Информационные процессы, Том 5, № 5, 2005, стр. 367–379.

Система массового обслуживания MAP/G/1/r с фоновыми заявками

П.П.Бочаров*, Л.О.Шлумпер**

**Российский университет дружбы народов, Москва, Россия*
***Московский физико-технический институт, Москва, Россия*
Поступила в редколлегию 27.09.2005

Аннотация—Исследуется однолинейная система массового обслуживания с марковским потоком основных заявок и фоновыми заявками, поступающими из бункера, где их запас не ограничен. Для основных заявок имеется накопитель конечной ёмкости. Длительности обслуживания как основных, так и фоновых заявок имеют произвольные функции распределения. Основные заявки имеют относительный приоритет при обслуживании по сравнению с фоновыми заявками. Выведен матричный алгоритм для расчёта стационарных вероятностей состояний системы как для произвольных моментов времени, так и для моментов поступления основных заявок. Получены преобразования Лапласа-Стилтьеса для времени ожидания основных заявок и времени их пребывания в системе.

1. ОПИСАНИЕ СИСТЕМЫ

Рассматривается однолинейная система массового обслуживания (СМО). На вход системы поступает марковский поток основных заявок (1-заявок) [1], характеризуемый матрицами Λ и N порядка l. Элементы матрицы Λ, не находящиеся на ее главной диагонали, задают интенсивности переходов марковского процесса, управляющего переходами в множестве его состояний $\{1, 2, \ldots, l\}$ — фаз генерации 1-заявок — и называемого в дальнейшем просто процессом генерации 1-заявок, при этом переходы с фазы на фазу не сопровождаются собственно поступлениями 1-заявок. В свою очередь, элементы матрицы N также задают интенсивности переходов процесса генерации 1-заявок в множестве фаз генерации, но такие переходы сопровождаются появлением новой 1-заявки.

Положим $\boldsymbol{\lambda} = -\Lambda\mathbf{1}$; здесь $\mathbf{1}$ — вектор из единиц, размерность которого определяется из контекста. Компонента λ_j вектора $\boldsymbol{\lambda}$ представляет собой интенсивность поступления 1-заявки, закончившей свою генерацию на фазе j. Заметим, что имеет место также равенство $\boldsymbol{\lambda} = N\mathbf{1}$.

Кроме того, положим $\Lambda^* = \Lambda + N$. Матрица Λ^* представляет собой матрицу интенсивностей переходов процесса генерации 1-заявок в множестве состояний (фаз генерации) $\{1, 2, \ldots, l\}$. В дальнейшем будем предполагать, что матрица N отлична от нулевой матрицы и, кроме того, матрица Λ^* неразложима.

Далее, пусть вектор $\mathbf{p}^*$ определяется из системы уравнений равновесия (СУР)

$$\mathbf{p}^{*\mathrm{T}}\Lambda^* = \mathbf{0}^{\mathrm{T}}$$

с условием нормировки

$$\mathbf{p}^{*\mathrm{T}}\mathbf{1} = 1;$$

здесь $\mathbf{0}$ — вектор из нулей, размерность которого определяется из контекста. Заметим, что компонента p_j^* вектора $\mathbf{p}^*$ определяет стационарную вероятность того, что процесс генерации

1-заявок в некоторый (произвольный) момент времени находится на фазе j. С учетом этого, можно записать выражение для стационарной интенсивности λ потока 1-заявок:

$$\lambda = \mathbf{p}^{*\mathrm{T}}\boldsymbol{\lambda}. \tag{1}$$

Для 1-заявок имеется накопитель ограниченной ёмкости r, $1 \le r \le \infty$. Если в момент поступления 1-заявки в накопителе нет свободных мест, то данная заявка теряется. Если после окончания обслуживания 1-заявки накопитель пуст, то на прибор из бункера поступает фоновая заявка (2-заявка). Предполагается, что запас фоновых заявок в бункере не ограничен, т.е. поток фоновых заявок является насыщенным. После окончания обслуживания 2-заявки на прибор поступает 1-заявка из накопителя, если он к этому моменту не пуст. В противном случае на прибор поступает из бункера новая 2-заявка. Поступившая во время обслуживания 2-заявки новая 1-заявка не прерывает обслуживание фоновой заявки и занимает свободное место в накопителе (если оно имеется). Таким образом, основные заявки обладают относительным приоритетом при обслуживании по сравнению с 2-заявками.

Длительности обслуживания как основных, так и фоновых заявок представляют собой независимые случайные величины. Время обслуживания j-заявки имеет функцию распределения (ФР) $B_j(x)$, $j = 1, 2$. Предполагается, что $B_j(0) = 0$ и, кроме того, $B_j(x)$ имеет конечное среднее значение, т.е. $\int_0^\infty x dB_j(x) = \frac{1}{\mu_j} < \infty$, $j = 1, 2$. Также предполагается, что для всех собственных значений $\{\sigma_i\}$ матрицы Λ выполнено условие $\beta_1(-\sigma_i) \ne 0$, где $\beta_j(s)$ — преобразование Лапласа-Стилтьеса (ПЛС) ФР $B_j(x)$, $j = 1, 2$.

Описанную выше СМО в обозначениях Кендалла будем классифицировать как $(MAP, ST)/\overrightarrow{G}_2/1/r$ (здесь символ "ST" есть сокращение от английского "saturated" — "насыщенный"). Кроме того, данную СМО можно отнести к системам с прогулками (vacations). Так как поступление фоновых заявок в рассматриваемую СМО происходит только только при освобождении СМО от основных заявок, то она соответствует системам с отключением прибора при их опустошении от заявок основного потока (exhaustive service discipline). Отличительной особенностью СМО $(MAP, ST)/\overrightarrow{G}_2/1/r$ является насыщенность потока фоновых заявок.

