OPERATIONS RESEARCH AND DECISIONS

No. 4 2017
DOI: 10.5277/ord170404

Edvinas GREIČIUS[1]
Saulius MINKEVIČIUS[1]

DIFFUSION LIMITS FOR THE QUEUE LENGTH OF JOBS IN MULTISERVER OPEN QUEUEING NETWORKS

A mathematical model of a multiserver open queueing network in heavy traffic is developed. This model is that of a multiserver computer system network in heavy traffic. A limit theorem for the length of the queue has been presented.

Keywords: *performance evaluation, multiserver open queueing network, limit theorem, queue length*

1. Introduction

Proofs of probability limit theorems (PLTs) have clear practical implications. In the article, the authors prove a PLT for the queue length (customers) in a multiserver open queueing network under heavy traffic conditions. Models of queueing networks have been extensively used for analysing the performance of manufacturing systems and transportation systems, as well as computer and communication networks. Therefore, many methods of approximation have emerged and PLT is among them. The history of investigations into diffusion approximations for queueing systems in heavy traffic is about forty years old, while the history of queueing networks is about twenty years old. Although Kolmogorov [29] had already proved in the fifties that it was possible to approximate the number of occupied phases in a queue with finite capacity by means of a diffusion process with reflection at the upper boundary, systematic study of the problem only started with the papers [26, 27, 35]. Similarly, methods of investigating single-phase (single-server) queueing systems in heavy traffic are consi-dered in [35, 21, 22, 3, 2], etc. Later on, a large number of papers were published aimed at various

[1]Institute of Mathematics and Informatics, Vilnius University, Akademijos 4, 08663 Vilnius, Lithuania, e-mail address: edvinas.greicius@gmail.com, minkevicius.saulius@gmail.com

aspects of diffusion approximations in models of queues (cf. the survey papers [19, 20, 40], and the monograph [25]). Investigations into diffusion approximations in queueing theory started in the papers [29, 17], and [36].

Due to serious technical difficulties, models of open queueing networks are not considered so often. The book of Bramson [8] contains the essentials of queueing networks from classical product-form theory to more recent developments, such as diffusion and fluid limits, stochastic comparisons, stability, dynamic scheduling, and optimization. Chen and Zhang [9] established a sufficient condition for the existence of a (conventional) diffusion approximation for multiclass queueing networks under various disciplines of ascribing priority . Chen and Ye [10] extended the work of Chen and Zhang [9] and established a new sufficient condition for the existence of a (conventional) diffusion approximation for multiclass queueing networks under various disciplines of ascribing priority. This sufficient condition is related to the weak stability of fluid networks and the stability of high priority classes of fluid networks that correspond to the queueing networks under consideration. Bramson and Dai [5] proved limit theorems under heavy traffic for six families of multiclass queueing networks (for example, the first three families are single-station systems operating under the first-in-first-out (FIFO), generalized-head-of-the-line proportional processor sharing (GHLPPS) and static buffer priority (SBP) service disciplines). Harrison [18] describes a general type of model of a stochastic system that involves three basic elements: activities, resources, and stocks of material. A system manager chooses activity levels dynamically, based on observations of the state, consuming some materials as inputs and producing other materials as outputs, subject to constraints on resource capacity. Mandelbaum and Stolyar [31] investigated a queueing system with multiple types of customers and flexible (multiskilled) servers that work in parallel. Such a system in heavy traffic is analyzed and it is shown that a very simple generalized $c\mu$ rule minimizes both instantaneous and cumulative queueing costs asymptotically, over essentially all of the scheduling disciplines, preemptive or non-preemptive. Kang et al. [24] consider a connection-level model of iInternet congestion control that represents the randomly varying number of flows present in a network. Here bandwidth is fairly shared amongst elastic document transfers according to a weighted bandwidth sharing policy. Rybko and Stolyar [37] study a stochastic network that consists of a set of servers processing multiple classes of jobs. Each class of jobs requires a concurrent occupancy of several servers while being processed, and each server is shared among the job classes according to a head-of-the-line processor-sharing mechanism.

