UNIFIED AND REFINED ANALYSIS OF THE RESPONSE TIME AND WAITING TIME IN THE M/M/m FCFS PREEMPTIVE-RESUME PRIORITY QUEUE

HIDEAKI TAKAGI
Professor Emeritus, University of Tsukuba
Faculty of Engineering, Information and Systems
1-1-1 Tennoudai, Tsukuba-shi, Ibaraki 305-8573, Japan

Abstract. We present a unified and refined analysis of the response time and waiting time in the M/M/m FCFS preemptive-resume priority queueing system in the steady state by scrutinizing and extending the previous studies such as Brosh (1969), Segal (1970), Buzen and Bondi (1983), Tatashev (1984), and Zeltyn et al. (2009). In particular, we analyze the durations of interleaving waiting times and service times during the response time of a tagged customer of each priority class that is preempted by the arrivals of higher-priority class customers. Our new contribution includes the explicit formulas for the second and third moments of the response time and the third moment of the waiting time. We also clarify the dependence between the waiting time and the total service time. Numerical examples are shown in order to demonstrate the computation of theoretical formulas.

1. Introduction. We consider a queueing system with m servers and an infinite capacity of the waiting room with several priority classes of customers. Customers of class p arrive in a Poisson process with rate λ_p (> 0) independently of customers of all other classes. Every customer requests a service which has the exponential distribution with mean 1/µ irrespective of his class. Classes are indexed 1, 2, . . . such that customers of class p have preemptive priority for service over customers of class q if p < q.

There are three cases which may happen when a customer of class p arrives:
- Unless all servers are busy, his service is started immediately.
- If all servers are busy serving customers of classes equal or higher than p, he must wait at the tail of waiting customers of class p.
- If all servers are busy serving customers, out of whom at least one of them is of class lower than p, let q (> p) be the lowest priority class of those customers being served. At this moment there are at most customers of classes q, q+1, . . . in the waiting room. In this case, the service to one of customers of class q is preempted and he is displaced from the service facility to the head of the

2010 Mathematics Subject Classification. Primary: 60K25; Secondary: 68M28, 90B22.
Key words and phrases. Multiserver priority queue, preemptive-resume, response time, waiting time, first passage time.

The reviewing process of the paper was handled by Wuyi Yue and Yutaka Takahashi as Guest Editors.

The author is supported by the Grant-in-Aid for Scientific Research (C) No. 26330354 from the Japan Society for the Promotion of Science (JSPS) in 2015.

1945
waiting room. We select such a customer of class $q$ for displacement that his service was started or resumed last among all the customers of class $q$ in service. Then the service to the arriving customer of class $p$ is started. This policy of selecting the customer to displace is assumed by Segal [7]. It is called Last-Come, First-Displaced (LCFD) by Fujiki [4].

As soon as a server becomes available, one of the customers of the highest priority class among those in the waiting room is called in for service. Within the same class, a customer is chosen on the first-come, first-served (FCFS) basis. When the service is resumed, a new sample of the service time is set up from the exponential distribution with mean $1/\mu$, irrespective of the amount of service given to him previously.

Thus we may call our system an “M/M/$m$ preemptive-resume priority queue with FCFS and LCFD within the same class.” The study of response time and waiting time of customers in this model dates back to old days including Brosh [1], Segal [7], Buzen and Bondi [2], Tatashev [10], and Zeltyn et al. [13]. The purpose of this paper is to review and refine these works from a unified viewpoint of the first passage time in a birth-and-death process with absorbing states [8]. We derive explicit formulas for the moments of the response time and waiting time of a tagged customer of each priority class in the steady state. We also clarify the dependence between the waiting time and the total service time. In the appendix, we summarize the results for the M/M/2 FCFS preemptive-resume priority queue.

We use the following notation for the analysis throughout this paper:

\[
\rho_p := \frac{\lambda_p}{m\mu}; \quad \lambda^+_p := \sum_{k=1}^{p} \lambda_k; \quad \rho^+_p := \sum_{k=1}^{p} \rho_k = \frac{\lambda^+_p}{m\mu} \quad p = 1, 2, \ldots.
\]

(we let $\lambda^+_0 = \rho^+_0 = 0$).

In the numerical examples in this paper, we assume that there are 4 classes of customers and that

\[
m = 5 \quad ; \quad \mu = 1 \quad ; \quad \lambda_p = \frac{\lambda}{4} \quad (1 \leq p \leq 4).
\]

For this setting of parameter values, we will show several performance measures against $\lambda$ for the range $0 \leq \lambda \leq 20$. Our formulas can be applied to systems with any number of servers, any number of classes, and any different distinct values of arrival rates. However we must assume that the service rates are identical for all customers of all classes and that the system is stable up to customers of class $p$ ($\rho^+_p < 1$).

2. Mean response and waiting times. We first follow Buzen and Bondi [2] for the neat derivation of mean response time $E[T_p]$ for customers of each class $p$.

Let us focus on customers of class $p$. Due to the service and preemption mechanism mentioned above, the behavior of a customer is never affected by customers of lower priority classes as well as customers of the same class who arrive after him. Therefore, we have only to consider customers of classes $1, 2, \ldots, p$.

We denote by $N^+_p$ the number of customers of classes $1, 2, \ldots, p$ present in the system at an arbitrary time in the steady state and define

\[
Q^+_{p,k} := P\{N^+_p = k\} \quad k = 0, 1, 2, \ldots.
\]
From the well-known analysis for the M/M/m queue [3, p. 90] with customers of classes 1, 2, ..., p, we get

\[
Q_{p,k}^+ = \begin{cases} 
Q_{p,0}^+ \frac{(m\rho_p^+)^k}{k!} & 1 \leq k \leq m, \\
Q_{p,m}^+ (m\rho_p^+)^{k-m} & k \geq m + 1, 
\end{cases}
\]  

(1)

where, from the normalization condition \(\sum_{k=0}^{\infty} Q_{p,k}^+ = 1\), we have

\[
\frac{1}{Q_{p,0}^+} = \sum_{k=0}^{m-1} \frac{(m\rho_p^+)^k}{k!} + \frac{(m\rho_p^+)^m}{m!(1 - \rho_p^+)}.
\]

where we assume that \(\rho_p^+ < 1\) for the system to be stable. Then we get

\[
E[N_p^+] = \sum_{k=1}^{\infty} kQ_{p,k}^+ = m\rho_p^+ + \frac{\rho_p^+ C(m, m\rho_p^+)}{1 - \rho_p^+},
\]

where

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

(2)

is Erlang’s C formula [3, p. 91]. In the present case, we have

\[
C(m, m\rho_p^+) = \sum_{k=m}^{\infty} Q_{p,k}^+ = \frac{Q_{p,m}^+}{1 - \rho_p^+} = \frac{Q_{p,0}^+ (m\rho_p^+)^m}{m!}
\]

as the probability that a customer of class \(p\) waits upon arrival.

We denote by \(N_p\) the number of customers of class \(p\) present in the system at an arbitrary time in the steady state. Then we get

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

From Little’s theorem \(E[N_p] = \lambda_p E[T_p]\) for customers of class \(p\), we obtain [2]

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

(3)

We denote by \(L_p\) the number of customers of class \(p\) present in the waiting room at an arbitrary time in the steady state. Then we have

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

which gives the mean waiting time [13]

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

(4)

We plot \(E[T_p]\) and \(E[W_p]\) in Figs. 1 and 2, respectively, for the numerical example described in Section 1.

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

\[
Q_{p,0}^+ = 1 - \rho_p^+ ; \quad C(1, a) = a
\]
so that we obtain

\[ E[T_1] = \frac{1}{\mu (1 - \rho_1)} \quad ; \quad E[T_p] = \frac{1}{\mu (1 - \rho_{p-1}^+)(1 - \rho_p^+)} \quad p \geq 2, \quad (5) \]

where \( \rho_p^+ = \lambda_p^+ / \mu \). This result can also be obtained from the observation in the FCFS system (White and Christie [12]) that the mean response time \( E[T_p] \) of a tagged customer of class \( p \) is composed of his own mean service time, the mean time to serve those customers of classes 1, 2, \ldots, \( p \) who are already present in the waiting room, and the mean workload brought by the subsequent arrivals of customers of classes 1, 2, \ldots, \( p - 1 \) during \( E[T_p] \). Thus we have the relation

\[ E[T_p] = \frac{1}{\mu} + \frac{1}{\mu} \sum_{k=1}^{p} E[N_k] + \sum_{k=1}^{p-1} \frac{\lambda_k}{\mu} E[T_p]. \]

Since we have \( E[N_k] = \lambda_k E[T_k] \) by Little’s theorem for customers of class \( k \) (1 \( \leq k \leq p \)), we get the recursion with respect to \( p \):

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

which yields Eq. (5). We note that this argument holds only for the case of \( m = 1 \).

3. First passage time to service completion. We now focus on a tagged customer of class \( p \) in state \( k \), which means that there are \( k \) other customers of classes
1, 2, . . . , \( p \) who compete for the servers with him at an arbitrary time in the steady state. Then the state transition diagram for our tagged customer is shown in Fig. 3. We consider the first passage time in this one-dimensional birth-and-death process with an absorbing state “service completion” denoted by “Sr”. See, for example, Kulkarni [6, Chapter 4] and Taylor and Karlin [11, Chapter III] for the general treatment of first passage times in discrete-time Markov chains with absorbing states.