Остановимся тут лишь на наиболее близких по типу рассматриваемой СМО работах. Работа [2] посвящена СМО $(PH, ST)/(PH, PH)/1/r$ с потоком основных заявок фазового типа и PH-распределениями длительностей обслуживания заявок обоих типов. Для этой СМО в [2] получены явные матричные выражения для стационарного распределения вероятностей её состояний. В [3] рассмотрено обобщение СМО $(PH, ST)/(PH, PH)/1/r$ на случай, когда процесс поступления основных заявок, а также процессы обслуживания как основных, так и фоновых заявок являются марковскими. В [3] получен устойчивый рекуррентно-матричный алгоритм для расчета стационарных вероятностей состояний системы $(MAP, ST)/(MSP, MSP)/1/r$, рассматриваемой в произвольные моменты времени, а также в моменты поступления основных заявок. Помимо этого, выведены выражения для ряда показателей производительности СМО в стационарном режиме. Также в [3] получено выражение для ПЛС времени ожидания начала обслуживания основных заявок. Обзор библиографии по СМО в непрерывном времени с прогулками при опустошении системы приводится в [3].

В настоящей работе для СМО $(MAP, ST)/\overrightarrow{G}_2/1/r$ получен рекуррентный матричный алгоритм для вычисления стационарного распределения вероятностей состояний системы, рассматриваемой как в произвольные моменты времени, так и в моменты поступления основных заявок. Выведено ПЛС времени ожидания начала обслуживания основных заявок, а также ПЛС времени их пребывания в системе.

Как и её аналоги с обслуживанием марковского типа в непрерывном [3] и в дискретном времени [4], данная система может быть эффективно использована для моделирования широкого спектра процессов в современных сетевых технологиях. СМО с фоновыми заявками описывают также поведение автоматизированной информационной системы (АИС) с оперативной реорганизацией баз данных (БД) [5, 6]. Такие системы управления БД поддерживают совместное обслуживание запросов и реорганизацию БД с относительным приоритетом запросов перед задачами реорганизации.

2. СТОХАСТИЧЕСКАЯ МОДЕЛЬ

Стохастическое поведение рассматриваемой СМО описывается линейчатым марковским процессом $\{\xi(t),\ t \geq 0\}$ над множеством состояний

$$\mathcal{X} = \left\{(i,k,j,x), \quad i=\overline{1,l}, \quad k=\overline{0,r}, \quad j=1,2, \quad x \geq 0\right\}.$$

Состояния процесса $\{\xi(t),\ t \geq 0\}$ имеют следующий физический смысл. Если для некоторого момента времени t $\xi(t)=(i,k,j,x)$, то это означает, что процесс генерации основной заявки на входе системы находится на фазе i, в накопителе имеется k 1-заявок, на приборе обслуживается j-заявка, и время, прошедшее с начала её обслуживания, равно x.

Рассмотрим последовательность $\{t_n,\ n \geq 0\}$ моментов начала обслуживания заявок (основных и фоновых) на приборе. Положим $\xi_n = \xi\left(t_n+0\right)$. Последовательность $\{\xi_n,\ n \geq 0\}$ образует однородную цепь Маркова, вложенную по моментам переходов в начало полупрямых (i,k,j,x). Нетрудно видеть, что эта цепь Маркова является неприводимой. Поэтому, согласно [1, гл. 1], процесс $\{\xi(t),\ t \geq 0\}$ является эргодическим и существует его единственное стационарное распределение, при этом стационарное распределение плотности вероятностей $p_{i,k,j}(x)$ состояний (i,k,j,x) представляются в виде

$$p_{i,k,j}(x) = \left[1 - B_j(x)\right] q_{i,k,j}(x), \tag{2}$$

где $q_{i,k,j}(x) < C < \infty$ для всех $i=\overline{1,l},\ k=\overline{0,r},\ j=1,2,\ x \geq 0$.

Введём векторы $\mathbf{q}_{k,j}^T(x) = \left(q_{1,k,j}, q_{2,k,j}, \ldots, q_{l,k,j}\right)$. Следуя [1, гл. 6, § 7], можно показать, что $\mathbf{q}_{k,j}(x)$ удовлетворяют системе дифференциальных уравнений

$$\frac{d}{dx}\mathbf{q}_{k,j}^T(x) = \mathbf{q}_{k,j}^T(x)\,\Lambda + u(k)\,\mathbf{q}_{k-1,j}^T(x)\,N, \quad k=\overline{0,r-1},\ j=1,2, \tag{3}$$

$$\frac{d}{dx}\mathbf{q}_{r,j}^T(x) = \mathbf{q}_{r,j}^T(x)\,\Lambda^* + \mathbf{q}_{r-1,j}^T(x)\,N, \quad j=1,2, \tag{4}$$

с граничными условиями

$$\mathbf{q}_{k,1}^T(0) = \sum_{j=1}^{2}\int_0^\infty \mathbf{q}_{k+1,j}^T(x)dB_j(x), \quad k=\overline{0,r-1}, \tag{5}$$

$$\mathbf{q}_{r,1}^T(0) = \mathbf{0}^T, \tag{6}$$

$$\mathbf{q}_{0,2}^T(0) = \sum_{j=1}^{2}\int_0^\infty \mathbf{q}_{0,j}^T(x)dB_j(x), \tag{7}$$

$$\mathbf{q}_{k,2}^T(0) = \mathbf{0}^T, \quad k=\overline{1,r}; \tag{8}$$

здесь $u(x)$ — функция Хевисайда.

