Available online at www.sciencedirect.com
SCIENCE DIRECT
Operations Research Letters 34 (2006) 199–204
Operations Research Letters
www.elsevier.com/locate/orl

A simple proof of diffusion approximations for LBFS re-entrant lines

Jiankui Yang[a], J.G. Dai[b,*], Jian-Gong You[a], Hanqin Zhang[c]

[a]*Department of Mathematics, Nanjing University, Nanjing, China*
[b]*School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, 30332-0205, USA*
[c]*Institute of Applied Mathematics, Academia Sinica, Beijing 100080, China*

Received 13 November 2004; accepted 10 March 2005
Available online 1 June 2005

Abstract

For a re-entrant line operating under the last-buffer–first-serve service policy, there have been two independent proofs of a heavy traffic limit theorem. The key to these proofs is to prove the uniform convergence of a critical fluid model. We give a new proof for the uniform convergence of the fluid model.

Keywords: Multiclass queueing networks; Heavy traffic; Reflecting Brownian motion; Fluid model; Critical stability; Linear Lyapunov function

1. Introduction

This paper is concerned with diffusion approximations for re-entrant lines operating under the last-buffer–first-serve (LBFS) service policy. A re-entrant line is a special type of multiclass queueing network that was first introduced in [8]. In a re-entrant line, all jobs follow a deterministic route. In this paper, we assume that all jobs make K visits to J stations of the network. Jobs in the kth stage of their visits are called class k jobs, and the station they visit is denoted by station $s(k)$. In other words, each job that enters from the outside of the network is a class 1 job; after it completes the service at station $s(1)$, it turns into a class 2 job and is routed to station $s(2)$, and so on. After service completion at station $s(K)$, a class K job leaves the network. We assume that each station has a single server and an infinite waiting room. We further assume that the network is operated under the LBFS service policy. Namely, when a server switches from one job to another, the new job will be taken from the leading (the longest waiting) job at the highest (the last) nonempty class at the server's station. For concreteness, we also assume that the service policy is preemptive resume. That is, when a job in a higher class than the one currently being

* Corresponding author.
E-mail address: dai@isye.gatech.edu (J.G. Dai).

0167-6377/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.orl.2005.03.004

served arrives at the server's station, the service of the current job is interrupted. When service of all jobs in higher classes is completed, the interrupted service continues from where it left off. For each class k, let $Q_k(t)$ indicate the number of class k jobs in the network at time t, $k = 1, 2, \ldots, K$. For each station j, let $W_j(t)$ denote the amount of work for server j (measured in units of remaining service time) embodied in those jobs that are at station j at time t. If no more arrivals (either external and internal) are allowed at station j after time t, server j needs to work $W_j(t)$ additional units of time before the station is empty. The diffusion approximation addresses the weak convergence of a scaled version of the K-dimensional queue length process $Q=\{Q(t), t\geqslant 0\}$ and the J-dimensional workload process $W = \{W(t), t\geqslant 0\}$ under a heavy traffic condition. It is typical the workload limit W^* is a J-dimensional reflecting Brownian motion, and the K-dimensional queue length limit Q^* is a constant multiple of the J-dimensional workload limit W^*. Namely, there exists some constant, $K \times J$ matrix Δ such that

$$Q^*(t) = \Delta W^*(t), \quad t\geqslant 0, \tag{1}$$

a condition known as state space collapse; see, for example, [1,10].

Bramson and Dai [2] and Chen and Ye [3] independently established the diffusion approximations for LBFS re-entrant lines by verifying the uniform convergence of the corresponding fluid model. Both proofs are long and difficult to follow. In this paper, we give a simple proof of the uniform convergence by constructing a suitable linear Lyapunov function. Our analysis of the Lyapunov function is a refinement of the one first done in [5]. Readers are referred to [5] for detailed discussions on fluid models and their analysis. We also refer readers to [2,3] and references therein for detailed discussions on diffusion approximations of queueing networks.

