PERGAMON

Mathematical and Computer Modelling 30 (1999) 65–78

MATHEMATICAL AND COMPUTER MODELLING

www.elsevier.nl/locate/mcm

Averaging Methods for Transient Regimes in Overloading Retrial Queueing Systems

V. V. ANISIMOV
Department of Industrial Engineering
Bilkent University, Ankara, Turkey
vlanis@bilkent.edu.tr
and
Applied Statistics Department
Kiev University, Kiev, Ukraine
vlanis@stoch.univ.kiev.ua

Abstract—A new approach is suggested to study transient and stable regimes in overloading retrial queueing systems. This approach is based on limit theorems of averaging principle and diffusion approximation types for so-called switching processes. Two models of retrial queueing systems of the types $\bar{M}/\bar{G}/\bar{1}/w.r$ (multidimensional Poisson input flow, one server with general service times, retrial system) and $M/M/m/w.r$ (m servers with exponential service) are considered in the case when the intensity of calls that reapply for the service tends to zero. For the number of re-applying calls, functional limit theorems of averaging principle and diffusion approximation types are proved.

Keywords—Retrial queueing systems, Averaging principle, Diffusion approximation, Switching processes, Markov processes.

1. INTRODUCTION

The complexity of a real model of information and computing systems leads to the necessity of the creation of new, more complicated models of queueing systems and developing new approaches in the investigations. Taking into account a complex structure of real systems, only in rare special cases is it possible to get analytic solutions for various characteristics. Therefore, asymptotic methods play a basic role in investigation and approximative modelling.

The paper is devoted to the development methods of asymptotic analysis such as averaging methods and methods of diffusion approximation for so-called retrial queueing systems. In these systems, customers finding the service busy may join the special retrial queue and repeat their attempts for service after some random time.

In recent years, there have appeared many publications concerning the development of theoretical, numerical, and approximating methods for different classes of retrial queueing models (see [1–15]).

Different asymptotic approaches for various classes of general queueing systems are considered in [16–22] (see also, references therein).

In the present paper, a new approach is suggested based on asymptotic results of averaging principle and diffusion approximation types for so-called switching processes.

0895-7177/99/$ - see front matter. © 1999 Elsevier Science Ltd. All rights reserved. Typeset by $\mathcal{A}_{\mathcal{M}}\mathcal{S}$-TEX
PII:S0895-7177(99)00132-6

The class of Switching Processes (SP) was introduced in the author's works (see [23,24]). SP are described as two-component processes $(x(t), \zeta(t))$, $t \geq 0$, with the property having a sequence of epochs $t_1 < t_2 < \cdots$ such that on each interval $[t_k, t_{k+1})$, $x(t) = x(t_k)$ and the behaviour of the process $\zeta(t)$ depends only on the value $(x(t_k), \zeta(t_k))$. This means that the character of development of SP varies spontaneously (switches) at certain epochs of time which can be random functionals of previous trajectory. Epochs t_k are switching times and $x(t)$ is the discrete switching component.

SP can be described in terms of constructive characteristics and they are very suitable in analyzing and investigating stochastic systems with "rare" and "fast" switches (see [18,24–27]), and they also can be adequate mathematical models at the analytical and approximate handling wide classes of queueing systems and networks.

We note that SP are in some sense the generalization of well-known classes of random processes such as Markov processes homogeneous in the second component [28]; processes with independent increments and semi-Markov switches [29], piecewise Markov aggregates [16]; Markov processes with semi-Markov interference of chance [30], and Markov and semi-Markov evolutions [31–34].

Two large classes of limit theorems were investigated for SP in triangular scheme. For the first class (the number of switches does not tend to infinity), convergence theorems in the class of SP were proved in [18,24]. These results made it possible to construct the theory of asymptotic enlargement (merging) of states and decreasing dimension for nonhomogeneous Markov and semi-Markov processes and to obtain various applications in queueing theory. In particular, several results devoted to the asymptotic analysis of flows of rare events on trajectories of SP were obtained in books [18,35] and applications to the analysis of highly-reliable systems with repeated calls were given by [12].

In the case of fast switches (number of switches tends to infinity), both the *averaging principle* and the *diffusion approximation* for different subclasses of SP were proved in [25,26,36–38]. Applications of these results to study asymptotic behaviour of characteristics of Markov queueing systems and networks under transient conditions and with large number of calls were investigated in [19,25–27]. Some applications to retrial queueing systems were obtained in [13,14].

In this paper, we give a general description of SP and we study some classes of retrial queueing systems in overloading case and in transient conditions. The method of investigation is based on averaging principle and diffusion approximation type theorems for SP which are given in the Appendix.

2. SWITCHING STOCHASTIC PROCESSES

In this section, we give the general definition and consider some subclasses of SP.

2.1. Simple Recurrent Process of Semi-Markov Type

Let $F_k = \{(\xi_k(\alpha), \tau_k(\alpha)), \alpha \in R^r\}$, $k \geq 0$, be jointly independent families of random variables with values in $R^r \times [0, \infty)$ and S_0 be independent of F_k, $k \geq 0$, random variable in R^r. Denote

$$t_0 = 0, \quad t_{k+1} = t_k + \tau_k(S_k), \quad S_{k+1} = S_k + \xi_k(S_k), \qquad k \geq 0, \tag{2.1}$$

and put

$$S(t) = S_k, \qquad \text{as } t_k \leq t < t_{k+1}, \quad t \geq 0. \tag{2.2}$$

Then, the process $S(t)$ forms a simple Recurrent Process of Semi-Markov type (RPSM) (see [25,26,36]).

In the case when the distributions of families F_k do not depend on the parameter k, the process $S(t)$ is a homogeneous SMP (Semi-Markov Process). Moreover, when the distributions of families F_k do not depend on the parameter α, epochs t_k, $k \geq 0$ form a recurrent flow, and $S(t)$ is a generalized renewal process. In particular, when the variables $\tau_k(\alpha)$ have exponential distributions, process $S(t)$ is an MP (Markov Process).

2.2. General Case of RPSM

Let

$$F_k = \{(\xi_k(x,\alpha), \tau_k(x,\alpha), \beta_k(x,\alpha)), x \in X, \alpha \in R^r\}, \qquad k \geq 0$$

be jointly independent families of random variables taking values in the space $R^r \times [0,\infty) \times X$, X be some measurable space, (x_0, S_0) be an initial value. We put

$$\begin{aligned} t_0 &= 0, & t_{k+1} &= t_k + \tau_k(x_k, S_k), \\ S_{k+1} &= S_k + \xi_k(x_k, S_k), & x_{k+1} &= \beta_k(x_k, S_k), \quad k \geq 0, \qquad (2.3)\\ S(t) &= S_k, & x(t) &= x_k, \qquad \text{as } t_k \leq t < t_{k+1}, \quad t \geq 0. \qquad (2.4) \end{aligned}$$

Then the pair $(x(t), S(t))$, $t \geq 0$ forms a general RPSM with feedback between both components.

In particular, when the distribution of the variables $\beta_k(x,\alpha)$ does not depend on the parameter α, the sequence x_k forms an MP and we obtain an RPSM with additional Markov switches.

