

М.Ю. КУЗНЕЦОВ

Інститут кібернетики ім. В.М. Глушкова НАН України, Київ, Україна,
e-mail: kuznetsov2016@icloud.com.

А.А. ШУМСЬКА

Національний технічний університет України «Київський політехнічний інститут
імені Ігоря Сікорського», Київ, Україна,
e-mail: shumskaa@ukr.net.

ПРИСКОРЕНЕ МОДЕЛЮВАННЯ СТАЦІОНАРНОЇ ЙМОВІРНОСТІ БЛОКУВАННЯ ВИКЛИКІВ У ДВОКАНАЛЬНІЙ СИСТЕМІ З ПОРОГОВИМИ СТРАТЕГІЯМИ ОБСЛУГОВУВАННЯ

Анотація. Досліджується модель системи, що складається з двох каналів обслуговування, до кожного з яких надходять декілька різнотипних Пуасонових потоків викликів. Виклик кожного потоку потребує певного ресурсу для свого обслуговування. Визначено порогові стратегії обслуговування, які дають змогу перенаправити виклик з одного каналу до іншого. Блокування викликів певного потоку відбувається за умови відсутності достатніх ресурсів у кожному каналі. Надходження виклику за такого стану каналів обслуговування призводить до його втрати. Для оцінювання стаціонарної ймовірності блокування викликів певного потоку запропоновано метод прискореного моделювання, ефективність якого ілюструється числовими прикладами. Досліджується можливість інваріантності цієї ймовірності стосовно виду функції розподілу часу обслуговування за фіксованою середньою тривалістю обслуговування.

Ключові слова: система обслуговування, стан блокування, ресурс, стаціонарна ймовірність, метод Монте-Карло, метод розширеної вибірки, прискорене моделювання, дисперсія оцінки.

ВСТУП

Проблема інваріантності стаціонарних ймовірностей станів систем обслуговування стосовно виду розподілу часу обслуговування за фіксованою середньою тривалістю обслуговування знаходиться в центрі уваги дослідників з моменту, коли спочатку Р. Форте [1], а потім Б.А. Севастьянов [2] за більш загальних умов довели цю властивість для системи Ерланга $M|G|n|0$. Суттєвий внесок у розвиток теорії інваріантності внесла і українська школа математичної теорії надійності. Тут доцільно згадати статтю Т.П. Мар'яновича [3] та узагальнювальну роботу І.М. Коваленка [4], де було знайдено необхідну і достатню умову інваріантності для певного класу резервованих систем, що дало змогу отримати всі попередні результати як частинні випадки. Суттєвий розвиток теорія отримала у роботі [5], де формули Ерланга та Енгсета були узагальнені на випадок стаціонарних маркованих точкових процесів. З часом ставало все складніше визначати моделі, в яких стаціонарні ймовірності станів знаходяться у вигляді добутку (product form), а саме ця особливість давала змогу доводити інваріантність стаціонарних характеристик стосовно виду певних функцій розподілу. Тому акцент у дослідженнях змістився в бік розроблення наближених методів. Зазначимо два загальні аналітико-статистичні методи [6–8] побудови числових оцінок неперервності характеристик систем обслуговування та резервованих систем. Ці методи поєднують статистичне моделювання вибірових траєкторій випадкових процесів, що визначають поведінку досліджуваних систем, з аналітичним обчисленням деяких функціоналів від цих траєкторій, пов'язаних із малоймовірними подіями. Застосування цих методів до знаходження стаціонарних ймовірностей станів конкретних систем обслуговування розглянуто в [9–11].

© М.Ю. Кузнєцов, А.А. Шумська, 2023

У запропонованій роботі акцентовано увагу на моделі системи, яка розглянута в [12]. Це двоканальна система з пороговими стратегіями обслуговування. Автори статті [12] допускають, що стаціонарні ймовірності блокування викликів вхідних потоків не залежать (або несуттєво залежать) від виду розподілів часу обслуговування для фіксованих середніх значень. При цьому наводяться три алгоритми отримання наближених формул для стаціонарних ймовірностей блокування. Прийнятну точність цих апроксимацій продемонстровано на числових прикладах. На жаль, числові приклади не можуть слугувати доказом справедливості висновків авторів, до того ж всі три запропоновані алгоритми ґрунтуються на формулах (7) та (8) з роботи [12], що може призвести до похибки у декілька порядків (див. приклад у розд. 1 представленої статті). Тому наведені в [12] формули не можна застосовувати до обчислення стаціонарних ймовірностей знаходження системи у стані блокування викликів того чи іншого потоку.

У подальших розділах даної статті для цієї моделі запропоновано альтернативний підхід, в основі якого є метод прискореного моделювання стаціонарних ймовірностей блокування і який використовує, зокрема, принцип розшарованої вибірки. Числові приклади ілюструють зміну відносної похибки оцінок у разі зменшення ймовірності блокування. Здійснено порівняння оцінок, отриманих прискореним моделюванням, з відповідними значеннями, обчисленими за наближеними формулами з [12]. Досліджено, чи може бути коректною гіпотеза про інваріантність ймовірностей блокування стосовно виду розподілів часу обслуговування за фіксованих середніх.

