СИСТЕМА ОБСЛУЖИВАНИЯ $MAP|PH|N|R-N$ КАК МОДЕЛЬ ФУНКЦИОНИРОВАНИЯ КОНТАКТ-ЦЕНТРА

С. Дудин, О. Тарамин*

Белорусский государственный университет
Минск, Беларусь
* taramin@mail.ru

В статье исследуется многолинейная система массового обслуживания с конечным буфером и нетерпеливыми запросами. На вход системы поступает марковский поток запросов. Времена обслуживания запросов каждым прибором имеют распределения фазового типа. Во время ожидания обслуживания в буфере запросы могут проявлять нетерпеливость и покидать систему.

Данная работа была частично поддержана Белорусским республиканским фондом фундаментальных исследований через грант № Ф10МЛД-010.

Ключевые слова: марковский поток, процесс обслуживания фазового типа, нетерпеливые запросы, контакт-центр.

1. ВВЕДЕНИЕ

В данной статье исследуется многолинейная система обслуживания с нетерпеливыми запросами. Эта система может использоваться для моделирования и оптимизации функционирования контакт-центра. Обслуживание абонентов осуществляется несколькими операторами. В случае занятости всех операторов абоненту предлагается подождать, при этом ему сообщается его номер в очереди и ориентировочное время ожидания. В зависимости от его номера в очереди абонент принимает решение, ожидать или нет обслуживания. Во время ожидания абонент также может разрывать соединение из-за нетерпеливости.

Считаем, что от обслуживания каждого абонента система получает прибыль, в то время как содержание оператора требует затрат. Необходимо определить такое число операторов, при котором прибыль центра была бы максимальной.

2. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

Рассматривается система обслуживания, состоящая из N приборов и буфера размера $R-N$, $R \geq N$. В систему поступает марковский входной поток (MAP – Markovian Arrival Process) запросов, заданный неприводимой цепью Маркова ν_t, $t \geq 0$, с непрерывным временем и конечным пространством состояний $\{0, ..., W\}$.

Время пребывания цепи в состоянии ν экспоненциально распределено с положительным параметром λ_ν. Когда время пребывания в состоянии ν истекло, с вероятностью $p^{(k)}_{\nu,\nu'}$ процесс ν_t переходит в состояние ν' и при этом генерируется группа из k запросов, $k = 0, 1, \nu, \nu' = \overline{0, W}$. Поведение MAP-потока полностью характеризуется матрицами $D_k, k = 0, 1$, которые определяются следующим образом

$$(D_k)_{\nu,\nu'} = \lambda_\nu p^{(k)}_{\nu,\nu'}, k = 1 \text{ и } k = 0, \nu \neq \nu', \ (D_0)_{\nu,\nu} = -\lambda_\nu, \nu = \overline{0, W}.$$

Матрица $D(1) = D_0 + D_1$ представляет собой инфинитезимальный генератор цепи $\nu_t, t \geq 0$.

Средняя интенсивность поступления запросов λ имеет вид

$$\lambda = \boldsymbol{\chi} D_1 \mathbf{e},$$

где $\boldsymbol{\chi}$ - вектор стационарного распределения цепи Маркова $\nu_t, t \geq 0$. Вектор $\boldsymbol{\chi}$ является единственным решением системы линейных алгебраических уравнений $\boldsymbol{\chi} D(1) = \mathbf{0}$, $\boldsymbol{\chi}\mathbf{e} = 1$. Здесь и далее $\mathbf{e}$ – вектор-столбец, состоящий из единиц, $\mathbf{0}$ – вектор-строка, состоящая из нулей.

Коэффициент вариации c_{var} длин интервалов между моментами поступления запросов

$$c^2_{var} = 2\lambda\boldsymbol{\chi}(-D_0)^{-1}\mathbf{e} - 1.$$

Коэффициент корреляции c_{cor} длин двух соседних интервалов между поступлением запросов

$$c_{cor} = (\lambda\boldsymbol{\chi}(-D_0)^{-1}(D(1) - D_0)(-D_0)^{-1}\mathbf{e} - 1)/c^2_{var}.$$

