

Laws of Little in an open queueing network

Saulius Minkevičius^{a,b}, Stasys Steišūnas^a

^aInstitute of Mathematics and Informatics, Vilnius University
Akademijos str. 4, LT-08663 Vilnius, Lithuania
minkevicius.saulius@gmail.com

^bFaculty of Mathematics and Informatics, Vilnius University
Naugarduko str. 24, LT-03225 Vilnius, Lithuania

Received: 10 October 2011 / **Revised:** 29 March 2012 / **Published online:** 5 July 2012

Abstract. The object of this research in the queueing theory is theorems about the functional strong laws of large numbers (FSLN) under the conditions of heavy traffic in an open queueing network (OQN). The FSLN is known as a fluid limit or fluid approximation. In this paper, FSLN are proved for the values of important probabilistic characteristics of the OQN investigated as well as the virtual waiting time of a customer and the queue length of customers. As applications of the proved theorems laws of Little in OQN are presented.

Keywords: queueing system, open queueing network, heavy traffic, the virtual waiting time of a customer, the queue length of a customers, a law of Little.

1 Introduction

The paper is devoted to the analysis of queueing systems in the context of the network and communications theory. We investigate FSLN about the virtual waiting time of a customer and the queue length of customers and theorems on the laws of Little in an OQN under the conditions of heavy traffic. Queueing networks have been extensively used for the analysis of manufacturing systems, transportation systems, and computer and communications networks. Therefore, many approximation methods have emerged, and FSLN are among them.

The investigation of delays arising in communications and computer systems is a very complicated problem which has not yet been solved in the general case. A valuable progress in this area has been achieved for models based on Kleinrock's hypothesis on the independence of transmission times of messages at different nodes (see [1–7]). The most fruitful approach to the calculation of delays in communication systems is based on limit theorems for heavy traffic and light traffic regimes. Some results have been obtained both for systems with finite (see [6, 8]) and infinite (see [2, 7, 9–19]) waiting rooms. The theoretical base for heavy traffic limit theorems includes the weak convergence results for stochastic processes (see [20–23]) as well as the martingale approach to limit theorems (see [24–26]). The limit theorems usually state that, in the heavy traffic regime,

properly normalized random variables (or random processes) describing queue lengths or waiting times converge in distribution to a normal random variable (or certain diffusion processes).

The first results for OQN in heavy traffic were obtained by Iglehart and Whitt [14, 15]. They considered a single station, multiserver and acyclic networks of queues. The limit process for networks in heavy traffic was described as a complicated functional of multidimensional Brownian motion. Harrison [10] considered the heavy traffic approximation to the stationary distribution of the waiting times in single server queues in series. His limit process was also given as a complicated functional of Brownian motion. In [11], Harrison again considered the diffusion approximation for tandem queues, described the limit process and found analytical solutions in several special cases. Reiman [27] proved the heavy traffic limit theorems for the queue length process associated with open queueing networks. These theorems state that the limit process is a reflected Brownian motion on the nonnegative orthant with constant directions for each boundary hyperplane. Harrison and Reiman [12] considered the properties of distribution of the multidimensional reflected Brownian motion. Harrison and Williams [13] also analysed Brownian models of open queueing networks with homogeneous customer populations. Reiman [16] studied a multiclass feedback queue in heavy traffic. A network of priority queues with one bottleneck station in heavy traffic was considered by Reiman and Simon [17]. One may note that the theory of heavy traffic analysis is rather well developed for systems that satisfy the Kleinrock hypothesis. Without this hypothesis the complexity of the problem dramatically increases.

Now let us review the latest achievements in open queueing networks theory after 2000. The book by Chen and Yao [28] contains the essentials of queueing networks from the classical product-form theory to more recent developments such as diffusion and fluid limits, stochastic comparisons, stability, dynamic scheduling, and optimization. The authors Chen and Zhang of paper [29] have established a sufficient condition for the existence of the (conventional) diffusion approximation to multiclass queueing networks under priority service disciplines. In [30], the authors Chen and Ye have extended the work of Chen and Zhang [29] and established a new sufficient condition for the existence of the (conventional) diffusion approximation to multiclass queueing networks under priority service disciplines. This sufficient condition is related to the weak stability of the fluid networks and the stability of high priority classes of fluid networks that correspond to the queueing networks under consideration. Using a slight modification method, the authors Bramson and Dai of [31] prove heavy traffic limit theorems for six families of multiclass queueing networks (for example, the first three families are single-station systems operating under first-in-first-out (FIFO) principle, generalized-head-of-the-line proportional processor sharing (GHLPPS) and static buffer priority (SBP) service disciplines). The paper of Harrison [32] describes a general type of the stochastic system model that involves three basic elements: activities, resources, and stocks of material. The system manager chooses activity levels dynamically based on state observations, consuming some materials as inputs and producing other materials as outputs, subject to resource capacity constraints. In [33], Mandelbaum and Stolyar investigate a queueing system with multitype customers and flexible (multiskilled) servers that work in parallel,

the 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 scheduling disciplines, preemptive or non-preemptive. In [34], Kang, Kelly, Lee and Williams consider a connection-level model of Internet congestion control that represents a 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. Ye and Yao in [35] 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 in a head-of-the-line processor-sharing mechanism.

Limit theorems (diffusion approximations) and the FSLLN for the queueing system under the conditions of heavy traffic are closely connected (they belong to the same field of research, i.e., investigations on the theory of queueing systems in heavy traffic). Therefore, first we shall try to trace the development of research on the general theory of a queueing system in heavy traffic. There is a vast literature on the diffusion approximation. Readers are referred to [7, 36, 37] and [38] for a general survey of the subject. The present work extends the studies of Iglehart and Whitt [14, 15] about a single station of multiserver queues, and of Reiman [27], Johnson [39], Chen and Mandelbaum [40] about networks of single server queues. Other closely related papers are by Harrison and Lemoine [41] on networks of infinite server queues, and Whitt [42] on a $GI/G/\infty$ queue.

