УДК 519.217

Є.О. Лебєдєв, В.Д. Пономарьов

Про оптимальне керування інтенсивністю обслуговування в системах з повторними викликами

(Представлено членом-кореспондентом НАН України В. В. Анісімовим)

Розглядається марківська модель системи з повторними викликами, в якій інтенсивність обслуговування залежить від довжини черги. Метод дослідження спирається на апроксимацію вихідної системи системою з урізаним простором станів. Як застосування розглянуто задачу пошуку оптимального керування інтенсивністю обслуговування в класі порогових стратегій.

При досліджені широкого класу стохастичних систем з повторними викликами виникає необхідність розрахунку характеристик систем з необмеженим фазовим простором. Марківський процес, який описує функціонування такої системи, має зліченну множину станів, а матриця переходів між станами, як правило, не має спеціальних властивостей, які б полегшили її перетворення для отримання явного розв'язку. Через ці обставини вдалося явно дослідити тільки декілька моделей (див., наприклад, [1]).

На сьогоднішній день для розрахунку ймовірнісних характеристик систем з повторними викликами застосовують або чисельні методи розв'язання системи рівнянь, або рекурентні алгоритми. Для їх побудови від систем із зліченним числом станів необхідно перейти до систем із скінченним фазовим простором. З цією метою, зазвичай, вважають, що черга повторних викликів не перевищує деякого заданого рівня M. Якщо вимога надходить до системи, коли всі прилади зайняті та вже є M повторних викликів, вона губиться. Інтуїтивно зрозуміло, що обравши M досить великим, можна оцінити ймовірнісні характеристики вихідної моделі з будь-якою наперед заданою точністю, використовуючи відповідні характеристики урізаної моделі.

У даній роботі розвивається саме такий підхід для системи з повторними викликами, пуассонівським вхідним потоком і керованою інтенсивністю обслуговування вимог. Її характеристики наближаються характеристиками урізаної моделі, які явно виписуються через параметри системи. Також досліджується задача вибору рівня M, за якого досягається задана точність обчислень.

Марківські моделі для вихідної та урізаної систем. Розглянемо ланцюг Маркова з неперервним часом $X(t) = (C(t); N(t))$, $C(t) \in \{0, 1, \ldots, c\}$, $N(t) \in \{0, 1, \ldots\}$, який задається інфінітезимальними характеристиками $a_{(i,j)(i',j')}$, $(i, j), (i', j') \in S(X) = \{0, 1, \ldots, c\} \times \times \{0, 1, \ldots\}$:

1) якщо $i = \{0, 1, \ldots, c-1\}$, то

$$a_{(i,j)(i',j')} = \begin{cases} \lambda, & (i',j') = (i+1,j); \\ j\mu, & (i',j') = (i+1,j-1); \\ i\nu_j, & (i',j') = (i-1,j); \\ -[\lambda + j\mu + i\nu_j], & (i',j') = (i,j); \\ 0 & \text{— в іншому випадку;} \end{cases}$$

2) якщо $i = c$, то

$$a_{(c,j)(i',j')} = \begin{cases} \lambda, & (i',j') = (c, j+1); \\ c\nu_j, & (i',j') = (c-1, j); \\ -[\lambda + c\nu_j], & (i',j') = (c, j); \\ 0 & \text{— в іншому випадку.} \end{cases}$$

Ланцюг Маркова $X(t)$ описує процес обслуговування в наступній системі з повторними викликами. Вхідний потік вимог є пуассонівським з параметром λ. Вимоги обслуговуються на c однакових приладах. Якщо є хоча б один вільний прилад, то вимога обслуговується негайно. Час обслуговування — показниково розподілена випадкова величина з параметром ν_j, який залежить від довжини черги j повторних викликів у поточний момент часу. Якщо всі прилади зайняті, то вимога повторно намагається отримати обслуговування через випадковий час, який має показниковий розподіл з параметром μ. Кількість зайнятих приладів у будь-який момент часу задається першою компонентою процесу $X(t)$, а кількість джерел повторних викликів — другою.

З'ясуємо умови існування стаціонарного режиму для процесу $X(t)$, $t > 0$.

