On the Stationary LCFS-PR Single-server Queue: A Characterization via Stochastic Intensity

Naoto Miyoshi

August 2001, B–372
Abstract

We consider a stationary single-server queue with preemptive-resume last-come, first-served (LCFS-PR) queueing discipline. The LCFS-PR single-server queue has some interesting properties and has been studied in the queueing literature. In this paper, we generalize the previous works such that the input process to the queueing system is described as a general stationary marked point process, and derive some formulas concerning the joint distributions of the queue length and the remaining service times of respective customers in the system at arbitrary time instances as well as at arrival instances. The formulas obtained here cover the previous results. The tool for derivation is the Palm-martingale calculus, that is, the connection between the notion of Palm probability and that of stochastic intensity.

Keywords: LCFS-PR single-server queues, Palm-martingale calculus, stochastic intensity kernel, joint distribution of remaining service times, queue length distribution.

1 Introduction

In this paper, we consider a stationary single-server queue with preemptive-resume last-come, first-served (LCFS-PR) queueing discipline. In the LCFS-PR discipline, a newly arriving customer is immediately served upon the arrival and, if there is one in service at his/her arrival epoch, this service is interrupted and is resumed just after the new customer leaves the system. Note that the service of the new customer can be also interrupted by subsequently customers’ arrivals.

The LCFS-PR single-server queue is known to have some interesting properties and has been studied in the queueing literature: When the arrival process is given by the superposition of independent Poisson processes, the LCFS-PR queue is characterized as a symmetric queue and the stationary joint distribution of the queue length and the remaining service times of respective customers in the system is derived by Kelly [8]. Fakinos [6] extends the results to the case with GI/GI inputs and derives the stationary joint distribution of the queue length and the remaining service times at arrival instances. The corresponding distributions at arbitrary time instances are then derived by Yamazaki [16], which is later extended by Fakinos [7] to the case with the service time distributions depending on the queue length. Furthermore, Takine [15] considers the LCFS-PR single-server queue fed by the multiple arrival streams governed by a Markov chain and derives the matrix-product form solution for the stationary joint distribution of the queue length and the remaining service times at arbitrary time instances.

In the current paper, we generalize the previous works such that the input process to the queueing system is described as a general stationary marked point process, and derive some formulas concerning the joint distributions of the queue length and the remaining service times at arbitrary time instances as well as at arrival instances. With the formulas obtained here, we can explain the previous results in the literature. The tool for derivation is the Palm-martingale calculus (see, e.g., Baccelli and Brémaud [1] and Brémaud [4]), that is, the connection between the notion of Palm probability and that of stochastic intensity. Recently, the author derives a general formula for the stationary workload distribution of a work-conserving single-server queue by Palm-martingale approach ([11]), and the current work is an extension of it.

This paper is organized as follows: In the next section, the input process to the queueing system is described as a stationary marked point process. The stochastic intensity kernel associated with the marked point process is then provided. In Section 3, some general formulas for the stationary joint distributions of the queue length and the remaining service times are derived. The relations with the
previous results are also discussed. Section 4 focuses on the distributions of the queue length and Section 5 provides the results for the stationary distributions of the workload in the system.

2 Description of the Arrival Process

The arrival process to the queueing system is defined on a probability space \((Ω, ℱ, P)\). On \((Ω, ℱ)\), a family of measurable shift operators \(\{θ_t\}_{t ∈ ℝ}\) is defined and satisfies \(P ◦ θ_t^{-1} = P\) for \(t ∈ ℝ\), that is, \(\{θ_t\}_{t ∈ ℝ}\) is stationary in \(P\). Let \(N\) denote a point process on \((ℝ, ℬ(ℝ))\) counting the number of time epochs at which customers arrive at and enter the system, and let \(\{T_n\}_{n ∈ ℤ}\) be the corresponding point sequence satisfying

\[
\cdots < T_0 \leq 0 < T_1 < \cdots ;
\]

\[
\lim_{n \to ±∞} T_n = ±∞,
\]

that is, \(N\) is \(P\)-a.s. simple and locally finite. Each arriving customer belongs to any one of classes \(1, \ldots, K\) and let \(C_n\) (\(∈ \{1, \ldots, K\}\)), \(n ∈ ℤ\), denote the class of the customer who arrives at \(T_n\). Also, let \(S_n ≥ 0\), \(n ∈ ℤ\), denote the service time required by the customer who arrives at \(T_n\). We assume that \(\{(T_n, C_n, S_n)\}_{n ∈ ℤ}\) is compatible with \(\{θ_t\}_{t ∈ ℝ}\) in the sense that

\[
\{(T_n, C_n, S_n)\}_{n ∈ ℤ} ◦ θ_t = \{(T_n - t, C_n, S_n)\}_{n ∈ ℤ} \quad t ∈ ℝ.
\]

