A First Passage Time Problem for Spectrally Positive Lévy Processes and Its Application to a Dynamic Priority Queue

Vahid Sarhangian[a], Barış Balcıoğlu[b,*]

[a] *University of Toronto, Joseph L. Rotman School of Management, 105 St. George Street, Toronto, M5S 3E6, Canada*
[b] *Sabancı University, Faculty of Engineering and Natural Sciences, Orhanlı-Tuzla, 34956 Istanbul, Turkey*

Abstract

We study a first passage time problem for a class of spectrally positive Lévy processes. By considering the special case where the Lévy process is a compound Poisson process with negative drift, we obtain the Laplace-Stieltjes transform of the steady-state waiting time distribution of low-priority customers in a two-class $M/GI/1$ queue operating under a dynamic non-preemptive priority discipline. This allows us to observe how the waiting time of customers is affected as the policy parameter varies.

Keywords: Lévy processes, Martingales, First passage time, Priority queues, Dynamic priority, Due-date scheduling

1. Introduction

Consider a two-class $M/GI/1$ queueing system with static non-preemptive priority (SNPP) in favor of class 1 customers. A new class 2 arrival has to wait for the total work she finds in the system as well as the work input by subsequent class 1 arrivals to be processed before she can start service. It follows that the steady-state waiting time of class 2 customers can be represented as the *first passage time* of class 1 workload to 0, initiated by the steady-state workload of the system [3]. While the static policy ensures the minimum possible wait for class 1 customers (when preemption is not allowed), it can lead to excessively long waiting times for class 2 customers, especially when the relative arrival rate of class 1 customers is high. To address this issue, one can consider a more flexible policy referred to as a *dynamic priority discipline*; if a class 2 customer has not started service by a pre-determined time threshold, which we shall call the *policy parameter* and denote by u, she is upgraded to class 1 and does no

*Corresponding author
Email address: `balcioglu@sabanciuniv.edu` (Barış Balcıoğlu)

longer have to wait for future class 1 arrivals who arrive during her remaining queueing time. In this case, the steady-state waiting time of class 2 customers can similarly be viewed as the first passage time of class 1 workload to 0 initiated by the steady-state workload, but where the class 1 input is *turned off* after u time units (see Section 3).

In this paper, we study the distribution of this first passage time under a general setting where the *net input* could be any spectrally positive Lévy process. By considering the special case where the net input is a compound Poisson process with a unit negative drift, we obtain the Laplace-Stieltjes Transform (LST) of the steady-state waiting time of class 2 customers in the dynamic priority queue. This allows us to observe, both analytically and numerically, how the waiting time of class 2 customers is affected as the policy parameter varies. We observe that while the mean and the variance of the class 2 waiting time are strictly increasing in the policy parameter u, the tail probability or the proportion of customers who wait longer than a certain threshold, say t_0, remains unchanged for all $u \geq t_0$. Clearly, the policy also affects the waiting time of class 1 customers, but that is not studied here; see the discussion in Subsection 3.4.

The dynamic priority discipline is appealing since it enables the decision maker to assign priorities not only by customer type, but also by the amount of wait a customer has experienced in the system. Furthermore, this is done through a controllable parameter which can be adjusted to manipulate the waiting time of both classes. The policy therefore finds applications in healthcare and customer contact centers where usually different levels of service must be provided for different types of customers.

