

Г.С. Осипов

МАТЕМАТИЧЕСКОЕ И ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ СИСТЕМ МАССОВОГО ОБСЛУЖИВАНИЯ

Сетевое научное издание

Рекомендовано УМО РАЕ
по классическому университетскому
и техническому образованию
в качестве учебного пособия для студентов
высших учебных заведений,
обучающихся по направлению подготовки
01.03.02 — «Прикладная математика
и информатика»

© Осипов Г.С., 2017

© ИД «Академия Естествознания»

© АНО «Академия Естествознания»

ISBN 978-5-91327-459-5

Москва
2017

УДК 519.872.8

ББК 22.18

О74

Осипов Г.С.

О74 Математическое и имитационное моделирование систем массового обслуживания. — М.: Издательский дом Академии Естествознания, 20176. — 56 с.

ISBN 978-5-91327-459-5

DOI 10.17513/np.234

Рассмотрены базовые основы математического и имитационного моделирования систем массового обслуживания. Проведена классификация типов систем обслуживания, определяемая классом решаемых задач и в свою очередь определяющая фундаментальные основы используемого математического аппарата. В едином комплексе реализовано взаимоувязанное решение задач методами классического математического моделирования с дискретно-событийной парадигмой имитационного моделирования. Большое внимание уделено решению практически значимых задач теории массового обслуживания в аспекте синергетического единства двух подходов к моделированию сложных трудно формализуемых систем. Освещены основные методологические принципы разработки приложений в среде имитационного моделирования *AnyLogic*.

Для студентов, обучающихся по направлению подготовки 01.03.02 «Прикладная математика и информатика». Может быть использовано в качестве учебного пособия студентами и аспирантами других направлений подготовки занимающимся моделированием, разработкой и исследованием систем массового обслуживания.

ISBN 978-5-91327-459-5

© Осипов Г.С., 2017

© ИД «Академия Естествознания»

© АНО «Академия Естествознания»

ОГЛАВЛЕНИЕ

Введение	4
1. Системы массового обслуживания с отказами	6
1.1. Математическая модель	6
1.2. Практическая реализация в среде имитационного моделирования	7
2. Системы массового обслуживания с ожиданием	10
2.1. Моделирование и оптимизация одноканальных СМО	10
2.2. Многоканальные СМО с ожиданием	17
2.3. Пример решения оптимизационной задачи	18
3. Системы массового обслуживания смешанного типа	23
3.1. Системы с ограниченной длиной очереди	23
3.1.1. Одноканальные СМО	23
3.1.2. Многоканальные СМО	25
3.1.3. Примеры практической реализации	27
3.2. Системы с ограниченным периодом ожидания	31
3.2.1. Необходимые сведения и аналитические выражения	31
3.2.2. Имитационное моделирование	33
4. Системы массового обслуживания с приоритетными заявками	40
4.1. Материал и методика	40
4.2. Практическая реализация	43
5. Модели с не пуассоновскими потоками	47
5.1. Основные сведения и формальные зависимости	47
5.2. Результаты имитационного моделирования	49
Список использованной литературы	54

ВВЕДЕНИЕ

По характеру обслуживания заявок все системы массового обслуживания (СМО) делятся на три типа [1]:

- 1) системы с отказами;
- 2) системы с ожиданием;
- 3) системы смешанного типа.

СМО первого типа характеризуются тем, что поступившие в данный момент времени заявка тут же принимается к обслуживанию, если имеется хотя бы один свободный канал, или получает отказ, если все каналы заняты. К системам с отказами относятся, например, аварийно-спасательные службы, в которых заявки (на оказание помощи) не могут становиться в очередь и ожидать освобождения средства спасания. Сюда также относятся информационные системы оперативного реагирования, в которых недопустима потеря актуальности информации.

В системах с ожиданием заявка в случае занятости всех каналов обслуживания становится в очередь и ожидает освобождения одного из них. Это наиболее представительный на практике класс систем. Типичным примером СМО с ожиданием являются системы с очередью типа *FIFO*, характерные для обслуживания покупателей в торговых центрах, последовательной пакетной обработкой информации и т.п.

Системам смешанного типа присущи особенности двух вышеописанных систем. Заявка в такой системе становится в очередь, если в ожидании находится ограниченное число заявок; если это число превышает допустимое ограничение, то заявка покидает систему (получает отказ). Ограниченной может быть и длительность ожидания. Принцип ограниченного времени ожидания (очереди) используется, например, при планировании обслуживания заявок на терминалах (грузовых, пассажирских, банковских). Превышение определенного

числа заявок, а также длительности ожидания, последствиями которых может быть потеря качества продукции или обслуживания, приводят к их перераспределению на другие терминалы.

По дисциплине обслуживания заявок выделяют системы без приоритета и с приоритетом в которых устанавливается определенная очередность обслуживания заявок.

Важнейшей характеристикой СМО является входящий поток заявок. Число заявок в единицу времени, обычно называют интенсивностью их поступления, или частотой и обозначают λ .

Если входящий поток является простейшим, то достаточно знать лишь один его параметр λ (или интервал между поступлением заявок $t_{\text{и}} = 1/\lambda$). В этом случае поток заявок распределен по закону Пуассона, а интервалы — по показательному закону, т.е. коэффициент вариации интервалов $v_{\lambda} = 1$.

В случае регулируемого потока, т.е. потока с последействием, для определения количественных характеристик обслуживания необходимо знать закон распределения входящего потока, в частности, для потока, описываемого двухпараметрическим распределением, необходимо знать среднее значение (λ или $t_{\text{и}}$) и дисперсионную характеристику интервала входящего на обслуживание потока (обычно v_{λ}).

Система обслуживания характеризуется числом каналов обслуживания n , длительностью обслуживания $t_{\text{об}}$ одной заявки и пропускной способностью μ — числом заявок, которое может обслужить один канал (или n каналов) в единицу времени.

Отношение

$$\rho = \frac{\lambda}{\mu} = \frac{t_{\text{об}}}{t_{\text{и}}}$$

называется коэффициентом использования пропускной способности системы, или приведенной плотностью потока.

1. СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ С ОТКАЗАМИ

1.1. Математическая модель

Имеется n каналов на которые поступает поток заявок с интенсивностью λ . Поток обслуживания имеет интенсивность μ , коэффициент использования пропускной способности системы оценивается величиной

$$\rho = \frac{\lambda}{\mu}.$$

Состояния системы S (СМО) удобно нумеровать по числу заявок, находящихся в системе, которое совпадает с числом занятых каналов:

S_0 — в СМО заявок нет;

S_1 — в СМО находится одна заявка;

S_2 — заняты два канала, остальные свободны;

...

S_k — в СМО находится k заявок (k каналов заняты, остальные свободны);

...

S_n — в СМО находится n заявок (заняты все n каналов).

Граф состояний СМО соответствует процессу гибели и размножения [1] и показан на рис. 1.1.

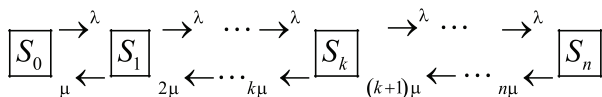


Рис. 1.1. Граф состояний СМО с отказами

Если потоки простейшие, то справедливы формулы Эрланга, на основе которых определяются основные показатели функционирования СМО, представленные в табл. 1.

Таблица 1

Показатели эффективности СМО

Показатель	Наименование	Формула
p_0	Предельная вероятность простоя	$p_0 = \left(\sum_{k=0}^n \frac{\rho^k}{k!} \right)^{-1}$
p_k	Предельные вероятности состояний	$p_k = \frac{\rho^k}{k!} p_0 \quad (k = \overline{1, n})$
$P_{\text{отк}}$	Вероятность отказа СМО (предельная вероятность того, что все n каналов системы будут заняты)	$P_{\text{отк}} = p_n = \frac{\rho^n}{n!} p_0$
Q	Относительная пропускная способность (вероятность того, что заявка будет обслужена)	$Q = P_{\text{обсл}} = 1 - P_{\text{отк}} = 1 - \frac{\rho^n}{n!} p_0$
A	Абсолютная пропускная способность (число заявок, обслуживаемых в единицу времени)	$A = \lambda Q = \lambda \left(1 - \frac{\rho^n}{n!} p_0 \right)$
$\bar{k}$	Среднее число занятых каналов	$\bar{k} = \frac{A}{\mu} = \rho \left(1 - \frac{\rho^n}{n!} p_0 \right) = \rho (1 - p_n)$

1.2. Практическая реализация в среде имитационного моделирования

Исследуем *пример* моделирования СМО с отказами [2].

Требуется определить минимально необходимое число аварийно-спасательных судов на водохранилище. Критерием является обеспечение вероятности отказа не более 10 %.

Известно, что в сутки в среднем поступает $\lambda = 2$ вызова на оказание помощи. Поток заявок носит случайный характер и может быть описан законом Пуассона, а длительность обслуживания аппроксимируется показательным распределением. Таким образом коэффициенты вариации интервалов входящего потока заявок и интервалов обслуживания, соответственно, равны единице: $v_\lambda = v_\mu = 1$.

Из статистических данных известно, что одно аварийно-спасательное судно может за сутки обслужить $\mu = 4$ заявки.

Решение.

Очевидно

$$\rho = \frac{\lambda}{\mu} = \frac{1}{2}.$$

Если поставить на выполнение работ одно ($n = 1$) судно, то вероятность отказа в обслуживании:

$$P_{\text{отк}} = \rho p_0 = \frac{1}{2} \left(1 + \frac{1}{2} \right)^{-1} = \frac{1}{3}.$$

Очевидно, что 33 % отказов в обслуживании – слишком большая величина, поэтому проведем расчет для двух ($n = 2$) судов:

Вероятность простоя:

$$p_0 = \left(\sum_{k=0}^n \frac{\rho^k}{k!} \right)^{-1} = \left(1 + \frac{1}{2} + \frac{(1/2)^2}{2} \right)^{-1} = \frac{8}{13}.$$

Вероятность отказа:

$$P_{\text{отк}} = \frac{\rho^n}{n!} p_0 = \frac{(1/2)^2}{2!} \cdot \frac{81}{2133} = \frac{1}{1} \approx 0,077 \text{ (8\%)}. \quad \text{2133}$$

Относительная пропускная способность:

$$Q = 1 - P_{\text{отк}} = \frac{12}{13}.$$

Абсолютная пропускная способность:

$$A = \lambda Q = 2 \cdot \frac{12}{13} = \frac{24}{13}.$$

Среднее число занятых судов определяем по формуле:

$$\bar{k} = \frac{A}{\mu} = \frac{24}{13} \cdot \frac{1}{4} = \frac{6}{13}.$$

Среднее время пребывания заявки в СМО:

$$T_s = Q \cdot \bar{t}_{об} = \frac{Q}{\mu} = \frac{12}{13} \cdot \frac{1}{4} = \frac{3}{13} \text{ сутки.}$$

На рис. 1.2 приведена принципиальная схема рассматриваемой системы в среде имитационного моделирования *AnyLogic* [2].



Рис. 1.2. Схема имитационной модели

Представленный фрагмент решения задачи свидетельствует о том, за определенный промежуток времени всего поступило 30 заявок на обслуживание, из которых одна получила отказ в те периоды, когда аварийно-спасательные суда были заняты.

На рис. 1.3 представлен временной график изменения количества аварийно-спасательных судов, задействованных в проведении спасательных мероприятий и гистограмма плотности вероятности распределения (с функцией распределения и средним значением) времени, затраченного на обслуживание.

