

© 2019 г. А.З. МЕЛИКОВ, чл.-корр. НАН Азербайджана, д-р техн. наук
(agassi.melikov@gmail.com)

(Институт систем управления НАН Азербайджана, Баку),
М.О. ШАХМАЛЫЕВ (mamed.shahmaliyev@gmail.com)

(Национальная академия авиации, Баку)

СИСТЕМА ОБСЛУЖИВАНИЯ М/М/1/ ∞ С ПОРТЯЩИМИСЯ ЗАПАСАМИ И ПОВТОРНЫМИ ЗАЯВКАМИ

Предложена модель системы обслуживания с одним сервером, портящимися запасами и повторными заявками, которые могут образовывать орбиту бесконечного размера. При отсутствии запасов в системе первичные заявки согласно схеме Бернулли либо становятся в очередь, либо уходят в орбиту. В системе используется (s, S) -политика пополнения запасов. Разработан метод расчета характеристик системы и решается задача минимизации суммарных штрафов за счет выбора критического уровня запасов.

Ключевые слова: система обслуживания–запасания, портящиеся запасы, повторные заявки, алгоритм расчета, оптимизация.

DOI: 10.1134/S0005231019010057

1. Введение

Первыми публикациями, посвященными изучению систем обслуживания–запасания (Queuing-Inventory Systems, QIS) с долговечными запасами при наличии повторных заявок, являются [1, 2]. В них авторы независимо друг от друга различными методами изучили марковскую модель QIS с мгновенным обслуживанием и (s, S) -политикой пополнения запасов. В этой модели считается, что если в момент поступления первичной заявки (p -заявки) в системе отсутствуют запасы, то она поступает в орбиту бесконечного размера; с орбиты повторные заявки (r -заявки) повторяют попытку быть обслуженными после некоторого случайного времени, которое имеет показательную функцию распределения (ф.р.). В качестве математической модели изучаемой системы используется двумерная цепь Маркова (ЦМ), при этом в [1] для нахождения характеристик системы применяется метод производящих функций, а в [2] используется метод преобразования Лапласа. Подобные модели с конечным размером орбиты для настойчивых r -заявок и при использовании $(S - 1, S)$ -политики изучены в [3, 4].

Модели QIS с долговечными запасами и положительным временем обслуживания заявок (т.е. время обслуживания заявок больше нуля) при наличии повторных заявок изучены в [5–9]. Так, в [5] изучены три класса Марковских моделей QIS с одним сервером, r -заявками и (s, S) -политикой пополнения запасов. Для каждого типа моделей построены соответствующие трехмерные ЦМ и показано, что они имеют трехдиагональную производящую матрицу (ПМ); с использованием матрично-геометрического метода [10] найдены стационарные распределения и характеристики изучаемых моделей QIS. В [6] изучается марковская модель QIS с (s, S) -политикой, где серверы системы рассматриваются как запасы, при этом запасы пополняются мгновенно.

Если в момент поступления заявки все серверы системы заняты, то заявка уходит в орбиту бесконечного размера. В [6] построена соответствующая трехмерная ЦМ и вычислены характеристики системы. В [7] изучена марковская модель QIS с конечной очередью высокоприоритетных p -заявок. Обслуживание низкоприоритетных заявок не приводит к уменьшению уровня запасов системы, при этом они обслуживаются лишь тогда, когда в моменты их поступления сервер системы свободен; иначе они уходят в орбиту конечного размера, при этом в орбите они становятся нетерпеливыми. В [7] построена четырехмерная ЦМ для нахождения совместного распределения числа p -заявок в очереди, числа r -заявок в орбите и уровня запасов системы. Искомое распределение находится в результате решения матричных уравнений. В [8] рассмотрена марковская модель QIS с ограниченной очередью p -заявок, в которой после завершения обслуживания согласно схеме Бернулли заявка либо уходит в орбиту, либо покидает систему. Орбита имеет конечную емкость и сервер обслуживает r -заявки лишь тогда, когда в момент ее генерации в системе нет p -заявок и/или нет запасов, при этом r -заявки не требуют запасы. Построена трехмерная ЦМ и с применением матричного метода [10] предложен алгоритм расчета ее стационарного распределения. В [9] рассмотрена марковская модель QIS с одним сервером и без мест для ожидания p -заявок. Если в момент поступления p -заявки сервер свободен и уровень запасов больше нуля, то эта заявка принимается на обслуживание; если в момент поступления p -заявки сервер свободен, но уровень запасов равно нулю, то эта заявка окончательно покидает систему. Поступившая p -заявка уходит в орбиту бесконечного размера, если в этот момент сервер занят; если в момент поступления r -заявки уровень запасов равен нулю или сервер занят, то она возвращается в орбиту. В [9] построена соответствующая трехмерная ЦМ и с применением матричного метода [10] предложен алгоритм расчета ее стационарного распределения.

Модели QIS с портящимися запасами (Perishable QIS, PQIS) и повторными заявками мало изучены. Авторам известны лишь публикации [11, 12]. В [11] изучена модель с одним сервером и МАР-поток, в которой время обслуживания заявок имеет ф.р. фазового типа. В [12] предполагается, что времена выполнения заказов, пребывания заявок в орбите и жизни запасов имеют показательные ф.р. В системе используется (s, S) -политика, и заявки уходят в орбиту, если в момент их поступления все места в ограниченной очереди заняты. Математической моделью системы является пятимерная ЦМ. С помощью матричного метода [10] разработан алгоритм расчета характеристик системы. В [12] изучена модель PQIS с отдыхающим сервером, простейшим потоком и мгновенным обслуживанием. Как и в [11] предполагается, что в системе используется (s, S) -политика и времена выполнения заказов, пребывания заявок в орбите и жизни запасов имеют показательные ф.р. Сервер уходит в многократный отдых, если уровень запасов равно нулю. Если в момент поступления заявки уровень запасов равен нулю или сервер находится в отдыхе, то она уходит в орбиту бесконечного размера. Математической моделью системы является трехмерная ЦМ и с помощью матричного метода разработан алгоритм расчета характеристик системы. Отметим, что в [11, 12] с целью обеспечения возможности применения матричного метода принима-

ется не реальное допущение о том, что интенсивность поступления заявок с орбиты является постоянной величиной и не зависит от числа заявок в орбите.

