СИСТЕМА M/PH/C С АДРЕСНОЙ СТРАТЕГИЕЙ ПОВТОРНЫХ ВЫЗОВОВ

В. В. Мушко

Белорусский государственный университет
Минск, Беларусь
E-mail: vmushko@tut.by

Исследована многолинейная система массового обслуживания с адресной стратегией первичных и повторных вызовов. В отличие от стандартной постановки задачи, предполагающей, что вызовы с орбиты обращены ко всем приборам одновременно, здесь предполагается, что в момент первичного поступления и момент попытки вызов выбирает один конкретный прибор (в соответствии с заданным распределением вероятностей). Время обслуживания вызова имеет распределение фазового типа. Получено условие эргодичности системы.

Ключевые слова: многолинейная система, распределение фазового типа, адресная стратегия повторных вызовов, стационарное распределение.

ВВЕДЕНИЕ

Важным разделом теории массового обслуживания является теория систем с повторными вызовами («retrials» или «repeated customers»). В таких системах вызов, поступивший в систему и заставший прибор занятым, не становится в очередь неограниченной или ограниченной длины, как в системах с ожиданием, и не уходит из системы навсегда, как в системах с потерей вызовов (с отказами). Этот вызов покидает систему на некоторое случайное время, иначе говоря, уходит на орбиту («orbit» или «retrial group»), а затем повторяет попытку попасть на обслуживание. Предполагается, что вызов повторяет попытки до тех пор, пока не поступит на обслуживание.

Важность этого раздела теории обусловлена его широкими практическими приложениями. Область приложений лежит в оценивании производительности и проектировании телефонных сетей, локальных вычислительных сетей с протоколами случайного множественного доступа, широковещательных радиосстей, мобильных, сотовых радиосетей. Наличие повторных попыток получить обслуживание является неотъемлемой чертой этих систем, игнорирование данного эффекта может привести к значительным погрешностям при принятии инженерных решений.

Первые математические результаты, касающиеся систем с повторными вызовами, были опубликованы в 40-х гг. [1]. Обзоры работ, посвященные данным системам, содержатся в статьях [2–6], а также монографиях известных специалистов в области теории систем с повторными вызовами Г. И. Фалина, Дж. Темплтона [7] и С. Н. Степанова [8].

Системы с повторными вызовами отличаются числом приборов, размерами орбиты, видом интенсивности повторных вызовов, стратегией повторов, стратегией первичного доступа.

Более интенсивно исследуются однолинейные системы, в то время как многолинейные системы исследованы в меньшей степени. Это можно объяснить тем, что стохастические процессы, описывающие поведение последних, имеют более сложную структуру. Аналитические результаты получены только в частных случаях, включая случай $c \le 2$, где c – число приборов (см., например, [9]). Для случая $c > 2$ применяют аппроксимационные и численные методы.

Что касается орбиты, то она может быть конечной (см., например, [10]) или бесконечной (такие системы встречаются чаще).

Иногда рассматривают системы, в которых повторные вызовы образуют стационарный пуассоновский поток с постоянной интенсивностью («constant retrial rate») $\alpha_i = \gamma, i \ge 1, \gamma > 0$, не зависящей от числа требований на орбите (см., например, [9]). Этот факт можно интерпретировать двояко. Первый вариант: при наличии i вызовов на орбите каждый вызов получает право совершать повторные попытки независимо от других вызовов через экспоненциально распределенное (с параметром γ/i) время. Второй вариант: только одному из вызовов разрешается совершать повторные попытки через экспоненциально распределенные (с параметром γ) интервалы времени.

Чаще рассматривают системы, в которых суммарная интенсивность повторных вызовов определяется по формуле $\alpha_i = i\alpha, i \ge 0, \alpha > 0$, где i – число вызовов на орбите, α – интенсивность потока повторных попыток одного вызова. Такой механизм повторных вызовов называется классическим («classical retrial rate»). См, например, [11].

