Controllable Queueing Systems: From the Very Beginning up to Nowadays[1]

V. Rykov

Doctor of Sci, Professor, Department of Applied Mathematics & Computer Modeling, Gubkin Russian State Oil & Gas University, Leninsky Prospect, 65, 119991 Moscow, Russia.
e-mail: vladimir_rykov@mail.ru

Abstract

The present paper represents a review of the Controllable Queueing Systems theory development from the very beginning up to nowadays. The main stages of this theory development are considered. Some new problems are mentioned. The review is devoted to those, who are interested in the creation and the development of the Controllable Queueing Systems theory from its generation up to nowadays, and who want to understand the tendency of its development and the new directions and problems of its study.

Keywords: *Controllable queueing systems, Markov decision processes, Optimality principle, monotonicity of optimal policies*

1 Introduction

The theory of Controllable Queueing Systems (CQS) is a special direction of investigations of a general theory of controllable stochastic processes from one side, and of a Queueing Theory (QT) from another side. The theory of controllable stochastic processes is a special topic, which we will not touch here, and will fix on CQS. Some papers devoted to the problems of Queueing Systems (QS) control have been arisen almost simultaneously with the first works about QS, but the special approach to CQS has been done by Rykov [48] in 1975. Several monographs devoted to this problem arisen thereafter [25, 26, 31, 64, 67].

The preliminary results of this theory development one can find in [48]. In this paper, we concentrate our attention on the new results and approaches in this theory, nevertheless some initial principal results, on which the theory is based also should be reminded a little bit.

The paper is organized as follows.

In the next section the definition of CQS will be done and some examples of CQS will be proposed. The elements of the theory of Discrete Time Controllable Semi-regenerative Processes (DTCSRP), which serves as a base for the CQS study, will be considered then. The optimality principle for CQS as a result of this theory development and the problems of the real optimal rules for CQS calculation including numerical methods of the optimal policies calculation will be discussed in the next two sections. The qualitative properties of the optimal policies in this section also considered. Sections 5 – 8 are devoted to special CQS. At least in the Conclusion some new problems, approaches and new problems settings will be discussed.

2 CQS. Definition and main properties

2.1 Definitions

In this review, we will use a little bit modified Kendall's system of notations for QS [29] and will consider QS as a mathematical object, consist of four components $\alpha \,|\, \beta \,|\, \gamma \,|\, \delta$, where

- α —input flow,
- β —service mechanism,
- γ —system structure,
- δ —service discipline.

The symbols α and β take usual values M, GI etc., for Markov, recurrent and others for input flow, and distributions of recurrent service mechanism, the system structure γ consists of two numbers $\gamma = (n, m)$, where the first symbol means the number of servers, and the second means size of the buffer, and it is omitted for buffer size equal to infinity. The last symbol δ is used for service discipline and it is omitted for FIFO (first in first out) discipline.

Each of these components is also a complex mathematical object, which is usually studied in details in QT and is determined by some parameters. QS investigation usually could be divided into three directions:

- **analysis** that means the calculation of some output system Quality of service (QoS) characteristics for completely determined system;
- **synthesis** that means finding of some of the input characteristics in order to provide needed QoS indexes, and
- **control** that means operating of the system during its work with a goal of its behavior optimization with respect to given criteria.

In the review we focus on the last problems that is most important for applications. Based on the above, the following definition of a CQS is reasonable.

Definition 1 *CQS is a QS, for which some parameters of its components admit dynamic variation during its operation. Naturally, this variation is admitted in some domain and serves to some goals, determined with some QoS functionals.*

2.2 Classification of CQS

Accordingly to this definition CQS's can be classified as systems with:

- controllable input flow,
- controllable service mechanism,
- controllable system structure,
- controllable service discipline and
- complex CQS, for which some parameters of several components admit control.

For the control problem setting one needs in the goal of control and control rules.

2.3 Control times, goals and rules

Note firstly that in most practical situation the change of control (decision making) is possible only in special times, for example in times of customers arrival or their service completion.

We will call these times as *control* or *decision times* (DT). From another side the control problem solution is usually accompanied with the *goal of control* and besides of control parameters also depends on the admissible domain of their variation, rules of their use as well as of process observation possibility.

Concerning the goals of control they should be given by some *control quality functionals* optimization. It might be some QoS characteristics optimization in a steady state regime of the system operation or optimization (minimization or maximization) of a time of some state or set of states attainment. In some other cases the goal of control could be formulated as an optimization of some economic indexes, connected with the system operation. In the last case some *structure of losses and rewards* (Loss-Reward Structure—LRS) connected with the system operation should be done. Following to the tradition we will consider as an optimality criterion the minimization of some *Loss Functional* (LF). Losses or rewards could be connected both with the system stay in different states and with transition from one state to another. The LRS usually includes:

  • random reward R_n of the system manager for the service of n-th customer with finite mean value m_R (*service cost*),

  • penalty $C_w(l)$ for sojourn time of l customers in the system (*waiting* or *holding cost*),

  • cost $C_u(a_k)$ for using k-th service mode (*using cost)*,

  • penalty $C_s(k, k')$ for switching from k-th service mode to the k'-th one (*switching cost*).

Using these data the LF should be constructed. The general form of the LF will be considered in the next section, and its special form jointly with concrete examples will be represented.

The control rules are usually determined with the help of *control strategies*, which define the manner to take the decisions by *a Decision Maker* (DM) and depends, generally speaking, on system behavior observability and can be realized in several ways:

  • taking into account whole history of the process,
  • taking into account only the system last state, or
  • without any information;
  • also the decision can be made randomly or not.

Specification of these strategies will be introduced in the section 3.2.2.

2.4 Examples

Consider some examples of CQS that will be studied in details later.

2.4.1 Arrival control

Consider a $M/GI/1/\delta$–QS with controllable input. The customers arrive accordingly to Poisson flow with intensity λ and are served during random times that are i.i.d. with general Cumulative Distribution Function (CDF) $B(t)$ with mean value $m_B = \mu^{-1}$ and variance σ_B^2. The LRS includes:

  • the random reward R_n of the system manager from the service of n-th customer with finite mean value m_R (service cost),

  • the linear penalty $C_w(l) = C_w l$ for sojourn time of l customers in the system (waiting cost),

The control times are the arrival times, and the decisions consist in the admission or rejection of an arriving customer into the system. The problem consists in the admission of customers in the

system organization aimed reward maximization (or loss minimization).

2.4.2 Control of service mechanism

Consider a $M/M/1/\delta$ –QS with controllable service rate. Customers arrive accordingly to Poisson input with intensity λ, and are served with exponentially distributed service time having one of finite number K values of parameter μ_k $(k = 1, 2, \ldots K)$. The LRS includes:

- the random cost R_n for the service of n-th customer with finite mean value m_R (service cost),
- the linear penalty $C_w(l) = C_w l$ for sojourn time of l customers in the system (waiting cost),
- cost $C_u(\mu_k)$ of using server in k-th regime with service rate μ_k (using cost), and
- penalty $C_s(\mu_k, \mu_l)$ for switching from k-th regime to the l-th one (switching cost).

Control times are both: arrival and service completion times, and the decision consist in the choice of the service regime (service rate) aiming at long run expected reward per unit of time maximization.

2.4.3 Control of system structure

Consider a $M/GI/1/\delta$ –QS with controllable system structure. Customers arrive accordingly to Poison input with intensity λ, and are served accordingly to recurrent service mechanism with generally distributed with CDF $B(t)$ service time. The LRS includes:

- penalties C_1 for switching on a server, and C_0 for its switching off,
- cost C_s of server using per unit of time (using cost),
- the linear penalty $C_w(l) = C_w l$ for sojourn time of l customers in the system (waiting, cost).

Control times are both: arrival and service completion times, and the decision consists in possibility to switch on a server in customer arrival time and switch off it in service completion time aiming long ran expected loss per unit of time minimization.

2.4.4 Service discipline control

Consider s $M_N/GI_N/1/\delta$ –QS with several (N) types of customers and controllable service discipline among the priority disciplines. Inside the classes the customers are served accordingly to the FIFO discipline.

The customers arrive from Poison input with intensity λ_i $(i = \overline{1, N})$ for i-th type of customer, and are served accordingly to the recurrent mechanism with general CDF $B_i(t)$ $(i = \overline{1, N})$. The LRS includes:

- the linear penalty $C_i(l) = C_i l$ for sojourn time of l customers of i-th type in system (waiting cost).

