THE WAITING TIME ANALYSIS OF MULTI-SERVER QUEUE WITH CONSTANT RETRIAL RATE AND DIFFERENT CONTROL POLICIES

by

Dmitry Efrosinin and Janos Sztrik

Number 564 November 2007
THE WAITING TIME ANALYSIS OF MULTI-SERVER QUEUE WITH CONSTANT RETRIAL RATE AND DIFFERENT CONTROL POLICIES

Efrosinin, Dmitry

Johannes Kepler University of Linz, Altenberger Str., 69,4040 Linz, Austria
[dmitry.efrosinin@jku.at]

Sztrik, Janos

University of Debrecen, P.O. Box 12, 4010 Debrecen, Hungary
[jsztrik@inf.unideb.hu]

Abstract

A controlled retrial queueing system with two heterogeneous servers, first-come first-served discipline for the customers on the orbit, possible direct access for the primary customers, Poisson arrivals and finite retrial group is considered. According to the optimal policy the faster server must be active whenever the customer enters the service area while the slower one can be active only when the number of customers in the orbit exceeds a prescribed threshold level. The problem of deriving the equilibrium waiting time and sojourn time distributions is formulated. We introduce the recursive method for the calculation of the corresponding Laplace-Stiltjes transforms and the inversion methods are used to get the distribution functions. Also the recursive method for the obtaining z-transforms is used to get some discrete distributions for the number of directly served primary customers and the number of retrials made by a customer until the service starts. The methods are applied also for other heuristic control policies, namely for the Scheduling threshold policy (STP), Fastest Free Server (FFS) or Random Server Selection (RSS)

Keywords: Retrial queue, Controlled queueing system, Steady-state probabilities, Threshold control policy, Waiting time distribution, Sojourn time distribution

1 Introduction

We consider controllable retrial queueing system with several exponential servers, functioning at different rates. The arriving customers form a Poisson process and can have a direct access to the service area. In our previous paper [8] we minimized the expected sojourn time over all customers and found that there is a threshold policy which uses a slow server only if the orbit size exceeds a certain threshold level. With respect to this policy we have applied the matrix-geometric method for the calculation of steady-states probabilities and different mean performance characteristics.

In this paper we formulate the problem of determining the stationary waiting and sojourn time distributions. For the uncontrolled classical queueing systems without retrials for the deriving of the waiting time distribution it is enough to consider the system state at the arrival time of a tagged customer, e.g. as is presented by Kleinrock [11]. In controlled case [20] it was shown that the waiting and sojourn time distributions correspond to the linear combination of Erlang distributions. The waiting time in uncontrolled retrial queues is more difficult. Some methods are described in the monograph by Falin and Tempelton [9]. For retrial systems (uncontrolled or controlled) with direct access of primary customers to the service area the waiting time analysis becomes more complicated because the waiting time of a customer in this case depends also on future arrivals. In systems with classical retrial policy, where the retrial rate depends on the number of orbiting customers, it is necessary to take into account that some later arrived customer can

---

1The investigation is supported by the Austro-Hungarian Cooperation Grant No 660:u1, 2006.
be earlier served according to the random order policy for the orbiting customers. The recursive scheme for the computation of the Laplace-Stiltjes transforms of the waiting time of a tagged customer is introduced by Artalejo et al. [4]. This paper motivate us to perform the waiting time analysis for the controlled queue with constant retrial policy, where the retrial rate is independent of the number of orbiting customers.

In systems with constant retrial rate with FCFS service discipline for the orbiting customers for a threshold policy future arrivals can influence the waiting time of the tagged customer by influencing the servers that can be active in the future. Therefore it is also necessary to consider the arrival process after the arrival of the tagged customer up to the time where service of this customer starts. For the system under consideration the calculation of the waiting time distribution is achieved by analysing the transient Markov process with absorption at the time the tagged customer starts service. This requires an extension of the state representation since we have to know at each time the position of the tagged customer in the list of orbiting customers.

In this paper we express the Laplace transforms of the waiting and sojourn time distributions. Representing the Laplace-Stiltjes transforms of the waiting or sojourn time densities of the tagged customer as vector we get that for threshold system it satisfies the threshold depended block-three-diagonal system. For uncontrolled queue with classical retrial rate such a structure was shown in [4]. But in case of FIFO discipline for the orbiting customers this system is recurrent with respect to the position of the tagged customer in the orbit.

The organization of the paper is as follows. In sections 2 and 3, the steady-state distributions are derived for a general threshold system and heuristic control policies. In sections 4 and 5, we develop recursive equations for the calculation of the stationary waiting and sojourn time distributions for a system under threshold and heuristic control policies as well as the corresponding moments of the arbitrary orders. In sections 6 and 7, we obtain, respectively, discrete distribution functions for the number of directly served customers and number of retrials made by a customer. In section 8 we give the results of numerical inversion of Laplace- and z-transforms and compare them for threshold and heuristic control policies.

In further sections we will use the notations \(e(n)\), \(e_j(n)\) and \(I_n\) for the column-vector of dimension \(n\) consisting of 1’s, column vector of dimension \(n\) with 1 in the \(j\)-th (beginning from 0-th) position and 0 elsewhere, and an identity matrix of dimension \(n \times n\). The notations without specifying the dimensions will be used for the suitably dimensioned vectors and matrices.

## 2 Description of the mathematical model

Consider the queueing model \(M/M/c\) in which primary customers arrive according to a Poisson stream with rate \(\lambda\), two heterogeneous exponential servers \(c = 2\) with rates \(\mu_1 > \mu_2\), constant retrial rate \(\gamma > 0\) and the number of places in the retrial orbit \(2 \leq K \leq \infty\). According to the control policy an arriving customer can join the orbit or have direct access to the accessible idle servers. The arrival process, service times and retrial times are assumed to be mutually independent.

Let \(Q(t)\) is the number of customers in the retrial orbit at time \(t\), \(D_1(t), D_2(t)\) describe the states of the servers at this time,

\[
D_j(t) = \begin{cases} 
0, & \text{if the } j\text{-th server is idle at time } t \text{ and} \\
1, & \text{if the } j\text{-th server is busy.} 
\end{cases}
\]

The observed process

\[
\{X(t)\}_{t \geq 0} = \{Q(t), D_1(t), D_2(t)\}_{t \geq 0}
\]

is a continuous-time Markov process with state space defined as

\[
E = \{x = (q, d_1, d_2); 0 \leq q \leq K, d_i = \{0, 1\}, i = 1, 2\} = \mathbb{N} \times \{0, 1\}^2,
\]

where \(q\) and \(d_i, i = 1, 2\) denote, respectively, the number of customers on the orbit and states of the servers.
Theorem 1 The optimal policy for the retrial queueing system $M/M/2$ with heterogeneous servers and constant retrial rate is of threshold and monotone type, i.e. the fastest idle server must be switched on whenever a primary or retrial customer arrives and another one must be switched on if and only if the fastest server is busy and the orbit length reaches the threshold level $q \geq q^*_2$.

The analytical representation of the threshold level is quite complicated, but by means of the value iteration algorithm it can be done numerically. If future arrivals are not taken into account (scheduling problem), i.e. when the objective is to minimize the sojourn time for an individual customer, the explicit solution for the threshold level exists

\begin{align*}
q^*_2 &= \left\lfloor \frac{\gamma}{\mu_1 + \gamma \left( \frac{\mu_1}{\mu_2} - 1 \right)} \right\rfloor + 1,
\end{align*}

such that if $q \geq q^*_2$ in state $x = (q, 1, 0)$ then upon retrial arrival it is optimal to dispatch a customer to the slower server, if $q \leq q^*_2 - 1$ then the slower server must be idle.

3 Steady-state distribution of the system under optimal policy

In this section we derive the equilibrium state distribution under the optimal threshold policy (OTP). The derivation works via a standard matrix-geometric approach [13], taking into account the special structure of the boundary states where not all servers are active. To distinguish the system under OTP from other control policies which will be discussed later we will use the upper index $(1)'$ for the concerned values. Let $q^*_2$ be the threshold level for activation of the second server. As it was mentioned above they can be found numerically (e.g. using the value iteration algorithm [10]).

Consider a Markov process $\{X(t)\}_{t \geq 0}$ defined by (22) with a state space $E$. This process is a QBD process with block - three-diagonal infinitesimal matrix. Note that the blocks have different sizes depending on the queue length.

3.1 Finite retrial group

First consider the system with finite retrial group, i.e. $K < \infty$. In this case the states are partitioned as follows:

- block 0 includes the single state: $(0, 0, 0)$;
- block 1 includes the states: $(0, 0, 1), (0, 1, 0), (1, 0, 0)$;
- block 2 includes the states: $(0, 1, 1), (1, 0, 1), (1, 1, 0), (2, 0, 0)$;
- blocks $i, \ 3 \leq i \leq K$ include the states:
  $(i - 2, 1, 1), (i - 1, 0, 1), (i - 1, 1, 0), (i, 0, 0)$;
- block $K + 1$ includes the states:
  $(K - 1, 1, 1), (K, 0, 1), (K, 1, 0)$;
- block $K + 2$ includes the single state: $(K, 1, 1)$;
Denote by $\Lambda^{(1)}$ the infinitesimal matrix of dimension $4(K + 1)$ for the system under OTP,

$$
\Lambda^{(1)} = \begin{pmatrix}
-\lambda & A_1^{(1)} & 0 & 0 & 0 & \ldots & 0 \\
D_0^{(1)} & -(C_1^{(1)} - B_1^{(1)}) & 0 & 0 & 0 & \ldots & 0 \\
0 & D_1^{(1)} & -(C_2^{(1)} - B_2^{(1)}) & A_2^{(1)} & 0 & \ldots & 0 \\
0 & 0 & D_2^{(1)} & -(C_3^{(1)} - B_3^{(1)}) & A_3^{(1)} & \ldots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & \ldots & 0 & D_3^{(1)} & -(C_4^{(1)} - B_4^{(1)}) & \ldots & A_4^{(1)} \\
0 & \ldots & 0 & 0 & D_4^{(1)} & \ldots & \ldots \\
0 & \ldots & 0 & 0 & 0 & \ldots & \ldots \\
0 & \ldots & 0 & 0 & 0 & \ldots & \ldots \\
0 & \ldots & 0 & 0 & 0 & \ldots & \ldots \\
\end{pmatrix}
$$

(3)

where

$$
\lambda + A_1^{(1)} e = D_0^{(1)} e - (C_1^{(1)} - B_1^{(1)}) e + A_2^{(1)} e = D_1^{(1)} e - (C_2^{(1)} - B_2^{(1)}) e + A_3^{(1)} e =
$$

$$
D_2^{(1)} e - (C_3^{(1)} - B_3^{(1)}) e + A_4^{(1)} e = D_3^{(1)} e - (C_4^{(1)} - B_4^{(1)}) e + A_5^{(1)} e =
$$

$$
D_4^{(1)} e - M = 0,
$$

$M = \mu_1 + \mu_2$. Matrices $A_i^{(1)}$ and $B_i^{(1)}$ represent primary and retrial arrivals, respectively, depending on whether the queue length are above or below threshold level:

$$
A_1^{(1)} = \begin{pmatrix} 0 & \lambda & 0 \end{pmatrix},
A_2^{(1)} = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix},
A_3^{(1)} = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix},
A_4^{(1)} = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix},
A_5^{(1)} = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix},
$$

$$
A_6^{(1)} = \begin{pmatrix} \lambda \\ \lambda \end{pmatrix}
$$

and

$$
B_1^{(1)} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},
B_2^{(1)} = \begin{pmatrix} \gamma & 0 & 0 \\ 0 & \gamma & 0 \\ 0 & 0 & \gamma \\ 0 & 0 & \gamma \end{pmatrix},
B_3^{(1)} = \begin{pmatrix} \gamma & 0 & 0 \\ 0 & \gamma & 0 \\ 0 & 0 & \gamma \\ 0 & 0 & \gamma \end{pmatrix},
B_4^{(1)} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \gamma & 0 & 0 \\ \gamma & 0 & 0 \end{pmatrix}
$$

Matrices $C_i^{(1)}$ represent cases when the system stays at certain states:

$$
C_1^{(1)} = \begin{pmatrix} \lambda + \mu_2 & 0 & 0 \\ 0 & \lambda + \mu_1 & 0 \\ 0 & 0 & \lambda + \gamma \end{pmatrix},
C_2^{(1)} = \begin{pmatrix} \lambda + M & 0 & 0 \\ 0 & \lambda + \mu_2 + \gamma & 0 \\ 0 & 0 & \lambda + \mu_1 + \gamma \end{pmatrix},
$$

$$
C_3^{(1)} = \begin{pmatrix} \lambda + M & 0 & 0 \\ 0 & \lambda + \mu_2 + \gamma & 0 \\ 0 & 0 & \lambda + \mu_1 + \gamma \end{pmatrix},
C_4^{(1)} = \begin{pmatrix} \lambda + M & 0 & 0 \\ 0 & \lambda + \mu_2 + \gamma & 0 \\ 0 & 0 & \lambda + \mu_1 + \gamma \end{pmatrix}.
$$

Matrices $D_i^{(1)}$ represent departures with elements depending on active servers:

$$
D_0^{(1)} = \begin{pmatrix} \mu_2 & \mu_1 \\ \mu_2 & \mu_1 \end{pmatrix},
D_1^{(1)} = \begin{pmatrix} \mu_1 & \mu_2 \\ 0 & \mu_2 \end{pmatrix},
D_2^{(1)} = \begin{pmatrix} 0 & \mu_1 & \mu_2 \\ 0 & 0 & \mu_2 \\ 0 & 0 & \mu_1 \end{pmatrix},
D_3^{(1)} = \begin{pmatrix} 0 & \mu_1 & \mu_2 \\ 0 & 0 & \mu_2 \\ 0 & 0 & \mu_1 \end{pmatrix},
D_4^{(1)} = \begin{pmatrix} 0 & \mu_1 & \mu_2 \\ 0 & 0 & \mu_2 \\ 0 & 0 & \mu_1 \end{pmatrix}.
$$

4
Denote by \( \pi^{(1)} = (\pi_0^{(1)}, \pi_1^{(1)}, \ldots) \) the row-vector of the steady-state probabilities,
\[
\pi^{(1)} = \{ \pi_x^{(1)} = \pi_{(q,d_1,d_2)}^{(1)} : x \in E \} = \lim_{t \to \infty} P\{X(t) = x\},
\]
by \( \{ \pi_k^{(1)} : k \geq 0 \} \) — the subvectors that specify the states with \( k \) jobs in the system.