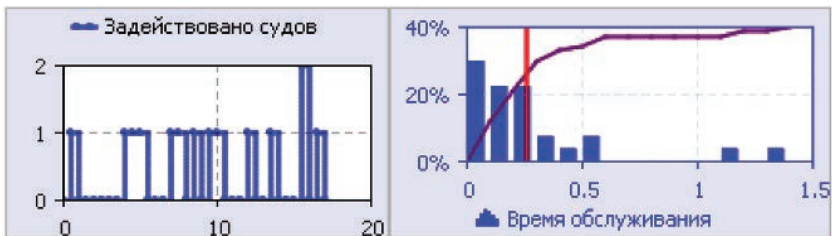


Рис. 1.3. Показатели функционирования СМО

2. СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ С ОЖИДАНИЕМ

Системы обслуживания с ожиданием являются наиболее распространенными и часто встречающимися на практике СМО.

2.1. Моделирование и оптимизация одноканальных СМО

На рис. 2.1 представлен граф состояний для одноканальной системы, где для удобства состояния системы $s_0, s_1, \dots, s_k, \dots$ нумеруются по числу заявок, находящихся в СМО.

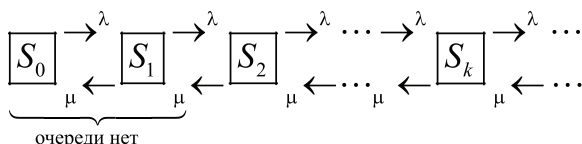


Рис. 2.1. Граф состояний для одноканальной СМО с неограниченной очередью

Система может находиться в одном из состояний:

s_0 — канал свободен;

s_1 — канал занят (обслуживает заявку), очереди нет;

s_2 — канал занят, одна заявка стоит в очереди;

...

s — канал занят, $(k - 1)$ заявок стоит в очереди;

...

По всем стрелкам поток заявок с интенсивностью λ переводит систему слева направо, а справа налево — поток обслуживания с интенсивность μ .

Основные характеристики рассматриваемых систем сведены в табл. 2 [3].

Таблица 2

Показатели работы СМО

№ п/п		Наименование	Характеристика/формула
1	λ	Интенсивность потока заявок	Среднее число заявок, поступающих в СМО в единицу времени [число заявок/единицу времени]
2	μ	Интенсивность потока обслуживания	Среднее число заявок, обслуживаемых в единицу времени [число заявок/единицу времени]
3	ρ	Приведенная интенсивность потока заявок (интенсивность нагрузки канала)	Среднее число заявок, приходящее за среднее время обслуживания одной заявки $\rho = \frac{\lambda}{\mu} < 1$
4	L_s	Среднее число заявок в системе	$L_s = \frac{\lambda}{\mu - \lambda} = \frac{\rho}{1 - \rho}$
5	L_q	Среднее число заявок в очереди (длина очереди)	$L_q = \frac{\rho^2}{1 - \rho} = \frac{\rho\lambda}{\mu - \lambda}$
6		Среднее число заявок в канале (обслуживания)	$L_s - L_q = \rho$
7	T_s	Среднее время пребывания заявки в системе	$T_s = \frac{L_s}{\lambda} = \frac{\rho}{\lambda(1 - \rho)} = \frac{1}{\mu(1 - \rho)} = \frac{1}{\mu - \lambda}$
8	T_q	Среднее время ожидания в очереди	$T_q = \frac{L_q}{\lambda} = \frac{\rho^2}{\lambda(1 - \rho)} = \frac{\rho}{\mu - \lambda} = \frac{\rho}{\mu(1 - \rho)}$
9		Среднее время обслуживания	$T_s - T_q = \frac{L_s - L_q}{\lambda} = \frac{\rho}{\lambda} = \frac{1}{\mu}$
10	γ	Отношение времени ожидания в очереди к времени обслуживания	$\gamma = \frac{T_q}{T_s - T_q} = \frac{L_q}{L_s - L_q} = \frac{\rho}{1 - \rho} = L_s$

Исследуем одноканальные системы массового обслуживания с очередью, на которую не наложено ограничений (ни по длине очереди, ни по времени ожидания) на примере входящего судопотока в морском порт, имеющий один канал — специализированный грузовой терминал и рейд для ожидания судов очереди на выполнение погрузочно-разгрузочных работ у терминала.

При проектировании новых портов с терминалом для обслуживания перспективного грузопотока (судопотока), а также для повышения эффективности функционирования существующих терминалов целесообразно предварительно построить имитационную модель системы, исследуя которую можно найти оптимальные параметры функционирования реальной системы. Таким образом, имитационная и оптимизационная составляющая модели должны составлять единый взаимоувязанный комплекс, обеспечивающий эффективное решение задачи минимизации приведенных совокупных затрат по эксплуатации терминала и по содержанию судов во время их пребывания в очереди и у терминала.

Моделирование систем массового обслуживания подчинено парадигме дискретно-событийного имитационного моделирования. Это совокупность операций или процессов с заявками, подчиненных причинно-следственным связям. На рис. 2.2 представлена процессная диаграмма одноканальной СМО [3].

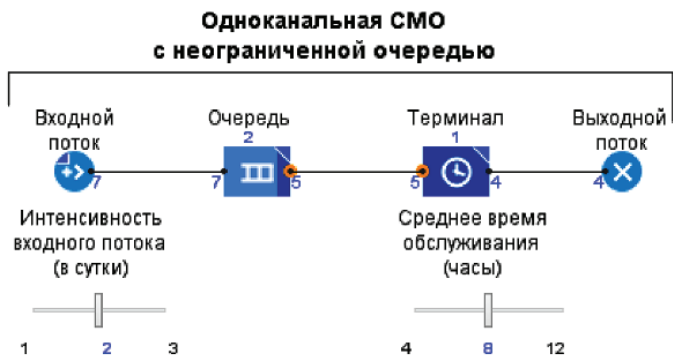


Рис. 2.2. Принципиальная схема одноканальной СМО

В одноканальной СМО с неограниченной очередью оптимизация подразумевает обеспечение такого соотношения интенсивностей поступления заявок (λ) и их обслуживания (μ) при котором обеспечивается минимум суммарных (совокупных) затрат по ожиданию заявок в очереди и обслуживания в канале. Очевидно, эти совокупные затраты f определяются приведенной интенсивностью потока заявок $\rho = \frac{\lambda}{\mu}$ (см. табл. 2), т.е. $f = f(\rho)$. Ясно, что стремление минимизировать затраты, связанные с ожиданием в очереди, приводит к увеличению расходов в канале (обслуживания). И наоборот минимизация затрат на обслуживание заявок (увеличение интенсивности нагрузки канала) приведет к возрастанию времени (и соответственно затрат) ожидания в очереди.

В рассматриваемой проблеме параметром оптимизации является интенсивность нагрузки терминала, а целевой функцией – суммарные приведенные затраты по судам и терминалу [4].

Минимизация расходов по терминалу возможна при условии его полного использования (отсутствия свободных резервов), а это, в свою очередь, ведет к возрастанию простоев судов. С другой стороны, минимизация расходов по судам может быть обеспечена за счет увеличения пропускной способности терминала, однако это, очевидно, связано с дополнительными затратами по устройству и содержанию терминала

Затраты по судам могут быть найдены так:

$$f_s = \lambda(t_{\text{гп}} + t_{\text{ож}} + t_T) Z_s = \lambda(t_{\text{гп}} + \gamma(\rho)t_{\text{гп}} + k_T t_{\text{гп}}) Z_s = \lambda t_{\text{гп}} (1 + \gamma(\rho) + k_T) Z_s, \quad (1)$$

где $t_{\text{гп}}$, $t_{\text{ож}}$, t_T – соответственно длительность грузовых операций, их ожидания и технических операций;

Z_s – затраты по судну в сутки;

k_T – коэффициент, учитывающий время на выполнение технических операций на судне.

Соответственно, затраты по терминалу:

$$f_T = k\rho Z_T^p(\Pi) + (1 - k\rho)Z_T^{np}(\Pi), \quad (2)$$

где k – коэффициент, учитывающий перерывы между сменами, различные технологии работы и т.д.;

Π – пропускная способность терминала (т/сут).

Z_T^p, Z_T^{np} – затраты по терминалу за время работы и простоя в сутки, соответственно.

Сложив выражения (1) и (2) и разделив на планируемый грузооборот $G = \rho\Pi$ получим суммарные приведенные затраты по судам и терминалу (УЕ/т):

$$\begin{aligned} f &= \frac{k\rho Z_T^p + (1 - k\rho)Z_T^{np} + \lambda t_{\text{тп}}(1 + \gamma(\rho) + k_T)Z_s}{\Pi\rho} = \\ &= \frac{kZ_T^p + \left(\frac{1}{\rho} - k\right)Z_T^{np} + (1 + \gamma(\rho) + k_T)Z_s}{\Pi}. \end{aligned}$$

Таким образом в среде *AnyLogic* решается следующая экстремальная задача:

$$\begin{aligned} (D, f): f(\rho) &= \frac{kZ_T^p(\Pi) + \left(\frac{1}{\rho} - k\right)Z_T^{np}(\Pi) + (1 + \gamma(\rho) + k_T)Z_s}{\Pi} \rightarrow \min; \\ D &= \{\rho \in R_+^n : \rho < 1\}. \end{aligned}$$

На рис. 2.3 представлены графики затрат по судну, терминалу их сумма и приведенных затрат в зависимости от интенсивности нагрузки терминала.

Увеличение пропускной способности терминала ($\Pi > G$) приводит к увеличению затрат по терминалу Z_T^p и Z_T^{np} и как следствие к уменьшению значения $\gamma = \gamma(\rho)$, а следовательно и затрат по судам.

На рис. 2.4 представлена принципиальная расчетная схема определения основных показателей функционирования системы.

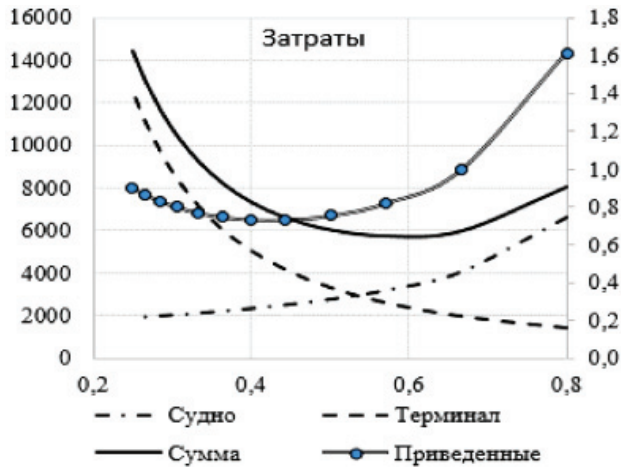


Рис. 2.3. Зависимость затрат в системе от загрузки терминала

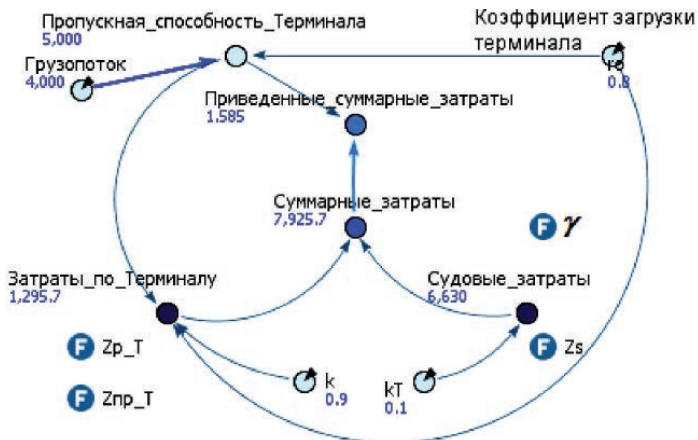


Рис. 2.4. Расчетная схема

Фрагмент результатов оптимизационного эксперимента, выполненного в среде *AnyLogic*, представлен на рис. 2.5. Видно, что приведенные суммарные затраты составляют 0,748 (YE/т), при этом интенсивность нагрузки терминала равна 0,44.

