ISSN 0005-1179, Automation and Remote Control, 2017, Vol. 78, No. 11, pp. 1991–2003. © Pleiades Publishing, Ltd., 2017.
Original Russian Text © A.Z. Melikov, A.M. Rustamov, L.A. Ponomarenko, 2017, published in Avtomatika i Telemekhanika, 2017, No. 11, pp. 64–78.

STOCHASTIC SYSTEMS

Approximate Analysis of a Queueing–Inventory System with Early and Delayed Server Vacations

A. Z. Melikov[*,a], **A. M. Rustamov**[*,b], **and L. A. Ponomarenko**[**,c]

[*] *Institute of Control Systems, Azerbaijan National Academy of Sciences, Baku, Azerbaijan*
[**] *International Research and Training Center for Information Technologies and Systems of the National Academy of Sciences of Ukraine, Kiev, Ukraine*
e-mail: [a] *agassi.melikov@rambler.ru,* [b] *anar.rustemov@gmail.com,* [c] *laponomarenko@ukr.net*

Received March 24, 2017

Abstract—We propose a model for a servicing system with perishable inventory and a finite queue of impatient claims where the server can be in one of three states: operational, early and delaying vacations. We develop a method for approximate computation of the system's characteristics. We show results of numerical experiments.

Keywords: queueing–inventory systems, perishable inventory, server vacation, computational algorithms, optimization.

DOI: 10.1134/S0005117917110054

1. INTRODUCTION

Inventory control systems where servicing time for the claims is a positive value are called Queueing-Inventory Systems (QIS). A detailed survey of the literature devoted to such systems can be found in [1].

Literature analysis has shown that QIS models with server with vacations have been studied very little, although this servicing model from an economic point of view is more profitable compared to classical schemes since if there is no queue and/or inventory the server can often be used to perform secondary jobs; such models have been studied in quite some detail in classical queueing systems theory, see, e.g., [2–4].

In latest years, QIS models with a server with vacations have been considered in [5–13]; these works assumed that system inventory is durable, and the server starts its vacation instantaneously if the system's warehouse is empty and (or) the number of claims in the system is zero [5–12]. At the same time, QIS models where inventory has a finite lifetime and the server goes to vacation not instantaneously but during some (random) time "ponders" a decision to go to vacation (early vacation) are also of interest. And if during this "pondering" there appears a possibility to service claims, the server becomes available; otherwise it goes to a delayed vacation. The work [13] considers a system with durable inventory and repeat claims with instantaneous servicing of claims; it uses a server vacation scheme close to the one we propose here. The work [13] has studied a system where the server begins a random downtime period as soon as the inventory level becomes zero. Here either at the end or inside that period it becomes available when the inventory replenished again; otherwise the server goes to vacation for a random time and becomes available for a replenished inventory only at the end of that time, and under no replenishment a new downtime period begins and so on. In other words, downtime and server vacation period alternate, and both periods have exponential probability distribution functions (p.d.f.) with different means.

The works [5–13] used different modifications of Neuts' matrix-geometric method [14] to compute the state probabilities for multidimensional Markov chains that are mathematical models of the systems in question. As the authors of these works note, these methods are very hard to implement in practice and often suffer from computational instability due to the ill-posedness of the large dimensional matrices used in these methods. Similar problems arise in the spectral decomposition method [15, 16].

In this work, we study a QIS model with perishable inventory (Perishable Queueing-Inventory System, PQIS) and with early and delayed vacations of the server. Such PQIS models are used to analyze the operation of blood banks, systems for processing expirable information, supply systems for foodstuffs and so on [17–19]. To compute state probabilities of the corresponding three-dimensional Markov chain (3-D MC), we develop an approximate method based on hierarchical merging of the states [20]. This method lets us present explicit formulas for approximate computation of the characteristics of the system in question, which are used further to perform cost analysis for the system. Previously, the method of merging the states of a 2-D MC had been used to study systems with durable [21] and perishable inventories [22] in the presence of an "immovable" server.

2. MODEL DESCRIPTION AND PROBLEM SETTING

The system has a warehouse of volume S, $0 < S < \infty$, and contains one server for servicing claims. The incoming flow of claims is Poisson with parameter λ, $0 < \lambda < \infty$, and every claim requires a certain inventory of unit size, i.e., after servicing a claim the level of inventory in the warehouse reduces by one. Servicing times for claims are independent and identically distributed (i.i.d.) random values (r.v.) with a joint exponential p.d.f. with parameter μ, $0 < \mu < \infty$. Claims are impatient only during their sojourn time in the queue, i.e., a claim present at the server is patient. Admissible waiting times for the claims in a queue are i.i.d. r.v. with exponential p.d.f. with parameter τ, $0 < \tau < \infty$.

Each unit of inventory independently of the others becomes unsuitable for use after a random time that has exponential p.d.f. with parameter γ. Here we assume that inventory which is already at the step of transmission for a claim cannot perish.

The inventory warehouse is replenished according to the (s, S) policy, i.e., the system makes an order if the level of inventory reduces from value $s + 1$ to s, and here to exclude repeat orders we assume that the ordering point is $s < S/2$. The times when inventory is replenished are positive i.i.d. r.v. with exponential p.d.f. with parameter ν.

The server is in the operational state (status) only if both values, the level of inventory in the system and the length of the queue of claims, are positive. If at least one of these values is zero, the server goes to vacation, and here we distinguish two different periods of vacation for the server: early and delayed.