We follow Segal [7] and Zeltyn et al. [13] for the following arguments. We denote by \( T_{p,k}^*(s) \) the Laplace-Stieltjes transform (LST) of the distribution function (DF) for the response time of a tagged customer of class \( p \) in state \( k \), where \( k = 0, 1, 2, . . . \). Referring to Fig. 3, we get the following set of equations for \( \{T_{p,k}^*(s); k \geq 0\} \):

\[
T_{p,0}^*(s) = B_{p,0}^*(s)[\beta_{p,0} + (1 - \beta_{p,0})T_{p,1}^*(s)],
\]

\[
T_{p,k}^*(s) = B_{p,k}^*(s)[\beta_{p,k} + (1 - \alpha_{p,k} - \beta_{p,k})T_{p,k+1}^*(s) + \alpha_{p,k}T_{p,k-1}^*(s)]
\]

\[
1 \leq k \leq m - 1,
\]

\[
T_{p,k}^*(s) = B_{p,k}^*(s)[(1 - \alpha'_{p})T_{p,k+1}^*(s) + \alpha'_{p}T_{p,k-1}^*(s)]
\]

\[ k \geq m, \]

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

\[
\alpha_{p,k} = \frac{k\mu}{\lambda_{p-1}^+ + (k + 1)\mu} \quad ; \quad \beta_{p,k} = \frac{\mu}{\lambda_{p-1}^+ + (k + 1)\mu}
\]

The LST of the DF for the time that the tagged customer of class \( p \) stays in state \( k \) is given by

\[
B_{p,k}^*(s) = \frac{\lambda_{p-1}^+ + (k + 1)\mu}{s + \lambda_{p-1}^+ + (k + 1)\mu} \quad 0 \leq k \leq m - 1.
\]

For \( k \geq m \), we use

\[
\alpha'_{p} = \frac{m\mu}{\lambda_{p-1}^+ + \mu} \quad ; \quad B_{p}^*(s) = \frac{\lambda_{p-1}^+ + m\mu}{s + \lambda_{p-1}^+ + m\mu},
\]
which do not depend on state $k$. Then the above set of equations can be written as

$$ (s + \lambda^+_{p-1} + \mu)T^*_{p,0}(s) = \mu + \lambda^+_{p-1}T^*_{p,1}(s), $$

$$ [s + \lambda^+_{p-1} + (k+1)\mu]T^*_{p,k}(s) = \mu + \lambda^+_{p-1}T^*_{p,k+1}(s) + k\mu T^*_{p,k-1}(s) \quad 1 \leq k \leq m-1, $$

$$ (s + \lambda^+_{p-1} + m\mu)T^*_{p,k}(s) = \lambda^+_{p-1}T^*_{p,k+1}(s) + m\mu T^*_{p,k-1}(s) \quad k \geq m. \tag{6} $$

From the homogeneous set of equations for $k \geq m$ in Eq. (6), we assume the solution in the form

$$ T^*_{p,k}(s) = [G^*_{p-1}(s)]^{k-m}T^*_{p,m}(s) \quad k \geq m, \tag{7} $$

where $G^*_{p-1}(s)$ is the solution to the quadratic equation

$$ \lambda^+_{p-1}[G^*_{p-1}(s)]^2 - (s + \lambda^+_{p-1} + m\mu)G^*_{p-1}(s) + m\mu = 0, $$

namely

$$ G^*_{p-1}(s) = \frac{s + \lambda^+_{p-1} + m\mu - \sqrt{(s + \lambda^+_{p-1} + m\mu)^2 - 4\lambda^+_{p-1}m\mu}}{2\lambda^+_{p-1}}. \tag{8} $$

We note that $G^*_{p-1}(s)$ is the LST of the DF for the length of a busy period in the $\text{M}/\text{M}/1$ queue with arrival rate $\lambda^+_{p-1}$ and service rate $m\mu$, which we denote by $G^+_{p-1}$. Then we have

$$ E[G^+_{p-1}] = \frac{1}{m\mu(1 - \rho^+_{p-1})}; \quad E[(G^+_{p-1})^2] = \frac{2}{(m\mu)^2(1 - \rho^+_{p-1})^3}, $$

$$ E[(G^+_{p-1})^3] = \frac{6(1 + \rho^+_{p-1})}{(m\mu)^3(1 - \rho^+_{p-1})^5}. $$

Substituting Eq. (7) into Eq. (6) with $k = m$, we get [7]

$$ T^*_{p,m}(s) = G^*_{p-1}(s)T^*_{p,m-1}(s). \tag{9} $$

We then have the non-homogeneous set of $m$ equations for $0 \leq k \leq m-1$ in Eq. (6), which makes it possible to express both $T^*_{p,m-1}(s)$ and $T^*_{p,m}(s)$ in terms of $T^*_{p,0}(s)$. Substituting them into Eq. (9), we can determine $T^*_{p,0}(s)$. Then we obtain all $\{T^*_{p,k}(s); k \geq 0\}$ in principle. Thus we get

$$ T^*_p(s) = \sum_{k=0}^{\infty} Q^+_p T^*_{p,k}(s) = \sum_{k=0}^{m-1} Q^+_p T^*_{p,k}(s) + \frac{Q^+_p}{1 - \rho^+_p} T^*_{p,m}(s), \tag{10} $$

where we define

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

$$ = \frac{(1 - \rho^+_p) T^*_{p,m}(s)}{1 - \rho^+_p G^*_{p-1}(s)}. \tag{11} $$
In particular, for customers of the highest priority class $p = 1$ who are never preempted ($\lambda_0^+ = 0$), we can get the simple solution to Eq. (6) as

$$T_{1,k}^*(s) = \begin{cases} \frac{\mu}{s + \mu} & 0 \leq k \leq m - 1, \\ \frac{\mu}{s + \mu}[G_0^*(s)]^{k+m} & k \geq m, \end{cases} \quad G_0^*(s) = \frac{m\mu}{s + m\mu},$$

which leads to

$$T_1^*(s) = W_1^*(s)\frac{\mu}{s + \mu},$$

where

$$W_1^*(s) = Q_{1,0}^+ \left[ \sum_{k=0}^{m-1} \frac{(m\rho_1)^k}{k!} + \frac{(m\rho_1)^m}{m!} \frac{s + m\mu - \lambda_1}{s + m\mu} \right]$$

$$= 1 - C(m, m\rho_1)(m\mu - \lambda_1)$$

is the well-known LST of the DF for the waiting time in the M/M/m FCFS queue (without priorities) with arrival rate $\lambda_1$ [3, p. 97].

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

$$Q_{p,k}^+ = (1 - \rho_p^+)(\rho_p^+)^k, \quad T_{p,k}^*(s) = [G_{p-1}^*(s)]^{k+1} \quad k \geq 0,$$

where $\rho_p^+ := \lambda_p^+ / \mu$ and then get [12]

$$T_p^* = (1 - \rho_p^+) \sum_{k=0}^{\infty} (\rho_p^+)^k [G_{p-1}^*(s)]^{k+1}$$

$$= \left( 1 - \rho_p^+ \right) G_{p-1}^*(s) \left( 1 - \rho_p^+ / G_{p-1}^*(s) - \rho_p^+ \right)$$

$$= \frac{2\mu(1 - \rho_p^+)}{s + \lambda_{p-1}^+ - 2\lambda_p^+ + \mu + \sqrt{(s + \lambda_{p-1}^+ + \mu)^2 - 4\lambda_{p-1}^+ \mu}}. \quad (14)$$

We then have the mean $E[T_p]$ in Eq. (5) and the second and third moments

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

$$E[T_p^3] = \frac{6[1 + \rho_p^+ - 4\rho_p^+ \mu + \rho_p^+ \mu^2 + (\rho_p^+ \mu^2)^2]}{\mu^3(1 - \rho_p^+)^5(1 - \rho_p^+)^3}. \quad (15)$$

4. **Mean response time.** Let us consider the mean response time in the framework of the first passage time. The mean response time for a customer of class $p$ who competes for the servers with $k$ others is given by $E[T_{p,k}] = -dT_{p,k}^*(s)/ds|_{s=0}, k \geq 0$. The complete set of $(m + 1)$ equations for the unknowns $\{E[T_{p,k}]; 0 \leq k \leq m\}$ is
derived from Eqs. (6) and (7) as
\[
-1 + (\lambda_{p-1}^+ + \mu)E[T_{p,o}] = \lambda_{p-1}^+ E[T_{p,1}],
\]
\[
-1 + [\lambda_{p-1}^+ + (k + 1)\mu]E[T_{p,k}] = \lambda_{p-1}^+ E[T_{p,k+1}] + k\mu E[T_{p,k-1}] \quad 1 \leq k \leq m - 1,
\]
\[
E[T_{p,m}] = E[T_{p,m-1}] + E[T_{p,m-1}].
\]
We can obtain \(\{E[T_{p,k}]; 0 \leq k \leq m\}\) explicitly as follows. First, by recursion we have
\[
E[T_{p,k}] - \xi_{k+1}E[T_{p,k+1}] = \eta_k(E[T_{p,k-1}] - \xi_k E[T_{p,k}]) + \frac{1}{(k+1)\mu}
\]
\[
= \eta_k \eta_{k-1} (E[T_{p,k-2}] - \xi_{k-1} E[T_{p,k-1}]) + \frac{2}{(k+1)\mu}
\]
\[
= \ldots
\]
\[
= \eta_k \eta_{k-1} \ldots \eta_1 (E[T_{p,0}] - \xi_1 E[T_{p,1}]) + \frac{k}{(k+1)\mu}
\]
\[
= \frac{1}{\mu} - \xi_{k+1} E[T_{p,0}] \quad 0 \leq k \leq m - 1, \quad (15)
\]
where we have defined
\[
\xi_k := \frac{\lambda_{p-1}^+}{k\mu} ; \quad \eta_k := \frac{k}{k+1} \quad 1 \leq k \leq m. \quad (16)
\]
By noting \(\xi_m = \rho_{p-1}^+\) in Eq. (15) with \(k = m - 1\), we get the relation
\[
E[T_{p,m-1}] - \rho_{p-1}^+ E[T_{p,m}] = \frac{1}{\mu} - \rho_{p-1}^+ E[T_{p,0}].
\]
From Eq. (9) we get another relation
\[
E[T_{p,m}] - E[T_{p,m-1}] = \frac{1}{m\mu(1 - \rho_{p-1}^+)}.
\]
Thus we obtain the expression for \(E[T_{p,m}]\) in terms of \(E[T_{p,0}]\)
\[
E[T_{p,m}] = \frac{1}{m\mu(1 - \rho_{p-1}^+)^2} + \frac{1}{\mu(1 - \rho_{p-1}^+)} - \frac{\rho_{p-1}^+}{1 - \rho_{p-1}^+} E[T_{p,0}], \quad (17)
\]
Now, from Eq. (15), we have
\[
E[T_{p,k}] = \frac{E[T_{p,k-1}]}{\xi_k} + E[T_{p,0}] - \frac{1}{\xi_k \mu}
\]
\[
= \frac{E[T_{p,k-2}]}{\xi_k \xi_{k-1}} + E[T_{p,0}] \left(1 + \frac{1}{\xi_k}\right) - \frac{1}{\mu} \left(1 + \frac{1}{\xi_k} \xi_{k-1} + \frac{1}{\xi_k}\right)
\]
\[
= \ldots
\]
\[
= E[T_{p,0}] \left(1 + \sum_{j=1}^{k} \prod_{l=1}^{j} \xi_l \right) - \frac{1}{\mu} \left(1 + \sum_{j=1}^{k-1} \prod_{l=1}^{j} \xi_l \right) \prod_{j=1}^{k} \xi_j ,
\]
which can be written as
\[
E[T_{p,k}] \frac{(m \rho_{p-1}^+)^k}{k!} = E[T_{p,0}] \sum_{j=0}^{k} \frac{(m \rho_{p-1}^+)^j}{j!} - \frac{1}{\mu} \sum_{j=0}^{k-1} \frac{(m \rho_{p-1}^+)^j}{j!} \quad 1 \leq k \leq m.
\]
From this expression with \( k = m \) and Eq. (17), we determine
\[
E[T_{p,0}] = \frac{1}{\mu} + \frac{(m\rho_{p-1}^+)^m}{m!(1 - \rho_{p-1}^+)}
\]
and get
\[
E[T_{p,k}] = \frac{1}{\mu} + \frac{k!(m\rho_{p-1}^+)^{m-k}}{m!(1 - \rho_{p-1}^+)} \sum_{j=0}^{k} \frac{(m\rho_{p-1}^+)^j}{j!} 
\]
\[
0 \leq k \leq m. \quad (18)
\]
We also have
\[
E[T_{p,k}] = E[T_{p,m}] + (k - m)E[\xi_{p-1}^+] \quad k \geq m.
\]
The mean response time for a customer of class \( p \) is then given by [13]
\[
E[T_p] = \sum_{k=0}^{\infty} Q_{p,k}E[T_{p,k}] = \sum_{k=0}^{m-1} Q_{p,k}E[T_{p,k}] + \frac{Q_{p,m}}{1 - \rho_p^+}E[F_{p,m}],
\]
where
\[
E[F_{p,m}] = -\frac{dF_{p,m}(s)}{ds} \bigg|_{s=0} = E[T_{p,m}] + \frac{\rho_p^+E[\xi_{p-1}^+]}{1 - \rho_p^+}.
\]
We have numerically confirmed that this yields the same result as Eq. (3).

