

ТЕХНИКА МОДЕЛИРОВАНИЯ МНОГОКАНАЛЬНЫХ ПРИОРИТЕТНЫХ СИСТЕМ*

Ю. И. Рыжиков (Санкт-Петербург)

Введение

Системы обслуживания любого рода (технические и организационные) и назначения (производственные процессы, передача данных, здравоохранение, бытовое обслуживание, охрана общественного порядка, военное дело) время от времени нуждаются в профилактическом осмотре, проведении регламентных и ремонтных работ и, наконец, в отдыхе персонала. Необходимость непрерывного выполнения ответственных функций, хотя бы и с уменьшенной производительностью, определяет применение *многоканальных* систем обслуживания. Такие системы, особенно в период функционирования в неполном составе, эксплуатируются в режиме, близком к насыщению. Поэтому для достижения приемлемой оперативности обслуживания хотя бы по наиболее важным заявкам приходится вводить *приоритеты*. Теория многоканальных систем обслуживания с приоритетами до настоящего времени практически не разработана. В обсуждающих ее немногих статьях делаются настолько стеснительные оговорки (как правило, об экспоненциальном и к тому же одинаковом для всех типов заявок распределении времени обслуживания [1]), что результаты оказываются бесполезными.

В работе автора [2] намечена свободная от этого ограничения схема решения задачи для средних значений ожидания (пребывания) в системе заявок каждого вида и демонстрируются удовлетворительные результаты ее применения. Однако некоторые элементы этой схемы построены на эвристиках, не гарантирующих достаточную точность во всем возможном диапазоне данных.

Изложенные соображения вынуждают считать имитационное моделирование (ИМ) единственным (пока) реальным инструментом исследования многоканальных систем с приоритетным обслуживанием. Среди инженеров и экономистов большой популярностью пользуются системы автоматизации ИМ (САИМ), прежде всего, созданная полвека назад GPSS и ее многочисленные клоны. Любая такая система избавляет потребителя от описания сложной логики процесса, предлагая ему транслятор, модифицирующий и настраивающий встроенный «универсальный» алгоритм моделирования. *Эта универсальность не может не быть ограниченной*, поскольку для любой системы ИМ существуют недоступные ей ситуации (например, для GPSS – моделирование вложенных прерываний и многоканальных систем с прерываниями). С другой стороны, претензии на универсальность усложняют фактически реализуемую модель. К примеру, в проведенном автором эксперименте по моделированию многоканальной системы обслуживания прогон GPSS-программы занял 4 с, а составленной на Фортране – 0.05 с. Разумеется, здесь важны не абсолютные значения трудоемкости, а разница в скорости – 80 раз!

По указанным причинам никакая САИМ, особенно при решении исследовательских задач, не заменит «прямого» моделирования на языках решения общематематических задач типа современного Фортрана, где возможности построения сложных моделей ограничены только *компетентностью программиста*. Заметим, кстати, что именно здесь проходит магистральный путь совершенствования САИМ. Ниже дается описание логики моделирования многоканальных приоритетных систем применительно к обоснованию будущих полуаналитических методов расчета. Соответственно целью мо-

*Работа поддержана программой фундаментальных исследований ОНИТ РАН (проект 2.3) и грантом РФФИ 10-08-00906-а.

делирования являлось определение по каждому типу заявок средних значений времени пребывания, кратности прерываний и их длительности.

Логика моделирования

При описании логики заявленной модели мы ограничимся указанием ее особенностей, имея в виду, что общая схема модели многоканальной системы читателю известна (см., например, [3]). Настройка модели на типы распределений интервалов между заявками и длительности обслуживания производилась посредством допускаемых современным Фортраном *внутренних процедур*. Генерировался суммарный рекуррентный поток внешних заявок с последующим случайным выбором вида заявки на основе их долей в суммарном потоке. Интенсивность потока для поддержания расчетного коэффициента загрузки умножалась на заданное число каналов. Распределение времени обслуживания для всех видов предполагалось однотипным с различными по видам заявок средними, причем соответствующий генератор предполагался сменным (рассматривались регулярное, показательное, эрлангово и гиперэкспоненциальное распределения).

Вышеуказанная цель моделирования определяет необходимость сбора соответствующей статистики отдельно для заявок каждого вида. В нашем случае на каждую заявку заводится паспорт, включающий в себя тип заявки z (он же – индекс приоритета), момент τ ее входа в систему, случайную трудоемкость θ . Для *прерванной* заявки в качестве θ записывается *остаток* времени обслуживания, вычисляемый как разность моментов планового завершения обслуживания и начала последнего прерывания его. В остальные позиции паспорта записываются кратность прерываний, начало последнего прерывания данной заявки и суммарная длительность ожидания.

