Queueing Systems 39, 157–182, 2001

Analysis of Markov Multiserver Retrial Queues with Negative Arrivals *

VLADIMIR V. ANISIMOV vlanis@bilkent.edu.tr
Department of Industrial Engineering, Bilkent University, Bilkent 06533, Ankara, Turkey and Applied Statistics Department, Kiev University, Kiev-17, 252017, Ukraine

JESUS R. ARTALEJO jesus_artalejo@mat.ucm.es
Department of Statistics and Operations Research, Complutense University of Madrid, Madrid 28040, Spain

Received 25 December 2000; Revised 11 June 2001

Abstract. Negative arrivals are used as a control mechanism in many telecommunication and computer networks. In the paper we analyze multiserver retrial queues; i.e., any customer finding all servers busy upon arrival must leave the service area and re-apply for service after some random time. The control mechanism is such that, whenever the service facility is full occupied, an exponential timer is activated. If the timer expires and the service facility remains full, then a random batch of customers, which are stored at the retrial pool, are automatically removed. This model extends the existing literature, which only deals with a single server case and individual removals. Two different approaches are considered. For the stable case, the matrix–analytic formalism is used to study the joint distribution of the service facility and the retrial pool. The approximation by more simple infinite retrial model is also proved. In the overloading case we study the transient behaviour of the trajectory of the suitably normalized retrial queue and the long-run behaviour of the number of busy servers. The method of investigation in this case is based on the averaging principle for switching processes.

Keywords: retrial queueing systems, negative arrivals, averaging principle, matrix–analytic methods, switching process

AMS subject classification: 60K25, 60J27

1. Introduction

Retrial queues are characterized by the feature that arriving customers who find all servers busy join the retrial group to try their luck again some time later. Queues in which customers are allowed to conduct retrials have been widely used to model many problems in telephone switching systems, computer, and communication systems. A complete description of situations where queues with retrial customers arise can be found in [19]. A classified bibliography is given in [8]. On the other hand, during the last

* This research was supported in part by DGES98-0837 and the European Commission through INTAS Project 96-0828.

decade there has been an increasing interest in queueing systems and networks with negative arrivals and their applications. In its simplest version, a negative arrival removes an ordinary positive customer according to some strategy. Extensions of this concept result when a negative arrival removes a random batch of customers, all the work from the queue or a random amount of work. For a comprehensive analysis of queueing networks with negative arrivals the reader is referred to the monographs [14,20]. A recent review of this topic can be found in [9]. We also mention a new approach [26] where the negative arrivals appear under the terminology string transition which consists in a string of instantaneous substractions or additions of units at the nodes.

A number of recent papers [10,12,13] deal with the queueing modelling of systems operating under the simultaneous presence of negative arrivals and repeated attempts. It should be noted that the existence of a flow of negative arrivals provides a mechanism to control an excessive congestion at the retrial group. Applications of retrial queues with negative arrivals are connected with the design and control of packet switching networks [10]. Previous papers consider single-server queues with individual removals operating in a stable regime. In this sense, the contribution of the paper is to extend the analysis to the multiserver case allowing batch removals in both transient and stable regimes. To this end, we employ the matrix–analytic formalism [23,24] and asymptotic methods based on the averaging principle for so called switching processes [1–4]. The consideration of both methodologies gives us an effective approach for studying transient and stable operating regimes of our complex retrial queue.

As related work we have to mention [4–7,15–17,21]. In [21] an approximation for time-dependent analysis of a multiserver retrial queue is proposed. The accuracy of the approximation is evaluated by making comparisons with a simulation. Papers [15–17] contain algorithmic methods for multiserver retrial queues with interarrival, interrepetition times and/or service times of the type *PH*, *MAP*, *SM*, etc. Applications of limit theorems for switching processes to overloaded Markov and semi-Markov type queueing models are studied in [4,5], applications to some classes of retrial models are considered in [6,7].

The rest of the paper is organized as follows. In section 2 we describe the mathematical model. The matrix–analytic approach for the study of the model at the stationary regime is described in section 3. In section 4 we investigate how the stationary distribution can be approximated with the help of more simple infinite approximate model with constant (after some level) retrial rate. In section 5 we prove the averaging principle in overloading case and transient conditions for the trajectory of a retrial queue and the long-run behaviour of the number of busy servers. Some general results about the asymptotic analysis of switching processes based on the averaging principle are summarized in the appendix.

2. The mathematical model

An initial model description at Markovian level is as follows. Customers arrive to a multiserver system according to a Poisson process with rate λ. The service facility consists of c identical servers, so an arriving customer who finds all servers busy is blocked and leaves temporary the service area. Such customers join a group of unsatisfied customers called orbit. We assume that the access from the retrial group to the service facility is governed by the linear retrial policy described in [11]; i.e., the probability of a repeated attempt during the interval $(t, t + \Delta t)$, given that j customers were in orbit at time t, is $(\alpha(1 - \delta_{j0}) + j\mu)\Delta t + o(\Delta t)$, where δ_{j0} is Kronecker's symbol. This linear retrial discipline provides a rule which incorporates simultaneously the classical and the constant retrial policies extensively studied in the literature. The service times have exponential distribution with rate ν both for primary customers and successful repeated attempts.

In addition to the regular customers, a second flow of negative arrivals following a Poisson process with rate δ is also considered. A negative arrival has the effect of removing a random batch of customers from the retrial group. Let p_k be the probability of delating k customers when a negative arrival occurs. We also denote $\delta_k = \delta p_k$ and $\delta_k^* = \sum_{i=k}^{\infty} \delta_i$, for $k \geqslant 1$. It should be pointed out that the introduction of a flow of negative arrivals provides a mechanism to control the congestion of the system. If at least one server is free, then any primary or orbiting customer may immediately join a server; so an excessive level of congestion at the retrial group is mainly caused by those customers arriving when the c servers are busy. Thus, we assume that negative customers only act when all servers are busy. In other words, the negative arrival can be viewed as a timer which is switched when the service facility is saturated and has effect if this state remains when the timer expires. Finally, we also assume that the input flows of positive and negative arrivals, intervals between repeated attempts and service times are mutually independent.

The system state can be described by the bivariate process $\{X(t);\ t \geqslant 0\} = \{(C(t), N(t));\ t \geqslant 0\}$, where $C(t)$ is the number of busy servers and $N(t)$ is the number of customers in the retrial group at time t. Note that the process $X(t)$ takes values in the semi-strip $S = \{0, \ldots, c\} \times \mathrm{N}$. The infinitesimal generator, $Q = (q_{ab})$, of the process $X(t)$ is as follows:

$$q_{(i,j),(m,n)} = \begin{cases} \lambda, & \text{if } (m, n) = (i + 1, j), \\ i\nu, & \text{if } (m, n) = (i - 1, j), \\ \alpha(1 - \delta_{j0}) + j\mu, & \text{if } (m, n) = (i + 1, j - 1), \\ -(\lambda + i\nu + \alpha(1 - \delta_{j0}) + j\mu), & \text{if } (m, n) = (i, j), \\ 0, & \text{otherwise,} \end{cases}$$

$$q_{(c,j),(m,n)} = \begin{cases} \lambda, & \text{if } (m, n) = (c, j + 1), \\ c\nu, & \text{if } (m, n) = (c - 1, j), \\ \delta_k, & \text{if } (m, n) = (c, j - k),\ 1 \leqslant k \leqslant j - 1, \\ \delta_j^*, & \text{if } (m, n) = (c, 0), \\ -(\lambda + c\nu + \delta), & \text{if } (m, n) = (c, j), \\ 0, & \text{otherwise.} \end{cases}$$

In contrast to the classical theory for random walks on a lattice semi-strip, the main feature of the generator described above is space-heterogeneity with respect to the second coordinate which is caused by the transitions $(i, j) \to (i+1, j-1)$. Note also that the evolution of our queueing model exhibits an alternating sequence of idle and busy periods of the servers. The distribution of the idle periods varies along the time as a consequence of the non-homogeneity introduced by the repeated attempts. This explains why the investigation of multiserver retrial queues is essentially more difficult that the one of single server models. In fact, explicit results and/or recursive exact methods are available only in a few particular cases [19], in particular when $c \leqslant 2$. Thus, in section 3 we assume the stable regime, and concentrate our efforts on the homogeneous case $\mu = 0$ and also on the operating approximation $\alpha(1-\delta_{j0}) + \min(j, M)\mu$ for the retrial rate (see [19,25]). In these cases, the analysis can be based on the matrix theory developed by Neuts [24]. The matrix–analytic methodology is nowadays a well-known technique among queueing specialists, so we omit unnecessary routines and just concentrate on that aspects that provide significant insight for our specific model with retrials and negative arrivals. In section 4 we prove the convergence of the approximated stationary distribution to the stationary distribution of the initial system with linear retrial rate.

The complexity of our queueing model leads to the necessity of study asymptotic results. Thus, in section 5, we investigate the asymptotic behaviour of the retrial group by employing the averaging principle for so-called switching processes. This versatile class of processes was introduced in [1]. A switching process is described as a bivariate process $\{(x(t), \zeta(t));\ t \geqslant 0\}$, with the property that there exists an increasing sequence of epochs t_k such that on each interval $[t_k, t_{k+1})$ we have $x(t) = x(t_k)$ and the behaviour of the second component $\zeta(t)$ depends on the value $(x(t_k), \zeta(t_k))$ only. The epochs t_k are called switching times and $x(t)$ is the discrete switching component. It should be noted that the process $X(t) = (C(t), N(t)),\ t \geqslant 0$, matches this definition when we choose t_k as the transition times in the queueing system.

