# Optimal Control of Requests Acceptance in the G|M|n|oo Queuing System with Impatient Customers to Minimize the Payback Period

Timofey Solovev
Applied Mathematics
National Research University Higher School of Economics
Moscow, Russia
tdsolovev@edu.hse.ru

**Abstract –** This article investigates the issue of optimal control strategy of incoming flow of customers to the G|M|n|∞ queuing system with impatient requests to minimize the payback period. Queuing systems are widely used for analysis and modeling of industrial, information, or social systems. One of the key performance indicators of the system is its payback time, since the faster the system pays off, the faster it starts to make a profit. Therefore, this criterion is investigated in this paper. Semi-Markov processes, controlled stochastic processes, Markov chains are frequently applied to explore queuing systems. The results of this research are the process of system functioning, a constructed mathematical model in the form of the controlled semi-Markov stochastic process, a solved optimization problem, and a selected optimal control of incoming flow of requests. A simulation model of this system will be also provided to confirm analytical result.

**Keywords –** attainment functional, controlled stochastic processes, control strategy, queuing systems.

## I. INTRODUCTION

Queuing system is a system designed to execute operations of the same type. In order to describe the queuing system correctly, it is necessary to define its main components:

- Incoming flow of requests (customers);
- Service process;
- Service devices;
- Storage (queue).

Queuing systems are frequently classified using Kendall's notation: in this work G|M|n|∞, where "G" is a distribution of intervals between arrivals of requests to the system, “M” is a distribution of customer service time, “n” is a number of service devices, “∞” is a capacity of the storage. Many types of queuing systems exist, for instance, there are systems with impatient customers, repeated calls, priority requests, infinite or finite storages.

The purpose of this study is to determine the optimal control strategy of input flow of customers into a non-Markov multidevice queuing system with infinite capacity storage and impatient requests to minimize a payback period.

To accomplish that, it is necessary to cover the following points: 1) to construct the mathematical model as controlled semi-Markov stochastic process; 2) to choose the efficiency criterion (attainment functional); 3) to formulate and solve the optimization problem; 4) to obtain the optimal strategy and analyze the results; 5) to construct the simulation model of the system.

The novelty and theoretical importance of the investigation is connected with the chosen efficiency criterion. In this paper, the average operating time is minimized, whereas average income is usually taken as an efficiency criterion. Generally, one of the most significant profitability indicators of any system is its ability to receive the income for a long period of time. However, the payback time of the system is also a crucial criterion, which is considered in this investigation.

Practical significance of the research is defined by the following things. The results obtained during this study can be utilized to optimize functioning, reduce costs, and maximize profits of any informational systems that can be represented as a described queuing system. Moreover, the results of current work can be applied in other studies.

This paper has the following structure. First of all, scientific literature on queuing theory is investigated. Secondly, the paper specifies stages and methods of the study and reports the results. Eventually, the article submits the conclusion.

## II. RELATED WORK

Study of queuing systems is a significant element in the optimization of real complex systems. As a result, a lot of studies have been done on this topic. The literature overview of the several investigations about queuing theory is below.

For example, the optimization problem of the main storage capacity in the G|M|1|K queuing system with the additional “M” capacity storage was researched by Agalarov (2020, [1]). If the main storage is not full, the incoming request is placed into it, otherwise the request is moved to the additional queue or lost.

Let $D(K)$ be the marginal (stationary) profit earned by the system per unit of time:

$$D(K) = \frac{1}{\bar{v}} \sum_{j=0}^{K+M} \pi_j(K) q_i(K) \quad (1)$$

$D(K)$ is a function of variable “K” and the other parameters are fixed. Here $\bar{v}$ is the average time between the arrivals of requests, $\pi_j(K)$ is stationary probabilities of the states $j$ (the state of the system is the total number of requests in both storages, when the new customer arrives), $q_i(K)$ is the average profit received by the system in the state $j$. Then the function $D(K)$ is unimodal for any finite value $M \geq 0$.

This investigation is substantial because the capacity of the main queue is one of the crucial characteristics of the system, which defines the profit. Therefore, the results of this work can be applied in real systems.

Afonon and Nikulin (2020, [2]) researched the optimization issue of multidevice queuing systems under the significant loads (it means that the intensity of service is much less than the intensity of incoming flow). Extreme situations frequently appear in real systems. Thus, the problem emerges: to minimize the number of service devices provided that the queuing system has guaranteed throughput. For this reason, this study is significantly important.

Authors considered the M|M|m system with λ is the incoming flow intensity and μ is the service intensity

Let $p_m$ be the probability that the request is lost:

$$p_m = 1 - \frac{R(m-1, \alpha)}{R(m, \alpha)} \quad (2)$$

Here $R(k, a)$ is integral of probabilities (Laplace functions). The probability $p_m$ depends on three parameters: λ, μ, $m$. The number of devices $m$ is unknown variable, so, the optimization problem can be written in the following way:

$$p_m = f(\lambda, \mu, m) \rightarrow \min_{\lambda,\mu>0} f(\lambda, \mu), m = 1,2, \ldots \quad (3)$$

The minimum value of $m$ for given λ, μ can be found by solving problem (3) numerically.

