

УДК 519.872

ПОЛНЫЙ РАСЧЕТ СИСТЕМЫ ОБСЛУЖИВАНИЯ С РАСПРЕДЕЛЕНИЯМИ КОКСА

Ю. И. Рыжиков,

доктор техн. наук, профессор

Военно-космическая академия им. А. Ф. Можайского

Разработан алгоритм полного расчета стационарных характеристик системы обслуживания вида $C_2/C_2/n$: стационарного распределения числа заявок, их распределения перед прибытием очередной заявки, моментов распределения интервалов между обслуженными заявками. Обсуждаются частные случаи и обобщения. Рассмотрены особенности программной реализации алгоритма, приведены результаты тестирования. Алгоритм предлагается использовать при расчете сетей обслуживания методом потокоэквивалентной декомпозиции.

An algorithm is proposed to compute the characteristics of the system $C_2/C_2/n$. This algorithm computes the stationary distribution of the number of demands, its distribution before arriving, and the moments of interdeparting intervals. Particular cases and generalizations, programming realization and testing results are discussed. The algorithm is planned to be used in the flow-equivalent network analysis.

Введение

Известно значительное влияние на результаты расчета сложных систем и сетей обслуживания не только средних интервалов между заявками и длительностей обслуживания, но и их распределений. Универсальным методом получения требуемых числовых характеристик является имитационное моделирование, существенными недостатками которого являются сложная логика программ, большой расход машинного времени и низкая точность определения малых вероятностей. Системы автоматизации имитационного моделирования типа GPSS (включая новейшую модификацию последней – GPSS World) не снимают перечисленных проблем, а лишь упрощают программирование ценой замедления работы моделей (как показали проведенные автором эксперименты – в десятки раз). Таким образом, несмотря на увеличение быстродействия ЭВМ, численные методы расчета СМО сохраняют свою актуальность.

Опубликовано множество методов расчета различных классов немарковских систем обслуживания, наиболее универсальные из которых опираются на подбираемые по методу моментов фазовые аппроксимации исходных распределений показательным, эрланговским, гиперэкспоненциальным, коксовым и общим распределением Ньютона. Для последнего, однако, отсутствуют алгоритмы расчета параметров аппроксимации. Эрланговские распределения из-за требования целочисленности фазового параметра не обеспечивают точного вы-

равнивания даже двух моментов. Поэтому наибольшую практическую ценность представляют гиперэкспоненциальное и коксово распределение. Однако для первого из них распределение Эрланга являются особыми случаями: система последовательных фаз не может быть сведена к тому же числу параллельных. Это гарантирует аварийное завершение программы подбора параметров аппроксимирующего распределения, вынуждает менять тип аппроксимации в зависимости от исходных данных и требует обширного набора программ, включающего все возможные комбинации. От данного недостатка свободно распределение Кокса, исследование свойств и систематическое использование которого в обеих позициях систем вида $GI/G/n$ впервые описываются ниже.

Распределение Кокса

К фазовым распределениям относятся порождаемые системой подлежащих прохождению фаз обслуживания (прибытия заявки) с показательной распределенной длительностью пребывания в каждой из них. При фиксации номера фазы такие распределения приобретают марковское свойство, что и определяет целесообразность их использования. Для практических целей применение двухфазного распределения C_2 , обеспечивающего выравнивание трех начальных моментов, представляется вполне достаточным. Диаграмма этого распределения представлена на рис. 1. Параметры μ_1 и μ_2 задают интенсивности потоков пе-

рехода, y – вероятность потребности во второй фазе, \bar{y} – вероятность отсутствия таковой.

Получим алгоритм расчета параметров распределения C_2 по его моментам $\{\tilde{f}\}$. Прежде всего заметим, что эти моменты с вероятностью y являются моментами свертки показательного распределенных задержек в фазах, а с дополнительной к ней суть моменты задержки в первой фазе. Таким образом, для расчета параметров C_2 имеем систему уравнений

$$\begin{aligned} \frac{\bar{y}}{\mu_1} + y \left(\frac{1}{\mu_1} + \frac{1}{\mu_2} \right) &= \tilde{f}_1; \\ \bar{y} \frac{2}{\mu_1^2} + y \left(\frac{2}{\mu_1^2} + 2 \frac{1}{\mu_1 \mu_2} + \frac{2}{\mu_2^2} \right) &= \tilde{f}_2; \\ \bar{y} \frac{6}{\mu_1^3} + y \left(\frac{6}{\mu_1^3} + 3 \frac{2}{\mu_1^2 \mu_2} + 3 \frac{1}{\mu_1 \mu_2^2} + \frac{6}{\mu_2^3} \right) &= \tilde{f}_3. \end{aligned}$$

Перейдем к обратным величинам $\{x_i = \mu_i^{-1}\}$ и нормированным моментам $\{f_i = \tilde{f}_i / i!\}$. Тогда исходная система может быть переписана как

$$\begin{aligned} \bar{y}x_1 + y(x_1 + x_2) &= f_1; \\ \bar{y}x_1^2 + y(x_1^2 + x_1x_2 + x_2^2) &= f_2; \\ \bar{y}x_1^3 + y(x_1^3 + x_1^2x_2 + x_1x_2^2 + x_2^3) &= f_3. \end{aligned}$$