Now we introduce some notation for our network. Let $\alpha > 0$ be the arrival rate of jobs from the outside and $m_k > 0$ be the mean service time of class k jobs, $k=1, \ldots, K$. We use $u(\ell)/\alpha$ to denote the interarrival time between the $(\ell-1)$th and the ℓth externally arriving job; we use $m_k v_k(\ell)$ to denote the service time of the ℓth class k job. We assume $u = \{u(\ell), \ell\geqslant 1\}$ and $v_k = \{v_k(\ell), \ell\geqslant 1\}$, $k = 1, \ldots, K$, are independent, identically distributed sequences, and these sequences are mutually independent. We assume that $\mathbb{E}(u(1))=1$, $\mathbb{E}(v_k(1))=1$, $a=\mathrm{var}(u(1)) < \infty$, and $b_k = \mathrm{var}(v_k(1)) < \infty$. We denote by $\mathscr{C}(j) = \{k : s(k) = j\}$ the set of classes that are served at station j, and by $C=(C_{jk})$ the $J\times K$ matrix with $C_{jk}=1$ when $s(k)=j$, and $C_{jk}=0$, otherwise. To avoid triviality, we assume that $\mathscr{C}(j) \neq \emptyset$ for all $j=1, \ldots, J$, i.e., all stations must be visited at least once. Let $P = (P_{k\ell})$ be the $K \times K$ routing matrix, i.e., $P_{k\ell}=1$ when $\ell=k+1$ and 0 otherwise. We use $m=(m_k)$ to denote the K-dimensional (column) vector of mean service times, and M to denote the corresponding diagonal matrix $\mathrm{diag}(m)$. Let $\rho = \alpha C m$ be the J-dimensional vector of traffic intensities. The jth component of ρ is $\rho_j = \alpha\sum_{k\in\mathscr{C}(j)} m_k$.

In the next section, we state the main theorem of this paper. The LBFS fluid model will be introduced in Section 3. There we prove that the fluid model is uniformly convergent, a fact that will be the key to the proof of our main theorem.

2. Main results

As a standard procedure, we consider a sequence of queueing networks, indexed by $n = 1, 2, \ldots$. We assume that the network topology and the routing in the sequence do not depend on n. Thus, each network in the sequence is the same one as described in the previous section except that, for the nth network, the external arrival rate is α^n and the mean service time for class k jobs is m_k^n. (The arrival rate and mean service times are the only network parameters that depend on n.) Thus, in the nth network, $\{u(\ell)/\alpha^n, \ell\geqslant 1\}$ and $\{v_k^n(\ell) = m_k^n v_k(\ell), \ell\geqslant 1\}$, $k = 1, \ldots, K$, are the interarrival and service time sequences. All processes associated with the nth network are appended with a superscript n. For example, $Q^n(t)$ and $W^n(t)$ denote the queue length vector and the workload vector, respectively, at time t in the nth network. In addition to the queue length and workload processes, we now introduce the J-dimensional idle-time process $Y^n = \{Y^n(t), t\geqslant 0\}$. For station j in the nth network, $Y_j^n(t)$ denotes the total amount of time that the server at station j has been idle over $[0, t]$.

Formally, the diffusion approximation is concerned with the following weak convergence result

$$(\hat{Q}^n, \hat{W}^n, \hat{Y}^n) \Longrightarrow (Q^*, W^*, Y^*) \quad \text{as } n \to \infty, \tag{2}$$

where

$$\hat{Q}^n(t) = \frac{1}{\sqrt{n}} Q^n(nt),$$
$$\hat{W}^n(t) = \frac{1}{\sqrt{n}} W^n(nt),$$
$$\hat{Y}^n(t) = \frac{1}{\sqrt{n}} Y^n(nt).$$

The weak convergence in (2), denoted by $\Longrightarrow$, is on the $(K+2J)$-dimensional path space $\mathbb{D}^{K+2J}[0,\infty)$ that is endowed with the Skorokhod J_1 topology; for weak convergence and the path space, see, for example, [7]. To state our theorem precisely, we first introduce assumptions on the sequence of networks.