Время обслуживания запроса каждым прибором имеет распределение фазового. Время обслуживания, имеющее распределение фазового типа (PH – Phase-type distribution) с неприводимым представлением $(\boldsymbol{\beta}, S)$, можно интерпретировать как время, за которое цепь Маркова с непрерывным временем $\eta_t, \ t \geq 0$, имеющая несущественные состояния $\{1, \dots, M\}$ и поглощающее состояние $M + 1$, достигнет поглощающего состояния. Начальное состояние цепи Маркова $\eta_t, \ t \geq 0$, в момент начала обслуживания запроса определяется вероятностным вектором $\boldsymbol{\beta} = (\beta_1, \dots, \beta_M)$. Переходы цепи $\eta_t, \ t \geq 0$, которые не приводят к окончанию обслуживания, задаются субгенератором S размера $M \times M$. Интенсивности переходов в поглощающее состояние описываются вектором $\mathbf{S}_0 = -S\mathbf{e}$. Среднее время обслуживания определяется формулой $b_1 = \boldsymbol{\beta}(-S)^{-1}\mathbf{e}$.

Если все приборы заняты и число запросов в буфере в момент прихода запроса равно i, $i = \overline{0, R - N - 1}$, то этот запрос с вероятностью q_i покидает систему, а с дополнительной вероятностью становится в очередь.

Если в момент прихода запроса свободные места в буфере отсутствуют, запрос покидает систему. Через экспоненциально распределенное с параметром $\alpha, 0 < \alpha < \infty$, время с момента попадания в буфер, запрос покидает систему, если он до этого не попал на обслуживание.

Качество функционирования системы описывается критерием

$$J(N) = a\lambda_{out} - b\lambda P^{(loss)} - dN,$$

который определяет прибыль системы в единицу времени. Здесь a – прибыль, полученная от обслуживания одного запроса, λ_{out} – среднее число запросов, обслуженных системой в единицу времени, b – штраф за потерю запроса, $P^{(loss)}$ – вероятность потери произвольного запроса, d – стоимость содержания одного прибора в единицу времени.

Необходимо определить такое число приборов N, при котором достигает максимума критерий $J(N)$.

3. ПРОЦЕСС ИЗМЕНЕНИЯ СОСТОЯНИЙ СИСТЕМЫ И СТАЦИОНАРНОЕ РАСПРЕДЕЛЕНИЕ ЧИСЛА ЗАПРОСОВ В СИСТЕМЕ

Пусть $i_t,\ i_t \in \{0,\ldots,R\}$, – число запросов в системе,

$\nu_t \in \{0,\ldots,W\}$, – состояние управляющего процесса MAP-потока поступления запросов,

$\eta_t^{(l)},\ \eta_t^{(l)} \in \{1,\ldots,M\}$, – состояние управляющего процесса обслуживания l-м занятым прибором в момент времени $t,\ t \geq 0,\ l \in \{1,\ldots,\min\{i_t,N\}\}$.

Процесс $\xi_t = \{i_t,\nu_t,\eta_t^{(1)},\ldots,\eta_t^{(\min\{i_t,N\})}\},\ t \geq 0$, является регулярной неприводимой цепью Маркова с непрерывным временем.

Занятые приборы предполагаются перенумерованными в порядке их занятия: номер 1 имеет наиболее долготекущий к данному моменту времени процесс обслуживания. Соответственно максимальный номер имеет процесс обслуживания, которые стартовал позже всех остальных. При завершении процесса обслуживания запроса и отсутствии очереди соответствующий ему номер прибора вычеркивается из записи, и производится перенумерация.

Перенумеруем состояния цепи ξ_t в лексикографическом порядке. Множество состояний, имеющих значение (i,ν) двух первых компонент цепи, будем называть макросостоянием (i,ν).

Пусть Q – генератор цепи Маркова $\xi_t,\ t \geq 0$, сформированный из блоков $Q_{i,j}$, состоящих из матриц $(Q_{i,j})_{\nu,\nu'}$ интенсивностей переходов цепи $\xi_t,\ t \geq 0$, из макросостояния (i,ν) в макросостояние $(j,\nu'),\ \nu,\nu' = \overline{0,W}$.

