

Block-structured supermarket models

Quan-Lin Li · John C. S. Lui

Received: 25 October 2013 / Accepted: 5 June 2014 / Published online: 29 June 2014
© Springer Science+Business Media New York 2014

Abstract Supermarket models are a class of parallel queueing networks with an adaptive control scheme that play a key role in the study of resource management of, such as, computer networks, manufacturing systems and transportation networks. When the arrival processes are non-Poisson and the service times are non-exponential, analysis of such a supermarket model is always limited, interesting, and challenging. This paper describes a supermarket model with non-Poisson inputs: Markovian Arrival Processes (MAPs) and with non-exponential service times: Phase-type (PH) distributions, and provides a generalized matrix-analytic method which is first combined with the operator semigroup and the mean-field limit. When discussing such a more general supermarket model, this paper makes some new results and advances as follows: (1) Providing a detailed probability analysis for setting up an infinite-dimensional system of differential vector equations satisfied by the expected fraction vector, where *the invariance of environment factors* is given as an important result. (2) Introducing the phase-type structure to the operator semigroup and to the mean-field limit, and a Lipschitz condition can be established by means of a unified matrix-differential algorithm. (3) The matrix-analytic method is used to compute the fixed point which leads to performance computation of this system. Finally, we use some numerical examples to illustrate how the performance measures of this supermarket model depend on the non-Poisson inputs and on the non-exponential service times. Thus the results of this paper give new highlight on understanding influence of non-Poisson inputs and of non-exponential service times on performance measures of more general supermarket models.

Q.-L. Li (✉)
School of Economics and Management Sciences, Yanshan University,
Qinhuangdao 066004, China
e-mail: liquanlin@mail.tsinghua.edu.cn

J. C. S. Lui
Department of Computer Science and Engineering, The Chinese University of Hong Kong,
Shatin, N.T, Hong Kong

Keywords Randomized load balancing · Supermarket model · Matrix-analytic method · Operator semigroup · Mean-field limit · Markovian arrival processes (MAP) · Phase-type (PH) distribution · Invariance of environment factors · Doubly exponential tail · *RG*-factorization.

1 Introduction

Supermarket models are a class of parallel queueing networks with an adaptive control scheme that play a key role in the study of resource management of, such as computer networks (e.g., see the dynamic randomized load balancing), manufacturing systems and transportation networks. Since a simple supermarket model was discussed by Mitzenmacher (1996), Vvedenskaya et al. (1996) and Turner (1996) through queueing theory as well as Markov processes, subsequent papers have been published on this theme, among which, see, Vvedenskaya and Suhov (1997, 2005), Turner (1998), Jacquet and Vvedenskaya (1998), Jacquet et al. (1999), Mitzenmacher (1999), Mitzenmacher and Upfal (2005), Kurtz (1981), Graham (2000a, b, 2004), Mitzenmacher et al. (2001), Luczak and Norris (2005), Luczak and McDiarmid (2006, 2007), Bramson et al. (2010, 2012, 2013), Li and Lui (2010), Li et al. (2011, 2013) and Li (2011, 2014). For the fast Jackson networks (or the supermarket networks), readers may refer to Martin and Suhov (1999), Martin (2001) and Suhov and Vvedenskaya (2002).

The available results of the supermarket models with non-exponential service times are still few in the literature. Important examples include an approximate method of integral equations by Vvedenskaya and Suhov (1997), the Erlang service times by Mitzenmacher (1999) and Mitzenmacher et al. (2001), the PH service times by Li et al. (2011) and Li and Lui (2010), and the ansatz-based modularized program for the general service times by Bramson et al. (2010, 2012, 2013).

Little work has been done on analysis of the supermarket models with non-Poisson inputs, which are more difficult and challenging due to the higher complexity of that N arrival processes are superposed. Li and Lui (2010) and Li (2011) used the superposition of N MAP inputs to study the infinite-dimensional Markov processes of supermarket modeling type. Comparing with the results given in Li and Lui (2010) and Li (2011), this paper provides more necessary phase-level probability analysis in setting up the infinite-dimensional system of differential vector equations, which leads some new results and methodologies in the study of block-structured supermarket models. Note that the PH distributions constitute a versatile class of distributions that can approximate arbitrarily closely any probability distribution defined on the nonnegative real line, and the MAPs are a broad class of renewal or non-renewal point processes that can approximate arbitrarily closely any stochastic counting process (e.g., see Neuts (1981, 1989) and Li (2010) for more details), thus the results of this paper are a key advance of those given in Mitzenmacher (1996) and Vvedenskaya et al. (1996) under the Poisson and exponential setting.

The main contributions of this paper are threefold. The first one is to use the MAP inputs and the PH service times to describe a more general supermarket model with non-Poisson inputs and with non-exponential service times. Based on the phase-type structure, we define the random fraction vector and construct an infinite-dimensional Markov process, which expresses the state of this supermarket model by means of an infinite-dimensional Markov process. Furthermore, we set up an infinite-dimensional system of differential vector

equations satisfied by the expected fraction vector through a detailed probability analysis. To that end, we obtain an important result: The invariance of environment factors, which is a key for being able to simplify the differential equations in a vector form. Based on the differential vector equations, we can provide a generalized matrix-analytic method to investigate more general supermarket models with non-Poisson inputs and with non-exponential service times. The second contribution of this paper is to provide the phase-type structure for the operator semigroup with respect to the MAP inputs and to the PH service times, and use the operator semigroup to provide the mean-field limit for the sequence of Markov processes who asymptotically approaches a single trajectory identified by the unique and global solution to the infinite-dimensional system of limiting differential vector equations. To prove the existence and uniqueness of solution through the Picard approximation, we provide a unified matrix-differential algorithm for establishing a Lipschitz condition, which is crucial in all the rigor proofs involved. The third contribution of this paper is to provide a generalized matrix-analytic method both for computing the fixed point and for analyzing the performance measures of this supermarket model. Furthermore, we use some numerical examples to indicate how the performance measures of this supermarket model depend on the non-Poisson MAP inputs and on the non-exponential PH service times. Therefore, the results of this paper gives new highlight on understanding influence of non-Poisson inputs and of non-exponential service times on performance measures of more general supermarket models.

The remainder of this paper is organized as follows. In Section 2, we first introduce a new MAP whose transition rates are controlled by the number of servers in the system. Then we describe a more general supermarket model of N identical servers with MAP inputs and PH service times. In Section 3, we define a random fraction vector and construct an infinite-dimensional Markov process, which expresses the state of this supermarket model. In Section 4, we set up an infinite-dimensional system of differential vector equations satisfied by the expected fraction vector through a detailed probability analysis, and establish an important result: The invariance of environment factors. In Section 5, we show that the mean-field limit for the sequence of Markov processes who asymptotically approaches a single trajectory identified by the unique and global solution to the infinite-dimensional system of limiting differential vector equations. To prove the existence and uniqueness of the solution, we provide a unified matrix-differential algorithm for establishing the Lipschitz condition. In Section 6, we first discuss the stability of this supermarket model in terms of a coupling method. Then we provide a generalized matrix-analytic method for computing the fixed point whose doubly exponential solution and phase-structured tail are obtained. Finally, we discuss some useful limits of the fraction vector $\mathbf{u}^{(N)}(t)$ as $N \rightarrow \infty$ and $t \rightarrow +\infty$. In Section 7, we provide two performance measures of this supermarket model, and use some numerical examples to indicate how the performance measures of this system depend on the non-Poisson MAP inputs and on the non-exponential PH service times. Some concluding remarks are given in Section 8.

To shorten this paper, three appendices are moved to Li and Lui (2014) as an online supplementary material (see <http://arxiv.org/pdf/1406.0285.pdf>), where Appendices A and C are respectively designed for the proofs of Theorems 1 and 3, and Appendix B contains the proof of Theorem 2, where the mean-field limit of the sequence of Markov processes in this supermarket model is given a detailed analysis through the operator semigroup. We hope that such an online material will also be convenient for readers, if caring for the rigor proofs.

2 Supermarket model description

In this section, we first introduce a new MAP whose transition rates are controlled by the number of servers in the system. Then we describe a more general supermarket model of N identical servers with MAP inputs and PH service times.

2.1 A new Markovian arrival process

Based on Chapter 5 in Neuts (1989), the MAP is a bivariate Markov process $\{(N(t), J(t)) : t \geq 0\}$ with state space $S = \{1, 2, 3, \dots\} \times \{1, 2, \dots, m_A\}$, where $\{N(t) : t \geq 0\}$ is a counting process of arrivals and $\{J(t) : t \geq 0\}$ is a Markov environment process. When $J(t) = i$, if the random environment shall go to state j in the next time, then the counting process $\{N(t) : t \geq 0\}$ is a Poisson process with arrival rate $d_{i,j}$ for $1 \leq i, j \leq m_A$. The matrix D with elements $d_{i,j}$ satisfies $D \geq 0$. The matrix C with elements $c_{i,j}$ has negative diagonal elements and nonnegative off-diagonal elements, and the matrix C is invertible, where $c_{i,j}$ is a state transition rate of the Markov chain $\{J(t) : t \geq 0\}$ from state i to state j for $i \neq j$. The matrix $Q = C + D$ is the infinitesimal generator of an irreducible Markov chain. We assume that $Qe = 0$, where e is a column vector of ones with a suitable size. Hence, we have

$$c_{i,i} = - \left[\sum_{j=1}^{m_A} d_{i,j} + \sum_{j \neq i}^{m_A} c_{i,j} \right].$$

Let

$$\mathbb{C} = \begin{pmatrix} -\sum_{j \neq 1}^{m_A} c_{1,j} & c_{1,2} & \cdots & c_{1,m_A} \\ c_{2,1} & -\sum_{j \neq 2}^{m_A} c_{2,j} & \cdots & c_{2,m_A} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m_A,1} & c_{m_A,2} & \cdots & -\sum_{j \neq m_A}^{m_A} c_{m_A,j} \end{pmatrix},$$

$$C(N) = \mathbb{C} - N \text{diag}(De),$$

$$D(N) = ND,$$

where

$$\text{diag}(De) = \text{diag} \left(\sum_{j=1}^{m_A} d_{1,j}, \sum_{j=1}^{m_A} d_{2,j}, \dots, \sum_{j=1}^{m_A} d_{m_A,j} \right).$$

Then

$$Q(N) = C(N) + D(N) = [\mathbb{C} - N \text{diag}(De)] + ND$$

is obviously the infinitesimal generator of an irreducible Markov chain with m_A states. Thus $(C(N), D(N))$ is the irreducible matrix descriptor of a new MAP of order m_A . Note that the new MAP is non-Poisson and may also be non-renewal, and its arrival rate at each environment state is controlled by the number N of servers in the system.

Note that

$$Q(N)e = [\mathbb{C} - N \text{diag}(De)]e + NDe = 0,$$

the Markov chain $Q(N)$ with m_A states is irreducible and positive recurrent. Let ω_N be the stationary probability vector of the Markov chain $Q(N)$. Then ω_N depends on the number $N \geq 1$, and the stationary arrival rate of the MAP is given by $N\lambda_N = N\omega_N De$.

2.2 Model description

Based on the new MAP, we describe a more general supermarket model of N identical servers with MAP inputs and PH service times as follows:

Non-Poisson inputs Customers arrive at this system as the MAP of irreducible matrix descriptor $(C(N), D(N))$ of size m_A , whose stationary arrival rate is given by $N\lambda_N = N\omega_N De$.

Non-exponential service times The service times of each server are i.i.d. and are of phase type with an irreducible representation (α, T) of order m_B , where the row vector α is a probability vector whose j th entry is the probability that a service begins in phase j for $1 \leq j \leq m_B$; T is a matrix of size m_B whose (i, j) th entry is denoted by $t_{i,j}$ with $t_{i,i} < 0$ for $1 \leq i \leq m_B$, and $t_{i,j} \geq 0$ for $i \neq j$. Let $T^0 = -Te = (t_1^0, t_2^0, \dots, t_{m_B}^0)^T \geq 0$, where “ A^T ” denotes the transpose of matrix (or vector) A . When a PH service time is in phase i , the transition rate from phase i to phase j is $t_{i,j}$, the service completion rate is t_i^0 , and the output rate from phase i is $\mu_i = -t_{i,i}$. At the same time, the mean of the PH service time is given by $1/\mu = -\alpha T^{-1}e$.

Arrival and service disciplines Each arriving customer chooses $d \geq 1$ servers independently and uniformly at random from the N identical servers, and waits for its service at the server which currently contains the fewest number of customers. If there is a tie, servers with the fewest number of customers will be chosen randomly. All customers in any server will be served in the FCFS manner. Figure 1 gives a physical interpretation for this supermarket model.

Remark 1 The block-structured supermarket models can have many practical applications to, such as, computer networks and manufacturing system, where it is a key to introduce the PH service times and the MAP inputs to such a practical model, because the PH distributions

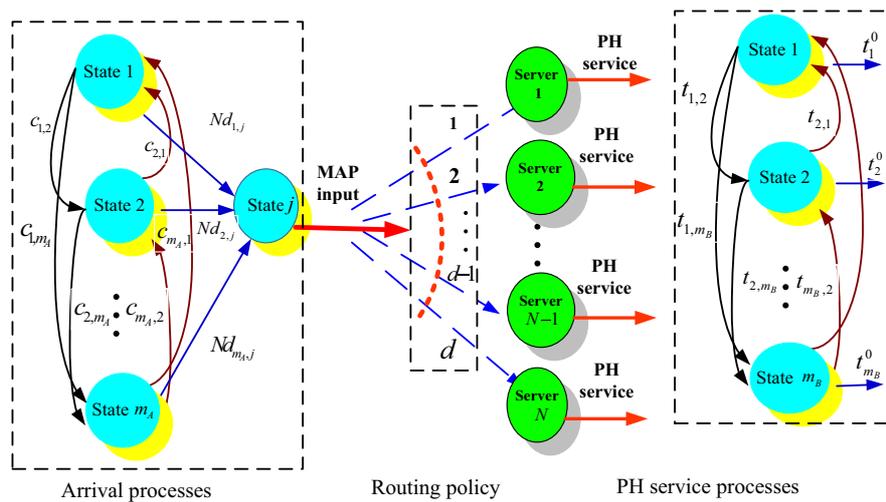


Fig. 1 The supermarket model with MAP inputs and PH service times

contain many useful distributions such as exponential, hyper-exponential and Erlang distributions; while the MAPs include, for example, Poisson process, PH-renewal processes, and Markovian modulated Poisson processes (MMPPs). Note that the probability distributions and stochastic point processes have extensively been used in most practical stochastic modeling. On the other hand, in many practical applications, the block-structured supermarket model is an important queueing model to analyze the relation between the system performance and the job routing rule, and it can also help to design reasonable architecture to improve the performance and to balance the load.

