ISSN 1392 – 124X INFORMATION TECHNOLOGY AND CONTROL, 2012, Vol. 41, No. 2

Modeling of the Phenomena in Multiserver Networks

Saulius Minkevičius[*], **Leonidas Sakalauskas**

Institute of Mathematics and Informatics of VU,
Akademijos 4, 08663 Vilnius, Lithuania
e-mail: minkevicius.saulius@gmail.com

crossref http://dx.doi.org/10.5755/j01.itc.41.2.917

Abstract. The paper is devoted to the analysis of queueing systems in the context of the network and communications theory. We investigate a theorem on the law of the iterated logarithm for a queue of jobs in an open multiserver queueing network and its applications to the mathematical models of the generalized Internet system and a multiserver computer system.

Keywords: Models of information systems, performance evaluation, open multiserver queueing network, queue length of jobs, the law of the iterated logarithm, heavy traffic.

1. Introduction

The paper is devoted to the analysis of queueing systems in the context of the network and communications theory. We investigate a theorem on the law of the iterated logarithm (LIL) for the queue of jobs in an open multiserver queueing network and its applications to the mathematical models of the generalized Internet system and a multiserver computer system. Queueing networks have been extensively used for the analysis of manufacturing systems, transportation systems, and computer and communication networks. Therefore, many approximation methods have emerged, and LIL are among them.

Limit theorems (diffusion approximations) and LIL under the conditions of heavy traffic belong to a special area of investigations on the queueing theory in heavy traffic. Therefore, first we shall try to trace the development of research on the general queueing theory in heavy traffic. The history of investigation of the diffusion approximation of a queueing system in heavy traffic has lasted for 40 years, and that of diffusion approximation of queueing networks – about 20 years. Though as far back as in the 50-ies Kolmogorov raised a hypothesis on an approximate description of the number of busy channels with failures by means of the diffusion process with reflection from the true boundary, a systematic investigation of the problem began only after publishing J. Kingman's works [21, 22] and especially that of Prokhorov [31]. The methods of investigation of single-server queueing systems in heavy traffic were considered in Borovkov [3], Kendall [20], Iglehart and Whitt [15, 16], etc. Later on, there appeared a large number of works devoted to various aspects of diffusion approximation of the queueing models (see survey papers by Iglehart [14] as well as Whitt [36], the book and the article of Karpelevich and Kreinin [18, 19]). The works of Reiman [33] and Kobyashi et al. [23] laid the basis for investigations of diffusion approximation of queueing networks. There is a vast literature on the diffusion approximation. Readers are referred to Whitt [36], Lemoine [24], Flores [6] and Glynn [7] for a general survey of the subject. The present work extends the studies by Iglehart and Whitt [15, 16] for a single station of multiserver queues, and by Reiman [33], Johnson [17], Chen and Mandelbaum [4] for networks of single server queues. Other closely related papers are written by Harrison and Lemoine [10] for networks of infinite server queues, and Whitt [37] for a $GI/G/\infty$ queue.

One can familiarize with the general theory on the LIL and its numerous applications in various fields of probability theory in the survey by Bingham [2]. The main part in the development of the theory on the LIL was played by Strassen [35] in the fundamental work where the functional variant of the LIL for a Wiener process was proved. The paper of Iglehart [11] can be considered as the first work on the LIL in the queueing theory. Applying the approach of Strassen [35], Iglehart [11] proved the LIL in it, under the conditions of heavy traffic, for the queue length of customers, waiting time of a customer, a virtual waiting time of a customer, and other important probability characteristics of the classical queueing system $G1/G/1$ and more general systems (e.g.,

[*] Corresponding author

for a multiple queueing system). Also, a functional variant of the LIL for a renewal process was proved by Iglehart [11]. Using the results of Iglehart [12, 13], the survey by Whitt [36] presented the proof of theorems on the LIL for the waiting time of a customer, the occupation time process, and the extreme value of the waiting time of a customer in the queueing system $G1/G/1$. Also, Glynn and Whitt [8, 9] proved theorems on the LIL for a cumulative process associated with the queue length of customers and waiting time of a customer in an ordinary queueing system $G1/G/1$.

Note that the research of the LIL in more general systems than the queueing system $G1/G/1$ or multiphase queues has just started (see Asmussen [1]). In Minkevičius [25, 26], the LIL was proved in heavy traffic for the queue length of customers, waiting time of a customer, a virtual waiting time of a customer in a multiphase queueing system. Sakalauskas and Minkevičius [34, 28] also give the proof of the theorem on the LIL under "overload conditions" of heavy traffic for a virtual waiting time of a customer and queue length of customers in the open Jackson network.

The classical papers of Iglehart and Whitt [15, 16] began the research of multiple (multichannel, multiserver) queueing systems in heavy traffic. Modern topics of these researches include performance of modeling techniques for server farms, scheduling and prioritization algorithms for multiserver systems, load balancing and load sharing, queueing analysis and sensitivity analysis of multiserver systems, workload characterization for multiserver systems, computational methods for multi-dimensional optimization, impact of heavy-tailed workloads on the multiserver system performance, etc. Modern applications of multiserver systems include web server farms, servers for high-performance computing and grid computing, manufacturing applications, etc. Modern methods of investigation of multiserver systems in heavy traffic include moment results, tail asymptotic results for some specialized systems, methods for reducing 2- dimensionally infinite Markov chains to 1-dimensionally infinite Markov chains, time and unit-scaling techniques, new results in scheduling theory and task assignment, etc. The books of Whitt [38] as well as Chen and Yao [5] present the results of the theory of multiserver queueing systems in heavy traffic, its present state, and the problems to be solved. Note that the research of the multiple queueing systems in more general systems than the queueing system $G1/G/1/n$ or multiphase/multiple queueing systems has just started (see papers of Whitt [39, 40], Puhalskii and Reed [32], Pang, Talreja and Whitt [30]). The present paper, which investigates the queue length of customers in the open multiserver network, extends the results of Sakalauskas, Minkevicius [28] in the open Jackson network under "overload conditions" of heavy traffic.

In this paper, we investigate an open multiserver queueing network model in heavy traffic. We present the LIL for a queue of jobs in an open multiserver queueing network. The main tool for the analysis of these queueing systems in heavy traffic is a functional LIL for the renewal process (the proof can be found in Strassen [35] and Iglehart [11]).

2. The network model

Consider a network of j stations, indexed by $j = 1, 2, \ldots, J$. Suppose that the station j has c_j servers, indexed by $(j, 1), \ldots, (j, c_j)$. A description of the primitive data and construction of processes of interest are the focus of this section. No probability space will be mentioned in this section, and of course, one can always think that all the variables and processes are defined on the same probability space.

