



UDC 519.872, 519.217

PACS 02.50.Ey, 02.50.Ga, 07.05.Tp

DOI: 10.22363/2658-4670-2026-34-3-316-333

EDN: RHUTHS

A two-queue JSQ–SLQ service system with absolute preemptive priority and non-zero switch-over times

Evgeny A. Khityaev¹, Ivan S. Zaryadov^{1,2}, Tatiana A. Milovanova¹

¹ RUDN University, 6 Miklukho-Maklaya St, Moscow, 117198, Russian Federation

² Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilova St, Moscow, 119333, Russian Federation

(received: May 10, 2026; revised: June 4, 2026; accepted: June 18, 2026)

Abstract. Purpose. To study a two-queue system with a single server operating under the Join-the-Shortest-Queue (JSQ) routing discipline and the Serve-the-Longest-Queue (SLQ) service discipline, where the destination queue for an arriving customer and the queue to be served are selected according to probabilistic tie-breaking rules when the queue lengths are equal, with absolute preemptive priority and non-zero server switch-over times. Methods. The system is studied using a Markov description, stationary balance equations, conditional probability generating functions, and numerical analysis. Results. The system dynamics are described by a two-dimensional Markov process whose components are the length of the first queue and a phase variable representing the difference between the queue lengths together with the server state. For this process, the stationary balance equations are derived. Conditional probability generating functions are introduced and used to obtain a matrix form of the balance equations and to consider the condition for the existence of a stationary regime. Based on the obtained generating functions, expressions are derived for the main stationary performance measures, including the mean queue lengths, variances, correlation coefficient, effective arrival rates, the probability that the server is in a switch-over state, and the probability that the server is idle. Numerical experiments are performed to evaluate the effect of the arrival rate, service rates, switch-over rates, and probabilistic tie-breaking parameters on the stationary characteristics of the system and on the phase distribution of the underlying Markov process. Conclusions. The results show that the combined use of the JSQ and SLQ disciplines contributes to balancing the queue lengths. However, the system behavior and its stationary characteristics significantly depend on the combination of the arrival rate, service rates, server switch-over rates, and probabilistic tie-breaking parameters.

Key words and phrases: queueing system, join-the-shortest-queue, serve-the-longest-queue, absolute preemptive priority, switch-over times, Markov process, conditional probability generating functions

For citation: E. A. Khityaev, I. S. Zaryadov, T. A. Milovanova “A two-queue JSQ–SLQ service system with absolute preemptive priority and non-zero switch-over times,” *Discrete and Continuous Models and Applied Computational Science*, vol. 34, no. 3, pp. 316–333, 2026. DOI: 10.22363/2658-4670-2026-34-3-316-333. EDN: RHUTHS.

1. Introduction

In queueing theory, an important role is played by models that incorporate control mechanisms for customer arrivals and the service process. Such models are widely used in applications where high throughput, efficient load balancing, and low delays are required, in particular in telecommunication networks, data centers, and transportation systems [1–5]. As the load on such systems grows and their structure becomes more complex, dynamic policies that take the current state of the system into account become especially important. These policies make it possible to distribute the customer flow more rationally and to use service resources more efficiently.

© 2026 E. A. Khityaev, I. S. Zaryadov, T. A. Milovanova



This work is licensed under a Creative Commons “Attribution-NonCommercial 4.0 International” license.

One of the classical approaches is the Join-the-Shortest-Queue (JSQ) routing discipline, under which an arriving customer is routed to the subsystem with the shortest queue. This model was first studied in 1958 for a system with two parallel queues and was described by a two-dimensional Markov process [6]. The classical setting assumes Poisson arrivals and exponentially distributed service times, which allows the system to be described in Markovian terms and makes Markov-process methods applicable to its analysis. Since then, the JSQ discipline has been studied in a wide range of queueing models [7–14], including more recent studies [15–19]. The main advantage of the JSQ discipline is that it balances the load by distributing customers among the queues, thereby reducing the risk of overflow in systems with finite buffers and decreasing the mean waiting time [3, 5, 7]. Moreover, it has been shown that in large systems with parallel queues and servers, the mean waiting time under JSQ may converge to zero as the number of queues and servers increases, which illustrates the high efficiency of this discipline [5]. At the same time, JSQ requires information about the current lengths of all queues, which may introduce additional overhead in real systems [20].

Another important approach for improving system performance is the Serve-the-Longest-Queue (SLQ) service discipline, which was considered by Cohen in 1987 [21]. Under this discipline, the server selects the longest queue for service at each decision epoch, which helps equalize queue lengths and prevents individual queues from growing disproportionately. The SLQ discipline was later studied in several related queueing models [22–25]. As a result, SLQ reduces the probability that individual queues become overloaded, which helps to decrease the mean waiting time and, in systems with finite buffers, the loss probability. This is especially important under heavy traffic and in systems with multiple servers, where an uneven distribution of customers may overload some queues even when other servers are idle. In this context, the SLQ discipline has also been studied in the Halfin–Whitt regime [26]. Thus, SLQ can effectively complement routing approaches that balance customer arrivals. At the same time, it requires information about the current queue lengths to select a queue for service.

Combining the JSQ and SLQ disciplines is of particular interest because it allows the routing of arriving customers to be coordinated with the dynamic control of the service process. Models of this type were studied by Perel et al., including a preemptive two-queue system with zero switch-over times [27], a non-preemptive two-queue system with zero and non-zero switch-over times [28], and a preemptive three-queue system with zero switch-over times [29]. In these works, the rule by which the server selects a queue for service is deterministic and based on the comparison of queue lengths: the server selects the longest queue for service. In the preemptive models, the server switches as soon as a non-attended queue becomes longer than the attended one, whereas in the non-preemptive model switching decisions are made only upon service completions. Ties between queues are treated separately. In the two-queue models, when the queue lengths become equal, the server remains at the current queue. In the three-queue model, the same deterministic longest-queue rule is used when there is a unique longest queue; however, if two non-attended queues share the longest length while the attended queue is shorter, the choice between them is made according to a probabilistic tie-breaking rule. Thus, systems in which the server may switch queues even when the queue lengths are equal, according to a probabilistic tie-breaking rule for selecting the queue to be served, are of interest and have not yet been considered. In systems with non-zero switch-over times, preemptive priority may be relative, meaning that an ongoing switch-over cannot be interrupted, or absolute, meaning that the server may interrupt the current switch-over and change its direction of motion.

The present paper considers non-zero switch-over times between the queues and, in contrast to previously studied models, introduces a probabilistic rule for selecting the queue to be served when the queue lengths are equal. This generalization broadens the scope of the model and makes it better suited for systems in which server switching is not purely deterministic.

The proposed model can be used to analyze and optimize systems of this type, including data centers, server clusters, telecommunication networks, and distributed computing systems, where both load distribution and server switch-over times play an important role.

The remainder of the paper is organized as follows. Section 2 describes the model, including the routing and service rules, the switch-over mechanism, and tie-breaking probabilities. In Section 3, a two-dimensional Markov process is introduced and the corresponding stationary balance equations are derived. Section 4 introduces the conditional probability generating functions and discusses the

stability condition for the existence of a stationary regime. Section 5 presents expressions for the main stationary performance measures. Section 6 contains numerical results showing how the main system parameters affect the stationary characteristics and phase probabilities. Section 7 provides a discussion and interpretation of the obtained numerical results. Finally, Section 5 summarizes the results and discusses possible extensions of the model.

2. System description

The system consists of two infinite queues Q_i , $i = \overline{1, 2}$, a single server, and non-zero server switch-over times between the queues. It operates under an absolute preemptive priority rule, under which the server may interrupt an ongoing switch-over and change its direction.

Customers arrive according to a Poisson process with rate λ . Under the JSQ discipline, each arriving customer joins the currently shortest queue. If the queue lengths are equal, a probabilistic tie-breaking rule is applied: the customer joins queue Q_i with probability p_i , $i = \overline{1, 2}$, where $p_1 + p_2 = 1$.

Throughout the paper, we use notation partly based on Perel et al. [27–29], adapted to the present model.

Let $L_i(t)$ denote the number of customers in queue Q_i , $i = \overline{1, 2}$, at time $t > 0$.

According to the SLQ discipline, the server selects the longest queue for service. The service time of a customer from queue Q_i is exponentially distributed with rate μ_i , $i = \overline{1, 2}$.

Customers in each queue are served according to the First-Come, First-Served (FCFS) discipline. When the server switches over to another queue, the customers remaining in the previous queue stay in the system and wait until the server returns.

The dynamic priority service rule depends on the current queue lengths and is defined as follows.

- If, while the server is serving a customer from queue Q_i , a new customer arrives at queue Q_j , $j \neq i$, and after this arrival $L_j(t) > L_i(t)$, then the server interrupts service in queue Q_i and starts a switch-over to queue Q_j . If after this arrival $L_i(t) = L_j(t)$, then the server continues the current service in queue Q_i with probability q_i or interrupts it and starts a switch-over to queue Q_j with probability $q_j = 1 - q_i$.
- If, after completing the service of a customer from queue Q_i , the condition $L_i(t) < L_j(t)$, $j \neq i$, is satisfied, then the server starts a switch-over to queue Q_j . If $L_i(t) = L_j(t)$, then the server either remains at the current queue Q_i with probability q_i or switches over to queue Q_j with probability $q_j = 1 - q_i$.
- Since the priority is absolute, if a new customer arrives at queue Q_i while the server is switching over from queue Q_i to queue Q_j , then, if $L_i(t) > L_j(t)$, the server interrupts the current switch-over and starts switching back from Q_j to Q_i . If $L_i(t) = L_j(t)$, the server either continues switching over to Q_j with probability $q_j = 1 - q_i$ or returns to queue Q_i with probability q_i .

