

© 2020 г. А.З. МЕЛИКОВ, чл.-корр. НАН Азербайджана
(agassi.melikov@gmail.com)
(Институт систем управления НАН Азербайджана, Баку),
С.Г. АЛИЕВА, канд. техн. наук (s@aliyeva.info)
(Бакинский государственный университет),
М.О. ШАХМАЛЫЕВ (mamed.shahmaliyev@gmail.com)
(Институт систем управления НАН Азербайджана, Баку)

МЕТОДЫ РАСЧЕТА СИСТЕМЫ С МГНОВЕННОЙ ОБРАТНОЙ СВЯЗЬЮ И ПЕРЕМЕННОЙ ИНТЕНСИВНОСТЬЮ ВХОДЯЩЕГО ПОТОКА

Изучена марковская модель системы обслуживания с одним сервером и мгновенной обратной связью. После завершения обслуживания часть вызовов согласно схеме Бернулли либо покидают систему, либо мгновенно возвращается для получения повторного обслуживания, при этом для повторного обслуживания потребуется положительное (случайное) время переключения сервера. Интенсивность входящего потока зависит от состояния сервера, который может быть в рабочем режиме или в режиме переключения. Найдено условие эргодичности соответствующей двумерной цепи Маркова и предложены три метода для ее исследования: метод производящих функций, метод спектрального расширения и метод фазового укрупнения состояний. Приводятся результаты численных экспериментов.

Ключевые слова: система обслуживания, мгновенная обратная связь, метод расчета.

DOI: 10.31857/S0005231020090056

1. Введение

В последние годы системы массового обслуживания (СМО) с обратной связью все больше привлекают внимание исследователей. Это объясняется тем, что учет эффекта обратной связи позволяет повысить адекватность изучаемых моделей к реальной ситуации, где получившие один раз обслуживание заявки могут потребовать повторного обслуживания по целому ряду различных причин. Так, например, в коммуникационных сетях ошибочно переданные данные передаются повторно. Другим примером служат производственные системы, где бракованные детали требуют повторной обработки. Третьим примером могут быть современные сервисные интеграторы, в которых получивший один раз качественное обслуживание клиент может еще раз обратиться для обслуживания именно к ним.

В классе СМО с обратной связью различают системы с мгновенной и отсроченной (задерживающей) обратной связью. История исследования этих систем берет свое начало от пионерских работ Такача [1, 2]. После опубликования указанных работ подобные системы некоторое время не привлекали

внимания исследователей СМО, исследования в этом направлении получили интенсивное продолжение лишь в последние два десятилетия. Здесь наряду с прочими следует отметить работы [3–13]. Подробный обзор работ можно найти в [12, 13], и поэтому здесь не будем останавливаться на изложении известных фактов.

В указанных выше работах, как правило, предполагается, что сервер может мгновенно начинать обслуживание заявки, которая требует повторного обслуживания. Вместе с тем это предположение зачастую плохо соотносится с действительностью. На практике зачастую для начала процесса обслуживания заявок указанного типа потребуется некоторое время “разогрева” сервера, т.е. время переключения сервера для обслуживания этих заявок является положительной случайной величиной (с.в.). Подобные ситуации, как правило, наблюдаются в производственных системах, где для обработки бракованной детали потребуется некоторое время переналадки сервера.

Здесь изучается модель СМО с одним сервером и мгновенной обратной связью, в которой время переключения сервера для повторного обслуживания является положительной с.в. Изучаются модели с неограниченной общей очередью для первичных заявок и заявок, которые требуют повторного обслуживания. Предлагаются и сравниваются три метода расчета характеристик модели: метод производящих функций, метод спектрального расширения [14, 15] и метод фазового укрупнения состояний многомерных цепей Маркова (ЦМ) [12, 13].

Отметим, что при соответствующем изменении термина “переключение сервера” получаются модели систем: а) с прогулками сервера, где независимо от длины очереди вызовов прогулки могут быть осуществлены после каждого акта обслуживания; б) с ненадежным сервером, где поломки сервера возможны лишь в моменты завершения обслуживания вызовов. Полученные в данной работе результаты могут быть использованы и для анализа указанных систем.

2. Описание модели системы и постановка задачи

Рассматривается система обслуживания с одним сервером и неограниченным числом мест для ожидания. На вход системы извне поступает пуассоновский поток первичных вызовов. Времена обслуживания этих вызовов являются независимыми и одинаково распределенными (н.о.р.) с.в. и имеют общую показательную функцию распределения со средним μ^{-1} . После завершения процесса обслуживания вызовы независимо друг от друга и согласно схеме Бернулли либо с вероятностью σ мгновенно требуют повторного обслуживания, либо с дополнительной вероятностью $1 - \sigma$ покидают систему. Времена обслуживания повторных вызовов также являются н.о.р. с.в. с тем же средним μ^{-1} .

Для начала процесса обслуживания вызовов, которые требуют повторного обслуживания, требуется некоторое положительное случайное время (время переключения сервера), которое имеет показательную функцию распределения со средним θ^{-1} . Иными словами, в любой момент времени сервер может находиться в одном из двух состояний: в рабочем режиме или в режиме пе-

реключения. Считается, что не допускается прерывать время переключения сервера. При этом вызов, который требует повторного обслуживания, из-за запрета на прерывание переключения сервера остается на нем.

Первичные вызовы имеют информацию о статусе сервера, т.е. если сервер находится в режиме переключения, то интенсивность p -вызовов равна λ_0 , иначе она равна λ_1 .

Считается, что случайные процессы поступления вызовов, их обслуживания и времена переключения сервера являются не зависимыми друг от друга.

