ISSN 0005-1179, Automation and Remote Control, 2006, Vol. 67, No. 2, pp. 173–220. © Pleiades Publishing, Inc., 2006.
Original Russian Text © V.M. Vishnevskii, O.V. Semenova, 2006, published in Avtomatika i Telemekhanika, 2006, No. 2, pp. 3–56.

REVIEWS

Mathematical Methods to Study the Polling Systems

V. M. Vishnevskii and O. V. Semenova

Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, Russia

Received September 12, 2005

Abstract—Reviewed were the mathematical methods that are used to investigate the polling systems which found wide application in modeling and design of various transport and industrial processes. Emphasis was made on the models of polling systems used to investigate the wireless broadband networks. The polling systems were classified; presented were stochastic models and methods of investigating discrete-time and continuous-time systems, systems with cyclic, periodic, and random queue polling, as well as the methods of their optimization.

PACS number: 02.50.Fz

DOI: 10.1134/S0005117906020019

1. INTRODUCTION

The models of polling systems whose study dates from the late 1950's found wide use in the public health systems, air and railway transportation, and communication systems. The number of works on the polling systems is quite large. More than 700 papers, conference proceedings, theses, and reports published before 1996 are listed on the H. Takagi's site `http://www.sk.tsukuba.ac.jp/~takagi/polling.html`. The studies of the polling system models that were carried out before 1994 are presented rather well in [58, 87, 232, 234, 236–238] which classified the polling systems and systematized the theoretical results obtained. The review of H. Levy and M. Sidi [192] is oriented to the readership interested in polling system applications. Analytical methods of investigating the queuing systems with multiple waiting lines are reviewed in [4, 25]. Applications of the polling systems to analysis of the telecommunication systems controlled by the *ATM* and *GigaEthernet* protocols are described in [82, 139, 160, 192, 235]. Vigorous development of the broadband wireless information transmission networks provoked interest in the models of polling systems [3, 6, 46, 271–273]. The polling models for investigating the characteristics of personal and local wireless networks are analyzed in [83, 170, 202, 212, 264]; those intended for the regional wireless broadband regional networks with centralized control, in [2, 5, 13, 265, 266]; those for the satellite communication systems, in [88]. Application of the Petri networks to analysis of the polling systems is described in [27, 152].

The present paper is aimed at reviewing the publications on the polling systems that appeared in the international journals after 1990, as well as the most important publications of the national researchers. Section 2 presents the basic definitions and describes the main parameters defining the polling system. The following Sections 3–7 consider the works on the discrete one-server polling systems. Section 3 reviews the discrete-time models, and Sections 4–6 consider the continuous-time models. Section 5 discusses the issues of optimization of the polling systems. Section 6 reviews models of the polling systems with multiple servers. The works on the polling networks are listed in Section 7. Section 8 is devoted to the models of continuous-time polling systems.

2. CLASSIFICATION OF THE POLLING SYSTEMS

The polling systems are varieties of the queuing systems with multiple queues which are divided into two classes. The systems of the first class have multiple servers, and the customers arriving to the system choose a preferable server. In the systems of the second class (*polling systems*), there is one (or more) server(s) which are common to all queues; they poll the queues and serve the queued customers.

Depending on the number of queues in the system, the polling systems may be *discrete* (the number of waiting places is finite or countable) or *continuous* (the number of waiting places is more than countable). In the latter case consideration is given to the systems where the customers are placed on a circle or n-dimensional domain.

Discrete polling systems are characterized by the number of queues, their capacity (the number of the waiting places), number of servers, processes of customer arrival and service, durations of server switchover between the queues, as well as the order and discipline of queue service. We assume that all queues are numerated from 1 to N, where $N \geq 2$ is the number of queues in the system. The queue with the number $i = \overline{1, N}$ will be denoted by Q_i.

By the *polling order* or visit order is meant the rule used by the server to choose the next queue. The polling order can be both static and dynamic. With the *static* order, the rule of choosing queues remains invariable over the entire course of system operation. With the *dynamic* order, the queue is chosen for service at certain decision-making instants on the basis of complete or partial information about system state.

Among the kinds of *static order*, specified are

(1) *Cyclic* order where the server polls the queues in the order Q_1, Q_2, $\ldots$, Q_N, Q_1, Q_2, $\ldots$, Q_N, $\ldots$. These polling systems are called the *cyclic* systems.

(2) *Periodic* order where the server polls the queues in the order $Q_{T(1)}, Q_{T(2)}, \ldots, Q_{T(M)}, Q_{T(1)}, Q_{T(2)}, \ldots, Q_{T(M)}, \ldots$ which is characterized by the so-called polling table $(T(1), T(2), \ldots, T(M))$ of length M $(M \geq N)$, $T(i) \in \{1, \ldots, N\}$, $i = \overline{1, M}$. It is assumed that the polling table comprises the numbers of all system queues.

(3) *Random* order where the queue Q_i is taken for service with the probability p_i, $i = \overline{1, N}$, $\sum_{i=1}^{N} p_i = 1$. Feasible is another variant of choosing the queue where after polling the queue Q_i the server switches over to Q_j with the probability p_{ij}, $i, j = \overline{1, N}$, $\sum_{j=1}^{N} p_{ij} = 1$, $i = \overline{1, N}$.

(4) *Priority* order where the system has queues of different priorities and some queue may be served only if all higher-priority queues have no customers.

Special cases of the periodic queue polling are represented by the star-type polling where the queues are served in the order Q_1, Q_2, Q_1, $Q_3, \ldots, Q_1$, Q_N), and the elevator-type polling where the queues are served in the order Q_1, $Q_2, \ldots, Q_{N-1}$, Q_N, Q_N, $Q_{N-1}, \ldots, Q_2$, Q_1.

Time periods called *cycles* are specified in the activity of the cyclic or periodic polling system. For the cyclic polling systems, by the cycle is meant the time required for the server to serve the queues from Q_1 to Q_N. For the periodic polling systems, by the cycle is meant the time required to serve queues from $Q_{T(1)}$ to $Q_{T(M)}$. In operation of some polling systems, the *Hamiltonian cycle*, that is, the time over which the server treats all queues only once, is specified.

By the *queue service discipline* is meant the number of customers treated by the server in one polling. Within the queue, the customers are served in the order defined by the *customer service discipline* which most frequently lies in serving them in the arrival order. The queue service disciplines may be *deterministic* and *random*.

For the *deterministic* discipline, the maximum number of customers treated by the server[1] in

[1] The expressions *server treats at most l customers* or *l customers may be treated* imply that either the server treats l customers or the queue is emptied, whereupon the server switches to another queue.

one visit to the queue is constant. Among the deterministic queue (say Q_i) service disciplines, the following disciplines are specified:

(1) *Exhaustive*, where the server treats customers until the queue is emptied.

(2) *Gated*, where the server treats only those customers that sojourned in the queue at polling instant. If the server treats only those customers which sojourned in the queue by the beginning of the cycle, this discipline is called the *globally-gated* discipline.

(3) *l_i-limited*, where the number of customers that can be treated by the server is limited by l_i, $l_i \geq 1$.

(4) *l_i-decrementing*, where the server treats queued customers until the queue length is decremented by l_i as compared with the polling instant, $l_i \geq 1$.

(5) A discipline where the time of server sojourn in the queue is limited.

For the *random* discipline, the number of customers that can be treated by the server in the queue Q_i is defined by the value of the discrete random variable ξ_i with the distribution law $\{a^i_j, j \geq 1\}$ which can vary with each visit to the queue. We mention some of the random disciplines:

(1) *Binomial* discipline with the random variable ξ_i having the binomial distribution with the parameters X_i and p_i, where X_i is the number of customers queued in Q_i at the polling instant and p_i is some number, $0 < p_i \leq 1$. For this discipline, $a_i^j = \mathrm{C}^j_{X_i} p_i^j (1 - p_i)^{X_i - j}$, $j = \overline{1, X_i}$, $a_i^j = 0$ for $j > X_i$.

(2) *Bernoulli* discipline where the first customer queued in Q_i is served with the probability 1 and each subsequent customer, with a given probability p_i. The server discharges the queue with the probability $1 - p_i$. For this discipline, $a_i^j = p^{j-1}$, $j \geq 1$.

Detailed classification of the service disciplines is given in [194] which also presents for various service disciplines the inequalities describing relations between the number of customers in the system.

If all queues of the polling system have identical service disciplines, we distinguish a polling system with the given (exhaustive, l-limited, or other) service discipline. If the service disciplines of the queue differ, then we distinguish a polling system with a *mixed* service discipline.

The order of queue polling and their service disciplines constitute the polling system *service policy*, that is, the rule for choosing the next customer from the queue connected to the server or from another queue.

Among the polling systems, we distinguish the *discrete-time* systems where time is divided into equal intervals called the slots and the *continuous-time* systems.

The polling system is called the *symmetric* or uniform system if the processes characterizing its queues (the processes of customer arrival and service, as well as those defining the durations of server switchovers between the queues) are stochastically equivalent. Otherwise, the system is called *nonsymmetric* or nonuniform system. If in the polling system the server needs no time for switching over between the queues, we can assert this it is a system with the zero switchover times; otherwise, we can state that it is a system with a nonzero switchover times.

We assume, except as otherwise noted, that the discrete polling system is nonsymmetric, the number of its queues is finite, the queues have an unlimited waiting space, and the server switchovers between the queues are nonzero. The server visits a queue, if empty, and then immediately switches over to another queue. We also assume that within the queue the customers are served in the order of their arrivals.

The intention of the majority of works on the polling systems is to determine the mean waiting time in each of the system queues. However, since it does not always happen that the explicit formulas to calculate these characteristics can be established, much attention is paid to determining approximate formulas and specifying the existing approximate values. Sometimes the problem of determining the mean waiting times comes to determining the weighted sum of these characteristics. By the weighted sum of the mean waiting times is meant the expression $\sum_{i=1}^{N}\rho_i\mathbf{M}W_i$, where $\mathbf{M}W_i$ is the mean waiting time in the queue Q_i, $\rho_i = \lambda_i b_i$ is the load of Q_i, λ_i is the customer flow rate, b_i is the mean time of queue service in Q_i, $i = \overline{1,N}$. The value $\rho = \sum_{i=1}^{N}\rho_i$ is called the *system load*.

3. DISCRETE-TIME POLLING SYSTEMS

The following models will be classified in terms of queue polling—cyclic, periodic, random, or priority. We consider separately the two-queue polling systems which are discussed in [106, 229]. The times of customer service are equal to one slot. The system examined in [106] has a correlated customer flow obeying the probabilities $\{a_{ij},\ i,j \geq 0\}$, that i customers arrive to the queue Q_1 during one slot and j customers, to the queue Q_2, $i,j \geq 0$. The queues are served exhaustively. For the given model, a system of linear algebraic equations of the mean waiting times was obtained. In the system of [229], the queue Q_1 has a 1-limited service and the number of customers treated in the queue Q_2 depends on the number of customers remaining in Q_1, as well as on the number of customers in Q_2 at the polling instant.

Systems with cyclic polling and Bernoulli customer flows in queues were studied in [14, 15, 240]. In the system considered in [14, 15], the server treats one customer from each queue (1-limited discipline). The study relied on the principle of model decomposition into N single-server queuing systems, which enabled determination of the cycle distribution function and the stationary distribution of the probabilities of the number of customers in the system. The polling system each of whose queues gets P priority Bernoulli flows was considered in [240]. The exhaustive, gated, 1-limited, and 1-decrementing queue services were examined. This model is a discrete counterpart of the systems examined in [129, 223]. For a fixed priority class, an expression of the weighted sum of the mean times of waiting in queues was obtained.

The cyclic polling system with exhaustive queue service and the zero server switchover was considered in [191] relying on the analysis [166] of the corresponding nonzero-switchover system whose results we present in more detail. The time is divided into equal-length intervals $\{[t,t+1),\ t = 0,1,2,\ldots\}$, slots. The input customer flow is characterized by a set of collectionwise independent random variables $\{X_i(t), t = 0,1,2,\ldots\}$, where $X_i(t)$ is the number of customers arriving over the interval $[t,t+1)$ to the queue Q_i. For each i, the random variables $X_i(t)$, $t \geq 0$, are distributed identically with the expectation μ_i and variance σ_i^2, $i = \overline{1,N}$. The customer service time is equal to one slot.

The duration of switchover from Q_i to Q_{i+1} within the interval $[t,t+1)$ is defined by the value of the random variable $S_i(t)$, $t = 0,1,\ldots$. The random variables $S_i(t)$, $t = 0,1,\ldots$, are independent and distributed identically with the expectation r_i and variance δ_i^2, $i = \overline{1,N}$.

The approach of Takagi [234] is used to determine the mean waiting time. The first moments and the matrix of covariance of the number of queued customers at the instants of polling are calculated. Let $L_i(t)$ be the number of customers in Q_i over the interval $[t,t+1)$ and $\tau_i(m)$ be the instant of the mth connection of the server to Q_i, $i = \overline{1,N}$. The generating function of the number of customers in the system $F_i(z_1,\ldots,z_n) = \mathbf{M}\left(\prod_{j=1}^{N} z_j^{L_j(\tau_i(m))}\right)$ is introduced,

$$f_i(j) = \mathbf{M}(L_j[\tau_i(m)]) = \left.\frac{\partial F_i(z_1,\ldots,z_n)}{\partial z_j}\right|_{z_1=\ldots=z_n=1},$$

$$f_i(j,k) = \mathbf{M}[L_j(\tau_i(m))L_k(\tau_i(m))] = \left.\frac{\partial^2 F_i(z_1,\ldots,z_n)}{\partial z_j \partial z_k}\right|_{z_1=\ldots=z_n=1}, \quad i,j,k = \overline{1,N}.$$

The expressions of $f_i(j)$, $i,j=\overline{1,N}$, were obtained in the explicit form:

$$f_i(i)=\frac{\mu_i(1-\mu_i)\sum\limits_{k=1}^{N} r_k}{1-\sum\limits_{k=1}^{N}\mu_k}, \tag{1}$$

$$f_i(j)=\mu_j\left(\sum_{k=j}^{i-1} r_k+\frac{\sum\limits_{k=j+1}^{i-1}\mu_k\sum\limits_{k=1}^{N} r_k}{1-\sum\limits_{k=1}^{N}\mu_k}\right),\quad j\neq i,\ i,j=\overline{1,N}.$$

The values of $f_i(j,k)$, $i,j,k=\overline{1,N}$, are established as a solution of the system of linear equations

$$\begin{aligned}
f_{i+1}(j,k)&=\mu_j\mu_k(\delta_i^2+r_i^2)+r_i\mu_k f_i(j)+r_i\mu_j f_i(k)\\
&+\frac{f_i(i,j)\mu_k+f_i(i,k)\mu_j}{1-\mu_i}+f_i(j,k)+f_i(i)\mu_j\mu_k\\
&+\left(\frac{2r_i}{1-\mu_i}+\frac{1}{(1-\mu_i)^2}+\frac{\sigma_i^2}{(1-\mu_i)^3}\right)+\frac{f_i(i,i)\mu_j\mu_k}{(1-\mu_i)^2},\quad i\neq j, i\neq k, j\neq k,\\
f_{i+1}(j,j)&=\mu_j^2(\delta_i^2+r_i^2)+r_i(\sigma_j^2-m_j)+2r_i\mu_j f_i(j)+f_i(j,j)+\frac{2f_i(i,j)\mu_j}{1-\mu_i}\\
&+\frac{f_i(i,i)\mu_j^2}{(1-\mu_j)^2}+f_i(i)\left\{\frac{\sigma_j^2-\mu_j}{1-\mu_j}+\mu_j^2\left(\frac{2r_i}{1-\mu_i}+\frac{1}{(1-\mu_i)^2}+\frac{\sigma_i^2}{(1-\mu_i)^3}\right)\right\},\ i\neq j,\\
f_{i+1}(i,k)&=\mu_i\mu_k(\delta_i^2+r_i^2)+r_i\mu_i\left(f_i(k)+\frac{f_i(i)\mu_k}{1-\mu_i}\right),\quad i\neq k,\\
f_{i+1}(i,i)&=\mu_i^2(\delta_i^2+r_i^2)+r_i(\sigma_i^2-\mu_i),\quad i,j,k=\overline{1,N}.
\end{aligned} \tag{2}$$

For the nonsymmetric system, the mean waiting time is as follows [234]:

$$\mathbf{M}W_i=\frac{f_i(i,i)+f_i(i)}{2\mu_i f_i(i)}+\frac{\sigma_i^2}{2\mu_i}\left(\frac{1}{1-\mu_i}-\frac{1}{\mu_i}\right),\quad i=\overline{1,N}, \tag{3}$$

and for the symmetric polling system,

$$\mathbf{M}W=\frac{\delta^2}{2r}+\frac{\sigma^2}{2\mu(1-N\mu)}+\frac{Nr(1-\mu)}{2(1-N\mu)}.$$

Now we pass to the results of [191]. In the model of the polling system with the zero switchover times it is assumed that as soon as the system is emptied (let it be at the slot t) the server chooses a queue with the probability $p=1/N$ and connects to it at $t+1$. If this queue is empty but other queues have customers, then the server circles the remaining queues beginning from the given one. If at the time instant $t+1$ the system remains empty, then the server repeats the procedure of choosing a queue for service. It was suggested in [191] to study this system by means of a model with almost zero server switchover which is defined as follows:

(1) When at time t the server ends serving some queue, at the same instant it is ready to start service of the next (on the circle) nonempty queue.

(2) If at time t the system is empty, then at time $t+1$ the server is ready to start service in the queue which is chosen as described above.

Therefore, $\mathbf{P}(S_i(t)=1)=p$, $\mathbf{P}(S_i(t)=0)=1-p=q$, whence $r_i=p$, $\delta_i^2=p(1-p)$, $i=\overline{1,N}$. By substituting the equalities for r_i,δ_i^2 in the equalities (1)–(3) and passing to the limit for $p\to 0$, one gets the mean waiting time in the system with the zero server switchover between the queues. In particular, for the symmetric system

$\mathbf{M}W = \dfrac{1}{2} + \dfrac{\sigma^2}{2\mu(1-N\mu)}$ for the exhaustive service,

$\mathbf{M}W = \dfrac{\delta^2}{2r} + \dfrac{\sigma^2}{2\mu(1-N\mu)} + \dfrac{Nr(1+\mu)}{2(1-N\mu)}$ for the gated service, and

$\mathbf{M}W = \dfrac{\delta^2}{2r} + \dfrac{\sigma^2(1+Nr)}{2\mu(1-(1+r)N\mu)} + \dfrac{N\delta^2\mu}{2(1-N\mu-Nr\mu)}$ for the 1-limited service.

The system with periodic polling was considered in [69]. The customer flow to the queue Q_n is a group flow defined by the set of random variables $\{x_j(n), j \geq 1\}$ and $\{p_i, i \geq 1\}$, where $x_j(n)$ is the number of groups arriving in the jth slot, p_i is the probability that the group consists of i customers, $\sum_{i=1}^{\infty} p_i = 1$, $n = \overline{1,N}$. Consideration was given to the exhaustive, gated, and 1-limited service of queues. Stability conditions were established, as well as an expression of the weighted sum of the mean waiting times.

The polling system with the zero server switchover and arbitrary polling order was considered in [246]. The inflow to the queue Q_n is characterized by the set of random variables $\{x_j(n), j \geq 1\}$, where $x_j(n)$ is the number of customers arriving within the jth slot, $n = \overline{1,N}$. The end of customer service in Q_n in the jth slot is defined by the state of the random variable $m_j(n)$ with the state space $\{0,1\}$ which changes its state only at the end of the slot. If $m_j(n) = 1$, then the server completes service; otherwise, service is continued. The queue must be connected to the server in order to be treated. Connection of the queue Q_n to the server is of random nature, and in the jth slot it is defined by the value of the random variable $c_j(n)$, $j \geq 1$: $c_j(n) = 0$ if the queue is connected to the server, and $c_j(n) = 1$, otherwise. The notion of system stabilizability was introduced: the system is referred to as stabilizable if there exists a queue service policy making the system *stable*, that is, the N-dimensional Markov chain describing the number of queued customers is irreducible and has a final probability distribution. The following system stabilizability criterion was established:

$$\sum_{n\in Q} \frac{\mathbf{M}(x_j(n))}{\mathbf{M}(m_j(n))} < 1 - \prod_{n\in Q}(1-\mathbf{M}(c_j(n))), \quad \text{for any} \quad Q \subset \{1,\dots N\}.$$

Consideration was also given to the problem of determining the optimal service policy minimizing the mean system waiting time.

4. CONTINUOUS-TIME POLLING SYSTEMS

4.1. Description of the Model

The basic model which is the subject of study of the majority of publications is as follows. The system has one server and N ($N \geq 2$) queues with an unlimited waiting space. A Poisson flow of customers with the parameter λ_i arrives to the queue Q_i where the customer service times are independent and identically distributed with the distribution function $B_i(t)$ with the mean $b_i = \int_0^{\infty} t dB_i(t)$, second moment $b_i^{(2)}$, and the Laplace–Stieltjes transform (LST) $\widetilde{B}_i(s)$, $i = \overline{1,N}$. We assume that the customer flows and the customer service times are independent. According to the Kendall classification, this system is called the $M/GI/1$-type polling system. If the times of customer service are distributed exponentially or the customer flows in queues are recurrent, then we deal with the $M/M/1$-type and $G/G/1$-type polling systems, respectively.

The server visits queues according to a certain order of polling and treats them according to the chosen service discipline. The time of passing from the queue Q_i to the queue Q_j, which is called the switchover time, has the distribution function $R_{ij}(t)$ with the mean r_{ij}, second moment $r_{ij}^{(2)}$, and the LST $\widetilde{R}_{ij}(s)$, $i, j = \overline{1, N}$. If the polling order is cyclic, then we omit the second subscript j. We denote by $\rho_i = \lambda_i b_i$ the load of the queue Q_i, by $\rho = \sum_{i=1}^{N} \rho_i$ the system load, and by r and $r^{(2)}$ the first and second moments of the total duration of server switchovers in one slot for the systems with cyclic or periodic polling.

We assume, except as otherwise noted, that the system has $M/GI/1$-type queues. The following subsection presents the results of investigating the two-queue polling systems. Subsections 4.3–4.6 present the respective research results for models with cyclic, periodic, random, and priority orders of queue polling by the server.

4.2. Two-queue Polling Systems

A two-queue system with the zero server switchover was considered in [79]. The first queue is of the $M/D/1$ type, and the second queue, of the $M/GI/1$ type. The service discipline is exhaustive. The LST of the distribution function of the system busy period was obtained.

Two symmetric $M/GI/1$-type polling systems were compared in [98]. In these models, the server spends time not only on inter-queue switchovers but also on warming up, that is, getting ready for queue service. In the first model, the server visits a queue if it is empty. In the second model, the server stops at the current queue if the other queue is empty. For these models, the mean waiting times were determined and compared for different parameters of the times of switchovers between the queues and warming-up times in order to verify whether they are constants or random variables.

