О стационарном распределении процесса обслуживания неординарных потоков с разделением времени при пороговом алгоритме переключения

А.В. Зорин

1 Введение

Приоритетные системы обслуживания с динамическими приоритетами (т.е., приоритетными индексами очередей, зависящим от их длин в момент принятия решения) и повторными требованиями, рассматривались в работах [1, 2, 3]. Они являются адекватными моделями для процессов обработки информации в вычислительных комплексах. Основной результат этих исследование состоит в следующем: алгоритм назначения приоритетных индексов, для которого будет минимальным математическое ожидание стоимости пребывания в системе всех требований (за единицу времени или за один рабочий акт обслуживающего устройства), является классическое правило постоянных приоритетных индексов, которые назначаются заранее по данным о средних длительностях обслуживания и о стоимостях пребывания индивидуальных требований за единицу времени.

В то же время, представляют интерес и другие задачи оптимизации. В работе [4] решалась задача минимизации среднего времени достижения процессом заданного множества состояний. При специальном виде областей разрешенной, финальной и запрещенной, наилучшие результаты (по сравнению с приоритетным алгоритмом и алгоритмом обслуживания самой длинной очереди) показывал «пороговый» алгоритм. В связи с этим, естественна задача анализа процесса обслуживания в классе пороговых алгоримов. В настоящей работе будет решаться задача определения (вычисления численными методами) стационарного распределения в терминах производящих функций, что в дальнейшем позволяет находить теоретические числовые характеристики процесса обслуживания и длин очередей численными методами теории функций комплексного переменного.

2 Постановка задачи

В систему поступают два неординарных пуассоновских потока Π_1, Π_2. Интенсивность поступления групп требований по потоку Π_j, $j = 1, 2$, равна $\lambda_j > 0$, а вероятность прихода группы размера n равна $f(b, j) \geqslant 0$, $b = 1, 2, \ldots$; $\sum\limits_{b=1}^{\infty} f(b, j) = 1$. Требования потока Π_j поступают в накопитель O_j неограниченной вместимости. Обслуживание требования из очереди O_j имеет экспоненциальное распределение с параметром β_j. Обслуженное требование из очереди O_j с вероятностью $p_{j,r}$ поступает на повторное обслуживание в очередь O_r, $r = 1, 2$, а с вероятностью $p_{j,0} = 1-p_{j,1}-p_{j,2} \geqslant 0$ покидает систему. После каждого акта обслуживания требования из очереди O_j обслуживающее устройство производит операцию внутренней переналадки, длительность которой распределена по показательному распределению с параметром $\bar{\beta}_j$. Если в момент окончания переналадки длины очередей описываются ненулевым вектором (x_1, x_2), то на обслуживание выбирается требование из очереди $s = h(x_1, x_2)$, где $h(\cdot, \cdot)$ есть заданное отображение неотрицательной целочисленной решетки $X = \{0,1,\ldots\} \times \{0,1,\ldots\}$ на множество $\{0,1,2\}$, причем равенство $s = h(x_1, x_2)$ влечет $x_s > 0$, а прообразом точки 0 является только нулевой вектор $\bar{0} = (0,0)$ из решетки X. Если после окончания переналадки очереди пусты, обслуживающее устройство переходит в режим ожидания поступления новых требований. При поступлении первой группы требований в пустую систему мгновенно начинается обслуживание одного и требований в группе, остальные занимают места в очереди, соответствующей потоку.

Пусть $\kappa_j(t)$ — число требований в очереди O_j в момент $t \geqslant 0$, $\kappa(t) = (\kappa_1(t), \kappa_2(t))$. Введем множество $\Gamma = \{\Gamma^{(0)}, \Gamma^{(1)}, \ldots, \Gamma^{(4)}\}$ состояний обслуживающего устройства; здесь $\Gamma^{(0)}$ есть состояние ожидания прихода нового требования, в состоянии $\Gamma^{(j)}$ при $j = 1$, 2 происходит обслуживание требования их очереди O_j, а при $j = 3$, 4 осуществляется акт переналадки после обслуживания требования из очереди O_{j-2}. Случайный элемент $\Gamma(t) \in \Gamma$ задает состояние обслуживающего устройства в момент $t \geqslant 0$.

