УДК 519.872

СИСТЕМИ ОБСЛУГОВУВАННЯ З ПОВТОРЕННЯМ І СКІНЧЕННОЮ КІЛЬКІСТЮ ДЖЕРЕЛ НАВАНТАЖЕННЯ

О. В. Коба, д-р фіз.-мат. наук, доц.

Інститут кібернетики імені В. М. Глушкова НАН України

e-koba@yandex.ru

Досліджено замкнені системи обслуговування з поверненням заявок через детермінований час і обмеженою кількістю джерел навантаження немарковського типу $D/M/1/0/n/D$ *і* $M/D/1/0/2/D$. *Побудовано вкладені ланцюги Маркова та системи рівнянь рівноваги. Розроблено методи розв'язання систем. Отримуються показники ефективності функціонування систем, зокрема продуктивність каналу обслуговування, середній час очікування заявки, середнє число повернень заявки.*

Ключові слова: замкнені системи обслуговування з повторенням заявок і обмеженою кількістю джерел навантаження; вкладений ланцюг Маркова; система рівнянь рівноваги, показники ефективності функціонування системи; продуктивність каналу обслуговування.

Closed queueing systems with deterministic retrial time and finite number of traffic sources on non-Markov type $D/M/1/0/n/D$ *and* $M/D/1/0/2/D$ *were investigated. Embedded Markov chains and the system of steady-state equations were built. Systems solutions methods were derived. System functioning characteristics, such as efficiency of service channel, mean waiting time of request, mean number returns of request and so on, were determined.*

Keywords: closed queueing systems with deterministic retrial time efficiency of service channel efficiency of service channel e and finite number of traffic sources; embedded Markov chains; system of steady-state equations; system functioning characteristics, efficiency of service channel.

Вступ

Багато реально існуючих технічних систем (телефонних, комп'ютерних, транспортних) не можна описати класичними системами масового обслуговування. Вони не враховують повторні спроби заявок обслуговуватися у випадку, коли попередні спроби були невдалими (наприклад, канали обслуговування були зайняті). У цьому випадку в системі виникає черга повторних заявок (потік заявок з другою, третю і т. д. спробою обслуговуватися).

Слід зазначити, що ця черга суттєво відрізняється від звичайної. По-перше, повторна заявка може не потрапити на канал обслуговування зараз після його звільнення. Заявка може потрапити на канал тільки тоді, коли відбулася повторна спроба заявки, і за умови, що в цей момент канал вільний. Повторні спроби кожна заявка робить через випадковий час, причому в багатьох випадках передбачається, що цей час розподілений за експоненціальним законом. Кажуть, що заявка між спробами обслуговуватись перебувала на орбіті, яка в свою чергу характеризується виключно часом перебування там заявки.

По-друге, після кожної невдалої спроби зайняти канал обслуговування (як заявка, що вперше надійшла, так і повторна) може піти з системи з імовірністю q .

Зрозуміло, що навіть при існуванні повторних заявок після звільнення каналу на нього може потрапити як заявка, що вперше надійшла, так і заявка з повторного потоку. Відмітимо, що якщо в системі з повторенням заявок дисципліна обслуговування *FIFO* (перший прийшов, перший обслужився), то така система отримала назву система типу Лакатоша (за ім'ям угорського математика Ласло Лакатоша, який вперше її дослідив при побудові математичної моделі процесу посадки повітряних суден) [1].

Постановка проблеми

Системи з повторенням заявок почали розвиватися в кінці XX ст. На сьогодні практично всі моделі систем обслуговування з повторенням вивчалися з експоненціальною орбітою (віртуальне середовище, куди відправляється блокована заявка, що характеризується часом перебування в ньому). Проте, як відомо, ці моделі не є адекватними багатьом реальним технічним системам, що призводить до отримання спотворених показників функціонування систем. Таким чином, дослідження систем обслуговування замкненого типу з неекспоненціальною орбітою на сьогодні є актуальним завданням.

