УДК 519.872

УПРАВЛЕНИЕ ПРИЕМОМ ЗАКАЗОВ В СИСТЕМЕ ДОСТАВКИ ТОВАРОВ С УЧЕТОМ КОНЕЧНОЙ ВМЕСТИМОСТИ СКЛАДА В ПУНКТЕ ВЫДАЧИ ПОСЫЛОК

А. Н. ДУДИН[1)]

[1)]*Белорусский государственный университет, пр. Независимости, 4, 220030, г. Минск, Беларусь*

Аннотация. Доставка посылок в пункты выдачи товаров стала очень популярной в связи с быстрым развитием интернет-торговли. Для эффективной работы интернет-торговли требуется создание новых моделей систем массового обслуживания. В настоящей статье процесс доставки товаров описывается как обработка посылки в двух зонах обслуживания. Обслуживание в первой зоне включает доставку посылки до пункта выдачи товаров, обслуживание во второй зоне – хранение посылки на складе до возможного ее получения клиентом. Вместимость склада является конечной. Если посылка прибывает на заполненный склад, то она теряется. В целях повышения эффективности работы системы доставки предлагается применять пороговое управление приемом посылок в первой зоне. Система доставки анализируется при достаточно общих предположениях о процессе прибытия посылок. Анализ выполняется путем рассмотрения соответствующим образом построенной многомерной цепи Маркова с непрерывным временем. Устанавливается и численно иллюстрируется зависимость основных показателей производительности системы доставки от вместимости склада и порога управления доступом. Формулируется и решается задача оптимизации.

Ключевые слова: система доставки посылок; пункт выдачи; тандемная очередь; управление доступом; многомерная цепь Маркова; марковский входной поток; оптимизация.

Образец цитирования:
Дудин АН. Управление приемом заказов в системе доставки товаров с учетом конечной вместимости склада в пункте выдачи посылок. *Журнал Белорусского государственного университета. Математика. Информатика.* 2025;1:68–78 (на англ.).
EDN: ZIAXUB

For citation:
Dudin AN. Admission control in a parcel delivery system with account of the finite capacity of the warehouse at the pick-up point. *Journal of the Belarusian State University. Mathematics and Informatics.* 2025;1:68–78.
EDN: ZIAXUB

Автор:
Александр Николаевич Дудин – доктор физико-математических наук, профессор; заведующий лабораторией прикладного вероятностного анализа кафедры теории вероятностей и математической статистики факультета прикладной математики и информатики.

Author:
Alexander N. Dudin, doctor of science (physics and mathematics), full professor; head of the laboratory of applied probabilistic analysis, department of probability theory and mathematical statistics, faculty of applied mathematics and computer science.
dudin@bsu.by

ADMISSION CONTROL IN A PARCEL DELIVERY SYSTEM WITH ACCOUNT OF THE FINITE CAPACITY OF THE WAREHOUSE AT THE PICK-UP POINT

A. N. DUDIN[a]

[a]*Belarusian State University, 4 Niezaliezhnasci Avenue, Minsk 220030, Belarus*

Abstract. Parcels delivering via pick-up points have become very popular due to the fast development of online marketplaces, customer-to-customer markets, and the use of parcel lockers as the last mile delivery solution. An adequate modelling of such delivery systems requires the creation of novel queueing models. In this paper, the process of delivering is described as the processing of a parcel in two service areas. The first area corresponds to processing until the parcel's arrival to the second area. The service in the second area corresponds to the storage of the parcel in a warehouse until it is picked up by a customer. The capacity of the warehouse is finite. A parcel arriving when the warehouse is full is lost. To enhance the efficiency of the system operation, threshold-type control by parcel admission at the first area is applied. Such a system is analysed under quite general assumptions about the parcel arrival flow. The analysis is implemented via consideration of the suitably constructed multidimensional continuous-time Markov chain. The dependence of the main performance measures of the delivery system on the warehouse capacity and the admission threshold are established and numerically illustrated. An optimisation problem is formulated and solved.

Keywords: parcel delivery system; pick-up point; tandem queue; admission control; multidimensional Markov chain; Markov arrival process; optimisation.

Introduction

As it is mentioned in [1], the business of parcel deliveries has been booming in recent years. From 64 bln parcels sent worldwide in 2016, the number of parcels has climbed up to over 161 bln in 2022 and it is forecast to reach 225 bln by 2028. Parcel delivery via the use of pick-up points, in particular the use of parcel lockers, is very popular as the solution for goods delivering by different online marketplaces (*Amazon*, *eBay*, *Rakuten*, *Shopee*, *AliExpress*, *Etsy*, *Walmart*, *Mercado Libre*, *Wildberries*, *Ozon*, etc.) and as the last mile delivery solution in a variety of out-of-home delivery. Therefore, the problem of enhancement of the organisation of this kind of delivery is intensively discussed in the existing literature (see, for example, [1–4]). The respective case studies can be found in [5–9].

Due to the uncertainty and randomness of the moments of parcel delivery to the pick-up point and the time until parcel withdrawal by the customer after its delivery to the pick-up point, it is evident that adequate mathematical models of parcel delivery processes can be constructed in terms of the theory of queues. Examples of application of this theory to this end can be found in [10–12].

In this paper, we propose and investigate the mathematical model of the order (parcels, goods, foods, etc.) delivering to the pick-up point and its storage there until receiving by the customer. Thus, we suppose that the lifetime of an arbitrary customer's order (parcel) in the delivery system consists of two parts, namely: (i) the time since the order generation in the system until its delivery to the target pick-up point and (ii) the sojourn time of the order in the warehouse of the target pick-up point until it is received by the customer or sent back to the sender. Correspondingly, an arbitrary order is sequentially processed in two service areas. The capacity of the second area, which models the pick-up point, is finite, and the orders arriving when it is exhausted are lost or returned to the sender. To mitigate the losses at the order entrance to this area, we propose the order admission control. New orders arriving in the first area are rejected if the total number of orders processed at two areas has the maximal admissible value (threshold). We explore the possibility of optimisation of operation of this delivery system via the proper choice of the capacity of the warehouse and the value of the threshold.