Лема 1. *Нехай $\nu = \varliminf_{j\to\infty} \nu_j$. Тоді при $\lambda/(c\nu) < 1$ ланцюг Маркова $X(t)$ — ергодичний і його граничний розподіл π_{ij}, $(i,j) \in S(X)$ збігається з единим стаціонарним.*

Доведення. Розглянемо як тест-функції Ляпунова

$$\varphi(i,j) = \alpha i + j, \qquad (i,j) \in S(X),$$

де параметр α буде визначено пізніше. Для обраних тест-функцій середній перенос

$$y_{ij} = \sum_{(i',j')\neq(i,j)} a_{(i,j)(i',j')}(\varphi(i',j') - \varphi(i,j))$$

дорівнює

$$y_{ij} = \begin{cases} \lambda\alpha - i\nu_j\alpha + j\mu(\alpha - 1), & 0 \leqslant i \leqslant c-1, \\ \lambda - c\nu_j\alpha, & i = c. \end{cases}$$

При $\lambda/(c\nu) < 1$ для будь-якого $\alpha \in (\lambda/(c\nu), 1)$ існує таке $\varepsilon > 0$, що $y_{ij} < -\varepsilon$ для всіх $(i,j) \in S(X)$ за винятком скінченного числа станів (i,j). Таким чином, для тест-функцій $\varphi(i,j) = \alpha i + j$, $\alpha \in (\lambda/(c\nu), 1)$ виконуються умови теореми Твіді [2, с. 97]. Лему доведено.

Розглянемо урізану систему з повторними викликами. Вона функціонує аналогічним чином, але має обмеження на максимальну кількість M джерел повторних викликів. Тобто, нові вимоги на обслуговування губляться назавжди, коли в системі зайняті всі прилади і існує M джерел повторних викликів. Формально функціонування цієї системи описується ланцюгом Маркова $X(t,M) = (C(t,M); N(t,M))$, де $C(t,M) \in \{0,1,\ldots,c\}$, $N(t,M) \in \{0,1,\ldots,M\}$, який має інфінітезимальні характеристики $a^{(M)}_{(i,j)(i',j')}$, (i,j), $(i',j') \in$ $\in S(X,M) = \{0,1,\ldots,c\} \times \{0,1,\ldots,M\}$:

1) якщо $i = \{0, 1, \ldots, c-1\}$, $j = \{0, 1, \ldots, M\}$, то

$$a^{(M)}_{(i,j)(i',j')} = \begin{cases} \lambda, & (i',j') = (i+1, j); \\ j\mu, & (i',j') = (i+1, j-1); \\ i\nu_j, & (i',j') = (i-1, j); \\ -[\lambda + j\mu + i\nu_j], & (i',j') = (i, j); \\ 0 & \text{— в іншому випадку;} \end{cases}$$

2) якщо $i = c$, $j = \{0, 1, \ldots, M-1\}$, то

$$a^{(M)}_{(c,j)(i',j')} = \begin{cases} \lambda, & (i',j') = (c, j+1); \\ c\nu_j, & (i',j') = (c-1, j); \\ -[\lambda + c\nu_j], & (i',j') = (c, j); \\ 0 & \text{— в іншому випадку;} \end{cases}$$

3) якщо $i = c$, $j = M$, то

$$a^{(M)}_{(c,M)(i',j')} = \begin{cases} c\nu_M, & (i',j') = (c-1, M); \\ -c\nu_M, & (i',j') = (c, M); \\ 0 & \text{— в іншому випадку.} \end{cases}$$

Оскільки фазовий простір $S(X, M)$ процесу $X(t, M)$ скінченний, то для $X(t, M)$ існує стаціонарний режим і через $\pi_{ij}(M)$, $(i, j) \in S(X, M)$ будемо позначати його стаціонарні ймовірності.

Розглянемо детальніше процес обслуговування в скінченній моделі.