The natural setting for functional limit theorems in this paper is the weak convergence of probability measures on the function space $D[0, 1]$. Since an excellent treatment of this subject has been recently published by Billingsley [20], we shall only make a few remarks here about our terminology and notation. Stochastic processes characterizing the queueing system give rise to sequences of random functions in D , the space of all right-continuous functions on $[0, 1]$ having left limits and endowed with Skorokhod metric, d . In [20], this metric is denoted by d_0 . With d , D is a complete, separable metric space. Let \mathcal{D} be the class of Borel sets of D . Then, if P_n and P are probability measures on \mathcal{D} which satisfy

$$\lim_{n \rightarrow \infty} \int_D f dP_n = \int_D f dP$$

for every bounded, continuous, real-valued function f on D , we say that P_n *weakly converges* to P , as $n \rightarrow \infty$, and write $P_n \Rightarrow P$. A *random function* X is a measurable mapping from some probability space $(\Omega, \mathcal{B}, \mathcal{P})$ into D having the *distribution* $P = \mathcal{P}X^{-1}$ on (D, \mathcal{D}) . We say 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 . A sequence of random functions $\{X_n\}$ *weakly converges* to X in probability if X_n and X are defined on a common domain and for all $\varepsilon > 0$, $\mathbf{P}\{d(X_n, X) \geq \varepsilon\} \rightarrow 0$. When X is a constant function (not random), the convergence in probability is equivalent to a weak convergence. In such cases, we write $d(X_n, X) \Rightarrow 0$ or $X_n \Rightarrow X$. If X_n and Y_n have a common domain, we also write $d(X_n, Y_n) \Rightarrow 0$, $\mathbf{P}\{d(X_n, Y_n) > \varepsilon\} \rightarrow 0$ as for all $\varepsilon > 0$. We also use the uniform metric ρ which is

defined by $\rho(x, y) = \sup_{0 \leq t \leq 1} |x(t) - y(t)|$ for $x, y \in D$. Also, note that $d(x, y) \leq \rho(x, y)$ for $x, y \in D$.

Next, we state two extremely useful theorems for obtaining weak convergence results in applications. The first has come to be known as the “converging together theorem”. For it we assume that X_n and Y_n are defined on a common domain and take values in a separable metric space (S, m) . This result can be found in [20, Thm. 4.1].

Theorem 1. *If $X_n \Rightarrow X$ and $d(X_n, Y_n) \Rightarrow 0$, then $Y_n \Rightarrow X$.*

Now, suppose h is a measurable mapping of S into S' , a second metric space with Borel sets \mathcal{B} . Each probability measure P on (S, \mathcal{B}) induces on (S', \mathcal{B}') a unique probability measure $Ph^{-1}(A) = P(h^{-1}A)$ for $A \in \mathcal{B}'$. Let D_h be a set of discontinuities of h . The next result, known as the continuous mapping theorem, is an analog of the Mann–Wald theorem for Euclidean spaces (see [20, Thm. 5.1]). Define $h \circ X = h(X)$, $X \in D$.

Theorem 2. *If $X_n \Rightarrow X$ and $\mathbf{P}\{X \in D_h\} = 0$, then $h \circ X_n \Rightarrow h \circ X$.*

In practice we use this result as follows. First we show $X_n \Rightarrow X$, often by just quoting the known results. Then, we find an appropriate mapping h which gives us the random elements we are really interested in, $h \circ X_n$, and finally, we apply Theorem 2.

In this paper, we investigate an OQN model in heavy traffic. We present the FSLLN about the virtual waiting time of a customer and the queue length of customers in an OQN.

2 The network model

First we consider OQN with the “first come, first served” service discipline at each station and general distributions of interarrival and service time. The basic components of the queueing network are arrival processes, service processes, and routing processes. Particularly, there are k nodes with the node j having a single server and a waiting room of unlimited capacity. The external input stream to the node j is a renewal process, the interarrival time of this process with the mean $\lambda_j = (\mathbf{E}[z_n^{(j)}])^{-1} > 0$, and finite variance $a_j = \mathbf{D}z_n^{(j)} > 0$, $j = 1, 2, \dots, k$. These external input streams at the various nodes are assumed to be independent. The service times at the node j are independent and have a common distribution with the mean $\mu_j = (\mathbf{E}[S_n^{(j)}])^{-1} > 0$ and finite variance $\sigma_j = \mathbf{D}S_n^{(j)} > 0$, $j = 1, 2, \dots, k$. The service times at the node j are also independent of all customer arrivals at the node j . A customer leaving the node j is immediately and independently routed to the node i with probability p_{ji} ; and the customer departs the system from the node j with probability $p_j = 1 - \sum_{i=1}^k p_{ji}$. The $k \times k$ matrix $\mathbf{P} = (p_{ij})$ is assumed to have a spectral radius strictly smaller than a unit (see [27]). Note that this system is quite general, encompassing the tandem system, acyclic networks of $GI/G/1$ queues, and networks of $GI/G/1$ queues with feedback.

In the context of the queueing network, the random variables $z_n^{(j)}$ function as interarrival times (from outside the network) at the station j , while $S_n^{(j)}$ is the n th service time

at the station j , and $\Phi_n^{(j)}$ is a routing indicator for the n th customer served at the station j . If $\Phi_n^{(i)} = j$ (which occurs with probability p_{ij}), then the n th customer, served at the station i , is routed to the station j . When $\Phi_n^{(i)} = 0$, the associated customer leaves the network. The matrix P is called a routing matrix.

To construct renewal processes, generated by the interarrival and service times, we assume $z_j(0) = 0$, $z_j(l) = \sum_{m=1}^l z_m^{(j)}$, $S_j(0) = 0$, $S_j(l) = \sum_{m=1}^l S_m^{(j)}$, $l \geq 1$, $j = 1, 2, \dots, k$. We now define

$$a_j(t) = \max(l \geq 0: z_j(l) \leq t), \quad x_j(t) = \max(l \geq 0: S_j(l) \leq t)$$

and denote by $\tau_j(t)$ the total number of customer service departure from the j th station of the network until time t , by $\tilde{\tau}_j(t)$ the total number of customer arrival at the j th station of the network until time t , by $\tau_{ij}(t)$ the total number of customers after service that depart from the i th station of the network and arrive at the j th station of the network until time t , and by $p_{ij}^t = \tau_{ij}(t)/\tau_i(t)$ a part of the total number of customers which, after service at the i th station of the network, arrive at the j th station; $i, j = 1, 2, \dots, k$ and $t > 0$.

