

**А. З. МЕЛИКОВ
Л. А. ПОНОМАРЕНКО
В. В. ПАЛАДЮК**

ТЕЛЕТРАФИК

**МОДЕЛИ, МЕТОДЫ,
ОПТИМИЗАЦИЯ**

**Под научной редакцией
академика НАН Украины В. С. Королюка**

Киев
ИПК «Политехника»
2007

М 47 Меликов А.З., Пономаренко Л.А., Паладюк В.В.
Телетрафик: Модели, методы, оптимизация. – К.: ИПК
«Политехника», 2007. – 256 с.: ил.

Монография посвящена исследованию традиционных и мультисервисных систем и сетей связи. Методы расчета и оптимизации основных характеристик сетей связи с различными стратегиями доступа разнотипных (эластичных и неэластичных) заявок базируются на использовании математического аппарата теории массового обслуживания (ТМО). Исследование марковских процессов принятия решений в задачах оптимизации моделей систем телетрафика проводится с применением методов и алгоритмов их фазового укрупнения.

Книга будет полезна специалистам в области математической теории телетрафика, а также студентам и аспирантам соответствующих специальностей.

ISBN 978-966-8373-81-7

Печатается по решению Ученого совета

Военного института Киевского национального университета
имени Тараса Шевченко (протокол № 28 от 31 мая 2007 г.)

Рецензенты:

Скопецкий В.В.,

д-р физ.-мат. наук, проф., член-корреспондент НАН Украины,
заведующий отделом Института кибернетики им В.М.Глушкова
НАН Украины

Гостев В.И.

д-р техн. наук, проф., Заслуженный деятель науки Украины, профессор
Государственного университета информационно-коммуникационных
технологий Министерства транспорта и связи Украины

ISBN 978-966-8373-81-7

© А.З.Меликов, Л.А.Пономаренко,
В.В.Паладюк, 2007

СОДЕРЖАНИЕ

СПИСОК ОСНОВНЫХ СОКРАЩЕНИЙ И ОБОЗНАЧЕНИЙ.....	6
ПРЕДИСЛОВИЕ НАУЧНОГО РЕДАКТОРА.....	8
ВВОДНАЯ ЧАСТЬ. МЕТОДЫ РАСЧЕТА МНОГОМЕРНЫХ МОДЕЛЕЙ СИСТЕМ ТЕЛЕТРАФИКА.....	12
В.1. Точные методы расчета многомерных моделей систем телетрафика.....	13
В.2. Приближенные методы расчета многомерных моделей систем телетрафика.....	21
Заключение и комментарий к вводной части.....	26
Литература к вводной части.....	27
ЧАСТЬ I. ИССЛЕДОВАНИЕ МОДЕЛЕЙ ТРАДИЦИОННЫХ БЕСПРОВОДНЫХ СОТОВЫХ СЕТЕЙ СВЯЗИ.....	29
Глава 1. МЕТОДЫ ИССЛЕДОВАНИЯ ХАРАКТЕРИСТИК БЕСПРОВОДНЫХ СОТОВЫХ СЕТЕЙ СВЯЗИ.....	30
1.1. Организация процессов обработки вызовов в беспроводных сетях сотовой связи.....	30
1.2. Методы расчета и оптимизации характеристик беспроводных сетей.....	39
Заключение и комментарий к главе 1 части I.....	48
Литература к главе 1 части I.....	49
Глава 2. ИССЛЕДОВАНИЕ МОДЕЛЕЙ ПРОЦЕССОВ ОБСЛУЖИВАНИЯ ВЫЗОВОВ В СЕТЯХ БЕЗ ОЧЕРЕДЕЙ.....	51
2.1. Расчет параметров модели соты с резервированием каналов для хэндовер-вызовов.....	51
2.2. Расчет параметров модели соты с индивидуальными каналами для хэндовер-вызовов.....	60
2.3. Численные эксперименты.....	69
2.4. Сравнительный анализ различных стратегий доступа.....	76
Заключение и комментарий к главе 2 части I.....	79
Литература к главе 2 части I.....	80
Глава 3. МЕТОДЫ РАСЧЕТА ХАРАКТЕРИСТИК МОДЕЛЕЙ ПРОЦЕССОВ ОБСЛУЖИВАНИЯ ВЫЗОВОВ В БЕСПРОВОДНЫХ СЕТЯХ С ОЧЕРЕДЯМИ.....	81
3.1. Расчет характеристик модели с очередями хэндовер-вызовов.....	81
3.2. Расчет характеристик модели с очередями новых вызовов.....	94
3.3. Численные эксперименты.....	99

Заключение и комментарии к главе 3 части I	113	Выводы и комментарии к главе 2 части II	209
Литература к главе 3 части I	114	Литература к главе 2 части II	210
Глава 4. ЗАДАЧИ УЛУЧШЕНИЯ ХАРАКТЕРИСТИК ПРОЦЕССОВ ОБСЛУЖИВАНИЯ ВЫЗОВОВ В БЕСПРОВОДНЫХ СЕТЯХ СОТОВОЙ СВЯЗИ.	115	Глава 3. ЗАДАЧИ НАХОЖДЕНИЯ ОПТИМАЛЬНЫХ ЗНАЧЕНИЙ ПАРАМЕТРОВ СТРАТЕГИЙ ДОСТУПА.	211
4.1. Задачи улучшения характеристик процессов обслуживания вызовов в сотах без очередей.	115	3.1. Задачи для моделей типа Гимпельсона.	211
4.2. Задачи улучшения характеристик процессов обслуживания вызовов в сотах с очередями.	123	3.1.1. Решение задачи эквивалентной производительности при CS-стратегии доступа.	211
Заключение и комментарии к главе 4 части I	127	3.1.2. Задачи нахождения оптимальных значений параметров TR-стратегии доступа.	213
Литература к главе 4 части I	128	3.1.3. Задачи нахождения оптимальных значений параметров стратегии доступа со специальными каналами для обслуживания широкополосных заявок.	219
ЧАСТЬ II. ИССЛЕДОВАНИЕ МОДЕЛЕЙ МУЛЬТИСЕРВИСНЫХ СЕТЕЙ СВЯЗИ	130	3.2. Задачи нахождения оптимальных значений параметров стратегии доступа для моделей с неэластичными и эластичными заявками.	221
Введение	130	3.2.1. Задачи нахождения оптимальных значений параметров стратегии доступа для моделей с одной эластичной заявкой на обслуживание.	221
Глава 1. ИССЛЕДОВАНИЕ МОДЕЛЕЙ МНОГОСКОРОСТНЫХ СИСТЕМ С НЕЭЛАСТИЧНЫМИ ЗАЯВКАМИ.	133	3.2.2. Задачи нахождения оптимальных значений параметров стратегии доступа для моделей со многими эластичными заявками на обслуживание.	224
1.1. Методы расчета характеристик моделей многоскоростных систем общего типа.	133	Заключение и комментарии к главе 3 части II	225
1.1.1. Полнодоступная стратегия доступа.	141	Литература к главе 3 части II	226
1.1.2. Полнодоступная стратегия доступа с уравниванием	147	ЧАСТЬ III. МАРКОВСКИЕ ПРОЦЕССЫ ПРИНЯТИЯ РЕШЕНИЙ (МППР) В ЗАДАЧАХ ОПТИМИЗАЦИИ МОДЕЛЕЙ СИСТЕМ ТЕЛЕГРАФИКА.	227
1.1.3. Стратегия доступа с резервированием каналов.	152	Глава 1. МЕТОДЫ ИССЛЕДОВАНИЯ ЗАДАЧ МППР.	227
1.2. Стратегия, основанная на схеме выделения специальной группы каналов для обслуживания широкополосных заявок	159	1.1. Классическая постановка задач МППР и методы их исследования.	227
1.3. Численные эксперименты.	175	1.2. Иерархический алгоритм типа фазового укрупнения для исследования МППР.	230
Заключение и комментарии к главе 1 части II	178	Заключение и комментарии к главе 1 части III	235
Литература к главе 1 части II	181	Литература к главе 1 части III	235
Глава 2. ИССЛЕДОВАНИЕ МОДЕЛЕЙ МНОГОСКОРОСТНЫХ СИСТЕМ С НЕЭЛАСТИЧНЫМИ И ЭЛАСТИЧНЫМИ ЗАЯВКАМИ.	181	Глава 2. ПРИМЕНЕНИЕ МППР В ЗАДАЧАХ ОПТИМИЗАЦИИ МНОГОСКОРОСТНЫХ СИСТЕМ ОБСЛУЖИВАНИЯ.	237
2.1. Модели с одной эластичной заявкой на обслуживание	182	2.1. Нахождение оптимальной стратегии доступа к многоскоростной системе обслуживания.	237
2.1.1. Модель с равноправным доступом для заявок обоих типов.	186	2.2. Нахождение субоптимальной стратегии доступа к многоскоростной системе обслуживания.	246
2.1.2. Модели с ограниченным доступом для неэластичных заявок.	189	Заключение и комментарии к главе 2 части III	249
2.2. Модели со многими эластичными заявками на обслуживание	195	Литература к главе 2 части III	250
2.2.1. Модели с непрерывной полосой для эластичных заявок.	200		
2.2.2. Модели с дискретной полосой для эластичных заявок.			
2.3. Численные эксперименты.			

СПИСОК ОСНОВНЫХ СОКРАЩЕНИЙ И ОБОЗНАЧЕНИЙ

- $A\Phi Y$ – алгоритм фазового укрупнения;
 MS – микросостояние;
 PM – производящая матрица;
 $SLAU$ – система линейных алгебраических уравнений;
 $SUGP$ – система уравнений глобального равновесия;
 $SULP$ – система уравнений локального равновесия;
 SUP – система уравнений равновесия;
 US – укрупненное состояние;
 ΦPC – фазовое пространство состояний;
 $ЦМ$ – цепь Маркова;
 \emptyset – пустое множество;
 \approx – приближенное равенство;
 $:=$ – равенство по определению $a := b$ означает, что a определяется выражением b ;
 $i = \overline{1, N}$ – i принимает значение от 1 до N , иначе $i = \overline{1, \dots, N}$;
 $[x]$ – целая часть числа x ;
 $x^+ = \max(0, x)$;
 $n! = 1 \cdot 2 \cdot \dots \cdot n$;
 h -вызов – хэндовер-вызов;
 o -вызов – новый вызов;
 i -вызов – вызов i -го типа или вызов из i -го трафика;
 $\lambda_o(\mu_o)$ – интенсивность поступления (обслуживания) новых вызовов;
 $\lambda_h(\mu_h)$ – интенсивность поступления (обслуживания) хэндовер-вызовов;
 $\lambda_i(\mu_i)$ – интенсивность поступления (обслуживания) i -вызовов;
 $\nu_o = \lambda_o / \mu_o$ – нагрузка, порожденная новыми вызовами;
 $\nu_h = \lambda_h / \mu_h$ – нагрузка, порожденная хэндовер-вызовами;
 $\nu_i = \lambda_i / \mu_i$ – нагрузка, порожденная i -вызовами;
 BDP – Birth Death Process (процесс размножения и гибели);
 BS – Base Station (базовая станция);
 CAC – Call Admission Control (управление доступом вызовов);
 DCA – Dynamic Channel Allocation (динамическое распределение каналов);
 FCA – Fixed Channel Allocation (фиксированное распределение каналов);

- $FDMA$ – Frequency Division Multiple Access (множественный доступ с разделением частот);
 $FIFO$ – First-In-First-Out (дисциплина обслуживания, при которой вызовы обслуживаются в порядке поступления с начала очереди);
 GC – Guard Channels («сторожевые» каналы; каналы, зарезервированные только для h -вызовов);
 GSM – Global Systems for Mobile Communications (глобальные системы мобильной связи);
 $I(A)$ – индикаторная функция события A ;
 MRQ – Multi Rate Queue (многоскоростная система обслуживания);
 MS – Mobile Subscriber, Mobile Station (подвижный абонент или подвижный аппарат);
 MSC – Mobile Switching Center (центр коммутации в мобильных системах);

- $P(A)$ – вероятность случайного события A ;
 PB – Probability of blocking (вероятность блокировки);
 P_o – вероятность блокировки новых вызовов;
 P_h – вероятность потери хэндовер-вызовов;
 QoS – Quality of Service (качество обслуживания);
 $TDMA$ – Time Division Multiple Access (множественный доступ с разделением времени);

$WCCN$ – Wireless Cellular Communication Network (беспроводная сотовая сеть связи);

$\rho(x)$, $\pi(x)$, $\rho(x)$ – стационарная вероятность состояния x ;

$x \rightarrow y$ – переход из состояния x в состояние y ;

$$\chi(x) = \begin{cases} 1, & \text{если } x > 0, \\ 0, & \text{если } x \leq 0; \end{cases} \quad \delta_{i,j} = \begin{cases} 1, & \text{если } i = j, \\ 0, & \text{если } i \neq j \end{cases}$$

– символы Кронекера;

e_i – i -й ортвектор евклидова пространства, размерность которого варьируется в каждом конкретном случае.

ПРЕДИСЛОВИЕ НАУЧНОГО РЕДАКТОРА

Стремительно растущий спрос на телекоммуникационные услуги и обмен информацией на фоне быстрого прогресса в области электроники, волоконной оптики и вычислительной техники привел к тому, что связь стала одной из наиболее динамичных отраслей инфраструктуры современного общества.

На рубеже XX – XXI веков началась и продолжается конвергенция телекоммуникационных, компьютерных и информационных технологий, а также взаимопроникновение и слияние соответствующих отраслей промышленности.

Теория телеграфика сформировалась в последние годы как самостоятельная научная дисциплина, которая представляет собой набор всеохватывающих методов решения проблем проектирования новых и эксплуатации действующих систем телекоммуникаций. Заметим, что под понятием «телеграфик» понимают теперь не только классические телефонные и телеграфные сообщения, но и потоки сообщений в новых информационных-вычислительных и телекоммуникационных сетях. При решении возникающих в указанной отрасли инженерной практики проблем широко используются как аналитические, так и численные методы моделирования. Однако здесь области приложения аналитических методов существенно ограничены из-за сложности моделей современных систем телеграфика. Поэтому в современной теории телеграфика особую актуальность приобретает разработка эффективных численных методов исследования. Специфические особенности различных систем телеграфика усугубляют проблемы разработки универсальных численных методов.

Функционирование сложных вероятностных систем описывается математическими моделями, учитывающими основные специфические особенности работы систем. В качестве таких математических моделей, как правило, используются различные классы случайных процессов: марковские, полумарковские, регенерирующие, с дискретным вмешательством случая и т. п.

Наиболее эффективный математический аппарат анализа разработанных систем, функционирование которых описывается однородными цепями или процессами Маркова. Основные характеристики марковских моделей определяются решениями систем линейных уравнений (алгебраических, дифференциальных или интегральных). Однако предположение о марковости функционирования системы является весьма ограничивающим, поэтому в качестве математических моделей достаточно сложных реальных систем используются более общие классы случайных процессов. Вместе с тем марковости модели иногда удается достичь усложнением фазового пространства моделирующего процесса.

Стремление к максимальной точности описания функционирования реальных сложных систем приводит к тому, что соответствующие математические модели становятся все более и более сложными. При этом

усложняется их математический анализ, аппарат анализа становится громоздким и часто недоступным в инженерной практике.

Основная трудность при моделировании и анализе сложных телекоммуникационных систем проявляется в непомерном возрастании числа возможных различных их состояний, что приводит к практической необозримости модели. Одной из актуальных задач теории сложных систем, в частности создаваемой теперь математической теории телеграфика, и является преодоление этой основной трудности – большой размерности фазового пространства состояний модели системы. Здесь проблема состоит в том, чтобы использовать большое число возможных состояний системы в качестве положительного фактора, позволяющего существенно упростить ее анализ (аналогично этому в статистической физике основные закономерности больших физических систем проявляются именно благодаря взаимодействию огромного числа составляющих их частиц).

Наиболее радикальный подход к преодолению сложности анализа реальной системы состоит в построении более простой укрупненной системы, анализ которой существенно проще анализа реальной, а основные характеристики могут быть приняты в качестве характеристик последней.

Идея исследования сложной системы по частям на уровне подсистем с последующим переходом к самой системе лежит в основе методов решения разнообразных задач. Наиболее последовательно и систематически эта идея проводится в теории, известной под названием диаоптики, развитой Г.Кроном в известной монографии «Исследование сложных систем по частям. Диаоптика» (Москва, Изд-во иностранной литературы, 1965). Методы диаоптики развиты по существу для построения реальных больших систем алгебраических уравнений. При этом основным здесь является тот факт, что окончательное решение системы по частям оказывается точным решением исходной системы уравнений.

В математической экономике используются также методы агрегирования, в которых строится точное или приближенное решение систем линейных уравнений разбиением на подсистемы.

Однако, методы диаоптики и агрегирования, несомненно доказавшие свою значимость, оказываются во многих случаях недостаточными для решения задач, возникающих в практике.

В настоящей книге последовательно разрабатываются и успешно применяются алгоритмы укрупнения состояний сложных систем телеграфика, математическим основанием которых служат предельные теоремы, разработанные в свое время В.С.Королюком и А.Ф.Турбиным. Идея этого метода состоит в том, что фазовое пространство состояний реальной системы расщепляется на конечное число непересекающихся классов. Состояния каждого из этих классов укрупняются (склеиваются) в одно состояние. В новом укрупненном фазовом пространстве, состоящем из конечного числа состояний, строится укрупненная система,

функционалирование которой в определенном смысле описывает функционирование исходной системы.

Замечательным результатом укрупнения является тот факт, что при достаточно общих предположениях функционалирование укрупненной системы описывается целью Маркова с непрерывным временем, в то время как функционалирование исходной системы описывается существованием более сложными случайными процессами, в частности – полумарковскими процессами. При использовании результатов И.Н.Коваленко применение алгоритмов укрупнения допускается при моделировании реальных систем значительно более общими процессами с дискретным вмешательством случая.

Еще одним важным достоинством алгоритмов укрупнения является возможность многократного их применения. Можно применить алгоритм укрупнения к уже укрупненной системе, т. е. можно строить иерархию укрупненных целей Маркова, более или менее детально описывающих функционалирование исходной реальной системы.

Из приведенного качественного описания алгоритма укрупнения следует, что основная количественная информация о реальной системе, которая требуется для применения алгоритмов укрупнения, содержится в матрицах переходных вероятностей вложенной цепи Маркова. В алгоритмах укрупнения используются также средние времена пребывания реальной системы в отдельных состояниях. Все эти количественные характеристики исходной системы имеют простую статистическую интерпретацию и потому могут быть всегда определены по результатам наблюдений функционалирования реальной системы. Конечно, для большей достоверности результатов математического анализа сложной телекоммуникационной системы желательны указанные выше количественные характеристики получить расчетным путем, минуя статистический эксперимент.

Отрадно, что в данной монографии разрабатываются численные методы исследования конкретных классов систем телетрафика. Они могут быть применены для исследования как традиционных беспроводных сетей передачи речевых сообщений, так и мультимедийных сетей связи. При этом эффективность предложенных методов демонстрируется на конкретных примерах. Важно отметить, что они могут быть успешно применены и в других отраслях инженерной практики.

Читатель данной книги должен быть знаком с основами традиционных курсов теории телекоммуникаций, теории массового обслуживания (ТМО), численных методов и методов математического программирования, в объеме программы инженерных факультетов технических университетов.

Книга состоит из четырех частей. Во вводной части даются необходимые сведения из теории многопоточковых СМО, описываются известные методы расчета их показателей качества обслуживания. Здесь же

предложен новый приближенный метод расчета моделей систем телетрафика, которые описываются двумерными цепями Маркова.

В части 1, которая состоит из четырех глав, исследуются модели традиционных беспроводных сетей связи, предназначенные для передачи речевых сообщений. В отличие от классических моделей здесь допускается, что новые вызовы и хэндовер-вызовы могут отличаться друг от друга по времени занятия радиоканалов соты. В главе 1 кратко описываются процессы обработки вызовов в исследуемых сетях. В главе 2 исследуются модели без очередей при различных схемах преимущественного обслуживания хэндовер-вызовов. В главе 3 аналогичные модели исследуются при наличии конечной и бесконечной очереди разнотипных вызовов, при этом хэндовер-вызовы могут иметь ограниченное время пребывания в соте. В последней главе этой части решаются задачи улучшения показателей качества обслуживания.

В части 2, которая состоит из трех глав, исследуются модели мультимедийных сетей связи, предназначенные для передачи разнородных сообщений. В главе 1 разработаны эффективные алгоритмы расчета показателей качества обслуживания в таких сетях при различных стратегиях доступа. При этом предполагается, что все заявки являются неэластичными. В главе 2 исследуются смешанные модели, в которых осуществляется совместное обслуживание неэластичных и эластичных заявок. При этом рассматриваются модели двух типов: с дискретной и непрерывной схемами распределения полосы между эластичными заявками. В главе 3 решаются задачи улучшения показателей качества обслуживания разнотипных заявок в мультимедийных сетях связи.

В части 3, которая состоит из двух глав, рассматриваются вопросы применения методов теории марковских процессов принятия решений (МППР) в системах телетрафика. В главе 1 приводится общая схема применения МППР в исследуемых системах и описываются точный и приближенный методы решения соответствующих задач оптимизации. В главе 2 описываются задачи нахождения оптимальной и субоптимальной стратегий доступа в моделях мультимедийных сетей связи, при этом критерием эффективности является максимизация загрузки каналов сети.

Книга содержит оригинальные результаты авторов, опубликованные за последние годы в известных научных журналах и доложенные на представительных международных научных конференциях. Каждая глава сопровождается комментарием и списком литературы, которые в совокупности дают достаточно полное представление о современном состоянии исследований по математической теории телетрафика. Книга будет полезна специалистам в области телекоммуникационных систем, а также студентам и аспирантам соответствующих специальностей.

Академик НАН Украины В.В.Королюк
Киев, 18 июня 2007 г.

Среди многомерных моделей систем телеграфика самой простой и в то же время важной (с практической и методологической точек зрения) является модель $M_K | M_K | N | 0$. Она служит естественным обобщением классической однопоточковой модели Эрланга $M | M | N | 0$.

Сначала рассмотрим разомкнутую модель $M_K | M_K | M | 0$. Краткое описание этой системы состоит в следующем. На вход СМО, содержащей $N \geq 1$ каналов, поступает K -мерный пуассоновский поток заявок первого рода, при этом i -й поток имеет параметр λ_i , $i = 1, \dots, K$. Функция распределения (ф.р.) времени обслуживания заявок i -го типа (i -заявок) является экспоненциальной с параметром μ_i . Если в момент поступления заявки любого типа в системе имеется хотя бы один свободный канал, то эта заявка принимается на обслуживание. В противном случае поступившая заявка теряется и в дальнейшем не оказывает влияния на работу системы.

Задача состоит в нахождении стационарного распределения вероятностей состояний системы.

С учетом вида ф.р. входящих трафиков и времени обслуживания разнотипных заявок можно сделать вывод, что функционирование этой системы можно описать K -мерной цепью Маркова (ЦМ). В стационарном режиме состояние системы в произвольный момент времени можно представить вектором $\mathbf{n} = (n_1, n_2, \dots, n_K)$, где n_i — число i -заявок в системе. Существование стационарного режима вытекает из конечности и неприводимости соответствующей ЦМ. Тогда множество всех возможных состояний системы, т.е. фазовое пространство состояний (ФПС) системы определяется следующим образом:

$$S := \left\{ \mathbf{n} : n_i = 0, 1, \dots, N; \sum_{i=1}^K n_i \leq N \right\}. \quad (B.1)$$

Поскольку изменения состояний системы происходят лишь в моменты поступления заявок и/или завершения их обслуживания, то в ФПС (B.1) возможны лишь переходы $\mathbf{n} \rightarrow \mathbf{n} + \mathbf{e}_i$ и $\mathbf{n} + \mathbf{e}_i \rightarrow \mathbf{n}$, где \mathbf{e}_i — i -й ортвектор K -мерного евклидова пространства, $i = 1, 2, \dots, K$.

Интенсивности этих переходов $q(\mathbf{n}, \mathbf{n}')$; $\mathbf{n}, \mathbf{n}' \in S$, определяются выражением:

$$q(\mathbf{n}, \mathbf{n}') = \begin{cases} \lambda_i, & \text{если } \mathbf{n}' = \mathbf{n} + \mathbf{e}_i, \\ n_i \mu_i, & \text{если } \mathbf{n}' = \mathbf{n} - \mathbf{e}_i, \\ 0 & \text{в остальных случаях.} \end{cases} \quad (B.2)$$

На основании выражения (B.2) можно составить систему уравнений глобального равновесия (СУГР) относительно стационарных вероятностей состояний $p(\mathbf{n})$, $\mathbf{n} \in S$ (в дальнейшем СУГР иногда коротко называ-

ВВОДНАЯ ЧАСТЬ

МЕТОДЫ РАСЧЕТА МНОГОМЕРНЫХ МОДЕЛЕЙ СИСТЕМ ТЕЛЕГРАФИКА

Возрастающая сложность систем телеграфика выдвигает на первый план проблемы разработки достаточно адекватных математических моделей этих систем с целью получения достоверных оценок их характеристик, реализации задач их оптимизации относительно выбранного критерия качества и разработки соответствующих алгоритмов управления ими.

Основным математическим аппаратом, позволяющим строить аналитические модели, адекватные распространённым системам телеграфика, является аппарат теории систем и сетей массового обслуживания. При таком подходе предполагается, что для описания эволюции рассматриваемой системы используется некоторая система массового обслуживания (СМО) с определенной степенью точности учитывающая ее основные особенности. При этом очевидно, что классические однопоточковые СМО могут служить лишь очень грубыми (приближенными) математическими моделями современных систем телеграфика, и поэтому такие модели могут быть только исходным, начальным приближением при математическом анализе таких систем.

Функционирование современных систем телеграфика, в которых сообщения существенно отличаются друг от друга (например, по уровню приоритетов, механизму обслуживания и т. д.) могут быть достаточно точно описаны с помощью моделей много-поточковых СМО, т.е. достаточно адекватные модели современных систем телеграфика обязательно являются многомерными. Необходимость исследования многомерных моделей особенно выпукло проявляется при изучении мультимедийных систем телеграфика, в которых в рамках единой сети объединяется передача разнородной информации — данных, речи, видеоинформации, телефаксов и др.

Напомним необходимые сведения о методах расчета многомерных моделей систем телеграфика.

ется СУР — система уравнений равновесия). С использованием (В.2) легко получить явный вид этой СУГР:

$$\left(\sum_{i=1}^K \lambda_i \chi \left(N - \sum_{j=1}^K n_j \right) + \sum_{i=1}^K n_i \mu_i \right) p(\mathbf{n}) = \sum_{i=1}^K \lambda_i p(\mathbf{n} - \mathbf{e}_i) \chi(n_i) + \sum_{i=1}^K n_i \mu_i p(\mathbf{n} + \mathbf{e}_i) \chi \left(N - \sum_{j=1}^K n_j - 1 \right), \quad \mathbf{n} \in S; \quad (\text{В.3})$$

$$\sum_{\mathbf{n} \in S} p(\mathbf{n}) = 1. \quad (\text{В.4})$$

Здесь и в дальнейшем $\chi(x) = \begin{cases} 1, & \text{если } x > 0, \\ 0, & \text{если } x \leq 0. \end{cases}$

Уравнение (В.4) называется условием нормировки. Решение СУГР имеет мультипликативный вид:

$$p(\mathbf{n}) = G^{-1}(N, K) \prod_{i=1}^K \frac{\lambda_i^{n_i}}{n_i!}, \quad \mathbf{n} \in S, \quad (\text{В.5})$$

где $\lambda_i := \lambda_i / \mu_i$, $G(N, K)$ — нормирующая константа, обеспечивающая выполнение условия (В.4):

$$G(N, K) = \sum_{\mathbf{n} \in S} \prod_{i=1}^K \frac{\lambda_i^{n_i}}{n_i!}. \quad (\text{В.6})$$

Из (В.5) определяем, что

$$p(\mathbf{0}) = G^{-1}(N, K). \quad (\text{В.7})$$

Для доказательства справедливости (В.5) составляется система уравнений локального равновесия (СУЛР) относительно соотношения (В.3).

Указанная СУЛР имеет следующий вид:

$$q(\mathbf{n}, \mathbf{n}') p(\mathbf{n}) = q(\mathbf{n}', \mathbf{n}) p(\mathbf{n}'), \quad \forall \mathbf{n}, \mathbf{n}' \in S. \quad (\text{В.8})$$

С использованием выражения (В.2) СУЛР (В.8) в явном виде записывается так:

$$\lambda_i p(\mathbf{n}) = (n_i + 1) \mu_i p(\mathbf{n} + \mathbf{e}_i), \quad \forall \mathbf{n}, \mathbf{n} + \mathbf{e}_i \in S. \quad (\text{В.9})$$

Теперь легко показать, что (В.5) является решением СУЛР (В.9). Действительно, из выражений (В.5) и (В.9) находим, что

$$\frac{p(\mathbf{n} + \mathbf{e}_i)}{p(\mathbf{n})} = \frac{\lambda_i}{n_i + 1} = \frac{\lambda_i}{(n_i + 1) \mu_i},$$

или, иными словами, (В.5) является решением СУЛР (В.9) и, следовательно, решением СУГР (В.3), (В.4).

Важной характеристикой данной модели является стационарная вероятность потери (блокировки) заявок (Probability of Blocking, PB). С учетом теоремы PASTA заключаем, что она является одинаковой для всех трафиков и определяется формулой:

$$PB := \sum_{\mathbf{n} \in S_d} p(\mathbf{n}), \quad (\text{В.10})$$

где $S_d := \left\{ \mathbf{n} \in S : \sum_{i=1}^K n_i = N \right\}$ — множество диагональных состояний (см.

рис. В.1).

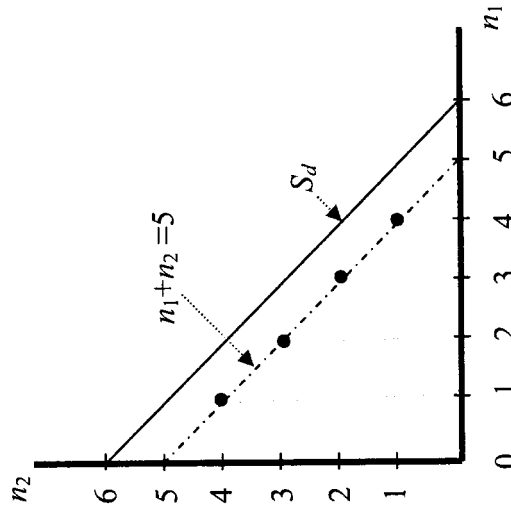


Рис. В.1. Вид ФПС (В.1) для $N = 6$, $K = 2$

Мультипликативный вид (В.5) стационарного распределения данной модели инвариантен относительно ф.р. времен обслуживания заявок при фиксированном среднем. Это утверждение остается в силе и для замкнутых моделей типа Энгсета, когда разнотипные заявки поступают от конечных источников.

Вычисление стационарного распределения с помощью (В.5) связано с огромными трудностями для моделей со многими каналами и типами трафиков. Это объясняется тем, что прямое вычисление $G(N, K)$ по формуле (В.6) сопровождается переполнением (при $i_j > 1$) и исчезновением порядка (при $i_j \rightarrow 0$).

Для преодоления указанных трудностей предложены различные подходы. Одним из эффективных подходов является использование алгоритма Бузена, основанного на двумерных рекуррентных формулах. При использовании указанного алгоритма нормирующая константа $G(N, K)$ вычисляется с помощью следующей рекурсии:

$$G(j, i) = \sum_{l=0}^j \frac{v_l}{l!} G(j-l, i-1), \quad i = \overline{2, K}, j = \overline{0, N}, \quad (\text{B.11})$$

$$G(j, 1) = \sum_{l=0}^j \frac{v_l}{l!}. \quad (\text{B.12})$$

Другой подход основан на принципах укрупнения ФПС (В.1). В указанном ФПС рассматривается следующее расщепление:

$$S = \prod_{j=0}^N S_j, \quad S_j \cap S_{j'} = \emptyset, \quad j \neq j', \quad (\text{B.13})$$

где $S_j := \left\{ \mathbf{n} \in S : \sum_{i=1}^K n_i = j \right\}$, т.е. подмножество S_j содержит все те

состояния из ФПС S , в которых общее число заявок равно $j, j = \overline{0, 1, \dots, N}$ (см. рис. В.1).

Определяется стационарная вероятность $\pi(j)$ укрупненного состояния S_j :

$$\pi(j) = \sum_{\mathbf{n} \in S_j} \prod_{i=1}^K \frac{v_i^{n_i}}{n_i!} G^{-1}(N, K), \quad j = \overline{0, N}. \quad (\text{B.14})$$

Утверждение 1. Введенные в (В.14) величины $\pi(j)$ удовлетворяют следующей системе уравнений:

$$\sum_{l=1}^N v_l \pi(j-l) = j \pi(j), \quad j = \overline{0, N}, \quad (\text{B.15})$$

$$\sum_{j=0}^N \pi(j) = 1, \quad (\text{B.16})$$

где $\pi(x) = 0$ при $x < 0$.

Для доказательства справедливости (В.15), (В.16) предварительно докажем следующую лемму.

Лемма 1.

$$v_i \pi(j-1) = M(n_i/j) \pi(j), \quad i = \overline{1, 2, \dots, K}; j = \overline{0, 1, \dots, N}, \quad (\text{B.17})$$

где M — знак математического ожидания.

Доказательство леммы. Перепишем СУЛР (В.9) в следующем виде:

$$v_i \pi(\mathbf{n}) p(\mathbf{n} - \mathbf{e}_i) = n_i p(\mathbf{n}), \quad i = \overline{1, 2, \dots, K}, \quad (\text{B.18})$$

где $\gamma_i(\mathbf{n}) = \begin{cases} 1, & \text{если } n_i \geq 1, \\ 0, & \text{если } n_i = 0. \end{cases}$

Суммируя (В.18) по всем $\mathbf{n} \in S_j$, получим:

$$v_i \sum_{\mathbf{n} \in S_j} \gamma_i(\mathbf{n}) p(\mathbf{n} - \mathbf{e}_i) = \sum_{\mathbf{n} \in S_j} n_i p(\mathbf{n}). \quad (\text{B.19})$$

Преобразуем левую часть равенства (В.19):

$$v_i \sum_{\mathbf{n} \in S_j} \gamma_i(\mathbf{n}) p(\mathbf{n} - \mathbf{e}_i) = v_i \sum_{\mathbf{n} \in S_j \cap \{n_i \geq 1\}} p(\mathbf{n} - \mathbf{e}_i), \quad (\text{B.20})$$

где

$$S_j \cap \{n_i \geq 1\} := \left\{ \mathbf{n} \in S_j : \sum_{l \neq i} n_l + (n_i - 1) = j - 1, n_l \geq 0, l \neq i \right\}.$$

Отсюда

$$v_i \sum_{\mathbf{n} \in S_j} \gamma_i(\mathbf{n}) p(\mathbf{n} - \mathbf{e}_i) = v_i \pi(j-1), \quad j \geq 1. \quad (\text{B.21})$$

Теперь преобразуем правую часть равенства (В.19):

$$\sum_{\mathbf{n} \in S_j} n_i p(\mathbf{n}) = \sum_{\mathbf{n} \in S_j} n_i \cdot \frac{p(\mathbf{n})}{\pi(j)} \pi(j), \quad j = \overline{0, 1, \dots, N}, i = \overline{1, 2, \dots, K}.$$

С учетом того, что

$$P(\mathbf{n}|j) = P \left(\left| \sum_{l=1}^K n_l = j \right. \right) = \begin{cases} \frac{p(\mathbf{n})}{\pi(j)}, & \text{если } \mathbf{n} \in S_j, \\ 0 & \text{в остальных случаях,} \end{cases}$$

находим:

$$\sum_{\mathbf{n} \in S_j} n_i p(\mathbf{n}) = \left(\sum_{\mathbf{n} \in S_j} n_i P(\mathbf{n}|j) \right) \pi(j) = M(n_i|j) \pi(j) \quad \forall j = \overline{0, 1, \dots, N}, i = \overline{1, 2, \dots, K}. \quad (\text{B.22})$$

Сравнивая (В.21) и (В.22), получим:

$$v_i \pi(j-1) = M(n_i|j) \pi(j), \quad j = \overline{0, 1, \dots, N}, i = \overline{1, 2, \dots, K}, \quad (\text{B.23})$$

Лемма доказана.

Для доказательства Утверждения 1 суммируются (B.23) по всем $i, i=1, 2, \dots, K$. Тогда получим:

$$\sum_{i=1}^K v_i \pi(j-1) = \left(\sum_{i=1}^K M(n_i | j) \right) \pi(j) = M \left(\sum_{i=1}^K n_i | j \right) \pi(j) = j \pi(j).$$

Утверждение 1 доказано.

Преимущество алгоритма (B.15), (B.16) перед алгоритмом Бузена (B.11), (B.12) очевидно. Во-первых, в отличие от алгоритма Бузена, алгоритм (B.15), (B.16) представляет собой одномерные рекурсии для любого K , которые тривиальным образом генерируют распределение $\pi(j), j=0, 1, \dots, N$. Во-вторых, в алгоритме (B.15), (B.16) нет зависимости от K . С использованием (B.7) и (B.10) из (B.15), (B.16) находим, что

$$G^{-1}(N, K) = \pi(0), \quad (B.24)$$

$$PB = \pi(N). \quad (B.25)$$

Алгоритм (B.15), (B.16) имеет простую вероятностную интерпретацию. Действительно, поскольку случайная величина j (общее число занятых каналов) описывается одномерным процессом размножения и гибели, то для данного процесса можно записать такое простое уравнение равновесия:

$$\lambda_j \pi(j-1) = n_i(j) \mu_j \pi(j)$$

или

$$v_j \pi(j-1) = n_i(j) \pi(j), \quad (B.26)$$

где $n_i(j)$ означает число i -заявок в системе, когда общее число заявок в системе равно j .

Сравнивая (B.26) с (B.17), убеждаемся, что они совпадают. Используя соотношения (B.15) и (B.16), легко выполнить тривиальное обобщение классических формул Эрланга для многопоточковых систем:

$$\pi(j) = \frac{v^j}{N! v^j} \sum_{i=0}^j \frac{v^i}{i!}, \quad j = 0, 1, \dots, N$$

где $v := \sum_{i=1}^K v_i$ — общая нагрузка системы.

Из (B.10) и (B.14) находим, что $PB = \pi(N)$. Описанные выше вычислительные процедуры тривиальным образом распространяются на модели многопоточковых систем обслуживания (Multi Rate Queues, MRQ). В многопоточковых системах i -заявки требуют для обслуживания одновременно $b_i \geq 1$ каналов, при этом все b_i каналов

начинают и завершают обслуживание одновременно (более подробно эти модели исследуются в последующих частях книги).

Для MRQ фазовое пространство состояний определяется так (см. рис. B.2):

$$S = \{n : n_i = 0, 1, \dots, [N/b_i], n \bullet b \leq N\}, \quad (B.27)$$

где $[x]$ означает целую часть x , $b = (b_1, \dots, b_K)$, $n \bullet b = \sum_{i=1}^K n_i b_i$.

Стационарное распределение данной модели также имеет мультипликативный вид (B.5), однако здесь нормирующая константа $G(N, K)$ определяется над ФПС (B.27).

В отличие от классической многомерной модели, в данной модели вероятности потери разнотипных заявок отличаются друг от друга. Так, вероятность потери i -заявок (PB_i) определяется следующим образом (см. рис. B.2):

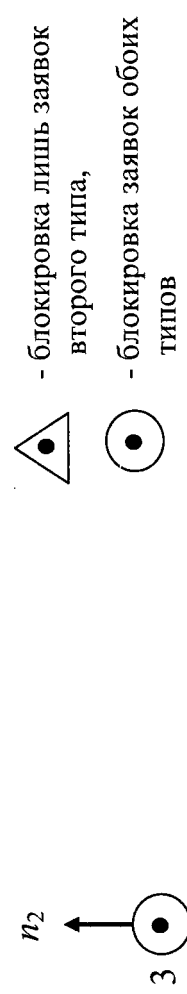


Рис. B.2. Вид ФПС (B.27) для $N=6, K=2, b=(1, 2)$

$$PB_i = \sum_{n \in S} p(n) I(f(n) < b_i), \quad (B.28)$$

где $f(n) := N - (n \bullet b)$ — число свободных каналов в состоянии $n \in S$, $I(A)$ — индикаторная функция события A .

Алгоритм Бузена (B.11), (B.12) для модели MRQ определяется так:

$$G(j, i) = \sum_{l=0}^{[j/b_i]} \frac{v_l^i}{l!} G(j - lb_i, i - 1), \quad i = 2, \dots, K, \quad j = 0, 1, \dots, N, \quad (B.29)$$

$$G(j, 1) = \sum_{l=0}^{[j/b_1]} \frac{v_l^1}{l!}. \quad (B.30)$$

Альтернативным подходом к расчету стационарного распределения модели MRQ являются одномерные рекуррентные формулы Кауфмана-Робертса.

Рассматривается следующее расщепление ФПС (В.27):

$$S = \sum_{j=0}^N S_j, \quad S_j \cap S_{j'} = \emptyset, \quad j \neq j', \quad (В.31)$$

где $S_j = \{n \in S : n \bullet b = j\}$, т.е. подмножество S_j содержит все те состояния из ФПС (В.27), в которых число занятых каналов равно j , $j = 0, 1, \dots, N$.

Утверждение 2. Стационарные вероятности укрупненных состояний $\pi(j)$ удовлетворяют следующие уравнения:

$$\sum_{i=1}^N v_i b_i \pi(j - b_i) = j \pi(j), \quad j = 0, \overline{N}, \quad (В.32)$$

$$\sum_{j=0}^N \pi(j) = 1, \quad (В.33)$$

где $\pi(x) = 0$ при $x < 0$.

Для доказательства утверждения 2 используется следующая лемма.

Лемма 2.

$$v_i \pi(j - b_i) = M(n_i/j) \pi(j), \quad i = 1, 2, \dots, K; \quad j = 0, 1, \dots, N.$$

Доказательство леммы 2 и утверждения 2 производится аналогично доказательствам леммы 1 и утверждения 1.

После определения $\pi(j)$, $j = 0, 1, \dots, N$, из соотношений (В.32), (В.33) вероятность потери i -заявок определяется так:

$$PB_i = \sum_{j=0}^{b_i-1} \pi(N - j), \quad i = 1, \dots, K. \quad (В.34)$$

Вероятностная интерпретация алгоритма (В.32), (В.33) совпадает с соответствующей интерпретацией для классической многомерной модели. Преимущество алгоритма (В.32), (В.33) перед алгоритмом (В.29), (В.30) очевидны.

В.2. Приближенные методы расчета многомерных моделей систем телеграфика

Во многих случаях для нахождения характеристик реальных систем телеграфика не удается использовать точные методы. Это объясняется тем, что соответствующие СУР не имеют аналитического решения (например, решения мультипликативного вида). В таких случаях необходимо использовать различные приближенные методы. Универсальным является метод имитационного моделирования.

Другим эффективным подходом к решению указанной проблемы является использование различных численных процедур.

Традиционный подход к использованию численных процедур для расчета характеристик марковских многомерных моделей систем телеграфика состоит в следующем.

На первом этапе определяется ФПС исследуемой модели, содержащее, как правило, векторы соответствующей размерности, описывающие микросостояния (МС) системы. На втором этапе при известных нагрузочных параметрах модели, исходя из принятой стратегии доступа различных заявок и механизма их обслуживания, определяется производящая матрица (ПМ) соответствующей многомерной цепи Маркова, описывающей функционирование системы. Определение ПМ эквивалентно составлению СУР для стационарных вероятностей состояний системы. На третьем этапе при наличии стационарного режима с применением определенных численных процедур находится стационарное распределение системы. И, наконец, на последнем этапе искомые характеристики системы выражаются через стационарные вероятности состояний. Это удается благодаря уникальному свойству марковских систем, для которых стационарная вероятность состояния представляет собой долю времени пребывания системы в соответствующем состоянии за достаточно большой интервал времени наблюдения.

Основная проблема применения указанного стандартного подхода связана с решением СУР (т.е. реализацией третьего этапа), так как с ростом числа классов разнотипных заявок и увеличением количества каналов, размерность ФПС модели растет с большой скоростью. В некоторых случаях с использованием специфической структуры соответствующей ПМ, а также учитывая соотношение ее элементов удается упростить эту проблему. Так, эта задача существенно упрощается для систем, которые описываются моделями обратимых марковских процессов.

Двумерный марковский процесс является обратимым, если не существует циркуляции между его состояниями. Циркуляция между состояниями определяется следующим образом [18]. Выбираются четыре соседних состояния (см. рис.В.3). Циркуляция по часовой стрелке и против нее определяются так:

по часовой стрелке:

$$(i, j) \rightarrow (i, j+1) : p(i, j) \lambda_2(i, j);$$

$$(i, j+1) \rightarrow (i+1, j+1) : p(i, j+1) \lambda_1(i, j+1);$$

$$\begin{aligned}
(i+1, j+1) &\rightarrow (i+1, j) : p(i+1, j+1)\mu_2(i+1, j+1); \\
(i+1, j) &\rightarrow (i, j) : p(i+1, j)\mu_1(i+1, j); \\
\text{против часовой стрелки:} \\
(i, j) &\rightarrow (i+1, j) : p(i, j)\lambda_1(i, j); \\
(i+1, j) &\rightarrow (i+1, j+1) : p(i+1, j)\lambda_2(i+1, j); \\
(i+1, j+1) &\rightarrow (i, j+1) : p(i+1, j+1)\mu_1(i+1, j); \\
(i, j+1) &\rightarrow (i, j) : p(i, j+1)\mu_2(i, j+1);
\end{aligned}$$

После сокращения вероятностей состояний, получается такое необходимое и достаточное условие обратимости двумерных марковских процессов:

$$\lambda_2(i, j)\lambda_1(i, j+1)\mu_2(i+1, j+1)\mu_1(i+1, j) = \lambda_1(i, j)\lambda_2(i+1, j)\mu_1(i+1, j)\mu_2(i, j+1). \quad (B.35)$$

Если условие (B.35) выполняется, то в системе удовлетворяется условие локального равновесия. Таким образом, необходимое условие обратимости состоит в том, что если существует переход из состояния i в состояние j , то должен существовать и обратный переход. Тогда имеется возможность выразить вероятность состояния (i, j) через вероятность состояния $(0, 0)$, при этом может быть выбран любой путь между этими двумя состояниями. Так, например, если выбирается путь $(0, 0), (1, 0), \dots, (i, 1), \dots, (i, j)$, то имеем:

$$p(i, j) = \frac{\lambda_1(0, 0)}{\mu_1(1, 0)} \cdot \frac{\lambda_1(1, 0)}{\mu_1(2, 0)} \cdot \dots \cdot \frac{\lambda_1(i-1)}{\mu_1(i, 0)} \cdot \frac{\lambda_2(i, 0)}{\mu_2(i, 1)} \cdot \frac{\lambda_2(i, 1)}{\mu_2(i, 2)} \cdot \dots \cdot \frac{\lambda_2(i, j-1)}{\mu_2(i, j)} \cdot p(0, 0).$$

Далее $p(0, 0)$ находится из условия нормировки.

Условие обратимости (B.35) удовлетворяется для многих классических систем обслуживания. Так, в частности, легко проверить, что оно верно для модели $M_2 | M_2 | N | 0$ (см. п.В.1), а также для ее обобщения — многократной системы обслуживания.

Задача расчета многомерных моделей систем телеграфика существенно усложняется, если соответствующий марковский процесс не является обратимым. Следовательно, актуальной становится проблема поиска путей, которые позволяют осуществить упрощенное описание многомерных моделей систем телеграфика. В связи с этим уместно процитировать В.С.Королюка [8]: "Одна из актуальных задач теории сложных систем состоит в построении более простых, укрупненных систем, анализ которых существенно проще анализа исходных математических моделей, а основные характеристики укрупненных моделей могут быть приняты в качестве характеристик исходных моделей".

Принципы фазового укрупнения весьма плодотворны и для преодоления "тирании размерности" в задачах расчета и оптимизации многомерных марковских и полумарковских систем. Ниже на их основе предлагается численный метод решения указанной выше проблемы для одного класса двумерных марковских процессов.

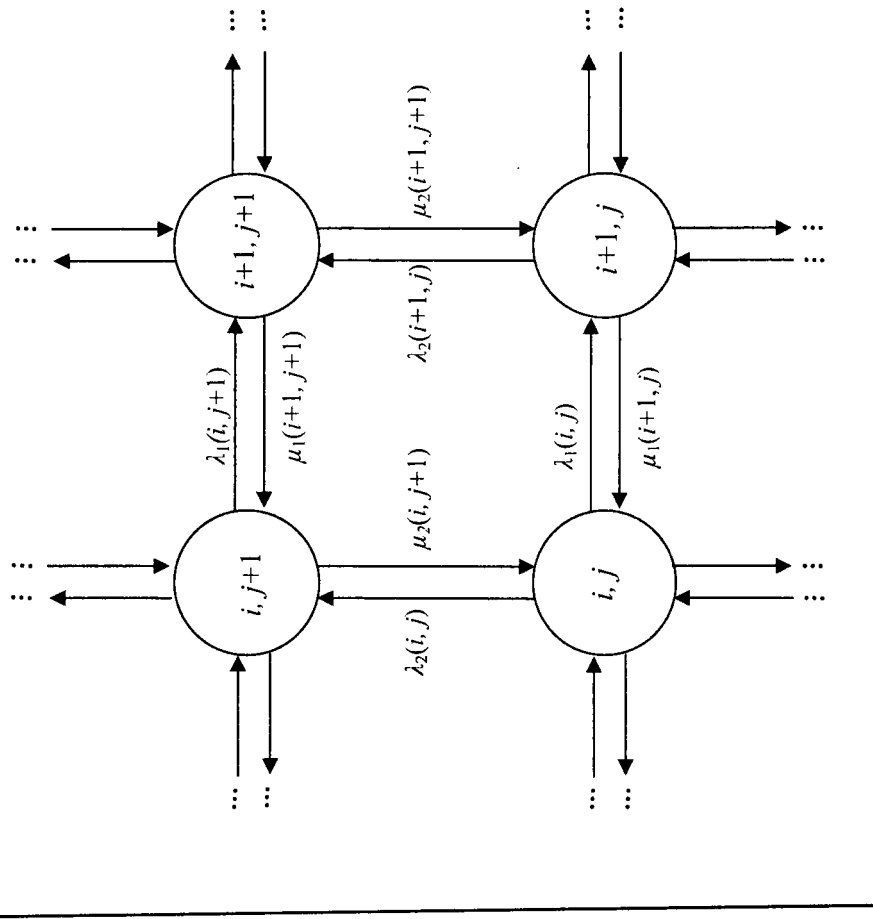


Рис. В.3. Схема для определения циркуляции между состояниями двумерных марковских процессов

Пусть ФПС двумерной ЦМ задается множеством $S := \{(i, j) : i = \overline{0, m}, j = \overline{0, n}\}$ где параметры m и/или n могут принимать бесконечные значения и $Q = \|q((i, j), (i', j'))\|$ является ее производящей матрицей. Предположим, что существует ее стационарное распределение $(p(i, j); (i, j) \in S)$.

Вводятся следующие определения.

Определение В.1. Двумерная цепь Маркова называется *строго непрерывной* по первой компоненте, если для всех возможных состояний (i, j) и $(i+1, j)$ выполняется следующее условие:

$$q((i, j), (i+1, j)) \cdot q((i+1, j), (i, j)) > 0. \quad (B.36)$$

Аналогичным образом определяются цепи, которые являются строго непрерывными по второй компоненте.

Определение В2. Двумерная цепь Маркова называется *слабо непрерывной* по второй компоненте, если для любого i существуют такие i' и j , что для некоторых l и k выполняется следующее условие:

$$q((i, j), (i', j+1)) \cdot q((l, j+1), (k, j)) > 0. \quad (B.37)$$

Аналогичным образом определяется *слабая непрерывность* двумерной цепи по первой компоненте.

Вербально условие (B.36) означает, что для всех возможных состояний (i, j) и $(i+1, j)$ существуют положительные интенсивности переходов между любыми двумя соседними состояниями в каждой строке; а условие (B.37) означает, что существуют положительные интенсивности переходов хотя бы между двумя парами состояний соседних строк (см. рис. В.4).

Пусть рассматриваемая двумерная цепь является строго непрерывной по первой компоненте и слабо непрерывной по второй компоненте.

Для корректного применения принципов теории фазового укрупнения ниже принимается следующее допущение.

Допущение.

$$q((i, j), (i', j)) \gg q((i, j), (i', j')), \text{ где } j \neq j'.$$

Это допущение означает, что интенсивности переходов между состояниями одной строки (т.е. при фиксированных значениях второй компоненты) намного больше интенсивностей между строками.

Замечание В1. Если рассматривается двумерная цепь, которая является строго непрерывной по второй компоненте, то это допущение очевидно образом модифицируется для столбцов (т.е. при фиксированных значениях первой компоненты).

Важно отметить, что принятое выше допущение не является экстремальным, и оно выполняется во многих реальных системах телеграфика, в которых интенсивность поступления вызовов одного типа намного превосходит интенсивность поступления вызовов другого типа. Более того, как будет видно из результатов последующих глав, это допущение почти не влияет на точность результатов вычислений характеристикисследуемых систем. Оно принимается лишь для корректности применения принципов теории фазового укрупнения.

Предлагаемый алгоритм содержит следующие шаги.

Шаг 1. Рассматривается расщепление ФПС S по значениям второй компоненты:

$$S = \prod_{j=0}^n S_j, \quad S_i \cap S_j = \emptyset, \quad i \neq j,$$

где $S_j := \{(i, j) \in S : i \in \overline{0, m}\}$, $j = \overline{0, n}$.

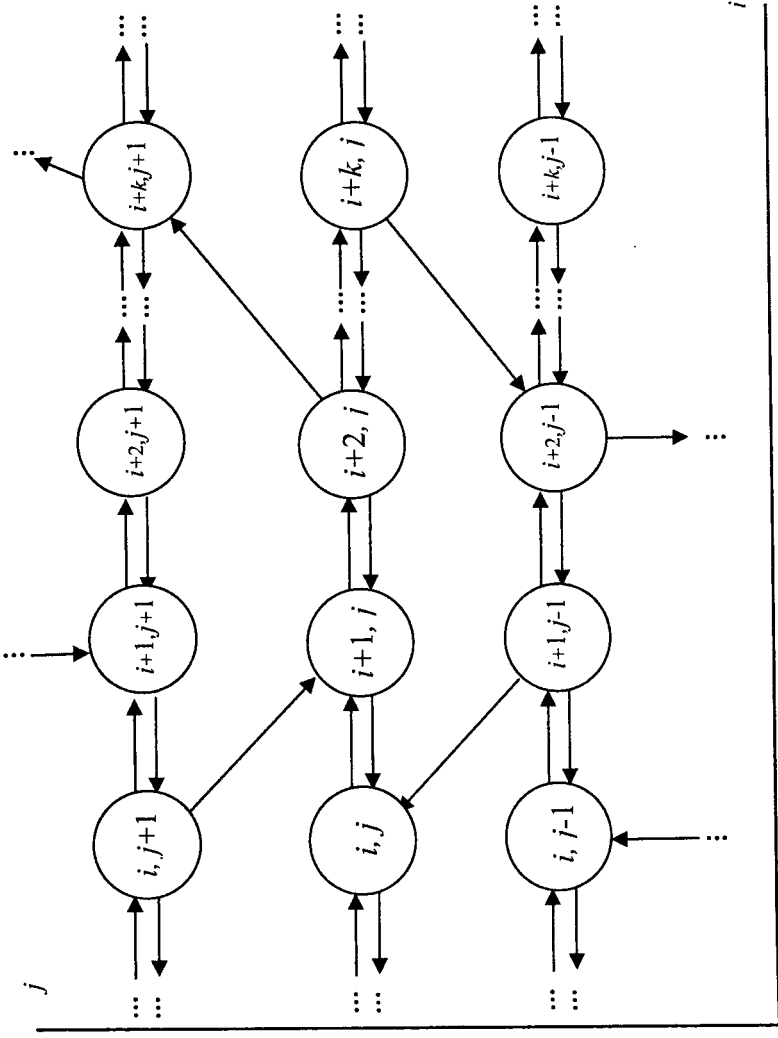


Рис. В.4. Вид диаграммы переходов между состояниями двумерной ЦМ, строго непрерывной по первой компоненте и слабо непрерывной по второй компоненте

Шаг 2. Все микросостояния из подкласса S_j объединяются в одно укрупненное состояние $\langle j \rangle$, $j = \overline{0, 1, \dots, n}$. Все такие укрупненные состояния образуют укрупненную модель с пространством состояний $\tilde{S} := \{\langle j \rangle : j = \overline{0, n}\}$.

Шаг 3. Находится стационарное распределение внутри расщепленных классов S_j , $j = \overline{0, 1, \dots, n}$. При этом элементы производящих матриц (Q_j) этих одномерных ЦМ $q_j(i, i)$ определяются так:

$$q_j(i, i) := q((i, j), (i, j)).$$

Стационарная вероятность состояния (i, j) внутри класса S_j обозначается $\rho^j(i)$, $j = \overline{0, 1, \dots, n}$.

Шаг 4. Элементы производящей матрицы $\tilde{Q} = \|q(\langle j \rangle, \langle j' \rangle)\|$ укрупненной модели определяются так:

$$q(\langle j \rangle, \langle j' \rangle) := \sum_{\substack{(i, j) \in S_j \\ (i', j') \in S_{j'}}} q((i, j), (i', j')) \rho^j(i).$$

Шаг 5. Определяется стационарное распределение укрупненной модели $\pi(<j>, <j>, <j> \in \tilde{S}$, где $\pi(<j>, <j>)$ означает стационарную вероятность состояния $<j> \in \tilde{S}$. При этом укрупненная модель представляет собой одномерная ЦМ с производящей матрицей \tilde{Q} .

Шаг 6. Стационарное распределение исходной двумерной цепи при-ближенно определяется так:

$$p(i, j) \equiv \rho^j(i) \pi(<j>), (i, j) \in S.$$

Замечание В2. На шаге 3 данного алгоритма предполагается, что существует стационарное распределение внутри классов $S_j, j=0, 1, \dots, n$. Для конечномерных цепей его существование следует из строгой непре-рывности исходной цепи по первой компоненте; и на шаге 5 существова-ние стационарного распределения укрупненной модели следует из сла-бой непрерывности исходной цепи по второй компоненте. Для бесконеч-номерных цепей требуются дополнительные условия, обеспечивающие эргодичность изучаемых цепей.

Очень важно отметить, что при использовании данного алгоритма не требуется генерация всего ФПС и решение соответствующей СУР ог-ромной размерности со сложной структурой матрицы коэффициентов. Более того, если использовать данный алгоритм, то на шагах 3 и 5 вы-числение соответствующих стационарных распределений одномерных цепей в большинстве случаев осуществляется с помощью хорошо из-вестных стандартных формул. Последнее означает, что вычислительная сложность разработанного алгоритма будет достаточно низкой по срав-нению с известными алгоритмами.

Заключение и комментарий к вводной части

Математическая теория телеграфика достаточно подробно изложена в много-численных монографиях и учебниках [1, 3, 6, 7, 9, 13, 15, 16, 21]. Раз-личные аспекты применения методов и моделей теории многопоточковых СМО в системах телеграфика также подробно исследованы в литературе [2, 4, 5, 10–12, 19].

Рассматриваемый в данной главе алгоритм вычисления нормирую-щей константы, основанный на двумерных рекуррентных уравнениях (В.11), (В.12) был предложен в работе [14]; алгоритм вычисления норми-рующей константы, основанный на одномерных рекуррентных уравне-ниях (В.32), (В.33), независимо друг от друга был предложен в работах [17, 20]. Разработанный в п. В2 приближенный алгоритм расчета двумер-ных ЦМ основан на принципах теории фазового укрупнения марковских и полумарковских систем [8]. Этот метод широко используется в после-дующих частях книги для исследования конкретных моделей систем те-леграфика.

Литература к вводной части

1. Башарин Г.П. Лекции по математической теории телеграфика. – М.: РУДН, 2004. – 190 с.
2. Башарин Г.П., Бочаров П.П., Коган Я.А. Анализ очередей в вычислительных сетях. Теория и методы расчета. – М.: Нау-ка, 1989. – 336 с.
3. Вишнеvский В.М. Теоретические основы проектирования компьютерных сетей. – М.: Техносфера, 2003. – 512 с.
4. Гнеденко Б.В., Коваленко И.Н. Введение в теорию массового обслуживания. – М.: Наука, 1987. – 336 с.
5. Дудин А.Н., Клименок В.И. Системы массового обслужива-ния с коррелированными потоками. – Минск: БГУ, 2000.
6. Ершов В.А., Кузнецов Н.А. Мультисервисные телекоммуни-кационные сети. – М.: МГТУ им. Н.Э. Баумана, 2003. – 432 с.
7. Корнышев Ю.Н., Пшеничников А.П., Харкевич А.Д. Теория телеграфика. – М.: Радио и связь, 1996. – 272 с.
8. Королюк В.С. Стохастические модели систем. – К.: Наук. думка, 1989. – 203 с.
9. Лагутин В.С., Степанов С.Н., Телеграфик мультисервисных сетей связи. – М.: Радио и связь, 2002. – 320 с.
10. Меликов А.З., Пономаренко Л.А., Рюшин Н.А. Математи-ческие модели много-поточковых систем обслуживания. – К.: Техника, 1991. – 265 с.
11. Назаров А.А. Асимптотический анализ марковизуемых сис-тем. – Томск, ТГУ, 1991.
12. Яшков С.Ф. Анализ очередей в ЭВМ. – М.: Радио и связь, 1989. – 216 с.
13. Н. Akimaru, К. Kawashima. Teletraffic. Theory and applications. – London: Springer Verlag, 1993. – 225 p.
14. Buzen J.P. Computational algorithms for closed networks with ex-ponential servers // Commun.Ass.Comput.Mach., 1973. – vol.16, no.9. – pp. 527–531.
15. Handbook “Teletraffic Engineering”, ITU-D, Study Group 2, Question 16/2. Geneva, January 2005.
16. Kaj I. Stochastic modelling in broadband communications sys-tems.–Philadelphia, 2002.–177 p.
17. Kaufman J.S. Blocking in shared resource environment // IEEE Trans. on Commun., 1981. –vol.10, no.10. – pp.1474–1481.
18. Kingman J.F.C. Markov population processes // J.Appl.Prob., 1969. – vol.6, no.1. – pp.1–18.

19. Neuts M.F. Matrix-geometric solution in stochastic models: An algorithmic approach. – Baltimore, John Hopkins University Press, 1981.
20. Roberts J.W. A service system with heterogeneous user requirements – applications to multi-service telecommunication systems // Performance of data communication systems and their applications. G.Pujole (Editor), North-Holland Publ.Co. –1981. – pp.423–431.
21. Ross K.W. Multiservice loss models for broadband communication networks. London: Springer Verlag, 1995. – 343 p.

