau < u)\right) d\mathrm{P}(D \leq t)\right),$$

where Z_l is the overshoot resulting from the first interval.

Observe that τ is equal to time $D \wedge \tau$ in the case in which the process has reached the origin at the end of the second interval before the timer D has expired. Hence, $\mathrm{E}[\int_{s=D\wedge\tau}^{\tau} \mathrm{e}^{-\alpha Z(s)} \, \mathrm{d}s] = 0$ with probability $\mathrm{P}_{Z_l}(\tau < D)$. Finally, letting $\tilde{\tau} := \tau - (D \wedge \tau)$ denote the length of the third interval, we derive from (3.27) that

$$\mathrm{E}[\tilde{\tau}] = \frac{\mathrm{E}[Z(D \wedge \tau)]}{\phi'_*(0)},$$

with $\mathrm{E}[Z(D \wedge \tau)]$ presented in (3.21). Hence, observing that (for $\phi'(0) \neq 0$)

$$\mathrm{E}[\tau] = \mathrm{E}[D \wedge \tau] + \mathrm{E}[\tilde{\tau}] = \left(1 - \frac{\phi'(0)}{\phi'_*(0)}\right) \mathrm{E}[D \wedge \tau] + \frac{\mathrm{E}[Z_l]}{\phi'_*(0)}, \quad (3.28)$$

with $\mathrm{E}[D \wedge \tau]$ given in (3.20) and $\mathrm{E}[Z_l]$ given in (3.6), we have determined all the terms on the right-hand side of (3.1).

Finally, the steady-state analysis of the Z-process may be summarized by combining the three intervals of Sections 3.1–3.3. In particular, applying (3.4), (3.7), (3.11), (3.27), and (3.28) to (3.1), we deduce the following theorem.

Theorem 3.1. *For $\phi'(0) \neq 0$, we have*

$$\mathrm{E}[\mathrm{e}^{-\alpha Z}] = \frac{(\alpha/\phi(\alpha)) \, \mathrm{E}[L(T_K)] + (1/\phi(\alpha) - 1/\phi_*(\alpha))(\mathrm{E}_{Z_l}[\mathrm{e}^{-\alpha Z(D\wedge\tau)}] - 1)}{(1/\phi'(0)) \, \mathrm{E}[L(T_K)] + (1/\phi'(0) - 1/\phi'_*(0))(\phi'(0) \, \mathrm{E}[D \wedge \tau] - \mathrm{E}[Z_l])}, \quad (3.29)$$

with $\mathrm{E}[L(T_K)]$, $\mathrm{E}[Z_l]$, $\mathrm{E}[D \wedge \tau]$, and $\mathrm{E}_{Z_l}[\mathrm{e}^{-\alpha Z(D\wedge\tau)}]$ given by (3.9), (3.6), (3.20), and (3.17), respectively.

Remark 3.3. In the case in which $\phi(\cdot) \equiv \phi_*(\cdot)$, (3.29) reduces to

$$\mathrm{E}[\mathrm{e}^{-\alpha Z}] = \phi'(0) \frac{\alpha}{\phi(\alpha)},$$

which corresponds to the LST of the steady-state version of a reflected Lévy process; see, e.g. [5, Corollary IX.3.4] or [13], and [27].

Remark 3.4. The LST in (3.29) can also be decomposed as the weighted sum of four terms, each of which can be inverted in a straightforward manner; (i) the LST $\alpha\phi'(0)/\phi(\alpha)$, which is the well-known LST of a reflected Lévy process with exponent $\phi(\alpha)$ (see Remark 3.3), (ii) the product of the term in (i) with the overshoot $\mathrm{E}_{Z_l}[\mathrm{e}^{-\alpha Z(D\wedge\tau)}]$, (iii) the LST $\alpha\phi'_*(0)/\phi_*(\alpha)$ similar to (i), and (iv) the product of the term in (iii) with the overshoot $\mathrm{E}_{Z_l}[\mathrm{e}^{-\alpha Z(D\wedge\tau)}]$. We note that a related type of decomposition can be found in, for example, [26]. In fact, this decomposition is particularly useful for determining the steady-state distribution of the Z-process, which is possible by inverting each of the above four terms separately. We refer the reader to [8] for details in the case in which $D = 0$.

