# Оптимизация параметров обслуживания одноканальной СМО

# Цель работы

Изучение:

1. возможностей аналитических и ими­та­ци­он­ных моде­лей для исследо­вания стохастических систем,
2. методов и приемов проведения анализа и оптимизации параметров сто­ха­с­тических систем.

# ­ Теоретические сведения

Рассмотрим одноканальную СМО с ожи­да­нием. Будем предполагать, что входящий поток заявок на обслу­живание есть про­стей­ший поток с интенсив­но­с­тью ***λ***. Интенсивность потока обслуживания равна μ и может иметь произ­вольное ра­спределение. Заявка, пришедшая на вход СМО и заставшая канал сво­бодным, немедленно поступает на обслу­жи­ва­ние. Заявка, пришед­шая на вход СМО и заставшая канал занятым, ста­но­вит­ся в оче­редь на обслуживание и ожидает неограниченно долго освобож­дения прибора.

По дис­ци­пли­не обслуживания, под которым будем понимать правило (ал­горитм) выборки заявок из очереди на обслуживание. Наиболее распро­стра­не­ны дисциплины:

* *С упорядоченным обслуживанием*

Заявки обслу­жи­ваются в порядке по­ступления: первый при­шел - первым обслуживается (*First In First Out, FIFO,* или *First Come First Served, FCFS*),

* *С неупорядоченным обслуживанием*

Заявки обслуживаются в случай­ном порядке или

* *Со стековым обслуживанием*

Первой из очереди выбирается по­следняя заявка: последний пришел – пер­вым обслуживается, *Last In First Out, LIFO,* или *Last Come First Served, LCFS*.

* *С приоритетным обслуживанием*

Заявки разных типов (приоритетов) образуют как бы несколько оче­ре­дей, по одной очереди на каждое значение при­­о­ритета, и в каждый мо­мент, когда сер­вер заканчивает обс­лу­жи­вание какой-то заявки, на обс­лу­­живание на­пра­в­ля­­ет­­ся заявка из первой в порядке убывания приори­тета непустой очереди.

Приоритет заявки может быть

* *статически­м* (не меняющимся с течением времени пребывания в системе)
* *динами­ческим* (например, может увеличиваться по мере времени ожи­да­ния заявки).

Для времени ожидания заявки в очереди справедлива формула Хинчина-Полачека:

  (1)

где

  – среднее время обслуживания заявки,

– среднее время ожидания заявки в очереди,

σ*обсл*   – среднеквадратическое (стандарт­ное) отклонение времени обслуживания в приборе,

ρ   – коэффициент использования прибора, ρ= 

 – средний интервал времени между поступлением заявок,

*cобсл*– коэффициент вариации времени обслуживания 

Число заявок, ожидающих обслуживания (среднюю длина очереди), можно на­й­ти, умножив на величину *λ*:

 (2)

что, с учетом равенства



дает

 (3)

Формула Хинчина-Полачека используется для оценивания длин очередей при про­ек­ти­ровании информацион­ных систем. Она применяется в случае экспо­нен­ци­а­ль­ного распределения времени поступления при любом ра­сп­ре­делении вре­мени об­с­лу­­жи­ва­ния и любой дисциплине управления, лишь бы выбор оче­ред­но­го сообще­ния для об­служи­вания не зависел от вре­­ме­ни обс­лу­живания.

При проектировании систем встречаются такие ситуации воз­никновения очередей, ко­гда дисциплина обслуживания выбирается в зависимости от вре­мени обслуживания. На­при­мер, в некоторых слу­чаях для перво­оче­ре­дного об­слу­­жи­ва­ния могут выбираться более короткие сообще­ния с тем, чтобы по­лу­чить меньшее среднее время обслу­жи­вания (среднее время пребывания в заявки системе). При уп­ра­в­ле­нии линией свя­зи (каналом Интернет) мож­но при­своить вход­ным сообщениям бо­лее вы­сокий приоритет, чем выходным, поскольку пер­вые короче. В таких случаях уже не­об­­хо­­­ди­мо использовать не урав­нение Хинчина — Пол­ла­че­ка или произ­вод­ные от не­го, а более сло­жные уравнения или использовать метод имитационного моделирования.

Особый интерес для практических применений имеют два особых случая.

1) Время обслуживания постоянно.

При регулярном характере потока рассеяние отсутствует, поэтому сред­не­ква­д­рати­чес­кое отклонение, и формулы (1), (2) прео­бра­зу­ются в



и

 (4)

2) Время обслуживания имеет экспоненциальное распределение.

