

МАРКОВСКИЕ МОДЕЛИ СИСТЕМ ОБСЛУЖИВАНИЯ–ЗАПАСАНИЯ С ПЕРЕМЕННЫМ ОБЪЕМОМ ЗАКАЗОВ

Аннотация. Изучены марковские модели систем обслуживания–запасания с переменным объемом заказов. Рассмотрены два класса моделей: с мгновенным обслуживанием и с ненулевым временем обслуживания. В модели с ненулевым временем обслуживания предполагается, что нетерпеливые расходуемые заявки могут образовать очередь конечной или бесконечной длины. Разработаны точный и приближенный методы для определения характеристик изучаемых систем при использовании предложенной политики пополнения запасов.

Ключевые слова: системы обслуживания–запасания, политика пополнения запасов, переменный объем заказа.

ВВЕДЕНИЕ

В классических моделях управления запасами, как правило, не учитывается работа сервера, который осуществляет обслуживание расходуемых заявок (p -заявок), и следовательно, в них изучается распределение (стационарное или нестационарное) лишь уровня запасов системы. Достаточно подробно основы теории управления запасами изложены, например, в [1–4]. Вместе с тем для детального анализа систем управления запасами требуется изучение совместного распределения уровня запасов системы и количества p -заявок в ней, так как зачастую системы управления запасами представляют собой системы обслуживания–запасания (СОЗ).

В последние годы интенсивно исследуются модели СОЗ при различных допущениях относительно:

- а) схемы обслуживания p -заявок;
- б) принятой политики пополнения запасов (ППЗ);
- в) времени доставки заказываемых запасов.

Под показателем а) зачастую понимается совокупность правил, которые определяют не только длительность обслуживания p -заявок, но также их поведение при отсутствии запасов системы. По длительности обслуживания p -заявок различают системы с нулевым (мгновенным) и ненулевым (положительным) временем обслуживания. По поведению p -заявок при отсутствии запасов различают системы с живой очередью, когда p -заявки ожидают непосредственно в системе, и системы с повторными заявками, когда p -заявки поступают в орбит и оттуда через некоторое случайное время повторяют свою попытку получения запасов.

Относительно показателя б) отметим, что ППЗ должна определять, когда и сколько необходимо заказывать запасов для пополнения склада системы. В известных работах предполагается, что моменты (точки) заказа запасов находятся, лишь исходя из текущего уровня запасов системы, так как на практике в большинстве случаев невозможно точно спрогнозировать интенсивности потоков p -заявок (этим фактором объясняется и трудность практической реализации результатов решения задач оптимального управления запасами). Для простоты реализации, а также для удобства математического анализа в известных публикациях

зачастую предполагается, что объемы заказов постоянные. Так, например, при использовании широко применяемой (s, S) -политики пополнения запасов принимается, что каждый раз, когда уровень запасов опускается ниже определенной величины s , необходимо заказывать запасы фиксированного размера $S - s$, где S — максимальный размер склада системы. Существуют и более сложные политики такого рода со многими параметрами.

Согласно показателю v) различают системы с нулевым и ненулевым временем доставки заказываемых запасов: в первом случае сделанный заказ выполняется мгновенно, а во втором — с некоторой задержкой. При этом в моделях с ненулевым временем доставки зачастую предполагают, что такое время является случайной величиной с известной ф.р.

Существует огромное количество публикаций, в которых изучаются модели СОЗ при различных допущениях относительно показателей а)–в). Достаточно подробный обзор известных работ по моделям СОЗ приведен в [5]. Однако почти не исследованы модели СОЗ, в которых объемы заказов являются переменными величинами, зависящими от текущего уровня запасов системы. (Авторы настоящей статьи не нашли работ, в которых изучаются подобные модели СОЗ, и с благодарностью примут ссылки на них). Лишь в [6] исследована аналогичная модель управления запасами, но в ней не учитывается наличие станции обслуживания, т.е. она не является моделью СОЗ.

Отметим, что, изучение моделей СОЗ с переменным объемом заказов представляет определенный теоретический интерес, так как они в ряде случаев обобщают известные ППЗ, а на практике использование таких ППЗ расширяет область определения надлежащей политики, которая в свою очередь позволяет увеличивать эффективность работы системы. Исходя из этого, в настоящей статье предложен новый класс ППЗ с переменным объемом заказов, а также точный и приближенный методы расчета характеристик изучаемых моделей СОЗ.

ОПИСАНИЕ МОДЕЛЕЙ СОЗ С ПЕРЕМЕННЫМ ОБЪЕМОМ ЗАКАЗОВ

Входящий поток p -заявок является пуассоновским с интенсивностью λ . Для простоты изложения предполагается, что каждая p -заявка требует ресурс единичного размера. Максимальный размер склада равен S , $0 < S < \infty$.

Далее рассмотрим два типа моделей СОЗ: (i) системы с мгновенным обслуживанием p -заявок; (ii) системы с положительным временем обслуживания p -заявок. В обоих типах моделей ППЗ одинаковы и определяются следующим образом.

Запасы по p -заявкам отпускаются до тех пор, пока склад системы не опустошится. Вводится некоторая (пороговая) величина s , $s < S$, и если уровень запасов на складе превысит ее, то система не делает заказов для пополнения запасами; а когда текущий уровень запасов системы m становится меньшим или равным величине s , на склад верхнего уровня отправляется заказ на поставку запасов объема $d_s(m)$, при этом очевидно, что выполняется соотношение $0 < d_s(m) \leq S - m$. Эту политику назовем политикой переменного объема заказа (Variable Size of Order, VSO).

Отметим, что известная политика пополнения запасов двух уровней (s, S) является частным случаем предложенной VSO-политики. Действительно, если предположить, что $s < S/2$ и $d_s(m)$ определяются по приведенной далее формуле, то получим упомянутую политику двух уровней:

$$d_s(m) = \begin{cases} S - s, & \text{если } m \leq s, \\ 0, & \text{если } m > s. \end{cases} \quad (1)$$

Более того, если в формуле (1) положим $s=0$, получим $(0, S)$ -политику пополнения запасов, т.е. СОЗ делает заказ лишь тогда, когда ее склад полностью опустошен и при этом объем заказа равен S .

Для определенности изложения в настоящей работе предполагается, что $s < S/2$ и величина заказываемого запаса $d_s(m)$ определяется так:

$$d_s(m) = \begin{cases} S - m, & \text{если } m \leq s, \\ 0, & \text{если } m > s. \end{cases} \quad (2)$$

Соотношение (2) означает, что, когда уровень запасов опускается до величины m , делается заказ такого объема, чтобы полностью заполнить склад системы, т.е. в отличие от политики двух уровней в данном случае предполагается, что объем заказа является переменной величиной и зависит от текущего уровня запасов системы.

Сделанный заказ выполняется с некоторой задержкой, т.е. указанное время имеет показательную ф.р. с параметром $\nu(m)$, который в общем случае зависит от текущего уровня m , $m=0, 1, \dots, s$, запасов на складе системы.

В моделях типа (i) невозможно образование очереди и для простоты изложения предполагается, что p -заявки, поступившие в моменты отсутствия запасов системы, окончательно теряются и не влияют на дальнейшее поведение системы. Однако существуют и другие схемы поведения потерянных p -заявок, например, они могут образовать источник повторных заявок.

В моделях типа (ii) предполагается, что в системе имеется один канал для обслуживания p -заявок, а время их обслуживания является случайной величиной, имеющей показательную ф.р. с конечным параметром μ , $0 < \mu < \infty$. После завершения обслуживания p -заявки уровень ресурсов на складе системы уменьшается на единицу. Рассматриваются модели с ограниченным и неограниченным буфером для ожидания p -заявок в очереди. В случае модели с неограниченным буфером любая p -заявка принимается в буфер, однако в модели с ограниченным буфером максимальная длина очереди p -заявок может равняться N , $N < \infty$. Если p -заявка поступила в момент, когда в очереди уже имеются N таких заявок, то независимо от уровня ресурсов на складе системы она теряется с вероятностью 1.

