

Министерство образования и науки Российской Федерации  
Ярославский государственный университет им. П. Г. Демидова  
Кафедра динамики электронных систем

**М. А. ДУБОВ**

# **ТЕОРИЯ СИСТЕМ МАССОВОГО ОБСЛУЖИВАНИЯ**

Практикум

*Рекомендовано  
Научно-методическим советом университета  
для студентов, обучающихся по направлению  
Инфокоммуникационные технологии и системы связи*

Ярославль  
ЯрГУ  
2015

УДК 519.872(076)  
ББК В183.5я73  
Д79

*Рекомендовано  
Редакционно-издательским советом университета  
в качестве учебного издания. План 2015 года*

Рецензент  
кафедра динамики электронных систем ЯрГУ

**Дубов, Михаил Андреевич.**  
Д79 Теория систем массового обслуживания : практи-  
кум / М. А. Дубов ; Яросл. гос. ун-т им. П. Г. Демидова.  
— Ярославль : ЯрГУ, 2015. — 44 с.

Практикум содержит краткие теоретические сведения по теории систем массового обслуживания и практические задания, предполагающие проведение аналитических расчетов и компьютерного моделирования.

Предназначен для студентов, обучающихся по направлению 11.03.02 (210700.62) Инфокоммуникационные технологии и системы связи (дисциплина «Теория телетрафика», цикл БЗ), очной формы обучения.

УДК 519.872(076)  
ББК В183.5я73

© ЯрГУ, 2015

## Введение

Теория систем массового обслуживания (СМО) является как самостоятельной наукой, так и основой более широкой и значимой в современных телекоммуникациях теории телетрафика. Задачей теории СМО является установление зависимостей между характером потока заявок, числом каналов (серверов), производительностью, дисциплиной и эффективностью обслуживания.

Основоположником теории является Агнер Краруп Эрланг, подробно изложивший в 1909 г. математическую модель телефонной станции. В статье приведены соотношения, связывающие количество каналов, нагрузку на станцию и вероятность отказа в обслуживании. Несмотря на более чем столетний прогресс в развитии телекоммуникаций, данные закономерности достаточно точно описывают процессы в сетях передачи речи и данных и различных иных СМО.

Математическая модель многих систем массового обслуживания строится на основе теории марковских процессов. В этом случае удастся относительно просто описать работу системы и получить аналитические зависимости. Главное условие применимости такого подхода — отсутствие значительного последствия (зависимости текущего состояния системы от ее поведения в прошлом).

В пособии преимущественно рассматриваются СМО, описываемые марковскими процессами: одноканальные, многоканальные, с очередями и без очередей, с «терпеливыми» и «нетерпеливыми» заявками. Приводятся краткие теоретические сведения и основные формулы, а также 33 практических задания, предполагающих проведение аналитических расчетов, небольших исследований и полноценного компьютерного моделирования СМО средствами Matlab и Simulink.

Последовательное успешное выполнение всех предложенных заданий способствует высокому уровню освоения основ теории СМО и возможности изучения более сложных вопросов теории телетрафика.

# 1. ПОТОКИ ВЫЗОВОВ

## 1.1. Способы задания потоков вызовов и их свойства

Существуют два способа задания входящих в СМО заявок на обслуживание (вызовов). В первом случае (рис. 1.1 а) задается вероятность  $P_k(t)$  поступления  $k$  вызовов на заданном интервале времени  $t$ . Во втором — плотность вероятности  $p(t)$  длительности пауз между вызовами (рис. 1.1 б). Данные способы равноправны, и используется более удобный.

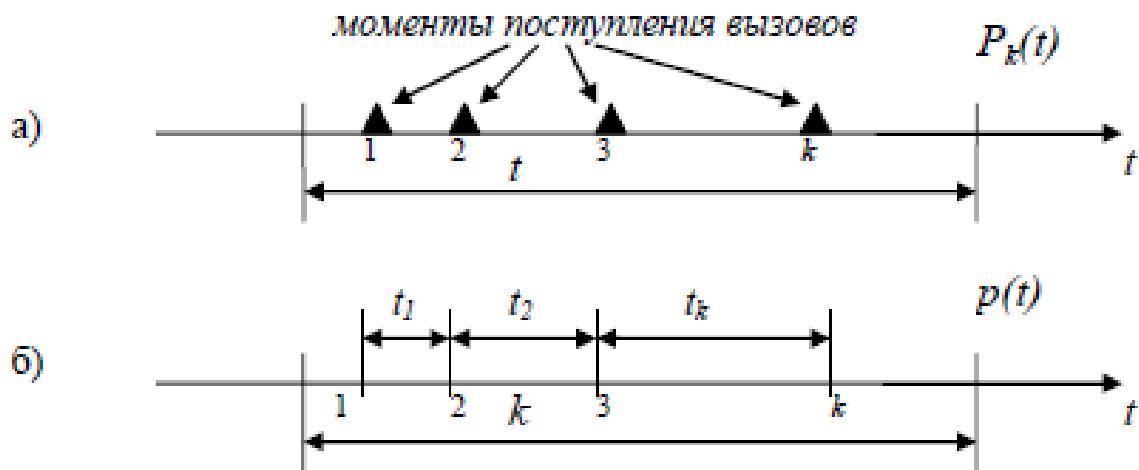


Рис. 1.1. Способы задания потока вызовов

Поток вызовов обладает следующими свойствами:

- ординарностью (вероятность поступления более чем одного вызова в конкретный момент времени ничтожно мала);
- стационарностью (независимость распределения вероятности от момента наблюдения);
- последствием (наличие зависимости от характера поступления предшествующих вызовов).

Если поток ординарный, стационарный и в нем отсутствует последствие, он называется *простейшим*. Основным параметром, характеризующим поток вызовов, является его *интенсивность*  $\lambda$ , или средняя длительность паузы между вызовами  $t_n$ . Очевидно, что эти величины являются математическими ожиданиями соответствующих распределений вероятности, описывающих поток.

Если число источников заявок достаточно мало, то суммарная интенсивность вызовов  $\lambda$  будет определяться выражением:

$$\lambda = (S - i)\alpha, \quad (1.1)$$

где  $S$  — общее количество источников,  $i$  — количество занятых источников,  $\alpha$  — интенсивность генерирования заявок каждым источником. Такой поток называется *примитивным*.

## **1.2. Основные распределения**

Многие потоки вызовов на своих интервалах стационарности хорошо попадают под классификацию простейших. При этом вероятность поступления определенного количества вызовов  $P_k(t)$  на заданном интервале времени описывается *распределением Пуассона*:

$$P_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}. \quad (1.2)$$

Математическое ожидание —  $\lambda$ , среднеквадратическое отклонение (СКО) —  $\sqrt{\lambda}$ , дисперсия —  $\lambda$ . Равенство математического ожидания и дисперсии является отличительной чертой распределения Пуассона.

Распределение длительностей пауз между вызовами простейшего потока описывается *экспоненциальным распределением*:

$$p(t) = \lambda e^{-\lambda t}. \quad (1.3)$$

Математическое ожидание —  $1/\lambda$ , среднеквадратическое отклонение (СКО) —  $1/\lambda$ , дисперсия —  $1/\lambda^2$ . Уникальным свойством экспоненциального распределения является отсутствие последствия, т. е. форма распределения не зависит от прошлого.

## **1.3. Генерирование потоков с заданным распределением**

Задача компьютерного моделирования потока вызовов с заданным распределением является отнюдь не простой, т. к. компьютер способен генерировать случайные числа только с равномерным распределением в диапазоне  $[0..1]$ . Соответственно, нужен способ преобразования равномерного распределения в желаемое.

Допустим, нужно сгенерировать случайные длительности пауз  $t_i$  между вызовами, подчиняющиеся экспоненциальному распределению. Воспользуемся тем, что интегральная функция вероятности монотонна и изменяется от 0 до 1. Тогда можно с помощью компьютера генерировать равномерно распределенные значения самой функции вероятности, а по ним восстанавливать значения по оси абсцисс (рис. 1.2).

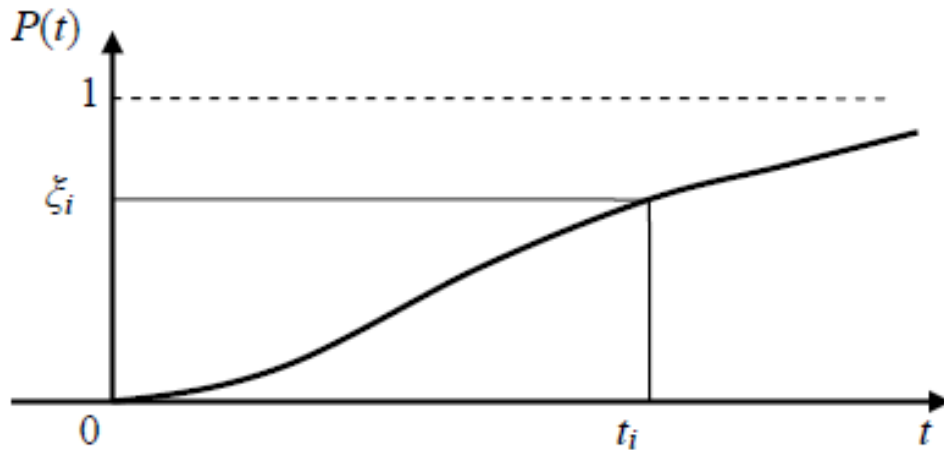


Рис. 1.2. Интегральная функция вероятности

В случае экспоненциального распределения удастся получить аналитическое выражение. Интегральная функция вероятности:

$$P(t) = \int_0^t \lambda e^{-\lambda t} dt = 1 - e^{-\lambda t}. \quad (1.4)$$

Пусть  $\xi_i$  — равномерно распределенная на интервале  $[0..1]$  случайная величина, тогда

$$1 - e^{-\lambda t_i} = \xi_i. \quad (1.5)$$

После небольшого преобразования можно еще несколько упростить правую часть, т. к. величина  $1 - \xi_i$  также распределена на  $[0..1]$ :

$$e^{-\lambda t_i} = 1 - \xi_i \longrightarrow e^{-\lambda t_i} = \xi_i. \quad (1.6)$$

В итоге получается выражение для искомой величины  $t_i$ , которая будет распределена экспоненциально:

$$t_i = -\frac{1}{\lambda} \ln \xi_i. \quad (1.7)$$

## 1.4. Практические задания

1. Дан результат 20 мин наблюдения за поступающими заявками на обслуживание (рис. 1.3). Чему равна интенсивность поступления заявок за время наблюдения? Какова точность полученной оценки интенсивности? Можно ли предположить, что поток заявок является простейшим?

