Stability of Multiclass Multiserver Models with Automata-type Phase Transitions

Alexander Rumyantsev[1,2] iD

[1] Institute of Applied Mathematical Research, Karelian Research Centre of RAS
[2] Petrozavodsk State University
Tel.: +7-8142-76-63-12
`ar0@krc.karelia.ru`

Abstract. In this paper we obtain sufficient conditions for the multiclass multiserver model with automata-type transitions to have a product-form type explicit form of the stability criterion. We use the method to obtain stability conditions of an interesting simultaneous service multiserver model describing a supercomputer, both in case of phase-dependent arrival rate and class-dependent service rate.

Keywords: Stochastic Stability · Multiclass Multiserver Model · Product Form · Supercomputer Model · Simultaneous Service Multiserver Model · Matrix-Analytic Method

1 Introduction

Matrix-analytic method (MAM) is an established powerful technique for the analysis of stochastic models which uses the special structure of a Markov chain to obtain its steady-state or transient distribution [15,11]. The basic step of a stochastic model analysis is to obtain the stability conditions of the model. The stability conditions within the MAM framework [15] are usually a version [20] of somewhat intuitive negative drift criterion [9,13]. However, the structure of a model may have additional properties that allow one to express the stability condition in an explicit form in terms of the model parameters, such as the input, service rates etc. Such properties are in the focus of this paper.

Consideration is given to multiclass multiserver queueing systems with the specific feature: automata-type transitions. Their structure allows to make a certain Markov chain embedding and, under certain assumptions, obtain the stability criterion involving only a product form (see (11) below). Surprisingly, it turns out that such a product form may be obtained even in the cases when some transition rates are not known (see (4) and discussion below).

This paper is inspired by the results on stability of a simultaneous service multiserver model describing a supercomputer, which were obtained in [17]. The model itself is of a separate interest. It is a multiserver model with two-dimensional customers, namely, each such a customer requires a random number of servers for a (same at each occupied server) random time, while servers dedicated to a customer are seized and released simultaneously. Such a model, to

the best of our knowledge, for the first time was treated by MAM in [12], while in [6] the stability condition of a two-server system was expressed in explicit form, whereas in [8] the condition was expressed for a two-server system with Markovian arrival process (MAP). In [17] the stability criterion in explicit form was obtained for arbitrary number of servers, Poisson arrivals and exponentially distributed class-independent service times, while in [1] it was formulated in a general case for a regenerative input flow (whereas the explicit form was given for exponential service times, and for a particular case of phase-type distribution). In the present paper, as an example, we extend the results given in [17] to a phase-dependent arrival flow and to a system with class-dependent service times.

Summarizing, the contribution of this paper is two-fold. Firstly we obtain such sufficient conditions for the multiclass multiserver model that the stability criterion can be obtained explicitly. Second, we extend the stability criterion of a simultaneous service multiserver model obtained in [17] towards phase-dependent arrival rate system and/or system with class-dependent service rate.

The structure of the paper is as follows. In Section 2 we review the stability criterion of a MAM known as the Neuts ergodicity condition. In Section 3 we give the description of a multiclass multiserver model with automata-type transitions and formulate the conditions for the stability condition to be given in product form. In Section 4 we demonstrate the application of the sufficient condition obtained in Section 3 to various models including supercomputer model with phase-dependent arrival rate and/or class-dependent service rate, and discuss the limitations of such conditions. The discussion of potential extensions of the work is given in Section 5.

2 Stability Criterion of a QBD Process

The so-called quasi-birth-death process (QBD process) is a two-dimensional discrete-state continuous-time Markov process $\{X(t), Y(t)\}_{t \geqslant 0}$, such that the *level*, $X(t)$, is a nonnegative integer-valued variable increased or decreased by at most one at each transition, whereas the *phase*, $Y(t)$, may take any value from the *phase state space*, $\mathcal{Y}$, upon transitions [11]. The QBD process is called level-independent if all the transition rates (except the boundary) do not depend on $X(t)$. The two-dimensional structure of the process allows to enumerate the states lexicographically into pairs (x, y), where x is the level value, and y is the phase value. Using this enumeration, the infinitesimal generator matrix, say Q, can be represented in the block-tridiagonal form:

$$
Q = \begin{pmatrix}
A^{0,0} & A^{0,1} & 0 & 0 & \cdots \\
A^{1,0} & A^{(1)} & A^{(0)} & 0 & \cdots \\
0 & A^{(2)} & A^{(1)} & A^{(0)} & \ddots \\
0 & 0 & A^{(2)} & A^{(1)} & \ddots \\
0 & 0 & \ddots & \ddots & \ddots
\end{pmatrix},
$$

where the element $Q[(x_0, y_0), (x_1, y_1)]$ of the matrix is the *transition rate* of the process $\{X(t), Y(t)\}_{t \geqslant 0}$ from the state (x_0, y_0) to the state (x_1, y_1) within adjacent levels, $|x_0 - x_1| \leqslant 1$. As such, $A^{0,0}$ is a square matrix consisting of transition rates at the (boundary) level 0, while $A^{0,1}$ (and $A^{1,0}$, respectively) are the (possibly rectangular) matrices of outgoing (incoming) transitions from (to) level 0. The matrices $A^{(2)}, A^{(1)}, A^{(0)}$ are square matrices consisting of transition rates from level x to $x - 1$, x and $x + 1$, respectively. It is worth noting that only the diagonal elements of Q are negative, and

$$(A^{(0)} + A^{(1)} + A^{(2)})\mathbf{1} = \mathbf{0}. \tag{1}$$

(In what follows, bold letter, say $\mathbf{x}$, denotes a vector, $\mathbf{x}_k$ is kth element of the vector, $\mathbf{1}$ is a vector of ones, and $\mathbf{0}$ is a vector of zeroes.) Commonly in the queueing applications, the level corresponds to the number of customers in the system or in the queue. Then transitions from level x to the level $x - 1$ and the level $x + 1$ correspond to a departure and an arrival, respectively.

