ФЕДЕРАЛЬНОЕ АГЕНТСТВО СВЯЗИ

ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«САНКТ-ПЕТЕРБУРГСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ТЕЛЕКОММУНИКАЦИЙ
им. проф. М. А. БОНЧ-БРУЕВИЧА»

А. Н. Соколов, Н. А. Соколов

ОДНОЛИНЕЙНЫЕ СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ

Учебное пособие

СПб ГУТ)))

Санкт-Петербург
2010

УДК [621.395:519.2](075.8)
ББК З882я73
С 59

Рецензенты:
доктор технических наук, профессор, зав. кафедрой
автоматической электросвязи СибГУТИ *В. В. Лебедянцев*,
доктор технических наук, профессор, зав. кафедрой
систем телекоммуникаций РУДН *К. Е.Самуйлов*

Рекомендовано к печати редакционно-издательским советом СПбГУТ в качестве учебного пособия

Соколов, А. Н.
С 59 Однолинейные системы массового обслуживания : учебное пособие / А. Н. Соколов, Н. А. Соколов. – СПб. : Изд-во «Теледом» ГОУВПО СПбГУТ, 2010. – 112 с.

Приводится теоретический материал, а также контрольные вопросы по дисциплине «Теория телетрафика». Дается анализ однолинейных систем массового обслуживания, которые служат адекватными математическими моделями для исследования процессов функционирования сетей следующего поколения и их основных элементов.

Предназначено для студентов, обучающихся по специальностям 210406 «Сети связи и системы коммутации», 210404 «Многоканальные телекоммуникационные системы», 210402 «Средства связи с подвижными объектами», 210407 «Эксплуатация средств связи».

УДК [621.395:519.2](075.8)
ББК З882я73

Содержание

Перечень основных используемых сокращений

ИПС – интерфейс пользователь–сеть
МСЭ – Международный союз электросвязи
СМО – система массового обслуживания
СеМО – сеть массового обслуживания
ССП – сеть следующего поколения
ТА – телефонный аппарат
ТФОП – телефонная сеть общего пользования
ФР – функция распределения
ЧНН – час наибольшей нагрузки

ETSI – European Telecommunications Standards Institute (Европейский институт телекоммуникационных стандартов)
IP[1] – Internet protocol (интернет-протокол)
FCFS – First come, first served (первым пришел – первым обслужен)
FIFO – First In, First Out (первым пришел – первым обслужен)
LCFS – Last come, first served (последним пришел – первым обслужен)
LIFO – Last In, First Out (последним пришел – первым обслужен)
NGN – Next Generation Network (сеть следующего поколения)
QoS – Quality of Service (качество обслуживания)

[1] Аббревиатура «IP» часто используется в более общем значении. Она указывает на сети, технологии и услуги, для реализации которых применяются пакетные способы коммутации, передачи и обработки информации.

ВВЕДЕНИЕ

Название учебного курса – «Теория телетрафика» – требует пояснений. По этой причине целесообразно начать «Введение» с комментариев к двум словам – «*теория*» и «*телетрафик*».

Слово «теория» пришло к нам из греческого языка. В переводе «theoria» означает наблюдение, рассмотрение, исследование, умозрение. В современных толковых словарях под теорией обычно понимается *высшая форма организации научного знания*, которая дает целостное представление о закономерностях и существенных связях в определенной предметной области. Теория основана на совокупности *базовых законов* и использует специально разработанный *понятийный аппарат*. Первый раздел этого учебного пособия посвящен основам «Теории телетрафика» и используемым терминам.

Слово «телетрафик» состоит из двух частей. Часть ряда сложных слов «*tele*» также заимствована из греческого языка. Буквальный перевод – вдаль, далеко. В электросвязи мы часто встречаемся со словами, которые начинаются с этих четырех букв: телеграфия, телефония, телевидение. Слово «*трафик*» переводится с английского языка по-разному: движение, транспорт, а применительно к электросвязи – нагрузка или обмен данными. Термин «*нагрузка*» часто использовался в отечественной технической литературе, но в последние годы кальки с английского языка – трафик и телетрафик – доминируют в монографиях, статьях и докладах.

Теория телетрафика возникла более ста лет назад. Она стимулировала развитие теории массового обслуживания, которая в свою очередь считается одним из разделов дисциплины «Исследование операций». Модели, рассматриваемые в теории телетрафика, обычно называются системами массового обслуживания.

Основоположником теории телетрафика по праву считается датский математик Агнер Краруп Эрланг. В 1909 г. им был опубликован фундаментальный труд «The Theory of Probabilities and Telephone Conversations» – теория вероятностей и телефонные разговоры. Он около 20 лет проработал в Копенгагенской телефонной компании, выполнил ряд исследовательских работ, имеющих важное теоретическое и практическое значение. В 40-х гг. XX в. название «*Эрланг*» присвоено единице измерения интенсивности трафика.

В данном учебном пособии рассматриваются системы с одним обслуживающим прибором. Они обычно называются *однолинейными*. Такое название связано с историей развития телефонии. В качестве первых обслуживающих приборов, попавших в сферу интересов специалистов

по теории телетрафика, были соединительные линии между телефонными станциями. Теперь системы с одним обслуживающим прибором используются в качестве моделей для расчетов ряда элементов современных телекоммуникационных сетей. Характерными примерами таких элементов можно считать тракт, по которому передаются IP-пакеты, и маршрутизатор, выполняющий функции распределения информации в телекоммуникационной сети.

Учебное пособие состоит из шести разделов. В конце каждого раздела содержатся список использованной в нем литературы, а также ряд контрольных вопросов и дополнительных заданий. Комментарии к ним помещены после раздела «Заключение». В лекциях по «теории телетрафика», как правило, используются материалы первых четырех разделов. Пятый и шестой разделы будут полезны желающим глубже изучить теорию телетрафика.

Содержание учебного пособия обсуждалось сотрудниками кафедр «Сети связи» и «Системы коммутации и распределения информации» СПбГУТ – Санкт-Петербургского государственного университета телекоммуникаций им. проф. М.А. Бонч-Бруевича. Ряд высказанных замечаний и предложений помог улучшить текст учебного пособия. Авторы благодарны коллегам за полезные советы. Существенный вклад в улучшение учебного пособия внесли рецензенты.

Замечания читателей будут использованы в дальнейшей работе. Все письма, которые касаются этого учебного пособия, направляйте, пожалуйста, по адресу: sokolov@niits.ru.

1. ОСНОВЫ ТЕОРИИ ТЕЛЕТРАФИКА

1.1. Модель системы массового обслуживания

Рассмотрим модель, называемую в ряде дисциплин «*черным ящиком*»[2] – рис. 1.1. Далее такой «черный ящик» будем называть системой массового обслуживания (СМО). В качестве синонима термина «СМО» в технической литературе по электросвязи встречается также словосочетание «система телетрафика». Наряду с бревиатурой СМО будем использовать слово «система».

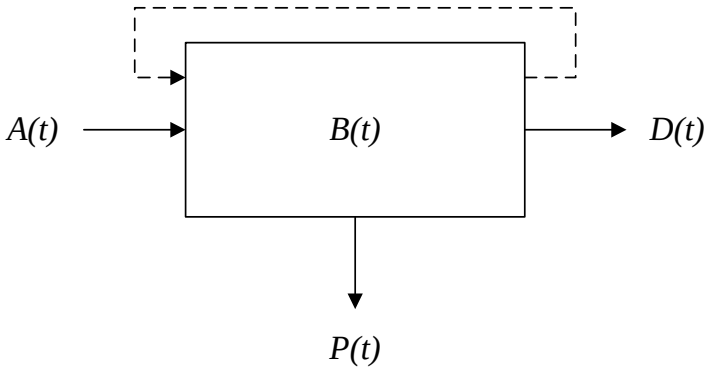


Рис. 1.1. Модель системы массового обслуживания

Процесс $A(t)$ на входе СМО связан с одним из самых важных понятий в теории телетрафика – потоком заявок. Вместо слова «заявка» иногда используют термины «требование» и «событие». Под заявкой понимается то, что должно быть обслужено. Например, вызов в сети телефонной связи для установления соединения или IP-пакет для передачи между терминалами.

Слово «*обслужено*» следует рассматривать как универсальное обозначение неких операций, которые могут заключаться в установлении соединения в телефонной сети, обработке информации или же в ином действии. В любом случае предполагается, что обслуживание может быть описано процессом $B(t)$.

В некоторых случаях заявка не может быть обслужена. Тогда она покидает СМО. Процесс $P(t)$ отображает события такого рода. Кроме того, процесс $P(t)$ может быть использован для учета влияния дисциплин обслуживания заявок в системе.

Успешно обслуженные заявки формируют поток, представимый процессом $D(t)$. Этот поток принято называть выходящим.

Пунктирной линией показана своего рода петля обратной связи. В некоторых СМО часть обслуженных заявок по каким-либо причинам

[2] Понятие «черный ящик» (black box) было введено для упрощения исследования сложных систем. Представление сложной системы в виде «черного ящика» не требует знания принципов ее работы. Как правило, достаточно изучить процессы на входе и на выходе системы.

может снова поступать на вход системы. Такие модели в учебном пособии не рассматриваются.

1.2. Потоки заявок

Поток заявок целесообразно рассматривать как последовательность, определяемую на оси «Время». На этой оси можно выделить моменты времени x_i, в которые заявки поступают на вход СМО. Значения x_i соответствуют, например, тем моментам времени, когда абоненты телефонной станции снимают трубку телефонного аппарата, намереваясь позвонить. Очевидно, что значения x_i следует рассматривать как случайные величины. Поток заявок также называется случайным. Существуют и детерминированные потоки заявок. Для них значения x_i заданы неким расписанием; следовательно, анализ потока заявок не связан с изучением случайных величин.

В данном учебном пособии рассматриваются случайные потоки заявок. Тем не менее в качестве первого примера целесообразно привести процесс поступления заявок в том случае, когда поток можно считать детерминированным. На рис. 1.2 показаны пять моментов времени x_i $(i = \overline{0, 4})$, в которые на вход СМО поступают заявки.

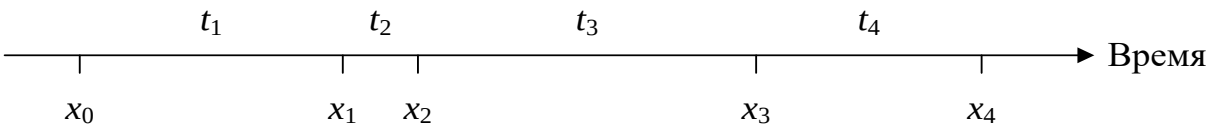


Рис. 1.2. Первая модель потока заявок

Длительность i-го промежутка времени между моментами поступления соседних заявок составляет t_i. Очевидно, что $t_i = x_i - x_{i-1}$ для всех $i \geq 1$. Модель, изображенная на рис. 1.2, содержит всю необходимую информацию о входящем потоке заявок лишь при одном условии. Оно заключается в том, что в любой момент времени x_i может поступить только одна заявка. В противном случае необходимо задать величины k_i $(i = \overline{0, 4})$, которые определяют количество заявок, поступающих в момент времени x_i. Для рассматриваемой модели $k_i \equiv 1$. Предположим, что в моменты времени x_1 и x_4 на вход СМО поступают две заявки сразу. Тогда лучше использовать другую модель потока заявок. Она приведена на рис. 1.3. В ней добавлена ось «Количество заявок», которая позволяет указать все значения k_i.

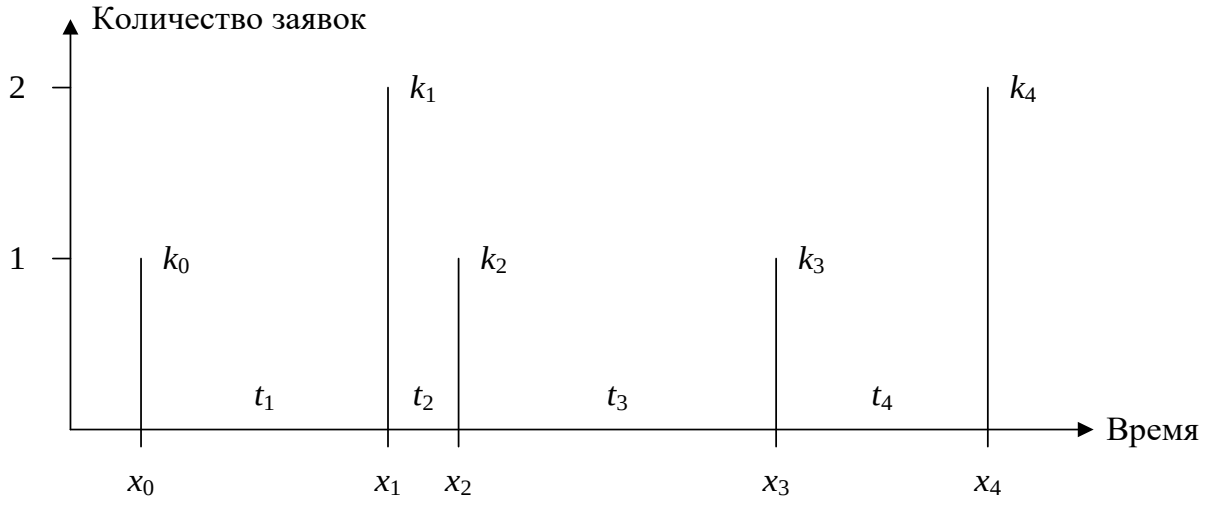


Рис. 1.3. Вторая модель потока заявок

Модели, приведенные на рис. 1.2 и 1.3, позволяют описать так называемые *однородные* потоки заявок. Слово «однородный» в данном случае используется для того, чтобы подчеркнуть факт общности заявок с точки зрения теории телетрафика. Это значит, что моменты x_i или величины t_i позволяют уяснить характер потока заявок. Подобная трактовка определения «однородный» напрямую не связана с природой заявок. Здесь будут рассматриваться только те виды СМО, на вход которых поступают однородные потоки заявок. Более того, исследуются модели телетрафика с *финитными потоками* заявок. Для финитных потоков всегда существует математическое ожидание количества заявок, поступающих на конечном интервале времени.

Математическое ожидание (среднее значение) количества заявок, поступающих на интервале времени $[0,t)$, принято называть ведущей функцией потока – $\Lambda(0,t)$. Данная функция, по определению, не может быть отрицательной и убывающей. Потоки заявок с монотонно возрастающей функцией $\Lambda(0,t)$ именуются *регулярными*. Для ряда моделей телетрафика кривую $\Lambda(0,t)$ целесообразно представлять в виде ступенчатой функции. Такие потоки заявок называют *сингулярными*.

Вернемся к модели, изображенной на рис. 1.1. Очевидно, что процесс $A(t)$ на входе СМО целесообразно рассматривать как функцию распределения (ФР) случайной величины. Далее для обозначения случайных величин используются прописные буквы латинского алфавита. Возможным значениям случайных величин соответствуют строчные буквы латинского алфавита. Тогда процесс $A(t)$ представляет собой ФР, определяемую следующим образом:

$$A(t) = P\{T \le t\}. \qquad (1.1)$$

Иногда используется дополнительная ФР: $P\{T > t\}$. Очевидно, что для вычисления этой функции используется соотношение

$$P\{T > t\} = 1 - P\{T \le t\}. \qquad (1.2)$$

Следует отметить, что в ряде публикаций ФР определяется строгим неравенством: $P\{T < t\}$. Различие в этих определениях существенно лишь для дискретных случайных величин.

Большинство моделей телетрафика основано на том, что отрезки времени t_i можно считать независимыми, одинаково распределенными случайными величинами. Тогда ФР (1.1) содержит полную информацию о потоке заявок, поступающих на вход СМО. Если значения t_i нельзя считать независимыми и одинаково распределенными случайными величинами, то задается совместный закон распределения n случайных величин:

$$P\left\{T_i \le t_i, i = \overline{1, n}\right\} = P\left\{T_1 \le t_1, T_2 \le t_2, ..., T_n \le t_n\right\}. \qquad (1.3)$$

В данном учебном пособии рассматриваются модели телетрафика, для которых входящий поток заявок представим ФР вида (1.1). Более того, для всех анализируемых СМО предполагается, что входящим потокам заявок присущи три важных свойства:

- *стационарность;*
- *ординарность;*
- *отсутствие последействия.*

Для стационарного потока вероятность поступления k заявок за некий промежуток времени от точки a до точки b зависит только от величины $(b-a)$. Эта вероятность инвариантна к значениям a и b на оси «Время».

Важная характеристика потока заявок – вероятность поступления хотя бы k вызовов на отрезке времени (a, b) – $\varphi_k(a,b)$. Эта вероятность позволяет сформулировать условие ординарности потока заявок. Пусть $\tau = b - a$. Тогда поток заявок будет ординарным, если при $\tau \to 0$ справедливо условие

$$\lim_{\tau \to 0} \frac{\varphi_2\left(a, a+\tau\right)}{\tau} = 0. \qquad (1.4)$$

С практической точки зрения свойство ординарности означает, что в любой момент времени на вход СМО не может поступить две (или более) заявки. Для большинства потоков заявок, исследуемых в теории те-

летрафика, допустима гипотеза об ординарности. Правда, ряд СМО не может быть представлен моделями с потоками заявок, для которых свойственна ординарность. Пачка телеграмм, принесенная в почтовое отделение, служит типичным примером модели с неординарным потоком заявок.

Допустим, что мы рассматриваем поток заявок после какого-то момента времени t_0. Если его характеристики не зависят от поведения потока для $t < t_0$, то можно говорить об отсутствии последействия.

Будем рассматривать в основном стационарные ординарные потоки заявок без последействия. Для них вводится только одна характеристика потока заявок – интенсивность. Обозначим ее греческой буквой λ, как принято в большинстве последних публикаций по теории телетрафика. Для потоков, которые не отвечают перечисленным выше свойствам, необходимо ввести параметр потока $\kappa(t)$. Он определяется как предел отношения вероятности поступления хотя бы одной заявки за период $[t, t+\tau)$ к длине этого отрезка времени при $\tau \to 0$.

Вернемся к рис. 1.2. Статистическая информация о величинах t_i позволяет определить ФР, обозначенную как $A(t)$. Предположим, что для этой функции существует преобразование Лапласа–Стилтьеса $\alpha(s)$. Тогда математическое ожидание длительности интервала между поступлениями соседних заявок $A^{(1)}$ определяется по одной из двух следующих формул:

$$A^{(1)} = \int_0^\infty t dA(t), \quad A^{(1)} = -\frac{d\alpha(s)}{ds}\Big|_{s=0}. \tag{1.5}$$

Величины λ и $A^{(1)}$ связаны между собой простым соотношением

$$\lambda = \frac{1}{A^{(1)}}. \qquad 1.6)$$

В теории телетрафика часто используется предположение, что входящий поток заявок может быть представлен при помощи пуассоновского закона распределения. Этот закон определяет вероятность поступления ровно k заявок p_k за промежуток времени длительностью t:

$$p_k = \frac{(\lambda t)^k}{k!} e^{-\lambda t}. \tag{1.7}$$

На рис. 1.4 приведены две гистограммы, иллюстрирующие изменение значений p_k для разных величин интенсивности входящего потока заявок. При построении обоих графиков принято, что $t = 1$.

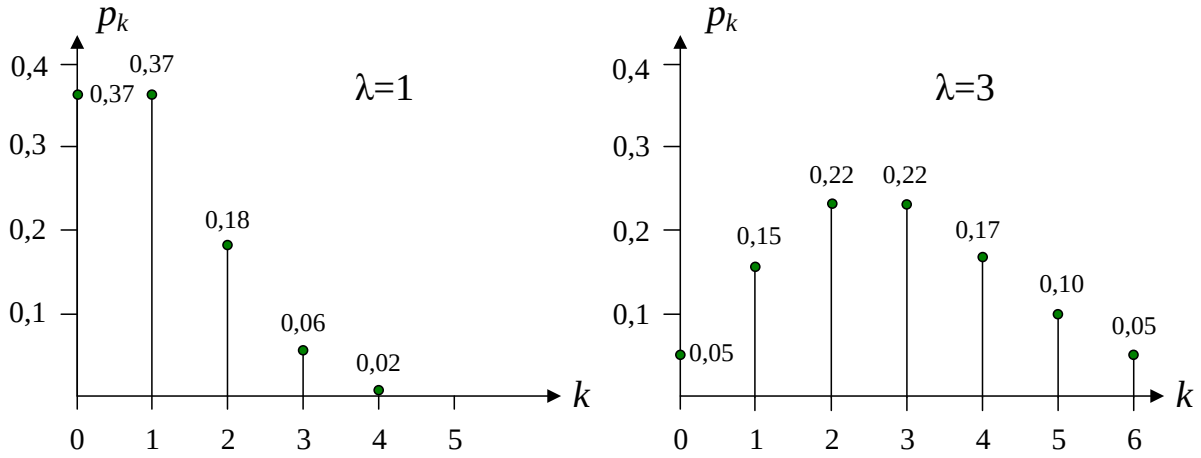


Рис. 1.4. Два примера пуассоновского закона распределения

Несложно показать, что для пуассоновского потока заявок распределение $A(t)$ подчиняется экспоненциальному закону:

$$A(t) = 1 - e^{-\lambda t}. \tag{1.8}$$

Для экспоненциального закона распределения известны величины дисперсии σ^2, коэффициента вариации C и p-квантиля t_p:

$$\sigma^2 = \frac{1}{\lambda^2}, \quad C = 1, \quad t_p = -\frac{1}{\lambda} ln(1 - p). \tag{1.9}$$

Три перечисленные характеристики случайных величин будут часто встречаться в следующих разделах. Это объясняется принципами нормирования показателей качества обслуживания, которые приняты международным сообществом.

В учебном пособии, помимо пуассоновского, акцентируется внимание еще на двух видах входящего потока заявок. Во-первых, анализируется СМО, на вход которой может поступать поток заявок с произвольным распределением $A(t)$. Правда, подробный анализ возможен только для узкого класса моделей. Во-вторых, исследуется СМО, для которой входящий поток заявок может быть задан некой гистограммой, что позволяет представить распределение $A(t)$ ступенчатой функцией – рис. 1.5. Предполагается, что каждый момент времени, когда не исключено поступление заявки, можно представить в виде произведения $i\tau$ $\left(i = \overline{0, N}\right)$. Ступенчатую функцию $A(t)$ проще представить при помощи ее преобразования Лапласа–Стилтьеса:

$$\alpha(s) = \sum_{i=0}^{N} P_i e^{-i\tau s}. \qquad (1.10)$$

В некоторых точках по оси «Время» функция $A(t)$ не имеет приращений. Например, для распределения $A(t)$, показанного на рис. 1.5, $P_2 = 0$.

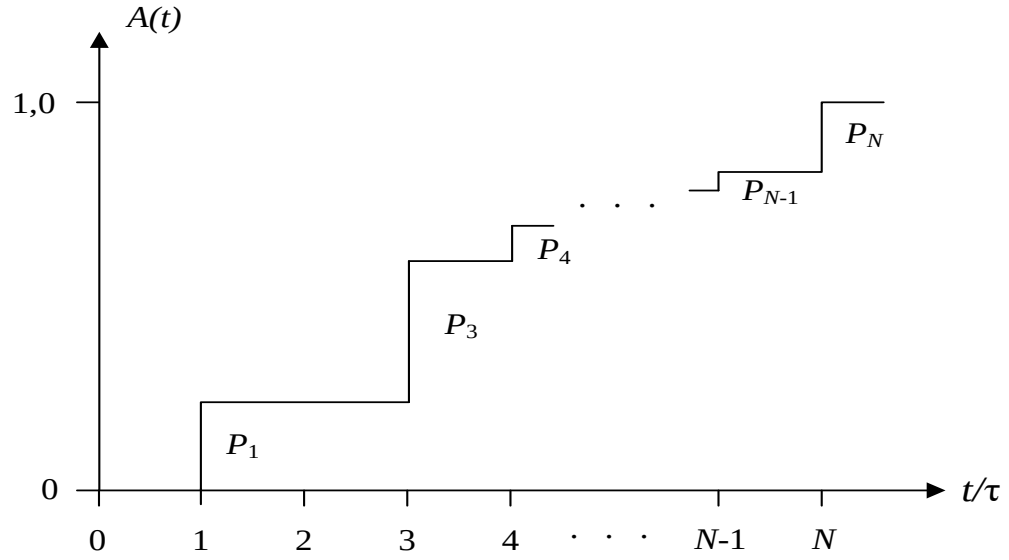


Рис. 1.5. Пример ступенчатой функции

Практическая ценность применения ступенчатых функций объясняется тем, что зачастую информация о распределении $A(t)$ извлекается из результатов измерений потока заявок, которые проводятся в эксплуатируемых телекоммуникационных сетях. Подобная информация обычно представляется в виде гистограммы, позволяющей с весьма высокой точностью оценить приращения P_i.

Информация о количестве заявок, поступающих в СМО, также формируется в результате измерений. На рис. 1.6 показано изменение количества вызовов (в данном примере они играют роль заявок), обслуживаемых телефонной станцией, в течение суток. Количество поступающих вызовов оценивалось за 1 мин. Усреднение измеряемой величины проводилось по 15-минутным интервалам.

Гистограмма, показанная на рис. 1.6, была получена в результате обработки статистических данных, которые собирались в течение 10 дней. Эти 10 дней соответствовали двум рабочим неделям. Конечно, такая выборка не позволяет судить об изменении количества поступающих вызовов в течение квартала или года. Более того, в некоторых случаях полезно выделить тренды, описывающие изменения исследуемого

процесса в течение нескольких лет. Тем не менее данные, подобные тем, что приведены на рис. 1.6, представляют большой практический интерес.

Характер изменения количества вызовов в течение суток говорит о том, что данный поток не обладает свойством стационарности. Однако для ряда задач, решаемых методами теории телетрафика, выбирается *час наибольшей нагрузки* (ЧНН). Для этого периода времени предположение о стационарности потока заявок считается приемлемым.

Источник: ITU-D. Teletraffic Engineering Handbook (edited by V.B. Iversen). – Geneva, 2003.

Рис. 1.6. Изменение количества вызовов за сутки

Для более подробного ознакомления с потоками заявок и их характеристиками целесообразно использовать монографии [1–5]. Будут полезны и книги по математике, в которых изложены важные положения по теории вероятностей [6, 7]. Некоторые сведения можно найти и в справочниках по математике [8, 9].

1.3. Длительность обслуживания заявок

Длительность обслуживания заявок в редких случаях может считаться постоянной величиной. Чаще всего длительность обслуживания заявок следует рассматривать как случайную величину. По этой причине процесс $B(t)$, который был показан на рис. 1.1, целесообразно интерпретировать как ФР длительности обслуживания заявок.

Преимущественно здесь будет рассматрен класс функций $B(t)$, для которых существует преобразование Лапласа–Стилтьеса $\beta(s)$.

Рассмотрим статистические данные о средней длительности телефонного разговора. На рис. 1.7 приведены соответствующие значения за сутки. Как и для предыдущей иллюстрации, усреднение измеряемой величины проводилось по 15-минутным интервалам.

Источник: ITU-D. Teletraffic Engineering Handbook (edited by V.B. Iversen). – Geneva, 2003.

Рис. 1.7. Изменение средней длительности телефонного разговора в сутки

Статистические данные, использованные для построения этой гистограммы, были собраны в 1973 г. Поэтому измеряемая величина связана с трафиком речи. Передача факсимильных сообщений в это время была большой редкостью, а трафик данных вообще отсутствовал.

Следует отметить, что длительность местных соединений при сборе статистических данных не учитывалась. Это означает, что рассматривался только междугородный и международный трафик речи. Необходимо также подчеркнуть, что измеряемая величина характерна для соединений, устанавливаемых со стационарных телефонных терминалов (в 1973 г. мобильной связи еще не было). Для сетей мобильной связи характерны иные распределения и численности вызовов и средней длительности разговоров.

Таким образом, рассматриваемый пример связан с заявками, роль которых играют вызовы, обслуживаемые телефонной сетью. Для других сетей связи закономерности, подобные тем, которые могут быть уста-

новлены в процессе анализа приведенной выше гистограммы, нельзя считать корректными.

График, показанный на рис. 1.7, иллюстрирует изменение среднего значения (математического ожидания) длительности обслуживания заявок. Очевидно, что процесс, наблюдаемый в течение суток, не может быть представлен функцией $B(t)$, для которой среднее значение длительности обслуживания заявок является постоянной величиной. Тем не менее в теории телетрафика обычно используются функции $B(t)$, для которых среднее значение $B^{(1)}$ не меняется. Такая гипотеза приемлема из-за целесообразности анализа поведения системы в течение ЧНН.

Среди тех распределений $B(t)$, которые представляют практический интерес, можно выделить семь законов. Ниже приводятся соответствующие распределения и средние значения длительности обслуживания заявок.

Во многих исследованиях, касающихся телефонного трафика, используется гипотеза об экспоненциальном распределении длительности обслуживания заявок:

$$B_1(t) = 1 - e^{-\mu t}, \quad B_1^{(1)} = \frac{1}{\mu}. \tag{1.11}$$

Экспоненциальное распределение длительности обслуживания заявок существенно упрощает исследование СМО. Кроме того, многие распределения $B(t)$, интересные с практической точки зрения, имеют коэффициент вариации менее единицы. Это означает, что результаты, полученные для экспоненциального распределения, позволяют оценить характеристики исследуемой СМО «сверху» – для пессимистического сценария. При этом для ряда моделей гипотезу об экспоненциальном распределении длительности обслуживания заявок следует считать очень грубым приближением. Данное утверждение справедливо и для некоторых законов распределения с модой в точке $t = 0$.

Распределения с коэффициентом вариации менее единицы часто описываются при помощи распределения Эрланга *k*-го порядка. Оно может рассматриваться как частный случай гамма-распределения. Напомним выражение для ФР и математического ожидания:

$$B_2(t) = 1 - e^{-\mu t} \sum_{i=0}^{k-1} \frac{(\mu t)^i}{i!}, \quad B_2^{(1)} = \frac{k}{\mu}. \tag{1.12}$$

При $k \to \infty$ распределение Эрланга вырождается. Этот случай рассматривается как самостоятельное распределение – постоянная длительность обслуживания, равная t_0. СМО с постоянным временем обслужи-

вания хорошо формализует процессы работы ряда устройств в сетях электросвязи, основанных на IP-технологии. Распределение времени обслуживания проще записать через преобразование Лапласа–Стилтьеса:

$$\beta_3(s) = e^{-t_0 s}, \quad B_3^{(1)} = t_0. \tag{1.13}$$

Распределения с коэффициентом вариации более единицы можно представить при помощи гиперэкспоненциального распределения второго порядка. Это распределение обычно представляют в такой форме:

$$B_4(t) = 1 - pe^{-2p\mu t} - (1-p)e^{-2(1-p)\mu t}, \quad B_4^{(1)} = \frac{1}{\mu}. \tag{1.14}$$

Распределение (1.14) задается с ограничением на величину p, которая называется параметром формы: $0 < p < 0{,}5$. Следующая функция $B(t)$, для которой коэффициент вариации меняется в широких пределах, – распределение Вейбулла–Гнеденко. Для него справедливы соотношения следующего вида:

$$B_5(t) = 1 - e^{-\left(\frac{t}{a}\right)^c}, \qquad B_5^{(1)} = a\Gamma\left(\frac{1}{c} + 1\right). \tag{1.15}$$

Величина a называется параметром масштаба, а переменная c – параметром формы. В этом распределении $a > 0$ и $c > 0$. В формулу для расчета коэффициента вариации входит гамма-функция. Следующее распределение – равномерное на отрезке времени $[a, b]$. Выражения для ФР и среднего значения длительности обслуживания заявок записываются в такой редакции:

$$B_6(t) = \frac{t-a}{b-a}, \quad B_6^{(1)} = \frac{a+b}{2}. \tag{1.16}$$

Последнее (седьмое) распределение относится к дискретным. Речь идет об уже упоминавшейся ступенчатой функции, которую удобно записывать через преобразование Лапласа–Стилтьеса. Пусть в точках $i\tau$ ($i = \overline{0, N}$) ФР имеет приращения P_i. Тогда искомые соотношения могут быть представлены следующим образом:

$$\beta_7(s) = \sum_{i=0}^{N} P_i\, e^{-i\tau s}, \quad B_7^{(1)} = \tau \sum_{i=0}^{N} iP_i\ . \tag{1.17}$$

Можно привести ряд других примеров. Подробный перечень распределений (не все используются в теории телетрафика) приведен, например, в [6]. В учебном пособии особое внимание будет уделено ступенчатым функциям. Дело в том, что их использование позволяет ре-

шить ряд практически важных задач. Правда, в отличие от экспоненциальной и некоторых других функций, для анализа СМО и их совокупности, которая образует сеть массового обслуживания (СеМО), необходимо использовать средства вычислительной техники. Все расчеты могут быть выполнены при помощи персонального компьютера, к которому не предъявляются особые требования в части производительности и/или объема оперативной памяти.

1.4. Дисциплины обслуживания заявок

Для ряда компонентов современных инфокоммуникационных систем предусмотрена возможность выбора дисциплины обслуживания заявок на этапе проектирования сети или в процессе ее эксплуатации. В некоторых случаях дисциплина (алгоритм) обслуживания заявок заранее определяется на основании международных или национальных стандартов. Иногда выбор дисциплины обслуживания невозможен. Подобная ситуация, как правило, свойственна техническим системам, в которых не используется программное обеспечение.

Дисциплины обслуживания заявок в СМО классифицируют разными способами. На рис. 1.8 приведен первый способ классификации. Он хорошо представляет алгоритмы, используемые в коммутационных станциях телефонной сети.