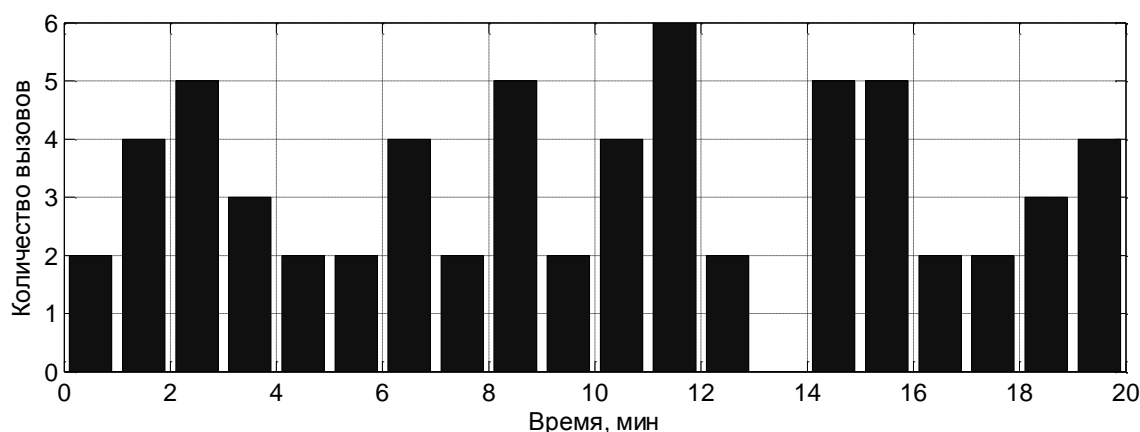


Рис. 1.3. Наблюдение за количеством поступающих вызовов

2. Дан результат измерения длительности пауз между последовательными 20 вызовами (рис. 1.4). Чему равна интенсивность потока на данном интервале наблюдения? Какова точность полученной оценки интенсивности? Можно ли предположить, что поток заявок является простейшим?

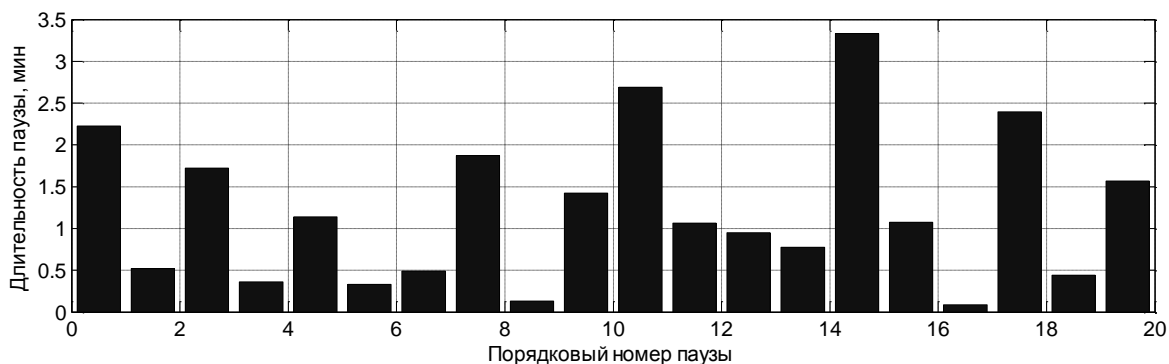


Рис. 1.4. Наблюдение за длительностью пауз между вызовами

3. На вход системы массового обслуживания поступает простейший поток заявок с интенсивностью 3 заявки/мин. Определите за время 5 мин вероятность непоступления ни одной заявки, поступления одной, двух и более трех.

4. Для простейшего потока с параметром  $\lambda = 299$  выз./ч определите значение  $k = k_m$ , при котором вероятность поступления  $k$  вызовов  $P_k(t)$  максимальна за время  $t = 89$  с. Рассчитайте величины вероятностей  $P_k(t)$  и постройте распределение вероятностей для  $k = k_m$ ;  $k = k_m \pm 2$ ;  $k = k_m \pm 4$ .

5. Концентратор обслуживает  $S = 4$  источника заявок. Для 5-минутного интервала времени  $t$  определите вероятность поступления ни одного, одного, двух и хотя бы одного вызова, если в начале интервала  $t$  все источники были свободны. Интенсивность свободного источника  $\alpha = 1$  выз./мин.

6. Реализуйте методику получения потока вызовов с экспоненциальным распределением длительности пауз между вызовами (п. 1.3) на языке Matlab. Сравните собственную реализацию с результатами работы соответствующей встроенной функции Matlab.

7. Предложите методику получения потока вызовов с распределением Пуассона количества поступающих вызовов в единицу времени и программный код на языке Matlab, ее реализующий. Сравните собственную реализацию с результатами работы соответствующей встроенной функции Matlab.

8. Используя средства пакета компьютерного моделирования Matlab, сгенерируйте простейший поток с интенсивностью  $\lambda = 3$  заявки/мин, состоящий из  $N = 1\,000$  заявок. Проведите статистическое исследование полученного потока заявок:

- найдите математическое ожидание, моду, дисперсию и СКО длительности паузы между вызовами;
- постройте гистограмму распределения длительности паузы между вызовами и сопоставьте ее графически с теоретическим распределением;
- обоснуйте выбор величины шага при построении гистограммы;
- проверьте статистическую гипотезу о виде распределения.

9. Постройте зависимость математического ожидания и дисперсии количества вызовов, приходящихся на интервал в 1 минуту, в зависимости от общей длительности наблюдаемого простейшего потока с интенсивностью  $\lambda = 3$  заявки/мин. Общую длительность потока рассматривать в пределах от 15 мин до 24 часов. Объясните полученный результат.



## 2. Марковские процессы

### 2.1. Определение марковского процесса. Расчет предельных вероятностей

Марковским процессом называется случайный процесс без памяти с дискретными состояниями и непрерывным временем. Число состояний может быть как конечным, так и бесконечным. Переходы между состояниями осуществляются в непрерывном времени с интенсивностями  $\lambda_{ij}$  [переходов/ед. времени], где  $i$  и  $j$  обозначают переход из состояния  $i$  в состояние  $j$ . Марковский процесс может быть представлен размеченным графом состояний (рис. 2.1).

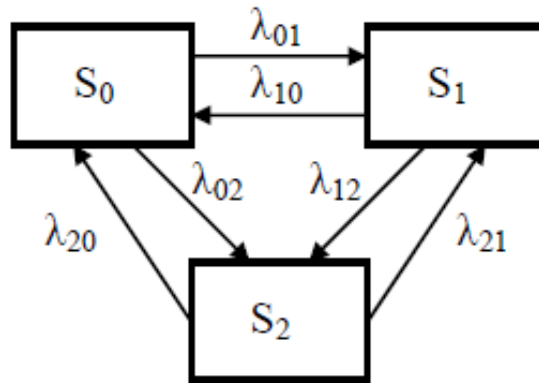


Рис. 2.1. Размеченный граф состояний

Основной характеристикой марковского процесса является вероятность пребывания системы в  $i$ -м состоянии в произвольный момент времени  $t - p_i(t)$ . Для нахождения вероятностей используется система линейных дифференциальных уравнений (уравнения Колмогорова), пример которой для графа с тремя состояниями приведен ниже:

$$\frac{d}{dt} \begin{pmatrix} p_1(t) \\ p_2(t) \\ p_3(t) \end{pmatrix} = \begin{pmatrix} -(\lambda_{12} + \lambda_{13}) & \lambda_{21} & \lambda_{31} \\ \lambda_{12} & -(\lambda_{21} + \lambda_{23}) & \lambda_{32} \\ \lambda_{13} & \lambda_{23} & -(\lambda_{31} + \lambda_{32}) \end{pmatrix} \begin{pmatrix} p_1(t) \\ p_2(t) \\ p_3(t) \end{pmatrix} \quad (2.1)$$

Например, изменение вероятности первого состояния  $\frac{d}{dt} p_1(t)$  складывается из исходящего потока из этого состояния  $-(\lambda_{12} + \lambda_{13})p_1(t)$  при условии, что система сейчас в нем находится, и входящего потока в это состояние  $\lambda_{21}p_2(t) + \lambda_{31}p_3(t)$  при условии, что система находится в других состояниях.

Для решения системы необходимо знание начальных условий  $p_i(0)$  и использование условия нормировки вероятностей:

$$p_1(t) + p_2(t) + p_3(t) = 1. \quad (2.2)$$

Если все состояния сообщаются друг с другом, то система переходит со временем в *стационарный режим* и вероятности состояний перестают меняться. Такие вероятности называются *предельными*:

$$p_i = \lim_{t \rightarrow \infty} (p_i(t)). \quad (2.3)$$

Смысл предельных вероятностей заключается в том, что они выражают относительное время пребывания системы в каждом из состояний. Найдены предельные значения могут быть путем решения системы уравнений:

$$\begin{cases} \begin{pmatrix} -(\lambda_{12} + \lambda_{13}) & \lambda_{21} & \lambda_{31} \\ \lambda_{12} & -(\lambda_{21} + \lambda_{23}) & \lambda_{32} \\ \lambda_{13} & \lambda_{23} & -(\lambda_{31} + \lambda_{32}) \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \\ p_1 + p_2 + p_3 = 1. \end{cases} \quad (2.4)$$

Начальные условия при нахождении предельных вероятностей значения не имеют.

## 2.2. Схема «гибели-размножения»

Частным случаем марковского процесса является схема «гибели-размножения», представленная на рис. 2.2. Отличительной чертой является возможность только последовательного перехода между состояниями. Такая схема хорошо подходит для представления систем массового обслуживания.

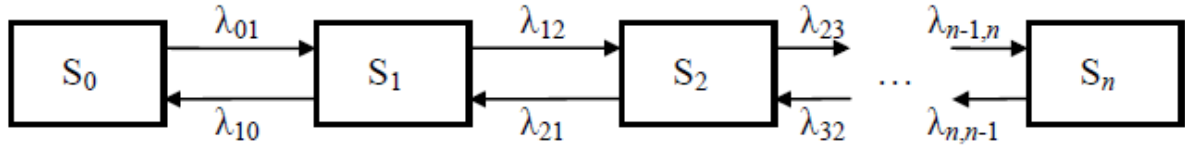


Рис. 2.2. Схема «гибели-размножения»

Поскольку все состояния сообщаются друг с другом, существуют предельные вероятности. Система (2.4) для их нахождения путем несложных преобразований в данном случае приходит к виду:

$$\begin{cases} \lambda_{01}p_0 = \lambda_{10}p_1; \\ \lambda_{12}p_1 = \lambda_{21}p_2; \\ \dots \\ \lambda_{n-1,n}p_{n-1} = \lambda_{n,n-1}p_n; \\ p_0 + p_1 + \dots + p_n = 1. \end{cases} \quad (2.5)$$

Предельную вероятность каждого состояния можно выразить через вероятность предыдущего состояния, что позволяет получить формулы для непосредственного вычисления предельных вероятностей:

$$p_0 = \left( 1 + \frac{\lambda_{01}}{\lambda_{10}} + \frac{\lambda_{01}\lambda_{12}}{\lambda_{10}\lambda_{21}} + \dots + \frac{\lambda_{01}\dots\lambda_{n-1,n}}{\lambda_{10}\dots\lambda_{n,n-1}} \right)^{-1}; \quad (2.6)$$

$$p_1 = \frac{\lambda_{01}}{\lambda_{10}} p_0; \quad p_2 = \frac{\lambda_{01}\lambda_{12}}{\lambda_{10}\lambda_{21}} p_0; \quad \dots; \quad p_n = \frac{\lambda_{01}\dots\lambda_{n-1,n}}{\lambda_{10}\dots\lambda_{n,n-1}} p_0.$$

### 2.3. Практические задания

1. Постройте в среде Matlab графики зависимости вероятностей состояний  $p_0(t)$  и  $p_1(t)$  от времени для марковского процесса, приведенного на рис. 2.3, при заданных интенсивностях переходов  $\lambda_{01}, \lambda_{10}$  [переходов/ед. времени] и начальных условиях  $p_0(0)$  и  $p_1(0)$ .

