

УМОВА СТІЙКОСТІ СИСТЕМИ ОБСЛУГОВУВАННЯ $GI/G/1$ З T -ПОВЕРНЕННЯМ ЗАЯВОК

О.В. КОБА

Розглянуто систему обслуговування $GI/G/1$, в якій у випадку зайнятості каналу заявка спрямовується на орбіту, з якої повертається через сталий час T . Нехай a — середній інтервал між заявками, τ — середній час обслуговування. Тоді за умови $(T + \tau)/a < 1$ існує стаціонарний режим системи. Доведено, що цю умову неможливо істотно покращити.

1. Вступні зауваження. Розв'язувана в даній статті задача належить до аналізу систем обслуговування з повторенням заявок (з повторними викликами) [1, 2]. У випадку, що розглядається, час, через який заявки повторюються, дорівнює сталому T із ймовірністю 1. Нашим завданням буде встановлення достатньої умови стійкості системи, якщо відомі значення T , середнього часу τ обслуговування заявки та середнього часу a між надходженням заявок.

Слід зазначити, що фактично всю теорію систем з поверненням заявок розвинено для випадку, коли час перебування заявки на орбіті має експоненціальну щільність розподілу. В цьому разі важливо знати тільки загальне число цих заявок, що і пояснює розвинення теорії систем саме з марковським поверненням.

Між тим, у задачах прикладного характеру, особливо у задачах з автоматизацією, марковська модель не є адекватною. Час перебування заявки на орбіті в таких системах часто є детермінованим [3] або має загальний розподіл.

Це і спричинило появу деяких інших робіт автора, наприклад, [4, 5].

2. Достатня умова стійкості системи. Розглянемо систему з поверненням заявок типу $GI/G/1$ з T -поверненням. Моменти надходження заявок в систему $t_n, n \geq 0, t_0 = 0$. Інтервали між надходженням заявок $\xi_n = t_n - t_{n-1}, n \geq 1$. Позначимо $A(x)$ функцію розподілу довжини інтервалу ξ_n між надходженням заявок, $B(x)$ — функцію розподілу часу обслуговування n -ї заявки. Нехай

$$a = \int_0^{\infty} x dA(x), \quad \tau = \int_0^{\infty} x dB(x),$$

T — сталий час перебування заявки на орбіті, W_n — час чекання n -ї заявки. Вважатимемо, що в початковий момент часу система вільна від заявок.

Теорема 1. За умови

$$\frac{T + \tau}{a} < 1$$

випадкова величина W_n має граничний розподіл, тобто

$$P\{W_n \leq x\} \xrightarrow{n \rightarrow \infty} F(x),$$

де $F(x)$ — деяка функція розподілу.

Доведення. Теорему буде доведено за допомогою порівняння даної системи з поверненням заявок з системою $GI/G/1$ з чеканням, у якій ξ_n — те ж саме, що і у вихідної системи, а час обслуговування $\eta_n = T + Y_n$.

Відомо [6], що за умови

$$\rho = \frac{E\{\eta_n\}}{E\{\xi_n\}} < 1$$

математичне сподівання числа заявок, які обслужені за період зайнятості системи $GI/G/1$, скінченне.

Без обмеження загальності покладемо, що період зайнятості системи $GI/G/1$ починається в момент t_0 надходження першої заявки. Якщо N_Q — число заявок, що обслужені за цей період зайнятості, то для будь-якого $n \geq 0$ маємо

$$P\{N_Q > n\} = P\{\eta_0 \geq \xi_1, \dots, \eta_0 + \dots + \eta_{n-1} \geq \xi_1 + \dots + \xi_n\}.$$

Повернемося тепер до системи обслуговування з поверненням заявок. Нехай N_{RQ} — число заявок, які обслужені за період зайнятості, що почався в момент t_0 . Цей період складається з часу обслуговування Y_k та інтервалів між ними довжиною U_k . Очевидно,

$$\begin{aligned} P\{N_{RQ} > n\} &= P\{Y_0 \geq \xi_1, Y_0 + U_1 + Y_1 \geq \xi_1 + \xi_2, \dots \\ &\dots, Y_0 + U_1 + Y_1 + \dots + U_{n-1} + Y_{n-1} \geq \xi_1 + \xi_2 + \dots + \xi_n\}. \end{aligned} \quad (1)$$