At first, let us define $Q_j(t)$ as the queue length of customers at the j th station of the queueing network at time t , $\beta_j = \lambda_j + \sum_{i=1}^k \mu_i p_{ij} - \mu_j > 0$, $\bar{\sigma}_j^2 = (\lambda_j)^3 \mathbf{D}z_n^{(j)} + \sum_{i=1}^k (\mu_i)^3 \mathbf{D}S_n^{(i)}(p_{ij})^2 + (\mu_j)^3 \mathbf{D}S_n^{(j)} > 0$, $j = 1, 2, \dots, k$.

Next, let us denote by $W_j(t)$ the virtual waiting time of a customer in the j th station of the queueing network at time t (time one must wait until a customer arrives at the j th station of the queueing network to be served at time t), $\hat{\beta}_j = (\lambda_j + \sum_{i=1}^k \mu_i p_{ij})/\mu_j - 1$, $\hat{\sigma}_j^2 = \sigma_j \{ \sum_{i=1}^k \mu_i p_{ij} + \lambda_j \} + \mu_j^{-2} \{ \sum_{i=1}^k \mu_i^3 p_{ij} / \sigma_i + \lambda_j^3 / a_j \}$, $j = 1, 2, \dots, k$.

Also, let us denote by $S_j(t)$ the time, which can be a sum of summary times of served customers, arriving at the j th station until time t ; $S_j(t) = \sum_{l=1}^{\tilde{\tau}_j(t)} S_l^{(j)}$, $y_j(t) = S_j(t) - t$, $\hat{x}_j(t) = \sum_{i=1}^k x_i(t)p_{ij} + a_j(t)$, $\tilde{x}_j(t) = \sum_{i=1}^k x_i(t)p_{ij}^t + a_j(t)$, $\hat{S}_j(t) = \sum_{l=1}^{\hat{x}_j(t)} S_l^{(j)}$, $\hat{y}_j(t) = \hat{S}_j(t) - t$, $f_t(y(\cdot)) = y(t) - \inf_{0 \leq s \leq t} y(s)$, $\hat{W}_j(t) = f_t(\hat{y}_j(\cdot))$, $v_j(t) = \sup_{0 \leq s \leq t} |\tilde{\tau}_j(s) - \hat{x}_j(s)|$, $j = 1, 2, \dots, k$ and $t > 0$. In addition, let $[x]$ be the integer part of number x . Note that, if $S_j(0) = V_j(0) = 0$ (OQN at time 0 is empty), then $W_j(t) = f_t(y_j(\cdot))$, $j = 1, 2, \dots, k$ (see, for example, [1, p. 41]).

Assume that the following conditions are fulfilled:

$$\lambda_j + \sum_{i=1}^k \mu_i p_{ij} > \mu_j, \quad j = 1, 2, \dots, k. \tag{1}$$

Note that these conditions guarantee that, with probability one there exists a queue length of customers and this queue length of customers is constantly growing. In addition, we assume throughout that

$$\max_{1 \leq j \leq k} \sup_{n \geq 1} \mathbf{E} \{ (z_n^{(j)})^{2+\varepsilon} \} < \infty \quad \text{for some } \varepsilon > 0, \tag{2}$$

$$\max_{1 \leq j \leq k} \sup_{n \geq 1} \mathbf{E} \{ (S_n^{(j)})^{2+\varepsilon} \} < \infty \quad \text{for some } \varepsilon > 0. \tag{3}$$

Conditions (2) and (3) imply the Lindeberg condition for the respective sequences, and are easier to verify in practice (usually $\varepsilon = 1$ works).

3 On a fluid limit of the virtual time of a customer in an OQN

At first we will prove the following theorem for a fluid limit of the virtual time of a customer in an OQN.

Theorem 3. *If conditions (1)–(3) are fulfilled, then*

$$\left(\frac{W_1(t)}{t}, \frac{W_2(t)}{t}, \dots, \frac{W_k(t)}{t} \right) \Rightarrow (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k).$$

Proof. Define $B_j(t)$ as the total amount of time that a server is busy at the station j in the interval $[0, t]$, $I_j(t) = t - B_j(t)$, $j = 1, 2, \dots, k$ and $t > 0$. Thus, we prove that

$$\frac{\sup_{0 \leq s \leq t} (x_j(s) - \tau_j(s))}{t} \Rightarrow 0, \quad j = 1, 2, \dots, k. \quad (4)$$

Note that

$$x_j(B_j(t)) \leq \tau_j(t) < x_j(B_j(t)) + 1, \quad j = 1, 2, \dots, k.$$

Also, note that $x_j(s) - \tau_j(s) \geq 0$, $j = 1, 2, \dots, k$. Thus, we have that

$$0 \leq \frac{x_j(t) - \tau_j(t)}{t} \leq \frac{x_j(t) - x_j(B_j(t))}{t} \leq \frac{x_j(t - B_j(t))}{t} = \frac{x_j(I_j(t))}{t},$$

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

Let us prove that, if conditions (1) are fulfilled, then

$$\frac{x_j(I_j(t))}{t} \Rightarrow 0, \quad j = 1, 2, \dots, k. \quad (5)$$

So, we have for $\varepsilon > 0$ and $y = \varepsilon^2 t$

$$\begin{aligned} & \mathbf{P} \left(\frac{x_j(I_j(t))}{t} > \varepsilon \right) \\ &= \mathbf{P} \left(\left\{ \frac{x_j(I_j(t))}{t} > \varepsilon \right\} \cap \{I_j(t) \leq \varepsilon^2 t\} \right) + \mathbf{P} \left(\frac{I_j(t)}{t} > \varepsilon \right) \\ &\leq \mathbf{P} \left(\left\{ \frac{x_j(\varepsilon^2 t)}{t} > \varepsilon \right\} \cap \{I_j(t) \leq \varepsilon^2 t\} \right) + \mathbf{P} \left(\frac{I_j(t)}{t} \geq \varepsilon^2 \right) \\ &\leq \mathbf{P} \left(\frac{x_j(\varepsilon^2 t)}{t} > \varepsilon \right) + \mathbf{P} \left(\frac{I_j(t)}{t} \geq \varepsilon^2 \right) \\ &\leq \mathbf{P} \left(\frac{x_j(y)}{y} > \frac{1}{\varepsilon} \right) + \mathbf{P} \left(\frac{I_j(t)}{t} \geq \varepsilon^2 \right), \quad j = 1, 2, \dots, k. \end{aligned} \quad (6)$$