ЧАСТЬ I

ИССЛЕДОВАНИЕ МОДЕЛЕЙ ТРАДИЦИОННЫХ БЕСПРОВОДНЫХ СОТОВЫХ СЕТЕЙ СВЯЗИ

В последнее десятилетие “индустриальный” период развития мирового сообщества переходит в “информационный”. Движению в этом направлении в последнее время способствует, а во многом его и определяет, повсеместное использование цифровой формы передачи информации в мобильных сетях связи и применение интеллектуальных средств обработки запросов пользователей. Технологически такая тенденция обеспечивается значительными успехами в исследовании и разработке средств и технологий связи для мобильных сетей и соответствующих систем коммутации и существенным продвижением в производстве элементной базы вычислительной техники, позволяющей разрабатывать более совершенные средства автоматизированного контроля и управления информационными потоками в мобильных сетях. Немаловажное значение имеют постоянно возрастающие потребности пользователей в услугах систем связи как по объему и качеству, так и по их разнообразию.

В конце прошлого века мы стали свидетелями очень быстрого развития и внедрения мобильных сетей связи. Первое и второе поколение мобильных сетей насчитывают на сегодняшний день более миллиарда пользователей по всему миру. Такому развитию также способствовала парадигма “коммуникации везде” вкупе с повсеместным распространением и доступностью таких высоко технологичных потребительских продуктов, как персональные переносные компьютеры, органайзеры, пейджеры, коммуникаторы, которые к тому же больше не являются ориентированными только на речь. Увеличиваются требования и к мультимедийным сервисам в мобильных сетях.

В этой части книги исследуются модели традиционных сетей сотовой связи, предназначенные, главным образом, для передачи речевых сообщений. Эти сети, как это отмечено выше, в настоящее время доминируют над всеми остальными сетями подобного типа и их изучение особенно полезно с методологической точки зрения, так как исследования моделей мультимедийных сетей базируются на результатах, полученных при изучении традиционных сетей.

МЕТОДЫ ИССЛЕДОВАНИЯ ХАРАКТЕРИСТИК БЕСПРОВОДНЫХ СОТОВЫХ СЕТЕЙ СВЯЗИ

1.1. Организация процессов обработки вызовов в беспроводных сетях сотовой связи

Рассмотрим кратко историю развития беспроводных сетей связи и вопросы организации процессов обработки вызовов в таких сетях.

Обращают внимание три ключевых этапа в истории развития телекоммуникационных технологий:

Этап 1. Появление автоматизированных телефонных станций (в конце 19-го века).

Этап 2. Переход к цифровым телекоммуникационным системам (начиная с 1970 по 1990-е года).

Этап 3. Интеграция технологий с коммутацией каналов и пакетов (начинающаяся с 1990 и продолжающаяся по сегодняшний день).

Указанные выше этапы характерны также и для мобильных систем. Мобильные коммуникации берут начало с 1861 года, когда Дж.М.С.Максвелл в Королевском Колледже в Лондоне предложил математическую теорию электромагнитных волн. Позже эта теория практически была продемонстрирована Н.Герцем в 1887 году в университете города Карлсруе. Несколькими годами позже итальянский изобретатель Маркони построил и летом 1895 года в университете Болоньи продемонстрировал беспроводное коммуникационное устройство. Это был первый радиотелеграф. Официально это и ознаменовало начало эры беспроводных коммуникаций. Для нужд гражданского населения использование беспроводных технологий началось в США с 2 МГц наземного мобильного радиотелефонного сервиса, разработанного в 1921 году для департамента полиции Детройта с целью контроля передвижения полицейских машин. Вскоре были осознаны преимущества беспроводных коммуникаций, однако распространению мешали ограничения на радиоресурсы (каналы) в спектрах низких частот, используемых в то время для связи. Пришла необходимость использования высоких частот. Ключевым прорывом в этом направлении стало изобретение Армстронгом частотной модуляции (Frequency Modulation, FM), которое сделало возможным качественную двунаправленную связь. Расширение FM технологии для большего количества пользователей требовало больших пропускных способностей каналов. Решением явилось разбиение области обслуживания на несколько более мелких областей, именуемых сотами, и использование одних и тех же частот каналов в разных сотах. Такую сотовую концепцию предложил в Bell Laboratories в 1947 году Д.Х.Ринг. С созданием сотовых сетей возникла одна из самых сложных проблем, так называемое явление хэндовер (*handover*) – переход абонента между сотами.

Такой переход должен был быть в идеале незаметным для подвижного абонента. Первый такой успешный звонок был реализован компанией AT&T в 1970 году в их аналоговой сотовой сети AMPS (Advanced Mobile Phone Services), которая работала в диапазоне частот 800 МГц. Эта система была коммерчески внедрена лишь в 1983 году.

В Европе первые мобильные сети стартовали в Скандинавии, где возникла проблема связи для населения, рассредоточенного по всем трудно доступным областям этих северных стран. Первая такая система возникла в 1978 году. Однако реальный бум пришелся на систему Nordic Mobile Telephony (NMT) в 1981 году. Все эти системы ознаменовали начало первого поколения мобильных сотовых сетей связи. Параллельно, в Великобритании ввели в строй свою собственную сеть TACS (Total Access Communication System), в Германии появилась C-system, в Японии – модификация TACS – JTACS (Japanese TACS). Основной общей чертой этих сетей было то, что все они были аналоговыми и не совместимыми с сетями других стран.

Таким образом, каждая страна в Европе фактически имела свою систему и набор стандартов, однако они были абсолютно несовместимы друг с другом, что, конечно же, накладывало ограничения на передвижения мобильных пользователей и сегментировало рынок для мобильных аппаратов. Европейские страны осознали эту проблему, и было решено создать панъвропейскую наземную мобильную систему. Поэтому в 1982 году формируется специальная исследовательская группа Groupe Spéciale Mobile (GSM). Позднее, в 1989 году проблема стандартизации GSM передается в Европейский Институт Стандартизации по Телекоммуникациям (European Telecommunication Standards Institute, ETSI).

Создание первой тестовой сети GSM в 1991 году ознаменовало начало развития мобильных сетей второго поколения. В этом же году GSM поменяло название на Global Systems for Mobile Communications. Вскоре GSM превалировало на рынке мобильных систем, обслуживая 700 миллионов пользователей и более 400 операторов по всему миру. Воплощая концепцию цифровых сетей интегрального обслуживания (ISDN), GSM изначально разрабатывалась как полностью цифровая система. Ещё одной отличительной чертой GSM стала возможность роуминга пользователей, т.е. GSM позволяла пользователям одной GSM сети пользоваться услугами других GSM операторов. Это и сделало GSM лидером сетей второго поколения.

Параллельно с GSM, Япония разрабатывает свою собственную похожую сеть PDC (Personal Digital Communications). В Северной Америке запускается D-AMPS (Digital AMPS) – цифровое дополнение к AMPS. Две эти системы занимают второе место в мире среди сетей второго поколения. Очень скоро возникла необходимость увеличения емкости цифровых мобильных сотовых сетей связи второго поколения и, как следствие, нового подхода к технологиям сотовых сетей. В 1993 году

в Соединенных Штатах вводят в строй стандарт CDMA (Code Division Multiple Access), предложенный компанией Qualcomm.

Таким образом, все сети второго поколения, которые сейчас занимают лидирующее положение в мире, изначально ориентированы на речь. Однако, они позволяют предоставлять и ряд других сервисов, включая низкоскоростную передачу данных (модемную), факс, короткие сообщения и так называемые дополнительные услуги, такие как предоставление номера звонящего, ограничения на звонки, конференц-звонок, перенаправление и т. д.

Третье поколение мобильных систем появилось в 2000-е годы (первые коммерческие системы стартовали в 2002 году в Японии и Южной Корее). Они созданы с целью поддержки доступа в Интернет и пакетной коммутации каналов кроме традиционной канальной коммутации. Эти системы допускали скорость передачи данных от 144 Кбс для быстро движущихся абонентов и до 2 Мбс для неподвижных абонентов. Будущие мобильные сети уже будут поддерживать IP соединение точка-точка, так называемые беспроводные IP сети.

Дальнейшее развитие мобильных сетей идет по пути объединения с наземными классическими телефонными сетями и Интернет (IP). Такой процесс называется конвергенцией. На арену также выходит понятие мобильная IP сеть.

Сеть связи будущего должна предоставить абоненту весь спектр информационных услуг, включая доступ к базам данных, дистанционное обучение, телеконференции, телебизнес, телемагазины, телемониторинг места проживания, состояния здоровья, получение в интерактивном режиме аудио- и видеoinформации и т. д. Необходимо отметить, что распространение новых услуг меняет характер нагрузки, передаваемой по цифровым линиям. Речевой трафик с коммутацией каналов вытесняется нагрузкой новых приложений, использующих протоколы пакетной коммутации.

Теперь рассмотрим архитектуру GSM сети и на примере GSM опишем организацию процессов обработки вызовов в беспроводных сотовых сетях связи.

Сеть GSM состоит из следующих подсистем: подсистема базовых станций (Base Station Subsystem, BSS), подсистема сети и коммутации (Network and Switching Subsystem, NSS) и подсистема оперативного управления и поддержки (Operation and Support Subsystem, OSS), как показано на рис. 1.1.

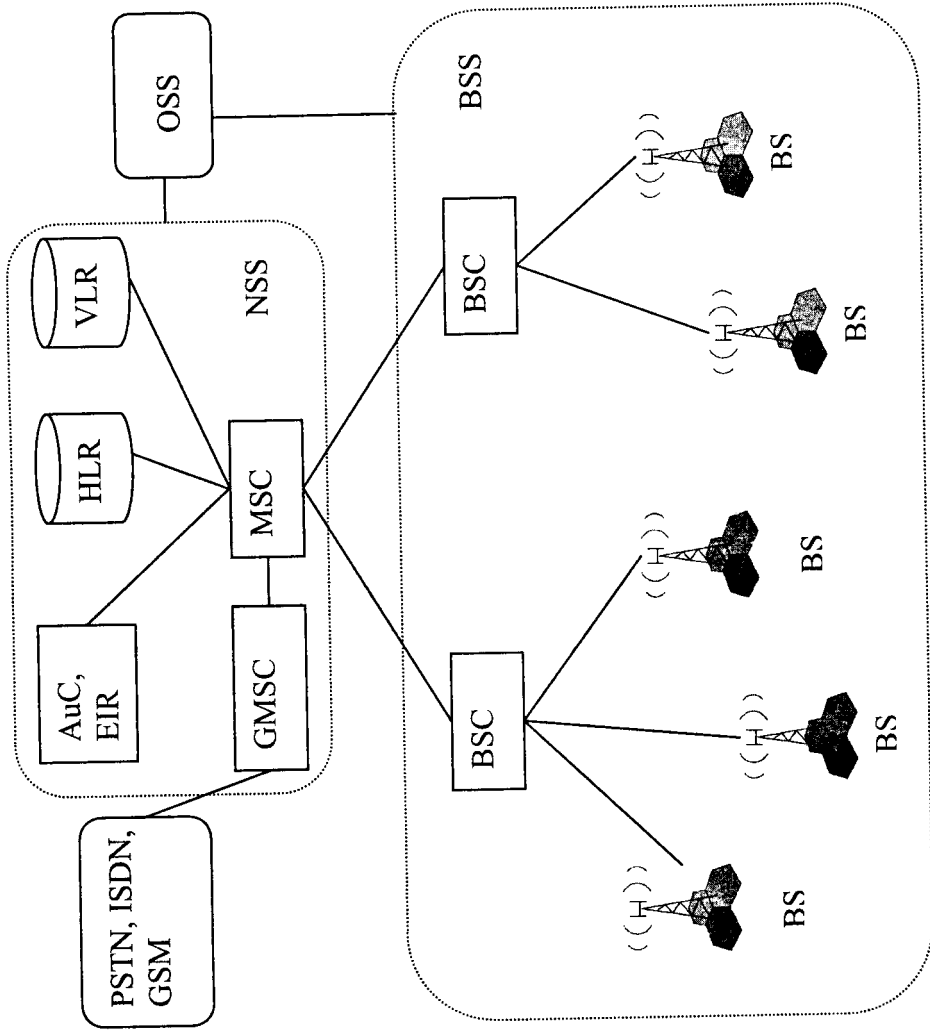


Рис. 1.1. Структура сети GSM

Подсистема BSS состоит из базовых станций передачи (Base Transceiver Stations, BTS или BS) и контроллера базовых станций (Base Station Controller, BSC). Функция BSS состоит в предоставлении средств связи между мобильными абонентами (Mobile Subscriber, MS) и подсистемой сети и коммутации. Базовые станции (BS) это фактически точки радио доступа для мобильных абонентов, и они имеют один или более передатчиков. Технология GSM – это комбинация двух технологий доступа: множественный доступ с разделением частот (Frequency Division Multiple Access, FDMA) и множественный доступ с разделением времени (Time Division Multiple Access, TDMA). Первые GSM системы функционировали на диапазоне частот 900 МГц. В каждом направлении передачи выделяется спектр частот в 25 МГц.

Технология FDMA используется для разбиения доступных 25 МГц спектра на 124 несущих частоты шириной в 200 kHz. Затем уже каждая час-

тота делится на 8 временных слотов (квантование по времени) с использованием TDMA технологии (см. рис. 1.2). Двухнаправленная связь достигается путем выделения тех же самых временных слотов на частотах, отстоящих друг от друга на диапазон в 45 МГц. Таким образом, пара несущих частот (в направлении вниз к MS – прием, и от MS – передача) называется абсолютным номером радиочастотного канала (Absolute Radio Frequency Channel Number, ARFCN). Например, ARFCN = 1 использует частоту 890.2 МГц для передачи и 935.2 МГц для приема, тогда как, скажем ARFCN 124, использует 915 МГц для передачи и 960 МГц для приема. Таким образом, в GSM спектр частот для передачи представляет собой диапазон от 890 МГц до 915 МГц, а спектр частот приема диапазон от 935 до 960 МГц.

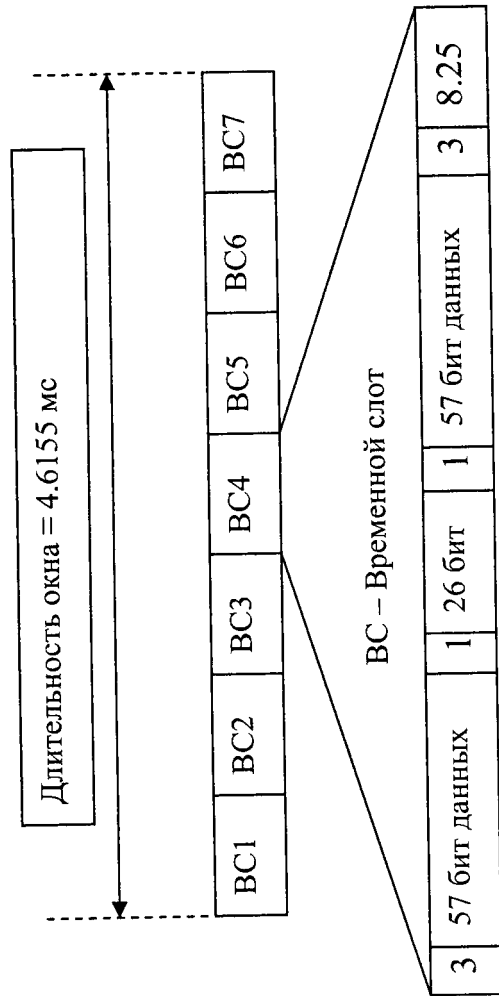


Рис 1.2. Разбиение частоты на временные слоты

Таким образом, сота имеет одну или несколько несущих частот. Пара временных слотов на одной из несущих частот в соте всегда выделяется для контрольной информации, а все остальные используются для передачи трафика. Несколько логических каналов могут быть мультиплексированы в один логический канал. Обычно в GSM в два временных слота, используемых для контрольных логических каналов, мультиплексировается до 10 логических каналов. Сегодня GSM системы используют кроме 900 МГц диапазона также и диапазон 1800 МГц, а в Америке – и 1900 МГц диапазоны. Это связано с регулированием частотных спектров в различных странах.

BSC производит мониторинг и контроль нескольких базовых станций (число их варьирует в зависимости от реализации и может доходить до нескольких сотен станций). Главными функциями BSC являются управление сотами, контроль базовых станций, а также функции комму-

тации. Физически BSC может располагаться рядом с центром коммутации (Mobile Switching Centre, MSC) или же в отдалении, когда несколько BSC соединены с одним центром коммутации.

Подсистема NSS включает в себя коммутацию (маршрутизацию), а также управление местоположением. Она состоит из MSC, базы данных для управления местоположением (реестра домашнего местоположения – Home Location Register, HLR и реестра местоположения посетителей – Visitor Location Register, VLR), шлюзовой MSC (Gateway MSC, GMSC), центра аутентификации (Authentication Center, AuC) и реестра идентификации аппаратов (Equipment Identity Register, EIR).

GMSC представляет собой интерфейс между домашней сетью и остальными сетями, включая классические телефонные сети (PSTN). HLR хранит данные мобильных абонентов, принадлежащих оператору вне зависимости от того, находятся ли они сейчас в домашней сети или же в сети других операторов (т.е. в роуминге). Эта информация представляет собой набор постоянных статических данных о MS, таких как уникальный номер MS (International Mobile Subscriber Identity, IMSI), так называемый номер мобильной станции (Mobile Station ISDN, MS-ISDN), ключ аутентификации (необходимый для защиты сети от несанкционированного доступа), дополнительные сервисы MS и некоторые другие временные данные. VLR содержит данные о мобильных абонентах, которые на данный момент находятся в зоне действия его центра коммутации. Статические данные VLR такие же, как и в HLR, в то время как временные данные слегка отличаются. Таким образом, VLR содержит данные всех домашних пользователей и всех пользователей других GSM сетей, находящихся в данный момент в зоне его действия. Также VLR следит за местоположением мобильного абонента. В свою очередь, MS хранят свои постоянные и текущие данные в специальных картах, называемых SIM (Subscriber Identification Module), которые используются непосредственно в мобильных аппаратах.

Как бы ни были развиты средства средства связи, как бы ни были высоки скорости передачи информации потоков в цифровых системах связи, это не исключает возможности образования узких мест в точках их концентрации. Наличие узкого места означает нехватку того или иного ресурса для обслуживания вызовов с заданным качеством. Отметим, что возможность появления узких мест заложена уже в основных принципах построения любых телекоммуникационных систем, в том числе и сотовых мобильных систем (а в дальнейшем мы покажем, что сотовые мобильные системы имеют и свои характерные проблемы), поскольку из экономических и статистических соображений количество коммутационного и линейного оборудования станций и узлов сети всегда проектируется исходя из предположения о том, что реальные потоки существуют не меньше, чем теоретически возможные. Это допущение может не выполняться на некоторых участках больших сетей из-за изменения профиля нагрузки. Поэтому, даже если ресурс выбран с многократным запа-

сом, его достаточность нельзя гарантировать, особенно когда речь идет о больших, динамично развивающихся сетях.

Необходимо отметить, что для проведенных здесь исследований конкретный тип технологии доступа или коммутации не имеет значения. Хотя исследования проводятся на примере процессов, происходящих в сетях GSM, созданные модели и подходы к их исследованию могут быть успешно применены и в других случаях, т.к. объект исследования тесно связан с фундаментальным для беспроводных сотовых сетей явлением. В связи с этим условимся называть беспроводным каналом объем пропускной способности, выделенный MS в конкретный момент времени, независимо от технологии доступа, будь то FDMA, TDMA, CDMA или их комбинация. При этом это вовсе не означает, что емкость соты обслуживается только канальной коммутацией.

Как было отмечено выше, беспроводные сети имеют свои специфические характеристики. Два наиболее важных различия между беспроводными и проводными сетями — это мобильность пользователей и битовые ошибки в беспроводных каналах, зависящие от местоположения. Эти две специфические особенности создают совершенно другие условия для требований к показателям качества обслуживания (Quality of Service, QoS).

Битовые ошибки в радиоинтерфейсе могут возникнуть вследствие нескольких причин, а именно: интерференции, шумов, затухания и экранирования. Затухание — одна из главных характеристик распространения сигнала в беспроводных каналах. С точки зрения шумов и экранирования затухание нежелательно, но именно затухание сделало возможным, избегая появления интерференции, использование одного и того же частотного ресурса. Именно затухание естественно ограничивает географическую область, делая возможным реализацию сотового принципа при построении беспроводных сетей.

Экранирование — это следствие препятствий на пути распространения радиосигнала (например, нет прямой видимости между MS и BS). Более того, из-за отражений сигнала от окружающих объектов различные части одного и того же сигнала могут достигать приемника различными путями. Этот эффект нежелателен в сетях GSM.

Интерференция — это следствие использования одних и тех же частот или же пограничных частот в одной и той же соте или соседних сотах.

Эти характеристики беспроводной среды создают более высокий коэффициент битовых потерь в беспроводных каналах по сравнению с проводными. Более того, коэффициент потерь также зависит от местоположения MS из-за инерционности движения и от текущего состояния канала в конкретном месте внутри соты. Таким образом, мы видим, как состояние беспроводной среды может влиять на показатели качества обслуживания вызовов.

Для того, чтобы позволить большому числу пользователей использовать зововую ту или иную технологию радиодоступа, необходимо использовать сотовый принцип (см. рис.1.3).

Беспроводная сотовая сеть состоит из точек радиодоступа, называемых базовыми станциями, каждая из которых покрывает определенную географическую область. Из-за того, что мощность радиоволн уменьшается с расстоянием (происходит затухание сигнала), становится возможным использовать одни и те же частоты в разных сотах, однако во избежание интерференции этот процесс необходимо планировать. Для лучшего использования частотного ресурса существующие несущие частоты группируются, и число сот, в которых эта группа частот используется, определяется так называемый фактор повторения частот. Например, в GSM системах есть несколько шаблонов повторения частот, таких как 3/9, 4/12 и 7/21. Запись x/y означает, что все доступные несущие частоты разделены на группы по y частот и распределены в x различных сотах. Далее такое разделение повторяется в остальных сотах (см. рис. 1.3). В густонаселенных районах (с большим количеством MS), следовательно, должны использоваться соты поменьше (пикосоты) из-за ограниченный объемов и фактора повторения частот.

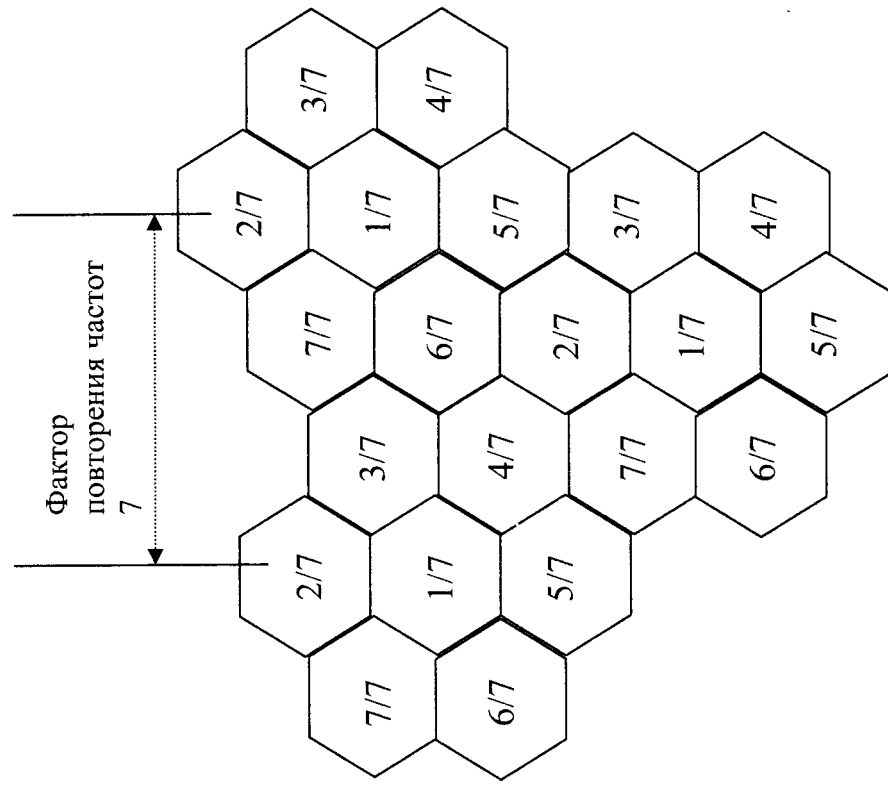


Рис. 1.3. Пример планирования частот

Для решения проблемы нехватки ресурсов можно использовать несколько подходов. Во-первых, если есть такая возможность, то можно просто увеличить количество ресурса (пропускную способность) за счет внедрения дополнительных линий или ресурсов связи или же за счет перехода на новую технологию связи. Однако при всей своей эффективности применения данного подхода требует больших капиталовложений и больших объемов исследований. В случае же с беспроводными сетями эта задача еще более усложняется, т.к. диапазон частотного спектра всегда ограничен. Поэтому число одновременных соединений в какой-то конкретной соте беспроводной сети очень ограничено возможностями технологии радиодоступа, тогда как в проводных сетях всегда есть возможность прокладывать новые линии связи.

Во-вторых, качество передачи нагрузки можно улучшить (здесь применимы и другие постановки задачи: гарантировать, выровнять и т.д.) и без увеличения скорости (пропускной способности) передачи, если воспользоваться какой-либо схемой преимущественного предоставления ресурса для сообщений части потоков (трафиков) или иначе стратегией управления доступом вызовов к сети (Call Admission Control, SAC). Реализация сформулированной идеи обычно осуществляется либо за счет уменьшения интенсивности обслуживаемой нагрузки, либо в результате задержки передачи сообщений (вызовов) тех потоков, которые по характеру передаваемой информации эту задержку допускают. Использование тех или иных стратегий доступа к сети не требует больших затрат и в противоположность первому из рассмотренных подходов данный метод можно считать оперативным и более оправданным средством решения проблемы нехватки ресурса.

В беспроводных сотовых сетях связи помимо инициированных новых потоков вызовов существует и дополнительный класс вызовов, требующих особого подхода. Это так называемые хэндовер-вызовы. Это явление присуще только беспроводным сотовым сетям связи.

Подвижность абонентов и сотовый принцип построения являются причинами такого фундаментального явления как хэндовер. Суть этого явления состоит в том, что MS, уже установивший связь с сетью (когда новый вызов успешно занял определенный ресурс сети, беспроводной канал), передвигаясь, проходит зоны между сотами и оказывается в области действия другой соты. С точки зрения новой соты — это хэндовер-вызов и, так как связь с MS уже установлена, переход обслуживания MS на новую соту должен произойти незамедлительно для него. Таким образом, наличие более приоритетное обслуживание хэндовер-вызовов в новой соте по отношению к новым вызовам в этой соте. Рассмотрим более детально это явление на примере GSM сети.

Решение о хэндовере может приниматься как мобильной станцией, так и BS при помощи мониторинга качества каналов. Если решение принято лишь мобильной станцией, она уведомляет об этом базовую станцию. В таком случае новую обслуживающую базовую станцию (следова-

тельно, соту) предлагает или сама мобильная станция, или же центр коммутации. Если ее предлагает мобильная станция, о кандидатуре новой базовой станции уведомляется MSC. Необходимо отметить, что все решения, принимаемые мобильной станцией, могут основываться лишь на измерениях радиоканалов, тогда как MSC может использовать несколько критериев, таких, например, как распределение трафика (нагрузки) в той или иной области.

С точки зрения сети, определение момента, когда необходимы хэндовер и его оперативное осуществление, является достаточно сложной задачей. Временные ошибки на каналах между MS и BS, о которых было сказано выше, могут возникнуть и внутри одной соты. Это означает, что необходимо определенное время, пока BS заранее сможет определить по уменьшению мощности сигнала, что абонент удалается от BS. С другой стороны, если действительно есть необходимость в хэндовере, то процесс должен быть инициирован как можно быстрее, иначе оборвется уже установившаяся связь, т.е. произойдет вынужденное прерывание разговора. Для того, чтобы определить необходимость хэндовера, MS должна проводить измерения качества канала, которым она сейчас пользуется, а также широкополосных каналов соседних сот. Различные стандарты сотовых сетей определяют различные процедуры такого измерения.

В стандарте GSM мобильная станция постоянно измеряет качество сигналов и периодически передает эти данные BS. В свою очередь, BS получает или только данные об измерениях, которые она должна оценить и принять решение о необходимости хэндовера, или же данные об измерениях и кандидатуру новой BS, выбранную самой MS. В первом случае BS инициирует запрос на хэндовер, если это необходимо, и посылает этот запрос центру коммутации. Затем центр коммутации выбирает наиболее подходящую BS для продолжения обслуживания вызова. Во втором случае базовая станция просто посылает MSC запрос на хэндовер в конкретную новую BS. В обоих случаях MSC уведомляет новую BS о запросе хэндовера. Новая BS, в зависимости от применяемой стратегии доступа, может этот хэндовер-вызов либо принять, либо блокировать, либо поставить в очередь. Она информирует MSC о статусе этого запроса. В зависимости от ответа, MSC разрешает, задерживает или вынужденно прерывает хэндовер-вызов (см. рис.1.4). Движение от BS1 к BS2 предполагается по прямой. Хэндовер должен произойти в период $[t_0, t_1]$.

1.2. Методы расчета и оптимизации характеристик беспроводных сетей

Беспроводные сети сотовой связи являются наиболее сложными телекоммуникационными системами с точки зрения разработки их адекватных математических моделей. Это, прежде всего, объясняется наличием в них эффекта хэндовера из-за их сотовой топологии.

В классических проводных коммуникационных сетях в период всего времени обслуживания вызов занимает лишь один канал (цифровой) системы, т.е. в них понятие «время обслуживания» связано лишь с вызовом. Однако в беспроводных сетях содовой связи на протяжении всего периода обслуживания вызов несколько раз может занимать каналы различных сот сети, т.е. в беспроводных сетях время занятия каналов и время обслуживания не являются одинаковыми.

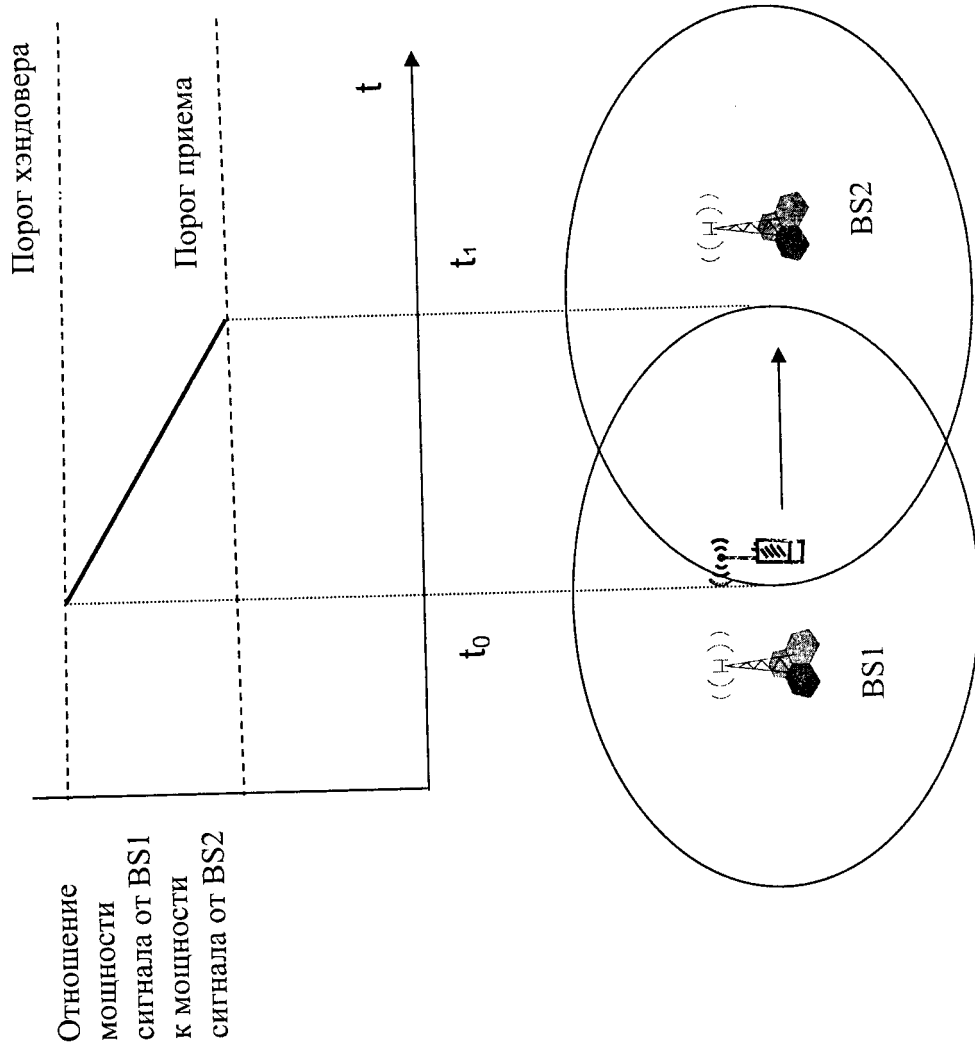


Рис. 1.4. Порог хэндовера и приема

Более того, если в проводных коммуникационных сетях с целью улучшения характеристик процессов обслуживания вызовов можно три-

виальным образом увеличить число линий (каналов), то в беспроводных сетях этот путь в принципе не может быть использован, так как в них спектр передачи всегда является ограниченным.

В связи с ограниченностью спектра передачи в беспроводных сетях очень актуальными являются проблемы распределения общего спектра между сотами сети. Единича беспроводного спектра, необходимого для обслуживания одиночного пользователя, называется каналом. Например, в TDMA временные слоты рассматриваются как каналы. Здесь рассматриваются сети с коммутацией каналов, таких как речевая коммуникация, в которых требуемая полоса для каждого соединения является фиксированной.

Существует два способа решения проблемы распределения каналов: фиксированное распределение каналов (Fixed Channel Allocation, FCA) и динамическое распределение каналов (Dynamic Channel Allocation, DCA).

При использовании способа FCA все каналы перманентно (постоянно) распределяются между сотами, т.е. клиенты в пределах определенной соты могут использовать только каналы этой соты. При этом для максимизации эффективности повторно используется каналы, одинаковое множество каналов повторно используется в сотах, находящихся в определенных расстояниях друг от друга. Это расстояние называется минимальным расстоянием повторного использования (Minimum Reuse Distance, MRD) и обозначается Δ . Пусть расстояние между базовыми станциями соты X и Y обозначается через $d(X, Y)$. Тогда одинаковые каналы могут быть использованы в сотах X и Y , если выполняется условие $d(X, Y) \geq \Delta$. Выполнение данного условия необходимо для предотвращения интерференции.

Отметим, что благодаря простоте реализации схемы FCA широко используются в существующих сетях содовой связи.

Более эффективным способом решения проблемы распределения каналов между сотами является использование схемы DCA. При использовании данной схемы, каналы фиксированным образом не привязываются к конкретным сотам, а назначаются при каждом новом запросе. Следовательно, характеристики соты оказываются лучшими при данной схеме по сравнению со схемой FCA за счет использования флуктуации трафика в различных сотах. Однако схема DCA является очень сложной в реализации, так как необходимо избежать нарушения требования по MRD, т.е. требуется координация использования каналов в пределах требуемых расстояний повторного использования. Более того, при использовании схемы DCA требуется максимизировать эффективность повторного использования, а также решить ряд достаточно сложных задач по глобальной оптимизации.

В литературе также известны публикации, в которых изучаются гибкие схемы распределения каналов, в которых при распределении каналов одновременно используются обе схемы FCA и DCA. В данной

книге используются модели беспроводных сетей, в которых используются схема FSA.

Можно утверждать, что достаточно адекватные математические модели процессов обработки вызовов в беспроводных сетях сотовой связи могут быть разработаны на основе использования теории сетей массового обслуживания с разнотипными заявками и произвольной топологией. В доступной литературе подобные модели освещены очень скудно. Это объясняется тем, что, несмотря на изысканность этих моделей, они могут быть использованы лишь для сетей очень малой размерности и при выполнении ряда упрощающих предположений, не соответствующих реальным условиям функционирования беспроводных сетей.

В связи с этими обстоятельствами в подавляющем большинстве работ исследуются модели изолированной соты. Поскольку в настоящей работе также исследуются подобные модели, то вкратце рассмотрим обзор основных работ в этом направлении.

Классической является работа [10]. В ней предполагается, что соты сети представляют собой правильные шестиугольники (см. рис.1.3.) и показано, что интенсивность новых вызовов (o -вызовов) в соте (λ_0) определяется так: $\lambda_0 = \frac{3\sqrt{3}}{2} R^2 \lambda_a$, где R – радиус сети (т.е. максимальное

расстояние от центра соты до ее границы), λ_a – интенсивность o -вызовов на единицу площади всей области.

Интенсивность хэндовер-вызовов (h -вызовов) обозначается λ_h (почти во всех исследованиях предполагается, что эта величина является известной).

В работе получены функции распределения случайных величин T_0 и T_h , где T_0 (T_h) – время, в течение которого мобильный o -вызов (h -вызов) остается внутри соты.

Исследуются две схемы организации разнотипных вызовов. В первой из них не разрешается организация очереди, и в целях защиты h -вызовов от частых потерь используется схема их приоритизации на основе стратегии резервирования каналов (Guard Channels). Это означает, что поступивший h -вызов теряется лишь тогда, когда все N каналов соты являются занятыми, а поступивший o -вызов принимается лишь тогда, когда число свободных каналов соты больше, чем g , $0 \leq g \leq N-1$, т.е. o -вызов блокируется, если число свободных каналов меньше или равно g .

Приоритетное обслуживание h -вызовов перед o -вызовами объясняется тем, что прерывание уже начавшегося разговора является более нежелательным, нежели блокировка нового вызова (см. §1.1). Предполагается, что время обслуживания o -вызовов имеет одинаковое показательное распределение со средним μ^{-1} . Поскольку показательное распределение не имеет памяти, то времена обслуживания h -вызовов также имеют одинаковое показательное распределение с тем же средним μ^{-1} .

Считается, что o -вызовы (h -вызовы) подчиняются пуассоновским законам распределения со средними λ_0 (λ_h). Состояние соты обозначается скалярной величиной j , которая указывает общее число разнотипных вызовов в базовой станции соты. Тогда функционирование соты описывается классическим процессом размножения и гибели (Birth-Death Processes, BDP). Граф этого процесса показан на рис 1.5.

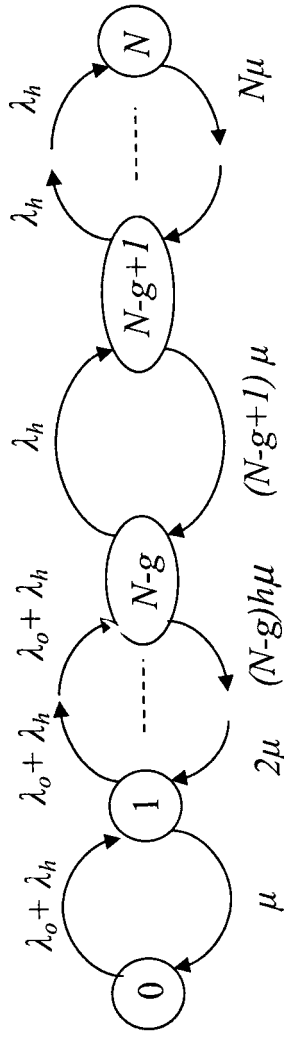


Рис. 1.5. Граф модели при использовании стратегии резервирования каналов

Пусть p_j обозначает стационарную вероятность состояния j , $j = 0, \dots, N$. С использованием стандартных формул для BDP имеем:

$$p_j = \begin{cases} \frac{\lambda_0 + \lambda_h}{\mu} p_{j-1}, & \text{если } j = 1, N-g, \\ \frac{\lambda_h}{\mu} p_{j-1}, & \text{если } j = \overline{N-g+1, N}, \end{cases} \quad (1.1)$$

Нормирующее условие имеет вид:

$$\sum_{j=0}^N p_j = 1. \quad (1.2)$$

Тогда с учетом (1.1) и (1.2) ищется стационарное распределение ($p_j, j = 0, \dots, N$) рассматриваемой системы:

$$p_j = \begin{cases} \frac{(\lambda_0 + \lambda_h)^j}{j! \mu^j} p_0, & \text{если } j = \overline{1, N-g}, \\ \frac{(\lambda_0 + \lambda_h)^{N-g} \lambda_h^{j-(N-g)}}{j! \mu^j} p_0, & \text{если } j = \overline{N-g+1, N}, \end{cases} \quad (1.3)$$

где

$$P_0 = \left(\sum_{k=0}^{N-g} \frac{(\lambda_0 + \lambda_h)^k}{k! \mu^k} + \sum_{k=N-g+1}^N \frac{(\lambda_0 + \lambda_h)^{N-g} \lambda_h^{k-(N-g)}}{k! \mu^k} \right)^{-1}. \quad (1.4)$$

После нахождения стационарного распределения $(p_j, j = 0, \dots, N)$ с использованием известной теоремы PASTA (Poisson Arrivals See Time Averages) находятся характеристики соты. Основными характеристиками в данной работе являются вероятность вероятности блокировки o -вызовов (P_o) и вероятность потери h -вызовов (P_h). Эти характеристики вычисляются так:

$$P_o(N, g) = \sum_{j=N-g}^N P_j, \quad (1.5)$$

$$P_h(N, g) = P_N. \quad (1.6)$$

Во второй приоритетной схеме указанной работы наряду с резервными каналами для h -вызовов допускается также образование бесконечной очереди таких вызовов, при этом h -вызовы в очереди являются нетерпеливыми, т.е. время ожидания в очереди h -вызовов является случайной величиной, подчиняющейся показательному распределению со средним γ . Граф соответствующего BDP показан на рис 1.6.

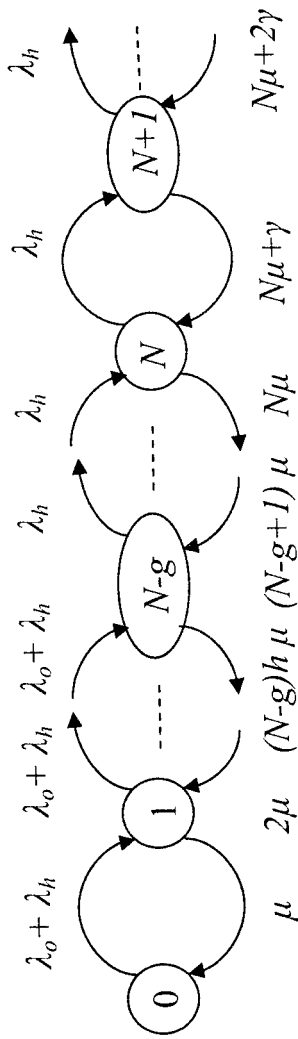


Рис. 1.6. Граф модели при использовании стратегии резервирования каналов и бесконечной очереди нетерпеливых h -вызовов

Стационарное распределение $(p_j, j = 0, 1, \dots)$ данного процесса определяется так:

$$P_j = \begin{cases} \frac{(\lambda_0 + \lambda_h)^j}{j! \mu^j} P_0, & \text{если } j = \overline{1, N-g}, \\ \frac{(\lambda_0 + \lambda_h)^{N-g} \lambda_h^{j-(N-g)}}{j! \mu^j} P_0, & \text{если } j = \overline{N-g+1, N}, \\ \frac{(\lambda_0 + \lambda_h)^{N-g} \lambda_h^{j-(N-g)}}{N! \mu^N \prod_{i=1}^{j-N} (N\mu + i\gamma)} P_0, & \text{если } j \geq N+1, \end{cases} \quad (1.7)$$

где

$$P_0 = \left(\sum_{j=0}^{N-g} \frac{(\lambda_0 + \lambda_h)^j}{j! \mu^j} + \sum_{j=N-g+1}^N \frac{(\lambda_0 + \lambda_h)^{N-g} \lambda_h^{j-(N-g)}}{j! \mu^j} + \sum_{j=N+1}^{\infty} \frac{(\lambda_0 + \lambda_h)^{N-g} \lambda_h^{j-(N-g)}}{N! \mu^N \prod_{i=1}^{j-N} (N\mu + i\gamma)} \right)^{-1} \quad (1.8)$$

Тогда, вероятность блокировки o -вызовов определяется так:

$$P_o = \sum_{j=N-g}^{\infty} P_j. \quad (1.9)$$

Вероятность потери h -вызовов находится из следующего равенства:

$$P_h = \sum_{k=0}^{\infty} P_{N+k} P_{j/k}, \quad (1.10)$$

где $P_{j/k} := \text{Prob} \{h\text{-вызов, поступивший в очередь в позиции } k, \text{ теряется}\}$,

$$P_{j/k} = 1 - \frac{\gamma}{N\mu + \gamma} \prod_{i=j}^k \left(1 - \frac{\mu}{N\mu + \gamma} \cdot 0.5^i \right) \quad (1.11)$$

В работе [9] предложены следующие быстрые рекуррентные формулы для вычисления P_o и P_h в модели без очереди (т.е. (1.5), (1.6)):

$$P_h(N-g+k, k) = \frac{P_h(N-g+(k-1), k-1)}{N} + \frac{P_h(N-g+(k-1), k-1)}{\alpha \cdot A}, \quad k = \overline{1, g}; \quad (1.12)$$

$$P_o(N-g+k, k) = \frac{N}{\alpha \cdot A} \frac{P_o(N-g+(k-1), k-1) + P_h(N-g+(k-1), k-1)}{N + P_h(N-g+(k-1), k-1)}, \quad k = \overline{1, g}; \quad (1.13)$$

где

$$A := \frac{\lambda_o + \lambda_h}{\mu}; \quad A_1 := \frac{\lambda_h}{\mu}; \quad \alpha := \frac{A_1}{A};$$

$$P_h(N-g, 0) := E_B(A, N-g);$$

$$P_o(N-g, 0) := E_B(A, N-g),$$

где $E_B(A, n)$ – В-формула Эрланга, т.е.

$$E_B(A, n) = \frac{A^n}{n!} \left(\sum_{k=0}^n \frac{A^k}{k!} \right)^{-1}.$$

В указанной работе также определены оптимальные значения числа резервных каналов (g) в различных постановках задач оптимизации. Так, например, при заданных значениях числа каналов (N) и нагрузочных параметров (λ_o, λ_h и μ) решена задача минимизации $P_o(g)$ при определенных ограничениях на $P_h(g)$. Более интересной является следующая задача: при заданных λ_o, λ_h и μ найти такие оптимальные значения N и g , чтобы

$$N \rightarrow \min \quad (1.14)$$

$$P_o(N, g) \leq P_{o0} \quad (1.15)$$

$$P_h(N, g) \leq P_{h0} \quad (1.16)$$

где P_{o0} и P_{h0} являются заданными параметрами.

В указанной работе предложен алгоритм решения задачи (1.14)–(1.16).

Модель соты с ограниченной очередью весьма терпеливых h -вызовов и резервными каналами исследованы в работе [26]. В ней показано, что функционирование модели описывается одномерным BDP и для вычисления P_o и P_h получены следующие формулы:

$$P_o = \sum_{j=N-g}^K P_j, \quad (1.17)$$

$$P_h = P_K, \quad (1.18)$$

где K – общая емкость соты (т.е. длина очереди равна $K - N$);

$$P_j = \begin{cases} \frac{(\lambda_o + \lambda_h)^j}{j! \mu^j} P_o, & \text{если } j = \overline{1, N-g}, \\ \frac{(\lambda_o + \lambda_h)^{N-g} \lambda_h^{j-(N-g)}}{j! \mu^j} P_o, & \text{если } j = \overline{N-g+1, N}, \end{cases} \quad (1.19)$$

$$P_o = \left(\sum_{j=0}^{N-g} \frac{(\lambda_o + \lambda_h)^j}{j! \mu^j} + \sum_{j=N-g+1}^N \frac{(\lambda_o + \lambda_h)^{N-g} \lambda_h^{j-(N-g)}}{j! \mu^j} + \sum_{j=N+1}^K \frac{(\lambda_o + \lambda_h)^{N-g} \lambda_h^{j-(N-g)}}{N! \mu^N (N\mu)^{j-N}} \right)^{-1} \quad (1.20)$$

С использованием (1.17)–(1.20) из модифицированной формулы Литтла находится среднее время ожидания в очереди h -вызовов (W_h):

$$W_h = \sum_{j=N}^K \frac{(j-N) P_j}{\lambda_h (1-P_d)}. \quad (1.21)$$

Модели процессов обработки в сотовых сетях при различных допущениях о длине очереди различных вызовов с применением спектральных методов и методов производящих функций были исследованы в работах [6, 17]. При этом использованы различные стратегии САС, основанные на модификации схемы резервирования каналов.

Задачи синтеза стратегии САС были рассмотрены в работе [15]. В ней решены три задачи. Первой из них является задача нахождения такой стратегии САС, при которой удается минимизировать некоторую линейную функцию потерь, оценивающую суммарные штрафы из-за потери различных вызовов. Показано, что оптимальная стратегия находится в классе стратегий, основанных на схеме резервирования каналов для h -вызовов. Во второй задаче требуется найти такую стратегию САС, при которой удается минимизировать вероятность блокировки o -вызовов при наличии жестких ограничений на вероятность потери h -вызовов. И, наконец, в третьей задаче находится стратегия, минимизирующая общее число каналов базовой станции/потери o - и h -вызовов, соответственно. Показано, что в последних двух задачах оптимальная стратегия САС находится в классе обобщенных стратегий резервирования каналов для h -вызовов (Fractional Guard Channels, FGS). Согласно стратегии FGS, если o -вызов поступает в тот момент, когда число занятых каналов больше или равно $N - g$, то этот вызов с некоторой положительной вероятностью принимается, а с дополнительной вероятностью

стью он получает отказ. Здесь разработаны алгоритмы нахождения параметров соответствующих стратегий доступа. Однако практическая реализация стратегий FGS наталкивается на определенные методологические проблемы, связанные с реализацией рандомизированной стратегии в реальных системах.

Заключение и комментарии к главе 1 части I

Выше было дано краткое описание современного состояния беспроводных сетей связи. Подробное изложение технологические основы этих сетей можно найти в специальной литературе [3, 12, 14, 16, 18-22, 25]. Математическая теория указанных сетей также изложена в монографической литературе [1, 2, 7, 11, 27]. В этих монографиях, а также в обзорных статьях [4, 5, 8, 13] можно найти необходимую информацию об известных моделях обслуживания разнотипных вызовов и методах и алгоритмах расчета их показателей качества обслуживания.

Здесь следует отметить, что почти во всех известных работах принимается допущение об идентичности длительности разговоров o - и h -вызовов. Однако из практических соображений можно предположить, что это допущение не вполне соответствует реальной ситуации. Действительно, в монографии [27] (глава 11) отмечается, что, вообще говоря, $\mu_o \neq \mu_h$, где μ_o^{-1} и μ_h^{-1} – средняя длительность разговоров o - и h -вызовов, соответственно. Более того, в работах [6, 17] с помощью обработки большого объема статистических данных показано, что в большинстве реальных сетях многих стран имеет место соотношение:

$$\lambda_h \gg \lambda_o, \mu_o \ll \mu_h. \quad (1.22)$$

С другой стороны, с теоретической точки зрения интерес представляют исследования моделей, в которых o - и h -вызовы не являются идентичными в смысле длительности их обработки, поскольку из них в частном случае получаются результаты для классических моделей.

В классических моделях принятие допущения $\mu_o = \mu_h$ позволяет использовать одномерные цепи Маркова для расчета их характеристик. Однако отказ от этого допущения приводит к необходимости исследования двумерных цепей Маркова (2-D MC). Исходя из этого, в следующей главе будут предложены новые модели и методы исследования процессов обработки вызовов в традиционных беспроводных сетях связи.

В ряде известных работ для исследования моделей рассматриваемых сетей используется метод имитационного моделирования [23, 24].

Литература к главе 1 части I

1. **Башарин Г.П.** Лекции по математической теории телеграфика. – М.: Изд. РУДН, 2004. – 190 с.
2. **Akimaru H., Kawashima M.** Teletraffic. Theory and applications. – London: Springer-Verlag, 1993. – 225 p.
3. **Andreadis A., Giambene G.** Protocols for high efficiency wireless networks. – Boston: Kluwer Academic Publishers, 2003. – 432 p.
4. **Boucherie R.J., Van Dijk N.M.** On a queuing network model for cellular mobile telecommunications networks // Oper. Res. – 2000. – Vol.48, № 1. – pp. 38 – 49.
5. **Boucherie R.J., Mandjes M.** Estimation of performance measures for product form cellular mobile communications networks // Telecomm. Systems. – 1998. – Vol.10. – pp. 321 – 354.
6. **Casares-Giner V.** Integration of dispatch and interconnect traffic in a land mobile trunking system. Waiting time distributions // Telecomm. Systems. – 2001. – Vol.16, № 3,4. – pp. 539 – 554.
7. **Chen H., Huang L., Kumar S., Kuo C.C.** Radio resource management for multimedia QoS support in wireless networks. – Boston: Kluwer Academic Publishers, 2004. – 256 p.
8. **Das Bit S., Mitra S.** Challenges of computing in mobile cellular environment – a survey // Computer Communications. – 2003. – Vol.26. – pp. 2090 – 2105.
9. **Haring G., Marie R., Puigjaner R., Trivedi K.** Loss formulas and their application to optimization for cellular networks // IEEE Transactions on Vehicular Technology. – 2001. – Vol.50, № 3. – pp. 664-673.
10. **Hong D., Rappoport S.S.** Traffic model and performance analysis of cellular mobile radio telephones systems with prioritized and nonprioritized handoff procedures // IEEE Transactions on Vehicular Technology. – 1986. – Vol.35, № 3. – pp. 77 – 92.
11. **Janevski T.** Traffic analysis and design of wireless IP networks. – Boston: Artech House, 2003. – 469 p.
12. **Korhonen J.** Introduction to 3G mobile systems. – Boston: Artech House, 2003. – 238 p.
13. **Li W., Chao X.** Modeling and performance evaluation of a cellular mobile networks // IEEE/ACM Tran. on Networking. – 2004. – Vol. 12, No. 1. – pp.131 – 145.
14. **Lin Y.-B., Chlamatac I.** Mobile networks. Protocol and Services. – N.Y.: Wiley, 2000. – 286 p.
15. **Lin Y.-B., Mohan S., Noerpel A.** Queuing priority channel assignment strategies for PCS handoff and initial access // IEEE Transactions on Vehicular Technology. – 1994. – Vol. 43, № 3. – pp.704 – 712.

16. Mehrotra A. Cellular Radio Analogy and Digital Systems. — Norwood, MA: Artech House, 1994. — 364 p.
17. Pla V., Cesares-Giner V. A spectral-based analysis of priority channel assignment schemes in mobile cellular communication systems // Int. J. of Wireless Inf.Networks. — 2005. — Vol. 12, No.2. — pp. 87 — 99.
18. Prasad R., Mohr W., Konhauser W. 3G mobile communication systems. — Norwood, MA: Artech House, 2000. — 280 p.
19. Prasad R., Prasad S. WLAN systems and wireless IP for next generation communications. Norwood, MA: Artech House, 2002. — 304 p.
20. Prasad R., Ruggieri M. Technology trends in wireless communications. — Boston: Artech House, 2003. — 144 p.
21. Rappoport S.S. Wireless communications, principles and practice. — N.Y.: Prentice-Hall, 1996. — 266 p.
22. Redl, S. M., Weber M.K., Oliphant M.W. An introduction to GSM, Norwood, MA: Artech House, 1995. — 288 p.
23. Tekinay S., Jabbari B. A measurement-based prioritization scheme for handovers in mobile cellular networks // IEEE JSAC. — 1992. — Vol.10, № 8. — pp.1343 — 1350.
24. Tekinay S., Jabbari B. Handover policies and channel assignment strategies in mobile cellular networks // IEEE Commun. Mag. — 1991. — Vol. 29, №11. — pp.42 — 46.
25. Schiller J. Mobile communications.— Norwood, MA: Addison-Wesley, 2000.— 276 p.
26. Yoon C.H., Un C.K. Performance of personal portable radio telephone systems with and without guard channels // IEEE JSAC. — 1993. — Vol.11, № 6. — pp.911 — 917.
27. Yue W., Matsumoto Y. Performance analysis of multi-channel and multi-traffic on wireless communication networks. — Boston: Kluwer Academic Publishers, 2002. — 320 p.

ИССЛЕДОВАНИЕ МОДЕЛЕЙ ПРОЦЕССОВ ОБСЛУЖИВАНИЯ ВЫЗОВОВ В СЕТЯХ БЕЗ ОЧЕРЕДЕЙ

В данной главе исследуются модели процессов обслуживания вызовов в беспроводных сетях связи при различных стратегиях доступа. Сначала изучаются модели, в которых используется наиболее известная SAC, основанная на GS-схеме доступа к общим каналам базовой станции соты. Далее исследуется модель изолированной соты, в которой принята стратегия управления доступом разнотипных вызовов в каналы базовой соты, основанная на выделении индивидуальных каналов для h -вызовов. Предложено также обобщение этой стратегии доступа к каналам. В ней вводится потолок на число h -вызовов в общей зоне каналов с целью защиты o -вызовов от частых потерь. Разработаны алгоритмы расчета характеристик системы при всех исследуемых стратегиях доступа и показана высокая точность предложенных зависимостей.

Выполнен сравнительный анализ характеристик соты при использовании различных стратегий доступа разнотипных вызовов. Показано, что стратегия одного типа позволяет улучшить некоторые характеристики, при этом ухудшаются другие характеристики соты. В связи с этим, отмечается актуальность проведения серьезных исследований по выбору соответствующей стратегии с целью удовлетворения заданных требований на показатели QoS разнотипных вызовов.

2.1. Расчет параметров модели соты с резервированием каналов для хэндовер-вызовов

Поскольку в первой главе достаточно подробно описана данная стратегия доступа, то здесь вкратце остановимся на конкретном изложении ее реализации. При пересечении границы некоторой соты движущийся пользователь освобождает канал данной соты и требует свободный канал в соседней соте. Это явление называется хэндовер. Если в соседней соте имеется хотя бы один свободный канал, то обслуживание хэндовер-вызова (h -вызова) возобновляется незаметно для пользователя; в противном случае хэндовер-вызов теряется. Обычно потери h -вызовов считаются более нежелательными, чем блокировка новых заявок (o -вызовов). Одной из эффективных превентивных мер для уменьшения потери хэндовер-вызовов является выделение специально для них определенного количества каналов. В литературе эти каналы называются сторожевыми (Guard Channels).

Как было отмечено в предыдущей главе, почти во всех известных печатных работах данная модель изучена при предположении о том, что хэндовер и новые заявки являются идентичными в смысле длительности занятия ими радиоканала. Благодаря данному допущению функциониро-

вание системы описывается одномерной цепью Маркова, и удается найти достаточно простые рекуррентные формулы для расчета вероятности потери хэндовер-вызова и блокировки новых заявок. Однако иногда предположение об идентичности разнотипных вызовов является ограничивающим и не всегда реальным. Исходя из этого, в данной работе рассматриваются обобщения модели, подробно изученной в доступной литературе.

Рассматривается модель работы изолированной соты беспроводной сотовой коммуникационной сети. Сота содержит N каналов, $1 < N < \infty$. На эти каналы поступают пуассоновские потоки новых вызовов (с интенсивностью λ_1) и хэндовер-вызовов (с интенсивностью λ_2). Если в момент поступления h -вызова имеется хотя бы один свободный канал, то он принимается и для его обслуживания назначается один из свободных каналов, в противном случае h -вызов теряется. Новый вызов принимается лишь тогда, когда имеется не меньше чем $g+1$ свободных каналов; в противном случае новая заявка блокируется.

Блок-схема алгоритма, реализующего данную САС, показана на рис. 2.1.

Функции распределения времени занятия канала вызовами обоих типов имеют показательный вид, но их параметры отличаются друг от друга, т.е. время занятия канала нового (хэндовер) вызова равно μ_1^{-1} (μ_2^{-1}), при этом, вообще говоря, $\mu_1 \neq \mu_2$.

Замечание 2.1. Важно отметить, что указанные интенсивности обслуживания канала включают в себя также соответствующие интенсивности перехода новых и хэндовер-вызовов в соседнюю соту. При этом, если предположить, что время пребывания разнотипных вызовов в сотах является показательно распределенной случайной величиной с различными, вообще говоря, средними, то именно эти параметры существенным образом влияют на время занятия канала разнотипными вызовами, в результате чего их длительности оказываются различными.

Замечание 2.2. Предполагается, что $g < N$, так как в противном случае новые вызовы никогда не принимались бы для обслуживания, т.е. в этом случае исходная модель сводится к классической модели Эрланга. В случае $g = 0$ отсутствуют ограничения для доступа новых вызовов, т.е. имеет место полностью доступная схема использования каналов. Поэтому в дальнейшем считается, что $0 < g < N$.