Стаціонарні ймовірності для скінченної системи. Для такої системи стаціонарні ймовірності задовольняють рівняння Колмогорова:

$$\begin{aligned} &[\lambda + j\mu + i\nu_j]\pi_{ij}(M) = (j+1)\mu\pi_{i-1j+1}(M) + \lambda\pi_{i-1j}(M) + (i+1)\nu_j\pi_{i+1j}(M), \\ &j = 0, \ldots, M-1, \quad i = 0, \ldots, c-1, \end{aligned} \tag{1}$$

$$[\lambda + M\mu + i\nu_M]\pi_{iM}(M) = \lambda\pi_{i-1M}(M) + (i+1)\nu_M\pi_{i+1M}(M), \quad i = 0, \ldots, c-1, \tag{2}$$

$$\begin{aligned} &[\lambda + c\nu_j]\pi_{cj}(M) = (j+1)\mu\pi_{c-1j+1}(M) + \lambda\pi_{c-1j}(M) + \lambda\pi_{cj-1}(M), \quad j = 0, \ldots, M-1, \\ &c\nu_M\pi_{cM}(M) = \lambda\pi_{c-1M}(M) + \lambda\pi_{cM-1}(M), \end{aligned} \tag{3}$$

$$\sum_{i=0}^{c}\sum_{j=0}^{M}\pi_{ij}(M) = 1.$$

Введемо такі позначення:

$$e_i(n) = (\delta_{i0}\delta_{i1}\ldots\delta_{in-1})^T, \qquad \delta_{ij} = \begin{cases} 1, & i = j; \\ 0, & i \neq j; \end{cases}$$

$1(c)$ — вектор розмірності c, складений з 1,

$$A_j = \|a^j_{ik}\|^{c-1}_{i,k=0}, \qquad a^j_{ik} = \begin{cases} -\lambda, & k = i-1, \\ \lambda + j\mu + i\nu_j, & k = i, \\ -(i+1)\nu_j, & k = i+1, \\ 0, & \text{в іншому випадку}; \end{cases}$$

якщо $i \neq 0,\ c-1$. При $i = 0$

$$a^j_{0k} = \begin{cases} \lambda + j\mu, & k = 0, \\ -\nu_j, & k = 1, \\ 0 & \text{— в іншому випадку}; \end{cases}$$

а при $i = c-1$

$$a^j_{c-1k} = \begin{cases} -\lambda, & k = c-2, \\ \lambda + j\mu + (c-1)\nu_j, & k = c-1, \\ 0 & \text{— в іншому випадку}; \end{cases}$$

$$B_j = \|b^j_{ik}\|^{c-1}_{i,k=0}, \qquad b^j_{ik} = \begin{cases} (j+1)\mu, & k = i-1, \\ 0 & \text{— в іншому випадку}, \end{cases}$$

якщо $i \neq 0,\ c-1$. При $i = 0$, $b^j_{0k} = 0$, $k = 0, 1, \ldots, c-1$, а при $i = c-1$

$$b^j_{c-1k} = \begin{cases} \dfrac{c(j+1)\mu\nu_j}{\lambda}, & k \neq c-2, \\ \dfrac{(j+1)\mu[\lambda + c\nu_j]}{\lambda}, & k = c-2; \end{cases}$$

$$C = \|c_{ik}\|^{c-1}_{i,k=0}, \qquad c_{ik} = \begin{cases} 1, & k = 0,\ i = 0, \\ a^M_{i-1k} & \text{— в іншому випадку}; \end{cases}$$

$$\Phi_j = \left(\prod_{i=j}^{M-1} A_i^{-1} B_i\right) C^{-1} e_0(c).$$

Лема 2. *Матриці A_j, $j = 0, 1, \ldots, M$ невироджені.*

Доведення. Перевіримо умову Адамара для стовпчиків матриць A_j, $j = 0, \ldots, M$ [3, с. 406]:

$$\begin{gathered} G^j_i \equiv |a^j_{ii}| - \sum_{\substack{k=0 \\ k \neq i}}^{c-1} |a^j_{ki}| > 0, \qquad i = 0, \ldots, c-1, \\ G^j_i = \lambda + j\mu + i\nu_j - i\nu_j - \lambda = j\mu, \qquad i = 0, \ldots, c-2, \\ G^j_{c-1} = \lambda + j\mu + (c-1)\nu_j - (c-1)\nu_j = \lambda + j\mu. \end{gathered} \tag{4}$$

Умова (4) виконуються для всіх $i = 0, \dots, c-1$, $j = 1, \dots, M$, що означає невиродженість матриць A_j, $j = 1, \dots, M$. Для A_0 виконується послаблена умова Адамара. Зважаючи на те, що A_0 є нерозкладною матрицею, маємо: A_0 теж невироджена.