Due to the stationarity of \(\{θ_t\}_{t ∈ ℝ}\), \(\{(T_n, C_n, S_n)\}_{n ∈ ℤ}\) then forms a stationary marked point process on the real line with mark space \(\{1, \ldots, K\} × ℝ^+\). We further assume that the intensity of \(N\) is positive and finite, that is, \(λ = E[N((0, 1))] ∈ (0, ∞)\). Then, the Palm probability with respect to \((N, P, θ_t)\) is defined by

\[
P_N^0(A) = \frac{1}{λ} E \left[ \int_0^1 1_A ◦ θ_t N(dt) \right], \quad A ∈ ℳ,
\]

where \(1_A\) is the indicator of event \(A\) (see, e.g., [1]). The expectation with respect to \(P_N^0\) is denoted by \(E_N^0\). Note that \(P_N^0(T_0 = 0) = 1\) and \(E_N^0[1_{T_{n+1} - T_n} = 1/λ\) for all \(n ∈ ℤ\). The traffic intensity of the queueing system is given by \(ρ = λ E_N^0[S_0]\).

Let \(N_k\), \(k = 1, \ldots, K\), denote the sub-point process of \(N\) counting the number of arriving customers in class \(k\), that is, \(N_k\) is defined by

\[
N_k(B) = \int_B 1_{\{C_n = k\} ◦ θ_t N(dt) }, \quad B ∈ ℬ(ℝ).
\]

Clearly, \(N = \sum_{k=1}^K N_k\), and by the definition of \(P_N^0\) in (1), the intensity of \(N_k\) is given by \(λ_k = E[N_k((0, 1))] = λ P_N^0(C_0 = k)\). We assume that \(P_N^0(C_0 = k) > 0\) for all \(k = 1, \ldots, K\). Then, the Palm probability \(P_N^0\) with respect to \(N_k\) is well defined for \(k = 1, \ldots, K\), and they satisfy \(λ P_N^0 = \sum_{k=1}^K λ_k P_{N_k}\). This immediately implies that \(ρ = \sum_{k=1}^K ρ_k\) with \(ρ_k = λ_k E_{N_k}[S_0]\).

Now, for discussions in the following sections, we define the stochastic intensity kernel associated with the marked point process \(\{(T_n, C_n, S_n)\}_{n ∈ ℤ}\). Let \(\{ℱ_t\}_{t ∈ ℝ}\) denote a history to which \(\{(T_n, C_n, S_n)\}_{n ∈ ℤ}\) is adapted, that is, \(ℱ_t, t ∈ ℝ\), is a sub-\(σ\)-field of \(ℱ\) such that \(ℱ_t ⊂ ℱ_s\) whenever \(s ≤ t\) and \(\int_B 1_{\{C_n\} \circ θ_s N(ds)}\) is \(ℱ_t\)-measurable whenever \(B × U ∈ ℬ((-∞, t] × ℝ^+))\) and \(k ∈ \{1, \ldots, K\}\). We assume that the point process \(N\) admits the \(ℱ_t\)-stochastic intensity \(\{λ(t)\}_{t ∈ ℝ}\), which is \(P\)-a.s. locally integrable and satisfies

\[
E[N((s, t)] | ℱ_s] = E\left[\int_s^t λ(u) du \bigg| ℱ_s\right], \quad (s, t) ∈ ℬ(ℝ),
\]
where we can always assume that \( \{\lambda(t)\}_{t \in \mathbb{R}} \) is \( \mathcal{F}_t \)-predictable (see [3]). We also assume that \( \{\mathcal{F}_t\}_{t \in \mathbb{R}} \) is compatible with \( \{\theta_t\}_{t \in \mathbb{R}} \) in the sense of \( \theta_t \mathcal{F}_s = \mathcal{F}_{s-t} \), and assume that the following conditional distributions with respect to \( \mathcal{F}_{0-} = \bigcup_{t < 0} \mathcal{F}_t \) exist for \( k = 1, \ldots, K; \)

\[
\tilde{q}_k = P_{\lambda}^0(C_0 = k \mid \mathcal{F}_{0-}), \quad (2)
\]

\[
\tilde{F}_k(x) = P_{\lambda}^0(S_0 \leq x \mid \mathcal{F}_{0-}). \quad (3)
\]

These conditional distributions enable us to consider the case where the distribution of the class of an arriving customer depends on the history until his/her arrival and the distribution of the service time also depends on both the history and the class. Then, \( \mathcal{F}_t \)-stochastic intensity kernel associated with \( \{(T_n, C_n, S_n)\}_{n \in \mathbb{Z}} \) is given by \( \{\lambda(t) \tilde{q}_k \tilde{F}_k \circ \theta_t; k = 1, \ldots, K\}_{t \in \mathbb{R}} \), where \( \lambda(0) \tilde{q}_k \tilde{F}_k \) is \( \mathcal{F}_{0-} \)-measurable. For simplicity of the notation, we also write \( \tilde{F}(x) = \sum_{k=1}^K \tilde{q}_k \tilde{F}_k(x) = P_{\lambda}^0(S_0 \leq x \mid \mathcal{F}_{0-}). \)

