СТАЦИОНАРНЫЕ ВЕРОЯТНОСТИ СОСТОЯНИЙ СИСТЕМЫ С ДВУМЯ ПРИОРИТЕТАМИ И РЕЗЕРВИРОВАНИЕМ КАНАЛОВ

А. Печинкин

Институт проблем информатики РАН
Москва, Россия
apechinkin@ipiran.ru

Рассматривается марковская модель двухприоритетной системы, обобщающей систему с резервированием каналов пучка, марковским входящим потоком и различными распределениями фазового типа времен обслуживания заявок каждого приоритета. Резервирование каналов пучка означает наличие некоторого числа каналов, которые могут занимать только приоритетные заявки. Показано, как можно вычислить стационарные вероятности состояний этой марковской модели.

Ключевые слова: система массового обслуживания; резервирование каналов пучка; стационарные вероятности состояний.

1. ВВЕДЕНИЕ

Одной из возможных схем обслуживания заявок нескольких приоритетов является система с резервированием каналов пучка (СРКП), в которой для заявок i-го приоритета имеется свой порог n_i, и заявки данного приоритета принимаются к обслуживанию только в том случае, когда число занятых каналов меньше n_i. Такая схема обслуживания заявок, получившая в литературе название системы с резервированием каналов пучка (СРКП), впервые была описана P. J. Burke (фирма Bell, США) в 1961 г. для обслуживания двухприоритетного пучка и впоследствии использована в ряде систем связи, в частности, в системе RITA (франция) [1], [2]. Различные модификации СРКП исследовались в работах [1], [3], [4], [5], [6], [7], [8], [9] и других.

В [10] была предложена методика расчета СРКП с произвольным числом приоритетов, пуассоновскими входящими потоками и одинаковым экспоненциальным обслуживанием заявок всех приоритетов. Эта методика позволила получить в [11] легко реализуемые алгоритмы вычисления основных стационарных характеристик двухприоритетной СРКП с марковским входящим потоком и одинаковым экспоненциальным обслуживанием заявок всех приоритетов. Однако предположения об экспоненциальности и одинаковой распределенности времен обслуживания заявок различных приоритетов существенно ограничивают сферу применения результатов [10] и [11]. В настоящей работе предложена свободная от этих недостатков марковская модель, обобщающая двухприоритетную СРКП. Показано, как для этой модели можно вычислить стационарные распределения очередей заявок разных приоритетов.

2. ОПИСАНИЕ МОДЕЛИ

Рассмотрим систему массового обслуживания марковского типа с входящим потоком заявок двух типов (далее заявки первого типа будем называть приоритетными, а второго — неприоритетными) и относительным приоритетом, описываемую следующим образом.

Имеются неотрицательные целые числа $n_0 \geq 1$, $n_1 \geq n_0$ и $n_2 \geq 0$. Эти числа интерпретируются так: n_0 — число приборов, доступных всем заявкам (иными словами, если занято менее n_0 приборов, то любая поступающая заявка немедленно начинает обслуживаться); n_1 — максимальное суммарное число неприоритетных заявок, находящихся на приборах, и приоритетных заявок, находящихся в системе; n_2 — число мест ожидания для неприоритетных заявок, т.е. максимальное число неприоритетных заявок, которые могут находиться в системе дополнительно к заявкам на приборах.

Процесс поступления и обслуживания заявок определяется матрицами $\Lambda_n^{(u)}$, $n = \overline{0, n_1}$, $u = 1, 2$, M_n, $n = \overline{1, n_1}$, N_n, $n = \overline{0, n_1}$, и Ω (размеры матриц определяются далее и задаются числами I_n, $n = \overline{0, n_1}$, которые будем называть числами фаз процесса поступления-обслуживания заявок в слоях n, или просто числами фаз) и протекает таким образом.

Если в системе находится n, $n = \overline{0, n_0 - 1}$, заявок (далее будем говорить, что процесс поступления-обслуживания находится в слое n) и фаза равна i, $i = \overline{1, I_n}$, то с интенсивностью $(\Lambda_n^{(u)})_{ij}$, $u = 1, 2$, $j = \overline{1, I_{n+1}}$, в систему поступает новая заявка u-го типа, которая тут же начинает обслуживаться, и фаза становится равной j. Кроме того, с интенсивностью $(N_n)_{ij}$, $j = \overline{1, I_n}$, $j \neq i$, в систему не поступают заявки и не заканчивается обслуживание ни одной из n заявок, но фаза становится равной j. Наконец, если дополнительно $n \geq 1$, то с интенсивностью $(M_n)_{ij}$, $j = \overline{1, I_{n-1}}$, заканчивается обслуживание одной из n заявок, процесс поступления-обслуживания переходит в слой $n - 1$ и фаза становится равной j.