3.3.1. *Brownian motion, M/G/1, and M/M/1 cases.* The results can be directly derived using the terms determined in the second interval. In particular, for Brownian motion, $\mathrm{E}_{Z_l}[\mathrm{e}^{-\alpha Z(D\wedge\tau)}]$ can be obtained from (3.23), with $\mathrm{P}(\tau < t)$ given in (3.22). The constant $\mathrm{E}[D \wedge \tau]$ can be found in (3.25) and $\mathrm{E}[Z_l] = K$. If X corresponds to the M/G/1 or M/M/1 queue then the general equations for $\mathrm{E}_{Z_l}[\mathrm{e}^{-\alpha Z(D\wedge\tau)}]$ and $\mathrm{E}[D \wedge \tau]$, i.e. (3.17) and (3.20), can be further specified using the Lévy exponents $\phi(\alpha) = \alpha r - \lambda(1 - \beta(\alpha))$ and $\phi(\alpha) = \alpha - \lambda\alpha/(\mu + \alpha)$, respectively, and the results on first exit times given at the end of Section 3.2.

4. The case of a subordinator

In this section we consider the case in which the process X is *nondecreasing* (a *subordinator*) during the first and second intervals (because of stability, this is not possible during the third interval). This case is of special interest in view of queueing models with a removable server; see, e.g. [4], [22], [26], [29], and [37]. In these papers it was assumed that the server started a new busy period only when the total amount of work had reached the level D (D-policy; note that this is a different D than our timer D). The analysis in the case of a monotone process during the first two intervals is in fact very similar to the analysis in Section 3, that is, we may consider each of the three successive intervals separately. Since the origin is not hit during the first two intervals, this case leads to more convenient expressions.

Following [26], the Lévy exponent of the subordinator X will be defined by $-\eta(\alpha) = \log \mathrm{E}[\mathrm{e}^{-\alpha X(1)}] = \phi(\alpha)$ (note the difference in the minus sign between $\eta(\alpha)$ and $\phi(\alpha)$). We first consider the exit position of the process when it leaves $[0, K)$. Denote the exit time from $[0, K)$ of X by T_K. The exit position is then given by (see, e.g. [1, Subsection 3.3])

$$\mathrm{E}_0[\exp(-\alpha Z(T_K))] = \mathrm{E}_0[\exp(-\alpha X(T_K))] = \eta(\alpha) \int_K^\infty \mathrm{e}^{-\alpha z}\, \mathrm{d}U(z), \qquad (4.1)$$

where the potential measure U is defined via

$$\int_0^\infty \mathrm{e}^{-\alpha z}\, \mathrm{d}U(z) = \frac{1}{\eta(\alpha)}.$$

We note that the potential measure U and 0-scale function $W(\cdot)$ are closely related; see, e.g. [1], [6], [11], and [12] for details.

Remark 4.1. Of particular interest for queueing systems with a removable server is the case in which X is a compound Poisson process with rate λ and jumps having distribution function $B(\cdot)$ and LST $\beta(\cdot)$. In that case $\eta(\alpha) = \lambda(1 - \beta(\alpha))$ and $U(z) = \sum_{n=0}^\infty B_n(z)/\lambda$, where $B_n(\cdot)$ is the n-fold convolution of $B(\cdot)$ with itself. The distribution of $X(T_K)$ can also be obtained from renewal theory [5, Chapter V].

Next, we briefly outline the analysis of the Z-process. As in Section 3, we apply the martingale (3.2) with stopping times T_K and $T_K + D \wedge \tau$ for the first and second intervals, respectively. Observe that in the case of a subordinator there is no reflection, yielding $L(t) \equiv 0$ in (3.2). Also, the LST of $Z(T_K)$ is given in (4.1). As in Section 3, we denote the value of the Z-process when it leaves $[0, K)$ by $Z_l = Z(T_K)$. This completes the analysis of the first interval, up to a constant.