Despite its wide application area, however, the literature on dynamic prioritization is limited. The policy described above, also known as the due-date scheduling, was originally introduced by Jackson [12, 13, 14] using *urgency numbers*, denoted by u_i, for a class i customer. Accordingly, a class i customer arriving at time t is assigned $t + u_i$ as its attribute. At each service completion epoch, the next customer to be served is the customer with the lowest attribute. In a two-class system letting $u \equiv u_2 - u_1$ ($u_2 \geq u_1$), the policy is the same as assigning t ($t + u$) as an attribute for class 1 (2) customers when they arrive at time t. Then, it is easy to see that for $u = 0$, the policy reduces to the first-come, first-served (FCFS) discipline and for $u = \infty$ the queue operates under the SNPP rule. By altering u, hence, one can control the priority assignment rule. Jackson studied a discrete-time queue and conducted simulation experiments for the continuous system. Bounds for the steady-state mean waiting times were obtained by Holtzman [11] for a queue with Poisson arrivals and multiple priority classes. Goldberg [9] analyzed the virtual waiting time process for class 1 and class 2 customers and derived an expression for the mean steady-state waiting times of each class in the same queueing system we consider in this paper. Then, Goldberg [10] provided a proof for Jackson's conjecture [13] on the shape of the tails of the waiting time distributions. Prabhu and Reseer [18] studied a sequence of two-class queueing systems with u ranging from 0 to ∞, and characterized the transient as well as the steady-state waiting time distribution of each class when u is a random variable (r.v.) with an exponential density on $(0,\infty)$ and discrete atoms at 0 and ∞. Koole et al. [17] recently considered a different implementation of the priority rule which is to assign customers of each class to a separate queue, and keep track of the waiting time of the first in line (FIL) customer in both queues. In this case, letting $W_i^{\mathrm{FIL}}(s)$ denote the waiting time of the FIL customer in queue i at time s, when a service is completed at time s_0, the next customer to be served is chosen from the queue with the higher value of $W_i^{\mathrm{FIL}}(s_0)-u_i$ (which is the customer with the lowest attribute in both queues). Koole et al. [17] introduced a discrete-time approximation to describe the waiting times of FIL customers and used it to approximate the waiting time distributions of class 1 and class 2 customers in an $M/M/c$ queue. There does not seem to be any other results for the distribution of the waiting times under the dynamic priority discipline. The importance of waiting-time distribution is however evident in many service systems such as contact centers where the service levels are expressed as the proportion of customers who wait longer than a certain time threshold [8].

In the next section, we present our main result for the distribution of the first passage time with Lévy net input. In Section 3, we use this result to obtain the LST of the distribution of class 2 waiting time in the dynamic priority queue and study the effect of changing the parameter on its mean, variance and tail probabilities.

2. The first passage time problem with Lévy net input

Let $X \equiv \{X(t); t \geq 0\}$ be a càdlàg (right-continuous with left limits) Lévy process with respect to some filtration satisfying the usual conditions (augmented by null sets and right-continuous), that is, X has independent and stationary increments, and $X(0)=0$ (for background on Lévy processes see e.g., [7] or chapter 1 of [5]). We further assume that X is spectrally positive (has no negative jumps) and is not non-decreasing (subordinator or identically zero). Let $\varphi(\theta) \equiv \log \mathbb{E}[e^{-\theta X(1)}]$, be the Laplace-Stieltjes exponent of X. Then it is well known that $\varphi(\theta)$ is convex, finite for all $\theta \geq 0$ and $\varphi(\theta) \to \infty$ as $\theta \to \infty$; see e.g., [7].

For a given $x, u > 0$ consider the process $Y \equiv \{Y(t); t \geq 0\}$ with

$$Y(t) = x + X(t \wedge u) - (t-u)^+, \tag{1}$$

where $a \wedge b = \min(a,b)$ and $a^+ = \max(a,0)$. When positive, Y behaves like a storage process which, starting from x, receives the net input X for u time units and then depletes linearly with unit rate. If X is comprised of a Lévy input $I \equiv \{I(t); t \geq 0\}$ (e.g., a Brownian motion or compound Poisson process) and a linear output with unit rate, i.e., $X(t) = I(t) - t$, then Y reduces to

$$Y(t) = x + I(t \wedge u) - t, \tag{2}$$

that is the input I is turned off after u time units. We are interested in the LST of the distribution of the stopping time $T_x \equiv \inf\{t \geq 0; Y(t) = 0\}$ or the first passage time of Y to zero.

Consider the stopping time $\tau_x \equiv \inf\{t \geq 0; x + X(t) = 0\}$ and define

$$\alpha_x(\theta) \equiv \mathbb{E}[e^{-\theta\tau_x} 1\{\tau_x \leq u\}] = \int_0^u e^{-\theta y} d\mathbb{P}(\tau_x \leq y). \tag{3}$$

Theorem 1. *The LST of the distribution of T_x, i.e., the first passage time of Y to zero is given by*

$$\mathbb{E}[e^{-\theta T_x}] = e^{-u(\theta-\varphi(\theta))}\Big(e^{-\theta x} - \alpha_x\big(\varphi(\theta)\big)\Big) + \alpha_x(\theta), \tag{4}$$

with $\alpha_x(\theta)$ defined in Eq. (3).