Being a special case of a continuous-time Markov chain, the irreducible QBD process has a specific form of the Foster negative drift criterion known as the Neuts ergodicity condition:

$$\boldsymbol{\alpha} A^{(0)} \mathbf{1} < \boldsymbol{\alpha} A^{(2)} \mathbf{1}, \tag{2}$$

where $\boldsymbol{\alpha} \geqslant \mathbf{0}$ is the unique *stochastic* vector satisfying the system of linear algebraic equations

$$\boldsymbol{\alpha}(A^{(0)} + A^{(1)} + A^{(2)}) = 0. \tag{3}$$

Thus the QBD process is essentially considered at high levels when its stability is investigated.

In what follows, we consider the stability condition (2) of a level-independent irreducible QBD with the finite phase space $\mathcal{Y}$. In general, the vector $\boldsymbol{\alpha}$ in (3) has to be found numerically whenever the matrices $A^{(0)}, A^{(1)}$ and $A^{(2)}$ are well defined. However, in the present paper the restrictions imposed on the possible phase transitions allow one to obtain $\boldsymbol{\alpha}$ explicitly, even in the case when the matrix $A^{(2)}$ is not completely defined. Indeed, let $A^{(0)}$ and $A^{(1)}$ be diagonal (in such a case, the arrival process is a Poisson process with possibly phase-dependent rate). Thus *at high levels* the phase, $Y(t)$, may change only at departure epochs. Denote

$$D = \text{diag}(\mathbf{d}),$$

where $\mathbf{d} = A^{(2)}\mathbf{1}$ is the nonnegative vector of row sums of $A^{(2)}$, i.e. $\mathbf{d}_y$ is the total departure rate from phase $y \in \mathcal{Y}$, and $\text{diag}(\mathbf{d})$ is the square matrix consisting of zeroes except the main diagonal that contains the vector $\mathbf{d}$. It then follows from (1) that

$$A^{(1)} = -A^{(0)} - D.$$

Thus (3) may be written as

$$\boldsymbol{\alpha} A^{(2)} = \boldsymbol{\alpha} D, \tag{4}$$

which means that $\boldsymbol{\alpha}$ depends only on the matrix $A^{(2)}$. Intuitively this is expected, since the phase may change only at departure epochs. Moreover, it can be seen from (4) that the matrix $A^{(2)}$ does not necessarily need to be given explicitly. Indeed, l.h.s. of the first equation in (4) is a weighted column sum, whereas r.h.s. uses the row sums of the matrix $A^{(2)}$.

Define the matrix $P = D^{-1}A^{(2)}$. Note that since $A^{(0)}$ and $A^{(1)}$ are diagonal, for the irreducible QBD the matrix D^{-1} exists. By noting that

$$D_{y,y}^{-1} = \mathbf{d}_y^{-1}, \quad y \in \mathcal{Y},$$

it can be easily seen that P is stochastic, since $P\mathbf{1} = \mathbf{1}$. The matrix P contains probabilities of phase transitions of the Markov chain *embedded at departure epochs*. Denote by $\boldsymbol{\pi}$ the steady-state probability vector of P, that is, the solution of $\boldsymbol{\pi}P = \boldsymbol{\pi}$ and $\boldsymbol{\pi}\mathbf{1} = 1$. Let us introduce two new quantities:

$$\boldsymbol{\alpha} = C\boldsymbol{\pi}D^{-1}, \quad C = (\boldsymbol{\pi}D^{-1}\mathbf{1})^{-1} = \left[\sum_{y \in \mathcal{Y}} \boldsymbol{\pi}_y \mathbf{d}_y^{-1}\right]^{-1}. \tag{5}$$

The expression for C follows from the last, balance equation of (4). It is easy to check by definition of D, P and $\boldsymbol{\pi}$ that $\boldsymbol{\alpha}$ defined in (5) satisfies (4). The key observation that follows from (5) is that $\boldsymbol{\alpha}$ depends only on the steady-state vector $\boldsymbol{\pi}$ and on the total departure rate vector $\mathbf{d}$, and thus detailed knowledge of all elements of the matrix $A^{(2)}$ (or P) is not necessary. Note that in such a case, the stability criterion (2) turns out to be

$$\boldsymbol{\alpha}A^{(0)}\mathbf{1} < C, \tag{6}$$

where C is given by (5) and depends only on the vectors $\boldsymbol{\pi}$ and $\mathbf{d}$. By rewriting (6) using $\boldsymbol{\alpha}$ from (5), recalling that both D^{-1} and $A^{(0)}$ are diagonal, we obtain the following intuitively clear stability criterion:

$$\sum_{y \in \mathcal{Y}} \boldsymbol{\pi}_y \rho_y < 1, \tag{7}$$

where

$$\rho_y = \frac{A_{y,y}^{(0)}}{\mathbf{d}_y}, \quad y \in \mathcal{Y}.$$

Indeed, (7) is a classical stability condition on the average load in the system that should be strictly less than one. In case the arrival process is Poisson, $A^{(0)} = \lambda I$ (where I is identity matrix) and (6) transforms into

$$\lambda < C. \tag{8}$$

However, the vector $\boldsymbol{\pi}$ has to be obtained given the matrix $A^{(2)}$ is not completely defined. In what follows, we obtain sufficient conditions for the model that allow $\boldsymbol{\pi}$ to be *guessed*.