Для более адекватного описания поведения расходуемых заявок в очереди в данном случае предполагается, что такие заявки являются нетерпеливыми, т.е. если уровень запасов системы равен m , то время ожидания p -заявок в очереди ограничено некоторой случайной величиной, имеющей показательную ф.р. со средним $\tau^{-1}(m)$. Последнее означает, что p -заявка покидает очередь (теряется), если до окончания допустимого времени ожидания не освобождается канал обслуживания.

Задача исследования моделей типа (i) состоит в определении распределения уровня запасов системы, а для моделей типа (ii) требуется найти совместное распределение уровня запасов системы и длины очереди p -заявок. Решение этих задач позволяет вычислить характеристики изучаемых СОЗ: S_{av} — средний уровень ресурсов на складе, P_p — вероятность потери p -заявок и L_{av} — среднюю длину очереди p -заявок (последнее относится лишь к моделям типа (ii)).

МЕТОДЫ РАСЧЕТА ХАРАКТЕРИСТИК МОДЕЛЕЙ СОЗ

Вначале рассмотрим модель СОЗ типа (i). Математической моделью данной СОЗ является одномерная цепь Маркова (ЦМ) с конечным множеством состояний, т.е. ее состояние задается параметром m , $m=0, 1, \dots, S$, обозначающим текущий уровень ресурсов на складе.

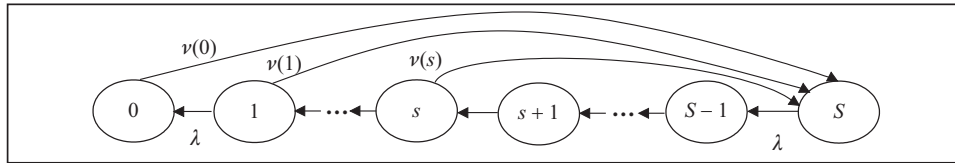


Рис. 1. Граф переходов между состояниями модели с мгновенным обслуживанием при VSO-политике пополнения запасов

Интенсивность перехода из состояния m в другое состояние m' обозначим $q(m, m')$. Тогда элементы производящей матрицы (Q -матрицы) данной ЦМ определяются так (рис. 1):

$$q(m, m') = \begin{cases} \nu(m), & \text{если } 0 \leq m \leq s, m' = S, \\ \lambda, & \text{если } m' = m - 1, \\ 0 & \text{в остальных случаях.} \end{cases} \quad (3)$$

Из соотношений (3) видно, что состояния этой конечной одномерной ЦМ являются сообщающимися, следовательно, в данной системе существует стационарный режим. Пусть $\pi(m)$ — стационарная вероятность состояния m , $m = 0, 1, \dots, S$. Эти вероятности находятся из системы уравнений равновесия (СУР), которая решается аналитически (здесь соответствующие преобразования опускаются):

$$\pi(m) = \begin{cases} \frac{1}{\lambda^m} \prod_{i=1}^m (\nu(i-1) + \lambda) \pi(0), & \text{если } 1 \leq m \leq s+1, \\ \frac{1}{\lambda^{s+1}} \prod_{i=1}^{s+1} (\nu(i-1) + \lambda) \pi(0), & \text{если } s+1 < m \leq S, \end{cases} \quad (4)$$

где $\pi(0)$ находится из условия нормировки, т.е. $\sum_{m=0}^S \pi(m) = 1$.

Замечание 1. В случае, когда время пополнения запасов является постоянным, т.е. $\nu(m) = \nu$ для любого $m = 0, 1, \dots, s$, формула (4) еще более упрощается:

$$\pi(m) = \begin{cases} \left(1 + \frac{\nu}{\lambda}\right)^m \pi(0), & \text{если } 1 \leq m \leq s+1, \\ \left(1 + \frac{\nu}{\lambda}\right)^{s+1} \pi(0), & \text{если } s+1 < m \leq S. \end{cases}$$

После нахождения вероятностей состояний из соотношений (4) можно вычислить усредненные характеристики исследуемой СОЗ. Так, средний уровень ресурсов на складе S_{av} определяется как математическое ожидание данной случайной величины, т.е.

$$S_{av} = \sum_{m=1}^S m \pi(m). \quad (5)$$

Потери p -заявок происходят вследствие отсутствия запасов системы, следовательно, для вычисления вероятности потери p -заявок P_p получим следующую формулу:

$$P_p = \pi(0). \quad (6)$$

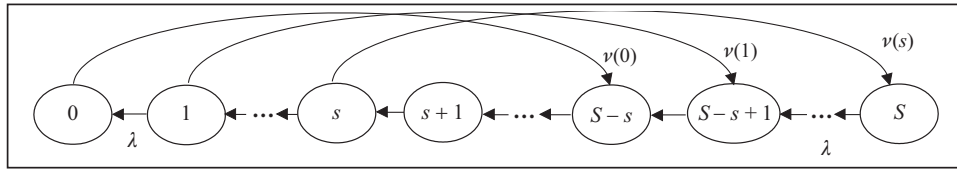


Рис. 2. Граф переходов между состояниями модели с мгновенным обслуживанием при (s, S) -политике пополнения запасов

Для сравнения предложенной ППЗ с известной (s, S) -политикой приведем формулы для последней схемы. Отметим, что при использовании (s, S) -политики работа системы также описывается конечной одномерной ЦМ, но при этом элементы ее Q -матрицы определяются так (рис. 2):

$$q(m, m') = \begin{cases} \nu(m), & \text{если } 0 \leq m \leq s, \quad m' = m + S - s, \\ \lambda, & \text{если } m' = m - 1, \\ 0 & \text{в остальных случаях.} \end{cases} \quad (7)$$

С использованием (7) можно показать, что вероятности состояний этой ЦМ определяются следующим образом (соответствующие преобразования также опускаются):

$$\pi(m) = \begin{cases} \alpha_m \pi(s+1), & \text{если } 0 \leq m \leq s, \\ \pi(s+1), & \text{если } s+1 \leq m \leq S-s, \\ \beta_m \pi(s+1), & \text{если } S-s+1 \leq m \leq S, \end{cases} \quad (8)$$

где $\alpha_m = \prod_{i=m+1}^{s+1} \frac{\lambda}{\lambda + \nu(i-1)}$, $\beta_m = \frac{1}{\lambda} \sum_{i=m-S+s}^s \alpha_i \nu(i)$ и $\pi(s+1)$ находится из условия нормировки.

Замечание 2. Формула (8) еще больше упрощается, когда $\nu(m) = \nu$ для любого $m = 0, 1, \dots, s$, так как в этом случае имеем

$$\alpha_m = \left(\frac{\lambda}{\lambda + \nu} \right)^{s-m+1}, \quad \beta_m = \frac{\nu}{\lambda} \sum_{i=m-S+s}^s \alpha_i.$$

Характеристики СОЗ при использовании (s, S) -политики пополнения запасов также определяются из формул (5) и (6).

Теперь рассмотрим модели СОЗ типа (ii). Вначале изучим модели с ограниченной очередью p -заявок. В отличие от модели типа (i) работу таких СОЗ нельзя описать одномерной ЦМ, так как в этих моделях, кроме уровня запасов системы, необходимо учитывать еще и число p -заявок в очереди. Следовательно, работа данной модели описывается конечной двумерной ЦМ с состояниями вида (m, n) , где, как и ранее, m — уровень ресурсов на складе, а n — число p -заявок в очереди. Фазовое пространство состояний (ФПС) определяется так (рис. 3, при $S = 4$ и $N = 3$):

$$E = \{(m, n) : m = 0, 1, \dots, S; n = 0, 1, \dots, N\}. \quad (9)$$

В этой модели интенсивность перехода из состояния $(m_1, n_1) \in E$ в другое состояние $(m_2, n_2) \in E$ обозначим $q((m_1, n_1), (m_2, n_2))$. Для вычисления элементов Q -матрицы данной двумерной ЦМ необходимо учитывать, что переходы между состояниями ФПС (9) связаны со следующими событиями: поступление p -заявок, завершение их обслуживания, а также уход из очереди вследствие их нетерпеливости и поступление запасов со склада верхнего уровня.