At the moment when the warehouse empties out, the system regardless of the queue state enters a period of early vacation of the server. The duration of this period is a random value that has exponential p.d.f. with parameter $\alpha > 0$. If during this period the inventory replenishes, and the queue has at least one claim, the server becomes available; otherwise it goes to delayed vacation. The server goes to early vacation with the same p.d.f. regardless of the inventory level if the system has no more claims, and here if inside that period a claim arrives and the level of inventory is greater than zero, the server becomes available; otherwise it goes for delayed vacation.

Since the incoming flow is Poisson, and time intervals between the orders are i.i.d. r.v. with exponential distribution, the true time the server spends in the state of early vacation is determined every time as a minimum of two exponentially distributed r.v. In other words, if the early vacation of the server is interrupted by an arriving claim, the true time of the server's early vacation has exponential distribution with parameter $\lambda + \alpha$, and if the early vacation is interrupted by arriving inventory, the true time the server spends on early vacation has an exponential distribution with parameter $\nu + \alpha$.

After the early vacation is over, the server goes for a delayed vacation whose duration is an r.v. with exponential p.d.f. with parameter $\beta > 0$. If during delayed vacation the level of inventory of this system and/or length of the queue of claims are zero, the server goes for delayed vacation again, and its duration has the same distribution; otherwise the server goes to operational mode.

The maximal system capacity, including the claim in the server, equals N, $0 < N < \infty$, i.e., a received claim is lost if at that moment the system already has N claims.

The problem is to find a joint distribution of the level of inventory for the system, the number of claims in the system, and the server's status. If we solve this problem we will know all necessary system characteristics. Due to space constraints, here we only consider three of them: average level of inventory in the warehouse (S_{av}); probability of losing a claim (PL); average intensity of the orders (RR).

3. MATHEMATICAL MODEL OF THE SYSTEM

System operation at an arbitrary moment of time is defined by a 3-D MC, and its states are defined by vectors $\mathbf{n} = (n_1, n_2, n_3)$, where the first and second components show respectively the current level of inventory and the number of claims in the system, and the third component denotes the server status, i.e.,

$$n_3 = \begin{cases} 1, & \text{if the server is in the operational state} \\ 2, & \text{if the server is in early vacation} \\ 3, & \text{if the server is in delayed vacation.} \end{cases}$$

The state space (SS) of this chain E is represented as a union of three mutually disjoint sets:

$$E = E_1 \bigcup E_2 \bigcup E_3, \tag{3.1}$$

where

$$\begin{aligned} E_1 &= \{\mathbf{n} : n_1 = 1, \ldots, S,\ n_2 = 1, \ldots, N,\ n_3 = 1\}, \\ E_2 &= \{\mathbf{n} : n_1 = 0,\ n_2 = 0, 1, \ldots, N,\ n_3 = 2;\ n_1 = 1, \ldots, S,\ n_2 = 0, n_3 = 2\}, \\ E_3 &= \{\mathbf{n} : n_1 = 0, 1, \ldots, S,\ n_2 = 0, 1, \ldots, N,\ n_3 = 3\}. \end{aligned}$$

The description of the system in question shows that transitions between states in the SS (3.1) are related to the following events:

(i) arrival of claims,

(ii) finishing the servicing process for a claim,

(iii) claims leaving the queue due to their impatience,

(iv) end of inventory lifetime,

(v) arrival of new inventory,

(vi) the server leaving for early vacation,

(vii) interrupting early vacation,

(viii) the server leaving for delayed vacation, and

(ix) the server returning from delayed vacation.

The intensity of transition from state $\mathbf{n}$ to state $\mathbf{n}'$ is denoted by $q(\mathbf{n}, \mathbf{n}')$, $\mathbf{n}, \mathbf{n}' \in S$. These values define the infinitesimal generator for this chain, and to construct it with regard to events (i)–(ix) it makes sense to distinguish the following cases in the choice of the original state: 1) $\mathbf{n} \in E_1$; 2) $\mathbf{n} \in E_2$; 3) $n \in E_3$.

Case $\mathbf{n} \in E_1$:

$$q(\mathbf{n}, \mathbf{n}') = \begin{cases} \lambda, & \text{if } \mathbf{n}' = \mathbf{n} + \mathbf{e}_2 \\ \mu, & \text{if } n_1 > 1, n_2 > 1,\ \mathbf{n}' = \mathbf{n} - \mathbf{e}_1 - \mathbf{e}_2 \text{ or} \\ & n_1 = 1 \text{ or } n_2 = 1, \mathbf{n}' = \mathbf{n} - \mathbf{e}_1 - \mathbf{e}_2 + \mathbf{e}_3 \\ (n_1 - 1)\,\gamma, & \text{if } \mathbf{n}' = \mathbf{n} - \mathbf{e}_1 \\ (n_2 - 1)\,\tau, & \text{if } \mathbf{n}' = \mathbf{n} - \mathbf{e}_2 \\ \nu, & \text{if } n_1 \leqslant s,\ \mathbf{n}' = \mathbf{n} + (S - s)\,\mathbf{e}_1 \\ 0 & \text{otherwise.} \end{cases} \tag{3.2}$$

