УДК 621.391:519.872
DOI: 10.31673/2412-9070.2023.060307

В. О. ВЛАСЕНКО, канд. техн. наук, доцент;
К. О. ТРЕНЬОВА, аспірантка,
Державний університет інформаційно-комунікаційних технологій, Київ

МЕТОДИКА РОЗРАХУНКУ ЧАСУ ЗАТРИМКИ В СИСТЕМІ КЕРУВАННЯ ІНФОКОМУНІКАЦІЙНИМИ МЕРЕЖАМИ НА БАЗІ ТЕОРІЇ МАСОВОГО ОБСЛУГОВУВАННЯ

Розглянуто методи розрахунку часу затримки в системі обслуговування типу M/M/m, досліджено основні компоненти моделі, зокрема інформаційні потоки та пристрій обслуговування, і вдосконалено формулу розподілу часу затримки на основі закону Літтла. Зазначено важливість цього показника у визначенні ефективності інфокомунікаціних систем та оптимізації їх функціонування.

Наведено приклади розрахунків часу затримки та проаналізовано фактори, що впливають на цей показник, такі як завантаженість телекомунікаційної системи та характеристики пристрою обслуговування. Крім того, сформульовано можливі шляхи вдосконалення системи, включно з використанням буферів та більш складних моделей.

M/M/m — це система з m обслуговувальними лініями, пуассонівським вхідним потоком, показовим розподілом часу обслуговування і дисципліною «першим прийшов – першим обслужений» (англ. FCFS — First-Come, First-Served), що означає, що заявки обслуговуються в тому порядку, в якому вони надходять.

Детальний аналіз моделі M/M/m та вивчення її основних складових дають змогу глибше розібратися з процесами в системах масового обслуговування. Час затримки в системі M/M/m відображається в його прямому впливі на якість обслуговування та задоволення користувачів таких мереж.

Ключові слова: мережа передавання даних; система керування; вузол комутації; навантаження; пакет даних; потік Пуассона; система масового обслуговування; інтенсивність дзвінків.

ВСТУП

Системи масового обслуговування нині використовуються в різних галузях, включно з телекомунікаціями, інформаційними технологіями, транспортом, торгівлею тощо. Однією з головних характеристик таких систем є час затримки, який визначає ефективність їх функціонування. У цій статті розглядається методика розрахунку часу затримки в системі обслуговування типу $M/M/m$, де M — це випадковий розподіл часу обслуговування та надходження викликів від абонентів, а m — кількість обслуговувальних пристроїв.

ОСНОВНА ЧАСТИНА

Розрахунок часу затримки в системі керування на основі моделі масового обслуговування без пріоритетів

У процесі вивчення мереж із комутацією пакетів постають проблеми черг. Пакети, що надходять на вхід, нагромаджуються, обробляються та вибирають канал передавання до наступного вузла. Затримка передавання в нагромаджувачі є ключовим показником роботи мережі, що позначається на користувачі. Час очікування залежить від оброблення пакету комутації у вузлі, довжини пакету, пропускної здатності та інтенсивності надходження пакетів [3]. Теорія черг вивчається в контексті мереж із комутацією каналів, включно з аналізом залежностей між кількістю каналів і ймовірністю блокування. Вона виникла в дослідженнях телефонних мереж і застосовується в інтегральних мережах, що поєднують комутацію пакетів і каналів.

Розглянемо найпростішу модель обслуговування, наведену на рис. 1.