1. ДВОКАНАЛЬНА СИСТЕМА З ПОРОГОВИМИ СТРАТЕГІЯМИ ОБСЛУГОВУВАННЯ

Розглянемо модель, запропоновану в [12]. Система обслуговування складається з двох каналів, $S^{(i)}$ — кількість ліній для обслуговування викликів, що надходять до i -го каналу, $i=1, 2$. До кожного каналу можуть надійти виклики K Пуасонових потоків, $\lambda_k^{(i)}$ — інтенсивність k -го потоку ($k=1, \dots, K$) до i -го каналу ($i=1, 2$), $\Lambda^{(i)} = \sum_{k=1}^K \lambda_k^{(i)}$ — сумарна інтенсивність усіх потоків до i -го каналу, $\Lambda = \Lambda^{(1)} + \Lambda^{(2)}$. Для свого обслуговування виклик k -го потоку потребує r_k ліній, $r = \max_{1 \leq k \leq K} r_k$. Час обслуговування η_k цього виклику має функцію розподілу $G_k(x)$, $M\eta_k < \infty$. Допустимо, що ці розподіли мають щільності $\{g_k(x)\}$.

Для кожного каналу задано два порогових рівня: $\Pi_1^{(i)}$ та $\Pi_2^{(i)}$, $1 \leq \Pi_1^{(i)} < \Pi_2^{(i)} \leq S^{(i)}$. Нехай $\rho^{(i)}$, $i=1, 2$, — це кількість зайнятих ліній в момент надходження виклику k -го потоку до i -го каналу обслуговування. Якщо $0 \leq \rho^{(i)} < \Pi_2^{(i)}$ та $\rho^{(i)} + r_k \leq S^{(i)}$, то кількість зайнятих ліній у i -му каналі збільшується на r_k і миттєво починається обслуговування цього виклику. Допустимо, що $\rho^{(i)} \geq \Pi_2^{(i)}$ або $\rho^{(i)} + r_k > S^{(i)}$. Якщо до того ж $0 \leq \rho^{(3-i)} < \Pi_1^{(3-i)}$ та $\rho^{(3-i)} + r_k \leq S^{(3-i)}$, то канал $3-i$ перебуває у нормальному режимі, при цьому надаються r_k каналів для обслуговування виклику, що надійшов. Якщо $\rho^{(3-i)} \geq \Pi_1^{(3-i)}$ або $\rho^{(3-i)} + r_k > S^{(3-i)}$, то виклик повертається до i -го каналу. У випадку $\rho^{(i)} + r_k \leq S^{(i)}$ починається його обслуговування. В усіх інших випадках виклик отримує відмову та залишає систему, тобто система обслуговування знаходиться у стані блокування викликів k -го потоку, які надходять до i -го каналу.

Метою дослідження є знаходження стаціонарної ймовірності π_k того, що система перебуває у стані блокування викликів k -го потоку, тобто

$$\pi_k = \frac{\lambda_k^{(1)}}{\lambda_k^{(1)} + \lambda_k^{(2)}} \pi_k^{(1)} + \frac{\lambda_k^{(2)}}{\lambda_k^{(1)} + \lambda_k^{(2)}} \pi_k^{(2)},$$

де $\pi_k^{(i)}$ — стаціонарна ймовірність знаходження системи у стані блокування до викликів k -го потоку, які надходять до i -го каналу.

У роботі [12] для знаходження наближених значень ймовірностей $\{\pi_k\}$ запропоновано три підходи, які ґрунтуються на апроксимаційних формулах (7) та (8). Розглянемо приклад, який ілюструє неможливість використання цих формул для наближених обчислень. Допустимо, що у систему обслуговування надходять три потоки ($K=3$), $r_1=1, r_2=2, r_3=5$. Нехай $S^{(1)}=30, S^{(2)}=40, \Pi_1^{(1)}=9, \Pi_2^{(1)}=18$ та $\Pi_1^{(2)}=12, \Pi_2^{(2)}=24$. Якщо вибрати $\lambda_3^{(2)}=10$ та $\lambda_k^{(i)}=10^{-6}$ для всіх інших значень i та k , то маємо такі оцінки: $\pi_1=0.1069, \pi_2=0.1069, \pi_3=0.2139$ (метод Монте-Карло, відносна похибка 1%); $\pi_1=0.1072, \pi_2=0.1071, \pi_3=0.2146$ (метод прискореного моделювання, відносна похибка 0.23%). Водночас рекурентний алгоритм [12] визначає такі оцінки: $\pi_1=8.46 \cdot 10^{-14}, \pi_2=8.46 \cdot 10^{-14}, \pi_3=1.69 \cdot 10^{-13}$. Тому дослідження наведеної вище моделі потребує розроблення принципово інших методів, один з яких запропоновано у наступних розділах цієї статті.

2. ПОБУДОВА МАРКОВСЬКОГО ПРОЦЕСУ

Стан системи обслуговування та її часова еволюція задається неперервним справа марковським процесом:

$$\zeta(t) = (\mu^{(i)}(t); \theta_j^{(i)}(t), \gamma_j^{(i)}(t), j=1, \dots, \mu^{(i)}(t), i=1,2), t \geq 0,$$

де $\mu^{(i)}(t)$ — кількість викликів, які перебувають на обслуговуванні у i -му каналі в момент t ; $\theta_j^{(i)}(t)$ — кількість ліній, які потребує j -й виклик у i -му каналі для обслуговування; $\gamma_j^{(i)}(t)$ — час, що залишився до закінчення обслуговування j -го виклику. Загальна кількість ліній, зайнятих у каналах у момент t , визначається рівністю $\rho^{(i)}(t) = \sum_{j=1}^{\mu^{(i)}(t)} \theta_j^{(i)}(t)$.