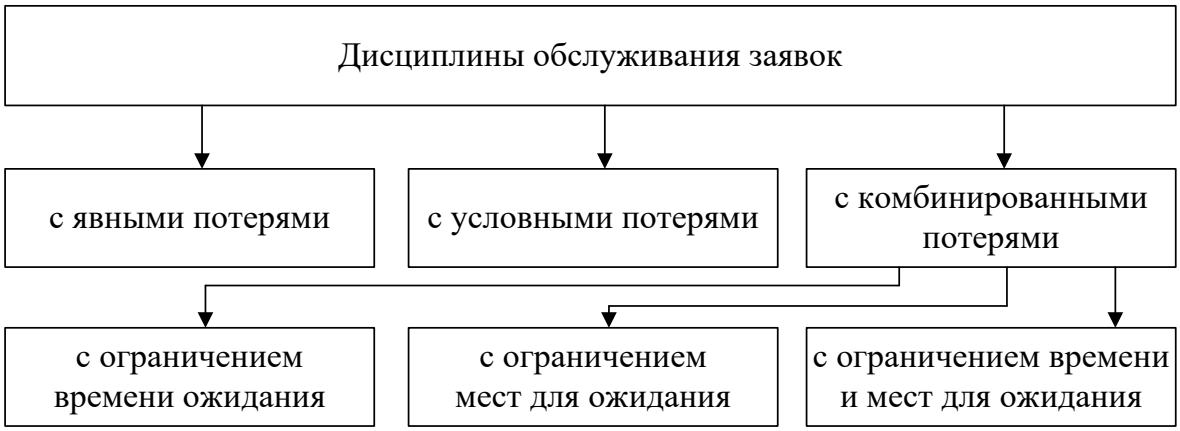


Рис. 1.8. Первый способ классификации дисциплин обслуживания заявок

Явные потери свойственны, в частности, алгоритмам функционирования декадно-шаговых телефонных станций. Если отсутствует свободный обслуживающий прибор, то вызов теряется. Абонент практически сразу же получает акустический сигнал «Занято». Дисциплина с явными потерями используется и в цифровых коммутационных станциях. В частности, при отсутствии свободных соединительных линий в требуе-

мом направлении (во всех возможных путях установления соединения) вызов также теряется.

Условные потери подразумевают, что при отсутствии свободного обслуживающего прибора заявка ожидает его освобождения. Обычно считается, что условные потери не приводят к отказу в обслуживании. С другой стороны, очевидно, что чрезмерное время ожидания может привести к тому, что абонент сам откажется от попытки вызова. Это значит, что условные потери – некая идеализация реальных процессов обслуживания заявок.

Комбинированные потери позволяют определить более реальные – с практической точки зрения – дисциплины обслуживания заявок. Три из них показаны в нижней части рис. 1.8. Обслуживание с ограниченным временем ожидания давно используется в телефонных станциях. Например, если вы слышите акустический сигнал «Ответ станции», но не набираете номер в течение некоторого интервала времени, то обслуживание будет прервано. Вы услышите акустический сигнал «Занято».

С дисциплиной, которая ограничивает число мест для ожидания, многие абоненты сталкиваются при попытке дозвониться до справочной службы местной телефонной сети. Если все места для ожидания заполнены, вежливый голос приносит вам свои извинения и просит повторить вызов позже. Некоторые справочные системы сочетают оба вида ограничений – по длительности ожидания и числу мест в очереди.

На рис. 1.9 приведен второй способ классификации. Он более подходит для алгоритмов, используемых в устройствах управления современных систем коммутации. В данном случае классификационным признаком (таксоном) служит тот тип приоритета, который используется для обслуживания заявки.

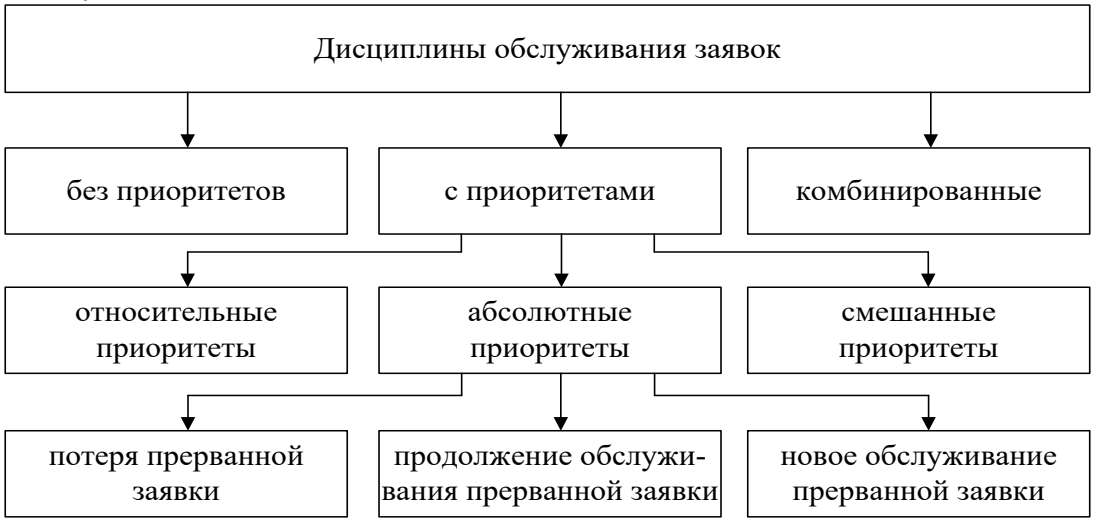


Рис. 1.9. Второй способ классификации дисциплин обслуживания заявок

Заявки могут обрабатываться без приоритетов. Дисциплина такого рода характерна для устройств управления электромеханических коммутационных станций. Приоритетные стратегии обслуживания заявок можно разделить на три группы, которые следует рассмотреть подробно. Заметим, что комбинированные дисциплины предусматривают переход к приоритетному обслуживанию при определенных условиях (например, резкий рост трафика, приводящий к снижению показателей качества обслуживания).

Для анализа приоритетных стратегий целесообразно ввести простую модель. Все заявки, поступающие в СМО, делятся на группы, которым присваивается приоритет от 1 до N. Заявка с приоритетом j имеет преимущество перед заявками, которым присвоены приоритеты под номерами от $j+1$ до N.

В СМО с относительными приоритетами обслуживание заявок не прерывается. Допустим, заявка с приоритетом j застала все обслуживающие приборы занятыми. Тогда она встает в очередь перед всеми заявками, имеющими более низкий приоритет. Среди заявок с приоритетом j она будет последней.

СМО с абсолютными приоритетами основаны на прерывании обслуживания заявок. Такая возможность предусмотрена для всех случаев, когда обслуживаются заявки более низкого приоритета. При этом (нижняя часть рис. 1.9) могут использоваться три основных варианта возобновления прерванного процесса обслуживания заявок.

В некоторых СМО используются смешанные приоритеты. Тогда множество $\{J\}$ разбивается на несколько классов: $\{J_1\},\{J_2\},...,\{J_L\}$. Чем меньше индекс у класса из множества $\{J\}$, тем выше абсолютный приоритет у обслуживаемых заявок. В пределах каждого класса $\{J_i\}$ заявкам могут назначаться относительные приоритеты.

Понятие «обслуживание заявок» включает также дисциплины их выбора из очереди. Классификация основных дисциплин выбора заявок на обслуживание приведена на рис. 1.10. Предлагаемая классификация очень проста. Она включает всего один уровень классификации используемых дисциплин.

Рис. 1.10. Классификация дисциплин выбора заявок на обслуживание

Случайный выбор заявок на обслуживание позволяет отказаться от каких-либо процедур формирования очереди. Этот алгоритм в инфокоммуникационных системах используется редко. Обслуживание в порядке очереди – классическая дисциплина выбора заявок на обслуживание. Он известен по англоязычным аббревиатурам *FIFO* (First In, First Out) и *FCFS* (First come, first served). Выбор заявки на обслуживание из конца очереди обычно используется в системах, подобных складам, но применяется также и в сетях связи. Эта дисциплина известна по аббревиатурам *LIFO* (Last In, First Out) и *LCFS* (Last come, first served).

Вернемся к рис. 1.8. Процесс $P(t)$ при обслуживании заявок с явными потерями целесообразно рассматривать как вероятностный. Обычно оценивается вероятность в течение ЧНН. По этой причине функцию $P(t)$ заменяют мерой p, представляющей собой вероятность отказа в обслуживании. Для однолинейных СМО оценка величины p – не сложная задача. При использовании дисциплин с условными, и особенно с комбинированными, потерями расчет значения p связан с решением весьма сложных задач. Кроме того, для этих двух дисциплин необходим расчет параметров времени ожидания и задержки (пребывания) заявок в СМО.

Дисциплины обслуживания заявок с приоритетами стимулируют введение сложных конструкций для формализации процесса $P(t)$. Аналогичная ситуация складывается с дисциплинами выбора заявок на обслуживание, классификация которых приведена на рис. 1.10. Способы описания процесса $P(t)$ рассматриваются в разд. 2.

1.5. Классификация систем массового обслуживания

Используемые ныне принципы классификации СМО базируются на предложениях Д. Кендалла, опубликованных в 1951 г. Для описания СМО используется запись следующего вида:

$$A/B/n. \tag{1.18}$$

Первый символ определяет характер входящего потока заявок. Распределение длительности обслуживания заявок идентифицируется вторым символом. Величина n указывает на количество обслуживающих приборов. В учебном пособии рассматриваются однолинейные СМО, т. е. $n \equiv 1$.

Для конкретизации характера входящего потока заявок и закона распределения длительности их обслуживания вводятся следующие символы:

• M – экспоненциальное распределение. Символ M в позиции A говорит о том, что входящий поток является пуассоновским. Если этот

символ стоит в позиции B, то в СМО длительность обслуживания заявок распределена по экспоненциальному закону;

• D – постоянная величина. Если этот символ поставлен в позиции A, то на вход СМО поступает поток заявок, который нельзя рассматривать как случайный процесс. Символ D в позиции B свидетельствует о том, что время обслуживания заявок постоянно;

• E_k – распределение Эрланга k-го порядка. Размещение этого символа в позиции A указывает на то, что распределение длительности интервалов между моментами поступления заявок в СМО подчиняется закону Эрланга k-го порядка. Если символ E_k находится в позиции B, то длительность обслуживания заявок распределена по этому же закону. Распределение Эрланга отличается важным свойством. Если $k=1$, то оно становится экспоненциальным, а при $k \to \infty$ – вырожденным. Тогда обозначение E_k следует заменить символами M и D соответственно;

• G – распределение общего вида (первая буква в слове «general»). Обычно этот символ указывается в позициях A или B, когда распределение для исследуемого процесса не может быть выражено известными законами. В ряде случаев в позиции A используется символ GI, чтобы подчеркнуть следующее: входящему потоку свойственно ограниченное последействие.

В тексте будут встречаться и другие обозначения в позициях A и B. В частности, обратившись к соотношениям (1.14)–(1.16), можно ввести следующие символы:

• H_2 – гиперэкспоненциальное распределение второго порядка (заменив цифру 2 буквой k, можно говорить об этом же законе, но k-го порядка);

• WG – распределение Вейбулла–Гнеденко;

• U – равномерное распределение на неком интервале $[a, b]$.

Классификация (1.18) была вполне приемлемой до появления систем с ожиданием, а также до использования сложных алгоритмов обслуживания заявок, приведенных на рис. 1.8–1.10. По мере развития телекоммуникационных сетей усложнились модели СМО, что потребовало дополнения классификации Кендалла. В отечественной литературе обычно используются дополнения, предложенные профессором Г. П. Башариным.

В модифицированной классификации Кендалла вводятся еще две позиции. Для систем с одним обслуживающим прибором можно использовать такое обозначение:

$$A/B/1/r/f_i^j. \qquad (1.19)$$

Символ r в четвертой позиции определяет количество мест для ожидания в очереди. Если $r = 0$, то места для ожидания отсутствуют. Следовательно, рассматривается СМО с явными потерями. Если в четвертой позиции поставлен символ ∞, то справедливо предположение о том, что заявки никогда не теряются. Целое число r $(0 < r < \infty)$ определяет конкретное значение количества мест для ожидания в очереди, которое имеется в исследуемой СМО.

Символ f_i^j в пятой позиции позволяет идентифицировать дисциплины постановки заявок в очередь и их выбора для обслуживания. Верхний индекс определяет дисциплину постановки заявок в очередь. При $j = 0$ данный процесс осуществляется без приоритета. Если $j = 2$, то из очереди «вытесняется» заявка, которая имеет более низкий приоритет. Нижний индекс характеризует дисциплину выбора заявок на обслуживание из очереди. Задействуются три значения i:

- $i = 0$ – обслуживание осуществляется без приоритетов;
- $i = 1$ – используются относительные приоритеты;
- $i = 2$ – применяется обслуживание с абсолютными приоритетами.

В некоторых зарубежных работах применяется иное дополнение классификации Кендалла. Вводятся три новые позиции. Для СМО с одним обслуживающим прибором типична такая запись:

$$A/B/1/K/N/X. \tag{1.20}$$

Символ K определяет количество мест для ожидания в очереди. Это означает, что он эквивалентен символу r, введенному в обозначении (1.19). Символ N указывает на общую численность обслуживаемых пользователей (например, терминалов, включенных в телефонную станцию). Символ X идентифицирует дисциплину обслуживания. Обычно он заменяется сокращением, которое связано с дисциплиной обслуживания. В частности, вместо символа X при использовании дисциплины «первым пришел – первым обслужен» указывается аббревиатура *FIFO*.

Далее будем использовать обозначения (1.18) и (1.19). С учетом записи вида (1.19) можно классифицировать СМО, как показано на рис. 1.11. В данную классификацию включены только системы с одним обслуживающим прибором. Более того, акцентируется внимание на том факте, что основной материал данного учебного пособия будет посвящен системам с ожиданием, в которых заявки обслуживаются без приоритетов.

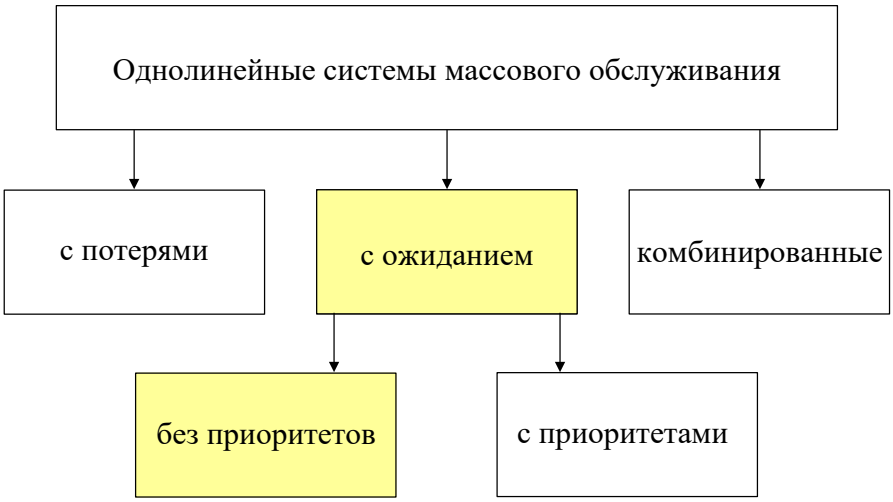


Рис. 1.11. Классификация систем массового обслуживания

Однолинейные СМО с потерями образуют класс моделей, который используется не очень часто. По этой причине здесь будут приведены только соотношения, позволяющие рассчитывать основные характеристики СМО. Для комбинированных систем также приводятся лишь основные результаты. Правда, интерес к соответствующим моделям существенно выше, но во многих случаях их можно исследовать как системы с ожиданием. Возникающая при таком допущении ошибка, как будет показано ниже, обычно находится в разумных пределах.

1.6. Аспекты качества обслуживания

В рекомендации Международного союза электросвязи (МСЭ) E.800[3] определены термины, прямо или косвенно относящиеся к качеству обслуживания. Термин «*качество*» определен Международной организацией по стандартизации. Она более известна по аббревиатуре ISO (International Organization for Standardization). Под качеством будем понимать совокупность характеристик объекта (или процесса), которые имеют отношение к его возможностям удовлетворять установленные или предполагаемые потребности. С этой точки зрения формулировка термина «*качество обслуживания*» требует только лишь конкретизации предыдущего определения. Во-первых, подразумеваются характеристики *обслуживания*. Во-вторых, речь должна идти о потребностях пользователя *в услугах связи*.

В английском языке качеству обслуживания соответствует термин *Quality of Service* (QoS). С точки зрения вопросов, которые рассматрива-

[3] С текстом рекомендаций МСЭ можно ознакомиться на официальном сайте этой международной организации – www.itu.int.

ются в данном учебном пособии, интересен ряд положений, изложенных в рекомендации МСЭ Е.800.

Во-первых, целесообразно остановиться на так называемом «*сквозном качестве обслуживания*». Это словосочетание – перевод с английского языка такого выражения: «end-to-end QoS». Суть термина «сквозное качество обслуживания» удачно иллюстрирует рис. 1.12, заимствованный из текста рекомендации Е.800. На этом рисунке показаны основные компоненты сети связи, используемые для обмена информацией между двумя пользователями. Окончания нижней стрелки свидетельствуют, что для оценки сквозного качества используется, в том числе, и мнение пользователя.

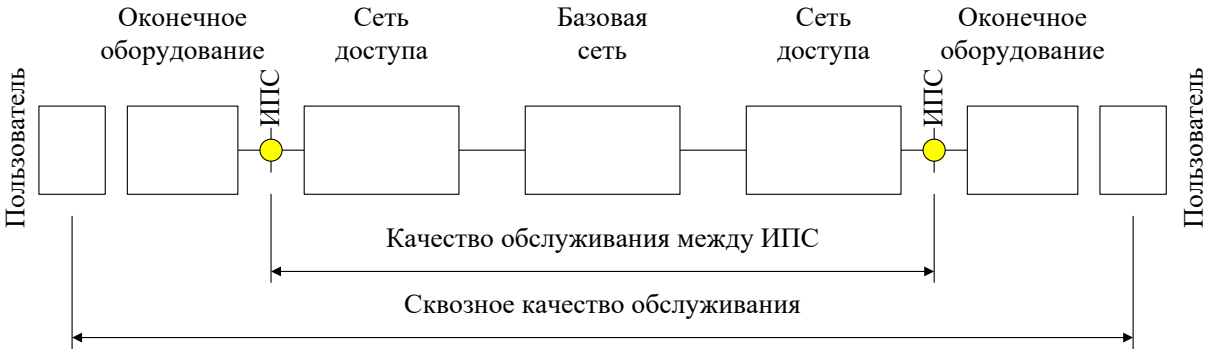


Рис. 1.12. Основные компоненты телекоммуникационной сети

Между компонентами «сеть доступа» и «оконечное оборудование» – в дополнение к модели МСЭ – указаны интерфейсы пользователь–сеть (ИПС). Они будут необходимы для постановки некоторых задач по расчету характеристик СеМО.

Во-вторых, следует подчеркнуть мнение МСЭ относительно уже использованного в этом разделе слова «характеристики». Каждая характеристика, так или иначе связанная с качеством обслуживания, должна быть наблюдаемой или измеряемой. В том случае, когда характеристики специфицированы, они становятся параметрами и могут быть выражены показателями (метриками).

Параметром обычно называется количественная характеристика обслуживания с конкретной областью применения и установленными границами. Параметры могут быть объективными и субъективными. В учебном пособии рассматриваются только задачи, связанные с объективными параметрами. Типичный пример параметра – число вызовов, потерянных в течение ЧНН.

В-третьих, для анализа некоторых показателей качества обслуживания используются методы, которые не связаны с теорией телетрафика. В качестве характерного примера подобных показателей следует назвать коэффициент ошибок по битам. Тем не менее ряд показателей такого рода влияет на оценку параметров, напрямую относящихся к теории теле-

трафика. Предположим, что все искаженные IP-пакеты переспрашиваются. С точки зрения теории телетрафика это означает, что возрастает среднее значение длительности обслуживания заявок.

В рекомендациях МСЭ, а также в документах ETSI (Европейского института телекоммуникационных стандартов) приведены показатели качества обслуживания для телекоммуникационных сетей различного назначения. Эти показатели связаны не только со сквозным качеством обслуживания. Нормируются также показатели, которые можно разделить на две основные группы. К первой группе относятся показатели, характерные для фрагмента или иерархического уровня телекоммуникационной сети. Примером таких показателей можно считать вероятность потери вызова между телефонными аппаратами одной местной (городской или сельской) телефонной сети. Эта сеть – часть всемирной системы телефонной связи. Во вторую группу входят показатели, относящиеся к процессу обслуживания заявок. В качестве примера такого показателя можно назвать среднее время установления соединения между телефонными аппаратами после завершения процесса набора номера вызывающим абонентом.

Традиционно и МСЭ, и ETSI предлагают нормировать события, заключающиеся в отказе в обслуживании, задавая соответствующую вероятность. В некоторых случаях эта вероятность определяется для разных условий функционирования сети. Когда необходимо нормировать длительность выполнения процесса обычно устанавливаются два параметра. Во-первых, определяется среднее значение длительности выполнения анализируемого процесса. Во-вторых, задается квантиль соответствующей ФР. Ранее чаще встречался 95%-й квантиль. В последнее время используются и другие квантили.

С учетом этих соображений основное внимание в учебном пособии уделяется тем задачам, которые связаны с оценкой параметров для показателей, перечисленных выше. Как было подчеркнуто в тексте, который предшествует рис. 1.11, основное внимание акцентируется на системах с ожиданием. По этой причине задачам расчета вероятностей, которые характеризуют отказы в обслуживании, посвящен только один разд. 2.

1.7. Несколько положений теории телетрафика

Этот подраздел включает только те положения теории телетрафика, которые важны с точки зрения вопросов, рассматриваемых в последующих разделах. Таким образом, его нельзя считать кратким изложением базовых положений. Читателям, которые не знакомы с теорией телетрафика, рекомендуется прочесть, как минимум, одну из книг, посвящен-

ных этой дисциплине [1–5, 10–12]. В данном подразделе мы ограничимся семью положениями.

Первое положение связано с реакцией пользователя на качество обслуживания, характерное для основных этапов предоставления инфокоммуникационных услуг. Далее в качестве примеров таких услуг будут, в основном, рассматриваться виды обслуживания, которые обеспечиваются телефонной сетью общего пользования (ТФОП). Для этой сети будет, при необходимости, конкретизироваться способ распределения информации. Пока в ТФОП используются два способа распределения информации: коммутация каналов (чаще) и коммутация пакетов (реже).

На рис. 1.13 показаны основные этапы обслуживания вызова в ТФОП, которые целесообразно выделить для анализа процесса организации связи между терминалами двух пользователей (абонентов). Предположим, что в качестве терминалов используются телефонные аппараты (ТА). Это означает, что процесс обслуживания вызова реализуется на основании типового алгоритма, принятого в ТФОП.

Рис. 1.13. Основные этапы обслуживания вызова в ТФОП

В момент времени t_1 абонент снимает трубку телефонного аппарата. В момент времени t_2 абонент получает информацию (как правило, в виде акустического сигнала, который называется «Ответ станции») о готовности к приему номера вызываемого абонента. К моменту t_3 все цифры, определяющие номер вызываемого абонента, получены. Далее осуществляется попытка подключения к линии вызываемого абонента. В период времени между моментами t_4 и t_5 могут выполняться некоторые дополнительные операции, обусловленные, например, аспектами оплаты соединения. В большинстве случаев этот период времени равен нулю.

Между моментами времени t_5 и t_6 в терминал вызывающего абонента передаются акустические сигналы «Контроль посылки вызова» или «Занято». Далее предполагается, что терминал вызываемого абонента свободен. В этом случае в терминал вызывающего абонента поступает сигнал «Контроль посылки вызова». В момент времени t_6 вызываемый абонент снимает микротелефонную трубку. На отрезке времени (t_6, t_7) должен быть создан разговорный тракт.

Отрезок времени (t_7, t_8) – обмен информацией между пользователями. Для ТФОП он эквивалентен обычному телефонному разговору. В момент времени t_8 оба пользователя (или один из них) решают завершить процесс обмена информацией. Промежуток (t_8, t_9) определяет время, которое необходимо для освобождения ресурсов ТФОП, используемых для организации связи между двумя ТА.

Длительность отрезка времени вида (t_i, t_{i+1}) для $1 \le i \le 8$ следует рассматривать как случайную величину. Для каждой величины (например, после проведения измерений) можно определить среднее значение $t_{i,i+1}^{(1)}$, дисперсию $\sigma_{i,i+1}^2$ и ряд других характеристик. Более того, для всех случайных величин ФР определена на конечном интервале аргумента t. Правда, для упрощения исследований эмпирические распределения иногда аппроксимируют функциями $F(t)$, которые заданы для $0 \le t < \infty$.

Если время (t_1, t_2) превышает приемлемую для абонента величину, то он не будет удовлетворен обслуживанием. В ряде случаев абоненты даже отказываются от попытки установления соединения. Чтобы избежать подобных ситуаций, обычно вводятся два показателя:

- допустимое среднее время задержки получения сигнала «Ответ станции» (т. е. длительности интервала $t_2 - t_1$);
- приемлемая величина 95%-го квантиля распределения этой же случайной величины.

К моменту времени t_3 вызывающий абонент заканчивает процесс набора требуемых цифр. Предположим, что дополнительные операции для установления соединения не нужны, т. е. $t_4 = t_5$. Если к моменту времени t_5 вызывающий абонент не получит ни один из акустических сигналов («Контроль посылки вызова» или «Занято»), то он не будет удовлетворен обслуживанием. Как и для периода времени (t_1, t_2), возможны случаи, когда абоненты отказываются от попытки установления соединения. Для минимизации числа таких ситуаций обычно вводятся два показателя:

- допустимое среднее время задержки после набора номера вызывающего абонента (т. е. длительности интервала $t_5 - t_3$);
- приемлемая величина 95%-го квантиля распределения этой же случайной величины.

Допустим, что в момент времени t_6 вызываемый абонент снял микротелефонную трубку. По очевидным психологическим причинам необ-

ходимо, чтобы достаточно быстро в период времени (t_6, t_7) между двумя терминалами был установлен разговорный тракт. По этой причине нормируется среднее значение времени установления разговорного тракта и 95%-й квантиль распределения этой случайной величины. Аналогичные нормы устанавливаются для отрезка времени (t_8, t_9), чтобы оперативно освободить ресурсы ТФОП, которые могут быть использованы для обслуживания новых вызовов.

Второе положение касается реакции оборудования сети электросвязи на поведение пользователя. Рассмотрим ситуацию, когда вызывающий абонент получает акустический сигнал «Ответ станции», но не начинает набор номера. В этом случае определенная часть ресурсов ТФОП, предназначенная для группового использования, простаивает. Очевидно, что время ожидания начала набора номера вызываемого абонента, как и длительность пауз между цифрами, следует ограничить. Обычно в коммутационных станциях подобные ограничения выбираются так, чтобы для подавляющего большинства абонентов время, отведенное для начала процесса, было вполне приемлемо.

При этом в процессе эксплуатации можно оценить долю заявок x_0, которая будет потеряна из-за превышения времени ожидания начала процесса $t_{\max}$. Величину $t_{\max}$ можно считать квантилем распределения времени ожидания, который определен для значения ФР, равного $1-x_0$. Ограничения такого рода могут применяться и на отрезке времени (t_7, t_8) при поддержке некоторых видов услуг. Характерным примером подобных услуг можно считать обращения абонентов в центры обслуживания вызовов.

Третье положение связано с понятием «период занятости». ФР соответствующей случайной величины обозначается как $G(t)$. Для объяснения сущности периода занятости обратимся к рис. 1.14. Предположим, что система находится в работоспособном состоянии. Тогда она может или обслуживать заявки или простаивать.

Интервалы (t_1, t_2), (t_3, t_4) и (t_5, t_6) называются периодами занятости. Интервалы (t_2, t_3), (t_4, t_5) и (t_6, t_7) именуются соответственно периодами простоя. Очевидно, что период занятости начинается в момент времени, когда в свободную систему поступает на обслуживание первая заявка. Он заканчивается, когда последняя заявка (из общего числа поступивших за период занятости) покидает систему, а очередь на обслуживание при этом отсутствует. Длительность периода занятости – случайная величина, для которой в ряде задач по исследованию СМО необходимо найти функцию $G(t)$.

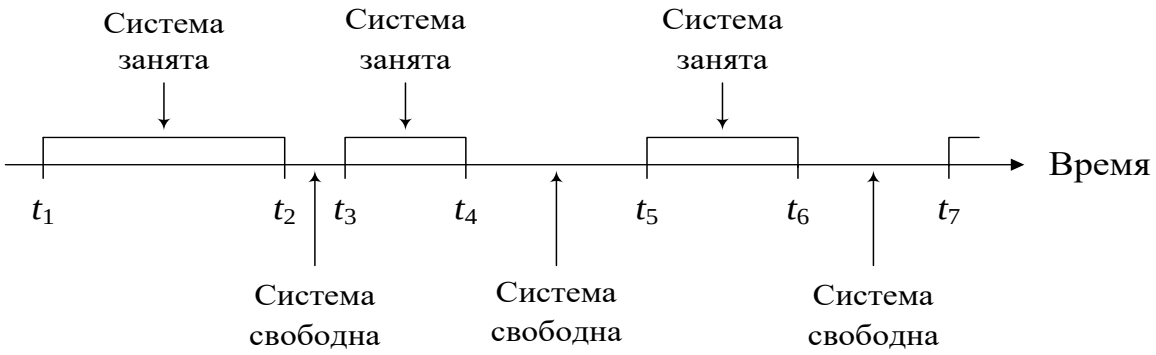


Рис. 1.14. Период занятости системы массового обслуживания

В некоторых случаях достаточно определить преобразование Лапласа–Стилтьеса данной функции распределения $\gamma(s)$. Для однолинейных систем вида $M/G/1$ искомое преобразование определяется так:

$$\gamma(s) = \beta\left[s + \lambda - \lambda\gamma(s)\right]. \tag{1.21}$$

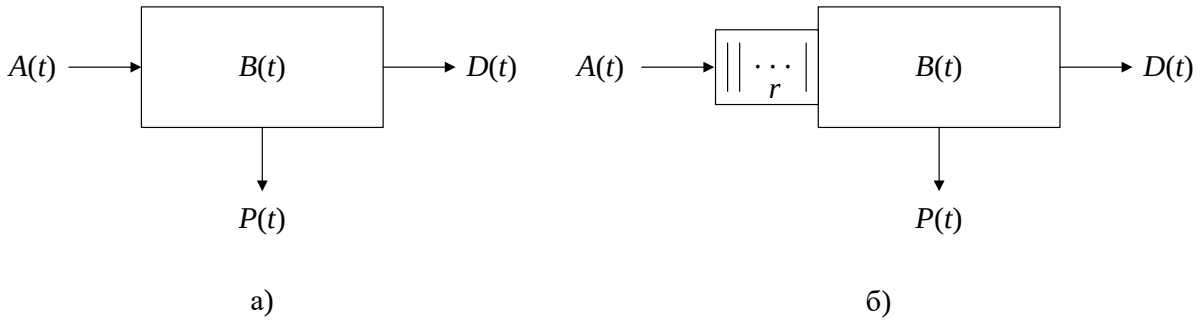
Величина λ – интенсивность входящего потока заявок. Вид соотношения (1.21) свидетельствует, что получение функции $G(t)$ представляет собой нетривиальную задачу. Для некоторых видов функции $B(t)$ распределения $G(t)$ приведены в разд. 3.

Четвертое положение иллюстрирует одно важное различие между алгоритмами обслуживания заявок с явными потерями и с ожиданием. Для объяснения этого различия обратимся к рис. 1.15. В его левой части показана модель СМО, которая не содержит мест для ожидания начала обслуживания. Если заявка приходит в течение периода занятости, то она теряется. В правой части изображена модель СМО, на входе которой установлен буферный накопитель емкостью r. Он хранит заявки, пришедшие в систему в период занятости.

Для модели системы с явными потерями допустимо условие такого рода: $\lambda > \mu$. В подобных случаях значительная часть заявок будет теряться (некоторые численные оценки приведены в разд. 2). Тем не менее можно утверждать, что СМО будет функционировать устойчиво. Это утверждение нельзя считать верным при обслуживании с ожиданием. При $\lambda > \mu$ очередь будет расти до величины r. Анализ систем, в которых используется дисциплина обслуживания с ожиданием, как правило, ограничивается следующим условием: $\lambda < \mu$. Основные соотношения для СМО с ожиданием получены именно для такого режима их функционирования. Во многих соотношениях для моделей с ожиданием фигурирует величина ρ – нагрузка (в некоторых работах она называется загрузкой):

$$\rho = \frac{\lambda}{\mu}. \qquad (1.22)$$

Неравенство $\lambda < \mu$ эквивалентно правилу $\rho < 1$. Все выражения, которые далее будут приведены, основаны на соблюдении условия $\rho < 1$.



а) б)

Рис. 1.15. Две модели систем массового обслуживания:
а) с явными потерями; б) с ожиданием

Пятое положение касается двух этапов функционирования СМО, в которых используется дисциплина обслуживания заявок с ожиданием. Подобные дисциплины предусматривают возможность формирования очереди. Очевидно, что для моделей с явными потерями время ожидания начала обслуживания и длина очереди тождественно равны нулю.

В СМО с ожиданием для всех заявок целесообразно выделить две фазы:

- ожидание в очереди W;
- обслуживание B.

Длительность каждой фазы целесообразно рассматривать как случайную величину. В учебном пособии будут рассматриваться модели, для которых можно вычислить конечные значения математического ожидания $W^{(1)}$ и $B^{(1)}$, а также дисперсии σ_W^2 и σ_B^2 обеих случайных величин.