С учетом $y + \bar{y} = 1$ она преобразуется в систему

$$\begin{aligned} x_1 + yx_2 &= f_1; \\ x_1^2 + yx_2(x_1 + x_2) &= f_2; \\ x_1^3 + yx_2(x_1^2 + x_1x_2 + x_2^2) &= f_3. \end{aligned} \quad (1)$$

Из первого уравнения системы исключаем

$$yx_2 = f_1 - x_1 \quad (2)$$

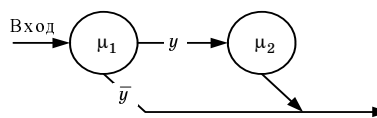
и подставляем в два оставшихся. Первое из них дает

$$x_2 = \frac{f_2 - x_1^2}{f_1 - x_1} - x_1.$$

Конечная форма задачи – квадратное уравнение с решением

$$x_{1,2} = \frac{f_3 - f_1f_2 \pm \sqrt{4f_1^3f_3 + 4f_2^3 + f_3^2 - 3f_1^2f_2^2 - 6f_1f_2f_3}}{2(f_2 - f_1^2)}. \quad (3)$$

Решение для x_1 берется с плюсом, y определяется из равенства (2). Заметим, что случай $f_2 = f_1^2$ соответствует одной экспоненте и должен выявляться на первом этапе алгоритма. Здесь есте-



■ Рис. 1. Двухфазное распределение Кокса

■ Таблица 1. Параметры коксовой аппроксимации

α	y	μ_1	μ_2
0.2	-10.196	268	3.732
0.4	-3.682	.488	3.512
0.6	-1.584	.677	3.323
0.8	-.577	.845	3.155
1.0	.000	1.000	1.000
1.2	.745	2.853	1.147
1.4	.815	2.707	1.293
1.6	.880	2.555	1.445
1.8	.940	2.378	1.622
2.0	1.000	2.000	2.000
2.2	1.061-.011i	2.000+.354i	2.000-.354i
2.4	1.111-.027i	2.000+.485i	2.000-.485i
2.6	1.154-.044i	2.000+.577i	2.000-.577i
2.8	1.190-.062i	2.000+.649i	2.000-.649i
3.0	1.222-.079i	2.000+.707i	2.000-.707i

ственно принять $y = 0$ и $\mu_1 = 1/f_1$, а значение μ_2 можно выбрать произвольно (например, равным μ_1).

Подставим в эти формулы моменты гамма-распределения с параметрами α и μ . Подкоренное выражение

$$D = \frac{\alpha^2(\alpha+1)(2-\alpha)(\alpha-1)^2}{18\mu^6}.$$

Таким образом, при $\alpha > 2$ аппроксимация должна иметь комплексные параметры. Случай $\alpha = 2$ является граничным: он соответствует распределению Эрланга второго порядка с $y = 1$ и $\mu_1 = \mu_2$.

В табл. 1 приведены параметры C_2 -аппроксимации гамма-распределения при единичном первом моменте. Таблица подтверждает сделанный прогноз и снимает опасения относительно возможной отрицательности $\{\mu_j\}$. Отметим парадоксальный, но ожидаемый результат: при $\alpha < 1$ вероятность выхода процесса во вторую фазу отрицательна.

■ Таблица 2. Распределение числа заявок в системе M/G/1

j	Параметр гамма-распределения							
	0.5		1.5		3.0		1e9	
	MG1	MCN	MG1	MCN	MG1	MCN	MG1	MCN
0	.300e-0	.300e-0	.300e-0	.300e-0	.300e-0	.300e-0	.300e-0	.300e-0
1	.165e-0	.167e-0	.233e-0	.233e-0	.263e-0	.263e-0	.304e-0	.306e-0
2	.120e-0	.118e-0	.159e-0	.159e-0	.174e-0	.173e-0	.190e-0	.187e-0
3	.912e-1	.903e-1	.106e-0	.106e-0	.106e-0	.106e-0	.101e-0	.101e-0
4	.706e-1	.703e-1	.696e-1	.696e-1	.637e-1	.638e-1	.517e-1	.524e-1
5	.550e-1	.550e-1	.457e-1	.456e-1	.379e-1	.380e-1	.263e-1	.267e-1
6	.430e-1	.431e-1	.299e-1	.299e-1	.226e-1	.226e-1	.134e-1	.135e-1
7	.336e-1	.338e-1	.196e-1	.196e-1	.134e-1	.134e-1	.682e-2	.682e-2
8	.264e-1	.265e-1	.128e-1	.128e-1	.798e-2	.798e-2	.347e-2	.344e-2
9	.207e-1	.208e-1	.842e-2	.841e-2	.474e-2	.474e-2	.177e-2	.173e-2
10	.162e-1	.167e-1	.551e-2	.551e-2	.282e-2	.281e-2	.899e-3	.873e-3
11	.127e-1	.128e-1	.361e-2	.361e-2	.168e-2	.167e-2	.457e-3	.440e-3
12	.996e-2	.100e-1	.237e-2	.237e-2	.996e-3	.993e-3	.233e-3	.222e-3
13	.782e-2	.784e-2	.155e-2	.155e-2	.592e-3	.590e-3	.118e-3	.118e-3
14	.613e-2	.614e-2	.101e-2	.101e-2	.352e-3	.350e-3	.603e-4	.562e-4
15	.481e-2	.481e-2	.665e-3	.666e-3	.209e-3	.208e-3	.307e-4	.283e-4
16	.377e-2	.377e-2	.436e-3	.436e-3	.124e-3	.124e-3	.156e-4	.143e-4
17	.296e-2	.296e-2	.285e-3	.286e-3	.740e-4	.735e-4	.794e-5	.719e-5
18	.232e-2	.232e-2	.187e-3	.187e-3	.440e-4	.436e-4	.404e-5	.362e-5

