Transient Analysis of the State Dependent M/M/1/K Queue

Ahmad Al Hanbali, Onno Boxma
EURANDOM, Eindhoven, The Netherlands
{a.al.hanbali,o.j.boxma}@tue.nl

July 8, 2009

Abstract

In this paper, we study the transient behavior of a state dependent M/M/1/K queue during the busy period. We derive in closed-form the joint transform of the length of the busy period, the number of customers served during the busy period, and the number of losses during the busy period. For two special cases called the threshold policy and the static policy we determine simple expressions for their joint transform. The performance metrics of the three random variables such as their expectations, variances, and covariances follow directly from the joint transform. Finally, we give additional results that contain the distribution of the maximum queue level reached during the busy period, the transform of the c-congestion period and the total number of such periods during the busy period.

1 Introduction

In practice, it is often the case that arrivals and their service times depend on the system state. For example, in telecommunication systems this happens at the packet switch (router): when its buffer size increases, a controller drops the arriving packets with an increasing probability. In human based service systems, it is known that there is a strong correlation between the volume of work demanded from a human and her/his productivity. Moreover, the transient performance measures of a system are important for understanding the system evolution. All these facts motivate us to study the transient measures of a state dependent queueing system.

The transient regime of queueing systems is much more difficult to analyze than the steady state regime. This explains the scarcity of transient research results in this field compared to the steady state regime. A good exception is the M/M/1 queue which has been well studied in both transient and steady state regimes. This paper is devoted to the study of the transient behavior of the state dependent M/M/1/K, i.e., the M/M/1 queue with finite waiting room of size $K-1$. In particular, we shall analyze the transient measures related to the busy period.

Takács in [13, Chap 1] was among the first to derive the state dependent probabilities of the M/M/1/K, referred to as $P_{ij}(t)$. Basically, these are the probabilities that at time t the queue length is j given it was i at time zero. Building on these probabilities Takács also determined the state dependent probabilities of the M/M/1 queue by taking the limit of $P_{ij}(t)$ for $K \to \infty$. For the M/G/1/K, Cohen [4, Chap III.6] computed the Laplace transform of $P_{ij}(t)$ and the bivariate transform of the number of customers served and number of losses due to overflow during the busy period. This is done using complex analysis theory. Specifically, the joint transform is presented as a fraction of two contour integrals that involve K and the Laplace-Stieltjes transform of the customers' service time. Rosenlund in [12] extended Cohen's result by deriving the joint transform of the busy period length, the number of customers served and the number of losses during the busy period. The approach of Rosenlund is more probabilistic than Cohen's analysis. However, Rosenlund's final result for the trivariate transform is represented as a fraction of two contour integrals. In an earlier work Rosenlund in [11] gave the trivariate transform of the M/M/1/K in terms of the roots of a specific quadratic equation. For more recent works on the busy period analysis of M/G/1/K we refer to [7, 14]. Recently, there was an increased interest in the expected number of losses during the busy period in the M/G/1/K queue with equal arrival and service rate; see, e.g., [1, 10, 15]. In this case, the interesting phenomenon is that the expected number of losses during the busy period in the M/G/1/K equals one for all values of $K \geq 1$.

In this paper we extend the results of Rosenlund in [11] for the M/M/1/K in several ways. First, we study a state dependent M/M/1/K with admission control. Second, we consider the residual busy period that is initiated with $n \geq 1$ customers. Moreover, we shall derive the distribution of the maximum number of customers during the busy period and other related performance measures. This is done using the theory of absorbing Markov chains. The key point is to model the event that the system becomes empty as absorbing. Keeping track of the evolution of the Markov chain before the absorption leads to the desired results such as the busy period length, the total number of customers served during the busy period, and the total number of losses during the busy period.

The paper is organized as follows. In Section 1.1, we give a detailed description of the model and the assumptions made. Section 2 reports our results that shall be presented in a number of different Theorems, Propositions, and Corollaries. More precisely, Theorem 1 gives our main result for the trivariate transform as function of the inverse of a specific matrix. Proposition 1 presents a numerical recursion to invert this matrix. In Corollaries 1 and 2, we derive the closed form expressions for the trivariate transform in two special cases that we shall refer to as the threshold policy and the static policy. The performance metrics of the three random variables such as their expectations, variances, and covariances are presented in Section 3. In Section 4, we give additional results that shall contain the distribution of the maximum queue level reached during the busy period, the transform of the c-congestion period and the total number of such periods during a busy period. Finally, in Section 5, we conclude the paper and give some research directions.

1.1 Model

We consider an M/M/1/K queueing system, i.e., an M/M/1 queue with finite waiting room of size $K-1$ customers. The arrival process is Poisson with rate λ_i and the service rate is μ_i in the case where the queue length is $i \in \{0, 1, \ldots, K\}$. We assume that an admission controller is installed at the entry of the queue that has the duty of dropping the arriving customers with probability p_i when the queue length is $i \in \{0, 1, \ldots, K\}$. In other words, the customers are admitted in the queue with probability $q_i = 1-p_i$ when its queue length is i. The arrivals to the queue of size K are all lost. It should be clear that in this case $p_K = 1$ and $q_K = 0$.

Let $N(t) \in \{0, 1, \ldots, K\}$ denote the Markov process that represents the queue length at time t. We are interested in the queue behavior during the *busy period* which is defined as: the time interval that starts with an arrival to an empty queue and ends for the first time the queue becomes empty again. Similarly, we define the residual busy period as the busy period initiated with $n \geq 1$ customers. Note that for $n = 1$ the residual busy period and the busy period are equal.

Consider an arbitrary residual busy period. Let B_n denote its length. Let S_n denote the total number of served customers during B_n. Let L_n denote the total number of losses, i.e. arrivals that are not admitted in the queue, during B_n. In this paper, we determine the joint transform $\mathbb{E}\big[e^{-wB_n} \cdot z_1^{S_n} \cdot z_2^{L_n}\big]$, $Re(w) \geq 0$, $|z_1| \leq 1$ and $|z_2| \leq 1$. We will use the theory of absorbing Markov chains. This is done by modeling the event that "the queue jumps to the empty state" as an absorbing event. Tracking the number of customers served and losses before the absorption occurs gives the desired result.

A word on the notation: throughout $x := y$ will designate that by definition x is equal to y, $1_{\{E\}}$ the indicator function of any event E ($1_{\{E\}}$ is equal to one if E is true and zero otherwise), x^T the transpose vector of x, e_i the unit row vector of appropriate dimension with all entries equal to zero except the i-th entry that is one, and $\mathbf{I}$ the identity matrix of appropriate dimension.

2 Results

Before reporting our results let us define the following matrices: the matrix $\mathbf{A}$ that is an upper bidiagonal matrix with upper diagonal equal to $(q_1\lambda_1, \ldots, q_{K-1}\lambda_{K-1})$ and diagonal equal to $-(\lambda_1+\mu_1, \ldots, \lambda_K+\mu_K)$, the matrix $\mathbf{B}$ that is a lower diagonal matrix with lower diagonal equal to $(\mu_2, \ldots, \mu_K)$, and the matrix $\mathbf{C}$ that is a diagonal matrix with diagonal equal to $(p_1\lambda_1, \ldots, p_{K-1}\lambda_{K-1}, \lambda_K)$. Moreover, let us denote $\mathbf{Q}_K(w, z_1, z_2) = w\mathbf{I} - \mathbf{A} - z_1\mathbf{B} - z_2\mathbf{C}$. For the sake of the ease of presentation, we shall refer to $\mathbf{Q}_K(w, z_1, z_2)$ as $\mathbf{Q}_K$. We are now ready to formulate our first result.