3 An infinite-dimensional Markov process

In this section, we first define the random fraction vector of this supermarket model. Then we use the random fraction vector to construct an infinite-dimensional Markov process, which describes the state of this supermarket model.

For this supermarket model, let $n_{k;i,j}^{(N)}(t)$ be the number of servers with at least k customers (note that the serving customer is also taken into account), and with the MAP be in phase i and the PH service time be in phase j at time $t \geq 0$. Clearly, $0 \leq n_{0;i}^{(N)}(t) \leq N$ and $0 \leq n_{k;i,j}^{(N)}(t) \leq N$ for $k \geq 1$, $1 \leq i \leq m_A$ and $1 \leq j \leq m_B$. Let

$$U_{0;i}^{(N)}(t) = \frac{n_{0;i}^{(N)}(t)}{N}, \quad 1 \leq i \leq m_A,$$

and for $k \geq 1$

$$U_{k;i,j}^{(N)}(t) = \frac{n_{k;i,j}^{(N)}(t)}{N}, \quad 1 \leq i \leq m_A, 1 \leq j \leq m_B.$$

Then $U_{k;i,j}^{(N)}(t)$ is the fraction of servers with at least k customers, and with the MAP be in phase i and the PH service time be in phase j at time t . Using the lexicographic order we write

$$U_0^{(N)}(t) = \left(U_{0;1}^{(N)}(t), U_{0;2}^{(N)}(t), \dots, U_{0;m_A}^{(N)}(t) \right)$$

for $k \geq 1$

$$U_k^{(N)}(t) = \left(U_{k;1,1}^{(N)}(t), U_{k;1,2}^{(N)}(t), \dots, U_{k;1,m_B}^{(N)}(t); \dots; \right. \\ \left. U_{k;m_A,1}^{(N)}(t), U_{k;m_A,2}^{(N)}(t), \dots, U_{k;m_A,m_B}^{(N)}(t) \right),$$

and

$$U^{(N)}(t) = \left(U_0^{(N)}(t), U_1^{(N)}(t), U_2^{(N)}(t), \dots \right). \quad (1)$$

Let $a = (a_1, a_2, a_3, \dots)$ and $b = (b_1, b_2, b_3, \dots)$. We write $a < b$ if $a_k < b_k$ for some $k \geq 1$; $a \leq b$ if $a_k \leq b_k$ for every $k \geq 1$.

For a fixed quaternary array (t, N, i, j) with $t \geq 0$, $N \in \{1, 2, 3, \dots\}$, $i \in \{1, 2, \dots, m_A\}$ and $j \in \{1, 2, \dots, m_B\}$, it is easy to see from the stochastic order that $n_{k;i,j}^{(N)}(t) \geq n_{k+1;i,j}^{(N)}(t)$ for $k \geq 1$. This gives

$$U_1^{(N)}(t) \geq U_2^{(N)}(t) \geq U_3^{(N)}(t) \cdots \geq 0 \quad (2)$$

and

$$1 = U_0^{(N)}(t) e \geq U_1^{(N)}(t) e \geq U_2^{(N)}(t) e \geq U_3^{(N)}(t) e \geq \cdots \geq 0. \quad (3)$$

Note that the state of this supermarket model is described as the random fraction vector $U^{(N)}(t)$ for $t \geq 0$, and $\{U^{(N)}(t), t \geq 0\}$ is a stochastic vector process for each $N = 1, 2, \dots$. Since the arrival process to this supermarket model is the MAP and the service times in each server are of phase type, $\{U^{(N)}(t), t \geq 0\}$ is an infinite-dimensional Markov process whose state space is given by

$$\begin{aligned} \tilde{\Omega}_N = \{ & (h_0^{(N)}, h_1^{(N)}, h_2^{(N)} \dots) : h_0^{(N)} \text{ is a probability vector of size } m_A, \\ & h_1^{(N)} \geq h_2^{(N)} \geq h_3^{(N)} \geq \dots \geq 0, h_k^{(N)} \text{ is a row vector of size } m_A m_B \text{ for } k \geq 1, \\ & 1 = h_0^{(N)} e \geq h_1^{(N)} e \geq h_2^{(N)} e \geq \dots \geq 0, \\ & \text{and } N h_k^{(N)} \text{ is a row vector of nonnegative integers for } k \geq 0 \}, \end{aligned} \tag{4}$$

We write

$$u_{0;i}^{(N)}(t) = E[U_{0;i}^{(N)}(t)]$$

and for $k \geq 1$

$$u_{k;i,j}^{(N)}(t) = E[U_{k;i,j}^{(N)}(t)].$$

Using the lexicographic order we write

$$u_0^{(N)}(t) = (u_{0;1}^{(N)}(t), u_{0;2}^{(N)}(t), \dots, u_{0;m_A}^{(N)}(t))$$

and for $k \geq 1$

$$\begin{aligned} u_k^{(N)}(t) = & (u_{k;1,1}^{(N)}(t), u_{k;1,2}^{(N)}(t), \dots, u_{k;1,m_B}^{(N)}(t); \dots; \\ & u_{k;m_A,1}^{(N)}(t), u_{k;m_A,2}^{(N)}(t), \dots, u_{k;m_A,m_B}^{(N)}(t)), \\ \mathbf{u}^{(N)}(t) = & (u_0^{(N)}(t), u_1^{(N)}(t), u_2^{(N)}(t), \dots). \end{aligned}$$

It is easy to see from Eqs. 2 and 3 that

$$u_1^{(N)}(t) \geq u_2^{(N)}(t) \geq u_3^{(N)}(t) \dots \geq 0 \tag{5}$$

and

$$1 = u_0^{(N)}(t) e \geq u_1^{(N)}(t) e \geq u_2^{(N)}(t) e \geq \dots \geq 0. \tag{6}$$

In the remainder of this section, for convenience of readers, it is necessary to explain the structure of this long paper which is outlined as follows. *Part one: The limit of the sequence of Markov processes.* It is seen from Eqs.(1) and (4) that we need to deal with the limit of the sequence $\{U^{(N)}(t)\}$ of infinite-dimensional Markov processes. This is organized in Appendix B (see <http://arxiv.org/pdf/1406.0285.pdf>) by means of the convergence theorems of operator semigroups, e.g., see Ethier and Kurtz (1986) for more details. *Part two: The existence and uniqueness of the solution.* As seen from Theorem 2 and Eq. (27), we need to study the two vectors $\mathbf{u}^{(N)}(t)$ and $\mathbf{u}(t) = \lim_{N \rightarrow \infty} \mathbf{u}^{(N)}(t)$. To that end, Section 4 sets up the system of differential vector equations satisfied by $\mathbf{u}^{(N)}(t)$, while Section 5 provides a unified matrix-differential algorithm for establishing the Lipschitz condition. *Part three: Computation of the fixed point and performance analysis.* Section 6 provides a generalized matrix-analytic method for computing the fixed point. Section 7 analyzes the performance measures of this supermarket model by means of some numerical examples.

4 The system of differential vector equations

In this section, we set up an infinite-dimensional system of differential vector equations satisfied by the expected fraction vector through a detailed probability analysis. Specifically, we obtain an important result: The invariance of environment factors, which is a key in rewriting the differential equations as a simple vector form.

To derive the system of differential vector equations, we first discuss an example with the number $k \geq 2$ of customers through the following three steps:

Step one: Analysis of the Arrival Processes In this supermarket model of N identical servers, we need to determine the change in the expected number of servers with at least k customers over a small time period $[0, dt)$. When the MAP environment process $\{J(t) : t \geq 0\}$ jumps from state l to state i for $1 \leq l, i \leq m_A$ and the PH service environment process $\{I(t) : t \geq 0\}$ sojourns at state j for $1 \leq j \leq m_B$, one arrival occurs in a small time period $[0, dt)$. In this case, the rate that any arriving customer selects d servers with at least $k - 1$ customers at random and joins the shortest one with $k - 1$ customers, is given by

$$\sum_{l=1}^{m_A} \left[u_{k-1;l,j}^{(N)}(t) d_{l,i} - u_{k;i,j}^{(N)}(t) (d_{i,1}, d_{i,2}, \dots, d_{i,m_A}) e \right] \times L_{k;l}^{(N)}(u_{k-1}(t), u_k(t)) N dt, \tag{7}$$

where

$$\begin{aligned} L_{k;l}^{(N)}(u_{k-1}(t), u_k(t)) &= \sum_{m=1}^d C_d^m \left\{ \sum_{j=1}^{m_B} [u_{k-1;l,j}^{(N)}(t) - u_{k;l,j}^{(N)}(t)] \right\}^{m-1} \left\{ \sum_{j=1}^{m_B} [u_{k;l,j}^{(N)}(t)] \right\}^{d-m} \\ &\quad + \sum_{m=1}^{d-1} C_d^m \left\{ \sum_{j=1}^{m_B} [u_{k-1;l,j}^{(N)}(t) - u_{k;l,j}^{(N)}(t)] \right\}^{m-1} \sum_{\substack{r_1+r_2+\dots+r_{m_A}=d-m \\ \sum_{i \neq l}^{m_A} r_i \geq 1 \\ 0 \leq r_j \leq d-m, 1 \leq j \leq m_A}} \\ &\quad \times \binom{d-m}{r_1, r_2, \dots, r_{m_A}} \prod_{i=1}^{m_A} \left\{ \sum_{j=1}^{m_B} [u_{k;i,j}^{(N)}(t)] \right\}^{r_i} + \sum_{m=2}^d C_d^m \sum_{m_1=1}^{m-1} \frac{m_1}{m} \\ &\quad \times C_{m_1}^{m_1} \left\{ \sum_{j=1}^{m_B} [u_{k-1;l,j}^{(N)}(t) - u_{k;l,j}^{(N)}(t)] \right\}^{m_1-1} \sum_{\substack{n_1+n_2+\dots+n_{m_A}=m-m_1 \\ \sum_{i \neq l}^{m_A} n_i \geq 1 \\ 0 \leq n_j \leq m-m_1, 1 \leq j \leq m_A}} \\ &\quad \times \binom{m-m_1}{n_1, n_2, \dots, n_{m_A}} \prod_{i=1}^{m_A} \left\{ \sum_{j=1}^{m_B} [u_{k-1;i,j}^{(N)}(t) - u_{k;i,j}^{(N)}(t)] \right\}^{n_i} \\ &\quad \times \sum_{\substack{r_1+r_2+\dots+r_{m_A}=d-m \\ 0 \leq r_j \leq d-m, 1 \leq j \leq m_A}} \binom{d-m}{r_1, r_2, \dots, r_{m_A}} \prod_{i=1}^{m_A} \left\{ \sum_{j=1}^{m_B} [u_{k;i,j}^{(N)}(t)] \right\}^{r_i}. \end{aligned} \tag{8}$$

Note that $\left[u_{k-1;l,j}^{(N)}(t) d_{l,i} - u_{k;i,j}^{(N)}(t) (d_{i,1}, d_{i,2}, \dots, d_{i,m_A}) e \right]$ is the rate that any arriving customer joins one server with the shortest queue length $k - 1$, where the MAP goes to phase i from phase l , and the PH service time is in phase j .

Now, we provide a detailed interpretation for how to derive Eq. (8) through a set decomposition of all possible events given in Fig. 2, where each of the d selected servers has at least $k - 1$ customers, the MAP arrival environment is in phase i or l , and the PH service environment is in phase j . Hence, the probability that any arriving customer selects d servers with at least $k - 1$ customers at random and joins a server with the shortest queue length $k - 1$ and with the MAP phase i or l is determined by means of Fig. 2 through the following three parts:

Part I: The probability that any arriving customer joins a server with the shortest queue length $k - 1$ and with the MAP phase l , and the queue lengths of the other selected $d - 1$ servers are not shorter than $k - 1$, is given by

$$\sum_{m=1}^d C_d^m \left\{ \sum_{j=1}^{m_B} \left[u_{k-1;l,j}^{(N)}(t) - u_{k;l,j}^{(N)}(t) \right] \right\}^{m-1} \left\{ \sum_{j=1}^{m_B} \left[u_{k;l,j}^{(N)}(t) \right] \right\}^{d-m},$$

where $C_d^m = d! / [m!(d - m)!]$ is a binomial coefficient, and

$$\left\{ \sum_{j=1}^{m_B} \left[u_{k-1;l,j}^{(N)}(t) - u_{k;l,j}^{(N)}(t) \right] \right\}^{m-1}$$

is the probability that any arriving customer who can only choose one server makes $m - 1$ independent selections during the $m - 1$ servers with the queue length $k - 1$ and with the MAP phase l at time t ; while $\left\{ \sum_{j=1}^{m_B} \left[u_{k;l,j}^{(N)}(t) \right] \right\}^{d-m}$ is the probability that there are $d - m$ servers whose queue lengths are not shorter than k and with the MAP phase l .

Each of the d selected servers is at the MAP phase l , and there is at least one server with the shortest queue length $k-1$. (Part I)	
In the d selected servers, there is only one server with the shortest queue length $k-1$ and with the MAP phase l ; there exists at least one server is at the MAP phase $i \neq l$; and all the other $d-1$ selected servers contain no less than k customers. (Part II)	In the d selected servers with no less than $k-1$ customers, there is at least one server with the shortest queue length $k-1$, and with the MAP phase l ; there is also at least one server with the shortest queue length $k-1$ and with the MAP phase $i \neq l$. (Part III)

Fig. 2 A set decomposition of all possible events

Part II: The probability that any arriving customer joins a server with the shortest queue length $k-1$ and with the MAP phase l ; and the queue lengths of the other selected $d-1$ servers are not shorter than $k-1$, and there exists at least one server with no less than k customers and with the MAP phase $i \neq l$, is given by

$$\sum_{m=1}^{d-1} C_d^m \left\{ \sum_{j=1}^{m_B} \left[u_{k-1;l,j}^{(N)}(t) - u_{k;l,j}^{(N)}(t) \right] \right\}^{m-1} \sum_{\substack{r_1+r_2+\dots+r_{m_A}=d-m \\ \sum_{i \neq l}^{m_A} r_i \geq 1 \\ 0 \leq r_j \leq d-m, 1 \leq j \leq m_A}} \times \binom{d-m}{r_1, r_2, \dots, r_{m_A}} \prod_{i=1}^{m_A} \left\{ \sum_{j=1}^{m_B} \left[u_{k;i,j}^{(N)}(t) \right] \right\}^{r_i},$$

where when $r_1 + r_2 + \dots + r_{m_A} = n$, $\binom{n}{r_1, r_2, \dots, r_{m_A}} = \frac{n!}{\prod_{i=1}^{m_A} r_i!}$ is a multinomial coefficient.