We assume that, as $n \to \infty$,

$$\alpha^n \to \alpha, \quad m_k^n \to m_k, \quad k = 1, \ldots, K, \tag{3}$$

and $\rho^n \to e$ at the rate

$$\sqrt{n}(\rho^n - e) \to \gamma, \tag{4}$$

where e is the J-dimensional vector of ones and γ is some J-dimensional vector. Note that (3) and (4) imply that each station is critically loaded in the limit, i.e.,

$$\rho_j = \alpha \sum_{k \in \mathscr{C}(j)} m_k = 1 \quad \text{for each station } j. \tag{5}$$

Let Δ be the $K \times J$ matrix given by

$$\Delta_{kj} = \begin{cases} 1/m_k & \text{if } k \in \mathscr{C}(j) \text{ and is the} \\ & \text{lowest priority class at station } j, \\ 0 & \text{otherwise.} \end{cases}$$

It serves as the lifting matrix in (1) under the LBFS service policy. For the existence of diffusion approximation, we assume that the initial data satisfy the following:

$$\hat{W}^n(0) \Rightarrow \xi \quad \text{as } n \to \infty \quad \text{for some nonnegative random vector } \xi \tag{6}$$

and

$$|\hat{Q}^n(0) - \Delta \hat{W}^n(0)| \to 0 \text{ in probability as } n \to \infty. \tag{7}$$

In defining the limit process in our theorem, we set

$$R = (I + CM(I - P')^{-1} P' \Delta)^{-1},$$
$$\theta = R\gamma, \tag{8}$$

$$\Gamma = \alpha RCM \Big(\operatorname{diag}(b_1, b_2, \ldots, b_K) + (I - P')^{-1} \operatorname{diag}(\alpha^2 a, 0, \ldots, 0) \times (I - P)^{-1} \Big) MC'R', \tag{9}$$

where prime denotes the transpose. It follows from Theorems 3.1 and 3.2 of [6] that the inverse in (8) exists, and thus the matrix R is defined. The matrix Γ is necessarily nonnegative definite. In the following theorem, we assume that the matrix Γ is positive definite (The formula Γ in (9) follows from (3.13) of [1]. There is a typo in (3.8) of [2]).

Theorem 2.1. *Assume that* (3)–(4) *and* (6)–(7) *all hold and Γ is positive definite. The weak convergence* (2) *holds as $n \to \infty$. Furthermore, the limit process* (Q^*, W^*, Y^*), *defined on some probability space* $(\Omega, \mathscr{F}, \mathbb{P})$, *satisfies the following relationships*:

(i) *$\mathbb{P}$-almost surely, W^* has continuous paths with $W^*(t) \in \mathbf{R}_+^J$ for $t \geqslant 0$ and*
(ii) *$W^*(t) = X^*(t) + RY^*(t)$ for $t \geqslant 0$,*
(iii) *Under $\mathbb{P}$, X^* is a Brownian motion with drift vector θ and covariance matrix Γ such that $X^*(0)$ has the distribution of ξ, and $\{X^*(t) - X^*(0) - \theta t, t \geqslant 0\}$ is martingale with respect tothe filtration generated by W^* and Y^*,*
(iv) *For each $j = 1, \ldots, J$, $\mathbb{P}$-almost surely,*
  (a) *$Y_j^*(0) = 0$,*
  (b) *Y_j^* is continuous and nondecreasing,*
  (c) *Y_j^* can increase only at times t where $W_j^*(t) = 0$,*
(v) *$\mathbb{P}$-almost surely, $Q^*(t) = \Delta W^*(t)$ for $t \geqslant 0$.*

The process W^* is known as a semimartingale reflecting Brownian motion with reflection matrix R, drift vector θ and covariance matrix Γ (see [9]). It follows from Theorems 3.1 and 3.2 of [6] that R is a completely-$\mathscr{S}$ matrix. Thus, W^* along with Y^* is well defined, and (W^*, Y^*) is unique in distribution. Relationship (v) is the state space collapse condition (1).

We are going to employ Theorem 5.3 of [2] to prove the theorem. To employ their theorem, we will check in the next section that a corresponding *fluid model* of the re-entrant network is *uniformly convergent* with lifting matrix Δ.

3. Fluid model and the simplified proof

In this section, we first introduce the fluid model of the sequence of reentrant lines. We then define the notion of uniform convergence for the fluid model. Finally, we prove that our fluid model is uniformly convergent with lifting matrix Δ, thus completing the proof of Theorem 2.1.

