Performance Evaluation 126 (2018) 22–38

Contents lists available at ScienceDirect

Performance Evaluation

journal homepage: www.elsevier.com/locate/peva

Alternating server with non-zero switch-over times and opposite-queue threshold-based switching policy

Amit Jolles [a], Efrat Perel [b], Uri Yechiali [a,*]

[a] Department of Statistics and Operations Research, School of Mathematical Sciences, Tel-Aviv University, Tel-Aviv, 6997801, Israel
[b] School of Industrial Engineering and Management, Afeka, Tel-Aviv Academic College of Engineering, Tel-Aviv, Israel

ARTICLE INFO

Article history:
Received 2 January 2018
Received in revised form 14 June 2018
Accepted 11 July 2018
Available online 24 July 2018

Keywords:
Alternating server
Non-zero switch-over times
Opposite-queue threshold policy
Derivation of the rate matrix R

ABSTRACT

A single server alternates between two Markovian queues with *non-zero* switch-over times. The server's switching instants are determined by the number of customers accumulated at the *unattended* queue. Specifically, when queue i ($i = 1, 2$) is attended and the number of customers in queue j ($j = 1, 2; j \neq i$) reaches a threshold, the server starts an exponentially distributed switch-over time to queue j, unless the number of customers in queue i is equal to or above queue i's threshold. However, if during a switch-over period from queue i to queue j the former reaches its threshold, the switch-over is **aborted**, and the server immediately returns to queue i and continues to serve the customers there. We analyze the system mainly via Matrix Geometric (MG) methods while deriving explicitly the rate matrix R, and thus eliminating the need for successive substitutions. We further reveal connections between the entries of R and the roots of polynomials related to the Probability Generating Functions (PGFs) of the system states. Expressions for the system's performance measures are obtained (e.g. mean queue size and mean sojourn time in queue 1, PGF and mean of the queue size in queue 2, as well as the Laplace Stieltjes transform and mean of the sojourn time in queue 2). Numerical results are presented and the effects of the various parameters, as well as the switch-over times, on the performance measures are examined. Seemingly counter-intuitive phenomena are discussed. Finally, various extreme cases are investigated.

1. Introduction

N-queue polling systems with a single server and switch-over times have been studied extensively in the queueing literature (see e.g. Takagi [1], Boxma and Groenendijk [2], Boxma, Levy and Yechiali [3], Browne and Yechiali [4], Yechiali [5], Resing [6] and many others). A recent survey (Boon, Van der Mei and Winands [7]) discussed a vast list of polling systems applications. Several papers focused on two-queue alternating-server system with *zero* switch-over times (see e.g. Takács [8], Boxma and Down [9] and Boxma, Schlegel and Yechiali [10]). Threshold-based systems, mostly depending on the queue level of the attended queue, were also investigated (see e.g. Lee [11], Lee and Sengupta [12], Haverkort, Idzenga and Kim [13], Boxma, Koole and Mitrani [14,15], Avram and Gómez-Corral [16] and many others).

Usually, the switching instants in the *non-zero* switch-over papers are determined by the occupancy level of the queue being attended by the server. Recently, in deviation, Avrachenkov, Perel and Yechiali [17] and Perel and Yechiali [18] studied two-queue polling systems with threshold-based switching policy determined mainly by the number of customers in the

* Corresponding author.
E-mail addresses: amitjolles@gmail.com (A. Jolles), efratp@afeka.ac.il (E. Perel), uriy@post.tau.ac.il (U. Yechiali).

https://doi.org/10.1016/j.peva.2018.07.004

unattended queue, but with *zero* switch-over times. In this work we generalize the model studied in [18] by including *non-zero* switch-over times. This expansion makes the model more realistic but also raises significantly the complexity of the probabilistic analysis: it requires doubling of the state space, which leads to expanded and non-symmetric steady-state equations, and yields some counter-intuitive results. In addition, we assume that if during a switch-over period from queue i to queue j ($i, j = 1, 2$; $i \neq j$) the former reaches its threshold, the switch-over is **aborted**, and the server immediately continues to serve queue i. This assumption is made in order to avoid additional analytical complication.

A notable example for such a system is a traffic light in an intersection that alternates right-of-way priority (see Meilijson and Yechiali [19]) according to the number of cars waiting in red. The duration of a yellow color is equivalent to a switch-over time that may be cut short. This involved switching procedure may also be considered as a non-cyclic polling system with state-dependent polling table. For determining efficient visit-order tables in polling systems see Boxma, Levy and Weststrate [20], and Wal v.d. and Yechiali [21]. Another example is the occupancy control of disks in data centers. When the amount of data on a given disk becomes large, causing an inefficient operation, the disk requires a clean-up action. Thirdly, this policy is suitable for queues with customers deadlines. A large unattended queue signals that many waiting customers there may miss their deadlines. Such a situation calls for the server's attention.

Our contribution is 3-fold: (*i*) We derive the joint probability distribution function of the queue sizes. (*ii*) When employing Matrix Geometric (MG) analysis, we obtain explicitly all the entries of the rate matrix R (which is the corner-stone of the MG analysis), thus eliminating the need to obtain R via successive substitutions. (*iii*) We reveal the connection between the entries of R and the roots of two matrices associated with the model related Probability Generating Functions (PGFs) defined in Section 3.1.

The structure of the paper is as follows: In Section 2 the model is described in detail. In Section 3 the system is defined as a three-dimensional QBD process and a Matrix Geometric approach is employed to derive the system's steady-state probabilities. It turns out that the elements of R are closely related to the roots of $|B(z)|$ and $|C(z)|$, where $B(z)$ and $C(z)$ are two matrices that stem from the above mentioned PGF approach. Notably, in Section 4 we express *explicitly* all the entries of the rate matrix R, thus allowing efficient calculation of the system's steady-state probabilities. In Section 5 we present numerical results and reveal a counter-intuitive behavior of the system's performance measures. In Section 6 we analyze extreme cases and discuss their implications, while Section 7 concludes the paper.

2. Model description

We study a two-queue Markovian system with a single alternating server, where the decisions on when to switch from an attended queue to its counterpart are determined by the queue size of the latter, and is based on a threshold policy. Furthermore, we consider the case where switch-over times are *non-zero*. Specifically, whenever the server attends queue i ($i = 1, 2$), it serves the customers there until the queue size in the opposite queue reaches its threshold level. At that instant the server starts a *non-zero* switch-over period to queue j ($j \neq i$), unless the number of customers in queue i is greater than or equal to its own threshold level. In the latter case the server remains in queue i until the number of customers there is reduced below its threshold level, and only then it starts switching to queue j. When a served queue is emptied while the other queue is not, the server immediately starts a switch-over period. If during the switch-over time from queue i to queue j the former reaches its threshold, the switch-over is **aborted**, and the server immediately switches back to queue i and continues to serve the customers there. Customers arrive to queue i ($i = 1, 2$) according to a Poisson process with rate λ_i, and the service time for each individual customer is exponentially distributed with mean $1/\mu_i$. Switch-over times in either direction are exponentially distributed with parameter α. All the above processes are mutually independent. The threshold levels are K for queue 1, and N for queue 2. Queue 2 is an $M/M/1$ system with an unlimited buffer, whereas Queue 1 is a limited buffer $M/M/1/C_1$ system with finite buffer $C_1 \geq K$. We treat the case $K = C_1$, where new arrivals to queue 1 are blocked and lost when the queue size is K. The case where $K < C_1 < \infty$ is similar but involves more equations and therefore will not be presented. Specifically, the matrices appearing in the generator matrix Q (to be defined shortly) will be of larger size. For example, the square matrices A_0, A_1 and A_2 will be of order $(2C_1 + 2)$, rather than of order $(2K + 2)$. We note that the assumption that C_1 is finite is imposed for tractability purpose. Let L_i denote the number of customers present in queue i ($i = 1, 2$) in steady-state (it will be shown that the system's stability condition is $\lambda_2 < \mu_2$). Let $I = 1$ if the server attends queue 1; $I = 2$ if the server attends queue 2; $I = S1$ if the server is in a switch-over move from queue 1 to queue 2, while $I = S2$ if the server is in a switch-over move from queue 2 to queue 1. The triple (L_1, L_2, I) defines a non reducible continuous-time Markov chain. The transition-rate diagram of the system's states is depicted in Fig. 2.1. Each box (k, n) there represents the four possible states (k, n, I) for $I = 1$, $I = 2$, $I = S1$ or $I = S2$. Let $P_{kn}(i) = \mathbb{P}(L_1 = k, L_2 = n, I = i)$, where $0 \leq k \leq K$; $0 \leq n$; $i = 1, 2, S1, S2$.