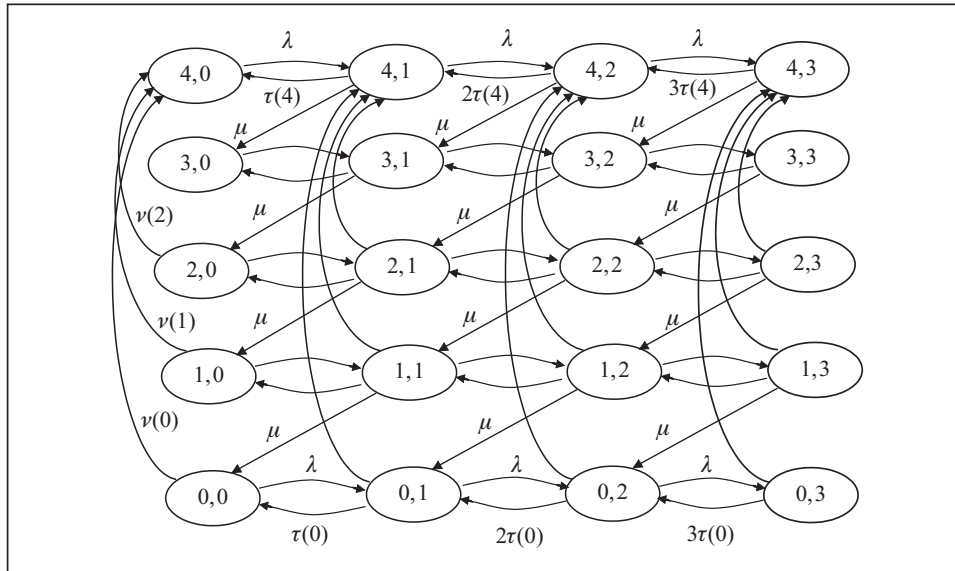


Рис. 3. Граф переходов между состояниями модели с ненулевым временем обслуживания при VSO-политике пополнения запасов

Исходя из принятой ППЗ, при определении исходного состояния $(m_1, n_1) \in E$ необходимо различать следующие случаи: $m_1 > s$ и $m_1 \leq s$. Тогда с учетом используемой схемы обслуживания p -заявок, их ухода из очереди вследствие нетерпеливости и предложенной ППЗ получаем, что элементы Q -матрицы данной модели определяются так (см. рис. 3):

- при $m_1 > s$

$$q((m_1, n_1), (m_2, n_2)) = \begin{cases} \lambda, & \text{если } m_2 = m_1, n_2 = n_1 + 1, n_1 \leq N - 1, \\ \mu, & \text{если } m_2 = m_1 - 1, n_2 = n_1 - 1, n_1 > 0, \\ n_1 \tau(m_1), & \text{если } m_2 = m_1, n_2 = n_1 - 1, n_1 > 0, \\ 0 & \text{в остальных случаях;} \end{cases} \quad (10)$$

- при $m_1 \leq s$

$$q((m_1, n_1), (m_2, n_2)) = \begin{cases} \lambda, & \text{если } m_2 = m_1, n_2 = n_1 + 1, n_1 \leq N - 1, \\ \mu, & \text{если } m_2 = m_1 - 1, n_2 = n_1 - 1, n_1 > 0, \\ n_1 \tau(m_1), & \text{если } m_2 = m_1, n_2 = n_1 - 1, n_1 > 0, \\ \nu(m_1), & \text{если } m_1 \leq s, m_2 = S, n_2 = n_1, \\ 0 & \text{в остальных случаях.} \end{cases} \quad (11)$$

С учетом соотношений (10) и (11) получаем, что все состояния этой конечной двумерной ЦМ являются сообщающимися (см. рис. 3), следовательно, в данной системе существует стационарный режим, т.е. стационарные вероятности состояний $(m, n) \in E$.

Пусть $p(m, n)$ обозначает вероятность состояния $(m, n) \in E$. Указанные вероятности находятся в результате решения соответствующей СУР для вероятностей состояний, которая составляется на основе соотношений (10) и (11) (явные формы этих уравнений здесь не приводятся ввиду очевидности их составления). В отличие от модели типа (i), в данной модели СУР имеет матрицу сложной структуры, отчего не удастся найти ее аналитического решения. Следовательно, возникает необходимость использования стандартных численных методов теории марковских процессов. Однако последние являются работоспособными лишь для моделей умеренной размерности и бесполезны для моделей большой размерности. Поэтому далее предлагается приближенный метод расчета стационар-

ного распределения данной модели с использованием подхода, разработанного в работе [5]. Отметим, что этот метод имеет высокую точность для моделей СОЗ, в которых интенсивность поступления p -заявок намного превосходит интенсивности их обслуживания. Если это допущение выполняется, то рассмотрим следующее расщепление исходного ФПС (9):

$$E = \bigcup_{m=0}^S E_m, \quad E_m \cap E_{m'} = \emptyset, \quad m \neq m', \quad (12)$$

где $E_m = \{(m, n) \in E : n = 0, 1, \dots, N\}$. Иными словами, проводится расщепление исходного ФПС (9) по строкам (см. рис. 3), т.е. класс состояний E_m содержит те состояния (m, n) из исходного ФПС (9), в которых уровень ресурсов равен m независимо от длины очереди p -заявок. Далее в исходном ФПС (9) определяется функция укрупнения

$$U((m, n)) = \langle m \rangle, \quad (13)$$

где $\langle m \rangle$ — укрупненное состояние, которое объединяет в себе класс состояний E_m , $m = 0, 1, \dots, S$. Полученное множество укрупненных состояний обозначается $\Omega = \{\langle m \rangle : m = 0, 1, \dots, S\}$.

Вероятность состояния (m, n) внутри расщепленной модели с пространством состояний E_m , $m = 0, 1, \dots, S$, обозначим $\rho_m(n)$. Следовательно, из соотношений (10) и (11) получим, что эти вероятности вычисляются как вероятности состояний классической модели Эрланга $M / M / N / 0$ с нагрузкой $\sigma_m = \lambda / \tau(m) \text{ erl.}$, т.е.

$$\rho_m(n) = \frac{\sigma_m^n}{n!} \rho_m(0), \quad (14)$$

где $\rho_m(0) = \left(\sum_{n=0}^N \frac{\sigma_m^n}{n!} \right)^{-1}$. Тогда с учетом (10), (11) и (14) после определенных

математических преобразований получаем, что интенсивности перехода из укрупненного состояния $\langle m_1 \rangle$ в другое укрупненное состояние $\langle m_2 \rangle$, которые обозначаются $q(\langle m_1 \rangle, \langle m_2 \rangle)$, $\langle m_1 \rangle, \langle m_2 \rangle \in \Omega$, вычисляются так:

$$q(\langle m_1 \rangle, \langle m_2 \rangle) = \begin{cases} x_{m_1}, & 0 \leq m_1 \leq s, \quad m_2 = S, \\ y_{m_1}, & 1 \leq m_1 \leq S, \quad m_2 = m_1 - 1, \\ 0 & \text{в остальных случаях,} \end{cases} \quad (15)$$

где $x_{m_1} = \nu(m_1)$, $y_{m_1} = \mu(1 - \rho_{m_1}(0))$.