### 3 Remaining Service Time Distributions

In the following, we assume that the stability condition \( \rho < 1 \) holds, and under this assumption, we also assume that the queueing system is in the steady state (see [9] or [1, Chap. 2]). Let \( L(t) \) denote the number of customers in the system at time \( t \). Note that, under the stability condition \( \rho < 1 \), we have \( P(L(0) > 0) = \rho \). When \( L(t) = l > 0 \), let \( I(t) = (I_1(t), \ldots, I_l(t)) \) and \( X(t) = (X_1(t), \ldots, X_l(t)) \), where \( I_j(t) \in \{1, \ldots, K\} \) and \( X_j(t) \geq 0 \), \( j = 1, \ldots, L(t) \), denote respectively the class and the remaining service time of the \( j \)th oldest customer in the system at time \( t \). Note that, since we consider the LCFS-PR queueing discipline, the oldest customer in the system is one that starts the current busy period, and the \( L(t) \)th oldest customer is the newest arriving customer in the system and in service at time \( t \).

First, we show the following:

**Theorem 3.1:** Let \( x = (x_1, x_2, \ldots) \in \mathbb{R}^\infty \) and \( k = (k_1, k_2, \ldots) \in \{1, \ldots, K\}^\infty \), and for \( l = 0, 1, 2, \ldots \), let

\[
A(l, k, x) = \{L(0-) = l; I_1(0-) = k_1, \ldots, I_l(0-) = k_l; X_1(0-) \leq x_1, \ldots, X_l(0-) \leq x_l\},
\]

with \( A(0, k, x) = \{L(0-) = 0\} \). Then, for \( l = 0, 1, 2, \ldots \),

\[
P(A(l, k, x)) = (1 - \rho) \prod_{j=1}^l G_j(k, x), \quad (4)
\]

\[
P_{\lambda}^0(A(l, k, x)) = \frac{(1 - \rho) \mathbb{E}[\lambda(0) | A(l, k, x)]}{\lambda} \prod_{j=1}^l G_j(k, x). \quad (5)
\]

where \( \prod_{j=1}^l \cdot = 1 \) and for \( j = 1, 2, \ldots \),

\[
G_j(k, x) = \int_0^{x_j} \mathbb{E}[\lambda(0) \tilde{q}_{k_j} (1 - \tilde{F}_{k_j}(y)) | A(j-1, k, x)] \, dy. \quad (6)
\]

In Theorem 3.1, \( x \) and \( k \) are defined on the infinite dimensional spaces only for notational convenience. Indeed, \( A(l, k, x) \) is concerning with \( \{k_1, \ldots, k_l\} \) and \( \{x_1, \ldots, x_l\} \), and \( G_j, j = 1, 2, \ldots \), is the mapping on \( \{1, \ldots, K\}^l \times \mathbb{R}^l \). Note that (4) represents the distribution of the system state under the time-stationary probability \( P \), that is, the distribution observed at arbitrary time epochs. While, (5) represents the distribution of the system state observed just before arrival instances.
Proof: The proof follows a similar line to that of Theorem 1 in [11], where the stationary workload distribution in a work-conserving single-server queue was derived. A subgoal is to have the relation between \( P(A(l, k, x)) \) and \( P(A(l-1, k, x)) \). Let \( N_{l-1,k,x} \) denote the point process counting the number of arrivals just before which the system is in \( A(l-1, k, x) \), that is,

\[
N_{l-1,k,x}(B) = \int_B 1_{A(l-1,k,x)} \circ \theta_t N(dt), \quad B \in \mathcal{B}(\mathbb{R}).
\]

The corresponding point sequence is denoted by \( \{T^{l-1,k,x}_n\}_{n \in \mathbb{Z}} \) and satisfies \( \cdots < T^{l-1,k,x}_0 < \cdots \) conventionally. The intensity of \( N_{l-1,k,x} \) is given by \( \lambda_{l-1,k,x} = E[N_{l-1,k,x}((0,1])] = \lambda P^0_l(A(l, k, x)) \). Since the queuing process is stationary and \( \lambda_1, \ldots, \lambda_K \) are all positive and finite, \( \lambda_{l-1,k,x} \) is also positive and finite. Let \( P^0_{l-1,k,x} \) denote the Palm probability with respect to \( N_{l-1,k,x} \) and let \( E^0_{l-1,k,x} \) be the corresponding expectation. Applying the Palm inversion formula (see [1]) to \( P(A(l, k, x)) \), we have

\[
P(A(l, k, x)) = \lambda_{l-1,k,x} E^0_{l-1,k,x} \left[ \int_{[0,T^{l-1,k,x}_1]} 1_{A(l,k,x)} \circ \theta_t dt \right].
\]