Obviously the row-vector \( \pi^{(1)} \) of the steady-state probabilities of the system under optimal policy satisfies the equations
\[
\pi^{(1)} \Lambda = 0, \pi^{(1)} e = 1. \tag{4}
\]
The probabilities for the system states can be represented in the form of recursive relation with some matrices \( M_k^{(1)} \),
\[
\pi_k^{(1)} = \pi_{k+1}^{(1)} M_k^{(1)}, \quad k = 0, 1, \ldots, K + 1,
\]
nevertheless
\[
\pi_0^{(1)} = \pi_1^{(1)} D_0^{(1)} \frac{1}{\lambda} = \pi_1^{(1)} M_0^{(1)}; \\
\pi_1^{(1)} = \pi_2^{(1)} D_1^{(1)} (C_1^{(1)} - B_1^{(1)} - D_0^{(1)} \frac{1}{\lambda} A_1^{(1)})^{-1} = \pi_2^{(1)} M_1^{(1)}; \\
\pi_2^{(1)} = \pi_3^{(1)} D_2^{(1)} (C_2^{(1)} - B_2^{(1)} - M_1^{(1)} A_2^{(1)})^{-1} = \pi_3^{(1)} M_2^{(1)}; \\
\pi_3^{(1)} = \pi_4^{(1)} D_3^{(1)} (C_3^{(1)} - B_3^{(1)} - M_2^{(1)} A_3^{(1)})^{-1} = \pi_4^{(1)} M_3^{(1)}; \\
\pi_i^{(1)} = \pi_{i+1}^{(1)} D_i^{(1)} (C_i^{(1)} - B_i^{(1)} - M_{i-1}^{(1)} A_i^{(1)})^{-1} = \pi_{i+1}^{(1)} M_i^{(1)}, \quad 4 \leq i \leq q^*_2; \\
\pi_{i}^{(1)} = \pi_{i+1}^{(1)} D_i^{(1)} (C_i^{(1)} - B_i^{(1)} - M_{i-1}^{(1)} A_i^{(1)})^{-1} = \pi_{i+1}^{(1)} M_i^{(1)}, \quad q^*_2 + 1 \leq i \leq K - 1; \\
\pi_{K}^{(1)} = \pi_{K+1}^{(1)} D_K^{(1)} (C_K^{(1)} - B_K^{(1)} - M_{K-1}^{(1)} A_K^{(1)})^{-1} = \pi_{K+1}^{(1)} M_N^{(1)}; \\
\pi_{K+1}^{(1)} = \pi_{K+2}^{(1)} D_{K+1}^{(1)} (C_{K+1}^{(1)} - B_{K+1}^{(1)} - M_{K}^{(1)} A_{K+1}^{(1)})^{-1} = \pi_{K+2}^{(1)} M_{K+1}^{(1)}. \tag{5}
\]

To calculate the values \( \pi_k^{(1)} \) it is necessary:

**Step 1.** Evaluate the matrices \( M_k^{(1)} \), \( k = 0, \ldots, K + 1 \) according to relation (5).

**Step 2.** Evaluate the value \( \pi_{K+2}^{(1)} = \pi_{(K,1,1)}^{(1)} \) from the normalization condition
\[
1 = \sum_{k=0}^{K+2} \pi_k^{(1)} e = \pi_{K+2}^{(1)} + \pi_{K+2}^{(1)} \prod_{i=0}^{K+1} M_j^{(1)} e = \pi_{K+2}^{(1)} \left[ 1 + \sum_{i=0}^{K+1} \prod_{j=K+1-i}^{K+1} M_j^{(1)} e \right]. \tag{6}
\]

**Step 3.** Substitute \( \pi_{K+2}^{(1)} \) into
\[
\pi_i^{(1)} = \pi_{K+2}^{(1)} \prod_{j=K+1-i}^{K+1} M_j^{(1)}, \quad i = 0, \ldots, K + 1.
\]

Note that this method works efficiently as long as \( K < \infty \) is not too large. But for large \( K \) the matrix geometric solution corresponding to \( K = \infty \) is a good approximation.

### 3.2 Infinite retrieval group

In case of infinite retrieval group \( K = \infty \) the infinitesimal matrix \( \Lambda^{(1)} \) has infinite size and is obtained from the matrix (3) by removing the last three rows.

We consider first the matrix-geometric part of the equations, above the threshold level \( q^*_2 \):\[
\pi_{q^*_2 + j}^{(1)} A_4^{(1)} + \pi_{q^*_2+j+2}^{(1)} D_2^{(1)} = \pi_{q^*_2+j+1}^{(1)} (C_3^{(1)} - B_3^{(1)}), \quad j \geq 0.
\]
Conjecturing the matrix-geometric form
\[ \pi_{q_2^i}^{(1)} = \pi_{q_2^i}^{(1)}(R^{(1)})^j \]
we substitute this guess into the last equation, then we get the following equation for matrix \( R^{(1)} \)
\[
(R^{(1)})^2 D_2^{(1)} - R^{(1)} (C_3^{(1)} - B_3^{(1)}) + A_4^{(1)} = 0. \tag{7}
\]
This is a quadratic equation in the matrix \( R^{(1)} \), which is typically solved numerically using the following iteration procedure:
\[
R^{(1)}(0) = 0, \tag{8}
R^{(1)}(n + 1) = A_4^{(1)}(C_3^{(1)} - B_3^{(1)})^{-1} + (R^{(1)})^2(n)D_2^{(1)}(C_3^{(1)} - B_3^{(1)})^{-1},
\]
where the iteration halts when entries in \( R^{(1)}(n + 1) \) and \( R^{(1)}(n) \) differ in absolute value by less that a given small constant.

The general theory [13] states that the necessary and sufficient condition for stability is
\[
p^{(1)} D_2^{(1)} e > p^{(1)} A_4^{(1)} e,
\]
where \( p^{(1)} = (p_0^{(1)}, p_1^{(1)}, p_2^{(1)}, p_3^{(1)}) \) is a stationary probability vector given by \( p^{(1)}(A_4^{(1)} - (C_3^{(1)} - B_3^{(1)}) + D_2^{(1)}) = 0, p^{(1)} e = 1. \)

**Theorem 3** For the system under optimal policy, the stationary vector \( p^{(1)} \) of \( A_4^{(1)} - (C_3^{(1)} - B_3^{(1)}) + D_2^{(1)} \) is given by
\[
p_0^{(1)} = \frac{(\lambda + \gamma)^2(\lambda + \mu_2 + \gamma)}{(\lambda + \mu_1 + \gamma)(\lambda + \gamma)(\lambda + 2\mu_1 + \gamma) + \mu_2 M},
p_1^{(1)} = \frac{\mu_1}{\lambda + \mu_2 + \gamma} p_0^{(1)},
p_2^{(1)} = \frac{\mu_2(\lambda + M + \gamma)}{(\lambda + \gamma)(\lambda + \mu_2 + \gamma)} p_0^{(1)},
p_3^{(1)} = \frac{\mu_1\mu_2(2(\lambda + \gamma) + M)}{(\lambda + \gamma)^2(\lambda + \mu_2 + \gamma)} p_0^{(1)}.
\]
The system is stable if and only if the load factor \( \rho^{(1)} \) satisfies
\[
\rho^{(1)} = \frac{\lambda(\lambda + \gamma)^2(\lambda + \mu_2 + \gamma)}{M\gamma(\lambda + \gamma)^2 + \gamma\mu_1\mu_2(3(\lambda + \gamma) + \mu_1) + \mu_2^2\gamma(\lambda + \mu_1 + \gamma)} < 1. \tag{9}
\]
**Proof:** By elementary calculations.

Equations for the boundary states below the threshold level are still to be solved, namely:
\[
\pi_0^{(1)} = \pi_1^{(1)} D_0^{(1)} \frac{1}{\lambda} = \pi_1^{(1)} M_0^{(1)};
\]
\[
\pi_1^{(1)} = \pi_2^{(1)} D_1^{(1)}(C_1^{(1)} - B_1^{(1)}) - D_0^{(1)} \frac{1}{\lambda} A_4^{(1)} - 1 = \pi_2^{(1)} M_1^{(1)};
\]
\[
\pi_2^{(1)} = \pi_3^{(1)} D_2^{(1)}(C_2^{(1)} - B_2^{(1)} - M_1^{(1)} A_3^{(1)}) - 1 = \pi_3^{(1)} M_2^{(1)};
\]
\[
\pi_3^{(1)} = \pi_4^{(1)} D_3^{(1)}(C_3^{(1)} - B_3^{(1)} - M_2^{(1)} A_4^{(1)}) - 1 = \pi_4^{(1)} M_3^{(1)};
\]
\[
\pi_i^{(1)} = \pi_{i+1}^{(1)} D_2^{(1)}(C_2^{(1)} - B_2^{(1)} - M_1^{(1)} A_3^{(1)}) - 1 = \pi_{i+1}^{(1)} M_i^{(1)}, \quad 4 \leq i \leq q_2^2 - 1;
\]
\[
\pi_{q_2^2}^{(1)} = \pi_{q_2^2 - 1}^{(1)} M_{q_2^2 - 1}^{(1)} + R^{(1)} D_2^{(1)}(C_2^{(1)} - B_2^{(1)})^{-1}. \tag{10}
\]
The following algorithm is used to in the calculations:

**Step 1.** Solve (7) for matrix $R^{(1)}$, using iterations (8).

**Step 2.** Evaluate the Matrices $M^{(1)}_j$ for $j = 0, \ldots, q^*_2 - 1$.

**Step 3.** Evaluate the value $\pi^{(1)}_{q^*_2}$ by solving the normalisation condition

$$1 = \sum_{k=0}^{\infty} \pi^{(1)}_k e = \pi^{(1)}_{q^*_2} \sum_{i=0}^{q^*_2-1} \prod_{j=q^*_2-1-i}^{q^*_2-1} M^{(1)}_j e + \pi^{(1)}_{q^*_2} \sum_{j=0}^{\infty} (R^{(1)})^j e$$

(11)

$$= \pi^{(1)}_{q^*_2} \sum_{i=0}^{q^*_2-1} \prod_{j=q^*_2-1-i}^{q^*_2-1} M^{(1)}_j e + (I - R^{(1)})^{-1} e,$$

with the equation

$$\pi^{(1)}_{q^*_2} (M^{(1)}_{q^*_2-1} A^{(1)}_3 - (C^{(1)}_2 - B^{(1)}_2) + R^{(1)} D^{(1)}_2) = 0.$$

**Step 4.** Substitute $\pi^{(1)}_{q^*_2}$ in

$$\pi^{(1)}_i = \pi^{(1)}_{q^*_2} \prod_{j=q^*_2-1-i}^{q^*_2-1} M^{(1)}_j$$

(12)

for the values $i = 0, \ldots, q^*_2 - 1$ and calculate

$$\pi^{(1)}_{q^*_2+j} = \pi^{(1)}_{q^*_2} (R^{(1)})^j$$

(13)

for $j > 0$.

### 4 Steady-state distributions of the system under heuristic policies

To measure the advantages of optimal threshold policy (OTP) three more servers’ selection disciplines will be considered, namely

- Scheduling threshold policy (STP)
- Fastest free server selection (FFS)
- Random server selection (RSS)

The policies STP, FFS and RSS will be denoted by indices $m = \{2, 3, 4\}$, respectively. The formulas for the calculation of the steady-state distribution of the system under the STP are exactly the same as for the optimal policy with only one exception that the threshold level may differ from the optimal one.

In the next subsections calculation procedures are treated for systems under the FFS and RSS policies.

#### 4.1 Finite retrial group

In STP case the fastest server must be busy whenever the arrival occurs whereas the slower server can be switched on only if upon arrival of a primary or retrial customer the orbit has length defined by (2). The policy FSS means that the fastest free server must be occupied upon an arrival of a primary or retrial customer. Under the policy RSS arrivals choose any free server with equal probability.

Since the system under FFS and RSS control policies is described by the same Markov process $\{X(t)\}_{t \geq 0}$ with the state space $E$ as under the optimal control policy, we have a similar states partitioning.
Analogously as in previous sections we can write down the three-diagonal infinitesimal block matrices \( \Lambda^{(m)} \), \( m = \{2, 3, 4\} \) of dimension \( 4(K + 1) \). It is obvious that for STP matrix has the same form as for the optimal policy. For policies FFS and RSS policies in case \( K < \infty \) the infinitesimal matrices are of the form,

\[
\Lambda^{(m)} = \begin{pmatrix}
-\lambda & A^{(m)}_1 & 0 & 0 & \cdots & 0 \\
D^{(m)}_0 & -(C^{(m)}_1 - B^{(m)}_1) & A^{(m)}_2 & 0 & \cdots & 0 \\
0 & D^{(m)}_1 & -(C^{(m)}_2 - B^{(m)}_2) & A^{(m)}_3 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & D^{(m)}_3 & -(C^{(m)}_4 - B^{(m)}_4) & A^{(m)}_5 \\
0 & \cdots & 0 & 0 & \cdots & -\frac{1}{\lambda} \\
0 & \cdots & 0 & 0 & \cdots & 0 \\
\end{pmatrix}
\]

where \( A^{(3)}_1 = A^{(1)}_1, A^{(3)}_2 = \left( \begin{array}{ccc}
\lambda & 0 & 0 \\
0 & \lambda & 0 \\
0 & 0 & \lambda \\
\end{array} \right), A^{(3)}_i = A^{(1)}_i, i = 4, 5, 6, \ B^{(3)}_1 = B^{(1)}_1, B^{(3)}_2 = B^{(1)}_2, B^{(3)}_3 = B^{(1)}_3, B^{(3)}_4 = B^{(1)}_4, B^{(3)}_5 = B^{(1)}_5, C^{(3)}_1 = C^{(1)}_1, C^{(3)}_2 = C^{(1)}_2, C^{(3)}_i = C^{(1)}_i, i = 3, 4, D^{(3)}_i = D^{(1)}_i, 0 \leq i \leq 4, \ A^{(4)}_1 = \left( \begin{array}{ccc}
\frac{1}{2} & \frac{1}{2} & 0 \\
0 & \frac{1}{2} & \frac{1}{2} \\
0 & \frac{1}{2} & \frac{1}{2} \\
\end{array} \right), A^{(4)}_2 = \left( \begin{array}{ccc}
\lambda & 0 & 0 \\
0 & \lambda & 0 \\
0 & 0 & \lambda \\
\end{array} \right), A^{(4)}_i = \left( \begin{array}{ccc}
\lambda & 0 & 0 \\
0 & \lambda & 0 \\
0 & 0 & \lambda \\
\end{array} \right), A^{(5)}_i = \left( \begin{array}{ccc}
\lambda & 0 & 0 \\
0 & \lambda & 0 \\
0 & 0 & \lambda \\
\end{array} \right). \]

The steady-state probability vector \( \pi^{(m)} \), \( m = \{2, 3, 4\} \) satisfies the system

\[
\pi^{(m)} \Lambda^{(m)} = 0, \quad \pi^{(m)} e = 1.
\]

