

УДК 004.01:006.72 (470.22)

МАТЕМАТИЧЕСКИЕ МОДЕЛИ И АЛГОРИТМЫ ОПТИМАЛЬНОГО УПРАВЛЕНИЯ FIFO-ОЧЕРЕДЯМИ В ОБЩЕЙ ПАМЯТИ

А. В. Соколов¹, А. М. Сазонов², Е. В. Морозов¹, Р. С. Некрасова¹,
Р. В. Разумчик³

¹ Институт прикладных математических исследований Карельского научного центра РАН

² Петрозаводский государственный университет

³ Институт проблем информатики ФИЦ ИУ РАН

В статье рассматриваются задачи оптимального распределения памяти для двух и произвольного числа очередей в общей памяти в дискретном времени. Приведены математические модели и алгоритмы оптимального распределения памяти между очередями, минимизирующие потери информации.

Ключевые слова: FIFO-очереди; теория массового обслуживания; дискретное время; оптимальное разделение памяти.

**A. V. Sokolov, A. M. Sazonov, E. V. Morozov, R. S. Nekrasova,
R. V. Razumchik. MATHEMATICAL MODELS AND
ALGORITHMS FOR THE OPTIMAL CONTROL OF FIFO-
QUEUES IN SHARED MEMORY**

The paper deals with the problems of optimal partition of shared memory for two and arbitrary number of queues in discrete time. Mathematical models and algorithms for the optimal partition of memory between queues, minimizing data loss, are presented.

Key words: FIFO-queues; queuing theory; discrete time; optimal partition of the shared memory.

ВВЕДЕНИЕ

Рассмотрим ресурс (например, внутреннюю память ПК, которая обладает свойством дискретности) конечной емкости $n \in \mathbb{N}$. Каждая единица емкости может быть в одном из двух состояний – свободна или занята. Будем считать, что время изменяется в фиксированных интервалах (слотах). Величину слота примем равной $h > 0$, единиц времени, и будем считать, что возможные изменения свободной емкости системы происходят случайным обра-

зом на величину ± 1 в моменты времени nh , $n = 1, 2, \dots$. Ресурс одновременно используется несколькими независимо и параллельно работающими системами, с целью организации в нем собственных очередей. Ввиду того, что емкость ресурса ограничена, очереди могут переполняться и в системах могут возникать потери. В данной работе речь пойдет об одной из задач, связанных с нахождением оптимального (с точки зрения минимума потерь) статического распределения емкости ресурса меж-

ду всеми системами, которые этот ресурс могут использовать. Это означает, что оптимальное разделение памяти производится один раз и может меняться только в случае значительного изменения вероятностных характеристик структур данных.

Во многих приложениях требуется работа с несколькими FIFO-очередями, расположенными в общем пространстве памяти. Для этого применяют различные программные или аппаратные решения [1–3]. В ряде работ построены математические модели работы с различными динамическими структурами данных в виде случайных блужданий по целочисленной решетке. Первоначально такие модели в виде случайного блуждания в треугольнике [4–9] были построены для решения задачи анализа процесса работы с двумя стеками, растущими навстречу друг другу, поставленной в [1]. В этих моделях предполагается, что на каждом шаге дискретного времени с заданными вероятностями происходят некоторые операции с очередями. Время выполнения этих операций — не случайная величина, а константа, поэтому фиксированным является и шаг времени. В [10–14] рассмотрены математические модели и алгоритмы оптимального управления FIFO-очередями в общей памяти. В данной статье мы рассмотрим задачу оптимального разделения общей памяти для нескольких FIFO-очереди в случае их последовательного циклического представления. Операции включения и исключения элементов в очереди выполняются за время $O(1)$, т. е. за фиксированное время при любом числе элементов. В качестве критерия оптимальности рассматривается минимальная доля потерянных элементов при бесконечном времени работы очереди. Эту величину разумно минимизировать, когда переполнение очереди является не аварийной, а стандартной ситуацией. То есть если очередь занимает всю предоставленную ей память, все последующие элементы, поступающие в нее, отбрасываются до тех пор, пока не появится свободная память (т. е. до тех пор, пока не произойдет исключение элемента из очереди). Такая схема применяется, например, в работе сетевых маршрутизаторов [3] в том случае, когда по мере увеличения трафика очередь маршрутизатора заполняется. Такое поведение маршрутизатора называется «сбросом хвоста». Потери пакетов приводят к снижению качества обслуживания, поэтому процент потерь, как правило, нужно держать на низком уровне.

Аналогичные постановки уже встречались в области моделирования узлов связи с помо-

щью систем/сетей обслуживания, функционирующих в непрерывном времени. Например, в различных постановках решались задачи (статического/ динамического/ полудинамического) распределения имеющейся свободной емкости $C \in \mathbf{N}$ между несколькими очередями таким образом, чтобы значение некоторого функционала было оптимальным.

Среди близких по постановке работ можно отметить [15, 16, 19–21], где прослеживается аналогия с рассматриваемой задачей (но в непрерывном времени). Так, например, в работе [19] рассматривается аналогичная система, но функционирующая в непрерывном времени, и для нее ставится (и в некоторых случаях решается) задача нахождения такого распределения свободной емкости, которое позволяет достичь заданных вероятностей потерь. Система также упоминается и в [20], но в этой работе внимание уделяется в основном численным экспериментам.

Неоптимальность (с точки зрения потерь) статического распределения свободной емкости обсуждается во многих работах. Чаще всего преимущество отдается полудинамическому распределению (см, например, [20]).

В [16], в более общей постановке, рассматривается случай, когда заявки могут иметь различный целочисленный объем (распределенный по геометрическому закону).

Алгоритмы (целочисленного) поиска оптимального (с точки зрения пропускной способности) распределения свободной емкости в сети массового обслуживания произвольной структуры (но состоящей из узлов типа $M/M/1$) обсуждаются в [17]. В работе упоминается, что решение задачи оптимизации затруднено тем, что для таких систем не существует явных формул для стационарного распределения вероятностей состояний. Одним из вариантов решения является применение эвристических алгоритмов. Однако чаще всего встречается прием, когда каждая система рассматривается независимо от остальных. Здесь подходы отличаются друг от друга тем, каким образом оцениваются исходные параметры каждой системы (интенсивность входящего потока и обслуживания). Подробнее с приближенными методами авторы предлагают ознакомиться в [18].