Приведем результаты расчета распределения числа заявок в системе M/G/1 (табл. 2). Процедура MG1 использовала общеизвестные соотношения, реализующие метод вложенных цепей Маркова на базе рекуррентного расчета распределения {q_j} числа заявок, прибывающих за случайное время обслуживания:

$$q_j = \int_0^{\infty} \frac{(\lambda t)^j}{j!} e^{-\lambda t} dB(t), \quad j = 0, 1, \dots,$$

для вырожденного и гамма-распределений [1]. Процедура MCN опиралась на C₂-аппроксимацию и решала описанные ниже векторно-матричные уравнения баланса. Индекс j указывает число заявок в системе. Параметр 1e9 фактически определяет вырожденное распределение, т. е. модель M/D/1.

Гамма-распределение было выбрано как удобный эталон по двум причинам:

1) параметры его плотности

$$b(t) = \frac{\mu(\mu t)^{\alpha-1}}{\Gamma(\alpha)} e^{-\mu t}$$

легко выражаются через среднее b₁ и дисперсию D согласно

$$\alpha = b_1^2 / D, \quad \mu = \alpha / b_1;$$

2) вышеупомянутые {q_j} для него вычисляются до удивления просто [1, с. 82]:

$$q_0 = \left(\frac{\mu}{\lambda + \mu} \right)^\alpha;$$

$$q_j = q_{j-1} \frac{\lambda}{\lambda + \mu} \frac{\alpha + j - 1}{j}, \quad j = 1, 2, \dots$$

Хорошее согласие прямого расчета модели $M/G/1$ и конечных результатов ее коковской аппроксимации в широком диапазоне условий, в том числе для комплексных и отрицательных параметров, убедительно свидетельствует о приемлемости обсуждаемой аппроксимации.

Достоинством C_2 -аппроксимации является естественное включение в нее как частных случаев M - и E_2 -моделей, а недостатками – трудность обобщения на большее число составляющих и усложнение диаграмм переходов.

Многофазное представление системы

Будем представлять состояние системы $C_2/C_2/n$ полным числом заявок в системе, фазой прибытия очередной заявки и расстановкой проходящих обслуживания заявок по его фазам («ключами» соответствующих микросостояний). Например, ключ (21) означает, что две заявки проходят первую фазу обслуживания с интенсивностями μ_1 , а одна – вторую с интенсивностью μ_2 . Диаграмма состояний разделяется по фазам прибытия на две вертикальные ветви с интенсивностями λ_1 и λ_2 соответственно, а по числу заявок в системе – на горизонтальные ярусы. Совокупности ключей в левой и правой ветвях диаграммы дублируются. На рис. 2 показаны переходы по завершению фазы обслуживания для половины диаграммы (для другой половины они тождественны).

Обозначим через S_j множество всех возможных микросостояний системы, при которых на обслуживании находится ровно j заявок, а через σ_j – количество элементов S_j . Нам потребуются следующие матрицы интенсивностей инфинитезимальных переходов:

$$A_j [\sigma_j \times \sigma_{j+1}] - \text{в } S_{j+1} \text{ (прибытие заявки),}$$

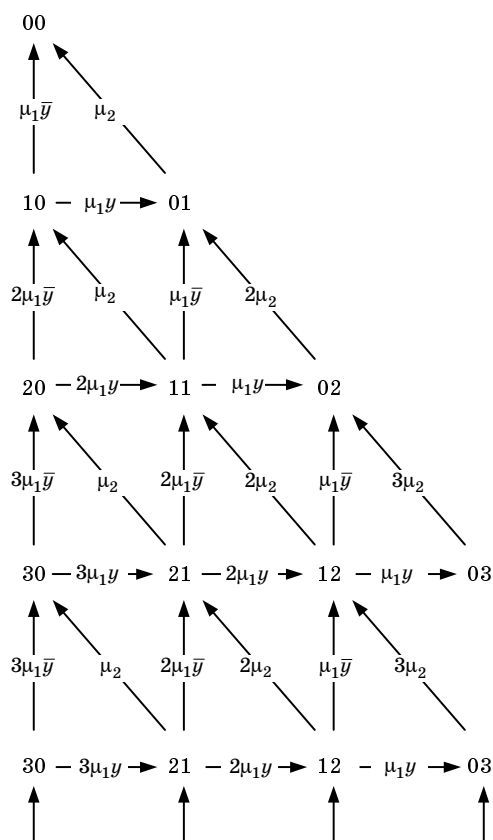
$$F_j [\sigma_j \times \sigma_j] - \text{в } S_j \text{ (конец промежуточной фазы обслуживания или прибытия заявки),}$$

$$B_j [\sigma_j \times \sigma_{j-1}] - \text{в } S_{j-1} \text{ (полное завершение обслуживания заявки),}$$

$$D_j [\sigma_j \times \sigma_j] - \text{ухода из состояний яруса } j.$$

В квадратных скобках здесь и далее указывается размер матриц. Элемент (i, k) любой из этих матриц представляет интенсивность перехода из i -го состояния j -го яруса в k -е состояние смежного (по переходам рассматриваемого типа) яруса. Заметим, что при $j > n$ все эти матрицы перестают зависеть от индекса. Ниже будут приведены примеры матриц для второго яруса.