3. Analysis of the system state in the stable case

In this section we study the process $X(t)$ operating under the stable regime. The analysis is based on the matrix–analytic theory for Markov models of $GI/M/1$ type [24]. Let us briefly describe the general formulation and concentrate on specific aspects such as the determination of stability abscissas and the effective computation of the matrix R. First, we consider the constant retrial policy; i.e., $\mu = 0$. Then, the generator Q can be re-expressed in the form

$$Q = \begin{pmatrix} B_0 & A_0 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots \\ B_1 & A_1 & A_0 & \mathbf{0} & \mathbf{0} & \dots \\ B_2 & A_2 & A_1 & A_0 & \mathbf{0} & \dots \\ B_3 & A_3 & A_2 & A_1 & A_0 & \dots \\ & \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix},$$

where all blocks are square matrices of order $c+1$. The level j in the matrix formalism corresponds to the subset of states $\{(i, j);\ 0 \leqslant i \leqslant c\}$. The elements of matrices A_k and B_k are as follows

$$(A_0)_{ij} = \begin{cases} \lambda, & \text{if } i = j = c, \\ 0, & \text{otherwise,} \end{cases}$$

$$(A_1)_{ij} = \begin{cases} \lambda, & \text{if } 0 \leqslant i \leqslant c-1,\ j = i+1, \\ i\nu, & \text{if } 1 \leqslant i \leqslant c,\ j = i-1, \\ -(\lambda + i\nu + (1-\delta_{ic})\alpha + \delta_{ic}\delta), & \text{if } 0 \leqslant i \leqslant c,\ j = i, \\ 0, & \text{otherwise,} \end{cases}$$

$$(A_2)_{ij} = \begin{cases} \alpha, & \text{if } 0 \leqslant i \leqslant c-1,\ j = i+1, \\ \delta_1, & \text{if } i = j = c, \\ 0, & \text{otherwise,} \end{cases}$$

$$(A_k)_{ij} = \begin{cases} \delta_{k-1}, & \text{if } i = j = c, \\ 0, & \text{otherwise,} \end{cases} \quad \text{for } k = 3, 4, \ldots,$$

$$(B_0)_{ij} = \begin{cases} \lambda, & \text{if } 0 \leqslant i \leqslant c-1,\ j = i+1, \\ i\nu, & \text{if } 1 \leqslant i \leqslant c,\ j = i-1, \\ -(\lambda + i\nu), & \text{if } 0 \leqslant i \leqslant c,\ j = i, \\ 0, & \text{otherwise,} \end{cases}$$

$$(B_1)_{ij} = \begin{cases} \alpha, & \text{if } 0 \leqslant i \leqslant c-1,\ \ j = i+1, \\ \delta_1^*, & \text{if } i = j = c, \\ 0, & \text{otherwise,} \end{cases}$$

$$(B_k)_{ij} = \begin{cases} \delta_k^*, & \text{if } i = j = c, \\ 0, & \text{otherwise,} \end{cases} \quad \text{for } k = 2, 3, \ldots.$$

Let us define the matrix $A = \sum_{k=0}^{\infty} A_k$. Then, we have

$$(A)_{ij} = \begin{cases} \lambda + \alpha, & \text{if } 0 \leqslant i \leqslant c-1,\ j = i+1, \\ i\nu, & \text{if } 1 \leqslant i \leqslant c,\ j = i-1, \\ -((1-\delta_{ic})(\lambda+\alpha) + i\nu), & \text{if } 0 \leqslant i \leqslant c,\ \ i = j, \\ 0, & \text{otherwise.} \end{cases}$$

Note that A is the generator of the system $M/M/c/c$ with arrival rate $\lambda + \alpha$ and service rate ν. Thus, the stationary probability vector of A is given by

$$\pi = \left(\sum_{k=0}^{c} \frac{1}{k!} \left(\frac{\lambda+\alpha}{\nu} \right)^k \right)^{-1} \left[1, \frac{\lambda+\alpha}{\nu}, \ldots, \frac{1}{c!} \left(\frac{\lambda+\alpha}{\nu} \right)^c \right]. \tag{3.1}$$

Following the general theory we observe that process $X(t)$ is positive recurrent if and only if

$$\pi A_0 \mathbf{e} < \sum_{k=2}^{\infty} (k-1) \pi A_k \mathbf{e},$$

where $\mathbf{e}$ denotes a column vector with all its elements equal to one. After some trivial algebra, the above equation reduces to

$$\lambda \pi_c < \alpha(1-\pi_c) + \delta g \pi_c, \tag{3.2}$$

where $g = \sum_{k=1}^{\infty} k p_k$. Then the stationary probability vector $\mathbf{x} = (\mathbf{x}_0, \mathbf{x}_1, \ldots)$ of Q exists. Components $\mathbf{x}_k$ are row vectors of dimension $c+1$ containing the distribution of level k.

Inequality (3.2) is a closed form expression but involves all system parameters in a nontrivial relationship. Thus, our next objective is to investigate the recurrence positive domain of the process with respect to any parameter, while the rest of system parameters are fixed.

Theorem 3.1. The following positively recurrent conditions hold:

1. Let us assume that λ, ν, δ and c are fixed. Then, as:
   1.1. If $\lambda \leqslant \delta g$ Then the system is positive recurrent (for all $\alpha > 0$).
   1.2. If $\delta g < \lambda < c\nu + \delta g$ Then the system is positive recurrent if and only if $\alpha > \alpha^*(\lambda, \nu, \delta, c) = \nu u^*(\lambda, \nu, \delta, c) - \lambda$, where $u^*(\lambda, \nu, \delta, c)$ is the unique root in $(\lambda/\nu, \infty)$ of the polynomial
$$F(u) = \frac{c\nu + \delta g - \lambda}{c!} u^c + \sum_{k=1}^{c-1} \frac{k\nu - \lambda}{k!} u^k - \lambda. \tag{3.3}$$
2. Let us assume that ν, α, δ and c are fixed. Now the system is positive recurrent if and only if $\lambda < \lambda^*(\nu, \alpha, \delta, c) = \nu u^*(\nu, \alpha, \delta, c) - \alpha$, where $u^*(\nu, \alpha, \delta, c)$ is the unique root in $(\alpha/\nu, \infty)$ of the polynomial
$$G(u) = \alpha \sum_{k=0}^{c-1} \frac{u^k}{k!} + \frac{\alpha + \delta g}{c!} u^c - \frac{\nu}{c!} u^{c+1}.$$
3. Let us assume that λ, ν, α and c are fixed. Then, the system is positive recurrent if and only if $\delta > \delta^*(\lambda, \nu, \alpha, c)$, where
$$\delta^*(\lambda, \nu, \alpha, c) = \left(\frac{\lambda}{c!} \left(\frac{\lambda + \alpha}{\nu} \right)^c - \alpha \sum_{k=0}^{c-1} \frac{1}{k!} \left(\frac{\lambda + \alpha}{\nu} \right)^k \right) \frac{c!}{g} \left(\frac{\nu}{\lambda + \alpha} \right)^c.$$

Proof. First we assume that λ, ν, δ and c are fixed. Intuition suggests that it should be a stability abscissa $\alpha^*(\lambda, \nu, \delta, c)$, such that the positive recurrent condition is fulfilled if $\alpha > \alpha^*(\lambda, \nu, \delta, c)$. To prove this we first observe that the positive recurrent condition (3.2) is trivially satisfied (with independence of α) when $\lambda \leqslant \delta g$. Thus, we now deal with the case $\lambda > \delta g$.

Let us define the auxiliary variable $u = (\lambda + \alpha)/\nu$. Then, the positive recurrent relationship (3.2) reduces, after some elementary algebra, to the polynomial relation $F(u) > 0$, where $F(u)$ is defined in (3.3).

For each $k \in \{0, 1, \ldots\}$ we define $S(k) = \{(i, j) \in S \mid i + j \leqslant k\}$. By equating the flow rate in and out of the subset $S(k)$, we have

$$\lambda = \sum_{i=0}^{c} \sum_{j=0}^{\infty} i\nu x_{ij} + \sum_{j=1}^{\infty} x_{cj} \sum_{k=1}^{j} \delta_k^*, \tag{3.4}$$

where x_{ij} is the stationary distribution of the system state.

The first term on the right side is bounded by $c\nu$. On the other hand, by interchanging the order of summation in the second term of the right hand side and taking into account that $\sum_{j=k}^{\infty} x_{cj} < 1$, we easily find that $\lambda < c\nu + \delta g$. We conclude that a necessary condition for the positive recurrence is $\lambda < c\nu + \delta g$. Descartes rule of signs states that the difference between the number of variations of sign in the sequence of coefficients of a polynomial function and the number of positive roots is a non-negative even integer. Thus, going back to $F(u)$ we now observe that the coefficients of the polynomial $F(u)$ have only one variation of sign. This implies that $F(u)$ has a unique root $u^*(\lambda, \nu, \delta, c)$ in the interval $(\lambda/\nu, \infty)$. Finally, the critical value α is given by $\alpha^*(\lambda, \nu, \delta, c) = \nu u^*(\lambda, \nu, \delta, c) - \lambda$. It proves part 1.

Now we fix α and δ. Similar arguments lead to prove that the positive recurrent condition holds if and only if

$$G(u) = \alpha \sum_{k=0}^{c-1} \frac{u^k}{k!} + \frac{\alpha + \delta g}{c!} u^c - \frac{\nu}{c!} u^{c+1} > 0.$$

The sequence of coefficients of $G(u)$ has only one variation of sign so, the polynomial $G(u)$ has only one root $u^*(\nu, \alpha, \delta, c)$ in $(\alpha/\nu, \infty)$.

Finally, we assume that λ, ν, α and c are fixed. Then, assumption 3 follows trivially from (3.2) after some algebra. □

We now turn our attention to the stationary vector $\mathbf{x}$. From [24, theorem 3.1.1] we can conclude that

$$\begin{aligned}
&\mathbf{x}_0 (I - R)^{-1} \mathbf{e} = 1, \\
&\mathbf{x}_0 \big(B_0 + RB_1 + B^*\big) = 0, \\
&\mathbf{x}_k = \mathbf{x}_0 R^k, \quad k \geqslant 0,
\end{aligned}$$

where

$$(B^*)_{ij} = \begin{cases} \displaystyle\sum_{k=2}^{\infty} \eta^k \delta_k^*, & \text{if } i = j = c, \\ 0, & \text{otherwise.} \end{cases}$$

R is the minimal non-negative solution to the equation $\sum_{k=0}^{\infty} R^k A_k = \mathbf{0}$, and η is its spectral radius. They key point is the computation of R. To this end, note that $R = -\left(A_0 + \sum_{k=2}^{\infty} R^k A_k\right) A_1^{-1}$. Then, the matrix R is given by $\lim_{n\to\infty} R_n$, where

$$R_0 = \mathbf{0}, \qquad R_n = -\left(A_0 + \sum_{k=2}^{\infty} R_{n-1}^k A_k\right) A_1^{-1}, \quad n \geqslant 1. \tag{3.5}$$

The above formulas provide a recursive method for computing R, but now the point is how to truncate the infinite series involved in the computations. Neuts [24, p. 37] describes a criterion for a possible choice of the level of truncation K. The application to our case leads to choose K as the first positive integer satisfying

$$\sum_{k=K+1}^{\infty} k\delta_{k-1} < \frac{10^{-8}}{\tau},$$

where $\tau = \max(\lambda + (c-1)\nu + \alpha, \lambda + c\nu + \delta)$. The truncation implies to compute R by using only the matrices A_i, for $0 \leqslant i \leqslant K$; i.e., we are neglecting the effect of matrices A_i, for $i > K$. It seems that this fact modifies essentially the matrix structure of $GI/M/1$ type of the model under study so, in what follows, we investigate an alternative procedure.