Требуется найти совместное распределение числа вызовов в системе и состояние сервера (задача А), а также среднее число вызовов в системе и производительность системы, т.е. интенсивность потока обслуженных вызовов, выходящих из системы (задача В).

3. Метод производящих функций

Состояние системы в произвольный момент времени определяется двумерным вектором (n, k) , где n — число вызовов в системе, k — состояние сервера, т.е.

$$k = \begin{cases} 0, & \text{если сервер находится в режиме переключения,} \\ 1, & \text{если сервер находится в рабочем режиме.} \end{cases}$$

Работа системы описывается двумерной ЦМ со следующим пространством состояний:

$$(3.1) \quad E = E_0 \bigcup_{n=1}^{\infty} E_n,$$

где $E_0 = \{(0, 1)\}$, $E_n = \{(n, 0), (n, 1)\}$, $n = 1, 2, \dots$

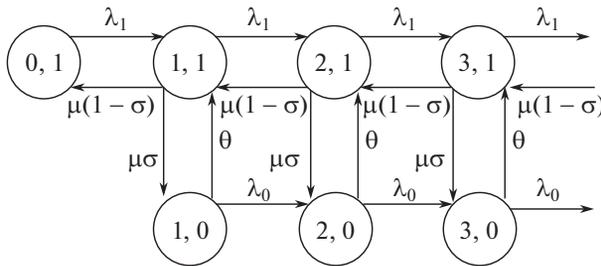


Рис. 1. Граф переходов.

Интенсивность перехода из состояния (n, k) в состояние (n', k') обозначим через $q((n, k), (n', k'))$. Эти величины определяются как (см. рис. 1):

$$(3.2) \quad q((n, 1), (n + 1, 1)) = \lambda_1;$$

$$(3.3) \quad q((n, 1), (n - 1, 1)) = \mu(1 - \sigma), \quad n > 0;$$

$$(3.4) \quad q((n, 1), (n, 0)) = \mu\sigma, \quad n > 0;$$

$$(3.5) \quad q((n, 0), (n, 1)) = \theta;$$

$$(3.6) \quad q((n, 0), (n + 1, 0)) = \lambda_0.$$

Пусть $p(n, k)$ обозначает стационарную вероятность состояния $(n, k) \in E$ (условие эргодичности устанавливается ниже). Эти величины удовлетворяют следующей бесконечной системе уравнений равновесия (СУР):

$$(3.7) \quad \lambda_1 p(0, 1) = \mu(1 - \sigma)p(1, 1);$$

Для случаев $n \geq 1$:

$$(3.8) \quad (\lambda_1 + \mu)p(n, 1) = \mu(1 - \sigma)p(n + 1, 1) + \lambda_1 p(n - 1, 1) + \theta p(n, 0);$$

$$(3.9) \quad (\lambda_0 + \theta)p(1, 0) = \mu\sigma p(1, 1);$$

Для случаев $n \geq 2$:

$$(3.10) \quad (\lambda_0 + \theta)p(n, 0) = \lambda_0 p(n - 1, 0) + \mu\sigma p(n, 1).$$

Введем следующие производящие функции:

$$P_k(z) = \sum_{n=\bar{k}}^{\infty} p(n, k) z^n, \quad k = 0, 1,$$

$$\text{где } \bar{k} = \begin{cases} 0, & \text{если } k = 1, \\ 1, & \text{если } k = 0. \end{cases}$$

Из (3.8) имеем

$$(3.11) \quad \begin{aligned} &(-\lambda_1 z^2 + (\lambda_1 + \mu)z - \mu(1 - \sigma)) P_1(z) - \theta z P_0(z) = \\ &= ((\lambda_1 + \mu)z - \mu(1 - \sigma)) p(0, 1) - \mu(1 - \sigma) z p(1, 1). \end{aligned}$$

С учетом (3.7) из (3.11) получаем

$$(3.12) \quad \begin{aligned} &(-\lambda_1 z^2 + (\lambda_1 + \mu)z - \mu(1 - \sigma)) P_1(z) - \theta z P_0(z) = \\ &= \mu(z - 1 + \sigma) p(0, 1). \end{aligned}$$

Аналогичным образом из уравнений (3.9) и (3.10) имеем

$$(3.13) \quad -\mu\sigma P_1(z) + (\theta + \lambda_0(1 - z)) P_0(z) = -\mu\sigma p(0, 1).$$

Из (3.13) получаем

$$(3.14) \quad P_1(z) = p(0, 1) + a(z) P_0(z),$$

где

$$a(z) = \frac{1}{\mu\sigma} (-\lambda_0 z + \lambda_0 + \theta).$$

С учетом (3.12) из (3.14) получаем

$$(3.15) \quad P_0(z) = \frac{\lambda_1 z^2 - \lambda_1 z}{a(z)b(z) - \theta z} p(0, 1),$$

где

$$b(z) = -\lambda_1 z^2 + (\lambda_1 + \mu)z - \mu(1 - \sigma).$$

Из (3.14) и (3.15) имеем

$$(3.16) \quad P_1(z) = \left(1 + a(z) \frac{\lambda_1 z^2 - \lambda_1 z}{a(z)b(z) - \theta z}\right) p(0, 1).$$

Для нахождения неизвестной величины $p(0, 1)$ используем следующее условие: $\lim_{z \rightarrow 1} (P_0(z) + P_1(z)) = 1$. При вычислении пределов получается неопределенность вида $\frac{0}{0}$, и согласно правилу Лопиталья находим, что

$$(3.17) \quad p(0, 1) = \left(1 + \frac{\lambda_1(\theta + \mu\sigma)}{\theta\mu(1 - \sigma) - \lambda_1\theta - \lambda_0\mu\sigma}\right)^{-1}.$$