To solve the system for FFS and RSS policies we represent the equations in the form

\[
\pi^{(m)}_0 = \frac{\pi^{(m)}_1 D^{(m)}_0}{\lambda} = \pi^{(m)}_1 M^{(m)}_0, \\
\pi^{(m)}_1 = \pi^{(m)}_2 D^{(m)}_1 (C^{(m)}_1 - B^{(m)}_1) - D^{(m)}_0 A^{(m)}_1 = \pi^{(m)}_2 M^{(m)}_1, \\
\pi^{(m)}_2 = \pi^{(m)}_3 D^{(m)}_2 (C^{(m)}_2 - B^{(m)}_2) - M^{(m)}_1 A^{(m)}_2 = \pi^{(m)}_3 M^{(m)}_2, \\
\pi^{(m)}_i = \pi^{(m)}_{i+1} D^{(m)}_i (C^{(m)}_i - B^{(m)}_i) - M^{(m)}_{i-1} A^{(m)}_i = \pi^{(m)}_{i+1} M^{(m)}_i, \quad 3 \leq i \leq K - 1; \\
\pi^{(m)}_K = \pi^{(m)}_{K+1} D^{(m)}_K (C^{(m)}_K - B^{(m)}_K) - M^{(m)}_{K-1} A^{(m)}_K = \pi^{(m)}_{K+1} M^{(m)}_K, \\
\pi^{(m)}_{K+1} = \pi^{(m)}_{K+2} D^{(m)}_{K+1} (C^{(m)}_{K+1} - B^{(m)}_{K+1}) - M^{(m)}_{K} A^{(m)}_K = \pi^{(m)}_{K+2} M^{(m)}_{K+1}, \quad m = \{3, 4\}. \quad (14)
\]

To calculate the values \( \pi^{(m)}_k \), \( m = \{3, 4\} \) it is necessary:

**Step 1.** Evaluate matrices \( M^{(m)}_k \), \( k = 0, \ldots, K + 1 \) by relation (14).

**Step 2** Evaluate the value \( \pi^{(m)}_{K+2} \) from the normalization condition

\[
1 = \sum_{k=0}^{K+2} \pi^{(m)}_k e = \pi^{(m)}_{K+2} \sum_{i=0}^{K+1} M^{(m)}_j e = \pi^{(m)}_{K+2} \left[ 1 + \sum_{i=0}^{K+1} \prod_{j=K+1-i}^{K+1} M^{(m)}_j e \right]. \quad (15)
\]

**Step 3.** Substitute \( \pi^{(m)}_{K+2} \) into

\[
\pi^{(m)}_i = \pi^{(m)}_{K+2} \prod_{j=K+1-i}^{K+1} M^{(m)}_j.
\]

for \( i = 0, \ldots, K + 1 \).
4.2 Infinite retrial group

In case of infinite retrial group, $K = \infty$, the infinitesimal matrices $\Lambda^{(m)}$, $m \in \{2, 3, 4\}$ are obtained from the above matrices by removing the last three rows. In this case we conjecture the matrix-geometric form

$$\pi_{2+j} = \pi_2^{(m)} (R^{(m)})^j, \quad j \geq 0$$

where the matrix $R^{(m)}$ is to be found by solving the following quadratic equation

$$(R^{(m)})^2 D_2^{(m)} - (\epsilon^{(m)} - B_3^{(m)}) + A_4^{(m)} = 0. \quad (16)$$

As before the necessary and sufficient condition for stability is $p^{(m)} D_2^{(m)} e > p^{(m)} A_4^{(m)} e$, where $p^{(m)}$ is a stationary probability vector given by $p^{(m)} (A_4^{(m)} - (\epsilon^{(m)} - B_3^{(m)}) + D_2^{(m)}) = 0$, $p^{(m)} e = 1$.

**Theorem 4** For the system under STP and FFS control policies, the stationary probability vector $p^{(m)}$ of $A_4^{(m)} - (\epsilon^{(m)} - B_3^{(m)}) + D_2^{(m)}$, $m \in \{2, 3\}$ is given by formulas 3. The system is stable if and only if the load factor $\rho^{(m)}$, defined by (9) satisfies

$$\rho^{(m)} < 1, \quad m \in \{2, 3\}.$$

**Theorem 5** For the system under RSS control policy, the stationary probability vector $p^{(4)}$ of $A_4^{(4)} - (\epsilon^{(4)} - B_3^{(4)}) + D_2^{(4)}$ is given by

$$
\begin{align*}
\pi_0^{(4)} &= \frac{(\lambda + \gamma)^2}{(\lambda + \gamma)(\lambda + M + \gamma) + 2\mu_1\mu_2}, \\
\pi_1^{(4)} &= \frac{\mu_1}{\lambda + \gamma} \pi_0^{(4)}, \\
\pi_2^{(4)} &= \frac{\mu_2}{\lambda + \gamma} \pi_0^{(4)}, \\
\pi_3^{(4)} &= \frac{2\mu_1\mu_2}{(\lambda + \gamma)^2} \pi_0^{(4)}.
\end{align*}
$$

The system is stable if and only if the load factor $\rho^{(4)}$ satisfies

$$\rho^{(4)} = \frac{\lambda(\lambda + \gamma)^2}{M\gamma(\lambda + \gamma) + 2\gamma\mu_1\mu_2} < 1. \quad (17)$$

**Proof:** By elementary calculations.

Proportions $\pi_0^{(m)}$ and $\pi_1^{(m)}$ satisfy relations

$$
\begin{align*}
\pi_3^{(m)} &= \pi_2^{(m)} D_0^{(m)} \frac{1}{\lambda} = \pi_1^{(m)} M_0^{(m)}, \\
\pi_1^{(m)} &= \pi_2^{(m)} D_1^{(m)} (\epsilon^{(m)} - B_1^{(m)} - D_0^{(m)} \frac{1}{\lambda} A_1^{(m)})^{-1} = \pi_2^{(m)} M_1^{(m)}.
\end{align*}
$$

In case $m \in \{3, 4\}$ the following algorithm is used for the calculations:

**Step 1.** Solve equations (16) for matrix $R^{(m)}$, using iterations starting from $R^{(m)}(0) = 0$.

**Step 2.** Evaluate matrices $M_0^{(m)}$ and $M_1^{(m)}$.

**Step 3.** Evaluate value $\pi_2^{(m)}$ by solving the normalisation condition

$$
\begin{align*}
1 &= \sum_{k=0}^{\infty} \pi_k^{(m)} e = \pi_2^{(m)} M_1^{(m)} M_0^{(m)} e + \pi_2^{(m)} M_1^{(m)} e + \pi_2^{(m)} \sum_{j=0}^{\infty} (R^{(m)})^j e \\
&= \pi_2^{(m)} \left[ M_1^{(m)} M_0^{(m)} e + M_1^{(m)} e + (I - R^{(m)})^{-1} e \right].
\end{align*}
$$
with the equation
\[ \pi_2^{(m)} (M_1^{(m)} A_2^{(m)} - (C_3^{(m)} - B_3^{(m)}) + R^{(m)} D_2^{(m)}) = 0. \]

Step 4. Substitute \( \pi_2^{(m)} \) into
\[ \pi_0^{(m)} = \pi_2^{(m)} M_1^{(m)} M_0^{(m)}, \pi_1^{(m)} = \pi_2^{(m)} M_1^{(m)} \] (20)
and calculate
\[ \pi_2^{(m)} j = \pi_2^{(m)} (R^{(m)})^j, \quad j > 0. \] (21)

5 Stationary distribution of the waiting time

In this section we find the distribution of the waiting time. For the controllable queue with heterogeneous servers the waiting time of some customer can depend on the future arrival because in this case the number of active servers can be changed. Therefore, the waiting time of the tagged customer in the system under OTP strongly depends not only on its position in the queue, but also on the queue length during its waiting time. In the alternative models we have also consider not only the system state at the arrival time of the tagged customer, but also the possibility that the customer who comes later will be served by free server.

Thus, to calculate the waiting time distribution we will consider the process just after the arrival of the tagged customer with absorption at the time when he starts the service. Let us introduce the transient Markov process
\[ X(t) = (Q(t), D_1(t), D_2(t), J(t)) \] (22)
with the same genenerators as the models of the previous section. The state spaces
\[ E = \{ x = (q, d_1, d_2, j); 0 \leq q \leq K, d_i = \{0, 1\}, i = 1, 2, 0 \leq j \leq q \}, \]
where the last component \( J(t) \) denotes the position of the fixed job in the list of waiting jobs at time \( t \). This component can take the values \( \{0, 1, 2, \ldots\} \), and decreases for the system under threshold policy
- at time of a retrial arrival when the first server is idle,
- at time of a retrial arrival when at least one of the servers is idle and the queue length is greater than \( q^*_2 \),
and for other systems - at time of a retrial arrival when at least one server is idle.

The process is absorbed when the component \( J(t) \) become equal to zero. Note that \( Q(t) \geq J(t) \) at any time \( t \) when the targeted customer has to wait in the orbit. We denote the state of the process \( X(t) \) by \( x = (q, d_1, d_2, j) \). At the point of time of a new arrival \( t^+ \) (the initial time for the transient Markov process) it is obvious that
\[ J(t^+) = Q(t^+) \]
if tagged customer has to wait in the orbit and \( J(t^+) = 0 \) if upon arrival the customer can be served immediately.

Define
\[ W^{(m)} \] - r.v. of the waiting time in the system under policy \( m \),
\[ W_x^{(m)} \] - r.v. of the residual waiting time of the tagged customer given that the system state is \( x \),
\[ w_x^{(m)}(t) \] - the dencity function of the residual waiting time,
\[ \tilde{w}_x^{(m)}(s) = \mathbb{E}[e^{-sW_x^{(m)}}] = \int_0^\infty e^{-st} w_x^{(m)}(t)dt, \quad Re[s] \geq 0 \] corresponding Laplace-Stieltjes transform.
Because of the markovity of the process $X(t)$, the residual waiting time in state $x$ consists of the time the system spend in state $x$ until the next transition with density $\lambda_x e^{-\lambda_x t}$ plus the residual time in a new state $y$ after possible transition from the state $x$, which take place with probability $\frac{\lambda_y}{\lambda_x}$. Thus from the low of total probability for the density $w_x(t)$ we get

$$w_x^{(m)}(t) = \sum_{y \neq x} \frac{\lambda_{xy}}{\lambda_x} \left[ \lambda_x e^{-\lambda_x t} * w_y^{(m)}(t) \right], \quad x \in E$$  \hfill (23)

where $*$ denotes convolution.

Applying the Laplace-Stieltjes transforms to the relation (23) we get

$$\tilde{w}_x^{(m)}(s) = \sum_{y \neq x} \frac{\lambda_{xy}}{s + \lambda_x} \tilde{w}_y^{(m)}(s), \quad x \in E.$$  \hfill (24)

We partition the above Laplace-Stieltjes transforms according to the partition of the system states: define the column-vectors $w_{i,j}^{(m)}(s)$ in which $i$ denotes the number of customers in the system and $j$ the position of the tagged customer:

$$\tilde{w}_{j,0}^{(m)}(s) = \tilde{w}_{(j,0,0,j)}^{(m)}(s), \quad 1 \leq j \leq K$$

$$\tilde{w}_{j+1,0}^{(m)}(s) = (\tilde{w}_{(j,0,1,j)}^{(m)}(s), \tilde{w}_{(j,1,0,j)}^{(m)}(s), \tilde{w}_{(j+1,0,0,j)}^{(m)}(s))^t, \quad 1 \leq j \leq K - 1,$$

$$\tilde{w}_{1,j}^{(m)}(s) = (\tilde{w}_{(1-2,1,j)}^{(m)}(s), \tilde{w}_{(1-1,0,j)}^{(m)}(s), \tilde{w}_{(1,0,0,j)}^{(m)}(s))^t, \quad 1 \leq j \leq i - 2 \leq K - 2,$$

$$\tilde{w}_{K-1,j}^{(m)}(s) = (\tilde{w}_{(K-1,1,j)}^{(m)}(s), \tilde{w}_{(K,0,1,j)}^{(m)}(s), \tilde{w}_{(K,1,0,j)}^{(m)}(s))^t, \quad 1 \leq j \leq K - 1,$$

$$\tilde{w}_{K+1,j}^{(m)}(s) = (\tilde{w}_{(K,0,1,K)}^{(m)}(s), \tilde{w}_{(K,1,0,K)}^{(m)}(s))^t,$$

$$\tilde{w}_{K+2,j}^{(m)}(s) = \tilde{w}_{(K+1,1,j)}^{(m)}(s), \quad 1 \leq j \leq K,$$

$$\tilde{w}_{j}^{(m)}(s) = (\tilde{w}_{j,j}^{(m)}(s), \tilde{w}_{j+1,j}^{(m)}(s), \ldots, \tilde{w}_{K+2,j}^{(m)}(s))^t, \quad 1 \leq j \leq K,$$

$$\tilde{w}_{K}^{(m)}(s) = (\tilde{w}_{1}^{(m)}(s), \tilde{w}_{2}^{(m)}(s), \ldots, \tilde{w}_{K}^{(m)}(s))^t.$$

The following theorem gives recurrent relation for the vectors $\tilde{w}_j^{(m)}(s)$, $1 \leq j \leq K$ of the Laplace-Stieltjes transforms $\tilde{w}_x^{(m)}(t)$.

**Theorem 6** The vectors of the Laplace-Stieltjes transforms $\tilde{w}_j^{(m)}(s)$, $1 \leq j \leq K$ of the conditional waiting time densities under the control policy $m = \{1, 2, 3, 4\}$ are related by the following recurrent block three-diagonal system

$$\Lambda_W^{(m)}(s) \tilde{w}_j^{(m)}(s) = -\Gamma_j^{(m)} e,$$  \hfill (26)

$$\Lambda_W^{(m)}(s) \tilde{w}_j^{(m)}(s) = -\Gamma_j^{(m)} \tilde{w}_{j-1}^{(m)}(s), \quad 2 \leq j \leq K.$$

