International Journal of Pure and Applied Mathematics
Volume 109 No. 2 2016, 325-370
ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version)
url: http://www.ijpam.eu
doi: 10.12732/ijpam.v109i2.12

ijpam.eu

ANALYSIS OF THE RESPONSE AND WAITING TIMES IN THE M/M/m LCFS PREEMPTIVE-RESUME PRIORITY QUEUE

Hideaki Takagi
Professor Emeritus, University of Tsukuba
Tsukuba Science City, Ibaraki 305-8573, JAPAN

Abstract: We study the response and waiting times in an M/M/m LCFS preemptive-resume priority queueing system following a "First-Come, First-Displaced" (FCFD) displacement policy in the steady state. In particular, we analyze the interactions of intermittent waiting times and service times during the response time of a tagged customer of each priority class, whose service may be preempted by the arrival of equal- or higher-priority-class customers. Numerical examples are provided to demonstrate the computation of the theoretical formulas and to compare the performance results with the M/M/m FCFS preemptive-resume priority queue following a "Last-Come, First-Displaced" (LCFD) displacement policy.

AMS Subject Classification: 60K25, 68M20, 90B22
Key Words: priority queue, multiserver, preemptive-resume, last-come first-served, first-come first-displaced, response time, waiting time, time in limbo, completion time, first passage time, absorbing state

1. Introduction

This study considers a queueing system consisting of m servers and a waiting room of infinite capacity, accommodating customers belonging to one of several priority classes. The arrival of class-p customers follows a Poisson process with a rate λ_p (> 0) that is independent of the arrival of customers in any of the

Received: June 29, 2016
Published: September 10, 2016

other classes. Upon arrival, every customer requests a service time which has an exponential class-independent distribution with mean $1/\mu$. Classes are indexed $1, 2, \ldots$ in order of decreasing priority, so that a class-p customer has priority for delivery of service over all customers of classes $p+1, p+2, \ldots$.

The arrival of a class-p customer entails three possibilities:

- Unless all servers are busy, service to this customer is started immediately.
- If all servers are busy serving customers belonging to higher-priority classes than p, this customer waits at the top of the queue of other class-p waiting customers.
- If all servers are already busy serving customers, at least one of whom being in a class of priority no higher than p, we define q ($\geq p$) as the lowest-priority class among all the customers being served at that moment. The service to one customer of class q is then preempted and that customer is displaced from the service facility to the top of the waiting room. The criterion for selecting this class-q customer is that his service was started or resumed before all the other class-q customers receiving service at that moment. Service to the incoming class-p customer then starts. This protocol for selecting which customer to displace is called *First-Come, First-Displaced* (FCFD) by Katzschner [6] and Fujiki [5].

As soon as a server becomes available, one of the highest-priority customers in the waiting room is called in to be served. Within a given class, a customer is chosen on a *Last-Come, First-Served* (LCFS) basis. When service resumes, a new sample of the service time is generated from the exponential distribution with mean $1/\mu$, regardless of the amount of service previously delivered to that customer. We may therefore call our system an "M/M/m immediate preemptive-resume priority queue with an LCFS service and FCFD displacement discipline within the same class." The phrase "immediate preemptive" emphasizes that the service, being delivered to a class-p customer, may be preempted by the arrival of another class-p customer when all servers are busy serving customers, including those of class p. Our protocol reduces to the one called "preemptive *Last-In, First-Out* (LIFO)" for a single-class M/G/1 queue by Wolff [14, p. 456]. Katzschner [6] mentions that "the FCFD discipline will be useful when the importance of a call decreases with the elapsing time. $\cdots$ Typical applications of this discipline are possible in the field of flight-traffic control and weather service."

This protocol for service preemption by a customer of the same class differs from the one assumed by Durr [4] in his study of an M/G/1 LCFS preemptive-resume priority queue. In his model (see also [8, p. 346]), the on-going service to a class-p customer can be preempted only by the arrival of a customer in a strictly higher-priority class $1, 2, \ldots, p-1$. We provide modification of his analysis of the M/G/1 queue for our preemption protocol in the Appendix.

The purpose of this paper is to derive explicit formulas for the mean and for some higher moments of the response and waiting times of an arbitrary customer from each priority class in the steady-state system. To the best of our knowledge, this is the first analysis done on this type of queueing system. The response and waiting times of customers in the M/M/m preemptive-resume priority queue with *First-Come, First-Served* (FCFS) service within the same class and *Last-Come, First-Displaced* (LCFD) displacement were studied by Brosh [1], Segal [7], Buzen and Bondi [2], Tatashev [11], Zeltyn et al. [15], and the present author [9, 10]. This paper refines the analysis presented in [10] by introducing more details in the relation between the received service time and the suspended waiting time.

We first define the quantities

$$\rho_p := \frac{\lambda_p}{m\mu} \quad ; \quad \lambda_p^+ := \sum_{k=1}^{p} \lambda_k \quad ; \quad \rho_p^+ := \sum_{k=1}^{p} \rho_k = \frac{\lambda_p^+}{m\mu}$$

for $p = 1, 2, \ldots$, while $\lambda_0^+ = \rho_0^+ = 0$.

Various decompositions of the response time for a tagged class-p customer are depicted in Fig. 1. The symbols in the figure are explained in the text where they appear respectively.

Our numerical examples assume four classes of customers, and also

$$m = 5 \quad ; \quad \mu = 1 \quad ; \quad \lambda_p = \frac{\lambda}{4} \quad (1 \le p \le 4).$$

Using these parameter settings, we will show how the performance varies for $0 \le \lambda \le 20$. Our formulas apply to systems with any number of servers and classes, and for any value of the class-dependent arrival rates. However, it is assumed that service rates are independent of the customer class and that the system is stable for customers of classes up to p ($\rho_p^+ < 1$).

2. Mean Response and Waiting Times

The *response time* of a class-p customer, denoted T_p, is defined as the time interval between his arrival in the queue and the service completion, i.e., the whole period spent in the system. The *waiting time*, denoted W_p, is defined as

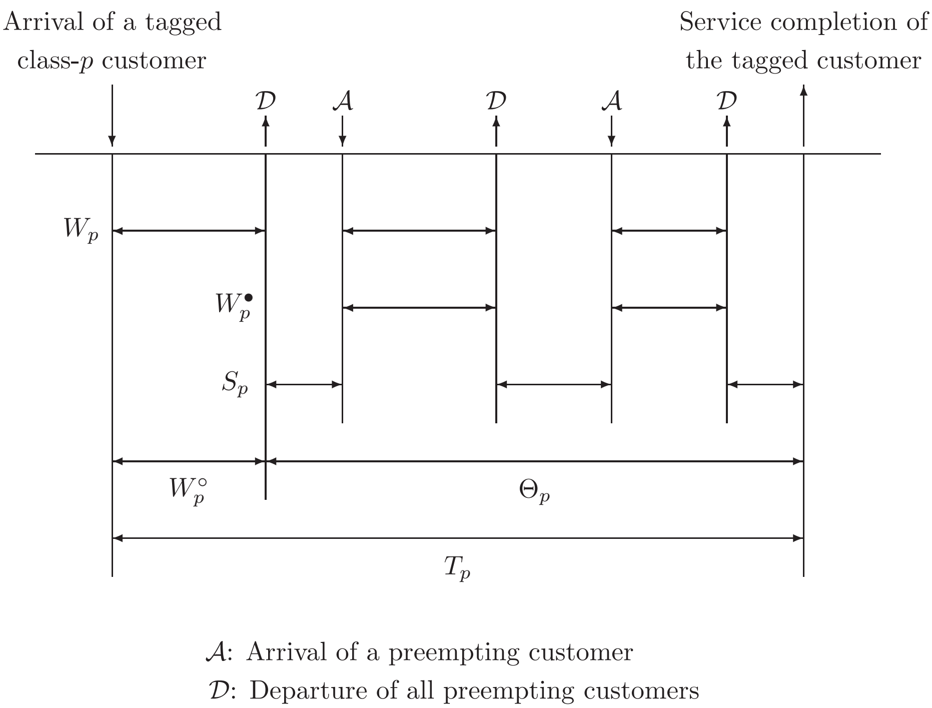

Figure 1: Decompositions of the response time for a class-p customer

the time interval, within the response time, in which this customer is not being served. In this section, we derive the mean values $E[T_p]$ and $E[W_p]$ by adapting the method of Buzen and Bondi [2] for the FCFS system to our LCFS system. The results for our LCFS system are the same as those for the FCFS system.

The service-and-preemption mechanism described in Section 1 is such that the response and waiting times of a given customer are unaffected by another customer belonging to a lower-priority class or by an existing customer of the same class. We therefore need only consider customers of classes $1, 2, \ldots, p$.

We denote by N^+_{p-1} the number of customers of classes $1, 2, \ldots, p-1$ who are present in the steady-state system at any given time, and define

$$Q^+_{p-1,k} := P\{N^+_{p-1} = k\} \qquad k = 0, 1, 2, \ldots.$$

From the well-known analysis of the M/M/m queue with customers of classes

$1, 2, \ldots, p-1$, we get [3, p. 90]

$$Q^+_{p-1,k} = \begin{cases} Q^+_{p-1,0} \dfrac{(m\rho^+_{p-1})^k}{k!} & 1 \le k \le m, \\ Q^+_{p-1,m} (\rho^+_{p-1})^{k-m} & k \ge m+1, \end{cases} \tag{1}$$

where, from the normalization condition $\sum_{k=0}^{\infty} Q^+_{p-1,k} = 1$,

$$\frac{1}{Q^+_{p-1,0}} = \sum_{k=0}^{m-1} \frac{(m\rho^+_{p-1})^k}{k!} + \frac{(m\rho^+_{p-1})^m}{m!(1-\rho^+_{p-1})}.$$

Then,

$$E[N^+_{p-1}] = \sum_{k=1}^{\infty} k Q^+_{p-1,k} = m\rho^+_{p-1} + \frac{\rho^+_{p-1} C(m, m\rho^+_{p-1})}{1-\rho^+_{p-1}},$$

where

$$C(m,a) := \frac{a^m}{m!} \Bigg/ \left[\left(1 - \frac{a}{m}\right) \sum_{k=0}^{m-1} \frac{a^k}{k!} + \frac{a^m}{m!} \right] \tag{2}$$

is *Erlang's C formula* [3, p. 91]. In the present case,

$$C(m, m\rho^+_{p-1}) = \sum_{k=m}^{\infty} Q^+_{p-1,k} = \frac{Q^+_{p-1,m}}{1-\rho^+_{p-1}} = \frac{Q^+_{p-1,0}}{1-\rho^+_{p-1}} \cdot \frac{(m\rho^+_{p-1})^m}{m!}.$$

By the *Poisson arrivals see time averages* (PASTA) property [14, p. 293], this is also the probability that a class-p customer is forced to wait upon arrival.

We denote by N_p the number of class-p customers present in the steady-state system at any given time. Then,

$$E[N_p] = E[N^+_p] - E[N^+_{p-1}] = \frac{\lambda_p}{\mu} + \frac{\rho^+_p C(m, m\rho^+_p)}{1-\rho^+_p} - \frac{\rho^+_{p-1} C(m, m\rho^+_{p-1})}{1-\rho^+_{p-1}}.$$

From Little's theorem [14, p. 235], which states that $E[N_p] = \lambda_p E[T_p]$, we obtain

$$E[T_p] = \frac{E[N_p]}{\lambda_p} = \frac{1}{\mu} + \frac{\rho^+_p C(m, m\rho^+_p)}{\lambda_p(1-\rho^+_p)} - \frac{\rho^+_{p-1} C(m, m\rho^+_{p-1})}{\lambda_p(1-\rho^+_{p-1})}. \tag{3}$$

We denote by L_p the number of class-p customers present in the waiting room at any given time in the steady state. Since the mean number of class-p

customers being served at any given time in the steady state is λ_p/μ, it follows that

$$E[L_p] = E[N_p] - \frac{\lambda_p}{\mu} = \frac{\rho_p^+ C(m, m\rho_p^+)}{1-\rho_p^+} - \frac{\rho_{p-1}^+ C(m, m\rho_{p-1}^+)}{1-\rho_{p-1}^+},$$

from which we obtain the mean waiting time

$$E[W_p] = \frac{E[L_p]}{\lambda_p} = \frac{\rho_p^+ C(m, m\rho_p^+)}{\lambda_p(1-\rho_p^+)} - \frac{\rho_{p-1}^+ C(m, m\rho_{p-1}^+)}{\lambda_p(1-\rho_{p-1}^+)}. \tag{4}$$

The mean response time in Eq. (3) and the mean waiting time in Eq. (4) for a class-p customer in the M/M/m LCFS preemptive-resume priority queue are the same as those in the corresponding M/M/m FCFS preemptive-resume priority queue. The latter results are derived in [2, 9, 10, 15]. Numerical results are plotted in [9, 10] for the example mentioned in Section 1.

For $m = 1$ (i.e., a single-server queue), we have

$$Q_{p-1,0}^+ = 1 - \rho_{p-1}^+ \quad ; \quad C(1, a) = a$$

and thus

$$E[T_p] = \frac{1}{\mu(1-\rho_{p-1}^+)(1-\rho_p^+)} \quad ; \quad E[W_p] = \frac{\rho_{p-1}^+ + \rho_p^+ - \rho_{p-1}^+\rho_p^+}{\mu(1-\rho_{p-1}^+)(1-\rho_p^+)}. \tag{5}$$

This result can also be obtained using an approach similar to White and Christie's method for a single-server FCFS system [12]. In the LCFS system, we observe that $E[T_p]$ comprises the customer's own mean service time, the mean time to serve the customers of classes $1, 2, \ldots, p-1$ who are already present in the waiting room, and the mean workload brought by the subsequent arrivals of customers of classes $1, 2, \ldots, p$ during the time course of $E[T_p]$. Thus,

$$E[T_p] = \frac{1}{\mu} + \frac{1}{\mu}\sum_{k=1}^{p-1} E[N_k] + \sum_{k=1}^{p} \frac{\lambda_k}{\mu} E[T_p], \tag{6}$$

where N_k denotes the number of class-k customers present in the system. From Little's theorem, $E[N_k] = \lambda_k E[T_k]$ for customers of class k $(1 \le k \le p-1)$, we derive the following recursive relation with respect to p:

$$E[T_1] = \frac{1}{\mu(1-\rho_1)} \quad ; \quad (1-\rho_p^+)E[T_p] = \frac{1}{\mu} + \sum_{k=1}^{p-1} \rho_k E[T_k] \qquad p \ge 2.$$

These equations in turn yield Eq. (5). We note that this argument for Eq. (6) holds only when $m = 1$.