The rest of the text is organised as follows. The model of the order delivery is formulated in the queueing theory language in section «Mathematical model». The tridimensional Markov chain describing the operation of the delivery system is introduced, and its generator is derived in section «The Markov chain describing the system operation and its generator». Formulas for the main performance measures computation are presented in section «Performance measures computation». Numerical illustrations are given in section «Numerical examples», including examples of potential applications of the obtained results for managerial goals. Finally, section «Conclusions» concludes the paper.

Mathematical model

The structure of the system under study is presented in fig. 1.

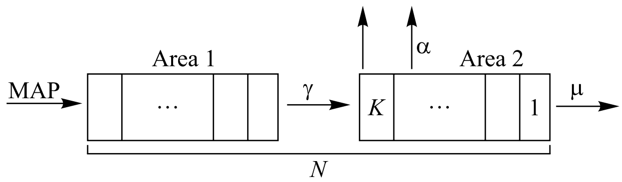

Fig. 1. The queueing system under study

Area 1 accommodates the orders that are in the process of delivery to the pick-up point. Area 2 models the stay of the order which is ready for picking up in the warehouse of the target pick-up point.

The process of parcel (order) arrival to the system is modelled by the MAP flow defined by the matrices D_0 and D_1 such that the matrix $D(1) = D_0 + D_1$ is the generator of the irreducible Markov chain with a continuous time ν_t, $t \geq 0$, having a finite state space $\{1, 2, \ldots, W\}$. The matrix D_1 consists of the intensities of transitions of the chain ν_t, $t \geq 0$, accompanied by the receiving of an order. The non-diagonal elements of the matrix D_0 determine the intensity of the corresponding transition of the chain ν_t, $t \geq 0$, without the receiving of an order, and the modules of the negative diagonal elements of the matrix D_0 determine the intensity of the exit of the process ν_t from the corresponding state of the Markov chain ν_t.

The average order rate λ is determined by the formula $\lambda = \boldsymbol{\theta} D_1 \boldsymbol{e}$ where $\boldsymbol{\theta}$ is a row vector of the invariant probabilities of the Markov chain ν_t, $t \geq 0$. This vector is the only solution to the system $\boldsymbol{\theta} D(1) = \boldsymbol{0}$, $\boldsymbol{\theta}\boldsymbol{e} = 1$. Here and below, $\boldsymbol{0}$ is a row vector of appropriate size consisting of zeros, and $\boldsymbol{e}$ is a column vector of appropriate size consisting of ones. Formulas for computation of other characteristics of the MAP, for example, the correlation coefficient of the successive interarrival times, variance, and higher moments of these times distribution, can be found in [13–18]. Let us also mention recent papers [19–29] devoted to the analysis of queues with the MAP.

Because the capacity of the real-world warehouses is limited, we assume that area 2 cannot accommodate more than K orders at the same time. Here, K is an arbitrarily fixed integer number. An order arriving when the area 2 capacity is exhausted, i. e. K orders are stored there, is not admitted to this area and is lost (sent back to the sender).

Due to this, aiming to reduce the probability of an arbitrary order loss after its service in area 1, it is desirable to apply some kind of order admission control in this area. We suggest that the reasonable control policy is as follows. A certain threshold N such as $N > K$ is fixed. An order arriving to area 1 is admitted for service if the total number of orders residing at the arrival moment in area 1 and area 2 is less than N. Otherwise, if this number is equal to N, the arriving order is rejected (lost at the entrance to area 1).

When the number of orders in the system at the arrival moment does not exceed $N - 1$, then a new order is admitted into the system, and the number of orders in area 1 increases by one; otherwise, the arriving order is lost.

The sojourn time of an order in area 1 (delivery time) is the exponentially distributed time with the parameter $\gamma > 0$. After this time expires, the order tries to enter area 2. If there are less than K orders residing in area 2 (the number of orders in the warehouse), the order arriving from area 1 is admitted for service in area 2. Otherwise, it is lost.

After the order is admitted to the warehouse, it is stored there the customer picks it up. The storage time of orders in the warehouse is limited. All customers of the system are divided into two groups. The first group is irresponsible customers that receive their orders only when the order storage time is almost complete. The second group is responsible customers who pick up their parcels earlier than the storage time runs out.

To take this into account, we assume that responsible customers pick up their orders after an exponentially distributed amount of time with the parameter $\mu > 0$. The order residence time in area 2 (the storage time of the order at the warehouse) is limited and it has an exponential distribution with the parameter $\alpha \geq 0$. When this time finishes, with a probability p, an irresponsible customer picks up her order (the order is considered to be successfully serviced), and with the complementary probability, the order is lost (returned from the warehouse to the sender).

It is obvious that the value of the threshold N has a profound effect on the performance of the considered system. If N is too small, then many arriving orders are not admitted to the system, and, due to the randomness of the arrival and picking-up processes, the starvation of area 2 can occur. As a consequence, the warehouse is underutilised, and the throughput of the system is low. If N is too large, there exists a high risk of an order rejection at the entrance to area 2 due to the lack of the place in the warehouse. Such a rejection again reduces the throughput of the system and leads to the waste of the system resources spent for the rejected order service in area 1. Therefore, the value of the threshold N has to be chosen optimally with respect to some fixed criterion. The main purpose of this paper is to provide an opportunity for computation of the optimal capacity K of the warehouse storage and the threshold N defining when the admission of new orders has to be temporarily suspended.

The Markov chain describing the system operation and its generator

Let the parameters K and N be fixed and let i_t be the number of orders in area 1, $i_t = \overline{0, N}$, k_t be the number of orders in area 2, $k_t = \overline{0, \min\{N - i_t, K\}}$, ν_t be the state of the underlying process of the MAP, $\nu_t = \overline{1, W}$, at time $t \geq 0$.