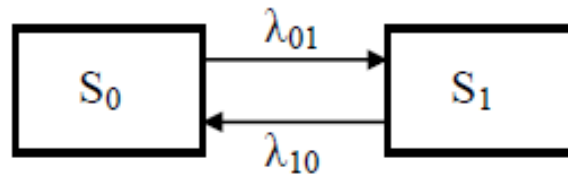


Рис. 2.3. Простейшая схема «гибели-размножения»

2. Используя результаты задания 1, докажите, что предельные вероятности состояний  $p_0$  и  $p_1$  не зависят от начальных условий  $p_0(0)$  и  $p_1(0)$ , а зависят только от интенсивности переходов  $\lambda_{01}, \lambda_{10}$ .

3. Постройте в среде Matlab гистограмму предельных вероятностей состояний  $p_0..p_n$  для схемы «гибели-размножения», приведенной на рис. 2.4, при заданных интенсивностях переходов  $\lambda, \mu$  и количестве состояний  $n$ .

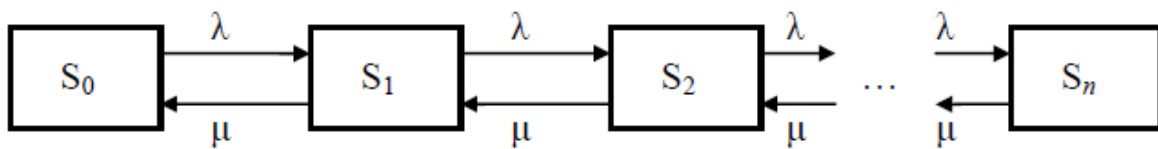


Рис. 2.4. Пример марковского процесса

### 3. Системы массового обслуживания с отказами

#### 3.1. Одноканальная СМО с отказами

Многие СМО можно представить в виде схемы «гибели-размножения». При этом под состояниями подразумеваются ситуации, когда система свободна, занят один или несколько каналов, занято несколько мест в очереди и т. д. Поскольку в самом определении марковского процесса заложено отсутствие памяти (последствия), это автоматически накладывает ограничение на поток переходов из состояния в состояние. Потoki должны быть простейшими (или примитивными). Поток, переводящий систему в состояния с большим номером, связывается с входящим потоком заявок с интенсивностью  $\lambda$ , а поток, переводящий систему в состояния с меньшим номером, связывается с потоком обслуживания с интенсивностью  $\mu$ , которая обратно пропорциональна средней длительности обслуживания одной заявки  $t_{\text{обс}}$ . Во многих реальных СМО такие потоки как раз являются простейшими.

На рис. 3.1 представлена схема одноканальной СМО с отказами. В соответствии с классификацией Кендалла — Башарина такая СМО обозначается  $M/M/1:0$ , где символы « $M$ » обозначают простейший входящий поток и поток обслуживания соответственно, «1» — один канал обслуживания, а «0» — отсутствие очереди. Состояние  $S_0$  соответствует свободной системе,  $S_1$  — состояние, когда единственный канал занят.

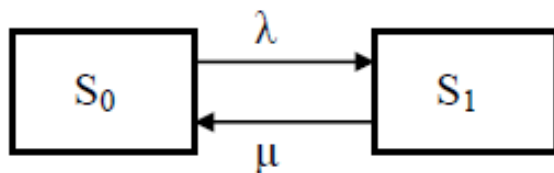


Рис. 3.1. Одноканальная СМО с отказами

Используя выражения (2.6), можно получить предельные вероятности состояний:

$$p_0 = \frac{\mu}{\mu + \lambda}; \quad p_1 = \frac{\lambda}{\mu + \lambda}. \quad (3.1)$$

Однако здесь и в следующих СМО удобнее использовать коэффициент нагрузки СМО  $\rho = \lambda/\mu$ . Тогда выражения (3.1) можно переписать так:

$$p_0 = \frac{1}{1+\rho}; \quad p_1 = \frac{\rho}{1+\rho}. \quad (3.2)$$

Вероятность  $p_1$  является самой главной характеристикой качества работы данной системы, а именно *вероятностью отказа*  $P_{\text{отк}} = p_1$ .

Следующая характеристика качества носит название *относительной пропускной способности*  $Q$  и отображает относительную долю заявок, получающих обслуживание:

$$Q = 1 - P_{\text{отк}} = p_0 = \frac{1}{1+\rho}. \quad (3.3)$$

*Абсолютная пропускная способность*  $A$  показывает абсолютное количество заявок, получающих обслуживание:

$$A = \lambda Q = \lambda(1 - P_{\text{отк}}) = \frac{\lambda}{1+\rho}. \quad (3.4)$$

Другие характеристики для такой простой системы определять нецелесообразно.

### 3.2. Многоканальная СМО с отказами

Система с отказами и числом каналов, равным  $n$ , приведена на рис. 3.2. Обозначается такая СМО символами  $M/M/n:0$ . Расшифровка состояний при этом следующая:  $S_0$  — система свободна,  $S_1$  — занят один канал (любой),  $S_2$  — заняты два канала (любых),  $S_n$  — заняты все  $n$  каналов. Интенсивность входящего потока не меняется в зависимости от занятости системы (считаем число источников заявок достаточно большим), а вот интенсивность обслуживания растет пропорционально числу занятых каналов. Это связано с тем, что для перехода системы в предыдущее состояние требуется освобождение не конкретного канала, а любого из занятых, что, естественно, произойдет с большей вероятностью, чем освобождение одного канала.

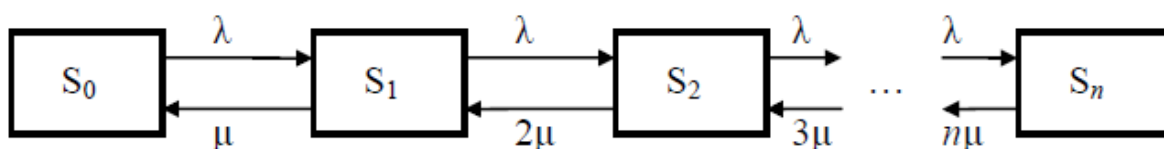


Рис. 3.2. Многоканальная СМО с отказами

Расчет основных характеристик, как и во всех остальных случаях, начинается с расчета предельных вероятностей:

$$p_0 = \left( 1 + \frac{\rho}{1!} + \frac{\rho^2}{2!} + \dots + \frac{\rho^n}{n!} \right)^{-1}; \quad p_k = \frac{\rho^k}{k!} p_0. \quad (3.5)$$

Формула для определения вероятности отказа носит название первой формулы Эрланга (В-формулы Эрланга):

$$P_{\text{отк}} = \frac{\rho^n / n!}{\sum_{k=0}^n \rho^k / k!}. \quad (3.6)$$

Относительная и абсолютная пропускная способности рассчитываются по аналогии с предыдущей системой.

Среднее количество занятых каналов  $n_3$  может быть найдено как отношение абсолютной пропускной способности к интенсивности обслуживания:

$$n_3 = \frac{A}{\mu} = \rho \left( 1 - \frac{\rho^n}{n!} p_0 \right). \quad (3.7)$$

Среднее количество заявок, находящихся в системе,  $N_{\text{СМО}}$  в данном случае равно числу занятых каналов, т. к. каждый канал может одновременно обслуживать только одну заявку, а очереди нет. Коэффициент занятости каналов равен отношению числа занятых каналов к общему числу каналов  $K_3 = n_3 / n$ .

Среднее время нахождения заявок в системе  $T_{\text{СМО}}$  может быть найдено через  $N_{\text{СМО}}$  по формуле Литтла:

$$T_{\text{СМО}} = \frac{N_{\text{СМО}}}{\lambda} = \frac{n_3}{\lambda}. \quad (3.8)$$

Коэффициент простоя каналов  $K_{\Pi}$  очевидно равен  $1 - K_3$ , причем эта величина обозначает еще и *вероятность занятости канала*  $P_{3K}$ , которая может быть выражена через *среднее время занятости канала*  $t_{3K}$  и *среднее время простоя канала*  $t_{3\text{пр}}$ :

$$P_{3K} = \frac{t_{3K}}{t_{3K} + t_{3\text{пр}}} = K_3. \quad (3.9)$$

Среднее время занятости канала  $t_{3K} = 1/\mu$ , тогда среднее время простоя одного канала с учетом и определяется выражением:

$$t_{3\text{пр}} = t_{3K} \frac{1 - P_{3K}}{P_{3K}} = \frac{1}{\lambda} \frac{n - \rho(1 - P_{\text{отк}})}{(1 - P_{\text{отк}})}. \quad (3.10)$$

### 3.3. Многоканальная СМО с отказами и ограниченным числом источников

В случае ограниченного числа источников заявок  $S$  поток на вход системы становится примитивным. Соответствующий граф состояний приведен на рис. 3.3. Интенсивность одного источника —  $\alpha$ , суммарная интенсивность всех источников —  $S\alpha$ . Подразумевается, что  $S > n$ , но при этом  $S$  достаточно мало и поступление одного вызова в систему значительно влияет на суммарный входящий поток.

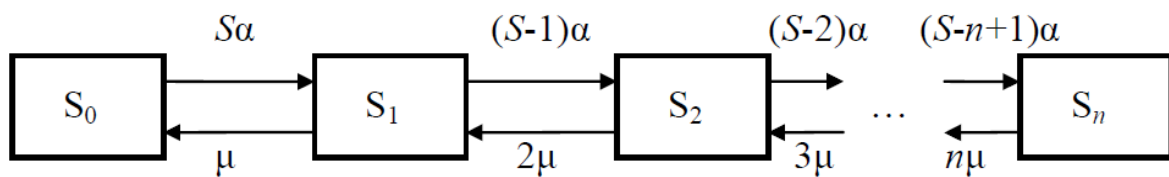


Рис. 3.3. Многоканальная СМО с отказами и ограниченным числом источников заявок

При обозначении нормированной нагрузки как  $\rho = \alpha/\mu$  выражения для предельных вероятностей принимают вид:



$$p_0 = \left( 1 + \frac{S\alpha}{1!\mu} + \frac{S(S-1)\alpha^2}{2!\mu^2} + \dots + \frac{S...(S-n+1)\alpha^n}{n!\mu^n} \right)^{-1} =$$

$$= \left( \sum_{k=0}^n C_S^k \rho^k \right)^{-1}; \quad (3.11)$$

$$p_k = \frac{S...(S-k+1)\alpha^k}{k!\mu^k} p_0 = C_S^k \rho^k p_0,$$

где  $C_S^k$  число сочетаний из  $S$  по  $k$ , выражаемое как

$$C_S^k = \frac{S!}{k!(S-k)!}. \quad (3.12)$$

Вероятность отказа определяется как вероятность последнего состояния, а сама формула носит название *формулы Энгсета*:

$$P_{\text{отк}} = \frac{C_S^n \rho^n}{\sum_{k=0}^n C_S^k \rho^k}. \quad (3.13)$$

Остальные характеристики системы определяются аналогично предыдущей.