The matrices $\Lambda_W^{(m)}(s) = (\Phi_j^{(m)} - sI_{4(K-j+1)})$ and $\Gamma_j^{(m)}$, $j \geq 1$ are of the dimension $4(K-j+1)$. All
matrices have $K + 3 - j$ block-columns and $K + 3 - j$ block-rows. The matrices $\Phi^{(m)}_j$ are of the form:

$$
\Phi^{(m)}_j = \left( \begin{array}{cccccc}
- (\lambda + \gamma) & A^{(m)}_1 & 0 & 0 & 0 & 0 \\
D^{(m)}_0 & - C^{(m)}_1 & A^{(m)}_2 & 0 & 0 & 0 \\
0 & 0 & D^{(m)}_1 & - C^{(m)}_2 & A^{(m)}_3 & 0 \\
0 & 0 & 0 & D^{(m)}_2 & - C^{(m)}_3 & A^{(m)}_4 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & \cdots & 0 & D^{(m)}_3 & - C^{(m)}_4 & A^{(m)}_5 \\
0 & \cdots & 0 & 0 & D^{(m)}_4 & - C^{(m)}_5 \\
0 & \cdots & 0 & 0 & 0 & D^{(m)}_5 \\
\end{array} \right)
$$

$$
\Phi^{(m)}_j = \left( \begin{array}{cccccc}
- (\lambda + \gamma) & A^{(m)}_1 & 0 & 0 & 0 & 0 \\
D^{(m)}_0 & - C^{(m)}_1 & A^{(m)}_2 & 0 & 0 & 0 \\
0 & 0 & D^{(m)}_1 & - C^{(m)}_2 & A^{(m)}_3 & 0 \\
0 & 0 & 0 & D^{(m)}_2 & - C^{(m)}_3 & A^{(m)}_4 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & \cdots & 0 & D^{(m)}_3 & - C^{(m)}_4 & A^{(m)}_5 \\
0 & \cdots & 0 & 0 & D^{(m)}_4 & - C^{(m)}_5 \\
0 & \cdots & 0 & 0 & 0 & D^{(m)}_5 \\
\end{array} \right)
$$

$$
\Phi^{(m)}_j = \left( \begin{array}{cccccc}
- (\lambda + \gamma) & A^{(m)}_1 & 0 & 0 & 0 & 0 \\
D^{(m)}_0 & - C^{(m)}_1 & A^{(m)}_2 & 0 & 0 & 0 \\
0 & 0 & D^{(m)}_1 & - C^{(m)}_2 & A^{(m)}_3 & 0 \\
0 & 0 & 0 & D^{(m)}_2 & - C^{(m)}_3 & A^{(m)}_4 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & \cdots & 0 & D^{(m)}_3 & - C^{(m)}_4 & A^{(m)}_5 \\
0 & \cdots & 0 & 0 & D^{(m)}_4 & - C^{(m)}_5 \\
0 & \cdots & 0 & 0 & 0 & D^{(m)}_5 \\
\end{array} \right)
$$

where

$$
A^{(m)}_1 = \left( \begin{array}{cc}
0 & \lambda \\
\lambda & 0 \\
\end{array} \right), \quad A^{(m)}_2 = \left( \begin{array}{cccc}
\lambda & 0 & 0 & 0 \\
0 & \lambda & 0 & 0 \\
\end{array} \right), \quad A^{(m)}_3 = \left( \begin{array}{cccc}
\lambda & 0 & 0 & 0 \\
0 & \lambda & 0 & 0 \\
\end{array} \right), \quad A^{(m)}_4 = \left( \begin{array}{cccc}
\lambda & 0 & 0 & 0 \\
0 & \lambda & 0 & 0 \\
\end{array} \right), \quad A^{(m)}_5 = \left( \begin{array}{cccc}
\lambda & 0 & 0 & 0 \\
0 & \lambda & 0 & 0 \\
\end{array} \right)
$$

$$
C^{(m)}_1 = \left( \begin{array}{cc}
\alpha + \beta & 0 \\
0 & \gamma \\
\end{array} \right), \quad C^{(m)}_2 = \left( \begin{array}{cccc}
\alpha + \beta & 0 & 0 & 0 \\
0 & \gamma & 0 & 0 \\
\end{array} \right), \quad C^{(m)}_3 = \left( \begin{array}{cccc}
\alpha + \beta & 0 & 0 & 0 \\
0 & \gamma & 0 & 0 \\
\end{array} \right), \quad C^{(m)}_4 = \left( \begin{array}{cccc}
\alpha + \beta & 0 & 0 & 0 \\
0 & \gamma & 0 & 0 \\
\end{array} \right)
$$

$$
D^{(m)}_1 = \left( \begin{array}{cccc}
0 & 0 & 0 & 0 \\
0 & \mu & 0 & 0 \\
\end{array} \right), \quad D^{(m)}_2 = \left( \begin{array}{cccc}
0 & 0 & 0 & 0 \\
0 & \mu & 0 & 0 \\
\end{array} \right), \quad D^{(m)}_3 = \left( \begin{array}{cccc}
0 & 0 & 0 & 0 \\
0 & \mu & 0 & 0 \\
\end{array} \right), \quad D^{(m)}_4 = \left( \begin{array}{cccc}
0 & 0 & 0 & 0 \\
0 & \mu & 0 & 0 \\
\end{array} \right)
$$

$$
\Phi^{(m)}_{K-1} = \left( \begin{array}{cccc}
- (\lambda + \gamma) & A^{(m)}_1 & 0 & 0 \\
D^{(m)}_0 & - C^{(m)}_1 & A^{(m)}_2 & 0 \\
0 & 0 & D^{(m)}_1 & - C^{(m)}_2 \\
\end{array} \right), \quad \Phi^{(m)}_K = \left( \begin{array}{cccc}
- (\lambda + \gamma) & \hat{A}^{(m)}_1 & 0 & 0 \\
\hat{D}^{(m)}_0 & - \hat{C}^{(m)}_1 & \hat{A}^{(m)}_2 & 0 \\
0 & 0 & \hat{D}^{(m)}_1 & - \hat{C}^{(m)}_2 \\
\end{array} \right)
$$

12
The matrices $\Gamma_j^{(m)}$ are of the form

$$
\Gamma_j^{(m)} = \begin{pmatrix}
B_0^{(m)} & 0 & 0 & 0 & 0 & \ldots & 0 \\
0 & B_2^{(m)} & 0 & 0 & 0 & \ldots & 0 \\
0 & 0 & B_2^{(m)} & 0 & 0 & \ldots & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & 0 & 0 & \ldots & B_3^{(m)} & 0 & \ldots & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & 0 & 0 & \ldots & 0 & 0 & 0 & 0
\end{pmatrix}
$$

$$
\begin{cases}
q_2^* - j - 1, 1 \leq j \leq q_2^* - 2, m \in \{1, 2\} \\
K - q_2^*
\end{cases}
$$

$$
\Gamma_j^{(m)} = \begin{pmatrix}
B_0^{(m)} & 0 & 0 & 0 & 0 & \ldots & 0 \\
0 & B_2^{(m)} & 0 & 0 & 0 & \ldots & 0 \\
0 & 0 & B_3^{(m)} & 0 & 0 & \ldots & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & 0 & 0 & \ldots & B_4^{(m)} & 0 & \ldots & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & 0 & 0 & \ldots & 0 & 0 & 0 & 0
\end{pmatrix}
$$

$$
\begin{cases}
K - j - 1, j = q_2^* - 1, m = \{1, 2\} \\
q_2^* \leq j \leq K - 2, m = \{1, 2\} \\
1 \leq j \leq K - 2, m = \{3, 4\}
\end{cases}
$$

$$
\Gamma_{K-1}^{(m)} = \begin{pmatrix}
B_0^{(m)} & 0 & 0 & 0 \\
0 & B_2^{(m)} & 0 & 0 \\
0 & 0 & B_3^{(m)} & 0 \\
0 & 0 & 0 & B_4^{(m)}
\end{pmatrix}, \quad \Gamma_K^{(m)} = \begin{pmatrix}
B_0^{(m)} & 0 & 0 & 0 \\
0 & B_4^{(m)} & 0 & 0 \\
0 & 0 & B_4^{(m)} & 0 \\
0 & 0 & 0 & 0
\end{pmatrix}, \quad \Gamma_{K+1}^{(4)} = \begin{pmatrix}
B_0^{(4)} & 0 & 0 & 0 \\
0 & B_4^{(4)} & 0 & 0 \\
0 & 0 & B_4^{(4)} & 0 \\
0 & 0 & 0 & 0
\end{pmatrix}
$$

where

$$
B_0^{(m)} = \begin{pmatrix}
0 & \gamma & 0 \\
0 & 0 & \gamma
\end{pmatrix}, \quad \tilde{B}_3^{(m)} = \begin{pmatrix}
\gamma & 0 & 0 & 0 \\
0 & \gamma & 0 & 0 \\
0 & 0 & \gamma & 0
\end{pmatrix}, m = \{1, 2, 3\}, \quad B_0^{(4)} = \begin{pmatrix}
\gamma & 0 & 0 & 0 \\
0 & \gamma & 0 & 0 \\
0 & 0 & \gamma & 0 \\
0 & 0 & 0 & \gamma
\end{pmatrix}, m = \{1, 2, 3, 4\}.
$$

**Proof:** Consider the model under optimal policy $m = 1$. Note that if $J(t) = 0$ then the fixed job should be immediately served, so the waiting time is 0, i.e.

$$
\tilde{w}_x^{(m)}(s) = 1, \quad x = (q, d_1, d_2, 0) \in E.
$$

For other states and positions of the tagged customer according to the equation (24) the problem reduces to the above case by backward induction. This is always possible since the position $J(t^+)$ is a decreasing
integer.

\[
\tilde{w}^{(1)}_{(q,1,0,1)}(s) = \frac{1}{s + \lambda + \mu_1 + \gamma} \left[ \lambda \tilde{w}^{(1)}_{(q,1,1,1)}(s) + \mu_1 \tilde{w}^{(1)}_{(q,0,0,1)}(s) + \gamma \right] \\
\text{for } q_2^{*} \leq q \leq K,
\]

\[
\tilde{w}^{(1)}_{(q,0,1,1)}(s) = \frac{1}{s + \lambda + \mu_2 + \gamma} \left[ \lambda \tilde{w}^{(1)}_{(q,1,1,1)}(s) + \mu_2 \tilde{w}^{(1)}_{(q,0,0,1)}(s) + \gamma \right] \\
\text{for } 1 \leq q \leq K,
\]

\[
\tilde{w}^{(1)}_{(q,0,0,1)}(s) = \frac{1}{s + \lambda + \gamma} \left[ \lambda \tilde{w}^{(1)}_{(q,1,0,0)}(s) + \mu_1 \tilde{w}^{(1)}_{(q,0,0,0)}(s) + \gamma \right] \\
\text{for } 1 \leq q \leq K,
\]

\[
\tilde{w}^{(1)}_{(q,1,0,j)}(s) = \frac{1}{s + \lambda + \mu_1} \left[ \lambda \tilde{w}^{(1)}_{(q+1,1,0,j)}(s) + \mu_1 \tilde{w}^{(1)}_{(q,0,0,j)}(s) + \gamma \tilde{w}^{(1)}_{(q-1,1,1,j)}(s) \right] \\
\text{for } 1 \leq j \leq q, 1 \leq q \leq q_2^{*} - 2,
\]

\[
\tilde{w}^{(1)}_{(q,1,0,j)}(s) = \frac{1}{s + \lambda + \mu_1} \left[ \lambda \tilde{w}^{(1)}_{(q,1,1,j)}(s) + \mu_1 \tilde{w}^{(1)}_{(q,0,0,j)}(s) + \gamma \tilde{w}^{(1)}_{(q-1,1,1,j-1)}(s) \right] \\
\text{for } 1 \leq j \leq q, q = q_2^{*} - 1,
\]

\[
\tilde{w}^{(1)}_{(q,0,0,j)}(s) = \frac{1}{s + \lambda + \mu_1 + \gamma} \left[ \lambda \tilde{w}^{(1)}_{(q,1,0,j)}(s) + \mu_1 \tilde{w}^{(1)}_{(q,0,0,j)}(s) + \gamma \tilde{w}^{(1)}_{(q-1,1,0,j-1)}(s) \right] \\
\text{for } 2 \leq j \leq K, j \leq q \leq K,
\]

\[
\tilde{w}^{(1)}_{(q,0,0,j)}(s) = \frac{1}{s + \lambda + \mu_2 + \gamma} \left[ \lambda \tilde{w}^{(1)}_{(q,1,1,j)}(s) + \mu_2 \tilde{w}^{(1)}_{(q,0,0,j)}(s) + \gamma \tilde{w}^{(1)}_{(q-1,1,1,j-1)}(s) \right] \\
\text{for } 2 \leq j \leq K, j \leq q \leq K,
\]

\[
\tilde{w}^{(1)}_{(q,1,1,j)}(s) = \frac{1}{s + \lambda + \mu_1 + \mu_2} \left[ \lambda \tilde{w}^{(1)}_{(q+1,1,1,j)}(s) + \mu_1 \tilde{w}^{(1)}_{(q,0,1,j)}(s) + \mu_2 \tilde{w}^{(1)}_{(q,1,0,j)}(s) \right] \\
\text{for } 1 \leq j \leq K - 1, j \leq q \leq K - 1,
\]

\[
\tilde{w}^{(1)}_{(K,1,1,j)}(s) = \frac{1}{s + \mu_1 + \mu_2} \left[ \mu_1 \tilde{w}^{(1)}_{(K,0,1,j)}(s) + \mu_2 \tilde{w}^{(1)}_{(K,1,0,j)}(s) \right] \\
\text{for } 1 \leq j \leq K.
\]

Now, after routine block identification, we may express the system for \( m = 1 \) in (27) as given in (26). For the STP the recurrent expressions are the same with corresponding threshold level \( q_2^{*} \). For the policy FFS it is enough to set \( q_2^{*} = 1 \). To get the expressions for RSS policy we take the system for FSS policy and the equation for the state \( q, 0, 0, j \) rewrite as follows

\[
\tilde{w}^{(4)}_{(q,0,0,j)}(s) = \frac{1}{s + \lambda + \gamma} \left[ \frac{\lambda}{2} \tilde{w}^{(4)}_{(q,1,0,j)}(s) + \tilde{w}^{(4)}_{(q,0,0,j)}(s) + \gamma \right] \\
\text{for } 1 \leq j \leq K, j \leq q \leq K.
\]

\[\square\]

The tagged customer must wait in orbit if upon arrival he finds the system in some state of the subset

\[ E_{W}^{(m)} = \{(q, 1, 0); 0 \leq q \leq q_2^{*} - 1\} \cup \{(q, 1, 1); 0 \leq q \leq K - 1\}, \quad m = \{1, 2\} \]

\[ E_{W}^{(m)} = \{(q, 1, 1); 0 \leq q \leq K - 1\}, \quad m = \{3, 4\} \]
Denote by \( \pi^{(m)}_W, m = \{1, 2, 3, 4\} \) the row-vectors of the dimension \(2K(K+1)\) which include the subvectors of steady-state probabilities in set \(E^{(m)}_W\).

\[
\pi^{(m)}_W = (\pi^{(m)}_1 e_1(3) e_2^t(4K) + \pi^{(m)}_2 e_0(4) e_2^t(4K), \pi^{(m)}_1 e_2(4) e_2^t(4(K-1)) + \pi^{(m)}_2 e_0(4) e_2^t(4(K-1)), \ldots, \\
\pi^{(m)}_{q_2+1} e_0(4) e_2^t(4(K - q_2^* + 1)), \ldots, \pi^{(m)}_{K+1} e_0(3) e_2^t(4), m = \{1, 2\}, \\
\pi^{(m)}_W = (\pi^{(m)}_2 e_0(4) e_2^t(4K), \pi^{(m)}_3 e_0(4) e_2^t(4(K-1)), \ldots, \pi^{(m)}_{K+1} e_0(3) e_2^t(4), m = \{3, 4\}.
\]

According to the PASTA property the conditional probability of the state \(x^-\) upon arrival coincides with the unconditional one. Hence for the Laplace transform of the unconditional waiting time distribution with respect to all possible initial states \(x\) of the process \(X(t)\) and the corresponding states \(x^-\) before an arrival we have

\[
\tilde{W}^{(m)}(s) = \frac{1}{s} (1 - \pi^{(m)}_W e + \pi^{(m)}_W \tilde{w}^{(m)}(s)). \quad (28)
\]