We note that A_0 can be expressed as $A_0 = \lambda \mathbf{e}_c \mathbf{e}_c'$, where $\mathbf{e}_c$ is a column vector of dimension $c+1$ whose elements are equal to zero excepts the last one which is equal to one. This special structure of A_0 leads to a matrix R_1 whose rows are equal to zero excepts the last row. Consequently, by iterating formula (3.5), we get

$$R = \begin{pmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{u} \end{pmatrix},$$

and now the problem reduces to the computation of $\mathbf{u}$. Note that $\mathbf{u}R = u_c \mathbf{u}$, where u_c is the last element of $\mathbf{u}$. Hence, R has a unique eigenvalue which is equal to u_c. In addition, according to the general theory u_c is also equal to the spectral radius of R; i.e., $\eta = u_c$. η can be computed by using a bisection method given in [24, pp. 39–40]. To this end, we define $-\Delta = \mathrm{diag}(A_1)$ and note that $R = \sum_{k=0}^{\infty} R^k B_k$, where $B_0 = A_0 \Delta^{-1}$, $B_1 = A_1 \Delta^{-1} + I$, $B_k = A_k \Delta^{-1}$, $k \geqslant 2$. Now η is the unique root in the interval (0, 1) of the equation $\eta = \chi(\eta)$, where $\chi(z)$ is the spectral radius of

$$B^*(z) = \sum_{k=0}^{\infty} B_k z^k = Iz + \left(\sum_{k=0}^{\infty} A_k z^k\right) \Delta^{-1}.$$

The elements of $\sum_{k=0}^{\infty} A_k z^k$ are

$$\left(\sum_{k=0}^{\infty} A_k z^k\right)_{ij} = \begin{cases} \lambda z + \alpha z^2, & \text{if } 0 \leqslant i \leqslant c-1,\ j = i+1, \\ i\nu z, & \text{if } 1 \leqslant i \leqslant c,\ j = i-1, \\ -(\lambda + i\nu + \alpha)z, & \text{if } 0 \leqslant i \leqslant c-1,\ j = i, \\ -(\lambda + c\nu + \delta)z + \lambda + \sum_{k=2}^{\infty} \delta_{k-1} z^k, & \text{if } i = j = c, \\ 0, & \text{otherwise.} \end{cases}$$

Now a recursive application of $\mathbf{u}R = \eta\mathbf{u}$ yields $\mathbf{u}R^k = \eta^k\mathbf{u}$. Then, multiplying the equation $\sum_{k=0}^{\infty} R^k A_k = \mathbf{0}$ by $\mathbf{u}$, we have $\mathbf{u}\sum_{k=0}^{\infty} \eta^k A_k = \mathbf{0}$. It provides the key to determine $(u_0, \ldots, u_{c-1})$ as a solution of a linear system of equations.

We now discuss the extension to the case of linear retrial policy. As usual, the first question to be investigated is the condition of positive recurrence.

Theorem 3.2. The queueing process $X(t) = (C(t), N(t))$, $t \geqslant 0$, operating under the linear retrial policy is positive recurrent if and only if

$$\lambda < c\nu + \delta g. \tag{3.6}$$

Proof. The necessity follows from the analysis done in the proof of theorem 3.1. Now we assume that $\mu > 0$ but equation (3.4) remains valid independently of the value of μ. Let $\lambda < c\nu + \delta g$. Consider the embedded Markov chain $\{Z_n;\ n \geqslant 0\}$ at the epochs when the process $X(t)$ changes its states. Note that $\beta = -\inf_{(i,j)\in S} q_{(i,j),(i,j)} > 0$, so a sufficient condition for the positive recurrence of $\{Z_n;\ n \geqslant 0\}$ is also sufficient for the process $X(t)$ in continuous time.

Now we use the classic Foster criterion: for an irreducible and aperiodic Markov chain $Z_n, n \geqslant 0$, with state space S, a sufficient condition for positive recurrence is the existence of a non-negative function $f(s)$, $s \in S$, a positive number ε and a finite subset $A \subset S$ such that the mean drift

$$\gamma_s = \mathbf{E}\big[f(Z_{n+1}) \mid Z_n = s\big] - f(s)$$

is finite for all $s \in A$ and $\gamma_s < -\varepsilon$ for all $s \notin A$.

In our case, we consider $f(i, j) = ai + j$, where a should be determined later. Then, we have

$$\gamma_{(i,j)} = \begin{cases} \dfrac{(\lambda - i\nu)a + (\alpha + j\mu)(a-1)}{\lambda + i\nu + \alpha + j\mu}, & \text{if } 0 \leqslant i \leqslant c-1, \\ \dfrac{\lambda - ac\nu - \sum_{k=1}^{j-1} k\delta_k - j\delta_j^*}{\lambda + c\nu + \delta}, & \text{if } i = c. \end{cases}$$

For $0 \leqslant i \leqslant c-1$, we have $\gamma_{(i,j)} \to a-1$, as $j \to \infty$. On the other hand, $\gamma_{(c,j)} \to (\lambda - ac\nu - \delta g)/(\lambda + c\nu + \delta)$ as $j \to \infty$.

To conclude we distinguish two cases. If $\lambda \leqslant \delta g$ then we choose any number a in the interval (0, 1) and let $\varepsilon = \min\{(1-a)/2, (ac\nu + \delta g - \lambda)/2(\lambda + c\nu + \delta)\}$. It is now clear that there exists $N(\varepsilon)$ such that if $j \geqslant N(\varepsilon)$ then $\gamma_{(c,j)} < -\varepsilon$ and also $\gamma_{(i,j)} < -\varepsilon$ for $0 \leqslant i \leqslant c-1$. In the case $\delta g < \lambda < c\nu + \delta g$ we choose the value a in such a way that $(\lambda - \delta g)/c\nu < a < 1$. □

Now to analyze the system state in the case of linear retrial policy we approximate the retrial rate by $\alpha(1-\delta_{j0}) + \min(j, M)\mu$ (see [25]). We mention that in (3.1), as $\alpha \to \infty$, $\pi_c \to 1$ and $\alpha(1-\pi_c) \to c\nu$. That means if the condition (3.6) is satisfied then at large enough M the condition (3.2) is also satisfied where instead of α we put $\alpha + M\mu$, and the system with constant retrial rate $\alpha + M\mu$ is ergodic. As the rates of the approximate system with retrial rate $\alpha(1-\delta_{j0}) + \min(j, M)\mu$ and the system with constant retrial rate $\alpha + M\mu$ differ only on the finite set $\{(i, j), i = 0, \dots, c, j \leqslant M\}$, then our approximate system is also ergodic. Thus, the corresponding infinitesimal generator leads to the following

$$Q^M = \begin{pmatrix} A_1^0 & A_0 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & & & & & \\ A_2^1 & A_1^1 & A_0 & \mathbf{0} & \mathbf{0} & \dots & & & & & \\ B_2 & A_2^2 & A_1^2 & A_0 & \mathbf{0} & \dots & & & & & \\ \vdots & & & \vdots & & & & & & & \\ B_{M-1} & A_{M-1} & \dots & A_3 & A_2^{M-1} & A_1^{M-1} & A_0 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots \\ B_M & A_M & \dots & A_4 & A_3 & A_2 & A_1 & A_0 & \mathbf{0} & \mathbf{0} & \dots \\ B_{M+1} & A_{M+1} & \dots & A_5 & A_4 & A_3 & A_2 & A_1 & A_0 & \mathbf{0} & \dots \\ & \ddots & \ddots & & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix},$$

where, for $1 \leqslant n \leqslant M$, we have

$$(A_1^n)_{ij} = \begin{cases} \lambda, & \text{if } 0 \leqslant i \leqslant c-1,\ j = i+1, \\ i\nu, & \text{if } 1 \leqslant i \leqslant c,\ j = i-1, \\ -(\lambda + i\nu + (1-\delta_{ic})(\alpha + n\mu) + \delta_{ic}\delta), & \text{if } 0 \leqslant i \leqslant c,\ j = i, \\ 0, & \text{otherwise,} \end{cases}$$

$$(A_2^n)_{ij} = \begin{cases} \alpha + n\mu, & \text{if } 0 \leqslant i \leqslant c-1,\ j = i+1, \\ \delta_{n1}\delta_1^* + (1-\delta_{n1})\delta_1, & \text{if } i = j = c, \\ 0, & \text{otherwise.} \end{cases}$$

In addition, we also have $A_1 = A_1^M$ and $A_2 = A_2^M$. The matrices A_0, $\{A_k\}_{k=3}^{\infty}$ and $\{B_k\}_{k=2}^{\infty}$ agree with those defined for the constant retrial case. Note that the generator Q^M corresponds to a matrix structure of the queueing model of type $GI/M/1$ with a large number of boundary states. Thus, it follows that the stationary probability vector $\mathbf{x}^M$ satisfies

$$\mathbf{x}_{M+k}^M = \mathbf{x}_{M-1}^M R^{k+1}, \quad \text{for } k \geqslant 0,$$

where R is the minimal non-negative solution of $\sum_{k=0}^{\infty} R^k A_k = \mathbf{0}$. The computation of R and the analysis of the stability condition match word by word the arguments of the constant retrial case by replacing α by $\alpha^* = \alpha + M\mu$.

Finally, we discuss the computation of vector $(\mathbf{x}_0^M, \ldots, \mathbf{x}_{M-1}^M)$. By partitioning Q^M, the problem reduces after some algebra to the solution of the following finite system:

$$\left(\sum_{i=0}^{M-2} \mathbf{x}_i^M + \mathbf{x}_{M-1}^M (I-R)^{-1}\right)\mathbf{e} = 1,$$
$$\left(\mathbf{x}_0^M, \ldots, \mathbf{x}_{M-1}^M\right)C + \eta^{-M}\mathbf{x}_{M-1}^M\left(\eta H^*, \eta^2 H_M^*, \ldots, \eta^{M-1}H_3^*, \eta^M\left(RA_2^* + H_2^*\right)\right) = \mathbf{0},$$

where C is the submatrix given by the first $(c+1)M$ rows and columns of Q^M, and A_2^*, H^* and $\{H_k^*\}_{k=2}^M$ are defined by

$$\left(A_2^*\right)_{ij} = \begin{cases} \alpha + M\mu, & \text{if } 0 \leqslant i \leqslant c-1,\ j = i+1, \\ 0, & \text{otherwise,} \end{cases}$$

$$\left(H^*\right)_{ij} = \begin{cases} \displaystyle\sum_{k=M}^{\infty} \eta^k \delta_k^*, & \text{if } i = j = c, \\ 0, & \text{otherwise,} \end{cases}$$

$$\left(H_k^*\right)_{ij} = \begin{cases} \displaystyle\sum_{n=k-1}^{\infty} \eta^n \delta_n, & \text{if } i = j = c, \\ 0, & \text{otherwise.} \end{cases}$$

The convergence of the vector $\mathbf{x}^M$ to the stationary probability vector of the initial system with linear retrial rate is investigated in the next section.

4. Approximation of the linear retrial model

On the way when we approximate the stationary distribution of the system with linear retrial rate by the stationary distribution of the approximate system, we need to prove that for any $(i, j) \in S$

$$\lim_{M\to\infty} \mathbf{x}_{ij}^M = \mathbf{x}_{ij}, \tag{4.1}$$

where $\mathbf{x}_{ij}^M$ is the stationary distribution of the approximate system. Some results in this direction (an approximation by a finite queueing model) for multiserver retrial queueing systems based on the notion of stochastic comparability of Markov processes and in particular migration processes are given in the book [19]. But in our case the process $X(t)$ is not a migration process and we have to prove the approximation by an infinite system. Therefore we provide a constructive proof based on the direct analysis of a return time to the initial state.

Consider the initial system QU with linear retrial policy. That means, if there are j customers in the retrial group then the total rate of repeated attempts is $\alpha_j = \alpha(1-\delta_{j0}) + j\mu,\ j \geqslant 0$.

Theorem 4.1. Assume that the condition of ergodicity (3.6) is satisfied. Then relation (4.1) is true.