Исходя из предположений о виде функций распределения входящих трафиков и времен занятия ими каналов, получаем, что функционирование соты описывается двумерной цепью Маркова (ЦМ), т.е. состояние соты в произвольный момент времени описывается двумерным вектором $\mathbf{n} = (n_1, n_2)$, где $n_i(n_2)$ указывает число новых (хэндовер) вызовов в каналах. Тогда фазовое пространство состояний соответствующей ЦМ определяется так:

$$S := \{ \mathbf{n} : n_1 = 0, \overline{N - g}, n_2 = 0, \overline{N}, n_1 + n_2 \leq N \}. \quad (2.1)$$

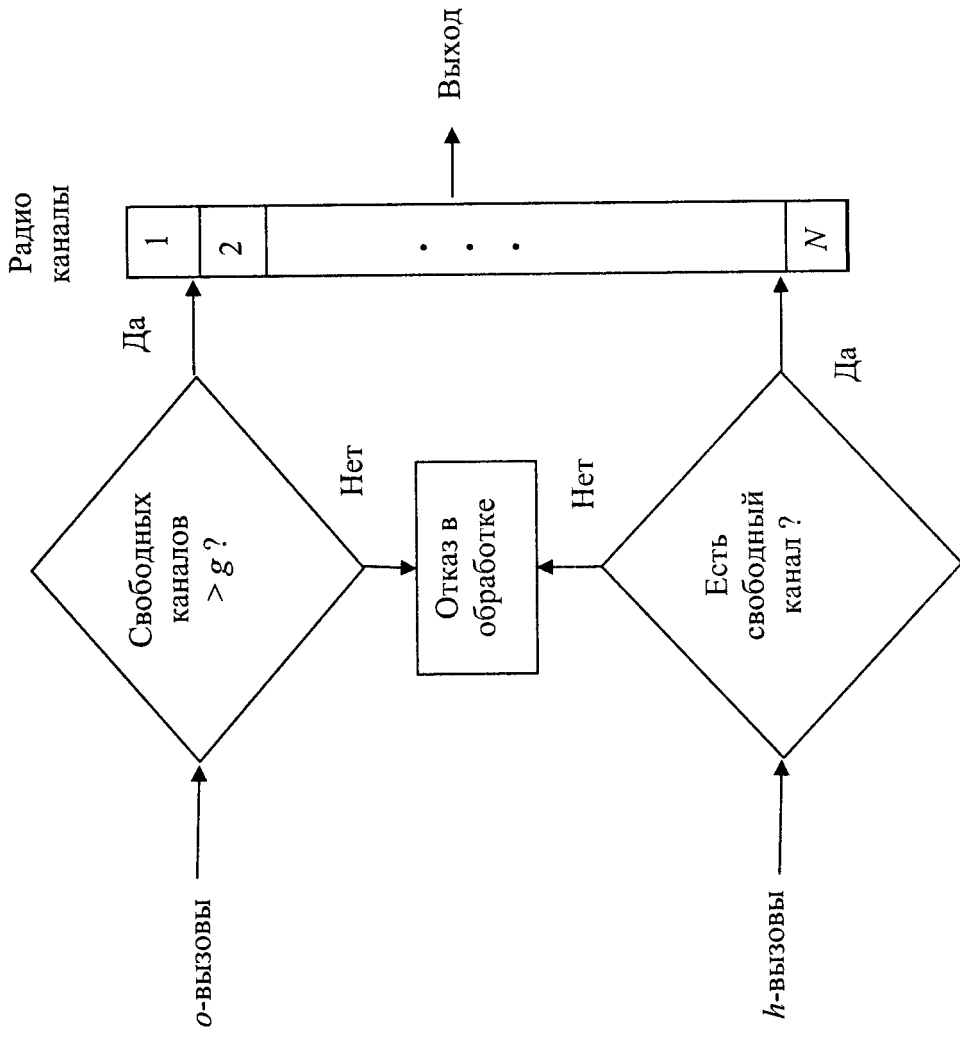


Рис. 2.1. Схема алгоритма САС на основе GS-стратегии доступа в каналы

Элементы производящей матрицы данной ЦМ, $q(\mathbf{n}, \mathbf{n}')$; $\mathbf{n}, \mathbf{n}' \in S$, определяются из следующих соотношений:

$$q(\mathbf{n}, \mathbf{n}') = \begin{cases} \lambda_1, & \text{если } n_1 + n_2 \leq N - g - 1, \mathbf{n}' = \mathbf{n} + \mathbf{e}_1, \\ \lambda_2, & \text{если } \mathbf{n}' = \mathbf{n} + \mathbf{e}_2, \\ n_i \mu_i, & \text{если } \mathbf{n}' = \mathbf{n} - \mathbf{e}_i, i = 1, 2, \\ 0, & \text{в остальных случаях,} \end{cases} \quad (2.2)$$

где $\mathbf{e}_1 = (1, 0)$, $\mathbf{e}_2 = (0, 1)$. Граф переходов данной цепи показан на рис. 2.2.

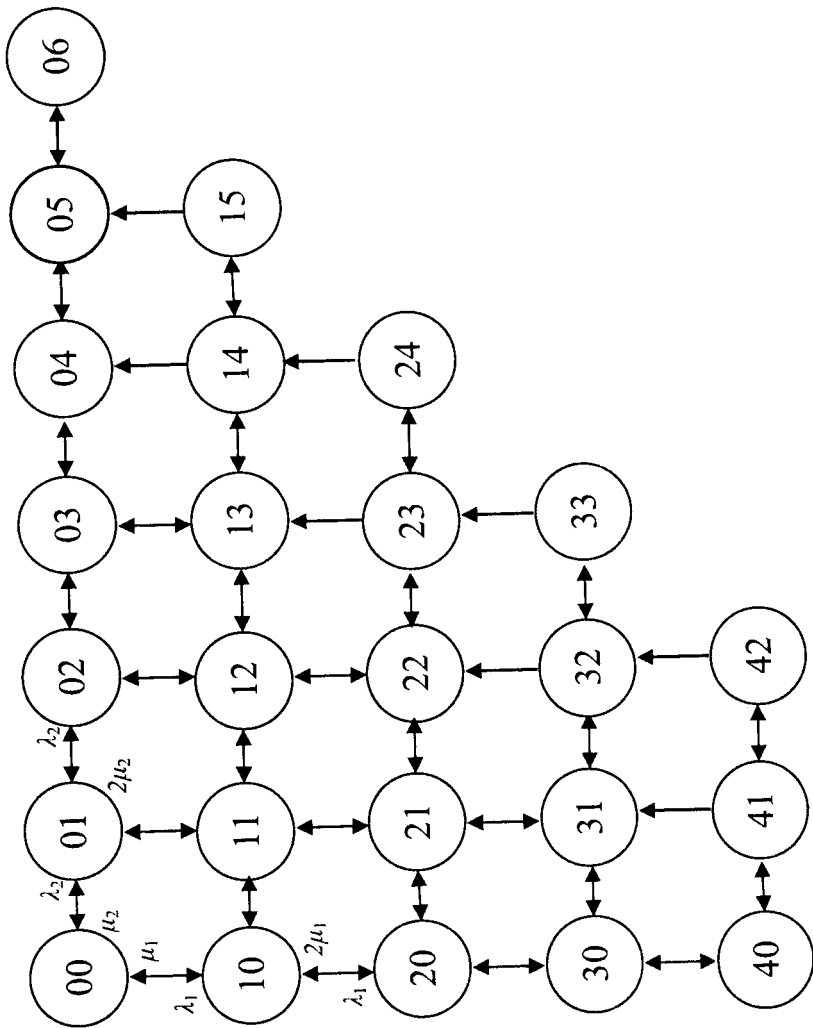


Рис.2.2. Граф переходов в модели с резервированием каналов для h -вызовов, $N = 6, g = 2$

Стационарную вероятность состояния $n \in S$ обозначим $p(n)$. Существование стационарного режима вытекает из конечности ФПС (2.1) и неприводимости соответствующего графа (см. рис. 2.2.)

Стационарное распределение находится как единственное решение соответствующей системы уравнений равновесия (СУР) данной ЦМ с пространством состояний (2.1) и производящей матрицей (2.2).

Используя стандартную технику теории конечномерных цепей Маркова, находим, что искомая СУР имеет следующий вид:

$$p(n)(\lambda_1(1 - \delta_{n_1+n_2, N}) + \lambda_2 I(n_1 + n_2 \leq N - g - 1) + \sum_{i=1}^2 n_i \mu_i) = \quad (2.3)$$

$$\sum_{i=1}^2 \lambda_i p(n - e_i) I(n_i > 0) + \sum_{i=1}^2 (n_i + 1) \mu_i p(n + e_i) (1 - \delta_{n_1+n_2, N-1}),$$

где δ_{ij} – символ Кронекера, $I(A)$ – индикаторная функция события A . К системе уравнений (2.3) добавляется нормирующее условие:

$$\sum_{n \in S} p(n) = 1. \quad (2.4)$$

Все желаемые показатели QoS могут быть вычислены с помощью стационарного распределения. Основными показателями QoS разнотипных вызовов являются вероятности их потери (блокировки) и среднее число занятых каналов базовой станции соты.

Вероятность потери хэндовер-вызовов и вероятность блокировки новых вызовов обозначаются $P_h(N, g)$ и $P_o(N, g)$, соответственно.

Эти величины определяются через стационарное распределение исходной модели. Согласно данной САС поступивший o -вызов теряется, если в момент его поступления число свободных каналов BS меньше, чем g (см. рис. 2.1). Тогда с помощью теоремы PASTA получаем, что вероятность блокировки o -вызовов определяется так:

$$P_o(N, g) := \sum_{n \in S} p(n) I(n_1 + n_2 \geq N - g), \quad (2.5)$$

Потеря h -вызова происходит лишь тогда, когда в момент его поступления все каналы BS заняты, т.е. опять с использованием указанной выше теоремы находим, что вероятность потери h -вызовов определяется так:

$$P_h(N, g) := \sum_{n \in S} p(n) I(n_1 + n_2 = N) \quad (2.6)$$

или

$$P_h(N, g) = \sum_{n \in S_g} p(n),$$

где $S_d := \{n \in S: n_1 + n_2 = N\}$ – множество диагональных состояний ФПС (2.1).

Формулы (2.5) и (2.6) могут быть записаны в явном виде следующим образом:

$$P_o(N, g) := \sum_{n_1=0}^{N-g} \sum_{n_2=N-g-n_1}^{N-n_1} p(n_1, n_2), \quad (2.7)$$

$$P_h(N, g) := \sum_{n_1=0}^{N-g} p(n_1, N - n_1). \quad (2.8)$$

Замечание 2.3. Из (2.7) и (2.8) видно, что при фиксированном значении N функция $P_h(N, g)$ является убывающей, а функция $P_o(N, g)$ является возрастающей функцией от g . Также из (2.7) и (2.8) имеем:

$$P_o(N, g) \geq P_h(N, g), \quad g = 0, \overline{N-1}, \quad (2.9)$$

$$P_n(N, 0) = P_n(N, 0) = \sum_{n_1=0}^N p(n_1, N - n_1). \quad (2.10)$$

Другой важной характеристикой соты (с точки зрения администрации) является коэффициент утилизации каналов BS соты, т.е. среднее число занятых каналов. Эту величину обозначим $\tilde{N}_{av}(N, g)$. Она определяется как математическое ожидание величины:

$$\tilde{N}_{av}(N, g) := \sum_{i=1}^N i \sum_{n \in S} p(n) \delta_{n_1+n_2, i}. \quad (2.11)$$

Стационарное распределение определяется в результате решения системы уравнений равновесия (2.3), (2.4). С использованием известной теоремы Колмогорова об обратимости двумерных цепей Маркова можно показать, что эта СУР имеет мультипликативное решение лишь в частном случае $g = 0$, а в остальных случаях ее решение связано с огромным вычислительными трудностями при больших значениях N (даже в случае $g = 0$ существуют известные вычислительные трудности). Последнее означает, что для вычисления стационарного распределения каждый раз при конкретных значениях нагрузочных параметров разнотипных трафиков и количества каналов соты необходимо решить соответствующую систему линейных алгебраических уравнений. Однако для реальных сот она имеет огромную размерность, и поэтому ее решение при больших размерностях связано порой с непреодолимыми вычислительными трудностями.

Для преодоления указанных трудностей здесь предлагается новый метод расчета стационарного распределения данной модели. Во всех исслеуемых в данной главе моделях предполагается, что $\lambda_2 \gg \lambda_1$ и $\mu_2 \gg \mu_1$ (см. (1.22)).

Очень важно отметить, что это допущение наиболее адекватно соответствует режиму работы реальных пикосот. Статистические исследования показали, что в таких сотах интенсивность трафика h -вызовов намного превосходит интенсивность трафика o -вызовов и длительность разговора h -вызовов намного меньше соответствующего показателя o -вызовов (см. Главу 1). Более того, выполнение данного допущения позволяет корректно применить алгоритм приближенного расчета стационарного распределения двумерных цепей Маркова, основанный на идеях фазового укрупнения (см. Вводную часть).

Из соотношений (2.2) видно, что рассматриваемая 2D MC является строго непрерывной по второй компоненте и слабо непрерывной по первой компоненте (см. также рис. 2.2). Исходя из этого, применение разработанного во вводной части приближенного алгоритма вполне корректно.

С этой целью рассматривается следующее расщепление ФПС (2.1):

$$S = \bigcup_{k=0}^{N-g} S_k, \quad S_k \cap S_{k'} = \emptyset, \quad k \neq k', \quad (2.12)$$

где $S_k := \{n \in S; n_1 = k\}$, т.е. класс состояний S_k содержит те состояния n из ФПС (2.1), в которых число o -вызовов равно k .

Классы состояний S_k объединяются в отдельные укрупненные состояния $\langle k \rangle$, и вводится следующая функция укрупнения в ФПС S :

$$U(n) = \langle k \rangle, \quad \text{если } n \in S_k, \quad k = 0, N - g. \quad (2.13)$$

Функция укрупнения (2.13) определяет укрупненную модель, которая является одномерной цепью Маркова с пространством состояний $\tilde{S} := \{\langle k \rangle; k = 0, N - g\}$.

Тогда согласно АФУ стационарное распределение исходной модели приближенно определяется так:

$$p(n_1, n_2) \approx \rho^{n_1}(n_2) \pi(\langle n_1 \rangle), \quad n \in S_n, \quad n_1 = 0, N - g. \quad (2.14)$$

где $\{\rho^{n_1}(n_2); n_1, n_2 \in S_n\}$ и $\{\pi(\langle n_1 \rangle); \langle n_1 \rangle \in \tilde{S}\}$ являются стационарными распределениями внутри класса S_{n_1} и укрупненной модели, соответственно.

Стационарная вероятность состояния (k, m) внутри класса S_k обозначается $\rho^k(m)$, где $k = 0, N - g, m = 0, N - k$. Тогда стационарное распределение соответствующего одномерного процесса размножения и гибели (см. рис.2.3) задается с помощью известных формул Эрланга для системы $M|M|N-k|0$ с нагрузкой v_2 эрл., $v_2 := \lambda_2 / \mu_2$:

$$\rho^k(m) = \frac{v_2^m}{m!} \rho^k(0), \quad k = 0, N - g, \quad m = 0, N - k. \quad (2.15)$$

где

$$\rho^k(0) = \left(\sum_{m=0}^{N-k} \frac{v_2^m}{m!} \right)^{-1}. \quad (2.16)$$

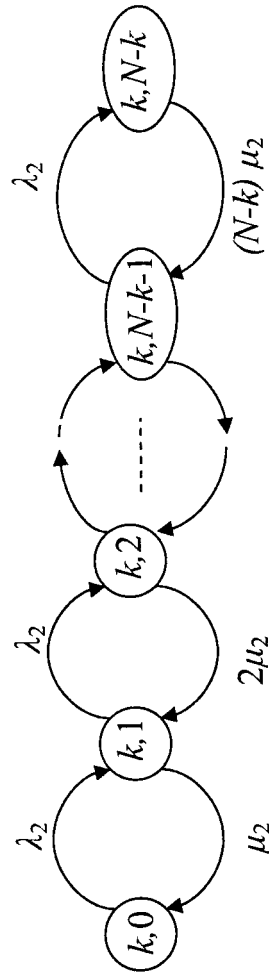


Рис. 2.3. Граф расщепленной модели с ФПС $S_k, k = 0, N - g$

С учетом (2.2), (2.15) и (2.16) получаем, что элементы производящей матрицы укрупненной модели $\hat{q}(x, y)$, $x, y \in \tilde{S}$, определяются так:

$$\hat{q}(x, y) = \begin{cases} \lambda_1 \rho^k(0) \sum_{i=0}^{N-g-k-1} \frac{v_2^i}{i!}, & \text{если } k' = k+1, \\ k\mu_1, & \text{если } k' = k-1, \\ 0 & \text{в остальных случаях.} \end{cases} \quad (2.17)$$

Тогда из (2.17) легко определяется стационарное распределение $\pi(\langle k \rangle)$, $\langle k \rangle \in \tilde{S}$, укрупненной модели, также описываемой одномерным процессом размножения и гибели (см. рис.2.4):

$$\pi(\langle k \rangle) = \frac{v_1^k}{k!} \prod_{i=1}^k \Lambda(i) \pi(\langle 0 \rangle), \quad (2.18)$$

где

$$\pi(\langle 0 \rangle) = \left(1 + \sum_{k=1}^{N-g} \frac{v_1^k}{k!} \prod_{i=1}^k \Lambda(i) \right)^{-1}, \quad (2.19)$$

$$\Lambda(i+1) = \rho^i(i, 0) \sum_{j=0}^{N-g-i-1} \frac{v_2^j}{j!}, \quad i = 0, \overline{N-g-1}. \quad (2.20)$$

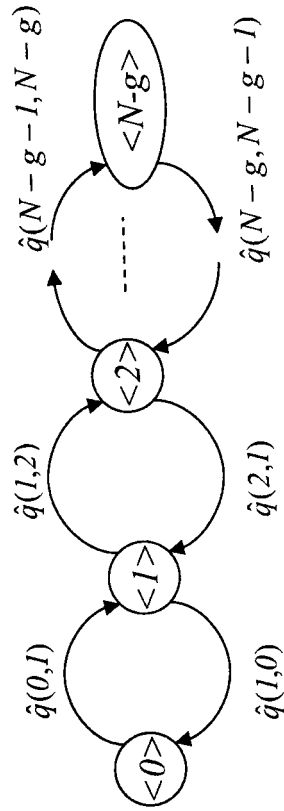


Рис. 2.4. Граф укрупненной модели с ФПС \tilde{S}

Следовательно, с помощью формул (2.14) – (2.16), (2.18) – (2.20) определяется стационарное распределение $(p(\mathbf{n}), \mathbf{n} \in S)$ исходной двумерной ЦМ с пространством состояний (2.1). Далее из формул (2.7), (2.8) и (2.11) найдутся искомые показатели QoS данной системы.

Опуская промежуточные математические преобразования, выпишем конечные формулы для вычисления указанных величин.

$$P_o(N, g) = \sum_{k=0}^{N-g} \sum_{m=N-g-k}^{N-k} \rho^k(k, m) \pi(\langle k \rangle), \quad (2.21)$$

$$P_n(N, g) = \sum_{k=0}^{N-g} E_B(v_2, N-k) \pi(\langle k \rangle), \quad (2.22)$$

$$\tilde{N}_{av}(N, g) = \sum_{k=1}^{N-g} k \sum_{i=0}^{k-1} \rho^i(k-i) \pi(\langle i \rangle) + \sum_{k=N-g+1}^{N-g} k \sum_{i=0}^{N-g-k} \rho^i(k-i) \pi(\langle i \rangle), \quad (2.23)$$

где $E_B(v, s)$ – В-формула Эрланга, т.е.

$$E_B(v, s) = \frac{v^s}{s!} \left(\sum_{k=0}^s \frac{v^k}{k!} \right)^{-1}. \quad (2.24)$$

Подводя итоги, можно предложить следующий алгоритм расчета величин (2.7), (2.8) и (2.11).

Шаг 1. Для $k = 0, \overline{N-g}$ и $m = 0, \overline{N-k}$ вычислить $\rho^k(m)$ из (2.15), (2.16).

Шаг 2. Для $k = 0, \overline{N-g-1}$ вычислить $\Lambda(k+1)$ из (2.20).

Шаг 3. Для $k = 0, \overline{N-g}$ вычислить $\pi(\langle k \rangle)$ из (2.18), (2.19).

Шаг 4. Вычислить $P_o(N, g)$, $P_n(N, g)$ и $\tilde{N}_{av}(N, g)$ из формул (2.21)–(2.23).

Замечание 2.4. Из (2.21), (2.22) в частности получаем, что

$$P_o(N, 0) = P_n(N, 0) = \sum_{k=0}^N E_B(v_2, N-k) \pi(\langle k \rangle). \quad (2.25)$$

Соотношение (2.25) еще раз подтверждает правомерность Замечания 2.3 (см. (2.10)).

Важно отметить, что конечные результаты не зависят от нагрузочных параметров трафиков (т.е. λ_i, μ_i), а определяются только через их отношение, т.е. через $v_i = \lambda_i/\mu_i$.

В предложенном алгоритме часто используется В-формула Эрланга, а также выражения, которые участвуют в этой формуле (см. (2.15)–(2.20)). Прямое вычисление по формуле (2.24) приводит к переполнению (или исчезновению порядка) при больших значениях N из-за больших факториалов и степеней нагрузки. Для преодоления этих трудностей, в частности можно использовать следующую эффективную рекуррентную процедуру.

• Полагается $F(1) = v$ и вычисляются

$$F(k) = F(k-1) \frac{v}{k}, \quad k = 2, \overline{N}.$$

- Полагается $G(0) = 1$ и вычисляются

$$G(k) = G(k-1) + F(k), \quad k = \overline{1, N}.$$

- $E_B(N, v) = F(N) / G(N)$.

2.2. Расчет параметров модели соты с индивидуальными каналами для хэндовер-вызовов

Теперь рассмотрим другую схему предоставления приоритета h -вызовам в сетях без очередей. Качественное описание предложеной стратегии САС состоит в следующем. Определенное количество каналов из общего пула каналов выделяется лишь для обслуживания h -вызовов, а оставшиеся каналы совместно и равноправно используются o - и h -вызовами. Изложим детальное описание этой стратегии доступа.

Ровно r каналов из общего числа N каналов выделяются лишь для обслуживания h -вызовов, $0 \leq r \leq N-1$, а остальные $N-r$ каналов используются совместно o - и h -вызовами согласно полноступенчатой схеме САС, т.е. весь пул из N каналов делится на две зоны: индивидуальную зону из r каналов (лишь для h -вызовов) и общую из $N-r$ каналов (для o - и h -вызовов). Иными словами, если в момент поступления h -вызова имеется хотя бы один свободный канал (либо в индивидуальной, либо в общей зоне), то он принимается на обслуживание; в противном случае хэндовер-вызов теряется. Поступивший o -вызов принимается лишь тогда, когда имеется хотя бы один свободный канал в общей зоне; в противном случае новый вызов блокируется.

Отметим, что процесс занятия каналов h -вызовами происходит следующим образом. Если в момент поступления h -вызова имеется свободный канал в обеих зонах, то он занимает канал индивидуальной зоны, т.е. h -вызовы в первую очередь занимают свою зону, и лишь при отсутствии свободного канала в собственной зоне они используют каналы общей зоны. Здесь также предполагается, что по завершении обслуживания h -вызова освобожденный канал, находящийся в индивидуальной зоне, передается в общую зону при наличии там h -вызова, а канал в общей зоне, осуществляющий обслуживание h -вызова, закрепляется за индивидуальной зоной. Эта процедура называется переупаковкой каналов. Схема алгоритма доступа показана на рис.2.5.

В стационарном режиме состояние соты в произвольный момент времени описывается двумерным вектором $\mathbf{n} = (n_1, n_2)$, где $n_i(n_i)$ указывает число новых (хэндовер) вызовов в каналах. Тогда фазовое пространство состояния соответствующей ЦМ определяется так:

$$S := \{ \mathbf{n} : n_1 = \overline{0, N-r}, n_2 = \overline{0, N}, n_1 + n_2 \leq N \}. \quad (2.26)$$

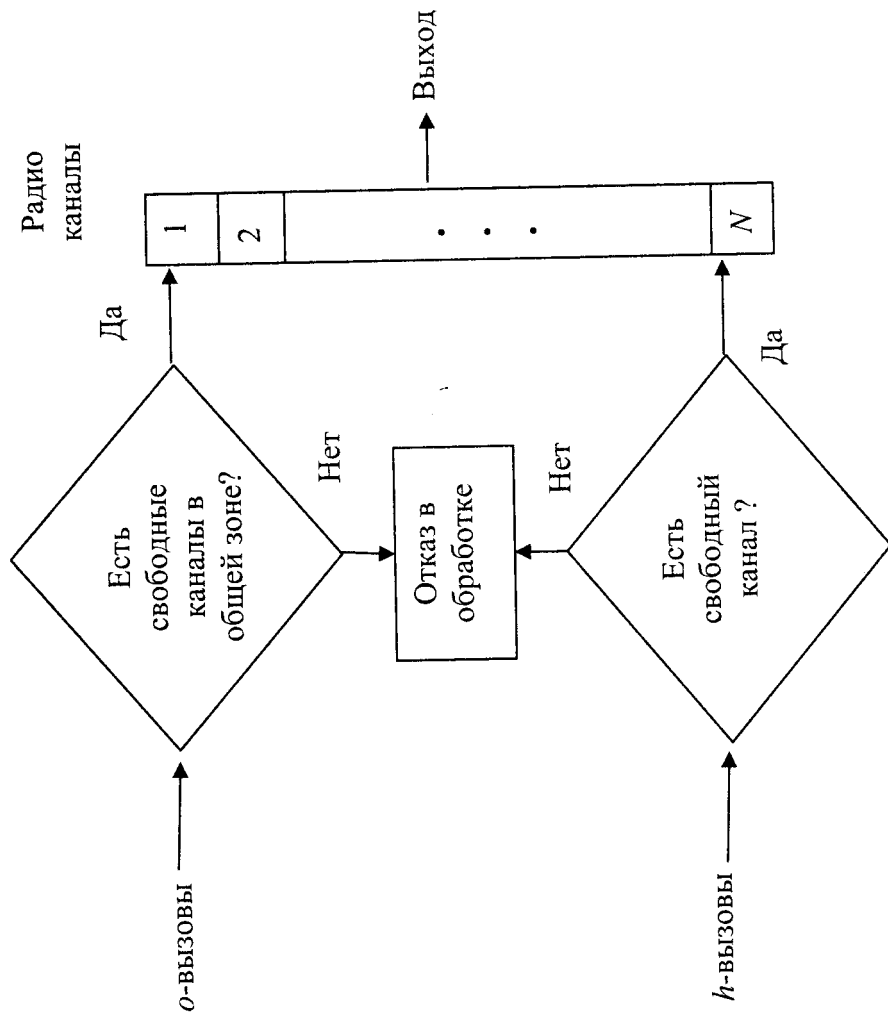


Рис. 2.5. Схема алгоритма САС, основанного на индивидуальных каналах для h -вызовов

Замечание 2.6. Здесь и в дальнейшем в целях упрощения изложения при всех стратегиях САС используются одинаковые обозначения для ФПС, стационарных распределений и показателей QoS разнотипных вызовов и характеристики соты. Однако из контекста, а также в необходимых случаях из буквенных индексов и списка аргументов рассматриваемых функций будет ясно, о каких именно моделях и стратегиях САС идет и идти речь.

С учетом работы принятой схемы обслуживания разнотипных вызовов делаем вывод, что интенсивности переходов между ее состояниями \mathbf{n} и \mathbf{n}' , $\mathbf{n}, \mathbf{n}' \in S$, обозначаемые $q(\mathbf{n}, \mathbf{n}')$, определяются из следующих соотношений:

$$q(\mathbf{n}, \mathbf{n}') = \begin{cases} \lambda_i, & \text{если } \mathbf{n}' = \mathbf{n} + \mathbf{e}_i, \\ n_i \mu_i, & \text{если } \mathbf{n}' = \mathbf{n} - \mathbf{e}_i, \\ 0 & \text{в остальных случаях.} \end{cases} \quad (2.27)$$

Граф с переходов данной ЦМ показан на рис.2.6. Основными характеристиками данной системы также являются вероятности потери разнотипных вызовов и среднее число занятых каналов.

Поскольку h -вызовы теряются лишь тогда, когда в момент их поступления все каналы BS заняты, то согласно теореме PASTA имеем:

$$P_h(N, r) := \sum_{\mathbf{n} \in S_d} p(\mathbf{n}), \quad (2.28)$$

где $p(\mathbf{n})$ — стационарная вероятность состояния $\mathbf{n} \in S$;

$S_d := \{\mathbf{n} \in S : n_1 + n_2 = N\}$ — множество диагональных состояний. Новые вызовы теряются в двух случаях: либо в момент их поступления все каналы BS заняты, либо в эти моменты в системе уже имеется r вызовов данного типа (т.е. все каналы индивидуальной зоны заняты).

Тогда, аналогично (2.28) имеем:

$$P_o(N, r) := \sum_{\mathbf{n} \in S_d} p(\mathbf{n}) + \sum_{\mathbf{n} \in S_{N-r}} p(\mathbf{n}), \quad (2.29)$$

где

$$S_{N-r} := \{\mathbf{n} \in S : n_1 = N - r\}.$$

Среднее число занятых каналов ($\tilde{N}_{av}(N, r)$) также определяется аналогично (2.11) через стационарное распределение $p(\mathbf{n})$, $\mathbf{n} \in S$:

$$\tilde{N}_{av}(N, r) := \sum_{k=1}^N k \sum_{\mathbf{n} \in S} p(\mathbf{n}) \delta_{n_1+n_2, k}. \quad (2.30)$$

Замечание 2.7. Непосредственно из формул (2.28) и (2.29) видно, что имеют место следующие соотношения:

$$P_h(N, r) \geq P_d(N, r), \quad r = 0, N-1; \quad (2.31)$$

$$P_h(N, 0) = P_d(N, 0) = \sum_{\mathbf{n} \in S_d} p(\mathbf{n}). \quad (2.32)$$

Для данной модели в отличие от модели с САС на базе GS-стратегии доступа существует мультипликативное решение СУР для стационарного распределения:

$$p(\mathbf{n}) = G^{-1}(N, r) \prod_{i=1}^2 \frac{\nu_i^{n_i}}{n_i}, \quad (2.33)$$

где $G^{-1}(N, r)$ — нормирующая константа над ФПС (2.26).

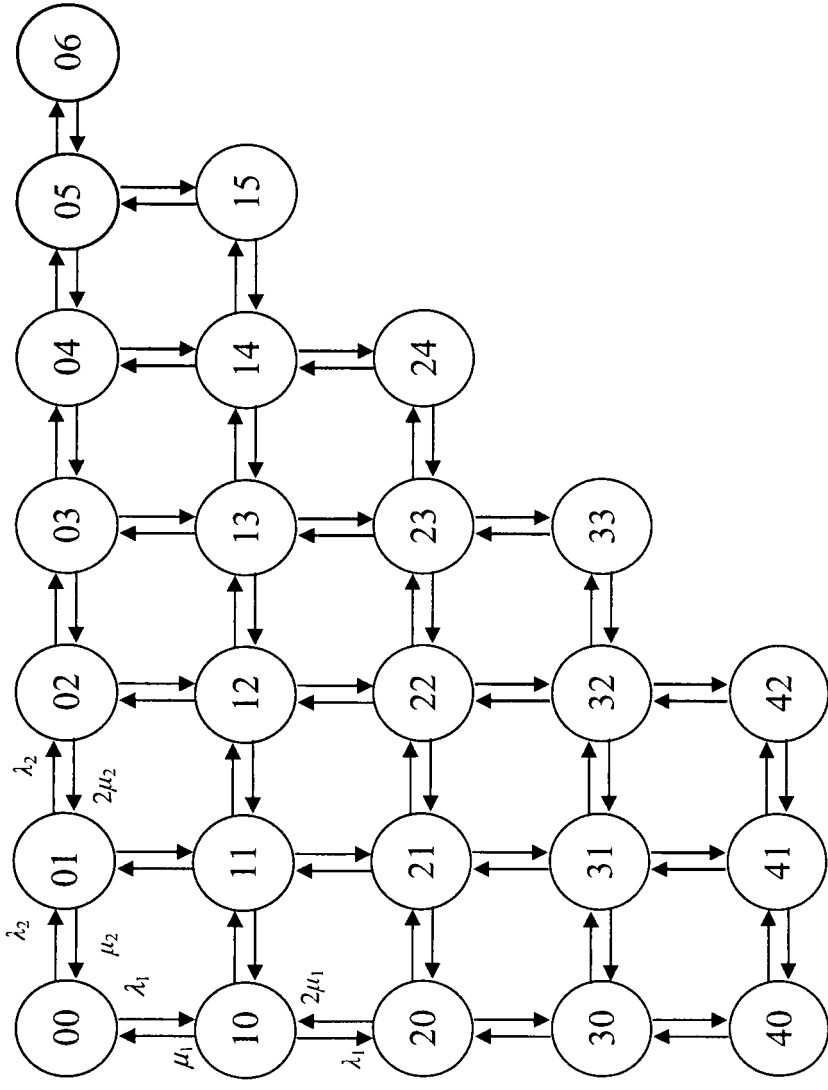


Рис.2.6. Граф модели с индивидуальными каналами для h -вызовов, $N = 6$, $r = 2$

Однако традиционный путь определения стационарного распределения через мультипликативное представление (2.33) при больших значениях N сталкивается с известными трудностями. Они связаны с тем, что при таком способе нахождения $p(\mathbf{n})$, $\mathbf{n} \in S$, потребуются генерация всего ФПС, а также вычисление факториалов больших чисел и степеней близких к нулю величин (при низких нагрузках) или больших величин (при высоких нагрузках), т.е. возникает проблема переполнения или исчезновения порядка.

В связи с этим, здесь предлагается другой подход к вычислению $p(\mathbf{n})$, $\mathbf{n} \in S$, который не требует генерации ФПС (2.26), а также позволяет более эффективно разрешить указанную проблему, т.к. в разработанных формулах используются уже табулированные величины.

Как и в §2.1 рассматривается следующее расщепление ФПС (2.26):

$$S = \prod_{k=0}^{N-r} S_k, \quad S_k \cap S_{k'} = \emptyset, \quad k \neq k', \quad (2.34)$$

где $S_k := \{\mathbf{n} \in S : n_1 = k\}$.

Далее классы состояний S_k объединяются в отдельные укрупненные состояния $\langle k \rangle$, и в ФПС (2.26) вводится следующая функция укрупнения:

$$U(n) = \langle k \rangle, \text{ если } n \in S_k, k = \overline{0, N-r}. \quad (2.35)$$

Стационарное распределение исходной модели приближенно определяется аналогично (2.14). При этом согласно расщеплению (2.34) и введенной функции укрупнения (2.35) стационарное распределение расщепленной модели с ФПС S_k определяется как стационарное распределение классической однопоточковой модели Эрланга $M|M|N-k|0$ с нагрузкой v_2 эрл., т.е. величины $\rho^k(m)$ определяются аналогично (2.15), (2.16):

$$\rho^k(m) = \frac{v_2^m}{m!} \rho^k(0), k = \overline{0, N-r}, m = \overline{0, N-k}, \quad (2.36)$$

где

$$\rho^k(0) = \left(\sum_{m=0}^{N-k} \frac{v_2^m}{m!} \right)^{-1}. \quad (2.37)$$

Тогда с учетом (2.27) и (2.36), (2.37) находим, что интенсивности переходов между состояниями $\langle k \rangle$ и $\langle k' \rangle$, $\langle k \rangle < k' \rangle \in \tilde{S}$, укрупненной модели $\hat{q}(\langle k \rangle, \langle k' \rangle)$ определяются следующим образом:

$$\hat{q}(\langle k \rangle, \langle k' \rangle) = \begin{cases} \lambda_1(1 - E_B(v_2, N-k)), & \text{если } k' = k+1, \\ k\mu_1, & \text{если } k' = k-1, \\ 0 & \text{в остальных случаях,} \end{cases} \quad (2.38)$$

где $E_B(v, s)$ – В-формула Эрланга.

Следовательно, укрупненная модель представляет собой одномерный процесс размножения и гибели с интенсивностями, определяемыми с помощью соотношений (2.38), т.е. стационарное распределение укрупненной модели $\pi(\langle k \rangle)$, $\langle k \rangle \in \tilde{S}$, в данной САС определяется так:

$$\pi(\langle k \rangle) = \frac{v_1^k}{k!} \prod_{i=0}^{k-1} (1 - E_B(v_2, N-i)) \pi(0), k = \overline{1, N-r}, \quad (2.39)$$

где

$$\pi(\langle 0 \rangle) = \left(1 + \sum_{k=1}^{N-r} \frac{v_1^k}{k!} \prod_{i=0}^{k-1} (1 - E_B(v_2, N-i)) \right)^{-1}. \quad (2.40)$$

Окончательно, с учетом (2.36)–(2.40) находим следующие формулы для расчета характеристик системы (2.28)–(2.30):

$$r'_o(N, r) = \pi(\langle N-r \rangle) + \sum_{k=0}^{N-r-1} E_B(v_2, N-k) \pi(\langle k \rangle); \quad (2.41)$$

$$P_h(N, r) = \sum_{k=0}^{N-r} E_B(v_2, N-k) \pi(\langle k \rangle); \quad (2.42)$$

$$\tilde{N}_{op}(N, r) = \sum_{k=1}^{N-r} k \sum_{i=0}^k \rho^i(k-i) \pi(\langle i \rangle) + \sum_{k=N-r+1}^N k \sum_{i=0}^k \rho^i(k-i) \pi(\langle i \rangle). \quad (2.43)$$

Резюмируя, можно предложить следующий алгоритм расчета характеристик системы при использовании САС, основанной на использовании индивидуальных каналов для h -вызовов.

Шаг 1. Для $k = \overline{0, N-r}$ и $m = \overline{0, N-k}$ вычислить $\rho^k(m)$ из (2.36), (2.37).

Шаг 2. Для $k = \overline{0, N-r}$ вычислить $\pi(\langle k \rangle)$ из (2.39), (2.40).

Шаг 3. Вычислить $P_o(N, r)$, $P_h(N, r)$ и $N_{op}(N, r)$ из (2.41)–(2.43).

Как видно из формул (2.41)–(2.43), для вычисления характеристик системы не потребуются генерация всего ФПС исходной модели и вычисление ее стационарного распределения, и их нахождение осуществляется с помощью простых вычислительных процедур. Указанные вычислительные процедуры являлись более простыми по сравнению с аналогичными процедурами для модели с САС на базе ГС-стратегии доступа (см. §2.1), и они также содержат хорошо известную В-формулу Эрланга, которая даже табулирована.

Замечание 2.8. Из (2.41) и (2.42) получаем, что

$$P_h(N, 0) = P_o(N, 0) = \sum_{k=0}^N E_B(v_2, N-k) \pi(\langle k \rangle). \quad (2.44)$$

Соотношение (2.44) еще раз подтверждает правомерность замечания 2.7 (см. (2.32)) и полностью совпадает с результатами §2.1.

Теперь рассмотрим обобщение данной САС. Вводится пороговый параметр b , ограничивающий число h -вызовов в общей зоне каналов, т.е. если в момент поступления h -вызова имеется хотя бы один свободный канал (либо в индивидуальной, либо в общей зоне), то он принимается для обслуживания лишь тогда, когда число h -вызовов в общей зоне меньше, чем b . При этом, очевидно, что должно выполняться условие $0 \leq b \leq N-r$.

Использование данной САС позволяет защитить o -вызовы от частых потерь вследствие ограничения доступа h -вызовов в общую зону.

Замечание 2.9. В частном случае $b = N-r$, из данной САС получается стратегия доступа, описанная выше; при $b = 0$ из данной САС получается стратегия доступа, согласно которой весь пул каналов полностью

разделен на две непересекающиеся части, т.е. в этом случае исходная система распадается на две изолированные системы, одна из которых обслуживает h -вызовы (с r каналами), а другая — o -вызовы (с $N - r$ каналами).

Схема алгоритма, реализующего данную стратегию САС, показана на рис. 2.7.

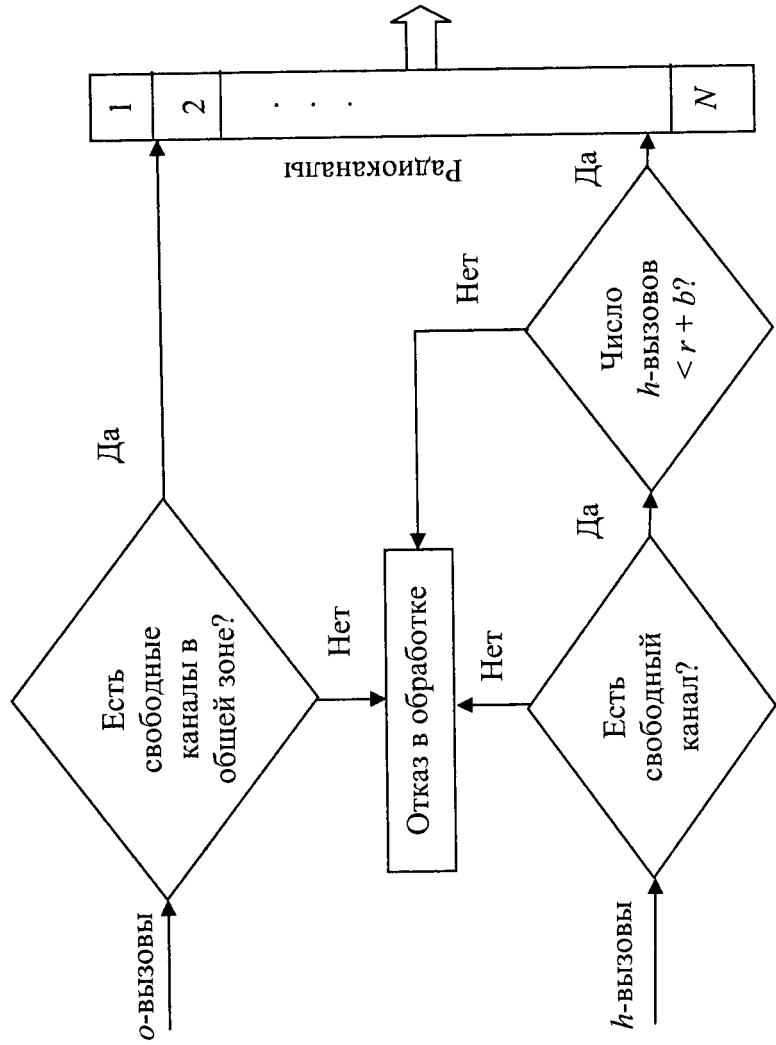


Рис. 2.7. Схема алгоритма САС, основанного на обобщенной стратегии индивидуальных каналов для h -вызовов

Фазовое пространство состояний ЦМ, описывающей работу соты при использовании данной обобщенной стратегии САС задается так

$$S := \{n : n_1 = \overline{0, N-r}, n_2 = \overline{0, r+b}, n_1 + n_2 \leq N\} \quad (2.45)$$

Элементы производящей матрицы ЦМ с пространством состояний (2.45) определяется точно так же, как и в (2.27).

Граф переходов этой ЦМ показан на рис. 2.8. Характеристики системы определяются через стационарное распределение соответствующей

ЦМ ($p(n), n \in S$), где S задается с помощью (2.45). Опуская подробности, ниже приводятся выражения для искомым характеристик системы:

$$P_o(N, r, b) := \sum_{n \in S_d} p(n) + \sum_{n \in S_{N-r} \setminus S_d} p(n), \quad (2.46)$$

$$P_h(N, r, b) := \sum_{n \in S_d} p(n) + \sum_{n \in S_{N-r} \setminus S_d} p(n) \delta_{n_2, r+b}, \quad (2.47)$$

где множество S определено в (2.45), $S_d := \{n \in S : n_1 + n_2 = N\}$, $S_{N-r} := \{n \in S : n_1 = N - r\}$.

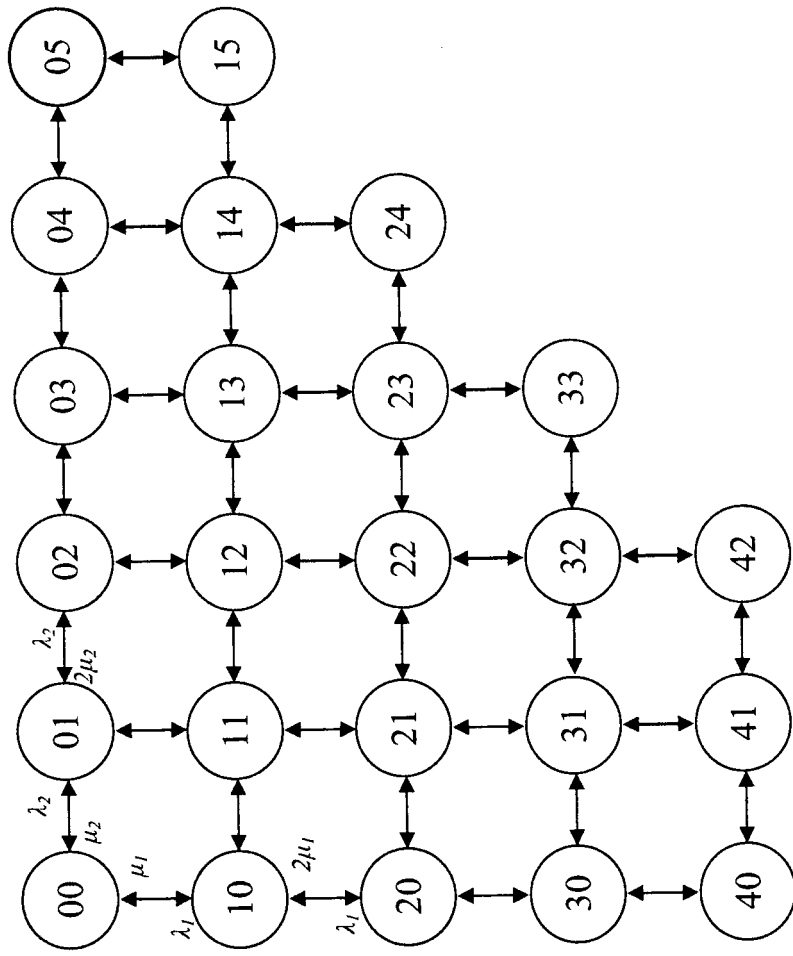


Рис. 2.8. Граф модели с обобщенной стратегией индивидуальных каналов для h -вызовов, $N = 6, r = 2, b = 3$

Среднее число занятых каналов ($\bar{N}_{av}(N, r, b)$) определяется точно так же, как в формуле (2.30).

Стационарное распределение данной цепи также имеет мультипликативный вид аналогично (2.33), но при этом нормирующая константа определяется над ФПС (2.45). Вместе с тем, вычислительные

трудности, которые присущи предыдущей модели, остаются и для данной модели.

Опуская уже известные промежуточные математические выкладки, ниже приводится алгоритм для вычисления характеристик системы при использовании данной САС, основанной на использовании обобщенной стратегии индивидуальных каналов для h -вызовов.

Шаг 1. Вычислить

$$\rho^k(m) = \begin{cases} \frac{v_2^m}{m!} \rho_1^k(0) & \text{если } \kappa = 0, \overline{N-r-b}, m = \overline{1, r+b}, \\ \frac{v_2^m}{m!} \rho_2^k(0), & \text{если } k = \overline{N-r-b+1, N-r}, m = \overline{1, N-k}, \end{cases}$$

$$\text{где } \rho_1^k(0) = \left(\sum_{m=0}^{r+b} \frac{v_2^m}{m!} \right)^{-1}, \rho_2^k(0) = \left(\sum_{m=0}^{N-k} \frac{v_2^m}{m!} \right)^{-1}.$$

Шаг 2. Вычислить

$$\pi(\langle\langle k \rangle\rangle) = \begin{cases} \frac{v_1^k}{k!} \pi(0), & k = \overline{1, N-r-b} \\ \frac{v_1^k}{k!} \prod_{i=0}^{k-N+r+b-1} (1 - E_B(v_2, r+b-i)) \pi(0), & k = \overline{N-r-b+1, N-r}, \end{cases}$$

$$\text{где } \pi(0) = \left(\sum_{k=0}^{N-r-b} \frac{v_1^k}{k!} + \sum_{k=N-r-b+1}^{N-r} \frac{v_1^k}{k!} \prod_{i=0}^{k-N+r+b-1} (1 - E_B(v_2, r+b-i)) \right)^{-1}$$

Шаг 3. Вычислить

$$P_o(N, r, b) = \pi(N-r) + E_B(v_2, r+b) \pi(N-r-b) + \sum_{k=N-r-b+1}^{N-r-1} E_B(v_2, N-k) \pi(k),$$

$$P_h(N, r, b) = E_B(v_2, r+b) \sum_{k=0}^{N-r-b} \pi(k) + \sum_{k=N-r-b+1}^{N-r} E_B(v_2, N-k) \pi(k),$$

$$\tilde{N}_{av}(N, r, b) = \sum_{k=1}^{N-r} k \sum_{i=0}^k \rho^i(k-i) \pi(\langle\langle i \rangle\rangle) + \sum_{k=N-r+1}^N k \sum_{i=0}^{N-r} \rho^i(k-i) \pi(\langle\langle i \rangle\rangle) I(i \geq k-r-b).$$

Отсюда видно, что в частном случае $b = N-r$ получается алгоритм расчета характеристик модели с индивидуальными каналами для h -вызовов, т.е. подтверждается справедливость замечания 2.9.

Одним из важных достоинств данной стратегии САС является то, что при ее использовании число степеней свободы (т.е. число варьируемых параметров) больше, чем в предыдущих двух стратегиях САС.

Действительно, если общее число каналов соты является фиксированной величиной, то при использовании САС, основанной на GS -стратегии, имеется один варьируемый параметр g (число резервированных каналов), а в САС, основанной на использовании индивидуальной зоны для h -вызовов, варьируемым параметром является r (размер индивидуальной зоны); а при использовании обобщенной стратегии имеются два варьируемых параметра r (размер индивидуальной зоны) и b (потолок для h -вызовов в общей зоне). Следовательно, в данном случае появляются дополнительные возможности для поддержания значений характеристик системы в желаемых пределах. Эти вопросы подробно рассматриваются в следующих параграфах, а также в главе 4 данной части книги.

2.3. Численные эксперименты

Предложенные выше алгоритмы расчета характеристики соты являются очень удобными в вычислительном отношении, так как в отличие от подхода, основанного на решении некоторой системы уравнений опделенной размерности, они не требуют генерации всего пространства состояний, и, более того, эти алгоритмы подразумевают использование табулированных величин (т.е. В-формулы Эрланга). В доступной литературе давно известны соответствующие таблицы, которые являются настольным инструментарием проектировщиков систем телекоммуникаций.

Отмеченные обстоятельства подчеркивают невысокую вычислительную сложность разработанных алгоритмов. На их основе разработаны соответствующие вычислительные программы и проведены объемные численные эксперименты.

Некоторые результаты указанных экспериментов для модели с резервированием каналов показаны на рис 2.9–2.14. Анализ этих результатов позволяет сделать следующие выводы.

1. При фиксированном общем количестве каналов базовой станции соты (N) с ростом числа резервных каналов для h -вызовов (g) функция P_h (вероятность потери h -вызовов) убывает, а функция P_o (вероятность блокировки o -вызовов), наоборот, растет (см. рис.2.9). В частном случае $g = 0$ (т.е. в случаях, когда между h - и o -вызовами не делается никаких различий) значение этих функций, как и следовало ожидать (см. формулы (2.10)), полностью совпадают. Эти факты свидетельствуют о том, что абсолютно справедливое обслуживание (в смысле разницы между значениями P_o и P_h) достигается при $g = 0$, а минимальное справедливое (или абсолютно несправедливое) обслуживание достигается при $g = N-1$.

2. Как и следовало ожидать, при фиксированных значениях параметров N и g вероятности блокировки (потерь) разнотипных вызовов с увеличением нагрузки трафиков любого типа также растут (см. рис.2.10 и

2.11). Однако характер изменения этих функций различный. Так, с ростом нагрузки трафика h -вызовов (т.е. v_2) обе функции P_o и P_h растут с достаточной высокой скоростью (см. рис. 2.10), а с ростом нагрузки трафика o -вызовов (т.е. v_1) функция P_h (вероятность потери h -вызовов) почти не изменяется, а функция P_o (вероятность блокировки o -вызовов) растет с заметной скоростью (см. рис. 2.11).

3. При фиксированных значениях N и нагрузки трафиков обоих типов (v_1 и v_2), функция \tilde{N}_{av} является монотонно убывающей относительно аргумента g . Этот факт также был ожидаемым, так как увеличение числа резервных каналов приводит к ухудшению утилизации общего пула каналов базовой станции соты. При этом скорость изменения функции \tilde{N}_{av} существенным образом зависит от общей нагрузки разнотипных вызовов, т.е. чем больше нагрузка, тем больше скорость убывания функций \tilde{N}_{av} (см. рис. 2.12.).

4. Зависимость функции \tilde{N}_{av} от нагрузок входящих трафиков носит несколько неожиданный, на первый взгляд, характер (см. рис. 2.13, 2.14), т.е. при фиксированном значении нагрузки одного (любого) трафика с ростом нагрузки другого трафика средняя утилизация каналов также растет, и после достижения определенного максимума, она начинает убывать. Это объясняется тем, что с ростом, например, нагрузки h -вызовов (рис. 2.13) они часто теряются и при фиксированной нагрузке o -вызовов и наличии определенных резервных каналов (g) общие каналы плохо используются разнотипными вызовами. Не менее интересным является следующий факт. Оказывается, что максимальное значение \tilde{N}_{av} определяется суммарной нагрузкой ($v = v_1 + v_2$) и не зависит от конкретных значений нагрузок каждого трафика. Так, например, при $N = 100$, $g = 40$ максимальное значение $\tilde{N}_{av} \approx 45$ достигается при $v_1 = 30$ эрл, $v_2 = 20$ эрл (т.е. $v_1 + v_2 = 50$ эрл), а также при $v_1 = 10$ эрл, $v_2 = 40$ эрл (т.е. $v_1 + v_2 = 50$ эрл); то же самое имеет место при $N = 100$, $g = 20$, т.е. в этом случае \tilde{N}_{av} максимальное значение $\tilde{N}_{av} \approx 75$ при $v_1 = 30$ эрл, $v_2 = 40$ эрл (т.е. $v_1 + v_2 = 70$ эрл) и при $v_1 = 40$ эрл, $v_2 = 30$ эрл (т.е. $v_1 + v_2 = 70$ эрл).

Другой важной целью выполнения численных экспериментов была оценка точности предложенных формул. При этом наши результаты сравнивались с результатами работы [5], где получены точные формулы для расчета вероятностей блокировки/потери для вызовов каждого типа в модели с идентичными (по длительности занятия каналов) разнотипными вызовами.

Точные значения P_o и P_h обозначим P'_o и P'_h , соответственно. В указанной работе для расчета этих величин предложены соответствующие формулы, которые приведены в Главе 1 (см. формулы (1.12), (1.13)).

Сравнительный анализ был выполнен в широком диапазоне изменения исходных данных (т.е. структурных и нагрузочных параметров) изучаемой модели. Часть этих сравнений показана в табл. 2.1.

Важно отметить, что при вычислении $P_d(N, g)$ абсолютная разница между нашими результатами и результатами работы [5], не превышает 0,1%, а при $g = 0$ они полностью совпадают. Небольшие отклонения имеют место при вычислении $P_b(N, g)$, но эти отклонения не превышают 10%, что является вполне допустимым при инженерных расчетах.

Таблица 2.1.

Сравнительный анализ результатов при $N = 100$, $\mu = 1.0$, $v_1 = v_2 = 40$ эрл.

g	Предложенный метод		Метод в работе [5]		Разница	
	P_o	P_h	P'_o	P'_h	$ P_o - P'_o $	$ P_h - P'_h $
0	0,003992029	0,003992029	0,003992029	0,003992029	8,67362E-19	0
10	0,064598467	0,001001117	0,042966815	2,70335E-06	0,021631652	0,000998414
20	0,180223675	8,77257E-05	0,13275316	8,75782E-10	0,047470515	8,77248E-05
30	0,334675423	2,4001E-06	0,258395244	1,78746E-13	0,076280179	2,4001E-06
35	0,419364998	2,69735E-07	0,32535538	2,30467E-15	0,094009618	2,69735E-07
40	0,506391932	2,43111E-08	0,391867208	2,84243E-17	0,114524724	2,43111E-08
75	0,895832397	9,27796E-16	0,789940932	6,76465E-31	0,105891465	9,27796E-16
85	0,999998024	7,3159E-16	0,880326743	7,90486E-25	0,119671282	7,3159E-16
90	0,999999996	7,31503E-16	0,922202126	8,47962E-37	0,07779787	7,31503E-16
95	1	7,31503E-16	0,962049928	9,05833E-39	0,037950072	7,31503E-16
99	1	7,31503E-16	0,992555831	2,39246E-40	0,007444169	7,31503E-16

Некоторые результаты численных экспериментов, выполненных на основе алгоритма расчета характеристик системы при использовании стратегии SAC, использующей индивидуальные каналы h -вызовов, показаны на рис. 2.15, 2.16. Анализ их результатов позволяет сделать следующие выводы.

1. Как и следовало ожидать, $P_h(N, r)$ является монотонно убывающей, а $P_o(N, r)$ – наоборот, монотонно возрастающей функцией аргумента r при фиксированном N (рис. 2.15). Иногда представляют практический интерес не только отдельные значения указанных функций, но и их взвешенная сумма – $WRB(N, r)$, которая определяет взвешенную сумму вероятностей потерь разнотипных вызовов:

$$WRB(N, r) := \frac{v_1}{v_1 + v_2} P_o(N, r) + \frac{v_2}{v_1 + v_2} P_h(N, r), \quad (2.48)$$

где в качестве весовых коэффициентов приняты доли соответствующих нагрузок вызовов каждого типа в общей нагрузке. Для полноты представления на рис 2.15 показан также график функции (2.48). При указанных исходных данных взвешенная функция потерь является монотонно возрастающей. Однако в общем случае нельзя делать вывод о монотонности функции (2.48), а так как ее поведение существенным образом зависит от скорости роста и убывания функций $P_o(N, r)$ и $P_h(N, r)$, соответственно, при конкретных значениях параметров (структурных и нагрузочных) системы.

2. Увеличение размера пула индивидуальных каналов для h -вызовов всегда (независимо от значений параметров системы) приводит к неэффективному использованию каналов системы. Этот эффект показан на рис. 2.16.

