STRATEGIES FOR OPTIMIZING SINGLE-CHANNEL MASS SERVICE SYSTEM WITH NON-STATIONARY INPUT FLOW

Oleg Zagurskiy[1], Liliya Savchenko[1], Oleksandr Domin[1], Maksym Odnorog[2]
[1]National University of Life and Environmental Sciences of Ukraine, Ukraine;
[2]Bila Tserkva National Agrarian University, Ukraine
zagurskiy_oleg@ukr.net, lilya_savchenko@ukr.net, demin31@gmail.com, odnorog_btnau@ukr.net

Abstract. The paper considers models of single-channel queuing systems, as well as methods for evaluating their efficiency under various operating conditions. A program has been developed that allows modeling the functioning of a single-channel system with a non-stationary input flow and random service time. Its use makes it possible to take into account the dynamics of changes in the intensity of the input flow over time and to choose a specific optimization strategy, taking into account the specifics of the system, in particular, its intended purpose, technical capabilities, and economic constraints. The results obtained are of practical importance for optimizing the operation of single-channel mass service systems in the logistics industry. Based on these results, recommendations are provided for selecting the optimal system parameters that will ensure the best compromise between service quality and resource efficiency.

Keywords: efficiency, logistics chain, optimization, mass service system, strategy, service quality.

Introduction

In the context of intensive development of the service sector, logistics and information technologies, the issue of effective organization of service processes is gaining particular importance. Single-channel systems, characterized by the presence of a single channel for servicing applications, are a fundamental element of the theory of queuing, which allows modeling a wide range of real processes of queuing – from loading/unloading a vehicle to the functioning of the supply network.

Queuing systems (QS) are widely used to model processes in transport, logistics, production and service systems, which increases the interest of scientists in their research. The theoretical foundations of QS analysis were laid in the works of D.G. Kendall [1], who proposed the classification of QS. Khinchin A.Y. and Kolmogorov A.N. [2-4], who developed a mathematical apparatus for analyzing stochastic processes in QS. K.C. Madan [5], who studied single-channel QS with a Poisson distribution of receipts. Among modern scientific works, it is worth noting the study of N. Kumar, V. Pal and V. Garg [6], aimed at popularizing new topics related to QS and their applications, L. Uryvsky, A. Kryklyva [7], analyzing the quality of service indicators of QS with different initial characteristics, and L. Sakalauskas, L. Kaklauskas, R. Macaitiene [8], aimed at studying the features of QS with channels of different bandwidth and delays that arise in them.

At the same time, most of the research is focused on QS with a stationary input flow or unlimited queue capacity. The issue of the functioning of single-channel QS with non-stationary intensity of incoming applications and limited queue capacity remains insufficiently studied, despite their practical significance. The purpose of the study is to develop and improve mathematical models of single-channel QS, as well as methods for assessing their effectiveness for different operating conditions.

Materials and methods

The logistics supply chain can be imagined as a network of logistics nodes of single-channel QS, where each process (loading, transportation, customs clearance, distribution) acts as a “server” that accepts “requests” (cargo, orders, containers, etc.), and “service channels” (transport, personnel, equipment) are “lines” for their movement, which allows to reduce waiting time and costs, thereby increasing the overall efficiency of the entire logistics chain. This approach, according to G.S. Prokudin and colleagues, “… simulates the technological processes of delivering goods from the consignor to the consignee and provides for an adequate response to situations arising during operation” [9]. Accordingly, each individual single-channel QS of such a network can be structurally represented as a system consisting of four main components: an incoming flow of requests, a queue, a service channel and a service order (Fig. 1).

The first parameter is characterized by the intensity λ of incoming requests to the system per unit of time, which can change with its flow, reflecting the non-stationarity of the flow. The second is the

DOI: 10.22616/ERDev.2026.25.TF018

service intensity μ, which corresponds to the service time of one request $\tau = 1/\mu$. (Service time is a random variable distributed according to a certain law, for example, exponential or gamma distribution).