Длительность ожидания и обслуживания образует некий отрезок времени, который называется задержкой заявок в СМО. В ряде публикаций используется термин «время пребывания». Далее этот отрезок времени обозначается буквой S.

Обычно время задержки заявок исследуется как случайная величина. Причем ее компоненты (ожидание и обслуживание) часто рассматриваются как взаимно независимые случайные величины. Практический интерес представляют математическое ожидание времени задержки заявок в системе $S^{(1)}$, дисперсия σ_S^2 и ФР $S(t)$. На основании правил, установленных для взаимно независимых случайных величин, можно записать три важных соотношения:

$$S^{(1)} = W^{(1)} + B^{(1)}, \quad \sigma_S^2 = \sigma_W^2 + \sigma_B^2, \quad \xi(s) = \omega(s) \cdot \beta(s). \quad (1.23)$$

Шестое положение связано с процессом $D(t)$, который определяет характер потока заявок, покидающих СМО. Этот процесс показан на рис. 1.1. Интерес к нему обусловлен тем, что зачастую он определяет характер потока заявок, который является входящим для следующей СМО. Соответствующая модель показана на рис. 1.16. Она включает три фазы обслуживания заявок. Каждая i-я фаза состоит из нескольких СМО – K, L и M соответственно.

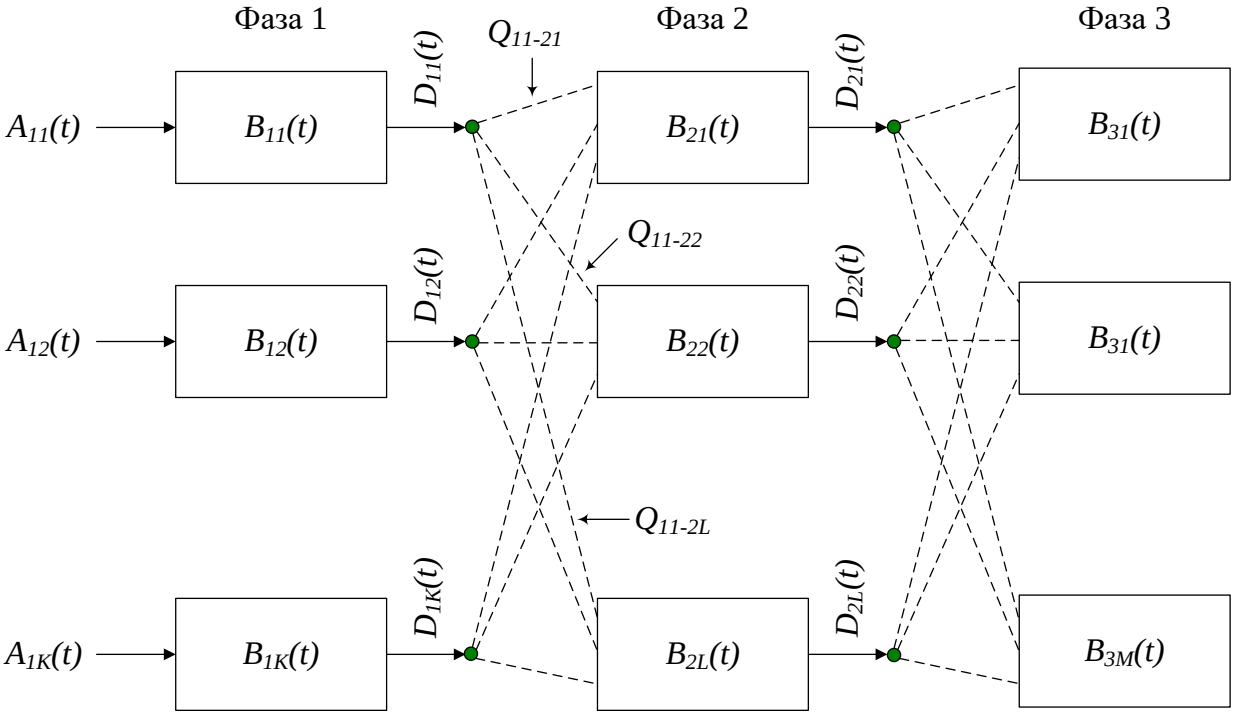


Рис. 1.16. Модель для определения процесса $D(t)$

На выходе СМО формируется выходящий поток заявок. Для распределения $D(t)$ получено преобразование Лапласа–Стилтьеса $\delta(s)$ при условии, что входящий поток является пуассоновским:

$$\delta(s) = \frac{\lambda + \rho s}{s + \lambda} \beta(s). \quad (1.24)$$

Из (1.24) можно получить формулу для коэффициента вариации длительности интервалов между моментами покидания заявками системы вида $M/G/1$:

$$C_D = \sqrt{1 - \rho^2 \left(1 - C_B^2\right)}. \quad (1.25)$$

Близость значения C_D к единице не может рассматриваться как достаточное условие для соответствия ФР $D(t)$ экспоненциальному закону

распределения. Кроме того, приближенное равенство $C_D \approx 1$ – необходимое условие для аппроксимации функции $A(t)$ экспоненциальным законом. Величина C_D близка к единице в двух случаях:

- нагрузка СМО очень мала ($\lambda << \mu$);
- коэффициент вариации времени обслуживания заявок в СМО близок к единице.

После некоторых СМО заявки продвигаются дальше. Для описания этих операций используется модель в виде СеМО. Именно такая модель показана на рис. 1.16. Для первой системы на фазе 1 обозначены вероятности Q_{ab-cd}. В рассматриваемом примере $a = b = 1$, $c = 2$, $d = \overline{1, L}$. Вероятности Q_{ab-cd} определяют доли заявок, которые будут направлены в другие системы (на следующие фазы обслуживания).

Седьмое положение посвящено процессам объединения и просеивания потоков. Для их анализа воспользуемся простой моделью, приведенной на рис. 1.17. На вход СМО поступает поток заявок, создаваемый N источниками.

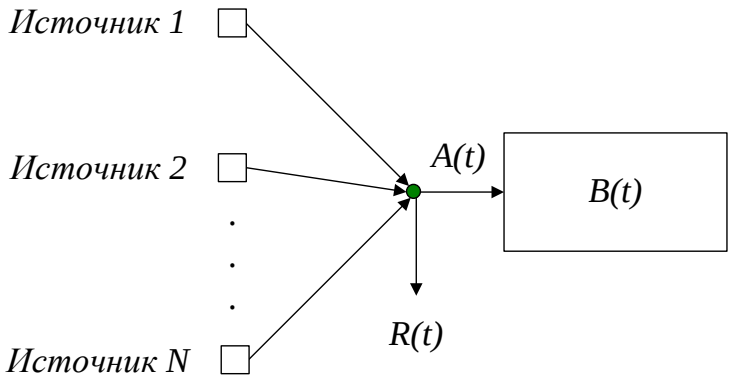


Рис. 1.17. Объединение и просеивание потоков

Процесс $R(t)$ определяет вид тех операций, которые изменяют характеристики суммарного потока заявок. Сначала рассмотрим случай, когда процесс $R(t)$ не реализуется. Это значит, что ФР $A(t)$ определяется для потока заявок, который генерируется N источниками. Допустим, что каждый из N потоков заявок может считаться пуассоновским. Пусть интенсивность i-го источника $\left(i = \overline{1, N}\right)$ равна λ_i. Тогда суммарный входящий поток заявок также будет пуассоновским с интенсивностью λ, определяемой следующей суммой:

$$\lambda = \sum_{i=1}^{N} \lambda_i . \tag{1.26}$$

Следовательно, распределение $A(t)$ будет экспоненциальным – выражение (1.8). Это свойство пуассоновских потоков заметно упрощает решение ряда практических задач. Еще одно важное свойство суммирования нескольких потоков заявок связано с тем, что при соблюдении неко-

торых условий распределение $A(t)$ также будет экспоненциальным. Предположим, что объединяется большое число стационарных ординарных потоков (с любым характером последействия). Необходимо, чтобы параметр каждого из N потоков (напомним, что этот показатель не всегда совпадает с интенсивностью) был невелик. В этом случае правомерна гипотеза о близости суммарного потока заявок к пуассоновскому. Оценку ошибок расчета характеристик СМО при замене распределения $A(t)$ экспоненциальным законом можно рассматривать как одно из важных направлений исследований в теории телетрафика.

Рассмотрим влияние двух видов процесса $R(t)$ на характер потока заявок. Первый вид процесса $R(t)$ заключается в том, что заявка, поступающая в систему, теряется с вероятностью $(1-p)$. Тогда с вероятностью p заявка попадает на обслуживание. Поток этих заявок называют просеянным. Процесс его получения именуется рекуррентной операцией просеивания. Второй вид процесса $R(t)$ основан на другом алгоритме просеивания. Ровно k заявок отбрасываются. Следующая заявка (под номером $k+1$) попадает на вход СМО. Затем снова отбрасываются k следующих заявок и т. д.

Предположим, что гипотеза о том, что суммарный поток заявок может считаться пуассоновским, вполне приемлема. Тогда при использовании рекуррентной операции просеивания на вход системы будет поступать пуассоновский поток заявок с интенсивностью $p\lambda$. Этот факт очень важен для исследования моделей СеМО. Часто заявка после завершения обслуживания с вероятностью p_j попадает в j-ю систему. Следовательно, на вход j-й системы будет поступать пуассоновский поток заявок с интенсивностью $p_j\lambda$. Реализация процесса $R(t)$ второго вида приводит к формированию потока Эрланга k-го порядка.

Контрольные вопросы и дополнительные задания

I. Попробуйте составить краткую историю развития теории телетрафика. Можно воспользоваться уже упомянутыми источниками, публикациями [13–19] и сведениями, которые размещены на сайтах в интернете.

II. Если рис. 1.3 сделать трехмерным, то какие дополнительные сведения стоит привести?

III. Записывайте в течение недели длительность каждого вашего разговора. Найдите среднее значение и дисперсию этой величины. Используя формулы, приведенные в подразд. 1.3, подберите подходящий закон распределения $B(t)$.

IV. Детализируйте рис. 1.13, подробно определив операции на отрезке времени между моментами t_4 и t_5, если необходимы дополнительные операции для уточнения, например, номера вызываемого абонента. Влияет ли длительность отрезка времени между моментами t_4 и t_5 на время обмена информацией?

V. Проанализируйте формулу (1.25). Постройте графики $C_D = f(C_B)$ при изменении C_B от нуля до двух и $C_D = f(\rho)$ для $0,05 \le \rho \le 0,95$. Попытайтесь дать количественную оценку положениям, приведенным после выражения (1.25):

- нагрузка СМО очень мала ($\lambda << \mu$);
- коэффициент вариации времени обслуживания заявок в СМО близок к единице.

Литература к разд. 1

1. *Лившиц, Б. С.* Теория телефонных и телеграфных сообщений / Б. С. Лившиц, Я. В. Фидлин, А. Д. Харкевич. – М. : Связь, 1971.
2. *Лившиц, Б. С.* Теория телетрафика / Б. С. Лившиц, А. П. Пшеничников, А. Д. Харкевич. – М. : Связь, 1979.
3. *Клейнрок, Л.* Теория массового обслуживания / Л. Клейнрок. – М. : Машиностроение, 1979.
4. *Корнышев, Ю. Н.* Теория телетрафика / Ю. Н. Корнышев, А. П. Пшеничников, А. Д. Харкевич. – М. : Радио и связь, 1996.
5. *Башарин, Г. П.* Лекции по математической теории телетрафика / Г. П. Башарин. – М. : РУДН, 2009.
6. *Вадзинский Р. Н.* Справочник по вероятностным распределениям / Р. Н. Вадзинский. – СПб. : Наука, 2001.
7. *Вентцель, Е. С.* Теория вероятностей / Е. С. Вентцель. – М. : Издательский центр «Академия», 2005.
8. *Корн, Т.* Справочник по математике для научных работников и инженеров / Т. Корн, Г. Корн. – М. : Наука, 1984.
9. *Бронштейн, И. Н.* Справочник по математике для инженеров и учащихся вузов / И. Н. Бронштейн, К. А Семендяев. – М. : Наука, 1986.
10. ITU-D. Teletraffic Engineering Handbook (edited by V.B. Iversen). – Geneva, 2003.
11. *Крылов, В. В.* Теория телетрафика и ее приложения / В. В. Крылов, С. С. Самохвалова. – СПб. : BHV, 2005.
12. *Степанов, С. Н.* Основы телетрафика мультисервисных сетей / С. Н. Степанов. – М. : Эко-Трендз, 2010.
13. *Коваленко, И. Н.* Теория массового обслуживания / И. Н. Коваленко // Итоги науки. Сер. Теория вероятностей. М. : ВИНИТИ, 1965.
14. *Климов, Г. П.* Стохастические системы обслуживания / Г. П. Климов. – М. : Наука, 1966.

15. *Башарин, Г. П.* Массовое обслуживание в телефонии / Г. П. Башарин, А. Д. Харкевич, М. А. Шнепс. – М.: Наука, 1968.

16. *Клейнрок, Л.* Вычислительные системы с очередями / Л. Клейнрок. – М. : Мир, 1979.

17. Развитие теории телетрафика в Советском Союзе / В. И Сифоров [и др.] // Модели информационных сетей и коммутационных схем. – М. : Наука, 1982.

18. *Бочаров, П. П.* Теория массового обслуживания / П. П. Бочаров, А. В. Печинкин. – М. : РУДН, 1995.

19. *Хинчин, А. Я.* Работы по математической теории массового обслуживания / А. Я. Хинчин. – М. : Либроком, 2010.

2. ОДНОЛИНЕЙНАЯ СИСТЕМА С ЯВНЫМИ ПОТЕРЯМИ

2.1. Сводка основных результатов

С практической точки зрения однолинейные системы с явными потерями интересны, пожалуй, только для анализа характеристик качества обслуживания при подключении нескольких телефонных терминалов к общей абонентской линии. В настоящее время в ТФОП применяется спаренное подключение двух терминалов к одной абонентской линии. Ранее, особенно в развивающихся странах, встречались случаи использования одной абонентской линии восемью терминалами. Однолинейные системы с явными потерями более интересны с теоретической точки зрения. Далее длительность обслуживания заявок в однолинейной системе с явными потерями предполагается случайной, распределенной по экспоненциальному закону.

Прежде всего необходимо ввести понятие *пуассоновского потока второго рода*. Этот поток генерируется конечным числом источников трафика. Согласно записи (1.20) количество источников трафика будет обозначаться как N. Для пуассоновского потока второго рода справедливо соотношение (1.8), но параметр потока λ_i зависит еще от двух переменных. Первая переменная – параметр потока заявок свободного источника *a*. В качестве второй переменной фигурирует количество активных источников трафика *i*. Величина λ_i определяется следующим соотношением [1–3]:

$$\lambda_i = (N - i)\alpha, \quad i \le N. \tag{2.1}$$

Из (2.1) следует, что когда все источники трафика активны, параметр потока равен нулю. Введем единый параметр интенсивности трафика Y. По аналогии с (1.22) эта величина вычисляется так:

$$Y = \frac{\alpha}{\mu}. \tag{2.2}$$

С учетом этих обозначений для однолинейной системы вероятность потери вызова π_C можно получить из формулы Энгсета [1–3] при условии, что количество приборов для обслуживания трафика (V) равно единице:

$$\pi_C = \frac{C_{N-1}^1 Y}{C_{N-1}^0 + C_{N-1}^1 Y} = \frac{(N-1)Y}{1 + (N-1)Y}. \tag{2.3}$$

Для расчета вероятности потерь по времени π_T и по нагрузке π_Y справедливы соотношения:

$$\pi_T = \frac{C_N^1 Y}{C_N^0 + C_N^1 Y} = \frac{NY}{1+NY}, \tag{2.4}$$

$$\pi_Y = \left(1 - \frac{1}{N}\right)\pi_T = \frac{(N-1)Y}{1+NY}. \tag{2.5}$$

Очевидно, что при $Y < \infty$ и $N < \infty$ справедливо неравенство $\pi_Y < \pi_C < \pi_T$. Этим системы с входящим пуассоновским потоком второго рода отличаются от модели Эрланга [1–3], для которой справедливо равенство потерь по вызовам (p_C), по времени (p_T) и по нагрузке (p_Y). По этой причине соответствующую вероятность обозначают буквой p без нижнего индекса. Потери для модели Эрланга при $V = 1$ и интенсивности трафика $Y_Э$ рассчитываются по формуле

$$p = \frac{Y_Э}{1+Y_Э}. \tag{2.6}$$

С учетом принятых обозначений $Y_Э = NY$. Тогда выражение (2.6) целесообразно переписать в ином виде:

$$p = \frac{NY}{1+NY}. \tag{2.7}$$

Следовательно, искомая величина совпадает с вероятностью потерь по времени для модели Энгсета. Эта вероятность, обозначенная как π_T, превышает аналогичные оценки потерь по вызовам и по нагрузке. Таким образом, более простая модель Эрланга позволяет получить точную (для потерь по времени) или верхнюю границу (для потерь по вызовам и по нагрузке) оцениваемой вероятности. В табл. 2.1 приведены результаты расчетов вероятностей p, π_T, π_C и π_Y для двух величин Y при четных значениях N от двух до восьми.

Несложно убедиться, что при $Y \to 0$ и конечных значениях N все вероятности стремятся к нулю. Также просто определить, что все вероятности стремятся к единице при конечных значениях Y и $N \to \infty$.

Интересны ошибки, возникающие при замене формул (2.3) и (2.5) соотношением (2.7), которое основано на менее точной модели. Относительные ошибки при расчетах вероятностей π_C и π_Y (δ_C и δ_Y соответственно) определяются соотношениями:

$$\delta_C = -\frac{1}{(N-1)(1+NY)}, \tag{2.8}$$

$$\delta_Y = -\frac{1}{N-1}. \qquad (2.9)$$

Таблица 2.1

Вероятности потерь для ряда значений Y и N

Вероятность	$N=2$	$N=4$	$N=6$	$N=8$
$Y=0,05$				
$p(\pi_T)$	0,090909	0,166667	0,230769	0,285714
π_C	0,047619	0,130435	0,200000	0,259259
π_Y	0,045455	0,125000	0,192308	0,250000
$Y=0,15$				
$p(\pi_T)$	0,230769	0,375000	0,473684	0,545454
π_C	0,130435	0,310345	0,428571	0,512195
π_Y	0,115385	0,281250	0,394737	0,477273

Из (2.8) и (2.9) следует, что $\delta_C \le \delta_Y$. Величины ошибок близки друг к другу при малых значениях Y. Относительная ошибка δ_C должна рассматриваться как функция, зависящая и от Y и от N. Относительная ошибка δ_Y определяется только величиной N. Функции $\delta_Y = f(N)$ и $\delta_C = f(Y,N)$ определены для $N \ge 2$. Из соотношения (2.9) можно определить количество источников трафика N_{MIN}, начиная с которого переход к модели Эрланга не приводит к ошибке (по модулю) более заданного порога δ_{MAX}:

$$N_{\text{MIN}} = \left\lceil \frac{1-\delta_{\text{MAX}}}{\delta_{\text{MAX}}} \right\rceil. \qquad (2.10)$$

Знак $\lceil\ \rceil$ указывает на то, что результат деления округляется до большего целого значения.

На рис. 2.1 приведены кривые для оценки относительной ошибки при расчетах вероятностей π_C и π_Y по (2.7).

Все кривые построены для модулей величин δ_C и δ_Y. Значения δ_C и δ_Y указаны по оси ординат в логарифмическом масштабе.

Обычно компоненты сетей электросвязи, исследуемые как однолинейные СМО, рассчитываются на сравнительно малые величины интенсивности нагрузки. По этой причине в качестве меры относительной ошибки, обусловленной переходом к модели Эрланга, следует использовать соотношение (2.9).

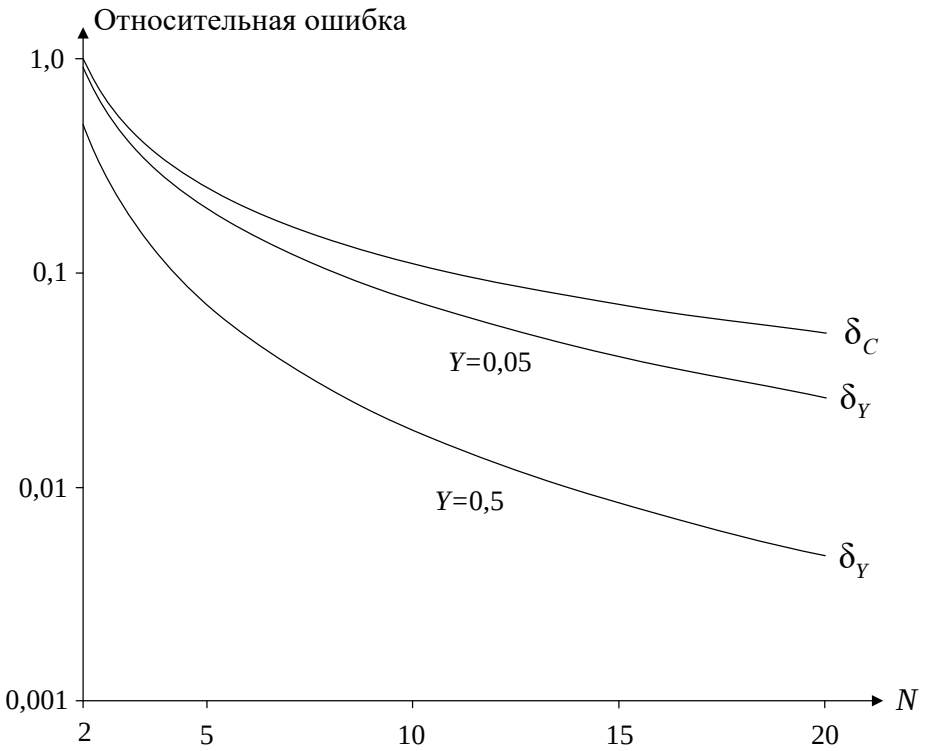


Рис. 2.1. Оценки относительной ошибки при расчете вероятностей π_C и π_Y

Контрольные вопросы и дополнительные задания

I. Проанализируйте поведение величин π_T, π_C и π_Y для спаренного включения двух телефонных аппаратов, т. е. для случая $N = 2$. Постройте функции $\pi_i(t_{АЛ})$, где $t_{АЛ}$ – время занятия абонентской линии. Величина $t_{АЛ}$ определяется как μ^{-1}.

II. Объясните значения π_C и π_Y, получаемые из (2.3) и (2.5) для одного источника нагрузки, т. е. при $N = 1$.

III. Проведите анализ мобильного телефона с двумя SIM-картами как однолинейной системы.

IV. Попробуйте провести анализ многолинейной системы ($V \geq 2$), используя методику, которая приведена в этом разделе. Постройте графики, отражающие зависимость π_T, π_C и π_Y от Y, N и V.

Литература к разд. 2

1. *Лившиц, Б. С.* Теория телефонных и телеграфных сообщений / Б. С. Лившиц, Я. В. Фидлин, А. Д. Харкевич. – М. : Связь, 1971.

2. *Лившиц Б. С.* Теория телетрафика / Б. С. Лившиц, А. П. Пшеничников, А. Д. Харкевич. – М. : Связь, 1979.

3. *Корнышев, Ю. Н.* Теория телетрафика / Ю. Н. Корнышев, А. П. Пшеничников, А. Д. Харкевич. – М. : Радио и связь, 1996.

3. СИСТЕМЫ С ПУАССОНОВСКИМ ВХОДЯЩИМ ПОТОКОМ

3.1. Общие положения

В этом разделе приведены соотношения, позволяющие оценивать качество работы тех технических средств, которые можно анализировать методами теории телетрафика. Подобные задачи будем называть *прямыми*. *Обратные* задачи связаны с расчетом характеристик СМО по заданным показателям качества их функционирования. Пусть y – некий гипотетический показатель качества работы компонента сети электросвязи, для вычисления которого приемлемы методы теории телетрафика. Предположим, что данный показатель связан прямо пропорциональной зависимостью с параметром x, который определяет, например, производительность исследуемого компонента сети электросвязи:

$$y = ax + b. \tag{3.1}$$

Если коэффициент пропорциональности a и константа b известны, то прямая задача представляет собой процесс вычисления по (3.1). Обратной задачей будет поиск параметра x. Для выбранного примера величина x находится элементарно:

$$x = \frac{y-b}{a}. \tag{3.2}$$

При исследовании СМО и прямым и обратным задачам свойственна практическая ценность. Эти задачи легче решать для простых моделей телетрафика. К таким моделям относятся системы, для которых приемлема гипотеза о том, что входящий поток заявок можно считать пуассоновским. В этой группе моделей обычно выделяют еще более простые СМО. В частности, если распределение $B(t)$ представимо экспоненциальным законом, то анализ модели упрощается. Такая модель, обозначаемая в классификации Кендалла как $M/M/1$, позволяет получить все характеристики СМО, интересные с практической точки зрения. Соответствующие формулы приведены в подразд. 3.2.

Чуть более сложная модель – однолинейная система с постоянной длительностью обслуживания заявок. Эта модель рассматривается в подразд. 3.3. Подразд. 3.4 посвящен СМО вида $M/G/1$, для которой не важен вид распределения $B(t)$. В остальных подразделах приведены результаты анализа для трех разновидностей модели $M/G/1$. Всем системам, рассматриваемым в этом разделе, присущи следующие свойства:

- количество мест для ожидания в очереди может считаться бесконечным;

•все заявки обслуживаются без приоритетов по правилу «первым пришел – первым обслужен».

В трех следующих подразделах приводятся не все характеристики, полученные для анализируемых систем. Для подробного изучения этих характеристик целесообразно воспользоваться монографиями [1–3] или другими публикациями.

3.2. Система массового обслуживания *M* / *M* / 1

Напомним, что анализ всех видов систем с ожиданием выполняется для следующей области изменения нагрузки:

$$0 \le \rho < 1. \tag{3.3}$$

Величины p_k ($k \ge 0$) определяют вероятности того, что в однолинейной системе находятся ровно k заявок. Они называются вероятностями состояний. Очевидно, что p_0 – вероятность того, что в системе нет ни одной заявки. Если $k = 1$, то рассматривается состояние, при котором в СМО находится одна заявка, которая обслуживается. Очередь отсутствует. При $k \ge 2$ вероятность p_k характеризует состояние СМО, когда очередь состоит из $(k-1)$ заявок. При этом одна заявка (самая первая из пришедших в систему) обслуживается. Для расчета вероятностей состояний справедливо такое соотношение:

$$p_k = (1-\rho)\rho^k. \tag{3.4}$$

Очевидно, что $p_k > p_{k+1}$. Распределение вероятностей p_k является геометрическим [4].

Из (3.4) несложно получить две важные характеристики для моделей телетрафика – среднее значение количества заявок, находящихся в системе, $N^{(1)}$, а также дисперсию этой случайной величины σ_N^2:

$$N^{(1)} = \frac{\rho}{1-\rho}, \tag{3.5}$$

$$\sigma_N^2 = \frac{\rho}{(1-\rho)^2}. \tag{3.6}$$

Для нахождения ряда других характеристик однолинейных систем полезна *формула Литтла*. Она связывает величину $N^{(1)}$, интенсивность пуассоновского потока заявок λ и время их задержки (ожидание плюс обслуживание) в системе $S^{(1)}$:

$$N^{(1)} = \lambda S^{(1)}. \tag{3.7}$$

Иногда формулу Литтла записывают в иной редакции. Если $N_w^{(1)}$ и $W^{(1)}$ обозначают средние значения длины очереди и длительности ожидания, то справедливо следующее соотношение:

$$N_w^{(1)} = \lambda W^{(1)}. \tag{3.8}$$

Из (3.5), (3.7) и (3.8) элементарно выводятся выражения для расчета средних значений длительности ожидания заявок в очереди и их пребывания в однолинейной системе:

$$W^{(1)} = \frac{\rho}{\mu(1-\rho)}, \tag{3.9}$$

$$S^{(1)} = \frac{1}{\mu(1-\rho)}. \tag{3.10}$$

Время ожидания начала обслуживания и время пребывания в системе по своей природе следует рассматривать как случайные величины. Полная информация об этих величинах может быть получена из соответствующих ФР: $W(t)$ и $S(t)$. Для исследуемой модели искомые распределения определяются так:

$$W(t) = 1 - \rho e^{-(1-\rho)\mu t}, \tag{3.11}$$

$$S(t) = 1 - e^{-(1-\rho)\mu t}. \tag{3.12}$$

Все приведенные выше выражения для расчета характеристик систем класса $M/M/1$ отличаются простотой. Для распределения длительности периода занятости это утверждение нельзя считать корректным. Первая производная от соответствующей ФР представима в следующей форме:

$$\frac{dG(t)}{dt} = \frac{1}{t\sqrt{\rho}} I_1(2t\sqrt{\lambda\mu}) e^{-(\lambda+\mu)t}. \tag{3.13}$$

Здесь $I_1(t)$ – модифицированная функция Бесселя первого рода и первого порядка [5].

На рис. 3.1 приведены два графика, которые иллюстрируют зависимость средних значений количества заявок в системе и времени пребывания при разных величинах нагрузки. Оба эти параметра при приближении нагрузки к единице возрастают до бесконечности. При построении графиков принято, что $\mu = 1$.

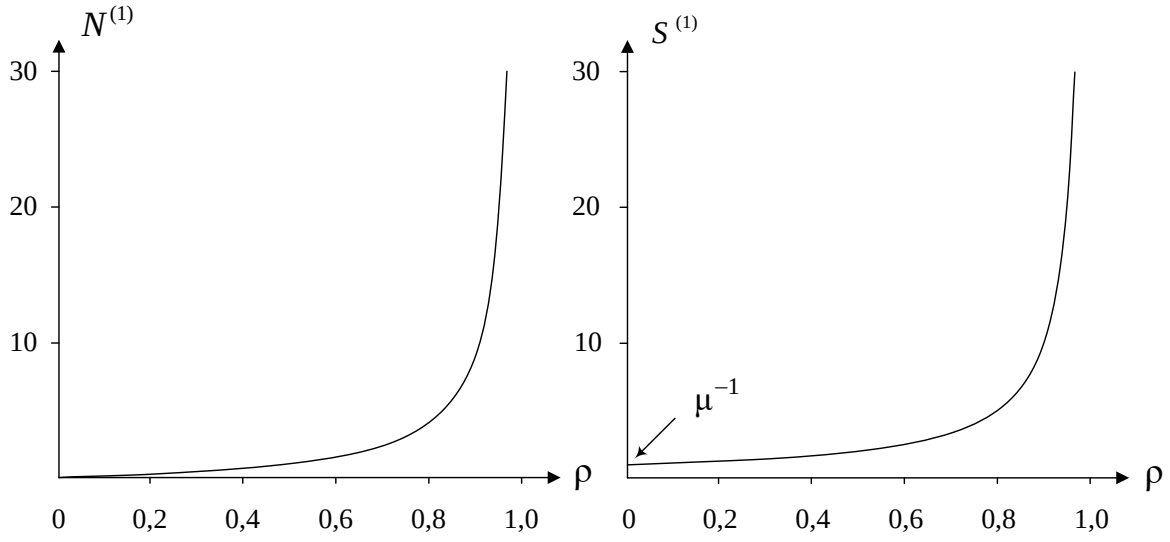


Рис. 3.1. Зависимость параметров $N^{(1)}$ и $S^{(1)}$ от нагрузки системы

На рис. 3.2 изображена дополнительная ФР длительности ожидания начала обслуживания для трех значений нагрузки ρ. По оси абсцисс отложено время, деленное на среднее значение длительности обслуживания заявок $B^{(1)}$. Оно принято за единицу. Тогда нагрузка ρ численно равна величине интенсивности входящего потока заявок λ. Подобный подход часто используется для построения графиков, иллюстрирующих поведение характеристик СМО.

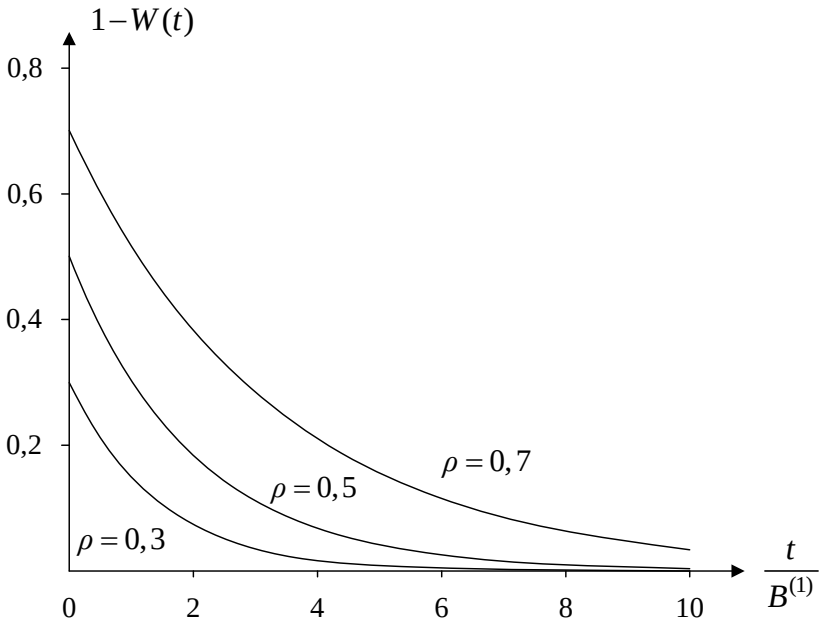


Рис. 3.2. Дополнительная функция распределения длительности ожидания

Для системы $M/M/1$ преобразование Лапласа–Стилтьеса времени обслуживания заявок определяется следующим соотношением:

$$\beta(s) = \frac{\mu}{s+\mu}. \tag{3.14}$$

Подставляя (3.14) в (1.24), получаем важное соотношение

$$\delta(s) = \frac{\lambda}{s+\lambda}. \tag{3.15}$$

Это означает, что поток заявок, которые покидают систему $M/M/1$ после их обслуживания, остается пуассоновским с интенсивностью λ. Данное обстоятельство существенно упрощает исследование СеМО. Интерес к модели $M/M/1$ обусловлен также и воспроизведением пуассоновского потока заявок на выходе системы.

