Математическая Теория Игр и её Приложения, т.11, в.3, с. 53–76

УДК 519.25, 004.75, 519.216.5
ББК 22.18

СИСТЕМА ОБСЛУЖИВАНИЯ-ЗАПАСАНИЯ ТИПА $M/M/1$ С КАНИКУЛАМИ, ПРЕРЫВАНИЯМИ КАНИКУЛ И ПОТЕРЯННЫМИ ПРОДАЖАМИ

Маникандан Рангасвами
Центральный университет штата Керала
Математический факультет
Касарагод, Индия
e-mail: mani@cukerala.ac.in

Саджив С. Наир
Государственный инженерный колледж
Математический факультет
Триссур, Индия
e-mail: sajeev.snair@rediffmail.com

Рассматривается односерверная система обслуживания-запасания в условиях стратегии пополнения (s, Q) с каникулами, прерываниями каникул и потерянными продажами. Во время каникул сервер предоставляет обслуживание с меньшей скоростью, чем в обычном режиме обслуживания. Если в режиме каникул некоторая заявка обслуживается сервером, то сервер продолжает и завершает ее обслуживание в этом же режиме. После этого он переключается на обычный режим обслуживания при условии, что в очереди находится хотя бы одна заявка. Если в очереди на данный момент времени не находится ни одной заявки, то сервер переключается на режим каникул. Будем также полагать, что если на момент завершения обслуживания в период каникул в системе находятся заявки, то сервер возвращается в обычный рабочий режим, иначе – остается в режиме каникул. Для системы с бесконечной вместимостью получено условие стационарности системы, а также вычислено значение вектора стационарного распределения. Приведены оценки различных критериев производительности. Кроме того, произведен анализ периода загрузки и получено стационарное распределение времени ожидания в очереди. Приведены численные иллюстрации производительности системы, а также предложена задача оптимизации.

Ключевые слова: обслуживание-запасание, положительное время обслуживания, каникулы сервера, утерянные продажи.

Поступила в редакцию: 19.07.18 *После доработки*: 09.09.19 *Принята к публикации*: 30.09.19

1. Введение

Вплоть до начала 90-х гг. прошлого века полагалось, что временем обслуживания в системе управления запасами можно пренебречь. Однако данное предположение плохо соотносится с действительностью. Работы по управлению запасами с положительным временем обслуживания были впервые представлены в литературе Меликовым и Молчановым [15], а также Сигманом и Симхи-Леви [27]. Исследования в данном направлении получили продолжение лишь к началу настоящего века. Здесь, наряду с прочими, следует отметить работы Шварца и др. [23, 24]; Баека и Муна [1, 2], Баека и др. [3]; Сафари и др. [22] (для произвольно распределенного времени пополнения запаса); Кришнамурти и др. [8, 9]; Кренцлера и Дадуны [7], Шахин и др. [25]. В одной из недавних работ авторы Шахин и Кришнамурти [26] рассматривают модель обслуживания-запасания с повторными вызовами и получают стационарное распределение в мультипликативном виде. В другой работе авторы Меликов и Шахмалиев [17] рассматривают односерверную систему обслуживания с портящимися запасами и повторными вызовами с орбиты неограниченной вместимости, решая задачу минимизации стоимости при выборе критического уровня запаса. В работе Кришнамурти и др. [10] представлен обзор литературы по теме обслуживания-запасания, объемлющий работы вплоть до 2009 г. Подробный обзор наработок представлен также в работе Кришнамурти и др. [11].

Существует очень мало исследований на тему каникул сервера в системах обслуживания-запасания. Вишванат и др. [19] исследуют модель запасания одного товара с положительным временем пополнения запаса, где сервер уходит в каникулы, когда товара не остается и очередь заявок пуста. Далее, Кришнамурти и Вишванат [9] исследуют модель производства-запасания с каникулами сервера, полагая, что входной поток заявок является марковским, а время обслуживания каждой заявки имеет распределение фазового типа. Работы Шивакумара [28], Джаярамана и др. [4], Падмавати и др. [21], а также недавняя статья Лакшми и Соуджанья [12] – примеры исследований систем обслуживания-запасания со множественными каникулами сервера. В еще более новой статье авторы Падмавати и Шивакумар [20] исследуют односерверную систему обслуживания-запасания с дискретным временем, отложенными запросами и каникулами сервера, выделяя три типа стратегий каникул, такие как модифицированные множественные каникулы (случайное время бездействия перед каждыми каникулами), единственные каникулы и множественные каникулы (следующие один за другим); далее, они численно исследуют систему для определения лучшей среди данных стратегий.

Работ по системам обслуживания-запасания с портящимися товарами и каникулами сервера также немного. Но в последние годы исследователи стали уделять этой области больше внимания. Так, Меликов и др. [16] рассматривают систему обслуживания с портящимися запасами и нетерпеливыми заявками; они также выделяют три типа стратегий каникул: рабочие, ранние и задерживающие. Далее авторы разрабатывают метод вычисления характеристик системы для модели. Королюк и др. [5, 6] рассматривают системы обслуживания-запасания с портящимися товарами и каникулами сервера, разрабатывают метод асимптотического системного анализа, точные и приблизительные методы вычисления характеристик системы.

В данной работе мы рассматриваем односерверную систему обслуживания-запасания при политике пополнения (s, Q) с рабочими каникулами и прерываниями каникул. В течение рабочих каникул сервер предоставляет обслуживание со сниженной интенсивностью. Мы полагаем, что сервер может вернуться из режима каникул в нормальный рабочий режим по завершении первого обслуживания в режиме каникул при условии, что в очереди находится хотя бы одна заявка и уровень запаса положителен к моменту окончания обслуживания, даже если каникулы еще не закончены (см. Ли и Тянь [14]). В противном случае он остается в режиме каникул. Для случая неограниченной вместимости системы получено условие стационарности системы, а также вычисляется стационарное распределение. Отметим, что получить стационарное распределение в мультипликативном виде не удалось ввиду сильной взаимозависимости таких компонентов системы, как число заявок, статус сервера и число товаров в запасе. Поэтому мы использовали алгоритмический подход для вычисления вектора стационарного распределения. Мы также приводим анализ периода загрузки и получаем некоторые ключевые характеристики производительности, такие как стационарное распределение времени ожидания в очереди и среднее время ожидания клиента.

Работа имеет следующую структуру: в разделе 2 приведена математическая модель системы обслуживания-запасания типа $M/M/1$. В разделе 3 представлены вычисления стационарного распределения, анализ периода загрузки, получено стационарное распределение времени ожидания в очереди, среднее время ожидания заявки и некоторые ключевые характеристики производительности. В разделе 4 приведены результаты численных экспериментов для иллюстрации характеристик производительности системы. Наконец, обсуждается задача оптимизации для поиска оптимального значения минимального уровня для запроса пополнения запаса и величины пополнения, а также соответственная минимальная стоимость, получаемая методом глобального поиска.

