ОБ АЛЬТЕРНАТИВНОМ АЛГОРИТМЕ РАСЧЕТА ХАРАКТЕРИСТИК СИСТЕМЫ ОБСЛУЖИВАНИЯ *BMAP|PH|N|0*

Д. Орловский

Белорусский государственный университет
Минск, Беларусь
odms@tut.by

Рассмотрена система массового обслуживания $BMAP|PH|N|0$, являющаяся обобщением известной системы Эрланга с отказами. Предлагается эффективный алгоритм расчета стационарных характеристик системы.

Ключевые слова: СМО $BMAP|PH|N|0$, стационарное распределение, система Эрланга с отказами.

1. ВВЕДЕНИЕ

Среди фундаментальных работ теории массового обслуживания особое место занимают пионерские работы А. К. Эрланга по исследованию многолинейных систем со стационарным пуассоновским входным потоком и экспоненциально распределенным временем обслуживания.

Изучая систему $M|M|N|0$, Эрланг получил формулу для вероятности отказа в обслуживании (из-за отсутствия свободных обслуживающих приборов) запросу, поступающему в систему. Построенные на ее основании таблицы вероятностей отказа для фиксированного числа обслуживающих приборов при заданной загрузке системы до сих пор используются при проектировании телефонных сетей.

Однако за последние несколько лет характер передаваемых по каналам связи данных существенно изменился. Так, вопреки используемым Эрлангом предпосылкам, потоки данных в современных сетях уже нельзя считать ни ординарными, ни стационарными, ни потоками без последействия. Это ставит под сомнение пригодность формулы Эрланга для адекватной оценки вероятности отказа в таких сетях и заставляет принимать к рассмотрению более общие системы массового обслуживания.

Среди различных модификаций системы Эрланга можно отметить, например, [1, 2]. В [1] рассматривалась многолинейная система $BMAP|PH|N|0$ с групповым марковским входным потоком (Batch Markovian Arrival Process – *BMAP*, подробнее см., например, [3]), фазовым распределением времени обслуживания (Phase(PH)-type distribution, см., например, [4]) и без буфера для трех различных дисциплин принятия заявок (частичное принятие, полное принятие и полный отказ). Был предложен алгоритм для вычисления стационарных вероятностей состояний системы и расчета важнейшей ее характеристики – вероятности отказа.

Однако для систем с более чем одной фазой PH-обслуживания предложенный алгоритм позволил получить результаты лишь при малом (порядка 10) числе обслуживающих приборов. Ограничивающим фактором стала размерность используемых при расчетах матриц. Причина этого кроется в характерной особенности всех систем с PH-обслуживанием – быстром росте размерности фазового пространства процесса изменений состояний системы с ростом числа приборов из-за необходимости отслеживать состояния управляющих процессов каждого из них.

В практических задачах число приборов обычно значительно больше 10, поэтому возникает необходимость создания алгоритмов, которые существенно расширили бы диапазон параметров поддающихся анализу систем. Один из возможных путей – пересмотр способа построения процесса изменений состояний системы.

Подход, используемый в [1], состоит в отслеживании номера фазы PH-обслуживания на каждом из i_t, $i_t = \overline{1,N}$, $t \geq 0$, занятых приборов. Для этого вводится многомерный процесс $m_t = \{m_t^{(1)}, \ldots, m_t^{(i_t)}\}$, где $m_t^{(k)}$ – номер фазы обслуживания, на которой в момент времени t находится k-й прибор, $t \geq 0$, $k = \overline{1,N}$. Положительная особенность такого подхода – прозрачность аналитических результатов, отрицательная – большая размерность пространства состояний процесса m_t, $t \geq 0$.

Так, для дисциплины частичного принятия заявок размерность инфинитезимального генератора (ИГ) системы оказалась равна $K = \bar{W}\frac{M^{N+1}-1}{M-1}$, где $\bar{W}$ – число состояний управляющего процесса $BMAP$. При $\bar{W} = 3$, $M = 3$, $N = 25$ это составляет $K \approx 4 \cdot 10^{12}$, а при $N = 30$ соответственно $K \approx 10^{15}$. Вычисления с матрицами такой размерности практически неосуществимы.