Part III: If there are m selected servers with the shortest queue length $k-1$ where there are m_1 servers with the MAP phase l and $m - m_1$ servers with the MAP phases $i \neq l$, then the probability that any arriving customer joins a server with the shortest queue length $k-1$ and with the MAP phase l is equal to m_1/m . In this case, the probability that any arriving customer joins a server with the shortest queue length $k-1$ and with the MAP phase l , the queue lengths of the other selected $d-1$ servers are not shorter than $k-1$, is given by

$$\sum_{m=2}^d C_d^m \sum_{m_1=1}^{m-1} \frac{m_1}{m} C_m^{m_1} \left\{ \sum_{j=1}^{m_B} \left[u_{k-1;l,j}^{(N)}(t) - u_{k;l,j}^{(N)}(t) \right] \right\}^{m_1-1} \sum_{\substack{n_1+n_2+\dots+n_{m_A}=m-m_1 \\ \sum_{i \neq l}^{m_A} n_i \geq 1 \\ 0 \leq n_j \leq m-m_1, 1 \leq j \leq m_A}} \times \binom{m-m_1}{n_1, n_2, \dots, n_{m_A}} \prod_{i=1}^{m_A} \left\{ \sum_{j=1}^{m_B} \left[u_{k-1;i,j}^{(N)}(t) - u_{k;i,j}^{(N)}(t) \right] \right\}^{n_i} \\ \times \sum_{\substack{r_1+r_2+\dots+r_{m_A}=d-m \\ 0 \leq r_j \leq d-m, 1 \leq j \leq m_A}} \binom{d-m}{r_1, r_2, \dots, r_{m_A}} \prod_{i=1}^{m_A} \left\{ \sum_{j=1}^{m_B} \left[u_{k;i,j}^{(N)}(t) \right] \right\}^{r_i}.$$

Using the above three parts, Eqs. (7) and (8) can be obtained immediately.

The following theorem gives an important result, called *the invariance of environment factors*, which will play an important role in setting up the infinite-dimensional system of differential vector equations. This enables us to apply the generalized matrix-analytic method to the study of more general supermarket models with non-Poisson inputs and non-exponential service times.

Theorem 1

$$L_{1;l}^{(N)} \left(u_0^{(N)}(t) \otimes \alpha, u_1^{(N)}(t) \right) = \sum_{m=1}^d C_d^m \left[\sum_{l=1}^{m_A} \sum_{j=1}^{m_B} \left(u_{0;l}^{(N)}(t) \alpha_j - u_{1;l,j}^{(N)}(t) \right) \right]^{m-1} \times \left[\sum_{l=1}^{m_A} \sum_{j=1}^{m_B} u_{1;l,j}^{(N)}(t) \right]^{d-m} \tag{9}$$

and for $k \geq 2$

$$L_{k;l}^{(N)} \left(u_{k-1}^{(N)}(t), u_k^{(N)}(t) \right) = \sum_{m=1}^d C_d^m \left[\sum_{l=1}^{m_A} \sum_{j=1}^{m_B} \left(u_{k-1;l,j}^{(N)}(t) - u_{k;l,j}^{(N)}(t) \right) \right]^{m-1} \times \left[\sum_{l=1}^{m_A} \sum_{j=1}^{m_B} u_{k;l,j}^{(N)}(t) \right]^{d-m} . \tag{10}$$

Thus $L_{1;l}^{(N)} \left(u_0^{(N)}(t) \otimes \alpha, u_1^{(N)}(t) \right)$ and $L_{k;l}^{(N)} \left(u_{k-1}^{(N)}(t), u_k^{(N)}(t) \right)$ for $k \geq 2$ are independent of the MAP phase $l \in \{1, 2, \dots, m_A\}$. In this case, we have

$$L_{1;l}^{(N)} \left(u_0^{(N)}(t) \otimes \alpha, u_1^{(N)}(t) \right) \stackrel{\text{def}}{=} L_1^{(N)} \left(u_0^{(N)}(t) \otimes \alpha, u_1^{(N)}(t) \right) \tag{11}$$

and for $k \geq 2$

$$L_{k;l}^{(N)} \left(u_{k-1}^{(N)}(t), u_k^{(N)}(t) \right) \stackrel{\text{def}}{=} L_k^{(N)} \left(u_{k-1}^{(N)}(t), u_k^{(N)}(t) \right) . \tag{12}$$

Proof See Appendix A (<http://arxiv.org/pdf/1406.0285.pdf>). □

For any two matrices $A = (a_{i,j})$ and $B = (b_{i,j})$, their Kronecker product is defined as $A \otimes B = (a_{i,j} b_{i,j})$, and their Kronecker sum is given by $A \oplus B = A \otimes I + I \otimes B$.

It is seen from the invariance of environment factors in Theorem 1 that Eq. 7 is rewritten as, in a vector form,

$$\left\{ u_{k-1}^{(N)}(t) (D \otimes I) - u_k^{(N)}(t) [\text{diag}(De) \otimes I] \right\} \times L_k^{(N)} \left(u_{k-1}^{(N)}(t), u_k^{(N)}(t) \right) N dt. \tag{13}$$

Note that $L_1^{(N)} \left(u_0^{(N)}(t) \otimes \alpha, u_1^{(N)}(t) \right)$ and $L_k^{(N)} \left(u_{k-1}^{(N)}(t), u_k^{(N)}(t) \right)$ are scale for $k \geq 2$.

Step two: Analysis of the Environment State Transitions in the MAP When there are at least k customers in the server, the rate that the MAP environment process jumps from state l to state i with rate $c_{l,j}$, and no arrival of the MAP occurs during a small time period $[0, dt)$, is given by

$$\left[\sum_{l=1}^{m_A} u_{k;l,j}^{(N)}(t) c_{l,i} + u_{k,i,j}^{(N)}(t) (d_{i,1}, d_{i,2}, \dots, d_{i,m_A}) e \right] N dt.$$

This gives, in a vector form,

$$u_k^{(N)}(t) \left([C + \text{diag}(De)] \otimes I \right) N dt. \tag{14}$$

Step three: Analysis of the Service Processes To analyze the PH service process, we need to consider the following two cases:

Case one: One service completion occurs with rate t_i^0 during a small time period $[0, dt)$. In this case, when there are at least $k + 1$ customers in the server, the rate that a customer is completed its service with entering PH phase j and the MAP is in phase i is given by

$$\left[u_{k+1;i,1}^{(N)}(t)t_1^0\alpha_j + u_{k+1;i,2}^{(N)}(t)t_2^0\alpha_j + \cdots + u_{k+1;i,m_B}^{(N)}(t)t_{m_B}^0\alpha_j \right] Ndt.$$

Case two: No service completion occurs during a small time period $[0, dt)$, but the MAP is in phase i and the PH service environment process goes to phase j . Thus, when there are at least k customers in the server, the rate of this case is given by

$$\left[u_{k;i,1}^{(N)}(t)t_{1,j} + u_{k;i,2}^{(N)}(t)t_{2,j} + u_{k;i,3}^{(N)}(t)t_{3,j} + \cdots + u_{k;i,m_B}^{(N)}(t)t_{m_B,j} \right] Ndt.$$

Thus, for the PH service process, we obtain that in a vector form,

$$\left[u_k^{(N)}(t) (I \otimes T) + u_{k+1}^{(N)}(t) (I \otimes T^0\alpha) \right] Ndt \quad (15)$$

Let

$$n_k^{(N)}(t) = \left(n_{k;1,1}^{(N)}(t), n_{k;1,2}^{(N)}(t), \dots, n_{k;1,m_B}^{(N)}(t); \dots; n_{k;m_A,1}^{(N)}(t), n_{k;m_A,2}^{(N)}(t), \dots, n_{k;m_A,m_B}^{(N)}(t) \right).$$

Then it follows from Eq. 13 to Eq.15 that

$$dE \left[n_k^{(N)}(t) \right] = \left\{ \left[u_{k-1}^{(N)}(t) (D \otimes I) - u_k^{(N)}(t) [\text{diag}(De) \otimes I] \right] L_k^{(N)} \left(u_{k-1}^{(N)}(t), u_k^{(N)}(t) \right) + u_k^{(N)}(t) \left\{ [C + \text{diag}(De)] \oplus T \right\} + u_{k+1}^{(N)}(t) (I \otimes T^0\alpha) \right\} Ndt.$$

Since $E \left[n_k^{(N)}(t)/N \right] = u_k^{(N)}(t)$ and $A \otimes I + I \otimes B = A \oplus B$, we obtain

$$\frac{du_k^{(N)}(t)}{dt} = \left\{ u_{k-1}^{(N)}(t) (D \otimes I) - u_k^{(N)}(t) [\text{diag}(De) \otimes I] \right\} L_k^{(N)} \left(u_{k-1}^{(N)}(t), u_k^{(N)}(t) \right) + u_k^{(N)}(t) \left\{ [C + \text{diag}(De)] \oplus T \right\} + u_{k+1}^{(N)}(t) (I \otimes T^0\alpha). \quad (16)$$

Using a similar analysis to Eq. 16, we obtain an infinite-dimensional system of differential vector equations satisfied by the expected fraction vector $\mathbf{u}^{(N)}(t)$ as follows:

$$\frac{du_1^{(N)}(t)}{dt} = \left\{ \left[u_0^{(N)}(t) \otimes \alpha \right] (D \otimes I) - u_1^{(N)}(t) [\text{diag}(De) \otimes I] \right\} L_1^{(N)} \left(u_0^{(N)}(t) \otimes \alpha, u_1^{(N)}(t) \right) + u_1^{(N)}(t) \left\{ [C + \text{diag}(De)] \oplus T \right\} + u_2^{(N)}(t) (I \otimes T^0\alpha), \quad (17)$$

and for $k \geq 2$

$$\frac{du_k^{(N)}(t)}{dt} = \left\{ u_{k-1}^{(N)}(t) (D \otimes I) - u_k^{(N)}(t) [\text{diag}(De) \otimes I] \right\} L_k^{(N)} \left(u_{k-1}^{(N)}(t), u_k^{(N)}(t) \right) + u_k^{(N)}(t) \left\{ [C + \text{diag}(De)] \oplus T \right\} + u_{k+1}^{(N)}(t) (I \otimes T^0\alpha), \quad (18)$$

with the boundary condition

$$\frac{du_0^{(N)}(t)}{dt} = u_0^{(N)}(t) (C + D), \tag{19}$$

$$u_0^{(N)}(t)e = 1; \tag{20}$$

and with the initial condition

$$u_k^{(N)}(0) = g_k, \quad k \geq 1, \tag{21}$$

where

$$g_1 \geq g_2 \geq g_3 \geq \dots \geq 0$$

and

$$1 = g_0e \geq g_1e \geq g_2e \geq \dots \geq 0.$$

Remark 2 It is necessary to explain some probability setting for the invariance of environment factors. It follows from Theorem 1 that

$$L_1^{(N)}(u_0^{(N)}(t) \otimes \alpha, u_1^{(N)}(t)) = \frac{[u_0^{(N)}(t)e]^d - [u_1^{(N)}(t)e]^d}{u_0^{(N)}(t)e - u_1^{(N)}(t)e}$$

and for $k \geq 2$

$$L_k^{(N)}(u_{k-1}^{(N)}(t), u_k^{(N)}(t)) = \frac{[u_{k-1}^{(N)}(t)e]^d - [u_k^{(N)}(t)e]^d}{u_{k-1}^{(N)}(t)e - u_k^{(N)}(t)e}.$$

Note that the two expressions will be useful in our later study, for example, establishing the Lipschitz condition, and computing the fixed point. Specifically, for $d = 1$ we have

$$L_1^{(N)}(u_0^{(N)}(t) \otimes \alpha, u_1^{(N)}(t)) = 1$$

and for $k \geq 2$

$$L_k^{(N)}(u_{k-1}^{(N)}(t), u_k^{(N)}(t)) = 1.$$

For $d = 2$ we have

$$L_1^{(N)}(u_0^{(N)}(t) \otimes \alpha, u_1^{(N)}(t)) = u_0^{(N)}(t)e + u_1^{(N)}(t)e > 1$$

and for $k \geq 2$

$$L_k^{(N)}(u_{k-1}^{(N)}(t), u_k^{(N)}(t)) = u_{k-1}^{(N)}(t)e + u_k^{(N)}(t)e.$$

This shows that $(L_1^{(N)}(u_0^{(N)}(t) \otimes \alpha, u_1^{(N)}(t)), L_2^{(N)}(u_1^{(N)}(t), u_2^{(N)}(t)), \dots)$ is not a probability vector.

5 The Lipschitz condition

In this section, we show that the mean-field limit of the sequence of Markov processes asymptotically approaches a single trajectory identified by the unique and global solution to the infinite-dimensional system of limiting differential vector equations. To that end, we provide a unified matrix-differential algorithm for establishing the Lipschitz condition, which is a key in proving the existence and uniqueness of the solution by means of the Picard approximation according to the basic results of the Banach space.

Let $\mathbf{T}_N(t)$ be the operator semigroup of the Markov process $\{\mathbf{U}^{(N)}(t), t \geq 0\}$. If $f : \Omega_N \rightarrow \mathbf{C}^1$, where $\Omega_N = \{\mathbf{g} \in \tilde{\Omega}_N : \mathbf{g}e < +\infty\}$, then for $\mathbf{g} \in \Omega_N$ and $t \geq 0$

$$\mathbf{T}_N(t)f(\mathbf{g}) = E[f(\mathbf{U}_N(t) | \mathbf{U}_N(0) = \mathbf{g})].$$

We denote by \mathbf{A}_N the generating operator of the operator semigroup $\mathbf{T}_N(t)$, it is easy to see that $\mathbf{T}_N(t) = \exp\{\mathbf{A}_N t\}$ for $t \geq 0$.