Обозначения:

В статье используются следующие обозначения и сокращения.
$\mathcal{N}(t)$: число заявок в системе в момент времени t.
$\mathcal{I}(t)$: уровень запаса в системе в момент времени t.
$\mathcal{S}(t)$: статус сервера – каникулы или обычный режим в момент времени t.

Таким образом, $\mathcal{S}(t) = \begin{cases} 0, & \text{сервер в режиме каникул в момент } t, \\ 1, & \text{сервер в обычном режиме в момент } t. \end{cases}$
$\boldsymbol{e} : (1, 1, ..., 1)'$ – вектор-столбец необходимой размерности, состоящий из единиц.
CTMC: Марковская цепь в непрерывном времени.
QBD: обобщенный процесс рождения и гибели.
LIQBD: обобщенный процесс рождения и гибели, не зависящий от уровня.

2. Математическая модель

Рассмотрим односерверную систему обслуживания-запасания с рабочими каникулами и прерываниями каникул. Сервер входит в режим каникул только при отсутствии заявок в системе, но не при исчерпании уровня запаса в момент завершения обслуживания. Мы полагаем, что если в системе остаются заявки в момент окончания обслуживания одной из них в период рабочих каникул, то сервер возвращается в обычный режим работы. С другой стороны, если в системе нет заявок на момент окончания обслуживания в период каникул, то сервер остается в режиме каникул. Случайная продолжительность каникул распределена экспоненциально с параметром θ.

Заявки поступают на единственный сервер в соответствии с пуассоновским процессом интенсивности λ. Внешним заявкам не разрешается войти в систему, если уровень запаса равен нулю. В таком случае заявки покидают систему, не взяв товар из запаса, то есть теряются. Будем называть их *потерянными продажами.* Время обслуживания подчиняется экспоненциальному распределению с параметром μ_v в период каникул и μ_b – в период обычной работы. Полагаем, что даже если режим каникул заканчивается в то время как некоторая заявка обслуживается, то переключение в обычный режим производится только начиная с обслуживания следующей заявки, если таковая ждет в очереди. Пополнение запаса управляется политикой (s, Q). Здесь s – минимальный уровень запаса, $Q = S - s$ — объем пополнения, S – максимальный объем запаса. Будем полагать $(S > 2s)$, чтобы избежать ситуации бесконечного пополнения запаса. Время пополнения запаса распределено экспоненциально с интенсивностью β. Тогда $\{\mathcal{X}(t) | t \geq 0\} = \{(\mathcal{N}(t), \mathcal{S}(t), \mathcal{I}(t)) | t \geq 0\}$ —

CTMC с пространством состояний Ω, заданным выражением

$$\Omega = \bigcup_{i=0}^{\infty} \mathcal{L}(i)$$

где пространство состояний CTMC разбито на уровни $\mathcal{L}(i)$, определяемые как

$$\mathcal{L}(0) = \{(0,0,0),(0,0,1),\ldots,(0,0,Q+s)\},$$

$\mathcal{L}(i) = \{(i,0,0),(i,0,1),\ldots,(i,0,Q+s),(i,1,1),\ldots,(i,1,Q+s)\}$, для $i \in \mathbb{N}$.

Опишем переходы между состояниями марковской цепи:

(a) Переходы, возникающие вследствие поступления заявок:

$(i,0,j) \to (i+1,0,j)$: интенсивность λ, для $i \in \mathbb{N}_0$; $j = 1,2,\ldots,S$.

$(i,1,j) \to (i+1,1,j)$: интенсивность λ, для $i \in \mathbb{N}_0$; $j = 1,2,\ldots,S$.

(b) Переходы, возникающие вследствие завершения обслуживания во время режима рабочего каникул:

$(1,0,j) \to (0,0,j-1)$: интенсивность μ_v, для $j = 1,2,\ldots,S$.

$(i,0,j) \to (i-1,1,j-1)$: интенсивность μ_v, для $i = 2,3,\ldots$; $j = 2,3,\ldots,S$.

$(i,0,1) \to (i-1,0,0)$: интенсивность μ_v, для $i = 2,3,\ldots$

(c) Переходы, возникающие вследствие завершения обслуживания во время обычного режима:

$(1,1,j) \to (0,0,j-1)$: интенсивность μ_b, для $j = 1,2,\ldots,S$.

$(i,1,j) \to (i-1,1,j-1)$: интенсивность μ_b, для $i = 2,3,\ldots$; $j = 2,3,\ldots,S$.

$(i,1,1) \to (i-1,0,0)$: интенсивность μ_b, для $i = 2,3,\dots$

(d) Переходы, возникающие вследствие пополнения запаса:

$(i,0,j) \to (i,0,Q+j)$: интенсивность β, для $i \in \mathbb{N}_0$; $j = 0,1,\dots,s$.

$(i,1,j) \to (i,1,Q+j)$: интенсивность β, для $i \in \mathbb{N}$; $j = 1,2,\dots,s$.

(e) Переходы, возникающие вследствие завершения каникул:

$(i,0,j) \to (i,1,j)$: интенсивность θ, для $i \in \mathbb{N}$; $j = 1,2,\dots,S$.

Прочие пары переходов имеют нулевую интенсивность. Инфинитезимальная образующая матрица $\boldsymbol{\mathcal{Q}}$ данной СТМС имеет блочную структуру:

$$\boldsymbol{\mathcal{Q}} = \begin{bmatrix} B_0 & B_1 & & & & \\ B_2 & A_1 & A_0 & & & \\ & A_2 & A_1 & A_0 & & \\ & & A_2 & A_1 & A_0 & \dots \\ & & & \ddots & \ddots & \ddots \end{bmatrix},$$

где B_0 суть квадратная матрица размерности $S+1$, содержащая интенсивности переходв во множестве состояний $\mathcal{L}(0)$; B_1 и A_0 соответствуют интенсивностям переходов с уровня $\mathcal{L}(i)$ на уровень $\mathcal{L}(i+1)$ при $i \in \mathbb{N}_0$, размерности матриц $(S+1)\times(2S+1)$ и $(2S+1)\times(2S+1)$, соответственно; B_2 имеет размерность $(2S+1)\times(S+1)$ и представляет собой интенсивности переходов с уровня $\mathcal{L}(1)$ на уровень $\mathcal{L}(0)$; A_1 суть квадратная матрица размера $2S+1$, содержащая интенсивности переходов внутри уровня $\mathcal{L}(i)$, для $i \in \mathbb{N}$, наконец, A_2 есть квадратная матрица размера $2S+1$, содержащая интенсивности переходов из состояний уровня $\mathcal{L}(i)$ на состояния уровня $\mathcal{L}(i-1)$, для $i = 2,3,\dots$.