3 Stability of a System with Automata-type Transitions

In this paper we consider the special type of queueing systems which we call the systems with *automata-type transitions*. Consider a queueing system in which customers of K classes arrive, $1 \leqslant K < \infty$. Customer's class is determined with a discrete distribution $\mathbf{p} = \{\mathbf{p}_1, \ldots, \mathbf{p}_K\}$ at arrival epochs, which are time epochs of a Poisson process with possibly phase-dependent rate $\lambda_y, y \in \mathcal{Y}$. Class-$k$ customer requires exponentially distributed with rate μ_k service time. The phase state space, $\mathcal{Y}$, is the M-ary Cartesian power of the set $\{1, \ldots, K\}$, i.e.

$$\mathcal{Y} \subseteq \{1, \ldots, K\}^M,$$

and the phase $\mathbf{y} \in \mathcal{Y}$ records the classes of M oldest (in the order of their arrival) customers in the system at time $t \geq 0$ (from now on, in notation we stress that $\mathbf{y}$ is a vector). Customers never change their classes, while being among these M oldest. The classes of all customers being served at time $t \geq 0$ are described in $Y(t)$ in such a way that the service status of each of M oldest customers is unambiguously determined by the phase. More formally, there is an injection

$$\boldsymbol{\sigma} : \mathcal{Y} \to \{0, 1\}^M,$$

such that $\boldsymbol{\sigma}_m(\mathbf{y}) = 1$ if mth customer is being served if the system is in phase $\mathbf{y}$, and $\boldsymbol{\sigma}_m(\mathbf{y}) = 0$ if the mth customer is in the queue, for $m = 1, \ldots, M$.

A particular example of such a system with automata-type transitions is the system with M servers in which each customer occupies at least one server. Here the information on classes of the M oldest customers present in the system (in the order of their arrival) and the total number in the system is sufficient to determine the next system state. Note that the classes of more recent customers (if any) are irrelevant for the system evolution. On the contrast, the two-class single-server system with absolute priorities is the example when the information on all the classes of customers present in the system is required, and thus such a system is not the one with the automata-type transitions and therefore is not in the focus of this paper.

The evolution of the considered system content can be described by the QBD process $\{X(t), Y(t)\}_{t \geqslant 0}$, where $X(t)$ is the total number of customers in the system, $Y(t) = (Y_1(t), \ldots, Y_M(t))$ the phase that contains *classes of the M oldest customers* in the system at time t. In what follows, the M-dimensional vector $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_M)$ is used to indicate the phase of the QBD in a lexicographical order, counting from the oldest customer. This QBD process has the following properties at high levels $X(t) > M$:

1. The phase is changed only at departure epochs.
2. At a departure epoch, the phase, $\mathbf{y}$, evolves in such a way that only one component, say m, (corresponding to the departing customer) of $\mathbf{y}$ is removed (note that this means $\boldsymbol{\sigma}_m(\mathbf{y}) = 1$), and the last component (corresponding to the class of Mth oldest customer after transition) is being sampled from a discrete distribution $\mathbf{p}$. Moreover, all the states in $\mathcal{Y}$ are communicating.

3. At a departure epoch, the total rate from state $(x, \mathbf{y})$ is $\sum_{m=1}^{M} \mu_{\mathbf{y}_m} \boldsymbol{\sigma}_m(\mathbf{y})$.

These properties significantly restrict the structure of the generator submatrices $A^{(i)}, i = 0, 1, 2$. Firstly, the matrices $A^{(0)}$ and $A^{(1)}$ must be diagonal (since the phase is changed only at departure epochs). Secondly, for $\mathbf{y} \in \mathcal{Y}$, the departure rate $\mathbf{d}_\mathbf{y}$ is given by

$$\mathbf{d}_\mathbf{y} = \sum_{m=1}^{M} \mu_{\mathbf{y}_m} \boldsymbol{\sigma}_m(\mathbf{y}).$$

Thirdly, the matrix $A^{(2)}$ is nonnegative with at least one positive element in each row.