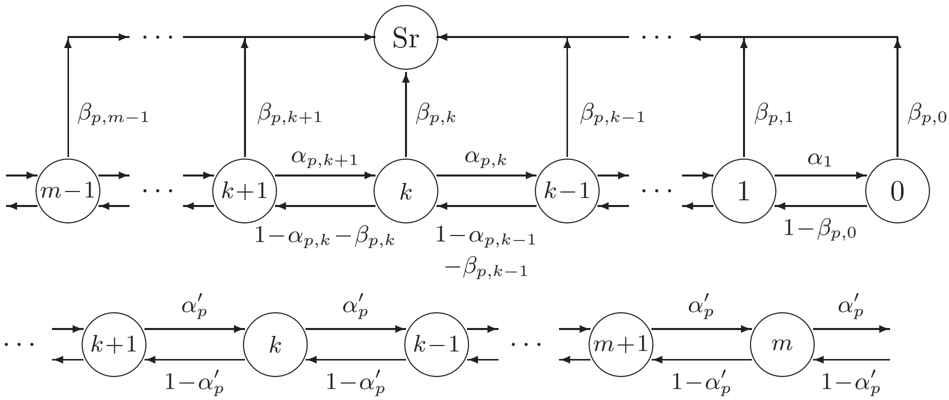

Figure 2: State transitions for a class-p customer until service completion

3. First Passage Time to Service Completion

We now focus on a tagged class-p customer in state k, signifying that there are k other customers of classes $1, 2, \ldots, p$ who compete against him for service at any given time while in the steady state. The state transitions for the tagged customer are depicted in Fig. 2. We consider the first passage time in this one-dimensional birth-and-death process with an absorbing state "Sr" representing the service completion.

We denote by $T^*_{p,k}(s)$ the Laplace-Stieltjes transform (LST) of the distribution function (DF) for the response time of this customer in state k, where $k = 0, 1, 2, \ldots$. Referring to Fig. 2, we obtain the following equations for $\{T^*_{p,k}(s); k \geq 0\}$:

$$\begin{aligned}
T^*_{p,0}(s) &= B^*_{p,0}(s)[\beta_{p,0} + (1-\beta_{p,0})T^*_{p,1}(s)], \\
T^*_{p,k}(s) &= B^*_{p,k}(s)[\beta_{p,k} + (1-\alpha_{p,k}-\beta_{p,k})T^*_{p,k+1}(s) + \alpha_{p,k}T^*_{p,k-1}(s)] \\
&\qquad\qquad 1 \leq k \leq m-1, \\
T^*_{p,k}(s) &= {B'}^*_{p,k}(s)[(1-\alpha'_{p,k})T^*_{p,k+1}(s) + \alpha'_{p,k}T^*_{p,k-1}(s)] \qquad k \geq m,
\end{aligned}$$

where, for $0 \leq k \leq m-1$, we use

$$\alpha_{p,k} = \frac{k\mu}{\lambda^+_p + (k+1)\mu} \quad ; \quad \beta_{p,k} = \frac{\mu}{\lambda^+_p + (k+1)\mu}.$$

The LST of the DF for the time spent by this customer in state k is given by

$$B^*_{p,k}(s) = \frac{\lambda_p^+ + (k+1)\mu}{s + \lambda_p^+ + (k+1)\mu} \qquad 0 \le k \le m-1.$$

For $k \ge m$, we use

$$\alpha'_{p,k} = \frac{m\mu}{\lambda_p^+ + m\mu} \quad ; \quad B'^*_{p,k}(s) = \frac{\lambda_p^+ + m\mu}{s + \lambda_p^+ + m\mu},$$

which are independent of k. Then the above equations can be rewritten as

$$\begin{aligned}
(s + \lambda_p^+ + \mu)T^*_{p,0}(s) &= \mu + \lambda_p^+ T^*_{p,1}(s), \\
[s + \lambda_p^+ + (k+1)\mu]T^*_{p,k}(s) &= \mu + \lambda_p^+ T^*_{p,k+1}(s) + k\mu T^*_{p,k-1}(s) \\
&\qquad\qquad 1 \le k \le m-1, \\
(s + \lambda_p^+ + m\mu)T^*_{p,k}(s) &= \lambda_p^+ T^*_{p,k+1}(s) + m\mu T^*_{p,k-1}(s) \\
&\qquad\qquad k \ge m.
\end{aligned} \tag{7}$$

From the homogeneous set of equations for the case $k \ge m$ in Eq. (7), we assume a solution of the form

$$T^*_{p,k}(s) = [G^*_p(s)]^{k-m} T^*_{p,m}(s) \qquad k \ge m, \tag{8}$$

where $G^*_p(s)$ is the solution to the quadratic equation

$$\lambda_p^+ [G^*_p(s)]^2 - (s + \lambda_p^+ + m\mu)G^*_p(s) + m\mu = 0,$$

namely

$$G^*_p(s) = \frac{s + \lambda_p^+ + m\mu - \sqrt{(s + \lambda_p^+ + m\mu)^2 - 4\lambda_p^+ m\mu}}{2\lambda_p^+}. \tag{9}$$

We note that $G^*_p(s)$ is the LST of the DF for the duration of a busy period in the M/M/1 queue with arrival rate λ_p^+ and service rate $m\mu$, which we denote by $\mathcal{G}_p^+$. Then,

$$E[\mathcal{G}_p^+] = \frac{1}{m\mu(1-\rho_p^+)} \quad ; \quad E[(\mathcal{G}_p^+)^2] = \frac{2}{(m\mu)^2(1-\rho_p^+)^3}$$

$$E[(\mathcal{G}_p^+)^3] = \frac{6(1+\rho_p^+)}{(m\mu)^3(1-\rho_p^+)^5}.$$

Substituting Eq. (8) into Eq. (7) with $k = m$, we get

$$T^*_{p,m}(s) = G^*_p(s) T^*_{p,m-1}(s). \tag{10}$$

We then have a non-homogeneous set of m equations for $0 \le k \le m-1$ in Eq. (7), which allows the expression of both $T^*_{p,m-1}(s)$ and $T^*_{p,m}(s)$ in terms of $T^*_{p,0}(s)$. Substituting these into Eq. (10), we can determine $T^*_{p,0}(s)$ and hence all $\{T^*_{p,k}(s); k \ge 0\}$:

$$T^*_p(s) = \sum_{k=0}^{\infty} Q^+_{p-1,k} T^*_{p,k}(s) = \sum_{k=0}^{m-1} Q^+_{p-1,k} T^*_{p,k}(s) + \frac{Q^+_{p-1,m}}{1-\rho^+_{p-1}} F^*_{p,m}(s) \qquad p \ge 2, \tag{11}$$

where we define

$$\begin{aligned} F^*_{p,m}(s) &:= (1-\rho^+_{p-1}) \sum_{k=m}^{\infty} (\rho^+_{p-1})^{k-m} T^*_{p,k}(s) \\ &= (1-\rho^+_{p-1}) T^*_{p,m}(s) \sum_{k=m}^{\infty} [\rho^+_{p-1} G^*_p(s)]^{k-m} \\ &= \frac{(1-\rho^+_{p-1}) T^*_{p,m}(s)}{1-\rho^+_{p-1} G^*_p(s)}. \end{aligned} \tag{12}$$

An arriving class-1 customer is always placed in state 0 in Fig. 2. Thus,

$$T^*_1(s) = T^*_{1,0}(s),$$

which is the LST of the DF for the response time of a customer in the M/M/m queue following a preemptive LIFO protocol (without priorities) [14, p. 456]. At present, there seems no closed-form expression, even for $T^*_1(s) = T^*_{1,0}(s)$, for the general case of m servers.

For $m = 1$ (a single-server queue), we have

$$Q^+_{p-1,k} = (1-\rho^+_{p-1})(\rho^+_{p-1})^k \quad ; \quad T^*_{p,k}(s) = [G^*_p(s)]^{k+1} \qquad k \ge 0,$$

where $\rho^+_{p-1} := \lambda^+_{p-1}/\mu$. This yields, explicitly,

$$T^*_p(s) \;=\; (1-\rho^+_{p-1}) \sum_{k=0}^{\infty} (\rho^+_{p-1})^k [G^*_p(s)]^{k+1}$$

$$= \frac{(1-\rho_{p-1}^+)G_p^*(s)}{1-\rho_{p-1}^+G_p^*(s)} = \frac{1-\rho_{p-1}^+}{1/G_p^*(s)-\rho_{p-1}^+}$$

$$= \frac{2\mu(1-\rho_{p-1}^+)}{s+\lambda_p^+-2\lambda_{p-1}^++\mu+\sqrt{(s+\lambda_p^++\mu)^2-4\lambda_p^+\mu}}. \quad (13)$$

We note that $T_1^*(s) = G_1^*(s)$ as expected. We then obtain the mean $E[T_p]$ in Eq. (5) and the second and third moments:

$$E[T_p^2] = \frac{2(1-\rho_{p-1}^+\rho_p^+)}{\mu^2(1-\rho_{p-1}^+)^2(1-\rho_p^+)^3},$$

$$E[T_p^3] = \frac{6[1+\rho_p^+-4\rho_{p-1}^+\rho_p^++(\rho_{p-1}^+)^2\rho_p^++(\rho_{p-1}^+\rho_p^+)^2]}{\mu^3(1-\rho_{p-1}^+)^2(1-\rho_p^+)^5}.$$

4. Mean Response Time

Consider the mean response time in the context of the first passage time for a class-p customer who competes for service with k other customers: $E[T_{p,k}] = -dT_{p,k}^*(s)/ds|_{s=0}$, $k \geq 0$. The complete set of $m+1$ equations for $\{E[T_{p,k}]; 0 \leq k \leq m\}$ is given by

$$\begin{aligned} -1+(\lambda_p^++\mu)E[T_{p,0}] &= \lambda_p^+E[T_{p,1}], \\ -1+[\lambda_p^++(k+1)\mu]E[T_{p,k}] &= \lambda_{p-1}^+E[T_{p,k+1}]+k\mu E[T_{p,k-1}] \\ &\qquad 1 \leq k \leq m-1, \\ E[T_{p,m}] &= E[\mathcal{G}_p^+]+E[T_{p,m-1}]. \end{aligned}$$

Using a similar calculation as for the FCFS system given in [9, 10], we obtain

$$E[T_{p,k}] = \frac{1}{\mu} + \frac{\dfrac{k!(m\rho_p^+)^{m-k}}{m!(1-\rho_p^+)}\displaystyle\sum_{j=0}^{k}\frac{(m\rho_p^+)^j}{j!}}{m\mu\left[\dfrac{(m\rho_p^+)^m}{m!}+(1-\rho_p^+)\displaystyle\sum_{j=0}^{m-1}\frac{(m\rho_p^+)^j}{j!}\right]} \quad 0 \leq k \leq m.$$

The mean response time for a class-p customer is then given by

$$E[T_p] = \sum_{k=0}^{m-1} Q_{p-1,k}^+E[T_{p,k}] + \frac{Q_{p-1,m}^+}{1-\rho_{p-1}^+}E[F_{p,m}],$$

where

$$E[F_{p,m}] = -\left.\frac{dF^*_{p,m}(s)}{ds}\right|_{s=0} = E[T_{p,m}] + \frac{\rho^+_{p-1}E[\mathcal{G}^+_p]}{1-\rho^+_{p-1}}.$$

We have confirmed numerically that this formula yields the same result as Eq. (3). In particular, for $p = 1$, we have

$$\begin{aligned} E[T_1] &= E[T_{1,0}] = \frac{1}{\mu} + \frac{\dfrac{(m\rho_1)^m}{m!(1-\rho_1)}}{m\mu\left[\dfrac{(m\rho_1)^m}{m!} + (1-\rho_1)\displaystyle\sum_{j=0}^{m-1}\dfrac{(m\rho_1)^j}{j!}\right]} \\ &= \frac{1}{\mu} + \frac{C(m, m\rho_1)}{m\mu - \lambda_1} \end{aligned}$$

in agreement with Eq. (3).

5. Second and Higher-Order Moments of the Response Time

The ℓth moment of the response time for a class-p customer who competes for service with k other customers is given by $E[T^\ell_{p,k}] = (-1)^\ell d^\ell T^*_{p,k}(s)/ds^\ell|_{s=0},\ k \geq 0$, for $\ell = 2, 3, \ldots$. The complete set of $m+1$ equations for the unknowns $\{E[T^\ell_{p,k}]; 0 \leq k \leq m\}$ is derived from Eqs. (7) and (8) in terms of the lower-order moments:

$$\begin{aligned} -\ell E[T^{\ell-1}_{p,0}] + (\lambda^+_p + \mu)E[T^\ell_{p,0}] &= \lambda^+_p E[T^\ell_1], \\ -\ell E[T^{\ell-1}_{p,k}] + [\lambda^+_p + (k+1)\mu]E[T^\ell_{p,k}] &= \lambda^+_p E[T^\ell_{p,k+1}] + k\mu E[T^\ell_{p,k-1}] \\ &\qquad 1 \leq k \leq m-1, \\ E[T^\ell_{p,m}] - E[T^\ell_{p,m-1}] &= \sum_{l=1}^{\ell}\binom{\ell}{l}E[(\mathcal{G}^+_p)^l]E[T^{\ell-l}_{p,m-1}]. \end{aligned}$$

We obtain $\{E[T^\ell_{p,k}]; 0 \leq k \leq m\}$ recursively with respect to ℓ in the same way as in [9, 10] as follows. Firstly, we have

$$\frac{E[T_{p,0}^{\ell}]}{\ell} = \frac{\left\{ \begin{array}{l} \displaystyle\sum_{j=0}^{m-1} E[T_{p,j}^{\ell-1}] \left[\frac{(m\rho_p^+)^m}{m!} + (1-\rho_p^+) \sum_{l=j+1}^{m-1} \frac{(m\rho_p^+)^l}{l!} \right. \\ \displaystyle + \frac{\lambda_p^+}{\ell} \cdot \frac{(m\rho_p^+)^m}{m!} \sum_{l=1}^{\ell} \binom{\ell}{l} E[(\mathcal{G}_p^+)^l] E[T_{p,m-1}^{\ell-l}] \end{array} \right\}}{\lambda_{p-1}^+ \left[\frac{(m\rho_p^+)^m}{m!} + (1-\rho_p^+) \sum_{j=0}^{m-1} \frac{(m\rho_p^+)^j}{j!} \right]}.$$

Then we calculate $\{E[T_{p,k}^{\ell}]; 1 \le k \le m\}$ using

$$E[T_{p,k}^{\ell}] \frac{(m\rho_p^+)^k}{k!} = E[T_{p,0}^{\ell}] \sum_{j=0}^{k} \frac{(m\rho_p^+)^j}{j!} - \frac{\ell}{\lambda_p^+} \sum_{j=1}^{k} \frac{(m\rho_p^+)^j}{j!} \sum_{l=0}^{j-1} E[T_{p,l}^{\ell-1}] \qquad 1 \le k \le m.$$

In particular,

$$\frac{E[T_{p,m}^{\ell}]}{\ell} = \frac{\left\{ \begin{array}{l} \displaystyle\sum_{j=0}^{m-1} E[T_{p,j}^{\ell-1}] \sum_{l=0}^{j} \frac{(m\rho_p^+)^l}{l!} \\ \displaystyle + \frac{m\mu}{\ell} \sum_{j=0}^{m} \frac{(m\rho_p^+)^j}{j!} \sum_{l=1}^{\ell} \binom{\ell}{l} E[(\mathcal{G}_p^+)^l] E[T_{p,m-1}^{\ell-l}] \end{array} \right\}}{m\mu \left[\frac{(m\rho_p^+)^m}{m!} + (1-\rho_p^+) \sum_{j=0}^{m-1} \frac{(m\rho_p^+)^j}{j!} \right]}.$$