Замечание 1. При выводе формулы (3.17) находим следующее необходимое условие эргодичности системы:

$$(3.18) \quad \lambda_1\theta + \lambda_0\mu\sigma < \theta\mu(1 - \sigma).$$

Далее, при выполнении условия эргодичности (3.18) условные средние значения числа вызовов в системе, когда сервер находится в рабочем режиме (L_1) и в режиме переключения (L_0), а также безусловное значение среднего числа вызовов в системе (L) определяются как

$$(3.19) \quad L_1 = P_1'(1), \quad L_0 = P_0'(1), \quad L = L_1 P_1(1) + L_0 P_0(1).$$

Производительность системы вычисляется следующим образом:

$$(3.20) \quad \Lambda = \mu(1 - \sigma)(P_1(1) - p(0, 1)).$$

Ввиду очевидности получения и громоздкости явный вид формул (3.19) и (3.20) здесь не приводится (лишь отметим, что для его получения правило Лопиталья применяется 2 раза).

Таким образом, метод производящих функций позволяет решить задачу В, но не позволяет решить задачу А.

4. Метод спектрального расширения

Применение метода спектрального расширения [14, 15] позволяет решить обе задачи, А и В. Поскольку этот метод подробно описан в указанных работах, то здесь описываются лишь его основные этапы.

Вводятся следующие матрицы:

A — элементы этой матрицы определяют одношаговые вертикальные переходы между состояниями (см. рис. 1), т.е.

$$A = \begin{pmatrix} 0 & q((n,0), (n,1)) \\ q((n,1), (n,0)) & 0 \end{pmatrix} = \begin{pmatrix} 0 & \theta \\ \mu\sigma & 0 \end{pmatrix};$$

B — элементы этой матрицы определяют одношаговые горизонтальные переходы вправо (см. рис. 1), т.е.

$$B = \begin{pmatrix} q((n,0), (n+1,0)) & 0 \\ 0 & q((n,1), (n+1,1)) \end{pmatrix} = \begin{pmatrix} \lambda_0 & 0 \\ 0 & \lambda_1 \end{pmatrix},$$

$$B_0 = \begin{pmatrix} 0 & 0 \\ 0 & q(n,1), (n+1,1) \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & \lambda_1 \end{pmatrix};$$

C — элементы этой матрицы определяют одношаговые горизонтальные переходы влево (см. рис. 1), т.е.

$$C = \begin{pmatrix} 0 & 0 \\ 0 & q((n,1), (n-1,1)) \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & \mu(1-\sigma) \end{pmatrix}.$$

Считается, что A_0 и C_0 являются нулевыми матрицами. На базе указанных выше матриц определяются диагональные матрицы D^A , D^B и D^C , в которых элементы главных диагоналей равны сумме элементов соответствующих строк матриц A , B и C , т.е.

$$D^A = \begin{pmatrix} \theta & 0 \\ 0 & \mu\sigma \end{pmatrix}, \quad D^B = \begin{pmatrix} \lambda_0 & 0 \\ 0 & \lambda_1 \end{pmatrix}, \quad D^C = \begin{pmatrix} 0 & 0 \\ 0 & \mu(1-\sigma) \end{pmatrix}.$$

Балансовые уравнения для вероятностей состояний из однородной части графа (см. рис. 1) с учетом введенных матриц записывается так:

$$(4.1) \quad \mathbf{v}_n (D^A + D^B + D^C) = \mathbf{v}_{n-1}B + \mathbf{v}_nA + \mathbf{v}_{n+1}C, \quad n \geq 1,$$

где

$$\mathbf{v}_n = (p(n,0), p(n,1)), \quad n \geq 1.$$

Уравнения (4.1) представляются в виде матрично-разностных уравнений второго порядка с постоянными коэффициентами:

$$(4.2) \quad \mathbf{v}_n Q_0 + \mathbf{v}_{n+1} Q_1 + \mathbf{v}_{n+2} Q_2 = 0, \quad n \geq 1,$$

где

$$Q_0 = B, \quad Q_1 = A - (D^A + D^B + D^C), \quad Q_2 = C.$$

Характеристический матричный полином уравнения (4.2) имеет следующий вид:

$$(4.3) \quad Q(\eta) = Q_0 + Q_1\eta + Q_2\eta^2.$$

Далее собственные значения η_0, η_1 и соответствующие им левые собственные векторы $\psi = (\psi_0(k), \psi_1(k))$, $k = 0, 1$, матрицы $Q(\eta)$ определяются из следующих уравнений:

$$(4.4) \quad \det(Q(\eta)) = 0,$$

$$(4.5) \quad \psi Q(\eta) = \mathbf{0},$$

где правая часть уравнения (4.5) является нулевой вектор-строкой.

В указанных выше работах разработаны эффективные алгоритмы решения уравнений (4.4) и (4.5). После нахождения решений уравнений (4.4) и (4.5) вероятности состояний определяются как

$$(4.6) \quad p(n, k) = a_0\psi_0(k)\eta_0^n + a_1\psi_1(k)\eta_1^n, \quad n = 0, 1, \dots; \quad k = 0, 1,$$

где a_0, a_1 являются параметрами, которые определяются с помощью балансовых уравнений для вероятностей состояний из неоднородной части графа (способ их нахождения также подробно описан в [14, 15]).