Consider now a transition at a departure epoch, governed by the matrix $P = D^{-1} A^{(2)}$. Fix $\mathbf{y} \in \mathcal{Y}$ as the phase right after the transition. Recall that $\mathbf{y}_M$ is the class of the Mth oldest customer in the system after the transition, and this value is sampled from $\mathbf{p}$. Denote by $\mathcal{Y}^-(\mathbf{y})$ the set of states leading to $\mathbf{y}$ by a one-step transition, i.e. for any $\mathcal{Y}^-(\mathbf{y}) = \{\mathbf{u} \in \mathcal{Y} : P(\mathbf{u}, \mathbf{y}) > 0\}$. Break the set $\mathcal{Y}^-(\mathbf{y})$ into the subsets

$$\mathcal{Y}^-(\mathbf{y}) = \bigcup_{k=1}^{K} \mathcal{Y}_k^-(\mathbf{y}),$$

using, as the index, the *class, k, of the departing customer* that caused the transition to $\mathbf{y}$. The following lemma shows that these subsets are disjoint.

Lemma 1. *For any $\mathbf{y} \in \mathcal{Y}$ it holds that $\mathcal{Y}_i^-(\mathbf{y}) \cap \mathcal{Y}_j^-(\mathbf{y}) = \varnothing, i \neq j$.*

Proof. For any $\mathbf{y} \in \mathcal{Y}$ define the characteristic vector

$$\mathbf{n}(\mathbf{y}) = (\mathbf{n}_1(\mathbf{y}), \dots, \mathbf{n}_K(\mathbf{y})), \quad \mathbf{n}_k(\mathbf{y}) = \sum_{m=1}^{M} I(\mathbf{y}_m = k), \quad k = 1, \dots, K, \quad (9)$$

where $I(\cdot)$ is the indicator function. Note that $\mathbf{n}_k(\mathbf{y})$ is the number of class-k customers among the M oldest in phase $\mathbf{y}$.

From (9) it follows that

$$\mathbf{n}(\mathbf{u}) = \mathbf{n}(k, \mathbf{y}_1, \dots, \mathbf{y}_{M-1}), \quad \mathbf{u} \in \mathcal{Y}_k^-(\mathbf{y}),$$

and $\mathbf{n}_k(\mathbf{u}) \geqslant 1$. Denote $\mathbf{n}^{(k)} = \mathbf{n}(\mathbf{u}), \mathbf{u} \in \mathcal{Y}_k^-(\mathbf{y})$. Then

$$\mathbf{n}(\mathbf{y}) = \mathbf{n}^{(k)} - \mathbf{e}_k + \mathbf{e}_{\mathbf{y}_M},$$

where $\mathbf{e}_k$ is an M-dimensional vector with 1 at the kth position and zero elsewhere. Thus for $i \neq j$, $\mathbf{n}^{(i)} - \mathbf{e}_i = \mathbf{n}^{(j)} - \mathbf{e}_j$, and hence $\mathbf{n}^{(i)} \neq \mathbf{n}^{(j)}$. $\qquad \square$

Due to the Lemma 1 from the balance equations $\boldsymbol{\pi} = \boldsymbol{\pi} P$ we have

$$\boldsymbol{\pi}_\mathbf{y} = \sum_{k=1}^{K} \sum_{\mathbf{u} \in \mathcal{Y}_k^-(\mathbf{y})} \boldsymbol{\pi}_\mathbf{u} P(\mathbf{u}, \mathbf{y}), \quad \mathbf{y} \in \mathcal{Y}. \qquad (10)$$

Take $\boldsymbol{\pi}_y$ in the *product form*:

$$\boldsymbol{\pi}_{\mathbf{y}} = \prod_{m=1}^{M} \mathbf{p}_{\mathbf{y}_m}, \quad \mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_M) \in \mathcal{Y}. \tag{11}$$

Note that any element (i.e. phase) in the set $\mathcal{Y}_k^-(\mathbf{y})$ contains classes $\mathbf{y}_1, \ldots, \mathbf{y}_{M-1}$ and the class of departing customer. Thus for any $\mathbf{u} \in \mathcal{Y}_k^-(\mathbf{y})$, according to (11),

$$\boldsymbol{\pi}_{\mathbf{u}} = \mathbf{p}_k \prod_{m=1}^{M-1} \mathbf{p}_{\mathbf{y}_m},$$

and (10) transforms into

$$\boldsymbol{\pi}_{\mathbf{y}} = \sum_{k=1}^{K} \mathbf{p}_k \prod_{m=1}^{M-1} \mathbf{p}_{\mathbf{y}_m} \left[\sum_{\mathbf{u} \in \mathcal{Y}_k^-(\mathbf{y})} P(\mathbf{u}, \mathbf{y}) \right].$$

Thus, if for all $k = 1, \ldots, K$, $\mathbf{y} \in \mathcal{Y}$,

$$\sum_{\mathbf{u} \in \mathcal{Y}_k^-(\mathbf{y})} P(\mathbf{u}, \mathbf{y}) = \mathbf{p}_{\mathbf{y}_M}, \tag{12}$$

then the product form (11) is the solution of the balance equations.

In the next lemma we give the sufficient condition for (12) to hold.