To sum up, the authors proposed a heuristic algorithm for finding the required minimum number of service devices to make the probability of losing requests small enough.

Agalarov, Agalarov and Shorgin (2016, [3]) investigated the issue of optimal storage capacity to maximize the average stationary income in the M|G|1 queuing system. An incoming customer is queued if the number of requests in storage is less than the specified value, otherwise the customer is lost. The request leaves the system only after its service.

The authors obtained the following results. The existence of optimal queue length was proven. The algorithm for calculating the lower bound for the optimal queue length and corresponding value of maximal profit is proposed. The results, for example, can be used to find optimal strategies of the M|G|1|r queuing systems.

This research is significant because the real systems usually do not have the infinite capacity storage, as a result, the optimal restriction on the queue length is necessary to minimize the losses.

Thereby, many types of queuing systems exist, some of them have been described above. Nevertheless, real systems are more complicated than the considered models. Consequently, the real issue here is to research more queuing systems with different criteria and characteristics. In addition, new projects, where new sophisticated systems are functioning, are launched every day. For this reason, the study of queuing theory methods is important.

## III. METHODS AND STAGES

The first stage is to describe the system operation. In this paper, the system with impatient customers and possibility of denial of service is investigated. In other words, a request can leave the storage with some probability before service. The queuing system includes $n$ service devices, therefore, if the system contains $n$ requests, the $n+1$ customer will be placed in storage and so on. Moreover, the system can refuse a request, in this case, the request is lost. Customers move to the service devices from the queue according to the FIFO rule. The model of this system is multidevice queuing system, which scheme is shown in Figure 1:

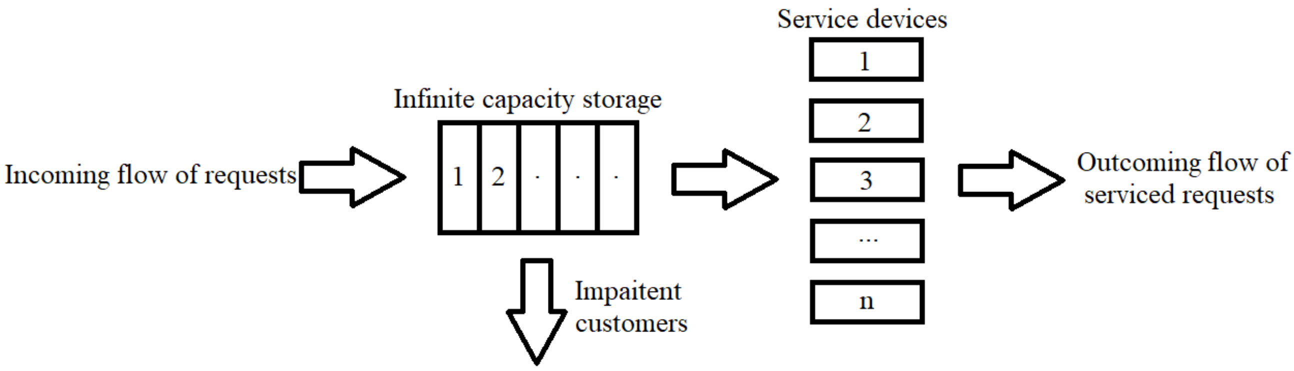


Fig. 1. The model of the queuing system with impatient customers.

The second stage is to formulate the problem. The incoming flow is an arbitrary flow with a given distribution function of the intervals between arrivals of requests, μ is a given parameter of the exponential distribution function of service time, the customers leave the queue with the specified intensity γ.

Within this problem, it is necessary to find the optimal control strategy of the input flow of customers into the system. The decision to accept (or not to accept) a request into the system is made when a new request arrives. A set of these decisions is the control strategy.

In this way, the problem is to optimize the work of the system by managing the incoming requests and to obtain the set profit in the minimum time.

The next step is to construct the mathematical model. The incoming flow will be regulated under the following principle: the number of customers in the system is counted at the moment when a new request arrives. Then the system decides to accept (or not to accept) a new request, where the choice depends on the quantity of requests in the system. A new customer is accepted with the certain probability $p_v$ or leaves with the probability $1 - p_v$ ($v$ is the number of customers in the system at the moment).

Let the stochastic process (Doob 1953 [7]) $\xi(t)$ be the quantity of requests at the $t$ time moment in the system. This process is non-Markov process. However, the sequence $\xi_n = \xi(t_n)$ forms a nested Markov chain, where $t_n$ are the moments of requests arrival. There have been several investigations into the nested Markov chains method ((Nummelin 1984 [4], Karlin 1968 [5]).