В случае экспоненциального распределения, как известно, сре­дне­ква­драти­чес­кое отклонение, поэтому (1), (2) приобретают вид:



и

 (5)

Большинство значений времени обслуживания в информационных систе­мах лежит где-то между этими двумя случаями. Вре­мена обслу­жи­вания, рав­ные постоянной величине, встречаются редко. Даже время доступа к твер­дому диску непосто­ян­­­но из-за различного положения массивов с данными на по­верхности. Од­ним из при­ме­ров, иллюстрирующих случай по­­сто­­­ян­­­ного вре­мени обслуживания может служить занятие линии связи для передачи сообще­ний фиксированной длины.

С другой стороны, разброс времени обслуживания обычно не так ве­лик, как в случае произ­вольного или экспонен­ци­аль­ного его ра­с­пределения, т.е.,  ре­д­ко достигает зна­че­ний . Этот случай иногда считают наихудшим и пото­му по­­ль­зу­ют­ся фо­р­му­ла­ми, от­носящимися к экспо­не­н­­­ци­а­льному расп­ре­де­ле­нию вре­мен обслу­жи­вания. Такой рас­чет может дать несколько за­вы­ше­н­ные раз­меры очередей и времен ожидания в них, но эта ошибка, по край­ней мере, не опасна.

Экс­поненциальное распределение времен обслуживания не наихудший слу­­чай, с которым приходится иметь де­ло в дей­стви­те­ль­ности. Однако, если вре­ме­на об­служивания, полученные при расчете очередей, оказы­ваются распре­де­лен­ны­ми хуже, чем времена с экспо­нен­циальным распределением, то это является пре­до­­­сте­ре­гающим сигналом для раз­ра­бот­чика. Если стандартное отклонение бо­льше сре­днего значения, то обычно воз­никает необ­ходимость в коррекции расчетов.

*Пример. Пусть имеются шесть типов сообщений со временами об­слу­­жива­ния 15, 20, 25, 30, 35, 40 и 300. Чис­ло сообщений ка­ждого типа одинаково. Ста­­н­дартное от­кло­­нение указанных времен несколько выше их среднего. Зна­че­­ние по­­­­сле­­д­него времени обслу­жи­­ва­ния намного бо­ль­­­ше других. Это приве­дет к тому, что со­общения будут нахо­диться в оче­ре­ди значительно дольше, чем, если бы вре­мена об­­слу­живания были одного порядка. В таком случае при про­ек­ти­ро­ва­нии це­ле­­­со­об­ра­зно принять меры для уменьшения длины очереди. На­­пример, если ука­занные ци­ф­ры связаны с дли­на­ми со­об­ще­ний, то, возмож­но, очень длинные со­об­щения стоит ра­зделить на части.*

*Получим количественную оценку системных показателей для перво­на­ча­льного ва­рианта организации информационного обмена и для двух вари­ан­тов, в* которых *длины сообщений передаваемых сообщений выровнены.*

*Вычислим величины, входящие в формулу для средней длины очереди. Бу­дем счи­тать, средняя продолжительность для всей группы сообщений со­ста­в­ля­ет 600 временных единиц. Тогда*

*коэффициент загрузки*

**

*математическое ожидание времени обслуживания*

**

*средне­ква­д­рати­чес­кое отклонение времени обслуживания*

**

*коэффициент вариации времени обслуживания*

**

*Подставляя в (*1)получим*, что средний размер очереди в первона­чаль­ном варианте будет равен*

**

*Вычислим теперь этот показатель для случая, когда последнее сооб­ще­ние раз­би­вается на два* сообщения *одинаковой длины. Поступая аналогич­ным об­ра­зом, будем иметь (очевидно, что значение остается тем же):*

**

**

**

**

*Разбивая длинное сообщение на три части, получаем для средней длины оче­ре­ди значение*



# ­ Содержание работы

1. Запустите *MS Excel* и подготовьте таблицу для занесения про­ме­жуточных расчетов и итоговых значений; формат таблицы показан на Рис. 1.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| АНАЛИТИЧЕСКАЯ МОДЕЛЬ | | | | | | | | | | | | | | | | | | | ИМИТАЦИОННАЯ МОДЕЛЬ | | | | | |
| Без приоритета | | | С приоритетом | | |
| *t*Σ | N | *t*a | V1 | V2 | V3 | V4 | V5 | V6 | V7 | V8 | V9 | M | D |  | cs^2 | n\_0 | n\_e | n\_d | n\_0 | n\_e | n\_d | n\_0 | n\_e | n\_d |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