Альтернативный подход к исследованию систем с идентичными обслуживающими приборами предложен Рамасвами в работе [5]. При этом рассматривается процесс $\psi_t = \{h_t^{(1)}, \ldots, h_t^{(M)}\}$, где $h_t^{(m)}$ – число занятых приборов, находящихся в момент t на m-й фазе обслуживания, $m = \overline{1,M}$, $h_t^{(1)} + \ldots + h_t^{(M)} = i_t$, $t \geq 0$.

В настоящей работе предлагается модификация с использованием подхода Рамасвами, представленного в [1] алгоритма поиска стационарных вероятностей и вероятности отказа для системы $BMAP|PH|N|0$ в случае дисциплины частичного принятия. Быстродействия исходного и модифицированного алгоритмов, а также верхние границы для числа обслуживающих приборов, при котором эти алгоритмы применимы на практике, будут сопоставлены в ходе вычислительных экспериментов.

2. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

Входящий $BMAP$-поток определяется неприводимой цепью Маркова с непрерывным временем (ЦМНВ) ν_t, $t \geq 0$, с пространством состояний $\{0, \ldots, W\}$ (т. н. управляющим процессом) и набором матриц D_k, $k \geq 0$, элементы которых характеризуют интенсивности переходов ν_t, $t \geq 0$, приводящих к генерации пакетов из k заявок. Пусть $D(z) = \sum_{k=0}^{\infty} D_k z^k$, $|z| \leq 1$. Матрица $D(1)$ – ИГ процесса ν_t. Вектор $\boldsymbol{\theta}$ стационарных вероятностей состояний ЦМ ν_t является решением системы $\boldsymbol{\theta}D(1) = \mathbf{0}$, $\boldsymbol{\theta}\boldsymbol{e} = 1$. Здесь и далее $\boldsymbol{e}$ – вектор-столбец, состоящий из единиц, размер которого определяется из контекста. Средняя интенсивность $BMAP$ $\lambda = \boldsymbol{\theta}D'(1)\boldsymbol{e}$.

В системе имеется N идентичных обслуживающих приборов. Распределение времени обслуживания – фазовое, определяется управляющим процессом m_t, $t \geq 0$, ЦМНВ с пространством состояний $\{0, \ldots, M\}$, где состояние 0 – поглощающее. Завершению обслуживания соответствует попадание в поглощающее состояние. Начальное состояние определяется вероятностным вектором-строкой $(0, \boldsymbol{\beta})$. ИГ процесса m_t, $t \geq 0$, есть матрица $\tilde{S} = \begin{pmatrix} 0 & 0 \\ S^0 & S \end{pmatrix}$, где $\boldsymbol{S}^0 = -S\boldsymbol{e}$.

Под состоянием системы в момент t будем понимать вектор $(i_t, \nu_t, \boldsymbol{h}_t)$, где i_t – число заявок в системе, ν_t – состояние управляющего процесса $BMAP$, $\boldsymbol{h}_t = (h_t^{(1)}, \ldots, h_t^{(M)})$ – вектор, компонента $h_t^{(m)}$ которого есть число занятых приборов, находящихся на m-й фазе обслуживания в момент времени t. Процесс изменения состояний системы есть многомерная ЦМНВ $\xi_t = \{i_t, \nu_t, \boldsymbol{h}_t\}$, $i_t = \overline{0, N}$, $\nu_t = \overline{0, W}$, $m = \overline{1, M}$, $h_t^{(1)} + \ldots + h_t^{(M)} = i_t$, $t \geq 0$.

Пусть $p(i, \nu, h^{(1)}, \ldots, h^{(M)}) = \lim\limits_{t \to \infty} P\{i_t = i, \nu_t = \nu, h_t^{(1)} = h^{(1)}, \ldots, h_t^{(M)} = h^{(M)}\}$.

Упорядочим состояния ξ_t, $t \geq 0$ в лексикографическом порядке по компонентам i_t, ν_t. Состояния с равными значениями i_t и ν_t, в свою очередь, доупорядочим в обратном лексикографическом порядке по компонентам $h^{(m)}$ вектора $\boldsymbol{h}$.

Введем вектор-строки $\boldsymbol{p}_i$, $i = \overline{0, N}$ стационарных вероятностей, соответствующих значению i первой компоненты процесса ξ_t, $t \geq 0$ и вектор $\boldsymbol{p} = (\boldsymbol{p}_0, \ldots, \boldsymbol{p}_N)$.