The control consists in the choice of a customer for service in any service completion (decision) time, aiming the minimization of a long ran expected loss per unit of time.

Consider firstly the common model for investigation those and many others CQS.

3 Discrete time controllable semi-regenerative processes

As it has been mentioned above in the section 2.3 in the most practical situation the change of control (decision making) is possible only at special times, for example in times of customers arrival or their service completion. We will call these times as *decision times* (DT) and denote by $\{S_n, n = 0, 1, \ldots, S_0 = 0\}$ the sequence of decision times. Of course they are some measurable functionals of the process, describing the system behavior. For modelling CQS it is possible to use so called *Discrete Time Controllable Semi-regenerative Process* (DCSRP). Detailed information about these processes and its applications one can find in [31]. Remind here some needed definition and properties of these processes.

3.1 Semi-regenerative processes

3.1.1 Definitions and main properties

Semi-regenerative processes are some mixture of regenerative and semi-Markov processes. Several authors (**G .Klimov, E .Nummelin, V. Rykov, M. Yastrebenetsky**) introduced them under different names. Remind here its contemporary definition. Let

- $X = \{X(t), t \in R\}$ be a stochastic process with measurable state space (E, E),

- $F_t^X = \sigma\{X(v), v \leq t\}$ be generated flow of σ-algebras, and

- $\{S_n, n = 0, 1, 2, \ldots\}$ be a sequence of its Markov times with $S_0 = 0$.

7

Definition 2 (Rykov, Yastrebenetsky (1971)) *A pair* $\{X(t), S_n\}$ *is called a (homogeneous) Semi-Regenerative Process (SRPr), if for any subset* $\Gamma \subset E$ *and for all* $n = 1, 2, \ldots$ *takes place*

$$P\{X(S_n + t) \in \Gamma \,|\, F_{S_n}\} = P\{X(S_n + t) \in \Gamma \,|\, X(S_n)\} =$$
$$= P\{X(S_1 + t) \in \Gamma \,|\, X(S_1)\}. \tag{1}$$

Here

- r.v.'s S_n are called *Regeneration Times* (RT's),

- intervals $T_n = (S_n, S_{n+1}]$ and their lengths $T_n = S_{n+1} - S_n$ are called *Regeneration Periods* (RP's),

- functional random elements W_n, where

$$W_n = \{(X(S_n + t), T_n), t \leq T_n, n = 1, 2, \ldots\}$$

are called *Regeneration Cycles* (RC), and

- random elements $X_n = X(S_n)$ are called *Regeneration States* (RS).

Remark 1 *Intuitively clear that the SRP behavior is fully determined with its regeneration cycles* W_n *that form a Markov Chain in functional space, and under an additional condition of regularity any SRP is reconstructed (up to equivalence) by it. However the real determination of a generator of Markov chain in functional space is not a simple procedure.*

Therefore, we focus on the most important consideration of the one-dimensional characteristics of SRP. For this consider also some complementary processes

$$Y_n = (X_n, T_n), \quad N(t) = \max\{n : S_n \leq t\}, \text{ and } \quad Y(t) = X(S_{N(t)}).$$

Theorem 3 (Jolkof, Rykov (1981) and Rykov (1997)) *Let* $\{(X(t), S_n), t \geq 0, n = 1, 2, \ldots\}$ *is an SRP. Then the sequences* $\{X_n, n = 1, 2, \ldots\}$ *and* $\{Y_n, n = 1, 2, \ldots\}$ *are homogeneous Embedded Markov (EMCh) and Semi-Markov chains (SMCh), and* $\{N(t), t \geq 0\}$ *is a Markov Renewal Process (MRP) while* $Y(t)$ *is a semi-Markov process (SMP).*

Proof of the theorem can be found in the remind papers (see also [20, 45] for generalization). Denote

• the Transition Matrix (TM) of MCh $\{X_n, n = 1, 2, \ldots\}$ and Semi-Markov Matrix (SMM) of SMCh $\{Y_n, n = 1, 2, \ldots\}$ as

$$P(x, y) = P\{X_{n+1} = y \mid X_n = x\},$$
$$Q(x, t, y) = P\{X_{n+1} = y, T_{n+1} \leq t \mid X_n = x\};$$

• one-dimensional SRP distribution of separate regeneration periods (*SRP transition function*) by

$$\phi(x, t, \Gamma) = P\{X(S_n + t) \in B\}, t < T_n \mid X(S_n) = x\}, \quad n \geq 1;$$

• one-dimensional SRP distribution given an initial state x by
$$\pi(x, t, B) = P\{X(t) \in B \mid X(0) = x\} \equiv P_x\{X(t) \in B\};$$

• Markov Renewal Matrix (MRM) [**Korolyuk, Turbin** (1972), **Jolkof, Rykov** (1981)] and [**Rykov** (1997)] by

$$K(x, t, y) = M_x \sum_{n \geq 0} 1_{\{([0, t], y)\}} (S_n, X_n).$$

The behavior of a SRP $\{(X(t), S_n), t \geq 0, n = 1, 2 \ldots\}$ does not fully determined by its transition function. But many useful properties and characteristics of SRP can be represented in terms of appropriate characteristics at its separate cycle and its MRM. Especially, for a most interesting in practice one-dimensional distributions of a SRP, the following theorem can be proved with the help of complete probability formulae:

Theorem 4 *One dimensional distributions of SRP satisfy to the following relations*

$$\pi(x, t, B) = \phi(x, t, B) + \sum_{y \in E} \int_0^t Q(x, du, y) \pi(y, t - u, B0, \tag{2}$$

$$\pi(x, t, B) = \phi(x, t, B) + \sum_{y \in E} \int_0^t K(x, du, y) \phi(y, t - u, B) = \tag{3}$$

$$= \phi(x, t, B) + K \star \phi(t, B), \tag{4}$$

where $\star$ denotes the matrix-functional convolution..

Remark 2 *One can see that the equality (4) is the solution of the Markov renewal equation (2).*

3.1.2 Renewal, Limit and Ergodic Theorems

A well- known Key Renewal, Limit and Ergodic Theorems are generalized for SRP's as follows.
Theorem 5 Key renewal theorem, *[**Rykov, Yastrebenetsky** (1971), **Jolkoff, Rykov** (1981)] Under*

usual for the renewal theorem conditions, the following limiting formula holds

$$\lim_{t\to\infty} K \star g(t,x) \equiv \lim_{t\to\infty} \int_0^t \sum_{y\in E} K(x,du,y)g(y,t-u)dt =$$

$$= m^{-1}\int_0^\infty \sum_{y\in E} \hat{\pi}(y)g(y,t)dt,$$

where $\hat{\pi} = \{\hat{\pi}(x), x\in E\}$ is the invariant distribution of the EMCh and

$$m = \int_0^\infty \sum_{x\in E} \hat{\pi}(x)Q(x,[s,\infty),E)ds$$

is a stationary mean of RP length.

The last statement provides the calculation of SRP stationary probability distribution in terms of its distributions at separate RP's and invariant measure of EMCh.

Theorem 6 Limit theorem *Under some addition assumption of uniform regularity, SRP with ergodic (positively recurrent) EMCh steady state probabilities does exist and is equal to*

$$\pi(\Gamma) = \lim_{t\to\infty} \pi(x,t,\Gamma) = m^{-1}\int_0^\infty \sum_{x\in E} \alpha(x)\phi(x,t,\Gamma)dt \tag{5}$$

Theorem 7 Ergodic theorem *For uniformly regular SRP with ergodic (positively recurrent) EMCh,*

$$\lim_{t\to\infty} 1t\int_0^t g(X(u))du = \frac{M_{\hat{\pi}}\int_0^{T_1} g(X(u))du}{M_{\hat{\pi}}T_1} =$$

$$= m^{-1}\sum_{x\in E} \hat{\pi}(y)g(x)q(x), \tag{6}$$

where $q(x) = \int_0^\infty (1-Q(x,u,E)du$ is the mean time of the semi-Markov process $Y(t)$ staying in the state x.

The **proof** of these theorems one can find in [20, 22, 45, 61].