Case $\mathbf{n} \in E_2$:

$$q(\mathbf{n}, \mathbf{n}') = \begin{cases} \lambda, & \text{if } n_1 = 0,\ \mathbf{n}' = \mathbf{n} + \mathbf{e}_2 \\ & \text{or } n_1 > 0,\ n_2 = 0,\ \mathbf{n}' = \mathbf{n} + \mathbf{e}_2 - \mathbf{e}_3 \\ \alpha, & \text{if } \mathbf{n}' = \mathbf{n} + \mathbf{e}_3 \\ n_2\tau, & \text{if } \mathbf{n}' = \mathbf{n} - \mathbf{e}_2 \\ \nu, & \text{if } n_2 > 0,\ \mathbf{n}' = \mathbf{n} + (S - s)\,\mathbf{e}_1 - \mathbf{e}_3 \\ & \text{or } n_1 \leqslant s,\ n_2 = 0,\ \mathbf{n}' = \mathbf{n} + (S - s)\,\mathbf{e}_1 \\ n_1\gamma, & \text{if } \mathbf{n}' = \mathbf{n} - \mathbf{e}_1 \\ 0 & \text{otherwise.} \end{cases} \tag{3.3}$$

Case $\mathbf{n} \in E_3$:

$$q(\mathbf{n}, \mathbf{n}') = \begin{cases} \lambda, & \text{if } \mathbf{n}' = \mathbf{n} + \mathbf{e}_2 \\ \beta, & \text{if } n_1 n_2 > 0,\ \mathbf{n}' = \mathbf{n} - 2\mathbf{e}_3 \\ n_1\gamma, & \text{if } \mathbf{n}' = \mathbf{n} - \mathbf{e}_1 \\ n_2\tau, & \text{if } \mathbf{n}' = \mathbf{n} - \mathbf{e}_2 \\ \nu, & \text{if } n_1 \leqslant s,\ \mathbf{n}' = \mathbf{n} + (S - s)\,\mathbf{e}_1 \\ 0 & \text{otherwise.} \end{cases} \tag{3.4}$$

Here $\mathbf{e}_i$ denotes the ith ortovector of the three-dimensional Euclidean space, $i = 1, 2, 3$.

Thus, the mathematical model for this PQIS is a three-dimensional Markov chain with SS (3.1), and its infinitesimal generator is defined from relations (3.2)–(3.4). This chain has a stationary distribution for all positive values of system parameters since it is finite and irreducible.

Let $p(\mathbf{n})$ denote the stationary probability of the state $\mathbf{n} \in E$. These values satisfy the system of equilibrium equations (SEE), composed on the basis of relations (3.2)–(3.4) (we do not show this SEE here explicitly since it is very cumbersome).

The system characteristics in question are defined via state probabilities of the above-described 3-D MC. The average level of inventory in the warehouse can be computed as

$$S_{av} = \sum_{k=1}^{S} k \sum_{\mathbf{n} \in E} p\,(\mathbf{n})\,\delta\,(n_1, k), \tag{3.5}$$

where $\delta\,(i, j)$ are Kroneker symbols.

Using full probability formulas, we find that the probability of losing a claim PL is defined as

$$PL = \sum_{k=1}^{3} P_k \times PL_k, \tag{3.6}$$

where P_1, P_2, P_3 is the probability that the server is in operational state, in early and delayed vacation respectively, PL_1, PL_2, PL_3 are probabilities of losing a claim when the server is in the operational state, in early and delayed vacation respectively. Probabilities of the server staying in various states are defined as follows:

$$P_k = \sum_{\mathbf{n}\in E} p\left(\mathbf{n}\right)\delta\left(n_3, k\right), \quad k = 1, 2, 3. \tag{3.7}$$

Probabilities of losing a claim when the server is in various states consist of two terms: (1) probability of losing a claim at its arrival time due to congestion in the buffer and (2) probability of losing a claim from the queue due to its impatience. In other words, we have

$$PL_k = \sum_{\mathbf{n}\in E_k} p\left(\mathbf{n}\right)\Big(\delta\left(n_2, N\right) + \overline{\delta}\left(n_2, N\right) P_k\left(n_1, n_2\right)\Big), \quad k = 1, 2, 3, \tag{3.8}$$

where $\overline{\delta}\left(u, v\right) = 1 - \delta\left(u, v\right)$ and $P_k(n_1, n_2)$ is the probability of the event that in state (n_1, n_2, k) the claim is lost due to impatience. Values $P_k(n_1, n_2)$, $k = 1, 2, 3$, can be computed from

$$\begin{aligned}
P_1(n_1, n_2) &= \frac{(n_2 - 1)\,\tau}{(n_2 - 1)\,\tau + \lambda I\left(n_2 < N\right) + (n_1 - 1)\,\gamma + \mu},\\
P_2(n_1, n_2) &= \frac{n_2\tau}{n_2\tau + \lambda I\left(n_2 < N\right)},\\
P_3(n_1, n_2) &= \frac{n_2\tau}{n_2\tau + \lambda I\left(n_2 < N\right) + n_1\gamma},
\end{aligned}$$

where $I(A)$ denotes the indicator function of A.

Since the inventory warehouse is replenished according to the (s, S) policy, to compute the average intensity of the orders we get the formula

$$RR = (\mu + s\gamma)\sum_{\mathbf{n}\in E_1} p(\mathbf{n})\delta(n_1, s+1) + (s+1)\gamma\sum_{i=2}^{3}\sum_{\mathbf{n}\in E_i} p(\mathbf{n})\delta(n_1, s+1). \tag{3.9}$$

We have noted above that to find state probabilities of this 3-D MC with matrix methods [14–16] proves to be inefficient or even impossible for models of large dimension. Therefore, in what follows we propose an alternative method for solving this problem.