3.3. Система массового обслуживания *M* / *D* / 1

Актуальность модели $M/D/1$ объясняется, как правило, двумя обстоятельствами. Во-первых, для распределений $B(t)$ с коэффициентом вариации менее единицы системы $M/M/1$ и $M/D/1$ позволяют получить верхнюю и нижнюю оценки для параметров, которые интересны с теоретической и практической точек зрения. Во-вторых, в сетях, основанных на пакетных технологиях, гипотеза о постоянном времени обслуживания заявок (передачи и обработки информационных блоков) выглядит вполне обоснованным предположением.

Среднее значение количества заявок, находящихся в системе вида $M/D/1$, рассчитывается по формуле

$$N^{(1)} = \frac{\rho}{1-\rho} - \frac{\rho^2}{2(1-\rho)}. \tag{3.16}$$

Второй член представляет собой максимум абсолютной ошибки, обусловленной заменой модели $M/G/1$ (при условии, что коэффициент вариации для распределения $B(t)$ меньше единицы) системами $M/M/1$ и $M/D/1$. В качестве относительной ошибки можно выбрать отношение второго члена в (3.16) к значению $N^{(1)}$, которое определяется выражением (3.5). Тогда искомая ошибка равна $0{,}5\rho$.

Важное свойство модели $M/D/1$ заключается в том, что длительность ожидания в очереди в ней ровно в два раза меньше, чем в системе $M/M/1$:

$$W^{(1)} = \frac{\rho}{2\mu(1-\rho)}. \tag{3.17}$$

Для вычисления среднего времени задержки заявок в системе $M/D/1$ необходимо к величине $W^{(1)}$ прибавить значение $B^{(1)}$, равное μ^{-1}:

$$S^{(1)} = \frac{2-\rho}{2\mu(1-\rho)}. \tag{3.18}$$

В системе $M/D/1$ задержка не может быть меньше величины $B^{(1)}$, что обусловлено логикой функционирования любого компонента в сетях электросвязи. Математически данное положение следует из (3.18) при $\rho = 0$.

ФР длительности ожидания заявок в очереди для систем с постоянным временем обслуживания была получена *Кроммелином*:

$$W(t) = (1-\rho)\sum_{k=0}^{\lfloor \mu t \rfloor} \frac{\left[\lambda\left(\frac{k}{\mu} - t\right)\right]^k}{k!} e^{\lambda\left(t - \frac{k}{\mu}\right)}. \tag{3.19}$$

Знак $\lfloor\ \rfloor$ указывает на то, что от результата деления берется целое значение. Зная распределение $W(t)$, несложно вычислить ФР длительности задержки заявок в системе вида $M/D/1$:

$$S(t) = \begin{cases} 0, \text{ при } t < B^{(1)}; \\ W\left(t - B^{(1)}\right), \text{ при } t \geq B^{(1)}. \end{cases} \tag{3.20}$$

Для рассматриваемой модели существенно проще – по сравнению с соотношением для системы $M/M/1$ – выглядит выражение для ФР длительности периода занятости [6]:

$$G(t) = \begin{cases} 0, \text{ при } t \leq B^{(1)}; \\ \sum\limits_{i=1}^{\left\lceil \frac{t}{B^{(1)}} \right\rceil} \frac{(i\rho)^{i-1}}{i!} e^{-i\rho}, \text{ при } t > B^{(1)}. \end{cases} \tag{3.21}$$

Равенство нулю функции $G(t)$ при $t \leq B^{(1)}$ обусловлено особенностью системы вида $M/D/1$. После поступления первой заявки она не может освободиться ранее, чем закончится время $B^{(1)}$.

Характер зависимости $S^{(1)}$ и $W^{(1)}$ от нагрузки для рассматриваемой модели такой же, как и для системы $M/M/1$. Заметно отличается поведение функции $W(t)$, рис. 3.3. На этом рисунке изображены три кривые

$W(t)$ для разных значений нагрузки. Ось абсцисс выбрана по правилам, принятым для предыдущей иллюстрации. Для оси ординат используется логарифмическая шкала.

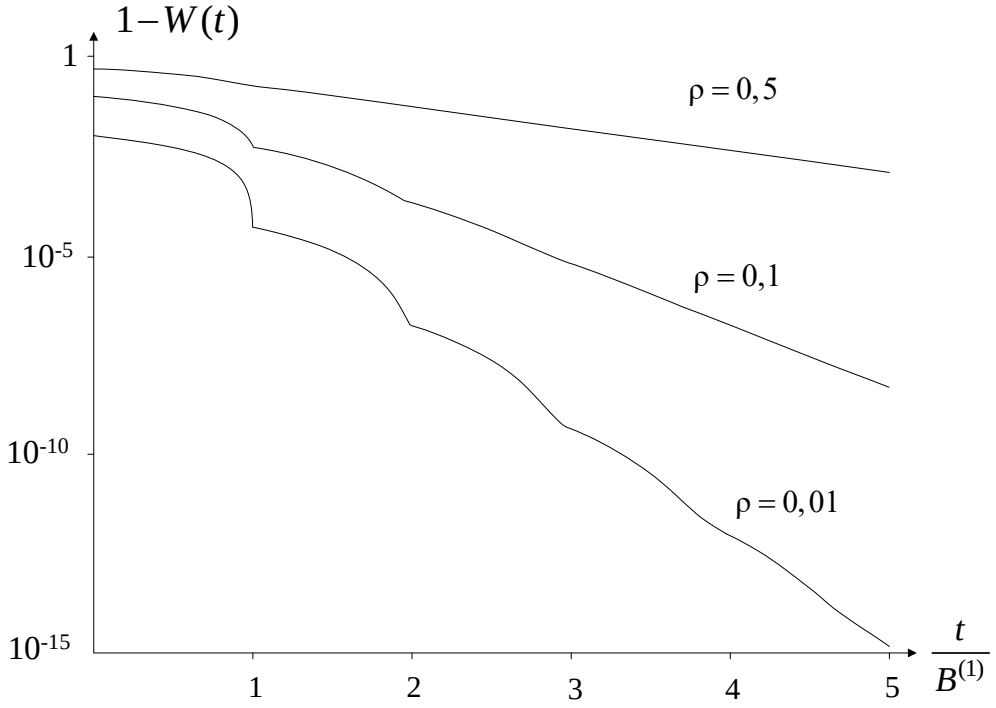


Рис. 3.3. Дополнительная функция распределения $W(t)$

Специфический характер кривых $W(t)$ обусловлен особенностью функции $B(t)$ для рассматриваемой модели СМО.

3.4. Система массового обслуживания *M* / *G* / 1

Система вида $M/G/1$ предусматривает произвольный характер функции $B(t)$. Естественно, что в процессе исследования данной модели не удается получить те же характеристики, которые известны для систем $M/M/1$ и $M/D/1$. Тем не менее, для модели $M/G/1$ выведены важные соотношения, позволяющие анализировать поведение многих компонентов сетей электросвязи.

Один из важнейших результатов исследования системы $M/G/1$ связан с формулой *Полячека–Хинчина*. В технической литературе это название применяется к нескольким соотношениям. Во-первых, формулой Полячека–Хинчина часто называют выражение для средней длины очереди:

$$N^{(1)} = \rho + \frac{\rho^2\left(1 + C_B^2\right)}{2(1-\rho)}. \qquad (3.22)$$

Величина C_B – коэффициент вариации длительности обслуживания заявок. Данный параметр часто встречается в соотношениях, справедливых для модели $M/G/1$.

Во-вторых, название «формула Полячека–Хинчина» используется также для средних значений длительности ожидания и задержки заявок:

$$W^{(1)} = \frac{\rho\left(1+C_B^2\right)}{2(1-\rho)} B^{(1)}, \tag{3.23}$$

$$S^{(1)} = \frac{2-\rho\left(1-C_B^2\right)}{2(1-\rho)} B^{(1)}. \tag{3.24}$$

В-третьих, соотношение (3.23) иногда записывают в другой форме, используя второй момент времени обслуживания заявок $B^{(2)}$:

$$W^{(1)} = \frac{W_0}{(1-\rho)}, \quad \text{где } W_0 = \frac{\lambda B^{(2)}}{2}. \tag{3.25}$$

Полученное выражение также именуют формулой Полячека–Хинчина. Напомним, что в ряде публикаций фамилия «Полячек» пишется иначе (например, Поллачек).

Еще две важные формулы известны по названию «уравнение Полячека–Хинчина». Они определяют преобразования Лапласа–Стилтьеса длительности ожидания и задержки заявок в СМО:

$$\omega(s) = \frac{s(1-\rho)}{s-\lambda+\lambda\beta(s)}, \tag{3.26}$$

$$\xi(s) = \frac{s(1-\rho)\beta(s)}{s-\lambda+\lambda\beta(s)}. \tag{3.27}$$

Получение распределений $W(t)$ и $S(t)$ на основании уравнений Полячека–Хинчина представляет собой сложную задачу. Для некоторых видов распределений $B(t)$ искомые функции выводятся в разд. 4. Напомним, что практическая ценность формул для расчета $W(t)$ и $S(t)$ объясняется нормированием квантиля ФР как одного из показателей качества обслуживания трафика. Выражения (3.26) и (3.27) можно использовать для нахождения моментов k-го порядка исследуемых случайных величин. Если известно преобразование Лапласа–Стилтьеса для ФР $\nu(s)$, то начальный момент k-го порядка случайной величины $V^{(k)}$ определяется следующим образом [7]:

$$V^{(k)} = (-1)^k \frac{d^k \nu(s)}{ds^k}\Big|_{s=0}. \qquad (3.28)$$

Получение функции $G(t)$ связано с нетривиальными преобразованиями. Правда, для некоторых задач достаточно найти моменты длительности периода занятости. Для модели $M/G/1$ три начальных момента распределения $G(t)$ определяются так:

$$G^{(1)} = \frac{B^{(1)}}{1-\rho}, \quad G^{(2)} = \frac{B^{(2)}}{(1-\rho)^3}, \quad G^{(3)} = \frac{B^{(3)}}{(1-\rho)^4} + \frac{3\lambda\left[B^{(2)}\right]^2}{(1-\rho)^5}. \qquad (3.29)$$

Влияние вида распределения $B(t)$ на характеристики системы $M/G/1$ хорошо иллюстрирует зависимость средней длительности задержки заявок от величины коэффициента вариации. На рис. 3.4 показана зависимость $S^{(1)} = f(p)$, построенная для функции $B(t)$, соответствующей гиперэкспоненциальному распределению второго порядка – формула (1.14). Величина p называется параметром формы. Она связана с коэффициентом вариации соотношением

$$C_B = \sqrt{1 + \frac{(1-2p)^2}{2p(1-p)}}. \qquad (3.30)$$

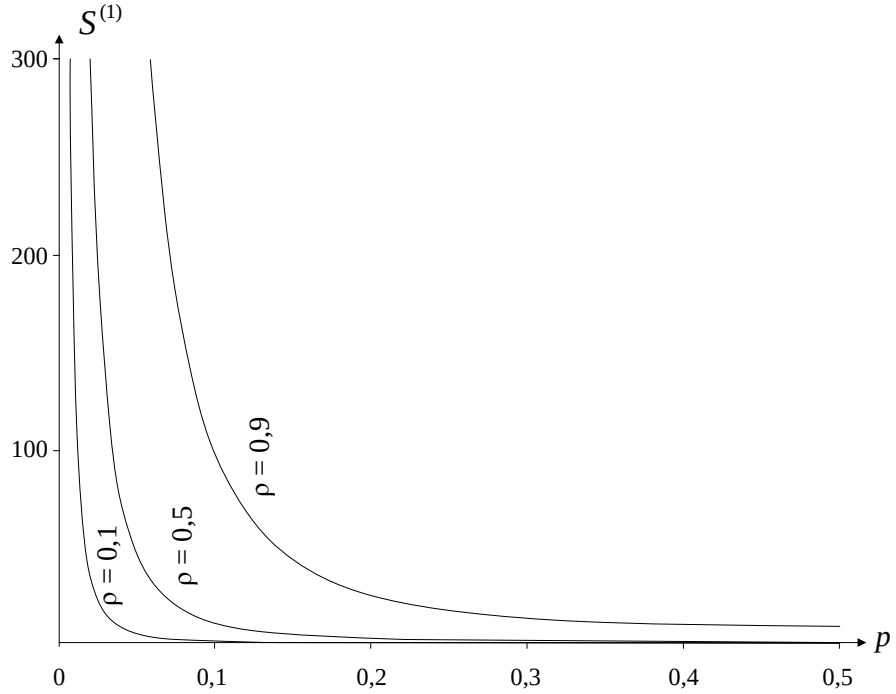


Рис. 3.4. Средняя длительность задержки заявок в системе $M/G/1$

Очевидно, что для модели $M/G/1$ влияние коэффициента вариации длительности обслуживания заявок на характеристики системы может

быть существенным. Правда, не для всех видов функции $B(t)$ коэффициент вариации может меняться в столь же широких пределах, как для гиперэкспоненциального распределения. В частности, параметр формы $p = 0{,}1$ соответствует коэффициенту вариации, равному примерно 2,13. При приближении p к 0,5 коэффициент вариации быстро стремится к единице. Если величина p становится близкой к нулю, то коэффициент вариации стремительно возрастает.

3.5. Разновидности модели вида *M / G / 1*

В следующих подразделах основное внимание уделяется двум характеристикам однолинейных систем, которые представляют собой разновидности модели $M/G/1$. Первая характеристика – среднее значение длительности задержки заявок. Обычно эта величина нормируется в качестве показателя качества обслуживания трафика при использовании дисциплин с ожиданием. Вторая характеристика – ФР длительности задержки заявок. Она нужна для оценки квантилей (одного или нескольких), которые также используются в качестве показателей качества обслуживания трафика для систем с ожиданием.

Расчет среднего значения длительности задержки заявок обычно осуществляется при исследовании процессов функционирования компонентов сети, модели которых можно рассматривать как СМО. При планировании сети чаще приходится решать обратную задачу. Допустимая величина средней длительности задержки заявок обычно заранее известна. Требуется решить одну из двух задач:

- найти максимальную величину входящего потока заявок λ при известной интенсивности обслуживания μ;
- рассчитать необходимую интенсивность обслуживания μ для известного уровня трафика λ.

Вторая задача встречается в практике проектирования сетей связи существенно чаще. Оценим величины минимально необходимой интенсивности обслуживания для двух систем: $M/M/1$ и $M/D/1$. Для модели вида $M/M/1$ искомая величина μ_{M1} может быть получена из (3.10):

$$\mu_{M1} = \frac{1 + \lambda S^{(1)}}{S^{(1)}}. \qquad (3.31)$$

Цифра 1 в нижнем индексе указывает на тот факт, что рассматривается первый из двух нормируемых показателей качества обслуживания трафика. При анализе второго показателя (квантиля) величины λ и μ будут снабжаться нижним индексом 2.

Минимально необходимая интенсивность обслуживания в системе $M/D/1$ – μ_{D1} может быть получена из (3.18):

$$\mu_{D1} = \frac{\left[1+\lambda S^{(1)}\right] + \sqrt{1+\left[\lambda S^{(1)}\right]^2}}{2S^{(1)}}. \qquad (3.32)$$

Очевидно, что $\mu_{M1} > \mu_{D1}$. Практический интерес представляет соотношение этих двух величин $\varphi_1(x)$. Переменная x определяется произведением $\lambda S^{(1)}$:

$$\varphi_1(x) = \frac{\mu_{M1}}{\mu_{D1}} = \frac{2(1+x)}{(1+x)+\sqrt{1+x^2}}. \qquad (3.33)$$

При малой интенсивности входящего потока $(\lambda \to 0)$ дробь близка к единице. Такое соотношение между величинами μ_{M1} и μ_{D2} имеет очевидный физический смысл. При малой нагрузке задержка заявки в системе определяется длительностью ее обслуживания. Эти величины в обеих системах одинаковы. При большой интенсивности входящего потока, когда $\lambda \to \mu$, можно считать, что $\lambda S^{(1)} >> 1$. Интересно, что и в этом случае результат деления μ_{M1} на μ_{D2} становится близким к единице. Суть такого соотношения понятна, если детально проанализировать поведение систем с ожиданием при $\rho \to 1$ [1]. Максимальное значение отношения (3.33) равно примерно 1,17. Из (3.10) и (3.18) следует, что для систем $M/M/1$ и $M/D/1$ максимум отношения средних задержек $\varphi_2(\rho)$ равен двум. Он наблюдается при $\rho \to 1$. Характер изменения функций $\varphi_1(x)$ и $\varphi_2(\rho)$ показан на рис. 3.5.

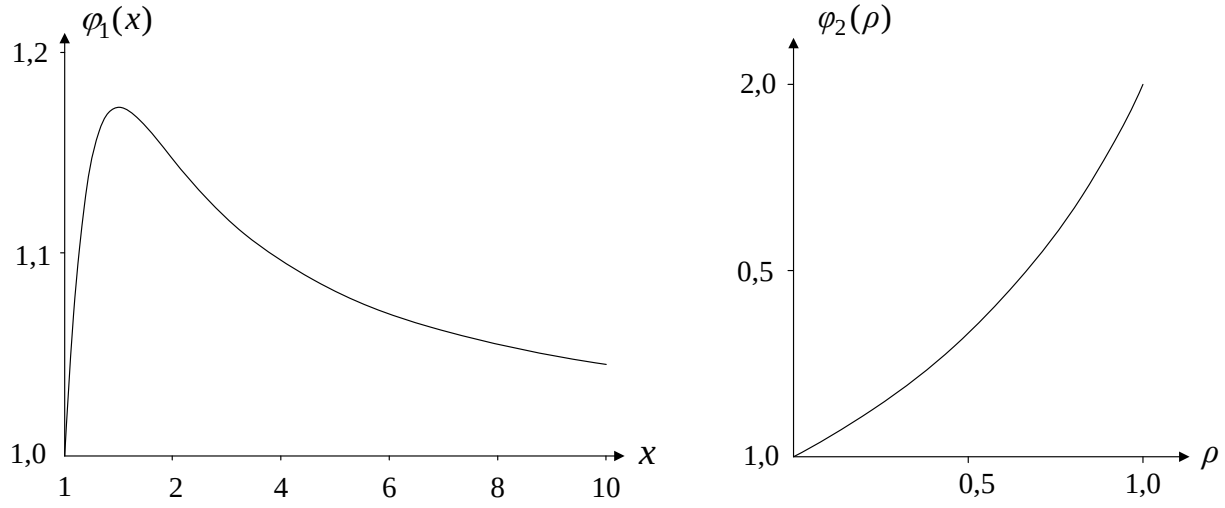


Рис. 3.5. Функции $\varphi_1(x)$ и $\varphi_2(\rho)$

Оценку квантилей распределения обычно получить труднее. Для модели $M/M/1$ данная задача решается элементарно. Формула для расчета t_p была приведена среди соотношений (1.9):

$$t_p = -\frac{1}{\lambda}\ln(1-p). \tag{3.34}$$

Вычисление квантилей t_p можно считать прямой задачей. Для планирования сети интересна обратная задача. Она, как и для среднего значения длительности задержки заявок в системе, подразумевает нахождение максимального порога λ_{M2} или необходимой интенсивности μ_{M2}. Вычислить эти значения при заданной норме t_p можно по (3.12):

$$\mu_{M2} = \lambda - \frac{ln(1-p)}{t_p}, \tag{3.35}$$

$$\lambda_{M2} = \mu + \frac{ln(1-p)}{t_p}. \tag{3.36}$$

Если нормированы оба показателя качества обслуживания трафика, то правила выбора величин λ и μ можно определить формулой

$$\lambda = \min\{\lambda_{M1}, \lambda_{M2}\}; \quad \mu = \max\{\mu_{M1}, \mu_{M2}\}. \tag{3.37}$$

Итак, с точки зрения задач по оценке показателей качества обслуживания трафика и планирования сети связи практический интерес представляют соотношения, позволяющие рассчитывать:

- среднее значение длительности задержки заявок;
- квантиль ФР длительности задержки заявок.

Соответствующие выражения позволяют решать прямые и обратные задачи. Правда, в ряде случаев обратные задачи (нахождение величин λ и μ по заданным нормам на среднее значение задержки заявок и квантиль) удается решить только численно. Так, формулу, позволяющую рассчитать *p*-квантили, можно вывести лишь для нескольких законов распределения случайных величин [4].

Модели, рассматриваемые ниже, относятся к классу систем вида $M/G/1/\infty/f_0^0$. Для оценки величины $S^{(1)}$ используется выражение (3.24), которое можно переписать в следующем виде:

$$S^{(1)} = \frac{2\mu - \lambda\left(1 - C_B^2\right)}{2\mu(\mu - \lambda)}. \tag{3.38}$$

Необходимо найти либо допустимый уровень λ, либо требуемую интенсивность μ. В этом и в следующих подразделах нижние индексы при величинах λ и μ не указываются. Искомые параметры определяются из выражения (3.38):

$$\lambda = \frac{2\mu\left(\mu S^{(1)} - 1\right)}{2\mu S^{(1)} - \left(1 - C_B^2\right)}, \tag{3.39}$$

$$\mu = \frac{\left(1 + \lambda S^{(1)}\right) + \sqrt{1 + 2\lambda S^{(1)} C_B^2 + \left(\lambda S^{(1)}\right)^2}}{2S^{(1)}}. \tag{3.40}$$

На рис. 3.6 показана зависимость минимальной интенсивности обслуживания заявок от коэффициента вариации времени обслуживания. Предполагается, что $\lambda = 1$. При этом нормируемая величина $S^{(1)}$ принимает такие значения: 1, 2 и 5.

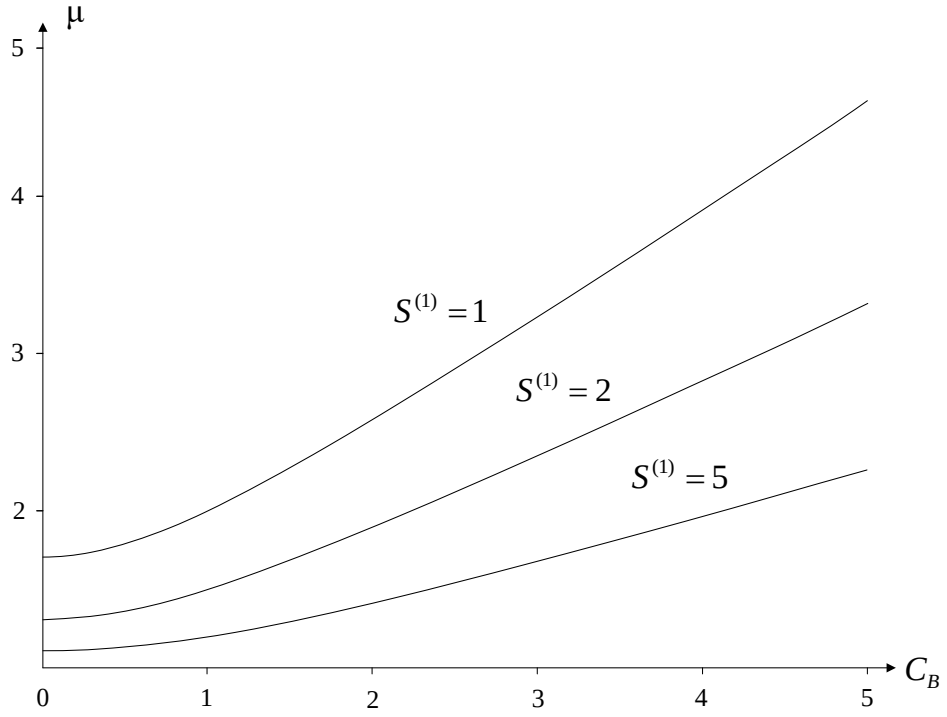


Рис. 3.6. Зависимость величины μ от коэффициента вариации C_B

Три кривые, изображенные на рис. 3.6, иллюстрируют интуитивно понятную тенденцию. Если допустима небольшая средняя задержка заявок (в частности, $S^{(1)} = 1$), то необходимая величина интенсивности обслуживания быстро возрастает при увеличении коэффициента вариации C_B. Для сравнительно больших допустимых средних значений задержки

заявок (например, при $S^{(1)} = 5$) требуемая величина μ растет медленнее при повышении коэффициента вариации C_B.

При использовании моделей вида $M/G/1/\infty/f_0^0$ возникает вопрос оценки точности допущения, связанного с неограниченным количеством мест для ожидания в очереди. Рассмотрим характеристики двух СМО: $M/M/1/\infty/f_0^0$ и $M/M/1/r/f_0^0$. Вероятность потери заявок для второй модели π определяется ее состоянием $(r+1)$, т. е. $\pi = p_{r+1}$. Искомая вероятность вычисляется по следующей формуле [8]:

$$\pi = \frac{(1-\rho)\rho^{r+1}}{1-\rho^{r+2}}. \tag{3.41}$$

Для системы с неограниченным количеством мест для ожидания в очереди (первая модель) вероятности состояний определяются соотношением (3.4). Вероятность того, что заявка застанет систему в состоянии $(r+1)$, будет определяться так:

$$p_{r+1} = (1-\rho)\rho^{r+1}. \tag{3.42}$$

Относительную ошибку в оценке вероятности состояний δ можно определить простым способом:

$$\delta = \frac{|\pi - p_{r+1}|}{\pi} = \rho^{r+2}. \tag{3.43}$$

На рис. 3.7 показана зависимость δ от r при различных значениях ρ. При сравнительно больших значениях r величины π и p_{r+1} быстро сближаются. Данный факт позволяет надеяться, что и для других характеристик СМО ошибка, которая возникает при использовании модели с неограниченным количеством мест для ожидания в очереди, не будет существенной. Конечно, такое предположение будет корректным при небольшой нагрузке системы и достаточном количестве мест для ожидания начала обслуживания.

Целесообразно проверить сформулированную выше гипотезу для средних значений длительности задержки заявок. Для модели $M/M/1/r/f_0^0$ данный показатель может быть получен в виде [9]

$$S^{(1)} = \frac{1}{\mu(1-\rho)} - \frac{(r+2)\rho^{r+2}}{\lambda\left(1-\rho^{r+2}\right)}. \tag{3.44}$$

Сравнение этого соотношения с выражением (3.10) показывает, что второй член – это абсолютная ошибка в оценке среднего значения вре-

мени задержки заявок при замене исследуемой модели системой вида $M/M/1/\infty/f_0^0$. Относительная ошибка δ может оцениваться следующим образом:

$$\delta = \frac{(r+2)\rho^{r+1}(1-\rho)}{(r+2)\rho^{r+1}-(r+1)\rho^{r+2}-1}. \qquad (3.45)$$

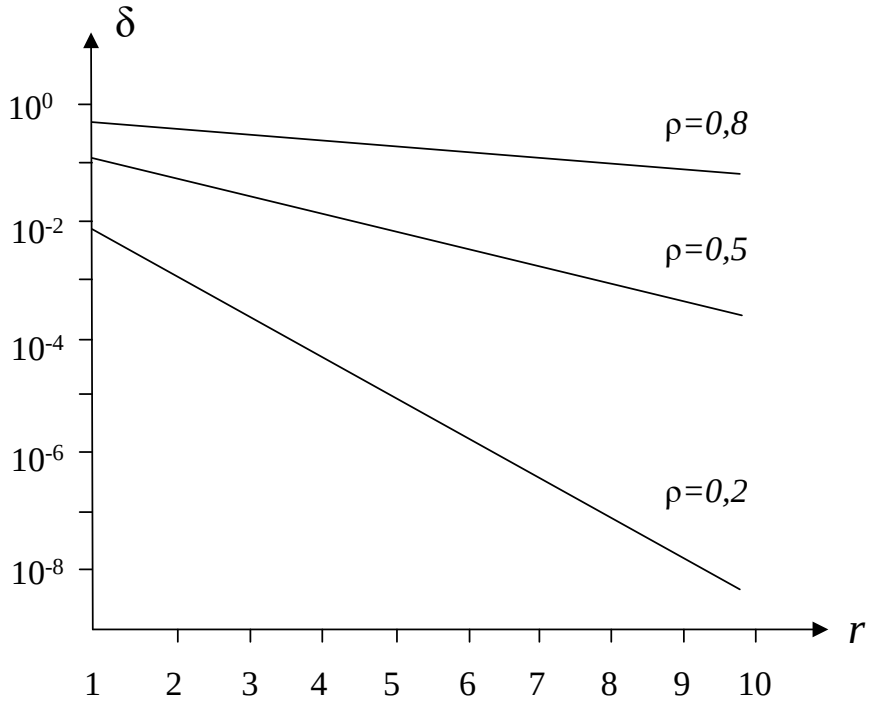


Рис. 3.7. Ошибки в расчете вероятностей состояний для модели $M/M/1/r/f_0^0$

На рис. 3.8 приведены кривые, иллюстрирующие зависимость модуля δ от r при различных значениях ρ. Эти кривые похожи на те, что были показаны на рис. 3.7. Для реально используемых значений нагрузки и количества мест для ожидания в очереди модель $M/M/1/\infty/f_0^0$ позволяет получить оценки показателей качества обслуживания с весьма высокой точностью. Для анализа режима функционирования сетей электросвязи при больших нагрузках и существенном ограничении количества мест для ожидания в очереди подобная замена модели приводит к заметным ошибкам в оценке показателей качества обслуживания.

Таким образом, модель вида $M/G/1/\infty/f_0^0$ позволяет с весьма высокой точностью исследовать СМО за исключением режимов, связанных с резким ростом нагрузки. Для однолинейных систем эти условия обычно соответствуют тем ситуациям, когда $\rho \geq 0,5$.

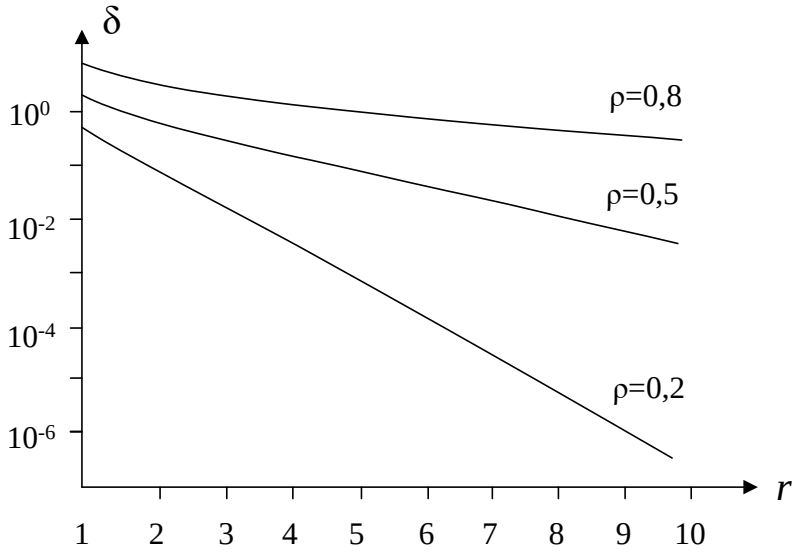


Рис. 3.8. Ошибки в расчете среднего значения длительности задержки заявок

3.6. Особенности расчета ФР длительности задержки заявок

Для исследуемых моделей телетрафика целесообразно выделить два отрезка времени с точки зрения оценки функции $S(t)$. Для первого отрезка времени – $(0, T_1)$ желательно найти точную формулу для анализируемого распределения. Если же $t > T_1$, то допустимо использовать приближение, которое позволяет оценивать функцию $S(t)$ с достаточно малой погрешностью. Обычно $T_1 \approx 5B^{(1)}$. Весьма точное приближение для распределения $S(t)$ можно получить при использовании разложения Хевисайда [7, 10]. Преобразование Лапласа–Стилтьеса для любой ФР можно представить как отношение полиномов $\eta_1(s)$ и $\eta(s)$. Если полином $\eta(s)$ имеет n простых корней, обозначаемых далее как s_i ($i = \overline{1, n}$), то оригинал можно представить в такой редакции:

$$S(t) = \sum_{i=1}^{n} \frac{\eta_1(s_i)}{\eta'(s_i)} e^{s_i t}. \tag{3.46}$$

Таким образом, разложение Хевисайда позволяет записать искомую функцию $S(t)$ в виде совокупности n экспоненциальных функций. При больших значениях t характер функции $S(t)$ будет определяться минимальным по модулю корнем s_x. Для поиска этого корня необходимо рассмотреть знаменатель выражения (3.27):

$$\eta(s) = s - \lambda + \lambda\beta(s). \tag{3.47}$$

Первый корень полинома $\eta(s)$ – ноль. Несложно убедиться, что именно этот корень определяет типичное первое слагаемое для ФР – единицу. Обозначим модуль s_x через z. В этом случае экспоненциальная функция будет выглядеть привычнее. С той же целью в числителе приближенной формулы будем использовать сомножитель $(\rho-1)$. Тогда после единицы будет стоять знак «минус», характерный для распределений случайных величин. Обозначим также значение производной от $\beta(s)$ в точке $s=-z$ через $f(-z)$. В результате этих операций приближенное выражение для оценки искомой функции приобретает следующий вид:

$$S(t) \approx 1 - \frac{(\rho-1)\beta(-z)}{1+\lambda f(-z)} e^{-zt} \approx 1 - Ae^{-zt}. \qquad (3.48)$$

Для системы $M/M/1$ после ряда несложных преобразований можно убедиться, что $z=\mu-\lambda$, $A=1$, а приближенная формула (3.48) совпадает с точной формулой (3.12). Для всех других моделей СМО формула (3.48) остается приближенной. В некоторых случаях $A>1$, что приводит к отрицательному значению ФР на начальном участке оси «Время». Этот казус приходится только на ту область изменения времени, которая не удовлетворяет неравенству $t>5B^{(1)}$, т. е. там, где применение приближенной формулы не рекомендуется. Тем не менее выражение (3.48) лучше представить в такой редакции:

$$S(t) \approx \begin{cases} 0 & \text{при } t \le t_0; \\ 1 - e^{-z(t-t_0)} & \text{при } t > t_0. \end{cases} \qquad (3.49)$$

Величина t_0 определяется при помощи простого соотношения, благодаря которому унифицируется форма представления $S(t)$:

$$t_0 = \frac{lnA}{z}. \qquad (3.50)$$

При $A<1$ величина t_0 становится отрицательной. Тогда целесообразно использовать другую трактовку хода кривой, построенной по приближенной формуле для расчета $S(t)$. В точке $t=0$ происходит скачок ФР $S(t)$. Очевидно, что величина этого скачка равна $(1-A)$, а при $t<0$ ФР времени задержки заявок равна нулю.

