ДВУХФАЗНАЯ СИСТЕМА ОБСЛУЖИВАНИЯ С ЦЕНТРАЛИЗОВАННОЙ СТРАТЕГИЕЙ ПОВТОРОВ

О.С. Тарамин

Белорусский государственный университет, кафедра ТВиМС
пр. Независимости, д. 4, Минск, Беларусь
телефон: + 375172095486; e-mail: taramin@bsu.by

Исследована двухфазная система массового обслуживания (СМО), первую фазу которой образует однолинейная система с групповым марковским входящим потоком и повторными вызовами, вторую фазу – многолинейная СМО без буфера. Рассмотрена централизованная стратегия повторов. В случае отсутствия свободных приборов на второй фазе заявка, обслужившаяся на первой фазе, либо ждет освобождения приборов, блокируя прибор первой фазы, либо уходит из системы недообслуженной.

Ключевые слова - групповой марковский поток, двухфазная система массового обслуживания, квазитеплицева цепь Маркова, централизованная стратегия повторов.

1 ВВЕДЕНИЕ

В связи с бурным развитием телекоммуникационных сетей и необходимостью создания адекватных средств математического моделирования этих сетей в современной науке активно изучаются многофазные системы обслуживания с повторными вызовами. Важность исследования данных систем обусловлена их широким практическим применением. Многофазные системы обслуживания с повторными вызовами являются математическими моделями многих информационно-телекоммуникационных сетей, они используются при проектировании и оценивании производительности телефонных сетей, локальных вычислительных сетей с протоколами случайного множественного доступа, широковещательных радиосетей, мобильных сотовых сетей.

В [1] рассмотрена двухфазная система с групповым марковским потоком (Batch Markovian Arrival Process – $BMAP$) и орбитой неограниченного объема для повторных вызовов на первой фазе. В данной статье исследуется аналогичная СМО, но в отличие от [1] предполагается, что суммарная интенсивность повторных попыток является постоянной, а не возрастающей функцией числа повторных вызовов, то есть рассматривается централизованная стратегия повторов.

В статье получено условие существования стационарного режима в рассматриваемой СМО, найдены стационарные распределения во вложенные и произвольные моменты времени, приведены формулы для нахождения основных характеристик производительности.

2 МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

Рассматривается двухфазная система с повторными вызовами. На первой фазе находится один обслуживающий прибор. Входящий поток первичных заявок является групповым марковским ($BMAP$). Поступление групп заявок в этом потоке возможно только в моменты скачков некоторой неприводимой цепи Маркова $\nu_t, t \geq 0$, с непрерывным временем и конечным пространством состояний $\{0,1,\dots,W\}$, которая называется управляющим процессом $BMAP$-потока. Интенсивности переходов процесса $\nu_t, t \geq 0$, сопровождающиеся поступлением групп заявок размера $k \geq 1$ (не сопровождающиеся поступлением заявок – при $k=0$), задаются элементами матриц D_k (недиагональными элементами матрицы D_0) или производящей функцией $D(z)=\sum_{k=0}^{\infty} D_k z^k, |z| \leq 1$. При этом матрица $D(1)$ является неприводимым инфинитезимальным генератором цепи Маркова $\nu_t, t \geq 0$. Вектор-строка $\boldsymbol{\theta}$ стационарного распределения этой цепи является решением системы уравнений $\boldsymbol{\theta} D(1)=\mathbf{0}, \boldsymbol{\theta e}=1$, где $\boldsymbol{e}$ – вектор-столбец, состоящий из единиц. Интенсивность λ поступления заявок в стационарном $BMAP$-потоке имеет вид $\lambda=\boldsymbol{\theta} D'(z)|_{z=1} \boldsymbol{e}$, интенсивность λ_b поступления групп заявок определяется как $\lambda_b=\boldsymbol{\theta}(-D_0)\boldsymbol{e}$. Коэффициент вариации c_{var} длины интервала между моментами поступления последовательных групп определяется формулой $c_{var}^2=2\lambda_b \boldsymbol{\theta}(-D_0)^{-1}\boldsymbol{e}-1$. Коэффициент корреляции c_{cor} длин двух соседних интервалов вычисляется следующим образом:

$$c_{cor}=(\lambda_b \boldsymbol{\theta}(-D_0)^{-1}(D(1)-D_0)(-D_0)^{-1}\boldsymbol{e}-1)/c_{var}^2 .$$

Более подробное описание $BMAP$-потока можно найти в [2].