3. Анализ системы

В данном разделе проведен анализ системы обслуживания-запасания в стационарном режиме. Получим условие стационарности модели. Обозначим $A{=}A_0 + A_1 + A_2$. Пусть вектор стационарных вероятностей, соответствующий образующей матрице A обозначен

$$\boldsymbol{\pi} = (\pi_0(0), \pi_0(1), \ldots, \pi_0(S), \pi_1(1), \pi_1(2), \ldots \pi_1(S)).$$

Тогда система для нахождения указанного вектора $\boldsymbol{\pi} A = 0$ и уравнение баланса $\boldsymbol{\pi}\mathbf{e} = 1$ преобразуются к следующему виду:

$$\pi_0(Q) = \frac{\beta}{\theta + \mu_v}\pi_0(0)$$

$$\pi_0(1) = \pi_0(2) = \cdots = \pi_0(Q{-}1) = \pi_0(Q{+}1) = \pi_0(Q{+}2) = \cdots = \pi_0(S) = 0$$

$$\pi_1(1) = \frac{\beta}{\mu_b}\pi_0(0)$$

$$\pi_1(2) = \frac{\beta(\beta + \mu_b)}{\mu_b^2}\pi_0(0)$$

$$\pi_1(3) = \frac{\beta(\beta + \mu_b)^2}{\mu_b^3}\pi_0(0)$$

$$\vdots$$

$$\pi_1(s+1) = \frac{\beta(\beta + \mu_b)^s}{\mu_b^{s+1}}\pi_0(0)$$

$$\pi_1(s+1) = \pi_1(s+2) = \cdots = \pi_1(Q)$$

$$\pi_1(1) + \pi_1(Q+1) = \pi_1(Q)$$

$$\pi_1(2) + \pi_1(Q+2) = \pi_1(Q)$$

$$\pi_1(3) + \pi_1(Q+3) = \pi_1(Q)$$

$$\vdots$$

$$\pi_1(s) + \pi_1(S) = \pi_1(Q)$$

LIQBD с инфинитезимальной образующей $\boldsymbol{\mathcal{Q}}$ является стационарным тогда и только тогда, когда $\boldsymbol{\pi} A_0\mathbf{e} < \boldsymbol{\pi} A_2\mathbf{e}$ (см., напр. [18]). Таким образом, справедливы выкладки

$$\mu_b\left(\pi_1(1) + \pi_1(2) + \cdots + \pi_1(S)\right) > \lambda\left(\pi_1(1) + \pi_1(2) + \cdots + \pi_1(S)\right) + \lambda\pi_0(Q),$$

$$\mu_b Q\pi_1(Q) > \lambda\left(\pi_0(Q) + Q\pi_1(Q)\right),$$

$$\mu_b Q \frac{\beta(\beta+\mu_b)^s}{\mu_b^{s+1}} \pi_0(0) > \lambda \left(\frac{\beta}{\theta+\mu_v} \pi_0(0) + Q \frac{\beta(\beta+\mu_b)^s}{\mu_b^{s+1}} \pi_0(0) \right),$$

$$\lambda < \frac{\mu_b Q \frac{(\beta+\mu_b)^s}{\mu_b^{s+1}}}{\frac{1}{\theta+\mu_v} + Q \frac{(\beta+\mu_b)^s}{\mu_b^{s+1}}},$$

$$\lambda < \frac{\mu_b}{1 + \frac{\mu_b^{s+1}}{(\theta+\mu_v) Q (\beta+\mu_b)^s}}.$$

Мы доказали следующий результат о стационарности системы.

Лемма 3.1. *Марковская цепь в непрерывном времени Ω стационарна тогда и только тогда, когда $\lambda < \frac{\mu_b}{1+\frac{\mu_b^{s+1}}{(\theta+\mu_v)Q(\beta+\mu_b)^s}}$.*

Можно заметить, что данное условие стационарности является менее ограничительным по сравнению с результатом, полученным для классической системы обслуживания типа $M/M/1$ с рабочими каникулами и прерываниями каникул в статье [14].

Далее, вычислим вектор стационарных вероятностей состояний системы $\boldsymbol{x}$, соответствующий образующей матрице $\boldsymbol{\mathcal{Q}}$ при выполнении условия стационарности. Вектор $\boldsymbol{x}$ может быть разделен на подвекторы в соответствии с уровнями:

$$\boldsymbol{x} = (\boldsymbol{x_0}, \boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{...}), \tag{3.1}$$

где подвекторы $\boldsymbol{x}$ далее разделены следующим образом

$$\boldsymbol{x_0} = (x_0(0,0), x_0(0,1), x_0(0,2), ..., x_0(0,S)), \tag{3.2}$$

$$\boldsymbol{x_i} = (x_i(0,0), x_i(0,1), x_i(0,2), ..., x_i(0,S), x_i(1,1), ..., x_i(1,S)), \ i \in \mathbb{N}. \tag{3.3}$$

Пусть $\boldsymbol{x_{i+1}} = \boldsymbol{x_1} R^i$, для $i \in \mathbb{N}$. Тогда из условия $\boldsymbol{x}\boldsymbol{\mathcal{Q}} = 0$ получим

$$\boldsymbol{x_1} A_0 + \boldsymbol{x_2} A_1 + \boldsymbol{x_3} A_2 = 0,$$

$$\boldsymbol{x_1} A_0 + \boldsymbol{x_1} R A_1 + \boldsymbol{x_1} R^2 A_2 = 0,$$

$$\boldsymbol{x_1} \left(A_0 + R A_1 + R^2 A_2 \right) = 0.$$

Выберем R как (минимальное неотрицательное) решение уравнения $A_0 + RA_1 + R^2A_2 = 0$. Кроме того,

$$\boldsymbol{x_0}B_0 + \boldsymbol{x_1}B_2 = 0,$$

$$\boldsymbol{x_0}B_1 + \boldsymbol{x_1}A_1 + \boldsymbol{x_2}A_2 = 0,$$

$$\boldsymbol{x_0}B_1 + \boldsymbol{x_1}(A_1 + RA_2) = 0,$$

$$\boldsymbol{x_1} = -\boldsymbol{x_0}B_1(A_1 + RA_2)^{-1} = \boldsymbol{x_0}V, \quad \text{где} \quad V = -B_1(A_1 + RA_2)^{-1}.$$

Следовательно, получим $\boldsymbol{x_0}(B_0 + VB_2) = 0$. Пусть $\boldsymbol{x_0}$ есть вектор стационарных вероятностей, соответствующий матрице B_0+VB_2. Тогда $\boldsymbol{x_1} = \boldsymbol{x_0}V$ и $\boldsymbol{x_{i+1}} = \boldsymbol{x_1}R^i$, для $i \in \mathbb{N}$. Тогда вектор стационарных вероятностей можно получить, разделив каждый подвектор $\boldsymbol{x_i}$ на нормализующую константу

$$[\boldsymbol{x_0} + \boldsymbol{x_1} + \cdots]\,\mathbf{e} = \left[\boldsymbol{x_0} + \boldsymbol{x_1}(I - R)^{-1}\right]\mathbf{e}.$$

После получения матрицы R, можно получить вектор $\boldsymbol{x}$ с учетом особенностей матричных коэффициентов. Для вычисления матрицы R можно использовать алгоритм логарифмического спуска (см. напр. [13]).