Вероятность укрупненного состояния $\langle m \rangle \in \Omega$ обозначим $\pi(\langle m \rangle)$. Из соотношений (15) с учетом решения СУР (4) находим, что эти вероятности вычисляются следующим образом:

$$\pi(\langle m \rangle) = \begin{cases} \gamma_m \pi(\langle 0 \rangle), & 1 \leq m \leq s+1, \\ \eta_m \pi(\langle 0 \rangle), & s+1 < m \leq S, \end{cases} \quad (16)$$

где $\gamma_m = \prod_{i=1}^m \frac{x_{i-1} + y_{i-1}}{y_i}$; $\eta_m = \frac{y_{s+1}}{y_m} \prod_{i=1}^{s+1} \frac{x_{i-1} + y_{i-1}}{y_i}$; $y_0 = 0$.

Вероятность $\pi(\langle 0 \rangle)$ находится из условия нормировки, т.е. $\sum_{m=0}^S \pi(\langle m \rangle) = 1$.

Наконец, при использовании (14) и (16) вероятности состояний исходной модели приближенно находятся так:

$$p(m, n) \approx \rho_m(n)\pi(\langle m \rangle).$$

Точные значения усредненных значений характеристик исследуемой системы определяются как математические ожидания соответствующих случайных величин. Так, используя результаты работы [5], находим среднюю длину очереди p -заявок и средний уровень ресурсов на складе

$$L_{av} = \sum_{n=1}^N n \sum_{m=0}^S p(m, n);$$

$$S_{av} = \sum_{m=1}^S m \sum_{n=0}^N p(m, n).$$

Как отмечалось ранее, p -заявки теряются в следующих случаях: когда в момент ее поступления в очереди не имеется свободного места и если до окончания допустимого времени ожидания не освобождается канал обслуживания. Следовательно, точное значение вероятности потери p -заявок вычисляется с помощью формулы

$$P_p = \sum_{m=0}^S p(m, N) + \sum_{m=0}^S \sum_{n=1}^N p(m, n) P_l(m, n),$$

где $P_l(m, n)$ — вероятность того, что в состоянии (m, n) p -заявка теряется вследствие нетерпеливости, т.е.

$$P_l(m, n) = \frac{n\tau(m)}{n\tau(m) + \mu I(m \geq 1) + \lambda I(n < N)},$$

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

Аналогично [5] далее с использованием предложенного подхода получены следующие формулы для приближенного расчета характеристик исследуемой системы:

$$L_{av} \approx \sum_{n=1}^N n \sum_{m=0}^S \rho_m(n)\pi(\langle m \rangle);$$

$$S_{av} \approx \sum_{m=1}^S m\pi(\langle m \rangle); \tag{17}$$

$$P_p \approx \sum_{m=0}^S \rho_m(N)\pi(\langle m \rangle) + \sum_{m=0}^S \sum_{n=1}^N \rho_m(n)\pi(\langle m \rangle) P_l(m, n).$$

Теперь рассмотрим модель типа (ii) с неограниченной очередью, т.е. предположим, что $N = \infty$. Работа данной системы также описывается двумерной, но в рассматриваемом случае бесконечной ЦМ с состояниями вида (m, n) , где m — уровень ресурсов на складе, n — число p -заявок в очереди, т.е. фазовое пространство состояний бесконечномерное по второй компоненте.

Прямым методом расчета точных значений вероятностей состояний этой бесконечной двумерной ЦМ является метод двумерных производящих функций. Вместе с тем использование этого метода для данной модели затруднительно

виду сложной структуры ее Q -матрицы. Поэтому здесь также предлагается использовать описанный приближенный метод.

Поскольку этапы применения данного метода уже подробно описаны, кратко изложим специфику его применения для модели с бесконечным ФПС.

Элементы Q -матрицы этой двумерной ЦМ определяются подобно (10) и (11). Как и ранее, рассматривается аналогичное (12) расщепление бесконечного ФПС модели и соответствующим образом строится функция укрупнения (см. (13)).

В модели типа (ii) вероятности состояний внутри расщепленной модели с ФПС E_m вычисляются как вероятности состояний модели $M / M / \infty$ с нагрузкой σ_m erl., т.е.

$$\rho_m(n) = \frac{\sigma_m^n}{n!} e^{-\sigma_m}, \quad n = 0, 1, 2, \dots$$

Интенсивности переходов между состояниями укрупненной модели определяются аналогично (15), при этом следует иметь в виду, что в рассматриваемой модели $\rho_{m_1}(0) = 1 - e^{-\sigma_{m_1}}$.

Далее вероятности состояний укрупненной модели вычисляются в соответствии с формулами (16).

Средний уровень ресурсов также определяется с помощью формулы (17). Средняя длина очереди в модели типа (ii) с неограниченной очередью находится так:

$$L_{av} \approx \sum_{m=0}^S \pi(\langle m \rangle) \sigma_m. \quad (18)$$

Вероятность потери p -заявок вследствие нетерпеливости в этой модели определяется следующим образом:

$$P_p \approx \sum_{m=0}^S e^{-\sigma_m} \pi(\langle m \rangle) \sum_{n=1}^{\infty} \frac{\sigma_m^n}{n!} P_l(m, n). \quad (19)$$

ЧИСЛЕННЫЕ РЕЗУЛЬТАТЫ

Представим результаты численных экспериментов, проведенных с помощью разработанных выше алгоритмов.

Вначале рассмотрим результаты для модели с мгновенным временем обслуживания, при этом исследуем две схемы изменения функции $\nu(m)$ относительно аргумента m , $m = 0, 1, \dots, s$. В первой схеме время выполнения заказа — возрастающая функция, а во второй — оно убывает относительно текущего уровня запасов системы. Отметим, что первая схема реалистичнее второй, так как, исходя из практических соображений, следует ожидать, что чем ниже текущий уровень запасов системы, тем быстрее они должны пополняться. Однако в целях более подробного анализа данной модели СОЗ далее рассматриваются обе схемы изменения времени выполнения заказа. Для конкретизации изложения предполагается, что в первой схеме $\nu^{-1}(m) = m + 1$, а во второй схеме $\nu^{-1}(m) = 1 + 1/(m + 1)$, $m = 0, 1, \dots, s$.

Далее принимаем, что максимальный размер склада системы $S = 50$, а интенсивность p -заявок равна 10 заявок за единицу времени, критический уровень запасов s — изменяемый параметр системы.

Поведение характеристик СОЗ при применении первой (реалистичной) схемы изменения времени выполнения заказа относительно критического уровня запасов показано на рис. 4, где на графиках приведены изучаемые характеристики систе-

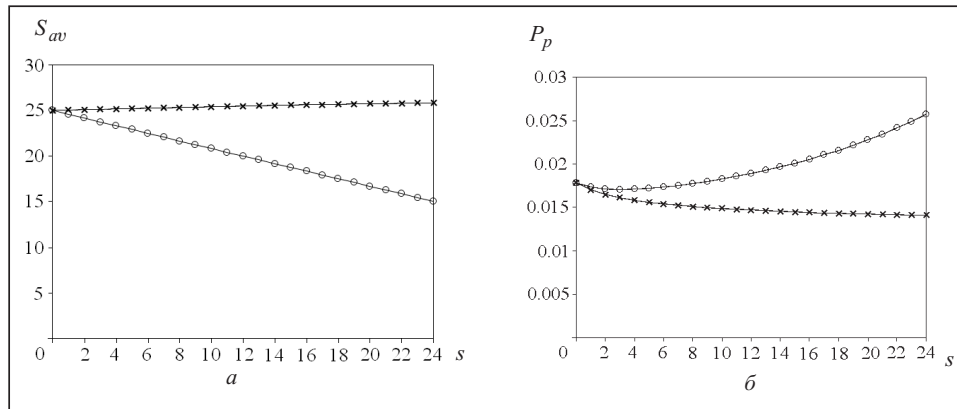


Рис. 4. Зависимости среднего уровня запасов системы (а) и вероятности потери p -заявок (б) от параметра s в модели с мгновенным обслуживанием при $v(m) = 1 / (m + 1)$

мы при использовании обеих ППЗ, которые определяются из формул (1) и (2) (здесь и далее символами \times и \circ отмечены кривые изменения показателя QoS при VSO- и (s, S) -политиках пополнения запасов соответственно).