Lemma 2. *For $m = 1, \ldots, M$, $k = 1, \ldots, K$ and $\mathbf{y} \in \mathcal{Y}$ construct the vectors $\boldsymbol{\varphi}^{(m,k)}(\mathbf{y}) = (\mathbf{y}_1, \ldots, \mathbf{y}_{m-1}, k, \mathbf{y}_m, \ldots, \mathbf{y}_{M-1})$. Then (12) holds if and only if for any $\mathbf{y} \in \mathcal{Y}$*

$$\sum_{m=1}^{M} \frac{\mu_k \boldsymbol{\sigma}_m(\boldsymbol{\varphi}^{(m,k)}(\mathbf{y}))}{\mathbf{d}_{\boldsymbol{\varphi}^{(m,k)}(\mathbf{y})}} = 1. \tag{13}$$

Proof. Fix $\mathbf{y} \in \mathcal{Y}$ and $k \in \{1, \ldots, K\}$. For any $\mathbf{u} \in \mathcal{Y}_k^-(\mathbf{y})$ there exists such $i \in \{1, \ldots, M\}$ that $\mathbf{u} = \boldsymbol{\varphi}^{(i,k)}(\mathbf{y})$ and $\boldsymbol{\sigma}_i(\mathbf{u}) = 1$ (possibly there are several such indices). Similarly, for a fixed i, by construction, $\boldsymbol{\varphi}^{(i,k)}(\mathbf{y}) \in \mathcal{Y}_k^-(\mathbf{y})$ if and only if $\boldsymbol{\sigma}_i(\boldsymbol{\varphi}^{(i,k)}(\mathbf{y})) = 1$. Thus,

$$\sum_{\mathbf{u} \in \mathcal{Y}_k^-(\mathbf{y})} P_{\mathbf{u},\mathbf{y}} = \sum_{m=1}^{M} \mathsf{P}^{(m)}_{\boldsymbol{\varphi}^{(m,k)}(\mathbf{y}),\mathbf{y}}, \tag{14}$$

where $\mathsf{P}^{(m)}_{\boldsymbol{\varphi}^{(m,k)}(\mathbf{y}),\mathbf{y}}$ is the probability of transition from the state $\boldsymbol{\varphi}^{(m,k)}(\mathbf{y})$ to $\mathbf{y}$ by departure of mth oldest customer. The latter equals, by construction of the considered embedded Markov chain,

$$\mathsf{P}^{(m)}_{\boldsymbol{\varphi}^{(m,k)}(\mathbf{y}),\mathbf{y}} = \mathbf{p}_{y_M} \frac{\mu_k \boldsymbol{\sigma}_m(\boldsymbol{\varphi}^{(m,k)}(\mathbf{y}))}{d_{\boldsymbol{\varphi}^{(m,k)}(\mathbf{y})}},$$

and thus (13) together with (14) give (12). $\qquad \square$

Now we formulate a sufficient condition for Lemma 2 to hold.

Corollary 1. *Let for any* $\mathbf{y} \in \mathcal{Y}$, $k = 1, \ldots, K$, *and* $\mathbf{u} \in \mathcal{Y}_k^-(\mathbf{y})$, $\mu_{\mathbf{u}_m} = \mu_k$ *hold for any* $m = 1, \ldots, M$ *such that* $\boldsymbol{\sigma}_m(\mathbf{u}) = 1$. *Let* $\boldsymbol{\sigma}_{k,\mathbf{y}} := \sum_{m=1}^M \boldsymbol{\sigma}_m(\mathbf{u})$ *be constant for all* $\mathbf{u} \in \mathcal{Y}_k^-(\mathbf{y})$. *Then* (13) *holds if*

$$\sum_{m=1}^M \boldsymbol{\sigma}_m(\boldsymbol{\varphi}^{(m,k)}(\mathbf{y})) = \boldsymbol{\sigma}_{k,\mathbf{y}}. \tag{15}$$

Proof. Under the assumptions made,

$$\mathbf{d}_{\boldsymbol{\varphi}^{(m,k)}(\mathbf{y})} = \mu_k \boldsymbol{\sigma}_{k,\mathbf{y}},$$

and hence

$$\sum_{m=1}^M \frac{\mu_k \boldsymbol{\sigma}_m(\boldsymbol{\varphi}^{(m,k)}(\mathbf{y}))}{d_{\boldsymbol{\varphi}^{(m,k)}(\mathbf{y})}} = \sum_{m=1}^M \frac{\boldsymbol{\sigma}_m(\boldsymbol{\varphi}^{(m,k)}(\mathbf{y}))}{\boldsymbol{\sigma}_{k,\mathbf{y}}} = 1.$$

$\square$

4 Model Examples

4.1 Oldest Customer Class Chunk Policy

Consider the M-server system with K classes of customers, Poisson arrival process of rate λ, class-dependent service rates μ_k and the following service policy: a group of the oldest customers of the same class are served, one customer per server. We may assume that a customer's class is sampled at Mth position in the system (ordered by arrival times). In such a case, $\mathcal{Y} = \{1, \ldots, K\}^M$ is the set of classes of M oldest customers in the system. Fix some $\mathbf{y} \in \mathcal{Y}$ and some class k. Then $\mathcal{Y}_k^-(\mathbf{y})$ contains only one state, $\mathbf{u} = (k, \mathbf{y}_1, \ldots, \mathbf{y}_{M-1})$. Moreover, $\boldsymbol{\sigma}(\boldsymbol{\varphi}^{(m,k)}(\mathbf{y})) = (1, 0, \ldots, 0)$ if $k \neq \mathbf{y}_1$, otherwise

$$\sum_{m=1}^M \boldsymbol{\sigma}_m(\boldsymbol{\varphi}^{(m,k)}(\mathbf{y})) = \sum_{m=1}^M \boldsymbol{\sigma}_m(\mathbf{u}) = \max\{i \leq M : \mathbf{y}_1 = \cdots = \mathbf{y}_i = k\} + 1.$$