A system with mixed service of queues was examined in [150]. One queue is served exhaustively, the other, by the 1-limited service discipline. We dwell in more detail on the results of this work. The system parameters were described in the last subsection. The study of this model relies on the scheme described in [151]. We first present the results concerning the mean cycle time c which is as follows:

$$c = s_1 + s_2 + r_1 + r_2, \tag{4}$$

where s_i is the mean time of service of Q_i in one cycle for which the equality $s_i = p_i b_i$ is valid, p_i is the probability that at an arbitrary instant a customer from Q_i, $i = \overline{1, 2}$, is served. Determination of the probability p_2 is based on the fact that the mean number of customers queued in Q_2 in one cycle is $\lambda_2 c$ and the mean number of customers served in Q_2 in one cycle is p_2, whence we get that $p_2 = \lambda_2 c$ and, as a result, $s_2 = \lambda_2 b_2 c$. Now, we determine the value of s_1 by noting that it is equal to the sum $\lambda_1 v_1$ of the busy periods of the $M/GI/1$ queuing system corresponding to the queue Q_1, where v_1 is the mean intervisit time for the queue Q_1. It is well-known that the mean duration of the busy period of the system $M/GI/1$ corresponding to the queue Q_1 obeys $g_1 = \frac{b_1}{1-\lambda_1 b_1}$, whence $s_1 = \lambda_1 v_1 g_1 = \frac{\lambda_1 v_1 b_1}{1-\lambda_1 b_1}$.

The mean time of visits to the queue Q_i follows $v_i = s_{3-i} + r_1 + r_2$; therefore, $s_1 = \frac{\rho_1(s_2+r_1+r_2)}{1-\rho_1} = \frac{\rho_1(p_2 c+r_1+r_2)}{1-\rho_1}$, where $\rho_i = \lambda_i b_i$. It follows from (4) that $c = \frac{r_1+r_2}{1-\rho}$; consequently, $p_2 = \frac{\lambda_2(r_1+r_2)}{1-\rho}$, $s_i = \frac{\rho_i(r_1+r_2)}{1-\rho}$, $v_i = \frac{(r_1+r_2)(1-\rho_i)}{1-\rho}$, where $\rho = \rho_1 + \rho_2$.

To determine the mean waiting time, the method of tagged customer is used. The queue to which such a customer arrives is called the tagged queue. We denote by W_i the waiting time of the tagged customer arriving to the queue Q_i. At the instant of its arrival, the server can be in one of the following states:

(1) Switchover to the queue Q_1 (with the probability p_{r1}).

(2) Switchover to the queue Q_2 (with the probability p_{r2}).

(3) Service of the queue Q_1 (with the probability p_{b1}).

(4) Service of the queue Q_2 (with the probability p_{b2}).

The mean waiting time of a customer queued in Q_1 which has l customers obeys the equality

$$\mathbf{M}[W_1|L_1 = l] = lb_1 + p_{r1}\left\{\frac{r_1^{(2)}}{2r_1}\right\} + p_{b2}\left\{\frac{b_1^{(2)}}{2b_1}\right\} + p_{r2}\left(\frac{r_2^{(2)}}{2r_2} + x_{21}\right) + p_{b2}\left(\frac{b_2^{(2)}}{2b_2} + r_1\right),$$

where L_1 is the number of customers in the system, x_{ij} is the mean time from the instant of polling the queue Q_i till the instant when the server leaves Q_j, provided that the tagged customer arrived to Q_j at the instant where service went on in Q_i, $i \neq j$, $i, j = 1, 2$.

The unconditional mean expectation in the queue Q_1 obeys the equality

$$\mathbf{M}[W_1] = \mathbf{M}[L_1]b_1 + p_{r1}\left\{\frac{r_1^{(2)}}{2r_1}\right\} + p_{b2}\left\{\frac{b_1^{(2)}}{2b_1}\right\} + p_{r2}\left(\frac{r_2^{(2)}}{2r_2} + x_{21}\right) + p_{b2}\left(\frac{b_2^{(2)}}{2b_2} + r_1\right).$$

By the Little theorem, we get $\mathbf{M}(L_1) = \lambda_1\mathbf{M}(W_1)$. Since $p_{ri} = r_i/c = r_i(1-\rho)/(r_1+r_2)$, $p_{bi} = s_i/c = \rho_i$, we get

$$\mathbf{M}(W_1) = \frac{(1-\rho)(r_1^{(2)} + r_2^{(2)})}{2(r_1+r_2)(1-\rho_1)} + \frac{r_2(1-\rho)x_{21}}{(r_1+r_2)(1-\rho_1)} + \frac{\rho_2 r_1}{1-\rho_1} + \frac{\lambda_1 b_1^{(2)} + \lambda_2 b_2^{(2)}}{2(1-\rho_1)}.$$

To calculate x_{ij}, we introduce the value q_{ij} denoting the probability that at the period of server connection to Q_j this queue was not empty, provided that the tagged customer arrived to Q_j at the instant of serving Q_i, $i \neq j$, $i, j = \overline{1,2}$. The total duration of this period in one cycle on the average is equal to $r_j^{(2)}/r_j$. The equality $x_{21} = r_1 + q_{21}b_2$ is valid. We also denote by c_{21} the mean cycle time for the polling system with the mean time of connection to the queue Q_2 equal to $r_2^{(2)}/r_2$, $c_{21} = r_2^{(2)}/r_2 + r_1 + q_{21}b_2 + q_{11}s_1$; then, $q_{21} = \lambda_2 c_{21}$. The mean number of customers arriving to Q_1 in time c_{21} is equal to $\lambda_1 c_{21}$. Since the mean time required to serve these customers is $q_{11}s_1$, the equality $q_{11}s_1 = \lambda_1 c_{21} b_1$ is valid, whence

$$q_{21} = \frac{\lambda_2(r_2^{(2)} + r_1 r_2)}{r_2(1-\rho)}, \quad x_{21} = \frac{\rho_2 r_2^{(2)} + r_1 r_2(1-\rho_1)}{r_2(1-\rho)}.$$

We, therefore, have a formula for the mean waiting time in the queue Q_1:

$$\mathbf{M}(W_1) = \frac{(1-\rho)r_1^{(2)} + (1-\rho_1+\rho_2)r_2^{(2)}}{2(r_1+r_2)(1-\rho_1)} + \frac{\lambda_1 b_1^{(2)} + \lambda_2 b_2^{(2)}}{2(1-\rho_1)} + \frac{\rho_2 r_1}{1-\rho_1} + \frac{r_1 r_2}{r}.$$

The equality for the mean waiting time in the queue Q_2 can be obtained from the expression for the weighted sum of the mean waiting times, the so-called *pseudoconservation law* [68, 75]:

$$\rho\mathbf{M}(W_1) + \rho\left(1 - \frac{\lambda_2(r_1+r_2)}{1-\rho}\right)\mathbf{M}(W_2)$$
$$= \rho\frac{(\lambda_1 b_1^{(2)} + \lambda_2 b_2^{(2)})}{2(1-\rho)} + \frac{\rho r^{(2)}}{2(r_1+r_2)} + \frac{r_1+r_2}{2(1-\rho)}(\rho^2 - \rho_1^2 + \rho_2^2).$$

The system with correlated customer flows was considered in [224]. Pairs of customers arrive to the system in addition to the basic flows of customers. One customer from a pair is sent to each queue. Consideration was given to the exhaustive and gated disciplines of queue service. For this system, the mean numbers of customers in each queue at the time of server connection it, the stationary distribution of the number of customers served in each queue in one cycle, the mean number of customers in each queue at the times of service completion, and the LST's of the distribution functions of the service waiting time in each queue were obtained.

The system with k_i-limited service of queues was studied in [90]. It was shown that under a heavy traffic (for $\rho \to \infty$) the queue lengths can grow according to either of the two scenarios:

(1) the length of only one queue is increased, the length of the second queue remaining small;

(2) the lengths of both queues increase.

Conditions for the system parameters under which the number of queued customers varies according to a certain scenario and also the formulas for approximate calculation of the probability distribution of the number of queued customers were obtained.

In [76] analysis was carried out for the two-queue $M/GI/1$-type polling system which has the following distinction. If at the instant of polling the queue Q_1 is empty, then the server waits for a customer during a random time. If at the end of this time no customer arrives, then the server switches to the queue Q_2. The following queue service disciplines were examined: one of the queues is served exhaustively, in the other queue only one customer is served; both queues are served exhaustively. The LST of the waiting time distribution function and the expression of the weighted sum of the mean waiting times were obtained. Consideration was given in [105] to a similar model without customer waiting by the server. At that, it is assumed that at least one of the random variables defining the durations of customer service or inter-queue switchover has the heavy-tail distribution with the parameter ν $(1 < \nu < 2)$. Approximate formulas for calculation of the mean waiting times were obtained:

$$1 - W_1(t) \sim \frac{1}{\nu - 1}\left(\frac{\lambda_1 b_1 + \lambda_2 b_2}{1 - \rho_1} + \frac{(1-\rho)(s_1 + s_2)}{(1-\rho_1)\sigma}\right) t^{\nu-1} L(t), \quad t \to \infty,$$

$$1 - W_2(t) \sim \frac{1}{\nu - 1}\left(\frac{\lambda_1 b_1 + \lambda_2 (s_1 + s_2 + b_2)}{(1-\rho_1)^{\nu-1}(1 - \rho - \lambda_2\sigma)} + \frac{s_1 + s_2}{(1-\rho_1)^{\nu-1}\sigma}\right) t^{\nu-1} L(t), \quad t \to \infty,$$

where λ_i is the intensity of the customer flow in Q_i, $\rho_i = \lambda\beta_i$, β_i is the time of service in Q_i, σ_i is the mean time switchover from Q_i to Q_{i+1}, $i = \overline{1,2}$, $\sigma = \sigma_1 + \sigma_2$, $\rho = \rho_1 + \rho_2$, the values of b_i and s_i, $i = \overline{1,2}$, being defined as follows:

$$1 - B_i(t) = [b_i + o(1)]t^{-\nu} L(t), \quad 1 - S_i(t) = [s_i + o(1)]t^{-\nu} L(t), \quad t \to \infty,$$

where $L(t)$ is an arbitrary function with $\lim\limits_{t\to\infty} L(at)/L(t) = 1$ for any $a > 0$.

Similar results were obtained in [71] for the $M/M/1$-type polling system. The case where the server interrupts service in Q_2 if the length of Q_1 exceeds L was considered.

The polling system with the Bernoulli service discipline was studied in [121]. The stationary distribution of the probabilities of system states at the instants of service completion, the LST's of the distribution functions of waiting times, and also mean waiting times were obtained. Similar results were established in [267] for the system where one queue is served exhaustively and the second queue has a Bernoulli service discipline.

A system with the zero server switchover was studied in [181]. The queues are served by the threshold service discipline defined as follows. Given is a number L. If at the instant of service completion the number of customers queued in Q_1 exceeds L, then the server switches over to Q_1. Otherwise, the server continues service of the current queue until its exhaustion. The probability generating function of the number of customers in the system at the instants of service completion was obtained. These results were extended in [67] to the case of nonzero server switchover between the stations, and the stability conditions, approximate formulas for calculation of the mean queue lengths at the instants of service completion, and expressions for the weighted sum of the mean waiting times were obtained in addition.

The polling system with mixed service and the zero server switchover was investigated in [178]. One queue has the exhaustive service, the second queue, the 1-limited service. These results were extended in [209] to the $M/GI/1$-type polling systems where one queue is served exhaustively and the second queue has the k-limited service discipline. The polling system where Q_1 is served exhaustively and Q_2 has the decrementing service discipline was considered in [159]. The stationary probability distribution of the number of customers in the system at the instants of service, the mean waiting times, and the relations for the mean waiting times for different k were established.

The two-queue priority polling system was examined in [116]. The queue Q_1 is served exhaustively whereupon the server switches to Q_2. If at the time of serving a customer in the second queue a customer arrives to the first queue, then the server completes service only of those customers in the second queue which arrived to it before this instant. The mean waiting time and the mean fraction of the time of serving the queues Q_1 and Q_2 were established.

The system where the rule of queue service is characterized by two thresholds (M, N) $(0 \leq M < N)$ was discussed in [123]. Upon completion of customer service, the queue to be served is chosen depending on the relation between the number of customers in Q_2 and the thresholds, as well as the number of the queue to which the server is connected. For this system, the stability conditions, the stationary distribution of the probabilities of system states at the instants of service completion, as well as the LST of the distribution functions of the waiting times were obtained. Similar results were obtained in [124] for the $M/M/1$-type polling system with the zero server switchover and two thresholds (R_i, F_i) corresponding to each queue, where $R_1 \leq F_1 < R_2 \leq F_2$. This model was generalized in [122] to the case of multiple servers.

For the system with time-limited queue service, approximations of the mean waiting times were established in [250]. Some properties of the stationary state distribution of the quasi-birth-and-death process describing, in particular, behavior of the two-queue polling system with the discipline of taking for service the shortest queue were proved in [239].

The $MAP/M/1$-type polling systems were examined in [86]. The number of queue waiting places is limited. If both queues have customers, then the server treats concurrently one customer from each queue; if one of the queues is empty, then the server treats customers from the other queue in the usual way. Consideration was given to a Markov chain describing the number of queued customers, and some system characteristics such as the probability of losing a customer because of overfilling of the corresponding queue, the mean queue lengths, and the mean customer waiting time were obtained. The readers can find in [231] an application of the two-queue polling model to the description of car traffic.

4.3. Cyclic Polling

4.3.1. Service discipline: exhaustive, gated, or limited. The works devoted to the $M/GI/1$-type polling systems with the zero server switchover were reviewed in [84].

The necessary and sufficient stability conditions obey the inequalities:

$\rho < 1$ for the exhaustive and gated service,

$\rho + \min\limits_i \lambda_i r < 1$ for the 1-limited service, and

$\rho + \min\limits_i \lambda_i (1 - \rho_i) r < 1$ for the 1-decrementing service [64].

The main methods and results of investigating the cyclic systems with exhaustive, gated, and globally-gated queue service and also of some systems with noncyclic polling order were expounded in [269]. We present some results of this work.

Gated queue service. Let X_i^j be the number of queued customers in Q_j at the instant of polling the queue Q_i, $i,j=\overline{1,N}$,

$$G_i(\mathbf{z}) = G_i(z_1,z_2,\ldots,z_{i-1},z_i,z_{i+1},\ldots,z_N) = \mathbf{M}\left[\prod_{j=1}^{N} z_j^{X_i^j}\right], \quad i=\overline{1,N},$$

are the generating functions of X_i^j, $i,j=\overline{1,N}$.

The functions $G_i(\mathbf{z})$, $i=\overline{1,N}$, satisfy the relations

$$G_{i+1}(\mathbf{z}) = G_i\left(z_1,z_2,\ldots,z_{i-1},\widetilde{B}_i\left[\sum_{j=1}^{N}\lambda_j(1-z_j)\right],z_{i+1},\ldots,z_N\right)\widetilde{R}_i\left[\sum_{j=1}^{N}\lambda_j(1-z_j)\right], \ i=\overline{1,N}.$$

The mean number of customers $f_i(j)=\mathbf{M}(X_i^j)$ queued in Q_j at the instant of polling Q_i obeys

$$f_i(j)=\mathbf{M}(X_i^j)=\left.\frac{\partial G_i(\mathbf{z})}{\partial z_j}\right|_{\mathbf{z}=\mathbf{1}}.$$

The values $f_i(j)$, $i,j=\overline{1,N}$, satisfy the equation system

$$\begin{aligned} f_{i+1}(j) &= f_i(j)+\lambda_j b_i f_i(i)+\lambda_j r_i, \quad j\neq i,\\ f_{i+1}(i) &= \lambda_i b_i f_i(i)+\lambda_i r_i, \quad i=\overline{1,N}, \end{aligned}$$

whose solution can be obtained explicitly as

$$f_i(j)=\begin{cases} \lambda_j\left(\sum\limits_{k=j}^{i-1}\left[\rho_k\dfrac{r}{1-\rho}+r_k\right]\right), & j\neq i\\ \lambda_i\dfrac{r}{1-\rho}, & j=i. \end{cases}$$

The second moments of X_i^j, $i,j=\overline{1,N}$, can also be obtained by means of the generating functions $G_i(\mathbf{z})$, $i=\overline{1,N}$, as

$$\begin{aligned} f_i(j,k) &= \mathbf{M}(X_i^jX_i^k)=\left.\frac{\partial^2 G_i(\mathbf{z})}{\partial z_j\partial z_k}\right|_{\mathbf{z}=\mathbf{1}},\\ f_i(i,i) &= \mathbf{M}(X_i^i(X_i^i-1))=\left.\frac{\partial^2 G_i(\mathbf{z})}{\partial z_i^2}\right|_{\mathbf{z}=\mathbf{1}}. \end{aligned}$$

The equation system for $f_i(j,k)$, $i,j,k=\overline{1,N}$, can be found in [234]. The mean cycle time in this system is $c=\frac{r}{1-\rho}$. The mean waiting time in the queue Q_i obeys the inequality

$$\mathbf{M}(W_i)=\frac{1+\rho_i f_i(i,i)}{2\lambda_i^2 c}.$$

Exhaustive queue service. The generating function $G_i(\mathbf{z})$ of the number of customers in the system at the instant of polling the queue Q_i is as follows:

$$G_{i+1}(\mathbf{z}) = G_i\left(z_1,z_2,\ldots,z_{i-1},\widetilde{\theta}_i\left[\sum_{\substack{j=1\\ j\neq i}}^{N}\lambda_j(1-z_j)\right],z_{i+1},\ldots,z_N\right)\widetilde{R}_i\left[\sum_{j=1}^{N}\lambda_j(1-z_j)\right],$$

where $\widetilde{\theta}_i(s)$ is the LST of the busy period of the $M/GI/1$-type system corresponding to the queue Q_i, $i = \overline{1, N}$.

The expectations of the random variables X_i^j, $i, j = \overline{1, N}$, satisfy the inequalities

$$f_i(j) = \begin{cases} \lambda_j \left(\sum\limits_{k=j+1}^{i-1} \rho_k \dfrac{r}{1-\rho} + \sum\limits_{k=j}^{i-1} r_k \right), & j \neq i \\ \lambda_i (1 - \rho_i) \dfrac{r}{1-\rho}, & j = i. \end{cases}$$

The mean waiting time obeys the formula

$$\mathbf{M}(W_i) = \frac{\lambda_i b_i^{(2)}}{2(1-\rho_i)} + \frac{f_i(i,i)}{2\lambda_i^2 (1-\rho_i)c}, \quad i = \overline{1, N}.$$

For the given polling system, the expression for the weighted sum of the mean waiting times for the exhaustive queue service is as follows:

$$\sum_{i=1}^{N} \rho_i \mathbf{M}(W_i) = \rho \frac{\sum\limits_{i=1}^{N} \lambda_i b_i^{(2)}}{2(1-\rho)} + \rho \frac{r^{(2)}}{2r} + \frac{r}{2(1-\rho)\left[\rho^2 - \sum\limits_{i=1}^{N} \rho_i^2\right]},$$

and for the gated service it is as follows:

$$\sum_{i=1}^{N} \rho_i \mathbf{M}(W_i) = \rho \frac{\sum\limits_{i=1}^{N} \lambda_i b_i^{(2)}}{2(1-\rho)} + \rho \frac{r^{(2)}}{2r} + \frac{r}{2(1-\rho)\left[\rho^2 + \sum\limits_{i=1}^{N} \rho_i^2\right]}.$$

Determination of the mean waiting times in the queues of the polling system with exhaustive service is discussed in [66, 95, 96, 120, 146, 156, 180, 205, 228]. An algorithm for approximate calculation of these characteristics in the case of zero server switchover between queues was proposed in [95]. The process describing dynamics of the number of customers in the system under heavy traffic was shown to converge to the Bessel diffusion process. This algorithm was improved in [228] where two kinds of server behavior were considered: either the server connects to a queue irrespective of the fact that it is empty or not or the server does not visit empty queues. For the system with a nonzero switchover of the server which does not connect to an empty queue, these results were generalized in [156] where exact formulas for the two-queue system were obtained. For the $G/G/1$-type polling system with the zero server switchover, approximate formulas for the mean waiting times were obtained in [96]. The properties of the polling systems were expounded in [101], the system in critical mode ($\rho = 1$) was considered in [201]. The polling system with finite source of customers was discussed in [241].

The main methods for investigating the model of polling systems with gated or exhaustive queue service (Section 4.1) were reviewed in [167]. As was shown in [97], the mean waiting time in queue can be decomposed into a sum of two addends of which one is a function of the sum of the mean times of server switchover between queues and the second is the mean waiting time in the corresponding $M/GI/1$-type system with a modified distribution function of service durations.

The characteristics of the polling systems with exhaustive service and zero or nonzero server switchover were jointly analyzed and compared in [228] which generalized [97] and [132]. Consideration was given to models of two kinds: the model where the server instantaneously switches over between the queues and remains at the current queue at the instant of system exhaust and the model with nonzero server switchover which at the instant of system exhaust continues to interrogate the queues. Since the models of the polling systems with zero and nonzero server switchover are studied separately, the purpose of [228] was to establish a joint result for both models and a relation for the waiting times in these systems. Joint analysis of the polling systems with zero and nonzero server switchover between the queues was also described in [60]. In the model considered in [29], the server connects to the queue if it has at least a certain (different for different queues) number of customers; at that, the server needs a random warming-up time before it can start service. Consideration was given to the gated and globally-gated service disciplines and the system with elevator-type polling order. The stationary distribution of the probabilities of the number of queued customers at the polling instants and also approximate values of the mean waiting times were obtained. The system where the server polls the queues until the system is exhausted and resumes polling as soon as the number of customers in the system exceeds some threshold M was considered in [142]. The stability conditions, the stationary distribution of the number of queues customers at the instants of polling, the expression for the weighted sum of the mean waiting times were established. For the globally-gated discipline, the mean waiting times were established explicitly. In the model of [141] the server visits only the nonempty queues and stops at the current queue when the system is exhausted. In [133] consideration was given to the system where after its exhaust the server switches to the queue Q_1. At the beginning of the busy period, each queue has a certain number of customers. The queue service discipline is exhaustive. For the given system, obtained were the stationary distribution of the number of customers in the system at the polling instants and also the mean waiting times.

A process describing operation of both the systems with server vacation and an individual queue in the exhaustive-service polling system was introduced in [31]. The mean cycle time (mean time of queue service and server vacation for the system with server vacations) and also the mean fraction of time when the server handles other queues (vacates) were obtained on the basis of the characteristics of this process.

For the polling systems with exhaustive or gated service of queues, the processes characterizing the number of customers and waiting durations in transition were considered in [222]. Ergodicity of these processes was demonstrated, and the convergence rate of their transition probabilities to the stationary distribution was estimated. It was shown in [219] that reduction in the mean times of service may result both in a decrease or an increase in the mean number of system customers.

For the symmetric system with 2, 3, and 4 queues and exhaustive or gated service discipline, [173] established formulas for the second moments of waiting times. The inequality $\sigma_1 \leq \sigma_N \leq \sigma_\infty$, where σ_N is the second moment of the waiting time in the polling system with N queues, $N \geq 1$, was shown to be valid.

In [92] an algorithm was proposed to calculate approximately the stationary distribution of the probabilities of system states and the moments of waiting for service in the system with different (gated, globally-gated, exhaustive, mixed) queue service disciplines, as well as for the system with the queue service order defined by the polling table.