Here, since the customers older than the customer arriving at \( T_0 \) (called customer 0) are never served until the customer 0’s service is completed, we have for \( t \in \{0, T^{l-1,k,x}_1 \} \) on \( \{T^{l-1,k,x}_0 = 0\} = \{T_0 = 0\} \cap A(l-1, k, x) \),

\[
\theta_t^{-1} A(l, k, x) = \left\{ \begin{array}{ll}
\text{customer 0 is from class } k_l & \text{and is in service at time } t \\
\text{with the remaining service time not greater than } x_l
\end{array} \right\}, \quad P^0_{l-1,k,x}-a.s.
\]

Namely,

\[
\int_{[0,T^{l-1,k,x}_1]} 1_{A(l,k,x)} \circ \theta_t dt = (S_0 \wedge x_l) 1\{C_0 = k_l\}, \quad P^0_{l-1,k,x}-a.s.,
\]

where \( a \wedge b = \min(a, b) \). Thus, noting that \( P^0_{l-1,k,x}(\cdot) = P^0_l(\cdot \mid A(l-1, k, x)) \),

\[
P(A(l, k, x)) = \lambda_{l-1,k,x} E^0_{l-1,k,x} [(S_0 \wedge x_l) 1\{C_0 = k_l\}] = \lambda E^0_N[(S_0 \wedge x_l) 1\{C_0 = k_l\} 1_{A(l-1,k,x)}]. \tag{7}
\]

Now, we transform the last expression in terms of the Palm probability \( P^0_N \) into an expression in terms of the time-stationary probability \( P \) via the stochastic intensity kernel. Using the conditional distributions (2) and (3) with respect to \( \mathcal{F}_{0-} \),

\[
E^0_N[(S_0 \wedge x_l) 1\{C_0 = k_l\} \mid \mathcal{F}_{0-}] = \tilde{q}_{k_l} \int_0^{x_l} (1 - \tilde{F}_{k_l}(y)) \, dy.
\]

Since \( 1_{A(l-1,k,x)} \) is also \( \mathcal{F}_{0-} \)-measurable, applying Papangelou’s formula (see, e.g., [1, 4, 13]) into (7), we have

\[
P(A(l, k, x)) = E \left[ \lambda(0) \tilde{q}_{k_l} \int_0^{x_l} (1 - \tilde{F}_{k_l}(y)) \, dy 1_{A(l-1,k,x)} \right] = P(A(l-1, k, x)) \int_0^{x_l} E \left[ \lambda(0) \tilde{q}_{k_l} (1 - \tilde{F}_{k_l}(y)) \mid A(l-1, k, x) \right] \, dy.
\]

Hence, noting that \( P(A(0, k, x)) = P(L(0) = 0) = 1 - \rho \), we have (4) inductively. Equality (5) is immediately obtained from (4) by Papangelou’s formula \( \lambda P^0_N(A(l, k, x)) = E[\lambda(0) 1_{A(l,k,x)}] \).

When we ignore the class of each customer, Theorem 3.1 reduces to the following:
Theorem 4.1: Let

\[ A(l, x) = \{ L(0-) = l; X_1(0-) \leq x_1, \ldots, X_l(0-) \leq x_l \}, \]

with \( A(0, x) = \{ L(0-) = 0 \} \). Then, for \( l = 0, 1, 2, \ldots, \)

\[
P(A(l, x)) = (1 - \rho) \prod_{j=1}^{l} G_j(x),
\]

\[
P_0^N(A(l, x)) = \frac{(1 - \rho) \mathbb{E}[\lambda(0) \mid A(l, x)]}{\lambda} \prod_{j=1}^{l} G_j(x),
\]

where

\[ G_j(x) = \int_{0}^{x_j} \mathbb{E}[\lambda(0) (1 - \tilde{F}(y)) \mid A(j - 1, x)] \, dy. \]

Remark 3.1: Both Theorem 3.1 and Corollary 3.2 are natural extensions of the previous results for LCFS-PR queues. If and only if the stochastic intensity kernel depends on the system state only through the queue length, the integrand of \( G_j(k, x) \) in (6) reduces to

\[
\mathbb{E}[\lambda(0) \tilde{q}_{k_j} (1 - \tilde{F}_{k_j}(y)) \mid L(0) = j - 1] = \frac{\lambda \mathbb{E}_N^0[\tilde{q}_{k_j} (1 - \tilde{F}_{k_j}(y)) 1_{\{L(0-) = j-1\}}]}{\mathbb{P}(L(0) = j - 1)} = \mathbb{E}[\lambda(0) \mid L(0) = j - 1] \mathbb{P}_N^0(C_0 = k_j \mid L(0) = j - 1) \mathbb{P}_{N_{k_j}}^0(S_0 > y \mid L(0) = j - 1),
\]

where Papangelou’s formula is used in the first equality and the second equality follows from \( \mathbb{P}_N^0(\cdot \mid C_0 = k) \). Both are used in the third equality. In addition, if \( \mathbb{E}[\lambda(0) \mid L(0) = 1] = \mathbb{E}[\lambda(0) \mid L(0) = 2] = \cdots = \mathbb{E}[\lambda(0) \mid L(0) > 0] \), that is the case where \( N \) is a renewal point process independent of \( \{(C_n, S_n)\}_{n \in \mathbb{Z}} \), Theorem 3.1 reduces to the multiclass version of the results by Fakinos [7].

