STEADY-STATE ANALYSIS OF A MULTICLASS MAP/PH/*c* QUEUE WITH ACYCLIC PH RETRIALS

TUĞRUL DAYAR * ** AND

M. CAN ORHAN,* *Bilkent University*

Abstract

A multiclass *c*-server retrial queueing system in which customers arrive according to a class-dependent Markovian arrival process (MAP) is considered. Service and retrial times follow class-dependent phase-type (PH) distributions with the further assumption that PH distributions of retrial times are acyclic. A necessary and sufficient condition for ergodicity is obtained from criteria based on drifts. The infinite state space of the model is truncated with an appropriately chosen Lyapunov function. The truncated model is described as a multidimensional Markov chain, and a Kronecker representation of its generator matrix is numerically analyzed.

Keywords: Markovian arrival process; phase-type service time distribution; acyclic phase-type retrial time distribution; Kronecker product; Markov chain

2010 Mathematics Subject Classification: Primary 60J22
Secondary 60J28; 68M20

1. Introduction

Retrial queues are queueing systems in which an arriving customer who finds all servers busy joins an infinite retrial queue (called the orbit) and retries to receive service. Such systems have been considered in various application areas, such as call centers, computer networks, and telecommunication systems. Detailed overviews and bibliographical information about retrial queues may be obtained from the surveys by Gómez-Corral [29] and Kim and Kim [33], the books by Artalejo and Gómez-Corral [4] and Falin and Templeton [26], and also from the bibliographies by Artalejo [1], [2].

Criteria based on Lyapunov functions have been widely used for the stability analysis of retrial queues. The following is a review to that end in which the list of references is not exhaustive but covers many related papers. For single- and multi-class M/M/*c* queues with exponential retrials respectively in [26] and [47], simple linear Lyapunov functions are used to show the sufficiency and necessity of an ergodicity condition. Artalejo and Phung-Duc [5] extended a single-class M/M/*c* queue with exponential retrial by allowing outgoing calls. Sakurai and Phung-Duc [44] considered a more general model than that in [5] with multiple types of outgoing calls. In both systems, simple linear Lyapunov functions are used in the sufficiency proofs of ergodicity conditions.

Analysis of a system in which arrival or service processes include phases is more complicated. For such systems, one possible approach is to obtain some of the parameters of the Lyapunov function using algebraic properties of the blocks of an auxiliary M/G/1-type matrix (see [38]). Phung-Duc and Kawanishi [40] considered an M/M/*c* retrial queue with exponential retrials

and after-call work, and allowed for customer abandonment in [41]. Diamond and Alfa [22] considered a retrial queue in which customers arrive according to a Markovian arrival processes (MAP) [37], service times follow phase-type (PH) [38] distributions, and retrial times follow exponential distributions. For these three systems, the auxiliary M/G/1-type matrices are obtained from their classical queueing system counterparts in which customers join infinite waiting lines instead of orbits. Dudin and Klimenok [24] defined asymptotically quasi-Toeplitz Markov chains (AQTMCs) whose blocks converge to the blocks of an M/G/1-type matrix as the level numbers increase. The sufficiency of an ergodicity condition is then obtained by choosing a Lyapunov function whose parameters are determined using the results for the blocks in the M/G/1-type matrix. They also showed that a single-server queue with batch MAP (BMAP) arrivals, a semi-Markovian service process, and exponential retrials is in the class of AQTMCs. Breur *et al.* [10] showed that a BMAP/PH/c queue with exponential retrials is in the class of AQTMCs and gave a sufficient condition for its ergodicity. He *et al.* [30] considered a BMAP/PH/*c* retrial queue with a waiting line and PH retrials. They obtained a sufficient ergodicity condition by choosing a Lyapunov function whose parameters follow from a classical BMAP/PH/*c* queue with no retrials.

Artalejo and Gómez-Corral [4, p. 33] indicated that mathematical analysis of multiclass retrial queues in which all customers can join an orbit is more difficult compared to its single-class counterpart since its joint queue length process is a random walk on the multidimensional integer lattice. Expected waiting time expressions are given for two- and multi-class M/G/1 queues with batch arrivals and exponential retrials in [34] and [25], respectively. Avrachenkov *et al.* [7] considered an M/G/*c* queue with waiting lines and constant retrial policy in which only one customer in the orbit can attempt to get service. Shin and Moon [47] showed that the stationary distribution of a multiclass M/M/*c* queue with exponential retrials converges to that of a classical multiclass M/M/*c* queue with discriminatory random order service policy as retrial rates tend to ∞, and they presented approximation formulae for some performance measures. Kim [31] considered a multiserver multiclass retrial queue in which customers arrive according to a class-dependent Poisson process, service and retrial times follow exponential distributions, and each server can serve a specific class of customers. They obtained a necessary and sufficient condition for positive recurrence by using the fluid limit approach. There are explicit results for a multiclass retrial queue with multiple servers and a few papers take a computational approach. The retrial queue in [16] has two customer classes in which customers either join an infinite waiting line or a finite orbit depending on their class if all servers are busy upon arrival. Therein, this system is modeled as an LDQBD and solved using the matrix-analytic method in [39]. Choi *et al.* [17] obtained several performance measures; they considered a different system than that in [16] in that the orbit is infinite and the waiting line is finite. Gharbi *et al.* [28] modeled a finite source retrial system using generalized stochastic Petri nets and analyzed it with the embedded Markov chain resolution algorithm in [15].

The analysis of a queueing system with PH retrials requires a state representation keeping the number of customers in all retrial phases. This leads to a random walk on the multidimensional integer lattice with more than one infinite dimension as for multiple customer classes. Therefore, there are only a handful of papers that consider PH retrials. In those, either simple arrival and service processes are assumed, or a methodology for numerical analysis is not proposed. M/M/*c* queues with PH retrials appear in [35], [45], and [46]. Kumar *et al.* [35] analyzed the waiting time distribution, and Shin and Moon [46] presented approximation formulae for the distributions of numbers of customers in service and orbit. Shin [45] described the model with two retrial phases as an LDQBD and gave an algorithmic solution. In [23], a stability condition for an M/PH/1 queue with PH retrials was obtained. Besides, a method was proposed for

approximating its stationary distribution and waiting time moments. In [3], a MAP/PH/c queue with PH retrials was numerically analyzed with a finite source assumption. Besides giving a sufficient condition for ergodicity, He *et al.* [30] showed that the condition is also necessary for the ergodicity of a BMAP/PH/c queue with a waiting line and PH retrials. However, a methodology for numerical analysis was not proposed. Finally, Chakravarthy [14] studied a MAP/PH/c queue with PH retrials via simulation due to its complexity. To the best of our knowledge, a multiclass queueing system with MAP arrivals, PH services, and PH retrials has not been analyzed previously.

In this paper a multiclass MAP/PH/c retrial queueing system with acyclic PH retrials is considered. The acyclic PH distribution is a subclass of the PH distribution, but is considered to be as powerful as a general PH distribution, since both are dense in the set of nonnegative distributions (see, for instance, [13]). Therefore, the system under consideration is quite general. A necessary and sufficient condition for its ergodicity is obtained from criteria based on drifts by choosing appropriate Lyapunov functions. The system includes multiple customer classes, and the construction of a useful auxiliary M/G/1-type matrix is not obvious. Hence, it seems that the approach taken to choose the Lyapunov function for the BMAP/PH/c retrial queue in [30] is not applicable when there are multiple customer classes. The infinite state space of the model is truncated with the help of the Lyapunov function chosen in the sufficiency proof, so that a finite state space including at least a given steady-state probability mass is determined [20]. Then the truncated model is described as a multidimensional Markov chain (MC), and a Kronecker representation of its underlying infinitesimal generator matrix is formed in a similar manner to those models in [8] and [19]. Finally, the steady-state distribution of the truncated model is computed iteratively using successive over-relaxation (SOR) [49]. Here, the truncated model is not modeled as an infinite LDQBD and is not solved using the matrix-analytic method of Bright and Taylor [11], although it is possible to do so by choosing an appropriate level definition, as in [8], [19], and [21]. The reason for this choice is that the method does not scale well as the number of dimensions in the multidimensional MC increases. This is due to the increase in the order of the diagonal blocks as the level number increases in multidimensional MCs.