Если поступившая группа первичных заявок застает обслуживающий прибор свободным, то одна из заявок группы начинает обслуживаться, а остальные идут на орбиту неограниченного объема. Если прибор занят в момент поступления группы, то все заявки этой группы идут на орбиту. Полагаем, что суммарная интенсивность повторных вызовов не зависит от числа заявок на орбите и равна $\gamma, \gamma > 0$. С практической точки зрения это можно интерпретировать двояко. Первый вариант: при наличии i заявок на орбите каждый запрос получает право совершать повторные попытки независимо от других запросов через экспоненциально распределенное с параметром $\frac{\gamma}{i}$ время. Второй вариант: только одному из запросов разрешается совершать повторные попытки через экспоненциально распределенные (с параметром γ) интервалы времени.

Времена обслуживания заявок (первичных и повторных) на первой фазе являются независимыми случайными величинами с общим распределением $B(t)$, имеющим конечный первый момент $b_1 = \int_0^\infty t dB(t)$.

После обслуживания на первой фазе заявка поступает на вторую фазу рассматриваемой системы, которая представляет собой N-линейную СМО без буфера. Время обслуживания прибором второй фазы имеет показательное распределение с параметром μ. Для обслуживания на второй фазе заявке требуется случайное число приборов ϕ. Здесь ϕ – целочисленная случайная величина с распределением

$$q_n = P\{\phi = n\},\, q_n \geq 0,\, n = \overline{0,N},\, \sum_{n=0}^{N} q_n = 1.$$

Если в момент окончания обслуживания заявки на первом приборе на второй фазе отсутствует необходимое число свободных приборов, то с вероятностью p, $0 \leq p \leq 1$, заявка уходит из системы недообслуженной (получает отказ в обслуживании на второй фазе, теряется), а с дополнительной вероятностью ждет, пока освободится нужное число приборов. Период ожидания сопровождается блокировкой прибора первой фазы.

Введем следующие обозначения:

• I – тождественная матрица, $\otimes$ и $\oplus$ – символы Кронекерова произведения и суммы матриц;

• $\hat{D}_k = I_{N+1} \otimes D_k,\, k \geq 0,\, \hat{D}(z) = \sum_{k=0}^{\infty} \hat{D}_k z^k,\, |z| \leq 1;$

• $P(n,t), n \geq 0,$ – коэффициенты матричного ряда $e^{D(z)t} = \sum_{n=0}^{\infty} P(n,t) z^n, |z| \leq 1$. Элемент матрицы $P(n,t)$, стоящий в ν-й строке и ν'-м столбце, определяет вероятность того, что на интервале $(0,t]$ поступит n заявок в $BMAP$ и управляющий процесс ν_t перейдет из состояния ν в состояние $\nu', \nu, \nu' = \overline{0,W}$;

• $H(t) = (H_{r,r'}(t))_{r,r'=\overline{0,N}}$, где $H_{r,r'}(t) = 0$ для $r \leq r'$, а при $r > r'$ $H_{r,r'}(t)$ – функция распределения с преобразованием Лапласа-Стилтьеса вида

$$h_{r,r'}(s) = \prod_{l=r'+1}^{r} l\mu(l\mu + s)^{-1};$$

• $Q_m, m = \overline{1,3},$ – квадратные матрицы вида:

$$Q_1 = \begin{pmatrix} q_0 & q_1 & \dots & q_N \\ 0 & q_0 & \dots & q_{N-1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & q_0 \end{pmatrix},\; Q_3 = \begin{pmatrix} 0 & \dots & 0 & q_N \\ 0 & \dots & 0 & q_{N-1} \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & q_0 \end{pmatrix},$$

$$Q_2 = \operatorname{diag}\{ \sum_{n=N-r+1}^{N} q_n, r = \overline{0,N}\};$$

• $R = \operatorname{diag}\{r, r = \overline{0,N}\}; \quad \overline{W} = W + 1;$