Proof. Let QU^M be the approximate system with retrial rate $\alpha_j^M = \alpha(1-\delta_{j0}) + \min(j, M)\mu,\ j \geqslant 0$. As it was mentioned earlier, if condition (3.6) is satisfied then there exists M^* such that at $M \geqslant M^*$ the system QU^M is also ergodic. Below suppose that $M \geqslant M^*$. Consider the process $X(t) = \{(C(t), N(t));\ t \geqslant 0\}$ introduced earlier. We can always assume that it is a right-continuous process. By $X^M(t) = \{(C^M(t), N^M(t));\ t \geqslant 0\}$ denote the corresponding process for the truncated system QU^M. Both processes $X(t)$ and $X^M(t)$ produce a random walk on the semi-strip $S = \{0, \ldots, c\} \times \mathrm{N}$.

Let us study first some properties of the initial system QU. Consider the subset of states $D_c = \{(c, j), j = 0, 1, \ldots\}$. Construct the embedded Markov chain (MC) at hitting times to D_c in the following way. Let $t_1 < t_2 < \cdots$ be the times of sequential jumps of $X(t)$ (changes of any component). Denote by $Z_k = \{(C_k, N_k);\ k \geqslant 0\}$ the embedded MC, where $(C_k, N_k) = (C(t_k), N(t_k))$. We put

$$u_1 = \min\{i\colon\ i \geqslant 0, C_i = c\}, \qquad u_{k+1} = \min\{i\colon\ i > u_k, C_i = c\}, \quad k \geqslant 1,$$

and denote $\tilde{t}_k = t_{u_k}$, $z_k = N(\tilde{t}_k)$, $\psi_k = z_{k+1} - z_k,\ k \geqslant 1$. Here u_k are the times of successive hits into D_c for the embedded MC Z_k, $\tilde{t}_k$ are the instants of times of successive hits into D_c for $X(t)$, z_k are the values of the component $N(t)$ at these instants, and ψ_k are the values of jumps of the component $N(t)$ on the line $C(t) \equiv c$. Consider the sequence z_k. It also forms an embedded MC for $X(t)$ with state space $\{0, 1, \ldots\}$. Actually this is a random walk on a half-line. Denote by $\psi(j)$ the value of a jump in state j:

$$\mathbf{P}\big\{\psi(j) = i\big\} = \mathbf{P}\{\psi_k = i \mid z_k = j\}, \quad -j \leqslant i \leqslant 1.$$

Put $d(j) = \mathbf{E}\psi(j)$. We can write a representation

$$\psi(j) = \begin{cases} 1, & \text{with probability } \lambda/\lambda_c, \\ -\min(\gamma, j), & \text{with probability } \delta/\lambda_c, \\ A_{c-1}(j), & \text{with probability } c\nu/\lambda_c, \end{cases} \tag{4.2}$$

where $\lambda_c = \lambda + c\nu + \delta$, γ is the size of a batch that can be deleted by negative customers, and the variables $A_i(j)$ are constructed in the following way. Introduce at first random variables:

$$\alpha_i(j) = \min\{k\colon\ k > 0, C_k = c \text{ given that } C_0 = i,\ N_0 = j\}, \quad i = 0, 1, \ldots, c-1.$$

Here $\alpha_i(j)$ is the number of steps of Z_k up to returning into the subset D_c starting from state (i, j). Then

$$\mathbf{P}\big\{A_i(j) = s\big\} = \mathbf{P}\{N_{\alpha_i(j)} = j + s \mid C_0 = i, N_0 = j\}, \quad s = 0, -1, \ldots, -j. \tag{4.3}$$

Let us prove that the variables $A_{c-1}(j)$ have the 2nd moment bounded uniformly in $j \geqslant 1$, that means

$$\sup_{j\geqslant 1} \mathbf{E} A_{c-1}(j)^2 < C. \tag{4.4}$$

This in particular implies uniform integrability of $A_{c-1}(j)$:

$$\lim_{L\to\infty} \sup_{j\geqslant 1} \mathbf{E}\big|A_{c-1}(j)\big|\chi\big(\big|A_{c-1}(j)\big| \geqslant L\big) = 0. \tag{4.5}$$

By the construction with probability one $|A_{c-1}(j)| \leqslant \alpha_{c-1}(j)$, $j \geqslant 0$. That means it is enough to prove uniform boundedness in $j \geqslant 0$ of the 2nd moment of the variables $\alpha_{c-1}(j)$. At first we mention that at $c = 1$ $\alpha_0(j) = 1$ a.s. for any $j \geqslant 0$. Let $c > 1$. If $N_k = j$ and $C_k = i, i \neq c$, then at $j \geqslant 0$

$$\mathbf{P}\{C_{k+1} = i + 1 \mid C_k = i,\ N_k = j\} \geqslant \min_{0\leqslant i\leqslant c-1, 0\leqslant k\leqslant j} \frac{\lambda + \alpha_k}{\lambda + \alpha_k + i\nu} = \frac{\lambda}{\lambda + (c-1)\nu} = q,$$

where $0 < q < 1$. This relation implies at $c > 1$ for any $i = 0, 1, \ldots, c-1$, $j \geqslant 0$

$$\mathbf{P}\big\{\alpha_i(j) \leqslant c\big\} \geqslant q^{c-i} \geqslant q^c.$$

Using this inequality and Markov property we can prove that for any $m \geqslant 1$

$$\mathbf{P}\big\{\alpha_{c-1}(j) > mc\big\} \leqslant \Big(\max_{0\leqslant i\leqslant c-1,\, 0\leqslant k\leqslant j} \mathbf{P}\big\{\alpha_i(k) > c\big\}\Big)^m \leqslant \big(1 - q^c\big)^m \tag{4.6}$$

uniformly in $j \geqslant 0$. Relation (4.6) implies (4.4) and (4.5).

Consider the behaviour of $\psi(j)$ when $j \to \infty$. According to (4.2) it is enough to study the behaviour of $A_{c-1}(j)$. It is easy to see that as $j \to \infty$ $\mathbf{P}\{A_{c-1}(j) = -1\} \to 1$. Using (4.5) we get $\mathbf{E}A_{c-1}(j) \to -1$ and according to (4.2)

$$\lim_{j\to\infty} d(j) = d_\infty = \lambda_c^{-1}(\lambda - c\nu - \delta g). \tag{4.7}$$

If (3.6) is satisfied then there exist some $\varepsilon > 0$ and M^* such that $d(j) < -\varepsilon$ as $j \geqslant M^*$. Note that at $M \geqslant M^*$ the system QU^M is ergodic.

Consider now the system QU^M at $M \geqslant M^*$. We can introduce by analogy the variables $Z_k^M = (C_k^M, N_k^M)$, z_k^M, ψ_k^M, $\psi^M(j)$ and in the same way prove that the variables $A_{c-1}^M(j)$, $j \geqslant 1$, $M \geqslant 1$, have uniformly bounded 2nd moment that means

$$\sup_{j\geqslant 1, M\geqslant 1} \mathbf{E}\big|A_{c-1}^M(j)\big|^2 < C < \infty.$$

Let T_{00} be the number of jumps of the embedded MC Z_k between two successive returns to state (0, 0), that means

$$T_{00} = \min\{k\colon\ k > 0,\ Z_k = (0,0) \text{given that } Z_0 = (0,0)\}, \tag{4.8}$$

and $\widehat{T}_{00}$ be the return time to state (0, 0) for the process $X(t)$: $\widehat{T}_{00} = t_{T_{00}}$, where t_k were introduced as times of sequential jumps of $X(t)$. Denote by T_{00}^M and $\widehat{T}_{00}^M$ the corresponding variables for the system QU^M.

We prove now that the variables T_{00}^M and $\widehat{T}_{00}^M$ at $M \geqslant M^*$ have uniformly bounded 2nd moments, that is $\mathbf{E}(T_{00}^M)^2 < C, \mathbf{E}(\widehat{T}_{00}^M)^2 < C$ as $M \geqslant M^*$.

Consider the system QU^M and the sequence $z_k^M,\ k \geqslant 1$, at $M \geqslant M^*$. Note that z_k^M forms a random walk on a half line. Denote by S_{M^*} the finite region: $S_{M^*} = \{0, 1, \ldots, M^*\}$. We construct an auxiliary embedded SMP (semi-Markov process) $\{\kappa^M(m);\ m \geqslant 0\}$ at discrete times of hits of z_k^M into the region S_{M^*} in the following way. Denote:

$$\tilde{u}^M(0) = 0, \qquad \tilde{u}^M(1) = \min\{k\colon\ k > 0,\ z_k^M \in S_{M^*}\},$$
$$\tilde{u}^M(s+1) = \min\{k\colon\ k > \tilde{u}^M(s), z_k^M \in S_{M^*}\}, \quad s \geqslant 1,$$

and put $\kappa^M(m) = z_{\tilde{u}^M(s)}^M$ as $\tilde{u}^M(s) \leqslant m < \tilde{u}^M(s+1), m = 0, 1, \ldots.$

Let $\tau_s^M = \tilde{u}^M(s+1) - \tilde{u}^M(s),\ s \geqslant 1$. Denote by $\tau^M(j),\ j \in S_{M^*}$, the random variables such that

$$\mathbf{P}\{\tau^M(j) = k\} = \mathbf{P}\{\tau_1^M = k \mid z_{\tilde{u}^M(1)}^M = j\}, \quad k = 1, 2, \ldots,\ j \in S_{M^*}.$$

Note that $\tau^M(j)$ is the sojourn time in state j for $\kappa^M(m)$ (time till the next hit of z_k^M into S_{M^*} starting from j). According to the structure of transitions $\tau^M(j) = 1$ as $j < M^*$. Let us prove that the variables $\tau^M(M^*)$ have the 2nd moment bounded uniformly in $M \geqslant M^*$. Denote

$$\tilde{\tau}^M(M^*) = \min\{k\colon\ k > 0, z_k^M \in S_{M^*} \text{ given that } z_0^M = M^* + 1\},$$

($\tilde{\tau}^M(M^*)$ is the return time to the region S_{M^*} starting from state $M^* + 1$). Then

$$\tau^M(M^*) = \begin{cases} \tilde{\tau}^M(M^*), & \text{with probability } \lambda/\lambda_c, \\ 1, & \text{with probability } 1 - \lambda/\lambda_c, \end{cases}$$

where $\lambda_c = \lambda + c\nu + \delta$. Now if we prove that the variables $\tilde{\tau}^M(M^*), M \geqslant M^*$, have the 2nd moment bounded uniformly in $M \geqslant M^*$ then the variables $\tau^M(M^*), M \geqslant M^*$, have the same property.

