On Finite Markovian Queues with Repeated Attempts

J. R. Artalejo [1] *V. Rajagopalan* [2] *R. Sivasamy* [2]

[1] Facultad de Matemáticas
Departamento de Estadística e I.O.
Universidad Complutense de Madrid
Madrid 28040

[2] Department of Statistics
Annamalai University
Annamalainagar
Tamilnadu
India 608 002

Abstract
This paper deals with the stochastic modelling of a wide class of finite retrial queueing systems in a Markovian environment. The main characteristic of this class is its versatility, so we discuss many applications associated with the specific choices of the system parameters. The infinitesimal generator of these queues can be reduced to a finite block-tridiagonal structure investigated in other papers [9, 10, 12, 13, 17]. The stationary distribution of the system state and first passage times are obtained by matrix methods. A numerical illustration is given.

Keywords: Finite block-tridiagonal; Queueing; Repeated attempts.

1 Introduction

Queueing systems with repeated attempts have wide practical use in designing local area networks and telecommunication systems. These queues are characterized by the following feature: a customer finding all servers and waiting positions busy upon arrival is obliged to leave the service area, but after some random time he repeats his demand. Between trials a customer is said to be in '*orbit*'. The growing interest on this topic is reflected in the publication during the last decades of an important number of papers. For a systematic account of the fundamental methods and results on this topic, we refer to the reader to the monograph by Falin and Templeton [7].

It should be pointed out that the infinitesimal generator of many Markovian queues with repeated attempts can be re-expressed as an infinite block-tridiagonal one that generalizes the so-called '*quasi birth and death processes*' [14]. The theory for infinite quasi birth and death processes assumes that transitions from a state (i, j) do not depend on j. This condition of limited space homogeneity implies a structural form where three matrix $(A_2,\ A_1,\ A_0)$ are recursively iterated. However, the main feature of a retrial queue is space heterogeneity which is caused by the transitions associated to the repeated attempts. That is the main difference between Markovian retrial queues and the theory of Neuts for infinite quasi birth and death processes. However, as far as we know, the relationship between retrial queues and finite quasi birth and death processes has not been observed in the existing literature. Thus, we hope that this paper would be useful to exploit in future research the matrix-geometric methodology for finite quasi birth and death processes as a powerful tool for the investigation of a wide class of finite retrial queues.

We are concerned with a versatile queueing system which is formalized as a Markovian bivariate process $\{(Q(t),\ O(t));\ t \geq 0\}$ on the finite state space $S = \{0, ..., M\} \times \{0, ..., N\}$. The variable $Q(t)$ denotes the number of customers in the service facility at time t, and $O(t)$ is the number of customers in orbit. We consider the most general setting by assuming that all rates involved in the model description depend on the system state.

First, in Section 2, we describe the mathematical model. In Section 3, we reduce the infinitesimal generator of our Markovian process to a finite block-tridiagonal one. Thus, the methodology developed in [9] is the mathematical key for investigating the steady state distribution and expectations of first passage times (Section 4). Our aim is to show a novel use of matrix-geometric methods. To that end, in Section 5, we illustrate how many retrial queues are connected with the model under consideration. Finally, a numerical illustration is provided in Section 6.

2 Model description

We consider a retrial queueing system with finite capacity in the service facility and in the orbit. The arrival input, service times and intervals between successive repeated attempts are assumed to be mutually independent. The state of the system can be described in terms of the process $\{(Q(t),\ O(t));\ t \geq 0\}$, where $Q(t)$ is the total number of servers and waiting positions occupied and $O(t)$ is the number of customers reapplying for service at time t. We assume a state dependent description in which the jumps of $(Q(t),\ O(t))$ are governed by exponential laws with parameters λ_{ij} (primary arrivals), μ_{ij} (service times) and ν_{ij} (repeated attempts). Taking into account the lack of memory property of the exponential distribution, we conclude that $\{(Q(t),\ O(t));\ t \geq 0\}$ is a time-homogeneous Markov chain with state space $S = \{0, ..., M\} \times \{0, ..., N\}$. Its infinitesimal transition rates are given by