Вторая особенность расчета ФР заключается в корректной оценке возникающих ошибок при использовании приближенных соотношений, а также в том случае, когда следует ограничить предел суммирования рядов, содержащих бесконечное количество членов. Относительную

ошибку в расчете значений $S(t)$ целесообразно оценивать для дополнительных ФР. Причина такого подхода заключается в том, что при сравнительно больших значениях t, когда функция $S(t)$ становится близкой к единице, практически любое приближение приводит к минимальной ошибке. С дополнительной ФР подобных случаев не происходит.

Одна из самых важных практических задач заключается в том, чтобы определить область разумного использования точных и приближенных методов при вычислении функций $S(t)$. Можно выделить два фактора, влияющих на точность расчетов. Во-первых, ошибка существенно зависит от соотношения t и $B^{(1)}$. Чем оно выше, тем более точные результаты можно получить при расчетах по приближенной формуле. Во-вторых, на ошибку заметно влияет уровень нагрузки СМО. Максимальная точность, как правило, присуща моделям с высокой нагрузкой.

Следует подчеркнуть еще один аспект аппроксимаций ФР $S(t)$ формулой (3.48). Другие виды приближений, приведенные, например, в [1, 2, 11], можно представить в аналогичной форме:

$$S(t) = 1 - A_X e^{-z_X t}. \tag{3.51}$$

Различие между двумя видами аппроксимаций целесообразно оценивать отношением

$$\frac{A}{A_X} e^{-(z - z_X)t}. \tag{3.52}$$

Из этого выражения следует, что точность аппроксимаций, в которых корень z не вычисляется, а выражается через параметры модели, падает с ростом t. Поэтому расчеты дополнительных ФР длительности задержки заявок целесообразно выполнять только по формуле (3.48), которая может быть представлена в редакции (3.49).

3.7. ФР длительности задержки заявок в системе $M / G_S / 1$

В ряде случаев функция $B(t)$ определяется в результате измерений. Если процесс измерений трафика организован в соответствии с установленными принципами [12], то полученная гистограмма позволяет достоверно определить функцию $B(t)$.

Число отсчетов – элементов гистограммы – может изменяться в широких пределах. Некоторые процессы представимы ступенчатой функцией, содержащей десятки и даже сотни приращений. Без применения средств вычислительной техники использование подобных гистограмм для дальнейшего анализа СМО не представлялось возможным. Поэтому

ранее все распределения, представленные в результате проведенных измерений ступенчатыми функциями, аппроксимировались непрерывными кривыми. Такой подход в некоторых случаях позволял упростить дальнейшие вычисления. Правда, аппроксимация гистограммы непрерывной кривой, определяемой обычно на интервале $[0,\infty)$, чревата ошибками по двум основным причинам:

- замена эмпирической ступенчатой ФР $B(t)$ непрерывной кривой, которая осуществляется, например, методом наименьших квадратов [13], не позволяет численно оценить последствия этой операции;
- «пролонгация» ФР времени обслуживания, ограниченной в реальной системе связи некой величиной t_{MAX}, до бесконечности (по оси «Время») может радикально изменить результаты последующих расчетов.

На рис. 3.9 приведен пример аппроксимации результатов измерений функции $B(t)$ экспоненциальным законом. Ниже, для этого же примера, будет проанализирована ошибка, обусловленная использованием непрерывной ФР.

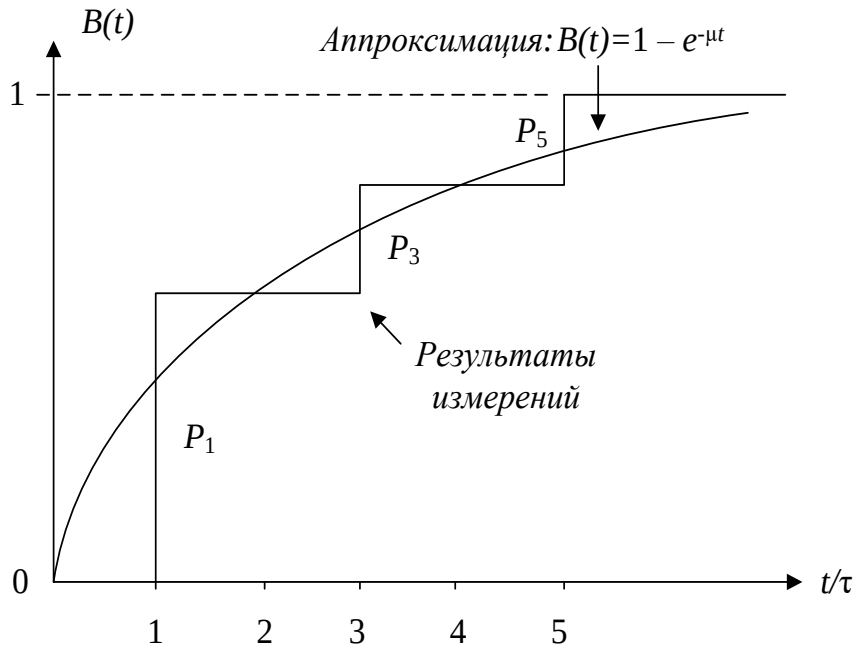


Рис. 3.9. Ступенчатая функция и ее аппроксимация непрерывной кривой

Предположим, что измерения длительности обслуживания заявок в исследуемой системе, показали следующее:

- с вероятностью P_1 заявка обслуживается за время τ;
- с вероятностью P_3 заявка обслуживается за время 3τ;
- с вероятностью P_5 заявка обслуживается за время 5τ.

Это означает, что при $t \geq 5\tau$ ФР длительности обслуживания заявок равна единице. Аппроксимация таким свойством не обладает: экспонента определена на отрезке времени $[0, \infty)$. Расчет интенсивности обслуживания заявок μ осуществляется одним из двух методов. Первый метод подразумевает оценку математического ожидания длительности обслуживания заявок $B_M^{(1)}$ по результатам проведенных измерений:

$$B_M^{(1)} = \tau(P_1 + 3P_3 + 5P_5). \tag{3.53}$$

Далее предполагается, что аппроксимирующая функция может быть представлена распределением с тем же средним значением длительности обслуживания заявок. Это означает, что справедливо равенство

$$\mu = \frac{1}{B_M^{(1)}}. \tag{3.54}$$

Второй метод основан на численном нахождении такой величины μ, чтобы ошибка аппроксимации была минимальна. Обычно применяется метод наименьших квадратов [13], апробированный для решения подобных задач.

Рассмотрим численный пример для таких результатов измерений: $P_1 = 0,6$, $P_2 = 0,3$ и $P_3 = 0,1$. Величину τ целесообразно принять равной единице. Для такого вида функции $B(t)$ получаем, что $\mu = 0,5$. При использовании метода наименьших квадратов величина интенсивности обслуживания составляет примерно 0,898. На рис. 3.10 приведены три функции $B(t)$. Нижние индексы в формулах аппроксимирующих ФР указывают на метод получения величины μ. Очевидно, что $\mu_1 = 0,5$ и $\mu_2 \approx 0,898$.

В точке τ_{-0} для обеих аппроксимаций в рассматриваемом примере наблюдается максимальная ошибка. Функция $B_1(t)$ отличается от истинного значения $B(t)$ примерно на 0,39. Для функции $B_2(t)$ эта ошибка составляет около 0,59. В третьей и в четвертой строках табл. 3.1 для обеих аппроксимаций $B(t)$ приведены численные значения относительных ошибок, возникающих при расчете некоторых характеристик времени обслуживания заявок.

Таблица 3.1

Ошибки, возникающие при аппроксимации функции $B(t)$

Исследуемые объекты	$B^{(1)}$	σ_B	C_B
Распределение $B(t)$	2	1,3416	0,6708

Ошибки для $B_1(t)$	0	49%	49%
Ошибки для $B_2(t)$	44%	17%	49%

Рис. 3.10. Две аппроксимации функции $B(t)$

Функции $B_1(t)$, для которой ошибка в расчете первого момента всегда равна нулю, свойственна большая погрешность в оценке дисперсии. Метод наименьших квадратов минимизирует дисперсию оценок. Поэтому погрешность в оценке дисперсии для функции $B_2(t)$ минимальная. Правда, ошибка при расчете математического ожидания времени обслуживания заявок становится весьма существенной.

Интересен анализ влияния аппроксимаций $B_i(t)$ на характеристики длительности задержки заявок в СМО со ступенчатой ФР времени обслуживания. Из (3.27) можно получить второй начальный момент длительности задержки заявок в СМО. Для этого нужно взять вторую производную от выражения $\xi(s)$ и найти ее значение при $s = 0$. Тогда можно составить табл. 3.2, в которой отражены ошибки оценок для трех характеристик исследуемой СМО. Следует подчеркнуть, что из-за различий в значениях μ – при одной и той же величине интенсивности входящего потока заявок нагрузка системы не будет одинаковой. По этой причине табл. 3.2 составлена для значения величины λ, которое равно 0,25. Тогда нагрузка исследуемой системы равна 0,5. Такая же величина нагрузки свойственна СМО вида $M/M/1$, для которой распределение

времени обслуживания аппроксимировано функцией $B_1(t)$. Если исходное распределение заменить функцией $B_2(t)$, то нагрузка СМО меняется. Она составляет примерно 0,278.

Таблица 3.2

Влияние ошибок аппроксимации на параметры времени задержки заявок

Исследуемые объекты	$S^{(1)}$	σ_S	C_S
Распределение $B(t)$	3,45	2,7269	0,7904
Ошибки для $B_1(t)$	16%	64%	42%
Ошибки для $B_2(t)$	55%	24%	69%

Следует подчеркнуть два обстоятельства. Во-первых, для распределений $B_1(t)$ и $B_2(t)$ значения $S^{(1)}$ и σ_S определяют соответственно верхнюю и нижнюю границы исследуемых характеристик СМО. Во-вторых, оба распределения дают оценку верхней границы для коэффициента вариации длительности задержки заявок в системе.

Ошибка в расчете среднего значения длительности задержки заявок при любой аппроксимации функции $B(t)$ зависит от нагрузки системы. Для первой аппроксимации величина относительной ошибки в оценке длительности задержки заявок δ_1 может быть вычислена по формуле

$$\delta_1 = \frac{\rho(C_B^2 - 1)}{2 + \rho(C_B^2 - 1)}. \qquad (3.55)$$

При малых величинах ρ ошибка в расчете среднего значения времени задержки заявок, обусловленная заменой истинного распределения длительности их обслуживания функцией $B_1(t)$, не столь существенна. При увеличении нагрузки ошибка δ_1 монотонно возрастает до своего максимума:

$$\frac{C_B^2 - 1}{C_B^2 + 1}. \qquad (3.56)$$

Примечательны некоторые закономерности для результатов, которые приведены в табл. 3.1 и 3.2. Данные закономерности справедливы при замене ступенчатой ФР длительности обслуживания заявок, для которой $C_B < 1$, экспоненциальным законом. Результаты двух вариантов аппроксимации функции $B(t)$ можно сформулировать так:

- метод наименьших квадратов позволяет определить нижнюю границу для среднего значения и дисперсии времени задержки заявок;

• замена ступенчатой функции $B(t)$ экспоненциальным распределением с тем же значением интенсивности позволяет получить верхнюю границу для среднего значения и дисперсии времени задержки заявок;

• использование метода наименьших квадратов приводит к меньшим ошибкам при расчете дисперсий длительности обслуживания и задержки заявок;

• замена ступенчатой функции $B(t)$ экспоненциальным распределением с тем же значением интенсивности позволяет уменьшить ошибку при вычислениях средних значений длительности обслуживания и задержки заявок;

• при любой аппроксимации исходной ФР экспоненциальным распределением рассчитанная величина C_S будет верхней границей для этой характеристики СМО.

Напомним, что для СМО вида $M/G_S/1$ математическое ожидание длительности задержки заявок рассчитывается по (3.38). Необходимо найти распределение $S(t)$. Оно может быть получено на основе результатов, которые содержатся в [14]. Автор этой публикации использовал другую форму выражения (3.27). Функция $\xi(s)$ представлена в [6] суммой членов бесконечно убывающей геометрической прогрессии:

$$\xi(s) = (1-\rho)s\sum_{i=0}^{\infty}(-1)^i \frac{\lambda^i}{(s-\lambda)^{i+1}}[\beta(s)]^{i+1}. \qquad (3.57)$$

При исследовании СМО вида $M/G_S/1$ в [6] сначала находится распределение для длительности ожидания начала обслуживания. Такой подход объясняется простой связью функций $\omega(s)$ и $\xi(s)$. Распределение длительности обслуживания заявок в исследуемой системе всегда можно представить значениями приращений P_j ($j=\overline{1,g}$) в точках $j\tau$:

$$\beta(s) = \sum_{j=1}^{g} P_j e^{-j\tau s}. \qquad (3.58)$$

Распределение задержки в системе на основании теоремы смещения [7, 10] будет определяться следующим образом:

$$S(t) = \sum_{j=1}^{g} P_j \Psi_+(t-i\tau)W(t-i\tau). \qquad (3.59)$$

Обозначение $\Psi_+(t-i\tau)$ использовано для функции, имеющей единичный скачок при нулевом значении аргумента:

$$\Psi_+(x) = \begin{cases} 1 & \text{при } x \geq 0; \\ 0 & \text{при } x < 0. \end{cases} \tag{3.60}$$

Это означает, что функция $W(t - j\tau)$ определена лишь для положительных значений аргумента. Для отрицательных значений аргумента она равна нулю. Таким образом, для системы $M/G_S/1$ достаточно вывести формулу, определяющую функцию $W(t)$. В [14] искомое выражение получено в таком виде:

$$W(t) = (1-\rho)\left\{ e^{\lambda t} + \sum_{i=1}^{\left\lfloor \frac{t}{\tau} \right\rfloor} e^{\lambda(t-i\tau)} \sum_{j=1}^{i} (-1)^j \frac{\lambda^j (t-i\tau)^j}{j!} R_{ij} \right\}. \tag{3.61}$$

Коэффициенты R_{ij} можно определять по алгоритму, приведенному в [14, таблица], или рассчитывать по рекуррентной формуле, которая предложена в [15]. Она получена для расчета коэффициентов при слагаемых вида $e^{-k\tau s}$ в результате возведения функции $\beta(s)$ в k-ю степень:

$$[\beta(s)]^k = \left(\sum_{j=1}^{g} P_j e^{-j\tau s} \right)^k = \sum_{n=k}^{kg} Q_{kn} e^{-n\tau s}. \tag{3.62}$$

В свою очередь коэффициенты Q_{kn} определяются по правилу возведения ряда в степень k:

$$Q_{kn} = \begin{cases} P_1^k \quad \text{при} \quad n = k; \\ \dfrac{\sum_{l=1}^{n-k} [l(k+1) - (n-k)] P_{l+1} Q_{k(n-l)}}{(n-k)P_1} \quad \text{при} \quad n = \overline{k+1, kg}. \end{cases} \qquad 3.63)$$

В данном случае предполагается, что $P_1 \neq 0$. Если такое условие не выполняется, то функцию $\beta(s)$ следует преобразовать. За знак суммы в (3.58) надо вынести множитель $e^{-c\tau s}$, в котором множитель c определяет количество периодов τ до первого ненулевого приращения функции $\beta(s)$. Коэффициенты R_{ij} и Q_{kn} связаны между собой простым соотношением

$$R_{ij} = Q_{ji}. \tag{3.64}$$

На рис. 3.11 приведены дополнительные ФР длительности задержки заявок в системе вида $M/G_S/1$. В качестве распределения $B(t)$ выбрана функция, изображенная на рис. 3.10. Результаты вычислений по точной формуле показаны лишь для нагрузки 0,05. Кривые, рассчитанные по приближенной формуле, построены для трех значений ρ: 0,05; 0,5 и 0,9.

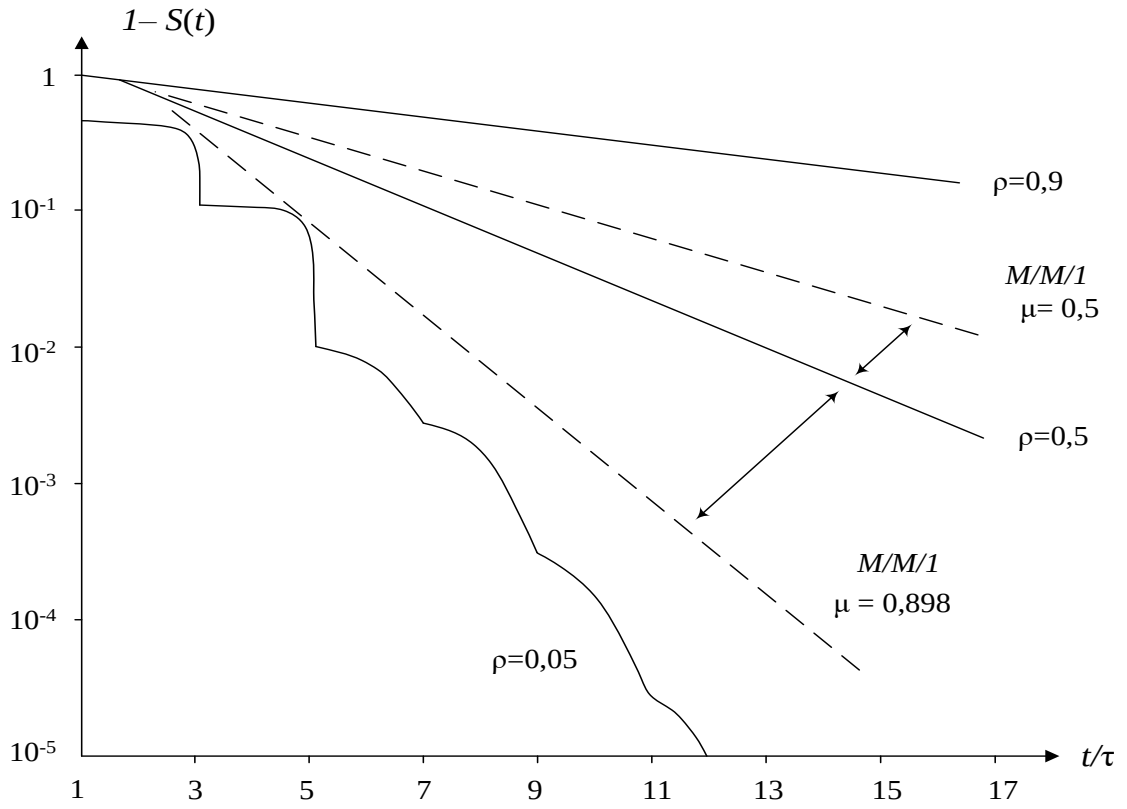


Рис. 3.11. Дополнительная ФР длительности задержки заявок в системе $M/G_S/1$

Для $\rho = 0{,}5$ на рис. 3.11 пунктирными линиями показаны две дополнительные ФР, полученные для модели $M/M/1$. Использование такой модели (как было показано на рис. 3.10) позволяет заменить ступенчатую функцию $B(t)$ экспоненциальной. Очевидно, что ошибки в расчете функции $1-S(t)$ при такой замене становятся весьма существенными. Кроме того, две рассматриваемые системы $M/M/1$ позволяют получить верхнюю и нижнюю границы для распределения времени задержки заявок.

Ошибка в расчете дополнительной ФР длительности задержки заявок уменьшается с ростом t. Для одного и того же значения t ошибка уменьшается с ростом нагрузки системы. Табл. 3.3 содержит сведения об относительных ошибках, обусловленных расчетами функции $1-S(t)$ по приближенной формуле. Очевидно, что в области $t > 5B^{(1)}$ использование

приближенной формулы позволяет выполнять вычисления с приемлемой относительной ошибкой.

Возможность использования приближенной формулы для оценки распределения длительности задержки заявок в СМО определяется видом нормируемого квантиля. Если определяется квантиль в точках, где $S(t)=0,5$ или $S(t)=0,9$, то чаще всего приходится выполнять вычисления по точной формуле. В рекомендациях МСЭ для показателей качества обслуживания пакетного трафика в ССП нормируется квантиль для $S(t)=0,999$. Тогда применение приближенных формул для расчета дополнительной ФР времени задержки заявок вполне приемлемо.

Таблица 3.3

Относительные ошибки при приближенном расчете функции $1-S(t)$

Время, $t/B^{(1)}$	Ошибка при расчете функции $1-S(t)$ для нагрузки ρ		
	$\rho=0,1$	$\rho=0,5$	$\rho=0,9$
1	–1,187	–0,444	$4,874\times10^{-3}$
2	–0,709	–0,113	$1,809\times10^{-2}$
3	–0,618	$-2,878\times10^{-2}$	$-5,781\times10^{-4}$
4	$3,871\times10^{-2}$	$1,095\times10^{-2}$	$-3,667\times10^{-4}$
5	0,169	$6,789\times10^{-3}$	$-3,725\times10^{-4}$
6	$4,075\times10^{-2}$	$1,304\times10^{-3}$	$-3,028\times10^{-4}$
7	$1,680\times10^{-2}$	$2,715\times10^{-4}$	$-2,921\times10^{-4}$
8	$-7,142\times10^{-2}$	$1,081\times10^{-4}$	$-3,569\times10^{-4}$

3.8. ФР длительности задержки заявок в системе $M/E_K/1$

Формулу (3.57) целесообразно использовать и для анализа ряда других моделей телетрафика. В этом разделе она применяется для получения функции $S(t)$ в системе вида $M/E_K/1$. В данной системе функция $B(t)$ подчиняется распределению Эрланга k-го порядка. Для этого распределения соотношение (3.57) может быть переписано следующим образом:

$$\xi(s)=(1-\rho)s\sum_{i=0}^{\infty}\frac{(-1)^i\lambda^i(k\mu)^{k(i+1)}}{(s-\lambda)^{i+1}(s+k\mu)^{k(i+1)}}. \qquad (3.65)$$

Дробь под знаком суммы может быть представлена в виде отношения полиномов $\eta_1(s)$ и $\eta(s)$. Полином $\eta(s)$ содержит кратные корни: $s_1=\lambda$ и $s_2=-k\mu$. В [7, 10] для подобных случаев предложено правило

нахождения оригинала, на основании которого функцию $S(t)$ можно представить так:

$$S(t) = (1-\rho)\sum_{i=o}^{\infty}(-1)^i \lambda^i (k\mu)^{k(i+1)} \sum_{n=1}^{2} e^{s_n t} \sum_{j=1}^{m_n} H_{nj} t^{m_n - j}. \qquad (3.66)$$

Величина m_n равна показателю степени полинома знаменателя $\eta(s)$. Возможны всего два значения этой величины: $i+1$ и $k(i+1)$. Коэффициенты H_{nj} определяются в соответствии с правилами, приведенными в [7, 10]:

$$H_{nj} = \frac{1}{(j-1)!(m_n - j)!} \frac{d^{j-1}}{ds^{j-1}} \left[\frac{(s - s_n)\eta_1(s)}{\eta(s)} \right]. \qquad (3.67)$$

После ряда преобразований выражение для расчета ФР времени задержки заявок в СМО вида $M/E_K/1$ может быть представлено в таком виде [16]:

$$S(t) = (1-\rho)\sum_{i=0}^{\infty}(-1)^i \lambda^i (k\mu)^{k(i+1)} \left[e^{\lambda t} \sum_{j=1}^{i+1} P_j t^{i-j+1} + (-1)^{i+1} e^{-k\mu t} \sum_{j=1}^{k(i+1)} Q_j t^{k(i+1)-j} \right]. \qquad (3.68)$$

Коэффициенты P_j и Q_j, входящие в формулу для искомого распределения, можно вычислить по формулам [16]:

$$P_j = \frac{(-1)^{j-1} C_{k(i+1)+j-2}^{j-1}}{(i+1-j)!(\lambda + k\mu)^{k(i+1)+j-1}}, \quad Q_j = \frac{C_{i+j-1}^{j-1}}{(k(i+1)-j)!(\lambda + k\mu)^{i+j}}. \qquad (3.69)$$

Для СМО вида $M/E_2/1$ корни s_1 и s_2 находятся в результате решения квадратного уравнения

$$s_{1,2} = \frac{(\lambda - 4\mu) \pm \sqrt{\lambda^2 + 8\lambda\mu}}{2}. \qquad (3.70)$$

Несложно убедиться, что оба корня – отрицательные величины при условии, что $\lambda < \mu$. Теперь функция $S(t)$ может быть представлена в следующей редакции:

$$S(t) = 1 - \left| \frac{s_1 e^{s_2 t} - s_2 e^{s_1 t}}{\sqrt{\lambda^2 + 8\lambda\mu}} \right|. \qquad (3.71)$$

Можно показать, что $0 \le S(t) \le 1$. Вычисление функции $S(t)$ для СМО вида $M/E_2/1$ позволяет приступить к задаче выбора верхнего предела суммирования по i в формуле (3.66). Предварительно целесообразно получить точные выражения для расчета $S(t)$ в двух других СМО – $M/E_3/1$ и $M/E_4/1$. Для этих систем искомые выражения можно представить в таком виде:

$$S(t) = 1 - \sum_{j=1}^{k} R(j,k)\,\mathrm{e}^{s(j,k)t}. \qquad (3.72)$$

Величина $s(j,k)$ – j-й корень знаменателя выражения (3.72) для распределения Эрланга третьего и четвертого порядка ($k = 3, 4$). Коэффициенты $R(j,k)$ вычисляются на основании разложения Хевисайда [7]:

$$R(j,k) = \frac{(\rho - 1)(k\mu)^k [s(j,k) + k\mu]}{\left[s(j,k) + k\mu\right]^{k+1} - \lambda k (k\mu)^k}. \qquad (3.73)$$

На рис. 3.12 показаны результаты вычисления функции $1 - S(t)$ по точной и по приближенной формулам для модели $M/E_3/1$. Расчеты выполнены для одного значения нагрузки СМО: $\rho = 0{,}1$ [17]. При более высокой нагрузке расхождение кривых, которые построены по точной и приближенной формулам, уменьшается. Ось времени нормирована относительно среднего времени обслуживания заявок $B^{(1)}$.

Численные оценки относительной ошибки при вычислениях функции $1 - S(t)$ по приближенной формуле приведены в табл. 3.4. Она составлена только для СМО вида $M/E_2/1$, так как с ростом k ошибка уменьшается. Вычисления выполнены для двух значений нагрузки: $\rho = 0{,}1$ и $\rho = 0{,}9$. Их можно считать границами диапазона реальных изменений нагрузки СМО в системах, исследуемых методами теории телетрафика.

Данные, приведенные в табл. 3.4, позволяют сделать весьма важный вывод: для выбранного диапазона изменения нагрузки СМО при $t > 5B^{(1)}$ для оценки значений функции $1 - S(t)$ можно использовать приближенную формулу. Относительная ошибка не превысит одного процента, что вполне приемлемо для решения практических задач. Следовательно, задача выбора предела суммирования по i в (3.68) интересна для диапазона $0 < t \le 5B^{(1)}$.

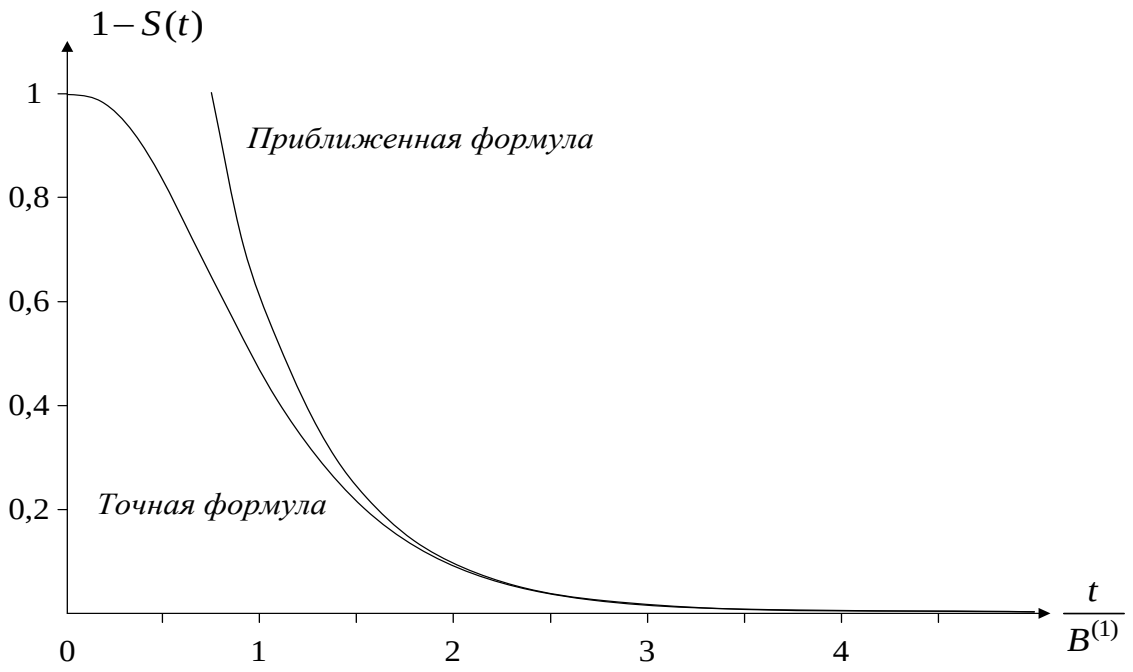


Рис. 3.12. Дополнительная ФР длительности задержки заявок в системе $M/E_3/1$

Таблица 3.4

Относительные ошибки при приближенном расчете функции $1-S(t)$

Время, $t/B^{(1)}$	Ошибка при расчете функции $1-S(t)$ для нагрузки ρ	
	$\rho=0{,}1$	$\rho=0{,}9$
1	$3{,}41\times10^{-1}$	$2{,}69\times10^{-3}$
2	$1{,}15\times10^{-1}$	$1{,}58\times10^{-4}$
3	$4{,}40\times10^{-2}$	$9{,}35\times10^{-6}$
4	$1{,}74\times10^{-2}$	$5{,}51\times10^{-7}$
5	$6{,}99\times10^{-3}$	$3{,}25\times10^{-8}$

Обозначим искомый предел суммирования через N. При малых значениях нагрузки СМО количество членов в сумме по i может быть небольшим. Например, для $\rho=0{,}1$ достаточно установить $N=4$, чтобы при $0<t\le 5B^{(1)}$ ошибка в оценке функции $1-S(t)$ не превышала одного процента. С ростом нагрузки СМО приходится увеличивать значение N. По этой причине исследование зависимости ошибок расчета функции $1-S(t)$ от величины N целесообразно проводить для высокой нагрузки СМО. Все технические устройства, исследуемые как СМО, не предназначены для работы с нагрузкой свыше 0,9. Именно этот уровень ρ целесообразно выбрать для определения значения N.

На рис. 3.13 показано поведение функции $1-S(t)$ при различных значениях N (от единицы до четырех).

Расчеты выполнены для СМО вида $M/E_3/1$ при $\rho = 0{,}9$. Четыре кривые, построенные при разных значениях верхнего предела суммирования по i в (3.68), показывают слабую зависимость ошибки в расчете функции $1-S(t)$ от N при малых значениях t. Совершенно иная картина наблюдается при $t > 2B^{(1)}$. Приемлемая ошибка в диапазоне $t \le 5B^{(1)}$ достигается при $N \ge 4$.

Величины ошибок для модели вида $M/E_2/1$ будут несколько выше. С учетом этого обстоятельства можно сформулировать такую рекомендацию: в той области изменения параметров СМО $M/E_K/1$, для которой оценку функции $1-S(t)$ следует выполнять по точной формуле, достаточно выбрать величину N, равную пяти.