For the second interval, we note that this interval always terminates owing to the expiration of the timer, because the process is nondecreasing there. In particular, for a timer with fixed duration t, it follows directly from the definition of the Lévy exponent that

$$\mathrm{E}[\exp(-\alpha Z(T_K + t \wedge \tau))] = \mathrm{E}[\exp(-\alpha(Z(T_K) + X(t)))] = \mathrm{E}[\exp(-\alpha Z_l)]\mathrm{e}^{-\eta(\alpha)t},$$

with $\mathrm{E}[\exp(-\alpha Z_l)]$ given by (4.1). By integrating over t, we obtain the following result for a general timer:

$$\mathrm{E}[\exp(-\alpha Z(T_K + D \wedge \tau))] = \mathrm{E}[\exp(-\alpha Z_l)] \int_{t=0}^\infty \mathrm{e}^{-\eta(\alpha)t}\, \mathrm{d}\mathrm{P}(D \le t). \qquad (4.2)$$

Finally, the third interval can easily be analyzed using another application of the martingale (cf. (3.27)) and the LST of $Z(T_K + D)$ (cf. (4.2)). The three intervals can now be easily combined. Specifically, combining (3.1), (3.3), (3.11), and (3.27) with the above yields

$$\mathrm{E}[\mathrm{e}^{-\alpha Z}] = \frac{1}{\mathrm{E}[T_K] + \mathrm{E}[\tau]}\left(\frac{1}{\eta(\alpha)} + \frac{1}{\phi_*(\alpha)}\right)\left(1 - \mathrm{E}[\exp(-\alpha Z_l)]\int_{t=0}^{\infty}\mathrm{e}^{-\eta(\alpha)t}\,\mathrm{dP}(D \leq t)\right),$$

where $\mathrm{E}[\exp(-\alpha Z_l)]$ is given by (4.1). The constant $\mathrm{E}[T_K] + \mathrm{E}[\tau]$ can be determined by letting $\alpha \downarrow 0$ and applying l'Hôpital's rule, yielding

$$\mathrm{E}[T_K] + \mathrm{E}[\tau] = \left(\frac{1}{\eta'(0)} + \frac{1}{\phi_*'(0)}\right)(\mathrm{E}[Z_l] + \eta'(0)\,\mathrm{E}[D]).$$

To obtain the constant $\mathrm{E}[Z_l]$, we may use the definition of U to rewrite (4.1) as $\mathrm{E}_0[\exp(-\alpha Z_l)] = 1 - \eta(\alpha)\int_0^K \mathrm{e}^{-\alpha z}\,\mathrm{d}U(z)$. Differentiating with respect to α and letting $\alpha \downarrow 0$ then gives $\mathrm{E}[Z_l] = \eta'(0)\int_0^K \mathrm{d}U(z)$.

5. Exponential timer

In Section 3 we determined the LST of the workload in the model where the duration of the timer has a general distribution function (and the process is not monotone during the first two intervals). In the special case of Brownian motion, tractable expressions for the steady-state workload appear. However, for the M/G/1 queue and even for the M/M/1 queue, the results become cumbersome, involving the complicated transient behavior of those queues. To obtain more tractable analytical results, we consider the case of an exponential timer in this section. The results can also be readily extended to cases where the duration of the timer consists of a mixture of exponential terms, as for Coxian (with different intensity parameters) and hyperexponential distributions (see the end of this section). Thus, in this section we first assume that D is exponentially distributed with intensity ξ, i.e. $\mathrm{P}(D < x) = 1 - \mathrm{e}^{-\xi x}$. The analysis in the first interval does not depend on the distribution of the timer. For convenience, we put the time origin at the end of this first interval and let Z_l again denote the overshoot over K. For the second interval, the integrals on the right-hand side of (3.17) and (3.19) can now be easily determined. After interchanging the integrals on the right-hand side of (3.17) and (3.19), and after some straightforward calculations, it follows that, for $\alpha \geq 0$ such that $\phi(\alpha) < \xi$,

$$\mathrm{E}_{Z_l}[\mathrm{e}^{-\alpha Z(D \wedge \tau)}] = \mathrm{E}[\exp(-\alpha Z_l)]\frac{\xi}{\xi - \phi(\alpha)} - \mathrm{E}_{Z_l}[\mathrm{e}^{-\xi\tau}]\frac{\phi(\alpha)}{\xi - \phi(\alpha)}, \tag{5.1}$$

and

$$\mathrm{E}\left[\int_{s=0}^{D \wedge \tau}\mathrm{e}^{-\alpha Z(s)}\,\mathrm{d}s\right] = \frac{-\mathrm{E}_{Z_l}[\mathrm{e}^{-\xi\tau}]}{\xi - \phi(\alpha)} + \frac{\mathrm{E}[\exp(-\alpha Z_l)]}{\xi - \phi(\alpha)}, \tag{5.2}$$

with $\mathrm{E}[\exp(-\alpha Z_l)]$ given by (3.5) and the constant $\mathrm{E}_{Z_l}[\mathrm{e}^{-\xi\tau}]$ given by (3.14). By analytic continuation, the results can be extended to all values of $\alpha \geq 0$. Note that this also holds for the α for which $\phi(\alpha) = \xi$, after an application of l'Hôpital's rule to (5.2). We note that, using (5.1), it is easy to analyze the third interval.