2.3. Switching Processes

Now we consider a general construction of an SP. Let

$$F_k = \{(\zeta_k(t,x,\alpha), \tau_k(x,\alpha), \beta_k(x,\alpha)), t \geq 0, x \in X, \alpha \in R^r\}, \qquad k \geq 0$$

be jointly independent parametric families where $\zeta_k(t,x,\alpha)$ for each fixed k, x, α is a random process belonging to Skorokhod space D^r_∞ and $\tau_k(x,\alpha), \beta_k(x,\alpha)$ be possibly dependent on $\zeta_k(\cdot,x,\alpha)$ random variables, $\tau_k(\cdot) > 0$, $\beta_k(\cdot) \in X$. Also, let (x_0, S_0) be an initial value. We put

$$\begin{gathered} t_0 = 0, \qquad t_{k+1} = t_k + \tau_k(x_k, S_k), \quad S_{k+1} = S_k + \xi_k(x_k, S_k), \\ x_{k+1} = \beta_k(x_k, S_k), \qquad k \geq 0, \end{gathered} \tag{2.5}$$

where $\xi_k(x,\alpha) = \zeta_k(\tau_k(x,\alpha), x, \alpha)$, and set

$$\zeta(t) = S_k + \zeta_k(t - t_k, x_k, S_k), \quad x(t) = x_k, \qquad \text{as } t_k \leq t < t_{k+1}, \quad t \geq 0. \tag{2.6}$$

Then, a two-component process $(x(t), \zeta(t))$, $t \geq 0$ is called an SP (see [18,23, 24]).

If variables $\beta_k(x,\alpha)$ do not depend on the argument α, then the sequence x_k forms itself an MP. Let, in addition, $\{\zeta(t,x), t \geq 0\}$ be a family of Markov processes and $\zeta(t,x,\alpha)$ denotes the process $\zeta(t,x)$ with initial value α. In that case, the process $(x(t), \zeta(t))$ forms a Markov (when the process $x(t)$ is an MP) or semi-Markov (when the process $x(t)$ is an SMP) random evolution (see [31–34]).

3. ASYMPTOTIC ANALYSIS OF SWITCHING RETRIAL SYSTEMS

In this section, we consider applications of limit theorems for SP given in the Appendix to the asymptotic analysis of some classes of retrial queueing systems.

3.1. System $\bar{M}/\bar{G}/\bar{1}/w.r$

Let us consider a one-server system with multiple Poisson input (a call of type i has the rate λ_i, $i = \overline{1,r}$, $r < \infty$). Let there also be given a family of distribution functions $\{F_i(x), i = \overline{1,r}\}$, $(F_i(0) = 0)$, values $\{q_i, i = \overline{1,r}\}$, $(0 \leq q_i \leq 1)$, and a family of continuous functions $\{\nu_i(\bar{s}), i = \overline{1,r}, \bar{s} \in R^r_+\}$. If a call of type i enters the system and finds the server idle, then with probability q_i the service immediately begins, the service times κ_i are independent random variables with distribution function $F_i(x)$. Otherwise, with probability $p_i = 1 - q_i$, it can get a refusal from a service, and in that case, it will attempt to be served later. Calls waiting to try again for service are said to be in the 'orbit'. If the call finds the server busy, it directly goes in 'orbit'.

Denote $\bar{S}_n(t) = \{S_n^{(i)}(t), i = \overline{1,r}\}$ where $S_n^{(i)}(t)$ is the number of calls of the type i in the orbit. If $\bar{S}_n(t) = n\bar{s}$, then on the small interval $[t, t+h]$, each call in the orbit independently of others can reapply for service with probability $(1/n)\nu(\bar{s})h + o(h)$. If a call finds the server idle, then the server immediately begins to serve it with service time κ_i. If a call finds the server busy, then it returns to the orbit.

Let $\bar{s} = (s_1, s_2, \dots, s_r)$ be a column vector. By symbol $\bar{s}^*$, we denote the conjugate vector. Suppose that there exist the expected values

$$\mathbf{E}\kappa_i = m_i, \qquad i = \overline{1,r}. \tag{3.1}$$

Let us introduce the following variables:

$$\lambda(\bar{s}) = \sum_{i=1}^{r} (\lambda_i q_i + s_i \nu_i(\bar{s})), \qquad \hat{m} = \sum_{i=1}^{r} m_i \lambda_i q_i. \tag{3.2}$$

Also, let $\bar{\lambda}, \bar{\alpha}(\bar{s})$ be column vectors with elements λ_i, $m_i\nu_i(\bar{s})$, respectively, and matrix $M = \bar{\lambda}\bar{\alpha}(\bar{s})^*$. We denote by I_P, I_Q, and G diagonal matrices with elements on the diagonal p_i, q_i, and $\nu_i(\bar{s})$, correspondingly, and put

$$g(\bar{s}) = 1 + \hat{m} + (\bar{\alpha}(\bar{s}), \bar{s}), \qquad \bar{A}(\bar{s}) = I_P\bar{\lambda} + \hat{m}\bar{\lambda} + (M - G)\bar{s}. \tag{3.3}$$

Below we give two theorems that show the asymptotic behaviour of the vector $\bar{S}_n(nt)$ when $n \to \infty$.

THEOREM 3.1. AVERAGING PRINCIPLE. *Suppose that $n^{-1}\bar{S}_n(0) \xrightarrow{P} \bar{s}_0$, functions $\nu_i(\bar{s})$ satisfy the local Lipschitz condition, and*

$$m_i > 0, \qquad i = \overline{1,r}. \tag{3.4}$$

Then, for any $T > 0$,

$$\sup_{0 \le t \le T} \left| n^{-1}\bar{S}_n(nt) - \bar{s}(t) \right| \xrightarrow{P} 0, \qquad n \to \infty, \tag{3.5}$$

where

$$\bar{s}(0) = \bar{s}_0, \qquad d\bar{s}(t) = g(\bar{s}(t))^{-1}\bar{A}(s(t))\,dt, \tag{3.6}$$

and a unique solution of the equation (3.6) exists on each interval. (Here, symbol $\xrightarrow{P}$ means convergence in probability.)

PROOF. At first, we represent the process $\bar{S}_n(t)$ as an SP. Let us denote by $t_{n1} < t_{n2} < t_{n3} < \cdots$ sequential times of finishing service. We consider these times as switching times. Denote $\bar{S}_{nk} = \bar{S}_n(t_{nk})$, $k > 0$, and introduce the family of random variables $\tau_n(\bar{s})$ such that

$$\mathbf{P}\{\tau_n(\bar{s}) \le x\} = \mathbf{P}\left\{t_{nk+1} - t_{nk} \le x \mid \bar{S}_{nk} = n\bar{s}\right\}.$$

If the server is free, under the condition that $\bar{S}_{nk} = n\bar{s}$, we have two flows of calls. In the first one, the call of type i has the intensity $q_i\lambda_i$, the second one consists of the calls in the orbit and the call of type i has the intensity $s_i\nu_i(\bar{s})$. Therefore, we can represent the variable $\tau_n(\bar{s})$ in the form

$$\tau_n(\bar{s}) = \eta(\lambda(\bar{s})) + \kappa(\bar{s}), \tag{3.7}$$

where $\eta(\lambda(\bar{s}))$ is an exponentially distributed random variable with parameter $\lambda(\bar{s})$ and η, $\kappa(\bar{s})$ are independent, and $\kappa(\bar{s})$ can be represented in the form

$$\kappa(\bar{s}) = \begin{cases} \kappa_j \text{ with probability } \lambda(\bar{s})^{-1}(\lambda_j q_j + s_j \nu_j(\bar{s})), \\ j = \overline{1,r} \end{cases}.$$