Из рис. 4, а видно, что при использовании (s, S) -политики с ростом критического уровня запасов их средний уровень уменьшается с достаточно высокой скоростью: от значения 25 до 15. Однако при использовании VSO-политики эта характеристика растет с очень малой скоростью (можно считать, что она является почти постоянной величиной, равной 25). Отметим, что при использовании (s, S) -политики средний уровень запасов системы оказывается меньшим, чем при VSO-политике. Этого следовало ожидать, так как при использовании VSO-политики каждый раз (при фиксированном значении критического уровня запасов) поставляется запасов больше, чем при (s, S) -политике.

Вместе с тем для вероятности потери p -заявок наблюдается обратная зависимость (см. рис. 4, б), т.е. при использовании (s, S) -политики с ростом критического уровня запасов вероятность потери p -заявок растет, а при использовании VSO-политики эта характеристика уменьшается. Первый факт объясняется тем, что в (s, S) -политике с ростом критического уровня запасов его средний уровень уменьшается (см. рис. 4, а), при этом вероятности потери p -заявок оказываются большими, чем при использовании VSO-политики. Последний факт также имеет логическое обоснование, так как при использовании VSO-политики средний уровень запасов системы больше, чем при (s, S) -политике (см. рис. 4, а), и следовательно, вероятность потери p -заявок будет меньшей, чем при использовании (s, S) -политики.

Отметим, что из рис. 4 видно, что скорости изменения характеристик системы при использовании (s, S) -политики существенно больше, чем при VSO-политике.

Поведение характеристик COЗ при применении второй (нереалистичной) схемы изменения времени выполнения заказа относительно критического уровня запасов показано на рис. 5. Как и в случае первой схемы изменения времени выполнения заказа, средний уровень запасов при использовании VSO-политики также растет, но с достаточно высокой скоростью (начиная от значения 25, возрастает до 32); однако, в отличие от первой схемы указанная характеристика при использовании (s, S) -политики при $0 \leq s \leq 11$ уменьшается с очень малой скоростью от значения 25 до 24, а при $12 \leq s \leq 25$ она также с малой скоростью возрастает до значения 28 (см. рис. 5, а). Такое поведение объясняется нереалистичнос-

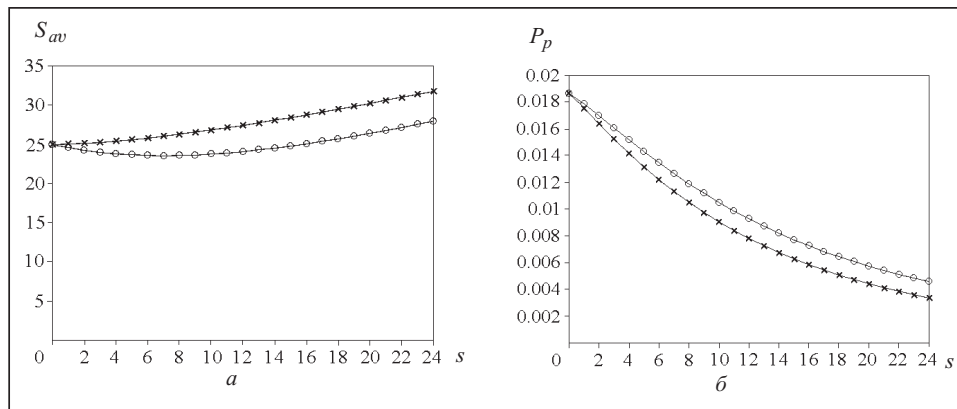


Рис. 5. Зависимости среднего уровня запасов системы (а) и вероятности потери p -заявок (б) от параметра s в модели с мгновенным обслуживанием при $\nu(m) = (m + 1) / (m + 2)$

тью второй схемы, согласно которой чем выше текущий уровень запасов системы, тем быстрее они пополняются. Заметим, что в данном случае также при использовании (s, S) -политики средний уровень запасов системы оказывается меньшим, чем при VSO-политике.

Отметим, что при применении второй схемы изменения времени выполнения заказа вероятность потери p -заявок с ростом критического уровня запасов уменьшается в обеих ППЗ (см. рис. 5, б). Этого следовало ожидать, так как в этой схеме средний уровень запасов растет при использовании обеих ППЗ (см. рис. 5, а). Как и в первой схеме, в данном случае также при использовании (s, S) -политики вероятности потери p -заявок оказываются большими, чем при VSO-политике.

Как видно из графиков на рис. 4 и 5, в упомянутых схемах изменения времени выполнения заказа при $s=0$ характеристики системы с использованием обеих ППЗ совпадают. Этого следовало ожидать, так как при $s=0$ VSO-политика совпадает с (s, S) -политикой.

Рассмотрим результаты сравнения характеристик системы при различных схемах изменения времени выполнения заказа. Относительно VSO-политики пополнения запасов отметим, что если в реалистичной схеме средний уровень запасов системы является почти постоянным (он равен 25), то в нереалистичной эта характеристика изменяется в интервале [25, 32]; иными словами, в нереалистичной схеме средний уровень запасов системы оказывается большим, чем в реалистичной. Вероятность потери p -заявок в реалистичной схеме изменяется в интервале [0.01785, 0.01412], а в нереалистичной эта характеристика изменяется в интервале [0.01869, 0.00338]; иными словами, в нереалистичной схеме вероятность потери p -заявок оказывается меньшей, чем в реалистичной. Последний факт объясняется тем, что в нереалистичной схеме средний уровень запасов системы оказывается большим, чем в реалистичной.

Для (s, S) -политики пополнения запасов отметим, что в реалистичной схеме средний уровень запасов системы изменяется в интервале [15, 25], а в нереалистичной эта характеристика изменяется в интервале [25, 27]; иными словами, в данном случае также в нереалистичной схеме средний уровень запасов системы оказывается большим, чем в реалистичной. Вероятность потери p -заявок в реалистичной схеме изменяется в интервале [0.01785, 0.02578], а в нереалистичной эта характеристика изменяется в интервале [0.01869, 0.00459]; иными словами, в нереалистичной схеме вероятность потери p -заявок оказывается меньшей, чем в реалистичной.

Эти сравнения позволяют сделать вывод о том, что в обеих ППЗ относительно уменьшения расходов системы на содержание запасов благоприятнее использование реалистичной схемы, в то время как относительно уменьшения расходов системы ввиду потери p -заявок предпочтительнее применение нереалистичной схемы времени выполнения заказа.

Проведенный анализ характеристик системы при различных политиках пополнения запасов и схемах изменения времени выполнения заказа дает возможность постановки и решения задач по определению надлежащих значений критического уровня запасов в целях удовлетворения заданных ограничений на указанные характеристики. Так, одна из возможных постановок подобных задач состоит в следующем. Пусть для выбранной политики (VSO- или (s, S) -политики) и схемы изменения времени выполнения заказа (реалистичной или нереалистичной) приняты ограничения на характеристики системы:

$$S_{av} \leq \delta, \quad (20)$$

$$P_p \leq \varepsilon, \quad (21)$$

где δ и ε являются заданными величинами.

Задача заключается в нахождении таких значений s , которые удовлетворяют ограничениям (20) и (21).

