УДК 004.724

ИЗУЧЕНИЕ МОДЕЛИ СМО С КВАНТОВАНИЕМ

ВРЕМЕНИ ОБСЛУЖИВАНИЯ

**В.А. Анохин,** студент четвертого курса РГРТУ, Рязань, Россия;

orcid.org/0000-0000-0000-000X, e-mail: anohin.va@yandex.ru

Рассматривается работа модели СМО с квантованием времени обслуживания. Целью работы является изучение влияния квантования времени обслуживания и назначение приоритетов на качество работы СМО. Искомыми зависимостями являются: зависимость загрузки от количества заявок; зависимость средней длины очереди от количества заявок и среднее время ожидания от количества заявок. Механизм квантования времени обслуживания отдельных задач применяется в многозадачных вычислительных системах, реализуя тем самым псевдопараллельность выполнения этих задач. Время обслуживания случайно и распределено по экспоненциальному закону.

**Ключевые слова:** система массового обслуживания, модель СМО с квантованием времени обслуживания, бесприоритетное и приоритетное обслуживание.

**DOI:**

Введение

Рассматривается нахождение зависимостей загрузки, средней длины очереди, среднего времени обслуживания путем моделирования системы массового обслуживания при помощи использования языка GPSS. Рассматриваемая система массового обслуживания состоит из трех источников заявок очереди и обслуживающего аппарата.

Теоретическая часть

*Система массового обслуживания* (СМО) — система, предназначенная для многократно повторяющегося (многоразового) использования при решении однотипных задач.

Механизм квантования времени обслуживания отдельных задач (заявок на обслуживание) обеспечивает не только псевдопараллельность выполнения задач, но и уменьшение среднего времени пребывания заявки в системе массового обслуживания за счет того, что менее трудоемкие заявки будут быстрее покидать СМО. Они не будут общее среднее время ожидания находиться в общей очереди.

Рассмотрим системы, которые обслуживают заявки от нескольких источников. Источники заявок (генераторы) вырабатывают заявки (транзакты) независимо друг от друга через случайные интервалы времени, распределенные по экспоненциальному закону (модель простейшего потока событий - ППС), и передают их на обработку в OPR. Поскольку OPR одновременно может обрабатывать только одну заявку, то остальные заявки помещаются в очередь QOPR, организованную на входе устройства OPR. Все заявки в очереди располагаются в соответствии со своим приоритетом.

Каждая заявка характеризуется временем обслуживания vi, необходимым для ее обслуживания в устройстве OPR, и приоритетом Pi. Время обслуживания также является случайной величиной, распределённой по экспоненциальному закону, что соответствует ППС обслуживания. На обслуживание каждой заявки устройство OPR выделяет квант процессорного времени q. Если обслуживаемая заявка не успевает выполниться за этот квант времени, то из времени обслуживания заявки vi вычитается квант времени q.

Процесс генерации, обслуживания и удаления заявок продолжается до тех пор, пока не истечет время, отведенное на моделирование, или не будет обслужено N заявок.

Экспериментальная часть

Исследуем модель системы массового обслуживания с квантованием времени обслуживания (рисунок 1).

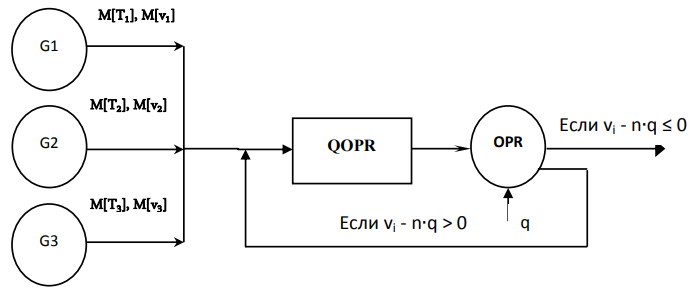


Рисунок 1 – Модель СМО

Исследуемая система массового обслуживания состоит из трех генераторов заявок G1, G2, G3, которые генерируют заявки с одинаковыми приоритетами. Интервалы между заявками Ti и необходимые времена обслуживания vi полагаются случайными с экспоненциальным законом распределения: M[T1]=15, M[T2]=50, M[T3]=100, M[v1]=5, M[v2]=10, M[v3]=40 единиц времени.