Introducing indicators $\chi_{j1}(\bar{s})$ (correspondingly, $\chi_{j2}(\bar{s})$) of the following events: after an idle period, a call of type j which comes from the input flow (correspondingly, from the orbit) occupies the server. It means that

$$\mathbf{P}\{\chi_{j1}(\bar{s}) = 1\} = 1 - \mathbf{P}\{\chi_{j1}(\bar{s}) = 0\} = \lambda(\bar{s})^{-1}\lambda_j q_j,$$
$$\mathbf{P}\{\chi_{j2}(\bar{s}) = 1\} = 1 - \mathbf{P}\{\chi_{j2}(\bar{s}) = 0\} = \lambda(\bar{s})^{-1}s_j\nu_j(\bar{s}).$$

According to these notations, we can write that

$$\tau_n(\bar{s}) = \eta(\lambda(\bar{s})) + \sum_{j=1}^{r} (\chi_{j1}(\bar{s}) + \chi_{j2}(\bar{s}))\,\kappa_j. \tag{3.8}$$

In our case, we have no discrete component $x(t)$. Let us define now the family of processes $\zeta_{nk}(t,\bar{s})$. According to the construction on the idle interval, we have a Poisson flow with parameter $\lambda_i p_i$ of calls of type i which go directly to the orbit, and on the busy period, we have a Poisson flow with parameter λ_i of calls of type i which go directly to the orbit. Denote by $\bar{\Pi}^{(k)}_{a_i}(t) = \{\Pi^{(k)}_{a_i}(t), i = \overline{1,r}\}$, $k \geq 0$, vector-valued jointly independent at different k Poisson processes for which components $\Pi^{(k)}_{a_i}(t)$ are independent Poisson processes with parameters a_i. Suppose, without loss of generality, that at time $t_{n0} = 0$, the server is idle. Then, on the interval $[t_{n0}, t_{n1}]$, we introduce the process

$$\begin{aligned} \bar{\zeta}_{n0}(t,\bar{s}) &= \bar{\Pi}^{(0)}_{p_i\lambda_i}(t), \qquad \text{as } t < \eta(\lambda(\bar{s})), \\ \bar{\zeta}_{n0}(t,\bar{s}) &= \bar{\Pi}^{(0)}_{p_i\lambda_i}(\eta(\lambda(\bar{s}))) - \sum_{j=1}^{r} \bar{e}_j\chi_{j2}(\bar{s}) + \sum_{j=1}^{r}(\chi_{j1}(\bar{s}) \\ &\quad + \chi_{j2}(\bar{s}))\,\bar{\Pi}^{(j)}_{\lambda_i}(t), \qquad \text{as } \eta(\lambda(\bar{s})) < t \leq \tau_n(\bar{s}), \end{aligned} \tag{3.9}$$

where $\bar{e}_j$ is the column vector in which the j^{th} component is equal to 1 and others are equal to 0. On each interval $[t_{nk}, t_{nk+1})$, the process $\bar{\zeta}_{nk}(t,\bar{s})$ is constructed in the same way.

Then, the process $\bar{S}_n(t)$ is equivalent to the SP constructed by families $\bar{\zeta}_{nk}(t,\bar{s})$ according to formulae (2.5),(2.6) (without component $x(\cdot)$).

Let us introduce the family of vector-valued variables $\bar{\xi}_n(\bar{s})$ such that

$$\mathbf{P}\left\{\bar{\xi}_n(\bar{s}) \leq x\right\} = \mathbf{P}\left\{\bar{S}_{nk+1} - \bar{S}_{nk} \leq x \mid \bar{S}_{nk} = n\bar{s}\right\}.$$

According to the construction, we can represent $\bar{\xi}_n(\bar{s})$ in the form

$$\bar{\xi}_n(\bar{s}) = \bar{\Pi}^{(0)}_{p_i\lambda_i}(\eta(\lambda(\bar{s}))) - \sum_{j=1}^{r} \bar{e}_j\chi_{j2}(\bar{s}) + \sum_{j=1}^{r}(\chi_{j1}(\bar{s}) + \chi_{j2}(\bar{s}))\,\bar{\Pi}^{(j)}_{\lambda_i}(\kappa_j). \tag{3.10}$$

Now we use Theorem A.1 from the Appendix. For simplicity, we omit index k and index n where it is possible. It is easy to calculate that

$$\begin{aligned} m(\bar{s}) &= \mathbf{E}\tau_n(\bar{s}) = \lambda(\bar{s})^{-1}\left(1 + \sum_{i=1}^{r}(\lambda_i q_i + s_i\nu_i(\bar{s}))m_i\right) = \lambda(\bar{s})^{-1}g(\bar{s}), \\ \bar{b}(\bar{s}) &= \mathbf{E}\bar{\xi}_n(\bar{s}) = a(\bar{s})^{-1}\left(I_P\lambda + \hat{m}\hat{\lambda} + (M - G)\bar{s}\right). \end{aligned} \tag{3.11}$$

Then, in our case, $g_n(\bar{s}) \le |\xi_n(\bar{s})| + 1$ and condition (A.5) automatically takes place. Now we prove that the convergence in (3.5) takes place for any $T > 0$. It is easy to see that

$$m(\bar{s}) \ge \lambda(\bar{s})^{-1} + \min_i m_i \ge \min_i m_i,$$

and according to the condition (3.4) in (A.9), $\int_0^\infty m(\eta(u))du = +\infty$.

Further, the function $\bar{A}(\bar{s}) = m(\bar{s})^{-1}b(\bar{s})$ satisfies the local Lipschitz condition and has no more than linear growth. This means that the solution of equation (3.6) exists on each interval and is unique and finally implies Theorem 3.1

Let us consider as an example a one-dimensional case (only one type of calls). We keep the previous notations and just omit the index i and symbol $^-$. Then we get

$$\begin{aligned} &\lambda(s) = \lambda q + s\nu, \qquad m(s) = \lambda(s)^{-1} + m, \qquad g(s) = 1 + \lambda qm + \nu ms, \\ &A(s) = \lambda p + \lambda^2 qm + (\lambda m - 1)s\nu(s). \end{aligned} \tag{3.12}$$

COROLLARY 3.1. *Suppose that* $\mathbf{E}\kappa = m > 0$, $n^{-1}S_n(0) \xrightarrow{P} s_0$, *and function* $\nu(s)$ *satisfies local Lipschitz condition. Then, the relation (3.5) takes place where in equation (3.6), the functions* $g(s), A(s)$ *are given in the expression (3.12).*

We mention that if $\lambda m < 1$ and $s\nu(s) \to \infty$ as $s \to \infty$, equation (3.6) has a point of stability s_*, which is the minimal solution of the equation

$$s\nu(s) = (1 - \lambda m)^{-1}\lambda(p + \lambda mq).$$

In particular, if $\nu(s) \equiv \nu$, then $s_* = (1 - \lambda m)^{-1}\nu^{-1}\lambda(p + \lambda mq)$ and $s(t) \to s_*$ as $t \to \infty$.