$$q_{ij} = \lim_{t\to\infty} (P\{Q(t) = m,\ O(t) = n \mid\ Q(0) = i,\ O(0) = j\} - \delta_{(i,j),(m,n)})t^{-1}$$

$$= \begin{cases} \lambda_{ij} p_i, & \text{if } (m,n) = (i+1,j), \\ (1-\delta_{jN})\lambda_{ij} q_i, & \text{if } (m,n) = (i,j+1), \\ \mu_{ij}, & \text{if } (m,n) = (i-1,j), \\ \nu_{ij} h_i, & \text{if } (m,n) = (i+1,j-1), \\ (1-\delta_{j0})\delta_{iM}\nu_{ij} h, & \text{if } (m,n) = (M,j-1), \\ -(\lambda_{ij}(p_i + (1\text{-}\delta_{jN})q_i) + \mu_{ij} + \nu_{ij} h_i + (1\text{-}\delta_{j0})\delta_{iM}\nu_{ij} h), & \text{if } (m,n) = (i,j), \\ 0, & \text{otherwise.} \end{cases} \tag{2.1}$$

It should be noted that p_i and q_i represent balking probabilities; i.e., an arriving customer finding the service facility at the state i decides either to join the service facility, with probability p_i, or to join the orbit and retry later, with probability q_i, or to leave the system, with probability $1 - p_i - q_i$. The probability h_i has the same mean that p_i but replacing primary arrivals by repeated attempts. In addition, a repeated attempt finding all servers and waiting positions occupied decides to leave the system with probability h. Obviously, we have $p_M = h_M = \mu_{0j} = \nu_{i0} = 0$.

Since the process $(Q(t),\ O(t))$ is irreducible and its state space is finite one concludes that the process is positive recurrent and regular; i.e., it takes only a finite number of jumps in any finite interval.

In Section 3 we will re-express the infinitesimal generator $\mathbf{Q} = [q_{ij}]$ in terms of a finite block-tridiagonal one, so the main probabilistic characteristics of our queueing system follow as an application of the methodology developed by several authors [9, 10, 12, 13, 17].

3 Mathematical solution

Let $S_n = \{(i,j) \in S \mid i + j = n\}$ for $0 \leq n \leq M + N$. Then $S = \bigcup_{n=0}^{M+N} S_n$. Such a grouping will not affect the basic properties of the process $\{(Q(t),\ O(t));\ t \geq 0\}$. We arrange the states of the diagonal subset S_n as $S_n = \{(n,0), (n-1,1), ..., (1,n-1), (0,n)\} \cap S$. Then, we observe that the generator $\mathbf{Q}$ can be re-expressed as follows:

$$\mathbf{Q} = \begin{array}{c|cccccccc|} & \mathrm{S}_0 & \mathrm{S}_1 & \mathrm{S}_2 & \mathrm{S}_3 & \ldots & \mathrm{S}_{M+N-2} & \mathrm{S}_{M+N-1} & \mathrm{S}_{M+N} \\ \mathrm{S}_0 & \mathbf{R}_0 & \mathbf{A}_0 & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathrm{S}_1 & \mathbf{D}_1 & \mathbf{R}_1 & \mathbf{A}_1 & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathrm{S}_2 & \mathbf{0} & \mathbf{D}_2 & \mathbf{R}_2 & \mathbf{A}_2 & \ldots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ . & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \mathrm{S}_{M+N-1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{D}_{M+N-1} & \mathbf{R}_{M+M-1} & \mathbf{A}_{M+N-1} \\ \mathrm{S}_{M+N} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{D}_{M+N} & \mathbf{R}_{M+N} \end{array}, \tag{3.1}$$