В работе [12], исследуется система с линейной зависимостью («linear retrial rate») интенсивности потока повторных вызовов с орбиты от текущей величины орбиты:

$$\alpha_i = \begin{cases} 0, & i = 0, \\ i\alpha + \gamma, \alpha > 0, \gamma > 0, & i > 0. \end{cases}$$

В [13] рассмотрена квадратичная зависимость интенсивности повторных вызовов от количества вызовов на орбите $\alpha_i = i^2\beta + i\alpha + \gamma, \beta > 0, \alpha > 0, \gamma > 0$ («quadratic retrial rate»).

Исследование управляемых систем подтолкнуло к созданию моделей, где мгновенная интенсивность повторных вызовов является линейной и зависит от состояния марковского управляющего процесса ξ_t с пространством состояний $\{0, \ldots, K\}$

$$\alpha_i = i\alpha^{(\xi)} + \gamma^{(\xi)}, i \ge 1,$$

где ξ – состояние управляющего процесса $\xi(t)$, $\xi \in \{0, \ldots, K\}$, i – число вызовов на орбите, $\alpha_0 = 0$ («modulated retrial rate»). Такие системы рассмотрены, например, в работе [14].

Для многолинейных систем могут быть рассмотрены различные стратегии доступа к приборам. Так, классическая стратегия предполагает, что если в момент повтора имеются один или несколько свободных приборов, то вызов занимает любой из них (см., например, [12]). Стратегия полного доступа заключается в том, что если в момент повтора имеются i свободных приборов, в то время как на орбите находится

k повторных вызовов, то $\min\{i,k\}$ вызовов занимают свободные приборы (см., например, [15]). Кроме перечисленных выше, также можно упомянуть о существовании различных схем приоритезации, которые в основном подразумевают использование зарезервированных каналов («guard channels») (см., например, [16]).

В данной работе будет рассмотрена адресная стратегия доступа к приборам.

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

Рассматривается многоканальная система, имеющая c идентичных приборов. Время обслуживания вызова имеет распределение фазового типа (общепринятый термин для такого процесса обслуживания – *PH* (*Phase-type*)-обслуживание (см., например, [17]). Процессом обслуживания управляет цепь Маркова m_t, $t \geq 0$ с непрерывным временем. Состояние процесса $m_t, t \geq 0$ в момент начала обслуживания определяется стохастическим вектором-строкой $\vec{\beta} = (\beta_1,...,\beta_M)$. Переходы процесса $m_t, t \geq 0$, которые не приводят к завершению обслуживания, определяются неприводимой матрицей S размерности $M \times M$. Диагональные элементы матрицы S отрицательны и $-S_{m,m}$ определяет параметр экспоненциально распределенного времени пребывания процесса в состоянии m, $|S_{m,m}| < \infty, m = \overline{1,M}$. Недиагональные элементы матрицы S определяют интенсивности переходов процесса $m_t, t \geq 0$ в пространстве состояний $\{1,\ldots,M\}$. Значения $-\sum\limits_{m'=1}^{M} S_{m,m'}$ определяют интенсивности перехода процесса $m_t, t \geq 0$ из состояния m в абсорбирующее состояние. Момент перехода процесса $m_t, t \geq 0$ в абсорбирующее состояние определяет момент завершения обслуживания. Интенсивности переходов, которые приводят к завершению обслуживания, определяются вектором $S_0 = -S\mathbf{e}$, где $\mathbf{e}$ – вектор-столбец, состоящий из единиц. Предполагается, что все компоненты вектора-столбца S_0 неотрицательны и, по крайней мере, одна из них положительна.

Поток, входящий в систему, является стационарным пуассоновским потоком с интенсивностью $\lambda, \lambda > 0$. В момент прибытия вызов выбирает для обслуживания r-й прибор с вероятностью $q_r, 0 < q_r < 1, r = \overline{1,c}, \sum\limits_{r=1}^{c} q_r = 1$. Если выбранный прибор свободен, то вызов занимает прибор и после обслуживания покидает систему. Если выбранный прибор занят, то вызов направляется в некоторую виртуальную динамическую область, называемую орбитой, и пытается получить обслуживание позже. Каждый вызов, находящийся на орбите, делает повторные попытки через интервалы времени, имеющие экспоненциально распределенную длину с параметром α, $\alpha>0$, независимо от других вызовов. В момент повтора вызов выбирает r-й прибор с вероятностью $\varphi_r, 0 < \varphi_r < 1, r = \overline{1,c}, \sum\limits_{r=1}^{c} \varphi_r = 1$. Если прибор свободен, то вызов занимает его и покидает систему после обслуживания. Если прибор занят, то вызов возвращается на орбиту, даже если несколько других приборов свободно в этот момент. Вызовы с орбиты пытаются получить обслуживание до тех пор, пока им не удастся занять прибор, выбранный при соответствующей попытке.