In all cases, whenever the server starts a switch-over to queue Q_j , the corresponding switch-over time is exponentially distributed with rate γ_j , $j = \overline{1, 2}$.

3. Two-dimensional Markov process and balance equations

Let $D(t) = L_1(t) - L_2(t)$ denote the difference between the numbers of customers in the first and second queues. To describe the position and switch-over state of the server, we introduce the phase variable $J(t)$, which encodes both the value of $D(t)$ and the position or switch-over state of the server. Then the state of the system at time t can be described by the Markov process $\xi(t) = \{L_1(t), J(t)\}$. The probability distribution of this process is defined as follows:

$$P_{n,d}(t) = \mathbb{P}\{L_1(t) = n, J(t) = d\}, \quad n \geq 0, d \in \mathcal{D},$$

where only admissible pairs (n, d) are considered, and

$$\mathcal{D} = \{1, 1_{s_1}, 0_1, 0_{s_1}, 0_{s_2}, 0_2, -1_{s_2}, -1\}.$$

In each phase, the numerical part of the notation indicates the value of $D(t)$, while the subscript, when present, specifies the position or switch-over state of the server. The phases d are defined as follows:

- 1 means that $D(t) = 1$, the server is located at Q_1 , and a new customer joins Q_2 .
- 1_{s_1} means that $D(t) = 1$, the server is switching over to Q_1 , and a new customer joins Q_2 .
- 0_1 means that $D(t) = 0$, the server is located at Q_1 , and a new customer joins Q_i with probability p_i , $i = \overline{1, 2}$, where $p_1 + p_2 = 1$.
- 0_{s_1} means that $D(t) = 0$ and the server is switching over to Q_1 .
- 0_{s_2} means that $D(t) = 0$ and the server is switching over to Q_2 .
- 0_2 means that $D(t) = 0$, the server is located at Q_2 , and a new customer joins Q_i with probability p_i , $i = \overline{1, 2}$, where $p_1 + p_2 = 1$.
- -1_{s_2} means that $D(t) = -1$, the server is switching over to Q_2 , and a new customer joins Q_1 .
- -1 means that $D(t) = -1$, the server is located at Q_2 , and a new customer joins Q_1 .

Under the stability condition, which ensures the existence of a stationary regime, the stationary probabilities are defined by

$$P_{n,d} = \lim_{t \rightarrow \infty} P_{n,d}(t), \quad n \geq 0, \quad d \in \mathcal{D}. \quad (1)$$

Then the stationary probabilities satisfy the following balance equations.

For the state $(n, 1)$:

$$(\lambda + \mu_1)P_{n,1} = \gamma_1 P_{n,1_{s_1}} + \lambda p_1 P_{n-1,0_1}, \quad n \geq 1.$$

For the state $(n, 1_{s_1})$:

$$(\lambda + \gamma_1)P_{n,1_{s_1}} = \lambda p_1 P_{n-1,0_{s_1}} + \lambda p_1 P_{n-1,0_{s_2}} + \lambda p_1 P_{n-1,0_2} + \mu_2 P_{n,0_2}, \quad n \geq 1.$$

For the state $(n, 0_1)$:

$$\begin{aligned} \lambda P_{0,0_1} &= \mu_1 q_1 P_{1,1} + \gamma_1 P_{0,0_{s_1}}, \quad n = 0, \\ (\lambda + \mu_1)P_{n,0_1} &= \lambda q_1 P_{n,1} + \mu_1 q_1 P_{n+1,1} + \gamma_1 P_{n,0_{s_1}}, \quad n \geq 1. \end{aligned}$$

For the state $(n, 0_{s_1})$:

$$\begin{aligned} (\lambda + \gamma_1)P_{0,0_{s_1}} &= \mu_2 q_1 P_{0,-1}, \quad n = 0, \\ (\lambda + \gamma_1)P_{n,0_{s_1}} &= \lambda q_1 P_{n,1_{s_1}} + \lambda q_1 P_{n-1,-1_{s_2}} + \lambda q_1 P_{n-1,-1} + \mu_2 q_1 P_{n,-1}, \quad n \geq 1. \end{aligned}$$

For the state $(n, 0_{s_2})$:

$$\begin{aligned} (\lambda + \gamma_2)P_{0,0_{s_2}} &= \mu_1 q_2 P_{1,1}, \quad n = 0, \\ (\lambda + \gamma_2)P_{n,0_{s_2}} &= \lambda q_2 P_{n,1} + \mu_1 q_2 P_{n+1,1} + \lambda q_2 P_{n,1_{s_1}} + \lambda q_2 P_{n-1,-1_{s_2}}, \quad n \geq 1. \end{aligned}$$

For the state $(n, 0_2)$:

$$\begin{aligned} \lambda P_{0,0_2} &= \gamma_2 P_{0,0_{s_2}} + \mu_2 q_2 P_{0,-1}, \quad n = 0, \\ (\lambda + \mu_2)P_{n,0_2} &= \gamma_2 P_{n,0_{s_2}} + \lambda q_2 P_{n-1,-1} + \mu_2 q_2 P_{n,-1}, \quad n \geq 1. \end{aligned}$$

For the state $(n, -1_{s_2})$:

$$(\lambda + \gamma_2)P_{n,-1_{s_2}} = \lambda p_2 P_{n,0_1} + \mu_1 P_{n+1,0_1} + \lambda p_2 P_{n,0_{s_1}} + \lambda p_2 P_{n,0_{s_2}}, \quad n \geq 0.$$

For the state $(n, -1)$:

$$(\lambda + \mu_2)P_{n,-1} = \lambda p_2 P_{n,0_2} + \gamma_2 P_{n,-1_{s_2}}, \quad n \geq 0.$$

As a result, we obtain the following system of balance equations:

$$\left\{ \begin{array}{lcl} (\lambda + \mu_1)P_{n,1} & = & \gamma_1 P_{n,1_{s_1}} + \lambda p_1 P_{n-1,0_1}, \quad n \geq 1, \\ (\lambda + \gamma_1)P_{n,1_{s_1}} & = & \lambda p_1 P_{n-1,0_{s_1}} + \lambda p_1 P_{n-1,0_{s_2}} + \lambda p_1 P_{n-1,0_2} + \mu_2 P_{n,0_2}, \quad n \geq 1, \\ \lambda P_{0,0_1} & = & \mu_1 q_1 P_{1,1} + \gamma_1 P_{0,0_{s_1}}, \quad n = 0, \\ (\lambda + \mu_1)P_{n,0_1} & = & \lambda q_1 P_{n,1} + \mu_1 q_1 P_{n+1,1} + \gamma_1 P_{n,0_{s_1}}, \quad n \geq 1, \\ (\lambda + \gamma_1)P_{0,0_{s_1}} & = & \mu_2 q_1 P_{0,-1}, \quad n = 0, \\ (\lambda + \gamma_1)P_{n,0_{s_1}} & = & \lambda q_1 P_{n,1_{s_1}} + \lambda q_1 P_{n-1,-1_{s_2}} + \lambda q_1 P_{n-1,-1} + \mu_2 q_1 P_{n,-1}, \quad n \geq 1, \\ (\lambda + \gamma_2)P_{0,0_{s_2}} & = & \mu_1 q_2 P_{1,1}, \quad n = 0, \\ (\lambda + \gamma_2)P_{n,0_{s_2}} & = & \lambda q_2 P_{n,1} + \mu_1 q_2 P_{n+1,1} + \lambda q_2 P_{n,1_{s_1}} + \lambda q_2 P_{n-1,-1_{s_2}}, \quad n \geq 1, \\ \lambda P_{0,0_2} & = & \gamma_2 P_{0,0_{s_2}} + \mu_2 q_2 P_{0,-1}, \quad n = 0, \\ (\lambda + \mu_2)P_{n,0_2} & = & \gamma_2 P_{n,0_{s_2}} + \lambda q_2 P_{n-1,-1} + \mu_2 q_2 P_{n,-1}, \quad n \geq 1, \\ (\lambda + \gamma_2)P_{n,-1_{s_2}} & = & \lambda p_2 P_{n,0_1} + \mu_1 P_{n+1,0_1} + \lambda p_2 P_{n,0_{s_1}} + \lambda p_2 P_{n,0_{s_2}}, \quad n \geq 0, \\ (\lambda + \mu_2)P_{n,-1} & = & \lambda p_2 P_{n,0_2} + \gamma_2 P_{n,-1_{s_2}}, \quad n \geq 0. \end{array} \right. \quad (2)$$

4. Conditional probability generating functions

For each phase

$$d \in \mathcal{D} = \{1, 1_{s_1}, 0_1, 0_{s_1}, 0_{s_2}, 0_2, -1_{s_2}, -1\},$$

we introduce the corresponding conditional probability generating function (PGF) $G_d(z)$:

$$G_d(z) = \sum_n P_{n,d} z^n, \quad |z| \leq 1. \quad (3)$$

Taking into account the admissible values of n for each phase, definition (3) gives

$$\left\{ \begin{array}{lcl} G_1(z) & = & \sum_{n \geq 1} P_{n,1} z^n, \\ G_{1_{s_1}}(z) & = & \sum_{n \geq 1} P_{n,1_{s_1}} z^n, \\ G_{0_1}(z) & = & \sum_{n \geq 0} P_{n,0_1} z^n, \\ G_{0_{s_1}}(z) & = & \sum_{n \geq 0} P_{n,0_{s_1}} z^n, \\ G_{0_{s_2}}(z) & = & \sum_{n \geq 0} P_{n,0_{s_2}} z^n, \\ G_{0_2}(z) & = & \sum_{n \geq 0} P_{n,0_2} z^n, \\ G_{-1_{s_2}}(z) & = & \sum_{n \geq 0} P_{n,-1_{s_2}} z^n, \\ G_{-1}(z) & = & \sum_{n \geq 0} P_{n,-1} z^n. \end{array} \right. \quad (4)$$