5. Second and higher moments of the response time. The \( \ell \)th moment of the response time for a customer of class \( p \) who competes for the servers with \( k \) others is given by \( E[T_{p,k}^\ell] = (-1)\ell d^\ell T_{p,k}^* (s)/ds^\ell \big|_{s=0}, \ k \geq 0, \ \text{for} \ \ell = 2,3,\ldots \). For the \( \ell \)th moment, the complete set of \( (m+1) \) equations for the unknowns \( \{E[T_{p,k}^\ell]; 0 \leq k \leq m\} \) is derived from Eqs. (6) and (7) in terms of the moments of lower degrees as follows:
\[
-\ell E[T_{p,0}^{\ell-1}] + (\lambda_{p-1}^+ + \mu)E[T_{p,0}^\ell] = \lambda_{p-1}^+ E[T_{p,1}^\ell],
\]
\[
-\ell E[T_{p,k}^{\ell-1}] + [\lambda_{p-1}^+ + (k+1)\mu]E[T_{p,k}^\ell] = \lambda_{p-1}^+ E[T_{p,k+1}^\ell] + k\mu E[T_{p,k-1}^\ell] \quad 1 \leq k \leq m - 1,
\]
\[
E[T_{p,m}^\ell] - E[T_{p,m-1}^\ell] = \sum_{l=1}^{\ell} \binom{\ell}{l} E[\xi_{p-1}^+]^l E[T_{p,m-l}^{\ell-l}].
\]
We can obtain \( \{E[T_{p,k}^\ell]; 0 \leq k \leq m\} \) successively with respect to \( \ell \) in the same way as in the preceding section. First, by recursion we have
\[
E[T_{p,k}^\ell] - \xi_{k+1}E[T_{p,k+1}^\ell] = \frac{1}{(k+1)\mu} \left( \ell \sum_{j=0}^{k} E[T_{p,j}^{\ell-1}] - \lambda_{p-1}^+ E[T_{p,0}^\ell] \right) \quad 0 \leq k \leq m - 1. \quad (19)
\]
The case $k = m - 1$ in this equation yields

$$E[T^\ell_{p,m-1}] - \rho^+_{p-1} E[T^\ell_{p,m}] = \frac{\ell}{m \mu} \sum_{j=0}^{m-1} E[T^\ell_{p,j}] - \rho^+_{p-1} E[T^\ell_{p,0}].$$

Thus we get

$$(1 - \rho^+_{p-1}) E[T^\ell_{p,m}] = \frac{\ell}{m \mu} \sum_{j=0}^{m-1} E[T^\ell_{p,j}] - \rho^+_{p-1} E[T^\ell_{p,0}] + \sum_{l=1}^{\ell} \left( \binom{\ell}{l} \right) E[(G^+_{p-1})^l] E[T^\ell_{p,m-1}].$$

By further recursion of Eq. (19) we have

$$E[T^\ell_{p,k}] \frac{(m \rho^+_{p-1})^k}{k!} = E[T^\ell_{p,0}] \sum_{j=0}^{k} \frac{(m \rho^+_{p-1})^j}{j!} - \frac{\ell}{\lambda^+_{p-1}} \sum_{j=1}^{k} \frac{(m \rho^+_{p-1})^j}{j!} \sum_{l=0}^{j-1} E[T^\ell_{p,l}]$$

$$1 \leq k \leq m. \tag{20}$$

From Eq. (20) and Eq. (21) with $k = m$, we determine

$$E[T^\ell_{p,0}] \frac{m \rho^+_{p-1}}{\ell} = \frac{\sum_{j=0}^{m-1} E[T^\ell_{p,j}] \left[ \frac{(m \rho^+_{p-1})^m}{m!} + (1 - \rho^+_{p-1}) \sum_{l=1}^{m-1} \frac{(m \rho^+_{p-1})^l}{l!} \right]}{\sum_{j=0}^{m-1} \frac{(m \rho^+_{p-1})^j}{j!} \sum_{l=1}^{j-1} \binom{\ell}{l} E[(G^+_{p-1})^l] E[T^\ell_{p,m-1}]} \lambda^+_{p-1} \left[ \frac{(m \rho^+_{p-1})^m}{m!} + (1 - \rho^+_{p-1}) \sum_{j=0}^{m-1} \frac{(m \rho^+_{p-1})^j}{j!} \right].$$

Thus we can calculate \{ $E[T^\ell_{p,k}]$; $1 \leq k \leq m$ \} by Eq. (21). In particular we have

$$E[T^\ell_{p,m}] \frac{m \rho^+_{p-1}}{\ell} = \frac{\sum_{j=0}^{m-1} E[T^\ell_{p,j}] \sum_{l=0}^{m} \frac{(m \rho^+_{p-1})^l}{l!}}{m \mu} \left[ \frac{(m \rho^+_{p-1})^m}{m!} + (1 - \rho^+_{p-1}) \sum_{j=0}^{m-1} \frac{(m \rho^+_{p-1})^j}{j!} \right].$$

Finally, the $\ell$th moment of the response time for a customer of class $p$ is

$$E[T^\ell_p] = \sum_{k=0}^{\infty} Q^+_{p,k} E[T^\ell_{p,k}] = \sum_{k=0}^{m-1} Q^+_{p,k} E[T^\ell_{p,k}] + \frac{Q^+_{p,m}}{1 - \rho^+_p} E[F^\ell_{p,m}], \tag{22}$$

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

$$E[F^\ell_{p,m}] = E[T^\ell_{p,m}] + \frac{\rho^+_p}{1 - \rho^+_p} \sum_{l=1}^{\ell} \binom{\ell}{l} E[F^\ell_{p,m-l}] E[(G^+_{p-1})^l], \quad \ell = 2, 3, \ldots.$$
The second moment of the response time for a customer of class \(p\) is given by \[13\]

\[
E[T_p^2] = \sum_{k=0}^{m-1} Q_{p,k}^+ E[T_{p,k}^2] + Q_{p,m}^+ \left\{ \frac{E[T_{p,m}^2]}{1 - \rho_p^+} + \frac{\rho_p^+ \{ E[(G_{p-1}^+)^2] + 2E[T_{p,m}]E[G_{p-1}^+] \}}{(1 - \rho_p^+)^2} \right. \\
+ \left. \frac{2(\rho_p^+ E[G_{p-1}^+])^2}{(1 - \rho_p^+)^3} \right\} .
\] (23)

The third moment is given by

\[
E[T_p^3] = \sum_{k=0}^{m-1} Q_{p,k}^+ E[T_{p,k}^3] + Q_{p,m}^+ \left\{ \frac{E[T_{p,m}^3]}{1 - \rho_p^+} + \frac{6(\rho_p^+ E[G_{p-1}^+])^3}{(1 - \rho_p^+)^4} \right. \\
+ \left. \rho_p^+ \{ E[(G_{p-1}^+)^3] + 3E[T_{p,m}]E[(G_{p-1}^+)^2] + 3E[T_{p,m}^2]E[G_{p-1}^+] \} \right. \\
+ \left. \frac{6(\rho_p^+)^2 \{ E[(G_{p-1}^+)^2]E[G_{p-1}^+] + E[T_{p,m}](E[G_{p-1}^+])^2 \}}{(1 - \rho_p^+)^3} \right\} .
\] (24)

We plot \(E[T_p^2]\) and \(E[T_p^3]\) in Figs. 4 and 5, respectively, for the numerical example described in Section 1.