In Appendix B (see <http://arxiv.org/pdf/1406.0285.pdf>), we will provide a detailed analysis for the limiting behavior of the sequence $\{\mathbf{U}^{(N)}(t), t \geq 0\}$ of Markov processes for $N = 1, 2, 3, \dots$, where two formal limits for the sequence $\{\mathbf{A}_N\}$ of generating operators and for the sequence $\{\mathbf{T}_N(t)\}$ of operator semigroups are expressed as $\mathbf{A} = \lim_{N \rightarrow \infty} \mathbf{A}_N$ and $\mathbf{T}(t) = \lim_{N \rightarrow \infty} \mathbf{T}_N(t)$ for $t \geq 0$, respectively.

We write

$$L_1(u_0(t) \otimes \alpha, u_1(t)) = \sum_{m=1}^d C_d^m \left[\sum_{l=1}^{m_A} \sum_{j=1}^{m_B} (u_{0,l}(t) \alpha_j - u_{1;l,j}(t)) \right]^{m-1} \\ \times \left[\sum_{l=1}^{m_A} \sum_{j=1}^{m_B} u_{1;l,j}(t) \right]^{d-m},$$

for $k \geq 2$

$$L_k(u_{k-1}(t), u_k(t)) = \sum_{m=1}^d C_d^m \left[\sum_{l=1}^{m_A} \sum_{j=1}^{m_B} (u_{k-1;l,j}(t) - u_{k;l,j}(t)) \right]^{m-1} \\ \times \left[\sum_{l=1}^{m_A} \sum_{j=1}^{m_B} u_{k;l,j}(t) \right]^{d-m}.$$

Let $\mathbf{u}(t) = \lim_{N \rightarrow \infty} \mathbf{u}^{(N)}(t)$ where $u_k(t) = \lim_{N \rightarrow \infty} u_k^{(N)}(t)$ for $k \geq 0$ and $t \geq 0$. Based on the limiting operator semigroup $\mathbf{T}(t)$ or the limiting generating operator \mathbf{A} , as $N \rightarrow \infty$ it follows from Eqs. 17 to 21 that $\mathbf{u}(t)$ is a solution to the system of differential vector equations as follows:

$$\frac{du_1(t)}{dt} = \{[u_0(t) \otimes \alpha](D \otimes I) - u_1(t)[\text{diag}(De) \otimes I]\} L_1(u_0(t) \otimes \alpha, u_1(t)) \\ + u_1(t) \{[C + \text{diag}(De)] \oplus T\} + u_2(t) (I \otimes T^0 \alpha), \quad (22)$$

and for $k \geq 2$

$$\frac{du_k(t)}{dt} = \{u_{k-1}(t)(D \otimes I) - u_k(t)[\text{diag}(De) \otimes I]\} L_k(u_{k-1}(t), u_k(t)) \\ + u_k(t) \{[C + \text{diag}(De)] \oplus T\} + u_{k+1}(t) (I \otimes T^0 \alpha), \quad (23)$$

with the boundary condition

$$u_0^{(N)}(t) = u_0^{(N)}(0) \exp\{(C + D)t\}, \quad (24)$$

$$u_0^{(N)}(t)e = 1, \quad (25)$$

and with initial condition

$$u_k(0) = g_k, \quad k \geq 0. \quad (26)$$

Based on the solution $\mathbf{u}(t, \mathbf{g})$ to the system of differential vector Eqs. (22) to (26), we define a mapping: $\mathbf{g} \rightarrow \mathbf{u}(t, \mathbf{g})$. Note that the operator semigroup $\mathbf{T}(t)$ acts in the space L , where $L = C(\tilde{\Omega})$ is the Banach space of continuous functions $f : \tilde{\Omega} \rightarrow \mathbf{R}$ with uniform metric $\|f\| = \max_{u \in \tilde{\Omega}} |f(u)|$, and

$$\tilde{\Omega} = \{\mathbf{u} : u_1 \geq u_2 \geq u_3 \geq \dots \geq 0; 1 = u_0^{(N)}e \geq u_1^{(N)}e \geq u_2^{(N)}e \geq \dots \geq 0\}$$

for the vector $\mathbf{u} = (u_0, u_1, u_2, \dots)$ with u_0 be a probability vector of size m_A and the size of the row vector u_k be $m_A m_B$ for $k \geq 1$. If $f \in L$ and $\mathbf{g} \in \tilde{\Omega}$, then

$$\mathbf{T}(t)f(\mathbf{g}) = f(\mathbf{u}(t, \mathbf{g})).$$

The following theorem uses the operator semigroup to provide the mean-field limit in this supermarket model. Note that the mean-field limit shows that there always exists the limiting process $\{U(t), t \geq 0\}$ of the sequence $\{U^{(N)}(t), t \geq 0\}$ of Markov processes, and also indicates the asymptotic independence of the block-structured queueing processes in this supermarket model.

Theorem 2 For any continuous function $f : \Omega \rightarrow \mathbf{R}$ and $t > 0$,

$$\lim_{N \rightarrow \infty} \sup_{\mathbf{g} \in \Omega} |\mathbf{T}_N(t)f(\mathbf{g}) - f(\mathbf{u}(t; \mathbf{g}))| = 0,$$

and the convergence is uniform in t with any bounded interval.

Proof See Appendix B (<http://arxiv.org/pdf/1406.0285.pdf>). □

Finally, we provide some interpretation on Theorem 2. If $\lim_{N \rightarrow \infty} U^{(N)}(0) = \mathbf{u}(0) = \mathbf{g} \in \Omega$ in probability, then Theorem 2 shows that $U(t) = \lim_{N \rightarrow \infty} U^{(N)}(t)$ is concentrated on the trajectory $\Gamma_{\mathbf{g}} = \{\mathbf{u}(t, \mathbf{g}) : t \geq 0\}$. This indicates the functional strong law of large numbers for the time evolution of the fraction of each state of this supermarket model, thus the sequence $\{U^{(N)}(t), t \geq 0\}$ of Markov processes converges weakly to the expected fraction vector $\mathbf{u}(t, \mathbf{g})$ as $N \rightarrow \infty$, that is, for any $T > 0$

$$\lim_{N \rightarrow \infty} \sup_{0 \leq s \leq T} \|U^{(N)}(s) - \mathbf{u}(s, \mathbf{g})\| = 0 \text{ in probability.} \tag{27}$$

In the remainder of this section, we provide a unified matrix-differential algorithm for establishing a Lipschitz condition for the expected fraction vector $f : \mathbf{R}_+^\infty \rightarrow \mathbf{C}^1(\mathbf{R}_+^\infty)$. The Lipschitz condition is a key for proving the existence and uniqueness of solution to the infinite-dimensional system of limiting differential vector Eqs. 22 to 26. On the other hand, the proof of the existence and uniqueness of solution is standard by means of the Picard approximation according to the basic results of the Banach space. Readers may refer to Li et al. (2013) for more details.

To provide the Lipschitz condition, we need to use the derivative of the infinite-dimensional vector $G : \mathbf{R}_+^\infty \rightarrow \mathbf{C}^1(\mathbf{R}_+^\infty)$. Thus we first provide some definitions and preliminaries for such derivatives as follows.

For the infinite-dimensional vector $G : \mathbf{R}_+^\infty \rightarrow \mathbf{C}^1(\mathbf{R}_+^\infty)$, we write $x = (x_1, x_2, x_3, \dots)$ and $G(x) = (G_1(x), G_2(x), G_3(x), \dots)$, where x_k and $G_k(x)$ are scalar for $k \geq 1$. Then the matrix of partial derivatives of the infinite-dimensional vector $G(x)$ is defined as

$$\mathcal{D}G(x) = \frac{\partial G(x)}{\partial x} = \begin{pmatrix} \frac{\partial G_1(x)}{\partial x_1} & \frac{\partial G_2(x)}{\partial x_1} & \frac{\partial G_3(x)}{\partial x_1} & \dots \\ \frac{\partial G_1(x)}{\partial x_2} & \frac{\partial G_2(x)}{\partial x_2} & \frac{\partial G_3(x)}{\partial x_2} & \dots \\ \frac{\partial G_1(x)}{\partial x_3} & \frac{\partial G_2(x)}{\partial x_3} & \frac{\partial G_3(x)}{\partial x_3} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \quad (28)$$

if each of the partial derivatives exists.

For the infinite-dimensional vector $G : \mathbf{R}_+^\infty \rightarrow \mathbf{C}^1(\mathbf{R}_+^\infty)$, if there exists a linear operator $A : \mathbf{R}_+^\infty \rightarrow \mathbf{C}^1(\mathbf{R}_+^\infty)$ such that for any vector $h \in \mathbf{R}_+^\infty$ and a scalar $t \in \mathbf{R}$

$$\lim_{t \rightarrow 0} \frac{\|G(x + th) - G(x) - thA\|}{t} = 0,$$

then the function $G(x)$ is called to be Gateaux differentiable at $x \in \mathbf{R}_+^\infty$. In this case, we

write the Gateaux derivative $A = \mathcal{D}G(x) = \frac{\partial G(x)}{\partial x}$.

Let $t = (t_1, t_2, t_3, \dots)$ with $0 \leq t_k \leq 1$ for $k \geq 1$. Then we write

$$\begin{aligned} & \mathcal{D}G(x + t\mathcal{O}(y-x)) \\ &= \begin{pmatrix} \frac{\partial G_1(x + t_1(y-x))}{\partial x_1} & \frac{\partial G_2(x + t_2(y-x))}{\partial x_1} & \frac{\partial G_3(x + t_3(y-x))}{\partial x_1} & \dots \\ \frac{\partial G_1(x + t_1(y-x))}{\partial x_2} & \frac{\partial G_2(x + t_2(y-x))}{\partial x_2} & \frac{\partial G_3(x + t_3(y-x))}{\partial x_2} & \dots \\ \frac{\partial G_1(x + t_1(y-x))}{\partial x_3} & \frac{\partial G_2(x + t_2(y-x))}{\partial x_3} & \frac{\partial G_3(x + t_3(y-x))}{\partial x_3} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \end{aligned}$$

If the infinite-dimensional vector $G : \mathbf{R}_+^\infty \rightarrow \mathbf{C}^1(\mathbf{R}_+^\infty)$ is Gateaux differentiable, then there exists a vector $t = (t_1, t_2, t_3, \dots)$ with $0 \leq t_k \leq 1$ for $k \geq 1$ such that

$$G(y) - G(x) = (y-x)\mathcal{D}G(x + t\mathcal{O}(y-x)). \quad (29)$$

Furthermore, we have

$$\|G(y) - G(x)\| \leq \sup_{0 \leq t \leq 1} \|\mathcal{D}G(x + t(y-x))\| \|y-x\|. \quad (30)$$

For convenience of description, Eqs. 22 to 26 are rewritten as an initial value problem as follows:

$$\begin{aligned} \frac{d}{dt} u_1 &= \{(u_0 \otimes \alpha)(D \otimes I) - u_1[\text{diag}(De) \otimes I]\} L_1(u_0 \otimes \alpha, u_1) \\ &+ u_1 \{[C + \text{diag}(De)] \oplus T\} + u_2(I \otimes T^0 \alpha) \end{aligned} \quad (31)$$

Property one

$$\frac{\partial x_k}{\partial x_k} = I, \quad \frac{\partial x_k S}{\partial x_k} = S,$$

where S is a matrix of size $m_A m_B$.

Note that

$$L_1(u_0 \otimes \alpha, x_1) = \frac{(u_0 e)^d - (x_1 e)^d}{u_0 e - x_1 e} = \frac{1 - (x_1 e)^d}{1 - x_1 e}$$

and for $k \geq 2$

$$L_k(x_{k-1}, x_k) = \frac{(x_{k-1} e)^d - (x_k e)^d}{x_{k-1} e - x_k e}.$$

Let $y_1 = x_1 e$. Then

$$\begin{aligned} \frac{\partial L_1(u_0 \otimes \alpha, x_1)}{\partial x_1} &= \frac{\partial y_1}{\partial x_1} \frac{\partial L_1(u_0 \otimes \alpha, x_1)}{\partial y_1} \\ &= e \frac{[(u_0 e)^d - (x_1 e)^d] - d(x_1 e)^{d-1}(u_0 e - x_1 e)}{(u_0 e - x_1 e)^2}. \end{aligned}$$

Similarly, for $k \geq 2$ we can obtain

$$\frac{\partial L_k(x_{k-1}, x_k)}{\partial x_{k-1}} = e \frac{d(x_{k-1} e)^{d-1}(x_{k-1} e - x_k e) - [(x_{k-1} e)^d - (x_k e)^d]}{(x_{k-1} e - x_k e)^2}$$

and

$$\frac{\partial L_k(x_{k-1}, x_k)}{\partial x_k} = e \frac{[(x_{k-1} e)^d - (x_k e)^d] - d(x_k e)^{d-1}(x_{k-1} e - x_k e)}{(x_{k-1} e - x_k e)^2}.$$