4. APPROXIMATE ANALYSIS OF THE MODEL

The approximate method proposed here can be used under certain asymptotical conditions, namely: in what follows we assume that the intensity of claim arrival significantly exceeds the intensity of the server leaving for vacation (this condition quite plainly corresponds to the operation mode of real PQIS, see, e.g., [8]). Then, on the first level of hierarchy in SS (3.1) we introduce the merging function

$$U\left(\mathbf{n}\right) = < n_3 >, \quad \text{if} \quad \mathbf{n} \in E_{n_3},$$

where $< n_3 >$ is an merged state that includes all states from the class E_{n_3}, $n_3 = 1, 2, 3$. We denote $\Omega = \{< n_3 >: n_3 = 1, 2, 3\}$.

Then state probabilities for the original model are defined as follows:

$$p(\mathbf{n}) \approx \rho_{n_3}(n_1, n_2)\, \pi(< n_3 >), \tag{4.1}$$

where $\rho_{n_3}(n_1, n_2)$ is the probability of state (n_1, n_2) inside the split model with state space E_{n_3}, $\pi(< n_3 >)$ is the probability of merged state $< n_3 > \in \Omega$.

Remark 1. Here and below by a split (with respect to the original) model with given SS we understand a model that takes into account only connections between states included in its SS and disregards connections between states from different classes occurring in the SS partition in the original model.

Now (4.1) shows that to compute the stationary distribution for the original chain we will need to find the probabilities of states for three 2-D MC and one 1-D MC with three states. For large dimensions of SS (3.1) computational obstacles arise also in the finding of a stationary distribution for the 2-D MC with state space E_{n_3}, $n_3 = 1, 2, 3$. Therefore, to these chains we can also apply the merging procedure (second level of hierarchy).

We first consider the split model with state space E_1. Here and below in order to correctly apply the method we assume that the intensity of claims arrival significantly exceeds the intensity of inventory perishing in the system (see [8]). Under this assumption we consider in the set E_1 the following partition:

$$E_1 = \bigcup_{i=1}^{S} E_1^i, \quad E_1^i \bigcap E_1^j = \varnothing, \quad \text{if} \quad i \neq j, \tag{4.2}$$

where $E_1^i = \{(n_1, n_2) \in E_1 : n_1 = i\}$, $i = 1, \ldots, S$.

Then, based on splitting (4.2) we define the merging function

$$U_1((n_1, n_2)) =< n_1 >, \quad \text{if} \quad (n_1, n_2) \in E_1^{n_1},$$

where $< n_1 >$ is an merged state that includes all states from the class $E_1^{n_1}$. We denote $\Omega_1 = \{< n_1 >: i = 1, \ldots, S\}$.

Similar to (4.1) we have:

$$\rho_1(n_1, n_2) \approx \rho_1^{n_1}(n_2)\, \pi_1(< n_1 >), \tag{4.3}$$

where $\rho_1^{n_1}(n_2)$ is the probability of state (n_1, n_2) in a split model with state space $E_1^{n_1}$, $\pi_1(< n_1 >)$ is the probability of merged state $< n_1 > \in \Omega_1$.

In the class of states E_1^i, $i = 1, \ldots, S$, the first component is constant and equals i. Therefore, when we study split models with SS E_1^i the state $(i, n_2) \in E_1^i$ can be defined with only the second component, i.e., for convenience of exposition the state (i, n_2) is simply is denoted by n_2, $n_2 = 1, \ldots, N$. The intensity of transition between states n_2 and n_2' in the split model with SS E_1^i is denoted by $q_1(n_2, n_2')$. Then relations (3.2)–(3.4) imply tat

$$q_1(n_2, n_2') = \begin{cases} \lambda, & \text{if } n_2' = n_2 + 1 \\ (n_2 - 1)\,\alpha, & \text{if } n_2' = n_2 - 1 \\ 0 & \text{otherwise.} \end{cases} \tag{4.4}$$

Now (4.4) implies that state probabilities of all split models with SS E_1^i, $i = 1, \ldots, S$, are computed in the same way, i.e.,

$$\rho_1^i(n_2) = \frac{\sigma_\alpha(n_2 - 1)}{\sum\limits_{j=0}^{N-1} \sigma_\alpha(j)}, \quad n_2 = 1, 2, \ldots, N, \tag{4.5}$$

where $\sigma_\alpha(j) = \frac{(\lambda/\alpha)^j}{j!}$.

Transition intensities between merged states $< i >$, $< j > \in \Omega_1$, denoted as $q_1(< i >, < j >)$, are computed as follows:

$$q_1(< i >, < j >) = \begin{cases} (i-1)\gamma + \mu(1 - \rho_1(1)), & \text{if } j = i - 1 \\ \nu, & \text{if } i \leqslant s,\ j = i + S - s \\ 0 & \text{otherwise.} \end{cases} \tag{4.6}$$

Remark 2. In (4.6) the superscript of probability $\rho_1(1)$ is omitted since we define it in the same way for all split models (see (4.5)).