The polling system with correlated group flows of customers was studied in [41]. Consideration was given to the exhaustive, gated, and limited service disciplines. It was proved that this system is stochastically decomposable, and approximate formulas of the mean waiting times were obtained.

For the $G/G/1$-type polling system with exhaustive service, the stability conditions were established in [9].

The polling system with gated service and Levy-type customer flow was investigated in [114]. A random process (stochastic Poincare map) describing the system states was introduced, and some its characteristics were obtained. Convergence of this process to the limit process was proved, and the limit process was shown to be stationary.

For the polling system with gated or globally-gated queue service discipline, [38] established a criterion for existence of the moments of any order of the random variables τ_i, $i = \overline{1, N}$, where τ_i is the sum of the service time of the queue Q_i, $i = \overline{1, N}$, and the time of server switchover to the next queue. For the system with gated service, an algorithm to calculate the marginal probabilities of the number of queued customers was obtained in [19]. The heavy-tail distribution of the waiting time in the symmetric system with gated queue service was investigated in [112].

For the symmetric $M/M/1$-type polling system with k_i-limited service and the zero server switchover, [50] presented an algorithm to calculate the stationary probabilities of the number of queued customers. In [140] consideration was given to the polling system with group Poisson customer flows in queues, constant service time, and the zero server switchover. The queue service discipline is 1-limited. For the given model, the mean waiting time was obtained.

The stability conditions for the system with k_i-limited service were obtained in [136]. Its results were generalized in [89] to the case where the distribution of the time of server switchover between queues depends not only on the number of the queue to which the server connects, but also on the number of its customers. Consideration was given to existence of the stationary mode for an individual queue in the polling system. The system with the group Poisson customer flows was examined in [24]. The greatest waiting time was determined.

Some characteristics of the polling systems such as the mean waiting times were established in [34, 43, 70].

4.3.2. Random service discipline. For the polling system with random service discipline, the LST's of the distribution functions of waiting time and time of sojourn in each queue were obtained in [188]. The polling system with correlated Levy customer flow was examined in [184].

The polling systems with k_i-limited service discipline where the numbers k_i, $i = \overline{1, N}$, are defined in each cycle by the values of the random variables were studied in [35, 194]. For these models, the stability conditions were obtained, and stochastic monotonicity of some system characteristics such as queue lengths or durations of cycles vs. the parameters of the customer flows, service processes, and server switchovers between the queues was analyzed.

4.3.3. Binomial service discipline. The polling systems with binomial service were considered in [189, 190]. We present some of the results. Let X_i be the number of customers in the queue at the instant when the server connects to it. The probability that the server will serve k_i customers is $C_{X_i}^{k_i} p_i^{k_i} (1 - p_i)^{X_i - k_i}$, $k_i = \overline{0, X_i}$.

The mean number of the queued customers in Q_i at the instant when the server connects to it obeys the following expressions:

$$f_i(i) = \mathbf{M}(X_i) = \lambda_i \sum_{j=1}^{N} r_j + \lambda_i \sum_{j=1}^{N} \mathbf{M}(X_j) b_j p_j + \mathbf{M}(X_i) \overline{p}_i, \tag{5}$$

where $\overline{p}_i = 1 - p_i$, $i = \overline{1, N}$.

Solution of system (5) provides the relation $\mathbf{M}(X_j) = \frac{\lambda_j}{p_j} \frac{p_i}{\lambda_j} \mathbf{M}(X_i)$, whence we get the equalities

$$f_i(i) = \mathbf{M}(X_i) = \frac{\lambda_i}{p_i} \frac{r}{1 - \rho}, \quad i = \overline{1, N}. \tag{6}$$

We denote by $f_i(j) = \mathbf{M}(X_i^j)$ the mean number of customers in the queue Q_j when the server connects to the queue Q_i, $i, j = \overline{1, N}$. The following relations are valid for them:

$$f_{i+1}(i) = (\overline{p}_i + p_i\rho_i)\frac{\lambda_i}{p_i}\frac{r}{1-\rho} + r_i\lambda_i,$$

$$f_{i+j+1}(i) = \lambda_i\left[\frac{r\sum\limits_{k=0}^{j}\rho_{i+k}}{1-\rho} + \frac{\overline{p}_i}{p_i}\frac{r}{1-\rho} + \sum_{k=0}^{j} r_{i+k}\right], \quad j = \overline{0, N-2}.$$

The system of equations of the second moments of the number of customers at the polling instants which follow the equalities

$$f_i(j,k) = \begin{cases} \mathbf{M}(X_i^j X_i^k), & j \neq k \\ \mathbf{M}[(X_i^j)^2 - \mathbf{M}(X_i^j)], & j = k, \end{cases}$$

is given by

$$f_{i+1}(j,k) = \lambda_j\lambda_k\left((r_i^{(2)})^2 + r_i^2\right) + r_i\lambda_k f_i(j) + r_i\lambda_j f_i(k) + f_i(i)\lambda_j\lambda_k\left[2b_ip_ir_i + p_ib_i^{(2)}\right]$$
$$+ f_i(j,k) + f_i(i,j)p_ib_i\lambda_k + f_i(i,k)b_ip_i\lambda_j + f_i(i,i)(p_ib_i)^2\lambda_j\lambda_k, \quad j \neq i, k \neq i,$$
$$f_{i+1}(i,k) = \lambda_i\lambda_k\left((r_i^{(2)})^2 + r_i^2\right) + r_i\lambda_i f_i(k) + f_i(i)\left[\lambda_i\lambda_k\left(2p_ib_ir_i + p_ib_i^{(2)}\right) + \overline{p}_i\lambda_k r_i\right]$$
$$+ f_i(i,k)(p_i\rho_i + \overline{p}_i) + f_i(i,i)\left[(p_ib_i)^2\lambda_i\lambda_k + \overline{p}_ip_ib_i\lambda_k\right], \quad i \neq k,$$
$$f_{i+1}(i,i) = \lambda_i^2\left((r_i^{(2)})^2 + r_i^2\right) + f_i(i)\left[\lambda_i^2\left(2p_ib_ir_i + p_ib_i^{(2)}\right) + 2\overline{p}_ir_i\lambda_i\right] + f_i(i,i)(\overline{p}_i + p_i\rho_i)^2.$$

The mean waiting time in Q_i obeys the formula

$$\mathbf{M}(W_i) = b_i + \frac{f_i(i,i)}{f_i(i)}\frac{1 + p_i\rho_i + \overline{p}_i}{2\lambda_i}, \quad i = \overline{1, N}.$$

In the case of symmetric system ($\lambda_i = \lambda$, $b_i = b$, $b_i^{(2)} = b^{(2)}$, $r_i = r$, $r_i^{(2)} = r^{(2)}$), the last equality takes the form

$$\mathbf{M}(W_i) = b + \frac{r^{(2)}}{2r} + \frac{N\lambda b^{(2)}}{2(1-N\rho)} + \frac{Nr(1+\rho)}{2(1-N\rho)} + \frac{\overline{p}}{p}\frac{Nr}{1-N\rho}.$$

4.3.4. Bernoulli service discipline. For the polling system with Bernoulli service discipline, the LST's of the functions of distributions of waiting times as well as the mean waiting times were obtained in [247]. Similar results were obtained in [242] for the symmetric system. The $M/M/1$-type polling system with the zero server switchover was considered in [49], and an algorithm for approximate calculation of the stationary probabilities of the number of customers in the system was developed. The results of [247] were extended in [52] to a nonzero server switchover between the queues.

4.3.5. Mixed service discipline. A stochastic decomposition for the system with group flows of customers and mixed queue service disciplines was obtained in [91]. The intervals between the instants of customer arrivals to the queue Q_i, $i = \overline{1, N}$ are distributed exponentially, the number of customers in a group is distributed arbitrarily. Queues may have any of the four service disciplines: exhaustive, gated, 1-limited, and half-exhaustive where the server handles the queue until its length becomes one less than it was at the instant of polling. The stochastic decomposition has the following sense: let the system operate in the stationary mode. The number of customers in the

$M/GI/1$-type system at an arbitrary time instant has the same distribution as the sum of the number of customers in the corresponding system and the number of customers in the polling system at an arbitrary instant of switchover between the queues. By the corresponding system is meant the queuing system with an input Poisson flow with the parameter $\Lambda = \sum_{i=1}^{N} \lambda_i$ and the distribution function $\frac{1}{\Lambda}\sum_{i=1}^{N} \lambda_i B_i(t)$ of the time of customer service.

In [255] consideration was given to the polling system with exhaustive and gated service under heavy traffic. Formulas for calculation of the moments of any order of the waiting times as well as approximate formulas for calculation of these values under low load were obtained. An explicit form of the LST of the distributions of the times of waiting under heavy traffic was obtained in [252]. A random process describing transient behavior of the system was introduced and examined in [148]. For the case of nonrandom durations r_i, $i = \overline{1, N}$, of server switchover between the queues, [251] proved that for $r = \sum_{i=1}^{N} r_i \to \infty$ the waiting times are distributed uniformly.

For the system with the zero server switchover and a mix of exhaustive and gated disciplines, expressions were established in [253] for any order of the moments of waiting times.

The polling system considered in [150] has queues of two categories: queues with exhaustive service and those with 1-limited service. Queues of the same class have identical parameters of customer arrivals, service, and server switchover between the queues. The results of investigating the two-queue systems were presented in Section 4.2 (page 179). Now, we extend them to the case of N queues where the queue Q_1 is served exhaustively, the rest of them having 1-limited service. The parameters of the queues $Q_2, \ldots, Q_N$ are determined like those of the queue Q_2 in the description of the two-queue system. A similar scheme is applied to the generalized system.

At the time of arrival of a tagged customer to the queue Q_i the server can be in any of the following states:

(1) switchover to the queue Q_1 (with the probability p_{r1});

(2) switchover to the queue Q_i (with the probability p_{ri}), $i = \overline{1, N}$;

(3) service of the queue Q_1 (with the probability p_{b1});

(4) service of the queue Q_i (with the probability p_{bi}), $i = \overline{1, N}$.

The mean waiting time of a customer arriving to the queue Q_1 which has l customers is given by the equality

$$\mathbf{M}[W_1 | L_1 = l] = l b_1 + p_{r1}\left(\frac{r_1^{(2)}}{2r_1}\right) + p_{b2}\left(\frac{b_1^{(2)}}{2b_1}\right) + \sum_{j=2}^{N}\left[p_{rj}\left[\frac{r_2^{(2)}}{2r_2} + y_{j1}\right] + p_{bj}\left[\frac{b_2^{(2)}}{2b_2} + x_{j1}\right]\right],$$

where x_{j1} is the mean time from the instant when the server leaves Q_j to the instant of beginning service in Q_1, provided that the tagged customer arrived to the system when the queue Q_j was served; y_{j1} is the mean time from the instant when the server leaves Q_j to the instant of starting service in Q_1 under the same condition, $j = \overline{1, N}$. The following equality was obtained using the Little formula:

$$(1-\rho_1)\mathbf{M}(W_1) = p_{r1}\left(\frac{r_1^{(2)}}{2r_1}\right) + p_{b2}\left(\frac{b_1^{(2)}}{2b_1}\right) + \sum_{j=2}^{N}\left[p_{rj}\left[\frac{r_2^{(2)}}{2r_2} + y_{j1}\right] + p_{bj}\left[\frac{b_2^{(2)}}{2b_2} + x_{j1}\right]\right].$$

The mean cycle time obeys the formula $c = \frac{r}{1-\rho}$. The equalities $p_{bj} = \rho_j$ and $p_{rj} = \frac{r_j}{c}$ are valid as well. It follows from them that

$$\mathbf{M}(W_1) = \frac{\lambda_1 b_1^{(2)} + (N-1)\lambda_2 b_2^{(2)}}{2(1-\rho_1)} + \frac{(1-\rho)\left(r_1^{(2)} + (N-1) r_2^{(2)}\right)}{2r(1-\rho_1)} + \frac{\rho_2}{1-\rho_1}\sum_{j=2}^{N} x_{j1} + \frac{(1-\rho) r_2}{r(1-\rho_1)\sum\limits_{j=2}^{N} y_{j1}}.$$

We denote by c_{xj} the mean cycle time for the polling system with the mean time $r_2^{(2)}/r_2$ of connection to the queue Q_j. Let p_{xl} stand for the probability that at the time of connection of the server to Q_j, which on the average is $r_j^{(2)}/r_j$, this queue is not empty, $l \neq j$. Valid is the equality

$$c_{xj} = \frac{b_2^{(2)}}{b_2} + r + p_{x1}h_1 + \sum_{\substack{l=2\\ l\neq j}}^{N} p_{xl}b_2,$$

where h_1 is the mean time of server sojourn at the queue Q_1, $p_{xl} = \lambda_2 c_{xj}$ and $p_{x1}h_1 = \lambda_1 c_{xj} b_1$, consequently, $c_{xj} = \frac{r+b_2^{(2)}/b_2}{1-\rho+\rho_2}$. Now, we get the equalities

$$x_{j1} = r_1 + (N-j)r_2 + \frac{(N-j)\left(\rho_2 r + \lambda_2 b_2^{(2)}\right)}{1-\rho+\rho_2},$$

$$y_{j1} = r_1 + (N-j)r_2 + \frac{(N-j+1)\rho_2\left(r - r_2 + r_2^{(2)}/r_2\right)}{1-\rho}$$

underlying the formula for mean time of waiting in Q_1

$$\mathbf{M}(W_1) = \frac{\lambda_1 b_1^{(2)}}{2(1-\rho_1)} + \frac{(N-1)\lambda_2 b_2^{(2)}}{2(1-\rho+\rho_2)} + \frac{(1-\rho)r_1^{(2)} + (N-1)(1-\rho+\rho_2)r_2^{(2)}}{2r(1-\rho_1)} + \frac{(N-1)\rho_2 r}{2(1-\rho_1)}$$
$$+ \frac{(N-1)(N-2)\rho_2^2 r}{2(1-\rho_1)(1-\rho+\rho_2)} + \frac{(N-1)(1-\rho)r_1 r_2}{r(1-\rho_1)} + \frac{(N-1)((N-2)(1-\rho)-N\rho_2)r_2^2}{2r(1-\rho_1)}.$$

To determine the mean time $\mathbf{M}(W_2)$ of waiting in the queues Q_j, $j = \overline{2,N}$, the pseudoconservation law was used [68, 75]:

$$\rho\mathbf{M}(W_1) + \frac{(N-1)\rho_2(1-\rho-\lambda_2 r)\mathbf{M}(W_2)}{1-\rho}$$
$$= \frac{\rho\left(\lambda_1 b_1^{(2)} + (N-1)\lambda_2 b_2^{(2)}\right)}{2(1-\rho)} + \frac{\rho r^{(2)}}{2r} + \frac{r\left(\rho^2 - \rho_1^2 + (N-1)\rho_2^2\right)}{2(1-\rho)}.$$

For the polling system with mixed exhaustive, gated, 1-limited, and 1-decrementing service, the pseudoconservation law is as follows:

$$\sum_{i\in E,G}\frac{\rho_i}{\rho}\mathbf{M}(W_i) + \sum_{i\in L}\frac{\rho_i}{\rho}\left(1 - \frac{\lambda_i r}{1-\rho}\right)\mathbf{M}(W_i) + \sum_{i\in D}\frac{\rho_i}{\rho}\left(1 - \frac{\lambda_i(1-\rho_i)r}{1-\rho}\right)\mathbf{M}(W_i)$$
$$= \frac{\sum\limits_{i=1}^{N}\lambda_i b_i^{(2)}}{2(1-\rho)} + \frac{r^{(2)}}{2r} + \frac{r\left(\rho - \sum\limits_{i=1}^{N}\rho_i^2\right)}{2\rho(1-\rho)} + \frac{r\sum\limits_{i\in E,G}\rho_i^2}{\rho(1-\rho)} - \frac{r\sum\limits_{i\in D}\rho_i\lambda_i^2 b_i^{(2)}}{2\rho(1-\rho)},$$

where E, G, L, D are the sets of the numbers of queues with exhaustive, gated, 1-limited, and 1-decrementing service, respectively.

A polling system which, except for the process of customer arrival, is described as in Section 4.1 was considered in [256]. The intervals between the instants of customer arrivals are distributed exponentially with the parameter λ. At each instant of arrival the system receives a group of customers defined by the vector $\mathbf{K} = (K_1, \ldots, K_N)$, where K_i is the number of customers intended for the queue Q_i. The queues are served using a mixed exhaustive–gated discipline. Approximate formulas for calculation of the mean waiting times were obtained. For the close-to-unity load ρ, the distribution of the random variable $(1-\rho)X_i$, where X_i is the number of customers in Q_i at an arbitrary instant of treating this queue by the server, was shown to be close to the gamma law.

The polling system where the server uses information about the system state to decides which queue to take for service was analyzed in [144].

4.3.6. Polling systems with unreliable server were examined in [28, 55, 77, 153, 165, 204]. A system with 1-limited service and breakdowns of the server was considered in [153]. The intervals between the server breakdowns are distributed exponentially, after breakdown the server does not handle customers during some random time interval called the repair time. If a breakdown occurs in the course of serving a customer, then service is interrupted, and after repair the customer is served repeatedly. Approximate values of the mean queue waiting times were established for the system. A similar system with the globally-gated service discipline was considered in [77] for the case where at the instant of breakdown the server goes on with service, and the breakdown takes effect only when customer service is completed. The paper [204] generalized the results of [77] and examined jointly the models of [77] and [153]. Consideration was given also to the variants of accumulation and loss of the customers arriving to the failed queue at the time of recovery.

A system with cyclic Bernoulli polling and exhaustive or gated service was studied in [165]. At occurrence of a breakdown, the server completes service of the current customer and switches to another queue. The aggrieved customer must be served for the second time. For this system, the stability conditions, stationary distribution of the system probabilities at the polling instants, and mean waiting times were obtained.

4.3.7. Polling systems with feedback. A two-queue exhaustive-service system was considered in [149]. On completion of service in the queue Q_i, the customer returns to the end of the queue with the probability ν_i or discharges the system with the probability $1-\nu_i$, $i=\overline{1,2}$. The mean waiting times were determined for this model. The results of [149] were generalized in [243] to the case of arbitrary number of queues. Consideration was given to the exhaustive, gated, and 1-limited service disciplines. For the symmetric system, the mean waiting time in queue and the mean number of customers in system at an arbitrary instant vs. the mean number of customers in the system at the instant of server disconnection from the queue were determined. For the nonsymmetric system, a system of linear algebraic equations of the mean sojourn times in queues was established.

In the model of [225], after completion of service in Q_i, the customer either discharges the system with the probability $p_{i,0}$ or goes to the queue Q_j, $i,j=\overline{1,N}$ with the probability $p_{i,j}$. The exhaustive and gated service disciplines were examined. For this system, obtained were the LST of the distribution function of the number of customers in the system at the starting instants of service, the mean time of customer sojourn in the system from the instant of leaving Q_i till the instant of leaving Q_j and an expression for the weighted sum of the mean waiting times; also the property of the stochastic decomposition was proved.

The following system with group customer flow was discussed in [193]. The intervals between the arrivals of groups are distributed exponentially with the parameter λ. The composition of a group of customers is defined by the random vector $\mathbf{k}=(k_1,\ldots,k_N)$, where k_i is the number of customers arriving to Q_i, $i=\overline{1,N}$. The exhaustive and gated service disciplines were considered. The first and second moments of the random variables X_i^j characterizing the numbers of queued customers at the instants when the server connects to a queue were obtained for this model.

An $M/GI/1$-type polling system with a limited waiting space and exhaustive service was studied in [137]. After service in Q_i, the customer goes to Q_j with the probability p_{ij} or discharges the system if there are no free places in it, $i,j=\overline{1,N}$. The stationary distribution of the probabilities of the number of customers in queues at the polling instants was determined.

4.3.8. Polling system with customer priority. In these systems, each queue has its individual priority. A queue may be served only if the higher priority queues are empty [85, 197]. If a customer arrives to a higher-priority queue, the server either completes service of the current queue or interrupts it and switches over to the higher-priority queue. The book [12] is devoted to such systems with absolute customer priority and a nonzero server switchover between the queues.

The symmetric system with two priority queues was considered in [216]. Two Poisson flows of priority customers arrive to each queue. The queue has an unlimited waiting space for higher-priority customers and only one place for lower-priority customers. The time of customer service is constant. If the queue has a higher-priority customer, the server treats it and then passes to another queue. If there are no priority customers, then the queue of lower-priority customers is served using the chosen (1-limited or exhaustive) discipline. For each priority customer class, the mean waiting time was obtained. A more general model where the system has an arbitrary number of queues with K priority Poisson inflows to each of them was considered in [129]. After connection to the queue, the server takes the highest-priority customers, treats them using the chosen (exhaustive, gated, 1-limited, or 1-decrementing) discipline, and switches over to the next queue. An expression for the weighted sum of mean waiting times was obtained for this system. The priority models of polling with group customer flows were examined in [210, 218].

The polling system of [244] has single-buffer queues each getting customers of P priority classes. After completing service of a customer in the queue Q_i, the server takes the higher-priority customer at the nearest (along its path) station. For this system, obtained were the stationary distribution of the states of the Markov chain describing system behavior at the instants of polling and also the LST's of the waiting time distribution functions.

An N-queue system where l_i independent simplest priority flows with the parameters λ_{ij}, $j = \overline{1, l_i}$, $i = \overline{1, N}$, arrive to the ith queue was considered in [138]. The time of service of the jth priority customer in the ith queue has the distribution function $B_{ij}(t)$. The server treats one customer from each queue. At connecting to a queue, the server takes the highest-priority customer. If a higher-priority customer arrives in the course of service, the service is not interrupted. The time of server switchover from Q_i to Q_{j+1} has the distribution function $R_i(t)$, $i = \overline{1, N}$. The mean waiting times of customers of each priority in each queue were obtained for this model.

4.3.9. Polling systems with finite buffers. The polling systems with single-buffer queues were considered in [93, 198]. For the $M/M/1/1$-type two-queue polling system, the latter paper established the distribution function for the interval between the instants when the customers leave each queue.

For an $M/G/1/n$-type polling system with exhaustive service, the LST of the distribution function of the interval from the instant of server disconnection from a queue to the next instant of its connection to the same queue and also the LST of the cycle time were obtained in [233]. In [157] a method of virtual buffer which connects to the queue when it is not served and accumulates the customers which fail to find a waiting place was suggested for this system. The buffer is deleted together with the stored customers at the instant of polling. This method enabled determination of the joint distribution of the number of queued customers at the instant of polling and also the LST of the distribution of the time between the instants of polling a fixed queue. For the system with single-buffer queues, formulas for calculation of the moments of any order of the waiting times were established in [21].

For the $G/D/1/n$-type polling systems, [7] presented an algorithm to calculate the stationary distribution of the number of customers in the system. The $G/G/1/n$-type polling system with 1-limited service discipline was considered in [248]. The stationary distributions of the probabilities of the number of queued customers at the polling instants and at arbitrary instants, as well as the probabilities of breakdown in each queue were obtained.