ОПИСАНИЕ ПОВЕДЕНИЯ СИСТЕМЫ В ТЕРМИНАХ ЦЕПИ МАРКОВА С НЕПРЕРЫВНЫМ ВРЕМЕНЕМ

Рассмотрим процесс $\zeta_t = \{i_t, m_t^{(1)}, ..., m_t^{(c)}\}, t \geq 0$, где $i_t, i_t \geq 0$ – число вызовов на орбите в момент времени t, $m_t^{(r)}$ – состояние r-го прибора в момент времени t, $r = \overline{1,c}$:

$$m_t^{(r)} = \begin{cases} 0, & \text{если } r\text{-й прибор свободен в момент времени } t, \\ m, & \text{если } r\text{-й прибор обслуживает вызов на } m\text{-ой фазе} \\ & \text{в момент времени } t,\ m = \overline{1,M}. \end{cases}$$

Очевидно, что процесс $\zeta_t, t \geq 0$ является цепью Маркова с непрерывным временем. Обозначим стационарное распределение этой цепи через

$$p(i, m_1, \ldots, m_c) = \lim_{t \to \infty} P\{i_t = i,\ m_t^{(1)} = m_1, \ldots, m_t^{(c)} = m_c\}, \tag{1}$$

$$i \geq 0,\ m_r = \overline{1,M},\ r = \overline{1,c}.$$

Условие существования предела в (1) приведено ниже и предполагается далее выполненным. Занумеруем состояния цепи Маркова $\zeta_t, t \geq 0$ в лексикографическом порядке и сформируем векторы-строки вероятностей $\vec{p}_i$, соответствующие состоянию i числа вызовов на орбите. Размерности этих векторов равны $(M+1)^c$. Сформируем также макровектор $\vec{p} = (\vec{p}_0, \vec{p}_1, ..., \vec{p}_i, ...)$.

Лемма. Вектор $\vec{p}$ является единственным решением системы:

$$\vec{p}Q = \vec{0},\ \vec{p}\mathbf{e} = 1,$$

где инфинитезимальный генератор Q цепи Маркова $\zeta_t, t \geq 0$ имеет следующую форму:

$$Q = \begin{pmatrix} Q_{0,0} & Q_{0,1} & 0 & 0 & \cdots \\ Q_{1,0} & Q_{1,1} & Q_{1,2} & 0 & \cdots \\ 0 & Q_{2,1} & Q_{2,2} & Q_{2,3} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \tag{2}$$

где блоки $Q_{i,j}$ вычисляются следующим образом:

$$Q_{i,i-1} = i\alpha L,\ i \geq 1,\ Q_{i,i} = A - i\alpha B,\ i \geq 0,\ Q_{i,i+1} = S,\ i \geq 0, \tag{3}$$

где

$$A = A_1,\ B = B_1,\ L = L_1,\ S = S_1$$

и матрицы A_1, B_1, L_1, S_1, размерности $(M+1)^c$, вычисляются из следующих рекуррентных отношений:

$$A_c = \begin{pmatrix} -\lambda & \lambda q_c \vec{\beta} \\ S_0 & S - \lambda I_M \end{pmatrix},$$

$$A_{c-k+1} = \begin{pmatrix} O_{(M+1)^{k-1}} & \lambda q_{c-k+1} \vec{\beta} \otimes I_{(M+1)^{k-1}} \\ S_0 \otimes I_{(M+1)^{k-1}} & S \otimes I_{(M+1)^{k-1}} \end{pmatrix} + I_{M+1} \otimes A_{c-k},\ k = \overline{2,c},$$