Consider the variable $\tilde{\tau}^M(M^*)$. For any $m \geqslant 0$, we have

$$\mathbf{P}\{\tilde{\tau}^M(M^*) > m + 1\} = \mathbf{P}\left\{\sum_{i=0}^{k} \psi_i^M \geqslant 0,\ k = 0, 1, \ldots, m \mid z_0^M = M^* + 1\right\}.$$

Denote $f(a, j, M) = \mathbf{E}\exp\{a\psi^M(j)\},\ a \geqslant 0,\ g(a, M) = \sup_{j > M^*} f(a, j, M),\ S_k = \sum_{i=0}^{k} \psi_i^M,\ k \geqslant 0.$

Taking into account that for any function $\varphi(x) \geqslant 0$ due to Chebyshev's inequality $\mathbf{P}\{\varphi(\xi) \geqslant 1\} \leqslant \mathbf{E}\varphi(\xi)$, we take $\varphi(x) = \mathrm{e}^{ax}$ and get recursively for any $a > 0, m \geqslant 0$

$$\begin{aligned}
&\mathbf{P}\{\psi_0^M \geqslant 0 \mid z_0^M = M^* + 1\} = \mathbf{P}\{\psi^M(M^* + 1) \geqslant 0\} \\
&\quad = \mathbf{P}\{\exp\{a\psi^M(M^* + 1)\} \geqslant 1\} \leqslant \mathbf{E}\exp\{a\psi^M(M^* + 1)\} \leqslant g(a, M); \\
&\mathbf{P}\{\psi_0^M \geqslant 0, \psi_0^M + \psi_1^M \geqslant 0 \mid z_0^M = M^* + 1\} \\
&\quad = \sum_{i \geqslant 0} \mathbf{P}\{\psi^M(M^* + 1) = i\}\mathbf{P}\{i + \psi^M(i + M^* + 1) \geqslant 0\} \\
&\quad \leqslant \sum_{i \geqslant 0} \mathbf{P}\{\psi^M(M^* + 1) = i\}\mathbf{E}\exp\{a(i + \psi^M(i + M^* + 1))\} \\
&\quad \leqslant g(a, M)\sum_{i \geqslant 0} \mathbf{P}\{\psi^M(M^* + 1) = i\}\exp\{ai\} \leqslant g(a, M)^2; \\
&\mathbf{P}\left\{\psi_0^M \geqslant 0, \ldots, \sum_{k=0}^m \psi_k^M \geqslant 0 \mid z_0^M = M^* + 1\right\} \\
&\quad = \sum_{i \geqslant 0} \mathbf{P}\{S_0 \geqslant 0, S_1 \geqslant 0, \ldots, S_{m-2} \geqslant 0, S_{m-1} = i, \\
&\qquad\qquad i + \psi^M(i + M^* + 1) \geqslant 0 \mid z_0^M = M^* + 1\} \\
&\quad \leqslant \sum_{i \geqslant 0} \mathbf{P}\{S_0 \geqslant 0, S_1 \geqslant 0, \ldots, S_{m-2} \geqslant 0, S_{m-1} = i \mid z_0^M = M^* + 1\} \\
&\qquad \times \mathbf{E}\exp\{a(i + \psi^M(i + M^* + 1))\} \\
&\quad \leqslant g(a, M)\sum_{i \geqslant 0} \mathbf{P}\{S_0 \geqslant 0, S_1 \geqslant 0, \ldots, S_{m-2} \geqslant 0, S_{m-1} = i \mid z_0^M = M^* + 1\}\exp\{ai\} \\
&\quad \leqslant g(a, M)\sum_{k \geqslant 0} \mathbf{P}\{S_0 \geqslant 0, S_1 \geqslant 0, \ldots, S_{m-2} = k \mid z_0^M = M^* + 1\} \\
&\qquad \times \mathbf{E}\exp\{a(k + \psi^M(k + M^* + 1))\} \\
&\quad \leqslant g(a, M)^2\sum_{k \geqslant 0} \mathbf{P}\{S_0 \geqslant 0, S_1 \geqslant 0, \ldots, S_{m-2} = k \mid z_0^M = M^* + 1\}\exp\{ak\} \\
&\quad \leqslant \cdots \leqslant g(a, M)^{m+1},
\end{aligned}$$

that means

$$\mathbf{P}\{\tilde{\tau}^M(M^*) > m + 1\} \leqslant g(a, M)^{m+1}, \quad a > 0,\ m \geqslant 0. \tag{4.9}$$

Now we prove that for some $a_0 > 0$ there exists q_0, $0 < q_0 < 1$, such that

$$\sup_{M > M^*} g(a_0, M) < q_0. \tag{4.10}$$

As $j \to \infty$, according to (4.2) $\psi(j)$ converges in distribution to $\psi(\infty)$ where:

$$\psi(\infty) = \begin{cases} 1, & \text{with probability } \lambda/\lambda_c, \\ -\gamma, & \text{with probability } \delta/\lambda_c, \\ -1, & \text{with probability } c\nu/\lambda_c, \end{cases}$$

and $\mathbf{E}\psi(\infty) = d_\infty = \lambda_c^{-1}(\lambda - c\nu - \delta g) < 0$ (see (4.7)). Now the function $f(a) = \mathbf{E}\exp\{a\psi(\infty)\}$ is the moment generating function of $\psi(\infty)$. As $\psi(\infty)$ takes values in the space $\{1, 0, -1, -2, \ldots\}$, $f(a)$ exists at $a \geqslant 0$ and $f'(x)|_{x=0} = d_\infty < 0$. That means for some $a_0 > 0$ and $q < 1$ $f(a_0) < q$. Note that $\psi^M(j)$ also satisfies relation (4.2) with the variable $A_{c-1}^M(j)$, which is constructed for the system QU^M in the same way as in (4.3). It can be easily seen that

$$\sup_{j \geqslant M_1, M \geqslant M_1} \left| \mathbf{P}\left\{A_{c-1}^M(j) = -1\right\} - 1 \right| \to 0, \quad \text{as } M_1 \to \infty, \tag{4.11}$$

that means

$$\sup_{j \geqslant 1, M \geqslant 1} \mathbf{P}\left\{\psi^M(j) < -N\right\} \to 0, \quad \text{as } N \to +\infty. \tag{4.12}$$

Relations (4.11), (4.12) imply that for any $k = 1, 0, -1, \ldots$

$$\sup_{j \geqslant M_1, M \geqslant M_1} \left| \mathbf{P}\left\{\psi^M(j) = k\right\} - \mathbf{P}\left\{\psi(\infty) = k\right\} \right| \to 0 \quad \text{as } M_1 \to \infty. \tag{4.13}$$

Now using relations (4.12), (4.13) and the inequality

$$\begin{aligned} \left| f(a_0, j, M) - f(a_0) \right| \leqslant{} & \mathrm{e}^{a_0} \left| \mathbf{P}\left\{\psi^M(j) = 1\right\} - \mathbf{P}\left\{\psi(\infty) = 1\right\} \right| \\ & + \sum_{k=0}^{-L} \left| \mathbf{P}\left\{\psi^M(j) = k\right\} - \mathbf{P}\left\{\psi(\infty) = k\right\} \right| \\ & + \mathbf{P}\left\{\psi^M(j) < -L\right\} + \mathbf{P}\left\{\psi(\infty) < -L\right\}, \end{aligned}$$

which is valid for any integer $L > 0$, we get $\sup_{j \geqslant M_1, M \geqslant M_1} |f(a_0, j, M) - f(a_0)| \to 0$, as $M_1 \to \infty$. This relation implies at some q_0, $q < q_0 < 1$ and at large enough M^* relation (4.10). According to (4.9) the variables $\tilde{\tau}^M(M^*)$ have geometrically bounded tail and correspondingly their 2nd moments are bounded uniformly in $M \geqslant M^*$.

Consider now an auxiliary SMP $\kappa^M(m)$, $m \geqslant 0$, constructed above. Denote by y_k^M its embedded MC and put $p_{ij}^M = \mathbf{P}\{y_2^M = j \mid y_1^M = i\}$, $i, j \in S_{M^*}$. It can be easy seen that according to (4.2) at $M > M^*$ $p_{j,j+1}^M = \lambda/\lambda_c > 0$, $0 \leqslant j \leqslant M^* - 1$, and as $1 \leqslant j \leqslant M^*$,

$$\begin{aligned} p_{j,j-1}^M &\geqslant \frac{c\nu}{\lambda + c\nu + \delta} \, \frac{\alpha + j\mu}{\lambda + (c-1)\nu + \alpha + j\mu} \\ &\geqslant \frac{c\nu}{\lambda + c\nu + \delta} \, \frac{\alpha}{\lambda + (c-1)\nu + \alpha} = \varepsilon_0 > 0. \end{aligned}$$

As these inequalities are true uniformly in $M > M^*$ and S_{M^*} is a finite region, then all states communicate and the number of jumps of y_k^M between two successive returns to state 0 has uniformly in $M \geqslant M^*$ geometrically bounded tail. Now we can represented the return time of $\kappa^M(m)$ to state 0 in the form

$$\sum_{k=0}^{\nu(M)-1} \tau_k^M\big(y_k^M\big), \tag{4.14}$$

given that $y_0^M = 0$, where $\tau_k^M(j)$ are jointly independent, $\tau_k^M(j)$ has the same distribution as a sojourn time of $\kappa^M(m)$ in state j and $\nu(M)$ is the number of jumps of y_k^M between two successive returns to state 0. As it was proved, the values $\tau^M(j)$, $j \leqslant M^*$, have the 2nd moment bounded uniformly in $M \geqslant M^*$. Then representation (4.14) implies that the return time of $\kappa^M(m)$ to state 0 has also uniformly bounded in $M \geqslant M^*$ 2nd moment.

Consider now MP $X^M(t)$, $t \geqslant 0$, with the embedded MC Z_k^M, $k \geqslant 0$. We remind that the sequence z_l^M, $l \geqslant 0$, forms an embedded MC for Z_k^M at hitting times to the subset D_c. As relation (4.6) has the same form for the system QU^M, we can prove in the same way that the number of jumps of Z_k^M between two successive hits into D_c also has the 2nd moment bounded uniformly in $j \geqslant 0$, $M \geqslant M^*$. That means, the 2nd moments of variables T_{00} (see (4.8)) and T_{00}^M are also bounded uniformly in $M \geqslant M^*$ because these variables can be represented similar to (4.14) as sums of random variables with uniformly bounded 2nd moments on the embedded MC which again has uniformly bounded in $M \geqslant M^*$ 2nd moment of return time. Note that the occupation times in states (i, j) of the process $X^M(t)$ have exponential distributions with uniformly bounded in $0 \leqslant i \leqslant c$, $j \geqslant 0$, 2nd moment. Finally, this implies that the 2nd moment of variables $\widehat{T}_{00}^M$ is also bounded uniformly in $M \geqslant M^*$.

Now as $M \to \infty$, in each bounded region the transition rates of $X^M(t)$ converge to corresponding rates of $X(t)$. That means for any $k > 0$

$$\lim_{M\to\infty} \mathbf{P}\big\{T_{00}^M = k\big\} = \mathbf{P}\{T_{00} = k\},$$

and correspondingly $\widehat{T}_{00}^M$ converges in distribution to $\widehat{T}_{00}$. Together with the uniform boundedness of the 2nd moment this implies that $\mathbf{E}\widehat{T}_{00}^M \to \mathbf{E}\widehat{T}_{00}$. Finally, relation (4.1) follows from the convergence of the expectation of return time and the renewal theorem. □

5. Averaging principle in the overloading case

In this section, we study the behaviour of the retrial system in the overloading case, i.e., we consider the system on a large interval and suppose that the initial value $N(0)$ is also large. In such a case $N(t)$ converges in probability to infinity as $t \to \infty$ and we study the behaviour of the trajectory of the suitably normalized $N(t)$.