Аналіз досліджень і публікацій

Слід також відмітити, що багато публікацій з цієї тематики розглядають завдання з нескінченною кількістю джерел навантаження [2, 3, 4, 5]. Проте у житті системи із скінченною кількістю джерел навантаження і немарковською орбітою трапляються дуже часто, наприклад, локальні комп'ютерні мережі, комп'ютерні системи, офісні телефонні станції і т. п.

Таким чином, потреба у дослідженні немарковських замкнених систем типу $A/B/m/n/k$ з поверненням заявок і неекспоненціальною орбітою є досить нагальною.

Зазначимо, що в Україні важливі результати для систем з повторними заявками були отримані І. М. Коваленко, В. В. Анісімовим, Є. О. Лебєдєвим, О. В. Кобою та їхніми учнями. Напрямки їх досліджень ергодичніть немарковських систем, багатокаскадні системи, системи зі змінними параметрами, телефонні системи, системи з різнотипними заявками, транспортні системи і т. д. Відмітимо також білоруську школу О. М. Дудіна та В. І. Кліменок, зокрема їх теоретичні результати з додатком до завдань моделювання функціонування комп'ютерних систем та мереж [1, 4, 5].

Мета статті — розглянути замкнені системи обслуговування типу $A/B/m/k/n$ за Кендалом [6] з поверненням заявок. За наявності вільних каналів обслуговування заявка, що надійшла, починає обслуговуватися негайно, у протилежному випадку за наявності вільних місць чекання заявка негайно відправляється у чергу. З другого боку, якщо під час надходження заявки на обслуговування всі канали і всі місця очікування зайняті, то заявка через певний період часу знову робить спробу обслуговуватись.

У статті передбачається розглянути замкненні системи $D/M/1/0/n/D$ та $M/D/1/0/2/D$ (тут 6-та позиція означає розподіл часу орбіти), побудувати відповідні марковські ланцюги та знайти показники функціонування систем.

Опис системи $D/M/1/0/n/D$

Розглянемо спочатку замкнену систему $D/M/1/0/n/D$ з поверненням заявок, тобто замкнену одноканальну систему обслуговування, що обслуговує заявки від n абонентів (тобто система має n джерел навантаження). Час обслуговування абонента — експоненціально розподілена випадкова величина з параметром μ. Кожний абонент, що одержав обслуговування, посилає нову заявку через час T. Якщо в момент надходження заявки канал зайнятий обслуговуванням заявки іншого абонента, то ця заявка повторюється через той самий проміжок часу T. Завдання полягає у визначенні продуктивності каналу обслуговування (середнього числа γ заявок, що обслуговуються за одиницю часу) та інших показників функціонування системи.

Марковський процес та показники функціонування системи $D/M/1/0/n/D$ **з поверненням.** Побудуємо марковський процес, що описує процес обслуговування заявок в системі. Будемо використовувати метод кусочно-лінійних марковських процесів [6].

Введемо дискретну змінну $\upsilon(t)$, що дорівнює одиниці, якщо в даний момент t деяка заявка обслуговується і дорівнює нулю в іншому випадку. Введемо також додаткові змінні $\xi_i(t)$ — тривалості часу до надходження чергових заявок абонентів, не беручи до уваги ту заявку, що, можливо, у даний момент обслуговується.

У результаті одержимо багатовимірний марковський процес зі станами $(1;\ \xi_1(t),...,\xi_{n-1}(t))$ і $(0;\ \xi_1(t),...,\xi_n(t))$, де першою компонентою вектора стану є значення змінної $\upsilon(t)$.

Нумерацію додаткових компонент не будемо зв'язувати з будь-яким упорядкуванням.

Розглянемо поведінку введеного процесу на малому інтервалі часу $(t, t+h)$.