Рис. 1 Формат основной таблицы

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| АНАЛИТИЧЕСКАЯ МОДЕЛЬ | | | | | | | | | | | | | | | | | | | ИМИТАЦИОННАЯ МОДЕЛЬ | | | | | |
| Без приоритета | | | С приоритетом | | |
| TΣ | N | Ta | V1 | V2 | V3 | V4 | V5 | V6 | V7 | V8 | V9 | M | D |  | cs^2 | nw\_0 | nw\_e | nw\_d | nw\_0 | nw\_e | nw\_d | nw\_0 | nw\_e | nw\_d |
| 600 | 7 | 85,71 | 15 | 20 | 25 | 30 | 35 | 40 | 300 |  |  | 66,43 | 9155,10 | 0,78 | 2,07 | **4,10** | **2,67** | **1,33** | **4,12** | **2,57** | **1,31** | **1,84** | **1,30** | **1,28** |
|  |  |  | 225 | 400 | 625 | 900 | 1225 | 1600 | 90000 |  |  | 13567,86 |  |  |  |  |  |  |  |  |  |  |  |  |
| 600 | 8 | 75,00 | 15 | 20 | 25 | 30 | 35 | 40 | 150 | 150 |  | 45,00 | 1900,00 | 0,60 | 0,94 | **0,87** |  |  | **0,84** |  |  | **0,57** |  |  |
|  |  |  | 225 | 400 | 625 | 900 | 1225 | 1600 | 22500 | 22500 |  | 3925,00 |  |  |  |  |  |  |  |  |  |  |  |  |
| 600 | 9 | 66,67 | 15 | 20 | 25 | 30 | 35 | 40 | 100 | 100 | 100 | 37,86 | 706,12 | 0,57 | 0,49 | **0,56** |  |  | **0,56** |  |  | **0,42** |  |  |
|  |  |  | 225 | 400 | 625 | 900 | 1225 | 1600 | 10000 | 10000 | 10000 | 2139,29 |  |  |  |  |  |  |  |  |  |  |  |  |

Рис. 2 Пример заполненной таблицы

*t*Σ  средняя продолжительность пакета сообщений (за­да­ет­ся в исход­ных данных задания)

*N* число сообщений в группе (в исходном варианте *N*=7)

*ta* средний интервал между сообщениями в потоке сообщений *T*Σ /*N*

*Vi*  длины сообщений в пакете сообщений (для первого варианта ра­сче­та задается как исход­ные данные к заданию) – первая строка,

квадраты длин сообщений в пакете сообщений – вторая строка (для расчета значения дисперсии)

*S*  итоговая сумма длин сообщений в пакете сообщений 

*M* математическое ожидание длин сообщений *S*/*N*

*D* дисперсия длин сообщений 

σ среднеквадратическое отклонение длин сообщений 

ρ коэффициент загрузки (*M* / *T*a)

*cs*^2 квадрат коэффициента вариации времени обслуживания 

*n\_a* расчетная величина средней длины очереди

*n\_*0 экспериментально полученная величина средней длины очереди для дисциплины обслуживания без приоритетов (*FIFO*)

*n\_*1 экспериментально полученная величина средней длины очереди для ди­с­циплины обслуживания с приоритетным обслуживанием корот­ких сообщений

1. В столбцах для параметров таблицы (*t*Σ, *N*, *V*1, …, *V*7) запишите исходные данные для своего варианта (номер варианта соответствует порядковому номеру в списке группы).
2. Под первой строкой разместите строку с квадратами значений длин сообщений, суммой квадратов длин и математического ожидания.
3. Заполните соответствующие клетки таблицы для расчета показателей мо­де­ли СМО. В столбцы с показателями аналитической модели впишите соот­вет­ст­вую­щие формулы.
4. Повторите шаги 2–4 с числом сообщений *N*=8 и *N*=9, разбивая последнее сообщение соответственно на два и три сообщения равной длины.
5. Проведите запуски программной модели для исходного варианта расп­ре­де­ле­­ния длин сообщений для бесприоритетной дисциплины обслужива­ния.

Запустите приложение *MS Visual Studio* и создайте проект (*Приложение B*. Сборка и настройка проекта *С++*), используя в качестве исходного модуля нужный шаблон (см. *Приложение А*. Модули исходных текстов для С++-проектов).

Откройте текст исходного модуля в и внимательно его изучите.

Установите в нужное значение параметр режима обслу­жи­ва­ния и за­­дайте для каждого запуска массивы, описывающие распреде­ление интер­ва­­лов (все элементы массива частот удобнее задавать равны­ми единице, эле­­менты массива значений задаются значениями *V*1, …, *V*7). Результаты за­пу­с­ков внесите в таблицу.

1. Повторите эксперимент для обслуживания с приоритетом коротких со­об­щений, переустановив в нужное значение параметр режима обслужива­ния, внесите результаты за­пу­с­ков в таблицу.
2. Повторите эксперимент (шаги 1 и 2) для *N*=8 и *N*=9.

# ­ Отчет по работе

Отчет по работе должен включать

* исходные данные,
* результаты расчетов,
* результаты эк­с­пе­ри­мен­тов с программной моделью

сведенные в таблицу по образцу Рис. 10‑2.