It is easy to check that

$$\begin{aligned} A_1(x) &= [C + \text{diag}(De)] \oplus T + [\text{diag}(De) \otimes I] \frac{(u_0 e)^d - (x_1 e)^d}{u_0 e - x_1 e} \\ &\quad + e x_1 [\text{diag}(De) \otimes I] \frac{[(u_0 e)^d - (x_1 e)^d] - d(x_1 e)^{d-1}(u_0 e - x_1 e)}{(u_0 e - x_1 e)^2}, \end{aligned} \quad (39)$$

$$\begin{aligned} B_1(x) &= (D \otimes I) \frac{(x_1 e)^d - (x_2 e)^d}{x_1 e - x_2 e} + e \{x_1 (D \otimes I) - x_2 [\text{diag}(De) \otimes I]\} \\ &\quad \times \frac{d(x_1 e)^{d-1}(x_1 e - x_2 e) - [(x_1 e)^d - (x_2 e)^d]}{(x_1 e - x_2 e)^2}; \end{aligned} \quad (40)$$

and for $k \geq 2$

$$C_k(x) = I \otimes T^0 \alpha, \quad (41)$$

$$\begin{aligned} B_k(x) &= (D \otimes I) \frac{(x_k e)^d - (x_{k+1} e)^d}{x_k e - x_{k+1} e} + e \{x_k (D \otimes I) - x_{k+1} [\text{diag}(De) \otimes I]\} \\ &\quad \times \frac{d(x_k e)^{d-1}(x_k e - x_{k+1} e) - [(x_k e)^d - (x_{k+1} e)^d]}{(x_k e - x_{k+1} e)^2}, \end{aligned} \quad (42)$$

$$\begin{aligned} A_k(x) &= [C + \text{diag}(De)] \oplus T + [\text{diag}(De) \otimes I] \frac{(x_{k-1} e)^d - (x_k e)^d}{x_{k-1} e - x_k e} \\ &\quad + e \{x_{k-1} (D \otimes I) - x_k [\text{diag}(De) \otimes I]\} \\ &\quad \times \frac{[(x_{k-1} e)^d - (x_k e)^d] - d(x_k e)^{d-1}(x_{k-1} e - x_k e)}{(x_{k-1} e - x_k e)^2}. \end{aligned} \quad (43)$$

Note that $\|A\| = \max_i \left\{ \sum_j |a_{i,j}| \right\}$, it follows from Eq. 38 that

$$\|\mathcal{DF}(x)\| = \max \left\{ \|A_1(x)\| + \|B_2(x)\|, \sup_{k \geq 2} \{ \|A_k(x)\| + \|B_k(x)\| + \|C_k(x)\| \} \right\}. \tag{44}$$

Since $u_0e \leq 1$ and $x_1e \leq 1$, we obtain

$$\frac{(u_0e)^d - (x_1e)^d}{u_0e - x_1e} = \sum_{j=0}^{d-1} (u_0e)^j (x_1e)^{d-1-j} \leq d,$$

$$\frac{[(u_0e)^d - (x_1e)^d] - d(x_1e)^{d-1}(u_0e - x_1e)}{(u_0e - x_1e)^2} = \sum_{k=0}^{d-2} \sum_{j=0}^k (u_0e)^j (x_1e)^{k-j} \leq \frac{(d-1)(d-2)}{2};$$

$$\frac{(x_{k-1}e)^d - (x_k e)^d}{x_{k-1}e - x_k e} \leq d,$$

$$\frac{[(x_{k-1}e)^d - (x_k e)^d] - d(x_k e)^{d-1}(x_{k-1}e - x_k e)}{(x_{k-1}e - x_k e)^2} \leq \frac{(d-1)(d-2)}{2}.$$

Thus it follows from Eqs. 39 and 40 that

$$\|A_1(x)\| \leq \|C + \text{diag}(De)\| + \frac{2d + (d-1)(d-2)}{2} \|D\| + \|T\|,$$

$$\|B_1(x)\| \leq [d + (d-1)(d-2)] \|D\|,$$

$$\|A_1(x)\| + \|B_1(x)\| \leq \|C + \text{diag}(De)\| + \left[2d + \frac{3(d-1)(d-2)}{2} \right] \|D\| + \|T\|.$$

It follows from Eqs. 41 to 43 that for $k \geq 2$

$$\|A_k(x)\| \leq \|C + \text{diag}(De)\| + [d + (d-1)(d-2)] \|D\| + \|T\|,$$

$$\|B_k(x)\| \leq [d + (d-1)(d-2)] \|D\|,$$

$$\|C_k(x)\| = \|T^0\alpha\|,$$

hence we have

$$\begin{aligned} & \|A_k(x)\| + \|B_k(x)\| + \|C_k(x)\| \\ & \leq \|C + \text{diag}(De)\| + 2[d + (d-1)(d-2)] \|D\| + \|T\| + \|T^0\alpha\|. \end{aligned}$$

Let

$$M = \max \left\{ \|C + \text{diag}(De)\| + 2[d + (d-1)(d-2)] \|D\| + \|T\| + \|T^0\alpha\| \right\}.$$

Then

$$\|A_1(x)\| + \|B_1(x)\| \leq M$$

and for $k \geq 2$

$$\|A_k(x)\| + \|B_k(x)\| + \|C_k(x)\| \leq M.$$

Hence, it follows from Eq. 44 that

$$\|\mathcal{DF}(x)\| \leq M.$$

Note that $x = \mathbf{u}$, this gives that for $\mathbf{u} \in \tilde{\Omega}$

$$\|\mathcal{DF}(\mathbf{u})\| \leq M. \tag{45}$$

For $\mathbf{u}, \mathbf{v} \in \tilde{\Omega}$,

$$\begin{aligned} \|F(\mathbf{u}) - F(\mathbf{v})\| &\leq \sup_{0 \leq t \leq 1} \|\mathcal{D}F(\mathbf{u} + t(\mathbf{v} - \mathbf{u}))\| \|\mathbf{u} - \mathbf{v}\| \\ &\leq M \|\mathbf{u} - \mathbf{v}\|. \end{aligned} \quad (46)$$

This indicates that the function $F(\mathbf{u})$ is Lipschitz for $\mathbf{u} \in \tilde{\Omega}$.

Note that $x = \mathbf{u}$, it follows from Eqs. 31 and 33 that for $\mathbf{u} \in \tilde{\Omega}$

$$\mathbf{u}(t) = \mathbf{u}(0) + \int_0^t F(\mathbf{u}(\xi)) d\xi,$$

this gives

$$\mathbf{u}(t) = \tilde{\mathbf{g}} + \int_0^t F(\mathbf{u}(\xi)) d\xi. \quad (47)$$

Using the Picard approximation as well as the Lipschitz condition, it is easy to prove that there exists the unique solution to the integral Eq. 47 according to the basic results of the Banach space. Therefore, there exists the unique solution to the system of differential vector Eqs. 31 to 33 (that is, Eq. 22 to Eq. 26).

6 A matrix-analytic solution

In this section, we first discuss the stability of this supermarket model in terms of a coupling method. Then we provide a generalized matrix-analytic method for computing the fixed point whose doubly exponential solution and phase-structured tail are obtained. Finally, we discuss some useful limits of the fraction vector $\mathbf{u}^{(N)}(t)$ as $N \rightarrow \infty$ and $t \rightarrow +\infty$.

6.1 Stability of this supermarket model

In this subsection, we provide a coupling method to study the stability of this supermarket model of N identical servers with MAP inputs and PH service times, and give a sufficient condition under which this supermarket model is stable.

Let Q and R denote two supermarket models with MAP inputs and PH service times, both of which have the same parameters $N, d, m_A, C, D, m_B, \alpha, T$, and the same initial state at $t = 0$. Let $d(Q)$ and $d(R)$ be two choice numbers in the two supermarket models Q and R , respectively. We assume $d(Q) = 1$ and $d(R) \geq 2$. Thus, the only difference between the two supermarket models Q and R is the two different choice numbers: $d(Q) = 1$ and $d(R) \geq 2$.

For the two supermarket models Q and R , we define two infinite-dimensional Markov processes $\{U_N^{(Q)}(t) : t \geq 0\}$ and $\{U_N^{(R)}(t) : t \geq 0\}$, respectively. The following theorem sets up a coupling between the two processes $\{U_N^{(Q)}(t) : t \geq 0\}$ and $\{U_N^{(R)}(t) : t \geq 0\}$.

Theorem 3 *For the two supermarket models Q and R , there is a coupling between the two processes $\{U_N^{(Q)}(t) : t \geq 0\}$ and $\{U_N^{(R)}(t) : t \geq 0\}$ such that the total number of customers in the supermarket model R is no greater than the total number of customers in the supermarket model Q at time $t \geq 0$.*

Proof See Appendix C (<http://arxiv.org/pdf/1406.0285.pdf>). □

Remark 3 Note that the N queueing processes in this supermarket model is symmetric, it is easy to see from Theorem 3 that the queue length of each server in the supermarket model R is no greater than that in the supermarket model Q at time $t \geq 0$.

Since this supermarket model with MAP inputs and PH service times is more general, it is necessary to extend the coupling method given in Turner (1996) and Martin and Suhov (1999) through a detailed probability analysis given in Appendix C. We show that such a coupling method can be applied to discussing stability of more general supermarket models.

Note that the stationary arrival rate of the MAP of irreducible matrix descriptor (C, D) is given by $\lambda = \omega D e$, and the mean of the PH service time is given by $1/\mu = -\alpha T^{-1} e$. The following theorem provides a sufficient condition under which this supermarket model is stable.

Theorem 4 *This supermarket model of N identical servers with MAP inputs and PH service times is stable if $\rho = \lambda/\mu < 1$.*

Proof From the two different choice numbers: $d(Q) = 1$ and $d(R) \geq 2$, we set up two different supermarket models Q and R , respectively. Note that the supermarket model Q is the set of N parallel and independent MAP/PH/1 queues. Obviously, the MAP/PH/1 queue is described as a QBD process whose infinitesimal generator is given by

$$Q = \begin{pmatrix} C & D \otimes \alpha & & & \\ I \otimes T^0 & C \oplus T & D \otimes I & & \\ & I \otimes (T^0 \alpha) & C \oplus T & D \otimes I & \\ & & & \ddots & \ddots & \ddots \end{pmatrix}.$$

Note that

$$A = A_{-1} + A_0 + A_1 = (C + D) \oplus (T + T^0 \alpha),$$

where

$$A_{-1} = I \otimes (T^0 \alpha), \quad A_0 = C \oplus T, \quad A_1 = D \otimes I,$$

thus it is easy to check that $\omega \otimes \theta$ is the stationary probability vector of the Markov chain A , where θ is the stationary probability vector of the Markov chain $T + T^0 \alpha$. Using Chapter 3 of Li (2010), it is clear that the QBD process Q is stable if $(\omega \otimes \theta) A_{-1} e > (\omega \otimes \theta) A_2 e$, that is, $\rho = \lambda/\mu < 1$. Hence, the supermarket model Q is stable if $\rho < 1$. It is seen from Theorem 3 and Remark 3 that the queue length of each server in the supermarket model R is no greater than that in the supermarket model Q at time $t \geq 0$, this shows that the supermarket model R is stable if the supermarket model Q is stable. Thus the supermarket model R is stable if $\rho = \lambda/\mu < 1$. This completes the proof. □

6.2 Computation of the fixed point

A row vector $\pi = (\pi_0, \pi_1, \pi_2, \dots)$ is called a fixed point of the infinite-dimensional system of differential vector Eqs. 22 to 26 satisfied by the fraction vector $\mathbf{u}(t)$ if $\pi = \lim_{t \rightarrow +\infty} \mathbf{u}(t)$, or $\pi_k = \lim_{t \rightarrow +\infty} u_k(t)$ for $k \geq 0$.

It is well-known that if π is a fixed point of the vector $\mathbf{u}(t)$, then

$$\lim_{t \rightarrow +\infty} \left[\frac{d}{dt} \mathbf{u}(t) \right] = 0.$$

Let

$$L_1(\pi_0 \otimes \alpha, \pi_1) = \sum_{m=1}^d C_d^m \left[\sum_{l=1}^{m_A} \sum_{j=1}^{m_B} (\pi_{0;l} \alpha_j - \pi_{1;l,j}) \right]^{m-1} \left[\sum_{l=1}^{m_A} \sum_{j=1}^{m_B} \pi_{1;l,j} \right]^{d-m}$$

and for $k \geq 2$

$$L_k(\pi_{k-1}, \pi_k) = \sum_{m=1}^d C_d^m \left[\sum_{l=1}^{m_A} \sum_{j=1}^{m_B} (\pi_{k-1;l,j} - \pi_{k;l,j}) \right]^{m-1} \left[\sum_{l=1}^{m_A} \sum_{j=1}^{m_B} \pi_{k;l,j} \right]^{d-m}.$$

Then

$$L_1(\pi_0 \otimes \alpha, \pi_1) = \frac{1 - (\pi_1 e)^d}{1 - \pi_1 e}$$

and for $k \geq 2$

$$L_k(\pi_{k-1}, \pi_k) = \frac{(\pi_{k-1} e)^d - (\pi_k e)^d}{\pi_{k-1} e - \pi_k e}.$$

To determine the fixed point $\pi = (\pi_0, \pi_1, \pi_2, \dots)$, as $t \rightarrow +\infty$ taking limits on both sides of Eqs. 22 to 26 we obtain the system of nonlinear vector equations as follows:

$$\pi_0(C + D) = 0, \quad \pi_0 e = 1, \quad (48)$$

$$\begin{aligned} & \{(\pi_0 \otimes \alpha)(D \otimes I) - \pi_1[\text{diag}(De) \otimes I]\} L_1(\pi_0 \otimes \alpha, \pi_1) \\ & + \pi_1 \{[C + \text{diag}(De)] \oplus T\} + \pi_2(I \otimes T^0 \alpha) = 0, \end{aligned} \quad (49)$$

and for $k \geq 2$

$$\begin{aligned} & \{\pi_{k-1}(D \otimes I) - \pi_k[\text{diag}(De) \otimes I]\} L_k(\pi_{k-1}, \pi_k) \\ & + \pi_k \{[C + \text{diag}(De)] \oplus T\} + \pi_{k+1}(I \otimes T^0 \alpha) = 0. \end{aligned} \quad (50)$$

Since ω is the stationary probability vector of the Markov chain $C + D$, then it follows from Eq. 48 that

$$\pi_0 = \omega. \quad (51)$$

For the fixed point $\pi = (\pi_0, \pi_1, \pi_2, \dots)$, $(\pi_0 e, \pi_1 e, \pi_2 e, \dots)$ is the tail vector of the stationary queue length distribution. The following theorem shows that the tail vector $(\pi_0 e, \pi_1 e, \pi_2 e, \dots)$ of the stationary queue length distribution is doubly exponential.