Theorem 1 (Level dependent) *The joint transform of B_n, S_n, and L_n is given by*[1]

$$\mathbb{E}_d\big[e^{-wB_n} z_1^{S_n} z_2^{L_n}\big] = \mu_1 z_1 e_n \big(\mathbf{Q}_K\big)^{-1} e_1^T.$$

Proof: In the following we model the event that the queue becomes empty, i.e. the end of the busy period, as an absorbing event. The trivariate transform is deduced by determining the last state visited before absorption.

Let $(N(t), S(t), L(t))$ denote the three-dimensional, continuous-time Markov process with discrete state-space $\xi := \{0, 1, \cdots, K\} \times \mathbb{N} \times \mathbb{N}$, where $N(t)$ represents the number of customers in the queue at time t, $S(t)$ the number of served customers from the queue until t, $L(t)$ the number of losses in the queue until t, and $\mathbb{N}$ the set of non-negative integers. States $(0, \cdot, \cdot)$ are absorbing. We refer to this absorbing Markov process by AMC. The absorption of AMC occurs when the queue becomes empty, i.e., $N(t) = 0$. By setting the initial state of AMC at $t = 0$ to $(n, 0, 0)$, $n \geq 1$, the time until absorption is equal to B_n, the residual busy period length. Moreover, it is clear that S_n (resp. L_n), the total number of departures (resp. losses) during the residual busy period, is equal to $S(B_n + \epsilon) = S_n$ (resp. $L(B_n + \epsilon) = L_n$), $\epsilon > 0$.

Let us denote

$$\pi_{i,j,l}(t) := \mathbb{P}\big((N(t), S(t), L(t)) = (i, j, l) \mid (n, 0, 0)\big).$$

The Laplace transform of $\pi_{i,j,l}(t)$ denotes

$$\tilde{\pi}_{i,j,l}(w) = \int_{t=0}^{\infty} e^{-wt} \pi_{i,j,l}(t) dt, \quad Re(w) \geq 0.$$

The Kolmogorov backward equations of AMC read

$$\begin{aligned}
\frac{d}{dt}\pi_{i,j,l}(t) &= -(\lambda_i + \mu_i)\pi_{i,j,l}(t) + \mu_{i+1}\pi_{i+1,j-1,l}(t) + \mathbf{1}_{\{i \geq 2\}} q_{i-1}\lambda_{i-1}\pi_{i-1,j,l}(t) \\
&\quad + p_i \lambda_i \pi_{i,j,l-1}(t), &(1)\\
\frac{d}{dt}\pi_{0,j,l}(t) &= \mu_1 \pi_{1,j-1,l}(t), &(2)
\end{aligned}$$

where $(i, j, l) \in \xi$, and by convention we assume that $\pi_{i',j',l'}(t) = 0$ for $(i', j', l') \notin \xi$. Since $(0, j, l)$ is an absorbing state it is easily seen that $\pi_{0,j,l}(t) = \mathbb{P}\big(B_n < t, S_n = j, L_n = l \mid (n, 0, 0)\big)$. Hence, the Laplace transform of the l.h.s. of (2) is equal to the joint transform $\mathbb{E}_d\big[e^{-wB_n} \cdot \mathbf{1}_{\{S_n = j\}} \cdot \mathbf{1}_{\{L_n = l\}}\big]$. Taking the Laplace transform on both sides in (2) gives that

$$\mathbb{E}_d\big[e^{-wB_n} \cdot \mathbf{1}_{\{S_n = j\}} \cdot \mathbf{1}_{\{L_n = l\}}\big] = \mu_1 \tilde{\pi}_{1,j-1,l}(w).$$

[1]The subscript d in $\mathbb{E}_d\big[e^{-wB_n} \cdot z_1^{S_n} \cdot z_2^{L_n}\big]$ is added to refer to the level dependent case.

Removing the condition on S_n and L_n we deduce that

$$\mathbb{E}_d\big[e^{-wB_n} z_1^{S_n} z_2^{L_n}\big] \quad = \quad \mu_1 z_1 \sum_{j=0}^{\infty} z_1^j \sum_{l=0}^{\infty} z_2^l \tilde{\pi}_{1,j,l}(w). \tag{3}$$

We now derive the r.h.s. of $\mathbb{E}_d\big[e^{-wB_n} z_1^{S_n} z_2^{L_n}\big]$. First, we shall compute the sum $\sum_{l=0}^{\infty} z_2^l \tilde{\pi}_{1,j,l}(w)$ and afterwards we shall sum the result over j.

Taking the Laplace transforms of the equations in (1) and writing them in matrix form we find that

$$\tilde{\Pi}_{0,0}(w)(w\mathbf{I} - \mathbf{A}) \quad = \quad e_n, \tag{4}$$

$$\tilde{\Pi}_{0,l}(w)(w\mathbf{I} - \mathbf{A}) \quad = \quad \tilde{\Pi}_{0,l-1}(w)\mathbf{C}, \quad l \geq 1, \tag{5}$$

$$\tilde{\Pi}_{j,l}(w)(w\mathbf{I} - \mathbf{A}) \quad = \quad \tilde{\Pi}_{j-1,l}(w)\mathbf{B} + \tilde{\Pi}_{j,l-1}(w)\mathbf{C}, \tag{6}$$

with $j \geq 1$ and $l \geq 1$, and where e_n represents the initial state vector with all entries equal to zero except the n-th entry that is 1, $\tilde{\Pi}_{j,l}(w)$ the Laplace transform vector that is equal to $(\tilde{\pi}_{1,j,l}(w), \cdots, \tilde{\pi}_{K,j,l}(w))$, $\mathbf{I}$ the identity matrix of order K, and where $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are introduced just before Theorem 1. Multiplying (6) by z_2^l and summing the result over all l we find that

$$\begin{aligned} \tilde{\Pi}_j(w, z_2) \quad &:= \quad \sum_{l=0}^{\infty} z_2^l \tilde{\Pi}_{j,l}(w) \\ &= \quad \tilde{\Pi}_{j-1}(w, z_2)\mathbf{B}(w\mathbf{I} - \mathbf{A} - z_2\mathbf{C})^{-1} = \cdots = \\ &= \quad \tilde{\Pi}_0(w, z_2)(\mathbf{B}\big(w\mathbf{I} - \mathbf{A} - z_2\mathbf{C})^{-1}\big)^j, \end{aligned} \tag{7}$$

where $j \geq 1$. Note that $(w\mathbf{I} - \mathbf{A} - z_2\mathbf{C})$ is invertible since it has a dominant main diagonal. The multiplication to the right of (7) with the column vector e_1^T yields that

$$\sum_{l=0}^{\infty} z_2^l \tilde{\pi}_{1,j,l}(w) = \tilde{\Pi}_0(w, z_2)(\mathbf{B}\big(w\mathbf{I} - \mathbf{A} - z_2\mathbf{C})^{-1}\big)^j e_1^T.$$