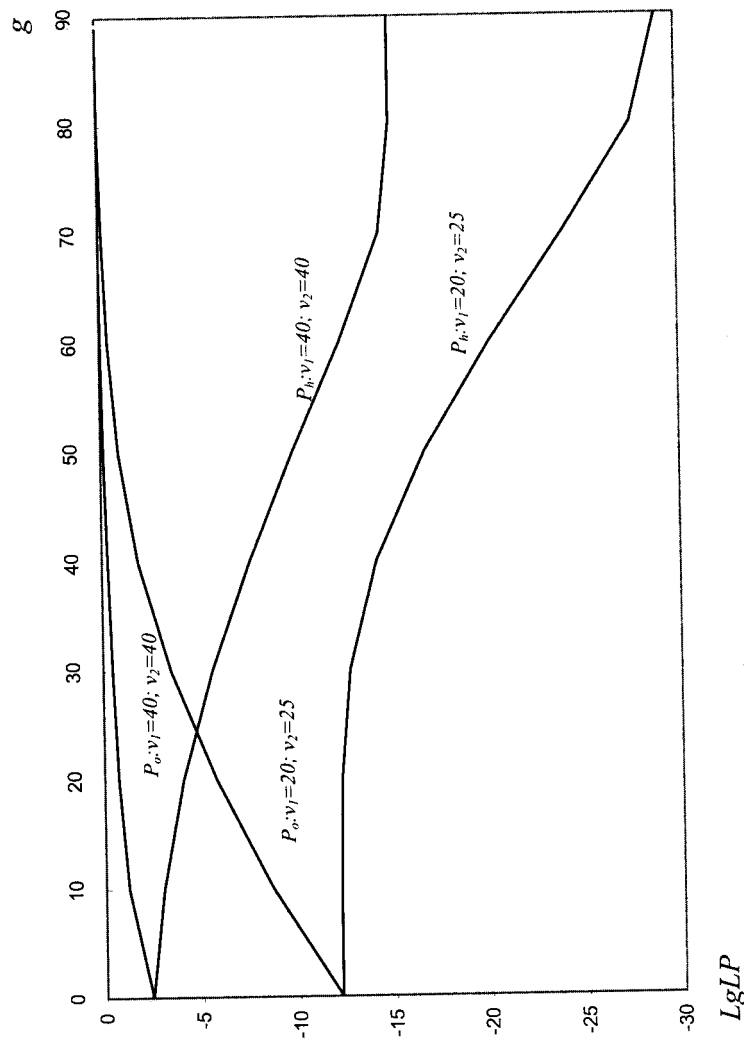


Рис. 2.9. Зависимость вероятностей потерь (LP) от g : $N = 100$

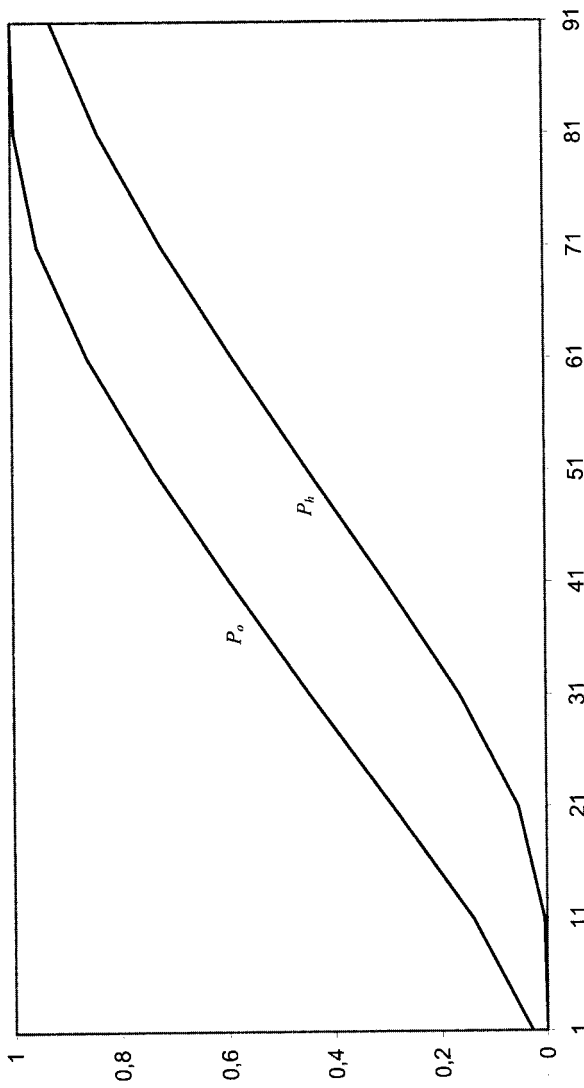


Рис. 2.10. Зависимость вероятностей потерь (LP) от v_1 :
 $N = 100, g = 40, v_1 = 50$

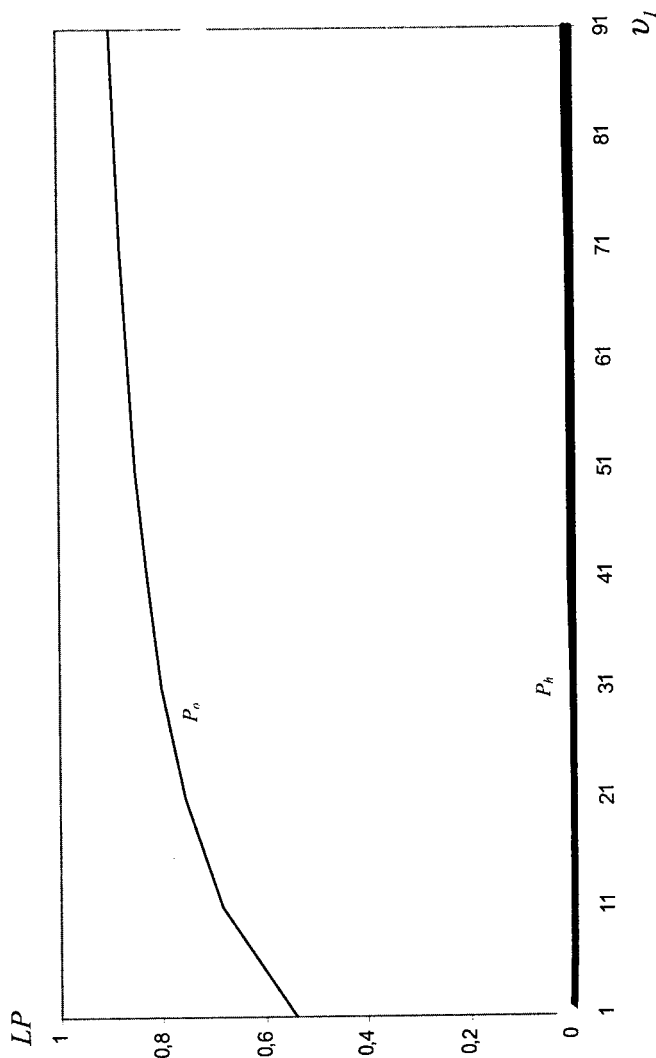


Рис. 2.11. Зависимость вероятностей потерь (LP) от v_1 :
 $N = 100, g = 40, v_2 = 60$

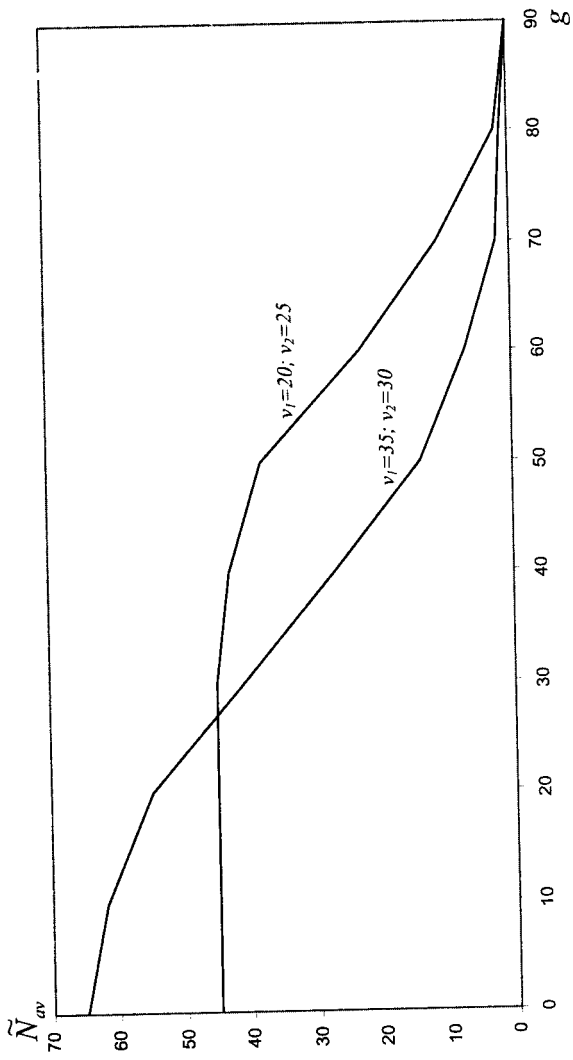


Рис. 2.12. Зависимость \tilde{N}_{av} от g : $N=100$

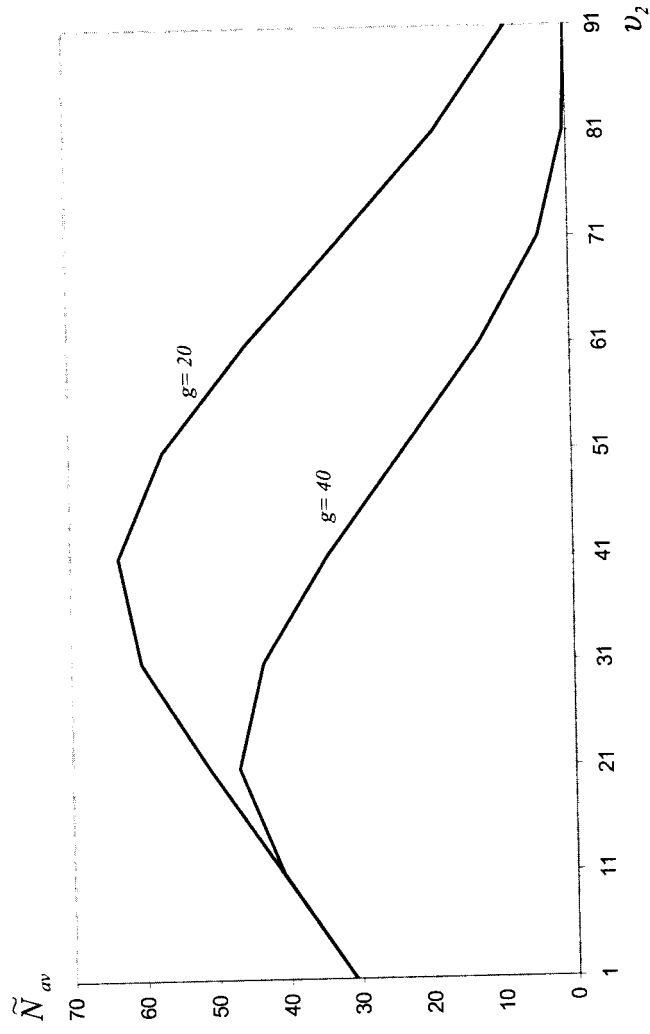


Рис. 2.13. Зависимость \tilde{N}_{av} от ν_2 : $N=100, \nu_1=30$

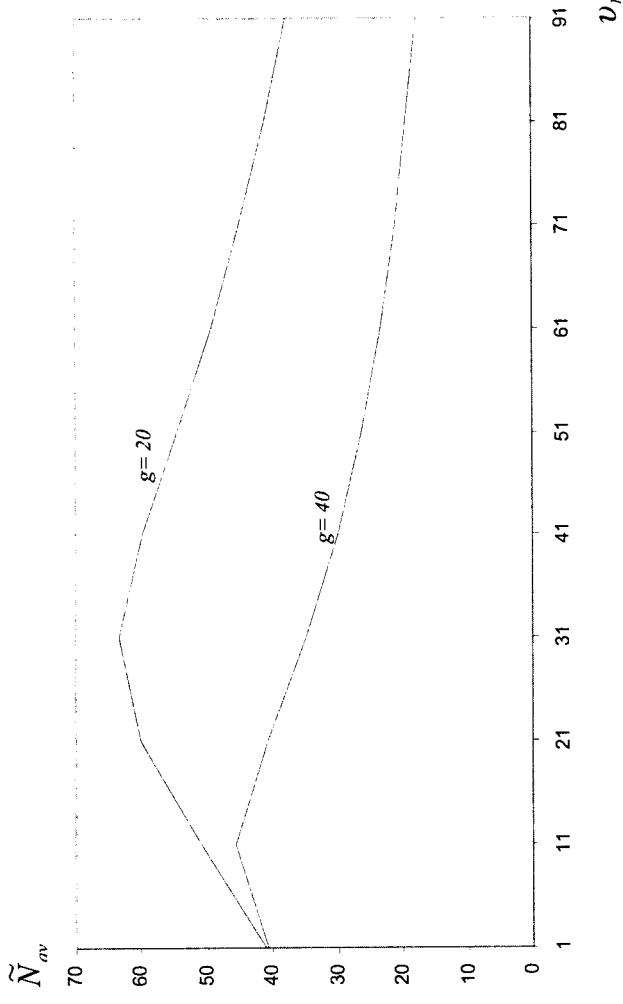


Рис. 2.14. Зависимость \tilde{N}_{av} от ν_1 : $N=100, \nu_2=40$

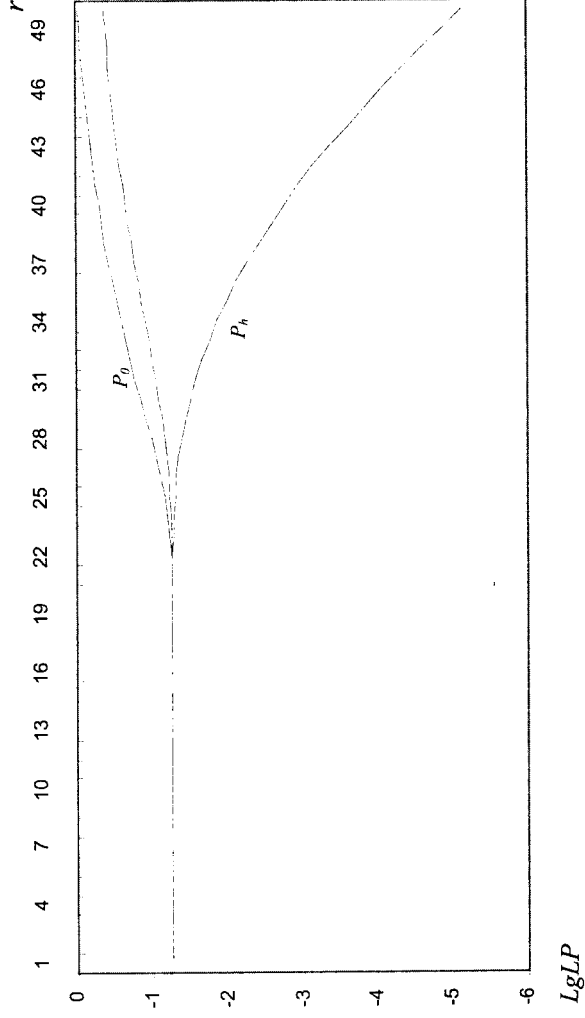


Рис. 2.15. Зависимость вероятностей потерь и соответствующей взвешенной суммы от r : $N=50; \nu_1=20; \nu_2=25$

ристик системы относительно изменения указанных варьируемых параметров являются идентичными.

Некоторые сравнения характеристик системы при различных стратегиях САС показаны в таблицах 2.2 – 2.4, где верхние индексы «*grd*» и «*ind*» обозначают значения соответствующие характеристике GS-стратегии и стратегии на базе индивидуальной зоны для *h*-вызовов, соответственно.

Таблица 2.2

Значения характеристик системы при $N = 100$, $v_1 = v_2 = 50$ эрл

$g(r)$	P_o^{grd}	P_o^{ind}	P_h^{grd}	P_h^{ind}	\tilde{N}_{av}^{grd}	\tilde{N}_{av}^{ind}
1	0.116	0.075	0.057	0.075	85.711	92.430
10	0.288	0.075	0.010	0.075	63.946	92.429
20	0.432	0.075	0.001	0.075	45.495	92.429
30	0.578	0.075	6.27E-05	0.075	30.827	92.429
40	0.721	0.077	1.71E-06	0.074	18.261	92.404
50	0.856	0.119	3.12E-08	0.052	8.377	91.399
60	0.923	0.250	7.82E-10	0.013	1.948	86.079
70	0.999	0.425	1.7E-10	9.8E-04	0.044	78.707
80	0.999	0.612	1.6E-10	2.3E-05	2.3E-05	69.394
90	0.999	0.804	1.6E-10	1.6E-08	6.29E-11	59.764
99	0.999	0.980	1.6E-10	1.6E-10	9.6E-21	50.98

Таблица 2.3

Значения характеристик системы при $N = 100$, $v_1 = 35$ эрл, $v_2 = 50$ эрл

$g(r)$	P_o^{grd}	P_o^{ind}	P_h^{grd}	P_h^{ind}	\tilde{N}_{av}^{grd}	\tilde{N}_{av}^{ind}
1	0.023	0.011	0.011	0.012	82.6	83.9
10	0.129	0.011	0.003	0.012	70	83.9
20	0.282	0.011	0.0004	0.0119	54.3	83.9
30	0.460	0.011	2.5E-05	0.0119	37.8	83.9
40	0.646	0.011	7.9E-07	0.0119	22.9	83.9
50	0.822	0.013	1.7E-08	0.116	10.1	83.9
60	0.957	0.056	5.8E-10	0.006	2.18	82
70	0.999	0.219	1.6E-10	0.0006	0.045	77.2
80	0.999	0.458	1.6E-10	1.5E-05	2.39E-05	68.1
90	0.999	0.724	1.6E-10	8.9E-08	8.9E-11	59
99	0.999	0.972	1.6E-10	3.2E-10	26E-21	50

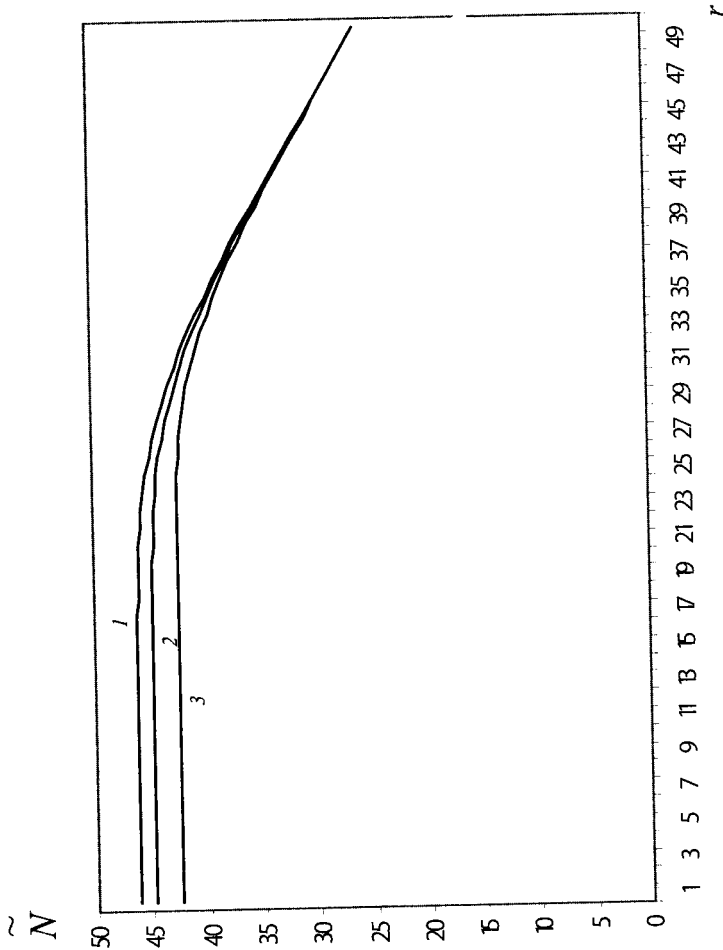


Рис. 2.16. Зависимость среднего числа занятых каналов от r ;

- 1 - $N = 50$; $v_1 = 30$; $v_2 = 25$;
- 2 - $N = 50$; $v_1 = 25$; $v_2 = 25$;
- 3 - $N = 50$; $v_1 = 20$; $v_2 = 25$

3. Следует отметить также очень высокую точность предложенного метода. Так, сравнительный анализ наших результатов и результатов, полученных с применением мультипликативного решения (при небольших значениях N) показывает, что они отличаются друг от друга не более 1%. А по простоте и эффективности предложенный подход намного превосходит подход, основанный на применении мультипликативного решения.

2.4. Сравнительный анализ различных стратегий доступа

В настоящем параграфе приведены результаты сравнительного анализа двух стратегий САС, а именно, стратегий, основанных на GS-стратегии или на использовании индивидуальных каналов для *h*-вызовов. При этом рассматривались случаи симметричных ($v_1 = v_2$) и асимметричных ($v_1 \neq v_2$) нагрузок.

В обеих стратегиях общее число каналов является фиксированным, а варьируемыми являются число резервных каналов g (для САС на базе GS-стратегии) и размер индивидуальной зоны r (для САС на базе индивидуальной зоны). Как известно (см. §§ 2.1, 2.2.), поведение характеристических каналов). Как известно (см. §§ 2.1, 2.2.), поведение характеристических каналов).

Таблица 2.4

Значения характеристик системы при $N = 100$, $v_1 = 50$ эрл, $v_2 = 35$ эрл

$g(r)$	P_o^{grd}	P_o^{ind}	P_h^{grd}	P_h^{ind}	\tilde{N}_{av}^{grd}	\tilde{N}_{av}^{ind}
1	0.021	0.011	0.009	0.011	82.62	83.98
10	0.101	0.011	0.001	0.011	72.26	83.98
20	0.217	0.011	8.2E-05	0.011	59.07	83.98
30	0.356	0.012	1.1E-06	0.011	344.51	83.97
40	0.505	0.029	5.5E-09	0.007	30.69	83.46
50	0.657	0.104	1.0E-11	0.0009	18.67	79.71
60	0.803	0.249	1.09E-14	1.5E-05	9.12	72.50
70	0.933	0.424	1.2E-17	3.8E-08	2.61	63.75
80	0.997	0.612	20E-19	2.1E-11	0.07	54.39
90	0.999	0.804	1.7E-19	1.3E-15	1.5E-06	47.76
99	0.999	0.980	1.7E-19	1.7E-19	2.2E-14	35.98

Сравнительный анализ этих таблиц, а также других результатов объемных вычислительных экспериментов позволяют сделать следующие важные выводы.

1. Для любых допустимых значений параметров N, v_1 и v_2 имеют место следующие соотношения:

$$P_o^{grd}(N, x) > P_o^{ind}(N, x),$$

$$P_h^{grd}(N, x) < P_h^{ind}(N, x),$$

$$\tilde{N}_{av}^{grd}(N, x) < \tilde{N}_{av}^{ind}(N, x),$$

$$WPB^{grd}(N, x) > WPB^{ind}(N, x),$$

где $x \in [1, N - 1]$.

2. Диапазоны изменения вероятностей блокировки/потерь для разнотипных вызовов для обоих САС являются очень близкими, однако скорости их изменения существенно отличаются друг от друга. Так, например, при $N = 100$, $v_1 = 50$ эрл, $v_2 = 35$ эрл диапазоны изменения характеристик системы определяются так:

$$P_o^{grd} \in [0.2142; 0.9999], P_o^{ind} \in [0.1199; 0.9804];$$

$$P_h^{grd} \in [1.72 * 10^{-19}; 0.0097], P_h^{ind} \in [4.86 * 10^{-19}; 0.0119];$$

$$WPB^{grd} \in [0.0166; 0.5882], WPB^{ind} \in [0.0119; 0.5767].$$

3. Стратегия GS использует общую пропускную способность хуже, чем стратегия на базе индивидуальных каналов для h -вызовов. Так, например, при тех же исходных данных ($N = 100$, $v_1 = 50$ эрл, $v_2 = 35$ эрл) лишь для $1 \leq g \leq 33$ при стратегии GS утилизация каналов чуть больше 40%, а уже для $g \geq 40$ утилизация каналов становится даже меньше 30%; однако при использовании САС на базе индивидуальных каналов h -вызовов для этих исходных данных утилизация каналов находится между 35% и 85%. Иными словами, при любых исходных данных стратегия доступа на базе индивидуальной зоны для h -вызовов оказывается лучше стратегии на базе GS в смысле утилизации каналов соты.

Эти выводы показывают актуальность проблемы выбора оптимальной (в заданном смысле) стратегии доступа в каналы соты разнотипных вызовов, т.е. при одних и тех же значениях числа каналов соты, нагрузках и желаемых диапазонах изменений значений показателей QoS , одна из этих стратегий может удовлетворять их, а другая – нет. Так, например, для $N = 100$, $v_1 = 50$ эрл, $v_2 = 35$ эрл и заданных требований $P_b \leq 0.1$, $P_d \leq 0.007$, $\tilde{N}_{av} \geq 80$ стратегия на базе GS ни при каком значении параметра g не удовлетворяет указанным ограничениям, в то время как стратегия на базе индивидуальной зоны для h -вызовов удовлетворяет указанным ограничениям при $r^* = 40$. Вместе с тем, при тех же исходных данных и заданных ограничениях $P_b \leq 0.3$, $P_d \leq 0.0001$, $\tilde{N}_{av} \geq 60$, стратегия на базе GS удовлетворяет этим ограничениям при $g^* = 20$; в то же время стратегия на базе индивидуальной зоны для h -вызовов ни при каком значении параметра r не удовлетворяет этим ограничениям.

Поскольку обе стратегии имеют одинаковую степень сложности в реализации, то при каждом конкретном случае необходимо провести серьезные исследования по выбору соответствующих стратегий доступа. Эти вопросы подробно рассматриваются в четвертой главе.

Заключение и комментарии к главе 2 части I

Поскольку в доступной литературе почти отсутствуют работы по изучению моделей процессов обработки разнотипных вызовов в беспроводных сетях связи с неидентичными новыми и хэндовер-вызовами и явными потерями, то результаты данной главы полностью базируются на исследованиях, выполненных в работах [1-4]. С целью оценки точности предложенных формул необходимые сравнения проводились с соответствующими результатами работы [5].

1. **Melikov A.Z., Babayev A.T.** A new method of performance analysis of queuing model with guard channels // Proceedings of Fifth International Workshop on Retrial Queues. Seoul, Korea, September 2004. — pp. 103 — 110.
2. **Melikov A.Z., Fattakhova M.I., Babayev A.T.** Calculation and optimization of call processing procedures in cellular wireless communication networks // Automatic control and computer sciences. — 2004. — Vol. 38, No.4. — pp. 55 — 63.
3. **Melikov A.Z., Fattakhova M.I., Babayev A.T.** Investigation of cellular communication networks with private channels for service of handover calls // Automatic Control and Computer Sciences. — 2005. — Vol.39, No.3. — pp.61 — 69.
4. **Melikov A.Z., Babayev A.T.** Refined approximations for performance analysis and optimization of queuing model with guard channels for handovers in cellular networks // Computer Communications. — 2006. — vol.29, no.9. — pp. 1386 — 1392.
5. **Haring G., Marie R., Puigjaner R., Trivedi K.** Loss formulas and their application to optimisation for cellular networks // IEEE Transactions on Vehicular Technology. — 2001. — Vol.50, № 3. — pp. 664 — 673.

МЕТОДЫ РАСЧЕТА ХАРАКТЕРИСТИК МОДЕЛЕЙ ПРОЦЕССОВ ОБСЛУЖИВАНИЯ ВЫЗОВОВ В БЕСПРОВОДНЫХ СЕТЯХ С ОЧЕРЕДЯМИ

В данной главе исследуются модели процессов обслуживания вызовов в беспроводных сетях сотовой связи, в которых предлагается организация очереди новых и/или хэндовер-вызовов.

Поскольку h -вызовы являются более чувствительными к возможным потерям и задержкам, чем новые o -вызовы, то в литературе предложены различные схемы предоставления приоритетов h -вызовам. Эти схемы, главным образом, подразумевают использование резервных или индивидуальных каналов для h -вызовов (см. главу 2) и/или организацию их очереди в BS. Организацию очереди h -вызовов можно реализовать в сетях, в которых микросоты покрываются некоторой макросотой, т.е. имеется некоторая зона (хэндовер-зона), внутри которой MS может обслуживаться в любой из соседних сот (см. рис. 1.4). Время, за которое MS пересекает хэндовер-зону, называется интервалом деградации. Когда MS поступает в хэндовер-зону, проверяется наличие свободных каналов в новой соте. Если в новой соте имеется свободный канал, то h -вызов немедленно занимает его, и процедура хэндовера считается успешно завершённой на данном этапе; в противном случае канал старой (т.е. предыдущей) соты продолжает использоваться данный h -вызов и одновременно он становится в очередь для ожидания освобождения какого-либо канала новой соты. Если до завершения интервала деградации в новой соте не становится доступным свободный канал, то происходит вынужденное прерывание разговора h -вызова.

Следует отметить, что с целью улучшения показателей качества обслуживания o -вызовов и компенсации их шансов необходима также организация очереди и для этого типа вызовов. Поэтому возникает необходимость исследования моделей обслуживания, в которых наряду с использованием резервных каналов для h -вызовов вводится буфер конечного или бесконечного размера для o -вызовов.

При использовании схемы с буферизацией вызовов зачастую предполагается, что при освобождении канала новой соты очередь вызовов (новых или хэндовер) обслуживается согласно дисциплине FIFO (First — In — First — Out).

3.1. Расчет характеристик модели с очередями хэндовер-вызовов

На рис. 3.1 показана структурная схема модели BS соты, которая содержит $N > 1$ радио каналов и буфер конечного размера $B < \infty$ лишь для ожидания в очереди h -вызовов.

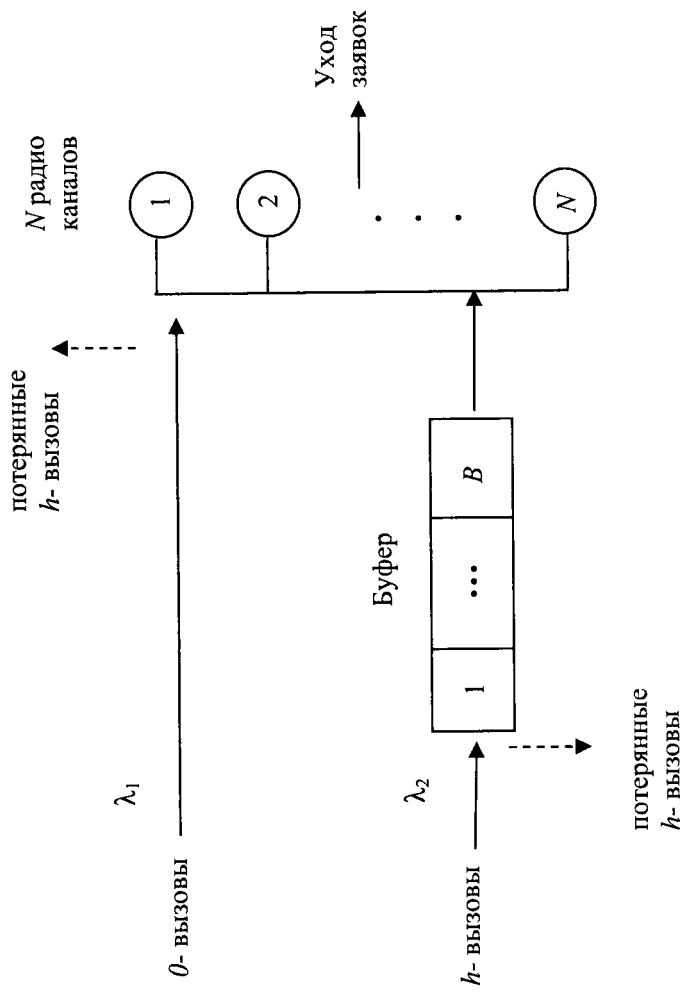


Рис. 3.1. Модель базовой станции соты с ограниченной очередью h -вызовов

Предполагается, что o -вызовы (h -вызовы) поступают в данную соту согласно закону Пуассона с интенсивностью λ_o (λ_h), и время, требуемое этими вызовами на занятие канала, (без учета процедуры хэндовер) является экспоненциально распределенной случайной величиной со средним μ^{-1} . Если в период обслуживания o -вызова происходит процедура хэндовер, то оставшееся время обслуживания данного вызова в новой соте (уже в качестве h -вызова) также имеет экспоненциальное распределение с тем же средним μ^{-1} с учетом отсутствия памяти у экспоненциального распределения.

Замечание 3.1. Для простоты изложения здесь предполагается, что разнотипные вызовы имеют одинаковые интенсивности обслуживания. Вместе с тем, описанный ниже математический аппарат остается неизменным и для моделей с различными интенсивностями обслуживания o - и h -вызовов.

Обслуживание вызовов осуществляется согласно схеме резервирования каналов, т.е. поступивший o -вызов принимается лишь тогда, когда число свободных каналов BS не меньше, чем $g+1$; в противном случае o -вызов теряется (блокируется); h -вызов принимается при наличии хотя бы одного свободного канала. Если все N каналов являются занятыми, то h -вызов присоединяется к очереди при наличии там хотя бы одного

свободного места; в противном случае поступивший h -вызов теряется. Он не может быть потерян из очереди, иными словами, предполагается, что ожидающие h -вызовы являются весьма терпеливыми.

В момент освобождения канала очередь h -вызовов (если она имеется) обслуживается согласно дисциплине FIFO; если очередь отсутствует, то освобожденный канал простаивает.

Схема алгоритма обслуживания разнотипных вызовов в данной модели показана на рис 3.2.

Для более детального описания работы соты используется двумерная цепь Маркова, т.е. состояние соты в произвольный момент времени задается вектором $k = (k_1, k_2)$, где k_i означает число o -вызовов (h -вызовов) в системе, $i = 1, 2$. Тогда множество всех возможных состояний системы определяется следующим образом:

$$S = \{k: k_1 = 0, g, k_2 = 0, N + B; k_1 + k_2 \leq N + B\}. \quad (3.1)$$

Поскольку o -вызовы обслуживаются в режиме блокировки, то в состоянии $k \in S$ число h -вызовов в каналах (k_2^g) и в очереди (k_2^q) определяется так:

$$k_2^g = \min\{N - k_1, k_2\}, \quad k_2^q = (k_1 + k_2 - N)^+, \quad (3.2)$$

где $x^+ = \max(0, x)$.

Замечание 3.2. В известных [3-8] работах состояние соты описывается скалярной величиной, которая указывает общее число занятых каналов BS, т.е. в качестве математической модели используется одномерная цепь Маркова. Такой способ описания состояния соты позволяет определить лишь вероятности блокировки вызовов каждого типа. Однако такой подход не позволяет определить характеристики индивидуальных потоков, что является особенно полезным, в частности, при определении оптимального (в некотором смысле) значения числа резервных каналов для h -вызовов. С другой стороны, если интенсивности обслуживания разнотипных вызовов отличаются друг от друга, то описание состояния соты скалярной величиной становится в принципе невозможным, т.е. предложенный здесь способ описания состояния соты является достаточно подробным.

Элементы производящей матрицы соответствующей цепи Маркова, $q(k, k')$, $k, k' \in S$, определяются так:

$$q(k, k') = \begin{cases} \lambda_o, & \text{если } k_1 + k_2 \leq N - g - 1, k' = k + e_1, \\ \lambda_h, & \text{если } k' = k + e_2, \\ k_1 \mu, & \text{если } k' = k - e_1, \\ k_2^g \mu, & \text{если } k' = k - e_2, \\ 0, & \text{в остальных случаях,} \end{cases} \quad (3.3)$$

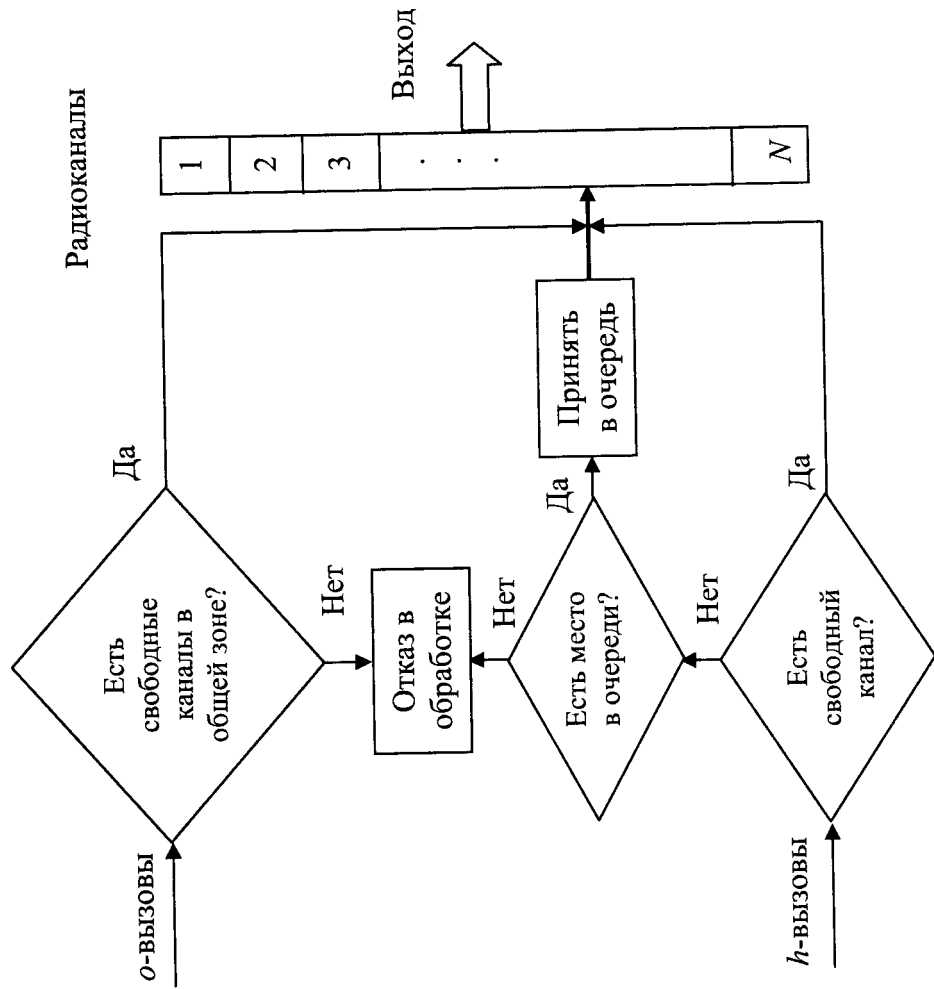


Рис. 3.2. Схема алгоритма САС с ограниченной очередью для ожидания терпеливых вызовов

Граф переходов данной цепи показан на рис. 3.3. Следовательно, математической моделью данной системы является двумерная цепь Маркова с фазовым пространством состояний (3.1), для которой элементы производящей матрицы определяются с помощью соотношений (3.3).

Стационарная вероятность состояния $\mathbf{k} \in S$ обозначается $p(\mathbf{k})$. Тогда исходя из того, что модель является марковской, согласно теореме PASTA вероятность потери h -вызовов (P_h) и вероятность блокировки o -вызовов (P_o) в данной системе определяются следующим образом:

$$P_h(N, g, B) := \sum_{\mathbf{k} \in S} p(\mathbf{k}) I(k_1 + k_2 = N + B), \quad (3.4)$$

$$P_o(N, g, B) := \sum_{\mathbf{k} \in S} p(\mathbf{k}) I(k_1 + k_2^s \geq N - g), \quad (3.5)$$

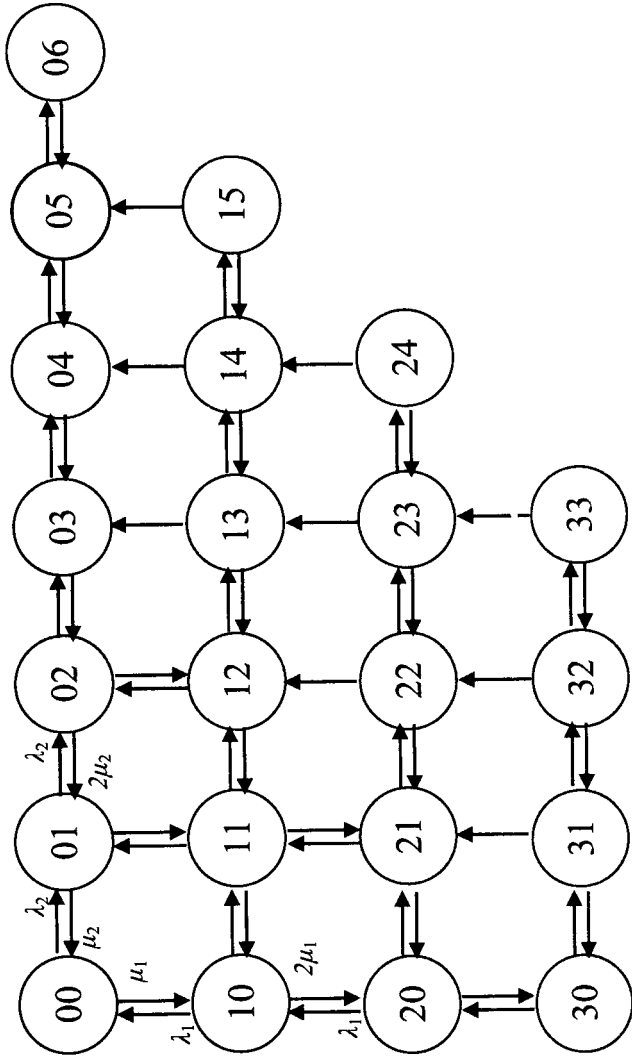


Рис. 3.3. Граф переходов для модели с терпеливыми h -вызовами при $N = 6, g = 3, B = 2$

Среднее число занятых каналов (\tilde{N}_{av}), o -вызовов (K_o) и h -вызовов (K_h) в системе определяются также через стационарное распределение модели:

$$\tilde{N}_{av}(N, g, B) := \sum_{j=1}^N j \sum_{\mathbf{k} \in S^j} p(\mathbf{k}), \quad (3.6)$$

$$K_o(N, g, B) := \sum_{j=1}^{N-g} j \sum_{\mathbf{k} \in S^j} p(\mathbf{k}), \quad (3.7)$$

$$K_h(N, g, B) := \sum_{r=1}^{N+B} r \sum_{\mathbf{k} \in S_r^j} p(\mathbf{k}), \quad (3.8)$$

где $S^j := \{\mathbf{k} \in S : k_1 + k_2^s = j\}$, $S_r^j := \{\mathbf{k} \in S : k_1 = j, i = 1, 2\}$.

С использованием (3.4) и (3.8) из модифицированной формулы Литтла находится среднее время ожидания h -вызовов в системе (W_h):

$$W_h(N, g, B) := \frac{K_h}{\lambda_h(1 - P_d)} - \mu^{-1}. \quad (3.9)$$

Следовательно, для нахождения характеристик (3.4) – (3.9) необходимо определить стационарное распределение модели ($p(\mathbf{k}), \mathbf{k} \in S$) из соответствующей СУР. Она имеет следующий вид:

$$p(\mathbf{k})(\lambda_0 I(k_1 + k_2^s \leq g - 1) + \lambda_{\eta}(1 - \delta_{k_1+k_2, N+B}) - (k_1 + k_2^s)\mu) =$$

$$= \lambda_0 p(\mathbf{k} - \mathbf{e}_1)(1 - \delta_{k_1, 0})I(k_1 + k_2^s \leq g - 2) + \lambda_{\eta} p(\mathbf{k} - \mathbf{e}_2)(1 - \delta_{k_2, 0}) +$$

$$+ (k_1 + 1)\mu p(\mathbf{k} + \mathbf{e}_1)I(k_1 < N - g - 1) + (k_2^s + 1)\mu p(\mathbf{k} + \mathbf{e}_2)I(k_2^s < N), \quad \mathbf{k} \in S, \quad (3.10)$$

$$\sum_{\mathbf{k} \in S} p(\mathbf{k}) = 1. \quad (3.11)$$

Однако поиск решения СУР (3.10), (3.11) сталкивается с огромными вычислительными трудностями при больших значениях N и B , так как она, вообще говоря, не имеет явного решения. Для преодоления указанных трудностей предлагается использовать приближенный метод расчета стационарного распределения данной модели.

Рассматривается следующее расщепление ФПС (3.1):

$$S = \bigcup_{j=0}^{N-g} S_j^j, \quad S_j^j \cap S_l^l = \emptyset, \quad j \neq l, \quad (3.12)$$

где множества S_j^j , $j = \overline{0, N-g}$ определены в (3.7), (3.8).

Множества S_j^j объединяются в отдельные укрупненные состояния $\langle j \rangle$, и вводится функция укрупнения с областью определения (3.1):

$$U(\mathbf{k}) = \langle j \rangle, \text{ если } \mathbf{k} \in S_j^j, j = \overline{0, N-g}, \quad (3.13)$$

Функция укрупнения (3.13) определяет укрупненную модель, которая является одномерной целью ЦМ с ФПС $\tilde{S} := \{ \langle j \rangle : j = \overline{0, N-g} \}$.

Согласно алгоритму фазового укрупнения для нахождения стационарного распределения исходной модели требуется предварительное определение стационарных распределений расщепленных и укрупненных моделей.

Стационарное распределение j -ой расщепленной модели с ФПС S_j^j обозначается $(\rho^j(m), m = \overline{0, N+B-j})$. Оно определяется как стационарное распределение системы обслуживания $M | M | N-j | B$ с нагрузкой v_{η}

э.р.л., $v_{\eta} := \lambda_{\eta} / \mu$:

$$\rho^j(m) = \begin{cases} \frac{v_{\eta}^m}{m!} \rho^j(0), & \text{если } m = \overline{1, N-j}, \\ \frac{v_{\eta}^m}{(N-j)!(N-j)^{m+j-N}} \rho^j(0), & \text{если } m = \overline{N-j+1, N-j+B}, \end{cases} \quad (3.14)$$

где

$$\rho^j(0) = \left(\sum_{m=0}^{N-j} \frac{v_{\eta}^m}{m!} + \frac{1}{(N-j)!} \sum_{m=N-j+1}^{N-j+B} \frac{v_{\eta}^m}{(N-j)^{m+j-N}} \right)^{-1}. \quad (3.15)$$

Для нахождения стационарного распределения $(\pi(\langle j \rangle), \langle j \rangle \in \tilde{S})$ укрупненной модели в первую очередь потребуется определение элементов производящей матрицы соответствующей укрупненной цепи. Обозначим их $q(\langle j' \rangle, \langle j'' \rangle), \langle j' \rangle, \langle j'' \rangle \in \tilde{S}$. Они определяются из следующих соотношений:

$$q(\langle j' \rangle, \langle j'' \rangle) = \begin{cases} \lambda_0 \cdot \Lambda(j' + 1), & \text{если } j'' = j' + 1, \\ j' \mu, & \text{если } j'' = j' - 1, \\ 0, & \text{в остальных случаях,} \end{cases} \quad (3.16)$$

$$\text{где } \Lambda(k+1) = \rho^k(0) \sum_{j=0}^{N-g-k-1} \frac{v_{\eta}^j}{j!}, \quad k = \overline{0, N-g-1}.$$

Далее стационарное распределение укрупненной модели $\pi(\langle j \rangle), \langle j \rangle \in \tilde{S}$ определяется как стационарное распределение классического процесса размножения и гибели с интенсивностями, заданными соотношениями (3.16):

$$\pi(\langle j \rangle) = \frac{v_{\eta}^j}{j!} \prod_{i=1}^j \Lambda(i) \pi(\langle 0 \rangle), \quad j = \overline{1, N-g}, \quad (3.17)$$

где $v_{\eta} := \lambda_{\eta} / \mu$,

$$\pi(\langle 0 \rangle) = \left(1 + \sum_{i=1}^{N-g} \frac{v_{\eta}^i}{i!} \prod_{j=1}^i \Lambda(j) \right)^{-1}. \quad (3.18)$$

Таким образом, стационарное распределение исходной модели приближенно определяется так:

$$p(k_1, k_2) \approx \rho^{k_1}(k_2) \pi(\langle k_1 \rangle), (k_1, k_2) \in S. \quad (3.19)$$

Следовательно, с учетом (3.14) – (3.19) находятся следующие приближенные формулы для расчета характеристик (3.4)–(3.8):

$$P_{\eta}(N, g, B) \approx \sum_{j=0}^{N-g} \rho^j(N+B-j) \pi(\langle j \rangle); \quad (3.20)$$

$$P_0(N, g, B) \approx \sum_{j=0}^{N-g} \sum_{i=N-g-j}^{N+B-j} \rho^j(i) \pi(\langle j \rangle); \quad (3.21)$$

$$\begin{aligned} \tilde{N}_{av}(N, g, \infty) &\approx \sum_{j=1}^{N-g} j \sum_{i=0}^j \rho^i(j-i) \pi(<i>) + \\ &+ \sum_{j=N-g+1}^{N-1} j \sum_{i=0}^{N-g} \rho^i(k-i) \pi(<i>) + N \left(1 - \sum_{i=0}^{N-g} \sum_{j=0}^{N-1-i} \rho^i(j) \pi(i) \right). \end{aligned} \quad (3.28)$$

Теперь рассмотрим сети с ограниченной очередью нетерпеливых h -вызовов. Они являются актуальными, поскольку из-за ограниченности размеров сот и конечной скорости мобильных пользователей в реальных сетях интервал деградации (т.е. время, за которое мобильный пользователь пересекает хэндовер-зону) всегда является случайной величиной с конечным средним, и этот факт существенным образом влияет на характеристики сетей сотовой связи.

Физическая модель сети идентична той, которая была исследована в предыдущем параграфе. Единственная разница состоит в том, что в данной модели нетерпеливый h -вызов может быть потерян и из очереди, если до окончания интервала его деградации не освобождается ни один канал новой соты. С целью получения обобщенных результатов предполагается, что интервалы деградации для всех h -вызовов являются независимыми случайными величинами и имеют одинаковые экспоненциальные распределения со средним γ^{-1} . Как и в случае, описанном выше, в момент освобождения канала очередь h -вызовов (если она имеется) обслуживается согласно дисциплине FIFO; если очередь отсутствует, то освобожденный канал простаивает.

Схема алгоритма обслуживания разнотипных вызовов в данной модели показана на рис. 3.4.

Состояние соты в произвольный момент времени также задается вектором $\mathbf{k} = (k_1, k_2)$, где k_i означает количество o -вызовов (h -вызовов) в системе, $i = 1, 2$. Тогда множество всех возможных состояний системы совпадает с (3.1). Поскольку o -вызовы обслуживаются в режиме блокировки, и система является консервативной (т.е. не допускается простой каналов при наличии очереди h -вызовов), то в состоянии $\mathbf{k} \in S$ число h -вызовов в каналах (k_2^h) и в очереди (k_2^g) определяются из (3.2).

Однако, элементы производящей матрицы $q(\mathbf{k}, \mathbf{k}')$, $\mathbf{k}, \mathbf{k}' \in S$, в данной модели отличаются от (3.3) и определяются с помощью следующих соотношений:

$$q(\mathbf{k}, \mathbf{k}') = \begin{cases} \lambda_o, & \text{если } k_1 + k_2 \leq N - g - 1, \mathbf{k}' = \mathbf{k} + \mathbf{e}_1, \\ \lambda_h, & \text{если } \mathbf{k}' = \mathbf{k} + \mathbf{e}_2, \\ k_1 \mu, & \text{если } \mathbf{k}' = \mathbf{k} - \mathbf{e}_1, \\ k_2^g \mu (k_2^g = 0) + (k_2^g \mu + k_2^g \gamma) I(k_2^g > 0), & \text{если } \mathbf{k}' = \mathbf{k} - \mathbf{e}_2, \\ 0, & \text{в остальных случаях.} \end{cases} \quad (3.29)$$

$$\begin{aligned} \tilde{N}_{av}(N, g, B) &\approx \sum_{j=1}^{N-g} j \sum_{i=0}^j \rho^i(j-i) \pi(<i>) + \\ &+ \sum_{j=N-g+1}^{N-1} j \sum_{i=0}^{N-g} \rho^i(k-i) \pi(<i>) + N \sum_{i=0}^{N-g} \sum_{j=N-1-i}^{N+B-1} \rho^i(j) \pi(<i>); \end{aligned} \quad (3.22)$$

$$K_o(N, g, B) \approx \sum_{j=1}^g j \pi(<j>); \quad (3.23)$$

$$K_h(N, g, B) \approx \sum_{j=1}^{N+B-g} j \sum_{i=0}^g \rho^i(j) \pi(<i>) + \sum_{j=N+B-g+1}^{N+B} j \sum_{i=0}^{N+B-j} \rho^i(j) \pi(<i>). \quad (3.24)$$

Если предположение об ограниченности длины очереди o -вызовов не соответствует действительности, то необходимо рассматривать модель с неограниченной очередью. Для исследования этой модели также может быть использован вышеописанный подход. Поскольку выбранное разбиение ФПС исходной модели полностью определяет структуру расщепленных и укрупненной модели, то, используя аналогичное (3.12) разбиение бесконечного пространства состояний данной модели и выполнив уже известные процедуры, находится стационарное распределение внутри классов $S'_j, j = 0, 1, 2, \dots$. В этом случае оно находится как стационарное распределение системы обслуживания $M|M|N-j | \infty$ с нагрузкой v_h , зрл.:

$$\rho^j(i) = \begin{cases} \frac{v_h^i}{i!} \rho^j(0), & i = \overline{1, N-j}, \\ \frac{v_h^i (N-j)^{N-j-i}}{(N-j)!} \rho^j(0), & i \geq N-j, \end{cases} \quad (3.25)$$

$$\rho^j(0) = \left(\sum_{i=0}^{N-j-1} \frac{v_2^i}{i!} + \frac{v_2^{N-j}}{(N-j)!} \cdot \frac{N-j}{N-j-v_2} \right)^{-1}. \quad (3.26)$$

Условием эргодичности модели является $V_h < 1$. Стационарное распределение укрупненной модели также вычисляется согласно формулам (3.16) – (3.18), однако здесь необходимо учитывать, что $\rho^j(0), j = 0, 1, 2, \dots$ определяются из формул (3.25), (3.26).

После выполнения необходимых математических преобразований находятся следующие формулы (приближенные) для вычисления характеристик модели с бесконечной очередью h -вызовов при наличии резервных каналов:

$$P_o(N, g, \infty) \approx 1 - \sum_{j=0}^{N-g-1} \sum_{i=0}^{N-g-1-j} \rho^j(i) \pi(j) + \pi(N-g), \quad (3.27)$$

$$P_h^{(v)} := \sum_{k \in S_h^g} p(k) + \frac{1}{\lambda_h} \sum_{i=1}^B i \gamma \sum_{j=0}^{N-g} p(j, N+i-j). \quad (3.31)$$

В последней формуле первый член суммы определяет вероятность появления события (i), а второй — вероятность появления события (ii).

Среднее число занятых каналов ($\tilde{N}_{av}^{(v)}$), о-вызовов ($K_o^{(v)}$) и h-вызовов ($K_h^{(v)}$) в сети определяются аналогично (3.6)–(3.8). Стационарное распределение модели определяется из соответствующей системы уравнений равновесия. Она составляется на основе (3.29) и имеет следующий вид:

$$\begin{aligned} p(k) (\lambda_o I(k_1 + k_2^s \leq g-1) + \lambda_h (1 - \delta_{k_1+k_2, N+B}) + (k_1 + k_2^s \mu + k_2^g \gamma)) = \\ = \lambda_o p(k - e_1) (1 - \delta_{k_1, 0}) I(k_1 + k_2^s \leq g-2) + \lambda_h p(k - e_2) (1 - \delta_{k_2, 0}) + \\ + (k_1 + 1) \mu p(k + e_1) I(k_1 < N-g-1) + (k_2^s + 1) \gamma p(k + e_2) I(k_2 < N-1) + \\ + k_2^g p(k + e_2), k \in S, \end{aligned} \quad (3.32)$$

$$\sum_{k \in S} p(k) = 1. \quad (3.33)$$

Однако нахождение стационарного распределения из системы уравнений (3.32) и (3.33) и в этом случае сталкивается с огромными вычислительными трудностями при больших значениях N и B , так как эта система не имеет явного решения. Зачастую решение подобных задач становится очевидным, если существует соответствующая цепь обладает свойством обратимости, и потому для нее существует стационарное распределение мультипликативного вида. Однако с использованием критерия Колмогорова для двумерных цепей легко можно показать, что данная цепь не является обратимой. Действительно, согласно этому критерию необходимо условие обратимости двумерной ЦМ состоит в том, что если существует переход из состояния (i, j) в состояние (i, j) , то должен существовать и обратный переход из состояния (i, j) в состояние (i, j) . Однако для рассматриваемой цепи это условие не выполняется. Так, например, согласно соотношениям (3.29) в данной ЦМ существует переход $(k_1, k_2) \rightarrow (k_1-1, k_2)$ с интенсивностью $k_1 \mu$, где $k_1 + k_2 > N-g$, однако обратный переход не существует.

Для преодоления указанных трудностей ниже используется описанный выше приближенный метод расчета стационарного распределения двумерных цепей Маркова.

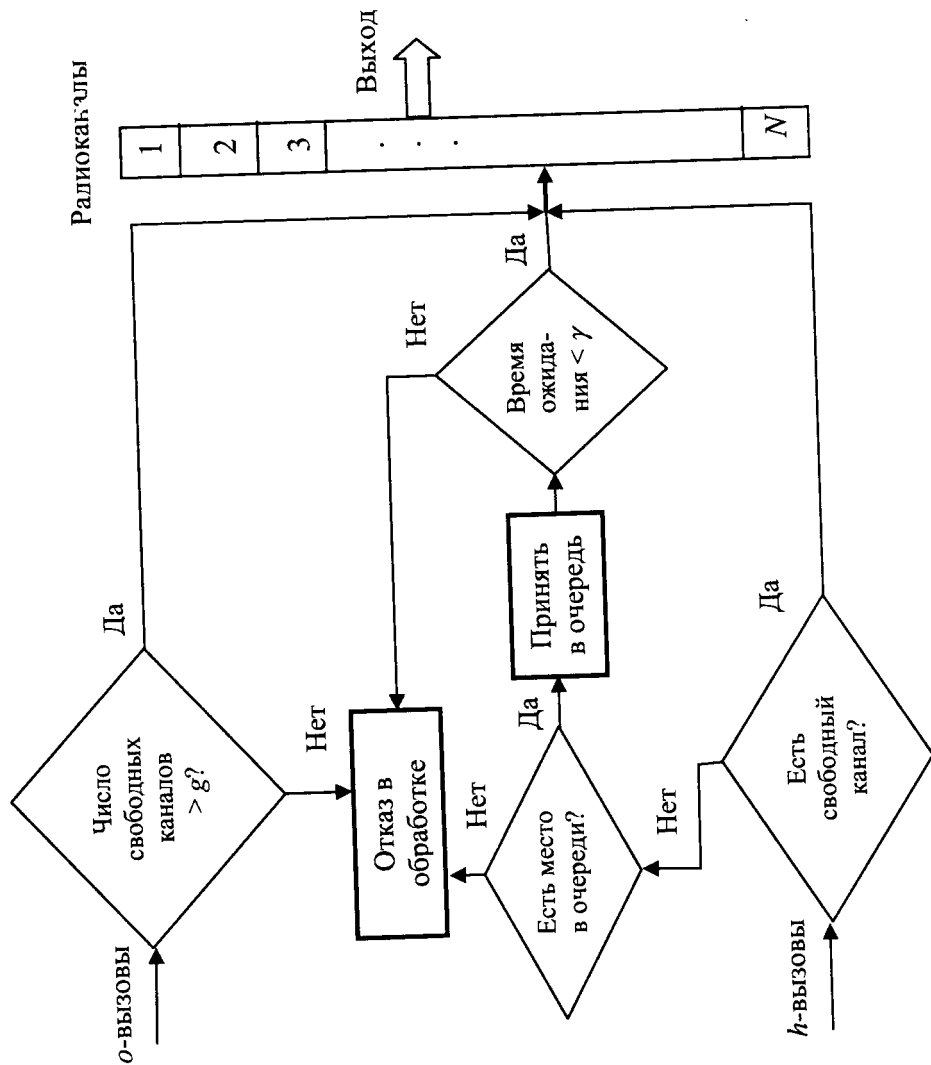


Рис. 3.4. Схема алгоритма САС с ограниченной очередью для ожидания нетерпеливых h-вызовов

Граф переходов данной цепи показан на рис 3.5. Вероятность блокировки о-вызовов в данной системе определяется аналогично (3.5):

$$P_o^{(v)} := \sum_{k \in S} p(k) I(k_1 + k_2^s \geq N-g), \quad (3.30)$$

Однако в этой модели метод расчета вероятности потери h-вызовов отличается от аналогичного метода модели с терпеливыми h-вызовами, так как здесь потери h-вызовов происходят при наступлении следующих событий:

- (i) в момент поступления h-вызова в очереди уже имеются B таких вызовов;
- (ii) интервал деградации h-вызова завершается раньше, чем он получает доступ к свободному каналу.

Следовательно, искомая характеристика определяется так:

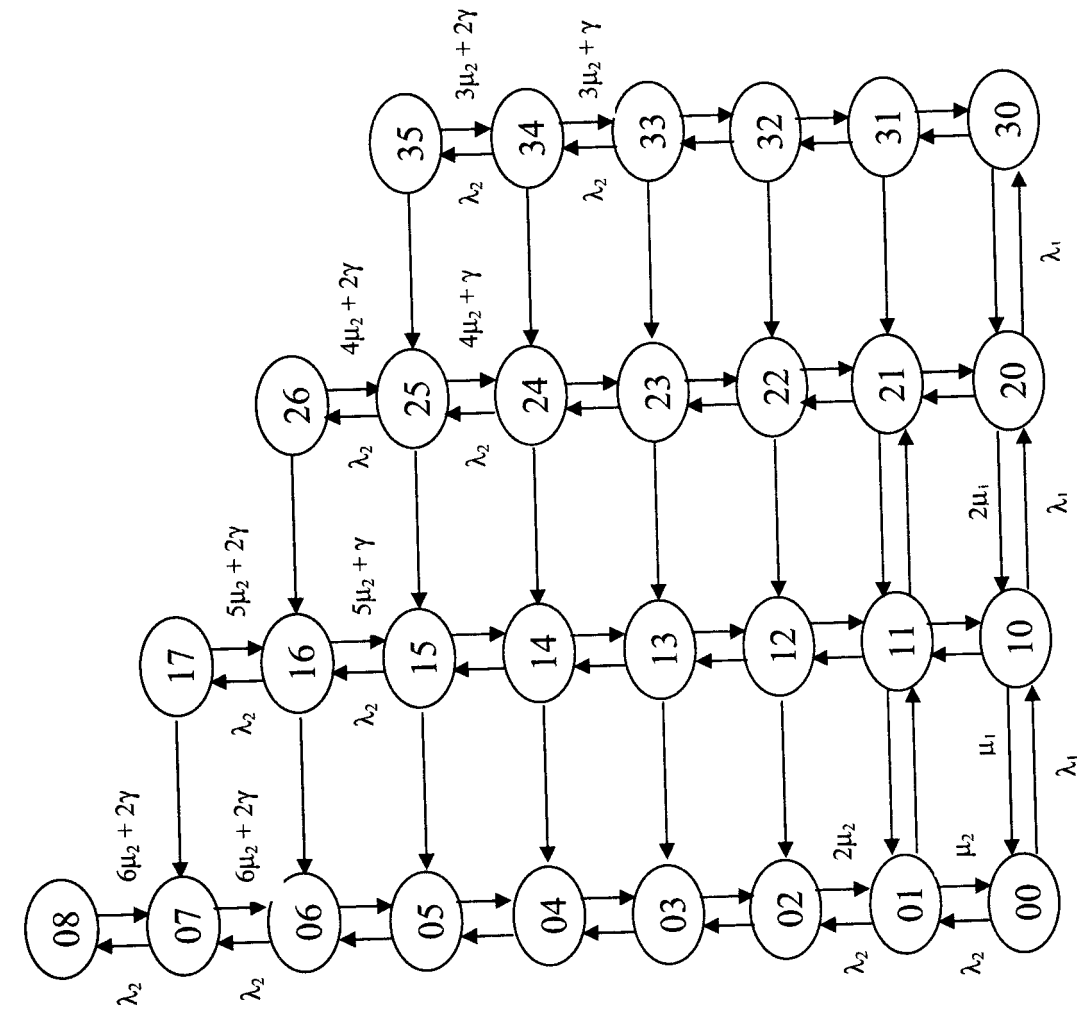


Рис.3.5. Граф переходов для модели с нетерпеливыми h -вызовами

Рассматривается расщепление (3.12) исходного ФПС (3.1) и далее на основе функции укрупнения (3.13) определяется укрупненная модель, которая является одномерной цепью Маркова с ФПС $\tilde{S} := \{< j >: j = \overline{0, N-g}\}$. В этом случае стационарное распределение j -й расщепленной модели отличается от (3.14), (3.15) и определяется так:

$$\rho^j(m) = \begin{cases} \frac{v_h^m}{m!} \rho^j(0), & \text{если } m = \overline{1, N-j}, \\ \frac{v_h^{N-j}}{(N-j)!} \prod_{i=1}^{m+j-N} \frac{\lambda_h^i}{(N-j)\mu + i\gamma} \rho^j(0), & \text{если } m = \overline{N-j+1, N-j+B}, \end{cases} \quad (3.34)$$

где

$$\rho^j(0) = \left(\sum_{m=0}^{N-j} \frac{v_h^m}{m!} + \frac{v_h^{N-j}}{(N-j)!} \sum_{m=N-j+1}^{m+j-N} \prod_{i=1}^{m+j-N} \frac{\lambda_h^i}{(N-j)\mu + i\gamma} \right)^{-1}, \quad j = \overline{0, N-g}. \quad (3.35)$$

Далее стационарное распределение укрупненной модели $\pi(< j >)$, $< j > \in \tilde{S}$ определяется аналогично (3.17), (3.18), и при этом важно учесть, что входящие в (3.17), (3.18) параметры $\rho^j(m)$ и $\Lambda(i)$ определяются с учетом (3.34), (3.35).

Резюмируя и опуская известные промежуточные математические выкладки, можно предложить следующие приближенные формулы для расчета характеристик модели с ограниченным интервалом деградации h -вызовов:

$$P_d^{(\gamma)} \approx \sum_{i=0}^{N-g} \rho^i(N+B-i)\pi(i) + \frac{1}{\lambda_h} \sum_{i=1}^B \sum_{j=0}^{N-g} i\gamma \rho^j(N+i-j)\pi(j); \quad (3.36)$$

$$P_b^{(\gamma)} \approx \sum_{j=0}^{N-g} \sum_{i=N-g-j}^{N+B-j} \rho^j(i)\pi(< j >); \quad (3.37)$$

$$\tilde{N}_{av}^{(\gamma)} \approx \sum_{j=1}^{N-g} j \sum_{i=0}^j \rho^i(j-i)\pi(< i >) + \sum_{j=N-g+1}^{N-1} j \sum_{i=0}^{N-g} \rho^i(j-i)\pi(< i >) + N \sum_{j=0}^{N-g} \pi(< j >) \sum_{i=N-j}^{N+B-j} \rho^i(i); \quad (3.38)$$

$$K_o^{(\gamma)} \approx \sum_{j=1}^{N-g} j\pi(< j >); \quad (3.39)$$

$$K_h^{(\gamma)} \approx \sum_{j=1}^{N+B-g} j \sum_{i=0}^g \rho^i(j)\pi(< i >) + \sum_{j=N+B-g+1}^{N+B} j \sum_{i=0}^{N+B-j} \rho^i(j)\pi(< i >). \quad (3.40)$$

Применение разработанного численного метода для исследования моделей с бесконечной очередью нетерпеливых h -вызовов представляется читателю.

3.2. Расчет характеристик модели с очередями новых вызовов

С целью компенсации шансов o -вызовов необходимо организовать их очереди (конечные или бесконечные), сохраняя при этом высокую приоритетность h -вызовов при поступлении в систему, которая обеспечивается за счет резервирования для них определенного числа радиоканалов. В связи с этим здесь исследуется модель соты, которая содержит конечный и/или бесконечный буфер лишь для ожидания в очереди o -вызовов.