![Figure 4](image1.png)

**Figure 4.** Second moment of the response time for a customer of class \(p\)

![Figure 5](image2.png)

**Figure 5.** Third moment of the response time for a customer of class \(p\)**
6. **Initial waiting time.** Let us call the time interval from the arrival of a customer of class $p$ to the instant at which he enters service for the first time his *initial waiting time*, and denote it by $W^\circ_{p,k}$. Brosh [1] derived the explicit expression for the mean initial waiting time $E[W^\circ_{p,k}]$. But we can obtain the LST of the DF for $W^\circ_{p,k}$ of a customer of class $p$ who competes for the servers with $k$ other customers of class $1, 2, \ldots, p$. By definition, we have

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

For $k \geq m$, we have the following set of equations:

$$W^\circ_{p,k}(s) = B^\circ_{p,k}(s) \big[ (1 - \alpha_p') W^\circ_{p,k+1}(s) + \alpha_p' W^\circ_{p,k-1}(s) \big] \quad k \geq m,$$

where $B^\circ_{p,k}(s)$ and $\alpha_p'$ are defined in Section 3. Thus we have

$$(s + \lambda_{p-1} + m\mu) W^\circ_{p,k}(s) = \lambda_{p-1} W^\circ_{p,k+1}(s) + m\mu W^\circ_{p,k-1}(s) \quad k \geq m.$$ 

These relations are satisfied by

$$W^\circ_{p,k}(s) = [G^\circ_{p-1}(s)]^{k-m+1} \quad k \geq m - 1,$$

where $G^\circ_{p-1}(s)$ is given in Eq. (8). Then we get

$$W^\circ_{p,k}(s) = \sum_{k=0}^{\infty} Q^\circ_{p,k} W^\circ_{p,k}(s) = \sum_{k=0}^{m-1} Q^\circ_{p,0} \frac{(\mu\rho_p^+)^k}{k!} + \sum_{k=m}^{\infty} Q^\circ_{p,m} (\rho_p^+)^{k-m} [G^\circ_{p-1}(s)]^{k-m+1} \quad k \geq m.$$ 

Thus we have

$$W^\circ_{p,k}(s) = Q^\circ_{p,0} \frac{(\mu\rho_p^+)^k}{k!} + \frac{Q^\circ_{p,m} G^\circ_{p-1}(s)}{1 - \rho_p^+ G^\circ_{p-1}(s)} = 1 - C(m, m\rho_p^+) + C(m, m\rho_p^+) \frac{(1 - \rho_p^+)(G^\circ_{p-1}(s))}{1 - \rho_p^+ G^\circ_{p-1}(s)} = 1 - \frac{C(m, m\rho_p^+)[1 - G^\circ_{p-1}(s)]}{1 - \rho_p^+ G^\circ_{p-1}(s)},$$

where $C(m, a)$ is given in Eq. (2). This leads to the Brosh’s result [1]:

$$E[W^\circ_{p,k}] = \frac{C(m, m\rho_p^+) \mu(1 - \rho_p^+)}{(1 - \rho_p^+)^3}.$$ 

We can also get the second and third moments of $W^\circ_{p,k}$ as

$$E[(W^\circ_{p,k})^2] = \frac{2C(m, m\rho_p^+)(1 - \rho_p^+)}{(\mu\rho_p^+)^2 (1 - \rho_p^+)^3 (1 - \rho_p^+)^2},$$

$$E[(W^\circ_{p,k})^3] = \frac{6C(m, m\rho_p^+)[1 + \rho_p^+ - 4\rho_p^+\rho_p^+ + \rho_p^+ (\rho_p^+)^2 + (\rho_p^+)^2]}{(\mu\rho_p^+)^3 (1 - \rho_p^+)^3 (1 - \rho_p^+)^3}.$$

We note that

$$W^\circ_{p,k}(s) = \frac{(1 - \rho_p^+ G^\circ_{p-1}(s))}{1 - \rho_p^+ G^\circ_{p-1}(s)} = \frac{m\mu(1 - \rho_p^+)[1 - G^\circ_{p-1}(s)]}{s - \lambda_p + \lambda_p G^\circ_{p-1}(s)} \quad (26)$$
is the LST of the DF for the initial waiting time of a customer of class \( p \) who arrives when all servers are busy serving customers of classes 1, 2, \ldots, \( p \). For \( m = 1 \) (a single-server queue), we observe that \( W_p^*(s) \equiv T_p^*(s) \) given in Eq. (14).

Gaver [5] defined the completion time for a customer of class \( p \) as the time interval from the instant at which he enters service (clearly, there are no customers of classes 1, 2, \ldots, \( p - 1 \) in the system at this moment) to the first instant after his service completion at which there are no customers of classes 1, 2, \ldots, \( p - 1 \) in the system. In our framework, the LST of the DF for the completion time of a customer of class \( p \) who competes for the servers with \( k \) other customers is given by \( T_{p,k}^*(s) \) for \( k \geq m \).

Since the initial waiting time and completion time are independent, we have the relation

\[
T_{p,k}^*(s) = W_{p,k}^*(s)T_{p,m-1}^*(s) = [G_{p-1}^*(s)]^{k-m+1}T_{p,m-1}^*(s) \quad k \geq m,
\]

which agrees with Eqs. (7) and (9).

7. **Waiting time.** After a customer of class \( p \) enters service for the first time, his service may be preempted several times before completion when he is pushed out of the service facility by the arrivals of customers of classes 1, 2, \ldots, \( p - 1 \). He stays in the waiting room until he again enters service. The total amount of the time a customer spends in the waiting room is called the waiting time, which is the response time minus the service time.

Tatashev [10] derived the LST of the DF for the waiting time \( W_p \) for customers of class \( p \). Later Zeltyn et al. [13] show the mean and the second moment of \( W_p \). Their analysis and result are reviewed in this section.

Let \( P_{p,k}\{Pr\} \) be the probability that a tagged customer of class \( p \) competing for the servers with \( k \) other customers (they are all customers of classes 1, 2, \ldots, \( p - 1 \) and those customers of class \( p \) who have arrived before the tagged customer) is preempted, where \( k = 0, 1, 2, \ldots, m - 1 \). The state transition diagram for our tagged customer is shown in Fig. 6. We consider the first passage time in this one-dimensional birth-and-death process with two absorbing states, namely “service preemption” denoted by “Pr” and “service completion” denoted by “Sr”.

![Figure 6. State transition diagram for a customer of class p until service preemption or completion.](image)

\[
\begin{align*}
\beta_{p,m-1} & \quad \beta_{p,k+1} & \beta_{p,k} & \beta_{p,k-1} & \beta_{p,1} & \beta_{p,0} \\
\alpha_{p,k+1} & \quad \alpha_{p,k} & \alpha_{p,k-1} & \alpha_{p,1} & \alpha_{p,0} \\
1 - \alpha_{p,k} - \beta_{p,k} & \quad 1 - \alpha_{p,k-1} & 1 - \alpha_{p,k} & 1 - \alpha_{p,0} \\
1 - \alpha_{p,m-1} - \beta_{p,m-1} & \quad 1 - \alpha_{p,k} & 1 - \beta_{p,k-1} & 1 - \beta_{p,0}
\end{align*}
\]
Referring to Fig. 6, we have the complete set of \( m \) equations for \( \{ P_{p,k} \} ; 0 \leq k \leq m - 1 \) as follows:

\[
P_{p,0} = (1 - \beta_{p,0}) P_{p,1},
\]

\[
P_{p,k} = (1 - \alpha_{p,k} - \beta_{p,k}) P_{p,k+1} + \alpha_{p,k} P_{p,k-1} \quad 1 \leq k \leq m - 2,
\]

\[
P_{p,m-1} = 1 - \alpha_{p,m-1} - \beta_{p,m-1} + \alpha_{p,m-1} P_{p,m-2},
\]

where \( \alpha_{p,k} \) and \( \beta_{p,k} \) (\( 0 \leq k \leq m - 1 \)) are given in Section 3. The solution is found as follows. By recursion we get

\[
P_{p,k} \eta_{k+1} P_{p,k+1} = \eta_k (P_{p,k-1} - \xi_k P_{p,k})
\]

\[= \ldots \]

\[= \eta_k \eta_{k-1} \cdots \eta_1 (P_{p,0} - \xi_1 P_{p,1})
\]

\[= -\xi_{k+1} P_{p,0} \quad 0 \leq k \leq m - 1,
\]

where \( \xi_k \) and \( \eta_k \) are defined in Eq. (16). Then we get

\[
P_{p,k} = \frac{P_{p,k-1} \eta_k}{P_{p,0}} + P_{p,0} = \ldots = P_{p,0} \left( 1 + \sum_{j=1}^{k} \frac{\prod_{l=1}^{j} \xi_l}{\prod_{j=1}^{k} \xi_j} \right) \frac{\left( \sum_{j=0}^{k} (m \rho_{p-1}^+) \right)^{j} \eta^j}{\prod_{j=1}^{k} \xi_j}
\]

\[= \frac{P_{p,0}}{\xi_k} \left( \sum_{j=0}^{k} \frac{(m \rho_{p-1}^+)^j}{j!} \right) \frac{\left( \sum_{j=0}^{k} (m \rho_{p-1}^+)^j \right)^{j}}{\prod_{j=1}^{k} \xi_j} \quad 0 \leq k \leq m - 1.
\]

From the convention \( P_{p,m} \{ \Pr \} = 1 \), we determine

\[
P_{p,0} \{ \Pr \} = \frac{(m \rho_{p-1}^+)^m}{m!} \left( \sum_{j=0}^{m} \frac{(m \rho_{p-1}^+)^j}{j!} \right) = B(m, m \rho_{p-1}^+),
\]

where

\[B(m, a) := \frac{a^m}{m!} \left( \sum_{k=0}^{m} \frac{a^k}{k!} \right)
\]

is the well-known Erlang’s \( B \) formula [3, p. 80]. Thus we obtain

\[
P_{p,k} \{ \Pr \} = \frac{B(m, m \rho_{p-1}^+)}{B(k, m \rho_{p-1}^+)} \quad 0 \leq k \leq m - 1.
\]

(27)

We note that

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

(28)

is the probability that the service for a customer of class \( p \) started after waiting is preempted (note that \( r_1 = 0 \)). The probability \( q_p \) that the service for a customer of class \( p \) started without waiting is preempted can be found as follows. When his service is started there are \( k \) other customers of classes \( 1, 2, \ldots, p \) with probability

\[
\frac{(m \rho_{p}^+)^k}{k!} \left( \sum_{j=0}^{m-1} \frac{(m \rho_{p}^+)^j}{j!} \right) \quad 0 \leq k \leq m - 1.
\]
Then his service is preempted with probability \( P_{p,k}\{Pr\} \). Therefore we get

\[
q_p = \sum_{k=0}^{m-1} \frac{B(m, m\rho^+_p) - B(m, m\rho^+_p)}{B(k, m\rho^+_p)} \frac{(m\rho^+_p)^k}{k!}.
\]

\[
= \frac{B(m, m\rho^+_p) - B(m, m\rho^+_p)}{\sum_{j=0}^{m-1} \frac{(m\rho^+_p)^j}{j!}} \sum_{k=0}^{m-1} \frac{(m\rho^+_p)^k}{k!} \sum_{j=0}^{k} \frac{(m\rho^+_p)^j}{j!}.
\]