Proof. Setting $J(t) \equiv 1\{t \leq u\}$ one can write Y as

$$Y(t) = x + \int_{(0,t]} J(s)dX(s) - \int_{(0,t]} (1 - J(s))ds,$$

where the second term on the right is a stochastic integral interpreted in the Itô sense. Then from Corollary 5.2.2 and Theorem 5.2.4 of [5] (pages 253-254) (see [16] for a multidimensional version), it is known that $M \equiv \{M(t); t \geq 0\}$ with

$$M(t) = e^{-\theta Y(t) - \varphi(\theta)\int_0^t J(s)ds - \theta\int_0^t (1-J(s))ds} = e^{-\theta Y(t) - \varphi(\theta)(t\wedge u) - \theta(t-u)^+},$$

is a Martingale. We proceed by applying the optional stopping theorem to the bounded stopping time $T_x \wedge \upsilon$ for some $\upsilon > 0$. Letting $\upsilon \to \infty$ and using the bounded convergence theorem, we arrive at

$$\mathbb{E}[M(T_x)] = \mathbb{E}[e^{-\theta Y(T_x) - \varphi(\theta)(T_x\wedge u) - \theta(T_x-u)^+}] = \mathbb{E}[M(0)] = e^{-\theta x}. \tag{5}$$

Noting that since u is finite, $\mathbb{P}(T_x < \infty) = 1$ and hence $Y(T_x) = 0$ with probability 1, Eq. (5) yields

$$\mathbb{E}[e^{-\varphi(\theta)(T_x\wedge u) - \theta(T_x-u)^+}] = e^{-\theta x}. \tag{6}$$

Next, observe that

$$\mathbb{E}[e^{-\varphi(\theta)(T_x\wedge u) - \theta(T_x-u)^+}] = \mathbb{E}[e^{-\varphi(\theta)T_x} 1\{T_x \leq u\}] + \\ e^{-u(\varphi(\theta)-\theta)}\mathbb{E}[e^{-\theta T_x} 1\{T_x > u\}]. \tag{7}$$

Furthermore,

$$\mathbb{E}[e^{-\theta T_x} 1\{T_x > u\}] = \mathbb{E}[e^{-\theta T_x}] - \mathbb{E}[e^{-\theta T_x} 1\{T_x \leq u\}]. \tag{8}$$

Substituting from Eqs. (8) and (6) in Eq. (7) after some algebra we get

$$\mathbb{E}[e^{-\theta T_x}] = e^{-u(\theta-\varphi(\theta))}(e^{-\theta x} - \mathbb{E}[e^{-\varphi(\theta)T_x} 1\{T_x \leq u\}]) + \\ \mathbb{E}[e^{-\theta T_x} 1\{T_x \leq u\}]. \tag{9}$$

Now, observe that

$$T_x = \tau_x \wedge u + (x + X(u))1\{\tau_x > u\}. \quad (10)$$

Thus, $T_x \le u$ if and only if $\tau_x \le u$ in which case $T_x = \tau_x$. It follows that,

$$\mathbb{E}[e^{-\theta T_x} 1\{T_x \le u\}] = \mathbb{E}[e^{-\theta \tau_x} 1\{\tau_x \le u\}] \equiv \alpha_x(\theta),$$

and thus we can rewrite Eq. (9) to obtain the LST of T_x as given in Eq. (4). □

Note that the LST of the distribution of τ_x is known e.g., from Kella and Whitt [15] to be

$$\mathbb{E}[e^{-\theta \tau_x}] = e^{-\psi(\theta)x},$$

where $\psi(\beta) \equiv \inf\{\theta; \varphi(\theta) > \beta\}$ for $\beta \ge 0$ is the right inverse of $\varphi(\theta)$. Therefore, in principal one has the distribution of τ_x and hence the value of $\alpha_x(\theta)$. If instead of u we set $U \sim \exp(\gamma)$ then $\alpha_x(\theta) = \mathbb{E}[e^{-(\theta+\gamma)\tau_x}] = e^{-\psi(\theta+\gamma)x}$. For a fixed u, $\alpha_x(\theta)$ can be computed numerically.