С помощью данного метода можно найти необходимое и достаточное условие эргодичности системы. Формулируется оно следующим образом [16]: цепь Маркова является эргодической тогда и только тогда, когда в однородной части графа суммарные интенсивности переходов вправо должны быть меньше, чем суммарные интенсивности переходов влево, т.е. требуется выполнение условия

$$(4.7) \quad \mathbf{v}B\mathbf{e} < \mathbf{v}C\mathbf{e},$$

где $\mathbf{e} = (1, 1)^T$, и $\mathbf{v} = (\nu_0, \nu_1)$ представляет собой вектор вероятностей состояний ЦМ с двумя состояниями и инфинитизимальным генератором A , т.е. $\nu_0 = \frac{\mu\sigma}{\theta + \mu\sigma}$, $\nu_1 = \frac{\theta}{\theta + \mu\sigma}$. Тогда из (4.7) получаем условие эргодичности, которое совпадает с условием (3.18).

Условные средние значения числа вызовов в системе, когда сервер находится в рабочем режиме (L_1) и в режиме переключения (L_0), определяются как

$$(4.8) \quad L_1 = \sum_{n=1}^{\infty} np(n, 1); \quad L_0 = \sum_{n=1}^{\infty} np(n, 0).$$

Безусловное значение среднего числа вызовов в системе (L) определяется как

$$(4.9) \quad L = L_1 \sum_{n=0}^{\infty} p(n, 1) + L_0 \sum_{n=1}^{\infty} p(n, 0).$$

Производительность системы вычисляется так:

$$(4.10) \quad \Lambda = \mu(1 - \sigma) \sum_{n=1}^{\infty} p(n, 1).$$

5. Метод фазового укрупнения

Хотя метод спектрального расширения позволяет решить обе задачи А и В, иногда при определенных значениях исходных данных системы появляются проблемы вычислительной неустойчивости. Ниже рассматривается применение метода фазового укрупнения состояний двумерных ЦМ [12, 13] для решения поставленных задач А и В. Он корректно может быть применен для систем, в которых время переключения сервера существенно меньше, чем интервалы между поступлениями вызовов, т.е. предполагается, что имеет место соотношение $\theta \gg \max(\lambda_0, \lambda_1)$. При выполнении этого условия интенсивности переходов внутри классов E_n (см. (3.1) и рис. 1) существенно больше, чем интенсивности переходов между состояниями из разных классов.

Согласно алгоритму метода фазового укрупнения состояний двумерных ЦМ, каждый класс E_n , $n = 0, 1, 2, \dots$ представляется в виде одного укрупненного состояния $\langle n \rangle$, $n = 0, 1, 2, \dots$. При этом класс E_0 содержит лишь одно состояние $(0, 1)$, а каждый класс E_n , $n > 0$ содержит два состояния $(n, 1)$ и $(n, 1)$.

Вероятности состояний внутри классов E_n , $n = 0, 1, 2, \dots$ обозначаются через $\rho_n(k)$, $k = 0, 1$. Поскольку класс E_0 содержит лишь одно состояние $(0, 1)$, то имеем $\rho_0(1) = 1$.

Интенсивности переходов между состояниями класса E_n , $n > 0$ определяются из (3.4) и (3.5), т.е. вероятности состояний внутри каждого класса E_n , $n > 0$ не зависят от индекса n и вычисляются как

$$(5.1) \quad \rho_n(0) = \frac{\mu\sigma}{\theta + \mu\sigma}, \quad \rho_n(1) = \frac{\theta}{\theta + \mu\sigma}.$$

Замечание 2. Поскольку вероятности состояний внутри классов E_n , $n = 0, 1, 2, \dots$, $\rho_n(k)$, $k = 0, 1$, не зависят от индекса n (см. (5.1)), то далее у этих величин индекс опускается.

С учетом (5.1) находим, что интенсивности переходов между укрупненными состояниями $\langle n \rangle$, $n = 0, 1, 2, \dots$ определяются следующим образом:

$$(5.2) \quad q(\langle 0 \rangle, \langle 1 \rangle) = \lambda_1.$$

Для случаев $n > 0$

$$(5.3) \quad q(\langle n \rangle, \langle n + 1 \rangle) = \tilde{\lambda}; \quad q(\langle n \rangle, \langle n - 1 \rangle) = \tilde{\mu}.$$

Здесь $\tilde{\lambda} = \frac{1}{\theta + \mu\sigma}(\lambda_1\theta + \lambda_0\mu\sigma)$, $\tilde{\mu} = \frac{\theta}{\theta + \mu\sigma}\mu(1 - \sigma)$.

Из соотношений (5.2) и (5.3) заключаем, что укрупненная модель описывается бесконечным процессом размножения — гибели с переменными параметрами. Условием эргодичности этого процесса является $\tilde{\lambda} < \tilde{\mu}$ или $\lambda_1\theta + \lambda_0\mu\sigma < \theta\mu(1 - \sigma)$.

Замечание 3. Полученное здесь условие эргодичности полностью совпадает с условием (3.18), найденным выше с помощью метода производящих функций. Там трудно было дать интерпретацию этого условия. Здесь, однако, можно дать его вероятностную интерпретацию. Когда сервер находится в

рабочем режиме, интенсивность поступления вызовов равна $\lambda_1 \rho(1)$, а когда сервер находится в режиме переключения, эта величина равна $\lambda_0 \rho(0)$. Иными словами, $\tilde{\lambda}$ представляет собой суммарную интенсивность поступления вызовов при различных режимах работы сервера. Вызовы обслуживаются лишь тогда, когда сервер находится в рабочем режиме, т.е. интенсивность обслуживания равна $\mu(1 - \sigma) \rho(1)$. Следовательно, условие эргодичности (3.18) имеет простую вероятностную интерпретацию: суммарная интенсивность поступления вызовов при различных режимах работы сервера должна быть меньше, чем интенсивность обслуживания.