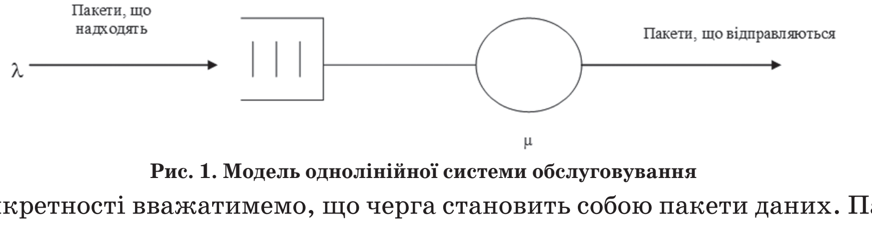

Рис. 1. Модель однолінійної системи обслуговування

Для більшої конкретності вважатимемо, що черга становить собою пакети даних. Пакети надходять випадково з середньою швидкістю μ пакетів за одиницю часу. Вони очікують обслуговування в нагромаджувачі й обслуговуються відповідно до деякої конкретної дисципліни із середньою швидкістю μ пакетів за одиницю часу.

© *В. О. Власенко, К. О. Треньова, 2023*

У контексті мережі передавання даних обслуговувальна лінія — це засіб передавання (вихідний канал або лінія), який передає дані із зазначеною швидкістю з блоків даних за одиницю часу [4].

Очевидно, якщо інтенсивність λ надходження пакетів наближається до швидкості μ оброблення пакетів, черга починає збільшуватись. У разі максимального нагромадження (реальний випадок), коли λ перевищить μ, черга досягне найбільшого допустимого значення і продовжуватиме зростати. Якщо нагромаджувач переповниться, надходження всіх наступних пакетів буде заблоковано. Для простоти передбачимо, що нагромаджувач є нескінченним, тоді черга при $\lambda \to \mu$ стає нестабільною.

Доведемо, що в аналізованій однолінійній системі обслуговування стабільність забезпечується при $\lambda < \mu$. Зокрема, визначимо параметр $\rho = \lambda/\mu$. Параметр ρ зазвичай називають *коефіцієнтом використання каналу* або *інтенсивністю навантаження*.

Здебільшого прийнято моделювати процес надходження пакетів або викликів за допомогою пуассонівського процесу, який переважно застосовується в теорії черг — методі опису вхідного потоку [3].

Найпростіша система обслуговування, яку буде досліджено, так звана черга типу $M/M/m$, є системою з пуассонівським вхідним потоком і показовим розподілом часу обслуговування (рис. 2) [2]. Для такої системи досить легко здобути ймовірності станів як у разі кінцевої, так і в разі нескінченної черги. Потім виведемо просту, але найбільш загальну залежність між середнім значенням часу затримки і середньою кількістю користувачів у черзі, названу формулою Літтла. Ця залежність є корисною в процесі розрахунку характеристик мереж різного типу.

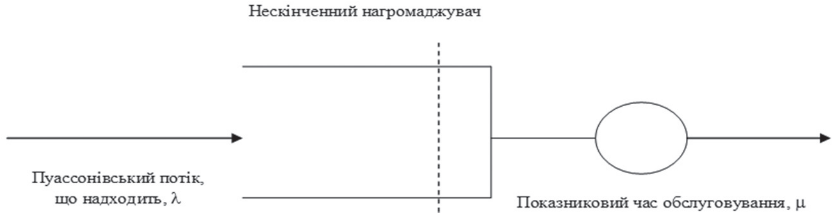

Рис. 2. Система обслуговування *M/M/m*

Розрахунок часу затримки в системі керування на базі системи обслуговування M/M/m

Як уже зазначалося, $M/M/m$ — це система з m обслуговувальними лініями, пуассонівським вхідним потоком, показовим розподілом часу обслуговування і дисципліною «першим прийшов – першим обслужений» (англ. **FCFS** — *First-Come, First-Served*), тобто заявки обслуговуються в тому порядку, в якому вони надходять. Як відомо з науково-технічної літератури [5], статистичні властивості системи $M/M/m$, наприклад, середній час зайняття, імовірність блокування для кінцевої черги, середня пропускна здатність тощо, досить легко можуть бути розраховані, якщо знайдено ймовірність p_n станів системи. За визначенням p_n — це ймовірність того, що в системі містяться n пакетів у мережі каналів з очікуванням або викликів у мережі з комутацією каналів, включно з користувачем, що перебуває на обслуговуванні.