Limit theorems and diffusion approximations for queueing systems under the conditions of heavy traffic are closely connected (they belong to the same field of research, i.e., investigations into the theory of queueing systems in heavy traffic). There is a vast literature on diffusion approximations. Readers are referred to [40, 30, 14] and [15] for a general survey of the subject. The natural setting for a limit theorem in the latter paper is the weak convergence of probability measures on the function space

$D[0,1](\equiv D)$. Since an excellent treatment of this subject has recently been published [1], we shall only make a few remarks here about our terminology and notation. Stochastic processes characterising a queueing system give rise to sequences of random functions in *D*; the space of all right-continuous functions on $[0,1]$ with left limits and endowed with a Skorohod metric, *d*. Let D be the class of Borel sets of *D*. If P_n and *P* are probability measures on D which satisfy

$$\lim_{n\to\infty}\int_D fdP_n = \int_D fdP$$

for every bounded, continuous, real-valued function *f* on *D*, then we say that P_n weakly converges to *P* as $n \to \infty$ and write $P_n \Rightarrow P$. A random function *X* is a measurable mapping from some probability space $(\Omega, \mathsf{B}, \mathsf{P})$ into *D* with the distribution $P = \mathsf{P}X^{-1}$ on (D, D). We assume that a sequence of random functions $\{X_n\}$ weakly converges to the random function *X*, and write $X_n \Rightarrow X$, if the distribution P_n of X_n converges to the distribution *P* of *X*.

We will repeatedly use an analogue of the following theorem on mutual convergence (see, for example, [1]):

Theorem 1.1. Let $\varepsilon > 0$ and $X_n, Y_n, X \in D$. If $P(\lim\limits_{n\to\infty} d(X_n, X) > \varepsilon) = 0$ and $P(\lim\limits_{n\to\infty} d(X_n, X_n) > \varepsilon) = 0$, then $P(\lim\limits_{n\to\infty} d(Y_n, X) > \varepsilon) = 0$. So, here we prove a PLT for the queue length in a multiserver open queueing network under heavy traffic conditions. The main tool for the analysis of such queueing systems under heavy traffic is a functional limit theorem for a complex renewal process (the proof can be found in [1]).

2. The model of the network

Consider a network of *j* stations, indexed by *j* = 1, 2, ..., *J*, such that any station *j* has c_j servers indexed by $(j, 1), \ldots, (j, c_j)$. First, $\{u_j(e), e \ge 1\}$, $j = 1, 2, \ldots, J$, are *J* sequences of exogenous interarrival times, where $u_j(e) \ge 0$ is the interarrival time between the $(e-1)$-th job and the *e*-th job that arrive at station *j* exogenously (from outside the network). Define

$$U_j(0) = 0,\ U_j(n) = \sum_{e=1}^{n} u_j(e),\ n \ge 1 \text{ and } A_j(t) = \sup\{n \ge 0 : U_j(n) \le t\}$$

where $A_j = \{A_j(t), t \ge 0\}$ is called the exogenous arrival process for station *j*, i.e., $A_j(t)$ counts the number of jobs that have arrived at station *j* from outside the network.

Second, $\{v_{jk_j}(e), e \geq 1\}$, $j = 1, 2, ..., J$, $k_j = 1, 2, ..., c_j$, are $c_1 + \ldots + c_J$ sequences of service times, where $v_{jk_j}(e) \geq 0$ is the service time for the e-th job served by server k_j at station j. Define

$$V_{jk_j}(0) = 0, V_{jk_j}(n) = \sum_{e=1}^{n} v_{jk_j}(e), n \geq 1, \text{ and } x_{jk_j}(t) = \sup\{n \geq 0 : V_{jk_j}(n) \leq t\},$$

where $x_{jk_j} = \{x_{jk_j}(t), t \geq 0\}$ is called the service process for server k_j at station j, i.e., $x_{jk_j}(t)$ counts the number of services completed by server k_j at station j over the periods where the server is occupied. We define $\mu_{jk_j} = (E[v_{jk_j}(e)])^{-1} > 0, \sigma_{jk_j} = D(v_{jk_j}(e)) > 0$ and $\lambda_j = \left(E\left[u_j(e)\right]\right)^{-1} > 0$, $a_j = D\left(u_j(e)\right) > 0$, $j = 1, 2, ..., k$; all of these terms are assumed to be finite.

In addition, let $\tilde{\tau}_j(t)$ be the total number of jobs routed to the j-th station of the network in the interval $[0, t]$, $\tau_j(t)$ be the total number of jobs that have been served at the j-th station of the network in the interval $[0, t]$, $\tilde{\tau}_{jk_j}(t)$ be the total number of jobs routed to the k_j-th server of the j-th station of the network in the interval $[0, t]$, let $\tau_{jk_j}(t)$ be the total number of customers whose requests have been satisfied by the k_j-th server of the j-th station of the network in the interval $[0, t]$, and $\tau_{ijk_i}(t)$ be the total number of jobs which have been served by the k_i-th server of the i-th station of the network and then routed to the k_j-th server of the j-th station of the network in the interval $[0, t]$. Let p_{ij} be the probability of a job being routed to the j-th station of the network after being served at the i-th station of the network. Denote by $p_{ijk_i}^t = \tau_{ijk_i}(t)/\tau_{ik_i}(t)$ the proportion of the number of jobs which, after service at the k_i-th server of the i-th station of the network, are routed to the j-th station of the network in the interval $[0, t]$, $i, j = 1, 2, ..., J$, $k_i = 1, ..., c_i$ and $t > 0$.