Імовірності $\pi_{ij}(M)$, $(i,j) \in S(X, M)$ можна подати через параметри системи в явному вигляді.

Теорема 1. *Стаціонарні ймовірності для скінченної системи мають такий вигляд:*

$$\pi_j(M) = \Phi_j \pi_{0M}(M), \qquad j = 0, \dots, M,$$

$$\pi_{cj}(M) = \frac{(j+1)\mu}{\lambda} 1(c)^T \Phi_{j+1} \pi_{0M}(M), \qquad j = 0, \dots, M-1,$$

$$\pi_{cM}(M) = \frac{\lambda e_{c-1}^T(c) + M\mu 1(c)^T}{c\nu_M} C^{-1} e_0(c) \pi_{0M}(M),$$

де

$$\pi_j(M) = (\pi_{0j}(M)\pi_{1j}(M)\dots\pi_{c-1j}(M))^T,$$

$$\pi_{0M}(M) = \left\{ \sum_{j=0}^{M} \left(1 + \frac{j\mu}{\lambda}\right) 1(c)^T \Phi_j + \frac{\lambda e_{c-1}^T(c) + M\mu 1(c)^T}{c\nu_M} \Phi_M \right\}^{-1}.$$

Доведення. Для пошуку $\pi_{ij}(M)$ використаємо теорему про рівність потоку ймовірностей через границю замкненої області в стаціонарному режимі [2, с. 49]. Для кожного $j = 1, 2, \dots, M$ побудуємо розбиття фазового простору $S(X, M) = S_j^{(1)}(X, M) \bigcup \overline{S}_j^{(1)}(X, M)$, $S_j^{(1)}(X, M) = \{(p, q) \in S(X, M) : q \leqslant j\}$. Прирівнюючи потоки ймовірностей через границю області $S_j^{(1)}(X, M)$, знаходимо

$$\lambda \pi_{cj-1}(M) = j\mu \sum_{i=0}^{c-1} \pi_{ij}(M), \qquad j = 1, \dots, M. \tag{5}$$

Подамо рівняння (1) у вигляді

$$-\lambda \pi_{i-1j}(M) + [\lambda + j\mu + i\nu_j]\pi_{ij}(M) - (i+1)\nu_j \pi_{i+1j}(M) = (j+1)\mu \pi_{i-1j+1}(M), \tag{6}$$

$j = 0, \dots, M-1$, $i = 0, \dots, c-2$.

Знайдемо з рівняння (5) імовірність $\pi_{cj}(M)$ і підставимо у (1) при $i = c-1$. Маємо:

$$-\lambda \pi_{c-2j}(M) + [\lambda + j\mu + (c-1)\nu_j]\pi_{c-1j}(M) = \frac{(j+1)c\mu\nu_j}{\lambda} \sum_{\substack{i=0 \\ i \neq c-2}}^{c-1} \pi_{ij+1}(M) +$$

$$+ \frac{(j+1)\mu[\lambda + c\nu_j]}{\lambda} \pi_{c-2j+1}(M), \qquad j = 0, \dots, M-1. \tag{7}$$

Систему рівнянь (6), (7) запишемо у векторно-матричній формі:

$$A_j \pi_j(M) = B_j \pi_{j+1}(M), \qquad j = 0, \dots, M-1.$$

З останнього рівняння знаходимо, що

$$\pi_j(M) = \left(\prod_{i=j}^{M-1} A_i^{-1} B_i\right) \pi_M(M), \qquad j = 0, \ldots, M-1. \tag{8}$$