Применение методов теории очередей к описанной задаче в дискретном времени найти не удалось, однако, как будет показано, данный подход является весьма естественным. В [1, параграф 2.2.2] Д. Кнут писал: «Провести математический анализ алгоритмов динамического распределения памяти чрезвычай-

но трудно... Не исключено, что для такого исследования можно применить некоторые математические результаты теории очередей, однако эта теория имеет, по-видимому, другую направленность». Таким образом, данная статья показывает, что Д. Кнут был не вполне прав и можно применять теорию массового обслуживания к анализу данной задачи в дискретном времени.

ПОСТАНОВКА ЗАДАЧИ ДЛЯ ДВУХ ОЧЕРЕДЕЙ

Рассмотрим систему обслуживания с буфером заданного размера n , разделенного между двумя очередями. Заявки из каждой очереди обрабатываются обслуживающим устройством (сервером). Рассматривается дискретная шкала времени, разделенного на слоты. В каждый момент времени (слот) возможно одно из следующих событий:

- приход заявки в очередь 1 с вероятностью p_1 ,
- уход из очереди 1 после завершения обслуживания с вероятностью q_1 ,
- приход заявки в очередь 2 с вероятностью p_2 ,
- уход из очереди 2 после завершения обслуживания с вероятностью q_2 ,
- отсутствие приходов и уходов с вероятностью r .

Кроме того, предполагается, что $p_1 + q_1 + p_2 + q_2 + r = 1$. Таким образом, появление двух событий в одном слоте невозможно. В этом состоит важное отличие данной постановки от классической очереди, где при фиксированной длине слота вероятность прихода и ухода (в течение слота), вообще говоря, является положительной. Эта вероятность оказывается равной нулю лишь при рассмотрении системы в непрерывном времени, при построении соответствующих уравнений Колмогорова для основного марковского процесса.

Итак, требуется найти такое значение k размера очереди 1 (тогда размер очереди 2 равен $n - k$), при котором вероятности потери заявок для обеих очередей будут минимальны.

АНАЛИЗ СТАЦИОНАРНОГО РЕЖИМА

Пусть целочисленные случайные величины (с. в.) $\tau^{(i)}$, $S^{(i)}$ есть типичный интервал между приходами заявок и типичное время обслуживания в системе $i = 1, 2$ соответственно. Легко

получить, что с. в. $\tau^{(i)}$ является геометрической,

$$P(\tau^{(i)} \geq j) = (1 - p_i)^j, j \geq 0 \quad (1)$$

со средней длиной $E\tau^{(i)} = \frac{1}{p_i}$ и интенсивностью прихода $\lambda_i = p_i, i = 1, 2$. Аналогично находим, что периоды обслуживания являются геометрическими с интенсивностями $\mu_i = 1/ES^{(i)} = q_i, i = 1, 2$. Таким образом, коэффициент загрузки системы i равен:

$$\rho_i = \frac{p_i}{q_i}, i = 1, 2. \quad (2)$$

Определим $r_1 = 1 - (p_1 + q_1) = p_2 + q_2 + r$ — вероятность того, что в системе 1 за 1 слот времени не было событий. Пусть ν_t есть число заявок в системе 1 в момент (слот) $t = 0, 1, \dots$. Процесс $\{\nu_t\}$ является цепью Маркова с множеством состояний $\{0, 1, \dots, k\}$, матрицей переходных вероятностей

$$P = \begin{pmatrix} 1 - p_1 & p_1 & 0 & \dots & 0 & 0 & 0 \\ q_1 & r_1 & p_1 & 0 & \dots & 0 & 0 \\ 0 & q_1 & r_1 & p_1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & 0 & q_1 & r_1 & p_1 \\ 0 & \dots & 0 & 0 & 0 & q_1 & 1 - q_1 \end{pmatrix}$$

и стационарным распределением

$$P_j = \lim_{t \rightarrow \infty} P(\nu_t = j), j = 0, 1, \dots, k. \quad (3)$$

В результате мы стандартным образом получаем такую систему уравнений для стационарного распределения числа заявок в системе 1

$$P_0 = P_0 \cdot (1 - p_1) + P_1 \cdot q_1, \quad (4)$$

$$P_j = P_{j-1} \cdot p_1 + P_j \cdot r_1 + P_{j+1} \cdot q_1, \quad j = 1, \dots, k - 1 \quad (5)$$

$$P_k = P_{k-1} \cdot p_1 + P_k \cdot (1 - q_1), \quad (6)$$

решение которой с условием нормировки $\sum_{j=0}^k P_j = 1$ имеет вид

$$P_j = P_0 \cdot \rho_1^j = \frac{(1 - \rho_1) \cdot \rho_1^j}{1 - \rho_1^{k+1}}, j = 0, \dots, k, \quad (7)$$

где при $\rho_1 \neq 1$ (т. е. при $p_1 \neq q_1$)

$$P_0 = \frac{1 - \rho_1}{1 - \rho_1^{k+1}} \quad (8)$$

и $P_0 = \frac{1}{k+1}$ при $p_1 = q_1$.