Thus, $\mathbb{E}_d\big[e^{-wB_n} \cdot z_1^{S_n} \cdot z_2^{L_n}\big]$ is equal to

$$\mu_1 z_1 \sum_{j=0}^{\infty} z_1^j \sum_{l=0}^{\infty} z_2^l \tilde{\pi}_{1,j,l}(w) = \mu_1 z_1 \tilde{\Pi}_0(w, z_2)(\mathbf{E})^{-1} e_1^T, \tag{8}$$

where $\mathbf{E} := \mathbf{I} - z_1\mathbf{B}(w\mathbf{I} - \mathbf{A} - z_2\mathbf{C})^{-1}$. It remains to find $\tilde{\Pi}_0(w, z_2)$. Equations (4) and (5) together give that

$$\tilde{\Pi}_0(w, z_2) := \sum_{l=0}^{\infty} z_2^l \tilde{\Pi}_{0,l}(w) = e_n(w\mathbf{I} - \mathbf{A} - z_2\mathbf{C})^{-1}. \tag{9}$$

Plugging (9) into (8) gives Theorem 1 rightaway. □

Let us denote $a_i = -q_i\lambda_i$, $b_i = w + \lambda_i(1 - z_2p_i) + \mu_i$, and $c_i = -z_1\mu_i$ for $i = 1, \ldots, K$. We note that the vectors $(a_1, \ldots, a_{K-1})$, $(b_1, \ldots, b_K)$, and $(c_2, \ldots, c_K)$ represent the upper-diagonal, diagonal, and lower-diagonal of the matrix $\mathbf{Q}_K$.

Proposition 1 *The joint transform B_1, S_1, and L_1 is given by*

$$\mathbb{E}_d\big[e^{-wB_1} \cdot z_1^{S_1} \cdot z_2^{L_1}\big] \quad = \quad u_1(w, z_1, z_2),$$

where $u_i(w, z_1, z_2)$, $i = 1, \ldots, K-1$, satisfies the following recursion

$$u_i(w, z_1, z_2) \quad = \quad -\frac{c_i}{b_i + a_i u_{i+1}(w, z_1, z_2)},$$

with $u_K(w, z_1, z_2) = b_K$.

Proof: According to Theorem 1 the joint transform B_1, S_1, and L_1 can be written as

$$\mathbb{E}_d\big[e^{-wB_1} \cdot z_1^{S_1} \cdot z_2^{L_1}\big] = \mu_1 z_1 q(1,1) = -c_1 q(1,1),$$

where $q(1,1)$ is the (1,1)-entry of $\mathbf{Q}_K^{-1}$. Let us partition the matrix $\mathbf{Q}_K$ as follows

$$\left(\begin{array}{c|c} b_1 & a_1 e_1 \\ \hline c_2 e_1^T & \mathbf{Q}_{K-1} \end{array}\right), \tag{10}$$

where the matrix $\mathbf{Q}_{K-1}$ is obtained from the matrix $\mathbf{Q}_K$ by removing its first row and first column. Therefore, $\mathbf{Q}_{K-1}$ is a tridiagonal matrix with upper-diagonal equal to $(a_2, \ldots, a_{K-1})$, diagonal equal to $(b_2, \ldots, b_K)$, and lower-diagonal equal to $(c_3, \ldots, c_K)$. A simple linear algebra gives that the inverse of $\mathbf{Q}_K$ reads

$$\left(\begin{array}{c|c} (q_K^*(1,1))^{-1} & -b_1^{-1} a_1 e_1 (\mathbf{Q}_{K-1}^*)^{-1} \\ \hline -c_2 \mathbf{Q}_{K-1}^{-1} e_1^T (q_K^*(1,1))^{-1} & (\mathbf{Q}_{K-1}^*)^{-1} \end{array}\right), \tag{11}$$

where $q_K^*(1,1) := b_1 - a_1 c_2 e_1 \mathbf{Q}_{K-1}^{-1} e_1^T$ and $\mathbf{Q}_{K-1}^* := \mathbf{Q}_{K-1} - \frac{a_1 c_2}{b_1} e_1^T e_1$. It is readily seen that

$$\mathbb{E}_d\big[e^{-wB_1} \cdot z_1^{S_1} \cdot z_2^{L_1}\big] = -c_1 q(1,1) = -\frac{c_1}{b_1 - a_1 c_2 e_1 \mathbf{Q}_{K-1}^{-1} e_1^T}. \tag{12}$$

Repeating the same way of partitioning of the matrix $\mathbf{Q}_K$ to $\mathbf{Q}_{K-1}$ one can show that

$$-c_2 e_1 \mathbf{Q}_{K-1}^{-1} e_1^T = -\frac{c_2}{b_2 - a_2 c_3 e_1 \mathbf{Q}_{K-2}^{-1} e_1^T},$$

where $\mathbf{Q}_{K-2}$ is obtained from the matrix $\mathbf{Q}_{K-1}$ by removing its first row and first column. For this reason, we deduce by induction that $\mathbb{E}_d\big[e^{-wB_1} \cdot z_1^{S_1} \cdot z_2^{L_1}\big]$ satisfies the recursion defined in Proposition 1. □

Remark 1 *The recursion in Proposition 1 has the following probabilistic interpretation. First, let us replace a_1, b_1, and c_1 by their values in (12). Second note that, by Theorem 1, $\mu_2 z_1 e_1 \mathbf{Q}_{K-1}^{-1} e_1^T$ is equal to the joint transform of B_1, S_1, and*

L_1 *in the M/M/1/K-1 with birth rate* $q_i\lambda_i$, $i = 1, \ldots, K-1$, *and death rate equal to* μ_i, $i = 2, \ldots, K$; *we shall refer to this transform as* $\mathbb{E}_{K-1}\big[e^{-wB_1} \cdot z_1^{S_1} \cdot z_2^{L_1}\big]$. *Therefore, we find that (12) can be written as follows*

$$\mathbb{E}_d\big[e^{-wB_1} z_1^{S_1} z_2^{L_1}\big] = \frac{\mu_1 z_1 + q_1\lambda_1 \mathbb{E}_d\big[e^{-wB_1} z_1^{S_1} z_2^{L_1}\big]\mathbb{E}_{K-1}\big[e^{-wB_1} z_1^{S_1} z_2^{L_1}\big]}{w + \mu_1 + \lambda_1(1 - p_1 z_2)}.$$

The previous equation can be readily derived by conditioning on the first event just after the starting time of the busy period B_1 *in the M/M/1/K. This event can be either a departure with rate* μ_1 *or an arrival with rate* $q_1\lambda_1$.

Proposition 2 (Threshold policy) *Let* $m \in \{1, \ldots, K\}$. *In the case where* $\lambda_i = \lambda^-$, $\mu_i = \mu^-$ *and* $p_i = p^-$ *for* $i \leq m-1$, *and* $\lambda_i = \lambda^+$, $\mu_i = \mu^+$ *for* $m \leq i \leq K$, $p_i = p^+$ *for* $m \leq i \leq K-1$, *and* $p_K = 1$, *the joint transform of* B_n, S_n, *and* L_n *in the M/M/1/K is given by*[2]
Case (A): *if* $n \leq m - 1$,

$$\mathbb{E}_{Th}\big[e^{-wB_n} \cdot z_1^{S_n} \cdot z_2^{L_n}\big] = \Big(\frac{\mu^- z_1}{q^-\lambda^-}\Big)^n \frac{P_{m-n}(s_1, s_2, \phi)}{P_m(s_1, s_2, \phi)}, \tag{13}$$

where for any tuple (a, b, c)

$$P_{i+1}(a, b, c) := a^{i+1} - b^{i+1} - c(a^i - b^i), \quad i \geq 0, \tag{14}$$