The fluid model for the sequence of re-entrant lines in Section 2 is defined through the following set of equations. Here, $\mu_k = 1/m_k$ and $\mu_0 = \alpha$.

$$\bar{Q}_k(t) = \bar{Q}_k(0) + \mu_{k-1}\bar{T}_{k-1}(t) - \mu_k \bar{T}_k(t) \quad \text{for } t \geqslant 0 \quad \text{and} \quad k = 1, \ldots, K, \tag{10}$$

$$\bar{Q}_k(t) \geqslant 0 \quad \text{for } t \geqslant 0 \quad \text{and} \quad k = 1, \ldots, K, \tag{11}$$

$$\bar{T}_k(0) = 0 \quad \text{and} \quad \bar{T}_k(\cdot) \quad \text{is nondecreasing for } k = 1, \ldots, K, \tag{12}$$

$\bar{I}_k(t)$ can increase only at times t when

$$\sum_{\ell \in \mathscr{C}(s(k)), \ell \geqslant k} \bar{Q}_\ell(t) = 0, \tag{13}$$

for $k = 1, 2, \ldots, K$, where

$$\bar{I}_k(t) = t - \sum_{\ell \in \mathscr{C}(s(k)), \ell \geqslant k} \bar{T}_\ell(t) \quad \text{for all } t \geqslant 0.$$

By convention, $\bar{T}_0(t) = t$ in (10). Eqs. (10)–(13) are known as the fluid model equations operating under the LBFS service policy. They define the LBFS fluid model. Since condition (5) holds, the fluid model is critically loaded. Each solution $(\bar{Q}, \bar{T})$ to (10)–(13) is a fluid model solution.

Since the rest of the paper is exclusively focused on the critically loaded LBFS fluid model, for notational convenience, we drop the bar from a fluid model solution.

Definition 3.1. The fluid model is said to be *uniformly convergent* with lifting matrix Δ if there exists a function h: $\mathbf{R}_+ \to \mathbf{R}_+$ with $h(t) \to 0$ as $t \to \infty$ such that for each fluid model solution (Q, T) with $|Q(0)| = 1$,

$$|Q(t) - Q(\infty)| \leqslant h(t), \quad \text{for all } t \geqslant 0,$$

for some $Q(\infty) \in \mathbf{R}_+^K$ satisfying

$$Q(\infty) = \Delta w \quad \text{for some } w \in \mathbf{R}_+^J.$$

In the definition, any norm $|\cdot|$ on $\mathbf{R}^K$ can be employed. For concreteness, for a vector $x = (x_k) \in \mathbf{R}^K$, we define $|x| = \sum_k |x_k|$. To prove that our fluid model is uniformly convergent, we introduce some more notation. For each station j with $1 \leqslant j \leqslant J$, define $K_j = \min\{k : k \in \mathscr{C}(j)\}$, the lowest class at station j. Without loss of generality, we assume that

$$K_1 < \cdots < K_J.$$

Let $\mathscr{K} = \{1, \ldots, K\}$, $\mathscr{L} = \{K_1, \ldots, K_J\}$, and $\mathscr{H} = \mathscr{K} \backslash \mathscr{L}$. For a fluid model solution (Q, T), a time t is said to be a *regular point* of (Q, T) if (Q, T) is differentiable at t. Whenever a derivative is considered at time t, it is assumed that t is a regular point of the corresponding fluid model solution.

Theorem 3.1. *Assume that* (5) *holds. There exists a* $\tau > 0$ *such that for any fluid model solution* (Q, T) *with* $|Q(0)| \leqslant 1$,

$$Q_\ell(\tau + t) = 0 \quad \text{for } t \geqslant 0 \quad \textit{and} \quad \ell \in \mathscr{H}, \tag{14}$$

$$Q_\ell(\tau + t) = Q_\ell(\tau) \quad \text{for } t \geqslant 0 \quad \textit{and} \quad \ell \in \mathscr{L}. \tag{15}$$

Condition (14) is known as the *SHP-condition* in [4]. Using a sufficient condition established in [4], Chen and Ye [3] proved Theorem 2.1 using a direct, lengthy proof of the SHP-condition.