Из соотношений (5.2) и (5.3) вычисляются вероятности укрупненных состояний $\pi(< n >)$, $n \geq 0$:

$$(5.4) \quad \pi(< n >) = \frac{\lambda_1}{\tilde{\lambda}} \left(\frac{\tilde{\lambda}}{\tilde{\mu}} \right)^n \pi(< 0 >), \quad n \geq 1,$$

где

$$\pi(< 0 >) = \left(1 + \frac{\lambda_1}{\tilde{\lambda}} \frac{\alpha}{1 - \alpha} \right)^{-1}, \quad \alpha = \frac{\tilde{\lambda}}{\tilde{\mu}}.$$

С учетом (5.1)–(5.4) находим приближенные значения вероятностей состояний исходной ЦМ:

$$(5.5) \quad \tilde{p}(0, 1) = \pi(< 0 >);$$

$$(5.6) \quad \tilde{p}(n, 0) = \rho(0) \pi(< n >);$$

$$(5.7) \quad \tilde{p}(n, 1) = \rho(1) \pi(< n >).$$

Приближенные значения (ПЗ) характеристики (4.8)–(4.10) определяются как

$$(5.8) \quad L_1 \approx \sum_{n=1}^{\infty} np(n, 1) = \rho(1) \sum_{n=1}^{\infty} n\pi(< n >) = \frac{\theta}{\theta + \mu\sigma} \frac{\lambda_1}{\tilde{\lambda}} \frac{\alpha}{(1 - \alpha)^2} \pi(< 0 >);$$

$$(5.9) \quad L_0 \approx \sum_{n=1}^{\infty} np(n, 0) = \rho(0) \sum_{n=1}^{\infty} n\pi(< n >) = \frac{\mu\sigma}{\theta + \mu\sigma} \frac{\lambda_1}{\tilde{\lambda}} \frac{\alpha}{(1 - \alpha)^2} \pi(< 0 >);$$

$$(5.10) \quad L \approx \frac{\theta}{\theta + \mu\sigma} L_1 + \frac{\mu\sigma}{\theta + \mu\sigma} L_0.$$

$$(5.11) \quad \Lambda \approx \mu(1 - \sigma) \sum_{n=1}^{\infty} \rho_n(1) \pi(< n >) = \mu(1 - \sigma) \rho(1) (1 - \pi(< 0 >)).$$

6. Численные результаты

Далее приводятся результаты расчета вероятностей состояний и характеристик изучаемой системы с помощью предложенных методов. Очевидно, что с методологической точки зрения, а также по времени выполнения метод спектрального расширения и метод фазового укрупнения существенно

выигрывают перед методом производящих функций. Метод фазового укрупнения позволяет приближенно решить поставленные задачи А и В, при этом он существенно выигрывает у метода спектрального расширения в части реализации, так как этим методом удается решить указанные задачи с помощью простых формул.

Основная цель проведения вычислительных экспериментов сводится к оценке точности результатов, полученных с помощью метода фазового укрупнения, а также к определению области его применения в зависимости от исходных параметров конкретной системы. При этом точные значения (ТЗ) вероятностей состояний и характеристик системы вычисляются с помощью метода спектрального расширения.

Точность вычисления вероятностей состояний оценивается с помощью двух норм подобия: подобия косинуса ($\|N\|_1$) и максимума разностей между точными и приближенными значениями этих величин ($\|N\|_2$), т.е.

$$(6.1) \quad \|N\|_1 = \frac{\sum_{(n,k) \in E} p(n,k) \tilde{p}(n,k)}{\left(\sum_{(n,k) \in E} (p(n,k))^2 \right)^{\frac{1}{2}} \left(\sum_{(n,k) \in E} (\tilde{p}(n,k))^2 \right)^{\frac{1}{2}}};$$

$$(6.2) \quad \|N\|_2 = \max_{(n,k) \in E} |p(n,k) - \tilde{p}(n,k)|.$$

Отметим, что подобие косинуса, как правило, применяется для оценки близости ориентации векторов, а не для оценки их длины. Однако здесь это применение правомерно, так как согласно условию

$$\sum_{(n,k) \in E} p(n,k) = \sum_{(n,k) \in E} \tilde{p}(n,k) = 1$$

конечные точки этих векторов находятся в одной гиперплоскости.

Некоторые результаты численных экспериментов, в которых имеет место соотношение $\theta \gg \max(\lambda_0, \lambda_1)$, показаны в табл. 1 (поскольку рассматривается проблема определения точности вычисления вероятностей состояний и характеристик системы, а не характер их поведения относительно исходных данных, то результаты численных экспериментов представлены в виде таблицы. Кроме того, в проводимых экспериментах зачастую точные и приближенные значения характеристик отличаются друг от друга в четвертом знаке после запятой. Именно с такой точностью и проводились вычисления). Здесь и далее принято, что $\sigma = 0,2$. Из таблицы видно, что точные и приближенные значения вероятностей состояний почти совпадают, так как норма подобия косинуса почти всегда равна единице и максимум разностей между точными и приближенными значениями является достаточно малой величиной. Более того, в этих экспериментах характеристики L_0 и L_1 также вычисляются с очень высокой точностью. Так, относительная погрешность (ОП) вычисления характеристики L_0 в наихудшем случае не превышает 0,1% (т.е. почти совпадает с точными значениями) и ОП при вычислении характеристики L_1 не превышает 4%.

