

УДК 519.872

А.З. Меликов, Л.А. Пономаренко, И.А. Алиев

АНАЛИЗ И ОПТИМИЗАЦИЯ МОДЕЛЕЙ СИСТЕМ ОБСЛУЖИВАНИЯ-ЗАПАСАНИЯ С ДВУМЯ ТИПАМИ ЗАЯВОК

Введение

В реальной жизни существует много систем обслуживания, в которых для обработки заявок кроме свободного сервера требуется и определенное количество запасов (ресурсов). Такие системы называются системами обслуживания запасания (Queuing Inventory Systems — QIS) [1].

В научной литературе подробно изучены модели QIS с идентичными (по всем показателям) заявками. Обзор работ, посвященных изучению моделей таких QIS, можно найти, например, в [2]. Следует отметить, что иногда допущение об идентичности заявок не соответствует реальным условиям работы QIS, так как зачастую заявки имеют различные размеры (под размером заявки понимается объем требуемого запаса). Другим примером QIS с неидентичными заявками являются системы, в которых некоторые заявки платят за один и тот же запас больше, чем другие.

Краткий обзор публикаций, посвященных изучению моделей QIS с различными типами заявок, дан в [3]. После появления указанной работы авторам стали известны публикации [4–7]. В них рассматривались «чистые» системы управления запасами без учета очереди заявок, при этом предполагалось, что изучаемые системы наблюдаются периодически, т.е. рассматриваемое время делится на последовательные интервалы равной длины, и система наблюдается лишь в начале или в конце каждого интервала. Более того, в них полагается, что время выполнения заказов равно нулю (т.е. заказы выполняются мгновенно). Детальный обзор работ, в которых изучались модели подобных систем, дан в [8]. Современный уровень технических средств позволяет непрерывно наблюдать как за уровнем запасов системы, так и за процессами поступления и обслуживания заявок. В связи с этим в последние годы интенсивно изучаются модели QIS с непрерывным временем наблюдения, т.е. в них решения относительно пополнения запасов системы, а также управления процессами поступления и обслуживания заявок принимаются в реальном времени. Одной из первых работ в этом направлении была [9], где изучалась модель с двумя типами заявок и детерминированным временем пополнения запасов. В дальнейшем эта модель была обобщена в [10–13].

В QIS с разнотипными заявками актуальны задачи выбора надлежащих схем обслуживания разнотипных заявок в целях оптимизации работы системы относительно выбранного критерия качества. Подобные задачи для моделей QIS с мгновенным обслуживанием и двумя типами заявок при использовании различных по-

литик пополнения запасов изучены в [14–17]. В [14–16] предполагается, что заявки низкого приоритета получают отказ, если в моменты их поступления уровень запасов системы меньше, чем некоторое пороговое значение; заявки высокого приоритета принимаются, если уровень запасов системы больше нуля. В отличие от [14–16] в [17] предполагается, что заявки низкого приоритета не получают отказа, а уходят в орбиту, если в моменты их поступления уровень запасов системы меньше, чем пороговое значение.

Вместе с тем задачи выбора надлежащих схем обслуживания разнотипных заявок мало изучены для моделей QIS с положительным временем обслуживания заявок. Одной из первых работ в этом направлении является [3], где предложена отличная от [14–17] схема доступа заявок в систему: заявки высокого приоритета принимаются при наличии хотя бы одного свободного места в буфере, а заявки низкого приоритета принимаются лишь тогда, когда общая длина очереди заявок меньше заданного порогового значения. Разработаны методы расчета и оптимизации характеристик системы при использовании двух политик пополнения запасов: политики двух уровней (т.е. (s, S) -политики) и политики переменного объема запасов (Variable Order Size — VOS) [18, 19]. В [3] для обозначения VOS-политики предложена символическая запись $(m, S - m)$, которая используется и в данной работе.

В настоящей публикации изучается модель QIS с двумя типами заявок и положительным временем обслуживания, в которой используется предложенная в [14–16] схема доступа. Подобная модель с конечной очередью была изучена в [20], где для расчета ее характеристик предложен точный метод, основанный на решении балансовых уравнений для двумерных цепей Маркова (ЦМ). Здесь изучаются модели QIS с общей конечной и бесконечной очередью разнотипных заявок и предложены методы анализа и оптимизации ее характеристик при использовании (s, S) -, $(S - 1, S)$ - и $(m, S - m)$ - политик пополнения запасов.

Описание моделей и постановка задачи

Рассматриваются модели QIS с двумя типами заявок при использовании различных политик пополнения запасов. Во всех моделях предполагается, что система имеет склад ограниченного объема S , $S < \infty$, и в эту систему поступают пуассоновские потоки заявок двух типов. Интенсивность потока заявок первого типа (обычные) равна λ_1 , а интенсивность потока заявок второго типа (приоритетные) равна λ_2 . Время обслуживания заявок обоих типов — экспоненциально распределенная случайная величина с общим средним μ . По завершении обслуживания заявки любого типа она с вероятностью σ_1 не получает запас, а с дополнительной вероятностью $\sigma_2 = 1 - \sigma_1$ получает. При этом если заявка любого типа получает запас, то уровень запасов системы уменьшается на единицу.

Рассматривается три политики пополнения запасов: 1) (s, S) — политика, согласно которой при снижении уровня запасов до величины s , $s < S/2$, делается заказ на поставку запасов объема $S - s$; 2) $(S - 1, S)$ — политика, в которой заказ делается каждый раз после уменьшения уровня запасов, при этом объем заказа равен единице; 3) $(m, S - m)$ — политика, согласно которой заказ на поставку запасов делается лишь тогда, когда уровень запасов опускается до величины s , при этом его объем зависит от текущего уровня m и равен $S - m$.

Во всех политиках заказ выполняется с некоторой задержкой, при этом для простоты изложения считается, что она распределена показательно с постоянным параметром ν .

В моделях с конечной очередью предполагается, что заявки разных типов ожидают в общей очереди с максимальной длиной N , $N < \infty$. Предполагается, что если в момент поступления заявки второго типа в очереди имеется свободное место, то она принимается в очередь; иначе она теряется. Заявка первого типа принимается в очередь лишь тогда, когда в момент ее поступления уровень запасов не меньше, чем s ; иными словами, если в момент поступления заявки первого типа уровень запасов меньше, чем s , то она теряется независимо от наличия свободных мест в очереди. Если в момент поступления заявки второго типа уровень запасов равен нулю, то она с вероятностью φ_1 присоединяется к очереди, а с вероятностью $\varphi_2 = 1 - \varphi_1$ уходит из системы необслуженной.