При однородных каналах простейшей дисциплиной выбора занимаемого канала является «первый свободный», выявляемый по намного превышающему допустимые значения таймера «запредельному» моменту освобождения. Содержание паспортов заносится в каналы (при наличии свободных) или в очереди – отдельно по типам заявок в порядке возрастания моментов прибытия. В «канальном» массиве 5-я позиция паспорта отводится для запоминания «физического» номера канала, обслуживающего данную заявку. Для ускорения поиска прерываемых и выбираемых на обслуживание после освобождения канала заявок после изменения состояния любого канала проводится переупорядочение списка «канальных» паспортов по убыванию приоритетов, а при равных приоритетах – по возрастанию времени прибытия в систему. В этом случае кандидат на прерывание всегда оказывается последним. Таким образом, в случае прибытия заявки в полностью занятую систему приоритет вновь прибывшей заявки достаточно сопоставить с приоритетом обслуживаемой в последнем канале.

Обозначим через π младший приоритет текущего обслуживания и сравним его с типом i новой заявки. При $i < \pi$ заявка, находящаяся в n -й позиции списка, прерывается; определяется остаток обслуживания; прерванная заявка помещается в очередь. При $i \geq \pi$ в i -очередь помещается новая заявка. Место вставки для прерванной заявки нужно искать в начале частной очереди, а для новой – в ее конце. Если общая очередь уже имеет максимальную длину, то (в зависимости от соотношения приоритетов) получает отказ либо новая заявка, либо последняя заявка в очереди. В паспорт прерывающей заявки заносится номер освобожденного канала, после чего для нее определяется позиция вставки. Затем выполняются вставки, вследствие чего сдвигаются нижележащие паспорта. Заявка, выбираемая из очереди после завершения обслуживания, всегда помещается в n -ю строку списка паспортов.

При выборе заявки на обслуживание *из очереди* проверяется текущая кратность ее прерываний. Если эта кратность нулевая, то в 6-ю позицию паспорта заносится время ожидания (разность между таймером и 2-й позицией – моментом прибытия заявки).

В противном случае фиксируется длительность завершаемого прерывания (разность между таймером и 5-й позицией паспорта), которая добавляется к 6-й позиции. Для заявки, закончившей свое пребывание в системе, в соответствии с ее типом (первая позиция паспорта) производится накопление результирующей статистики: случаев завершения, кратностей прерываний, суммарного ожидания и длительности прерываний.

Дальнейшие обобщения

Технология формирования потока заявок с просеиванием суммарного потока может быть заменена на работу с *неоднородным рекуррентным потоком*: отдельную (с разными видами распределений) генерацию интервалов между заявками каждого вида.

При нескольких свободных каналах может потребоваться *равновероятный* выбор из них. Если таких каналов m , то для определения выбираемого канала следует вычислить $N = [m \cdot U] + 1$, где U – случайное число, равномерно распределенное на полуинтервале $[0,1)$, а квадратные скобки означают взятие целой части. Далее начинается просмотр моментов освобождения каналов, и для каждого свободного – в счетчик добавляется единица. Заявка принимается в тот канал, для которого содержимое счетчика сравнивается с N . Дальнейшим обобщением может явиться полная или частичная специализация каналов обслуживания по видам обслуживаемых заявок.

Выше была описана работа модели с *дообслуживанием прерванной заявки*. Отметим дополнительные особенности моделирования приоритетной системы с прерыванием и возобновлением обслуживания со старой или новой реализациями его случайной длительности (схемы RW и RS соответственно, даже в одноканальном случае приводящие к чрезвычайно сложным расчетным схемам [4]). В первом варианте требуемое время обслуживания формируется один раз (при прибытии заявки) и в случае прерывания пересчету не подлежит. Во втором варианте целесообразно формировать требуемое время обслуживания в момент занятия канала (независимо от кратности захода) и соответствующую компоненту из паспорта заявки исключить.

Логика модели с относительным приоритетом (без прерываний) существенно проще и легко реализуется как упрощенный вариант описанной выше.

Опыт применения

Описанная модель была применена для проверки многообещающей методики «инвариантов отношения» [1]. Применительно к интересующей нас задаче «инвариант отношения» имеет вид символической пропорции, составленной из средних для соответствующих систем величин:

$$\frac{\bar{M} / \bar{G} / n}{\bar{M} / \bar{G} / 1} \approx \frac{M / G / n}{M / G / 1},$$

откуда следует, что

$$\bar{M} / \bar{G} / n \approx \bar{M} / \bar{G} / 1 \cdot \frac{M / G / n}{M / G / 1}.$$

Поскольку *численные* методы расчета систем из правой части пропорции к настоящему времени (в отличие от года написания [1]) известны [3], возникла идея использовать это отношение для расчета средних времен ожидания и пребывания заявок каждого типа *порознь*. При проведении численных экспериментов базовые интенсивности входящих простейших потоков принимались равными $\{0,222, 0,333, 0,44\}$, а средние длительности обслуживания – $\{0,45, 0,90, 1,35\}$. Для получения заданного суммарного коэффициента загрузки 0,9 интенсивности потоков умножались на этот коэффициент и на количество каналов обслуживания. Длительности обслуживания предпола-

