

УДК 689.13.19/16(045)

В. М. Синеглазов, д-р.техн.наук, проф.,
А. С. Юрченко, канд. техн. наук

РАЗРАБОТКА АНАЛИТИЧЕСКОЙ МОДЕЛИ ДИНАМИЧЕСКОГО РАСПРЕДЕЛЕНИЯ НЕСТРАНИЧНОЙ ПАМЯТИ

Институт аэрокосмических систем управления НАУ, email: AYurchenko@yahoo.com

Рассмотрен подход к построению аналитической модели процессов распределения оперативной памяти в вычислительных системах, использующих сегментное распределение памяти.

Ключевые слова: алгоритмы распределения памяти, фрагментация.

Введение. Алгоритмы динамического распределения нестраничной памяти используют либо сегментное распределение, либо распределение памяти алгоритмами «близнецов». Всю совокупность алгоритмов сегментного распределения памяти можно разделить на два класса: алгоритмы предоставления памяти из резерва свободной памяти (первого подходящего, наиболее подходящего, наименее подходящего и алгоритмы перераспределения памяти) [1]. В литературных источниках приводятся противоречивые результаты исследований эффективности различных алгоритмов распределения памяти. Ввиду сложности описания процессов распределения памяти аналитические модели таких процессов вообще отсутствуют. Поэтому задачей данной работы является разработка аналитической модели процессов распределения нестраничной памяти.

Аналитическая модель распределения памяти. При построении модели распределения памяти систему динамического распределения памяти будем рассматривать как обслуживающее устройство одноканальной системы массового обслуживания (СМО) с отказами, структура которой представлена на рис. 1. Источником заявок будут служить исполняющиеся программы с их заявками к операционной системе на предоставление памяти. Процессы массового обслуживания в этой модели являются дискретными процессами с конечным или счетным множеством состояний и непрерывным временем. Переход из одного состояний в другое происходит скачком в момент, когда наступает какое-то событие, вызывающее такой переход (поступление нового требования, начало или конец обслуживания, уход требования из очереди т. д.). Для процессов массового обслуживания с пуассоновским потоком требований и экспоненциальным распределением времени обслуживания развитие зависит от состояния только в настоящий момент и не зависит от того, как происходило развитие в прошлом [2]. Обозначим через $S_{N,M}$ состояние памяти, в которой N занятых сегментов и M свободных сегментов ($S_{0,1}$ – состояние памяти, в котором вся память свободная). Обозначим через $P_{N,M}$ вероятность того, что в момент t система памяти находится в состоянии $S_{N,M}$ ($N = 0, 1, 2, \dots, N_{\max}; M = 1, 2, \dots, N_{\max} + 1$), где N_{\max} – максимальное количество сегментов информации, которые могут быть размещены в памяти. Очевидно, для любого момента времени t сумма вероятностей состояний равна единице (нормировочное условие), т. е.

$$\sum_N \sum_M P_{N,M}(t) = 1,$$

так как события, состоящие в том, что в момент t система памяти находится в состояниях $S_{0,1}; \dots; S_{N_{\max}, N_{\max}+1}$ несовместимы и образуют полную систему событий. Задача состоит в том, чтобы определить вероятности состояний $P_{0,1}(t), \dots, P_{N_{\max}, N_{\max}+1}(t)$ как функции времени.



Рис. 1. Структура системы массового обслуживания

Марковский процесс описывается относительно вероятностей $P_{0,1}(t), \dots, P_{N_{\max}, N_{\max}+1}(t)$ системой дифференциальных уравнений, называемых уравнениями Колмогорова. При составлении этих уравнений удобно воспользоваться графом состояний, вершины которого соответствуют состояниям, а дуги – возможным переходам из состояния в состояние. Для рассматриваемой системы массового обслуживания граф состояния $S_{N,M}$ показан на рис. 2, на котором представлены возможные варианты занятия свободных сегментов, а также освобождение занятых свободных сегментов. Зафиксируем момент t и найдем вероятность $P_{N,M}(t + \Delta t)$ того, что в момент $t + \Delta t$ система будет в состоянии $S_{N,M}$. Так как система может оставаться в прежнем состоянии или переходить только в соседние состояния, то

$$P_{N,M}(t + \Delta t) = P(A) + P(B) + P(C) + P(D) + P(E),$$

где A, B, C, D, E – несовместимые события. Событие A означает, что система за время Δt не изменила своего состояния $S_{N,M}$, а события B, C, D, E – что переход в $S_{N,M}$ произошел соответственно из состояний $S_{N-1,M}, S_{N-1,M+1}, S_{N+1,M+1}, S_{N+1,M}$.