3.1. Анализ периода занятости

Для исследуемой системы определим период занятости как время между приходом клиента в пустую систему с положительным уровнем запаса и ближайшим последующим временем ухода клиента, оставляющим систему пустой (без клиентов). Таким образом, это в точности время первого попадания из состояния $(1, 0, j)$, для $j = 1, 2, \ldots, S$ в состояние $(0, 0, \widetilde{j})$, для $\widetilde{j} = 0, 1, \ldots, S-1$. Цикл занятости при этом равен времени между уходами, оставляющими систему пустой (без клиентов). Поэтому цикл занятости есть время первого возвращения во множества состояний $(0, 0, \widetilde{j})$, где $\widetilde{j} = 0, 1, \ldots, S$ после хотя бы одного перехода в состояния не из этого множества. Прежде, чем продолжить анализ, введем понятие фундаментального периода. Для исследуемого процесса QBD фундаментальный период есть время первого достижения уровня $i-1$ с уровня i, где $i = 2, 3, \ldots$. Граничные случаи $i = 1$ и $i = 0$ рассмотрим отдельно.

Необходимо заметить, что, в связи со структурой исследуемого QBD процесса, распределение времени первого достижения не зависит от i $(i = 2, 3, \dots)$.

Пусть $G_{j\widetilde{j}}(k,\tau)$ есть условная вероятность того, что процесс QBD, стартуя во множестве $(i, 0, j)$, где $j = 1, 2, \dots, S$ и $i = 2, 3, \dots,$ в момент времени 0, впервые достигает одного из состояний $(i-1, 0, \widetilde{j})$, где $\widetilde{j} = 0, 1, \dots, S-1$ ровно за k переходов за время не более τ. Следовательно,

$$G_{j\widetilde{j}}(k,\tau) = P[\tau < \infty : \chi(\tau) = \widetilde{j}/\chi(0) = j]$$

где τ есть время первого достижения уровня $i-1$ с уровня i и χ есть исследуемый QBD процесс. В связи с особенностями структуры $\boldsymbol{\mathcal{Q}}$, вероятность $G_{j\widetilde{j}}(k,\tau)$ не зависит от i. Матрицу с элементами $G_{j\widetilde{j}}(k,\tau)$ обозначим $G(k,\tau)$. Для удобства, запишем результат двойного преобразования элементов матрицы

$$\widetilde{G}_{j\widetilde{j}}(z,\theta) = \sum_{k=1}^{\infty} z^k \int_0^{\infty} e^{-\theta\tau} \mathrm{d}G_{j\widetilde{j}}(k,\tau) \quad ; \quad |z| \le 1, \theta > 0,$$

и матрицу

$$\widetilde{G}(z,\theta) = (\widetilde{G}_{j\widetilde{j}}(z,\theta)).$$

Матрица $\widetilde{G}(z,\theta)$ является единственным решением следующего уравнения (см., напр. [18]):

$$\widetilde{G}(z,\theta) = z(\theta I - A_1)^{-1} A_2 + (\theta I - A_1)^{-1} A_0 \widetilde{G}^2(z,\theta). \tag{3.4}$$

При этом матрица $G = \widetilde{G}(1,0)$ содержит времена первого достижения, за исключением граничных состояний. На основе матрицы R, матрица G может быть вычислена следующим образом (см., напр. [13]):

$$G = -(A_1 + RA_2)^{-1} A_2. \tag{3.5}$$

Воспользуемся методом логарифмического спуска для вычисления G. Для граничных уровней 1 и 0 обозначим $G_{j\widetilde{j}}^{(1,0,j)}(k,\tau)$, где $j = 1, 2, \dots, S$ и $G_{\widetilde{j}\widetilde{j}}^{(0,0,\widetilde{j})}(k,\tau)$, где $\widetilde{j} = 0, 1, \dots, S-1$, условную вероятность первого достижения уровня 0 с уровня 1 и время первого возвращения в 0, соответствено. Далее, по аналогии с (3.4), получим

$$\widetilde{G}^{(1,0,j)}(z,\theta) = z(\theta I - A_1)^{-1}B_2 + (\theta I - A_1)^{-1}A_0\widetilde{G}(z,\theta)\widetilde{G}^{(1,0,j)}(z,\theta), \quad (3.6)$$

где $j = 1, 2, \ldots, S.$, и

$$\widetilde{G}^{(0,0,\widetilde{j})}(z,\theta) = [\lambda/(\lambda+\theta), 0, \widetilde{j}]\widetilde{G}^{(1,0,j)}(z,\theta), \quad (3.7)$$

где $j = 1, 2, \ldots, S$; и $\widetilde{j} = 0, 1, \ldots, S-1$.

Заметим, что $\widetilde{G}^{(1,0,j)}(z,\theta)$, где $j = 1, 2, \ldots, S$, является матрицей размера $(2S+1)\times(S+1)$. Поэтому преобразование Лапласа–Стилтьеса для периода занятости есть первый элемент матрицы $\widetilde{G}^{(1,0,j)}(1,0)$. Для дальнейшего использования обозначим $G_{10} = \widetilde{G}^{(1,0,j)}(1,0)$, $G_{00} = \widetilde{G}^{(0,0,\widetilde{j})}(1,0)$, где $j = 1, 2, \ldots, S$ и $\widetilde{j} = 0, 1, \ldots, S-1$. В связи с положительной возвратностью процесса QBD, матрицы G , G_{10} и G_{00} являются стохастическими. Если $B_0 = (-A_1)^{-1}A_2$ и $B_2 = (-A_1)^{-1}A_0$, то G есть минимальное неотрицательное решение (см., напр. [18]) матричного уравнения $G = B_0 + B_2G^2$. Из уравнений (3.6) и (3.7) получим

$$G_{10} = -(A_1 + A_0G)^{-1}B_2 \quad (3.8)$$

и

$$G_{00} = [1, 0, \widetilde{j}]G_{10}, \quad (3.9)$$

где $\widetilde{j} = 1, 2, \ldots, S-1$, соответственно. Уравнение (3.4) эквивалентно

$$zA_2 - (\theta I - A_1)\widetilde{G}(z,\theta) + A_0\widetilde{G}^2(z,\theta) = 0. \quad (3.10)$$

Пусть

$$D = -\left.\frac{\partial\widetilde{G}(z,\theta)}{\partial\theta}\right|_{z=1,\theta=0}$$