$q^- = 1 - p^-$ *and* $q^+ = 1 - p^+$, s_1 *and* s_2 *are the distinct roots of*

$$q^-\lambda^- s^2 - \big(w + \mu^- + \lambda^-(1 - z_2 p^-)\big)s + \mu^- z_1 = 0, \tag{15}$$

$$\phi = \frac{\mu^+ z_1}{q^+\lambda^+} \times \frac{P_{K-m+1}(t_1, t_2, z_2)}{P_{K-m+2}(t_1, t_2, z_2)}, \tag{16}$$

and where t_1 *and* t_2 *are the distinct roots of*

$$q^+\lambda^+ t^2 - \big(w + \mu^+ + \lambda^+(1 - z_2 p^+)\big)t + \mu^+ z_1 = 0. \tag{17}$$

Case (B): *if* $n \geq m$, $\mathbb{E}_{Th}\big[e^{-wB_n} \cdot z_1^{S_n} \cdot z_2^{L_n}\big]$ *is equal to*

$$\frac{\Big(\frac{\mu^+ z_1}{q^+\lambda^+}\Big)^{n-m+1}\Big(\frac{\mu^- z_1}{q^-\lambda^-}\Big)^{m-1} P_1(s_1, s_2, \phi) P_{K-n+1}(t_1, t_2, z_2)}{P_m(s_1, s_2, \phi) P_{K-m+2}(t_1, t_2, z_2)}. \tag{18}$$

Proof: The application of Theorem 1 for the special case of the M/M/1/K queue with $\lambda_i = \lambda^-$, $\mu_i = \mu^-$ and $p_i = p^-$ for $i \leq m-1$, and $\lambda_i = \lambda^+$, $\mu_i = \mu^+$ for $m \leq i \leq K$ and $p_i = p^+$ for $m \leq i \leq K-1$, and $p_K = 1$, gives that

$$\mathbb{E}_{Th}\big[e^{-wB_n} \cdot z_1^{S_n} \cdot z_2^{L_n}\big] \quad = \quad \mu^- z_1 e_n \mathbf{Q}_K^{-1} e_1^T,$$

[2] The subscript Th in $\mathbb{E}_{Th}\big[e^{-wB_n} \cdot z_1^{S_n} \cdot z_2^{L_n}\big]$ is added to refer to the threshold policy.

where $\mathbf{Q}_K$ has the following canonical form

$$\mathbf{Q}_K = \left(\begin{array}{c|c} \mathbf{T}_{11} & \mathbf{T}_{12} \\ \hline \mathbf{T}_{21} & \mathbf{T}_{22} \end{array} \right). \tag{19}$$

The matrix $\mathbf{T}_{11}$ is a tridiagonal Toeplitz $(m-1)$-by-$(m-1)$ matrix with diagonal entries equal to $w+\lambda^-(1-p^-z_2)+\mu^-$, upper-diagonal entries equal to $-q^-\lambda^-$, and lower-diagonal entries equal to $-z_1\mu^-$. The matrix $\mathbf{T}_{22}$ is the sum of $(-q^+\lambda^+z_2)h\cdot h^T$, where h is the column vector of dimension $K-m+1$ with all entries equal to zero except the $(K-m+1)$-st that is 1, and of a tridiagonal Toeplitz $(K-m+1)$-by-$(K-m+1)$ matrix with diagonal entries equal to $w+\lambda^+(1-p^+z_2)+\mu^+$, upper-diagonal entries equal to $-q^+\lambda^+$, and lower-diagonal entries equal to $-z_1\mu^+$. The matrix $\mathbf{T}_{12}$ is equal to $-(q^-\lambda^-)u\cdot v$, where u is the column vector of dimension $m-1$ with all entries equal to zero except the last entry that is 1 and v the row vector of dimension $K-m+1$ with all entries equal to zero except the first that is 1. Finally, the matrix $\mathbf{T}_{21}$ is equal to $-(z_1\mu^+)v^T\cdot u^T$.

Note that the joint transform

$$\mathbb{E}_{Th}\left[e^{-wB_n}\cdot z_1^{S_n}\cdot z_2^{L_n}\right] \quad = \quad \mu^-z_1q(n,1),$$

where $q(n,1)$ is the $(n,1)$-entry of $\mathbf{Q}_K^{-1}$. By analogy with the derivation of the inverse of $\mathbf{Q}_K$ in (11) we find that

$$\mathbf{Q}_K^{-1} = \left(\begin{array}{c|c} (\mathbf{T}_{11}^*)^{-1} & -\mathbf{T}_{11}^{-1}\mathbf{T}_{12}(\mathbf{T}_{22}^*)^{-1} \\ \hline -\mathbf{T}_{22}^{-1}\mathbf{T}_{21}(\mathbf{T}_{11}^*)^{-1} & (\mathbf{T}_{22}^*)^{-1} \end{array} \right),$$

where $\mathbf{T}_{11}^* := \mathbf{T}_{11}-\mathbf{T}_{12}\mathbf{T}_{22}^{-1}\mathbf{T}_{21}$ and $\mathbf{T}_{22}^* := \mathbf{T}_{22}-\mathbf{T}_{21}\mathbf{T}_{11}^{-1}\mathbf{T}_{12}$.

Case (A) $n \le m-1$. In this case $q(n,1)$ is equal to the $(n,1)$-entry of $(\mathbf{T}_{11}^*)^{-1}$. We note that

$$q(n,1) = e_n\big(\mathbf{T}_{11} - \big(q^-\lambda^-\mu^+z_1t_{22}(1,1)\big)u\cdot u^T\big)^{-1}e_1^T, \tag{20}$$

where $t_{22}(1,1)$ is the (1,1)-entry of $\mathbf{T}_{22}^{-1}$. Let us denote

$$\phi = \mu^+z_1t_{22}(1,1). \tag{21}$$

Note that $\mathbf{T}_{11}$ is a tridiagonal Toeplitz $(m-1)$-by-$(m-1)$ matrix. Therefore, the value of $q(n,1)$ can be deduced from Lemma 3 in the Appendix (with $a=-z_1\mu^-$, $b=w+\lambda^-(1-p^-z_2)+\mu^-$, $c=-q^-\lambda^-$, $M=m-1$, and $\alpha=q^-\lambda^-\phi$). Plugging $q(n,1)$ into (20) gives (13). Note that by Rouché's Theorem it is easily checked that Eq. (15) has a unique root within the unit disk. In the case where $z_2=1$, the root with smallest absolute value can be interpreted as the joint transform of the busy period length and the number of departures in the busy period in the M/M/1 queue with arrival rate $q^-\lambda^-$ and service rate μ^- [8, Chap. 3, Sec. 3]. Moreover, it is easily verified that these roots are equal 1 and $\mu^-/(q^-\lambda^-)$ for $w=0$ and $z_1=z_2=1$.

It remains to find ϕ. Note that $t_{22}(1,1)$ is the (1,1)-entry of $\mathbf{T}_{22}^{-1}$. By definition, the matrix $\mathbf{T}_{22}$ is the sum of a tridiagonal Toeplitz $(K-m+1)$-by-$(K-m+1)$ matrix and of $(-q^+\lambda^+z_2)h\cdot h^T$. Plugging the value of $t_{22}(1,1)$ given in Lemma 4 in the Appendix (with $a=-z_1\mu^+$, $b=w+\lambda^+(1-p^+z_2)+\mu^+$, $c=-q^+\lambda^+$, $\alpha=q^+\lambda^+z_2$, $M=K-m+1$, and $n=1$) into (21) yields (16). This completes the proof of the case where $n\le m-1$.