Положим $(i,k,j)=\bigcup\limits_{x\geq 0}(i,k,j,x)$. Стационарная вероятность $p_{i,k,j}$ состояния (i,k,j) СМО, когда не учитывается прошедшее время обслуживания заявки, находящейся на приборе, имеет вид

$$p_{i,k,j}=\int\limits_0^\infty p_{i,k,j}(x)dx.$$

Положим $\mathbf{p}^T_{k,j}=(p_{1,k,j},p_{2,k,j},\dots,p_{l,k,j})$. Тогда условие нормировки запишется в виде

$$\mathbf{p}^T_{\cdot,\cdot}\mathbf{1}=1; \tag{9}$$

здесь и далее символ "$\cdot$" в качестве нижнего индекса означает суммирование по всем возможным значениям данного индекса.

3. РЕШЕНИЕ УРАВНЕНИЙ РАВНОВЕСИЯ

По аналогии с марковскими процессами с дискретным множеством состояний будем называть систему уравнений (3)–(8) системой уравнений равновесия (СУР).

Приступим к решению СУР. Рассмотрим сначала систему дифференциальных уравнений (3) и (4). Для её решения нам потребуются следующие вспомогательные матричные функции ([1, гл. 6, § 7]):

$$\begin{aligned}F_0(x)&=e^{\Lambda x},\\ F_k(x)&=\int\limits_0^\infty F_{k-1}(y)Ne^{\Lambda(x-y)}dy,\quad k\geq 1,\end{aligned} \tag{10}$$

с граничными условиями

$$F_0(0)=I,\quad F_k(0)=0,\quad k\geq 1. \tag{11}$$

Функции $F_k(x)$ удовлетворяют системе дифференциальных уравнений

$$\frac{d}{dx}F_k(x)=F_k(x)\Lambda+u(k)F_{k-1}(x)N,\quad k\geq 0, \tag{12}$$

с граничными условиями (11).

Матрицы $F_k(x)$ имеют следующий вероятностный смысл: $(F_k(x))_{i,m}$ есть вероятность того, что за интервал времени $[0,x)$ поступит k 1-заявок и процесс их генерации в момент x будет находиться на фазе m при условии, что в момент времени 0 он находился на фазе i.

Для сокращения записи в дальнейшем примем обозначение $\mathbf{q}_{k,j}\equiv\mathbf{q}_{k,j}(0)$.

Теорема 1. *Решение системы дифференциальных уравнений (3) и (4) с граничными условиями (5)–(8) имеет вид*

$$\mathbf{q}^T_{k,1}(x)=\sum_{m=0}^{k}\mathbf{q}^T_{k-m,1}F_m(x),\quad k=\overline{0,r-1}, \tag{13}$$

$$\mathbf{q}^T_{r,1}(x)=\mathbf{q}^T_{0,1}e^{\Lambda^* x}-\sum_{k=0}^{r-1}\mathbf{q}^T_{k,1}(x), \tag{14}$$

$$\mathbf{q}^T_{k,2}(x)=\mathbf{q}^T_{0,2}F_k(x),\quad k=\overline{0,r-1}, \tag{15}$$

$$\mathbf{q}^T_{r,2}(x)=\mathbf{q}^T_{0,2}e^{\Lambda^* x}-\sum_{k=0}^{r-1}\mathbf{q}^T_{k,2}(x). \tag{16}$$

Доказательство. Справедливость формул (13) и (15) доказывается непосредственной подстановкой их в уравнение (3) с учётом (5), (8) и (10)–(12).

Для доказательства (14) и (16) просуммируем уравнения (3) и (4) по $k = \overline{0, r}$ при фиксированном $j = 1, 2$. В результате получим дифференциальное уравнение

$$\frac{d}{dx}\mathbf{q}_{\cdot,j}^T(x) = \mathbf{q}_{\cdot,j}^T(x)\Lambda^*, \quad j = 1, 2. \tag{17}$$

Отсюда с учётом (8) следует, что

$$\mathbf{q}_{\cdot,1}(x)^T = \mathbf{q}_{\cdot,1}^T e^{\Lambda^* x}, \tag{18}$$
$$\mathbf{q}_{\cdot,2}(x)^T = \mathbf{q}_{0,2}^T e^{\Lambda^* x}. \tag{19}$$

Из (18) и (19) непосредственно следуют формулы (14) и (16). Таким образом, теорема доказана.

Перейдём теперь к отысканию неизвестных векторов $\mathbf{q}_{k,1}$, $k = \overline{0, r-1}$, и $\mathbf{q}_{0,2}$.

Положим $B_{k,j} = \int_0^\infty F_k(x) dB_j(x)$. Матрица $B_{k,j}$ представляет собой матричный аналог экспоненциального момента порядка k ФР $B_j(x)$, и её элемент $(B_{k,j})_{i,m}$ есть вероятность того, что за время обслуживания j-заявки поступит k 1-заявок и процесс генерации 1-заявок перейдёт на фазу m при условии, что в момент начала обслуживания данной заявки он находился на фазе i. Для вычисления матриц $B_{k,j}$ можно использовать методы, предложенные в [1, гл. 6, § 7].