3. Application: the dynamic priority queue

3.1. The class 2 waiting time distribution

In this subsection, we use the above result to obtain the LST of the distribution of the steady-state waiting time of class 2 customers W_2, in the dynamic priority queue described in the introduction. For a class i customer, $i = 1, 2$, let λ_i denote the Poisson arrival rate, and $b_i, b_i^{(2)}$, and $\widetilde{b}_i(\theta)$ the mean, the second moment, and the LST of the cumulative distribution function (CDF) of the service time, respectively. In addition, define $\rho_i \equiv \lambda_i b_i$ and $\lambda \equiv \lambda_1 + \lambda_2$, and denote the workload process by $\{V(t) : t \ge 0\}$. That is, $V(t)$ is the total amount of work in the system at time t, including the work submitted by both classes of customers. Assuming $\rho \equiv \rho_1 + \rho_2 < 1$, as $t \to \infty$, the distribution of $V(t)$ converges to that of the steady-state workload V, i.e., $\lim_{t\to\infty} P(V(t) \le x) = P(V \le x)$ for all $x \ge 0$. Note that V has a discrete atom at 0 and a continuous density on $(0, \infty)$. The LST of the CDF of V for any work-conserving and non-idling service discipline is given by (e.g., [19])

$$\tilde{v}(\theta) \equiv \mathbb{E}[e^{-\theta V}] = \frac{(1-\rho)\theta}{\theta - \lambda + \lambda_1 \widetilde{b}_1(\theta) + \lambda_2 \widetilde{b}_2(\theta)}.$$

Consider a tagged class 2 customer arriving in the steady-state. Then the customer has to wait for the total work she finds in the system plus the amount of work brought by class 1 customers in the next u time units, before she starts receiving service [9]. Since Poisson Arrivals See Time Averages (PASTA) [20] the amount of work observed by the customer upon arrival is the steady-state workload V. Thus, the steady-state waiting time of class 2 customers W_2, is the first passage time of class 1 workload to 0, which is initiated by the steady-state workload V, and is turned off after u time units.

It is now easy to see, in light of Eq. (2), that the above is a special case of the problem studied in Section 2 where the net input is a compound Poison process (with rate λ_1 and jump size distribution having LST $\widetilde{b}_1(\theta)$) with a unit negative drift. Also the initiating level is a random variable with LST $\tilde{v}(\theta)$. Denote by $\{A(t); t \geq 0\}$ the Poisson process associated with class 1 arrivals, and by $\{S_n; n \geq 1\}$ the sequence of independent and identically distributed (i.i.d) class 1 service times. If we set the net input process to $X(t) = \sum_{i=1}^{A(t)} S_i - t$, in Eq. (1) then we have

$$Y(t) = x + \sum_{i=1}^{A(t \wedge u)} S_i - t,$$

and hence $T_x \equiv \inf\{t \geq 0; Y(t) = 0\}$ is the first passage time of class 1 workload to zero starting from level x and with input turned off after u time units. It follows for the LST of the distribution of W_2 that

$$\mathbb{E}[e^{-\theta W_2}] = \int_0^\infty \mathbb{E}[e^{-\theta T_x}] d\mathbb{P}(V \leq x). \tag{11}$$

For this case the exponent of X can be easily verified to be $\varphi(\theta) = \lambda_1(\widetilde{b}_1(s) - 1) + \theta$. Also the LST of $\tau_x \equiv \inf\{t \geq 0 : x + X(t) = 0\}$ is readily available by observing that it is the first passage time to 0 for class 1 workload starting from level x. We have (e.g., [3])

$$\mathbb{E}[e^{-\theta \tau_x}] = e^{-x(\theta + \lambda_1 - \lambda_1 \widetilde{g}(\theta))},$$

where $\widetilde{g}(\theta)$ is the solution to the Kendall functional equation $\widetilde{g}(\theta) = \widetilde{b}_1(\theta + \lambda_1 - \widetilde{g}(\theta))$, and can be computed by iteration; see [1]. If we replace x with the random variable V then

$$\widetilde{f}(\theta) \equiv \mathbb{E}[e^{-\theta \tau_V}] = \mathbb{E}[e^{-(\theta + \lambda_1 - \lambda_1 \widetilde{g}(\theta))V}] = \tilde{v}(\theta + \lambda_1 - \lambda_1 \widetilde{g}(\theta)), \tag{12}$$

which is the LST of the waiting time distribution of class 2 customers under the SNPP policy.

Figure 1 depicts a sample path of $x + X(t)$ and $Y(t)$ and the associated realization of stopping times T_x, τ_x. Marks on the time axis indicate class 1 arrivals. Note that sample paths are identical up until the fourth class 1 arrival which happens after u time units pass from the tagged class 2 arrival at time 0.