3. Matrix geometric

In this section we use Matrix Geometric methodology to derive the probability distribution function of the system's state, $\{P_{kn}(i)\}_{0 \leq k \leq K,\, 0 \leq n,\, i=1,2,S1,S2}$. We construct a quasi birth-and-death (QBD) process (Neuts [22], Latouche and Ramaswami [23]) with an infinite state space S under the order:

$S = \{(0, 0, 1), (0, 0, 2), (1, 0, 1), (1, 0, S2), (2, 0, 1), (2, 0, S2), \ldots, (K, 0, 1), (K, 0, S2);$
$(0, 1, 2), (0, 1, S1), (1, 1, 1), (1, 1, 2), (1, 1, S1), (1, 1, S2), \ldots, (K-1, 1, 1), (K-1, 1, 2), (K-1, 1, S1), (K-1, 1, S2),$

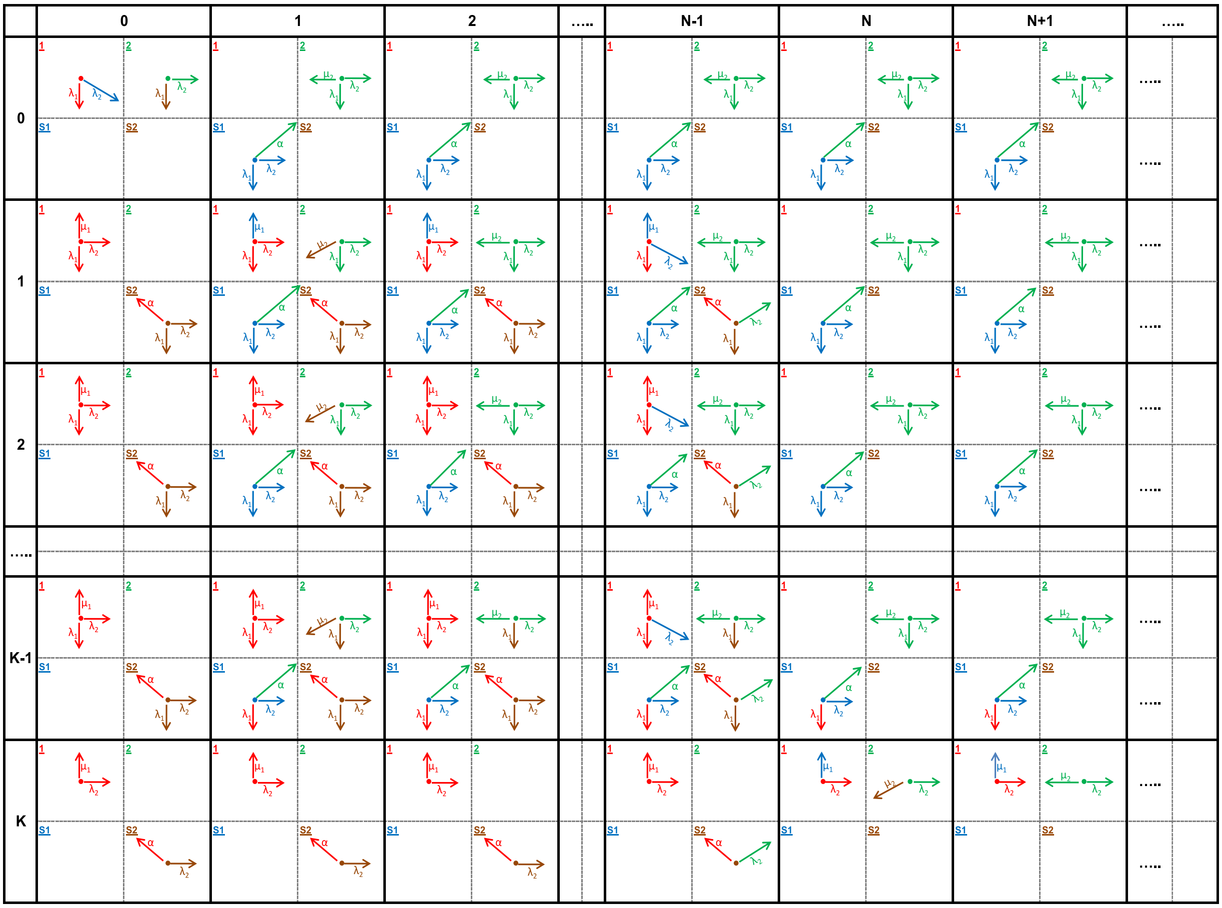


Fig. 2.1. Transition rate diagram of (L_1, L_2, I). The indicators 1, 2, $S1$ or $S2$ appearing in each cell indicate whether $I = 1, I = 2, I = S1$ or $I = S2$, respectively. The arrow colors *Red*, *Green*, *Blue* or *Brown* indicate the transition's target state: $I = 1, I = 2, I = S1$ or $I = S2$, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$(K, 1, 1), (K, 1, S2); \dots;$
$(0, N-1, 2), (0, N-1, S1), (1, N-1, 1), (1, N-1, 2), (1, N-1, S1), (1, N-1, S2), \dots,$
$(K-1, N-1, 1), (K-1, N-1, 2), (K-1, N-1, S1), (K-1, N-1, S2), (K, N-1, 1), (K, N-1, S2);$
$(0, N, 2), (0, N, S1), (1, N, 2), (1, N, S1), \dots, (K-1, N, 2), (K-1, N, S1), (K, N, 1), (K, N, 2);$
$(0, N+1, 2), (0, N+1, S1), (1, N+1, 2), (1, N+1, S1), \dots, (K-1, N+1, 2), (K-1, N+1, S1), (K, N+1, 1), (K, N+1, 2); \dots\}.$
The generator matrix Q is given by

$$Q = \begin{pmatrix}
B_1^0 & B_0^0 & \mathbf{0} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
B_2^1 & B_1 & B_0 & \mathbf{0} & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
\mathbf{0} & B_2 & B_1 & B_0 & \mathbf{0} & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
\vdots & \ddots & \mathbf{0} & B_2 & B_1 & B_0 & \mathbf{0} & \ddots & \ddots & \ddots & \ddots & \ddots \\
\vdots & \ddots & \ddots & \mathbf{0} & B_2 & B_1 & B_0^{N-1} & \mathbf{0} & \ddots & \ddots & \ddots & \ddots \\
\vdots & \ddots & \ddots & \ddots & \mathbf{0} & A_2^N & A_1 & A_0 & \mathbf{0} & \ddots & \ddots & \ddots \\
\vdots & \ddots & \ddots & \ddots & \ddots & \mathbf{0} & A_2 & A_1 & A_0 & \mathbf{0} & \ddots & \ddots \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \mathbf{0} & A_2 & A_1 & A_0 & \mathbf{0} & \ddots \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots
\end{pmatrix},$$

where **0** is a matrix of zeros, and the matrices on the upper diagonal are: B_0^0 is of size $(2K+2)\times 4K$, B_0 is of size $4K\times 4K$, B_0^{N-1} is of size $4K\times(2K+2)$ and A_0 is of size $(2K+2)\times(2K+2)$. The matrices on the main diagonal are: B_1^0 is of size $(2K+2)\times(2K+2)$, B_1 is of size $4K\times 4K$ and A_1 is of size $(2K+2)\times(2K+2)$. The matrices on the lower diagonal are: B_2^1 is of size $4K\times(2K+2)$, B_2 is of size $4K\times 4K$, A_2^N is of size $(2K+2)\times 4K$, and A_2 is of size $(2K+2)\times(2K+2)$. The matrices A_0, A_1 and A_2 are detailed below while the rest of the matrices are detailed in Appendix.

$$A_0 = diag(\lambda_2),$$

$$A_1 = \begin{pmatrix}
-\beta_5 & 0 & \lambda_1 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0\\
\alpha & -\beta_2 & 0 & \lambda_1 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0\\
0 & 0 & -\beta_5 & 0 & \lambda_1 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0\\
0 & 0 & \alpha & -\beta_2 & 0 & \lambda_1 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0\\
0 & 0 & 0 & 0 & -\beta_5 & 0 & \lambda_1 & 0 & \cdots & \cdots & \cdots & \cdots & 0\\
0 & 0 & 0 & 0 & \alpha & -\beta_2 & 0 & \lambda_1 & 0 & \cdots & \cdots & \cdots & 0\\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots\\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots\\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots\\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & -\beta_5 & 0 & 0 & \lambda_1\\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \alpha & -\beta_2 & \lambda_1 & 0\\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \mu_1 & -\beta_3 & 0\\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & -\beta_6
\end{pmatrix},$$

where $\beta_2 = \lambda_1+\lambda_2+\alpha$; $\beta_3 = \lambda_2+\mu_1$; $\beta_5 = \lambda_1+\lambda_2+\mu_2$; and $\beta_6 = \lambda_2+\mu_2$.

$$A_2 = \begin{pmatrix}
\mu_2 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0\\
0 & 0 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \vdots\\
0 & 0 & \mu_2 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \vdots\\
0 & 0 & 0 & 0 & 0 & \cdots & \cdots & \cdots & \cdots & \vdots\\
0 & 0 & 0 & 0 & \mu_2 & 0 & \cdots & \cdots & \cdots & \vdots\\
0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots\\
0 & \ddots & \ddots & \ddots & \ddots & \ddots & \mu_2 & 0 & \ddots & \vdots\\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \vdots\\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0\\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & \mu_2
\end{pmatrix}.$$

Let $A = A_0 + A_1 + A_2$. Then,

$$A = \begin{pmatrix}
-\lambda_1 & 0 & \lambda_1 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0\\
\alpha & -(\lambda_1+\alpha) & 0 & \lambda_1 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \vdots\\
0 & 0 & -\lambda_1 & 0 & \lambda_1 & 0 & \cdots & \cdots & \cdots & \cdots & \vdots\\
0 & 0 & \alpha & -(\lambda_1+\alpha) & 0 & \lambda_1 & \cdots & \cdots & \cdots & \cdots & \vdots\\
0 & 0 & 0 & 0 & -\lambda_1 & 0 & \lambda_1 & 0 & \cdots & \cdots & \vdots\\
0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots\\
0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots\\
0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & -\lambda_1 & 0 & 0 & \lambda_1\\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \alpha & -(\lambda_1+\alpha) & \lambda_1 & 0\\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \mu_1 & -\mu_1 & 0\\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0
\end{pmatrix}.$$