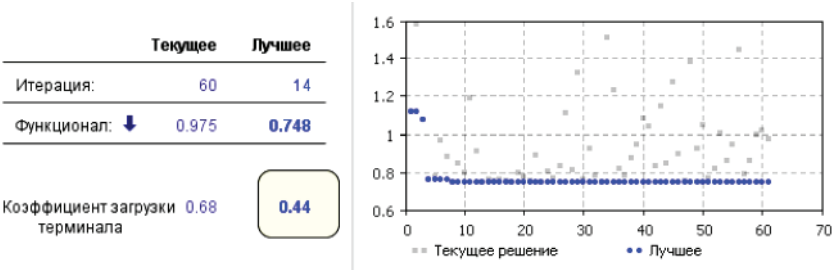


Рис. 2.5. Результаты решения оптимизационной задачи

Следует отметить, что дополнением к процедуре решения оптимизационных задач в среде *AnyLogic* выступает надстройка, позволяющая проводить параметрический анализ решения. Это дает возможность проанализировать влияние того или иного параметра на целевую функцию, выявить качественные и количественные закономерности этого влияния. В качестве примера на рис. 2.6 представлены результаты работы процедуры параметрического эксперимента — выявлена зависимость целевой функции задачи от интенсивности нагрузки терминала.

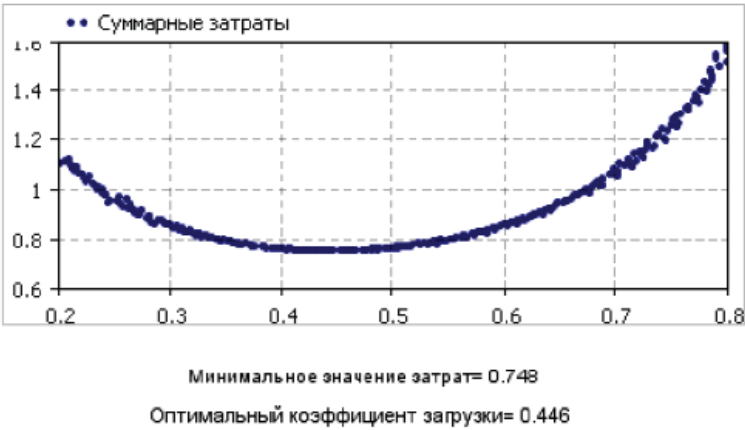


Рис. 2.6. Результат параметрического эксперимента

2.2. Многоканальные СМО с ожиданием

Рассмотрим n -канальную (многоканальную) систему с неограниченной очередью. Поток заявок, поступающих в СМО, имеет интенсивность λ , а поток обслуживаний — интенсивность μ . На рис. 2.7 представлен граф состояний системы.

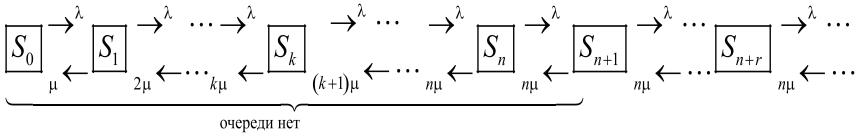


Рис. 2.7. Граф состояний для многоканальной СМО с неограниченной очередью

Система может находиться в одном из состояний:

S_0 — в СМО заявок нет (все каналы свободны);

S_1 — один канал занят, остальные свободны;

S_2 — заняты два канала, остальные свободны;

...

S_k — занято k каналов, остальные свободны;

...

S_n — заняты все n каналов, (очереди нет);

S_{n+1} — заняты все n каналов, одна заявка стоит в очереди;

...

S_{n+r} — заняты все n каналов, r заявок стоит в очереди;

...

Можно показать, что при $\frac{\rho}{n} \geq 1$ очередь будет расти до бесконечности, иначе справедливы формулы (например, [1]), представленные в табл. 3.

Таблица 3

Показатели работы СМО с неограниченной очередью

№ п/п		Наименование	Формула
1	p_0	Вероятность того, что система находится в состоянии S_0	$p_0 = \left(\sum_{i=0}^n \frac{\rho^i}{i!} + \frac{\rho^{n+1}}{n!(n-\rho)} \right)^{-1}$
2	P_q	Вероятность того, что заявка окажется в очереди	$P_q = \frac{\rho^{n+1}}{n!(n-\rho)} p_0$
3	L_q	Среднее число заявок в очереди (длина очереди)	$L_q = \frac{\rho^{n+1} p_0}{n \cdot n! \left(1 - \frac{\rho}{n} \right)^2}$
4	L_s	Среднее число заявок в системе	$L_s = L_q + \rho$
5	T_q	Среднее время пребывания заявки в очереди	$T_q = \frac{L_q}{\lambda} >$
6	T_s	Среднее время пребывания заявки в системе	$T_s = \frac{L_s}{\lambda}$
7	γ	Отношение времени ожидания в очереди к времени обслуживания	$\gamma = \frac{L_q}{L_s - L_q}$
8	Q	Относительная пропускная способность — вероятность того, что заявка будет обслужена	$Q = 1$
9	A	Абсолютная пропускная способность системы (среднее число заявок, обслуживаемых в единицу времени)	$A = \lambda Q = \lambda$
10	$\bar{k}$	Среднее число занятых каналов	$\bar{k} = \frac{\lambda}{\mu} = \rho$

2.3. Пример решения оптимизационной задачи

Произведем оптимизацию многоканальной системы с неограниченной очередью по критерию минимальных совокупных приведенных затрат в зависимости от количества причалов и интенсивности их нагрузки [5].

На рис. 2.8 приведена принципиальная схема моделируемой системы с элементами управления — интенсивностью входного потока, количеством причалов и средним временем обслуживания одного судна.



Рис. 2.8. Принципиальная схема СМО

Заготовка для простейшей 2-D анимационной схемы приведена на рис. 2.9.

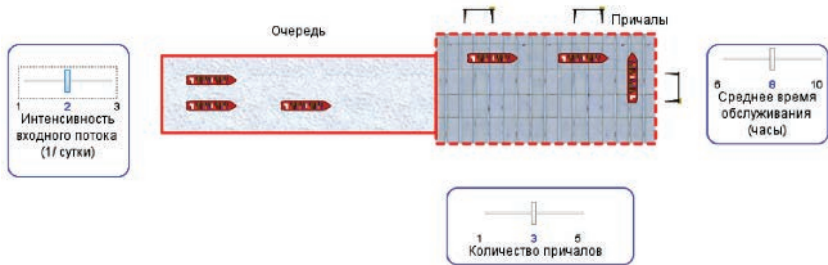


Рис. 2.9. 2-D схема системы

При известном соотношении интенсивности потока судов и интенсивности их обслуживания (т.е. при заданной приведенной интенсивности потока судов, или интенсивности нагрузки терминала) $\rho = \frac{\lambda}{\mu}$ задача сводится к определению оптимального числа задействованных причалов для обеспечения минимума целевой функции — совокупных приведенных затрат.

Формула расчета затрат по терминалу (2) преобразуется к виду:

$$f_T = n \left(k \frac{\rho}{n} Z_T^p(\Pi) + \left(1 - k \frac{\rho}{n} \right) Z_T^{np}(\Pi) \right).$$

Соответственно изменится и вид целевой функции:

$$f = \frac{(1 + \gamma(\rho) + k_T) Z_s + k Z_T^p(\Pi) + (1/\varphi - k) Z_T^{mp}(\Pi)}{\Pi} \rightarrow \min,$$

где $\varphi = \frac{\rho}{n}$ – интенсивность нагрузки причалов.

Пусть $\Pi = 5000$, $\rho = 1,8$, тогда минимальное количество задействованных причалов будет $n = 2$, при этом интенсивность их нагрузки составит $\varphi = \frac{\rho}{n} = 0,9$.

В табл. 4 представлены результаты расчета показателей функционирования системы при различных (в диапазоне от 2 до 5) значениях количества задействованных причалов.

Таблица 4

Основные показатели работы СМО

Характеристики обслуживания	Число причалов n			
	2	3	4	5
Вероятность простоя причалов p_0	0,05	0,15	0,16	0,17
Средняя длина очереди L_q	7,67	0,53	0,11	0,02
Затраты по судам f_s	6972	1814	1506	1446
Затраты по причалам f_T	1289	1839	2389	2939
Относительные затраты f	1,65	0,73	0,78	0,88

Из данных, приведенных в табл. 4 следует, что необходимо задействовать 3 причала, при этом интенсивность нагрузки причалов составит $\varphi = \frac{\rho}{3} = 0,6$.

На рис. 2.10 и 2.11 представлены основные характеристики работы системы, полученные во время моделирования в среде *AnyLogic* и соответствующие предельные значения, рассчитанные по формулам, приведенным в табл. 3.

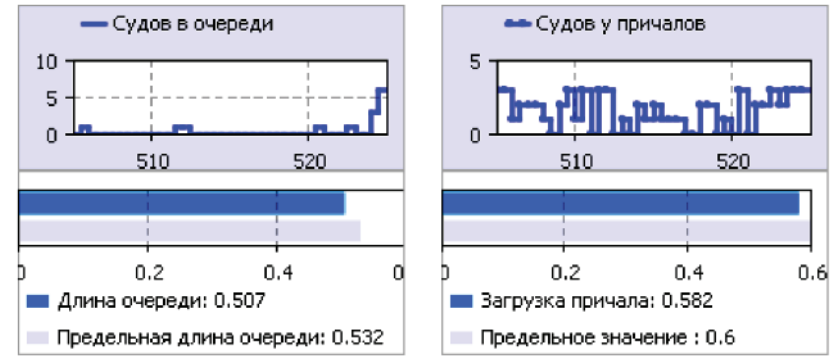


Рис. 2.10. Данные о длине очереди и загрузки причала

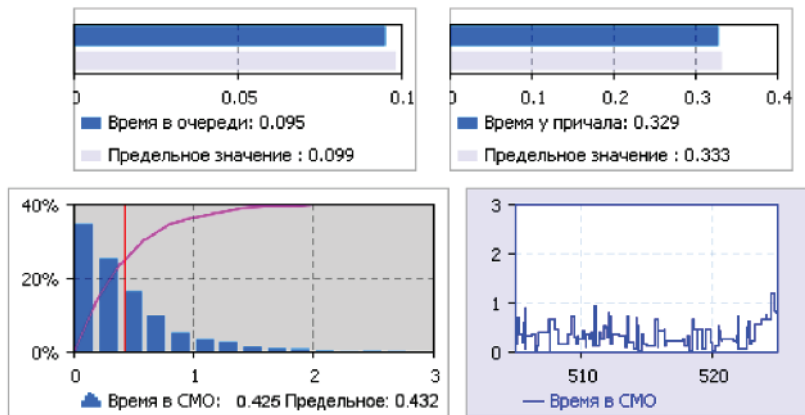


Рис. 2.11. Показатели времени в очереди, обслуживания и в системе

Процедура оптимизации в *AnyLogic* позволяет решать многопараметрические задачи. Так можно определить при каких сочетаниях приведенной интенсивности потока судов и числом задействованных причалов будет обеспечен минимум совокупных приведенных расходов. На рис. 2.12 представлен фрагмент поиска оптимума по двум параметрам — ρ и n .

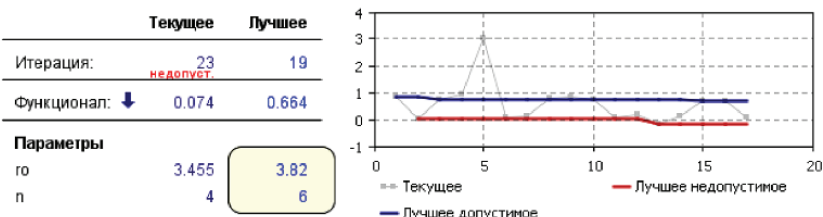


Рис. 2.12. Поиск оптимума по двум параметрам

Аналогично может выполняться оптимизация по любым другим параметрам и их комбинации.

3. СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ СМЕШАННОГО ТИПА

3.1. Системы с ограниченной длиной очереди

Если новая заявка поступает в момент, когда все m мест в очереди заняты, она покидает СМО необслуженной, т.е. получает отказ.

3.1.1. Одноканальные СМО

Пронумеруем состояния СМО по числу заявок, находящихся в системе:

S_0 — канал свободен

S_1 — канал занят, очереди нет

S_2 — канал занят, одна заявка стоит в очереди

... ..

S_k — канал занят, $k - 1$ заявок стоят в очереди

... ..

S_{m+1} — канал занят, m заявок стоят в очереди

На рис. 3.1 представлен граф состояний одноканальной СМО с ограниченной очередью.

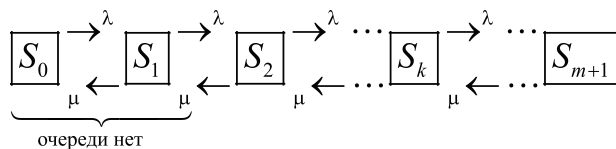


Рис. 3.1. Граф состояний одноканальной СМО при ограниченной длине очереди

Очевидно

$$\left\{ \begin{array}{l} p_1 = \rho p_0; \\ p_2 = \rho^2 p_0; \\ \dots; \\ p_k = \rho^k p_0; \\ \dots; \\ p_{m+1} = \rho^{m+1} p_0; \\ p_0 = (1 + \rho + \rho^2 + \dots + \rho^{m+1})^{-1}. \end{array} \right.$$

Тогда

$$\begin{aligned} p_0 &= \left(\frac{1 \cdot (1 - \rho^{m+2})}{1 - \rho} \right)^{-1} = \frac{1 - \rho}{1 - \rho^{m+2}} \quad (\rho = 1 \Rightarrow p_0 = (m+2)^{-1}); \\ P_{\text{отк}} = p_{m+1} &= \frac{\rho^{m+1} (1 - \rho)}{1 - \rho^{m+2}}; \quad Q = 1 - P_{\text{отк}} = 1 - \frac{\rho^{m+1} (1 - \rho)}{1 - \rho^{m+2}}; \quad A = \lambda Q; \\ L_q &= 1 \cdot p_2 + 2 \cdot p_3 + \dots + (k-1) \cdot p_k + \dots + m \cdot p_{m+1} = \\ &= 1 \cdot \rho^2 p_0 + 2 \cdot \rho^3 p_0 + \dots + (k-1) \cdot \rho^k p_0 + \dots + m \cdot \rho^{m+1} p_0 = \\ &= \rho^2 p_0 [1 + 2 \cdot \rho + \dots + (k-1) \cdot \rho^{k-2} + \dots + m \cdot \rho^{m-1}] = \\ &= \rho^2 p_0 \sum_{k=1}^m k \cdot \rho^{k-1} = \rho^2 p_0 \sum_{k=1}^m \frac{d}{d\rho} \rho^k = \\ &= \rho^2 p_0 \frac{d}{d\rho} \sum_{k=1}^m \rho^k = \rho^2 p_0 \frac{d}{d\rho} \frac{\rho(1 - \rho^m)}{1 - \rho} = \rho^2 p_0 \frac{1 - \rho^m (m+1 - m\rho)}{(1 - \rho)^2}. \end{aligned}$$

Таким образом, длина очереди определится следующим образом:

$$L_q = \frac{\rho^2 [1 - \rho^m (m+1 - m\rho)]}{(1 - \rho^{m+2})(1 - \rho)}.$$

3.1.2. Многоканальные СМО

Рассмотрим n -канальную СМО с ожиданием, на которую поступает поток заявок с интенсивностью λ ; интенсивность обслуживания m (для одного канала); число мест в очереди m .

Состояния системы нумеруются по числу заявок, находящихся в системе:

- S_0 — все каналы свободны
- S_1 — занят один канал, остальные свободны
- ...
- S_k — заняты k -каналов, остальные свободны
- ...
- S_n — заняты все n -каналов, свободных нет
- S_{n+1} — заняты все n -каналов, одна заявка стоит в очереди
- ...
- S_{n+r} — заняты все n -каналов, r -заявок в очереди
- ...
- S_{n+m} — заняты все n -каналов, m -заявок в очереди

Граф состояний многоканальной СМО с ограничением на длину очереди приведен на рис. 3.2 [5].

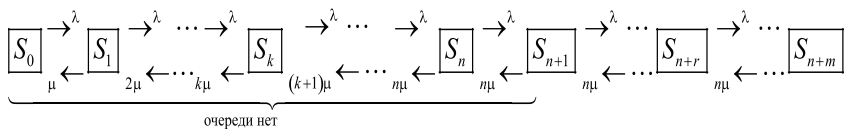


Рис. 3.2. Граф состояний многоканальной СМО при ограниченной длине очереди

У каждой стрелки проставлены соответствующие интенсивности потоков событий. Причем по стрелкам справа налево систему переводит поток обслуживаний интенсивность которого равна m , умноженному на число занятых каналов.

Очевидно в данном случае:

$$\begin{aligned}
 p_0 &= \left(1 + \frac{\rho}{1!} + \frac{\rho^2}{2!} + \dots + \frac{\rho^n}{n!} + \underbrace{\frac{\rho^{n+1}}{n \cdot n!} + \frac{\rho^{n+2}}{n^2 \cdot n!} + \dots + \frac{\rho^{n+m}}{n^m \cdot n!}}_{\text{убывающая геометрическая прогрессия со знаменателем } \rho/n} \right)^{-1} = \\
 &= \left(\sum_{k=0}^n \frac{\rho^k}{k!} + \frac{\rho^{n+1} \left(1 - \left(\frac{\rho}{n} \right)^m \right)}{n \cdot n! \left(1 - \frac{\rho}{n} \right)} \right)^{-1}; \\
 p_1 &= \frac{\rho}{1!} p_0, \dots, p_k = \frac{\rho^k}{k!} p_0, \dots, p_n = \frac{\rho^n}{n!} p_0; \quad p_{n+1} = \frac{\rho^{n+1}}{n \cdot n!} p_0, \dots, p_{n+m} = \frac{\rho^{n+m}}{n^m \cdot n!} p_0; \\
 p_{\text{отк}} &= p_{n+m} = \frac{\rho^{n+m}}{n^m \cdot n!} p_0; \quad Q = 1 - p_{\text{отк}} = 1 - \frac{\rho^{n+m}}{n^m \cdot n!} p_0; \\
 A &= \lambda Q = \lambda \left(1 - \frac{\rho^{n+m}}{n^m \cdot n!} p_0 \right); \quad \bar{k} = \frac{A}{\mu} = \frac{\lambda \left(1 - \frac{\rho^{n+m}}{n^m \cdot n!} p_0 \right)}{\mu} = \rho \left(1 - \frac{\rho^{n+m}}{n^m \cdot n!} p_0 \right).
 \end{aligned}$$

Среднее число заявок в очереди:

$$\begin{aligned}
 L_q &= 1 \cdot p_{n+1} + 2 \cdot p_{n+2} + \dots + m \cdot p_{n+m} = p_0 \left[1 \cdot \frac{\rho^{n+1}}{n \cdot n!} + 2 \cdot \frac{\rho^{n+2}}{n^2 \cdot n!} + \dots + m \cdot \frac{\rho^{n+m}}{n^m \cdot n!} \right] = \\
 &= \frac{\rho^{n+1}}{n \cdot n!} p_0 \left[1 + 2\varphi + 3\varphi^2 + \dots + m\varphi^{m-1} \right] = \frac{\rho^{n+1}}{n \cdot n!} p_0 \sum_{r=1}^m r \cdot \varphi^{r-1} = \\
 &= \frac{\rho^{n+1}}{n \cdot n!} p_0 \sum_{r=1}^m \frac{d}{d\varphi} \varphi^r = \frac{\rho^{n+1}}{n \cdot n!} p_0 \frac{d}{d\varphi} \sum_{r=1}^m \varphi^r = \\
 &= \frac{\rho^{n+1}}{n \cdot n!} p_0 \frac{d}{d\varphi} \frac{\varphi(1-\varphi^m)}{1-\varphi} = \frac{\rho^{n+1}}{n \cdot n!} p_0 \frac{1-\varphi^m(m+1-m\varphi)}{(1-\varphi)^2} = \\
 &= \frac{\rho^{n+1} p_0 \left[1 - \left(m+1-m\frac{\rho}{n} \right) \left(\frac{\rho}{n} \right)^m \right]}{n \cdot n! \left(1 - \frac{\rho}{n} \right)^2}.
 \end{aligned}$$

Среднее число заявок в системе:

$$L_s = L_q + \bar{k}.$$

В табл. 5 приведены основные показатели работы СМО с ограниченной по длине очередью.

Таблица 5

Показатели работы многоканальной СМО с ограниченной очередью

№ п/п		Наименование	Формула
1	p_0	Вероятность того, что система находится в состоянии S_0	$p_0 = \left(\sum_{i=0}^n \frac{\rho^i}{i!} + \frac{\rho^{n+1} \left(1 - \left(\frac{\rho}{n} \right)^m \right)}{n \cdot n! \left(1 - \frac{\rho}{n} \right)} \right)^{-1}$
2	L_q	Среднее число заявок в очереди	$L_q = \frac{\rho^{n+1} p_0 \left(1 - \left(m + 1 - m \frac{\rho}{n} \right) \left(\frac{\rho}{n} \right)^m \right)}{n \cdot n! \left(1 - \frac{\rho}{n} \right)^2}$
3	$\bar{k}$	Среднее число заявок под обслуживанием (среднее число занятых каналов)	$\bar{k} = \rho \left(1 - \frac{\rho^{n+m}}{n^m \cdot n!} p_0 \right)$
4	L_s	Среднее число заявок в системе	$L_s = L_q + \bar{k}$
5	Q	Относительная пропускная способность	$Q = 1 - \frac{\rho^{n+m}}{n^m n!} p_0$
6	A	Абсолютная пропускная способность системы	$A = \lambda Q$

3.1.3. Примеры практической реализации

Пример 1.

В порту имеется один специализированный причал. Сравнить показатели функционирования причала в случае неограниченной

очереди и при наличии ограничения на ее длину (не может превосходить m судов). Используются формулы, представленные в табл. 5.

Результаты сравнения основных показателей функционирования СМО для различных вариантов ограничений на длину очереди представлены в табл. 6 (расчеты выполнены при $\rho = 0,8$).

Таблица 6

Влияние ограничения на длину очереди

m	p_0	Q	A	L_q	T_q	L_s	T_s
1	0,41	0,738	0,295	0,262	0,656	0,852	2,131
3	0,297	0,878	0,351	0,861	2,151	1,563	3,908
5	0,253	0,934	0,373	1,396	3,489	2,142	5,356
10	0,215	0,982	0,393	2,329	5,823	3,115	7,786
∞	0,2	1	0,4	3,2	8	4	10

Пример 2

В порту имеется один специализированный причал. Интенсивность входящего потока судов равна 0,4 (судов в сутки). Среднее время обработки одного судна составляет 2 суток. Найти показатели эффективности работы причала при неограниченной длине очереди и при условии, что приходящее судно покидает причал (без разгрузки), если очередь полностью заполнена — в ней находится 3 судна.

Решение

А. Длина очереди не ограничена ($m = \infty$).