Multiplying the corresponding equations of system (2) by z^n , summing over n , and using relations (4), we obtain the following system of equations for the generating functions:

$$\begin{cases} (\lambda + \mu_1)G_1(z) - \gamma_1 G_{1_{s_1}}(z) - \lambda p_1 z G_{0_1}(z) & = 0, \\ (\lambda + \gamma_1)G_{1_{s_1}}(z) - \lambda p_1 z G_{0_{s_1}}(z) - \lambda p_1 z G_{0_{s_2}}(z) - (\lambda p_1 z + \mu_2)G_{0_2}(z) & = -\mu_2 P_{0,0_2}, \\ -(\lambda z + \mu_1)q_1 G_1(z) + (\lambda + \mu_1)z G_{0_1}(z) - \gamma_1 z G_{0_{s_1}}(z) & = \mu_1 z P_{0,0_1}, \\ -\lambda q_1 G_{1_{s_1}}(z) + (\lambda + \gamma_1)G_{0_{s_1}}(z) - \lambda q_1 z G_{-1_{s_2}}(z) - (\lambda z + \mu_2)q_1 G_{-1}(z) & = 0, \\ -(\lambda z + \mu_1)q_2 G_1(z) - \lambda q_2 z G_{1_{s_1}}(z) + (\lambda + \gamma_2)z G_{0_{s_2}}(z) - \lambda q_2 z^2 G_{-1_{s_2}}(z) & = 0, \\ -\gamma_2 G_{0_{s_2}}(z) + (\lambda + \mu_2)G_{0_2}(z) - (\lambda z + \mu_2)q_2 G_{-1}(z) & = \mu_2 P_{0,0_2}, \\ -(\lambda p_2 z + \mu_1)G_{0_1}(z) - \lambda p_2 z G_{0_{s_1}}(z) - \lambda p_2 z G_{0_{s_2}}(z) + (\lambda + \gamma_2)z G_{-1_{s_2}}(z) & = -\mu_1 P_{0,0_1}, \\ -\lambda p_2 G_{0_2}(z) - \gamma_2 G_{-1_{s_2}}(z) + (\lambda + \mu_2)G_{-1}(z) & = 0, \end{cases}$$

which can be written in matrix form as

$$A(z)\vec{G}(z) = \vec{P}(z), \quad (5)$$

where

$$\vec{G}(z) = (G_1(z), G_{1_{s_1}}(z), G_{0_1}(z), G_{0_{s_1}}(z), G_{0_{s_2}}(z), G_{0_2}(z), G_{-1_{s_2}}(z), G_{-1}(z))^T$$

is the vector of generating functions,

$$\vec{P}(z) = (0, -\mu_2 P_{0,0_2}, \mu_1 z P_{0,0_1}, 0, 0, \mu_2 P_{0,0_2}, -\mu_1 P_{0,0_1}, 0)^T$$

is the right-hand side vector containing the unknown boundary probabilities $P_{0,0_1}$ and $P_{0,0_2}$, and the matrix $A(z)$ is given by

$$\begin{pmatrix} \lambda + \mu_1 & -\gamma_1 & -\lambda p_1 z & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda + \gamma_1 & 0 & -\lambda p_1 z & -\lambda p_1 z & -(\lambda p_1 z + \mu_2) & 0 & 0 \\ -(\lambda z + \mu_1)q_1 & 0 & (\lambda + \mu_1)z & -\gamma_1 z & 0 & 0 & 0 & 0 \\ 0 & -\lambda q_1 & 0 & \lambda + \gamma_1 & 0 & 0 & -\lambda q_1 z & -(\lambda z + \mu_2)q_1 \\ -(\lambda z + \mu_1)q_2 & -\lambda q_2 z & 0 & 0 & (\lambda + \gamma_2)z & 0 & -\lambda q_2 z^2 & 0 \\ 0 & 0 & 0 & 0 & -\gamma_2 & \lambda + \mu_2 & 0 & -(\lambda z + \mu_2)q_2 \\ 0 & 0 & -(\lambda p_2 z + \mu_1) & -\lambda p_2 z & -\lambda p_2 z & 0 & (\lambda + \gamma_2)z & 0 \\ 0 & 0 & 0 & 0 & 0 & -\lambda p_2 & -\gamma_2 & \lambda + \mu_2 \end{pmatrix}.$$

By Cramer's rule, the generating functions, which are the solutions of system (5), are given by

$$G_d(z) = \frac{|A_d(z)|}{|A(z)|}, \quad d \in \mathcal{D},$$

where $|A(z)|$ is the determinant of the matrix $A(z)$, and $|A_d(z)|$ is the determinant of the matrix obtained by replacing the corresponding column of $A(z)$ with the right-hand side vector $\vec{P}(z)$.

We note that the generating functions $G_d(z)$, $d \in \mathcal{D}$, depend on the boundary probabilities $P_{0,0_1}$ and $P_{0,0_2}$, whereas the remaining probabilities at $n = 0$ are not independent and can be expressed in terms of these boundary probabilities using the stationary balance equations (2).

The determinant $|A(z)|$ is a sixth-degree polynomial in z and can be written as

$$|A(z)| = z(1 - z)h(z),$$

where $h(z)$ is a fourth-degree polynomial in z whose coefficients depend on the system parameters.

The explicit expression for $h(z)$ is omitted because of its considerable length and complex structure.

Setting the determinant $|A(z)|$ equal to zero, we obtain that its roots are $z = 0$, $z = 1$, and the roots of the polynomial $h(z)$. The root $z = 1$ corresponds to the normalization condition for the stationary probabilities:

$$\sum_{d \in \mathcal{D}} G_d(1) = G_1(1) + G_{0_1}(1) + G_{1_{s_1}}(1) + G_{0_{s_1}}(1) + G_{0_2}(1) + G_{0_{s_2}}(1) + G_{-1_{s_2}}(1) + G_{-1}(1) = \sum_{d \in \mathcal{D}} \lim_{z \rightarrow 1} \frac{|A_d(z)|}{|A(z)|} = 1, \quad (6)$$

whereas the factor z corresponding to the root $z = 0$ appears in both the numerator and the denominator of the expressions for the generating functions $G_d(z)$. Since the multiplicity of this root is the same in both, this factor cancels out, and no singularity arises at $z = 0$. Therefore, the root $z = 0$ does not affect the analytic properties of these functions.

Since the generating functions $G_d(z)$ must be analytic and finite for $|z| \leq 1$, all roots of the equation $|A(z)| = 0$ that lie inside the unit circle and are different from $z = 1$ must also be roots of $|A_d(z)|$.

Thus, the boundary probabilities are determined from two equations: one follows from the normalization condition (6), while the other follows from the requirement that, after substituting a root of the denominator $|A(z)|$ lying inside the unit circle, the numerators $|A_d(z)|$ also vanish. This ensures that the generating functions $G_d(z)$ remain finite and allows the boundary probabilities $P_{0,0_1}$ and $P_{0,0_2}$ to be determined uniquely.

We further note that a stationary regime exists only if the polynomial $h(z)$ has exactly one root inside the unit circle $|z| < 1$. The location of the remaining roots outside the unit circle does not affect the existence of the stationary regime.

Since, in general, an analytical study of the number of roots of the polynomial $h(z)$ inside the unit circle is difficult due to the complicated dependence of its coefficients on the system parameters, the subsequent analysis is carried out under the assumption that a stationary regime exists (see Section 5).

To determine the stability boundary, we consider the value of polynomial $h(z)$ at $z = 1$ and introduce the notation

$$g(\lambda) = h(1),$$

where $g(\lambda)$ is an eighth-degree polynomial in the parameter λ :

$$g(\lambda) = \sum_{i=0}^8 \alpha_i \lambda^i = \alpha_8 \lambda^8 + \sum_{i=1}^7 \alpha_i \lambda^i + \alpha_0 = (p_1 q_2 + q_1 p_2 + p_1 q_2 q_1 p_2) \lambda^8 + \sum_{i=1}^7 \alpha_i \lambda^i - g_1^2 g_2^2 \mu_1^2 \mu_2^2. \quad (7)$$

Since the stability boundary corresponds to a change in the sign of the polynomial $g(\lambda)$, it is necessary to consider the positive roots of the equation

$$g(\lambda) = 0. \quad (8)$$

However, an analytical determination of these roots is also difficult in the general case due to the complicated dependence of the coefficients of $g(\lambda)$ on the system parameters.

At the same time, it follows from (7) that

$$g(0) < 0, \quad \lim_{\lambda \rightarrow +\infty} g(\lambda) = +\infty.$$

Therefore, by the Bolzano–Cauchy theorem, equation (8) has at least one positive root. Since the arrival rate λ is a non-negative parameter, the stability boundary corresponds to the smallest positive root of this equation:

$$\lambda^* = \min\{\lambda > 0 : g(\lambda) = 0\}.$$

Thus, the interval $0 < \lambda < \lambda^*$ defines the range of arrival rates for which the existence of a stationary regime is considered. Moreover, for all λ in this interval, the condition

$$g(\lambda) < 0 \quad (9)$$

is satisfied.

In the numerical experiments presented in Section 6, we considered parameter sets for which condition (9) is satisfied. For each such set, we additionally checked that the polynomial $h(z)$ has exactly one root inside the unit circle $|z| < 1$, which ensures the existence of a stationary regime.