Then relations (4.6) for computing the probabilities of merged states $\pi_1(< n_1 >)$, $< n_1 > \in \Omega_1$, imply the following expressions (see [22]):

$$\pi_1(< n_1 >) = \begin{cases} a_1(n_1)\pi_1(< s+1 >), & \text{if } 1 \leqslant n_1 \leqslant s \\ b_1(n_1)\pi_1(< s+1 >), & \text{if } s+1 \leqslant n_1 \leqslant S-s \\ c_1(n_1)\pi_1(< s+1 >), & \text{if } S-s+1 \leqslant n_1 \leqslant S, \end{cases} \tag{4.7}$$

where

$$a_1(n_1) = \prod_{i=n_1+1}^{s+1} \frac{\Lambda_1(i)}{\nu + \Lambda_1(i-1)}; \quad b_1(n_1) = \frac{\Lambda_1(s+1)}{\Lambda_1(n_1)}; \quad c_1(n_1) = \frac{\nu}{\Lambda_1(n_1)} \sum_{i=n_1-S+s}^{s} a_1(i),$$

$$\Lambda_1(i) = \begin{cases} 0, & i = 1 \\ (i-1)\gamma + \mu(1 - \rho_1(1)), & 2 \leqslant i \leqslant S. \end{cases}$$

Probability $\pi_1(< s+1 >)$ can be found from the normalization condition, i.e.,

$$\sum_{i=1}^{S} \pi_1(< i >) = 1.$$

Further, using (4.5) and (4.7) we find from (4.3) the stationary distribution in the split model with SS E_1.

Now consider the split model with SS E_2. For this model we consider the following partition of space states:

$$E_2 = E_2^0 \bigcup E_2^1, \quad E_2^0 \bigcap E_2^1 = \varnothing, \tag{4.8}$$

where

$$E_2^0 = \{(n_1, n_2) \in E_2 : n_1 = 0; n_2 = 0, 1, \ldots, N\},$$
$$E_2^1 = \{(n_1, n_2) \in E_2 : n_1 = 1, \ldots, S; n_2 = 0\}.$$

Further, based on splitting (4.8) we define the merging function:

$$U_2(n_1, n_2) = \begin{cases} < 0 >, & \text{if } (n_1, n_2) \in E_2^0 \\ < 1 >, & \text{if } (n_1, n_2) \in E_2^1, \end{cases}$$

where $< k >$ is the merged state that includes all states from the class E_1^k, $k = 0, 1$. We denote $\Omega_2 = \{< k >: k = 0, 1\}$.

In the class of states E_2^0 the first component is constant and equal to zero, i.e., in models with SS E_2^0 state $(0, n_2)$ can be defined with only the second component n_2, $n_2 = 1, \dots, N$. The intensity of transition between states n_2 and n_2' in the split model with SS E_2^0 is denoted by $q_2^0\,(n_2, n_2')$. Relations (3.2)–(3.4) imply that

$$q_2^0\,(n_2, n_2') = \begin{cases} \lambda, & \text{if } n_2' = n_2 + 1 \\ n_2\tau, & \text{if } n_2' = n_2 - 1 \\ 0 & \text{otherwise.} \end{cases} \tag{4.9}$$

Now (4.9) implies that probabilities of states n_2 in the model with SS E_2^0, denoted by $\rho_2^0(n_2)$, are computed as state probabilities for the Erlang model $M/M/N/0$ with load λ/τ (Erl), i.e.,

$$\rho_2^0\,(n_2) = \frac{\sigma_\tau\,(n_2)}{\sum\limits_{j=0}^{N} \sigma_\tau\,(j)}, \quad n_2 = 0, 1, \dots, N, \tag{4.10}$$

where $\sigma_\tau\,(j) = \frac{(\lambda/\tau)^j}{j!}$.

In the class of states E_2^1, the second component is constant and equal to zero, i.e., in models with SS E_2^1 state $(n_1, 0)$ can be defined with only the first component n_1, $n_1 = 1, \dots, S$. The intensity of transition between states n_1 and n_1' in the split model with SS E_2^1 is denoted by $q_2^1\,(n_1, n_1')$. Here relations (3.2)–(3.4) imply that

$$q_2^1\,(n_1, n_1') = \begin{cases} \nu, & \text{if } n_1 \leqslant s,\ n_1' = n_1 + S - s \\ n_1\gamma, & \text{if } n_2' = n_2 - 1 \\ 0 & \text{otherwise.} \end{cases} \tag{4.11}$$

Similar to (4.7), (4.11) implies that probabilities of states n_1 of the model with SS E_2^1, denoted by $\rho_2^1\,(n_1)$, are computed as follows:

$$\rho_2^1(n_1) = \begin{cases} a_2(n_1)\rho_2^1(s+1), & \text{if } 1 \leqslant n_1 \leqslant s \\ b_2(n_1)\rho_2^1(s+1), & \text{if } s+1 \leqslant n_1 \leqslant S - s \\ c_2(n_1)\rho_2^1(s+1), & \text{if } S - s + 1 \leqslant n_1 \leqslant S, \end{cases} \tag{4.12}$$

where

$$a_2(n_1) = \prod_{i=n_1}^{s} \frac{\Lambda_2\,(i)}{\nu + \Lambda_2\,(i-1)}; \quad b_2(n_1) = \frac{s+1}{n_1}; \quad c_2(n_1) = \frac{\nu}{n_1\gamma} \sum_{i=n_1-S+s}^{s} a_2\,(i);$$

$$\Lambda_2\,(i) = \begin{cases} 0, & i = 0 \\ (i+1)\,\gamma, & 1 \leqslant i \leqslant s. \end{cases}$$

Probability $\rho_2^1(s+1)$ can be found from the normalization condition, taking into account that $\sum_{i=1}^{S} \rho_2^1\,(i) = 1$. Then (4.10) and (4.12) imply that intensities of transitions between merged states $< k >, < k' > \in \Omega_2$ can be found as follows:

$$q_2(< k >, < k' >) = \begin{cases} \nu\rho_2^0\,(0)\,, & \text{if } k = 0, \quad k' = 1 \\ \gamma\rho_2^1\,(1)\,, & \text{if } k = 1, \quad k' = 0. \end{cases} \tag{4.13}$$

Probabilities of merged states $\pi_2\left(<k>\right)$, $<k>\in\Omega_2$ are easy to compute from (4.13). Further, using (4.10) and (4.12) we can find, similar to (4.3), the stationary distribution in the split model with SS E_2.