Таблица 1. Результаты численных экспериментов для системы, в которой удовлетворяется условие $\theta \gg \max(\lambda_0, \lambda_1)$

(μ, θ)	(λ_0, λ_1)	$\ N\ _1$	$\ N\ _2$	L_1			L_0		
				ТЗ	ПЗ	ОП	ТЗ	ПЗ	ОП
(50, 75)	(3, 5)	1	0,0006	0,1436	0,1436	0,0004	0,0198	0,0191	0,0338
	(4, 10)	1	0,0012	0,3329	0,3327	0,0007	0,0461	0,0444	0,0385
	(5, 15)	0,99	0,0018	0,5972	0,5966	0,0011	0,0829	0,0795	0,0400
(55, 75)	(3, 5)	1	0,0006	0,1289	0,1288	0,0004	0,0196	0,0189	0,0342
	(4, 10)	1	0,0013	0,2936	0,2934	0,0007	0,0448	0,0430	0,0396
	(5, 15)	0,99	0,0019	0,5141	0,5135	0,0011	0,0786	0,0753	0,0421
(60, 75)	(3, 5)	1	0,0006	0,1169	0,1168	0,0004	0,0194	0,0187	0,0346
	(4, 10)	1	0,0013	0,2626	0,2624	0,0007	0,0438	0,0420	0,0405
	(5, 15)	0,99	0,0020	0,4513	0,4508	0,0011	0,0754	0,0721	0,0438
(50, 80)	(3, 5)	1	0,0005	0,1436	0,1435	0,0004	0,0185	0,0179	0,0318
	(4, 10)	1	0,0011	0,3330	0,3327	0,0006	0,0432	0,0416	0,0361
	(5, 15)	1	0,0016	0,5973	0,5967	0,0010	0,0775	0,0746	0,0376
(55, 80)	(3, 5)	1	0,0005	0,1288	0,1288	0,0004	0,0183	0,0177	0,0322
	(4, 10)	1	0,0011	0,2936	0,2935	0,0006	0,0419	0,0403	0,0372
	(5, 15)	1	0,0017	0,5142	0,5137	0,0010	0,0736	0,0706	0,0396
(60, 80)	(3, 5)	1	0,0005	0,1168	0,1168	0,0004	0,0181	0,0175	0,0325
	(4, 10)	1	0,0011	0,2626	0,2625	0,0006	0,0409	0,0394	0,0381
	(5, 15)	1	0,0018	0,4515	0,4510	0,0010	0,0706	0,0677	0,0412
(50, 85)	(3, 5)	1	0,0004	0,1435	0,1435	0,0003	0,0174	0,0169	0,0300
	(4, 10)	1	0,0010	0,3330	0,3328	0,0006	0,0405	0,0391	0,0341
	(5, 15)	1	0,0015	0,5974	0,5969	0,0009	0,0728	0,0702	0,0355
(55, 85)	(3, 5)	1	0,0004	0,1288	0,1287	0,0003	0,0172	0,0167	0,0303
	(4, 10)	1	0,0010	0,2936	0,2935	0,0006	0,0394	0,0380	0,0351
	(5, 15)	1	0,0015	0,5144	0,5139	0,0009	0,0691	0,0665	0,0373
(60, 85)	(3, 5)	1	0,0004	0,1168	0,1167	0,0003	0,0170	0,0165	0,0306
	(4, 10)	1	0,0010	0,2626	0,2625	0,0006	0,0384	0,0371	0,0359
	(5, 15)	1	0,0016	0,4516	0,4512	0,0009	0,0663	0,0637	0,0389

Интересными являются результаты численных экспериментов, в которых не выполняется соотношение $\theta \gg \max(\lambda_0, \lambda_1)$. Соответствующие результаты показаны в табл. 2, откуда видно, что в этих экспериментах точные и приближенные значения вероятностей состояний также вычисляются с высокой точностью. Здесь однако при вычислении характеристик L_0 и L_1 имеются большие погрешности. Так, при вычислении характеристики L_0 относительная погрешность в наихудшем случае составляет почти 28%, а эта же величина при вычислении характеристики L_1 достигает 45%. Эти результаты доказывают, что принятое условие $\theta \gg \max(\lambda_0, \lambda_1)$ является важным для корректного применения метода фазового укрупнения состояний.

В связи с приведенными выше результатами численных экспериментов интерес представляет изучение зависимостей значений указанных выше норм подобия (6.1) и (6.2) относительно изменения параметра θ . Некоторые резуль-

Таблица 2. Результаты численных экспериментов для системы, в которой не удовлетворяется условие $\theta \gg \max(\lambda_0, \lambda_1)$