First, let $\{u_j(e), e \geq 1\}, j = 1, 2, \ldots, J$, be J sequences of exogenous interarrival times, where $u_j(e) \geq 0$ is the interarrival time between the $(e-1)$-st job and the e-st job that arrive at the station j exogenously (from the outside of the network). Define $U_j(0) = 0$, $\quad U_j(n) = \sum_{e=1}^{n} u_j(e), \ n > 1$ and $A_j(t) = \sup\{n > 0 : U_j(n) \leq t\}$, where $A_j = \{A_j(t), \ t \geq 0\}$ is called an exogenous arrival process of the station j, i.e., $A_j(t)$ counts the number of jobs that arrived at the station j from the outside of the network at the moment t.

Second, let $\{v_{jk_j}(e), e \geq 1\}, \ j = 1, 2, \ldots, J,$ $k_j = 1, 2, \ldots, c_j$, be $c_1 + \ldots + c_J$ sequences of service times, where $v_{jk_j}(e) \geq 0$ is the service time for the e-th job served by the server k_j of the station j. Define $V_{jk_j}(0) = 0$, $\quad V_{jk_j}(n) = \sum_{e=1}^{n} v_{jk_j}(e), \ n \geq 1$ 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 a service process for the server k_j at the station j, i.e., $x_{jk_j}(t)$ counts the number of services completed by the server k_j at the station j during the server's busy time. We define $\mu_{jk_j} = \left(E\left[v_{jk_j}(e)\right]\right)^{-1} > 0$, $\sigma_{jk_j} = D\left(v_{jk_j}(e)\right) > 0$ and $\lambda_j = \left(E\left[u_j(e)\right]\right)^{-1} > 0$, $a_j = D\left(u_j(e)\right) > 0$, $j = 1, 2, ..., k$; with all of these terms assumed finite.

Also, let $\tilde{\tau}_j(t)$ be the total number of jobs routed to the jth station of the network in the interval $[0, t]$, $\tau_j(t)$ be the total number of jobs after service departure from the jth 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 server of the jth station of the network in the interval $[0,t]$. Let $\tau_{jk_j}(t)$ be the total number of customers after service departure from the k_j server of the jth station of the network in the interval $[0,t]$, and $\tau_{ijk_i}(t)$ be the total number of jobs after service departure from the k_i server of the ith station of the network and routed to the k_j server of the jth station of the network in the interval $[0,t]$. Let p_{ij} be a probability of the job after service at the ith station of the network routed to the jth station of the network. Denote $p^t_{ijk_i} = \frac{\tau_{ijk_i}(t)}{\tau_{ik_i}(t)}$ to be a part of the total number of jobs which, after service at the k_i server of the ith station of the network, are routed to the jth station of the network in the interval $[0,t]$, $i,j = 1,2,\ldots,J$, $k_i = 1,\ldots,c_i$ and $t > 0$.

The process of primary interest is 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 the station j at time t. Now we introduce the processes $Q_{jk_j} = \{Q_{jk_j}(t),\ t \geq 0\}$, where $Q_{jk_j}(t)$ indicates the number of jobs waiting to be served by the server k_j of the station j at time t; clearly, we have $Q_j(t) = \sum_{k_i=1}^{c_j} Q_{jk_i}(t),\ j = 1,2,\ldots,J$.

The dynamics of the queueing system (to be specified) depends on the service discipline at each service station. To be more precise, "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 one of the servers with the smallest index. We assume that the service station is work-conserving; namely, not all servers at a station can be idle when there are customers waiting for service at that station. In particular, we assume that a station must serve at its full capacity when the number of waiting jobs is equal to or exceeds the number of servers at that station. We assume that the queue of jobs in each station of the open queueing network is unlimited.

3. The main results

Let $\beta_j = \sum_{i=1}^{J}\sum_{k_i=1}^{c_i} \mu_{ik_i} \cdot p_{ij} + \lambda_j - \sum_{k_j=1}^{c_j} \mu_{jk_j}$, $\hat{\sigma}_j^2 = \sum_{i=1}^{J}\sum_{k_i=1}^{c_i} \mu^3_{ik_i} \cdot \sigma_{ik_i} \cdot p^2_{ij} + \lambda^3_j \cdot a_j + \sum_{k_j=1}^{c_j} \mu^3_{jk_j} \cdot \sigma_{jk_j} > 0$, $j = 1,2,\ldots,J$.

We assume that the following conditions are fulfilled:

$$\sum_{i=1}^{J}\sum_{k_i=1}^{c_i} \mu_{ik_i} \cdot p_{ij} + \lambda_j > \sum_{k_j=1}^{c_j} \mu_{jk_j},\ j = 1,2,\ldots,J. \tag{1}$$

Note that these conditions quarantee that there exists a queue of jobs and it is constantly growing. One of the results of the paper is the following theorem on the LIL for the total queue of jobs in an open queueing network.

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

$$P\left(\overline{\lim_{t\to\infty}} \frac{Q_j(t) - \beta_j \cdot t}{\hat{\sigma}_j \cdot a(t)} = 1\right) =$$

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

$j = 1,2,\ldots,J$ *and* $a(t) = \sqrt{2t \ln\ln t}.$

Proof. First, define $\hat{x}_j(t) = \sum_{i=1}^{J}\sum_{k_i=1}^{c_i} x_{ik_i}(t) \cdot p_{ij} + A_j(t) - \sum_{k_j=1}^{c_j} x_{jk_j}(t), w(t) = \sum_{j=1}^{J}\sum_{i=1}^{J}\sum_{k_i=1}^{c_i} x_{ik_i}(t) \cdot |p^t_{ijk_i} - p_{ij}|, \gamma(t) = \sum_{i=1}^{J}\sum_{k_i=1}^{c_i} \sup_{0\leq s\leq t} (x_{ik_i}(s) - \tau_{ik_i}(s)),\ j = 1,2,\ldots,J.$