Remark 3.2: Furthermore, if and only if the stochastic intensity kernel is independent of the system state, which is known as the lack of bias assumption for ASTA (arrivals see time averages) property (see e.g., [10]), \( \mathbb{E}[\lambda(0) \mid A(l, k, x)] = \mathbb{E}[\lambda(0)] = \lambda \) in (5) and the integrand of \( G_j \) in (6) reduces to

\[
\mathbb{E}[\lambda(0) \tilde{q}_{k_j} (1 - \tilde{F}_{k_j}(y))] = \lambda_{k_j} (1 - F_{k_j}(y)),
\]

where \( F_{k_j}(y) = \mathbb{P}_{N_{k_j}}^0(S_0 \leq y) \) for \( k = 1, \ldots, K \). Then, both (4) and (5) also reduce to the well known formula for the LCFS-PR queue with Poisson arrivals (see [8]);

\[
P(A(l, k, x)) = (1 - \rho) \prod_{j=1}^{l} \lambda_{k_j} \int_{0}^{x_j} (1 - F_{k_j}(y)) \, dy.
\]

4 Queue Length Distributions

In this section, we focus on the queue length in the LCFS-PR queue and provide the formulas for various distributions. First, letting \( x_j, j = 1, 2, \ldots, \) go to infinity in (4) and (5), we have the following:

Theorem 4.1: Let \( k = (k_1, k_2, \ldots) \in \{1, \ldots, K\}^\infty \), and let for \( l = 0, 1, 2, \ldots, \)

\[ A(l, k, \infty) = \{ L(0-) = l; I_1(0-) = k_1, \ldots, I_l(0-) = k_l \}, \]
with \( A(0, k, \infty) = \{ L(0-) = 0 \} \). Then, for \( l = 0, 1, 2, \ldots, \)

\[
P(A(l, k, \infty)) = (1 - \rho) \prod_{j=1}^{l} g_j(k),
\]

\[
P_N^0(A(l, k, \infty)) = \frac{(1 - \rho) E[\lambda(0) \mid A(l, x, \infty)]}{\lambda} \prod_{j=1}^{l} g_j(k),
\]

where \( \prod_{j=1}^{l} \cdot = 1 \) conventionally and

\[
g_j(k) = E[\lambda(0) \mid A(j - 1, k, \infty)] P_N^0(C_0 = k_j \mid A(j - 1, k, \infty)) E_{N_k}^0 [S_0 \mid A(j - 1, k, \infty)].
\]

**Proof:** It suffices to show that, when each \( x_i, i = 1, \ldots, j, \) goes to infinity, \( G_j(k, x) \) in (6) converges to the right-hand side of (10). Noting that \( \int_0^{\infty} P_N^0(S_0 > y \mid A) \, dy = E_N^0[S_0 \mid A] \) for any event \( A \in \mathcal{F} \) and any \( k = 1, \ldots, K, \)

\[
\lim_{x_1, \ldots, x_j \to \infty} G_j(k, x) = \int_0^{\infty} E[\lambda(0) \tilde{g}_{k_j} (1 - \tilde{F}_{k_j}(y)) \mid A(j - 1, k, \infty)] \, dy
\]

\[
= \frac{\lambda}{P(A(j - 1, k, \infty))} \int_0^{\infty} E_N^0[\tilde{g}_{k_j} (1 - \tilde{F}_{k_j}(y)) 1_{A(j - 1, k, \infty)}] \, dy
\]

\[
= \frac{\lambda P_N^0(A(j - 1, k, \infty)) P_N^0(C_0 = k_j \mid A(j - 1, k, \infty)) E_{N_k}^0 [S_0 \mid A(j - 1, k, \infty)]}{P(A(j - 1, k, \infty))},
\]

where we use Papangelou’s formula in the second equality and use \( P_N^0(C_0 = k_j \mid A(j - 1, k, \infty)) \) in the third equality. Since \( \lambda P_N^0(A(j - 1, k, \infty)) = E[\lambda(0) 1_{A(j - 1, k, \infty)}] \), the proof is completed. \( \square \)

Next, we consider the joint distribution of the number of customers in each class. Let \( L(t) = (L_1(t), \ldots, L_K(t)) \), where \( L_k(t) (\in \mathbb{Z}_+) \), \( k = 1, \ldots, K \), denotes the number of class \( k \) customers in the system at time \( t \). Then, we have the following:

**Theorem 4.2:** Let \( l = (l_1, \ldots, l_K) \in \mathbb{Z}^K_+ \) and \( 0 = (0, \ldots, 0) \) on \( \mathbb{R}^K \). The distribution of \( L(0) \) under \( P \) satisfies the recursion such that \( P(L(0) = 0) = 1 - \rho \) and,

\[
P(L(0) = l) = \sum_{k=1}^{K} P(L(0) = l - e_k) g_{l-e_k}(k) 1_{\{l_k \geq 1\}},
\]

where \( e_k, k = 1, \ldots, K \), denotes the unit vector on \( \mathbb{R}^K \) such that only the \( k \)th element is equal to one and others are zero, and

\[
g_{l}(k) = E[\lambda(0) \mid L(0) = l] P_N^0(C_0 = k \mid L(0-) = l) E_{N_k}^0 [S_0 \mid L(0-) = l].
\]

Furthermore, the distribution of \( L(0-) \) under \( P_N^0 \) satisfies the recursion such that

\[
P_N^0(L(0-) = 0) = \frac{(1 - \rho) E[\lambda(0) \mid L(0) = 0]}{\lambda},
\]

and

\[
P_N^0(L(0-) = l) = E[\lambda(0) \mid L(0) = l] \sum_{k=1}^{K} \frac{P_N^0(L(0-) = l - e_k)}{E[\lambda(0) \mid L(0) = l - e_k]} g_{l-e_k}(k) 1_{\{l_k \geq 1\}}.
\]
Proof: Let $B(l, k) = \{L(0) = l, I_k(0) = k\}$, where $|l| = \sum_{k=1}^{K} l_k$, that is, $B(l, k)$ represents the event such that the queue length vector is given by $l$ and the class of the customer in service is $k$ at time $0$.

When $|l| > 0$, similar to the proof of Theorem 3.1, we have

$$P(B(l, k)) = \begin{cases} 0, & l_k = 0, \\ P(L(0) = l - e_k) \int_0^\infty E[\lambda(0) \tilde{q}_k (1 - \tilde{F}_k(z)) \mid L(0) = l - e_k] \, dz, & l_k \geq 1. \end{cases}$$

Similar to the proof of Theorem 4.1, we have $\int_0^\infty E[\lambda(0) \tilde{q}_k (1 - \tilde{F}_k(z)) \mid L(0) = l - e_k] \, dz = g_l - e_l(k)$. Thus, summing up over $k = 1, \ldots, K$, we have (11). Also, (12) follows from Papangelou’s formula $\lambda P_N^0(L(0) = 1) = E[\lambda(0) 1_{\{L(0) = 1\}}].$

Finally, when we ignore the class of each customer, both Theorems 4.1 and 4.2 reduce to the stationary distribution of the total number of customers in the system:

**Corollary 4.3:** For $l = 0, 1, 2, \ldots$,

$$P(L(0) = l) = (1 - \rho) \prod_{j=1}^{l} g_{j-1},$$

$$P_N^0(L(0) = l) = \frac{(1 - \rho) E[\lambda(0) \mid L(0) = l]}{\lambda} \prod_{j=1}^{l} g_{j-1},$$

where $\prod_{j=1}^{0} \cdot = 1$ conventionally, and for $j = 0, 1, 2, \ldots$,

$$g_j = E[\lambda(0) \mid L(0) = j] E_N^0[S_0 \mid L(0-) = j].$$

Theorems 4.1, 4.2 and Corollary 4.3 are also extensions of the previous results as seen in the following:

**Remark 4.1:** From Theorems 3.1 and 4.1, we have immediately

$$P(X_1(0) \leq x \mid L(0) = 1, I_1(0) = k) = \frac{1}{E_N^0[S_0 \mid L(0-0) = 0]} \int_0^x P_N^0(S_0 > y \mid L(0-) = 0) \, dy.$$

The right-hand side represents the stationary residual life time distribution of the service time of a class $k$ customer provided that he/she arrives at the empty system. This is indeed the observation discussed in [7, 14].

**Remark 4.2:** We can see from (14) that, for $l = 0, 1, 2, \ldots$,

$$\frac{P_N^0(L(0-) = l + 1)}{P_N^0(L(0-) = l)} = \frac{E[\lambda(0) \mid L(0) = l + 1]}{E[\lambda(0) \mid L(0) = l]} g_l = E[\lambda(0) \mid L(0) = l + 1] E_N^0[S_0 \mid L(0-) = l].$$

The last expression above represents the expected number of the interruptions of an $(l + 1)$st oldest customer’s service. Indeed, an $(l + 1)$st oldest customer sees $l$ customers in the system at his/her arrival epoch and the total time that the queue length is just equal to $l + 1$ during the sojourn time of the $(l + 1)$st oldest customer is given by his/her service time. Also from (14),

$$P_N^0(L(0-) = 0) = \frac{(1 - \rho) E[\lambda(0) \mid L(0) = 0]}{\lambda} = \frac{\lambda - E[\lambda(0) 1_{\{L(0) > 0\}}]}{\lambda} = 1 - E[\lambda(0) \mid L(0) > 0] E_N^0[S_0],$$