Процесс $\{(\Gamma(t), \kappa(t)); t \geqslant 0\}$ является однородным марковским. Его пространство состояний можно взять в виде $\{(\Gamma^{(0)}, 0, 0)\} \cup \{\Gamma^{(1)}, \Gamma^{(2)}, \Gamma^{(3)}, \Gamma^{(4)}\} \times X \times X$. Обозначим

$$Q(r, x_1, x_2; t) = \mathbf{P}(\{\Gamma(t) = \Gamma^{(r)}, \kappa(t) = (x_1, x_2)\}); \qquad f_j(z) = \sum_{b=1}^{\infty} z^b f(b, j), \qquad |z| \leqslant 1;$$

$$\Psi(z_1, z_2, r; t) = \mathbf{E}\big(z_1^{\kappa_1(t)} z_2^{\kappa_2(t)} I(\Gamma(t) = \Gamma^{(r)})\big) = \sum_{x_1=0}^{\infty} \sum_{x_2=0}^{\infty} z_1^{x_1} z_2^{x_2} Q(r, x_1, x_2; t), \quad |z_1| < 1.|z_2| < 1.$$

Теорема 1. *Для $r = 1$, 2 имеют место уравнения*

$$\frac{\partial}{\partial t}\Psi(z_1, z_2, r; t) = \Psi(z_1, z_2, r; t)(\lambda_1(f_1(z_1) - 1) + \lambda_2(f_2(z_2) - 1) - \beta_r) +$$
$$+ \sum_{j=1}^{2} \bar{\beta}_j \mathbf{E}\big(z_1^{\kappa_1(t)} z_2^{\kappa_2(t)} I(\{\Gamma(t) = \Gamma^{(2+j)}, h(\kappa_1(t), \kappa_2(t)) = r\})\big) + \lambda_r f_r(z_r) Q(0, 0, 0; t), \tag{1}$$

$$\frac{\partial}{\partial t}\Psi(z_1, z_2, 2 + r; t) = \Psi(z_1, z_2, 2 + r; t)(\lambda_1(f_1(z_1) - 1) + \lambda_2(f_2(z_2) - 1) - \bar{\beta}_r) +$$
$$+ \beta_r z_r^{-1}(1 + p_{r,1}(z_1 - 1) + p_{r,2}(z_2 - 1))\Psi(z_1, z_2, r; t), \tag{2}$$

$$\frac{d}{dt} Q(0, 0, 0, t) = -(\lambda_1 + \lambda_2) Q(0, 0, 0; t) + \bar{\beta}_1 Q(3, 0, 0, t) + \bar{\beta}_2 Q(4, 0, 0, t). \tag{3}$$

3 О решении стационарных уравнений

В дальнейшем нас будут интересовать стационарные вероятности, выбрав которые в качестве начальных получим

$$Q(r, x_1, x_2) = \lim_{t \to \infty} Q(r, x_1, x_2; t), \qquad \Psi(z_1, z_2, r) = \lim_{t \to \infty} \Psi(z_1, z_2, r; t).$$

Обозначим через $\mu_j = f_j'(1)$ средний размер группы по потоку Π_j. Введем векторы и матрицу

$$\beta = (\beta_1^{-1}, \beta_2^{-1}), \quad \bar{\beta} = (\bar{\beta}_1^{-1}, \bar{\beta}_2^{-1}), \quad \bar{\lambda} = \begin{pmatrix} \lambda_1 \mu_1 \\ \lambda_2 \mu_2 \end{pmatrix}, \quad \mathbf{Q} = \begin{pmatrix} p_{1,1} & p_{1,2} \\ p_{2,1} & p_{2,2} \end{pmatrix}, \quad \mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Пусть матрица $(\mathbf{I} - \mathbf{Q})$ обратима.