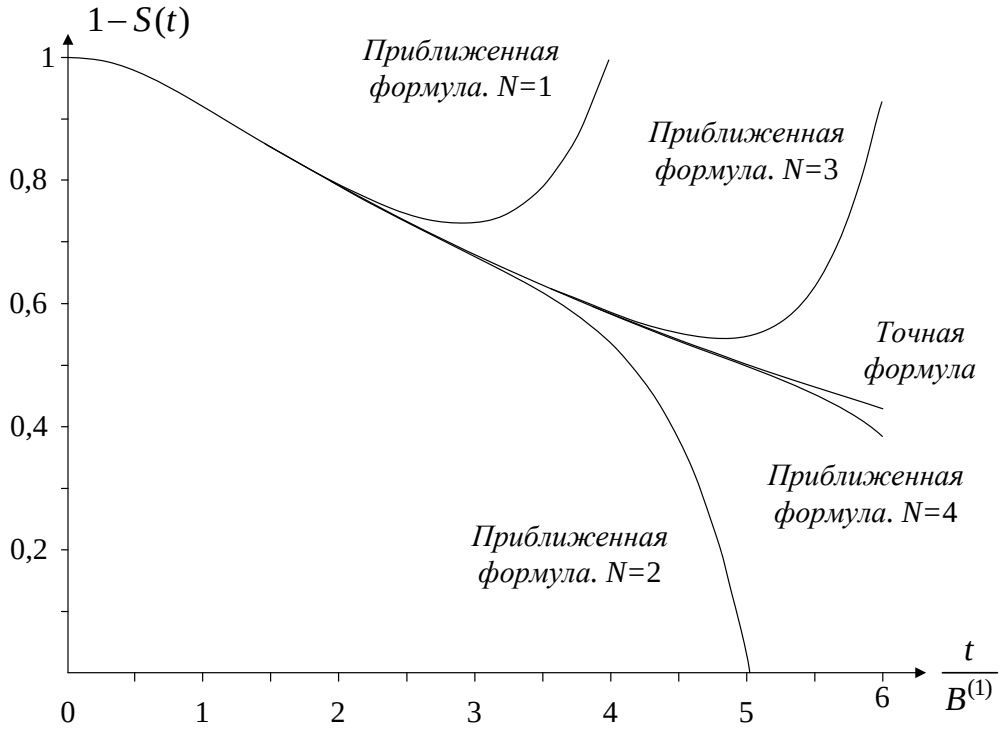


Рис. 3.13. Ошибка точности расчета функции $1-S(t)$ для модели $M/E_3/1$

Модель вида $M/E_K/1$ интересна для исследования систем, в которых длительность обслуживания заявок представляет собой случайную величину с диапазоном изменения коэффициента вариации от нуля до единицы. В частности, распределение Эрланга k-го порядка хорошо описывает продолжительность телефонного разговора.

3.9. ФР длительности задержки заявок в системе *M / U / 1*

Последняя модель, рассматриваемая в этом разделе, представляет собой систему с равномерно распределенной длительностью обслуживания заявок на отрезке времени $[0, \tau]$. Это распределение полезно, в частности, для описания процессов, похожих на операции сканирования. Слово «равномерный» обычно переводится на английский язык как «uniform». Поэтому соответствующую СМО в классификации Кендалла целесообразно обозначить $M/U/1$.

Анализ системы $M/U/1$ можно начать с функции $W(t)$ – распределения времени ожидания начала обслуживания. Формулу для расчета $W(t)$ целесообразно записать в таком виде [18]:

$$W(t) = (1-\rho)\sum_{k=0}^{\infty}\rho^k \left\{ \frac{1}{B^{(1)}}\int_0^t [1-B(x)]dx \right\}^{*(k)}. \qquad (3.74)$$

Выражение в фигурных скобках удобно представить при помощи преобразования Лапласа–Стилтьеса. Такой подход позволяет упростить вычисление *k*-кратной свертки. Преобразование для функции $B(t)$ представимо в следующей редакции:

$$\beta(s) = \frac{1-e^{-\tau s}}{s\tau}. \qquad (3.75)$$

Выражение в фигурных скобках – ФР остаточного времени обслуживания [1]. Для преобразования Лапласа–Стилтьеса этой функции справедливо равенство

$$\hat{\beta}(s) = \frac{1-\beta(s)}{sB^{(1)}} = 2\frac{\tau s-(1-e^{-\tau s})}{(\tau s)^2}. \qquad (3.76)$$

В [19] было получено выражение для расчета *k*-кратной свертки преобразования Лапласа–Стилтьеса ФР остаточного времени обслуживания. После возведения в степень *k* получаем

$$\left[\hat{\beta}(s)\right]^k = 2^k \sum_{i=0}^{k} C_k^i e^{-i\tau s} \sum_{j=0}^{k-i} \frac{(-1)C_{k-i}^j}{(\tau s)^{k+i+j}}. \qquad (3.77)$$

Из полученного выражения следует [7], что оригинал будет содержать слагаемые с множителями такого рода:

$$\frac{t^{n-1}}{(n-1)!}. \qquad (3.78)$$

Эти множители будут сдвинуты по оси «Время» вправо в соответствии с теоремой смещения [10] на величину $i\tau$. После приведения подобных членов искомое выражение приобретает вид

$$W(t) = (1-\rho)\sum_{k=0}^{\infty}(2\rho)^k \sum_{i=0}^{k} C_k^i \Psi_+(t-i\tau)\sum_{j=0}^{k-1}\frac{(-1)^j C_{k-i}^j \left(\frac{t-i\tau}{\tau}\right)^{k+i+j}}{(k+i+j)!}. \quad (3.79)$$

Очевидно, что функция $S(t)$ может быть представлена в виде

$$S(t) = F(t) - \Psi_+(t-\tau)F(t-\tau), \quad \text{где} \quad F(t) = \frac{1}{\tau}\int_0^t W(x)dx. \quad (3.80)$$

После интегрирования искомое распределение $S(t)$ запишем так:

$$S(t) = (1-\rho)\sum_{k=0}^{\infty}(2\rho)^k \sum_{i=0}^{k} C_k^i \Psi_+(t-i\tau)\sum_{j=0}^{k-i}\frac{(-1)^j C_{k-i}^j \left(\frac{t-i\tau}{\tau}\right)^{k+i+j+1}}{(k+i+j+1)!}. \quad (3.81)$$

Вычисления этой функции в диапазоне $0,05 \le \rho \le 0,95$ показали, что суммирование по k может быть ограничено 15 членами. В этом случае при $t \le 5B^{(1)}$ ошибка в расчете дополнительной ФР длительности задержки заявок в СМО не превысит 1%.

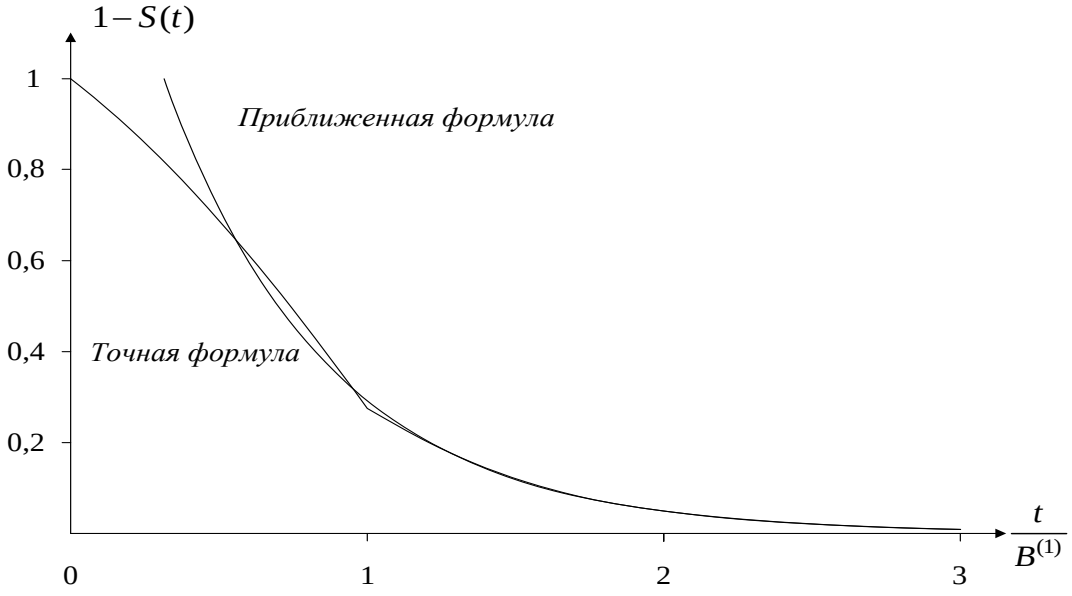


Рис. 3.14. Дополнительная ФР длительности задержки заявок в системе $M/U/1$

На рис. 3.14 показаны графики дополнительной ФР длительности задержки заявок в СМО, построенные по точной и по приближенной формулам. Вычисления выполнены для нагрузки $\rho = 0,5$. Ход двух кривых иллюстрирует хорошее совпадение результатов расчета с ростом t.

Методы получения ФР длительности задержки заявок, использованные в этом разделе, полезны для анализа других моделей вида $M/G/1$. Необходимое условие заключается в том, что функция $B(t)$ должна иметь преобразование Лапласа–Стилтьеса. Конечно, данное условие нельзя считать достаточным для получения точной формулы, которая позволяет рассчитать функцию $S(t)$.

Контрольные вопросы и дополнительные задания

I. Выведите формулу для вычисления дисперсии длительности ожидания начала обслуживания применительно к модели вида $M/M/1$.

II. Из (3.22)–(3.24) найдите аналогичные характеристики для моделей $M/M/1$ и $M/D/1$, подставляя соответствующие значения коэффициента C_B.

III. Используя правило (3.28), получите формулы (3.23) и (3.24) из соответствующих преобразований Лапласа–Стилтьеса.

IV. Найдите три начальных момента длительности периода занятости для моделей $M/M/1$ и $M/D/1$. Сравните полученные величины для нагрузки $\rho = 0,5$.

V. По аналогии с (3.31) получите выражения для максимально возможной интенсивности потока заявок λ в системе вида $M/M/1$.

VI. Докажите, что числитель в (3.39) всегда положителен.

VII. Повторите вывод основных соотношений, приведенных в подразд. 3.7, для $P_1 = 1$. Покажите, что полученное распределение длительности ожидания начала обслуживания совпадает с формулой Кроммелина.

VIII. Получите функцию $S(t)$ для модели $M/H_2/1$, используя подход, который был предложен в подразд. 3.5.

IX. Постройте графики функции $1 - S(t)$ для модели $M/U/1$ при разных значениях нагрузки.

Литература к разд. 3

1. *Клейнрок, Л.* Теория массового обслуживания / Л. Клейнрок. – М. : Машиностроение, 1979.

2. *Клейнрок, Л.* Вычислительные системы с очередями / Л. Клейнрок. – М. : Мир, 1979.

3. ITU-D. Teletraffic Engineering Handbook (edited by V.B. Iversen). – Geneva, 2003.

4. *Вадзинский, Р. Н.* Справочник по вероятностным распределениям / Р. Н. Вадзинский. – СПб. : Наука, 2001.

5. *Бейтмен, Г.* Высшие трансцендентные функции. Функции Бесселя, функции параболического цилиндра, ортогональные многочлены / Г. Бейтмен, А. Эрдейи. – М. : Наука, 1966.

6. *Риордан, Д.* Вероятностные системы обслуживания / Д. Риордан. – М. : Связь, 1966.

7. *Диткин, В. А.* Интегральные преобразования и операционное исчисление / В. А. Диткин, А. П.Прудников. – М. : Наука, 1974.

8. *Вентцель, Е. С.* Теория вероятностей / Е. С. Вентцель. – М. : Издательский центр «Академия», 2005.

9. *Крылов, В. В.* Теория телетрафика и ее приложения / В. В. Крылов, С. С. Самохвалова. – СПб. : BHV, 2005.

10. Деч Г. Руководство к практическому применению преобразования Лапласа и Z-преобразования / Г. Деч. – М. : Наука, 1971.

11. *Володин, С. В.* Об аппроксимации распределения длительности ожидания заявок в одноканальных системах массового обслуживания / С. В. Володин, К. К. Колин // Системы распределения информации. – М. : Наука, 1972.

12. *Корнышев, Ю. Н.* Теория телетрафика / Ю. Н. Корнышев, А. П. Пшеничников, А. Д. Харкевич. – М. : Радио и связь, 1996.

13. *Корн, Т.* Справочник по математике для научных работников и инженеров / Т. Корн, Г. Корн. – М. : Наука, 1984.

14. *Васильченко, А.И.* Исследование задержек сообщений в общем канале сигнализации и определение их влияния на качество обслуживания абонентов ГТС : автореф. дис. … канд. техн. наук / А. И. Васильченко. – М. : ЦНИИС, 1974.

15. *Соколов, Н. А.* Время ожидания сигнальных единиц второго относительного приоритета в общем канале сигнализации / Н. А. Соколов // Квазиэлектронная и электронная коммутационная техника : сборник научных трудов ЦНИИС. – М. : ЦНИИС, 1980.

16. *Соколов, Н. А.* Распределение длительности задержки заявок в однолинейных системах массового обслуживания / Н. А. Соколов // Модели распределения информации и методы их анализа. – М. : Наука, 1988.

17. *Яновский, Г. Г.* Оценка квантиля функции распределения времени задержки заявок в однолинейных системах массового обслуживания / Г. Г. Яновский, А. Н. Соколов. – Инфокоммуникационные технологии. – 2008. – № 4.

18. *Штойян Д.* Качественные свойства и оценки стохастических моделей / Д. Штойян. – М. : Мир, 1979.

19. *Соколов, Н. А.* Однолинейная система массового обслуживания с равномерно распределенной длительностью обслуживания заявок/ Н. А. Соколов // Модели систем информатики. – М. : Наука, 1987.

4. СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ С ПРИОРИТЕТАМИ

4.1. Актуальные задачи

Анализ СМО с приоритетным обслуживанием относится к задачам, которые можно считать одними из самых сложных в теории телетрафика. В этом разделе приводится ряд простых положений, связанных с изучением СМО, в которых применяются дисциплины приоритетного обслуживания заявок. Результаты исследования СМО с приоритетами можно найти в уже упоминавшихся монографиях [1–3]. Для желающих ознакомиться с этим направлением в теории телетрафика более подробно будут полезны также и книги других авторов [4–7].

Для СМО с приоритетами актуальны все задачи, которые были рассмотрены выше. К ним следует добавить задачи, суть которых обусловлена обслуживанием с приоритетами. Во-первых, следует выбрать подходящую дисциплину в соответствии с классификацией, приведенной на рис. 1.9. Во-вторых, необходимо выбрать уровень приоритета для заявок определенного рода. В-третьих, целесообразно сформулировать правила управления трафиком разного приоритета при различных условиях работы СМО.

Введение приоритетного обслуживания позволяет улучшить показатели качества обслуживания заявок, которым присваивается высокий приоритет. Если ресурсы СМО остаются неизменными, то показатели качества обслуживания заявок низкого приоритета снижаются (по сравнению с теми дисциплинами, которые не предусматривают никакого преимущества).

В подразд. 4.2 и 4.3 будут приведены основные результаты исследования однолинейных СМО с абсолютными и с относительными приоритетами. Для СМО обоих видов вводятся два допущения. Во-первых, предполагается, что потоки заявок любого приоритета можно считать пуассоновскими. Во-вторых, количество мест для ожидания не ограничено. В некоторых случаях накладывается ограничение на закон распределения длительности обслуживания заявок.

4.2. СМО с относительными приоритетами

Допустим, что на вход СМО поступает r потоков заявок. Интенсивность потока заявок i-го приоритета ($i = \overline{1, r}$) равна λ_i. Длительность обслуживания заявок i-го приоритета – случайная величина с распределением $B_i(t)$. Предполагается также, что для распределения $B_i(t)$ суще-

ствуют среднее значение длительности обслуживания заявок $B_i^{(1)}$ и второй момент $B_i^{(2)}$. Нагрузка СМО ρ определяется следующим правилом:

$$\rho = \sum_{i=1}^{r} \lambda_i B_i^{(1)}. \tag{4.1}$$

Это выражение иногда удобнее представлять в другой редакции, вводя нагрузку для заявок i-го приоритета ρ_i, равную $\lambda_i B_i^{(1)}$:

$$\rho = \sum_{i=1}^{r} \rho_i. \tag{4.2}$$

Суммарная интенсивность потока заявок, который обслуживается в СМО, λ, и усредненная длительность обслуживания $B^{(1)}$ рассчитываются по формулам:

$$\lambda = \sum_{i-1}^{r} \lambda_i, \tag{4.3}$$

$$B^{(1)} = \frac{1}{\lambda} \sum_{i=1}^{r} \lambda_i B_i^{(1)}. \tag{4.4}$$

Среднее значение длительности задержки заявок i-го относительного приоритета $S_i^{(1)}$ равно сумме соответствующих значений времени ожидания $W_i^{(1)}$ и обслуживания $B_i^{(1)}$:

$$S_i^{(1)} = W_i^{(1)} + B_i^{(1)}. \tag{4.5}$$

Величины $B_i^{(1)}$ для $i = \overline{1, r}$ рассчитываются элементарно. Следовательно, для оценки значений $S_i^{(1)}$ необходимо получить выражения для вычисления средних значений $W_i^{(1)}$. Их можно получить после ряда преобразований, приведенных, например, в [2–4]. Для компактной записи формулы, позволяющей рассчитать величину $W_i^{(1)}$, обычно вводится параметр υ_i:

$$\upsilon_i = \sum_{k=0}^{i} \rho_i. \tag{4.6}$$

Следует отметить, что $\rho_0 = 0$. Теперь выражение для вычисления $W_i^{(1)}$ может быть представлено в виде

$$W_i^{(1)} = \frac{\sum_{k=1}^{r} \lambda_k B_k^2}{2(1-\upsilon_i)(1-\upsilon_{i-1})}. \tag{4.7}$$

Для СМО с экспоненциальным распределением длительности обслуживания заявок (4.7) упрощается:

$$W_i^{(1)} = \frac{\rho B^{(1)}}{(1-\upsilon_i)(1-\upsilon_{i-1})}. \tag{4.8}$$

Для этого условия сравнительно компактный вид приобретает также и выражение для расчета дисперсии длительности ожидания в очереди заявок i-го относительного приоритета σ_i^2:

$$\sigma_i^2 = \rho \frac{2(1-\upsilon_i\upsilon_{i-1}) - \rho(1-\upsilon_{i-1})}{(1-\upsilon_i)^2(1-\upsilon_{i-1})^3}\left[B^{(1)}\right]^2. \tag{4.9}$$

В принципе можно получить соотношения для вычисления моментов длительности ожидания начала обслуживания любого порядка, даже если распределение $B_i(t)$ является произвольным. Для этого придется дифференцировать преобразование Лапласа–Стилтьеса $\omega_i(s)$:

$$\omega_i(s) = \frac{\left(1 - \sum_{k=1}^{r} \rho_k\right) x_i + \sum_{k=i+1}^{r} \lambda_k \left[1 - \beta_k(x_i)\right]}{s - \lambda_i + \lambda_i \beta_i(x_i)}. \tag{4.10}$$

В данное выражение входит величина x_i, которая определяется через распределение длительности периода занятости. Преобразование Лапласа–Стилтьеса этой функции уже было введено в разд. 1. Если $i = 1$, то $x_i = s$. Для остальных случаев ($r \geq i \geq 2$) величина x_i определяется так:

$$x_i = s + (1-\gamma_{i-1}) \sum_{k=1}^{i-1} \lambda_k. \tag{4.11}$$

Преобразование Лапласа–Стилтьеса γ_{i-1} определяется для распределения периода занятости СМО заявками (i – 1)-го относительного приоритета.

На рис. 4.1 показаны кривые, иллюстрирующие зависимость $W_i^{(1)}$ от нагрузки для СМО с тремя приоритетами. Предполагается, что соотношения между величинами λ_1, λ_2 и λ_3 заданы значениями 10, 30 и 60%. Распределения $B_i(t)$ считаются экспоненциальными. Средние значения длительности обслуживания заявок всех трех приоритетов полагаются идентичными. Пунктирной линией показано изменение средней длительности ожидания заявок при их обслуживании без приоритетов $W^{(1)}$.

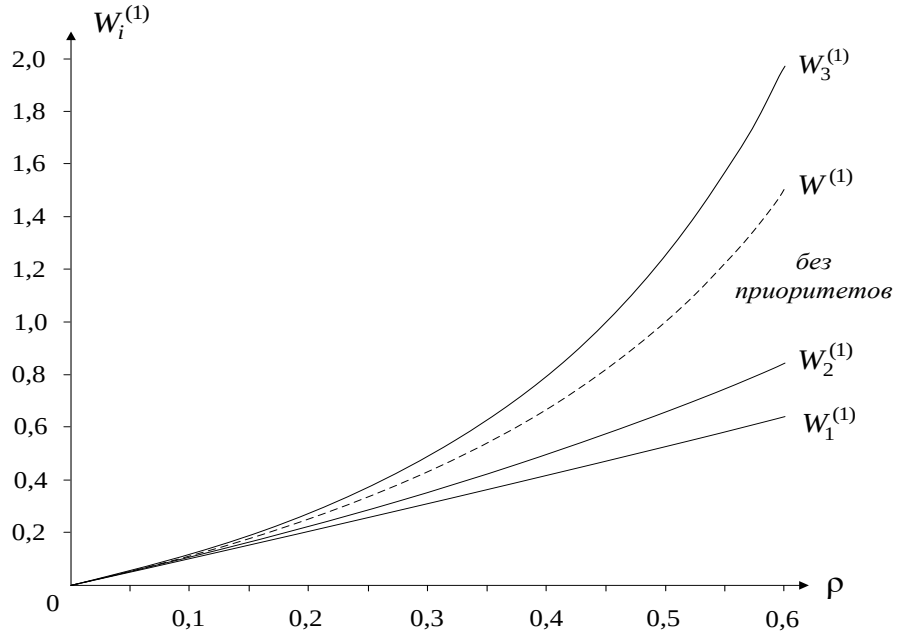


Рис. 4.1. Средние значения $W_i^{(1)}$ для СМО с тремя относительными приоритетами

Очевидно, что для выбранного соотношения между величинами λ_i средние значения длительности ожидания заявок первого и второго относительных приоритетов снижаются по сравнению с обслуживанием без преимущества. Этот вывод остается справедливым во всем диапазоне изменения нагрузки (на рис. 4.1 величина ρ ограничена порогом 0,6). Такой выигрыш достигается заметным повышением средней длительности ожидания для заявок третьего относительного приоритета. В том случае, если для заявок этого рода большие задержки допустимы (с точки зрения качества обслуживания соответствующего трафика), то введение приоритетного обслуживания можно считать оправданным.

Кроме среднего значения случайной величины всегда интересна оценка дисперсии. На рис. 4.2 приведены кривые, показывающие изменение σ_i^2 при росте нагрузки в СМО с тремя приоритетами. Используются те же условия, которые были выбраны для построения графиков на рис. 4.1. Пунктирная линия определяет изменение дисперсии при обслу-

живании заявок без приоритетов σ^2. Область изменения величины ρ ограничена, как и на рис. 4.1, порогом 0,6.

Как и следовало ожидать, введение обслуживания с преимуществом позволяет снизить дисперсию длительности ожидания для заявок первого и второго относительного приоритета. Для заявок третьего относительного приоритета дисперсия длительности ожидания возрастает (по сравнению с дисциплиной обслуживания, которая не основана на введении преимущества).

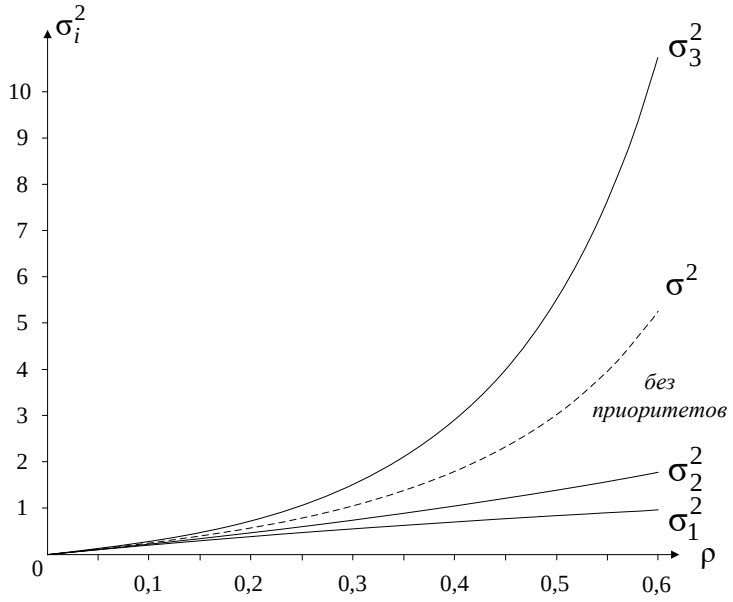


Рис. 4.2. Дисперсии σ_i^2 для СМО с тремя относительными приоритетами

Из (4.10) можно получить и набор функций распределения длительности ожидания в очереди для заявок i-го относительного приоритета $W_i(t)$. Необходимые преобразования громоздки и в этом учебном пособии они не рассматриваются.

4.3. СМО с абсолютными приоритетами

Использование относительных приоритетов означает, что обслуживание заявок не прерывается ни в каких случаях. Предположим, что в некий момент времени t_0 в СМО поступила заявка первого (самого высокого) приоритета. Возможно, что в этот момент в СМО обслуживается единственная заявка r-го (самого низкого) приоритета. Пусть T_0 – время обслуживания этой заявки. Заметим, что оно может заметно отличаться от среднего значения длительности обслуживания заявок r-го приоритета. Для рассматриваемого примера время освобождения СМО можно считать случайной величиной, распределенной равномерно на отрезке $[0, T_0]$.

Для некоторых компонентов телекоммуникационной сети возникающие задержки будут чрезмерными. Тогда могут оказаться эффективными дисциплины обслуживания с абсолютными приоритетами. Очевидно, что длительность ожидания начала обслуживания для заявок с высоким приоритетом удастся снизить. Следствием этой операции становится повышение длительности ожидания в очереди для заявок низших приоритетов.

Среднее значение ожидания заявок i-го абсолютного приоритета – $W_i^{(1)}$ может быть вычислено из соотношения

$$W_i^{(1)} = B^{(1)} \frac{\sum_{k=1}^{i} \rho_k + (1-\upsilon_i)\upsilon_{i-1}}{(1-\upsilon_i)(1-\upsilon_{i-1})}. \qquad (4.12)$$

На рис. 4.3 показаны кривые, иллюстрирующие зависимость $W_i^{(1)}$ от нагрузки для СМО с тремя приоритетами. Предполагается, что соотношения между величинами λ_1, λ_2 и λ_3 заданы теми же значениями, которые были выбраны для кривых, приведенных на рис. 4.1 и 4.2. Пунктирными линиями изображены значения $W_i^{(1)}$, полученные для модели с относительными приоритетами.

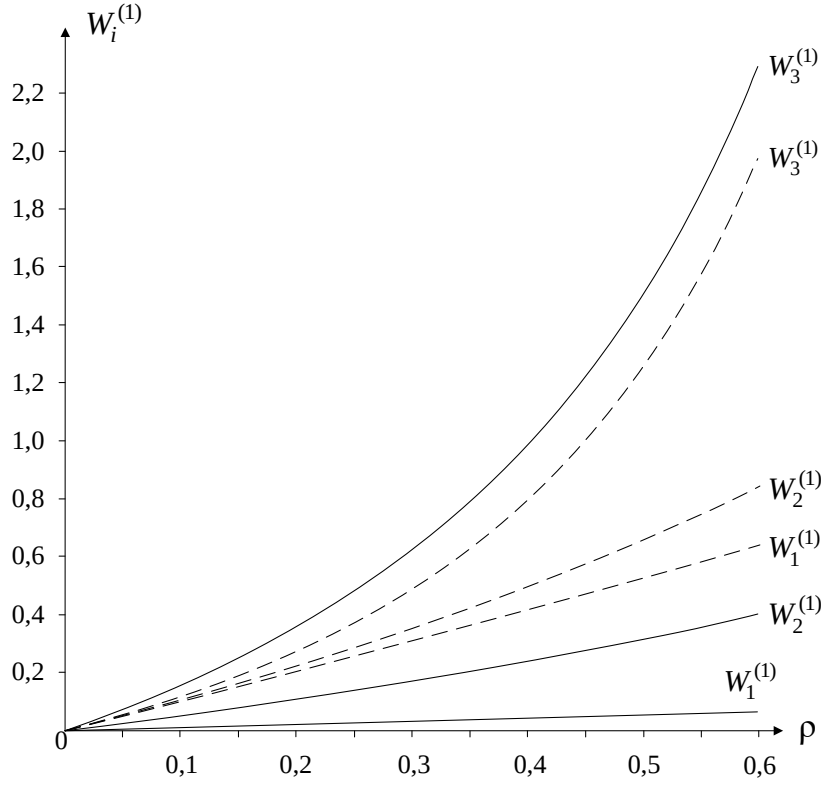


Рис. 4.3. Средние значения $W_i^{(1)}$ для СМО с тремя абсолютными приоритетами

Ход кривых хорошо иллюстрирует интуитивно понятные закономерности, которые характерны для дисциплин обслуживания заявок с абсолютными приоритетами. Заявки высших приоритетов обслуживаются в среднем быстрее. Этот эффект компенсируется ростом среднего значения длительности ожидания в очереди для заявок, которым был присвоен низкий приоритет. Очевидно, что выбор типов приоритетов должен быть сделан с учетом показателей качества обслуживания заявок разного рода. При решении подобных задач приходится учитывать и ряд других факторов.

Контрольные вопросы и дополнительные задания

I. Используя (4.7), проанализируйте СМО с двумя приоритетами, в которой функция $B_1(t)$ подчиняется экспоненциальному закону распределения, а $B_2(t)$ – закону Эрланга k-го порядка. Постройте график $W^{(1)} = f(k)$ при неименных значениях ρ_1 и ρ_2; дайте объяснение полученной зависимости.

II. Выведите формулу для расчета коэффициента вариации длительности ожидания начала обслуживания заявок i-го приоритета на основании соотношений (4.8) и (4.9).

III. Постройте графики, подобные приведенным на рис. 4.1, при соотношении между величинами λ_1, λ_2 и λ_3, которое задано такими значениями: 60, 30 и 10%. Проведите качественный анализ эффективности приоритетного обслуживания.

Литература к разд. 4

1. *Клейнрок, Л.* Теория массового обслуживания / Л. Клейнрок. – М. : Машиностроение, 1979.

2. *Клейнрок, Л.* Вычислительные системы с очередями / Л. Клейнрок. – М. : Мир, 1979.

3. ITU-D. Teletraffic Engineering Handbook (edited by V.B. Iversen). – Geneva, 2003.

4. Приоритетные системы обслуживания / Б. В. Гнеденко [и др.]. – М. : МГУ, 1973.

5. *Джейсуол, Н.* Очереди с приоритетами / Н. Джейсуол. – М. : Мир, 1973.

6. *Саати, Т. Л.* Элементы теории массового обслуживания и ее приложения / Т. Л. Саати. – М. : Либроком, 2010.

7. *Алиев, Т. И.* Основы моделирования дискретных систем / Т. И. Алиев. – СПб. : СПбГУ ИТМО, 2009.

5. АНАЛИЗ МОДЕЛЕЙ С ВХОДЯЩИМ ПОТОКОМ ПРОИЗВОЛЬНОГО ВИДА

5.1. Система массового обслуживания *G / M / 1*

Эта модель представляет теоретический и практический интерес по двум причинам. Во-первых, она позволяет изучать характеристики систем, на вход которых поступает поток заявок, отличающийся от пуассоновского. Во-вторых, ряд результатов анализа этой модели полезно сопоставить с соотношениями, полученными для систем вида $M/M/1$ и $M/G/1$.

Для получения характеристик модели $G/M/1$ необходимо сначала найти величину υ, которая в области $0<\upsilon<1$ является единственным корнем следующего уравнения:

$$\upsilon = \alpha(\mu - \mu\upsilon). \tag{5.1}$$

Выражения для средних значений длительности ожидания и задержки, а также распределения $W(t)$ и $S(t)$ похожи на аналогичные формулы для модели вида $M/M/1$. Различие состоит в том, что вместо нагрузки ρ следует использовать переменную υ:

$$W^{(1)} = \frac{\upsilon}{\mu(1-\upsilon)}, \tag{5.2}$$

$$S^{(1)} = \frac{1}{\mu(1-\upsilon)}, \tag{5.3}$$

$$W(t) = 1 - \upsilon e^{-(1-\upsilon)\mu t}, \tag{5.4}$$

$$S(t) = 1 - e^{-(1-\upsilon)\mu t}. \tag{5.5}$$

Несложно убедиться, что для пуассоновского входящего потока заявок решением уравнения (5.1) будет $\upsilon = \rho$. Тогда выражения (5.2)–(5.5) совпадают с аналогичными соотношениями, полученными для модели $M/M/1$.

5.2. Основные результаты для модели *G / G / 1*

В этом подразделе анализируются модели более общего вида. В идеале исследуемая СМО представляет собой модель $G/G/1$. Для этой модели в [1] получены кумулянты *n*-го порядка W_n времени ожидания начала обслуживания:

$$W_n = \sum_{k=1}^{\infty} \frac{1}{k} \int_0^{\infty} x^n dF^{(k)}(x). \qquad (5.6)$$

Преобразование Лапласа–Стилтьеса функции $F(x)$, обозначаемое далее как $\varphi(s)$, определяется следующим образом:

$$\varphi(s) = \alpha(-s)\beta(s). \qquad (5.7)$$

Получить кумулянты W_n для СМО вида $G/G/1$ преобразованием выражения (5.6) невозможно. Кроме того, для ряда более простых моделей задача упрощается. В этом разделе кумулянты W_n и некоторые другие характеристики СМО выводятся для одной модели, которая представляет практический интерес для сетей, основанных на технологии передачи и коммутации пакетов.

Предположим, что информация о функции $A(t)$ основана на результатах измерений. Пусть с некоторым периодом квантования, равным τ, определяются соответствующие величины приращений исследуемой функции P_i. Если первое приращение функции $A(t)$ зафиксировано в точке $z\tau$, то ее преобразование Лапласа–Стилтьеса $\alpha(s)$ можно записать в таком виде:

$$\alpha(s) = e^{-z\tau s} \sum_{i=0}^{m} P_i e^{-i\tau s}. \qquad (5.8)$$

Величина m определяет ту точку на оси «Время», в которой обнаружено последнее приращение функции $A(t)$.

