

МОДЕЛЮВАННЯ ОБРОБКИ ЗАПИТІВ У ІНФОРМАЦІЙНИХ СИСТЕМАХ

к.т.н. Г.А. Кучук, к.т.н. М.І. Гіневський, к.т.н. Ю.О. Акімова
(подав проф. А.В. Корольов)

Розглянуто підхід до побудови моделей обробки запитів у системах $M/G/1/\infty$.

Аналіз пріоритетних напрямків розвитку передових країн світу показує, що інформаційні системи (ІС) – це інтелектуальна основа забезпечення оборонної, економічної та наукової могутності держави. А так як при проектуванні ІС використання адекватних моделей дозволяє уникнути багатьох нерациональних витрат, то проблеми моделювання набувають великої актуальності. Сьогодні раціональне забезпечення якості інформаційних систем неможливе без застосування моделей і методів, що дозволяють оцінювати й оптимізувати процеси збору, збереження, обробки та подання інформації. Одним з базових елементів відповідних моделей є підмодель обробки запитів у інформаційних системах [1 – 6].

Розглянемо моделювання процесів обробки за допомогою пріоритетних і безпріоритетних систем масового обслуговування (СМО) $M/G/1/\infty$ [1], формалізуємо у відповідності зі схемою (рис. 1). На схемі позначено наступні типи запитів: **Z1** – на виконання технологічних операцій; **Z2** – на одержання вихідної інформації; **Z3** – на введення інформації в базу даних (БД); **Z4** – на пересилання повідомлень. У результаті виконання запиту звичайно одержують: **O1** – перелік виконаних технологічних операцій; **O2** – вихідні документи; **O3** – колекцію БД; **O4** – передані повідомлення.

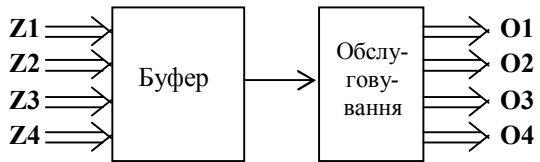


Рис. 1. Укрупнена схема формалізації обробки запитів

При обслуговуванні запитів можливі різні варіанти обробки [2]. Найбільш простий варіант (**V1**) припускає безпріоритетну однозадачну послідовну обробку (у всіх запитів однаковий пріоритет і обробляються вони за порядком надходження). У варіанті **V2** на відміну від **V1** обслуговування проводиться **N** пристроями. Варіант **V3** припускає обробку запитів з відносними пріоритетами відповідно до дисципліни FIFO без переривань обробки обслуговуваних запитів. Варіант **V4**, на відміну від **V3**, дозволяє переривати обробку. У варіанті **V5** передбачається пакетна обробка зі звичайним формуванням пакетів і дотриманням дисципліни FIFO усередині пакета.

Для всіх розглянутих варіантів обробки **V1 - V5** можна апроксимувати ймовірність своєчасної обробки запиту типу **i**, $i = \overline{1, I}$ (імовірність того, що

час реакції системи на запити i - го типу $t_{\Pi,i}$, що включає як час очікування в черзі, так і час безпосередньої обробки, не перевершує припустимого часу реакції системи $T_{Д,i}$ на розглянутий тип запиту), за допомогою неповної Γ - функції таким чином:

$$P_{co,i} = P(t_{\Pi,i} \leq T_{Д,i}) = \frac{\gamma_i T_{Д,i}}{T_{\Pi,i}^{(1)}} \int_0^{T_{Д,i}} t^{\gamma_i-1} \cdot e^{-t} dt \Big/ \int_0^{\infty} t^{\gamma_i-1} \cdot e^{-t} dt, \quad (1)$$

де $\gamma_i = \frac{T_{\Pi,i}^{(1)}}{\sqrt{(T_{\Pi,i}^{(2)})^2 - (T_{\Pi,i}^{(1)})^2}}$; $T_{\Pi,i}^{(1)}$, $T_{\Pi,i}^{(2)}$ - перший і другий моменти середньо-

го часу перебування запиту в системі.