При $n = \overline{n_0 + 1, n_1 - 1}$ будем говорить, что процесс поступления-обслуживания находится в слое n в том случае, когда суммарное число неприоритетных заявок, находящихся на приборах, и приоритетных заявок, находящихся в системе, равно n и, возможно, имеется еще какое-то число (но не более n_2) неприоритетных заявок в очереди. При этом если фаза процесса поступления-обслуживания равна i, $i = \overline{1, I_n}$, то с интенсивностью $(\Lambda_n^{(1)})_{ij}$, $j = \overline{1, I_{n+1}}$, в систему поступает приоритетная заявка, причем, как и прежде, процесс поступления-обслуживания переходит в слой $n+1$ и фаза становится равной j. С интенсивностью $(N_n)_{ij}$, $j = \overline{1, I_n}$, $j \neq i$, в систему не поступают заявки и не заканчивается обслуживание заявок, но с i-й на j-ю изменяется фаза, а с интенсивностью $(M_n)_{ij}$, $j = \overline{1, I_{n-1}}$, заканчивается обслуживание одной из заявок и фаза становится равной j. Однако если в систему поступает неприоритетная заявка (с интенсивностью $(\Lambda_n^{(2)})_{ij}$, $j = \overline{1, I_n}$), то слой

n не меняется, а фаза становится равной j. Поступающая неприоритетная заявка либо становится в очередь неприоритетных заявок (если там имеются свободные места), либо теряется.

Единственным отличием случая $n = n_1$ от предыдущего является то, что поступающая приоритетная заявка теряется, изменяя фазу в том же слое. Естественно, матрица $\Lambda_{n_1}^{(1)}$ в этом случае является квадратной порядка I_{n_1}.

Обратимся к последнему случаю: $n = n_0$. Здесь возможны два варианта, связанные с окончанием обслуживания. Первый вариант появляется, если в системе в очереди отсутствуют неприоритетные заявки. Тогда, как и раньше, с интенсивностью $(M_{n_0})_{ij}$, $i = \overline{1, I_{n_0}}$, $j = \overline{1, I_{n_0-1}}$, заканчивается обслуживание одной из заявок, процесс поступления-обслуживания переходит в слой $n_0 - 1$ и фаза становится равной j. Прежний смысл имеют также матрицы $\Lambda_{n_0}^{(u)}$, $u = 1, 2$, и N_{n_0}. Если же в системе в очереди имеются неприоритетные заявки (второй вариант), то с интенсивностью $(M_{n_0})_{ij}$, $i = \overline{1, I_{n_0}}$, $j = \overline{1, I_{n_0-1}}$, заканчивается обслуживание заявки, но на освободившийся прибор мгновенно поступает заявка из очереди неприоритетных заявок, с вероятностью Ω_{jl}, $l = \overline{1, I_{n_0}}$, меняя фазу с j-й на l-ю и возвращая процесс поступления-обслуживания в слой n_0. Матрицы $\Lambda_{n_0}^{(u)}$ имеют прежний смысл.

В тех случаях, когда суммирование возможно, положим:

$$N_n^{(0)} = N_n + \Lambda_n^{(1)} + \Lambda_n^{(2)}; \quad N_n^{(1)} = N_n + \Lambda_n^{(2)}; \quad N_n^{(2)} = N_n + \Lambda_n^{(1)}.$$