Имеем

$$\rho = \frac{\lambda}{\mu} = \lambda \bar{t}_{\text{об}} = 0,4 \cdot 2 = 0,8.$$

Вероятность того, что причал свободен:

$$p_0 = 1 - \rho = 1 - 0,8 = 0,2.$$

Вероятность того, что причал занят:

$$P_{\text{зан}} = 1 - p_0 = 1 - 0,2 = 0,8.$$

Среднее число судов, ожидающих разгрузки:

$$L_q = \frac{\rho^2}{1-\rho} = \frac{0,8^2}{1-0,8} = 3,2.$$

Среднее время ожидания разгрузки:

$$T_q = \frac{L_q}{\lambda} = \frac{3,2}{0,4} = 8 \text{ (сутки)}.$$

Среднее число судов, находящихся у причала:

$$L_s = \frac{\rho}{1-\rho} = \frac{0,8}{1-0,8} = 4 \text{ (сутки)}.$$

Среднее время пребывания судна у причала по формуле:

$$T_s = \frac{L_s}{\lambda} = \frac{4}{0,4} = 10 \text{ (сутки)}.$$

Б. Длина очереди ограничена ($m = 3$)

Вероятность того, что причал свободен:

$$p_0 = \frac{1-\rho}{1-\rho^{m+2}} = \frac{1-0,8}{1-0,8^{3+2}} = 0,297.$$

Вероятность того, что приходящее судно покинет причал без разгрузки:

$$P_{\text{отк}} = p_{m+1} = \rho^{m+1} p_0 = 0,8^{3+1} \cdot 0,297 = 0,122.$$

Относительная пропускная способность причала:

$$Q = 1 - P_{\text{отк}} = 1 - 0,122 = 0,878.$$

Абсолютная пропускная способность причала

$$A = \lambda Q = 0,4 \cdot 0,878 = 0,351.$$

Среднее число судов, ожидающих разгрузку:

$$L_q = \frac{\rho^2 [1 - \rho^m (m + 1 - m\rho)]}{(1 - \rho^{m+2})(1 - \rho)} = \frac{0,8^2 [1 - 0,8^3 (3 + 1 - 3 \cdot 0,8)]}{(1 - 0,8^{3+2})(1 - 0,8)} = 0,861.$$

Среднее время ожидания разгрузки:

$$T_q = \frac{L_q}{\lambda} = \frac{0,861}{0,4} = 2,153 \text{ (сутки)}.$$

Среднее число судов, находящихся у причала:

$$L_s = 0,861 + (1 - 0,297) = 1,564.$$

Среднее время пребывания судна у причала по:

$$T_s = \frac{L_s}{\lambda} = \frac{1,564}{0,4} = 3,91 \text{ (сутки)}.$$

В табл. 7 приведены результаты расчетов показателей функционирования СМО с неограниченной длиной очереди и с длиной очереди не более 3 судов.

Таблица 7

Сравнение расчетов

Показатель	$m = \infty$	$m = 3$
p_0	0,2	0,297
$P_{\text{отк}}$	0	0,122
Q	1	0,878
A	0,4	0,351
L_q	3,2	0,861
T_q	8	2,153
L_s	4	1,564
T_s	10	3,91

Очевидно ограничение по длине очереди приводит к увеличению вероятности того, что причал свободен(простаивает) и отказа в обслуживании, но позволяет существенно снизить время пребывания судов в системе.

3.2. Системы с ограниченным периодом ожидания

3.2.1. Необходимые сведения и аналитические выражения

Основными характеристиками многоканальных систем массового обслуживания (СМО) с ожиданием [1] является вероятность p_0 простоя системы и средняя длина L_q очереди:

$$p_0 = \left(\sum_{k=0}^n \frac{\rho^k}{k!} + \frac{\rho^{n+1}}{n!(n-\rho)} \right)^{-1}; \quad (3)$$

$$L_q = \frac{\rho^{n+1} p_0}{n \cdot n! \left(1 - \frac{\rho}{n}\right)^2},$$

где n — число каналов обслуживания;

$\rho = \frac{\lambda}{\mu}$ — коэффициент загрузки системы — отношение интенсивности входящего потока заявок к интенсивности их обслуживания в СМО.

Для формализации описания функционирования систем с ограниченным временем ожидания их удобно представлять в виде графа состояний, который в простейшем варианте является схемой гибели и размножения (см. рис. 3.3).

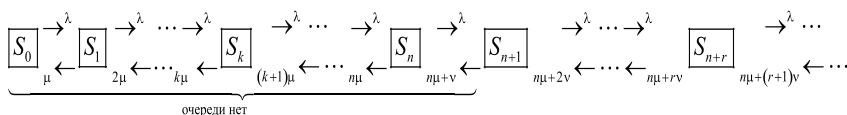


Рис. 3.3. Граф состояний СМО с ограниченным временем ожидания

В этом случае предельные вероятности определяются по следующим формулам:

$$p_0 = \left(\sum_{k=0}^n \frac{\rho^k}{k!} + \frac{\rho^n}{n!} \sum_{r=1}^{\infty} \frac{\rho^r}{\prod_{i=1}^r (n+i\beta)} \right)^{-1}; \quad (4)$$

$$p_k = \frac{\rho^k}{k!} p_0 \left(k = \overline{1, n} \right); \quad p_{n+r} = \frac{\rho^n}{n!} \sum_{r=1}^{\infty} \frac{\rho^r}{\prod_{i=1}^r (n+i \cdot \beta)} p_0,$$

где $\beta = \frac{\nu}{\mu}$;

ν – интенсивность уходящего (из очереди, не дождавшись обслуживания) потока заявок;

r – число заявок, находящихся в очереди;

k – число занятых каналов.

Исследуем формулу (4) вероятности того, что система находится в состоянии S_0 (все каналы свободны) [6]. Здесь в отличие от формулы (3) второе слагаемое в скобках есть бесконечный ряд, который не является прогрессией, но его элементы быстро убывают с ростом их номера.

Представим бесконечную сумму в виде двух слагаемых, в первом учитывается конечное число $(q - 1)$ ее элементов, а второе (бесконечная сумма) – остаток.

$$\sum_{r=1}^{\infty} \frac{\rho^r}{\prod_{i=1}^r (n+i \cdot \beta)} = \sum_{r=1}^{q-1} \frac{\rho^r}{\prod_{i=1}^r (n+i \cdot \beta)} + \sum_{r=q}^{\infty} \frac{\rho^r}{\prod_{i=1}^r (n+i \cdot \beta)}. \quad (5)$$

Оценим остаток R , очевидно

$$R = \sum_{r=q}^{\infty} \frac{\rho^r}{\prod_{i=1}^r (n+i \cdot \beta)} < \sum_{r=q}^{\infty} \frac{\rho^r}{\prod_{i=1}^r i \cdot \beta} = \sum_{r=q}^{\infty} \frac{(\rho/\beta)^r}{r!}.$$

Можно показать, что

$$R < \frac{(\rho/\beta)^q}{q!} e^{\rho/\beta}. \quad (6)$$

Действительно

$$\frac{(\rho/\beta)^q}{q!} \left(1 + \frac{\rho/\beta}{q+1} + \frac{(\rho/\beta)^2}{(q+1)(q+2)} + \dots \right) < \frac{(\rho/\beta)^q}{q!} \left(1 + \frac{\rho/\beta}{1!} + \frac{(\rho/\beta)^2}{2!} + \dots \right)$$

и, соответственно

$$\frac{\rho^n}{n!} \sum_{r=q}^{\infty} \frac{(\rho/\beta)^r}{r!} < \frac{\rho^n}{n!} \frac{(\rho/\beta)^q}{q!} e^{\rho/\beta}.$$

Длину очереди можно найти по формуле:

$$L_q = \frac{\rho - \bar{k}}{\beta},$$

где $\bar{k} = \sum_{k=1}^{n-1} k p_k + n \left(1 - \sum_{i=0}^{n-1} p_i \right)$ – среднее число занятых каналов

3.2.2. Имитационное моделирование

Применим методику исследования СМО с ограничением на время ожидания заявок в очереди для решения практической задачи. Для моделирования используем специализированную аналитическую платформу по имитационному моделированию [6, 7].

Морской грузовой терминал состоит из $n = 3$ специализированных причалов для грузообработки (разгрузки/погрузки/обработки) судов. Интенсивность входящего потока судов $\lambda = 4$ (судов в сутки). Интенсивность разгрузки судов на каждом причале $\mu = 2$ (судна в сутки).

Проведем сравнение показателей функционирования терминала для трех вариантов работы:

- ограничений на очередь нет – СМО функционирует как многоканальная система с ожиданием;
- суда покидают очередь (уходят по таймауту) если длительность ожидания превышает некоторую величину;
- известна интенсивность уходящего из очереди потока судов.

1. Если ограничений на очередь нет, то имеем многоканальную систему с ожиданием. Принципиальная схема такой системы в среде *AnyLogic* представлена на рис. 3.4.

По представленной схеме к терминалу всего подошло 23 судна, два из них ожидают в очереди, три находятся у причалов на обработке и 18 покинули терминал после обработки.

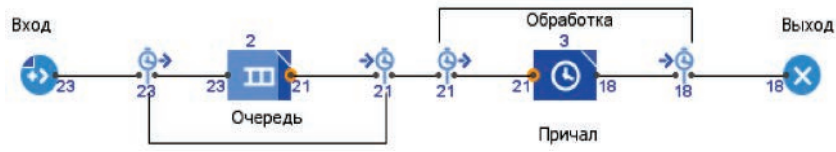


Рис. 3.4. Схема СМО с ожиданием

Для рассматриваемой системы

$$\rho = \frac{\lambda}{\mu} = 2.$$

Определяем вероятность простоя причалов по формуле (3):

$$p_0 = \left(1 + \frac{2}{1!} + \frac{2^2}{2!} + \frac{2^3}{3!} + \frac{2^4}{3!(3-2)} \right)^{-1} = \frac{1}{9}.$$

Среднее число судов в очереди:

$$L_q = \frac{2^4 \cdot \frac{1}{9}}{3 \cdot 6 \left(1 - \frac{2}{3} \right)^2} = \frac{8}{9}.$$

На рис. 3.5 представлены диаграммы, характеризующие длину очереди по результату имитационного моделирования и рассчитанную по аналитическому выражению.

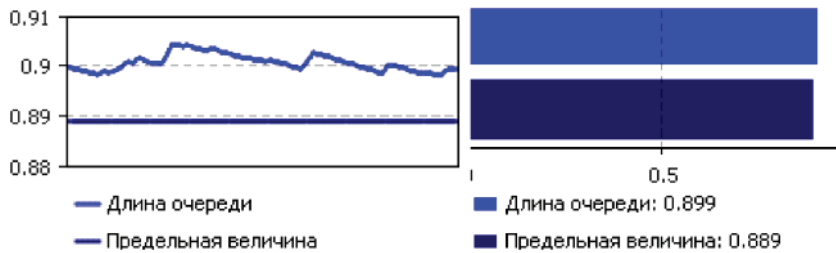


Рис. 3.5. Модельное и предельное значение длины очереди

Среднее время ожидания в очереди на обслуживание:

$$T_q = \frac{L_q}{\lambda} = \frac{8/9}{49} = \frac{2}{49} \approx 0,22.$$

Данные о текущем количестве судов в очереди, времени пребывания в очереди и его предельному значению представлены на рис. 3.6.



Рис. 3.6. Данные разmere очереди и времени пребывания в ней

Рис. 3.7 характеризует соотношение времени, проведенного в очереди к эффективному времени грузообработки (разгрузки).

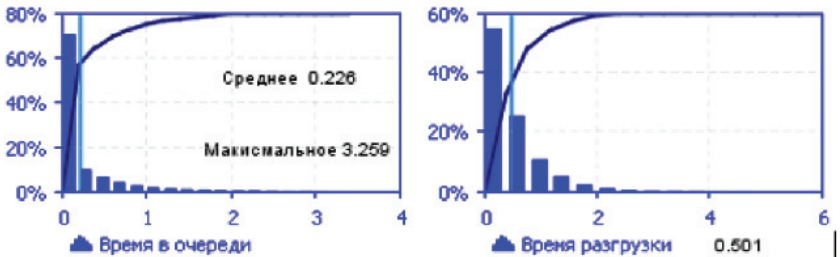


Рис. 3.7. Соотношение времени в очереди и обработки

Отметим, что в данном режиме максимальное время пребывания в очереди превышает 3 суток.