Алгоритм моделювання вибіркового траєкторій процесу $\zeta(t)$, $t \geq 0$, є цілком очевидним, тому розглядати його не будемо. Позначимо $\pi_k(t)$ ймовірність того, що виклик k -го потоку, який надійде в момент t , отримає відмову і буде втрачений. Метою дослідження є стаціонарні ймовірності

$$\pi_k = \lim_{t \rightarrow \infty} \pi_k(t), \quad k=1, \dots, K.$$

Для обчислення ймовірностей $\{\pi_k\}$ скористаємось результатами роботи [13]. На відміну від методу Монте-Карло оцінка для π_k будується окремо для кожного k . Тому значення k вважатимемо фіксованим. Нехай t — деякий момент часу, T — тривалість проміжку часу, що починається в момент t (вибір T описано у подальшому). До моменту t моделюємо систему обслуговування методом Монте-Карло і знаходимо стан $\zeta(t)$, який є початковим для подальшого прискореного моделювання траєкторій системи обслуговування у проміжку $[t, t+T]$ з метою потрапляння системи у стан блокування викликів k -го потоку. Для такого моделювання використовуємо метод розшарованої вибірки. Отже, моделювання системи здійснюємо в проміжку $[0, T]$, вважаючи початковим $\zeta(0)=\zeta_0$ той стан, в якому опиниться система за час t .

Допустимо, що в проміжку $[0, T]$ в систему обслуговування надійшло n викликів (розподіл Пуассона з параметром Λ). Моменти $\omega_1, \dots, \omega_n$ надходження цих викликів є рівномірно розподіленими в $[0, T]$ випадковими величинами.

На відміну від методу Монте–Карло не будемо впорядковувати ці моменти за зростанням. Знайдемо ймовірність p того, що обслуговування виклику не закінчиться до моменту T :

$$p = \varphi(T) = \mathbf{P}\{\omega_j + \delta_j > T\} = \frac{1}{T} \int_0^T [1 - G(T - x)] dx = \frac{1}{T} \int_0^T [1 - G(x)] dx, \quad (1)$$

де $\{\delta_j\}$ — незалежні однаково розподілені випадкові величини (тривалості обслуговування), що мають функцію розподілу

$$G(x) = \mathbf{P}\{\delta_j < x\} = \sum_{i=1}^2 \sum_{m=1}^K \frac{\lambda_m^{(i)}}{\Lambda} G_m(x). \quad (2)$$

Момент T можна вибирати різними способами. Скористаємось найпростішим з них. Як момент T візьмемо розв’язок рівняння

$$1 - G(T) = \varepsilon, \quad (3)$$

де ε — деяка мала величина, наприклад $\varepsilon = 10^{-8}$. Це означає вкрай низьку ймовірність того, що обслуговування триватиме довше за час T , тобто вкрай мало ймовірно, що обслуговування викликів, які перебувають у системі в початковий момент часу, не закінчиться до моменту T . Такий вибір T нівелює вплив стану системи в початковий момент часу, отриманого методом Монте–Карло, на стан системи в момент T . Для відомого T із співвідношення (1) знаходимо ймовірність p .

3. АЛГОРИТМ ПРИСКОРЕНОГО МОДЕЛЮВАННЯ ЙМОВІРНОСТІ БЛОКУВАННЯ

Визначимо мінімальну кількість викликів, які мають бути присутніми на обслуговуванні для того, щоб система щонайменше на одному каналі могла б знаходитись у стані блокування викликів k -го потоку. Оскільки r — це максимальна кількість ліній, що може зайняти один виклик, то i -й канал може бути блокованим для викликів k -го потоку, якщо: а) у цьому каналі на обслуговуванні знаходиться щонайменше $V^{(i)} = \min\{j: j \cdot r \geq S^{(i)} - r_k + 1\}$ викликів, б) в каналі $3 - i$ на обслуговуванні знаходиться щонайменше $W^{(3-i)} = \min\{j: j \cdot r \geq \Pi_1^{(3-i)}\}$ викликів.

Позначимо

$$M = \min\{V^{(1)} + W^{(2)}, V^{(2)} + W^{(1)}\} \quad (4)$$

кількість викликів, одночасно присутніх у системі обслуговування, мінімально необхідних для потрапляння у стан блокування викликів k -го потоку.

Нехай L — це кількість викликів, які можуть надійти у проміжку $[0, T]$. Ця випадкова величина має розподіл Пуассона з параметром ΛT . Якщо $L < M$, то система навіть теоретично не може потрапити у стан блокування. Для фіксованого $L = n$ кількість викликів v_T , обслуговування яких не закінчиться до моменту T , має біноміальний розподіл з параметрами n та p . Як оцінку для стаціонарної ймовірності π_k вибираємо $\pi_k(T)$, коли початковий стан в момент $\tau_0 = 0$

$$\xi(\tau_0) = \xi_0 = (\mu^{(i)}(\tau_0); \theta_j^{(i)}(\tau_0), \gamma_j^{(i)}(\tau_0), j = 1, \dots, \mu^{(i)}(\tau_0), i = 1, 2) \quad (5)$$

визначається методом Монте–Карло. Мають місце такі співвідношення:

$$\pi_k(T) = \sum_{l=M}^{\infty} \pi_k^{(l)}(T), \quad (6)$$

$$\pi_k^{(l)}(T) = \mathbf{P}\{v_T = l, L \geq l, A_k(T)\}, \quad (7)$$

де $A_k(T) = \{\text{виклик } k\text{-го потоку, який надійде в момент } T, \text{ отримає відмову}\}$. Формули (6) та (7) реалізують принцип розшарованої вибірки. Параметром