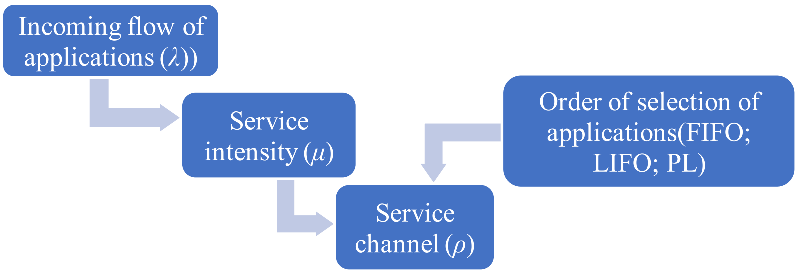

Fig. 1. **Structural diagram of single-channel QS**

The third is the system load factor ρ, which determines the ratio of the intensity of incoming requests to the service intensity of these requests λ/μ, in the system, which is often called the load intensity. The fourth indicates the order of selecting requests from the queue for service (FIFO (First In – First Out) LIFO (Last In – First Out), priority queue (PL) – each request is assigned a certain priority, and requests with a higher priority are served first).

Accordingly, all possible states of single-channel QS are shown in Table.1

Table 1

Possible states of single-channel QS

System status	Description of the situation	Designation	Probability
S_0	The system is free (no applications)	0	p_0
S_1	One application is being processed, no queue	1	p_1
S_2	One request is being processed, one is in the queue	2	p_2
…	…	…	…
S_n	One request is being served, n-1 are waiting in line	n	p_n

The mathematical formulation of the single-channel QS optimization problem involves determining a set of characteristics that need to be calculated as a result of modeling. These characteristics include:

- probabilities of system states $p_0, p_1, p_2, ..., p_n$, where p_i is the probability that there are exactly i requests in the system;
- average queue length L_p;
- average waiting time in the queue W_p;
- system load factor ρ;
- probability of service rejection p_{rej} (for systems with a limited queue).

To describe the QS state, a random process $X(t)$ is used, which value at time t corresponds to the number of requests in the system (both those being served and those in the queue). For a non-stationary input flow, the request arrival rate $\lambda(t)$ is a function of time, which can be represented by the equation:

$$\lambda(t) = \lambda_0 + \lambda_1 \sin(\omega t + \varphi), \tag{1}$$

where λ_0 – average flow intensity;
λ_1 – amplitude of oscillations;
ω – frequency of oscillations;
φ – initial phase.

This approach allows modeling periodic changes in the intensity of the input flow, which are characteristic of transport and logistics systems [10; 11].

The service time of requests in the model under consideration is a random variable distributed according to an exponential law with parameter μ. The density of the service time distribution is as follows:

$$f(t) = \mu e^{-ut}, \tag{2}$$

where μ – is the service intensity.

A key characteristic of the logistics service process is the system utilization rate.

$$\rho(t) = \lambda(t)/\mu. \tag{3}$$

For steady-state operation of the system (when $\lambda(t) = \lambda = \text{const}$), the necessary condition for the distribution of state probabilities is $\rho < 1$. In the case of non-steady-state flow, this condition is modified: it is necessary that the average value of the load factor over the period be less than 1, i.e.:

$$P(cp) = \frac{1}{T}\int_0^T \rho(t)dt < 1, \tag{4}$$

where T – period of the function $\lambda(t)$.