where $\mathbf{R}_n = [r^n_{(i,j),(m,n)}]$ are square matrices of order $\sharp(S_n)$ and their elements are given by

$$r^n_{(i,j),(m,n)}$$

$$= \begin{cases} \nu_{ij}h_i, & \text{if } (m,n) = (i+1, j-1), \\ -(\lambda_{ij}(p_i + (1\text{-}\delta_{jN})q_i) + \mu_{ij} + \nu_{ij}h_i + (1\text{-}\delta_{j0})\delta_{iM}\nu_{ij}h), & \text{if } (m,n) = (i,j), \\ 0, & \text{otherwise.} \end{cases} \tag{3.2}$$

The matrices $\mathbf{A}_n = [a^n_{(i,j),(m,n)}]$ govern the transitions from the subset S_n into the states of S_{n+1} and hence are rectangular $\sharp(S_n) \times \sharp(S_{n+1})$ with the following elements

$$a^n_{(i,j),(m,n)} = \begin{cases} \lambda_{ij}p_i, & \text{if } (m,n) = (i+1, j), \\ (1-\delta_{jN})\lambda_{ij}q_i, & \text{if } (m,n) = (i, j+1), \\ 0, & \text{otherwise.} \end{cases} \tag{3.3}$$

Finally, the matrices $\mathbf{D}_n = [d^n_{(i,j),(m,n)}]$ are also rectangular $\sharp(S_n) \times \sharp(S_{n-1})$. The matrix $\mathbf{D}_n$ governs the transitions from the states of S_n to S_{n-1}, so we have

$$d^n_{(i,j),(m,n)} = \begin{cases} \mu_{ij}, & \text{if } (m,n) = (i-1, j), \\ (1\text{-}\delta_{j0})\delta_{iM}\nu_{ij}h, & \text{if } (m,n) = (M, j-1), \\ 0, & \text{otherwise.} \end{cases} \tag{3.4}$$

The above finite block-tridiagonal generator is called *finite birth and death process.* Existing literature for infinite birth and death models assumes spatial homogeneity assumptions as we mention in Section 1. However, the finite case with a general matrix structure can be investigated using a variety of mathematical approaches. We next briefly review the main results for the finite case. These results provide the mathematical framework for the determination of the main performance characteristics of our versatile Markovian retrial queue.

The literature devoted to finite quasi birth and death processes has mostly focused on the determination of the stationary distribution, the study of first passage times and the analysis of sojourn times. Sometimes the results are given for the case of a discrete time process but the developments can usually be extended to the continuous time case.

Gaver *at al.* [9] developed a direct approach for computing the stationary distribution. This method is similar to Gaussian elimination algorithms. The interested reader may also found information about other previous papers and comparisons among the numerical efficiency of several approaches. The homogeneous case has been studied by Hajek [10] and Ye and Li [17]. The method suggested by Ye and Li is based on a bisection approach, leading to significant improvements on the computational efforts.

Gaver *et al.* [9] also analyzed the first passage times for general finite quasi birth and death processes. They obtained two systems of recurrence equations for the Laplace-Stieltjes transforms of passage times to higher and lower levels. Recently, Latouche and Ramaswani [12] developed more efficient algorithms for the homogeneous case.

Finally, we mention the paper by Li and Sheng [13] who investigated a highly stable folding algorithm for sojourn times of finite quasi birth and death processes.

Of course, we do not pretend to describe here the technical characteristics and differences among the existing algorithms. More detailed information on the computational complexity, memory requirements and another approaches and queueing applications can be obtained from the above references.

4 The stationary distribution and first passage times

For the sake of completeness, we next summarize some useful results for computing the stationary distribution and the expected value of some upper and lower first passage times. These results are the key for developing the numerical example given in Section 6.