Case (B) $n\ge m$. Let us denote $l=n-m+1$. Since $n\ge m$, $q(n,1)$ is equal to the $(l,1)$-entry of $-\mathbf{T}_{22}^{-1}\mathbf{T}_{21}(\mathbf{T}_{11}^*)^{-1}$. Note that

$$-e_l\mathbf{T}_{22}^{-1}\mathbf{T}_{21}=\mu^+z_1t_{22}(l,1)u^T,$$

where $t_{22}(l,1)$ is the $(l,1)$-entry of $\mathbf{T}_{22}^{-1}$ and u was defined as the column vector of dimension $m-1$ with all entries equal to zero except the last one that is 1. Therefore, we find that

$$\mathbb{E}_{Th}\big[e^{-wB_n}z_1^{S_n}z_2^{L_n}\big]=\mu^+z_1t_{22}(l,1)\mu^-z_1u^T(\mathbf{T}_{11}^*)^{-1}e_1, \tag{22}$$

where $\mu^-z_1u^T(\mathbf{T}_{11}^*)^{-1}e_1$ is given in (13) with $n=m-1$, which reads

$$\mu^-z_1u^T(\mathbf{T}_{11}^*)^{-1}e_1=\frac{\left(\frac{\mu^-z_1}{q^-\lambda^-}\right)^{m-1}(s_1-s_2)}{s_1^m-s_2^m-\phi(s_1^{m-1}-s_2^{m-1})},$$

where ϕ is given in (16). Thus, it remains to find $t_{22}(l,1)$ to complete the proof. By analogy with the derivation of $t_{22}(1,1)$, the entry $t_{22}(l,1)$ is given in Lemma 4 in the Appendix (with $a=-z_1\mu^+$, $b=w+\lambda^+(1-p^+z_2)+\mu^+$, $c=-q^+\lambda^+$, $\alpha=q^+\lambda^+z_2$, $M=K-m+1$, and $n=l$). Plugging the values of $t_{22}(l,1)$ and $\mu^-z_1(\mathbf{T}_{11}^*)^{-1}u^Te_1$ into (22) we find (18). This completes the proof. □

Corollary 1 (Static policy) *In the level independent M/M/1/K queue with $\lambda_i=\lambda$, $\mu_i=\mu$, and $p_j=p$ for $i=1,\dots,K$ and $j=1,\dots,K-1$, the joint transform of B_n, S_n, and L_n is given by*

$$\mathbb{E}_S\big[e^{-wB_n}\cdot z_1^{S_n}\cdot z_2^{L_n}\big]=\Big(\frac{z_1\mu}{q\lambda}\Big)^n\frac{P_{K-n+1}(r_1,r_2,z_2)}{P_{K+1}(r_1,r_2,z_2)}, \tag{23}$$

where $P_i(r_1,r_2,z_2)$, $i=K-n+1,K+1$, is given in (14) (with $a=r_1$, $b=r_2$ and $c=z_2$), and r_1 and r_2 are the distinct roots of

$$q\lambda r^2-\big(w+\mu+\lambda(1-z_2p)\big)r+\mu z_1=0. \tag{24}$$

Proof: Proposition 2 applied to the M/M/1/K queue with $\lambda^+=\lambda^-=\lambda$, $\mu^+=\mu^-=\mu$, and $p^+=p^-=p$ readily proves Corollary 1. □

Remark 2 *We emphasize that Corollary 1 extends the result of Rosenlund on the M/M/1/K in [11] in two ways. First, it gives the joint transform of B_n, S_n, L_n for the case when $n>1$. Second, it allows the dropping of customers even when the queue is not full.*

3 Performance metrics

Let us first consider the level independent M/M/1/K given in Corollary 1. Let ρ denote the load, i.e. $\rho = q\lambda/\mu$. Moreover, let us denote $\rho_0 := \lambda/\mu$, $\mathbb{V}[X]$ the variance of the rv X, $\mathbb{C}\text{ov}[X,Y]$ the covariance of the rvs X and Y. Taking the derivative of the $\mathbb{E}_S\big[e^{-wB_n} \cdot z_1^{S_n} \cdot z_2^{L_n}\big]$ given in (23) according to w, z_1, and z_2 we find the variances of the three rvs B_n, S_n, and L_n and their covariances. We note that the formulae of $\mathbb{V}_S[S_n]$, $\mathbb{V}_S[L_n]$, $\mathbb{C}\text{ov}_S[B_n,S_n]$, $\mathbb{C}\text{ov}_S[B_n,L_n]$, and $\mathbb{C}\text{ov}_S[S_n,L_n]$ for $\rho \neq 1$ and $n > 1$, are lengthy; for this reason we shall just report these measures for $n = 1$.

1- Marginal measures: $\rho \neq 1$

$$\mathbb{E}_S[B_n] = \frac{1}{\mu} \cdot \frac{n(1-\rho) - (1-\rho^n)\rho^{K-n+1}}{(1-\rho)^2},$$

$$\begin{aligned}\mathbb{V}_S[B_n] \quad = \quad & \frac{1}{\mu(1-\rho)^4}\Big[n(1-\rho^2) + 4(K+1)\rho^{K+1} - 4K\rho^{K+2} \\ & -4(K-n+1)\rho^{K-n+1} + 4(K-n)\rho^{K-n+2} - \rho^{2K-2n+2} \\ & +\rho^{2K+2}\Big],\end{aligned}$$

$$\mathbb{E}_S[S_n] = \mu\mathbb{E}_S[B_n] = \frac{n(1-\rho) - (1-\rho^n)\rho^{K-n+1}}{(1-\rho)^2},$$

$$\mathbb{V}_S[S_1] = \frac{(1+\rho)\big(\rho + (1-2K)(1-\rho)\rho^K - \rho^{2K}\big)}{(1-\rho)^3},$$

$$\mathbb{E}_S[L_n] = \frac{n(1-\rho)(\rho_0-\rho) + (1-\rho_0)(1-\rho^n)\rho^{K-n+1}}{(1-\rho)^2},$$

$$\begin{aligned}\mathbb{V}_S[L_1] \quad = \quad & \frac{-1}{(1-\rho)^3}\big[(1-\rho_0)^2(1+\rho)\rho^{2K} + (3-\rho_0-4K+4K\rho_0)\rho^{K+2} \\ & -2(1-2K+\rho_0+2\rho_0^2)\rho^{K+1} - (1-3\rho_0+4K\rho_0-4K\rho_0^2) \\ & -(3-\rho_0)\rho^2 + (1+4\rho_0-\rho_0^2)\rho - \rho_0(1+\rho_0))\big].\end{aligned}$$