Finally, the constant $\mathrm{E}[D \wedge \tau]$ can be obtained either from (3.20) or by taking $\alpha = 0$ in (5.2), yielding

$$\mathrm{E}[D \wedge \tau] = \frac{1}{\xi}(1 - \mathrm{E}_{Z_l}[\mathrm{e}^{-\xi\tau}]). \tag{5.3}$$

Now, combining the above with (3.29), the LST of the steady-state workload in the case of an exponential timer is summarized in the following corollary.

Corollary 5.1. *Assume that* $P(D < x) = 1 - e^{-\xi x}$. *Then, for* $\phi'(0) \neq 0$,

$$
\begin{aligned}
&E[e^{-\alpha Z}] \\
&= \left(\frac{\alpha}{\phi(\alpha)} E[L(T_K)] + \left(\frac{1}{\phi(\alpha)} - \frac{1}{\phi_*(\alpha)} \right) \right. \\
&\qquad\qquad \times \left. \left(\frac{\xi}{\xi - \phi(\alpha)} E[\exp(-\alpha Z_l)] - \frac{\phi(\alpha)}{\xi - \phi(\alpha)} E_{Z_l}[e^{-\xi\tau}] - 1 \right) \right) \\
&\quad \times \left(\frac{1}{\phi'(0)} E[L(T_K)] + \left(\frac{1}{\phi'(0)} - \frac{1}{\phi'_*(0)} \right) \left(\frac{\phi'(0)}{\xi} - \frac{\phi'(0)}{\xi} E_{Z_l}[e^{-\xi\tau}] - E[Z_l] \right) \right)^{-1},
\end{aligned}
$$

with $E[L(T_K)]$, $E[Z_l]$, $E[\exp(-\alpha Z_l)]$, *and* $E_{Z_l}[e^{-\xi\tau}]$ *given by (3.9), (3.6), (3.5), and (3.14), respectively.*

Remark 5.1. Another way to analyze the LST of the workload in the second interval, that is, (3.18) in Section 3, is to consider the case of an exponential timer first and exploit the lack-of-memory property of an exponential timer. More specifically, let D denote a generic exponential random variable with mean $1/\xi$. Using (3.12), it follows that

$$
\begin{aligned}
E_{Z(0)}[e^{-\alpha Z(D \wedge \tau)}] &= P_{Z(0)}(\tau < D) + E[e^{-\alpha(Z(0) + X(D))}] - E_{Z(0)}[e^{-\alpha(Z(0) + X(D))}(\tau < D)] \\
&= E_{Z(0)}[e^{-\xi\tau}] + E[e^{-\alpha Z(0)}] \frac{\xi}{\xi - \phi(\alpha)} - E_{Z(0)}[e^{-\xi\tau}] \frac{\xi}{\xi - \phi(\alpha)},
\end{aligned}
$$

where the last term in the second step follows from the fact that at time τ, that is, the first epoch at which $Z(0) + X(\tau) = 0$, the time until the timer expires is still exponentially distributed. Using the above, it is easily seen that $E[\int_{s=0}^{D \wedge \tau} e^{-\alpha Z(s)} \, ds]$ is given by (5.2). Finally, note that integrating over an exponential timer is equivalent to multiplying by ξ and taking the LST. It may be checked that dividing by ξ and inverting (5.2) with respect to ξ provides (3.18).