In the case where $\lambda m = 1$, we get unusual behaviour for $s(t)$,

$$s(t) = \frac{1}{m\nu}\left(\sqrt{2\lambda m\nu(\lambda qm + p)t + (1 + \lambda qm + m\nu s_0)^2} - 1 - \lambda qm\right).$$

Now we will study the diffusion approximation. We keep the notations of Theorem 3.1. Suppose that there exist $\mathbf{E}\kappa_i^2$, $i = \overline{1,r}$. Put $\sigma_i^2 = \operatorname{Var} \kappa_i$ and introduce the following variables:

$$\begin{aligned} &\hat{m}(\bar{s}) = \sum_{i=1}^r m_i s_i \nu_i, \qquad \alpha_j(\bar{s}) = \lambda(\bar{s})^{-1}\left(\lambda_j q_j + s_j\nu_j\right), \\ &\hat{\sigma}^2(\bar{s}) = \sum_{j=1}^r \alpha_j(\bar{s})\left(\sigma_j^2 + \left(m_j - \sum_{i=1}^r \alpha_i(\bar{s})m_i\right)^2\right). \end{aligned} \tag{3.13}$$

Put $f(\bar{s}) = \lambda(\bar{s})^{-1}(\hat{m} + \hat{m}(\bar{s}))$ and introduce vectors

$$\bar{j}(\bar{s}) = -g(\bar{s})^{-1}f(\bar{s})I_Q\bar{\lambda} + \lambda(\bar{s})^{-1}G\bar{s}, \qquad \bar{a}(\bar{s}) = g(\bar{s})^{-1}\left(I_Q\bar{\lambda} + G\bar{s}\right).$$

Let $\bar{\beta}(\bar{s})$ and $\bar{\beta}_m(\bar{s})$ be column vectors with components $\beta_i(\bar{s})$ and $\beta_i(\bar{s})m$, correspondingly, where $\beta_i(\bar{s}) = \lambda(\bar{s})^{-1}s_i\nu_i(\bar{s})$. We put $B(\bar{s}) = \bar{j}(\bar{s})\bar{\beta}(\bar{s})^* + \bar{a}(\bar{s})\bar{\beta}_m(\bar{s})^*$. Also, let Λ and $\Lambda_1(\bar{s})$ be diagonal matrices with elements λ_i on the diagonal and $p_i\lambda_i + s_i\nu_i$, correspondingly. Denote

$$\begin{aligned} D^2(\bar{s}) = {} & g(\bar{s})^{-2}\hat{\sigma}^2(\bar{s})\left(I_Q\bar{\lambda} + G\bar{s}\right)\left(I_Q\bar{\lambda} + G\bar{s}\right)^* + \lambda(\bar{s})^{-2}G\bar{s}\bar{s}^*G \\ & + g(\bar{s})^{-2}\lambda(\bar{s})^{-2}\left(\lambda(\bar{s})f(\bar{s})I_Q\bar{\lambda} - G\bar{s}\right)\left(\lambda(\bar{s})f(\bar{s})I_Q\bar{\lambda} - G\bar{s}\right)^* \\ & - B(\bar{s}) - B(\bar{s})^* + \lambda(\bar{s})^{-1}\Lambda_1(\bar{s}) + f(\bar{s})\Lambda. \end{aligned} \tag{3.14}$$

Further, suppose that functions $\nu_i(\bar{s})$ are continuously differentiable and denote by $Q(\bar{s}) = (g(\bar{s})^{-1}A(\bar{s}))'$ a matrix derivative of the vector $g(\bar{s})^{-1}A(\bar{s})$. Put

$$\hat{\gamma}_n(t) = \frac{1}{\sqrt{n}}\left(\bar{S}_n(nt) - n\bar{s}(t)\right), \qquad t \in [0, T].$$

THEOREM 3.2. DIFFUSION APPROXIMATION. *Suppose that conditions of Theorem 3.1 hold and* $n^{-1/2}(\bar{S}_n(0) - \bar{s}_0) \stackrel{w}{\Rightarrow} \bar{\gamma}_0$.

Then the sequence of processes $\bar{\gamma}_n(t)$ *J-converges on any interval* $[0,T]$ *to the diffusion process* $\bar{\gamma}(t)$ *which satisfies the following stochastic differential equation:*

$$d\bar{\gamma}(t) = Q(\bar{s}(t))\bar{\gamma}(t)\,dt + D(\bar{s}(t))m(\bar{s}(t))^{-1/2}\,d\bar{w}(t), \qquad \bar{\gamma}(0) = \bar{\gamma}_0, \tag{3.15}$$

where the function $\bar{s}(\cdot)$ *satisfies equation (3.6), the function* $m(\bar{s})$ *is given in (3.11), and* $\bar{w}(t)$ *is the standard Wiener process in* R^r.

PROOF. We use the same representation for variables $\tau_n(\bar{s})$ and $\xi_n(\bar{s})$ (see (3.8),(3.10)). According to Theorem A.2, we need to calculate the variance of the variable

$$\bar{\rho}_n(\bar{s}) = \bar{\xi}_n(\bar{s}) - \bar{b}_n(\bar{s}) - g(\bar{s})^{-1}\bar{A}(\bar{s})(\tau_n(\bar{s}) - m(\bar{s})),$$

(see (3.11)). For convenience, we can split $\bar{\rho}_n(\bar{s})$ into two independent parts,

$$\bar{\rho}_n^{(1)}(\bar{s}) = \bar{\Pi}^{(0)}_{\lambda_i p_i}(\eta(\lambda(\bar{s}))) - \lambda(\bar{s})^{-1}I_P\bar{\lambda} - g(\bar{s})^{-1}\bar{A}(\bar{s})\left(\eta_n(\lambda(\bar{s})) - \lambda(\bar{s})^{-1}\right),$$

and the remaining part. After calculations, we obtain that $\mathbf{E}\bar{\rho}_n(\bar{s})\bar{\rho}_n(\bar{s})^* = D^2(\bar{s})$, which implies the statement of Theorem 3.2.

REMARK 3.1. If equation (3.6) has the point of stability $\bar{s}_*$ and $\bar{s}_0 = \bar{s}_*$, then we have so-called quasi-stationary regime in which $\bar{s}(t) \equiv \bar{s}_*$ and the process $\bar{\gamma}(t)$ in Theorem 3.2 satisfies the equation

$$d\bar{\gamma}(t) = Q(\bar{s}_*)\bar{\gamma}(t)\,dt + D(\bar{s}_*)m(\bar{s}_*)^{-1/2}\,dw(t), \qquad \bar{\gamma}(0) = \bar{\gamma}_0,$$

which describes the Ornstein-Uhlenbeck process.

We mention that some similar models for a one-dimensional case were considered in [13,14,19].

3.2. System $M/M/m/w.r$

Now we consider a system with m identical servers and service intensity μ. The input is a Poisson flow of identical calls with parameter λ. Denote by $Q_n(t)$ the number of busy servers at time t. Let there be given families $\{p_i(s), q_i(s), r_i(s), i = \overline{0,m}\}$ and $\{\nu(s), \alpha(s), g(s)\}$, $s \geq 0$, of continuous nonnegative functions. Here,

$$p_i(s) + q_i(s) + r_i(s) = 1, \quad \alpha(s) + g(s) = 1, \qquad \text{for any } s \geq 0, \quad i = \overline{0,m}.$$

Let $S_n(t)$ denote the number of calls in the orbit at time t. If a call enters the system at time t and $(Q_n(t), S_n(t)) = (i, ns)$, $(i \leq r)$, then with probability $p_i(s)$, service immediately begins and with probability $q_i(s)$, the call directly goes to the orbit, and with probability $r_i(s)$, the call gets a refusal and leaves the system. If $i = m$, we put $p_m(s) \equiv 0$.

If $S_n(t) = ns$, each call in the orbit independently of others can reapply for service with local intensity at time t $(1/n)\nu((1/n)S_n(t)) = (1/n)\nu(s)$. If a call finds an idle server, service immediately begins. If a call finds all servers busy, then it returns to the orbit with probability $\alpha(s)$ or with probability $g(s)$, it leaves the system.