Передбачається, що система працює у сталому режимі, і тому зазначені ймовірності не змінюються в часі. Нехай процес обслуговування (довжина пакету або тривалість з'єднання) описується показовим розподілом із параметром μ (визначає швидкість оброблення вимог на обслуговування). Тоді ймовірність $p_n(t + \Delta t)$ того, що в момент часу $t + \Delta t$ у системі буде розміщено n пакетів (викликів), легко можна дістати як функцію відповідних імовірностей у момент часу t. Із діаграми зміни станів у часі (рис. 3) випливає, якщо система в момент часу $t + \Delta t$ перебуває в стані n, то в момент t вона може бути тільки в станах $n - 1$, n або $n + 1$ (при $n \geq 1$).

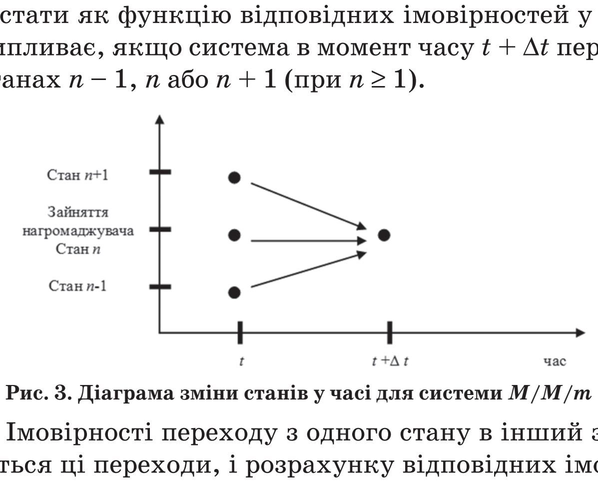

Рис. 3. Діаграма зміни станів у часі для системи *M/M/m*

Імовірність $p_n(t+\Delta t)$ того, що в момент часу $t+\Delta t$ система перебуває в стані n, має бути сумою (що взаємно виключається) імовірностей станів $n-1$, n або $n+1$, в яких система могла бути в момент t, помножених відповідно на (незалежні) ймовірності надходження в стан n, що відбувається за Δt одиниць часу:

$$p_n(t+\Delta t) = p_n(t)[(1-\lambda\Delta t)(1-\mu\Delta t) + \mu\Delta t\cdot\lambda\Delta t + 0(\Delta t)] + \\ + p_{n-1}(t)[\lambda\Delta t(1-\mu\Delta t) + 0(\Delta t)] + \\ + p_{n+1}(t)[(1-\lambda\Delta t)\mu\Delta t + 0(\Delta t)]. \quad (1)$$

Імовірності переходу з одного стану в інший здобуто внаслідок розгляду шляхів, за якими відбуваються ці переходи, і розрахунку відповідних імовірностей із використанням властивостей потоку, що надходить, та розподілом часу обслуговування. Наприклад, якщо система залишилася в стані n, $n \geq 1$, то могло статися: одне посилання й одне надходження з імовірністю $\mu\Delta t\cdot\lambda\Delta t$, або жодного посилання та надходження з імовірністю $(1 - \lambda\Delta t)$ $(1 - \mu\Delta t)$. Аналогічно обчислюємо й інші члени рівності (1), оскільки $0(\Delta t)$ містить у собі члени порядку $(\Delta t)^2$ і вищого.