Обслуживание разнотипных вызовов осуществляется по схеме резервирования каналов. Эта схема в отличие от схемы с индивидуальными каналами для h -вызовов (см. параграф 2.2 этой части книги) не является изолированным резервированием, т.е. поступивший h -вызов принимается при наличии хотя бы одного свободного канала из общего числа $m+n$ каналов; в противном случае, h -вызов теряется. Поступивший o -вызов принимается для обслуживания лишь тогда, когда число свободных каналов больше n , где n — число резервных каналов; в противном случае, o -вызов присоединяется к очереди, при этом o -вызовы выбираются из очереди для обслуживания лишь тогда, когда число свободных каналов становится больше n . При этом очередь o -вызовов обслуживается согласно дисциплине FIFO.

Альтернативное описание этой схемы занятия каналов разнотипными вызовами состоит в следующем. Все $m+n$ каналов разделяются на две группы: первая группа содержит m каналов, а вторая группа — n каналов. Для обслуживания поступившего h -вызова поиск свободного канала сначала осуществляется в первой группе; если все m каналов этой группы заняты, то поиск свободного канала осуществляется во второй группе. Если все каналы обеих групп заняты, то h -вызов теряется. Новые вызовы могут обслуживаться лишь в первой группе этой группы заняты, если в момент поступления o -вызова все каналы этой группы заняты, этот вызов становится в очередь. При освобождении канала первой группы (т.е. после завершения обслуживания одного o - или h -вызова в данной группе) один h -вызов, обслуживаемый в этот момент во второй группе каналов, перекладывается в первую группу независимо от длины очереди o -вызовов (т.е. происходит переназначение h -вызова из второй группы на первую).

Основными характеристиками системы являются вероятность потерь при h -вызовах (P_h), средняя длина очереди o -вызовов (L_o^0), а также их среднее время ожидания в очереди (W_o^0).

Состояние данной системы в произвольный момент времени также описывается двумерным вектором $\mathbf{k} = (k_1, k_2)$, где k_1 указывает число o -вызовов в системе, $k_1 = 0, 1, 2, \dots$ и k_2 означает общее число занятых каналов, $k_2 = 0, 1, \dots, m+n$. Отметим, что фазовое пространство состояний S этой системы не содержит векторов $\mathbf{k} = (k_1, k_2)$, где $k_1 > 0, k_2 < m$.

Исходя из механизма функционирования исследуемой системы, полагаем, что элементы производящей матрицы $q(\mathbf{k}, \mathbf{k}')$, $\mathbf{k}, \mathbf{k}' \in S$, соответствующей двумерной цепи Маркова определяются из следующих соотношений (см. рис.3.6):

$$q(\mathbf{k}, \mathbf{k}') = \begin{cases} \lambda_0 + \lambda_h, & \text{если } k_1 = 0, \\ \lambda_h, & \text{если } k_2 \leq m-1, \\ \lambda_0, & \text{если } k_2 \geq m, \\ k_2 \mu, & \text{если } k_2 \geq m, \\ m\mu, & \text{если } k_2 \neq m, \\ 0 & \text{если } k_2 = m, \end{cases} \quad \text{в остальных случаях.} \quad (3.41)$$

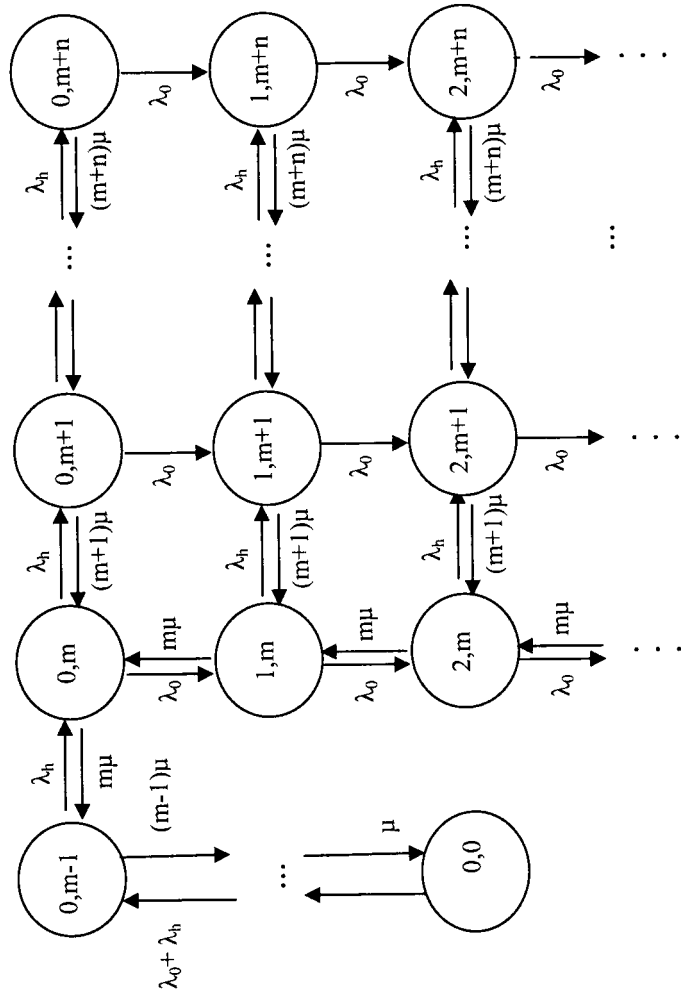


Рис.3.6. Граф переходов между состояниями модели с бесконечной очередью o -вызовов

Искомые характеристики системы выражаются с помощью стационарного распределения модели:

находим, что стационарные распределения расщепленных моделей определяются так:
для $i = 0$:

$$\rho^0(j) = \begin{cases} \frac{V^j}{j!} \rho_0 & \text{если } j = \overline{1, m}, \\ \left(\frac{V}{V_h}\right)^m \frac{V_h^j}{j!} \rho_0 & \text{если } j = \overline{m+1, m+n}, \end{cases} \quad (3.49)$$

для $i > 0$:

$$\rho^i(j) = \frac{m! V_h^j}{V_h^m j!} \rho_1, \quad j = \overline{m+1, m+n}, \quad (3.50)$$

где

$$\rho_0 = \left(\sum_{i=0}^m \frac{V^i}{i!} + \left(\frac{V}{V_h}\right)^m \sum_{i=m+1}^{m+n} \frac{V_h^i}{i!} \right)^{-1}, \quad \rho_1 = \left(\frac{m!}{V_h^m} \sum_{i=m}^{m+n} \frac{V_h^i}{i!} \right)^{-1}, \quad V := V_0 + V_h.$$

Элементы производящей матрицы укрупненной модели, обозначаемые $q(<i>, <i'>)$, $<i>, <i'> \in \tilde{S}$, определяются с помощью (3.41), (3.49), (3.50):

$$q(<i>, <i'>) = \begin{cases} \lambda_0 \sum_{j=m}^{m+n} \rho^0(j), & \text{если } i=0, \quad i'=i+1, \\ \lambda_0, & \text{если } i>0, \quad i'=i+1, \\ m\mu\rho_1, & \text{если } i>0, \quad i'=i-1, \\ 0 & \text{в остальных случаях.} \end{cases} \quad (3.51)$$

Из (3.51) находится следующее условие эргодичности укрупненной модели:

$$a := \frac{V_0}{m\rho^1(m)} < 1$$

или в явном виде

$$\frac{V_0}{m} \cdot \frac{m!}{V_h^m} \left(\sum_{i=m}^{m+n} \frac{V_h^i}{i!} \right) < 1. \quad (3.52)$$

При выполнении условия (3.52) стационарное распределение укрупненной модели ($\pi(<i>) : <i> \in \tilde{S}$) определяется следующим образом:
 $\pi(<i>) = a^i b \pi(<0>), \quad i = 1, 2, \dots$ (3.53)

где

$$P_h = \sum_{k=0}^{\infty} P(k_1, m+n), \quad (3.42)$$

$$L_q^0 = \sum_{k_1=1}^{\infty} \sum_{k_2=m}^{m+n} k_2 P(k_1, k_2), \quad (3.43)$$

$$W_q^0 = L_q^0 / \lambda_0. \quad (3.44)$$

С целью определения приближенных значений характеристик (3.42)–(3.44) рассматривается следующее разбиение ФПС исходной модели:

$$S = \bigcup_{i=0}^{\infty} S_i, \quad S_i \cap S_j = \emptyset, \quad i \neq j, \quad (3.45)$$

где $S_i := \{k \in S : k_1 = i\}$.

Классы микросостояний S_i представляются в виде изолированных укрупненных состояний $<i>$, и на исходном ФПС строится следующая функция укрупнения:

$$U(k) = <i>, \quad \text{если } k \in S_i, \quad i = 0, 1, 2, \dots \quad (3.46)$$

Функция укрупнения (3.46) определяет укрупненную модель, которая также является цепью Маркова с бесконечным пространством состояний $\tilde{S} := \{<i> : i = 0, 1, 2, \dots\}$.

Элементы производящей матрицы расщепленных моделей с ФПС S_i , обозначаемые $q_i(k, k')$, $k, k' \in S_i$, определяются так:
для модели с ФПС S_0 :

$$q_0(k, k') = \begin{cases} \lambda_0 + \lambda_h, & \text{если } k_2 \leq m-1, \quad k'_2 = k_2 + 1, \\ \lambda_h, & \text{если } m \leq k_2 \leq m+n-1, \quad k'_2 = k_2 + 1, \\ k_2 \mu, & \text{если } k_2 = k_2 - 1, \\ 0 & \text{в остальных случаях,} \end{cases} \quad (3.47)$$

для модели с ФПС $S_h, i > 0$:

$$q_0(k, k') = \begin{cases} \lambda_h, & \text{если } k'_2 = k_2 + 1, \\ k_2 \mu, & \text{если } k_2 = k_2 - 1, \\ 0 & \text{в остальных случаях.} \end{cases} \quad (3.48)$$

Стационарную вероятность состояния $(i, j) \in S$ внутри расщепленной модели с ФПС S_i обозначим $\rho^i(j)$. Тогда, с помощью (3.47), (3.48)

$$b := \sum_{i=m}^{m+n} \rho^0(i), \quad \pi(<0>) = \frac{1-a}{1-a+ab}. \quad (3.54)$$

Замечание 3.3. Важно заметить, что условие эргодичности (3.52) полностью совпадает с аналогичным условием, найденным в [6]. Однако, в [6] это условие было найдено с помощью сложных вероятностных рассуждений, а здесь оно установлено исходя из эргодичности некоторого одномерного процесса размножения и гибели.

После выполнения необходимых математических преобразований найдя следующие приближенные формулы для вычисления характеристик (3.42)–(3.44) исследуемой модели:

$$P_h \approx \frac{1}{1-a+ab} ((1-a)E_B(V_h, m+n) + abE_B^T(V_h, m)), \quad (3.55)$$

$$L_q^0 \approx \frac{ab}{(1-a+b)(1-a)}, \quad (3.56)$$

$$W_c^0 \approx L_q^0 / \lambda_0, \quad (3.57)$$

где $E_B(V_h, m+n)$ – В-формула Эрланга для системы $M | M | m+n$ с нагрузкой V_h эрл., $E_B^T(V_h, m)$ – усеченная В-формула Эрланга:

$$E_B^T(V_h, m) := \frac{V_h^{m+n}}{(m+n)!} \left(\sum_{i=m}^{m+n} \frac{V_h^i}{i!} \right)^{-1}.$$

Из формулы (3.55) видно, что P_h является выпуклой комбинацией двух функций $E_B(V_h, m+n)$ и $E_B^T(V_h, m)$. Следовательно, при любых нагрузках имеет место следующее соотношение:

$$E_B(V_h, m+n) \leq P_h \leq E_B^T(V_h, m). \quad (3.58)$$

В неравенстве (3.58) границы достигаются и становятся одинаковыми только в случае $n = 0$, т.е. при отсутствии резервных каналов для h -вызовов. Этот результат имеет вполне логичный смысл.

Предложенный метод позволяет определить характеристики данной системы и при наличии лишь ограниченного буфера для ожидания в очереди o -вызовов. Пусть максимально допустимая длина очереди o -вызовов равна $R < \infty$. Тогда, при любых значениях нагрузочных и структурных параметров в системе существует стационарный режим (т.е. не требуется выполнения условия эргодичности (3.52)).

Стационарные распределения расцепленных моделей определяются согласно (3.49), (3.50), но в данном случае число таких моделей конечно и равно $R+1$. Применяя описанный выше подход и опуская промежуточные

ные математические преобразования, находим, что для данной модели стационарное распределение укрупненной модели определяется так:

$$\pi(<i>) = a^i b \pi(<0>), \quad i = \overline{1, R}, \quad (3.59)$$

где

$$\pi(<0>) = \left(1 + ab \frac{1-a^R}{1-a} \right)^{-1}. \quad (3.60)$$

Следовательно, приближенные значения характеристик (3.42)–(3.44) для модели с ограниченной очередью вычисляются следующим образом:

$$P_h(R) \approx \pi(<0>) E_B(V_h, m+n) + (1-\pi(<0>)) E_B^T(V_h, m), \quad (3.61)$$

$$L_q^0(R) \approx ab \frac{1-a^R(R+1+Ra)}{(1-a)^2} \pi(<0>), \quad (3.62)$$

$$W_q^0(R) \approx \frac{L_q^0(R)}{\lambda_0(1-P_0(R))}, \quad (3.63)$$

где $P_0(R)$ означает вероятность потери o -вызовов, которая для данной модели вычисляется так:

$$P_0(R) \approx \pi(R) \quad \text{или} \quad P_0(R) \approx a^R b \pi(<0>). \quad (3.64)$$

Применение разработанного численного метода для исследования моделей с очередями нетерпеливых o -вызовов, а также моделей с более сложными схемами приоритетного назначения каналов для h -вызовов (см., например, [5, 7, 8]) предоставляется читателю.

3.3. Численные эксперименты

Разработанные выше алгоритмы расчета характеристик изучаемых моделей реализованы на языке Object Pascal в интегрированной среде разработки Delphi 6. Целью выполнения численных экспериментов было изучение поведения характеристик системы в зависимости от изменения ее параметров. Результаты этих экспериментов для мс., если с ограниченной очередью терпеливых h -вызовов показаны на рис. 3.7–3.12. Их анализ позволяет сделать ряд выводов.

1. Как и следовало ожидать, величина P_h является монотонно убывающей, а величина P_0 – монотонно возрастающей относительно параметра g при всех фиксированных значениях остальных структурных и нагрузочных параметров системы (здесь и в дальнейшем при анализе зависимости каждой характеристики от конкретного параметра все остальные параметры считаются фиксированными). Из рис. 3.7 видно, что в точке $g = 1$ абсолютная разница между P_0 и P_h достигает минимального значения, т.е. при $g = 1$ в системе достигается справедливое (абсолютное)

обслуживание разнотипных вызовов в смысле разницы между значениями вероятностей их потери и блокировки. Отметим, что это обстоятельство имеет место при любых значениях параметров системы. Увеличение размера буфера существенным образом влияет на значение вероятностей потери h -вызовов; но вместе с тем вероятность блокировки o -вызовов почти не зависит от размера буфера (см. рис.3.8). Так, например, $lgP_o = -0.802799$ при $B = 1$, а при $B = 25$ соответствующее значение $lgP_o = -0.802716$ для исходных данных, указанных на рис. 3.8. Аналогичные результаты имеют место в очень широком диапазоне изменения значений исходных данных модели.

2. При фиксированном значении буфера среднее число вызовов обоих типов в системе монотонно убывает с ростом числа резервных каналов для h -вызовов (см. рис.3.9), при этом среднее число o -вызовов убывает с большей скоростью по сравнению с аналогичной характеристикой h -вызовов. Убывание K_o было вполне ожидаемым явлением, но убывание K_h (хоть и с малой скоростью) несколько неожиданно. Однако более детальное изучение составляющих этой величины показало, что с ростом числа резервных каналов h -вызовы имеют высокие шансы на доступ к каналам обслуживания (т.е. растет число h -вызовов в каналах), и, таким образом, уменьшается число таких вызовов в буфере, т.е. одна составляющая для выбранных исходных данных выше, чем скорость увеличения, то величина K_h также является убывающей для этих исходных данных системы.

Среднее число h -вызовов в системе растет с увеличением размера буфера, а соответствующая характеристика для o -вызовов является убывающей функцией аргумента B (см. рис.3.10). При этом скорости изменений этих величин для выбранных исходных данных являются достаточно низкими. Отметим, что такое поведение K_o и K_h полностью согласуется с поведением P_o и P_h относительно изменения аргумента B (см. рис.3.8).

3. Среднее время ожидания в очереди h -вызовов в системе уменьшается с ростом числа резервных каналов (рис. 3.11), но оно увеличивается с ростом размера буфера (рис. 3.12). Такое поведение W_h было вполне ожидаемым, так как с увеличением g растут шансы h -вызовов на доступ к каналам (и тем самым уменьшается среднее время ожидания в очереди), а с увеличением B растет длина очереди h -вызовов (и тем самым увеличивается среднее время ожидания в очереди).

Другая цель выполнения численных экспериментов заключалась в оценке точности предложенных формул. При этом результаты сравниваются с результатами работы [4], где получены точные формулы для расчета вероятностей блокировки вызовов каждого типа в модели с идентичными (по длительности занятия каналов) разнотипными вызовами и ограниченной очередью h -вызовов.

Точные значения P_o и P_h обозначим P'_o и P'_h , соответственно (они

показаны в п.1.2 данной части книги). Сравнительный анализ был выполнен в широком диапазоне изменения исходных данных (т.е. структурных и нагрузочных параметров) изучаемой модели.

Важно отметить высокую точность предложенных формул, так как при вычислении вероятности потери h -вызовов их точные и приближенные значения полностью совпадают. Некоторые совсем несущественные расхождения наблюдаются при вычислении вероятности блокировки o -вызовов, и при этом максимальное значение погрешности во всех рассматриваемых экспериментах составляет менее 2%, а в подавляющем большинстве случаев она равна нулю. Так, например, при $N = 50$, $B = 10$, $\lambda_1 = 2$, $\lambda_2 = 1$, $\mu = 10$ максимальная погрешность имеет место при $g = 49$ и составляет 1.6%, т.е. при этих исходных данных точное значение равно $P'_o = 0.233$, а ее приближенное значение (вычисленное по предложенному алгоритму) равно $P_o = 0.249$.

Результаты численных экспериментов для модели с нетерпеливыми h -вызовами показаны на рис. 3.13 – 3.18. Как и следовало ожидать, величина P_h является монотонно убывающей, а величина P_o – монотонно возрастающей относительно параметра g при всех фиксированных значениях остальных структурных и нагрузочных параметров системы (здесь и в дальнейшем при анализе зависимости каждой характеристики от конкретного параметра все остальные параметры считаются фиксированными). На приведенном рис. 3.13 соответствующие графики «перескакают» в некоторой точке (при $g = 26$), т.е. в этой точке абсолютная разница между P_o и P_h имеет минимальное значение, и, таким образом, для указанных исходных данных модели, при $g = 26$, в системе достигается справедливое (абсолютное) обслуживание разнотипных вызовов в смысле разницы между значениями вероятностей их потери и блокировки. Очевидно, что в общем случае нельзя утверждать, что эти графики обязательно «пересекаются» при всех исходных данных модели. Вместе с тем, характер зависимости P_o и P_h от параметра g сохраняется для данной модели при всех исходных данных.

Поведение P_o относительно параметра γ является совершенно понятным (она убывает с ростом γ), так как увеличивается интенсивности ухода нетерпеливых h -вызовов из очереди благоприятно действует на процесс доступа o -вызовов в каналы (рис. 3.14). Вместе с тем, характер зависимости P_h от параметра γ на первый взгляд кажется аномальным (она также убывает с ростом γ). Однако, детальное изучение составляющих этой величины показало, что с ростом γ h -вызовы часто теряются из очереди (т.е. растет вероятность потери h -вызовов из-за их нетерпеливости), и, таким образом, освобождается вероятность потери вновь поступающих h -вызовов (т.е. уменьшается вероятность потери h -вызовов из-за переполненности буфера). То есть одна составляющая P_h растет, а другая – уменьшается, а поскольку скорость уменьшения для выбранных исходных данных выше, чем скорость увеличения, то величина P_h также является убывающей (рис. 3.14).

Значение величины \tilde{N}_{av} с ростом g уменьшается (см. рис. 3.15), т.е. с ростом числа резервных каналов для h -вызовов общая утилизация каналов падает в результате того, что они не эффективно используются o -вызовами; увеличение значений \tilde{N} с ростом параметра γ объясняется тем, что при этом каналы эффективно используются o -вызовами (рис. 3.16). Следует отметить, что на значение величины \tilde{N}_{av} очевидно сильнее влияет параметр g , нежели параметр γ . Так, например, если для выбранных исходных данных значения \tilde{N} при $g = 1$ и $g = 24$ равны 21.53 и 6.02 соответственно (т.е. общая утилизация каналов растет почти в 3.5 раза), то ее аналогичные значения при $\gamma = 1$ и $\gamma = 25$ равны 19.51 и 21.52 (т.е. общая утилизация каналов возрастает всего на 10%).

Величина K_0 является монотонно убывающей относительно параметра g , но она монотонно возрастает относительно параметра γ . Обратная картина наблюдается при исследовании величины K_h (см. рис. 3.17 и 3.18). Эти результаты также были ожидаемыми, так как увеличение числа резервных каналов для h -вызовов увеличивает шансы доступа вызовов данного типа в систему, и, тем самым, уменьшает шансы o -вызовов на доступ к каналам системы; а уменьшение длины интервала деградации для h -вызовов приводит к уменьшению среднего числа вызовов данного типа вследствие их частых потерь из очереди, что, таким образом, увеличивает среднее число o -вызовов в системе.

Отметим, что параметры γ и g в неодинаковой степени влияют на скорости изменения величин K_0 и K_h . Так, например, величина K_h очень медленно растет с увеличением g , в то время как она с большей скоростью убывает при увеличении «степени нетерпеливости» h -вызовов (т.е. с уменьшением длины интервала деградации).

На рис. 3.19 – 3.25 показаны зависимости характеристик системы с бесконечными и конечными очередями o -вызовов от числа резервных каналов для h -вызовов. В численных экспериментах для модели с бесконечной очередью o -вызовов исходные данные выбирались так: $N = 40$, $\lambda_h = 15$, $\mu = 1$. Результаты численных экспериментов полностью подтвердили все теоретические ожидания. Так, функция вероятности потери h -вызовов является убывающей (рис.3.19), а функции среднее число o -вызовов в очереди (рис.3.20) и среднее время их ожидания (рис.3.21) являются возрастающими относительно числа резервных каналов. При этом все эти функции являются возрастающими относительно интенсивности трафика o -вызовов (рис.3.19 – 3.21). Отметим, что для указанных исходных данных свойство эргодичности модели теряется при $n \geq 24$, потому на этих графиках значения n указаны в интервале [1, 23].

Некоторые результаты численных экспериментов для модели с конечной очередью o -вызовов показаны на рис.3.22 – 3.25. В них исходные данные выбирались так: $N = 40$, $R = 3$, $\lambda_h = 15$, $\mu = 1$. Для этой модели функция вероятности потери h -вызовов также является убывающей

относительно числа резервных каналов (рис.3.22). Сравнение результатов соответствующих экспериментов показали, что введение ограниченного буфера для o -вызовов почти не влияет на значень вероятности потери h -вызовов (рис.3.19 и 3.22). Этот факт объясняется тем, что в выбранных исходных данных интенсивность трафика o -вызовов намного меньше, чем интенсивность трафика h -вызовов. Функция вероятности потери o -вызовов возрастает относительно числа резервных каналов, при этом с ростом размера буфера, как и следовало ожидать, она уменьшается (рис.3.23). Также с увеличением числа резервных каналов увеличиваются средняя длина очереди o -вызовов (рис.3.24) и среднее время их ожидания (рис.3.25), при этом эти функции являются возрастающими относительно интенсивности трафика o -вызовов. Сравнение результатов численных экспериментов также показали, что введение ограниченного буфера для o -вызовов почти не влияет на величину средней длины их очереди, особенно, при малых интенсивностях этого трафика (рис.3.20 и 3.24). Те же самые комментарии относятся к поведению функции среднего времени ожидания o -вызовов в различных моделях (см.рис.3.21 и 3.25). Эти факты также объясняются достаточно низкой интенсивностью o -вызовов по сравнению с интенсивностью h -вызовов.

И в этом случае выполнение численных экспериментов помогло оценить точность полученных формул (3.55)–(3.57). Так, при выполнении вышеуказанного допущения относительно соотношений интенсивности трафиков o - и h -вызовов (т.е. $\lambda_h \gg \lambda_o$), наши результаты почти полностью совпадают с результатами работы [6] (ее результаты являются точными). Некоторые сравнения даны в табл. 3.1, где в третьем столбце указаны значения P_h , вычисленные с помощью алгоритма работы [6], а в четвертом столбце указаны соответствующие значения этой величины, рассчитанные с применением предложенного здесь алгоритма. Аналогичные результаты имеют место и для других характеристик исследуемых моделей.

Таблица 3.1

Сравнительный анализ результатов при $\mu = 2.0$, $\lambda_o = 1$, $\lambda_h = 10$

$m+n$	n	P_h , [6]	P_h , (3.55)
20	3	8.43E-07	2.98E-07
20	5	7.02E-07	2.97E-07
20	7	5.88E-07	2.96E-07
30	3	1.92E-13	2.79E-14
30	5	1.59E-13	2.75E-14
30	7	1.32E-13	2.72E-14
40	20	4.07E-14	2.71E-14
40	7	1.07E-21	9.42E-23
40	20	3.17E-22	8.54E-23

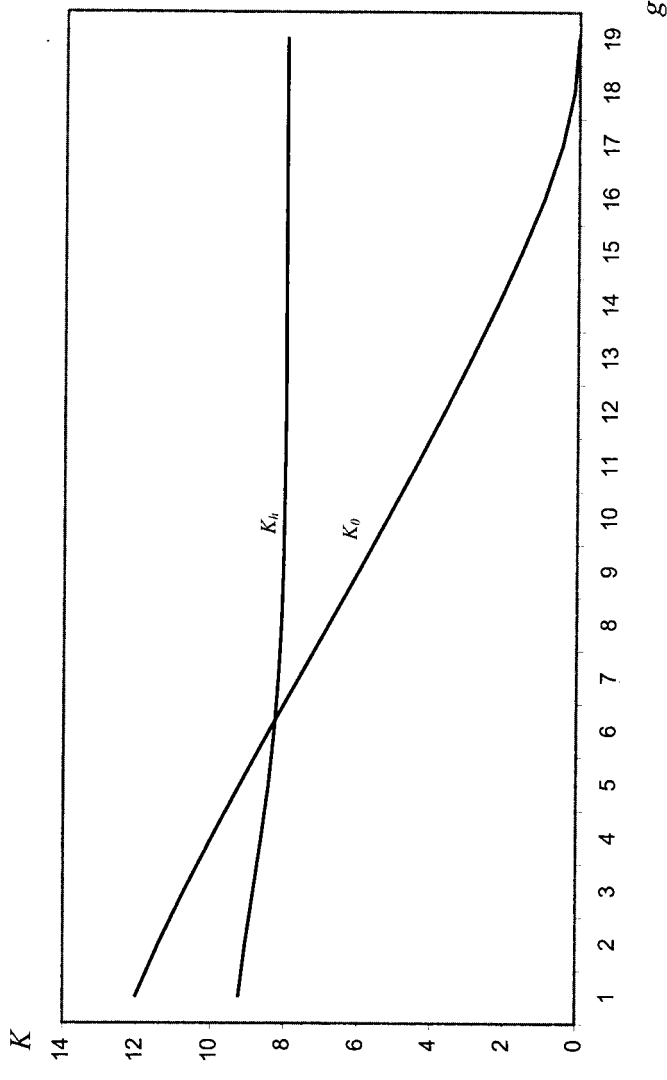


Рис. 3.9. Зависимость среднего числа разнотипных вызовов от g :
 $N = 20; B = 6; \mu = 1; \lambda_1 = 50; \lambda_2 = 8$

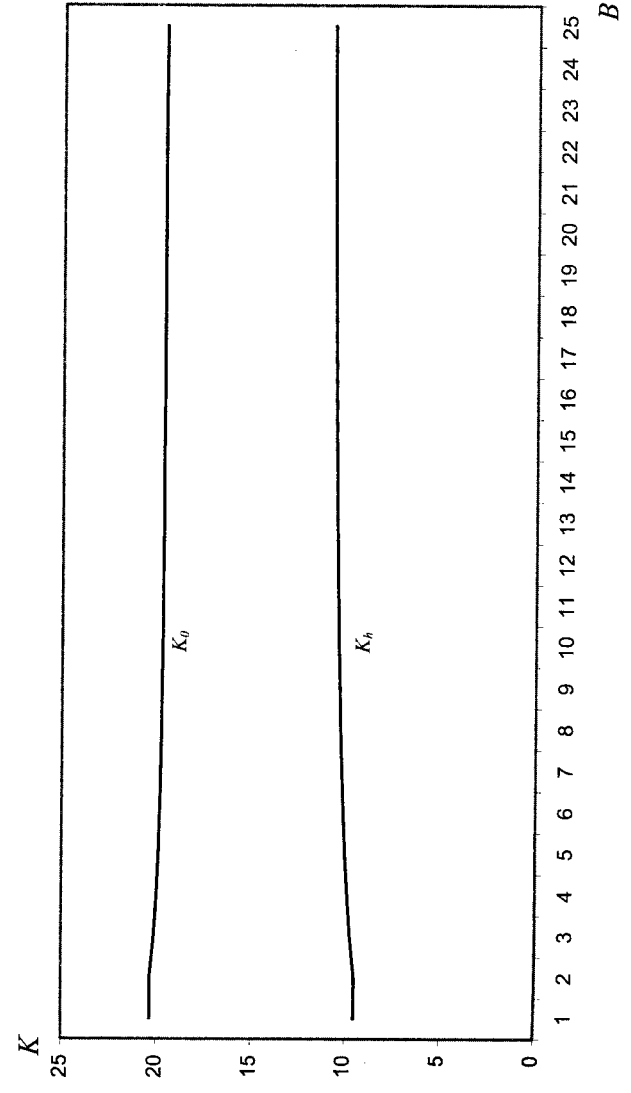


Рис. 3.10. Зависимость среднего числа разнотипных вызовов от B :
 $N = 35; g = 1; \mu = 2; \lambda_1 = 50; \lambda_2 = 20$

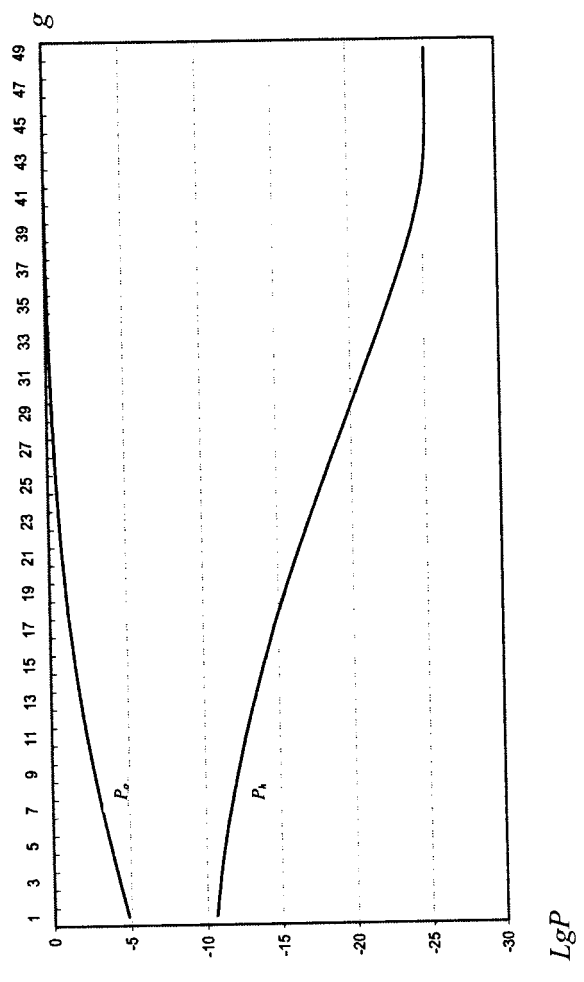


Рис. 3.7. Зависимость вероятностей потерь разнотипных вызовов от g :
 $N = 50; B = 25; \mu = 2; \lambda_1 = 20; \lambda_2 = 30$

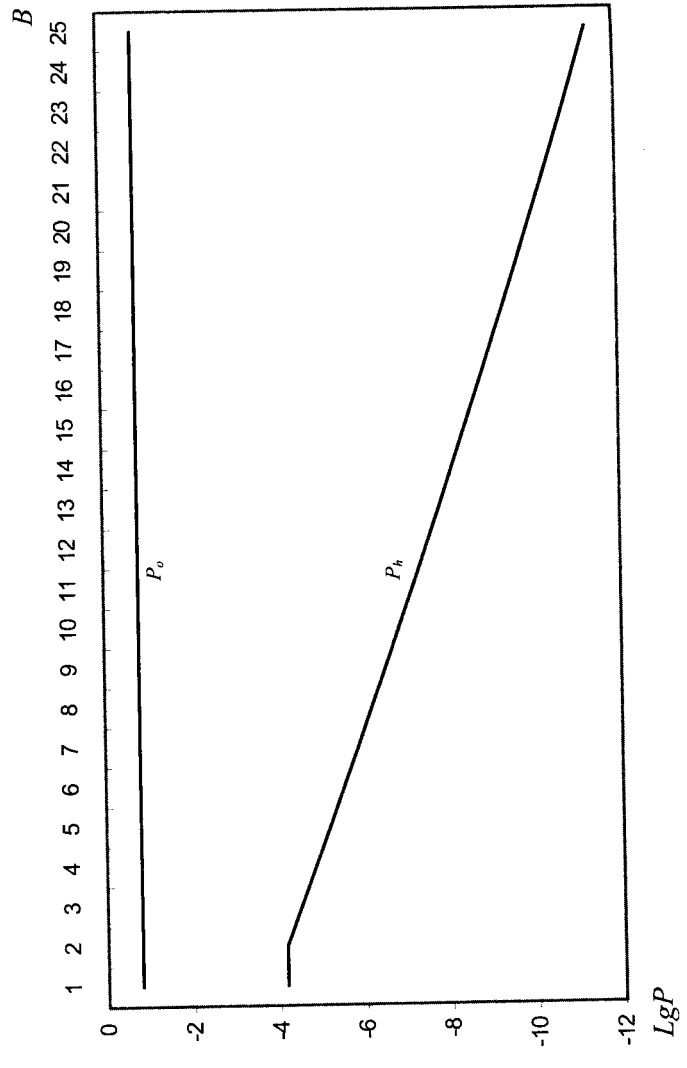


Рис. 3.8. Зависимость вероятностей потерь разнотипных вызовов от B :
 $N = 35; g = 10; \mu = 1; \lambda_1 = 15; \lambda_2 = 8$

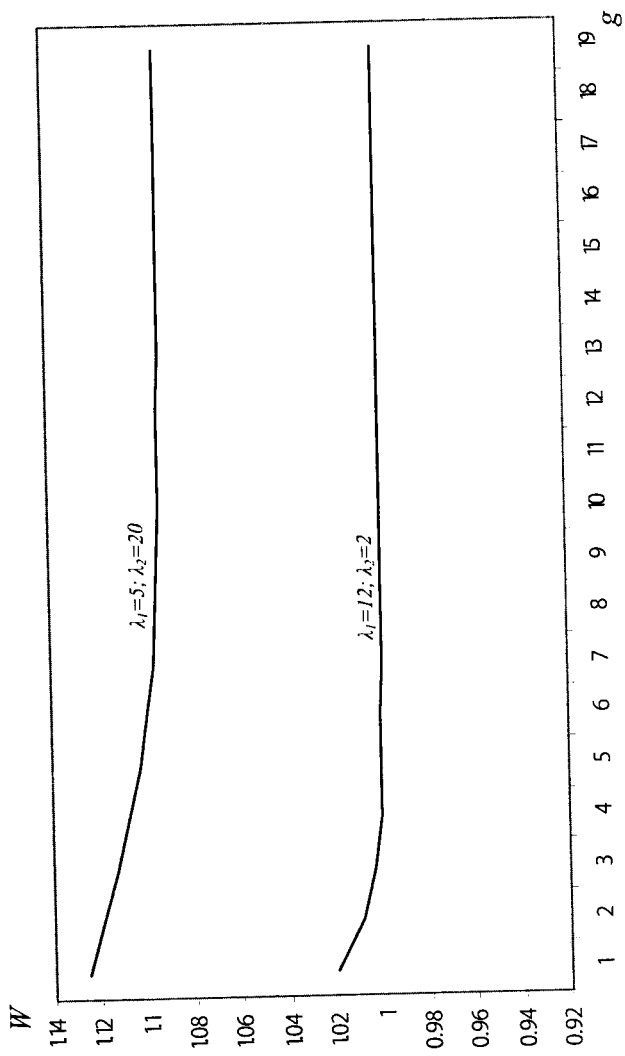


Рис. 3.11. Зависимость среднего времени ожидания h -вызовов от g :
 $N = 20$; $B = 6$; $\mu = 1$

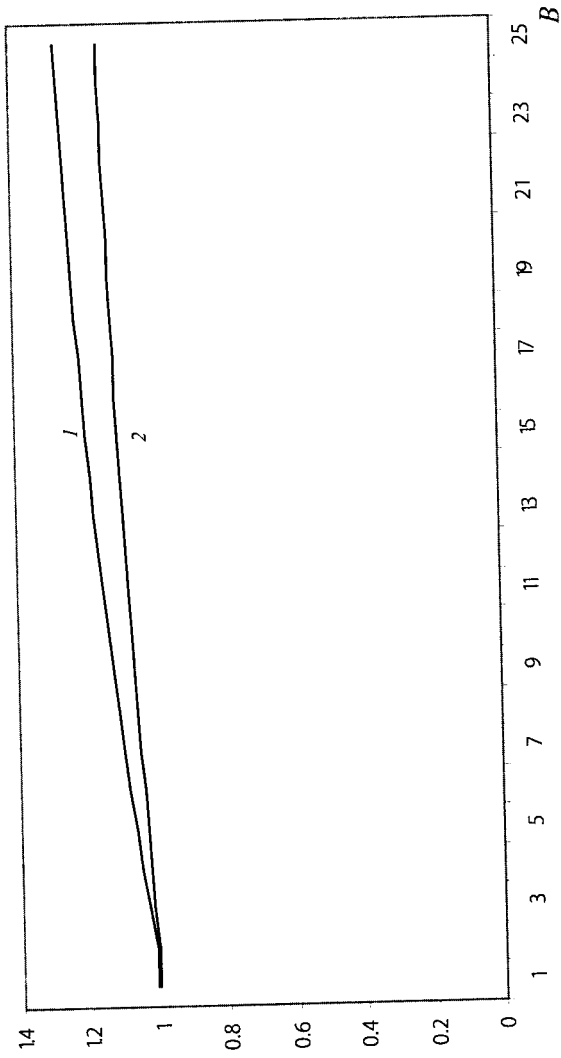


Рис. 3.12. Зависимость среднего времени ожидания h -вызовов от B :

1- $N = 30$; $g = 2$; $\mu = 1$; $\lambda_2 = 40$; $\lambda_1 = 10$;
 2- $N = 40$; $g = 1$; $\mu = 1$; $\lambda_1 = 30$; $\lambda_2 = 20$

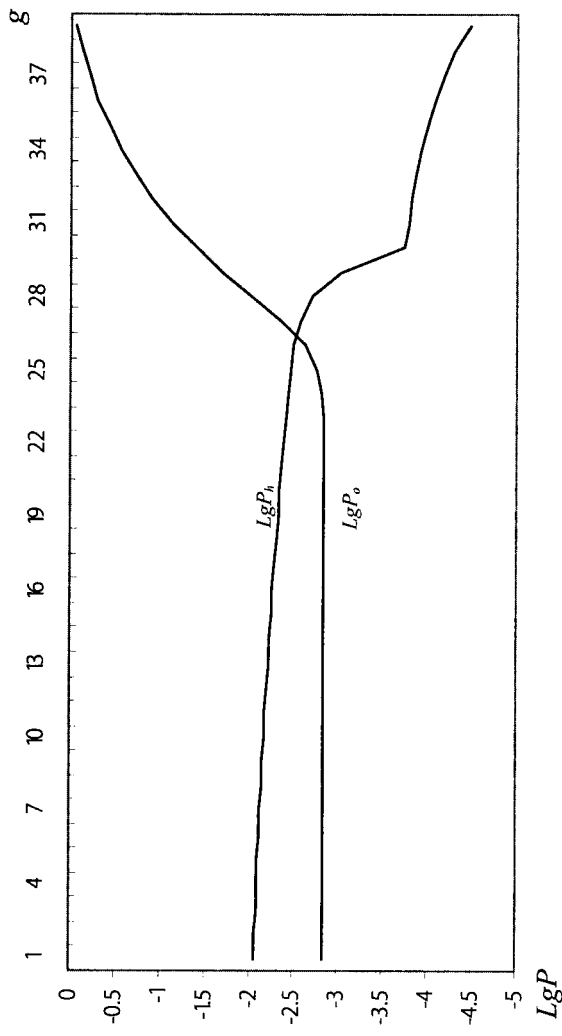


Рис. 3.13. Зависимость вероятностей потерь разнотипных вызовов от g :
 $\lambda_1 = 20$; $\lambda_2 = 8$; $N = 40$; $B = 10$; $\mu = 5$; $\gamma = 20$

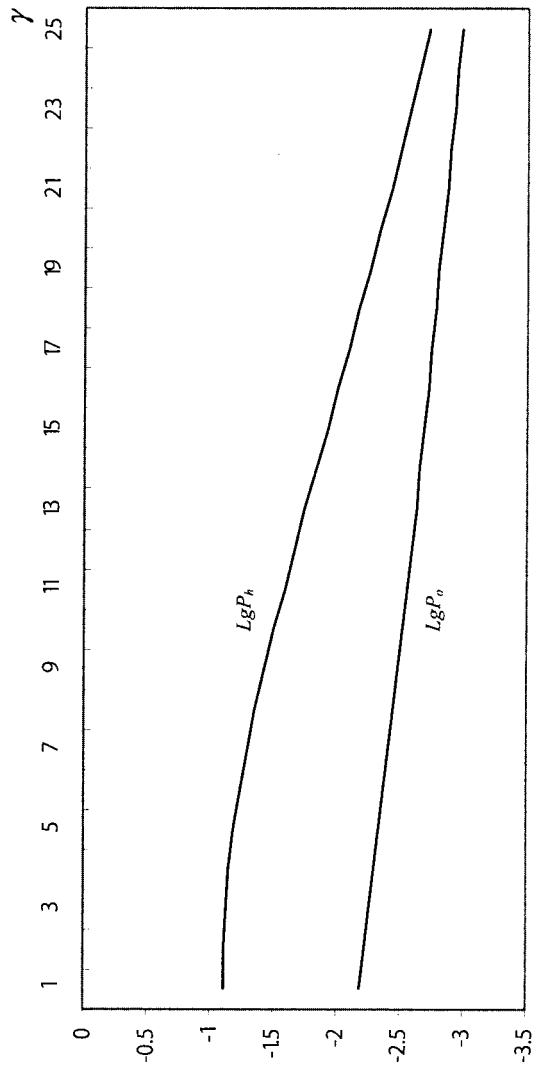


Рис. 3.14. Зависимость вероятностей потерь разнотипных вызовов от γ :
 $\lambda_1 = 20$; $\lambda_2 = 8$; $N = 40$; $B = 10$; $g = 20$; $\mu = 5$

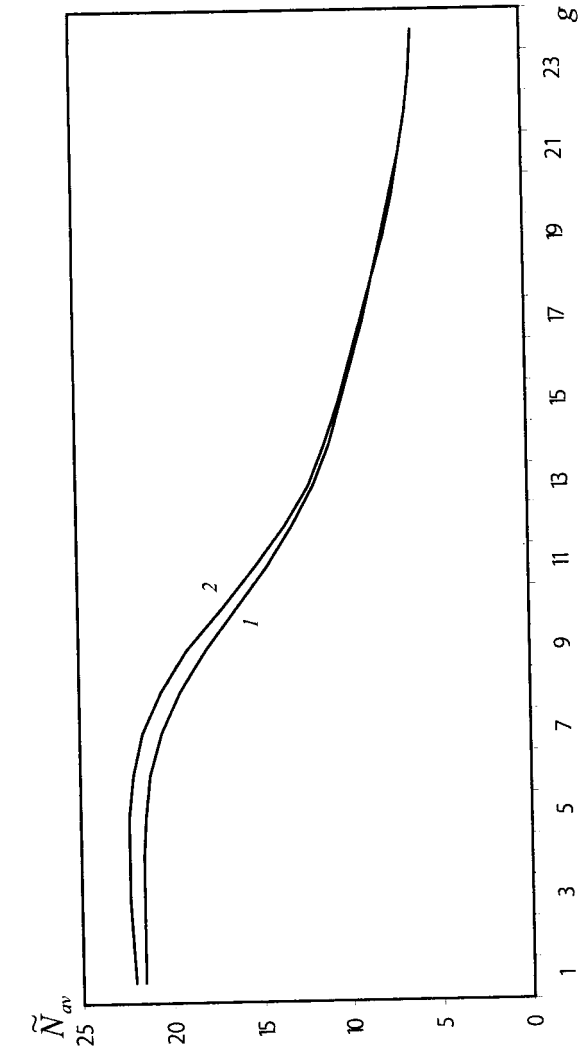


Рис. 3.15. Зависимость среднего числа занятых каналов

в системе от g :

1 - $\lambda_1 = 10$; $\lambda_2 = 6$; $N = 25$; $B = 6$; $\mu = 1$; $\gamma = 80$;

2 - $\lambda_1 = 11$; $\lambda_2 = 6$; $N = 25$; $B = 3$; $\mu = 1$; $\gamma = 50$

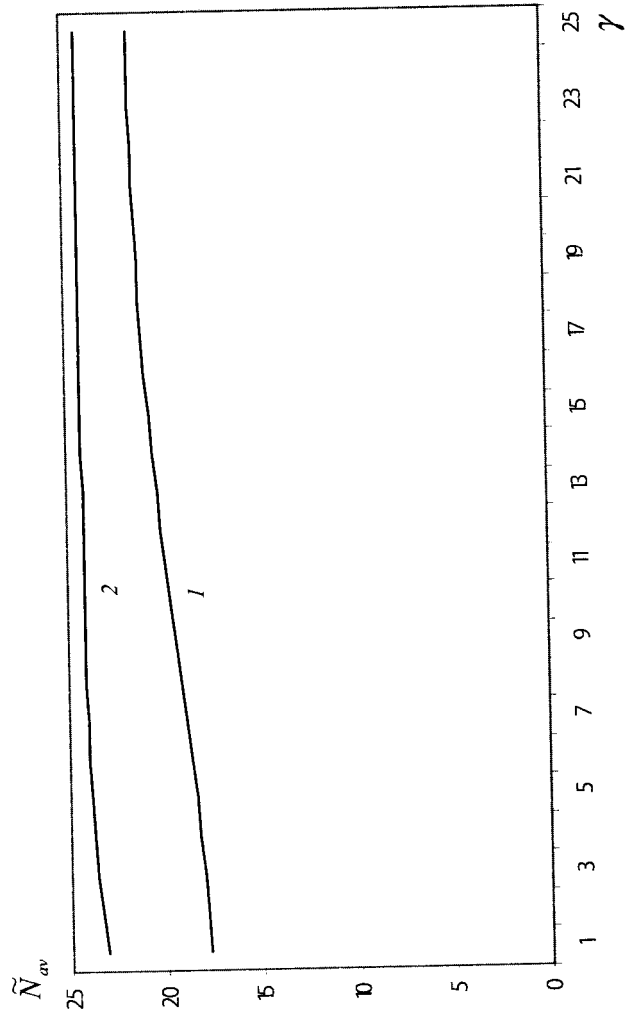


Рис. 3.16. Зависимость среднего числа занятых каналов

в системе от γ :

1 - $\lambda_1 = 10$; $\lambda_2 = 5$; $N = 25$; $B = 4$; $g = 3$; $\mu = 1$;

2 - $\lambda_1 = 15$; $\lambda_2 = 4$; $N = 25$; $B = 4$; $g = 5$; $\mu = 1$

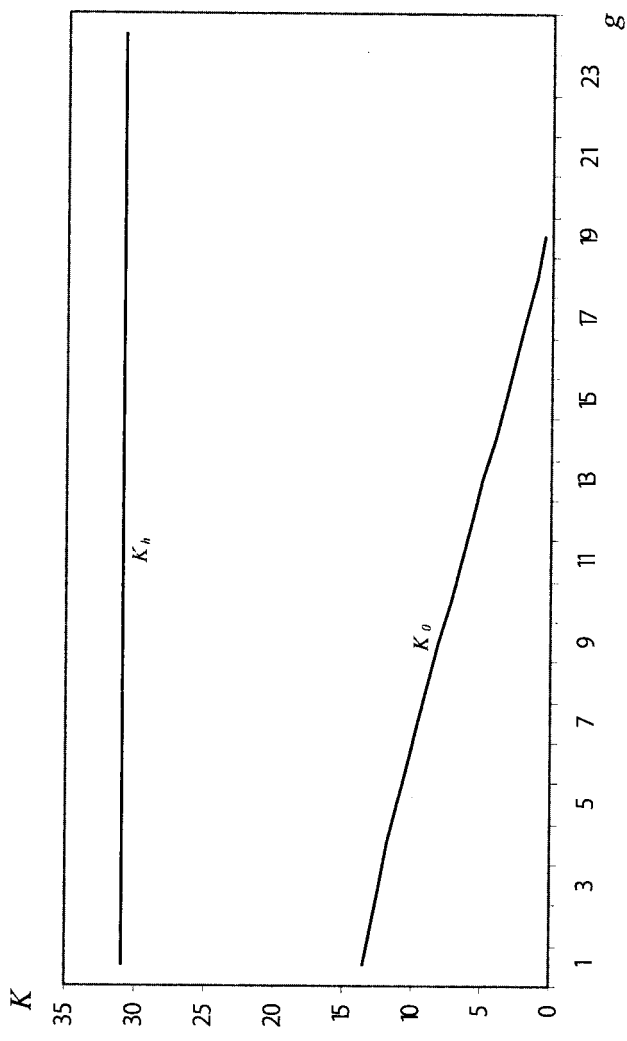


Рис. 3.17. Зависимость среднего числа разнотипных вызовов от g ;

K_0 - $\lambda_1 = 10$; $\lambda_2 = 3$; $N = 20$; $B = 6$; $\mu = 1$; $\gamma = 10$;

K_h - $\lambda_1 = 15$; $\lambda_2 = 12$; $N = 25$; $B = 6$; $\mu = 1$; $\gamma = 10$

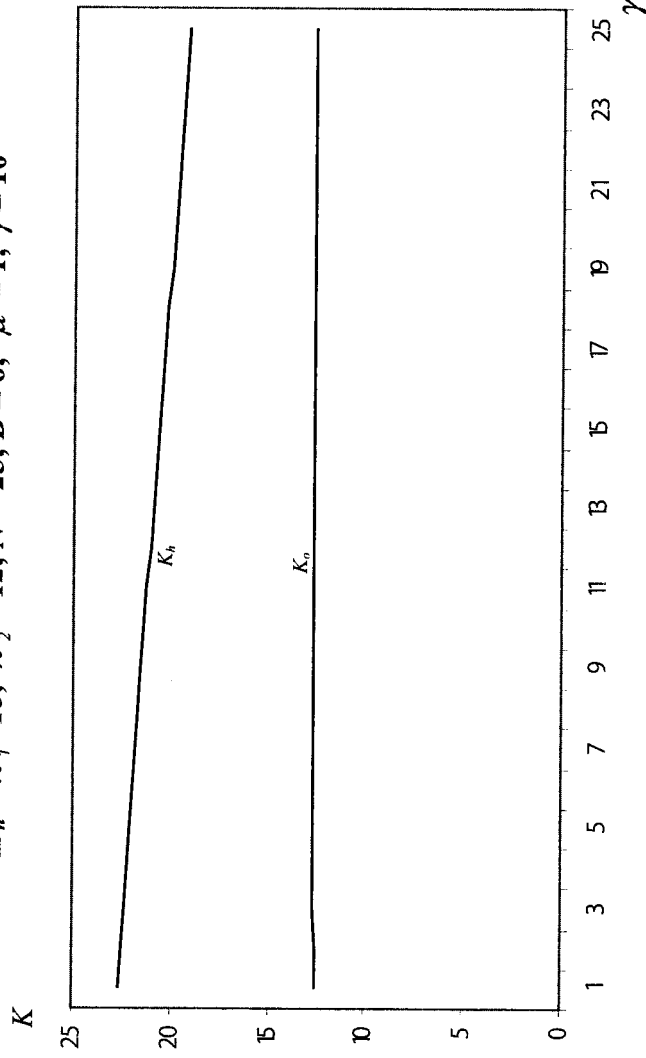


Рис. 3.18. Зависимость среднего числа разнотипных вызовов от γ ;

K_0 - $\lambda_1 = 20$; $\lambda_2 = 3$; $N = 20$; $B = 3$; $g = 3$; $\mu = 1$;

K_h - $\lambda_1 = 20$; $\lambda_2 = 15$; $N = 20$; $B = 3$; $g = 3$; $\mu = 1$

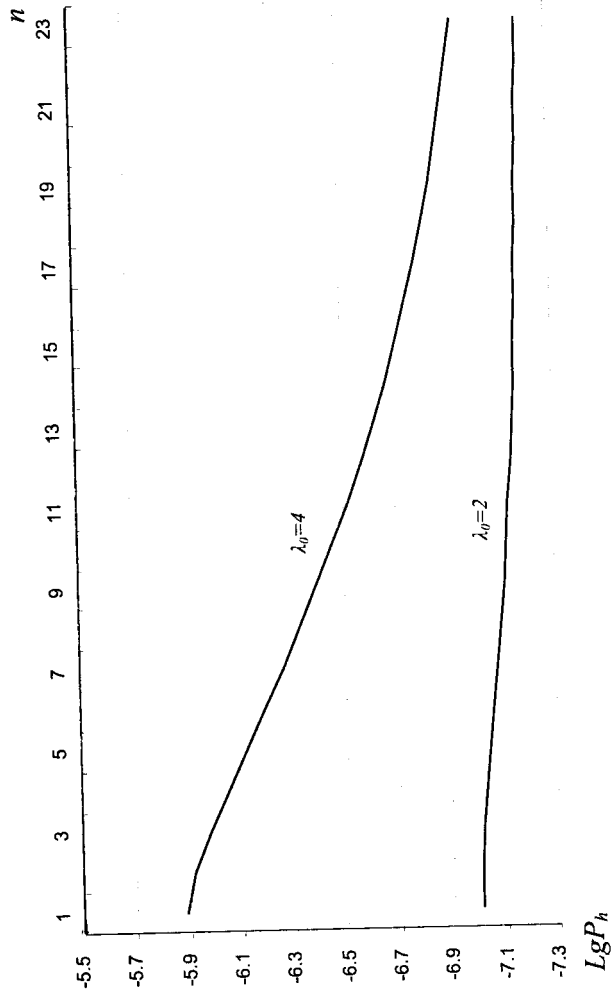


Рис.3.19. Зависимость вероятности потери h -вызовов от числа резервных каналов в модели с бесконечной очередью

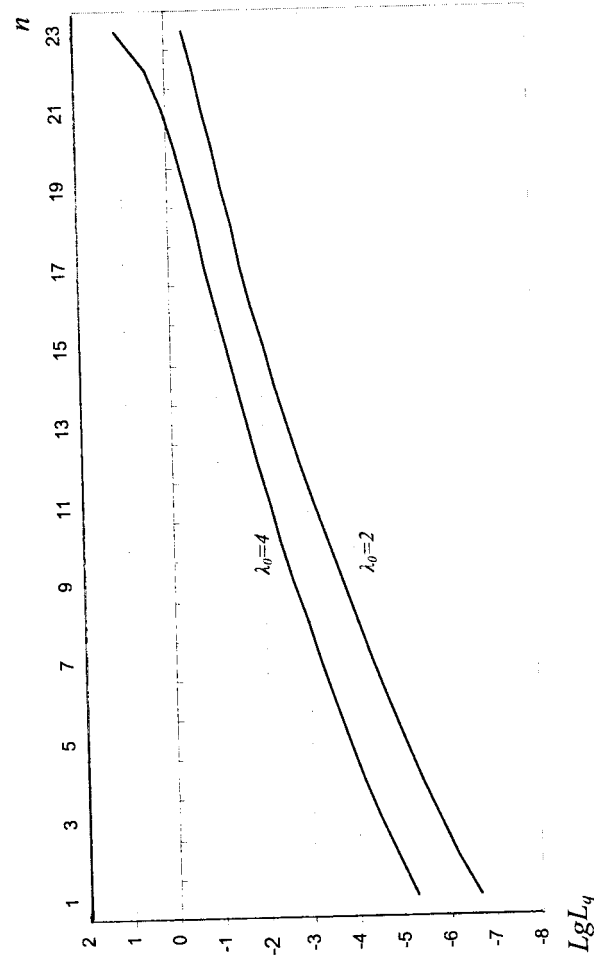


Рис.3.20. Зависимость средней длины очереди o -вызовов от числа резервных каналов в модели с бесконечной очередью

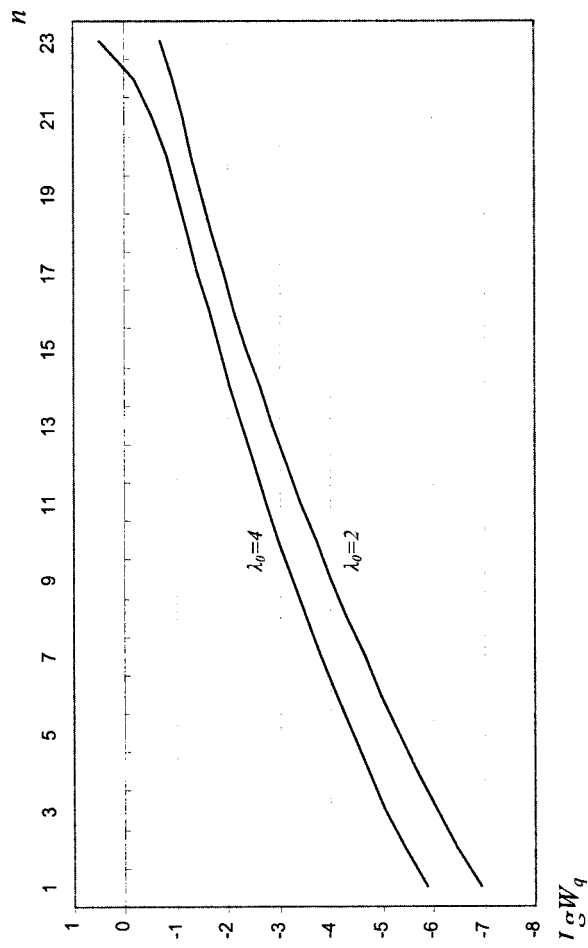


Рис. 3.21. Зависимость среднего времени ожидания o -вызовов от числа резервных каналов в модели с бесконечной очередью

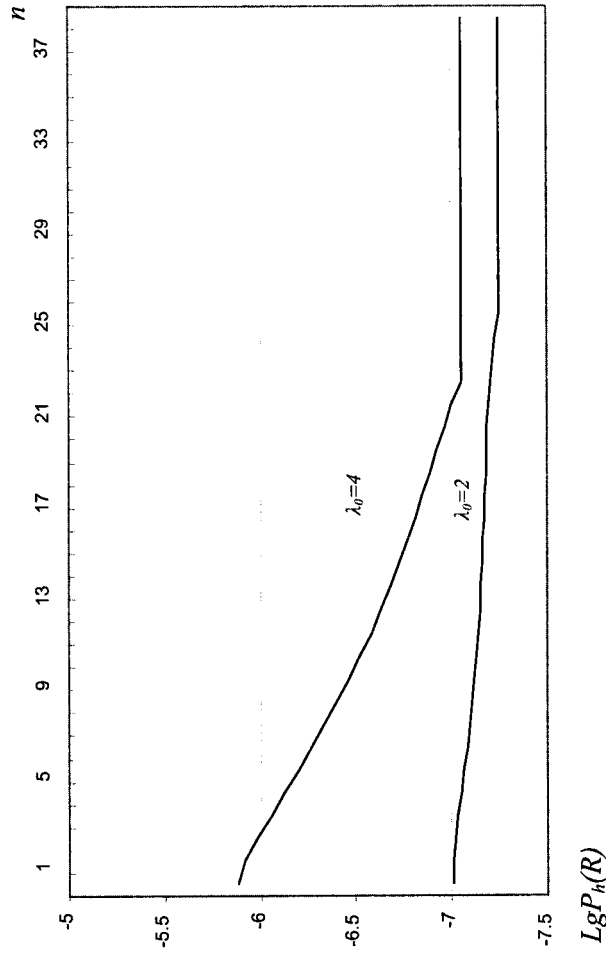


Рис. 3.22. Зависимость вероятности потери h -вызовов от числа резервных каналов в модели с конечной очередью

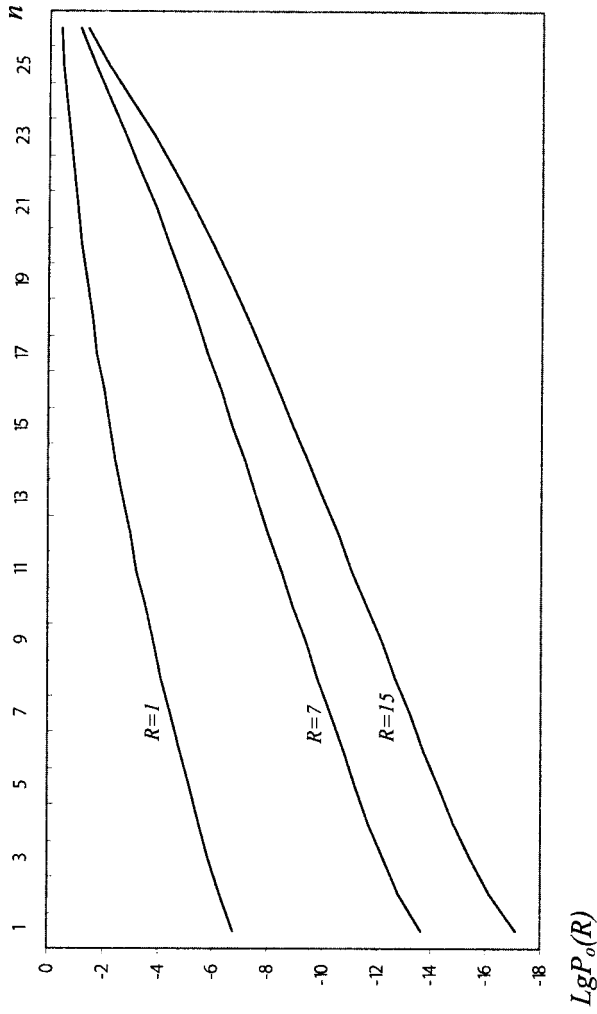


Рис. 3.23. Зависимость вероятности потери o -вызовов от числа резервных каналов в модели с конечной очередью при $\lambda_0=2$

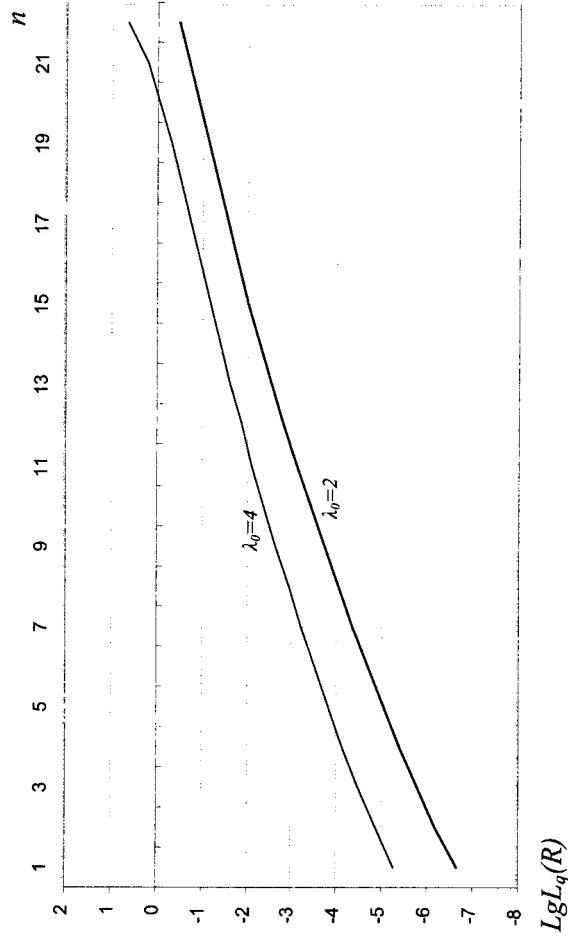


Рис. 3.24. Зависимость средней длины очереди o -вызовов от числа резервных каналов в модели с конечной очередью

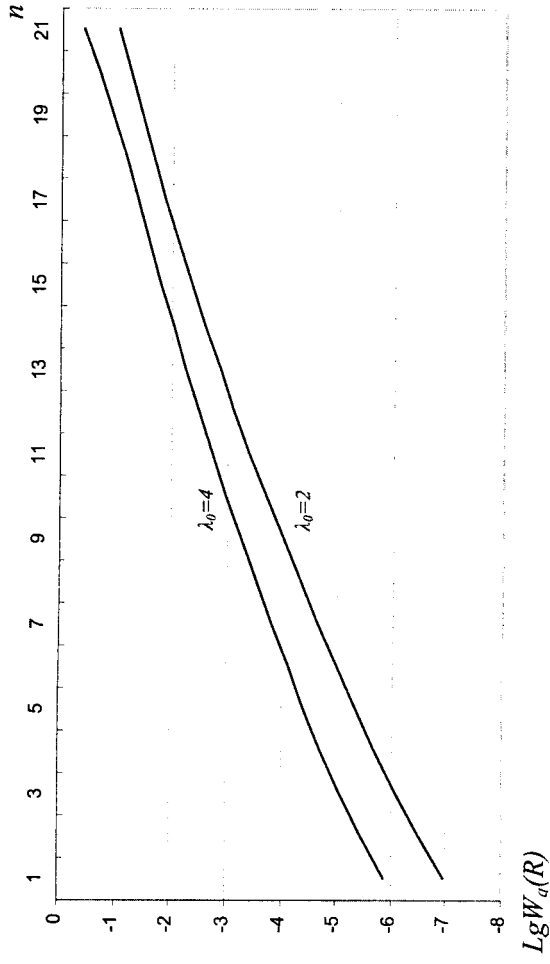


Рис. 3.25. Зависимость среднего времени ожидания o -вызовов от числа резервных каналов в модели с конечной очередью

Важно отметить, что даже при невыполнении вышеуказанного допущения относительно соотношений интенсивностей разнотипных вызовов, разница между результатами приближенного расчета и соответствующими результатами работы [6] в наихудших случаях не превосходит 10^{-5} . Так, например, для модели с 20-ти каналов ($m+n=20$) и нагрузочными параметрами $\mu=2$, $\lambda_0=\lambda_n=7$, максимальная разница получена при $n=1$, т.е. для этих параметров значения вероятности потери i -вызовов, вычисленные с помощью алгоритма работы [6] и предложенного здесь алгоритма равны $1.91E-05$ и $4.13E-06$, соответственно. Аналогичные результаты были получены и при других соотношениях интенсивностей разнотипных трафиков.