Теорема 2. *Имеют место соотношения*

$$Q(0, 0, 0) = 1 - (\beta + \bar{\beta})(\mathbf{I} - \mathbf{Q}^T)^{-1} \bar{\lambda},$$

$$\Psi(1,1,1) = \frac{\beta_2\beta(\mathbf{I}-\mathbf{Q}^T)^{-1}\bar\lambda - (1,1)(\mathbf{I}-\mathbf{Q}^T)^{-1}\bar\lambda}{\beta_2-\beta_1},$$
$$\Psi(1,1,2) = \frac{(1,1)(\mathbf{I}-\mathbf{Q}^T)^{-1}\bar\lambda - \beta_1\beta(\mathbf{I}-\mathbf{Q}^T)^{-1}\bar\lambda}{\beta_2-\beta_1},$$
$$\Psi(1,1,3) = \frac{\bar\beta_2\bar\beta(\mathbf{I}-\mathbf{Q}^T)^{-1}\bar\lambda - (1,1)(\mathbf{I}-\mathbf{Q}^T)^{-1}\bar\lambda}{\bar\beta_2-\bar\beta_1},$$
$$\Psi(1,1,4) = \frac{(1,1)(\mathbf{I}-\mathbf{Q}^T)^{-1}\bar\lambda - \bar\beta_1\bar\beta(\mathbf{I}-\mathbf{Q}^T)^{-1}\bar\lambda}{\bar\beta_2-\bar\beta_1}.$$

Эти формулы не зависят от функции переключения $h(\cdot)$.

Пользуясь методом из работы [4], можно доказать, что условие $(\beta+\bar\beta)(\mathbf{I}-\mathbf{Q}^T)^{-1}\bar\lambda < 1$ является необходимым и достаточным для существования стационарного распределения.

Рассмотрим алгоритм порогового типа: пусть $L \geqslant 0$ — целое

$$h(x_1,x_2) = \begin{cases} 1, & \text{если } x_1 > L \text{ или } x_2 = 0, x_1 > 0 \\ 2, & \text{если } x_1 \leqslant L, x_2 > 0 \\ 0, & \text{если } x_1 = x_2 = 0 \end{cases}$$

В этом случае, имеют место соотношения (в стационарном случае):

$$\mathbf{E}\big(z_1^{\kappa_1(t)} z_2^{\kappa_2(t)} I(\{\Gamma(t) = \Gamma^{(2+j)}, h(\kappa_1(t),\kappa_2(t)) = 1\})\big) = \Psi(z_1,z_2,2+j) - \Psi(0,z_2,2+j) -$$
$$- \sum_{k=1}^{L} \frac{z_1^k}{k!}\frac{\partial^k}{\partial z_1^k}(\Psi(z_1,z_2,2+j) - \Psi(z_1,0,2+j))\Big|_{z_1=0},$$
$$\mathbf{E}\big(z_1^{\kappa_1(t)} z_2^{\kappa_2(t)} I(\{\Gamma(t) = \Gamma^{(2+j)}, \kappa(t) \in X_2\})\big) = \Psi(0,z_2,2+j) - \Psi(0,0,2+j) +$$
$$+ \sum_{k=1}^{L} \frac{z_1^k}{k!}\frac{\partial^k}{\partial z_1^k}(\Psi(z_1,z_2,2+j) - \Psi(z_1,0,2+j))\Big|_{z_1=0},$$
$$\Psi(0,0,2+j) = Q(2+j,0,0).$$

Решение стационарных уравнений в данном случае ведет к громоздким вычислениями, в следствие чего оказывается затруднительно аналитически выяснять вопросы о разрешимости некоторых вспомогательных уравнений, о числе их решений. В докладе обсуждается возможность использования систем символьных вычислений и компьютерной алгебры для решения указанной задачи.

Список литературы

[1] Климов Г. П. Системы обслуживания с разделением времени. I // Теория вероятностей и ее применения. — 1974. — Т. 19, Вып. 3. — С. 558–576.

[2] Китаев М.Ю., Рыков В.В. Системы обслуживания с ветвящимися потоками вторичных требований // Автоматика и телемеханика. — 1980. — № 9. — С. 52–61.

[3] Федоткин М.А. Оптимальное управление конфликтными потоками и маркированные точечные процессы с выделенной дискретной компонентой, I // Литовский математический сборник. — 1988. — Т. 28, №. 4. — С. 784–794.

[4] Zorine A.V. On ergodicity conditions in a polling model with Markov modulated input and state-dependent routing / A.V. Zorine // Queueing systems. — 2014. — V. 76. — № 2. — P. 223–241.