3.2 Discrete time controllable SRP

3.2.1 Definition

In controllable stochastic processes usually two factors should be considered: actions of Nature and will of the Decision Maker (DM). As a system behavior modeling process, we consider the SRP and suppose that the decision times coincide with the regeneration times that leads to the definition of a discrete time controllable process (DTCSRP).

Definition 8 *DTCSRP is a triple* $\{(X(t),S_n,U_n),n=1,2,\ldots\}$, *where* $\{(X(t),S_n),n=1,2,\ldots\}$ *is a SRP for which Regeneration Times (RT's)* $\{S_n\}$ *are also Decision Times (DT's), and* U_n *denotes the decision in time* S_n. *As before for DTSRP the intervals* $T_n = (S_n,S_{n+1}]$ *and their lengths* $T_n = S_{n+1} - S_n$ *are called Regeneration Periods (RP's).*

Many concrete CQS could be modeled with DTCSRP, including those that has been proposed in the section 2.4. We will return to them in the sections 5 – 8. See also [31].

As for usual SRP the main role for DTSRP play the controllable embedded Markov (CMCh) $\{(X_n, U_n), n = 1, 2, \ldots\}$ and semi-Markov $\{(Y_n, U_n), n = 1, 2, \ldots\}$ (CSMCH) chains. Its family (with respect to the decision $a \in A$) of transition matrices (described the Nature actions) is denoted as

$$P(x, y; a) = P\{X(S_n) = y \mid X(S_{n-1}) = x, U(S_n) = a\},$$
$$Q(x, t, y; a) = P\{X(S_n) = y, T_n \leq t \mid X(S_{n-1}) = x, U(S_n) = a\}.$$

Remark 3 *Besides these two processes one could consider also controllable functional random elements* $\{(W_n, U_n), n = 1, 2, \ldots\}$ *where* $W_n = \{(X(S_n + t), T_n), t \leq T_n, n = 1, 2, \ldots\}$ *is a Regeneration Cycle (RC), and* U_n *is the appropriate decision in time* S_n.

The behavior of the controllable process determines not only its transition probabilities but also by the controller or Decision Maker (DM). The control rules are determined by the strategies.

3.2.2 Strategies

Definition 9 *The manner of decision making is called strategy.*

As it was mentioned in the section 2.3, the strategy depends on system behavior, observability and can be realized by several ways. In order to formalize mentioned there possibilities we need to consider the history of the DTCSRP. Denoting the decision at the time S_n by U_n and putting $T_n = S_n - S_{n-1}$ the random history of the process up to time S_{n+1} can be represented as a sequence

$$H_n = \{X_0, U_0, T_1, X_1, \ldots U_{n-1}, T_n, X_n\}$$

and their realization with appropriate small letters.

A trajectory of DTCSRP is presented in the diagram

(-25, 10) X_0

The mostly general decision rule in the case of fully observable controllable process trajectory is determined by the *random measurable strategy*

$$\delta = \{d_i(u_i \mid h_i), i = 0, 1, \ldots\},$$

where $d_i(u_i \mid h_i)$ is a distribution of the DTCSRP decisions admissible at the trajectory. This class of strategies is denoted by Δ. Beside this class, there exists another classes, and a mostly popular one is a class of *simple Markov strategies*, for which decisions depends only on the last state of the process and appropriate distribution is degenerated one at some decision

$$d_i(u_i \mid h_i) = \varepsilon(f(x_i))$$

where x_i is the last state of the trajectory h_i and $f(x_i) = u_i^\star$ determines an optimal non-randomized decision in the state x_i. The appropriate strategy can be represented as $\delta = \{f, f, \ldots f, \ldots\} = f^\infty$, and appropriate class of simple Markov strategies as Δ_0.

Suppose, for simplicity that the initial time $S_0 = 0$ is a decision and regeneration time. Then, with an initial distribution α of the process and given strategy δ on the set of process trajectories, there exists some probability measure P_α^δ. Appropriate to this measure expectation will be denoted by E_α^δ. For degenerated at the state x initial distribution $\alpha = \varepsilon(x)$ appropriate

probability and expectation will be denoted as P_x^δ and E_x^δ, respectively.

3.2.3 Quality functional and optimization problem

The loss functional associated with the system RLS will be denoted as $Z(t)$ and it is supposed that it can be represented in the form

$$Z(t) = \sum_{n \geq 0} 1_{\{S_n \leq t\}} Z_n(t - S_n) \tag{7}$$

where $Z_n(t)$ represents the appropriate loss functional resulting only from the decision time S_n and which does not depend on another decisions. It is calculated based on the LRS for the concrete models.

There are several approaches for control of the process. The most popular are two of them.
  • Expected discounted loss minimization,

$$w(x, \delta) \equiv E_x^\delta \int_0^\infty e^{-st} Z(t) dt \Rightarrow \text{infinum}; \tag{8}$$

  • Expected long run loss minimization,

$$g(x, \delta) \equiv \lim_{t \to \infty} 1t E_x^\delta Z(t) \Rightarrow \text{infinum}. \tag{9}$$

For both cases the optimal strategy δ^* is that, for which

$$w(x, \delta^*) = \inf\{w(x, \delta), \delta \in \Delta\} = w(x), \tag{10}$$

$$g(x, \delta^*) = \inf\{g(x, \delta), \delta \in \Delta\} = g(x). \tag{11}$$

3.2.4 Optimality equations

One of the main results of the Markov Decision Processes theory, which also holds for DTCSRP is (see [31]) that the functions (10) satisfy to so called *optimality*, or *Bellman equations* that have the forms:
  • For the discounted loss minimization

$$w(x) = \inf_{a \in A_x} \left\{ \tilde{c}(x, a) + \sum_{y \in E} \tilde{q}(x, a, y) w(y) \right\}, \tag{12}$$

where $\tilde{c}(x, a)$ is one step discounted expected lost function with the initial state x and decision a,

$$\tilde{c}(x, a) = E_x^a \int_0^\infty e^{-st} Z_n(t);$$

and $\tilde{q}(x, a, y)$ is a probability generating function of a RP (inter decision times) under decision a for initial state x at given DT S_n,

$$\tilde{q}(x, a, y) = \int_0^\infty e^{-st} Q_{x,y}(dt; a).$$

  • For long run criterion optimization, the optimality equation besides the *model price* $g(x)$ includes also so called *value function* $v = \{v(x), x \in E\}$,

$$v(x) = \inf_{a \in A_x} \left\{ c(x,a) - g(x)m(x,a) + \sum_{y \in E} P(x,a,y)v(y) \right\}, \tag{13}$$

where $c(x,a)$ is a one step expected lost function with an initial state x and decision a,

$$c(x,a) = E_x^a Z_n(\infty),$$

and $P(x,a,y)$ is a transition probability during a RP (inter decision times) under decision a for initial state x at given DT S_n,

$$P(x,a,y) = Q_{x,y}(\infty;a).$$

For given functions $w(x)$ or $g(x)$ and $v(x)$, the right side hand of the equations (12, 13) are known as a *Bellman function* $b(x,a)$. Another name of these equations is the *Bellman equations*, which in general case could be represented as

$$v(x) = \inf_{a \in A_x} b(x,a). \tag{14}$$

4 Optimal strategy construction

4.1 Optimality Principle

The main result of the DTCSRP theory, as well as the theory of DMP, consists in the validity of *optimality principle* in the framework of some assumptions that takes place for mostly applicable situations. In [31] it was shown that under enough reasonable from application point of view conditions for DTCSRP, the optimality principle holds. i.e. there exists a simple Markov optimal strategy, and appropriate policies can be found from optimality equations. The optimality principle means that:

  • an optimal strategy exists and belongs to the class of simple Markov strategies,
therefore it is determined by the policy $f = \{f(x : x \in E)$, and

  • it could be found as the solution of optimality equation

$$f(x) = argmin\{b(x,a) : a \in A(x)\}.$$

As a result of these investigations the problem of QS control is reduced to the problems of
  • solution of Bellman equation (14) and
  • the Bellman function $b(x,a)$ minimization with respect to $a \in A(x)$.

4.2 Numerical methods

There are two main methods for the solution of Bellman equation
  • iteration algorithm due to **Howard** [28]
  • linear programming algorithm due to Wolf and Danzig .