Випадок 1

Якщо $\upsilon(t)=1$; $\xi_1(t)=x_1,...,\xi_{n-1}(t)=x_{n-1}$ і протягом часу h поточне обслуговування не закінчиться, у цьому випадку одержимо $\upsilon(t+h)=1$; $\xi_1(t+h)=x_1-h,...,\xi_{n-1}(t+h)=x_{n-1}-h$ при досить малому h.

Якщо ж обслуговування закінчиться, то $\upsilon(t+h)=0$. При цьому в момент закінчення обслуговування список додаткових компонент поповниться компонентою, що дорівнює T. Будемо вважати, що ця компонента може з рівною ймовірністю зайняти будь-яке місце у векторі стану, тобто новий вектор може мати вигляд $(T, x_1, x_2, ..., x_{n-1})$, $(x_1, T, x_2, ..., x_{n-1})$, …, $(x_1, x_2, ..., x_{n-1}, T)$ з імовірністю $1/n$.

Випадок 2

Якщо $\upsilon(t)=0$; $\xi_1(t)=x_1,...,\xi_n(t)=x_n$, то за досить малого значення h маємо $\upsilon(t+h)=0$; $\xi_1(t+h)=x_1-h,...,\xi_n(t+h)=x_n-h$.

Розглянемо тепер подію, коли одна з компонент $\xi_i(t)$ менша за h.

Якщо $\upsilon(t)=1$, $\xi_i(t)<h$, то в якийсь момент інтервалу $(t,t+h)$ i-а додаткова компонента заміниться значенням T. Якщо $\upsilon(t)=0$, $\xi_i(t)<h$, то ця додаткова компонента зникне. Імовірність того, що $\xi_i(t)<h$ для двох значень i, зневажливо мала.

Побудуємо систему рівнянь ергодичних щільностей та розв'яжемо її.

З урахуванням відзначеного вище, а також з використанням методу, описаного в праці [6], до системи диференціальних рівнянь для ергодичних щільностей $p(x_1,...,x_{n-1})$ і $q(x_1,...,x_n)$ випадкового вектора додаткових компонент, що відповідають випадкам $\upsilon(t)=1$ і $\upsilon(t)=0$, одержимо:

$$\left[\frac{\partial}{\partial x_1}+...+\frac{\partial}{\partial x_{n-1}}-\mu\right]p(x_1,...,x_{n-1})+$$
$$+nq(x_1,...,x_{n-1},0)=0, \quad (1)$$

$$\left[\frac{\partial}{\partial x_1}+...+\frac{\partial}{\partial x_n}\right]q(x_1,...,x_n)=0. \quad (2)$$

Виходячи з нумерації компонент, робимо висновок, що функції p і q симетричні щодо своїх аргументів. Враховуючи цю властивість, складаємо рівняння для граничних значень змінних:

$$q(x_1,...,x_{n-1},T)=\frac{\mu}{n}p(x_1,...,x_{n-1}), \quad (3)$$

$$p(x_1,...,x_{n-2},T)=p(x_1,...,x_{n-2},0). \quad (4)$$

До цього долучимо умову нормування

$$\int_0^T...\int_0^T p(x_1,...,x_{n-1})dx_1...dx_{n-1}+$$
$$+\int_0^T...\int_0^T q(x_1,...,x_n)dx_1...dx_n=1. \quad (5)$$

Прямою підстановкою можна переконатися у тому, що системі (1)–(5) задовольняє такий розв'язок: $p(x_1,...,x_{n-1})=\frac{n}{T^{n-1}(n+\mu T)}$,

$q(x_1,...,x_n)=\frac{\mu}{T^{n-1}(n+\mu T)}$, де $0<x_i<T$ для всіх i.