Введем следующие обозначения:

- $A^{\oplus\, i} = \underbrace{A \oplus A \oplus \ldots \oplus A}_{i}, i > 0,\ A^{\oplus\, 0} = O_{1\times 1};$
- $\bar{W} = W + 1;$
- $\alpha_i = (i - N)\alpha.$

Лемма 1. *Генератор Q имеет следующую блочно-трехдиагональную структуру:*

$$Q = \begin{pmatrix} Q_{0,0} & Q_{0,1} & O & \dots & O & O \\ Q_{1,0} & Q_{1,1} & Q_{1,2} & \dots & O & O \\ O & Q_{2,1} & Q_{2,2} & \dots & O & O \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ O & O & O & \dots & Q_{R-1,R-1} & Q_{R-1,R} \\ O & O & O & \dots & Q_{R,R-1} & Q_{R,R} \end{pmatrix}.$$

Ненулевые блоки генератора имеют следующий вид:

$$Q_{i,i} = D_0 \oplus S^{\oplus i},\ i = \overline{0, N-1},$$

$$Q_{N,N} = D_0 \oplus S^{\oplus N} + q_0 D_1 \otimes I_{M^N},$$

$$Q_{i,i} = D_0 \oplus S^{\oplus N} + q_{i-N} D_1 \otimes I_{M^N} - \alpha_i I_{\bar{W}M^N},\ i = \overline{N+1, R-1},$$

$$Q_{R,R} = D(1) \oplus S^{\oplus N} - \alpha_i I_{\bar{W}M^N},$$

$$Q_{i,i-1} = I_{\bar{W}} \otimes \mathbf{S_0}^{\oplus i},\ i = \overline{1, N},$$

$$Q_{i,i-1} = I_{\bar{W}} \otimes (\mathbf{S_0}\boldsymbol{\beta})^{\oplus N} + \alpha_i I_{\bar{W}M^N},\ i = \overline{N+1, R},$$

$$Q_{i,i+1} = D_1 \otimes I_{M^i} \otimes \boldsymbol{\beta},\ i = \overline{0, N-1},$$

$$Q_{i,i+1} = (1 - q_{i-N}) D_1 \otimes I_{M^N},\ i = \overline{N, R-1}.$$

Доказательство. Доказательство леммы опирается на анализ переходов цепи Маркова ξ_t, $t \geq 0$, за бесконечно малый интервал времени с последующей группировкой интенсивностей соответствующих переходов в блоки матрицы Q. □

Так как цепь Маркова $\xi_t = \{i_t, \nu_t, \eta_t^{(1)}, \dots, \eta_t^{(\min\{i_t, N\})}\}$, $t \geq 0$, регулярная и неприводимая, а также имеет конечное пространство состояний, то существуют следующие пределы (стационарные вероятности):

$$\pi(i, \nu, \eta^{(1)}, \dots, \eta^{(\min\{i,N\})}) = \lim_{t\to\infty} P\{i_t = i,\ \nu_t = \nu,\ \eta_t^{(1)} = \eta^{(1)}, \dots, \eta_t^{(\min\{i_t,N\})} = \eta^{(\min\{i,N\})}\},$$

$$i = \overline{0, R},\ \nu = \overline{0, W},\ \eta^{(l)} = \overline{1, M},\ l \in \{1, \dots, \min\{i, N\}\}.$$

Перенумеруем стационарные вероятности цепи ξ_t в лексикографическом порядке и сформируем из них векторы $\boldsymbol{\pi}(i, \nu)$, состоящие из вероятностей $\pi(i, \nu, \eta^{(1)}, \dots, \eta^{(\min\{i,N\})})$, записанных в лексикографическом порядке по компонентам $(\eta^{(1)}, \dots, \eta^{(\min\{i,N\})})$, и векторы

$$\boldsymbol{\pi}_i = (\boldsymbol{\pi}(i, 0), \boldsymbol{\pi}(i, 1), \dots, \boldsymbol{\pi}(i, W)),\ i = \overline{0, R}.$$

Известно, что вектор-строка $(\boldsymbol{\pi}_0, \dots, \boldsymbol{\pi}_R)$ – единственное решение системы

$$(\boldsymbol{\pi}_0, \dots, \boldsymbol{\pi}_R) Q = \mathbf{0},\ \ (\boldsymbol{\pi}_0, \dots, \boldsymbol{\pi}_R)\mathbf{e} = 1. \qquad (1)$$

В случае, когда размерность системы (1) невелика, эту систему легко решить с помощью компьютера стандартными методами. В противном случае, может быть использован специальный устойчивый алгоритм, решающий систему такого вида, который был разработан в [1].

С учетом того, что генератор Q в данной модели имеет блочно-трехдиагональную форму, в то время как в [1] допускается блочная верхне-Хессенбергова форма, алгоритм из [1] можно записать в более простом виде.