The processes of primary interest are the queue length process $Q = (Q_j)$ with $Q_j = \{Q_j(t), t \geq 0\}$, where $Q_j(t)$ indicates the number of jobs at station j at time t. We use the notation $Q_{jk_j} = \{Q_{jk_j}(t), t \geq 0\}$, where $Q_{jk_j}(t)$ indicates the number of jobs waiting to be served by server k_j of station j at time t. Clearly, we have

$$Q_j(t) = \sum_{k_i=1}^{c_j} Q_{jk_i}(t), \ \ j = 1, 2, \ ..., J$$

To be more precise, the first come, first served' (FCFS) service discipline is assumed for all *J* stations. When a job arrives at a station and finds more than one server available, it will join the server with the smallest index. Suppose that the queue of jobs at each station of such an open queueing network is continuous in both time and space. Under these assumptions, we prove a PLT for the queue length in a multiserver open queueing network. Similarly as in [32], we investigate multiserver open queueing networks in series: $v_{jk_j,n}(e)$ and $u_{j,n}(e)$ are independent identically distributed random variables describing the *n*-th network; $j = 1, 2, ..., J$, $k_j = 1, 2, ..., c_j$, $e \geq 1$ and $n \geq 1$.

Let us denote

$$\beta_{j,n} = \sum_{i=1}^{J}\sum_{k_i=1}^{c_i}\mu_{ik_i,n}p_{ij} + \lambda_{j,n} - \sum_{k_j=1}^{c_j}\mu_{jk_j,n},\ \hat{\sigma}^2_{j,n} = \sum_{i=1}^{J}\sum_{k_i=1}^{c_i}\mu^3_{ik_i,n}\sigma_{ik_i,n}p^2_{ij} + \lambda^3_{j,n}a_{j,n}$$

$$+\sum_{k_j=1}^{c_j}\mu^3_{jk_j,n}\sigma_{jk_j,n} > 0,\ j = 1, 2, ..., J,\ \beta_n = \sum_{j=1}^{J}\beta_{j,n},\ \hat{\sigma}^2_n = \sum_{j=1}^{J}\hat{\sigma}^2_{j,n}$$

Also, we define

$$\mu_{jk_j,n} = (Ev_{jk_j,n}(e))^{-1} > 0,\ \sigma_{jk_j,n} = Dv_{jk_j,n}(e) > 0,\ \lambda_{j,n} = (Eu_{j,n}(e))^{-1} > 0$$

$$a_{j,n} = Du_{j,n}(e) > 0,\ j = 1, 2, ..., J,\ k_j = 1, 2, ..., c_j$$

In this work, we assume that the following overload conditions are fulfilled

$$\sum_{i=1}^{J}\sum_{k_i=1}^{c_i}\mu_{ik_i,n}p_{ij} + \lambda_j > \sum_{k_j=1}^{c_j}\mu_{jk_j,n} \tag{1}$$

and

$$\hat{\sigma}^2_{j,n} \to \hat{\sigma}^2_j > 0,\ \ j = 1, 2, ..., J \tag{2}$$

Note that condition (1) guarantees that with probability one there exists a queue of jobs which is constantly growing.

In addition, we assume throughout that

$$\max_{1\leq j\leq J}\max_{1\leq k_j\leq c_j}\sup_{e\geq 1}E\left(v_{jk_j,n}(e)\right)^{2+\gamma} < \infty \text{ for some } \gamma > 0 \tag{3}$$

$$\max_{1\leq j\leq J}\max_{1\leq k_j\leq c_j}\sup_{e\geq 1}E\left(u_{j,n}(e)\right)^{2+\gamma} < \infty \text{ for some } \gamma > 0 \tag{4}$$

Conditions (3) and (4) imply the Lindeberg conditions for the respective sequences, and are easier to verify in practice (usually $\gamma = 1$ works).

3. Main results

Applying the results of [34], we prove the following theorem about the diffusion limit of the queue length in a multiserver open queueing network. Also, we note that conditions (1)–(4) are used in the proof of the results of [34].

Theorem 3.1. If conditions (1)–(4) are fulfilled, then

$$\left(\frac{Q_1(nt)-\beta_1 nt}{\sqrt{n}}, \frac{Q_2(nt)-\beta_2 nt}{\sqrt{n}}, \ ..., \ \frac{Q_J(nt)-\beta_J nt}{\sqrt{n}} \right)$$

$\Rightarrow (\hat{\sigma}_1 z_1(t), \hat{\sigma}_2 z_2(t), ..., \hat{\sigma}_J z_J(t)),$ where $z_j(t),\ j=1, 2, ..., J,\ 0 \le t \le 1$ are independent standard Wiener processes.