7
where we use $P(L(0) > 0) = \rho = \lambda E_N^0[S_0]$. Now, consider the case where the service times are i.i.d. and independent of the arrival point process $N$, and furthermore, suppose that $E[\lambda(0) \mid L(0) = 1] = E[\lambda(0) \mid L(0) = 2] = \cdots = E[\lambda(0) \mid L(0) > 0]$, that is the case of the GI/GI/1 queue. Then, (15) reduces to

$$E[\lambda(0) \mid L(0) > 0] E_N^0[S_0] =: \gamma,$$

and combining with (16), we can verify that the distribution of $L(0)$ under $P_N^0$ reduces to the geometric distribution with parameter $\gamma$, as is derived in [6].

On the other hand, we have from (13) that $P(L(0) = l + 1)/P(L(0) = l) = g_l$ for $l = 0, 1, 2, \ldots$ and

$$P(L(0) = 1) = (1 - \rho) E[\lambda(0) \mid L(0) = 0] E_N^0[S_0] \mid L(0) = 0 = \rho \lambda E_N^0[S_0] \mid L(0) = 0 = \lambda E_N^0[S_0] \mid L(0) = 0 \mid (1 - \gamma).$$

In the case where the service times are i.i.d. and independent of $N$, and further $E[\lambda(0) \mid L(0) = l] = E[\lambda(0) \mid L(0) > 0]$ for $l = 1, 2, \ldots$, then $g_l = \gamma$ for $l = 1, 2, \ldots$ and we have also the well-known result (see [14, 16]):

$$P(L(0) = l) = \rho (1 - \gamma) \gamma^{l-1}, \quad l = 1, 2, \ldots.$$

**Remark 4.3:** Finally, if and only if the stochastic intensity kernel is independent of the queue length and the class of each customer in the system, both (8) and (9) reduce to $(1 - \rho) \prod_{j=1}^{l} \rho_{k_j}$, which is also the well-known result for the LCFS-PR queue with Poisson arrivals (see [8]).

## 5 Workload Distributions

In this final section, we discuss the stationary distribution of the remaining amount of work in the system. Let $V(t) = (V_1(t), \ldots, V_K(t))$, where $V_k(t) (\geq 0)$, $k = 1, \ldots, K$ denotes the remaining work in the system of class $k$ customers at time $t$, that is,

$$V_k(t) = \sum_{j=1}^{L(t)} X_j(t) 1_{\{I_j(t) = k\}}, \quad k = 1, \ldots, K.$$

Then, we have the following:

**Theorem 5.1:** For $y = (y_1, \ldots, y_K) \in \mathbb{R}_+^K$,

$$P(V(0) \leq y) = (1 - \rho) \sum_{l=0}^{\infty} G^{(l)}(y), \quad (17)$$

$$P_N^0(V(0-) \leq y) = \frac{(1 - \rho) E[\lambda(0) \mid V(0) \leq y]}{\lambda} \sum_{l=0}^{\infty} G^{(l)}(y), \quad (18)$$

where vector inequalities represent the coordinatewise ordering, that is, $V(0) \leq y$ means $V_k(t) \leq y_k$ for $k = 1, \ldots, K$. Also, $G^{(0)} \equiv 1$ and, for $l = 1, 2, \ldots$, $G^{(l)}$ is recursively given by

$$G^{(l)}(y) = \sum_{k=1}^{K} \int_0^{y_k} G^{(l-1)}(y - z e_k) E[\lambda(0) \bar{q}_k (1 - \bar{F}_k(z)) \mid L(0) = l - 1, V(0) \leq y - z e_k] \, dz.$$
Hence, summing up the above over $l$,

$$C(l, y) = \{ L(0-) = l; V_1(0-) \leq y_1, \ldots, V_K(0-) \leq y_K \},$$

$$C'(l, k, y, z) = \{ L(0-) = l, I_l(0-) = k; V_1(0-) \leq y_1, \ldots, V_{k-1}(0-) \leq y_{k-1}, V_k(0-) - X_l(0-) \leq y_k, X_l(0-) \leq z, V_{k+1}(0-) \leq y_{k+1}, \ldots, V_K(0-) \leq y_K \}.$$

Then, similar to the proof of Theorem 3.1,

$$P(C'(l, k, y, z)) = P(C(l - 1, y)) E \left[ \lambda(0) \tilde{q}_k \int_0^z (1 - \tilde{F}_k(w)) \, dw \mid C(l - 1, y) \right].$$

Thus, considering the conditional convolution of $V_k(0-) - X_l(0-) \leq y_k - z$ and $X_l(0-) \leq z$ over $z \in [0, y_k]$, and then summing up over $k = 1, \ldots, K$, we have

$$P(C(l, y)) = \sum_{k=1}^{K} \int_0^{y_k} P(C(l - 1, y - ze_k)) E \left[ \lambda(0) \tilde{q}_k (1 - \tilde{F}_k(z)) \mid C(l - 1, y - ze_k) \right] \, dz.$$

Since $P(C(0, y)) = 1$, we have inductively,

$$P(C(l, y)) = (1 - \rho) G^{(l)}(y).$$

Hence, summing up the above over $l = 0, 1, 2, \ldots$, we have (17). Again, (18) follows from Papangelou’s formula.