4.2.1 Iteration algorithm

The iteration algorithm has been proposed firstly by **R.A.Howard** for MDP [28]. For both discounted and long run cost minimization, it consists of two procedures.

The algorithm

  • **Beginning:** Choose some (for example, one-step optimal policy f_0)

  • **Policy evaluation:** For a given policy f_k , solve the Bellman equations (12) or (13) in

order to find value functions $w_k = \{w_k(x), x \in E\}$ or $g_k = \{g_k(x), x \in E\}$ and
$v_k = \{v_k(x), x \in E$

- **Policy improvement:** For the given value functions w_k or g_k and v_k , find for each
state x the decision $f_{k+1}(x)$ that minimizes the value of Bellman function
$$f_{k+1}(x) = argmin\{b(x,a;v_k) : a \in A(x)\}.$$
Construct the new policy $f_{k+1} = \{f_{k+1}(x); x \in E\}$, leaving in each state x the previous decision
$f_k(x)$ if it coincides with the new one: $f_{k+1}(x) = f_k(x)$.

- **End:** Compare two successive policies f_{k+1} and f_k . If they coincide STOP and the
last policy is optimal, if no go to the step **Policy evaluation**.

It has been proved in [28] that at each step k of the algorithm the value functions w_k or g_k
and v_k

do not increase and therefore for the case of the decision process with finite number of states,
the algorithm stops for the finite number of steps, and for the case of system with denumerable states
space the algorithm converges.

There are different improvements and specification of the algorithm (see, for example,
Puterman [46], **Rykov** [49], and others.

4.2.2 Linear programming algorithm

Consider firstly LPA for the discounted losses model. For an event
$$B_n(x,u) = \{\ldots, X(S_n) = x, U(S_n) = u, \ldots\}$$
denote by $\tilde{\pi}_n(x,u)$ the quantity
$$\tilde{\pi}_n(x,u) = E_\mu^\delta e^{-sS_n} 1_{\{B_n(x,u)\}}$$
for simple Markov strategy $\delta = f^\infty$ and for initial process distribution μ . Than the functional (8)
with the process initial distribution μ due to (7) can be represented as follows

$$
\begin{aligned}
w(\mu, \delta) &\equiv E_\mu^\delta \int_0^\infty e^{-st} Z(t) dt = \\
&= E_\mu^\delta \int_0^\infty e^{-st} \sum_{n \geq 0} \sum_{x \in E} \sum_{u \in A(x)} 1_{\{B_n(x,u), S_n \leq t\}} Z_n(t - S_n) dt = \\
&= E_\mu^\delta \sum_{n \geq 0} \sum_{x \in E} \sum_{u \in A(x)} e^{-sS_n} 1_{\{B_n(x,u)\}} E_x^u \int_0^\infty e^{-sv} Z_n(v) dv = \\
&= \sum_{n \geq 0} \sum_{x \in E} \sum_{u \in A(x)} \tilde{\pi}_n(x,u) \tilde{c}(x,u),
\end{aligned}
\tag{15}
$$

where $\tilde{c}(x,u) = E_x^u \int_0^\infty e^{-sv} Z_n(v) dv$.

Using the backward Kolmogorov equation for DTCSRP, one can show that the function
$\tilde{\pi}_n(x,u)$ satisfies the equation

$$
\sum_{u \in A_x} \tilde{\pi}_n(y,u) = \begin{cases} \mu(y) & \text{for } n = 0 \\ \sum_{x \in E} \sum_{u \in A_x} \tilde{\pi}_{n-1}(x,u) \tilde{q}(x,u,y) & \text{for } n > 0, \end{cases}
\tag{16}
$$

If we introduce the variables $\xi(x,u) = \sum_{n \geq 0} \tilde{\pi}_n(x,u)$, then the problem of the discounted cost minimization can be represented as

$$\sum_{x \in E} \sum_{u \in A(x)} \tilde{c}(x,u)\xi(x,u) \Rightarrow \text{infinum} \tag{17}$$

under the restrictions that arise from the equations (16)

$$\sum_{x \in E} \sum_{u \in A(x)} (\delta_{x,y} - \tilde{q}(x,u,y))\xi(x,u) = \mu(y), \quad y \in E, \tag{18}$$

and

$$\sum_{x \in E} \sum_{u \in A_x} \xi(x,u) = 1 + \tilde{\tau}_1(s)(1 + \tilde{\tau}_2(s)(1 + \ldots)\ldots, \quad \xi(x,u) \geq 0. \tag{19}$$

The connection of the Linear Program solution and optimal control policy to an appropriate DTCSRP is contained in the following theorem:

Theorem 10 (Wolf Danzig) *For DTSRP with any simple Markov strategy $\delta = f^\infty$ the variables $\xi(x,u)$ of the Linear Program (17–rest-lp-discopt-2) have the following property: $\xi(x,u) > 0$ only for one, say $u^*(x)$ value $u \in A(x)$ for any $x \in E$, being $\xi(x,u) = 0$ for all another $u \in A(x)$ and vice versa. To to any solution y of Linear Program with this property corresponds a simple Markov strategy $f = \{f(x) = u^*(x), x \in E\}$.*

Let us turn now to the problem of long run loss minimization. The Linear Programming Algorithm (LPA) for the criteria of long run expected loss minimization looks like a little bit complicated. For any simple Markov strategy $\delta = f^\infty$, denote by $\Pi(\delta)$ and $H(\delta)$ the limiting and the fundamental matrices of the embedded Markov chain (MC) $X = \{X_n\}$,

$$\Pi(\delta) = \ln \sum_{0 \leq k \leq n-1} P^k(\delta) = [\pi(x,y;f)], H(\delta) = (I - P(\delta) + \Pi(\delta))^{-1} = [h(x,y;f)].$$

The LPA for long run expected loss minimization looks like the following

$$\sum_{x \in E} \sum_{u \in A(x)} c(x,u)\xi(x,u) \Rightarrow \text{infinum} \tag{20}$$

with respect to variables

$$\xi(y,u) = \sum_{x \in E} \mu(x)\pi(x,y;\delta)d(u \mid y), \quad \text{and} \quad \eta(y,u) = \sum_{x \in E} \mu(x)h(x,y;\delta)d(u \mid y).$$

under the restrictions

$$\sum_{x \in E} \sum_{u \in A(x)} (\delta_{x,y} - p(x,u,y)\xi(x,u) = 0,$$

$$\sum_{x \in E} \sum_{u \in A(x)} \xi(x,u) + \sum_{x \in E} \sum_{u \in A(x)} (\delta_{x,y} - p(x,u,y)\eta(x,u) = \mu(y) \, \forall \, y \in E. \tag{21}$$

Connections of the LPA solution to optimal policy is established with the help of Wolf and Danzig theorem.

Remark 4 *Additional attractive feature of the LPA consists in the possibility to use the dual linear programming algorithms for constructing of the optimality domains of some given simple control rules, as it will be shown in the section 8.2.*

4.3 Qualitative properties of optimal policies

The knowledge of some qualitative properties of optimal policies allows significantly simplify their calculation. For example, for monotone policies it is enough to find only the levels for

switching the policy from one regime to another. The optimality principle validation allows to investigate some qualitative properties of optimal policies, for example its monotonicity. Because the optimal policy is the optimizer of the Bellman function $b(x,u)$,

$$f(x) = argmin\{b(x,u) : u \in A(x)\},$$

one can investigate the conditions for it, which allow to provide the monotonicity of the optimal policy $f = \{f(x) : x \in E\}$.