розширення є кількість викликів, обслуговування яких не закінчиться до моменту T (деякі з них можуть отримати відмову в момент надходження). Нехай $M_{\max}$ — довільне ціле число таке, що $M_{\max} \geq M$ (рекомендації щодо вибору $M_{\max}$ надано у подальшому). Тоді формулу (6) запишемо у вигляді

$$\pi_k(T) = \sum_{l=M}^{M_{\max}} \pi_k^{(l)}(T) + \pi_k^*(T), \quad (8)$$

де

$$\pi_k^*(T) = \mathbf{P}\{v_T > M_{\max}, L \geq v_T, A_k(T)\}. \quad (9)$$

Алгоритми моделювання ймовірностей $\pi_k^{(l)}(T)$ та $\pi_k^*(T)$ є подібними. Для фіксованого $l \in \{N, \dots, N_{\max}\}$ маємо

$$\pi_k^{(l)}(T) = \mathbf{M}_{\zeta(\tau_0)} \left[\sum_{n=l}^{\infty} \frac{(\Lambda T)^n}{n!} e^{-\Lambda T} C_n^l p^l (1-p)^{n-l} \mathbf{P}\{A_k(T) | L=n, v_T=l\} \right], \quad (10)$$

$$\begin{aligned} \pi_k^*(T) = \mathbf{M}_{\zeta(\tau_0)} \left[\sum_{j=M_{\max}+1}^{\infty} \sum_{n=j}^{\infty} \frac{(\Lambda T)^n}{n!} e^{-\Lambda T} C_n^j p^j (1-p)^{n-j} \times \right. \\ \left. \times \mathbf{P}\{A_k(T) | L=n, v_T=j\} \right], \quad (11) \end{aligned}$$

де символ $\mathbf{M}_{\zeta(\tau_0)}$ — усереднення за станом системи обслуговування в початковий момент часу. Формули (10) та (11) відповідають випадкам, коли у проміжку $[0, T]$ надійде l викликів, обслуговування яких не закінчиться до моменту T , та відповідно не менше, ніж $M_{\max} + 1$ таких викликів.

На підставі (10) та (11) ґрунтуються алгоритми побудови незміщених оцінок імовірностей $\pi_k^{(l)}(T)$ та $\pi_k^*(T)$. Сформулюємо алгоритм побудови оцінки $\hat{\pi}_{k1}^{(l)}(T)$ в одній реалізації для $\pi_k^{(l)}(T)$ (зміни до цього алгоритму для побудови оцінки $\hat{\pi}_{k1}^*(T)$ наведено далі у зауваженнях до алгоритму).

1. Методом Монте-Карло моделюємо стан системи обслуговування в початковий момент $\tau_0 = 0$ часу: $\zeta(0) = \zeta_0$ (див. (5)).

2. Обчислюємо (див. (10))

$$B^{(l)}(T) = \sum_{n=l}^{\infty} b_n^{(l)}(T), \quad (12)$$

$$b_n^{(l)}(T) = \frac{(\Lambda T)^n}{n!} e^{-\Lambda T} C_n^l p^l (1-p)^{n-l}, \quad n \geq l. \quad (13)$$

3. Покладемо $\hat{\pi}_{k1}^{(l)}(T) = B^{(l)}(T)$ (початкове значення оцінки).

4. Моделюємо кількість L викликів, що надійдуть у проміжку $[0, T]$, а саме $L = n$ з імовірністю $b_n^{(l)}(T)/B^{(l)}(T)$. Нехай $L = n$. Це означає, що у проміжку $[0, T]$ надійшло n викликів, при цьому в l викликах обслуговування не закінчиться до моменту T , однак у $n - l$ викликів обслуговування закінчиться до моменту T .

5. Моделюємо моменти надходження викликів та відповідні тривалості обслуговування. Для цього моделюємо рівномірно розподілені у проміжку $[0, T]$ випадкові величини $\{\omega_j\}$ та тривалості їхніх обслуговувань $\{\delta_j\}$, які мають функцію розподілу $G(x)$, що визначається згідно з (2). Якщо $\omega_j + \delta_j > T$, то ω_j — момент надходження виклику, обслуговування якого не закінчиться до моменту

T (загальна кількість таких моментів становить l). Якщо $\omega_j + \delta_j < T$, то обслуговування такого виклику закінчиться до моменту T (загальна кількість цих моментів складає $n - l$).

6. Впорядковуємо моменти надходження в порядку зростання: $0 = \tau_0 < \tau_1 < \dots < \tau_n$. Відповідні тривалості обслуговування позначимо $y_1, \dots, y_n$.

7. Для відомих моментів надходження викликів $0 = \tau_0 < \tau_1 < \dots < \tau_n < \tau_{n+1} = T$ та тривалостей обслуговування $y_1, \dots, y_n$ знаходимо ймовірність $q_{jm}^{(i)}$ того, що в момент τ_j до каналу i надійде виклик m -го потоку:

$$q_{jm}^{(i)} = \frac{\lambda_m^{(i)} g_m(y_j)}{\sum_{i=1}^2 \sum_{s=1}^K \lambda_s^{(i)} g_s(y_j)}.$$

8. Допустимо, що відомий стан системи в момент τ_j ($0 \leq j \leq n$): $\zeta(\tau_j) = \zeta_j$. У цьому випадку моделюємо стан системи в момент $\tau_{j+1} - 0$. Він відрізняється від попереднього стану лише тим, що систему покидають виклики, залишкова тривалість обслуговування яких була меншою за $\tau_{j+1} - \tau_j$. У решта викликів залишкова тривалість обслуговування зменшується на цю величину.

9. Нехай $j < n$ та $\tau_{j+1} + \delta_{j+1} < T$. Тоді згідно з розподілом $\{q_{j+1,m}^{(i)}\}$ моделюємо номер каналу, до якого надійде виклик, та номер потоку, до якого він належить. Далі знаходимо стан системи обслуговування у момент $\tau_{j+1} + 0$, збільшимо j на одиницю та повертаємось на крок 8 алгоритму.

10. Допустимо, що $j < n$ та $\tau_{j+1} + \delta_{j+1} > T$. Цей крок має декілька етапів.