(μ, θ)	(λ_0, λ_1)	$\ N\ _1$	$\ N\ _2$	L_1			L_0		
				ТЗ	ПЗ	ОП	ТЗ	ПЗ	ОП
(50, 4)	(3, 5)	0,99	0,0818	0,1843	0,1616	0,1233	0,6692	0,4040	0,3962
	(4, 10)	0,97	0,1136	0,4545	0,3636	0,2000	1,5909	0,9091	0,4286
	(5, 15)	0,94	0,1002	1,0269	0,7385	0,2809	3,2885	1,8462	0,4386
(55, 4)	(3, 5)	0,99	0,0832	0,1649	0,1445	0,1233	0,6617	0,3975	0,3993
	(4, 10)	0,96	0,1188	0,3953	0,3162	0,2000	1,5415	0,8696	0,4359
	(5, 15)	0,94	0,1111	0,8417	0,6053	0,2809	3,0359	1,6646	0,4517
(60, 4)	(3, 5)	0,99	0,0843	0,1491	0,1307	0,1233	0,6556	0,3922	0,4019
	(4, 10)	0,96	0,1231	0,3497	0,2797	0,2000	1,5035	0,8392	0,4419
	(5, 15)	0,94	0,1202	0,7131	0,5128	0,2809	2,8606	1,5385	0,4622
(50, 7)	(3, 5)	0,99	0,0384	0,1589	0,1519	0,0439	0,2984	0,2171	0,2727
	(4, 10)	0,99	0,0649	0,3668	0,3391	0,0755	0,6920	0,4844	0,3000
	(5, 15)	0,98	0,0734	0,6979	0,6189	0,1131	1,2789	0,8842	0,3086
(55, 7)	(3, 5)	0,99	0,0390	0,1424	0,1361	0,0439	0,2951	0,2139	0,2753
	(4, 10)	0,99	0,0674	0,3214	0,2971	0,0755	0,6731	0,4669	0,3063
	(5, 15)	0,98	0,0790	0,5894	0,5228	0,1131	1,2082	0,8215	0,3201
(60, 7)	(3, 5)	0,99	0,0394	0,1289	0,1233	0,0439	0,2925	0,2113	0,2774
	(4, 10)	0,99	0,0694	0,2860	0,2644	0,0755	0,6584	0,4533	0,3115
	(5, 15)	0,98	0,0837	0,5102	0,4524	0,1131	1,1565	0,7756	0,3293
(50, 10)	(3, 5)	0,99	0,0220	0,1522	0,1488	0,0220	0,1879	0,1488	0,2079
	(4, 10)	0,99	0,0404	0,3478	0,3344	0,0385	0,4348	0,3344	0,2308
	(5, 15)	0,99	0,0500	0,6375	0,6000	0,0588	0,7875	0,6000	0,2381
(55, 10)	(3, 5)	0,99	0,0223	0,1364	0,1334	0,0220	0,1857	0,1467	0,2101
	(4, 10)	0,99	0,0418	0,3055	0,2938	0,0385	0,4230	0,3231	0,2361
	(5, 15)	0,99	0,0534	0,5426	0,5106	0,0588	0,7468	0,5617	0,2479
(60, 10)	(3, 5)	0,99	0,0226	0,1236	0,1209	0,0220	0,1840	0,1450	0,2118
	(4, 10)	0,99	0,0430	0,2724	0,2619	0,0385	0,4138	0,3143	0,2405
	(5, 15)	0,99	0,0562	0,4722	0,4444	0,0588	0,7167	0,5333	0,2558

таты этих исследований показаны на рис. 2, где исходные данные выбраны так: $\mu = 50$, $\lambda_0 = 5$, $\lambda_1 = 10$. Из этих графиков видно, что с ростом значений параметра θ норма подобия (6.1) систематически растет и приближается к единице, а норма подобия (6.2) систематически убывает и приближается к нулю. Эти результаты были вполне ожидаемыми, так как с ростом значений параметра θ увеличиваются интенсивности переходов между состояниями внутри классов E_n , $n \geq 0$ и тем самым уменьшаются интенсивности переходов между состояниями из разных классов (см. рис. 1).

Предложенные простые формулы позволяют изучить поведение характеристик системы (3.19) и (3.20) относительно любых параметров. Для конкретности изложения на рис. 3 показано поведение указанных характеристик относительно параметра θ . Эти результаты также были ожидаемыми, т.е. с ростом параметра θ увеличивается доля времени пребывания сервера в рабо-

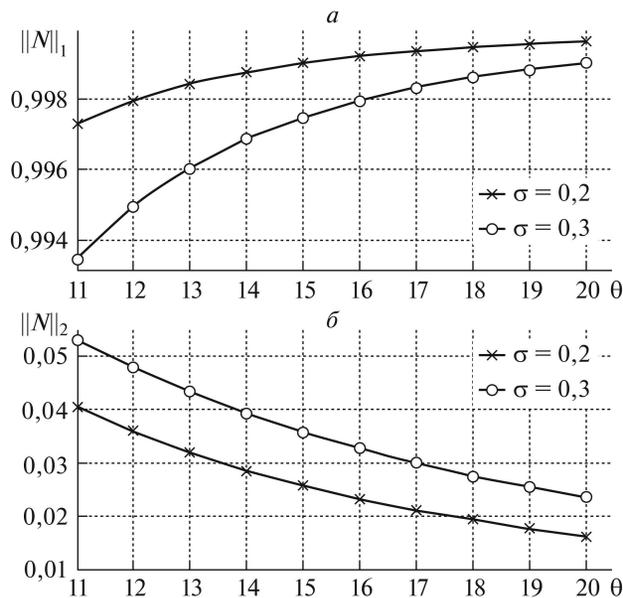


Рис. 2. Зависимость норм подобиа (6.1) (а) и (6.2) (б) от параметра θ .

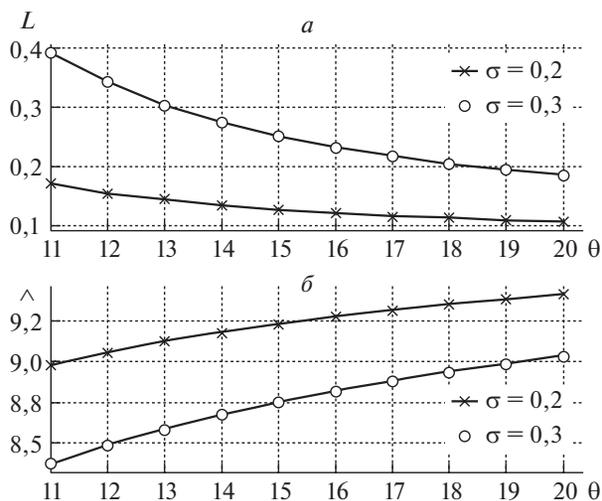


Рис. 3. Зависимость характеристик (3.19) (а) и (3.20) (б) от параметра θ .