В данной модели за вероятность потери принимается непосредственно потеря заявки

(т. е. приход новой заявки в полностью загруженную систему). Поскольку моменты приходов не зависят от числа заявок в очереди, то

$$P_{loss}^1 = P_k \cdot p_1. \quad (9)$$

Таким образом, вероятность потери в системе 1 при $\rho_1 \neq 1$ есть

$$P_{loss}^1 = P_{k+1} = P_k \cdot p_1 = \frac{(1 - \rho_1) \cdot \rho_1^k \cdot p_1}{1 - \rho_1^{k+1}}. \quad (10)$$

Аналогично вероятность потери в системе 2 при $\rho_2 \neq 1$ есть

$$P_{loss}^2 = \frac{(1 - \rho_2) \cdot \rho_2^{n-k} \cdot p_2}{1 - \rho_2^{n-k+1}}. \quad (11)$$

При $\rho_1 = 1$ вероятность потери в системе 1 равна

$$P_{loss}^1 = \frac{p_1}{k+1}, \quad (12)$$

и в системе 2, при $\rho_2 = 1$:

$$P_{loss}^2 = \frac{p_2}{n-k+1}. \quad (13)$$

ОПТИМАЛЬНОЕ РАЗБИЕНИЕ

Далее задача сводится к минимизации функции

$$\varphi(k) = P_{loss}^1 + P_{loss}^2. \quad (14)$$

Рассмотрим случай равных вероятностей: $p_i = q_i, i = 1, 2$. В данном случае функция $\varphi(k)$ имеет вид:

$$\varphi(k) = \frac{p_1}{k+1} + \frac{p_2}{n-k+1}. \quad (15)$$

Перейдем к непрерывной функции $\varphi(x)$, заданной на $[0, n]$.

Найдем ее минимум, для чего решим уравнение

$$\frac{\partial \varphi(x)}{\partial x} = -\frac{p_1}{(x+1)^2} + \frac{p_2}{(n-x+1)^2} = 0. \quad (16)$$

Получим

$$x^* = \frac{\sqrt{p_1} \cdot (n+1) - \sqrt{p_2}}{\sqrt{p_1} + \sqrt{p_2}}. \quad (17)$$

Поскольку

$$\frac{\partial^2 \varphi(x)}{\partial x^2} = \frac{2p_1}{(x+1)^3} + \frac{2p_2}{(n-x+1)^3} > 0, \quad (18)$$

то x^* есть точка минимума. Теперь достаточно сравнить значения $\varphi(\lfloor x^* \rfloor)$ и $\varphi(\lceil x^* \rceil)$ и выбрать наименьшее. Существование двух решений означает наличие двух оптимальных разбиений памяти, отличающихся на одну единицу. В случае $x^* \leq 0$ можно либо выделить всю

память очереди 2, тем самым не сохраняя очередь 1, либо выделить очереди 2 $n-1$ и оставить очередь 1 размера 1. Случай $x^* \geq n$ аналогичен предыдущему с точностью до перестановки очередей.

Рассмотрим случай $p_i \neq q_i, i = 1, 2$. В данном случае

$$\varphi(k) = \frac{(1 - \rho_1) \cdot \rho_1^k \cdot p_1}{1 - \rho_1^{k+1}} + \frac{(1 - \rho_2) \cdot \rho_2^{n-k} \cdot p_2}{1 - \rho_2^{n-k+1}}. \quad (19)$$

Перейдем к непрерывной функции $\varphi(x)$, заданной на $[0, n]$.

Теорема 1. *Функция $\varphi(x)$ имеет единственный минимум, и этот минимум является глобальным.*

Доказательство. Для того чтобы доказать существование единственного минимума, являющегося глобальным, достаточно доказать, что функция $\varphi(x)$ выпуклая. Дифференцируем функцию $\varphi(x)$ дважды, находим

$$\frac{\partial^2 \varphi(x)}{\partial x^2} =$$

$$= p_1 \cdot (1 - \rho_1) \cdot \ln^2 \rho_1 \cdot \frac{\rho_1^x \cdot (1 + \rho_1^{x+1})}{(1 - \rho_1^{x+1})^3} + \quad (20)$$

$$+ p_2 \cdot (1 - \rho_2) \cdot \ln^2 \rho_2 \cdot \frac{\rho_2^{n-x} \cdot (1 + \rho_2^{n-x+1})}{(1 - \rho_2^{n-x+1})^3}.$$

Так как $(1 - \rho_1)$ и $(1 - \rho_1^{x+1})^3$, $(1 - \rho_2)$ и $(1 - \rho_2^{n-x+1})^3$ имеют одинаковый знак, а остальные множители в каждом из слагаемых положительны, то $\frac{\partial^2 \varphi(x)}{\partial x^2} > 0$.

Таким образом, функция $\varphi(x)$ выпуклая и имеет единственный глобальный минимум. \square

Обозначим точку минимума функции $\varphi(x)$ через x_{min} . Поскольку функция $\varphi(x)$ — выпуклая, то рассматриваемая на множестве целых чисел функция $\varphi(k)$ может иметь не больше двух точек минимума — ближайшие целые точки к x_{min} . Таким образом, достаточно сравнить значения $\varphi(\lfloor x_{min} \rfloor)$ и $\varphi(\lceil x_{min} \rceil)$ и выбрать наименьшее.

Замечание 1. В случае равенства только одной пары $p_i = q_i$ функция φ также является выпуклой, как доказано выше (см. (18), (20)). Таким образом, функция будет иметь единственный глобальный минимум.

На основании выпуклости функции $\varphi(k)$ составим алгоритм для нахождения оптимального разбиения. Данный алгоритм по сути является перебором, заканчивающимся по условию. Если на текущем шаге i алгоритма $\varphi(i) < \varphi(i-1)$ и $\varphi(i) < \varphi(i+1)$, то значение $k = i$ является оптимальным решением. Таким образом, достаточно рассмотреть не все значения функции $\varphi(k)$, а только их часть, расположенную по одной ветви.

ПОСТАНОВКА ЗАДАЧИ ДЛЯ ПРОИЗВОЛЬНОГО ЧИСЛА ОЧЕРЕДЕЙ

Рассматривается система из m очередей (систем) с общим ресурсом размера n . Заявки из каждой очереди обрабатываются одним обслуживающим устройством (сервером). Рассматривается дискретная шкала времени, разделенного на слоты. В каждый момент времени (слот) возможно одно из следующих событий:

- приход заявки в очередь i с вероятностью p_i
- уход из очереди i после завершения обслуживания с вероятностью q_i
- отсутствие приходов и уходов с вероятностью r .