В моделях с бесконечной очередью предполагается, что заявка высокого приоритета всегда принимается в очередь, а заявка низкого приоритета принимается лишь тогда, когда в момент ее поступления уровень запасов больше, чем s .

Во всех моделях заявки являются нетерпеливыми, т.е. если во время их ожидания в очереди уровень запасов системы опускается до нулевого значения, то заявка уходит из системы после случайного времени, которое имеет показательное распределение с параметром τ .

Задача исследования всех моделей заключается в следующем: 1) найти совместное распределение уровня запасов системы и количества заявок в системе; 2) найти средний уровень запасов (S_{av}), среднюю интенсивность заказов (RR) и вероятности потери заявок каждого типа (PB_1, PB_2); 3) решить задачу максимизации прибыли системы за счет выбора надлежащих значений критического уровня запасов s .

Метод решения задачи

Сразу отметим, что во всех моделях работа QIS с конечной очередью описывается двумерной ЦМ с состояниями вида (m, n) , где m — уровень запасов в складе, n — общее число заявок в системе. Пространство состояний (ПС) определяется как $E = \{0, 1, \dots, S\} \times \{0, 1, \dots, N\}$.

Сначала рассмотрим модель QIS при использовании (s, S) -политики. Интенсивность перехода от состояния $(m, n) \in E$ в состояние $(m', n') \in E$ обозначим $q((m, n), (m', n'))$. Эти величины составляют производящую матрицу (ПМ) данной ЦМ. Для их определения необходимо различать следующие случаи: 1) $m \leq s$; 2) $m > s$.

В первом случае возможные выходы из состояния (m, n) связаны с наступлением следующих событий: (i) поступлением заявок; (ii) завершением обслуживания заявок; (iii) уходом заявок из очереди из-за нетерпеливости и (iv) пополнением запасов.

Если наступает событие типа (i), где $m > 0$, при этом заявка низкоприоритетна, то система не меняет своего состояния, поскольку в таких состояниях заявки данного типа теряются; если заявка высокоприоритетна, то она принимается системой при наличии хотя бы одного свободного места в очереди с вероятностью 1. Тогда следующим состоянием системы будет $(m, n + 1)$ и интенсивность такого перехода равна λ_2 .

Если наступает событие типа (i), где $m = 0$, при этом заявка низкоприоритетна, то, как и выше, система не меняет своего состояния; если заявка высокоприоритетна, то она принимается системой при наличии хотя бы одного свободного места в очереди с вероятностью φ_1 . Тогда следующим состоянием системы будет $(0, n+1)$ и интенсивность такого перехода равна $\lambda_2\varphi_1$.

Если наступает событие типа (ii), при этом заявка получает запасы, то система переходит в состояние $(m, n-1)$ с интенсивностью μ_1 ; иначе следующим состоянием системы будет $(m-1, n-1)$, при этом интенсивность такого перехода равна μ_2 .

Если наступает событие типа (iii), то осуществляется переход в состояние $(0, n-1)$, при этом интенсивность такого перехода равна $n\tau$.

Если наступает событие типа (iv), то происходит переход в состояние $(m+S-s, n)$, при этом интенсивность такого перехода равна ν .

Следовательно, в случае $m < s$ положительные элементы ПМ данной двумерной ЦМ определяются следующим образом (рис.: \times , \diamond и \circ относятся к результатам для политик (s, S) , $(m, S-m)$ и $(S-1, S)$ соответственно).

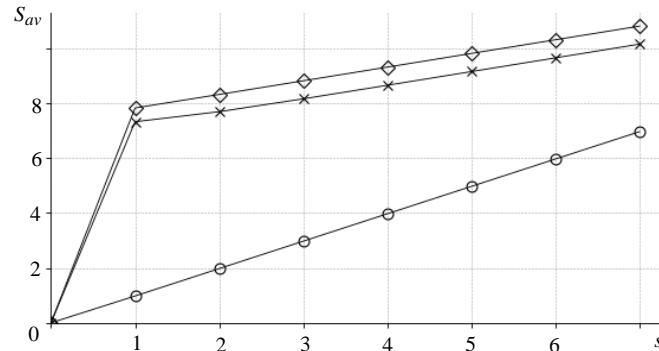


Рис. 1

$$q((m, n), (m', n')) = \begin{cases} \lambda_2, & \text{если } m > 0, m' = m, n' = n+1, \\ \lambda_2\varphi_1, & \text{если } m' = m = 0, n' = n+1, \\ \mu_1, & \text{если } m > 0, m' = m, n' = n-1, \\ \mu_2, & \text{если } m > 0, m' = m-1, n' = n-1, \\ n\tau, & \text{если } m' = m = 0, n' = n-1, \\ \nu, & \text{если } m' = m+S-s, n' = n. \end{cases} \quad (1)$$

Здесь и в дальнейшем приняты следующие обозначения: $\mu_k = \mu\sigma_k, k = 1, 2$.

Во втором случае возможные выходы из состояния (m, n) связаны лишь с наступлением событий типа (i) и (ii), поскольку в таких состояниях невозможен уход заявок из очереди из-за нетерпеливости, а также не осуществляется пополнение запасов. При этом если наступает событие типа (i), то независимо от типа заявки она принимается системой при наличии хотя бы одного свободного места в очереди с вероятностью 1, иными словами, следующим состоянием системы будет $(m, n+1)$ и интенсивность такого перехода равна λ , где $\lambda = \lambda_1 + \lambda_2$. При наступлении события типа (ii) переходы между состояниями определяются аналогично первому случаю.

Таким образом, в случае $m \geq s$ элементы ПМ данной двумерной ЦМ определяются следующим образом (см. рис. 1):

$$q((m, n), (m', n')) = \begin{cases} \lambda, & \text{если } m' = m, n' = n + 1, \\ \mu_1, & \text{если } m' = m, n' = n - 1, \\ \mu_2, & \text{если } m' = m - 1, n' = n - 1. \end{cases} \quad (2)$$

Из соотношений (1) и (2) получаем, что состояния исследуемой конечномерной ЦМ сообщаются одно с другим, т.е. в ней существует стационарный режим.

Стационарную вероятность состояния $(m, n) \in E$ обозначим $p(m, n)$. Эти вероятности удовлетворяют системе уравнений равновесия (СУР), которая составляется на основе (1) и (2) и представляет собой систему линейных алгебраических уравнений размерности $(S+1)(N+1)$. Явный вид этой СУР достаточно очевиден [20], поэтому она здесь не приводится.