Среднее время пребывания судна в системе:

$$T_s = T_q + \frac{1}{\mu} \approx 0,22 + 0,5 = 0,72.$$

2. Известно, что суда покидают очередь, если время нахождения в ней превышает 1 сутки.

Принципиальная схема такой системы с уходом из очереди по таймауту представлена на рис. 3.8.



Рис. 3.8. Схема СМО с «нетерпеливыми» заявками

По текущему состоянию системы можно констатировать, что из 183 пришедших судов, 172 покинули систему после грузообработки, 3 находятся на обработке, 6 – в очереди и 2 покинули терминал, не дождавшись обработки из-за превышения допустимого времени пребывания в очереди.

На рис. 3.9 представлены результаты моделирования по времени, проведенном судами в очереди и длине очереди. Понятно, что максимальное время в очереди ограничено 1 сутками (ранее 3,259), а среднее время уменьшилось до 0,139 (0,226) за счет того, что часть судов уходит по таймаут. Очевидно средняя длина очереди также стала меньше, чем была в режиме ожидания без ограничений.

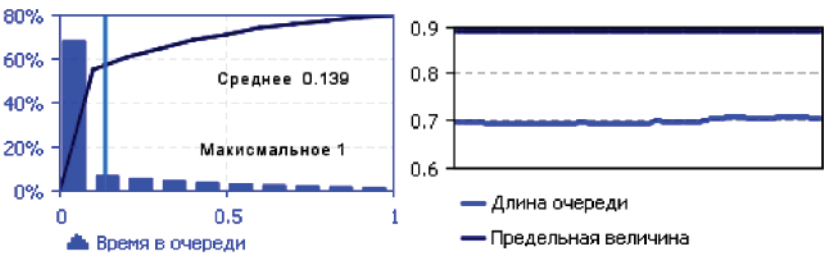


Рис. 3.9. Время в очереди и ее длина с учетом таймаута

3. На основании статистических данных известна интенсивность досрочного ухода судов из очереди $\nu = 0,15$. Исследуем как изменяются основные показатели функционирования СМО при различных значениях q количества элементов, учитываемых в разложении (5) для вероятности простоя системы.

Пусть $q = 6$. Зависимости $p_0 = f(r)$ и $R = f(r)$ представлены на рис. 3.10.

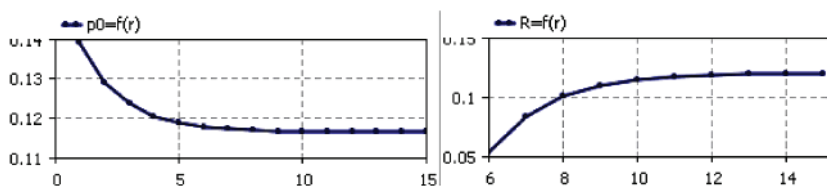


Рис. 3.10. Влияние числа заявок в очереди на показатели СМО

Очевидно, в данном случае:

$$\beta = \frac{\nu}{\mu} = 0,075; \quad p_0 = 0,119.$$

Тогда среднее число занятых каналов:

$$\bar{k} = \sum_{k=1}^{n-1} k p_k + n \left(1 - \sum_{i=0}^{n-1} p_i \right) = 1,931,$$

длина очереди:

$$L_q = \frac{\rho - \bar{k}}{\beta} = 0,915.$$

На рис. 3.11 представлена информация о модельном значении длины очереди и ее предельном значении (0,915). Отличия в значениях объясняются погрешностью, которая вносится за счет учета только 5 элементов в разложении по формуле (5).

Рис. 3.12 содержит информацию о модельном значении времени, проводимом судами в очереди и его предельном значении

$$T_q = \frac{L_q}{\lambda} = \frac{0,915}{4} \approx 0,229.$$

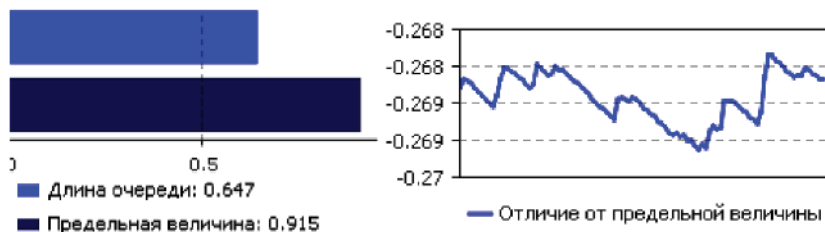


Рис. 3.11. Длина очереди и ее предельное значение



Рис. 3.12. Сравнение времени в очереди и его предельного значения

Очевидно, при $q = 6$ остаток $R \rightarrow 0,12$ (рис. 3.10). Таким образом, оценка (6) завышена.

На основании данных, представленных на графиках рис. 3.10 можно сделать вывод о том, что при $r \geq 15$ показатели функционирования СМО остаются практически неизменными.

Тогда

$$p_0 = 0,117; \quad \bar{k} = \sum_{k=1}^{n-1} k p_k + n \left(1 - \sum_{i=0}^{n-1} p_i \right) = 1,951;$$

$$L_q = \frac{\rho - \bar{k}}{\beta} = 0,65.$$

В этом случае показатели (расчетные и предельные) длины очереди становятся практически одинаковыми (см. рис. 3.13).

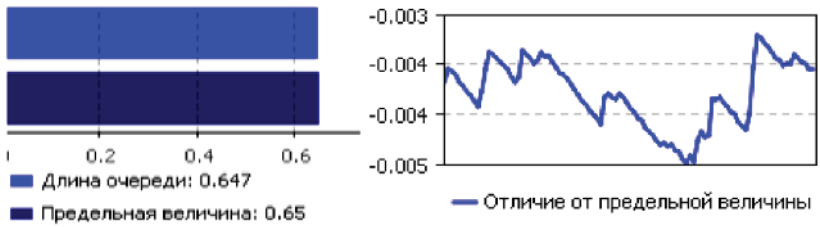


Рис. 3.13. Информация о длине очереди



Рис. 3.14. Значения времени ожидания в очереди

На рис. 3.14 приведена информация о среднем времени, проведенном судами в очереди. Расчетное 0,139 и предельное:

$$T_q = \frac{L_q}{\lambda} = \frac{0,65}{4} \approx 0,163.$$

4. СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ С ПРИОРИТЕТНЫМИ ЗАЯВКАМИ

4.1. Материал и методика

Для одноканальных систем массового обслуживания с ожиданием с простейшим потоком заявок и временем обслуживания, распределенным по показательному закону, средняя продолжительность ожидания обслуживания определяется по известному выражению:

$$T_q = \frac{\rho}{\mu - \lambda}, \quad (7)$$

где ρ — приведенная интенсивность потока заявок (интенсивность нагрузки канала);

λ, μ — интенсивности потока заявок и обслуживания, соответственно.

На практике часто используются СМО с приоритетным обслуживанием заявок.

Пусть в систему поступают заявки двух рангов (заявки первого ранга обслуживаются в первую очередь). В таких случаях (поток заявок уже не является простейшим) среднюю длительность ожидания обслуживания для заявок первого t_1 и второго t_2 рангов можно определить по следующим эмпирическим выражениям:

$$t_1 = k_1 T_q; \quad t_2 = k_2 T_q,$$

где $k_1 = k_1(a, \rho); k_2 = k_2(a, \rho)$.

Здесь $a = \frac{\lambda_1}{\lambda}$ (λ_1 — интенсивность потока заявок первого ранга);

$$k_1 = \frac{1-\rho}{1-a\rho}; \quad k_2 = \frac{1}{1-a\rho}.$$

Окончательно:

$$t_1 = \frac{1-\rho}{1-a\rho} T_q; \quad t_2 = \frac{1}{1-a\rho} T_q. \quad (8)$$

На рис. 4.1 приведены графики зависимости коэффициентов k_1 и k_2 от величины a при фиксированном значении нагрузки канала.

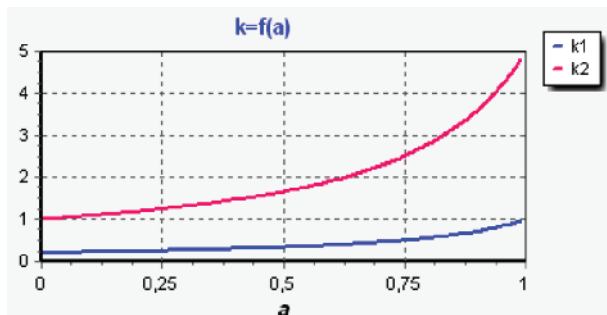


Рис. 4.1. Графики эмпирических коэффициентов

Одной из возможностей организации имитационного моделирования СМО с учетом наличия приоритетных заявок двух рангов является разделение (блок «Выбор» на рис. 4.2) входящего потока на две очереди в соответствии с процентным соотношением приоритетных и простых заявок. В одну из очередей направляются заявки с приоритетом.

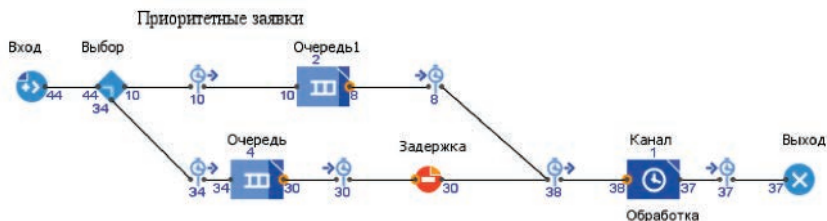


Рис. 4.2. Схема СМО с приоритетами в AnyLogic

В этом случае при нахождении в очереди приоритетных заявок альтернативная очередь блокируется до тех пор, пока обрабатываются эти заявки (блок «Задержка» на рис. 4.2).

Таким же образом можно организовать обслуживание многоканальных (три и более) приоритетных заявок.

Более унифицированным методом является подход, основанный на явном задании для каждой заявки ее приоритета. Тогда из очереди на обслуживание в первую очередь забираются все пришедшие к текущему моменту времени заявки с наиболее высоким приоритетом и т.д.

На рис. 4.3 представлен фрагмент имитационной модели. Поток с заявками наивысшего приоритета обозначен как «Красный».



Рис. 4.3. Модель трехканального потока с приоритетами

Особую практическую значимость имеет СМО в которых время обслуживания заявок в канале различно. В этом случае необходимо решать оптимизационную задачу по упорядочению потока заявок на обслуживание для минимизации среднего (суммарного) времени ожидания в очереди и оперативно формировать систему приоритетов.

В более общем случае для минимизации стоимостных затрат за время ожидания порядок обслуживания должен учитывать не только длительность обслуживания, но и «стоимость простоя» (C) заявки в единицу времени. Можно показать, что, в данном случае приоритеты в обслуживании заявок должны быть упорядочены по условию:

$$(\mu_1 C_1) > (\mu_2 C_2) > (\mu_3 C_3) > \dots$$

4.2. Практическая реализация

Решим задачу определения времени ожидания обслуживания при простейших потоках заявок. Пусть в одноканальную систему поступает простейший поток заявок λ , обслуживание распределено по показательному закону со средней длительностью $t_{\text{обсл}} = 0,5$ и коэффициент загрузки системы $\rho = 0,8$, тогда в соответствии с формулой (7) средняя длительность ожидания обслуживания составит:

$$T_q = \frac{\rho}{\mu - \lambda} = \frac{0,8}{2 - 1,6} = 2.$$

На рис. 4.4 приведены информация о модельных и предельных значениях времени ожидания в очереди.