Theorem 5 *If $\rho = \lambda/\mu < 1$, then the tail vector $(\pi_0 e, \pi_1 e, \pi_2 e, \dots)$ of the stationary queue length distribution is doubly exponential, that is, for $k \geq 0$*

$$\pi_k e = \rho^{\frac{d^k - 1}{d - 1}}. \quad (52)$$

Proof Multiplying both sides of the Eq. 50 by the vector e , and noting that $[C + \text{diag}(De)]e = 0$ and $Te = -T^0$, we obtain

$$[(\pi_0 \otimes \alpha)(De \otimes e) - \pi_1(De \otimes e)] L_k(\pi_0 \otimes \alpha, \pi_1) - \mu \left[\pi_1(e \otimes T^0) - \pi_2(e \otimes T^0) \right] = 0 \quad (53)$$

and for $k \geq 2$,

$$[\pi_{k-1} (De \otimes e) - \pi_k (De \otimes e)] L_k (\pi_{k-1}, \pi_k) - \mu [\pi_k (e \otimes T^0) - \pi_{k+1} (e \otimes T^0)] = 0. \tag{54}$$

Let $\pi_k = \eta_k (\omega \otimes \theta)$ for $k \geq 1$. Note that $\lambda = \omega De$, $\mu = \theta T^0$ and $\rho = \lambda/\mu$, it follows from Eqs. 53 and 54 that

$$\rho (1 - \eta_1^d) - (\eta_1 - \eta_2) = 0$$

and

$$\rho (\eta_{k-1}^d - \eta_k^d) - (\eta_k - \eta_{k+1}) = 0.$$

This gives

$$\pi_k e = \eta_k = \rho^{\frac{d^k-1}{d-1}}.$$

This completes the proof. □

Let $\zeta_1 = L_1 (\pi_0 \otimes \alpha, \pi_1)$ and $\zeta_k = L_k (\pi_{k-1}, \pi_k)$ for $k \geq 2$. Then

$$\zeta_k = \frac{\rho^{\frac{d^k-d}{d-1}} - \rho^{\frac{d^{k+1}-d}{d-1}}}{\rho^{\frac{d^{k-1}-1}{d-1}} - \rho^{\frac{d^k-1}{d-1}}}, \quad k \geq 1.$$

We write

$$B_k = [C + (1 - \zeta_k) \text{diag} (De)] \oplus T$$

and

$$Q = \begin{pmatrix} B_1 & \zeta_2 (D \otimes I) & & & \\ I \otimes (T^0 \alpha) & B_2 & \zeta_3 (D \otimes I) & & \\ & I \otimes (T^0 \alpha) & B_3 & \zeta_4 (D \otimes I) & \\ & & \ddots & \ddots & \ddots \end{pmatrix}.$$

Then the level-dependent QBD process is irreducible and transient, since

$$\zeta_1 > \zeta_2 > \zeta_3 > \dots > 0,$$

$$[B_1 + \zeta_2 (D \otimes I)] e = -(\zeta_1 - \zeta_2) [(De) \otimes e] - e \otimes T^0 \preceq 0$$

and

$$[I \otimes (T^0 \alpha) + B_k + \zeta_k (D \otimes I)] e = -(\zeta_k - \zeta_{k+1}) [(De) \otimes e] \preceq 0.$$

In what follows we will express the fixed point $(\pi_0, \pi_1, \pi_2, \dots)$. To that end, we need to provide the UL-type of *RG*-factorization of the QBD process Q according to Chapter 1 in Li (2010) or Li and Cao (2004). Applying the UL-type of *RG*-factorization, we can give the maximal non-positive inverse of matrix Q , which leads to the matrix-product solution of the fixed point $(\pi_0, \pi_1, \pi_2, \dots)$ by means of the *R*- and *U*-measures.

Let the matrix sequence $\{R_k, k \geq 1\}$ be the minimal nonnegative solution to the nonlinear matrix equations

$$\xi_{k+1} (D \otimes I) + R_k B_{k+1} + R_k R_{k+1} \left[I \otimes (T^0 \alpha) \right] = 0,$$

and the matrix sequence $\{G_k, k \geq 2\}$ be the minimal nonnegative solution to the nonlinear matrix equations

$$I \otimes (T^0 \alpha) + B_k G_k + \xi_{k+1} (D \otimes I) G_{k+1} G_k = 0.$$

Let the matrix sequence $\{U_k, k \geq 0\}$ be

$$\begin{aligned} U_k &= B_{k+1} + [\zeta_{k+2} (D \otimes I)] [-U_{k+1}]^{-1} \left[I \otimes (T^0 \alpha) \right] \\ &= B_{k+1} + R_{k+1} \left[I \otimes (T^0 \alpha) \right] \\ &= B_{k+1} + [\zeta_{k+2} (D \otimes I)] G_{k+1}. \end{aligned}$$

Hence we obtain

$$R_0 = \zeta_1 (D \otimes I) (-U_1)^{-1}$$

and

$$G_1 = (-U_0)^{-1} \left[I \otimes (T^0 \alpha) \right].$$

Based on the R -measure $\{R_k, k \geq 0\}$, G -measure $\{G_k, k \geq 1\}$ and U -measure $\{U_k, k \geq 0\}$, we can get the UL-type of RG -factorization of the matrix Q as follows

$$Q = (I - R_U) U_D (I - G_L),$$

where

$$R_U = \begin{pmatrix} 0 & R_0 & & & \\ & 0 & R_1 & & \\ & & 0 & R_2 & \\ & & & \ddots & \ddots \\ & & & & \ddots & \ddots \end{pmatrix},$$

$$U_D = \text{diag}(U_0, U_1, U_2, \dots)$$

and

$$G_L = \begin{pmatrix} I & & & & \\ G_1 & I & & & \\ & G_2 & I & & \\ & & & \ddots & \ddots & \ddots \end{pmatrix}.$$

Using the RG -factorization, we obtain the maximal non-positive inverse of the matrix Q as follows

$$Q_{max}^{-1} = (I - G_L)^{-1} U_D^{-1} (I - R_U)^{-1}, \quad (55)$$

where

$$(I - R_U)^{-1} = \begin{pmatrix} I & X_1^{(0)} & X_2^{(0)} & X_3^{(0)} & \cdots \\ & I & X_1^{(1)} & X_2^{(1)} & \cdots \\ & & I & X_1^{(2)} & \cdots \\ & & & I & \cdots \\ & & & & \ddots \end{pmatrix},$$

$$X_k^{(l)} = R_l R_{l+1} R_{l+2} \cdots R_{l+k-1}, \quad k \geq 1, l \geq 0;$$

$$U_D^{-1} = \text{diag} \left(U_0^{-1}, U_1^{-1}, U_2^{-1}, \dots \right);$$

$$(I - G_L)^{-1} = \begin{pmatrix} I & & & & \\ Y_1^{(1)} & I & & & \\ Y_2^{(2)} & Y_1^{(2)} & I & & \\ Y_3^{(3)} & Y_2^{(3)} & Y_1^{(3)} & I & \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

$$Y_k^{(l)} = G_l G_{l-1} G_{l-2} \cdots G_{l-k+1}, \quad l \geq k \geq 1.$$

The following theorem illustrates that the fixed point $(\pi_0, \pi_1, \pi_2, \dots)$ is matrix-product.

Theorem 6 *If $\rho < 1$, then the fixed point $\pi = (\pi_0, \pi_1, \pi_2, \dots)$ is given by*

$$\begin{aligned} \pi_0 &= \omega, \\ \pi_1 &= \zeta_1 (\omega \otimes \alpha) (D \otimes I) (-U_0)^{-1} \end{aligned} \tag{56}$$

and for $k \geq 2$

$$\pi_k = \zeta_1 (\omega \otimes \alpha) (D \otimes I) (-U_0)^{-1} R_0 R_1 \cdots R_{k-2}. \tag{57}$$

Proof It follows from Eq. 54 that

$$\begin{aligned} (\pi_1, \pi_2, \pi_3, \dots) & \begin{pmatrix} B_1 & \zeta_2 (D \otimes I) & & & \\ I \otimes (T^0 \alpha) & B_2 & \zeta_3 (D \otimes I) & & \\ & I \otimes (T^0 \alpha) & B_3 & \zeta_4 (D \otimes I) & \\ & & & \ddots & \ddots & \ddots \end{pmatrix} \\ &= -(\zeta_1 (\omega \otimes \alpha) (D \otimes I), 0, 0, \dots). \end{aligned}$$

This gives

$$(\pi_1, \pi_2, \pi_3, \dots) = -(\zeta_1 (\omega \otimes \alpha) (D \otimes I), 0, 0, \dots) (I - G_L)^{-1} U_D^{-1} (I - R_U)^{-1}.$$

Thus we obtain

$$\pi_1 = \zeta_1 (\omega \otimes \alpha) (D \otimes I) (-U_0)^{-1}$$

and for $k \geq 2$

$$\pi_k = \zeta_1 (\omega \otimes \alpha) (D \otimes I) (-U_0)^{-1} R_0 R_1 \cdots R_{k-2}.$$

This completes the proof. □

To further understand the fixed point, we consider a special block-structured supermarket model with Poisson inputs and PH service times. In this case, we can give an interesting explicit expression of the fixed point.

Note that $C = -\lambda$, $D = \lambda$, it is clear that $\omega = 1$ and $\pi_0 = 1$. It follows from Eqs. 49 and 50 that

$$\lambda(\theta - \pi_1) \frac{1 - (\pi_1 e)^d}{1 - (\pi_1 e)} + \pi_1 T + \pi_2 T^0 \alpha = 0$$

and for $k \geq 2$

$$\lambda(\pi_{k-1} - \pi_k) \frac{(\pi_{k-1} e)^d - (\pi_k e)^d}{(\pi_{k-1} e) - (\pi_k e)} + \pi_k T + \pi_{k+1} T^0 \alpha = 0.$$

Thus we obtain

$$(\pi_1, \pi_2, \pi_3, \dots) \Theta = \lambda \left((\theta - \pi_1) \frac{1 - (\pi_1 e)^d}{1 - (\pi_1 e)}, (\pi_1 - \pi_2) \frac{(\pi_1 e)^d - (\pi_2 e)^d}{(\pi_1 e) - (\pi_2 e)}, \dots \right), \quad (58)$$

where

$$\Theta = \begin{pmatrix} -T & & & & \\ -T^0 \alpha & -T & & & \\ & -T^0 \alpha & -T & & \\ & & & \ddots & \ddots \\ & & & & \ddots & \ddots \end{pmatrix}.$$

Since

$$\Theta^{-1} = \begin{pmatrix} (-T)^{-1} & & & & \\ (e\alpha)(-T)^{-1} & (-T)^{-1} & & & \\ (e\alpha)(-T)^{-1} & (e\alpha)(-T)^{-1} & (-T)^{-1} & & \\ (e\alpha)(-T)^{-1} & (e\alpha)(-T)^{-1} & (e\alpha)(-T)^{-1} & (-T)^{-1} & \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

It follows from Eq. 58 that

$$\pi_1 \left[I + \lambda \frac{1 - (\pi_1 e)^d}{1 - (\pi_1 e)} (-T)^{-1} \right] = \lambda \frac{1 - (\pi_1 e)^d}{1 - (\pi_1 e)} \theta (-T)^{-1} + \lambda \alpha (-T)^{-1} (\pi_1 e)^d \quad (59)$$

and for $k \geq 2$

$$\pi_k \left[I + \lambda \frac{(\pi_{k-1} e)^d - (\pi_k e)^d}{(\pi_{k-1} e) - (\pi_k e)} (-T)^{-1} \right] = \lambda \frac{(\pi_{k-1} e)^d - (\pi_k e)^d}{(\pi_{k-1} e) - (\pi_k e)} \theta (-T)^{-1} + \lambda \alpha (-T)^{-1} (\pi_k e)^d. \quad (60)$$

Note that the matrices $I + \lambda \frac{1 - (\pi_1 e)^d}{1 - (\pi_1 e)} (-T)^{-1}$ and $I + \lambda \frac{(\pi_{k-1} e)^d - (\pi_k e)^d}{(\pi_{k-1} e) - (\pi_k e)} (-T)^{-1}$ for $k \geq 2$ are all invertible, it follows from Eqs. (59) and (60) that

$$\pi_1 = \left[\lambda \frac{1 - (\pi_1 e)^d}{1 - (\pi_1 e)} \theta (-T)^{-1} + \lambda \alpha (-T)^{-1} (\pi_1 e)^d \right] \left[I + \lambda \frac{1 - (\pi_1 e)^d}{1 - (\pi_1 e)} (-T)^{-1} \right]^{-1}.$$

and for $k \geq 2$

$$\pi_k = \left[\lambda \frac{(\pi_{k-1}e)^d - (\pi_k e)^d}{(\pi_{k-1}e) - (\pi_k e)} \omega (-T)^{-1} + \lambda \alpha (-T)^{-1} (\pi_k e)^d \right] \times \left[I + \lambda \frac{(\pi_{k-1}e)^d - (\pi_k e)^d}{(\pi_{k-1}e) - (\pi_k e)} (-T)^{-1} \right]^{-1}.$$

Thus we obtain

$$\pi_1 = \left[\lambda \zeta_1 \omega (-T)^{-1} + \lambda \alpha (-T)^{-1} \rho^d \right] \left[I + \lambda \zeta_1 (-T)^{-1} \right]^{-1} \tag{61}$$

and for $k \geq 2$

$$\pi_k = \left[\lambda \zeta_k \omega (-T)^{-1} + \lambda \alpha (-T)^{-1} \rho^{\frac{dk+1-d}{d-1}} \right] \left[I + \lambda \zeta_k (-T)^{-1} \right]^{-1}. \tag{62}$$

Remark 4 For this block-structured supermarket model, the fixed point is matrix-product and depends on the R-measure $\{R_k, k \geq 0\}$, see Eqs. (56) and (57). However, when the input is a Poisson process, we can give the explicit expression of the fixed point by Eqs. (61) and (62). This explains the reason why the MAP input makes the study of block-structured supermarket models more difficult and challenging.

6.3 The double limits

In this subsection, we discuss some useful limits of the fraction vector $\mathbf{u}^{(N)}(t)$ as $N \rightarrow \infty$ and $t \rightarrow +\infty$. Note that the limits are necessary for using the stationary probabilities of the limiting process to give an effective approximate performance of this supermarket model.