For $0 \leq n \leq M+N$, we observe the process $\{(Q(t), O(t));\, t \geq 0\}$ during those intervals of time spent at S_n, before it enters S_{n+1} for the first time. We denote the infinitesimal generator of this restricted process by $\mathbf{C}_n$. $\mathbf{C}_n$ can be determined as follows

$$\mathbf{C}_n = \begin{cases} \mathbf{R}_0, & \text{for } n=0, \\ \mathbf{R}_n + \mathbf{D}_n(-\mathbf{C}_{n-1}^{-1})\mathbf{A}_{n-1}, & \text{for } 1 \leq n \leq M+N, \end{cases} \tag{4.1}$$

where $\mathbf{C}_n$-processes, for $0 \leq n \leq M+N-1$, are all transient while $\mathbf{C}_{M+N}$-process is positive recurrent.

On the other hand, for $0 \leq n \leq M+N$, we denote by $\widehat{\mathbf{C}}_n$ the infinitesimal generator of the restriction of the process $\{(Q(t), O(t));\ t \geq 0\}$, observed during those intervals of time spent at S_n before the original process moves to S_{n-1} for the first time. Then, all $\widehat{\mathbf{C}}_n$-processes, $1 \leq n \leq M+N$, are transient while $\widehat{\mathbf{C}}_0$ is positive recurrent. The matrices $\widehat{\mathbf{C}}_n$ are recursively computed as follows

$$\widehat{\mathbf{C}}_n = \begin{cases} \mathbf{R}_n + \mathbf{A}_n(-\widehat{\mathbf{C}}_{n+1}^{-1})\mathbf{D}_{n+1}, & \text{for } 0 \leq n \leq M+N-1, \\ \mathbf{R}_{M+N} & \text{for } n = M+N. \end{cases} \tag{4.2}$$

Let us denote the stationary probability vector of the process $\{(Q(t), O(t));\, t \geq 0\}$ by $\mathbf{P} = (\mathbf{P}_0, ..., \mathbf{P}_{M+N})$ so that $\mathbf{P}\mathrm{Q} = 0$ and $\mathbf{P}\mathbf{e} = 1$, where $\mathbf{e}$ denotes a column vector with all its elements equal to one, and $\mathbf{P}_n$, $0 \leq n \leq M+N$, are row vectors of dimension $\sharp(S_n)$. Then, $\mathbf{P} = (\mathbf{P}_0, ..., \mathbf{P}_{M+N})$ satisfies the following

$$\mathbf{P}_{M+N}\mathbf{C}_{M+N} = \mathbf{0}, \tag{4.3}$$

$$\mathbf{P}_n = \mathbf{P}_{n+1}\mathbf{D}_{n+1}(-\mathbf{C}_n^{-1}), \text{ for } 0 \leq n \leq M+N-1, \tag{4.4}$$

$$\sum_{n=0}^{M+N} \mathbf{P}_n\mathbf{e} = 1. \tag{4.5}$$

From (4.3), vector $\mathbf{P}_{M+N}$ could be determined uniquely, up to a multiplicative constant. This constant is decided by (4.4) and (4.5).

Now we turn our attention to the computation of the first passage times. Let the expected values of the first passage time from the states of S_{n-1} to the states of S_n be $u^{(n)}$ and that of the states of S_{n+1} to S_n be $v^{(n)}$. Then, they satisfy the following recurrence relations

$$u^{(n)} = \begin{cases} -\mathbf{C}_0^{-1}\mathbf{e}, & \text{for } n = 1, \\ -\mathbf{C}_{n-1}^{-1}(\mathbf{e} + \mathbf{D}_{n-1}u^{(n-1)}), & \text{for } 2 \leq n \leq M+N, \end{cases} \tag{4.6}$$

$$v^{(n)} = \begin{cases} -\widehat{\mathbf{C}}_{n+1}^{-1}(\mathbf{e} + \mathbf{A}_{n+1}v^{(n+1)}), & \text{for } 0 \leq n \leq M+N-2, \\ -\widehat{\mathbf{C}}_{M+N}^{-1}\mathbf{e}, & \text{for } n = M+N-1. \end{cases} \tag{4.7}$$

5 Particular cases

In this section, we illustrate the usefulness of our retrial queueing system by describing a variety of queueing phenomena which are obtained as particular cases of our model description.