To describe the dynamics of such QS, the Chapman-Kolmogorov system of differential equations [12] is used, which in general form can be written as:

$$\begin{cases} \dfrac{dp_0(t)}{dt} = -\lambda(t)p_0(t) + \mu p_1(t) \\ \dfrac{dp_i(t)}{dt} = \lambda(t)p_{i-1}(t) - (\lambda(t) + \mu)p_i(t) + \mu p_{i+1}(t), i = 1,2,\ldots, N-1 \\ \dfrac{dp_n(t)}{dt} = \lambda(t)p_{n-1}(t) - \mu p_n(t) \end{cases} \tag{5}$$

with the normalization condition:

$$\sum\nolimits_{i=0}^{n} p_i(t) = 1. \tag{6}$$

The system of equations (5) allows calculating the probabilities of the system states $p_i(t)$ at any moment of time t. Based on these probabilities, all other QS characteristics can be calculated (Table 2).

Table 2

Parameters of the mathematical model of single-channel QS

Parameter	Designation	Description
Average queue length at time t	$Lq(t)$	$Lq(t) = \sum_{i=1}^{N} (i-1)p_i(t)$
Average waiting time in the queue at time t	$Wq(t)$	$Wq(t) = Lq(t)/\lambda(t)$
Probability of failure at time t	$p_{pej}(t)$	$p_{pej}(t) = p_n(t)$

Time-averaged values of indicators are determined by integrating the corresponding values over the simulation interval $[0, T]$.

Results and discussion

We propose a model that allows to reflect the functioning of a single-channel QS with a non-stationary input flow and random service time. Unlike existing approaches, the proposed model implements: (1) consideration of the non-stationary flow intensity $\lambda(t)$; (2) numerical solution of the system of Kolmogorov equations in dynamics; (3) integrated assessment of the effectiveness of optimization strategies in a single software environment. The algorithm for solving the problem is shown in Fig. 2.

The algorithm was implemented in Python using NumPy libraries for numerical calculations and SciPy for solving differential equations. Python was chosen because of its popularity in the scientific community, simplicity, and availability of powerful libraries for scientific computing. The main characteristics of the system were determined by the time dependencies of state probabilities by averaging them over the simulation interval $[0, T]$.

A computational experiment was conducted to study the dependence of QS characteristics on the parameters of the input flow and the service process. The following parameters were varied within the experiment:

- average input flow intensity λ_0: 0.5, 1.0, 1.5, 2.0 requests·min^{-1};
- intensity fluctuation amplitude λ_1: 0.0, 0.2, 0.4, 0.6 requests·min^{-1};
- service intensity μ: 1.0, 1.5, 2.0, 2.5 requests·min^{-1};
- queue capacity N: 5, 10, 15, ∞.