и

$$\widetilde{D} = \left.\frac{\partial\widetilde{G}(z,\theta)}{\partial z}\right|_{z=1,\theta=0}.$$

Дифференцируя (3.10) по θ и по z, а также подставляя $z = 1$ и $\theta = 0$, получим (см. [18])

$$D = -A_1^{-1}G + B_2(GD + DG)$$

и

$$\widetilde{D} = B_0 + B_2(G\widetilde{D} + \widetilde{D}G).$$

При начальном значении $\mathbf{0}$ для D и $\widetilde{D}$, методом последоватльной подстановки получим значения для D и $\widetilde{D}$ из вышеуказанных выражений. По аналогии с (3.6) и (3.7), получим

$$D_{10} = -(A_1 + A_0G)^{-1}(I + A_0D)G_{10},$$

и

$$D_{00} = [1/\lambda, 0, \widetilde{j}]G_{10} + [1, 0, \widetilde{j}]D_{10}, \ \widetilde{j} = 0, 1, \ldots, S-1,$$

где

$$D_{10} = -\left.\frac{\partial \widetilde{G}^{(1,0,j)}(z,\theta)}{\partial \theta}\right|_{z=1,\theta=0}, \text{ for } j = 1, 2, \ldots, S$$

$$D_{00} = -\left.\frac{\partial \widetilde{G}^{(0,0,\widetilde{j})}(z,\theta)}{\partial \theta}\right|_{z=1,\theta=0}, \text{ for } \widetilde{j} = 0, 1, \ldots, S-1.$$

Первые элементы векторов D_{10} и D_{00} есть средняя длина периода занятости и цикла занятости, соответственно. Обозначив

$$\widetilde{D}_{10} = \left.\frac{\partial \widetilde{G}^{(1,0,j)}(z,\theta)}{\partial z}\right|_{z=1,\theta=0},$$

получим из (3.6), что

$$\widetilde{D}_{10} = -(A_1 + A_0G)^{-1}(C_2 + A_0DG_{10}).$$

Первая компонента вектора $\widetilde{D}_{10}$ есть среднее число завершений обслуживания на цикле занятости.

3.2. Стационарное распределение времени ожидания в очереди

В данном разделе получено преобразование Лапласа–Стилтьеса для стационарного времени ожидания в очереди, и среднее время ожидания клиента. Распределение стационарного времени ожидания в системе обслуживания-запасания в общем случае невозможно получить аналитически. Тем не менее, удается получить преобразование Лапласа–Стилтьеса для времени ожидания клиента в очереди и вывести выражение для его среднего значения. Заметим сначала, что приходящий клиент попадает на обслуживание без ожидания с вероятностью $\boldsymbol{z_0} = \boldsymbol{x_0}\mathbf{e}$. С вероятностью $1 - \boldsymbol{z_0}$ клиент вынужден ожидать. Любой такой клиент может быть обслужен только в обычном режиме работы сервера. Если отмеченный клиент будет первым в очереди, его время ожидания будет равно незавершенному времени обслуживания клиента, находящегося на обслуживании. Таким образом, среднее время ожидания такого клиента будет равно $\frac{1}{\mu_v}$ или $\frac{1}{\mu_b}$, в зависимости от типа обслуживания, предоставляемого клиенту, который предшествует отмеченному. Время ожидания может быть рассмотрено как время до поглощения цепи Маркова с разряженной структурой. Пространство состояний такой цепи (которое включает также пришедшего клиента при подсчете) определяется следующим образом:

$$\widetilde{\Omega} = \{*\} \bigcup \{(i,j)|\ i = 2,3,\ldots, j = 1,2,\ldots,S\}. \tag{3.11}$$

Состояние $*$ означает, что отмеченный клиент поступил на обслуживание. Это значит, что $*$ есть объединение состояний $\{(0,j)|\ j = 1,2,\ldots,S\}$. Образующая матрица цепи $\boldsymbol{\mathcal{H}}$ задается следующим образом:

$$\boldsymbol{\mathcal{H}} = \begin{pmatrix} 0 & \boldsymbol{0} & & & \\ \boldsymbol{a} & \widetilde{A}_1 & & & \\ & A_2 & \widetilde{A}_1 & & \\ & & A_2 & \widetilde{A}_1 & \\ & & & \ddots & \ddots \end{pmatrix}, \tag{3.12}$$

где

$$\widetilde{A}_1 = A_1 + \lambda I, \quad \boldsymbol{a} = A_2\mathbf{e}. \tag{3.13}$$

Вектор начального распределения для цепи, определяемой $\boldsymbol{\mathcal{H}}$, обозначается $\boldsymbol{\sigma}$ и задается следующим образом:

$$\boldsymbol{\sigma} = (\boldsymbol{z_0}, \boldsymbol{z_1}, \boldsymbol{z_2}, \boldsymbol{z_3} \cdots),$$

где $\boldsymbol{z_0} = \boldsymbol{x_0}\mathbf{e}$ и $\boldsymbol{z_i}$ суть стационарная вероятность прихода клиента в систему с занятым сервером, работающим в обычном режиме, с числом клиентов в системе (включая вновь вошедшего) равным i.

$$\boldsymbol{z_i} = \left(\mathbf{0}_{1\times S}, \boldsymbol{x_i}(1,1), \boldsymbol{x_i}(1,2)\cdots, \boldsymbol{x_i}(1,S)\right), \quad \text{для} \quad i = 2,3,\dots. \tag{3.14}$$

Пусть $\widetilde{\mathbf{W}}(t),\ t \geq 0$ есть вероятность того, что клиент, входящий в систему при сервере, работающем в обычном режиме, попадет на обслуживание за время, не превышающее t с момента прихода. Получим преобразование Лапласа–Стилтьеса, $\widetilde{w}(\theta)$, для указанной вероятности. Используя особенности структуры матрицы $\boldsymbol{\mathcal{H}}$, можно убедиться, что верна теорема.