Applying the Chebyshev inequality, we estimate the first item in inequality (6)

$$\mathbf{P}\left(\frac{x_j(y)}{y} > \frac{1}{\varepsilon}\right) \leq \mathbf{E}\left|\frac{x_j(y)}{y}\right| \varepsilon. \quad (7)$$

But, for $y > 0$, $\mathbf{E}|x_j(y)/y| < \infty$, so for every $\varepsilon_1 > 0$,

$$\mathbf{P}\left(\frac{x_j(y)}{y} > \frac{1}{\varepsilon}\right) \leq \varepsilon_1, \quad j = 1, 2, \dots, k. \quad (8)$$

Thus, for $\varepsilon > 0$ (see (8)),

$$\lim_{t \rightarrow \infty} \mathbf{P}\left(\frac{x_j(y)}{y} > \frac{1}{\varepsilon}\right) = \lim_{y \rightarrow \infty} \mathbf{P}\left(\frac{x_j(y)}{y} > \frac{1}{\varepsilon}\right) = 0, \quad j = 1, 2, \dots, k. \quad (9)$$

Applying the proposition from the paper of Reiman [27], we see that

$$\frac{I_j(t)}{t} \Rightarrow 0, \quad j = 1, 2, \dots, k. \quad (10)$$

Thus, the first term in (6) converges to zero (see (8)). The second term in (6) also converges to zero (see (10)). So, applying (9)–(10), we obtain that, for $\varepsilon > 0$,

$$\lim_{t \rightarrow \infty} \mathbf{P}\left(\frac{x_j(I_j(t))}{t} > \varepsilon\right) = 0, \quad j = 1, 2, \dots, k.$$

Thus, we have proved (4). Next, note that

$$|W_j(t) - \hat{y}_j(t)| \leq |W_j(t) - \hat{W}_j(t)| + |\hat{W}_j(t) - \hat{y}_j(t)|.$$

So, we achieve that, for $\varepsilon > 0$,

$$\begin{aligned} & \mathbf{P}\left(\frac{\sup_{0 \leq s \leq t} |W_j(s) - \hat{y}_j(s)|}{t} > \varepsilon\right) \\ & \leq \mathbf{P}\left(\frac{\sup_{0 \leq s \leq t} | - \inf_{0 \leq s \leq l} \hat{y}_j(s) |}{t} > \frac{\varepsilon}{2}\right) \\ & \quad + \mathbf{P}\left(\frac{\sup_{0 \leq l \leq t} \sup_{0 \leq s \leq l} |\hat{y}_j(s) - y_j(s)|}{t} > \frac{\varepsilon}{4}\right) \\ & \leq \mathbf{P}\left(\frac{\sup_{0 \leq s \leq t} |\hat{y}_j(s) - y_j(s)|}{t} > \frac{\varepsilon}{4}\right) \\ & \quad + \mathbf{P}\left(\frac{\sup_{0 \leq l \leq t} | - \inf_{0 \leq s \leq l} \hat{y}_j(s) |}{t} > \frac{\varepsilon}{2}\right), \quad j = 1, 2, \dots, k. \end{aligned} \quad (11)$$

If conditions (1) are fulfilled (see [43]), then

$$\frac{\inf_{0 \leq s \leq t} y_j(s)}{\sqrt{t}} \Rightarrow 0, \quad j = 1, 2, \dots, k. \quad (12)$$

But applying Theorem 2 to $|\cdot|$ and the supremum function, we get (see, for example, [20])

$$\frac{\sup_{0 \leq l \leq t} |\inf_{0 \leq s \leq l} y_j(s)|}{\sqrt{t}} \Rightarrow 0, \quad j = 1, 2, \dots, k. \quad (13)$$

So, the second term in (11) converges to zero. Finally, we prove that

$$\frac{\sup_{0 \leq s \leq t} |\hat{y}_j(s) - y_j(s)|}{t} \Rightarrow 0, \quad j = 1, 2, \dots, k. \quad (14)$$

We obtain for $\varepsilon > 0$

$$\begin{aligned} & \mathbf{P} \left(\frac{\sup_{0 \leq s \leq t} |\hat{y}_j(s) - y_j(s)|}{t} > \varepsilon \right) \\ & \leq \mathbf{P} \left(\frac{\sup_{0 \leq s \leq t} \sum_{m=[\hat{x}_j(s)]-v_j(s)}^{[\hat{x}_j(s)]+v_j(s)} S_m^{(j)}}{t} > \varepsilon \right) \\ & = \mathbf{P} \left(\frac{\sup_{0 \leq s \leq t} \sum_{m=[\hat{x}_j(s)]-v_j(s)}^{[\hat{x}_j(s)]+v_j(s)} (S_m^{(j)} - \frac{1}{\mu_j})}{t} > \frac{\varepsilon}{2} \right) + \mathbf{P} \left(\frac{v_j(t)}{t} > \frac{\mu_j \varepsilon}{2} \right), \quad (15) \end{aligned}$$

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

In order that the first and second term in inequality (15) converge to zero (see [43]), it suffices that

$$\lim_{t \rightarrow \infty} \mathbf{P} \left(\frac{v_j(t)}{t} > \varepsilon \right) = 0, \quad j = 1, 2, \dots, k. \quad (16)$$