После нахождения стационарных вероятностей состояний можно вычислить усредненные характеристики исследуемой QIS. Действительно, анализ работы изучаемой системы показывает, что искомые характеристики определяются следующим образом [20]:

$$S_{av} = \sum_{m=1}^S m \sum_{n=0}^N p(m, n); \quad (3)$$

$$RR = \mu_2 \sum_{n=1}^N p(s+1, n); \quad (4)$$

$$PB_1 = \sum_{m=s}^S p(m, N) + \sum_{m=0}^{s-1} \sum_{n=0}^N p(m, n); \quad (5)$$

$$PB_2 = \sum_{m=0}^S p(m, N) + \sum_{n=1}^N p(0, n) \cdot \frac{n\tau}{\lambda_1 + n\tau}. \quad (6)$$

Отметим, что при использовании политик (s, S) и $(S-1, S)$ объем заказываемого запаса всегда является постоянным, т.е. при использовании политики (s, S) объем заказа равен $S-s$, а при использовании политики $(S-1, S)$ эта величина равна единице. Вместе с тем, в отличие от них, при использовании $(m, S-m)$ -политики объем поставляемого заказа — переменная величина, и поэтому для нее необходимо ввести дополнительную характеристику — средний объем заказа, V_{av} . Он определяется так:

$$V_{av} = \sum_{m=S-s}^S m \sum_{n=0}^N p(S-m, n). \quad (7)$$

При использовании $(S-1, S)$ -политики элементы ПМ соответствующей ЦМ определяется аналогично соотношениям (1) и (2), при этом имеются следующие отличия: в обоих случаях $m < s$ и $m \geq s$ в момент поступления заказа происходит переход из состояния (m, n) в состояние $(m+1, n)$ с интенсивностью $(S-m)v$.

При использовании $(m, S-m)$ -политики элементы ПМ соответствующей ЦМ определяются аналогично соотношениям (1) и (2), при этом единственное отличие состоит в том, что в случаях $m < s$ (см. формулы (1)) в момент поступления заказа происходит переход из состояния (m, n) в состояние (S, n) с интенсивностью v .

Отметим, что и при использовании последних двух политик пополнения запасов состояния соответствующих ЦМ сообщаются между собой, поэтому в них также существует стационарный режим. Аналогично [20] можно составить соответствующие СУР размерности $(S+1)(S+1)(N+1)$ для вероятностей состояний. При использовании $(m, S-m)$ -политики характеристики системы также определяются из (3)–(6), а для $(S-1, S)$ -политики средняя интенсивность заказов вычисляется из следующей формулы:

$$RR = \mu_2 \left(1 - \sum_{m=0}^S p(m, 0) \right). \quad (8)$$

Следовательно, задача заключается в нахождении решений соответствующих СУР. Для моделей умеренной размерности эта задача легко реализуется с помощью известных пакетов программ. Вместе с тем для моделей большой размерности эта задача нетривиальна.

Ниже предложен унифицированный алгоритм решения задачи нахождения вероятностей состояний системы при использовании всех трех политик пополнения запасов. Этот алгоритм может корректно применяться для моделей QIS, в которых интенсивность поступления заявок существенным образом превышает интенсивности пополнения запасов.

Сначала рассмотрим задачу для модели с (s, S) -политикой пополнения запасов. Имеем следующее разбиение исходного ПС:

$$E = \bigcup_{m=0}^S E_m, E_m \cap E_{m'} = \emptyset, m \neq m', \quad (9)$$

где $E_m = \{(m, n) \in E : n = 0, 1, \dots, N\}$, $m = 0, 1, \dots, S$.

Микросостояния $(m, n) \in E$ объединяются в одно укрупненное состояние $\langle m \rangle$, и на исходном ПС составляется функция укрупнения $U : E \rightarrow \Omega$, где $U(m, n) = \langle m \rangle$, $\Omega = \{\langle m \rangle : m = 0, 1, \dots, S\}$.

На основе разбиения (9) составляется $S+1$ расщепленных моделей с ПС E_m , $m = 0, 1, \dots, S$, где в каждой модели учитываются лишь переходы между состояниями, входящими в соответствующее ПС. При этом интенсивности указанных переходов определяются из соотношений (1) и (2).

Вероятность состояния $(m, n) \in E_m$ внутри расщепленной модели с ПС E_m обозначим $\rho_m(n)$. Тогда согласно (1) и (2) получаем, что вероятности состояний внутри расщепленной модели с ПС E_0 совпадают с соответствующими вероятностями состояний системы обслуживания $M/M/N/N$ с нагрузкой $\omega = \lambda_2 \phi_1 / \tau$, т.е.

$$\rho_0(n) = \frac{\omega^n}{n!} \bigg/ \left(\sum_{i=0}^N \frac{\omega^i}{i!} \right), n = 0, 1, \dots, N. \quad (10)$$

Из (1) и (2) находим, что вероятности состояний внутри расщепленных моделей с ПС E_m при $m = 1, \dots, s$ и $m = s+1, \dots, S$ совпадают с вероятностями состояний системы обслуживания $M/M/1/N$ с нагрузкой $\psi_1 = \lambda_2 / \mu_1$ и $\psi_2 = \lambda / \mu_1$ соответственно, т.е.:

• в случаях $m = 1, \dots, s$:

$$\rho_m(n) = \begin{cases} \frac{(1-\psi_1)\psi_1^n}{1-\psi_1^{N+1}}, & \text{если } \psi_1 \neq 1, \\ \frac{1}{N+1}, & \text{если } \psi_1 = 1. \end{cases} \quad (11)$$

• в случаях $m = s+1, \dots, S$:

$$\rho_m(n) = \begin{cases} \frac{(1-\psi_2)\psi_2^n}{1-\psi_2^{N+1}}, & \text{если } \psi_2 \neq 1, \\ \frac{1}{N+1}, & \text{если } \psi_2 = 1. \end{cases} \quad (12)$$

Для нахождения вероятностей укрупненных состояний $\pi(\langle m \rangle), \langle m \rangle \in \Omega$, необходимо вычислить интенсивности переходов между этими состояниями $q(\langle m \rangle, \langle m' \rangle)$. С учетом (1), (2) и (10)–(12) находим, что эти величины определяются так:

$$q(\langle m \rangle, \langle m' \rangle) = \begin{cases} \nu, & \text{если } 0 \leq m \leq s, m' = m + S - s, \\ \mu_2(1 - \rho_m(0)), & \text{если } 1 \leq m \leq S, m' = m - 1, \\ 0 & \text{в остальных случаях.} \end{cases} \quad (13)$$