### 3.4. Практические задания

1. Методика экспериментального расчета вероятности отказа основывается на измерении реального количества обслуженных заявок в единицу времени  $A$ , т. к. входящий поток заявок  $\lambda$  не всегда можно непосредственно измерить. Определите вероятность отказа одноканальной справочной системы, в которой экспериментально установлено, что в среднем обслуживается  $A = 9$  заявок в час, при этом среднее время обслуживания одной заявки  $t_{\text{обс}}$  составляет 90 с.

2. В 5-канальный call-центр поступает поток вызовов с интенсивностью  $\lambda = 25$  вызовов/час. Средняя длительность разговора составляет 4 минуты. Определите параметры эффективности работы call-центра: вероятность отказа, вероятность простоя всех 5 каналов, вероятность занятости не более 3 каналов одно-

временно, относительную пропускную способность, абсолютную пропускную способность, среднее количество занятых каналов, среднее время простоя каждого канала за 1 час работы.

3. Разработайте в среде Matlab алгоритм, рассчитывающий прибыль  $P$  (profit) [у. е.] от работы СМО с отказами в зависимости от числа каналов  $n$ , при следующих входных параметрах: доход от обслуживания одной заявки  $I$  (income) [у. е.], расход на содержание одного канала  $E$  (expense) [у. е.], интенсивность входящего потока заявок  $\lambda$  [заявок/час] и среднее время обслуживания  $t_{\text{обс}}$  [мин]. В результате работы алгоритма должны быть получены график зависимости прибыли  $P$  от числа каналов  $n$ , значение оптимального числа каналов  $n_{\text{опт}}$  и само значение максимальной прибыли  $P_{\text{макс}}$ .

4. Постройте графики зависимости абсолютной пропускной способности многоканальных СМО с отказами от нагрузки  $\rho$  при фиксированной интенсивности потока входящих заявок  $\lambda$ . Число каналов от 1 до 10.

5. Разработайте в среде Matlab алгоритм, позволяющий определить минимально необходимое число каналов СМО с отказами при заданной нагрузке  $\rho$  и вероятности отказа  $P_{\text{отк}}$ .

6. Сравните системы  $M/M/n:0$  с ограниченным и неограниченным числом источников заявок. Приведите оценку числа источников  $S$  в соотношении с числом каналов  $n$ , при котором необходимо пользоваться для нахождения вероятности отказа формулой Энгсета, а не Эрланга.

## 4. Системы массового обслуживания с очередями

### 4.1. Одноканальная СМО с ограниченной очередью

Одноканальная СМО с ограниченным числом мест для ожидания (очередью) обозначается как  $M/M/1:m$ , где  $m$  — количество мест в очереди. Соответствующий ей размеченный граф состояний приведен на рис. 4.1. Состояние  $S_0$  обозначает, что система свободна,  $S_1$  — занят один канал,  $S_2$  — канал занят и одна заявка в очереди,  $S_{m+1}$  — канал занят и все  $m$  мест в очереди также заняты, последующие заявки получают отказ в обслуживании. Таким образом, СМО является гибридной: с очередью и отказами. Интенсивность переходов в предыдущие состояния в отличие от многоканальной системы остается неизменной, т. к. длина очереди не влияет на скорость продвижения заявок по ней.

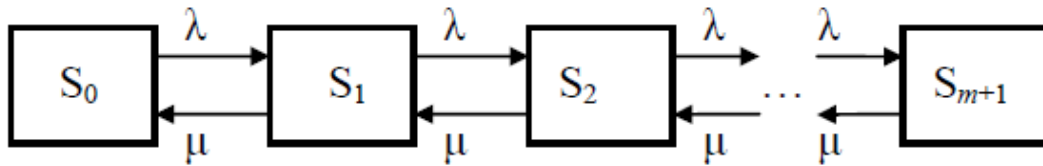


Рис. 4.1. Одноканальная СМО с ограниченной очередью

Предельные вероятности состояний:

$$p_0 = \left(1 + \rho + \rho^2 + \dots + \rho^{m+1}\right)^{-1} = \frac{1 - \rho}{1 - \rho^{m+2}}; \quad p_k = \rho^k p_0; \quad (4.1)$$

а в случае  $\rho = 1$  преобразуются к виду  $p_0 = p_k = 1/(m+2)$ .

Вероятность отказа в обслуживании:

$$P_{\text{отк}} = p_{m+1} = \rho^{m+1} p_0. \quad (4.2)$$

Вероятность попадания в очередь складывается из вероятностей попадания на первую, вторую и т. д. позицию в очереди, вплоть до последней, дальше наступает отказ в обслуживании:

$$P_{\text{оч}} = p_1 + p_2 + \dots + p_m = \frac{1 - \rho^m}{1 - \rho} \rho p_0 = \frac{1 - \rho^m}{1 - \rho^{m+2}} \rho. \quad (4.3)$$

Среднее число заявок, находящихся в очереди,  $N_{оч}$  можно найти, подставив предельные вероятности в следующее выражение:

$$N_{оч} = 1 \cdot p_2 + 2 \cdot p_3 + \dots + m \cdot p_{m+1}. \quad (4.4)$$

В результате средняя длина очереди равна:

$$N_{оч} = \rho^2 \frac{1 - \rho^m [1 + m(1 - \rho)]}{(1 - \rho)^2} p_0. \quad (4.5)$$

При  $\rho = 1$  формула преобразуется к виду:

$$N_{оч} = \frac{m(m+1)}{2(m+2)}. \quad (4.6)$$

Среднее число заявок на обслуживании  $N_{обс}$ , которое всегда равно среднему числу занятых каналов  $n_3$ , в данном случае проще определить по классической формуле:

$$N_{обс} = n_3 = \sum_{k=1}^n k p_k + \sum_{i=1}^m n p_i = 1 \cdot p_1 + 1 \cdot (p_2 + \dots + p_{m+1}) = 1 - p_0. \quad (4.7)$$

Среднее число заявок в СМО складывается из среднего числа заявок в очереди и среднего числа заявок на обслуживании:

$$N_{СМО} = N_{оч} + N_{обс}. \quad (4.8)$$

Среднее время ожидания в очереди  $T_{оч}$  находится с помощью формулы Литтла:

$$T_{оч} = \frac{N_{оч}}{\lambda}, \quad (4.9)$$

а среднее время обслуживания заявок  $T_{обс}$  равно

$$T_{обс} = \frac{Q}{\mu}. \quad (4.10)$$

Тогда среднее время нахождения заявки в СМО равно:

$$T_{СМО} = T_{оч} + T_{обс}. \quad (4.11)$$

Среднее время простоя канала определяется аналогично (3.10):

$$t_{пр} = \frac{1 - \rho(1 - P_{отк})}{\lambda(1 - P_{отк})}. \quad (4.12)$$

При  $m = 0$  (очередь отсутствует) все формулы переходят к виду для системы с отказами, а при  $m \rightarrow \infty$  система превращается в систему с неограниченной очередью.

## **4.2. Одноканальная СМО с неограниченной очередью**

При снятии ограничения на количество мест в очереди получается безотказная одноканальная СМО, приведенная на рис. 4.2. Согласно классификации Кендалла — Башарина такая СМО обозначается как  $M/M/1:\infty$  или сокращенно  $M/M/1$ . По-прежнему состояние  $S_0$  соответствует свободной системе,  $S_1$  — занятому каналу и отсутствию очереди,  $S_2$  — занятому каналу и одной заявке в очереди и т. д.

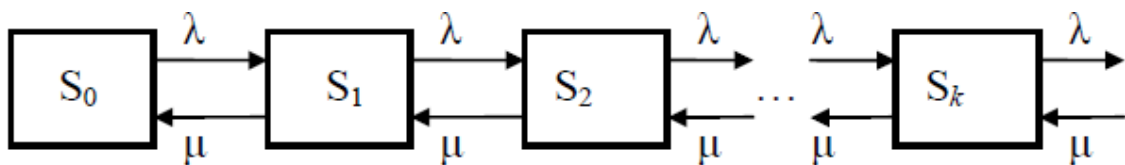


Рис. 4.2. Одноканальная СМО с неограниченной очередью

Основные характеристики обслуживания могут быть получены при переходе  $m \rightarrow \infty$  из формул п. 4.1. Однако при  $\rho \geq 1$  возникает ситуация неограниченного роста очереди. Очевидно, что в этом случае поток входящих заявок превышает поток обслуживания и один канал не в состоянии справиться с работой. Такой случай на практике не рассматривают, полагая  $\rho < 1$ .

Предельные вероятности состояний:

$$p_0 = 1 - \rho; \quad p_k = \rho^k (1 - \rho). \quad (4.13)$$

Поскольку очередь неограниченна:

$$P_{\text{отк}} = 0; \quad Q = 1; \quad A = \lambda. \quad (4.14)$$

Вероятность попадания в очередь в данном случае является главной характеристикой и определяется из :

$$P_{\text{оч}} = p_1 + p_2 + \dots = \rho. \quad (4.15)$$

Среднее количество заявок в очереди и на обслуживании:

$$N_{\text{оч}} = \frac{\rho^2}{1-\rho}; \quad N_{\text{обс}} = n_3 = \rho. \quad (4.16)$$

Общее число заявок в системе:

$$N_{\text{СМО}} = N_{\text{оч}} + N_{\text{обс}} = \frac{\rho}{1-\rho}. \quad (4.17)$$

Среднее время ожидания в очереди и на обслуживании:

$$T_{\text{оч}} = \frac{N_{\text{оч}}}{\lambda} = \frac{\rho^2}{\lambda(1-\rho)}; \quad T_{\text{обс}} = \frac{N_{\text{обс}}}{\lambda} = \frac{1}{\mu}. \quad (4.18)$$

Среднее время нахождения заявки в системе:

$$T_{\text{СМО}} = \frac{N_{\text{СМО}}}{\lambda} = \frac{1}{\mu - \lambda}. \quad (4.19)$$

Среднее время простоя канала:

$$t_{\text{пр}} = \frac{1-\rho}{\lambda} = \frac{\mu - \lambda}{\mu\lambda}. \quad (4.20)$$

### 4.3. Многоканальная СМО с очередью

Случай многоканальной системы с ограниченной  $M/M/n:m$  или неограниченной очередью  $M/M/n:\infty$  ( $M/M/n$ ) является обобщением ранее рассмотренных СМО. Граф состояний для системы с ограниченной очередью приведен на рис. 4.3. Интенсивность обслуживания растет с ростом числа занятых каналов, но прекращает свой рост при поступлении заявок в очередь. Состояние  $S_{n+1}$  соответствует ситуации, когда заняты все каналы и одна заявка в очереди,  $S_{n+m}$  — когда заняты все каналы и места для ожидания.