Модель PQIS с мгновенным обслуживанием и повторными заявками изучена в [13], где можно найти подробный обзор публикаций, посвященных изучению моделей PQIS.

На русском языке непрерывные модели систем управления с портящимися запасами были исследованы в [14, 15]. Так, в [15] изучена модель системы управления с непрерывно портящимися запасами, в которой пополнение запасов происходит также непрерывно. Считается, что моменты поступления заявок образуют простейший поток с интенсивностью, зависящей от цены продажи запасов, при этом размер покупок является непрерывной случайной величиной. В [15] при релейном управлении интенсивностью потока и определенных асимптотических условиях, налагаемых на скорости пополнения запасов, найдена плотность распределения уровня запасов системы.

В настоящей статье изучается модель PQIS с одним сервером, простейшим потоком, положительным временем обслуживания заявок, неограниченной очередью первичных заявок и повторными заявками. Считается, что при отсутствии запасов поступившая p -заявка согласно схеме Бернулли либо становится в очередь, либо уходит в орбиту бесконечного размера. При отсутствии запасов заявки в очереди проявляют нетерпеливость, а r -заявки в орбите являются весьма настойчивыми. После завершения обслуживания заявка согласно схеме Бернулли либо получает запас, либо отказывается получить запас. Такая схема обслуживания впервые независимо друг от друга была введена в [7, 16] и далее она применена в [17]. В статье предложена математическая модель системы в виде трехмерной ЦМ, для изучения которой применяется метод иерархического фазового укрупнения состояний системы [18].

Статья организована следующим образом. В разделе 2 приводится описание модели и дана постановка задачи исследования. В разделе 3 построена математическая модель изучаемой системы в виде трехмерной цепи Маркова, разработан алгоритм построения ее производящей матрицы и получены формулы для расчета точных значений характеристик системы. Приближенный метод расчета характеристик системы разработан в разделе 4. Результаты численных экспериментов по расчету и оптимизации модели приводятся в разделе 5. В разделе 6 даны заключительные замечания.

2. Описание модели и постановка задачи

Рассмотрим модель $M/M/1/\infty/PQIS/RQ$. В систему извне поступает простейший поток идентичных (по размеру) p -заявок, т.е. после обслуживания каждой заявки уровень запасов на складе уменьшается на единицу. Кроме того, запасы системы независимо друг от друга становятся непригодными после случайного времени, которые имеют показательные ф.р. с конечным средним значением. Считается, что запас, который уже находится на этапе отпуска по заявке, не может портиться.

Если в момент поступления p -заявки уровень запасов положительный, то она с вероятностью единица принимается на обслуживание при условии, что в этот момент сервер является свободным; иначе p -заявка становится в очередь бесконечной длины. Заявки присоединяются к очереди согласно схеме Бернулли даже тогда, когда уровень запасов равен нулю, т.е. если в момент поступления очередной p -заявки в системе отсутствуют запасы. В этом случае p -заявка либо с определенной (положительной) вероятностью становится в очередь, либо с дополнительной вероятностью уходит в орбиту бесконечного размера для “обдумывания” и повторения своего запроса через некоторое время.

Считается, что заявки (первичные и повторные) в очереди становятся нетерпеливыми при отсутствии запасов системы. Интервалы времени между уходами заявок из очереди имеют показательную ф.р. Уходящая из очереди нетерпеливая заявка согласно схеме Бернулли либо с определенной (положительной) вероятностью окончательно уходит из системы, либо с дополнительной вероятностью уходит в орбиту, при этом имеется возможность многократного ухода в орбиту.

Заявки в орбите не имеют информации как об уровне запасов системы, так и о состоянии очереди. Если в орбите имеются заявки, то интервалы времени между их поступлениями имеют показательную ф.р. Повторные заявки (r -заявки) являются настойчивыми, т.е. если в момент поступления r -заявки уровень запасов системы равен нулю, то она опять мгновенно возвращается в орбиту с вероятностью единица. После завершения обслуживания заявки возможны два решения: (1) заявка отказывается получить запас и уходит из системы; (2) заявка получает запас и уходит из системы.

Время обслуживания как первичных, так и повторных заявок является одинаковым, при этом длительность обслуживания зависит от того, получила ли заявка запас или нет; считается, что в обоих случаях это время имеет показательную ф.р., но с различными средними значениями.

В системе используется (s, S) -политика пополнения запасов, т.е. предполагается, что когда уровень запасов опускается до некоторой пороговой величины s , то отправляется заказ на поставку запасов объема $S - s$. При этом для предотвращения случаев многократных заказов необходимо выполнение соотношения $s < S/2$. Считается, что заказы выполняются с некоторыми задержками, при этом эти задержки являются независимыми друг от друга положительными случайными величинами, которые имеют общую показательную ф.р. с конечным средним значением.

Задача состоит в нахождении совместного распределения уровня запасов системы, суммарного числа заявок в системе и числа r -заявок в орбите. Кроме того, требуется определить усредненные характеристики изучаемой PQIS и минимизировать суммарные штрафы, связанные с ее функционированием.

3. Точный метод расчета характеристик системы

Введем обозначения:

S — максимальный объем склада системы;

s — точка заказа запасов, $s < S/2$;

γ^{-1} — средняя длительность жизни единицы запаса;
 λ^{-1} — среднее время между поступлениями p -заявок;
 τ^{-1} — среднее время пребывания заявок в системе при отсутствии запасов;
 ϕ_1 — вероятность того, что поступившая p -заявка становится в очередь, если в этот момент в системе отсутствуют запасы;
 ϕ_2 — вероятность того, что поступившая p -заявка уходит в орбиту, если в этот момент в системе отсутствуют запасы;
 ψ_1 — вероятность того, что покидающая очередь из-за нетерпеливости заявка уходит из системы;
 ψ_2 — вероятность того, что покидающая очередь из-за нетерпеливости заявка уходит в орбиту;
 σ_1 — вероятность того, что заявка не получает запас и уходит из системы;
 σ_2 — вероятность того, что заявка получает запас и уходит из системы;
 μ_1^{-1} — среднее время обслуживания заявок, которые не получают запас;
 μ_2^{-1} — среднее время обслуживания заявок, которые получают запас;
 ν^{-1} — среднее время задержки выполнения заказа;
 η^{-1} — среднее время пребывания заявки в орбите.