We study an averaging principle for the process $(1/n)S_n(nt)$. Denote

$$\rho(j,s) = \hat{c}(s)^{-1}\frac{1}{j!\mu^j}\prod_{i=0}^{j-1}(p_i(s)\lambda + s\nu(s)), \qquad j = \overline{0,m},$$

where (3.16)

$$\hat{c}(s) = \sum_{j=0}^{m}\frac{1}{j!\mu^j}\prod_{i=0}^{j-1}(p_i(s)\lambda + s\nu(s)), \qquad \text{and} \qquad \prod_{i=0}^{-1} = 1.$$

Let us define the function

$$\hat{b}(s) = \lambda \sum_{i=0}^{m} \rho(i,s) q_i(s) - s\nu(s)(1 - (1 - g(s))\rho(m,s)). \tag{3.17}$$

THEOREM 3.3. *Suppose that $n^{-1}\bar{S}_n(0) \xrightarrow{P} \bar{s}_0$, functions $p_i(s)$, $q_i(s)$, $g(s)$, $\nu(s)$ satisfy local Lipschitz condition and function $\nu(s)$ is bounded. Then, for any $T > 0$,*

$$\sup_{0 \le t \le T} \left| n^{-1} S_n(nt) - s(t) \right| \xrightarrow{P} 0, \qquad n \to \infty, \tag{3.18}$$

where

$$s(0) = s_0, \qquad \mathrm{d}s(t) = \hat{b}(s(t))\,\mathrm{d}t, \tag{3.19}$$

and a unique solution of equation (3.19) exists on any interval.

PROOF. At first we represent the process $(Q_n(t), S_n(t))$ as an SP. In our case, process $(Q_n(t), S_n(t))$ is a Markov process with values in $\{0,1,\dots,m\} \times \{0,1,\dots\}$. Denote by $t_{n1} < t_{n2} < \cdots$ the sequential times of any transition in the system. We mention that some transitions may not cause the changing of the state, but they are connected with some service processes (for instance, loss of an input call). We consider times $t_{n1} < t_{n2} < \cdots$ as switching times. In the scale of time nt, the first component is quickly varying and we will use Theorem A.3.

Define the process $(Q_n(t), S_n(t))$ as a right-continuous process and let us write the intensities of transition probabilities. Put $\lambda_i(s) = \lambda + i\mu + s\nu(s)$, $i \le m$. If $S_n(t) = ns$, then transition probabilities do not depend on n and we omit index n for simplicity. Let $\lambda((i,ns),(j,ny))$ denote the intensity of transition probability from the state $(Q_n(t), S_n(t)) = (i, ns)$ to the state (j, ny). Then,

$$\lambda((i,s),(j,y)) = \begin{cases} i\mu, & \text{if } j = i-1,\ y = s, \\ p_i(s)\lambda, & \text{if } j = i+1,\ y = s, \\ r_i(s)\lambda, & \text{if } j = i,\ y = s, \\ q_i(s)\lambda, & \text{if } j = i,\ y = s+1, \\ s\nu(s), & \text{if } j = i+1,\ y = s-1, \\ 0, & \text{otherwise},\ 0 \le i < m, \end{cases}$$

$$\lambda((m,s),(j,y)) = \begin{cases} m\mu, & \text{if } j = m-1,\ y = s, \\ r_m(s)\lambda, & \text{if } j = m,\ y = s, \\ q_m(s)\lambda, & \text{if } j = m,\ y = s+1, \\ g(s)s\nu(s), & \text{if } j = m,\ y = s-1, \\ \alpha(s)s\nu(s), & \text{if } j = m,\ y = s, \\ 0, & \text{otherwise}. \end{cases}$$

Let us introduce the family of random variables $\xi(i,s)$ such that

$$\mathbf{P}\{\xi(i,s) \in C\} = \mathbf{P}\left\{S_n(t_{n2}) - S_n(t_{n1}) \in C \mid (Q_n(t_{n1}), S_n(t_{n1})) = (i, ns)\right\}.$$

Then the variable $\xi(i,s)$ can be represented in the form: for $i < m$,

$$\xi(i,s) = \begin{cases} 1, & \text{with probability } \lambda_i(s)^{-1} q_i(s), \\ -1, & \text{with probability } \lambda_i(s)^{-1} s\nu(s), \\ 0, & \text{otherwise}, \end{cases}$$

$$\xi(m,s) = \begin{cases} 1, & \text{with probability } \lambda_m(s)^{-1} q_m(s), \\ -1, & \text{with probability } \lambda_m(s)^{-1} g(s)s\nu(s), \\ 0, & \text{otherwise}. \end{cases}$$

Now, we can describe the process $(Q_n(t), S_n(t))$ as an SP in the same way as in the Appendix, Section A.2. In our case, the variable $\tau_{nk}(i,s)$ has exponential distribution with parameter $\lambda_i(s)$ and the variable $\xi_{nk}(i,s)$ has the same distribution as the variable $\xi(i,s)$ introduced above. Further, denote at each fixed $s \geq 0$ by $\tilde{x}_k(s)$, $k \geq 0$ a Markov process with transition probabilities

$$p_{ij}(s) = \begin{cases} \lambda_i(s)^{-1} i\mu, & \text{if } j = i-1, \\ \lambda_i(s)^{-1}(r_i(s) + q_i(s))\lambda, & \text{if } j = i, \\ \lambda_i(s)^{-1}(p_i(s)\lambda + s\nu(s)), & \text{if } j = i+1, \\ = 0, & \text{otherwise, } i = \overline{0, m-1}, \end{cases}$$

$$p_{mj}(s) = \begin{cases} \lambda_m(s)^{-1} m\mu, & \text{if } j = m-1, \\ \lambda_m(s)^{-1}(\lambda + s\nu(s)), & \text{if } j = m, \\ = 0, & \text{otherwise.} \end{cases}$$

At any $s \geq 0$, the state space forms one essential class. Denote by $\{\pi(i,s), i = \overline{0,m}\}$ a stationary distribution of $\tilde{x}_k(s)$, $k \geq 0$. It is easy to see that in each bounded region $\{0 \leq s \leq L\}$, the process $\tilde{x}_k(s)$ is uniformly ergodic.

Let us introduce functions

$$\begin{gathered} m(s) = \sum_{i=0}^{m} \pi(i,s)\lambda_i(s)^{-1}, \qquad \hat{q}(s) = \sum_{i=0}^{m} \pi(i,s) q_i(s)\lambda_i(s)^{-1}, \\ b(s) = \lambda\hat{q}(s) - s\nu(s)m(s) + s\nu(s)(1 - g(s))\pi(m,s)\lambda_m(s)^{-1}. \end{gathered} \tag{3.20}$$

Denote $\tilde{b}(s) = m(s)^{-1}b(s)$. Then, according to Theorem A.3, the relation (3.18) holds where $s(t)$ is a solution of equation (A.19). Further, as the function $\nu(s)$ is bounded for some $c_0 > 0$,

$$\liminf_{s\to\infty} s m(s) > c_0.$$

This relation implies that

$$\int_0^\infty m(\eta(u))\, du = +\infty,$$

and convergence in (3.18) takes place for any $T > 0$.

Now, let us calculate the function $\tilde{b}(s)$ in an explicit form. We mention that values $m(s)^{-1}\pi(i,s) \times\lambda(i,s)^{-1}$, $i = \overline{0,m}$, at each fixed s are stationary probabilities for the Markov process with continuous time $\tilde{x}(t,s)$, $t \geq 0$, given by intensities $\tilde{a}_{ij}(s) = \lambda(i,s)p_{ij}(s)$ (here we allow transitions back to the same state). But the process $\tilde{x}(t,s)$, $t \geq 0$, is equivalent to the birth and death process $x(t,s)$ with intensities of birth $c_i(s)$ and death $d_i(s)$ in the state i, respectively, $c_i(s) = p_i\lambda + s\nu(s)$, $i < m$, $d_i(s) = i\mu$, $i \leq m$. Therefore, stationary probabilities of the process $\tilde{x}(t,s)$ are given by expression (3.16) and after simple calculations, we get that $\tilde{b}(s) = \hat{b}(s)$, which finally proves Theorem 3.3.