Proof. Denote $c = 2J^2(c_1 + c_2 + ... + c_J), W_j(t) = Q_j(t) - \hat{x}_j(t), j = 1, 2, ..., J$ and $t > 0$. Hence, for $\varepsilon > 0$ we obtain

$$P\left(\frac{\sup\limits_{0 \le t \le 1} |W_j(nt)|}{\sqrt{n}} > \varepsilon \right) \le P\left(\frac{\sup\limits_{0 \le t \le 1} \sup\limits_{0 \le s \le nt} \left\{ \sum_{j=1}^{J} \sum_{i=1}^{J} \sum_{k_i=1}^{c_i} x_{ik_i}(s) \cdot | p_{ijk_i}^{s} - p_{ij} | \right\}}{\sqrt{n}} > \frac{\varepsilon}{2} \right)$$

$$+P\left(\frac{\sup\limits_{0 \le t \le 1} \{ \sum_{i=1}^{J} \sum_{k_i=1}^{c_i} \sup\limits_{0 \le s \le nt} (x_{ik_i}(s) - \tau_{ik_i}(s) \}}{\sqrt{n}} > \frac{\varepsilon}{2} \right)$$

$$\le P\left(\frac{\sup\limits_{0 \le t \le 1} \left\{ \sum_{j=1}^{J} \sum_{i=1}^{J} \sum_{k_i=1}^{c_i} \sup\limits_{0 \le s \le nt} \{ x_{ik_i}(s) | p_{ijk_i}^{s} - p_{ij} | \} \right\}}{\sqrt{n}} > \frac{\varepsilon}{2} \right)$$

$$+P\left(\frac{\sum_{i=1}^{J} \sum_{k_i=1}^{c_i} \sup\limits_{0 \le t \le 1} \sup\limits_{0 \le s \le nt} (x_{ik_i}(s) - \tau_{ik_i}(s))}{\sqrt{n}} > \frac{\varepsilon}{2} \right)$$

$$\leq P\left(\frac{\sum_{j=1}^{J}\sum_{i=1}^{J}\sum_{k_i=1}^{c_i}\sup_{0\le t\le 1}\sup_{0\le s\le nt}\{x_{ik_i}(s)\,|\,p_{ijk_i}^{s}-p_{ij}\,|\}}{\sqrt{n}}>\frac{\varepsilon}{2}\right)$$

$$+P\left(\frac{\sum_{i=1}^{J}\sum_{k_i=1}^{c_i}\sup_{0\le t\le 1}\sup_{0\le s\le nt}(x_{ik_i}(s)-\tau_{ik_i}(s))}{\sqrt{n}}>\frac{\varepsilon}{2}\right)$$

$$\leq P\left(\frac{\sum_{j=1}^{J}\sum_{i=1}^{J}\sum_{k_i=1}^{c_i}\sup_{0\le s\le n}\{x_{ik_i}(s)\,|\,p_{ijk_i}^{s}-p_{ij}\,|\}}{\sqrt{n}}>\frac{\varepsilon}{2}\right)$$

$$+P\left(\frac{\sum_{i=1}^{J}\sum_{k_i=1}^{c_i}\sup_{0\le s\le n}(x_{ik_i}(s)-\tau_{ik_i}(s))}{\sqrt{n}}>\frac{\varepsilon}{2}\right)$$

$$\leq \sum_{j=1}^{J}\sum_{i=1}^{J}\sum_{k_i=1}^{c_i}P\left(\frac{\sup_{0\le s\le n}\{x_{ik_i}(s)\,|\,p_{ijk_i}^{s}-p_{ij}\,|\}}{\sqrt{n}}>\frac{\varepsilon}{c}\right)$$

$$+\sum_{i=1}^{J}\sum_{k_i=1}^{c_i}P\left(\frac{\sup_{0\le s\le n}(x_{ik_i}(s)-\tau_{ik_i}(s))}{\sqrt{n}}>\frac{\varepsilon}{c}\right) \tag{5}$$