The matrix A represents the infinitesimal generator of a specific continuous time Markov-chain with absorbing state at point $2K+1$. Indeed, let $\vec{\pi} = (\pi_0, \pi_1, \ldots, \pi_{2K+1})$ be the stationary probability vector of the matrix A, i.e. $\vec{\pi}A = 0$ and $\vec{\pi}\vec{e} = 1$, where $\vec{0}$ is a vector of 0's and $\vec{e}$ is a vector of 1's. From $\vec{\pi}A = \vec{0}$ we have that $\vec{\pi} = (\underbrace{0, 0, \dots, 0}_{2K+1 \text{ times}}, \pi_{2K+1})$. From $\vec{\pi}\vec{e} = 1$ we have that $\pi_{2K+1} = 1$, namely, $\vec{\pi} = (\underbrace{0, 0, \dots, 0}_{2K+1 \text{ times}}, 1)$. Hence, the stability condition $\vec{\pi}A_0\vec{e} < \vec{\pi}A_2\vec{e}$ (see [22]) becomes:

$$\lambda_2 < \mu_2. \tag{3.1}$$

Define the steady-state probability vectors:

$$\begin{aligned}
\vec{P}_0 &= (P_{00}(1), P_{00}(2), P_{10}(1), P_{10}(S2), \dots, P_{K0}(1), P_{K0}(S2)),\\
\vec{P}_n &= (P_{0n}(2), P_{0n}(S1), P_{1n}(1), P_{1n}(2), P_{1n}(S1), P_{1n}(S2), \dots, P_{K-1,n}(1),\\
&\quad P_{K-1,n}(2), P_{K-1,n}(S1), P_{K-1,n}(S2), P_{Kn}(1), P_{Kn}(S2)), \qquad 1 \le n \le N-1,\\
\vec{P}_n &= (P_{0n}(2), P_{0n}(S1), \dots, P_{K-1,n}(2), P_{K-1,n}(S1), P_{Kn}(1), P_{Kn}(2)), \qquad n \ge N.
\end{aligned}$$

Then,

$$\vec{P}_n = \vec{P}_N R^{n-N}, \quad n \ge N, \tag{3.2}$$

where R is the minimal non-negative solution of the matrix quadratic equation [22,23]

$$A_0 + RA_1 + R^2A_2 = \mathbf{0}. \tag{3.3}$$

In most cases of Matrix Geometric analysis, the matrix R is obtained via successive substitutions. However, in this study we are able to derive explicitly all the entries of R, thus reducing considerably the computational efforts. The expressions for the entries of R are given in Section 4.

The vectors $\vec{P}_0, \vec{P}_1, \dots, \vec{P}_N$, are the solution of the following linear system of equations:

$$\begin{aligned}
&\vec{P}_0B_1^0 + \vec{P}_1B_2^1 = \vec{0}\\
&\vec{P}_0B_0^0 + \vec{P}_1B_1 + \vec{P}_2B_2 = \vec{0}\\
&\vec{P}_{n-1}B_0 + \vec{P}_nB_1 + \vec{P}_{n+1}B_2 = \vec{0}, \quad 2 \le n \le N-2\\
&\vec{P}_{N-2}B_0 + \vec{P}_{N-1}B_1 + \vec{P}_NA_2^N = \vec{0}\\
&\vec{P}_{N-1}B_0^{N-1} + \vec{P}_NA_1 + \vec{P}_{N+1}A_2 = \vec{0}\\
&\sum_{n=0}^{N-1}\vec{P}_n\vec{e} + \vec{P}_N[\mathcal{I} - R]^{-1}\vec{e} = 1
\end{aligned}$$

where $\mathcal{I}$ is the identity matrix.

$\mathbb{E}[L_i]$, the mean total number of customers in queue i (Q_i), $i = 1, 2$, is given by

$$\mathbb{E}[L_1] = \vec{P}_0\vec{Z}_1 + \sum_{n=1}^{N-1}\vec{P}_n\vec{Z}_2 + \sum_{n=N}^{\infty}\vec{P}_n\vec{Z}_1 = \vec{P}_0\vec{Z}_1 + \sum_{n=1}^{N-1}\vec{P}_n\vec{Z}_2 + \vec{P}_N[\mathcal{I} - R]^{-1}\vec{Z}_1 \tag{3.4}$$

$$\begin{aligned}
\mathbb{E}[L_2] &= \sum_{n=1}^{\infty} n\vec{P}_n\vec{e} = \sum_{n=1}^{N-1} n\vec{P}_n\vec{e} + \sum_{n=N}^{\infty} n\vec{P}_N R^{n-N}\vec{e}\\
&= \sum_{n=1}^{N-1} n\vec{P}_n\vec{e} + N\vec{P}_N[\mathcal{I} - R]^{-1}\vec{e} + \vec{P}_N R[\mathcal{I} - R]^{-2}\vec{e}
\end{aligned} \tag{3.5}$$

where, $\vec{Z}_1 = (0, 0, 1, 1, 2, 2, \dots, K-1, K-1, K, K)^T$ and $\vec{Z}_2 = (0, 0, 1, 1, 1, 1, 2, 2, 2, 2, \dots, K-1, K-1, K-1, K-1, K, K)^T$.

Define $P_{K\bullet} = P_{K\bullet}(1) + P_{K\bullet}(2) + P_{K\bullet}(S2)$. $P_{K\bullet}$ is the probability of having K customers in Q_1, and is the loss probability of type 1 customers, P_{loss}.

Then, by Little's law, the mean sojourn time of a customer in Q_i, $\mathbb{E}[W_i]$, $i = 1, 2$ is

$$\mathbb{E}[W_1] = \frac{\mathbb{E}[L_1]}{\lambda_1^{eff}} \tag{3.6}$$

$$\mathbb{E}[W_2] = \frac{\mathbb{E}[L_2]}{\lambda_2}, \tag{3.7}$$

where,

$$\lambda_1^{eff} = \lambda_1(1 - P_{loss}) = \lambda_1 \left(1 - \sum_{n=0}^{N-1} (P_{Kn}(1) + P_{Kn}(S2)) - \sum_{n=N}^{\infty} (P_{Kn}(1) + P_{Kn}(2)) \right)$$
$$= \lambda_1 \left(1 - \sum_{n=0}^{N-1} (P_{Kn}(1) + P_{Kn}(S2)) - \vec{P}_N [\mathcal{I} - R]^{-1} \vec{v} \right)$$

with $\vec{v} = (\underbrace{0, 0, \dots, 0}_{2K \text{ times}}, 1, 1)^t$.

Let R_{lm}, for $1 \le l \le 2K + 2$ and $1 \le m \le 2K + 2$, denote the elements of the matrix R. Due to the structure of the matrices A_0, A_1 and A_2, the matrix R is almost an upper triangular matrix, with only $K + 2$ non-zero elements beneath the main diagonal, $R_{2k,2k-1}$, for all $1 \le k \le K$, $R_{2K+1,2K-1}$ and $R_{2K+1,2K}$. Therefore, by solving Eq. (3.3) we derive *closed form expressions* for the elements of R. It will be shown in the sequel that the entries of R are related to the roots of $|B(z)|$ and $|C(z)|$, where $B(z)$ and $C(z)$ are defined in the following sub-section.

Remark 3.1. Consider $\vec{P}_n$, $n \ge 0$, as defined in Section 3, let $P(L_2 = n) = \sum_{k=0}^{K} (\sum_i P_{kn}(i)) = \vec{P}_n \vec{e}$ (with some $P_{kn}(i) = 0$, see Fig. 2.1). Applying the argument leading to the distributional form of Little's law, namely that the customers left behind a departing customer are those that arrived during the latter's sojourn time, W_2, we get the Laplace Stieltjes transform (LST) of W_2, denoted $\tilde{W}_2(\cdot)$, in terms of the PGF of L_2:

$$\hat{L_2}(z) \equiv \mathbb{E}[z^{L_2}] = \sum_{n=0}^{\infty} P(L_2 = n) z^n = \sum_{n=0}^{\infty} \left(\vec{P}_n \vec{e} \right) z^n = \sum_{n=0}^{\infty} \left(\int_{t=0}^{\infty} e^{-\lambda_2 t} \frac{(\lambda_2 t)^n}{n!} f_{W_2}(t) dt \right) z^n = \tilde{W}_2\left(\lambda_2 (1 - z) \right).$$

Now,

$$\hat{L_2}(z) = \sum_{n=0}^{N-1} (\vec{P}_n \vec{e}) z^n + \sum_{n=N}^{\infty} (\vec{P}_n \vec{e}) z^n = \sum_{n=0}^{N-1} (\vec{P}_n \vec{e}) z^n + \sum_{n=N}^{\infty} (\vec{P}_N R^{n-N}) \vec{e} z^n$$
$$= \sum_{n=0}^{N-1} (\vec{P}_n \vec{e}) z^n + \vec{P}_N z^N \sum_{n=N}^{\infty} (zR)^{n-N} \vec{e} = \sum_{n=0}^{N-1} (\vec{P}_n \vec{e}) z^n + \vec{P}_N z^N [I - zR]^{-1} \vec{e}.$$

3.1. Probability generating functions

We define four sets of PGFs:

For $I = 1$,

$$G_k(z) = \begin{cases} \sum_{n=0}^{N-1} P_{kn}(1) z^n, & 1 \le k \le K-1, \qquad |z| < \infty \\ \sum_{n=0}^{\infty} P_{kn}(1) z^n, & k = K, \qquad |z| \le 1. \end{cases}$$

For $I = 2$ and for all $|z| \le 1$,

$$F_k(z) = \begin{cases} \sum_{n=0}^{\infty} P_{kn}(2) z^n, & k = 0 \\ \sum_{n=1}^{\infty} P_{kn}(2) z^n, & 1 \le k \le K-1 \\ \sum_{n=N}^{\infty} P_{kn}(2) z^n, & k = K. \end{cases}$$

In the same manner, for $I = S1$,

$$H_k(z) = \sum_{n=1}^{\infty} P_{kn}(S1) z^n, \quad 0 \le k \le K-1, \qquad |z| \le 1.$$

Finally, for $I = S2$,

$$T_k(z) = \sum_{n=0}^{N-1} P_{kn}(S2) z^n, \quad 1 \le k \le K, \qquad |z| < \infty.$$

By writing the Balance Equations, multiplying each equation by z^n, summing over n and rearranging the terms, we obtain four sets of linear equations where the unknowns are the sought for PGFs:

For $I = 1$, we construct a set of linear equations of the form

$$A(z)\vec{G}(z) = \vec{P}(z), \tag{3.8}$$

where the K-dimensional column vectors $\vec{G}(z)$ and $\vec{P}(z) = (P_1(z), P_2(z), \dots, P_K(z))^t$, and the matrix $A(z)_{K\times K}$ are defined as follows:

$$\vec{G}(z) = (G_1(z), G_2(z), \dots, G_K(z))^t,$$

$$P_k(z) = \begin{cases} \alpha T_2(z) - \lambda_2 z P_{1,N-1}(1)z^{N-1} + \lambda_1 P_{00}(1), & k = 1 \\ \alpha T_k(z) - \lambda_2 z P_{k,N-1}(1)z^{N-1}, & 2 \le k \le K-2 \\ \alpha T_{K-1}(z) - \lambda_2 z P_{K-1,N-1}(1)z^{N-1} + \mu_1 \sum_{n=0}^{N-1} P_{Kn}(1)z^n, & k = K-1 \\ \alpha T_K(z) + \lambda_1 H_{K-1}(z), & k = K \end{cases}$$

and

$$A(z) = \begin{pmatrix} \alpha(z) & -\mu_1 & 0 & \cdots & \cdots & \cdots & 0 \\ -\lambda_1 & \alpha(z) & -\mu_1 & 0 & \cdots & \cdots & 0 \\ 0 & -\lambda_1 & \alpha(z) & -\mu_1 & 0 & \cdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\lambda_1 & \alpha(z) & -\mu_1 & 0 \\ 0 & \ddots & \ddots & \ddots & -\lambda_1 & \alpha(z) & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & -\lambda_1 & \alpha_K(z) \end{pmatrix},$$

where,

$$\alpha(z) = \lambda_1 + \mu_1 + \lambda_2(1-z),$$
$$\alpha_K(z) = \mu_1 + \lambda_2(1-z).$$

Similarly, for $I = 2$,

$$B(z)\vec{F}(z) = \vec{Q}(z), \tag{3.9}$$

where the $(K+1)$-dimensional column vectors $\vec{F}(z)$ and $\vec{Q}(z) = (Q_0(z), Q_1(z), \dots, Q_K(z))^t$, and the matrix $B(z)_{(K+1)\times(K+1)}$ are defined as follows:

$$\vec{F}(z) = (F_0(z), F_1(z), \dots, F_K(z))^t,$$

$$Q_k(z) = \begin{cases} \alpha H_0(z) + \mu_2(1 - \frac{1}{z})P_{00}(2), & k = 0 \\ -\mu_2 \frac{1}{z} P_{11}(2)z - \lambda_1 P_{00}(2) + \alpha H_1(z) + \lambda_2 P_{1,N-1}(S2)z^N, & k = 1 \\ -\mu_2 \frac{1}{z} P_{k1}(2)z + \alpha H_1(z) + \lambda_2 P_{k,N-1}(S2)z^N, & 2 \le k \le K-1 \\ -\lambda_1 \sum_{n=1}^{N-1} P_{K-1,n}(2)z^n - \mu_2 \frac{1}{z} P_{KN}(2)z^N + \lambda_2 P_{K,N-1}(S2)z^N, & k = K \end{cases}$$

and

$$B(z) = \begin{pmatrix} \beta(z) & 0 & 0 & \cdots & \cdots & 0 \\ -\lambda_1 & \beta(z) & 0 & 0 & \cdots & 0 \\ 0 & -\lambda_1 & \beta(z) & 0 & 0 & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & -\lambda_1 & \beta(z) & 0 \\ 0 & \cdots & \cdots & 0 & -\lambda_1 & \beta_K(z) \end{pmatrix},$$

where,

$$\beta(z) = \lambda_1 + \lambda_2(1-z) + \mu_2(1-\frac{1}{z}),$$
$$\beta_K(z) = \lambda_2(1-z) + \mu_2(1-\frac{1}{z}).$$

In the same manner, for $I = S1$, we get

$$C(z)\vec{H}(z) = \vec{R}(z), \tag{3.10}$$

where the K-dimensional column vectors $\vec{H}(z)$ and $\vec{R}(z)$, and the matrix $C(z)_{K\times K}$ are the following:

$$\vec{H}(z) = (H_0(z), H_1(z), \dots, H_{K-1}(z))^t,$$

$$R_k(z) = \begin{cases} \mu_1 F_1(z) - \mu_1 P_{10}(1) + \lambda_2 P_{00}(1)z, & k=0 \\ \lambda_2 P_{k,N-1}(1)z^N, & 1 \le k \le K-2 \\ \mu_1 F_K(z) - \mu_1 \sum_{n=0}^{N-1} P_{K,n}(1)z^n + \lambda_2 P_{K-1,N-1}(1)z^N, & k = K-1 \end{cases}$$

and

$$C(z) = \begin{pmatrix} \gamma(z) & 0 & 0 & \cdots & \cdots & 0 \\ -\lambda_1 & \gamma(z) & 0 & 0 & \cdots & 0 \\ 0 & -\lambda_1 & \gamma(z) & 0 & 0 & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & -\lambda_1 & \gamma(z) & 0 \\ 0 & \cdots & \cdots & 0 & -\lambda_1 & \gamma(z) \end{pmatrix},$$

where,

$$\gamma(z) = \lambda_1 + \lambda_2(1-z) + \alpha.$$

Finally, for $I = S2$, we have

$$D(z)\vec{T}(z) = \vec{V}(z), \tag{3.11}$$

where the column vectors $\vec{T}(z)$ and $\vec{V}(z)$, and the matrix $D(z)_{K\times K}$ are the following:

$$\vec{T}(z) = (T_1(z), T_2(z), \dots, T_K(z))^t,$$

$$V_k(z) = \begin{cases} -\lambda_2 z P_{1,N-1}(S2)z^{N-1} + \lambda_1 P_{00}(2) + \mu_2 P_{11}(2), & k=1 \\ -\lambda_2 z P_{k,N-1}(S2)z^{N-1} + \mu_2 P_{k1}(2), & 2 \le k \le K-1 \\ -\lambda_2 z P_{K,N-1}(S2)z^{N-1} + \lambda_1 \sum_{n=1}^{N-1} P_{K-1,n}(2)z^n + \mu_2 P_{K,N}(2)z^{N-1}, & k=K \end{cases}$$

and

$$D(z) = \begin{pmatrix} \delta(z) & 0 & 0 & \cdots & \cdots & 0 \\ -\lambda_1 & \delta(z) & 0 & 0 & \cdots & 0 \\ 0 & -\lambda_1 & \delta(z) & 0 & 0 & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & -\lambda_1 & \delta(z) & 0 \\ 0 & \cdots & \cdots & 0 & -\lambda_1 & \delta_K(z) \end{pmatrix},$$

where,

$$\delta(z) = \lambda_1 + \lambda_2(1-z) + \alpha,$$
$$\delta_K(z) = \lambda_2(1-z) + \alpha.$$

We first discuss the roots of each of the polynomials $|A(z)|$ and $|B(z)|$.

Theorem 3.1. *Given $\lambda_1, \lambda_2, \mu_1 > 0$, for $K \geq 1$ the polynomial $|A(z)|$ is of degree K and possesses K distinct roots in $(1, \infty)$. One of which is $z_K = 1 + \frac{\mu_1}{\lambda_2}$.*

Theorem 3.2. *Given $\lambda_1, \lambda_2, \mu_2 > 0$, for $K \geq 1$ the polynomial $|B(z)|$ is of degree $2(K+1)$. It has a root at $z^* = 1$, a root of multiplicity K, $z_1 = \frac{\lambda_2+\mu_2+\lambda_1-\sqrt{(\lambda_2+\mu_2+\lambda_1)^2-4\lambda_2\mu_2}}{2\lambda_2}$ in $(0, 1)$, another root of multiplicity K, $z_2 = \frac{\lambda_2+\mu_2+\lambda_1+\sqrt{(\lambda_2+\mu_2+\lambda_1)^2-4\lambda_2\mu_2}}{2\lambda_2}$ in $(1, \infty)$, and a single root, $z_3 = \frac{\mu_2}{\lambda_2}$ in $(0, 1)$ iff $\lambda_2 > \mu_2$.*

The proof of Theorem 3.1, based on an interlacing argument regarding the roots of $|A(z)|$, and the proof of Theorem 3.2 are similar to proofs given in [18], and henceforth omitted.