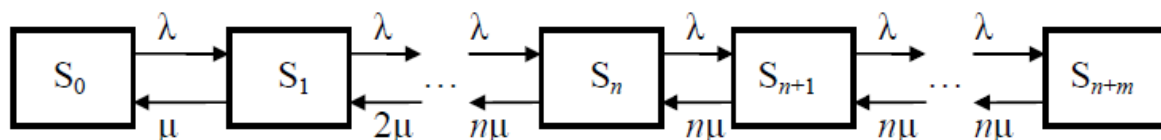


Рис. 4.3. Многоканальная СМО с очередью

Для многоканальной системы удобно введение нормированной нагрузки  $\chi = \rho/n$ . Тогда предельные вероятности могут быть выражены следующим образом:

$$p_0 = \left( \sum_{k=0}^{n-1} \frac{\rho^k}{k!} + \frac{\rho^n}{n!} \frac{1-\chi^{n+1}}{1-\chi} \right)^{-1}; \quad (4.21)$$

$$p_k = \frac{\rho^k}{k!} p_0 \quad (k = 0..n); \quad p_{n+k} = \frac{\rho^{n+k}}{n^k n!} p_0.$$

Вероятность отказа в системе с ограниченной очередью, как всегда, равна вероятности попадания в последнее состояние системы:

$$P_{\text{отк}} = p_{n+m} = \frac{\rho^{n+m}}{n^m n!} p_0. \quad (4.22)$$

В случае бесконечной очереди и при условии  $\chi < 1$  (очередь не будет бесконечно расти)  $p_0$  преобразуется к виду:

$$p_0 = \left( \sum_{k=0}^{n-1} \frac{\rho^k}{k!} + \frac{\rho^n}{n!} \frac{1}{1-\chi} \right)^{-1}. \quad (4.23)$$

Вероятность попадания в очередь (на любую позицию) в случае бесконечной очереди описывается *второй формулой Эрланга (С-формулой Эрланга)*:

$$P_{\text{оч}} = \frac{\frac{\rho^n}{n!} \frac{1}{1-\chi}}{\sum_{k=0}^{n-1} \frac{\rho^k}{k!} + \frac{\rho^n}{n!} \frac{1}{1-\chi}}. \quad (4.24)$$

Среднее число заявок в ограниченной очереди:

$$N_{\text{оч}} = \chi \frac{\rho^n}{n!} \frac{1-\chi^n [1+m(1-\chi)]}{(1-\chi)^2} p_0, \quad (4.25)$$

а в неограниченной очереди:

$$N_{\text{оч}} = \frac{\rho^n}{n!} \frac{\chi}{(1-\chi)^2} p_0. \quad (4.26)$$

Среднее время простоя канала при ограниченной очереди эквивалентно , а в случае неограниченной преобразуется к виду:

$$t_{\text{пр}} = \frac{n - \rho}{\lambda}. \quad (4.27)$$

Остальные характеристики находятся аналогично ранее описанным СМО.

#### 4.4. СМО с «нетерпеливыми» заявками

В реальных СМО вполне возможна ситуация, когда заявки из очереди будут уходить самостоятельно, не дождавсь обслуживания («нетерпеливые» заявки). Такой поток заявок также считается простейшим и характеризуется *средним временем терпения заявки в очереди*  $t_{\text{тер}}$  или *интенсивностью уходов*  $\nu = 1/t_{\text{тер}}$ . Интенсивность уходов будет возрастать пропорционально росту очереди. Граф состояний, соответствующий такой системе, представлен на рис. 4.4.

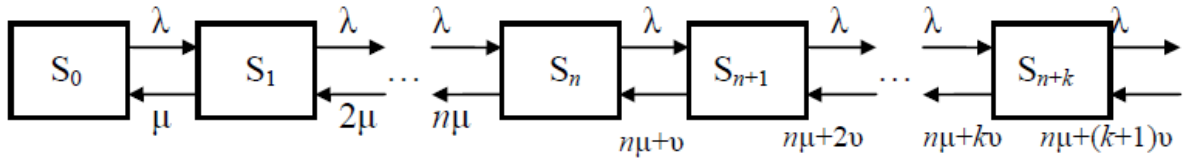


Рис. 4.4. СМО с «нетерпеливыми» заявками

После введения *нормированной интенсивности уходов*  $\beta = \nu/\mu$  предельные вероятности состояний выражаются следующим образом:

$$p_0 = \left( \sum_{k=0}^n \frac{\rho^k}{k!} + \frac{\rho^n}{n!} \sum_{k=1}^{\infty} \frac{\rho^k}{\prod_{l=1}^k (n-l\beta)} \right)^{-1}; \quad (4.28)$$

$$p_k = \frac{\rho^k}{k!} p_0 \quad (k = 0..n); \quad p_{n+k} = \frac{\rho^n}{n!} \frac{\rho^k}{\prod_{l=1}^k (n-l\beta)} p_0 \quad (k = 1..\infty).$$



Суммы, входящие в выражения для определения вероятностей, будут сходиться при любом  $\rho$ , следовательно, бесконечная очередь в такой системе никогда не образуется по сравнению с системой с «терпеливыми» заявками. Однако, поскольку не все заявки получают обслуживание, абсолютная и относительная пропускные способности будут выражаться следующим образом:

$$A = \lambda - \nu N_{yx}, \quad (4.29)$$

$$Q = \frac{A}{\lambda} = 1 - \frac{\nu}{\lambda} N_{yx}, \quad (4.30)$$

где  $N_{yx}$  — среднее количество уходящих из очереди заявок. Вероятность ухода заявок из очереди будет эквивалентна вероятности отказа  $P_{отк} = 1 - Q$ .

Среднее число занятых каналов:

$$n_3 = \frac{A}{\mu} = \rho - N_{yx}\beta. \quad (4.31)$$

Тогда среднее число уходящих заявок может быть выражено через  $n_3$ :

$$N_{yx} = \frac{\rho - n_3}{\beta}, \quad (4.32)$$

а количество занятых каналов найдено по классической формуле:

$$n_3 = \sum_{k=1}^n kp_k + \sum_{k=1}^m np_{n+k} = \sum_{k=1}^n kp_k + n \left( 1 - \sum_{k=0}^n p_k \right). \quad (4.33)$$

Среднее число заявок в очереди находится по классической формуле:

$$N_{оч} = \sum_{k=1}^{\infty} kp_{n+k}, \quad (4.34)$$

суммирование в которой может быть ограничено конечным, но достаточно большим числом слагаемых.

Остальные характеристики находятся аналогично предыдущим случаям.

## 4.5. Практические задания

1. Разработайте в среде Matlab алгоритм, решающий следующую задачу. В гипермаркете  $n_{\text{макс}}$  касс, одновременно работают не все. Среднее время обслуживания одного покупателя  $t_{\text{обс}}$  [мин]. Посетители сами выбирают кассу таким образом, что загрузка касс примерно одинакова. Требуется построить график зависимости минимально необходимого числа работающих касс  $n_{\text{мин}}$  от суммарного потока посетителей на входе гипермаркета  $\lambda$  [посетителей/мин]. Критерий оптимизации — средняя длина очереди не более  $N_{\text{оч}}$  человек.

2. В условии задания 3 из п. 3.4 добавьте конечное число мест для ожидания в очереди  $m$ . Затраты на содержание одного места для ожидания —  $E_m$  [у. е.]. Модернизируйте алгоритм получения графика зависимости прибыли  $P$  от числа каналов  $n$  при различных  $m$ , расчета значения оптимального числа каналов  $n_{\text{опт}}$  и мест для ожидания  $m_{\text{опт}}$  и самого значения максимальной прибыли  $P_{\text{макс}}$ .

3. Для системы  $M/M/n$  с неограниченной очередью, входным простейшим потоком  $\lambda$  [заявок/мин] и средним временем обслуживания  $t_{\text{обс}}$  [мин] разработайте алгоритм, на выходе которого будет гистограмма вероятности времени ожидания в очереди.

4. Разработайте алгоритм, отображающий в результате своей работы зависимость минимально необходимого числа операторов почтового отделения  $n_{\text{мин}}$  от потока посетителей  $\lambda$  [человек/час]. Среднее время обслуживание одного человека —  $t_{\text{обс}}$  [мин]. Среднее время терпения в очереди —  $t_{\text{тер}}$  [мин]. Очередь не ограничена. Критерий оптимизации — число уходящих без обслуживания не более  $N_{\text{ух}}$  [человек/час].

5. Рассчитайте аналитически основные характеристики качества работы многоканальной СМО замкнутого типа. Общее число источников —  $S$ , интенсивность каждого источника —  $\alpha$ , количество каналов —  $n$ , количество мест в очереди  $m = S - n$ , размеченный граф состояний для системы приведен на рис. 4.5.

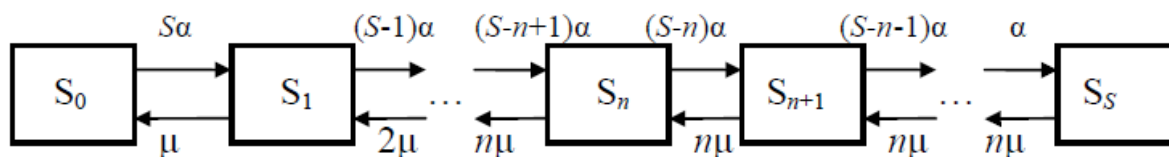


Рис. 4.5. Замкнутая СМО

## 5. Сравнение различных систем массового обслуживания

В данном разделе предлагается решить три типичные задачи теории СМО, опираясь на знания, полученные при освоении предыдущих разделов. Задания предполагают проведение теоретических расчетов и построение в зависимости от нагрузки на систему соответствующих графиков, иллюстрирующих преимущество одной из сравниваемых систем над другой.

1. Сравните по основным характеристикам (вероятность отказа, пропускная способность) систему массового обслуживания с двумя обслуживающими серверами (рис. 5.1 слева) и систему массового обслуживания с одним обслуживающим сервером, но с удвоенной производительностью (рис. 5.1 справа). Все системы с отказами.

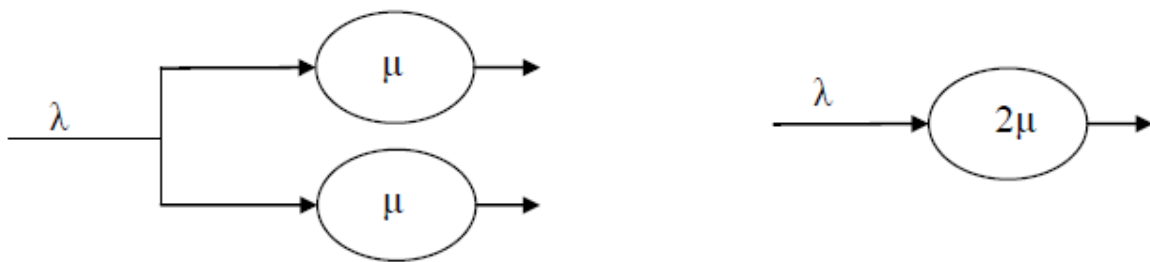


Рис. 5.1. Сравнение одноканальной и многоканальной СМО  
с отказами

2. То же, что в задании 1, только все системы с неограниченной очередью (рис. 5.2). Характеристики для сравнения — среднее время ожидания в очереди и среднее число занятых каналов.