Note that $\tilde{\tau}_j(t) = \sum_{i=1}^k \tau_i(t) p_{ij}^t + a_j(t)$, $j = 1, 2, \dots, k$. So, we have for $\varepsilon > 0$ that

$$\begin{aligned} & \mathbf{P} \left(\frac{v_j(t)}{t} > \varepsilon \right) \\ & \leq \mathbf{P} \left(\frac{\sup_{0 \leq s \leq t} |\tilde{\tau}_j(s) - \tilde{x}_j(s)|}{t} > \frac{\varepsilon}{2} \right) + \mathbf{P} \left(\frac{\sup_{0 \leq s \leq t} |\tilde{x}_j(s) - \hat{x}_j(s)|}{t} > \frac{\varepsilon}{2} \right), \quad (17) \end{aligned}$$

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

Next, we derive that, for $\varepsilon > 0$,

$$\begin{aligned} & \mathbf{P} \left(\frac{\sup_{0 \leq s \leq t} |\tilde{\tau}_j(s) - \tilde{x}_j(s)|}{t} > \frac{\varepsilon}{2} \right) \\ & \leq \mathbf{P} \left(\frac{\sup_{0 \leq s \leq t} \sum_{i=1}^k (x_i(s) - \tau_i(s))}{t} > \frac{\varepsilon}{2} \right) \\ & \leq \sum_{i=1}^k \mathbf{P} \left(\frac{\sup_{0 \leq s \leq t} (x_i(s) - \tau_i(s))}{t} > \frac{\varepsilon}{2k} \right) = 0, \quad j = 1, 2, \dots, k. \quad (18) \end{aligned}$$

Hence, we obtain that, for $\varepsilon > 0$ (see (4)),

$$\lim_{t \rightarrow \infty} \mathbf{P} \left(\frac{\sup_{0 \leq s \leq t} |\tilde{\tau}_j(s) - \tilde{x}_j(s)|}{t} > \varepsilon \right) = 0, \quad j = 1, 2, \dots, k. \quad (19)$$

Now we prove that for $\varepsilon > 0$

$$\mathbf{P}\left(\frac{\sup_{0 \leq s \leq t} |\tilde{x}_j(s) - \hat{x}_j(s)|}{t} > \varepsilon\right) = 0, \quad j = 1, 2, \dots, k. \quad (20)$$

For $\varepsilon > 0$, we get that

$$\begin{aligned} & \mathbf{P}\left(\frac{\sup_{0 \leq s \leq t} |\tilde{x}_j(s) - \hat{x}_j(s)|}{t} > \varepsilon\right) \\ & \leq \mathbf{P}\left(\sum_{i=1}^k x_i(t) \sup_{0 \leq s \leq t} |p_{ij}^s - p_{ij}| > \varepsilon t\right) \\ & \leq \sum_{i=1}^k \mathbf{P}\left(\left\{x_i(t) \sup_{0 \leq s \leq t} |p_{ij}^s - p_{ij}| > \frac{\varepsilon}{2k}\right\} \cap \left\{\frac{x_i(t) - \mu_i t}{t} \leq \varepsilon\right\}\right) \\ & \quad + \mathbf{P}\left(\frac{x_i(t) - \mu_i t}{t} > \varepsilon\right) \\ & \leq \sum_{i=1}^k \left\{\mathbf{P}\left(\{\varepsilon t + \mu_i t\} \sup_{0 \leq s \leq t} |p_{ij}^s - p_{ij}| > \frac{\varepsilon t}{k}\right) + \mathbf{P}\left(\frac{x_i(t) - \mu_i t}{t} > \varepsilon\right)\right\} \\ & \leq \sum_{i=1}^k \left\{\mathbf{P}\left(\sup_{0 \leq s \leq t} |p_{ij}^s - p_{ij}| > \frac{\varepsilon t}{\varepsilon t + \mu_i t}\right) + \mathbf{P}\left(\frac{x_i(t) - \mu_i t}{t} > \varepsilon\right)\right\} \\ & \leq \sum_{i=1}^k \left\{\mathbf{P}\left(\sup_{0 \leq s \leq t} |p_{ij}^s - p_{ij}| > \frac{\varepsilon}{\varepsilon + \mu_i}\right) + \mathbf{P}\left(\frac{\sup_{0 \leq s \leq t} (x_i(s) - \mu_i s)}{t} > \varepsilon\right)\right\}, \end{aligned}$$

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

Hence, we obtain for $\varepsilon > 0$

$$\begin{aligned} & \mathbf{P}\left(\frac{\sup_{0 \leq s \leq t} |\tilde{x}_j(s) - \hat{x}_j(s)|}{t} > \varepsilon\right) \\ & \leq \left\{\mathbf{P}\left(\frac{\sup_{0 \leq s \leq t} (x_i(s) - \mu_i s)}{t} > \varepsilon\right) + \sum_{i=1}^k \mathbf{P}\left(\sup_{0 \leq s \leq t} |p_{ij}^s - p_{ij}| > \frac{\varepsilon}{\varepsilon + \mu_i}\right)\right\}, \quad (21) \end{aligned}$$

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

Denote $\mu = \max_{1 \leq i \leq k} \mu_i$, $\varepsilon_2 = \varepsilon / (\varepsilon + \mu)$. Using the lemma from [44], we obtain

$$\begin{aligned} & \lim_{t \rightarrow \infty} \mathbf{P}\left(\sup_{0 \leq s \leq t} |p_{ij}^s - p_{ij}| > \varepsilon_2\right) \\ & \leq \lim_{t \rightarrow \infty} \mathbf{P}\left(\limsup_{\varepsilon_2 \downarrow 0} \sup_{0 \leq s \leq t} |p_{ij}^s - p_{ij}| > \varepsilon_2\right) \\ & \leq \lim_{\varepsilon_2 \downarrow 0} \mathbf{P}\left(\limsup_{t \rightarrow \infty} \sup_{0 \leq s \leq t} |p_{ij}^s - p_{ij}| > \varepsilon_2\right) = 0, \quad j = 1, 2, \dots, k. \quad (22) \end{aligned}$$

So, the first term in (21) converges to zero. Using the probability limit theorem for a renewal process (see, for example, [20]), we get, for $\varepsilon > 0$,

$$\mathbf{P}\left(\lim_{t \rightarrow \infty} \sup_{0 \leq s \leq t} \frac{(x_j(s) - \mu_j s)}{t} > \varepsilon\right) = 0, \quad j = 1, 2, \dots, k. \quad (23)$$