5. Stationary performance measures

We use the stationary probabilities defined earlier in (1) and the generating functions introduced in (3).

Then, at $z = 1$, we obtain

$$G_d(1) = \sum_{n \geq 0} P_{n,d}, \quad d \in \mathcal{D}.$$

We determine the values of the first and second derivatives of the generating functions $G_d(z)$ at $z = 1$:

$$\begin{aligned} G'_d(z) &= \frac{d}{dz} \left(\sum_{n \geq 0} P_{n,d} z^n \right) = \sum_{n \geq 0} n P_{n,d} z^{n-1} \Rightarrow G'_d(1) = \sum_{n \geq 0} n P_{n,d}, \\ G''_d(z) &= \frac{d}{dz} \left(\sum_{n \geq 0} n P_{n,d} z^{n-1} \right) = \sum_{n \geq 0} n(n-1) P_{n,d} z^{n-2} \Rightarrow G''_d(1) = \sum_{n \geq 0} n(n-1) P_{n,d}. \end{aligned}$$

The stationary marginal probability of phase d is given by

$$P_{\cdot,d} = \lim_{t \rightarrow \infty} \mathbb{P}\{J(t) = d\} = \sum_{n \geq 0} P_{n,d} = G_d(1), \quad d \in \mathcal{D}.$$

The stationary marginal probability that queue Q_1 contains n customers is given by

$$P_{n,\cdot} = \lim_{t \rightarrow \infty} \mathbb{P}\{L_1(t) = n\} = \sum_{d \in \mathcal{D}} P_{n,d}, \quad n \geq 0.$$

We now compute the mean queue lengths $\mathbb{E}(L_1)$ and $\mathbb{E}(L_2)$, as well as the mean difference $\mathbb{E}(D)$.

Since the numerical part of each phase $d \in \mathcal{D} = \{1, 1_{s_1}, 0_1, 0_{s_1}, 0_{s_2}, 0_2, -1_{s_2}, -1\}$ corresponds to the value of the queue-length difference, by the definition of expectation

$$\mathbb{E}(D) = \sum_{d \in \mathcal{D}} d G_d(1) = \sum_{d \in \mathcal{D}} d P_{\cdot,d}. \quad (10)$$

Expanding (10) in terms of the generating functions, we obtain

$$\begin{aligned} \mathbb{E}(D) &= 1 \cdot G_1(1) + 1 \cdot G_{1_{s_1}}(1) + 0 \cdot G_{0_1}(1) + 0 \cdot G_{0_{s_1}}(1) + 0 \cdot G_{0_{s_2}}(1) + 0 \cdot G_{0_2}(1) - \\ &\quad - 1 \cdot G_{-1_{s_2}}(1) - 1 \cdot G_{-1}(1) = G_1(1) + G_{1_{s_1}}(1) - G_{-1_{s_2}}(1) - G_{-1}(1). \end{aligned}$$

Equivalently, expanding (10) in terms of the marginal phase probabilities, we obtain

$$\begin{aligned} \mathbb{E}(D) &= 1 \cdot P_{\cdot,1} + 1 \cdot P_{\cdot,1_{s_1}} + 0 \cdot P_{\cdot,0_1} + 0 \cdot P_{\cdot,0_{s_1}} + 0 \cdot P_{\cdot,0_{s_2}} + 0 \cdot P_{\cdot,0_2} - \\ &\quad - 1 \cdot P_{\cdot,-1_{s_2}} - 1 \cdot P_{\cdot,-1} = P_{\cdot,1} + P_{\cdot,1_{s_1}} - P_{\cdot,-1_{s_2}} - P_{\cdot,-1}. \end{aligned}$$

The mean length of the first queue is given by

$$\mathbb{E}(L_1) = \sum_{n \geq 0} n P_{n,\cdot} = \sum_{n \geq 0} n \sum_{d \in \mathcal{D}} P_{n,d} = \sum_{d \in \mathcal{D}} \sum_{n \geq 0} n P_{n,d} = \sum_{d \in \mathcal{D}} G'_d(1).$$

The mean length of the second queue can be obtained from the relation between the queue lengths. Since $D = L_1 - L_2$, we have $L_2 = L_1 - D$. Hence, by the linearity of expectation,

$$\mathbb{E}(L_2) = \mathbb{E}(L_1) - \mathbb{E}(D) = \sum_{d \in \mathcal{D}} G'_d(1) - \sum_{d \in \mathcal{D}} d G_d(1) = \sum_{d \in \mathcal{D}} (G'_d(1) - d G_d(1)).$$

The covariance of the queue lengths $\text{Cov}(L_1, L_2)$ is defined as

$$\text{Cov}(L_1, L_2) = \mathbb{E}(L_1 L_2) - \mathbb{E}(L_1) \mathbb{E}(L_2).$$

Since $L_2 = L_1 - D$, we have

$$\mathbb{E}(L_1 L_2) = \mathbb{E}(L_1 (L_1 - D)) = \mathbb{E}(L_1^2 - L_1 D) = \mathbb{E}(L_1^2) - \mathbb{E}(L_1 D).$$

Therefore,

$$\text{Cov}(L_1, L_2) = \mathbb{E}(L_1^2) - \mathbb{E}(L_1 D) - \mathbb{E}(L_1) \mathbb{E}(L_2),$$

where

$$\begin{aligned} \mathbb{E}(L_1^2) &= \sum_{n \geq 0} n^2 P_{n,\cdot} = \sum_{d \in \mathcal{D}} \sum_{n \geq 0} n^2 P_{n,d} = \sum_{d \in \mathcal{D}} \sum_{n \geq 0} n(n-1) P_{n,d} + \sum_{d \in \mathcal{D}} \sum_{n \geq 0} n P_{n,d} = \\ &= \sum_{d \in \mathcal{D}} G_d''(1) + \sum_{n \geq 0} n P_{n,\cdot} = \sum_{d \in \mathcal{D}} G_d''(1) + \mathbb{E}(L_1) = \sum_{d \in \mathcal{D}} G_d''(1) + \sum_{d \in \mathcal{D}} G_d'(1), \end{aligned}$$

$$\begin{aligned} \mathbb{E}(L_1 D) &= \sum_{d \in \mathcal{D}} \sum_{n \geq 0} n d P_{n,d} = \sum_{n \geq 0} n P_{n,1} + \sum_{n \geq 0} n P_{n,1_{s_1}} - \sum_{n \geq 0} n P_{n,-1_{s_2}} - \sum_{n \geq 0} n P_{n,-1} = \\ &= G_1'(1) + G_{1_{s_1}}'(1) - G_{-1_{s_2}}'(1) - G_{-1}'(1). \end{aligned}$$

The variances of the numbers of customers in the first and second queues, $\text{Var}(L_1)$ and $\text{Var}(L_2)$, are computed as follows:

$$\begin{aligned} \text{Var}(L_1) &= \mathbb{E}(L_1^2) - (\mathbb{E}(L_1))^2 = \sum_{d \in \mathcal{D}} G_d''(1) + \sum_{d \in \mathcal{D}} G_d'(1) - \left(\sum_{d \in \mathcal{D}} G_d'(1) \right)^2 = \\ &= \sum_{d \in \mathcal{D}} G_d''(1) + \left(1 - \sum_{d \in \mathcal{D}} G_d'(1) \right) \sum_{d \in \mathcal{D}} G_d'(1), \end{aligned}$$

$$\begin{aligned} \text{Var}(L_2) &= \mathbb{E}(L_2^2) - (\mathbb{E}(L_2))^2 = \mathbb{E}((L_1 - D)^2) - (\mathbb{E}(L_1 - D))^2 = \mathbb{E}(L_1^2 - 2L_1 D + D^2) - \\ &- (\mathbb{E}(L_1) - \mathbb{E}(D))^2 = \mathbb{E}(L_1^2) - 2\mathbb{E}(L_1 D) + \mathbb{E}(D^2) - (\mathbb{E}(L_1) - \mathbb{E}(D))^2. \end{aligned}$$

Moreover, since

$$D^2 = \begin{cases} 1, & d \in \{1, 1_{s_1}, -1_{s_2}, -1\}, \\ 0, & d \in \{0_1, 0_{s_1}, 0_{s_2}, 0_2\}, \end{cases}$$

we have

$$\mathbb{E}(D^2) = G_1(1) + G_{1_{s_1}}(1) + G_{-1_{s_2}}(1) + G_{-1}(1).$$

The correlation coefficient $\text{Cor}(L_1, L_2)$ is computed as

$$\text{Cor}(L_1, L_2) = \frac{\text{Cov}(L_1, L_2)}{\sqrt{\text{Var}(L_1) \text{Var}(L_2)}}.$$

Next, we determine the effective arrival rates λ_i^{eff} , $i = \overline{1, 2}$, of customers joining queue Q_i , $i = \overline{1, 2}$.

For queue Q_1 : If $L_1 < L_2$, then an arriving customer joins queue Q_1 with probability 1, and the corresponding contribution is $\lambda(P_{\cdot, -1_{s_2}} + P_{\cdot, -1})$. If $L_1 = L_2$, then the customer joins queue Q_1 with probability p_1 , and the corresponding contribution is $\lambda p_1(P_{\cdot, 0_1} + P_{\cdot, 0_{s_1}} + P_{\cdot, 0_{s_2}} + P_{\cdot, 0_2})$.