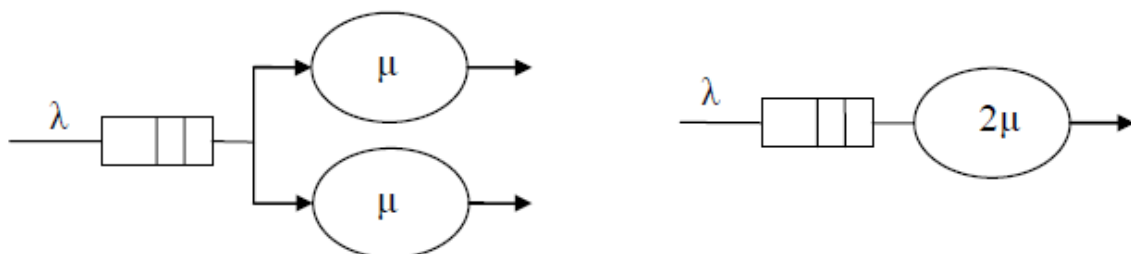


Рис. 5.2. Сравнение одноканальной и многоканальной СМО  
с очередями

3. Сравните по среднему времени ожидания в очереди системы массового обслуживания с общей очередью к двум серверам (рис. 5.3 слева) и с отдельной очередью к каждому серверу (рис. 5.3 справа). Производительность серверов одинаковая.

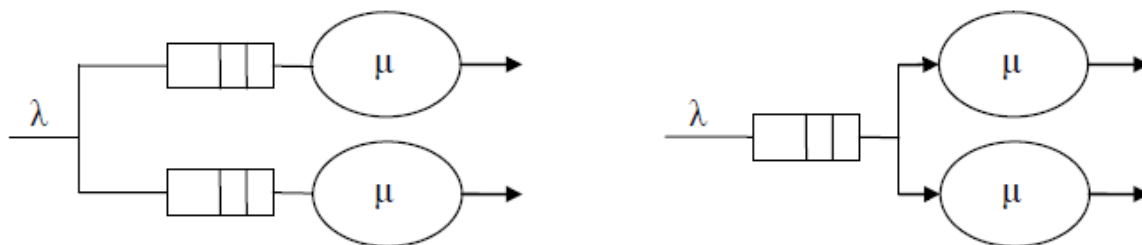


Рис. 5.3. Сравнение общей и отдельных очередей

## 6. Моделирование СМО в среде Simulink

### 6.1. Описание библиотеки SimEvents

Во всех предыдущих случаях внимание было уделено лишь средним вероятностным характеристикам качества работы систем массового обслуживания. Для более глубокого понимания процессов, происходящих в СМО, может потребоваться полноценное моделирование ее работы в течение определенного интервала времени, непосредственное наблюдение за входящими, обслуженными, отклоненными и ожидающими в очереди заявками. Такое исследование может быть проведено средствами библиотеки SimEvents пакета Simulink.

Рассмотрим основные функциональные блоки, которые потребуются при моделировании. Первым необходимым абсолютно для любой системы является блок, генерирующий входящие в систему заявки, «Time-Based Entity Generator» (рис. 6.1).

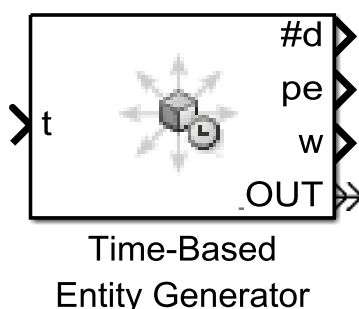


Рис. 6.1. Блок генерации заявок

Первая вкладка его настроек «Entity Generation» содержит несколько опций. «Generate entities upon» позволяет выбрать источник информации о скорости генерирования заявок: диалог («Intergeneration time from dialog») или внешний порт («Intergeneration time from port t»). Если выбрать первый вариант, то станут активны следующие опции. «Distribution» — вид распределения длительности пауз между генерируемыми заявками (фиксированное время «Constant», равномерное распределение «Uniform», экспоненциальное распределение «Exponential»). В рамках данного практикума используется только экспоненци-

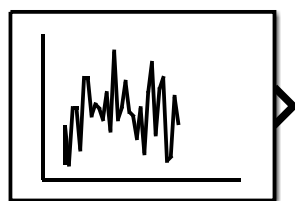
альное распределение длительности пауз, поэтому все остальные настройки будут приведены для него. «Initial seed» — число, инициализирующее генератор случайных чисел. Выбирается произвольно. «Mean» — средняя длительность паузы, равна  $1/\lambda$ . «Generation event priority» — приоритет генерируемой заявки. Поскольку работа с приоритетами не предусмотрена, то значение произвольно. «Generate entity at simulation start» позволяет сразу сгенерировать заявку при старте моделирования. В большинстве случаев не требуется.

Вторая вкладка настроек «Blocking» позволяет задать реакцию генератора заявок на заблокированный выход. «Response when blocked» определяет действие при заблокированном выходе: приостановить заявку или сообщить об ошибке. «Response when unblocked» описывает реакцию на разблокировку приемника заявок. Либо немедленно сгенерировать заявку, либо подождать в соответствии с заданным законом распределения длительности пауз. Для всех моделей практикума используются настройки «Pause generation» и «Delayed restart».

Третья вкладка «Entity type» задает тип генерируемого события. Варианты «Blank» и «Standard» отличаются наличием или отсутствием атрибутов по умолчанию у генерируемой заявки. Достаточно выбрать вариант «Blank».

Четвертая вкладка содержит функции включения возможных портов вывода информации от источника заявок: «Number of entities departed» — текущее количество отправленных заявок, «Pending entity present in block» — текущее количество задержанных к отправке заявок из-за блокировки выхода, «Average intergeneration time» — фактическое среднее время длительности паузы между генерируемыми заявками. Результаты могут быть выведены как на блок «Display» из библиотеки Simulink, отображающий численное значение, так и на осциллограф «Signal Scope» из библиотеки SimEvents.

Блок «Event-Based Random Number» является генератором случайных чисел с заданным законом распределения (рис. 6.2). Используется преимущественно для задания временных интервалов, соответствующих длительности паузы между заявками или длительности обслуживания заявки.



Event-Based  
Random Number

Рис. 6.2. Генератор случайных чисел

В настройках можно выбрать число «Initial seed», инициализирующее генератор случайных чисел, для получения при необходимости различных реализаций случайного процесса. Остальные настройки связаны с заданием параметров выбранного распределения. Самих распределений достаточно много, но в рамках практикума используется только экспоненциальное.

Блок «*Signal Scope*» представляет собой специальный осциллограф, предназначенный для работы с другими блоками библиотеки SimEvents (рис. 6.3). Его настройки доступны по щелчку по кнопке «Parameters» в левом верхнем углу осциллографа. В них можно подписать получаемый график, задать масштаб осей, тип линии и некоторые другие параметры.



Signal Scope

Рис. 6.3. Анализатор потока событий

Блок «*Single Server*» моделирует работу одиночного обслуживающего сервера (рис. 6.4). На первой вкладке настроек «Single Server» можно выбрать источник информации о длительности обслуживания заявок: диалог («Dialog»), сигнальный порт («Signal port t») или атрибут самой заявки («Attribute»). В первом случае время обслуживания каждой заявки будет фиксировано. Во втором — через внешний порт можно задать сигнал, определяющий время обслуживания в соответствии с каким-либо видом статистического распределения. В третьем

время обслуживания заложено в атрибуте самой заявки. Очевидно, что среднее время обслуживания заявок —  $1/\mu$ .

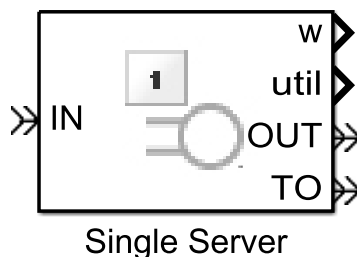


Рис. 6.4. Одноканальный сервер

Вторая вкладка «Preemption» и опция «Service completion event priority» первой вкладки не рассматриваются, т. к. не предусмотрена работа с приоритетными заявками.

Третья вкладка «Timeout» позволяет включить дополнительный выход «ТО» из сервера для заявок, у которых истекло время терпения нахождения в системе («нетерпеливые» заявки).

Четвертая вкладка «Statistics» содержит список возможных информационных сигналов с сервера: «Number of entities departed» — текущее число покинувших сервер заявок; «Number of entities in block» — текущее количество заявок на обслуживании; «Pending entity present in block» — текущее количество задержанных в блоке из-за блокировки выхода заявок; «Average wait» — фактическое среднее время нахождения заявки на обслуживании; «Utilization» — коэффициент занятости сервера, он же среднее число занятых каналов; «Number of entities timed out» — текущее число заявок, покинувших сервер по причине завершения времени терпения.

Блок «N-Server» моделирует работу  $N$  независимых параллельных серверов (рис. 6.5). Многие настройки аналогичны настройкам блока «Single Server». Дополнительно появляется опция «Number of servers», позволяющая задать число серверов. Вторая вкладка «Service control» позволяет извне управлять работой серверов, например приостанавливать их работу по некому событию, но в данной работе не используется. На вкладке «Statistics» появляются новые выходы для анализа: «Pending entity present in block» — сигнал присутствия задержанных



в блоке из-за блокировки выхода заявок и «Server occupancy» — сигнал занятости каждого из серверов ( $N$ -размерный сигнал).

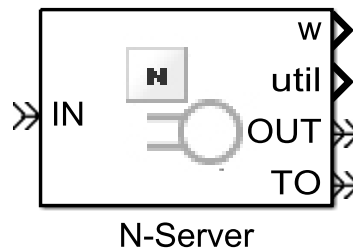


Рис. 6.5. Многоканальный сервер

Блок «*Entity Sink*» является необходимым элементом каждой модели и представляет собой приемник заявок неограниченной емкости, если включена опция «Input port available for entity arrivals» (рис. 6.6). Блок позволяет показать количество поступивших в него заявок после включения опции «Report number of entities arrived».

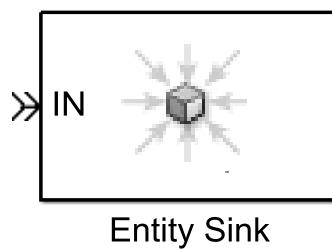


Рис. 6.6. Приемник заявок

Блок «*Set Attribute*» позволяет задать заявке на обслуживание дополнительные атрибуты, по которым в дальнейшем система будет определять, каким образом обрабатывать эту заявку (рис. 6.7). Например, как атрибут заявки можно задать длительность ее обслуживания. Такой вариант равносильен тому, что время обслуживания задается в самом сервере для заявки без установки атрибута, но является более предпочтительным.

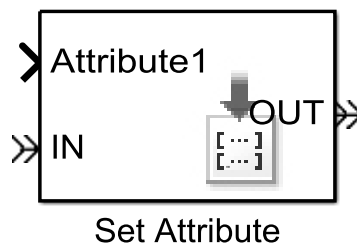


Рис. 6.7. Блок установки атрибутов

На первой вкладке «Attribute» приведен список атрибутов заявки. Каждому можно задать имя и значение, которое может определяться как в диалоге, так и через внешний порт.