# ­ Контрольные вопросы

* 1. Дайте краткое описание модели СМО с неограниченной очередью.
  2. Какими показателями характеризуется функционирование СМО с не­ог­ра­ни­ченной очередью?
  3. Что такое дисциплина обслуживания? Назовите примеры наиболее извест­ных дисциплин.
  4. Что позволяет определить формула Хинчина-Полачека? Для каких случаев справедлива формула?
  5. Каким образом можно определить показатели СМО для случаев, преду­смат­ривающих использование приоритетных дисциплин обслуживания?
  6. Какие распределения времени обслуживания представляют наибольший ин­терес для практики? Опишите эти случаи.
  7. От каких факторов зависят в основном значения показателей СМО с не­ог­ра­­ниченной очередью? Как можно влиять на эти факторы для улуч­шения показателей?

# ­ Варианты исходных данных

| № | *t*пост | *t*обсл | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|  | 2900 | 200 | 220 | 235 | 240 | 280 | 290 | 960 |
|  | 3000 | 190 | 220 | 240 | 250 | 270 | 280 | 930 |
|  | 2300 | 130 | 135 | 140 | 150 | 160 | 175 | 870 |
|  | 1800 | 120 | 130 | 135 | 140 | 150 | 175 | 600 |
|  | 3300 | 260 | 295 | 310 | 320 | 330 | 340 | 960 |
|  | 1900 | 120 | 140 | 130 | 135 | 150 | 175 | 600 |
|  | 1700 | 140 | 120 | 130 | 135 | 150 | 155 | 570 |
|  | 2800 | 220 | 240 | 230 | 210 | 195 | 240 | 930 |
|  | 1500 | 95 | 100 | 110 | 115 | 90 | 85 | 570 |
|  | 2100 | 120 | 140 | 150 | 135 | 130 | 155 | 780 |
|  | 2700 | 230 | 220 | 210 | 240 | 205 | 200 | 960 |
|  | 4200 | 395 | 380 | 400 | 410 | 415 | 405 | 1200 |
|  | 3600 | 320 | 340 | 330 | 295 | 260 | 310 | 1200 |
|  | 2200 | 115 | 110 | 120 | 125 | 140 | 145 | 930 |
|  | 1700 | 90 | 100 | 95 | 105 | 110 | 120 | 660 |
|  | 2500 | 120 | 140 | 130 | 135 | 150 | 155 | 990 |
|  | 5000 | 442 | 451 | 456 | 459 | 473 | 484 | 1500 |
|  | 2100 | 153 | 161 | 165 | 173 | 180 | 186 | 720 |
|  | 1800 | 130 | 135 | 140 | 145 | 150 | 155 | 600 |
|  | 2800 | 215 | 220 | 240 | 230 | 210 | 195 | 930 |
|  | 2700 | 150 | 135 | 160 | 175 | 185 | 190 | 990 |
|  | 1800 | 100 | 110 | 120 | 140 | 130 | 135 | 750 |
|  | 3500 | 243 | 254 | 266 | 271 | 280 | 283 | 1320 |
|  | 2200 | 164 | 174 | 178 | 180 | 186 | 182 | 810 |
|  | 2300 | 145 | 155 | 160 | 165 | 175 | 185 | 870 |

# ­ Задания для самостоятельного решения

1. Сборочный участок производит в один час в среднем 90 блоков. На этом участке ра­ботает контролер, который проверяет все собран­ные блоки, сре­д­няя про­должительность контрольных опе­ра­ций составляет 1,25 ми­нут. Опре­де­лить среднее число изделий, ожидающих внимания конт­ро­лера, и среднее вре­мя ожида­ния изделиями в очереди на проверку.
2. Железнодорожную станцию обслуживает касса с одним окном. В летние выход­ные дни интен­сивность потока пассажиров к кассе составляет 0,45 человек в минуту. Кассир затра­чивает на об­луживания пассажира в сред­нем 2 мин. Определить среднее число пассажиров у кассы и среднее вре­мя, затрачиваемое пассажиром на приобретение билета.
3. Порт имеет один грузовой причал для разгрузки судов. Интенсивность потока судов со­став­ляет 1,5 судов в сутки. Разгрузка судна требует в сре­днем 0,5 суток. Пре­д­по­ла­гается, что суда могут ждать в очереди неогра­ни­ченно долго. Найти показатели эффек­тив­ности ра­бо­ты причала.
4. К базе данных сервера продажи билетов поступает 10 запросов в минуту. Среднее вре­мя об­работки каждого запроса со­­­став­ляет 1 секунду. Запрос, поступивший в мо­мент обработки одного из ранее поступивших запросов, становится в очередь. Опре­де­лить среднюю длину очереди и среднее вре­мя, которое проведут запросы до обслуживания.