Here we have

\[
\sum_{j=0}^{m-1} \frac{(m\rho^+_p)^j}{j!} = 1 - \frac{B(m, m\rho^+_p)}{B(m, m\rho^+_p)} \frac{(m\rho^+_p)^m}{m!}
\]

and

\[
\sum_{k=0}^{m-1} \frac{(m\rho^+_p)^k}{k!} \sum_{j=0}^{k} \frac{(m\rho^+_p)^j}{j!} = \frac{B(m, m\rho^+_p) - B(m, m\rho^+_p)}{\sum_{j=0}^{m-1} \frac{(m\rho^+_p)^j}{j!}} \sum_{j=0}^{m-1} \frac{(m\rho^+_p)^j}{j!}.
\]

Thus we can write [13] (note that \( q_1 = 0 \))

\[
q_p = \frac{\rho^+_{p-1}}{\rho_p} \frac{B(m, m\rho^+_p) - B(m, m\rho^+_p)}{B(m, m\rho^+_p)} \frac{(m\rho^+_p)^m}{m!}.
\] (29)

Similarly, let \( P_{p,k}\{Sr\} \) be the probability that a tagged customer of class \( p \) competing for the servers with \( k \) other customers is completed without preemption, where \( k = 0, 1, 2, \ldots, m - 1 \). Referring to Fig. 6, we have the complete set of equations for \( \{P_{p,k}\{Sr\}; 0 \leq k \leq m - 1\} \) as follows:

\[
P_{p,0}\{Sr\} = \beta_{p,0} + (1 - \beta_{p,0})P_{p,1}\{Sr\},
\]

\[
P_{p,k}\{Sr\} = \beta_{p,k} + (1 - \alpha_{p,k} - \beta_{p,k})P_{p,k+1}\{Sr\} + \alpha_{p,k}P_{p,k-1}\{Sr\}
\]

\[
1 \leq k \leq m - 2,
\]

\[
P_{p,m-1}\{Sr\} = \beta_{p,m-1} + \alpha_{p,m-1}P_{p,m-2}\{Sr\}.
\]

The solution is given by

\[
P_{p,k}\{Sr\} = 1 - P_{p,k}\{Pr\} = 1 - \frac{B(m, m\rho^+_p)}{B(k, m\rho^+_p)} \quad 0 \leq k \leq m - 1.
\] (30)

Then we can numerically confirm the relation

\[
\sum_{k=0}^{m-1} Q^+_p P_{p,k}\{Sr\} = [1 - C(m, m\rho^+_p)](1 - q_p)
\]

as the probability that an arriving customer of class \( p \) is started service immediately upon arrival and his service is not preempted until completion. We can also confirm
the relation
\[
\sum_{k=0}^{m-1} Q_{p,k}^{+} P_{p,k} \{ \Pr \} = [1 - C(m, m\rho_p^+)] q_p
\]
as the probability that an arriving customer of class \( p \) is started service immediately upon arrival and his service is preempted before completion.

Upon arrival of a tagged customer of class \( p \), the following cases occur:

- If less than \( m \) servers are busy for serving customers of classes 1, 2, \ldots, \( p \), his service is started immediately. This case occurs with probability \( 1 - C(m, m\rho_p^+) \).
  - If his service is not preempted, his waiting time is zero. This subcase occurs with probability \( 1 - q_p \).
  - If his service is preempted, he waits \( G_{p-1}^+ \) time units for his service to be resumed. 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 making him wait \( G_{p-1}^+ \) time units.

- If \( m \) servers are busy for 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 \( C(m, m\rho_p^+) \). His service is preempted \( i \) times with probability \( (1 - r_p)(r_p)^i \) \((i = 0, 1, 2, \ldots)\) with each preemption making him wait \( G_{p-1}^+ \) time units. Note that the LST of the DF for \( W_p^+ \) is given in Eq. (26).

Therefore, the LST of the DF for the waiting time of a tagged customer of class \( p \) is given by \([10, 13]\)

\[
W_p^+(s) = [1 - C(m, m\rho_p^+)] \left\{ 1 - q_p + q_p G_{p-1}^+(s) \sum_{i=0}^{\infty} (1 - r_p)(r_p)^i [G_{p-1}^+(s)]^i \right\}
+ C(m, m\rho_p^+) W_p^+(s) \sum_{i=0}^{\infty} (1 - r_p)(r_p)^i [G_{p-1}^+(s)]^i
= [1 - C(m, m\rho_p^+)] \left\{ 1 - q_p + \frac{q_p(1 - r_p)G_{p-1}^+(s)}{1 - r_p G_{p-1}^+(s)} \right\}
+ C(m, m\rho_p^+) \frac{(1 - r_p)W_p^+(s)}{1 - r_p G_{p-1}^+(s)}
\]
\[(31)\]

The service for customers of the highest priority class \( p = 1 \) is never preempted so that we have

\[
W_1^+(s) = W_1^+(s) = 1 - C(m, m\rho_1) + C(m, m\rho_1)W_1^+(s) \quad ; \quad W_1^+(s) = \frac{m\mu - \lambda_1}{s + m\mu - \lambda_1}
\]
as given in Eq. (13).

The mean waiting time for a customer of class \( p \) is given by

\[
E[W_p] = \frac{[1 - C(m, m\rho_p^+)] q_p}{m\mu(1 - r_p)(1 - \rho_p^+)} + \frac{C(m, m\rho_p^+)(1 - r_p\rho_p^+)}{m\mu(1 - r_p)(1 - \rho_{p-1}^+)(1 - \rho_p^+)}.
\]
We have numerically confirmed that this yields the same result as Eq. (4). The second moment of the waiting time for a customer of class \( p \) is given explicitly by

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

This expression yields the same numerical values as those from the following expression derived by Zeltyn et al. [13]:

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

The agreement of Eqs. (32) and (33) can be proved algebraically by using the relation [3, p. 92]

\[
C(m, m\rho_{p-1}^+) = \frac{\rho_{p-1}^+ - r_p}{\rho_{p-1}(1 - r_p)} = \frac{B(m, m\rho_{p-1}^+)}{1 - \rho_{p-1}^+ + \rho_{p-1}B(m, m\rho_{p-1}^+)}.
\]

The third moment of the waiting time for a customer of class \( p \) is given by

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

where

\[
w_1(\rho_{p-1}^+, \rho_p^+) = 3 - \rho_{p-1}^+ - 3\rho_p^+ - 2\rho_{p-1}^+\rho_p^+ + (\rho_p^+)^2 + (\rho_{p-1}^+)^2\rho_p^+ + \rho_{p-1}^+\rho_p^+ + (\rho_p^+)^2, \\
w_2(\rho_{p-1}^+, \rho_p^+) = -3 - 2\rho_{p-1}^+ + \rho_p^+ + (\rho_p^+)^2 + 10\rho_{p-1}^+\rho_p^+ - 3(\rho_{p-1}^+)^2\rho_p^+ - 4\rho_{p-1}^+(\rho_p^+)^2, \\
w_3(\rho_{p-1}^+, \rho_p^+) = 1 + \rho_{p-1}^+ - 4\rho_{p-1}^+\rho_p^+ + \rho_{p-1}^+\rho_p^+ + (\rho_p^+)^2 + (\rho_{p-1}^+)^2\rho_p^+.
\]

We plot \( E[W_p^2] \) and \( E[W_p^3] \) in Figs. 7 and 8, respectively, for the numerical example described in Section 1. In Table 1, we show the difference in the mean and the second and third moments between the initial waiting time \( W_p^\text{I} \) and the waiting time \( W_p \) for a customer of the lowest priority class \( p = 4 \).

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

\[
C(1, \rho_p^+) = \rho_p^+, \quad r_p = q_p = \frac{\rho_{p-1}^+}{1 + \rho_{p-1}^+}.
\]
Then we get

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

(35)
which yields

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

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

\[
\quad \left. - \frac{(\rho_p^+ - 1)^3}{(1 - \rho_p^+)^3} \right],
\]

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

\[
\quad \left. - \frac{(\rho_p^+ - 1)^3[1 - 2(\rho_p^+)^2]}{(1 - \rho_p^+)^3(1 - \rho_p^-)^2} - \frac{(\rho_p^+ - 1)^4(2 - 3 \rho_p^+)}{(1 - \rho_p^-)^5(1 - \rho_p^-)^4} + \frac{(\rho_p^+ - 1)^3}{(1 - \rho_p^-)^5} \right].
\]

8. **Service time.** We are also interested in the total service time that each customer of class \(p\) receives before service completion in the M/M/m FCFS preemptive-resume priority queue. The total service time consists of several partial service times of two types, which we look at separately in the following.

Let \(V_{p,k}(s)\) be the LST of the DF for the time to preemption for a customer of class \(p\) who competes for the servers with \(k\) other customers, where \(0 \leq k \leq m - 1\). Referring to Fig. 6, we have the complete set of \(m\) equations for \(\{V_{p,k}(s); 0 \leq k \leq m - 1\}\) as follows:

\[
(s + \lambda_{p-1}^+ + \mu)V_{p,0}(s) = \lambda_{p-1}^+ V_{p,1}(s),
\]

\[
[s + \lambda_{p-1}^+ + (k+1)\mu]V_{p,k}(s) = \lambda_{p-1}^+ V_{p,k+1}(s) + k\mu V_{p,k-1}(s) \quad 1 \leq k \leq m - 2,
\]

\[
(s + \lambda_{p-1}^+ + m\mu)V_{p,m-1}(s) = \lambda_{p-1}^+ + (m-1)\mu V_{p,m-2}(s).
\]

We can obtain \(\{V_{p,k}(s); 0 \leq k \leq m - 1\}\) algebraically. We note that \(P_{p,k}\{Pr\} = V_{p,k}(0)\) as given in Eq. (27) for \(0 \leq k \leq m - 1\).