• $\hat{Q}_m = Q_m \otimes I_{\overline{W}},\, m = \overline{1,3}; \quad \overline{Q} = \hat{Q}_1 + p\hat{Q}_2;$

• $Q = \overline{Q} + (1-p)\int_0^\infty (dH(t) \otimes e^{D_0 t})\hat{Q}_3.$

3 СТАЦИОНАРНОЕ РАСПРЕДЕЛЕНИЕ ВЕРОЯТНОСТЕЙ СОСТОЯНИЙ ВЛОЖЕННОЙ ЦЕПИ МАРКОВА

Пусть t_n – n-ый момент окончания обслуживания на первой фазе. Рассмотрим процесс $\xi_n = \{i_n, r_n, \nu_n\}, n \geq 1$, где i_n – число заявок на орбите в момент t_n, $i_n \geq 0$; r_n – число занятых приборов на второй фазе в момент $t_n - 0$, $r_n = \overline{0,N}$; ν_n – состояние управляющего процесса $BMAP$ в момент t_n, $\nu_n = \overline{0,W}$.

Процесс $\xi_n, n \geq 1$, является трехмерной цепью Маркова. Перенумеруем состояния этой цепи в лексикографическом порядке и сформируем $(N+1)\overline{W} \times (N+1)\overline{W}$-матрицы $P_{i,j}, i,j \geq 0$, вероятностей переходов цепи $\xi_n, n \geq 1$, из состояний со значением i первой компоненты в состояния со значением j этой компоненты.

Лемма 1. Матрица P одношаговых вероятностей переходов цепи Маркова $\xi_n, n \geq 1$, имеет блочную структуру

$$P = (P_{i,j})_{i,j \geq 0} = \begin{pmatrix} V_0 & V_1 & V_2 & V_3 & \cdots \\ Y_0 & Y_1 & Y_2 & Y_3 & \cdots \\ 0 & Y_0 & Y_1 & Y_2 & \cdots \\ 0 & 0 & Y_0 & Y_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

где

$$V_i = \sum_{k=1}^{i+1}[-Q(\Delta \oplus D_0)^{-1}\hat{D}_k + (1-p)H_k\hat{Q}_3\gamma A]\Omega_{i-k+1} +$$

$$+(1-p)\sum_{n=1}^{i} H_n\hat{Q}_3 A \sum_{k=1}^{i-n+1}\hat{D}_k\Omega_{i-n-k+1},$$

$$Y_i = \bar{Q}A[\gamma\Omega_i + \sum_{k=1}^{i}\hat{D}_k\Omega_{i-k}] +$$

$$+(1-p)[\sum_{k=0}^{i} H_k\hat{Q}_3 A\gamma\Omega_{i-k} + \sum_{m=0}^{i-1} H_m\hat{Q}_3 A \sum_{k=1}^{i-m}\hat{D}_k\Omega_{i-m-k}], i \ge 0,$$

$$A = \int_0^\infty e^{-\gamma t}e^{(\Delta\oplus D_0)t}dt = (\gamma I - \Delta \oplus D_0)^{-1},$$

$$\Omega_n = \int_0^\infty e^{\Delta t}\otimes P(n,t)dB(t),\ H_n = \int_0^\infty dH(t)\otimes P(n,t), n \ge 0,$$

$$\Delta = \begin{pmatrix} 0 & 0 & \dots & 0 & 0 \\ \mu & -\mu & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & N\mu & -N\mu \end{pmatrix}.$$

Из Леммы 1 следует, что матрицы переходных вероятностей $P_{i,j}$ цепи Маркова $\xi_n, n \ge 1$, при $i > 0$ зависят от значений счетных компонент i и j только через их разность $j - i$. Согласно определению, данному в статье [3], из этого следует, что рассматриваемая цепь $\xi_n, n \ge 1$, принадлежит классу многомерных квазитеплицевых цепей Маркова.