Из (13) с учетом результатов [3] получаем, что вероятности укрупненных состояний $\pi(\langle m \rangle), \langle m \rangle \in \Omega$, вычисляются следующим образом:

$$\pi(\langle m \rangle) = \begin{cases} \alpha_m \pi(\langle s+1 \rangle), & \text{если } 0 \leq m \leq s, \\ \pi(\langle s+1 \rangle), & \text{если } s+1 \leq m \leq S-s, \\ \beta_m \pi(\langle s+1 \rangle), & \text{если } S-s+1 \leq m \leq S. \end{cases} \quad (14)$$

Здесь и далее приняты следующие обозначения:

$$\alpha_m = \left(\frac{\theta_1}{\nu + \theta_1} \right)^{s+1-m}, \quad \beta_m = \frac{\nu}{\theta_2} \sum_{i=m-S+s}^s \alpha_i; \quad \theta_k = \mu_2 \frac{\psi_k - \psi_k^{N+1}}{1 - \psi_k^{N+1}}, \quad k = 1, 2.$$

Вероятность $\pi(\langle s+1 \rangle)$ находится из условия нормировки, т.е.

$$\pi(\langle s+1 \rangle) = \left(S - 2s + \sum_{m=0}^s \alpha_m + \sum_{m=S-s+1}^S \beta_m \right)^{-1}.$$

С учетом (10)–(12) и (14) вероятности состояний исходной двумерной ЦМ определяются так:

$$p(m, n) \approx \rho_m(n) \pi(\langle m \rangle). \quad (15)$$

После выполнения определенных преобразований получим следующие формулы для вычисления характеристик (3)–(8):

$$S_{av} \approx \sum_{m=1}^S m \pi(\langle m \rangle); \quad (16)$$

$$RR \approx \theta_2 \pi(< s+1 >); \quad (17)$$

$$PB_1 \approx \sum_{m=s}^S \rho_m(N) \pi(< m >) + \sum_{m=0}^{s-1} \pi(< m >); \quad (18)$$

$$PB_2 \approx \sum_{m=0}^S \rho_m(N) \pi(< m >) + \pi(< 0 >) \sum_{n=1}^N \rho_0(n) \cdot \frac{n\tau}{\lambda_1 + n\tau}. \quad (19)$$

Предложенный подход можно применить и при использовании $(S-1, S)$ - и $(m, S-m)$ -политик пополнения запасов. При этом в обеих политиках вероятности состояний внутри расщепленных моделей также определяются с помощью формул (10)–(11), (15) а вероятности укрупненных состояний вычисляются следующим образом.

Для $(S-1, S)$ -политики интенсивности переходов между укрупненными состояниями вычисляются так:

$$q(< m >, < m' >) = \begin{cases} \theta_1, & \text{àñèè } m \leq s, m' = m-1, \\ \theta_2, & \text{àñèè } m > s, m' = m-1, \\ (S-m)v, & \text{àñèè } m' = m+1, \\ 0 & \text{à ñòàèüíüð ñëó÷àýð.} \end{cases} \quad (20)$$

Тогда из (20) получаем, что вероятности укрупненных состояний при использовании $(S-1, S)$ -политики находятся из следующих соотношений:

$$\pi(< m >) = \begin{cases} \frac{S!}{(S-m)!} \left(\frac{v}{\theta_1}\right)^m \pi(< 0 >), & \text{àñèè } 0 \leq m \leq s, \\ \frac{S!}{(S-m)!} \left(\frac{\theta_2}{\theta_1}\right)^s \left(\frac{v}{\theta_2}\right)^m \pi(< 0 >), & \text{àñèè } s+1 \leq m \leq S, \end{cases} \quad (21)$$

$$\text{где } \pi(< 0 >) = \left(\sum_{m=0}^s \frac{S!}{(S-m)!} \left(\frac{v}{\theta_1}\right)^m + \left(\frac{\theta_2}{\theta_1}\right)^s \sum_{m=s+1}^S \frac{S!}{(S-m)!} \left(\frac{v}{\theta_2}\right)^m \right)^{-1}.$$

Далее для этой политики приближенные значения характеристик S_{av} , PB_1 и PB_2 вычисляются из формул (16), (18) и (19) соответственно. Приближенное значение характеристики RR определяется так:

$$RR \approx \mu_2 \left(1 - \sum_{m=0}^S \rho_m(0) \pi(< m >) \right). \quad (22)$$

При использовании $(m, S-m)$ -политики интенсивности переходов между укрупненными состояниями вычисляются так:

$$q(< m >, < m' >) = \begin{cases} \theta_1, & \text{àñèè } m \leq s, m' = m-1, \\ \theta_2, & \text{àñèè } m > s, m' = m-1, \\ v, & \text{àñèè } m \leq s, m' = S, \\ 0 & \text{à ñòàèüíüð ñëó÷àýð.} \end{cases} \quad (23)$$

Следовательно, из (23) получаем, что вероятности укрупненных состояний при использовании $(m, S-m)$ -политики находятся из следующих соотношений:

$$\pi(< m >) = \begin{cases} \frac{v}{\theta_1} \left(1 + \frac{v}{\theta_1}\right)^{m-1} \pi(< 0 >), & \text{если } 1 \leq m \leq s, \\ \frac{v}{\theta_2} \sum_{i=1}^s \left(1 + \frac{v}{\theta_1} \left(1 + \frac{v}{\theta_1}\right)^{i-1}\right) \pi(< 0 >), & \text{если } s+1 \leq m \leq S, \end{cases} \quad (24)$$

$$\text{где } \pi(< 0 >) = \left(1 + \frac{v}{\theta_1} \sum_{m=1}^s \left(1 + \frac{v}{\theta_1}\right)^{m-1} + (S-s) \frac{v}{\theta_2} \sum_{i=1}^s \left(1 + \frac{v}{\theta_1} \left(1 + \frac{v}{\theta_1}\right)^{i-1}\right)\right)^{-1}.$$

Далее для указанных политик приближенные значения характеристик (3)–(6) вычисляются из формул (16)–(19) с учетом (22). Для $(m, S-m)$ -политики средний объем заказов приближенно определяется так:

$$V_{av} \approx \sum_{m=0}^s (S-m) \pi(< m >). \quad (25)$$

Теперь рассмотрим методы решения рассмотренных выше задач для моделей QIS с бесконечной очередью, т.е. предположим, что $N = \infty$. Поскольку разработанный метод подробно описан для моделей QIS с конечной очередью, то ниже приводятся лишь окончательные виды формул для расчета вероятностей состояний и характеристик системы.