Рис. 2. Граф состояния

Пусть система в момент t находилась в состоянии $S_{N,M}$ и вероятность того, что за время Δt она перейдет в состояние S_{N_1, M_1} равно $P_{N,M}^{N_1, M_1}(\Delta t)$. Величину

$$\lambda_{N,M}^{N_1, M_1} = \lim_{\Delta t \rightarrow 0} \frac{P_{N,M}^{N_1, M_1}(\Delta t)}{\Delta t}$$

называют плотностью вероятности перехода. При достаточно малом Δt справедливо приближенное соотношение

$$P_{N,M}^{N1,M1}(\Delta t) \approx \lambda_{N,M}^{N1,M1} \Delta t.$$

Очевидно, вероятность того, что система за время Δt не перейдет из состояния $S_{N,M}$ в состояние $S_{N1,M1}$ выражается

$$1 - P_{N,M}^{N1,M1}(\Delta t) \approx 1 - \lambda_{N,M}^{N1,M1} \Delta t.$$

Выразим вероятность событий A, B, C, D и E через вероятности состояний и плотности вероятностей перехода (членами высших порядков малости по сравнению с Δt пренебрегаем):

$$\begin{aligned} P(A) &\approx P_{N,M}(t)(1 - \lambda_{N,M}^{N-1,M-1} \Delta t)(1 - \lambda_{N,M}^{N-1,M} \Delta t)(1 - \lambda_{N,M}^{N-1,M+1} \Delta t)(1 - \lambda_{N,M}^{N+1,M} \Delta t)(1 - \lambda_{N,M}^{N+1,M-1} \Delta t) \approx \\ &\approx P_{N,M}(t) \left[1 - \left(\lambda_{N,M}^{N-1,M-1} + \lambda_{N,M}^{N-1,M} + \lambda_{N,M}^{N-1,M+1} + \lambda_{N,M}^{N+1,M} + \lambda_{N,M}^{N+1,M-1} \right) \Delta t \right]; \end{aligned}$$

$$P(B) \approx P_{N-1,M}(t) \lambda_{N-1,M}^{N,M} \Delta t;$$

$$P(C) \approx P_{N-1,M+1}(t) \lambda_{N-1,M+1}^{N,M} \Delta t;$$

$$P(D) \approx P_{N+1,M+1}(t) \lambda_{N+1,M+1}^{N,M} \Delta t;$$

$$P(E) \approx P_{N+1,M}(t) \lambda_{N+1,M}^{N,M} \Delta t.$$

На основании этих соотношений имеем:

$$\begin{aligned} P_{N,M}(t + \Delta t) &= P_{N,M}(t) \left[1 - \left(\lambda_{N,M}^{N-1,M-1} + \lambda_{N,M}^{N-1,M} + \lambda_{N,M}^{N-1,M+1} + \lambda_{N,M}^{N+1,M} + \lambda_{N,M}^{N+1,M-1} \right) \Delta t \right] + \\ &+ P_{N-1,M}(t) \lambda_{N-1,M}^{N,M} \Delta t + P_{N-1,M+1}(t) \lambda_{N-1,M+1}^{N,M} \Delta t + P_{N+1,M+1}(t) \lambda_{N+1,M+1}^{N,M} \Delta t + \\ &+ P_{N+1,M}(t) \lambda_{N+1,M}^{N,M} \Delta t \end{aligned}$$