Finally, the ℓth moment of the response time is given by

$$E[T_p^{\ell}] = \sum_{k=0}^{\infty} Q_{p-1,k}^+ E[T_{p,k}^{\ell}] = \sum_{k=0}^{m-1} Q_{p-1,k}^+ E[T_{p,k}^{\ell}] + \frac{Q_{p-1,m}^+}{1-\rho_{p-1}^+} E[F_{p,m}^{\ell}], \qquad (14)$$

where $E[F_{p,m}^{\ell}] = (-1)^{\ell} d^{\ell} F_{p,m}^*(s)/ds^{\ell}|_{s=0}$ is obtained recursively using

$$E[F_{p,m}^{\ell}] = E[T_{p,m}^{\ell}] + \frac{\rho_{p-1}^+}{1-\rho_{p-1}^+} \sum_{l=1}^{\ell} \binom{\ell}{l} E[F_{p,m}^{\ell-l}] E[(\mathcal{G}_p^+)^l]$$

for $\ell = 2, 3, \ldots$. Thus, the second moment of the response time for a class-p customer is given by

$$\begin{aligned} E[T_p^2] &= \sum_{k=0}^{m-1} Q^+_{p-1,k} E[T^2_{p,k}] + \frac{Q^+_{p-1,m}}{1-\rho^+_{p-1}} \left\{ E[T^2_{p,m}] + \frac{2(\rho^+_{p-1} E[\mathcal{G}^+_p])^2}{(1-\rho^+_{p-1})^2} \right. \\ &+ \left. \frac{\rho^+_{p-1}\{E[(\mathcal{G}^+_p)^2] + 2E[T_{p,m}]E[\mathcal{G}^+_p]\}}{1-\rho^+_{p-1}} \right\}, \end{aligned} \tag{15}$$

and the third moment by

$$\begin{aligned} E[T_p^3] &= \sum_{k=0}^{m-1} Q^+_{p-1,k} E[T^3_{p,k}] + \frac{Q^+_{p-1,m}}{1-\rho^+_{p-1}} \left\{ E[T^3_{p,m}] + \frac{6(\rho^+_{p-1} E[\mathcal{G}^+_p])^3}{(1-\rho^+_{p-1})^3} \right. \\ &+ \frac{\rho^+_{p-1}\{E[(\mathcal{G}^+_p)^3] + 3E[T_{p,m}]E[(\mathcal{G}^+_p)^2] + 3E[T^2_{p,m}]E[\mathcal{G}^+_p]\}}{1-\rho^+_{p-1}} \\ &+ \left. \frac{6(\rho^+_{p-1})^2\{E[(\mathcal{G}^+_p)^2]E[\mathcal{G}^+_p] + E[T_{p,m}](E[\mathcal{G}^+_p])^2\}}{(1-\rho^+_{p-1})^2} \right\}. \end{aligned} \tag{16}$$

The second and third moments of the response time, $E[T_p^2]$ and $E[T_p^3]$, are plotted in Figs. 3 and 4, respectively, for the numerical example described in Section 1. Table 1 compares the $E[T_p^2]_{\text{FCFS}}$ values calculated from [9, 10, 15] and the $E[T_p^2]_{\text{LCFS}}$ values obtained from Eq. (15) for the same numerical example. We observe that $E[T_p^2]_{\text{LCFS}}$ is much greater than $E[T_p^2]_{\text{FCFS}}$ for the same set of parameter values.

6. Initial Waiting Time

The *initial waiting time* for a class-p customer, denoted W_p°, is defined as the time interval spanning his arrival and the instant at which he first receives service (see Fig. 1). Clearly, $W_1^\circ = 0$ for a highest-priority customer. For a customer of class p (≥ 2), W_p° is determined by the customers of classes $1, 2, \ldots, p-1$ who are present in the system at the time of his arrival and the customers of classes $1, 2, \ldots, p$ who arrive after him.

We denote by $W^\circ_{p,k}(s)$ the LST of the DF of the initial waiting time $W^\circ_{p,k}$ of a class-p customer who, upon arrival, competes for service with k other customers of classes $1, 2, \ldots, p-1$. By definition,

$$W^\circ_{p,k}(s) = 1 \qquad 0 \leq k \leq m-1.$$

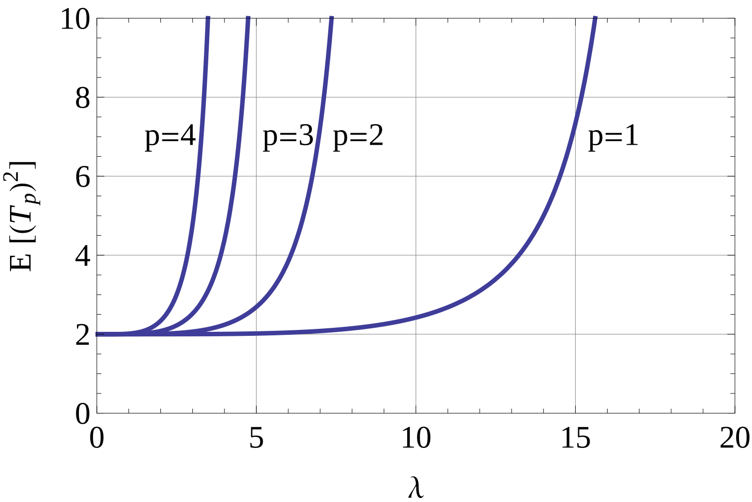

Figure 3: Second moment of the response time in the M/M/m LCFS preemptive-resume priority queue

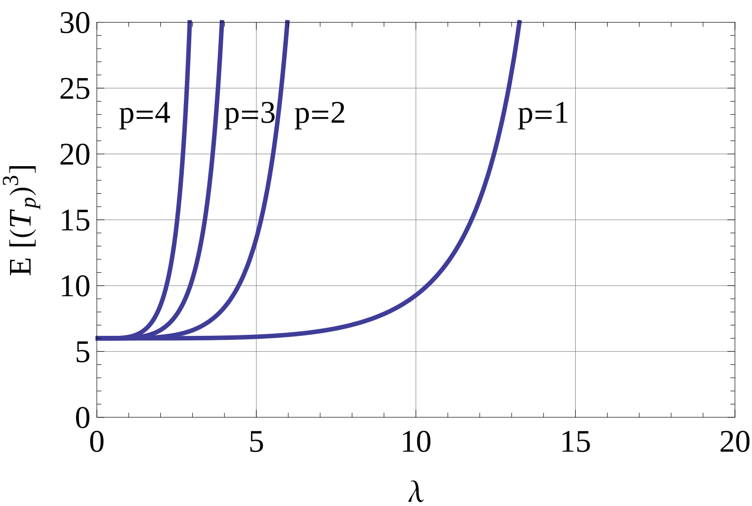

Figure 4: Third moment of the response time in the M/M/m LCFS preemptive-resume priority queue

Table 1: Comparison of the second moment of the response time in the M/M/m FCFS and LCFS preemptive-resume priority queues

λ	$E[T_1^2]_{\text{FCFS}}$	$E[T_1^2]_{\text{LCFS}}$
0	2.00000	2.00000
5	2.00656	2.01823
10	2.14602	2.42076
15	3.32995	7.33793
17	6.07226	26.7504
18	11.1499	84.1473

λ	$E[T_2^2]_{\text{FCFS}}$	$E[T_2^2]_{\text{LCFS}}$
0	2.00000	2.00000
2	2.00572	2.01065
4	2.13013	2.23911
6	2.91820	3.84812
7	4.28566	7.25239
8	8.49564	21.3787

λ	$E[T_3^2]_{\text{FCFS}}$	$E[T_3^2]_{\text{LCFS}}$
0	2.00000	2.00000
2	2.05481	2.08164
3	2.34530	2.52038
4	3.47013	4.38878
5	8.45801	15.0244
6	61.5514	234.766

λ	$E[T_4^2]_{\text{FCFS}}$	$E[T_4^2]_{\text{LCFS}}$
0	2.00000	2.00000
2	2.25244	2.33933
3	3.91468	4.77432
4	18.7710	35.2459
4.5	96.0743	302.523
4.8	714.629	4936.70

For $k \geq m$, we have

$$W^{\circ}_{p,k}(s) = B'^{*}_{p}(s)[(1-\alpha'_p)W^{\circ}_{p,k+1}(s) + \alpha'_p W^{\circ}_{p,k-1}(s)],$$

where $B'^{*}_{p}(s)$ and α'_p are as defined in Section 3. Thus,

$$(s+\lambda_p^+ + m\mu)W^{\circ}_{p,k}(s) = \lambda_p^+ W^{\circ}_{p,k+1}(s) + m\mu W^{\circ}_{p,k-1}(s) \qquad k \geq m.$$

These relations are satisfied by

$$W^{\circ}_{p,k}(s) = [G^*_p(s)]^{k-m+1} \qquad k \geq m-1,$$

where $G^*_p(s)$ is given by Eq. (9). Thus,

$$\begin{aligned} W^{\circ}_p(s) &= \sum_{k=0}^{\infty} Q^+_{p-1,k} W^{\circ}_{p,k}(s) \\ &= \sum_{k=0}^{m-1} Q^+_{p-1,0}\frac{(m\rho^+_{p-1})^k}{k!} + \sum_{k=m}^{\infty} Q^+_{p-1,m}(\rho^+_{p-1})^{k-m}[G^*_p(s)]^{k-m+1} \end{aligned}$$

$$= Q^+_{p-1,0} \sum_{k=0}^{m-1} \frac{(m\rho^+_{p-1})^k}{k!} + \frac{Q^+_{p-1,m} G^*_p(s)}{1-\rho^+_{p-1} G^*_p(s)}$$

$$= 1 - C(m, m\rho^+_{p-1}) + C(m, m\rho^+_{p-1}) \frac{(1-\rho^+_{p-1}) G^*_p(s)}{1-\rho^+_{p-1} G^*_p(s)}$$

$$= 1 - \frac{C(m, m\rho^+_{p-1})[1-G^*_p(s)]}{1-\rho^+_{p-1} G^*_p(s)}, \qquad (17)$$

where $C(m, a)$ is given by Eq. (2). The mean and the second and third moments of W°_p are then given by

$$E[W^\circ_p] = \frac{C(m, m\rho^+_{p-1})}{m\mu(1-\rho^+_{p-1})(1-\rho^+_p)},$$

$$E[(W^\circ_p)^2] = \frac{2C(m, m\rho^+_{p-1})(1-\rho^+_{p-1}\rho^+_p)}{(m\mu)^2(1-\rho^+_{p-1})^2(1-\rho^+_p)^3},$$

$$E[(W^\circ_p)^3] = \frac{6C(m, m\rho^+_{p-1}) \left\{ \begin{array}{l} 1+\rho^+_p - 4\rho^+_{p-1}\rho^+_p \\ +(\rho^+_{p-1})^2\rho^+_p + (\rho^+_{p-1}\rho^+_p)^2 \end{array} \right\}}{(m\mu)^3(1-\rho^+_{p-1})^3(1-\rho^+_p)^5}.$$

We note that

$$W^+_p(s) = \frac{(1-\rho^+_{p-1})G^*_p(s)}{1-\rho^+_{p-1}G^*_p(s)} = \frac{m\mu(1-\rho^+_{p-1})[1-G^*_p(s)]}{s+\lambda_p-\lambda_p G^*_p(s)} \qquad (18)$$

is the LST of the DF of the initial waiting time W^+_p for a customer of class p who arrives when all servers are already busy serving customers of classes $1, 2, \ldots, p-1$. We have

$$E[W^+_p] = \frac{1}{m\mu(1-\rho^+_{p-1})(1-\rho^+_p)},$$

$$E[(W^+_p)^2] = \frac{2(1-\rho^+_{p-1}\rho^+_p)}{(m\mu)^2(1-\rho^+_{p-1})^2(1-\rho^+_p)^3}.$$

Thus, we can write Eq. (17) as

$$W^\circ_p(s) = 1 - C(m, m\rho^+_{p-1}) + C(m, m\rho^+_{p-1}) W^+_p(s).$$

For $m = 1$ (a single-server queue), we have

$$W_p^\circ(s) = 1 - \rho_{p-1}^+ + \rho_{p-1}^+ W_p^+(s) = \frac{1-\rho_{p-1}^+}{1-\rho_{p-1}^+ G_p^*(s)},$$

$$E[W_p^\circ] = \frac{\rho_{p-1}^+}{\mu(1-\rho_{p-1}^+)(1-\rho_p^+)}. \tag{19}$$

See the final result in the Appendix.

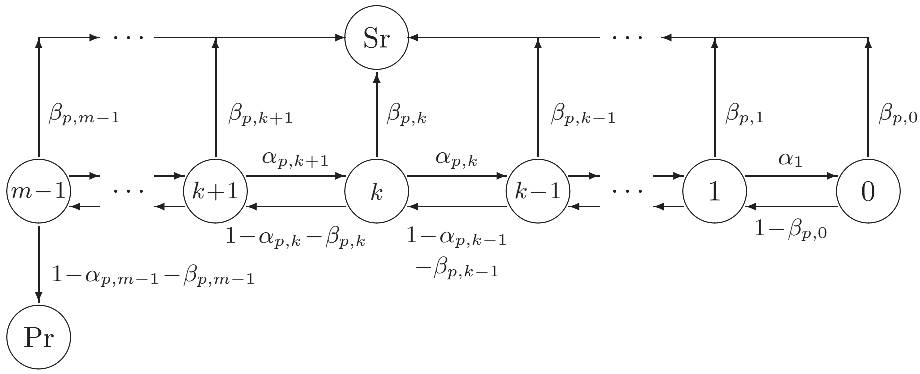

Figure 5: State transitions for a class-p customer until service preemption or completion

7. Total Waiting Time

Service to a class-p customer may be preempted several times before it is completed. He is then pushed out of the service facility by an arriving customer of a higher- or equal-priority class $1, 2, \ldots, p$. The first customer stays in the waiting room until he is called in to be served again. The total time spent by a customer in the waiting room is called the *total waiting time*, which equals the response time minus the total service time (see Fig. 1). We follow the approach of Tatashev [11] to derive the distribution of the total waiting time for a class-p customer.

Let $P_{p,k}\{\text{Pr}\}$ be the probability that a tagged class-p customer is preempted as he competes for service with k other customers $(0 \le k \le m-1)$ of either higher priority (i.e., of class $1, 2, \ldots, p-1$), or of equal priority (class p) but who arrived after him. Figure 5 shows the state transitions for the tagged customer of class p. We consider the first passage time in this one-dimensional birth-and-death process which contains two absorbing states "Pr" representing the service preemption and "Sr" the service completion.

Referring to Fig. 5, the complete set of m equations for $\{P_{p,k}\{\text{Pr}\};$ $0 \le k \le m-1\}$ is as follows:

$$\begin{aligned} P_{p,0}\{\text{Pr}\} &= (1-\beta_{p,0})P_{p,1}\{\text{Pr}\}, \\ P_{p,k}\{\text{Pr}\} &= (1-\alpha_{p,k}-\beta_{p,k})P_{p,k+1}\{\text{Pr}\} + \alpha_{p,k}P_{p,k-1}\{\text{Pr}\} \\ &\qquad\qquad\qquad\qquad\qquad\qquad 1 \le k \le m-2, \\ P_{p,m-1}\{\text{Pr}\} &= 1-\alpha_{p,m-1}-\beta_{p,m-1}+\alpha_{p,m-1}P_{p,m-2}\{\text{Pr}\}, \end{aligned}$$

where $\alpha_{p,k}$ and $\beta_{p,k}$ $(0 \le k \le m-1)$ are specified in Section 3. The solution is

$$P_{p,k}\{\text{Pr}\} = \frac{B(m, m\rho_p^+)}{B(k, m\rho_p^+)} \qquad 0 \le k \le m-1 \tag{20}$$

with the well-known *Erlang's B formula* [3, p. 80]

$$B(m,a) := \frac{a^m}{m!} \bigg/ \sum_{k=0}^{m} \frac{a^k}{k!} .$$

We note that

$$r_p := P_{p,m-1}\{\text{Pr}\} = \frac{B(m, m\rho_p^+)}{B(m-1, m\rho_p^+)} = \rho_p^+ \left[1 - B(m, m\rho_p^+)\right]$$

is the probability that the service to a class-p customer, started after a period of waiting, is preempted.