Теорема 3.1. *Преобразование Лапласа–Стилтьеса, $\widetilde{w}(\theta)$, величины $\widetilde{\boldsymbol{W}}(t)$, имеет следующий вид:*

$$\widetilde{w}(\theta) = \boldsymbol{z_0} + \sum_{i=0}^{\infty} \boldsymbol{z_i}[(\theta I - \widetilde{A}_1)^{-1}A_2]^i(\theta I - \widetilde{A}_1)^{-1}A_2\boldsymbol{e}. \tag{3.15}$$

Следствие 3.1. *Среднее время ожидания клиента в очереди, μ'_W, может быть получено следующим образом:*

$$\mu'_W = [\boldsymbol{z_2}(I-R)^{-1} - \boldsymbol{z_2}\sum_{k=0}^{\infty} R^k P^{k+1} + \boldsymbol{z_2}(I-R)^{-2}\widetilde{P}](I-P+\widetilde{P})^{-1}(-\widetilde{A}_1)^{-1}\boldsymbol{e}, \tag{3.16}$$

где

$$P = (-\widetilde{A}_1)^{-1}A_2, \quad \widetilde{P} = \boldsymbol{ep}, \tag{3.17}$$

и $\boldsymbol{p}$ есть вектор стационарных вероятностей для матрицы P, т.е.

$$\boldsymbol{p}P = \boldsymbol{p}, \quad \boldsymbol{pe} = 1. \tag{3.18}$$

Замечание: При вычислении среднего времени ожидания μ'_W, необходимо получить значение суммы ряда $\sum_{k=0}^{\infty} R^k P^{k+1}$. Поскольку P есть стохастическая матрица, $\boldsymbol{z_2}\sum_{k=0}^{\infty} R^k P^{k+1}\mathbf{e} = 1 - \boldsymbol{z_0}$, и потому можно воспользоваться частичной суммой из N^* слагаемых, выбирая число слагаемых из условия $|\boldsymbol{z_2}\sum_{k=0}^{N^*} R^k P^{k+1}\mathbf{e} - (1-\boldsymbol{z_0})| < \epsilon$, где ϵ есть наперед заданная малая постоянная величина.

3.3. Прочие характеристики производительности

- Среднее число клиентов в системе,
$$E(N) = \sum_{i=1}^{\infty}\left(\sum_{j=1}^{S} i x_i\,(1,j) + \sum_{j=0}^{S} i x_i\,(0,j)\right).$$

- Средний уровень запаса,
$$E(I) = \sum_{i=0}^{\infty}\sum_{j=1}^{S} j x_i\,(0,j) + \sum_{i=1}^{\infty}\sum_{j=1}^{S} j x_i\,(1,j).$$

- Среднее число пополнений запаса в единицу времени,
$$E(R) = \beta\left(\sum_{j=0}^{s}\left(\sum_{i=0}^{\infty} x_i\,(0,j) + \sum_{i=1}^{\infty} x_i\,(1,j)\right)\right).$$

- Вероятность занятости сервера,
$$P_{busy} = \sum_{i=1}^{\infty}\sum_{j=1}^{S} x_i\,(1,j) + \sum_{i=1}^{\infty}\sum_{j=1}^{S} x_i\,(0,j).$$

- Интенсивность обслуживания в обычном режиме,
$$R_{SNM} = \mu_b\left(\sum_{i=1}^{\infty}\sum_{j=1}^{S} x_i\,(1,j)\right).$$

- Интенсивность обслуживания в режиме каникул,
$$R_{SVM} = \mu_v\left(\sum_{i=1}^{\infty}\sum_{j=1}^{S} x_i\,(0,j)\right).$$

- Средняя интенсивность запросов на пополнение,
$$E(RO) = \mu_b\left(\sum_{i=1}^{\infty} x_i\,(1,s+1)\right) + \mu_v\left(\sum_{i=1}^{\infty} x_i\,(0,s+1)\right).$$

- Интенсивность ухода в режим каникул,
$$\Gamma = \tfrac{\mu_v}{\lambda+\mu_v+\beta}\left(\sum_{i=1}^{\infty} x_i(0,1) + \sum_{j=1}^{s} x_1(1,j)\right)$$
$$+\tfrac{\mu_b}{\lambda+\mu_b+\beta}\left(\sum_{i=1}^{\infty} x_i(1,1) + \sum_{j=1}^{s} x_1(1,j)\right)$$
$$+\tfrac{\mu_v}{\lambda+\mu_v}\left(\sum_{j=s+1}^{S} x_1(1,j)\right) + \tfrac{\mu_b}{\lambda+\mu_b}\left(\sum_{j=s+1}^{S} x_1(1,j)\right).$$

- Интенсивность окончания каникул, $R_{VR} = \theta \left(\sum_{i=1}^{\infty} \sum_{j=1}^{S} x_i (0, j) \right).$

- Средняя интенсивность потери клиентов, $E_{loss} = \lambda \left(\sum_{i=0}^{\infty} x_i (0, 0) \right).$

- Среднее число ожидающих клиентов при непустом запасе, $W_{inv} = \left(\sum_{i=1}^{\infty} \sum_{j=1}^{S} i x_i(0, j) + \sum_{i=1}^{\infty} \sum_{j=1}^{S} i x_i(1, j) \right).$

- Среднее число ожидающих клиентов при пустом запасе, $\widetilde{W_{inv}} = \left(\sum_{i=1}^{\infty} i x_i(0, 0) \right).$

4. Численные эксперименты

В разделе представлены результаты численных экспериментов для расчета характеристик производительности системы при политике пополнения запасов (s, Q); в конце данного раздела сформулирована задача подбора оптимальных параметров (s^*, Q^*) и соответствующей минимальной цены методом глобального поиска. В табл. 1 представлено влияние интенсивности приходов λ на характеристики производительности. Видно, что с увеличением λ, вероятность занятости сервера, среднее число пополнений ресурса и среднее число клиентов в системе возрастают, в то время как средний уровень запаса убывает, как и ожидается. В табл. 2 представлены результаты расчетов избранных характеристик производительности при растущей интенсивности обслуживания (в период рабочих каникул), μ_v. Анализ результатов позволяет сделать вывод, что среднее число пополнений запаса и интенсивность запроса пополнений возрастают, вероятность занятости и среднее число клиентов в системе убывают. Заметим, что подобные эффекты наблюдаются и при изменении параметра μ_b (см. табл. 3); однако средняя интенсивность потери клиентов убывает и средний уровень возрастает с ростом интенсивности обслуживания в обычном режиме, μ_b.

λ	P_{busy}	$E(I)$	$E(R)$	$E(N)$	$E(RO)$	E_{loss}
1	0.0524083	8.66341	0.13196005	0.2580666	0.0677713	0.00035206
1.2	0.065652	8.59858704	0.1547746	0.31453803	0.7752265	0.0003091
1.4	0.079377	8.5305395	0.17724235	0.37214845	0.08633815	0.00073044
1.6	0.093688	8.46489525	0.19800134	0.43188059	0.09408338	0.00096384
1.8	0.109027	8.39699173	0.2181963	0.49292067	0.1009889	0.0018059
2	0.124578	8.32856941	0.23739675	0.55564225	0.10704409	0.00312592

Таблица 1. Влияние интенсивности λ при $(\mu_v, \mu_b, \theta, \beta, s, S) = (3, 10, 2, 3, 5, 12)$ на избранные характеристики производительности.