Переходы по завершению фазы обслуживания не меняют фазы прибытия, что позволяет ограничиться изображением для них половины диаграммы (рис. 2). Здесь окончание первой фазы с интенсивностью $y\mu_1$ переводит заявку во вторую фазу обслуживания, а с интенсивностью $\bar{y}\mu_1$ – заверша-



■ Рис. 2. Интенсивности переходов по завершению обслуживания в фазе

ет обслуживание и поднимает изображающую точку на вышележащий ярус. Последнее при завершении второй фазы (интенсивность μ_2) происходит всегда. Интенсивности переходов проставлены с учетом числа заявок, находящихся в каждом фазе исходного микросостояния.

Например, для полной диаграммы переходов по завершению обслуживания («удвоенный» рис. 2) имеем клеточно-диагональную матрицу интенсивностей переходов со второго на первый ярус:

$$B_2 = \begin{bmatrix} 2\mu_1 \bar{y} & 0 & 0 & 0 \\ \mu_2 & \mu_1 \bar{y} & 0 & 0 \\ 0 & 2\mu_2 & 0 & 0 \\ 0 & 0 & 2\mu_1 \bar{y} & 0 \\ 0 & 0 & \mu_2 & \mu_1 \bar{y} \\ 0 & 0 & 0 & 2\mu_2 \end{bmatrix}.$$

Диаграмма переходов по фазам прибытия заявок в трехканальной системе оказывается весьма запутанной; поэтому мы ограничимся описанием логики ее построения. Если интервалы между заявками подчинены C_2 -распределению с параметра-

ми $\{u, \lambda_1, \lambda_2\}$, то все переходы из первой фазы с интенсивностями $u\lambda_1$ приводят в аналогичное микросостояние второй фазы прибытия того же яруса, а с интенсивностями $\bar{u}\lambda_1$ – в первую фазу нижележащего яруса. Переходы по завершению второй (всегда последней) фазы имеют интенсивность λ_2 и также приводят в первую фазу нижележащего яруса. Соответственно интенсивности «окончательных прибытий» описывает матрица

$$A_2 = \begin{bmatrix} \bar{u}\lambda_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \bar{u}\lambda_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \bar{u}\lambda_1 & 0 & 0 & 0 & 0 & 0 \\ \lambda_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda_2 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Переходы в пределах яруса могут быть связаны с выходом из первой во вторую фазу обслуживания (в пределах каждой ветви) либо с прибытием заявки (переход из левой ветви в одноименное микросостояние правой):

$$F_2 = \begin{bmatrix} 0 & 2\mu_1 y & 0 & u\lambda_1 & 0 & 0 \\ 0 & 0 & \mu_1 y & 0 & u\lambda_1 & 0 \\ 0 & 0 & 0 & 0 & 0 & u\lambda_1 \\ 0 & 0 & 0 & 0 & 2\mu_1 y & 0 \\ 0 & 0 & 0 & 0 & 0 & \mu_1 y \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Для диагональной матрицы D_2 суммарных интенсивностей ухода из микросостояний яруса мы приведем только ее диагональ:

$$\text{diag } D_2 = [\lambda_1 + 2\mu_1, \lambda_1 + \mu_1 + \mu_2, \lambda_1 + 2\mu_2, \lambda_2 + 2\mu_1, \lambda_2 + \mu_1 + \mu_2, \lambda_2 + 2\mu_2].$$

Уравнения глобального баланса и их решение

Получим стационарные вероятности микросостояний, представленных на диаграммах предыдущего раздела.

Введем векторы-строки $\gamma_j = \{\gamma_{j,1}, \gamma_{j,2}, \dots, \gamma_{j,\sigma_j}\}$ нахождения системы в состоянии $(j, i), j = 0, 1, \dots$. Теперь можно записать векторно-матричные уравнения баланса переходов между состояниями

$$\begin{aligned} \gamma_0 D_0 &= \gamma_0 F_0 + \gamma_1 B_1; \\ \gamma_j D_j &= \gamma_{j-1} A_{j-1} + \gamma_j F_j + \gamma_{j+1} B_{j+1}, \quad j = 1, 2, \dots \end{aligned} \quad (4)$$

Система (4), дополненная условием нормировки, даже для моделей с ограниченной очередью