Коэффициент вариации длительности интервалов между моментами поступлений соседних заявок C_A может меняться в широких пределах. Его значение не сказывается на дальнейших преобразованиях.

Время обработки пакетов можно считать постоянной величиной, равной $B^{(1)}$. Эту величину удобно определять произведением $l\tau$. Функцию распределения длительности обслуживания заявок $B(t)$ удобно выражать через преобразование Лапласа–Стилтьеса:

$$\beta(s) = e^{-l\tau s}. \qquad (5.9)$$

Такое представление функции $\beta(s)$ не будет вносить заметную ошибку во все дальнейшие вычисления, если величина τ достаточно мала. При необходимости оценка влияния величины τ на результаты расчета исследуемых характеристик СМО может быть выполнена простым путем. Надо взять два значения времени обслуживания: $l_1\tau$ и $l_2\tau$. Вели-

чина l_1 представляет собой результат округления частного от деления $B^{(1)}$ на τ до ближайшего целого значения в меньшую сторону. Тогда $l_2 = l_1 + 1$. Все характеристики, вычисленные для значений $l_1\tau$ и $l_2\tau$, будут определять верхнюю и нижнюю границы для характеристик исследуемой СМО.

Предложенную модель в классификации Кендалла можно обозначить так: $G_S / D / 1$. Символ G_S подчеркивает следующий факт: рассматривается распределение общего вида, но с возможными изменениями лишь в некоторые моменты времени $i\tau$ $\left(i = \overline{0, N}\right)$.

Для данной модели функция $\varphi(s)$, определяемая соотношением (5.7), будет рассчитываться по формуле

$$\varphi(s) = e^{(z-l)\tau s} \sum_{i=0}^{m} P_i e^{i\tau s}. \tag{5.10}$$

В соотношение (5.6) входит k-кратная свертка функции $F(x)$. Для преобразования Лапласа–Стилтьеса эта свертка вычисляется возведением правой части выражения (5.10) в степень k [2]:

$$\left[\varphi(s)\right]^k = e^{k(z-l)\tau s} \sum_{i=0}^{km} q_i(k) e^{i\tau s}. \tag{5.11}$$

Коэффициенты $q_i(k)$ определяются на основании правила возведения ряда в степень [3]:

$$q_i(k) = \begin{cases} P_0^k, \text{ если } i = 0; \\ \dfrac{1}{ip_0} \displaystyle\sum_{j=1}^{i} (jk - i + j) \cdot P_j \cdot q_{i-j}(k), \text{ если } i = \overline{1, mk}. \end{cases} \tag{5.12}$$

Введя функцию $\psi_+(x)$, определяемую соотношением (4.30), выражение для расчета кумулянтов W_n можно представить в таком виде:

$$W_n^C = \tau^n \sum_{k=1}^{\infty} \frac{1}{k} \sum_{i=0}^{km} q_i(k) [i + k(z-l)]^n \psi_+[i + k(z-l)]. \tag{5.13}$$

Кумулянты W_1 и W_2 определяют среднее значение и дисперсию длительности ожидания заявок в очереди для СМО вида $G_S / D / 1$. Среднее значение времени задержки заявок (первый момент) $S^{(1)}$ и дисперсия этой величины σ_S^2 рассчитываются следующим образом:

$$S^{(1)} = W_1 + l\tau, \quad \sigma_S^2 = W_2. \tag{5.14}$$

Для вычисления значений $S^{(1)}$ и σ_S^2 необходимо выбрать верхний предел при суммировании по k, т. е. заменить символ ∞ неким конечным значением M. Величина M определяется видом функции $A(t)$ и значением нагрузки СМО ρ. Для расчета параметра ρ следует предварительно вычислить величины λ и μ:

$$\lambda = \frac{1}{z + \tau \sum_{i=0}^{m} iP_i}, \quad \mu = \frac{1}{l\tau}. \tag{5.15}$$

Если $\rho \geq 0,7$, то целесообразно использовать приближенные оценки, полученные для работы СМО вида $G/G/1$ при большой нагрузке [4]. Следовательно, применение метода, предложенного в этом разделе, будет полезным в таком диапазоне нагрузки: $0 < \rho < 0,7$. Именно для этой области изменения параметра ρ необходимо определить верхний предел суммирования по k. Численный анализ показывает, что при $0 < \rho < 0,7$ достаточно установить $m = 100$. Тогда для любых распределений $A(t)$ ошибки в расчете первых четырех кумулянтов не превысят одного процента. Сравнение результатов расчета $S^{(1)}$ и σ_S^2 в СМО вида $G_S/D/1$ с аналогичными оценками для модели вида $M/D/1$ приведено в [5].

5.3. Оценка квантиля

Кумулянты не позволяют получить ФР длительности задержки заявок в СМО. Тем не менее они полезны для приближенной оценки квантилей распределения t_p.

В рекомендации МСЭ Y.1541 [6] приведена методика оценки 0,999-го квантиля ФР длительности задержки IP-пакетов. Эти IP-пакеты правомерно рассматривать как заявки, поступающие в СМО. Задача, которая была поставлена при разработке рекомендации Y.1541, заключалась в оценке квантиля для ФР длительности задержки заявок между ИПС. Это значит, что адекватной моделью служит СеМО. Подробнее модели СеМО будут рассматриваться в разд. 6. В этом разделе мы ограничимся лишь использованием метода, предложенного МСЭ, для одной СМО.

Метод МСЭ предполагает, что сначала для каждой СМО оцениваются два основных параметра:

- математическое ожидание длительности задержки заявок $S^{(1)}$;
- среднеквадратическое отклонение длительности задержки заявок $\sigma_S(i)$.

Далее вводится допущение, что для анализируемой СМО (например, в результате проведения измерений) получено значение квантиля t_p. Кроме того, для вероятности p определяется величина x_p. Она представляет собой значение квантиля для стандартного нормального распределения [7]. Используя эти параметры, можно вычислить асимметрию исследуемой случайной величины γ [6]:

$$\gamma \approx 6\frac{x_p - \dfrac{t_p - S^{(1)}}{\sigma_S}}{1-\left(x_p\right)^2}. \tag{5.16}$$

Кумулянты времени задержки заявок в СМО позволяют рассчитать величину γ непосредственно. Тогда из (5.16) несложно найти квантиль

$$t_p \approx \sigma_S x_p + S^{(1)} - \frac{\sigma_S \gamma\left(1-x_p^2\right)}{6}. \tag{5.17}$$

Соотношение (5.16), положенное в основу метода МСЭ, является приближенным. Погрешность, обусловленная использованием приближенных формул, в рекомендации Y.1541 не приводится. Следовательно, для применения выражения (5.17) в практической работе необходимо оценить возникающие ошибки.

Для оценки ошибок при расчете квантиля t_p можно использовать такой подход. Для некоторой совокупности распределений вычисляются точные значения квантиля, которые сравниваются с оценкой, полученной из (5.17). При этом в качестве значения p выбирается одна из типичных величин, хотя метод, предложенный МСЭ, ориентирован на нормируемый показатель для ССП $p = 0,999$. В табл. 5.1 приведены значения квантиля x_p для часто используемых значений вероятности p.

Таблица 5.1

Квантили нормального распределения

Вероятность p	0,5	0,9	0,95	0,99	0,999
Квантиль x_p	0	1,282	1,645	2,326	3,090

В табл. 5.2 сведены результаты расчета квантилей по формуле (5.17) и точные значения, взятые из справочника [7]. Вычислены и возникаю-

щие относительные ошибки. В качестве примеров взяты два распределения. Для обоих распределений $S^{(1)} = 1$.

Относительная ошибка оценки квантиля для экспоненциального распределения в диапазоне $0,999 \geq p \geq 0,9$ снижается. Для параболического распределения наблюдается противоположная зависимость: 50%-му квантилю соответствует нулевая ошибка. По мере роста вероятности *p* до уровня 0,999 ошибка монотонно возрастает.

Таблица 5.2

Квантили *t*(*p*) и ошибки их вычисления по формуле (5.17)

Значение вероятности p	Точное значение квантиля t_p	Приближенное значение квантиля t_p	Относительная ошибка
Экспоненциальное распределение при условии, что $S^{(1)} = 1$			
0,5	0,693	0,667	0,038
0,9	2,303	2,497	0,084
0,95	2,996	3,214	0,073
0,99	4,605	4,796	0,041
0,999	6,908	6,939	0,004
Параболическое распределение с параметрами $\alpha = 0$ и $\beta = 2$			
0,5	1,0	1,0	0,000
0,9	1,608	2,282	0,419
0,95	1,729299	2,645	0,530
0,99	1,882194	3,326	0,767
0,999	1,963259	4,090	1,083

Численные оценки, приведенные в табл. 5.2, а также полученные для ряда других распределений, позволяют сделать три важных вывода. Во-первых, соотношение (5.17), позволяющее оценивать квантили распределения длительности задержки на основе метода МСЭ, не следует считать приемлемым для всех видов функции $S(t)$. Во-вторых, предложенный подход позволяет с весьма высокой точностью оценивать квантили ряда распределений при условии, что $S(t) \geq 0,99$. При этом распределение $S(t)$ должно быть определено на всей положительной полуоси параметра «Время», т. е. для конечного значения t всегда справедливо такое неравенство: $S(t) < 1$. В-третьих, для распределений с возможными значениями на ограниченном интервале параметра «Время» соотношение (5.17) дает приемлемые оценки только для квантилей, близких к медиане.

Следовательно, для распределений с возможными значениями на ограниченном интервале параметра «Время» целесообразно разработать другой метод для вычисления квантилей. Не исключено, что весьма про-

дуктивным методом может стать аппроксимация функции $S(t)$ рядом Эджворта [8]. Можно разработать и другие методы расчета, которые позволят сразу же оценивать квантили, минуя аппроксимацию распределения $S(t)$.

Величина относительной ошибки в расчете квантилей t_p определяется видом распределения и его параметрами. Среди этих параметров практический интерес связан с коэффициентом вариации, который зависит от двух важных величин: среднего значения и дисперсии случайной величины. Интерес к коэффициенту вариации объясняется тем, что он не имеет размерности и фактически представляет собой нормированную дисперсию.

На рис. 5.1 приведена зависимость относительной ошибки в расчете квантиля t_p от коэффициента вариации времени задержки заявок C_S. Аппроксимацией ФР длительности задержки заявок служит распределение Вейбулла–Гнеденко [4]. Выбор этого распределения обусловлен тем, что для него величина коэффициента вариации изменяется в широких пределах.

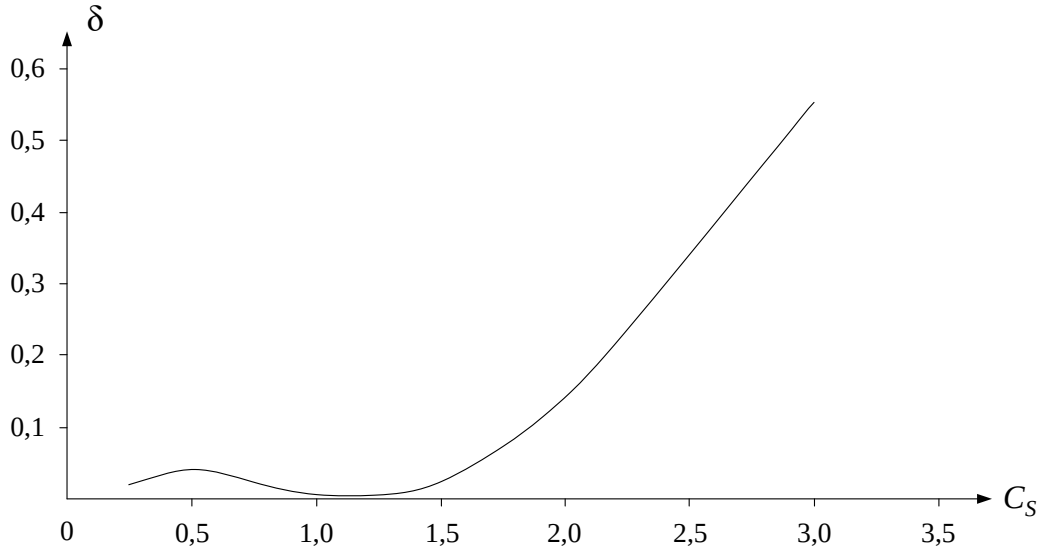


Рис. 5.1. Зависимость относительной ошибки δ от коэффициента вариации

Из графика следует, что в диапазоне изменения коэффициента вариации от нуля до (примерно) двух, относительная ошибка в оценке квантиля остается приемлемой для инженерных расчетов. Для распределений с высоким значением коэффициента вариации целесообразно разработать более точный метод расчета квантиля. Распределения такого рода, как показал ряд измерений трафика в современных телекоммуникационных сетях, характерны для некоторых видов услуг.

5.4. Приближенный анализ СМО вида *G* / *D* / 1

При исследовании некоторых видов СМО (в частности, проектируемых) невозможно провести измерения для получения функции $A(t)$, но относительно ее характера можно предложить обоснованную гипотезу. В подобных случаях функция $A(t)$ определяется одним из известных законов распределения случайных величин [3]. В этом разделе предложен приближенный метод анализа СМО вида $G/D/1$ на основе результатов, полученных в [5]. Метод основан на дискретизации функции $A(t)$ с неким периодом τ, что позволяет перейти к модели $G_S/D/1$.

Для описания ступенчатой функции $A(t)$ удобно использовать ее преобразование Лапласа–Стилтьеса – выражение (5.8). Очевидно, что точность результатов исследования СМО при замене распределения $A(t)$ ступенчатой функцией будет зависеть от величины τ. С этой точки зрения рассматриваемая задача похожа на выбор периода дискретизации аналогового сигнала. Для определения требований к величине τ следует провести анализ нескольких моделей с разными видами функций $A(t)$.

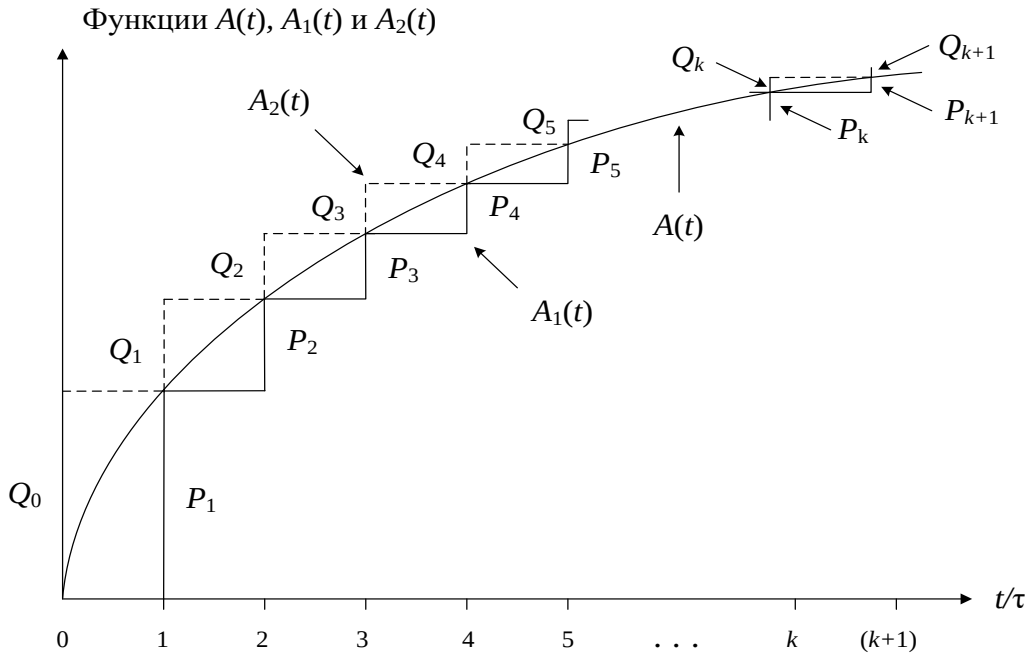


Рис. 5.2. Две ступенчатые функции, заменяющие распределение $A(t)$

В теории телетрафика часто используется модель с пуассоновским входящим потоком заявок с интенсивностью λ, для которого справедливо соотношение (1.8). На рис. 5.2 показано распределение $A(t)$ и две ступенчатые функции $A_1(t)$ и $A_2(t)$. Они позволяют оценивать верхнюю и нижнюю границы ряда параметров для систем с пуассоновским вхо-

дящим потоком. В точке $k\tau$ приращения функций $A_1(t)$ и $A_2(t)$ определяются величинами P_k и Q_k соответственно. Заметим, что $P_0 = 0$.

Величины P_k и Q_k для распределений $A_1(t)$ и $A_2(t)$ вычисляются по следующим формулам:

$$P_k = (e^{\lambda\tau} - 1)e^{-k\lambda\tau}, \quad Q_k = (e^{\lambda\tau} - 1)e^{-(k+1)\lambda\tau}. \qquad (5.18)$$

Преобразования Лапласа–Стилтьеса этих двух ступенчатых функций $\alpha_1(s)$ и $\alpha_2(s)$ связаны между собой простым соотношением:

$$\alpha_2(s) = \alpha_1(s)e^{\tau s}. \qquad (5.19)$$

Введем также функцию $A_3(t)$, которая получается из ступенчатых функций $A_1(t)$ или $A_2(t)$ их смещением по оси абсцисс на $0{,}5\tau$ влево или вправо соответственно. Функции, подобные $A_3(t)$, получаются при выборе аппроксимирующей зависимости методом наименьших квадратов [9]. Преобразование Лапласа–Стилтьеса третьей функции $\alpha_3(s)$ удобно выражать через формулы для изображений $\alpha_1(s)$ или $\alpha_2(s)$:

$$\alpha_3(s) = \alpha_1(s)e^{0{,}5\tau s} = \alpha_2(s)e^{-0{,}5\tau s}. \qquad (5.20)$$

После подстановки выражений для расчета приращений P_k и Q_k функции $\alpha_1(s)$, $\alpha_2(s)$ и $\alpha_3(s)$ определяются как суммы членов геометрической прогрессии [9]:

$$\alpha_1(s) = \frac{\left(1 - e^{-\lambda\tau}\right)e^{-s\tau}}{1 - e^{-(\lambda+s)\tau}}; \quad \alpha_2(s) = \frac{1 - e^{-\lambda\tau}}{1 - e^{-(\lambda+s)\tau}}; \quad \alpha_3(s) = \frac{\left(1 - e^{-\lambda\tau}\right)e^{-0{,}5\tau s}}{1 - e^{-(\lambda+s)\tau}}. \qquad (5.21)$$

По правилу нахождения моментов случайной величины k-го порядка из формул для изображений ФР можно определить средние значения длительности интервалов между поступлениями заявок $A_1^{(1)}$, $A_2^{(1)}$ и $A_3^{(1)}$:

$$A_1^{(1)} = \frac{\tau}{1 - e^{-\lambda\tau}}, \quad A_2^{(1)} = \frac{\tau e^{-\lambda\tau}}{1 - e^{-\lambda\tau}}, \quad A_3^{(1)} = \frac{\tau\left(1 + e^{-\lambda\tau}\right)}{2\left(1 - e^{-\lambda\tau}\right)}. \qquad (5.22)$$

Это же правило позволяет определить второй момент, необходимый для расчета дисперсии. Как и следовало ожидать, для всех трех ступенчатых функций значения дисперсии (σ_1^2, σ_2^2 и σ_3^2) идентичны. По этой

причине нижний индекс можно опустить, используя привычное обозначение σ^2:

$$\sigma^2 = \frac{\tau^2 e^{-\lambda\tau}}{\left(1-e^{-\lambda\tau}\right)^2} = \frac{\tau^2}{4sh\left(0,5\lambda\tau\right)}. \qquad (5.23)$$

Для оценки квантилей ряда распределений используется коэффициент асимметрии Sk [6]. Следуя правилам расчета данного показателя [7] для трех рассматриваемых распределений, можно убедиться, что он одинаков:

$$Sk = e^{0,5\lambda\tau} + e^{-0,5\lambda\tau}. \qquad (5.24)$$

Выражения (5.22) и (5.23) позволяют найти коэффициенты вариации исследуемой случайной величины C_1, C_2 и C_3:

$$C_1 = e^{-0,5\lambda\tau}, \quad C_2 = e^{0,5\lambda\tau}, \quad C_3 = \frac{2}{e^{0,5\lambda\tau} + e^{-0,5\lambda\tau}}. \qquad (5.25)$$

При $\tau \to 0$ все три коэффициента вариации стремятся к единице. Можно показать, что при $\tau \to 0$ $A_1^{(1)}$, $A_2^{(1)}$ и $A_3^{(1)}$ стремятся к λ^{-1}, σ^2 – к λ^{-2}, C_1, C_2 и C_3 – к единице, а Sk – к двум. Значения λ^{-1}, λ^{-2}, «1» и «2» свойственны соответственно математическому ожиданию, дисперсии, коэффициентам вариации и асимметрии случайной величины, которая подчиняется экспоненциальному закону распределения.

Для значений τ, отличных от нуля, существует относительная ошибка в оценке моментов случайной величины, которая обусловлена аппроксимацией распределения $A(t)$ ступенчатой функцией. Для функций $A_1(t)$, $A_2(t)$ и $A_3(t)$ эти ошибки обозначаются так: $\delta_1(j)$, $\delta_2(j)$ и $\delta_3(j)$. Переменная j идентифицирует ту характеристику случайной величины, для которой рассчитывается относительная ошибка.

Проще всего найти ошибки $\delta_1(C)$, $\delta_2(C)$ и $\delta_3(C)$ для коэффициентов вариации, так как точное значение этой характеристики для рассматриваемого распределения $A(t)$ равно единице. Несложно показать, что минимальной ошибкой является величина $\delta_3(C)$. Формулы для расчета ошибок $\delta_1(C)$, $\delta_2(C)$ и $\delta_3(C)$ представимы в следующей форме:

$$\delta_1(C) = 1 - e^{-0,5\lambda\tau}; \quad \delta_2(C) = 1 - e^{0,5\lambda\tau}; \delta_3(C) = 1 - \frac{2}{e^{0,5\lambda\tau} + e^{-0,5\lambda\tau}}. \qquad (5.26)$$

Для достаточно малых значений τ величина относительной ошибки оценивается слагаемым $0{,}5\lambda\tau$. Для доказательства этого утверждения функции $\delta_1(C)$, $\delta_2(C)$ и $\delta_3(C)$ следует разложить в ряд Маклорена [9]. Если задан уровень допустимой относительной ошибки для оценки коэффициента вариации $\delta(C)$ и известен параметр λ, то для выбора величины τ справедливо неравенство

$$\tau \le \frac{2\delta(C)}{\lambda}. \tag{5.27}$$

Определение требований к величине τ с учетом допустимой относительной ошибки при оценке средних значений исследуемой случайной величины связано с более громоздкими преобразованиями. После их выполнения несложно получить неравенство вида (5.27) с одним отличием. Вместо сомножителя $\delta(C)$ в нем будет фигурировать величина допустимой относительной ошибки для оценки среднего значения длительности интервала между поступлениями соседних заявок $\delta(A)$.

Величина коэффициента вариации представляет собой нормированную дисперсию. Это означает, что при использовании аппроксимаций (5.21) для расчета двух основных характеристик исследуемой случайной величины (среднего значения и дисперсии) период τ выбирается из неравенства

$$\tau \le \frac{2\delta}{\lambda}. \tag{5.28}$$

Множитель δ в данном случае целесообразно рассматривать как допустимый уровень относительной ошибки при оценке среднего значения и дисперсии случайной величины, распределение которой подчиняется экспоненциальному закону. Несложно убедиться, что условие (5.28) позволяет оценивать с заданной точностью и коэффициент асимметрии.

Исследование возникающих ошибок для функций $A(t)$, заданных на ограниченном интервале времени, можно провести на примере равномерного закона распределения длительности интервалов между моментами поступления заявок [7]. На рис. 5.3 показано такое распределение $A(t)$ на нормированном к τ интервале времени от 0 до $(n-1)$. Ступенчатая функция $A_3(t)$ имеет одинаковые приращения в точках $k\tau$, равные n^{-1}.

Для функции $A(t)$ среднее значение длительности интервалов между моментами поступления заявок $A^{(1)}$, дисперсия σ^2, коэффициент вариации C определяются по хорошо известным формулам [7]:

$$A^{(1)} = \frac{n\tau}{2}, \quad \sigma^2 = \frac{n^2\tau^2}{12}, \quad C = \frac{1}{\sqrt{3}}. \tag{5.29}$$

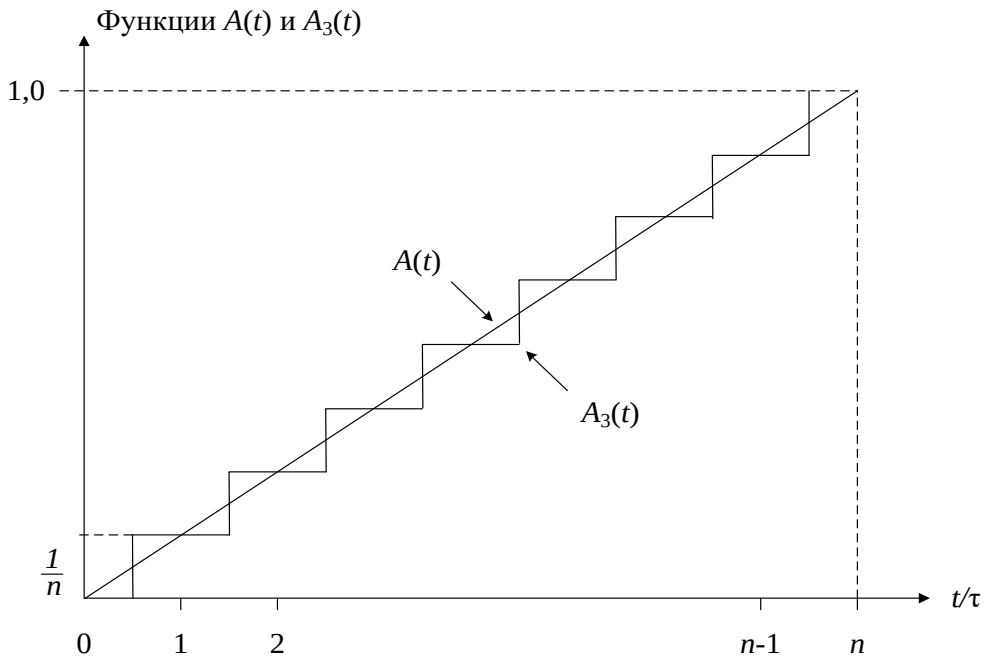


Рис. 5.3. Равномерное распределение и его аппроксимация

В качестве ступенчатой функции $A(t)$ целесообразно использовать распределение, которое соответствует третьему виду функции, введенному в предыдущем разделе. Для ступенчатой функции вида $A_3(t)$ исследуемые характеристики (они снабжены нижним индексом «3») вычисляются по формулам [7]:

$$A_3^{(1)} = \frac{(n-1)\tau}{2}; \qquad \sigma_3^{(2)} = \frac{(n^2-1)\tau^2}{12}; \qquad C_3 = \frac{\sqrt{n^2-1}}{\sqrt{3}(n-1)}. \qquad (5.30)$$

Коэффициенты асимметрии для обоих распределений равны нулю. По этой причине соответствующие ошибки далее не рассматриваются. Относительная ошибка при оценке среднего значения исследуемой случайной величины $\delta(A)$ определяется следующим образом:

$$\delta(A) = \frac{A^{(1)} - A_3^{(1)}}{A^{(1)}} = 0,5\lambda\tau. \qquad (5.31)$$

Это означает, что величина τ должна выбираться из (5.28), которое было получено для другого закона распределения исследуемой случайной величины. Можно показать, что соблюдение условия (5.28) приводит к выбору меньшего значения τ, чем те неравенства, из которых вычисляются ошибки для σ^2 или C. Иными словами, для равномерного

закона распределения $A(t)$ выбор величины τ должен осуществляться на основе (5.28).

Третий пример связан с распределениями, для которых коэффициент вариации длительности интервалов между моментами поступления заявок превышает единицу. На профессиональном сленге они называются распределениями с «тяжелыми хвостами» [10]. Одним из удачных примеров такого распределения (с точки зрения простоты последующего анализа) следует считать гиперэкспоненциальное. Для гиперэкспоненциального распределения второго порядка [7] с параметром формы p преобразование Лапласа–Стилтьеса ФР длительности интервалов между поступлениями заявок $\alpha(s)$ представимо в таком виде:

$$\alpha(s) = p\frac{2p\lambda}{s+2p\lambda} + (1-p)\frac{2(1-p)\lambda}{s+2(1-p)\lambda}. \qquad (5.32)$$

Математическое ожидание времени между моментами поступления заявок $A^{(1)}$ связано с величиной интенсивности потока заявок λ соотношением (1.6). При замене гиперэкспоненциального распределения дискретным с отсчетами, взятыми с периодом τ, целесообразно использовать функцию вида $A_3(t)$. Для нее среднее значение времени между моментами поступления заявок $A_3^{(1)}$ может быть получено виде

$$A_3^{(1)} = \frac{\tau p\left(1+e^{-2p\lambda\tau}\right)}{2\left(1-e^{-2p\lambda\tau}\right)} + \frac{\tau(1-p)\left(1+e^{-2(1-p)\lambda\tau}\right)}{2\left(1-e^{-2(1-p)\lambda\tau}\right)}. \qquad (5.33)$$

Относительную ошибку оценки $A^{(1)}$ целесообразно исследовать как функцию $f(p,\lambda,\tau)$, которая зависит от трех параметров p, λ и τ:

$$f(p,\lambda,\tau) = 1-\lambda\tau\frac{p\left(1+e^{-2p\lambda\tau}\right)\left(1-e^{-2(1-p)\lambda\tau}\right)+(1-p)\left(1+e^{-2(1-p)\lambda\tau}\right)\left(1-e^{-2p\lambda\tau}\right)}{2\left(1-e^{-2p\lambda\tau}\right)\left(1-e^{-2(1-p)\lambda\tau}\right)}. \qquad (5.34)$$

На рис. 5.4 приведены графики функций $f(p,\lambda,\tau)$, построенные для трех значений τ в диапазоне изменений параметра формы p от 0,01 до 0,49 при $\lambda = 1$. Выбор такого диапазона обусловлен тем, что распределение вида (5.32) определено при условии, что $0 < p < 0,5$.

Для каждой кривой указаны также значения относительной ошибки, вычисленные на основании соотношения (5.28): $\delta = 0,5\lambda\tau$. Очевидно, что $f(p,\lambda,\tau) << \delta$. Данный факт объясняется характером изменений

функции $A(t)$. Распределения с «тяжелыми хвостами» с ростом t приближаются к единице не столь быстро, как, например, экспоненциальное. По этой причине отсчеты, взятые с периодом τ, который вычислен по (5.28), более точно представляют ФР. Из этого следует вывод о том, что (5.28) вполне приемлемо для распределений с «тяжелыми хвостами». Можно показать, что данный вывод справедлив и для оценок других параметров ФР длительности интервалов между моментами поступления заявок. Соответствующие формулы очень громоздки и в этом учебном пособии не приводятся.

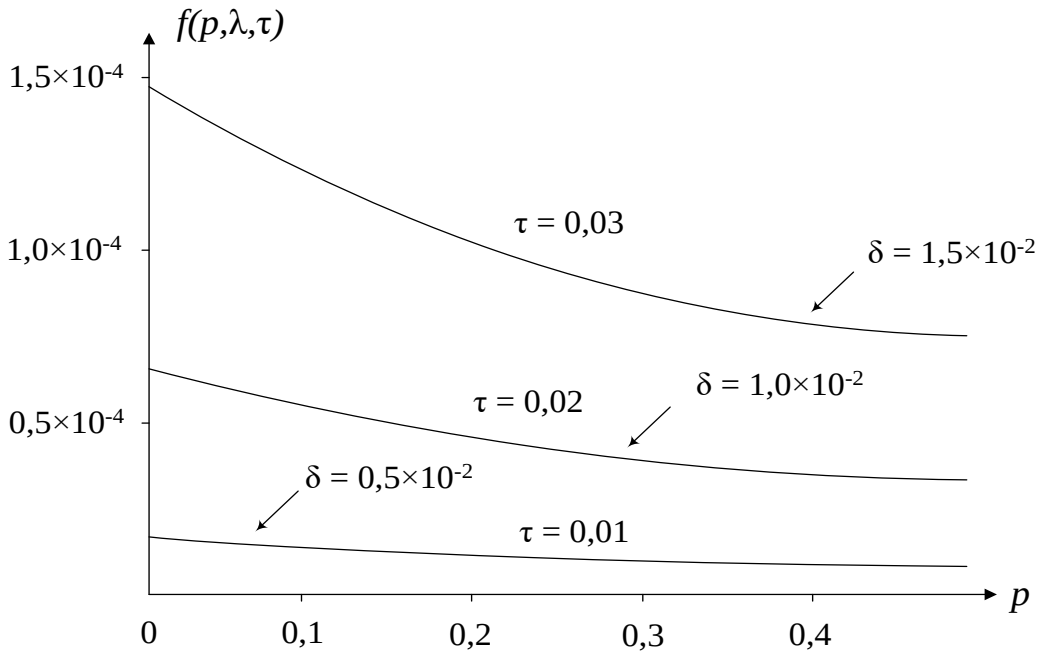


Рис. 5.4. Зависимость относительной ошибки $f(p,\lambda,\tau)$ от параметра формы p

Возможность исследования СМО посредством дискретизации ФР длительности интервалов между моментами поступления заявок, для которой величина коэффициента вариации превышает (иногда – весьма существенно) единицу, позволяет анализировать процессы функционирования мультисервисных сетей. Именно для сетей такого рода характерны распределения с «тяжелыми хвостами» [10].