А. За відсутності надходження викликів у моменти $\tau_{j+1}, \dots, \tau_n$ перевіряємо, чи буде знаходитись система у стані блокування стосовно викликів k -го потоку, які можуть надійти у той чи інший канал в момент T . Якщо відповідь позитивна, то покладемо $J_m^{(i)} = 1$ для будь-яких $i = 1, 2, m = 1, \dots, K$ та перейдемо на етап С. Якщо відповідь негативна, то переходимо на етап В (зауважимо, що значення $J_m^{(i)} = 1$ дає дозвіл на надходження виклику m -го потоку в канал i в момент τ_{j+1} ; якщо $J_m^{(i)} = 0$, то за умов надходження такого виклику система за жодних обставин не зможе потрапити у стан блокування викликів k -го потоку в момент T).

В. Для всіх можливих значень i та m досліджуємо вплив події $D_m^{(i)} = \{\text{у момент } \tau_{j+1} \text{ до каналу } i \text{ надійшов виклик } m\text{-го потоку}\}$ на стан блокування в момент T . Для цього обчислюємо мінімальну кількість викликів, які мають надійти, щоб система щонайменше на одному каналі могла знаходитись у стані блокування викликів k -го потоку. Нехай $\alpha^{(1)}$ та $\alpha^{(2)}$ — кількість зайнятих ліній у відповідних каналах в момент T за умови, що відбулась подія $D_m^{(i)}$, та за відсутності надходження викликів у моменти $\tau_{j+2}, \dots, \tau_n$. Тоді канал v може бути блокованим для викликів k -го потоку, якщо: а) до цього каналу на обслуговування додатково надійде щонайменше $V^{(v)}(\alpha^{(v)}) = \min \{j: j \cdot r \geq S^{(v)} - \alpha^{(v)} - r_k + 1\}$ викликів, б) до каналу $3 - v$ на обслуговування надійде щонайменше $W^{(3-v)}(\alpha^{(3-v)}) = \min \{j: j \cdot r \geq \Pi_1^{(3-v)} - \alpha^{(3-v)}\}$ викликів. Нехай

$$M(\alpha^{(1)}, \alpha^{(2)}) = \min \{V^{(1)}(\alpha^{(1)}) + W^{(2)}(\alpha^{(2)}), V^{(2)}(\alpha^{(2)}) + W^{(1)}(\alpha^{(1)})\}$$

— мінімальна кількість викликів, які мають надійти у систему обслуговування, щоб була можливість потрапити у стан блокування викликів k -го потоку. Якщо $M(\alpha^{(1)}, \alpha^{(2)}) \leq n - (j + 1)$, то покладемо $J_m^{(i)} = 1$; в іншому випадку маємо $J_m^{(i)} = 0$.

С. Обчислюємо $Q_{j+1} = \sum_{i=1}^2 \sum_{m=1}^K J_m^{(i)} q_{j+1,m}^{(i)}$ та $\hat{\pi}_{k1}^{(l)}(T) := \hat{\pi}_{k1}^{(l)}(T) \cdot Q_{j+1}$ (сим-

вол := означає, що нове значення оцінки $\hat{\pi}_{k1}^{(l)}(T)$ дорівнює попередньому значенню, помноженому на Q_{j+1}).

Д. Згідно з розподілом $\{J_m^{(i)} q_{j+1,m}^{(i)} / Q_{j+1}\}$ моделюємо номер каналу, до якого надійде виклик, та номер потоку, до якого він належить. Далі знаходимо стан системи обслуговування в момент $\tau_{j+1} + 0$, збільшуємо j на одиницю та повертаємось на крок 8 алгоритму.

11. Допустимо, що $j = n$. Перевіряємо, чи знаходиться система у стані блокування викликів k -го потоку, які надходять до i -го каналу. Нехай $I_k^{(i)}$ — індикатор цієї події. Тоді як оцінку $\hat{\pi}_{k1}^{(l)}(T)$ в одній реалізації для ймовірності блокування викликів k -го потоку вибираємо

$$\hat{\pi}_{k1}^{(l)}(T) := \hat{\pi}_{k1}^{(l)}(T) \cdot \left(\frac{\lambda_k^{(1)}}{\lambda_k^{(1)} + \lambda_k^{(2)}} I_k^{(1)} + \frac{\lambda_k^{(2)}}{\lambda_k^{(1)} + \lambda_k^{(2)}} I_k^{(2)} \right).$$

Алгоритм закінчено.

Зауваження 1. Алгоритм прискореного моделювання ґрунтується на виконанні умов, необхідних (але не достатніх) для потрапляння системи обслуговування у стан блокування викликів k -го потоку в момент T . Втім суттєво збільшується ймовірність потрапляння системи у стан блокування, але вона не досягає значення одиниці, оскільки під час надходження деякі виклики можуть бути заблокованими і втраченими.

Зауваження 2. Алгоритм побудови оцінки $\hat{\pi}_{k1}^*(T)$ в одній реалізації для $\pi_k^*(T)$ формулюється аналогічно наведеному вище. При цьому кроки 2 та 3 замінюються на такі.

2'. Обчислюємо

$$B(T) = \sum_{j=M_{\max}+1}^{\infty} B^{(j)}(T),$$

де $\{B^{(j)}(T)\}$ визначаються згідно з (12), (13). Як початкове значення оцінки вибираємо $\hat{\pi}_{k1}^*(T) = B(T)$.

3'. Моделюємо випадкову величину, яка приймає значення l з ймовірністю $B^{(l)}(T)/B(T)$.

За умов фіксованого значення l , починаючи з кроку 4, алгоритм повністю повторює попередній.