характеризуется чрезвычайно высокой размерностью, и стандартные методы решения систем линейных алгебраических уравнений применительно к ней оказываются малоэффективными. Для расчета разомкнутых систем обслуживания можно применить восходящий к Ивэнсу [5] метод матрично-геометрической прогрессии. Этот метод основан на представлении векторов вероятностей микросостояний для $j > n$ формулой

$$\gamma_j = \gamma_n R^{j-n}, \quad j = n, n+1, \dots, \quad (5)$$

где R – некоторая матрица (знаменатель прогрессии). Перепишем второе из уравнений (4) с учетом (5) для значения j , при котором произошла вышеупомянутая стабилизация матриц, так что их индексы можно опустить:

$$\gamma_{j-1} R D = \gamma_{j-1} A + \gamma_{j-1} R F + \gamma_{j-1} R^2 B. \quad (6)$$

Теперь ясно, что R удовлетворяет матричному квадратному уравнению

$$R^2 B + R E + A = 0, \quad (7)$$

где $E = F - D$. Применим к его решению метод итераций. Будем искать поправку Δ к очередному приближению из уравнения (7), заменив в нем R на $R + \Delta$. Пренебрегая членом, содержащим квадрат поправки, имеем

$$R^2 B + \Delta R B + R \Delta B + R E + \Delta E + A \approx 0$$

или

$$\Delta(RB + E) + R \Delta B \approx -(R^2 B + RE + A).$$

Последнее уравнение приводится к уравнению

$$\Delta G + R \Delta B = H,$$

которое можно расписать как систему линейных алгебраических уравнений относительно компонент матричной поправки Δ и решить любым стандартным методом. В качестве начального приближения можно принять $R = \rho^{2/(v_A^2 + v_B^2)} I$, где ρ – коэффициент загрузки системы; v_A и v_B – коэффициенты вариации распределений интервалов между смежными заявками и длительности обслуживания соответственно; I – единичная матрица.

Далее усеченная система (4) для $j = 0, n$, в последнем из уравнений которой вектор γ_{n+1} заменен на $\gamma_n R$, расписывается относительно компонент векторов $\{\gamma_j\}, j = 0, n$. Затем одно из уравнений заменяется на условие баланса заявок

$$\sum_{j=0}^{n-1} (n-j) \gamma_j \mathbf{1}_j = n - b/a,$$

где $\mathbf{1}_j$ – вектор-столбец из σ_j единиц; a и b – средние интервал между заявками и длительность обслуживания. Наконец, находится решение системы для компонент начальных векторов.

Последующие векторы вероятностей микростояний вычисляются согласно формуле (5).

Для решения системы (4) можно также применить менее быстродействующий, но более универсальный итерационный метод, впервые предложенный Такахаси и Таками [7] и описанный в работах [1, 2] в более общей форме, с детальной проработкой расчетных зависимостей и с частными вариантами.

Расчет выходящего потока

В сетях обслуживания потоки на входе узлов формируются из выходящих потоков других узлов. В общем случае выходящий поток сохраняет среднюю интенсивность (и средний интервал между заявками) входящего, но меняет вид распределения интервалов между заявками. Последнее удобно характеризовать коэффициентами немарковости

$$\xi_i = d_i / d_1^i - i!, \quad i = 2, 3, \dots,$$

где $\{d_j\}$ – моменты распределения интервалов между заявками выходящего потока, который предполагается рекуррентным. Для простейшего потока упомянутые коэффициенты согласно данному определению равны нулю.

Ключом к проблеме преобразования потока в коковом узле обслуживания является расчет марковской системы.

Марковские системы. В системе $M/M/n$, загруженной полностью, частное преобразование Лапласа–Стилтьеса (ПЛС) интервала до очередного обслуживания при наличии в системе $j \geq n$ заявок

$$\delta_j(s) = n\mu / (n\mu + s), \quad j = n, n+1, \dots$$

При $j < n$ интенсивность обслуживания меняется с прибытием каждой очередной заявки. Будем интерпретировать s как параметр простейшего потока «катастроф». Тогда $\delta_j(s)$ можно считать вероятностью отсутствия катастроф за интервал от ухода заявки, оставившей в системе ровно j заявок, до следующего завершения обслуживания. Вероятности появления событий каждого из рассматриваемых простейших потоков (катастроф, завершений обслуживания и прибытия новых заявок) пропорциональны их интенсивностям, т. е. s , $j\mu$ и λ соответственно. Теперь ясно, что

$$\delta_j(s) = \frac{j\mu}{j\mu + \lambda + s} + \frac{\lambda}{j\mu + \lambda + s} \delta_{j+1}(s),$$

$$j = n-1, n-2, \dots, 0.$$

Здесь первое слагаемое соответствует завершению обслуживания, а второе – прибытию новой заявки (отчего число заявок в системе увеличилось на единицу) и отсутствию катастроф до следующего завершения обслуживания при новом начальном условии.