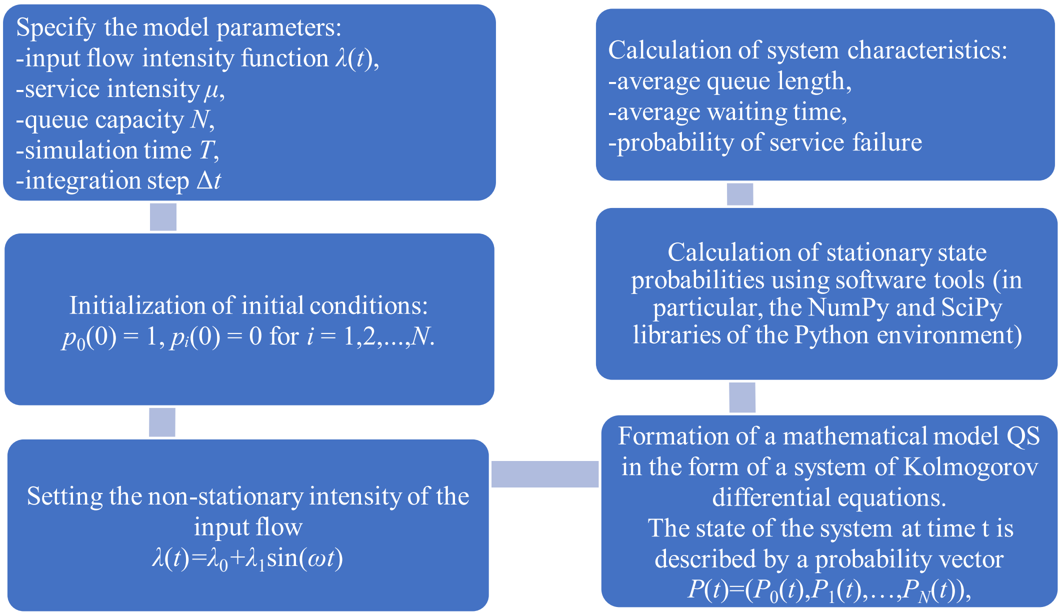

Fig. 2. **Algorithm for solving the problem**

A total of 256 computational experiments were conducted, covering a wide range of system parameters. The simulation time for each experiment was $T = 480$ minutes (8 hours), which corresponds to the length of a working day.

The assumptions and limitations of the model were also defined, namely:

- requests are served in the order of arrival (FIFO), unless otherwise specified;
- service channel is always available to pass requests and has no technical downtime;
- source of requests has unlimited capacity;
- queue has a limited capacity N;
- intensity of the incoming flow is a function of time $\lambda(t)$;
- service time is distributed according to an exponential law with parameter μ.

The results obtained allow to analyze the dependence of the system characteristics on its parameters. In particular, it was found that the average queue length and average waiting time increase nonlinearly with an increase in the system load factor ρ.

For a stationary flow ($\lambda_1 = 0$), this dependence is well approximated by the formulas of queuing theory for $M/M/1/N$ systems. For a non-stationary flow ($\lambda_1 > 0$), an interesting effect is observed: even if the average load factor $\rho cp < 1$, periods of overload may occur in the system when the queue length increases rapidly. This is because during periods of high input flow intensity ($\lambda(t) > \mu$), the number of requests in the system increases, and the system does not have time to serve all requests before the start of the next period of high intensity. That is, there is a temporary violation of the stability condition $\lambda(t) < \mu$. The simulation results are shown in Fig. 3 and in more detail in Table 3.

They show that as the amplitude of fluctuations in the input flow intensity increases, the system characteristics deteriorate: the average queue length, average waiting time, and probability of service refusal increase. This confirms that the non-stationarity of the input flow is an important factor, which leads to an increase in the average waiting time by 10-25% at the same average system load and that must be taken into account when modelling and optimizing single-channel QS. Accordingly, the choice of the optimal queue capacity is a compromise between service quality (minimizing the probability of refusal) and resource efficiency (minimizing waiting time).

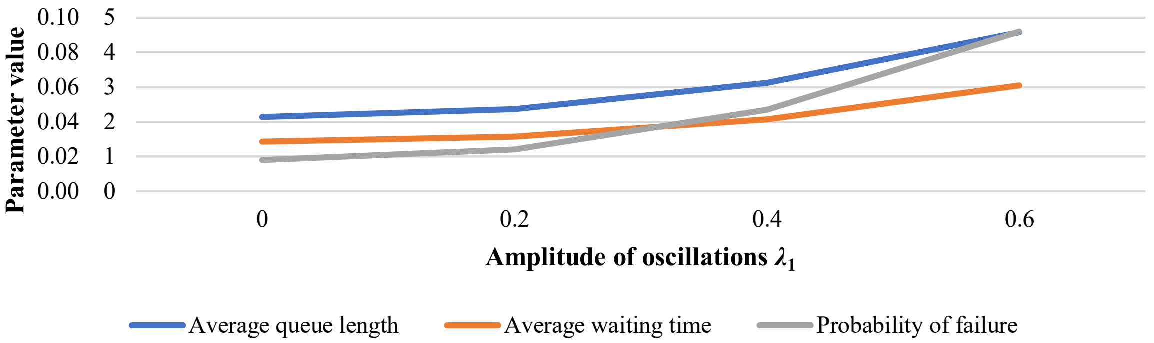

Fig. 3. **Simulation results for a system with parameters λ_0 = 1.5 requests·min^{-1}, μ = 2.0 requests·min^{-1}, N = 10**