The system discussed in [161] has M finite-capacity queues of which N queues are the so-called predecessors and the rest of them are successors. The queues are numerated so that each predecessor is followed by a successor. The predecessors are served by the gated discipline, and at visit to a successor only the customers that were present at the previous cycle are served. The LST and the first two moments of the waiting times were determined for this system, as well as the optimal system topology minimizing the weighted sum of the mean waiting times.

The polling system of [245] has $(M+1)$ queues of which M queues have unit capacity and the $(M+1)$st queue has an unlimited waiting space and is served exhaustively. The customer inflow to Q_i is the group Poisson flow. The number of customers in a group has a geometrical distribution with the parameter p_i, that is, the probability that the arriving group has size k is $(1-p_i)p_i^{k-1}$, $i=\overline{1,M+1}$. Formulas were obtained for the mean waiting times and system throughput.

4.3.10. Polling systems with limited queue service times are the those where the time of server sojourn in a queue is limited. The server handles a queue either until completion of operation according to the accepted service discipline or expiration of its time when it leaves the queue [53].

An $M/M/1$-type polling system with gated or limited service discipline was considered in [102]. The time of server sojourn at the queue Q_i is limited by the constant T_i, $i=\overline{1,N}$. The stationary distribution of the probabilities of system states at the polling instants was determined. An $M/GI/1$-type polling system and exhaustive service was studied in [117]. The time of server sojourn at a queue is bounded by an exponentially distributed random variable. For the case where the time of server sojourn at a queue expired and customer is not completed, the following variants of server behavior were considered:

(1) server completes all customers accumulated in the queue before this instant,

(2) server completes the current customer,

(3) server interrupts service.

After that the server goes to another queue. The stability conditions, the stationary distribution of the number of customers in queues at the polling instants and at an arbitrary time instant, and also some performance characteristics such as the mean time of queue service and the mean number of the served customers were obtained for the given model. An expression for the weighted sum of the mean waiting times was determined for variants 1 and 2. It was shown how the considered queue service discipline can be reduced to the exhaustive, gated, and 1-limited disciplines. The polling system with exhaustive service of queues was studied in [187] where a stationary distribution of the probabilities of system states at the instants of service completion was obtained. For the system with constant times of server sojourn at queues, approximate formulas for the mean waiting times were obtained in [133]. In [226] the results of [187] were generalized using the approach presented in [167] under the assumption that the duration of server sojourn at queue has an arbitrary distribution, warming-up time being required after server connection to the queue. The stationary distribution of the probabilities of system states at the instants of server connection to the queue, as well as the formulas for calculation of the moments of the waiting time of any order were obtained. A method of analysis of the waiting time in the symmetric $M/GI/1$-type polling system was suggested in [158].

A polling system with the service discipline featuring the so-called branching property [132, 215] was considered in [259]. An equation system for the mean waiting time in queues under heavy traffic was established. A parameter $f_i = 1 - \mathbf{M}L_i$, where L_i is the fraction of time of serving Q_i in the cycle $i=\overline{1,N}$, is assigned to each service discipline from the class under study. The mean waiting times in the corresponding queues were shown to be equal under heavy traffic and different service disciplines with identical values of f_i, $i=\overline{1,N}$.

For the $MAP/PH/1$-type polling system with exhaustive service, the stationary distribution of the number of customers in the system and the mean waiting times were determined in [147].

4.3.11. Polling systems with server vacations. A system with exhaustive service of queues was studied in [145]. After its exhaustion, the server goes on vacation whose duration has the distribution function $V_j(t)$, where j is the number of the queue where the server was at the instant of system exhaustion. The LST's of the distribution functions of the waiting times and also the mean fraction of the server vacation time were established.

4.3.12. Polling systems with retrial customers. A polling system with group Poisson flow of customers was examined in [175]. Each group is decomposed in the subgroups of customers intended each for a certain queue. During the server sojourn at the queue Q_i, each waiting customer tries to occupy the server after time intervals distributed exponentially. Having connected to a queue, the server waits for a customer to make request for service or a new customer to arrive. The server waiting time is distributed exponentially. After completion of service, the server again waits for a certain time, and if none of the customers makes an attempt to be served or no new customer arrives, the server switches to the next queue. For the stationary mode, the mean number of customers in each queue was determined. Similar results were obtained in [176] for a more general model where it was assumed that on connection to a queue the server treats the residing (primary) customers in the order of their arrival, whereas those arriving in the course of serving the primary customers become retrying and try to occupy the server independently of each other. When waiting for a retried customer, the server behaves as above. The model of polling with mixed discipline of serving the queues was investigated in [177].

4.3.13. Closed polling systems. In these polling systems, a constant number of customers circulates, no customer comes to the system from outside or leaves it. A model of the cyclic polling system where each queue has one customer was studied in [18]. After completion of service by the server, the customer must be served by an external device after which it is returned to the queue. The LST of the distribution of waiting time was obtained.

A polling system with arbitrary number of customers was considered in [40]. The queues are served in a random order with the probabilities p_{ij}, $i, j = \overline{1, N}$. Consideration was given to the gated and globally-gated service disciplines. The mean time of cycle is $\mathbf{M}(C) = c\sum_{i=1}^{N} \pi_i b_i + r$ or $\mathbf{M}(C) = M\sum_{i=1}^{N} \pi_i b_i + r$, respectively, for the gated and globally-gated service, where $c = \frac{M}{\sum_{k=1}^{N} \pi_k \sum_{j=1}^{k} p_{kj}}$, M is the number oh customers in the system, π_i, $i = \overline{1, N}$, is the stationary distribution of the Markov chain with the transition probability matrix p_{ij}, $i, j = \overline{1, N}$. The system performance for the gated and globally-gated services was established to be $\frac{c}{\mathbf{M}(C)}$ or $\frac{M}{\mathbf{M}(C)}$, respectively. The generating function of the number of customers in the system at the polling instants and at arbitrary time instants was determined, as well as the mean number of customers served in unit time. The optimal order of queue service for the globally-gated service which makes up the Hamiltonian cycle and minimizes the penalty for waiting in unit time was established. A similar model where in addition breakdowns may occur under which the server interrupts service of the current queue and switches over to the next queue was considered in [108]. At that, each queue Q_i must be accessible to the server with the probability p_i and inaccessible with the complementary probability $1 - p_i$, $i = \overline{1, N}$. If at the instant of polling, the queue is inaccessible, the server switches over to the next queue. Consideration was given to exhaustive, gated, and globally-gated service disciplines.

A polling model with customers of two types—permanent customers circulating in the system and temporary customers discharging the system after service—was examined in [270].

4.4. Periodic Polling

An $M/GI/1$-type polling system with mixed (exhaustive and gated) service discipline was studied in [254] which is a generalization of [258] where the cyclic polling was considered. The system was studied for the heavy-load case. A system of linear algebraic equation of the mean waiting times was determined. A similar model was considered in [149] where approximate formulas for the mean waiting times were obtained and optimization of the order of polling and choice of the service discipline for each queue were considered.

For the polling system with exhaustive service, variants of server's behavior after emptying the system were examined in [113]:

(1) The server stops at the current queue and at arrival of a customer to the system begins polling in the prescribed order.

(2) The server stops at the current queue and at arrival of a customer to the system connects to the queue to which the customer came.

(3) The server moves to some queue called the basic queue and with beginning of the busy period moves to the queue to which the customer came.

The LST of the distribution functions of the waiting times was obtained for these models.

Asymptotic behavior of the polling system with exhaustive service and increased mean times of switchover between queues was studied in [206].

For a $G/G/1$-type polling system with a mix of exhaustive and gated service disciplines, the limit expression for the distributions of waiting times under heavy traffic was obtained in [208] which generalized the results of [207]. For the same system (without mixed disciplines), approximate formulas were obtained for calculation of the mean number of customers and the mean waiting times.

A $MAP/PH/1$-type polling system with the zero switchover between the queues and exhaustive service was studied in [130]. The time of server sojourn is limited. If this time expires and the customer service is not yet completed, then this customer must be served for the second time. For the given system, the distribution of the busy period was obtained, and an algorithm to calculate the mean number of queued customers was developed.

A polling system with globally-gated service discipline and elevator-type polling was considered in [33]. The instants when the server starts moving from queue Q_1 to queue Q_N and the instant when the server starts its backward movement are the gate ones. For the elevator-type order of service, the mean waiting times were shown to be independent of the queue number and obey

$$\mathbf{M}(W) = \frac{1}{1-\rho}\left((1-\rho)\frac{r^{(2)}}{2r} + r\rho + \frac{1}{2}\sum_{i=1}^{N}\lambda_i b_i^{(2)}\right) + \frac{r}{2},$$

which enables a “fairer” service than for cyclic polling. A similar model where the server treats Q_i if at the polling instant there were at least k_i, $i = \overline{1,N}$, customers was considered in [29], before starting to treat a queue the server needs time to warm up. For this model, a stationary distribution of the number of queued customers and approximate values of the mean waiting times were obtained.

4.5. Random Polling

A cyclic Bernoulli-type polling was introduced in [39]. With this order, the server connecting to the queue Q_i either treats it according to the given service discipline with the probability p_i and with the probability $1 - p_i$ switches over to the queue Q_{i+1}, $i = \overline{1,N}$. After connection to Q_i, the server needs a random warming-up time with the mean d_i and the second moment $d_i^{(2)}$. The service discipline is gated, exhaustive, and partially exhaustive where served are the customers that sojourned in the queue at the polling instant and the customers that arrived during the time of server warming-up. We present the main results of investigating this model. Conditions of existence of a stationary mode $\rho < 1$, $p_i > 0$, $i = \overline{1,N}$.

We denote by X_i^j the number of customers in the queue Q_j at the instant of polling Q_i, $i,j = \overline{1,N}$; let also $f_k(i) = \mathbf{M}(X_i^k)$, $f_k(i,j) = \mathbf{M}(X_k^j X_k^i)$ for $i \neq j$, $j \neq k$, $i \neq k$, $f_i(i,i) = \mathbf{M}(X_i^i(X_i^i - 1))$.

The values $f_k(i)$, $i,k=\overline{1,N}$, obey the systems of linear equations

$$\begin{aligned}&f_{k+1}(k)=\lambda_k\overline{r}_k+[p_k\rho_k+(1-p_k)]f_k(k),\\&f_{k+1}(i)=\lambda_i\overline{r}_k+f_k(i)+p_k\lambda_i b_k f_k(k),\quad i\neq k,\ i,k=\overline{1,N},\end{aligned}\tag{7}$$

for the gated service,

$$\begin{aligned}&f_{k+1}(k)=\lambda_k\overline{r}_k+(1-p_k)f_k(k),\\&f_{k+1}(i)=\lambda_i\overline{r}_k+f_k(i)+p_k\lambda_i\omega_k f_k(k),\quad i\neq k,\ i,k=\overline{1,N},\end{aligned}\tag{8}$$

for the exhaustive service, and

$$\begin{aligned}&f_{k+1}(k)=\lambda_k\overline{r}_k+(1-p_k)f_k(k),\\&f_{k+1}(i)=\lambda_i\overline{r}_k+f_k(i)+p_k\lambda_i\omega_k(\lambda_k d_k+f_k(k)),\quad i\neq k,\ i,k=\overline{1,N},\end{aligned}\tag{9}$$

for the partially exhaustive service, where $\overline{r}_k=r_k+p_kd_k$, $\omega_i=b_i(1-\rho_i)^{-1}$. It follows from (7)–(9) that

$$p_i(1-\rho_i)f_i(i)=\lambda_i\left[\sum_{k=1}^{N}\overline{r}_k+\sum_{\substack{k=1\\k\neq i}}^{N}\frac{b_k}{1-\rho_k}p_k(1-\rho_k)f_k(k)\right],\quad i=\overline{1,N},$$

for the gated service,

$$p_if_i(i)=\lambda_i\left[\sum_{k=1}^{N}\overline{r}_k+\sum_{\substack{k=1\\k\neq i}}^{N}p_k\omega_kf_k(k)\right],\quad i=\overline{1,N},$$

for the exhaustive service, and

$$p_i(\lambda_id_i+f_i(i))=\lambda_i\left[\sum_{k=1}^{N}\overline{r}_k+\sum_{\substack{k=1\\k\neq i}}^{N}p_k\omega_k(\lambda_kd_k+f_k(k))\right],\quad i=\overline{1,N},$$

for the partially exhaustive service.

The values $f_k(i,j)$, $i,j,k=\overline{1,N}$, are the solutions of the system of linear equations.

The mean time of the cycle c obeys the formula $c=\frac{\sum_{i=1}^N r_i}{1-\rho}$. The mean waiting time in Q_i is defined as follows:

$$\mathbf{M}(W_i)=\frac{\rho_i}{\lambda_i}+\frac{f_i(i,i)(1+\rho_i)}{2\lambda_if_i(i)}+d_i-b_i,\quad i=\overline{1,N},$$

for the gated service,

$$\mathbf{M}(W_i)=\frac{\rho_i}{\lambda_i}+\frac{\lambda_ib_i^{(2)}}{2(1-\rho_i)}+\frac{f_i(i,i)}{2\lambda_if_i(i)}+d_i-b_i,\quad i=\overline{1,N},$$

for the partially exhaustive service, and

$$\mathbf{M}(W_i)=\frac{\rho_i}{\lambda_i}+\frac{\lambda_ib_i^{(2)}}{2(1-\rho_i)}+\frac{f_i(i,i)+2f_i(i)\lambda_id_i+\lambda_i^2b^{(2)}}{2\lambda_i(f_i(i)+\lambda_id_i)}-b_i,\quad i=\overline{1,N},$$

for the exhaustive service.

For the two-queue system with the zero server switchover, similar results were presented in [179]. If in this system both queues are nonempty, then with the probability p_i the server treats a customer from the current queue and with the complementary probability $(1-p_i)$, a customer from the other queue, $i = \overline{1,2}$.

For the $M/GI/1$-type polling system with exhaustive service, the stability conditions were established in [128]. The random process characterizing the joint number of customers in the polling system in the transient mode was studied in [104].

The polling system with mixed service discipline was considered in [227], and the stability conditions as well as an expression for the weighted sum of the mean waiting times were obtained for it. A technique for calculation of the mean waiting time was presented for the system with exhaustive and gated service.

Consideration was given in [185] to the polling system with unit-capacity queues where the server can vacate after exhaustion of the system. The priority order of queue service was discussed in addition to the random order. The LST's of the distributions of the times of service, switchover, and server vacation.

Using a new interpretation of the cycle time, the polling systems with zero and nonzero server switchover were analyzed jointly in [185]. The system queues are the single-buffer ones. The server does not treat customers in the following cases:

(1) Switchover between the queues.

(2) Waiting for a customer in the empty system.

(3) Simple server.

By the server cycle is meant the duration of service of customer or waiting for it and vacation or switchover between queues. Consideration was given to the priority order of service where the server takes the highest-priority queue, the random order where the server takes the queue Q_i with the probability γ_i, and the cyclic order. The system with noncorrelated customer flows, noncyclic service, and server vacations depending on the system state was considered for the first time. The LST's of the distribution functions of service time, server switchover, and server vacation were determined.

The $G/G/1$-type polling system was analyzed in [32] relying on the results obtained for the $G/G/1$-type polling system with server vacations. In [200] the stability existence conditions were established. The system with the 1-limited service discipline was considered in [119]. In this model, on visiting the queue Q_i the server switches over to the queue Q_j with the probability p_{ij}, provided that Q_j is nonempty, and with the probability $\widetilde{p}_{ij}$, otherwise. For this system, obtained were the existence condition for the stationary mode, the stationary distribution of the server state, and the mean waiting time. The $G/G/1$-type polling system with mixed exhaustive and gated service discipline was studied in [44] where the stability conditions and the expressions relating the mean number of customers and the mean waiting times with the corresponding characteristics of the one-server queuing system were determined. Approximate formulas for the mean waiting times in the heavy-load environment were obtained. The $G/G/1$-type polling system queues where the flow arriving to the queue Q_i is defined by the set of values $\{A_i(s,t],\ 0 \leq s < t\}$, with $A_i(s,t]$ for the number of customers arriving to the queue over the time interval $(s,t]$, was studied in [127]. The set $N(i)$ of the numbers of queues which are regarded as neighboring to the queue Q_i is assigned to each queue Q_i. After visiting Q_i, the server switches over to the longest Q_j, that is, $j \in N(i)$, $i = \overline{1,N}$. The server handles at most B queued customers and stops service if the queue length becomes smaller than C. The pair (B,C) is defined by the two-dimensional random variable whose distribution depends on the queue number and the number of customers queued at the instant of polling. The sufficient stability conditions were established for this system.

The symmetric polling system with the simplest customer flow which is common to all queues was considered in [78]. The arriving customer is placed on the shortest queue. One of the customers is a special customer. The queue where it sojourns is taken with the probability p, the rest of the queues being taken equiprobably, that is, with the probability $\frac{1-p}{N-1}$. At each time instant, the system can have at most one special customer. The customer queued at the end can move to another, smaller queue. The server treats one customer from each queue. The service is distributed exponentially. For the special customer, the waiting time distribution function was established.

The polling system with the server route defined by a sequence of pairs of random variables $\{v_j, w_j\}_{j=-\infty}^{\infty}$, where v_j is the number of the queue and w_j is the duration of server switchover to Q_{v_j}, was considered in [23]. In one visit to Q_i, the server treats $f_i(x)$ customers, where x is the number of queued customers at the polling instant. The necessary and sufficient conditions for boundedness in probability were obtained in this work, and existence of the stationary mode was proved for the case where these conditions are met. This model was studied further in [22] under the assumption that the number of customers treated by the server in one visit to Q_i is $f_i(x, D)$, where D is some random control parameter and the number of queues is assumed to be countable.

For the systems with probability-limited service and with infinite number of queues, a criterion for existence of the stationary mode was established in [56, 57], respectively, as $\omega = \sum_i \pi_i \sum_j p_{ij}\omega_{ij} < \infty$ and $\lambda_i\omega < (1-\rho)\pi_i d_i$, where λ_i is the rate of the customer flow to the ith queue, p_{ij} is the probability that on completing treatment of Q_i the server switches over to Q_j, ω_{ij} is the mean time of this switchover, d_i is the mean number of the customers treated in Q_i, $i, j = \overline{1, N}$, and π_i, $i = \overline{1, N}$, is the stationary distribution of the Markov chain with the transition probability matrix $(p_{ij},\ i, j = \overline{1, N})$.

4.6. Priority Polling

The polling models with priority service are described in [162, 174]. The polling system of [174] has three queues. The flow of customers to the queue Q_1 is controlled by the Markov chain with the state space $\{0, 1\}$. If at the beginning of the slot the Markov chain is in state 1, then a customer arrives to the queue Q_1; otherwise, no customer arrives. The Bernoulli flows arrive to the queues Q_2 and Q_3. The queue Q_1 has an absolute priority, and if at the instant of arrival of a customer to Q_1 the server treats another queue, at the beginning of the next slot it switches over to Q_1 and after emptying it returns to the interrupted queue. The queues Q_2 and Q_3 are treated using, respectively, the gated and 1-limited disciplines. For the given model, the mean waiting times were determined. Similar results were obtained in [162] for the polling system with queues numerated in the descending order of priorities. In this system, the queue Q_i may be treated only if the queues $Q_1, \ldots, Q_{i-1}$, $i = \overline{2, N}$, are empty.

For the $M/D/1$-type polling system consisting of $N+1$ queue, a priority reservation discipline was introduced in [203]. The queue Q_{N+1} has a priority and can reserve the server which on treating the current queue Q_i passes to the queue Q_{N+1} and then to the queue Q_{i+1}. The mean waiting times and system throughput were determined. The system with the probabilistic priority of queue service was considered in [155].

The system of unit-capacity queues with correlated customer flows was examined in [186]. The order of service obeys the probabilities γ_i, $i = \overline{1, N}$, of taking the queues for service. Consideration was given to two models where (i) the queue accepts a new customer only after completion of service of the preceding customer and (ii) the queue becomes accessible to the newly arrived customer after the start of service of the preceding customer. Obtained were a system of linear algebraic equations for calculation of the mean waiting times, and for the symmetric system, exact formulas for the mean waiting times, the probability of losing customers, and the system throughput.

The polling system with two $PH/PH/1$ queues, phase-type flows, and phase service was considered in [1]. The queue Q_1 has a relative priority. An algorithm to calculate the stationary probabilities of system states was developed.

5. SERVICE OPTIMIZATION IN THE POLLING SYSTEMS

This section reviews the publications on optimization in the polling systems. The works are grouped in the subsections in terms of the optimized system parameters (or characteristics) such as the order of queue, discipline of queue service, server behavior, and so on.

5.1. Polling Order

Optimization of the order of queue service can be static (the optimal in a sense order of polling is predefined and remains unchanged in the course of operation) and dynamic (the order of polling depends on the system state).

For the polling system of Section 4.1 with random order of queue service defined by the probabilities p_i, $i = \overline{1, N}$, and exhaustive or gated service, the optimal values of the probabilities p_i^*, $i = \overline{1, N}$, minimizing the weighted sum of the mean waiting times were obtained in [65]:

$$\sum_{i=1}^{N} \rho_i \mathbf{M}(W_i) = \rho \frac{\sum\limits_{i=1}^{N} \rho_i b_i^{(2)}}{2(1-\rho)} - \frac{\sigma}{1-\rho} \sum_{k \in e} \frac{\rho_k^2}{p_k} + \frac{\sigma}{1-\rho} \sum_{k=1}^{N} \frac{\rho_k}{p_k} - \rho\sigma + \rho \frac{\sigma^{(2)}}{\sigma},$$

where $\sigma = \sum_{i=1}^{N} p_i r_i$, $\sigma^{(2)} = \sum_{i=1}^{N} p_i r_i^{(2)}$, and e is the set of the numbers of queues with exhaustive service. The optimization problem

$$\sum_{i=1}^{N} \rho_i \mathbf{M}(W_i) \xrightarrow[p_1, \ldots, p_N]{} \min, \quad \sum_{i=1}^{N} p_i = 1, \; p_1 \geq 0, \ldots, \; p_N \geq 0,$$

is the classics of nonlinear optimization with linear constraints which yields to the method of Lagrangian multipliers. Its solution is as follows:

$$p_k^* = \frac{\sqrt{\rho_k(1-\rho_k)}}{\sum\limits_{j \in e} \sqrt{\rho_j(1-\rho_j)} + \sum\limits_{j \in g} \sqrt{\rho_j}}, \quad k \in e, \tag{10}$$

$$p_k^* = \frac{\sqrt{\rho_k}}{\sum\limits_{j \in e} \sqrt{\rho_j(1-\rho_j)} + \sum\limits_{j \in g} \sqrt{\rho_j}}, \quad k \in g, \tag{11}$$

where g is the set of the numbers of queues gated service.

For the polling system of Section 4.1, the paper [72] posed the problem of determining the optimal polling table minimizing the penalty for waiting for customers in unit time $\sum_{i=1}^{N} c_i \lambda_i \mathbf{M} W_i$, where c_i is the penalty for waiting for customer in the queue Q_i in unit time, $\mathbf{M}W_i$ is the mean time of waiting in Q_i, $i = \overline{1, N}$. We note that according to the Little formula the product $\lambda_i \mathbf{M} W_i$ is equal to the mean number of queued customers Q_i. Three service disciplines were considered: exhaustive, gated, and 1-limited. The numbers of the queues with exhaustive, gated, and 1-limited service disciplines are accumulated in the respective sets e, g, and $(1 - L)$.