The formula (28) includes two contributions:

\[
1 - \pi^{(m)}_W e = 1 - \left[ \sum_{q=0}^{q_2^*-1} \pi^{(m)}_{(q,1,0)} + \sum_{q=q_2^*-1}^{K-1} \pi^{(m)}_{(q,1,1)} \right] = \sum_{q=0}^{K-1} \pi^{(m)}_{(q,0,0)} + \sum_{q=q_2^*-1}^{K} \pi^{(m)}_{(q,1,0)}, m = \{1, 2\}
\]

\[
1 - \pi^{(m)}_W e = 1 - \sum_{q=0}^{K-1} \pi^{(m)}_{(q,1,1)} = \sum_{q=0}^{K-1} \pi^{(m)}_{(q,0,1)} + \pi^{(m)}_{(q,1,0)} + \pi^{(m)}_{(q,0,0)}, m = \{3, 4\}
\]

is a steady-state probability that the tagged customer does not have to wait for the service; the transform

\[
\pi^{(m)}_W \tilde{w}^{(m)}(s) = \sum_{q=0}^{q_2^*-2} \pi^{(m)}_{(q,1,0)} \tilde{w}^{(m)}_{(q+1,1,q+1)}(s) + \sum_{q=0}^{K-1} \pi^{(m)}_{(q,1,1)} \tilde{w}^{(m)}_{(q+1,1,q+1)}(s), m = \{1, 2\}
\]

\[
\pi^{(m)}_W \tilde{w}^{(m)}(s) = \sum_{q=0}^{K-1} \pi^{(m)}_{(q,1,1)} \tilde{w}^{(m)}_{(q+1,1,q+1)}(s), m = \{3, 4\}
\]

of the contribution of the waiting time with density function \(w^{(m)}_e(t)\) given \(W^{(m)} > 0\). Note that

\[
\int_0^\infty w^{(m)}_e(u) du = \pi^{(m)}_W e.
\]

The limit property of the Laplace-Stiltjes transform allows us to get the value of the function \(w^{(m)}_e(t)\) at point \(t = 0\):

\[
\lim_{s \to \infty} s \tilde{u}^{(m)}_{(q,d_1,d_2,j)}(s) = \begin{cases} 
\gamma, & \text{if } q_2^* \leq q \leq K, d_1 = 1, d_2 = 0, j = 1 \\
\gamma, & \text{if } 1 \leq q \leq K, d_1 = 0, d_2 = 1, j = 1 \\
\gamma, & \text{if } 1 \leq q \leq K, d_1 = 0, d_2 = 0, j = 1, \quad m = \{1, 2\} \\
0, & \text{otherwise}
\end{cases}
\]

\[
\lim_{s \to \infty} s \tilde{u}^{(m)}_{(q,d_1,d_2,j)}(s) = \begin{cases} 
\gamma, & \text{if } 1 \leq q \leq K, 0 \leq d_1 + d_2 \leq 1, \quad m = \{3, 4\} \\
0, & \text{otherwise}
\end{cases}
\]

Therefore we have

\[
\lim_{t \to 0} w^{(m)}_e(t) = \lim_{s \to \infty} s \pi^{(m)}_W \tilde{w}^{(m)}(s) = 0.
\]
The inversion of the Laplace transform $\frac{1}{\pi} \pi_W^{(m)} \tilde{W}^{(m)}(s)$ of the function $W^{(m)}_c(t) = \int_0^t w^{(m)}_c(u)du$ is used to get the distribution function of the waiting time

$$W^{(m)}(t) = P[W^{(m)} \leq t] = 1 - \pi_W^{(m)}e + W^{(m)}_c(t), \ t \geq 0.$$ 

Now we obtain the $n$-th moments of $W^{(m)}_c$ which we denote by $W^{(m)}_c(n) = E[(W^{(m)}_c)^n]$, $m = \{1, 2, 3, 4\}$. According to the introduced partitioning of the conditional Laplace transforms we denote by $W^{(m)}(n)$ the vector of the corresponding $n$-th moments:

$$W^{(m)}_{i,j}(n) = (W^{(m)}_{i,j}(d_1, d_2, j)(n) | d_i = \{0, 1\}, q + d_1 + d_2 = i)^t, 1 \leq i \leq K + 2, 1 \leq j \leq \min\{i, K\},$$

$$W^{(m)}_{j}(n) = (W^{(m)}_{j,j}(n), W^{(m)}_{j+1,j}(n), \ldots, W^{(m)}_{K+2,j}(n))^t, 1 \leq j \leq K,$$

$$W^{(m)}(n) = (W^{(m)}_{1}(n), W^{(m)}_{2}(n), \ldots, W^{(m)}_{K}(n))^t.$$ 

By differentiation the expressions (26) over the parameter $s$ we get

$$\Lambda^{(m)}_{w,1}(s) \frac{d^n}{ds^n} W^{(m)}_1(s) - n \frac{d^{n-1}}{ds^{n-1}} W^{(m)}_1(s) = 0,$$

$$\Lambda^{(m)}_{w,j}(s) \frac{d^n}{ds^n} W^{(m)}_j(s) - n \frac{d^{n-1}}{ds^{n-1}} W^{(m)}_j(s) = -\Gamma^{(m)}_{j} \frac{d^n}{ds^n} W^{(m)}_{j-1}(s), 2 \leq j \leq K.$$ 

Since $W^{(m)}_j(n) = (-1)^n \frac{d^n}{ds^n} W^{(m)}_j(s) \bigg|_{s=0}$ we get

$$\Phi^{(m)}_1 W^{(m)}_1(n) = -n W^{(m)}_1(n - 1), n \geq 1,$$

$$\Phi^{(m)}_j W^{(m)}_j(n) = -n W^{(m)}_j(n - 1) - \Gamma^{(m)}_j W^{(m)}_{j-1}(n), 2 \leq j \leq K, n \geq 1,$$

$$\tilde{W}^{(m)}_j(0) = e(4(K - j + 1)), 1 \leq j \leq K.$$ 

Therefore for the unconditional moments of the waiting time we have

$$E[(W^{(m)})^n] = \pi_W^{(m)} \tilde{W}^{(m)}(n), n \geq 0, m = \{1, 2, 3, 4\}.$$ 

### 6 Stationary distribution of the sojourn time

To carry out the calculation of a sojourn time distribution we apply the same Markov process (22) with only exception that the absorption takes place at the time when the tagged customer completes the service that will correspond to the case when $J(t) = -1$. At the point of time of a new arrival $t^+$ to some of the state from the set $E^{(m)}_W$ the customer has to wait for the service and

$$J(t^+) = Q(t^+).$$

The arrival to the state from the set $E^{(m)}_W \setminus E^{(m)}_W$ implies

$$J(t^+) = 0$$

that means that the customer must be served immediately on the first or second server according to the control policy. Let the process is absorbed when the component $J(t)$ become equal to $-1$. It occurs when the tagged customer leaves the system.

Define

$T^{(m)}$ - r.v. of the sojourn time in the system,
$T_x$ - r.v. of the residual sojourn time of the tagged customer given that the system state is $x$, $t_{x}$ the conditional density function of the residual sojourn time, $\tilde{t}_x^m(s) = E[e^{-sT_x^m}]$, $Re[s] \geq 0$ corresponding Laplace-Stiltjes transform.

As above we partition the conditional Laplace-Stiltjes transforms of the sojourn time densities in the following way:

$\tilde{t}_{i,j}^m(s) = \tilde{t}^m_{i,0,j}(s)$,  $1 \leq j \leq K$. \hfill (29)

$\tilde{t}^m_{0,j}(s) = (\tilde{t}^m_{-1,0,1,0}(s),\tilde{t}^m_{0,1,0,0}(s))$,  $1 \leq j \leq K+1$,

$\tilde{t}^m_{i,j}(s) = (\tilde{t}^m_{i,0,1,0}(s),\tilde{t}^m_{i,1,0,0}(s),\tilde{t}^m_{i,j+1,0,0}(s))$,  $1 \leq j \leq K-1$,

$\tilde{t}^m_{i,0}(s) = (\tilde{t}^m_{i,-1,0,1,0}(s),\tilde{t}^m_{i,-1,0,1,0}(s),\tilde{t}^m_{i,j+1,0,0}(s))$,  $1 \leq j \leq i-2 \leq K-2$,

$\tilde{t}^m_{K+1,j}(s) = (\tilde{t}^m_{K+1,-1,1,0}(s),\tilde{t}^m_{K+1,0,1,0}(s),\tilde{t}^m_{K+1,1,0,0}(s))$,  $1 \leq j \leq K-1$,

$\tilde{t}^m_{K+1,K}(s) = (\tilde{t}^m_{K+1,0,1,K}(s),\tilde{t}^m_{K+1,1,0,K}(s))$.

Theorem 7 The vectors of the Laplace-Stiltjes transforms $\tilde{t}_j^m(s)$,  $0 \leq j \leq K$ of the conditional sojourn time densities under the control policy $m = \{1, 2, 3, 4\}$ are related by the following recurrent block three-diagonal system

$\Lambda_{j,0}^m(s)\tilde{t}_0^m(s) = -\hat{1}_0^m e$,

$\Lambda_{j,1}^m(s)\tilde{t}_1^m(s) = -\hat{1}_1^m\tilde{t}_0^m(s)$,

$\Lambda_{j,j}^m(s)\tilde{t}_j^m(s) = -\hat{1}_j^m\tilde{t}_{j-1}^m(s)$,  $2 \leq j \leq K$.

where $\Lambda_{j,0}^m(s) = \Phi_{j,0}^m - sI_{2(K+1)}$ and $\Gamma_{0}^m$ are the block diagonal matrices of the dimension $2(K+1)$. These matrices have $K+1$ block-rows and block-columns. The matrices $\Lambda_{j,j}^m(s) = \Lambda_{W_{j}}^m(s)$ for $j \geq 1$, $\hat{1}_j^m$ are diagonal block matrices of the dimension $2(K+1) \times 4K$. These matrices have $K+1$ block-rows and $K+2$ block-columns. The matrices $\Phi_{j,0}^m$ have the blocks $\left( \begin{array}{cc} -\mu_2 & 0 \\ 0 & -\mu_1 \end{array} \right)$ and the matrices $\Gamma_{0}^m$ have
the blocks $\tilde{D}_0^{(m)}$. The matrices $\tilde{\Gamma}_1^{(m)}$ are of the form:

\[
\tilde{\Gamma}_1^{(m)} = \begin{pmatrix}
G_0^{(m)} & 0 & 0 & 0 & 0 & \cdots & 0 \\
0 & G_1^{(m)} & 0 & 0 & 0 & \cdots & 0 \\
0 & 0 & G_2^{(m)} & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & G_3^{(m)} & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 0 & \ddots & \ddots & \ddots \\
0 & 0 & 0 & \cdots & 0 & 0 & \ddots & G_4^{(m)} \\
\end{pmatrix}
\]

\[
\tilde{\Gamma}_4^{(m)} = \begin{pmatrix}
G_0^{(m)} & 0 & 0 & 0 & \cdots & 0 \\
0 & 0 & G_1^{(m)} & 0 & 0 & \cdots & 0 \\
0 & 0 & 0 & G_2^{(m)} & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 0 & \ddots & \ddots \\
0 & 0 & 0 & \cdots & 0 & 0 & \ddots & G_4^{(m)} \\
\end{pmatrix}, \quad m = \{3, 4\}
\]

where $G_0^{(m)} = \begin{pmatrix} 0 & \gamma \end{pmatrix}, m = \{1, 2, 3\}$. $G_4^{(4)} = \begin{pmatrix} 0 & \gamma & 1 \\
0 & \gamma & 0 \\
0 & \gamma & 0 \\
0 & \gamma & 0 \\
\end{pmatrix}$, $G_3^{(4)} = \begin{pmatrix} 0 & \gamma & 1 \\
0 & \gamma & 0 \\
0 & \gamma & 0 \\
0 & \gamma & 0 \\
0 & \gamma & 0 \\
\end{pmatrix}$, $G_2^{(4)} = \begin{pmatrix} 0 & \gamma & 1 \\
0 & \gamma & 0 \\
0 & \gamma & 0 \\
0 & \gamma & 0 \\
0 & \gamma & 0 \\
0 & \gamma & 0 \\
\end{pmatrix}$, $G_1^{(4)} = \begin{pmatrix} 0 & \gamma & 0 \\
0 & \gamma & 0 \\
0 & \gamma & 0 \\
0 & \gamma & 0 \\
0 & \gamma & 0 \\
0 & \gamma & 0 \\
0 & \gamma & 0 \\
\end{pmatrix}$, $m = \{1, 2, 3, 4\}$.

**Proof:** Consider again the system under OTP $m = \{1\}$. If $J(t) = -1$ then the served tagged customer leaves the system and the sojourn time is 0, i.e.

$\tilde{t}_x^{(1)}(s) = 1, x = (q, d_1, d_2, -1) \in E$.

For the states where the tagged customer is in service area the service time does not depend on the future arrivals or service time on the other server if it is busy. Thus we have

$\tilde{t}_x^{(1)}(s) = \frac{\mu_1}{s + \mu_1}, 0 \leq q \leq K$

$\tilde{t}_x^{(1)}(s) = \frac{\mu_2}{s + \mu_2}, 0 \leq q \leq K$.

For the states where the tagged customer stays just before the servers $J(t) = 1$ then we have the following recursive expressions:

\[
\tilde{t}_{(q, 1, 0, 0)}^{(1)}(s) = \frac{1}{s + \lambda + \mu_1 + \gamma} \left[ \lambda \tilde{t}_{(q, 1, 1, 1)}^{(1)}(s) + \mu_1 \tilde{t}_{(q, 0, 0, 1)}^{(1)}(s) + \gamma \tilde{t}_{(q-1, 0, 1, 0)}^{(1)}(s) \right]
\]

for $q_2^{\ast} \leq q \leq K$.

\[
\tilde{t}_{(q, 0, 1, 1)}^{(1)}(s) = \frac{1}{s + \lambda + \mu_2 + \gamma} \left[ \lambda \tilde{t}_{(q, 1, 1, 1)}^{(1)}(s) + \mu_2 \tilde{t}_{(q, 0, 0, 1)}^{(1)}(s) + \gamma \tilde{t}_{(q-1, 1, 0, 0)}^{(1)}(s) \right]
\]

for $1 \leq q \leq K$.

\[
\tilde{t}_{(q, 0, 0, 1)}^{(1)}(s) = \frac{1}{s + \lambda + \gamma} \left[ \lambda \tilde{t}_{(q, 1, 0, 1)}^{(1)}(s) + \gamma \tilde{t}_{(q-1, 1, 0, 0)}^{(1)}(s) \right]
\]

for $1 \leq q \leq K$.