The probability q_p that the service to a class-p customer, started without any waiting, is preempted can be derived as follows. When his service is started, there are k other customers of classes $1, 2, \ldots,$
$p-1$ with probability (named *Erlang's loss distribution* [3, p. 80]):

$$\frac{(m\rho_{p-1}^+)^k}{k!} \bigg/ \sum_{j=0}^{m-1} \frac{(m\rho_{p-1}^+)^j}{j!} \qquad 0 \le k \le m-1.$$

His service is then preempted with probability $P_{p,k}\{\text{Pr}\}$. Therefore,

$$q_p = \sum_{k=0}^{m-1} P_{p,k}\{\text{Pr}\} \frac{(m\rho_{p-1}^+)^k/k!}{\sum_{j=0}^{m-1}(m\rho_{p-1}^+)^j/j!}$$

$$= \sum_{k=0}^{m-1} \frac{B(m, m\rho_p^+)}{B(k, m\rho_p^+)} \cdot \frac{(m\rho_{p-1}^+)^k/k!}{\sum_{j=0}^{m-1}(m\rho_{p-1}^+)^j/j!}.$$

After some manipulation (see [9, 10, 11, 15]), we get

$$q_p = \frac{\rho_p^+ \left[B(m, m\rho_p^+) - B(m, m\rho_{p-1}^+)\right]}{\rho_p \left[1 - B(m, m\rho_{p-1}^+)\right]}.$$

Similarly, let $P_{p,k}\{\mathrm{Sr}\}$ be the probability that service to a tagged class-p customer, competing with k other customers ($0 \le k \le m-1$), is completed without preemption. Referring to Fig. 5, the complete set of m equations for $\{P_{p,k}\{\mathrm{Sr}\}; 0 \le k \le m-1\}$ is

$$\begin{aligned} P_{p,0}\{\mathrm{Sr}\} &= \beta_{p,0} + (1-\beta_{p,0})P_{p,1}\{\mathrm{Sr}\}, \\ P_{p,k}\{\mathrm{Sr}\} &= \beta_{p,k} + (1-\alpha_{p,k}-\beta_{p,k})P_{p,k+1}\{\mathrm{Sr}\} + \alpha_{p,k}P_{p,k-1}\{\mathrm{Sr}\} \\ &\qquad 1 \le k \le m-2, \\ P_{p,m-1}\{\mathrm{Sr}\} &= \beta_{p,m-1} + \alpha_{p,m-1}P_{p,m-2}\{\mathrm{Sr}\}. \end{aligned}$$

The solution is given by

$$P_{p,k}\{\mathrm{Sr}\} = 1 - P_{p,k}\{\mathrm{Pr}\} = 1 - \frac{B(m, m\rho_p^+)}{B(k, m\rho_p^+)} \qquad 0 \le k \le m-1. \tag{21}$$

We can then numerically confirm the relation

$$\sum_{k=0}^{m-1} Q_{p-1,k}^+ P_{p,k}\{\mathrm{Sr}\} = [1 - C(m, m\rho_{p-1}^+)](1-q_p)$$

as the probability that an arriving class-p customer receives service immediately upon arrival and that this service is not preempted before completion. We can also confirm the relation

$$\sum_{k=0}^{m-1} Q_{p-1,k}^+ P_{p,k}\{\mathrm{Pr}\} = [1 - C(m, m\rho_{p-1}^+)]q_p$$

as the probability that an arriving class-p customer receives service immediately upon arrival and that this service is preempted before completion.

Upon arrival of the tagged class-p customer, the following cases may arise:

- If fewer than m servers are busy serving customers of classes $1, 2, \ldots, p-1$, his service is started immediately. This case occurs with probability $1 - C(m, m\rho_{p-1}^{+})$.
  - If this service is not preempted, his waiting time is zero. This subcase occurs with probability $1 - q_p$.
  - If this service is preempted, he waits $\mathcal{G}_p^{+}$ time units for his service to resume. This subcase occurs with probability q_p. The resumed service is preempted i times with probability $(1 - r_p)(r_p)^i$ $(i = 0, 1, 2, \ldots)$, with each preemption incurring a waiting time of $\mathcal{G}_p^{+}$ time units.
- If all servers are busy serving customers of classes $1, 2, \ldots, p$, he waits W_p^{+} time units for his service to be started for the first time. This case occurs with probability $C(m, m\rho_{p-1}^{+})$. His service is then preempted i times with probability $(1 - r_p)(r_p)^i$ $(i = 0, 1, 2, \ldots)$, with each preemption incurring a waiting time of $\mathcal{G}_p^{+}$ time units. We note that the LST of the DF of W_p^{+}, $W_p^{+}(s)$, is given by Eq. (18).

Therefore, the LST of the DF for the waiting time of a tagged class-p customer is

$$\begin{aligned} W_p^*(s) &= [1 - C(m, m\rho_{p-1}^{+})] \\ &\quad \times \left\{ 1 - q_p + q_p G_p^*(s) \sum_{i=0}^{\infty} (1 - r_p)(r_p)^i [G_p^*(s)]^i \right\} \\ &\quad + C(m, m\rho_{p-1}^{+}) W_p^{+}(s) \sum_{i=0}^{\infty} (1 - r_p)(r_p)^i [G_p^*(s)]^i \\ &= [1 - C(m, m\rho_{p-1}^{+})] \left\{ 1 - q_p + \frac{q_p(1 - r_p)G_p^*(s)}{1 - r_p G_p^*(s)} \right\} \\ &\quad + C(m, m\rho_{p-1}^{+}) \frac{(1 - r_p)W_p^{+}(s)}{1 - r_p G_p^*(s)}. \end{aligned} \tag{22}$$

The mean total waiting time for a class-p customer is given by

$$E[W_p] = \frac{[1 - C(m, m\rho_{p-1}^{+})]q_p}{m\mu(1 - r_p)(1 - \rho_p^{+})} + \frac{C(m, m\rho_{p-1}^{+})(1 - r_p\rho_{p-1}^{+})}{m\mu(1 - r_p)(1 - \rho_{p-1}^{+})(1 - \rho_p^{+})}.$$

Numerically, this can be shown to yield the same result as Eq. (4).

The second moment of the total waiting time is given by

$$\begin{aligned} E[W_p^2] &= \frac{2[1-C(m,m\rho_{p-1}^+)]q_p(1-r_p\rho_p^+)}{(m\mu)^2(1-r_p)^2(1-\rho_p^+)^3} \\ &+ \frac{2C(m,m\rho_{p-1}^+)}{(m\mu)^2}\left[\frac{1-\rho_{p-1}^+\rho_p^+}{(1-\rho_{p-1}^+)^2(1-\rho_p^+)^3}\right. \\ &+ \left.\frac{r_p[2-\rho_{p-1}^+-\rho_p^+-r_p(1-\rho_{p-1}^+\rho_p^+)]}{(1-r_p)^2(1-\rho_{p-1}^+)(1-\rho_p^+)^3}\right]. \end{aligned} \tag{23}$$

This expression simplifies to

$$\begin{aligned} E[W_p^2] &= \frac{2}{(m\mu)^2}\left[\frac{(1-\rho_{p-1}^+\rho_p^+)C(m,m\rho_{p-1}^+)}{(1-\rho_{p-1}^+)^2(1-\rho_p^+)^3}\right. \\ &+ \frac{\rho_p^+C(m,m\rho_{p-1}^+)[1-C(m,m\rho_p^+)]}{(1-\rho_{p-1}^+)(1-\rho_p^+)^3} \\ &+ \left.\frac{[q_p+(r_p-q_p)C(m,m\rho_{p-1}^+)](1-r_p\rho_p^+)}{(1-r_p)^2(1-\rho_p^+)^3}\right] \end{aligned} \tag{24}$$

by using the relation [3, p. 92]

$$C(m,m\rho_p^+) = \frac{B(m,m\rho_p^+)}{1-\rho_p^+ + \rho_p^+B(m,m\rho_p^+)} = \frac{\rho_p^+ - r_p}{\rho_p^+(1-r_p)}.$$

The third moment is given by

$$\begin{aligned} E[W_p^3] &= \frac{6[1-C(m,m\rho_{p-1}^+)]q_p[1+\rho_p^+-4r_p\rho_p^++r_p^2\rho_p^++(r_p\rho_p^+)^2]}{(m\mu)^3(1-r_p)^3(1-\rho_p^+)^5} \\ &+\frac{6C(m,m\rho_{p-1}^+)}{(m\mu)^3}\left[\frac{1+\rho_p^+-4\rho_{p-1}^+\rho_p^++(\rho_{p-1}^+)^2\rho_p^++(\rho_{p-1}^+\rho_p^+)^2}{(1-\rho_{p-1}^+)^3(1-\rho_p^+)^5}\right. \\ &+ \left.\frac{\sum_{i=1}^3 r_p^i w_i(\rho_{p-1}^+,\rho_p^+)}{(1-r_p)^3(1-\rho_{p-1}^+)^2(1-\rho_p^+)^5}\right], \end{aligned} \tag{25}$$

where

$$w_1(\rho_{p-1}^+,\rho_p^+) := 3-3\rho_{p-1}^+-\rho_p^+-2\rho_{p-1}^+\rho_p^++(\rho_{p-1}^+)^2$$

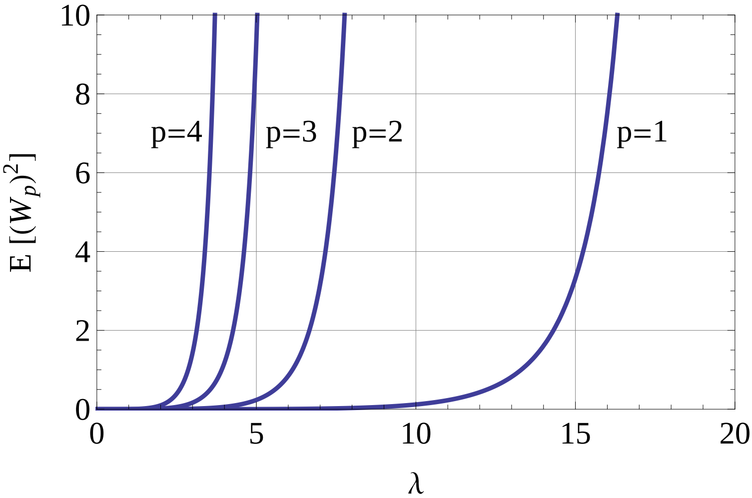

Figure 6: Second moment of the waiting time in the M/M/m LCFS preemptive-resume priority queue

$$
\begin{aligned}
& + \ (\rho_{p-1}^{+})^2\rho_p^{+} + \rho_{p-1}^{+}(\rho_p^{+})^2, \\
w_2(\rho_{p-1}^{+}, \rho_p^{+}) \ &:= \ -3 + \rho_{p-1}^{+} - 2\rho_p^{+} + (\rho_p^{+})^2 + 10\rho_{p-1}^{+}\rho_p^{+} \\
& - \ 3\rho_{p-1}^{+}(\rho_p^{+})^2 - 4(\rho_{p-1}^{+})^2\rho_p^{+}, \\
w_3(\rho_{p-1}^{+}, \rho_p^{+}) \ &:= \ 1 + \rho_p^{+} - 4\rho_{p-1}^{+}\rho_p^{+} + (\rho_{p-1}^{+})^2\rho_p^{+} + (\rho_{p-1}^{+})^2(\rho_p^{+})^2.
\end{aligned}
$$

Figures 6 and 7 plot $E[W_p^2]$ and $E[W_p^3]$, respectively, for the numerical example described in Section 1. Table 2 compares the $E[W_p^2]_{\mathrm{FCFS}}$ values calculated from [9, 10, 15] and the $E[W_p^2]_{\mathrm{LCFS}}$ values obtained from Eq. (23) for the same numerical example. We observe that $E[W_p^2]_{\mathrm{LCFS}}$ is much larger than $E[W_p^2]_{\mathrm{FCFS}}$ for the same set of parameter values.

For $m = 1$ (a single-server queue), we have

$$
C(1, \rho_{p-1}^{+}) = \rho_{p-1}^{+}, \quad r_p = q_p = \frac{\rho_p^{+}}{1 + \rho_p^{+}}.
$$

and hence

$$
W_p^*(s) = \frac{1 - \rho_{p-1}^{+}}{[1 - \rho_{p-1}^{+}G_p^*(s)][1 + \rho_p^{+} - \rho_p^{+}G_p^*(s)]}, \tag{26}
$$

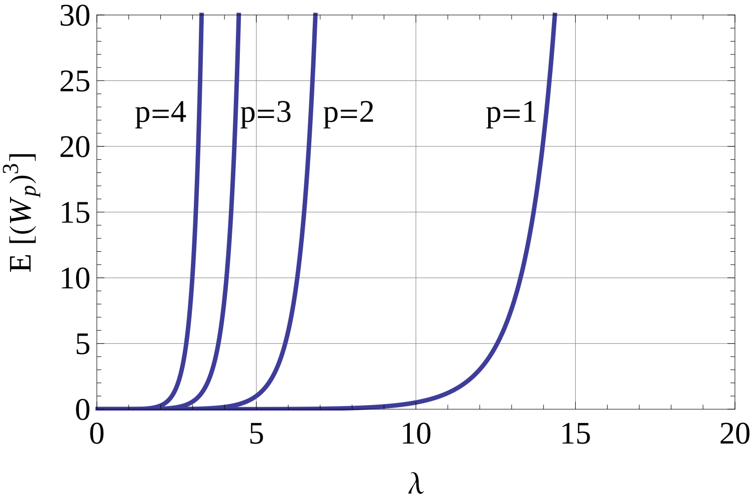

Figure 7: Third moment of the waiting time in the M/M/m LCFS preemptive-resume priority queue

which yields

$$E[W_p] = \frac{\rho_{p-1}^+ + \rho_p^+ - \rho_{p-1}^+\rho_p^+}{\mu(1-\rho_{p-1}^+)(1-\rho_p^+)},$$

$$E[W_p^2] = \frac{2}{\mu^2}\left[\frac{\rho_{p-1}^+ + \rho_p^+ - \rho_{p-1}^+\rho_p^+ - (\rho_{p-1}^+)^2\rho_p^+}{(1-\rho_{p-1}^+)^2(1-\rho_p^+)^3}\right.$$
$$\left. + \frac{(1-2\rho_{p-1}^+)(\rho_p^+)^2}{(1-\rho_{p-1}^+)(1-\rho_p^+)^3} - \frac{(\rho_p^+)^3}{(1-\rho_p^+)^3}\right],$$

$$E[W_p^3] = \frac{6}{\mu^3}\left[\frac{\rho_{p-1}^+ + \rho_p^+[1-4(\rho_{p-1}^+)^2 + (\rho_{p-1}^+)^3]}{(1-\rho_{p-1}^+)^3(1-\rho_p^+)^5}\right.$$
$$+ \frac{(\rho_p^+)^2[3-10\rho_{p-1}^+ + 9(\rho_{p-1}^+)^2 - (\rho_{p-1}^+)^3]}{(1-\rho_{p-1}^+)^3(1-\rho_p^+)^5}$$
$$- \frac{(\rho_p^+)^3[1-2(\rho_{p-1}^+)^2]}{(1-\rho_{p-1}^+)^2(1-\rho_p^+)^5} - \frac{(\rho_p^+)^4(2-3\rho_{p-1}^+)}{(1-\rho_{p-1}^+)(1-\rho_p^+)^5}$$