The behaviour of the system under study is described by a regular irreducible Markov chain with a continuous time

$$\xi_t = \{i_t, k_t, \nu_t\}, t \geq 0.$$

For easier analysis of this Markov chain, let us enumerate the states of the Markov chain in lexicographic order and call sub-level (i, k) the set of the states $(i, k, 1), (i, k, 2), \dots, (i, k, W)$. Level i is the set of sub-levels $(i, 0), (i, 1), \dots, (i, \min\{N - i, K\}), \overline{0, N}$.

Theorem 1. *The generator Q of the Markov chain ξ_t, $t \geq 0$, has the following block tridiagonal structure:*

$$Q = \begin{pmatrix} Q_{0,0} & Q_{0,1} & O & O & O & \dots & O \\ Q_{1,0} & Q_{1,1} & Q_{1,2} & O & O & \dots & O \\ O & Q_{2,1} & Q_{2,2} & Q_{2,3} & O & \dots & O \\ O & O & Q_{3,2} & Q_{3,3} & Q_{3,4} & \dots & O \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ O & O & O & O & O & \dots & Q_{N-1,N} \\ O & O & O & O & O & \dots & Q_{N,N} \end{pmatrix},$$

where the non-zero blocks $Q_{i,j}$, $|i - j| \leq 1$, containing the intensities of transitions from the states of level i to the states of level j are defined as follows:

$$Q_{i,i} = I_{K+1} \otimes D_0 - i\gamma I_{(K+1)W} - (\mu + \alpha) C_K \otimes I_W + (\mu + \alpha) C_K E_K^- \otimes I_W,\ i = \overline{0, N - K - 1}, \tag{1}$$

$$Q_{i,i} = I_{N-i+1} \otimes D_0 - i\gamma I_{(N-i+1)W} - (\mu + \alpha) C_{N-i} E_{N-i}^- \otimes I_W + \hat{I}_{N-i} \otimes D_1 + \\ + (\mu + \alpha) C_{N-i} E_{N-i}^- \otimes I_W,\ i = \overline{N - K, N}, \tag{2}$$

$$Q_{i,i+1} = I_{K+1} \otimes D_1,\ i = \overline{0, N - K - 1}, \tag{3}$$

$$Q_{i,i+1} = \tilde{E}_{N-i}^- \otimes D_1,\ i = \overline{N - K, N - 1}, \tag{4}$$

$$Q_{i,i-1} = i\gamma E^+ \otimes I_W,\ i = \overline{1, N - K}, \tag{5}$$

$$Q_{i,i-1} = i\gamma \tilde{E}_{N-i}^+ \otimes I_W,\ i = \overline{N - K + 1, N}. \tag{6}$$

Here, $\otimes$ is the Kronecker product of matrices (see, for example, [30]); I is the identity matrix, and O is the zero matrix, with the dimensionality in a subscript if it is necessary; $\delta_{\text{condition}}$ is the Kronecker delta, that is $\delta_{\text{condition}} = 1$ if condition is true and $\delta_{\text{condition}} = 0$ otherwise; $C_l = \text{diag}\{0, 1, 2, \dots, l\}$, i. e. C_l is the diagonal matrix with the diagonal entries $\{0, 1, 2, \dots, l\}$, $l = \overline{0, K}$; $\hat{I}_l$, $l = \overline{0, K}$, is a square matrix of size $l + 1$ with all zero elements except the element $(\hat{I}_l)_{l,l} = 1$; E_l^-, $l = \overline{0, K}$, is a square matrix of size l + 1 with all zero elements except the elements $(E_l^-)_{k,k-1} = 1$, $k = \overline{1, l-1}$; $\tilde{E}_l^-$, $l = \overline{1, K}$, is a matrix of size $(l + 1) \times l$ with all zero elements except the *elements* $\left(\tilde{E}_l^-\right)_{k,\,k} = 1,\ k = \overline{0,\ l-1}$; E^+ *is a square matrix of size* $K+1$ *with all zero elements except the elements* $\left(E^+\right)_{k,\,k+1} = 1,\ k = \overline{0,\ K-1}$, *and* $\left(E^+\right)_{K,\,K} = 1$; $\tilde{E}_l^+,\ l = \overline{0,\ K-1}$, *is a matrix of size* $(l+1)\times(l+2)$ *with all zero elements except the elements* $\left(\tilde{E}_l^+\right)_{k,\,k+1} = 1,\ k = \overline{0,\ l}$.

P r o o f. Proof of theorem 1 is performed via the analysis of all possible transitions of the Markov chain ξ_t during the time interval of an infinitesimal length. The block tridiagonal form of the generator is explained by the fact that the arrival and departure of the orders to area 1 occur one by one.

Let us first prove formulas (1) and (2) for the matrices $Q_{i,\,i}$. The size of these matrices is defined by the cardinality of the state space of the two-dimensional process $\{k_t,\ \nu_t\}$. The component ν_t has W possible states. The state space of the component k_t depends on the value of the component i_t. If the number i_t of orders at area 1 is small, namely, $i_t = \overline{0,\ N-K-1}$, then the component k_t admits the values in the range from 0 to K. If $i_t > N-K-1$, then the component k_t admits the values in the range from 0 to $N-i_t$. Correspondingly, the size of the matrices $Q_{i,\,i}$ is equal to $(K+1)\times W$ if $i = \overline{0,\ N-K-1}$ and to $(N-i+1)\times W$ if $i = \overline{N-K,\ N}$.

The diagonal entries of these matrices are negative. The modules of these entries define the rates of the exit of the Markov chain ξ_t from its states. The exit from the state $\{i,\ k,\ \nu\}$ is possible due to (i) service completion of one of i orders in area 1 (the corresponding rate is $i\gamma$), (ii) departure of one of k orders from area 2 (the corresponding rate is $k(\mu+\alpha)$), (iii) exit of the underlying process ν_t from its state (the corresponding rates are the modules of the diagonal entries of the matrix D_0).