The following theorem gives the limit of the vector $\mathbf{u}(t, \mathbf{g})$ as $t \rightarrow +\infty$, that is,

$$\lim_{t \rightarrow +\infty} \mathbf{u}(t, \mathbf{g}) = \lim_{t \rightarrow +\infty} \lim_{N \rightarrow \infty} \mathbf{u}^{(N)}(t, \mathbf{g}).$$

Theorem 7 *If $\rho < 1$, then for any $\mathbf{g} \in \Omega$*

$$\lim_{t \rightarrow +\infty} \mathbf{u}(t, \mathbf{g}) = \pi.$$

Furthermore, there exists a unique probability measure φ on Ω , which is invariant under the map $\mathbf{g} \mapsto \mathbf{u}(t, \mathbf{g})$, that is, for any continuous function $f : \Omega \rightarrow \mathbf{R}$ and $t > 0$

$$\int_{\Omega} f(\mathbf{g}) d\varphi(\mathbf{g}) = \int_{\Omega} f(\mathbf{u}(t, \mathbf{g})) d\varphi(\mathbf{g}).$$

Also, $\varphi = \delta_{\pi}$ is the probability measure concentrated at the fixed point π .

Proof It is seen from Theorem 6 that the condition $\rho < 1$ guarantees the existence of solution in Ω to the system of nonlinear Eqs. (48) to (50). This indicates that if $\rho < 1$, then as $t \rightarrow +\infty$, the limit of $\mathbf{u}(t, \mathbf{g})$ exists in Ω . Since $\mathbf{u}(t, \mathbf{g})$ is the unique and global solution to the infinite-dimensional system of differential vector Eqs. (22) to (26) for $t \geq 0$, the vector $\lim_{t \rightarrow +\infty} \mathbf{u}(t, \mathbf{g})$ is also a solution to the system of nonlinear Eqs. (48) to (50). Note that π is the unique solution to the system of nonlinear Eqs. (48) to (50), hence we obtain that $\lim_{t \rightarrow +\infty} \mathbf{u}(t, \mathbf{g}) = \pi$. The second statement in this theorem can be immediately given by the probability measure of the limiting process $\{U(t), t \geq 0\}$ on state space Ω . This completes the proof. □

The following theorem indicates the weak convergence of the sequence $\{\varphi_N\}$ of stationary probability distributions for the sequence $\{U^{(N)}(t), t \geq 0\}$ of Markov processes to the probability measure concentrated at the fixed point π .

Theorem 8 (1) If $\rho < 1$, then for a fixed number $N = 1, 2, 3, \dots$, the Markov process $\{U^{(N)}(t), t \geq 0\}$ is positive recurrent, and has a unique invariant distribution φ_N .

(2) $\{\varphi_N\}$ weakly converges to δ_π , that is, for any continuous function $f : \Omega \rightarrow \mathbf{R}$

$$\lim_{N \rightarrow \infty} E_{\varphi_N} [f(\mathbf{g})] = f(\pi).$$

Proof (1) From Theorem 3, this supermarket model of N identical servers is stable if $\rho < 1$, hence this supermarket model has a unique invariant distribution φ_N .

(2) Since $\tilde{\Omega}$ is compact under the ordinary metric $\rho(\mathbf{u}, \mathbf{u}')$, so is the set $\mathcal{P}(\tilde{\Omega})$ of probability measures. Hence the sequence $\{\varphi_N\}$ of invariant distributions has limiting points. A similar analysis to the proof of Theorem 5 in Martin and Suhov (1999) shows that $\{\varphi_N\}$ weakly converges to δ_π and $\lim_{N \rightarrow \infty} E_{\varphi_N} [f(\mathbf{g})] = f(\pi)$. This completes the proof. \square

Based on Theorems 7 and 8, we obtain a useful relation as follows

$$\lim_{t \rightarrow +\infty} \lim_{N \rightarrow \infty} \mathbf{u}^{(N)}(t, \mathbf{g}) = \lim_{N \rightarrow \infty} \lim_{t \rightarrow +\infty} \mathbf{u}^{(N)}(t, \mathbf{g}) = \pi.$$

Therefore, we have

$$\lim_{\substack{N \rightarrow \infty \\ t \rightarrow +\infty}} \mathbf{u}^{(N)}(t, \mathbf{g}) = \pi,$$

which justifies the interchange of the limits of $N \rightarrow \infty$ and $t \rightarrow +\infty$. This is necessary in many practical applications when using the stationary probabilities of the limiting process to give an effective approximation for the performance analysis of this supermarket model.

7 Performance computation

In this section, we provide two performance measures of this supermarket model, and use some numerical examples to show how the two performance measures of this supermarket model depend on the non-Poisson MAP inputs and on the non-exponential PH service times.

7.1 Performance measures

For this supermarket model, we provide two simple performance measures as follows:

(1) *The mean of the stationary queue length in any server* The mean of the stationary queue length in any server is given by

$$E[Q_d] = \sum_{k=1}^{\infty} \pi_k e = \sum_{k=1}^{\infty} \rho \frac{d^k - 1}{d - 1}. \quad (63)$$

(2) *The expected sojourn time that any arriving customer spends in this system* Note that $u_0^{(N)}(0) \geq 0$ and $u_0^{(N)}(0) e = 1$, it is clear that

$$\lim_{t \rightarrow +\infty} u_0^{(N)}(t) = \lim_{t \rightarrow +\infty} u_0^{(N)}(0) \exp\{(C + D)t\} = \omega.$$

For the PH service times, any arriving customer finds k customer in any server whose probability is given by $(\omega \otimes \alpha - \pi_1) L_d(\omega \otimes \alpha, \pi_1) e$ for $k = 0$ and $(\pi_k - \pi_{k+1}) L_d(\pi_k, \pi_{k+1}) e$

for $k \geq 1$. When $k \geq 1$, the head customer in the server has been served, and so its service time is residual and is denoted as X_R . Let X be of phase type with irreducible representation (α, T) . Then X_R is also of phase type with irreducible representation (θ, T) , where θ is the stationary probability vector of the Markov chain $T + T^0\alpha$. Clearly, we have

$$E[X] = \alpha(-T)^{-1}e, \quad E[X_R] = \theta(-T)^{-1}e.$$

Thus it is easy to see that the expected sojourn time that any arriving customer spends in this system is given by

$$\begin{aligned} E[T_d] &= (\omega \otimes \alpha - \pi_1) L_d(\omega \otimes \alpha, \pi_1) e E[X] \\ &\quad + \sum_{k=1}^{\infty} (\pi_k - \pi_{k+1}) L_d(\pi_k, \pi_{k+1}) e \{E[X_R] + kE[X]\} \\ &= [1 - (\pi_1 e)^d] E[X] + \sum_{k=1}^{\infty} [(\pi_k e)^d - (\pi_{k+1} e)^d] \{E[X_R] + kE[X]\} \\ &= (\pi_1 e)^d [E[X_R] - E[X]] + E[X] \left[1 + \sum_{k=1}^{\infty} (\pi_k e)^d \right] \end{aligned} \tag{64}$$

From Eqs. 63 and 64, we obtain

$$E[T_d] = E[X] \{1 + E[Q_d^*]\} + \rho^d \{E[X_R] - E[X]\}, \tag{65}$$

where

$$E[Q_d^*] = \sum_{k=1}^{\infty} (\pi_k e)^d$$

is the mean of stationary queue length of any server which is seen by an arriving customer. Specifically, if $E[X_R] = E[X]$ (for example, the exponential service times), then

$$E[T_d] = E[X] \{1 + E[Q_d^*]\}, \tag{66}$$

which is the Little’s formula in this supermarket model from the sojourn time setting.

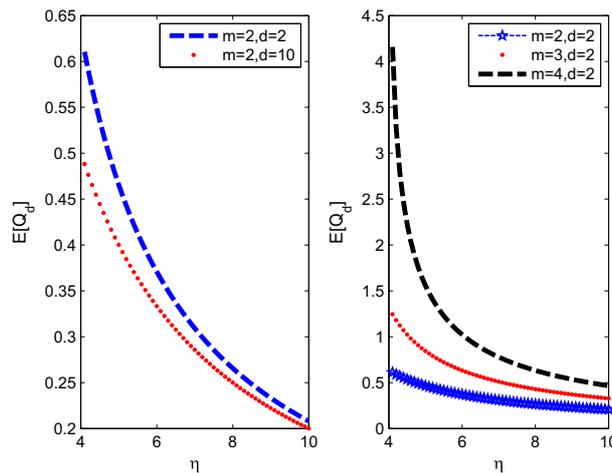


Fig. 3 $E[Q_d]$ vs η for $(m, d) = (2, 2), (3, 2), (4, 2)$ and $(2, 10)$

It is seen from Eq. 63 that $E[Q_d]$ only depends on the traffic intensity $\rho = \lambda/\mu$, where $\lambda = \omega D e$ and $\mu = -\alpha T^{-1} e$; and from Eq. 64 that $E[T_d]$ depends not only on the traffic intensity ρ but also on the mean $E[X_R]$ of the residual PH service time, where $E[X_R] = \theta(-T)^{-1} e$. Based on this, it is clear that the performance computation of this supermarket model can be given easily for more general MAP inputs and PH service times, although here our numerical examples are simple.

7.2 Numerical examples

In this subsection, we provide some numerical examples which are used to indicate how the performance measures of this supermarket model depend on the non-Poisson MAP inputs and on the non-exponential PH service times.

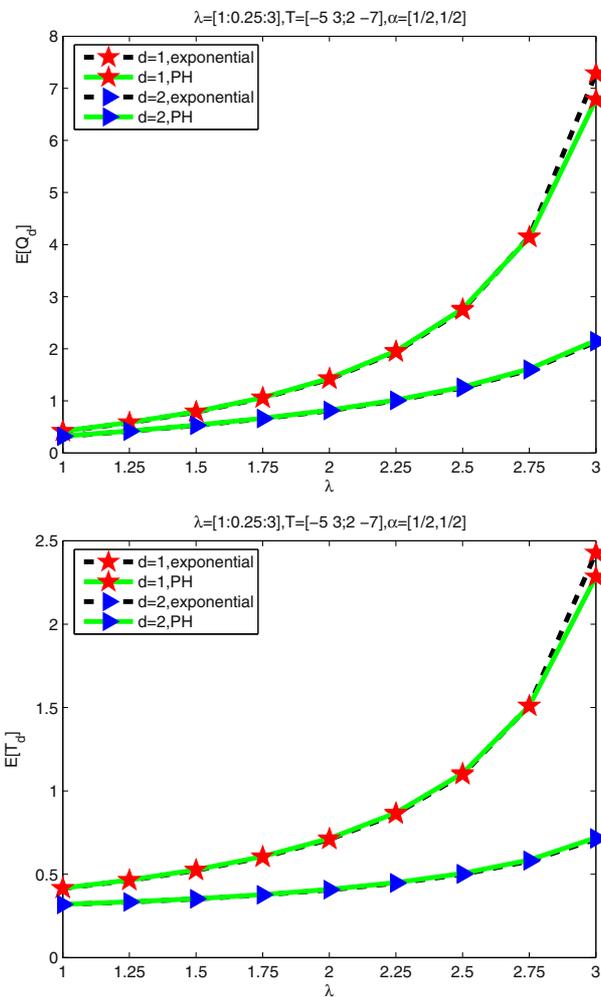


Fig. 4 Performance comparison between the exponential and PH service times

Example one: The Erlang service times In this supermarket model, the customers arrive at this system as a Poisson process with arrival rate $N\lambda$, and the service times at each server are an Erlang distribution $E[m, \eta]$. Let $\lambda = 1$. Then $\rho = m/\eta$. When $\rho < 1$, we have $\eta > m$. Figure 3 shows how $E[Q_d]$ depends on the different parameter pairs $(m, d) = (2, 2), (3, 2), (4, 2)$ and $(2, 10)$, respectively. It is seen that $E[Q_d]$ decreases as d increases or as η increases, and it increases as m increases.

Example two: Performance comparisons between the exponential and PH service times We consider two related supermarket models with Poisson inputs of arrival rate $N\lambda$: one with exponential service times, while another with PH service times. For the two supermarket models, our goal is to observe the influence of different service time distributions on the performance of this supermarket model. To that end, the parameters of this system are taken as

$$\mu = 3.4118, \alpha = \left(\frac{1}{2}, \frac{1}{2}\right), T = \begin{pmatrix} -5 & 3 \\ 2 & -7 \end{pmatrix}.$$

Under the exponential and PH service times, Fig. 4 depicts how $E[Q_d]$ and $E[T_d]$ depend on the arrival rate $\lambda \in [1, 3]$ with $\lambda < \mu$, and on the choice number $d = 1, 2$. It is seen that $E[Q_d]$ and $E[T_d]$ decrease as d increases, while $E[Q_d]$ and $E[T_d]$ increase as λ increases.