Очевидно, что существование решения задачи (18) и (19) зависит от значений параметров δ и ε . Так, например, для VSO-политики и реалистичной схемы изменения времени выполнения заказа она не имеет решений, если $\delta = 20$ и $\varepsilon = 0.02$. В то же время при этих ограничениях она имеет решения для (s, S) -политики, здесь решением задачи являются $s = 13$ и $s = 14$. Аналогично можно сформулировать другие задачи выбора критического уровня запасов изучаемой системы.

Теперь рассмотрим результаты численных экспериментов для системы с ненулевым временем обслуживания. Поскольку в данной системе количество исходных параметров больше, чем в предыдущей, для простоты анализа результатов численных экспериментов предполагается, что время выполнения заказа является постоянной величиной, т.е. принимается $\nu(m) = 1$ для всех m , $m = 0, 1, \dots, s$. Вместе с тем здесь также изучаются две схемы изменения функции $\tau^{-1}(m)$ относительно аргумента m , $m = 0, 1, \dots, S$, т.е. предполагается, что в первой схеме допустимое время ожидания в очереди p -заявок является возрастающей функцией, а во второй — оно убывает относительно текущего уровня запасов системы. В данном случае первая схема также является более реалистичной, чем вторая, так как, исходя из практических соображений, следует ожидать, что чем ниже текущий уровень запасов системы, тем более нетерпеливыми оказываются p -заявки в очереди.

Далее рассматриваются обе схемы изменения параметра $\tau(m)$, который оценивает степень нетерпеливости заявок. Для конкретизации изложения вид этих функций здесь определяется аналогично предыдущим моделям, т.е. в первой схеме $\tau^{-1}(m) = m + 1$, а во второй схеме $\tau^{-1}(m) = 1 + 1/(m + 1)$, $m = 0, 1, \dots, S$.

Вначале рассмотрим результаты численных экспериментов для модели с ограниченной очередью p -заявок, т.е. предположим, что максимальный размер буфера для ожидания в очереди p -заявок $N = 100$. Максимальный размер склада системы, как и ранее, $S = 50$, интенсивность поступления p -заявок равна $\lambda = 0.1$ заявок за единицу времени, а интенсивность их обслуживания $\mu = 0.05$. Изменяемым параметром системы также является критический уровень запасов, т.е. s .

В целях изучения эффективности применения вводимой здесь VSO-политики приведем сравнительный анализ характеристик системы с аналогичными характеристиками при использовании (s, S) -политики. (Методы и алгоритмы нахождения характеристик изучаемой системы при использовании (s, S) -политики пополнения запасов разработаны в [5]).

Поведение характеристик CO3 при использовании реалистичной схемы изменения функции $\tau^{-1}(m)$ относительно аргумента $m, m = 0, 1, \dots, S$, показано на рис. 6.

Из рис. 6, *a* видно, что при использовании VSO-политики с ростом критического уровня запасов их средний уровень растет почти линейно от значения 1 до 25; однако при использовании (s, S) -политики указанная характеристика вначале растет с очень большой скоростью, т.е. при $s = 0$ ее значение равно 20, а при $s = 1$ средний уровень запасов становится почти равным 50 (склад системы почти всегда оказывается заполненным полностью), а далее она убывает почти линейно до значения 38. Иными словами, при использовании (s, S) -политики средний уровень запасов системы оказывается существенно бóльшим, чем при VSO-политике.

Зависимость вероятности потери p -заявок от критического уровня запасов показана на рис. 6, *б*. Из этого рисунка видно, что при использовании (s, S) -политики при начальных значениях критического уровня запасов вероятность потери p -заявок растет с очень малой скоростью и далее она убывает с почти линейной скоростью: а именно при $s = 0$ ее значение равно 0.3121, при $s = 2$ вероятность потери равна 0.3803 и при $s \geq 3$ указанная характеристика начинает уменьшаться до значения 0.3744. При использовании VSO-политики этот показатель систематически возрастает с достаточно высокой скоростью от значения 0.1409 до 0.3625. Иными словами, при использовании (s, S) -политики вероятность потери p -заявок оказывается существенно бóльшей, чем при использовании VSO-политики.

Зависимость средней длины очереди p -заявок от критического уровня запасов показана на рис. 6, *в*. Очевидно, что эти графики идентичны графикам на рис. 6, *a*, т.е. при использовании (s, S) -политики средняя длина очереди p -заявок оказывается бóльшей, чем при VSO-политике. Следовательно, для реалистичной схемы изменения функции $\tau^{-1}(m)$ относительно аргумента $m, m = 0, 1, \dots, S$, все характеристики системы являются существенно лучшими при использовании VSO-политики пополнения запасов.

Поведение характеристик CO3 при использовании нереалистичной схемы изменения функции $\tau^{-1}(m)$ относительно аргумента $m, m = 0, 1, \dots, S$, показано на рис. 7.

Из рис. 7, *a* видно, что при использовании обеих политик пополнения запасов поведение характеристики S_{av} относительно роста критического уровня запасов почти полностью совпадает с реалистичной схемой изменения функции $\tau^{-1}(m)$ относительно аргумента $m, m = 0, 1, \dots, S$ (см. рис. 6, *a*), при этом почти совпадают их абсолютные значения, т.е. в данном случае также при использовании (s, S) -политики средний уровень запасов системы оказывается существенно бóльшим, чем при VSO-политике.

Из рис. 7, *б* видно, что при использовании (s, S) -политики вероятность потери p -заявок является практически постоянной и равна 0.0844, а при использовании VSO-политики указанная характеристика систематически уменьшается с малой скоростью от значения 0.1147 до 0.0857; иными словами, при использовании

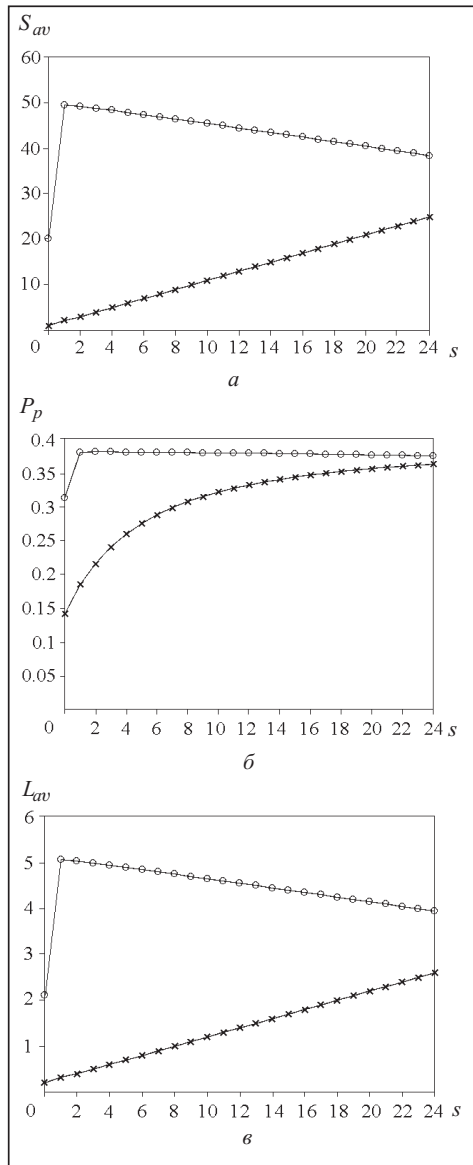


Рис. 6. Зависимости среднего уровня запасов системы (а), вероятности потери p -заявок (б) и средней длины очереди p -заявок (в) от параметра s в модели с ограниченной очередью и ненулевым временем обслуживания при $\tau(m) = 1 / (m + 1)$