$j=1, 2, ..., k$

Let us estimate the first term in (5). Note that for $\varepsilon>0$

$$P\left(\frac{\sup_{0\le s\le n}\{x_{ik_i}(s)\,|\,p_{ijk_i}^{s}-p_{ij}\,|\}}{\sqrt{n}}>\varepsilon\right)\leq P\left(\frac{\{x_{ik_i}(n)\sup_{0\le s\le n}|\,p_{ijk_i}^{s}-p_{ij}\,|\}}{\sqrt{n}}>\varepsilon\right)$$

$$\le P\left(\left\{\frac{\{x_{ik_i}(n)\sup\limits_{0\le s\le n}|\,p^s_{ijk_i}-p_{ij}\,|\}}{\sqrt{n}}>\varepsilon\right\}\right.$$

$$\left.\cap\left(\{x_{ik_i}(n)\le\varepsilon n+\mu_{ik_i}n\}\right)+P\left(\frac{x_{ik_i}(n)}{n}-\mu_{ik_i}>\varepsilon\right)\right.$$

$$\le P((\varepsilon n+\mu_{ik_i}n)\sup_{0\le s\le n}|\,p^s_{ijk_i}-p_{ij}\,|>\varepsilon\sqrt{n})$$

$$+P\left(\frac{x_{ik_i}(n)}{n}-\mu_{ik_i}>\varepsilon\right)\le P\left(\sup_{0\le s\le n}|\,p^s_{ijk_i}-p_{ij}\,|>\frac{\varepsilon\sqrt{n}}{\varepsilon n+\mu_{ik_i}n}\right)$$

$$+P\left(\sup_{0\le s\le n}\left|\frac{x_{ik_i}(s)}{s}-\mu_{ik_i}\right|>\varepsilon\right)\le P(\sup_{0\le s\le n}|\,p^s_{ijk_i}-p_{ij}\,|>0)$$

$$+P\left(\sup_{0\le s\le n}\left|\frac{x_{ik_i}(s)}{s}-\mu_{ik_i}\right|>\varepsilon\right),\ i=1,2,\ldots,J,\ k_i=1,2,\ldots,c_i \tag{6}$$

Finally, we find establish that for $\varepsilon>0$ (see (5), (6))

$$\lim_{n\to\infty}P\left(\frac{\sup\limits_{0\le t\le 1}|W_j(nt)|}{\sqrt{n}}>\varepsilon\right)$$

$$\le\sum_{j=1}^{J}\sum_{i=1}^{J}\sum_{k_i=1}^{c_i}\lim_{n\to\infty}P(\sup_{0\le s\le n}|\,p^s_{ijk_i}-p_{ij}\,|>0)$$

$$+\sum_{j=1}^{J}\sum_{i=1}^{J}\sum_{k_i=1}^{c_i}\lim_{n\to\infty}P\left(\sup_{0\le s\le n}\left|\frac{x_{ik_i}(s)}{s}-\mu_{ik_i}\right|>\frac{\varepsilon}{c}\right)$$

$$+\sum_{i=1}^{J}\sum_{k_i=1}^{c_i}\lim_{n\to\infty}P\left(\frac{\sup\limits_{0\le s\le n}(x_{ik_i}(s)-\tau_{ik_i}(s))}{\sqrt{n}}>\frac{\varepsilon}{c}\right)$$

$$\le\sum_{j=1}^{J}\sum_{i=1}^{J}\sum_{k_i=1}^{c_i}P(\lim_{n\to\infty}\sup_{0\le s\le n}|\,p^s_{ijk_i}-p_{ij}\,|>0)$$

$$+\sum_{j=1}^{J}\sum_{i=1}^{J}\sum_{k_i=1}^{c_i} P\left(\lim_{n\to\infty}\sup_{0\le s\le n}\left|\frac{x_{ik_i}(s)}{s}-\mu_{ik_i}\right|>\frac{\varepsilon}{c}\right)$$

$$+\sum_{i=1}^{J}\sum_{k_i=1}^{c_i} P\left(\lim_{n\to\infty}\frac{\sup_{0\le s\le n}(x_{ik_i}(s)-\tau_{ik_i}(s))}{\sqrt{n}}>\frac{\varepsilon}{c}\right), \quad j=1,2,\ldots,k \tag{7}$$

Let us prove that the first term in (7) converges to zero. Note that (see [34]),

$$P(\lim_{n\to\infty}\sup_{0\le s\le n}|p_{ijk_i}^{s}-p_{ij}|>0)$$

$$\le P(\lim_{n\to\infty}\lim_{\delta\downarrow 0}\sup_{0\le s\le n}|p_{ijk_i}^{s}-p_{ij}|>\delta)$$

$$=\lim_{\delta\downarrow 0}P\,(\lim_{n\to\infty}\sup_{0\le s\le n}|p_{ijk_i}^{s}-p_{ij}|>\delta)=0,\; i,\; j=1,2,\ldots,J \tag{8}$$