As the process $N(t)$ is not in general a Markov process, it is rather difficult to use the general results of the convergence in [18], and also it is not possible to apply directly the martingale approach [22]. Therefore we apply results given in the appendix on the convergence of switching processes [2–4], which are oriented towards the asymptotic analysis of recurrent type stochastic processes.

Let us generalize the model considered in section 2 by assuming that system parameters depend on the current value of the normalized queue $N(t)$. Suppose that we have the dependence on some scaling parameter n ($n \to \infty$) in the following way: if $n^{-1}N_n(t) = s$ then $\lambda(s)$ is a regular arrival rate, $\nu(s)$ equals the service rate, $\alpha(s)$ denotes the retrial rate, $\delta(s)$ is a negative arrival rate, $\gamma(s)$ denotes the batch size of customers which may be deleted by a negative arrival and $g(s) = \mathbf{E}\gamma(s)$.

With the help of a limit theorem given in appendix we study the averaging principle for the process $n^{-1}N_n(nt)$ as $n \to \infty$. Denote

$$\pi_i(s) = \frac{1}{A(s)i!}\left(\frac{\lambda(s) + \alpha(s)}{\nu(s)}\right)^i, \quad 0 \leqslant i \leqslant c, \tag{5.1}$$

where

$$A(s) = \sum_{k=0}^{c} \frac{1}{k!}\left(\frac{\lambda(s) + \alpha(s)}{\nu(s)}\right)^k,$$

and

$$\hat{b}(s) = \big(\lambda(s) - \delta(s)g(s)\big)\pi_c(s) - \big(1 - \pi_c(s)\big)\alpha(s). \tag{5.2}$$

Theorem 5.1. Suppose that $n^{-1}N_n(0) \xrightarrow{\mathrm{P}} s_0$, as $n \to \infty$, where $s_0 > 0$ a.s., and functions $\lambda(s)$, $\nu(s)$, $\alpha(s)$, $\delta(s)$ and $g(s)$ satisfy local Lipschitz condition and

$$\inf_{s \geqslant 0}\big(\lambda(s) + \alpha(s)\big) > 0, \qquad \inf_{s \geqslant 0} \nu(s) > 0.$$

Suppose there exist constants C_1, C_2 such that for any $s \geqslant 0$

$$\lambda(s) + \nu(s) + \alpha(s) + \delta(s) \leqslant C_1 + sC_2, \tag{5.3}$$

the function $g(s)$ is bounded, and for some $T > 0$ $s(t) > 0$, $t \in [0, T]$ a.s., where $s(t)$ is a solution of the equation

$$s(0) = s_0, \quad \mathrm{d}s(t) = \hat{b}\big(s(t)\big)\,\mathrm{d}t. \tag{5.4}$$

Then, as $n \to \infty$, we have

$$\sup_{0 \leqslant t \leqslant T} \left|n^{-1}N_n(nt) - s(t)\right| \xrightarrow{\mathrm{P}} 0,\text{[1]} \tag{5.5}$$

where a unique solution of (5.4) exists on each interval.

Proof. The process $X_n(t) = (C_n(t), N_n(t))$, $t \geqslant 0$, is a Markov process taking values on $\{0, \ldots, c\} \times \mathrm{N}$. Let $\{t_{nk}\}_{k=1}^{\infty}$ be the sequential times of jumps of $X_n(t)$. To investigate the asymptotic behaviour of $N_n(t)$ we first construct an auxiliary recurrent process

[1] Symbol $\xrightarrow{\mathrm{P}}$ means the convergence in probability.

of semi-Markov type. To this end, we define jointly independent families of random variables $F_k = \{(\xi_k(i,s), \tau_k(i,s), \beta_k(i,s));\ 0 \leqslant i \leqslant c,\ s \geqslant 0\}$, $k \geqslant 0$. Here $\tau_k(i,s)$ is independent of $(\xi_k(i,s), \beta_k(i,s))$ and follows an exponential law with parameter

$$\lambda_i(s) = \begin{cases} \lambda(s) + i\nu(s) + \alpha(s), & \text{if } 0 \leqslant i \leqslant c-1, \\ \lambda(s) + c\nu(s) + \delta(s), & \text{if } i = c. \end{cases}$$

The variables $(\xi_k(i,s), \beta_k(i,s))$ have the same distribution as the following ones:

$$\big(\xi(0,s), \beta(0,s)\big) = \begin{cases} (0,1), & \text{with probability } \lambda(s)/\lambda_0(s), \\ (-1,1), & \text{with probability } \alpha(s)/\lambda_0(s), \end{cases} \tag{5.6}$$

$$\big(\xi(i,s), \beta(i,s)\big) = \begin{cases} (0,i+1), & \text{with probability } \lambda(s)/\lambda_i(s), \\ (-1,i+1), & \text{with probability } \alpha(s)/\lambda_i(s), \\ (0,i-1), & \text{with probability } i\nu(s)/\lambda_i(s), \end{cases} \tag{5.7}$$
$$\text{for } 1 \leqslant i \leqslant c-1,$$

$$\big(\xi(c,s), \beta(c,s)\big) = \begin{cases} (1,c), & \text{with probability } \lambda(s)/\lambda_c(s), \\ (0,c-1), & \text{with probability } c\nu(s)/\lambda_c(s), \\ (-\gamma(s),c), & \text{with probability} \delta(s)/\lambda_c(s). \end{cases} \tag{5.8}$$

Following expressions (A.1)–(A.3), we now construct a recurrent process of semi-Markov type $\{(x_n(t), S_n(t));\ t \geqslant 0\}$. In our case, the variables $(\xi_k(\cdot), \tau_k(\cdot), \beta_k(\cdot))$ do not depend on index n. We put $(x_n(0), S_n(0)) = (C_n(0), N_n(0))$. According to the construction, the trajectories of $S_n(t)$ and $N_n(t)$ are coincident in any interval $[0, L]$ such that $S_n(t) > 0$, $t \in [0, L]$. It should be pointed out that the only difference in the construction of trajectories can be associated to the negative arrival epochs. That means, we take the value $N - \gamma(n^{-1}N)$ for a trajectory of $S_n(t)$ and the value $\max\{N - \gamma(n^{-1}N), 0\}$ for a trajectory of $N_n(t)$.

We now use theorem A.1 (see the appendix) to prove an averaging principle for process $(x_n(t), S_n(t))$, $t \geqslant 0$, constructed above.

First, we consider the sequence $\{x_{nk}\}_{k=0}^{\infty}$. Note that it is not, in general, a Markov process. Then, according to theorem A.1, we need to calculate at each fixed $s > 0$ a stationary distribution of an auxiliary Markov process $\{\tilde{x}_{nk}(s);\ k \geqslant 0\}$ with transition probabilities

$$p_n(i,j,s) = \mathbf{P}\big\{\beta_1(i,s) = j\big\}, \quad 0 \leqslant i, j \leqslant c,\ s > 0.$$

In our case, the above probabilities $p_n(i,j,s) = p(i,j,s)$ do not depend on index n and according to (5.6)–(5.8) we have

$$p(0,1,s) = 1,$$
$$p(i,j,s) = \begin{cases} (\lambda(s) + \alpha(s))\lambda_i(s)^{-1}, & \text{if } j = i+1, \\ i\nu(s)\lambda_i(s)^{-1}, & \text{if } j = i-1, \\ 0, & \text{otherwise}, \end{cases} \quad \text{for } 1 \leqslant i \leqslant c-1,$$

$$p(c,j,s)=\begin{cases}(\lambda(s)+\delta(s))\lambda_c(s)^{-1}, & \text{if } j=c,\\ c\nu(s)\lambda_c(s)^{-1}, & \text{if } j=c-1,\\ 0, & \text{otherwise.}\end{cases}$$

Denote $\tilde{x}_k(s)=\tilde{x}_{nk}(s)$. It can be easy checked that at our conditions for any fixed L there exists $\varepsilon>0$ such that as $0\leqslant s\leqslant L$ we have $p(i,i+1,s)\geqslant\varepsilon>0,\ i=0,1,\dots,c-1$ and $p(i,i-1,s)\geqslant\varepsilon>0,\ i=1,\dots,c$. That means the process $\tilde{x}_k(s)$ is ergodic uniformly in $0\leqslant s\leqslant L$.

The stationary distribution $\{\tilde{\pi}(i,s);\ 0\leqslant i\leqslant c\}$ of $\tilde{x}_k(s)$ satisfies the system of balance equations:

$$\tilde{\pi}(i,s)\frac{\lambda(s)+\alpha(s)}{\lambda_i(s)}=\tilde{\pi}(i+1,s)\frac{(i+1)\nu(s)}{\lambda_{i+1}(s)},\quad 0\leqslant i\leqslant c-1,$$

whose solution is

$$\tilde{\pi}(i,s)=\frac{\lambda_i(s)}{i!}\left(\frac{\lambda(s)+\alpha(s)}{\nu(s)}\right)^i\frac{\tilde{\pi}(0,s)}{\lambda_0(s)},\quad 1\leqslant i\leqslant c,$$

where

$$\tilde{\pi}(0,s)=\lambda_0(s)B(s)^{-1},\quad B(s)=\sum_{i=0}^{c}\frac{\lambda_i(s)}{i!}\left(\frac{\lambda(s)+\alpha(s)}{\nu(s)}\right)^i.$$

After some algebra, we now get the values of $m(s)$ and $b(s)$ (see (A.5), (A.6)), which are given by

$$m(s)=\sum_{i=0}^{c}\tilde{\pi}(i,s)\lambda_i(s)^{-1}, \tag{5.9}$$

$$b(s)=\big(\lambda(s)+\alpha(s)-\delta(s)g(s)\big)\tilde{\pi}(c,s)\lambda_c(s)^{-1}-\alpha(s)\sum_{i=0}^{c}\tilde{\pi}(i,s)\lambda_i(s)^{-1}.$$

It should be noted that values $\tilde{\pi}(i,s)\lambda_i(s)^{-1}m(s)^{-1},\ i=0,1,\dots,c$, are the stationary probabilities of the corresponding Markov process $\tilde{x}(t,s),\ t\geqslant 0$, in continuous time constructed for any fixed s with the help of imbedded Markov process $\tilde{x}_k(s)$ and exponential times with parameters $\lambda_i(s)$. As by the construction the transition from state c with rate $\lambda(s)+\delta(s)$ leaves the process in the same state c, then the process $\tilde{x}(t,s)$ is equivalent to a birth-and-death process with states $\{0,1,\dots,c\}$ and rates $\lambda(s)+\alpha(s)$ and $i\nu(s)$ of birth and death correspondingly. Therefore $\pi_i(s),\ i=0,1,\dots,c$, in (5.1) are the stationary probabilities of the process $\tilde{x}(t,s)$ and we can express $\hat{b}(s)=m(s)^{-1}b(s)$ as we claimed in (5.2). It is straightforward to verify that $\hat{b}(s)$ has no more than linear growth. In addition, it also satisfies a local Lipschitz condition. It therefore follows that a solution of (5.4) exists and is unique on any interval.