We are now in the position to present the LST of the distribution of W_2. First, from Eq. (4) we have

$$\mathbb{E}[e^{-\theta T_x}] = e^{-u\left(-\lambda_1(\widetilde{b}_1(\theta) - 1)\right)} \left(e^{-\theta x} - \alpha_x\big(\lambda_1(\widetilde{b}_1(\theta) - 1) + \theta\big)\right) + \alpha_x(\theta).$$

Removing the condition on $V = x$, via Eq. (11) we obtain the result stated in the following corollary.

Corollary 1. *The LST of the CDF of the class 2 waiting time is given by*

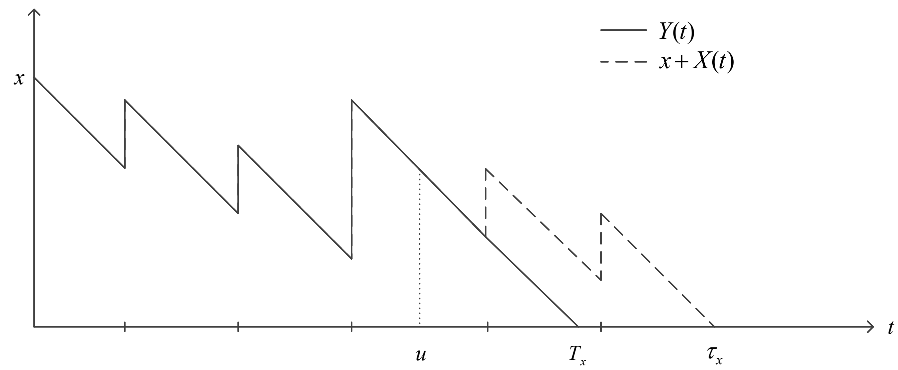

Figure 1: A sample path of $x + X(t)$ and $Y(t)$.

$$\widetilde{w}_2(\theta) \equiv \mathbb{E}[e^{-\theta W_2}] = e^{-u(-\lambda_1(\widetilde{b}_1(\theta)-1))}\Big(\tilde{v}(\theta) - \alpha_V\big(\lambda_1(\widetilde{b}_1(\theta)-1)+\theta\big)\Big) + \alpha_V(\theta), \tag{13}$$

where $\alpha_V(\theta)$ is the restricted transform

$$\alpha_V(\theta) = \int_0^u e^{-\theta y} d\mathbb{P}(\tau_V \le y). \tag{14}$$

Given the above result one can compute the complimentary distribution or tail probabilities of W_2 by numerically inverting $(1-\widetilde{w}_2(\theta))/\theta$, e.g., using the methods due to Abate and Whitt [2]. This, however, requires evaluating the restricted transform which in turn requires computing $d\mathbb{P}(\tau_V \le y)$. In what follows we shall give an expression for $\alpha_V(\theta)$ based on $\mathbb{P}(\tau_V > y)$ instead of $d\mathbb{P}(\tau_V \le y)$. This way since $\mathbb{P}(\tau_V > y) < 1$ for all $y > 0$, the error bound for the discretization error of the inversion algorithm given in Abate and Whitt [2] applies; see also Abate et al. [4] for more details.

Consider random variable $\tau_V \wedge u$ and observe that

$$\begin{aligned} \mathbb{E}[e^{-\theta(\tau_V \wedge u)}] &= \int_0^u e^{-\theta y} d\mathbb{P}(\tau_V < u) + \int_u^\infty e^{-\theta u} d\mathbb{P}(\tau_V < u) \\ &= \alpha_V(\theta) + e^{-\theta u}\mathbb{P}(\tau_V > u). \end{aligned} \tag{15}$$

Furthermore,

$$\int_0^{\tau_V \wedge u} \theta e^{-\theta y} dy = 1 - e^{-\theta(\tau_V \wedge u)}.$$

Taking the expectation of both sides, we have

$$\mathbb{E}[e^{-\theta(\tau_V \wedge u)}] = 1 - \theta \int_0^u e^{-\theta y}\mathbb{P}(\tau_V > y) dy,$$

which together with Eq. (15), gives

$$\alpha_V(\theta) = 1 - \theta \int_0^u e^{-\theta y}\mathbb{P}(\tau_V > y)dy - e^{-\theta u}\mathbb{P}(\tau_V > u). \quad (16)$$

Given Eq. (16) and having $\widetilde{f}(\theta)$, i.e., the LST of the CDF of τ_V (see Eq. (12)) one can compute $\mathbb{P}(\tau_V > y)$ for any $y > 0$ by numerically inverting $(1 - \widetilde{f}(\theta))/\theta$, and use it together with numerical integration to evaluate the restricted transform.