Заключение и комментарии к главе 3 части I

Представленные в данной главе результаты полностью базируются на идеях работ [1,2]. Здесь предложены численные алгоритмы расчета характеристик процессов обслуживания разнотипных вызовов в мобильных беспроводных сетях связи с очередями. Показано, что адекватными математическими моделями этих процессов являются двумерные цепи Маркова с соответствующими производящими матрицами, элементы которых определяются исходя из конкретной схемы обслуживания разнотипных вызовов. Эти модели позволяют определить не только вероятность блокировки/потери разнотипных вызовов, но также с их помощью найти среднее число вызовов каждого типа в системе и/или в

ЗАДАЧИ УЛУЧШЕНИЯ ХАРАКТЕРИСТИК ПРОЦЕССОВ ОБСЛУЖИВАНИЯ ВЫЗОВОВ В БЕСПРОВОДНЫХ СЕТЯХ СОТОВОЙ СВЯЗИ

В предыдущих двух главах получены формулы для расчета характеристик процессов обслуживания разнотипных вызовов при использовании разных стратегий доступа. В большинстве случаев эти стратегии содержат различные параметры, значения которых подаются регулированию. Так, например, при использовании стратегии доступа на основе резервных каналов в моделях соты с явными потерями, легко видеть, что с увеличением количества резервируемых каналов для h -вызовов (т.е. g) вероятность потери h -вызовов должна уменьшаться, а вероятность блокировки o -вызовов должна увеличиваться, но при этом должна уменьшаться утилизация каналов. Эти теоретические ожидания полностью подтвердились с помощью вычислительных экспериментов. Иными словами, использование SAC на базе GS-стратегии доступа приводит к серьезным ухудшениям утилизации и без этого дефицита ресурсов соты. Последнее обстоятельство особенно остро проявляется в городских сотах и макросотах, в которых h -вызовы имеют низкую интенсивность. Такие же проблемы возникают при исследовании моделей с очередями.

С учетом изложенного интерес представляют задачи выбора таких экстремальных значений параметров регулирования, при которых не только удовлетворились бы ограничения на вероятность потерь разнотипных вызовов, но также и коэффициент использования каналов находился бы в желаемых пределах. В данной главе рассматриваются некоторые задачи подобного типа.

4.1. Задачи улучшения характеристик процессов обслуживания вызовов в сотах без очередей

Сначала рассмотрим задачи выбора оптимальных значений количества резервируемых каналов в модели изолированной соты.

оды не позволяют определить элементов показана высокая точность расчета характеристик значений изучаемых характеристик из формул, общепринятых лиз показал, что результаты аналитических расчетов, приведенных совпадают, а при вычислении склонение в наихудших случаях делей в предположении идентичности времени занятия канала

Бабаев А.Т. Численный метод сетей связи с ограниченной управленения и информатики. —

Бабаев А.Т. Исследование с ограниченной очередью обслуживания и информатики. — Model and performance analysis of systems with prioritized and nonprioritized access // IEEE Trans. on Vehicular technology. —

personal portable radio telephone channels // IEEE J. Selected Areas in Commun. — 1988. — 6(4). — С. 911 — 917.

queuing priority channel assignment // IEEE Trans. on Vehicular technology. — 1988. — 37(4). — С. 704 — 712.

$$P_o(N, 0) \leq P_o(N, g) \leq P_o(N, N-1); \quad (4.1)$$

$$P_h(N, N-1) \leq P_h(N, g) \leq P_h(N, 0). \quad (4.2)$$

Из (4.1) и (4.2) видно, что максимально справедливое обслуживание достигается при $g = 0$, так как в этом случае при любых нагрузках и значениях числа каналов (N) имеем:

$$P_h(N, 0) = P_o(N, 0)$$

А минимально справедливое обслуживание достигается при $g = N-1$, так как в этом случае разница $P_o(N, g) - P_h(N, g)$ достигает максимального значения, т.е.

$$\max_g \{P_o(N, g) - P_h(N, g)\} = P_o(N, N-1) - P_h(N, N-1) \quad (4.3)$$

Более интересными являются задачи, в которых требуется найти такие (экстремальные) значения параметра g (при фиксированном N), чтобы обеспечить заданный уровень качества обслуживания разнотипных вызовов. При этом уровень качества обслуживания зачастую определяется уровнем потерь (блокировки) хэндовер и новых вызовов, соответственно. Одна из возможных постановок задач такого типа состоит в следующем.

Пусть при заданных N, v_o и v_h требуется найти такое оптимальное значение g , чтобы минимизировать $P_o(N, g)$ при известном ограничении на $P_h(N, g)$, т.е. требуется решить следующую задачу:

$$P_o(N, g) \xrightarrow{g} \min \quad (4.4)$$

$$P_h(N, g) \leq P_{h_0}, \quad (4.5)$$

где P_{h_0} является заданным числом.

Задача (4.4), (4.5) легко решается с учетом свойства монотонности функций $P_o(N, g)$ и $P_h(N, g)$ относительно аргумента g при фиксированном N . Действительно, с учетом этого свойства одним из известных методов одномерного поиска (в частности методом деления пополам) находится такое минимальное значение $g^* \in [0, N-1]$, которое удовлетворяет ограничению (4.5), именно это значение g^* и является решением задачи (4.4), (4.5). При этом очевидно, что если $P_{h_0} < P_h(N, N-1)$, то задача (4.4), (4.5) не имеет решения.

Наиболее интересной является следующая задача: при фиксированных N, v_o и v_h требуется найти такой промежуток изменения параметра g , внутри которого удовлетворяются заданные ограничения на величины

$P_h(N, g)$ и $P_o(N, g)$, т.е. требуется найти такие $\underline{g}, \bar{g} \in [0, N-1]$, $\underline{g} \leq \bar{g}$, чтобы

$$\bar{g} - \underline{g} \rightarrow \max, \quad (4.6)$$

$$P_o(N, g) \leq P_{o_0}, \quad (4.7)$$

$$P_h(N, g) \leq P_{h_0}, \forall g \in [\underline{g}, \bar{g}], \quad (4.8)$$

где P_{o_0} и P_{h_0} являются заданными числами.

Для решения задачи (4.6)-(4.8) с учетом сделанных выше замечаний относительно монотонности функций P_o и P_h , а также соотношений (4.1), (4.2) можно предложить следующий алгоритм.

Шаг 1. Если $P_{h_0} \leq P_h(N, N-1)$ и/или $P_{o_0} < P_o(N, 0)$, то задача (4.6)-(4.8) не имеет решения.

Шаг 2. Параллельно решаются следующие одномерные задачи:

$$g_o^* := \max \{g \mid P_o(N, g) \leq P_{o_0}\} \quad (4.9)$$

$$g_h^* := \min \{g \mid P_h(N, g) \leq P_{h_0}\} \quad (4.10)$$

Шаг 3. Если $g_o^* \leq g_h^*$, то задача (4.6)-(4.8) не имеет решения; в противном случае решением задачи являются $\underline{g} := g_o^*$, $\bar{g} := g_o^*$.

Теперь рассмотрим следующую задачу. Необходимо найти такое максимальное значение g , чтобы удовлетворялись ограничения на $P_o(N, g)$, $P_h(N, g)$ и, вдобавок, на $N_{av}(N, g)$. Формально эта задача может быть описана так:

$$g \rightarrow \max, \quad (4.11)$$

$$P_o(N, g) \leq \varepsilon_o, \quad (4.12)$$

$$P_h(N, g) \leq \varepsilon_h, \quad (4.13)$$

$$N_{av}(N, g) \geq \tilde{N}_{av}, \quad (4.14)$$

где ε_o , ε_h и \tilde{N}_{av} заданные значения.

Оптимальное решение проблемы (4.11) - (4.14), если оно существует, обозначим g^* . Для решения этой проблемы можно использовать следующий алгоритм.

Шаг 1. Найти такое максимальное значение g , $g \in [0, N-1]$, чтобы удовлетворялось условие (4.14). Это значение обозначается как g_o .

Шаг 2. Если $P_h(N, g_o) > \varepsilon_h$, задача не имеет решения.

Шаг 3. Если $P_o(N, g_o) \leq \varepsilon_o$, то $g^* := g_o$.

Шаг 4. В интервале $[0, g_o]$ найти такое максимальное значение \tilde{g} , чтобы условие $P_o(N, \tilde{g}) \leq \varepsilon_o$ было верным. Если $P_h(N, \tilde{g}) \leq \varepsilon_h$, то $g^* := \tilde{g}$; иначе задача не имеет решения.

Сложность предложенного алгоритма для решения оптимизационных задач (без учета сложности вычисления $P_h(N, g)$, $P_o(N, g)$ и $N_{av}(N, g)$) достаточно невысокая, так как эти алгоритмы требуют вычисления значений указанных выше функций максимум в $[\log_2 N]$ точках, где $[x]$ — целая часть x , а также проведения сравнения со значениями ε_o , ε_d и \tilde{N}_{av} . Таким образом алгоритмы имеют логарифмическую сложность и не трудны для реализации.

Результаты решения оптимизационных задач (4.4), (4.5), (4.6)–(4.8) и (4.11)–(4.14) показаны в табл. 4.1, 4.2 и 4.3, соответственно, где символ \emptyset означает, что задача не имеет решения.

Таблица 4.1

Результаты решения задачи (4.4),(4.5) при $N = 100$, $v_o = 40$ эрл.

v_h	40	40	40	40	40	40	40	40	50	50	50	50
P_{h_0}	E-02	E-03	E-04	E-05	E-06	E-04	E-03	E-02	E-03	E-04	E-05	E-06
g^*	0	11	20	27	33	7	19	27	34	41		

Таблица 4.2

Результаты решения задачи (4.6)–(4.8) при $N = 100$, $v_o = 40$ эрл.

v_h	20	20	40	40	40	50	50	50	50	60
P_{o_0}	E-02	E-02	4E-01	4E-01	E-02	5E-01	5E-01	5E-01	5E-01	5E-01
P_{h_0}	E-07	E-08	E-03	E-04	E-03	E-03	E-03	E-03	E-03	E-02
$[g, \tilde{g}]$	[16,28]	[22,22]	[11,33]	[20,33]	\emptyset	\emptyset	[19,29]	\emptyset	\emptyset	[13,19]

Таблица 4.3

Результаты решения задачи (4.11)–(4.14)

v_o	25	30	35	40	40	45	45	60	60	65
v_h	25	35	30	40	40	50	50	40	40	70
ε_o	E-02	E-02	E-03	E-03	E-02	E-03	E-02	E-03	E-02	E-03
ε_h	E-03	E-04	E-04	E-04	E-04	E-05	E-05	E-06	E-06	E-06
N	80	100	100	120	120	150	150	120	120	120
\tilde{N}_{av}	48	60	60	70	70	80	80	70	75	75
g^*	13	16	8	10	19	23	23	6	\emptyset	19

Как видно из таблицы 4.1, с ростом P_{h_0} значение g^* уменьшается, а g^* увеличивается с ростом v_h (при фиксированных значениях остальных параметров). Эти результаты были ожидаемыми, так как предъявление более строгих требований к уровню потерь хэндовер-вызовов, а также увеличение трафика таких заявок требуют увеличения числа сторожевых каналов для обслуживания заявок данного типа.

Из таблицы 4.2. видно, что в интервале существования решения задачи (4.6)–(4.8) с ростом P_{h_0} значение \underline{g} уменьшается, а с ростом P_{o_0} значение \underline{g} также растет (при фиксированных значениях остальных параметров).

Оптимальное решение задачи (4.11)–(4.14) зависит от значений шести начальных параметров, поэтому достаточно сложно делать какие-либо выводы по поводу изменений значения g^* в ответ на изменение начальных параметров. Однако есть некоторые явные выводы, например, с увеличением \tilde{N}_{av} значение оптимального решения, если оно существует, уменьшается.

Теперь рассмотрим задачи улучшения характеристик модели соты с индивидуальными каналами для h -вызовов (изолированное резервирование). В данной схеме предоставления приоритетов h -вызовам при фиксированном числе каналов обслуживания участвует также параметр r — число индивидуальных каналов для h -вызовов. Очевидно, что с помощью варьирования значения этого параметра иногда можно удовлетворить заданные ограничения на характеристики системы. Увеличение

параметра r приводит к уменьшению вероятности потери h -вызовов, но при этом увеличивается вероятность потери o -вызовов, а также уменьшается коэффициент эффективного использования каналов.

Здесь аналогично задаче (4.11)–(4.14) рассматривается задача нахождения оптимальных значений параметра r , при которых удается удовлетворить заданные ограничения на характеристики согты.

Пусть общее число каналов N и нагруженные параметры разнотипных трафиков v_o и v_h являются заданными величинами. Пусть также заданы ограничения на вероятность потерь разнотипных вызовов, а также на коэффициент эффективного использования каналов, т.е. аналогично (4.12)–(4.14) указанные величины должны удовлетворять следующим ограничениям:

$$P_o(N, r) \leq \varepsilon_o, \quad (4.15)$$

$$P_h(N, r) \leq \varepsilon_h, \quad (4.16)$$

$$\frac{1}{N} \tilde{N}_{av}(N, r) \geq \varepsilon_c, \quad (4.17)$$

где ε_o , ε_h и ε_c являются заданными числами.

Задача формулируется так. Требуется найти максимальное (\bar{r}) и минимальное (\underline{r}) значения параметра r (если таковые существуют), чтобы удовлетворялись ограничения (4.15)–(4.17), т.е. требуется решить следующую задачу:

$$\bar{r} - \underline{r} \rightarrow \max, \quad (4.18)$$

при ограничениях (4.15)–(4.17), $\bar{r}, \underline{r} \in [0, N - 1]$.

Эту задачу можно разделить на две сепарабельные задачи:

Задача I. Нахождение \bar{r} при ограничениях (4.15)–(4.17),

$$\bar{r} \in [0, N - 1].$$

Задача II. Нахождение \underline{r} при тех же ограничениях.

Если обе задачи одновременно имеют решение, то оптимальное решение исходной задачи (4.15)–(4.18) строится очевидным образом; в противном случае (т.е. если хотя бы одна из указанных задач не имеет решения) исходная задача тоже не имеет решения.

Теперь рассмотрим метод решения указанных сепарабельных задач. Отметим, что при решении этих задач существенным образом используются свойства монотонности функций $P_o(N, r)$, $P_h(N, r)$ и $\tilde{N}_{av}(N, r)$ относительно аргумента r при фиксированном значении N .

Из условия (4.17) видно, что если $\frac{1}{N} \tilde{N}_{av}(N, 0) < \varepsilon_c$, то данная задача не имеет решения. Действительно, поскольку максимальное использование каналов достигается при $r = 0$, и если даже при такой схеме не выполняется условие (4.17), то задача I не имеет решения. В

противном случае (т.е. если $\frac{1}{N} \tilde{N}_{av}(N, 0) \geq \varepsilon_c$) оптимальное решение данной задачи (если оно существует) находится в интервале $[0, r^*]$, где r^* — максимальное значение $r \in [0, N - 1]$, при котором условие (4.17) удовлетворяется. Для нахождения значения r^* исходя из свойства монотонности функции $\tilde{N}_{av}(N, r)$, можно применить метод деления пополам (дихотомии).

Если $P_o(N, r^*) \leq \varepsilon_o$ и $P_h(N, r^*) \leq \varepsilon_h$, то решением задачи I является $\bar{r} := r^*$. Если в точке $r = r^*$ условие (4.16) не выполняется, то независимо от выполнения условия (4.15), задача I не имеет решения. Это объясняется тем, что поскольку функция $P_h(N, r)$ является монотонно убывающей относительно аргумента r , и условие (4.16) не удовлетворяется в точке $r = r^*$, то оно может выполняться лишь при $r > r^*$, но при $r > r^*$ нарушается условие (4.17).

Если $P_o(N, r^*) > \varepsilon_o$ и $P_h(N, r^*) \leq \varepsilon_h$, то в интервале $[0, r^*]$ находится такое максимальное r^{**} , чтобы $P_o(N, r^{**}) \leq \varepsilon_o$. Для нахождения r^{**} также с учетом монотонности функции $P_o(N, r)$ относительно аргумента r может быть использован метод дихотомии. Тогда, если $P_h(N, r^{**}) \leq \varepsilon_h$, то решением задачи I является $\bar{r} := r^{**}$; в противном случае данная задача не имеет решения.

Резюмируя, можно предложить следующий алгоритм решения задачи I.

Шаг 1. Если $\frac{1}{N} \tilde{N}_{av}(N, 0) < \varepsilon_c$, то задача I не имеет решения.

Шаг 2. Находится максимальное значение r^* в интервале $[0, N]$, чтобы выполнялось условие (4.17).

Шаг 3. Если $P_h(N, r^*) > \varepsilon_h$, то задача I не имеет решения.

Шаг 4. Если $P_o(N, r^*) \leq \varepsilon_o$ и $P_h(N, r^*) \leq \varepsilon_h$, то решением задачи I является $\bar{r} := r^*$.

Шаг 5. В интервале $[0, r^*]$ находится максимальное r^{**} , чтобы $P_o(N, r^{**}) \leq \varepsilon_o$. Если $P_h(N, r^{**}) \leq \varepsilon_h$, то решением задачи I является $\bar{r} := r^{**}$; в противном случае задача не имеет решения.

Используя вышеописанную схему решения задачи I, можно предложить следующий алгоритм решения задачи II.

Шаг 1. Совпадает с шагом 1 алгоритма решения задачи I.

Шаг 2. Совпадает с шагом 2 алгоритма решения задачи I.

Шаг 3. Совпадает с шагом 3 алгоритма решения задачи I.

Шаг 4. Если $P_o(N, r^*) \leq \varepsilon_o$ и $P_d(N, r^*) \leq \varepsilon_d$, то в интервале $[0, r^*]$ находится такое минимальное r_{**} , чтобы $P_d(N, r_{**}) \leq \varepsilon_d$. Тогда решением задачи II является $\underline{r} := r_{**}$.

4.2. Задачи улучшения характеристик процессов обслуживания вызовов в сотах с очередями

В этом параграфе рассмотрим задачи улучшения характеристик процессов обслуживания вызовов в сотах, в которых допускается образование очереди для разнотипных вызовов.

Сначала рассмотрим модель, в которой h -вызовы в ограниченной очереди являются весьма терпеливыми. Полученные в § 3.1 формулы для приближенного расчета характеристик изучаемых процессов позволяют также произвести их оптимизацию относительно выбранного критерия качества обслуживания разнотипных вызовов. При выводе указанных формул было принято допущение о том, что нет ограничений на время ожидания h -вызовов в очереди. Теперь рассмотрим способ, который позволяет учитывать не только эти ограничения, но также ограничения на вероятность потери разнотипных вызовов.

Предложенный способ основывается на варьировании размера буфера для ожидания в очереди h -вызовов. Как было указано в § 3.1, при увеличении размера буфера также увеличивается и среднее время ожидания h -вызовов в системе, т.е. с помощью выбора соответствующего значения параметра B можно добиться того, чтобы значения указанной характеристики (т.е. W_h) находились в желаемых пределах. С другой стороны, уменьшение значения параметра B приводит к возрастанию вероятности потери h -вызовов (и уменьшению вероятности потери o -вызовов). Таким образом, возникает следующая задача оптимизации данной модели: при фиксированных нагрузках разнотипных вызовов (т.е. ν_o и ν_h) и при известных значениях общего (N) и резервного (g) числа каналов требуется найти такое максимальное значение размера буфера (B^*), при котором удовлетворятся заданные ограничения на время ожидания h -вызовов в очереди, а также на вероятность потери разнотипных вызовов. Формально эта задача записывается так:

$$B \rightarrow \max, \quad (4.19)$$

при ограничениях

$$W_h(B) \leq \delta, \quad (4.20)$$

$$P_h(B) \leq \epsilon_h, \quad (4.21)$$

$$P_o(B) \leq \epsilon_o, \quad (4.22)$$

где δ , ϵ_o и ϵ_h являются заданными величинами.

Замечание 4.1. С целью выделения параметра оптимизации здесь и в дальнейшем в обозначениях соответствующих функций аргумент указывается явным образом (в данной задаче этим аргументом является B , т.к. другие параметры модели фиксированы).

Укрупненное описание одного из возможных алгоритмов решения задачи (4.19)–(4.22) состоит в следующем. В первую очередь находится

Шаг 5. В интервале $[0, r^*]$ находится минимальное \tilde{r}_{**} , чтобы $P_h(N, \tilde{r}_{**}) \leq \epsilon_h$. Если $P_o(N, \tilde{r}_{**}) \leq \epsilon_o$, то решением задачи II является $\underline{r} := \tilde{r}_{**}$; в противном случае задача не имеет решения.

Очевидно, что задача (4.15)–(4.18) имеет решение лишь тогда, когда обе задачи I и II одновременно имеют решения, т.е. если хотя бы одна из этих задач не имеет решения, то исходная задача также не имеет решения.

Результаты решения сформулированных выше оптимизационных задач показаны в табл. 4.4.

На основе анализа результатов решений рассматриваемой задачи можно сделать следующие выводы:

- с ростом суммарной нагрузки длина оптимального интервала $(r - \underline{r})$ сокращается, и, в конечном счете, задача не имеет решения;
- с уменьшением ϵ_o оптимальный интервал (если он существует) перемещается влево внутри общего интервала $[0, N - 1]$;
- с уменьшением ϵ_h оптимальный интервал (если он существует) перемещается вправо внутри общего интервала $[0, N - 1]$;
- с уменьшением ϵ_o оптимальный интервал (если он существует) перемещается влево внутри общего интервала и его длина сокращается.

Таблица 4.4

Результаты решения задачи (4.15)–(4.18) при $N = 40$

ν_o	ν_h	ϵ_o	ϵ_h	ϵ_c	\bar{r}	\underline{r}
10	10	E-02	E-02	E-01	22	10
50	35	E-02	E-02	E-01	∅	∅
20	15	E-03	E-01	E-02	∅	∅
15	5	E-02	E-05	E-01	16	12
15	8	E-01	E-04	E-03	22	7
20	8	E-01	E-04	E-03	∅	∅
15	12	E-01	E-04	E-03	∅	∅
20	5	E-01	E-03	E-02	17	8
20	5	E-02	E-03	E-02	10	8
20	5	E-02	E-02	E-02	10	0
20	5	E-02	E-03	E-03	10	8
5	10	E-02	E-03	E-03	29	0
5	15	E-03	E-04	E-03	26	0
5	15	E-03	E-05	E-03	∅	∅

такое максимальное значение \underline{B} , чтобы удовлетворялось ограничение (4.20). Обозначим это значение \underline{B} .

Примечание. Для нахождения \underline{B} может быть использован следующий способ. Функция W_h аппроксимируется снизу величиной \underline{W}_h , представляющей собой среднее время ожидания в классической системе обслуживания $M | M | N | B$ с нагрузкой v_h эрл., и находится наименьшее целое решение (относительно B) неравенства $\underline{W}_h \leq \delta$, обозначаемое через B_0 . Далее из интервала $[1, B_0]$ находится искомое \underline{B} , при этом, исходя из свойства монотонности функции $W_h(B)$, может быть использован, в частности, метод дихотомии.

Аналогичным образом может быть решена задача нахождения минимального значения \underline{B} (т.е. \underline{B}^*), при котором удовлетворяются ограничения (4.19)–(4.22). Таким образом, объединяя решение этих двух задач оптимизации можно указать концы интервала изменения значений B , внутри которого удовлетворяются заданные ограничения на характеристики процессов обслуживания разнотипных вызовов.

Некоторые результаты решения указанных задач оптимизации показаны в табл. 4.5.

Таблица 4.5

Результаты решения задачи (4.19)–(4.22) при $N = 25, g = 5$

λ_0	λ_h	μ	ε_0	ε_h	δ	\underline{B}^*	\underline{B}^*
10	40	3	0.2	E-04	2.00	100	12
60	25	3	0.6	E-03	0.35	100	21
60	25	3	0.5	E-02	0.34	14	4
60	25	2	0.8	E-02	0.55	19	11
60	25	1	0.8	E-02	0.55	\emptyset	\emptyset
50	15	2	0.5	E-02	0.50	\emptyset	\emptyset

Анализ результатов рассматриваемых задач оптимизации позволяет сделать ряд общих выводов:

- с увеличением нагрузки трафика любого типа (одновременно или в отдельности) значение \underline{B}^* (если оно существует) уменьшается;
- с увеличением скорости радиоканалов значение \underline{B}^* также увеличивается;
- при ослаблении требований к вероятностям потери разнотипных вызовов (т.е. при увеличении значений ε_0 и ε_h) длина интервала $[\underline{B}^*, \underline{B}^*]$ увеличивается;

– при ослаблении требований на время ожидания в очереди h -вызовов (т.е. при увеличении δ) длина интервала $[\underline{B}^*, \underline{B}^*]$ уменьшается.

Теперь рассмотрим задачи улучшения характеристик процессов обслуживания вызовов в сотах, в которых допускается образования ограниченной очереди для нетерпеливых h -вызовов. При этом в качестве управляемого параметра могут быть использованы размер буфера, общее число каналов и другие параметры системы. Однако здесь предполагается, что единственно варьируемым (управляемым) параметром является число резервируемых каналов для h -вызовов (т.е. g).

Как показали результаты численных экспериментов для этой модели, с увеличением числа резервных каналов при фиксированных значениях остальных параметров сети также увеличивается вероятность блокировки o -вызовов и одновременно уменьшается вероятность потери h -вызовов. Следовательно, с помощью увеличения числа резервных каналов можно достичь желаемого уровня потерь для более важных h -вызовов. Однако это приводит к неэффективному использованию всего пула каналов, так как с увеличением числа резервных каналов уменьшается коэффициент использования радиоканалов. Последнее обстоятельство особенно остро проявляется в городских сотах, где интенсивность h -вызовов низкая.

Таким образом, возникает следующая задача улучшения характеристик процессов обслуживания разнотипных вызовов: при известных нагрузочных (т.е. v_0, v_h и γ) и структурных (т.е. N и B) параметрах соты требуется найти такой интервал изменения значений числа резервных каналов (т.е. g), внутри которого удовлетворятся заданные ограничения на вероятности потерь разнотипных вызовов, а также на среднее число занятых каналов.

Решение этой задачи находится в результате определения решений следующих сепарабельных задач оптимизации:

$$\bar{g} := \arg \max_{g \in \{1, N\}} \{P_h(g) \leq \varepsilon_h, P_o(g) \leq \varepsilon_o, \tilde{N}(g) \geq N_o\}, \quad (4.23)$$

$$\underline{g} := \arg \min_{g \in \{1, N\}} \{P_h(g) \leq \varepsilon_h, P_o(g) \leq \varepsilon_o, \tilde{N}(g) \geq N_o\}, \quad (4.24)$$

где $\varepsilon_h, \varepsilon_o$ и N_o являются заданными величинами (с целью выделения параметра оптимизации в данной задаче в обозначениях функций P_o, P_h и \tilde{N} явным образом указывается аргумент g).

Очевидно, что задачи (4.23), (4.24) одновременно либо обе имеют решение, либо обе не имеют решения.

Учитывая свойство монотонности участвующих в задачах (4.23), (4.24) функций $P_o(g), P_h(g)$ и $\tilde{N}(g)$ можно предложить следующий алгоритм их решения. Для конкретности изложения рассмотрим алгоритм решения задачи (4.23).

Если $\tilde{N}(I) \leq N_o$, то данная задача (4.23) не имеет решения; в противном случае находится решение следующей задачи.

$$N_1 := \arg \max_{g \in [1, N]} \{ \tilde{N}(g) \geq N_o \}. \quad (4.25)$$

Если $P_h(N_1) > \varepsilon_h$, то данная задача (4.23) не имеет решения; в противном случае если $P_o(N_1) \leq \varepsilon_o$, то решением задачи (4.23) является $\bar{g} := N_1$.

Если $P_o(N_1) > \varepsilon_o$, то решается следующая задача:

$$N_2 := \arg \max_{g \in [1, N_1]} \{ P_o(g) \leq \varepsilon_o \}. \quad (4.26)$$

Тогда, если $P_h(N_2) \leq \varepsilon_h$, то решением задачи (4.23) является $\bar{g} := N_2$; в противном случае задача (4.23) не имеет решения.

Отметим, что для решения задач (4.25), (4.26) можно использовать, в частности, метод дихотомии. Аналогичным способом решается задача (4.24), и, таким образом, объединяя решения этих двух задач оптимизации, можно указать концы интервала изменения значений g , внутри которого удовлетворяются заданные ограничения на показатели QoS различных вызовов.

Некоторые результаты решения задач (4.23), (4.24) показаны в табл. 4.6.

Таблица 4.6

Результаты решения задач (4.23), (4.24) при $N = 40, B = 10, \gamma = 20, \mu = 5$

λ_o	10	10	10	20	20	5	5	5	6	4
λ_h	8	8	8	8	10	10	10	10	10	6
ε_h	E-02	E-04	E-06	E-02	E-02	1.5E-02	E-01	E-01	E-07	E-04
ε_o	E-02	10 ⁻³	E-06	E-03	E-03	3E-03	3E-03	3E-03	3E-03	3E-03
N_o	2	3	3	4	6	2	2	3	2	2
\underline{g}	1	1	1	27	∅	32	32	32	34	1
\bar{g}	30	27	24	28	∅	32	34	33	34	18

На основе анализа результатов решения рассматриваемой задачи сделаем несколько общих выводов:

- при ослаблении требований к вероятностям потерь (т.е. с ростом ε_h и ε_o) \underline{g} также увеличивается, а \bar{g} уменьшается;
- при ослаблении требований к среднему числу занятых каналов (т.е. с ростом N_o) длина интервала $\bar{g} - \underline{g}$ сокращается и, в конечном счете, задача не имеет решения;
- с уменьшением нагрузки трафика любого типа (т.е. λ_1 и λ_2) длина интервала $\bar{g} - \underline{g}$ растет.

Задачи выбора оптимальных значений используемой стратегии доступа и структурных параметров моделей обслуживания при возможности образования очереди (конечной или бесконечной длины) о-вызовов ставятся и решаются подобным образом. Возможность самостоятельного исследования этих задач представляется читателю.

Заключение и комментарии к главе 4 части I

Здесь формулированы и решены некоторые задачи улучшения характеристик процессов обслуживания разнотипных вызовов как в изолированной соте, так и в сетях связи, в которых допускается образование их очереди. В различных моделях в качестве управляемых параметров выбираются как параметры соты, так и параметры стратегии управления доступом разнотипных вызовов.

Предложены алгоритмы решения трех задач по улучшению характеристик процессов обслуживания в изолированной соте, в которой стратегия доступа основана на GS-схеме резервирования каналов (неизолированное резервирование). В одной из них найдено оптимальное значение числа резервных каналов, при котором минимизируется вероятность блокировки о-вызовов и удовлетворяется заданное ограничение на вероятность потери h-вызовов. Во второй задаче найдены пределы изменения числа резервных каналов для h-вызовов, при которых удовлетворяются заданные ограничения на вероятность блокировки о-вызовов и вероятность потери h-вызовов. Наконец, в третьей задаче дополнительно учитывается ограничение на утилизацию пула каналов.

Предложен также алгоритм решения задачи нахождения оптимальных пределов изменения размера индивидуальной зоны для стратегии доступа, основанной на выделении специальной зоны для обслуживания h-вызовов (изолированное резервирование). При этом оптимальность понимается в смысле максимизации длины указанного интервала, при котором удовлетворяются не только заданные ограничения на вероятность блокировки/потери о- и h-вызовов, но также на утилизацию суммарной пропускной способности каналов соты.

Для модели соты с ограниченной очередью терпеливых h -вызовов предложен алгоритм решения задачи нахождения максимального размера буфера для h -вызова, при котором удовлетворяются ограничения на вероятность блокировки/потери o - и h -вызовов и на время ожидания в очереди h -вызовов.

Для модели соты с ограниченной очередью нетерпеливых h -вызовов предложен алгоритм решения задачи нахождения интервала изменения значений числа резервных каналов с максимальной длиной, внутри которого удовлетворяются заданные ограничения на вероятность потери разнотипных вызовов, а также на среднее число занятых каналов.

Приведены результаты вычислительных экспериментов для всех задач оптимизации характеристик процессов обслуживания разнотипных вызовов в исследуемых моделях. Их подробный анализ позволяет оптимальным образом определить параметры используемой стратегии доступа в зависимости от нагрузок трафиков и заданных ограничений на характеристики соты, а также определить вид необходимой стратегии доступа, позволяющей оптимальным образом удовлетворить заданные ограничения на характеристики системы.

Результаты данной главы базируются на работах [1–6]. В доступной литературе рассматриваемые здесь задачи мало исследованы (см., например, [7–9]), хотя они являются важными и интересными и в научном плане, и в плане практических приложений.

Литература к главе 4 части I

1. Пономаренко Л.А., Меликов А.З., Бабаев А.Т. Численный метод исследования моделей сотовых сетей связи с ограниченной очередью h -вызовов // Проблемы управления и информатики. – 2005. – N 3. – С.76 – 88.
2. Пономаренко Л.А., Меликов А.З., Бабаев А.Т. Исследование характеристик сетей сотовой связи с ограниченной очередью нетерпеливых h -вызовов // Проблемы управления и информатики. – 2006. – N 4. – С.97 – 107.
3. Melikov A.Z., Babayev A.T. A new method of performance analysis of queueing model with guard channels // Proceedings of Fifth International Workshop on Retrial Queues. Seoul, Korea, September 2004. – pp. 103 – 110.
4. Melikov A.Z., Babayev A.T. Analysis of call handling in single cell in wireless networks with impatient handover calls // Applied and Computational Mathematics. – 2005. – No.2. – pp. 3 – 9.
5. Melikov A.Z., Fattakhova M.I., Babayev A.T. Calculation and optimization of call processing procedures in cellular wireless communication networks // Automatic control and computer sciences. Allerton Press Inc. – 2004. – Vol. 38, No.4. – pp. 55 – 63.

6. Melikov A.Z., Fattakhova M.I., Babayev A.T. Studies of cellular communication networks with individual channels for service of handover-calls // Automatic Control and Computer Sciences. Allerton Press Inc. – 2005. – Vol.39, No.3. – pp.67 – 74.
7. DasBit S., Mitra S. Challenges of computing in mobile cellular environment – a survey // Computer Communications. – 2003. – Vol.26. – pp. 2090 – 2105.
8. Ramjee R., Towsley D., Nagarajan R. On optimal call admission controls in cellular networks // Wireless Networks. – 1997. – Vol. 3. – pp. 29 – 41.
9. Sidi M., Starobinski D. New call blocking versus handoff blocking in cellular networks // Wireless Networks. – 1997. – Vol. 3. – pp. 15 – 27.

ИССЛЕДОВАНИЕ МОДЕЛЕЙ МУЛЬТИСЕРВИСНЫХ СЕТЕЙ СВЯЗИ

Введение

Современные системы телекоммуникации являются мультисервисными, так как в них осуществляется обработка широкого спектра различных сообщений — речь, видео, данные, факс, электронная почта и т.д. Высокая сложность этих систем выдвигает на первый план проблемы разработки достаточно адекватных математических моделей с целью получения достоверных оценок их характеристик, реализации задач обеспечения необходимого уровня качества обслуживания различных сообщений, а также решения задач их оптимизации относительно выбранного критерия качества и разработки соответствующих алгоритмов управления системами.

В настоящее время в мире издается довольно много авторитетных научных журналов, где ежегодно публикуются сотни статей, посвященных исследованию различных аспектов указанной глобальной проблемы. В последние годы появился ряд монографий по этой тематике. Во всех работах отмечается, что основным математическим аппаратом, позволяющим строить аналитические модели, адекватные мультисервисным системам телекоммуникации, является аппарат теории систем и сетей массового обслуживания. При этом очевидно, что классические однополосные системы обслуживания не могут служить адекватными математическими моделями реальных систем телекоммуникации, так как в указанных системах предполагается, что все заявки являются идентичными по всем показателям (например, по значениям нагрузочных параметров, уровню приоритетов, механизмам обслуживания и т.д.).

Вместе с тем, в современных мультисервисных системах телекоммуникации сообщения существенно отличаются друг от друга почти по всем показателям, и поэтому их работа может быть достаточно точно описана лишь моделями многопоточковых систем. Классическая теория многоканальных и многопоточковых систем базируется на многих предположениях, одна из которых является основной: «одна заявка — один канал». Иными словами, в них предполагается, что заявка любого типа во всем периоде обслуживания использует не более одного канала (ресурса) системы.

Однако, в современных мультисервисных системах телекоммуникации эта предпосылка не выполняется. Так в них, например, видеоинформация требует более широкой полосы в цифровой линии передачи, чем данные или речевая информация. В литературе по теории телеграфика запросы, требующие большого числа каналов в линии передачи, называются широкополосными (*wide-band*), а запросы, требующие меньшего

числа каналов — узкополосными (*narrow-band*). В литературе по теории систем обслуживания многопоточковые системы, в которых различные заявки требуют для обслуживания одновременно случайное число каналов называются многопоточковыми. (*Multi-Rate Queue, MRQ*).

В многопоточковых системах необходимо различать два типа заявок: неэластичные и эластичные. Неэластичными называются заявки, которые имеют фиксированную ширину. Это означает, что при обслуживании неэластичной заявки все каналы, обслуживающие данную заявку, включаются одновременно, и канал, окончивший свою работу, блокируется до момента, когда последний канал из множества каналов, обслуживающих данную заявку, закончит работу. Иными словами, обслуживание неэластичной заявки может начинаться лишь тогда, когда имеется необходимое число свободных каналов, требующихся для ее обслуживания.

В отличие от неэластичных заявок, эластичные заявки не имеют фиксированной ширины, т.е. для эластичных заявок задаются только пределы изменения их ширины. Последнее означает, что начало и/или конец акта обслуживания эластичной заявки на разных каналах могут не совпадать. Иными словами, при обслуживании эластичной заявки время начала обслуживания разными каналами может не совпадать, и при этом каждый канал имеет индивидуальное задание по обслуживанию заявки и после его выполнения становится доступным для других заявок.

Существуют реальные мультисервисные системы телекоммуникации, в которых одновременно производится обработка неэластичных и эластичных заявок. Такие системы называются смешанными.

Поскольку в *MRQ* с неэластичными заявками при отсутствии необходимого числа свободных каналов не может быть начато обслуживание заявки, то следует ожидать, что широкополосные заявки будут теряться (блокироваться) чаще, чем узкополосные заявки. Поэтому, в таких системах с целью поддержания показателей качества обслуживания (*Quality of service, QoS*) различных заявок на желаемом уровне необходимо опделить соответствующую стратегию доступа (*Call Admission Control, SAC*).

Любая стратегия доступа в *MRQ* с чистыми потерями задает правила для приема в систему поступившей заявки. Эти правила базируются на альтернативных решениях: принимать поступившую заявку в систему или отказать ей в этом; нужно ли для принятия поступившей заявки прерывать обслуживание некоторой (или некоторых) заявок (заявок) или не делать этого и т.д.

Наиболее простыми *SAC* являются фиксированные стратегии доступа. При использовании таких стратегий названные выше решения принимаются на основе некоторого заранее разработанного регламента и не учитывают текущего состояния системы, причем состояние системы может описываться различными способами. Важно отметить, что с точки зрения удобства практического использования более предпочтительными

являются фиксированные САС, не требующие при их реализации особых затрат на математическое обеспечение.

Однако, реальные MRQ функционируют в условиях существенной неопределенности относительно параметров входящих трафиков, что делает неэффективным во многих случаях использование фиксированных стратегий доступа. Эти обстоятельства сделали актуальными вопросы изучения MRQ с управляемой стратегией доступа. При использовании таких стратегий указанные выше решения, влияющие на судьбу поступившей заявки, принимаются с учетом текущего состояния системы, сложившейся на данный момент ситуации.

Возможны и другие схемы классификации MRQ с чистыми потерями. Однако с учетом круга исследуемых здесь проблем мы ограничимся указанными выше принципами классификации.

Предметом исследования данной части книги являются много-скоростные системы обслуживания с чистыми (явными) потерями (т.е. без очередей).

ИССЛЕДОВАНИЕ МОДЕЛЕЙ МНОГОСКОРОСТНЫХ СИСТЕМ С НЕЭЛАСТИЧНЫМИ ЗАЯВКАМИ

В данной главе разрабатываются эффективные алгоритмы расчета характеристик многоскоростных систем с неэластичными заявками. Предлагаются модифицированные алгоритмы Кауфмана-Робертса для расчета характеристик общей модели многоскоростной системы. Эти алгоритмы предназначены для расчета желаемых характеристик общих моделей MRQ , в которых приняты три наиболее часто используемые стратегии доступа: простая полностью доступная стратегия, полностью доступная стратегия с уравниванием и стратегия, основанная на схеме резервирования каналов. Показывается, что разработанные алгоритмы существенным образом превосходят соответствующие аналоги по степени сложности.

Здесь также дан метод расчета характеристик многоскоростных систем типа Гимпельсона, в которых обслуживаются неэластичные заявки двух типов — узкополосные (n -заявки) и широкополосные (w -заявки). С учетом специфики этих моделей здесь предлагается новая стратегия доступа, основанная на схеме выделения специальной группы каналов для обслуживания w -заявок.

Приводятся результаты вычислительных экспериментов, выполненных с помощью разработанных алгоритмов, а также дан содержательный анализ этих экспериментов.

1.1. Методы расчета общих моделей многоскоростных систем

1.1.1. Полностью доступная стратегия доступа

Подробное описание общей модели исследуемых MRQ состоит в следующем. На вход многоканальной системы, содержащей $N > 1$ каналов, поступает пуассоновский поток различных заявок с суммарной интенсивностью Λ . Каждая вновь поступившая заявка с вероятностью σ_i требует для обслуживания одновременно b_i каналов, $1 \leq b_i \leq N$, $i = 1, \dots, K$, при этом

$$\sum_{i=1}^K \sigma_i = 1. \quad (1.1)$$

Считается, что в момент поступления заявки становится известным количество каналов, затребованных ею для обслуживания.

Тогда с учетом условия (1.1) находим, что на вход N -канальной системы обслуживания поступают K типов пуассоновских потоков заявок, при этом интенсивность i -го потока равна $\lambda_i := \Lambda \sigma_i$, $i = 1, \dots, K$; при этом заявки i -го типа (i -заявки) требуют одновременно b_i каналов, $i = 1, \dots, K$.

В настоящем подразделе предполагается, что в системе принята полnodоступная стратегия доступа в каналы (*Complete Sharing, CS*). Согласно данной стратегии доступа, все потоки заявок равноправно используют каналы, т.е. для обслуживания i -заявки могут быть выделены любые b_i свободных каналов, $i = 1, \dots, K$. При этом все каналы одновременно начинают и завершают обслуживание данной заявки. Если в момент поступления заявки любого типа нет достаточного числа свободных каналов, то она с вероятностью 1 получает отказ (блокируется) и в дальнейшем не влияет на работу системы. Схема работы системы при использовании *CS*-стратегии доступа показана на рис. 1.1.

Время занятия каналов i -заявкой является случайной величиной, подчиненной экспоненциальному закону распределения с параметром μ_i , $i = 1, \dots, K$.

Поскольку входящие потоки являются пуассоновскими и времена их обслуживания подчинены экспоненциальному закону, то состояние системы в произвольный момент времени может быть описано с помощью K -мерного вектора $\mathbf{n} = (n_1, \dots, n_K)$, где n_i означает число i -заявок в системе, $i = 1, \dots, K$. Иными словами, функционирование данной *MRQ* описывается K -мерной цепью Маркова со следующим фазовым пространством состояний:

$$S_{CS} := \{\mathbf{n}: n_i = 0, \dots, [N/b_i], (\mathbf{n}, \mathbf{b}) \leq N\}, \quad (1.2)$$

где $\mathbf{b} = (b_1, \dots, b_K)$, $[x]$ — целая часть x , (\mathbf{n}, \mathbf{b}) означает скалярное произведение векторов \mathbf{n} и \mathbf{b} :

$$(\mathbf{n}, \mathbf{b}) := \sum_{i=1}^K n_i b_i. \quad (1.3)$$

Условие $(\mathbf{n}, \mathbf{b}) \leq N$ в определении (1.2) означает, что в любом состоянии $\mathbf{n} \in S_{CS}$ число занятых каналов, которое определяется из (1.3), не может превосходить N (т.е. общего числа каналов системы).

Вид ФПС модели (1.2) показан на рис. 1.2 (а).

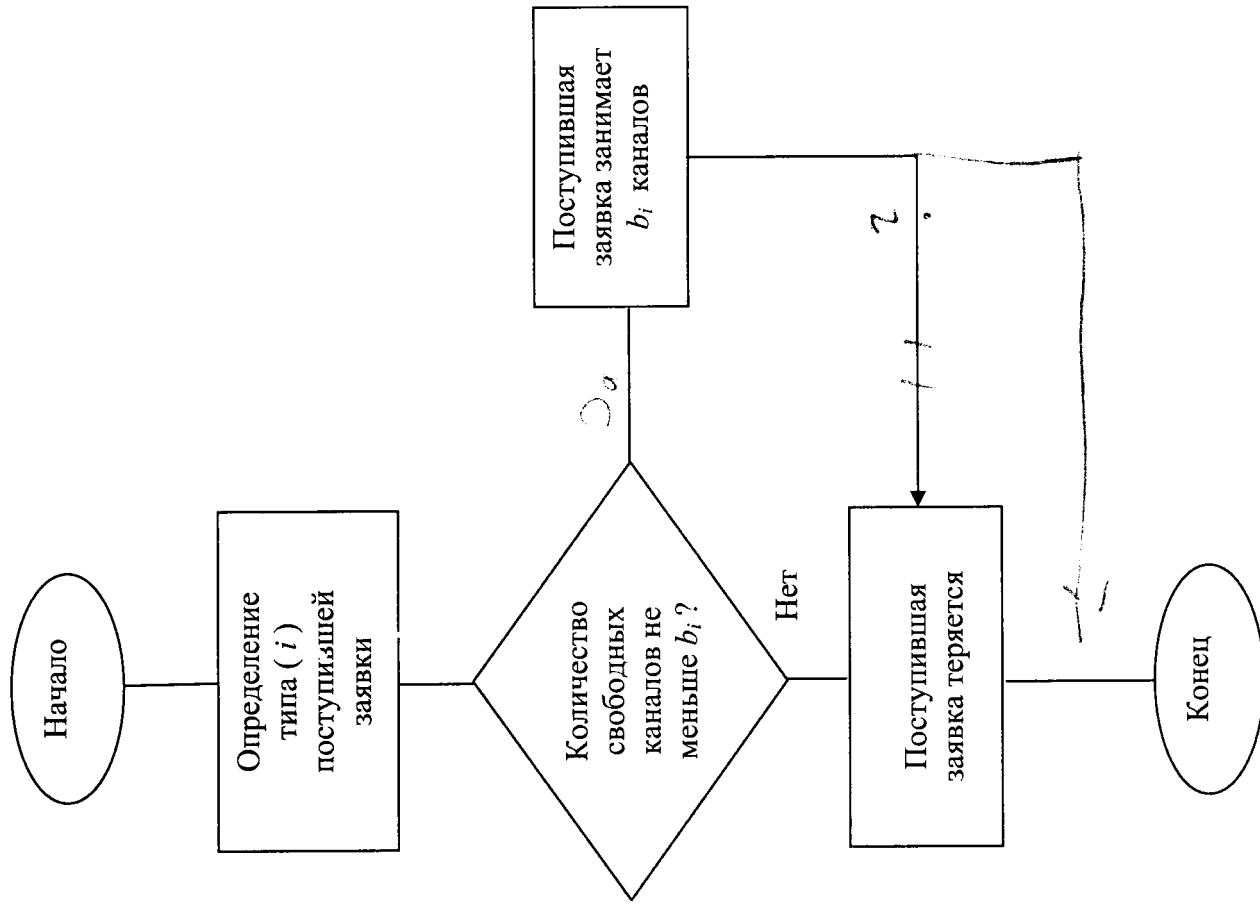


Рис. 1.1. Схема *CS*-стратегии доступа в каналы

Переходы между состояниями \mathbf{n} , $\mathbf{n}' \in S_{CS}$ происходят лишь в момент поступления заявок и ухода их из системы после завершения обслуживания. С учетом этого элементы производящей матрицы данной многомерной ЦМ, $q_{CS}(\mathbf{n}, \mathbf{n}')$, $\mathbf{n}, \mathbf{n}' \in S_{CS}$ определяются из следующих соотношений:

$$q_{CS}(\mathbf{n}, \mathbf{n}') = \begin{cases} \lambda_i, & \text{если } \mathbf{n}' = \mathbf{n} + \mathbf{e}_i, \\ n_i \mu_i, & \text{если } \mathbf{n}' = \mathbf{n} - \mathbf{e}_i, \\ 0 & \text{в остальных случаях,} \end{cases} \quad (1.4)$$

где \mathbf{e}_i — i -й орт вектор K -мерного евклидова пространства, $i = 1, \dots, K$.
Граф переходов между состояниями показан на рис. 1.2 (б).

При любых положительных значениях параметров входящих графов (т.е. $\lambda_i, \mu_i, i = 1, \dots, K$) все состояния $\mathbf{n} \in S_{CS}$ являются сообщающимися, и следовательно, система является эргодической.

Стационарную вероятность состояния $\mathbf{n} \in S_{CS}$ обозначим через $p(\mathbf{n})$. Известно, что стационарное распределение модели находится из системы уравнений равновесия. Она составляется с помощью соотношений (1.4):

$$\left(\sum_{i=1}^K \lambda_i I(f(\mathbf{n}) \geq b_i) + \sum_{i=1}^K n_i \mu_i \right) p(\mathbf{n}) = \sum_{i=1}^K \lambda_i p(\mathbf{n} - \mathbf{e}_i) I(n_i > 0) + \sum_{i=1}^K (n_i + 1) \mu_i p(\mathbf{n} + \mathbf{e}_i) I(f(\mathbf{n}) \geq b_i), \quad \mathbf{n} \in S_{CS}; \quad (1.5)$$

$$\sum_{\mathbf{n} \in S_{CS}} p(\mathbf{n}) = 1, \quad (1.6)$$

где $f(\mathbf{n}) := N - (\mathbf{n}, \mathbf{b})$ — число свободных каналов в состоянии $\mathbf{n} \in S_{CS}$.

В указанной выше СУР уравнение (1.6) называется условием нормировки.

Стационарное распределение системы $(p(\mathbf{n}); \mathbf{n} \in S_{CS})$, при CS-стратегии доступа имеет мультипликативный вид:

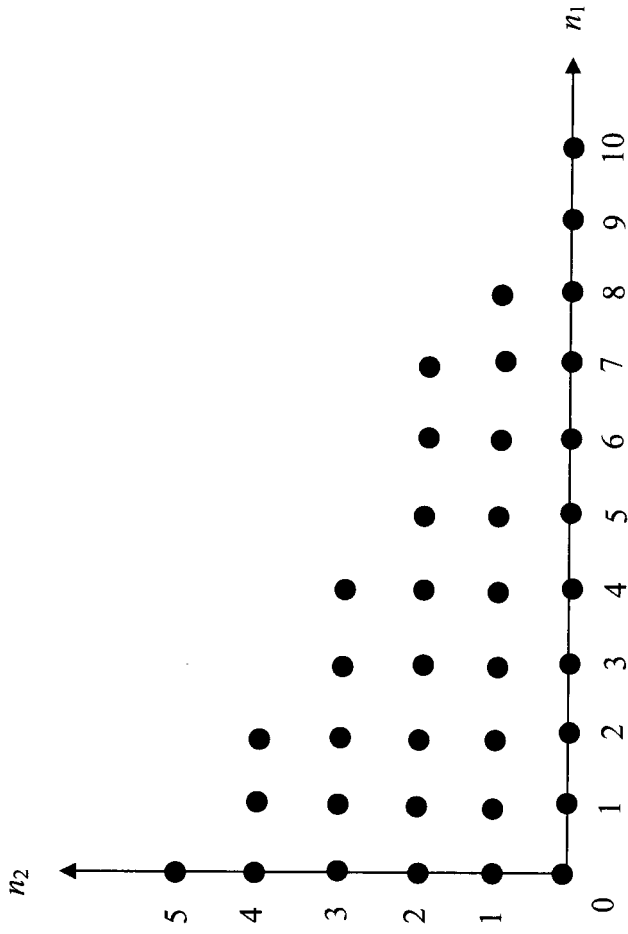
$$p(\mathbf{n}) = G^{-1}(N, K) \prod_{i=1}^K \frac{V_i^{n_i}}{n_i!}, \quad (1.7)$$

$$\text{где } V_i := \lambda_i / \mu_i, \quad G(N, K) := \sum_{\mathbf{n} \in S_{CS}} \prod_{i=1}^K \frac{V_i^{n_i}}{n_i!}.$$

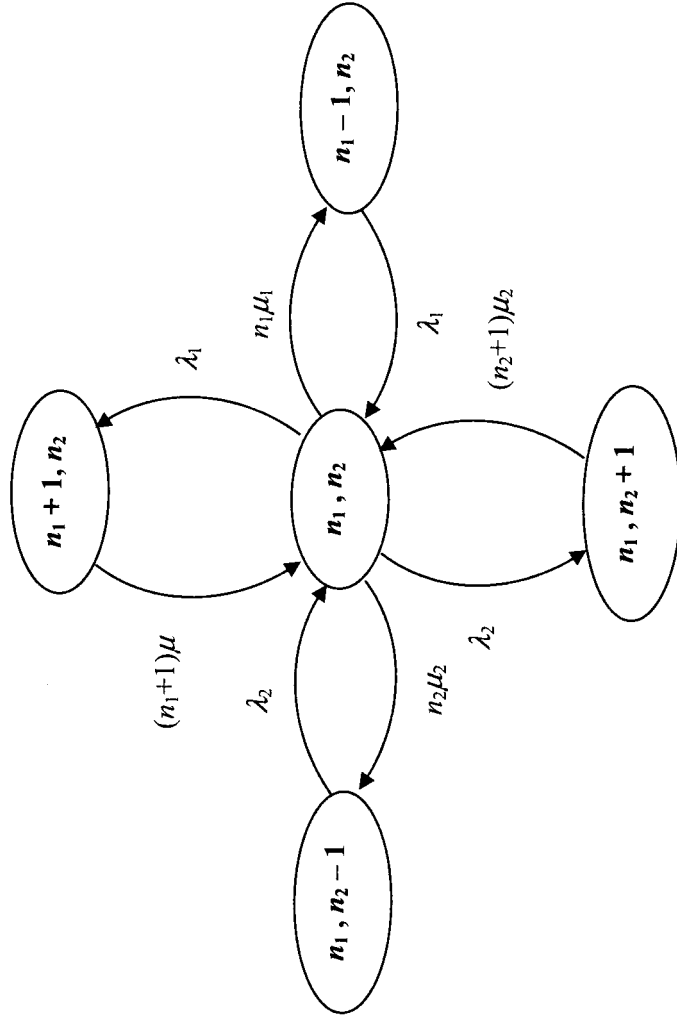
В формуле (1.7) параметр $G(N, K)$ называется нормирующей константой над ФПС (1.2). Из этой же формулы видно, что

$$G^{-1}(N, K) = p(\mathbf{0}),$$

где $\mathbf{0}$ — K -мерный нулевой вектор.



(а)



(б)

Рис. 1.2. Вид ФПС с CS-стратегией доступа при $N = 10, K = 2, b = (1, 2)$ (а) и граф переходов между ее состояниями (б)

Вычисление стационарного распределения ($p(\mathbf{n})$; $\mathbf{n} \in S_{CS}$) является достаточным условием для нахождения желаемых показателей QoS разнотипных заявок, а также показателей эффективности использования каналов системы (утилизации каналов).

Основным показателем QoS заявок i -го типа является стационарная вероятность их блокировки (потери). Эту вероятность обозначим через PB_i (Probability of Blocking, PB), $i = 1, \dots, K$. Эти показатели QoS вычисляются так:

$$PB_i(CS) := \sum_{r=0}^{b_i-1} p(\mathbf{n}), i = 1, \dots, K. \quad (1.8)$$

Замечание 1.1. Из последней формулы (1.8) получаем, что если $b_i = b_j$ для некоторых i, j ; $i \neq j$, то $PB_i(CS) = PB_j(CS)$ для любых значений нагрузочных параметров модели, т.е. если заявки двух трафиков требуют одинакового количества каналов, то вероятности их потери равны друг другу.

Исходя из (1.8) легко можно вычислить другие показатели QoS разнотипных заявок. Так, средняя интенсивность обслуживания i -заявок, RS_i (Rate of Served Calls), $i = 1, \dots, K$, определяется так:

$$RS_i(CS) := \lambda_i(1 - PB_i(CS)), i = 1, \dots, K. \quad (1.9)$$

Основным показателем эффективности использования каналов является среднее число занятых каналов, обозначаемое через \tilde{N} . Этот параметр определяется так:

$$\tilde{N}(CS) := \sum_{\mathbf{n} \in S_{CS}} (\mathbf{n}, \mathbf{b}) p(\mathbf{n}). \quad (1.10)$$

Из (1.10) легко определить коэффициент утилизации каналов (N_0):

$$N_0(CS) := \tilde{N}(CS) / N. \quad (1.11)$$

Прямое вычисление нормирующей константы $G(N, K)$ и дальнейшее определение стационарного распределения ($p(\mathbf{n})$; $\mathbf{n} \in S_{CS}$) из (1.7) и (1.8) при больших размерах ФПС (1.2) связано с огромными вычислительными трудностями. Дело в том, что с ростом N и K размерность ФПС (1.2) растет экспоненциально. Так, например, при $N = 20$ и $\mathbf{b} = (1, 2, 3, \dots, 10, 12, 14, \dots, 20)$ размерность ФПС равна 2544, а при $N = 30$ и $\mathbf{b} = (1, 2, 3, \dots, 10, 12, 14, \dots, 30)$ эта величина равна 23960.

Для преодоления указанных трудностей ниже предлагается следующий эффективный алгоритм.

Рассматривается следующее расщепление ФПС (1.2):

$$S_{CS}^r = \bigcup_{r=0}^N S_{CS}^r, S_{CS}^r \cap S_{CS}^{r'} = \emptyset, r \neq r', \quad (1.12)$$

где $S_{CS}^r := \{\mathbf{n} \in S_{CS} : (\mathbf{n}, \mathbf{b}) = r\}$, т.е. S_{CS}^r содержит те состояния $\mathbf{n} \in S_{CS}$, в которых общее число занятых каналов равно r , $r = 0, \dots, N$.