Теорема 2. *Векторы* $\mathbf{q}_{k,1}$, $k = \overline{0, r-1}$, *определяются соотношениями*

$$\mathbf{q}_{k,1}^T = \mathbf{q}_{0,2}^T Q_k, \quad k = \overline{0, r-1}, \tag{20}$$

где матрицы Q_k, $k = \overline{0, r-1}$, *задаются следующими рекуррентными формулами:*

$$\begin{aligned} &Q_0 = (I - B_{0,2})\, B_{0,1}^{-1}, \\ &Q_k = \left(Q_{k-1} - \sum_{m=1}^{k} Q_{k-m} B_{m,1} - B_{k,2} \right) B_{0,1}^{-1}, \quad k = \overline{1, r-1}, \end{aligned} \tag{21}$$

а неизвестный вектор $\mathbf{q}_{0,2}$ *является единственным решением системы алгебраических уравнений*

$$\mathbf{q}_{0,2}^T Q = \mathbf{0}^T, \tag{22}$$
$$\mathbf{q}_{0,2}^T \left(\frac{1}{\mu_1} Q_{\cdot} + \frac{1}{\mu_2} I \right) \mathbf{1} = 1, \tag{23}$$

где

$$Q = I - B_{0,2}^* - Q_{\cdot} \left(I - B_{0,1}^* \right) \tag{24}$$

и $B_{0,j}^* = \int\limits_0^\infty e^{\Lambda^* x} dB_j(x)$.

Доказательство. Подставим выражения (13) и (15) в уравнение (4) при $k = \overline{0, r-2}$ и (7). В результате с учётом (8) получим

$$
\begin{aligned}
\mathbf{q}_{k,1}^T &= \sum_{m=0}^{k+1} \mathbf{q}_{k-m+1,1}^T B_{m,1} + \mathbf{q}_{0,2}^T B_{k+1,2}, \quad k = \overline{0, r-2}, \\
\mathbf{q}_{0,2}^T &= \sum_{j=1}^{2} \mathbf{q}_{0,j}^T B_{0,j}\,.
\end{aligned} \tag{25}
$$

Система алгебраических уравнений (25) имеет $l(r+1)$ неизвестных $q_{i,k,1}$, $i = \overline{1, l}$, $k = \overline{0, r-1}$ и $q_{i,0,2}$, $i = \overline{1, l}$, и такой же порядок. Так как стационарное распределение процесса $\{\xi(t), t \geq 0\}$ единственно, то ранг системы уравнений (25) равен $l(r+1) - 1$. Следовательно, её решение находится с точностью до константы, которая определяется из дополнительного условия нормировки (9).

Считая вектор $\mathbf{q}_{0,2}$ известным, представим систему уравнений (25) как систему уравнений относительно неизвестных $\mathbf{q}_{k,1}$, $k = \overline{0, r-1}$:

$$
\begin{gathered}
\left(\mathbf{q}_{0,1}^T, \mathbf{q}_{1,1}^T, \ldots, \mathbf{q}_{r-1,1}^T\right) \cdot \begin{pmatrix} B_{0,1} & B_{1,1} - I & B_{r,1} & \cdots & B_{r-1,1} \\ 0 & B_{0,1} & B_{1,1} - I & \cdots & B_{r-2,q} \\ 0 & 0 & B_{0,1} & \cdots & B_{r-3,1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & B_{0,1} \end{pmatrix} = \\
= \left(\mathbf{q}_{0,2}^T \left(I - B_{0,2}\right), -\mathbf{q}_{0,2}^T B_{1,2}, -\mathbf{q}_{0,2}^T B_{2,2}, \ldots, -\mathbf{q}_{0,2}^T B_{r-1,2}\right).
\end{gathered} \tag{26}
$$

Матрица коэффициентов системы уравнений (26) является блочно-треугольной. Поэтому, если матрица $B_{0,1}$ не вырождена, все неизвестные $\mathbf{q}_{k,1}$, $k = \overline{0, r-1}$, можно выразить через вектор $\mathbf{q}_{0,2}$.

Заметим, что

$$
B_{0,1} = \int_0^\infty e^{\Lambda x} dB_1(x) = \beta_1(-\Lambda),
$$

где $\beta_1(s)$ — ПЛС ФР $B_1(x)$. В силу того, что собственное число σ_i матрицы Λ является собственным числом матрицы $\beta_1(-\Lambda) = B_{0,1}$ [7], и с учётом предположения, сделанного в разделе 1 о том, что $\beta_1(-\sigma_i) \neq 0$ для всех $i = \overline{0, l}$, имеем, что обратная матрица $B_{0,1}^{-1}$ существует. Поэтому будем искать решение системы уравнений в виде (20), где Q_k — неизвестные матрицы. Подставляя (20) в (26), методом неопределённых коэффициентов получаем, что матрицы Q_k, $k = \overline{0, r-1}$, удовлетворяют соотношениям (21).

Неизвестный вектор $\mathbf{q}_{0,2}$ можно определить из оставшегося неиспользованным уравнения (5) при $k = r-1$ после подстановки в него (13) - (16). Однако для определения $\mathbf{q}_{0,2}$ можно использовать не получаемую таким образом систему уравнений, а эквивалентную ей систему уравнений (22) и (23), которая выводится следующим образом.

Просуммируем все уравнения в (5) - (8). В результате получим равенство

$$
\mathbf{q}_{0,2}^T + \mathbf{q}_{\cdot,1}^T = \sum_{j=1}^{2} \int_0^\infty \mathbf{q}_{\cdot,j}^T(x) dB_j(x). \tag{27}
$$

Подставляя в (27) выражения (18) и (19) для $\mathbf{q}^T_{\cdot,j}(x)$, приходим к следующей системе уравнений:

$$\mathbf{q}^T_{0,2} + \mathbf{q}^T_{\cdot,1} = \mathbf{q}^T_{\cdot,1} B^*_{0,1} + \mathbf{q}^T_{0,2} B^*_{0,2}. \tag{28}$$

Подставляя теперь в (28) выражение (20) для $\mathbf{q}_{k,1}$, $k = \overline{0, r-1}$, приходим к системе уравнений (22), из которой вектор $\mathbf{q}_{0,2}$ находится с точностью до константы. Последнюю можно определить из условия нормировки (9) либо эквивалентного (9) уравнения, получаемого следующим образом.