Table 2: Comparison of the second moment of the waiting time in the M/M/m FCFS and LCFS preemptive-resume priority queues

λ	$E[W_1^2]_{\rm FCFS}$	$E[W_1^2]_{\rm LCFS}$
0	0.00000	0.00000
5	0.00138	0.00230
10	0.04172	0.11972
15	0.59109	3.31875
17	2.32700	20.0696
18	6.09995	74.0385

λ	$E[W_2^2]_{\rm FCFS}$	$E[W_2^2]_{\rm LCFS}$
0	0.00000	0.00000
2	0.00114	0.00141
4	0.03741	0.05966
6	0.38795	0.84614
7	1.18006	3.16948
8	4.12679	14.8981

λ	$E[W_3^2]_{\rm FCFS}$	$E[W_3^2]_{\rm LCFS}$
0	0.00000	0.00000
2	0.01304	0.01641
3	0.11454	0.16641
4	0.67756	1.17949
5	4.07815	9.35307
6	50.1305	217.735

λ	$E[W_4^2]_{\rm FCFS}$	$E[W_4^2]_{\rm LCFS}$
0	0.00000	0.00000
2	0.07668	0.09806
3	0.92966	1.42828
4	12.3304	27.0361
4.5	81.9532	283.096
4.8	676.740	4881.80

$$+ \frac{(\rho_p^+)^5}{(1-\rho_p^+)^5}\Bigg] .$$

The *limbo*, or purgatory, is a term proposed by Wolff [13] to designate the time period, within the total waiting time, during which service to a customer is suspended. The total waiting time W_p for a class-p customer consists of the initial waiting time W_p° and the time in limbo, denoted $W_p^\bullet$ (see Fig. 1):

$$W_p = W_p^\circ + W_p^\bullet .$$

The joint LST of the DF for W_p° and $W_p^\bullet$ is given by

$$\begin{aligned} \tilde{W}_p^*(s,s') &= E\left[e^{-sW_p^\circ - s'W_p^\bullet}\right] \\ &= [1 - C(m, m\rho_{p-1}^+)]\left\{1 - q_p + \frac{q_p(1-r_p)G_p^*(s')}{1-r_pG_p^*(s')}\right\} \end{aligned}$$

$$+ \quad \frac{C(m, m\rho_{p-1}^{+})(1-r_p)W_p^{+}(s)}{1-r_pG_p^*(s')},$$

which is consistent with

$$W_p^{\circ}(s) = \tilde{W}_p^*(s,0) \quad ; \quad W_p^*(s) = \tilde{W}_p^*(s,s).$$

The LST of the DF for $W_p^{\bullet}$ is given by

$$\begin{aligned} W_p^{\bullet}(s) &= \tilde{W}_p^*(0,s) \\ &= [1 - C(m, m\rho_{p-1}^{+})] \left\{ 1 - q_p + \frac{q_p(1-r_p)G_p^*(s)}{1-r_pG_p^*(s)} \right\} \\ &+ \frac{C(m, m\rho_{p-1}^{+})(1-r_p)}{1-r_pG_p^*(s)}. \end{aligned} \tag{27}$$

The mean and second moment of $W_p^{\bullet}$ are

$$\begin{aligned} E[W_p^{\bullet}] &= \frac{[1 - C(m, m\rho_{p-1}^{+})]q_p + C(m, m\rho_{p-1}^{+})r_p}{m\mu(1-r_p)(1-\rho_p^{+})} \\ &= E[W_p] - E[W_p^{\circ}], \\ E[(W_p^{\bullet})^2] &= \frac{2(1-r_p\rho_p^{+})\{[1 - C(m, m\rho_{p-1}^{+})]q_p + C(m, m\rho_{p-1}^{+})r_p\}}{(m\mu)^2(1-r_p)^2(1-\rho_p^{+})^3}. \end{aligned}$$

We also have

$$E[W_p^{\circ}W_p^{\bullet}] = \frac{C(m, m\rho_{p-1}^{+})r_p}{(m\mu)^2(1-r_p)(1-\rho_{p-1}^{+})(1-\rho_p^{+})^2}, \tag{28}$$

which confirms the relation

$$E[W_p^2] = E[(W_p^{\circ} + W_p^{\bullet})^2] = E[(W_p^{\circ})^2] + 2E[W_p^{\circ}W_p^{\bullet}] + E[(W_p^{\bullet})^2].$$

For $m = 1$ (a single-server queue), we have

$$\tilde{W}_p^*(s,s') = \frac{1-\rho_{p-1}^{+}}{[1-\rho_{p-1}^{+}G_p^*(s)][1+\rho_p^{+}-\rho_p^{+}G_p^*(s')]} = W_p^{\circ}(s)W_p^{\bullet}(s'),$$

where

$$W_p^{\bullet}(s) = \frac{1}{1+\rho_p^{+}-\rho_p^{+}G_p^*(s)} = \frac{\mu}{\lambda_p^{+}+\mu-\lambda_p^{+}G_p^*(s)}.$$

Therefore, W_p° and $W_p^{\bullet}$ are independent.

8. Service Times

Another quantity of interest is the service time received by each class-p customer before service completion. We consider two types of service time.

Let $V^*_{p,k}(s)$ be the LST of the DF for the time to preemption, denoted $V_{p,k}$, for a class-p customer who competes for service with k other customers, where $0 \le k \le m-1$. Referring to Fig. 5, the complete set of m equations for $\{V^*_{p,k}(s); 0 \le k \le m-1\}$ is

$$\begin{aligned}
(s+\lambda_p^+ + \mu)V^*_{p,0}(s) &= \lambda_p^+ V^*_{p,1}(s),\\
[s+\lambda_p^+ + (k+1)\mu]V^*_{p,k}(s) &= \lambda_p^+ V^*_{p,k+1}(s) + k\mu V^*_{p,k-1}(s)\\
&\qquad\qquad 1 \le k \le m-2,\\
(s+\lambda_p^+ + m\mu)V^*_{p,m-1}(s) &= \lambda_p^+ + (m-1)\mu V^*_{p,m-2}(s).
\end{aligned} \tag{29}$$

We note that $P_{p,k}\{\mathrm{Pr}\} = V^*_{p,k}(0)$ for $0 \le k \le m-1$. The mean $E[V_{p,k}] = -dV^*_{p,k}(s)/ds|_{s=0}$ is given by

$$\begin{aligned}
E[V_{p,0}] &= \sum_{j=1}^{m} \frac{(m\rho_p^+)^j}{j!} \sum_{l=0}^{j-1} P_{p,l}\{\mathrm{Pr}\} \Bigg/ \lambda_p^+ \sum_{j=0}^{m} \frac{(m\rho_p^+)^j}{j!},\\
E[V_{p,m-1}] &= \sum_{j=0}^{m-1} \frac{(m\rho_p^+)^j}{j!} \sum_{l=j}^{m-1} P_{p,l}\{\mathrm{Pr}\} \Bigg/ m\mu \sum_{j=0}^{m} \frac{(m\rho_p^+)^j}{j!},
\end{aligned} \tag{30}$$

and

$$E[V_{p,k}] = \frac{\left\{ \begin{array}{l} \left[\displaystyle\sum_{j=k+1}^{m} \frac{(m\rho_p^+)^j}{j!} \sum_{l=0}^{j-1} P_{p,l}\{\mathrm{Pr}\}\right] \displaystyle\sum_{j=0}^{k} \frac{(m\rho_p^+)^j}{j!} \\ -\left[\displaystyle\sum_{j=1}^{k} \frac{(m\rho_p^+)^j}{j!} \sum_{l=0}^{j-1} P_{p,l}\{\mathrm{Pr}\}\right] \displaystyle\sum_{j=k+1}^{m} \frac{(m\rho_p^+)^j}{j!} \end{array} \right\}}{\lambda_p^+ \dfrac{(m\rho_p^+)^k}{k!} \displaystyle\sum_{j=0}^{m} \frac{(m\rho_p^+)^j}{j!}} \qquad 1 \le k \le m-2. \tag{31}$$

The second moment $E[V^2_{p,k}] = d^2V^*_{p,k}(s)/ds^2|_{s=0}$ is obtained using

$$E[V^2_{p,0}] = 2\sum_{j=1}^{m} \frac{(m\rho_p^+)^j}{j!} \sum_{l=0}^{j-1} E[V_{p,l}] \Bigg/ \lambda_p^+ \sum_{j=0}^{m} \frac{(m\rho_p^+)^j}{j!},$$

$$E[V_{p,m-1}^2] = 2\sum_{j=0}^{m-1}\frac{(m\rho_p^+)^j}{j!}\sum_{l=j}^{m-1}E[V_{p,l}] \Bigg/ m\mu\sum_{j=0}^{m}\frac{(m\rho_p^+)^j}{j!} ,$$

and

$$\frac{E[V_{p,k}^2]}{2} = \frac{\left\{\begin{array}{l}\left[\sum_{j=k+1}^{m}\frac{(m\rho_p^+)^j}{j!}\sum_{l=0}^{j-1}E[V_{p,l}]\right]\sum_{j=0}^{k}\frac{(m\rho_p^+)^j}{j!} \\ -\left[\sum_{j=1}^{k}\frac{(m\rho_p^+)^j}{j!}\sum_{l=0}^{j-1}E[V_{p,l}]\right]\sum_{j=k+1}^{m}\frac{(m\rho_p^+)^j}{j!}\end{array}\right\}}{\lambda_p^+\frac{(m\rho_p^+)^k}{k!}\sum_{j=0}^{m}\frac{(m\rho_p^+)^j}{j!}} \qquad 1 \le k \le m-2.$$

Let $U_{p,k}^*(s)$ be the LST of the DF for the time to service completion, denoted $U_{p,k}$, for a class-p customer who competes for service with k other customers, where $0 \le k \le m-1$. Referring to Fig. 5, the complete set of m equations for $\{U_{p,k}^*(s); 0 \le k \le m-1\}$ is

$$\begin{aligned}(s+\lambda_p^+ +\mu)U_{p,0}^*(s) &= \mu + \lambda_p^+ U_{p,1}^*(s),\\ [s+\lambda_p^+ +(k+1)\mu]U_{p,k}^*(s) &= \mu + \lambda_p^+ U_{p,k+1}^*(s) + k\mu U_{p,k-1}^*(s)\\ &\qquad 1 \le k \le m-2,\\ (s+\lambda_p^+ + m\mu)U_{p,m-1}^*(s) &= \mu + (m-1)\mu U_{p,m-2}^*(s).\end{aligned} \tag{32}$$

We note that $P_{p,k}\{\mathrm{Sr}\} = U_{p,k}^*(0)$ for $0 \le k \le m-1$. The mean $E[U_{p,k}] = -dU_{p,k}^*(s)/ds|_{s=0}$ is given by

$$E[U_{p,0}] = \sum_{j=1}^{m}\frac{(m\rho_p^+)^j}{j!}\sum_{l=0}^{j-1}P_{p,l}\{\mathrm{Sr}\} \Bigg/ \lambda_p^+\sum_{j=0}^{m}\frac{(m\rho_p^+)^j}{j!} ,$$

$$E[U_{p,m-1}] = \sum_{j=0}^{m-1}\frac{(m\rho_p^+)^j}{j!}\sum_{l=j}^{m-1}P_{p,l}\{\mathrm{Sr}\} \Bigg/ m\mu\sum_{j=0}^{m}\frac{(m\rho_p^+)^j}{j!} ,$$

and

$$E[U_{p,k}] = \frac{\left\{ \begin{array}{l} \left[\sum_{j=k+1}^{m} \frac{(m\rho_p^+)^j}{j!} \sum_{l=0}^{j-1} P_{p,l}\{\mathrm{Sr}\}\right] \sum_{j=0}^{k} \frac{(m\rho_p^+)^j}{j!} \\ -\left[\sum_{j=1}^{k} \frac{(m\rho_p^+)^j}{j!} \sum_{l=0}^{j-1} P_{p,l}\{\mathrm{Sr}\}\right] \sum_{j=k+1}^{m} \frac{(m\rho_p^+)^j}{j!} \end{array} \right\}}{\lambda_p^+ \frac{(m\rho_p^+)^k}{k!} \sum_{j=0}^{m} \frac{(m\rho_p^+)^j}{j!}} \qquad 1 \le k \le m-2.$$

The second moment $E[U_{p,k}^2] = d^2U_{p,k}^*(s)/ds^2|_{s=0}$ is obtained using

$$E[U_{p,0}^2] = 2\sum_{j=1}^{m} \frac{(m\rho_p^+)^j}{j!} \sum_{l=0}^{j-1} E[U_{p,l}] \Bigg/ \lambda_p^+ \sum_{j=0}^{m} \frac{(m\rho_p^+)^j}{j!} ,$$

$$E[U_{p,m-1}^2] = 2\sum_{j=0}^{m-1} \frac{(m\rho_p^+)^j}{j!} \sum_{l=j}^{m-1} E[U_{p,l}] \Bigg/ m\mu \sum_{j=0}^{m} \frac{(m\rho_p^+)^j}{j!} ,$$

and

$$\frac{E[U_{p,k}^2]}{2} = \frac{\left\{ \begin{array}{l} \left[\sum_{j=k+1}^{m} \frac{(m\rho_p^+)^j}{j!} \sum_{l=0}^{j-1} E[U_{p,l}]\right] \sum_{j=0}^{k} \frac{(m\rho_p^+)^j}{j!} \\ -\left[\sum_{j=1}^{k} \frac{(m\rho_p^+)^j}{j!} \sum_{l=0}^{j-1} E[U_{p,l}]\right] \sum_{j=k+1}^{m} \frac{(m\rho_p^+)^j}{j!} \end{array} \right\}}{\lambda_p^+ \frac{(m\rho_p^+)^k}{k!} \sum_{j=0}^{m} \frac{(m\rho_p^+)^j}{j!}} \qquad 1 \le k \le m-2.$$

We note the following relations:

$$E[U_{p,k}] + E[V_{p,k}] = \frac{P_{p,k}\{\mathrm{Sr}\}}{\mu} \quad ; \quad E[U_{p,k}^2] + E[V_{p,k}^2] = \frac{2E[U_{p,k}]}{\mu} \qquad 0 \le k \le m-1. \tag{33}$$