Блок «*Output Switch*» позволяет разделить входящий в него поток заявок по нескольким направлениям в соответствии с правилом (рис. 6.8). На первой вкладке задается непосредственно число выходных путей «Number of entity output ports» и правило выбора пути «Switching criterion». Возможные правила: «Round robin» — каждая следующая заявка подается на следующий свободный порт по кругу; «Equiprobable» — равновероятный выбор; «First port that is not blocked» — первый свободный порт; «From signal port p» — по внешнему сигналу; «From attribute» — по соответствующему атрибуту.

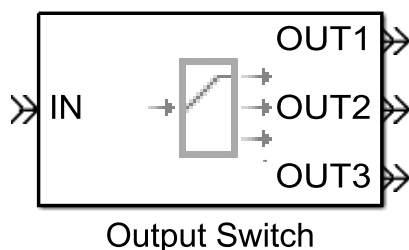


Рис. 6.8. Коммутатор потоков

На второй вкладке «Timeout» имеется возможность включить дополнительный выход «ТО» для заявок с завершившимся временем терпения. На третьей вкладке появляется новый сигнал для анализа «Last entity departure port», отображающий порт, который был выбран для покинувшей блок заявки.

Блок «*Path Combiner*» позволяет объединить заявки, приходящие разными путями, на один выход (рис. 6.9). На первой вкладке задается число объединяемых путей «Number of entity input ports». На второй вкладке «Status Notification» задается правило открытия входящих портов для приема заявок «Input port precedence»: «IN1 port» — первый порт; «Round robin» — открывается следующий за предыдущим принявшим заявку портом по кругу; «Equiprobable» — равновероятный выбор; «From signal port p» — по внешнему сигналу. На третьей вкладке интерес представляет сигнал для анализа «Last entity arrival port», отображающий порт, с которого была взята текущая заявка.

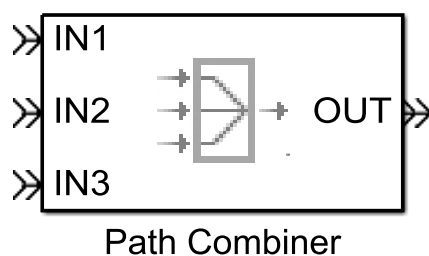


Рис. 6.9. Агрегатор потоков

Блок «*FIFO Queue*» позволяет организовать очередь из заявок по принципу «первый вошел — первый вышел» (рис. 6.10). На первой вкладке настроек задается размер очереди «Capacity», который, в частности, может принимать значение «inf», что соответствует неограниченной очереди. На второй вкладке «Timeout» включается порт выхода заявок «ТО», потерявших терпение при ожидании. Именно в данном блоке эта опция наиболее актуальна, т. к. заявки на практике уходят именно из очереди, а не в процессе обслуживания.

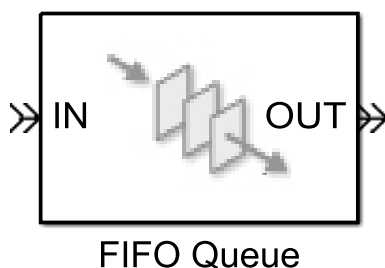


Рис. 6.10. Блок очереди

На третьей вкладке приведен список доступных для анализа сигналов: «Number of entities departed» — текущее количество выпущенных из очереди заявок; «Number of entities in queue» — текущее число заявок в очереди; «Average wait» — среднее время ожидания в очереди; «Average queue length» — средняя длина очереди; «Number of entities timed-out» — текущее число заявок, покинувших очередь, не дождавшись обслуживания.

Блок «*Start Timer*» запускает таймер для каждой проходящей через него заявки (рис. 6.11). Практически не имеет настроек, но содержит имя таймера, т. к. таймеров для одной заявки может быть несколько.

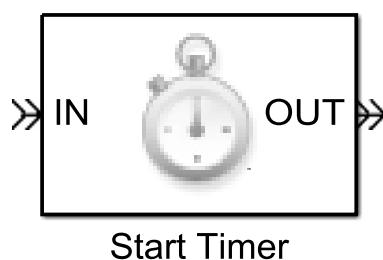


Рис. 6.11. Блок запуска таймера

Блок «*Read Timer*» считывает данные таймера, прикрепленного к заявке, для определения длительности пребывания заявки в части системы между блоками «*Start Timer*» и «*Read Timer*» (рис. 6.12). На второй вкладке интерес представляют два сигнала: «*Elapsed time*» — текущее значение таймера для каждой проходящей заявки и «*Average elapsed time*» — среднее значение таймера по всем заявкам. Такая связка наилучшим образом подходит для измерения времени нахождения заявки в СМО.

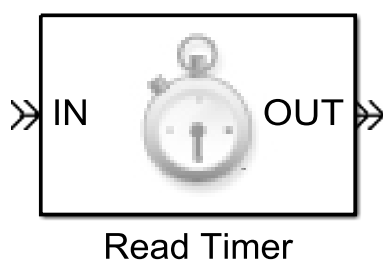


Рис. 6.12. Блок чтения значения таймера

Блок «*Schedule Timeout*» задает старт времени терпения и само время терпения каждой проходящей заявки (рис. 6.13). Обязательно наличие ранее блока «*Start Timer*». Затем заявка может покинуть следующие блоки системы через порты «*Timed-Out*» или дойти до блока «*Cancel Timeout*», который отменяет ограниченное время терпения заявки. В настройках блока задается имя таймера «*Timeout tag*», к которому будет применена операция, и источник определения длительности терпения заявки (диалог, внешний порт, атрибут). Предпочтительнее определение через атрибут, но могут использоваться и другие варианты.

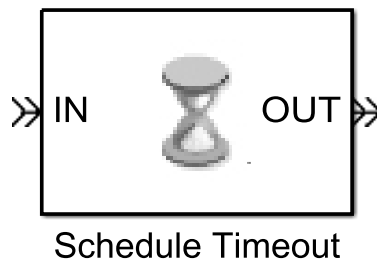


Рис. 6.13. Блок установки времени терпения заявки

Блок «*Cancel Timeout*» отменяет ограниченное время терпения заявки, завершая действие блока «*Schedule Timeout*», с которым используется в паре (рис. 6.14).

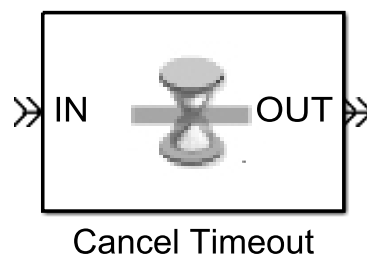


Рис. 6.14. Блок остановки таймера

Остальные блоки библиотеки SimEvents не являются необходимыми в данном практикуме, поэтому не рассматриваются.

## 6.2. Примеры моделирования СМО

Попытка создать первую версию модели одноканальной СМО с отказами, скорее всего, привела бы к модели, состоящей из генератора заявок, напрямую подключенного к одиночному серверу. Заявки генерируются с экспоненциальным распределением длительности пауз со средним значением  $t_{\text{паузы}}$ , а сервер обслуживает заявки в течение времени, также подчиненного экспоненциальному закону со средним временем обслуживания  $t_{\text{обс}}$ . Однако наблюдение за сигналом фактической средней длительности пауз между заявками и коэффициентом занятости сервера, определяемым согласно формуле , показало бы, что они оказываются завышенными и не соответствуют теоретически предполагаемым результатам.

Причина такого поведения модели кроется в принципе работы генератора заявок. Когда сервер занят обслуживанием предыдущей заявки, а генератор пытается послать новую заявку, происходит не отказ в обслуживании, как предполагается в теории, а приостановка генератора. Когда сервер освободится, он разблокирует генератор, и тот отправит задержанную заявку. Другими словами, налицо наличие памяти (последствия) в системе. Генерирование заявки зависит от занятости сервера, чего не должно быть.

Для того чтобы преодолеть эту проблему и создать СМО, функционирующую согласно описанному в третьей главе принципу, потребуется добавить в модель коммутатор заявок и приемник отклоненных заявок. В итоге получится корректная версия одноканальной СМО с отказами, приведенная на рис. 6.15.

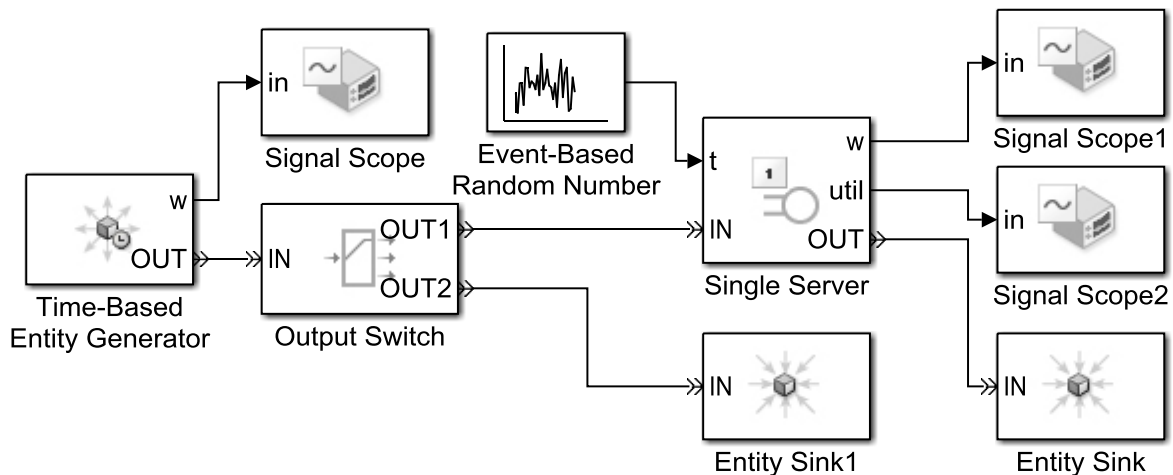


Рис. 6.15. Модель одноканальной СМО с отказами

В данном случае блокировки источника заявок не происходит, и при условии, что сервер занят, заявка уходит на второй порт коммутатора. Фактическая и теоретическая длительность паузы между заявками при продолжительном наблюдении за системой совпадают, а количество задержанных заявок в генераторе равно нулю все время моделирования.

Рассмотрим альтернативную версию модели одноканальной СМО с отказами и использованием атрибутов заявки для задания длительности обслуживания (рис. 6.16). В данном случае время обслуживания привязано к каждой заявке в виде ее атрибута и сервер берет это время именно из чтения атрибута. До-

полнительно отображена возможность подсчета числа сгенерированных, отклоненных и обслуженных заявок, что позволяет рассчитать экспериментальную вероятность отказа системы.