5.1. Brownian motion case

For Brownian motion, a generally distributed timer already yields explicit results; see Section 3. For an exponential timer, the result can be further simplified to the transform given in Corollary 5.1, with $E[\exp(-\alpha Z_l)] = e^{-\alpha K}$, $E[Z_l] = K$, and $\phi(\alpha)$ and $E[L(T_K)]$ given in Section 3.1. Finally (see Section 3.2),

$$
E_{Z_l}[e^{-\xi\tau}] = \exp\left(-\frac{\mu + \sqrt{\mu^2 + 2\sigma^2\xi}}{\sigma^2} \right) K.
$$

5.2. M/G/1 and M/M/1 cases

In the M/G/1 case the scale function $W(\cdot)$ has an explicit form, related to the steady-state workload distribution in the M/G/1 queue; see Section 3.1. As noted there, in the M/M/1 queue we have the simple form

$$
E[\exp(-\alpha Z_l)] = e^{-\alpha K} \frac{\mu}{\mu + \alpha} \quad \text{and} \quad E[Z_l] = K + \frac{1}{\mu}.
$$

Also, $E[L(T_K)]$ can be obtained from the results in Section 3.1. Finally, the constant $E_{Z_l}[e^{-\xi\tau}]$ in the M/M/1 case is given by

$$E_{Z_l}[e^{-\xi\tau}] = e^{-\Phi(\xi)K}\frac{\mu}{\mu + \Phi(\xi)},$$

with $\Phi(\xi)$ presented in (3.26). We note that

$$\frac{\mu}{\mu + \Phi(\xi)} = \frac{\xi + \lambda + \mu - \sqrt{(\mu + \lambda + \xi)^2 - 4\lambda\mu}}{2\lambda},$$

corresponding to the LST of the M/M/1 busy period (having parameter ξ).

The results for an exponential timer can also be extended to cases where the distribution of the timer duration consists of a mixture of exponential terms. This extension is rather straightforward in cases where all exponentials have a different intensity parameter, as for Coxian (with different parameters) and hyperexponential distributions, for example. In particular, assume that $P(D < x) = 1 - \sum_{i=1}^{k} p_i \exp(-\xi_i x)$ with $\sum_{i=1}^{k} p_i = 1$ (where some p_is are allowed to be negative). In this case (3.17) reduces to

$$E_{Z_l}[e^{-\alpha Z(D\wedge\tau)}] = \sum_{i=1}^{k} p_i\left(E[\exp(-\alpha Z_l)]\frac{\xi_i}{\xi_i - \phi(\alpha)} - E_{Z_l}[\exp(-\xi_i\tau)]\frac{\phi(\alpha)}{\xi_i - \phi(\alpha)}\right).$$

After similar calculations as in the case of an exponential timer (cf. Corollary 5.1), we deduce the following corollary.

Corollary 5.2. *Assume that* $P(D < x) = 1 - \sum_{i=1}^{k} p_i \exp(-\xi_i x)$ *with* $\sum_{i=1}^{k} p_i = 1$. *Then, for* $\phi'(0) \neq 0$,

$$E[e^{-\alpha Z}]$$
$$= \left(\frac{\alpha}{\phi(\alpha)}E[L(T_K)] + \left(\frac{1}{\phi(\alpha)} - \frac{1}{\phi_*(\alpha)}\right)\right.$$
$$\left.\times \sum_{i=1}^{k} p_i\left(\frac{\xi_i}{\xi_i - \phi(\alpha)}E[e^{-\alpha Z_l}] - \frac{\phi(\alpha)}{\xi_i - \phi(\alpha)}E_{Z_l}[e^{-\xi_i\tau}] - 1\right)\right)$$
$$\times \left(\frac{1}{\phi'(0)}E[L(T_K)] + \left(\frac{1}{\phi'(0)} - \frac{1}{\phi_*'(0)}\right)\left(\phi'(0)\sum_{i=1}^{k}\frac{p_i}{\xi_i}(1 - E_{Z_l}[e^{-\xi_i\tau}]) - E[Z_l]\right)\right)^{-1},$$

with $E[L(T_K)]$, $E[Z_l]$, $E[\exp(-\alpha Z_l)]$, *and* $E_{Z_l}[\exp(-\xi_i\tau)]$ *given by* (3.9), (3.6), (3.5), *and* (3.14), *respectively.*