4. ЗАСТОСУВАННЯ МЕТОДУ РОЗШАРОВАНОЇ ВИБІРКИ ДЛЯ ОЦІНЮВАННЯ ЙМОВІРНОСТІ $\pi_k(T)$

Формула (8) зручно вписується в схему застосування методу розширеної вибірки. Допустимо, що N — загальна кількість реалізацій, які використаємо для побудови оцінки ймовірності $\pi_k(T)$, причому $N = N_M + \dots + N_{M_{\max}} + N^*$, де N_l та N^* — кількості реалізацій, що використовуються для побудови оцінок для $\pi_k^{(l)}(T)$ та $\pi_k^*(T)$. Тоді оцінку $\hat{\pi}_k(T)$ будемо за формулою

$$\hat{\pi}_k(T) = \sum_{l=M}^{M_{\max}} \frac{1}{N_l} \sum_{j=1}^{N_l} \hat{\pi}_{kj}^{(l)}(T) + \frac{1}{N^*} \sum_{j=1}^{N^*} \hat{\pi}_{kj}^*(T). \quad (14)$$

Оскільки всі оцінки є незалежними випадковими величинами, то

$$\mathbf{D} \hat{\pi}_k(T) = \sum_{l=M}^{M_{\max}} \frac{1}{N_l} [\sigma_k^{(l)}(T)]^2 + \frac{1}{N^*} [\sigma_k^*(T)]^2, \quad (15)$$

де $[\sigma_k^{(l)}(T)]^2 = \mathbf{D} \hat{\pi}_{k1}^{(l)}(T)$, $[\sigma_k^*(T)]^2 = \mathbf{D} \hat{\pi}_{k1}^*(T)$. Відомо, що найменше значення дисперсії досягається у разі, коли відповідна кількість реалізацій пропорційна кореню з дисперсії, а саме

$$N_l = N \cdot \frac{\sigma_k^{(l)}(T)}{\sigma_k(T)}, \quad N^* = N \cdot \frac{\sigma_k^*(T)}{\sigma_k(T)}, \quad (16)$$

де $\sigma_k(T) = \sum_{l=M}^{M_{\max}} \sigma_k^{(l)}(T) + \sigma_k^*(T)$. Підставивши (16) у (15), отримуємо

$$\mathbf{D} \hat{\pi}_k(T) = \frac{1}{N} \sigma_k^2(T).$$

Згідно з центральною граничною теоремою оцінка $\hat{\pi}_k(T)$ має асимптотично нормальний розподіл. Тому відносна похибка оцінки, побудованої за формулою (14) з достовірністю $1-\gamma$, визначається у такий спосіб:

$$R_k(T) = \frac{\alpha(\gamma) \sigma_k(T)}{\sqrt{N} \pi_k(T)} 100\%, \quad (17)$$

де $\alpha(\gamma)$ є розв'язком рівняння $\frac{2}{\sqrt{2\pi}} \int_0^{\alpha(\gamma)} e^{-u^2/2} du = 1-\gamma$.

Оскільки значення $\{\sigma_k^{(l)}(T)\}$, $\sigma_k^*(T)$ та $\{\pi_k(T)\}$ невідомі, то для розрахунків використовують їхні статистичні аналоги $\{\hat{\sigma}_k^{(l)}(T)\}$, $\hat{\sigma}_k^*(T)$ та $\{\hat{\pi}_k(T)\}$, за яких на попередньому етапі можна наближено оцінювати $\{N_l\}$ та N^* . Далі на основному етапі оцінка будується згідно з (14), причому моделювання можна організувати з урахуванням даних, отриманих на попередньому етапі.

Оцінка (14) залишається незміщеною для будь-якого вибору $M_{\max} \geq M$. Із збільшенням $M_{\max}$ дисперсія $\sigma_k^2(T)$ спадає, але разом з тим збільшується час моделювання, витрачений на попередньому етапі. З іншого боку, чим більші значення приймає $M_{\max}$, тим меншим буде вплив на $\sigma_k^2(T)$. Будемо вибирати $M_{\max}$ з умови

$$M_{\max} = \min \{l: \sigma_k^*(T) \leq \sigma_k^{(l)}(T)\}. \quad (18)$$

Вибір величин ε , T та $M_{\max}$ розглянемо більш детально на числовому прикладі.

5. ЧИСЛОВИЙ ПРИКЛАД

Розглянуті далі приклади ілюструють точність оцінок, отриманих запропонованим методом прискореного моделювання. Проведено їхнє порівняння з відповідними оцінками, отриманими методом Монте-Карло та за наближеними формулами з [12]. Досліджено залежність ймовірності блокування від виду функцій розподілу тривалості обслуговування за фіксованими середніми.

Розглянемо систему обслуговування, що складається з двох каналів, до яких надходять три потоки викликів ($K=3$), при цьому $r_1=1$, $r_2=2$, $r_3=5$. Інтенсивності вхідних потоків задаємо таким чином: $\lambda_1^{(1)}=0.5$, $\lambda_2^{(1)}=2$, $\lambda_3^{(1)}=0.1$, $\lambda_1^{(2)}=1$, $\lambda_2^{(2)}=0.1$, $\lambda_3^{(2)}=10$. Порогові значення визначаються так: $\Pi_1^{(i)}=0.3 \cdot S^{(i)}$, $\Pi_2^{(i)}=0.6 \cdot S^{(i)}$, $i=1, 2$. Ємність каналів $S^{(1)}$ та $S^{(2)}$ є змінною,