We first consider the mean \(E[V_{p,k}] = -dV_{p,k}(s)/ds|_{s=0}\). The complete set of \(m\) equations for \(\{E[V_{p,k}]; 0 \leq k \leq m - 1\}\) is given by

\[
-P_{p,0}\{Pr\} + (\lambda_{p-1}^+ + \mu)E[V_{p,0}] = \lambda_{p-1}^+ E[V_{p,1}],
\]

\[
-P_{p,k}\{Pr\} + [\lambda_{p-1}^+ + (k+1)\mu]E[V_{p,k}] = \lambda_{p-1}^+ E[V_{p,k+1}] + k\mu E[V_{p,k-1}] \quad 1 \leq k \leq m - 2,
\]

\[
-P_{p,m-1}\{Pr\} + (\lambda_{p-1}^+ + m\mu)E[V_{p,m-1}] = (m-1)\mu E[V_{p,m-2}].
\]
By recursion as in Section 4, we get the relation

\[ E[V_{p,k}] - \xi_{k+1} E[V_{p,k+1}] = \eta_k (E[V_{p,k-1}] - \xi_k E[V_{p,k}]) + \frac{P_{p,k} \{ Pr \}}{(k + 1) \mu} \]

\[ = \cdots \]

\[ = \eta_k \eta_{k-1} \cdots \eta_1 (E[V_{p,0}] - \xi_1 E[V_{p,1}]) + \frac{\sum_{j=0}^{k} P_{p,j} \{ Pr \}}{(k + 1) \mu} \]

\[- \xi_{k+1} E[V_{p,0}] + \frac{\sum_{j=0}^{k} P_{p,j} \{ Pr \}}{(k + 1) \mu} \quad 0 \leq k \leq m - 2.\]

Then we get

\[ E[V_{p,k}] = \frac{E[V_{p,k-1}]}{\xi_k} + E[V_{p,0}] - \frac{\sum_{j=0}^{k-1} P_{p,j} \{ Pr \}}{\lambda_{p-1}^+} \]

\[ = \cdots \]

\[ = E[V_{p,0}] \left( 1 + \sum_{j=1}^{k} \prod_{l=1}^{j} \xi_l \right) - \frac{1}{\lambda_{p-1}^+} \sum_{j=1}^{k} \left( \prod_{l=1}^{j} \xi_l \right) \sum_{l=0}^{j-1} P_{p,l} \{ Pr \} \]

\[ \bigg/ \prod_{j=1}^{k} \xi_j , \]

which can be written as

\[ E[V_{p,k}] \frac{(m \rho_{p-1}^+)^k}{k!} = E[V_{p,0}] \sum_{j=0}^{k} \frac{(m \rho_{p-1}^+)^j}{j!} - \frac{1}{\lambda_{p-1}^+} \sum_{j=1}^{k} \frac{(m \rho_{p-1}^+)^j}{j!} \sum_{l=0}^{j-1} P_{p,l} \{ Pr \} \]

\[ 1 \leq k \leq m - 1. \]

From the convention \( E[V_{p,m}] = 0 \), we determine

\[ E[V_{p,0}] = \sum_{j=1}^{m} \frac{(m \rho_{p-1}^+)^j}{j!} \sum_{l=0}^{j-1} P_{p,l} \{ Pr \} / \lambda_{p-1}^+ \sum_{j=0}^{m} \frac{(m \rho_{p-1}^+)^j}{j!} . \]

We then get

\[ E[V_{p,k}] = \left\{ \begin{array}{l}
\sum_{j=k+1}^{m} \frac{(m \rho_{p-1}^+)^j}{j!} \sum_{l=0}^{j-1} P_{p,l} \{ Pr \} \\
- \sum_{j=k+1}^{m} \frac{(m \rho_{p-1}^+)^j}{j!} \sum_{l=0}^{j-1} P_{p,l} \{ Pr \}
\end{array} \right\} \frac{k}{\lambda_{p-1}^+} \sum_{j=0}^{m} \frac{(m \rho_{p-1}^+)^j}{j!} \]

\[ 1 \leq k \leq m - 1, \quad (37) \]

where \( P_{p,l} \{ Pr \} \) is given in Eq. (27). In particular, we have

\[ E[V_{p,m-1}] = \sum_{j=0}^{m-1} \frac{(m \rho_{p-1}^+)^j}{j!} \sum_{l=j}^{m-1} P_{p,l} \{ Pr \} / m_{\mu} \sum_{j=0}^{m} \frac{(m \rho_{p-1}^+)^j}{j!} . \quad (38) \]
We next consider the second moment \( E[V_{p,k}^2] = d^2 V^*_{p,k}(s)/ds^2 \). The complete set of \( m \) equations for \( \{E[V_{p,k}^2]; 0 \leq k \leq m - 1\} \) is given by

\[
-2E[V_{p,0}] + (\lambda_{p-1}^+ + \mu)E[V_{p,0}^2] = \lambda_{p-1}^+ E[V_{p,1}^2],
\]

\[
-2E[V_{p,k}] + [\lambda_{p-1}^+ + (k + 1)\mu]E[V_{p,k}^2] = \lambda_{p-1}^+ E[V_{p,k+1}^2] + k\mu E[V_{p,k-1}^2] \quad 1 \leq k \leq m - 2,
\]

\[
-2E[V_{p,m-1}] + (\lambda_{p-1}^+ + m\mu)E[V_{p,m-1}^2] = (m - 1)\mu E[V_{p,m-2}^2].
\]

We can derive

\[
E[V_{p,k}^2] \frac{(m\rho_{p-1}^+)^k}{k!} = E[V_{p,0}] \sum_{j=0}^{k} \frac{(m\rho_{p-1}^+)^j}{j!} - 2 \frac{\lambda_{p-1}^+}{j+1} \sum_{j=0}^{k} \frac{(m\rho_{p-1}^+)^j}{j!} \sum_{l=0}^{j-1} E[V_{p,l}],
\]

\[1 \leq k \leq m - 1.\]

We then obtain

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

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

and

\[
E[V_{p,k}^2] / 2 = \left\{ \begin{array}{c} \sum_{j=k+1}^{m} \frac{(m\rho_{p-1}^+)^j}{j!} \sum_{l=0}^{j-1} E[V_{p,l}] / \lambda_{p-1}^+ \sum_{j=0}^{m} \frac{(m\rho_{p-1}^+)^j}{j!} \sum_{l=0}^{j-1} E[V_{p,l}] \\ - \frac{k}{j} \sum_{j=0}^{k} \frac{(m\rho_{p-1}^+)^j}{j!} \sum_{l=0}^{j-1} E[V_{p,l}] / m\mu \sum_{j=0}^{m} \frac{(m\rho_{p-1}^+)^j}{j!} \end{array} \right\} \frac{m}{j!} \frac{(m\rho_{p-1}^+)^k}{k!} \sum_{j=0}^{m} \frac{(m\rho_{p-1}^+)^j}{j!},
\]

\[1 \leq k \leq m - 2,\]

where \( E[V_{p,l}] \) is given in Eq. (37).

Let \( U^*_{p,k}(s) \) be the LST of the DF for the time to service completion for a customer of class \( p \) who competes for the servers with \( k \) other customers, where \( 0 \leq k \leq m - 1 \). Referring to Fig. 6, we have the complete set of \( m \) equations for \( \{U^*_{p,k}(s); 0 \leq k \leq m - 1\} \) as follows:

\[
(s + \lambda_{p-1}^+ + \mu)U^*_{p,0}(s) = \mu + \lambda_{p-1}^+ U^*_{p,1}(s),
\]

\[
[s + \lambda_{p-1}^+ + (k + 1)\mu]U^*_{p,k}(s) = \mu + \lambda_{p-1}^+ U^*_{p,k+1}(s) + k\mu U^*_{p,k-1}(s) \quad 1 \leq k \leq m - 2,
\]

\[
(s + \lambda_{p-1}^+ + m\mu)U^*_{p,m-1}(s) = \mu + (m - 1)\mu U^*_{p,m-2}(s).
\]

We note that \( P_{p,k}(Sr) = U^*_{p,k}(0) \) as given in Eq. (30) for \( 0 \leq k \leq m - 1 \).
We first consider the mean $E[U_{p,k}] = -dU_{p,k}^*(s)/ds_{s=0}$. The complete set of $m$ equations for $\{E[U_{p,k}]; 0 \leq k \leq m - 1\}$ is given by

\[-P_{p,0}\{\text{Sr}\} + (\lambda_{p-1}^+ + \mu)E[U_{p,0}] = \lambda_{p-1}^+ E[U_{p,1}],\]

\[-P_{p,k}\{\text{Sr}\} + [\lambda_{p-1}^+ + (k + 1)\mu]E[U_{p,k}] = \lambda_{p-1}^+ E[U_{p,k+1}] + k\mu E[U_{p,k-1}]\]

\[1 \leq k \leq m - 2,\]

\[-P_{p,m-1}\{\text{Sr}\} + (\lambda_{p-1}^+ + m\mu)E[U_{p,m-1}] = (m - 1)\mu E[U_{p,m-2}].\]

This set has the same structure as the set for $\{E[V_{p,k}]; 0 \leq k \leq m - 1\}$ with $P_{p,k}\{\text{Pr}\}$ replaced by $P_{p,k}\{\text{Sr}\}$. Therefore we have

\[E[U_{p,0}] = \sum_{j=1}^{m} \frac{(m\rho_{p-1}^+)^j}{j!} \sum_{l=0}^{j-1} P_{p,l}\{\text{Sr}\} / \lambda_{p-1}^+ \sum_{j=0}^{m} \frac{(m\rho_{p-1}^+)^j}{j!},\]

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

and

\[E[U_{p,k}] = \frac{\lambda_{p-1}^+ (m\rho_{p-1}^+)^k}{k!} \sum_{j=0}^{m} \frac{(m\rho_{p-1}^+)^j}{j!} \sum_{j=0}^{m} \frac{(m\rho_{p-1}^+)^j}{j!} \sum_{l=0}^{j-1} P_{p,l}\{\text{Sr}\} / \lambda_{p-1}^+ \sum_{j=0}^{m} \frac{(m\rho_{p-1}^+)^j}{j!},\]

\[1 \leq k \leq m - 2.\]

The second moments $E[U_{p,k}^2] = d^2U_{p,k}^*(s)/ds^2|_{s=0}, 0 \leq k \leq m - 1$, are given by

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

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

and

\[E[U_{p,k}^2] = \frac{1}{2} \left\{ \left[ \sum_{j=0}^{m} \frac{(m\rho_{p-1}^+)^j}{j!} \sum_{l=0}^{j-1} E[U_{p,l}] \right] \right. \sum_{j=0}^{m} \frac{(m\rho_{p-1}^+)^j}{j!} \sum_{l=0}^{j-1} E[U_{p,l}] / \left. \lambda_{p-1}^+ \sum_{j=0}^{m} \frac{(m\rho_{p-1}^+)^j}{j!} \right\},\]

\[1 \leq k \leq m - 2.\]