Finally, the expressions become more involved for the case in which the distribution of the timer involves the sum of two exponentials with the same intensity parameter. As an example, assume that $dP(D < x)/dx = \xi^2 x e^{-\xi x}$, i.e. the duration of the timer has an Erlang-2 distribution function. Instead of direct substitution into (3.17) and (3.20), we may use the results for the case of an exponential timer. More specifically, interchanging integral and differentiation, we have the relation $\int_0^\infty t e^{-\xi t} f(t)\,dt = -(d/d\xi)(\int_0^\infty e^{-\xi t} f(t)\,dt)$. Thus, in the case of an Erlang-2 timer, dividing (5.1) by ξ, taking derivatives with respect to ξ, and

finally multiplying by ξ^2 yields

$$\mathrm{E}[\mathrm{e}^{-\alpha Z(D \wedge \tau)}] = \frac{1}{\xi - \phi(\alpha)}\left(\frac{\xi^2}{\xi - \phi(\alpha)}\,\mathrm{E}[\exp(-\alpha Z_l)] - \xi\phi(\alpha)\,\mathrm{E}[\tau\mathrm{e}^{-\xi\tau}]\right.$$
$$\left. - \phi(\alpha)\frac{2\xi - \phi(\alpha)}{\xi - \phi(\alpha)}\,\mathrm{E}[\mathrm{e}^{-\xi\tau}]\right).$$

Similarly, from (5.3), we obtain the constant

$$\mathrm{E}[D \wedge \tau] = \frac{2}{\xi}(1 - \mathrm{E}[\mathrm{e}^{-\xi\tau}]) - \mathrm{E}[\tau\mathrm{e}^{-\xi\tau}].$$

The LST of the steady-state workload in the case of an Erlang-2 timer follows directly by substituting the above into (3.29).

Acknowledgements

References

[1] ALILI, L. AND KYPRIANOU, A. E. (2005). Some remarks on first passage of Lévy processes, the American put and pasting principles. *Ann. Appl. Prob.* **15,** 2062–2080.

[2] ALTMAN, E., BASAR, T. AND SRIKANT, R. (1997). Multi-user rate-based flow control with action delays: a team-theoretic approach. In *Proc. 36th IEEE Conf. Decision Control* (San Diego, 1997), IEEE, New York, pp. 2387–2392.

[3] ALTMAN, E., KOFMAN, D. AND YECHIALI, U. (1995). Discrete time queues with delayed information. *Queueing Systems* **19,** 361–376.

[4] ARTALEJO, J.R. (2001). On the M/G/1 queue with *D*-policy. *Appl. Math. Modelling* **25,** 1055–1069.

[5] ASMUSSEN, S. (2003). *Applied Probability and Queues*, 2nd edn. Springer, New York.

[6] AVRAM, F., KYPRIANOU, A. E. AND PISTORIUS, M. R. (2004). Exit problems for spectrally negative Lévy processes and applications to (Canadized) Russian options. *Ann. Appl. Prob.* **14,** 215–238.

[7] BEKKER, R. (2005). Queues with state-dependent rates. Doctoral Thesis, Department of Mathematics and Computer Science, Eindhoven University of Technology.

[8] BEKKER, R. (2007). Queues with Lévy input and hysteretic control. Preprint WS2008-1, Department of Mathematics, Vrije Universiteit Amsterdam.

[9] BEKKER, R., BOXMA, O. J. AND RESING, J.A.C. (2007). Lévy processes with adaptable exponent. Res. Rep. 2007-014, EURANDOM.

[10] BEKKER, R., BORST, S. C., BOXMA, O. J. AND KELLA, O. (2004). Queues with workload-dependent arrival and service rates. *Queueing Systems* **46,** 537–556.

[11] BERTOIN, J. (1996). *Lévy Processes*, Cambridge University Press.

[12] BERTOIN, J. (1997). Exponential decay and ergodicity of completely asymmetric Lévy processes in a finite interval. *Ann. Appl. Prob.* **7,** 156–169.

[13] BINGHAM, N. H. (1975). Fluctuation theory in continuous time. *Adv. Appl. Prob.* **7,** 705–766.

[14] BROCKWELL, P. J., RESNICK, S. I. AND TWEEDIE, R. L. (1982). Storage processes with general release rule and additive inputs. *Adv. Appl. Prob.* **14,** 392–433.

[15] COHEN, J. W. (1976). On the optimal switching level for an M/G/1 queueing system. *Stoch. Process. Appl.* **4,** 297–316.