Так, при использовании всех политик пополнения запасов вероятности состояний внутри расщепленной модели с ПС E_0 совпадают с соответствующими вероятностями состояний системы обслуживания $M/M/\infty$ с нагрузкой $\omega = \lambda_2 \phi_1 / \tau$, т.е.

$$\rho_0(n) = \frac{\omega^n}{n!} e^{-\omega}, \quad n = 0, 1, \dots$$

Вероятности состояний внутри расщепленных моделей с ПС E_m для $m = 1, \dots, s$ и $m = s+1, \dots, S$ существуют лишь при выполнении условия эргодичности $\psi_2 < 1$ и совпадают с вероятностями состояний системы обслуживания $M/M/1/\infty$ с нагрузкой ψ_1 и ψ_2 соответственно, т.е.:

- в случаях $m = 1, \dots, s$:

$$\rho_m(n) = \psi_1^n (1 - \psi_1), \quad n = 0, 1, \dots;$$

- в случаях $m = s+1, \dots, S$:

$$\rho_m(n) = \psi_2^n (1 - \psi_2), \quad n = 0, 1, \dots$$

При использовании (s, S) - $(S-1, S)$ - и $(m, S-m)$ -политик пополнения запасов вероятности укрупненных состояний находятся из соотношений (14), (21) и (24) соответственно, где θ_k заменяются величинами $\mu_2 \psi_k$, $k = 1, 2$.

Средний уровень запасов, средняя интенсивность заказов (кроме $(S-1, S)$ -политики) и средний объем заказов (для $(m, S-m)$ -политики) определяются из (16), (17) и (25) соответственно.

Средняя интенсивность заказов для $(S-1, S)$ -политики вычисляется следующим образом:

$$RR \approx \mu_2 \left(1 - \psi_1 \sum_{m=1}^s \pi(< m >) - \psi_2 \sum_{m=s+1}^S \pi(< m >) \right).$$

Вероятности потери разнотипных заявок в данной модели для всех политик пополнения запасов вычисляются так:

$$PB_1 \approx \sum_{m=0}^{s-1} \pi(< m >);$$

$$PB_2 \approx \pi(< 0 >) e^{-\omega} \sum_{n=1}^{\infty} \frac{\omega^n}{n!} \cdot \frac{n\tau}{\lambda_1 + n\tau}. \quad (26)$$

Замечание 1. Здесь не удастся получить явную формулу для вероятности поте-

ри заявок второго типа. Вместе с тем бесконечный ряд в формуле (26) сходится, так как соответствующий мажорантный ряд $\sum_{n=1}^{\infty} \frac{\omega^n}{n!}$ сходится.

Поэтому при проведении численных экспериментов верхняя граница этой суммы может быть заменена произвольной достаточно большой величиной.

Численные результаты

Здесь приводится часть результатов вычислительных экспериментов, которые выполнены с применением разработанных алгоритмов. Эти эксперименты сходны с экспериментами, проводимыми в [3] для моделей QIS с другой схемой доступа разнотипных заявок. Поэтому здесь лишь вкратце описываются результаты экспериментов.

Как и в [3], точность разработанных приближенных алгоритмов вычисления вероятностей состояний моделей QIS для всех политик пополнения запасов оценивается с помощью двух норм близости: подобия косинуса $\|N\|_1$ и максимум абсолютных значений разностей $\|N\|_2$.

Для краткости изложения приводятся результаты лишь для модели с конечной очередью и (s, S) -политикой. В табл. 1 все значения норм подобия получены для разных значений S, S, N но при фиксированных значениях $\sigma_1 = 0,6; \varphi_1 = 0,7; v = 2, \tau = 1$. Из таблицы видно, что разрабо-

Таблица 1

(S, S, N)	(λ_1, λ_2)	μ	Значения норм подобия	
			$\ N\ _1$	$\ N\ _2$
(10,2,5)	(55,50)	15	0,97385	0,03177
	(60,55)	20	0,97395	0,03373
	(65,60)	25	0,97494	0,03258
(10,2,10)	(55,50)	15	0,96662	0,04114
	(60,55)	20	0,96287	0,04262
	(65,60)	25	0,95979	0,04480
(10,2,15)	(55,50)	15	0,96491	0,04582
	(60,55)	20	0,95986	0,04785
	(65,60)	25	0,95506	0,04811
(15,5,5)	(55,50)	15	0,95895	0,02811
	(60,55)	20	0,94481	0,03108
	(65,60)	25	0,93238	0,03234
(15,5,10)	(55,50)	15	0,95133	0,03434
	(60,55)	20	0,93354	0,03611
	(65,60)	25	0,91732	0,03818
(15,5,15)	(55,50)	15	0,94986	0,03814
	(60,55)	20	0,93163	0,03996
	(65,60)	25	0,91492	0,04022
(20,7,5)	(55,50)	15	0,96667	0,02188
	(60,55)	20	0,95233	0,02429
	(65,60)	25	0,93785	0,02535
(20,7,10)	(55,50)	15	0,96069	0,02676
	(60,55)	20	0,94339	0,02817
	(65,60)	25	0,92589	0,02986
(20,7,15)	(55,50)	15	0,95947	0,02968
	(60,55)	20	0,94169	0,03123
	(65,60)	25	0,92372	0,03153

танный приближенный алгоритм вычисления вероятностей состояний указанной модели имеет высокую точность (аналогичные результаты получены и для других политик пополнения запасов).

Нами проведены исследования также по установлению точности вычисления характеристик системы (табл. 2). Очевидно, что и в этом случае наблюдается высокая точность разработанных приближенных алгоритмов.