Работа системы описывается трехмерной ЦМ с состояниями вида (m, n, k) , где m — уровень запасов на складе, n — число заявок в очереди и k — число r -заявок в орбите. Пространство состояний (ПС) этой цепи определяется так:

$$(3.1) \quad E = \bigcup_{k=0}^{\infty} E_k, \quad E_k \cap E_{k'} = \emptyset, \quad k \neq k',$$

где

$$E_k = \{(m, n, k) : m = 0, 1, \dots, S, n = 0, 1, \dots\}, \quad k = 0, 1, 2, \dots$$

Переходы между состояниями внутри класса ПС (3.1) происходят при поступлении p -заявок и пополнении запасов системы, а также после завершения обслуживания заявок, времени жизни запасов и ухода заявок из очереди из-за их нетерпеливости. Переходы между состояниями из различных классов E_k и $E_{k'}$, $k \neq k'$ ПС (3.1) связаны с уходами заявок в орбиту и поступлением r -заявок с орбиты.

Интенсивность перехода из состояния (m_1, n_1, k_1) в состояние (m_2, n_2, k_2) обозначим через $q((m_1, n_1, k_1), (m_2, n_2, k_2))$. Совокупность этих величин определяет ПМ данной цепи. Представим предложенный авторами псевдо-код алгоритма генерации элементов ПМ изучаемой цепи Маркова.

```

function QELEM( $m_1, n_1, k_1, m_2, n_2, k_2$ )
  define  $q := 0$ 
  if  $k_2 = k_1$  and  $m_1 > 0$  then
    if  $m_2 = m_1$  and  $n_2 = n_1 + 1$  then  $q := \lambda$ 
    else if  $m_2 = m_1$  and  $n_2 = n_1 - 1$  then  $q := \mu_1 \sigma_1$ 
    else if  $m_2 = m_1 - 1$  and  $n_2 = n_1 - 1$  then  $q := \mu_2 \sigma_2$ 

```

```

else if  $m_2 = m_1 - 1$  and  $n_2 = n_1 = 0$  then  $q := m_1\gamma$ 
else if  $m_2 = m_1 - 1$  and  $n_2 = n_1 > 0$  then  $q := (m_1 - 1)\gamma$ 
else if  $m_1 \leq s$  and  $m_2 = m_1 + S - s$  and  $n_2 = n_1$  then  $q := \nu$ 
end if
else if  $k_2 = k_1$  and  $m_1 = 0$  then
  if  $m_2 = 0$  and  $n_2 = n_1 + 1$  then  $q := \lambda\phi_1$ 
  else if  $m_2 = 0$  and  $n_2 = n_1 - 1$  then  $q := n_1\tau\psi_1$ 
  else if  $m_2 = S - s$  and  $n_2 = n_1$  then  $q := \nu$ 
  end if
else if  $k_2 \neq k_1$  then
  if  $n_2 = n_1$  and  $k_2 = k_1 + 1$  and  $m_1 = m_2 = 0$  then  $q := \lambda\phi_2$ 
  else if  $n_2 = n_1 - 1$  and  $k_2 = k_1 + 1$  and  $m_1 = m_2 = 0$  then  $q := n_1\tau\psi_2$ 
  else if  $n_2 = n_1 + 1$  and  $k_2 = k_1 - 1$  and  $m_1 = m_2 > 0$  then  $q := k_1\eta$ 
  end if
end if
return  $q$ 
end function

```

Стационарную вероятность состояния (m, n, k) обозначим через $p(m, n, k)$.
Условие существования стационарного режима определяется далее.

Основными характеристиками системы являются следующие величины: средний уровень запасов на складе (S_{av}), средняя интенсивность порчи запасов системы (Γ_{av}), средняя интенсивность заказов (RR); средняя длина очереди заявок в системе (L_s); среднее число r -заявок в орбите L_o , интенсивности потери заявок из очереди из-за их нетерпеливости (RL_q). Они находятся через стационарные вероятности состояний. Так, средний уровень запасов, среднее число заявок в системе и в орбите определяются как математические ожидания соответствующих с.в.:

$$(3.2) \quad S_{av} = \sum_{(m,n,k) \in E} mp(m, n, k);$$

$$(3.3) \quad L_s = \sum_{(m,n,k) \in E} np(m, n, k);$$

$$(3.4) \quad L_o = \sum_{(m,n,k) \in E} kp(m, n, k).$$

Средняя интенсивность порчи запасов вычисляется как

$$(3.5) \quad \Gamma_{av} = \gamma \left(\sum_{m=1}^S m \sum_{(m,0,k) \in E} p(m, 0, k) + \sum_{m=2}^S (m-1) \sum_{(m,n,k) \in E} p(m, n, k) I(n > 0) \right),$$

где $I(A)$ – индикаторная функция события A .

Заказы запасов осуществляются тогда, когда их уровень опускается до значения s независимо от состояния очереди и числа заявок в орбите, т.е.:

$$(3.6) \quad RR = \gamma(s+1) \sum_{(s+1,0,k) \in E} p(s+1,0,k) + \\ + (\mu_2\sigma_2 + s\gamma) \sum_{(s+1,n,k) \in E} p(s+1,n,k)I(n > 0).$$

Интенсивность ухода заявок из очереди из-за их нетерпеливости определяется так:

$$(3.7) \quad RL_q = \tau\psi_1 \sum_{(0,n,k) \in E} np(0,n,k).$$

Нахождение характеристик (3.2)–(3.7) требует определения вероятностей состояний, которые удовлетворяют системе уравнений равновесия (СУР). Она составляется на основе ПМ и представляет собой бесконечную систему линейных уравнений.