Using the limit theorem for a renewal process, we see that the second term in (7) converges to zero (see [1]). Hence, we obtain that the third term in (7) also converges to zero. Thus, we have proven that (see (7), (8))

$$P\left(\frac{\sup_{0\le t\le 1}|W_j(nt)|}{\sqrt{n}}>\varepsilon\right)\Rightarrow 0,\;\; j=1,2,\ldots,J \tag{9}$$

Note that (see, for example, [2])

$$\frac{\hat{x}_j(nt)-\beta_j nt}{\sqrt{n}}=\sum_{i=1}^{J}\sum_{k_i=1}^{c_i}\frac{(x_{ik_i}(t)-\mu_{ik_i}nt)p_{ij}}{\sqrt{n}}$$

$$+\frac{(A_j(t)-\lambda_j nt)-\left(\sum_{k_i=1}^{c_i}(x_{ik_i}(t)-\mu_{ik_i}nt)\right)}{\sqrt{n}}\Rightarrow\hat{\sigma}_j z_j(t) \tag{10}$$

where $z_j(t)$ are independent standard Wiener processes, $j=1,2,\ldots,J$, $0\le t\le 1$.

Thus, using the mutual convergence theorem (see [1]) and from (9), (10), we derive that

$$\frac{Q_j(nt)-\beta_j nt}{\sqrt{n}}\Rightarrow\hat{\sigma}_j z_j(t) \tag{11}$$

where $z_j(t)$ are independent standard Wiener processes, $j = 1, 2, ..., J$, $0 \le t \le 1$. By [32], for vector convergence to hold it suffices that a limit process is continuous in $C[0, 1]$. Using this result together with (11), we complete the proof of the theorem.

Let us denote $V(t) = \sum_{j=1}^{J} Q_j(t),\ t > 0$. Finally, we will prove the following limit theorem on the total queue length in a multiserver open queueing network.

We also assume that the following condition is satisfied:

$$\hat{\sigma}_n^2 \to \hat{\sigma}^2 \tag{12}$$

Theorem 3.2. If conditions (1)–(4), (12) are fulfilled, then

$$\frac{V(nt) - \beta_n nt}{\sqrt{n}} \Rightarrow \hat{\sigma} z(t)$$

where $z(t)$ are independent standard Wiener processes, $0 \le t \le 1$.

Proof. It suffices to note that $| V(nt) - \sum_{j=1}^{J} \hat{x}_j(nt) | \le \sum_{j=1}^{J} | W_j(nt) |$ and apply Theorem 3.1. The proof of Theorem 3.2 is complete.

An interesting situation occurs when jobs in the queue belong to different classes. This is an example of a multi-class queuing network (see fundamental works on multiclass queuing networks in heavy traffic [6, 11], etc.). In this paper, diffusion limits for the queue length are investigated in the environment of a multiserver open queuing network. The stability of multiclass queuing networks has been investigated for a long time. However, the most important advances were only made in the 2000s. One of the most important papers of this kind, [12], analyzes recent achievements in the analysis of the stability of queuing networks via fluid models. Significant aspects of stability have already been investigated by a number of authors, including [6, 42], and others. Various service disciplines are investigated in the paper [12] and that paper makes a clear distinction between local stability for a service discipline and the global stability region of a network. Generally speaking, it is known that the stability of a queuing network is closely related to the model provided in [37]. The paper [12] additionally states that for a given service discipline an open queuing network is stable if the state of the corresponding fluid model eventually converges to zero starting from any initial condition. The stability of fluid models and the latest advances presented in [12] conclude that it is not an easy task to derive the region of stability for any given service discipline. It must be mentioned that linear Lyapunov functions are by far the most powerful tool for determining the stability regions of queuing networks with different service disciplines.

However, following the latest developments in the analysis of the FCFS discipline, it is recommended to consider using the techniques provided in the paper [11].

4. Applications of the main results

First we present a theorem about the fluid approximation for the length of a queue of jobs in a multiserver open queueing network under heavy traffic conditions.

Theorem 4.1. If conditions (1)–(4) are fulfilled, then

$$\left(\frac{Q_1(t)}{t}, \frac{Q_2(t)}{t}, ..., \frac{Q_J(t)}{t} \right) \Rightarrow \left(\beta_1, \beta_2, ..., \beta_J \right), \ 0 \le t \le 1$$

Next, we present the law of the iterated logarithm for the queue length in a multiserver open queueing network under heavy traffic conditions.