μ_v	P_{busy}	$E(I)$	$E(R)$	$E(N)$	$E(RO)$	E_{loss}
2	0.142511	8.32898903	0.2297449	0.67837971	0.08146109	0.00325903
2.2	0.138591	8.32946396	0.23133451	0.64966559	0.08712611	0.00309318
2.4	0.134877	8.32984734	0.23283148	0.62336963	0.09248553	0.00295406
2.6	0.131036	8.33016109	0.23424277	0.59919232	0.09756339	0.00283633
2.8	0.127728	8.32806683	0.23621041	0.57621247	0.10247333	0.00327543
3	0.124578	8.32856941	0.23739675	0.55564225	0.10704409	0.00312592

Таблица 2. Влияние интенсивности μ_v при $(\lambda, \mu_b, \theta, \beta, s, S) = (2, 10, 2, 3, 5, 12)$ на избранные характеристики производительности.

μ_b	P_{busy}	$E(I)$	$E(R)$	$E(N)$	$E(RO)$	E_{loss}
9	0.139829	8.32816219	0.23238914	0.57965511	0.10519363	0.00317645
9.2	0.136489	8.32825947	0.23347327	0.57428849	0.10559878	0.00316335
9.4	0.133303	8.32834721	0.23451371	0.56922758	0.10598521	0.00315204
9.6	0.130263	8.32842731	0.23551297	0.56444746	0.10635413	0.00314224
9.8	0.127357	8.3285017	0.23647314	0.55992573	0.10670675	0.00313349
10	0.124578	8.32856941	0.23739675	0.55564225	0.10704409	0.00312592

Таблица 3. Влияние интенсивности μ_b при $(\lambda, \mu_v, \theta, \beta, s, S) = (2, 3, 2, 3, 5, 12)$ на избранные характеристики производительности.

θ	P_{busy}	$E(I)$	$E(R)$	$E(N)$	$E(RO)$	E_{loss}
2	0.124578	8.32856941	0.23739675	0.55564225	0.10704469	0.00312592
2.2	0.126890	8.32891941	0.23743591	0.5431757	0.10379495	0.00301799
2.4	0.129062	8.32921982	0.23743591	0.53166908	0.10073796	0.00292426
2.6	0.131108	8.32947826	0.23747325	0.52101725	0.09785652	0.00284229
2.8	0.133037	8.32970238	0.23747492	0.5111292	0.09513572	0.00277052
3	0.134860	8.32989788	0.23746681	0.50192672	0.09256257	0.00270711

Таблица 4. Влияние интенсивности θ при $(\lambda, \mu_v, \mu_b, \beta, s, S) = (2, 3, 10, 3, 5, 12)$ на избранные характеристики производительности.

β	P_{busy}	$E(I)$	$E(R)$	$E(N)$	$E(RO)$	E_{loss}
3	0.124439	8.32856941	0.23739675	0.55564225	0.10704409	0.00312592
3.2	0.124482	8.37062931	0.23654588	0.55563289	0.10708673	0.00257989
3.4	0.124515	8.40771294	0.23571137	0.55562615	0.1071198	0.00216231
3.6	0.124541	8.44065475	0.23489445	0.55562139	0.10714592	0.001838
3.8	0.124562	8.47011089	0.23409598	0.55561811	0.1071667	0.00158269
4	0.124578	8.49660873	0.23331647	0.5556159	0.10718346	0.00137916

Таблица 5. Влияние интенсивности β при $(\lambda, \mu_v, \mu_b, \theta, s, S) = (2, 3, 10, 2, 5, 12)$ на избранные характеристики производительности.

В табл. 4 представлены результаты расчетов избранных характеристик производительности при возрастающем параметре длительности каникул θ. Результаты численного эксперимента показывают, что вероятность занятости и средний уровень запаса возрастают, но число клиентов в системе, средняя интенсивность запроса пополнения и средняя интенсивность потерь клиентов убывают. Для растущих значений интенсивности пополнения запаса β (см. табл. 5), средняя интенсивность запроса пополнения возрастает, в то время как среднее число пополнений в единицу времени убывает. Заметим, что в экпериментах не отмечено значимого влияния параметра β на прочие характеристики производительности, такие как вероятность занятости сервера, средний уровень запаса и среднее число клиентов в системе, как следует из табл. 4.

Пусть необходимо найти оптимальную пару управляющих переменных, (s^*, Q^*), минимизирующую стоимость $\mathcal{F}(s, Q)$, определенную следующим образом:

$$\mathcal{F}(s, Q) = h.I_m + c_1.E_{loss} + c_2.\widetilde{W_{inv}} + (K + Q.c_3)\,.R_r,$$

где K есть фиксированная стоимость размещения требования, c_1 есть стоимость, вызванная потерей клиента, c_2 есть стоимость единицы времени ожидания клиента, вызванного опустошением запаса, c_3 есть стоимость поставки единицы товара и h есть стоимость хранения на складе единицы товара за единицу времени. Хотя оптимальные значения параметров s и Q не удалось получить аналитически (в связи со сложностью стоимостной функции), данные параметры могут быть получены численно. Пусть значения параметров системы таковы: $\lambda = 5, \mu_v = 3, \mu_b = 10, \beta = 3, K = \$500, h = \$5, c_1 = \$100, c_2 = \$50$ и $c_3 = \$50$, тогда оптимальной парой параметров (s^*, Q^*) является пара $(4, 15)$, при этом стоимость принимает наименьшее значение \$134.9468, что подтверждено методом глобального поиска.

5. Заключение

В работе исследована модель системы обслуживания-запасания типа $M/M/1$ с политикой пополнения запаса (s, Q), рабочими каникулами, прерываниями каникул и потерянными продажами. Получено стационарное распределение модели с использованием обобщенного процесса рождения-гибели, обнаружено, что полученное условие стационарности является менее ограничительным по сравнению с классической системой обслуживания типа $M/M/1$ с рабочими каникулами и прерыванием каникул (см. [14]). Представлен метод расчета стационарного распределения. Проведен анализ периода занятости; получены важные характеристики системы, такие как стационарное распределение времени ожидания в очереди, среднее время ожидания клиента. Наконец, исследовано влияние параметров системы на характеристики производительности, в ходе численного эксперимента найдена оптимальная конфигурация системы при заданных параметрах. Необходимо отметить, что данная модель может быть обобщена для Марковского входного потока, распределений времени обслуживания и времени пополнения запаса фазового типа. В то же время, вычислительная сложность существенно возрастает с ростом размерности пространства состояний системы.

Благодарности

Авторы благодарят профессора А. Кришнамурти за ценные замечания. Авторы выражают благодарность двум анонимным рецензентам за их комментарии и конструктивную критику. Это помогло улучшить представление статьи.

СПИСОК ЛИТЕРАТУРЫ

1. Baek J.W., Moon S.K. *The $M/M/1$ queue with a production-inventory system and lost sales* // Appl. Math. Comput. 2014. V. 233. P. 534–544.

2. Baek J.W., Moon S.K. *A production-inventory system with a Markovian service queue and lost sales* // J. Korean Stat. Soc. 2016. V. 45. P. 14–24.