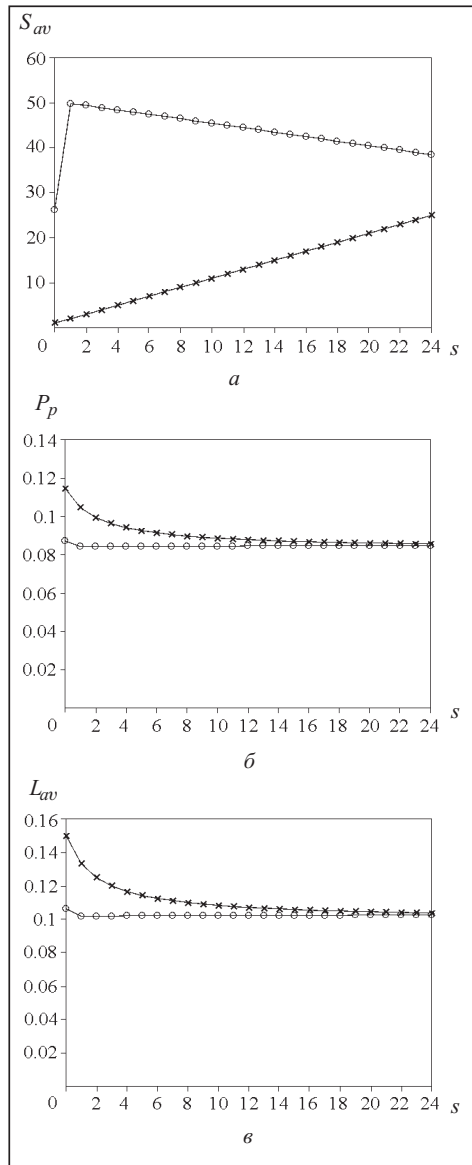


Рис. 7. Зависимости среднего уровня запасов системы (а), вероятности потери p -заявок (б) и средней длины очереди p -заявок (в) от параметра s в модели с ограниченной очередью и ненулевым временем обслуживания при $\tau(m) = (m + 1) / (m + 2)$

(s, S)-политики вероятность потери p -заявок оказывается чуть меньшей, чем при VSO-политике при малых значениях критического уровня запасов, а при $s \geq 10$ они почти уравниваются.

Зависимость средней длины очереди p -заявок от критического уровня запасов показана на рис. 7, в. Из него видно, что эти графики идентичны графикам на рис. 7, б, т.е. при использовании (s, S)-политики средняя длина очереди p -заявок оказывается чуть меньшей, чем при VSO-политике. Следовательно, для нереалистичной схемы изменения функции $\tau^{-1}(m)$ относительно аргумента m , $m = 0, 1, \dots, S$, только характеристика S_{av} является существенно лучшей при ис-

пользовании VSO-политики пополнения запасов, а остальные две характеристики: P_p и L_{av} , почти одинаковы при обеих политиках пополнения запасов.

Сравнение абсолютных значений характеристик системы при различных схемах изменения времени выполнения заказа показало, что лишь средний уровень запасов системы оказывается существенно меньшим при VSO-политике пополнения запасов, если используется реалистичная схема, т.е. средний уровень запасов системы при VSO-политике пополнения запасов изменяется в интервале [1, 24], в то время как при (s, S) -политике эта характеристика изменяется в интервале [26, 38]. Остальные две характеристики системы почти не различаются в обеих ППЗ в приведенных схемах изменения времени выполнения заказа.

Теперь рассмотрим результаты численных экспериментов для системы с ненулевым временем обслуживания и неограниченной очередью p -заявок, т.е. предположим, что $N = \infty$. Значения нагрузочных параметров модели остаются неизменными.

Поведение характеристик модели CO3 с неограниченной очередью p -заявок при использовании реалистичной схемы изменения функции $\tau^{-1}(m)$ относительно аргумента m , $m = 0, 1, \dots, S$, показано на рис. 8.

Из рис. 8, *а* видно, что при использовании обеих политик пополнения запасов поведение характеристики S_{av} существенно отличается от модели с ограниченной очередью (см. рис. 7, *а*). Заметим, что при использовании (s, S) -политики средний уровень запасов системы оказывается существенно меньшим, чем при VSO-политике.

Из рис. 8, *б* видно, что при использовании VSO-политики вероятность потери p -заявок является почти постоянной и равна 0.3808, а при использовании (s, S) -политики указанная характеристика растет с малой скоростью от значения 0.3121 до 0.3743 и далее также с малой скоростью уменьшается до значения 0.3701; иными словами, при использовании (s, S) -политики вероятность потери p -заявок оказывается чуть меньшей, чем при VSO-политике при малых значениях критического уровня запасов, а при $s \geq 10$ их значения почти совпадают.

Зависимость средней длины очереди p -заявок от критического уровня запасов показана на рис. 8, *в*. Из него видно, что эти графики идентичны графикам на рис. 8, *а*, т.е. при использовании (s, S) -политики средняя длина очереди p -заявок оказывается меньшей, чем при VSO-политике. Следовательно, для данной модели при реалистичной схеме изменения функции $\tau^{-1}(m)$ относительно аргумента m , $m = 0, 1, \dots, S$, все характеристики являются существенно лучшими при использовании (s, S) -политики пополнения запасов.

Поведение характеристик модели CO3 с неограниченной очередью p -заявок при применении нереалистичной схемы изменения функции $\tau^{-1}(m)$ относительно аргумента m , $m = 0, 1, \dots, S$, показано на рис. 9.

Из рис. 9, *а* видно, что при использовании VSO-политики с ростом критического уровня запасов их средний уровень растет почти линейно от значения 25 до 37; однако при использовании (s, S) -политики указанная характеристика вначале растет с очень большой скоростью, т.е. при $s = 0$ ее значение равно 26, а при $s = 10$ средний уровень запасов становится равным 41, а далее он убывает до значения 32. Иными словами, при использовании (s, S) -политики средний уровень запасов системы оказывается существенно большим, чем при VSO-политике, если $s \leq 20$, а при $s > 20$ средний уровень запасов системы является большим при использовании VSO-политики.

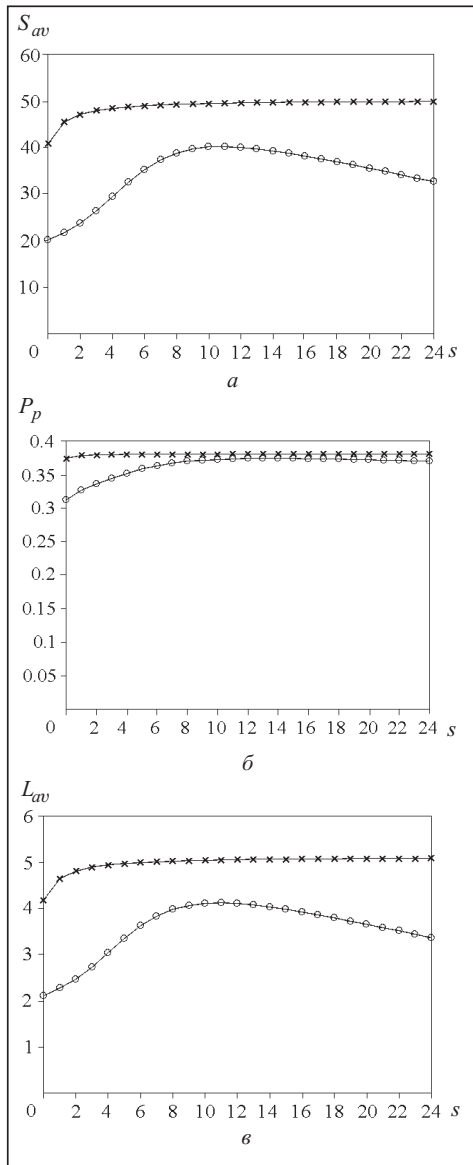


Рис. 8. Зависимости среднего уровня запасов системы (а), вероятности потери p -заявок (б) и средней длины очереди p -заявок (в) от параметра s в модели с неограниченной очередью и ненулевым временем обслуживания при $\tau(m) = 1 / (m + 1)$