Because for CQS the states and decision sets E and A are usually multi-dimensional, the problem of the optimal policy monotonicity investigation consists in finding the conditions for monotonicity of the multidimensional optimization problem solution. These conditions usually have a form of sub- or super-modularity for the optimizing function [69]. For the special case of QS optimal control an appropriate condition has been proposed by **Rykov** in [53]. In order to explain the optimal policy monotonicity conditions, consider firstly the problem of a smooth function $b(x,a)$ minimization in an enough good domain $G \subset E \times A$.

$$\min\{b(x,a) : a \in A_x\} \tag{22}$$

The necessary condition of a local minimum is

$$b_a(x,a) = 0, \quad b_{aa}(x,a) > 0, \tag{23}$$

where the notations are used:

$$b_a(x,a) = \frac{\partial b(x,a)}{\partial a}, \quad b_{aa}(x,a) = \frac{\partial^2 b(x,a)}{(\partial a)^2}, \quad b_{ax}(x,a) = \frac{\partial^2 b(x,a)}{\partial a \partial x}.$$

Therefore the problem solution is a function $a = f(x)$ from the first equation of (23). Its derivative can be found from the equation

$$b_{ax}(x,a) + b_{aa}(x,a)f'(x) = 0,$$

or

$$f'(x) = -\frac{b_{ax}(x,a)}{b_{aa}(x,a)}.$$

Thus, the monotonicity condition of the smooth optimization problem can be formulated as a theorem

Theorem 11 , *If the minimizer $f(x)$ of a smooth function $b(x,a)$ minimization is monotone, then $b_{ax}(.,.)$ preserve the sign along this solution. At that it is non-positive for non-decreasing solution $f(x)$ and it is non-negative for non non-increasing solution. The inverse conversation also holds: if the second mixed derivative of smooth function $b(.,.)$ in enough good domain G preserves the sign, then the solution of the minimization problem (22) can be chosen monotone.*

It is necessary to note that because the solution of problem (22) may be **non-unique,** the **monotone choice** should be used.

For the discrete problem of minimization, appropriate conditions are almost similar, however instead of the second mixed derivative now the second difference of the Bellman function with respect to variables of state and control spaces is used. If it will be denoted the same as before by , $b(x,a)$ the conditions of the discrete monotonicity problem look like in the previous theorem. For detailed formulation and proof see [53].

It succeeds not too often to prove the monotonicity property of an optimal policy. Nevertheless, because for queueing systems monotonicity leads to the threshold type of strategy, and the attraction of this kind simple strategies and the simplicity of their realization leads to many investigations of CQS under strategies with monotone (threshold) policies. In a series of papers of Dudin and his colleagues [11] – [16], the systems with multi-threshold control policies for QS with Markov Arrival (MAP) and Batch Markov Arrival Processes (BMAP) have been investigated. The

numerical investigation of optimal control of a multi-server system with heterogeneous servers has been done in the paper [17]. In **Rykov & Efrosinin** [57], the optimal control policy of systems on their lifetime has been found in the class of threshold policies.

In the next sections we demonstrate the above considered methods for some of CQS examples.

5 Arrival control

Control of arrivals to QS is a traditional problem in QT and has a long story. An excellent review of the earliest works about the problem one can find in **Stidhan Jr.** [65], to which we will follow in some parts of this section. Consider a model of arrival control $M/GI/1$, proposed in the example of the section 2.4.1. In this model the customers arrive accordingly to Poisson flow with intensity λ and are served during random times that are i.i.d. with CDF $B(t)$, mean value $m_B = \mu^{-1}$ and variance σ_B^2. The LRS usually includes: (a) reward R_n for each served customer with mean value m_R and (b) penalty $C_w(l)$ for l customers staying in the system per unit of time (holding or waiting cost). The control times are the arrival times, and the decisions consist in the admission or rejection of an arriving customer into the system.

The problem consists in admission of the arriving to the system customers (jobs) for service. There are different possibilities for control of arrivals to QS: (a) static, or (b) dynamic for (c) single-server, (d) multi-server, or (e) network queueing models.

5.1 Static flow control

5.1.1 Single-server static arrival control

In [65] a single-server static flow control model $M/GI/1$ has been reviewed. For this model the DM admits each job with probability p and rejects it with additional probability $1-p$. Therefore the real admitted flow of jobs has an intensity λp. Under this control rule for any control parameter p the system is a usual $M/GI/1$ QS with Poisson input intensity λp and generally distributed service time $B, B(t) = P\{B \le t\}$ with mean service time b and variance σ_b^2.

The problem consists in admission of jobs to the system in order to maximize the value of long run mean reward that due to proposed LRS has the form

$$\lambda p m_R - \lambda p C_w w(p) \quad \text{under the condition} \quad 0 \le \lambda p \le b, \tag{24}$$

where $w(p)$ is a stationary customer sojourn time in the system. Therefore, for the considered system one has

$$\lambda p w(p) = l(p) = \lambda p b + \frac{\lambda^2 p^2 (b^2 + \sigma_b^2)}{2(1-\lambda p b)}.$$

In terms of $\lambda b = \rho$ and $V_b = \sigma_b b^{-1}$ the problem (24) can be represented in the form

$$\lambda p m_R - \lambda p C_w w(p) = \lambda m_R p - c_w \left[\rho p + \frac{\rho^2 (1+V_b^2) p}{2(1-\rho p)} \right] \Rightarrow \max \tag{25}$$

under the condition $0 \le p \le \rho^{-1}$.

This is a simple optimization problem that could be simply solved numerically. Moreover, the last equation allows to obtain some interesting theoretical results proposed in [65].

5.1.2 Multi-server static arrival control

In this case the r-server system $M/G/r$ is considered. A Poisson input flow of customers with the intensity λ arrives into the system. The customers are served by servers during random times with general distribution $B_i(\cdot)$ with mean value b_i and variance σ_i^2 for i-th server $(i = \overline{1, r})$. In order to provide the stationary regime's existence, it is supposed that $\lambda \sum_{1 \leq i \leq r} b_i < 1$ and the servers are ordered such that

$$b_1 \leq b_2 \leq \ldots \leq b_r.$$

The (static) control rule is determined by the probability vector $p_1, p_2, \ldots, p_r$, where p_i is the probability to admit the arrived job to the i-th server.

In this case in [65] the problem of the static optimization is considered with respect to minimization of summary number of customers in the queue

$$\text{minimize} \sum_{1 \leq i \leq r} l_i(p_i) \quad \text{subject to} \quad \sum_{1 \leq i \leq r} p_1 = 1, \, 0 \leq p_i < \rho_i^{-1},$$

where $l_i(p_i)$ is the stationary number of customers in the buffer of i-th server, and $\rho_i = \lambda p_i$.

5.2 Dynamic arrival control

The dynamic arrival control of this model also in [65] has been considered. Some different generalization of the dynamic flow control model in **Kitaev & Rykov** (1995) [31] has been done.

The Bellman equation for the problem has been proposed and it was shown that the Bellman function satisfies to the theorem 11, from which it follows that the optimal policy belongs to the class of monotone polices and therefore the optimal strategy has a threshold property. Thus, there exists the threshold level, say l^* such that all arriving customers, which find in te

queue more than l^* another customers should be rejected. The optimal threshold level l^* can be found by investigation of the loss functional for the system $M/GI/(1, l)$ with finite buffer.

6 Service mechanism control

Control of the system service mechanism is the most frequently considered area of control. There are many diverse settings of the problem. The distinguish consists in input and service mechanism as well as in LRS.

One of the earliest work devoted to the service mechanism control has been proposed by **Sabeti** [62]. He considers the $M/M/(1, n)$ system with controllable service rate without waiting cost C_w and proves the monotonicity property of the optimal policy. For the $M/GI/(1, n)$ system, this result has been generalized in [63]. For closed queueing systems with controllable service rate the analogous results have been done in [7, 8]. More detailed review of the earliest works on the topic one can find in [9, 66, 48]. In the framework of DTCSRP, the system has been considered in [31]. The numerical investigation of the optimal service rate policies has been proposed in [73].

The hysteresis phenomena of optimal policy arise when the switching costs are taken into account. The $M/M/(1, n)$ system, with $n \leq \infty$, with controllable both input and service rate and a switching cost in [42] has been considered. As a control parameter the pair $a = (\lambda, \mu)$ is considered and the total order of the control parameters is supposed to be: $a \leq a'$ iff both $\lambda \leq \lambda', \mu \leq \mu'$. The control epochs for the model are both arrival and service completion times. The switching cost $C_s(a, a')$ has been included into the model. The conditions that provide optimality of monotone

hysteresis policy are studied. One step optimization problem of the Bellman function for the same model was studied in [27]. The optimality of monotone hysteresis policy under a little bit more general conditions is proved in [50]. The same mode under more general conditions in the framework of DTCSRP has been considered in [31]. The optimality of monotone hysteresis policy has been established there.