Controlled semi-Markov stochastic process $X(t)$ with the countable set of states $E = \{0, 1, 2 \dots\}$ is introduced. This type of processes is determined as three-dimensional Markov chain (Kashtanov and Medvedev 2010 [6]) or as Markov recovery process $(\xi_n, \theta_n, u_n), n \geq 0$, where:

- $\xi_n \in E$ is a nested Markov chain;
- $\theta_n \in R^+ = [0, +\infty)$ is time;
- $u_n \in U_i$ is management, $i \in E$.

Thereby, the controlled semi-Markov process $X(t)$ can be defined as a couple:

$$X(t) = \big(\xi(t), u(t)\big), \tag{4}$$

$$\xi(t) = \xi_{v(t)-1}, \quad u(t) = u_{v(t)}$$

Here $v(t) = inf\{n\colon \sum_{k \leq n} \theta_k > t\}, \theta_0 = 0$ is called the counting process. The second component of the controlled semi-Markov process $u(t)$ defines the decision-making trajectory.

The last step is to formulate the optimization problem. In this problem, the set of decisions consists of only two elements $u(t) \in U = \{0, 1\}$: $U_0$ is the request is lost, $U_1$ is the system accepts the request. $W(t)$ functional is equal to the profit received by the system during the operation time *t*. This functional is called the attainment functional. It is assumed that the process $\xi(t)$ is in $i_0$ state at the initial moment of time $t = 0$: $\xi(0) = i_0$.

Let the variable $S$ be a given number that system needs to earn, let the random variable $\eta$ be the time to earn income $S$:

$$\eta = min(t \mid W(t) = S) \tag{5}$$

Then the introduced conditional mathematical expectation of the variable $\eta$ is determined by the following expression:

$$L_{i_0} = E(\eta \mid \xi(0) = i_0), i_0 = 0 \tag{6}$$

It is necessary to find the minimum of this mathematical expectation $L_{i_0}$. Then the optimization problem can be rewritten in this form:

$$L_0 = E(\eta \mid \xi(0) = 0) \to \min_{p_v} \tag{7}$$

## IV. RESULTS ANTICIPATED

The results are yielded in this research:

1. The mathematical model of G|M|n|∞ queuing system with impatient customers as controlled semi-Markov stochastic process (4) was developed (using nested Markov chains method).

2. The efficiency criterion was chosen, and the optimization problem (7) was formulated: to reach the set profit during the minimum operating time.

In the future, this optimization problem will be solved, the control strategy will be found, and the simulation model of this system also will be provided to confirm analytical results.

## V. CONCLUSION

Queuing systems are broadly used to reflect different real complex systems. Queuing systems are very convenient models because it is easy to analyze and optimize them. In this study, the mathematical model of the system was constructed, and optimization problem was formulated. Next, this problem will be solved, the optimal control strategy of incoming flow of customers will be obtained, and the simulation model of the system will be created. These results can be utilized to improve the real systems which can be represented as G|M|n|∞ queuing system with impatient customers or to investigate other efficiency criterions of this system.

## REFERENCES

[1] Ya. M. Agalarov. 2020. «Optimizaciya emkosti osnovnogo nakopitelya v sisteme massovogo obsluzhivaniya tipa G|M|1|K s dopolnitel'nym nakopitelem»

[Optimozation of the capacity of the main storage in G|M|1|k queueing system with an additional storage device]. – Informatika i eyo Primeneniya – Inform. Appl. 14(2):72–79.

[2] V. V. Afonon, V. V. Nikulin. 2020. «Optimizaciya mnogokanal'nyh sistem massovogo obsluzhivaniya pri bol'shih zagruzkah» [Optimization of quruing systems under significant loads]. – Vestnik of Astrakhan State Technical University. Series: Management, Computer Science, Informatics. 3:105-115.

[3] Ya. M. Agalarov, M. Ya. Agalarov, V. S. Shorgin. 2016. «Ob optimal’nom porogovom znachenii dliny ocheredi v odnoy zadache maksimizatsii dokhoda sistemy massovogo obsluzhivaniya tipa M/G/1» [About the optimal threshold of queue length in particular problem of profit maximization in M/G/1 queueing system]. – Informatika i eyo Primeneniya – Inform. Appl. 10(2):70–79.

[4] E. Nummelin. 1984, 2004. «General irreducible Markov chains and non-negative operators». – Cambridge University Press. 156p.

[5] S. Karlin. 1968. «A first course in stochastic processes». – New York and London: Academic press. 502p.

[6] V. A. Kashtanov, A. I. Medvedev. 2010. «Teoriya nadezhnosti slozhnyh sistem» [Reliability theory of complex systems]. – M.: Fizmatlit. 608 p.

[7] J. L. Doob. 1953. «Stochastic Processes». – New York: John Wiley and Sons. 664p.

*Word count 1790*