Таблица 2

(S, s, N)	(λ_1, λ_2)	μ	S_{AV}		RR		PB_1		PB_2	
			ТЗ	ПЗ	ТЗ	ПЗ	ТЗ	ПЗ	ТЗ	ПЗ
(10,2,5)	(55,50)	15	5,06673	4,92769	0,73350	0,71244	0,85679	0,91591	0,83275	0,90101
	(60,55)	20	4,94267	4,75574	0,96565	0,93988	0,82962	0,89908	0,79708	0,87960
	(65,60)	25	4,82616	4,63699	1,19029	1,16573	0,80845	0,88520	0,76793	0,86186
(10,2,10)	(55,50)	15	5,25035	4,92764	0,74391	0,71243	0,83494	0,91591	0,81845	0,89756
	(60,55)	20	5,17450	4,75564	0,98689	0,93985	0,80157	0,89908	0,77908	0,87590
	(65,60)	25	5,10166	4,63681	1,22648	1,16568	0,77464	0,88518	0,74622	0,85811
(10,2,15)	(55,50)	15	5,32444	4,92764	0,74678	0,71243	0,81552	0,91591	0,80281	0,89370
	(60,55)	20	5,26971	4,75564	0,99300	0,93985	0,77732	0,89908	0,76007	0,87178
	(65,60)	25	5,21661	4,63681	1,23728	1,16568	0,74608	0,88518	0,72428	0,85395
(15,5,5)	(55,50)	15	9,00647	8,06838	0,59879	0,58953	0,85686	0,92179	0,83526	0,89359
	(60,55)	20	8,84327	7,61139	0,79708	0,77963	0,82901	0,90866	0,79791	0,86720
	(65,60)	25	8,68163	7,26644	0,99389	0,96773	0,80706	0,89846	0,76601	0,84483
(15,5,10)	(55,50)	15	9,23039	8,06832	0,59954	0,58951	0,83925	0,92179	0,82542	0,89238
	(60,55)	20	9,14050	7,61120	0,79911	0,77956	0,80623	0,90865	0,78684	0,86561
	(65,60)	25	9,05070	7,26605	0,99843	0,96760	0,77935	0,89845	0,75413	0,84299
(15,5,15)	(55,50)	15	9,31462	8,06832	0,59970	0,58951	0,82372	0,92179	0,81317	0,89103
	(60,55)	20	9,25277	7,61120	0,79946	0,77956	0,78682	0,90865	0,77229	0,86386
	(65,60)	25	9,19090	7,26605	0,99912	0,96760	0,75645	0,89845	0,73777	0,84099
(20,7,5)	(55,50)	15	12,50611	11,31339	0,46101	0,45803	0,85692	0,92164	0,84025	0,89580
	(60,55)	20	12,34181	10,70099	0,61429	0,60761	0,82834	0,90878	0,80420	0,86902
	(65,60)	25	12,17802	10,21320	0,76716	0,75585	0,80546	0,89906	0,77336	0,84594
(20,7,10)	(55,50)	15	12,73045	11,31333	0,46129	0,45801	0,84338	0,92164	0,83273	0,89527
	(60,55)	20	12,64050	10,70076	0,61494	0,60754	0,81081	0,90878	0,79587	0,86822
	(65,60)	25	12,55049	10,21270	0,76853	0,75571	0,78411	0,89905	0,76465	0,84492
(20,7,15)	(55,50)	15	12,81468	11,31333	0,46137	0,45801	0,83143	0,92164	0,82332	0,89468
	(60,55)	20	12,75284	10,70076	0,61508	0,60754	0,79588	0,90878	0,78469	0,86735
	(65,60)	25	12,69095	10,21270	0,76877	0,75571	0,76649	0,89905	0,75211	0,84383

Ниже приводятся графики, показывающие поведение характеристик системы относительно изменения параметра s при использовании различных политик пополнения запасов. Результаты численных экспериментов для модели с конечной очередью показаны на рис. 1–3, где обозначения \times , \diamond и \circ на кривых относятся к результатам для политики (s, S) , $(m, S - m)$ и $(S - 1, S)$ соответственно.

Исходные параметры модели выбраны следующим образом:

$$S = 15, N = 10, \lambda_1 = 60, \lambda_2 = 50, \mu = 30, \sigma_1 = 0,6, \phi_1 = 0,7, \nu = 3, \tau = 1.$$

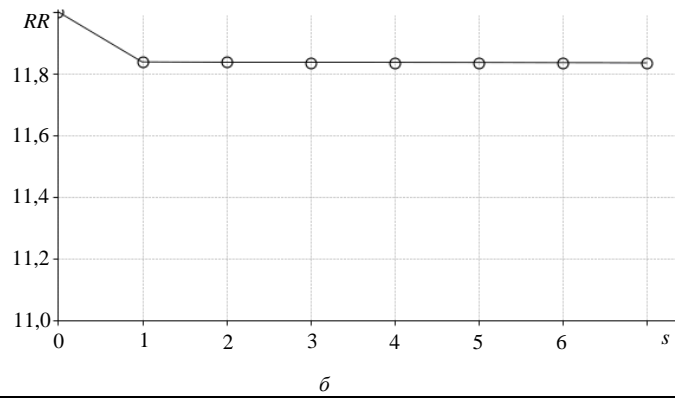
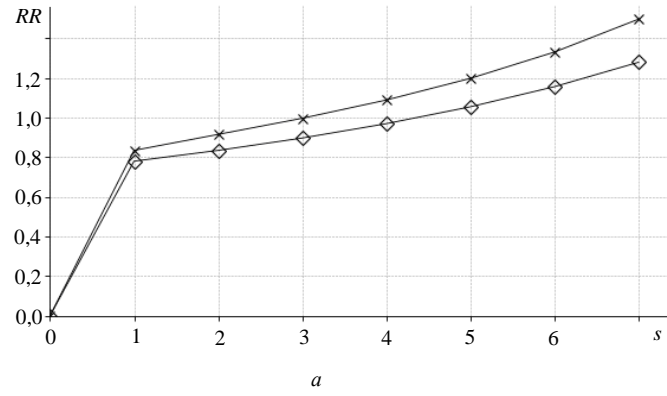


Рис. 2

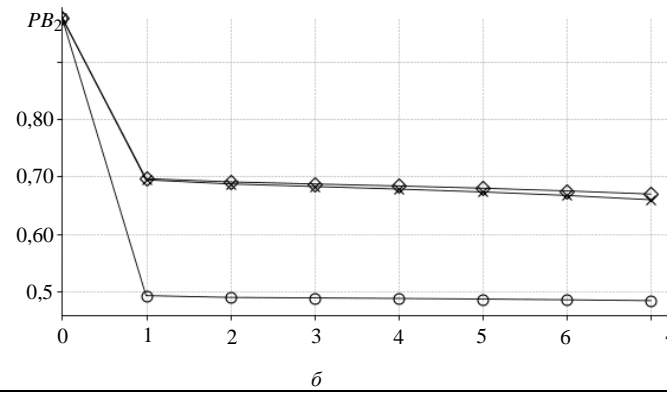
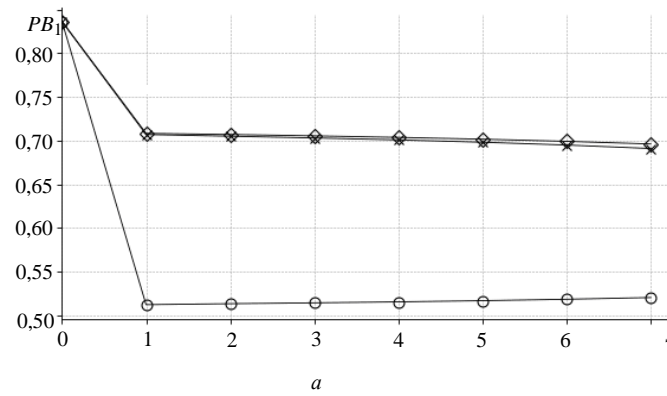