Note that the root z_1 in Theorem 3.2 is the Laplace Stieltjes transform of the busy period of an $M/M/1$ queue with arrival rate λ_2 and service rate μ_2, evaluated at λ_1. It also expresses the probability that the duration of a busy period in a regular $M(\lambda_2)/M(\mu_2)/1$ queue will fall short of the inter-arrival time at Q_1.

We now investigate the roots of $|C(z)|$.

Theorem 3.3. *For any $\lambda_1 > 0$, $\lambda_2 > 0$, $\alpha > 0$ and $K \geq 1$, $|C(z)|$ is a polynomial of degree K, possessing a root of multiplicity K, z_4, in the open interval $(1, \infty)$.*

Proof. The matrix $C(z)$ possesses non-zero elements on the main diagonal and on the lower main diagonal. All other entries are 0. Therefore,

$$|C(z)| = (\gamma(z))^K, \tag{3.12}$$

The polynomial $\gamma(z)$ has only one root: $z_4 = \frac{\lambda_1+\lambda_2+\alpha}{\lambda_2} > 1$. Therefore, z_4 is a root of $|C(z)|$, of multiplicity K, in the open interval $(1, \infty)$. □

Finally, we consider $|D(z)|$.

Theorem 3.4. *For any $\lambda_1 > 0$, $\lambda_2 > 0$, $\alpha > 0$ and $K \geq 1$, $|D(z)|$ is a polynomial of degree K, possessing a root of multiplicity $K-1$, $z_4 = \frac{\lambda_1+\lambda_2+\alpha}{\lambda_2} > 1$, in the open interval $(1, \infty)$, and another root $z_5 = \frac{\lambda_2+\alpha}{\lambda_2} > 1$, also in the open interval $(1, \infty)$.*

Proof. The matrix $D(z)$ possesses non-zero elements on the main diagonal and on the lower main diagonal. All other entries are 0. Therefore,

$$|D(z)| = (\delta(z))^{K-1}\delta_K(z), \tag{3.13}$$

The polynomial $\delta(z)$ has only one root: $z_4 = \frac{\lambda_1+\lambda_2+\alpha}{\lambda_2} > 1$. Therefore, z_4 is a root of $|D(z)|$, of multiplicity $K-1$, in the open interval $(1, \infty)$. The polynomial $\delta_K(z)$ has also only one root $z_5 = \frac{\lambda_2+\alpha}{\lambda_2} > 1$. Therefore, z_5 is a root of $|D(z)|$, in the open interval $(1, \infty)$. □

4. Explicit expressions for the entries of the rate matrix *R*

Below we present the outcome when solving Eq. (3.3) (with $\sum_{i=1}^{0}(\cdot) \triangleq 0$):

$$R_{2k-1,2k-1} = \frac{2\lambda_2}{\lambda_2+\mu_2+\lambda_1+\sqrt{(\lambda_2+\mu_2+\lambda_1)^2-4\lambda_2\mu_2}} = \frac{1}{z_2}, \text{ for } 1 \leq k \leq K,$$
$$R_{2k,2k} = \frac{\lambda_2}{\lambda_1+\lambda_2+\alpha} = \frac{1}{z_4}, \text{ for } 1 \leq k \leq K-1,$$
$$R_{2K,2K} = \frac{\lambda_2\beta_3}{\beta_2\beta_3-\lambda_1\mu_1},$$
$$R_{2K+1,2K+1} = \frac{\lambda_2\beta_2}{\beta_2\beta_3-\lambda_1\mu_1},$$
$$R_{2K+2,2K+2} = \frac{\lambda_2}{\mu_2} = \frac{1}{z_3},$$

$$R_{1,2k-1} = \frac{\lambda_1 R_{1,2k-3} + \mu_2 \sum_{i=1}^{k-2} R_{1,2i+1} R_{1,2(k-i)-1}}{\beta_5 - 2\mu_2/z_2}, \text{ for } 2 \le k \le K,$$

$$R_{2k-1,2(k+j)-1} = R_{1,2j+1}, \ 2 \le k \le K\ ;\ 1 \le j \le K-k,$$

$$R_{2,2k} = R_{22}\left(\frac{\lambda_1}{\beta_2}\right)^{k-1} = \left(\frac{1}{z_4}\right)\left(\frac{\lambda_1}{\lambda_2 z_4}\right)^{k-1}, \text{ for } 2 \le k \le K-1,$$

$$R_{21} = \frac{\alpha R_{22}}{\beta_5 - \mu_2\left(\frac{1}{z_2} + \frac{1}{z_4}\right)},$$

$$R_{2,2k-1} = \frac{\lambda_1 R_{2,2k-3} + \alpha R_{2,2k} + \mu_2\left(R_{21}\left(R_{1,2k-1} + R_{2,2k}\right) + \sum_{i=1}^{k-2}\left(R_{2,2i+1}R_{1,2(k-i)-1} + R_{2,2(i+1)}R_{2,2(k-i)-1}\right)\right)}{\beta_5 - \mu_2\left(\frac{1}{z_2} + \frac{1}{z_4}\right)},$$

for $2 \le k \le K-1$,

$$R_{2k,j} = R_{2,j-2(k-1)}, \ 2 \le k \le K-1\ ;\ 2k-1 \le j \le 2(K-1),$$

$$R_{2k,2K} = R_{2K,2K}\left(\frac{\lambda_1}{\beta_2}\right)^{K-k} = \left(\frac{\lambda_2\beta_3}{\beta_2\beta_3 - \lambda_1\mu_1}\right)\left(\frac{\lambda_1}{\lambda_2 z_4}\right)^{K-k}, \text{ for } 1 \le k \le K-1,$$

$$R_{2k,2K+1} = R_{2K+1,2K+1}\left(\frac{\lambda_1}{\beta_2}\right)^{K-k+1} = \left(\frac{\lambda_2\beta_2}{\beta_2\beta_3 - \lambda_1\mu_1}\right)\left(\frac{\lambda_1}{\lambda_2 z_4}\right)^{K-k+1}, \text{ for } 1 \le k \le K-1,$$

$$R_{2K,2K+1} = \frac{\lambda_1}{\beta_3}R_{2K,2K} = \frac{\lambda_1\lambda_2}{\beta_2\beta_3 - \lambda_1\mu_1} = \frac{\lambda_1}{\beta_2}R_{2K+1,2K+1},$$

$$R_{2K+1,2K} = \frac{\mu_1}{\beta_2}R_{2K+1,2K+1} = \frac{\mu_1\lambda_2}{\beta_2\beta_3 - \lambda_1\mu_1} = \frac{\mu_1}{\beta_3}R_{2K,2K},$$

$$R_{2K,2K-1} = \frac{\alpha R_{2K,2K}\left(\beta_5 - \mu_2(R_{2K+1,2K+1} + R_{2K-1,2K-1})\right) + \alpha\mu_2 R_{2K,2K+1}R_{2K+1,2K}}{\left(\beta_5 - \mu_2\left(R_{2K,2K} + \frac{1}{z_2}\right)\right)\left(\beta_5 - \mu_2\left(R_{2K+1,2K+1} + \frac{1}{z_2}\right)\right) - \mu_2^2\alpha R_{2K,2K+1}R_{2K+1,2K}},$$

$$R_{2K+1,2K-1} = \frac{\alpha R_{2K+1,2K} + \mu_2 R_{2K+1,2K}R_{2K,2K-1}}{\beta_5 - \mu_2\left(R_{2K+1,2K+1} + \frac{1}{z_2}\right)},$$

$$R_{2k-1,2K-1} = R_{1,2(K-k)-1}, \text{ for } 2 \le k \le K-1,$$

$$R_{2k,2K-1} = \frac{\lambda_1 R_{2,1} + \alpha R_{2k,2K} + \mu_2\left(R_{21}R_{1,2(K-k)+1} + \sum_{i=1}^{K-k-1} R_{2,i}R_{2k+i,2K-1} + R_{2k,2K}R_{2K,2K-1} + R_{2k,2K+1}R_{2K+1,2K-1}\right)}{\beta_2 - \mu_2\left(\frac{1}{z_2} + \frac{1}{z_4}\right)},$$

for $2 \le k \le K-2$,

$$R_{2k-1,2K+2} = \frac{\lambda_1 R_{2k-1,2K-1} + \mu_2 \sum_{i=1}^{K-k} R_{2k-1,2(k+i)-1}R_{2(k+i)-1,2K+2}}{\beta_6 - \mu_2\left(\frac{1}{z_2} + \frac{1}{z_3}\right)}, \text{ for } 1 \le k \le K,$$

$$R_{2K,2K+2} = \frac{\left(\lambda_1 + \mu_2 R_{2K-1,2K+2}\right)\left(R_{2K,2K-1}\left(\beta_6 - \mu_2\left(R_{2K+1,2K+1} + \frac{1}{z_3}\right)\right) + \mu_2 R_{2K,2K+1}R_{2K+1,2K-1}\right)}{\left(\beta_6 - \mu_2\left(R_{2K,2K} + \frac{1}{z_3}\right)\right)\left(\beta_6 - \mu_2\left(R_{2K+1,2K+1} + \frac{1}{z_3}\right)\right) - \mu_2^2 R_{2K,2K+1}R_{2K+1,2K}},$$

$$R_{2K+1,2K+2} = \frac{\lambda_1 R_{2K+1,2K-1} + \mu_2\left(R_{2K+1,2K-1}R_{2K-1,2K+2} + R_{2K+1,2K}R_{2K,2K+2}\right)}{\beta_6 - \mu_2\left(R_{2K+1,2K+1} + \frac{1}{z_3}\right)},$$

$$R_{2k,2K+2} = \frac{\lambda_1 R_{2k,2K-1} + \mu_2\left(R_{2k,2k-1}R_{2k-1,2K+2} + \sum_{i=1}^{K-k+2} R_{2k,2k+i}R_{2k+i,2K+2}\right)}{\beta_6 - \mu_2\left(\frac{1}{z_4} + \frac{1}{z_3}\right)}, \text{ for } 1 \le k \le K-1,$$

The other elements of R are equal to zero.