Доповнимо систему (2) при $i = 0, 1, \ldots, c-2$ рівністю $\pi_{0M}(M) = \pi_{0M}(M)$ і запишемо у векторно-матричній формі:

$$C\pi_M(M) = e_0(c)\pi_{0M}(M).$$

Звідси знаходимо, що

$$\pi_M(M) = C^{-1} e_0(c) \pi_{0M}(M). \tag{9}$$

Остаточно, з рівнянь (8), (9) маємо:

$$\pi_j(M) = \left(\prod_{i=j}^{M-1} A_i^{-1} B_i\right) C^{-1} e_0(c) \pi_{0M}(M), \qquad j = 0, \ldots, M.$$

Підставляючи вираз для $\pi_{j+1}(M)$ в рівняння (5), знаходимо $\pi_{cj}(M)$, $j = 0, \ldots, M-1$. Імовірність $\pi_{cM}(M)$ отримуємо з рівняння (3), а ймовірність $\pi_{0M}(M)$ — з умови нормування. Теорему доведено.

Подібні результати були отримані в [4] для систем з повторними викликами і обмеженим числом джерел первинних вимог.

Наступний крок зробимо у напрямку того, щоб показати, що система з обмеженням на число повторних викликів наближає вихідну систему.

Обгрунтування апроксимації і застосування до розв'язання оптимізаційних задач. Для строгого доведення того, що характеристики скінченної системи наближають відповідні характеристики вихідної системи, використаємо поняття стохастичної впорядкованості (див. [5]).

Лема 3. *Нехай виконуються умови леми 1. Тоді:*

1) якщо $X(0, M) \leqslant {}_{st}X(0)$, то $X(t, M) \leqslant {}_{\mathrm{st}}X(t)$ для всіх $t \geqslant 0$ і $X(M) \leqslant {}_{\mathrm{st}}X$, де $X = (C, N)$, $X(M) = (C(M), N(M))$ — випадкові вектори, розподіл яких збігається з π_{ij}, $(i, j) \in S(X)$ та $\pi_{ij}(M)$, $(i, j) \in S(X, M)$ відповідно;

2) якщо $X(0, M) \leqslant {}_{\mathrm{st}}X(0, M+1)$, то $X(t, M) \leqslant {}_{\mathrm{st}}X(t, M+1)$ для всіх $t \geqslant 0$ і $X(M) \leqslant$ $\leqslant {}_{\mathrm{st}}X(M+1)$.

Лема 3 випливає з результатів про стохастичну впорядкованість процесів міграції [5, с. 111–116]. В свою чергу безпосереднім наслідком цієї леми є наступний результат.

Теорема 2. *Нехай виконуються умови леми 1. Тоді для будь-яких $(i, j) \in S(X)$ $\pi_{ij} =$ $= \lim\limits_{M\to\infty} \pi_{ij}(M)$.*

Таким чином, ми показали, що стаціонарні ймовірності $\pi_{ij}(M)$ із ростом M прямують до ймовірностей π_{ij}. Внаслідок монотонності при фіксації точності апроксимації можна обчислювати рівень M, який забезпечить задану точність. При додаткових обмеженнях на характер залежності інтенсивності обслуговування від черги повторних викликів для M можна отримувати верхні оцінки.

Розглянемо тепер задачу оптимізації прибутку в системі з повторними викликами і пуассонівським вхідним потоком. Нехай керування інтенсивністю обслуговування здійснюється на основі порогової стратегії. В цьому випадку, якщо кількість джерел повторних викликів в системі не перевищує H, то система функціонує в першому режимі і інтенсивність обслуговування дорівнює $\nu^{(1)}$. Якщо кількість джерел повторних викликів стає більшою за H, то система переходить у другий режим з інтенсивністю обслуговування $\nu^{(2)}$. Поріг H набуває значення $-1, 0, 1, \ldots$. При $H = -1$ система весь час функціонує у другому режимі, $\nu = \nu^{(2)}$.