чем режиме, тем самым уменьшается среднее число вызовов в системе (см. рис. 3,а) и увеличивается производительность системы (см. рис. 3,б). При этом с ростом вероятности мгновенного повторения обслуживания (т.е. σ) увеличивается среднее число вызовов в системе (см. рис. 3,а) и одновременно уменьшается производительность системы (см. рис. 3,б). Действительно, с ростом указанной вероятности увеличивается доля времени пребывания сервера в режиме переключения (при котором не производится обслуживание

вызовов), тем самым увеличивается среднее число вызовов в системе и, следовательно, уменьшается ее производительность.

7. Заключение

В работе изучается модель системы обслуживания с одним сервером, неограниченным числом мест для ожидания и мгновенной обратной связью. После завершения обслуживания часть вызовов согласно схеме Бернулли либо покидает систему, либо мгновенно требует повторного обслуживания. Для начала процесса повторного обслуживания вызовов серверу потребуется некоторое случайное время переключения, которое имеет показательную функцию распределения. Считается, что когда сервер находится в статусе переключения, он не может осуществлять обслуживание вызовов. При этом не допускается прерывание периода переключения. Интенсивность поступления вызовов зависит от статуса сервера.

Показано, что математической моделью изучаемой системы является двумерная ЦМ с бесконечномерным пространством состояний. Найдено условие эргодичности модели и предложены три метода исследования соответствующей двумерной ЦМ: метод производящих функций, метод спектрального расширения и метод фазового укрупнения состояний двухмерных ЦМ. Первые два из них являются точными методами, а третий — приближенный. С помощью численных экспериментов показана высокая точность разработанных приближенных формул для решения поставленной задачи.

СПИСОК ЛИТЕРАТУРЫ

1. *Takacs L.* A single-server queue with feedback // *Bell Syst. Tech. J.* 1963. V. 42. P. 505–519.
2. *Takacs L.* A queuing model with feedback // *Oper. Res.* 1977. V. 11. P. 345–354.
3. *Назаров А.А., Мусеева С.П., Морозова А.С.* Исследования СМО с повторным обслуживанием и неограниченным числом обслуживающих приборов методом предельной декомпозиции // *Выч. технол.* 2008. Т. 13. Вып. 5. С. 88–92.
4. *Dudin A.N., Kazimirsky A.V., Klimenok V.I., et.al.* The queuing model MAP/PH/1/N with feedback operating in a Markovian random environment // *Austrian J. Stat.* 2005. V. 34. Iss. 2. P. 101–110.
5. *Krishnamoorthy A., Manjunath A.S.* On queues with priority determined by feedback // *Calcutta Stat. Assos. Bul.* 2018. V. 70. Iss. 1. P. 33–56.
6. *Boxma O.J., Yechiali U.* An M/G/1 queue with multiple types of feedback and gated vacations // *J. Appl. Probab.* 1997. V. 34. Iss. 3. P. 773–784.
7. *Choi B.D., Kim B.* M/G/1 queueing system with fixed feedback policy // *The ANZIAM J.* 2002. V. 44. Iss. 2. P. 283–297.
8. *Choi B.D., Kim B., Choi S.H.* An M/G/1 queue with multiple types of feedback, gated vacations and FCFS policy // *Comput. & Oper. Res.* 2003. V. 30. Iss. 9. P. 1289–1309.
9. *Krishna Kumar B., Madheswari S.P., Vijayakumar A.* The M/G/1 retrial queue with feedback and starting failures // *Appl. Math. Model.* 2002. V. 26. Iss. 11. P. 1057–1075.

10. *Krishna Kumar B., Rukmani R., Thangaraj V.* On multiserver feedback retrial queue with finite buffer // *Appl. Math. Model.* 2009. V. 33. Iss. 4. P. 2062–2083.
11. *Krishna Kumar B., Vijayalakshmi G., Krishnamoorthy A., et.al.* A single server feedback retrial queue with collisions // *Comput. & Oper. Res.* 2010. V. 37. Iss. 7. P. 1247–1255.
12. *Koroliuk V.S., Melikov A.Z., Ponomarenko L.A., et.al.* Methods for analysis of multi-channel queuing models with instantaneous and delayed feedbacks // *Cyber. Syst. Anal.* 2016. V. 52. Iss. 1. P. 58–70.
13. *Melikov A.Z., Aliyeva S.H., Sztrik J.* Analysis of queuing system MMPP/M/ K/K with delayed feedback // *Mathematics.* 2019. V. 7. 1128. 14 p.
doi:10.3390/math7111128
14. *Mitrani I., Chakka R.* Spectral expansion solution for a class of Markov models: application and comparison with the matrix-geometric method // *Perform. Eval.* 1995. V. 23. P. 241–260.
15. *Chakka R.* Spectral expansion solution for some finite capacity queues // *Oper. Res.* 1998. V. 79. P. 27–44.
16. *Neuts M.F.* Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach. Baltimore: John Hopkins Univer. Press, 1981. 332 p.

Статья представлена к публикации членом редколлегии В.М. Вишневым.

Поступила в редакцию 09.12.2019

После доработки 02.03.2020

Принята к публикации 25.05.2020