Theorem 4.2. If conditions (1)–(4) are fulfilled, then

$$P\left(\overline{\lim_{t\to\infty}} \frac{Q_j(t) - \beta_j t}{\hat{\sigma}_j a(t)} = 1 \right) = P\left(\underline{\lim_{t\to\infty}} \frac{Q_j(t) - \beta_j t}{\hat{\sigma}_j a(t)} = -1 \right) = 1, \ j = 1, 2, ..., J$$

First denote

$$\beta_{jk_j} = \sum_{k_i=1}^{J} \mu_{ik_i} p_{ij} + \lambda_j - \mu_{jk_j}, \ j = 1, 2, ..., J, \, k_j = 1, 2, ..., c_j$$

Also in this section, we present the following corollary on the probability that a computer network fails due to overload.

Corollary 4.2. If $t \ge \max_{1\le j\le J} \max_{1\le k_j \le c_j} \frac{m_{jk_j}}{\beta_{jk_j}}$ and condition (1) is fulfilled, then the computer network becomes unreliable (all of the computers fail). The proof of Corollary 4.2 is presented in [38], and we omit it. The proof of Corollary 4.2 is complete.

5. Concluding remarks and future research

• If the conditions of the theorem on the PLT are fulfilled (i. e., condition (1) is satisfied), the network is occupied at first and if condition (1) is satisfied later on, the network becomes uncontrollable after a certain time, since $t \geq \max_{1 \leq j \leq J} \max_{1 \leq k_j \leq c_j} \frac{m_{jk_j}}{\beta_{jk_j}}$.

• Condition (1) is fundamental, the behaviour of the whole network and its evolution is not clear when condition (1) is not satisfied. Therefore, this fact is the object of further research and discussion.

• Note that a computer with a Windows operating system functions steadily if the number of jobs does not exceed 5 (therefore, $m_{jk_j} >= 5$). Otherwise, the computer fails.

• The theorems of this paper are proved for a class of multiserver open queueing networks in heavy traffic with the FCFS service principle, independent waiting times for customers at each node of the queueing system, and the times between the arrivals of customers of the multiserver open queueing network are independent identically distributed random variables. However, analogous theorems can be applied to a wider class of multiserver open queueing networks in heavy traffic, e.g., when there are various groups of customers in a queue, when the interarrival times of customers of the multiserver open queueing network are weakly dependent random variables.

Acknowledgements

This research has been supported by the National Complex Programme and European Union Structural funds *SMART improvement of researchers qualifications by implementing world-class R&D projects*.

References

[1] BILLINGSLEY P., *Convergence of Probability Measures*, Wiley, 1968.
[2] BOROVKOV A.A., *Asymptotic methods in queueing theory*, Nauka, Moscow 1980 (in Russian).
[3] BOROVKOV A.A., *Stochastic processes in queueing theory*, Nauka, Moscow 1972 (in Russian).
[4] BOXMA O., *On a tandem queueing model with identical service times at both counters*, *I, II*, Adv. Appl. Prob., 1979, 11 (3), 616–643, 644–659.
[5] BRAMSON M., DAI J.G., *Heavy traffic limits for some queueing networks*, Ann. Appl. Prob., 2001, 11 (1), 49–90.
[6] BRAMSON M., *State space collapse with application to heavy traffic limits for multiclass queuing networks*, Queuing Sys., 1998, 30 (1, 2), 89–140.
[7] CHEN H., MALDELBAUM A., *Stochastic discrete flow networks. Diffusion approximations and bottlenecks*, Ann. Appl. Prob., 1991, 19 (4), 1463–1519.
[8] CHEN H., YAO D.D., *Fundamentals of Queueing Networks. Performance, Asymptotics and Optimization*, Springer-Verlag, New York 2001.

[9] CHEN H., ZHANG H., *A sufficient condition and a necessary condition for the diffusion approximations of multiclass queueing networks under priority service disciplines*, Queing Syst., Theory Appl., 2000, 34 (1–4), 237–268.

[10] CHEN H., YE H.Q., *Existence condition for the diffusion approximations of multiclass priority queueing networks*, Queueing Syst., Theory Appl., 2001, 38, 435–470.

[11] CHEN H., *Fluid approximations and stability of multiclass queuing networks: work-conserving disciplines*, Ann. Applied Probability., 1995, 5 (3), 637–665.

[12] DAI J.G., *Stability of open multiclass queuing networks via fluid models*, F. Kelly, R. Williams (Eds.), Proc. IMA Workshop on Stochastic Networks, Springer-Verlag, New York 2000, 100–121.

[13] DAI J.G., *On positive Harris recurrence of multiclass queueing networks. A unified approach via unified fluid limit models*, Ann. Appl. Prob., 1995, 5 (1), 49–77.

[14] FLORES C., *Diffusion approximations for computer communications networks*, [In:] B. Gopinath (Ed.), Comp. Comm., Proc. Syrup. Appl. Math., American Mathematical Society, 1985, 83–124.