Let us study the cases when equation (3.19) has stability points.

CASE 1. Suppose that

$$\inf_{s\geq 0} g(s) = g_0 > 0, \quad \inf_{s\geq 0} \nu(s) = \nu_0 > 0, \quad \sum_{i=0}^{m} q_i(0) > 0. \tag{3.21}$$

This means that there always is a flow of lost calls in the state m (when all servers are busy). Then, $\hat{b}(0) > 0$ and $\hat{b}(s) \leq \lambda - s\nu_0 g_0 \to -\infty$ as $s \to \infty$. Denote by s_* the minimal root of the equation

$$\hat{b}(s) = 0, \tag{3.22}$$

(it exists according to continuity). In some small neighborhood of s_* $\hat{b}(s) > 0$ at $s < s_*$ and $\hat{b}(s) < 0$ at $s > s_*$. This means that the point s_* is a point of stability for solutions with initial value s_0 in some neighborhood of s_*.

We mention that in this case, a stable solution exists for any values of λ and μ. This fact can be explained in the following way: if s is large, then the flow of lost calls also has a large intensity no less than $s\nu_0 g_0$.

CASE 2. Suppose that $g(s) \equiv 0$, $q_m(s) \equiv 1$, $\sum_{i=0}^{m} q_i(0) > 0$, $s\nu(s) \to \infty$ as $s \to \infty$ and

$$\lambda < m\mu. \tag{3.23}$$

This means that if a call finds all servers busy, it goes with probability one into the orbit and there is no flow of lost calls in the state m. It is not difficult to calculate that

$$\lim_{s\to\infty} \rho(m,s) = 1, \qquad \lim_{s\to\infty} \hat{b}(s) = \lambda - m\mu. \tag{3.24}$$

As $\hat{b}(0) > 0$, relations (3.23),(3.24) imply that the minimal root of the equation (3.22) exists and it is the point of stability.

In particular, if $m = 1$, functions $p_i(\cdot), q_i(\cdot), \nu_i(\cdot)$ do not depend on s, and $g(s) \equiv 0$, then

$$\hat{b}(s) = \frac{\lambda^2 + s\nu(\lambda - \mu)}{\lambda + \mu + s\nu},$$

which corresponds to expressions in (3.12) for the case $q = 1$, $p = 0$, $m = \lambda^{-1}$.

These results show that a technique based on limit theorems of averaging principle and diffusion approximation for switching processes gives us a new effective approach for studying transient and stable regimes of operating for rather complex retrial queueing systems in overloading conditions.

We mention that some asymptotic results for similar systems on the base of another technique were obtained in [8].

APPENDIX

A.1. Averaging Principle and Diffusion Approximation for SP

Here we give limit theorems for SP in the case of fast switches. We consider the process on the interval $[0, nT]$ and suppose that its characteristics depend on the parameter of series n in such a way that the number of switches tends, in probability, to infinity. Then, under some natural assumptions, the normed trajectory of SP uniformly converges, in probability, to some function which is the solution of a nonstochastic differential equation, and the normalized deviation between trajectory and this solution weakly converges in Skorokhod space $D_{[0,T]}$ to some diffusion process (see [25,26,36–38]).

Let us consider a simple case of SP (characteristics do not depend on the parameter x, that means there is no discrete component).

At each $n > 0$, let $F_{nk} = \{(\zeta_{nk}(t,\alpha), \tau_{nk}(\alpha)),\ t > 0,\ \alpha \in R^m\}$, $k \geq 0$, be jointly independent families of random functions, S_{no} be an initial value. We put $\xi_{nk}(\alpha) = \zeta_{nk}(\tau_{nk}(\alpha), \alpha)$ and construct SP $\zeta_n(t)$ according to formulae (2.5),(2.6),

$$\begin{aligned} &t_{n0} = 0, \quad t_{nk+1} = t_{nk} + \tau_{nk}(S_{nk}), \quad S_{nk+1} = S_{nk} + \xi_{nk}(S_{nk}), \qquad k \geq 0, \\ &\zeta_n(t) = S_{nk} + \zeta_{nk}(t - t_{nk}, S_{nk}), \quad \text{as } t_{nk} \leq t < t_{nk+1}, \quad t \geq 0. \end{aligned} \tag{A.1}$$

Denote

$$g_{nk}(\alpha) = \sup_{t<\tau_{nk}(\alpha)} |\zeta_{nk}(t,\alpha)|.$$

Suppose that the distributions of variables $(\zeta_{nk}(\cdot), \tau_{nk}(\cdot))$ do not depend on index k. Assume that functions $m_n(\alpha) = \mathbf{E}\tau_{n1}(n\alpha)$, $b_n(\alpha) = \mathbf{E}\xi_{n1}(n\alpha)$ exist.

THEOREM A.1. AVERAGING PRINCIPLE. *Suppose that the following conditions hold.*

(1) *For any* $N > 0$,

$$\lim_{L\to\infty} \limsup_{n\to\infty} \sup_{|\alpha|<N} \{\mathbf{E}\tau_{n1}(n\alpha)\chi(\tau_{n1}(n\alpha) > L) + \mathbf{E}|\xi_{n1}(n\alpha)|\chi(|\xi_{n1}(n\alpha)| > L)\} = 0. \tag{A.2}$$

(2) *As* $\max(|\alpha_1|, |\alpha_2|) < N$,

$$|m_n(\alpha_1) - m_n(\alpha_2)| + |b_n(\alpha_1) - b_n(\alpha_2)| < C_N|\alpha_1 - \alpha_2| + \alpha_n(N), \tag{A.3}$$

where C_N *are bounded constants,* $\alpha_n(N) \to 0$ *uniformly on* $|\alpha_1| < N$, $|\alpha_2| < N$, $n \to \infty$.

(3) *There exist functions* $m(a) > 0$ *and* $b(\alpha)$ *such that for any* $\alpha \in R^r$ *as* $n \to \infty$,

$$m_n(\alpha) \to m(\alpha), \qquad b_n(\alpha) \to b(\alpha). \tag{A.4}$$

(4) *For any* $\varepsilon > 0$, $N > 0$,

$$\lim_{n\to\infty} n \sup_{|\alpha|<N} \mathbf{P}\left\{\frac{1}{n} g_{n1}(n\alpha) > \varepsilon\right\} = 0. \tag{A.5}$$

(5) *Moreover,*

$$n^{-1} S_{n0} \xrightarrow{P} s_0. \tag{A.6}$$

Then,

$$\sup_{0\le t\le T} \left|n^{-1}\zeta_n(nt) - s(t)\right| \xrightarrow{P} 0, \tag{A.7}$$

where

$$s(0) = s_0, \qquad \mathrm{d}s(t) = m(s(t))^{-1} b(s(t))\,\mathrm{d}t, \tag{A.8}$$

and T *is any positive number such that* $y(+\infty) > T$ *with probability 1 where*

$$\begin{aligned} y(t) &= \int_0^t m(\eta(u))\,\mathrm{d}u, \\ \eta(0) &= s_0, \quad \mathrm{d}\eta(u) = b(\eta(u))\,\mathrm{d}u, \end{aligned} \tag{A.9}$$

(it is assumed that a solution of equation (A.8) exists on each interval and is unique).