С использованием результатов [3] доказана следующая

Теорема 1. Необходимым и достаточным условием существования стационарного распределения цепи Маркова $\xi_n, n \ge 1$, является выполнение неравенства

$$\rho = \boldsymbol{x}[F'(1)(\gamma I - \hat{D}_0) + F(1)(\hat{D}'(1) + (\gamma I - \hat{D}_0)\Omega'(1))]\boldsymbol{e} < 1, \quad (1)$$

где вектор $\boldsymbol{x}$ – единственное решение системы линейных алгебраических уравнений

$$\boldsymbol{x}F(1)(\gamma I + \hat{D}(1) - \hat{D}_0)\Omega(1) = \boldsymbol{x},\ \ \boldsymbol{x}\boldsymbol{e} = 1.$$

Здесь $F(z) = (\bar{Q} + (1-p)H(z)\hat{Q}_3)A$, $\Omega(z) = \sum_{n=0}^{\infty}\Omega_n z^n =$

$$= \int_0^\infty e^{\Delta t}\otimes e^{D(z)t}dB(t), H(z) = \sum_{n=0}^{\infty} H_n z^n = \int_0^\infty dH(t)\otimes e^{D(z)t},$$

$|z| \le 1$.

Далее будем считать, что неравенство (1) выполняется. Пусть $\pi(i,r,\nu), i \ge 0, r = \overline{0,N}, \nu = \overline{0,W}$, – искомые стационарные вероятности. Введем обозначения для векторов-строк этих вероятностей:

$$\boldsymbol{\pi}(i,r) = (\pi(i,r,0), \pi(i,r,1), \dots, \pi(i,r,W));$$
$$\boldsymbol{\pi}_i = (\boldsymbol{\pi}(i,0), \boldsymbol{\pi}(i,1), \dots, \boldsymbol{\pi}(i,N)), i \ge 0.$$

Для вычисления векторов $\boldsymbol{\pi}_i, i \ge 0$, используется алгоритм (см. [3]), разработанный для вычисления стационарного распределения многомерных квазитеплицевых цепей Маркова.

4 СТАЦИОНАРНОЕ РАСПРЕДЕЛЕНИЕ ВЕРОЯТНОСТЕЙ СОСТОЯНИЙ СИСТЕМЫ В ПРОИЗВОЛЬНЫЙ МОМЕНТ ВРЕМЕНИ

Состояние системы в произвольный момент времени t определим как $(i_t, r_t, \nu_t, m_t), t \ge 0$, где i_t – число заявок на орбите, r_t – число занятых приборов на второй фазе, ν_t – состояние управляющего процесса $BMAP$-потока, m_t –случайная величина, принимающая значения 0, 1, 2 в зависимости от того, свободен, занят или заблокирован прибор первой фазы в момент времени t.

Рассмотрим процесс $\xi_t = \{i_t, r_t, \nu_t, m_t\}, t \ge 0$. Введем обозначения для стационарных вероятностей этого процесса:

$$p(i,r,\nu,m) = \lim_{t\to\infty} P\{i_t = i, r_t = r, \nu_t = \nu, m_t = m\},$$
$$i \ge 0, r = \overline{0,N}, \nu = \overline{0,W}, m = 0,1,2.$$

Пусть также $\boldsymbol{p}_i(m)$ вектор-строка вероятностей $p(i,r,\nu,m)$, упорядоченных в лексикографическом порядке компонент $(r,\nu), i \ge 0, m = 0,1,2$.

Теорема 2. Ненулевые векторы $\boldsymbol{p}_i(m)$ стационарных вероятностей процесса $\xi_t, t \ge 0$, выражаются через стационарное распределение вложенной цепи Маркова $\boldsymbol{\pi}_i, i \ge 0$, следующим образом:

$$\boldsymbol{p}_0(0) = -\tau^{-1}\boldsymbol{\pi}_0 Q(\Delta \oplus D_0)^{-1},$$

$$\boldsymbol{p}_i(0) = \tau^{-1}[\boldsymbol{\pi}_i\bar{Q} + (1-p)\sum_{l=0}^{i}\boldsymbol{\pi}_l H_{i-l}\hat{Q}_3]A, i \ge 1,$$