[15] GLYNN E.M., *Diffusion approximations*, [In:] D.P. Heyman, M.J. Sobel (Eds.), *Handbooks in Operations Research and Management Science*, Vol. 2. *Stochastic Models*, Elsevier, 1990.

[16] HARRISON J.M., LEMOINE A.J., *A note on networks of infinite-server queues*, J. Appl. Prob., 1981, 18, 561–567.

[17] HARRISON J.M., *Brownian Motion and Stochastic Flow Processes*, Wiley, New York 1985.

[18] HARRISON J.M., *A broader view of Brownian networks*, Ann. Appl. Prob., 2003, 13 (3), 1119–1150.

[19] HARRISON J.M., NQUYEN V., *Brownian models of multiclass queueing networks: current status and open problems*, Queueing Syst., Theory Appl., 1993, 13, 5–40.

[20] IGLEHART D., *Weak convergence in queueing theory*, Adv. Appl. Prob., 1973, 5, 570–594.

[21] IGLEHART D., WHITT W., *Multiple channel queues in heavy traffic*, Adv. Appl. Prob., 1970, 2, 150–177.

[22] IGLEHART D., WHITT W., *Multiple channel queues in heavy traffic. II. Sequences, networks and batches*, Adv. Appl. Prob., 1970, 2, 355–369.

[23] JOHNSON D.P., *Diffusion Approximations for Optimal Filtering of Jump Processes and for Queueing Networks*, PhD Diss., University of Wisconsin, 1983.

[24] KANG W.N., KELLY F.P., LEE N.H., WILLIAMS R.J., *State space collapse and diffusion approximation for a network operating under a fair bandwidth sharing policy*, Ann. Appl. Probab., 2009, 19 (5), 1719–1780.

[25] KARPELEVICH F., KREININ A., *Heavy traffic limits for multiphase queues*, Amer. Math. Soc., Providence, Rhode Island, 1994.

[26] KINGMAN J., *On queues in heavy traffic*, J. R. Statist. Soc., 1962, 24, 383–392.

[27] KINGMAN J., *The single server queue in heavy traffic*, Proc. Camb. Phil. Soc., 1962, 57, 902–904.

[28] KOBYASHI H., *Application of the diffusion approximation to queueing networks*, J. ACM, 1974, 21, 316–328.

[29] KOLMOGOROV A.N., *Foundations of the Theory of Probability*, 2nd Ed., Chelsea Publishing Company, New York 1956.

[30] LEMOINE A.J., *Network of queues. A survey of weak convergence results*, Manage. Sci., 1978, 24, 1175–1193.

[31] MANDELBAUM A., STOLYAR A.L., *Scheduling flexible servers with convex delay costs: heavy-traffic optimality of the generalized $c\mu$ rule*, Oper. Res., 2004, 52, 6, 836–855.

[32] MINKEVIČIUS S., *Weak convergence in multiphase queues*, Liet. Mat. Rink., 1986, 26, 717–722 (in Russian).

[33] MINKEVIČIUS S., *Complex transient phenomena in multiphase queueing systems*, Liet. Mat. Rink., 1999, 39 (3), 343–356 (in Russian).

[34] MINKEVIČIUS S., *On the law of the iterated logarithm in multiserver open queueing networks*, Stochastics, 2013, 86 (1), 46–59.
[35] PROHOROV Y., *Transient phenomena in queues*, Liet. Mat. Rink., 1963, 3, 199–206 (in Russian).
[36] REIMAN M., *Open queueing networks in heavy traffic*, Math. Oper. Res., 1984, 9, 441–459.
[37] RYBKO A.N., STOLYAR A.L., *Ergodicity of stochastic processes describing the operation of open queuing networks*, Probl. Peredachi Inf., 1992, 28 (3), 3–26, Probl. Inf. Trans., 1992, 28 (3), 199–220.
[38] SAKALAUSKAS L., MINKEVIČIUS S., *On the law of the iterated logarithm in open queueing networks*, Eur. J. Oper. Res., 2000, 120, 632–640.
[39] STRASSEN V., *An invariance principle for the law of the iterated logarithm*, Z. Wahrscheinlichkeitstheorie Verwandte Gebiete, 1964, 3, 211–226.
[40] WHITT W., *Heavy traffic limit theorems for queues: a survey*, [In:] *Lecture Notes in Economics and Mathematical Systems*, Springer, Berlin, 1974, 98, 307–350.
[41] WHITT W., *On the heavy-traffic limit theorem for GIG/∞ queues*, Adv. Appl. Prob., 1982, 14, 171–190.
[42] WHITT W., *Large fluctuations in a deterministic multiclass network of queues*, Manage. Sci., 1993, 39, 1020–1028.

Received 5 March 2017
Accepted 25 November 2017