2. Mathematical model

The system under consideration is a multiclass MAP/PH/c queue with acyclic PH retrials, where c is the number of servers. We model this system as a multidimensional MC and give the generator matrix of the underlying Markov process.

Recall that a MAP can be viewed as a counting process or an irreducible MC with some marked transitions (describing arrivals), as in [13]. We will be using the definition of a MAP with the latter interpretation given below.

Definition 1. A MAP with representation $(\boldsymbol{C}, \boldsymbol{D})$ of order m is an irreducible MC with a state space of size m and irreducible generator matrix $(\boldsymbol{C} + \boldsymbol{D})$, where $\boldsymbol{C}$ is a nonsingular matrix with negative diagonal and nonnegative off-diagonal elements, and $\boldsymbol{D} \geq 0$.

A state of the MAP in Definition 1 is said to be a phase. Without loss of generality, we assume that the phases of the MAP are numbered 1 through m. The MAP characterizes a stochastic process, where $\boldsymbol{C}$ includes transitions without an arrival and $\boldsymbol{D}$ includes transitions with one customer arrival. The definition of PH distribution which will be used to model service and retrial times is given next.

Definition 2. Let $m \in \mathbb{Z}_{\geq 0}$, let $\boldsymbol{\beta} \in \mathbb{R}^{1 \times m}_{\geq 0}$, and let $\boldsymbol{T} \in \mathbb{R}^{m \times m}_{\geq 0}$ be a nonsingular matrix with negative diagonal and nonnegative off-diagonal elements. A PH distribution with representation

$(\boldsymbol{\beta}, \boldsymbol{T})$ of order m is the distribution of time until absorption in a finite state space MC with generator matrix

$$\hat{\boldsymbol{T}} = \begin{bmatrix} \boldsymbol{T} & \boldsymbol{T}^0 \\ \boldsymbol{0} & 0 \end{bmatrix}_{(m+1) \times (m+1)},$$

and initial probability vector $(\boldsymbol{\beta}, 1 - \boldsymbol{\beta e}) \in \mathbb{R}_{\geq 0}^{1 \times (m+1)}$, where $\boldsymbol{e}$ represents a column vector of 1s.

Without loss of generality, we assume that the state space of $\hat{\boldsymbol{T}}$ is $\{1, \ldots, m+1\}$, where $m+1$ is the absorbing state and the other states are transient. The transient states in a PH distribution are called phases. We assume that the process does not start in the absorbing state; hence, $\boldsymbol{\beta e} = 1$ holds. Since $\hat{\boldsymbol{T}}$ is the generator matrix of an MC, $\boldsymbol{T}^0 = -\boldsymbol{T e}$ is a nonnegative column vector. Next, the definition of acyclic PH distribution is given.

Definition 3. A PH distribution with representation $(\boldsymbol{\xi}, \boldsymbol{U})$ is said to be acyclic if its states can be ordered in such a way that $\boldsymbol{U}$ is an upper-triangular matrix.

We consider a retrial queueing system with $c \geq 1$ homogeneous servers and $K \geq 1$ customer classes. Customers of class $k \in \{1, \ldots, K\}$ arrive according to a MAP with representation $(\boldsymbol{C}_k, \boldsymbol{D}_k)$ of order m_k^A. Since $(\boldsymbol{C}_k + \boldsymbol{D}_k)$ is irreducible by Definition 1, there exists a nonnegative row vector $\boldsymbol{\theta}_k \in \mathbb{R}_{\geq 0}^{1 \times m_k^A}$ such that $\boldsymbol{\theta}_k (\boldsymbol{C}_k + \boldsymbol{D}_k) = \boldsymbol{0}$ and $\boldsymbol{\theta}_k \boldsymbol{e} = 1$. Furthermore, the average arrival rate is given by $\lambda_k = \boldsymbol{\theta}_k \boldsymbol{D}_k \boldsymbol{e}$. If all servers are busy upon arrival, an arriving customer of class k joins orbit k and retries to capture a server after a random amount of time. A retrial customer is blocked if it attempts to receive service when there are no idle servers. The service time of a class-k customer follows a PH distribution with representation $(\boldsymbol{\beta}_k, \boldsymbol{T}_k)$ of order m_k^S and $\boldsymbol{T}_k^0 = -\boldsymbol{T}_k \boldsymbol{e}$. The retrial time of a class-k customer follows an acyclic PH distribution with representation $(\boldsymbol{\xi}_k, \boldsymbol{U}_k)$ of order m_k^R and $\boldsymbol{U}_k^0 = -\boldsymbol{U}_k \boldsymbol{e}$. Hence, we assume that $\boldsymbol{U}_k$ is upper triangular. For a customer of class k, the average service rate is given by $\mu_k = [-\boldsymbol{\beta}_k (\boldsymbol{T}_k)^{-1} \boldsymbol{e}]^{-1}$ and the average retrial rate is given by $\delta_k = [-\boldsymbol{\xi}_k (\boldsymbol{U}_k)^{-1} \boldsymbol{e}]^{-1}$.

In [14], a single-class MAP/PH/c queue with PH retrials was modeled using a multidimensional MC. The multiclass counterpart can also be modeled similarly. To that end, we let $X_k(t)$, $X_{b_k^R + i_k^R}(t)$, and $X_{b_k^S + i_k^S}(t)$ respectively denote the phase of the arrival process of class-k customers, the number of class-k retrial customers in phase i_k^R, and the number of busy servers serving class-k customers in phase i_k^S for $i_k^R = 1, \ldots, m_k^R$ and $i_k^S = 1, \ldots, m_k^S$, where

$$m^R = \sum_{k=1}^{K} m_k^R, \qquad m^S = \sum_{k=1}^{K} m_k^S, \qquad b_k^R = K + \sum_{k'=1}^{k-1} m_{k'}^R, \qquad b_k^S = K + m^R + \sum_{k'=1}^{k-1} m_{k'}^S.$$

Then the multidimensional MC $X(t) = \{X_1(t), \ldots, X_{K+m^R+m^S}(t) : t \geq 0\}$ has the state space $\mathcal{S} = \mathcal{S}^A \times \mathcal{S}^R \times \mathcal{S}^S$, where

$$\mathcal{S}^A = \bigtimes_{k=1}^{K} \{1, \ldots, m_k^A\}, \qquad \mathcal{S}^R = \mathbb{Z}_{\geq 0}^{m^R}, \qquad \mathcal{S}^S = \{\boldsymbol{y} = (y_1, \ldots, y_{m^S}) \in \mathbb{Z}_{\geq 0}^{m^S} \mid \boldsymbol{y e} \leq c\},$$

and a possible state representation of the model is $\boldsymbol{x} = (x_1, \ldots, x_{K+m^R+m^S}) \in \mathcal{S}$. We let $\boldsymbol{n}(\boldsymbol{x})$ denote the number of busy servers in state $\boldsymbol{x}$; that is, $\boldsymbol{n}(\boldsymbol{x}) = \sum_{k=1}^{K} \sum_{i=1}^{m_k^S} x_{b_k^S + i}$. The set $\mathcal{S}^S$ is defined so that the number of busy servers does not exceed the number of servers, and its size

is given by

$$|\mathcal{S}^S| = \sum_{i=0}^{c} \frac{(i + m^S - 1)!}{i!\,(m^S - 1)!}.$$

Note that another possible approach for modeling this system is to keep the phase of each server in a single but different dimension of the state. However, that approach leads to a larger state space, as discussed in [32] and [43].