In all other states the transforms $\tilde{t}_{x}^{(1)}(s)$ satisfy the equations (31). By expressing these equations in matrix form we obtain the expressions (30) for optimal policy. Analogously we obtain the expressions for the systems under other control policies.
The tagged customer joins the system if upon arrival he finds the system in one of the states in the set
\[ E_T^{(m)} = \{(q, d_1, d_2); d_1 + d_2 = 0, 1\} \cup \{(q, d_1, d_2); d_1 + d_2 = 2, 0 \leq q \leq K - 1\}, m = \{1, 2, 3, 4\} \]
Denote by \( \pi_T^{(m)} \) the row-vectors of the dimension \( 2(N + 1)^2 \) which include the steady-state probabilities of the states in the set \( E_T^{(m)} \):
\[
\pi_T^{(m)} = \left[ \sum_{i=q_2^1}^{K} \pi_i^{(m)} e_2(4) + \pi_{K+1}^{(m)} e_2(3) \right] e_0(2(K + 1))
\]
\[ + \left[ \pi_0^{(m)} + \pi_1^{(m)} (e_0(3) + e_2(3)) + \sum_{i=2}^{K} \pi_i^{(m)} (e_1(4) + e_3(4)) + \pi_{K+1}^{(m)} e_1(3) \right] e_1'(2(K + 1)), \pi_W^{(m)} \],
\[
\pi_T^{(3)} = \left[ \pi_1^{(3)} e_1(3) + \sum_{i=2}^{K} \pi_i^{(3)} e_2(4) + \pi_{K+1}^{(3)} e_2(3) \right] e_0(2(K + 1))
\]
\[ + \left[ \pi_0^{(3)} + \pi_1^{(3)} (e_0(3) + e_2(3)) + \sum_{i=2}^{K} \pi_i^{(3)} (e_1(4) + e_3(4)) + \pi_{K+1}^{(3)} e_1(3) \right] e_1'(2(K + 1)), \pi_W^{(3)} \],
\[
\pi_T^{(4)} = \left[ \pi_1^{(4)} e_1(3) + \sum_{i=2}^{K} \pi_i^{(4)} e_2(4) + \pi_{K+1}^{(4)} e_2(3) \right] e_0(2(K + 1))
\]
\[ + \left[ \pi_0^{(4)} + \pi_1^{(4)} e_0(3) + \sum_{i=2}^{K} \pi_i^{(4)} e_1(4) + \pi_{K+1}^{(4)} e_1(3) \right] e_1'(2(N + 1))
\]
\[ + \frac{1}{2} \left[ \pi_1^{(4)} e_2(3) + \sum_{i=2}^{K} \pi_i^{(4)} e_3(4) \right] (e_0'(2(K + 1)) + e_1'(2(K + 1))), \pi_W^{(4)} \].

Then for the unconditional Laplace transform of the sojourn time distribution with respect to all possible initial states \( x \) of the Process \( X^{(m)}(t) \) and corresponding states before an arrival \( x^- \) we get
\[
\tilde{T}^{(m)}(s) = \frac{1}{s} \pi_T^{(m)} \tilde{t}^{(m)}(s),
\]
and componentwise
\[
\pi_T^{(m, q_1, q_2)}(s) = \sum_{q=0}^{K} \left[ \pi_i^{(m)}(q_1, q_2, 0, 0) + \pi_i^{(m)}(q_1, q_2, 1, 0) \right] e_0(2(K + 1))
\]
\[ + \left[ \pi_0^{(m)}(q_1, q_2, 0, 1) + \pi_1^{(m)}(q_1, q_2, 1, 1) \right] e_1'(2(K + 1)), \pi_W^{(m)} \],
\[
\pi_T^{(3, q_1, q_2)}(s) = \sum_{q=0}^{K} \left[ \pi_i^{(3)}(q_1, q_2, 0, 1) + \pi_i^{(3)}(q_1, q_2, 1, 1) \right] e_0(2(K + 1))
\]
\[ + \left[ \pi_0^{(3)}(q_1, q_2, 0, 0) + \pi_1^{(3)}(q_1, q_2, 1, 0) \right] e_1'(2(K + 1)), \pi_W^{(3)} \],
\[
\pi_T^{(4, q_1, q_2)}(s) = \sum_{q=0}^{K} \left[ \pi_i^{(4)}(q_1, q_2, 0, 1) + \pi_i^{(4)}(q_1, q_2, 1, 1) \right] e_0(2(K + 1))
\]
\[ + \left[ \pi_0^{(4)}(q_1, q_2, 0, 0) + \pi_1^{(4)}(q_1, q_2, 1, 0) \right] e_1'(2(K + 1)), \pi_W^{(4)} \].
Let \( t^{(m)}(\tau), m = \{1, 2, 3, 4\} \) denotes the unconditional density associated with the Laplace-Stiltjes transform \( \bar{\ell}^{(m)}(s) \). Its value at point \( \tau = 0 \) satisfies
\[
\lim_{s \to \infty} s^{j} \bar{\ell}^{(m)}_{(q,d_{1},d_{2},j)}(s) = \begin{cases} 
\mu_{1}, & \text{if } 0 \leq q \leq K, d_{1} = 0, d_{2} = 1, j = 0 \\
\mu_{2}, & \text{if } 0 \leq q \leq K, d_{1} = 1, d_{2} = 0, j = 0 \\
0, & \text{otherwise}
\end{cases}
\]

Therefore,
\[
\lim_{\tau \to 0} t^{(m)}(\tau) = \lim_{s \to \infty} s \pi^{(m)}_{T} \bar{\ell}^{(m)}(s) = \mu_{2} \left[ \sum_{i = 0}^{K} \pi^{(m)}_{1} e_{2}(4) + \pi^{(m)}_{K+1} e_{2}(3) \right] + \mu_{1} \left[ s \pi^{(m)}_{0} + s \pi^{(m)}_{1} (e_{0}(3) + e_{2}(3)) + \sum_{i = 2}^{K} \pi^{(m)}_{i} (e_{1}(4) + e_{3}(4)) + \pi^{(m)}_{K+1} e_{1}(3) \right], m = \{1, 2\}
\]
\[
\lim_{\tau \to 0} t^{(3)}(\tau) = \lim_{s \to \infty} s \pi^{(m)}_{T} \bar{\ell}^{(3)}(s) = \mu_{2} \left[ \pi^{(3)}_{1} e_{1}(3) + \sum_{i = 2}^{K} \pi^{(3)}_{i} e_{2}(4) + \pi^{(3)}_{K+1} e_{2}(3) \right] + \mu_{1} \left[ s \pi^{(3)}_{0} + s \pi^{(3)}_{1} (e_{0}(3) + e_{2}(3)) + \sum_{i = 2}^{K} \pi^{(3)}_{i} (e_{1}(4) + e_{3}(4)) + \pi^{(3)}_{K+1} e_{1}(3) \right],
\]
\[
\lim_{\tau \to 0} t^{(4)}(\tau) = \lim_{s \to \infty} s \pi^{(m)}_{T} \bar{\ell}^{(4)}(s) = \mu_{2} \left[ \pi^{(4)}_{1} e_{1}(3) + \sum_{i = 2}^{K} \pi^{(4)}_{i} e_{2}(4) + \pi^{(4)}_{K+1} e_{2}(3) \right] + \mu_{1} \left[ s \pi^{(4)}_{0} + s \pi^{(4)}_{1} (e_{0}(3) + e_{2}(3)) + \sum_{i = 2}^{K} \pi^{(4)}_{i} (e_{1}(4) + e_{3}(4)) + \pi^{(4)}_{K+1} e_{1}(3) \right] + \frac{M}{2} \left[ \pi^{(4)}_{1} e_{2}(3) + \sum_{i = 2}^{K} \pi^{(4)}_{i} e_{3}(4) \right].
\]

Now it suffices to invert the Laplace transform \( \bar{\ell}^{(m)}(s) \) to get the distribution function of the sojourn time
\[
T^{(m)}(\tau) = \mathbb{P}[T^{(m)} \leq \tau] = \int_{0}^{\tau} t^{(m)}(u) du, \tau \geq 0.
\]

We now find the \( n \)-th moment of \( T^{(m)}_{x}(n) \), \( m = \{1, 2, 3, 4\} \) which is denoted by \( \bar{T}^{(m)}_{x}(n) = \mathbb{E}[(T^{(m)}_{x})^{n}] \) for \( n \geq 0 \). Let \( T^{(m)}(n) \) denote the vector containing the moments partitioned as a corresponding Laplace transforms:
\[
\bar{T}^{(m)}_{j,0}(n) = (\bar{T}_{j-1,0,0}^{(m)}(n), \bar{T}_{j-1,1,0}^{(m)}(n)), 1 \leq j \leq N + 1,
\]
\[
\bar{T}^{(m)}_{j,i}(n) = (\bar{T}_{(q,d_{1},d_{2},j)}^{(m)}(n))_{d_{1} = \{0, 1\}, q + d_{1} + d_{2} = i}^{t}, 1 \leq i \leq K + 2, 1 \leq j \leq \min\{i, K\},
\]
\[
\bar{T}^{(m)}_{0}(n) = (\bar{T}_{1,0}^{(m)}(n), \bar{T}_{2,0}^{(m)}(n), \ldots, \bar{T}_{K+2,0}^{(m)}(n)),
\]
\[
\bar{T}^{(m)}_{j}(n) = (\bar{T}_{j,0}^{(m)}(n), \bar{T}_{j,1}^{(m)}(n), \ldots, \bar{T}_{j,K+2}^{(m)}(n))^{t}, 1 \leq j \leq K,
\]
\[
\bar{T}^{(m)}(n) = (\bar{T}^{(m)}_{0}(n), \bar{T}^{(m)}_{1}(n), \ldots, \bar{T}^{(m)}_{K}(n))^{t}.
\]

By differentiation the Laplace transforms (30) over the parameter \( s \) since \( \bar{T}^{(m)}(n) = (-1)^{n} \frac{d^{n}}{ds^{n}} \ell^{(m)}(s) \bigg|_{s=0} \) we get
\[
\bar{T}^{(m)}_{0}(n) = \left( \frac{n!}{\mu_{2}}, \frac{n!}{\mu_{1}}, \ldots, \frac{n!}{\mu_{2}}, \frac{n!}{\mu_{1}} \right)^{t}, n \geq 1,
\]
\[
\Phi_{j}^{(m)} \bar{T}^{(m)}_{j}(n) = -n \bar{T}^{(m)}_{j}(n-1) - \bar{T}^{(m)}_{j-1}(n), 1 \leq j \leq K, n \geq 1
\]
\[
\bar{T}^{(m)}_{j}(0) = C(4(K - j + 1)), 1 \leq j \leq K.
\]
For the unconditional moments of the sojourn time we have
\[ \mathbb{E}[(T^{(m)})^n] = \pi_T^{(m)} \mathbf{T}^{(m)}(n), \quad n \geq 0. \]

## 7 The number of customers served by direct access

In this section we calculate the distribution for the discrete random value \( \Theta^{(m)} \) of the number of customers that have been served by direct access to some idle server until the tagged customer on the orbit reaches the service area. Denote by \( \theta \) where the first term represents the arrival of a primary customer that can be immediately served and the second term represents the service of this customer.

In this section we calculate the distribution for the discrete random value \( \Theta^{(m)} \) of the number of primary customers served by direct access to some idle server, \( \Theta_x^{(m)} \) - r.v. of the number of primary customers that will be served until the tagged customer reaches the service area, given that the system is in state \( x \).

\[ \theta_x^{(m)}(k) = \mathbb{P}[\Theta_x^{(m)} = k] \quad \text{the conditional dencity function of the r.v.} \quad \Theta_x^{(m)}, \]

\[ \tilde{\theta}_x^{(m)}(z) = \mathbb{E}[z \Theta_x^{(m)}] \quad \text{the conditional density function of the r.v. of the number of primary customers served by direct access to the idle server,} \]

\[ \tilde{\theta}_x^{(m)}(z) = \sum_{k=0}^{\infty} \theta_x^{(m)}(k)z^k, \mid z \mid \leq 1 \quad \text{corresponding } z - \text{transform.} \]

By the low of total probability for the Markov process \( X(t) \) the conditional dencity function \( \theta_x(k) \) has the form

\[ \theta_x^{(m)}(k) = \frac{a_{xy} \theta_y^{(m)}(1)}{a_x} \theta_x^{(m)}(k-1) + \sum_{y \neq x,y'} a_{xy} \theta_y^{(m)}(k), \quad k \geq 1, \quad x = 1, \ldots, m, \]

where the first term represents the arrival of a primary customer that can be immediately served and the second one corresponds to all other possible transitions that do not change the event under consideration.

Applying \( z - \) transforms to the relation (33) we get

\[ \tilde{\theta}_x^{(m)}(z) = \frac{z a_{xy} \tilde{\theta}_y^{(m)}(z)}{a_x} + \sum_{y \neq x,y'} a_{xy} \tilde{\theta}_y^{(m)}(z). \]

We partition the above \( z - \) transforms according to the partition of the system states: define the column-vectors \( \tilde{\theta}_{i,j}^{(m)}(z) \) in which \( i \) denotes the number of customers in the system and \( j \) the position of the tagged customer:

\[
\begin{align*}
\tilde{\theta}_{1,j}^{(m)}(z) &= \tilde{\theta}_{(j,0,0,j)}^{(m)}(z), \quad 1 \leq j \leq K, \\
\tilde{\theta}_{j+1,j}^{(m)}(z) &= (\tilde{\theta}_{(j,0,1,j)}^{(m)}(z), \tilde{\theta}_{(j,1,0,j)}^{(m)}(z), \tilde{\theta}_{(j+1,0,0,j)}^{(m)}(z))^t, \quad 1 \leq j \leq K - 1, \\
\tilde{\theta}_{i,j}^{(m)}(z) &= (\tilde{\theta}_{(i-2,1,1,j)}^{(m)}(z), \tilde{\theta}_{(i-1,0,1,j)}^{(m)}(z), \tilde{\theta}_{(i,1,0,j)}^{(m)}(z), \tilde{\theta}_{(i+1,0,0,j)}^{(m)}(z))^t, \quad 1 \leq j \leq i - 2 \leq K - 2, \\
\tilde{\theta}_{K+1,j}^{(m)}(z) &= (\tilde{\theta}_{(K,0,1,1,j)}^{(m)}(z), \tilde{\theta}_{(K,0,0,1,j)}^{(m)}(z), \tilde{\theta}_{(K,1,0,j)}^{(m)}(z))^t, \quad 1 \leq j \leq K - 1, \\
\tilde{\theta}_{K+2,j}^{(m)}(z) &= (\tilde{\theta}_{(K,0,1,K)}^{(m)}(z), \tilde{\theta}_{(K,1,0,K)}^{(m)}(z))^t, \\
\tilde{\theta}_{j}^{(m)}(z) &= (\tilde{\theta}_{j}^{(m)}(z), \tilde{\theta}_{j+1,j}^{(m)}(z), \ldots, \tilde{\theta}_{K+2,j}^{(m)}(z))^t, \quad 1 \leq j \leq K, \\
\tilde{\theta}_{j}^{(m)}(z) &= (\tilde{\theta}_{1}^{(m)}(z), \tilde{\theta}_{2}^{(m)}(z), \ldots, \tilde{\theta}_{K}^{(m)}(z))^t.
\end{align*}
\]

**Theorem 8** The vectors of \( z \)-transforms \( \tilde{\theta}_{j}^{(m)}(z) \), \( 1 \leq j \leq K \) of the conditional dencities under policy \( m \in \{1, 2, 3, 4\} \) are related by the following recurrent block three-diagonal system

\[ \Lambda_{\Theta_1}^{(m)}(z) \tilde{\theta}_1^{(m)}(z) = -I_1^{(m)} e, \]

\[ \Lambda_{\Theta_j}^{(m)}(z) \tilde{\theta}_j^{(m)}(z) = -I_j^{(m)} \tilde{\theta}_{j-1}^{(m)}(z), \quad 2 \leq j \leq K. \]