3. СТАЦИОНАРНЫЕ ВЕРОЯТНОСТИ СОСТОЯНИЙ

Пусть $\nu(t)$ обозначает слой, в котором находится процесс поступления-обслуживания в момент t. Предположим, что в начальный момент значение процесса $\nu(t)$ равнялось n, $n = \overline{n_0, n_1}$, отсутствовала очередь неприоритетных заявок и фаза была i. Введем матрицы $F_k(n)$, $k = \overline{0, n_2 - 1}$, и $\tilde{F}_k(n)$, $k = \overline{0, n_2}$. Элементы $(F_k(n))_{ij}$ и $(\tilde{F}_k(n))_{ij}$, $i = \overline{1, I_n}$, $j = \overline{1, I_{n-1}}$, матриц $F_k(n)$ и $\tilde{F}_k(n)$ представляют собой вероятности того, что непосредственно после момента, когда впервые значение процесса $\nu(t)$ станет равным $n - 1$ (в случае $n = n_0$ таким моментом может быть также момент, когда на прибор впервые поступит неприоритетная заявка из очереди неприоритетных заявок, однако эта неприоритетная заявка пока еще не учитывается как поступившая на прибор), фаза будет j и в системе в очереди неприоритетных заявок будет находиться ровно k и, соответственно по крайней мере k, заявок. Следующие соотношения позволяют рекуррентно по n и k, начиная с $n = n_1$ и $k = 0$, вычислять матрицы $F_k(n)$ и $\tilde{F}_k(n)$:

$$F_0(n_1) = (-N_{n_1}^{(2)})^{-1} M_{n_1}; \tag{1}$$

$$\tilde{F}_0(n_1) = (-N_{n_1}^{(0)})^{-1} M_{n_1}; \tag{2}$$

$$F_k(n_1) = (-N_{n_1}^{(2)})^{-1} \Lambda_{n_1}^{(2)} F_{k-1}(n_1), \quad k = \overline{1, n_2 - 1}; \tag{3}$$

$$\tilde{F}_k(n_1) = (-N_{n_1}^{(2)})^{-1} \Lambda_{n_1}^{(2)} \tilde{F}_{k-1}(n_1), \quad k = \overline{1, n_2}; \tag{4}$$

$$F_0(n) = -N_n^{-1}\left[\Lambda_n^{(1)} F_0(n+1) F_0(n) + M_n\right], \quad n = \overline{n_0, n_1 - 1}; \tag{5}$$

$$\tilde{F}_0(n) = (-N_n^{(1)})^{-1}\left[\Lambda_n^{(1)} \tilde{F}_0(n+1) \tilde{F}_0(n) + M_n\right], \quad n = \overline{n_0, n_1 - 1}; \tag{6}$$

$$F_k(n) = -N_n^{-1}\left[\Lambda_n^{(1)} \sum_{i=0}^{k} F_i(n+1) F_{k-i}(n) + \right.$$

$$\left. + \Lambda_n^{(2)} F_{k-1}(n)\right], \quad k = \overline{1, n_2 - 1}, \quad n = \overline{n_0, n_1 - 1}; \tag{7}$$

$$\tilde{F}_k(n) = -N_n^{-1}\left[\Lambda_n^{(1)}\left(\sum_{i=0}^{k-1} F_i(n+1)\tilde{F}_{k-i}(n) + \tilde{F}_k(n+1)\tilde{F}_0(n)\right) + \right.$$

$$\left. + \Lambda_n^{(2)} \tilde{F}_{k-1}(n)\right], \quad k = \overline{1, n_2}, \quad n = \overline{n_0, n_1 - 1}. \tag{8}$$

Положим $F_k = F_k(n_0)$, $k = \overline{0, n_2 - 1}$, и $\tilde{F}_k = \tilde{F}_k(n_0)$, $k = \overline{0, n_2}$.

Вложенную цепь Маркова для данной системы можно ввести различными способами. Удобно определить ее таким образом. Рассмотрим следующие моменты: моменты освобождения каких-либо приборов или поступления заявок в систему в случае, когда перед этими моментами было занято менее n_0 приборов; моменты освобождения каких-либо приборов в случае, когда перед этими моментами было занято ровно n_0 приборов. Последовательность таких моментов обозначим через $\{\tau_l, \quad l \geq 1\}$. Отметим, что непосредственно после момента второго типа либо в системе остается ровно $n_0 - 1$ заявок (все они обслуживаются на приборах), либо на прибор поступает неприоритетная заявка из очереди. Множество состояний вложенной цепи Маркова представляет собой пару (i, m), $i = \overline{1, I_m}$, $m = \overline{0, n_0 + n_2 - 1}$. Здесь i — фаза (процесса поступления-обслуживания) соответствующего слоя. Более хитро определяется второй компонент m множества состояний. А именно, если $m = \overline{0, n_0 - 1}$, то, как обычно, m — число заявок в системе. Однако, если $m = \overline{n_0, n_0 + n_2 - 1}$, то число занятых приборов равно n_0, а $m - n_0$ — число неприоритетных заявок в очереди. Саму цепь Маркова образуют фаза и число занятых приборов (номер слоя) или число занятых приборов (номер слоя) плюс число неприоритетных заявок в очереди непосредственно после моментов τ_l.