Now we give the generator matrix underlying the model in terms of matrices associated with arrival, service, and retrial. The matrix $\boldsymbol{Q}_k^A$ includes transitions associated with the arrival of class-k customers and is given elementwise as

$$\boldsymbol{Q}_k^A(\boldsymbol{x}, \boldsymbol{y}) = \begin{cases} \boldsymbol{C}_k(x_k, i) & \text{if } i \neq x_k \text{ and } \boldsymbol{y} = \boldsymbol{x} + (i - x_k)\boldsymbol{e}_k^\top, \\ \boldsymbol{D}_k(x_k, i)\boldsymbol{\xi}_k(j) & \text{if } \boldsymbol{n}(\boldsymbol{x}) = c \text{ and } \boldsymbol{y} = \boldsymbol{x} + (i - x_k)\boldsymbol{e}_k^\top + \boldsymbol{e}_{b_k^R+j}^\top, \\ \boldsymbol{D}_k(x_k, i)\boldsymbol{\beta}_k(j') & \text{if } \boldsymbol{n}(\boldsymbol{x}) < c \text{ and } \boldsymbol{y} = \boldsymbol{x} + (i - x_k)\boldsymbol{e}_k^\top + \boldsymbol{e}_{b_k^S+j'}^\top, \\ 0 & \text{otherwise,} \end{cases}$$

for $i = 1, \ldots, m_k^A$, $j = 1, \ldots, m_k^R$, $j' = 1, \ldots, m_k^S$, and $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{S}$, where $\boldsymbol{e}_k$ represents the kth principal axis vector. The matrix $\boldsymbol{Q}_k^R$ includes transitions associated with the retrial of class-k customers and is given elementwise as

$$\boldsymbol{Q}_k^R(\boldsymbol{x}, \boldsymbol{y}) = \begin{cases} x_{b_k^R+i}\boldsymbol{U}_k(i, j) & \text{if } i \neq j \text{ and } \boldsymbol{y} = \boldsymbol{x} - \boldsymbol{e}_{b_k^R+i}^\top + \boldsymbol{e}_{b_k^R+j}^\top, \\ x_{b_k^R+i}\boldsymbol{U}_k^0(i)\boldsymbol{\beta}_k(j') & \text{if } \boldsymbol{n}(\boldsymbol{x}) < c \text{ and } \boldsymbol{y} = \boldsymbol{x} - \boldsymbol{e}_{b_k^R+i}^\top + \boldsymbol{e}_{b_k^S+j'}^\top, \\ 0 & \text{otherwise,} \end{cases}$$

for $i, j = 1, \ldots, m_k^R$, $j' = 1, \ldots, m_k^S$, and $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{S}$. Finally, the matrix $\boldsymbol{Q}_k^S$ includes transitions associated with the service of class-k customers and is given elementwise as

$$\boldsymbol{Q}_k^S(\boldsymbol{x}, \boldsymbol{y}) = \begin{cases} llx_{b_k^S+i}\boldsymbol{T}_k(i, j) & \text{if } i \neq j \text{ and } \boldsymbol{y} = \boldsymbol{x} - \boldsymbol{e}_{b_k^S+i}^\top + \boldsymbol{e}_{b_k^S+j}^\top, \\ x_{b_k^S+i}\boldsymbol{T}_k^0(i) & \text{if } \boldsymbol{y} = \boldsymbol{x} - \boldsymbol{e}_{b_k^S+i}^\top, \\ 0 & \text{otherwise,} \end{cases}$$

for $i, j = 1, \ldots, m_k^S$ and $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{S}$. Then the generator matrix underlying $X(t)$ becomes

$$\boldsymbol{Q} = \boldsymbol{Q}^{\text{off}} + \text{diag}(-\boldsymbol{Q}^{\text{off}}\boldsymbol{e}) \quad \text{with} \quad \boldsymbol{Q}^{\text{off}} = \sum_{k=1}^{K}(\boldsymbol{Q}_k^A + \boldsymbol{Q}_k^R + \boldsymbol{Q}_k^S).$$

Example 1. Now consider an example with $K = 2$ and $c = 2$. Let the vectors and matrices describing the arrivals, services, and retrials be given by

$$\boldsymbol{C}_1 = \begin{bmatrix} -0.8 & 0.8 \\ 0 & -0.8 \end{bmatrix}_{2\times 2}, \quad \boldsymbol{D}_1 = \begin{bmatrix} 0 & 0 \\ 0.8 & 0 \end{bmatrix}_{2\times 2}, \quad \boldsymbol{C}_2 = [-0.3]_{1\times 1}, \quad \boldsymbol{D}_2 = [0.3]_{1\times 1},$$

$$\boldsymbol{\xi}_1 = [1, 0]_{1\times 2}, \quad \boldsymbol{U}_1 = \begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix}_{2\times 2}, \quad \boldsymbol{\xi}_2 = [1]_{1\times 1}, \quad \boldsymbol{U}_2 = [-0.5]_{1\times 1},$$

$$\boldsymbol{\beta}_1 = [0.75, 0.25]_{1\times 2}, \quad \boldsymbol{T}_1 = \begin{bmatrix} -1 & 0.25 \\ 0 & -0.25 \end{bmatrix}_{2\times 2}, \quad \boldsymbol{\beta}_2 = [1]_{1\times 1}, \quad \boldsymbol{T}_2 = [-0.5]_{1\times 1}.$$

Hence,

$$\boldsymbol{U}_1^0 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}_{2\times 1}, \qquad \boldsymbol{U}_2^0 = [0.5]_{1\times 1}, \qquad \boldsymbol{T}_1^0 = \begin{bmatrix} 0.75 \\ 0.25 \end{bmatrix}_{2\times 1}, \qquad \boldsymbol{T}_2^0 = [0.5]_{1\times 1}.$$

Here $\lambda_1 = 0.4, \lambda_2 = 0.3, \delta_1 = \delta_2 = 0.5, \mu_1 = 0.4, \mu_2 = 0.5, m_1^A = 2, m_2^A = 1, m_1^R = 2, m_2^R = 1, m_1^S = 2$, and $m_2^S = 1$. This is an eight-dimensional model with $m^R = 3, m^S = 3, b_1^R = 2, b_2^R = 4, b_1^S = 5, b_2^S = 7$, and $\boldsymbol{n}(\boldsymbol{x}) = x_6 + x_7 + x_8$. Therefore, the state space of the MC is given by $\mathcal{S} = \mathcal{S}^A \times \mathcal{S}^R \times \mathcal{S}^S$, where $\mathcal{S}^A = \{1, 2\} \times \{1\}, \mathcal{S}^R = \mathbb{Z}_{\geq 0}^3, \mathcal{S}^S = \{(0, 0, 0), (0, 0, 1), (0, 0, 2), (0, 1, 0), (0, 1, 1), (0, 2, 0), (1, 0, 0), (1, 0, 1), (1, 1, 0), (2, 0, 0)\}$.

3. Ergodicity condition

With the help of Lyapunov functions, we show that the inequality

$$\sum_{k=1}^K \frac{\lambda_k}{\mu_k} < c \tag{1}$$

is a necessary and sufficient ergodicity condition for the multiclass MAP/PH/c queue with acyclic PH retrials considered. The Lyapunov function used to show the sufficiency of this condition will also aid us in finding a truncated state space with a given steady-state probability mass. When the arrival, service, or retrial processes include phases, Lyapunov functions need to be chosen so that the phases are carefully taken into consideration. Otherwise, a necessary and sufficient ergodicity condition may not be found. In order to obtain Lyapunov functions leading to such a condition, we start with Lyapunov functions that work for simple models, and then add terms to these functions for the additional complexities of the model.

The following two lemmas introduce two vectors which include variables that will be used in the additional terms.