2- Joint measures: $\rho \neq 1$

$$\begin{aligned}\mathbb{C}\text{ov}_S[B_1,S_1] \quad = \quad & \frac{1}{\mu(1-\rho)^3}\big[-\rho^{2K+1} - \rho^{2K} + K\rho^{K+2} + (2K-1)\rho^{K+1} \\ & -(3K-1)\rho^K + 2\rho\big],\end{aligned}$$

$$\begin{aligned}\mathbb{C}\text{ov}_S[B_1,L_1] \quad = \quad & \frac{1}{\mu(1-\rho)^3}\big[(1-\rho_0)(1+\rho)\rho^{2K} - (2K-1)\rho^{K+2} + 4K\rho_0\rho^{K+1} \\ & -(1-2K+4K\rho_0)\rho^K - \rho^2 - (1-\rho_0)\rho + \rho_0\big],\end{aligned}$$

$$\begin{aligned}\mathbb{C}\text{ov}_S[S_1, L_1] &= \frac{1}{(1-\rho)^3}\big[(1-\rho_0)(1+\rho)\rho^{2K} + (2-3K+K\rho_0)\rho^{K+2} \\ &\quad -(1-2K+\rho_0-2K\rho_0)\rho^{K+1} - (1-\rho_0-K+3K\rho_0)\rho^K \\ &\quad +2\rho_0\rho - 2\rho^2\big],\end{aligned}$$

$$\mathbb{E}\Big[\frac{L_1}{S_1}\Big] = \begin{cases} \rho_0 & , \quad K=1 \\ \rho_0 + \frac{\rho_o p}{1+\rho} + \frac{1}{q}\ln\left(\frac{1}{1+\rho}\right) & , \quad K=2 \end{cases}$$

In the case where $K=1$, the rv S_1 becomes equal to one w.p. 1, and B_1 is distributed exponentially with parameter μ. Moreover, note that $p_{K=1}=1$; this means that during the busy period all the arriving customers are dropped w.p. 1. Therefore, we deduce that $\mathbb{E}[L_1/S_1] = \mathbb{E}[L_1] = \lambda/\mu$.

Remark 3 *It is easy to see that $\mathbb{E}_S[L_n] = n$ for $\rho_0 = 1$. This result extends Abramov's result in [1] on the expected number of losses during the busy period for M/M/1/K in two ways. First, it allows the early drop of the customers even when the queue is not full. Second, it gives the expected number of losses during the residual busy period that is initiated with n customers, $n > 1$.*

Case B: $\rho = 1$

1- Marginal measures: $\rho = 1$

$$\mathbb{E}_S[B_n] = \frac{1}{\mu}\frac{n(2K-n+1)}{2},$$

$$\begin{aligned}\mathbb{V}_S[B_n] &= \frac{n}{6\mu^2}\big[4K^3+6K^2+4K-6K^2n+4Kn^2 \\ &\quad -6Kn-n^3+2n^2-2n+1\big],\end{aligned}$$

$$\mathbb{E}_S[S_n] = \frac{n(2K-n+1)}{2},$$

$$\mathbb{V}_S[S_n] = \frac{n(4K^3-6K^2-2K+4Kn^2-n^3+n)}{6},$$

$$\mathbb{E}_S[L_n] = \frac{n(2Kp-np-p+2)}{2q},$$

$$\begin{aligned}\mathbb{V}_S[L_n] &= \frac{n}{6q^2}\big[4K^3+6K^2+4K+12K^2nq-6K^2nq^2+6Knq^2+4Kn^2q^2 \\ &\quad -8Kn^2q-6Kn-2Kq^2-6K^2q^2-8K^3q+4K^3q^2-6K^2n+4Kn^2 \\ &\quad +10Kq-n^3q^2-5qn-n^3+2n^2-2n^2q^2+2n^3q+q^2n-2n+2q^2 \\ &\quad +3q+1\big].\end{aligned}$$

2- Joint measures: $\rho = 1$

$$\mathbb{C}\mathrm{ov}_S[B_n, S_n] = \frac{1}{6\mu}\big[4K^3 + 3K^2 - 2K - 6K^2n - 3Kn + 4Kn^2 - n^3 + n^2 + n - 1\big],$$

$$\mathbb{C}\mathrm{ov}_S[B_n, L_n] = \frac{n}{6\mu q}\big[4K^3 + 6K^2 + 4K - 6K^2n - 6Kn + 4Kn^2 - 4K^3q + 6K^2nq - 4Kn^2q + 2Kq + n^3q - qn - n^3 - 2n + 2n^2 + 1\big],$$

$$\mathbb{C}\mathrm{ov}_S[S_n, L_n] = \frac{n}{6q}\big[4K^3 + 3K^2 - 2K + 6K^2nq - 4Kn^2q - 3Knq - 4K^3q + 3K^2q - 6K^2n - 3Kn + 4Kn^2 - n^3 - qn + n^2q + n^3q + n^2 + n + 2qK - q - 1\big],$$

$$\mathbb{E}\Big[\frac{L_1}{S_1}\Big] = \begin{cases} \frac{1}{q} & , \quad K = 1 \\ \frac{3 - 2\ln(2) - q}{2q} & , \quad K = 2 \end{cases}$$

Let us now consider the threshold policy given in Proposition 2. For the sake of the ease of presentation, we shall restrict ourselves to the case where $n = 1$ and to deriving the first moment marginal measures. Let $\rho^- = q^-\lambda^-/\mu^-$ and $\rho^+ = q^+\lambda^+/\mu^+$. Taking the derivative of the $\mathbb{E}_{Th}\big[e^{-wB_1} \cdot z_1^{S_1} \cdot z_2^{L_1}\big]$ given in (13) according to w, z_1, and z_2 we find that, for $\rho^+ \neq 1$ and $\rho^- \neq 1$,

$$\mathbb{E}_{Th}[B_1] = \frac{1}{\rho^-\mu^+\mu^-(1-\rho^-)(1-\rho^+)}\big[\mu^-\rho^+(1-\rho^-) \times (\rho^+)^{K-m}(\rho^-)^m + (\rho^+\mu^+ - \rho^-\mu^- + \mu^- - \mu^+)(\rho^-)^m + (1-\rho^+)\mu^+\rho^-\big],$$

$$\mathbb{E}_{Th}[S_1] = \frac{1}{\rho^-(1-\rho^-)(1-\rho^+)}\big[\rho^+(\rho^- - 1)(\rho^+)^{K-m} \times (\rho^-)^m + (\rho^+ - \rho^-)(\rho^-)^m - \rho^-(\rho^+ - 1)\big],$$

$$\mathbb{E}_{Th}[L_1] = \frac{1}{q^-q^+\rho^-(1-\rho^-)(1-\rho^+)}\big[q^-(\rho^- - 1)(\rho^+)^2 \times (\rho^+)^{K-m}(\rho^-)^m + q^-q^+\rho^+(1-\rho^-) \times (\rho^+)^{K-m}(\rho^-)^m + ((q^+ - q^-)\rho^+\rho^- + (\rho^- - \rho^+)q^-q^+ + \rho^+q^- - \rho^-q^+)(\rho^-)^m - (\rho^-)^2q^-q^+ + (\rho^-)^2q^+ + \rho^+(\rho^-)^2q^-q^+ - \rho^+(\rho^-)^2q^+\big].$$

4 Miscellaneous results

For simplicity we consider in this section the level independent M/M/1/K with $\lambda_i = \lambda$, $\mu_i = \mu$, and $p_j = p$ for $i = 1, \ldots, K$ and $j = 1, \ldots, K-1$, and $p_K = 1$.

Distribution of the maximum number of customers simultaneously present during the residual busy period: Let M_n^K denote the maximum number of customers simultaneously present in the M/M/1/K during its residual busy period B_n.