Матрицу переходных вероятностей вложенной цепи Маркова будем обозначать через $P = (P_{m_1 m_2})$, $m_1, m_2 = \overline{0, n_0 + n_2 - 1}$. Отметим, что в силу принятого нами соглашения элементы $P_{m_1 m_2}$ матрицы P сами являются матрицами размеров, определяемых числами фаз соответствующих слоев. Эти матрицы задаются формулами:

$$P_{m,m-1} = -N_m^{-1} M_m, \quad m = \overline{1, n_0 - 1};$$

$$P_{m,m+1} = -N_m^{-1}(\Lambda_m^{(1)} + \Lambda_m^{(2)}), \quad m = \overline{0, n_0 - 1};$$

$$P_{m_1 m_2} = F_{m_2 - m_1 + 1}\Omega, \quad m_1 = \overline{n_0, n_0 + n_2 - 1}, \quad m_2 = \overline{m_1 - 1, n_0 + n_2 - 2}, \quad m_2 \neq n_0 - 1;$$

$$P_{n_0, n_0 - 1} = F_0;$$

$$P_{m,n_0+n_2-1} = \tilde{F}_{n_0+n_2-m}\Omega, \quad m = \overline{n_0, n_0+n_2-1}.$$

Остальные матрицы $P_{m_1 m_2}$ являются нулевыми.

Обозначим через $\vec{p}^{\,*}$ вектор-строку стационарных вероятностей вложенной цепи Маркова. Естественно, координаты $\vec{p}^{\,*}_m$, $m = \overline{0, n_0+n_2-1}$, вектора $\vec{p}^{\,*}$ сами являются вектор-строками с координатами $(\vec{p}^{\,*}_m)_i$, где индекс i означает фазу.

Вектор $\vec{p}^{\,*}$ удовлетворяет системе уравнений равновесия (СУР)

$$\vec{p}^{\,*} = \vec{p}^{\,*} P$$

с условием нормировки

$$\sum_{m=0}^{n_0+n_2-1} \vec{p}^{\,*}_m \vec{1} = 1.$$

Здесь через $\vec{1}$ обозначен вектор-столбец $(1, \ldots, 1)^T$, размерность которого определяется размером умножаемой на него слева матрицы. Для решения СУР удобно применить алгоритм, основанный на последовательном упрощении цепи Маркова с помощью исключения множеств состояний (см. [12], стр. 22).

Предположим теперь, что в начальный момент значение процесса $\nu(t)$ равнялось n, $n = \overline{n_0, n_1}$, отсутствовала очередь неприоритетных заявок и фаза была i. Обозначим через $S_k(n)$, $k = \overline{n, n_1}$, матрицу, элементом $(S_k(n))_{ij}$, $j = \overline{1, I_k}$, которой является среднее время, проведенное системой в состояниях, когда значение процесса $\nu(t)$ равнялось k и фаза была j, до того момента, когда впервые значение процесса $\nu(t)$ стало равным $n-1$ (в случае $n = n_0$ таким моментом может быть также момент, когда на прибор впервые поступит неприоритетная заявка из очереди неприоритетных заявок), а через $\vec{t}_k(n)$, $k = \overline{0, n_2-1}$, и $\vec{T}_k(n)$, $k = \overline{0, n_2}$, — вектор-столбцы, координатами $(\vec{t}_k(n))_i$ и $(\vec{T}_k(n))_i$, $i = \overline{1, I_n}$, которых являются средние времена, проведенные системой в состояниях, когда в очереди неприоритетных заявок имелось ровно k и, по крайней мере k, заявок (без учета фазы), до того момента, когда впервые значение процесса $\nu(t)$ стало равным $n-1$ (с прежним замечанием относительно случая $n = n_0$). Матрицы $S_k(n)$ и векторы $\vec{t}_k(n)$ и $\vec{T}_k(n)$ удовлетворяют уравнениям, аналогичным (1)–(3), и также легко вычисляются рекуррентно.