Рис. 3

Из рис. 1 видно, что функция S_{av} является возрастающей при всех трех политиках пополнения запасов, при этом при использовании политик (s, S) и $(m, S - m)$ средние уровни запасов близки один другому, и с точки зрения поддержания невысокого уровня запасов предпочтительной является $(S - 1, S)$ -политика. На первый взгляд при использовании $(S - 1, S)$ -политики средний уровень запасов не должен зависеть от параметра s , поскольку при такой политике заказ делается после каждого отпуска запаса, т.е. не зависит от критического уровня запасов s . Однако и при такой политике на значения этой характеристики влияет принятая схема доступа разнотипных заявок, которая зависит от параметра s .

Функция RR является возрастающей для политик (s, S) и $(m, S - m)$ (см. рис. 2, а), при этом значение этой функции при использовании (s, S) -политики оказывается чуть больше, чем при использовании $(m, S - m)$ -политики. Вместе с тем эта функция является невозрастающей для $(S - 1, S)$ -политики (см. рис. 2, б), при этом ее значения оказываются существенно большими, чем при использовании остальных двух политик. Этого следовало ожидать, так как при использовании $(S - 1, S)$ -политики заказы делаются каждый раз после отпуска запасов по заявкам.

Интересным является тот результат, что вероятности потери разнотипных вызовов (функции PB_1 и PB_2) почти совпадают при использовании политик (s, S) и $(m, S - m)$, при этом с точки зрения уменьшения вероятности предпочтительной является $(S - 1, S)$ -политика (см. рис. 3). Как и следовало ожидать, вероятности потери приоритетных заявок оказываются меньшими, чем вероятность потери обычных заявок.

Теперь рассмотрим задачу максимизации прибыли системы с бесконечной очередью за счет выбора надлежащих значений параметра s при использовании различных политик пополнения запасов.

Доходы системы (RV) определяются так:

$$RV(s) = \sum_{k=1}^2 \lambda_k (1 - PB_k(s)) C_{rev}^k PS_k(s), \quad (27)$$

где $PS_k(s)$ оценивает вероятность того, что обслуженная заявка k -го типа покупает запас; C_{rev}^k — доходы системы из-за продажи единицы запаса по заявкам k -го типа, $k = 1, 2$. Следует ожидать, что доходы от обслуживания приоритетных заявок намного больше, чем от обслуживания обычных, т.е. $C_{rev}^2 > C_{rev}^1$.

Величины $PS_k(s)$, $k = 1, 2$, определяются следующим образом:

$$PS_1(s) = \frac{\mu_2}{\lambda + \mu_1 + \mu_2} \sum_{m=s+1}^S (1 - \rho_m(0)) \pi(< m >);$$

$$PS_2(s) = \frac{\mu_2}{\lambda_2 + \mu_1 + \mu_2} \sum_{m=1}^S (1 - \rho_m(0)) \pi(< m >).$$

Замечание 2. Здесь и далее в обозначениях характеристик системы и функционалов в скобках указан оптимизируемый параметр s .

Суммарные штрафы (TC) в системе определяются аналогично [3]:

$$TC(s) = (K + c_r V_{av}(s))RR(s) + c_h S_{av}(s) + c_l^1 \lambda_1 PB_1(s) + c_l^2 \lambda_2 PB_2(s), \quad (28)$$

где K — фиксированная цена одного заказа; c_r — цена единицы объема заказа; c_h — цена хранения единицы объема запасов за единицу времени; c_l^k — штраф за потерю одной заявки k -го типа, $k=1, 2$. Считается, что штрафы за потери приоритетных заявок намного больше, чем обычных, т.е. $c_l^2 \gg c_l^1$.

Замечание 3. При использовании (s, S) - и $(S-1, S)$ -политики суммарные штрафы в системе определяются согласно (28), где V_{av} заменяется на $S-s$ и 1 соответственно.

Задача оптимизации ставится следующим образом: требуется максимизировать прибыль (PT) системы за счет выбора надлежащих значений критического уровня запасов, т.е. решить следующую задачу:

$$s^* = \arg \max_s PT(s), \quad (29)$$

где $PT(s) = RV(s) - TC(s)$.

Задача (29) имеет решение, так как множество допустимых решений $\{0 \leq s < S/2\}$ конечно и дискретно.

В табл. 3 приводятся результаты решения задачи (29) для трех политик пополнения запасов, где PT^* обозначает максимальное значение функционала $PT(s)$.

Таблица 3

S	(λ_1, λ_2)	(s, S) - политика		$(S-1, S)$ - политика		$(m, S-m)$ - политика	
		s^*	PT^*	s^*	PT^*	s^*	PT^*
10	(10,5)	0	0,53	0	-6,41	1	-0,28
	(15,5)	0	2,29	0	-2,09	1	0,08
	(20,10)	0	4,88	0	4,62	1	-1,58
15	(10,5)	0	0,48	0	-6,9	1	0,16
	(15,5)	0	2,32	0	-2,45	1	0,98
	(20,10)	0	5,13	0	5,67	1	0,38
20	(10,5)	0	0,33	0	-7,4	1	0,31
	(15,5)	0	2,21	0	-2,95	1	1,42
	(20,10)	0	5,13	0	5,22	1	1,57

Фиксированные параметры системы выбирались так:

$$\mu = 60, \sigma_1 = 0,6, \varphi_1 = 0,7, v = 3, \tau = 1.$$

Коэффициенты в функционалах (27) и (28) определяются следующим образом:

$$C_{rev}^1 = 2; C_{rev}^2 = 4; K = 0,5; c_r = 0,1; c_h = 0,1; c_l^1 = 0,5; c_l^2 = 1.$$

Из табл. 3 видно, что во многих случаях при использовании $(S-1, S)$ -политики суммарные штрафы оказываются большими, чем доходы системы (система имеет «отрицательную» прибыль; это объясняется тем, что при использовании этой политики интенсивность заказов системы существенно больше, чем при использовании остальных двух политик (см. рис. 2). Интересен результат, что в

очень широком диапазоне изменения параметров системы оптимальное значение критического уровня запасов для всех политик пополнения запасов — постоянная величина. Следует также отметить, что при одинаковых исходных данных максимальные значения $PT(s)$ существенно отличаются в различных политиках пополнения запасов.

