ДИНАМИЧЕСКИЕ ПРИОРИТЕТЫ В СИСТЕМАХ МАССОВОГО ОБСЛУЖИВАНИЯ

Статья В.В. Рыкова и Э.Е. Лемберга

Легиньких Г.А.

Российский университет дружбы народов, Москва, Россия

Основная проблема

Вопрос:
Дают ли динамические приоритеты преимущество по сравнению с обычными (статическими) приоритетами?

Динамический приоритет:
Правило выбора требования зависит от текущего состояния системы

Обычный приоритет:
Фиксированное правило (например, “всегда сначала тип А”)

Математическая модель (1)

Параметры системы: 1. Один обслуживающий прибор 2. N входящих потоков требований

$$\lambda_i > 0, \quad \sum_{i=1}^{N} \lambda_i = 1$$

λ_i – интенсивность потока типа i

Нормировка: λ_i = вероятность принадлежности вновь поступившего требования к типу i

Математическая модель (2)

Время обслуживания:

$$B_i(x) = P\{\text{время обслуживания требования типа } i \leq x\}$$

Штрафы: $a_i > 0$ — потери за единицу времени ожидания требования типа i

Целевая функция:

$$L = \sum_{i=1}^{N} a_i \lambda_i v_i$$

v_i — среднее время ожидания требования типа i

Формальное определение динамических приоритетов

Пространство состояний:

$$x = (x_1, x_2, \dots, x_N), \quad x_i \geq 0$$

x_i — количество требований типа i в системе (очередь + обслуживание)

Разбиение на области управления:

$$\mathcal{E} = \{E_1, E_2, \dots, E_N\}, \quad \bigcup_{i=1}^{N} E_i = X$$

В каждой области E_i правило: **“выбирать требование типа** i**”**

Функция управления

Формальная запись правила:

$$d(x) = i, \quad \text{если } x \in E_i$$

Пример для двух потоков: - Если $(q_1, q_2) \in E_1$ => обслуживаем тип 1 - Если $(q_1, q_2) \in E_2$ => обслуживаем тип 2

Задача: найти оптимальное разбиение $\mathcal{E}$, минимизирующее L

Анализ для двух потоков ($N = 2$)

Рекуррентные соотношения:

$$\begin{cases} q_i' = \xi_{ji} & \text{с вероятностью } \lambda_j, \text{ если } t - 0\text{-момент} \\ q_i' = q_i + \xi_{ji} - \delta_{ji} & \text{если } t - j\text{-момент} \end{cases}$$

Обозначения: - ξ_{ji} — число требований типа i, пришедших за время обслуживания требования типа j - δ_{ji} — символ Кронекера ($\delta_{ji} = 1$ при $j = i$, иначе 0)

Производящие функции

Определение:

$$G_i(z_1, z_2) = M(z_1^{q_1} z_2^{q_2} \mid E_i)$$

$G_i(z_1, z_2)$ — производящая функция совместного распределения длин очередей при условии E_i

Преобразование Лапласа-Стилтьеса:

$$b_i(u) = \int_0^{\infty} e^{-ux} dB_i(x)$$

Связь:

$$M(z_1^{\xi_{j1}} z_2^{\xi_{j2}}) = b_j(\lambda_1(1 - z_1) + \lambda_2(1 - z_2))$$

Основное уравнение баланса

Ключевое уравнение:

$$\pi_0\left[1-\lambda_1 b_1(u)-\lambda_2 b_2(u)\right]+\pi_1 G_1\left(1-\frac{b_1(u)}{z_1}\right)+\pi_2 G_2\left(1-\frac{b_2(u)}{z_2}\right)=0$$

где:

$$u=\lambda_1(1-z_1)+\lambda_2(1-z_2)$$

Составляющие: - π_0 — вероятность 0-момента (прибор свободен) - π_i — вероятность i-момента (закончили обслуживание типа i)

Стационарные вероятности

Коэффициенты загрузки:

$$\rho_i = \lambda_i b_i, \quad b_i = \int_0^\infty x dB_i(x)$$

ρ_i – доля времени, когда прибор занят обслуживанием требований типа i

Общая загрузка:

$$R = \rho_1 + \rho_2$$

Вероятности:

$$\pi_0 = 1 - R, \quad \pi_1 = \lambda_1 R, \quad \pi_2 = \lambda_2 R$$

Условные математические ожидания

Определение:

$$g_{ij} = M(q_j \mid E_i)$$

g_{ij} – среднее число требований типа j при условии, что система в области E_i

Дифференцируем уравнение баланса:

Берём производные от уравнения (слайд 8) по z_1 и z_2, подставляем $z_1 = z_2 = 1$

Система уравнений для g_{ij}

$$\begin{cases} (1-\rho_1)g_{11} - \rho_2 g_{21} = \lambda_1 C + (1-\rho_1) \\ -\rho_1 g_{12} + (1-\rho_2)g_{22} = \lambda_2 C + (1-\rho_2) \end{cases}$$

Константа C:

$$C = \frac{b^{(2)}}{2R}, \quad b^{(2)} = \lambda_1 b_1^{(2)} + \lambda_2 b_2^{(2)}$$

Вторые моменты:

$$b_i^{(2)} = \int_0^\infty x^2 dB_i(x)$$

Среднее время ожидания

Связь v_i и g_{ii}:

$$v_i = R\frac{g_{ii} - 1}{\lambda_i}$$

Вывод: 1. По формуле Литтла: $Mq_i = \lambda_i v_i$ 2. Раскладываем по полной вероятности:

$$Mq_i = \pi_0 M(q_i \mid E_0) + \pi_1 M(q_i \mid E_1) + \pi_2 M(q_i \mid E_2)$$

3. $M(q_i \mid E_i) = g_{ii}$ включает обслуживаемое требование 4. Среднее число в очереди: $g_{ii} - 1$

Фундаментальное ограничение

Важнейшее равенство:

$$\rho_1 v_1 + \rho_2 v_2 = \frac{CR^2}{1-R}$$

Выполняется для ЛЮБОГО разбиения $\mathcal{E}$!

Следствия: 1. Нельзя одновременно уменьшить v_1 и v_2 2. Уменьшение v_1 ведёт к увеличению v_2, и наоборот 3. $\rho_1 v_1 + \rho_2 v_2$ — инвариант системы

Обычные приоритеты (тип 1 высший)

Условие: $g_{12} = 0$

(Если обслуживаем тип 1, в очереди нет требований типа 2)

Решение системы:

$$g_{11} = \frac{\rho_1 C}{1 - \rho_1}$$

Время ожидания:

$$v_1^{(1)} = \frac{CR}{1 - \rho_1}, \quad v_2^{(1)} = \frac{CR}{(1 - \rho_1)(1 - R)}$$

Обычные приоритеты (тип 2 высший)

Условие: $g_{21} = 0$

(Если обслуживаем тип 2, в очереди нет требований типа 1)

Время ожидания:

$$v_1^{(2)} = \frac{CR}{(1-\rho_2)(1-R)}, \quad v_2^{(2)} = \frac{CR}{1-\rho_2}$$

Область достижимых значений:

$$v_1^{(1)} \le v_1 \le v_1^{(2)}, \quad v_2^{(2)} \le v_2 \le v_2^{(1)}$$

Минимизация потерь (шаг 1)

Выражаем v_2 через v_1: Из $\rho_1 v_1 + \rho_2 v_2 = \frac{CR^2}{1-R}$:

$$v_2 = \frac{1}{\rho_2}\left(\frac{CR^2}{1-R} - \rho_1 v_1\right)$$

Подставляем в L:

$$L = a_1\lambda_1 v_1 + a_2\lambda_2 \cdot \frac{1}{\rho_2}\left(\frac{CR^2}{1-R} - \rho_1 v_1\right)$$

Минимизация потерь (шаг 2)

Упрощаем:

$$L = \left(a_1\lambda_1 - a_2\lambda_2\frac{\rho_1}{\rho_2}\right)v_1 + \frac{a_2\lambda_2 CR^2}{\rho_2(1-R)}$$

Так как $\rho_i = \lambda_i b_i$**:**

$$a_1\lambda_1 - a_2\lambda_2\frac{\rho_1}{\rho_2} = a_1\lambda_1 - a_2\lambda_2\frac{\lambda_1 b_1}{\lambda_2 b_2}$$

Минимизация потерь (шаг 3)

Преобразуем коэффициент:

$$a_1\lambda_1 - a_2\lambda_2\frac{\lambda_1 b_1}{\lambda_2 b_2} = \lambda_1\left(a_1 - a_2\frac{b_1}{b_2}\right)$$