Applying (21)–(23), we obtain for $\varepsilon > 0$

$$\lim_{t \rightarrow \infty} \mathbf{P}\left(\frac{\sup_{0 \leq s \leq t} |\tilde{x}_j(s) - \hat{x}_j(s)|}{t} > \varepsilon\right) = 0, \quad j = 1, 2, \dots, k. \quad (24)$$

Consequently, (14) is proved. Again, applying the limit theorem to complex renewal processes, we get that

$$\frac{\hat{y}_j(t) - \hat{\beta}_j t}{t} = \frac{\sum_{i=1}^{\hat{x}_j(t)} S_i^{(j)} - \frac{\lambda_j + \sum_{i=1}^k \mu_i p_{ij}}{\mu_j} t}{\hat{\sigma}_j t} \Rightarrow 0, \quad j = 1, 2, \dots, k. \quad (25)$$

Finally, applying Theorem 1, we achieve that

$$\frac{W_j(t)}{t} \Rightarrow \hat{\beta}_j > 0, \quad j = 1, 2, \dots, k. \quad (26)$$

So, the proof of the theorem is complete. \square

4 On a fluid limit of the queue length of customers in an OQN

Now we prove the following theorem for a fluid limit of the queue length of customers in an OQN.

Theorem 4. *If conditions (1)–(3) are fulfilled, then*

$$\left(\frac{Q_1(t)}{t}, \frac{Q_2(t)}{t}, \dots, \frac{Q_k(t)}{t}\right) \Rightarrow (\beta_1, \beta_2, \dots, \beta_k).$$

Proof. Minkevičius, Kulvietis [45] have proved that,

$$|Q_j(t) - \bar{x}_j(t)| \leq w_j(t) + \gamma_j(t), \quad (27)$$

where $\bar{x}_j(t) = \sum_{i=1}^k x_i(t)p_{ij} + a_j(t) - x_j(t)$, $w(t) = \sum_{i=1}^k x_i(t)\{\sum_{l=1}^k |p_{il}^t - p_{il}|\}$, $\gamma(t) = \sum_{i=1}^k \sup_{0 \leq s \leq t} (x_i(s) - \tau_i(s))$, $j = 1, 2, \dots, k$ and $t > 0$.

Using (27) we can find that, for $\varepsilon > 0$,

$$\begin{aligned}
& \mathbf{P}\left(\left|\frac{Q_j(t)}{t} - \beta_j\right| > \varepsilon\right) \\
& \leq \mathbf{P}\left(\left|\frac{Q_j(t)}{t} - \frac{\bar{x}_j(t)}{t}\right| > \frac{\varepsilon}{2}\right) + \mathbf{P}\left(\left|\frac{\bar{x}_j(t)}{t} - \beta_j\right| > \frac{\varepsilon}{2}\right) \\
& \leq \sum_{i=1}^k \sum_{l=1}^k \mathbf{P}\left(\frac{x_i(t)}{t} |p_{il}^t - p_{il}| > \frac{\varepsilon}{2k^2}\right) \\
& \quad + \sum_{i=1}^k \mathbf{P}\left(\frac{\sup_{0 \leq s \leq t} (x_i(s) - \tau_i(s))}{t} > \frac{\varepsilon}{2k}\right) + \mathbf{P}\left(\left|\frac{\hat{x}_j(t)}{t} - \beta_j\right| > \varepsilon\right) \\
& \leq \sum_{i=1}^k \sum_{l=1}^k \mathbf{P}\left(|p_{il}^t - p_{il}| > \frac{\varepsilon}{2k^3}\right) + \sum_{i=1}^k \mathbf{P}\left(\frac{\sup_{0 \leq s \leq t} (x_i(s) - \tau_i(s))}{t} > \frac{\varepsilon}{2k}\right) \\
& \quad \times \sum_{i=1}^k \mathbf{P}\left(\frac{\sup_{0 \leq s \leq t} |x_i(s) - \mu_i s|}{t} > \frac{\varepsilon}{2k^2}\right) + \mathbf{P}\left(\left|\frac{\hat{x}_j(t)}{t} - \beta_j\right| > \frac{\varepsilon}{2}\right), \quad (28)
\end{aligned}$$

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

Applying the results of the lemma in [44] we obtain that the first term in (28) converges to zero. Also, applying the limit theorem for a renewal process, we see that the second and fourth terms in (28) also converge to zero. Thus, applying (4), we achieve that the third term converges to zero. From these notes we obtain that, for $\varepsilon > 0$,

$$\lim_{t \rightarrow \infty} \mathbf{P}\left(\left|\frac{Q_j(t)}{t} - \beta_j\right| > \varepsilon\right) = 0, \quad j = 1, 2, \dots, k.$$

The proof of the theorem is complete. \square

5 The laws of Little for extreme values in an OQN

Here we present applications of the proved theorems (the laws of Little).

Theorem 5. *If conditions (1)–(3) are fulfilled, then*

$$\lim_{t \rightarrow \infty} \frac{Q_j(t)}{W_j(t)} \Rightarrow \mu_j > 0, \quad j = 1, 2, \dots, k.$$

Proof. To this end, we make use of Theorems 4 and 3 on fluid limits for the queue length of customers and a virtual waiting time of a customer in an OQN.