Lemma 1. *There exists a unique vector $\boldsymbol{u}_k \in \mathbb{R}^{m_k^A \times 1}$ for a MAP with representation $(\boldsymbol{C}_k, \boldsymbol{D}_k)$ and $\lambda_k = \boldsymbol{\theta}_k \boldsymbol{D}_k \boldsymbol{e}$ such that $(\boldsymbol{C}_k + \boldsymbol{D}_k)\boldsymbol{u}_k = \lambda_k \boldsymbol{e} - \boldsymbol{D}_k \boldsymbol{e}$ and $\boldsymbol{u}_k \boldsymbol{e} = 1$.*

Proof. The reduced linear system of equations $\hat{\boldsymbol{A}}_k \boldsymbol{u}_k = \hat{\boldsymbol{b}}_k$ with $\hat{\boldsymbol{A}}_k = \boldsymbol{M}_k(\boldsymbol{C}_k + \boldsymbol{D}_k) = (\boldsymbol{I}_{m_k^A} - \boldsymbol{e}_{m_k^A} \boldsymbol{e}_{m_k^A}^\top)(\boldsymbol{C}_k + \boldsymbol{D}_k)$ and $\hat{\boldsymbol{b}}_k = \boldsymbol{M}_k(\lambda_k \boldsymbol{e} - \boldsymbol{D}_k \boldsymbol{e}) = (\boldsymbol{I}_{m_k^A} - \boldsymbol{e}_{m_k^A} \boldsymbol{e}_{m_k^A}^\top)(\lambda_k \boldsymbol{e} - \boldsymbol{D}_k \boldsymbol{e})$, where $\boldsymbol{M}_k = \boldsymbol{I}_{m_k^A} + \boldsymbol{e}_{m_k^A}(-\boldsymbol{e}_{m_k^A}^\top + \boldsymbol{\theta}_k)$ and $\boldsymbol{I}_{m_k^A}$ denotes the identity matrix of order m_k^A, has only its last equation altered to make the equation 0 on both sides. Because $(\boldsymbol{C}_k + \boldsymbol{D}_k)$ is an irreducible generator matrix, this implies that $\hat{\boldsymbol{A}}_k$ is of rank $m_k^A - 1$. Thus, $(\boldsymbol{C}_k + \boldsymbol{D}_k)\boldsymbol{u}_k = \lambda_k \boldsymbol{e} - \boldsymbol{D}_k \boldsymbol{e}$ is consistent (see [36, Chapter 2.3]), and there exists a unique vector $\boldsymbol{u}_k$ under the normalization condition $\boldsymbol{u}_k \boldsymbol{e} = 1$. $\square$

Since transition rates describing arrivals in a MAP depend on the phase of the process, elements of $\boldsymbol{u}_k$ will be used in additional terms to obtain an ergodicity condition based on the average arrival rate λ_k instead of phase-dependent arrival rates in $\boldsymbol{D}_k$.

Lemma 2. *There exists a unique vector $\boldsymbol{v}_k \in \mathbb{R}_{\geq 0}^{m_k^S \times 1}$ for a PH service distribution with representation $(\boldsymbol{\beta}_k, \boldsymbol{T}_k)$ and $\mu_k = [-\boldsymbol{\beta}_k(\boldsymbol{T}_k)^{-1}\boldsymbol{e}]^{-1}$ such that $\boldsymbol{v}_k = -\mu_k(\boldsymbol{T}_k)^{-1}\boldsymbol{e}$ and $\boldsymbol{\beta}_k \boldsymbol{v}_k = 1$.*

Proof. The matrix $-\boldsymbol{T}_k$ is a nonsingular M-matrix, $-\boldsymbol{T}_k^{-1} \geq 0$ [36, p. 626], and, therefore, $\boldsymbol{v}_k \geq 0$ exists. This implies that $\boldsymbol{\beta}_k \boldsymbol{v}_k = 1$ since

$$\boldsymbol{\beta}_k \boldsymbol{v}_k = -\mu_k(\boldsymbol{\beta}_k(\boldsymbol{T}_k)^{-1}\boldsymbol{e}) = -\mu_k(-\mu_k)^{-1}. \qquad \square$$

Since transition rates describing PH service depend on the phase of the customer, elements of $\boldsymbol{v}_k$ will be used in additional terms to obtain an ergodicity condition based on the average service rate μ_k instead of phase-dependent service rates in $\boldsymbol{T}_k^0 = -\boldsymbol{T}_k\boldsymbol{e}$.

3.1. Necessary condition

The condition in (1) is necessary for ergodicity if it can be shown that the system is nonergodic when the condition does not hold. Fortunately, the following theorem provides a nonergodicity condition for MCs. Note that the theorem was originally proved for discrete-time MCs (see [6, p. 22] and [27, p. 30]), but here we give a continuous-time version which is obtained by considering the embedded MC.

Theorem 1. *An MC with generator matrix $\boldsymbol{Q}$ is nonergodic if there exists two constants $\tau, \sigma \in \mathbb{R}$ and a Lyapunov function $f : \mathcal{S} \to \mathbb{R}$ such that*

(i) $\sum_{\boldsymbol{y}\in\mathcal{S}} \boldsymbol{P}(\boldsymbol{x}, \boldsymbol{y})|f(\boldsymbol{y}) - f(\boldsymbol{x})| \leq \tau$ *for $\boldsymbol{x} \in \mathcal{S}$, and*

(ii) $\sum_{\boldsymbol{y}\in\mathcal{S}} \boldsymbol{P}(\boldsymbol{x}, \boldsymbol{y})(f(\boldsymbol{y}) - f(\boldsymbol{x})) \geq 0$ *for $\boldsymbol{x} \in \mathcal{R}$,*

where the sets $\mathcal{R} = \{\boldsymbol{x} \in \mathcal{S} \mid f(\boldsymbol{x}) > \sigma\}$ and $(\mathcal{S} \setminus \mathcal{R})$ are nonempty, and the matrix $\boldsymbol{P} \in \mathbb{R}_{\geq 0}^{|\mathcal{S}|\times|\mathcal{S}|}$ is given elementwise as

$$\boldsymbol{P}(\boldsymbol{x}, \boldsymbol{y}) = \begin{cases} \dfrac{\boldsymbol{Q}(\boldsymbol{x}, \boldsymbol{y})}{|\boldsymbol{Q}(\boldsymbol{x}, \boldsymbol{x})|} & \text{if } \boldsymbol{y} \neq \boldsymbol{x}, \\ 0 & \text{otherwise,} \end{cases} \quad \text{for } \boldsymbol{x}, \boldsymbol{y} \in \mathcal{S}.$$

This theorem is used to prove a nonergodicity condition for an M/M/c retrial queue with exponential retrial in [26, p. 98]. Therein, a Lyapunov function $f(\boldsymbol{x})$ linear in the infinite variables (i.e. numbers of different class customers in the orbits) is used. We have also chosen a linear Lyapunov function, but have added constant terms, including elements of vectors $\boldsymbol{u}_k$ and $\boldsymbol{v}_k$ from Lemmas 1 and 2. Thus, we consider

$$f(\boldsymbol{x}) = \sum_{k=1}^{K} \frac{1}{\mu_k}\left(\sum_{i=1}^{m_k^R} x_{b_k^R+i}\right) + \sum_{k=1}^{K} \frac{\boldsymbol{u}_k(x_k)}{\mu_k} + \sum_{k=1}^{K} \frac{1}{\mu_k}\left(\sum_{i=1}^{m_k^S} \boldsymbol{v}_k(i)x_{b_k^S+i}\right).$$

The first term of $f(\boldsymbol{x})$ is the initial function, whereas the other two terms are added to obtain a phase-independent condition. Note that each of the three terms is in the form of a summation of K other terms, each corresponding to a different customer class.

Since $\boldsymbol{Q}$ is a generator matrix, $\boldsymbol{P}(\boldsymbol{x}, \boldsymbol{y}) \leq 1$ for $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{S}$. The value of $|f(\boldsymbol{y}) - f(\boldsymbol{x})|$ is finite if $\boldsymbol{P}(\boldsymbol{x}, \boldsymbol{y}) > 0$ for $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{S}$. Besides, each row of the matrix $\boldsymbol{P}$ includes a finite number of nonzero elements due to the form of the particular model considered. Therefore, there exists some constant τ such that Theorem 1(i) is satisfied.