Після спрощення й об'єднання членів $0(\Delta t)$ матимемо:

$$p_n(t + \Delta t) = [1-(\lambda + \mu)\Delta t]\cdot p_n(t) + \lambda\Delta t p_{n-1}(t) + \mu\Delta t + p_{n+1}(\mathrm{t})\mu\Delta t. \quad (2)$$

Рівність (2) може бути використана для дослідження перехідного (що залежить від часу) режиму роботи системи $M/M/m$, за умови, що ця робота починається в момент $t_0 = 0$ з деякого відомого стану або набору станів. Інакше через розклад $p_n(t + \Delta t)$ у ряд Тейлора відносно t зі збереженням лише перших двох членів можна знайти диференційне рівняння, що описує зміну $p_n(t)$ у часі:

$$p_n(t + \Delta t) = p_n(t) + \frac{dp_n(t)}{dt}\Delta t. \quad (3)$$

Скориставшись виразами (2) і (3) і виконавши спрощення, дістанемо таке рівняння:

$$\frac{dp_n(t)}{dt} = -(\lambda + \mu)p_n(t) + \lambda p_{n-1}(t) + \mu p_{n+1}(t). \quad (4)$$

Це диференційне рівняння, розв'язання якого дає результат — залежність $p_n(t)$ від реального часу, тобто ймовірність $p_n(t)$ має наближатися до сталого стаціонарного значення $p_n(t)$ зі збільшенням часу. Отже, у стаціонарному значенні $p_n(t)$ маємо $dp_n(t)/dt = 0$. Тоді рівняння (4) для випадку стаціонарних, незмінних у часі ймовірностей, спрощується і для системи $M/M/m$ набирає вигляду

$$(\lambda + \mu)p_n(t) = \lambda p_{n-1}(t) + \mu p_{n+1}(t),\ n \geq 1. \quad (5)$$

Як унаочнює діаграма станів (див. рис. 3), на якій показано систему $M/M/m$, з інтенсивністю λ відбувається перехід на один стан праворуч у разі надходження користувача в систему, а з інтенсивністю μ — перехід на один стан ліворуч під час завершення обслуговування або відходження пакету. Інакше кажучи, якщо інтенсивність помножити на Δt, матимемо ймовірність $\lambda\Delta t$ переходу на один стан праворуч завдяки надходженню, або ймовірність $\mu\Delta t$ переходу на один стан ліворуч через завершення обслуговування.

Рівняння (5) для ймовірностей станів може бути розв'язане кількома способами. Потік, що надходить у ділянку, яка охоплює всю безліч точок від 0 до n такий, що дорівнює μp_{n+1}, а потік, який покидає її, — λp_n. Прирівнявши ці два потоки, дістанемо:

$$\lambda p_n = \mu p_{n+1}. \quad (6)$$

Рівновага інтенсивності виходу з деякого стану і повернення в цей стан не тільки дає можливість записати рівняння рівноваги (5), а й здобути його розв'язок:

$$p_n = \rho_n \cdot p_0; \quad \rho \equiv \lambda/\mu. \quad (7)$$

Для того, щоб визначити невідому ймовірність p_0, потрібно звернутися до умови для ймовірностей:

$$\sum_n p_n = 1.$$

Для випадку нескінченної черги типу $M/M/m$ досить просто відшукаємо $p_0 = (1 - \rho)$, при $\rho < 1$, що приведе до розв'язку для сталого режиму у вигляді

$$p_n = (1 - \rho)\rho^n, \quad \rho = \lambda/\mu < 1. \quad (8)$$

Але ця ймовірність має бути такою самою, як і ймовірність блокування, тобто ймовірність того, що пакети або виклики не можуть бути прийняті системою [1]. Запропонуємо припущення, яке доцільно використати під час розрахунку кількісних характеристик інтелектуальної мережі.

Навантаження λ, що визначається як середня кількість надходжень пакетів за секунду, зображено на рис. 4. У разі ймовірності блокування $p_{БЛ}$ чиста інтенсивність надходжень дорівнює $\lambda(1-p_{БЛ})$. Але це не що інше, як пропускна здатність γ або кількість пакетів, що обслуговуються за секунду стаціонарною системою.