Классы состояний S_{CS}^r описываются одним укрупненным состоянием $\langle r \rangle$, а его стационарная вероятность обозначается через $\pi(\langle r \rangle)$.

Граф переходов между укрупненными состояниями $\langle r \rangle$, $\langle r' \rangle \in \tilde{S}_{CS}$, где $\tilde{S}_{CS} := \{\langle r \rangle : r = 0, \dots, N\}$, показан на рис. 1.3.

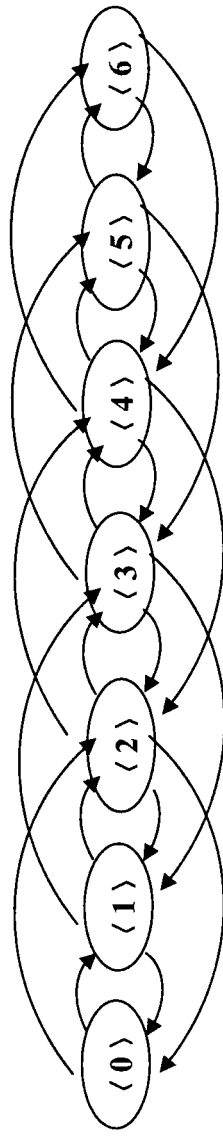


Рис. 1.3. Граф переходов между укрупненными состояниями ФПС \tilde{S}_{CS}

при $N = 6$, $K = 2$, $\mathbf{b} = (1, 2)$

Интенсивности переходов между укрупненными состояниями $\langle r \rangle$, $\langle r' \rangle \in \tilde{S}_{CS}$, обозначаемые через $\tilde{q}(\langle r \rangle, \langle r' \rangle)$, определяются так:

$$\tilde{q}(\langle r \rangle, \langle r' \rangle) = \sum_{\substack{\mathbf{n} \in S_{CS}^r \\ \mathbf{n}' \in S_{CS}^{r'}}} q(\mathbf{n}, \mathbf{n}') p(\mathbf{n}). \quad (1.13)$$

Как видно из формулы (1.13), определение стационарного распределения укрупненной модели ($\pi(\langle r \rangle)$; $\langle r \rangle \in \tilde{S}_{CS}$) не представляется возможным без предварительного вычисления стационарного распределения исходной модели ($p(\mathbf{n})$; $\mathbf{n} \in S_{CS}$). При этом очевидно, что для стационарного распределения исходной модели удовлетворяется следующее соотношение:

$$\pi(\langle r \rangle) = \sum_{\mathbf{n} \in S_{CS}^r} p(\mathbf{n}), \sum_{r=0}^N \pi(\langle r \rangle) = 1.$$

Стационарное распределение ($\pi(\langle r \rangle)$; $r = 0, \dots, N$) удовлетворяет следующей системе уравнений:

$$\sum_{i=1}^K v_i \pi(\langle r - b_i \rangle) = r \pi(\langle r \rangle), r = 1, \dots, N. \quad (1.14)$$

$$\sum_{r=0}^N \pi(\langle r \rangle) = 1, \quad (1.15)$$

где $\pi(\langle x \rangle) = 0$ при $x < 0$.

Соотношения (1.14), (1.15) позволяют вычислить стационарное распределение исходной модели, с помощью которого желаемые характеристики системы (1.8), (1.10) определяются так:

$$PB_i(CS) = \sum_{j=0}^{b_i-1} \pi(< N - j >), i = 1, \dots, K, \quad (1.16)$$

$$\tilde{N}(CS) = \sum_{r=1}^N r \pi(< r >). \quad (1.17)$$

Определяем вначале следующие множества:

$$A(r) := \{i : i \in \overline{1, K} : i \text{ - заявки требуют } r \text{ каналов}\}, r = 1, \dots, N. \quad (1.18)$$

С помощью множеств $A(r)$ в (1.18) осуществляется процедура укрупнения потоков по числу требуемых каналов обслуживания, т.е. все потоки заявок, требующие одинаковое число каналов, укрупняются в один поток.

Далее находим укрупненные нагрузки:

$$\hat{v}_r = \begin{cases} \sum_{i \in A(r)} r_i, & \text{если } A(r) \neq \emptyset, \\ 0, & \text{если } A(r) = \emptyset. \end{cases} \quad (1.19)$$

С учетом (1.18) и (1.19) находим, что система уравнений (1.14), (1.15) имеет следующую расширенную матрицу:

$$\begin{pmatrix} \hat{v}_1 & -1 & 0 & \dots & 0 & 0 & 0 \\ 2\hat{v}_2 & \hat{v}_1 & -2 & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ N\hat{v}_N & (N-1)\hat{v}_{N-1} & (N-2)\hat{v}_{N-2} & \dots & \hat{v}_1 & -N & 0 \\ 1 & 1 & 1 & \dots & 1 & 1 & 1 \end{pmatrix}.$$

Отсюда легко получается, что

$$\pi(< r >) = g_r \pi(< 0 >), r = 1, \dots, N. \quad (1.20)$$

где g_r вычисляется с помощью следующих одномерных рекуррентных формул:

$$\begin{aligned} g_0 &:= 1 \\ g_r &= \frac{1}{r} \sum_{i=1}^r i \hat{v}_i g_{r-i}, r = \overline{1, N}. \end{aligned} \quad (1.21)$$

Следовательно, с учетом (1.16), (1.17) и (1.20), (1.21) после определенных математических преобразований находим, что характеристики модели (1.8) и (1.10) вычисляются так:

$$PB_i(CS) = \left(\sum_{j=N-r+1}^N g_j \right) / \left(\sum_{j=0}^N g_j \right), i \in A(r), \quad (1.22)$$

$$\tilde{N}(CS) = \left(\sum_{i=1}^N i g_i \right) / \left(\sum_{i=0}^N g_i \right). \quad (1.23)$$

Далее с помощью (1.9) и (1.11) вычисляется средняя интенсивность обслуженных заявок каждого типа, а также коэффициент утилизации каналов системы.

Резюмируя можно предложить следующий модифицированный алгоритм Кауфмана-Робертса для расчета характеристик общей модели многократных систем обслуживания.

Шаг 1: Укрупнение потоков.

1.1. Определение множеств $A(r)$ с помощью (1.18).

1.2. Нахождение укрупненных нагрузок \hat{v}_r из (1.19).

Шаг 2: Расчет характеристик системы.

2.1. Вычисление параметров $g_r, r = 0, \dots, N$ из системы соотношений (1.21).

2.2. Вычисление характеристик системы $PB_i(CS), i = 1, \dots, K$ и $\tilde{N}(CS)$ из формул (1.22) и (1.23), соответственно.

Замечание 1.2. Основное отличие данного алгоритма от аналогичных алгоритмов состоит в том, что он имеет меньшую сложность, т.к. здесь в отличие от известных алгоритмов его вычислительная сложность не зависит от общего числа типов заявок (т.е. от K) и оценивается как $O(N)$. Такая инвариантность обеспечивается на шаге 1 за счет укрупнения потоков по числу требуемых каналов.

1.1.2. Полностью доступная стратегия доступа с уравниванием

В предыдущем параграфе предполагалось, что все потоки заявок равноправно используют каналы системы. Однако, при использовании CS-стратегии доступа следует ожидать, что с ростом ширины заявок вероятности их потерь также растут, т.е.

$$\text{если } b_i > b_j, \text{ то } PB_i(CS) \geq PB_j(CS), i, j = 1, \dots, K. \quad (1.24)$$

Справедливость соотношений (1.24) непосредственно следует из (1.8).

Отсюда возникает проблема организации абсолютно справедливого обслуживания разнотипных заявок в многократных системах в смысле выравнивания вероятностей блокировок (потерь) разнотипных заявок. Это означает, что следует организовывать такое обслуживание разнотипных заявок, чтобы их шансы принятия в каналы были одинаковыми, т.е. требуется найти такую стратегию доступа, чтобы

$$PB_i = PB_j \quad \forall i, j = 1, \dots, K. \quad (1.25)$$

Самый простой способ обеспечения выполнения условия (1.25) состоит в выравнивании ширины всех потоков заявок. Эта стратегия доступа называется полностью доступной стратегией доступа с выравниванием (*Complete Sharing with Equalization, CSE*).

Указанная CSE-стратегия доступа определяется так: поступающая заявка любого типа принимается для обслуживания, если в этот момент число свободных каналов больше или равно b , где $b = \max \{b_j\}$.

Схема алгоритма работы CSE-стратегии доступа показана на рис. 1.4.

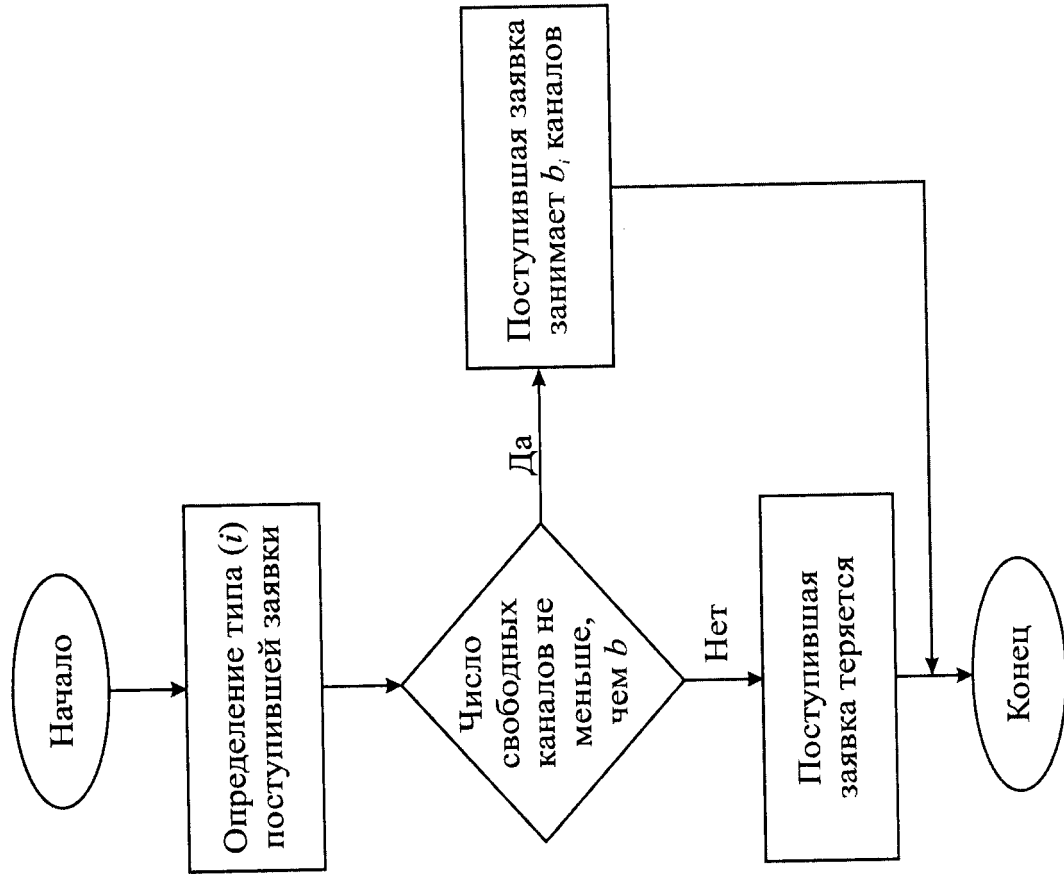


Рис.1.4. Схема CSE-стратегии доступа в каналы.

Состояние системы при использовании данной стратегии доступа, также как и в случае CS-стратегии доступа, в произвольный момент задается с помощью K -мерного вектора $\mathbf{n} = (n_1, \dots, n_K)$, где n_i означает число i -заявок в системе, $i = 1, \dots, K$. Однако, при использовании CSE-стратегии доступа максимально возможное число i -заявок в системе

равно $\left[\frac{N-b}{b_i} \right] + 1$, т.е. ФПС модели в данном случае определяется так:

$$S_{CSE} := \left\{ \mathbf{n} : n_i = 0, \left[\frac{N-b}{b_i} \right] + 1, (\mathbf{n}, \mathbf{b}) \leq N \right\} \quad (1.26)$$

Вид ФПС модели (1.26) показан на рис.1.5 (а).

Элементы производящей матрицы соответствующей цепи Маркова с ФПС (1.26) при использовании CSE-стратегии доступа определяются следующим образом:

$$q_{CSE}(\mathbf{n}, \mathbf{n}') = \begin{cases} \lambda_i I(f(\mathbf{n}) \geq b), & \text{если } \mathbf{n}' = \mathbf{n} + \mathbf{e}_i, \\ n_i \mu_i, & \text{если } \mathbf{n}' = \mathbf{n} - \mathbf{e}_i, \\ 0 & \text{в остальных случаях.} \end{cases} \quad (1.27)$$

Граф переходов между состояниями системы при использовании CSE-стратегии доступа показан на рис.1.5 (б).

Поскольку рассматриваемая ЦМ является конечномерной и все ее состояния являются сообщающимися, то в ней существует стационарный режим при любых положительных значениях параметров входящих трафиков.

Система уравнений равновесия для данной ЦМ составляется с помощью соотношений (1.27). Она имеет следующий вид:

$$\begin{aligned} & \left(\sum_{i=1}^K \lambda_i I(f(\mathbf{n}) \geq b) \right) \cdot \sum_{i=1}^K n_i \mu_i p(\mathbf{n}) = \\ & = \sum_{i=1}^K \lambda_i p(\mathbf{n} - \mathbf{e}_i) I(f(\mathbf{n} - \mathbf{e}_i) \geq b) + \\ & + \sum_{i=1}^K (\mu_i + 1) \mu_i p(\mathbf{n} + \mathbf{e}_i) I(f(\mathbf{n}) \geq b), \quad \mathbf{n} \in S_{CSE}; \end{aligned} \quad (1.28)$$

$$\sum_{\mathbf{n} \in S_{CSE}} p(\mathbf{n}) = 1. \quad (1.29)$$

Основной показатель QoS разнотипных заявок – вероятность их потерь, является одинаковой для всех типов заявок. Обозначим этот показатель через $PB(CSE)$. Он определяется через стационарное распределение следующим образом:

$$PB(CSE) := \sum_{f(\mathbf{n})=0}^{b-1} p(\mathbf{n}) \quad (1.30)$$

Средняя интенсивность обслуженных i -заявок ($RS_i(CSE)$), $i = 1, \dots, K$, определяется так:

$$RS_i(CSE) = \lambda_i (1 - PB(CSE)), \quad i = 1, \dots, K. \quad (1.31)$$

Среднее число занятых каналов ($\tilde{N}(CSE)$) и коэффициент утилизации каналов ($N_0(CSE)$) определяются аналогично (1.10) и (1.11), соответственно:

$$\tilde{N}(CSE) := \sum_{\mathbf{n} \in S_{CSE}} (\mathbf{n}, \mathbf{b}) p(\mathbf{n}), \quad N_0(CSE) := \tilde{N}(CSE)/N. \quad (1.32)$$

Из формул (1.30) – (1.32) видно, что необходимым этапом вычисления желанных характеристик модели является нахождение стационарного распределения ($p(\mathbf{n}) : \mathbf{n} \in S_{CSE}$) модели из СУР (1.28), (1.29).

В отличие от модели с CS-стратегией доступа, для модели с CSE-стратегией доступа не существует мультипликативного решения СУР (1.28), (1.29) (см. формулы (1.7)). Доказательство данного факта вытекает из необратимости рассматриваемой ЦМ [11]. Согласно данному свойству, необходимым условием существования мультипликативного представления стационарного распределения ЦМ является следующий легко проверяемый факт: если существует переход из состояния \mathbf{n} в состояние \mathbf{n}' , то должен существовать обратный переход из состояния \mathbf{n}' в состояние \mathbf{n} .

Действительно, с использованием (1.27) заключаем, что существует переход из состояния $\mathbf{n} + \mathbf{e}_i$ в состояние \mathbf{n} , однако обратный переход существует лишь тогда, когда $f(\mathbf{n}) \geq b$, т.е. при использовании CSE-стратегии доступа не существует мультипликативного решения СУР (1.28), (1.29). Это обстоятельство существенным образом затрудняет поиск решения СУР (1.28), (1.29) при больших размерностях ФПС (1.26), и таким образом, вычисление характеристик модели (1.30), (1.31) при использовании CSE-стратегии доступа в каналы.

Для преодоления указанных трудностей можно использовать подход, предложенный в предыдущем параграфе при исследовании модели с CS-стратегией доступа в каналы.

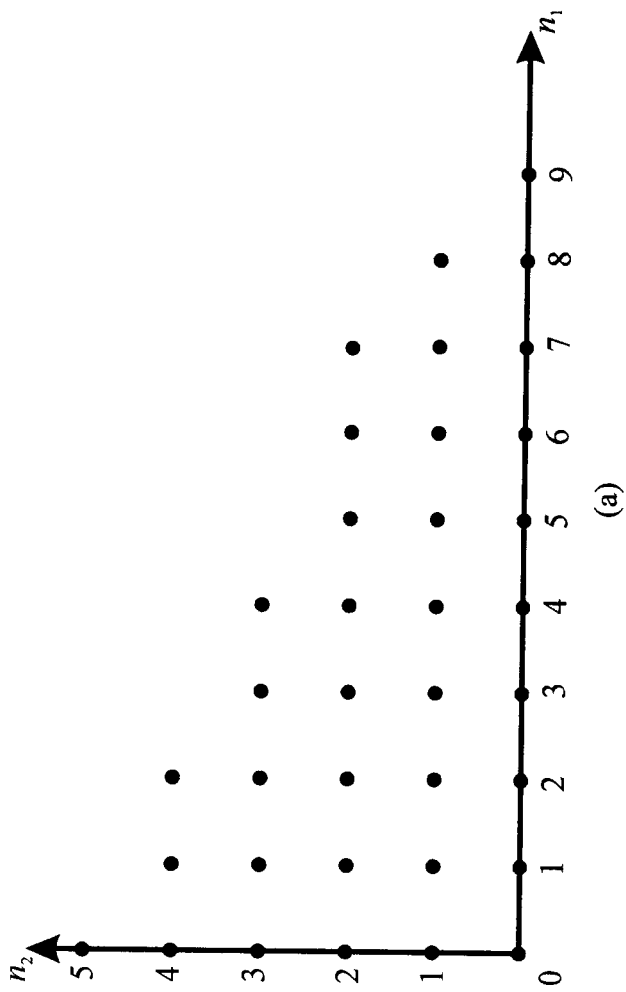
Аналогично (1.12) рассматривается следующее расщепление ФПС (1.26):

$$S_{CSE} = \bigcup_{i=0}^N S_{CS}^i, \quad S_{CSE}^i \cap S_{CSE}^{i'} = \emptyset, \quad i \neq i', \quad (1.33)$$

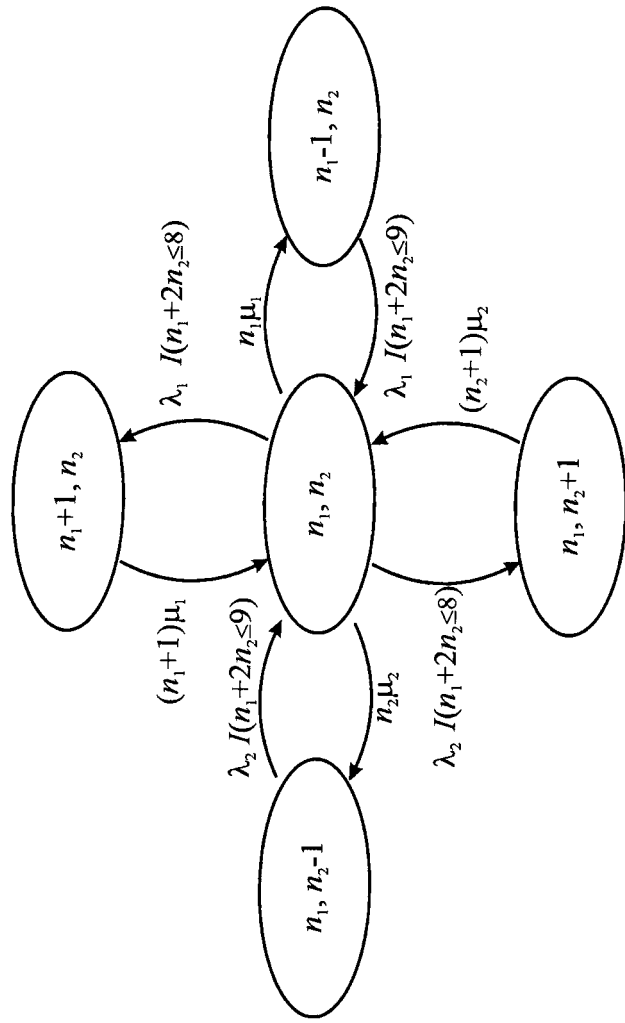
где $S_{CSE}^i := \{\mathbf{n} \in S : (\mathbf{n}, \mathbf{b}) = i\}$, т.е. класс состояний S_i содержит те микросостояния $\mathbf{n} \in S$, в которых число занятых каналов равно $i, i = 0, 1, 2, \dots, N$.

Далее все микросостояния из класса S_i склеиваются в одно укрупненное состояние $\langle i \rangle$, и на исходном ФПС (1.26) строится функция укрупнения $U : S \rightarrow \hat{S}_{CSE}, \hat{S}_{CSE} := \{\langle 0 \rangle, \langle 1 \rangle, \dots, \langle N \rangle\}$, которая определяется так:

$$U(\mathbf{n}) = \langle i \rangle, \text{ если } \mathbf{n} \in S_{CSE}^i. \quad (1.34)$$



(а)



(б)

Рис.1.5. Вид ФПС модели с CSE-стратегией доступа при $N = 10, K = 2, b = (1, 2)$ (а) и граф переходов между ее состояниями (б)

Граф переходов между укрупненными состояниями $\langle r \rangle, \langle r' \rangle \in \tilde{S}_{CSE}$, показан на рис.1.6.

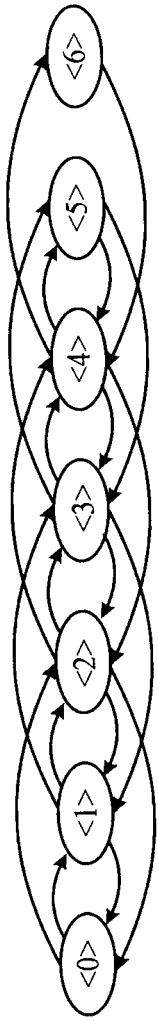


Рис.1.6. Граф переходов между укрупненными состояниями ФПС \tilde{S}_{CSE} при $N=6, K=2, b=(1,2)$

Следовательно, с учетом (1.30) и (1.32) находим, что имеют место следующие соотношения:

$$PB(CSE) = \sum_{i=N-b+1}^N \pi(\langle i \rangle), \quad (1.35)$$

$$\tilde{N}(CSE) = \sum_{j=1}^N j \pi(\langle j \rangle). \quad (1.36)$$

Стационарное распределение укрупненной модели ($\pi(\langle r \rangle): r=0, \dots, N$) удовлетворяет следующим соотношениям:

$$\pi(\langle i \rangle) = \frac{1}{i} \sum_{j=1}^K v_j b_j \pi(\langle i - b_j \rangle) F_j(i - b_j), \quad i=1, \dots, N; \quad (1.37)$$

$$\sum_{i=0}^N \pi(\langle i \rangle) = 1, \quad (1.38)$$

где $v_i = \lambda_i / \mu_i$, $\pi(x) = 0$, если $x < 0$.

$$F_j(i - b_j) = \begin{cases} 1, & \text{если } i - b_j \leq N - b, \\ 0 & \text{в остальных случаях.} \end{cases} \quad (1.39)$$

На основе системы уравнений (1.37) – (1.39) разработаны следующие формулы для расчета (1.30) и (1.32):

$$PB(CSE) = \left(\sum_{j=N-b+1}^N g_j / \sum_{j=0}^N g_j \right), \quad (1.40)$$

$$\tilde{N}(CSE) = \left(\sum_{i=1}^N i g_i / \sum_{i=0}^N g_i \right), \quad (1.41)$$

где

$$g_0 := 1, \quad g_l = \frac{1}{l} \sum_{i=1}^l i v_i g_{l-i} F_i(l-i), \quad l=1, \dots, N. \quad (1.42)$$

Замечание 1.4. Предложенные в работе [13] формулы содержат ошибку, так как в них аналогичная формула для вычисления функций $F_j^*(i - b_j)$ определяется так:

$$F_j^*(i - b_j) = \begin{cases} 1, & \text{если } i \leq N - b \\ 0 & \text{в остальных случаях.} \end{cases} \quad (1.43)$$

Однако, если функцию (1.39) определить с помощью формулы (1.43), то из (1.37) получается, что $\pi(\langle i \rangle) = 0$ для всех $i > N - b$. Но последнее является неверным, так как $\pi(\langle i \rangle) > 0$ для всех $i = 0, \dots, N$.

1.1.3. Стратегия доступа с резервированием каналов

Введенные в предыдущих параграфах CS- и CSE-стратегии доступа не позволяют удовлетворить слишком разные требования разнотипных трафиков, предъявляемые к показателям качества их обслуживания, а также эффективно использовать пропускные способности каналов системы. В связи с этим применяются более сложные стратегии доступа в каналы.

Среди этих стратегий особое место занимают стратегии резервирования каналов (*Trunk Reservation, TR*). Согласно им заявки данного типа принимаются лишь тогда, когда число каналов, остающихся свободными после принятия данной заявки, превышает некоторую величину. Эта величина называется параметром резервирования для заявок данного типа. При этом его значение может либо быть постоянной величиной, либо быть зависимой от типа заявки.

Основной целью введения стратегий резервирования является защита широкополосных заявок, так как при CS-стратегии доступа заявки такого типа часто получают отказ в обслуживании по сравнению с узкополосными заявками.

Здесь предлагается эффективный алгоритм расчета желаемых характеристик общей модели MRQ со стратегией резервирования каналов, в которой параметры резервирования зависят от типа заявок. Показывает, что данная стратегия является обобщением двух предыдущих стратегий доступа – CS и CSE.

Рассматриваемая стратегия резервирования каналов определяется следующим образом. Если в момент поступления заявки i -го типа число свободных каналов не меньше $b_i + r_i$, где $0 \leq r_i \leq N - b_i$, то она принимается для обслуживания; в противном случае поступающая заявка блокируется (теряется) с вероятностью 1. Параметр r_i называется параметром резервирования каналов для заявок i -го типа, $i = 1, \dots, K$.

Будем считать, что если $b_i = b_j$, то $r_i = r_j$, $i \neq j$, т.е. заявки, требующие одинакового числа каналов для обслуживания, имеют также одинаковые параметры резервирования каналов.

Схема работы TR-стратегии доступа в каналы показана на рис.1.7.

Замечание 1.5. Если $r_i = 0$ для всех $i = 1, \dots, K$, то получается модель MRQ с полностью доступной стратегией (CS-стратегией) доступа.

Замечание 1.6. Если $r_i = b - b_i$ для всех $i = 1, \dots, K$, где $b := \max\{b_i\}$, то получается модель MRQ с полностью доступной стратегией с выравниванием.

Функционирование данной модели также описывается K-мерной цепью Маркова с состояниями $\mathbf{n} = (n_1, \dots, n_K)$, где n_i означает число заявок i-го типа в системе. Можно показать, что при использовании TR-стратегии доступа максимально возможное число i-заявок в системе

$$\text{равно } \left[\frac{N - r_i - 1}{b_i} \right] + 1.$$

Следовательно, фазовое пространство состояний соответствующей ЦМ задается так:

$$S_{TR} = \left\{ \mathbf{n} : n_i = 0, 1, \dots, \left[\frac{N - r_i - 1}{b_i} \right] + 1, i = \overline{1, K}; (\mathbf{n}, \mathbf{b}) \leq N \right\}. \quad (1.44)$$

Элементы производящей матрицы данной ЦМ с пространством состояний (1.44) при использовании TR-стратегии доступа в каналы определяются так:

$$q_{TR}(\mathbf{n}, \mathbf{n}') = \begin{cases} \lambda_i I(f(\mathbf{n}) \geq b_i + r_i), & \text{если } \mathbf{n}' = \mathbf{n} + \mathbf{e}_i \\ n_i \mu_i, & \text{если } \mathbf{n}' = \mathbf{n} + \mathbf{e}_i \\ 0 & \text{в остальных случаях.} \end{cases} \quad (1.45)$$

Вид ФПС модели (1.44) и граф переходов между ее состояниями показаны на рис.1.8. (а) и (б), соответственно.

Стационарные вероятности состояний $(p(\mathbf{n}) : \mathbf{n} \in S_{TR})$ удовлетворяют следующей системе уравнений (существование стационарного режима вытекает из неприводимости и апериодичности данной конечномерной цепи Маркова):

$$\begin{aligned} & \left(\sum_{i=1}^K \lambda_i I(f(\mathbf{n}) \geq b_i + r_i) + \sum_{i=1}^K n_i \mu_i \right) p(\mathbf{n}) = \\ & = \sum_{i=1}^K \lambda_i p(\mathbf{n} - \mathbf{e}_i) I(f(\mathbf{n} - \mathbf{e}_i) \geq b_i + r_i) + \\ & + \sum_{i=1}^K (n_i + 1) \mu_i p(\mathbf{n} + \mathbf{e}_i) I(f(\mathbf{n}) \geq b_i + r_i), \quad \mathbf{n} \in S_{TR}; \end{aligned} \quad (1.46)$$

$$\sum_{\mathbf{n} \in S_{TR}} p(\mathbf{n}) = 1. \quad (1.47)$$

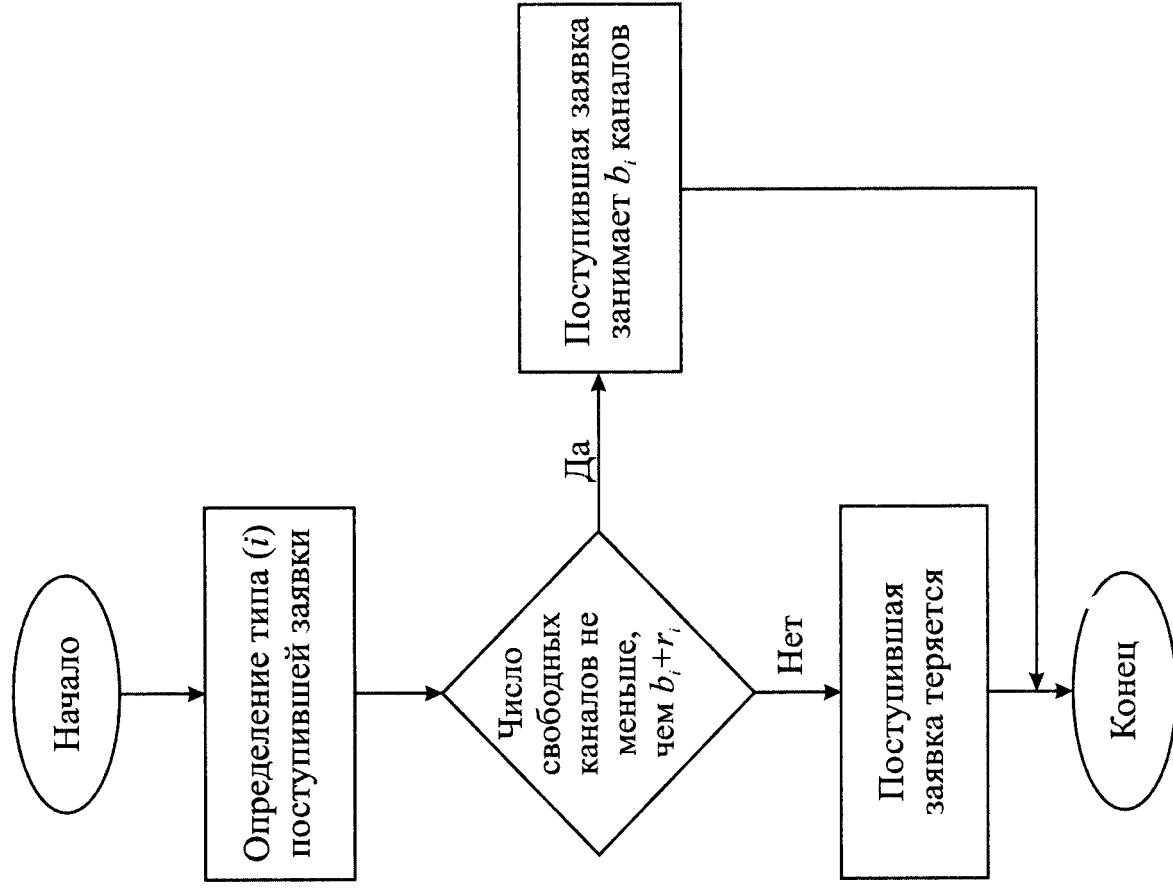


Рис.1.7. Блок-схема TR-стратегии доступа в каналы.

Основные характеристики модели – вероятность блокировки заявок i-го типа $(PB_i(TR))$ и среднее число занятых каналов $(\tilde{N}(TR))$, определяются через стационарное распределение системы:

$$PB_i(TR) = \sum_{f(\mathbf{n})=0}^{b_i + n_i - 1} p(\mathbf{n}), i = \overline{1, K}, \quad (1.48)$$

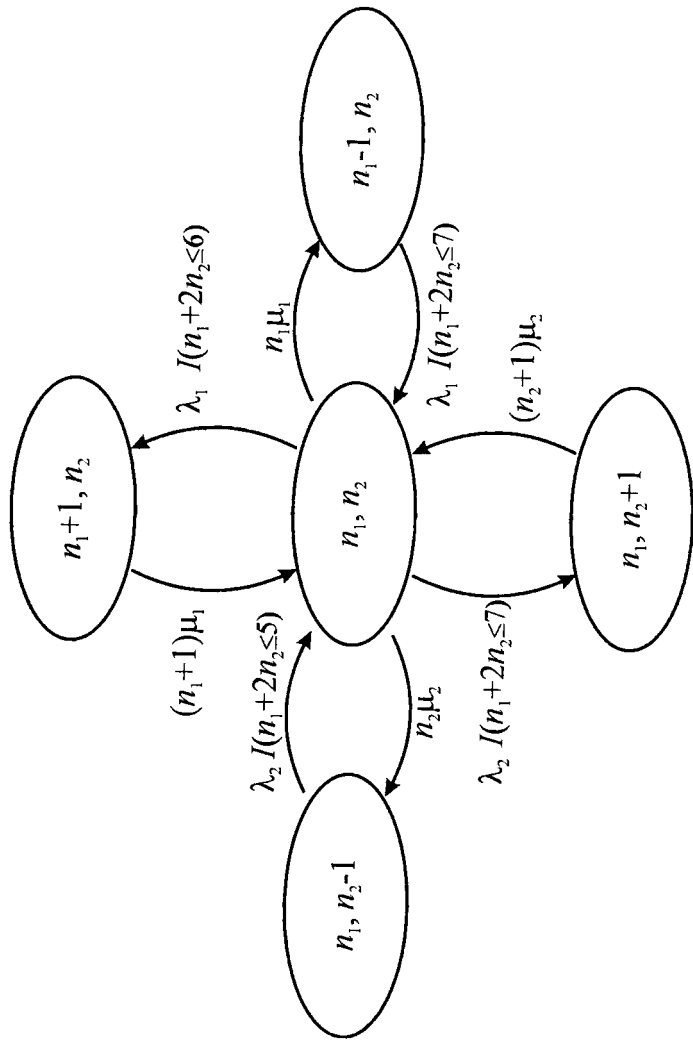
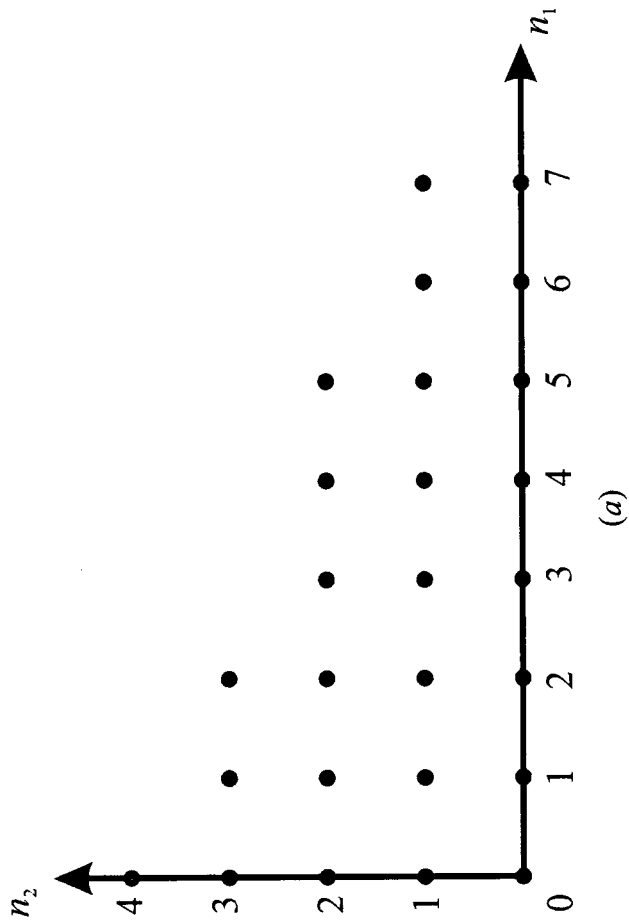
$$\tilde{N}(TR) = \sum_{\mathbf{n} \in S_{TR}} (\mathbf{n}, \mathbf{b}) p(\mathbf{n}). \quad (1.49)$$

Другие характеристики модели при данной стратегии доступа (т.е. $RS_i(TR)$ и $N_0(TR)$) определяются аналогично (1.9) и (1.11).

Замечание 1.7. Из формул (1.48) определяем, что если $b_i = b_j$, $i \neq j$, то $PB_i(TR) = PB_j(TR)$ при любых значениях структурных и нагрузочных параметров системы.

Основная проблема при расчете характеристик (1.48), (1.49) состоит в нахождении $p(\mathbf{n})$, $\mathbf{n} \in S$, так как в отличие от CS-стратегии в данном случае, как и в случае использования CSE-стратегии доступа, не существует стационарного распределения модели мультипликативного вида. Последнее обстоятельство затрудняет решение рассматриваемой проблемы при больших размерностях ФПС (1.44).

В связи с этим здесь предлагается другой метод решения указанной проблемы, основанный на подходе, изложенном в параграфах 1.1.1 и 1.1.2. Поскольку выше дано детальное описание используемого подхода, промежуточные математические выкладки опускаются.



(б)

Рис.1.8. Вид ФПС модели с TR-стратегией доступа при $N = 10$, $K = 2$, $\mathbf{b} = (1, 2)$, $\mathbf{r} = (3, 1)$ (а) и граф переходов между ее состояниями (б).

В данном случае характеристики (1.48) и (1.49) выражаются через стационарное распределение укрупненной модели $(\pi(x) : x = 0, \dots, N)$ следующим образом:

$$PB_i(TR) = \sum_{j=0}^{b_i + n_i - 1} \pi(N - j), i = \overline{1, K}, \quad (1.50)$$

$$\tilde{N}(TR) = \sum_{j=1}^N j \pi(j). \quad (1.51)$$

Стационарное распределение укрупненной модели удовлетворяет следующие соотношения:

$$\pi(i) = \frac{1}{i} \sum_{j=1}^K v_j b_j \pi(i - b_j) G_j(i - b_j), i = 1, \dots, N; \quad (1.52)$$

$$\sum_{i=0}^N \pi(i) = 1, \quad (1.53)$$

где $v_i = \lambda_i / \mu_i$, $\pi(x) = 0$, если $x < 0$.

$$G_i(j) = \begin{cases} 1, & \text{если } j \leq N - b_i - r_i, \\ 0 & \text{в остальных случаях.} \end{cases}$$

Тогда используя описанный выше подход, находим, что стационарное распределение укрупненной модели определяется так:

$$\pi(i) = g_i(0), \quad i = 1, \dots, N, \quad (1.54)$$

где $g(i)$ определяются с помощью одномерных рекуррентных формул:

$$g_0 := 1, \quad (1.55)$$

$$g_i = \frac{1}{i} \sum_{j=1}^i j v_j g_{i-j} G_j(i-j), \quad i = 1, \dots, N. \quad (1.56)$$

С учетом (1.50) и (1.51) находим, что

$$PB_i(TR) = \left(\sum_{j=N-r_i+1}^N g_j \right) / \left(\sum_{j=0}^N g_j \right), \quad (1.57)$$

$$\tilde{N}(TR) = \left(\sum_{i=1}^N i g_i \right) / \left(\sum_{j=0}^N g_j \right). \quad (1.58)$$

1.2. Стратегия, основанная на схеме выделения специальной группы каналов для обслуживания широкополосных заявок

В настоящем параграфе предлагается новая стратегия доступа в *MRQ* типа Гимпельсона. В ней превентивной мерой защиты *w*-заявок от частых потерь является использование индивидуальной группы каналов специально для обслуживания заявок данного типа (*Special Group of Channels, SGC*-стратегия доступа). Иными словами, в данной стратегии доступа определенная часть каналов выделяется специально для *w*-заявок, а в оставшейся общей зоне вводится ограничение либо на число заявок одного типа, либо обоих типов. Важно отметить, что, несмотря на достаточную сложность предложенной стратегии доступа, здесь с применением подхода, использованного в предыдущей части данной книги, удается получить простые алгоритмы расчета желаемых показателей характеристик исследуемой системы.

Описание предлагаемой *SGC*-стратегии доступа состоит в следующем. С целью защиты *w*-заявок от частых потерь, из общего множества каналов специальная часть выделяется для их обслуживания. Очевидно, что размер специальной зоны должен быть кратным m , т.к. в противном случае эта часть каналов использовалась бы не эффективно.

Предположим, что размер специальной зоны для обслуживания *w*-заявок равен mA , где A является целым положительным числом, $A = 1, 2, \dots, A$,

$$\bar{A} = \begin{cases} \lfloor N/m \rfloor - 1, & \text{если } \text{mod}(N, m) = 0, \\ \lfloor N/m \rfloor & \text{в противном случае.} \end{cases}$$

Оставшаяся часть каналов размером $N - mA$ совместно используется заявками обоих типов. Однако, поскольку образование специальной зоны увеличивает шансы *w*-заявок быть принятыми в систему, то здесь с целью компенсации шансов *n*-заявок вводится ограничение на число *w*-заявок в общей зоне. Иными словами, считается, что число *w*-заявок в общей зоне не должно превышать некоторого порогового значения \bar{k}_w . Очевидно, что

$$\bar{k}_w \leq \left\lfloor \frac{N}{m} - A \right\rfloor.$$

Замечание 1.8. При $\bar{k}_w = \left\lfloor \frac{N}{m} - A \right\rfloor$ из данной стратегии как частный

случай получается стратегия, которая ограничивает доступ *n*-заявок в систему с пороговым параметром равным $N - mA$. Здесь же отметим, что если $\left\lfloor \frac{N}{m} - A \right\rfloor = 0$, то данная стратегия соответствует схеме полного

разделения всего множества каналов между *n*- и *w*-заявками. Поэтому в дальнейшем считается, что параметр A выбирается таким образом, чтобы выполнялось условие $k_w \geq 1$.

Рассмотрим задачу расчета вероятностей потерь разнотипных заявок при использовании предложенной стратегии доступа к каналам.

В стационарном режиме функционирование данной *MRQ* описывается цепью Маркова с состояниями вида $\mathbf{k} = (k_n, k_w)$, где $k_n(k_w)$ означает число *n*-заявок (*w*-заявок) в системе. Тогда согласно описанной стратегии доступа, фазовое пространство состояний данной ЦМ определяется так:

$$S^S := \{ \mathbf{k} : k_n = 0, 1, \dots, N - mA; k_w = 0, 1, \dots, k_w^{\max}(k_n) \}, \quad (1.59)$$

где $k_w^{\max}(k_n)$ означает максимально допустимое число *w*-заявок в системе, если число *n*-заявок в системе равно k_n . Легко показать, что

$$k_w^{\max}(k_n) = A + \min \left\{ \bar{k}_w, \left\lfloor \frac{N - mA - k_n}{m} \right\rfloor \right\}. \quad (1.60)$$

Схема *SGC*-стратегии доступа в каналы показана на рис.1.9, где f_0 означает число свободных каналов в общей зоне.

Вид ФСП модели показан на рисунке 1.10 (а). С учетом работы системы с учетом предложенной стратегии заключаем, что данная ЦМ пред-

ставляет собой модель двумерного процесса размножения и гибели в пространстве состояний S^g . Элементы ее производящей матрицы $q(\mathbf{k}, \mathbf{k}')$, $\mathbf{k}, \mathbf{k}' \in S^g$ определяются так (см. рис. 1.10 (б)):

$$q(\mathbf{k}, \mathbf{k}') = \begin{cases} \lambda_n, & \text{если } \mathbf{k}' = \mathbf{k} + \mathbf{e}_1, \\ \lambda_w, & \text{если } \mathbf{k}' = \mathbf{k} + \mathbf{e}_2, \\ k_n \mu_n, & \text{если } \mathbf{k}' = \mathbf{k} - \mathbf{e}_1, \\ k_w \mu_w, & \text{если } \mathbf{k}' = \mathbf{k} - \mathbf{e}_2, \\ 0 & \text{в остальных случаях.} \end{cases} \quad (1.61)$$

На основе (1.61) можно составить СУР для стационарного распределения $(p(\mathbf{k}); \mathbf{k} \in S^g)$ данной модели. Однако, в виду очевидности их составления явный вид этой СУР здесь не приводится. Отмеченные выше трудности имеют место и при нахождении стационарного распределения данной модели на основе решения СУР. Поэтому используем приближенный подход.

Для дальнейшего изложения необходимо рассмотреть следующее представление ФПС (1.59) модели:

$$S_g = \prod_{i=0}^{N-m} S_i^g, \quad S_i^g = \emptyset, \quad i \neq j, \quad (1.62)$$

где $S_i^g := \{\mathbf{k} \in S : k_n = i\}$.

Рассмотрим проблемы определения мощностей множеств S_i^g в расщеплении (1.62). Очевидно, что наибольшее число элементов имеют множества S_i^g для $i = 0, k_n^1$, где k_n^1 определяется как решение уравнения $k_n + m \bar{k}_w = N - mA$ относительно k_n , т.е.

$$k_n^1 = N - m(A + \bar{k}_w). \quad (1.63)$$

Иными словами, мощность всех множеств S_i^g , для $i = 0, k_n^1$, равна $A + \bar{k}_w$. Поскольку уменьшение числа w -заявок в системе на единицу позволяет увеличить число n -заявок ровно на m единиц, то легко найти, что мощность множества S_i^g для $i = k_n^1 + (j-1)m + 1, k_n^1 + jm$ равна $A + \bar{k}_w - j$, где $j = 1, 2, \dots, \bar{k}_w$.

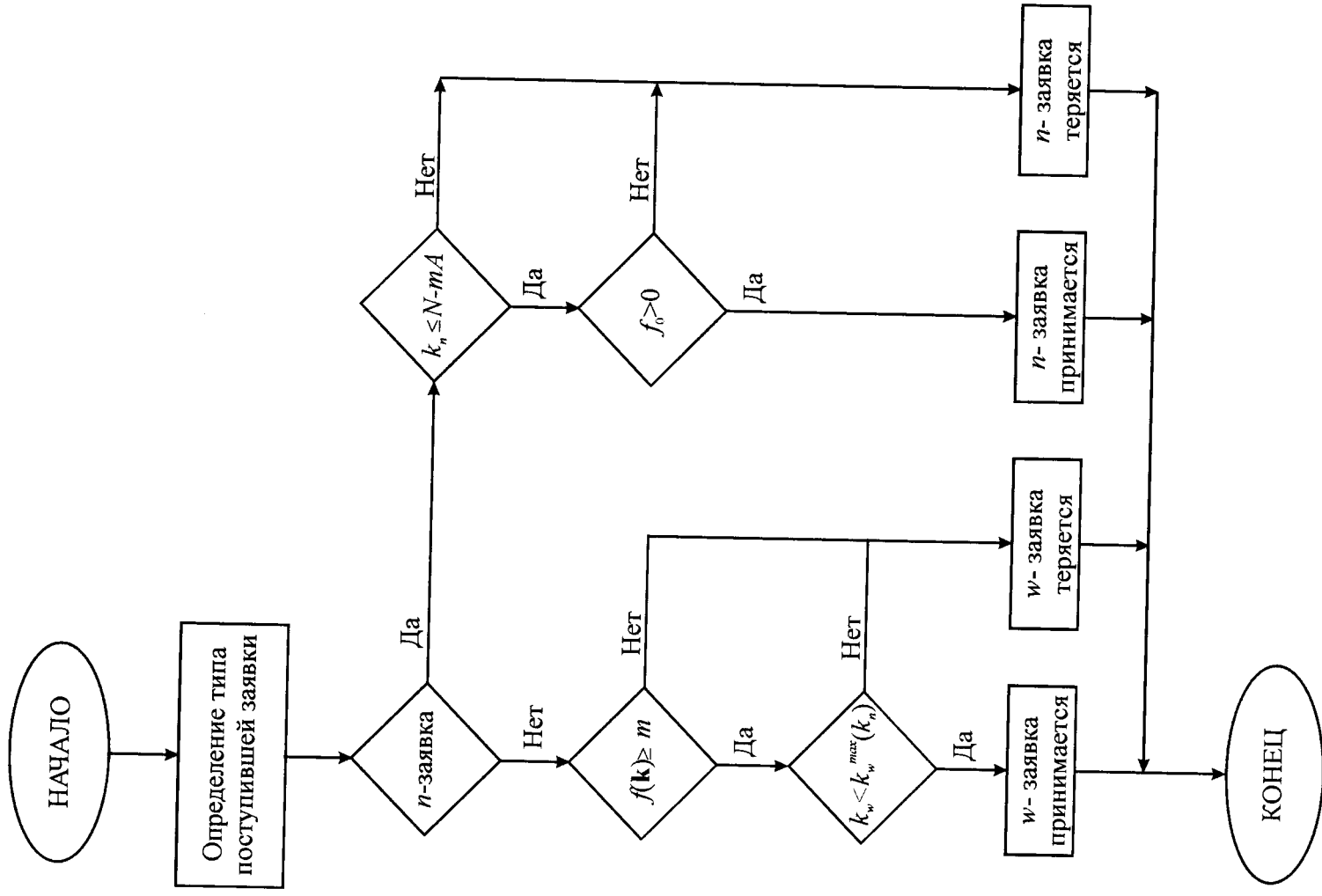
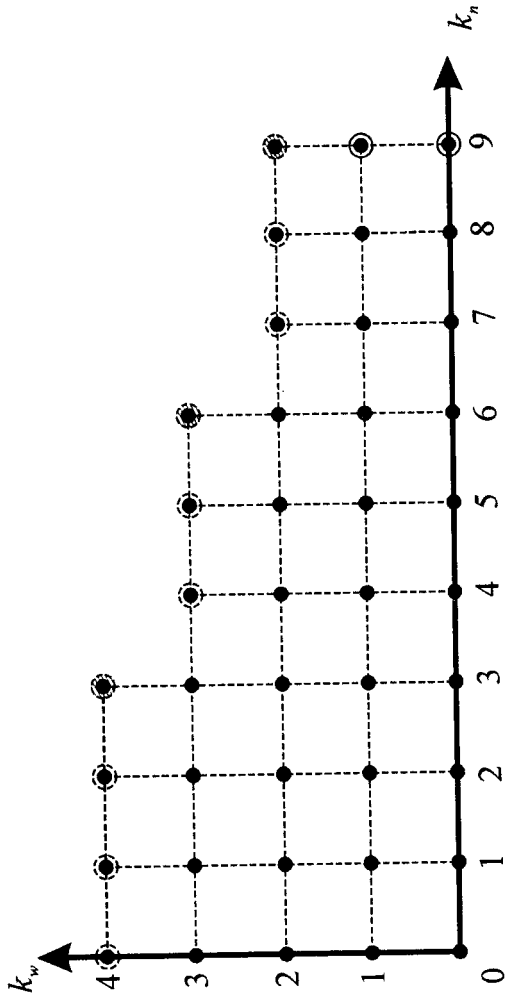
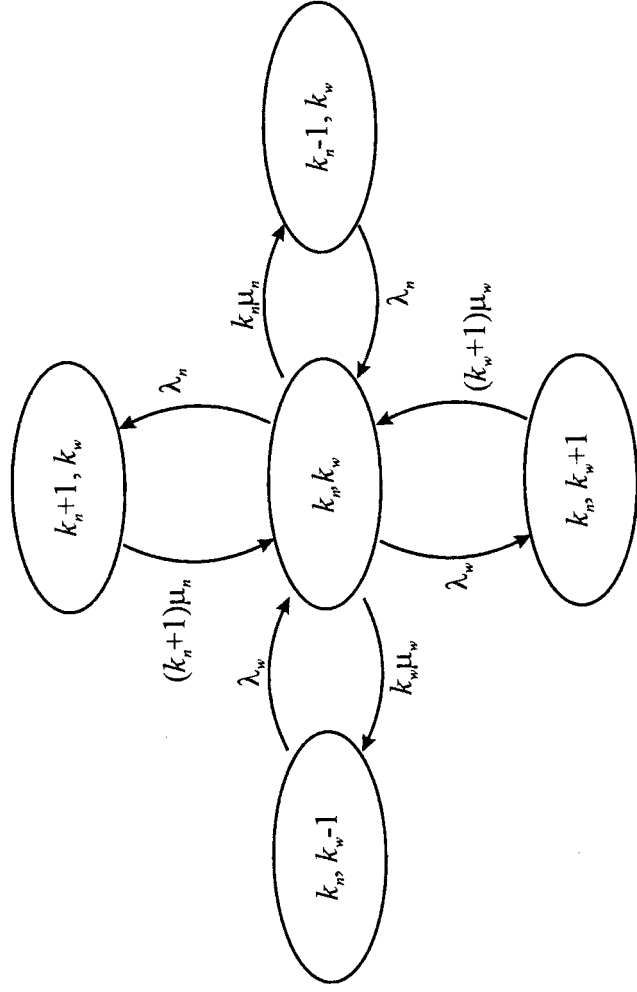


Рис.1.9. Схема SGC-стратегии доступа в каналы



- - блокируются лишь n -заявки;
- - блокируются лишь w -заявки;
- ⊙ - блокируются заявки обоих типов.

(a)



(б)

Рис.1.10. Вид ФПС модели для SGC-стратегии доступа при $N = 15$, $m = 3$, $A = 2$, $k_w = 2$ (a) и граф переходов между состояниями системы (б)

Очевидно, что мощность множества S_i^g в расщеплении (1.62) одновременно определяет максимально возможное число w -заявок в данном подмножестве состояний. Следовательно, с учетом выше изложенного, заключаем, что формула (1.60) может быть записана так:

$$k_w^{\max}(k_n) = \begin{cases} A + \bar{k}_w, & \text{если } 0 \leq k_n \leq k_n^1, \\ A + \bar{k}_w - i, & \text{если } k_n^1 + (i-1)m + 1 \leq k_n \leq k_n^1 + im, \quad i = 1, 2, \dots, \bar{k}_w, \end{cases} \quad (1.64)$$

где k_n^1 определяется из (1.63).

Замечание 1.9. Максимальное значение параметра k_n в формуле (1.64) определенное как $k_n^1 + mk_w$, согласно (1.63) равно $N - mA$, т.е. с помощью формулы (1.64) определяются все допустимые значения параметра $k_w^{\max}(k_n)$ из формулы (1.60).

Вероятности потери n -заявок (PB_n) и w -заявок (PB_w) определяются через стационарное распределение $p(\mathbf{k})$, $\mathbf{k} \in S^g$, данной цепи. С учетом работы предложенной стратегии доступа искомые вероятности потери вычисляются так:

$$PB_n(SGC) = \sum_{i=0}^{\bar{k}_w} p(k_n^1 + im, A + \bar{k}_w - i) + \sum_{\mathbf{k} \in S_{N-m}^g} p(\mathbf{k}), \quad (1.65)$$

$$PB_w(SGC) = \sum_{k_n=0}^{N-mA} p(k_n, k_w^{\max}(k_n)). \quad (1.66)$$

Теперь рассмотрим эффективный метод, который, минуя решения соответствующей системы балансовых уравнений большой размерности, позволяет найти стационарное распределение данной модели с помощью простых вычислительных процедур.

В расщеплении (1.62) классы состояний S_i^g описываются одним укрупненным состоянием $\langle i \rangle$, и на исходном ФПС (1.59) определяется следующая функция укрупнения:

$$U(\mathbf{k}) = \langle i \rangle, \text{ если } \mathbf{k} \in S_g. \quad (1.67)$$

Функция укрупнения (1.67) определяет укрупненную (относительно исходной) ЦМ с пространством состояний $\hat{S} := \{\langle i \rangle : i = 0, \dots, N - mA\}$.

Согласно алгоритмам фазового укрупнения стационарное распределение рассматриваемой цепи Маркова приближенно определяется так:

$$p(i, j) \approx \rho^j(j) \pi(\langle i \rangle), \quad (i, j) \in S^g, \quad (1.68)$$

где $\rho^j(j)$, $(i, j) \in S^g$ и $\pi(\langle i \rangle)$, $\langle i \rangle \in \hat{S}$, являются стационарными распределениями внутри класса S_i^g и укрупненной модели, соответственно.

Стационарное распределение внутри класса S_i^g определяется как известное распределение классической системы обслуживания $M | M | k_w^{\max}(i) | 0$ с нагрузкой ν_w эрл., $\nu_w := \lambda_w / \mu_w$, т.е.

$$\rho^i(j) = \frac{\nu_w^j}{j!} \rho^i(0), \quad j = \overline{1, k_w^{\max}(i)}, \quad (1.69)$$

$$\rho^i(0) = \left(\sum_{j=0}^{k_w^{\max}(i)} \frac{\nu_w^j}{j!} \right)^{-1}. \quad (1.70)$$

С использованием (1.61), (1.69) и (1.70) определяются элементы производящей матрицы укрупненной модели $q(<i>, <j>), <i>, <j> \in \hat{S}$.

После выполнения определенных математических преобразований находят следующие формулы для вычисления стационарного распределения укрупненной ЦМ, которая представляет собой модель одномерного процесса размножения и гибели:

$$\pi(x) = \begin{cases} \frac{\nu_w^x}{x!} \pi(<0>), & \text{если } 1 \leq x \leq k_n^1, \\ \frac{\nu_w^x}{x!} \prod_{i=0}^{x-1} (1 - E_B(\nu_w, A + \bar{k}_w - i)) \pi(<0>), & \text{если } k_n^1 + (j-1)m + 1 \leq x \leq k_n^1 + jm, \quad j = \overline{1, \bar{k}_w}, \end{cases} \quad (1.71)$$

где

$$\pi(<0>) = \left(1 + \sum_{x=1}^{k_n^1} \frac{\nu_w^x}{x!} + \sum_{x=k_n^1+1}^{k_n^1 + m \bar{k}_w} \frac{\nu_w^x}{x!} \prod_{i=0}^{x-1} (1 - E_B(\nu_w, A + \bar{k}_w - i)) \right)^{-1}, \quad (1.72)$$

$$y_x = \begin{cases} \left[\frac{x - k_n^1}{m} \right], & \text{если } \text{mod}(x - k_n^1, m) = 0, \\ \left[\frac{x - k_n^1}{m} \right] + 1, & \text{в противном случае;} \end{cases} \quad (1.73)$$

$E_B(\nu, r) - B$ - формула Эрланга.

С учетом (1.65), (1.66) и (1.71) - (1.73) находятся следующие формулы для приближенного вычисления вероятностей потери разнотипных заявок:

$$PB_n \approx \pi(<k_n^1 + m \bar{k}_w >) + \sum_{i=0}^{\bar{k}_w-1} E_B(\nu_w, A + \bar{k}_w - i) \pi(<k_n^1 + im >); \quad (1.74)$$

$$PB_w \approx E_B(\nu_w, A + \bar{k}_w) \sum_{i=0}^{k_n^1} \pi(<i>) + \sum_{i=1}^{\bar{k}_w} E_B(\nu_w, A + \bar{k}_w - 1) \sum_{j=k_n^1+(i-1)m+1}^{k_n^1+im} \pi(<j>). \quad (1.75)$$

Важно отметить сравнительно невысокую сложность предложенного алгоритма вычисления искомых вероятностей потерь, так как в формулах (1.74), (1.75) широко используется B - формула Эрланга, которая табулирована.

1.3. Численные эксперименты

Сначала рассмотрим результаты численных экспериментов для обшей модели MRQ при всех трех исследуемых стратегиях доступа в каналы в отдельности. Отметим, что разработка соответствующих вычислительных программ на основе алгоритмов, предложенных в параграфах 1.1.1-1.1.3, не представляет никаких трудностей при любых размерах систем ФПС моделей. Вычислительные программы разработаны на языке Object Pascal в интегрированной среде разработки Delphi 6. Цель выполнения этих экспериментов состоит в определении поведения характеристик модели при изменении ее нагрузочных и структурных параметров.

Для наглядности представления результатов численных экспериментов в настоящем параграфе представлены соответствующие результаты для модели Гимпельсона, в которой на N каналах обслуживаются заявки двух типов - узкополосные (narrow-band, n -заявки) и широкополосные (wide-band, w -заявки). При этом n -заявка обслуживается лишь одним каналом, а для обслуживания одной w -заявки требуется одновременно m каналов, $1 < m \leq N$.

Ниже данная модель при всех трех стратегиях доступа исследуется с помощью разработанных модифицированных алгоритмов Кауфмана Робертса.

1. CS-стратегия доступа в каналы

Здесь введем следующие обозначения: $\lambda_n(\lambda_w)$ - интенсивность потока n -заявок (w -заявок); $\mu_n(\mu_w)$ - интенсивность обслуживания n -заявок (w -заявок).

При CS-стратегии доступа в каналы поступившая заявка любого типа принимается, если в момент ее поступления имеются необходимые число свободных каналов. При этом обслуживание w -заявки начинается и завершается всеми m каналами одновременно, т.е. w -заявки занимают

все m каналов на весь период обслуживания, $1 \leq m \leq N$, где N – общее число каналов.

В этом параграфе для удобства изложения основные показатели QoS – вероятности потери n -заявок и w -заявок обозначаются ($PB_n(CS)$) и ($PB_w(CS)$), соответственно.

Очевидно, что для данной стратегии при любых значениях нагрузочных и структурных параметров модели имеет место соотношение (см. также (1.24)):

$$PB_w(CS) > PB_n(CS).$$

Среднее число занятых каналов системы обозначается через $\bar{N}(N, m)$.