Умножим обе части (18) и (19) справа на вектор $\mathbf{1}$. В результате получим

$$q_{\cdot,1}(x) = q_{\cdot,1}, \tag{29}$$

$$q_{\cdot,2}(x) = q_{0,2}, \tag{30}$$

где $q_{\cdot,j}(x) \equiv \mathbf{q}^T_{\cdot,j}(x)\mathbf{1}$, $q_{\cdot,1} \equiv \mathbf{q}^T_{\cdot,1}\mathbf{1}$ и $q_{0,2} \equiv \mathbf{q}^T_{0,2}\mathbf{1}$. Далее, умножая соотношения (29) и (30) на $1-B_1(x)$ и $1-B_2(x)$ соответственно, а затем интегрируя их по x в пределах от 0 до ∞, получим

$$p_{\cdot,1} = \frac{1}{\mu_1} q_{\cdot,1}, \tag{31}$$

$$p_{\cdot,2} = \frac{1}{\mu_2} q_{0,2}. \tag{32}$$

Отсюда с учётом условия нормировки (9) и равенства (20) получаем уравнение (23). Следовательно, теорема 2 доказана.

Нам осталось теперь найти выражение для векторов $\mathbf{p}_{k,j}$, $k = \overline{0, r}$, $j = 1, 2$.

Положим $G_{k,j} = \int\limits_0^\infty [1 - B_j(x)]\, F_k(x) dx$, $k = \overline{1, r-1}$, $j = 1, 2$, и $G^*_{0,j} = \int\limits_0^\infty [1 - B_j(x)]\, e^{\Lambda^* x} dx$, $j = 1, 2$.

Теорема 3. *Стационарное распределение вероятностей* $\mathbf{p}_{k,j}$, $k = \overline{0, r}$, $j = 1, 2$, *определяется следующими выражениями:*

$$\mathbf{p}^T_{k,1} = \sum_{m=0}^{k} \mathbf{q}^T_{k-m,1} G_{m,1}, \quad k = \overline{0, r-1}, \tag{33}$$

$$\mathbf{p}^T_{r,1} = \mathbf{q}^T_{\cdot,1} G^*_{0,1} - \sum_{k=0}^{r-1} \mathbf{p}^T_{k,1}, \tag{34}$$

$$\mathbf{p}^T_{k,2} = \mathbf{q}^T_{0,2} G_{k,2}, \quad k = \overline{0, r-1}, \tag{35}$$

$$\mathbf{p}^T_{r,2} = \mathbf{q}^T_{0,2} G^*_{0,2} - \sum_{k=0}^{r-1} \mathbf{p}^T_{k,2}. \tag{36}$$

Доказательство. Доказательство формул (33) - (36) получаем, умножая соотношения (13), (14) и (15), (16) на $1 - B_{1(x)}$ и $1 - B_2(x)$ соответственно и интегрируя их по x в пределах от 0 до ∞.

Вычисление матриц $G_{k,j}$, используемых в формулах (33) - (36), можно осуществлять рекуррентным образом, используя матрицы $B_{k,1}$. Действительно, умножая обе части уравнения (12) на $1 - B_j(x)$ и интегрируя по x в пределах от 0 до ∞, приходим к соотношениям

$$\int\limits_0^\infty F'_k(x)\, [1 - B_j(x)]\, dx = G_{k,j}\Lambda + u(k) G_{k-1,j} N, \quad k \geq 0, \; j = 1, 2.$$

Отсюда, интегрируя интеграл в левой части равенства по частям и учитывая (11), получаем

$$\begin{aligned} &-I + B_{0,j} = G_{0,j}\Lambda, \quad j = 1,2, \\ &B_{k,j} = G_{k,j}\Lambda + G_{k-1,j}N, \quad k \geq 0, \ j = 1,2. \end{aligned}$$

Последние равенства приводят к следующим рекуррентным соотношениям для $G_{k,j}$:

$$\begin{aligned} &G_{0,j} = (B_{0,j} - I)\,\Lambda^{-1}, \quad j = 1,2, \\ &G_{k,j} = (G_{k-1,j}N - B_{k,j})\,\Lambda^{-1}, \quad k \geq 0, \ j = 1,2. \end{aligned} \tag{37}$$

Заметим, что если нас интересует только распределение вероятностей длины очереди основных заявок без учёта вида обслуживаемой заявки $\mathbf{p}_k \equiv \mathbf{p}_{k,\cdot}$, то для вычисления $\mathbf{p}_k$, $k = \overline{0,r}$, можно использовать более простые соотношения.

Положим $\mathbf{q}_k \equiv \mathbf{q}_{k,\cdot}$. Умножая обе части уравнения (3) на $1 - B_j(x)$, а затем суммируя по $j = 1,2$ и интегрируя полученное равенство по x в пределах от 0 до ∞, получим следующие соотношения:

$$\begin{aligned} &-\mathbf{q}_k^T + \sum_{j=1}^{2}\int_0^\infty \mathbf{q}_{k,j}^T(x)dB_j(x) = \mathbf{p}_k^T\Lambda + u(k)\mathbf{p}_{k-1}^T N, \quad k = \overline{0,r-1}, \\ &\sum_{j-1}^{2}\int_0^\infty \mathbf{q}_{r,j}^T(x)dB_j(x) = \mathbf{p}_r^T\Lambda^* + \mathbf{p}_{r-1}^T N. \end{aligned}$$