Інтегруванням за $x_1,x_2,..$ одержуємо ергодичні ймовірності укрупнених станів:

$$\lim_{t\to\infty}P\{\upsilon(t)=1\}=\frac{n}{n+\mu T},$$

$$\lim_{t\to\infty}P\{\upsilon(t)=0\}=\frac{\mu T}{n+\mu T}.$$

Визначимо продуктивність каналу обслуговування. Середнє число γ заявок, що обслуговуються каналом за одиницю часу, можна знайти з наступного ергодичного співвідношення. Протягом досить тривалого часу s канал зайнятий $\frac{ns}{n+\mu T}$; час обслуговування однієї заявки визначається в середньому $1/\mu$. Таким чином,

$$\gamma s\sim\frac{ns}{n+\mu T}\Big/\frac{1}{\mu},$$

звідки

$$\gamma=\frac{n\mu}{n+\mu T}. \quad (6)$$

Отриманий результат має такий імовірнісний сенс. Подамо вираз (6) в іншому вигляді:

$$\frac{1}{\gamma}=\frac{1}{\mu}+\frac{T}{n}.$$

На основі теорії потоків однорідних подій [6] величина $1/\gamma$ (тобто зворотна величина до інтенсивності потоку обслужених заявок) дорівнює середній довжині інтервалу між прийняттям заявок до обслуговування. Цей інтервал складається з часу обслуговування (середнє значення якого дорівнює $1/\mu$) і інтервалу від закінчення обслуговування до надходження наступної заявки. Оскільки повторення заявок кожного абонента відбувається через час T, то природно припустити, що перша заявка надійде в середньому через час T/n. Це міркування пояснює формулу (6).

Визначимо середнє число повернень заявки. Як було відзначено вище, за час s у системі обслуговується в середньому $\frac{n\mu s}{n+\mu T}$ заявок. При $\upsilon(t)=1$ на одиницю реального часу приходиться ($n-1$) одиниць сумарного часу чекання заявок; при $\upsilon(t)=0$ ця кількість складає n одиниць часу. Таким чином, за час s на чекання витрачено $(n-\mathrm{P}\{\upsilon(t)=1\})s$ одиниць часу, тобто,

$$(n-P\{\upsilon(t)=1\})\Big/\frac{n\mu}{n+\mu T}=T+\frac{n-1}{\mu} \quad (7)$$

одиниць часу чекання в розрахунку на одну заявку.

Перший доданок T у правій частині виразу (7) є час з моменту закінчення обслуговування до подачі наступної заявки. Другий доданок, $\max(x,0)$, дорівнює середньому числу $\overline{N}$ повернень заявки даного абонента, помноженому на час T між поверненнями.

Звідси виводимо формулу

$$\overline{N}=\frac{n-1}{\mu T}.$$

Середній час $\overline{W}$ очікування заявки між її подачею і початком обслуговування визначають формулою

$$\overline{W}=T\overline{N}=\frac{n-1}{\mu}.$$

Опис системи $M/D/1/0/2/D$

Тепер розглянемо замкнену систему $M/D/1/0/2/D$ з поверненням заявок, тобто замкнену одноканальну систему обслуговування, що обслуговує заявки від 2-х абонентів (канал обслуговує 2 джерела навантаження). Час обслуговування — стала величина τ. Якщо в даний момент канал зайнятий обслуговуванням заявки, то заявка, що надходить, блокується і повторюється через сталий час $T>\tau$. Якщо в даний момент заявка абонента не блокована і не знаходиться в процесорі, то за малий час h абонент може подати заявку з імовірністю $\lambda h+o(h)$, незалежно від стану процесора (зайнятий — не зайнятий) і попереднього процесу обслуговування.

Завдання полягає в тому, щоб визначити продуктивність процесора γ, де γ — середнє число обслугованих заявок за одиницю часу.

Марковський процес та показники функціонування системи $M/D/1/0/2/D$ **з поверненням.** Позначимо t_n n-й момент закінчення обслуговування заявки. Покладемо $\xi_n = 0$, якщо в момент t_n заявка не блокована; у противному випадку покладемо $\xi_n = (k, x)$, де x — час, що пройшов після моменту блокування заявки; k — число заявок, повністю або частково обслугованих протягом цього часу. За побудовою ξ_n є однорідним ланцюгом Маркова з континуальною множиною станів. Можливі значення змінної k: $1 \le k \le T/\tau$ за цілого T/τ, $1 \le k \le [T/\tau]+1$ за дробового T/τ.