или

$$\begin{aligned} \frac{P_{N,M}(t + \Delta t) - P_{N,M}(t)}{\Delta t} &= P_{N-1,M}(t) \lambda_{N-1,M}^{N,M} + P_{N-1,M+1}(t) \lambda_{N-1,M+1}^{N,M} - \\ &- P_{N,M}(t) \left(\lambda_{N,M}^{N-1,M-1} + \lambda_{N,M}^{N-1,M} + \lambda_{N,M}^{N-1,M+1} + \lambda_{N,M}^{N+1,M} + \lambda_{N,M}^{N+1,M-1} \right) + \\ &+ P_{N+1,M+1}(t) \lambda_{N+1,M+1}^{N,M} + P_{N+1,M}(t) \lambda_{N+1,M}^{N,M}. \end{aligned}$$

Переходя к пределу при $\Delta t \rightarrow 0$, получаем дифференциальное уравнение относительно производной вероятности $S_{N,M}$ состояния (для простоты аргументы t вероятностей состояний опускаем):

$$\begin{aligned} \frac{dP_{N,M}}{dt} &= P_{N-1,M} \lambda_{N-1,M}^{N,M} + P_{N-1,M+1} \lambda_{N-1,M+1}^{N,M} - \\ &- P_{N,M} \left(\lambda_{N,M}^{N-1,M-1} + \lambda_{N,M}^{N-1,M} + \lambda_{N,M}^{N-1,M+1} + \lambda_{N,M}^{N+1,M} + \lambda_{N,M}^{N+1,M-1} \right) + \\ &+ P_{N-1,M+1} \lambda_{N-1,M+1}^{N,M} + P_{N+1,M} \lambda_{N+1,M}^{N,M} \end{aligned}$$

Записав аналогичные выражения для всех состояний, получим систему дифференциальных уравнений Колмогорова в матричной форме:

$$\frac{dP}{dt} = \Lambda P,$$

где $P = (P_{0,1,\dots, P_{N_{\max}, N_{\max}+1}})$ – вектор вероятностей состояний и Λ – матрица плотностей вероятностей переходов. Систему уравнений Колмогорова легко записать непосредственно из размеченного графа системы, в котором каждой дуге приписан вес, равный соответствующей плотности вероятности перехода. Так, сравнивая рис. 3 с уравнением для $S_{N,M}$ состояния, легко вывести простое правило. Производная вероятности $S_{N,M}$ состояния равна сумме членов, каждый из которых представляет собой произведения веса дуги, инцидентной $S_{N,M}$ вершине, на вероятность того состояния, к которому она направлена. При этом вес дуги принимается положительным, если дуга направлена от $S_{N,M}$ вершины, и отрицательным, если дуга направлена к $S_{N,M}$ вершине. Это правило применимо для записи уравнений Колмогорова по размеченному графу любого марковского процесса. На основе приведенного правила можно записать непосредственно и матрицу Λ .