гались подчиненными гамма-распределению с параметрами формы $\alpha=3$ и $\alpha=0,25$, что соответствовало коэффициентам вариации 0,577 и 2,0. Высшие моменты $\{b_i\}$ распределений обслуживания вычислялись через первые согласно

$$b_i = b_{i-1} \cdot b_1 \cdot [1 + (i-1) / \alpha], \quad i = 2, 3.$$

В процессе расчета по этим моментам подбиралась H_2 -аппроксимация (для первого α ее параметры оказывались комплексными, а для второго – вещественными). В табл. 1 и 2 приведены результаты расчета среднего времени пребывания заявок типов 1-3 в системе $\bar{M} / \bar{G} / n$ при относительном и абсолютном приоритетах в сопоставлении с результатами имитационных экспериментов. Прогон моделей осуществлялся до обработки 300 тыс. заявок первого типа.

Таблица 1

Относительный приоритет

Тип обслуж.	Метод	n=1			n=3			n=5		
		1	2	3	1	2	3	1	2	3
E_3	Имит.	1,195	2,057	11,992	0,716	1,326	4,520	0,607	1,159	3,082
	Инвар.	1,192	2,059	11,931	0,680	1,260	4,634	0,579	1,102	3,195
H_2	Имит.	3,203	5,180	37,517	1,093	2,099	12,540	0,757	1,518	7,672
	Инвар.	3,233	5,246	41,028	1,276	2,189	13,121	0,888	1,584	7,599

Таблица 2

Абсолютный приоритет

Тип обслуж.	Метод	n=1			n=3			n=5		
		1	2	3	1	2	3	1	2	3
E_3	Имит.	0,480	1,313	12,743	0,450	0,948	4,865	0,450	0,911	3,296
	Инвар.	0,480	1,313	12,689	0,450	0,826	4,846	0,450	0,799	3,407
H_2	Имит.	0,560	2,186	35,197	0,449	1,022	11,652	0,450	0,924	7,163
	Инвар.	0,561	2,204	41,786	0,451	0,894	13,332	0,450	0,811	7,791

Из этих таблиц можно сделать следующие выводы:

1. Расхождение результатов «инвариантного» счета и имитационного моделирования не превышает 15%, что свидетельствует о приемлемости идеи расчета и корректности как расчетных схем обоих сравниваемых подходов, так и реализующих их компьютерных программ.

2. Упомянутая корректность дополнительно подтверждается: качественным соответствием результатов разумным ожиданиям: уменьшением по числу каналов среднего времени пребывания и разброса его по типам заявок, а также заметным увеличением этого разброса при переходе от относительных к абсолютным приоритетам.

3. Расхождение результатов допустимо в очень широком диапазоне коэффициентов вариации распределений длительности обслуживания (в рассмотренных примерах – от 0,577 до 2,0) и не обнаруживает тенденцию к росту при увеличении числа каналов.

Заключение

Разработанный метод опирается на максимальную приближенность опорных вариантов к обчислываемой ситуации и дифференцированный учет влияния вида приоритетов. Так, для относительного приоритета по методу инвариантов вычислялось среднее время ожидания, к которому затем прибавлялось известное достоверно среднее время обслуживания. Поэтому результаты счета для относительного приоритета согласуются с имитационным моделированием заметно лучше, чем для абсолютного. Таким образом, *инварианты отношения следует применять для оценки только той части результирующей величины, которая не может быть вычислена точно*. При практическом использовании инвариантного подхода необходимости в трудоемком как на этапе программирования, так и в процессе счета имитационном моделировании нет.

Основная доля трудоемкости метода приходится на расчет модели $M/H_2/n$, повторяемый для абсолютного приоритета по числу типов заявок, а для относительного приоритета – однократно.

Предложенная в данной статье схема использования инвариантов отношения, в отличие от предложений в работе [1], опирается не на приближенные оценки средней длины очереди, а на надежные численные алгоритмы. Она позволяет с достаточной для практических целей точностью рассчитывать средние времена пребывания заявок в *многоканальной системе с произвольными распределениями длительности обслуживания*.

Литература

1. **Бронштейн О. И., Духовный И. М.** Модели приоритетного обслуживания в информационно-вычислительных системах. М.: Наука, 1976. 220 с.
2. **Рыжиков Ю. И.** Средние времена ожидания и пребывания в многоканальных приоритетных системах // Информационно-управляющие системы. 2006. № 6(25). С. 43–49.
3. **Рыжиков Ю. И.** Компьютерное моделирование систем с очередями: Курс лекций. СПб.: ВКА им. А. Ф. Можайского, 2007. 164 с.
4. **Конвей Р. В., Максвелл В. Л., Миллер Л. В.** Теория расписаний / Пер. с англ. М.: Наука, 1975. 359 с.