Вычисление вероятностей состояний с помощью метода многомерных производящих функций связано с известными методологическими и вычислительными трудностями. Применение популярных матрично-геометрического метода [10] или метода спектрального расширения [19] потребует введения некоторых условий относительно элементов производящей матрицы, изучаемой трехмерной ЦМ с целью ее унификации. Часто эти допущения не соответствуют реальным условиям работы системы. Так, в данной модели для их применения необходимо принимать нереальные допущения о постоянстве интенсивностей поступления r -заявок с орбиты, а также интенсивностей ухода p -заявок из очереди из-за их нетерпеливости. Кроме того, указанные методы часто страдают вычислительной неустойчивостью из-за плохой обусловленности используемых матриц большой размерности. Поэтому здесь для решения указанной проблемы используется альтернативный (приближенный) метод.

4. Приближенный метод расчета характеристик системы

Этот метод корректно можно применить для систем, в которых интенсивности переходов между состояниями, входящими в различные классы E_k и $E_{k'}$, $k \neq k'$, оказываются малыми по сравнению с интенсивностями переходов между состояниями внутри классов. Это допущение, в частности, выполняется в системах, в которых интенсивность p -заявок существенно больше интенсивности r -заявок, т.е. когда выполняется условие $\eta \ll \lambda$.

На основе расщепления (3.1) строится функция укрупнения $U_1(m,n,k) = \langle k \rangle$, где $\langle k \rangle$ означает укрупненное состояние, которое включает в себя все состояния из класса E_k . Множество укрупненных состояний обозначается через $\Omega_1 = \{\langle k \rangle : k = 0, 1, \dots\}$. Тогда имеем:

$$(4.1) \quad p(m,n,k) \approx \rho^k(m,n)\pi_1(\langle k \rangle),$$

где $\rho^k(m, n)$ — вероятность состояния (m, n) внутри расщепленной модели с ПС E_k и $\pi_1(\langle k \rangle)$ — вероятность укрупненного состояния $\langle k \rangle$, $\langle k \rangle \in \Omega_1$.

Из (4.1) заключаем, что для решения поставленной задачи потребуется нахождение вероятностей состояний бесконечного числа двумерных и одной одномерной цепи Маркова, при этом все эти цепи имеют бесконечное пространство состояний.

Для нахождения вероятностей состояний $\rho^k(m, n)$ предлагается к указанным двумерным цепям Маркова с ПС E_k , $k = 0, 1, \dots$, еще раз применить алгоритм фазового укрупнения состояний. Расщепленные модели с ПС E_k , $k = 0, 1, \dots$, являются идентичными двумерными ЦМ, поэтому далее фиксируется индекс k и рассматривается следующее расщепление:

$$(4.2) \quad E = \bigcup_{m=0}^S E_k^m, \quad E_k^{m_1} \cap E_k^{m_2} = \emptyset, \quad m_1 \neq m_2,$$

где

$$E_k^m = \{(m, n, k) \in E_k : n = 0, 1, \dots\}, \quad m = 0, \dots, S.$$

На основе расщепления (4.2) строится функция укрупнения $U_2(m, n, k) = \langle m \rangle$, где $\langle m \rangle$ означает укрупненное состояние, которое включает в себя все состояния из класса E_k^m . Обозначим $\Omega_2 = \{\langle m \rangle : m = 0, 1, \dots, S\}$. Тогда имеем:

$$(4.3) \quad \rho^k(m, n) \approx \rho_m^k(n) \pi_2^k(\langle m \rangle),$$

где $\rho_m^k(n)$ — вероятность состояния (m, n) внутри расщепленной модели с пространством состояний E_k^m и $\pi_2^k(\langle m \rangle)$ — вероятность укрупненного состояния $\langle m \rangle$, $\langle m \rangle \in \Omega_2$.

Интенсивности переходов между состояниями расщепленной модели с ПС E_k^m не зависят от индекса k , поэтому далее в обозначениях вероятностей состояний расщепленных моделей $\rho_m^k(n)$ и укрупненных состояний $\pi_2^k(\langle m \rangle)$ верхний индекс k опускается. Из алгоритма построения ПМ исходной модели (см. псевдо-код) заключаем, что вероятности состояний внутри всех расщепленных моделей с ПС E_k^m , $m = 1, \dots, S$, совпадают с вероятностями состояний модели $M/M/1/\infty$ с загрузкой $a = \lambda/(\mu_1 \sigma_1)$ Эрланг (Эрл), т.е.

$$(4.4) \quad \rho_m(n) = a^n (1 - a), \quad m = 1, \dots, S.$$

Замечание 1. Здесь предполагается, что выполняется условие эргодичности модели, т.е. $a < 1$.

Замечание 2. Поскольку величины $\rho_m(n)$, $m = 1, \dots, S$, не зависят от индекса m , то далее этот индекс опускается.

Также из алгоритма построения ПМ исходной модели (см. псевдо-код) заключаем, что вероятности состояний внутри расщепленной модели с ПС E_k^0 совпадают с вероятностями состояний модели $M/M/\infty$ с загрузкой $b = \lambda \phi_1 / \tau \psi_1$ Эрл, т.е.

$$(4.5) \quad \rho_0(n) = \theta_n(b), \quad n = 0, 1, \dots$$

Здесь и в дальнейшем

$$\theta_n(b) = \frac{b^n}{n!} e^{-b}, \quad , n = 0, 1, \dots$$

Интенсивности переходов между укрупненными состояниями $\langle m_1 \rangle, \langle m_2 \rangle \in \Omega_2$ обозначаются через $q_1(\langle m_1 \rangle, \langle m_2 \rangle)$. Эти величины вычисляются так:

$$(4.6) \quad q_1(\langle m_1 \rangle, \langle m_2 \rangle) = \sum_{(m_1, n_1, k) \in E_k^{m_1}} q((m_1, n_1, k), (m_2, n_2, k)) \rho_{m_1}(n_1).$$