Заключение

В работе предложены модели системы обслуживания-запасания с двумя типами заявок, где часть заявок после завершения обслуживания не покупает запасы. Разнотипные заявки могут образовать общую очередь конечной или бесконечной длины, и время их обслуживания — положительная случайная величина. Заявки высокого приоритета принимаются при наличии хотя бы одного свободного места в очереди, в то время как заявки низкого приоритета принимаются лишь тогда, когда уровень запасов системы выше определенной критической величины. В системе принята одна из трех политик пополнения запасов: (s, S) , $(S-1, S)$ и $(m, S-m)$. В случаях отсутствия запасов заявки в очереди являются нетерпеливыми. Для всех политик время выполнения заказов системы для пополнения запасов имеет экспоненциальные распределения с конечным средним. Разработан унифицированный приближенный метод для вычисления характеристик изучаемых моделей и решены задачи оптимизации работы изучаемых систем.

А.З. Меліков, Л.А. Пономаренко, І.А. Алієв

АНАЛІЗ І ОПТИМІЗАЦІЯ МОДЕЛЕЙ СИСТЕМ ОБСЛУГОВУВАННЯ-ЗАПАСАННЯ З ДВОМА ТИПАМИ ЗАЯВОК

Запропоновано моделі систем обслуговування-запасання з двома типами заявок, де заявки низького пріоритету приймаються лише тоді, коли поточний рівень запасів вище певного критичного значення. Передбачається, що в системі може бути використана одна із наступних політик поповнення запасів: політика двох рівнів, політика змінного об'єму замовлень і політика, згідно з якою замовлення одиничного розміру виконуються після кожної видачі запасів. Розроблено точний і наближений методи розрахунку характеристик системи. Наведено результати числових експериментів щодо розрахунку моделей і вирішено задачу їх оптимізації.

A.Z. Melikov, L.A. Ponomarenko, I.A. Aliyev

ANALYSIS AND OPTIMIZATION OF MODELS OF QUEUEING-INVENTORY SYSTEMS WITH TWO KINDS OF CUSTOMERS

Models of queuing-inventory systems with two kinds of customers are proposed. It is assumed that customers of low priorities should be accepted if current inventory level is higher than a given critical threshold value. In system it might be used one of the following policies: two level policy; policy of variable order size and policy in which inventory of unit size is ordered after completion of service of each customer. Exact and approximate methods to calculate the systems characteristics are proposed. Results of numerical experiments related to calculation and optimization of systems characteristics are demonstrated.

1. Schwarz M., Sauer C., Daduna H., Kulik R., Szekli R. M/M/1 queuing systems with inventory // Queuing Systems. — 2006. — **54**, N 1. — P. 55–78.
2. Krishnamoorthy A., Lakshmy B., Manikandan R. A survey on inventory models with positive service time // OPSEARCH. — 2011. — **48**. — P. 153–169.
3. Melikov A.Z., L.A.Ponomarenko, I.A.Aliyev. Markov models of systems with two types of customers and different replenishment policies // Cybernetics and Systems Analysis. — 2018. — **54**, N 6. — P. 900–917.
4. Kaplan A. Stock rationing // Management Science. — 2007. — **15**, N 5. — P. 260–267.
5. Vienott A.F. Optimal policy in a dynamic, single product, non-stationary inventory model with several demand classes // Operations Research. — 1965. — **13**, N 5. — P. 761–778.
6. Evans R.V. Sales and restocking policies in a single item inventory system // Management Science. — 1968. — **14**, N 7. — P. 463–472.
7. Topkis D.M. Optimal ordering and rationing policies in a non-stationary dynamic inventory model with n demand classes // Ibid. — 1968. — **15**, N 3. — P. 160–176.
8. Kleijn M.J., Dekker R. An overview of inventory systems with several demand classes // Lecture Notes in Economics and Mathematical Systems. — 1999. — **480**. — P. 253–265.
9. Nahmias S., Demmy W.S. Operating characteristics of an inventory system with rationing // Management Science. — 1981. — **27**, N 11. — P. 1236–1245.
10. Moon I., Kang S. Rationing policies for some inventory systems // Journal of the Operational Research Society. — 1998. — **49**. — P. 509–518.
11. Arsalan H., Graves S.C., Roemer T.A. A single-product inventory model for multiple demand classes // Management Science. — 2007. — **53**, N 9. — P. 1486–1550.
12. Sivakumar B., Arivarignan G. A modified lost sales inventory system with two types of customers // Quality Technology and Quantitative Management. — 2008. — **5**, N 4. — P. 339–349.
13. Deshpande V., Cohen M.A., Donohue K. A threshold inventory rationing policy for service-differentiated demand classes // Management Science. — 2003. — **49**, N 6. — P. 683–703.
14. Isotupa K.P.S. An (S, Q) Markovian inventory system with lost sales and two demand classes // Mathematical and Computer Modelling. — 2006. — **43**. — P. 687–694.
15. Isotupa K.P.S. An (S, Q) inventory system with two demand classes of customers // International Journal of Operational Research. — 2011. — **12**, N 1. — P. 12–19.
16. Isotupa K.P.S. Cost analysis of an $(S-1, S)$ inventory system with two demand classes and rationing // Annals of Operations Research. — 2015. — **233**. — P. 411–421.
17. Karthick T., Sivakumar B., Arivarignan G. An inventory system with two types of customers and retrial demands // International Journal of Systems Science: Operations & Logistics. — 2015. — **2**, N 2. — P. 90–112.
18. Melikov A.Z., Ponomarenko L.A., Bagirova S.A. Markov models of queuing-inventory systems with variable order size // Cybernetics and Systems Analysis. — 2017. — **53**, N 3. — P. 373–386.
19. Krishnamoorthy A., Manikandan R., Lakshmy B. Revisit to queuing-inventory system with positive service time // Annals of Operations Research. — 2015. — **233**, N 5. — P. 221–236.
20. Алиев И.А. Об одной модели обслуживания-запасания с двумя типами заявок // Вестник Бакинского университета. — 2017. — № 4. — С. 105–110.

Получено 21.06.2018

Статья представлена к публикации членом редколлегии акад. НАН Украины Чикрием А.А.