3. Baek J.W., Bae Y.H., Lee H.W. and Ahn S. *Continuous-type (s, Q)-inventory model with an attached $M/M/1$ queue and lost sales* // Perform. Eval. 2018. V. 125. P. 68–79.

4. Jayaraman B., Sivakumar B. and Arivarignan G. *A Perishable Inventory System with Postponed Demands and Multiple Server Vacations* // Model. Simul. Eng. 2012. V. 2012. Article ID 620960. 17 p.

5. Koroliuk V.S., Melikov A.Z., Ponomarenko L.A. and Rustamov A.M. *Models of Perishable Queueing-Inventory Systems with Server Vacations* // Cybern. Syst. Anal. 2018. V. 54. N. 1. P. 31–44.

6. Koroliuk V.S., Melikov A.Z., Ponomarenko L.A. and Rustamov A.M. *Asymptotic Analysis of the System with Server Vacation and Perishable Inventory* // Cybern. Syst. Anal. 2017. V. 53. N. 4. P. 543–553.

7. Krenzler R., Daduna H. *Loss systems in a random environment: steady-state analysis* // Queueing Syst. 2014. V. 80. P. 127–153.

8. Krishnamoorthy A., Viswanath C. Narayanan. *Stochastic Decomposition in Production Inventory with Service Time* // European Journal of Operational Research. 2013. V. 228. P. 358–366.

9. Krishnamoorthy A., Viswanath C. Narayanan. *Production inventory with service time and vacation to the server* // IMA Journal of Management Mathematics. 2010.

10. Krishnamoorthy A., Lakshmy B., Manikandan R. *A survey on inventory models with positive service time* // OPSEARCH. 2011. V. 48 N. 2. P. 153–169.

11. Krishnamoorthy A., Shajin D. and Viswanath C. Narayanan. *Inventory with positive service time: a survey* // To appear as a chapter in the book titled: “Advanced Trends in Queueing Theory". Editors: V. Anisimov & N. Limnios, Publisher: ISTE & J.Wiley, London. 2019.

12. Laxmi P.V., Soujanya M.L. *Retrial inventory model with negative customers and multiple working vacations* // International Journal of Management Science and Engineering Management. 2017.

13. Latouche G., Ramaswami V. *A logarithmic reduction algorithm for quasi-birth-and-death processes* // Journal of Applied Probability. 1993. V. 30. N. 3. P. 650–674.

14. Li J. and Tian N. *The* $M/M/1$ *queue with working vacations and vacation interruptions* // J. Syst. Sci. Syst. Eng. 2007. V. 16. N. 1. P. 121–127.

15. Melikov A.Z., Molchanov A.A. *Stock Optimization in Transport/ Storage* // Cybern. Syst. Anal. 1992. V. 28. N. 3. P. 484–487.

16. Melikov A.Z., Rustamov A.M., Ponomarenko L. A. *Approximate Analysis of a Queueing-Inventory System with Early and Delayed Server Vacations* // Automation and Remote Control. 2017. V. 78. N. 11. P. 1991–2003.

17. Melikov A.Z., Shahmaliyev M.O. *Queueing System* $M/M/1/\infty$ *with Perishable Inventory and Repeated Customers* // Automation and Remote Control. 2019. V. 80. N. 1. P. 53–65.

18. Neuts M.F.. *Matrix-Geometric Solutions in Stochastic Models - An Algorithmic Approach.* 2nd ed., Dover Publications, Inc., New York. 1994.

19. Viswanath C. Narayanan, Deepak T.G., Krishnamoorthy A. and Krishnakumar B. *On an (s, S) Inventory Policy with Service Time, Vacation to Server and Correlated Lead Time* // Quality Technology & Quantitative Management. 2008. V. 5. N. 2. P. 129–143.

20. Padmavathi I., Sivakumar B. *Comparison of three vacation policies in discrete time inventory system with postponed demands* // Int. J. of Systems Science: Operations & Logistics. 2019.

21. Padmavathi I., Sivakumar B. and Arivarignan G. *A Retrial Inventory System with Single and Modified Multiple Vacation for Server* // Ann. Oper. Res. 2015. V. 233. N. 1. P. 335–364.

22. Saffari M., Asmussen S. and Haji R. *The $M/M/1$ queue with inventory, lost sale, and general lead times* // Queueing Syst. 2013.

23. Schwarz M., Sauer C., Daduna H., Kulik R. and Szekli R. *$M/M/1$ queueing systems with inventory* // Queueing Syst. 2006. V. 54. P. 55–78.

24. Schwarz M., Wichelhaus C. and Daduna H. *Product Form Models for Queueing Networks with an Inventory* // Stochastic Models. 2007. V. 23. No. 4. P. 627–663.

25. Shajin D., Benny B., Razumchik R. V., Krishnamoorthy A. *Discrete Product Inventory Control System with Positive Service Time and Two Operation Modes* // Automation and Remote Control. 2018. V. 79. No. 9. P. 1593–1608.

26. Shajin D., Krishnamoorthy A. *Stochastic decomposition in retrial queueing inventory system* // RAIRO Oper. Res. (в печати) 2018.

27. Sigman K., Simchi-Levi D. *Light traffic heuristic for an $M/G/1$ queue with limited inventory* // Ann. Oper. Res. 1992. V. 40. P. 371–380.

28. Sivakumar B. *An Inventory System with Retrial Demands and Multiple Server Vacation* // Quality Technology & Quantitative Management. 2011. V. 8. No. 2. P. 125–146.

AN $M/M/1$ QUEUEING-INVENTORY SYSTEM WITH WORKING VACATIONS, VACATION INTERRUPTIONS AND LOST SALES

Manikandan Rangasvami, Department of Mathematics, Central University of Kerala, Ph.D. (mani@cukerala.ac.in),
Sajeev S. Nair, Department of Mathematics, Government Engineering College (sajeev.snair@rediffmail.com).

Abstract: We consider a single server queueing-inventory system under (s, Q) replenishment strategy with working vacations, vacation interruptions and lost sales. The server provides service at a lower rate during working vacations than while in normal mode of service. If working vacation realizes while providing service in that mode, then the server continues in the present status until the current service is completed. Upon that he switches to normal mode of service, provided there is at least one customer waiting. If no customer is waiting at this point of time, he goes for vacation. Also we assume that if there are customers in the system at a service completion epoch during working vacation, the server will comeback to the normal working mode, else the server stays in the working vacation mode. With the system having infinite capacity, the condition for stability of the system is obtained, followed by computation of the steady-state probability vector is discussed. Various performance measures are evaluated. In addition, busy period analysis and the stationary waiting time distribution in the queue is derived. Numerical illustrations are provided to illustrate the system performance and an optimization problem is also discussed.

Keywords: queueing-inventory, positive service time, server vacation, lost sales.