Сформулированное условие для выбора величины τ приемлемо для получения численных характеристик функции $A(t)$. Необходимо определить подобное условие для моментов времени задержки. С практической точки зрения интерес представляют среднее значение времени задержки $S^{(1)}$ и дисперсия этой случайной величины σ_S^2. Вместо дисперсии можно использовать также коэффициент вариации C_S.

Если функция $A(t)$ описывается законом (1.8), то рассматриваемой моделью служит система вида $M/D/1$. Для нее известны выражения, позволяющие рассчитать $S^{(1)}$, σ_S^2 и C_S. Они приведены, например, в [4]. Для компактной записи этих выражений следует использовать параметр нагрузки ρ.

С учетом (5.9) интенсивность обслуживания заявок будет равна $(l\tau)^{-1}$. Тогда формулы для вычисления характеристик $S^{(1)}$, σ_S^2 и C_S можно представить в следующей редакции:

$$S^{(1)} = \frac{2-\rho}{2[1-\rho]} l\tau, \qquad \sigma_S^2 = \frac{\rho(4-\rho)}{12(1-\rho)^2}(l\tau)^2, \quad C_S = \frac{\sqrt{\rho(4-\rho)}}{\sqrt{3}(2-\rho)}. \qquad (5.35)$$

При использовании ступенчатых функций для аппроксимации распределения $A(t)$ оценки $S^{(1)}$, σ_S^2 и C_S приведены выше. Эти оценки вместе с результатами, полученными при расчетах по (5.35), позволяют определить то максимальное значение τ, для которого относительная ошибка в оценке исследуемых характеристик времени задержки заявок не превышает установленный порог δ.

Проведенные вычисления показали, что величина τ зависит от нагрузки ρ. Для задач проектирования технических средств, исследуемых как СМО, интересен диапазон нагрузки от 0,1 до 0,7. Именно для таких значений ρ целесообразно оценить величину относительной ошибки, возникающей при расчетах параметров времени задержки заявок. В перечень исследуемых параметров этой случайной величины целесообразно включить среднее значение и коэффициент вариации. Эти два параметра часто используются для оценки показателей качества обслуживания в СМО.

В формулы (5.35) входит величина нагрузки ρ, а неравенство (5.28) получено для интенсивности потока заявок λ. Для дальнейшего анализа целесообразно предположить, что время обслуживания заявок равно единице. Тогда значения ρ и λ будут численно совпадать.

Результаты вычисления ошибок, возникающих при оценке исследуемых параметров времени задержки заявок, приведены в табл. 5.3. Величина τ определялась из (5.28) при условии, что $\delta = 2{,}5\%$.

Таблица 5.3

Относительные ошибки в оценке параметров времени задержки заявок

Ошибка при расчете	$\rho = 0{,}1$	$\rho = 0{,}3$	$\rho = 0{,}5$	$\rho = 0{,}7$
Среднее значение	0,010	0,002	0,001	0,011
Коэффициент вариации	0,009	0,006	0,003	0,015

Данные, приведенные в табл. 5.3, свидетельствуют, что в выбранном диапазоне изменений нагрузки ошибки оценки двух исследуемых параметров времени задержки заявок не превышают порога $\delta = 2{,}5\%$. Это значение допустимой относительной ошибки было использовано для выбора периода дискретизации функции $A(t)$. Следовательно, условие (5.28) можно считать достаточным и для приближенного анализа параметров времени задержки заявок в однолинейных СМО с постоянным временем обслуживания заявок.

Контрольные вопросы и дополнительные задания

I. Вычислите среднее значение длительности задержки заявок для моделей двух видов: $M/E_2/1$ и $E_2/M/1$. Проанализируйте характер изменений этой характеристики для обеих моделей при разной нагрузке СМО.

II. Оцените точность получения среднего значения длительности ожидания заявок в очереди в СМО вида $G_S/D/1$ для времени обслуживания, равного $l_1\tau$ и $l_2\tau$. Используйте аргументы, приведенные после формулы (5.9).

III. Попробуйте объяснить следующий факт: почему в табл. 5.2 не фигурирует нагрузка.

IV. Получите оценки, которые аналогичны приведенным в табл. 5.2, для других видов распределений, перечисленных, например, в [7].

V. Попытайтесь найти аппроксимацию распределений Эрланга второго порядка при помощи ряда Эджворта. Найдите ошибки в оценке квантилей.

VI. Найдите сходства и различия между методами дискретизации функции $A(t)$, которые предложены в подразд. 5.4, и преобразования аналогового сигнала в системах передачи с импульсно-кодовой модуляцией.

Литература к разд. 5

1. *Штойян, Д.* Качественные свойства и оценки стохастических моделей / Д. Штойян. – М. : Мир, 1979.

2. *Диткин, В. А.* Интегральные преобразования и операционное исчисление / В. А. Диткин, А. П. Прудников. – М. : Наука, 1974.

3. *Градштейн, И. С.* Таблицы интегралов, сумм, рядов и произведений / И. С. Градштейн, И. М. Рыжик. – М. : Наука, 1971.

4. *Клейнрок, Л.* Вычислительные системы с очередями / Л. Клейнрок. – М. : Мир, 1979.

5. *Соколов, А. Н.* Метод оценки задержки IP пакетов в узле коммутации / А. Н. Соколов // Научно-технические ведомости СПбГПУ. – 2009. – № 4 (82).

6. ITU-T. Network Performance Objectives for IP-Based Services. Recommendation Y.1541. – Geneva, 2006.

7. *Вадзинский, Р. Н.* Справочник по вероятностным распределениям / Р. Н. Вадзинский. – СПб. : Наука, 2001.

8. *Крамер, Г.* Математические методы статистики / Г. Крамер. – М. : Мир, 1975.

9. *Корн, Т.*Справочник по математике для научных работников и инженеров / Т. Корн, Г. Корн. – М. : Наука, 1984.

10. *Шелухин, О. И.* Моделирование информационных систем / О. И. Шелухин, А. М. Тенякшев, А. В. Осин. – М. : Радиотехника, 2005.

11. *Соколов, А. Н.* Приближенный метод анализа однолинейных систем массового обслуживания с постоянным временем обработки заявок / А. Н. Соколов // Проблемы информатики. – 2010. – № 3.

6. СЕТИ МАССОВОГО ОБСЛУЖИВАНИЯ

6.1. Модель сети массового обслуживания

Некоторые задачи, решаемые методами теории телетрафика, связаны с моделями СеМО. Типичным примером задач такого рода служит оценка характеристик качества обслуживания между двумя ИПС. Обычно между ними располагаются несколько узлов коммутации, каждый из которых может быть представлен в виде СМО.

Практический интерес связан с характеристиками качества обслуживания между парой ИПС, для которых известна совокупность используемых СМО. Эта совокупность, как правило, задается правилами построения технических систем, анализируемых при помощи методов теории телетрафика. Например, при оценке характеристик качества обслуживания вызовов в ТФОП количество используемых СМО вычисляется на основе правил маршрутизации для установления местных, междугородных и международных соединений.

На рис. 6.1 приведена гипотетическая модель СеМО. Ее можно рассматривать как модель, иллюстрирующую принципы установления междугородного соединения в ТФОП. В этом случае СМО1 и СМО5 соответствуют тем узлам коммутации, в которые включены терминалы абонентов, участвующих в организации связи. СМО2, СМО3 и СМО4 – модели тех транзитных узлов, через которые может быть установлено требуемое соединение.

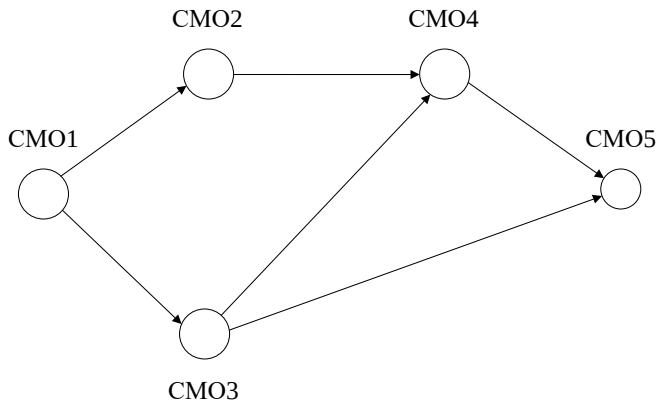


Рис. 6.1. Модель сети массового обслуживания

На вход i-й СМО (для рассматриваемой модели $i = \overline{1, 5}$) поступает поток заявок с интенсивностью λ_i. Для многих практических приложений интересны так называемые линейные СеМО. Условие линейности состоит в том, что величины λ_k и λ_l связаны между собой посредством коэффициента p_{kl}:

$$\lambda_k = p_{kl}\lambda_l. \tag{6.1}$$

Заметим, что в общем случае $p_{kl} \neq p_{lk}$. Коэффициенты p_{kl} образуют матрицу. Для модели, изображенной на рис. 6.1, часть коэффициентов p_{kl} равна нулю:

$$\begin{Vmatrix} 0 & p_{12} & p_{13} & 0 & 0 \\ 0 & 0 & 0 & p_{24} & 0 \\ 0 & 0 & 0 & p_{34} & p_{35} \\ 0 & 0 & 0 & 0 & p_{45} \\ 0 & 0 & 0 & 0 & 0 \end{Vmatrix}. \qquad (6.2)$$

Нелинейный характер связи величин λ_k и λ_l может быть обусловлен различными причинами. Например, некоторые заявки, попадающие в i-й узел, порождают сложные процессы, нарушающие соотношение (6.1).

Принято также различать три вида СеМО: открытые (разомкнутые), замкнутые и комбинированные. В этом разделе рассматриваются только открытые СеМО, для которых суммарное количество заявок (обслуживаемых и находящихся в очереди) нельзя считать постоянной величиной. Это свойство обусловлено возможностью поступления в каждую СМО заявок от внешних источников нагрузки. Далее рассматриваются однородные СеМО, обслуживающие заявки одного типа. Кроме того, предполагается, что заявкам не назначается приоритет и используются дисциплины обслуживания класса FIFO.

СеМО состоит из N элементов. Каждый элемент – отдельная СМО. Предполагается, что длительность перехода заявки из одной СМО в другую равна нулю. Если гипотеза такого рода не может быть принята, то исследуемую модель приходится усложнять. Типичный пример – необходимость учета времени распространения сигнала между двумя узлами коммутации. Это время обычно постоянно. Его можно учесть за счет введения дополнительной СМО вида $M/D/\infty$. Эта модель предполагает, что для каждой заявки может быть сразу же выделено обслуживающее устройство. Следовательно, длительность задержки заявки в СМО $S^{(1)}$ равна времени ее обслуживания $B^{(1)}$, а оно для модели $M/D/\infty$ неизменно.

Ряд полезных моделей СеМО рассматривается, например, в [2–8]. Можно найти и другие монографии, посвященные соответствующему разделу теории телетрафика.

6.2. Основные результаты анализа простейших СеМО

Решение практических задач, использующих модели СеМО, зачастую связано с нахождением трех показателей: вероятности отказа в обслуживании, среднего значения длительности задержки заявок и квантиля одноименной ФР. Это утверждение основано на правилах нормирования качественных показателей, принятых МСЭ и ETSI. Вернемся к модели, показанной на рис. 6.1. Пусть показатели качества обслуживания трафика нормированы для ИПС, расположенных за СМО под номерами 1 и 5. Это значит, что все нормы должны выполняться при использовании любого из возможных путей прохождения заявок из СМО1 в СМО5. Будем считать, что максимальные величины вероятности отказа в обслуживании и задержки заявок наблюдаются для маршрута СМО1–СМО2–СМО4–СМО5. Для этого маршрута, содержащего m СМО ($m = 4$), должны выполняться три показателя качества обслуживания трафика:

- вероятность потери заявок $P(m)$;
- среднее значение длительности задержки заявок $S^{(1)}(m)$;
- квантиль ФР длительности задержки заявок $t_p(m)$.

Если процессы потери заявок на всех фазах обслуживания взаимно независимы, то величина $P(m)$ вычисляется по известным вероятностям потери заявок во всех СМО π_k:

$$P(m) = 1 - \prod_{k=1}^{m} (1 - \pi_k). \tag{6.3}$$

Еще проще – в силу аддитивности математического ожидания – оценивается среднее значение длительности задержки заявок. Для этого надо знать аналогичные величины для каждой k-й СМО $S_k^{(1)}$:

$$S^{(1)}(m) = \sum_{k=1}^{m} S_k^{(1)}. \tag{6.4}$$

К сожалению, вычисление квантиля $t_p(m)$ осуществляется более сложным образом. Сначала следует найти ФР длительности задержки заявок для маршрута, состоящего из m СМО $S(m,t)$. Пусть известны аналогичная ФР для каждой k-й СМО $S_k(t)$ и ее преобразование Лапласа–Стилтьеса $\xi_k(s)$. В этом случае при условии независимости задержек на каждой фазе обслуживания справедливо следующее соотношение:

$$\xi(m,s)=\prod_{k=1}^{m}\xi_k(s). \tag{6.5}$$

Распределение $S(m,t)$ определяется из своего изображения $\xi(m,s)$. Квантиль $t_p(m)$ находится численно в результате решения такого уравнения:

$$S\left(m,t_p(m)\right)=p. \tag{6.6}$$

Вычисления по (6.3) и (6.4) не представляются сложными задачами. Иная ситуация складывается с оценкой квантиля, а точнее – с получением распределений вида $S(m,t)$. Ниже приводятся два примера, иллюстрирующие способы получения искомой функции.

Первый пример связан с так называемой экспоненциальной СеМО, в которой каждая фаза обслуживания может быть представлена моделью $M/M/1$. Будем считать, что известны величины интенсивности входящего потока заявок λ_k и их обслуживания μ_k в каждой СМО. Это позволяет записать преобразование Лапласа–Стилтьеса $\xi(m,s)$ в такой редакции:

$$\xi(m,s)=\prod_{k=1}^{m}\frac{\mu_k-\lambda_k}{s+(\mu_k-\lambda_k)}. \tag{6.7}$$

Далее целесообразно рассматривать частный случай, когда параметры λ_k и μ_k для всех СМО одинаковы. Это позволяет опустить индекс k и трансформировать (6.7):

$$\xi(m,s)=\left[\frac{\mu-\lambda}{s+(\mu-\lambda)}\right]^m. \tag{6.8}$$

Используя таблицы преобразования Лапласа [9], несложно получить распределение $S(m,t)$:

$$S(m,t)=1-e^{-(\mu-\lambda)t}\sum_{k=0}^{m-1}\frac{\left[(\mu-\lambda)t\right]^{m-k-1}}{(m-k-1)!}. \tag{6.9}$$

Если величины λ и μ на всех фазах будут различны, то формула для вычисления функции $S(m,t)$ становится более громоздкой, но ее вывод не вызывает затруднений. В ряде случаев выражение (6.9) целесообразно представлять в другой форме, используя отношение λ к μ, т. е. величину ρ:

$$S(m,t) = 1 - e^{-(1-\rho)\mu t} \sum_{k=0}^{m-1} \frac{\left[(1-\rho)\mu t\right]^{m-k-1}}{(m-k-1)!}. \qquad (6.10)$$

Характер функции $S(m,t)$ зависит и от нагрузки СМО и количества систем в маршруте. На рис. 6.2 показано изменение плотности исследуемого распределения – производной от функции $S(m,t)$. Поведение плотности ФР нагляднее иллюстрирует различие между возможными задержками при разном количестве фаз обслуживания. При вычислении производных были приняты два условия: $\rho = 0,5$ и $\mu = 1$.

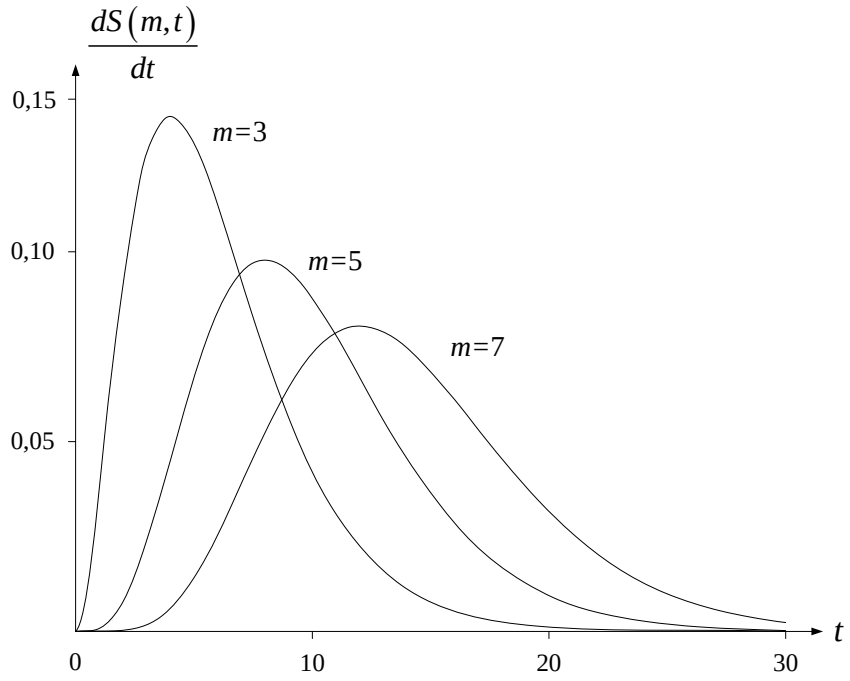


Рис. 6.2. Производные функции $S(m,t)$ для экспоненциальной СеМО

Характер кривых, изображенных на рис. 6.2, показывает, что при росте числа фаз обслуживания огибающая плотности распределения меняет свой характер. Она все более становится похожа на плотность нормального закона распределения случайной величины. По всей видимости, эта закономерность послужила основанием для разработки метода оценки квантиля, который рассматривался в разд. 5.

При получении распределений $S_k(t)$ в процессе измерений приходится оперировать со случайной величиной, распределенной на конечном интервале времени. Рассмотрим случай равномерного распределения на отрезке $(t_{\text{MIN}}, t_{\text{MAX}})$. Можно считать, что $t_{\text{MIN}} = 0$. Для всех других значений t_{MIN} исследуемое распределение можно сдвинуть по оси «Вре-

мя» так, чтобы выполнялось условие: $t_{\text{MIN}} = 0$. В этом случае вместо значения t_{MAX} следует использовать разность $t_{\text{MAX}} - t_{\text{MIN}}$.

Предположим, что заявки проходят через m фаз с равными значениями величины t_{MAX}, обозначаемой далее через x. Очевидно, что тогда индекс k далее можно опустить. Необходимо найти распределение исследуемой величины mx. Для каждой СМО распределение длительности задержки заявок определяется преобразованием Лапласа–Стилтьеса такого вида [10]:

$$\xi(s) = \frac{1 - e^{-xs}}{xs}. \tag{6.11}$$

Для нахождения ФР суммы случайных величин следует воспользоваться правилом свертки изображений [9, 10]:

$$\xi(m,s) = \frac{\sum_{i=0}^{m} (-1)^i C_m^i e^{-isx}}{(xs)^m}. \tag{6.12}$$

Следует подчеркнуть, что применение данного правила для рассматриваемой модели связано с допущением о взаимной независимости процессов, происходящих во всех компонентах исследуемой СеМО. Для экспоненциальной СеМО существует строгое доказательство правомерности использования правила свертки [1].

Оригинал от правой части выражения (6.12) на основании теоремы смещения [9, 10] представим в такой форме:

$$S(m,t) = \frac{1}{m!\,x^m} \begin{cases} t^m, \quad \text{при } 0 \le t < x; \\ t^m - C_m^1 (t-x)^m, \quad \text{при } x \le t < 2x; \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ \sum_{i=0}^{m-1} (-1)^i C_m^i (t - ix)^m, \quad \text{при } (m-1)x \le t < mx; \\ m!\,x^m, \quad t \ge mx. \end{cases} \tag{6.13}$$

Если величины t_{MAX} для всех фаз не идентичны, то оригинал функции $\xi(m,s)$ будет представлен выражением более громоздкого вида. Качественный характер кривых $S(m,t)$ при этом не меняется. Графики плотности распределения приведены на рис. 6.3 для нескольких значе-

ний m – количества тех СМО, через которые проходят заявки. Величина x принята равной единице.

Графики плотности трех распределений менее похожи на аналогичные кривые для нормального закона. Это обусловлено тем, что функция $S(m,t)$ определена на конечном интервале времени. При использовании таких функций могут возникать большие ошибки в оценке квантиля. Данное утверждение иллюстрирует табл. 6.2, хотя в ней содержатся результаты расчета для параболического распределения.

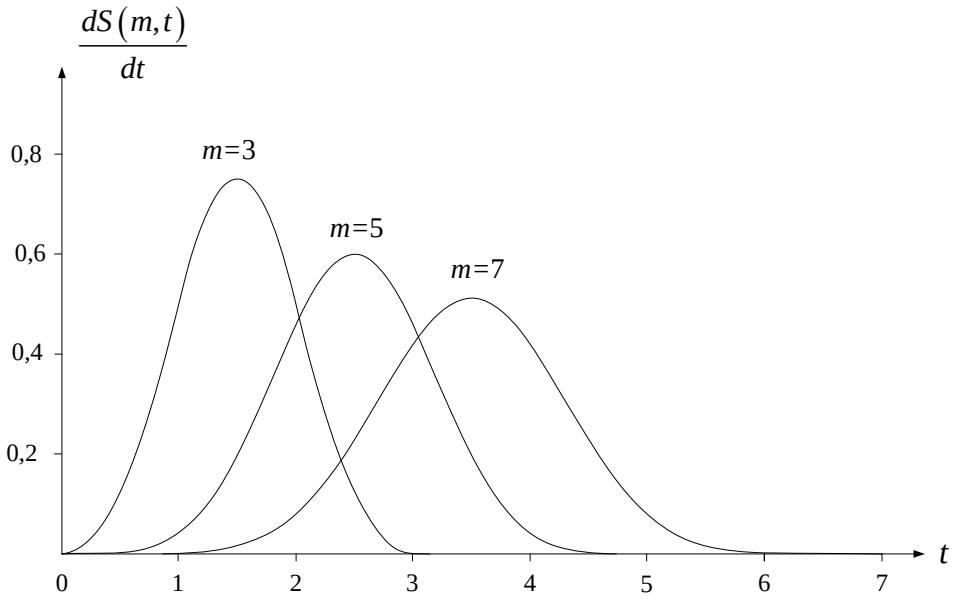


Рис. 6.3. Производные функций $S(m,t)$ для сети, состоящей из СМО вида $M/U/1$

6.3. Некоторые направления исследования СеМО

В этом подразделе кратко перечислены те направления исследования СеМО, которые, по всей видимости, будут полезны для решения задач, относящихся к характеристикам качества обслуживания мультисервисного трафика в сетях следующего поколения. Для задач подобного рода целесообразно выделить пять направлений исследований СеМО.

Во-первых, практический интерес представляет изучение СеМО, начавшееся с работ Дж. Джексона [2]. Исследованные им модели часто называют сетями Джексона. Другое название – открытые сети. Этими словами подчеркивается тот факт, что в СеМО могут поступать заявки от внешних источников трафика. Проще всего исследовать модели, которые называются экспоненциальными. Для них функции $A(t)$ и $B(t)$ можно представить выражениями (1.8) и (1.11) соответственно. В насто-

ящее время практический интерес смещается к моделям СеМО, для которых функции $A(t)$ и $B(t)$ заметно отличаются от экспоненциального закона.

Во-вторых, для некоторых компонентов инфокоммуникационных систем интересны модели СеМО, в которых циркулирует постоянное количество заявок. Подобные сети, как правило, называют замкнутыми. Иногда их именуют сетями Гордона–Ньюелла [2] – по фамилиям авторов, опубликовавших одну из первых работ по исследованию замкнутых СеМО. Первоначально анализировались характеристики экспоненциальных замкнутых сетей. Естественно, что процессы работы современных инфокоммуникационных систем стимулируют изучение замкнутых СеМО более общего вида.

В-третьих, некоторые процессы, протекающие в инфокоммуникационных системах, могут быть исследованы при помощи моделей, называемых СеМО с отрицательными заявками [6]. В ряде публикаций соответствующие модели называют G-сетями, что связано с фамилией известного специалиста в области теории телетрафика – Gelenbe. Он одним из первых начал изучение СеМО с отрицательными заявками. Появление в СеМО отрицательной заявки меняет работу сети специфическим образом. Из сети «уходит» одна обычная (ее называют положительной) заявка. После этого события отрицательная заявка покидает СеМО без обслуживания.

В-четвертых, следует отметить сравнительно новое направление в исследовании СеМО, известное по аббревиатуре BCMP. Это сокращение образовано из первых букв фамилий авторов предложенного метода анализа СеМО: Baskett, Chandy, Muntz, Palacios. Метод BCMP позволяет получить уравнения равновесия (глобального баланса) для сетей, обслуживающих заявки разных классов (приоритетов). При этом сеть включает в себя узлы нескольких типов, которые различаются количеством обслуживающих приборов и дисциплиной обслуживания заявок. Метод BCMP позволяет исследовать модели более сложные, чем сети Джексона.

В-пятых, нельзя не упомянуть об имитационном моделировании – важном методе исследования СеМО. Многие модели, важные с практической точки зрения, невозможно анализировать при помощи только аналитических методов. Кроме того, моделирование следует рассматривать как один из эффективных инструментов проверки результатов, которые получены аналитически за счет введения ряда допущений.

Контрольные вопросы и дополнительные задания

I. Рассмотрите СеМО, в которой заявки проходят через пять систем вида $M/D/1$ с идентичными величинами нагрузки. Воспользуйтесь со-

отношением (6.5), чтобы получить формулы для среднего значения длительности задержки заявок и для дисперсии этой случайной величины. Величину μ положите равной единице.

II. Решите предыдущую задачу при условии, что соотношение между нагрузками пяти СМО составляет такой ряд: 0,3; 0,4; 0,5; 0,6; 0,7. Рассчитайте величину коэффициента вариации длительности задержки заявок и сравните его с аналогичной величиной для задачи I при условии, что нагрузка каждой СМО равна 0,5.

III. Попробуйте найти величину $t_p(m)$ из формулы (6.6) для маршрута в СеМО, состоящего из двух систем вида $M/M/1$ с одинаковыми величинами нагрузки.

Литература к разд. 6

1. *Клейнрок, Л.* Вычислительные системы с очередями / Л. Клейнрок. – М. : Мир, 1979.

2. *Башарин, Г. П.* Теория сетей массового обслуживания и ее приложения к анализу информационно-вычислительных систем / Г. П. Башарин, А. Л. Толмачев // Итоги науки и техники. Сер. Теория вероятностей. Математическая статистика. Теоретическая кибернетика. – 1983. – Т. 21.

3. *Жожикашвили, В. А.* Сети массового обслуживания. Теория и применение к сетям ЭВМ / В. А. Жожикашвили, В. М. Вишневский. – М. : Радио и связь, 1988.

4. *Башарин, Г. П.* Анализ очередей в вычислительных сетях. Теория и методы расчета / Г. П. Башарин, П. П. Бочаров, Я. А. Коган. – М. : Наука, 1989.

5. *Бочаров, П. П.* Теория массового обслуживания / П. П. Бочаров, А. В. Печинкин. – М. : РУДН, 1995.

6. *Вишневский, В. М.* Теоретические основы проектирования компьютерных сетей / В. М. Вишневский. – М. : Изд-во «Техносфера», 2003.

7. *Ивницкий, В. А.* Теория сетей массового обслуживания / В. А. Ивницкий. – М. : Физматлит, 2004.

8. *Алиев, Т. И.* Основы моделирования дискретных систем / Т. И. Алиев. – СПб. : СПбГУ ИТМО, 2009.

9. *Деч, Г.* Руководство к практическому применению преобразования Лапласа и *Z*-преобразования / Г. Деч. – М. : Наука, 1971.

10. *Диткин, В. А.* Интегральные преобразования и операционное исчисление / В. А. Диткин, А. П. Прудников. – М. : Наука, 1974.

Заключение

Изложены результаты исследования однолинейных систем массового обслуживания, связанные в основном с теми показателями качества обслуживания трафика разного рода, которые нормируются в эксплуатируемых и вновь создаваемых телекоммуникационных сетях. Однако учебное пособие не следует рассматривать как справочный материал по всестороннему анализу однолинейных систем. Для детального изучения этих моделей телетрафика необходимо воспользоваться источниками, список которых приводится в конце каждого раздела.

У некоторых читателей могут возникнуть сложности с математическим аппаратом, используемым в учебном пособии (правда, авторы стремились максимально упростить все приведенные результаты). Вспомнить базовые положения теории вероятностей помогут монографии, на которые сделаны ссылки в разд. 1, или другие источники. Следует подчеркнуть, что надо очень осторожно относиться к материалам, размещенным на сайтах в интернете. Они не всегда содержат корректные результаты.

Авторы рекомендуют не пренебрегать вопросами и заданиями, которые приведены в конце каждого раздела. Они помогут лучше усвоить теоретический материал. Если у вас возникнут вопросы, обращайтесь к нам по адресу: sokolov@niits.ru.

Комментарии к вопросам и заданиям

Разд. 1

К заданию II. Если говорить о сетях следующего поколения, то дополнительными сведениями можно считать тип заявки. Например, под заявкой понимается IP-пакет. Тогда полезно знать вид информации (речь, данные, видео и т. д.), содержащейся в IP-пакете. Это позволит эффективно обслуживать заявки.

К заданию IV. Ответ на вопрос, поставленный в конце этого задания не так прост, как может показаться. Чаще всего время обмена информацией не зависит от длительности интервала (t_4, t_5). Кроме того, возможны ситуации, когда абоненты располагают ограниченным временем на процесс обмена информацией. Тогда затянувшиеся операции по уточнению номера вызываемого абонента приведут к сокращению времени для обмена информацией.

К заданию 4. По всей видимости, необходимо задать величину допустимой ошибки, что позволит ввести количественные оценки.

Разд. 2

К заданию II. Очевидно, что операции, связанные с концентрацией трафика, при $N = 1$ не используются. По этой причине $\pi_C = \pi_Y = 0$.

К заданию III. При проведении этого исследования целесообразно воспользоваться результатами, полученными при выполнении задания I.

К заданию IV. Один из интересных и полезных результатов, которые могут быть получены, состоит в том, чтобы определить те пороговые значения переменных Y, N и V, при которых вычисление искомых вероятностей (с заданной погрешностью) можно осуществлять по формуле Эрланга.

Разд. 3

К заданию I. Решить эту задачу можно двумя способами. Первый способ состоит в том, чтобы на основании соотношения (3.11) найти первый – $W^{(1)}$ и второй $W^{(2)}$ моменты длительности ожидания начала обслуживания. Тогда дисперсия исследуемой случайной величины σ_w^2 вычисляется так:

$$\sigma_w^2 = W^{(2)} - \left[W^{(1)}\right]^2.$$

Второй способ основан на использовании соотношений (3.23) при условии, что $\beta(s)$ задано выражением (3.14). Дважды дифференцируя соот-

ношение (3.23), несложно найти значения $W^{(1)}$ и $W^{(2)}$, по которым вычисляется дисперсия σ_w^2.

К заданию VI. Воспользуйтесь соотношением $\mu = B^{(1)}$ и очевидным неравенством $S^{(1)} \geq B^{(1)}$.

К заданию IX. Желательно построить графики функций $S(t)$ и $1 - S(t)$. Попробуйте для наглядности использовать логарифмический масштаб для оси ординат.

Разд. 4

К заданию III. Очевидно, что для предложенного соотношения между величинами λ_1, λ_2 и λ_3 эффективность приоритетного обслуживания будет низкой.

Разд. 5

К заданию I. После определения средних значений длительности задержки заявок для исследуемых моделей постарайтесь оценить влияние вида функций $A(t)$ и $B(t)$ на полученные результаты.

К заданию III. На самом деле нагрузка в этой таблице задана косвенно. В том случае, когда рассматривается экспоненциальное распределение, нагрузка входит в соотношение для $S^{(1)}$. Для параболического распределения она определяется через параметры α и β. Можно дать и другую трактовку, если отвлечься от терминологии теории телетрафика: рассматриваются однопараметрическое и двухпараметрическое распределения без связи с величиной «нагрузка».

К заданию IV. В качестве объектов исследования можно взять следующие виды распределений: логистическое, Рэлея, арксинуса, Симпсона.

К заданию VI. По всей видимости, следует отметить три момента. Во-первых, нет прямой аналогии между спектром частот и функцией распределения $A(t)$. Во-вторых, при преобразовании аналогового сигнала говорится о его ограничении верхней частотой *F*. Для функции $A(t)$ такое ограничение не требуется. В-третьих, аналоговый сигнал после его преобразования восстанавливается в первозданном виде. При исследовании СМО всегда оговаривается ошибка в оценке ее характеристик.

Разд. 6

К заданию III. Можно воспользоваться выражением (6.9), подставив в него $m = 2$. После этого несложно найти значение $t_p(m)$.

Соколов Андрей Николаевич
Соколов Николай Александрович

ОДНОЛИНЕЙНЫЕ СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ

Учебное пособие

Редактор И. И. Щенсняк

План 2010 г., п. 6

Подписано к печати 30.12.2010
Объем 7,0 усл.-печ. л. Тираж 150 экз. Зак. 108

Издательство «Теледом» ГОУВПО СПбГУТ. 191186 СПб., наб. р. Мойки, 61

Отпечатано в ГОУВПО СПбГУТ