Результирующее распределение интервалов между заявками выходящего потока задается ПЛС

$$D(s) = \sum_{j=0}^{n-1} \pi_j \delta_j(s) + \left(1 - \sum_{j=0}^{n-1} \pi_j \right) \delta_n(s),$$

где финальные вероятности $\{\pi_j\}$ наличия в системе сразу после завершения обслуживания ровно j заявок при $t \rightarrow \infty$ подсчитываются согласно уравнению

$$\pi_j = \mu_{j+1} p_{j+1} / \sum_{i=1}^{\infty} \mu_i p_i, \quad j = 0, 1, \dots \quad (8)$$

Здесь $\{p_j\}$ – стационарное распределение числа заявок в системе, а μ_j равно $j\mu$ при $j < n$ и $n\mu$ – в противном случае. Знаменатель формулы (8) сводится к равенству

$$\sum_{i=1}^{\infty} \mu_i p_i = \mu \left[\sum_{i=1}^{n-1} i p_i + n p_n / (1 - \rho) \right].$$

Моменты искомого распределения можно получить многократным численным дифференцированием в нуле таблично заданной функции $D(s)$, $s = 0, h, 2h, \dots$ с последующей сменой знаков у нечетных производных. Однако предпочтительнее вычислять их непосредственно. Обозначим $E(x)$ набор моментов показательного распределения с параметром x , имеющий компоненты

$$E_i(x) = i! / x^i, \quad i = 1, 2, \dots$$

Символом $*$ будем обозначать свертку распределений в моментах. Для наборов моментов распределений между выходящими заявками сохраним индексацию прежних обозначений, заменив лишь $\delta(s)$ на d . В модели $M/M/n$ при числе заявок в системе $j \geq n$

$$d_j = E(n\mu), \quad j = n, n+1, \dots$$

Перепишем формулу для $\delta_j(s)$ при $j < n$ в виде

$$\delta_j(s) = \frac{j\mu + \lambda}{j\mu + \lambda + s} \left(\frac{j\mu}{j\mu + \lambda} \cdot 1 + \frac{\lambda}{j\mu + \lambda} \delta_{j+1}(s) \right) =$$

$$= \frac{j\mu}{j\mu + \lambda} \frac{j\mu + \lambda}{j\mu + \lambda + s} + \frac{\lambda}{j\mu + \lambda} \frac{j\mu + \lambda}{j\mu + \lambda + s} \delta_{j+1}(s).$$

Заменяя преобразование Лапласа показательных плотностей их моментами и произведение ПЛС – сверткой, убеждаемся, что

$$d_j = \frac{j\mu}{j\mu + \lambda} E(j\mu + \lambda) + \frac{\lambda}{j\mu + \lambda} E(j\mu + \lambda) * d_{j+1},$$

$$j = n-1, n-2, \dots, 0,$$

в частности:

$$d_0 = E(\lambda) * d_1.$$

Заметим, что свертка в моментах распределений A и B может быть выполнена на основе символического разложения

$$c^n = (a + b)^n \quad (9)$$

путем перевода показателей степени в индексы соответствующих моментов. Для выполнения сверток полезно составить автономную процедуру.

Результаты практически совпали с моментами распределения интервалов между входящими заявками, а коэффициенты немарковости с точностью до пяти знаков оказались нулевыми (выходящий поток – простейший). Это согласуется с классическим результатом Берке, позволяет считать основную идею расчета правильной и распространить ее на более сложные немарковские системы, для которых теория расчета выходящего потока до сих пор отсутствовала (в частности, на интересующую нас «дважды коксову»).

Коксова система. Введем обозначения для j -го яруса:

$M_{j,i}$ – суммарная интенсивность обслуживания в i -м микросостоянии каждой ветви;

$\widehat{M}_{j,i}$ – интенсивность переходов вверх по завершению обслуживания;

$M_{j,i}$ – интенсивность переходов вправо по завершению фазы обслуживания (сумма двух последних интенсивностей равна $M_{j,i}$);

$d_{j,i}^{(1)}, d_{j,i}^{(2)}$ – наборы моментов распределения времени до ближайшего обслуживания при соответствующем исходном состоянии (верхние индексы указывают левую и правую ветвь диаграммы соответственно);

σ_j – количество микросостояний в одной ветви диаграммы.

В этих обозначениях для $j \geq n$ и правой ветви диаграммы имеем

$$d_{n,i}^{(2)} = \frac{\widehat{M}_{n,i}}{M_{n,i} + \lambda_2} E(M_{n,i} + \lambda_2) + E(M_{n,i} + \lambda_2) * \left[\frac{\overline{M}_{n,i}}{M_{n,i} + \lambda_2} d_{n,i}^{(2)} + \frac{\lambda_2}{M_{n,i} + \lambda_2} d_{n,i}^{(1)} \right],$$

$$i = \sigma_n, \sigma_n - 1, \dots, 1.$$

Для тех же значений j и левой ветви диаграммы надо дополнительно учитывать возможные переходы в правую ветвь по завершению фазы прибытия заявки:

$$d_{n,i}^{(1)} = \frac{\widehat{M}_{n,i}}{M_{n,i} + \lambda_1} E(M_{n,i} + \lambda_1) + E(M_{n,i} + \lambda_1) * \left[\frac{\overline{M}_{n,i}}{M_{n,i} + \lambda_1} d_{n,i}^{(1)} + \frac{u\lambda_1}{M_{n,i} + \lambda_1} d_{n,i}^{(2)} + \frac{\bar{u}\lambda_1}{M_{n,i} + \lambda_1} d_{n,i}^{(1)} \right],$$

$$i = \sigma_n, \sigma_n - 1, \dots, 1.$$

После выполнения сверток в моментах согласно формуле (9) замечаем, что для каждого i искомые $\{d_{n,i}\}$ входят в обе части приведенных уравнений, причем «с перекрестом». Поэтому для их определения приходится последовательно решать σ_n систем из шести линейных алгебраических уравнений (по три момента в обеих ветвях на каждое микросостояние).