Кроме того, предполагается, что $p_1 + q_1 + \dots + p_m + q_m + r = 1$. Таким образом, появление двух и более событий в одном слоте невозможно. Пусть k_1, \dots, k_{m-1} — максимальные размеры очереди 1, ..., очереди $m-1$ соответственно. Тогда максимальный размер очереди m равен $n - \sum_{i=1}^{m-1} k_i$. Требуется найти такие значения $k_i, i = 1, \dots, m-1$, чтобы потери заявок были минимальны. Обозначим $k_m = \sum_{i=1}^{m-1} k_i$ для удобства записи.

АНАЛИЗ СТАЦИОНАРНОГО РЕЖИМА

Для вывода уравнений для данной модели можно провести аналогичные рассуждения как в случае двух очередей. Пусть все $p_i \neq q_i$. Тогда для $i = 1, \dots, m-1$

$$P_j^i = \frac{(1 - \rho_i) \cdot \rho_i^j}{1 - \rho_i^{k_i+1}}, j = 0, \dots, k_i, \quad (21)$$

и для $i = m$:

$$P_j^m = \frac{(1 - \rho_m) \cdot \rho_m^j}{1 - \rho_m^{n-k_m+1}}, j = 0, \dots, n - k_m. \quad (22)$$

Вероятность потери в системе $i, i < m$

$$P_{loss}^i = P_{k_i}^i \cdot p_i = \frac{(1 - \rho_i) \cdot \rho_i^{k_i} \cdot p_i}{1 - \rho_i^{k_i+1}}. \quad (23)$$

Вероятность потери в системе m :

$$\begin{aligned} P_{loss}^m &= P_{n-k_m}^m \cdot p_m = \\ &= \frac{(1 - \rho_m) \cdot \rho_m^{n-k_m} \cdot p_m}{1 - \rho_m^{n-k_m+1}}. \end{aligned} \quad (24)$$

При $p_i = q_i$

$$P_{loss}^i = \frac{p_i}{k_i + 1}, i = 1, \dots, m-1, \quad (25)$$

и

$$P_{loss}^m = \frac{p_m}{n - k_m + 1}. \quad (26)$$

Дальнейшие результаты совпадают с результатами, полученными в работах [11] и [13].

ОПТИМАЛЬНОЕ РАЗБИЕНИЕ

Как и выше, будем минимизировать функцию:

$$\begin{aligned} \varphi(k) &= \sum_{i=1}^m P_{loss}^i = \\ &= \sum_{i=1}^{m-1} \frac{(1 - \rho_i) \cdot \rho_i^{k_i} \cdot p_i}{1 - \rho_i^{k_i+1}} + \\ &+ \frac{(1 - \rho_m) \cdot \rho_m^{n-k_m} \cdot p_m}{1 - \rho_m^{n-k_m+1}}, \text{ где } k = (k_1, \dots, k_{m-1}). \end{aligned} \quad (27)$$

В случае $p_i = q_i$ i -е слагаемое имеет вид: $\frac{p_i}{k_i+1}$. В случае $p_m = q_m$ последнее слагаемое имеет вид: $\frac{p_m}{n-k_m+1}$.

Рассмотрим случай в (27) $p_i = q_i, i = 1, \dots, m$. Тогда функция имеет вид

$$\varphi(k) = \sum_{i=1}^{m-1} \frac{p_i}{k_i + 1} + \frac{p_m}{n - k_m + 1}. \quad (28)$$

Перейдем к непрерывной функции $\varphi(x)$, заданной на R^{m-1} .

Чтобы найти точку минимума x^* , получаем систему уравнений

$$\begin{aligned} \frac{\partial \varphi(x)}{\partial x_i} &= -\frac{p_i}{(x_i + 1)^2} + \\ &+ \frac{p_m}{(n - \sum_{j=1}^{m-1} x_j + 1)^2} = 0, \end{aligned} \quad (29)$$

решение которой имеет вид:

$$x_i^* = \frac{(n + m) \sqrt{p_i}}{\sum_{j=1}^m \sqrt{p_j}} - 1. \quad (30)$$

Докажем, что x^* — точка минимума. Найдем частные производные 2-го порядка:

$$\frac{\partial^2 \varphi(x)}{\partial x_i \partial x_j} = \frac{2p_m}{\left(n - \sum_{l=1}^{m-1} x_l + 1\right)^3} = A, \quad (31)$$

$$\frac{\partial^2 \varphi(x)}{\partial x_i^2} = \frac{2p_i}{(x_i + 1)^3} + \frac{2p_m}{\left(n - \sum_{l=1}^{m-1} x_l + 1\right)^3} = B_i + A. \quad (32)$$

Так как $0 < x_i$ и $\sum_{i=1}^{m-1} x_i < n$, то $A > 0$, $B_i > 0$. Для того чтобы доказать, что x^* — точка ми-

нимума, нужно, чтобы матрица

$$\varphi'' = \begin{pmatrix} A + B_1 & A & \dots & A \\ A & A + B_2 & \dots & A \\ \dots & \dots & \dots & \dots \\ A & A & \dots & A + B_m \end{pmatrix} \quad (33)$$

была положительно определена, т. е. по критерию Сильвестра все ее угловые миноры должны быть положительны.

Докажем это методом математической индукции.

- База ($i = 1$): $\varphi''_1 = A + B_1 > 0$.
- Пусть $\varphi''_i > 0$. Докажем, что $\varphi''_{i+1} > 0$.