The non-diagonal entries of the matrices $Q_{i,\,i}$ define the rates of transition of the Markov chain ξ_t to another state without the change of the value i of the component i_t. There are only two variants of such transitions: (i) transitions of the underlying process ν_t (rates of which are defined by the non-diagonal entries of the matrix D_0), (ii) departure of one of k orders from area 2. The rates of the departure are equal to $k(\mu+\alpha)$, and, as the consequence of this departure, the number k_t of orders at area 2 decreases by one. Transition probabilities of the component k_t from the state k to the state $k-1,\ k = \overline{1,\ K}$, are given by the entries of the matrix E_K^-.

As the result of these considerations, we obtain formula (1). Formula (2) which is valid for $i = \overline{N-K,\ N}$ is obtained via the same considerations. An additional summand $\hat{I}_{N-i}\otimes D_1$ appears here because a new order arrival (the corresponding rates are given by the entries of the matrix D_1) when the number of orders in area 2 is equal to $N-i$ (thus, the total number of orders in the system is equal to N) does not change the state of the system because this order is not admitted to the system.

The operation of the Kronecker product of matrices is used in (1), (2), and other formulas to describe transition rates of two independent Markov components $\{k_t,\ \nu_t\}$ of the Markov chain ξ_t.

Formulas (3), (4) are clear because the increase of the component i_t can occur only when a new order is admitted to area 1 which happens with the rates defined by the entries of the matrix D_1. A new order admission to area 1 does not imply any change in the number of orders in area 2. This explains the presence of the Kronecker multiplier I_{K+1} in formula (3), valid for $i = \overline{0,\ N-K-1}$. When $i = \overline{N-K,\ N-1}$, the increase in the value i implies the reduction of the state space of the component k_t of the chain. This reduction is implemented via the Kronecker multiplication of the matrix D_1 from the left by the non-square matrix $\tilde{E}_{N-i}^-$. This explains formula (4).

Transitions from the level i to the level $i-1$ occur with the rate $i\gamma$ when the service of one order at area 1 finishes. If $i = \overline{0,\ N-K}$, this order starts service at area 2, and the number of orders at this area increases by 1. The matrix E^+ describes transition probabilities of the component k_t of the Markov chain ξ_t at this moment. When $i = \overline{N-K+1,\ N}$, the decrease in the value i implies the extension of the state space of the component k_t of the chain. This extension is reflected by the non-square transition probability matrix $\tilde{E}_{N-i}^+$.

As the result of these considerations, we obtain formulas (5) and (6).

Theorem 1 is proven.

Because the Markov chain ξ_t is irreducible and has a finite state space, the stationary probabilities of its states

$$\pi(i,\ k,\ \nu) = \lim_{t\to\infty} P\{i_t = i,\ k_t = k,\ \nu_t = \nu\},\ i = \overline{0,\ N},\ k = \overline{0,\ \min\{N-i,\ K\}},\ \nu = \overline{1,\ W},$$

exist for any set of the system parameters.

Corresponding to the introduced enumeration of the states of the Markov chain ξ_t, let us introduce the row vectors $\pi(i, k)$ of the stationary probabilities of the states that belong to the sub-level (i, k) and vectors $\boldsymbol{\pi}_i$ of the stationary probabilities of the states that belong to the level i. The vectors $\boldsymbol{\pi}_i$, $i = \overline{0, N}$, satisfy the system of linear algebraic equations (Chapman – Kolmogorov equations)

$$(\pi_0, \pi_1, \ldots, \pi_N)Q = \mathbf{0}, (\pi_0, \pi_1, \ldots, \pi_N)\boldsymbol{e} = 1.$$

This system is finite. However, in potential real-life applications of the model, the capacity K of the warehouse can be large, and the number of equations in this system can be large as well. Therefore, to solve this system it is necessary to use algorithms that effectively account for the block tridiagonal structure of the generator Q and the fact that the solution of the system represents a probability distribution. The numerically stable algorithms obtained in [31; 32] are recommended for further use.

Performance measures computation

Having an opportunity to compute the vectors $\boldsymbol{\pi}_i$, $i = \overline{0, N}$, it is possible to calculate the values of the basic performance characteristics of the considered queueing model, in particular:

- the mean number of orders residing at an arbitrary moment in area 1

$$L_1 = \sum_{i=1}^{N} i\pi_i \boldsymbol{e};$$

- the mean number of orders residing at an arbitrary moment in area 2

$$L_2 = \sum_{i=1}^{N-1} \sum_{k=1}^{\min\{N-i, K\}} k\pi(i, k)\boldsymbol{e};$$

- the mean number of orders processed at an arbitrary moment in the system

$$L = \sum_{i=0}^{N} \sum_{k=0}^{\min\{N-i, K\}} (i + k)\pi(i, k)\boldsymbol{e} = L_1 + L_2;$$

- the rate of the output flow of serviced orders is

$$\lambda_{\text{out}} = \sum_{i=0}^{N-1} \sum_{k=1}^{\min\{N-i, K\}} k(\mu + (1-p)\alpha)\pi(i, k)\boldsymbol{e};$$

- the loss probability of an order at the entrance to the system

$$P_{\text{ent}} = \frac{1}{\lambda} \sum_{i=N-K}^{N} \pi(i, N-i) D_1 \boldsymbol{e}; \quad (7)$$

- the loss probability of an order when moving from area 1 to area 2

$$P_{\text{loss, 1-2}} = \frac{1}{\lambda} \sum_{i=1}^{N-K} \gamma i \pi(i, K)\boldsymbol{e}; \quad (8)$$

- the loss probability of an order due to lack of demand by the customer (expiration of the storage time)

$$P_{\text{imp}} = \frac{1}{\lambda} \sum_{i=0}^{N-1} \sum_{k=1}^{\min\{N-i, K\}} pk\alpha\pi(i, k)\boldsymbol{e}; \quad (9)$$

- the loss probability of an arbitrary order

$$P_{\text{loss}} = P_{\text{ent}} + P_{\text{loss, 1-2}} + P_{\text{imp}} = 1 - \frac{\lambda_{\text{out}}}{\lambda}.$$

The latter relation giving two alternative ways for computation of P_{loss} is useful for verification of the analytical results and computer implementation of procedures for computing the invariant distribution of the system states.