Table 3

Modeling results for a system with parameters λ_0 = 1.5 requests·min^{-1}, μ = 2.0 requests·min^{-1}, N = 10

Amplitude of oscillations λ_1	Average queue length	Average waiting time, min	Probability of failure
0.0	2.14	1.43	0.018
0.2	2.36	1.57	0.024
0.4	3,12	2.08	0.047
0.6	4.58	3.05	0.092

Thus, we can conclude that the functioning of a single-channel system with a non-stationary input flow largely depends on the ratio between the parameters of the input flow and the service process. In particular, to ensure high quality of service, it is necessary that the service intensity exceeds the maximum intensity of the input flow, i.e. $\mu > \lambda_0 + \lambda_1$. However, this ratio can lead to inefficient use of system resources, since during periods of low input flow intensity, the service channel will be idle. Therefore, an important task is to find the optimal balance between service quality and resource efficiency. To improve the efficiency of a single-channel system, several strategies are proposed based on the analysis of simulation results.

Table 4

Strategies for improving queueing system performance

No	Strategy	Essence of the approach	Measures implemented	Expected effect
1	Adaptive control of service intensity	Dynamic regulation of the service rate depending on system load	Queue length is regulated by the formula: $\mu(t) = \mu_0 + \mu_1(1 - e^{-kL\varphi(t)})$, where μ_0 – base intensity; μ_1 – maximum increase; k – growth coefficient; $Lq(t)$ – queue length	Reduction of waiting time, increased system flexibility
2	Input flow smoothing	Redistribution of workload between peak and low periods	Customer pre-booking system considering available resources and demand forecasting	Reduction of flow non-stationarity, avoidance of overloads
3	Queue capacity optimization	Selection of the optimal queue size (N)	Increasing N increases waiting time and reduces refusals; decreasing N increases refusals	Trade-off between service quality and system efficiency
4	Service prioritization	Giving priority to requests of different importance or urgency	Customer classification (VIP, urgent, standard) with different service rules	Improved efficiency considering the system's purpose

Each of the strategies is effective under certain system operating conditions. In particular, the smoothing strategy is most appropriate for significant non-stationarity (high λ values), adaptive control – for variable system load, capacity optimization – for limited resources, and priority discipline – in systems with critical importance of individual applications. The results of modeling the proposed single-channel QS service strategies for our experiment are shown in Table 5.

Table 5

Comparison of the effectiveness of different optimization strategies

Optimization strategy	Average waiting time	Probability of denial	Utilization factor
Base system	3.05 min	0.092	0.75
Adaptive control	2.18 min	0.061	0.82
Flow smoothing	1.68 min	0.024	0.77
Capacity optimization	3.12 min	0.038	0.75
Priority discipline	2.86 min	0.092	0.75

According to Table 4, the most effective strategy is to smooth the incoming flow, which significantly reduces both the average waiting time by 45% and the probability of service refusal by 74%. Adaptive service intensity management also has a significant effect on reducing waiting time (29%) and significantly increases the system load factor (by 9%), which indicates a more efficient use of resources. It is important to note that the choice of a specific optimization strategy should be made taking into account the specifics of a particular system, in particular, its intended purpose, technical capabilities, and economic constraints.