Формально керування $X(t)$ на основі порогової стратегії означає, що інтенсивність обслуговування має такий вигляд:

$$\nu_j = \begin{cases} \nu^{(1)}, & j = 0, 1, \ldots, H; \\ \nu^{(2)}, & j = H+1, H+2, \ldots. \end{cases}$$

При такій стратегії керування задача оптимізації прибутку формулюється наступним чином. Нехай $f_i(t, H)$ — число вимог, обслуговування яких завершено до моменту t при роботі системи у i-му режимі, $i = 1, 2$; $f_3(t, H)$ — число вимог, які отримали відмову в обслуговуванні і стали повторними викликами; $f_4(t, H)$ — число перемикань інтенсивності обслуговування. При виконанні умов леми 1 границі $\lim\limits_{t\to\infty} t^{-1} f_i(t, H)$ існують. Будемо позначати їх через $f_i(H)$, $i = 1, \ldots, 4$.

Розглянемо оптимізаційну задачу:

$$F(H) = C_1 f_1(H) + C_2 f_2(H) - C_3 f_3(H) - C_4 f_4(H) \to \max,$$
$$H = \{-1, 0, 1, \ldots\},$$

де C_i, $i = 1, 2$ — прибуток, пов'язаний з обслуговуванням одного виклику при роботі системи в i-му режимі; C_3 — штраф за відмову в обслуговуванні; C_4 — штраф за перемикання інтенсивності обслуговування.

Подібні оптимізаційні задачі для одноканальних систем з повторними викликами розглядалися в роботі [6].

Граничні функціонали $f_i(H)$, $i = 1, \ldots, 4$, можуть бути виписані через стаціонарні ймовірності:

$$f_1(H) = \nu^{(1)} \sum_{i=1}^{c} \sum_{j=0}^{H} i\pi_{ij}, \qquad f_2(H) = \nu^{(2)} \sum_{i=1}^{c} \sum_{j=H+1}^{\infty} i\pi_{ij},$$

$$f_3(H) = \lambda \sum_{j=0}^{\infty} \pi_{cj}, \qquad f_4(H) = \lambda\pi_{cH} + (H+1)\mu \sum_{i=0}^{c-1} \pi_{iH+1}.$$

Спираючись на апроксимацію, що дає теорема 2, для підрахунку цільової функції $F(H)$ можна використовувати векторно-матричні формули з теореми 1.

На закінчення відзначимо, що у випадку перевантаженого режиму функціонування керованих систем з повторними викликами ефективним методом аналізу процесу обслуговування є метод дифузійної апроксимації (див. [6]). Підходи до аналізу немарківських систем з повторними викликами можна знайти в [7, 8].

1. *Artalejo J. R., Gomez-Corral A.* Retrial queueing systems. – Berlin: Springer, 2008. – 317 p.
2. *Уолрэнд Дж.* Введение в теорию сетей массового обслуживания. – Москва: Мир, 1993. – 336 с.

3. *Гантмахер Ф. Р.* Теория матриц. – Москва: Наука, 1967. – 576 с.
4. *Лебедев Є. О., Пономарьов В. Д.* Оптимізація систем з повторами і скінченним числом джерел вимог // Вісн. Київ. ун-ту. Сер.: фіз. – мат. науки. – 2008. – № 2. – С. 91–97.
5. *Falin G. I., Templeton J. G. C.* Retrial queues. – London: Chapman and Hall, 1997. – 317 p.
6. *Anisimov V. V.* Switching processes in queueing models. – New York: Wiley, 2008. – 352 p.
7. *Гнеденко Б. В., Коваленко И. Н.* Введение в теорию массового обслуживания. 3-е изд. – Москва: КомКнига, 2005. – 301 с.
8. *Коба Е. В., Коваленко И. Н.* Условие эргодичности для системы с повторными вызовами при нерешетчатом распределении цикла на орбите // Доп. НАН України. – 2004. – № 8. – С. 70–77.

Київський національний університет
ім. Тараса Шевченка

Надійшло до редакції 07.10.2010

E. O. Lebedev, V. D. Ponomarov

On the optimal control over the service rate in retrial queues

The paper deals with the Markov model of a retrial queueing system in which the service rate depends on the queue length. The method is based on the approximation of the input system by a system with finite state space. The problem of finding the optimal control policy in the class of threshold policies is discussed as an example of practical applications.