Домножаем и делим на b_1/b_2:

$$= \lambda_1\frac{b_1}{b_2}\left(\frac{a_1}{b_1} - \frac{a_2}{b_2}\right)$$

Так как $\lambda_1\frac{b_1}{b_2} > 0$, знак коэффициента определяется знаком:

$$\frac{a_1}{b_1} - \frac{a_2}{b_2}$$

Оптимальное правило для N=2

Если $\frac{a_1}{b_1} \geq \frac{a_2}{b_2}$**:** Коэффициент при $v_1 \geq 0$ => минимизируем v_1 => $v_1^{\text{min}} = v_1^{(1)}$ (приоритет типу 1)

Если $\frac{a_1}{b_1} \leq \frac{a_2}{b_2}$**:** Коэффициент при $v_1 \leq 0$ => максимизируем v_1 => $v_1^{\text{max}} = v_1^{(2)}$ (приоритет типу 2)

Итог: Даём приоритет типу с бóльшим отношением $\frac{a_i}{b_i}$

Обобщение на N потоков

Общая система уравнений:

Для любых k, l $(k < l)$:

$$\lambda_l \sum_{i=1}^{N} \rho_i g_{ik} + \lambda_k \sum_{i=1}^{N} \rho_i g_{il} - \lambda_k g_{lk} - \lambda_l g_{kl} = \lambda_l \rho_k + \lambda_k \rho_l - \lambda_k \lambda_l \frac{b^{(2)}}{R}$$

Для каждого l:

$$\sum_{i=1}^{N} \rho_i g_{il} - g_{ll} = -(1 - \rho_l) - \lambda_l \frac{b^{(2)}}{2R}$$

Множество достижимых значений

Определение:

$$V = \{v(\mathcal{E}) = (v_1(\mathcal{E}), \dots, v_N(\mathcal{E})) \mid \mathcal{E} - \text{разбиение}\}$$

где

$$v_i(\mathcal{E}) = R\frac{g_{ii} - 1}{\lambda_i}$$

Свойства: 1. $V \subset \mathbb{R}^N$ 2. V — выпуклое множество 3. Крайние точки V соответствуют обычным приоритетам

Главная теорема

Теорема: В классе систем с динамическими приоритетами **оптимальной является система с обыкновенными приоритетами**.

Алгоритм определения оптимальных приоритетов:

1. Для каждого типа i вычислить:

$$r_i = \frac{a_i}{b_i}$$

2. Упорядочить типы по убыванию r_i:

$$r_{i_1} \geq r_{i_2} \geq \cdots \geq r_{i_N}$$

3. Назначить приоритеты: тип i_1 — высший, ..., тип i_N — низший

Практический пример — call-центр

Данные:

Тип звонка	Описание	λ_i (зв./час)	b_i (мин)	a_i (руб/мин)
1	VIP-клиенты	2	5	100
2	Постоянные клиенты	4	8	40
3	Новые клиенты	6	10	10

Практический пример — call-центр

Вычисляем отношения:

$$\frac{a_1}{b_1} = \frac{100}{5} = 20, \quad \frac{a_2}{b_2} = \frac{40}{8} = 5, \quad \frac{a_3}{b_3} = \frac{10}{10} = 1$$

Оптимальный порядок: VIP => Постоянные => Новые

Время ожидания в call-центре

Для типа с приоритетом k:

$$v_{i_k} = \frac{CR}{(1 - \sum_{j=1}^{k-1} \rho_{i_j})(1 - \sum_{j=1}^{k} \rho_{i_j})}$$

Результаты: - VIP ($k = 1$): $v_1 = 8.2$ мин - Постоянные ($k = 2$): $v_2 = 12.7$ мин - Новые ($k = 3$): $v_3 = 77.4$ мин (≈ 1 час 17 мин)

Экономические потери

Потери при оптимальной стратегии:

$$L_{\text{опт}} = \sum a_i \lambda_i v_i = 8319.6 \text{ руб/час}$$

Сравнение: - FIFO: $L_{\text{FIFO}} = 21132$ руб/час - Обратные приоритеты: $L_{\text{обр}} = 17993.4$ руб/час

Экономия: - От FIFO: **60.6%** (12812.4 руб/час) - От обратных приоритетов: **53.8%** (9673.8 руб/час)

Выводы

Основные результаты:

1. Динамические приоритеты **не дают преимущества** перед обычными
2. Оптимальное правило — обычные приоритеты с упорядочением по a_i/b_i
3. Критерий $\frac{a_i}{b_i}$ имеет ясную экономическую интерпретацию

Практическая значимость:

- Простые правила оптимальны
- Универсальный алгоритм для любых систем
- Применимо в телекоммуникациях, медицине, ИТ