The problem of determining the optimal polling table is decomposed in the following subproblems:

(1) The optimal frequency $f_i = \frac{m_i}{M}$ of visits to the queue Q_i, where M is the length of the polling table and m_i is the number of visits to the queue Q_i in this table, $i = \overline{1, N}$, is determined using the optimization criterion. The values of M and m_i, $i = \overline{1, N}$, are unknown.

(2) Determined is the length of the polling table M, that is, the minimal number for which Mf_1, $\dots$, Mf_N are either integers or differ from integers at most by a small value ε, provided that the sum of these values is an integer. With a knowledge of f_i, $i = \overline{1, N}$, and M, one can calculate the values of m_i, $i = \overline{1, N}$.

(3) Now, determined is the order of queue service realizing the polling table with the established M and m_i, $i = \overline{1, N}$. This procedure was presented in [154].

The following approximation of the mean waiting times was used to determine the objective function vs. m_i, $i = \overline{1, N}$:

$$\mathbf{M}(W_i) \approx A(1-\rho_i)\frac{\sum\limits_{j=1}^{N} m_j r_j}{m_i}, \quad i \in e,$$

$$\mathbf{M}(W_i) \approx A(1+\rho_i)\frac{\sum\limits_{j=1}^{N} m_j r_j}{m_i}, \quad i \in g,$$

$$\mathbf{M}(W_i) \approx A\frac{1-\rho+\rho_i}{1-\rho-\lambda_i\sum\limits_{j=1}^{N} m_j r_j/m_i}\frac{\sum\limits_{j=1}^{N} m_j r_j}{m_i}, \quad i \in 1-L,$$

where A is a positive constant independent of $m_1, \dots, m_N$. The following optimal values of $m_1, \dots, m_N$ were determined to within the constant factor:

$$m_i \sim \sqrt{c_i\lambda_i(1-\rho_i)/r_i}, \quad i \in e,$$

$$m_i \sim \sqrt{c_i\lambda_i(1+\rho_i)/r_i}, \quad i \in g,$$

$$m_i \sim \lambda_i + \left(1-\rho-\sum_{k=1}^{N}\lambda_k r_k\right)\frac{\sqrt{c_i\lambda_i(1-\rho+\rho_i)/r_i}}{\sum\limits_{j=1}^{N} r_j\sqrt{c_j\lambda_j(1-\rho+\rho_j)/r_j}}, \quad i \in 1-L.$$

For the system with exhaustive, gated, or limited discipline, the optimal frequencies of visits to the queues f_i, $i = \overline{1, N}$, minimizing the number of system customers were obtained in [269]. The mean waiting times $\mathbf{M}W_i$, $i = \overline{1, N}$, are replaced by the approximate values from [72] which enabled determination of the optimal values of m_i, $i = \overline{1, N}$.

An N-queue system with Poisson flow of customers and group service was considered in [260]. The time of visit to queue consists of the three periods: (i) server connection to a queue, (ii) customer service, and (iii) disconnection. The queues are served using the exhaustive, gated, or globally-gated disciplines. To ensure fair service, in a cycle the server polls all nonempty queues once and only once. Needed is to determine the optimal order of polling, that is, the permutation $\boldsymbol{\pi} = (\pi(1), \dots, \pi(L))$, where L is the number of nonempty queues and $\pi(l)$ is the number of the queue which is polled lth in turn in the cycle, $l = \overline{1, L}$. Different optimization criteria were considered:

(1) Minimization of the weighted sum of the mean times of customer sojourn in the system.

(2) Minimization of the mean time of the next cycle.

(3) Maximization of the system throughput (the mean number of customers served in a cycle).

For criterion 1, the optimization problem was solved for the globally-gated discipline; for the rest of the service disciplines, the problem was reduced to that of minimization of a function with L integer variables. For criterion 2, it was shown that for the globally-gated discipline all permutations $\boldsymbol{\pi}$ are optimal. For the rest of the disciplines, the problem was solved in some special cases. For criterion 3, the optimization problem was solved for the gated and exhaustive service disciplines.

For the symmetric $M/GI/1$-type polling system with gated or exhaustive service, consideration was given in [118] to optimization of the mean cycle time, provided that the polling order makes up a Hamiltonian cycle. The mean time of cycle was shown to be maximal for the cyclic polling; and for the problem of minimization of the mean time of cycle, the optimal order of queue service was determined which is other than the cyclic order.

For the polling system with server breakdowns, the problem of static and dynamic optimization of the service order which minimizes the waiting penalty was solved in [77]. In the case of dynamic optimization, the decision about the order of service is made at the beginning a cycle. It was also shown that for the elevator-type order of service, the mean waiting times are equal.

For the $M/G/1/1$-type polling system, consideration was given in [81] to the problem of optimization of the order of service with the aim of minimizing the total waiting penalty and loss of customers.

The $G/G/1$-type polling system was considered in [195]. The time of server sojourn in each queue is defined by a random variable. The server cannot leave the queue until expiration of the sojourn time. If the sojourn time expired and there are queued customers, then either all remaining customers or only those that were queued before polling discharge the queue. Consideration was given to the problem of optimal determination of the queue to be served with the aim of minimizing the number of system customers. The cyclic polling was shown to be optimal if there is no information about the system state; if there exists some partial or full information, then the optimal choice is represented by the queue with the greatest number of customers.

Application of the polling model to the Internet search engines can be found in [94]. There are N Web-pages polled by the search engine with the aim of updating the Web-page database in compliance with some polling table. The content of the ith page is modified after time intervals that are distributed exponentially with the parameter μ_i, $i = \overline{1, N}$. From the instant of engine access to the next instant of modification of page contents, the page is regarded as updated after which it is regarded as outdated until the next access of the search engine. Let r_i be the fraction of time during which the ith page is outdated. Needed is to determine a polling table minimizing the value of the objective function $C = \sum_{i=1}^{N} c_i r_i$. In the polling system describing such a model, the instants of changes of the page contents are interpreted as those of customer arrivals, and the durations of access of the search engine to the page, as the switchover durations. The time of queue service is assumed to be zero. For this system, the exact greatest lower bound of the objective function and the optimal frequencies of the visits of server (search engine) to the queues (Web-pages) were determined under the assumption that c_i is proportional to μ_i. Similar results were obtained for $N \to \infty$. Some algorithms [94] of Web-page polling which realize the optimal visit frequencies are analyzed in [17].

A polling system with finite waiting space and group customer service was studied in [268]. It is required to determine the optimal sequence of queue service that maximizes the system throughput. The choice of a queue with the greatest number of customers is optimal if the queues have the same waiting space are equal.

The $M/GI/1$-type system with elevator-type polling and globally-gated service discipline was studied in [42]. It is desired to numerate the queues so as to reduce the variance of the mean waiting times $\sum_{k=1}^{N} |\Delta_k|$, where Δ_k is the difference between the mean waiting times of a customer during the period when the server polls the queues from Q_N to Q_1 and the customer arriving to Q_k during the time where the server polls the queues from Q_1 to Q_N. The properties of the optimal permutation $\boldsymbol{\pi} = (\pi(1), \dots, \pi(N))$ were described. The optimal permutations for an even N were obtained.

5.2. Queue Service Disciplines

For the cyclic polling, the problem of optimal service disciplines was solved in [199] for the following structure of polling. The first N^c queues receive customers representing custom-built goods, the rest of the queues getting the standard goods. Service of the custom-built goods needs a permission for service to arrive to the queue. Each permission has a certain lifetime during which the goods may be served. The standard goods need no permission and are served as usual. Optimization is carried out with the aim of minimizing the penalty for server switchover between the queues. The system has N queues.

The polling system with the k_i-limited service discipline (see Section 4.1, page 178) was considered in [62]. Minimization of the mean penalty $\sum_{i=1}^{N} c_i \lambda_i \mathbf{M}W_i$ on the set of parameters $k_1, k_2, \dots, k_N$ was considered for waiting a customer during a unit time, provided that $\sum_{i=1}^{N} \gamma_i k_i \le K$, where γ_i, $i = \overline{1, N}$, and K are some positive constants. By using in place of $\mathbf{M}W_i$, $i = \overline{1, N}$, the approximate values determined using different approaches, the optimal set of $k_1, k_2, \dots, k_N$ was determined.

(1) *Approach based on the polling table with 1-limited service.* The cyclic service of queues with k_i-limited service discipline is interpreted as the 1-limited periodic service with the polling table

$$\{\underbrace{Q_1, \dots, Q_1}_{k_1}, \dots, \underbrace{Q_N, \dots, Q_N}_{k_N}\}.$$

The results of [72] were used to obtain the following optimal values of k_i, $i = \overline{1, N}$:

$$k_i^* = \frac{\lambda_i r}{1-\rho} + \left(K - \sum_{j=1}^{N} \gamma_j \frac{\lambda_j r}{1-\rho}\right) \frac{\sqrt{c_i \lambda_i (1-\rho+\rho_i)/\gamma_i}}{\sum\limits_{j=1}^{N} \gamma_j \sqrt{c_j \lambda_j (1-\rho+\rho_j)/\gamma_j}}, \quad i = \overline{1, N}.$$

(2) *Approximation of the mean waiting times for the system with k_i-limited service* makes use of the results of [72, 75]. In this case,

$$k_i^* = \frac{\lambda_i r}{1-\rho} + \left(K - \sum_{j=1}^{N} \gamma_j \frac{\lambda_j r}{1-\rho}\right) \lambda_i \frac{\sqrt{c_i (1-\rho+\rho_i)/\gamma_i}}{\sum\limits_{j=1}^{N} \lambda_j \gamma_j \sqrt{c_j (1-\rho+\rho_j)/\gamma_j}}, \quad i = \overline{1, N}.$$

(3) *Approximation of the mean waiting times as obtained in* [134],

$$\mathbf{M}(W_i) \approx \frac{(1-\rho_i)(1-\rho) + \dfrac{\rho_i}{k_i}(2-\rho)}{1 - \rho - \dfrac{\lambda_i r}{k_i}} \mathbf{M}(RC_i), \quad i = \overline{1, N},$$

where $\mathbf{M}(RC_i)$ is the mean time of the cycle which starts with service of the queue Q_i. The following approximate value was used for $\mathbf{M}(RC_i)$:

$$\mathbf{M}(RC_i) \approx \frac{D + \dfrac{r}{1-\rho} \sum\limits_{j=1}^{N} \dfrac{\rho_j^2}{k_j}}{\sum\limits_{j=1}^{N} \left[\rho_j (1-\rho_j) + \dfrac{\rho_j^2}{k_j} \dfrac{2-\rho}{1-\rho}\right]}, \quad i = \overline{1, N}.$$

However, the problem of optimization on the basis of this approach was not yet solved analytically; the optimal values of $k_1^*, \dots, k_N^*$ can be determined only numerically. The following optimal values were obtained in [62] using the approximate value of $\mathbf{M}(RC_i) \approx B\sum_{j=1}^{N} m_j r_j/(1-\rho)$:

$$k_i^* = \frac{\lambda_i r}{1-\rho} + \left(K - \sum_{j=1}^{N} \gamma_j \frac{\lambda_j r}{1-\rho}\right) \frac{\sqrt{c_i\lambda_i[\rho_i(2-\rho)+\lambda_i r(1-\rho_i)]/\gamma_i}}{\sum\limits_{j=1}^{N} \gamma_j \sqrt{c_j\lambda_j[\rho_j(2-\rho)+\lambda_j r(1-\rho_j)]/\gamma_j}}, \quad i = \overline{1,N}.$$

Optimization was further considered in [65] where the optimal values of $k_1^*, k_2^*, \ldots, k_N^*$ were defined more precisely. For the system with group Poisson flows of customers, a similar optimization problem was considered in [24] using the mean waiting time as an optimization criterion. The upper and lower boundaries of the optimal values of the parameters $k_1^*, k_2^*, \ldots, k_N^*$ were obtained.

For the polling system of Section 4.1, the problem of determining the optimal mixed (exhaustive or gated) service discipline in order to minimize the penalty for waiting for customer in unit time was posed in [258]. Obtained were approximate values of the mean waiting times which underlay solution of the optimization problem for the system with the zero server switchover; for the system with a nonzero switchover only partial solution was obtained.

For the polling system with periodic service, [61] introduced a fixed-time scheme of visits to queues which is defined by the pair of vectors (P, T), where $P = (P(1), \ldots, P(M))$ is the polling table of length M, $T = (T_1, \ldots, T_M)$, $T_i \geq 0$, $i = \overline{1,M}$, with T_k for the times of starting service of the queues $Q_{P(k)}$ and $Q_{P(k+1)}$. Approximate values of the mean waiting time were established, and the problem of determining the optimal scheme (P^*, T^*) minimizing the cumulative penalty for customer sojourn in the system was solved.

The problem of determining the optimal Bernoulli-type service parameters $(p_1, \ldots, p_N)$ for the cyclic polling scheme was solved in [54]. The weighted sum of the mean waiting times was used as the criterion for optimization. Partial and approximate solutions of this system were determined.

5.3. Optimization of the Service Policy

Optimization of the service policy for the $M/M/1/n$-type two-queue polling system with the zero server switchover was discussed in [230]. Established was the optimal rule of server switching which minimizes the cumulative waiting penalty and loss of customers for which there were no waiting space. This model was further considered in [168] with the aim of analyzing the limit behavior of the switchover curve which decomposes the two-dimensional domain of system states into the subdomains on getting to which the server connects to one or another queue or vacates. In a similar model of [163], it is assumed that a penalty is imposed for each customer lost because of buffer overflow. A problem was posed of determining the optimal policy in the class of stationary nonradomized Markovian policies, solution about taking a customer for service being based on the information about the T connections of the server. It was shown that in the case of penalty for lost customers, the optimal policy belongs to the class of threshold policies, that is, the server connects to a queue if the number of its customers exceeds a certain threshold. In the case where no penalty for lost customers is imposed, the problem of determining the optimal policy is much more complicated. The properties of the optimal policies were described. The results of [110, 111, 168, 196] were generalized in [230]. Optimization of the service policies in some queuing systems for the purpose of investigating the optimal policies for monotonicity was discussed in [8, 16, 217].

For the $M/M/1$-type two-queue polling system where the penalties for customer waiting in unit time and for inter-queue switchover are defined, the form of the function of cumulative penalty imposed in unit time was determined in [169]. The system with two finite-capacity queues was considered in [249]. The inflow to Q_i is the Poisson flow with the parameter $\lambda_i(x_1, x_2, k)$, $i = \overline{1,2}$, depending on the number of queued customers x_1 and x_2 and the number k of the queue treated by the server. It is desired to determine the optimal service policy maximizing the system throughput, provided that the mean waiting time is bounded by T. Consideration was given to the cases of full and incomplete information about the system states; in the latter case, only the state of the queue to which the server is connected is known.

The $G/G/1$–type two-queue polling system was considered in [36]. Each queue may be inaccessible to the server for a random time, but the customers still are queued. If a queue is inaccessible, then the server connects to another queue. The stability conditions were obtained for a wide class of service policies. The scheme of determination of the optimal policy was derived under certain constraints on the cost coefficients under heavy traffic.

The symmetric $M/D/1/1$-type polling system with constant time of server switchover between the queues was considered in [147]. The server polls cyclically the queues. It is required to determine the optimal server policy (simple, customer service, or switchover to the next queue); at that, the server may move backward. The aim of optimization lies in maximizing the system throughput, provided that the full information about the system state is available.

The problem of optimizing service of the $G/D/1$-type polling system was solved in [47] for the l-limited service discipline.

The optimal service policy in the $M/GI/1$-type polling system was determined in [43] as a sequence of taking customers for service which minimizes the weighted sum of the mean waiting times. Consideration was given to the permissible policies for which a stationary mode exists, the server does not interrupt customer treatment, and the decision about taking a customer for service relies only on the information about the past and current system states. For the permissible policies, the boundaries of the mean waiting times were determined. For the random polling order, one of the feasible optimal sets of the probabilities $\{p_{ij}\}$ of server switchover from Q_i to Q_j, $i, j = \overline{1, N}$, was given. Determination of the optimal polling table, that is, of the integers h_{ij} meaning the number of switchovers from Q_i to Q_j in the polling table, was reduced to conditional minimization of the function N^2 in integer variables. In [111] the weighted sum of the mean waiting times plus the cost of server switchover between the queues was used as the optimization criterion. The properties of the optimal service policy were listed, and a heuristic policy for the two-queue system was presented.

For the $M/GI/1$-type polling system, minimization of the mean penalty for customer waiting and server switchover in unit time was studied in [110]. It is assumed that the server decides to take a queue for service on the basis of information only about the former and current system states. The characteristics of the optimal policies were presented, as well as the heuristic policy defining the rules for inter-queue switchover and for server vacation. A similar problem was considered in [164] for the $M/M/1$-type polling system.

Optimization of the system with feedback was considered in [45]. It was assumed that after service in the queue Q_i the customer discharges the system with the probability p_{i0} and with the probability p_{ij} goes to the queue Q_j, $i, j = \overline{1, N}$. Permissible are the service policies where without interrupting customer treatment the server chooses a queue to serve only on the basis of the information about the past and current system states. Both static and dynamic policies were considered. Posed was the problem of determining the minimum of the cost function for the chosen class of policies and constructing a policy for which the cost function has a value sufficiently close to the minimum. For this system, obtained were the optimal frequencies of queue visits and switchovers between the queues for cyclic polling, periodic polling, and random order of queue service.

The paper [196] was devoted to optimization of the service policy for a system consisting of the $G/G/1$-type queues. The optimal policy should minimize the number of customers in system. The problem of optimization was decomposed into three subproblems:

(1) Determine the optimal action of the server connected to a nonempty queue (service, vacation, or switchover).

(2) Determine the optimal action of the server connected to an empty queue (vacation or switchover).

(3) Optimization of the choice of taking the next queue for service if the server decided to leave the current queue.

As was shown for problem 1, it is the exhaustive service discipline that is optimal. Problems 2 and 3 were solved only for symmetric systems. It was proved that at the instant of emptying the entire system the server must remain at the current queue in the case of problem 2. For the discrete-time system and the server switchover time equal to one slot, the server must switch over between the queues at the instant of emptying the system. For problem 3, the optimal choice of a queue for service depends on availability of the information about system state. If the system state is known at each time instant, then the longest queue will be the optimal choice. If no such information exists, then the cyclic polling is optimal. A similar problem of optimization for the $M/GI/1$-type polling system was studied in [73]. The weighted sum of the mean waiting times $\sum_{i=1}^{N} \rho_i \mathbf{M} W_i$ is used as the optimization criterion. The exhaustive procedure of queue service was shown to be the optimal solution of problem 1. Problem 3 was solved for the random order of taking the queues for service under exhaustive or gated service procedures. Therefore, the optimal set $(p_1^*, \ldots, p_N^*)$ of the probabilities of choosing queues for service was determined. For the systems with a polling table-defined order of service, this problem was considered in [72].

For the $G/G/1$-type two-queue polling system with the zero server switchover, [213] described the properties of the optimal service policy in the system with possible customer service interrupt minimizing the waiting penalty in unit time under heavy traffic. A system with an arbitrary number of queues that operates in the transient mode was discussed in [261]. Given is the dimensional function $\mathbf{C}(t) = (C_1(t), \ldots, C_N(t))$, where $C_i(t)$ defines the cost of the t-long time of waiting in the queue Q_i, $i = \overline{1, N}$. Needed is to determine the optimal policy defined by the vector $\mathbf{T}(t) = (T_1(t), \ldots, T_N(t))$ of queue service durations over the time interval $[0, t]$. The function $J(t) = \sum_{k=1}^{N} \sum_{i=1}^{A_k(t)} C_k(\tau_{k,i})$, where $A_k(t)$ is the number of customers in Q_k that arrived over the interval $[0, t]$ and $\tau_{k,i}$ is the time spent by k customers in the queue Q_i, was used as the optimization criterion. Under heavy traffic, the asymptotically optimal policy is that under which at the time t the server treats a queued customer with the greatest $\mu_k C_k(a_k(t))$, where μ_k is the mean time of service in Q_k and $a_k(t)$ is the time spent by the customer waiting at the beginning of Q_k over the time interval $[0, t]$.

The polling system of [109] has N queues with unlimited waiting space. Queue service is cyclic. The simplest customer flow arrives to the queue Q_1. The customer served in Q_i goes to the end of the queue Q_{i+1}, and so on, and after treatment in the queue Q_N it discharges the system. The time of service in the ith queue has the distribution function $B_i(t)$, $i = \overline{1, N}$. The customer waiting in the queue Q_i is penalized by h_i in time unit, at that $h_i \le h_{i+1}$, $i = \overline{1, N}$. It is required to determine the optimal rule g of server behavior (service, vacation at the current queue, or switchover to the next queue) for the purpose of minimizing the loss function:

$$J(g) = \lim_{T\to\infty} \sup \frac{1}{T} \mathbf{M} \left(\int\limits_0^T \sum_{i=1}^{N} h_i X_i^g(t) dt \right),$$

where h_i is the penalty for customer waiting in Q_i at unit time and $X_i^g(t)$ is the number of customers in Q_i for the policy g, $i = \overline{1, N}$. It was shown for $N = 2$ that the optimal policy is that for which queue service is determined by means of a nondecreasing function $f(X_1(t))$: if $X_2(t) > f(X_1(t))$, then the server switches over to the queue Q_2; otherwise, the queue Q_1 is served. The properties of the optimal policy were determined for an arbitrary number of queues.

For the polling system of Section 4.1, [262] studied the characteristics of system performance vs. variations of the mean durations of server switchover between the queues. As was shown, if the heavily loaded queues have lower cost of connection to them, then with the optimal policy-based service the system characteristics improve with reduction in the mean switching times.

For the discrete-time system with cyclic polling of queues and the server treating all queued customers simultaneously during one slot, [263] considered the problem of optimizing server behavior (service or vacation). Only the information about the queue to which the server is connected is available to it. The server decides about the further action by the end of the slot depending on the previous decisions about choice of action and the information about the states of queues it visited before. The optimization criterion includes the penalty for customer waiting in unit time and the cost of service and server switchover. The optimal queue service policy was proved to be the threshold one, that is, the server handles Q_i if at the polling instant it has at least θ_i customers, $i = \overline{1, N}$.

5.4. Minimization of the Mean Waiting Time

For the polling system of Section 4.1 (page 178) with the exhaustive or gated service disciplines, [100] determined the value z^* of the cumulative mean time of server switchover in one cycle which has the following property. If the cumulative mean time z of server switchover in one cycle is such that $z < z^*$, then the mean waiting times assume the least values if in each cycle the server vacates forcedly over $(z^* - z)$ time units, that is, it was shown that reduction in the mean time of server switchover may increase the mean waiting time and the forced server vacation can be used to optimize the system characteristics. However, if the server does not vacate when the system is empty and goes on with polling the queues, then the mean waiting time decreases. This fact was explained in [99]. Studies in this direction were continued in [211] where some policies of forced server vacation were considered and compared in terms of mean waiting times. Optimization of the polling systems for transportation applications was considered in [135].