а саме $S^{(1)} = 15j$, $S^{(2)} = 20j$, $j = 1, \dots, 6$. Час обслуговування η_m виклику m -го потоку має функцію розподілу $G_m(x) = 1 - e^{-\pi x^2/4}$, $x \geq 0$, $M\eta_m = 1$. Досліджуватимемо точність оцінок ймовірності блокування викликів другого потоку ($k=2$) в залежності від $S^{(1)}$ та $S^{(2)}$. Аналогічні оцінки мають місце і для $k=1, 3$. Як верхню оцінку ймовірності того, що обслуговування буде тривати довше за T , виберемо $\varepsilon = 10^{-8}$. Згідно з (3) маємо $T = 4.84$. Із співвідношення (1) визначаємо $p = 0.206$. Як початковий проміжок «входження» системи обслуговування в стаціонарний режим візьмемо $[0, 10]$, тобто $t = 10$. Як показали обчислення, збільшення t та T жодним чином не впливає на результат, окрім збільшення часу моделювання. На попередньому етапі оцінюємо дисперсії $\{\sigma_k^{(l)}(T)\}$ та $\{\sigma_k^*(T)\}$. Для цього використовуємо $N_0 = 10000$ реалізацій для кожного $l = M, \dots, M_{\max}$ (параметр M визначається значеннями $S^{(1)}$, $S^{(2)}$, r_k та r (див. (4)); параметр $M_{\max}$ знаходиться на підставі співвідношення (18). На основному етапі використовуємо $N = 10^6$ реалізацій.

У табл. 1 залежно від кількості ліній обслуговування у каналах наведено оцінку $\hat{\pi}_k(T)$ ймовірності блокування викликів другого потоку ($k=2$), оцінку $\hat{\sigma}_k(T)$ дисперсії, оцінку $\hat{R}_k(T)$ відносної похибки. Окрім того, вказано значення параметрів M та $M_{\max}$, які визначають кількість шарів у разі застосування методу розшарованої вибірки.

Із зростанням $S^{(1)}$, $S^{(2)}$ ймовірність блокування зменшується з $4.24 \cdot 10^{-2}$ до $1.63 \cdot 10^{-8}$, а саме на шість порядків. У той же час відносна похибка оцінки зростає тільки в два рази. Якщо скористатися методом Монте-Карло і провести обчислення для перших двох варіантів з відносними похибками 0.57 та 0.39, то отримаємо оцінки $4.24 \cdot 10^{-2}$ та $5.91 \cdot 10^{-3}$ відповідно. Якщо порівняти з методом прискореного моделювання, то у першому випадку час моделювання практично однаковий, а у другому випадку час моделювання зростає в 20.4 рази. З подальшим зменшенням ймовірності $\pi_k(T)$ перевага прискореного моделювання стає ще набагато переконливішою. Якщо скористатися наближеними формулами із [12], то отримаємо такі апроксимації для ймовірності $\pi_k(T)$: $2.22 \cdot 10^{-2}$, $1.09 \cdot 10^{-3}$, $6.34 \cdot 10^{-5}$, $8.68 \cdot 10^{-7}$, $7.23 \cdot 10^{-9}$, $1.20 \cdot 10^{-11}$. Вочевидь, що ці результати свідчать про хибність запропонованих апроксимальних формул, бо похибка становить від 1.9 до 1358 разів.

Таблиця 1

Кількість ліній у каналах	M	$M_{\max}$	$\hat{\pi}_k(T)$	$\hat{\sigma}_k(T)$	$\hat{R}_k(T)$, %
$S^{(1)} = 15$, $S^{(2)} = 20$	5	27	$4.24 \cdot 10^{-2}$	$1.50 \cdot 10^{-8}$	0.57
$S^{(1)} = 30$, $S^{(2)} = 40$	9	27	$5.92 \cdot 10^{-3}$	$1.36 \cdot 10^{-10}$	0.39
$S^{(1)} = 45$, $S^{(2)} = 60$	13	27	$1.16 \cdot 10^{-3}$	$8.61 \cdot 10^{-12}$	0.50
$S^{(1)} = 60$, $S^{(2)} = 80$	17	28	$8.05 \cdot 10^{-5}$	$6.35 \cdot 10^{-14}$	0.61
$S^{(1)} = 75$, $S^{(2)} = 100$	21	30	$2.04 \cdot 10^{-6}$	$7.57 \cdot 10^{-17}$	0.84
$S^{(1)} = 90$, $S^{(2)} = 120$	26	35	$1.63 \cdot 10^{-8}$	$9.79 \cdot 10^{-21}$	1.19

Таблиця 2

Функції розподілу $\{G_m(x)\}$	$\hat{\pi}_k(T)$	$\hat{R}_k(T), \%$	Довірчий інтервал
$1 - e^{-\pi x^2/4}$	$8.05 \cdot 10^{-5}$	0.61	$(8.00 \cdot 10^{-5}, 8.10 \cdot 10^{-5})$
$1 - e^{-x}$	$8.08 \cdot 10^{-5}$	0.61	$(8.03 \cdot 10^{-5}, 8.13 \cdot 10^{-5})$
$1 - \frac{1}{\left(1 + \frac{x}{9}\right)^{10}}$	$8.08 \cdot 10^{-5}$	0.61	$(8.04 \cdot 10^{-5}, 8.13 \cdot 10^{-5})$
$1 - \frac{1}{\left(1 + \frac{x}{6}\right)^7}$	$8.08 \cdot 10^{-5}$	0.61	$(8.04 \cdot 10^{-5}, 8.13 \cdot 10^{-5})$
$1 - \frac{1}{\left(1 + \frac{x}{3}\right)^4}$	$8.08 \cdot 10^{-5}$	0.60	$(8.03 \cdot 10^{-5}, 8.12 \cdot 10^{-5})$

Допустимо тепер, що тривалість обслуговування має фіксоване середнє значення, наприклад $M\eta_m = 1$, $m = 1, \dots, K$. На числовому прикладі перевіряємо, наскільки функціональний вид розподілу тривалості обслуговування впливає на ймовірність блокування $\pi_k(T)$. Нехай $S^{(1)} = 60$, $S^{(2)} = 80$. Таблиця 2 містить оцінки та відповідні довірчі інтервали, які побудовано для п'яти варіантів функцій розподілу $\{G_m(x)\}$ з достовірністю 0.99 за $N = 10^6$ реалізаціями (перші два варіанти — розподіл Вейбулла, решта три варіанти — розподіл Парето).