1. *Truncated Poisson input* (Falin and Templeton [7], Pearce [15] and Stepanov [16])

The arrival process can be modelled by taking $\lambda_{ij} = \lambda$ for $0 \leq i \leq M-1$, $0 \leq j \leq N$, $\lambda_{Mj} = \lambda$ for $0 \leq j \leq N-1$ and $\lambda_{MN} = 0$.

2. *Service facility with c servers and M-c waiting positions*

The service time parameters are now given by $\mu_{ij} = \min(i,c)\mu$ for $0 \leq i \leq M$, $0 \leq j \leq N$.

3. *Linear retrial policy* (Artalejo and Gómez-Corral [2] and Fayolle [8])

The retrial policy studied in [2] considers simultaneously the classical retrial rate which depends on the orbit size and the homogeneous one introduced in [8]. Thus, we have $\nu_{ij} = \alpha(1-\delta_{j0}) + j\mu$ for $0 \leq j \leq N$.

4. *Balking/retrial policy* (Artalejo [1] and Falin and Artalejo [4])

If we choose $p_i + q_i < 1$ then an arriving customer may leave the system without receive service. On the other hand, $p_i = h_i = 1$, for $0 \leq i \leq c-1$, means that any customer finding a server free starts automatically to be served. Many practical applications satisfy that p_i is a decreasing function of the number of customers at the service facility.

5. *Retrial queues with priority customers* (Choi and Park [3] and Falin *et al.* [6])

We now consider $c = 1$, $p_0 = 1$, $p_i = p \in (0,1)$ for $0 \leq i \leq M-1$ and $q_i = 1 - p_i$ for $0 \leq i \leq M$. A priority customer joins the service facility whereas a non priority one only joins the server if it is idle upon arrival. Otherwise, he joins the orbit. We also assume that the priority discipline is of the *head-of-the-line* type. Note that $h_0 = 1$, $h_i = 0$ for $0 \leq i \leq M-1$.

6. *Non persistent customers* (Falin and Templeton [7])

In this case a customer after some unsuccessful retrials gives up further repeated attempts and leaves the system. It occurs with a probability h which does not depend on the number of previous repeated attempts.

7. *Quasi-random input* (Falin and Artalejo [5], Falin and Templeton [7] and Kornyshev [11])

We consider a system with K identical sources that request service according to an exponential distribution with rate λ. Then, the arrival parameters are $\lambda_{ij} = \lambda(K-i-j)$ for $0 \leq i \leq M < K$ and $0 \leq j \leq N = K-c$. c denotes again the number of servers. We assume that $p_i = h_i = 1$, for $0 \leq i \leq c-1$, to guarantee that customers finding any server idle automatically get it. In this case, the state space reduces to $S = \{(i,j) \mid 0 \leq i \leq c,\ 0 \leq j \leq N\} \cup \{(i,j) \mid c+1 \leq i \leq M,\ 0 \leq j \leq N,\ i+j \leq K\}$.

6 Numerical example

It is clear that the size of the queue length at the service facility and the size of the retrial group are influenced by the decisions of retrial customers. Hence the way of joining the service facility by any unit in orbit can be considered as one of the most important characteristics of a retrial queue. This ensures that $\{h_i;\ 0 \leq i \leq M-1\}$ and the value of h are significant parameters for the above system.

We next assume that in some local area network, customers are classified into primary types and retrial types as described in Sections 1 and 2. It is reasonable to incorporate a reward structure and to choose the best value of h (probability of leaving the system when a repeated attempt finds the service facility full) optimizing the superimposed reward function. In view of the fact that for any given queueing problem the cost/reward structure is not uniquely defined, we have not pursued optimization too far. Thus the reward structure chosen below and the ensuing analysis should be considered as a simple sample of possible approaches to optimization and not as a vital stage in the development of the study.