Proof of Theorem 3.1. In view of $\sum_{\ell \in \mathscr{H}} Q_\ell(t) = \sum_{j=1}^{J} \sum_{\ell = K_j+1}^{K_{j+1}-1} Q_\ell(t)$, with convention that $K_{J+1} = K + 1$, it suffices to prove that there exist constants $\tau_1 > \tau_2 > \cdots > \tau_J$ such that for any fluid model solution (Q, T) with $|Q(0)| \leqslant 1$ and $j = 1, \ldots, J$,

$$\sum_{\ell = K_j+1}^{K_{j+1}-1} Q_\ell(t) = 0 \quad \text{for each } t \geqslant \tau_j, \tag{16}$$

$$\mu_\ell \dot{T}_\ell(t) = \alpha \quad \text{for } \ell \geqslant K_j \text{ and each regular point } t \geqslant \tau_j. \tag{17}$$

We now use the backward induction on j to prove (16) and (17). We first need to prove that (16) and (17) hold for $j = J$. Since this proof is identical to the one in the induction step, we omit it. Assume that (16) and (17) hold for $J, J-1, \ldots, i+1$ with $i \geqslant 1$. We now prove that (16) holds for $j = i$. Let

$$F_i(t) = \sum_{\ell=K_i+1}^{K_{i+1}-1} Q_\ell(t) \quad \text{and} \quad L_i(t) = \sum_{\ell=1}^{K_{i+1}-1} Q_\ell(t).$$

We claim that for any regular $t \geqslant \tau_{i+1}$,

$$F_i(t) > 0 \quad \text{implies } \dot{L}_i(t) \leqslant -\varepsilon, \tag{18}$$

where

$$\varepsilon = \alpha \min_{1 \leqslant i \leqslant J} \left(1 - \alpha \sum_{\ell \in \mathscr{C}(i), \ell \neq K_i} m_\ell\right).$$

Note that $\varepsilon > 0$ by (5). To see (18), assume that $F_i(t) > 0$. Let k be the last class ℓ in $K_i < \ell < K_{i+1}$ with $Q_\ell(t) > 0$. We calculate $\mu_k \dot{T}_k(t)$ first. Since $Q_\ell(\cdot)$ is a Lipschitz continuous function, by Lemma 2.2 of [5], $Q_\ell(t) = 0$ implies $\dot{Q}_\ell(t) = 0$. Then from the definition of k,

$$\mu_\ell \dot{T}_\ell(t) = \mu_k \dot{T}_k(t) \quad \text{for } k \leqslant \ell < K_{i+1}.$$

By the induction assumption,

$$\mu_\ell \dot{T}_\ell(t) = \alpha \quad \text{for } \ell \geqslant K_{i+1}.$$

By (13), $\sum_{\ell \in \mathscr{C}(j), \ell \geqslant k} \dot{T}_\ell(t) = 1$, where $j = s(k)$ and $k > K_j$. Therefore,

$$\mu_k \dot{T}_k(t) = \frac{1 - \alpha \sum_{\ell \in \mathscr{C}(j), \ell \geqslant K_{i+1}} m_\ell}{\sum_{\ell \in \mathscr{C}(j), k \leqslant \ell < K_{i+1}} m_\ell}. \tag{19}$$

From (10) and (19),

$$\begin{aligned}
\dot{L}_i(t) &= \sum_{\ell=1}^{k} \dot{Q}_\ell(t) = \alpha - \mu_k \dot{T}_k(t) \\
&= \frac{\alpha \sum_{\ell \in \mathscr{C}(j), \ell \geqslant k} m_\ell - 1}{\sum_{\ell \in \mathscr{C}(j), k \leqslant \ell < K_{i+1}} m_\ell} \\
&\leqslant \frac{\alpha \sum_{\ell \in \mathscr{C}(j), \ell \neq K_j} m_\ell - 1}{\sum_{\ell \in \mathscr{C}(j), k \leqslant \ell < K_{i+1}} m_\ell} \\
&\leqslant \alpha \left(\alpha \sum_{\ell \in \mathscr{C}(j), \ell \neq K_j} m_\ell - 1\right) \leqslant -\varepsilon.
\end{aligned}$$