By the definition of the queue of jobs at the stations of the network, we get that, for $j = 1,2,\ldots,J$, $k_j = 1,2,\ldots,c_j$ and $t > 0$

$$\begin{aligned}
Q_j(t) &= \tilde{\tau}_j(t) - \tau_j(t) = \sum_{k_i=1}^{c_j} Q_{ik_i}(t) = \sum_{k_i=1}^{c_j} \tilde{\tau}_{ik_i}(t) \\
&- \sum_{k_i=1}^{c_j} \tau_{ik_i}(t) = \sum_{k_i=1}^{c_j} \tilde{\tau}_{ik_i}(t) - \sum_{k_i=1}^{c_j} x_{ik_i}(t) \\
&+ \sum_{k_i=1}^{c_j} x_{ik_i}(t) - \sum_{k_i=1}^{c_j} \tau_{ik_i}(t) \leq \sum_{k_i=1}^{c_j} \tilde{\tau}_{ik_i}(t) \\
&- \sum_{k_i=1}^{c_j} x_{ik_i}(t) + \sum_{k_i=1}^{c_j} \sup_{0\leq s\leq t} (x_{ik_i}(s) - \tau_{ik_i}(s)) \\
&= \sum_{i=1}^{J}\sum_{k_i=1}^{c_i} \tau_{ijk_i}(t) + A_j(t) - \sum_{k_i=1}^{c_j} x_{ik_i}(t) \\
&+ \sum_{k_i=1}^{c_j} \sup_{0\leq s\leq t} (x_{ik_i}(s) - \tau_{ik_i}(s)) \leq A_j(t) \\
&- \sum_{k_j=1}^{c_j} x_{jk_j}(t) + \sum_{i=1}^{J}\sum_{k_i=1}^{c_i} \tau_{ik_i}(t) \cdot \frac{\tau_{ijk_i}(t)}{\tau_{ik_i}(t)} \\
&+ \sum_{k_i=1}^{c_j} \sup_{0\leq s\leq t} (x_{ik_i}(s) - \tau_{ik_i}(s))
\end{aligned}$$

$$\leq \sum_{i=1}^{J}\sum_{k_i=1}^{c_i} x_{ik_i}(t)\cdot p_{ijk_i}^t + A_j(t) - \sum_{k_j=1}^{c_j} x_{jk_j}(t)$$

$$+ \sup_{0\leq s\leq t}(x_{jk_j}(s) - \tau_{jk_j}(s)) = A_j(t) - \sum_{k_i=1}^{c_j} x_{ik_i}(t)$$

$$+ \sum_{i=1}^{J}\sum_{k_i=1}^{c_i} x_{ik_i}(t)\cdot (p_{ijk_i}^t - p_{ij} + p_{ij}) \leq$$

$$\leq \sum_{i=1}^{J}\sum_{k_i=1}^{c_i} x_{ik_i}(t)\cdot p_{ij} + A_j(t) - \sum_{k_i=1}^{c_j} x_{ik_i}(t) +$$

$$+ \sum_{i=1}^{J}\sum_{k_i=1}^{c_i} x_{ik_i}(t)\cdot |p_{ijk_i}^t - p_{ij}|$$

$$+ \sum_{k_i=1}^{c_j} \sup_{0\leq s\leq t}(x_{ik_i}(s) - \tau_{ik_i}(s))$$

$$= \hat{x}_j(t) + w(t) + \gamma(t).$$

Hence it follows that

$$Q_j(t) \leq \hat{x}_j(t) + w(t) + \gamma(t),\ j = 1,2,\ldots,J \quad (2)$$

and $t > 0$. Also, note that

$$Q_j(t) \geq \tilde{\tau}_j(t) - \sum_{k_i=1}^{c_j} x_{ik_i}(t) = A_j(t) - \sum_{k_i=1}^{c_j} x_{ik_i}(t)$$

$$\sum_{i=1}^{J}\sum_{k_i=1}^{c_i} \tau_{ik_i}(t)\cdot p_{ijk_i}^t + = A_j(t) - \sum_{k_i=1}^{c_j} x_{ik_i}(t)$$

$$+ \sum_{i=1}^{J}\sum_{k_i=1}^{c_i} (x_{ik_i}(t) + \tau_{ik_i}(t) - x_{ik_i}(t))\cdot p_{ijk_i}^t$$

$$= \sum_{i=1}^{J}\sum_{k_i=1}^{c_i} x_{ik_i}(t)\cdot p_{ijk_i}^t + A_j(t) - \sum_{k_i=1}^{c_j} x_{ik_i}(t)$$

$$+ \sum_{i=1}^{J}\sum_{k_i=1}^{c_i} (\tau_{ik_i}(t) - x_{ik_i}(t))\cdot p_{ijk_i}^t = A_j(t) +$$

$$+ \sum_{i=1}^{J}\sum_{k_i=1}^{c_i} x_{ik_i}(t)\cdot p_{ijk_i}^t - \sum_{k_j=1}^{c_j} x_{jk_j}(t)$$

$$- \sum_{i=1}^{J}\sum_{k_i=1}^{c_i} (x_{ik_i}(t) - \tau_{ik_i}(t))\cdot p_{ijk_i}^t$$

$$\geq \sum_{i=1}^{J}\sum_{k_i=1}^{c_i} x_{ik_i}(t)\cdot p_{ijk_i}^t + A_j(t) - \sum_{k_j=1}^{c_j} x_{jk_j}(t)$$

$$- \sum_{i=1}^{J}\sum_{k_i=1}^{c_i} (x_{ik_i}(t) - \tau_{ik_i}(t)) \geq +A_j(t)$$

$$- \sum_{k_j=1}^{c_j} x_{jk_j}(t) + \sum_{i=1}^{J}\sum_{k_i=1}^{c_i} x_{ik_i}(t)\cdot p_{ijk_i}^t$$

$$- \sup_{0\leq s\leq t} \sum_{i=1}^{J}\sum_{k_i=1}^{c_i} (x_{ik_i}(s) - \tau_{ik_i}(s)) \geq$$

$$\sum_{i=1}^{J}\sum_{k_i=1}^{c_i} x_{ik_i}(t)\cdot p_{ijk_i}^t + A_j(t) - \sum_{k_j=1}^{c_j} x_{jk_j}(t) -$$

$$\sum_{i=1}^{J}\sum_{k_i=1}^{c_i} \sup_{0\leq s\leq t}(x_{ik_i}(s) - \tau_{ik_i}(s)) = - \sum_{k_j=1}^{c_j} x_{jk_j}(t)$$

$$- \sum_{i=1}^{J}\sum_{k_i=1}^{c_i} x_{ik_i}(t)\cdot (p_{ijk_i}^t - p_{ij} + p_{ij}) + A_j(t)$$

$$- \sum_{i=1}^{J}\sum_{k_i=1}^{c_i} \sup_{0\leq s\leq t}(x_{ik_i}(s) - \tau_{ik_i}(s))$$

$$= \sum_{i=1}^{J}\sum_{k_i=1}^{c_i} x_{ik_i}(t)\cdot p_{ij} + A_j(t) - \sum_{k_i=1}^{c_j} x_{ik_i}(t)$$

$$+ \sum_{i=1}^{J}\sum_{k_i=1}^{c_i} x_{ik_i}(t)\cdot (p_{ijk_i}^t - p_{ij})$$

$$- \sum_{i=1}^{J}\sum_{k_i=1}^{c_i} \sup_{0\leq s\leq t}(x_{ik_i}(s) - \tau_{ik_i}(s)) \geq \hat{x}_j(t)$$

$$- \sum_{i=1}^{J}\sum_{k_i=1}^{c_i} x_{ik_i}(t)\cdot |p_{ijk_i}^t - p_{ij}|$$

$$- \sum_{i=1}^{J}\sum_{k_i=1}^{c_i} \sup_{0\leq s\leq t}(x_{ik_i}(s) - \tau_{ik_i}(s)) = \hat{x}_j(t)$$

$$- w(t) - \gamma(t),\ j = 1,2,\ldots,J \text{ and } t > 0. \quad (3)$$

Hence it follows that

$$Q_j(t) \geq \hat{x}_j(t) - \sum_{i=1}^{k} w_i(t) - \sum_{i=1}^{k} \gamma_i(t), \quad (4)$$

$j = 1,2,\ldots,J$ and $t > 0$.