Table 5.1
Performance measures as functions of λ_1, when $\lambda_2 = 3$, $\mu_1 = 3$, $\mu_2 = 4$ and $\alpha = 5$.

Values of λ_1	P_{loss}	λ_1^{eff}	$\mathbb{E}[L_1]$	$\mathbb{E}[L_2]$	$\mathbb{E}[W_1]$	$\mathbb{E}[W_2]$
0.01	1.6×10^{-8}	0.01	0.048	3.0362	4.7980	1.0121
0.1	0.0004	0.0999	0.6055	3.3292	6.0575	1.1097
0.5	0.1986	0.4007	5.4876	4.4380	13.694	1.4793
1	0.5344	0.4655	8.5837	5.2767	18.438	1.7590
2	0.7620	0.4759	9.5894	5.9541	20.148	1.9847
4	0.8741	0.5036	9.8400	6.6118	19.537	2.2039
10	0.9431	0.5686	9.9389	7.9745	17.481	2.6582
100	0.9924	0.7586	9.9924	25.907	13.173	8.6356
1000	0.9983	1.7325	9.9983	203.88	5.7709	67.961
10000	0.9997	2.7703	9.9997	2001.7	3.6096	667.23
100000	0.9999	2.9750	9.9999	20001	3.3613	6667.1
1000000	0.99999	2.9975	10	200001	3.3361	66667

Table 5.2
Performance measures as functions of λ_2, when $\lambda_1 = 2$, $\mu_1 = 3$, $\mu_2 = 4$ and $\alpha = 5$.

Values of λ_2	P_{loss}	λ_1^{eff}	$\mathbb{E}[L_1]$	$\mathbb{E}[L_2]$	$\mathbb{E}[W_1]$	$\mathbb{E}[W_2]$
0.01	0.0059	1.9881	1.8891	0.0224	0.9502	2.2407
0.1	0.0074	1.9851	2.0625	0.2113	1.0390	2.1136
0.5	0.0352	1.9297	3.2757	0.8230	1.6975	1.6459
1	0.1357	1.7286	5.4280	1.5370	3.1400	1.5370
2	0.4727	1.0546	8.5720	3.2026	8.1286	1.6013
2.5	0.6267	0.7467	9.2212	4.2734	12.350	1.7094
3	0.7620	0.4760	9.5894	5.9541	20.148	1.9847
3.5	0.8848	0.2305	9.8260	10.256	42.635	2.9302
3.75	0.9411	0.1178	9.9155	17.890	84.186	4.7707

Note that $2K$ (out of $2K + 2$) elements on the main diagonal of R are equal to the inverse of the roots of $|B(z)|$ and $|C(z)|$ in the open interval $(1, \infty)$, as described in Theorems 3.2 and 3.3, while the other elements depend both on those roots and on other parameters of the system. By obtaining explicit expressions for all elements of the rate matrix R, we can by-pass the sequential substitution method commonly used to calculate numerically R, and efficiently study problems with large values of K and N.

5. Numerical results

In this section we present numerical calculations of $P_{loss} = P_{K\bullet}$, $\lambda_1^{eff} = \lambda_1(1 - P_{loss})$, $\mathbb{E}[L_i]$ (Eqs. (3.4) and (3.5)) and $\mathbb{E}[W_i]$ (Eqs. (3.6) and (3.7)), $i = 1, 2$, as follows:

Tables 5.1–5.5 exhibit results for the performance measures when $K = 10$ and $N = 3$, for different values of λ_1, λ_2, μ_1, μ_2 and α. In each table one of the parameters changes while all other parameters remain fixed. The basic values are $\lambda_1 = 2$, $\lambda_2 = 3$, $\mu_1 = 3$, $\mu_2 = 4$ and $\alpha = 5$, respectively.

Investigating Table 5.1 it is seen that, when λ_1 increases, both queue sizes increase, as well as $\mathbb{E}[W_2]$. However, $\mathbb{E}[W_1]$ first increases monotonically as λ_1 increases ($\lambda_1 \leq 2$) and then monotonically decreases. The explanation for this seemingly counter-intuitive phenomenon is the following: Since Q_1's buffer is bounded ($K = 10$), large values of λ_1 almost do not affect $\mathbb{E}[L_1]$, nor λ_1^{eff}. When L_1 reduces from $L_1 = 10$ to $L_1 = 9$, while $L_2 \geq N$, the server starts switching to Q_2, but this move is immediately aborted with a new arrival to Q_1, thus eliminating potential waiting times in Q_1 had the server completed switching to Q_2. This is in *contrast* to the results in Perel and Yechiali [18], where $\mathbb{E}[W_1]$ increases as λ_1 increases, since switching there is instantaneous and thus allowing the server to remain for a while in Q_2. Note also that $E[W_1]$ approaches the value $K \cdot \frac{1}{\mu_1} = 3\frac{1}{3}$, as $\lambda_1 \longrightarrow \infty$, since Q_1 is loaded and almost all customers that join the queue are admitted in the Kth position.

Table 5.2 exhibits an interesting direction of change of $\mathbb{E}[W_2]$ when λ_2 increases. It first decreases with increasing values of λ_2, and then increases. As $\mathbb{E}[L_2]$ increases and approaches the threshold N, the server spends more time in Q_2, pushing $\mathbb{E}[W_2]$ down, despite the increasing λ_2. However, as $\mathbb{E}[L_2]$ grows well beyond N, the server stays most of the time in Q_2, but high values of λ_2 cause new customers to wait longer. This phenomenon is depicted in Fig. 5.1.

In Table 5.3, when μ_1 increases, both $\mathbb{E}[W_1]$ and $\mathbb{E}[W_2]$ decrease. This follows since Q_1 often stays below its threshold, allowing the server to switch to Q_2 without being aborted, thus decreasing its size.

Table 5.4 demonstrates that increasing values of μ_2 decrease the values of all performance measures.

The results of Table 5.5 emphasize the impact of relatively small threshold of Q_2 ($N = 3$, compared to $K = 10$). The server stays most of the time in Q_2 and $\mathbb{E}[L_1]$ is close to K for any value of α. Increasing values of α (rapid switches) decrease the values of all performance measures, except $\lambda_1^{eff} = \lambda(1 - P_{loss})$.

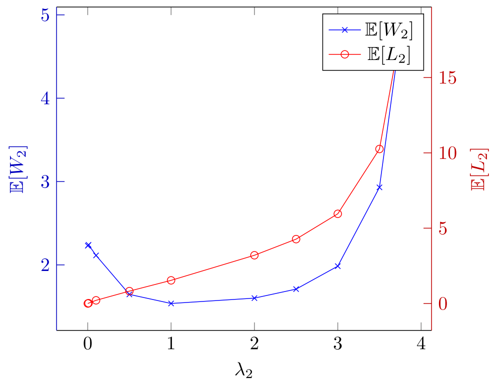


Fig. 5.1. $\mathbb{E}[L_2]$ and $\mathbb{E}[W_2]$ as functions of λ_2.

Table 5.3
Performance measures as functions of μ_1, when $\lambda_1 = 2$, $\lambda_2 = 3$, $\mu_2 = 4$ and $\alpha = 5$.

Values of μ_1	P_{loss}	λ_1^{eff}	$\mathbb{E}[L_1]$	$\mathbb{E}[L_2]$	$\mathbb{E}[W_1]$	$\mathbb{E}[W_2]$
0.001	0.9995	0.0010	9.9995	4200.1	10397	1400.0
0.01	0.9964	0.0073	9.9964	421.50	1375.0	140.50
0.1	0.9847	0.0306	9.9847	46.149	326.36	15.383
0.5	0.9449	0.1103	9.9439	12.962	90.189	4.3206
1	0.9005	0.1990	9.8935	8.7647	49.720	2.9216
2	0.8272	0.3455	9.7749	6.6644	28.289	2.2215
4	0.6988	0.6025	9.3026	5.5780	15.440	1.8593
10	0.4267	1.1467	6.9733	4.6262	6.0813	1.5421
100	0.2230	1.5543	4.3079	3.8023	2.7715	1.2674
1000	0.2084	1.5831	4.0838	3.7323	2.5796	1.2441
10000	0.2070	1.5859	4.0620	3.7255	2.5613	1.2418
100000	0.2069	1.5861	4.0598	3.7248	2.5595	1.2416

Table 5.4
Performance measures as functions of μ_2, when $\lambda_1 = 2$, $\lambda_2 = 3$, $\mu_1 = 3$ and $\alpha = 5$.