Formula (7) is valid because the loss of an order at the entrance occurs when the order sees N orders presenting in the system. The components of the row vector $\sum_{i=N-K}^{N} \pi(i, N-i)$ define the distribution of states of the underlying process of arrivals at such an arrival moment. The column of vector $\frac{1}{\lambda} D_1 \boldsymbol{e}$ defines the probabilities of an order arrival at an arbitrary moment under the fixed state of the underlying process [17]. Now, formula (7) evidently follows from the formula of total probability.

The probability $P_{\text{loss},\,1-2}$ is computed as the ratio of the rate of the lost orders during their transfer from area 1 to area 2 to the arrival rate which is equal to λ. The rate of customers transfer from area 1 is proportional to the number of orders at this area. The loss of the customers, which transfer to area 2, can occur only if the number of orders at area 1 is less or equal to $N - K$ while the number of orders at area 2 is equal to K. Therefore, the rate of the lost orders is $\sum_{i=1}^{N-K} \gamma i \pi(i, K)\boldsymbol{e}$. Formula (8) is proven.

The probability P_{imp} is also computed as the ratio of the rate of the customers which are lost due to expiration of the storage time to the arrival rate λ. Under the fixed numbers (i, k) of orders at areas 1 and 2, the rate of the loss is equal to $pk\alpha$. By implementing summation of the products of these rates by the probabilities $\pi(i, k)\boldsymbol{e}$ over the values of (i, k) under which the corresponding loss can happen we obtain formula (9).

Numerical examples

The aims of this section are to demonstrate the feasibility of the proposed algorithms for computation of the basic performance characteristics of the system, graphical illustration of their dependencies on the parameters K and N, and to show the possibility to use the obtained result for managerial purposes.

Let the MAP flow of orders be determined by the matrices

$$D_0 = \operatorname{diag}\{-8, -2\}, \; D_1 = \begin{pmatrix} 7.5 & 0.5 \\ 0.1 & 1.9 \end{pmatrix}.$$

This MAP has the mean arrival rate λ equal to 3, the coefficients of correlation and variation of the successive interarrival times are $c_{\text{cor}} = 0.17$ and $c_{\text{var}} = 1.625$ respectively.

The parameters of the exponential distributions of the service time of an order in area 1 (γ), the time till an order pick-up by the responsible customer (μ), and the storage time of an order at area 2 (α) are chosen as follows: $\gamma = 0.03$, $\mu = 0.02$, $\alpha = 0.004$. The probability p that an irresponsible customer picks up her order when the storage time expires is assumed to be equal to 0.03.

We vary the parameter N in the interval $[50, 500]$ with step 50 and the parameter K in the interval $[25, N]$ with step 25.

The dependence of the average numbers L_1 and L_2 of orders at areas 1 and 2 on the parameters N and K is presented in fig. 2 and 3.

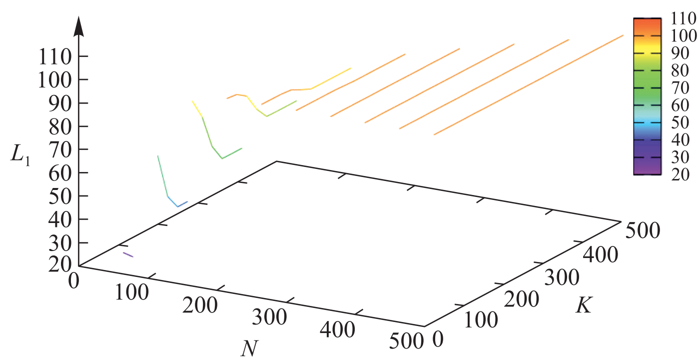

Fig. 2. The dependence of L_1 on the parameters N and K

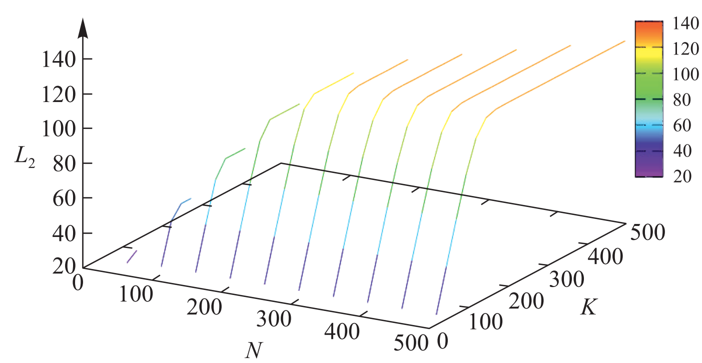

Fig. 3. The dependence of L_2 on the parameters N and K

One can see from these figures that the numbers of orders at both areas grow with the increase in the system capacity N. The number of orders in area 1 decreases, and the number of orders in area 2 increases with the increase in the parameter K. This can be explained as follows. The increase in K means that the capacity of area 2 increases, and more orders can be stored in this area which leads to the increase in L_2. At the same time, under the fixed value of N, the increase in the number of orders in area 2 leads to a shortness of capacity of area 1, and more arriving orders are rejected upon arrival. Thus, L_1 decreases with the growth in K.

The dependence of the loss probability P_{ent} of an order upon arrival on the parameters N and K is presented in fig. 4.