[16] COHEN, J. W. (1982). *The Single Server Queue*. North-Holland, Amsterdam.

[17] DENTENEER, D. AND VAN LEEUWAARDEN, J. S. H. (2005). The delayed bulk service queue: a model for a reservation process. In *Proc. 19th ITC*, North-Holland, Amsterdam, pp. 909–918.

[18] DENTENEER, D., VAN LEEUWAARDEN, J. S. H. AND ADAN, I. J. B. F. (2007). The acquisition queue. *Queueing Systems* **56,** 229–240.

[19] DE TURCK, K. AND WITTEVRONGEL, S. (2005). Delay analysis of the go-back-*N* ARQ protocol over a time-varying channel. In *Proc. 2nd Europ. Perf. Eval. Workshop* (Lecture Notes Comput. Sci. **3670**), Springer, Berlin, pp. 124–138.

[20] DE VUYST, S., WITTEVRONGEL, S. AND BRUNEEL, H. (2004). Delay analysis of the stop-and-wait ARQ protocol over a correlated channel. In *Proc. HET-NETs '04*, pp. P21/1–P21/11.

[21] DSHALALOW, J. H. (1997). Queueing systems with state dependent parameters. In *Frontiers in Queueing: Models and Applications in Science and Engineering*, CRC, Boca Raton, FL, pp. 61–116.

[22] FEINBERG, E. A. AND KELLA, O. (2002). Optimality of *D*-policies for an M/G/1 queue with a removable server. *Queueing Systems* **42,** 355–376.

[23] GAVER, D. P. AND MILLER, R. G. (1962). Limiting distributions for some storage problems. In *Studies in Applied Probability and Management Science*, Stanford University Press, pp. 110–126.

[24] HARRISON, J. M. (1985). *Brownian Motion and Stochastic Flow Systems*. John Wiley, New York.

[25] JACOBSON, V. (1988). Congestion avoidance and control. In *Proc. ACM SIGCOMM 1988*, pp. 314–329.

[26] KELLA, O. (1998). An exhaustive Lévy storage process with intermittent output. *Stoch. Models* **14,** 979–992.

[27] KELLA, O. AND WHITT, W. (1992). Useful martingales for stochastic storage processes with Lévy input. *J. Appl. Prob.* **29,** 396–403.

[28] KYPRIANOU, A. E. AND PALMOWSKI, Z. (2005). A martingale review of some fluctuation theory for spectrally negative Lévy processes. In *Séminaire de Probabilités XXXVIII* (Lecture Notes Math. **1857**), Springer, Berlin, pp. 16–29.

[29] LEE, E. Y. AND AHN, S. K. (1998). P_λ^M-policy for a dam with input formed by a compound Poisson process. *J. Appl. Prob.* **35,** 482–488.

[30] LEE, J. AND KIM, J. (2007). Workload analysis of an M/G/1 queue under the P_λ^M policy with a set-up time. *Appl. Math. Modelling* **31,** 236–244.

[31] MALHOTRA, R., VAN HAALEN, R., MANDJES, M. R. H. AND NÚÑEZ-QUEIJA, R. (2005). Modeling the interaction of IEEE 802.3x hop-by-hop flow control and TCP end-to-end flow control. In *Proc. NGI 2005 Conf. Next Generation Internet Networks Traffic Eng.* (Rome).

[32] NGUYEN-NGOC, L. AND YOR, M. (2005). Some martingales associated to reflected Lévy processes. In *Séminaire de Probabilités XXXVIII* (Lecture Notes Math. **1857**), Springer, Berlin, pp. 42–69.

[33] PADHYE, J., FIROIU, V., TOWSLEY, D. AND KUROSE, J. (1998). Modeling TCP throughput: a simple model and its empirical validation. In *Proc. ACM SIGCOMM 1998*, ACM, New York, pp. 303–314.

[34] PRABHU, N. U. (1965). *Queues and Inventories*. John Wiley, New York.

[35] PRABHU, N. U. (1980). *Stochastic Storage Processes*. Springer, New York.

[36] SHARMA, V. (2001). Queues with service rate controlled by a delayed feedback. *Queueing Systems* **39,** 303–315.

[37] TIJMS, H. C. (1976). Optimal control of the workload in an M/G/1 queueing system with removable server. *Math. Operat. Statist.* **7,** 933–944.