Values of μ_2	P_{loss}	λ_1^{eff}	$\mathbb{E}[L_1]$	$\mathbb{E}[L_2]$	$\mathbb{E}[W_1]$	$\mathbb{E}[W_2]$
3.25	0.9264	0.1472	9.8830	14.882	67.120	4.9606
3.5	0.8649	0.2703	9.7796	9.0270	36.186	3.0090
3.75	0.8103	0.3795	9.6819	6.9922	25.512	2.3307
4	0.7620	0.4760	9.5894	5.9541	20.148	1.9847
10	0.3048	1.3903	7.8261	2.5769	5.6290	0.8590
100	0.0922	1.8157	5.2089	1.4877	2.8689	0.4959
1000	0.0797	1.8406	4.9433	1.4114	2.6856	0.4705
10000	0.0785	1.8430	4.9173	1.4041	2.6681	0.4680

Table 5.5
Performance measures as functions of α, when $\lambda_1 = 2$, $\lambda_2 = 3$, $\mu_1 = 3$ and $\mu_2 = 4$.

Values of α	P_{loss}	λ_1^{eff}	$\mathbb{E}[L_1]$	$\mathbb{E}[L_2]$	$\mathbb{E}[W_1]$	$\mathbb{E}[W_2]$
3.25	0.7950	0.4100	9.6784	6.3537	23.608	2.1179
3.5	0.7895	0.4211	9.6642	6.2747	22.951	2.0916
3.75	0.7842	0.4316	9.6505	6.2053	22.362	2.0685
4	0.7793	0.4415	9.6373	6.1439	21.831	2.0480
10	0.7120	0.5761	9.4332	5.5400	16.375	1.8467
100	0.6375	0.7249	9.1523	5.1009	12.625	1.7003
1000	0.6279	0.7442	9.1116	5.0520	12.244	1.6840
10000	0.6269	0.7461	9.1074	5.0471	12.206	1.6824
100000	0.6268	0.7463	9.1070	5.0466	12.202	1.6822

Remark 5.1. : when $\alpha \to \infty$, our model converges to the model studied in [18] for the work conserving scenario. Comparing in Table 5.5 the numbers appearing when $\alpha \geq 100$ to the corresponding numbers in Table 1 of [18] for the case where $\lambda_1 = 2$, the results for $\mathbb{E}[L_1]$, $\mathbb{E}[L_2]$, $\mathbb{E}[W_1]$ and $\mathbb{E}[W_2]$ are close, as expected.

Remark 5.2. The above numerical results are for given values of the thresholds and for a given switching policy. One may consider optimization issues, such as modeling the system as a Markovian Decision process including costs aspects and directed at determining optimal switching instances and optimal threshold levels. We leave those research directions for future work. We also note that control and optimization issues of polling systems were dealt in [19,20,5], Gandhi and Cassandras [24,21,16].

6. Extreme cases

In this section we examine the impact of extreme values of λ_1, λ_2, μ_1, μ_2 and α (as they reach 0 or ∞) on the system's performance measures.

$\lambda_2 \to \infty$ *or* $\mu_2 \to 0$

As the stability condition is $\lambda_2 < \mu_2$, these two cases are not stable.

$\lambda_1 \to \infty$

This case leads to an unstable system as well. Q_1 is (almost) always at its maximum capacity. That is, $L_1 \equiv K$ and $P_{\text{loss}} = 1$. In such a case, at the next instant when the server switches from state $I = 1$ to state $I = S1$, the next arrival to Q_1 occurs almost always before the switching time is over, causing an immediate switch back to Q_1. Therefore, the server will hardly ever attend Q_2.

$\mu_1 \to 0$

This case is also unstable. As soon as the server attends Q_1 and $L_1 = K$, L_1 will not reduce below the threshold level and the server will never switch back to Q_2 even when the number of customers in Q_2 reaches its threshold, N. As a result, L_2 will increase to ∞.

$\lambda_1 \to 0$

In this case $\mathbb{P}(I = 1) = 0$ and $\mathbb{P}(I = 2) = 1$. That is, Q_2 operates as an $M(\lambda_2)/M(\mu_2)/1$ system. Hence, $\mathbb{P}(L_1 = 0) = 1$. When $\lambda_1 \to 0$, $\mathbb{E}[L_2] \to \frac{\rho_2}{1-\rho_2} = 3$. See Table 5.1 where $\mathbb{E}[L_2] = 3.0362$ when $\lambda_1 = 0.01$.

$\lambda_2 \to 0$

Here $\mathbb{P}(I = 1) = 1$, and $\mathbb{P}(I = 2) = 0$. Thus, Q_1 operates as an $M(\lambda_1)/M(\mu_1)/1/K$ system for which $P_{\text{loss}} \to \frac{\rho_1^K(1-\rho_1)}{1-\rho_1^{K+1}} = 0.0058$, and $\mathbb{E}[L_1] \to \frac{\rho_1}{1-\rho_1} - \frac{(K+1)\rho_1^{K+1}}{1-\rho_1^{K+1}} = 1.8713$, where $\rho_1 = \frac{\lambda_1}{\mu_1}$. See Table 5.2 where $P_{\text{loss}} = 0.0059$ and $\mathbb{E}[L_1] = 1.8891$ when $\lambda_2 = 0.01$.

$\mu_1 \to \infty$

Whenever the server is at Q_1, the number of customers there immediately reduces to 0. Then, if Q_2 is empty, the server remains at Q_1 until the first moment thereafter when a customer arrives at Q_2. Otherwise, if Q_2 is not empty, the server immediately starts a switch-over to Q_2. When the server is at Q_2, it stays there until Q_1 reaches its threshold while Q_2 is still below its own threshold, N. When Q_1 reaches its threshold and Q_2 is *not* below its threshold, the server stays at Q_2 until the number of customers there falls short of Q_2's threshold, upon which the server starts a switch-over to Q_1. Note that in this case $P_{\text{loss}} = P_{K\bullet}(2) + P_{K\bullet}(S2)$.

$\mu_2 \to \infty$

Once the server arrives at Q_2, it immediately empties it. If Q_1 is empty, the server stays in Q_2 until a customer arrives at Q_1. Else, the server immediately starts a switch-over to Q_1. Whenever the server is at Q_1, it stays there until Q_2 reaches its threshold and Q_1 is below its own threshold K. If Q_2 reaches its threshold and Q_1 is *not* below its threshold, the server stays at Q_1 until the number of customers there decreases below K, upon which the server starts a switch-over to Q_2. Note that in this case $P_{\text{loss}} = P_{K\bullet}(1) + P_{K\bullet}(S2)$.

7. Summary and conclusions

This paper presents a 3-fold contribution to the literature on polling systems, in particular on 2-queue alternating server models: (i) Switch-over decisions are threshold-based and depend on the queue which is *not* being served, where in the majority of polling systems, such as exhaustive, gated or globally-gated regimes, these decisions depend on the queue *being* served. (ii) It investigates more deeply the role of the switch-over durations, and (iii) by explicitly determining the entries of the rate matrix R, it renders a reduced computational effort for the calculations of the system's performance measures. We reveal that the entries of the rate matrix R are expressed in terms of the roots of the determinants of two matrices, $B(z)$ and $C(z)$. Those matrices satisfy $B(z)\vec{F}(z) = \vec{Q}(z)$ and $C(z)\vec{H}(z) = \vec{R}(z)$ respectively, where $\vec{F}(z)$ and $\vec{H}(z)$ are each a vector whose entries are the sought-for PGFs of the system's steady-state probabilities. $B(z)$ and $C(z)$ are finite square matrices with entries constructed from the parameters of the system. $\vec{Q}(z)$ and $\vec{R}(z)$ are finite-dimensional vectors. In deviation from many cases in which the rate matrix is calculated numerically, we are able to derive closed-form expressions for all the elements of R.

As for further research, one question is whether or not the model can be extended to multiple queues. Another possible extension is to analyze a non-work-conserving system. Such a policy may be used when switching costs are high. A further possible extension is to study the non-preemptive case. Finally, one can study non-exponential switch-over times as follows: let X_i denote the switch-over time from Q_i to Q_j $(i \neq j)$ with probability density function $f_i(t)$. Then, the probability of successful switching from Q_i to Q_j is $\alpha_i = \int_{t=0}^{\infty} e^{-\lambda_i t} f_i(t)dt = \widetilde{X}_i(\lambda_i)$, $i = 1, 2$.

Acknowledgment

We thank Nir Perel for very valuable discussions.