Вектор $\boldsymbol{p}$ удовлетворяет системе линейных алгебраических уравнений

$$\boldsymbol{p}Q = \boldsymbol{0}, \quad \boldsymbol{p}\boldsymbol{e} = 1, \tag{1}$$

где Q – ИГ ЦМ ξ_t, $t \geq 0$.

Найдем для состояния $z = (i, \nu, h_1, \ldots, h_M)$ процесса ξ_t, $t \geq 0$ его номер p_z при таком упорядочении. Нумерацию начинаем с нуля. Пусть $L(\hat{i}, \hat{\nu}, \hat{h}_1, \ldots, \hat{h}_k)$ – число состояний вида $(\hat{i}, \hat{\nu}, \hat{h}_1, \ldots, \hat{h}_k, h_{k+1}, \ldots, h_M)$, $\hat{i} = \overline{0, N}$, $\hat{\nu} = \overline{0, W}$, $k = \overline{0, M-1}$, $\hat{h}_1 + \ldots + \hat{h}_k \leq i$. Несложно убедиться, что эта величина не зависит от $\hat{\nu}$ и равна $L\left(\hat{i}, \hat{\nu}, \hat{h}_1, \ldots, \hat{h}_k\right) = L\left(\hat{i}, \hat{h}_1, \ldots, \hat{h}_k\right) = C^{M-\hat{k}-1}_{\hat{i}-(\hat{h}_1+\ldots+\hat{h}_k)+M-\hat{k}-1}$. Тогда $p_z = \bar{W}C^{M}_{i+M-1} + \hat{\nu}C^{M-1}_{i+M-1} + \sum_{k=1}^{M-1} C^{M-k}_{i-(h_1+\ldots+h_k+1)+M-k}$.

Пусть в момент времени t процесс ξ_t, $t \geq 0$, находится в состоянии $z_t = (i, \nu, \boldsymbol{h})$. Рассмотрим всевозможные события, которые могут произойти за бесконечно малый промежуток времени $[t, t+\delta[$ с вероятностью, превосходящей $o(\delta)$.

1) Переход ν_t из состояния ν в состояние ν' с генерацией пакета из k заявок.
2) Переход ν_t из состояния ν в состояние ν' без генерации заявок.
3) Изменения номера фазы PH-обслуживания на одном из занятых приборов с m на m' без завершения обслуживания.
4) Завершение обслуживание заявки прибором, находящимся на m-й фазе PH-обслуживания.

Интенсивности этих событий и состояния, устанавливающиеся в результате их наступления, приведены в табл. 1. Символом $\boldsymbol{e}_l$ в таблице обозначен вектор-строка размера M, l-й элемент которого равен 1, а остальные элементы – нули.

Используя табл. 1, ИГ Q процесса ξ_t, $t \geq 0$, можно построить поэлементно, однако удобнее воспользоваться для этого предложенным Лукантони и Рамасвами в работах [5, 6] алгоритмом.

Лемма. *ИГ процесса* ξ_t, $t \geq 0$, *имеет блочную структуру* $Q = (Q_{i,j})$, $i, j = \overline{0, N}$. *Блоки* $Q_{i,j}$ *размера* $\bar{W}L(i) \times \bar{W}L(j)$ *определяются как:*

$$Q_{i,j} = 0, \quad i = \overline{2, N},\ j = \overline{i-2, N-2}, \qquad Q_{i,i-1} = I_{\bar{W}} \otimes L_{N-i}(N, \tilde{S}), \quad i = \overline{1, N},$$

$$Q_{i,i} = D_0 \oplus A(i, S) + \Delta_i, \quad Q_{i,N} = \sum_{k=0}^{\infty} D_{N-i+k} \otimes P_{i,N}(\beta), \quad i = \overline{0, N-1},$$

$$Q_{i,j} = D_{j-i} \otimes P_{i,j}(\beta), \quad i = \overline{1, N-1},\ j = \overline{i+1, N-1}, \qquad Q_{N,N} = \sum_{k=0}^{\infty} D_k \oplus A(N, S) + \Delta_N,$$

где $I_{\bar{W}}$ *– единичная матрица размерности* $\bar{W}$, $P_{i,j}(\beta) = P_i(\beta)P_{i+1}(\beta)\ldots P_{j-1}(\beta)$, $0 \leq i < j \leq N$, Δ_i, $i = \overline{0, N}$ *– диагональные матрицы, элементы которых подобраны так, чтобы* $Qe = \mathbf{0}$.