При цьому відносна частина вчасно оброблених у системі запитів усіх типів розраховується як

$$k_{\Sigma} = \sum_{i=1}^I \lambda_i P_{co,i} / \lambda, \quad \lambda = \sum_{i=1}^I \lambda_i,$$

де λ_i – інтенсивність запитів i - го типу на обробку; λ - сумарна інтенсивність запитів ІС, а відносна частина оброблених запитів за критерієм для кожної групи з множини $\{\text{crit}_j, j = 1, J\}$ (наприклад, за середнім часом обробки $T_{\Pi,i} \leq T_{\Pi}$ чи за ймовірністю своєчасної обробки $P_{co,i} \geq P_{co}$), обраному при вході до системи, дорівнює

$$k_{cr} = \sum_{i=1}^I \lambda_i P_{co,i} \cdot \sum_{j=1}^J \text{ind}_j(i) / \lambda, \quad (2)$$

де $\text{ind}_j(i)$ - булева функція, що дорівнює одиниці тоді і тільки тоді, коли для запиту i виконано критерій j .

Розрахунок першого і другого моментів величини $T_{\Pi,i}$ розглянемо на найбільш прийнятному на практиці варіанті обробки запитів, при якому для різних груп можуть вибиратися варіанти $V1 - V5$. Для цього розіб'ємо всі можливі типи питань на n груп, причому в кожній групі з номером g пріоритети всіх запитів попадають в інтервал $[S_{g-1} + 1; S_g]$. Розташуємо групи в порядку спадання пріоритетів. Усередині кожної групи пріоритети запитів будемо розглядати як відносні. Кожній групі g призначимо свій варіант обробки запитів $D(g)$. У нашому випадку $D(g) \in V = \{V1, \dots, V5\}$, однак множина V у загальному випадку може мати істотно більшу потужність. Між запитами двох груп g і f призначаються або відносні, або абсолютні пріоритети v_{gf} .

Схематично варіант вищеописаної обробки запитів подано на рис. 2.

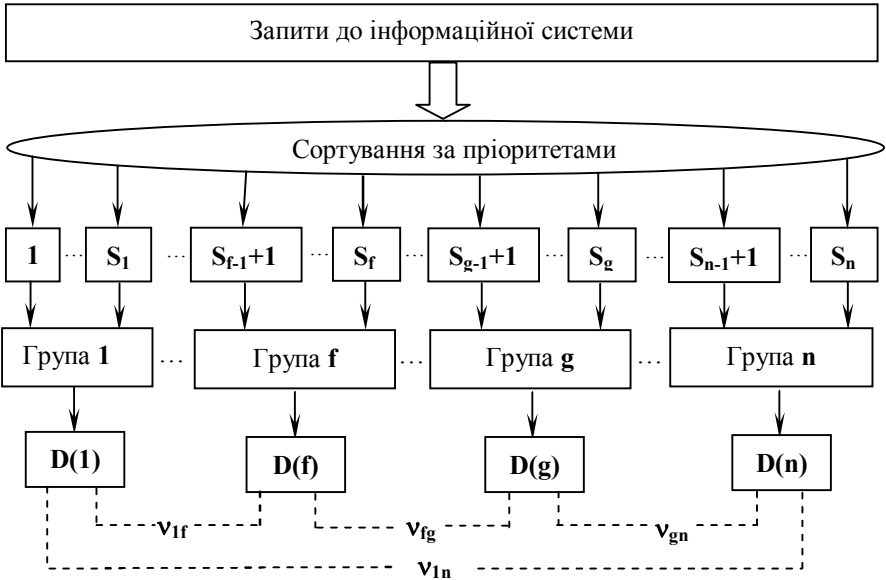


Рис. 2. Схема визначення міжгрупових пріоритетів

Позначимо очікувані перші моменти обробки запитів i - го типу у вільній системі як $\beta_i^{(1)}$ ($i = \overline{1, S_n}$; $S_n \equiv 1$). Нехай $\beta_{ig}^{(1)}$ ($g = \overline{1, n}$) - середній час обробки того ж запиту із групи g . Для типу запиту з номером пріоритету i з групи g перший і другий моменти середнього часу на обробку в прийнятих позначеннях для однозадачних режимів $D(g) \neq V2$ розраховуються так [3, 4]:

$$T_{\Pi, ig}^{(1)} = W_{ig}^{(1)} + \frac{\beta_i^{(1)}}{1 - R_{S_{g-1}}}; \quad (3)$$

$$T_{\Pi, ig}^{(2)} = W_{ig}^{(2)} + \frac{2T_{\Pi, ig}^{(1)} \cdot \beta_i^{(1)}}{1 - R_{S_{g-1}}} + \frac{\beta_i^{(2)}}{(1 - R_{S_{g-1}})^2} + \frac{\beta_i^{(1)} \cdot R_{S_{g-1}}^{(2)}}{(1 - R_{S_{g-1}})^3}, \quad (4)$$