We construct a reward function of the simplest possible type: The contributions of the components to the total profit are assumed to be linear with respect to their averages values. Essentially we must consider two such components: The expected number of customers in the service facility, $Q = \lim_{t \longrightarrow \infty} E[Q(t)]$, and the expected number of customers in orbit, $O = \lim_{t \longrightarrow \infty} E[O(t)]$. Let the amount of collections per unit time be made as α and β monetary units per customer from the service facility and the orbit, respectively. Thus, we obtain

$$T(h) = \alpha Q + \beta O. \tag{6.1}$$

As the explicit expressions for Q and O are not available in terms of h, one could trace the optimal value of h from the values of $T(h)$ by varying the value of h. For this purpose, we next choose the system parameters as follows

$$M = 3,\ N = 2,\ \alpha = 1000,\ \beta = 370,$$

$$\lambda_{ij} = \lambda(i+1)^2(j+1)^2,\ \text{for } 0 \le i \le M,\ 0 \le j \le N,$$

$$\mu_{ij} = \mu i^2(j+1)^2,\ \text{for } 1 \le i \le M,\ 0 \le j \le N,$$

$$\nu_{ij} = \nu(i+1)^2 j^2,\ \text{for } 0 \le i \le M,\ 1 \le j \le N, \tag{6.2}$$

$$p_0 = 0.35,\ p_1 = 0.45,\ p_2 = 0.55,\ p_3 = 0.0,$$

$$q_i = 1 - p_i,\ 0 \le i \le M-1,\ q_3 = 0.35,$$

$$h_i = 2(i+1)/M(M+1),\ \text{for } 0 \le i \le M-1,\ h_3 = 0,$$

where $(\lambda,\ \mu,\ \nu) = (60,\ 30,\ 25)$.

With the help of the general methodology summarized in Section 4, we may compute the infinitesimal generators $\{C_n;\ 0 \le n \le M+N\}$ and then the stationary probability vector $\mathbf{P} = (\mathbf{P}_0, ..., \mathbf{P}_{M+N})$. Once vector $\mathbf{P}$ is computed, it is easy to derive the expectations Q and O, and many other probabilistic descriptors.

h	Q	O	$T(h)$
0.41	1.6397	1.6200	2239.1280
0.42	1.6419	1.6143	2239.1370
0.43	1.6440	1.6085	2239.1430
0.44	1.6441	1.6028	2239.1460
0.45	1.6482	1.5972	2239.1460
0.46	1.6503	1.5915	2239.1430
0.47	1.6524	1.5858	2239.1370
0.48	1.6545	1.5802	2239.1290
0.49	1.6565	1.5745	2239.1180
0.50	1.6586	1.5689	2239.1050

Table 1. Q, O and $T(h)$, for $h \in [0.41,\ 0.5]$

In Table 1, the value of Q, O and $T(h)$ are given for various values of h. In addition, in Figure 1, we plot the total profit $T(h)$ versus h. It is concluded that, for the set of parameters under consideration, $T(h)$ attains its maximum in the interval $[0.44,\ 0.45]$.