Усі розподіли відрізняються один від одного досить суттєво, хоча середні тривалості обслуговування є однаковими. Разом з тим усі довірчі інтервали мають спільні точки. Отже, можна зробити висновок: наведений приклад не суперечить гіпотезі, що ймовірність блокування не залежить від виду розподілів часу обслуговування, а визначається лише середніми характеристиками. Приклади, що не увійшли в цю статтю, також не спростовують таку гіпотезу. Отже, автори вважають цілком обґрунтованими спроби математично строго довести запропоновану гіпотезу інваріантності.

ВИСНОВКИ

У роботі показано, що запропоновані раніше наближені формули для знаходження стаціонарної ймовірності блокування викликів у двоканальній системі з пороговими стратегіями обслуговування є помилковими. Автори пропонують альтернативний підхід, який ґрунтується на сумісному використанні методу розширеної вибірки та спеціальних алгоритмів моделювання, спрямованих на суттєве підвищення ймовірності потрапляння системи обслуговування у стан блокування викликів конкретного потоку. Наведені приклади підтверджують обґрунтованість такого підходу. Запропонований метод можна використовувати для оптимізації кількості ліній у кожному каналі для забезпечення заданих імовірностей блокування для кожного потоку викликів.

СПИСОК ЛІТЕРАТУРИ

1. Fortet R.M. Random distributions with an application to telephone engineering. Proc. Berkeley Sympos. Math. Stat. and Prob. Berkeley. 1956. Vol. 2. P. 81–88.

2. Севастьянов Б.А. Предельная теорема для марковских процессов и ее приложение к телефонным системам с отказами. *Теория вероятностей и ее применения*. 1957. Т. 2, № 1. С. 106–116.
3. Марьянович Т.П. Обобщение формул Эрланга на случай, когда приборы могут выходить из строя и восстанавливаться. *Укр. мат. журн.* 1960. Т. 12, № 3. С. 279–286.
4. Коваленко И.Н. Об условии независимости стационарных распределений от вида закона распределения времени обслуживания. *Проблемы передачи информации*. 1963. Вып. 11. С. 147–151.
5. König D., Matthes K., Nawrotzki K. Verallgemeinerungen der Erlangschen und Engsettschen Formeln. Berlin: Akademie-Verlag, 1967. 123 S.
6. Коваленко И.Н. К расчету поправок к характеристикам СМО. *Проблемы устойчивости стохастических моделей*. Москва: ВНИИСИ, 1986. С. 45–48.
7. Кузнецов Н.Ю. Аналитико-статистический метод построения количественных оценок непрерывности характеристик систем массового обслуживания и резервированных систем. *Проблемы устойчивости стохастических моделей*. Москва: ВНИИСИ, 1986. С. 54–62.
8. Коваленко И.Н., Кузнецов Н.Ю. Методы расчета высоконадежных систем. Москва: Радио и связь, 1988. 176 с.
9. Кузнецов Н.Ю. Нахождение стационарных вероятностей состояний системы $GI|G|n|0$ с входящим потоком требований, близким к пуассоновскому. *Кибернетика*. 1984. № 2. С. 74–79.
10. Кузнецов Н.Ю., Шумская А.А. Оценка отклонения стационарных вероятностей системы $GI|G|m|r$ от вероятностей состояний системы $M|M|m|r$ аналитико-статистическим методом. *Кибернетика и системный анализ*. 2013. № 5. С. 51–60.
11. Кузнецов М.Ю., Кузнецов І.М., Шумська А.А. Знаходження стохастичного градієнта для оптимізації ефективності систем, які описуються деревами відмов з ефективною. *Міжнародний науково-технічний журнал «Проблеми керування та інформатики»*. 2022. № 4. С. 76–88.
12. Sagkriotis S.G., Pantelis S.K., Moscholios I.D., Vassilakis V.G. Call blocking probabilities in a two-link multirate loss system for Poisson traffic. *IET Networks*. 2018. Vol. 7, N 4. P. 233–241.
13. Кузнецов Н.Ю., Кузнецов И.Н. Ускоренное моделирование вероятности блокировки требований в сетях обслуживания с множественным доступом. *Кибернетика и системный анализ*. 2021. Т. 57, № 4. С. 30–43.

M.Yu. Kuznetsov, A.A. Shumska

FAST SIMULATION OF STEADY-STATE CALL BLOCKING PROBABILITY IN A TWO-CHANNEL SYSTEM WITH THRESHOLD SERVICE STRATEGIES

Abstract. A model of a queueing system consisting of two service channels is considered. Each channel receives calls from several Poisson flows of different types. A call of each flow requires a certain resource for its service. Threshold service strategies are defined that make it possible to redirect a call from one channel to another. Blocking of calls of a certain flow occurs if each channel does not have sufficient service resources. All calls coming in a system-blocking state are lost. To evaluate the steady-state probability of blocking calls of a certain flow, a fast simulation method is proposed. The accuracy of the estimates is illustrated by numerical examples. The possibility of the insensitivity of the steady-state blocking probabilities with respect to the form of service time distributions by fixed means is investigated.

Keywords: queueing system, blocking state, resource, steady-state probability, Monte–Carlo method, stratified sampling, fast simulation, variance of estimate.

Надійшла до редакції 27.03.2023