It follows from (18) that for any $s \geqslant \tau_{i+1}$, $F_i(t) = 0$ for some $t \in (s, s + L_i(s)/\varepsilon)$. In particular,

$$F_i(t) = 0 \quad \text{for some } t \in (\tau_{i+1}, \tau_{i+1} + L_i(\tau_{i+1})/\varepsilon), \tag{20}$$

and $F_i(t)$ equals to 0 infinitely many times after time τ_{i+1}. We next show that for any two time points t_1 and t_2 with $\tau_{i+1} \leqslant t_1 < t_2$,

$$\begin{aligned}
&F_i(t_1) = F_i(t_2) = 0 \quad \text{implies that} \\
&F_i(t) = 0 \quad \text{for all } t \in (t_1, t_2).
\end{aligned} \tag{21}$$

To see this, we first claim that $\dot{L}_i(t) \leqslant 0$ for $t \geqslant \tau_{i+1}$. This claim holds trivially when $L_i(t) = 0$. Suppose that $L_i(t) > 0$. The claim can be proved by mimicking the proof of (18) with k being defined to be the last class ℓ with $1 \leqslant \ell < K_{i+1}$ and $Q_\ell(t) > 0$. Suppose that (21) does not hold, i.e., $F_i(t) > 0$ for some $t \in (t_1, t_2)$. Then there exists some interval $(s_1, s_2) \subset (t_1, t_2)$ such that $F_i(t) > 0$ for all $t \in (s_1, s_2)$. It follows (18) that

$$L_i(t_1) \geqslant L_i(s_1) > L_i(s_2) \geqslant L_i(t_2). \tag{22}$$

On the other hand, let

$$G_i(t) = \sum_{k \in \mathscr{C}(i)} m_k \sum_{\ell=1}^{k} Q_\ell(t)$$

be the total workload at station i. It follows from fluid model equations (10)–(13) and (5) that

$$\dot{G}_i(t) = 1 - \sum_{\ell \in \mathscr{C}(i)} \dot{T}_\ell(t) \geqslant 0,$$

thus

$$G_i(t_1) \leqslant G_i(t_2). \tag{23}$$

By the induction assumption, $Q_\ell(t_1) = 0$ for $\ell > K_i$ and $\ell \in \mathscr{C}(i)$. Thus,

$$\begin{aligned}
G_i(t_1) &= \sum_{k \in \mathscr{C}(i)} m_k \sum_{\ell=1}^{K_i} Q_\ell(t_1) \\
&= \left(\sum_{\ell=1}^{K_i} Q_\ell(t_1)\right)\left(\sum_{k \in \mathscr{C}(i)} m_k\right) \\
&= L_i(t_1) \left(\sum_{k \in \mathscr{C}(i)} m_k\right).
\end{aligned} \tag{24}$$

Similarly, we have

$$G_i(t_2) = L_i(t_2) \left(\sum_{k \in \mathscr{C}(i)} m_k\right). \tag{25}$$

It follows from (22) and (24)–(25) that

$$G_i(t_1) = L_i(t_1)\left(\sum_{k\in\mathscr{C}(i)} m_k\right) > L_i(t_2)\left(\sum_{k\in\mathscr{C}(i)} m_k\right) = G_i(t_2),$$

contradicting (23). Thus, we have proved (21).

It follows from (20) and (21) that $F_i(t)=0$ for $t \geqslant \tau_{i+1} + L_i(\tau_{i+1})/\varepsilon$. Note that $L_i(\tau_{i+1}) \leqslant L_i(0) + \alpha\tau_{i+1} \leqslant 1 + \alpha\tau_{i+1}$. By setting

$$\tau_i = \frac{1+(\alpha+\varepsilon)\tau_{i+1}}{\varepsilon},$$

which is independent of fluid model solutions, we have $F_i(t)=0$ for $t \geqslant \tau_i$, thus proving (16) for $j=i$.