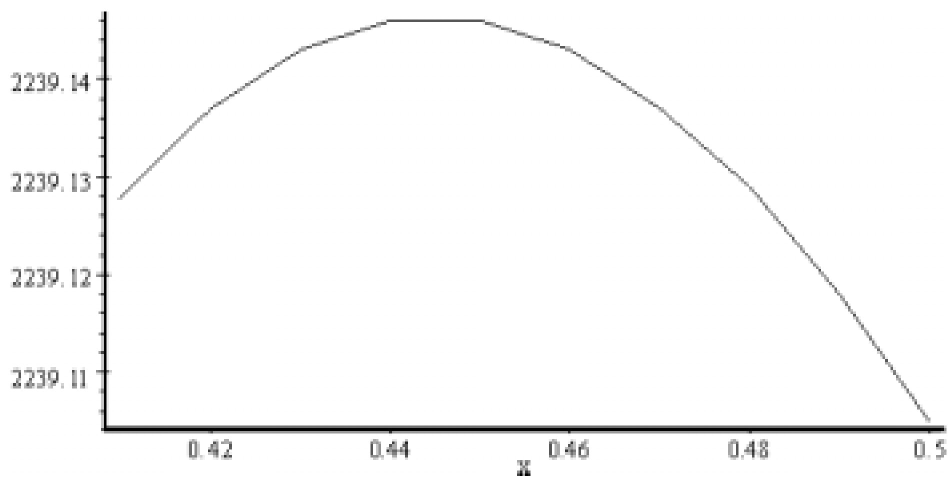

Fig. 1: The total profit T(h) *versus* h.

Acknowledgements

The work of J.R. Artalejo is supported by the European Commission under INTAS 96-0828 and by the DGICYT PB98-0837.

References

[1] ARTALEJO, J.R., "A queueing system with returning customers and waiting line", *Operations Research Letters,* 17 (1995), 191-199.

[2] ARTALEJO, J.R. and A. GOMEZ-CORRAL, "Steady state solution of a single-server queue with linear repeated request", *Journal of Applied Probability,* 34 (1997), 223-233.

[3] CHOI, B.D. and K.K. PARK, "The M/G/1 retrial queue with Bernoulli schedule", *Queueing Systems,* 7 (1990), 219-228.

[4] FALIN, G.I. and J.R. ARTALEJO, "Approximations for multiserver queues with balking/retrial discipline", *OR Spektrum,* 17 (1995), 239-244.

[5] FALIN, G.I. and J.R. ARTALEJO, "A finite source retrial queue", *European Journal of Operational Research,* 108 (1998), 409-424.

[6] FALIN, G.I., J.R. ARTALEJO and M. MARTIN, "On the single server retrial queue with priority customers", *Queueing Systems,* 14 (1993), 439-455.

[7] FALIN, G.I. and J.G.C. TEMPLETON, *Retrial Queues,* Chapman and Hall, London , 1997.

[8] FAYOLLE, G., "A simple telephone exchange with delayed feedbacks", in: *Teletraffic Analysis and Computer Performance Evaluation*, eds. O.J. Boxma, J.W. Cohen and H.C. Tijms, Elsevier.

[9] GAVER, D.P., P.A. JACOBS and G. LATOUCHE, "Finite birth-and-death models in randomly changing environments", *Advances in Applied Probability,* 16 (1984), 715-731.

[10] HAJEK, B., "Birth-and-death processes on the integers with phases and general boundaries", *Journal of Applied Probability,* 19 (1982), 488-499.

[11] KORNYSHEV, Y.N., "Design of a fully accessible switching system with repeated calls", *Telecommunications,* 23 (1969), 46-52.

[12] LATOUCHE, G. and V. RAMASWANI, "Expected passage times in homogeneous quasi-birth-and-death processes", *Stochastic Models,* 11 (1995), 103-122.

[13] LI, S-Q. and H-D. SHENG, "Generalized folding-algorithm for sojourn time analysis of finite QBD processes and its queueing applications", *Stochastic Models,* 12 (1996), 507-522.

[14] NEUTS, M.F., *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach,* The John Hopkins University Press, 1981.

[15] PEARCE, C.E.M., "On the problem of re-attempted calls in teletraffic", *Stochastic Models,* 3 (1987), 393-407.

[16] STEPANOV, S.N., "Markov models with retrials: the calculation of stationary performance measures based on the concept of truncation", *Mathematical and Computer Modelling,* (to appear).

[17] YE, J. and S-Q. LI, "Analysis of multi-media traffic queues with finite buffer and overload control - PART I: algorithm", *Proceedings IEEE Infocom'91,* Bal Harbour,

1464-1474.