Теорема 1. *Компоненты* $\boldsymbol{\pi}_i,\ i = \overline{0, R}$, *вектора решения системы (1) связаны соотношениями:*

$$\boldsymbol{\pi}_i = \boldsymbol{\pi}_0 F_i, i = \overline{1, R},$$

где матрицы F_i *вычисляются рекуррентно:*

$$F_0 = I, \quad F_i = -F_{i-1}Q_{i-1,i}(Q_{i,i} + Q_{i,i+1}G_i)^{-1}, i = \overline{1, R},$$

матрицы $G_i, i = \overline{0, R-1}$, *вычисляются с помощью обратной рекурсии:*

$$G_i = -(Q_{i+1,i+1} + Q_{i+1,i+2}G_{i+1})^{-1}Q_{i+1,i}, \quad i = R-2, R-3, \ldots, 0,$$

при начальном условии

$$G_{R-1} = -(Q_{R,R})^{-1}Q_{R,R-1},$$

а вектор $\boldsymbol{\pi}_0$ *является единственным решением следующей системы:*

$$\boldsymbol{\pi}_0(Q_{0,0} + Q_{0,1}G_0) = \mathbf{0}, \quad \boldsymbol{\pi}_0 \sum_{l=0}^{R} F_l \mathbf{e} = 1.$$

Доказательство. Доказательство вытекает непосредственно из результатов, полученных в [1]. □

4. ХАРАКТЕРИСТИКИ ПРОИЗВОДИТЕЛЬНОСТИ СИСТЕМЫ

Найдя векторы стационарных вероятностей $\boldsymbol{\pi}_i,\ i = \overline{0, R}$, можно вычислить различные характеристики производительности системы.

Распределение вероятностей числа запросов в системе вычисляется как

$$\lim_{t\to\infty} P\{i_t = i\} = \boldsymbol{\pi}_i \mathbf{e},\ i = \overline{0, R}.$$

Среднее число занятых приборов

$$\bar{N} = \sum_{i=1}^{N} i\boldsymbol{\pi}_i \mathbf{e} + N \sum_{i=N+1}^{R} \boldsymbol{\pi}_i \mathbf{e}.$$

Интенсивность выходящего потока обслуженных запросов

$$\lambda_{out} = \sum_{i=0}^{N} \boldsymbol{\pi}_i(\mathbf{e}_{\bar{W}} \otimes \sum_{l=0}^{i-1}(\mathbf{e}_{M^l} \otimes \mathbf{S}_0 \otimes \mathbf{e}_{M^{i-l-1}})) + \sum_{i=N+1}^{R} \boldsymbol{\pi}_i(\mathbf{e}_{\bar{W}} \otimes \sum_{l=0}^{N-1}(\mathbf{e}_{M^l} \otimes \mathbf{S}_0 \otimes \mathbf{e}_{M^{N-l-1}})).$$

Вероятность потери запроса на входе в систему из-за заполненности буфера определяется как

$$P^{(ent-loss)} = \lambda^{-1}\boldsymbol{\pi}_R(D_1 \otimes I_{M^N})\mathbf{e}.$$

Вероятность ухода запроса на входе в систему из-за нежелания ожидать:

$$P^{(esc-loss)} = \lambda^{-1}\sum_{i=N}^{R-1} q_{i-N}\boldsymbol{\pi}_i(D_1 \otimes I_{M^N})\mathbf{e}.$$

Вероятность потери запроса:

$$P^{(loss)} = 1 - \frac{\lambda_{out}}{\lambda}.$$

Вероятность ухода запроса из очереди из-за нетерпеливости:

$$P^{(imp-loss)} = P^{(loss)} - P^{(ent-loss)} - P^{(esc-loss)}.$$

5. ЗАКЛЮЧЕНИЕ

В данной статье исследована многолинейная система массового обслуживания с конечным буфером и нетерпеливыми запросами. Построен процесс функционирования данной системы, найдено стационарное распределение вероятностей состояний системы, получены формулы для нахождения основных характеристик производительности.

ЛИТЕРАТУРА

1. *Klimenok V., Kim C.S., Orlovsky D., Dudin A.* Lack of invariant property of Erlang loss model in case of the MAP input // Queueing Systems. 2005. V. 49. P. 187–213.