Розглянемо стан системи з поверненням заявок в момент закінчення деякого обслуговування. Якщо до цього моменту період зайнятості не закінчився, то деяке число заявок знаходиться на орбіті, тому повернення з орбіти відбудеться раніше, ніж через час T . Таким чином, $U_k \leq T$. Тоді з (1) випливає, що

$$\begin{aligned} P\{N_{RQ} > n\} &\leq P\{(Y_0 + T) \geq \xi_1, \dots, (Y_0 + T) + \dots + (Y_{n-1} + T) \geq \xi_1 + \dots + \xi_n\} = \\ &= P\{\eta_0 \geq \xi_1, \dots, \eta_0 + \dots + \eta_{n-1} \geq \xi_1 + \dots + \xi_n\} = P\{N_Q > n\}. \end{aligned}$$

Підсумовуючи по n , знайдемо

$$E\{N_{RQ}\} = \sum_{n=0}^{\infty} P\{N_{RQ} > n\} \leq \sum_{n=0}^{\infty} P\{N_Q > n\} = E\{N_Q\} < \infty. \quad (2)$$

Якщо в моменти t_n заявка, що надходить, застає систему вільною, то відбувається деяка рекурентна подія. З (2) випливає, що вона додатна.

Оцінимо умовну ймовірність q події {в момент $(t_{n+1}-)$ система вільна} при умові {в момент (t_n-) система вільна}.

Маємо

$$q = P\{Y_{n-1} \leq \xi_n\}.$$

Якщо припустити, що $q = 0$, то

$$P\left\{\frac{1}{n}(Y_0 + \dots + Y_{n-1}) \leq \frac{1}{n}(\xi_1 + \dots + \xi_n)\right\} = 0$$

для будь-якого n , що неможливо, оскільки за законом великих чисел

$$\frac{1}{n}(Y_0 + \dots + Y_{n-1}) \xrightarrow{n \rightarrow \infty} \tau,$$

$$\frac{1}{n}(\xi_1 + \dots + \xi_n) \xrightarrow{n \rightarrow \infty} a \quad (\text{за ймовірністю})$$

і $\tau < a$. Таким чином, $q > 0$, а тому рекурентна подія аперіодична. Оскільки вона також додатна, то є ергодичною. Існування граничного розподілу W_n виводиться з теорії рекурентних подій [7].

Теорему доведено.

3. Неможливість поліпшення отриманої оцінки.

Теорема 2. Для будь-якого $\varepsilon > 0$ існує така система RQ з T -поверненням, для якої

$$\frac{T + \tau}{a} = 1 + \varepsilon$$

і час чекання $W_n \xrightarrow{n \rightarrow \infty} \infty$ за ймовірністю.

Доведення. Візьмемо довільне $p \in (0, 1)$ і припустимо, що $T = 1$, $P\{\xi_n = 1\} = p$, $P\{\xi_n = 0\} = q = 1 - p$. Відносно Y_n припустимо лише, що вони додатні, з математичним сподіванням $\tau < \infty$.

У будь-який момент $t = n$ в систему надходить група S_n заявок (припустимо, що S_n незалежні випадкові величини), де

$$P\{S_n = k\} = pq^{k-1}, \quad k \geq 1,$$

звідки

$$E\{S_n\} = \frac{1}{p}.$$

Позначимо L_n число заявок на орбіті в момент $(n+)$. Для цієї величини виконується рекурентна нерівність

$$L_n \geq (L_{n-1} + S_n - 1)_+, \quad (3)$$

оскільки в момент n повертаються всі заявки, відправлені на орбіту в момент $n-1$ і до них приєднується нова група S_n заявок, за виключенням однієї, що вже обслуговується, якщо $L_{n-1} + S_n > 0$ і попереднє обслуговування закінчено. (Якщо $Y_n \leq 1$ з ймовірністю 1, то (3) буде рівністю.)

Рівняння

$$L_n = (L_{n-1} + S_n - 1)_+$$

можна інтерпретувати як співвідношення для величини черги в системі $GI/D/1$, в якій час обслуговування дорівнює 1, і в будь-який момент $t = n$ надходить геометрично розподілена група заявок. Оскільки навантаження такої системи

$$\rho = \frac{E\{S_n\}E\{Y_n\}}{E\{\text{інтервал між пучками}\}} = \frac{(1/p)1}{1} = \frac{1}{p} > 1,$$

то [7]

$$L_n \xrightarrow[n \rightarrow \infty]{} \infty \quad (4)$$

з ймовірністю 1. (Перехід від рівності до нерівності (3) лише підсилює твердження.)