Thus the stability criterion follows, after some algebra, from (7):

$$\lambda \sum_{k=1}^K \frac{1}{\mu_k} \left[(1 - \mathbf{p}_k) \sum_{m=1}^{M-1} \frac{\mathbf{p}_k^m}{m} + \frac{\mathbf{p}_k^M}{M} \right] < 1.$$

4.2 Weird Policy

Consider a two-server system with $K = 2$ types of customers. For $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2) \in \mathcal{Y} = \{1, 2\}^2$ define the following service rule: $\boldsymbol{\sigma}(1, 1) = (0, 1), \boldsymbol{\sigma}(1, 2) = (1, 0)$, $\boldsymbol{\sigma}(2, 1) = (1, 0), \boldsymbol{\sigma}(2, 2) = (1, 1)$. Thus, similar to the model in Section 4.1, for any $\mathbf{y} \in \mathcal{Y}$ and any $k = 1, 2$, the set $\mathcal{Y}_k^-(\mathbf{y})$ contains only one state. Finally, the conditions of Corollary 1 are satisfied (which can be checked for each state $\mathbf{y} \in \mathcal{Y}$) and the product form stability criterion can be stated as

$$\lambda \left[\frac{\mathbf{p}_1}{\mu_1} + \frac{2\mathbf{p}_2 - \mathbf{p}_2^2}{2\mu_2} \right] < 1.$$

4.3 Supercomputer Model with Phase-Dependent Arrival Rate

In this section we reformulate the result first obtained in [17]. Consider an M-server sytem with $K = M$ classes of customers, where the input follows Poisson process of rate λ, and a customer of class k (arriving with probability $\mathbf{p}_k$) requires k servers simultaneously for a random time, exponentially distributed with rate μ. The servers dedicated to a customer all are seized and released simultaneously. Clearly the service discipline is not work-conserving (a customer may be waiting in the queue when there are idle servers). As it was shown in [12,17], the system evolution may be described by the QBD process with $X(t)$ being the number of customers in the system and $Y(t)$ containing the classes of M oldest customers in the system (placed in the order of arrival). Thus for stability we consider the case $X(t) > M$ and following the Corollary 1 check the sufficient condition (15) for the product form of $\boldsymbol{\pi_y}$.

Fix $\mathbf{y} \in \mathcal{Y} = \{1, \ldots, M\}^M$ and define $\boldsymbol{\sigma}_m(\mathbf{y}) = 1, m \leqslant m(\mathbf{y})$, where

$$m(\mathbf{y}) = \max \left\{ i \leq M : \sum_{j=1}^{i} \mathbf{y}_j \leq M \right\}.$$

Fix $k \leq M$ and observe that for any $\mathbf{u} \in \mathcal{Y}_k^-(\mathbf{y})$, $\sum_{m=1}^{M} \boldsymbol{\sigma}_m(\mathbf{u}) = \boldsymbol{\sigma}_{k,\mathbf{y}}$ and is constant. Indeed, for any such $\mathbf{u}$,

$$m(\mathbf{u}) = 1 + \max \left\{ i \leq M : k + \sum_{j=1}^{i} \mathbf{y}_j \leq M \right\}, \tag{16}$$

and hence $m(\mathbf{u})$ depends only on k and $\mathbf{y}$. It can be seen that $m(\mathbf{u}) = \boldsymbol{\sigma}_{k,\mathbf{y}}$. Finally, it follows from (16) that for any $m \leq M$ the value $\boldsymbol{\sigma}_m(\boldsymbol{\varphi}^{(m,k)}(\mathbf{y})) = 1$ if and only $m \leq \boldsymbol{\sigma}_{k,\mathbf{y}}$, and thus (15) follows.

Interestingly, the stability criterion of the supercomputer model in [17] was given in the form (8), where, as it was shown in [16], the value C may be obtained, after some algebra, in the form

$$C = \left[\sum_{m=1}^{M} \frac{1}{m\mu} \sum_{j=m}^{M} \mathbf{p}_j^{*m} \sum_{k=M-j+1}^{M} \mathbf{p}_k \right]^{-1},$$

where the summation is performed over the number m of customers served, the number j of servers occupied by m customers, and the number k of servers required by the first customer waiting in the queue, and $\mathbf{p}_j^{*m}$ is the jth component of m times discrete convolution of the vector $\mathbf{p}$ with itself. However, the form of the result (7) allows to generalize the criterion to the model with *phase-dependent input rate*.

Indeed, let $\lambda_{\mathbf{y}}$ be the input rate in the system if $\mathbf{y} \in \mathcal{Y}$ is the phase of the system (we avoid the discussion of the service rate for the case $X(t) < M$ since it is not used in the stability analysis, for completeness we may fix it as λ). Then

$A_{\mathbf{y},\mathbf{y}}^{(0)} = \lambda_{\mathbf{y}}$. It follows from (7) that the following stability criterion holds for the system with phase-dependent input rate:

$$\sum_{\mathbf{y}\in\mathcal{Y}} \prod_{m=1}^{M} \mathbf{p}_{\mathbf{y}_m} \frac{\lambda_{\mathbf{y}}}{m(\mathbf{y})\mu} < 1,$$

which is reduced to (8) if $\lambda_{\mathbf{y}} \equiv \lambda$.