By combining (2) and (4), we can write

$$|Q_j(t) - \hat{x}_j(t)| \leq \sum_{i=1}^{k} w_i(t) + \sum_{i=1}^{k} \gamma_i(t), \quad (5)$$

$j = 1,2,\ldots,J$ and $t > 0$. The further proof is the same as in Minkevičius and Kulvietis [27]. The proof of the theorem is complete. □

Note that inequality (5) is the key inequality used to prove several laws (fluid approximations, functional limit theorems and LIL) for a queue of jobs in open multiserver queueing networks under heavy traffic conditions.

4. On the model of the generalized Internet network

In this section, we consider a generalized Internet network with three resources and six routes (see Figure 1). We now assume that the network is composed of three nodes, each modelled as a multiserver queue. In other words, external arrivals of jobs at the first node are $\lambda_{11}, \lambda_{12}$ and λ_{13}. The packages are served at the first node in the servers of node with the rate $\mu_{11}, \mu_{12}, \mu_{13}$ and μ_{14}, afterwards Internet packages of data (jobs, queries or messages) with probability $p_{12} = p$ (usually $p = 0.1$) are sent to the second node and with probability $p_{10} = 1-p$ leave λ_{22}. The packages are served in the second node with the intensity values μ_{21} and μ_{22}. Then Internet packages, with probability $p_{23} = q$ (usually $q = 0.1$), arrive at the third node in one direction, and with probability $1-q$ they leave the network in another direction. External routes to the third node are characterized by intensity values λ_3. The packages are served in the third node with the intensity values μ_{31} and μ_{32}. Then Internet packages, with probability $p_{30} = r$ (usually $r = 0.9$), leave the network in one direction and they leave the network in another direction (with probability $1-r$).

Next, denote by $\bar{Q}_j(t),\ j = 1,2,3$ the total queue of Internet packages in the jth node of the Internet-type network at the time moment $t,\ t > 0$. Define

$$\begin{aligned}&\bar{\beta}_1 = \lambda_{11} + \lambda_{12} + \lambda_{13} - \mu_{11} - \mu_{12} - \mu_{13} \\ &- \mu_{14} > 0, \quad \bar{\beta}_2 = \lambda_{21} + \lambda_{22} - \mu_{21} - \mu_{22} + \\ &+ (\mu_{11} + \mu_{12} + \mu_{13} + \mu_{14}) \cdot p_{12} > 0,\ \bar{\beta}_3 = \\ &= \lambda_3 + (\mu_{21} + \mu_{22}) \cdot p_{23} - \mu_{31} - \mu_{32} > 0, \\ &\bar{\sigma}_1^2 = a_1 \cdot (\lambda_{11}^3 + \lambda_{12}^3 + \lambda_{13}^3) + \sigma_1 \\ &\cdot (\mu_{11}^3 + \mu_{12}^3 + \mu_{13}^3 + \mu_{14}^3) > 0,\ \bar{\sigma}_2^2 = a_1 \cdot \\ &\cdot (\lambda_{21}^3 + \lambda_{22}^3) + \sigma_1 \cdot (\mu_{11}^3 + \mu_{12}^3 + \mu_{13}^3 + \mu_{14}^3) \cdot \\ &\cdot p_{12}^2 + \sigma_2 \cdot (\mu_{21}^3 + \mu_{22}^3) > 0, \quad \bar{\sigma}_3^2 = a_1 \cdot \lambda_3^3 + \\ &\sigma_1 \cdot (\mu_{21}^3 + \mu_{22}^3) \cdot p_{23}^2 + \sigma_3 \cdot (\mu_{31}^3 + \mu_{32}^3) > 0,\end{aligned} \tag{6}$$

Applying Theorem 3.1, we obtain the following theorem on the total queue of packages in the generalized Internet network system.

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

$$\mathbf{P}\left(\overline{\lim_{n\to\infty}} \frac{\bar{Q}_j(t) - \bar{\beta}_j \cdot t}{\tilde{\sigma}_j \cdot a(t)} = 1\right) = 1, \quad j = 1,2,3.$$

Similarly as in Minkevičius and Kulvietis [27], we can obtain

$$\frac{M\bar{Q}_j(t)}{t} = \bar{\beta}_j + \bar{\sigma}_j \cdot \sqrt{\frac{2\ln\ln t}{t}}, \quad j = 1,2,3. \tag{7}$$

Note that $M\bar{Q}_j(t),\ j = 1,2,3$ is the total average queue of packages in the generalized Internet network at the time moment $t,\ t > 0$.

Now we give an example from the generalized network. Assume that packages are routed to the first node W_1 at the rate λ_{11} of 13700, λ_{12} of 13800 and λ_{13} of 13600 per second. These packages are served at the rate μ_{11} of 10000, μ_{12} of 10000, μ_{13} of 10000 and μ_{14} of 10000 per second. After service in the node W_1 the packages are routed to the second node W_2 at the rates of $(\mu_{11}+\mu_{12}+\mu_{13}+\mu_{14}) \cdot p_{12}$ of 4000, λ_{21} of 8500 and λ_{22} of 8500 per second. The packages are served at the rates μ_{21} of 10000 and μ_{22} of 10000 per second in the node W_2, then the packages are routed to the third node at the rates $(\mu_{21}+\mu_{22}) \cdot p_{23}$ of 2000 and λ_3 of 19000 per second. The packages are served at the rates μ_{31} of 10000 and μ_{32} of 10000 per second in the node W_3. After the packages have been served in the node W_3, they leave the system. So, $\bar{\beta}_1 = 1100,\ \bar{\sigma}_1 = 31033,\ \bar{\beta}_2 = 1000,\ \bar{\sigma}_2 = 16620,\ \bar{\beta}_3 = 1000$ and $\bar{\sigma}_3 = 23727$.