Let $S_{p,k}^*(s)$ be the LST of the DF for the total service time, denoted $S_{p,k}$, received by a customer of class $p$ who competes for the servers with $k$ other customers, where
\[ k \geq 0. \] For a customer of class \( p \) who waits upon arrival or resumes service after preemption, the LST of the DF for the total service time of a customer of class \( p \) 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 we have

\[
S_{p,k}^*(s) = \begin{cases} 
U_{p,k}^*(s) + V_{p,k}^*(s) & 0 \leq k \leq m - 1, \\
\frac{U_{p,m-1}^*(s)}{1 - V_{p,m-1}^*(s)} & k \geq m.
\end{cases}
\]

From Eqs. (36) and (39), we derive the following set of equations for \( \{S_{p,k}^*(s); 0 \leq k \leq m - 1\} \):

\[
(s + \lambda_{p-1}^+ + \mu)S_{p,0}^*(s) = \mu + \lambda_{p-1}^+ S_{p,1}^*(s),
\]

\[
[s + \lambda_{p-1}^+ + (k + 1)\mu]S_{p,k}^*(s) = \mu + \lambda_{p-1}^+ S_{p,k+1}^*(s) + k\mu S_{p,k-1}^*(s),
\]

\[
1 \leq k \leq m - 2,
\]

\[
(s + m\mu)S_{p,m-1}^*(s) = \mu + (m - 1)\mu S_{p,m-2}^*(s).
\]

The solution

\[
S_{p,k}^*(s) = \frac{\mu}{s + \mu} \quad 0 \leq k \leq m - 1
\]

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

\[
\frac{U_{p,k}^*(s)}{1 - V_{p,k}^*(s)} = \frac{\mu}{s + \mu} \quad 0 \leq k \leq m - 1 ; \quad S_{p,k}^*(s) = \frac{\mu}{s + \mu} \quad k \geq m.
\]

This result implies that \( S_{p,k}^* \) is exponentially distributed with a mean \( 1/\mu \) regardless of the number \( k \) of competing customers present in the system. Then the LST of the DF for the total service time of a customer of class \( p \) is given by

\[
S_p^*(s) = \sum_{k=0}^{\infty} Q_{p,k}^+ S_{p,k}^*(s) = \sum_{k=0}^{m-1} Q_{p,k}^+ S_{p,k}^*(s) + \sum_{k=m}^{\infty} Q_{p,k}^+ S_{p,k}^*(s)
\]

\[
= \sum_{k=0}^{m-1} Q_{p,k}^+ U_{p,k}^*(s) + \left\{ \sum_{k=0}^{m-1} Q_{p,k}^+ V_{p,k}^*(s) + C(m,m\rho_p^+ \Bigr) \right\} \frac{U_{p,m-1}^*(s)}{1 - V_{p,m-1}^*(s)}
\]

\[
= \frac{\mu}{s + \mu},
\]

signifying that the service time \( S_p^* \) is exponentially distributed with mean \( 1/\mu \).
We note that $W_p^\circ$ and $S_p$ are independent, because the joint LST of their DF is given as the product of the LSTs of their individual DFs:

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

$$= \sum_{k=0}^{m-1} Q_{p,k}^+ \left[ U_{p,k}^*(s') + V_{p,k}^*(s') \frac{\mu}{s' + \mu} \right] + C(m,m\rho_p^+)W_p^+(s) \frac{\mu}{s' + \mu}$$

$$= \left[ 1 - C(m,m\rho_p^+) + C(m,m\rho_p^+)W_p^+(s) \right] \frac{\mu}{s' + \mu} = W_p^\circ(s)S_p^*(s').$$

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.

9. Joint distribution of the waiting time and service time. By combining the arguments for deriving the LST of the DF for the total waiting time $W_p$ in Eq. (31) and the LST of the DF for the total service time $S_p$ in Eq. (41), the joint LST of the DF for $W_p$ and $S_p$ of a customer of class $p$ is given by

$$\tilde{T}_p^*(s, s') = \sum_{k=0}^{m-1} Q_{p,k}^+ U_{p,k}^*(s') + \left\{ G_{p-1}^*(s) \sum_{k=0}^{m-1} Q_{p,k}^+ V_{p,k}^*(s') + C(m,m\rho_p^+)W_p^+(s) \right\} \frac{U_{p,m-1}^*(s')}{1 - G_{p-1}^*(s)V_{p,m-1}(s')}$$

The marginal distributions yield

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

which agree with Eqs. (31) and (41), respectively.

The LST of the DF for the response time $T_p$ of a customer of class $p$ is given by

$$T_p^*(s) = \tilde{T}_p^*(s, s) = \sum_{k=0}^{m-1} Q_{p,k}^+ U_{p,k}^*(s) + \left\{ G_{p-1}^*(s) \sum_{k=0}^{m-1} Q_{p,k}^+ V_{p,k}^*(s) + C(m,m\rho_p^+)W_p^+(s) \right\} \frac{U_{p,m-1}^*(s)}{1 - G_{p-1}^*(s)V_{p,m-1}(s)}$$

Clearly it does not hold that $T_p^*(s) = W_p^*(s)S_p^*(s)$ for $p \geq 2$, which means that the total waiting time $W_p$ and the total service time $S_p$ are not independent for $p \geq 2$. In fact, they are positively correlated as we get from Eq. (42) the covariance of the
total waiting time and the total service time

\[
\text{Cov}[W_p, S_p] = \frac{\sum_{k=0}^{m-1} Q_{p,k}^+ E[V_{p,k}]}{m\mu(1-r_p)(1-\rho_{p-1}^+)} + \frac{\{[1-C(m,m\rho_p^+)]q_p + C(m,m\rho_p^+)\} E[V_{p,m-1}]}{m\mu(1-r_p)^2(1-\rho_{p-1}^+)} \tag{44}
\]

where \( E[V_{p,k}] \) and \( E[V_{p,m-1}] \) are given in Eqs. (37) and (38), respectively. For customers of the highest priority class (\( p = 1 \)), we have \( T^*_t(s) \) given in Eq. (12) and \( E[W_1, S_1] = 0 \), which means that \( W_1 \) and \( S_1 \) are independent, because they are never preempted. We plot \( \text{Cov}[W_p, S_p] \) in Fig. 9 for \( p \geq 2 \) for the numerical example described in Section 1.

From Eq. (43), we get the mean response time \( E[T_p] \) already given in Eq. (3). The second moment of the response time is given by

\[
E[T_p^2] = E[(W_p + S_p)^2] = E[W_p^2] + 2\text{Cov}[W_p, S_p] + 2E[W_p]E[S_p] + E[S_p^2]
\]

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

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

\[
+ \frac{2C(m,m\rho_p^+)}{(m\mu)^2} \left[ \frac{1-\rho_{p-1}^+\rho_p^+}{(1-\rho_{p-1}^+)^3(1-\rho_p^+)^2} + \frac{r_p(2-\rho_{p-1}^+ - \rho_p^+ - r_p^2(1-\rho_{p-1}^+\rho_p^+))}{(1-r_p)^2(1-\rho_{p-1}^+)^3(1-\rho_p^+)} + \frac{m(1-r_p\rho_{p-1}^+)}{(1-r_p)(1-\rho_{p-1}^+)(1-\rho_p^+)} \right] + \frac{2}{\mu^2} \tag{45}
\]

We have numerically confirmed that this yields the same result as Eq. (23) shown in Fig. 4.
For \( m = 1 \) (a single-server queue), we have
\[
\tilde{T}^*_p(s, s') = \frac{\mu(1 - \rho^+_p)}{[1 - \rho^+_p G^*_p(s)][s' + \lambda^+_p + \mu - \lambda^+_p G^*_p(s)]},
\]
which results in \( \tilde{T}^*_p(s, 0) = W^*_p(s) \) in Eq. (35) and \( \tilde{T}^*_p(0, s) = S^*_p(s) \) in Eq. (41), but \( \tilde{T}^*_p(s, s') \neq W^*_p(s)S^*_p(s') \) for \( p \geq 2 \). Therefore, \( W_p \) and \( S_p \) are not independent (except for \( p = 1 \)), although \( S_p \) is exponentially distributed. In fact, we get
\[
\text{Cov}[W_p, S_p] = \frac{\rho^+_{p-1}}{\mu^2(1 - \rho^+_p)}.
\]

10. **Concluding remarks.** In summary, we have reviewed the past studies on the analysis of the response time and waiting time in the \( M/M/m \) FCFS preemptive-resume priority queue from a unified viewpoint of the first passage time in a birth-and-death process with absorbing states. Our new contribution includes the explicit formulas for the second moment of the response time given in Eqs. (23) and (46) as well as those for the third moments of the response time and the waiting time given in Eq. (24) and in Eq. (34), respectively. The LST of the DF for the waiting time, \( W^*_p(s) \), is given in Eq. (31) [10, 13]. We have also clarified the dependence between the waiting time and the total service time. As noted in Section 3, we can obtain an explicit formula for the LST of the DF of the response time, \( T^*_p(s) \), for any system with the specific number \( m \) of servers in principle. However, the generic form for any \( m \) is not available so far.

Finally, we note that a similar analysis is possible for the \( M/M/m \) preemptive-resume priority queue with Last-Come First-Served (LCFS) and First-Come First-Displaced (FCFD) disciplines within the same class [9].