Наконец, обозначим через $\vec{p}_n$, $n = \overline{0, n_1}$, — вектор-строку с координатами $(\vec{p}_n)_i$, $i = \overline{1, I_n}$, представляющими собой стационарные вероятности того, что значение процесса $\nu(t)$ равно n и фаза равна i, а через $p^{(2)}_n$, $n = \overline{0, n_2}$, — стационарную вероятность того, что в системе в очереди неприоритетных заявок имеется n заявок (фаза не учитывается). Тогда:

$$\vec{p}_n = -\frac{1}{m}\vec{p}^{\,*}_n N_n^{-1}, \quad n = \overline{0, n_0-1}; \qquad \vec{p}_n = \frac{1}{m}\sum_{k=n_0}^{n_0+n_2-1} \vec{p}^{\,*}_k S_n, \quad n = \overline{n_0, n_1};$$

$$p^{(2)}_0 = \frac{1}{m}\left(-\sum_{k=0}^{n_0-1} \vec{p}^{\,*}_k N_k^{-1}\vec{1} + \vec{p}^{\,*}_{n_0}\vec{t}_0 \right); \qquad p^{(2)}_n = \frac{1}{m}\sum_{k=0}^{n} \vec{p}^{\,*}_{n_0+k}\vec{t}_{n-k}, \quad n = \overline{1, n_2-1};$$

$$p_{n_2}^{(2)} = \frac{1}{\overline{m}} \sum_{k=0}^{n_2-1} \vec{p}^{\,*}_{n_0+k} \vec{T}_{n_2-k},$$

где:

$$\overline{m} = \sum_{n=0}^{n_0+n_2-1} \vec{p}^{\,*}_n \vec{m}_n;$$

$$\vec{m}_n = -N_n^{-1}\vec{1}, \quad n = \overline{0, n_0-1}; \qquad \vec{m}_n = \vec{T}_0, \quad n = \overline{n_0, n_0+n_2-1}.$$

ЛИТЕРАТУРА

1. *Grandjean C. H.* Traffic Calculations in Suturation Routing with Priorities // Electr. Commun. 1974. V. 49. № 1, P. 72–79.
2. *Людвиг Г., Рой Р.* Ограничения для сетей с волновым поиском сетей // Тр. ин-та инженеров электроники и радиоэлектроники. 1977. Т. 65. № 9. С. 154–165.
3. *Weber J. H.* Some Traffic Characteristics of Communications Networks with Automatic Alternate Routing // Bell System Techn. J. 1962. March. P. 1201–1247.
4. *Weber J. H.* Simulation Study of Routing and Control in Communications Networks // Bell System Techn. J. 1964. Nov. P. 2639–2676.
5. *Grandjean C. H.* Call Routing Strategies in Telecommunications Networks // Electr. Commun. 1967. V. 42, № 3. P. 380–391.
6. *Джейсуол Н.* Очереди с приоритетами. М.: Мир, 1973.
7. *Esoqbue A. O., Singh A. J.* A Stochastic Model for a Optimal Priority Bed Distribution in a Hospital // Oper. Res. 1976. № 24. P. 884–889.
8. *Otterman J.* Grande of Service Direct Traffic mixed with Store-and-Forward Traffic // Bell System Techn. J. 1962. Apr. P. 1415–1437.
9. *Liu F. K.* A Combined Delay and Loss System with Priority // ICC. 1973. V. 39. № 7. P. 39-7–39-13.
10. *Печинкин А. В., Федоров В. М.* Методика расчета многоканальной системы приоритетного обслуживания с резервированием каналов // Системное моделирование. Вып. 15. Новосибирск: ВЦ СО АН СССР, 1990.
11. *Бурыгин С. В., Глазунов А. С., Печинкин А. В.* Система приоритетного обслуживания с резервированием каналов и марковским входящим потоком // Вестник Российского университета дружбы народов. Сер. Прикладная математика и информатика. 2001. № 1. С. 80–89.
12. *Bocharov P. P., D'Apice C., Pechinkin A. V., Salerno S.* Queueing Theory. Utrecht–Boston: VSP, 2004.