Thus, from (6) we get (see Figures 2-4)

$$\begin{aligned}\frac{M\bar{Q}_1(t)}{t} &= 1150 + 31033 \cdot \sqrt{\frac{2\ln\ln t}{t}}, \\ \frac{M\bar{Q}_2(t)}{t} &= 1000 + 16620 \cdot \sqrt{\frac{2\ln\ln t}{t}}, \\ \frac{M\bar{Q}_3(t)}{t} &= 1000 + 23727 \cdot \sqrt{\frac{2\ln\ln t}{t}}.\end{aligned}$$

5. On the model of the multiserver computer system

In this section, we consider a multiserver computer system (see Figure 5). We assume that this system is composed of two multiserver nodes. In other words, we now have an open multiserver Jackson network, where there is external arrival λ_{11} at the first node at the rates μ_{11} and μ_{12}. Afterward the packages of data are sent with probability $p_{12} = p$ (usually $p = 0.1$) to the second node and leave the network with probability $p_{10} = 1-p$. The packages are served in the second node at the rates μ_{21} and μ_{22}. We denote by $\tilde{Q}(t),\ j = 1,2$, the total of packages in nodes at the time moment $t,\ t > 0$.

Let us denote $\tilde{\beta}_1 = \lambda_{11} - \mu_{11} - \mu_{12} > 0,\ \tilde{\sigma}_1^2 = a_1 \cdot \lambda_{11}^3 + \sigma_1 \cdot (\mu_{11}^3 + \mu_{12}^3) > 0$,
$\tilde{\beta}_2 = (\mu_{11} + \mu_{12}) \cdot p_{12} - \mu_{21} - \mu_{22} > 0,\ \tilde{\sigma}_2^2 = \sigma_1 \cdot (\mu_{11}^3 + \mu_{12}^3) \cdot p_{12}^2 + \sigma_2 \cdot (\mu_{11}^3 + \mu_{22}^3) > 0$,

Similarly as in (7) we can obtain

$$\frac{M\tilde{Q}_j(t)}{t} = \tilde{\beta}_j + \tilde{\sigma}_j \cdot \sqrt{\frac{2\ln\ln t}{t}}, \quad j = 1,2.$$

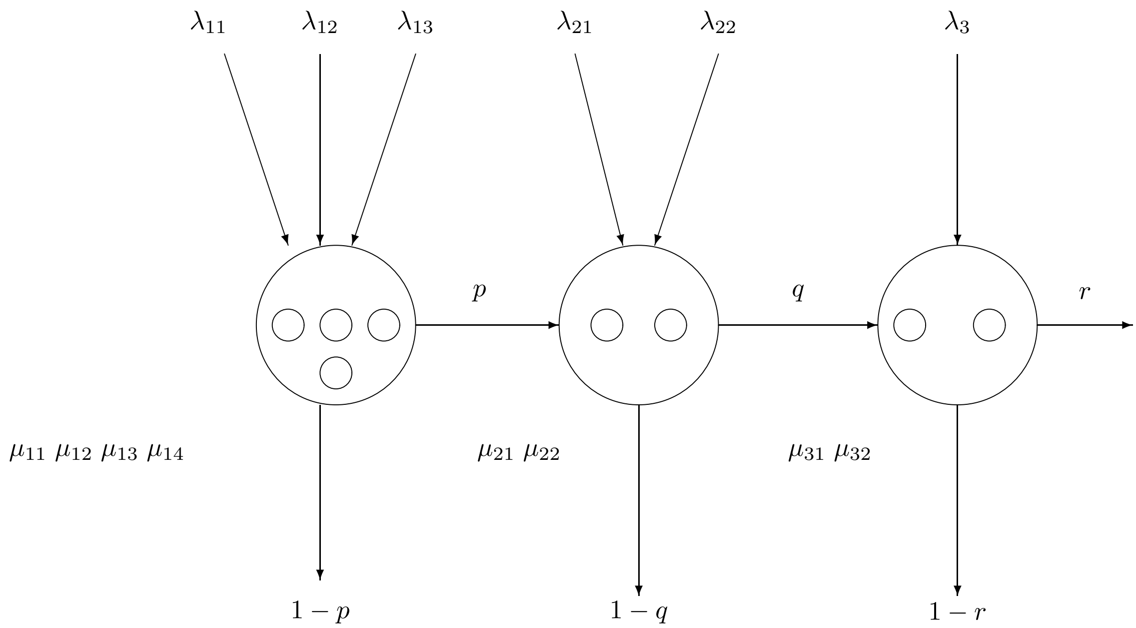

Figure 1. Model of the generalized Internet network

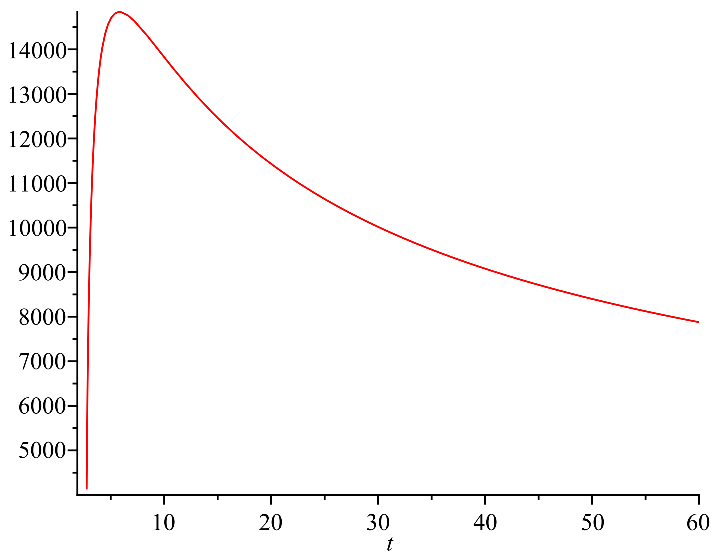

Figure 2. Values for $\frac{M\tilde{Q}_1(t)}{t}$

Now we present an example from the network practice. Assume that packages of data are routed to the first device V_1 at the rate λ_{11} of 2200 per second. These packages are served at the rates μ_{11} of 10000 and μ_{12} of 10000 per second. Then the packages of data arrive at the second node V_2 with probability $p_{12} = p$ (usually $p = 0.9$). The packages are served at the rates μ_{21} of 8500 and μ_{22} of 8500 per second. After service in the node V_2, the packages are routed to the first node V_1. So, $\tilde{\beta}_1 = 2000$, $\tilde{\sigma}_1 = 26153$, $\tilde{\beta}_2 = 1000$, $\tilde{\sigma}_1 = 17507$.