21
The matrices $\Lambda_{\Theta_j}(z) = (\Phi_j + (1 - z)Q_j)$ and $\Gamma_{\tilde{\lambda}}(z)$, $j \geq 1$ are of the dimension $4(K - j + 1)$. All matrices have $K - 3 - j$ block-columns and $K - 3 - j$ block-rows. The matrices $Q_j$ are of the form:

$$Q_j = -\begin{pmatrix}
0 & H_1 & 0 & 0 & 0 & \ldots & 0 \\
0 & 0 & H_2 & 0 & 0 & \ldots & 0 \\
0 & 0 & 0 & H_3 & 0 & \ldots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\
0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\
0 & \ldots & 0 & 0 & 0 & \ldots & 0
\end{pmatrix}, \quad q_j - 2, 1 \leq j \leq q_2 - 2, m = \{1, 2\}
$$

$$Q_j = -\begin{pmatrix}
0 & H_1 & 0 & 0 & \ldots & 0 \\
0 & 0 & H_2 & 0 & \ldots & 0 \\
0 & 0 & 0 & H_3 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & \ldots & 0 & 0 & \ldots & 0 \\
0 & \ldots & 0 & 0 & \ldots & 0 \\
0 & \ldots & 0 & 0 & \ldots & 0
\end{pmatrix}, \quad K - q_2 - 1
$$

$$Q_{K-1} = -\begin{pmatrix}
0 & H_1 & 0 & 0 \\
0 & 0 & H_2 & 0 \\
0 & 0 & 0 & H_3 \\
0 & 0 & 0 & 0
\end{pmatrix}, \quad Q_{K} = -\begin{pmatrix}
0 & H_1 & 0 \\
0 & 0 & H_2 \\
0 & 0 & 0 \\
0 & 0 & 0
\end{pmatrix}
$$

where

$$H_1 = \tilde{A}_1, \quad H_2 = \begin{pmatrix}
\lambda & 0 & 0 & 0 \\
0 & \lambda & 0 & 0 \\
0 & 0 & \lambda & 0 \\
0 & 0 & 0 & \lambda
\end{pmatrix}, \quad H_3 = \begin{pmatrix}
\lambda & 0 & 0 & 0 \\
0 & \lambda & 0 & 0 \\
0 & 0 & \lambda & 0 \\
0 & 0 & 0 & \lambda
\end{pmatrix}, \quad H_4 = \begin{pmatrix}
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{pmatrix}, \quad H_5 = \begin{pmatrix}
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{pmatrix}, \quad H_6 = \begin{pmatrix}
0 & \lambda & 0 & 0 \\
\lambda & 0 & 0 & 0 \\
\lambda & 0 & 0 & 0 \\
\frac{\lambda}{2} & \frac{\lambda}{2} & \frac{\lambda}{2} & \frac{\lambda}{2}
\end{pmatrix}, \quad H_{i} = \begin{pmatrix}
0 & 0 & 0 & 0 \\
\lambda & 0 & 0 & 0 \\
\lambda & 0 & 0 & 0 \\
\frac{\lambda}{2} & \frac{\lambda}{2} & \frac{\lambda}{2} & \frac{\lambda}{2}
\end{pmatrix}, \quad H_{i} = \begin{pmatrix}
0 & 0 & 0 & 0 \\
\lambda & 0 & 0 & 0 \\
\lambda & 0 & 0 & 0 \\
\frac{\lambda}{2} & \frac{\lambda}{2} & \frac{\lambda}{2} & \frac{\lambda}{2}
\end{pmatrix}
$$

Proof:

$$\tilde{\Theta}_{(q,1,0,0)}(z) = \frac{1}{\lambda + \mu_1} \left[ \lambda \tilde{\Theta}_{(q+1,1,0,0)}(z) + \mu_1 \tilde{\Theta}_{(q,0,0,0)}(z) \right]
$$

for $1 \leq j \leq q$, $1 \leq q \leq q_2 - 2$.

$$\tilde{\Theta}_{(q,1,0,0)}(z) = \frac{1}{\lambda + \mu_1} \left[ \lambda \tilde{\Theta}_{(q+1,1,1,0)}(z) + \mu_1 \tilde{\Theta}_{(q,0,0,1)}(z) \right]
$$

for $1 \leq j \leq q$, $q = q_2 - 1$.

$$\tilde{\Theta}_{(q,1,0,1)}(z) = \frac{1}{\lambda + \mu_1 + \gamma} \left[ \lambda \tilde{\Theta}_{(q+1,1,1,1)}(z) + \mu_1 \tilde{\Theta}_{(q,0,0,1)}(z) + \gamma \right]
$$

for $q_2^* \leq q \leq N$.

$$\tilde{\Theta}_{(q,0,1,1)}(z) = \frac{1}{\lambda + \mu_2 + \gamma} \left[ \lambda \tilde{\Theta}_{(q+1,1,1,1)}(z) + \mu_2 \tilde{\Theta}_{(q,0,0,1)}(z) + \gamma \right]
$$

for $1 \leq q \leq K$. 

22
\[ \tilde{\theta}^{(1)}_{(q,0,0,j)}(z) = \frac{1}{\lambda + \gamma} \left[ z\lambda \tilde{\theta}^{(1)}_{(q,1,0,1)}(z) + \gamma \right] \]
for \( 1 \leq q \leq K, \)
\[ \tilde{\theta}^{(1)}_{(q,1,0,j)}(z) = \frac{1}{\lambda + \mu_1 + \gamma} \left[ z\lambda \tilde{\theta}^{(1)}_{(q,1,1,j)}(z) + \mu_1 \tilde{\theta}^{(1)}_{(q,0,0,j)}(z) + \gamma \tilde{\theta}^{(1)}_{(q-1,1,1,j-1)}(z) \right] \]
for \( q^* \leq j \leq K, j \leq q \leq K, \)
\[ \tilde{\theta}^{(1)}_{(q,0,j,j)}(z) = \frac{1}{\lambda + \mu_2 + \gamma} \left[ z\lambda \tilde{\theta}^{(1)}_{(q,1,1,j)}(z) + \mu_2 \tilde{\theta}^{(1)}_{(q,0,0,j)}(z) + \gamma \tilde{\theta}^{(1)}_{(q-1,1,1,j-1)}(z) \right] \]
for \( 2 \leq j \leq K, j \leq q \leq K, \)
\[ \tilde{\theta}^{(1)}_{(q,1,j,j)}(z) = \frac{1}{\lambda + \mu_1 + \mu_2} \left[ \lambda \tilde{\theta}^{(1)}_{(q+1,1,1,j)}(z) + \mu_1 \tilde{\theta}^{(1)}_{(q,0,1,j)}(z) + \mu_2 \tilde{\theta}^{(1)}_{(q,0,0,j)}(z) \right] \]
for \( 1 \leq j \leq K - 1, j \leq q \leq K - 1, \)
\[ \tilde{\theta}^{(1)}_{(K,1,j,j)}(z) = \frac{1}{\mu_1 + \mu_2} \left[ \mu_1 \tilde{\theta}^{(1)}_{(K,0,1,j)}(z) + \mu_2 \tilde{\theta}^{(1)}_{(K,1,0,j)}(z) \right] \]
for \( 1 \leq j \leq K. \)

For the unconditional \( z \)-transform of the number of directly served customers we get
\[ \tilde{\theta}^{(m)}(z) = 1 - \pi^{(m)}_W e + \pi^{(m)}_W \theta^{(m)}(z), \quad (37) \]
where \( 1 - \pi^{(m)}_W e \) as before means the probability that the customer upon an arrival goes immediately to the service area. The inversion of the contribution \( \pi^{(m)}_W \theta^{(m)}(z) \) leads to the density function \( \theta^{(m)}_c(k) \), that together with \( \Theta^{(m)}_c(n) = \sum_{k=0}^{n} \theta^{(m)}_c(k) \) implies the relation for the distribution function
\[ \Theta^{(m)}(n) = 1 - \pi^{(m)}_W e + \Theta^{(m)}_c(n), \quad n \geq 0. \]
The probability that no primary customers will be directly served while another are waiting on the orbit can be expressed as follows
\[ \Theta^{(m)}(0) = 1 - \pi^{(m)}_W e + \pi^{(m)}_W \theta^{(m)}(0). \]
The \( n \)-th factorial moment of the random value \( \Theta^{(m)}_x(n) \) denote by \( \Theta^{(m)}_x(n) = \mathbb{E}[\Theta^{(m)}_x(\Theta^{(m)}_x - 1) \ldots (\Theta^{(m)}_x - n + 1)], \quad n \geq 1 \). We partition the conditional factorial moments in the same way as before:
\[ \Theta^{(m)}_{i,j}(n) = (\Theta^{(m)}_{q+d_1,d_2,j}(n)) \quad \text{for} \quad 1 \leq i \leq K + 2, 1 \leq j \leq \min\{i, K\}, \]
\[ \Theta^{(m)}_{j,j}(n) = (\Theta^{(m)}_{j,j}(n)), \Theta^{(m)}_{j+1,j}(n), \ldots, \Theta^{(m)}_{K+2,j}(n)) \quad \text{for} \quad 1 \leq j \leq K, \]
\[ \Theta^{(m)}_{j,i}(n) = (\Theta^{(m)}_{i,n}), \Theta^{(m)}_{2,n}(n), \ldots, \Theta^{(m)}_{K,n}(n)). \]

By differentiating the relation (35) we get
\[ \Lambda^{(m)}_{\Theta,1}(z) \frac{d^n}{dz^n} \tilde{\theta}^{(m)}_{i,j}(z) - nQ^{(m)}_{i,j} \frac{d^{n-1}}{dz^{n-1}} \tilde{\theta}^{(m)}_{1,j}(z) = 0, \]
\[ \Lambda^{(m)}_{\Theta,j}(z) \frac{d^n}{dz^n} \tilde{\theta}^{(m)}_{j,j}(z) - nQ^{(m)}_{j,j} \frac{d^{n-1}}{dz^{n-1}} \tilde{\theta}^{(m)}_{j-1,j}(z) = -\Gamma^{(m)}_{j} \frac{d^n}{dz^n} \tilde{\theta}^{(m)}_{j-1,j}(z), \quad 2 \leq j \leq K. \]
Since $\Theta_j^{(m)}(n) = \frac{d^n}{dz^n} \tilde{\Theta}_j^{(m)}(z) \bigg|_{z=1}$ we get

$$
\Phi_1^{(m)} \Theta_1^{(m)}(n) = n Q_1^{(m)} \Theta_1^{(m)}(n-1), \ n \geq 1,
$$

$$
\Phi_j^{(m)} \Theta_j^{(m)}(n) = n Q_j^{(m)} \Theta_j^{(m)}(n-1) - \Gamma_j^{(m)} \Theta_{j-1}^{(m)}(n), \ 2 \leq j \leq K, \ n \geq 1,
$$

$$
\Theta_j^{(m)}(0) = e(4(K-j+1)), \ 1 \leq j \leq K.
$$

The moments for the unconditional random values have the form

$$
E[\Theta^{(m)}(\Theta^{(m)} - 1) \ldots (\Theta^{(m)} - n + 1)] = \pi_{\Psi}^{(m)} \Theta^{(m)}(n), \ n \geq 0.
$$

## 8 The number of retrials made by a customer

Denote by

$\Psi^{(m)} = \text{r.v. of the number of retrials made by a tagged customer until the service starts},$

$\Psi^{(m)} = \text{r.v. of the number of retrials made by a tagged customer given that the system is in state } x,$

$\psi^{(m)}_x(k) = \mathbb{P}[\Psi^{(m)} = k] \text{ the conditional density function of the r.v. } \Psi^{(m)},$

$\tilde{\psi}^{(m)}_x(z) = E[z^{\Psi^{(m)}_x}(k)] = \sum_{k=0}^{\infty} \psi^{(m)}_x(k) z^k, \ |z| \leq 1 \text{ corresponding } z-\text{transform}.$

By the law of total probability for the Markov process $X(t)$ the conditional density function $\psi^{(m)}_x(k)$ has the form

$$
\psi^{(m)}_x(k) = \frac{a_{xy}^{(m)}}{a_x} \psi^{(m)}_y(k - 1) + \sum_{y \neq x, y'} \frac{a_{xy}}{a_x} \psi^{(m)}_y(k), \quad (38)
$$

where the first term represents the retrial of a tagged customer.

Applying $z-$ transforms to the relation (38) we get

$$
\tilde{\psi}^{(m)}_x(z) = \frac{a_{xy}^{(m)}}{a_x} \tilde{\psi}^{(m)}_y(z) + \sum_{y \neq x, y'} \frac{a_{xy}}{a_x} \tilde{\psi}^{(m)}_y(z), \quad (39)
$$

$$
\tilde{\psi}^{(m)}_{j,j}(z) = \tilde{\psi}^{(m)}_{j,00,j}(z), \ 1 \leq j \leq K
$$

$$
\tilde{\psi}^{(m)}_{j,j+1,j}(z) = \tilde{\psi}^{(m)}_{j,01,j}(z), \ \tilde{\psi}^{(m)}_{j,j+1,0,j}(z), \ \tilde{\psi}^{(m)}_{j,j+1,0,j}(z) \bigg|_{z=1}, \ 1 \leq j \leq K - 1,
$$

$$
\tilde{\psi}^{(m)}_{j,j}(z) = \bigg( \tilde{\psi}^{(m)}_{j,11,j}(z), \ \tilde{\psi}^{(m)}_{j,11,j}(z), \ \tilde{\psi}^{(m)}_{j,11,j}(z) \bigg) \bigg|_{z=1}, \ 1 \leq j \leq K - 1,
$$

$$
\psi^{(m)}_{K+1,j}(z) = \psi^{(m)}_{K+1,1,j}(z), \ \psi^{(m)}_{K+1,1,j}(z), \ \psi^{(m)}_{K+1,1,j}(z) \bigg|_{z=1}, \ 1 \leq j \leq K - 1,
$$

$$
\psi^{(m)}_{K,0j}(z) = \psi^{(m)}_{K,0j}(z), \ \psi^{(m)}_{K,0j}(z), \ \psi^{(m)}_{K,0j}(z) \bigg|_{z=1}, \ 1 \leq j \leq K,
$$

$$
\psi^{(m)}_{j}(z) = \psi^{(m)}_{j}(z), \ \psi^{(m)}_{j+1,j}(z), \ \psi^{(m)}_{j+2,j}(z) \bigg|_{z=1}, \ 1 \leq j \leq K,
$$

$$
\psi^{(m)}_{j}(z) = \psi^{(m)}_{j}(z), \ \psi^{(m)}_{j+1,j}(z), \ \psi^{(m)}_{j+2,j}(z) \bigg|_{z=1}, \ 1 \leq j \leq K,
$$

Theorem 9 The vectors of $z$-transforms $\tilde{\psi}^{(m)}_{j}(z), \ 1 \leq j \leq K$ of the conditional densities under the control policy $m = \{1, 2, 3, 4\}$ are related by the following recurrent block three-diagonal system

$$
\Lambda^{(m)}_{\Psi^{(m)}}(z) \tilde{\psi}^{(m)}_{j}(z) = -\Gamma^{(m)}_{\tilde{\psi}^{(m)}}(z) \tilde{\psi}^{(m)}_{j-1}(z), \ 2 \leq j \leq K.
$$

