

**ИМИТАЦИОННАЯ МОДЕЛЬ ПРИОРИТЕТНОЙ МНОГОКАНАЛЬНОЙ
СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ****С. В. Кокорин (Санкт-Петербург)**

Интерес к приоритетным системам имеется в различных областях социальной деятельности: моделирование экстремальных ситуаций в больницах, потока запросов в информационных системах, ремонтных работ. Приоритетные системы сложны для обработки как с точки зрения аналитического, так и имитационного подхода.

Аналитические методы позволяют решать данную задачу только в средних и только для некоторых распределений времени обслуживания и интервалов между заявками. В большинстве случаев входной поток предполагается простейшим [1], при этом лишь в некоторых случаях различным для разных классов заявок [2]. В то же время специализированные имитационные приложения, такие как GPSS, требуют значительных вычислительных мощностей для получения достаточно точных оценок. Длительность моделирования должна быть такой, чтобы результаты моделирования были достоверными для каждого класса заявок, включая класс заявок, создающий наименьшую нагрузку [3].

Рассмотрим реализацию построения приоритетной многоканальной системы очередей со смешанной статической системой приоритетов. Типы заявок с различными относительными приоритетами объединяются в классы с различными абсолютными приоритетами. Тип заявки определяет распределения времени обслуживания и промежутков между появлением заявок. Возможна имитация систем с конечной и бесконечной очередью. При бесконечной очереди ограничением является только доступная приложению память. Таким образом, чтобы описать имитационную систему, необходимо задать распределение времени обслуживания и интервалов между заявками для каждого типа заявок в отдельности, количество каналов, максимальное количество заявок в системе и правило объединения типов заявок с относительными приоритетами в группы с абсолютными приоритетами.

При имитации используется событийное моделирование времени. В системе может произойти два принципиально различных события: появление новой заявки либо завершение обслуживания в одном из каналов. Для всех каналов и всех типов заявок приложение хранит информацию о ближайшем событии в будущем, относительно текущего модельного времени. В случае коллизий первыми обрабатываются появления заявок в порядке уменьшения приоритетов, затем обрабатывается обслуживание, также в порядке уменьшения приоритетов.

При появлении новой заявки для нее ищется место в очереди; если такого не оказалось (в случае конечной очереди), то в очереди ищется заявка с наименьшим относительным приоритетом; если таковых несколько, то выбирается та, которая находится в системе меньше время. Прибывшая заявка сравнивается с выбранной; если её приоритет выше, то она замещает собой выбранную заявку. После добавления заявки в очередь производится поиск свободного канала обслуживания; если он нашёлся, то заявка поступает на обслуживание, если же нет, то среди заявок в обслуживании выбирается одна с наименьшим абсолютным приоритетом, находящаяся в обслуживании наименьшее время (если имеется несколько с одним приоритетом). Учитывается совокупное время пребывания заявки в обслуживании, если заявка уже вытеснялась до этого момента. Если абсолютный приоритет прибывшей заявки выше, она замещает собой выбранную.

Продолжительность обслуживания конкретной заявки генерируется единожды во время первого поступления заявки в канал и сохраняется для каждой заявки в систе-

ме. В случае вытеснения – обновляется остаточное время обслуживания у вытесненной заявки, заявка возвращается в очередь (для неё всегда будет место в случае конечной очереди системы). Для заявок, находящихся в обслуживании, сделан специальный индекс для быстрого поиска среди всех заявок, находящихся в системе. По умолчанию заявки, вытесненные с прибора, возвращаются в начало очереди соответствующего класса. Любая заявка может быть вытеснена произвольное количество раз безотносительно того, как часто она вытеснялась раньше.

Когда какая-либо заявка покидает систему, заявка с наивысшим относительным приоритетом выбирается из очереди; если таковых несколько, то из них выбирается заявка, появившаяся в системе первой, при этом количество прерываний и продолжительность уже полученного обслуживания не учитывается.