Рис. 4. Співвідношення між продуктивністю і навантаженням

Отже маємо:

$$\gamma = \lambda(1 - p_{БЛ}). \quad (9)$$

У літературі [3] показано, що в кінцевій системі $M/M/m$ імовірність блокування фактично дорівнює

$$p_{БЛ} = p_N = (1 - \rho)\rho^N/(1 - \rho^{N-1}). \quad (10)$$

Цей вираз для ймовірності блокування може бути використано в простих розрахунках у процесі проєктування. За невеликих значень імовірності блокування вираз (10) можна спростити.

Доцільно розглянути приклад при $\rho < 1$. Нехай $E(n)$ — середня кількість користувачів (пакетів або викликів у черзі) в системі середнього значення випадкової величини.

На основі визначення середнього значення випадкової величини дістанемо

$$E(n)=\sum_{n=0}^{\infty} np_n = \rho/(1-\rho). \qquad (11)$$

Коли навантаження системи відносно невелике (наприклад, $\rho = \lambda/\mu \le 0{,}5$), середня кількість користувачів у системі відносно мала (при $\rho < 0{,}5$ вона менша за одиницю). Коли ж ρ збільшується, наближаючись до одиниці, то середня кількість користувачів різко зростає завдяки члену $(1 - \rho)$ у знаменнику. У реальній системі обслуговування з кінцевою чергою ця кількість, очевидно, в ділянці $\rho < 1$ зростає не так різко, але рівність (11) для нескінченної черги дає прийнятну модель. Середню довжину черги в системі $M/M/m$ (11) показано на рис. 5.

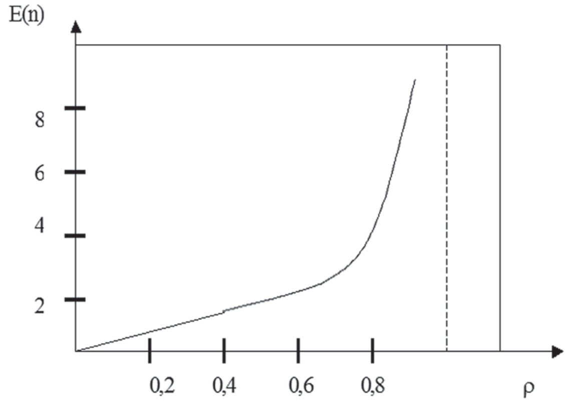

Рис. 5. Середня довжина черги в системі *M/M/m*

Зі зростанням навантаження системи нарощується й її продуктивність. Однак при цьому блокується все більша кількість користувачів, отже, швидко росте середня кількість $E(n)$ користувачів у черзі. Збільшення $E(n)$ зумовлює збільшення часу затримки в черзі. Це типовий випадок обміну характеристик: із підвищенням навантаження росте продуктивність (важлива характеристика), але одночасно зростає блокування і час затримки [3].

У системі керування дуже важливо знайти загальний час затримки команд керування отриманням послуги (зокрема часу очікування в черзі і, до того ж, часу обслуговування або передавання).

Щоб здобути наближені кількісні характеристики, доцільно скористатися формулою Літтла:

$$\lambda E(T) = E(n). \qquad (12)$$

Параметр λ інтерпретується як інтенсивність надходжень у систему і відповідає продуктивності ν.

Застосувавши формулу (12) до системи $M/M/m$, дістанемо:

$$E(T) = E(n)/\lambda = 1/\mu(1 - \rho). \qquad (13)$$

У разі $\rho << 1$ точне значення середнього часу обслуговування дорівнює $E(T) = 1/\mu$.

Це випадок, коли в черзі в середньому перебуває невелика кількість користувачів. Тому на очікування в черзі приблизно витрачається мало часу, і затримка майже завжди пояснюється лише обслуговуванням або часом передавання. Однак зі збільшенням нормованого навантаження або інтенсивності тракту звичайний вплив черги починає відчуватися і $E(T)$ швидко зростає.