Let $S_{p,k}^*(s)$ be the LST of the DF for the total service time, denoted $S_{p,k}$, received by a class-p customer who competes for service with k other customers, where $k \geq 0$. If the customer waits upon arrival, or if his service resumes after a preemption, the LST of the DF of the service time until completion is given by

$$\sum_{i=0}^{\infty} U^*_{p,m-1}(s)[V^*_{p,m-1}(s)]^i = \frac{U^*_{p,m-1}(s)}{1 - V^*_{p,m-1}(s)}.$$

Thus,

$$S^*_{p,k}(s) = \begin{cases} U^*_{p,k}(s) + V^*_{p,k}(s)\dfrac{U^*_{p,m-1}(s)}{1 - V^*_{p,m-1}(s)} & 0 \leq k \leq m-1, \\ \dfrac{U^*_{p,m-1}(s)}{1 - V^*_{p,m-1}(s)} & k \geq m. \end{cases}$$

From Eqs. (29) and (32), we derive the following equations for $\{S^*_{p,k}(s); 0 \leq k \leq m-1\}$:

$$\begin{aligned} (s + \lambda_p^+ + \mu)S^*_{p,0}(s) &= \mu + \lambda_p^+ S^*_{p,1}(s), \\ [s + \lambda_p^+ + (k+1)\mu]S^*_{p,k}(s) &= \mu + \lambda_p^+ S^*_{p,k+1}(s) + k\mu S^*_{p,k-1}(s) \\ & \qquad\qquad 1 \leq k \leq m-2, \\ (s + m\mu)S^*_{p,m-1}(s) &= \mu + (m-1)\mu S^*_{p,m-2}(s). \end{aligned} \tag{34}$$

The solution

$$S^*_{p,k}(s) = \frac{\mu}{s+\mu} \qquad 0 \leq k \leq m-1$$

does not depend on k. It follows that

$$\frac{U^*_{p,m-1}(s)}{1 - V^*_{p,m-1}(s)} = S^*_{p,m-1}(s) = \frac{\mu}{s+\mu},$$

which yields the relation

$$\begin{aligned} \frac{U^*_{p,k}(s)}{1 - V^*_{p,k}(s)} &= \frac{\mu}{s+\mu} \qquad 0 \leq k \leq m-1, \\ S^*_{p,k}(s) &= \frac{\mu}{s+\mu} \qquad k \geq m. \end{aligned}$$

This result implies that $S_{p,k}$ is exponentially distributed with a mean $1/\mu$ regardless of the number k (≥ 0) of competing customers present in the system.

Then the LST of the DF of the overall service time received by a class-p customer is given by

$$
\begin{aligned}
S_p^*(s) &= \sum_{k=0}^{\infty} Q_{p-1,k}^+ S_{p,k}^*(s) = \sum_{k=0}^{m-1} Q_{p-1,k}^+ S_{p,k}^*(s) + \sum_{k=m}^{\infty} Q_{p-1,k}^+ S_{p,k}^*(s) \\
&= \sum_{k=0}^{m-1} Q_{p-1,k}^+ U_{p,k}^*(s) \\
&+ \left\{ \sum_{k=0}^{m-1} Q_{p-1,k}^+ V_{p,k}^*(s) + C(m, m\rho_{p-1}^+) \right\} \frac{U_{p,m-1}^*(s)}{1 - V_{p,m-1}^*(s)} \\
&= \frac{\mu}{s+\mu},
\end{aligned}
\tag{35}
$$

signifying that the service time S_p is exponentially distributed with mean $1/\mu$.

We note that W_p° and the S_p are independent, because the joint LST of their DFs is given as the product of their individual LSTs of the DF:

$$
\begin{aligned}
& E\left[e^{-sW_p^\circ - s'S_p}\right] = \sum_{k=0}^{m-1} Q_{p-1,k}^+ U_{p,k}^*(s') \\
& + \left\{ \sum_{k=0}^{m-1} Q_{p-1,k}^+ V_{p,k}^*(s') + C(m, m\rho_{p-1}^+) W_p^+(s) \right\} \frac{U_{p,m-1}^*(s')}{1 - V_{p,m-1}^*(s')} \\
& = \sum_{k=0}^{m-1} Q_{p-1,k}^+ \left[U_{p,k}^*(s') + V_{p,k}^*(s') \frac{\mu}{s'+\mu} \right] + C(m, m\rho_{p-1}^+) W_p^+(s) \frac{\mu}{s'+\mu} \\
& = \left[1 - C(m, m\rho_{p-1}^+) + C(m, m\rho_{p-1}^+) W_p^+(s) \right] \frac{\mu}{s'+\mu} \\
& = W_p^\circ(s) S_p^*(s').
\end{aligned}
$$

The independence of the service time and the waiting time prior to his first service is a reasonable conclusion, because the service time that any customer receives is independent of what has occurred in the system before his service is started for the first time. It follows that $\mathrm{Cov}[W_p^\circ, S_p] = 0$. Thus,

$$
\begin{aligned}
\mathrm{Cov}[W_p, S_p] &= \mathrm{Cov}[W_p^\circ + W_p^\bullet, S_p] = \mathrm{Cov}[W_p^\circ, S_p] + \mathrm{Cov}[W_p^\bullet, S_p] \\
&= \mathrm{Cov}[W_p^\bullet, S_p],
\end{aligned}
\tag{36}
$$

which is given in Eq. (39) below.

For $m = 1$ (a single-server queue), we have explicitly

$$V^*_{p,0}(s) = \frac{\lambda_p^+}{s + \lambda_p^+ + \mu} \quad ; \quad U^*_{p,0}(s) = \frac{\mu}{s + \lambda_p^+ + \mu}.$$

9. Response Time

The response time is the sum of the waiting and service times (see Fig. 1). The response time of a class-p customer is

$$T_p = W_p + S_p.$$

Combining the arguments for deriving the LST of the DF for W_p in Eq. (22) and the LST of the DF for S_p in Eq. (35), we obtain the joint LST of the DFs for W_p and S_p:

$$\begin{aligned}\tilde{T}^*_p(s, s') &:= E\left[e^{-sW_p - s'S_p}\right] = \sum_{k=0}^{m-1} Q^+_{p-1,k} U^*_{p,k}(s') \\ &+ \left\{ G^*_p(s) \sum_{k=0}^{m-1} Q^+_{p-1,k} V^*_{p,k}(s') + C(m, m\rho^+_{p-1}) W^+_p(s) \right\} \\ &\times \frac{U^*_{p,m-1}(s')}{1 - G^*_p(s) V^*_{p,m-1}(s')}. \end{aligned} \tag{37}$$

The marginal distributions yield

$$W^*_p(s) = \tilde{T}^*_p(s, 0) \quad ; \quad S^*_p(s) = \tilde{T}^*_p(0, s)$$

in agreement with Eqs. (22) and (35), respectively.

The LST of the DF for T_p is

$$\begin{aligned}T^*_p(s) &= \tilde{T}^*_p(s, s) = \sum_{k=0}^{m-1} Q^+_{p-1,k} U^*_{p,k}(s) \\ &+ \left\{ G^*_p(s) \sum_{k=0}^{m-1} Q^+_{p-1,k} V^*_{p,k}(s) + C(m, m\rho^+_{p-1}) W^+_p(s) \right\} \\ &\times \frac{U^*_{p,m-1}(s)}{1 - G^*_p(s) V^*_{p,m-1}(s)}. \end{aligned} \tag{38}$$

Clearly, $T_p^*(s)$ is not identical to $W_p^*(s)S_p^*(s)$, signifying that W_p and S_p are not independent. In fact, they are positively correlated, as expressed by their covariance, obtained using Eq. (37):

$$\mathrm{Cov}[W_p, S_p] = \frac{\sum_{k=0}^{m-1} Q_{p-1,k}^{+} E[V_{p,k}]}{m\mu(1-r_p)(1-\rho_p^{+})} + \frac{\{[1-C(m, m\rho_{p-1}^{+})]q_p + C(m, m\rho_{p-1}^{+})\}E[V_{p,m-1}]}{m\mu(1-r_p)^2(1-\rho_p^{+})}, \tag{39}$$

which equals $\mathrm{Cov}[W_p^{\bullet}, S_p]$ via Eq. (36), where $E[V_{p,k}]$ and $E[V_{p,m-1}]$ are specified by Eqs. (31) and (30), respectively. We also have

$$\mathrm{Cov}[T_p, S_p] = \mathrm{Cov}[W_p + S_p, S_p] = \mathrm{Cov}[W_p, S_p].$$

Figure 8 plots $\mathrm{Cov}[W_p, S_p]$ for the numerical example described in Section 1. Table 3 compares the $\mathrm{Cov}[W_p, S_p]_{\mathrm{FCFS}}$ values calculated from [9, 10, 15] (note that $\mathrm{Cov}[W_1, S_1]_{\mathrm{FCFS}} = 0$) and the $\mathrm{Cov}[W_p, S_p]_{\mathrm{LCFS}}$ values obtained from Eq. (39) for the same numerical example. We observe that $\mathrm{Cov}[W_p, S_p]_{\mathrm{LCFS}}$ is much greater than $\mathrm{Cov}[W_p, S_p]_{\mathrm{FCFS}}$.

Equation (38) yields the mean response time $E[T_p]$ already given in Eq. (3). The second moment of the response time is given by

$$E[T_p^2] = E[(W_p + S_p)^2]$$

$$= E[W_p^2] + 2\mathrm{Cov}[W_p, S_p] + 2E[W_p]E[S_p] + E[S_p^2]$$

$$= \frac{2\sum_{k=0}^{m-1} Q_{p-1,k}^{+} E[V_{p,k}]}{m\mu(1-r_p)(1-\rho_p^{+})} + \frac{2\{[1-C(m, m\rho_{p-1}^{+})]q_p + C(m, m\rho_{p-1}^{+})\}E[V_{p,m-1}]}{m\mu(1-r_p)^2(1-\rho_p^{+})}$$

$$+ \frac{2[1-C(m, m\rho_{p-1}^{+})]q_p}{(m\mu)^2}\left[\frac{1-r_p\rho_p^{+}}{(1-r_p)^2(1-\rho_p^{+})^3} + \frac{m}{(1-r_p)(1-\rho_p^{+})}\right]$$

$$+ \frac{2C(m, m\rho_{p-1}^{+})}{(m\mu)^2}\left[\frac{1-\rho_{p-1}^{+}\rho_p^{+}}{(1-\rho_{p-1}^{+})^2(1-\rho_p^{+})^3}\right.$$

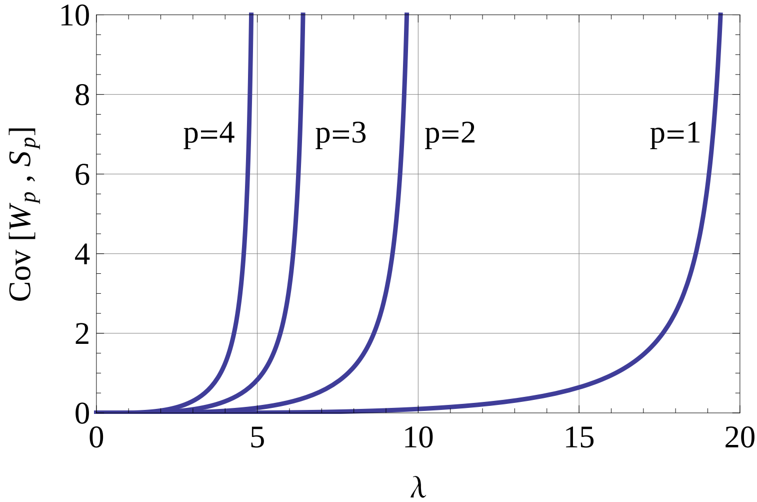

Figure 8: Covariance of the total waiting and service times in the M/M/m LCFS preemptive-resume priority queue

$$+ \frac{r_p[2-\rho_{p-1}^{+}-\rho_p^{+}-r_p(1-\rho_{p-1}^{+}\rho_p^{+})]}{(1-r_p)^2(1-\rho_{p-1}^{+})(1-\rho_p^{+})^3}$$

$$\left. + \frac{m(1-r_p\rho_{p-1}^{+})}{(1-r_p)(1-\rho_{p-1}^{+})(1-\rho_p^{+})}\right] + \frac{2}{\mu^2}. \tag{40}$$

We have confirmed numerically that this formula yields the same result as that shown in Fig. 3 using Eq. (15).

For $m = 1$ (a single-server queue),

$$\tilde{T}_p^*(s,s') = \frac{\mu(1-\rho_{p-1}^{+})}{[1-\rho_{p-1}^{+}G_p^*(s)][s'+\lambda_p^{+}+\mu-\lambda_p^{+}G_p^*(s)]}.$$

Thus, we get $\tilde{T}_p^*(s,0) = W_p^*(s)$ in Eq. (26) and $\tilde{T}_p^*(0,s) = S_p^*(s)$ in Eq. (35), but $\tilde{T}_p^*(s,s') \neq W_p^*(s)S_p^*(s')$. Therefore, W_p and S_p are not independent although S_p is exponentially distributed. In fact, we have

$$\mathrm{Cov}[W_p, S_p] = \frac{\rho_p^{+}}{\mu^2(1-\rho_p^{+})}.$$

Table 3: Comparison of covariance of the total waiting and service times in the M/M/m FCFS and LCFS preemptive-resume priority queues

λ	FCFS $\mathrm{Cov}[W_1, S_1]$	LCFS $\mathrm{Cov}[W_1, S_1]$
0	0.00000	0.00000
5	0.00000	0.00538
10	0.00000	0.09838
15	0.00000	0.64016
17	0.00000	1.46777
18	0.00000	2.52944

λ	FCFS $\mathrm{Cov}[W_2, S_2]$	LCFS $\mathrm{Cov}[W_2, S_2]$
0	0.00000	0.00000
2	0.00041	0.00275
4	0.00752	0.05088
6	0.03473	0.27059
7	0.06022	0.54888
8	0.09610	1.15200

λ	FCFS $\mathrm{Cov}[W_3, S_3]$	LCFS $\mathrm{Cov}[W_3, S_3]$
0	0.00000	0.00000
2	0.00554	0.01727
3	0.02782	0.08943
4	0.08186	0.29022
5	0.18593	0.83170
6	0.37162	3.17664

λ	FCFS $\mathrm{Cov}[W_4, S_4]$	LCFS $\mathrm{Cov}[W_4, S_4]$
0	0.00000	0.00000
2	0.02554	0.05829
3	0.11927	0.29978
4	0.35807	1.24268
4.5	0.58664	3.23941
4.8	0.79021	9.29258

10. Completion Time

Finally, we consider the *completion time* for a class-p customer, denoted Θ_p, which is defined as the time interval between initiation and completion of his service. Clearly, it is the sum of the time in limbo and the service time (see Fig. 1):

$$\Theta_p = W_p^{\bullet} + S_p.$$

Therefore, the joint LST of the DF for $W_p^{\bullet}$ and S_p is given by

$$\begin{aligned}\tilde{\Theta}_p^*(s, s') := E\left[e^{-sW_p^{\bullet} - s'S_p}\right] &= \sum_{k=0}^{m-1} Q_{p-1,k}^{+} U_{p,k}^*(s') \\ &+ \left\{ G_p^*(s) \sum_{k=0}^{m-1} Q_{p-1,k}^{+} V_{p,k}^*(s') + C(m, m\rho_{p-1}^{+}) \right\} \frac{U_{p,m-1}^*(s')}{1 - G_p^*(s) V_{p,m-1}^*(s')}\end{aligned}$$

with marginal distributions

$$W_p^{\bullet}(s) = \tilde{\Theta}_p^*(s,0) \quad ; \quad S_p^*(s) = \tilde{\Theta}_p^*(0,s) = \frac{\mu}{s+\mu}.$$