$$\begin{aligned} \varphi''_{i+1} &= \begin{vmatrix} A + B_1 & A & \dots & A & A \\ A & A + B_2 & \dots & A & A \\ \dots & \dots & \dots & \dots & \dots \\ A & A & \dots & A + B_i & A \\ A & A & \dots & A & A + B_{i+1} \end{vmatrix} = \\ &= \begin{vmatrix} A + B_1 & A & \dots & A & A \\ A & A + B_2 & \dots & A & A \\ \dots & \dots & \dots & \dots & \dots \\ A & A & \dots & A + B_i & A \\ A & A & \dots & A & A \end{vmatrix} + \begin{vmatrix} A + B_1 & A & \dots & A & 0 \\ A & A + B_2 & \dots & A & 0 \\ \dots & \dots & \dots & \dots & \dots \\ A & A & \dots & A + B_i & 0 \\ A & A & \dots & A & B_{i+1} \end{vmatrix} = \\ &= \begin{vmatrix} B_1 & 0 & \dots & 0 & A \\ 0 & B_2 & \dots & 0 & A \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & B_i & A \\ 0 & 0 & \dots & 0 & A \end{vmatrix} + \begin{vmatrix} A + B_1 & A & \dots & A & 0 \\ A & A + B_2 & \dots & A & 0 \\ \dots & \dots & \dots & \dots & \dots \\ A & A & \dots & A + B_i & 0 \\ A & A & \dots & A & B_{i+1} \end{vmatrix} = \\ &= AB_1 \dots B_i + B_{i+1} \varphi''_i > 0 \end{aligned} \quad (34)$$

Таким образом, x^* есть точка минимума функции $\varphi(x)$.

Перейдем к целочисленному решению. Рассмотрим функцию $\psi(x_i) = \varphi(x_1^*, \dots, x_i, \dots, x_{m-1}^*)$ при фиксированном i . Тогда

$$\psi'(x_i) = \frac{\partial \varphi(x)}{\partial x_i} = -\frac{p_i}{(x_i + 1)^2} + \frac{p_m}{\left(n - \sum_{j=1}^{m-1} x_j + 1\right)^2}. \quad (35)$$

Как доказано ранее, $x_i^* = \frac{(n+m)\sqrt{p_i}}{\sum_{j=1}^m \sqrt{p_j}} - 1$ — един-

ственная точка минимума функции $\psi(x_i)$ на $[0, n]$. На промежутке $[0, x_i^*)$ функция $\psi(x_i)$ убывает, поэтому

$$\psi(\lfloor x_i^* \rfloor) < \psi(k_i) \forall k_i = 0, \dots, \lfloor x_i^* \rfloor - 1. \quad (36)$$

На промежутке $(x_i^*, n]$ функция $\psi(x_i)$ возрастает, поэтому

$$\psi(\lceil x_i^* \rceil) < \psi(k_i) \forall k_i = \lceil x_i^* \rceil + 1, \dots, n. \quad (37)$$

Авторами предложен следующий алгоритм разбиения памяти: округлить вниз значения x_i^* и распределить оставшуюся память (размера меньше m) между очередями по методу, описанному далее. Можно показать, что $x_i^* \leq n + 1$. Так как значения x_i^* округляются вниз, то $k_i^* \leq n$. В случае, если несколько $k_i^* = n$, можно решать задачу только для этих очередей, выделив остальным либо по 1, либо 0.

Рассмотрим общий случай, когда есть очереди, для которых $p_i \neq q_i$. Обозначим $k_m = n - \sum_{i=1}^{m-1} k_i$ размер m -й очереди. Как указано

выше, функция φ будет иметь вид:

$$\varphi(k) = \sum_{i=1}^m P_{loss}^i = \sum_{i=1}^m \frac{(1 - \rho_i) \cdot \rho_i^{k_i} \cdot p_i}{1 - \rho_i^{k_i+1}}, \quad (38)$$

где $k = (k_1, \dots, k_m)$.

В случае $p_i = q_i$ i -е слагаемое имеет вид $\frac{p_i}{k_i+1}$.

Для решения данной задачи используем алгоритм, предложенный в [13].

Будем пошагово увеличивать выделяемую очередям память от 0 до n на 1, так чтобы на каждом шаге минимизировать вероятность потери. Будем увеличивать на 1 размер очереди, для которой значение $d = (P_{loss}^i(k_i) - P_{loss}^i(k_i + 1))$ максимально, т. е. при увеличении размера данной очереди достигается наибольшее уменьшение вероятности потери.

Покажем, что данный алгоритм применим для решения данной задачи, т. е. приводит к оптимальному решению.

Обозначим $\varphi_n(k)$ вероятность потери при оптимальном разбиении памяти размера n между m очередями.

$$\begin{aligned} \varphi_n(k) &= \min_{k_1+\dots+k_m=n} \sum_{i=1}^m P_{loss}^i = \quad (39) \\ &= \min_{k_1+\dots+k_m=n} \sum_{i=1}^m \frac{(1 - \rho_i) \cdot \rho_i^{k_i} \cdot p_i}{1 - \rho_i^{k_i+1}}. \end{aligned}$$

Рассмотрим рекуррентную формулу:

$$\begin{aligned} \varphi_n(k) &= \varphi_{n-1}(k) - \\ &- \max_{1 \leq i \leq m} (P_{loss}^i(k_i) - P_{loss}^i(k_i + 1)). \quad (40) \end{aligned}$$

Вычислим начальное значение $\varphi_0(k)$:

$$\varphi_0(k) = \sum_{i=1}^m \frac{(1 - \rho_i) \cdot p_i}{1 - \rho_i} = \sum_{i=1}^m p_i. \quad (41)$$

При размере памяти равном 0 попытка включения элемента в любую очередь будет приводить к его потере. Таким образом, применяя рекуррентную формулу (40), получаем разбиение памяти $k^* = (k_1^*, \dots, k_m^*)$. Покажем, что данное разбиение оптимально, т. е. нужно доказать, что $\varphi(k^*) \leq \varphi(k^{**})$ для любого $k^{**} = (k_1^{**}, \dots, k_m^{**})$. Введем обозначение:

$$s = \max_{1 \leq i \leq m} (P_{loss}^i(k_i) - P_{loss}^i(k_i + 1)). \quad (42)$$

Покажем, что $P_{loss}^i(k_i) - P_{loss}^i(k_i + 1) > P_{loss}^i(k_i + a) - P_{loss}^i(k_i + a + 1)$, где $a > 0$ — некоторое число.