It remains to prove that $y(+\infty)>T$ a.s. for any $T>0$. To this end, we first note that under the conditions of theorem 5.1 we have that $\sup_{s>0}|b(s)|\leqslant C$. This implies that equation (A.8) has a unique solution on each interval and $\eta(u)$ a.s. satisfies the relation

$$\big|\eta(u)\big| \leqslant s_0 + Cu, \qquad u \geqslant 0. \tag{5.10}$$

Suppose first that $s_0 > 0$ is a deterministic value. It is easy to see that

$$m(s) \geqslant \min_{0 \leqslant i \leqslant c} \lambda_i(s)^{-1}. \tag{5.11}$$

Then, in the region $\eta(u) > 0$, we can combine (5.3), (5.10) and (5.11) to get

$$m\big(\eta(u)\big) \geqslant \frac{1}{C_3 + C_4\eta(u)} \geqslant \frac{1}{C_5 + C_6 u}, \tag{5.12}$$

where C_i are some positive constants.

Consider now an interval $[0, T]$ such that $s(t) > 0$, $t \in [0, T]$. Following [2] we have the representation $s(t) = \eta(y^{-1}(t))$. Then, for any u such that $y(u) < T$, we get $\eta(u) > 0$. If we suppose that $y(u) < T$ for all $u > 0$, then we obtain a contradiction, because relation (5.12) together with (A.7) yields $y(+\infty) = +\infty$. This proves that $y(+\infty) > T$. If s_0 is a random variable and $s(t) > 0, t \in [0, T]$ a.s., then for almost all realizations $s_0(\omega)$ we obtain that $y(+\infty, \omega) = +\infty$. This finally proves that $y(+\infty) > T$ a.s. Hence, all conditions of theorem A.1 are satisfied. Consequently, relation (5.5) is true for the normalized trajectory $n^{-1}S_n(nt)$ of the auxiliary recurrent process of semi-Markov type.

Finally, we show that the trajectories of $n^{-1}S_n(nt)$ and $n^{-1}N_n(nt)$ are asymptotically equivalent in the region $s(t) > 0, t \in [0, T]$. We can construct on the same probability space processes $S_n(nt)$ and $N_n(nt)$ in a recurrent way. First, we put $S_n(0) = N_n(0)$. Further, following the standard simulation techniques we can construct simultaneously on the same sequence of uniformly distributed random variables $\{\omega_k\}_{k=0}^{\infty}$ the trajectories of $S_n(nt)$ and $N_n(nt)$. Here $S_n(nt)$ is constructed directly according to formulas (A.1)–(A.3) and if $S_n(nt) > 0$, $t \in [0, T]$, then $S_n(nt) = N_n(nt)$, $t \in [0, T]$. Now, taking into account (A.9), we see that $s(t) > 0, t \in [0, T]$, implies

$$\mathbf{P}\big\{S_n(nt) > 0,\ t \in [0, T]\big\} \to 1, \quad \text{as } n \to \infty. \tag{5.13}$$

Then, for any measurable set $A \in B_{D_T}$, we get according to (5.13) that

$$\begin{aligned}
&\big|\mathbf{P}\big\{n^{-1}S_n(nt) \in A,\ t \in [0, T]\big\} - \mathbf{P}\big\{n^{-1}N_n(nt) \in A,\ t \in [0, T]\big\}\big| \\
&\quad \leqslant \big|\mathbf{P}\big\{n^{-1}S_n(nt) \in A,\ S_n(nt) > 0,\ t \in [0, T]\big\} \\
&\qquad - \mathbf{P}\big\{n^{-1}N_n(nt) \in A,\ S_n(nt) > 0,\ t \in [0, T]\big\}\big| \\
&\qquad + 2\mathbf{P}\big\{\text{exists } u,\ u \in [0, T], \text{such that } S_n(nu) \leqslant 0\big\} \\
&\quad = 2\mathbf{P}\big\{\text{exists } u,\ u \in [0, T], \text{such that } S_n(nu) \leqslant 0\big\} \to 0, \quad \text{as } n \to \infty.
\end{aligned}$$

The above relations show that the asymptotic behaviour of the trajectories of $S_n(nt)$ and $N_n(nt)$ is the same. Thus, we finally prove the relation (5.5). □

We next consider some particular cases. Suppose that $n^{-1}N_n(0) \xrightarrow{P} s_0 > 0$, where s_0 is some deterministic value. Assume that input, service rates and the rate of negative customers do not depend on the value of the queue and on the parameter n. That means $\lambda(s) \equiv \lambda$, $\nu(s) \equiv \nu$, $\delta(s) \equiv \delta$ and $g(s) \equiv g$ (these functions do not depend on s).

Consider first a constant retrial policy with rate α. That means in previous notation $\alpha(s) \equiv \alpha$. Denote $\hat{b} = (\lambda - \delta g)\pi_c - (1-\pi_c)\alpha$, where π_c is calculated according to (5.1).

Corollary 5.2. As $n \to \infty$, the relation (5.5) holds where $s(t) = s_0 + \hat{b}t,\ t \in [0, T]$. If $\hat{b} \geqslant 0$, it is true for any $T > 0$, if $\hat{b} < 0$, the value T should satisfy the relation $s_0 + \hat{b}T > 0$.

Consider now the model with linear retrial rate described in section 2. Suppose that system parameters depend on the scaling factor n in such a way that the retrial rate in state j has the form $\alpha(1 - \delta_{j0}) + \mu j/n$, where $\alpha > 0,\ \mu > 0$. In this case $\alpha(s) = \alpha + \mu s,\ s \geqslant 0$. Let $\pi_c(s)$ be calculated according to (5.1) with the function $\alpha(s) = \alpha + \mu s,\ s \geqslant 0$.

Corollary 5.3. As $n \to \infty$, the relation (5.5) holds for any $T > 0$ such that $s(t) > 0$, $t \in [0, T]$, where in (5.2) $\hat{b}(s) = (\lambda - \delta g)\pi_c(s) - (\alpha + \mu s)(1 - \pi_c(s))$.

Let us consider the behaviour of the function $s(t)$ in this case.

Proposition 5.4.

(1) Let $\lambda\pi_c \leqslant \alpha(1 - \pi_c) + \delta g\pi_c$. Then $\hat{b}(s) < 0$ for any $s > 0$. That means, the function $s(t)$ is monotonically decreasing to 0 as $t \to \infty$ and the system is stable in this sense. Relation (5.5) holds for any $T > 0$ such that $s(T) > 0$.

(2) Let $\lambda\pi_c > \alpha(1 - \pi_c) + \delta g\pi_c$ and $\lambda < c\nu + \delta g$. Then there exists a unique root s_* of the equation $\hat{b}(s) = 0$ which is a stable point of the equation. That means, relation (5.5) holds for any $T > 0$ and $s(t) \to s_*$ as $t \to \infty$. The value s_* is in some sense a quasi-stable point for the queue, at large n and t $N_n(nt) \asymp ns_*$ and we have quasi-stationary behaviour of the queue.

(3) Let $\lambda \geqslant c\nu + \delta g$. Then for all $s \geqslant 0$ $\hat{b}(s) > 0$. That means, relation (5.5) holds for any $T > 0$, $s(t)$ is monotonically increasing and the system is not stable.

Proof. We use similar arguments as at the proof of theorem 3.1. Define the auxiliary variable $u(s) = (\lambda + \alpha + \mu s)/\nu$. Then, the relation $\hat{b}(s) > 0$ (or $\geqslant 0$) is equivalent to the polynomial relation $R(u(s)) > 0$ (or $\geqslant 0$) where

$$R(u) = \frac{\lambda - c\nu - \delta g}{c!}u^c + \sum_{k=1}^{c-1} \frac{\lambda - k\nu}{k!}u^k + \lambda.$$

Consider first the case when $\lambda \geqslant c\nu + \delta g$. Then obviously $R(u) > 0$ for any $u \geqslant 0$. Consider now the 1st case. As it was shown in theorem 3.1, the relation $\lambda\pi_c < \alpha(1-\pi_c) + \delta g\pi_c$ is equivalent to the relation $R(u(0)) < 0$ or $\hat{b}(0) < 0$. This means the ergodicity of the system with constant retrial rate α. Then the system with constant retrial rate $\alpha + \mu s$ is also ergodic for any $s > 0$, that means $R(u(s)) < 0$ and $\hat{b}(s) < 0$ for any $s > 0$. Let now $\lambda\pi_c = \alpha(1-\pi_c) + \delta g\pi_c$. That means $R(u(0)) = 0$. This relation implies $\lambda < c\nu + \delta g$ (otherwise $R(u) > 0$, $u \geqslant 0$ and we get a contradiction). But now the coefficients of $R(u)$ have only one variation of sign. This means according to Descartes' rule of signs that $R(u)$ has only one positive root and $R(u(s)) < 0$ as $s > 0$ (correspondingly $\hat{b}(s) < 0$ as $s > 0$). Consider now the 2nd case. If $\lambda < c\nu + \delta g$ then, as it was mentioned, $R(u)$ has only one positive root. But if $\lambda\pi_c > \alpha(1-\pi_c) + \delta g\pi_c$, then according to theorem 3.1 $\hat{b}(0) > 0$ and $R(u(0)) > 0$. As $R(+\infty) = -\infty$, there exists a unique positive root s_* of the equation $\hat{b}(s) = 0$ and this root is a stable point. □

We mention that the results of averaging principle and diffusion approximation types for some models of overloading one and multi-server retrial queues (without negative customers) are obtained using the asymptotic technique for switching processes in [6,7].

Consider now the component $C_n(t)$ and study its long-run behaviour.

Theorem 5.5. Suppose that the conditions of theorem 5.1 hold. Then for any discrete function $f(i), i = 0, 1, \ldots, c$, and for any T chosen as in theorem 5.1, as $n \to \infty$ uniformly in $0 \leqslant t \leqslant T$

$$\frac{1}{n}\int_0^{nt} f\big(C_n(u)\big)\,\mathrm{d}u \xrightarrow{\mathrm{P}} \sum_{i=0}^{c} f(i)\int_0^t \pi_i\big(s(v)\big)\,\mathrm{d}v, \tag{5.14}$$

$$\frac{1}{n}\int_0^{nt} \mathbf{P}\big\{C_n(u) = i\big\}\,\mathrm{d}u \longrightarrow \int_0^t \pi_i\big(s(v)\big)\,\mathrm{d}v, \quad i = 0, 1, \ldots, c, \tag{5.15}$$

where $\pi_i(s)$ are calculated according to (5.1) and $s(t)$ is a solution of (5.4).