Also, the response time is the sum of the initial waiting time and the completion time (see Fig. 1):

$$T_p = W_p^{\circ} + \Theta_p.$$

The joint LST of the DF for W_p° and Θ_p is given by

$$\begin{aligned}\tilde{T}_p^{**}(s,s') := E\left[e^{-sW_p^{\circ} - s'\Theta_p}\right] = \sum_{k=0}^{m-1} Q_{p-1,k}^{+} U_{p,k}^*(s') \\ + \left\{ G_p^*(s') \sum_{k=0}^{m-1} Q_{p-1,k}^{+} V_{p,k}^*(s') + C(m, m\rho_{p-1}^{+}) W_p^{+}(s) \right\} \\ \times \frac{U_{p,m-1}^*(s')}{1 - G_p^*(s') V_{p,m-1}^*(s')}\end{aligned}$$

with

$$W_p^{\circ}(s) = \tilde{T}_p^{**}(s,0) \quad ; \quad T_p^*(s) = \tilde{T}_p^{**}(s,s)$$

in agreement with Eqs. (17) and (38), respectively. The LST of the DF for Θ_p is then given by

$$\begin{aligned}\Theta_p^*(s) = \tilde{\Theta}_p^*(s,s) = \tilde{T}_p^{**}(0,s) = \sum_{k=0}^{m-1} Q_{p-1,k}^{+} U_{p,k}^*(s) \\ + \left\{ G_p^*(s) \sum_{k=0}^{m-1} Q_{p-1,k}^{+} V_{p,k}^*(s) + C(m, m\rho_{p-1}^{+}) \right\} \frac{U_{p,m-1}^*(s)}{1 - G_p^*(s) V_{p,m-1}^*(s)}.\end{aligned}$$

The mean completion time is given by

$$\begin{aligned} E[\Theta_p] &= \frac{[1 - C(m, m\rho_{p-1}^{+})]q_p + C(m, m\rho_{p-1}^{+}) r_p}{m\mu(1-r_p)(1-\rho_p^{+})} + \frac{1}{\mu} \\ &= E[W_p^{\bullet}] + E[S_p] = E[T_p] - E[W_p^{\circ}] \end{aligned}$$

and its second moment by

$$E[\Theta_p^2] = E[(W_p^{\bullet} + S_p)^2] = E[(W_p^{\bullet})^2] + 2\mathrm{Cov}[W_p^{\bullet}, S_p] + 2\frac{E[W_p^{\bullet}]}{\mu} + \frac{2}{\mu^2}.$$

The mutual independence of the initial waiting and service times (see Section 8) implies

$$\begin{aligned}\mathrm{Cov}[W_p^\circ,\Theta_p] &= \mathrm{Cov}[W_p^\circ, W_p^\bullet + S_p] = \mathrm{Cov}[W_p^\circ, W_p^\bullet] + \mathrm{Cov}[W_p^\circ, S_p] \\ &= \mathrm{Cov}[W_p^\circ, W_p^\bullet],\end{aligned}$$

which can be computed from Eq. (28). We also get

$$\mathrm{Cov}[\Theta_p, S_p] = \mathrm{Cov}[W_p^\bullet + S_p, S_p] = \mathrm{Cov}[W_p^\bullet, S_p] = \mathrm{Cov}[W_p, S_p],$$

which is given in Eq. (39).

We can decompose T_p into three non-overlapping periods W_p°, $W_p^\bullet$, and S_p:

$$T_p = \overbrace{W_p^\circ + W_p^\bullet}^{W_p} \!\!\!\!\!\!\!\!\underbrace{ + S_p}_{\Theta_p}$$

with the joint LST of the three DFs given by

$$\begin{aligned}\tilde{T}_p^{***}(s,s',s'') &:= E\left[e^{-sW_p^\circ - s'W_p^\bullet - s''S_p}\right] = \sum_{k=0}^{m-1} Q_{p-1,k}^+ U_{p,k}^*(s'') \\ &+ \left\{G_p^*(s') \sum_{k=0}^{m-1} Q_{p-1,k}^+ V_{p,k}^*(s'') + C(m, m\rho_{p-1}^+) W_p^+(s)\right\} \\ &\times \frac{U_{p,m-1}^*(s'')}{1 - G_p^*(s') V_{p,m-1}^*(s'')}.\end{aligned}$$

For $m = 1$ (a single-server queue), we have

$$\begin{aligned}\tilde{\Theta}_p^*(s,s') &= \frac{\mu}{s' + \lambda_p^+ + \mu - \lambda_p^+ G_p^*(s)}, \\ \tilde{T}_p^{**}(s,s') &= \frac{\mu(1-\rho_{p-1}^+)}{[1-\rho_{p-1}^+ G_p^*(s)][s' + \lambda_p^+ + \mu - \lambda_p^+ G_p^*(s')]} \\ &= W_p^\circ(s)\Theta_p^*(s'), \\ \tilde{T}_p^{***}(s,s',s'') &= \frac{\mu(1-\rho_{p-1}^+)}{[1-\rho_{p-1}^+ G_p^*(s)][s'' + \lambda_p^+ + \mu - \lambda_p^+ G_p^*(s')]}\end{aligned}$$

$$= W_p^\circ(s)\tilde{\Theta}_p^*(s', s'').$$

Therefore, W_p° and Θ_p are independent and

$$\Theta_p^*(s) = \frac{\mu}{s + \lambda_p^+ + \mu - \lambda_p^+ G_p^*(s)} = G_p^*(s)$$

as expected.

11. Concluding Remarks

We have done a detailed study of the waiting, service, and response times of an arbitrary customer belonging to each class in an M/M/m *immediate* preemptive-resume priority queue with an LCFS service and FCFD displacement discipline applied within the same class. Our result for the special case of a single-server queue differs from that of Durr [4] for an M/M/1 LCFS *deferred* preemptive-resume priority queue. In his model, it is assumed that an arriving class-p customer does not preempt the service of another class-p customer. That is to say, only if the arriving customer has a strictly lower class number than the customer being served, then the server interrupts the current service and immediately starts serving the arriving customer. In the Appendix of this paper, we provide modification of his analysis for our preemption protocol.

We have derived explicit expressions for the LST of the DF for the initial waiting time, the time in limbo, and the total waiting time, and also their mean and moments. We have shown that the received service time is exponentially distributed and that it is independent of the initial waiting time. However, the received service time has a positive correlation with the time in limbo and the total waiting time. We have presented a recursive algorithm for calculating the moments of the response and completion times. A remaining task is to find the LST of the DF explicitly for the response and completion times in the case of a generic number of servers.

We have validated the theoretical formulas using numerical examples and compared the performance of the present queue with that of the corresponding M/M/m FCFS preemptive-resume priority queue with LCFD displacement discipline.

Appendix
Waiting Time in the M/G/1 LCFS Preemptive-Resume Priority Queue: Modification to Durr (1969)

We consider an M/G/1 immediate preemptive-resume priority queue with an LCFS service and FCFD displacement discipline within the same class. Classes are indexed $1, 2, \ldots$ in order of decreasing priority, so that a class-p customer has priority of service over all customers of classes $p+1, p+2, \ldots$. The arrival of class-p customers follows a Poisson process with a rate λ_p (> 0), independent of the arrival of customers of any other classes. The LST of the DF for the service time of a class-p customer is denoted $H_p^*(s)$ with mean h_p and ℓth moment $h_p^{(\ell)}, \ell = 2, 3, \ldots$. The LST of the DF for the service time of an arbitrary customer of classes $1, 2, \ldots, p$ is denoted $H_p^+(s)$ with mean h_p^+ and ℓth moment $h_p^{+(\ell)}, \ell = 2, 3, \ldots$. We then introduce the notation

$$\lambda_p^+ := \sum_{k=1}^{p} \lambda_k \quad ; \quad H_p^+(s) := \frac{1}{\lambda_p^+} \sum_{k=1}^{p} \lambda_k H_k^*(s),$$

$$h_p^+ := \frac{1}{\lambda_p^+} \sum_{k=1}^{p} \lambda_k h_k \quad ; \quad h_p^{+(\ell)} := \frac{1}{\lambda_p^+} \sum_{k=1}^{p} \lambda_k h_k^{(\ell)} \qquad \ell = 2, 3, \ldots,$$

$$\rho_p := \lambda_p h_p \quad ; \quad \rho_p^+ := \sum_{k=1}^{p} \rho_k = \lambda_p^+ h_p^+.$$

Unlike Durr's model [4], it is assumed in our service preemption protocol that the arrival of a class-p customer preempts the on-going service to another class-p customer. In other respects, our analysis proceeds in the same fashion as in [4].

(1) We denote by $\mathcal{D}_p$ the duration of a busy period generated by customers of class $1, 2, \ldots, p$. The LST of the DF, mean, and the moments for $\mathcal{D}_p$ are given by

$$D_p^*(s) = H_p^+[s + \lambda_p^+ - \lambda_p^+ D_p^*(s)],$$

and

$$E[\mathcal{D}_p] = \frac{h_p^+}{1 - \rho_p^+} \quad ; \quad E[\mathcal{D}_p^2] = \frac{h_p^{+(2)}}{(1 - \rho_p^+)^3},$$

$$E[\mathcal{D}_p^3] = \frac{h_p^{+(3)}}{(1 - \rho_p^+)^4} + \frac{3\lambda_p^+ [h_p^{+(2)}]^2}{(1 - \rho_p^+)^5},$$

$$E[\mathcal{D}_p^4] = \frac{h_p^{+(4)}}{(1-\rho_p^+)^5} + \frac{10\lambda_p^+ h_p^{+(2)} h_p^{+(3)}}{(1-\rho_p^+)^6} + \frac{15(\lambda_p^+)^2 [h_p^{+(2)}]^3}{(1-\rho_p^+)^7}.$$

(2) We denote by $\mathcal{G}_p$ a time interval starting with the service time of a class-p customer and ending with the service completion of that customer. The LST of the DF for $\mathcal{G}_p$ is given by

$$G_p^*(s) = H_p^*[s + \lambda_p^+ - \lambda_p^+ D_p^*(s)],$$

where we note the relation

$$D_p^*(s) = \frac{\lambda_p}{\lambda_p^+} G_p^*(s) + \frac{\lambda_{p-1}^+}{\lambda_p^+} D_{p-1}^*[s + \lambda_p - \lambda_p G_p^*(s)].$$

The mean and moments of $\mathcal{G}_p$ are given by

$$E[\mathcal{G}_p] = h_p(1 + \lambda_p^+ E[\mathcal{D}_p]) = \frac{h_p}{1-\rho_p^+},$$

$$E[\mathcal{G}_p^2] = \frac{h_p^{(2)}}{(1-\rho_p^+)^2} + \frac{\lambda_p^+ h_p h_p^{+(2)}}{(1-\rho_p^+)^3},$$

$$E[\mathcal{G}_p^3] = \frac{h_p^{(3)}}{(1-\rho_p^+)^3} + \frac{\lambda_p^+[h_p h_p^{+(3)} + 3h_p^{(2)} h_p^{+(2)}]}{(1-\rho_p^+)^4} + \frac{3h_p[\lambda_p^+ h_p^{+(2)}]^2}{(1-\rho_p^+)^5}.$$

(3) We denote by $\mathcal{M}_p$ the duration of a time interval from the arrival time of a tagged class-p customer (an arbitrary time) to the moment at which all customers of classes $1, 2, \ldots, p-1$ who are present in the system at the arrival time of the tagged customer have been served. This duration is equivalent to the waiting time of an arbitrary customer in an FCFS M/G/1 queue (without priorities) with customers of classes $1, 2, \ldots, p-1$ and the service time for which the LST of the DF is given by $H_{p-1}^+(s)$ [3, p. 217]. Therefore, the LST of the DF, mean, and the moments for $\mathcal{M}_p$ are given by

$$M_p^*(s) = \frac{(1-\rho_{p-1}^+)s}{s - \lambda_{p-1}^+ + \lambda_{p-1}^+ H_{p-1}^+(s)},$$

and

$$
\begin{aligned}
E[\mathcal{M}_p] &= \frac{\lambda_{p-1}^{+} h_{p-1}^{+(2)}}{2(1-\rho_{p-1}^{+})}, \\
E[\mathcal{M}_p^2] &= \frac{\lambda_{p-1}^{+} h_{p-1}^{+(3)}}{3(1-\rho_{p-1}^{+})} + \frac{[\lambda_{p-1}^{+} h_{p-1}^{+(2)}]^2}{2(1-\rho_{p-1}^{+})^2}, \\
E[\mathcal{M}_p^3] &= \frac{\lambda_{p-1}^{+} h_{p-1}^{+(4)}}{4(1-\rho_{p-1}^{+})} + \frac{(\lambda_{p-1}^{+})^2 h_{p-1}^{+(2)} h_{p-1}^{+(3)}}{(1-\rho_{p-1}^{+})^2} + \frac{3[\lambda_{p-1}^{+} h_{p-1}^{+(2)}]^3}{4(1-\rho_{p-1}^{+})^3}.
\end{aligned}
$$

This characterization of $\mathcal{M}_p$ differs from Durr [4] in the sense that the influence of customers of classes $p, p+1, \ldots$ does not appear in $\mathcal{M}_p$ in the above.