When ignoring the class of each customer, Theorem 5.1 reduces to the stationary distribution of the total amount of work in a work-conserving single-server queueing system, which is derived in [11]:

**Corollary 5.2 ([11]):** Let $V(t)$ denote the stationary total work in the system at time $t$. Then,

$$P(V(0) \leq y) = (1 - \rho) \sum_{l=0}^{\infty} G^{(l)}(y),$$

$$P_0(V(0) \leq y) = \frac{(1 - \rho)}{\lambda} \frac{E[\lambda(0) \mid V(0) \leq y]}{\sum_{l=0}^{\infty} G^{(l)}(y)},$$

where $G^{(0)} \equiv 0$ and $G^{(l)}$ for $l = 1, 2, \ldots$, is recursively given by

$$G^{(l)}(y) = \int_0^y G^{(l-1)}(y - z) E[\lambda(0) (1 - \tilde{F}(z)) \mid L(0) = l - 1, V(0) \leq y - z] \, dz.$$

**Remark 5.1:** Corollary 5.2 is an extension of well-known Beneš’s formula for the M/GI/1 queue (see [2, 5]) and its extension to the GI/GI/1 queue (see [12]).

**References**

[1] F. Baccelli and P. Brémaud, *Elements of Queueing Theory: Palm-Martingale Calculus and Stochastic Recurrences*. Springer-Verlag, Berlin, 1994.

[2] V.E. Beneš, “On queues with Poisson arrivals,” *Ann. Math. Statist.*, vol. 28, pp. 670–677, 1957.

[3] P. Brémaud, *Point Processes and Queues: Martingale Dynamics*. Springer-Verlag, New York, 1981.

[4] P. Brémaud, “Characteristics of queueing systems observed at events and the connection between stochastic intensity and Palm probability,” *Queueing Syst.: Theory and Appl.*, vol. 5, pp. 99–112, 1989.
[5] R.B. Cooper and S.-C. Niu, “Beneš's formula for $M/G/1$-FIFO ‘explained’ by preemptive-resume LIFO,” J. Appl. Prob., vol. 23, pp. 550–554, 1986.

[6] D. Fakinos, “The $G/G/1$ queueing system with a particular queue discipline,” J. Royal Statist. Soc. Ser. B, vol. 43, pp. 190–196, 1981.

[7] D. Fakinos, “The single server queue with service depending on queue size and with preemptive resume last-come-first-served queue discipline,” J. Appl. Prob., vol. 24, pp. 758–767, 1987.

[8] F.P. Kelly, Reversibility and Stochastic Networks. John Wiley & Sons, Chichester, 1979.

[9] R.M. Loynes, “The stability of queues with non-independent inter-arrival and service times,” Proc. of Cambridge Phil. Soc., vol. 58, pp. 497–520, 1962.

[10] B. Melamed and W. Whitt, “On arrivals that see time averages: A martingale approach,” J. Appl. Prob., vol. 27, pp. 376–384, 1990.

[11] N. Miyoshi, “On the stationary workload distribution of work-conserving single-server queues: A general formula via stochastic intensity,” J. Appl. Prob., vol. 38, 2001. to appear.

[12] S.-C. Niu, “Representing workloads in $GI/G/1$ queues through the preemptive-resume LIFO queue discipline,” Queueing Sys.: Theory and Appl., vol. 3, pp. 157–178, 1988.

[13] F. Papangelou, “Integrability of expected increments and a related random change of time scale,” Trans. American Math. Soc., vol. 165, pp. 483–506, 1972.

[14] N.-Q. Rudesindo, “Note on the $GI/GI/1$ queue with LCFS-PR observed at arbitrary times,” Prob. Engr. and Informat. Sci., vol. 15, pp. 179–187, 2001.

[15] T. Takine, “Matrix product-form solution for an LCFS-PR single-server queue with multiple arrival streams governed by a Markov chain,” Queueing Sys.: Theory and Appl., 2002. to appear.

[16] G. Yamazaki, “Invariance relations of $GI/G/1$ queueing systems with preemptive-resume last-come first-served queue discipline,” J. Operat. Res. Soc. Japan, vol. 27, pp. 338–347, 1984.