5.5. Optimization of Customer Routing

The $M/M/1$-type polling system with $N + 1$ queue, cyclic polling, and exhaustive service was considered in [221]. The customers arriving to the queue Q_{N+1} have one of the following characteristics:

(1) on completion of service in the queue Q_{N+1}, the customer goes to the queue Q_i, $i = \overline{1, N+1}$, with the probability p_i or

(2) the queue Q_{N+1} is not treated by the server; therefore, the customers arriving to it are distributed between other queues with the probabilities q_i, $i = \overline{1, N}$.

It is desired to minimize the function of mean cost in unit time on the set of parameters p_i, $i = \overline{1, N+1}$, (in case 1) or q_i, $i = \overline{1, N}$, (in case 2). For case 2, the problem was solved.

The $M/GI/1$-type polling system with exhaustive service and the zero server switchover was studied in [10]. After completion of customer service, the set $(k_1, \ldots, k_N)$ of secondary customers going to the corresponding queues arises in the queue Q_i with the probability $q_i(k_1, \ldots, k_N)$. The priority service minimizing the loss function was shown to be the optimal policy of queue service. For a similar model with exhaustive service, the problem of determining the optimal order of queue polling making up a Hamiltonian cycle minimizing the objective function was solved in [20].

6. MULTIPLE-SERVER POLLING SYSTEMS

6.1. Independent Servers

The present subsection reviews the works on polling system models with servers visiting the queues independently of each other.

6.1.1. Identical servers. This subsection lists the models with identical servers, that is, the parameters of service are independent of the number of the server.

In the model of [103] each server treats one customer from each queue. After visiting the queue Q_i, the server with the probabilities p_{ij} connects to the queue Q_j, $i, j = \overline{1, N}$. For this system, obtained were the stability conditions and the stationary probability distribution of the number of queued customers which has the multiplicative form, that is, $p(n_1, n_2, \ldots, n_N) = \prod_{i=1}^{N} p_i(n_i)$, where $p(n_i)$ is the probability that the length of Q_i is n_i, $i = \overline{1, N}$.

For the $G/G/1$-type two-server polling system with exhaustive service, the stability conditions were established in [126]. The $G/G/1$-type multiple-server polling system was considered in [26]. Each customer has its destination (queue) which can be reached only by means of a server. The queue Q_j is the destination of the customer queued in Q_i with the probability p_{ij}. At connecting to a queue, the server takes only one customer and switches over to the destination queue. After reaching the destination, the customer discharges the system. If the server is connected to an empty queue, then it waits for arrival of a customer. By the system state is meant the state of the process $Q(n) = (q_j(n), x_j(n), j = \overline{1, N})$, $n \geq 1$, where $q_j(n)$ is the number of customers in the jth queue at the instant $(t_n - 0)$ of arrival of the nth customer and $x_j(n)$ is the number of servers at the jth station at the instant $(t_n - 0)$. A system classification in terms of the properties of the process $Q(n)$, $n \geq 1$, was introduced. The random process $Q(n)$, $n \geq 1$, is transient or zero-recurrent. The system queues were classified as well. This model is applicable to the transportation systems.

6.1.2. Nonidentical servers. A polling system of M nonidentical servers was considered in [11]. If the mth server at the nth visit to the queue Q_i meets with k customers, then the number of customers treated at this visit follows $f_{i,n}^{(m)}(k)$. The duration of customer service in the ith queue by the mth server is $\sigma_i^{(m)}$. After treating the queue Q_i, the mth server switches over to the queue Q_j in time $T_{ij,n}^{m}$ with the probability p_{ij}^{m}, where n is the number of the server passage from Q_i to Q_j. For system stability and instability which is understood in the sense of positive recurrence of the random process describing system behavior, the sufficient conditions were obtained. The problem of minimization of the system load ρ on the set of parameters $(p_1, \ldots, p_N)$ was solved. In [107] consideration was given to a more general $G/G/1$-type polling system where each queue is treated by one of the S possible disciplines. The route of the mth server is determined as follows. After visiting the queue Q_i with the sth service discipline, the server moves with the probability $p_{js,j's'}^{m}$ to the queue $Q_{j'}$ with the s'th service discipline, the time of inter-queue switchover being $\delta_{js,j's'}^{m}(n)$, where n is the number of the server switchover. A liquid model was constructed for this system, and the necessary and sufficient conditions for system stability and instability were established. The advantages of this fluid model were emphasized.

The $M/M/1$-type polling system with m servers and Bernoulli service discipline was studied in [257]. Each server polls the queues according to its polling table of length n. The study of this model was continued in [63] for the case where the times of customer service are distributed arbitrarily and the queues are treated using the exhaustive or gated service discipline. The mean times between server visits and approximate formulas for calculation of the mean waiting times and weighted sum were determined.

6.1.3. Coupled servers. In these polling systems, the servers jointly poll the queues and after connecting to a queue, treat the customers independently of each other so that the total number of served customers is in compliance with the chosen queue service discipline.

The $M/M/1$-type polling system with c servers and the zero switchover times was considered in [80]. After connection of the servers to a queue, its customers are treated in parallel. If at the polling instant their number is below $c-1$, then this queue is not served at the current cycle. If at treatment of a queue some servers vacate, they cannot help other servers. The servers poll queues in an arbitrary order making up the Hamiltonian cycle, that is, the cycle where the server visits each queue but only once. Let the system start from the initial state $Q(0) = (Q_1(0), \dots, Q_N(0))$, $Q_i(0) \ge c-1$, where $Q_i(t)$ is the number of customers queued in Q_i at time t. We denote by $\pi_0 = (i_1, i_2, \dots, i_N)$ the order of queue service.

If the servers visit queues cyclically ($\pi_0 = (1, 2, \dots, N)$), then for the *gated* service the mean time $\mathbf{M}(X_i(n_i))$ of server sojourn at the queue Q_i which had n_i customers at the polling instant obeys the equality

$$\mathbf{M}(X_i(n_i)) = \frac{n_i + c(H_c - 1)}{c} b_i, \tag{12}$$

where $H_c = \sum_{j=1}^{c} 1/j$.

Let now the order of queue service be arbitrary, $\pi_0 = (i_1, \dots, i_N)$. We denote by X_i the time of server sojourn at the queue Q_i in the cycle π_0 and by $S_i = \sum_{k=1}^{i} X_k$, the service time of the queues $Q_1, \dots, Q_i$ in the cycle, $Z_i = \mathbf{M}(S_i)$, $i = \overline{1, N}$. The system states at the instants of polling the queue Q_i meet the following system of stochastic equations:

$$Q_r(S_{i-1}) = \begin{cases} N_r(S_{i-1} - S_{r-1}), & r \le i-1 \\ Q_r(0) + N_r(S_{i-1}), & r \ge i, \end{cases}$$

where the random variable $N_r(t)$ has the Poisson distribution with the parameter $\lambda_r t$. The relation

$$\mathbf{M}(X_i \mid Q_i(S_{i-1}), S_{i-1}) = b_i(Q_i(S_{i-1}) + c(H_c - 1))/c$$

was established using (12) and the inequality $Q_i(S_{i-1}) \ge Q_i(0) \ge c-1$. By adding S_{i-1} to both sides of this equality, we obtain that the expected instant of completion of service in Q_i during the cycle π_0 satisfies the relation

$$Z_i = Z_{i-1} + b_i(Q_i(0) + \lambda_i Z_{i-1} + c(H_c - 1))/c$$

which can be rearranged in the difference equation

$$Z_0 = 0, \quad Z_i - (1 + \rho_i/c) Z_{i-1} = b_i(Q_i(0) + c(H_c - 1))/c, \quad i = \overline{1, N},$$

having the following solution:

$$Z_i = \sum_{k=1}^{i} b_k(Q_k(0) + c(H_c - 1))/c \prod_{r=k+1}^{i} (1 + \rho_r/c), \quad i = \overline{1, N}.$$

If the queues are served *exhaustively*, then the mean time of the Hamiltonian cycle π_0 starting with a visit to the queue Q_i obeys the equality

$$Z_m(\pi_0) = \sum_{i=1}^{m} b_i \frac{Q_i(0) + \gamma_c(\rho_i) + c(H_c - 1)}{c - \rho_i} \prod_{r=i+1}^{m} \left(1 + \frac{\rho_r}{c - \rho_r}\right), \quad m = \overline{1, N},$$

where γ_c is the following polynomial:

$$\gamma_c(x) = x\left[\frac{1}{(c-1)(c-2)} + \ldots + \frac{c-2}{2 \times 1}\right] + x^2\left[\frac{1}{(c-1)(c-2)(c-3)} + \cdots + \frac{c-3}{3 \times 2 \times 1}\right] + \ldots$$
$$+ x^{c-2}\left[\frac{1}{(c-1)(c-2)\ldots 1}\right].$$

Optimal Hamiltonian cycle. The Hamiltonian cycle was shown to take the least time if the servers visit queues in the ascending order of $\frac{Q_i(0)+c(H_c-1)}{\lambda_i}$ for the gated service and $\frac{Q_i(0)+c(H_c-1)+\gamma_c(\rho_i)}{\lambda_i}$ for the exhaustive service.

Similar results were obtained for systems with a nonzero server switchover. At that, the time of switchover between the queues Q_i and Q_j is the sum of the values of random variables θ_i (the time of disconnection from Q_i) and τ_j (the time of connection to Q_j), $i, j = \overline{1, N}$.

A polling system which resembles that of [80] and has an infinite number of servers and globally-gated service discipline was studied in [182]. At the beginning of each cycle, decision about the order of queue service in the cycle is made depending on the system state, provided that the order of polling makes up a Hamiltonian cycle. The distribution of the server warming-up time before service is given in addition to the distribution of the times of inter-queue switchover. The mean waiting times were obtained for this system.

The polling system with infinite number of servers, correlated customer flows, and exhaustive or gated disciplines was considered in [183]. The queues are treated randomly: the server takes the queue Q_i, $i = \overline{1, N}$, with the probability γ_i. The mean number of queued customers at an arbitrary time instant and the mean waiting times were determined.

The m-server system (Section 4.1, page sec4.1) with the cyclic queue polling and the Bernoulli queue service discipline was considered in [59]. For this model, a system of equations relating the generating functions of the stationary probabilities of the number of queued customers at the polling instants and those of server disconnection from the queue was obtained. Consideration was given to the cases where these functions can be determined explicitly.

7. NETWORKS OF POLLING SYSTEMS

A network consisting of polling systems with customers of two types and one server was considered in [214]. The customers arriving from recurrent flows are placed in the queue Q_1, after service in it they pass to the queue Q_2, and so on. Each type of queued customers is served using the exhaustive discipline. Connection of the server for treatment of customers of some type requires some random time. The distribution of the customer sojourn in the system was determined for the case of heavy traffic. The paper [220] determined the stability conditions for the polling networks.

8. CONTINUOUS POLLING MODELS

A circular polling system whose customers are placed on the circle along which the server moves with a constant speed and treats the encountered customers was considered in [172]. The arriving customers make a Poisson flow, they are uniformly distributed along the circle. As was shown in [171], under heavy traffic the distribution of the number of waiting customers approaches the gamma-distribution. The circular polling system with constant service time was examined in [48]. For this system, the LST's of the distribution function of the time between the instants of customer departures were determined. The model of polling with random speed of server motion was studied in [115].

A polling system on a compactum with the Poisson flow of customers was considered in [30]. The arriving customers are uniformly distributed over the compactum over which the server moves along a certain trajectory and treats the nearby customers. The paper [37] extended the results of [30] to the case of convex bounded multidimensional domain. The server has full information about locations of the customers. It was noted that the service disciplines such as first-come-first-served (the customers are treated in the order of their arrival) or service of the nearest customer that were used in the aforementioned publications are either inefficient or unjustified. Two new disciplines were suggested in [37]. The time of work is divided into cycles. According to the first discipline, the server treats only those customers which were in the system at the instant of cycle initiation. According to the second discipline, the server either treats the nearest of the customers that must be served in this cycle or moves along some trajectory and treats the nearby customers. The stability conditions, the mean cycle time, and the mean number of customers in the system were obtained.

9. ACKNOWLEDGMENTS

Special thanks are due to Profs. O.J. Boxma (Eindhoven Technological University, Netherlands), U. Yechiali (Tel Aviv University, Israel), and J.H. Dshalalow (Florida Technological Institute, Melbourne, USA) for the materials shared with us, as well as to Prof. V.V. Rykov for careful reading of the manuscript and helpful remarks.

REFERENCES

1. Al'bores, F.Kh. and Bocharov, P.P., Analysis of Two Limited Relative-priority Queues in the One-server Queuing System with Phase-type Distribution, *Avtom. Telemekh.*, 1993, no. 4, pp. 96–107.
2. Bakanov, A.S., Vishnevskii, V.M., and Lyakhov, A.I., A Method for Evaluating Performance of Wireless Communication Networks with Centralized Control, *Avtom. Telemekh.*, 2000, no. 4, pp. 97–105.
3. Vishnevskii, V.M., Wireless Networks for Broadband Access to the Internet Resources, *Elektrosvyaz'*, 2000, no. 10. pp. 9–13.
4. Vishnevskii, V.M., *Teoreticheskie osnovy proektirovaniya komp'yuternykh setei* (Theoretical Fundamentals of Computer Network Design), Moscow: Tekhnosfera, 2003.
5. Vishnevskii, V.M., Lyakhov, A.I., and Guzakov, N.N., Estimation of the Maximum Throughput of the Wireless Access to Internet, *Avtom. Telemekh.*, 2004, no. 9, pp. 52–70.
6. Vishnevskii, V.M., Lyakhov, A.I., Portnoi, S.L., and Shakhnovich, I.V., *Shirokopolosnye besprovodnye seti peredachi informatsii* (Broadband Wireless Information Transmission Networks), Moscow: Tekhnosfera, 2005.
7. Gavrilov, A.F. and Krasil'nikov, Yu.P., Cyclic Service with Direct Information Connection, *Avtom. Telemekh.*, 1976, no. 10, pp. 17–22.
8. Efrosinin, D.V. and Rykov, V.V., Numerical Study of the Optimal Control of a System with Heterogeneous Servers, *Avtom. Telemekh.*, 2003, no. 2, pp. 143–151.
9. Zhdanov, V.S. and Saksonov, E.A., Existence Conditions for Stable Modes in Cyclic Queuing Systems, *Avtom. Telemekh.*, 1979, no. 2, pp. 29–38.
10. Kitaev, M.Yu. and Rykov, V.V., Queuing System with Branching Flows of Secondary Customers, *Avtom. Telemekh.*, 1980, no. 9, pp. 52–61.
11. Kovalevskii, A.P., Positive Recurrence and Optimization of Polling Systems with Multiple Servers, in *Aktual'nye problemy sovremennoi matematiki*, Novosibirsk: NII MIOO NGU, 1997, vol. 3, pp. 75–86.
12. Klimov, G.P. and Mishkoi, G.K., *Prioritetnye sistemy obsluzhivaniya s orientatsiei* (Priority Queuing Systems with Orientation), Moscow: Mosk. Gos. Univ., 1979.
13. Lakontsev, D.V. and Semenova, O.V., Mathematical Models of Centralized Control in Wireless Networks IEEE 802.11, in *Raspredelennye komp'yuternye i telekommunikatsionnye seti (DCCN-2005)* (Distributed Computer and Telecommunication Networks (DCCN-2005)), Moscow: Tekhnosfera, 2005, pp. 77–83.
14. Nazarov, A.A. and Urazbaeva, S.U., Study of the Decomposed Model of the Multi-Packet Mode of Communication Network with DQDB Protocol, *Vest. Tomsk. Gos. Univ., Matematika. Kibernetika. Informatika*, 2002, no. 275, pp. 199–201.
15. Nazarov, A.A. and Urazbaeva, S.U., Study of Discrete-Time Queuing Systems and Their Application to Analysis of Fiber-Optic Communication Networks, *Avtom. Telemekh.*, 2002, no. 12, pp. 59–70.
16. Rykov, V.V., On the Monotonicity Conditions for the Optimal Control Policies for Queuing Systems, *Avtom. Telemekh.*, 1999, no. 9, pp. 92–106.

17. Rykov, V.V. and Verbitskii, S.N., Analysis of Some Algorithms for Polling Web-pages, *Vest. Ross. Univ. Druzhby Narodov, Prikladn. Mat. Inform.*, 2000, no. 1, pp. 96–104.

18. Saksonov, E.A., Study of Multi-Channel Closed Cyclic Queuing System, *Avtom. Telemekh.*, 1979, no. 12, pp. 80–86.

19. Saksonov, E.A., A Method for Calculation of the Marginal State Probabilities of the Cyclic Queuing Systems, *Avtom. Telemekh.*, 1997, no. 1, pp. 85–89.

20. Timofeev, E.A., Optimization of Mean Queue Lengths in a Queuing System with Branching Flows of Secondary Customers, *Avtom. Telemekh.*, 1995, no. 3, pp. 60–67.

21. Titenko, I.M., On Cyclically Served Multi-Channel Systems with Losses, *Avtom. Telemekh.*, 1984, no. 10, pp. 88–95.

22. Foss, S.G. and Chernova, N.I., On Polling Systems with Infinite Number of Stations, *Sib. Mat. Zh.*, 1996, vol. 37, no. 4, pp. 940–956.

23. Foss, S.G. and Chernova, N.I., Comparison Theorems and Ergodic Properties of the Polling Systems, *Probl. Peredachi Inf.*, 1996, no. 4, pp. 46–71.

24. Yakushev, Yu.F., On Optimal Protocol of Multiple Marker-Type Access in Local Computer Network, *Avtom. Telemekh.*, 1990, no. 10, pp. 125–134.

25. Adan, I.J.B.F., Boxma, O.J., and Resing, J.A.C., Queuing Models with Multiple Waiting Lines, *Queuing Syst.*, 2001, vol. 37, no. 1–3, pp. 65–98.

26. Afanassieva, L.G., Delcoigne, F., and Fayolle, G., On Polling Systems where Servers Wait for Customers, *Markov Process and Related Fields*, 1997, vol. 3, no. 4, pp. 527–545.

27. Ajmone Marsan, M., Donatelli, S., and Neri, F., GSPN Models of Markovian Multiserver MultiQueue Systems, *Performance Evaluat.*, 1990, vol. 11, no. 4, pp. 227–240.

28. Almási, B., A Queuing Model for a Non-homogeneous Polling System Subject to Breakdowns, *Ann. Univ. Sci. Budapest, Sect. Comp.*, 1999, vol. 18, pp. 11–23.

29. Altman, E., Blanc, H., Khamisy, A., and Yechiali, Y., Gated-type Polling Systems with Walking and Switch-in Times, *Commun. Stat.: Stochastic Models*, 1994, vol. 10, no. 4, pp. 741–763.

30. Altman, E. and Foss, S., Polling on a Graph with General Arrival and Service Time Distribution, *Lett. Oper. Res.*, 1997, vol. 20, no. 4, pp. 187–194.

31. Altman, E., Foss, S., Riehl, E., and Stidham, S., Performance Bounds and Pathwise Stability for Generalized Vacation and Polling Systems, *Oper. Res.*, 1998, vol. 46, no. 1, pp. 137–148.

32. Altman, E., Gaujal, B., and Hordijk, A., Optimal Open-loop Control of Vacations, Polling and Service Assignment, *Queuing Syst.*, 2000, vol. 36, no. 4, pp. 303–325.

33. Altman, E., Khamisy, A., and Yechiali, U., On Elevator Polling with Globally Gated Regime, *Queuing Syst.*, 1992, vol. 11, no. 1–2, pp. 85–90.

34. Altman, E. and Kofman, D., Bounds for Performance Measures of Token Rings, *IEEE/ACM Trans. Networking*, 1996, vol. 4, no. 2, pp. 292–299.

35. Altman, E., Konstantopoulos, P., and Liu, Z., Stability, Monotonicity and Invariant Quantities in General Polling Systems, *Queuing Syst.*, 1992, vol. 11, no. 1–2, pp. 35–57.

36. Altman, E. and Kushner, H., Control of Polling in Presence of Vacations in Heavy Traffic with Applications to Satellite and Mobile Radio Systems, *SIAM J. Contr. Optimiz.*, 2002, vol. 41, no. 1, pp. 217–252.

37. Altman, E. and Levy, H., Queuing in Space, *Advances Appl. Prob.*, 1995, vol. 26, no. 4, pp. 1095–1116.

38. Altman, E. and Spieksma, F., Polling Systems: Moment Stability of Station Times and Central Limit Theorems, *Commun. Stat.: Stochastic Models*, 1996, vol. 12, no. 2, pp. 307–328.

39. Altman, E. and Yechiali, U., Cyclic Bernoulli Polling, *ZOR—Methods and Models Oper. Res.*, 1993, vol. 38, no. 1, pp. 55–76.

40. Altman, E. and Yechiali, U., Polling in a Closed Network, *Prob. Eng. Inf. Sci.*, 1994, vol. 8, pp. 327–343.

41. Baba, Y., Analysis of Batch Arrival Cyclic Service Multiqueue Systems with Limited Service Discipline, *J. Oper. Res. Soc. Japan*, 1991, vol. 34, no. 1, pp. 93–104.

42. Baxter, L.A., Harche, F., and Yechiali, U., A Note on Minimizing the Variability of the Waiting Times in a Globally-gated Elevator Polling System, *Naval Res. Logistics*, 1997, vol. 44, no. 6, pp. 605–611.

43. Bertsimas, D., The Achievable Region Method in the Optimal Control of Queuing Systems: Formulations, Founds and Policies, *Queuing Syst.*, 1995, vol. 21, no. 3–4, pp. 337–389.

44. Bertsimas, D. and Mourtzinou, G., Decomposition Results for General Polling Systems and Their Applications, *Queuing Syst.*, 1999, vol. 31, no. 3–4, pp. 295–316.

45. Bertsimas, D. and Niño-Mora, J., Optimization of Multiclass Queuing Networks with Changeover Times via the Achievable Region Approach: I. The Single-Station Case, *Math. Oper. Res.*, 1999, vol. 24, no. 2, pp. 331–361.

46. Bing, B., *Wireless Local Area Networks: The New Wireless Revolution*, New York: Wiley-Interscience, 2002.

47. Birman, A., Gail, H.R., Hantler, S.L., Rosber, Z., and Sidi, M., An Optimal Service Policy for Buffer Systems, *J. Association Comput. Machinery*, 1995, vol. 42, no. 3, pp. 641–657.

48. Bisdikian, C. and Merakos, L., Output Process from a Continuous Tokenring Local Area Network, *IEEE Trans. Commun.*, 1992, vol. 40, no. 12, pp. 1796–1799.

49. Blanc, J.P.C., A Numerical Approach to Cyclic-Service Queuing Models, *Queuing Syst.*, 1990, vol. 6, no. 2, pp. 173–188.

50. Blanc, J.P.C., An Algorithmic Solution of Polling Models with Limited Service Disciplines, *IEEE Trans. Commun.*, 1992, vol. 40, no. 7, pp. 1152–1155.

51. Blanc, J.P.C., Performance Evaluation of Polling Systems by Means of the Power-series Algorithm, *Ann. Oper. Res.*, 1992, vol. 35, no. 1–4, pp. 155–186.