С учетом приведенного псевдо-кода, выражений (4.4) и (4.5) из (4.6) после определенных преобразований находим, что $q_1(\langle m_1 \rangle, \langle m_2 \rangle)$ определяются из соотношений:

$$q_1(\langle m_1 \rangle, \langle m_2 \rangle) = \begin{cases} \Lambda_1(m_1), & \text{если } m_2 = m_1 - 1, \\ \nu, & \text{если } m_1 \leq s, m_2 = m_1 + S - s, \\ 0, & \text{в остальных случаях,} \end{cases}$$

где

$$\Lambda_1(m_1) = m_1 \gamma (1 - a) + a(\mu_2 \sigma_2 + (m_1 - 1) \gamma), \quad m_1 = 1, 2, \dots, S.$$

Отсюда имеем (см. [13]):

$$(4.7) \quad \pi_2(\langle m \rangle) = \begin{cases} \alpha_m \pi_2(\langle s+1 \rangle), & \text{если } 0 \leq m \leq s, \\ \beta_m \pi_2(\langle s+1 \rangle), & \text{если } s+1 \leq m \leq S-s, \\ \chi_m \pi_2(\langle s+1 \rangle), & \text{если } S-s+1 \leq m \leq S, \end{cases}$$

где

$$\begin{aligned} \alpha_m &= \prod_{i=m+1}^{s+1} \frac{\Lambda_1(i)}{\nu + \Lambda_1(i-1)}, \quad \beta_m = \frac{\Lambda_1(s+1)}{\Lambda_1(m)}, \\ \chi_m &= \frac{\nu}{\Lambda_1(m)} \sum_{i=m-S+s}^s \alpha_i, \quad \Lambda_1(0) = 0; \\ \pi_2(\langle s+1 \rangle) &= \left(\sum_{m=0}^s \alpha_m + \sum_{m=s+1}^{S-s} \beta_m + \sum_{m=S-s+1}^S \chi_m \right)^{-1}. \end{aligned}$$

Интенсивности переходов между укрупненными состояниями $\langle k_1 \rangle, \langle k_2 \rangle \in \Omega_1$, обозначаются через $q_2(\langle k_1 \rangle, \langle k_2 \rangle)$. Они вычисляются следующим образом:

$$(4.8) \quad q_2(\langle k_1 \rangle, \langle k_2 \rangle) = \sum_{(m, n, k) \in E_k} q((m, n, k), (m', n', k')) \rho^k(m, n).$$

С учетом псевдо-кода, выражений (4.3)–(4.5) и (4.7) из (4.8) после определенных преобразований находим, что

$$(4.9) \quad q_2(\langle k_1 \rangle, \langle k_2 \rangle) = \begin{cases} \Lambda_2, & \text{если } k_2 = k_1 + 1, \\ k_1 M_2, & \text{если } k_2 = k_1 - 1, \\ 0, & \text{в остальных случаях,} \end{cases}$$

где

$$\Lambda_2 = \lambda \left(\phi_2 + \phi_1 \frac{\psi_2}{\psi_1} \right), \quad M_2 = \eta(1 - \pi_2(\langle 0 \rangle)).$$

Следовательно, из (4.9) заключаем, что вероятности укрупненных состояний $\pi_1(\langle k \rangle)$, $\langle k \rangle \in \Omega_1$, совпадают с вероятностями состояний модели $M/M/\infty$ с загрузкой $c = \Lambda_2/M_2$ Эрл, т.е

$$(4.10) \quad \pi_1(\langle k \rangle) = \theta_k(c), \quad k = 0, 1, \dots$$

Окончательно, с учетом (4.1), (4.4), (4.5), (4.7) и (4.10) находим, что искомые вероятности состояний изучаемой трехмерной ЦМ определяются так:

$$(4.11) \quad p(m, n, k) \approx \rho_m(n) \pi_2(\langle m \rangle) \pi_1(\langle k \rangle).$$

Далее с учетом (4.11) после определенных преобразований из (3.2)–(3.7) получим формулы для приближенного расчета характеристик изучаемой системы:

$$(4.12) \quad S_{av} \approx \sum_{m=1}^S m \pi_2(\langle m \rangle);$$

$$(4.13) \quad \Gamma_{av} \approx \gamma \sum_{m=1}^S (m - a) \pi_2(\langle m \rangle);$$

$$(4.14) \quad RR \approx \pi_2(\langle s + 1 \rangle) [(s + 1) \gamma (1 - a) + (s \gamma + \mu_2 \sigma_2) a];$$

$$(4.15) \quad L_s \approx \pi_2(\langle 0 \rangle) b + (1 - \pi_2(\langle 0 \rangle)) \frac{a^2}{1 - a};$$

$$(4.16) \quad L_o \approx c;$$

$$(4.17) \quad RL_q \approx \lambda \phi_1 \pi_2(\langle 0 \rangle).$$

5. Численные результаты

Сначала рассмотрим задачу минимизации суммарных штрафов (Total Cost, TC), связанных с функционированием системы. Эти штрафы включают расходы, связанные с заказом, хранением и порчей запасов, а также штрафы из-за ожидания заявок в очереди и их потери. Иными словами, указанные штрафы определяются так:

$$(5.1) \quad TC(s) = (K + c_r(S - s))RR + c_h S_{av} + c_p \Gamma_{av} + c_l RL_q + c_w(L_s + L_o),$$

где K — фиксированная цена одного заказа, c_r — цена единицы объема запаса; c_h — цена хранения единицы объема запасов за единицу времени; c_p — цена порчи единицы запаса; c_l — штрафы за потери одной заявки; c_w — цена за единицу времени пребывания в системе (или в орбите) одной заявки.

Замечание 3. Считаем, что все параметры системы, кроме критического уровня запасов (т.е. s), являются фиксированными величинами. Для компактности записи в (5.1) лишь в левой части (в выражении функционала TC) в скобках указан оптимизируемый параметр s . Очевидно, что все функции в правой части (5.1), которые обозначают характеристики системы, также зависят от этого параметра.