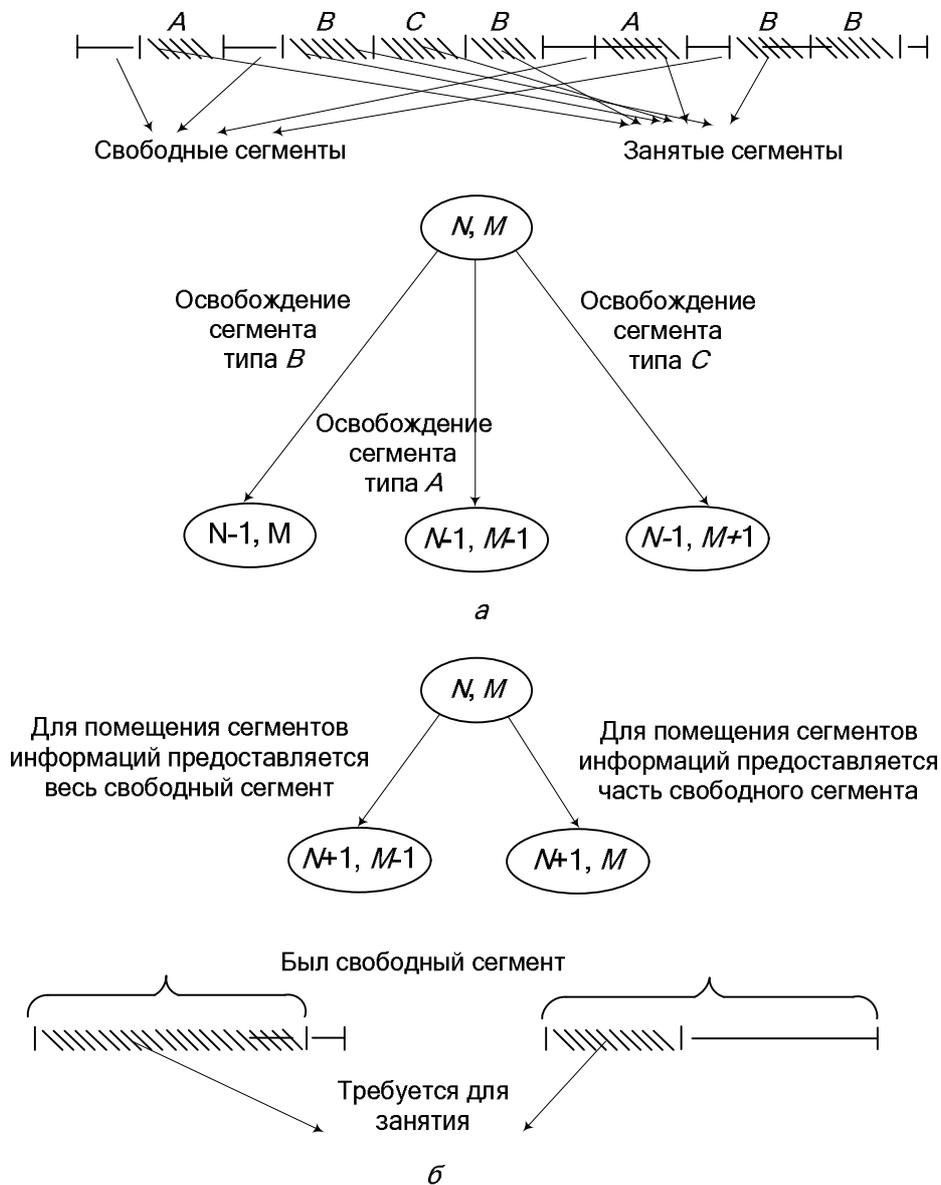


Рис. 3. Возможные типы занятия и освобождения сегментов: *a* – освобождение; *б* – занятие свободных

Различные алгоритмы будут отличаться матрицей Λ . Основная трудность при решении полученных уравнений Колмогорова заключается в задании матриц Λ для различных алгоритмов. Решение поставленной задачи в настоящее время не представляется возможным, доказательством тому – безусловные попытки решить аналогичную задачу даже для наиболее известных алгоритмов FIRST_FIT и BEST_BIT. Поэтому исследование различных алгоритмов распределения памяти стараются выполнить путем получения всевозможных приближенных оценок, накладывая определенные ограничения на характеристики потока требований: размеры, «времена жизни» сегментов и другие, а основное представление о алгоритмах распределения на нестраничной памяти возможно с помощью имитационного моделирования.

Выводы. Предлагаемая модель описания процессов распределения нестраничной памяти сводится к решению дифференциальных уравнений Колмогорова. Решение этих уравнений представляет достаточно сложную задачу, поскольку необходимо задавать вероятности переходов между различными состояниями, определяемыми количеством занятых и свободных сегментов. Поэтому основной упор при исследовании алгоритмов управления памятью делается на имитационное моделирование различных алгоритмов.

Список литературы

1. Дональд Кнут. Искусство программирования / Дональд Кнут // Т.1. Основные алгоритмы = The Art of Computer Programming, Vol. 1. Fundamental Algorithms. – 3-е изд. – М.: Вильямс, 2006. – С. 720.
2. Клейнрок Л. Теория массового обслуживания / Л. Клейнрок. – М.: Машиностроение, 1979. – С. 432.

В. М. Синеглазов, О. С. Юрченко

Розроблення аналітичної моделі динамічного розподілу несторінкової пам'яті

Розглянуто підхід до побудови аналітичної моделі процесів розподілу оперативної пам'яті в обчислювальних системах, які використовують сегментний розподіл пам'яті.

V. M. Sineglazov, A. S. Jurchenko

To develop an analytical model of the dynamic distribution of non-paged memory

An approach to the construction of analytical models of the processes of memory allocation in computer systems using segmented memory allocation.