$$L_c = \begin{pmatrix} 0 & \varphi_c \vec{\beta} \\ \bar{0}_M & O_M \end{pmatrix},$$

$$L_{c-k+1} = \begin{pmatrix} O_{(M+1)^{k-1}} & \varphi_{c-k+1} \vec{\beta} \otimes I_{(M+1)^{k-1}} \\ O_{M(M+1)^{k-1} \times (M+1)} & O_{M(M+1)^{k-1}} \end{pmatrix} + I_{M+1} \otimes L_{c-k},\ k = \overline{2,c},$$

$$B_c = diag\{\varphi_1,\ 0_M\},$$

$$B_{c-k+1} = B_{c-k} \otimes I_{M+1} + \varphi_{c-k+1} I_{(M+1)^{k-1}} \otimes diag\{1,\ 0_M\},\ k = \overline{2,c},$$

$$S_c = diag\{0,\ \lambda q_1 e_M\},$$

$$S_{c-k+1} = S_{c-k} \otimes I_{M+1} + \lambda q_{c-k+1} I_{(M+1)^{k-1}} \otimes diag\{0,\ e_M\},\ k = \overline{2,c},$$

где $\vec{0}$ – вектор-строка, состоящая из нулей, O_l – нулевая матрица размерности l, $O_{l,m}$ – нулевая матрица размерности $l \times m$, I_l – тождественная матрица размерности l, 0_l – вектор-строка размерности l, состоящая из нулей, $\bar{0}_l$ – вектор-столбец размерности l, состоящий из нулей, e_l – вектор-строка размерности l, состоящая из единиц, $\otimes$ – символ кронекерова произведения матриц, $diag\{a_1, \ \ldots, \ a_c\}$ – диагональная матрица с диагональными элементами (блоками) $a_1, \ldots, a_c$.

СТАЦИОНАРНОЕ РАСПРЕДЕЛЕНИЕ ЦЕПИ МАРКОВА. УСЛОВИЕ ЭРГОДИЧНОСТИ

Для установления условий существования стационарного распределения и вычисления вектора $\vec{p}$ стационарных вероятностей применим аппарат многомерных асимптотически квазитеплицевых цепей Маркова [12, 18]. Для этого рассмотрим формальную цепь Маркова ξ_n, $n \geq 1$ с дискретным временем, которую получим, разделив каждую строку генератора Q на модуль максимального диагонального элемента этого генератора и добавив единицу к диагональному элементу.

Введем диагональную матрицу $R_i = F + i\alpha\hat{I}$, где матрица F состоит из диагональных элементов матрицы A_1, матрица $\hat{I}$ получена из единичной матрицы I, заменой диагональных элементов, номера которых совпадают с номерами нулевых диагональных элементов матрицы B_1, на 0.

Таким образом, цепь Маркова $\xi_n,\ n \geq 1$ с дискретным временем характеризуется матрицей P одношаговых переходных вероятностей, имеющей следующий вид:

$$P = \begin{pmatrix} P_{0,0} & P_{0,1} & 0 & 0 & \cdots \\ P_{1,0} & P_{1,1} & P_{1,2} & 0 & \cdots \\ 0 & P_{2,1} & P_{2,2} & P_{2,3} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \tag{4}$$

где

$$\begin{gathered} P_{i,i-1} = i\alpha(F + i\alpha\hat{I})^{-1} L_1,\ i \geq 1, \\ P_{i,i} = I + (F + i\alpha\hat{I})^{-1}(A_1 - i\alpha B_1),\ i \geq 0, \\ P_{i,i+1} = (F + i\alpha\hat{I})^{-1} S_1,\ i \geq 0. \end{gathered} \tag{5}$$

Формулы (4), (5) получены из (2), (3) умножением блоков (3) на матрицу R_i^{-1}.