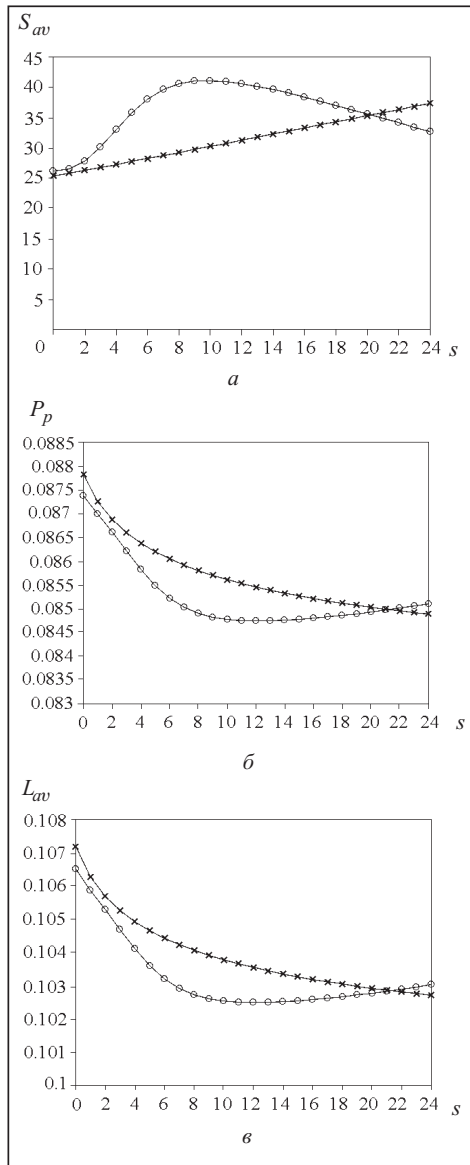


Рис. 9. Зависимости среднего уровня запасов системы (а), вероятности потери p -заявок (б) и средней длины очереди p -заявок (в) от параметра s в модели с неограниченной очередью и ненулевым временем обслуживания при $\tau(m) = (m + 1) / (m + 2)$

Из рис. 9, б видно, что вероятности потери p -заявок при использовании обеих политик пополнения запасов почти совпадают, т.е. они отличаются лишь в третьем знаке после десятичной точки. При этом значения этой характеристики оказываются чуть меньшими при использовании (s, S) -политики, если $s \leq 20$.

Зависимость средней длины очереди p -заявок от критического уровня запасов показана на рис. 9, в. Эти графики идентичны графикам на рис. 9, б, т.е. при использовании (s, S) -политики средняя длина очереди p -заявок оказывается чуть меньшей, чем при VSO-политике, $s \leq 20$. Следовательно, для данной модели при нереалистичной схеме изменения функции $\tau^{-1}(m)$ относительно аргумента m ,

$m = 0, 1, \dots, S$, характеристика S_{av} является существенно лучшей при использовании VSO-политики, $s \leq 20$, а остальные две характеристики оказываются чуть лучшими при (s, S) -политике пополнения запасов.

Сравнение абсолютных значений характеристик системы с неограниченной очередью p -заявок при различных схемах изменения времени выполнения заказа показало, что в обеих ППЗ все характеристики системы оказываются лучшими при нереалистичной схеме изменения функции $\tau^{-1}(m)$ относительно аргумента m , $m = 0, 1, \dots, S$.

ЗАКЛЮЧЕНИЕ

В работе предложена новая политика пополнения запасов в системах обслуживания–запасания. Отличительной особенностью этой политики является то, что в ней объемы заказов являются переменными величинами, которые зависят от текущего уровня запасов. Рассмотрены модели двух типов: с мгновенным обслуживанием расходующих заявок и с ненулевым временем обслуживания. Получены явные формулы для расчета характеристик моделей СОЗ с мгновенным обслуживанием расходующих заявок при использовании предложенной политики, а также известной политики двух уровней.

В моделях с ненулевым временем обслуживания предполагается, что нетерпеливые расходующие заявки могут образовывать очередь конечной или бесконечной длины, при этом интенсивности потери расходующих заявок из очереди вследствие нетерпеливости зависят от текущего уровня запасов системы. Показано, что математическими моделями изучаемых систем являются двумерные цепи Маркова, и разработаны точный и приближенный методы для определения характеристик изучаемых СОЗ при использовании предложенной политики пополнения запасов. Полученные результаты позволяют провести сравнительный анализ характеристики изучаемых СОЗ при использовании различных политик пополнения запасов.

СПИСОК ЛИТЕРАТУРЫ

1. Прабху Н. Методы теории массового обслуживания и управления запасами (изучение основных случайных процессов). Пер. с англ. Москва: Машиностроение, 1969. 356 с.
2. Прабху Н. Стохастические процессы теории запасов. Пер. с англ. Москва: Мир, 1984. 184 с.
3. Рыжиков Ю.И. Теория очередей и управление запасами. С.-Петербург: Питер, 2001. 384 с.
4. Рубальский Г.Б. Вероятностные и вычислительные методы управления запасами. Москва: Знание, 1987. 115 с.
5. Melikov A.Z., Ponomarenko L.A., Bagirova S.A. Analysis of queueing-inventory systems with impatience consume customers. *Journal of Automation and Information Sciences*. 2016. Vol. 48, N 1. P. 53–68.
6. Demchenko I.Yu. Stationary distribution in inventory control models. *Cybernetics and Systems Analysis*. 2016. Vol. 52, N 2. P. 319–322.

Надійшла до редакції 02.09.2016

А.З. Меліков, Л.А. Пономаренко, С.А. Багірова
МАРКОВСЬКІ МОДЕЛІ СИСТЕМ ОБСЛУГОВУВАННЯ–ЗАПАСАННЯ ЗІ ЗМІННИМ
ОБСЯГОМ ЗАМОВЛЕНЬ

Анотація. Досліджено марковські моделі систем обслуговування–запасання зі змінним обсягом замовлень. Розглянуто два класи моделей: з миттєвим обслуговуванням і з ненульовим часом обслуговування. У моделі з ненульовим часом обслуговування передбачено, що нетерплячі витрачальні вимоги можуть утворювати чергу скінченної або нескінченної довжини. Розроблено точний і наближений методи для визначення характеристик таких систем з використанням запропонованої політики поповнення запасів.

Ключові слова: системи обслуговування–запасання, політика поповнення запасів, змінний обсяг замовлення.

A.Z. Melikov, L.A. Ponomarenko, S.A. Bagirova
MARKOV MODELS OF QUEUING-INVENTORY SYSTEMS
WITH VARIABLE SIZE OF ORDER

Abstract. Markov models of the queuing-inventory systems with variable order size are investigated. Two classes of models, with instantaneous and non-zero service times of customers are considered. The model with non-zero service time assumes that impatient customers can form a queue of either finite or infinite length. Exact and approximate methods are developed to calculate the characteristics of the systems under given lead policies.

Keywords: queuing-inventory systems, lead policy, variable order size.

Меліков Агаси Зарбали оглы,
чл.-кор. НАН Азербайджана, доктор техн. наук, професор, зав. лабораторией Института систем управления НАН Азербайджана, Баку, e-mail: agassi.melikov@rambler.ru.

Пonomаренко Леонид Анатольевич,
доктор техн. наук, профессор, главный научный сотрудник Международного научно-учебного центра информационных технологий и систем НАН Украины и МОН Украины, Киев,
e-mail: laronomarenko@ukr.net.

Багірова Севидж Алакбер кызы,
докторант Бакинского государственного университета, Азербайджан, e-mail: aseva90@gmail.com.