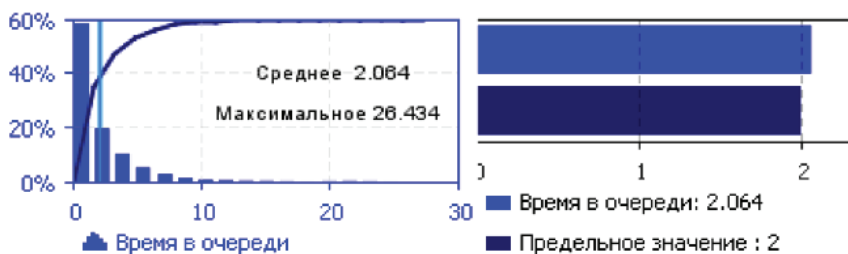


Рис. 4.4. Результаты имитации простейшего потока заявок

Исследуем поток с приоритетами [8]. Пусть в соответствии с формулой (8) 20 % ($a = 0,2$) заявок имеют право на приоритетное обслуживание, тогда:

$$t_1 = \frac{1-\rho}{1-a\rho} T_q = \frac{1-0,8}{1-0,2 \cdot 0,8} \cdot 2 = 0,4762; \quad t_2 = \frac{t_1}{1-\rho} = 2,381.$$

На рис. 4.5 представлена гистограмма распределения времени ожидания в очереди для потока с высшим приоритетом и сравнение экспериментального времени с эмпирической оценкой (8).

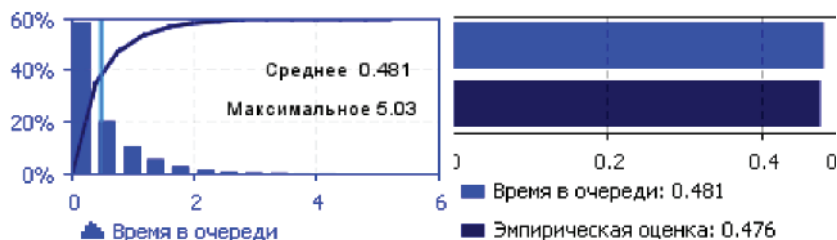


Рис. 4.5. Характеристики времени ожидания для заявок с приоритетом

На рис. 4.6 представлены аналогичные показатели для потока заявок без приоритета.

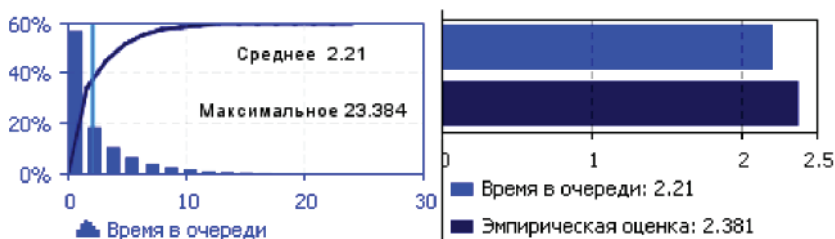


Рис. 4.6. Характеристики времени ожидания для заявок без приоритета

Результаты моделирования свидетельствуют об удовлетворительности эмпирических оценок (8).

Как уже отмечалось универсальным способом организации обработки заявок является явное задание приоритета при формировании заявки. Проведем моделирование СМО в соответствии со схемой, представленной на рис. 4.3.

Пусть число каналов $n = 4$, наивысший приоритет имеют «красные» заявки.

Интервал времени между прибытиями задан треугольными распределениями (соответственно для «красных», «желтых» и «зеленых заявок»):

$$A_1 = (2,5; 2,75; 3); \quad A_2 = (1; 1,25; 1,5); \quad A_3 = (0,8; 0,9; 1).$$

На рис. 4.7 представлен фрагмент имитационного моделирования функционирования СМО. В данный момент времени в очереди две «зеленые» заявки, а обслуживаются в каналах одна «красная», одна «желтая» и две «зеленых заявки» (см. рис. 4.3).

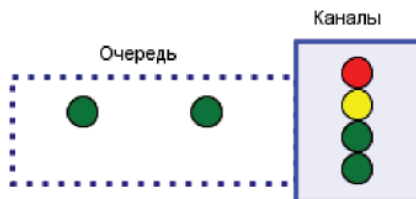


Рис. 4.7. Текущее состояние системы

Очевидно при освобождении одного из каналов в первую очередь на обслуживание будут забираются «красные» заявки, если их в очереди не окажется, то «желтые» и, наконец, «зеленые», обладающие самым низким приоритетом.

Рис. 4.8 содержит информацию об изменении среднего времени ожидания заявок в очереди.

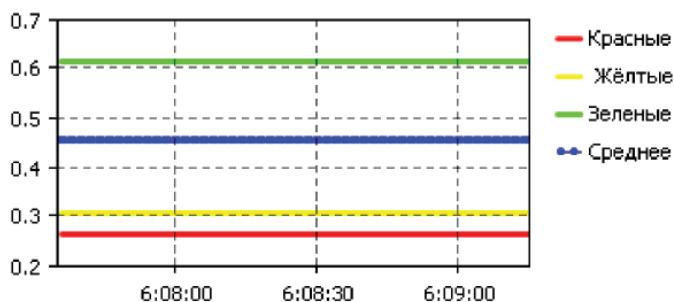


Рис. 4.8. Временной график изменения среднего времени ожидания

В соответствии с представленным графиком среднее время ожидания в очереди для «красных» заявок составляет 0,263, для «желтых» — 0,305, а для «зеленых» — 0,614.

В данном случае продолжительность обслуживания была постоянной $B = (1,25; 1,8; 2,2)$ для любых заявок вне зависимости от их ранга.

Исследуем влияние установки приоритетов на среднюю продолжительность пребывания заявок в очереди при разных временах обслуживания заявок.

Пусть интервал времени между прибытиями одинаков для всех трех групп заявок: $A = (1; 1,25; 1,5)$.

Продолжительность обслуживания определяется различными треугольными законами распределения:

$$B_1 = (1; 1,1; 1,2); \quad B_2 = (1,25; 1,625; 2); \quad B_3 = (2; 2,25; 2,5),$$

т.е. быстрее обслуживаются заявки из первой группы.

Если положить, что наивысшим приоритетом обладают заявки первой группы («быстрые заявки»), то соотношение времен пребывания (по уменьшению приоритета) в очереди будет таким:

$$\frac{(t_1 \parallel t_2 \parallel t_3)}{\bar{t}} = \frac{(0,223 \parallel 0,455 \parallel 1,124)}{0,601},$$

здесь $\bar{t}$ — среднее время пребывания заявок в очереди.

Если ранги групп заявок сменить на противоположные (первыми будут обслуживаться заявки с наибольшим временем обслуживания), то получим следующие результаты:

$$\frac{(t_1 \parallel t_2 \parallel t_3)}{\bar{t}} = \frac{(0,257 \parallel 0,462 \parallel 1,573)}{0,764},$$

т.е. среднее время пребывания в очереди увеличилось на 27 %.

Соотношение времен пребывания заявок в СМО с изначальным приоритетом:

$$\frac{(1,323 \parallel 2,077 \parallel 3,374)}{2,258},$$

и после изменения приоритетов:

$$\frac{(2,507 \parallel 2,086 \parallel 2,673)}{2,422}.$$

5. МОДЕЛИ С НЕ ПУАССОНОВСКИМИ ПОТОКАМИ

5.1. Основные сведения и формальные зависимости

В простейших потоках интервал времени между двумя соседними событиями распределен по показательному закону [1] (см. рис. 5.1, а):

$$f(t) = \mu e^{-\mu t} \quad (t > 0),$$

где μ — параметр распределения.

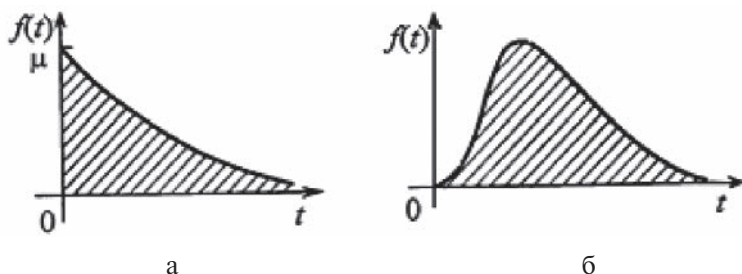


Рис. 5.1. Распределения интервалов времени

На практике же потоки событий в СМО часто существенно отличаются от простейших, а для потока обслуживания такое отличие — норма. Более типичным является случай, когда закон распределения времени обслуживания $f(t)$ отличен от показательного, и его наивероятнейшее значение не равно нулю (см. рис. 5.1, б). В практически значимых задачах аналитические формулы для исследования характеристик СМО в рамках классического математического аппарата удается получить только для самых простых случаев [1].

Очевидно, простейшие потоки являются «наименее регулярными» так как для них коэффициент вариации интервалов между событиями равен единице. Однако на практике наиболее часто

встречаются регулируемые потоки, т.е. потоки с последствием для которых коэффициент вариации заключен между нулем (регулярный поток) и единицей. Такие потоки могут быть описаны с помощью треугольных и трапециевидных чисел, обладающих очевидной простой использования и «понятной» семантикой [9].

Исследование вариантов описания потоков обслуживания и поступления заявок проведем для следующих режимов:

1. Оба потока простейшие.

В этом случае длина очереди определяется по известному выражению:

$$L_q = \frac{\rho^2}{1-\rho},$$

где $\rho = \frac{\lambda}{\mu}$ — приведенная интенсивность потока заявок;

λ, μ — соответственно интенсивности входящего потока и потока обслуживания.

2. Время обслуживания задается нечетким числом, тогда:

$$L_q = \frac{\rho^2 (1 + v_\mu^2)}{2(1-\rho)},$$

где $v_\mu = \frac{\sigma_\tau^\mu}{M_\tau^\mu}$ — коэффициент вариации времени обслуживания.

3. Время обслуживания — константа.

$$L_q = \frac{\rho^2 (1 + v_\mu^2)}{2(1-\rho)} = \frac{\rho^2}{2(1-\rho)}.$$

4. Оба потока (и поток поступления заявок и обслуживания) описываются нечеткими числами.

В этом случае точных аналитических выражений не получено, доказано, что:

$$\frac{\rho^2 (v_\lambda^2 + v_\mu^2)}{2(1-\rho)} - \frac{\rho(1-v_\lambda^2)}{2} \leq L_q \leq \frac{\rho^2 (v_\lambda^2 + v_\mu^2)}{2(1-\rho)} + \frac{(1-\rho)(1-v_\lambda^2)}{2},$$

где $v_\lambda = \frac{\sigma_\tau^\lambda}{M_\tau^\lambda}$ — коэффициент вариации интервалов между заявками.

5.2. Результаты имитационного моделирования

Проведем имитационное моделирование и исследование вариантов описания потоков обслуживания и поступления заявок в соответствии с ранее сформулированными режимами. Пусть для примера интенсивности входящего потока заявок и потока обслуживаний имеют следующие значения: $\lambda = 2$, $\mu = 3$ (в день).

1. Оба потока простейшие.

В данном случае предельная длина очереди и соответственно время ожидания в очереди составят:

$$L_q = \frac{\rho^2}{1-\rho} = \frac{4}{3}; \quad T_q = \frac{L_q}{\lambda} = \frac{2}{3}.$$

На рис. 5.2 представлен фрагмент моделирования потоков с информацией о длине очереди.

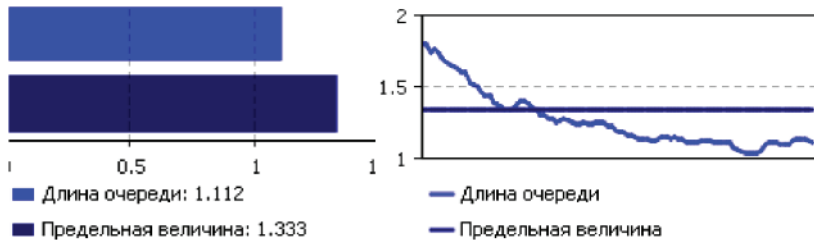


Рис. 5.2. Изменение длины очереди

Очевидно расчетная длина очереди в процессе имитации может отличаться от предельного значения. На рисунке представлен фрагмент времени, в течение которого расчётная длина являлась и больше и меньше своего предельного значения.