Обозначим через $\vec{\pi} = (\vec{\pi}_0, \vec{\pi}_1, \ldots, \vec{\pi}_i, \ldots)$ вектор-строку стационарных вероятностей цепи Маркова $\xi_n,\ n \geq 1$. Он удовлетворяет следующим уравнениям:

$$\vec{\pi} = \vec{\pi} P,\ \ \vec{\pi}\,\mathbf{e} = 1\,.$$

Тогда компоненты вектора $\vec{p}$ вычисляются следующим образом:

$$\vec{p}_i = \bar{c}\ \vec{\pi}_i R_i^{-1}, \quad i \geq 0\,,$$

где $\bar{c}$ находится из условия нормировки.

Проанализировав структуру блоков переходных вероятностей матрицы P, представим их искусственно в форме:

$$P_{i,l} = Q_1^{(i)}\, Y_{l-i+1}^{(1)} + Q_2^{(i)}\, Y_{l-i+1}^{(2)}, \quad l = i-1, i, i+1\,, \tag{6}$$

где

$$\begin{gathered} Q_1^{(i)} = i\alpha\hat{I}(F + i\alpha\hat{I})^{-1},\ Q_2^{(i)} = F(F + i\alpha\hat{I})^{-1}, \\ \lim_{i\to\infty} Q_1^{(i)} = \hat{I}\,,\ \lim_{i\to\infty} Q_2^{(i)} = \bar{I} = I - \hat{I}\,,\ Y_0^{(1)} = L_1\,,\ Y_0^{(2)} = O_{(M+1)^c}\,, \\ Y_1^{(1)} = I - B_1\,,\ Y_1^{(2)} = I + F^{-1}A_1\,,\ Y_2^{(1)} = O_{(M+1)^c}\,,\ Y_2^{(2)} = F^{-1}S_1\,. \end{gathered} \tag{7}$$

Сравнивая (6), (7) с определением многомерной асимптотически квазитеплицевой цепи Маркова (*AQTMC* – asymptotically quasi-Toeplitz Markov chain) в [12], заключаем, что цепь Маркова $\xi_n,\ n \geq 1$ принадлежит к классу *AQTMC*, поэтому, результаты из [12, 18] могут быть применены для ее исследования.

Предельная (по отношению к *AQTMC* $\xi_n,\ n \geq 1$) квазитеплицева цепь Маркова $\tilde{\xi}_n,\ n \geq 1$ имеет блоки $\tilde{P}_{i,l},\ l = i-1, i, i+1$ переходных вероятностей следующей формы:

$$P_{i,i-1} = \tilde{Y}_0 = \hat{I}\, L_1,\ i \geq 1\,,$$

$$P_{i,i} = \widetilde{Y}_1 = I - \hat{I}B_1 + \bar{I}F^{-1}A_1, i \ge 0,$$

$$P_{i,i+1} = \widetilde{Y}_2 = \bar{I}F^{-1}S_1, i \ge 0.$$

Обозначим $\widetilde{Y}(z) = \widetilde{Y}_0 + \widetilde{Y}_1 z + \widetilde{Y}_2 z^2$, $|z| \le 1$.

Как следует из [12], достаточное условие существования стационарного распределения цепи Маркова ξ_n, $n \ge 1$ совпадает с достаточным условием существования стационарного распределения цепи Маркова $\widetilde{\xi}_n$, $n \ge 1$ и имеет следующую форму:

$$\vec{X}\widetilde{Y}'(1)\mathbf{e} < 1,$$

где вектор $\vec{X}$ удовлетворяет уравнениям:

$$\vec{X}\,\widetilde{Y}(1) = \vec{X}, \ \vec{X}\mathbf{e} = 1.$$

Используя это утверждение, можно получить достаточное условие существования стационарного распределения $\vec{\pi}$.

Теорема. Стационарное распределение $\vec{\pi}$ существует, если следующее условие выполнено:

$$\vec{X}[I - B_1 + \bar{I}F^{-1}(A_1 + 2S_1)]\mathbf{e} < 1,$$

где вектор $\vec{X}$ *удовлетворяет уравнениям:*

$$\vec{X}[L_1 - B_1 + \bar{I}F^{-1}(A_1 + S_1)]\mathbf{e} = \vec{0}, X\mathbf{e} = 1.$$

Замечание. Очевидно (см. [18]), что данное условие является и достаточным условием существования стационарного распределения $\vec{p}$. Стационарные распределения $\vec{\pi}$ и $\vec{p}$ могут быть найдены с помощью алгоритма, предложенного в [18].