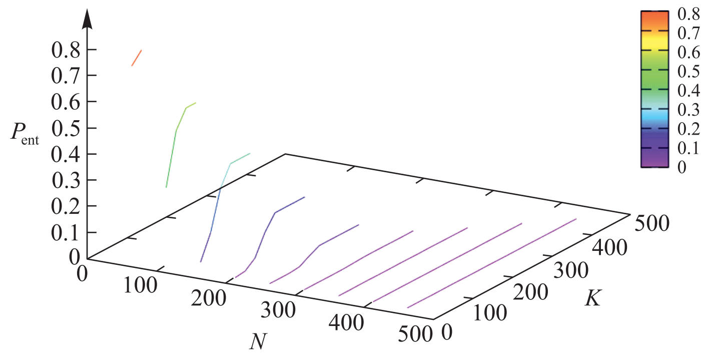

Fig. 4. The dependence of P_{ent} on the parameters N and K

This probability increases with an increase in K and decreases with an growth in N. The reason for this is explained in the same way as for L_1. Note that in the considered example for $N = 500$ the value of P_{ent} is less than $2.6 \cdot 10^{-12}$ for all values of K.

The dependence of the probability $P_{loss,\,1-2}$ of an order loss during the transfer from area 1 to area 2 on the parameters N and K is presented in fig. 5.

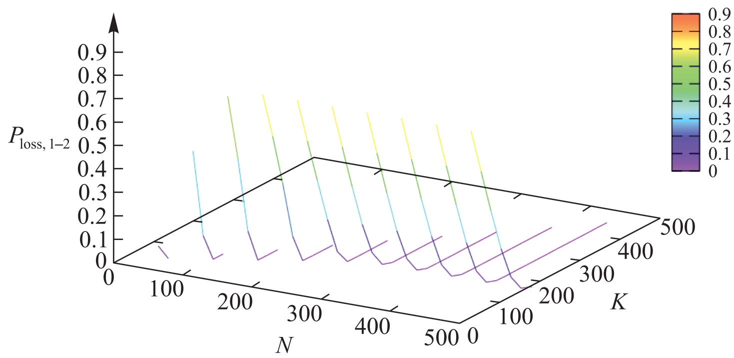

Fig. 5. The dependence of $P_{loss,\,1-2}$ on the parameters N and K

The loss probability $P_{loss,\,1-2}$ increases with the growth in the parameter N because an increase in N leads to an increase in the number of orders in area 1, and more orders arrive at area 2. Thus, when the capacity of the buffer K is not large, more orders are rejected due to the impossibility of placing them in area 2. When the capacity K increases, the loss probability $P_{loss,\,1-2}$ evidently decreases. In the considered example, for $K \geq 275$, the loss probability $P_{loss,\,1-2}$ becomes negligible (less than 10^{-10}).

The dependence of the probability P_{imp} of an order loss due to the storage time expiration in area 2 on the parameters N and K is presented in fig. 6. This probability increases with the increase in N and K because this probability is strongly related to the number of orders L_2 in area 2. Note that in the considered example, for large values of N and K, the loss probability P_{imp} takes the fixed value. This value shows us that in this example, regardless of how large capacities N and K are chosen, 0.5 % of orders are not picked up. When N and K are not large, this probability is less than 0.005 because the orders that won't be picked up can be lost for other reasons (upon arrival or during the transfer between the areas).

The dependence of the total probability P_{loss} of an order loss on the parameters N and K is presented in fig. 7. The loss probability P_{loss} decreases with the growth in N and K since the increase in buffer capacities creates better conditions for orders, and fewer of them are lost.

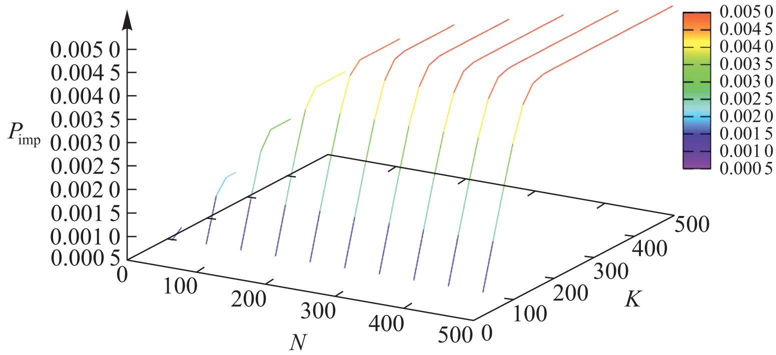

Fig. 6. The dependence of P_{imp} on the parameters N and K

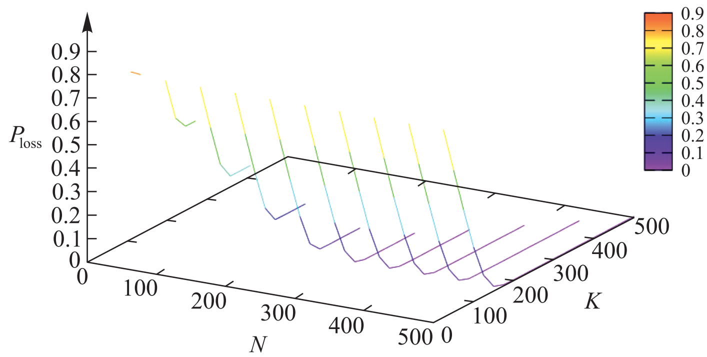

Fig. 7. The dependence of P_{loss} on the parameters N and K

Having highlighted the dependencies of performance measures of the system on the parameters N and K, it is possible to formulate an optimisation problem. Let us assume that the cost criterion, which evaluates the quality of the system operation, has the following form:

$$E = E(N, K) = a\lambda_{out} - c_1\lambda P_{ent} - c_2\lambda P_{loss,\,1-2} - c_3\lambda P_{imp} - dK.$$

Here, a is the system's revenue for each order service; c_1 is the penalty for an order loss upon arrival; c_2 is the penalty for an order loss due to the overflow of area 2; c_3 is the penalty for an order loss due to sending it back; d is the charge that is paid for maintenance of one place in area 2 per unit of time. Thus, the criterion E defines the average revenue of the system per unit time.

In this example, let us fix the following values of the cost coefficients:

$$a = 2,\ c_1 = 1,\ c_2 = 10,\ c_3 = 0.5,\ d = 0.01.$$

The dependence of the criterion E on the parameters N and K is presented in fig. 8.