Можливі значення x: $(k-1)\tau \le x \le \min\{k\tau, T\}$.

Ланцюг Маркова ξ_n має ергодичний розподіл, оскільки, якщо $\xi_n = 0$, то $\xi_{n+1} = 0$ з імовірністю $e^{-\lambda\tau}$, а якщо $\xi_n = (k, x)$, то $\xi_{n+1} = 0$ з імовірністю $e^{-\lambda(T-x+\tau)} \ge e^{-\lambda(T+\tau)}$.

Позначимо через p_0 ергодичну ймовірність стану $\xi_n = 0$, через $\mathrm{p}_k(x)h + o(h)$ — стаціонарну ймовірність відрізка станів ξ_n від (k, x) до $(k, x+h)$. Шукана величина γ — продуктивність каналу — може бути виражена через характеристики p_0 і $\mathrm{p}_k(x)$. Дійсно, при $\xi_n = 0$ час простою процесора, що почався, є показова випадкова величина з параметром 2λ; тобто середній час простою дорівнює $1/(2\lambda)$; якщо ж $\xi_n = (k, x)$, то час простою більше z з імовірністю $e^{-\lambda z}$ при $0 < z < T-x$ і з нульовою ймовірністю при $z > x$, а отже, середній час простою складає $\int_0^{T-x} e^{-\lambda z}dz = (1-e^{-\lambda(T-x)})/\lambda$. Звідси випливає, що середній час простою на один цикл є

$$T_0 = \frac{p_0}{2\lambda} + \frac{1}{\lambda}\sum_{k=1}^{[T/\tau]+1}\int_{(k-1)\tau}^{\min\{k\tau,T\}} \mathrm{p}_k(x)(1-e^{-\lambda(T-x)})dx, \quad (8)$$

і тоді шукану величину γ можна визначити формулою

$$\gamma = \frac{1}{T_0+\tau}.$$

Визначимо стаціонарні ймовірності системи.

1. Нехай $\xi_n = 0$. Тоді $\xi_{n+1} = 0$ з імовірністю $e^{-\lambda\tau}$, оскільки для цієї події необхідно і достатньо, щоб за час τ наступного обслуговування абонента заявка від другого абонента не надходила. Якщо ж така заявка надійде через час $\tau - x$ після початку наступного обслуговування, то буде $\xi_{n+1} = (1, x)$.

2. Позначимо через $t_n + \theta$ момент надходження чергової заявки. Якщо $\theta > T - x + \tau$, то $\xi_{n+1} = 0$. Якщо $T - x < \theta < T - x + \tau$, то $\xi_{n+1} = (1, T - x + \tau - \gamma)$.

Якщо $T - x - \tau < \theta < T - x$, то $\xi_{n+1} = (1, \theta + \tau - T + x)$. Якщо $\gamma < T - x - \tau$, то $\xi_{n+1} = (k+1, x + \gamma + \tau)$.

Довизначимо $p_k(x)$ як 0 при неможливих значеннях x; тоді в рівнянні (8) інтеграли можна записати в межах від 0 до T.

Покладемо також

$$p(t) = p_1(t) + p_2(t) + ... \quad (9)$$

Унаслідок наведеного вище закону переходу від ξ_n до ξ_{n+1}, виконуються такі рівняння:

$$p_0 = p_0 e^{-\lambda\tau} + \int_0^T p(t)e^{-\lambda(T-t+\tau)}dt; \quad (10)$$

$$p_1(x) = p_0\lambda e^{-\lambda(\tau-x)} + \lambda\int_0^T p(t)e^{-\lambda(T+\tau-t-x)}dt + \lambda\int_0^{T-\tau+x} p(t)e^{-\lambda(T-t)}dt, \quad (11)$$