де для варіюємих значень пріоритетів u, m ($u > m$; $u, m = \overline{1, S_n}$; $r \in (m, u)$) розрахунки моментів тимчасових витрат ресурсу IC $W_{ig}^{(1)}$, $W_{ig}^{(2)}$, завантаження системи різними комбінаціями типів запитів ρ_λ (λ - підмножина множини запитів Σ) і ймовірнісних оцінок закінчення обробки нижчої пріоритетної ланки найближчої більш пріоритетної групи $R_{S_{g-1}}$ виконуються за наступними формулами [7]:

$$\rho_{\text{mu}(f)} = \sum_{j=m}^i \lambda_j \beta_{jf} + \sum_{j=i+1}^u \lambda_j \beta_{jf} \theta_{ij}^{(g)}, \quad \text{де } \theta_{ij}^{(g)} = \begin{cases} 1, & \text{при } S_{g-1} + 1 \leq m \leq S_g; \\ 0, & \text{для інших груп;} \end{cases}$$

$$R_{S_{g-1}}^{(f)} = \sum_{j=1}^{S_{g-1}} (1 - y_{jf}) \cdot \sum_{l=S_{g-1}+1}^{S_g} \lambda_l \cdot \beta_{lf}; \quad y_{jf} \in \{0,1\}; \quad \tilde{\rho}_{\text{mu}(f)} = \rho_{\Sigma} - \rho_{\text{mu}(f)};$$

$$E_i = \rho_{S_{g-1},i}^{(2)} \cdot \tilde{\rho}_{S_{g-1}} + \rho_{S_{g-1}+1,i} \cdot \rho_{S_{g-1}}^{(1)}; \quad C_i = \frac{\left(\rho_{S_{g-1}}^{(1)} + \rho_{S_{g-1}+1,i-1} \right)^3 - \rho_{S_{g-1}+1,i}^3}{\left(\tilde{\rho}_{S_{g-1}}^{(1)} \right)^3 - \rho_{S_{g-1}+1,S_g}^3};$$

$$W_{ig}^{(1)} \stackrel{D(g) = v^3}{=} \left(\rho_{S_g}^{(1)} \right)^2 / \left(2\tilde{\rho}_{i-1}^{(1)} \cdot \tilde{\rho}_i^{(1)} \right);$$

$$W_{ig}^{(1)} \stackrel{D(g) = v^5}{=} \frac{\left(\rho_{S_g}^{(1)} \right)^2 \cdot \left(\rho_{S_{g-1}} + \rho_{S_{g-1}+1,i-1} + \rho_{S_{g-1}+1,i} \right)}{2\tilde{\rho}_{i-1}^{(1)} \cdot \tilde{\rho}_i^{(1)} \cdot \left(\tilde{\rho}_{S_{g-1}}^{(1)} + \rho_{S_{g-1}+1,S_g} \right)};$$

$$W_{ig}^{(2)} \stackrel{D(g) = v^3}{=} \frac{\frac{1}{3} \left(\rho_{S_g}^{(1)} \right)^3 \cdot \tilde{\rho}_i \cdot \tilde{\rho}_{i-1} + \frac{1}{2} \left(\rho_{S_g}^{(1)} \cdot \rho_i^{(1)} \right)^2 \cdot \tilde{\rho}_{i-1} + \frac{1}{2} \left(\rho_{S_g}^{(1)} \cdot \rho_{i-1}^{(1)} \right)^2 \cdot \tilde{\rho}_i}{2\left(\rho_{i-1}^{(1)} \right)^3 \cdot \left(\rho_i^{(1)} \right)^2};$$

$$W_{ig}^{(2)} \stackrel{D(g) = v^5}{=} \frac{\frac{1}{3} C_r \cdot B_1 + \left(\rho_{S_g}^{(1)} \right)^2 \cdot B_2 + \frac{1}{2} \left(\rho_{S_g}^{(2)} \right)^2 \cdot B_3 \cdot \lambda_i}{2\left(\tilde{\rho}_{S_{g-1}}^{(1)} \right)^3};$$

$$\text{де } B_1 = \left(\rho_{S_g}^{(1)} \right)^3 \cdot \tilde{\rho}_{S_{g-1}} + 3 \left(\rho_{g-1}^{(1)} \cdot \tilde{\rho}_{S_g}^{(1)} \right)^2;$$

$$B_2 = \frac{E_{S_g} C_i \tilde{\rho}_{S_{g-1}+1,S_g}^{(1)} + E_{i-1} \left(\tilde{\rho}_{S_{g-1}}^{(1)} + \rho_{S_{g-1}+1,S_g} \right) - E_i \cdot \rho_{S_{g-1}+1,i-1}}{\tilde{\rho}_{S_g}^{(1)} \cdot \left(\tilde{\rho}_{S_{g-1}}^{(1)} + \rho_{S_{g-1}+1,S_g} \right)};$$