Remark 5 *This system also can be considered as a complex system, in which two parameters input and service rates are controllable.*

The oore general case of a $M/GI/(1,n)$ system with controllable service rate, however without switching cost] has been studied in [73]. Customers arrive with Poisson flow of intensity λ and are served with one of finite numbers of rates (service modes) $a \in A = \{a_k : k = 0,1,2,\ldots r\}$, such that the service time has a CDF

$$B_k(t) = B(t;a_k) = B(a_k t), k = \overline{0,r},$$

where $B(t)$ is a given CDF with finite mean m_B. As the control times (epochs) the service beginning times are supposed. At that the service delay it is convenient to consider as service mode with rate $a_0 = 0$. The control consists in choice in the control epochs, one of the service modes in order to maximize the discounted or long run expected reward or minimize appropriated loses. The monotonicity of the optimal policy has been studied numerically.

For multi-server systems the control rules consist usually in choice of servers for service. This setting is closed to the problems of the system structure control and will be considered in the next section.

7 System structure control

7.1 Servers switching on and off problem

The problems of system structure control are very closed to the service mechanism control and consist in switching on and off servers depending on the system state with the goal to optimize some given functional. Such type of systems have been considered by **Heyman** [24] and **Deb** [10].

Consider the system $M/GI/1$ with controllable system structure, which consists in the possibility to switch on and off the server. Customers arrive accordingly to Poison flow of the intensity λ and are served during random service time with general CDF $B(t)$. Reward structure includes:

- penalties C_1 and C_0 for switching on and off the server;

- penalty c_s per unit of server operating time (using cost); and

- penalty $c_w(i)$ for waiting i customers per unit of time.

Control epochs are both the arrival and service completion times and the control consist in switching on and off the server with the goal long run average losses minimization.

The monotonicity of the optimal policy has been proved that leads to its threshold property.

7.2 Slow server problem

So called slow server problem also can be considered as a system structure control problem. The problem is the following. For the $M/M/r$ QS with Poisson input intensity λ and r heterogeneous servers with service intensities $\mu_i (i = \overline{1,r})$, ordered such that

$$\mu_1 > \mu_2 > \ldots > \mu_r \qquad (26)$$

the conservative discipline, for which any idle server should be used, will be not expedient for example, with respect to minimization of the mean sojourn (holding) time (or mean number) of customers in the system. Therefore, the problem of optimal servers using arise.

The problem firstly has been considered for two servers in the static regime by **B. Krishnamoorthy** in [39], then also for two servers in the dynamic regime it is studied by **Hajek, Lin&Kumar and Koole** in [23, 41, 37]. It was shown that the optimal rule for using of the slow server has a threshold character and the optimal level of queue length q^* for using the slow server is that for which

$$q^* = \left[\mu_1 \mu_2^{-1}\right] + 1.$$

For the system with several heterogeneous servers the analogous rule has been proved by **Rykov** in [54], see also [53]. For the problem investigation consider a Markov decision process with set of space states $E = \{x = (q, d_1, \ldots d_r)\}$, where q is the queue length and d_k $(k = \overline{1, r})$ is an indicator of the k-th server state: $d_k = 0$ if the k th server free, and $d_k = 1$ if it is busy. A system of sets

$$A(x) = \{\text{set of indexes of free serves in the state } x\}$$

will be used as action sets.

It was proved that the optimal servers using rule also have the threshold structure. For any system state $x = (x_1, \ldots, x_r)$, there exists a threshold level $q^*(x)$ such that the new server should be used iff only the queue length q is greater than $q^*(x), q > q^*(x)$. At that the server of the most intensity among the free servers should be switched on. However the levels of the servers switching on and off should be calculated numerically, and one example of this calculation one can be found in [17].

For the generalized setting of the problem, which include also the waiting cost c_0 for any customer waiting in queue per unit of time and the using servers cots c_k $(k = \overline{1, r})$ for k-th server using per unit of time, the analogous optimal service rule has been found in [56, 55]. In this case the monotonicity of the optimal policy preserves if for servers ordering accordingly to (26) also the following condition holds:

$$c_1^{-1}\mu_1 > c_2^{-1}\mu_2 > \ldots > c_r^{-1}\mu_r. \qquad (27)$$

It should be noted that the optimal rule does not depend on the input intensity λ and waiting cost c_0. Also the threshold levels $q^*(x)$ should be found numerically.

8 Service discipline control

One of the first investigations on CQS does not touch the delicate problems regarding the MDP and deals only with the optimization problems in the framework of simple priority systems.

8.1 Priority optimization

The problem of priority system optimization can be considered as the problem of service discipline optimization in small class of decision rules that does not depend on the process trajectory observation and on the system state.

One of the first papers, devoted to QS optimal control, was the problem of optimal priority assignment [2]. The problem is the following. Consider a single server $M_r/GI_r/1$ queuing system with r independent Poison inputs of intensity $\lambda_k, (k = \overline{1, r})$. Random service time B_k of k-th type customers has CDF $B_k(t), (k = \overline{1, r})$ with mean value $b_k = \mu_k^{-1}$ and its waiting unit of time

penalties with c_k units. The problem consists in choice of the service priorities in order to minimize the expected long run loses

$$\lim_{t \to \infty} 1tE \int_0^t \sum_{1 \le k \le r} c_k L_k(s)ds = \sum_{1 \le k \le r} c_k l_k,$$

where $L_k(t)$ is the k-th type of customers queueing length in time t, l_k its stationary mean value, w_k is the stationary waiting time of the k-th type customers. The priority systems has been investigated in detail by **Klimov** [32, 35], **Gnedenko** and others [19], **Jaiswal** [21] and others, where the steady state system characteristics have been calculated. Based on the stationary characteristics of the priority systems, the solution has been proposed by **Bronstein and Rykov** in [2] (see also [6]) with the help of a simple method that many years after got the name *perturbation method* [3], and the rule now known as the $c\mu$-rule. It is the following. The priority should be organized in a such manner that

$$b_1 c_1 \le b_2 c_2 \le \dots \le b_r c_r$$

or in the order

$$c_1 \mu_1 \ge c_2 \mu_2 \ge \dots \ge c_r \mu_r,$$

which gives the name $c\mu$ *-rule* to this discipline.

Further, it was shown [59] that this discipline is also optimal inside essentially awide class of so called dynamic priorities, where the idea is the following: the whole set of states E is divided into classes $E_i, (i = 0,1,\dots r)$, in which the decision $a = i$ in the decision epoch should be taken. This class of decisions in fact represents a class of Markov decision strategies, and therefore the results show that the priority is the optimal discipline in the class of Markov decisions. These results have been developed further for system with dynamic preemptive resume priorities in [70, 71, 72] and others.

8.2 System with feedback

For more general settings the problem of service discipline control has been proposed by **Klimov** [33] and **Kitaev&Rykov** [30]. Remind more general system $M_r / GI_F 1 / \delta$ with feedback, proposed in [30]: r types of customers arrive into the single-server system from Poisson flows with intensities λ_i for i-th type of customers (i-customer); the customers are served with i.i.d. service times distributed accordingly to the CDF $B_i(t)$ with mean value b_i. Each served i-customer leaves the system with probability $q_i(0)$, or generates the set of the same type of customers $(n_1,\dots,n_r)$ with probability $q_i(n_1,\dots,n_r)$ that should be served also in the system. The LRS includes only a linear waiting (holding) cost $C_w(l,i) = c_i l_i$ for the i-th type of customer unit of time spent in the system.

The problem consists in service discipline construction that minimize the long run expected loss for the system exploitation. In [31] the existence of a simple Markov optimal strategy has been proved that gave the possibility to consider the problem in the framework of dynamic priority setting. As the result, the problem has been reduced to the linear program

$$\sum_{1 \le i,j \le r} b_i c_j x_{ij} \Rightarrow \min \tag{28}$$

with respect to x_{ij}, under restrictions

$$\sum_{1 \le i \le r} (a_{ij} x_{ik} + (a_{ik} x_{ij}) = \gamma_{jk} \quad (j,k = \overline{1,r}). \tag{29}$$

Here $x_{ij} = E_i L_j$ is the stationary mean value of j-th type of customers in the i-th type decision

epochs, $a_{ij} = \delta_{ij} - b_i \lambda_j - q_{ij}$, and q_{ij} is mean number of j-th type of customers generated by the i-th one. The model constant $\gamma_{jk} = \gamma_{kj} > 0$ does not depend on the control rule.