After some algebraic operations, we obtain

$$\frac{1}{|\boldsymbol{Q}(\boldsymbol{x}, \boldsymbol{x})|} \sum_{\boldsymbol{y}\in\mathcal{S}} \boldsymbol{Q}(\boldsymbol{x}, \boldsymbol{y})(f(\boldsymbol{y}) - f(\boldsymbol{x})) = \frac{1}{|\boldsymbol{Q}(\boldsymbol{x}, \boldsymbol{x})|}\left(\sum_{k=1}^{K} \frac{\lambda_k}{\mu_k} - \boldsymbol{n}(\boldsymbol{x})\right).$$

The right-hand side of this equation is nonnegative if $\sum_{k=1}^{K}(\lambda_k/\mu_k) \geq c$ since $|\boldsymbol{Q}(\boldsymbol{x}, \boldsymbol{x})| > 0$ and $\boldsymbol{n}(\boldsymbol{x}) \leq c$ hold for $\boldsymbol{x} \in \mathcal{S}$. Besides, the sets $\mathcal{R}$ and $(\mathcal{S} \setminus \mathcal{R})$ are nonempty when $\sigma = \max_{\boldsymbol{x}\in\mathcal{S}}(\sum_{k=1}^{K} \boldsymbol{u}_k(x_k)/\mu_k)$. Therefore, Theorem 1(ii) also holds.

3.2. Sufficient condition

In coming up with a Lyapunov function to show the sufficiency of the condition in (1) for ergodicity, we benefit from the next theorem in [48]. Note that the Lyapunov function is

assumed to be nonnegative in [48], but the theorem is also valid if the Lyapunov function is bounded from below (see [26, p. 97]). This theorem will also be used to bound the steady-state probability mass associated with the truncated state space from below, as proposed in [20].

Theorem 2. *The MC is ergodic if and only if there exists a Lyapunov function* $g: \mathcal{S} \to \mathbb{R}$ *that is bounded from below and a finite set* $\mathcal{C} \subset \mathcal{S}$ *satisfying the three conditions*

   (i) $\{x \in \mathcal{S} \mid g(x) \leq \tau\}$ *is finite for all* $\tau < \infty$,

   (ii) $d(x) \leq -\gamma$ *for all* $x \in \mathcal{S} \setminus \mathcal{C}$ *and some* $\gamma > 0$, *and*

   (iii) $d(x) < \infty$ *for all* $x \in \mathcal{S}$,

where $d(x) = \sum_{y \in \mathcal{S}} Q(x, y)(g(y) - g(x))$ *is called the drift in state* $x \in \mathcal{S}$.

We will be choosing a quadratic Lyapunov function $g(x)$ that is similar to those discussed in [8] to obtain a drift function $d(x)$ whose infinite variables have negative coefficients for $x \in \mathcal{S}$. If customers in the orbits attempt to receive service in all PH retrial phases, adding terms with variables each corresponding to the number of customers in a service phase leads to the condition in (1). However, if the PH retrial process of a class-k customer includes a phase, say i, in which no attempt is made to receive service (i.e. if $U_k^0(i) = 0$, where $U_k^0 = -U_k e$), then the coefficient of the corresponding variable, $x_{b_k^R + i}$, is positive in $d(x)$ at states with no busy servers. Therefore, the value of $g(x)$ should also depend on the number of customers in retrial phases. This can be managed by adding carefully chosen terms so that $d(x)$ at states with no busy servers become negative for a sufficiently large number of customers in the orbits. Based on this explanation, the following lemma gives the vector that includes variables to be used in the term added due to PH retrials of class-k customers.

Lemma 3. *For an acyclic PH retrial distribution with representation* (ξ_k, U_k), *let* $\hat{U}_k = U_k + \mathrm{diag}(U_k^0)$ *and* $\eta_k \in \mathbb{R}^{m_k^R \times 1}$ *be given elementwise as*

$$\eta_k(i) = \begin{cases} \dfrac{-c}{\mu_k} & \text{if } i \in \mathcal{I}_k, \\ 0 & \text{otherwise}, \end{cases} \quad \text{for } i = 1, \ldots, m_k^R,$$

where $\mathcal{I}_k = \{i \in \{1, \ldots, m_k^R\} \mid U_k^0(i) = 0\}$. *Then there exists* $w_k \in \mathbb{R}^{m_k^R \times 1}$ *such that* $\hat{U}_k w_k = \eta_k$.

Proof. Note that all elements in row i of $\hat{U}_k$ are 0 if and only if $\hat{U}_k(i, i) = 0$. Then the result follows from the row echelon form of the upper-triangular matrix $\hat{U}_k$ obtained by interchanging all zero rows with nonzero rows below them (see [36, Chapter 2.1]), that its rank is equal to its number of nonzero rows, and $\hat{U}_k w_k = \eta_k$ is consistent and has infinitely many solutions (see [36, Chapter 2.3]). Note that negative elements of η_k could take a smaller value than $-c/\mu_k$ that may lead to a smaller truncated state space. We choose this value in order to bound coefficients of infinite variables from above by $(\sum_{k=1}^{K} (\lambda_k/\mu_k) - c)$ in $d(x)$ for all states. $\square$

There exist infinitely many solutions to $\hat{U}_k w_k = \eta_k$. Hence, additional constraints need to be imposed. Here, we choose to set $w_k(i)$ to 1 if row i of $\hat{U}_k$ is 0. The elements of vector η_k contribute to coefficients of infinite variables corresponding to class-k retrial customers in $d(x)$; hence, its elements need to be nonpositive with at least one negative element. If PH retrials are allowed to be nonacyclic, it is a possibility that $\hat{U}_k w_k = \eta_k$ is inconsistent since $\hat{U}_k$ is a singular M-matrix. Hence, the acyclicity assumption is necessary due to the form of the chosen $g(x)$. We conjecture that the condition in (1) is also necessary and sufficient when PH retrials are nonacyclic; however, we have not managed to provide a suitable $g(x)$ for such a system.

Having defined all necessary variables, as the Lyapunov function we consider

$$g(\boldsymbol{x}) = \sum_{k'=1}^{K} \sum_{l'=1}^{K} \frac{1}{2\mu_{k'}\mu_{l'}} \left(\sum_{i=1}^{m_{k'}^R} x_{b_{k'}^R+i} \right) \left(\sum_{i=1}^{m_{l'}^R} x_{b_{l'}^R+i} \right) + \sum_{k'=1}^{K} \left(\sum_{l'=1}^{K} \frac{\boldsymbol{u}_{l'}(x_{l'})}{\mu_{k'}\mu_{l'}} \right) \left(\sum_{i=1}^{m_{k'}^R} x_{b_{k'}^R+i} \right)$$

$$+ \sum_{k'=1}^{K} \left(\sum_{l'=1}^{K} \frac{1}{\mu_{k'}\mu_{l'}} \left(\sum_{j=1}^{m_{l'}^S} \boldsymbol{v}_{l'}(j) x_{b_{l'}^S+j} \right) \right) \left(\sum_{i=1}^{m_{k'}^R} x_{b_{k'}^R+i} \right)$$

$$+ \sum_{k'=1}^{K} \sum_{i=1}^{m_{k'}^R} \boldsymbol{w}_{k'}(i) x_{b_{k'}^R+i} + \sum_{k'=1}^{K} \alpha_{k'}(\boldsymbol{x}) \left(\sum_{i=1}^{m_{k'}^S} x_{b_{k'}^S+i} \right),$$

where

$$a_k(i, \boldsymbol{x}) = \boldsymbol{w}_k(i) + \frac{1}{2\mu_k^2} + \sum_{k'=1}^{K} \frac{\boldsymbol{u}_{k'}(x_{k'})}{\mu_k \mu_{k'}} + \sum_{k'=1}^{K} \frac{1}{\mu_k \mu_{k'}} \left(\sum_{j'=1}^{m_{k'}^S} \boldsymbol{v}_{k'}(j') x_{b_{k'}^S+j'} \right) \quad \text{for } i \notin \mathcal{I}_k$$

and $\alpha_k(\boldsymbol{x}) = \min_{i \notin \mathcal{I}_k} (a_k(i, \boldsymbol{x}) - c/(\boldsymbol{U}_k^0(i)\mu_k))$ for $\boldsymbol{x} \in \mathcal{S}$ and $k = 1, \ldots, K$. The first term of $g(\boldsymbol{x})$ is the initial quadratic Lyapunov function. The second and third terms are added to obtain a phase-independent condition, and the last two terms are added due to PH retrials. Note that $\alpha_k(\boldsymbol{x})$ is well defined since $\mathcal{I}_k \neq \{1, \ldots, m_k^R\}$ for $k = 1, \ldots, K$ due to Definitions 2 and 3. Then $g(\boldsymbol{x})$ is a quadratic polynomial in which the coefficients of all infinite variables are finite and the coefficient of each quadratic term $(x_{b_k^R+i} x_{b_{k'}^R+i'})$ is positive for $i = 1, \ldots, m_k^R$, $i' = 1, \ldots, m_{k'}^R$, and $k, k' = 1, \ldots, K$. Hence, the function g is bounded from below and condition (i) holds.