$$B_3 = \frac{\left(\beta_i^{(2)} \cdot \tilde{\rho}_{S_{g-1}}^{(1)} + \beta_i^{(1)} \cdot \left(\rho_{S_{g-1}}^{(1)} \right)^2 \right) \cdot \left(\tilde{\rho}_{S_{g-1}}^{(1)} + \rho_{S_{g-1}+1,i-1} + \rho_{S_{g-1}+1,i} \right)}{\tilde{\rho}_{S_g}^{(2)} \cdot \left(\rho_{S_{g-1}}^{(1)} + \rho_{S_{g-1}+1,S_g}^{(1)} \right)}.$$

Варіанти V1 і V4 розглядаються аналогічно варіантам V3 і V5, але з більш простими умовами. Для багатозадачного режиму у варіанті V2 перший і другий моменти можна розрахувати таким чином [3]:

$$T_{n,ig}^{(1)} = \frac{\beta_{ig}^{(1)}}{\rho_{\Sigma} - \rho_g} ; \quad (5)$$

$$T_{n,ig}^{(2)} = \int_0^{\infty} \frac{2t^2}{(\rho_{\Sigma} - \rho_g)^2} dH_i(t) - \frac{2t}{\rho_{\Sigma} - \rho_g} \cdot \int_0^{\infty} dH_i(t) \cdot \int_0^{\infty} \delta(u) du , \quad (6)$$

де $H_i(t)$ – функція розподілу запитів i -го типу у вільній системі;

$$H(t) = \left(\sum_{i=1}^I \lambda_i H_i(t) \right) / \lambda ; \quad \delta(u) = u + \int_0^u (u-z) \sum_{n=1}^{\infty} \rho_g^n \cdot dF^{*n}(z) ;$$

$$F^{*(n)}(z) - n\text{-на згортка функції розподілу } F(z) = \int_0^z (1 - H(t)) dt \Big/ \int_0^z t dH(t) .$$

На основі виразів (1) – (6) побудована підмодель процесів обробки запитів у інформаційній системі, що дозволяє провести детальний аналіз завантаження системи в залежності від застосування параметрів потоку запитів. Зокрема, по різних групах і пріоритетах можна одержати середній час сподівання, середній час реакції системи на запити, ймовірність виконання обмежень за часом, відсоток вчасно оброблених запитів.

ЛІТЕРАТУРА

1. Клейнрок Л. Теория массового обслуживания. – М.: Машиностроение, 1979. – 432 с.
2. Martin I. System Analysis for Data Transmission. – Prentice Hall, 1972. – 286 p.
3. Kostogzyzov A.I. Models for Information Systems Operation Quality Investigation // Proceedings I Workshop MALOPD. – Moscow. – 1999. – p. 20-31/
4. Кучук Г.А. Метод синтезу логічної структури мережевої бази даних // Системи обробки інформації. – X. : НАНУ, ПАНМ, ХВУ, 2001. – Вип. 2(12). – С. 32-36.
5. Кучук Г.А. Формалізація предметної області багатовимірних баз даних // Системи обробки інформації. – X. : ХВУ, 2001. – Вип. 1(11). – С. 110-114.
6. Кучук Г.А. Минимизация загрузки каналов связи вычислительной сети // Системи обробки інформації. – X. : НАНУ, ПАНМ, ХВУ, 1998. – Вип. 1(5). – С. 149-154.
7. Гнеденко Б.В., Данизян Э.А., Димитров Б.Н. Приоритетные системы обслуживания. – М.: МГУ, 1973. – 448 с.

Надійшла 05.02.2002

КУЧУК Георгій Анатолійович, канд. техн. наук, ст. науч. співроб., нач. НДВ наукового центру при ХВУ. У 1977 році закінчив мехмат Московського держуніверситету. Область наукових інтересів – оптимізація інформаційних систем.

ГІНЕВСЬКИЙ Михайло Іванович, канд. техн. наук, ст. науч. співроб., нач. ІОЦ ХВУ. У 1967 році закінчив Харківське ВВКІУ. Область наукових інтересів – оптимізація інформаційних систем.

АКИМОВА Юлія Олександрівна, канд. техн. наук, доцент кафедри ХВУ. У 1996 році закінчила мехмат Харківського держуніверситету. Область наукових інтересів – оптимізація інформаційних систем.