ЗАКЛЮЧЕНИЕ

В данной работе была исследована многолинейная система с адресными первичными и повторными вызовами. Поток, входящий в систему, является стационарным пуассоновским потоком. Время обслуживания вызова имеет распределение фазового типа. Поведение данной системы описывается многомерной цепью Маркова с непрерывным временем. Соответствующая многомерная цепь Маркова с дискретным временем, вложенная по моментам скачков исходной цепи, принадлежит к классу асимптотически квазитеплицевых цепей Маркова. Используя известные [12, 18] результаты для таких цепей, для рассмотренной системы получено достаточное условие существования стационарного распределения.

ЛИТЕРАТУРА

1. *Kosten L.* On the influence of repeated calls in the theory of probabilities of blocking // De Ingenier. 1947. V. 59. P. 1–25.

2. *Artalejo J. R.* A classified bibliography of research on retrial queues: Progress in 1990–1999 // TOP. 1999. V. 7. N. 2. P. 187–211.

3. *Artalejo J. R.* Accessible bibliography on retrial queues // Mathematical and Computer Modelling. 1999. V. 30. P. 223–233.

4. *Falin G. I.* A survey of retrial queues // Queueing Systems. 1990. V. 7. P. 127–167.

5. *Yang T., Templeton J. G. C.* A survey of retrial queues // Queueing Systems. 1987. V. 2. P. 201–233.

6. *Kulkarni V. G., Liang H. M.* Retrial queues revisited. Frontiers in Queueing: Models and Applications in Science and Engineering. – ed. J. H. Dshalalow (CRC Press, Inc, Boca Raton), 1997. P. 19–34.

7. *Falin G. I., Templeton J. G. C.* Retrial Queues. London: Chapman&Hall, 1997. 328 p.

8. *Степанов С. Н.* Численные методы расчета систем с повторными вызовами. М.: Наука, 1983. 230 с.

9. *Artalejo J. R.* Stationary analysis of the characteristics of the M/M/2 queue with constant repeated attempts // Opsearch. 1996. V. 33. N. 2. P. 83–95.

10. *Choi B. D., Chang Y.* $MAP_1, MAP_2/M/c$ retrial queue with the retrial group of finite capacity and geometric loss // Mathematical and Computer Modelling. 1999. V. 30. P. 99–113.

11. *Diamond J. E., Alfa A. S.* The MAP/PH/1 retrial queue // Stochastic Models. 1998. V. 14. P. 1151–1177.

12. *Dudin A. N., Klimenok V. I.* A retrial BMAP/SM/1 system with linear repeated requests // Queueing Systems. 2000. V. 34. N. 1–4. P. 47–66.

13. *Artalejo J. R., Lopez-Herrero M. J.* On the M/G/1 queue with quadratic repeated attemps // Statist. Methods. 2001. V. 3. N. 2. P. 60–78.

14. *Dudin A. N., Klimenok V. I.* BMAP/SM/1 model with Markov modulated retrials // TOP. 2000. V. 7. N. 2. P. 267–278.

15. *Artalejo J. R., Gomez-Corral A., Neuts M. F.* Numerical analysis of multiserver retrial queues operating under a full access policy. Advances in Matrix Algorithmic Methods for Stochastic Models: Proc. of the 3rd Int. Conf. on Matrix Analytic Methods, eds. G. Latouche and P. Taylor. Notable Publications, Inc., New Jersey, 2000. P. 1–19.

16. *Choi B. D., Chang Y.* $MAP_1, MAP_2/M/c$ retrial queue with guard channel and its application to cellular networks // TOP. 1999. V. 7. N. 2. P. 231–248.

17. *Бочаров П. П., Печинкин А. В.* Теория массового обслуживания. М.: УДН, 1995. 528 с.

18. *Breuer L., Dudin A. N., Klimenok V. I.* A retrial BMAP/PH/N system // Queueing Systems. 2002. V. 40. N. 4. P. 433–457.