Отсюда с учётом граничных условий (5) и (8) получаем

$$\begin{aligned} &-\mathbf{q}_0^T + \mathbf{q}_{0,2}^T = \mathbf{p}_0^T\Lambda, \\ &-\mathbf{q}_k^T + \mathbf{q}_{k-1,1}^T = \mathbf{p}_k^T\Lambda + \mathbf{p}_{k-1}^T N, \quad k = \overline{1,r-1}, \\ &\mathbf{q}_{r-1,1}^T = \mathbf{p}_r^T\Lambda^* + \mathbf{p}_r^T\Lambda^* + \mathbf{p}_{r-1}^T N. \end{aligned}$$

Из этих равенств с помощью несложных преобразований получаем следующие рекуррентные соотношения для $\mathbf{p}_k$:

$$\begin{aligned} &\mathbf{p}_0^T = \left(-\mathbf{q}_0^T + \mathbf{q}_{0,2}^T\right)\Lambda^{-1}, \\ &\mathbf{p}_k^T = -\left(\mathbf{q}_k^T + \sum_{m=0}^{k-1}\mathbf{p}_m^T\Lambda^*\right)\Lambda^{-1}, \quad k = \overline{1,r-1}. \end{aligned} \tag{38}$$

Рассмотрим теперь уравнение (17). Умножая обе части этого уравнения на $1 - B_j(x)$, а затем суммируя их по $j = 1,2$ и интегрируя по x в пределах от 0 до ∞, получим

$$-\mathbf{q}_{\cdot}^T + \sum_{j=1}^{2}\int_0^\infty \mathbf{q}_{\cdot,j}^T(x)dB_j(x) = \mathbf{p}_{\cdot}^T\Lambda^*. \tag{39}$$

Суммируя теперь уравнения (5) - (8) для граничных условий по всем значениям k, получаем равенство

$$\mathbf{q}_{\cdot}^T = \sum_{j=1}^{2}\int_0^\infty \mathbf{q}_{\cdot,j}^T(x)dB_j(x). \tag{40}$$

Из (39) и (40) получаем уравнение

$$\mathbf{p}_{\cdot}^T \Lambda^* = \mathbf{0}^T. \tag{41}$$

Отсюда с учётом равенства (9), в частности, вытекает, что

$$\mathbf{p}_{\cdot}^T = \mathbf{p}^*. \tag{42}$$

Таким образом, вычислив $\mathbf{p}_{\cdot}$ из системы уравнений (41) и (9), мы можем теперь определить $\mathbf{p}_r$ с помощью соотношения

$$\mathbf{p}_r = \mathbf{p}_{\cdot} - \sum_{k=0}^{r-1} \mathbf{p}_k. \tag{43}$$

4. НЕКОТОРЫЕ СООТНОШЕНИЯ ДЛЯ ПОКАЗАТЕЛЕЙ ПРОИЗВОДИТЕЛЬНОСТИ

Обозначим через λ_A интенсивность потока 1-заявок, принятого в систему, через λ_L — потерянного на входе системы и через $\lambda_{D,1}$ — обслуженного системой.

В силу равенства (42) интенсивность λ поступающего на систему потока 1-заявок, определяется равенствами

$$\lambda = \mathbf{p}^{*T} \boldsymbol{\lambda} = \mathbf{p}_{\cdot}^T \boldsymbol{\lambda}. \tag{44}$$

Рассмотрим теперь вектор $\mathbf{p}_{\cdot} - \mathbf{p}_{r,0}$. Заметим, что i-я компонента этого вектора есть стационарная вероятность того, что процесс генерации 1-заявок находится на фазе i и в накопителе находится по крайней мере одно свободное место, т.е. система доступна для поступающих на неё 1-заявок. Тогда для интенсивности λ_A принятого потока 1-заявок получаем следующее выражение:

$$\lambda_A = \left(\mathbf{p}_{\cdot}^T - \mathbf{p}_r^T\right) \boldsymbol{\lambda}. \tag{45}$$

Так как в стационарном режиме $\lambda_A = \lambda_{D,1}$, то из (45) следует, что

$$\lambda_{D,1} = \left(\mathbf{p}_{\cdot}^T - \mathbf{p}_r^T\right) \boldsymbol{\lambda}. \tag{46}$$

Основываясь на эргодичности процесса $\{\xi(t),\ t \geq 0\}$, можно выписать другое выражение для $\lambda_{D,1}$:

$$\lambda_{D,1} = \mu_1 \mathbf{p}_{\cdot,\mathbf{1}}^{\mathbf{T}} \mathbf{1},$$

где μ_j — интенсивность обслуживания j-заявок.

Далее учитывая, что $\lambda = \lambda_A + \lambda_L$, из (44) и (45) получаем

$$\lambda_L = \mathbf{p}_r^T \boldsymbol{\lambda}. \tag{47}$$

Теперь можно определить вероятность потери 1-заявки π в стационарном режиме функционирования СМО:

$$\pi = \frac{\lambda_L}{\lambda}. \tag{48}$$

Выражение (48) для вероятности потери с учётом (47) можно записать в виде

$$\pi = \frac{1}{\lambda}\mathbf{p}_r^T\boldsymbol{\lambda}. \tag{49}$$

Рассмотрим теперь равенства (38). Умножая обе части этих равенств справа на $\boldsymbol{\lambda}$ и суммируя их по всем значениям k, получим

$$\lambda - \mathbf{p}_r^T\boldsymbol{\lambda} = q_{\cdot,1}. \tag{50}$$

Отсюда с учётом (49) и (31) получаем соотношение

$$\lambda(1-\pi) = \mu_1 p_{\cdot,1}, \tag{51}$$

отражающее равенство интенсивностей принятого в систему и обслуженного ей потоков 1-заявок.

Равенство (51) можно использовать для контроля расчётов стационарного распределения вероятностей состояний системы, осуществляемых с помощью полученного выше алгоритма.