Lemma 1 *For $n \in \{1, \cdots, K-1\}$, we have:*

$$\begin{aligned}
\mathbb{P}[M_n^K = h] &= 0, \quad 1 \le h \le n-1, \\
\mathbb{P}[M_n^K = h] &= \frac{(1-\rho)\rho^{h-n}(1-\rho^n)}{(1-\rho^h)(1-\rho^{h+1})}, \quad n \le h \le K-1, \\
\mathbb{P}[M_n^K = K] &= \rho^{K-n}\frac{1-\rho^n}{1-\rho^K}.
\end{aligned}$$

Proof: In this proof, we will exclude the trivial case where $h < n+1$ which induces that $\mathbb{P}[M_n^K < h] = 0$. Let us focus on the event $\{M_n^K < h\}$ in the case where $n < h \le K$. Let us consider the queue length process $N(t) \in \{0, 1, \cdots, K\}$ during the residual busy period B_n. By assumption, let us force the state 0 and the set $\{h, \cdots, K\}$ to be absorbing states. It is then clear that $\mathbb{P}[M_n^K < h]$ is equal to the probability that the absorption occurs in the set $\{h, \cdots, K\}$. Let $\mathbf{G}_{h-1}$ denote the transient generator of the previous absorbing Markov process. Thus, $\mathbf{G}_{h-1}$ is a $(h-1)$-by-$(h-1)$ tridiagonal Toeplitz matrix with upper-diagonal elements equal to $q\lambda$, diagonal elements $-(q\lambda + \mu)$, and lower diagonal elements μ. The theory of absorbing Markov processes yields (see, e.g., [6])

$$\mathbb{P}[M_n^K < h] = -\mu e_n \mathbf{G}_{h-1}^{-1} e_1^T = -\mu g_{n1}, \tag{25}$$

where g_{n1} is the (n,1)-entry of $\mathbf{G}_{h-1}^{-1}$. Lemma 3 with $a = \mu$, $b = -(q\lambda + \mu)$, $c = q\lambda$, $M = h-1$ gives that

$$-\mu g_{n1} = -\mu \frac{(x_1^{-1} - x_2^{-1})(x_1^h x_2^n - x_2^h x_1^n)}{q\lambda(x_1 - x_2)(x_1^h - x_2^h)}, \tag{26}$$

where x_1 and x_2 are the roots of $q\lambda x^2 - (q\lambda + \mu)x + \mu$ which yields that $x_1 = 1$ and $x_2 = \mu/(q\lambda) = 1/\rho$. Plugging $x_1 = 1$ and $x_2 = 1/\rho$ into (26) we find that

$$\mathbb{P}[M_n^K < h] = \frac{1-\rho^{h-n}}{1-\rho^h}. \tag{27}$$

Knowing that $\mathbb{P}[M_n^K = h] = \mathbb{P}[M_n^K < h+1] - \mathbb{P}[M_n^K < h]$ readily gives the desired result. Now, given that $\mathbb{P}[M_n^K \le K] = \mathbb{P}[M_n^K < K] + \mathbb{P}[M_n^K = K] = 1$ one can easily derive $\mathbb{P}[M_n^K = K]$. $\square$

Remark 4 *Observe that* $\mathbb{P}[M_n^K < K] = \mathbb{P}[M_n^\infty < K]$*, where* M_n^∞ *is the maximum number of customers simultaneously present in the M/M/1/*∞ *during its residual busy period* B_n*. The probability distribution of* M_n^∞ *has been derived in Cohen [4, Chap. II.2, p. 192] and it agrees with our result. Therefore, given that* $\mathbb{P}[M_n^K \le K] = 1$ *one can immediately find* $\mathbb{P}[M_n^K = K]$*. We note that the advantage of our proof is that it avoids the use of the complex function theory required in Cohen's derivation. For more general results about the distribution of the maximum number of customers during the busy period in the level dependent M/M/1/K we refer to [2] and [3, p. 73].*

The c-congestion period: Given that the queue length process has hit level c during the busy period at 0, let O_c denote the first time that the queue length hits level $c-1$ after 0, S_c denote the number of customers served during O_c, and L_c denote the number of customers dropped during O_c. We emphasize that O_1 is equal to the busy period B_1. Moreover, it is easy to see that the joint transform $\mathbb{E}\big[e^{-wO_c} \cdot z_1^{S_c} \cdot z_2^{L_c}\big] = \mathbb{E}_S\big[e^{-wB_1} \cdot z_1^{S_1} \cdot z_2^{L_1}\big]$ with queue size equal to $K-c+1$ customers. Therefore, replacing K by $K-c+1$ and n by one in (23) gives that

$$\mathbb{E}\big[e^{-wO_c} \cdot z_1^{S_c} \cdot z_2^{L_c}\big] \quad = \quad \frac{z_1\mu}{q\lambda} \times \frac{P_{K-c+1}(r_1, r_2, z_2)}{P_{K-c+2}(r_1, r_2, z_2)},$$

where $P_i(r_1, r_2, z_2)$, $i = K-c+1, K-c+2$, is given in (14) (with $a = r_1$, $b = r_2$ and $c = z_2$), and r_1 and r_2 are the distinct roots of (24).

Distribution of the number of c-congestion periods during the residual busy period: Let N_c denote the total number of visits to state c in the M/M/1/K during its residual busy period B_n. Due to the Markov property of the queue length process it is clear that N_c is a modified geometric rv, i.e., the distribution of N_c reads

$$\mathbb{P}[N_c = 0] = 1 - f_0, \quad \mathbb{P}[N_c = h] = f_0 f^{h-1}(1-f),$$

where $h = 1, 2, \ldots$, f_0 is the first passage probability of the queue length process to state c during the remaining busy period that is initiated with n customers, and f is the first passage probability of the queue length process to state c during the remaining busy period that is initiated with $c-1$ customers. The probabilities f_0 and f can be written in terms of the distribution of the maximum number of customers present in the queue during its residual busy period B_n, M_n^K, as follows

$$\begin{aligned} f_0 &= 1 - \mathbb{P}[M_n^K < c] = \begin{cases} 1 & , \quad n \ge c \\ \rho^{c-n}\frac{1-\rho^n}{1-\rho^c} & , \quad n < c \end{cases} \\ f &= 1 - \mathbb{P}[M_{c-1}^K < c] = \rho\frac{1-\rho^{c-1}}{1-\rho^c}, \end{aligned}$$

where the second equality in the previous equations follows directly from Lemma 1. Note that when $n \ge c$ the probability $\mathbb{P}[N_c = 0] = 0$ and when $n < c$ the probability $\mathbb{P}[N_c = 0] > 0$.

P.g.f. of the number of visits to state h during the residual busy period: Let V_h denote the total number of visits to state h in the M/M/1/K during its residual busy period B_n that is initiated with n customers.

Lemma 2 *The p.g.f. of V_h is given by*

$$\mathbb{E}\big[z^{V_h}\big] = \begin{cases} -\mu e_n \big(\mathbf{G}^*\big)^{-1} e_1^T & , \quad h \neq n, \\ -\mu z e_n \big(\mathbf{G}^*\big)^{-1} e_1^T & , \quad h = n, \end{cases}$$

where $G := \mathbf{G}^*{}_K + \lambda e_K^T e_K - z e_{h-1}^T e_h$ and e_i is the unit row vector with all elements equal to zero except the i-th element that is one.