4.4 Supercomputer Model with Different Service Rates

In this section we further generalize the supercomputer model by considering the case when classes have different service rates. It will be clear from subsequent analysis that the stability condition can be explicitly obtained for the model with $M = 2$ servers/classes. In this regard, we mention the pioneering work [6] where the 2-server model of this kind was first studied in a rather general context, for the general distribution of the class-2 customers (although the stability condition of such a case was not discussed extensively).

Let customer of class m occupy m servers for exponentially distributed service time of rate μ_m, $m = 1, 2$. Then the set $\mathcal{Y}$ contains only 4 phases, namely, the pairs $(\mathbf{y}_1, \mathbf{y}_2), \mathbf{y}_1, \mathbf{y}_2 \in \{1, 2\}$. It is also clear that $\boldsymbol{\sigma}(\mathbf{y}) = (1, 1)$ only if $\mathbf{y} = (1, 1)$, and otherwise $\boldsymbol{\sigma}(\mathbf{y}) = (1, 0)$. The subsequent analysis does not differ from Section 4.3, we only note that additionally we need to check the precondition of Corollary 1 that for any $\mathbf{y} \in \mathcal{Y}$ and $k \leq 2$, the rate $\mu_{\mathbf{u}_m} = \mu_k$, for $\mathbf{u} \in \mathcal{Y}_k^-(\mathbf{y})$ and $m = 1, 2$ such that $\boldsymbol{\sigma}_m(\mathbf{u}) = 1$. This is straightforward, since the only phase where both customers are served is $(1, 1)$, and the both have service rate μ_1. However, this can not be generalized towards the system with $M > 2$ servers (if all classes are non-degenerate). It then follows that the stability condition of such a model with $M = 2$ servers is

$$\frac{\mathbf{p}_1^2 \lambda_{(1,1)}}{2\mu_1} + \frac{\mathbf{p}_1 \mathbf{p}_2 \lambda_{(1,2)}}{\mu_1} + \frac{\mathbf{p}_1 \mathbf{p}_2 \lambda_{(2,1)}}{\mu_2} + \frac{\mathbf{p}_2^2 \lambda_{(2,2)}}{\mu_2} < 1,$$

which becomes $\lambda < 2\mu/(2 - \mathbf{p}_1^2)$ (firstly appeared in [6]), if $\lambda_{\mathbf{y}} \equiv \lambda$ and $\mu_m \equiv \mu$, $\mathbf{y} \in \mathcal{Y}, m = 1, 2$.

5 Discussion

In the present paper we have introduced the special type of queueing systems with automata-type transitions. The idea is inspired by the concept of natural numbers and finite automata [2,3], where, loosely speaking, the output of a finite automation under certain conditions may have similar stochastic properties as the input. This theory was further developed into a concept of Stochastic Automata Networks [10] which also possess the product form solution under certain conditions on the automata interaction. We note that product form in many Markovian frameworks is tightly related to reversibility, either in time [5,18,21]

or in structure [14]. In the present context, the reversibility may be thought of as time reversal, if for a state $\mathbf{y}$ we first sample the class of departing customer as k, and then sample the previous state as $\mathbf{u} \in \mathcal{Y}_k^-(\mathbf{y})$ (note that the conditions of the Corollary 1 give the probabilities for such a sampling). This, however, requires proving that the probability of a departing (which corresponds to arrival backward in time) customer class k is $\mathbf{p}_k$, which we leave beyond the scope of this paper.

A natural question arises: is there an extension of the automata-type policy? An essential component is the possibility to make backward-in-time transitions, i.e. to define for any $\mathbf{y} \in \mathcal{Y}$ the sets of states $Y_k^-(\mathbf{y})$. This may be hard or even impossible, if the customers are ordered not in the order of arrival, or in any other deterministic way.

Another point of discussion is the connection between the QBD model and a more general simulation framework. Namely, if $Y(t)$ is the phase process, and $X(t)$ is the level process, then one can consider the so-called Generalized Semi-Markov Process related to the process $\{X(t), Y(t)\}_{t \geq 0}$ in case the transitions are not Markovian. In such a case, the values $\boldsymbol{\sigma}_i(\mathbf{y})\mu_{\mathbf{y}_i}$ may be thought of as *clock speeds*, and we need to keep track of the remaining work (which is not exponential in general), and possibly track the remaining interarrival time. Moreover, intuitively, the row-wise transformation of $A^{(2)}$ to P by multiplying on D^{-1} is inspired by the connection between time- and event-stationarity in a MAM framework, see e.g. [4]. At the same time, this is a more general property related to Palm theory, see [7,19]. This shows a direction for further study of the applicability of the obtained results to stochastic models beyond exponential distributions.

The possible generalizations of the model come from the point that at arrivals the phase is not changed. Thus, it seems technically straightforward to generalize the result towards MAP arrival processes (or even batch MAP arrivals), since the arrival phases will be unrelated to the phase of the system, and the rest will be done with Kronecker algebra. This is a promising direction of future research.