Значения vi записываются в первый параметр транзактов, моделирующих заявки. Каждому транзакту, занявшему OPR, выделяется квант процессорного времени, равный 1 единицы. После ухода транзакта из OPR из значения его первого параметра вычитается величина кванта. Если квант был меньше остатка времени обслуживания (заявка не будет обслужена полностью), транзакт возвращается в очередь в соответствии с приоритетом (контроль приоритетов осуществляется автоматически), в противном случае происходит обслуживание заявки на остаток времени обслуживания, происходит увеличение счетчика обслуженных заявок, а сама заявка уничтожается.

Программа на языке GPSS имеет вид:

**TIME TABLE M1,50,50,10** ;Задание таблицы табулирования времени пребывания транзакта в системе

;Задание (кусочно-линейное) функции экспоненты EXP1

**EXP1 FUNCTION RN1,C24**

**0,0/.1,.104/.2,.222/.3,.335/.4,.509/.5,.69/.6,.915**

**.7,1.2/.75,1.38/.8,1.6/.84,1.85/.88,2.12/.9,2.3**

**.92,2.52/.94,2.81/.95,2.99/.96,3.2/.97,3.5/.98,3.9**

**.99,4.6/.995,5.3/.998,6.2/.999,7/.9998,8**

**QUANT VARIABLE 1** ;Квант процессорного времени q=1

;Генератор заявок с приоритетом 1, M[T1]=20, M[v1]=4

**GENERATE 15,FN$EXP1,,,1** ;Генерируем заявки c приоритетом 1 через интервалы времени (EXP1,M[T1]=20)

**ASSIGN 1,5,EXP1** ;Присваиваем первому параметру транзакта (заявки) время обслуживания (EXP1, M[v1]=4)

**TRANSFER ,MET1** ;Пересылаем транзакт в очередь (безусловная передача транзакта на метку MET1)

;Генератор заявок с приоритетом 1, M[T2]=100, M[v2]=50

**GENERATE 50,FN$EXP1,,,1** ;Генерируем заявки c приоритетом 1 через интервалы времени (EXP1,M[T2]=100)

**ASSIGN 1,10,EXP1** ;Присваиваем первому параметру транзакта (заявки) время обслуживания (EXP1, M[v2]=50)

**TRANSFER ,MET1** ;Пересылаем транзакт в очередь

;Генератор заявок с приоритетом 1, M[T3]=50, M[v3]=10

**GENERATE 100,FN$EXP1,,,1** ;Генерируем заявки c приоритетом 1 через интервалы времени (EXP1,M[T3]=50)

**ASSIGN 1,40,EXP1** ;Присваиваем первому параметру транзакта (заявки) время обслуживания (EXP1, M[v3]=10)

**MET1 QUEUE QOPR** ;Вход транзакта в очередь

**SEIZE OPR** ;Занимаем устройство OPR

**DEPART QOPR** ;выход транзакта из очереди

**TEST L V$QUANT,P1,MET2** ;Если кванта не хватило q < остатка vi, то

**ADVANCE V$QUANT** ;Задерживаем заявку на время кванта q иначе не метку MET2

**RELEASE OPR** ;Освобождаем устройство OPR

**ASSIGN 1-,V$QUANT** ;Вычитаем из времени обслуживания заявки vi квант времени q

**TRANSFER ,MET1** ;Передаем недообслуженный транзакт в очередь

**FIN TABULATE TIME** ;занести значение времени пребывания транзакта в таблицу TIME

**TERMINATE 1** ;Регистрация обслуженной заявки

;(увеличение счетчика обслуженных заявок и удаление заявки)

**MET2 ADVANCE P1** ;Если кванта хватило, то задержка на остаток времени обслуживания

**RELEASE OPR** ;Освобождаем устройство OPR

**TRANSFER ,FIN** ;Конец обслуживания заявки

**START 200000**

Последний операторзадает время моделирования, путем задания начального значения счетчика завершений.

Промоделируем данную СМО для 200000 единиц времени. Зафиксируем результаты моделирования системы массового обслуживания. Результаты приведены на рисунке 2, где показан фрагмент отчета GPSS.