Схема вычисления матриц $P_c(\boldsymbol{b})$ приведена в [6], матриц $A(c, T)$ и $L_i(c, T)$ – в [5]. Доказательство основывается на анализе переходов ЦМНВ ξ_t, $t \geq 0$, с использованием результатов [5, 6]. Оно достаточно прозрачно и здесь нами опущено.

Размерность построенного таким образом ИГ Q равна $K = \bar{W} \sum_{i=0}^{N} L(i) = \bar{W} C_{N+M}^{M}$. При $\bar{W} = 3$, $M = 3$, $N = 25$ это составляет $K = 9828$, при $N = 30$ получаем $K = 16368$. Напомним, что для этих же наборов параметров предложенный в [1] подход давал соответственно $K \approx 4 \cdot 10^{12}$ и $K \approx 10^{15}$.

Для решения системы (1) применим приведенный в [1] алгоритм.

1) Вычисляем матрицы $G_i, i = \overline{0, N-1}$ из обратной рекурсии:

$$G_i = (-Q_{i+1,i+1} - \sum_{l=1}^{N-i-1} Q_{i+1,i+1+l} G_{i+l} G_{i+l-1} \ldots G_{i+1})^{-1} Q_{i+1,i},\ i = N-1, \ldots, 0.$$

2) Определяем матрицы $\bar{Q}_{i,l}$, $i = \overline{0, l}$, $l = \overline{0, N}$ из обратной рекурсии:

$$\bar{Q}_{i,N} = Q_{i,N},\ i = \overline{0, N}, \qquad \bar{Q}_{i,l} = Q_{i,l} + \bar{Q}_{i,l+1} G_l,\ i = \overline{0, l},\ l = N-1, N-2, \ldots, 0.$$

3) Вычисляем матрицы F_l, $l = \overline{1, N}$ из соотношений

$$F_l = (\bar{Q}_{0,l} + \sum_{i=1}^{l-1} F_i \bar{Q}_{i,l})(-\bar{Q}_{l,l})^{-1},\ l = \overline{1, N-1}, \qquad F_N = (Q_{0,N} + \sum_{i=1}^{N-1} F_i Q_{i,N})(-Q_{N,N})^{-1}.$$

4) Находим вектор $\boldsymbol{p}_0$ как единственное решение системы линейных уравнений $\boldsymbol{p}_0 \bar{Q}_{0,0} = \mathbf{0}$, $\quad \boldsymbol{p}_0(\sum_{l=1}^{N} F_l \boldsymbol{e} + \boldsymbol{e}) = 1$.

5) Вычисляем векторы $\boldsymbol{p}_l = \boldsymbol{p}_0 F_l$, $l = \overline{1, N}$. Вероятность отказа P_{loss} можно найти по формуле: $P_{loss} = 1 - \lambda^{-1} \sum_{i=0}^{N-1} \vec{p}_i \sum_{k=0}^{N-i} (k + i - N)(D_k \otimes I_{L(i)})\mathbf{e}$.

Описанный в [1] алгоритм расчета стационарных вероятностей состояний системы (далее – алгоритм А) и алгоритм, представленный в данной статье (алгоритм Б), были реализованы в рамках пакета прикладных программ “Sirius-M”. С их помощью была произведена серия экспериментов, имевших целью сравнение быстродействия алгоритмов и сопоставление диапазона параметров исследуемых систем, при которых эти алгоритмы практически применимы.

Для экспериментов использовался $BMAP$, заданный матрицами D_k, $k = \overline{0, 10}$ третьего порядка и PH-распределения с двумя, тремя и четырьмя фазами, обозначенные соответственно как PH_2, PH_3 и PH_4.

Конфигурация ЭВМ: Intel Pentium 4 (1.8GHz), 512 MB RAM, 3000 MB pagefile.

Результаты экспериментов представлены в табл. 2. Колонка 'алгоритм' определяет используемый алгоритм (А/Б), N – число обслуживающих приборов, K – размерность ИГ. '+' в колонке 'результат' означает успешное завершение алгоритма, '–' – отказ из-за нехватки памяти, 'время работы' – астрономическое время работы алгоритма до завершения/отказа. Колонка 'расход памяти' содержит максимальное за время работы алгоритма количество (в МБ) используемой им памяти.