Перейдем к непрерывному случаю. Рассмотрим функцию $h(x_i) = P_{loss}^i(x_i) - P_{loss}^i(x_i + 1)$ на $(0, n)$. При $p_i = q_i$,

$$h(x_i) = \frac{p_i}{x_i + 1} - \frac{p_i}{x_i + 2}, \quad (43)$$

$$h'(x_i) = -\frac{p_i}{(x_i + 1)^2} + \frac{p_i}{(x_i + 2)^2}. \quad (44)$$

Поскольку $x_i + 1 < x_i + 2$, то $h'(x_i) < 0$. Если $p_i \neq q_i$, то

$$\begin{aligned} h(x_i) &= p_i \cdot (1 - \rho_i) \cdot \frac{\rho_i^{x_i}}{1 - \rho_i^{x_i+1}} - \quad (45) \\ &- p_i \cdot (1 - \rho_i) \cdot \frac{\rho_i^{x_i+1}}{1 - \rho_i^{x_i+2}}, \end{aligned}$$

$$\begin{aligned} h'(x_i) &= p_i \cdot (1 - \rho_i) \cdot \ln \rho_i \cdot \frac{\rho_i^{x_i}}{(1 - \rho_i^{x_i+1})^2} - \quad (46) \\ &- p_i \cdot (1 - \rho_i) \cdot \ln \rho_i \cdot \frac{\rho_i^{x_i+1}}{(1 - \rho_i^{x_i+2})^2}. \end{aligned}$$

Заметим, что в слагаемых в (46) совпадают множители, кроме $R_1 := \frac{\rho_i^{x_i}}{(1 - \rho_i^{x_i+1})^2}$ и $R_2 := \frac{\rho_i^{x_i+1}}{(1 - \rho_i^{x_i+2})^2}$. Рассмотрим отношение:

$$\frac{R_1}{R_2} = \frac{(1 - \rho_i^{x_i+2})^2}{\rho_i \cdot (1 - \rho_i^{x_i+1})^2} = \left(\frac{1 - \rho_i^{x_i+2}}{1 - \rho_i^{x_i+\frac{3}{2}}} \right)^2. \quad (47)$$

- Пусть $\rho_i < 1$. Тогда, как нетрудно видеть, $h'(x_i) < 0$ (т.к. $\frac{R_1}{R_2} > 0$, $\ln \rho_i < 0$).
- Пусть $\rho_i > 1$. В этом случае также $h'(x_i) < 0$ (т.к. $\frac{R_1}{R_2} > 0$, $\ln \rho_i > 0$).

Таким образом, $h'(x_i) < 0$, следовательно, $h(x_i)$ монотонно убывает на $(0, n)$. Очевидно, что дискретная функция $h(k_i)$ также монотонно убывает.

По построению $P_{loss}^i(a) - P_{loss}^i(a + 1) \geq s$ при $0 \leq a \leq k_i$ и $P_{loss}^i(a) - P_{loss}^i(a + 1) \leq s$ при $a \geq k_i$.

$$\begin{aligned} P_{loss}^i(a) - P_{loss}^i(a + b) &= \sum_{j=0}^{b-1} (P_{loss}^i(a + j) - \quad (48) \\ &- P_{loss}^i(a + j + 1)) > bs \text{ при } a + b \leq k_i. \end{aligned}$$

Аналогично

$$P_{loss}^i(a) - P_{loss}^i(a + b) < bs \text{ при } a + b > k_i. \quad (49)$$

Разобьем множество индексов $I = \{1, \dots, n\}$ следующим образом:

- $I_1 = \{i : k_i^* < k_i^{**}\}$
- $I_2 = \{i : k_i^* = k_i^{**}\}$
- $I_3 = \{i : k_i^* > k_i^{**}\}$

$$\sum_{i \in I_1} (k_i^{**} - k_i^*) = \sum_{i \in I_3} (k_i^* - k_i^{**}) = d, \quad (50)$$

$$0 < d \leq n.$$

Следовательно,

$$\varphi(k^*) - \varphi(k^{**}) \leq \sum_{i \in I_1} (k_i^{**} - k_i^*) \cdot s - \quad (51)$$

$$- \sum_{i \in I_3} (k_i^* - k_i^{**}) \cdot s = ds - ds = 0.$$

Таким образом, $\varphi(k^*) \leq \varphi(k^{**})$, поэтому разбиение k^{**} не лучше k^* , т. е. разбиение k^* — оптимальное. Возможно существование нескольких оптимальных разбиений. В этом случае можно выбрать любое из них.

Данным способом можно найти оптимальное разбиение памяти за время $O(mn)$, так как для вычисления $\varphi_1, \dots, \varphi_n$ требуется n шагов, и на каждом шаге осуществляется поиск максимума выражения $(P_{loss}^i(k_i) - P_{loss}^i(k_i + 1))$ среди m очередей.

ЗАКЛЮЧЕНИЕ

В будущем мы предполагаем перенести подход, предложенный в данной статье, на задачу оптимального управления настраиваемыми очередями (Custom Queuing) в случае последовательного циклического представления входящих в механизм FIFO-очередей. Настраиваемые очереди — это один из аппаратно-независимых механизмов управления перегрузками, который похож на механизм приоритетных очередей, но, в отличие от него, не блокирует низкоприоритетный трафик. Настраиваемые очереди обслуживаются по очереди через некоторый промежуток времени.

Работа А. В. Соколова, Р. С. Некрасовой и А. М. Сазонова поддержана РФФИ (грант № 15-01-03404-а), а также Программой стратегического развития ПетрГУ в рамках реализации комплекса мероприятий по развитию научно-исследовательской деятельности.

Работа Е. В. Морозова частично поддержана РФФИ (проекты 15-07-02341, 15-07-02354, 15-07-02360), а также Программой стратегического развития ПетрГУ в рамках реализации комплекса мероприятий по развитию научно-исследовательской деятельности.

Работа Р. В. Разумчика поддержана РФФИ (грант 15-07-03007).

ЛИТЕРАТУРА

1. *Кнут Д.* Искусство программирования для ЭВМ. Т. 1. М.: Вильямс, 2001. 712 с.

2. *Седжвик Р.* Фундаментальные алгоритмы на C++. Киев: Диасофт, 2001. 687 с.