Appendix: Explicit results for the response and waiting times in the \( M/M/2 \) FCFS preemptive-resume priority queue

We summarize the explicit results for the response time and waiting time in the \( M/M/2 \) FCFS preemptive-resume priority queue with the following parameters:
\[
\rho_p := \frac{\lambda_p}{2\mu}; \quad \lambda^+_p := \sum_{k=1}^p \lambda_k; \quad \rho^+_p := \sum_{k=1}^p \rho_k = \frac{\lambda^+_p}{2\mu} \quad p = 1, 2, \ldots
\]

A.1 Response time

The distribution for the number of customers in the system in Eq. (1):
\[
Q^+_{p,0} = \frac{1 - \rho^+_p}{1 + \rho^+_p}, \quad Q^+_{p,1} = \frac{2(1 - \rho^+_p)\rho^+_p}{1 + \rho^+_p}, \quad Q^+_{p,2} = \frac{2(1 - \rho^+_p)(\rho^+_p)^2}{1 + \rho^+_p}.
\]

The LST of the DF for the first passage time in Eq. (6):
\[
T^*_{p,0}(s) = \frac{\mu[s + 2\lambda^+_{p-1} + 2\mu - \lambda^+_p G^*_p(s)]}{(s + \lambda^+_p + \mu)[s + \lambda^+_p - 1 - G^*_p(s)] + 2\mu},
\]
\[
T^*_{p,1}(s) = \frac{\mu(s + \lambda^+_p + 2\mu)}{(s + \lambda^+_p + \mu)[s + \lambda^+_p - 1 - G^*_p(s)] + 2\mu},
\]
\[
T^*_{p,2}(s) = \frac{\mu(s + \lambda^+_p + 2\mu)G^*_p(s)}{(s + \lambda^+_p + \mu)[s + \lambda^+_p - 1 - G^*_p(s)] + 2\mu}.
\]
where \( G^*_p(s) \) is the solution to the quadratic equation

\[
\lambda^+_p[G^*_p(s)]^2 - (s + \lambda^+_p + 2\mu)G^*_p(s) + 2\mu = 0.
\]

The LST of the DF for the response time of a customer of class \( p \):

\[
T^+_p(s) = Q^+_{p,0} T^+_p(s) + Q^+_{p,1} T^+_p(s) + \frac{Q^+_{p,2} T^+_p(s)}{1 - \rho_p^+ G^*_p(s)}
\]

\[
= \frac{1 - \rho_p^+}{(1 + \rho_p^+)[1 - \rho_p^+ G^*_p(s)]}
\]

\[
2\rho_p^+(s + \lambda^+_p + 2\mu) + [s + 2\lambda^+_p + 2\mu - \lambda^+_p G^*_p(s)][1 - \rho_p^+ G^*_p(s)]
\]

\[
(s + \lambda^+_p + \mu)\{s + \lambda^+_p[1 - G^*_p(s)] + 2\mu\} - \lambda^+_p \mu
\]

Moments for the response time of a customer of class \( p \):

\[
E[T^+_p] = \frac{1 + \rho_p^+ \rho_p^+}{\mu[1 - (\rho_p^+)^2][1 - (\rho_p^+)^2]}
\]

\[
E[T^+_p] = \frac{\{2(1 - \rho_p^+)^2 - \rho_p^+ [2 - 6\rho_p^+ + 3(\rho_p^+)^2]\}}{\mu^2(1 - \rho_p^+)[1 - (\rho_p^+)^2]^2[1 - (\rho_p^+)^2]}
\]

\[
\begin{align*}
&\{24 - 48\rho_p^+ + 41(\rho_p^+)^2 - 10(\rho_p^+)^3 - (\rho_p^+)^4\} \\
&-\rho_p^+ [48 - 161\rho_p^+ + 157(\rho_p^+)^2 - 53(\rho_p^+)^3 + 3(\rho_p^+)^4] \\
&+(\rho_p^+)^2 [89 - 241\rho_p^+ + 201(\rho_p^+)^2 - 59(\rho_p^+)^3 - 2(\rho_p^+)^4] \\
&-(\rho_p^+)^3 [39 - 141\rho_p^+ + 187(\rho_p^+)^2 - 71(\rho_p^+)^3 - 2(\rho_p^+)^4] \\
&+(\rho_p^+)^4 [15 - 55\rho_p^+ + 62(\rho_p^+)^2 - 19(\rho_p^+)^3 + 3(\rho_p^+)^4] \\
&-(\rho_p^+)^5 [1 - 2\rho_p^+ + 4(\rho_p^+)^3 + (\rho_p^+)^4] \\
&4\mu^3(1 - \rho_p^+)^2[1 - (\rho_p^+)^2]^2(1 - \rho_p^+)^2]
\end{align*}
\]

A.2 Waiting Time

Erlang’s C formula:

\[
C(2, a) = \frac{a^2}{2 + a} ; \quad C(2, 2\rho_p^+) = \frac{2(\rho_p^+)^2}{1 + \rho_p^+}
\]

Probabilities related with service preemption in Eqs. (28) and (29):

\[
r_p = \frac{\rho_p^+ + 1 + 2\rho_p^+}{1 + 2\rho_p^+ + 2(\rho_p^+)^2}, \quad q_p = \frac{2\rho_p^+ + 2\rho_p^+ + 2\rho_p^+}{1 + 2\rho_p^+ + 2(\rho_p^+)^2}.
\]

LST of the DF for the waiting time of a customer of class \( p \):

\[
W^*_p(s) = \frac{(1 - \rho_p^+)(2\rho_p^+ + 1 + \rho_p^+ - 1 + 2\rho_p^+ - \rho_p^+ G^*_p(s))}{(1 + \rho_p^+)(1 - \rho_p^+ G^*_p(s))} \{1 + 2\rho_p^+ - \rho_p^+ G^*_p(s)\} + 2(\rho_p^+)^2
\]
Moments for the waiting time of a customer of class $p$:

\[
E[W_p] = \frac{\rho_p^+ - \rho_p^- \rho_p^+ + (\rho_p^+ - \rho_p^-)^2}{\mu(1 - (\rho_p^-)^2)[1 - (\rho_p^+)^2]},
\]

\[
E[W_p^2] = \frac{(\rho_p^+)^2[5 + 2\rho_p^+ + (\rho_p^+)^2] + \rho_p^- \rho_p^+[5 + 17\rho_p^+ - 19(\rho_p^+)^2 + 3(\rho_p^+)^3] + (\rho_p^+ - \rho_p^-)^2[5 + 17\rho_p^+ - 9(\rho_p^+)^2 - 41(\rho_p^+)^3 + 16(\rho_p^+)^4] + (\rho_p^+ - \rho_p^-)^3[27 - 9\rho_p^+ - 85(\rho_p^+)^2 + 53(\rho_p^+)^3 + 2(\rho_p^+)^4] + (\rho_p^+ - \rho_p^-)^4[45 - 85\rho_p^+ - 52(\rho_p^+)^2 + 143(\rho_p^+)^3 - 45(\rho_p^+)^4]}{4\mu^2(1 - (\rho_p^-)^2)[1 - (\rho_p^+)^2]^2(1 - \rho_p^+)[1 - (\rho_p^+)^2]},
\]

A.3 Service time

LST's of the DF related with service preemption in Eqs. (36) and (39):

\[
V_{p,0}^*(s) = \frac{(\lambda_{p-1}^+)^2}{(s + \lambda_{p-1}^+ + \mu)(s + \lambda_{p-1}^+ + 2\mu) - \lambda_{p-1}^+ \mu},
\]

\[
V_{p,1}^*(s) = \frac{\lambda_{p-1}^+(s + \lambda_{p-1}^+ + \mu)}{(s + \lambda_{p-1}^+ + \mu)(s + \lambda_{p-1}^+ + 2\mu) - \lambda_{p-1}^+ \mu},
\]

\[
U_{p,0}^*(s) = \frac{\mu(s + 2\lambda_{p-1}^+ + 2\mu)}{(s + \lambda_{p-1}^+ + \mu)(s + \lambda_{p-1}^+ + 2\mu) - \lambda_{p-1}^+ \mu},
\]

\[
U_{p,1}^*(s) = \frac{\mu(s + \lambda_{p-1}^+ + 2\mu)}{(s + \lambda_{p-1}^+ + \mu)(s + \lambda_{p-1}^+ + 2\mu) - \lambda_{p-1}^+ \mu},
\]

which lead to Eq. (41).
A.4 Joint distribution of the waiting time and the service time in Eq. (42):

\[
\hat{T}_p(s, s') = \frac{1 - \rho_p^+}{(1 + \rho_p^+)[1 - \rho_p^+ G_{p-1}^*(s)]} \times \lambda_p^+(s' + \lambda_{p-1}^+ + 2\mu) + \mu[s' + 2\lambda_{p-1}^+ + 2\mu - \lambda_{p-1}^+ G_{p-1}^*(s)][1 - \rho_p^+ G_{p-1}^*(s)] \{s' + \lambda_{p-1}^+ [1 - G_{p-1}^*(s)] + 2\mu \} - \lambda_{p-1}^+ \mu.
\]

Covariance of the waiting time and the service time for a customer of class \( p \):

\[
\text{Cov}[W_p, S_p] = \frac{\rho_{p-1}^+ [\rho_p^+ + \rho_{p-1}^+ (3 + 2\rho_{p-1}^+)(1 + \rho_p^+)]}{2\mu^2 (1 + \rho_{p-1}^+)[1 - (\rho_{p-1}^+)^2][1 - \rho_p^+ G_{p-1}^*(s)]}.\]

REFERENCES

[1] I. Brosh, Preemptive priority assignment in multichannel systems, Operations Research, 17 (1969), 526–535.
[2] J. P. Buzen and A. B. Bondi, The response times of priority classes under preemptive resume in M/M/m queues, Operations Research, 31 (1983), 456–465.
[3] R. B. Cooper, Introduction to Queueing Theory, 2nd edition, Elsevier North Holland, New York, 1981.
[4] M. Fujiki, (Japanese) Fundamental theory and application on communication traffic. 5 queueing theory (part 2), Transactions of the Institute of Electronics and Communication Engineers of Japan, 55 (1972), 1194–1200.
[5] D. P. Gaver, Jr., A waiting line with interrupted service, including priorities, Journal of the Royal Statistical Society, Series B (Methodological), 24 (1962), 73–90.
[6] V. G. Kulkarni, Modeling and Analysis of Stochastic Systems, Chapman & Hall, Boca Raton, Florida, 1995.
[7] M. Segal, A multiserver system with preemptive priorities, Operations Research, 18 (1970), 316–323.
[8] H. Takagi, Detailed analysis of the response time and waiting time in the M/M/m FCFS preemptive-resume priority queue, in Queueing Theory and Network Applications (eds. T. V. Do, Y. Takahashi, W. Yue and V.-Ha Nguen), Springer, (2016), 3–17.
[9] H. Takagi, Analysis of the response and waiting times in the M/M/m LCFS preemptive-resume priority queue, International Journal of Pure and Applied Mathematics, 109 (2016), 325–370.
[10] 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, 1983, 163–166).
[11] H. M. Taylor and S. Karlin, An Introduction to Stochastic Modeling, 3rd edition, Academic Press, San Diego, California, 1998.
[12] H. White and L. S. Christie, Queuing with preemptive priorities or with breakdown, Operations Research, 6 (1958), 79–95.
[13] 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.

Received September 2015; 1st revision February 2016; final revision June 2016.

E-mail address: takagi@sk.tsukuba.ac.jp