Consequently, (see Figures 6 and 7)

$$\frac{M\tilde{Q}_1(t)}{t} = 2000 + 26153 \cdot \sqrt{\frac{2 \ln \ln t}{t}},$$
$$\frac{M\tilde{Q}_2(t)}{t} = 1000 + 17507 \cdot \sqrt{\frac{2 \ln \ln t}{t}}.$$

Remark 5.1. When modelling an Internet network system, we apply an heuristic argument, that in

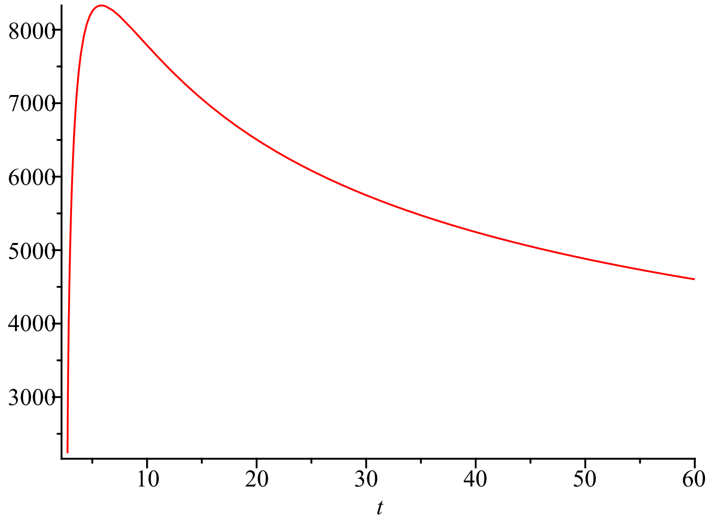

Figure 3. Values for $\frac{M\bar{Q}_2(t)}{t}$

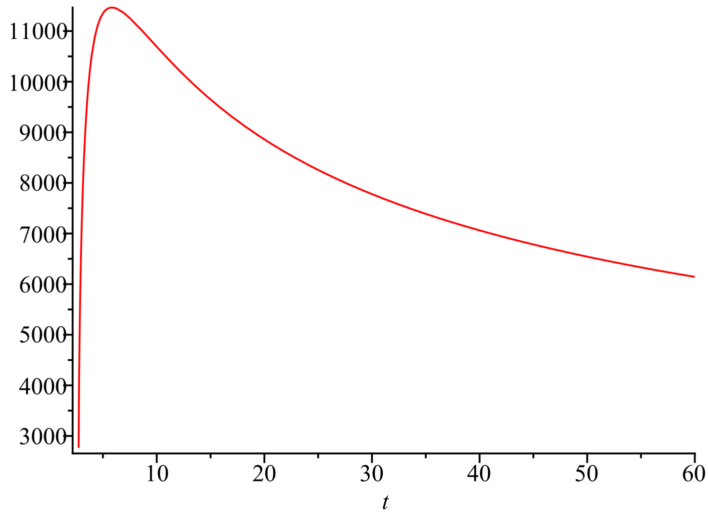

Figure 4. Values for $\frac{M\bar{Q}_3(t)}{t}$

real conditions, an average Internet network system receives 40 Mg data per second. The size of average IP package of data is about 1100 bytes. Thus, the average number of packages of data in a system is about 40000 packages per second.

6. Application of the main results

At first we present a theorem about fluid approximation for a queue of jobs in multiserver queueing networks in heavy traffic conditions.

Theorem 6.1. *If conditions* (1) *are fulfilled, then*

$$\left(\frac{Q_1(t)}{t}; \frac{Q_2(t)}{t}; \ldots; \frac{Q_J(t)}{t}\right) \Rightarrow (\beta_1; \beta_2; \ldots; \beta_J),$$

$0 \leq t \leq 1$.

Next, we present a functional limit theorem for a queue of jobs in multiserver queueing networks in heavy traffic conditions.

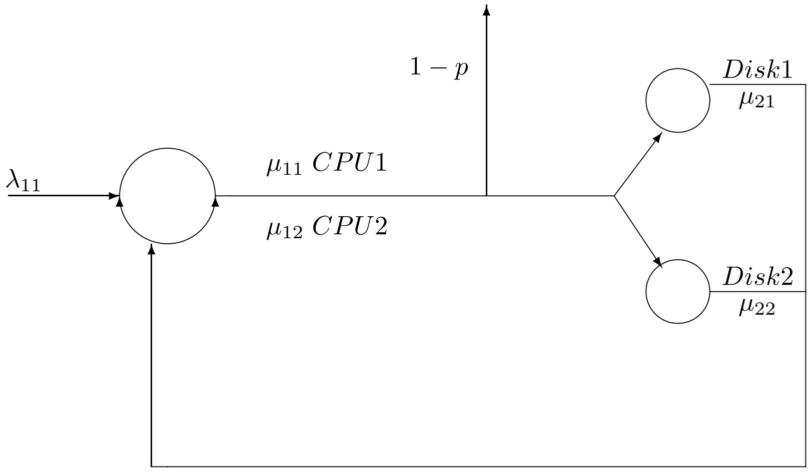

Figure 5. Model of the multiserver computer system

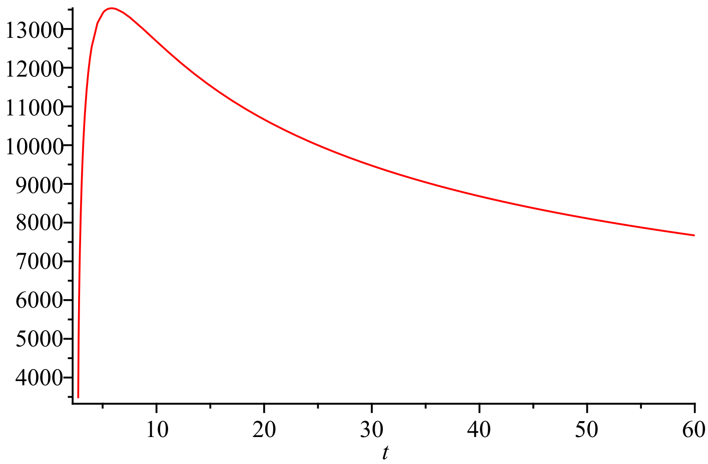

Figure 6. Values for $\frac{M\tilde{Q}_1(t)}{t}$

Theorem 6.2. *If conditions* (1) *are satisfied, then*

$$\left(\frac{Q_1(nt)-\beta_1\cdot n\cdot t}{\sigma_1\cdot\sqrt{n}};\frac{Q_2(nt)-\beta_2\cdot n\cdot t}{\sigma_2\cdot\sqrt{n}};\dots;\right.$$

$$\left.\frac{Q_J(nt)-\beta_J\cdot n\cdot t}{\sigma_J\cdot\sqrt{n}}\right)\Rightarrow(z_1(t);z_2(t);\dots;z_J(t))\,,$$

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

Proof. The proof of Theorems 6.1 and 6.2 is based on the proof of Theorem 4.1, and we omit it. □

Remark 6.1. We see that the constants $\hat{\sigma}_j>0,\ j=1,2,\dots,J,$ are the same in the formulation of Theorems 3.1 and 6.2.