52. Blanc, J.P.C., The Power-series Algorithm Applied to Cyclic Polling Systems, *Commun. Stat.: Stochastic Models*, 1991, vol. 7, no. 4, pp. 527–545.

53. Blanc, J.P.C., The Power-series Algorithm for Polling Systems with Time Limits, *Prob. Eng. Inf. Sci.*, 1998, vol. 12, pp. 221–237.

54. Blanc, J.P.C. and van Der Mei, R.D., Optimization of Polling Systems with Bernoulli Schedules, *Performance Evaluat.*, 1995, vol. 22, no. 2, pp. 139–158.

55. Blanc, J.P.C. and van der Mei, R.D., The Power-series Algorithm Applied to Polling Systems with a Dormant Server, in *The Fundamental Role of Teletraffic in the Evolution of Telecommunication Networks*, Labetoulle, J. and Roberts, J.W., Eds., Amsterdam: Elsevier, 1994, pp. 865–874.

56. Borovkov, A., Korshunov, D., and Schassberger, R., Ergodicity of a Polling Network with an Infinite Number of Stations, *Queuing Syst.*, 1999, vol. 32, no. 1, pp. 169–193.

57. Borovkov, A. and Schassberger, R., Ergodicity of a Polling Network, *Stochastic Proc. Appl.*, 1994, vol. 50, no. 2, pp. 253–262.

58. Borst, S.C., *Polling Systems*, Amsterdam: Stichting Mathematisch Centrum, 1996.

59. Borst, S.C., Polling Systems with Multiple Coupled Servers, *Queuing Syst.*, 1995, vol. 20, no. 3–4, pp. 369–394.

60. Borst, S.C. and Boxma, O.J., Polling Models with and without Switchover Times, *Oper. Res.*, 1997, vol. 47, pp. 536–543.

61. Borst, S.C., Boxma, O.J., Harink, J.H.A., and Huitema, G.B., Optimization of Fixed-time Polling Schemes, *Telecommun. Syst.*, 1994, vol. 3, pp. 31–59.

62. Borst, S.C., Boxma, O., and Levy, H., The Use of Service Limits for Efficient Operation of Multi-Station Single-medium Communication Systems, *IEEE/ACM Trans. Networking*, 1995, vol. 3, no. 5, pp. 602–612.

63. Borst, S.C. and van der Mei, R.D., Waiting Time Approximations for Multiple-server Polling Systems, *Performance Evaluat.*, 1998, vol. 31, no. 3–4, pp. 163–182.

64. Boxma, O.J., Models of Two Queues: A Few New Views, in *Teletraffic Analysis and Computer Performance Evaluation*, Boxma, O.J., Cohen, J.W., and Tijms, H., Eds., Amsterdam: Elsevier, 1986, pp. 75–98.

65. Boxma, O.J., Static Optimization of Queuing Systems, in *Recent Trends in Optimization Theory and Applications*, Agarwal, R.P., Ed., Singapore: World Scientific, 1995, pp. 1–16.

66. Boxma, O.J., Deng, Q., and Resing, J.A.C., Polling Systems with Regularly Varying Service and/or Switchover Times, *Advances Performance Anal.*, 2000, vol. 3, pp. 71–107.

67. Boxma, O.J. and Down, D.G., Dynamic Server Assignment in a Two-queue Model, *Eur. J. Oper. Res.*, 1997, vol. 103, pp. 595–609.

68. Boxma, O.J. and Groenendijk, W.P., Pseudo Conservation Laws in Cyclic-service Systems, *J. Appl. Prob.*, 1987, vol. 24, pp. 949–964.

69. Boxma, O.J., Groenendijk, W.P., and Weststrate, J.A., A Pseudoconservation Law for Service Systems with a Polling Table, *IEEE Trans. Commun.*, 1990, vol. 38, no. 10, pp. 1865–1870.

70. Boxma, O.J. and Kelbert, M., Stochastic Bounds for a Polling System, *Ann. Oper. Res.*, 1994, vol. 48, no. 1–4, pp. 295–310.

71. Boxma, O.J., Koole, G.M., and Mitrani, I., Polling Models with Threshold Switching, in *Quantitative Methods in Parallel Systems*, New York: Springer-Verlag, 1995, pp. 129–140.

72. Boxma, O.J., Levy, H., and Weststrate, J.A., Efficient Visit Frequencies for Polling Tables: Minimization of Waiting Cost, *Queuing Syst.*, 1991, vol. 9, no. 1–2, pp. 133–162.

73. Boxma, O.J., Levy, H., and Weststrate, J.A., Efficient Visit Orders for Polling Systems, *Performance Evaluat.*, 1993, vol. 18, no. 2, pp. 103–123.

74. Boxma, O.J., Levy, H., and Yechiali, U., Cyclic Reservation Schemes for Efficient Operation of Multiple-queue Single-server Systems, *Ann. Oper. Res.*, 1992, vol. 35, no. 1–4, pp. 187–208.

75. Boxma, O.J. and Meister, B., Waiting-time Approximations for Cyclic–Service Systems with Switchover Times, *Performance Evaluat.*, 1987, vol. 7, pp. 299–308.

76. Boxma, O.J., Schlegel, S., and Yechiali, U., Two-queue Polling Models with a Patient Server, *Ann. Oper. Res.*, 2002, vol. 112, no. 1–4, pp. 101–121.

77. Boxma, O.J., Weststrate, J.A., and Yechiali, U., A Globally Gated Polling System with Server Interruptions, and Applications to the Repairman Problem, *Prob. Eng. Inf. Sci.*, 1993, vol. 7, no. 2, pp. 187–208.

78. Brill, P.H. and Hlynka, M., A Single Server N-line Queue in which a Customer May Receive Special Treatment, *Commun. Stat.: Stochastic Models*, 1998, vol. 14, no. 4, pp. 905–931.

79. Browne, S. and Kella, O., Parallel Service with Vacations, *Oper. Res.*, 1995, vol. 43, no. 5, pp. 870–878.

80. Browne, S. and Weiss, G., Dynamic Priority Rules for Polling with Multiple Servers, *Oper. Res. Lett.*, 1992, vol. 12, no. 3, pp. 129–138.

81. Browne, S. and Yechiali, U., Dynamic Scheduling in Single Server Multiclass Service Systems with Unit Buffers, *Naval Res. Logistics*, 1991, vol. 38, no. 3, pp. 383–396.

82. Bruneel, H. and Kim, B.G., *Discrete-time Models for Communication Systems Including ATM*, Boston: Kluwer, 1993.

83. Bruno, R., Conti, M., and Gregory, E., Bluetooth: Architecture, Protocols and Scheduling Algorithms, *Cluster Comput.*, 2002, vol. 5, pp. 117–131.

84. Campbell, G.M., Cyclic Queuing Systems, *Eur. J. Oper. Res.*, 1991, vol. 51, no. 2, pp. 155–167.

85. Chakravarthy, S.R., Analysis of a Priority Polling System with Group Services, *Commun. Stat.: Stochastic Models*, 1998, vol. 14, no. 1, pp. 25–49.

86. Chakravarthy, S.R. and Thiagarajan, S., Two Parallel Finite Queues with Simultaneous Services and Markovian Arrivals, *J. Appl. Math. Stochastic Anal.*, 1997, vol. 10, no. 10, pp. 383–405.

87. Chang, K.-H., First-come-first-served Polling Systems, *Asia-Pacific J. Oper. Res.*, 2001, vol. 18, no. 1, pp. 1–11.

88. Chang, K.S., Stability Conditions for a Pipeline Polling Scheme in Satellite Communications, *Queuing Syst., Theory Appl.*, 1993, vol. 14, no. 3–4, pp. 339–348.

89. Chang, R.K.C. and Lam, S., A Novel Approach to Queue Stability Analysis of Polling Models, *Performance Evaluat.*, 2000, vol. 40, no. 1–3, pp. 27–46.

90. Chang, W. and Down, D.G., Exact Asymptotics for k_i-limited Exponential Polling Models, *Queuing Syst.*, 2002, vol. 42, no. 4, pp. 401–419.

91. Chiarawongse, J. and Srinivasan, M.M., On Pseudo-conservation Laws for the Cyclic Server System with Compound Poisson Arrivals, *Oper. Res. Lett.*, 1991, vol. 10, no. 8, pp. 453–459.

92. Choudhury, G.L. and Whitt, W., Computing Distributions and Moments in Polling Models by Numerical Transform Inversion, *Performance Evaluat.*, 1996, vol. 25, no. 4, pp. 267–292.

93. Chung, H., Un, C., and Jung, W., Performance Analysis of Markovian Polling Systems with Single Buffers, *Performance Evaluat.*, 1994, vol. 19, no. 4, pp. 303–315.

94. Coffman, E.G., Jr., Liu, Z., and Weber, R.R., Optimal Robot Scheduling for Web Search Engines, *J. Scheduling*, 1998, vol. 1, no. 1, pp. 15–29.

95. Coffman, E.G., Jr., Puhalskii, A.A., and Reiman, M.I., Polling Systems in Heavy Traffic: A Bessel Process Limit, *Math. Oper. Res.*, 1998, vol. 23, no. 2, pp. 257–304.

96. Coffman, E.G., Jr., Puhalskii, A.A., and Reiman, M.I., Polling Systems with Zero Switchover Times: A Heavy-Traffic Averaging Principle, *Ann. Appl. Prob.*, 1995, vol. 5, no. 3, pp. 681–719.

97. Cooper, R.B., Niu, S.-C., and Srinivasan, M.M., A Decomposition Theorem for Polling Models: The Switchover Times are Effectively Additive, *Oper. Res.*, 1996, vol. 44, no. 4, pp. 629–633.

98. Cooper, R.B., Niu, S.-C., and Srinavasan, M.M., Setups in Polling Models: Does it Make Sense to Set up if no Work is Waiting?, *J. Appl. Prob.*, 1999, vol. 36, pp. 585–592.

99. Cooper, R.B., Niu, S.-C., and Srinavasan, M.M., Some Reflections on the Renewal-theory Paradox in Queuing Theory, *J. Appl. Math. Stochastic Anal.*, 1998, vol. 11, no. 3, pp. 355–368.

100. Cooper, R.B., Niu, S.-C., and Srinivasan, M.M., When Does Forced Idle Time Improve Performance in Polling Models?, *Manag. Sci.*, 1998, vol. 44, no. 8, pp. 1079–1086.

101. Daganzo, C.F., Some Properties of Polling Systems, *Queuing Syst.*, 1990, vol. 6, no. 2, pp. 137–154.

102. de Souza e Silva, E., Gail, R.H., and Muntz, R.R., Polling Systems with Server Timeouts and Their Application to Token Passing Networks, *IEEE/ACM Trans. Networking*, 1995, vol. 3, no. 5, pp. 560–575.

103. Delcoigne, F. and Fayolle, G., Thermodynamical Limit and Propagation of Chaos in Polling Systems, *Markov Processes and Related Fields*, 1999, vol. 5, no. 1, pp. 89–124.

104. Delcoigne, F. and la Fortelle, A., Large Deviations Rate Function for Polling Systems, *Queuing Syst.*, 2002, vol. 41, no. 1–2, pp. 13–44.

105. Deng, Q., A Two-queue E/1-L Polling Model with Regularly Varying Service and/or Switchover Times, *Commun. Stat.: Stochastic Models*, 2003, vol. 19, no. 4, pp. 507–526.

106. Dou, C. and Chang, J.-F., Serving Two Correlated Queues with a Synchronous Server under Exhaustive Service Discipline and Nonzero Switchover Time, *IEEE Trans. Commun.*, 1991, vol. 39, no. 11, pp. 1582–1589.

107. Down, D., On the Stability of Polling Models with Multiple Servers, *J. Appl. Prob.*, 1998, vol. 35, no. 4, pp. 925–935.

108. Dror, H. and Yechiali, U., Closed Polling Models with Failing Nodes, *Queuing Syst.*, 2000, vol. 35, no. 1–4, pp. 55–81.

109. Duenyas, I., Gupta, D., and Olsen, T.L., Control of a Single-server Tandem Queuing System with Setups, *Oper. Res.*, 1998, vol. 46, no. 2, pp. 218–230.

110. Duenyas, I. and van Oyen, M.P., Heuristic Scheduling of Parallel Heterogeneous Queues with Set-ups, *Manag. Sci.*, 1996, vol. 42, no. 6, pp. 814–829.

111. Duenyas, I. and van Oyen, M.P., Stochastic Scheduling of Parallel Queues with Set-up Costs, *Queuing Syst.*, 1995, vol. 19, pp. 421–444.

112. Duffield, N.G., Exponents for the Tails of Distributions in Some Polling Models, *Queuing Syst.*, 1997, vol. 26, no. 1–2, pp. 105–119.

113. Eisenberg, M., The Polling System with a Stopping Server, *Queuing Syst.*, 1994, vol. 18, no. 3–4, pp. 387–431.

114. Eliazar, I., Gated Polling Systems with Levy Inflow and Inter-dependent Switchover Times: A Dynamical-Systems Approach, *Queuing Syst.*, 2005, vol. 49, no. 1, pp. 49–72.

115. Eliazar, I., The Snowblower Problem, *Queuing Syst.*, 2003, vol. 45, no. 4, pp. 357–380.

116. Eliazar, I., Fibich, G., and Yechiali, U., A Communication Multiplexer Problem: Two Alternating Queues with Dependent Randomly-timed Gated Regime, *Queuing Syst.*, 2002, vol. 42, no. 4, pp. 325–353.

117. Eliazar, I. and Yechiali, U., Polling under the Randomly-timed Gated Regime, *Commun. Stat.: Stochastic Models*, 1998, vol. 14, no. 1–2, pp. 79–93.

118. Fabian, O. and Levy, H., Pseudo-Cyclic Policies for Multi-queue Single Server Systems, *Ann. Oper. Res.*, 1994, vol. 48, no. 1–4, pp. 127–152.

119. Fayolle, G. and Lasgouttes, J.-M., A State-dependent Polling Model with Markovian Routing, *IMA Volumes Math. Appl.*, 1995, vol. 71, pp. 283–301.

120. Federgruen, A. and Katalan, Z., Approximating Queue Size and Waiting-time Distributions in General Polling Systems, *Queuing Syst.*, 1994, vol. 18, no. 3–4, pp. 353–386.

121. Feng, W., Kowada, M., and Adachi, K., A Two-queue Model with Bernoulli Service Schedule and Switching Times, *Queuing Syst.*, 1998, vol. 30, no. 3–4, pp. 405–434.

122. Feng, W., Kowada, M., and Adachi, K., Analysis of a Multi-server Queue with Two Priority Classes and (M, N)-threshold Service Schedule I: Non-preemptive Priority, *Int. Trans. Oper. Res.*, 2000, vol. 7, no. 6, pp. 653–671.

123. Feng, W., Kowada, M., and Adachi, K., Performance Analysis of a Two-queue Model with an (M, N)-threshold Service Schedule, *J. Oper. Res. Soc. Japan*, 2001, vol. 44, no. 2, pp. 101–124.

124. Feng, W., Kowada, M., and Adachi, K., Two-queue and Two-server Model with a Hysteretic Control Service Policy, *Sci. Math. Japonicae*, 2001, vol. 54, no. 1, pp. 93–107.

125. Fischer, M.J., Harris, C.M., and Xie, J., An Interpolation Approximation for Expected Wait in a Time-limited Polling System, *Comput. Oper. Res.*, 2000, vol. 27, no. 4, pp. 353–366.

126. Foss, S. and Kovalevskii, A., A Stability Criterion via Fluid Limits and Its Application to a Polling System, *Queuing Syst.*, 1999, vol. 32, no. 1–3, pp. 131–168.

127. Foss, S. and Last, G., On the Stability of Greedy Polling Systems with General Service Policies, *Prob. Eng. Inf. Sci.*, 1998, vol. 12, no. 1, pp. 49–68.

128. Foss, S. and Last, G., Stability of Polling Systems with Exhaustive Service Policies and State-dependent Routing, *Ann. Appl. Prob.*, 1996, vol. 6, no. 1, pp. 116–137.

129. Fournier, L. and Rosberg, Z., Expected Waiting Times in Cyclic Service Systems under Priority Disciplines, *Queuing Syst.*, 1991, vol. 9, no. 4, pp. 419–439.

130. Frigui, I. and Alfa, A.S., Analysis of a Discrete Time Table Polling System with MAP Input and Time-limited Service Discipline, *Telecommunication Syst.*, 1999, vol. 12, no. 1, pp. 51–77.

131. Frigui, I. and Alfa, A.S., Analysis of Time-limited Polling System, *Comput. Commun.*, 1998, vol. 21, no. 6, pp. 558–571.

132. Fuhrmann, S.W., A Decomposition Result for a Class of Polling Models, *Queuing Syst.*, 1992, vol. 11, no. 1–2, pp. 109–120.

133. Fuhrmann, S.W. and Moon, A., Queues Served in Cyclic Order with an Arbitrary Start-up Distribution, *Naval Res. Logistics*, 1990, vol. 37, no. 1, pp. 123–133.

134. Fuhrmann, S.W. and Wang, Y.T., Analysis of Cyclic Service Systems with Limited Service: Bounds and Approximations, *Performance Evaluat.*, 1988, vol. 8, pp. 35–54.

135. Gandhi, A.D. and Cassandras, C.G., Optimal Control of Polling Models for Transportation Applications, *Math. Comput. Modelling*, 1996, vol. 23, no. 11–12, pp. 1–23.

136. Georgiadis, L. and Szpankowski, W., Stability of Token Passing Rings, *Queuing Syst.*, 1992, vol. 11, no. 1–2, pp. 7–33.

137. Grasman, S.E., Olsen, T.L., and Birge, J.R., Finite Buffer Polling Models with Routing, *Eur. J. Oper. Res.*, 2005, vol. 165, no. 3, pp. 794–809.

138. Grelá-M'Poko, B., Mehmet Ali, M., and Hayes, J.F., Approximate Analysis of Asymmetric Single-Service Prioritized Token Passing Systems, *IEEE Trans. Commun.*, 1991, vol. 39, no. 7, pp. 1037–1040.

139. Grillo, D., Polling Mechanism Models in Communication Systems—Some Application Examples, in *Stochastic Analysis of Computer and Communication Systems*, Takagi, H., Ed., Amsterdam: North-Holland, 1990, pp. 659–698.

140. Groenendijk, W.P. and Levy, H., Performance Analysis of Transaction Driven Computer Systems via Queuing Analysis of Polling Models, *IEEE Trans. Comput.*, 1992, vol. 41, no. 4, pp. 455–466.

141. Günalay, Y. and Gupta, D., Polling System with Patient Server and State-dependent Setup Times, *IIE Trans.*, 1997, vol. 29, no. 6, pp. 469–480.

142. Günalay, Y. and Gupta, D., Threshold Start–up Control Policy for Polling Systems, *Queuing Syst.*, 1998, vol. 29, no. 2–4, pp. 399–421.

143. Gupta, D. and Buzacott, J.A., A Production System with Two Job Classes, Changeover Times and Revisitation, *Queuing Syst.*, 1990, vol. 6, no. 4, pp. 353–368.

144. Gupta, D. and Günalay, Y., Recent Advances in the Analysis of Polling Systems, in *Advances in Combinatorial Methods with Applications to Probability and Statistics, Special Edition*, Boston: Birkhauser, 1996, pp. 339–360.

145. Gupta, D., Günalay, Y., and Srinivasan, M.M., The Relationship Between Preventive Maintenance and Manufacturing System Performance, *Eur. J. Oper. Res.*, 2001, vol. 132, no. 1, pp. 146–162.

146. Gupta, D. and Srinivasan, M.M., Polling Systems with State-dependent Setup Times, *Queuing Syst.*, 1996, vol. 22, no. 3–4, pp. 403–423.

147. Harel, A. and Stulman, A., Polling, Greedy and Horizon Servers on a Circle, *Queuing Syst.*, 1995, vol. 43, no. 1, pp. 177–186.

148. Hirayama, T., Hong, S.J., and Krunz, M.M., A New Approach to Analysis of Polling Systems, *Queuing Syst.*, 2004, vol. 48, no. 1–2, pp. 89–102.

149. Hwang, L.-C. and Chang, C.-J., An Exact Analysis of an Asymmetric Polling System with Mixed Service Discipline and General Service Order, *Comput. Commun.*, 1997, vol. 20, pp. 1292–1300.

150. Ibe, O.C., Analysis of Polling Systems with Mixed Service Discipline, *Commun. Stat.: Stochastic Models*, 1990, vol. 6, no. 4, pp. 667–689.

151. Ibe, O.C. and Cheng, X., Approximate Analysis of Asymmetric Single–service Token Passing Systems, *IEEE Trans. Commun.*, 1989, vol. 37, no. 6, pp. 572–577.

152. Ibe, O.C. and Trivedi, K.S., Stochastic Petri Net Models of Polling Systems, *IEEE J. Selected Areas Commun.*, 1990, vol. 8, no. 9, pp. 1649–1657.

153. Ibe, O.C. and Trivedi, K.S., Two Queues with Alternating Service and Server Breakdown, *Queuing Syst.*, 1990, vol. 7, no. 3, pp. 253–268.

154. Itai, A. and Rosberg, Z., A Golden Ratio Control Policy for a Multiple-acess Channel, *IEEE Trans. Automat. Control.*, 1984, vol. 29, pp. 712–718.

155. Jiang, Y., Tham, C.-K., and Ko, C.-C., Delay Analysis of a Probabilistic Priority Discipline, *Eur. Trans. Telecommun.*, 2002, vol. 13, no. 6, pp. 563–577.

156. Jirachiefpattana, A., County, P., Dillon, T.S., and Lai, R., Performance Evaluation of PC Routers Using a Single-server Multi-queue System with a Reflection Technique, *Comput. Commun.*, 1997, vol. 20, no. 1, pp. 1–10.

157. Jung, W.Y. and Un, C.K., Analysis of a Finite-buffer Polling System with Exhaustive Service Based on Virtual Buffering, *IEEE Trans. Commun.*, 1994, vol. 42, no. 12, pp. 3144–3149.

158. Karvelas, D., Leon-Garcia, A., Delay Analysis of Various Service Disciplines in Symmetric Token Passing Networks, *IEEE Trans. Commun.*, 1993, vol. 41, no. 9, pp. 1342–1355.

159. Katayama, T., Performance Analysis and Optimization of a Cyclic-Service Tandem Queuing System with Multi-class Customers, *Comput. Math. Appl.*, 1992, vol. 24, no. 1/2, pp. 25–33.

160. Khalid, M., Vyavahare, P.D., and Kerke, H.B., Analysis of Asymmetric Polling Systems, *Comput. Oper. Res.*, 1997, vol. 42, no. 4, pp. 317–333.

161. Khamisy, A., Altman, E., and Sidi, M., Polling Systems with Synchronization Constraints, *Ann. Oper. Res.*, 1992, vol. 35, no. 1–4, pp. 231–267.

162. Khamisy, A. and Sidi, M., Discrete-time Priority Queues with Two–state Markov Modulated Arrivals, *Commun. Stat.: Stochastic Models*, 1992, vol. 8, no. 2, pp. 337–357.