References

1. Afanaseva, L., Bashtova, E., Grishunina, S.: Stability Analysis of a Multi-server Model with Simultaneous Service and a Regenerative Input Flow. Methodology and Computing in Applied Probability (Jul 2019). https://doi.org/10/ggnsgj, `http://link.springer.com/10.1007/s11009-019-09721-9`
2. Agafonov, V.: Normal sequences and finite automata. Soviet Mathematics Doklady **9**, 324–325 (1968)

3. Becher, V., Heiber, P.A.: Normal numbers and finite automata. Theoretical Computer Science **477**, 109–116 (Mar 2013). https://doi.org/10/f4rn5f, `https://linkinghub.elsevier.com/retrieve/pii/S0304397513000698`

4. Bladt, M., Nielsen, B.F.: Matrix-Exponential Distributions in Applied Probability, Probability Theory and Stochastic Modelling, vol. 81. Springer US, Boston, MA (2017). https://doi.org/10.1007/978-1-4939-7049-0, `http://link.springer.com/10.1007/978-1-4939-7049-0`

5. Brill, P.H., Cheung, C.h., Hlynka, M., Jiang, Q.: Reversibility Checking for Markov Chains. Communications on Stochastic Analysis **12**(2) (Oct 2018). https://doi.org/10.31390/cosa.12.2.02, `https://digitalcommons.lsu.edu/cosa/vol12/iss2/2`

6. Brill, P.H., Green, L.: Queues in which customers receive simultaneous service from a random number of servers: a system point approach. Management Science **30**(1), 51–68 (1984). https://doi.org/10.1287/mnsc.30.1.51, `http://pubsonline.informs.org/doi/abs/10.1287/mnsc.30.1.51`

7. Brémaud, P.: Characteristics of queueing systems observed at events and the connection between stochastic intensity and Palm probability. Queueing systems **5**(1-3), 99–111 (1989). https://doi.org/10.1007/BF01149188, `http://link.springer.com/article/10.1007/BF01149188`

8. Chakravarthy, S.R., Karatza, H.D.: Two-server parallel system with pure space sharing and Markovian arrivals. Computers & Operations Research **40**(1), 510 – 519 (2013). https://doi.org/10.1016/j.cor.2012.08.002

9. Foster, F.G.: On the Stochastic Matrices Associated with Certain Queuing Processes. Ann. Math. Statist. **24**(3), 355–360 (Sep 1953). https://doi.org/10.1214/aoms/1177728976, `https://projecteuclid.org:443/euclid.aoms/1177728976`, publisher: The Institute of Mathematical Statistics

10. Fourneau, J., Plateau, B., Stewart, W.: Product form for Stochastic Automata Networks. In: Proceedings of the 2nd International ICST Conference on Performance Evaluation Methodologies and Tools. ICST, Nantes, France (2007). https://doi.org/10.4108/valuetools.2007.1980, `http://eudl.eu/doi/10.4108/valuetools.2007.1980`

11. He, Q.M.: Fundamentals of Matrix-Analytic Methods. Springer New York (2014)

12. Kim, S.: M/M/s Queueing System Where Customers Demand Multiple Server Use. Ph.D. thesis, Southern Methodist University (1979)

13. Meyn, S.P., Tweedie, R.L.: Markov chains and stochastic stability. Springer Science & Business Media (2012)

14. Miyazawa, M.: Reversibility in Queueing Models. In: Wiley Encyclopedia of Operations Research and Management Science, pp. 1–19. American Cancer Society (2013). https://doi.org/10.1002/9780470400531.eorms1079, `https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470400531.eorms1079`

15. Neuts, M.F.: Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach. Johns Hopkins University Press, Baltimore (1981)

16. Rumyantsev, A., Morozov, E.: Accelerated Verification of Stability of Simultaneous Service Multiserver Systems. In: Proceedings of 2015 7th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), Brno, Czech Republic, 6-8 October 2015. pp. 239–242. IEEE (2015)

17. Rumyantsev, A., Morozov, E.: Stability criterion of a multiserver model with simultaneous service. Annals of Operations Research **252**(1), 29–39 (2017). https://doi.org/10.1007/s10479-015-1917-2, `https://doi.org/10.1007/s10479-015-1917-2`

18. Taylor, P.: Quasi-reversibility and networks of queues with nonstandard batch movements. Mathematical and Computer Modelling **31**(10-12), 335–341 (May 2000). https://doi.org/10.1016/S0895-7177(00)00104-7, `https://linkinghub.elsevier.com/retrieve/pii/S0895717700001047`

19. Thorisson, H.: On time- and cycle-stationarity. Stochastic Processes and their Applications **55**(2), 183–209 (Feb 1995). https://doi.org/10.1016/0304-4149(94)00038-U, `http://linkinghub.elsevier.com/retrieve/pii/030441499400038U`

20. Tweedie, R.L.: Relations between ergodicity and mean drift for markov chains. Australian Journal of Statistics **17**(2), 96–102 (Jun 1975). https://doi.org/10.1111/j.1467-842X.1975.tb00945.x, `https://doi.org/10.1111/j.1467-842X.1975.tb00945.x`, publisher: John Wiley & Sons, Ltd

21. Walrand, J., Varaiya, P.: Interconnections of Markov chains and quasi-reversible queuing networks. Stochastic Processes and their Applications **10**(2), 209–219 (Sep 1980). https://doi.org/10.1016/0304-4149(80)90022-8, `http://www.sciencedirect.com/science/article/pii/0304414980900228`