Finally, we denote $P(Q_j(t)>0)$ as a probability of blocking of the node j in a multiserver computer network. So, $P(\min\limits_{1\le j\le J}Q_j(t)>0)$ is the probability of blocking of the whole multiserver computer network (because if $\min\limits_{1\le j\le J}Q_j(t)>0$, then

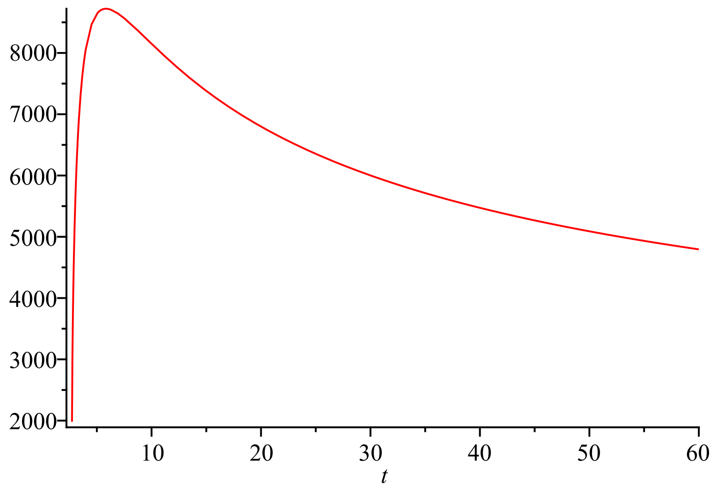

Figure 7. Values for $\frac{M\tilde{Q}_2(t)}{t}$

$Q_j(t) > 0,\ j = 1, 2, \dots, J$). We will prove the following corollary about the probability of blocking of multiserver computer network.

Corollary 6.1. *If conditions* (1) *are fulfilled, then*

$$\lim_{t\to\infty} P(\min_{1\le j\le J} Q_j(t) > 0) = 1.$$

Proof. Let us A^c denote as a complement of set A. We can write that

$$\begin{aligned}
&\lim_{t\to\infty} P((\min_{1\le j\le J} Q_j(t) > 0)^c) \\
&= \lim_{t\to\infty} P\left(\left\{\bigcap_{j=1}^{J}(Q_j(t) > 0)\right\}^c\right) \\
&= \lim_{t\to\infty} P\left(\bigcup_{j=1}^{J}(Q_j(t) > 0)^c\right) \\
&\le \lim_{t\to\infty} \sum_{j=1}^{J} P(\{Q_j(t) > 0\}^c) \\
&= \sum_{j=1}^{J} \lim_{t\to\infty} P(\{Q_j(t) > 0\}^c) \\
&= \sum_{j=1}^{J} \left(1 - \lim_{t\to\infty} P(Q_j(t) > 0)\right).
\end{aligned} \tag{8}$$

Suppose conditions (1) are satisfied. Then we prove that $\lim_{t\to\infty} P(Q_j(t) > 0) = 1,\ j = 1, 2, \dots, J$. For $j = 1, 2, \dots, J$, by applying Theorem 6.2 we obtain

$$\begin{aligned}
&\lim_{t\to\infty} P(Q_j(t) > 0) \\
&= \lim_{t\to\infty} P\left(\frac{Q_j(t) - \beta_j \cdot t}{\sigma_j \cdot \sqrt{t}} > -\frac{\beta_j \cdot t}{\sigma_j \cdot \sqrt{t}}\right) \\
&= 1 - \lim_{t\to\infty} P\left(\frac{Q_j(t) - \beta_j \cdot t}{\sigma_j \cdot \sqrt{t}} \le -\frac{\beta_j \cdot t}{\sigma_j \cdot \sqrt{t}}\right) \\
&= 1 - \Phi(-\infty) = 1,
\end{aligned} \tag{9}$$

where $\Phi(\cdot)$ is the normal distribution function.

Consequently, we conclude that (see (8) and (9))

$$\lim_{t\to\infty} P((\min_{1\le j\le J} Q_j(t) > 0)^c) = 0$$

or

$$\lim_{t\to\infty} P(\min_{1\le j\le J} Q_j(t) > 0) = 1.$$

The proof of Corollary 6.1 is complete. □

Thus, if conditions (1) are fulfilled, then the whole multiserver network is busy.

Let $\beta_{jk_j} = \sum_{k_i=1}^{J} \mu_{ik_i} \cdot p_{ij} + \lambda_j - \mu_{jk_j} > 0,\ ,\ j = 1, 2, \dots, J,\ k_j = 1, 2, \dots, c_j$. In this section, we also present the following corollary on the probability that a computer network fails due to overload.

Corollary 6.2. *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 conditions* (1) *are fulfilled, then the computer network becomes unreliable (all computers fail).*

Proof. The proof of Corollary 6.2 is presented in Minkevičius and Kulvietis [29], and we omit it. The proof of Corollary 6.2 is complete. □

7. Concluding remarks and future research

1. If the conditions of the theorem on LIL are fulfilled (i. e., conditions (1) are satisfied), then the network is occupied (see Corollary 6.1) and if conditions (1) are satisfied later on, the network becomes uncontrollable after a certain time (as $t \geq \max_{1\leq j\leq J} \max_{1\leq k_j\leq c_j} \frac{m_{jk_j}}{\beta_{jk_j}}$) (see Corollary 6.2).

2. Conditions (1) are fundamental - the behaviour of the whole network and its evolution is not clear if conditions (1) are not satisfied. Therefore, this fact is the object of further research and discussion.

3. Note that the computer with the Windows operation system functions steadily if the number of jobs does not exceed 5 (therefore, $m_{jk_j} \geq 5$). In other cases, the computer fails (see paragraph 1).

4. The theorems of this paper are proved for a class of open multiserver queueing network in heavy traffic with the service principle "first come, first served", endless waiting time of a customer in each node of the queueing system, and the times between the arrival of customers at the open multiserver queueing networks being independent identically distributed random variables. However, analogous theorems can be applied to a wider class of open multiserver queueing networks in heavy traffic: when the arrival and service of customers in a queue is by groups, when interarrival times of customers at the open multiserver queueing network are weakly dependent random variables, etc.

References

[1] **S. Asmussen.** Extreme value theory for queues via cycle maxima. *Extremes*. 1998, 1, 137-168.