It remains to prove that (17) holds for $j=i$. Let $t \geqslant \tau_i$ be a regular point. Since $Q_\ell(t)=0$ for $K_i < \ell < K_{i+1}$, $\mu_\ell \dot{T}_\ell(t) = \mu_{K_i}\dot{T}_{K_i}(t)$ for $K_i < \ell < K_{i+1}$. It suffices to prove that

$$\mu_{K_i}\dot{T}_{K_i}(t) = \alpha. \tag{26}$$

We first assume that $Q_{K_i}(t) > 0$. In this case, (26) follows from (5) and (19), with $k = K_i$ and $j = i$ in (19).

Clearly, if $Q_\ell(t) = 0$ for all $1 \leqslant \ell \leqslant K_i$, (26) holds. Now we assume that $Q_{K_i}(t) = 0$, but $\sum_{1\leqslant \ell < K_i} Q_\ell(t) > 0$. Since $\sum_{\ell\in\mathscr{C}(i)} \dot{T}_\ell(t) \leqslant 1$, similar to the derivation of (19), with $k = K_i$ and $j = i$, we have

$$\mu_{K_i}\dot{T}_{K_i}(t) \leqslant \frac{1-\alpha\sum_{\ell\in\mathscr{C}(i),\ell\geqslant K_{i+1}} m_\ell}{\sum_{\ell\in\mathscr{C}(i),K_i\leqslant\ell<K_{i+1}} m_\ell} = \alpha.$$

On the other hand, let k be the last class ℓ with $1 \leqslant \ell < K_i$ and $Q_\ell(t) > 0$. Then $\mu_k \dot{T}_k(t)$ is given by (19), which is greater than or equal to α. Therefore, we have proved (26), thus completing the induction step. □

Proof of Theorem 2.1. We employ Theorem 5.3 of [2] to prove the theorem. To employ the theorem, one needs to check that the matrix R in (8) is completely-$\mathscr{S}$ and that the fluid model is uniformly convergent with lifting matrix Δ. The completely-$\mathscr{S}$ condition is satisfied due to Theorems 3.1 and 3.2 of [6]. From (15) of Theorem 3.1, the fluid model is uniformly convergent with lifting matrix Δ. Thus, Theorem 2.1 follows from Theorem 5.3 of [2]. □

Acknowledgements

This research has been supported in part by the Chinese National Science Foundation. In addition, J. G. Dai's research has been supported in part by National Science Foundation grant DMI-0300599, and by TLI-Asia Pacific, a partnership between National University of Singapore and Georgia Institute of Technology.

References

[1] M. Bramson, State space collapse with application to heavy traffic limits for multiclass queueing networks, Queueing Systems Theory Appl. 30 (1998) 89–148.
[2] M. Bramson, J.G. Dai, Heavy traffic limits for some queueing networks, Ann. Appl. Probab. 11 (2001) 49–90.
[3] H. Chen, H.Q. Ye, Existence condition for the diffusion approximations of multiclass priority queueing networks, Queueing Systems Theory Appl. 38 (4) (2001) 435–470.
[4] H. Chen, H. Zhang, A sufficient condition and a necessary condition for the diffusion approximations of multiclass queueing networks under priority service disciplines, Queueing Systems Theory Appl. 34 (1–4) (2000) 237–268.
[5] J.G. Dai, G. Weiss, Stability and instability of fluid models for re-entrant lines, Math. Oper. Res. 21 (1996) 115–134.
[6] J.G. Dai, D.H. Yeh, C. Zhou, The QNET method for re-entrant queueing networks with priority disciplines, Oper. Res. 45 (1997) 610–623.
[7] S.N. Ethier, T.G. Kurtz, Markov Processes: Characterization and Convergence, Wiley, New York, 1986.
[8] J.M. Harrison, Brownian models of queueing networks with heterogeneous customer populations, in: W. Fleming, P.L. Lions (Eds.), Stochastic Differential Systems, Stochastic Control Theory and Their Applications, The IMA volumes in mathematics and its applications, vol. 10, Springer, New York, 1988, pp. 147–186.
[9] L.M. Taylor, R.J. Williams, Existence and uniqueness of semimartingale reflecting Brownian motions in an orthant. Probab. Theory Related Fields 96 (1993) 283–317.
[10] R.J. Williams, Diffusion approximations for open multiclass queueing networks: sufficient conditions involving state space collapse, Queueing Systems Theory Appl. 30 (1998) 27–88.