Proof. The proof follows the same steps as in theorem 5.1. We keep the previous notation for sequences t_{nk}, S_{nk}, constructed by variables $(\xi_k(i,s), \tau_k(i,s), \beta_k(i,s))$ according to relations (A.1)–(A.3). Denote $\zeta_n(t) = \int_0^{nt} f(C_n(u))\,\mathrm{d}u$, $\zeta_{nk} = \zeta_n(t_{nk})$, $k \geqslant 0$. The following representation is true

$$\zeta_n(t) = \zeta_{nk} + (t - t_{nk})f(x_{nk}), \quad \text{as } t_{nk} \leqslant t < t_{n,k+1}.$$

Put $\tilde{\zeta}_n(t) = \zeta_{nk}$ as $t_{nk} \leqslant t < t_{n,k+1}$. Then $\{(x_n(t), (S_n(t), \tilde{\zeta}_n(t)));\ t \geqslant 0\}$ is a recurrent process of semi-Markov type constructed by the families $((\xi_k(i,s), f(i)\tau_k(i,s)), \tau_k(i,s), \beta_k(i,s))$. According to theorem A.1

$$\sup_{0\leqslant t\leqslant T}\big|n^{-1}\tilde{\zeta}_n(nt) - v(t)\big| \xrightarrow{\mathrm{P}} 0, \quad n \to \infty, \tag{5.16}$$

and the function $v(t)$ satisfies the equation:

$$v(0) = 0, \qquad \mathrm{d}v(t) = m\big(s(t)\big)^{-1} h\big(s(t)\big)\, \mathrm{d}t,$$

where $m(s)$ is calculated as in (5.9) and $h(s) = \sum_{i=0}^{c} \tilde{\pi}(i, s) f(i) \lambda_i(s)^{-1}$. Using [3, theorem 1] we get that the limiting behaviour of the normalized trajectories of $n^{-1}\zeta_n(nt)$ and $n^{-1}\tilde{\zeta}_n(nt)$ is the same. Taking into account that $\tilde{\pi}(i, s)\lambda_i(s)^{-1}/m(s) = \pi_i(s)$, we get relation (5.14). As $n^{-1}|\zeta_n(nt)| \leqslant t \max_{i=0,1,\ldots,c} |f(i)|$ a.s., the convergence in probability implies the convergence of expectations and relation (5.15) is also true. □

Appendix A. Averaging principle for switching processes

The averaging principle for switching processes (SPs) provides an elegant mathematical approach for the investigation of the asymptotic behaviour when the number of switches tends to infinity. Under some natural assumptions, the normalized trajectory of SP uniformly converges in probability to some function which is a solution of some differential equation.

Let us consider an important subclass of SPs (see [1,4]) which is useful in our queueing application. For any $n > 0$, let $F_{nk} = \{(\xi_{nk}(x, s), \tau_{nk}(x, s), \beta_{nk}(x, s));\ x \in X, s \in \mathcal{R}^r\}$, $k \geqslant 0$, be jointly independent families of random vectors taking values in $\mathcal{R}^r \times [0, \infty) \times X$, where X be some measurable space. Let also (x_{n0}, S_{n0}) be an initial value independent of $\{F_{nk}\}_{k=0}^{\infty}$. Then, we construct the following recurrent sequences

$$t_{n0} = 0, \qquad t_{n,k+1} = t_{nk} + \tau_{nk}\big(x_{nk}, n^{-1}S_{nk}\big), \tag{A.1}$$

$$x_{n,k+1} = \beta_{nk}\big(x_{nk}, n^{-1}S_{nk}\big), \qquad S_{n,k+1} = S_{nk} + \xi_{nk}\big(x_{nk}, n^{-1}S_{nk}\big), \quad k \geqslant 0, \tag{A.2}$$

and denote

$$x_n(t) = x_{nk}, \qquad S_n(t) = S_{nk}, \quad \text{for } t_{nk} \leqslant t < t_{n,k+1}. \tag{A.3}$$

The process $\{(x_n(t), S_n(t));\ t \geqslant 0\}$ is a recurrent process of semi-Markov type with feedback between both components (a special subclass of SPs). In what follows, we assume that the distributions of variables in F_{nk} do not depend on index k. Note, that in general a component x_{nk} is not itself a Markov process, and a component $x_n(t)$ also in general is not a Markov and even a semi-Markov process.

Let us assume that moment functions $m_n(x, s) = \mathbf{E}\tau_{n1}(x, s)$ and $b_n(x, s) = \mathbf{E}\xi_{n1}(x, s)$ exist. Denote $p_n(x, A, s) = \mathbf{P}\{\beta_{n1}(x, s) \in A\}$, $x \in X$, $A \in B_X$, $s \in \mathcal{R}^r$. Suppose that at each $s \in \mathcal{R}^r$ there exists a family of transition probabilities $p_0(x, A, s)$, $x \in X$, $A \in B_X$, which are continuous in s uniformly in each bounded region $|s| \leqslant L$ with respect to $x \in X$, $A \in B_X$, and for any $L > 0$ as $n \to \infty$

$$\sup_{x, A, |s| \leqslant L} \big|p_n(x, A, s) - p_0(x, A, s)\big| \longrightarrow 0. \tag{A.4}$$

For any fixed $s \in \mathcal{R}^r$ consider an auxiliary Markov process $\{\tilde{x}_k(s); k \geqslant 0\}$ with state space X and transition probabilities $p_0(x, A, s)$, $x \in X$, $A \in B_X$. Assume that the process $\{\tilde{x}_k(s);\ k \geqslant 0\}$ is uniformly ergodic with stationary measure $\tilde{\pi}_0(A, s)$ uniformly in s in each bounded region. Denote

$$m_n(s) = \int_X m_n(x, s)\tilde{\pi}_0(\mathrm{d}x, s), \qquad b_n(s) = \int_X b_n(x, s)\tilde{\pi}_0(\mathrm{d}x, s). \tag{A.5}$$

Theorem A.1. Suppose that (A.4) holds and

(1) for any $N > 0$

$$\lim_{L\to\infty} \overline{\lim_{n\to\infty}} \sup_{|s|\leqslant N} \sup_x \bigl\{\mathbf{E}\tau_{n1}(x, s)\chi\bigl(\tau_{n1}(x, s) > L\bigr) + \mathbf{E}\bigl|\xi_{n1}(x, s)\bigr|\chi\bigl(\bigl|\xi_{n1}(x, s)\bigr| > L\bigr)\bigr\} = 0;$$

(2) for any x, as $\max(|s_1|, |s_2|) < N$

$$\bigl|m_n(x, s_1) - m_n(x, s_2)\bigr| + \bigl|b_n(x, s_1) - b_n(x, s_2)\bigr| \leqslant C_N|s_1 - s_2| + \alpha_n(N),$$

where C_N are constants, and $\alpha_n(N) \to 0$ uniformly in $|s_1| < N$, $|s_2| < N$;

(3) there exist functions $m(s) > 0$, $b(s)$ and a value s_0 (possibly random) such that as $n \to \infty$

$$m_n(s) \to m(s), \qquad b_n(s) \to b(s) \quad \text{and} \quad n^{-1}S_{n0} \xrightarrow{\mathrm{P}} s_0; \tag{A.6}$$

(4) there exists T such that $y(+\infty) > T$ a.s., where

$$y(t) = \int_0^t m\bigl(\eta(u)\bigr)\,\mathrm{d}u \tag{A.7}$$

and

$$\eta(0) = s_0, \qquad \mathrm{d}\eta(u) = b\bigl(\eta(u)\bigr)\,\mathrm{d}u. \tag{A.8}$$

Then

$$\sup_{0\leqslant t\leqslant T} \bigl|n^{-1}S_n(nt) - s(t)\bigr| \xrightarrow{P} 0, \quad n \to \infty, \tag{A.9}$$

where $s(t)$ is a unique solution of the equation

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

The proof of theorem A.1 follows from the averaging principle for general recurrent sequences and SPs with feedback [2,3].

References

[1] V.V. Anisimov, Switching processes, Cybernetics 13(4) (1977) 590–595.

[2] V.V. Anisimov, Averaging principle for switching recurrent sequences, Theory Probab. Math. Statist. 45 (1992) 1–8.

[3] V.V. Anisimov, Averaging principle for switching processes, Theory Probab. Math. Statist. 46 (1993) 1–10.
[4] V.V. Anisimov, Switching processes: averaging principle, diffusion approximation and applications, Acta Appl. Math. 40 (1995) 95–141.
[5] V.V. Anisimov, Diffusion approximation for processes with semi-Markov switches and applications in queueing models, in: *Semi-Markov Models and Applications*, eds. J. Janssen and N. Limnios (Kluwer, Dordrecht, 1999) pp. 77–101.
[6] V.V. Anisimov, Averaging methods for transient regimes in overloading retrial queuing systems, Math. Comput. Modelling 30(3/4) (1999) 65–78.
[7] V.V. Anisimov, Switching stochastic models and applications in retrial queues, Top 7(2) (1999) 169–186.
[8] J.R. Artalejo, A classified bibliography of research on retrial queues: Progress in 1990–1999, Top 7(2) (1999) 187–211.
[9] J.R. Artalejo, G-networks: a versatile approach for work removal in queueing networks, European J. Oper. Res. 126 (2000) 233–249.
[10] J.R. Artalejo and A. Gomez-Corral, Stochastic analysis of the departure and quasi-input processes in a versatile single-server queue, J. Appl. Math. Stochastic Anal. 9 (1996) 171–183.
[11] J.R. Artalejo and A. Gomez-Corral, Steady state solution of a single-server queue with linear repeated requests, J. Appl. Probab. 34 (1997) 223–233.
[12] J.R. Artalejo and A. Gomez-Corral, Generalized birth and death processes with applications to queues with repeated attempts and negative arrivals, OR Spektrum 20 (1998) 5–14.
[13] J.R. Artalejo and A. Gomez-Corral, On a single server queue with negative arrivals and request repeated, J. Appl. Probab. 36 (1999) 907–918.
[14] X. Chao, M. Miyazawa and M. Pinedo, *Queueing Networks: Customers, Signals and Product Form Solutions* (Wiley, Chichester, 1999).
[15] B.D. Choi, Y. Chang and B. Kim, $MAP_1/MAP_2/M/c$ retrial queue with guard channels and its applications to cellular networks, Top 7 (1999) 231–248.
[16] J.E. Diamond and A.S. Alfa, Matrix analytical methods for multi-server retrial queues with buffers, Top 7 (1999) 249–266.
[17] A.N. Dudin and V.I. Klimenok, A retrial $BMAP/SM/1$ system with linear requests, Queueing Systems 34 (2000) 47–66.
[18] S.N. Ethier and T.G. Kurtz, *Markov Processes, Characterization and Convergence* (Wiley, New York, 1986).
[19] G.I. Falin and J.G.C. Templeton, *Retrial Queues* (Chapman and Hall, London, 1997).
[20] E. Gelenbe and G. Pujolle, *Introduction to Queueing Networks* (Wiley, Chichester, 1998).
[21] N. Grier, W.A. Massey, T. McKoy and W. Whitt, The time-dependent Erlang loss model with retrials, Telecommunication Systems 7 (1997) 253–265.
[22] J. Jacod and A.N. Shiryaev, *Limit Theorems for Stochastic Processes* (Springer, Berlin, 1987).
[23] G. Latouche and V. Ramaswani, *Introduction to Matrix Analytic Methods in Stochastic Modeling* (ASA–SIAM, Philadelphia, 1999).
[24] M.F. Neuts, *Matrix-Geometric Solutions in Stochastic Models – An Algorithmic Approach* (Johns Hopkins Univ. Press, Baltimore, MD, 1981).
[25] M.F. Neuts and B.M. Rao, Numerical investigation of a multiserver retrial model, Queueing Systems 7 (1990) 169–190.
[26] R. Serfozo, *Introduction to Stochastic Networks* (Springer, New York, 1999).