Now we consider the convergence of the process

$$\gamma_n(t) = \frac{1}{\sqrt{n}}(\zeta_n(nt) - ns(t)), \qquad t \in [0, T]$$

to some diffusion process. Denote

$$\begin{aligned} \tilde{b}_n(\alpha) &= m_n(\alpha)^{-1} b_n(\alpha), \qquad \tilde{b}(\alpha) = m(\alpha)^{-1} b(\alpha), \\ \rho_n(\alpha) &= \xi_{n1}(n\alpha) - b_n(\alpha) - \tilde{b}(\alpha)(\tau_{n1}(n\alpha) - m_n(\alpha)), \\ D_n^2(\alpha) &= \mathbf{E}\rho_n(\alpha)\rho_n(\alpha)^*, \end{aligned}$$

(we denote the conjugate vector by the symbol $*$),

$$q_n(\alpha, z) = \sqrt{n}\left(\tilde{b}_n\left(\alpha + \frac{1}{\sqrt{n}} z\right) - \tilde{b}(\alpha)\right).$$

THEOREM A.2. DIFFUSION APPROXIMATION. *Let conditions (A.3)–(A.6) be satisfied where in (A.3),* $\sqrt{n}\alpha_n(N) \to 0$, *and in addition, the following conditions hold.*

(1) *There exist continuous, matrix-valued functions* $D^2(\alpha)$ *and* $Q(\alpha)$ *and vector-valued function* $g(a)$ *such that uniformly in each bounded region* $|\alpha| \le N$, *as* $n \to \infty$,

$$D_n^2(\alpha) \to D^2(\alpha), \tag{A.10}$$

and

$$q_n(\alpha, z) \to Q(\alpha)z + g(\alpha), \qquad \textit{at any } z \in R^r. \tag{A.11}$$

(2) *For any* $N > 0$,

$$\lim_{L\to\infty} \limsup_{n\to\infty} \sup_{|\alpha|<Nn} \left\{ \mathbf{E}\tau_{n1}^2(\alpha)\chi(\tau_{n1}(\alpha) > L) + \mathbf{E}|\xi_{n1}(\alpha)|^2\chi(|\xi_{n1}(a))| > L) \right\} = 0. \tag{A.12}$$

(3) *For any* $\varepsilon > 0$, $N > 0$,

$$\lim_{n\to\infty} n \sup_{|\alpha|<N} \mathbf{P}\left\{ \frac{1}{\sqrt{n}} g_{n1}(n\alpha) > \varepsilon \right\} = 0. \tag{A.13}$$

(4) *Moreover,*

$$\gamma_n(0) \overset{w}{\Rightarrow} \gamma_0. \tag{A.14}$$

Then, the sequence of processes $\gamma_n(t)$ *J-converges on any interval* $[0, T]$, *such that* $y(+\infty) > T$, *to the diffusion process* $\gamma(t)$ *which satisfies the following stochastic differential equation:*

$$\mathrm{d}\gamma(t) = \{Q(s(t))\gamma(t) + g(s(t))\}\,\mathrm{d}t + D(s(t))m(s(t))^{-1/2}\,\mathrm{d}w(t), \quad \gamma(0) = \gamma_0, \tag{A.15}$$

where $s(\cdot)$ *satisfies equation (A.8) (J-convergence denotes the weak convergence of measures in Skorokhod space* $D_{[0,T]}$*) and* $w(t)$ *is the standard Wiener process in* R^r.

The proof of these two theorems can be found in [26].

A.2. Averaging Principle for General RPSM

Now, we give an averaging principle for a general RPSM. For any $n > 0$, let there be given jointly independent families of random vectors

$$F_{nk} = \{(\xi_{nk}(x,\alpha), \tau_{nk}(x,\alpha), \beta_{nk}(x,\alpha)), x \in X, \alpha \in R^r\}, \qquad k \ge 0,$$

with values in the space $R^r \times [0,\infty) \times X$, where X is some measurable space. Also, let (x_{n0}, S_{n0}) be an initial value independent of F_{nk}, $k \ge 0$. We put

$$t_{n0} = 0, \qquad t_{nk+1} = t_{nk} + \tau_{nk}(x_{nk}, S_{nk}),$$
$$S_{nk+1} = S_{nk} + \xi_{nk}(x_{nk}, S_{nk}), \quad x_{nk+1} = \beta_{nk}(x_{nk}, S_{nk}), \qquad k \ge 0, \tag{A.16}$$
$$S_n(t) = S_{nk}, \quad x_n(t) = x_{nk}, \qquad \text{as } t_{nk} \le t < t_{nk+1}, \quad t \ge 0. \tag{A.17}$$

Then the pair $(x_n(t), S_n(t))$, $t \ge 0$ forms a general RPSM with feedback between both components. Suppose for simplicity that distributions of families F_{nk} do not depend on the index $k \ge 0$ and let the moment functions exist

$$m_n(x,\alpha) = \mathbf{E}\tau_{n1}(x, n\alpha), \qquad b_n(x,\alpha) = \mathbf{E}\xi_{n1}(x, n\alpha).$$

Denote $p_n(x, A, \alpha) = \mathbf{P}\{\beta_{n1}(x,\alpha) \in A\}$, $x \in X$, $A \in B_X$, $\alpha \in R^r$ and let, for any fixed α $\tilde{x}_{nk}(\alpha)$, $k \ge 0$, be an MP in X with transition probabilities

$$\mathbf{P}\{\tilde{x}_{nk+1}(\alpha) \in A \mid \tilde{x}_{nk}(\alpha) = x\} = p_n(x, A, \alpha).$$

Suppose that there exists a family of transition probabilities $q(x, A, \alpha)$, $x \in X$, $A \in B_X$, $\alpha \in R^r$ uniformly continuous in α in each bounded region $|\alpha| \le L$ uniformly in $x \in X$, $A \in B_X$ and for any $L > 0$,

$$\sup_{x,A,|\alpha|\le L} |p_n(x, A, \alpha) - q(x, A, \alpha)| \to 0. \tag{A.18}$$

Further, suppose that MP $\tilde{x}_{nk}(\alpha)$, $k \geq 0$, is uniformly ergodic with stationary measure $\pi(A, \alpha)$ uniformly in α in each bounded region and in $n > 0$. Denote

$$m_n(\alpha) = \int_X m_n(x, \alpha)\pi_n(\mathrm{d}x, \alpha), \qquad b_n(\alpha) = \int_X b_n(x, \alpha)\pi_n(\mathrm{d}x).$$

THEOREM A.3. *Suppose that (A.18) holds.*

(1)

$$\lim_{L\to\infty} \limsup_{n\to\infty} \sup_{|\alpha|<N} \sup_x \{\mathbf{E}\tau_{n1}(x, n\alpha)\chi(\tau_{n1}(x, n\alpha) > L) + \mathbf{E}|\xi_{n1}(x, n\alpha)|\chi(|\xi_{n1}(x, n\alpha)| \leq L)\} = 0.$$

(2) *For any x as* $\max(|\alpha_1|, |\alpha_2|) < N$,

$$|m_n(x, \alpha_1) - m_n(x, \alpha_2)| + |b_n(x, \alpha_1) - b_n(x, \alpha_2)| < C_N|\alpha_1 - \alpha_2| + \alpha_n(N),$$

where C_N are some constants, $\alpha_n(N) \to 0$ uniformly in $|\alpha_1| < N$, $|\alpha_2| < N$, $n \to \infty$.