Hewer, we have

$$\lim_{t \rightarrow \infty} \frac{Q_j(t)}{W_j(t)} = \lim_{t \rightarrow \infty} \frac{\frac{Q_j(t)}{t}}{\frac{W_j(t)}{t}} \Rightarrow \mu_j > 0, \quad j = 1, 2, \dots, k.$$

The proof of the theorem is complete. \square

In this section, we use some conditions:

$$\beta_k > \beta_{k-1} > \cdots > \beta_2 > \beta_1 > 0, \quad (29)$$

$$\hat{\beta}_k > \hat{\beta}_{k-1} > \cdots > \hat{\beta}_2 > \hat{\beta}_1 > 0, \quad (30)$$

$$\mu_k > \mu_{k-1} > \cdots > \mu_2 > \mu_1 > 0. \quad (31)$$

These conditions were used in [46]. We prove that from conditions (29) and (31) follow conditions (30) and (31):

$$\hat{\beta}_j = \frac{\beta_j}{\mu_j} > \frac{\beta_{j-1}}{\mu_{j-1}} = \hat{\beta}_{j-1}, \quad j = 1, 2, \dots, k. \quad (32)$$

Theorem 6. *If conditions (29) and (31) are satisfied, then*

$$\frac{\max_{1 \leq j \leq k} \sup_{0 \leq s \leq t} Q_j(s)}{\max_{1 \leq j \leq k} \sup_{0 \leq s \leq t} W_j(s)} \Rightarrow \mu_k > 0.$$

Proof. Applying the results of Minkevičius [46] we obtain that

$$\frac{\max_{1 \leq j \leq k} \sup_{0 \leq s \leq t} Q_j(ns) - \beta_k nt}{\sqrt{n}} \Rightarrow \bar{\sigma}_k z(t), \quad (33)$$

where $z(t)$ is a standard Wiener process, $0 \leq t \leq 1$ and

$$\frac{\max_{1 \leq j \leq k} \sup_{0 \leq s \leq t} W_j(ns) - \hat{\beta}_k nt}{\sqrt{n}} \Rightarrow \hat{\sigma}_k z(t), \quad (34)$$

where $z(t)$ is a standard Wiener process, $0 \leq t \leq 1$.

Thus, applying (33), (34) and the methods of the present paper (see the proof of Theorem 3 and Theorem 4), we derive that

$$\frac{\max_{1 \leq j \leq k} \sup_{0 \leq s \leq t} Q_j(s)}{t} \Rightarrow \beta_k, \quad (35)$$

and

$$\frac{\max_{1 \leq j \leq k} \sup_{0 \leq s \leq t} W_j(s)}{t} \Rightarrow \hat{\beta}_k. \quad (36)$$

Using (35) and (36), we complete the proof of Theorem 5.2. \square

Theorem 7. *If conditions (29) and (31) are satisfied, then*

$$\frac{\max_{1 \leq j \leq k} Q_j(t)}{\max_{1 \leq j \leq k} W_j(t)} \Rightarrow \mu_k > 0.$$

Proof. Applying similar methods as in (33), (34), we see that

$$\frac{\max_{1 \leq j \leq k} Q_j(nt) - \beta_k nt}{\sqrt{n}} \Rightarrow \bar{\sigma}_k z(t), \quad (37)$$

where $z(t)$ is a standard Wiener process, $0 \leq t \leq 1$ and

$$\frac{\max_{1 \leq j \leq k} W_j(nt) - \hat{\beta}_k nt}{\sqrt{n}} \Rightarrow \hat{\sigma}_k z(t), \quad (38)$$

where $z(t)$ is a standard Wiener process, $0 \leq t \leq 1$.

Thus, applying (37), (38) and the methods of the present paper (see the proof of Theorem 3 and Theorem 4), we achieve that

$$\frac{\max_{1 \leq j \leq k} Q_j(t)}{t} \Rightarrow \beta_k, \quad (39)$$

and

$$\frac{\max_{1 \leq j \leq k} W_j(s)}{t} \Rightarrow \hat{\beta}_k. \quad (40)$$

Using (39) and (40), we complete the proof of Theorem 7. \square

Theorem 8. *If conditions (29) and (31) are satisfied, then*

$$\frac{\min_{1 \leq j \leq k} Q_j(t)}{\min_{1 \leq j \leq k} W_j(t)} \Rightarrow \mu_1 > 0.$$

Proof. The proof is similar as that in Theorem 7. The proof of Theorem 8 is complete. \square

References

1. A.A. Borovkov, *Stochastic Processes in Queueing Theory*, Springer, Berlin, 1976.
2. A.A. Borovkov, *Asymptotic Methods in Queueing Theory*, Wiley, New York, 1984.
3. A.A. Borovkov, Limit theorems for queueing networks, *Theory Probab. Appl.*, **31**, pp. 413–427, 1986.
4. F.I. Karpelevitch, A.Ya. Kreinin, Joint distributions in Poissonian tandem queues, *Queueing Sys.*, **12**, pp. 274–286, 1992.
5. G.P. Klimov, Several solved and unsolved problems of the service by queues in series, *Izv. Akad. Nauk SSSR, Tekh. Kibern.*, **6**, pp. 88–92, 1970 (in Russian).
6. Ya.A. Kogan, A.A. Pukhalsky, Tandem queues with finite intermediate waiting room and blocking in heavy traffic, *Probl. Control Inf. Theory*, **17**, pp. 3–13, 1988.

7. W. Whitt, Heavy traffic limit theorems for queues: A survey, in: *Mathematical Methods in Queueing Theory, Proceedings of a Conference at Western Michigan University, May 10–12, 1973*, Lect. Notes Econ. Math. Syst., Vol. 98, Springer-Verlag, Berlin, Heidelberg, New York, 1974, pp. 307–350.
8. F.P. Kelly, An asymptotic analysis of blocking, *Modelling and Performance Evaluation Methodology, Proceedings of the International Seminar Paris, France, January 24–26, 1983*, Lect. Notes Control Comput. Sci., Vol. 60, Springer, Berlin, 1984, pp. 3–20.
9. P.W. Glynn, W. Whitt, A new view of the heavy-traffic limit theorems for infinite-server queues, *Adv. Appl. Probab.*, **23**, pp. 188–209, 1991.
10. J.M. Harrison, The heavy traffic approximation for single server queues in series, *Adv. Appl. Probab.*, **10**, pp. 613–629, 1973.
11. J.M. Harrison, The diffusion approximation for tandem queues in heavy traffic, *Adv. Appl. Probab.*, **10**, pp. 886–905, 1978.
12. J.M. Harrison, M.I. Reiman, On the distribution of multidimensional reflected Brownian motion, *SIAM J. Appl. Math.*, **41**, pp. 345–361, 1981.
13. J.M. Harrison, R.J. Williams, Brownian models of open queueing networks with homogeneous customer populations, *Stochastics*, **22**, pp. 77–115, 1987.
14. D.L. Iglehart, W. Whitt, Multiple channel queues in heavy traffic. I, *Adv. Appl. Probab.*, **2**, pp. 150–175, 1970.
15. D.L. Iglehart, W. Whitt, Multiple channel queues in heavy traffic. II. Sequences, networks and batches, *Adv. Appl. Probab.*, **2**, pp. 355–364, 1970.
16. M.I. Reiman, A multiclass feedback queue in heavy traffic, *Adv. Appl. Probab.*, **20**, pp. 179–207, 1988.
17. M.I. Reiman, B. Simon, A network of priority queues in heavy traffic: One bottleneck station, *Queueing Syst.*, **6**, pp. 33–58, 1990.
18. W. Szczotka, F.P. Kelly, Asymptotic stationarity of queues in series and the heavy traffic approximation, *Ann. Probab.*, **18**, pp. 1232–1248, 1990.
19. W. Whitt, Weak convergence theorems for priority queues: Preemptive resume discipline, *J. Appl. Probab.*, **8**, pp. 79–94, 1971.
20. P. Billingsley, *Convergence of Probability Measures*, Wiley, New York, 1968.
21. A.A. Borovkov, Weak convergence of functionals of random sequences and processes defined on the hole axis, *Proc. Steklov Inst. Math.*, **128**, pp. 41–65, 1972.
22. Y.V. Prokhorov, Convergence of random processes and limit theorems in probability theory, *Theory Probab. Appl.*, **1**, pp. 157–214, 1956.
23. A.V. Skorokhod, *Studies in the Theory of Random Processes*, Addison-Wesley, New York, 1965.