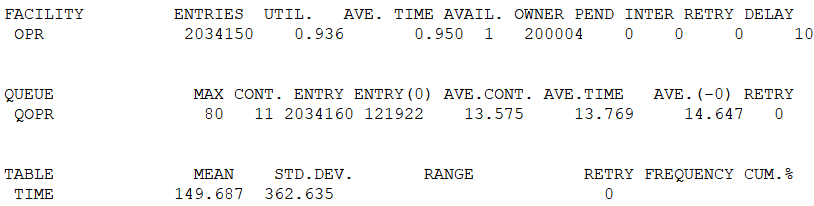


Рисунок 2 – Результаты моделирования

Рассчитаем теоретическое значение коэффициента загрузки обслуживающего аппарата и сравним его с экспериментальным значением.

Значение коэффициента загрузки обслуживающего аппарата вычисляется по формуле:

Результат:

Экспериментальное значение: ρ = 0,936

Снимем зависимость среднего времени пребывания заявки (транзакта) в системе от величины кванта q = 1, 4, 16, 64. Результат показан в таблице 1.

**Таблица 1 – Полученные экспериментальные данные**

|  |  |
| --- | --- |
| **q** | **mean** |
| 1 | 118,68 |
| 4 | 160,41 |
| 16 | 203,66 |
| 64 | 282,74 |

Построим график зависимости среднего времени пребывания заявки в системе от величины кванта. График показан на рисунке 3.

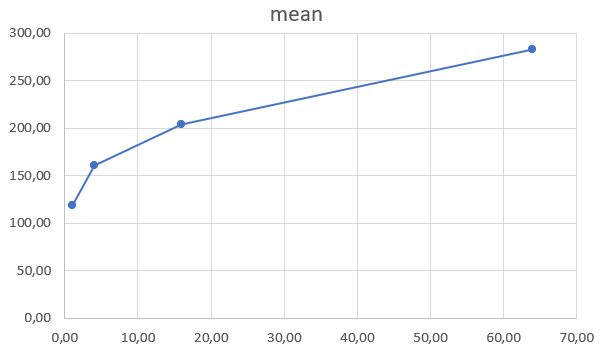


Рисунок 3 – Зависимость среднего времени пребывания

заявки в системе от величины кванта

Выключим в данной модели квантование времени обслуживания. Для этого зададим значение кванта на порядок больше, чем время обслуживания vi наиболее трудоемкой заявки.

При величине кванта q = 400 среднее время пребывания заявки в системе ū = 309,29. Данная величина превышает максимальный показатель времени из таблицы 1.

Полученные результаты показывают, что использование квантования времени обслуживания позволяет системе массового обслуживания работать более эффективно.

Чтобы изучить влияние приоритета обслуживания заявок установим q = 1 и назначим приоритеты так, чтобы более высокий приоритет был у заявок с меньший трудоемкостью. Запустим программу на выполнение. Время пребывания заявки в системе составляет ū = 88.91. Данная величина меньше минимального показателя времени из таблицы 1.

Полученные результаты показывают, что использование повышенного приоритета для заявок с меньшей длительностью выполнения позволяет повысить эффективность системы.

**Заключение**

В статье изучена зависимость основных параметров конкретной СМО замкнутого типа от количества заявок в системе на примере конкретных значений времени обслуживания.

Библиографический список

1. Афонин В.В., Федосин С.А., Моделирование систем. Практикум по GPSS/PC. –  
Саранск: Изд-во Мордов. Ун-та, 2001.   
2. Шрайбер Т. Дж. Моделирование на GPSS. М.:Машиностроение, 1980.  
3. Феррари Д. Оценка производительности вычислительных систем. М.: Мир, 1981.  
4. КудрявцевЕ. М. GPSS World. Основы имитационного моделирования различных  
систем. – М.: ДМК Пресс, 2004.– 320 с.: ил. (Серия «Проектирование»).  
5. Гнеденко Б.В, Коваленко И.Н. Введение в теорию массового обслуживания. – М.:  
Наука, 1987.  
6. Ермаков С.М., Мелос В.Б., Математический эксперимент с моделями сложных  
стохастических систем. – СПб.: Изд. ГУ, 1992.  
7. Колбанев М.О., Яковлев С.А. Модели и методы оценки характеристик обработки  
информации в интеллектуальных сетях связи. – СПб.: Изд-во Госуниверситета,  
2002.