(3) *There exist functions $b(\alpha), m(\alpha) > 0$ and possibly random variable s_0 such that as $n \to \infty$,*

$$b_n(\alpha) \to b(\alpha), \quad m_n(\alpha) \to m(\alpha), \quad \alpha \in R^r, \quad \text{and} \quad n^{-1}S_{n0} \xrightarrow{P} s_0.$$

Then,

$$\sup_{0\leq t\leq T} \left|n^{-1}S_n(nt) - s(t)\right| \xrightarrow{P} 0,$$

where

$$s(0) = s_0, \qquad \mathrm{d}s(t) = m(s(t))^{-1}b(s(t))\,\mathrm{d}t, \tag{A.19}$$

and T is any positive number such that $y(+\infty) > T$ with probability 1 where

$$y(t) = \int_0^t m(\eta(u))\,\mathrm{d}u, \tag{A.20}$$
$$\eta(0) = s_0, \quad \mathrm{d}\eta(u) = b(\eta(u))\,\mathrm{d}u.$$

The proof of this result follows from the averaging principle for general switching recurrent sequences and switching processes (see [37,38]).

REFERENCES

1. T. Yang and J.G.C. Templeton, A survey on retrial queues, *Queueing Systems* **2**, 203–233 (1987).
2. G.I. Falin, A survey of retrial queues, *Queueing Systems* **7**, 127–168 (1990).
3. V.G. Kulkarni and H.M. Liang, Retrial queues revisited, In *Frontiers in Queueing Models and Applications in Science and Engineering*, (Edited by J.H. Dshalalow), pp. 19–34, CRC Press, (1997).
4. M.F. Neuts and M.F. Ramalhoto, A service model in which the server is required to search for customers, *Journal of Applied Probability* **21**, 157–166 (1984).
5. G.I. Falin, J.R. Artalejo and M. Martin, On the single server retrial queue with priority customers, *Queueing Systems* **14**, 439–455 (1993).
6. J.R. Artalejo, A queueing system with returning customers and waiting line, *Operations Research Letters* **17**, 191–199 (1995).
7. G.I. Falin, Estimation of retrial rate in a retrial queue, *Queueing Systems* **19**, 231–246 (1995).
8. G.I. Falin and J.R. Artalejo, Approximation for multiserver queues with balking/retrial discipline, *OR Spektrum* **17**, 239–244 (1995).
9. M. Martin and J.R. Artalejo, Analysis of an $M/G/1$ queue with two types of impatient units, *Advances in Applied Probability* **27**, 840–861 (1995).
10. J.R. Artalejo and G.I. Falin, On the orbit characteristics of the $M/G/1$ retrial queue, *Naval Research Logistics* **43**, 1147–1161 (1996).
11. J.R. Artalejo, Analysis of an $M/G/1$ queue with constant repeated attempts and server vacations, *Computer and Operations Research* **24**, 493–504 (1997).
12. V.V. Anisimov and Kh.L. Atadzhanov, Asymptotic analysis of highly-reliable systems with repeated call (in Russian), *Issledov. Oper. i ASU* **37**, 32–36 (1991).

13. V.V. Anisimov and Kh.L. Atadzhanov, Diffusion approximation of systems with repeated calls, *Theory of Probab. and Math. Statistics* **44**, 3–8 (1991).
14. V.V. Anisimov and Kh.L. Atadzhanov, Diffusion approximation of systems with repeated calls and unreliable server, *Journal of Mathematical Sciences* **72**, 3032–3034 (1994).
15. G.I. Falin and J.G.C. Templeton, *Retrial Queues*, Chapman and Hall, (1997).
16. N.P. Buslenko, V.V. Kalashnikov and I.N. Kovalenko, *Lectures on the Theory of Complex Systems*, (in Russian), Sov. Radio, Moscow, (1973).
17. G.P. Basharin, P.P. Bocharov and Ja.A. Kogan, *Analysis of Queues in Computing Networks*, (in Russian), Nauka, Moscow, (1989).
18. V.V. Anisimov, *Random Processes with Discrete Component, Limit Theorems*, (in Russian), Kiev University, (1988).
19. V.V. Anisimov and E.A. Lebedev, *Stochastic Queueing Networks. Markov Models*, (in Russian), Kiev University, (1992).
20. H. Chen and A. Mandelbaum, Hierarchical modelling of stochastic networks, Part I, II, In *Stochastic Modeling and Analysis of Manufacturing Systems*, (Edited by D.D. Yao), pp. 47–105; 107–131, Springer-Verlag, (1994).
21. J.M. Harrison, Balanced fluid models of multiclass queueing network: A heavy traffic conjecture, In *Stochastic Networks*, IMA Volumes in Mathematics and its Appl., Volume 71, pp. 1–20, (1995).
22. J.M. Harrison and R.J. Williams, A multiclass closed queueing network with unconventional heavy traffic behavior, *Ann. Appl. Probab.* **6** (1), 1–47 (1996).
23. V.V. Anisimov, Switching processes, *Cybernetics* **13** (4), 590–595 (1977).
24. V.V. Anisimov, Limit theorems for switching processes and their applications, *Cybernetics* **14** (6), 917–929 (1978).
25. V.V. Anisimov, Limit theorems for switching processes, In *Functional Analysis III*, Proc. Postgrad. School and Conf., Dubrovnik, Yugoslavia, (1989); *Var. Publ. Ser., Aarhus Univ.* (40), 235–262 (1992).
26. V.V. Anisimov, Switching processes: Averaging principle, diffusion approximation and applications, *Acta Applicandae Mathematicae*, Kluwer, The Netherlands **40**, 95–141 (1995).
27. V.V. Anisimov, Asymptotic analysis of switching queueing systems in conditions of low and heavy loading, In *Matrix-Analytic Methods in Stochastic Models*, Lecture Notes in Pure and Appl. Mathem. Series, Volume 183, pp. 241–260, Marcel Dekker, (1996).
28. I.I. Ežov and A.V. Skorokhod, Markov processes which are homogeneous in the second component, *Theor. Probab. Appl.* **14**, 679–692 (1969).
29. V.V. Anisimov, Asymptotic consolidation of the states of random processes, *Cybernetics* **9** (3), 494–504 (1973).
30. I.I. Gikhman and A.V. Skorokhod, *Theory of Random Processes II*, (in Russian), Nauka, Moscow, (1973).
31. R. Hersh, Random evolutions: Survey of results and problems, *Rocky Mount. J. Math.* **4** (3), 443–475 (1974).
32. T. Kurtz, A limit theorem for perturbed operator semigroups with applications to random evolutions, *J. Funct. Anal.* **12**, 55–67 (1973).
33. M. Pinsky, *Random Evolutions*, Lecture Notes in Math., Volume 451, pp. 89–100, Springer-Verlag, New York, (1975).
34. V.S. Korolyuk and A.V. Swishchuk, *Random Evolutions*, Kluwer Academic, (1994).
35. V.V. Anisimov, O.K. Zakusilo and V.S. Dontchenko, *The Elements of Queueing Theory and Asymptotic Analysis of Systems* (in Russian), p. 248, Visca Scola, Kiev, (1987).
36. V.V. Anisimov and A.O. Aliev, Limit theorems for recurrent processes of semi-Markov type, *Theor. Probab. and Math. Statist.* **41**, 7–13 (1990).
37. V.V. Anisimov, Averaging principle for switching recurrent sequences, *Theor. Probab. and Math. Statist.* **45**, 1–8 (1991).
38. V.V. Anisimov, Averaging principle for switching processes, *Theory Probab. and Math. Statist.* **46**, 1–10 (1992).