На рис. 5.3 приведена информация о среднем времени ожидания заявки в очереди и времени ее обслуживания.

2. Время обслуживания задается нечетким числом $\tilde{A} = (4, 5, 15)$ (часы).

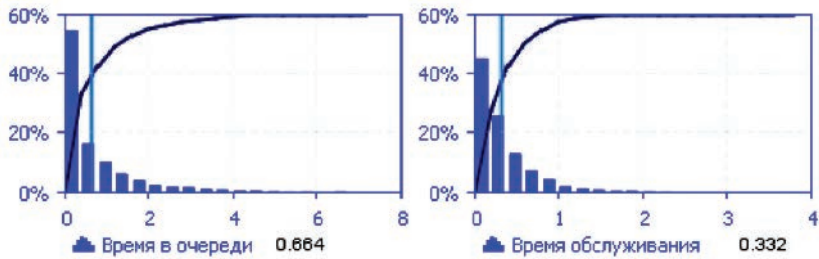


Рис. 5.3. Среднее время ожидания в очереди и обслуживания

Очевидно:

$$L_q = \frac{\rho^2 (1 + v_\mu^2)}{2(1 - \rho)} \approx 0,731.$$

На рис. 5.4 представлен фрагмент имитационного моделирования с информацией о длине очереди. Длина очереди в данном случае меньше, чем в предыдущем случае, т.к. поток обслуживания ближе к регулярному.

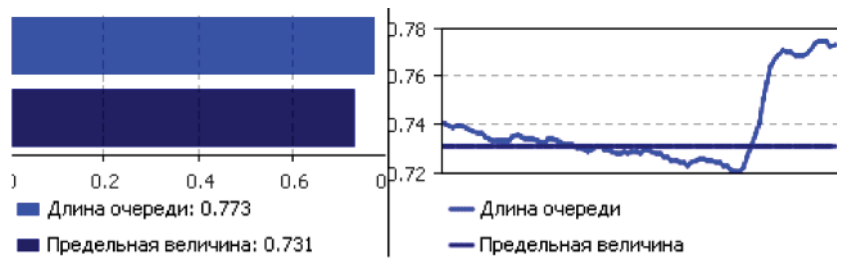


Рис. 5.4. Графики изменения длины очереди

На рис. 5.5 приведены распределения для времени ожидания в очереди и времени обслуживания. Гистограмма для интервалов обслуживания качественно отличается от аналогичной, приведенной на рис. 5.3, где моделировался показательный закон распределения.

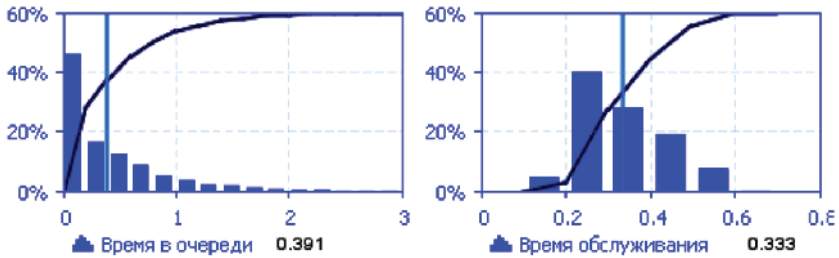


Рис. 5.5. Модельные значения времени обслуживания и обслуживания

3. Время обслуживания – константа, равная его математическому ожиданию (8 часов). Очевидно:

$$L_q = \frac{\rho^2 (1 + v_\mu^2)}{2(1 - \rho)} = \frac{\rho^2}{2(1 - \rho)} = \frac{2}{3}.$$

На рис. 5.6 и 5.7 представлены результаты моделирования для случая, когда время обслуживания постоянно. Очевидно, полученные данные близки по величине к варианту задания времени обслуживания в виде треугольного нечеткого числа.

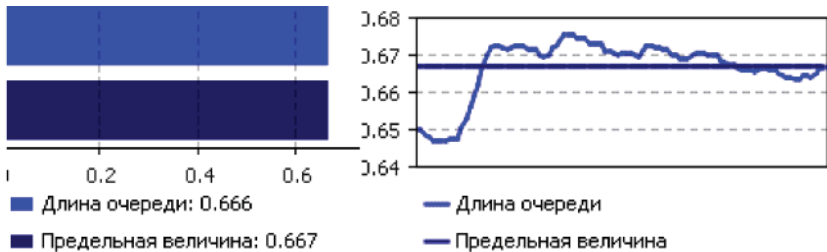


Рис. 5.6. Данные о длине очереди

Предельная длина очереди в два раза меньше, чем для простейшего закона обслуживания, определяемого показательным законом распределения.

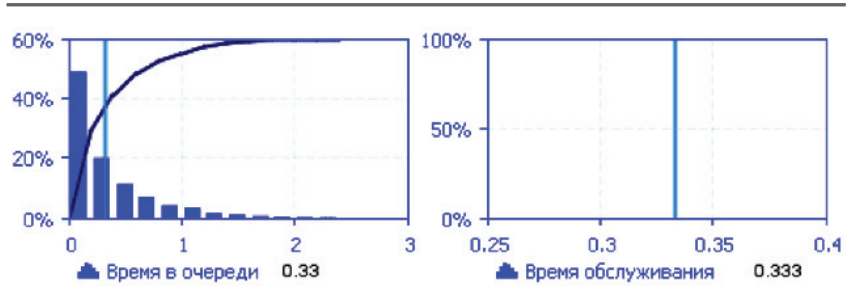


Рис. 5.7. Данные о времени проведенной заявкой в СМО

4. Оба потока описываются нечеткими числами.

Пусть время обслуживания задается нечетким числом $\tilde{A} = (4, 5, 15)$ (часы), а для интервала времени между поступлением заявок используется нечеткое число $\tilde{b} = (0,1; 0,6; 0,8)$ (дни).

В этом случае:

$$L_q \approx \frac{\rho^2 (v_{\lambda}^2 + v_{\mu}^2)}{2(1-\rho)} = 0,122.$$

На рис. 5.8 и 5.9 представлены результаты имитационного эксперимента по моделированию СМО в условиях, когда в основу распределения времени обслуживания и интервалов времени поступления заявок, положены нечеткие числа.

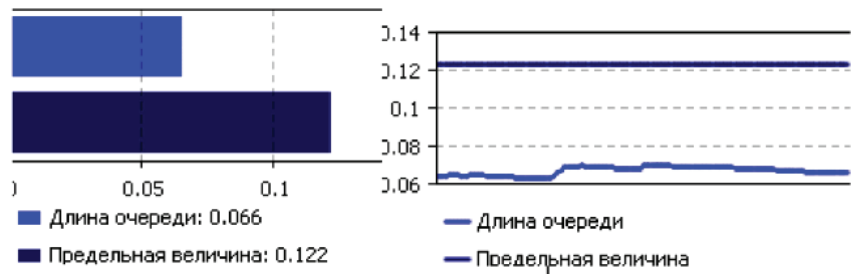


Рис. 5.8. Результаты моделирования по длине очереди

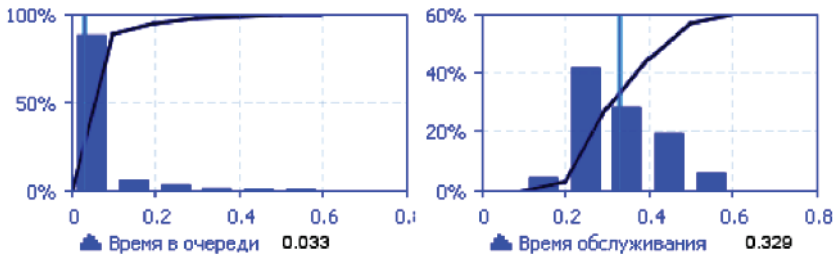


Рис. 5.9. Данные о времени обслуживания и ожидании в очереди

В данном случае расчетная длина очереди существенно меньше приближенной оценки ее предельной величины. Интервалы изменения возможных значений оценки длины очереди слева и справа, соответственно равны:

$$\frac{\rho(1-\nu_{\lambda}^2)}{2} \approx 0,3; \quad \frac{(1-\rho)(1-\nu_{\lambda}^2)}{2} \approx 0,15.$$

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Вентцель Е. С. Исследование операций: задачи, принципы, методология. М.: Наука, 1980. 208 с.

2. Осипов Г. С. Моделирование систем массового обслуживания с отказами // Бюллетень науки и практики. Электрон. журн. 2016. № 11 (12). С. 154–165. Режим доступа: <http://www.bulletennauki.com/osipov-g> (дата обращения 01.02.2017). DOI: 10.5281/zenodo.166801.

3. Осипов Г. С. Одноканальные системы массового обслуживания с неограниченной очередью в *AnyLogic* // Бюллетень науки и практики. Электрон. журн. 2016. № 8 (9). С. 92–95. Режим доступа: <http://www.bulletennauki.com/osipov> (дата обращения 15.08.2016). DOI: 10.5281/zenodo.60245.

4. Осипов Г. С. Оптимизация одноканальных систем массового обслуживания с неограниченной очередью // Бюллетень науки и практики. Электрон. журн. 2016. № 9 (10). С. 63–71. Режим доступа: <http://www.bulletennauki.com/osipov-gs> (дата обращения 15.09.2016). DOI: 10.5281/zenodo.154304.

5. Осипов Г. С. Исследование систем массового обслуживания с ожиданием в *AnyLogic* // Бюллетень науки и практики. Электрон. журн. 2016. № 10 (11). С. 139–151. Режим доступа: <http://www.bulletennauki.com/osipov-g-s> (дата обращения 15.10.2016). DOI: 10.5281/zenodo.161072.

6. Осипов Г. С. Системы массового обслуживания с ограниченной длительностью ожидания // Бюллетень науки и практики. Электрон. журн. 2016. № 12 (13). С. 28–36. Режим доступа: <http://www.bulletennauki.com/osipov-g-1> (дата обращения: 15.12.2016). DOI: 10.5281/zenodo.204623

7. Осипов Г.С., Осипова Е.В. Исследование систем массового обслуживания с «нетерпеливыми» заявками. // В сборнике: Материалы конференций ГНИИ «НАЦРАЗВИТИЕ». Октябрь 2016 Сборник

избранных статей. Ответственный за выпуск Л.А. Павлов. 2016. С. 41-44.

8. Осипов Г. С. Исследование систем массового обслуживания с приоритетными заявками // Бюллетень науки и практики. Электрон. журн. 2017. № 1 (14). С. 17–24. Режим доступа: <http://www.bulletennauki.com/osipovgs> (дата обращения 15.01.2017). DOI: 10.5281/zenodo.244131.

9. Осипов Г. С. Модели систем массового обслуживания с нечеткими треугольными и трапециевидными параметрами // 3-я Международная научно–практическая конференция «Актуальные вопросы науки и практики XXI в.» (Нижевартовск, 27-30 ноября 2016 г.): материалы. Нижневартовск: Издательский центр «Наука и практика», 2016. С. 11-20. Режим доступа: <http://www.konferenc.com/osipov> (дата обращения 10.12.2016). DOI: 10.5281/zenodo.199215.

Научное издание

Осипов Геннадий Сергеевич

**МАТЕМАТИЧЕСКОЕ И ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ
СИСТЕМ МАССОВОГО ОБСЛУЖИВАНИЯ**

Сетевое научное издание



© Осипов Г.С., 2017
© ИД «Академия Естествознания»
© АНО «Академия Естествознания»