Для «ненасыщенных» ярусов $j = n-1, n-2, \dots, 1$ в правой ветви диаграммы прибытие заявки приводит в одноименные состояния левой ветви нижележащего яруса. Соответственно имеем рекуррентные формулы

$$d_{j,i}^{(2)} = \frac{\widehat{M}_{j,i}}{M_{j,i} + \lambda_2} E(M_{j,i} + \lambda_2) + E(M_{j,i} + \lambda_2) * \left[\frac{\overline{M}_{j,i}}{M_{j,i} + \lambda_2} d_{j,i+1}^{(2)} + \frac{\lambda_2}{M_{j,i} + \lambda_2} d_{j+1,i}^{(1)} \right],$$

$$i = \sigma_j, \sigma_j - 1, \dots, 1.$$

Для левой ветви зоны ненасыщенных состояний завершение фазы прибытия заявки может привести как в правую ветвь диаграммы, так и в нижележащий ярус:

$$d_{j,i}^{(1)} = \frac{\widehat{M}_{j,i}}{M_{j,i} + \lambda_1} E(M_{j,i} + \lambda_1) + E(M_{j,i} + \lambda_1) * \left[\frac{\overline{M}_{j,i}}{M_{j,i} + \lambda_1} d_{j,i+1}^{(1)} + \frac{u\lambda_1}{M_{j,i} + \lambda_1} d_{j,i}^{(2)} + \frac{\bar{u}\lambda_1}{M_{j,i} + \lambda_1} d_{j+1,i}^{(1)} \right],$$

$$i = \sigma_j, \sigma_j - 1, \dots, 1.$$

Здесь первое слагаемое соответствует непосредственному завершению обслуживания, а последующие – переходам по завершению первой фазы обслуживания и фазы прибытия заявки (окончательной или с переходом во вторую фазу ожидания прибытия) с рекуррентным учетом моментов распределения времени до ближайшего обслуживания во вновь возникшем состоянии.

Наконец, в случае полного освобождения системы до очередного завершения обслуживания сначала придется дожидаться прибытия хотя бы одной заявки. Здесь

$$d_{0,1}^{(2)} = E(\lambda_2) * d_{1,1}^{(1)};$$

$$d_{0,1}^{(1)} = E(\lambda_1) * (ud_{0,1}^{(2)} + \bar{u}d_{1,1}^{(1)})d_{1,1}^{(1)}.$$

В «ненасыщенной» зоне искомые наборы моментов определяются последовательно через ранее вычисленные: для каждого i сначала вычисляются «правые», а затем «левые» моменты.

Распределение времени ожидания

Интересующее нас распределение можно задать его первым моментом \tilde{w}_1 и набором коэффициентов немарковости. Тогда высшие моменты можно вычислить по формуле

$$\tilde{w}_i = \tilde{w}_1^i (\xi_i + i!), \quad i = 2, 3, \dots \quad (10)$$

Известно, что условное распределение времени ожидания заявки в системе $GI/M/n$ подчинено показательному закону. Это обстоятельство позволяет считать, что вид распределения времени ожидания не зависит от распределения интервалов между смежными заявками и числа каналов и определяется только распределением времени обслуживания. В работе [4] предложен конструктивный метод решения этой задачи, основанный на сохранении коэффициентов немарковости системы $M/G/1$.

Проблемы программной реализации

Зависимость структуры матриц переходов от типа и порядка аппроксимирующих фазовых распределений ставит ценность программной реализации расчетной схемы в прямую зависимость от возможности автоматического построения матриц переходов. Эту проблему в общем случае можно решить следующим образом:

- для каждого j -го яруса системы, $j = \overline{0, n}$, автоматически сгенерировать последовательность ключей микросостояний;
- формировать матрицы переходов, сопоставляя «ключи-источники» j -го яруса и совместимые с ними по выбранному типу переходов ключи-«преемники» для матрицы B_j на $(j-1)$ -м ярусе, для F_j – на j -м, для A_j – на $(j+1)$ -м.

В обсуждаемом случае C_2 -аппроксимации структура матриц и правила их формирования относительно очевидны. Однако программная реализация метода оказалась значительно сложнее, чем это представляется при ознакомлении с его идеей. Дело в том, что метод предполагает работу с переменным числом матриц перехода изменяемой размерности, определяемой в процессе счета. Соответственно при программировании на Фортране пришлось упаковывать матрицы каждого вида (и их половинки для треугольных матриц) в линейные массивы, выделять под эти объекты динамическую ([allocatable]) память, фиксировать отдельно их размеры и координаты в линейных структурах, явно переадресовывать операнды матричных операций.

На эти проблемы наложились порождаемые специфической структурой упомянутых матриц: A – блочная с ненулевыми блоками в виде диагональных матриц, B – блочно-диагональная с одинаковыми блоками, F – верхняя треугольная, D – диагональная.

Учет этих особенностей при программировании требуемых матричных операций оказался нетривиальным и весьма поучительным.