5. СТАЦИОНАРНЫЕ ВЕРОЯТНОСТИ ПО МОМЕНТАМ ПОСТУПЛЕНИЯ ОСНОВНЫХ ЗАЯВОК

Обозначим через $\pi^-_{A,(i,k,j)}(x)$ стационарную плотность вероятности состояния (i,k,j,x) системы, рассматриваемой непосредственно перед поступлением 1-заявки, а через $\pi^-_{A,(i,k,j)}$ — стационарную вероятность состояния $(i,k,j) = \bigcup\limits_{x\geq 0}(i,k,j,x)$; здесь физический смысл индексов i, k, j, x такой же, как и для состояний процесса $\{\xi(t),\ t\geq 0\}$. Введём векторы $\boldsymbol{\pi}^-_{A,(k,j)}(x)$ и $\boldsymbol{\pi}^-_{A,(k,j)}$, аналогичные по структуре и размерности векторам $\mathbf{p}_{k,j}(x)$ и $\mathbf{p}_{k,j}$.

Теорема 4. *Имеют место следующие равенства:*

$$\boldsymbol{\pi}^{T-}_{A,(k,j)}(x) = \frac{1}{\lambda}\mathbf{p}^T_{k,j}(x)N, \quad k=\overline{0,r},\ j=1,2, \tag{52}$$

$$\boldsymbol{\pi}^{T-}_{A,(k,j)} = \frac{1}{\lambda}\mathbf{p}^T_{k,j}N, \quad k=\overline{0,r},\ j=1,2, \tag{53}$$

где λ — интенсивность потока 1-заявок.

Доказательство. Рассмотрим стационарный режим функционирования системы. Для обоснования формулы (52) замечаем, что $N_{i,m}\Delta$ есть элементарная вероятность того, что за время $(t,\ t+\Delta)$ поступит 1-заявка и процесс генерации 1-заявок в момент $t+\Delta$ будет находиться на фазе m при условии, что в момент t он находился на фазе i. Тогда нетрудно видеть, что

$$\boldsymbol{\pi}^{T-}_{A,(k,j)}(x) = \lim_{\Delta\to 0}\frac{\mathbf{p}^T_{k,j}(x)N\Delta}{\int\limits_0^\infty \mathbf{p}^T_{\cdot,\cdot}(x)\,dx\,N\mathbf{1}\Delta} = \frac{\mathbf{p}^T_{k,j}(x)N}{\mathbf{p}^T_{\cdot}\boldsymbol{\lambda}}.$$

Отсюда с учётом равенств (44) вытекает формула (52). Интегрируя обе части равенства (52) по x в пределах от 0 до ∞, получаем формулу (53). Тем самым теорема 4 доказана.

Далее, обозначим через $\pi^-_{A,(k,j)}(x)$ и $\pi^-_{A,(k,j)}$ стационарную плотность вероятностей состояния $(k,j,x) = \bigcup\limits_{i=1}^{l}(i,k,j,x)$ и стационарную вероятность состояния $(k,j) = \bigcup\limits_{x\geq 0}(k,j,x)$, рассматриваемых непосредственно перед поступлением 1-заявки в систему. Умножая обе части равенства (44) справа на вектор $\mathbf{1}$, получаем

$$\pi^-_{A,(k,j)}(x) = \frac{1}{\lambda}\mathbf{p}^T_{k,j}(x)\boldsymbol{\lambda}, \quad k=\overline{0,r},\ j=1,2. \tag{54}$$

Интегрируя обе части равенства (54) по x в пределах от 0 до ∞, получим

$$\pi^-_{A,(k,j)} = \frac{1}{\lambda}\mathbf{p}^T_{k,j}\boldsymbol{\lambda}, \quad k=\overline{0,r},\ j=1,2. \tag{55}$$

Положим $\pi^-_{A,k} = \pi^-_{A,(k,\cdot)}$. Тогда, суммируя равенство (55) по $j=1,2$, получим

$$\pi^-_{A,k} = \frac{1}{\lambda}\mathbf{p}^T_{k}\boldsymbol{\lambda}, \quad k=\overline{0,r},\ j=1,2. \tag{56}$$

В заключение заметим, что вероятность $\pi^-_{A,r}$, определяемая формулой (56), есть вероятность потери 1-заявки, и выражение для неё совпадает с выражением (49), которое было получено из других соображений.

6. СТАЦИОНАРНОЕ РАСПРЕДЕЛЕНИЕ ВРЕМЕНИ ОЖИДАНИЯ ОСНОВНЫХ ЗАЯВОК ПРИ ДИСЦИПЛИНЕ FCFS

Рассмотрим стационарный режим функционирования СМО. В предположении, что 1-заявки выбираются из очереди на обслуживание в порядке их поступления в систему, т.е. в соответствии с дисциплиной FCFS, изучим время ожидания 1-заявок, принятых в систему.

Обозначим через $W_1(t|(i,k,j,x))$ условную ФР времени ожидания некоторой (выделенной) 1-заявки, принятой в систему и заставшую её непосредственно перед поступлением в состоянии (i,k,j,x), а через $\omega_1(s|(i,k,j,x))$ — ПЛС этой ФР.