The linear problem (28, 29) solution is attained as usual in the extreme points of manifold, determined by the restrictions (29). This statement allows to show that the optimal policy determines the priority discipline. The dual linear program has been used in order to find the optimal priority rule. It is not enough simple and it is determined with an algorithm, which one can find in [30, 31].

Remark 6 *It is well known the connection between iteration algorithm of optimal policy construction and some linear program. In this problem, using a dual linear program allows to construct the optimal control policy. The possibilities of the dual linear program does not fully used up today. Really using the dual linear program allows to search parameters of the model, in which some given simple rule (or service discipline) will be optimal (see also remarks in the Conclusion). This approach has not yet exhausted itself and should be developed in future.*

The further investigation of the priority queues has been done by **Miscoy** and others [43], where especially numerical methods for generalized Kendall equation has been proposed. It also can be used as a method for optimal priority construction for the priority queues with switching times.

8.3 Closed queuing systems

The optimality of the priority rule for closed queueing system has been proved by **Koole** in [38]. A closed system $< M_r/M/1/\delta >$ with r sources, one server and controllable service discipline is considered. Sources have exponentially distributed life and repair times with parameters $\lambda_i \mu_i \ (i = \overline{1,r})$.. The system includes a waiting (holding) cost $c_k \ (k = \overline{1,r})$ for every unit of time that the k-th source is not functioning. The preemptive service discipline that minimizes the total average waiting cost is the goal of investigation.

The problem is formulated in the framework of the decision Markov model. In [38] it is proved that if the sources are ordered in such a manner that the conditions hold

$$\lambda_1 \leq \ldots \leq \lambda_r \quad \text{and} \quad c_1 \mu_1 \geq \ldots \geq c_r \mu_r \tag{30}$$

then the priority rule: *serve the source with minimal index* is optimal.

In the case when all $c_k = 1$ this result shows that this rule is also optimal for minimizing the average queue length. For the case when $\lambda_k = \lambda$ for all $k = \overline{1,r}$, the results show that the $c\mu$-rule also holds for the closed system with homogeneous sources. It is possible to see that the results also coincide with the analogous results for the open queueing system with heterogeneous servers considered in subsection 7.2.

9 Conclusion

To model CQS in the paper the DTCSRP is used. Of course, the assumption about the possibility to change the control in special control epochs is a some restriction. However, this restriction is natural for many practical situations. Moreover, for MDP due to the memoryless property of exponential distribution the epochs of any states changing are the natural decision times. However, there are some examples (and works), where the choose of control epochs is a subject of investigation. So in [4, 72, 58] the problem of optimal service interrupting has been investigated.

Another approach that could be mentioned as a subject for further investigation consists of construction of the domains into system parameters set such that some natural control rule would be optimal if the system parameters lie in the domain (*domains of optimality*). Besides the practical

importance of such approach, it allows in some cases to find an optimal policy for all possible values of the system parameters (example of such approach has been proposed in [30, 31].

Further development of the above approach for optimal control of QS is the following. Because for CQS the states and decision sets E and A are usually multi-dimensional, the problem of the optimal rules qualitative properties investigation should be stated as follows: Is it possible to introduce into sets E and A such a partial order, in which the optimal policy would be monotone?

References

[1]   A.K. Agrawala, E.G. Coffman Jr., M.R. Garey, and S.K. Tripathi (1984). A stochastic optimization algorithm, minimizing expected flow times on uniform processors.// IEEE Trans. Comput., V. C-33, pp. 351-356.

[2]   O.I. Bronstein, V.V. Rykov (1965). On optimal priority rules in queueing systems.// Izv. AS USSR. Techn. Cibern. No. 6, pp. 28–37 (in Russian).

[3]   Buyukkoc C C., Varaiya P., Walrand J. (1985) The $c\mu$ rule revisited.// Advances in Applied Probability. V. 17, pp. 237-238.

[4]   S.S. Chitgopepeker (1969). Continious time Markovian sequential control processes. // SIAM J. Contr. V. 7, No. 3, pp. 367-389.

[5]   R.W. Conway, W.L. Maxwell, L.W. Miller (1967). Theory of scheduling.// Addison. Wisley Mass. 1967.

[6]   D. Cox, W. Smith (1966). Queueing theory.// M.: "Mir", 1966.

[7]   T.B. Crabill (1972). Optimal control facility with variable exponential service times and constant arrival rate.// Management Sci. V.18, No. 9, pp. 560-566.

[8]   T.B. Crabill (1974). Optimal control of a maintenance system with variable service rate. // Oper. Res. V.22, No. 4, pp. 736-745.

[9]   T.B. Crabill, D. Gtross, and M. Nagazine (1977). A classified bibliography of research on optimal design and control queues.// Oper. Res. V.25, pp. 219-232.

[10] R.Deb (1976). Optimal control of batch service queues with switching costs.// Adv. Appl. Prob. V. 8, pp. 177-194.

[11] A. Dudin (1998). Optimal multithreshold control for a $BMAP/G/1$ queue with N service modes.// Queueing Systems. V. 30, pp. 273-287.

[12] A. Dudin, I. Khalaf (1988). Optimal control by the current value of the threshold: Data communication/switching communication under the adaptive communication, in: XIII All-Union Workshop in Computer Networks. VINITI: Moscow, 1988.

[13] Che Soong Kim, Valentina Klimenok, Alexander Birukov, Alexander Dudin (2006). Optimal multi-threshold control by the $BMAP/SM/1$ retrial system.// Ann. Oper. Res. V. 141, pp 193-210.

[14] Che Soong Kim, Sergey Dudin, Valentina Klimenok (2009). The $MAP/PH/1/N$ queue with flows of customers as a model for traffic control in telecommunication networks.// Performance Evaluation V. 66, pp. 564-579.

[15] Bin Sun, Moon Ho Lee, Sergey A. Dudin, and Alexander N. Dudin (2014). Analysis of Multiserver Queueing System with Opportunistic Occupation and Reservation of Servers. Hindawi Publishing Corporation. Mathematical Problems in Engineering. Volume 2014, Article ID 178108, 13 pages, http://dx.doi.org/10.1155/2014/178108.

[16] Bin Sun, Moon Ho Lee, Alexander N. Dudin, and Sergey A. Dudin (2014). $MAP+MAP/M_2/N/\infty$ Queueing System with Absolute Priority and Reservation of Servers.// Hindawi Publishing Corporation. Mathematical Problems in Engineering. Volume 2014, Article ID 813150, 15 pages, http://dx.doi.org/10.1155/2014/813150.

[17] D.V. Efrosinin, V.V. Rykov (2003). Numerical investigation 0ptimal control of a system with heterogeneous servers.// Autom and Remote Control, No. 2, pp. 143-151.

[18] A. Emphermides, P. Varaiya and J. Walrand (1980). A simple dynamic routine problem.// IEEE

Trans on AC. V. AC-25, No.4, pp. 690-693.

[19] B.V. Gnedenko and others (1973). Priority queueing systems.// M.: Moscow State Univ. Publ. House, 1973.

[20] J. Jacod (1971). Theoreme de renouvellemant et classificasion pour les hains semi-Markoviennes.// Ann. inst. Henri Poincare, sect B. V. 7, pp. 85-129.

[21] N. Jaiswall. Queues with priority. M.: "Mir", 1973.

[22] S. Jolkoff, V. Rykov (1981). Generalized regenerative processes with embedded regenerative periods and their applications.// MOS, Ser. Optimization. 1981, pp. 575-591.

[23] B. Hajek (1984). Optimal control of two interacting service stations.// IEEE Trans. Automat. Control. V.29, pp. 491-499.

[24] D.P. Heyman (1968). Optimal operating policies for $M/GI/1$ queueing systems.// Oper. Res. V. 16, pp. 362-382.

[25] D.P. Heyman, M. Sobel (1982). Stochastic Models in Operation Research. Vol. 1. McGraw-Hill, New York, 1982.

[26] D.P. Heyman, M. Sobel (1984). Stochastic Models in Operation Research. Vol. 2. McGraw-Hill, New York, 1984.

[27] S.K. Hipp, U.D. Holzbaur. (1988) Decision process with monotone histeretic policies.// Oper. Res. V. 36, No. 4, pp. 585-588.

[28] R.A. Howard (1960). Dynamic Programming and Markov Processes. MITPress, 7-th edition (1972).