Finally, we mention that our results are also applicable to the multiserver $M/M/c$ queue with c identical servers where service times for both classes are exponentially distributed with rate μ, i.e., $1/b_1 = 1/b_2 = \mu$. In this case, the total workload has the same distribution as the waiting time of customers in a single class FCFS $M/M/c$ queue with arrival rate λ and service rate μ and $\widetilde{g}(\theta)$ is the LST of the busy period distribution in an $M/M/1$ queue with arrival rate λ_1 and service rate $c\mu$, which is available in closed form (see e.g., [6], page 105).

3.2. Monotonicity results

In this subsection, we discuss the monotonicity of the expected value, variance and tail probability of W_2 with respect to the policy parameter u.

First, by differentiating Eq. (13) we have

$$\mathbb{E}[W_2] = \mathbb{E}[V] + \rho_1 \int_0^u \mathbb{P}(\tau_V > t)dt, \quad (17)$$

which is the same result given in [9]. Clearly, we have $d\mathbb{E}[W_2]/du = \rho_1\mathbb{P}(\tau_V > u) > 0$, and hence the expected waiting time for class 2 customers is strictly increasing in u.

Next we obtain the second moment by evaluating the second derivative of $\widetilde{w}_2(\theta)$ at $\theta = 0$. After some algebra, we get

$$\begin{aligned}\mathbb{E}[W_2^2] = \widetilde{w}_2''(0) = \mathbb{E}[V^2] + 2u\rho_1\mathbb{E}[V] + \alpha_V''(0)\rho_1(2-\rho_1) \\ + \alpha_V'(0)\big(2u\rho_1(1-\rho_1) - \lambda_1 b_1^{(2)}\big) + (1-\alpha_V(0))\left(\lambda_1 u b_1^{(2)} + u^2\rho_1^2\right). \quad (18)\end{aligned}$$

Note that from Eq. (14), $1 - \alpha_V(0) = \mathbb{P}(\tau_V > u)$. Also, using integration by parts we have

$$\begin{aligned}\alpha_V'(0) &= u\mathbb{P}(\tau_V > u) - \int_0^u \mathbb{P}(\tau_V > t)dt, \\ \alpha_V''(0) &= -u^2\mathbb{P}(\tau_V > u) + 2\int_0^u t\mathbb{P}(\tau_V > t)dt,\end{aligned}$$

which are substituted in Eq. (18), and after some simplification, we obtain

$$\begin{aligned}\mathbb{E}[W_2^2] = \mathbb{E}[V^2] + 2u\rho_1\mathbb{E}[V] + \big(\lambda_1 b_1^{(2)} - 2u\rho_1(1-\rho_1)\big)\int_0^u \mathbb{P}(\tau_V > t)dt \\ + 2\rho_1(2-\rho_1)\int_0^u t\mathbb{P}(\tau_V > t)dt. \quad (19)\end{aligned}$$

Having the first moment from Eq. (17), from $\mathrm{Var}[W_2] = \mathbb{E}[W_2^2] - \mathbb{E}[W_2]^2$, we obtain

$$\begin{aligned}\mathrm{Var}[W_2] = \mathrm{Var}[V] + 2\rho_1 u\mathbb{E}[V] + \\ \big(\lambda_1 b_1^{(2)} - 2u\rho_1(1-\rho_1) - 2\rho_1\mathbb{E}[V]\big)\int_0^u \mathbb{P}(\tau_V > t)dt - \\ \big(\rho_1 \int_0^u \mathbb{P}(\tau_V > t)dt\big)^2 + 2\rho_1(2-\rho_1)\int_0^u t\mathbb{P}(\tau_V > t)dt. \quad (20)\end{aligned}$$