Finally, we consider the split model with SS E_3. Similar to (4.2) here we consider the following partition:

$$E_3=\bigcup_{i=0}^{S}E_3^i,\quad E_3^i\bigcap E_3^j=\varnothing,\quad \text{if}\quad i\neq j, \tag{4.14}$$

where $E_3^i=\{(n_1,n_2)\in E_3 : n_1=i\}$, $i=0,1,\ldots,S$. Based on splitting (4.14) we define the merging function:

$$U_3\left((n_1,n_2)\right)=<n_1>,\quad \text{if}\quad (n_1,n_2)\in E_3^{n_1}, \tag{4.15}$$

where $<n_1>$ is the merged state that includes all states from the class $E_3^{n_1}$. We denote $\Omega_3=\{<n_1>: i=0,1,\ldots,S\}$.

Further, we repeat these procedures in the same way for the split model with SS E_1. Therefore, we will not go into the details of these procedures here but will only mention the differences. Note that probabilities of states $\rho_3^i\left(n_2\right)$, $n_2=0,1,\ldots,N$, in all split models with SS E_3^i, $i=0,1,\ldots,S$, are computed in the same way (do not depend on the superscript) as state probabilities for the classical Erlang model $M/M/N/0$ with load $\lambda/\tau(\mathrm{Erl})$ (see (4.10)).

Transition intensities between merged states $<i>$, $<j>\in\Omega_3$ of this model are defined as follows:

$$q_3(<i>,<j>)=\begin{cases} i\gamma, & \text{if } j=i-1\\ \nu, & \text{if } i\leqslant s,\quad j=i+S-s\\ 0 & \text{otherwise.}\end{cases} \tag{4.16}$$

Thus, relations (4.16) imply that probabilities of merged states $\pi_3\left(<n_1>\right)$, $<n_1>\in\Omega_3$, can be computed as follows:

$$\pi_3(<n_1>)=\begin{cases} a_3(n_1)\pi_3(<s+1>), & \text{if } 0\leqslant n_1\leqslant s\\ b_3(n_1)\pi_3(<s+1>), & \text{if } s+1\leqslant n_1\leqslant S-s\\ c_3(n_1)\pi_3(<s+1>), & \text{if } S-s+1\leqslant n_1\leqslant S,\end{cases} \tag{4.17}$$

where

$$a_3(n_1)=\prod_{i=n_1+1}^{s+1}\frac{i\gamma}{\nu+(i-1)\gamma};\quad b_3(n_1)=\frac{s+1}{n_1};\quad c_3(n_1)=\frac{\nu_0}{n_1\gamma}\sum_{i=n_1-S+s}^{s}a_3\left(i\right).$$

Probability $\pi_3(<s+1>)$ can be found from the corresponding normalization condition. Further, due to relations (4.10) and (4.17) similarly to (4.3) we can find stationary probabilities for the states of the split model with SS E_3.

Now to compute the stationary distribution for the original 3-D MC we will need to find the probabilities of merged states, i.e., $\pi(<k>)$, $<k>\in\Omega$ (see (4.1)). After certain transformations we get that intensities of transitions between states $<k>$, $<k'>\in\Omega$ are computed as follows:

$$q\left(<k>,<k'>\right)=\begin{cases} \mu\left(\pi_1(<1>)+\rho_1\left(1\right)\left(1-\pi_1\left(<1>\right)\right)\right), & \text{if } k=1,\quad k'=2\\ \nu\pi_2\left(<0>\right)\left(1-\rho_2^0\left(0\right)\right)+\lambda\pi_2\left(<1>\right), & \text{if } k=2,\quad k'=1\\ \alpha, & \text{if } k=2,\quad k'=3\\ \beta\left(1-\rho_3\left(0\right)\right)\left(1-\pi_3\left(<0>\right)\right), & \text{if } k=3,\quad k'=1,\\ 0 \quad \text{otherwise.} & \end{cases} \tag{4.18}$$

Remark 3. In (4.18) and below superscripts of probabilities $\rho_1(i)$ and $\rho_3(i)$ are omitted since they are defined in the same way for all split models.

Relations (4.18) easily yield the necessary probabilities $\pi(<k>)$, $<k>\in\Omega$.