24
The matrices \( \Lambda_{q,j}^{(m)}(z) = (\Phi_1^{(m)} + (1 - z)V^{(m)}, \Lambda_{q,j}^{(m)} = \Phi_j^{(m)}, j \geq 2 \) and \( \Gamma_{1}^{(m)}(z) = z\Gamma_{1}^{(m)}, \Gamma_{j}^{(m)}, j \geq 2 \) are of the dimension \( 4(N - j + 1) \). All matrices have \( N + 3 - j \) block-columns and \( N + 3 - j \) block-rows. The matrix \( V^{(m)} \) is of the form

\[
V^{(m)} = -\gamma\begin{pmatrix}
0 & 0 & 0 & \ldots & 0 \\
0 & (e_0 + e_2)(e_0 + e_2)^t & 0 & \ldots & 0 \\
0 & 0 & 0 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & 0 \\
0 & \ldots & 0 & 0 & 0 \\
0 & \ldots & 0 & 0 & 0 \\
0 & \ldots & 0 & 0 & 0 \\
0 & \ldots & 0 & 0 & 0 \\
0 & \ldots & 0 & 0 & 0 \\
\end{pmatrix}
\]

\[
N - q_2^\
\]
\[ \psi_{(q,0,1,j)}(z) = \frac{1}{\lambda + \mu_2 + \gamma} \left[ \lambda \psi_{(q,1,1,j)}(z) + \mu_2 \psi_{(q,0,0,j)}(z) + \gamma \psi_{(q-1,1,1,j-1)}(z) \right] \]
for \( 2 \leq j \leq K, \ j \leq q \leq K, \)
\[
\psi_{(q,0,0,j)}(z) = \frac{1}{\lambda + \gamma} \left[ \lambda \psi_{(q,1,0,j)}(z) + \gamma \psi_{(q-1,1,0,j-1)}(z) \right]
\]
for \( 2 \leq j \leq K, \ j \leq q \leq K, \)
\[
\psi_{(q,1,1,j)}(z) = \frac{1}{\lambda + \mu_1 + \mu_2} \left[ \lambda \psi_{(q+1,1,1,j)}(z) + \mu_1 \psi_{(q,0,1,j)}(z) + \mu_2 \psi_{(q,1,0,j)}(z) \right]
\]
for \( 2 \leq j \leq K - 1, \ j \leq q \leq K - 1, \)
\[
\psi_{(K,1,1,j)}(z) = \frac{1}{\mu_1 + \mu_2} \left[ \mu_1 \psi_{(K,0,1,j)}(z) + \mu_2 \psi_{(K,1,0,j)}(z) \right]
\]
for \( 2 \leq j \leq K. \)

For the unconditional \( z \)-transform we get
\[
\psi^{(m)}(z) = 1 - \pi_W^{(m)} e + \pi_W^{(m)} \psi^{(m)}(z), \quad (42)
\]
where \( 1 - \pi_W^{(m)} e \) denotes the probability that the customer upon an arrival goes immediately to the service area. By inversion of the contribution \( \pi_W^{(m)} \psi^{(m)}(z) \) we get the density function \( \psi^{(m)}(k) \) and by the relation \( \psi^{(m)}(n) = \sum_{k=0}^{n} \psi^{(m)}(k) \) we derive the distribution function
\[
\psi^{(m)}(n) = 1 - \pi_W^{(m)} e + \psi^{(m)}(n), \ n \geq 0.
\]
The probability that the customer makes no retrials coincides with the probability that it will be directly served, i.e.
\[
\psi^{(m)}(0) = 1 - \pi_W^{(m)} e.
\]
The \( n \)-th factorial moment of the random value \( \psi^{(m)}(n) \) denote by \( \tilde{\psi}_{x}^{(m)}(n) = \mathbb{E}[\psi^{(m)}(\psi^{(m)} - 1) \ldots (\psi^{(m)} - n + 1)] \), \( n \geq 1 \). We partition the conditional value \( \psi^{(m)}(n) \) denote by \( \tilde{\psi}_{x}^{(m)}(n) = \mathbb{E}[\psi^{(m)}(\psi^{(m)} - 1) \ldots (\psi^{(m)} - n + 1)] \), \( n \geq 1 \). We partition the conditional value \( \psi^{(m)}(n) \) denote by \( \tilde{\psi}_{x}^{(m)}(n) = \mathbb{E}[\psi^{(m)}(\psi^{(m)} - 1) \ldots (\psi^{(m)} - n + 1)] \), \( n \geq 1 \). We partition the conditional value \( \psi^{(m)}(n) \) denote by \( \tilde{\psi}_{x}^{(m)}(n) = \mathbb{E}[\psi^{(m)}(\psi^{(m)} - 1) \ldots (\psi^{(m)} - n + 1)] \), \( n \geq 1 \).

By differentiating the relation (35) we get
\[
A^{(m)}_{i,j}(z) = \frac{d^n}{dz^n} \psi^{(m)}(z) - n V^{(m)}_{j} \frac{d^{n-1}}{dz^{n-1}} \psi^{(m)}(z) = 0,
\]
\[
A^{(m)}_{j,j} \frac{d^n}{dz^n} \psi^{(m)}(z) = - \Gamma^{(m)}_{j} \frac{d^n}{dz^n} \psi^{(m)}(z), \ 2 \leq j \leq K.
\]
Since \( \psi^{(m)}(n) = \frac{d^n}{dz^n} \psi^{(m)}(z) \bigg|_{z=1} \) we get
\[
\Phi^{(m)}_{11}(n) = n V^{(m)}_{1} \psi^{(m)}(n - 1), \ n \geq 1,
\]
\[
\Phi^{(m)}_{j1}(n) = - \Gamma^{(m)}_{j} \psi^{(m)}(n - 1), \ 2 \leq j \leq K, \ n \geq 1,
\]
\[
\Phi^{(m)}_{jj}(0) = e(4(K - j + 1)), \ 1 \leq j \leq K.
\]
The moments for the unconditional random values have the form
\[
\mathbb{E}[\psi^{(m)}(\psi^{(m)} - 1) \ldots (\psi^{(m)} - n + 1)] = \pi_W^{(m)} \psi^{(m)}(n), \ n \geq 0.
\]
9 Numerical results and comparison analysis

Consider the system $M/M/2$ with primary arrival rate $\lambda$, retrial rate $\gamma$ and service rates $\mu_1$ and $\mu_2$. By inversion of the derived Laplace transforms (28) and (32) it is possible to evaluate the waiting and sojourn time distributions. There are two practical algorithms for the inversion of Laplace transform: The conventional algorithm and the Fourier-series algorithm. Mathematical packages such as Mathematica, Mathlab, Mathcad, etc. include the standard functions that apply the conventional one. It is based on the partial function expansion method and works efficiently only in case of small order rational functions. So we can obtain on the basis of this method an accurate representation of the functions $W^{(m)}(t)$ and $T^{(m)}(t)$ only for small $t$ but in symbolic form. The algorithms based on Furier-series represent the numerical inversion methods, e.g. Euler and Post-Widder [2], that can be used for large $t$ as well.

The mentioned mathematical software allow also to invert $z$-transforms (37) and (42) to get the distribution functions $\Theta(n)$ and $\Psi(n)$. However there are problems by inversion of the functions of higher order. Therefore we implement the numerical inversion of $z$-transforms using the Lattice-Poisson algorithm [1].

By means of Mathematica package we have created the procedures:

- for the calculation of steady-state probabilities under optimal and heuristic service disciplines, formulas (10–13) and (18)–(21),
- for the numerical inversion of the Laplace transforms $\tilde{W}(s)$ and $\tilde{T}(s)$ using the Euler and Post-Widder algorithms,
- for the numerical inversion of the $z$-transforms $\tilde{\Theta}(z)$ and $\tilde{\Psi}(z)$ using the Lattice-Poisson algorithm.

In Figures 1–4 we have indicated the waiting time (the figures labelled by letter "a") and the sojourn time (the figures labelled by letter "b") for different values of the primary customer arrival rate $\lambda$ and retrial customer rate $\gamma$. In our examples we fix the service rates $\mu_1 = 2.2$, $\mu_2 = 0.3$. The following observation can be noticed from these figures:

1. The curves of the waiting time distributions $W^{(m)}(t)$ for the system under threshold control policies (OTP, STP) lie below the other curves (FFS, RSS) that specifies that the waiting time of a customer in the orbit is larger for the threshold systems. This does not contradict the optimality of the threshold policy since it minimizes the sojourn time. On the figures with sojourn time distributions $T^{(m)}(t)$ one can notice that for some small values of argument $t$ the curves for the OTP can lie below other graphs but starting from some point of time $t$ they are above the others. Nevertheless the mean sojourn time for the optimal policy turns out to be the least. The curves of the waiting time distribution for the system under FFS control policy lie above the other curves that illustrates that this policy minimizes the waiting time of a customer in the orbit.

At the same time the largest sojourn time belongs to the system under RSS policy. It can be explained by the fact that this policy with equal probability assigns a customer to the faster or slower server that in turn makes a significant contribution to the sojourn time increasing.

2. In Figures 1 and 2 the primary customer arrival rate $\lambda$ is varied and the corresponding load factors are $\rho^{(m)} = 0.196$, $m = \{1, 2, 3\}$, $\rho^{(4)} = 0.204$ and $\rho^{(m)} = 0.410$, $m = \{1, 2, 3\}$, $\rho^{(4)} = 0.424$. In Figure 3 and 4 the retrial customer rate $\gamma$ is varied and the load factors are, respectively, $\rho^{(m)} = 0.199$, $m = \{1, 2, 3\}$, $\rho^{(4)} = 0.200$ and $\rho^{(m)} = 0.375$, $m = \{1, 2, 3\}$, $\rho^{(4)} = 0.377$. As $\lambda$ or $\gamma$ increases then the load factor $\rho^{(m)}$ also increases that leads to the distributions with heavier tails. While in Figure 1 and 3 the curves for the threshold systems (OTP, STP) look very similar, in Figure 2 and 4 the sufficient difference can be noticed. Thus we can assume that if the load factor is sufficiently small, i.e. the system has a so-called "light traffic", the curves for the OTP can lie below other graphs but starting from some point of time $t$ they are above the others.
then the scheduling threshold policy may be a good approximation for the optimal one.

Figure 1: Distribution functions (a) $W^{(m)}(t)$ (b) $T^{(m)}(t)$ for $\lambda=0.5, \mu_1=2.2, \mu_2=0.3, \gamma=2.5$

Figure 2: Distribution functions (a) $W^{(m)}(t)$ (b) $T^{(m)}(t)$ for $\lambda=0.9, \mu_1=2.2, \mu_2=0.3, \gamma=2.5$

Figure 3: Distribution functions (a) $W^{(m)}(t)$ (b) $T^{(m)}(t)$ for $\lambda=0.5, \mu_1=2.2, \mu_2=0.3, \gamma=8.5$
Figure 4: Distribution functions (a) $W^{(m)}(t)$ (b) $T^{(m)}(t)$ for $\lambda=0.9$, $\mu_1=2.2$, $\mu_2=0.3$, $\gamma=8.5$

The following Figures 5 ($\lambda = 0.5$ and $\lambda = 0.9$ in figures labelled, respectively, by letter "a" and "b") and 6 ($\gamma = 2.5$ and $\gamma = 8.5$ in figures labelled, respectively, by letter "a" and "b") represent the discrete distribution functions $\Theta(n)$ as a stepped curves for the number of primary arrivals that will be directly served before an orbiting customer reaches the service area. The presented diagrams reveals the following observations

1. At point $t = 0$ the presented functions equal to the probability that there are no customers that will be served directly. As it was mentioned above, this probability equals to the probability $W^{(m)}(0)$ plus some probability that the orbiting customer will be served earlier then a primary arrival reaches some server.

2. As it to be expected in the system under FFS policy the orbiting customer passes fewer primary customers. It coincides with the observation that the waiting time distribution of an orbiting customer under this policy lie above other graphs. The policies (OTP, STP) turn out to be the worst with respect to the value of interest that can be also explained by the sufficiently larger waiting time of a customer under threshold policies in comparison with others.

3. For small load factor $\rho^{(m)}$ (the concrete values are given in previous example) the number of directly served primary arrivals is quite small, the concrete examples show that almost surely not more then 2, i.e. $\Theta^{(m)}(3) > 0.99$ (Figure 5(a)) and not more then 1, $\Theta^{(m)}(2) > 0.99$ (Figure 6(a)) primary arrivals will be served directly while the orbiting customer is waiting for the service. As the load factor increases, more customers are served directly, that can be confirmed by the figures observation, namely, $\Theta^{(m)}(7) > 0.99$ (Figure 5(b)) and $\Theta^{(m)}(3) > 0.99$ (Figure 6(b)).
The next Figures 7 (λ = 0.5 and λ = 0.9 in figures labelled, respectively, by letter "a" and "b") and 8 (γ = 2.5 and γ = 8.5 in figures labelled, respectively, by letter "a" and "b") illustrate the discrete distribution functions Ψ(n) for the number of retrials made by an orbiting customer until it reaches the service area. The following conclusions can be done by observing the graphs.

1. At point \( t = 0 \) the jump of the functions corresponds to the case when no retrials will be made by a customer and equals to the probability that the customer will be served directly, i.e. \( \Psi^{(m)}(0) = W^{(m)}(0) \).

2. As \( \lambda \) and \( \gamma \) increase the distributions reveal the heavier tails. Under the FFS and RSS control policies the customer makes less retrials as under threshold policies (OTP, STP) because first two disciplines imply the shorter waiting time.

3. The number of retrials strongly depends on the retrial rate. In case when \( \gamma \) is small, Figure 7(a,b), then the number of retrials with large probability will be not very large, e.g. in this example \( \Psi^{(m)}(6) > 0.99 \) and \( \Psi^{(m)}(7) > 0.99 \). Otherwise, when \( \gamma \) is large, Figure 8(a,b), then the number of retrial significantly increases, e.g. \( \Psi^{(m)}(10) > 0.96 \) and \( \Psi^{(m)}(10) > 0.94 \).
Figure 7: The distribution function $\Psi(n)$ (a) $\lambda = 0.5$ (b) $\lambda = 0.9$, $\mu_1=2.2$, $\mu_2=0.3$, $\gamma = 2.5$

Figure 8: The distribution function $\Psi(n)$ (a) $\lambda = 0.5$ (b) $\lambda = 0.9$, $\mu_1=2.2$, $\mu_2=0.3$, $\gamma = 8.5$

10 Conclusion

The presented in the paper results show that for controlled retrial queues it is also possible to perform quite detailed performance analysis. The presented methods can be extended by increasing the number of servers $c > 2$ and by considering more bursty arrival and service processes. We develop the algorithms for the calculation of the Laplace transforms of the waiting and sojourn time distributions. To get the distribution functions we use the appropriate methods for the inversion of Laplace- and z-transforms based on the Fourier-series methods. It is demonstrated that while the OTP shows the best sojourn time, the FFS policy reveals the best waiting time for the orbiting customer. In heavy traffic case, when the load factor is large, then the differences between the policies can be neglected. In light traffic case the results for the OPT and STP coincide, thus for the optimal threshold levels as approximations one can take the corresponding levels of the scheduling problem.
References