For $\boldsymbol{n}(\boldsymbol{x}) = c$, the drift is given by

$$d(\boldsymbol{x}) = \sum_{k=1}^{K} \frac{1}{\mu_k} \left(\sum_{k'=1}^{K} \frac{\lambda_{k'}}{\mu_{k'}} - c \right) \left(\sum_{i=1}^{m_k^R} x_{b_k^R+i} \right) - \sum_{k=1}^{K} \frac{c}{\mu_k} \left(\sum_{i \in \mathcal{I}_k} x_{b_k^R+i} \right)$$

$$+ \sum_{k=1}^{K} \sum_{i=1}^{m_k^A} \boldsymbol{D}_k(x_k, i) \left[\frac{\boldsymbol{u}_k(i)}{\mu_k^2} + \sum_{k'=1, k' \neq k}^{K} \frac{\boldsymbol{u}_{k'}(x_{k'})}{\mu_k \mu_{k'}} + \sum_{k'=1}^{K} \frac{1}{\mu_k \mu_{k'}} \left(\sum_{j'=1}^{m_{k'}^S} \boldsymbol{v}_{k'}(j') x_{b_{k'}^S+j'} \right) \right]$$

$$+ \sum_{k=1}^{K} \sum_{i=1}^{m_k^A} \boldsymbol{D}_k(x_k, i) \left[\left(\sum_{j=1}^{m_k^R} \boldsymbol{\xi}_k(j) \boldsymbol{w}_k(j) \right) + \frac{1}{2\mu_k^2} \right] - \sum_{k=1}^{K} \alpha_k(\boldsymbol{x}) \left(\sum_{i=1}^{m_k^S} x_{b_k^S+i} \boldsymbol{T}_k^0(i) \right);$$

on the other hand, for $\boldsymbol{n}(\boldsymbol{x}) < c$, the drift is given by

$$d(\boldsymbol{x}) = \sum_{k=1}^{K} \frac{1}{\mu_k} \left(\sum_{k'=1}^{K} \frac{\lambda_{k'}}{\mu_{k'}} - \sum_{k'=1}^{K} \sum_{i=1}^{m_{k'}^S} x_{b_{k'}^S+i} \right) \left(\sum_{i=1}^{m_k^R} x_{b_k^R+i} \right) - \sum_{k=1}^{K} \frac{c}{\mu_k} \left(\sum_{i \in \mathcal{I}_k} x_{b_k^R+i} \right)$$

$$+ \sum_{k=1}^{K} \sum_{i \notin \mathcal{I}_k} x_{b_k^R+i} \boldsymbol{U}_k^0(i) (-a_k(i, \boldsymbol{x}) + \alpha_k(\boldsymbol{x}))$$

$$+ \sum_{k=1}^{K} \alpha_k(\boldsymbol{x}) \left(\sum_{i=1}^{m_k^A} \boldsymbol{D}_k(x_k, i) - \sum_{i=1}^{m_k^S} x_{b_k^S+i} \boldsymbol{T}_k^0(i) \right) \quad \text{for } \boldsymbol{x} \in \mathcal{S}.$$

When $\sum_{k=1}^{K}(\lambda_k/\mu_k) < c$ holds, coefficients of all infinite variables are negative; hence, Theorem 2(iii) holds. Besides, $\mathcal{C} = \{x \in \mathcal{S} \mid d(x) > -\gamma\}$ is finite for any arbitrarily chosen $\gamma > 0$. Therefore, Theorem 2(ii) also holds and the model is ergodic if $\sum_{k=1}^{K}(\lambda_k/\mu_k) < c$. As discussed in [20], for some given positive $\varepsilon < 1$, we have $\sum_{x \in \mathcal{C}} \pi(x) \geq 1 - \varepsilon$, where $\gamma = \sup_{x \in \mathcal{S}} d(x)(1/\varepsilon - 1)$ and π is the steady-state solution.

Example 2. (*Example 1 continued.*) In this example,

$$u_1 = \begin{bmatrix} 0.25 \\ 0.75 \end{bmatrix}_{2 \times 1}, \qquad v_1 = \begin{bmatrix} 0.8 \\ 1.6 \end{bmatrix}_{2 \times 1}, \qquad w_1 = \begin{bmatrix} 6 \\ 1 \end{bmatrix}_{2 \times 1}, \qquad u_2 = v_2 = w_2 = [1]_{1 \times 1}.$$

Then $\mathfrak{l}_1 = \{1\}$, $\mathfrak{l}_2 = \varnothing$,

$$a_1(2, x) = 4.125 + 6.25u_1(x_1) + 5u_2(x_2) + 5x_6 + 10x_7 + 5x_8, \qquad \alpha_1(x) = a_1(2, x) - 5,$$
$$a_2(1, x) = 3 + 5u_1(x_1) + 4u_2(x_2) + 4x_6 + 8x_7 + 4x_8, \qquad \alpha_2(x) = a_2(1, x) - 8.$$

The Lyapunov function is given by

$$\begin{aligned}
g(x) = {}& 3.125(x_3 + x_4)^2 + 5(x_3 + x_4)x_5 + 2x_5^2 + 6x_3 + x_4 + x_5 \\
& + (6.25u_1(x_1) + 5u_2(x_2) + 5x_6 + 10x_7 + 5x_8)(x_3 + x_4) \\
& + (5u_1(x_1) + 4u_2(x_2) + 4x_6 + 8x_7 + 4x_8)x_5 + \alpha_1(x)(x_6 + x_7) + \alpha_2(x)x_8.
\end{aligned}$$

For $n(x) = 2$, the drift is given by

$$\begin{aligned}
d(x) = {}& -6x_3 - x_4 - 0.8x_5 + \mathbf{1}_{\{x_1=2\}}(4u_2(x_2) + 4x_6 + 8x_7 + 4x_8 + 8.55) \\
& + \mathbf{1}_{\{x_2=1\}}(1.5u_1(x_1) + 1.2x_6 + 2.4x_7 + 1.2x_8 + 2.1) \\
& + \alpha_1(x)(-0.75x_6 - 0.25x_7) + \alpha_2(x)(-0.5x_8),
\end{aligned}$$

where $\mathbf{1}_{\{\cdot\}}$ denotes the indicator function, and, for $n(x) < 2$, the drift is given by

$$\begin{aligned}
d(x) = {}& 2.5(-0.4 - x_6 - x_7 - x_8)(x_3 + x_4) + 2(-0.4 - x_6 - x_7 - x_8)x_5 \\
& + \alpha_1(x)(\mathbf{1}_{\{x_1=2\}} 0.8 - 0.75x_6 - 0.25x_7) + \alpha_2(x)(\mathbf{1}_{\{x_2=1\}} 0.3 - 0.5x_8).
\end{aligned}$$