(4) We denote by $\mathcal{Q}_p$ the duration of a time interval from the arrival time of a tagged class-p customer to the moment at which all customers of classes $1, 2, \ldots, p-1$ who are present in the system at the arrival time of the tagged customer as well as those of classes $1, 2, \ldots, p-1$ who arrive after the tagged customer have been served. The LST of the DF, mean, and the moments for $\mathcal{Q}_p$ are given by

$$
\begin{aligned}
Q_p^*(s) &= M_p^*[s + \lambda_{p-1}^{+} - \lambda_{p-1}^{+} D_{p-1}^*(s)] \\
&= (1-\rho_{p-1}^{+}) \left\{ 1 + \frac{\lambda_{p-1}^{+}[1 - D_{p-1}^*(s)]}{s} \right\},
\end{aligned}
$$

and

$$
\begin{aligned}
E[\mathcal{Q}_p] &= E[\mathcal{M}_p](1 + \lambda_{p-1}^{+} E[\mathcal{D}_{p-1}]) = \frac{\lambda_{p-1}^{+}}{2}(1-\rho_{p-1}^{+}) E[\mathcal{D}_{p-1}^2] \\
&= \frac{\lambda_{p-1}^{+} h_{p-1}^{+(2)}}{2(1-\rho_{p-1}^{+})^2}, \\
E[\mathcal{Q}_p^2] &= \frac{\lambda_{p-1}^{+}}{3}(1-\rho_{p-1}^{+}) E[\mathcal{D}_{p-1}^3] \\
&= \frac{\lambda_{p-1}^{+} h_{p-1}^{+(3)}}{3(1-\rho_{p-1}^{+})^3} + \frac{[\lambda_{p-1}^{+} h_{p-1}^{+(2)}]^2}{(1-\rho_{p-1}^{+})^4},
\end{aligned}
$$

$$E[\mathcal{Q}_p^3] = \frac{\lambda_{p-1}^+}{4}(1-\rho_{p-1}^+)E[\mathcal{D}_{p-1}^4]$$

$$= \frac{\lambda_{p-1}^+ h_{p-1}^{+(4)}}{4(1-\rho_{p-1}^+)^4} + \frac{5(\lambda_{p-1}^+)^2 h_{p-1}^{+(2)} h_{p-1}^{+(3)}}{2(1-\rho_{p-1}^+)^5} + \frac{15[\lambda_{p-1}^+ h_{p-1}^{+(2)}]^3}{4(1-\rho_{p-1}^+)^6}.$$

(5) The waiting time of a class-p customer, denoted $\mathcal{W}_p^\circ$, is the duration of a time interval from the arrival time of a class-p customer to the moment at which there are no customers of classes $1, 2, \ldots, p$ present in the system. Therefore, the LST of the DF for $\mathcal{W}_p^\circ$ is given by

$$W_p^\circ(s) = Q_p^*[s+\lambda_p-\lambda_p G_p^*(s)]$$

$$= (1-\rho_{p-1}^+)\left\{1 + \frac{\lambda_{p-1}^+\{1-D_{p-1}^+[s+\lambda_p-\lambda_p G_p^*(s)]\}}{s+\lambda_p-\lambda_p G_p^*(s)}\right\}$$

$$= \frac{(1-\rho_{p-1}^+)[s+\lambda_p^+-\lambda_p^+ D_p^*(s)]}{s+\lambda_p-\lambda_p G_p^*(s)}.$$

Alternatively, $\mathcal{W}_p^\circ$ is a time interval to serve all customers of classes $1, 2, \ldots, p$ who arrive during $\mathcal{M}_p$. Thus the LST of the DF for $\mathcal{W}_p^\circ$ is also given by

$$W_p^\circ(s) = M_p^*[s+\lambda_p^+-\lambda_p^+ D_p^*(s)]$$

$$= \frac{(1-\rho_{p-1}^+)[s+\lambda_p^+-\lambda_p^+ D_p^*(s)]}{s+\lambda_p^+-\lambda_p^+ D_p^*(s)-\lambda_{p-1}^+ + \lambda_{p-1}^+ H_{p-1}^+[s+\lambda_p^+-\lambda_p^+ D_p^*(s)]}.$$

The two expressions for $W_p^\circ(s)$ are actually the same. From the relation

$$\lambda_p^+ H_p^+(s) = \lambda_{p-1}^+ H_{p-1}^+(s) + \lambda_p H_p^*(s),$$

we observe that

$$\lambda_p^+ D_p^*(s) = \lambda_{p-1}^+ H_{p-1}^+[s+\lambda_p^+-\lambda_p^+ D_p^*(s)] + \lambda_p G_p^*(s),$$

which leads to the same expression for $W_p^\circ(s)$ given above.

The mean and moments of $\mathcal{W}_p^\circ$ are given by

$$E[W_p^\circ] = E[\mathcal{Q}_p](1+\lambda_p E[\mathcal{G}_p]) = E[\mathcal{M}_p](1+\lambda_p^+ E[\mathcal{D}_p])$$

$$
\begin{aligned}
&= \frac{\lambda_{p-1}^{+} h_{p-1}^{+(2)}}{2(1-\rho_{p-1}^{+})(1-\rho_p^{+})}, \\
E[(W_p^{\circ})^2] &= \frac{\lambda_{p-1}^{+} h_{p-1}^{+(3)}}{3(1-\rho_{p-1}^{+})(1-\rho_p^{+})^2} + \frac{\lambda_{p-1}^{+}\lambda_p^{+} h_{p-1}^{+(2)} h_p^{+(2)}}{2(1-\rho_{p-1}^{+})(1-\rho_p^{+})^3} \\
&+ \frac{[\lambda_{p-1}^{+} h_{p-1}^{+(2)}]^2}{2(1-\rho_{p-1}^{+})^2(1-\rho_p^{+})^2}, \\
E[(W_p^{\circ})^3] &= \frac{\lambda_{p-1}^{+} h_{p-1}^{+(4)}}{4(1-\rho_{p-1}^{+})(1-\rho_p^{+})^3} \\
&+ \frac{\lambda_{p-1}^{+}\lambda_p^{+}[h_{p-1}^{+(2)} h_p^{+(3)} + 2h_p^{+(2)} h_{p-1}^{+(3)}]}{2(1-\rho_{p-1}^{+})(1-\rho_p^{+})^4} \\
&+ \frac{(\lambda_{p-1}^{+})^2 h_{p-1}^{+(2)} h_{p-1}^{+(3)}}{(1-\rho_{p-1}^{+})^2(1-\rho_p^{+})^3} + \frac{3[\lambda_{p-1}^{+} h_{p-1}^{+(2)}]^3}{4(1-\rho_{p-1}^{+})^3(1-\rho_p^{+})^3} \\
&+ \frac{3\lambda_{p-1}^{+} h_{p-1}^{+(2)}[\lambda_p^{+} h_p^{+(2)}]^2}{2(1-\rho_{p-1}^{+})(1-\rho_p^{+})^5} + \frac{3[\lambda_{p-1}^{+} h_{p-1}^{+(2)}]^2 \lambda_p^{+} h_p^{+(2)}}{2(1-\rho_{p-1}^{+})^2(1-\rho_p^{+})^4}.
\end{aligned}
$$

(6) The response time of a class-p customer, denoted $\mathcal{T}_p$, is the sum of the waiting time and the time interval from the start of his service to its completion, which are independent of each other. The latter is given by $\mathcal{G}_p$. Therefore, the LST of the DF for $\mathcal{T}_p$ is given by

$$T_p^*(s) = W_p^{\circ}(s) G_p^*(s).$$

The mean response time of a class-p customer is given by

$$E[T_p] = E[W_p^{\circ}] + E[\mathcal{G}_p] = \frac{\lambda_{p-1}^{+} h_{p-1}^{+(2)}}{2(1-\rho_{p-1}^{+})(1-\rho_p^{+})} + \frac{h_p}{1-\lambda_p^{+}},$$

which can also be derived from the recursive relation

$$E[T_p] = h_p + E[M_p] + \sum_{k=1}^{p} h_k \lambda_k E[T_p].$$

The higher-order moments of $\mathcal{T}_p$ can be obtained from those of W_p° and $\mathcal{G}_p$.

In a special case in which customers of all classes have the same service time distribution, we let

$$H_p^*(s) = H_p^+(s) = H^*(s), \quad h_p = h_p^+ = h, \quad h_p^{(\ell)} = h_p^{+(\ell)} = h^{(\ell)}$$

for $\ell = 2, 3, \ldots$, and $\rho_p^+ := \lambda_p^+ h$. Therefore we observe that

$$D_p^*(s) = G_p^*(s),$$

which leads to

$$\begin{aligned} E[\mathcal{D}_p] &= E[\mathcal{G}_p] = \frac{h}{1-\rho_p^+} \quad ; \quad E[\mathcal{D}_p^2] = E[\mathcal{G}_p^2] = \frac{h^{(2)}}{(1-\rho_p^+)^3}, \\ E[\mathcal{D}_p^3] &= E[\mathcal{G}_p^3] = \frac{h^{(3)}}{(1-\rho_p^+)^4} + \frac{3\lambda_p^+[h^{(2)}]^2}{(1-\rho_p^+)^5}, \\ E[\mathcal{D}_p^4] &= E[\mathcal{G}_p^4] = \frac{h^{(4)}}{(1-\rho_p^+)^5} + \frac{10\lambda_p^+ h^{(2)} h^{(3)}}{(1-\rho_p^+)^6} + \frac{15(\lambda_p^+)^2[h^{(2)}]^3}{(1-\rho_p^+)^7}. \end{aligned}$$

The LST of the DF for the waiting time of a class-p customer is then given by

$$\begin{aligned} W_p^\circ(s) &= \frac{(1-\rho_{p-1}^+)[s + \lambda_p^+ - \lambda_p^+ G_p^*(s)]}{s + \lambda_p - \lambda_p G_p^*(s)} \\ &= (1-\rho_{p-1}^+)\left\{1 + \frac{\lambda_{p-1}^+[1 - G_p^*(s)]}{s + \lambda_p - \lambda_p G_p^*(s)}\right\} \end{aligned}$$

with mean and moments

$$\begin{aligned} E[W_p^\circ] &= \frac{\lambda_{p-1}^+ h^{(2)}}{2(1-\rho_{p-1}^+)(1-\rho_p^+)}, \\ E[(W_p^\circ)^2] &= \frac{\lambda_{p-1}^+ h^{(3)}}{3(1-\rho_{p-1}^+)(1-\rho_p^+)^2} \\ &+ \frac{\lambda_{p-1}^+[h^{(2)}]^2(\rho_{p-1}^+ + \rho_p^+ - 2\rho_{p-1}^+\rho_p^+)}{2h(1-\rho_{p-1}^+)^2(1-\rho_p^+)^3}, \\ E[(W_p^\circ)^3] &= \frac{\lambda_{p-1}^+ h^{(4)}}{4(1-\rho_{p-1}^+)(1-\rho_p^+)^3} \end{aligned}$$

$$+ \frac{\lambda_{p-1}^{+} h^{(2)} h^{(3)} (2\rho_{p-1}^{+} + 3\rho_p^{+} - 5\rho_{p-1}^{+}\rho_p^{+})}{2h(1-\rho_{p-1}^{+})^2(1-\rho_p^{+})^4}$$

$$+ \frac{3\lambda_{p-1}^{+}[h^{(2)}]^3 \left\{ \begin{array}{l} 2\rho_{p-1}^{+}\rho_p^{+}(1-3\rho_p^{+}) + 2(\rho_p^{+})^2 \\ +(\rho_{p-1}^{+})^2[1-4\rho_p^{+}+5(\rho_p^{+})^2] \end{array} \right\}}{4h^2(1-\rho_{p-1}^{+})^3(1-\rho_p^{+})^5}.$$

In a further special case in which customers of all classes have the same exponentially distributed service times with mean $1/\mu$, we have

$$H^*(s) = \frac{\mu}{s+\mu}, \quad h = \frac{1}{\mu}, \quad h^{(2)} = \frac{2}{\mu^2}, \quad h^{(3)} = \frac{6}{\mu^3}, \quad h^{(4)} = \frac{24}{\mu^4}.$$

Then the quadratic equation for $G_p^*(s)$ is

$$\lambda_p^{+}[G_p^*(s)]^2 - (s + \lambda_p^{+} + \mu)G_p^*(s) + \mu = 0,$$

which yields

$$E[\mathcal{G}_p] = \frac{1}{\mu(1-\rho_p^{+})}, \quad E[\mathcal{G}_p^2] = \frac{2}{\mu^2(1-\rho_p^{+})^3}, \quad E[\mathcal{G}_p^3] = \frac{6(1+\rho_p^{+})}{\mu^3(1-\rho_p^{+})^5}.$$

The LST of the DF for the waiting time of a class-p customer is given by

$$W_p^{\circ}(s) = \frac{1-\rho_{p-1}^{+}}{1-\rho_{p-1}^{+}G_p^*(s)}$$

with mean and moments

$$E[W_p^{\circ}] = \frac{\rho_{p-1}^{+}}{\mu(1-\rho_{p-1}^{+})(1-\rho_p^{+})},$$

$$E[(W_p^{\circ})^2] = \frac{2\rho_{p-1}^{+}(1-\rho_{p-1}^{+}\rho_p^{+})}{\mu^2(1-\rho_{p-1}^{+})^2(1-\rho_p^{+})^3},$$

$$E[(W_p^{\circ})^3] = \frac{6\rho_{p-1}^{+}[1+\rho_p^{+}-4\rho_{p-1}^{+}\rho_p^{+}+(\rho_{p-1}^{+})^2\rho_p^{+}+(\rho_{p-1}^{+}\rho_p^{+})^2]}{\mu^3(1-\rho_{p-1}^{+})^3(1-\rho_p^{+})^5}.$$

This result agrees with Eq. (19).

Acknowledgments

This work was supported by the Grant in Aid for Scientific Research (C) No. 26330354 from the Japan Society for the Promotion of Science (JSPS) in 2016.

References

[1] I. Brosh, Preemptive priority assignment in multichannel systems, *Opns. Res.*, **17** (1969), 526–535, http://dx.doi.org/10.1287/opre.17.3.526

[2] J. P. Buzen and A. B. Bondi, The response times of priority classes under preemptive resume in M/M/m queues, *Opns. Res.*, **31** (1983), 456–465, http://dx.doi.org/10.1287/opre.31.3.456

[3] R. B. Cooper, *Introduction to Queueing Theory*, Second edition, Elsevier North Holland, 1981.

[4] L. Durr, A single-server priority queuing system with general holding times, Poisson input, and reverse-order-of-arrival queuing discipline. *Opns. Res.*, **17** (1969), 351–358, http://dx.doi.org/10.1287/opre.17.2.351

[5] M. Fujiki, Fundamental theory and application on communication traffic. 5 queueing theory (part 2), *Transactions of the Institute of Electronics and Communication Engineers of Japan*, Vol.55, No.9, pp.1194–1200, September 1972 (in Japanese).

[6] L. Katzschner, Loss systems with displacing priorities. *Sixth Int. Teletraffic Cong.*, pp.224/1–224/8, Munich, 1970.

[7] M. Segal, A multiserver system with preemptive priorities, *Opns. Res.*, **18** (1970), 316–323, http://dx.doi.org/10.1287/opre.18.2.316

[8] H. Takagi, *Queueing Analysis: A Foundation of Performance Evaluation*, Volume 1: *Vacation and Priority Systems*, North-Holland, Elsevier Science Publishers B. V., 1991.

[9] H. Takagi, Detailed analysis of the response time and waiting time in the M/M/m FCFS preemptive-resume priority queue, *Queueing Theory and Network Applications*, edited by T. V. Do, Y. Takahashi, W. Yue, and V.-H. Nguyen, pp.3–17, Advances in Intelligent Systems and Computing 383, Springer International Publishing Switzerland 2016.

[10] H. Takagi, Unified and refined analysis of the response time and waiting time in the M/M/m FCFS preemptive-resume priority queue. To appear in the *Journal of Industrial and Management Optimization*, 2016.

[11] A. G. Tatashev, Calculation of the distribution of the waiting time in a multiple-channel queueing system with fixed priorities, *Engineering Cybernetics*, **22** (1984), 59–62. (Originally published in *Tekhnicheskaya Kibernetika*, **6** (1983), 163–166.

[12] H. White and L. S. Christie, Queuing with preemptive priorities or with breakdown. *Opns. Res.*, **6** (1958), 79–95, http://dx.doi.org/10.1287/opre.6.1.79

[13] R. W. Wolff, Work-conserving priorities. *J. Appl. Prob.*, **7** (1970), 327–337. http://dx.doi.org/10.2307/3211968

[14] R. W. Wolff, *Stochastic Modeling and the Theory of Queues*, Prentice Hall, 1989.

[15] S. Zeltyn, Z. Feldman, and S. Wasserkrug, Waiting and sojourn times in a multi-server queue with mixed priorities, *Queueing Systems*, **61** (2009), 305–328, http://dx.doi.org/10.1007/s11134-009-9110-4