Appendix

$$B_0^0 = \begin{pmatrix} 0 & \lambda_2 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ \lambda_2 & 0 & 0 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & \lambda_2 & 0 & 0 & \ddots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \lambda_2 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \lambda_2 & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & \lambda_2 \end{pmatrix},$$

$B_0 = diag(\lambda_2)$,

$$B_0^{N-1} = \begin{pmatrix} \lambda_2 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & 0 & \lambda_2 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & \lambda_2 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & 0 & \lambda_2 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & \lambda_2 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda_2 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \lambda_2 & 0 & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \lambda_2 & 0 & 0 & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \lambda_2 & 0 & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \lambda_2 & 0 & 0 & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \lambda_2 & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & \lambda_2 \end{pmatrix},$$

$$B_1^0 = \begin{pmatrix} -\beta_0 & 0 & \lambda_1 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & -\beta_0 & 0 & \lambda_1 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ \mu_1 & 0 & -\beta_1 & 0 & \lambda_1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & \alpha & -\beta_2 & 0 & \lambda_1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & \mu_1 & 0 & -\beta_1 & 0 & \lambda_1 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 0 & 0 & \alpha & -\beta_2 & 0 & \lambda_1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \mu_1 & 0 & -\beta_3 & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \alpha & -\beta_4 \end{pmatrix},$$

$$B_1 = \begin{pmatrix}
-\beta_5 & 0 & 0 & \lambda_1 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
\alpha & -\beta_2 & 0 & 0 & \lambda_1 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
0 & \mu_1 & -\beta_1 & 0 & 0 & 0 & \lambda_1 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
0 & 0 & 0 & -\beta_5 & 0 & 0 & 0 & \lambda_1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\
0 & 0 & 0 & \alpha & -\beta_2 & 0 & 0 & 0 & \lambda_1 & 0 & \cdots & \cdots & \cdots & 0 \\
0 & 0 & \alpha & 0 & 0 & -\beta_2 & 0 & 0 & 0 & \lambda_1 & 0 & \cdots & \cdots & 0 \\
0 & 0 & \mu_1 & 0 & 0 & 0 & -\beta_1 & 0 & 0 & 0 & \lambda_1 & 0 & \cdots & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 & 0 & \lambda_1 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 & \lambda_1 & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 & \lambda_1 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \mu_1 & 0 & 0 & 0 & -\beta_3 & 0 \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \alpha & -\beta_4
\end{pmatrix},$$

where $\beta_0 = \lambda_1 + \lambda_2$; $\beta_1 = \lambda_1 + \lambda_2 + \mu_1$; $\beta_2 = \lambda_1 + \lambda_2 + \alpha$; $\beta_3 = \lambda_2 + \mu_1$; $\beta_4 = \lambda_2 + \alpha$ and $\beta_5 = \lambda_1 + \lambda_2 + \mu_2$.

$$B_2^1 = \begin{pmatrix}
0 & \mu_2 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
0 & 0 & 0 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \vdots \\
0 & 0 & 0 & 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & \mu_2 & 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & 0 & 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & 0 & 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & 0 & 0 & \mu_2 & 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \mu_2 & 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & \mu_2 & 0 & 0 & 0 \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0
\end{pmatrix},$$

$$B_2 = \begin{pmatrix}
\mu_2 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
0 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \vdots \\
0 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \vdots \\
0 & 0 & 0 & \mu_2 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \vdots \\
0 & 0 & 0 & 0 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \vdots \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \vdots \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \mu_2 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \vdots \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \mu_2 & 0 & 0 & 0 & 0 \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0
\end{pmatrix},$$

Note that all the entries of the last 4 rows in matrices B_2^1 and B_2 are zeros.

$$A_2^N = \begin{pmatrix}
\mu_2 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
0 & 0 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \vdots \\
0 & 0 & 0 & \mu_2 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \vdots \\
0 & 0 & 0 & 0 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \mu_2 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \vdots \\
0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \mu_2 & 0 & \ddots & \ddots & \vdots \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \vdots \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \vdots \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \mu_2
\end{pmatrix},$$

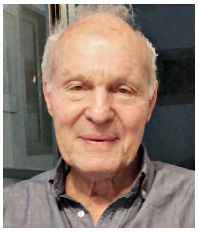
References

[1] H. Takagi, Analysis of Polling Systems, The MIT Press, 1986.
[2] O. Boxma, W. Groenendijk, Pseudo-conservation laws in cyclic service systems, Adv. Appl. Probab. 24 (1987) 949–964.
[3] O. Boxma, H. Levy, U. Yechiali, Cyclic reservation schemes for efficient operation of multiple-queue single-server systems, Ann. Oper. Res. 35 (1–4) (1992) 187–208.
[4] S. Browne, U. Yechiali, Dynamic priority rules for cyclic-type queues, Adv. Appl. Probab. 21 (2) (1989) 432–450.
[5] U. Yechiali, Analysis and control of polling systems, in: L. Donatiello, R. Nelson (Eds.), Performance Evaluation of Computer and Communication Systems, Springer-Verlag, 1993, pp. 630–650.
[6] J. Resing, Polling systems and multitype branching processes, Queueing Syst. 13 (1993) 409–426.
[7] M. Boon, R. Van der Mei, E. Winands, Applications of polling systems, Surv. Oper. Res. Manag. Sci. 16 (2) (1993) 67–82.
[8] L. Takács, Two queues attended by a single server, Oper. Res. 16 (1968) 639–650.
[9] O. Boxma, D. Down, Dynamic server assignment in a two-queue model, European J. Oper. Res. 103 (1997) 595–609.
[10] O. Boxma, S. Schlegel, U. Yechiali, Two-queue polling models with a patient server, Ann. Oper. Res. 112 (2002) 101–121.
[11] D.S. Lee, A two-queue model with exhaustive and limited service disciplines, Commun. Stat. Stoch. Models 12 (1996) 285–305.
[12] D.S. Lee, B. Sengupta, Queueing analysis of a threshold based priority scheme for ATM networks, IEEE/ACM Trans. Netw. 1 (1993) 709–717.
[13] B. Haverkort, H. Idzenga, B. Kim, Performance evaluation of threshold-based ATM cell scheduling policies under Markov modulated Poisson traffic using stochastic Petri nets, in: Chapman, Hall (Eds.), Performance Modelling and Evaluation of ATM Networks, in: Proceedings IFIP '95, 1995, pp. 553–572.
[14] O. Boxma, G. Koole, I. Mitrani, A two-queue polling model with a threshold service policy, in: P. Dowd, E. Gelenbe (Eds.), Proceedings MASCOTS '95, IEEE Computer Society Press, Los Alamitos, CA, 1995, pp. 84–89.
[15] O. Boxma, G. Koole, I. Mitrani, Polling models with threshold switching, in: F. Baccelli, A. Jean-Marie, I. Mitrani (Eds.), Quantitative Methods in Parallel Systems '95, Springer Verlag, Berlin, 1995, pp. 129–140.

[16] F. Avram, A. Gómez-Corral, On the optimal control of a two-queue polling model, Oper. Res. Lett. 34 (2006) 339–348.
[17] K. Avrachenkov, E. Perel, U. Yechiali, Finite-buffer polling systems with threshold-based switching policy, TOP 24 (3) (2016) 541–571.
[18] E. Perel, U. Yechiali, Two-queue polling systems with switching policy based on the queue that is not being served, Stoch. Models 33(3) (2017) 430–450.
[19] I. Meilijson, U. Yechiali, On optimal right-of-way policies at a single-server station when insertion of idle times is permitted, Stochastic Process. Appl. 6(1) (1977) 25–32.
[20] O. Boxma, H. Levy, J. Weststrate, Efficient visit frequencies for polling tables: minimization of waiting cost, Queueing Syst. 9 (1991) 133–162.
[21] J.v.d. Wal, U. Yechiali, Dynamic visit-order rules for batch-service polling, Probab. Engrg. Inform. Sci. 17 (3) (2003) 351–367.
[22] M. Neuts, Matrix Geometric Solutions in Stochastic Models — an Algorithmic Approach, The Johns Hopkins University Press, Baltimore and London, 1981.
[23] G. Latouche, V. Ramaswami, Introduction to matrix analytic methods in stochastic modeling, in: ASA-SIAM Series on Statistics and Applied Mathematics, ASA-SIAM, Philadelphia, 1999.
[24] A. Gandhi, C. Cassandras, Optimal control of polling models for transportation applications, Math. Comput. Modelling 23 (11) (1996) 1–23.

Amit Jolles is a researcher in the field of Machine Learning. He received a B.Sc. (Cum Laude) in Industrial Engineering from the Technion, Israel, and an M.Sc. (with distinction) in Operations Research from Tel-Aviv University, Israel.

Efrat Perel is a researcher in the fields of Queueing Theory and Stochastic Processes. She received a B.Sc. in Mathematical Sciences, M.Sc. in Operations Research (Cum Laude) and a Ph.D. in Operations Research, all from Tel-Aviv University, Israel. Efrat is a full faculty member and a senior lecturer at Afeka College of Engineering, and an adjunct lecturer at Tel-Aviv University. Dr. Perel has published studies in top journals in the fields of queueing systems, stochastic processes and service systems.



Emeritus Uri Yechiali is a world expert in the fields of Queueing Theory and Stochastic Modeling. He received a B.Sc. (Cum Laude) in Industrial Engineering and an M.Sc. in Operations Research - both from the Technion, Israel, and a Ph.D. in Operations Research from Columbia University, New York. He joined Tel-Aviv University in 1971 and promoted to a full professor in 1981. Professor Yechiali was a Visiting Professor at Columbia University, NYU, INRIA (France), and held the Beta Chair in Eurandom (Holland). His 140 scientific publications include seminal works on vacation models in queues, queues in random environment, analysis and control of polling systems, tandem Jackson networks, optimal policies for live organ transplants, modeling genetic regulatory networks, analysis of asymmetric inclusion processes, and statistical mathematical programming. Prof. Yechiali supervised 41 Master and 13 Ph.D. students (most of them hold faculty positions in prestigious universities). For his academic excellence he received in 2004 a Life Achievement Award from the Operations Research Society of Israel. His current research interests are tandem stochastic networks; ASIP models; optimal allocation policies for live-organ transplants; queues with preliminary services; and QBD processes.