24. B. Grigelionis, R. Mikulevičius, Diffusion approximation in queueing theory, in: *Fundamentals of Teletraffic Theory, Proceedings of the Third International Seminar, Moscow, USSR, June 20–26, 1984*, pp. 147–158, 1984.
25. E.V. Krichagina, R.Sh. Liptzer, A.A. Pukhalsky, The diffusion approximation for queues with input flow, depending on a queue state and general service, *Theory Prob. Appl.*, **33**, pp. 124–135, 1988.
26. R. Sh. Liptzer, A.N. Shiryaev, *Theory of Martingales*, Kluwer, Boston, 1989.
27. M.I. Reiman, Open queueing networks in heavy traffic, *Math. Oper. Res.*, **9**, p 441-458, 1984.
28. H. Chen, H. Zhang, A sufficient condition and a necessary condition for the diffusion approximations of multiclass queueing networks under priority service disciplines, *Queueing Syst.*, **34**, pp. 237–268, 2000.
29. H. Chen, D.D. Yao, *Fundamentals of Queueing Networks: Performance, Asymptotics and Optimization*, Springer-Verlag, New York, 2001.
30. H. Chen, Q. Ye, Existence condition for the diffusion approximations of multiclass priority queueing networks, *Queueing Sys.*, **38**, pp. 435–470, 2002.
31. M. Bramson, J.G. Dai, Heavy traffic limits for some queueing networks, *Ann. Appl. Probab.*, **11**, pp. 49–90, 2001.
32. J.M. Harrison, A broader view of brownian networks, *Ann. Appl. Probab.*, **3**, pp. 1119–1150, 2003.
33. A. Mandelbaum, A.L. Stolyar, Scheduling flexible servers with convex delay costs: Heavy-traffic optimality of the generalized $c\mu$ -rule, *Oper. Res.*, **52**, pp. 836–855, 2004.
34. W.N. Kang, F.P. Kelly, N.H. Lee, R.J. Williams, State space collapse and diffusion approximation for a network operating under a fair bandwidth sharing policy, *Ann. Appl. Probab.*, **19**, pp. 1719–1780, 2009.
35. H.Q. Ye, D.D. Yao, Heavy traffic optimality of stochastic networks under utility-maximizing resource control, *Oper. Res.*, **56**, pp. 453–470, 2008.
36. A.J. Lemoine, Network of queues – A survey of weak convergence results, *Manage. Sci.*, **24**, pp. 1175–1193, 1978.
37. C. Flores, Diffusion approximations for computer communications networks, in: B. Gopinath (Ed.), *Computer Communications Vol. 31, Proceedings of Symposia in Applied Mathematics*, American Mathematical Society, 1985, pp. 83–124.
38. P.W. Glynn, Diffusion approximations, in: D.P. Heyman, M.J. Sobel (Eds.), *Handbooks in Operations Research and Management Science, Vol. 2, Stochastic Models*, Elsevier Science Publishers B.V., North-Holland, 1990.
39. D.P. Johnson, *Diffusion Approximations for Optimal Filtering of Jump Processes and for Queueing Networks*, Ph.D. Dissertation, University of Wisconsin, 1983.
40. H. Chen, A. Mandelbaum, Stochastic discrete flow networks: Diffusion approximations and bottlenecks, *Ann. Probab.*, **19**, pp. 1463–1519, 1991.

41. J.M. Harrison, A.J. Lemoine, A note on networks of infinite-server queues, *J. Appl. Probab.*, **18**, pp. 561–567, 1981.
42. W. Whitt, On the heavy-traffic limit theorem for $GI/G/\infty$ queues, *Adv. Appl. Probab.*, **14**, pp. 171–190, 1982.
43. D.L. Iglehart, Multiple channel queues in heavy traffic. IV. Law of the iterated logarithm, *Z. Wahrscheinlichkeitstheor. Verw. Geb.*, **17**, pp. 168–180, 1971.
44. L. Sakalauskas, S. Minkevičius, On the law of the iterated logarithm in open queueing networks, *Eur. J. Oper. Res.*, **120**, pp. 632–640, 2000.
45. S. Minkevičius, G. Kulvietis, Application of the law of the iterated logarithm in open queueing networks, *WSEAS Trans. Syst.*, **6**, pp. 643–651, 2007.
46. S. Minkevičius, On the global values of the queue length in open queueing networks, *Int. J. Comput. Math.*, pp. 1029–1065, 2009.
47. V. Strassen, An invariance principle for the law of the iterated logarithm, *Z. Wahrscheinlichkeitstheor. Verw. Geb.*, **3**, pp. 211–226, 1964.