4. Numerical results

Once a Kronecker representation of the truncated generator matrix $\bar{Q}$ is obtained as in [8] and [19], we can employ a memory-efficient Kronecker-based iterative solver in which the nonzeros of $\bar{Q}$ are not stored and no factorization takes place during the course of computing $\bar{\pi}$. In practice, this approach always saves a significant amount of memory. Another approach is to employ an LDQBD model, where level l has the state space

$$\mathcal{S}(l) = \{x = (x_1, \ldots, x_{K+m^R+m^S}) \in \mathcal{S} \mid l = \max\{x_{K+1}, \ldots, x_{K+m^R}\}\},$$

and then use the matrix-analytic method proposed in [11]. In this method, conditional expected sojourn time matrices between two given truncation levels need to be computed and stored. For our system, the conditional expected sojourn time matrix at level l includes about $\bar{M}^2 l^{2(m^R-1)}$ nonzeros, where $\bar{M} = |\mathcal{S}^S|(\prod_{k=1}^{K} m_k^A)m^R$. Hence, the memory allocated to store these matrices becomes extremely large. Phung-Duc *et al.* [42] proposed an algorithm to compute the conditional expected sojourn time matrices of an LDQBD with a smaller memory usage (see [42, Algorithm 1]). However, the stationary distribution was computed using the algorithm

TABLE 1: Numerical results.

Model	ρ	$|\bar{\mathcal{S}}|$	$E_{1,1}$	$E_{1,2}$	$E_{2,1}$	$\mathbb{P}_{block}$	$||\bar{\boldsymbol{\pi}}\,\bar{\boldsymbol{Q}}||_1$	$||\tilde{\boldsymbol{\pi}}\,\boldsymbol{Q}||_1$
ERL$_1$	0.8	29 932 260	0.2584	2.0165	2.8985	0.6784	$9e-14$	$9e-14$
ERL$_2$	0.6	811 800	0.1097	0.4422	0.6756	0.4164	$2e-14$	$2e-12$
ERL$_3$	0.4	12 800	0.0317	0.0860	0.1511	0.2048	$6e-15$	$7e-6$
EXP$_1$	0.8	270 400	3.1663		2.5486	0.6755	$8e-14$	$8e-14$
EXP$_2$	0.6	25 600	0.7154		0.6302	0.4142	$2e-14$	$2e-11$
EXP$_3$	0.4	1600	0.1464		0.1461	0.2038	$4e-15$	$6e-6$

proposed in [11] (see [42, Algorithm 3]), which requires computing and storing the conditional expected sojourn time matrices (see [42, Remark 3.5]). Hence, memory usage is expected to be large in this algorithm. On the other hand, the Kronecker-based iterative solver is expected to solve the truncated system with comparable accuracy if the value of the stopping tolerance it uses is sufficiently small.

We implemented a Kronecker solver [18] built on the Nsolve package [12] of the APNN toolbox [9]. The solver obtains the truncated state space of the model, generates the Kronecker structured matrix of the truncated model, and computes its steady-state solution by using the SOR method of Nsolve. All computations were carried out on an Intel® Core™2 Duo 2.4 GHz processor with 4 GB of main memory. Iterations are stopped when the infinity norm of the residual vector of the truncated model (i.e. $||\bar{\boldsymbol{\pi}}\,\bar{\boldsymbol{Q}}||_\infty$) becomes smaller than 10^{-15}, and the relaxation parameter of SOR is chosen as 0.9.

We considered six different models with $\varepsilon = 0.2$. In Table 1 we report the results of numerical experiments with these models. The first column gives the name of the model. ERL$_1$ is the model introduced in Example 1. ERL$_2$ and ERL$_3$ are obtained by multiplying the matrices describing the arrival processes of ERL$_1$ by 0.75 and 0.5, respectively. EXP$_1$ differs from the first model in that the retrial time of customer class 1 is exponentially distributed with average retrial rate 0.5. EXP$_2$ and EXP$_3$ are obtained by multiplying the matrices describing the arrival processes of EXP$_1$ by 0.75 and 0.5, respectively. The second column gives the traffic intensity $\rho = (\sum_{k=1}^{K} \lambda_k/\mu_k)c^{-1}$. The third column gives the number of states in the truncated state space. Columns $E_{1,1}$, $E_{1,2}$, and $E_{2,1}$ give the average number of class-1 customers in retrial phases 1 and 2, and the average number of class-2 customers in retrial phase 1, respectively. Retrial process of customer class 1 has a single phase in EXP$_1$, EXP$_2$, and EXP$_3$; hence, $E_{1,2}$ is undefined for those models. Column $\mathbb{P}_{block}$ provides the probability that an arriving customer finds all servers busy. Column $||\bar{\boldsymbol{\pi}}\,\bar{\boldsymbol{Q}}||_1$ provides the 1-norm of the residual vector of the truncated model which is an indicator of the accuracy of the solution to the truncated model. Column $||\tilde{\boldsymbol{\pi}}\,\boldsymbol{Q}||_1$ provides the 1-norm of the residual vector of the infinite model and is obtained from

$$||\tilde{\boldsymbol{\pi}}\,\boldsymbol{Q}||_1 = \sum_{\boldsymbol{x}\in\bar{\mathcal{S}}} \left| r(\boldsymbol{x}) - \sum_{\boldsymbol{y}\notin\bar{\mathcal{S}}} \bar{\pi}(\boldsymbol{x})\boldsymbol{Q}(\boldsymbol{x},\boldsymbol{y}) \right| + \sum_{\boldsymbol{x}\in\bar{\mathcal{S}}}\sum_{\boldsymbol{y}\notin\bar{\mathcal{S}}} \bar{\pi}(\boldsymbol{x})\boldsymbol{Q}(\boldsymbol{x},\boldsymbol{y}) \quad \text{with} \quad \boldsymbol{r} = \bar{\boldsymbol{\pi}}\,\bar{\boldsymbol{Q}}.$$

This value is an indicator of the accuracy of the solution to the infinite model.

The relative difference between the average numbers of retrial customers in ERL$_i$ and EXP$_i$ becomes relatively large as the traffic intensity increases. This difference ranges from 0.03 (customer class 2 in ERL$_3$ and EXP$_3$) to 0.39 (customer class 1 in ERL$_1$ and EXP$_1$). The relative difference between the blocking probabilities is around 0.005. In ERL$_1$ and EXP$_1$,

residuals of truncated and infinite models are about the same; hence, the truncation error is not larger than the numerical error. In other models, the truncation error is larger than the numerical error. Therefore, as the traffic intensity increases, choosing a smaller ε value does not introduce additional inaccuracy to the computed solution.

Acknowledgement

References

[1] ARTALEJO, J. R. (1999). Accessible bibliography on retrial queues. *Math. Comput. Modelling* **30,** 1–6.

[2] ARTALEJO, J. R. (2010). Accessible bibliography on retrial queues: progress in 2000–2009. *Math. Comput. Modelling* **51,** 1071–1081.

[3] ARTALEJO, J. R. AND GÓMEZ-CORRAL, A. (2007). Modelling communication systems with phase type service and retrial times. *IEEE Commun. Lett.* **11,** 955–957.

[4] ARTALEJO, J. R. AND GÓMEZ-CORRAL, A. (2008). *Retrial Queueing Systems: A Computational Approach*. Springer, Berlin.

[5] ARTALEJO, J. R. AND PHUNG-DUC, T. (2012). Markovian retrial queues with two way communication. *J. Ind. Manag. Optimization* **8,** 781–806.

[6] ASMUSSEN, S. (2003). *Applied Probability and Queues* (Appl. Math. (New York) **51**). Springer, New York.

[7] AVRACHENKOV, K., MOROZOV, E. AND STEYAERT, B. (2016). Sufficient stability conditions for multi-class constant retrial rate systems. *Queueing Systems* **82,** 149–171.

[8] BAUMANN, H., DAYAR, T., ORHAN, M. C. AND SANDMANN, W. (2013). On the numerical solution of Kronecker-based infinite level-dependent QBD processes. *Performance Evaluation* **70,** 663–681.