Consequently, the effective arrival rate to queue Q_1 is

$$\lambda_1^{\text{eff}} = \lambda(P_{\cdot, -1_{s_2}} + P_{\cdot, -1} + p_1(P_{\cdot, 0_1} + P_{\cdot, 0_{s_1}} + P_{\cdot, 0_{s_2}} + P_{\cdot, 0_2})).$$

For queue Q_2 : If $L_1 > L_2$, then an arriving customer joins queue Q_2 with probability 1, and the corresponding contribution is $\lambda(P_{\cdot, 1} + P_{\cdot, 1_{s_1}})$. If $L_1 = L_2$, then the customer joins queue Q_2 with probability p_2 , and the corresponding contribution is $\lambda p_2(P_{\cdot, 0_1} + P_{\cdot, 0_{s_1}} + P_{\cdot, 0_{s_2}} + P_{\cdot, 0_2})$.

Consequently, the effective arrival rate to queue Q_2 is

$$\lambda_2^{\text{eff}} = \lambda(P_{\cdot, 1} + P_{\cdot, 1_{s_1}} + p_2(P_{\cdot, 0_1} + P_{\cdot, 0_{s_1}} + P_{\cdot, 0_{s_2}} + P_{\cdot, 0_2})).$$

Since the sum of the effective arrival rates to queues Q_1 and Q_2 equals the total arrival rate λ , we have

$$\lambda_1^{\text{eff}} + \lambda_2^{\text{eff}} = \lambda.$$

The effective load factor of queue Q_i , interpreted as the fraction of time during which the server serves this queue, is given by

$$\rho_i^{\text{eff}} = \frac{\lambda_i^{\text{eff}}}{\mu_i}, \quad i = \overline{1, 2}.$$

The mean fraction of time during which the server is in a switch-over state is given by

$$P_{\text{sw}} = P_{\cdot, 1s_1} + P_{\cdot, 0s_1} + P_{\cdot, 0s_2} + P_{\cdot, -1s_2}.$$

The probability that the server is idle, that is, neither serving nor switching over, is defined as follows:

$$P_{\text{idle}} = P_{0,0_1} + P_{0,0_2} = 1 - \rho_1^{\text{eff}} - \rho_2^{\text{eff}} - P_{\text{sw}}.$$

Since

$$G_d(z) = \sum_{n=0}^{\infty} P_{n,d} z^n = P_{0,d} + P_{1,d}z + P_{2,d}z^2 + \dots, \quad |z| \leq 1,$$

substituting $z = 0$ into the expression for $G_d(z)$ gives

$$G_d(0) = P_{0,d} + P_{1,d} \cdot 0 + P_{2,d} \cdot 0^2 + \dots = P_{0,d}.$$

It follows that the mean fraction of time during which the server is in a switch-over state while the system is empty is given by

$$P_{\text{sw}0} = P_{0,0s_1} + P_{0,0s_2} = G_{0s_1}(0) + G_{0s_2}(0).$$

Then the probability that the system is empty, regardless of the server state, is given by

$$P_0 = P_{0,0_1} + P_{0,0s_1} + P_{0,0s_2} + P_{0,0_2} = P_{\text{idle}} + P_{\text{sw}0}.$$

6. Numerical results

This section presents the numerical results in two groups of tables. Tables 1–4 show how changes in the customer arrival rate, service rates, server switch-over rates, and tie-breaking probabilities affect the main stationary characteristics of the system. These characteristics include the mean queue lengths, the variances of the queue lengths, the correlation coefficient, the effective customer arrival rates, and the fractions of time during which the server is serving, switching over, or idle for different parameter values. Tables 5–7 present the stationary probabilities of the process phases, each of which corresponds to a particular value of the queue-length difference D and a particular server state, and show how these probabilities change as the customer arrival rate λ and the tie-breaking probabilities vary.

We introduce the load parameter

$$\rho = \frac{\lambda}{\lambda^*}, \quad \rho \in (0, 1),$$

and set $\lambda = \rho\lambda^*$ for $\rho = 0.3, 0.5, 0.7, 0.85, 0.95$ to evaluate the effect of the arrival rate on the stationary characteristics of the system (see Table 1).

In Tables 2–7, the results are presented for a moderate load, $\rho = 0.7$, that is, for $\lambda = 0.7\lambda^*$, where λ^* is the corresponding stability boundary. This choice makes it possible to illustrate the effect of the model parameters on the system characteristics more clearly. We recall that the parameter $\rho = \frac{\lambda}{\lambda^*}$ characterizes the load level relative to the stability boundary, whereas the quantities ρ_i^{eff} represent the fractions of time during which the server serves the corresponding queues Q_i , $i = 1, 2$.

Table 1

Dependence of the main performance measures on the load parameter ρ for $\lambda^* \approx 1.491719$, $\mu_1 = \mu_2 = 3$, $\gamma_1 = \gamma_2 = 3$, $p_1 = p_2 = q_1 = q_2 = 0.5$.

ρ	λ	$\mathbb{E}(L_1)$	$\mathbb{E}(L_2)$	$\text{Var}(L_1)$	$\text{Var}(L_2)$	$\text{Cor}(L_1, L_2)$	P_{idle}
0.30	0.448	0.172	0.172	0.169	0.169	0.404	0.698
0.50	0.746	0.416	0.416	0.470	0.470	0.667	0.496
0.70	1.044	0.996	0.996	1.630	1.630	0.875	0.295
0.85	1.268	2.459	2.459	7.510	7.510	0.969	0.146
0.95	1.417	8.326	8.326	74.062	74.062	0.997	0.049

Table 2

Influence of the service rates on the main performance measures at $\rho = 0.7$, $\gamma_1 = \gamma_2 = 2$, $p_1 = p_2 = q_1 = q_2 = 0.5$.

μ_1	μ_2	λ^*	$\mathbb{E}(L_1)$	$\mathbb{E}(L_2)$	$\text{Cor}(L_1, L_2)$	ρ_1^{eff}	ρ_2^{eff}	P_{sw}	P_{idle}
0.1	0.2	0.12815	1.189	1.088	0.897	0.403	0.247	0.053	0.297
1	3	0.87537	1.073	0.969	0.881	0.274	0.113	0.320	0.293
5	5	1.51539	1.037	1.037	0.887	0.106	0.106	0.505	0.283
0.5	10	0.76393	1.256	1.017	0.907	0.407	0.033	0.286	0.274

Table 2 shows the effect of the service rates μ_1 and μ_2 on the stationary characteristics of the system. The influence of the server switch-over rates γ_1 and γ_2 on the system characteristics is shown in Table 3.

Table 4 presents the system characteristics for different combinations of the tie-breaking probabilities p_1 and q_1 .

Next, we consider the stationary probabilities of the process phases for different values of the load parameter ρ (see Table 5). Due to rounding, the sums of probabilities in the rows may differ slightly from one.

The effect of the tie-breaking probability p_1 on the stationary distribution over the process phases is shown in Table 6.

Finally, Table 7 shows how the tie-breaking probability q_1 affects the stationary distribution over the process phases.

The numerical results presented in this section are further discussed and interpreted in Section 7.

7. Discussion

The results presented in Table 1 show that, as the load parameter ρ increases, the mean queue lengths $\mathbb{E}(L_1)$ and $\mathbb{E}(L_2)$ grow significantly. As ρ approaches one, the mean queue lengths increase sharply and nonlinearly. This is because the arrival rate λ becomes close to the stability boundary λ^* , so that the server spends less time idle and more time serving customers or switching over between the queues. Under such conditions, even a small increase in the load may lead to customer accumulation in the queues and, consequently, to a sharp increase in the queue lengths.

At the same time, the variances of the queue lengths grow faster than the corresponding mean values, especially for high values of ρ . This indicates that the queue lengths become more variable as the load increases. Together with the growth of the mean queue lengths, this indicates a decrease in the efficiency of the system.

Table 3

Influence of the switch-over rates on the main performance measures at $\rho = 0.7$, $\mu_1 = \mu_2 = 2$, $p_1 = p_2 = q_1 = q_2 = 0.5$.

γ_1	γ_2	λ^*	$\mathbb{E}(L_1)$	$\mathbb{E}(L_2)$	$\text{Cor}(L_1, L_2)$	ρ_1^{eff}	ρ_2^{eff}	P_{sw}	P_{idle}
0.1	0.2	0.15689	1.269	1.174	0.919	0.025	0.030	0.688	0.257
1	3	0.89238	1.063	0.978	0.882	0.143	0.170	0.401	0.287
5	5	1.38724	1.020	1.020	0.877	0.243	0.243	0.213	0.301
0.5	10	0.78691	1.228	1.027	0.904	0.110	0.165	0.463	0.262

Table 4

Influence of tie-breaking probabilities on the main performance measures at $\rho = 0.7$, $\mu_1 = \mu_2 = 2$, $\gamma_1 = \gamma_2 = 2$.

p_1	q_1	λ^*	$\mathbb{E}(L_1)$	$\mathbb{E}(L_2)$	$\text{Cor}(L_1, L_2)$	ρ_1^{eff}	ρ_2^{eff}	P_{sw}	P_{idle}
0.2	0.2	1.02190	0.931	0.932	0.868	0.114	0.244	0.323	0.320
	0.5	0.99603	0.931	1.057	0.880	0.134	0.214	0.355	0.296
	0.8	0.93088	0.877	1.127	0.895	0.146	0.180	0.393	0.281
0.5	0.2	0.97267	1.037	0.909	0.878	0.148	0.192	0.362	0.298
	0.5	0.99447	0.996	0.996	0.875	0.174	0.174	0.357	0.295
	0.8	0.97267	0.909	1.037	0.878	0.192	0.148	0.362	0.298
0.8	0.2	0.93088	1.127	0.877	0.895	0.180	0.146	0.393	0.281
	0.5	0.99603	1.057	0.931	0.880	0.214	0.134	0.355	0.296
	0.8	1.02190	0.932	0.931	0.868	0.244	0.114	0.323	0.320

Table 5

Stationary phase probabilities for different values of ρ at $\lambda^* \approx 1.491719$, $\mu_1 = \mu_2 = 3$, $\gamma_1 = \gamma_2 = 3$, $p_1 = p_2 = q_1 = q_2 = 0.5$.