Однолінійна система має дотримуватися простого співвідношення між середнім часом очікування $E(W)$ і середньою затримкою $E(T)$ у системі:

$$E(T) = E(W) + 1/\mu. \qquad (14)$$

Теорема Літтла дає нам змогу знайти в реальному вигляді співвідношення для середньої кількості користувачів $E(q)$, які очікують в черзі:

$$E(q) = \lambda E(W) = \lambda E(T) - \lambda\mu = E(n) - \rho. \qquad (15)$$

Формула Літтла є загальною і вона може бути застосована також до частини системи обслуговування.

Висновки

У статті наведено розрахунок часу затримки в системі обслуговування $M/M/m$, що є важливим показником для оцінювання ефективності та продуктивності таких систем. Побудова моделі, розрахунок формули часу затримки дало змогу здобути результати аналізу впливу різних факторів на час очікування клієнтів у черзі.

Модель $M/M/m$ дає можливість уявити ситуацію, де перебування клієнтів та час їх обслуговування є випадковими подіями. Формула для розрахунку часу затримки базується на законі Літтла, що уможливлює прогнозування та оптимізацію роботи системи.

Було взято до уваги фактори, що впливають на час затримки, зокрема завантаженість системи, інтенсивність потоку клієнтів та характеристики пристрою обслуговування. Удосконалення системи через використання буферів та більш складних моделей сприяло зменшенню часу очікування та покращенню загальної продуктивності мережі передавання даних.

Визначення часу затримки в системі обслуговування $M/M/m$ є ключовим елементом керування та оптимізації процесів. Розрахунок цього показника дав можливість удосконалювати мережу та надавати користувачам більш якісні послуги, що є важливим аспектом у сучасному світі щораз вищого попиту на швидкі та ефективні вирішення.

Список використаної літератури

*1. **Інтелектуальна** система керування інфокомунікаційними мережами / Л. Беркман, О. Барабаш, О. Ткаченко [та ін.] // Міжнар. журн. нових тенденцій в інженерних дослідженнях. 2020. № 8 (5). С. 1920–1925.*

*2. **Козловський В. В., Туровський О. Л., Кулінський В. Д.** Формалізація вимог до системи керування телекомунікаційними мережами // Зб. наук. праць «Проблеми інформатизації та керування». 2020. №64.*

*3. **Толубко В. Б., Беркман Л. Н.** Методи оптимізації. Київ: ДУТ, 2016. С. 442.*

*4. **Кучук Г. А., Коваленко А. А., Лукова-Чуйко Н. В.** Метод мінімізації середньої затримки пакетів у віртуальних з'єднаннях мережі підтримки хмарного сервісу // Системи керування, навігації та зв'язку. 2017. Вип. 2(42). С. 117–120.*

*5. **Dimitrakopoulos Y., Burnetas A.** The value of service rate flexibility in an M/M/1 queue with admission control // Transactions IISE. 2017.*

V. O. Vlasenko, K. O. Trenova

METHODOLOGY FOR CALCULATING THE DELAY TIME IN THE MANAGEMENT SYSTEM INFOCOMMUNICATION NETWORKS BASED ON THE THEORY OF MASS SERVICE

The article is devoted to the consideration of delay time calculation methods in the M/M/m service system, the main components of the model, such as information flows and the service device, are investigated, and the delay time distribution formula based on Little's law is improved. The importance of this indicator in determining the effectiveness of information communication systems and optimizing their functioning is noted.

The article provides examples of delay time calculations and considers factors that affect this indicator, such as the workload of the telecommunications system and the characteristics of the service device. In addition, possible ways to improve the system are explored, including the use of buffers and more complex models.

M/M/m is a system with m service lines, a Poisson input flow, an indicative service time distribution, and a First-Come-First-Serve (FCFS) discipline, which means that requests are served in the order in which they are received.

A detailed analysis of the M/M/m model and the study of its main components allow a deeper understanding of the processes in mass service systems. The delay time in the M/M/m system is reflected in its direct impact on the quality of service and satisfaction of users of such networks

Keywords: data transmission network; control system; load switching node; data packet; Poisson flow; mass service system; call intensity.