[9] BAUSE, F., BUCHHOLZ, P. AND KEMPER, P. (1998). A toolbox for functional and quantitative analysis of DEDS. In *Computer Preformance Evaluation* (Lecture Notes Comput. Sci. **1469**), Springer, Berlin, pp. 356–359.

[10] BREUER, L., DUDIN, A. AND KLIMENOK, V. (2002). A retrial BMAP/PH/N system. *Queueing Systems* **40,** 433–457.

[11] BRIGHT, L. AND TAYLOR, P. G. (1995). Calculating the equilibrium distribution in level dependent quasi-birth-and-death processes. *Commun. Statist. Stoch. Models* **11,** 497–525.

[12] BUCHHOLZ, P. The Nsolve package. Available at http://ls4-www.cs.tu-dortmund.de/download/buchholz/struct-matrix-market.html.

[13] BUCHHOLZ, P., KRIEGE, J. AND FELKO, I. (2014). *Input Modeling with Phase-Type Distributions and Markov Models: Theory and Applications*. Springer, Cham.

[14] CHAKRAVARTHY, S. R. (2013). Analysis of MAP/PH/c retrial queue with phase type retrials – simulation approach. In *Modern Probabilistic Methods for Analysis of Telecommunication Networks* (Commun. Comput. Inf. Sci. **356**), Springer, Berlin, pp. 37–49.

[15] CHIOLA, G., DUTHEILLET, C., FRANCESCHINIS, G. AND HADDAD, S. (1993). Stochastic well-formed colored nets and symmetric modeling applications. *IEEE Trans. Comput.* **42,** 1343–1360.

[16] CHOI, B. D. AND CHANG, Y. (1999). MAP_1, MAP_2/M/c retrial queue with the retrial group of finite capacity and geometric loss. *Math. Comput. Modelling* **30,** 99–113.

[17] CHOI, B. D., CHANG, Y. AND KIM, B. (1999). MAP_1, MAP_2/M/c retrial queue with guard channels and its application to cellular networks. *Top* **7,** 231–248.

[18] DAYAR, T. AND ORHAN, M. C. RetrialQueueSolver. Available at http://www.cs.bilkent.edu.tr/~tugrul/software.html.

[19] DAYAR, T. AND ORHAN, M. C. (2012). Kronecker-based infinite level-dependent QBD processes. *J. Appl. Prob.* **49,** 1166–1187.

[20] DAYAR, T., HERMANNS, H., SPIELER, D. AND WOLF, V. (2011). Bounding the equilibrium distribution of Markov population models. *Numer. Linear Algebra Appl.* **18,** 931–946.

[21] DAYAR, T., SANDMANN, W., SPIELER, D. AND WOLF, V. (2011). Infinite level-dependent QBD processes and matrix-analytic solutions for stochastic chemical kinetics. *Adv. Appl. Prob.* **43,** 1005–1026.

[22] DIAMOND, J. E. AND ALFA, A. S. (1998). The MAP/PH/1 retrial queue. *Commun. Statist. Stoch. Models* **14,** 1151–1177.

[23] DIAMOND, J. E. AND ALFA, A. S. (1999). Approximation method for M/PH/1 retrial queues with phase type inter-retrial times. *Europ. J. Operat. Res.* **113,** 620–631.

[24] DUDIN, A. AND KLIMENOK, V. (2000). A retrial BMAP/SM/1 system with linear repeated requests. *Queueing Systems Theory Appl.* **34,** 47–66.

[25] FALIN, G. I. (1988). On a multiclass batch arrival retrial queue. *Adv. Appl. Prob.* **20,** 483–487.

[26] FALIN, G. I. AND TEMPLETON, J. G. C. (1997). *Retrial Queues*. Chapman & Hall, London.

[27] FAYOLLE, G., MALYSHEV, V. A. AND MENSHIKOV, M. V. (1995). *Topics in the Constructive Theory of Countable Markov Chains*. Cambridge University Press.

[28] GHARBI, N., DUTHEILLET, C. AND IOUALALEN, M. (2009). Colored stochastic Petri nets for modelling and analysis of multiclass retrial systems. *Math. Comput. Modelling* **49,** 1436–1448.

[29] GÓMEZ-CORRAL, A. (2006). A bibliographical guide to the analysis of retrial queues through matrix analytic techniques. *Ann. Operat. Res.* **141,** 163–191.

[30] HE, Q.-M., LI, H. AND ZHAO, Y. Q. (2000). Ergodicity of the BMAP/PH/s/$s+K$ retrial queue with PH-retrial times. *Queueing Systems Theory Appl.* **35,** 323–347.

[31] KIM, B. (2011). Stability of a retrial queueing network with different classes of customers and restricted resource pooling. *J. Ind. Manag. Optimization* **7,** 753–765.

[32] KIM, C. S., MUSHKO, V. AND DUDIN, A. N. (2012). Computation of the steady state distribution for multi-server retrial queues with phase type service process. *Ann. Operat. Res.* **201,** 307–323.

[33] KIM, J. AND KIM, B. (2016). A survey of retrial queueing systems. To appear in *Ann. Operat. Res.*

[34] KULKARNI, V. G. (1986). Expected waiting times in a multiclass batch arrival retrial queue. *J. Appl. Prob.* **23,** 144–154.

[35] KUMAR, M. S., SOHRABY, K. AND KISEON, K. (2013). Delay analysis of orderly reattempts in retrial queueing system with phase type retrial time. *IEEE Commun. Lett.* **17,** 822–825.

[36] MEYER, C. (2000). *Matrix Analysis and Applied Linear Algebra*. SIAM, Philadelphia, PA.

[37] NEUTS, M. F. (1979). A versatile Markovian point process. *J. Appl. Prob.* **16,** 764–779.

[38] NEUTS, M. F. (1981). *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*. Johns Hopkins University Press, Baltimore, MD.

[39] NEUTS, M. F. AND RAO, B. M. (1990). Numerical investigation of a multiserver retrial model. *Queueing Systems* **7,** 169–189.

[40] PHUNG-DUC, T. AND KAWANISHI, K. (2011). Multiserver retrial queues with after-call work. *Numer. Algebra Control Optimization* **1,** 639–656.

[41] PHUNG-DUC, T. AND KAWANISHI, K. (2014). Performance analysis of call centers with abandonment, retrial and after-call work. *Performance Evaluation* **80,** 43–62.

[42] PHUNG-DUC, T., MASUYAMA, H., KASAHARA, S. AND TAKAHASHI, Y. (2010). A simple algorithm for the rate matrices of level-dependent QBD processes. In *Proceedings of the 5th International Conference on Queueing Theory and Network Applications*, ACM, New York, pp. 46–52.

[43] RAMASWAMI, V. AND LUCANTONI, D. M. (1985). Algorithms for the multiserver queue with phase-type service. *Commun. Statist. Stoch. Models* **1,** 393–417.

[44] SAKURAI, H. AND PHUNG-DUC, T. (2015). Two-way communication retrial queues with multiple types of outgoing calls. *TOP* **23,** 466–492.

[45] SHIN, Y. W. (2011). Algorithmic solution for M/M/c retrial queue with PH$_2$-retrial times. *J. Appl. Math. Inform.* **29,** 803–811.

[46] SHIN, Y. W. AND MOON, D. H. (2011). Approximation of M/M/c retrial queue with PH-retrial times. *Europ. J. Operat. Res.* **213,** 205–209.

[47] SHIN, Y. W. AND MOON, D. H. (2014). M/M/c retrial queue with multiclass of customers. *Methodol. Comput. Appl. Prob.* **16,** 931–949.

[48] TWEEDIE, R. L. (1975). Sufficient conditions for regularity, recurrence and ergodicity of Markov processes. *Math. Proc. Camb. Phil. Soc.* **78,** 125–136.

[49] UYSAL, E. AND DAYAR, T. (1998). Iterative methods based on splittings for stochastic automata networks. *Europ. J. Operat. Res.* **110,** 166–186.