3. *Боллапрагада В., Мэрфи К., Уайт Р.* Структура операционной системы Cisco IOS. М.: Вильямс, 2002. 197 с.

4. *Соколов А. В.* О распределении памяти для двух стеков // Автоматизация эксперимента и обработки данных. 1980. С. 65–71.

5. *Yao A. C.* An Analysis of a Memory Allocation Scheme for Implementing Stacks // SIAM J. Comput. 1981. 10(2) P. 398–403. doi: 10.1137/0210029

6. *Louchard G.* Some distributed algorithms revisited // Communications in Statistics. Stochastic Models. 1995. Vol. 11, iss. 4. doi: 10.1080/15326349508807361

7. *Louchard G., Schott R., Tolley M., Zimmermann P.* Random walks, heat equation and distributed algorithms // Journal of Computational and Applied Mathematics. 1994. Vol. 53, iss. 2, P. 243–274. doi: 10.1016/0377-0427(94)90048-5

8. *Louchard G., Schott R.* Probabilistic analysis of some distributed algorithms // Random Structures & Algorithms, 1991. Vol. 2, iss. 2. P. 151–186. doi: 10.1002/rsa.3240020203

9. *Flajolet Ph.* The evolution of two stacks in bounded space and random walks in a triangle // Mathematical Foundations of Computer Science, Lecture Notes in Computer Science. 1986. Vol. 233. P. 325–340. doi: 10.1007/BFb0016257

10. *Аксенова Е. А.* Оптимальное управление FIFO-очередями на бесконечном времени // Стохастическая оптимизация в информатике. СПб.: С.-Петербургский университет, 2006. № 2 С. 71–76.

11. *Аксенова Е. А., Драц А. В., Соколов А. В.* Оптимальное управление n FIFO-очередями на бесконечном времени // Информационно-управляющие системы. 2009. № 6. С. 401–415.

12. *Drac A. V., Sokolov A. V.* The linked list representation of n LIFO-stacks and/or FIFO-queues in the single-level memory // Information Processing Letters. Elsevier Science Publishing Company Inc. 2013. Vol. 13, no. 19–21. P. 832–835.

13. *Соколов А. В., Драц А. В.* Моделирование некоторых методов представления n FIFO очередей в памяти одного уровня // Эвристические алгоритмы и распределенные вычисления. 2014. Т. 1, № 1. С. 40–52.

14. *Барковский Е. А., Соколов А. В.* Оптимальное управление двумя параллельными

FIFO-очередями на бесконечном времени // Информационно-управляющие системы. 2015. № 5. С. 65–71.

15. *Irland M. I.* Buffer management in packet switch // IEEE Transactions on Communications. Vol. COM-26. March 1978. P. 328–337. doi: 10.1109/TCOM.1978.1094076

16. *Rich M., Schwartz M.* Buffer sharing in computer-communication network nodes // IEEE trans. commun. Vol. COM-25. 1977. P. 958–970.

17. *Yamashita H., Onvural R. O.* Allocation of buffer capacities in queueing networks with arbitrary topologies // Annals of Operations Research. August 1994. Vol. 48, iss. 4. P. 313–332. doi: 10.1007/BF02024519

18. *Perros H. G.* Open queueing networks with blocking // Stochastic Analysis of Computer and

Communications Systems / Ed. H. Takagi. North-Holland, 1990.

19. *Yum T. P., Dou C.* Buffer allocation strategies with blocking requirements // Performance Evaluation. 1984. Vol. 4, iss. 4. P. 285–295. doi: 10.1016/0166-5316(84)90013-0

20. *Kamoun E., Kleinrock L.* Analysis of Shared Finite Storage in a Computer Network Node Environment Under General Traffic Schemes // IEEE Transactions on Communications. July 1980. Vol. COM-28, no. 7. P. 992–1003.

21. *Ozel O., Uysal-Biyikoglu E., Girici T.* Optimal Buffer Partitioning on a Multiuser Wireless Link // Proceedings of Information Theory and Applications Workshop, UCSD, San Diego, CA, Jan. 31 – Feb. 5, 2010. doi: 10.1109/ITA.2010.5454079

Поступила в редакцию 30.05.2016

REFERENCES

1. *Knuth D.* Искусство программирования для ЭВМ [The art of computer programming]. Vol. 1. Moscow: Vilyams, 2001. 712 p.

2. *Sedgewick R.* Fundamentalnye algoritmy na C++ [Fundamental algorithms in C++]. Kiev: Diasoft, 2001. 687 p.

3. *Bollapragada V., Murphi C., White R.* Структура операционной системы Cisco IOS [The structure of Cisco IOS]. Moscow: Vilyams, 2002. 197 p.

4. *Sokolov A. V.* O raspredelenii pamyati dlya dvuh stekov [On memory distribution for 2 stacks]. Avtomatizatsiya eksperimenta i obrabotki dannyh [Experiment automation and data processing]. 1980. P. 65–71.

5. *Yao A. C.* An Analysis of a Memory Allocation Scheme for Implementing Stacks. *SIAM J. Comput.* 10(2). 1981. P. 398–403. doi: 10.1137/0210029

6. *Louchard G.* Some distributed algorithms revisited. Communications in Statistics. *Stochastic Models.* 1995. Vol. 11, iss. 4. doi: 10.1080/15326349508807361

7. *Louchard G., Schott R., Tolley M., Zimmermann P.* Random walks, heat equation and distributed algorithms. *Journal of Computational and Applied Mathematics.* 1994. Vol. 53, iss. 2. P. 243–274. doi: 10.1016/0377-0427(94)90048-5

8. *Louchard G., Schott R.* Probabilistic analysis of some distributed algorithms. *Random Structures & Algorithms.* Summer 1991. Vol. 2, iss. 2. P. 151–186. doi: 10.1002/rsa.3240020203

9. *Flajolet Ph.* The evolution of two stacks in bounded space and random walks in a triangle.