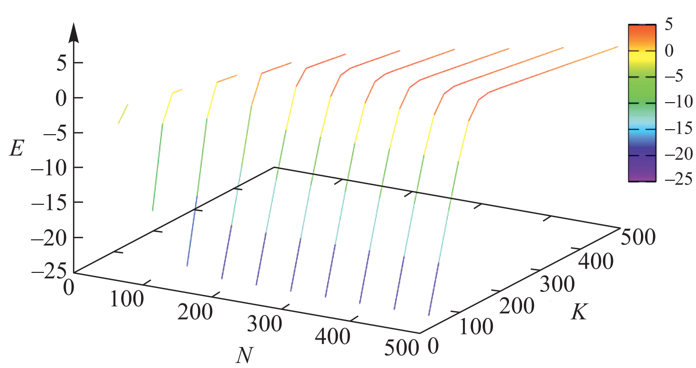

Fig. 8. The dependence of the cost criterion E
on the parameters N and K

Some additional information about the results of computation of the values of the cost criterion is presented in table.

The results of computation of values of the cost criterion (values of criterion $E(N, K)$ for different values of *K*, optimal value $N^*(K)$ of *N* for the given value of *K*, and values of the cost criterion $E(N, K)$ for *N* = *K* and *N* = 500)

K	$N^*(K)$	$E(N^*(K), K)$	$E(K, K)$	$E(500, K)$
25	50	–1.5247	–	–23.1304
50	50	–3.4191	–3.4191	–16.3107
75	75	0.1732	0.1732	–9.6164
100	150	1.6009	–0.0707	–3.3341
150	250	4.0153	1.3108	3.8594
200	400	3.9609	2.4601	3.9609
250	500	3.4625	3.0524	3.9609

N o t e. For $K > 250$, the optimal values of $N^*(K)$ are greater than 500.

Based on fig. 8 and table, we can conclude that the values of the criterion E strongly depend on the parameters N and K. For small values of K, due to the high probability of order rejection, the system's revenue is negative, i. e. the system suffers losses. The value of the losses depends on the parameter N. A greater value of N implies greater losses for example: $E = -3.626798$ for $N = 50$, $E = -15.159633$ for $N = 100$, and $E = -23.13035$ for $N \geq 250$.

When the parameter K of area 2 increases, the probabilities of order rejection decrease, and the throughput of the system increases, which implies the increase of the system revenue. If $K = 50$, then $E = -1.524699$ for $N = 50$, $E = -3.419056$ for $N = 100$, $E = -12.891237$ for $N = 150$, $E = -16.17527$ for $N = 200$ and $E = -16.13$ for $N \geq 250$. However, starting from a certain value of K, the cost criterion stops the increase with the growth of K (due to the high payment for maintenance of the warehouse of a large capacity) and further decreases with the growth of K. The optimal value of the cost criterion $E(N, K)$ in this example is equal to 4.165264 and is achieved when $K = 175$ and $N = 350$.

Conclusions

In this paper, the process of a customer's order delivery to the pick-up point and its stay there is described in terms of queueing theory. Each order has to be sequentially processed in two areas. Area 2 models the order maintenance in a warehouse of finite capacity before being picked up by the customer. To avoid overflow of the warehouse and sending the order back to the sender, order admission control in area 1 is proposed. This is the threshold type control. The arrival of orders to the system is described by a quite general Markov arrival process, which allows us to fit not only the mean arrival rate, as the stationary Poisson process does, but also the variance of interarrival times and their correlation in real-world systems. The existence of two types of customers, responsible and irresponsible, characterised by different distribution of time until an order withdrawal in pick-up point is taken into account.

Under the fixed values K of the capacity of the warehouse and admission control threshold N, the performance of the order delivering system is evaluated via the analysis of a suitably constructed multidimensional Markov chain. The possibility of computation of the values of various performance measures and the solution of optimisation problems is numerically illustrated.

The obtained results can be extended to the delivery system with the batch arrival of orders and their group picking up. Retrials of the customers that failed to make the order due to the system overflow can be considered as well. The results can be used for solving more global problem of routing the flows of parcel in a network between the alternative neighbouring pick-up points aiming to optimally match the rates of flows to the capacities of the warehouses at different pick-up points and their throughput.

References

1. Janinhoff L, Klein R, Sailer D, Schoppa JM. Out-of-home delivery in last-mile logistics: a review. *Computers and Operations Research.* 2024;168:106686. DOI: 10.1016/j.cor.2024.106686.

2. Nguyen TTT, Cabani A, Cabani I, De Turck K, Kieffer M. Analysis and forecasting of the load of parcel pick-up points: contribution of C2c e-commerce. *SSRN.* 2024:4887323. DOI: 10.2139/ssrn.4887323.

3. Wang L, Xu M, Qin H. Joint optimization of parcel allocation and crowd routing for crowdsourced last-mile delivery. *Transportation Research Part B: Methodological*. 2023;171:111–135. DOI: 10.1016/j.trb.2023.03.007.

4. Cardenas ID, Dewulf W, Vanelslander T, Smet C, Beckers J. The e-commerce parcel delivery market and the implications of home B2C deliveries *vs* pick-up points. *International Journal of Transport Economics.* 2017;44(2):235–256. DOI: 10.19272/201706702004.

5. Lamas-Fernandez C, Martinez-Sykora A, McLeod F, Bektas T, Cherrett T, Allen J. Improving last-mile parcel delivery through shared consolidation and portering: a case study in London. *Journal of the Operational Research Society*. 2024;75(6):1043–1054. DOI: 10.1080/01605682.2023.2231095.

6. van Duin JHR, Wiegmans BW, van Arem B, van Amstel Y. From home delivery to parcel lockers: a case study in Amsterdam. *Transportation Research Procedia*. 2020;46:37–44. DOI: 10.1016/j.trpro.2020.03.161.

7. Molin E, Kosicki M, van Duin R. Consumer preferences for parcel delivery methods: the potential of parcel locker use in the Netherlands. *European Journal of Transport and Infrastructure Research*. 2022;22(2):183–200. DOI: 10.18757/ejtir.2022.22.2.6427.