Имитационная модель реализована на языке Фортран в виде расширяемого пакета. Сбор статистических данных о системе реализован в виде обработки событий: инициализация, появление новой заявки в системе, завершение обслуживания заявки, завершение имитации. Точка останова генерации реализаций может быть задана произвольно. Пакет предоставляет возможность использования следующих условий: количество прибывших заявок, количество обслуженных заявок по достижении определённого модельного времени. В докладе рассматривается остановка по пороговому значению общего числа обслуженных заявок. При моделировании использовались генераторы равномерного, показательного, гиперэкспоненциального, эрлангового распределений, регулярный поток.

Из интересующих характеристик системы были выбраны следующие вычисляемые для каждого типа заявки: распределение числа заявок в системе, коэффициент загрузки, моменты времени пребывания и ожидания, распределение числа вытеснений для одной заявки, количество обслуженных заявок данного типа, доля потерянных заявок в случае конечной очереди.

В случае исследования приоритетных систем интересно влияние различных стратегий выбора приоритетов и их связь с другими характеристиками. Так, ввод абсолютных приоритетов сокращает среднее время пребывания для заявок с высокими приоритетами за счёт увеличения времени пребывания заявок с низкими в значительно большей степени, чем в случае относительных приоритетов. В случае приоритетных систем очень ярко проявляется эффект расщепления – изменение мощности и числа каналов с сохранением коэффициента загрузки сети. Влияние расщепления в первую очередь проявляется для заявок с высокими приоритетами и в значительно меньшей степени для низших приоритетов. При этом среднее совокупное время пребывания заявок в системе растёт с увеличением числа каналов. Разделение заявок на несколько типов с различными относительными приоритетами увеличивает среднее время пребывания заявки в системе за исключением разбиения заявок из класса с максимальным абсолютным приоритетом. Распределение числа прерываний обслуживания для типа заявок определяется только загруженностью системы заявками с более высоким абсолютным приоритетом.

Условия имитационного моделирования задавались таким образом, чтобы при различных условиях существовала одна или более инварианта. В большинстве случаев такой инвариантой выступал общий коэффициент загрузки системы, конечно, за исключением, когда рассматривалось его влияние. Количество итераций выбиралось по принципу минимально необходимого для получения оценок достаточной точности. Достоверность оценок обеспечивалась по значению их стандартного отклонения в ряде экспериментов для заданного количества итераций. В большинстве случаев 10^6 экспериментов оказывается достаточным, разработка более чем 10^9 экспериментов возможна с точки зрения скорости работы, но не желательна, так как требует использования бо-

лее сложных базовых генераторов случайных чисел и подвержена сильному влиянию точности арифметических операций.

Результаты имитации совпадают с результатами численных расчётов полученных с помощью пакета МОСТ [4] и аналогичными имитационными моделями GPSS. В случае сравнения с МОСТом расчёты производились для различных одноканальных приоритетных систем, для которых имеется аналитическое решение в средних. Аналитическая модель позволяет вести расчёт в средних для одноканальных приоритетных систем со статическими приоритетами, простейшим входящим потоком и произвольным распределением времени обслуживания, заданным начальными моментами.

Удалось получить более быстрые (в сотни раз) процедуры расчёта конкретного класса задач по сравнению со специализированной системой расчёта имитационных моделей GPSS. Сбор характеристик, не учтённых в базовых отчётах GPSS, приводит к еще более значительному выигрышу в производительности. Разработанная имитационная модель позволяет учитывать практически все особенности модификаций в рамках моделирования приоритетных многоканальных систем: виды распределений, набор собираемой статистики, управление логикой поступления заявок в очередь и на обслуживание, логикой вытеснения.

Исследования выполнялись при финансовой поддержке гранта РФФИ № 07-07-00169, программы ОНИТ РАН (проект 2.3).

Литература

1. **Bramson M.** Stability of queuing networks. USA: Probability surveys, 2008.
2. **Рыжиков Ю. И.** Средние времена ожидания и пребывания в многоканальных приоритетных системах. Информационные управляющие системы. СПб., 2006.
3. **Алиев Т. И.** Основы моделирования дискретных систем. СПб: СПбГУ ИТМО, 2009.
4. **Рыжиков Ю. И.** Компьютерное моделирование систем с очередями. Курс лекций. СПб: ВКА им. А. Ф. Можайского, 2007.