Задача оптимизации записывается так:

$$(5.2) \quad s^* = \operatorname{argmin} \{TC(s) : 0 \leq s \leq s'\},$$

где

$$s' = \begin{cases} \left\lceil \frac{S}{2} \right\rceil, & \text{если } S - \text{нечетное число,} \\ \frac{S}{2} - 1, & \text{если } S - \text{четное число,} \end{cases} \quad [x] - \text{целая часть } x.$$

При любых значениях входных параметров задача (5.2) имеет решение, так как множество возможных решений является дискретным и конечным. Вместе с тем из-за многочисленности параметров, от которых зависят функции, входящие в правую часть (5.1), не удастся делать общее заключение о выпуклости (или вогнутости) функционала TC , поэтому для решения (5.2) далее использован метод полного перебора.

Предположим, что система может заказывать запасы у нескольких поставщиков, при этом времена выполнения заказов и стоимости за единицу заказов у различных поставщиков отличаются друг от друга. При этом считается, что с ростом скорости выполнения заказа растет и стоимость за единицу заказов.

Постоянные значения исходных данных системы и коэффициентов в выражении функционала (5.1) выбирались так:

$$\begin{aligned} S &= 15, \quad \lambda = 20, \quad \mu_1 = 60, \quad \mu_2 = 10, \quad \eta = 5, \quad \gamma = 2, \\ \tau &= 1,5, \quad \sigma_1 = 0,4, \quad \phi_1 = 0,3, \quad \psi_1 = 0,8; \\ c_h &= 0,5, \quad c_p = 0,2, \quad c_l = 2,0, \quad c_w = 1. \end{aligned}$$

Результаты решения задачи (5.2) показаны в таблице, где TC^* — минимальное значение функционала (5.2). Из таблицы видно, что для данной системы целесообразно пользоваться услугами второго поставщика. Анализ результатов оптимизационной задачи (5.2) показывает, что с ростом интенсивности пополнения запасов значение критического уровня запасов не увеличивается. Иными словами, использование услуг “быстрых” поставщиков позволяет поддерживать достаточно низкий критический уровень запасов.

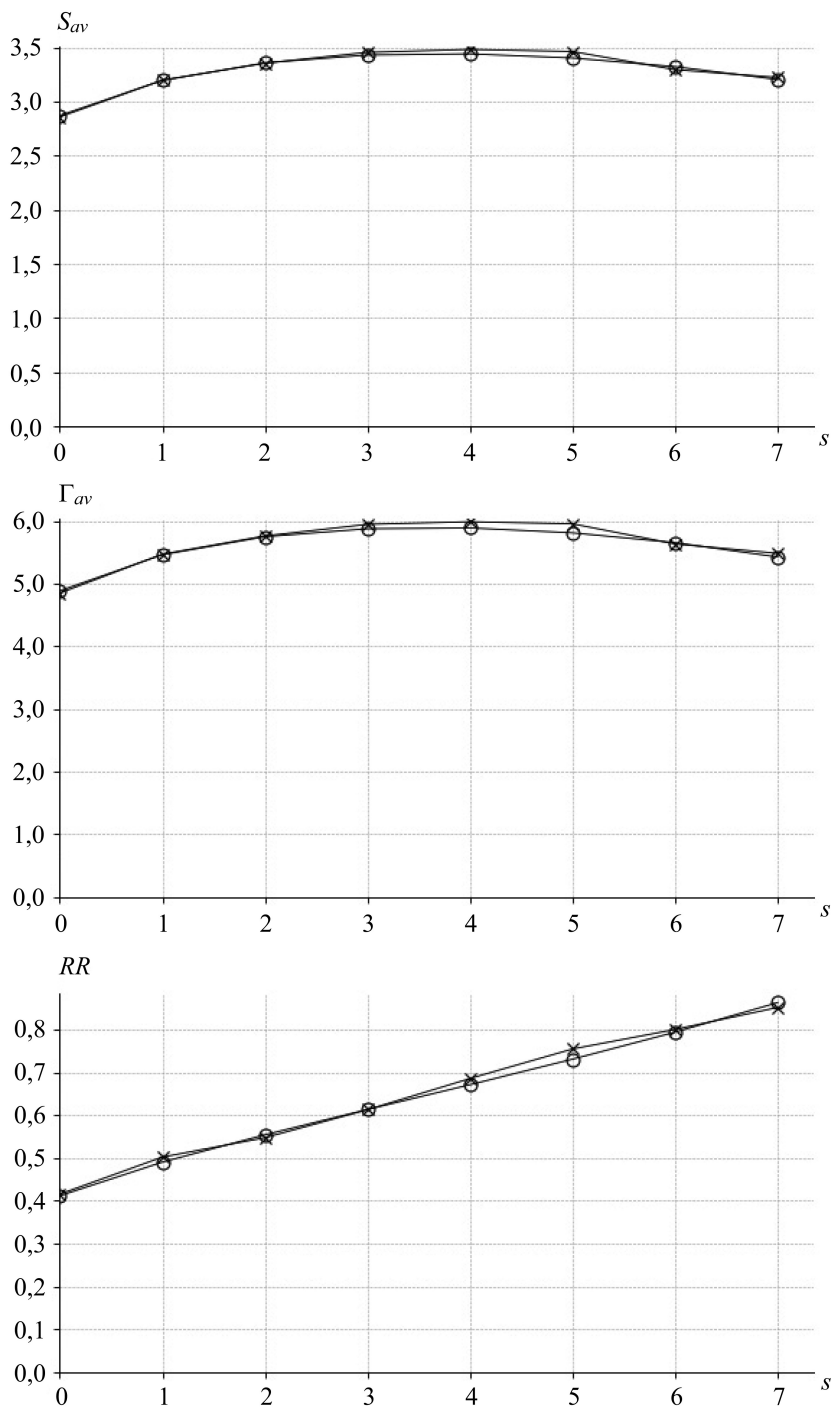


Рис. 1. Сравнение характеристик первой группы при различных методах их вычисления.