[29] D.G. Kendall (1953) Stochastic processes occurring in the theory of queues and their analysis by the method of embedded Markov chains.// Annals of Math. Stat. Vol. 24, pp. 338-354.

[30] V.Yu. Kitaev and V.V. Rykov (1980) A service system with a branching flow of secondary customers.// Autom and Remote Control, No. 9, pp. 52-61.

[31] V.Yu. Kitaev and V.V. Rykov (1995) Controlled queueing systems. CRC Press, N.Y, 1995.

[32] G.P. Klimov (1966). Stochastic queueing systems.// M.: "Nauka', 1966 (In Russian).

[33] G.P. Klimov (1974). Time sharing service systems I.// Theory Prob. Appl. V. 39, pp. 479-490.

[34] G.P. Klimov (2011). Probability theory and mathematical statistics.// M.: Moscow State Univ. Publ. House, 2011 (in Russian).

[35] G.P. Klimov (2011). Queueing theory.// M.: Moscow State Univ. Publ. House, 2011 (in Russian).

[36] G.M. Koole (1995) Stochastic scheduling and dynamic programming. CWI TRACT. V.113. Amsterdam.

[37] G.M. Koole (1995) A simple proof of the optimality of the threshold policy in two-server queueing system.// Syst.&Control Lett. V. 26, pp. 301-303.

[38] G.M. Koole (1996) Scheduling a Repairman in a Finite Source System.// Mat. Meth. in Oper. Res., V. 44, pp. 333-344.

[39] B. Krishnamoorthy (1963). On Poisson queue with two heterogeneous servers.// Oper. Res. V. 11, pp. 321-330.

[40] P. Langrock and W.W. Rykow (1984) Methoden und Modelle zur Steurung von Bedienungssystemen. In: Handbuch der Bedienungstheorie. Berlin, Akademie-Verlag, V. 2, pp. 422-486.

[41] W. Lin, P.R. Kumar (1984) Optimal control of a queueing system with two heterogeneous servers.// IEEE Transactions on Automatic Control. V.29, pp. 696-703.

[42] F.V. Lu, R.F. Serfozo. (1984) $M/M/1$ Queueing decision process with monotone histeretic optimal policies.// Oper. Res. V.32, No. 5, pp. 1116-1129.

[43] G.K Miscoy, S Giordano, A.Yu. Bejan, V.V. Rykov(2008). Multi-dimensional analogs of Kendal's equations for priority queueing systems: numerical aspects.// Autom. and Remote Control No. 6 (2008), pp. 82-94.

[44] Ph. Nain, P. Tsoucas, J. Walrand (1969) Interchange arguments in stochastic scheduling.// J. Appl. Prob., V. 27, pp. 815-826.

[45] E. Nummelin (1978). Uniform and ratio-limit theorems for Markov-renewal and semi-

regenerative processes on a general state space.// Ann. inst. Henri Poincare, sect. B, V. 14, pp. 119-143.

[46] M. Puterman (1964). Markov Decision Processes. Wiley, N.Y., 1994.

[47] S. Ross (1970). Applied Probability Models with Optimization Applications. Holden-Day, San Francisko, 1970.

[48] V.V. Rykov (1975). Controllable queueing systems.In: Itogi nauki i techniki. Teoria verojatn. Matem. Statist. Teoretich. Kibern. V. 12, pp. 45-152. (In Russian). There is English translation in Journ. of Soviet Math.

[49] V.V. Rykov (1977). Controllable stochastic processes and systems.// Gubkin Moscow state institute of oil and gas industry. Moscow. 106 p. (In Russian).

[50] V.V. Rykov (1995) Histeretic phenomena in controllable queueing systyems.// Vestnik RUDN. Ser. Applied mat. and informatics, No. 1, pp. 101-111.

[51] V.V. Rykov (1997) Two approaches to complex hierarchical systems decomposition. Continuously interrupted systems.// Autom and Remote Control, No. 10, pp. 91-104.

[52] V.V. Rykov (1997) Two approaches to complex hierarchical systems decomposition. Aggregative systems.// Autom and Remote Control, No. 12, pp. 140-149.

[53] V.V. Rykov (1999) On monotonicity conditions for optimal control policies of queueing systems (in Russian).// Autom. and Remoute Control, No. 9, pp. 92-106.

[54] V.V. Rykov (2001). Monotone Control of Queueing Systems with Heyerogeneous Servers.// Quaueing Systems. V. 37, pp. 391-403.

[55] V.V. Rykov (2013). On a Slow Server Problem.// In: Stochastic Orders in Reliability and Risk (edd. by H. Li and X. Li). Springer, N.Y., 2013, pp. 351-361.

[56] V.V. Rykov, D.V. Efrosinin (2009). To the slow server problem.// Autom and Remote Control, No. 12, pp. 81-91.

[57] V.V. Rykov, D.V. Efrosinin (2012). On optimal control of systems on their life time. Recent Advances in System Reliability. Springer Series in Reliability Engineering. Springer, (Edds. by A.Lisnianski and I.Frenkel), 2012, pp. 307-319.

[58] V.V. Rykov, D.V. Kozyrev. (2014) Optimal control for scheduling with correction.// In: Proceedings of the XII All-Russian Workshop on Control Problems (Moscow 6-19 June 2014). M.: ITP Ras, 2014, pp. 8850-8854.

[59] V.V. Rykov, E.E. Lemberg. On optimal dynamic priorites in single server queueing systems. Izv. AS USSR. Techn. Cybern. No. 1, pp. 25–34 (in Russian).

[60] V.V. Rykov, E. Levner (1998) On optimal allocation of jobs between heterogeneous servers.// In Distributed Computer Communication Networks. Proceedings of the International Workshop, June 16-19 1998, Moscow. IITP RAS, Moscow, 1998, pp. 20-28.

[61] V.V. Rykov, M.A. Yastrebenetsky (1971). On regenerative processes with several types of regeneration points.// Cybernetics, No. 3 (1971), pp. 82-86. Kiev (in Russian).

[62] H.Sabeti (1973). Optimal selection of service rates in queueing with different costs.// J.Oper.Res.Soc. Jpn. V.16, No. 1, pp. 15-35.

[63] R.Schasberger (1975). A note on optimal service selection in a single server queue.// Management Sci. V.21, No. 11, pp. 1321-1331.

[64] L.I. Sennott (1999). Stochastic Dynamic Programming and the Control of Queueing Systems.// John Wiley & Sons, N.Y., 1999, 358p.

[65] Stidham, Sh. Jr. (1985). Optimal control of admission to a queueing system.// IEEE Transactions on Automatic Control. V. AC - 30, No. 8, pp. 705-713.

[66] Sh. Stidham, Jr. and R. Weber, R (1993) A survey of Markov decision models for control of networks of queues.// Queueing Systems. V. 13, pp. 291-314.

[67] H. Tilms (1994). Stochastic Models. An Algorithmic Approach. Wiley, N.Y., 1994.

[68] E.A. Timofeev. (1991). On optimal choise mean waiting times in queueing system $GI/M_n/1$ .// Autom. and Remote Control, 1991. No. 6, pp. 77–83.

[69] D. Topkis (1978) Minimizing a submodular function on a lattice.// Operations Research. V.26,

No. 2, pp. 305-321.

[70] E.B. Veklerov (1967) On optimal pre-emptive dynamic priority disciplines in queueing systems.// Izv. AN SSSR, Techn. cyb., No. 2, pp. 87-90. (In Russian).

[71] E.B. Veklerov (1971) On optimal priority disciplines in queueing system. *Autom. and Remout Control* No. 6, 149-153. (In Russian).

[72] E.B. Veklerov, V.V. Rykov (1971) On optimal synthesis of queueing systems.// In. Adaptive systems. Complex systems. M.:"Nauka", 1971, pp. 384-389.

[73] S.N. Verbitskiy, V.V. Rykov. (1998) Numerical investigation of servixe rate optiman policies.// Autom. and Remote Control. No. 11 (1998), pp. 59-70 (In Russian. There is English translation).

[74] W. Winston (1977) Optimality of the shortest line discipline. J. Aapl. Prob. V. 14, pp. 181-189.

[75] R. Weber (1993) On a conjecture about assigning jobs to processors of different speeds. IEEE Transactions on Automatic Control. V. 38, No. 1, pp. 166-170.