якщо $0 \le x \le \tau$;

$$p_1(x) = 0, \ \tau \le x \le T;$$

$$p_k(x) = \lambda\int_0^{x-\tau} p_{k-1}(t)e^{-\lambda(x-t-\tau)}dt, \ k \ge 2; \quad (12)$$

$$p_0 + \int_0^T p(t)dt = 1 . \quad (13)$$

Наведемо алгоритм розв'язання системи рівнянь (8)–13). Тимчасово припустимо, що p_0 — відома стала величина. Покладемо $p_k^{(0)}(x) = 0$.

При $n = 1,2,...$ визначимо $\mathrm{p}_k^{(n)}(x)$ і $p^{(n)}(x) = \sum_k p_k^{(n)}(x)$ рекурентними формулами

$$p_1^{(n)}(x) = p_0\lambda e^{-\lambda(\tau-x)} + \lambda\int_0^T p^{(n-1)}(t)e^{-\lambda(T+\tau-t-x)}dt + \lambda\int_0^{T-\tau+x} p^{(n-1)}(t)e^{-\lambda(T-t)}dt,$$

$$0 \le x \le \tau; \quad (14)$$

$$p_1^{(n)}(x) = 0, \ \tau \le x \le T;$$

$$p_k^{(n)}(x) = \lambda\int_0^{x-\tau} p_{k-1}^{(n)}(t)e^{-\lambda(x-t-\tau)}dt, \ k \ge 2. \quad (15)$$

За індукцією маємо, що $p_k^{(n)}(x)$ і $p^{(n)}(x)$ не спадають по n і не перевищують, відповідно, $p_k(x)$, $p(x)$. Отже, існує границя при $n \to \infty$; це і є шуканий розподіл $p_k(x)$, $p(x)$.

Ітераційний процес можна реалізувати, заміняючи в формулах (14) і (15) p_0 одиницею; отримане рішення нормується умовою (13), чим і визначається p_0.

Висновок

Досліджено замкнені системи обслуговування з скінченною кількістю джерел навантаження і з поверненням заявок через детермінований час, тобто немарковські системи типу $D/M/1/0/n$ і $M/D/1/0/2/D$. Побудовано вкладені ланцюги Маркова та системи рівнянь рівноваги. Отримано показники ефективності функціонування систем: продуктивність каналів обслуговування, середнє число повернень заявки від одного джерела, середній час очікування обслуговування заявки, ймовірності перебування систем в станах. Результати досліджень можуть слугувати для побудови математичних моделей і оцінки показників ефективності функціонування реальних систем, таких як локальні комп'ютерні мережі, комп'ютерні системи, офісні телефонні станції та інші.

ЛІТЕРАТУРА

1. *Коба Е. В.* К классификации систем массового обслуживания с повторением вызовов / Е. В. Коба, И. Н. Коваленко // Кибернетика и системный анализ. — № 3, 2010. — С. 84–91.

2. *Yang T.* A survey on retrial queues / T. Yang, J.G.C. Templeton // Queueing Systems. — 1987. — № 3. — P. 201–233.

3. *Artalejo J.* Standart and retrial queueing systems: a comparative analysis / J. Artalejo, G. Falin // Revista matemática complutense. — 2002. — XV, № 1. — P. 101–129.

4. *Artalejo J.* A classified bibliography of research in retrial queueing. Progress in 1990–1999 / J. Artalejo // Top. — 1999. — № 7. — P. 187–211.

5. *Artalejo J.* A classified bibliography of research in retrial queueing. Progress in 2000–2009 / J. Artalejo // Mathematical and Computer Modeling. — 2010, Vol. 51. — P. 1071–1081.

6. *Гнеденко Б. В.* Введение в теорию массового обслуживания / Б. В. Гнеденко, И. Н. Коваленко. — М. : ЛКИ, 2007. — 400 с.

Стаття надійшла до редакції 16.11.2015