Finally, for approximate computation of the stationary distribution for the original 3-D MC we find the following formulas:

for the cases $n_3 = 1$ and $n_3 = 3$:

$$p(n_1, n_2, n_3) \approx \rho_{n_3}(n_2)\pi_{n_3}(<n_1>)\pi(<n_3>); \tag{4.19}$$

for the case $n_3 = 2$:

$$p(n_1, n_2, 2) \approx \begin{cases} \rho_2^0(n_2)\,\pi_2(<0>)\,\pi(<2>), & \text{if } n_1 = 0 \\ \rho_2^1(n_1)\,\pi_2(<1>)\,\pi(<2>), & \text{if } n_1 > 0. \end{cases} \tag{4.20}$$

With (4.19) and (4.20), after certain transformations we get the following formulas for approximate computation of characteristics (3.5)–(3.9):

$$\begin{aligned} S_{av} &\approx \pi(<1>)\sum_{k=1}^{S} k\pi_1(<k>) + \pi(<2>)\,\pi_1(<1>)\sum_{k=1}^{S} k\rho_2^1(k) \\ &\quad + \pi(<3>)\sum_{k=1}^{S} k\pi_3(<k>); \\ P_k &\approx \pi(<k>), \quad k = 1,2,3; \\ PL_1 &\approx \pi(<1>)\left(\rho_1(N) + \sum_{k=1}^{S}\pi_1(<k>)\sum_{i=2}^{N-1}\rho_1(i)\,P_1(k,i)\right); \\ PL_2 &\approx \pi(<2>)\,\pi_2(<0>)\left(\rho_2^0(N) + \sum_{k=1}^{N-1}\rho_2^0(k)\,P_2(0,k)\right); \\ PL_3 &\approx \pi(<3>)\left(\rho_3(N) + \sum_{k=0}^{S}\pi_3(<k>)\sum_{i=1}^{N-1}\rho_3(i)\,P_3(k,i)\right); \\ RR &\approx (\mu + s\gamma)\,\pi(<1>)\,\pi_1(<s+1>) \\ &\quad + (s+1)\gamma\pi(<2>)\pi_2(<1>)\rho_2^1(s+1) + (s+1)\gamma\pi(<3>)\pi_3(s+1). \end{aligned}$$

5. NUMERICAL RESULTS

The resulting formulas let us study the behaviour of system characteristics with respect to the changes of any its load or structural parameters. At the same time, due to space constraints here we consider only one optimization problem for this system.

Suppose that all structural and load system parameters, apart from the intensity of the incoming flow of claims, have fixed values, and the only controlled parameter is the order point (i.e., s). Let us consider a rater important problem of finding such values of the parameter s that minimize the total costs (TC) related to system operation.

In stationary mode, the said costs are defined as follows:

$$TC(s) = c_r RR + c_h S_{av} + c_l \lambda PL, \tag{5.1}$$

where c_r is the cost of one order of inventory; c_h is the price of storing a unit of inventory volume per unit of time; c_l is the penalty for losing one claim.

Results of solving problem (5.2)

λ	5	6	7	8	11	13	16	19	22	25
s^*	8	9	10	11	12	13	14	15	16	17
TC^*	28.94	31.37	34.33	37.55	47.97	55.37	67.46	81.91	99.86	121.72
TC^*_{sim}	26.12	32.87	31.56	32.28	49.63	58.21	63.33	78.25	95.76	125.32

Thus, the optimization problem can be written as follows:

$$s^* = \arg\min \{TC(s) : 0 \leqslant s \leqslant \overline{s}\}, \tag{5.2}$$

where

$$\overline{s} = \begin{cases} \lceil S/2 \rceil, & \text{if } S \text{ is odd} \\ \lceil S/2 \rceil - 1, & \text{if } S \text{ is even,} \end{cases} \qquad \lceil x \rceil \quad \text{denotes the whole part of } x.$$

For all values of input parameters problem (5.2) has a solution since the set of possible solutions is discrete and finite. To solve (5.2) one can use standard minimization methods for functions of a discrete argument, including full enumeration.

The table shows the results of solving problem (5.2) for the following values of system parameters: $S = 90$, $N = 25$, $\mu = 4$, $\gamma = 1.4$, $\nu = 4$, $\tau = 6$, $\beta = 1.3$, $\alpha = 1$. Coefficients in the functional (5.1) were chosen as follows [8]: $c_r = 15$, $c_h = 0.2$, $c_l = 5$.

In the table, TC^* and TC^*_{sim} are minimal values of functional (5.1) computed with the approach we proposed and with the imitational approach respectively. In numerical experiments, the values of parameter λ changed with step one, and the table shows those values of this parameter for which the optimal solution of the problem changes. It is important to note that optimal solutions of problem (5.2) in both approaches are the same, which is not true for the values of functional (5.1).

The table shows that the optimal solution of problem (5.2) is piecewise constant and increases at a sufficiently slow rate with respect to the change of intensity of the incoming flow. It is also important to note that it is invariant in a relatively wide range of changing the intensity of the incoming flow of claims, which is very important from the practical point of view. Here we also note that for chosen original data for $\lambda > 28$ the optimal solution of problem (5.2) is always $s^* = 0$.

As for the time needed to solve problem (5.2) under various approaches, we note that the proposed formulas let us solve it in a few seconds while the imitational approach needs, in order to get more reliable results (with five percent confidence interval) several tens of hours on a computer with the following characteristics: memory command in MatLab; maximum possible array: 562 MB (5.890e+008 bytes) (limited by contiguous virtual address space available); memory available for all arrays: 2163 MB (2.268e+009 bytes) (limited by virtual address space available); memory used by MATLAB: 1743 MB (1.828e+009 bytes); physical memory (RAM): 8092 MB (8.485e+009 bytes).