Позначимо Q_n число заявок на орбіті в момент (t_n) . Тоді $Q_n \geq L_{t_n-1}$. Очевидно, $t_n \rightarrow \infty$ при $n \rightarrow \infty$ з ймовірністю 1. Звідси і з (4)

$$Q_n \xrightarrow[n \rightarrow \infty]{} \infty \quad (5)$$

з ймовірністю 1. Якщо $Q_n \geq K$, то також $Q_{n+i} \geq K - i$, $i > 0$, а тому ймовірність того, що n -а заявка не буде взята на обслуговування за час k ,

$$P\{W_n > k\} \geq \frac{K - k - 1}{K}. \quad (6)$$

Внаслідок (5) права частина (6) прямує до 1 при $n \rightarrow \infty$ при будь-якому k , тобто $W_n \rightarrow \infty$, $n \rightarrow \infty$ за ймовірністю.

Оскільки в даному прикладі $T = 1$, $E\{\xi_n\} = a = p$, то

$$\frac{T + \tau}{a} = \frac{1 + \tau}{p}.$$

Щоб задовольнити умову теореми, достатньо взяти $\tau < \varepsilon$ і покласти $p = \frac{1 + \tau}{1 + \varepsilon}$.

Теорему доведено.

4. Умова, за якої кожна заявка буде обслужена. У п. 2 було виведено достатню умову стійкості системи обслуговування з T -поверненням типу $GI/G/1$, яку в загальному випадку покращити неможливо (теорема 2). Проте у випадку конкретних розподілів потоку та обслуговування, напевне, можливі й точніші оцінки, але ця задача складна. Так, точна умова стійкості системи $M/D/1$ з T -поверненням авторіві невідома.

У цьому параграфі доведемо, що у випадку системи $M/D/1$ за умови $\rho = \lambda\tau < 1$ і будь-якого $T \geq \tau$ кожна заявка буде обслужена із ймовірністю 1.

Нехай t_n — момент надходження n -ї заявки. Введемо подію $C_{nk} : \{n\text{-та заявка не взята на обслуговування до моменту } t + kt \text{ включно}\}$. Якщо ця подія наступила, то в кожному з моментів $t_n, t_n + T, \dots, t_n + kT$ обслуговується або яка-небудь з перших $n-1$ заявок, або заявка, що надійшла в деякому інтервалі

$$(t_n jT - \tau, t_n + jT), \quad 1 \leq j \leq k.$$

Загальне число таких заявок — пуассонівська випадкова величина N_{nk} з параметром $k\rho$.

Таким чином, маємо

$$P\{C_{nk}\} \leq P\{N_{nk} > k - n\}.$$

Оскільки математичне сподівання

$$E\{N_{nk}\} = k\rho$$

і дисперсія

$$\sigma^2\{N_{nk}\} = k\rho,$$

то за нерівністю Чебишева при $k > n/(1 - \rho)$

$$P\{N_{nk} > k - n\} = P\{N_{nk} - E\{N_{nk}\} > k(1 - \rho) - n\} \leq \frac{k\rho}{(k(1 - \rho) - n)^2} \xrightarrow{k \rightarrow \infty} 0.$$

За аксіомою неперервності знаходимо, що із ймовірністю 1 знайдеться таке k_n , при якому n -та заявка обслужить в інтервалі $(t_n + k_n T, t_n + k_n T + \tau)$, що і треба було довести.

ЛІТЕРАТУРА

1. Yang T., Templeton J.G.C. A survey on retrial queues // Queueing Systems. — 1987. — № 2. — P. 201–233.
2. Falin G.I., Templeton J.G.C. Retrial Queues. — London: Chapman & Hall, 1997. — 328 p.
3. Lakatos L. A probability model connected with landing of airplanes // Safety and Reliability. — A.A.Balkema / Rotterdam / Brookfield. — 1999. — 1. — P. 151–154.
4. Коба О.В. Достаточное условие эргодичности системы $M/D/1$ с T -возвращением и приоритетом задержанных заявок // Доповіді НАН України. — 2003. — № 5. — С.17–20.
5. Коба О.В. Стационарные характеристики системы массового обслуживания $GI/G/1$ из T -поверненням при обслуговуванні в порядку черги // Вісник національного авіаційного ун-ту. — 2003. — № 1. — С. 122–125.
6. Гнеденко Б.В., Коваленко И.Н. Введение в теорию массового обслуживания. — М.: Наука, 1987. — 336 с.
7. Феллер В. Введение в теорию вероятностей и ее приложения. 1. — М.: Мир, 1984. — 528 с.

Надійшла 05.04.2004