To show that $\mathrm{Var}[W_2]$ is also strictly increasing in u, we differentiate the former with respect to u. Then, after simplifying we have

$$\begin{aligned}\frac{d}{du}\mathrm{Var}[W_2] &= 2\rho_1\big(1 - \mathbb{P}(\tau_V > u)\big)\mathbb{E}[W_2] - \\ &\quad 2\rho_1\big(\int_0^u \mathbb{P}(\tau_V > t)dt - u\mathbb{P}(\tau_V > u)\big) + \lambda_1 b_1^{(2)}\mathbb{P}(\tau_V > u) \\ &> 0,\end{aligned}$$

where the inequality follows from the fact that $(1 - \mathbb{P}(\tau_V > u))\,\mathbb{E}[W_2] > \int_0^u \mathbb{P}(\tau_V > t)dt - u\mathbb{P}(\tau_V > u)$. To see this, note that for $u = 0$ the inequality holds, also both sides are strictly increasing in u and converge to $\mathbb{E}[\tau_V]$ as $u \to \infty$.

Similarly one might expect the tail probability $\mathbb{P}(W_2 > t)$ to be also strictly increasing in u. However, From Eq. (10), we can see that $T_x > t_0$ if and only if $\tau_x > t_0$ and hence $\mathbb{P}(W_2 > t_0) = \mathbb{P}(\tau_V > t_0)$ for any $t_0 \leq u$. Since the distribution of τ_V, as given in Eq. (12), is independent of u, it follows that for a fixed $t = t_0$, the tail probability $\mathbb{P}(W_2 > t_0)$ remains identical for any u such that $u \geq t_0$. We also demonstrate this via numerical examples in the following subsection.

3.3. Numerical example

In this subsection, using a numerical example, we demonstrate some of our findings.

We consider a multiserver system with $c = 8$ servers and exponentially distributed service times with $\mu = 1$. Poisson arrival rates for classes 1 and 2 are $\lambda_1 = 3.5$ and $\lambda_2 = 2.5$, respectively. Table 1 lists the mean and variance of class 2 waiting times computed for different values of u. One can see that they are both increasing in u.

In Figure 2 (a) we plot the complimentary distribution function of class 2 waiting times for different values of u. As shown analytically in Subsection 3.2, and also observed in the figure, changing the policy parameter u does not affect the distribution in all percentiles. For instance, we observe that a service level in form of $P(W_2 > 0.5)$ remains identical for u =1, 2, and ∞. Thus, to improve the service level at level 0.5, a policy parameter $u < 0.5$ must be considered. We also used simulation to estimate the tail probabilities for class 1 waiting times. Figure 2 (b) presents the point estimates which shows that, contrary to what we observe for class 2 customers, changing the parameter u affects all parts of the

u	0	0.5	1.0	1.5	2.0	2.5	3.0	∞
$\mathbb{E}[W_2]$	0.178	0.234	0.265	0.285	0.297	0.305	0.310	0.317
$\text{Var}[W_2]$	0.147	0.223	0.298	0.364	0.417	0.459	0.492	0.575

Table 1: Mean and variance of class 2 waiting times for different values of u

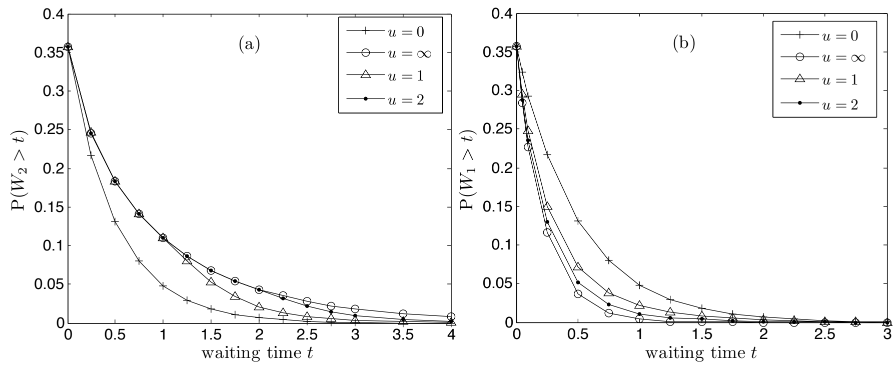

Figure 2: Complimentary distribution function of the waiting time of (a) class 2 and (b) class 1 customers for different values of u

waiting time distribution. Hence, for example, even though changing u from 2 to ∞ does not affect the service level for class 2 customers at level 0.5, it improves the performance for class 1 customers.

3.4. Discussions

Here, we only characterize the class 2 waiting time distribution. In Goldberg [9] an expression for the mean waiting time of class 1 customers is derived using a simple probabilistic argument. It remains for future research to characterize the distribution of class 1 customers for a given policy parameter.