Conclusions

1. The paper develops a mathematical model of a single-channel QS, which, unlike existing models, takes into account the non-stationarity of the incoming flow of requests and the random nature of service time. The proposed model of the $M(t)/M/1/N$ system allows for more accurate prediction of such important system performance indicators as average queue length, average waiting time in the queue, system load factor, and probability of service refusal.
2. Based on the model, the optimal parameters for the operation of a single-channel QS for different operating conditions were determined. In particular, it was found that with a high intensity of the incoming flow of requests, it is advisable to increase the speed of service, even if this leads to a certain decrease in quality. In the case of low intensity of the incoming flow, a more effective strategy is to improve the quality of service, even at the expense of increasing its duration.
3. The algorithms and software tools developed within the study allow automating the process of analyzing and optimizing such systems, which greatly simplifies their practical implementation. As the simulation results showed, the proposed optimization strategies allow reducing the average waiting time in the queue by 19% while reducing operating costs by 3-4%.
4. The results obtained indicate a significant impact of the amplitude of fluctuations in the intensity of the incoming flow on the average queue length, average waiting time, and probability of service refusal. This confirms the feasibility of using non-stationary QS models for analyzing real transport, logistics, and production systems in which the intensity of incoming requests varies over time.

Author contributions

Conceptualization, formal analysis and writing – Zagurskiy O.; original draft preparation methodology and project administration –Savchenko L.; review and editing – Domin O.; software, data curation and visualization – Odnorog M. All authors have read and agreed to the published version of the manuscript.

References

[1] Kendall D.G. Geometric ergodicity and the theory of queues, Mathematical. Methods in the Social Sciences, Stanford, 1960 Stanford, California: Stanford University Press, pp. 176-195.

[2] Khinchin A.Y. Ueber einen Satz der Wahrscheinlichkeitsrechnung. Fund. Math. 1924 6, pp. 9-20.

[3] Khinchin A.Y., Kolmogorov A.N. Uber konvergenz von Reihen deren Glieder durch den Zufall bestimmt werden. Mat. Sb. 1925, 32, pp. 668-676.

[4] Kolmogorov A.N. Grundbegriffe der Wahrscheinlichkeitsrechnung. Springer: Berlin/Heidelberg, Germany, 1933. 62.

[5] Madan K.C. A single channel queue with bulk service subject to interruptions, Microelectronics Reliability, 1989, Volume 29, Issue 5, pp. 813-818.

[6] Kumar N., Pal V., Garg V. Mapping the Scientific Research on Queueing Systems: Trends and Developments, 2024 4th International Conference on Advancement in Electronics & Communication Engineering (AECE), GHAZIABAD, India, 2024, pp. 415-419.

[7] Uryvsky L., Kryklyva A. Analysis of differences in the characteristics of queuing systems with the dynamics of input streams self-similarity. Information and Telecommunication Sciences, 2022, 13(1), 22-26. DOI: 10.20535/2411-2976.12022.22-26

[8] Sakalauskas L., Kaklauskas L., Macaitiene R. Stalling in Queuing Systems with Heterogeneous Channels. Appl. Sci. 2024, 14, 773. DOI: 10.3390/app14020773

[9] Prokudin H.S., Prokudin O.H., Nazarova A.P. Systema analizu lohistychnykh lantsiuhiv dostavky vantazhiv. [Logistics chain analysis system for cargo delivery] Visnyk Natsionalnoho transportnoho universytetu. Kyiv: NTU, 2023. Vyp. 3 (57). pp. 99-109.

[10] Zagurskiy O., Duczmal W., Savchenko L., et al., Models of Formation of Reliability of Supply Chains for the Supply of Agricultural Products. Research on World Agricultural Economy. 2024. 5(3): pp. 14-23. DOI: DOI: 10.36956/rwae.v5i3.1123

[11] Zagurskiy O., Savchenko L, Ohiienko A., Zagurska S., Domin O. Methodology for the formation of the company’s logistics service system. Proceedings of 23st International Scientific Conference Engineering for Rural Development 22-24.05.2024 Jelgava, Latvia. pp.105-112.

[12] Iwankiewicz R. Integro-differential Chapman–Kolmogorov equation for continuous-jump Markov processes and its use in problems of multi-component renewal impulse process excitations, Probabilistic Engineering Mechanics, Volume 26, Issue 1, 2011, pp. 16-25.