ρ	$P_{\cdot,1}$	$P_{\cdot,1_{s_1}}$	$P_{\cdot,0_1}$	$P_{\cdot,0_{s_1}}$	$P_{\cdot,0_{s_2}}$	$P_{\cdot,0_2}$	$P_{\cdot,-1_{s_2}}$	$P_{\cdot,-1}$
0.30	0.0596	0.0413	0.3640	0.0351	0.0351	0.3640	0.0413	0.0596
0.50	0.0854	0.0710	0.2867	0.0569	0.0569	0.2867	0.0710	0.0854
0.70	0.1031	0.1009	0.2185	0.0776	0.0776	0.2185	0.1009	0.1031
0.85	0.1120	0.1229	0.1725	0.0925	0.0925	0.1725	0.1229	0.1120
0.95	0.1164	0.1373	0.1441	0.1022	0.1022	0.1441	0.1373	0.1164

Table 6

Stationary phase probabilities for different values of p_1 at $\rho = 0.7$, $\mu_1 = \mu_2 = 3$, $\gamma_1 = \gamma_2 = 3$, $q_1 = 0.5$.

p_1	λ^*	$P_{\cdot,1}$	$P_{\cdot,1_{s_1}}$	$P_{\cdot,0_1}$	$P_{\cdot,0_{s_1}}$	$P_{\cdot,0_{s_2}}$	$P_{\cdot,0_2}$	$P_{\cdot,-1_{s_2}}$	$P_{\cdot,-1}$
0.2	1.459	0.1200	0.1455	0.0901	0.0268	0.1364	0.3444	0.0533	0.0835
0.5	1.491	0.1031	0.1009	0.2185	0.0776	0.0776	0.2185	0.1009	0.1031
0.8	1.459	0.0835	0.0533	0.3444	0.1364	0.0268	0.0901	0.1455	0.1200

Table 7

Stationary phase probabilities for different values of q_1 at $\rho = 0.7$, $\mu_1 = \mu_2 = 3$, $\gamma_1 = \gamma_2 = 3$, $p_1 = 0.5$.

q_1	λ^*	$P_{\cdot,1}$	$P_{\cdot,1_{s_1}}$	$P_{\cdot,0_1}$	$P_{\cdot,0_{s_1}}$	$P_{\cdot,0_{s_2}}$	$P_{\cdot,0_2}$	$P_{\cdot,-1_{s_2}}$	$P_{\cdot,-1}$
0.2	1.494	0.0660	0.0745	0.2081	0.0962	0.0589	0.2298	0.1257	0.1407
0.5	1.491	0.1031	0.1009	0.2185	0.0776	0.0776	0.2185	0.1009	0.1031
0.8	1.494	0.1407	0.1257	0.2298	0.0589	0.0962	0.2081	0.0745	0.0660

The correlation coefficient between the queue lengths also increases with ρ and becomes close to one as the stability boundary is approached. This suggests that the queue lengths vary in a more coordinated way, indicating effective balancing due to the combined use of the JSQ and SLQ disciplines.

The server idle probability, on the contrary, decreases as ρ increases, since the server is less likely to be idle when the arrival rate becomes higher.

Table 2 shows that, in the symmetric case $\mu_1 = \mu_2 = 5$, the mean queue lengths are equal, $\mathbb{E}(L_1) = \mathbb{E}(L_2) = 1.037$, as are the effective load factors, $\rho_1^{\text{eff}} = \rho_2^{\text{eff}} = 0.106$. This is consistent with the symmetry of the model parameters and reflects the symmetric behavior of the two queues.

When the service rates are asymmetric, the load is redistributed between the queues. If the first queue is served more slowly than the second one, its mean length increases. This effect is clearly seen for $\mu_1 = 0.5$, $\mu_2 = 10$, where $\mathbb{E}(L_1) = 1.256$, whereas $\mathbb{E}(L_2) = 1.017$. Along with the increase in the mean queue length, the effective load factor of the first queue also increases to $\rho_1^{\text{eff}} = 0.407$, while the effective load factor of the second queue is much smaller, $\rho_2^{\text{eff}} = 0.033$. Thus, the queue with the lower service rate becomes more heavily loaded.

The values of the correlation coefficient $\text{Cor}(L_1, L_2)$ remain rather high in all the cases considered, which indicates that a positive dependence between the queue lengths is preserved regardless of whether the service rates are symmetric or asymmetric. This effect may be explained by the combined use of the JSQ and SLQ disciplines, which tend to balance the queue lengths by routing new customers to the shortest queue and giving service priority to the longest queue.

It is also important to note how the fraction of time during which the server is in a switch-over state, P_{sw} , changes. In the symmetric case ($\mu_1 = \mu_2 = 5$), it attains the largest value among all the considered cases, $P_{\text{sw}} = 0.505$. This is because, for equal service rates, the queue lengths become equal more often, and the server therefore switches between the queues more frequently. In the asymmetric case, for example, when $\mu_1 = 0.5$ and $\mu_2 = 10$, the fraction of switch-over time decreases to $P_{\text{sw}} = 0.286$, since the server mostly remains at the slower and, consequently, more heavily loaded queue, and switches over less frequently than in the symmetric case.

The idle probability P_{idle} also depends on the combination of service rates, although in the considered cases it remains approximately at the same level, in the range 0.274–0.297. The smallest value, $P_{\text{idle}} = 0.274$, is observed under strong asymmetry, $\mu_1 = 0.5$ and $\mu_2 = 10$. In this case, the first queue is served much more slowly than the second one and therefore becomes more heavily loaded, as reflected by $\rho_1^{\text{eff}} = 0.407$, while $\rho_2^{\text{eff}} = 0.033$. Thus, in the considered examples, changes in the service rates affect the idle probability only moderately.

Table 3 shows that, for small values of γ_1 and γ_2 , the server spends most of its time in a switch-over state. For example, when $\gamma_1 = 0.1$ and $\gamma_2 = 0.2$, we have $P_{\text{sw}} = 0.688$. This is due to the long switch-over process. Moreover, because the system operates under absolute preemptive priority, longer switch-overs increase the probability of arrivals during a switch-over, which further increases P_{sw} , the fraction of time spent in switch-over. As a result, in this case the effective load factors of the queues are small, $\rho_1^{\text{eff}} = 0.025$ and $\rho_2^{\text{eff}} = 0.030$, while the mean queue lengths are relatively large.

As the switch-over rates increase, the fraction of time spent in switch-over generally decreases. For example, it decreases from $P_{\text{sw}} = 0.688$ for $\gamma_1 = 0.1$, $\gamma_2 = 0.2$ to $P_{\text{sw}} = 0.213$ for $\gamma_1 = \gamma_2 = 5$. This indicates that faster switch-overs reduce the time spent in switch-over and allow the server to serve

customers more efficiently. Consequently, the mean queue lengths also decrease: from $\mathbb{E}(L_1) = 1.269$ and $\mathbb{E}(L_2) = 1.174$ for $\gamma_1 = 0.1$, $\gamma_2 = 0.2$ to $\mathbb{E}(L_1) = \mathbb{E}(L_2) = 1.020$ for $\gamma_1 = \gamma_2 = 5$.

When the switch-over rates are asymmetric, the fraction of time spent in switch-over may remain rather high, especially when one of the switch-over rates is much smaller than the other. For example, when $\gamma_1 = 0.5$ and $\gamma_2 = 10$, we have $P_{\text{sw}} = 0.463$, which is considerably higher than in the symmetric case $\gamma_1 = \gamma_2 = 5$, where $P_{\text{sw}} = 0.213$. In this case, since $\gamma_1 < \gamma_2$, switch-overs to Q_1 are slower, and queue Q_1 becomes more heavily loaded in terms of its mean length: $\mathbb{E}(L_1) = 1.228$, whereas $\mathbb{E}(L_2) = 1.027$. The effective load factors are also unequal, $\rho_1^{\text{eff}} = 0.110$ and $\rho_2^{\text{eff}} = 0.165$.

The idle probability P_{idle} changes only moderately as the switch-over rates vary. In the considered cases, it remains in the range 0.257–0.301. The largest value, $P_{\text{idle}} = 0.301$, is observed in the symmetric case $\gamma_1 = \gamma_2 = 5$, where the fraction of time spent in switch-over is the smallest. Thus, in these examples, changes in the switch-over rates mainly affect the fraction of time spent in switch-over, whereas the idle probability varies only slightly.

The correlation coefficient $\text{Cor}(L_1, L_2)$, as in Table 2, remains high, which indicates that the queue lengths vary in a coordinated way. This can similarly be explained by the tendency of the JSQ and SLQ disciplines to balance the queue lengths even in the presence of non-zero switch-over times.

When the queue lengths are equal, the parameters p_1 and q_1 define two different tie-breaking mechanisms. The parameter p_1 controls the routing of arriving customers, whereas q_1 controls the server's choice of the queue to be served. As Table 4 shows, when these parameters differ from 0.5, the load tends to shift toward one of the queues. For example, for $(p_1, q_1) = (0.2, 0.8)$, the second queue is more heavily loaded: $\mathbb{E}(L_2) = 1.127$ and $\rho_2^{\text{eff}} = 0.180$, whereas $\mathbb{E}(L_1) = 0.877$ and $\rho_1^{\text{eff}} = 0.146$. Since the remaining parameters are symmetric, the case $(p_1, q_1) = (0.8, 0.2)$ exhibits mirror symmetry: the same numerical values are interchanged between the first and second queues, namely $\mathbb{E}(L_1) = 1.127$, $\rho_1^{\text{eff}} = 0.180$, $\mathbb{E}(L_2) = 0.877$, and $\rho_2^{\text{eff}} = 0.146$.