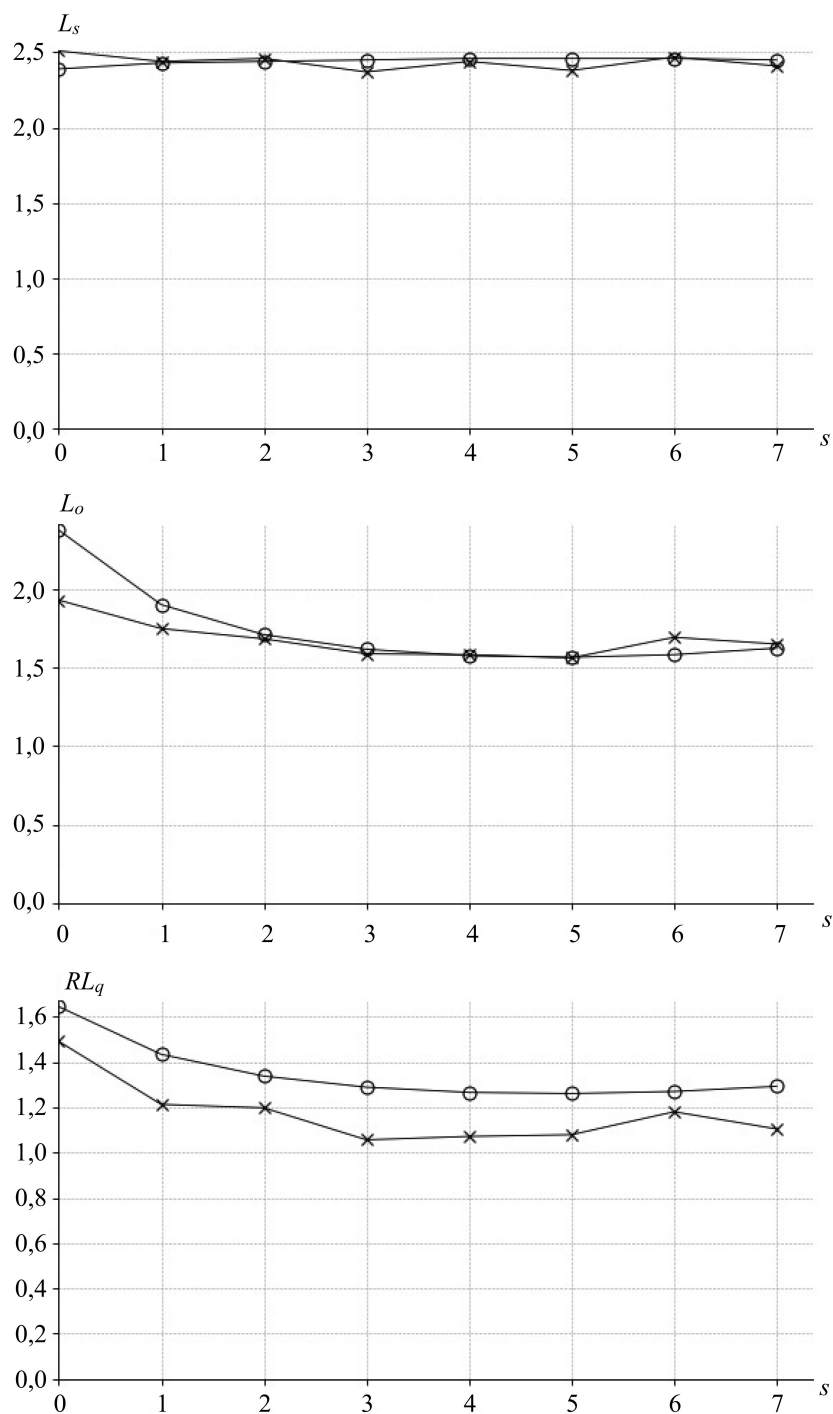


Рис. 2. Сравнение характеристик второй группы при различных методах их вычисления.

Результаты решения задачи (5.2)

Номер поставщика	ν	(k, c_r)	s^*	TC^*
1	0,5	(2; 0,1)	6	22,92385
2	1	(3; 0,1)	5	20,26252
3	1,5	(4,5; 0,15)	1	21,09278
4	2	(5; 0,2)	1	21,83515
5	2,5	(5,5; 0,2)	0	21,95194
6	3	(7; 0,35)	0	25,28343

Теперь рассмотрим вопросы точности формул (4.12)–(4.17), при этом их точность оценивается с помощью имитационного моделирования (при разработке программы имитационного моделирования использованы алгоритмы, предложенные в [20]).

Характеристики системы разделены на две группы: первая группа оценивает работу части “системы управления запасами” (формулы (4.12)–(4.14)), а вторая группа — “системы обслуживания заявок” (формулы (4.15)–(4.17)). На рис. 1 и 2 приводятся результаты сравнения значений характеристик системы при использовании различных подходов для указанных выше значений параметров системы.

На рис. 1 и 2 символы о и × означают результаты, полученные с помощью разработанных приближенных формул (4.12)–(4.17) и имитационного моделирования соответственно. Отметим, что эти формулы имеют очень высокую точность для первой группы (см. рис. 1), так как здесь результаты обоих подходов почти совпадают. Иными словами, максимальная относительная погрешность для характеристик (4.12)–(4.14) при использовании предложенных приближенных формул и имитационного моделирования составляет 0,007, 0,012 и 0,023 соответственно. Для второй группы предложенные приближенные формулы также имеют достаточно высокую точность, при этом небольшая погрешность наблюдается для характеристики (4.16) (см. рис. 2). Для этой группы максимальная относительная погрешность для характеристик (4.15)–(4.17) при использовании предложенных приближенных формул и имитационного моделирования составляет 0,048, 0,193 и 0,182 соответственно.