Mathematical Foundations of Computer Science, Lecture Notes in Computer Science. 1986. Vol. 233. P. 325–340. doi: 10.1007/BFb0016257

10. *Aksenova E. A.* Optimalnoe upravlenie FIFO-ocheredyami na beskonechnom vremeni [Optimal control of the FIFO-queues for infinity time]. Stokhasticheskaya optimizatsiya v informatike [Stochastic optimization in informatics]. St.-Petersburg: S-Peterburgskii un., 2006. No. 2. C. 71–76.

11. *Aksenova E. A., Drac A. V., Sokolov A. V.* Optimalnoe upravlenie n FIFO-ocheredyami na beskonechnom vremeni [Optimal control of the n FIFO-queues for infinity time]. *Informacionno-upravlyayushchie sistemy* [Information and control systems]. 2009. No. 6. C. 401–415.

12. *Drac A. V., Sokolov A. V.* The linked list representation of n LIFO-stacks and/or FIFO-queues in the single-level memory. *Information Processing Letters.* Elsevier Science Publishing Company Inc. 2013. Vol. 13, no. 19–21. P. 832–835.

13. *Sokolov A. V., Drac A. V.* Modelirovanie nekotorykh metodov predstavleniya n FIFO ocheredej v pamyati odnogo urovnya [Simulation of some methods of representation of n FIFO-queues in the single-level memory]. *Evrysticheskie algoritmy i raspredelennye vychisleniya* [Heuristic algorithms and distributed computing]. 2014. Vol. 1, no. 1. C. 40–52.

14. *Barkovsky E. A., Sokolov A. V.* Optimalnoe upravlenie dvumya parallelnymi FIFO-ocheredyami na beskonechnom vremeni [Optimal control of two parallel FIFO-queues on an infinite time]. *Informacionno-upravlyayushchie sistemy* [Information and control systems]. 2015. No. 5. C. 65–71.

15. *Ireland M. I.* Buffer management in packet switch. *IEEE Transactions on Communications*. Vol. COM-26. March 1978. P. 328–337. doi: 10.1109/TCOM.1978.1094076
16. *Rich M., Schwartz M.* Buffer sharing in computer-communication network nodes. *IEEE trans. commun.* Vol. COM-25. 1977. P. 958–970.
17. *Yamashita H., Onvural R. O.* Allocation of buffer capacities in queueing networks with arbitrary topologies. *Annals of Operations Research*. August 1994. Vol. 48, iss. 4. P. 313–332. doi: 10.1007/BF02024519
18. *Perros H. G.* Open queueing networks with blocking, in: *Stochastic Analysis of Computer and Communications Systems*, ed. H. Takagi. North-Holland, 1990.
19. *Yum T. P., Dou C.* Buffer allocation strategies with blocking requirements. *Performance Evaluation*. 1984. Vol. 4, iss. 4. P. 285–295. doi: 10.1016/0166-5316(84)90013-0
20. *Kamoun E., Kleinrock L.* Analysis of Shared Finite Storage in a Computer Network Node Environment Under General Traffic Schemes. *IEEE Transactions on Communications*. July 1980. Vol. COM-28, no. 7. P. 992–1003.
21. *Ozel O., Uysal-Biyikoglu E., Girici T.* Optimal Buffer Partitioning on a Multiuser Wireless Link, In proceedings of Information Theory and Applications Workshop, UCSD, San Diego, CA, Jan. 31 – Feb. 5, 2010. doi: 10.1109/ITA.2010.5454079

Received May 30, 2016

СВЕДЕНИЯ ОБ АВТОРАХ:

Соколов Андрей Владимирович
ведущий научный сотрудник, к. т. н.
Институт прикладных математических
исследований Карельского научного центра РАН
ул. Пушкинская, 11, Петрозаводск,
Республика Карелия, Россия, 185910
эл. почта: sokavs@gmail.com
тел.: (8142) 766313

Сазонов Александр Михайлович
студент
Петрозаводский государственный университет
пр. Ленина, 33, Петрозаводск,
Республика Карелия, Россия, 185910
эл. почта: sazonov@cs.karelia.ru

Морозов Евсей Викторович
ведущий научный сотрудник, д. ф.-м. н., проф.
Институт прикладных математических
исследований Карельского научного центра РАН
ул. Пушкинская, 11, Петрозаводск,
Республика Карелия, Россия, 185910
эл. почта: emorozov@krc.karelia.ru
тел.: (8142) 766313

Некрасова Руслана Сергеевна
младший научный сотрудник, к. ф.-м. н.
Институт прикладных математических
исследований Карельского научного центра РАН
ул. Пушкинская, 11, Петрозаводск,
Республика Карелия, Россия, 185910
эл. почта: nekrasova@krc.karelia.ru
тел.: (8142) 766313

Разумчик Ростислав Валерьевич
старший научный сотрудник, к. ф.-м. н.
Институт проблем информатики ФИЦ ИУ РАН
ул. Вавилова, 44, Москва, Россия, 119333
эл. почта: rrazumchik@gmail.com

CONTRIBUTORS:

Sokolov, Andrey
Institute of Applied Mathematical Research,
Karelian Research Centre,
Russian Academy of Sciences
11 Pushkinskaya St., 185910
Petrozavodsk, Karelia, Russia
e-mail: sokavs@gmail.com
tel.: (8142) 766313

Sazonov, Alexandr
Petrozavodsk State University
33 Lenin St., 185910 Petrozavodsk,
Karelia, Russia
e-mail: sazonov@cs.karelia.ru

Morozov, Evsey
Institute of Applied Mathematical Research,
Karelian Research Centre,
Russian Academy of Sciences
11 Pushkinskaya St., 185910 Petrozavodsk,
Karelia, Russia
e-mail: emorozov@krc.karelia.ru
tel.: (8142) 766313

Nekrasova, Ruslana
Institute of Applied Mathematical Research,
Karelian Research Centre,
Russian Academy of Sciences
11 Pushkinskaya St., 185910 Petrozavodsk,
Karelia, Russia
e-mail: nekrasova@krc.karelia.ru
tel.: (8142) 766313

Razumchik, Rostislav
Institute of Informatics Problems,
Russian Academy of Science
44 Vavilova St., 119333 Moscow, Russia
e-mail: rrazumchik@gmail.com