$$\boldsymbol{p}_i(1) = \tau^{-1}\{\boldsymbol{\pi}_0\sum_{k=1}^{i+1}[-Q(\Delta\oplus D_0)^{-1}\hat{D}_k + (1-p)H_k\hat{Q}_3\gamma A]\hat{\Omega}_{i-k+1} +$$

$$+(1-p)\boldsymbol{\pi}_0\sum_{k=1}^{i} H_k\hat{Q}_3 A\sum_{m=1}^{i-k+1}\hat{D}_m\hat{\Omega}_{i-k-m+1} +$$

$$+\sum_{l=1}^{i}\boldsymbol{\pi}_l[\bar{Q}A\sum_{k=1}^{i-l+1}\hat{D}_k\hat{\Omega}_{i-l-k+1} +$$

$$+(1-p)\sum_{k=0}^{i-l} H_k\hat{Q}_3 A\sum_{m=1}^{i-l-k+1}\hat{D}_m\hat{\Omega}_{i-l-k-m+1}] +$$

$$+\sum_{l=1}^{i+1}\boldsymbol{\pi}_l[\tilde{Q}A\gamma\hat{\Omega}_{i-l+1}+(1-p)\sum_{k=0}^{i-l+1}H_k\hat{Q}_3A\gamma\hat{\Omega}_{i-l-k+1}]\}, i\geq 0,$$

$$\boldsymbol{p}_i(2)=\tau^{-1}(1-p)\sum_{l=0}^{i}\boldsymbol{\pi}_l\sum_{k=0}^{i-l}(H_k+\delta_{0,k}I)\hat{Q}_2\times$$

$$\times\int_0^\infty e^{-\mu Rt}\otimes P(i-l-k,t)dt, i\geq 0,$$

где $\delta_{i,j}=\begin{cases}0, & i\neq j\\ 1, & i=j\end{cases}$, $\hat{\Omega}_n=\int_0^\infty e^{\Delta t}\otimes P(n,t)(1-B(t))dt, n\geq 0,$

τ – средняя величина интервала между двумя соседними моментами окончания обслуживания на первой фазе:

$$\tau=b_1+(1-p)\sum_{i=0}^{\infty}\boldsymbol{\pi}_i(I_{N+1}\otimes\boldsymbol{e}_{\bar{W}})\int_0^\infty tdH(t)Q_3\boldsymbol{e}+$$

$$+\boldsymbol{\pi}_0Q(\boldsymbol{e}_{N+1}\otimes I_{\bar{W}})(-D_0)^{-1}\boldsymbol{e}_{\bar{W}}+$$

$$+(1-p)\boldsymbol{\pi}_0\sum_{n=1}^{\infty}H_n\hat{Q}_3(\boldsymbol{e}_{N+1}\otimes I_{\bar{W}})(\gamma I-D_0)^{-1}\boldsymbol{e}_{\bar{W}}+$$

$$+\sum_{i=1}^{\infty}\boldsymbol{\pi}_i[\tilde{Q}+(1-p)\sum_{n=0}^{\infty}H_n\hat{Q}_3](\boldsymbol{e}_{N+1}\otimes I_{\bar{W}})(\gamma I-D_0)^{-1}\boldsymbol{e}_{\bar{W}}.$$

Следствие 1. Векторы $\boldsymbol{p}_i$ стационарных вероятностей процесса $\{i_t^*, r_t, \nu_t\}$, где i_t^* – число заявок на первой фазе (на орбите и на приборе, включая заблокированную), вычисляются следующим образом:

$$\boldsymbol{p}_0=\boldsymbol{p}_0(0),\quad \boldsymbol{p}_i=\boldsymbol{p}_i(0)+\sum_{m=1}^{2}\boldsymbol{p}_{i-1}(m), i\geq 1.$$

5 ХАРАКТЕРИСТИКИ ПРОИЗВОДИТЕЛЬНОСТИ

Вычислив стационарные распределения вероятностей состояний системы, можно найти различные характеристики ее производительности:

• Среднее число заявок на первой фазе в моменты окончания обслуживания и в произвольный момент времени: $L=\boldsymbol{\Pi}'(1)\boldsymbol{e}$ и $\bar{L}=\boldsymbol{P}'(1)\boldsymbol{e}$, где

$$\boldsymbol{\Pi}(z)=\sum_{i=0}^{\infty}\boldsymbol{\pi}_iz^i,\ \boldsymbol{P}(z)=\sum_{i=0}^{\infty}\boldsymbol{p}_iz^i, |z|\leq 1.$$

• Векторы стационарного распределения числа занятых приборов на второй фазе в моменты окончания обслуживания на первой фазе и в произвольный момент времени:

$$\boldsymbol{r}=\boldsymbol{\Pi}(1)(I_{N+1}\otimes\boldsymbol{e}_{\bar{W}}),\ \bar{\boldsymbol{r}}=\boldsymbol{P}(1)(I_{N+1}\otimes\boldsymbol{e}_{\bar{W}}).$$

• Среднее число занятых приборов на второй фазе в моменты окончания обслуживания на первой фазе и в произвольный момент времени:

$$N_{busy}=\boldsymbol{r}R\boldsymbol{e},\ \bar{N}_{busy}=\bar{\boldsymbol{r}}R\boldsymbol{e}.$$

• Вероятность P_{loss} того, что произвольная заявка уйдет из системы (потеряется) после обслуживания на первой фазе, и вероятность P_{block} того, что произвольная заявка вызовет блокировку прибора первой фазы: $P_{loss}=p\boldsymbol{\Pi}(1)\hat{Q}_2\boldsymbol{e},\ P_{block}=(1-p)\boldsymbol{\Pi}(1)\hat{Q}_2\boldsymbol{e}.$

• Вероятности $P_{idle}, P_{serve}, P_{block}$ того, что прибор первой фазы свободен, занят или заблокирован:

$$P_{idle}=\sum_{i=0}^{\infty}\boldsymbol{p}_i(0)\boldsymbol{e},\ P_{serve}=\tau^{-1}b_1,\ P_{block}=1-P_{idle}-P_{serve}.$$

• Вероятность того, что произвольная заявка, поступившая в систему, сразу пойдет на обслуживание (не посетив орбиту): $P_{imm}=-\lambda^{-1}\sum_{i=0}^{\infty}\boldsymbol{p}_i(0)(\boldsymbol{e}_{N+1}\otimes D_0\boldsymbol{e}).$

• Вероятность того, что произвольная заявка, поступившая в систему, успешно обслужится на обеих фазах (не попадет на орбиту, не потеряется и не вызовет блокировку прибора первой фазы):

$$P_{success}=-\lambda^{-1}\sum_{i=0}^{\infty}\boldsymbol{p}_i(0)(I_{N+1}\otimes D_0\boldsymbol{e})\int_0^\infty e^{\Delta t}dB(t)Q_1\boldsymbol{e}.$$

Полученные в статье аналитические результаты были реализованы в виде модуля пакета прикладных программ «Sirius++» (см. работу [4]), который позволяет вычислять стационарные распределения системы во вложенные и произвольные моменты времени, а также различные характеристики ее производительности.

ЛИТЕРАТУРА

[1] Двухфазная система обслуживания с групповым марковским потоком и повторными вызовами / В.И. Клименок, О.С. Тарамин // Автоматика и телемеханика. – 2009. –Т. 70. –№ 12.

[2] New results on the single server queue with a batch Markovian arrival process / D. Lucantoni // Communication in Statistics-Stochastic Models. – 1991. – V. 7. – P. 1-46.

[3] Multi-dimensional asymptotically quasi-Toeplitz Markov chains and their application in queueing theory / V.I. Klimenok, A.N. Dudin // Queueing Systems. – 2006. – V. 54. - P. 245-259.

[4] Software "Sirius++" for performance evaluations of modern communication networks / A.N. Dudin, V.I. Klimenok, G.V. Tsarenkov // Modelling and Simulation 2002: Proc. of the 16th European Simulation Multiconference, Darmstadt, June 3-5, 2002, Netherlands: SCS. – 2002. – P. 489-493.