Regarding applicability, we should mention that the class 2 waiting time distribution is sufficient for a decision maker who, given a fixed capacity, wants to provide the lowest possible wait to class 1 customers while assuring a certain service level for class 2 customers. If for some u, the desired service level is attained for class 2 customers then the same service level is clearly guaranteed for class 1 customers as well. To have an estimate of the tail probabilities for class 1 customers in the case of exponential service times one can use the approximation developed in Koole et al. [17].

Acknowledgements

We would like to thank the anonymous referee for his/her invaluable comments. We are grateful to the associate editor who brought to our attention the possibility of generalizing our results to include Lévy processes and provided helpful and detailed comments regrading this generalization. This work was supported in part by Natural Sciences and Engineering Research Council (NSERC) of Canada.

References

[1] J. Abate and W. Whitt. 1992. “Solving probability transform functional equations for numerical inversion”, *Operations Research Letters*, Vol. 12, 275–281.

[2] J. Abate and W. Whitt. 1995. “Numerical inversion of Laplace transforms of probability distributions”, *ORSA Journal of Computing*, Vol. 7, 36–43.

[3] J. Abate and W. Whitt. 1997. “Asymptotics for M/G/1 low-priority waiting-time tail probabilities”, *Queueing Systems*, Vol. 25, 173–233.

[4] J. Abate, G. L. Choudhury, W. Whitt. 2000. “An introduction to numerical transform inversion and its application to probability models”, in: W.K. Grassmann (Ed.), Computational Probability, Kluwer Academic Publishers, Boston, MA, 257–323.

[5] D. Applebaum, *Lévy Processes and Stochastic Calculus*, second ed, Cambridge: Cambridge University Press, 2009.

[6] S. Asmussen. 2003. *Applied Probability and Queues*. 2nd edition, Springer-Verlag, New York.

[7] N. H. Bingham. 1975. “Fluctuation theory in continuous time”, *Advances in Applied Probability*, Vol. 7, No. 4, 705–766.

[8] N. Gans, G. M. Koole, and A. Mandelbaum. 2003. “Telephone call centers: tutorial, review, and research prospects”, *Manufacturing & Service Operations Management*, Vol. 5, 79–141.

[9] H. M. Goldberg. 1977. “Analysis of the earliest due date scheduling rule in queueing systems”, *Mathematics of Operations Research*, Vol. 2, No. 2, 145–154.

[10] H. M. Goldberg. 1980. “Jackson’s conjecture on earliest due date scheduling”, *Mathematics of Operations Research*, Vol. 5, No. 3, 460–466.

[11] J. M. Holtzman. 1971. “Bounds for a dynamic priority queue”, *Operations Research*, Vol. 19, 461–468.

[12] J. R. Jackson. 1960. “Some problems in queueing with dynamic priorities”, *Naval Research Logistics Quarterly*, Vol. 7, 235–249.

[13] J. R. Jackson 1961. “Queues with dynamic priority discipline”, *Management Science*, Vol. 8, 18–34.

[14] J. R. Jackson. 1962. “Waiting time distributions for queues with dynamic priorities”, *Naval Research Logistics Quarterly*, Vol. 9, 31–36.

[15] O. Kella and W. Whitt. 1992. "Useful martingales for stochastic storage processes with Lévy input, *Journal of Applied Probability*, Vol. 29, No. 2, 396–403.

[16] O. Kella, and O. Boxma. 2013. "Useful martingales for stochastic storage processes with Lévy-type input", *Journal of Applied Probability*, Vol. 50, No. 2, 439–449.

[17] G. M. Koole , B. F. Nielsen, and T. B. Nielsen. 2012. "First in line waiting times as a tool for analyzing queueing systems", *Operations Research*, Vol. 5, 1258–1266.

[18] N. U. Prabhu and P. K. Reeser. 1985. "A random family of queueing systems with a dynamic priority discipline", *Mathematics of Operations Research*, Vol. 10, No. 1, 24–32.

[19] H. Takagi, T. Takine, and O. J. Boxma. 1992. "Distribution of the workload in multiclass queueing systems with server vacations", *Naval Research Logistics*, Vol. 39, 41–52.

[20] R. W. Wolff. 1982. Poisson Arrivals See Time Averages. *Operations Research*, Vol. 30, No. 2, 223–231.