8. Cheng X, Liao S, Hua Z. A policy of picking up parcels for express courier service in dynamic environments. *International Journal of Production Research*. 2017;55(9):2470–2488. DOI: 10.1080/00207543.2016.1231431.

9. de Assis TF, de Abreu VHS, da Costa MG, D'Agosto MDA. Methodology for prioritizing best practices applied to the sustainable last mile – the case of a Brazilian parcel delivery service company. *Sustainability*. 2022;14(7):3812. DOI: 10.3390/su14073812.

10. Zabielska A, Jacyna M, Lasota M, Nehring K. Evaluation of the efficiency of the delivery process in the technical object of transport infrastructure with the application of a simulation model. *Eksploatacja i Niezawodnosc – Maintenance and Reliability*. 2023;25(1):1–12. DOI: 10.17531/ein.2023.1.1.

11. Hideyama S, Phung-Duc T, Okada Y. Queueing analysis of home delivery services with parcel lockers. *Lecture Notes in Computer Science*. 2019;11688:351–368. DOI: 10.1007/978-3-030-27181-7_21.

12. Liu Y, Escribano Macias JJ, Zhang K, Ye Q, Angeloudis P. Customer behaviour analysis for coordinated delivery service using mobile parcel lockers and autonomous robots: a congestion-game model. *SSRN*. 2023:4549554. DOI: 10.2139/ssrn.4549554.

13. Chakravarthy SR. The batch Markovian arrival process: a review and future work. *Advances in Probability Theory and Stochastic Processes.* 2001;1(1):21–49.

14. Chakravarthy SR. *Introduction to matrix-analytic methods in queues 1: analytical and simulation approach – basics*. London: John Wiley and Sons; 2022. 341 p. DOI: 10.1002/9781394165421.

15. Chakravarthy SR. *Introduction to matrix-analytic methods in queues 2: analytical and simulation approach – queues and simulation*. London: John Wiley and Sons; 2022. 448 p. DOI: 10.1002/9781394174201.

16. Lucantoni DM. New results on the single server queue with a batch Markovian arrival process. *Communication in Statistics. Stochastic Models*. 1991;7(1):1–46. DOI: 10.1080/15326349108807174.

17. Dudin AN, Klimenok VI, Vishnevsky VM. *The theory of queuing systems with correlated flows*. Cham: Springer; 2020. 410 p. DOI: 10.1007/978-3-030-32072-0.

18. Vishnevskii VM, Dudin AN. Queueing systems with correlated arrival flows and their applications to modeling telecommunication networks. *Automation and Remote Control*. 2017;78:1361–1403. DOI: 10.1134/S000511791708001X.

19. Klimenok VI. A queueing system with a batch Markovian arrival process and varying priorities. *Journal of the Belarusian State University. Mathematics and Informatics.* 2022;2:47–56. Russian. DOI: 10.33581/2520-6508-2022-2-47-56.

20. Dudina OS. Analytical modelling of systems with a ticket queue. *Journal of the Belarusian State University. Mathematics and Informatics.* 2024;2:40–53. Russian.

21. Dudin A, Dudin S, Melikov A, Dudina O. Framework for analysis of queueing systems with correlated arrival processes and simultaneous service of a restricted number of customers in scenarios with infinite buffer and retrials. *Algorithms*. 2024;17(11):493. DOI: 10.3390/a17110493.

22. Dudin A, Dudin S, Klimenok V, Dudina O. Stability of queueing systems with impatience, balking and non-persistence of customers. *Mathematics*. 2024;12(14):2214. DOI: 10.3390/math12142214.

23. Chakravarthy S, Melikov A. A new admission control scheme in queueing-inventory system with two priority classes of demands. *Opsearch*. 2024;11:1–34. DOI: 10.1007/s12597-024-00877-8.

24. Melikov A, Lawrence S, Sivakumar B. Analysis and optimization of hybrid replenishment policy in a double-sources queueing-inventory system with MAP arrivals. *Annals of Operations Research*. 2023;331(2):1249–1267. DOI: 10.1007/s10479-023-05646-2.

25. Mathew N, Joshua VC, Krishnamoorthy A, Melikov A, Mathew AP. A production inventory model with server breakdown and customer impatience. *Annals of Operations Research*. 2023;331(2):1269–1304. DOI: 10.1007/s10479-023-05659-x.

26. Melikov A, Chakravarthy SR, Aliyeva S. A retrial queueing model with feedback. *Queueing Models and Service Management*. 2023;6(1):63–95.

27. Dudin A, Dudin S, Manzo R, Rarità L. Queueing system with batch arrival of heterogeneous orders, flexible limited processor sharing and dynamical change of priorities. *AIMS Mathematics*. 2024;9(5):12144–12169. DOI: 10.3934/math.2024593.

28. Dudin SA, Dudina OS, Kostyukova OI. Analysis of a queuing system with possibility of waiting customers jockeying between two groups of servers. *Mathematics*. 2023;11(6):1475. DOI: 10.3390/math11061475.

29. Nair DV, Krishnamoorthy A, Melikov A, Aliyeva S. MMAP/(PH, PH)/1 queue with priority loss through feedback. *Mathematics*. 2021;9(15):1797. DOI: 10.3390/math9151797.

30. Graham A. *Kronecker products and matrix calculus with applications*. New York: Dover Publications; 2018. 144 p.

31. Baumann H, Sandmann W. Multi-server tandem queue with Markovian arrival process, phase-type service times, and finite buffers. *European Journal of Operational Research*. 2017;256(1):187–195. DOI: 10.1016/j.ejor.2016.07.035.

32. Dudin SA, Dudina OS. Call center operation model as a MAP/PH/N/N – R system with impatient customers. *Problems of Information Transmission*. 2011;47:364–377. DOI: 10.1134/S0032946011040053.

Received 12.12.2024 / revised 02.03.2025 / accepted 02.03.2025.