The correlation coefficient $\text{Cor}(L_1, L_2)$ remains high and positive in all considered cases. This indicates a coordinated behavior of the queue lengths and can be explained in the same way as in the previous tables.

The mean fraction of time during which the server is in a switch-over state, P_{sw} , also depends on the choice of p_1 and q_1 , but varies within a moderate range. In the considered cases, it ranges from 0.323 to 0.393. The idle probability P_{idle} is less sensitive to changes in these parameters and remains between 0.281 and 0.320.

Table 5 shows that, for a low load value ($\rho = 0.3$), the largest probabilities correspond to the phases 0_1 and 0_2 . This means that the system spends most of the time in a balanced state, where the queue lengths are equal.

As the load increases, the probabilities of the phases 0_1 and 0_2 , where the queue lengths are equal and the server is at one of the queues, decrease. At the same time, the probabilities of the phases 1 and -1 , as well as those of the switch-over phases 1_{s_1} , 0_{s_1} , 0_{s_2} , and -1_{s_2} , increase. This is because, as the arrival rate increases, the system has less time to restore equality between the queue lengths.

It is important to note that, even as the load increases, a substantial part of the probability mass remains concentrated in the phases with $D = 0$. This reflects the effect of the JSQ and SLQ disciplines, which tend to equalize the queue lengths: arriving customers are routed to the shortest queue, while the server gives priority to the longest queue. As a result, the system regularly returns to phases with equal queue lengths.

As the load parameter ρ approaches one, the probability distribution becomes more evenly spread across the phases, while the probability that the server is in a switch-over phase increases. This is due to more frequent changes in the queue-length difference, which cause the server to switch between the queues more often.

It should also be noted that, due to the symmetry of the model parameters, the probabilities of the corresponding phases for the first and second queues coincide.

Table 6 shows that changing the parameter p_1 leads to a redistribution of probabilities among the process phases. For $p_1 = 0.5$, the distribution is symmetric: the probabilities of the corresponding phases for the first and second queues coincide.

When p_1 deviates from the symmetric value, the probability distribution shifts toward one of the queues. For example, for $p_1 = 0.2$, the total probability of the phases associated with the second queue increases, while the first queue becomes less loaded. For $p_1 = 0.8$, the opposite pattern is observed: the distribution shifts toward the first queue, while the second queue becomes less loaded.

It can also be seen that changing p_1 affects the probabilities of the switch-over phases. As p_1 increases, the probabilities of the phases 0_{s_1} and -1_{s_2} become larger, whereas the probabilities of the phases 1_{s_1} and 0_{s_2} decrease. When p_1 decreases, the opposite pattern is observed: the probabilities of the phases 0_{s_1} and -1_{s_2} become smaller, whereas the probabilities of the phases 1_{s_1} and 0_{s_2} increase.

Thus, the parameter p_1 has a significant effect on the distribution of phase probabilities and determines the direction of the load shift in the system.

Table 7 shows that the parameter q_1 affects the distribution of probabilities over the process phases, although this effect is weaker than that of the parameter p_1 in Table 6.

For $q_1 = 0.5$, as in Table 6, the distribution is symmetric: the probabilities of the corresponding phases for the first and second queues coincide. When q_1 differs from 0.5, the probability mass shifts toward one of the queues. For $q_1 = 0.2$, the system is more often in phases associated with the second queue, whereas for $q_1 = 0.8$, the shift is toward the first queue. These changes are most noticeable in the phases with unequal queue lengths.

Thus, compared with p_1 , the parameter q_1 plays a secondary role in the load distribution between the queues.

8. Conclusion

This paper studies a two-queue queueing system with the JSQ–SLQ disciplines, absolute preemptive priority, and non-zero server switch-over times. A Markov model is constructed in terms of the length of the first queue and a phase variable that encodes both the queue-length difference and the server state. This makes it possible to apply the probability generating function method and to obtain analytical expressions for the main stationary characteristics, as well as a condition for the existence of a stationary regime. The obtained analytical results were used to numerically evaluate how the arrival rate, service rates, server switch-over rates, and tie-breaking parameters affect the stationary characteristics and phase probabilities.

The numerical results show that the behavior of the system is shaped by its load, service rates, tie-breaking parameters, and server switch-over times. As the load increases, the mean queue lengths and their variances grow rapidly, which is related to the arrival rate approaching the stability boundary. At the same time, the positive correlation between the queue lengths increases, reflecting the balancing effect of the jointly applied JSQ and SLQ disciplines.

It is shown that the service rates noticeably affect the load distribution between the queues. The queue with the lower service rate becomes more heavily loaded, which is reflected in an increase in its mean length and effective load factor. However, in the considered examples, the idle probability changes only moderately as the service rates vary. The server switch-over times also have a significant effect on the system characteristics. When switching is slow, the server spends a substantial fraction of time in switch-over, which leads to larger mean queue lengths. Faster switching, in contrast, reduces the time spent in switch-over and improves service efficiency.

The tie-breaking parameters also affect the system behavior. The parameter p_1 plays an important role in shifting the load between the queues and noticeably affects the distribution of phase probabilities. The parameter q_1 also affects the phase probabilities, but its effect is weaker than that of p_1 , and it plays a secondary role in the load distribution between the queues.

Overall, the combined use of the JSQ and SLQ disciplines provides stable equalization of queue lengths even in the presence of parameter asymmetry. However, the system performance is sensitive to the combination of the routing, service, and switch-over parameters.

Further research may be directed toward generalizing the proposed model to systems with a larger number of queues and servers. Such a generalization would make it possible to analyze multi-server systems with a complex interaction structure, which is typical of distributed computing and telecommunication systems. In addition, it is of particular interest to study other rules for selecting a queue when the queue lengths are equal, as well as systems with different priority mechanisms, including preemptive and non-preemptive variants.

Another relevant direction is the study of systems with more general distributions of service and server switch-over times, as well as non-Poisson arrival processes. This would make it possible to go beyond the Markovian assumptions and extend the applicability of the model.

The model considered in this paper can be used to analyze and optimize queueing systems in which the routing of arriving customers and the selection of a queue for service depend on the current state of the system, for example, in telecommunication networks, data centers, and distributed computing systems.

Author Contributions: Conceptualization, Evgeny A. Khityaev, Ivan S. Zaryadov and Tatiana A. Milovanova; methodology, Evgeny A. Khityaev, Ivan S. Zaryadov and Tatiana A. Milovanova; software, Evgeny A. Khityaev; validation, Evgeny A. Khityaev and Ivan S. Zaryadov; formal analysis, Evgeny A. Khityaev, Ivan S. Zaryadov and Tatiana A. Milovanova; investigation, Evgeny A. Khityaev and Ivan S. Zaryadov; resources, Evgeny A. Khityaev and Ivan S. Zaryadov; data curation, Evgeny A. Khityaev and Ivan S. Zaryadov; writing—original draft preparation, Evgeny A. Khityaev; writing—review and editing, Evgeny A. Khityaev and Ivan S. Zaryadov; visualization, Evgeny A. Khityaev and Ivan S. Zaryadov; supervision, Ivan S. Zaryadov; project administration, Ivan S. Zaryadov. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No new data were created or analysed during this study. Data sharing is not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Declaration on Generative AI: During the preparation of this work, the authors used ChatGPT-5 for text translation and language editing, as well as for auxiliary assistance with selected mathematical formulations and computational aspects of the study. All AI-assisted materials were reviewed, verified, and edited as needed by the authors, who take full responsibility for the content of the publication.

References

- [1] M. A. A. Boon, R. D. Van der Mei, and E. M. M. Winands, “Applications of Polling Systems,” *Surveys in Operations Research and Management Science*, vol. 16, no. 2, pp. 67–82, 2011. DOI: 10.1016/j.sorms.2011.01.001
- [2] S. Bodas, S. Shakkottai, L. Ying, and R. Srikant, “Scheduling in Multi-Channel Wireless Networks: Rate Function Optimality in the Small-Buffer Regime,” *IEEE Transactions on Information Theory*, vol. 60, no. 2, pp. 1101–1125, 2014. DOI: 10.1109/TIT.2013.2293216
- [3] V. Gupta, M. Harchol-Balter, K. Sigman, and W. Whitt, “Analysis of Join-the-Shortest-Queue Routing for Web Server Farms,” *Performance Evaluation*, vol. 64, no. 9–12, pp. 1062–1081, 2007. DOI: 10.1016/j.peva.2007.06.012
- [4] Y. Lu, Q. Xie, G. Kliot, A. Geller, J. R. Larus, and A. Greenberg, “Join-Idle-Queue: A Novel Load Balancing Algorithm for Dynamically Scalable Web Services,” *Performance Evaluation*, vol. 68, no. 11, pp. 1056–1071, 2011. DOI: 10.1016/j.peva.2011.07.015
- [5] M. Van der Boor, S. C. Borst, J. S. H. Van Leeuwen, and D. Mukherjee, “Scalable Load Balancing in Networked Systems: A Survey of Recent Advances,” *SIAM Review*, vol. 64, no. 3, pp. 554–622, 2022. DOI: 10.1137/20M1323746
- [6] F. A. Haight, “Two Queues in Parallel,” *Biometrika*, vol. 45, no. 3–4, pp. 401–410, 1958. DOI: 10.1093/biomet/45.3-4.401
- [7] W. Winston, “Optimality of the Shortest Line Discipline,” *Journal of Applied Probability*, vol. 14, no. 1, pp. 181–189, 1977. DOI: 10.2307/3213271