[2] **N. H. Bingham.** Variants of the law of the iterated logarithm. *Bulletin London Mathematical Society*. 1986, 18, 433-467.

[3] **A. Borovkov.** Probability Processes in Theory of Queues. *Nauka, Moscow*, 1972.

[4] **H. Chen and A. Mandelbaum.** Stochastic discrete flow networks: Diffusion approximations and bottlenecks. *The Annals of Probability*. 1991, 19, 1463-1519.

[5] **H. Chen and D. Yao.** Fundamentals of queueing networks: performance, asymptotics, and optimization. *Springer*. 2001, 405 p.

[6] **C. Flores.** Diffusion approximations for computer communications networks. *In B. Gopinath, editor. Computer Communications, Proc. Syrup. Appl. Math., American Mathematical Society*. 1985, 83- 124.

[7] **P. W. Glynn.** Diffusion approximations. *In D. P. Heyman and M. J. Sobel, editors, Handbooks in Operations Research and Management Science, Vol. 2 of Stochastic Models. North-Holland*. 1990.

[8] **P. W. Glynn and W. Whitt.** Sufficient conditions for functional-limit-theorem versions of $L = \lambda W$. *Queueing Systems*. 1987, 1, 279-287.

[9] **P. W. Glynn and W. Whitt.** An LIL version of $L = \lambda W$. *Mathematics of Operations Research*. 1988, 13(4), 693-710.

[10] **J. M. Harrison and A. J. Lemoine.** A note on networks of infinite-server queues. *Journal of Applied Probability*. 1981, 18, 561-567.

[11] **D. L. Iglehart.** Multiple channel queues in heavy traffic. IV. Law of the iterated logarithm. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*. 1971, 17, 168-180.

[12] **D. L. Iglehart.** Functional limit theorems for the $GI/G/1$ queue in light traffic. *Advances in Applied Probability*. 1971, 3, 269-281.

[13] **D. L. Iglehart.** Extreme values in the $GI/G/1$ queue. *Annals of Mathematical Statistics*. 1972, 43, 627-635.

[14] **D. L. Iglehart.** Weak convergence in queueing theory. *Advances in Applied Probability*. 1973, 5, 570-594.

[15] **D. L. Iglehart and W. Whitt.** Multiple channel queues in heavy traffic I. *Advances in Applied Probability*. 1970, 2, 150-175.

[16] **D. L. Iglehart and W. Whitt.** Multiple channel queues in heavy traffic II: Sequences, networks and batches. *Advances in Applied Probability*. 1970, 2, 355-364.

[17] **D. P. Johnson.** Diffusion Approximations for Optimal Filtering of Jump Processes and for Queueing Networks. *Ph.D. dissertation, University of Wisconsin*. 1983.

[18] **F. I. Karpelevich and A. I. Kreinin.** Heavy Traffic Limits for Multiphase Queues. *American Mathematical Society, Providence, Rhode Island*. 1994.

[19] **F. I. Karpelevich and A. I. Kreinin.** Asymptotic analysis of queueing systems with identical service. *J. Applied Probability*. 1996, 33, 267-281.

[20] **D. Kendall.** Some problems in the theory of queues. *J. R. Statist. Soc.* 1961, Ser. B, 13, 161-185.

[21] **J. Kingman.** On queues in heavy traffic. *J. R. Statist. Soc.* 1962, 24, 383-392.

[22] **J. Kingman.** The single server queue in heavy traffic. *Proc. Camb. Phil. Soc.* 1962, 57, 902- 904.

[23] **H. Kobyashi.** Application of the diffusion approximation to queueing networks. *Journal of ACM*. 1974, 21, 316-328.

[24] **A. J. Lemoine.** Network of queues – A survey of weak convergence results. *Management Science*. 1978, 24, 1175-1193.

[25] **S. Minkevičius.** On the law of the iterated logarithm in multiphase queueing systems. *Lietuvos Matematikos Rinkinys*. 1995, 35 (3), 360–366.

[26] **S. Minkevičius.** On the law of the iterated logarithm

in multiphase queueing systems. II. *Informatica*. 1997, 8 (3), 367–376.

[27] **S. Minkevičius, G. Kulvietis.** Application of the law of the iterated logarithm in open queuing net-works. *WSEAS Transactions on Systems*. 2007, 6, 643-651.

[28] **S. Minkevičius, L. Sakalauskas.** Nonlinear phenomena in digital networks. *Information Technology Control*. 2010, 39, Issue 3, 187-194.

[29] **S. Minkevičius, G. Kulvietis.** Investigation of the reliability of multi-server computer networks. *Proceedings of ASMTA 2011, 18th International Conference on Analytical and Stochastic Modeling Techniques and Applications, 20-22 June 2011, Venice, Italy, Lecture Notes of Computer Sciences.* 2011, 6751, 249-256.

[30] **G. Pang, R. Talreja, W. Whitt.** Martingale proofs of many-server heavy-traffic limits for Markovian queues. *Probability Surveys.* 2007, 4, 193-267.

[31] **Y. Prokhorov.** Transient phenomena in queues. *Lietuvos Matematikos Rinkinys.* 1963, 3, 199-206.

[32] **A. Puhalskii, J. Reed.** On many-server queues in heavy traffic. *The Annals of Applied probability.* 2010, 20 (1), 129-195.

[33] **M. Reiman.** Open queuing networks in heavy traffic. *Mathematics of Operations Research.* 1984, 9, 441-459.

[34] **L. Sakalauskas, S. Minkevičius.** On the law of the iterated logarithm in open queuing networks. *European Journal of Operational Research.* 2000, 120, 632–640.

[35] **V. Strassen.** An invariance principle for the law of the iterated logarithm. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete.* 1964, 3, 211–226.

[36] **W. Whitt.** Heavy traffic limit theorems for queues: a survey. In: *Lecture Notes in Economics and Mathematical Systems.* 1974, 98, 307-350. Springer-Verlag, Berlin, Heidelberg, NewYork.

[37] **W. Whitt.** On the heavy-traffic limit theorem for *GI/G/∞* queues. *Advances in Applied Probability.* 1982, 14, 171-190.

[38] **W. Whitt.** Stochastic-Process Limits. An Introduction to Stochastic-Process Limits and their Application to Queues. *Springer Series in Operations Research and Financial Engineering*. 2002, 602 p.

[39] **W. Whitt.** Heavy-traffic limits for the G/H2*/n/m queue. *Mathematics of Operations Research*. 2005, 30 (1), 1-27.

[40] **W. Whitt.** Fluid models for multiserver queues with abandonments. *Operations Research.* 2006, 54 (1), 37-54.

Received May 2011.