Тестирование метода

Для проверки корректности метода и его программной реализации оценивалась средняя длина очереди (показатель компактный, наглядный, чувствительный к входным данным и инвариантный к масштабу времени) при коэффициенте загрузки системы $\rho = 0,7$ для различных видов исходных распределений и числа каналов. Для моделей (в обозначениях Кендалла) $M/M/1$, $M/E_3/1$, $M/D/1$, $M/\Gamma_{1,5}/1$, $E_4/M/1$, $D/M/1$, $\Gamma_{1,5}/M/1$, $M/M/2$, $M/D/2$, $M/E_3/2$, $E_4/D/2$ были получены соответственно значения 1,633, 1,089, 0,817, 1,361, 0,861, 0,599, 1,292, 1,345, 0,692, 0,911, 0,113. Эти результаты практически совпали с полученными ранее посредством других численных методов, в том числе не использующих фазовые аппроксимации (например, метода Кроммелина для систем $M/D/n$, Быкадорова для $E_k/D/n$, Такача для $GI/M/n$).

Правильность расчета выходящего потока подтверждается:

- равенством средних интервалов между заявками входящего и выходящего потоков;
- получением практически нулевых коэффициентов немарковости выходящего потока для моделей вида $M/M/n$, обработанных по общей схеме;
- приближением этих коэффициентов к такому для распределения времени обслуживания – в случае одноканальных систем при устремлении коэффициента загрузки к единице;
- совпадением результатов обсчета вышеперечисленных моделей с полученными посредством менее универсальной аппроксимации $H_2/H_2/n$.

Частные случаи и обобщения

Самостоятельную ценность имеет наиболее типичный частный случай рассмотренной модели – с простейшим входящим потоком. В этом случае диаграмма переходов сводится к одной ветви и программная реализация сильно упрощается.

Как обобщение алгоритма можно рассматривать его вариант для системы с ограниченной очередью. В этом случае в микросостояния нижнего яруса диаграммы нет переходов снизу. Соответственно для этого яруса корректируется матрица $D-F$ и отпадает надобность в расчете одного из вспомогательных векторов итерационного метода (β_j''). Для определения вероятности свободного состояния на заключительном этапе итерационного метода здесь условие баланса заявок заменяется условием нормировки вероятностей. Дальнейшим обобщением является расчет замкнутой системы с интенсивностью входящего (простейшего) потока, зависящей от числа заявок в системе – в типичном случае линейно убывающей до нуля. Здесь дополнительно модифицируются числовые значения матриц.

Заключение

Конкретные результаты данной статьи сводятся к следующим.

1. Предложен алгоритм расчета параметров распределения Кокса C_2 , выравнивающий три заданных момента.

2. Проведено исследование алгоритма; выявлена область входных данных, порождающая «патологические» вероятности переходов (комплексные и отрицательные); показано, что при обратном переходе от вероятностей микросостояний к укрупненным вероятностям патология исчезает и конечные результаты согласуются с полученными существенно иными методами. Указаны преимущества C_2 -аппроксимации в сравнении с другими фазовыми распределениями.

3. Впервые предложены диаграммы переходов для «дважды коксовой» системы, на основе которой получены матрицы интенсивностей переходов.

4. Разработан способ расчета распределения интервалов между заявками выходящего потока,

в частности – непосредственно моментов этого распределения.

5. Даны рекомендации по программной реализации основных этапов алгоритма.

Эти результаты вместе с методами решения уравнений баланса и расчета временных характеристик позволяют, наконец, сравнительно просто и достаточно точно проводить полный расчет многоканальных систем с произвольными распределениями интервалов между заявками и времени обслуживания.

В сравнении с имитационным моделированием метод существенно выигрывает по объему и точности вычислений. Он избавляет от изучения и использования других фазовых аппроксимаций и (если не говорить о подготовке будущих теоретиков) – вообще всех прочих методов расчета немарковских систем обслуживания. Метод может использоваться как составная часть процедур расчета сетей обслуживания методом потокоэквивалентной декомпозиции [1].

Литература

1. Рыжиков Ю. И. Теория очередей и управление запасами: Учебник. СПб.: Питер, 2001. 384 с.
2. Рыжиков Ю. И. Алгоритм расчета многоканальной системы с эрланговским обслуживанием // Автоматика и телемеханика. 1980. № 5. С. 30–37.
3. Рыжиков Ю. И., Хомоненко А. Д. Итеративный метод расчета многоканальных систем с произвольным распределением времени обслуживания // Проблемы управления и теории информации. 1980. № 3. С. 32–38.
4. Рыжиков Ю. И. Три метода расчета временных характеристик разомкнутых систем массового обслуживания // Автоматика и телемеханика. 1993. № 2. С. 127–133.
5. Evans R. D. Geometric Distribution in Some Two Dimensional Queuing Systems // Operat. Res. 1967. Vol. 15. N 5. P. 830–846.
6. Khomonenko A. D., Bubnov V. P. A Use of Coxian Distribution for Iterative Solution of $M/G/n/R \leq \infty$ queuing systems // Problems of Control and Information Theory. 1985. Vol. 14. N 2. P. 143–153.
7. Takahashi Y., Takami Y. A Numerical Method for the Steady-state Probabilities of a GI/G/c Queuing System in a General Class // J. of the Operat. res. soc. of Japan. 1976. Vol. 19. N 2. P. 147–157.