Proof: By analogy with the proof of Theorem 1, let $(N(t), H(t))$ denote the two dimensional, continuous-time Markov process with discrete state-space $\xi := \{0, 1, \cdots, K\} \times \mathbb{N}$, where $N(t)$ represents the number of customers in the queue at time t, and $H(t)$ the number of visits to state h until t. States $(0, \cdot)$ are absorbing. We refer to this absorbing Markov process by AMC$_h$. The absorption of AMC$_h$ occurs when the queue becomes empty, i.e., $N(t) = 0$. By setting the initial state of AMC$_h$ at $t = 0$ to $(n, 0)$, $n \geq 1$ and $n \neq h$, the time until absorption is equal to B_n, the residual busy period length. Moreover, it is clear that V_h, the total number of visits to state h during the residual busy period, is equal to $H(B_n + \epsilon)$, $\epsilon > 0$. In the case when $n = h$, i.e., the initial state of the AMC$_h$ is $(n, 1)$, we have that $V_h = H(B_n + \epsilon) + 1$. Following the footprints of the proof of Theorem 1, i.e., first writing down the Kolomogorov equations of AMC$_h$, second taking their transforms and presenting them in a matrix form, and finally solving the matrix equations, gives the desired result in Lemma 2.

5 Conclusion and future research

In this paper, we determined the closed-form expression for the joint transform of the length of the busy period of the state dependent M/M/1/K queue, the number of customers served during the busy period, and the number of losses during the busy period. For two different policies referred to as the threshold policy and the static policy we derived simple expressions for their joint transform. Moreover, we derived the distribution of the maximum queue level reached during the busy period, the transform of the c-congestion period and the total number of such periods during a busy period.

In future research, we aim at generalizing Theorem 1 for the PH/PH/1/K queue where PH stands for phase type distribution. Basically, in Theorem 1 the tridiagonal fundamental matrix becomes a tridiagonal block matrix. There are two ways to invert this tridiagonal block matrix. The first one is similar to the numerical recursion reported in Proposition 1. The other approach is a transform based approach that requires the computation of the determinant of a matrix. The result in this case is a function of the roots of a polynomial and its form is a fraction of two polynomials of these roots of order K-1 and K. For the PH/M/1/K and M/PH/1/K these roots are simply identified as solution of a function that involves the LST of inter-arrival times/service times.

Appendix

A Toeplitz matrix is a matrix in which all the diagonal elements are equal. Let $\mathbf{T}$ denote the M-by-M tridiagonal Toeplitz matrix with lower-diagonal elements equal to a, diagonal elements equal to b, and upper-diagonal elements equal to c.

Lemma 3 *The (i,j)-entry of $\mathbf{T}^{-1}$ equals*

$$t_{ij} = \begin{cases} -\frac{(x_1^i - x_2^i)(x_1^{M+1-j} - x_2^{M+1-j})}{c(x_1-x_2)(x_1^{M+1}-x_2^{M+1})} & , \ i \le j \le M \\ \frac{(x_1^{-j} - x_2^{-j})(x_1^{M+1}x_2^i - x_2^{M+1}x_1^i)}{c(x_1-x_2)(x_1^{M+1}-x_2^{M+1})} & , \ j \le i \le M \end{cases} \tag{28}$$

where x_1 and x_2 are the roots of

$$cx^2 + bx + a = 0.$$

Proof: See [5, Sec. 3.1]. □

Let $\mathbf{T}^* := \mathbf{T} - \alpha e_M^T \cdot e_M$, where e_M is the unit row vector with all entries equal to zero except the M-th entry that is one.

Lemma 4 *The $(n,1)$-entry of $(\mathbf{T}^*)^{-1}$ equals*

$$\begin{aligned} t_{n1}^* &= \frac{-1}{c}\left(\frac{a}{c}\right)^{n-1} \frac{1}{c(x_1^{M+1} - x_2^{M+1}) + \alpha(x_1^M - x_2^M)} \\ &\quad \times \left[c(x_1^{M-n+1} - x_2^{M-n+1}) + \alpha(x_1^{M-n} - x_2^{M-n})\right]. \end{aligned} \tag{29}$$

Proof: The application of the Sherman-Morrison formula [9, p. 76] to $(\mathbf{T}^*)^{-1}$ gives that

$$t_{n1}^* = t_{n1} + \alpha \frac{t_{nM} t_{M1}}{1 - \alpha t_{MM}}.$$

Plugging the values of t_{ij} given in Lemma 3 into the previous equation gives that

$$\begin{aligned} t_{n1}^* &= \frac{1}{cx_1x_2(x_1^{M+1} - x_2^{M+1})}\Big(-(x_1^{M+1}x_2^n - x_2^{M+1}x_1^n) + \\ &\quad \alpha \frac{(x_1^n - x_2^n)(x_1^{M+1}x_2^M - x_2^{M+1}x_1^M)}{c(x_1^{M+1} - x_2^{M+1}) + \alpha(x_1^M - x_2^M)}\Big) \\ &= \frac{-1}{c} \times \frac{(x_1x_2)^{n-1}(c(x_1^{M-n+1} - x_2^{M-n+1}) + \alpha(x_1^{M-n} - x_2^{M-n}))}{c(x_1^{M+1} - x_2^{M+1}) + \alpha(x_1^M - x_2^M)} \\ &= \frac{-1}{c}\left(\frac{a}{c}\right)^{n-1} \times \frac{c(x_1^{M-n+1} - x_2^{M-n+1}) + \alpha(x_1^{M-n} - x_2^{M-n})}{c(x_1^{M+1} - x_2^{M+1}) + \alpha(x_1^M - x_2^M)}, \end{aligned} \tag{30}$$

which completes the proof.

Acknowledgement

The authors would like to thank Alain Jean-Marie and Yoni Nazarathy for stimulating discussions about the paper.

References

[1] V. Abramov. On a property of refusal systems. *J. Appl. Prob.*, 34:800–805, 1997.

[2] R. Boucherie. On the maximum number of customers simultaneously present in a queue. Technical report, CWI, BS-R9423, 1994.

[3] K. Chung. *Markov chains with stationary transition probabilities.* Springer-Verlag, 1967.

[4] J. W. Cohen. *The single server queue.* North-Holland, 1982.

[5] M. Dow. Explicit inverses of Toeplitz and associated matrices. *ANZIAM J.*, 44(E):E185–E215, Jan. 2003.

[6] D. P. Gaver, P. A. Jacobs, and G. Latouche. Finite birth-and-death models in randomly changing environments. *Adv. Appl. Prob.*, 16:715–731, 1984.

[7] T. Lee. M/G/1/K queue with vacation time and exhaustive service discipline. *Operations Research*, 32(4):774–784, 1984.

[8] N. U. Prabhu. *Stochastic storage processes.* Springer-Verlag, New York, 1998.

[9] W. Press, B. Flannery, S. Teukolsky, and W. Vetterling. *Numerical recipes in C: The art of scientific computing.* Cambridge University Press, 1992.

[10] R. Righter. A note on losses in M/G/1/n queues. *J. Appl. Prob.*, 36:1240–1243, 1999.

[11] S. I. Rosenlund. On the length and number of served customers of the busy period of a generalised M/G/1 queue with finite waiting room. *Adv. Appl. Prob.*, 5:379–389, 1973.

[12] S. I. Rosenlund. Busy periods in time-dependent M/G/1 queue. *Adv. Appl. Prob.*, 8:195–208, 1976.

[13] L. Takács. *Introduction to the theory of queues.* Oxford University Press, 1962.

[14] H. Takagi and R. Lamaire. On the busy period of an M/G/1/K queue. *Operations Research*, 42(1):192–193, 1994.

[15] R. Wolff. Losses per cycle in a single-server queue. *J. Appl. Prob.*, 39:905–909, 2002.