- [8] A. Hordijk and G. Koole, “On the Optimality of the Generalized Shortest Queue Policy,” *Probability in the Engineering and Informational Sciences*, vol. 4, no. 4, pp. 477–487, 1990. DOI: 10.1017/S0269964800001777
- [9] S. Halfin, “The Shortest Queue Problem,” *Journal of Applied Probability*, vol. 22, no. 4, pp. 865–878, 1985. DOI: 10.2307/3213954
- [10] I. J. B. F. Adan, J. Wessels, and W. H. M. Zijm, “Analysis of the Asymmetric Shortest Queue Problem,” *Queueing Systems*, vol. 8, no. 1, pp. 1–58, 1991. DOI: 10.1007/BF02412240
- [11] J. W. Cohen, “Analysis of the Asymmetrical Shortest Two-Server Queueing Model,” *International Journal of Stochastic Analysis*, vol. 11, no. 2, pp. 115–162, 1998. DOI: 10.1155/S1048953398000112
- [12] G.-J. Van Houtum, I. Adan, J. Wessels, and W. H. M. Zijm, “Performance Analysis of Parallel Identical Machines with a Generalized Shortest Queue Arrival Mechanism,” *OR Spektrum*, vol. 23, no. 4, pp. 411–427, 2001. DOI: 10.1007/PL00013360
- [13] H. Yao and C. Knessl, “On the Infinite Server Shortest Queue Problem: Symmetric Case,” *Stochastic Models*, vol. 21, no. 1, pp. 101–132, 2005. DOI: 10.1081/STM-200046493
- [14] H. Yao and C. Knessl, “On the Infinite Server Shortest Queue Problem: Non-Symmetric Case,” *Queueing Systems*, vol. 52, no. 2, pp. 157–177, 2006. DOI: 10.1007/s11134-006-5500-z
- [15] I. J. B. F. Adan, O. J. Boxma, S. Kapodistria, and V. G. Kulkarni, “The Shorter Queue Polling Model,” *Annals of Operations Research*, vol. 241, no. 1, pp. 167–200, 2016. DOI: 10.1007/s10479-013-1495-0
- [16] P. Eschenfeldt and D. Gamarnik, “Join the Shortest Queue with Many Servers. The Heavy-Traffic Asymptotics,” *Mathematics of Operations Research*, vol. 43, no. 3, pp. 867–886, 2018. DOI: 10.1287/moor.2017.0887
- [17] D. A. Dawson, J. Tang, and Y. Q. Zhao, “Performance Analysis of Joining the Shortest Queue Model Among a Large Number of Queues,” *Asia-Pacific Journal of Operational Research*, vol. 36, no. 4, p. 1950019, 2019. DOI: 10.1142/S0217595919500192
- [18] A. Braverman, “Steady-State Analysis of the Join-the-Shortest-Queue Model in the Halfin–Whitt Regime,” *Mathematics of Operations Research*, vol. 45, no. 3, pp. 1069–1103, 2020. DOI: 10.1287/moor.2019.1023
- [19] I. Dimitriou, “Analysis of the Symmetric Join the Shortest Orbit Queue,” *Operations Research Letters*, vol. 49, no. 1, pp. 23–29, 2021. DOI: 10.1016/j.orl.2020.10.011
- [20] W. Weng and W. Wang, “Achieving Zero Asymptotic Queueing Delay for Parallel Jobs,” *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, vol. 4, no. 3, pp. 1–36, 2020. DOI: 10.1145/3428327
- [21] J. W. Cohen, “A Two-Queue, One-Server Model with Priority for the Longer Queue,” *Queueing Systems*, vol. 2, no. 3, pp. 261–283, 1987. DOI: 10.1007/BF01158902
- [22] L. Flatto, “The Longer Queue Model,” *Probability in the Engineering and Informational Sciences*, vol. 3, no. 4, pp. 537–559, 1989. DOI: 10.1017/S0269964800001376
- [23] G.-J. Van Houtum, I. Adan, and J. Van der Wal, “The Symmetric Longest Queue System,” *Stochastic Models*, vol. 13, no. 1, pp. 105–120, 1997. DOI: 10.1080/15326349708807416
- [24] C. Knessl and H. Yao, “On the Nonsymmetric Longer Queue Model: Joint Distribution, Asymptotic Properties, and Heavy Traffic Limits,” *Advances in Operations Research*, vol. 2013, no. 1, p. 680 539, 2013. DOI: 10.1155/2013/680539
- [25] R. Pedarsani and J. Walrand, “Stability of Multiclass Queueing Networks under Longest-Queue and Longest-Dominating-Queue Scheduling,” *Journal of Applied Probability*, vol. 53, no. 2, pp. 421–433, 2016. DOI: 10.1017/jpr.2016.10
- [26] R. Atar and S. Saha, “A Note on Non-Existence of Diffusion Limits for Serve-the-Longest-Queue when the Buffers Are Equal in Size,” *Electronic Communications in Probability*, vol. 21, pp. 1–10, 2016. DOI: 10.1214/16-ECP4370
- [27] E. Perel, N. Perel, and U. Yechiali, “A Polling System with ‘Join the Shortest–Serve the Longest’ Policy,” *Computers & Operations Research*, vol. 114, p. 104809, 2020. DOI: 10.1016/j.cor.2019.104809
- [28] E. Perel, N. Perel, and U. Yechiali, “The Non-Preemptive ‘Join the Shortest Queue–Serve the Longest Queue’ Service System with or without Switch-Over Times,” *Mathematical Methods of Operations Research*, vol. 99, no. 1, pp. 3–38, 2024. DOI: 10.1007/s00186-023-00848-7

- [29] E. Perel, N. Perel, and U. Yechiali, “A 3-Queue Polling System with Join the Shortest–Serve the Longest Policy,” *Indagationes Mathematicae*, vol. 34, no. 5, pp. 1101–1120, 2023. DOI: 10.1016/j.indag.2022.11.001

Information about the authors

Khityaev, Evgeny A.—Postgraduate Student of the Department of Probability Theory and Cyber Security of RUDN University (e-mail: evgenykhityaev@yandex.ru, ORCID: 0000-0002-2375-1873, ResearcherID: PYW-1479-2026)

Zaryadov, Ivan S.—Candidate of Physical and Mathematical Sciences, Assistant Professor of the Department of Probability Theory and Cyber Security of RUDN University; Senior Researcher of the Institute of Informatics Problems of the Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences (e-mail: zaryadov-is@rudn.ru, ORCID: 0000-0002-7909-6396, ResearcherID: B-8154-2018, Scopus Author ID: 35294470000)

Milovanova, Tatiana A.—Candidate of Physical and Mathematical Sciences, Assistant Professor of the Department of Probability Theory and Cyber Security of RUDN University (e-mail: milovanova-ta@rudn.ru, ORCID: 0000-0002-9388-9499, ResearcherID: X-6718-2018)

RUSSIAN METADATA

УДК 519.872, 519.217

PACS 02.50.Ey, 02.50.Ga, 07.05.Tp

DOI: 10.22363/2658-4670-2026-34-3-316-333

EDN: RHUTHS

Двухочередная система массового обслуживания с дисциплинами JSQ–SLQ, абсолютным приоритетом и ненулевыми временами перехода прибора

Е. А. Хитяев¹, И. С. Зарядов^{1,2}, Т. А. Милованова¹

¹ Российский университет дружбы народов, ул. Миклухо-Маклая, д. 6, Москва, 117198, Российская Федерация

² Институт проблем информатики, Федеральный исследовательский центр «Информатика и управление» РАН, ул. Вавилова, д. 44, кор. 2, Москва, 119333, Российская Федерация

Аннотация. Цель. Исследовать двухочередную систему массового обслуживания с одним прибором, функционирующую в соответствии с дисциплинами выбора кратчайшей очереди и обслуживания длиннейшей очереди, с вероятностным выбором очереди для поступления и обслуживания при равенстве их длин, прерываемым обслуживанием по абсолютному приоритету и ненулевыми временами перехода прибора между очередями. Методы. Для анализа системы используются марковское описание, уравнения глобального баланса для стационарных вероятностей, метод условных производящих функций и численный анализ. Результаты. Построен двумерный марковский процесс, компонентами которого являются длина первой очереди и фазовая переменная, отражающая разность длин очередей и состояние прибора. Для построенного процесса представлена система уравнений глобального баланса для стационарных вероятностей. Введены условные производящие функции, с помощью которых получена матричная форма уравнений баланса и рассмотрено условие существования стационарного режима. Получены выражения для основных стационарных характеристик системы, включая средние длины очередей, дисперсии, коэффициент корреляции, эффективные интенсивности поступления заявок, долю времени, в течение которого прибор находится в состоянии перехода, и вероятность его простоя. Проведён численный анализ влияния интенсивности поступления заявок, интенсивностей обслуживания, интенсивностей перехода прибора и вероятностей выбора очереди при равенстве длин на характеристики системы и распределение фаз процесса. Выводы. Показано, что совместное использование дисциплин выбора кратчайшей очереди и обслуживания длиннейшей очереди способствует выравниванию длин очередей, однако поведение системы и её стационарные характеристики существенно зависят от сочетания параметров маршрутизации, обслуживания, переходов прибора и вероятностей выбора очереди при равенстве длин.

Ключевые слова: система массового обслуживания, выбор кратчайшей очереди, обслуживание длиннейшей очереди, абсолютный приоритет, переход прибора, марковский процесс, условные производящие функции