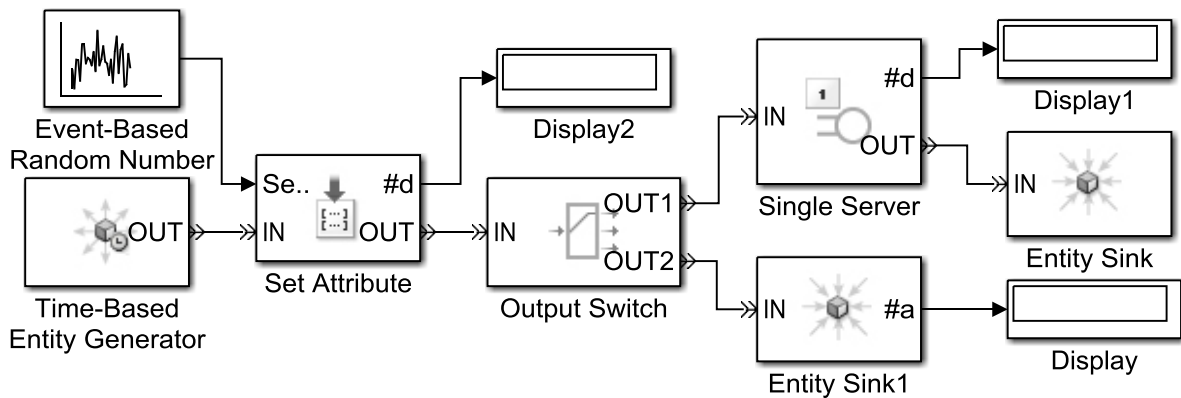


Рис. 6.16. Пример использования атрибутов заявки

На рис. 6.17 представлена модель многоканальной СМО с бесконечной очередью. Здесь задействованы таймеры, позволяющие определить среднее время пребывания заявок в системе. Дополнительно подключен выход из очереди в связи с завершением времени терпения, который в данном случае будет содержать всегда нулевой сигнал, т. к. все заявки имеют неограниченное терпение.

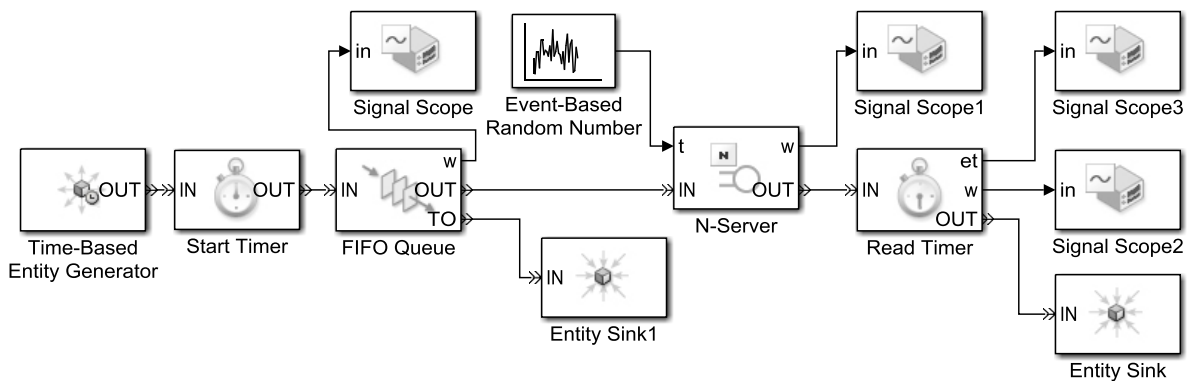


Рис. 6.17. Модель многоканальной СМО с очередью

В случае ограниченного времени терпения заявок модель преобразуется к виду, показанному на рис. 6.18. Блок «Schedule Timeout» задает время терпения каждой заявки, распределенное по экспоненциальному закону, а блок «Cancel Timeout» завершает отсчет времени терпения, т. к. считается, что в процессе обслуживания заявки имеют бесконечное терпение и пройдут его до конца.

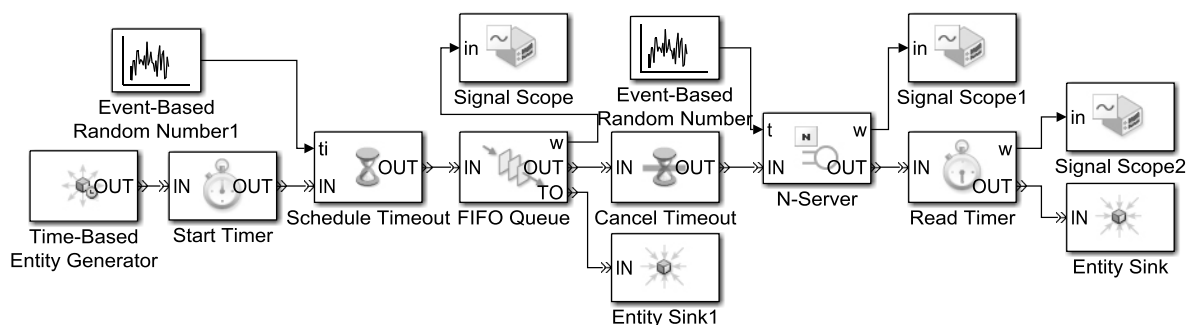


Рис. 6.18. Модель с «нетерпеливыми» заявками

Дополнительно следует отметить, что справка к библиотеке SimEvents содержит большое число примеров с готовыми моделями, что дополнительно может помочь при создании собственных моделей.

### 6.3. Практические задания

1. Исследуйте одноканальную систему с отказами  $M/M/1:0$ .
2. Исследуйте многоканальную систему с отказами  $M/M/n:0$ .
3. Исследуйте одноканальную систему с неограниченной очередью  $M/M/1$ .
4. Исследуйте многоканальную систему с неограниченной очередью  $M/M/n$ .
5. Исследуйте многоканальную систему с ограниченной очередью  $M/M/n:m$ .
6. Исследуйте многоканальную систему с неограниченной очередью  $M/M/n$  и ограниченным временем ожидания.
7. Проведите экспериментальное сравнение СМО из главы 5.



## Контрольные вопросы

1. Приведите распределение Пуассона и его основные характеристики: математическое ожидание, дисперсию, моду. Объясните связь этих характеристик с параметрами потока вызовов.

2. Приведите экспоненциальное распределение непрерывной случайной величины и его основные характеристики: математическое ожидание, дисперсию, моду. Объясните связь этих характеристик с параметрами потока вызовов.

3. Как связаны экспоненциальное распределение и распределение Пуассона?

4. Какова методика моделирования простейшего потока вызовов? Каким образом можно сгенерировать данные по экспоненциальному распределению?

5. Дайте определение системы массового обслуживания. Назовите основные виды СМО.

6. Способы задания потоков вызовов. Свойства потоков.

7. Марковские процессы в непрерывном времени. Уравнения Колмогорова.

8. Расчет предельных вероятностей в марковских процессах. Каков смысл предельных вероятностей?

9. Схема «гибели-размножения» на примере  $M/M/1:0$ .

10. Система  $M/M/1:\infty$ . Расчет основных характеристик.

11. Система  $M/M/n:\infty$ . Расчет основных характеристик.

12. Система  $M/M/n:m$ . Расчет основных характеристик.

13. Система  $M/M/n:0$ . Расчет основных характеристик.

14. Отличие системы  $M/M/n:\infty$  с «терпеливыми» и «нетерпеливыми» заявками.

15. В чем особенности систем массового обслуживания с ограниченным числом источников заявок?

16. Формула Литтла, С-формула Эрланга, В-формула Эрланга, формула Энгсета. Для каких систем применимы? Что выражают?

17. Сравните эффективность работы двух параллельных серверов и одного с удвоенной производительностью (случай системы с отказами).

18. Сравните эффективность работы двух параллельных серверов и одного с удвоенной производительностью (случай системы с очередью).

19. Сравните эффективность работы системы с общей очередью к серверам и с отдельной очередью к каждому серверу (на примере двух серверов).

### **Список рекомендуемой литературы**

1. Крылов, В. В. Теория телетрафика и ее приложения / В. В. Крылов, С. С. Самохвалова. — СПб. : BHV-Санкт-Петербург, 2005.

2. Авсиевич, А. В. Теория массового обслуживания : Потoki требований, системы массового обслуживания / А. В. Авсиевич, Е. Н. Авсиевич. — Самара : СамГАПС, 2004.

3. Авсиевич, А. В. Методические указания к выполнению лабораторного практикума по дисциплине «Теория массового обслуживания» / А. В. Авсиевич. — Самара : СамГАПС, 2004.

4. Кузнецова, В. А. Введение в теорию массового обслуживания : текст лекций / В. А. Кузнецова Е. В. Никулина. — Ярославль : ЯрГУ, 2005.

5. Бочаров, П. П. Теория массового обслуживания : учебник для вузов / П. П. Бочаров А. В. Печинкин. — М. : РУДН, 1995.

6. Кремер, Н. Ш. Теория вероятностей и математическая статистика : учебник для вузов / Н. Ш. Кремер. — М. : ЮНИТИ, 2012.

7. Файзутдинов, Р. Н. Математическое моделирование сложных систем : лабораторный практикум / Р. Н. Файзутдинов. — Казань : КАИ, 2013.

8. ITU-D. Teletraffic Engineering Handbook. — Geneva, 2003. — URL : <http://www.itu.int>.

## Оглавление

Введение.....	3
1. Потоки вызовов .....	4
1.1. Способы задания потоков вызовов и их свойства.....	4
1.2. Основные распределения .....	5
1.3. Генерирование потоков с заданным распределением ...	5
1.4. Практические задания .....	7
2. Марковские процессы.....	9
2.1. Определение марковского процесса.	
Расчет предельных вероятностей .....	9
2.2. Схема «гибели-размножения» .....	10
2.3. Практические задания .....	11
3. Системы массового обслуживания с отказами.....	13
3.1.Одноканальная СМО с отказами.....	13
3.2. Многоканальная СМО с отказами .....	14
3.3. Многоканальная СМО с отказами и ограниченным числом источников .....	16
3.4. Практические задания .....	17
4. Системы массового обслуживания с очередями .....	19
4.1. Одноканальная СМО с ограниченной очередью.....	19
4.2. Одноканальная СМО с неограниченной очередью .....	21
4.3. Многоканальная СМО с очередью.....	22
4.4. СМО с «нетерпеливыми» заявками .....	24
4.5. Практические задания .....	26
5. Сравнение различных систем массового обслуживания	27
6. Моделирование СМО в среде Simulink .....	29
6.1. Описание библиотеки SimEvents .....	29
6.2. Примеры моделирования СМО .....	37
6.3. Практические задания .....	40
Контрольные вопросы .....	41
Список рекомендуемой литературы .....	42

Учебное издание

**Дубов Михаил Андреевич**

**ТЕОРИЯ СИСТЕМ  
МАССОВОГО ОБСЛУЖИВАНИЯ**

Практикум

Редактор, корректор М. Э. Левакова  
Верстка М. Э. Леваковой

Подписано в печать 14.10.2015. Формат 60×84<sup>1</sup>/<sub>16</sub>.  
Усл. печ. л. 2,56. Уч.-изд. л. 2,0.  
Тираж 40 экз. Заказ

Оригинал-макет подготовлен  
в редакционно-издательском отделе ЯрГУ.

Ярославский государственный университет  
им. П. Г. Демидова.  
150000, Ярославль, ул. Советская, 14.