Example three: The role of the PH service times In this supermarket model with $d = 2$, the customers arrive at this system as a Poisson process with arrival rate $N\lambda$, and the service times at each server are a PH distribution with irreducible representation $(\alpha, T(i))$, $\alpha = (1/2, 1/2)$,

$$T(1) = \begin{pmatrix} -5 & 3 \\ 2 & -7 \end{pmatrix}, T(2) = \begin{pmatrix} -4 & 3 \\ 2 & -7 \end{pmatrix}, T(3) = \begin{pmatrix} -4 & 4 \\ 2 & -7 \end{pmatrix}.$$

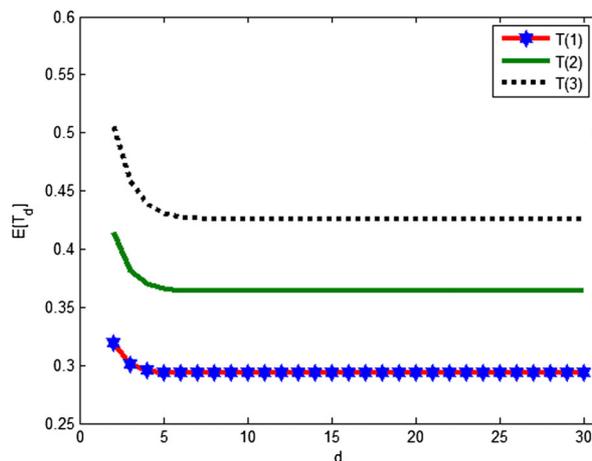
It is seen that some minor changes are designed in the first rows of the matrices $T(i)$ for $i = 1, 2, 3$. Let $\lambda = 1$. Then

$$\rho(1) = 0.2931, \rho(2) = 0.3636, \rho(3) = 0.4250.$$

This gives

$$\rho(1) < \rho(2) < \rho(3).$$

Fig. 5 $E[T_d(i)]$ vs the transition rate matrices $T(i)$ for $i = 1, 2, 3$



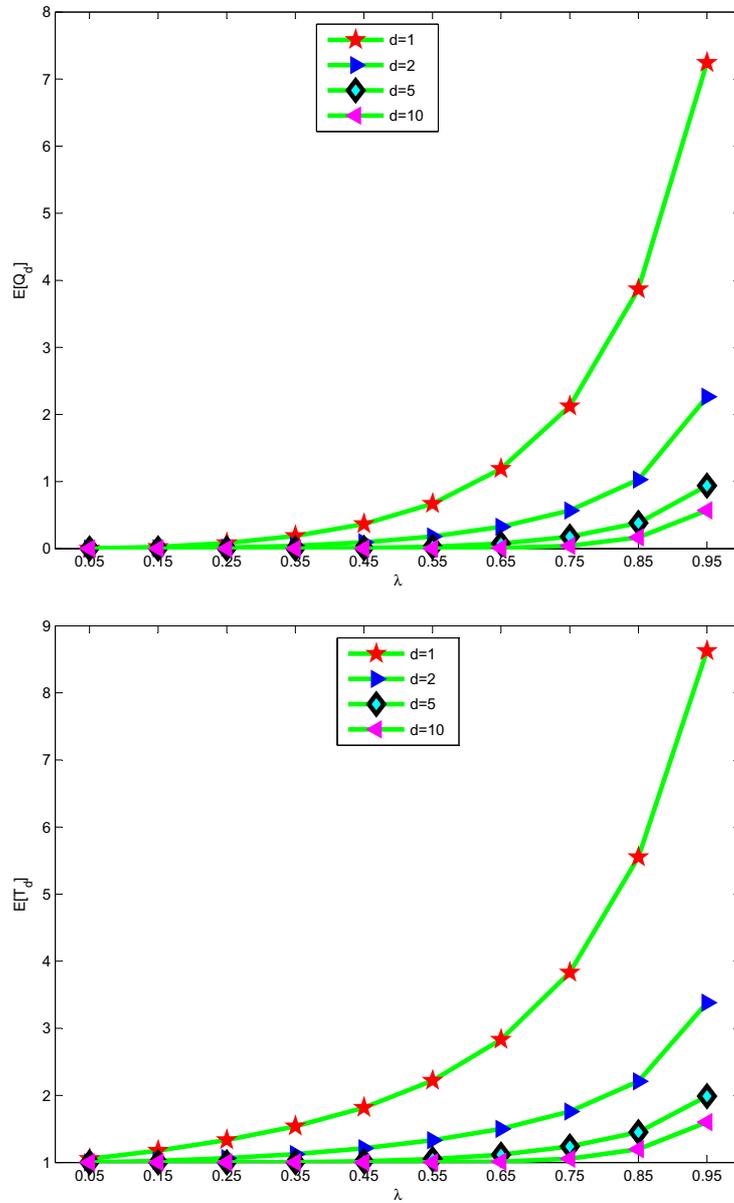


Fig. 6 The role of the MAP inputs

Figure 5 indicates how $E [T_d]$ depends on the different transition rate matrices $T (i)$ for $i = 1, 2, 3$, and

$$E [T_d (1)] < E [T_d (2)] < E [T_d (3)].$$

It is seen that $E [T_d]$ decreases as d increases.

Example four: The role of the MAP inputs In this supermarket model, the service time distribution is exponential with service rate $\mu = 1$, and the arrival processes are the MAP

of irreducible matrix descriptor $(C(N), D(N))$, where

$$C = \begin{pmatrix} -5 - \frac{2}{7}\lambda & 5 \\ 7 & -7 - 2\lambda \end{pmatrix}, \quad D = \begin{pmatrix} \frac{2}{7}\lambda & 0 \\ 0 & 2\lambda \end{pmatrix}.$$

It is easy to check that $\omega = (7/12, 5/12)$, and the stationary arrival rate $\lambda^* = \omega D e = \lambda$. If $\mu = 1$ and $\rho = \lambda^*/\mu = \lambda < 1$, then $\lambda \in (0, 1)$.

Figure 6 shows how $E[Q_d]$ and $E[T_d]$ depend on the parameter λ of the MAP under different choice numbers $d = 1, 2, 5, 10$. It is seen that $E[Q_d]$ and $E[T_d]$ decrease as d increases, while $E[Q_d]$ and $E[T_d]$ increase as λ increases.

8 Concluding remarks

In this paper, we analyze a more general block-structured supermarket model with non-Poisson MAP inputs and with non-exponential PH service times, and set up an infinite-dimensional system of differential vector equations satisfied by the expected fraction vector through a detailed probability analysis, where an important result: The invariance of environment factors is obtained. We apply the phase-structured operator semigroup to proving the phase-structured mean-field limit, which indicates the asymptotic independence of the block-structured queueing processes in this supermarket model. Furthermore, we provide an effective algorithm for computing the fixed point by means of the generalized matrix-analytic method. Using the fixed point, we provide two performance measures of this supermarket model, and use some numerical examples to illustrate how the two performance measures depend on the non-Poisson MAP inputs and on the non-exponential PH service times. From many practical applications, the block-structured supermarket model is an important queueing model to analyze the relation between the system performance and the job routing rule, and it can also help to design reasonable architecture to improve the performance and to balance the load.

Note that this paper provide a clear picture for how to use the phase-structured mean-field model as well as the generalized matrix-analytic method to analyze performance measures of more general supermarket models. We show that this picture is organized as three key parts: (1) Setting up system of differential equations, (2) necessary proofs of the phase-structured mean-field limit, and (3) performance computation of this supermarket model through the fixed point. Therefore, the results of this paper give new highlight on understanding influence of non-Poisson inputs and of non-exponential service times on performance measures of more general supermarket models. Along such a line, there are a number of interesting directions for potential future research, for example:

- analyzing non-Poisson inputs such as renewal processes;
- studying non-exponential service time distributions, for example, general distributions, matrix-exponential distributions and heavy-tailed distributions; and
- discussing the bulk arrival processes, such as BMAP inputs, and the bulk service processes, where effective algorithms for the fixed point are necessary and interesting.

Up to now, we believe that a larger gap exists when dealing with either renewal inputs or general service times in a supermarket model, because a more challenging infinite-dimensional system of differential equations need be established, a more complicated mean-field limit need be proved, and computation of the fixed point will be more interesting, difficult and challenging.

Acknowledgments The authors thank the Associate Editor and two reviewers for many valuable comments to sufficiently improve the presentation of this paper. At the same time, Q.L. Li was supported by the National Natural Science Foundation of China under grant No. 71271187, the Hebei Natural Science Foundation of China under grant No. A2012203125, and the Fostering Plan of Innovation Team and Leading Talent in Hebei Universities under grant No. LJRC027.

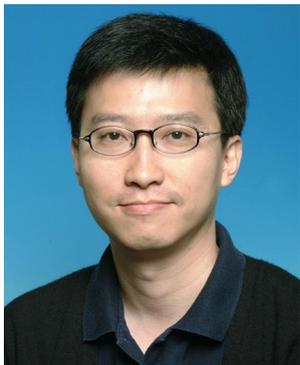
References

- Bramson M, Lu Y, Prabhakar B (2010) Randomized load balancing with general service time distributions. In: Proceedings of the ACM SIGMETRICS international conference on measurement and modeling of computer systems, pp 275–286
- Bramson M, Lu Y, Prabhakar B (2012) Asymptotic independence of queues under randomized load balancing. *Queueing Syst* 71:247–292
- Bramson M, Lu Y, Prabhakar B (2013) Decay of tails at equilibrium for FIFO join the shortest queue networks. *Ann Appl Probab* 23:1841–1878
- Ethier SN, Kurtz TG (1986) Markov processes: characterization and convergence. Wiley, New York
- Graham C (2000a) Kinetic limits for large communication networks. In: Bellomo N., Pulvirenti M. (eds) Modelling in applied sciences. Birkhäuser, pp 317–370
- Graham C (2000b) Chaoticity on path space for a queueing network with selection of the shortest queue among several. *J Appl Probab* 37:198–201
- Graham C (2004) Functional central limit theorems for a large network in which customers join the shortest of several queues. *Probab Theory Relat Fields* 131:97–120
- Jacquet P, Vvedenskaya ND (1998) On/off sources in an interconnection networks: Performance analysis when packets are routed to the shortest queue of two randomly selected nodes, technical report N^o 3570, INRIA Rocquencourt, France
- Jacquet P, Suhov YM, Vvedenskaya ND (1999) Dynamic routing in the mean-field approximation, technical report N^o 3789, INRIA Rocquencourt, France
- Kurtz TG (1981) Approximation of population processes. SIAM
- Li QL (2010) Constructive computation in stochastic models with applications: The *RG*-factorizations, Springer and Tsinghua Press
- Li QL (2011) Super-exponential solution in Markovian supermarket models: Framework and challenge, Available: arXiv:1106.0787
- Li QL (2014) Tail probabilities in queueing processes. *Asia Pac J Oper Res* 31(2):1–31
- Li QL, Cao J (2004) Two types of *RG*-factorizations of quasi-birth-and-death processes and their applications to stochastic integral functionals. *Stoch Model* 20:299–340
- Li QL, Dai G, Lui JCS, Wang Y (2013) The mean-field computation in a supermarket model with server multiple vacations. *Discret Event Dyn Syst* 24(2014):1–50. Publishing Online
- Li QL, Lui JCS (2010) Doubly exponential solution for randomized load balancing models with Markovian arrival processes and PH service times, Available: arXiv:1105.4341
- Li QL, Lui JCS (2014) Block-structured supermarket models: Three appendices, Available: arXiv:1406.0285
- Li QL, Lui JCS, Wang Y (2011) A matrix-analytic solution for randomized load balancing models with PH service times. In: Performance evaluation of computer and communication systems: milestones and future challenges. Lecture notes in computer science, 6821, pp 240–253
- Luczak MJ, McDiarmid C (2006) On the maximum queue length in the supermarket model. *Ann Probab* 34:493–527
- Luczak MJ, McDiarmid C (2007) Asymptotic distributions and chaos for the supermarket model. *Electron J Probab* 12:75–99
- Luczak MJ, Norris JR (2005) Strong approximation for the supermarket model. *Ann Appl Probab* 15:2038–2061
- Martin JB (2001) Point processes in fast Jackson networks. *Ann Appl Probab* 11:650–663
- Martin JB, Suhov YM (1999) Fast Jackson networks. *Ann Appl Probab* 9:854–870
- Mitzenmacher MD (1996) The power of two choices in randomized load balancing. PhD Thesis, department of computer science, University of California, Berkeley
- Mitzenmacher MD (1999) On the analysis of randomized load balancing schemes. *Theory Comput Syst* 32:361–386
- Mitzenmacher MD, Richa A, Sitaraman R (2001) The power of two random choices: a survey of techniques and results. In: Handbook of randomized computing, vol 1, pp 255–312

- Mitzenmacher MD, Upfal E (2005) Probability and computing: randomized algorithms and probabilistic analysis. Cambridge University Press
- Neuts MF (1981) Matrix-geometric solutions in stochastic models: an algorithmic approach. Johns Hopkins University Press
- Neuts MF (1989) Structured stochastic matrices of $M/G/1$ Type and their applications. Marcel Decker Inc., New York
- Suhov YM, Vvedenskaya ND (2002) Fast Jackson Networks with Dynamic Routing. *Probl Inf Transm* 38:136–153
- Turner SRE (1996) Resource Pooling in Stochastic Networks. Ph.D. Thesis, Statistical Laboratory, Christ's College, University of Cambridge
- Turner SRE (1998) The effect of increasing routing choice on resource pooling. *Prob Eng Information Sci* 12:109–124
- Vvedenskaya ND, Dobrushin RL, Karpelevich FI (1996) Queueing system with selection of the shortest of two queues: an asymptotic approach. *Probl Inf Transm* 32:20–34
- Vvedenskaya ND, Suhov YM (1997) Dobrushin's mean-field approximation for a queue with dynamic routing. *Markov Process Relat Fields* 3:493–526
- Vvedenskaya ND, Suhov YM (2005) Dynamic routing queueing systems with vacations. *Inf Process Electr Sci J* 5:74–86. The Keldysh institute of applied mathematics. The institute for information transmission problems



Quan-Lin Li is Full Professor in School of Economics and Management Sciences, Yanshan University, Qinhuangdao, China. He received the Ph.D. degree in Institute of Applied Mathematics, Chinese Academy of Sciences, Beijing, China in 1998. He has published a book (*Constructive Computation in Stochastic Models with Applications: The RG-Factorizations*, Springer, 2010) and over 40 research papers in a variety of journals, such as, *Advances in Applied Probability*, *Queueing Systems*, *Stochastic Models*, *European Journal of Operational Research*, *Computer Networks*, *Performance Evaluation*, *Discrete Event Dynamic Systems*, *Computers & Operations Research*, *Computers & Mathematics with Applications*, *Annals of Operations Research*, and *International Journal of Production Economics*. His main research interests concern with Queueing Theory, Stochastic Models, Matrix-Analytic Methods, Manufacturing Systems, Computer Networks, Network Security, and Supply Chain Risk Management.



John C. S. Lui (M93-SM02-F10) was born in Hong Kong. He received the Ph.D. degree in computer science from the University of California, Los Angeles, 1992. He is currently a Professor with the Department of Computer Science and Engineering, The Chinese University of Hong Kong (CUHK), Hong Kong. He was the chairman of the Department from 2005 to 2011. His current research interests are in communication networks, network/system security (e.g., cloud security, mobile security, etc.), network economics, network sciences (e.g., online social networks, information spreading, etc.), cloud computing, large-scale distributed systems, and performance evaluation theory. Professor Lui is a Fellow of the Association for Computing Machinery (ACM), a Fellow of IEEE, a Croucher Senior Research Fellow, and an elected member of the IFIP WG 7.3. He serves on the Editorial Board of *IEEE/ACM Transactions on Networking*, *IEEE Transactions on Computers*, *IEEE Transactions on Parallel and Distributed Systems*, *Journal of Performance Evaluation* and *International Journal of Network Security*. He received various departmental teaching awards and the CUHK Vice-Chancellors Exemplary Teaching Award. He is also a co-recipient of the IFIP WG 7.3 Performance 2005 and IEEE/IFIP NOMS 2006 Best Student Paper Awards.