163. Kim, E. and van Oyen, M.P., Beyond the $c\mu$ Rule: Dynamic Scheduling of a Two-class Loss Queue, *Math. Methods Oper. Res.*, 1997, vol. 48, no. 1, pp. 17–36.

164. Kim, E., van Oyen, M.P., and Rieders, M., General Dynamic Programming Algorithms Applied to Polling Systems, *Commun. Stat.: Stochastic Models*, 1998, vol. 14, no. 5, pp. 1197–1221.

165. Kofman, D. and Yechiali, U., Polling with Stations Breakdowns, *Performance Evaluat.*, 1996, vol. 27–28, no. 4, pp. 647–672.

166. Konheim, A.G. and Meister, B., Waiting Lines and Times in a System with Polling, *J. ACM*, 1974, vol. 21, no. 3, pp. 470–490.

167. Kohneim, A.G., Levy, H., and Srinivasan, M.M., Descendant Set: An Efficient Approach for the Analysis of Polling Systems, *IEEE Trans. Commun.*, 1994, vol. 42, no. 2–4, pp. 1245–1253.

168. Koole, G., Assigning a Single Server to Inhomogeneous Queues with Switching Costs, *Theoretical Comput. Sci.*, 1997, vol. 182, no. 1–2, pp. 203–216.

169. Koole, G. and Nain, Ph., On the Value Function of a Priority Queue with an Application to a Controlled Polling Model, *Queuing Syst.*, 2000, vol. 34, no. 1–4, pp. 199–214.

170. Kopsel, A., Ebert, J.-P., and Wolisz, A., A Performance Comparison of Point and Distributed Function of an IEEE 802.11 WLAN in the Presence of Real-time Requirements, *Proc. Inf. Workshop MoMuc2000*, Waseda, Japan, October 2000.

171. Kroese, D.P., Heavy Traffic Analysis for Continuous Polling Models, *J. Appl. Prob.*, 1997, vol. 34, no. 3, pp. 720–732.

172. Kroese, D.P. and Schmidt, V., A Continuous Polling System with General Service Times, *Ann. Appl. Prob.*, 1992, vol. 2, no. 4, pp. 906–927.

173. Kudoh, S., Takagi, H., and Hashida, O., Second Moments of the Waiting Time in Symmetric Polling Systems, *J. Operat. Res. Soc. Japan*, 2000, vol. 43, no. 2, pp. 306–316.

174. Landry, R. and Stavrakakis, I., Queuing Study of 3–priority Policy with Distinct Service Strategies, *IEEE/ACM Trans. Networking*, 1993, vol. 1, no. 5, pp. 576–589.

175. Langaris, C., A Polling Model with Retrial Customers, *J. Oper. Res. Soc. Japan*, 1997, vol. 40, no. 4, pp. 489–508.

176. Langaris, C., Gated Polling Models with Customers in Orbit, *Math. Comput. Modeling*, 1999, vol. 30, no. 3–4, pp. 171–187.

177. Langaris, C., Markovian Polling Systems with Mixed Service Discipline and Retrial Customers, *TOP*, 1999, vol. 7, pp. 305–322.

178. Lee, D.-S., A Two-queue Model with Exhaustive and Limited Service Disciplines, *Commun. Stat.: Stochastic Models*, 1996, vol. 12, no. 2, pp. 285–305.

179. Lee, D.–S., Analysis of a Two-queue Model with Bernoulli Schedules, *J. Appl. Prob.*, 1997, vol. 34, no. 1, pp. 176–191.

180. Lee, D-S. and Sengupta, B., An Approximate Analysis of a Cyclic Server Queue with Limited Service and Reservations, *Queuing Syst.*, 1992, vol. 11, no. 1–2, pp. 153–178.

181. Lee, D-S. and Sengupta, B., Queuing Analysis of a Threshold Based Priority Scheme for ATM Networks, *IEEE Trans. Networking*, 1993, vol. 1, no. 6, pp. 709–717.

182. Lee, T., Analysis of Infinite Servers Polling Systems with Correlated Input Process and State Dependent Vacations, *Eur. J. Oper. Res.*, 1999, vol. 115, no. 2, pp. 392–412.

183. Lee, T., Analysis of Random Polling Systems with Infinite Coupled Servers and Correlated Input Process, *Comput. Oper. Res.*, 2003, vol. 30, no. 13, pp. 2003–2020.

184. Lee, T., A Closed Form Solution for the Asymmetric Random Polling System with Correlated Levy Input Process, *Math. Oper. Res.*, 1997, vol. 22, no. 2, pp. 432–457.

185. Lee, T., Models for Design and Control of Single Server Polling Computer and Communication Systems, *Oper. Res.*, 1998, vol. 46, pp. 515–531.

186. Lee, T. and Sunjaya, J., Exact Analysis of Asymmetric Random Polling Systems with Single Buffers and Correlated Levy Input Process, *Queuing Syst.*, 1996, vol. 23, no. 3–4, pp. 131–156.

187. Leung, K.K., Cyclic Service Systems with Non-preemptive, Time-limited Service, *IEEE Trans. Commun.*, 1994, vol. 42, no. 8, pp. 2521–2524.

188. Leung, K.K., Cyclic Service Systems with Probabilistically-limited Service, *IEEE J. Selected Areas Commun.*, 1991, vol. 9, no. 2, pp. 185–193.

189. Levy, H., Analysis of Cyclic Polling Systems with Binomial-gated Service, in *Performance of Distributed Parallel Systems*, Amsterdam: Elsevier, 1989, pp. 127–139.

190. Levy, H., Binomial-gated Service: A Method for Effective Operation and Optimization of Polling Systems, *IEEE Trans. Commun.*, 1991, vol. 39, no 9, pp. 1341–1350.

191. Levy, H. and Kleinrock, L., Polling Systems with Zero Switch-over Periods: A General Method for Analysis the Expected Delay, *Performance Evaluat.*, 1991, vol. 13, no. 2, pp. 97–107.

192. Levy, H. and Sidi, M., Polling Systems: Applications, Modeling and Optimization, *IEEE Trans. Commun.*, 1990, vol. 38, no. 10, pp. 1750–1760.

193. Levy, H. and Sidi, M., Polling Systems with Simultaneous Arrivals, *IEEE Trans. Commun.*, 1991, vol. 39, no. 6, pp. 823–827.

194. Levy, H., Sidi, M., and Boxma, O.J., Dominance Relations in Polling Systems, *Queuing Syst.*, 1990, vol. 6, no. 2, pp. 155–171.

195. Liu, Z. and Nain, P., Optimal Scheduling in Some Multi-queue Single-server Systems, *IEEE Trans. Automat. Control*, 1992, vol. 37, no. 2, pp. 247–252.

196. Liu, Z., Nain, P., and Towsley, D., On Optimal Polling Policies, *Queuing Syst.*, 1992, vol. 11, no. 1–2, pp. 59–83.

197. Lye, K. and Seah, K., Random Polling Scheme with Priority, *Electronic Lett.*, 1992, vol. 28, no. 14, pp. 1290, 1291.

198. Magalhaes, M.N., McNickle, D.C., and Salles, M.C.B., Outputs from a Loss System with Two Stations and a Smart (Cyclic) Server, *Investigacion Oper.*, 1998, vol. 16, no. 1–3, pp. 111–126.

199. Markowitz, D.M. and Wein, L.M., Heavy Traffic Analysis of Dynamic Cyclic Policies: A Unified Treatment of the Single Machine Scheduling Problem, *Oper. Res.*, 2001, vol. 49, no. 2, pp. 246–270.

200. Massoulie, L., Stability of Non-Markovian Polling Systems, *Queuing Syst.*, 1995, vol. 21, no. 1–2, pp. 67–95.

201. Menshikov, M. and Zuyev, S., Polling Systems in Critical Regime, *Stoch. Proc. Appl.*, 2001, vol. 92, pp. 201–218.

202. Miorandi, D., Zanella, A., and Pierobon, G., Performance Evaluation of Bluetooth Polling Schemes: An Analytical Approach, *ACM Mobile Networks Appl.*, 2004, vol. 9, no. 2, pp. 63–72.

203. Murata, M., Shiomoto, K., and Miyahara, H., Performance Analysis of Token Ring Networks with a Reservation Priority Discipline, *IEEE Trans. Commun.*, 1990, vol. 38, no. 10, pp. 1844–1853.

204. Nakdimon, O. and Yechiali, U., Polling Systems with Breakdowns and Repairs, *Eur. J. Oper. Res.*, 2001, vol. 149, no. 3, pp. 588–613.

205. Olsen, T.L., Approximations for the Waiting Time Distribution in Polling Models with and without State-dependent Setups, *Oper. Res. Lett.*, 2001, vol. 28, no. 3, pp. 113–123.

206. Olsen, T.L., Limit Theorems for Polling Models with Increasing Setups, *Prob. Eng. Inf. Sci.*, 2001, vol. 15, no. 1, pp. 35–55.

207. Olsen, T.L. and van Der Mei, R.D., Polling Systems with Periodic Server Routing in Heavy Traffic: Distribution of the Delay, *J. Appl. Prob.*, 2003, vol. 40, no. 2, pp. 305–326.

208. Olsen, T.L. and van Der Mei, R.D., Polling Systems with Periodic Server Routing in Heavy-Traffic: Renewal Arrivals, *Oper. Res. Lett.*, 2005, vol. 33, no. 1, pp. 17–25.

209. Ozawa, T., Alternating Service Queues with Mixed Exhaustive and k-limited Service, *Performance Evaluat.*, 1990, vol. 11, no. 3, pp. 165–175.

210. Park, B.U., Ryu, W., Kim, D.-U., Lee, B.L., and Chung, J.-W., Two Priority Class Polling Systems with Batch Poisson Arrivals, *Korean Commun. Stat.*, 1999, vol. 6, no. 3, pp. 881–891.

211. Peköz, E., More on Using Forced to Idle Time to Improve Performance in Polling Models, *Prob. Eng. Inf. Sci.*, 1999, vol. 13, no. 4, pp. 489–496.

212. Qiao, D., Choi, S., Soomoro, A., and Shin, K.G., Energy-efficient PCF Operation of IEEE 802.11a Wireless LAN, *Proc. INFOCOM 2002*, New York, June 2002.

213. Reiman, M. and Wein, L., Dynamic Scheduling of a Two-class Queue with Setups, *Oper. Res.*, 1998, vol. 46, no. 4, pp. 532–547.

214. Reiman, M. and Wein, L., Heavy Traffic Analysis of Polling Systems in Tandem, *Oper. Res.*, 1999, vol. 47, no. 4, pp. 524–534.

215. Resing, J.A.C., Polling Systems and Multitype Branching Processes, *Queuing Syst.*, 1993, vol. 13, no. 4, pp. 409–426.

216. Rubin, I. and Tsai, Z., Performance of Token Schemes Supporting Delay Constrained Priority Traffic Streams, *IEEE Trans. Commun.*, 1990, vol. 38, no. 11, pp. 1994–2003.

217. Rykov, V.V., Monotone Control of Queuing Systems with Heterogeneous Servers, *Queuing Syst.*, 2001, vol. 37, no. 4, pp. 391–403.

218. Ryu, W., Jun, K.P., Kim, D.W., and Park, B.U., Waiting Times in Priority Polling Systems with Batch Poisson Arrivals, *Korean Commun. Stat.*, 1998, vol. 5, no. 3, pp. 809–817.

219. Sarkar, D., Zangwill, W.I., Variance Effects in Cyclic Production Systems, *Manag. Sci.*, 1991, vol. 37, no. 4, pp. 444–453.

220. Schassberger, R., Stability of Polling Networks with State-dependent Server Routing, *Prob. Eng. Inf. Sci.*, 1995, vol. 9, no. 4, pp. 539–550.

221. Sharafali, M., Co, H.C., and Goh, M., Production Scheduling in a Flexible Manufacturing System under Random Demand, *Eur. J. Oper. Res.*, 2004, vol. 158, no. 1, pp. 89–102.

222. Sharma, V., Stability and Continuity of Polling Systems, *Queuing Syst.*, 1994, vol. 16, no. 1–2, pp. 115–137.

223. Shimogawa, S. and Takahashi, T., A Note on the Pseudo-conservation Law for a Multi-queue with Local Priority, *Queuing Syst.*, 1992, vol. 11, no. 1–2, pp. 145–151.

224. Shiozawa, Y., Takine, T., Takahashi, Y., and Hasegawa, T., Analysis of a Polling System with Correlated Input, *Computer Networks ISDN Syst.*, 1990, vol. 20, no. 1–5, pp. 297–308.

225. Sidi, M., Levy, H., and Fuhrmann, S.W., A Queuing Network with a Single Cyclically Roving Server, *Queuing Syst.*, 1992, vol. 11, no. 1–2, pp. 121–144.

226. Singh, M.P. and Srinivasan, M.M., Exact Analysis of the State Dependent Polling Model, *Queuing Syst.*, 2002, vol. 41, no. 4, pp. 371–399.

227. Srinivasan, M.M., Nondeterministic Polling Systems, *Manag. Sci.*, 1991, vol. 37, no. 6, pp. 667–681.

228. Srinivasan, M.M., Niu, S.-C., and Cooper, R.B., Relating Polling Models with Nonzero Switchover Times, *Queuing Syst.*, 1995, vol. 19, pp. 149–168.

229. Stavrakakis, I. and Tsakiridou, S., Study of a Class of Partially Ordered Service Strategies for a System of Two Discrete-time Queues, *Performance Evaluat.*, 1997, vol. 29, no. 1, pp. 15–33.

230. Suk, J.B. and Cassandras, C.G., Optimal Scheduling of Two Competing Queues with Blocking, *IEEE Trans. Automat. Control*, 1991, vol. 36, no. 9, pp. 1086–1091.

231. Suzuki, S. and Yamashita, H., Mean Waiting Times of the Alternating Traffic Starting Delays, *J. Oper. Res. Soc. Japan*, 1998, vol. 41, no. 3, pp. 442–454.

232. Takagi, H., Analysis and Applications of Polling Models, in *Performance Evaluation: Origins and Directions. Lecture Notes Comput. Sci.*, Haring, G., Lindemann, Ch., and Reiser, M., Eds., 2000, vol. 1769, pp. 423–442.

233. Takagi, H., Analysis of an $M/G/1//N$ Queue with Multiple Server Vacations, and Its Application to a Polling Model, *J. Oper. Res. Soc. Japan*, 1992, vol. 35, pp. 300–315.

234. Takagi, H., *Analysis of Polling Systems*, Cambridge: MIT Press, 1986.

235. Takagi, H., Applications of Polling Models to Computer Networks, *Comput. Networks ISDN Syst.*, 1991, vol. 22, no. 3, pp. 193–211.

236. Takagi, H., Queuing Analysis of Polling Systems, *ACM Comput. Surveys*, 1988, vol. 20, no. 1, pp. 5–28.

237. Takagi, H., Queuing Analysis of Polling Models: An Update, in *Stochastic Analysis of Computer and Communication Systems*, Takagi, H., Ed., Amsterdam: North-Holland, 1990, pp. 267–318.

238. Takagi, H., Queuing Analysis of Polling Models: Progress in 1990–1994, in *Frontiers in Queuing*, Dshalalow, J.H., Ed., Boca Raton: CRC, 1997, pp. 119–146.

239. Takahashi, Y., Fujimoto, K., and Makimoto, N., Geometric Decay of the Steady-state Probabilities in a Quasi-birth-and-death Process with a Countable Number of Phases, *Commun. Stat.: Stochastic Models*, 2001, vol. 17, no. 1, pp. 1–24.

240. Takahashi, Y. and Kumar, B.K., Pseudo-conservation Law for Discrete-time Multi-queue System with Priority Disciplines, *J. Oper. Res. Soc. Japan*, 1995, vol. 38, no. 4, pp. 450–466.

241. Takine, T. and Hasewaga, T., A Cyclic-Service Finite Source Model with Round-robin Scheduling, *Queuing Syst.*, 1992, vol. 11, no. 1–2, pp. 91–108.

242. Takine, T. and Hasewaga, T., Average Waiting Time of a Symmetrical Polling System under Bernoulli Scheduling, *Oper. Res. Lett.*, 1991, vol. 10, no. 9, pp. 535–539.

243. Takine, T., Takagi, H., and Hasegawa, T., Sojourn Times in Vacation and Polling Systems with Bernoulli Feedback, *J. Appl. Prob.*, 1991, vol. 28, no. 2, pp. 422–432.

244. Takine, T., Takagi, H., Takahashi, Y., and Hasegawa, T., Analysis of Asymmetric Single-buffer Polling and Priority Systems without Switchover Times, *Performance Evaluat.*, 1990, vol. 11, no. 4, pp. 253–264.

245. Takine, T., Takahashi, Y., and Hasegawa, T., Modeling and Analysis of a Single-buffer Polling System Interconnected with External Networks, *INFOR*, 1990, vol. 28, no. 3, pp. 166–177.

246. Tassiulas, L. and Ephremides, A., Dynamic Server Allocation to Parallel Queues with Randomly Varying Connectivity, *IEEE Trans. Inf. Theory*, 1993, vol. 39, no. 2, pp. 466–478.

247. Tedijanto, T.E., Exact Results for the Cyclic-service Queue with a Bernoulli Schedule, *Performance Evaluat.*, 1990, vol. 11, no. 2, pp. 107–115.

248. Tran-Gia, P., Analysis of Polling Systems with General Input Process and Finite Capacity, *IEEE Trans. Commun.*, 1992, vol. 40, no. 2, pp. 337–344.

249. Tseng, K.H. and Hsiao, M.-T.T., Optimal Control of Arrivals to Token Ring Networks with Exhaustive Service Discipline, *Oper. Res.*, 1995, vol. 43, no. 1, pp. 89–101.

250. van der Heijden, M.C., Harten, A., and Ebben, M.J.R., Waiting Times at Periodically Switched One-way Traffic Lanes—A Periodic, Two-queue Polling System with Random Setup Times, *Prob. Eng. Inf. Sci.*, 2001, vol. 15, no. 4, pp. 495–517.

251. van der Mei, R.D., Delay in Polling Systems with Large Switch-over Times, *J. Appl. Prob.*, 1999, vol. 36, no. 1, pp. 232–243.

252. van der Mei, R.D., Distribution of the Delay in Polling Systems in Heavy Traffic, *Performance Evaluat.*, 1999, vol. 38, no. 2, pp. 133–148.

253. van der Mei, R.D., Polling Systems in Heavy Traffic: Higher Moments of the Delay, *Queuing Syst.*, 1999, vol. 31, no. 3–4, pp. 265–294.

254. van der Mei, R.D., Polling Systems with Periodic Server Routing in Heavy Traffic, *Commun. Stat.: Stochastic Models*, 1999, vol. 15, no. 2, pp. 273–297.

255. van der Mei, R.D., Polling Systems with Switch-over Times under Heavy Load: Moments of the Delay, *Queuing Syst.*, 2000, vol. 36, no. 4, pp. 381–404.

256. van der Mei, R.D., Waiting-time Distributions in Polling Systems with Simultaneous Batch Arrivals, *Ann. Oper. Res.*, 2002, vol. 113, no. 1–4, pp. 155–173.

257. van der Mei, R.D. and Borst, S.C., Analysis of Multiple-server Polling Systems by Means of the Power-series Algorithm, *Commun. Stat.: Stochastic Models*, 1997, vol. 13, no. 2, pp. 339–369.

258. van der Mei, R.D. and Levy, H., Expected Delay Analysis of Polling Systems in Heavy Traffic, *Advances Appl. Prob.*, 1998, vol. 30, no. 2, pp. 586–602.

259. van der Mei, R.D. and Levy, H., Polling Systems in Heavy Traffic: Exhaustiveness of Service Policies, *Queuing Syst.*, 1997, vol. 27, no. 3–4, pp. 227–250.

260. van der Wal, J. and Yechiali, U., Dynamic Visit-order Rules for Batch-service Polling, *Prob. Eng. Inf. Sci.*, 2003, vol. 17, no. 3, pp. 351–367.

261. van Mieghem, J.A., Dynamic Scheduling with Convex Delay Costs: The Generalized $c\mu$ Rule, *Ann. Appl. Prob.*, 1995, vol. 5, no. 3, pp. 809–833.

262. van Oyen, M.P., Monotonicity of Optimal Performance Measures for Polling Systems, *Prob. Eng. Inf. Sci.*, 1997, vol. 11, no. 2, pp. 219–228.

263. van Oyen, M.P. and Teneketzis, D., Optimal Batch Service of a Polling System under Partial Information, *Zeitschrift Oper. Res.*, 1996, vol. 44, no. 3, pp. 401–419.

264. Vishnevsky, V.M. and Lyakhov, A.I., Adaptive Features of IEEE 802.11 Protocol: Utilization, Tuning and Modifications, *Proc. of 8th HP-OVUA Conf.*, Berlin, June 2001.

265. Vishnevsky, V.M., Lyakhov, A.I., and Bakanov, A.S., Method for Performance Evaluation of Wireless Networks with Centralized Control, *Proc. Int. Conf. "Distributed Computer Communication Networks (Theory and Applications)" (DCCN'99)*, Tel-Aviv, Israel, November 9–13, 1999, pp. 189–194.

266. Vishnevsky, V.M., Lyakhov, A.I., and Guzakov, N.N., An Adaptive Polling Strategy for IEEE 802.11 PCF, *Proc. 7th Int. Symp. on Wireless Personal Multimedia Communications (WPMC'04)*, Abano Terme, Italy, September 12–15, 2004, vol. 1, pp. 87–91.

267. Weststrate, J.A. and van der Mei, R.D., Waiting Times in a Two-queue Model with Exhaustive and Bernoulli Service, *Zeitschrift für Operations Research (ZOR)—Math. Methods Oper. Res.*, 1994, vol. 40, no. 3, pp. 289–303.

268. Xia, C.H., Michailidis, G., Bambos, N., and Glynn, P.W., Optimal Control of Parallel Queues with Batch Service, *Prob. Eng. Inf. Sci.*, 2002, vol. 16, no. 3, pp. 289–307.

269. Yechiali, U., Analysis and Control of Polling Systems, in *Performance Evaluation of Computer and Communication Systems*, Donatiello, L. and Nelson, R., Eds., Berlin: Springer, 1993, pp. 630–650.

270. Yechiali, U. and Armony, R., Polling Systems with Permanent and Transient Customers, *Commun. Stat.: Stochastic Models*, 1999, vol. 15, no. 3, pp. 395–427.

271. Ziouva, E. and Antonakopoulos, T., Improved IEEE 802.11 PCF Performance Using Silence Detection and Cyclic Shift on Stations Polling, *IEE Proc. Commun.*, 2003, vol. 150, no. 1, pp. 45–51.

272. Ziouva, E. and Antonakopoulos, T., Efficient Voice Communications over IEEE802.11 WLANs Using Improved PCF Procedures, *Proc. INC*, Plymouth, 2002/2007.

273. Ziouva, E. and Antonakopoulos, T., Improved IEEE802.11 PCF Performance Using Silence Detection and Cyclic Shift on Stations Polling, *IEE Proc. Commun.*, 2003, vol. 150, no. 1, pp. 45–51.

This paper was recommended for publication by V.V. Rykov, a member of the Editorial Board