Если поступающая заявка застала систему в состоянии (i,k,j,x), $i=\overline{0,j}$, $k=\overline{0,r-1}$, $j=1,2$, $x\geq 0$, то время её ожидания начала обслуживания равно сумме длительностей обслуживания k 1-заявок, находящихся в очереди впереди неё и остаточного времени обслуживания j-заявки, находящейся на приборе. Так как длительности обслуживания заявок любого типа независимы, то для ПЛС $\omega_1(s|(i,k,j,x))$ имеем

$$\omega_1(s|(i,k,j,x)) = \beta_1^k(s)\int\limits_0^\infty e^{-sy}\frac{d_yB_j(x+y)}{1-B_j(x)}, \quad i=\overline{0,l},\ k=\overline{0,r-1},\ j=1,2,\ x\geq 0.$$

Далее, обозначим через $W_{i,1}(t)$ стационарную вероятность того, что 1-заявка принята в систему, время её ожидания меньше t и в момент поступления этой заявки процесс генерации 1-заявок находился на фазе i, а через $\omega_{i,1}(s)$ обозначим ПЛС для $W_{i,1}(t)$. Кроме того, введём вектор $\boldsymbol{\omega}_1^T(s) = (\omega_{1,1}(s),\ldots,\omega_{l,1}(s))$. Тогда выражение для $\boldsymbol{\omega}_1(s)$ запишется в виде

$$\boldsymbol{\omega}_1^T(s) = \frac{1}{1-\pi}\sum_{j=1}^{2}\sum_{k=0}^{r-1}\beta_1^k(s)\int\limits_0^\infty \boldsymbol{\pi}^{T-}_{A,(k,j)}(x)\int\limits_0^\infty e^{-sy}\frac{d_yB_j(x+y)}{1-B_j(x)}dx,$$

где π — вероятность потери 1-заявки. С учётом формул (52) и (2) последнее равенство примет вид

$$\boldsymbol{\omega}_1^T(s) = \frac{1}{\lambda(1-\pi)} \sum_{j=1}^{2} \sum_{k=0}^{r-1} \beta_1^k(s) \int\limits_0^\infty \int\limits_0^\infty \mathbf{q}_{k,j}^T(x) e^{-sy} d_y B_j(x+y) dx N.$$

Отсюда с учётом (13), (15) и (12) после несложных преобразований получаем

$$\begin{aligned}\boldsymbol{\omega}_1^T(s) = \frac{1}{\lambda(1-\pi)} \Bigg\{ \sum_{k=0}^{r-1} \beta_1^k(s) \Bigg[\sum_{m=0}^{k} \mathbf{q}_{k-m,1}^T \Bigg[(-1)^{m+1} \beta_1(s)(sI+\Lambda)^{-1} \left(N(sI+\Lambda)^{-1}\right)^m \\ + \sum_{n=0}^{m} (-1)^n B_{m-n,1}(sI+\Lambda)^{-1} \left(N(sI+\Lambda)^{-1}\right)^n \Bigg] \\ + \mathbf{q}_{0,2}^T \Bigg[(-1)^{k+1} \beta_2(s)(sI+\Lambda)^{-1} \left(N(sI+\Lambda)^{-1}\right)^k \\ + \sum_{n=0}^{k} (-1)^n B_{k-n,2}(sI+\Lambda)^{-1} \left(N(sI+\Lambda)^{-1}\right)^n \Bigg] \Bigg] \Bigg\} N.\end{aligned} \quad (57)$$

Заметим, что в силу сделанных выше предположений относительно матриц Λ и N матрица $(sI+\Lambda)^{-1}$ существует.

ПЛС $\omega_1(s)$ для стационарной ФР времени ожидания 1-заявки, принятой в систему, определяется соотношением

$$\omega_1(s) = \boldsymbol{\omega}_1^T(s)\mathbf{1},$$

где $\boldsymbol{\omega}_1(s)$ задаётся формулой (57).

Далее, пусть $\varphi_1(s)$ — ПЛС ФР времени пребывания 1-заявки в системе. Тогда $\varphi_1(s)$ определяется формулой

$$\varphi_1(s) = \omega_1(s)\beta_1(s), \quad (58)$$

вывод которой очевиден.

7. ЗАКЛЮЧЕНИЕ

Рассмотрена однолинейная СМО с марковским потоком основных заявок и насыщенным потоком фоновых заявок (запас которых не ограничен). Обслуживание обоих типов заявок рекуррентное. Для основных заявок имеется накопитель конечной емкости, и они обслуживаются с относительным приоритетом по сравнению с фоновыми заявками. Основным результатом работы является матричный рекуррентный алгоритм для вычисления стационарного распределения вероятностей состояний системы для произвольных моментов времени. На основе стационарных вероятностей для произвольных моментов времени вычисляются также стационарные вероятности по моментам поступления заявок, что, в свою очередь, позволяет получить характеристики (ПЛС) времени ожидания основных заявок и времени их пребывания в системе.

СПИСОК ЛИТЕРАТУРЫ

1. Бочаров П.П., Печинкин А.В. *Теория массового обслуживания.* М.: Изд-во РУДН, 1995.
2. Бочаров П.П. Об обслуживании двух типов заявок с относительным приоритетом и распределениями фазового типа. В кн.: *Теория телетрафика в системах информатики*, М.: Наука, 1989, стр. 45-50.
3. Бочаров П.П., Шлумпер Л.О. Однолинейная система массового обслуживания с фоновыми заявками, *Автоматика и телемеханика*, 2005, № 6, стр. 74-88.
4. Бочаров П.П., Шлумпер Л.О. Однолинейная система массового обслуживания с фоновыми заявками в дискретном времени, *Информационные процессы*, 2005, Том 5, № 3, стр. 236-246.
5. Литвин В.Г., Лысяков П.К. О влиянии реорганизации базы данных на время ответа информационной системы. *Управляющие системы и машины*, 1981, № 2, стр. 105-109.
6. Литвин В.Г., Лысяков П.К. Дискретно-непрерывная модель информационной системы с совместным обслуживанием запросов и реорганизацией баз данных. *Автоматика и телемеханика*, 1983, № 5, стр. 149-158.
7. Като Т. *Теория возмущений линейных операторов.* М.: Изд-во Мир, 1972.