3. ТАБЛИЦЫ

Таблица 1

Возможные переходы системы

События	$z_{t+\delta}$	Интенсивность перехода	Условие
1	$(i+k, \nu', \boldsymbol{h} + \boldsymbol{e}_{l_1} + \ldots + \boldsymbol{e}_{l_k})$	$(D_k)_{\nu\nu'} \boldsymbol{\beta}_{l_1} \ldots \boldsymbol{\beta}_{l_k}$	$i < N$, $1 \le k < N - i$
	$(N, \nu', \boldsymbol{h} + \boldsymbol{e}_{l_1} + \ldots + \boldsymbol{e}_{l_{N-i}})$	$(\sum\limits_{k=N-i}^{\infty} D_k)_{\nu\nu'} \boldsymbol{\beta}_{l_1} \ldots \boldsymbol{\beta}_{l_{N-i}}$	$i < N$
1,2	$(N, \nu', \boldsymbol{h})$	$(\sum\limits_{k=0}^{\infty} D_k)_{\nu\nu'}$	$i = N$, $\nu \ne \nu'$
2	$(i, \nu', \boldsymbol{h})$	$(D_0)_{\nu\nu'}$	$i < N$, $\nu \ne \nu'$
3	$(i, \nu, \boldsymbol{h} - \boldsymbol{e}_m + \boldsymbol{e}_{m'})$	$h^{(m)} S_{mm'}$	$m \ne m'$, $\boldsymbol{h}\boldsymbol{e}_m \ge 1$
4	$(i-1, \nu, \boldsymbol{h} - \boldsymbol{e}_m)$	$h^{(m)} (S_0)_m$	$i \ge 1$, $\boldsymbol{h}\boldsymbol{e}_m \ge 1$

Таблица 2

Результаты экспериментов

Алгоритм	PH	N	K	Результат	Время работы	Расход памяти
А	PH_2	10	8188	+	3:09:11	571.8
А	PH_2	11	16380	–	0:07:06	1295.2
А	PH_3	6	4372	+	0:08:45	264.2
А	PH_3	7	13120	–	0:08:41	1247.6
А	PH_4	5	5460	+	1:36:59	264.2
А	PH_4	6	21844	–	0:01:15	860.1
Б	PH_2	10	312	+	0:00:00.03	0.4
Б	PH_3	6	480	+	0:00:00.06	0.8
Б	PH_4	5	840	+	0:00:00.25	1.9
Б	PH_2	154	48360	+	0:55:01	1942.8
Б	PH_2	155	48984	–	0:49:49	1960.7
Б	PH_3	30	21824	+	2:09:14	1844.6
Б	PH_3	31	23936	–	2:11:44	1915.2

Окончание табл. 2

Б	PH_4	15	15504	+	2:56:38	1326.3
Б	PH_4	16	19380	–	5:22:36	1829.7

4. ЗАКЛЮЧЕНИЕ

Из результатов экспериментов видно, что приведенный в данной работе алгоритм работает намного быстрее и позволяет исследовать системы со значительно большим числом обслуживающих приборов, чем алгоритм [1]. Это свидетельствует о целесообразности использования подхода Рамасвами при исследовании некоторых других многолинейных систем с PH-обслуживанием.

ЛИТЕРАТУРА

1. *Klimenok V. I., Kim C. S., Orlovsky D. S., Dudin A. N.* Lack of invariant property of Erlang BMAP|PH|N|0 model // Queueing Systems (submitted 10.2003).
2. *Krieger U. R., Naumov V.* Analysis Of A Versatile Queuing Model With State-Dependent Service Times // MMB, 1999. С. 121–135.
3. *Lucantoni D.* New results on the single server queue with a batch Markovian arrival process // Comm. Statist.-Stochastic Models. 1991. V. 7. P. 1–46.
4. *Neuts M.* Matrix-geometric solutions in stochastic models. Baltimore: John Hopkins University Press, 1981.
5. *Ramaswami V.* Independent Markov Processes in parallel // Comm. Statist.-Stochastic Models. 1985. V. 1(3). P. 419–432.
6. *Ramaswami V., Lucantoni D. M.* Algorithms for the multi-server queue with phase-type service. Independent Markov Processes in parallel // Comm. Statist.-Stochastic Models. 1985. V. 1(3). P. 393–417.