6. Заключение

Предложена марковская модель системы обслуживания–запасания с портящимися запасами и положительным временем обслуживания заявок с помощью одного сервера. При отсутствии запасов первичные заявки согласно схеме Бернулли либо уходят в орбиту бесконечного размера, либо присоединяются к очереди бесконечной длины. Заявки являются нетерпеливыми лишь тогда, когда уровень запасов системы равен нулю, при этом интенсивность ухода заявок из очереди, а также интенсивность поступления повторных заявок с орбиты зависят от числа заявок в очереди и в орбите соответственно. В системе принята (s, S) -политика пополнения запасов, при этом время выполнения заказа является положительной случайной величиной.

Предложена трехмерная цепь Маркова, которая является математической моделью изучаемой системы, и разработан алгоритм построения ее производящей матрицы. Разработан приближенный метод вычисления вероятностей состояний этой цепи и найдены явные формулы вычисления ее характеристик. Решена задача минимизации суммарных штрафов системы, связанных с заказом, хранением и порчей запасов, а также с ожиданием и потерей заявок.

В качестве направлений дальнейших исследований следует указать изучение подобных моделей при наличии МАР потока заявок, произвольной ф.р. времени их обслуживания, а также при использовании других политик пополнения запасов, в частности, рандомизированной политики, политики переменного объема заказов и т.д. Большой научный и практический интерес представляют также вопросы решения задач минимизации суммарных штрафов при наличии определенных ограничений на характеристики системы, а также изучение инвариантности оптимальных решений относительно изменения нагрузочных параметров системы.

СПИСОК ЛИТЕРАТУРЫ

1. *Artalejo J.R., Krishnamoorthy A., Lopez-Herrero M.J.* Numerical Analysis of (s, S) Inventory System with Repeated Attempts // *Ann. Oper. Res.* 2006. V. 141. P. 67–83.
2. *Ushakumari P.V.* On (s, S) Inventory System with Random Lead Time and Repeated Demands // *J. Appl. Math. Stoch. Anal.* 2006. Article ID 81508.
3. *Lopez-Herrero M.J.* Waiting Time and Other First-Passage Time Measures in an (s, S) Inventory System with Repeated Attempts and Finite Retrial Group // *Comput. Oper. Res.* 2010. V. 37. P. 1256–1261.
4. *Anbazhagan N., Wang J., Gomathi D.* Base Stock Policy with Retrial Demands // *Appl. Math. Model.* 2013. V. 37. P. 4464–4473.
5. *Krishnamoorthy A., Jose K.P.* Comparision of Inventory Systems with Service, Positive Lead-Time, Loss, and Retrial of Customers // *J. Appl. Math. Stoch. Anal.* (Hindawi Publishing Corporation). V. 2007. Article ID 37848.
6. *Nair A.N., Jacob M.J.* (s, S) Inventory System with Positive Service Time and Retrial of Demands: an Approach Through Multi-Server Queues // *ISRN Oper. Res.* (Hindawi Publishing Corporation). V. 2014. Article ID 596031.
7. *Yadavalli V.S.S., Anbazhagan N., Jeganathan K.* A Retrial Inventory System with Impatient Customers // *Appl. Math. Inform. Sci.* 2015. V. 9. No. 2. P. 637–650.
8. *Amirthakodi M., Sivakumar B.* An Inventory System with Service Facility and Finite Orbit for Feedback Customers // *OPSEARCH.* 2015. V. 52. No. 2. P. 225–255.
9. *Manikandan R., Nair S.S.* M/M/1/1 Queuing-Inventory System with Retrial of Unsatisfied Customers // *Comm. Appl. Anal.* 2017. V. 21. No. 2. P. 217–236.
10. *Neuts M.F.* Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach. Baltimore: John Hopkins Univ. Press, 1981.
11. *Manuel P., Sivakumar B., Arivarignan G.* A Perishable Inventory System with Service Facilities and Retrial Customers // *Comput. Ind. Eng.* 2008. V. 54. P. 484–501.
12. *Sivakumar B.* An Inventory System with Retrial Demands and Multiple Server Vacation // *Quality Technol. Quantit. Management.* 2011. V. 8. No. 2. P. 125–146.

13. *Melikov A.Z., Ponomarenko L.A., Shahmaliyev M.O.* Models of Perishable Queuing-Inventory Systems with Repeated Customers // J. Autom. Inform. Sci. 2016. V. 48. No. 6. P. 22–38.
14. *Новицкая Е.В., Терпугов А.Ф.* Оптимизация розничной продажи скоропортящейся продукции. Томск: Изд-во Том. ун-та, 2004.
15. *Лившиц К.И., Ульянова Е.С.* Диффузионная аппроксимация процесса производства и сбыта скоропортящейся продукции // Изв. вузов. Физика. 2015. Т. 58. № 11/2. С. 281–285.
16. *Krishnamoorthy A., Manikandan R., Lakshmy B.* Revisit to Queuing-Inventory System with Positive Service Time // Ann. Oper. Res. 2015. V. 233. P. 221–236.
17. *Melikov A.Z., Ponomarenko L.A., Shahmaliyev M.O.* Analysis of Perishable Queuing-Inventory Systems with Different Types of Requests // J. Autom. Inform. Sci. 2017. V. 49. No. 9. P. 42–60.
18. *Melikov A.Z., Ponomarenko L.A., Rustamov A.M.* Approximate Analysis of Queuing-Inventory System with Earlier and Delayed Vacations // Autom. Remote Control. 2017. V. 78. No. 11. P. 1991–2003.
Меликов А.З., Пономаренко Л.А., Рустамов А.М. Приближенный анализ системы обслуживания-запасания с ранним и запаздывающими отдыхами сервера // АиТ. 2017. № 11. С. 64–78.
19. *Mitrani I., Chakka R.* Spectral Expansion Solution for a Class of Markov Models: Application and Comparison with the Matrix-Geometric Method // Performance Evaluat. 1995. V. 23. P. 241–260.
20. *Gillespie D.T.* A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions // J. Comput. Phys. 1976. V. 22. P. 403–434.

Статья представлена к публикации членом редколлегии В.М. Вишневским.

Поступила в редакцию 09.04.2018

После доработки 23.07.2018

Принята к публикации 08.11.2018