6. CONCLUSION

In this work, we have proposed a model for a servicing system with one server and perishable inventory where impatient claims can form queues of limited length. The system adheres to the (s, S) policy of replenishing inventory, and the times of servicing claims and fulfilling orders are positive random values. The server is in the operational state if the system has at least one claim and inventory of size at least one. In the absence of inventory and/or claims the server goes for an early vacation over random time, and the server becomes available if during this period inventory arrives and the system has claims. From the early vacation state the server goes to the delayed vacation state and sojourns there for a random time. After the time in the state of delayed vacation is over, the server begins servicing claims if inventory is present; otherwise it goes for a vacation period again. We have shown that the mathematical model of this system is a three-dimensional Markov chain. We have developed both an exact and an approximate method to find out its characteristics, and here the exact method is based on solving the SEE for state probabilities and is efficient only for systems of moderate dimension. The approximate approach is based on hierarchical merging of states in the three-dimensional Markov chain and can be applied for asymptotic analysis of systems of any dimension.

REFERENCES

1. Krishnamoorthy, A., Lakshmy, B., and Manikandan, R., A Survey on Inventory Models with Positive Service Time, *OPSEARCH*, 2011, vol. 48, no. 2, pp. 153–169.
2. Doshi, B.T., Queueing Systems with Vacations: A Survey, *Queueing Syst.*, 1986, vol. 1, no. 1, pp. 29–66.
3. Tian, N. and Zhang, Z.G., *Vacation Queueing Models*, Heidelberg: Springer, 2006.
4. Ke, J.C., Wu, C.H., and Zhang, Z.G., Recent Developments in Vacation Queueing Models, *Int. J. Oper. Res.*, 2010, vol. 7, no. 4, pp. 3–8.
5. Daniel, J.K. and Ramanarayanan, R., An (s, S)-Inventory System with Two Servers and Rest Periods to the Server, *Naval Res. Logistics*, 1988, vol. 35, pp. 119–123.
6. Narayanan, V.C., Deepak, T.G., Krishnamoorthy, A., et al., On an (s, S)-Inventory Policy with Service Time, Vacation to Server and Correlated Lead Time, *Quality Technol. Quantit. Manage.*, 2008, vol. 5, pp. 129–143.
7. Sivakumar, B., An Inventory System with Retrial Demands and Multiple Server Vacations, *Quality Technol. Quantit. Manage.*, 2011, vol. 8, pp. 125–146.
8. Jajaraman, B., Sivakumar, B., and Arivarignan, G., A Perishable Inventory System with Postponed Demands and Multiple Server Vacations, *Model. Simul. Eng.* (Hindawi Publishing Corporation), 2012, vol. 2012, article ID 620960, 17 p.
9. Jeganathan, K., Perishable Inventory System at Service Facilities with Multiple Server Vacations and Impatient Customers, *J. Statist. Appl. Probab. Lett.*, 2014, vol. 3, no. 3, pp. 63–73.
10. Kathiresan, S., Anbazhagan, N., and Jeganathan, K., An Inventory System with Retrial Demands and Working Vacation, *Int. J. Sci. Res. Publ.*, 2014, vol. 4, no. 12, pp. 1–25.
11. Jeganathan, K., Anbazhagan, N., and Vigneshwaran, B., Perishable Inventory System with Server Interruptions, Multiple Server Vacations and Policy, *Int. J. Oper. Res. Inform. Syst.*, 2015, vol. 6, no. 2, pp. 32–52.
12. Yadavalli, V.S.S. and Jeganathan, K., Perishable Inventory Model with Two Heterogeneous Servers Including One with Multiple Vacations and Retrial Customers, *J. Control Syst. Eng.*, 2015, vol. 3, no. 1, pp. 10–34.
13. Padmavathi, I., Sivakumar, B., and Arivarignan, G., A Retrial Inventory System with Single and Modified Multiple Vacation for Server, *Ann. Oper. Res.*, 2015, vol. 233, no. 1, pp. 335–364.
14. Neuts, M.F., *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*, Baltimore: John Hopkins Univ. Press, 1981.
15. Mitrani, I. and Chakka, R., Spectral Expansion Solution for a Class of Markov Models: Application and Comparison with the Matrix-Geometric Method, *Perform. Evaluation*, 1995, vol. 23, pp. 241–260.
16. Chakka, R., Spectral Expansion Solution for Some Finite Capacity Queues, *Ann. Oper. Res.*, 1998, vol. 79, pp. 27–44.
17. Nahmias, S., Perishable Inventory Theory: A Review, *Oper. Res.*, 1982, vol. 30, pp. 680–708.
18. Nahmias, S., *Perishable Inventory Theory*, Heidelberg: Springer, 2011.

19. Baron, O., Berman, O., and Perry, D., Continuous Review Inventory Models for Perishable Items Ordered in Batches, *Math. Method. Oper. Res.*, 2010, vol. 72, pp. 217–247.

20. Koroliuk, V.S., Melikov, A.Z., Ponomarenko, L.A., et al., Methods for Analysis of Multi-Channel Queueing Models with Instantaneous and Delayed Feedbacks, *Cybernet. Systems Anal.*, 2016, vol. 52, no. 1, pp. 58–70.

21. Melikov, A.Z., Ponomarenko, L.A., and Bagirova, S.A., Analysis of Queueing-Inventory Systems with Impatience Consume Customers, *J. Autom. Inform. Sci.*, 2016, vol. 48, no. 1, pp. 53–68.

22. Melikov, A.Z., Ponomarenko, L.A., and Shahmaliyev, M.O., Perishable Queueing-Inventory Models with Repeated Consumer Customers, *J. Autom. Inform. Sci.*, 2016, vol. 48, no. 6, pp. 22–38.

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