С применением описанного в параграфе 1.1.1 алгоритма после выполнения соответствующих математических преобразований получают следующие выражения для характеристик модели:

$$PB_n(CS) = g_N / \left(\sum_{r=0}^N g_r \right),$$

$$PB_w(CS) = \left(\sum_{r=N-m+1}^N g_r \right) / \left(\sum_{r=0}^N g_r \right),$$

$$N^{(CS)} = \left(\sum_{r=1}^N r g_r \right) / \left(\sum_{r=0}^N g_r \right).$$

где

$$g_r = \begin{cases} \frac{V_n^r}{r!}, & \text{если } r = 1, m-1, \\ \frac{1}{r} \left(\frac{V_n^r}{r!} + r V_w \right), & \text{если } r = m, \\ \frac{1}{r!} (V_n \cdot g_{r-1} + m V_w g_{r-m}), & \text{если } r = \overline{m+1, N}, \end{cases}$$

$$V_n := \lambda_n / \mu_n, \quad V_w := \lambda_w / \mu_w.$$

Часть результатов численных экспериментов показана на рис.1.11–1.16.

Их анализ позволяет сделать следующие выводы:

– значения функций PB_n и PB_w , вычисленные для исходных данных работы [15] полностью совпадают с соответствующими результатами указанной работы, но при этом разработанный здесь алгоритм по простоте и скорости превосходит соответствующую вычислительную процедуру [15];

– при фиксированной общей нагрузке функции PB_n и PB_w от V_n (или V_w), вообще говоря, не являются монотонными (см. рис.1.11 и 1.12);

– при фиксированных N , V_n , V_w функция PB_w является монотонно возрастающей относительно аргумента m , но аналогичное свойство не выполняется для функции PB_n (см. рис.1.13);

– при фиксированных нагрузках входящих графиков и m , функция PB_w является монотонно убывающей относительно аргумента N , а функция PB_n монотонно убывает лишь в полуоткрытых интервалах $N \in [Rm, (R+1)m)$, где R – натуральное число, т.е. в точках типа $N = Rm$ значения функции PB_n резко увеличиваются по сравнению с предыдущей точкой $N = Rm - 1$. Такое поведение функции PB_n имеет следующее объяснение: при значениях N , кратных m , появляется дополнительная возможность для w -заявок быть принятыми на обслуживание, чем ухудшается соответствующая возможность для n -заявок (см. рис.1.14);

– поведение функции \bar{N} в зависимости от m также отличается своей монотонностью и существенным образом зависит от соотношения значений нагрузочных параметров модели (см. рис.1.15);

– при любых значениях интенсивностей поступления разнотипных заявок и количества каналов функции PB_n и PB_w являются монотонно убывающими относительно интенсивностей их обслуживания (см. рис.1.16).

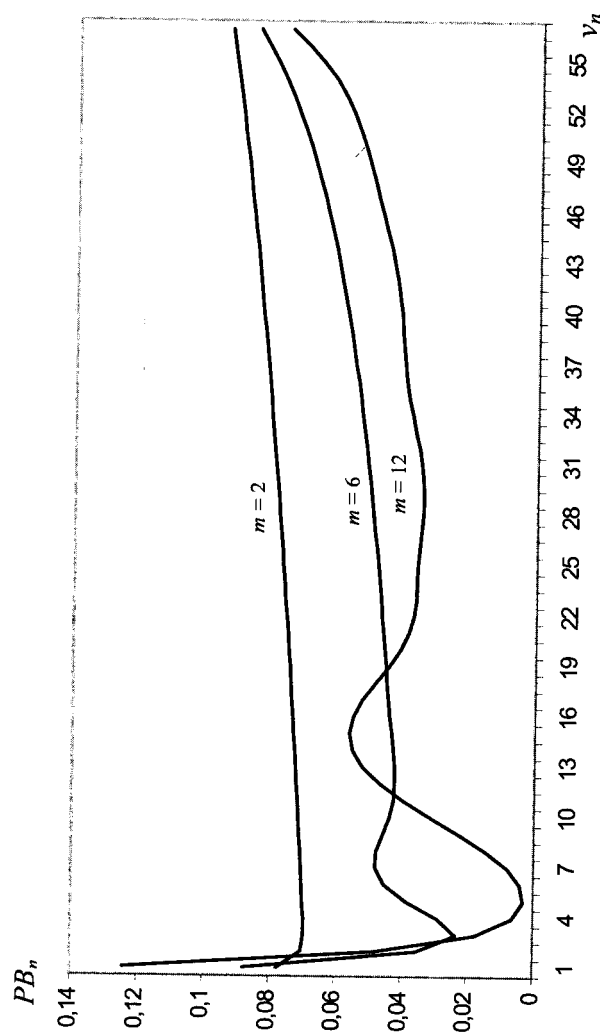


Рис.1.11. Зависимость вероятности потери n -заявок от V_n при CS стратегии доступа: $B = 60$; $V_n + mV_w = 60$

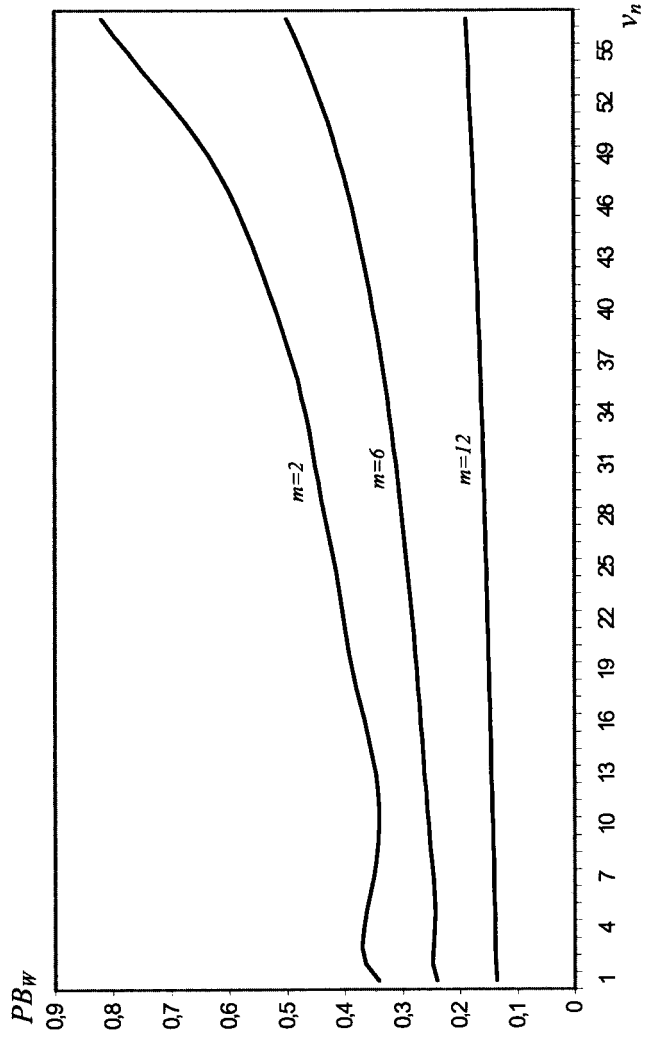


Рис. 1.12. Зависимость вероятности потери w -заявок от ν_n при CS-стратегии доступа: $B = 60$; $\nu_n + m\nu_w = 60$

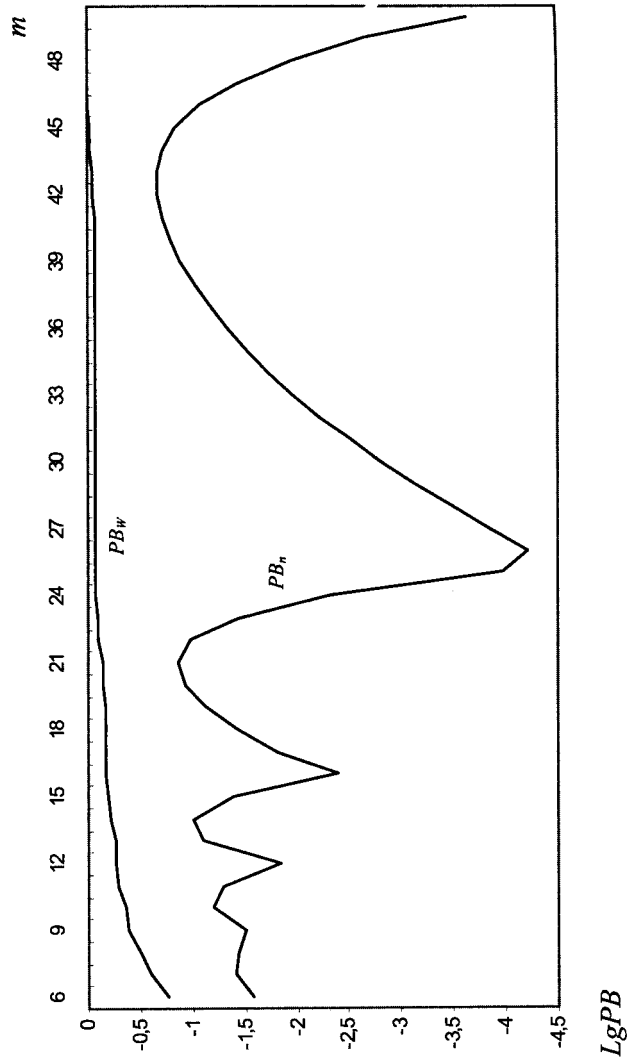


Рис. 1.13. Зависимость вероятности потери заявок от m при CS-стратегии доступа: $B = 50$; $\nu_n = 10$, $\nu_w = 5$

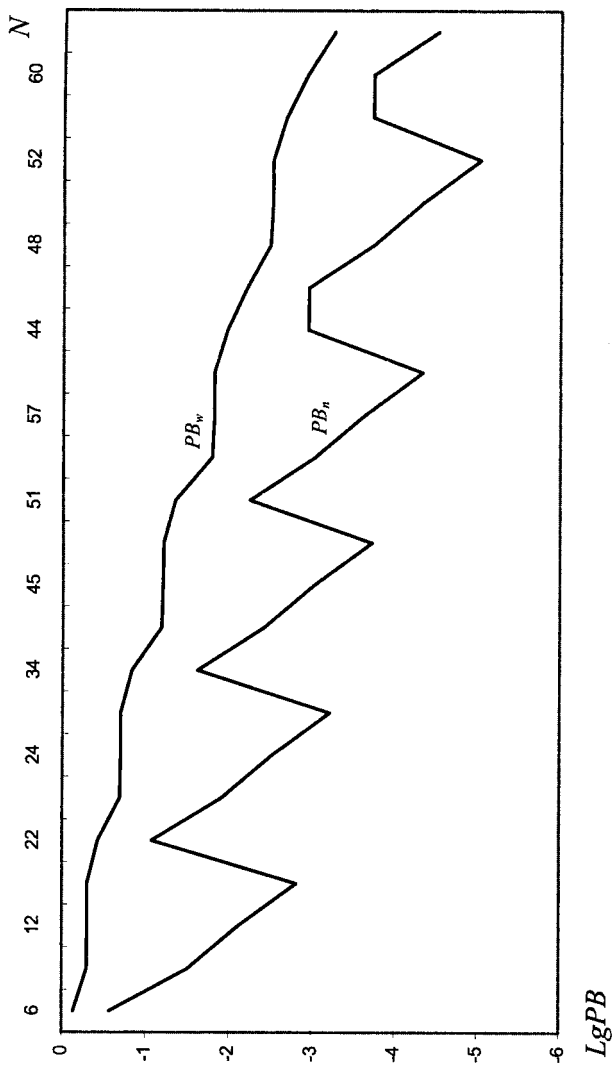


Рис. 1.14. Зависимость вероятности потери заявок от N при CS-стратегии доступа: $m = 6$; $\nu_n = 1$; $\nu_w = 1$

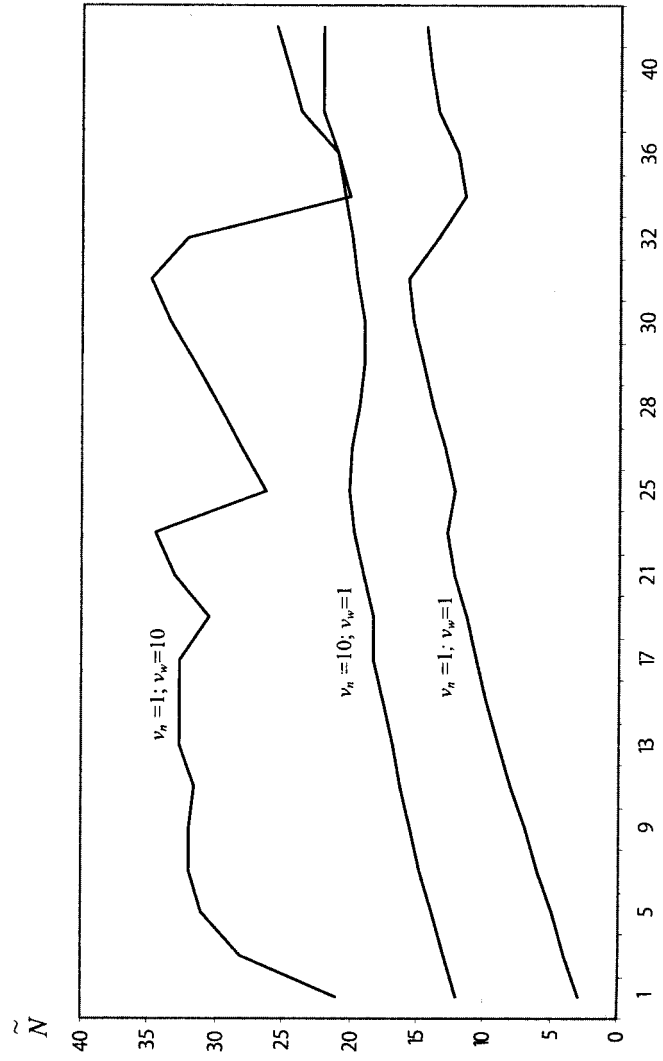


Рис. 1.15. Зависимость среднего числа занятых каналов от m при CS-стратегии доступа $N = 40$

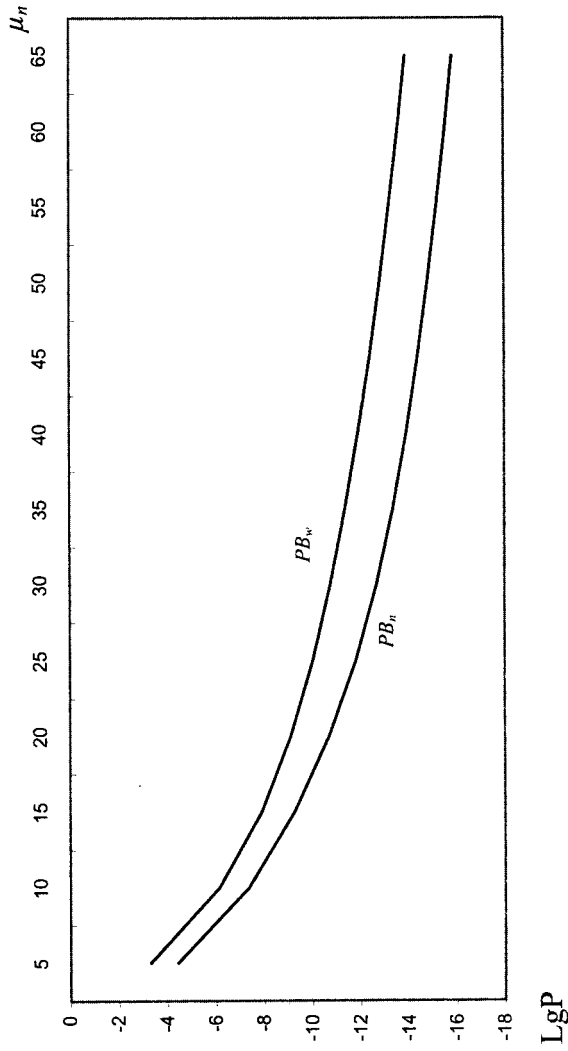


Рис. 1.16. Зависимость вероятности потери заявок от μ_n при CS-стратегии доступа: $N = 60$; $m = 6$; $\lambda_n = 50$; $\lambda_w = 50$

2. CSE-стратегия доступа в каналы

При применении CSE-стратегии доступа в модели Гимпельсона по-ступившие n -заявка или w -заявка принимаются для обслуживания лишь тогда, когда число свободных каналов больше или равно m , где m число каналов, требуемых для обслуживания w -заявок.

Вероятности блокировки n -заявок и w -заявок при использовании данной стратегии доступа обозначается через $P_{B_n}(CSE)$ и $P_{B_w}(CSE)$, соответственно. При этом для данной стратегии при любых значениях нагрузочных и структурных параметров модели имеем:

$$P_{B_n}(CSE) = P_{B_w}(CSE) = PB(CSE).$$

Используя алгоритм, разработанный в §1.1.2, после выполнения определенных математических преобразований удается намного упростить промежуточные вычислительные процедуры для нахождения $P_{B_n}(CSE)$ (или $P_{B_w}(CSE)$). При этом необходимо различать два случая.

$$\text{Случай } m \leq \left[\frac{N+1}{2} \right]:$$

$$g_r = \begin{cases} \frac{V_n^r}{r!}, & \text{если } r = 0, m-1, \\ \frac{1}{r} (V_n g_{r-1} I(m \leq N-r+1) + m V_w), & \text{если } r = m, N; \end{cases}$$

Случай $m > \left[\frac{N+1}{2} \right]:$

$$g_r = \begin{cases} \frac{V_n^r}{r!}, & \text{если } r = 0, m-1, \\ \frac{1}{r} m V_w g_{r-m}, & \text{если } r = m, N. \end{cases}$$

Часть результатов численных экспериментов, проведенных с помощью разработанных формул, показаны на рис.1.17-1.20.

Их анализ позволяет сделать следующие выводы:

- вероятность потери заявок с ростом значений параметра m (т.е. с увеличением ширины w -заявок) также растет, что является вполне ожидаемым результатом (см. рис.1.17). Однако, при фиксированной общей нагрузке (в наших экспериментах $V_n + V_w = 12$ Эрл) скорость изменения данной функции существенным образом зависит от соотношения интенсивностей входящих трафиков. Вместе с тем, начиная с определенного значения параметра m значения этой функции приближается к определенному пределу (в наших экспериментах указанное значение параметра m равно 18);

- среднее число занятых каналов является убывающей функцией аргумента m , т.е. с увеличением ширины w -заявок уменьшается коэффициент использования каналов (см. рис.1.18). При этом следует отметить, что при фиксированной общей нагрузке (в наших экспериментах $V_n + V_w = 12$ Эрл) коэффициент утилизации каналов тем выше, чем выше доля w -заявок в общей нагрузке. Последний факт также соответствует теоретическим ожиданиям;

- как и следовало ожидать, вероятность потери заявок является убывающей функцией от аргумента N (общее число каналов), при этом при фиксированной общей нагрузке вероятность потерь тем ниже, чем ниже доля w -заявок в общей нагрузке (см. рис.1.19). Так, например, при $N = 25$ ($m = 5$ является фиксированной величиной) $P_{B(4, 8)/P_{B(8, 4)}} \approx 1.3$, а при $N = 35$ соответствующее отношение приблизительно равно 2 (здесь запись $P_{B(x, y)}$ означает вероятность потери при $V_n := x$ и $V_w := y$);

- среднее число занятых каналов является возрастающей функцией аргумента N (общее число каналов), при этом при фиксированной общей нагрузке коэффициент использования тем выше, чем выше доля w -заявок в общей нагрузке (см. рис.1.20).

3. TR-стратегия доступа в каналы.

Вначале рассмотрим применения разработанного в §1.1.3 алгоритма для частной модели Гимпельсона с N каналами, в которой

$$\mathbf{b} = \left(\underbrace{1, \dots, 1}_{k_1}, \underbrace{N, \dots, N}_{k_2} \right).$$

Очевидно, что для данной модели допустимыми значениями параметров резервирования каналов являются $r_N = 0$ и $1 \leq r_1 \leq N - 1$ (при $r_1 = 0$ получаются результаты для CS-стратегии доступа).

С использованием указанного выше алгоритма легко находим, что

$$PB_N(TR) = v_m \pi(0), \quad PB_N(TR) = 1 - \pi(0),$$

$$\text{где } \hat{v}_1 := \sum_{i=1}^{k_1} v_i, \quad \hat{v}_2 := \hat{v}_3 := \dots := \hat{v}_{N-1} := 0, \quad \hat{v}_N := \sum_{i=k_1+1}^{k_1+k_2} v_i, \quad \pi(0) = \left(\hat{v}_N + \sum_{i=0}^{N-1} \hat{v}_i \right)^{-1}.$$

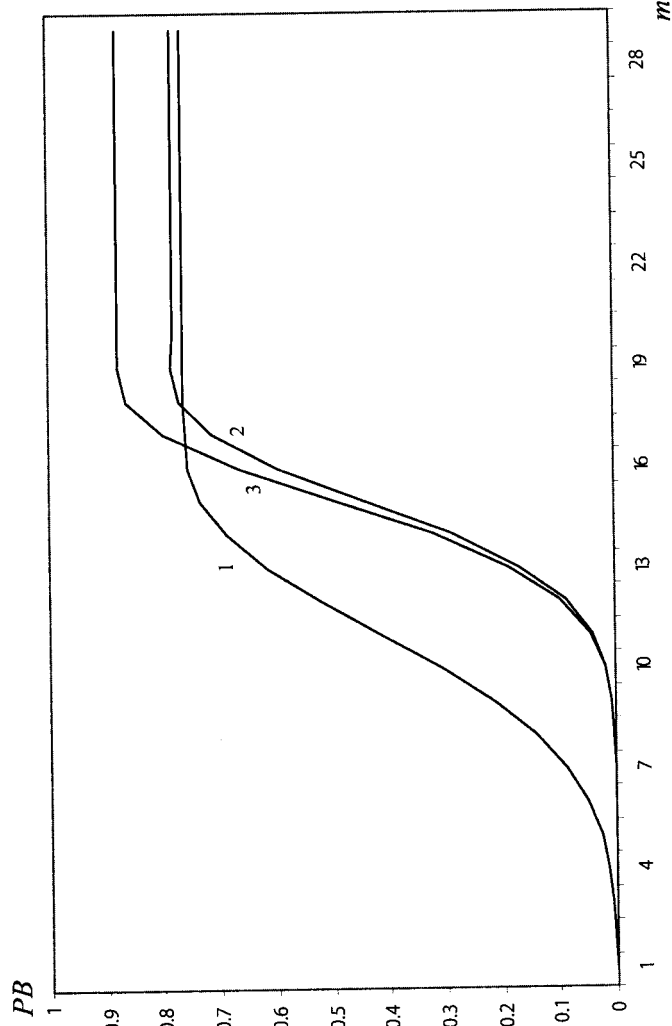


Рис. 1.17. Зависимость вероятности потерь заявок от m при CSE-стратегии доступа: $N = 50$; 1 - $v_1 = 8$; $V_m = 4$; 2 - $v_1 = 4$; $V_m = 4$; 3 - $v_1 = 4$; $V_m = 8$

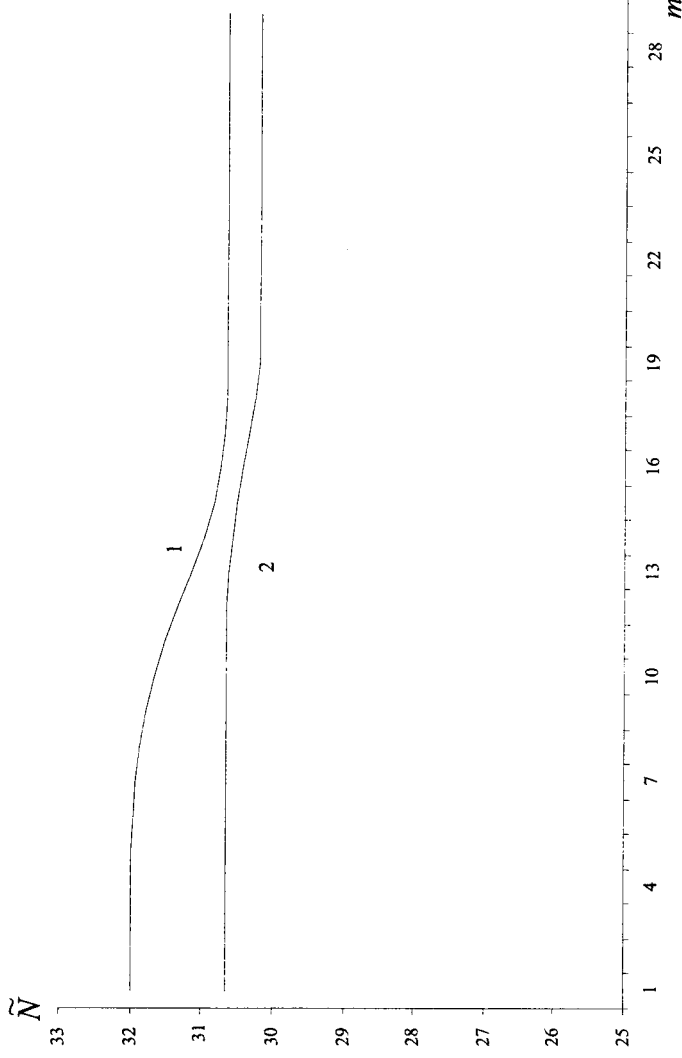


Рис. 1.18. Зависимость среднего числа занятых каналов от m при CSE-стратегии доступа: $N = 50$; 1 - $v_1 = 8$; $V_m = 4$; 2 - $v_1 = 4$; $V_m = 8$

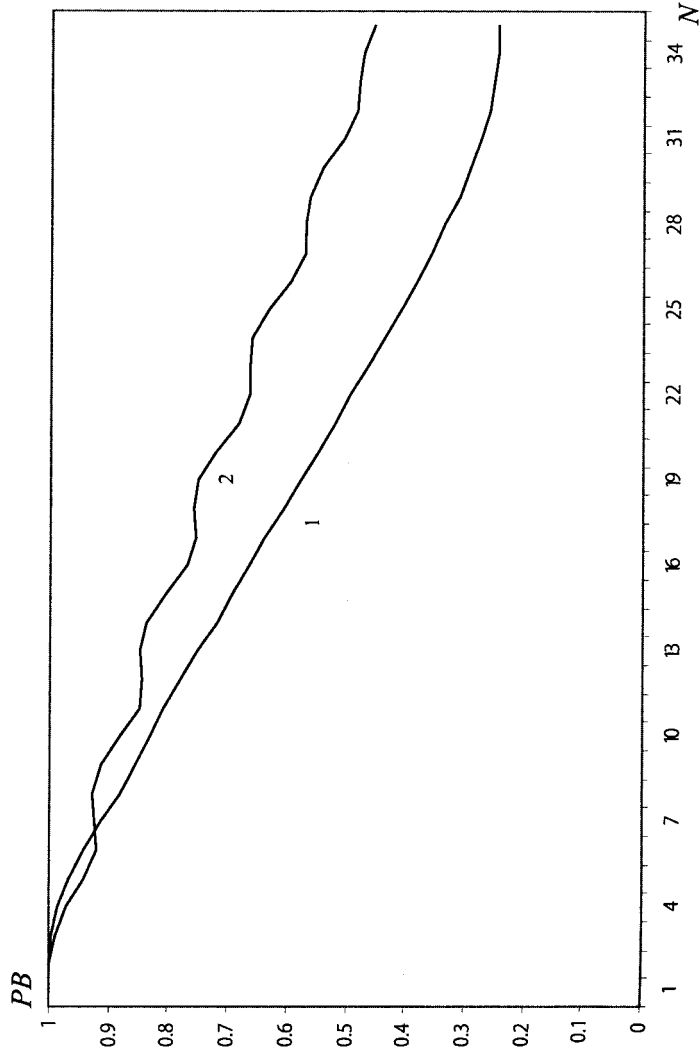


Рис. 1.19. Зависимость вероятности потери заявок от N при CSE-стратегии доступа: $m = 7$; 1 - $v_1 = 8$; $V_m = 4$; 2 - $v_1 = 4$; $V_m = 8$

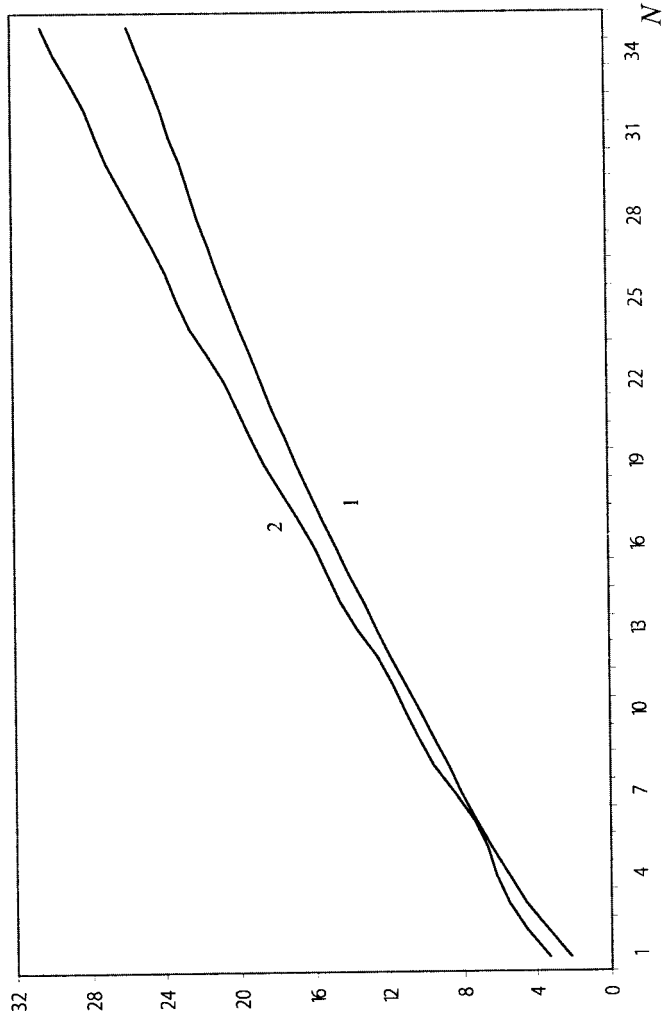


Рис.1.20. Зависимость среднего числа занятых каналов от N при CSE-стратегии доступа: $m = 7$; $1 - \nu_1 = 8$; $\nu_m = 4$; $2 - \nu_1 = 4$; $\nu_m = 8$

Здесь PB_n (PB_w) означает вероятность блокировки заявок любого типа, требующих одновременно лишь одного канала (всех каналов).

Из последних формул можно сделать следующие выводы:

- при фиксированных $\hat{\nu}_1, \hat{\nu}_N$ и r_1 с ростом N функция PB_w растет, а функция PB_n убывает;
- при фиксированных $\hat{\nu}_1, \hat{\nu}_N$ и N с ростом r_1 функция PB_n растет, а функция PB_w убывает;
- $PB_w > PB_n$ при любых $\hat{\nu}_1, \hat{\nu}_N, N$ и r_1 , так как это неравенство эквивалентно неравенству $\sum_{i=0}^{N-1} \frac{\hat{\nu}_i}{i!} > 1$;

$$\sum_{i=0}^{N-1} \frac{\hat{\nu}_i}{i!} > 1;$$

$$-\min_r \{PB_w - PB_n\} = PB_w(N-1) - PB_n(N-1) = \hat{\nu}_1 / (1 + \hat{\nu}_1 + \hat{\nu}_N).$$

Теперь рассмотрим общую модель Гимпельсона. Для данной модели, кроме рассмотренной выше частной модели Гимпельсона, не удается получить простые явные формулы для расчета ее характеристик. Вместе с тем, соответствующий алгоритм расчета указанных показателей является достаточно простым и позволяет исследовать их поведение практи-

чески в любых диапазонах изменения значений нагрузочных и структурных параметров системы.

На рисунках 1.21-1.28 показаны графики зависимостей характеристик модели от параметров резервирования r_1 и r_m (нами получены и другие зависимости этих показателей от нагрузочных параметров ν_1 и ν_2 , а также от N и m , которые из-за ограниченности объема книги здесь не приводятся).

Из рис. 1.21-1.22 видно, что при фиксированных значениях r_1 с ростом аргумента r_m функция PB_n также растет, а функция PB_w , наоборот, убывает. Эти результаты являются вполне ожидаемыми, так как с ростом значений параметра резервирования для широкополосных заявок (т.е. заявок второго типа) уменьшаются их шансы быть принятыми в канал, и одновременно увеличиваются шансы для принятия на обслуживание узкополосных заявок (т.е. заявок первого типа). При этом уменьшается коэффициент утилизации каналов (рис.1.23) и взвешенная сумма вероятностей блокировки WPB разнотипных трафиков (рис.1.24).

Функция PB_n систематически уменьшается с ростом аргумента r_1 (см. рис.1.25), что является вполне ожидаемым результатом, так как с ростом параметра резервирования уменьшаются шансы n -заявок для принятия в каналы. Следует отметить, что эта функция при фиксированных значениях параметра r_1 является монотонно убывающей относительно параметра резервирования (r_m) для w -заявок (см. также рис.1.25). Этот факт также соответствует теоретическим ожиданиям.

Неадекватным является поведение функции PB_w относительно аргумента r_1 при фиксированных значениях параметра резервирования (r_m) для w -заявок и при значениях r_1 близких к N (см. рис.1.26). Здесь также следует заметить, что эта функция при фиксированных значениях параметра r_1 является монотонно возрастающей относительно параметра резервирования (r_m) для w -заявок (см. также рис.1.26).

Из рис.1.27 видно, что взвешенная функция вероятности потери разнотипных заявок является монотонно возрастающей относительно аргумента r_1 при любых нагрузках.

Немонотонность функции \tilde{N} относительно аргумента r_1 объясняется неадекватным поведением функции PB_w (см. рис.1.28), при этом утилизация каналов улучшается при уменьшении параметра резервирования (r_m) для w -заявок (см. также рис.1.28), что является вполне ожидаемым результатом.

Теперь рассмотрим результаты численных экспериментов, выполненных с помощью разработанного в §1.2 алгоритма расчета параметров модели при SGC-стратегии доступа. Часть результатов показана на рис.1.29 и 1.30. Цель выполнения этих экспериментов состоит в определении поведения вероятностей потерь n - и w -заявок при изменении нагрузочных параметров рассматриваемой стратегии доступа в каналы. Их анализ позволяет сделать следующие выводы:

— при фиксированных N, m, V_n, V_w функция PB_n^{SGC} является монотонно возрастающей, а функция PB_w^{SGC} — наоборот, монотонно убывающей относительно обоих параметров A и k_w ;

— несмотря на то, что в отдельных интервалах небольшой длины функция $PB_w(SGC)$ относительно обоих аргументов близка к линейной, но во всей области определения она имеет нелинейный характер. Так, например, в интервале $[1, 10]$ функция $PB_w(A)$ очень хорошо аппроксимируется линейной функцией $PB_w(A) = -0.04 \cdot A + 0.89$, а уже при $A = 15$ точное значение $PB_w(A) = 0.36$, а ее приближенное значение, вычисленное по указанному линейному выражению, равно 0.29 , и это расхождение увеличивается с ростом A .

Функция $PB_n(SGC)$ относительно обоих аргументов является нелинейной, при этом это свойство особенно ярко проявляется при больших значениях A и малых значениях k_w ;

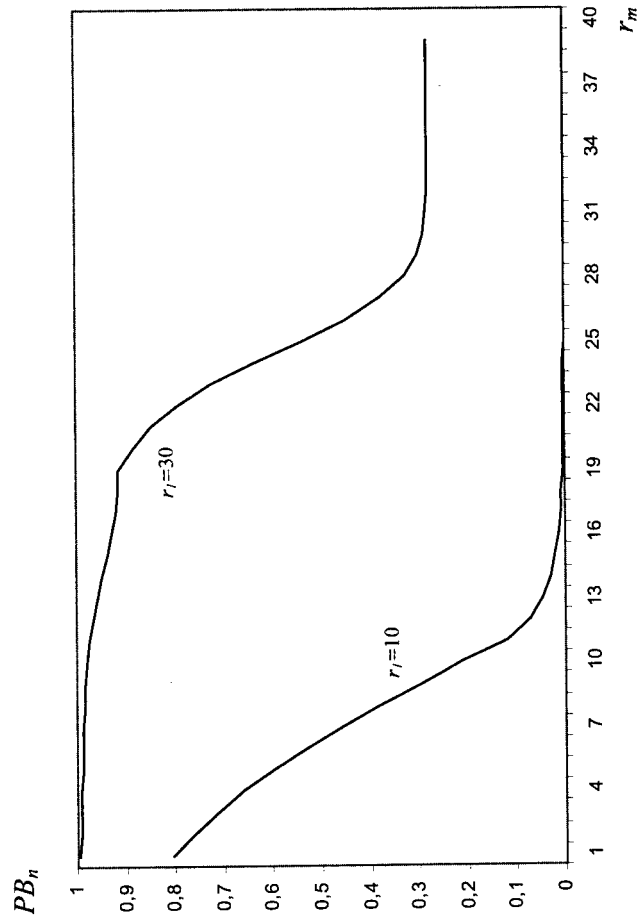


Рис. 1.21. Зависимость вероятности потери n -заявок от r_m при $N=50; m=10; v_1=25; v_2=20$

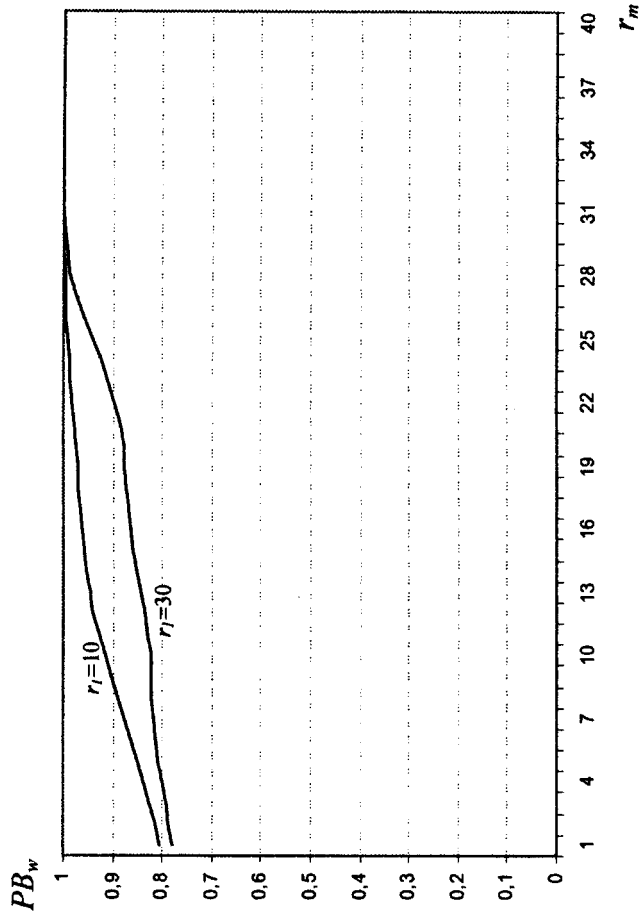


Рис. 1.22. Зависимость вероятности потери w -заявок от r_m при $N=50; m=10; v_1=25; v_2=20$

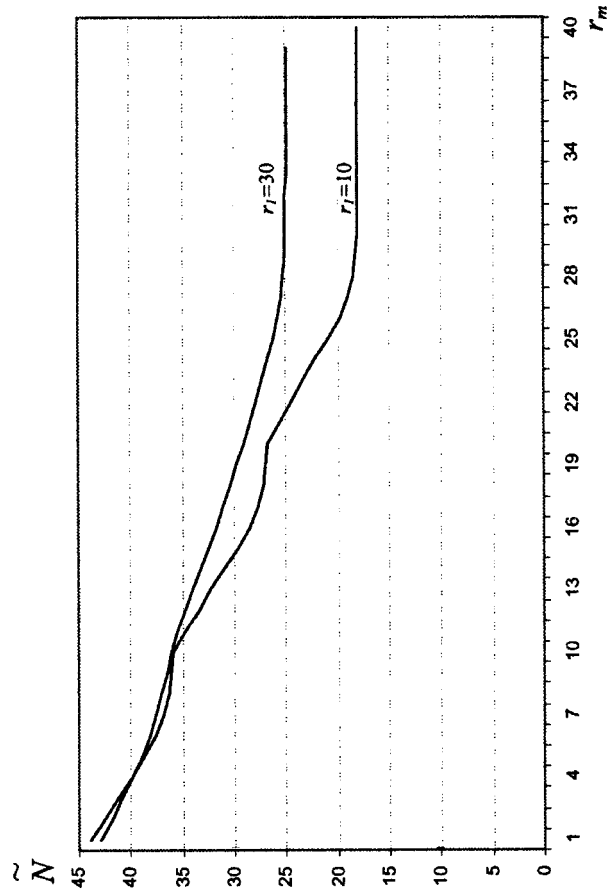


Рис. 1.23. Зависимость среднего числа занятых каналов от r_m при $N=50; m=10; v_1=25; v_2=20$

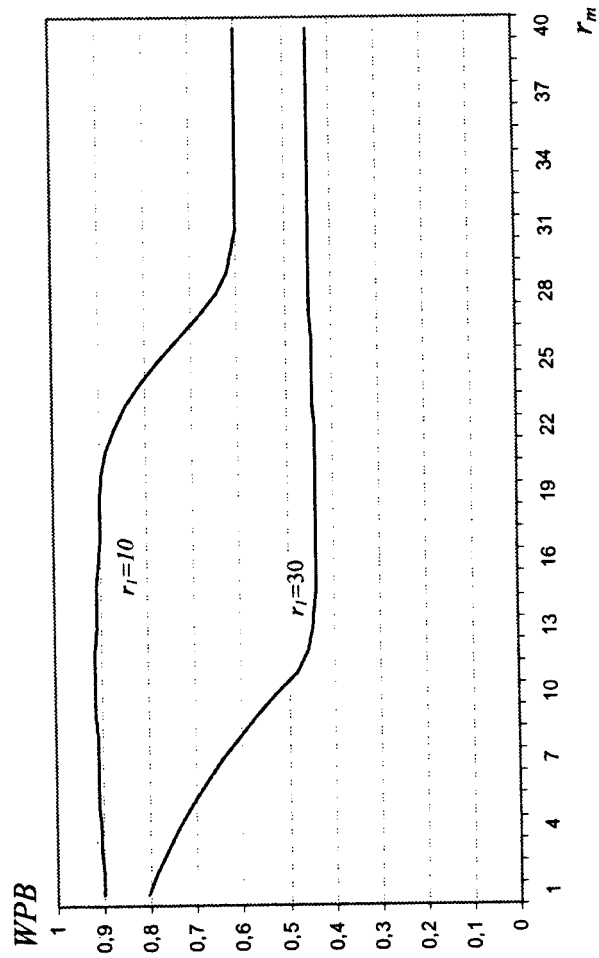


Рис. 1.24. Зависимость взвешенной суммы вероятностей потери от r_m при $N=50$; $m=10$; $\nu_1=25$; $\nu_2=20$

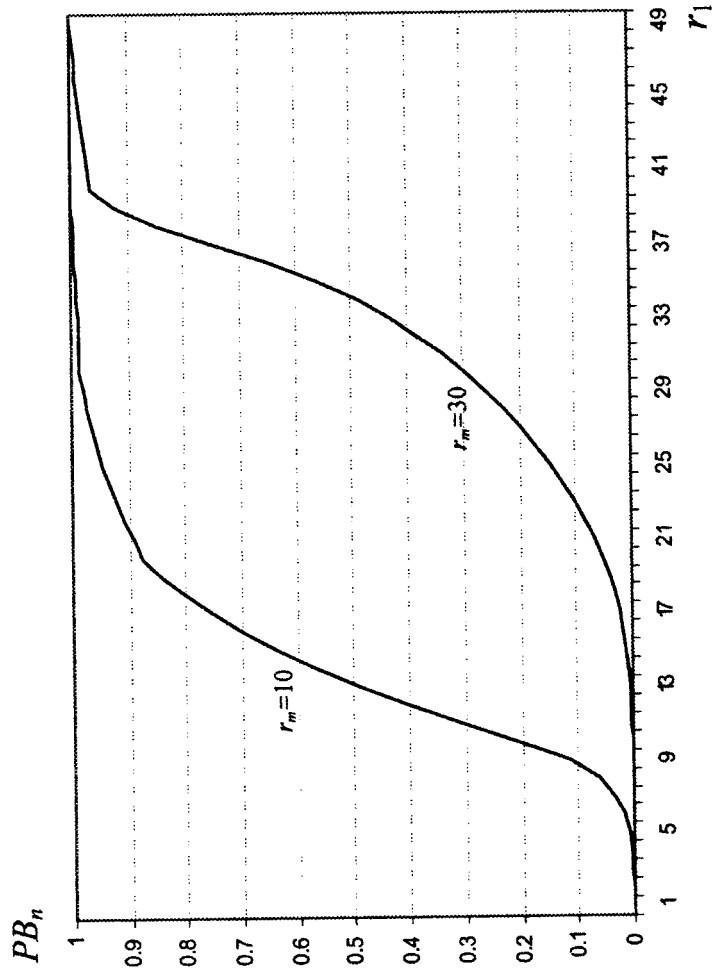


Рис. 1.25. Зависимость вероятности потери n -заявок от r_l при $N=50$; $m=10$; $\nu_1=25$; $\nu_2=20$

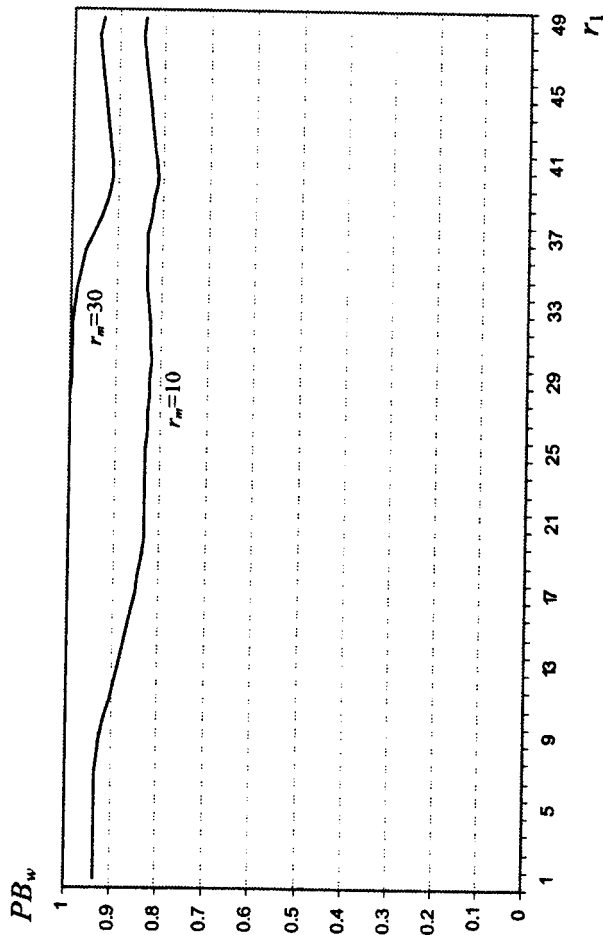


Рис. 1.26. Зависимость вероятности потери w -заявок от r_l при $N=50$; $m=10$; $\nu_1=25$; $\nu_2=20$

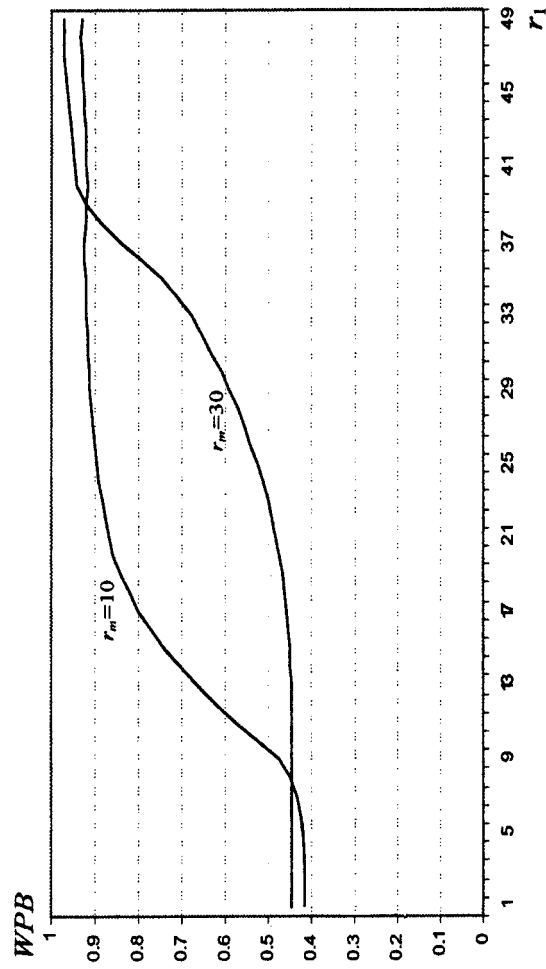


Рис. 1.27. Зависимость взвешенной суммы вероятностей потерь от r_l при $N=50$; $m=10$; $\nu_1=25$; $\nu_2=20$

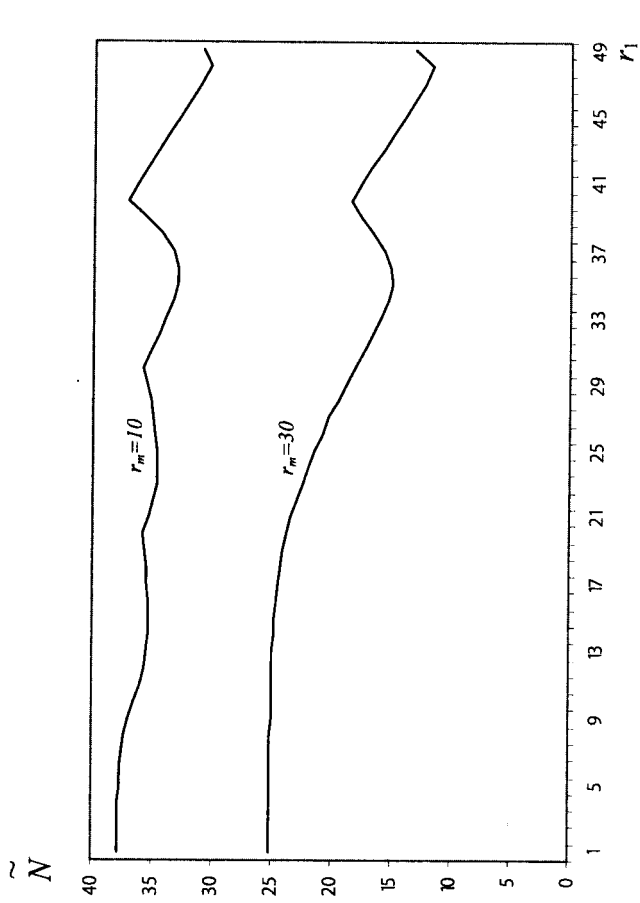


Рис. 1.28. Зависимость среднего числа занятых каналов от r_1 при $N = 50$; $m = 10$; $v_1 = 25$; $v_2 = 20$

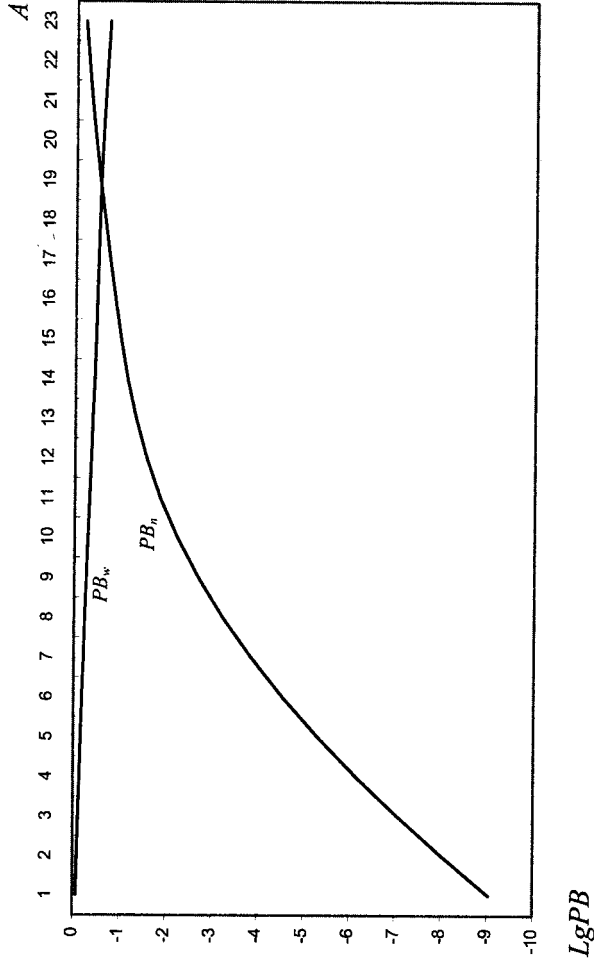


Рис. 1.29. Зависимость вероятности потери разнотипных заявок от параметра A при SGC-стратегии доступа:

$N = 80$, $m = 3$, $v_n = 30$, $v_w = 25$, $\bar{k}_w = 3$

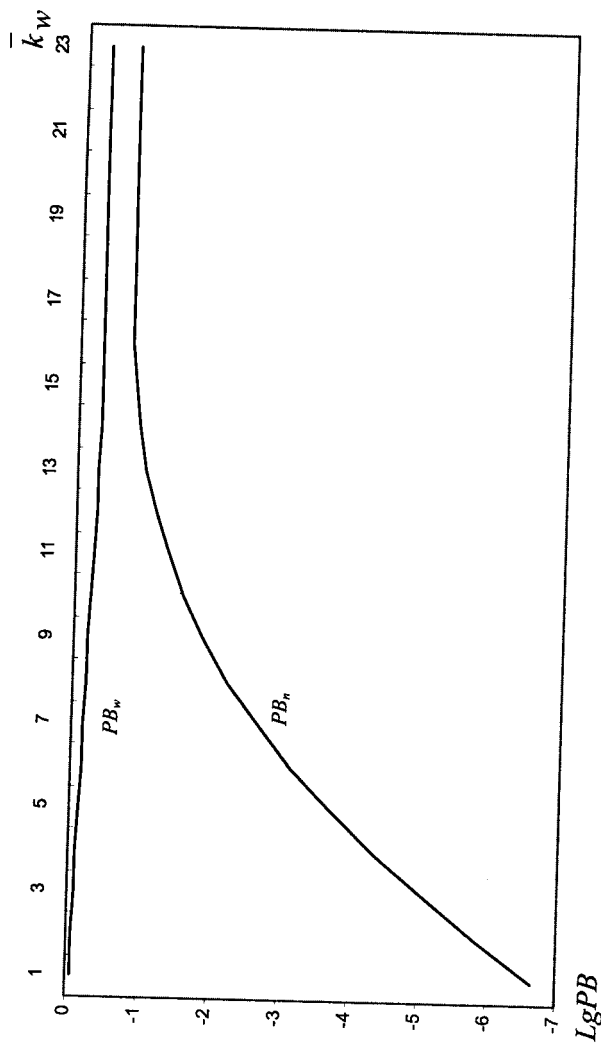


Рис. 1.30. Зависимость вероятности потери разнотипных заявок от параметра \bar{k}_w при SGC-стратегии доступа:

$N = 80$, $m = 3$, $A = 3$, $v_n = 30$, $v_w = 30$

— диапазон изменений функций $PB_n(SGC)$ и $PB_w(SGC)$ очень сильно отличается друг от друга. Так, максимальное значение $PB_w(A)$ (т.е. $PB_w(1) = 0.85$) почти в 4.5 раза больше ее минимального значения (т.е. $PB_w(23) = 0.19$), а соответствующие значения $PB_n(23) = 0.65$ и $PB_n(1) = 0.1E-10$, т.е. $PB(23) / PB(1) \approx 10^{10}$. Такая же картина имеет место при изучении зависимости этих функций от k_w .

— наиболее справедливое обслуживание в смысле максимальной близости значений $PB_n(A)$ и $PB_w(A)$ достигается при изучении зависимостей этих функций от k_w .

— наиболее справедливое обслуживание в смысле максимальной близости значений $PB_n(A)$ и $PB_w(A)$ достигается при $A = 19$, и при этом $PB_n(19) = 0.31$ и $PB_w(19) = 0.29$.

Заключение и комментарии к главе 1 части II

Обзор работ по MRQ по состоянию на 1995 г. можно найти в [3, 17]. Анализ литературы после указанного срока полностью подтвердил высказанное в [3] предположение о том, что исследование моделей MRQ окажутся центральными в прикладной теории систем массового обслуживания. Наиболее изученными являются модели MRQ с неэластичными заявками. Сохраняя хронологическую последовательность появления работ, следует в первую очередь отметить работу L. Gimpelson [15]. В ней исследуется следующая мо-

дель марковской MRQ . На вход N -канальной системы поступают узкополосные и широкополосные заявки, при этом широкополосные заявки требуют в m раз больше каналов, чем узкополосные (m — известное натуральное число). В системе принята полноступенчатая SAC (*Complete Sharing, CS-схема*), т.е. если в момент поступления заявки любого типа имеется необходимое число свободных каналов, то поступившая заявка принимается в систему; в противном случае поступившая заявка теряется. Для нахождения вероятностей потерь заявок каждого типа используется система уравнений равновесия, при этом для численного решения этой СУР используется итерационная процедура Гаусса-Зейделя. Исследуется также модель, в которой для ожидания в очереди широкополосных заявок имеется ограниченный накопитель. В работе приведены результаты объемных вычислительных экспериментов.

В [16, 26] показано, что стационарное распределение общей модели MRQ имеет мультипликативный вид. Согласно [10] мультипликативное решение существует и в тех случаях, когда функция распределения времени обслуживания i -заявок является произвольной со средним значением μ_i , $i = 1, \dots, M$. Оно остается в силе и для моделей MRQ типа Энгсета, когда разнотипные заявки поступают из конечных источников.

В работах Kaufman [16] и Roberts [26] почти одновременно и независимо друг от друга предложен оригинальный подход для расчета показателей QoS многоскоростных систем. Краткое описание этого подхода состоит в следующем.

Все микросостояния $n \in S$, в которых общее число занятых каналов равно r , описываются одним укрупненным состоянием r , а стационарная вероятность укрупненного состояния r обозначается через $\rho(r)$, $r = 0, \dots, N$. Тогда, показатели QoS вычисляются через стационарное распределение $\rho(r)$, $r = 0, \dots, N$ следующим образом:

$$PB_i = \sum_{j=0}^{b_i-1} \rho(N-j), \quad i = 1, \dots, M;$$

$$\tilde{N} = \sum_{r=1}^N r \rho(r)$$

В указанных работах Kaufman и Roberts доказали, что стационарное распределение $\rho(r)$, $r = 0, \dots, N$ удовлетворяет следующей системе уравнений:

$$\sum_{i=1}^M v_i b_i \rho(r - b_i) = r \rho(r), \quad r = 1, \dots, N,$$

$$\sum_{r=0}^N \rho(r) = 1,$$

где $\rho(x) = 0$ при $x < 0$.

В дальнейшем этот алгоритм был использован в работе [4] для разработки более эффективного алгоритма расчета указанных выше показате-

лей QoS . Совсем недавно в работе [29] был предложен подобный [4] алгоритм, состоящий из следующих шагов.

Шаг 1. Вычисляются следующие величины:

$$g_0 := 1; \quad g_j = 0 \quad \text{при } j < 0;$$

$$g_r = \frac{1}{r} \sum_{j=1}^M b_j v_j g_{r-b_j}, \quad r = \overline{1, N}.$$

Шаг 2. Вычисляются вероятности потери разнотипных заявок:

$$PB_i = \left(\sum_{j=N-b_j+1}^N g_j \right) / \left(\sum_{j=0}^N g_j \right), \quad i = \overline{1, M}.$$

Как видно из описания данного алгоритма, он почти полностью совпадает с алгоритмом, разработанным в [4]. Далее в [7] разработан усовершенствованный вариант алгоритма [4], который был опубликован почти одновременно с работой [29]. Как показано в 1.1, данный алгоритм по простоте не проигрывает алгоритму [29], хотя в [29] авторы отмечают, что их алгоритм является лучшим среди всех известных алгоритмов решения данной задачи.

Подобные алгоритмы также предложены в главе 2 книги [2], но эти алгоритмы, по словам самих авторов, при очень больших значениях N и M работают неэффективно. Вместе с тем, предложенный в [7] алгоритм позволяет осуществить решение указанной задачи при любых значениях N и M .

Модели MRQ с чистыми потерями были исследованы при более сложных SAC в работах [1, 6, 9]. В работе [24] исследуется модель MRQ , в которой SAC основана на стратегии резервирования каналов. С использованием алгоритма Kaufman-Roberts в работе [24] показано, что стационарное распределение $\rho(r)$, $r = 0, \dots, N$ удовлетворяет следующей системе уравнений:

$$\sum_{j=i}^M v_j b_j \rho(i - b_j) G_j(i - b_j) = i \rho(i), \quad i = \overline{1, N},$$

$$\sum_{i=0}^N \rho(i) = 1,$$

где $\rho(x) = 0$ при $x < 0$;

$$G_i(j) = \begin{cases} 1, & \text{если } j \leq N - b_i - r_i, \\ 0 & \text{в остальных случаях.} \end{cases}$$

В работе [18] на основе данного алгоритма был разработан эффективный способ расчета показателей QoS модели MRQ при использовании SAC , основанной на TR -схеме доступа. В дальнейшем в работе [8] был разработан усовершенствованный вариант данного алгоритма.

С целью организации абсолютно справедливо обслуживания (в смысле равенства вероятностей потерь разнотипных заявок) в MRQ воз-

можно использование *SAC*, основанной на полностью доступной стратегии с уравниванием (*Complete Sharing with Equalization, CSE*). Такая модель была исследована в работе [13]. В работе [13] показано, что стационарное распределение ($\rho(r)$, $r = 0, \dots, N$) удовлетворяет следующей системе уравнений:

$$\sum_{j=1}^M v_j b_j \rho(i - b_j) F_j(i - b_j) = i \rho(i), \quad i = 1, \dots, N;$$

$$\sum_{i=0}^N \rho(i) = 1,$$

где $\rho(x) = 0$ при $x < 0$;

$$F_j(i - b_j) = \begin{cases} 1, & \text{если } i \leq N - b_j, \\ 0 & \text{в остальных случаях.} \end{cases}$$

(В 1.1 показано, что последняя формула содержит ошибку, см. Замечание 1.4).

В подавляющем большинстве работ, посвященных исследованию *MRQ*, изучаются модели, в которых ресурсы (каналы) системы полностью коммутируются, т.е. для которых можно считать, что для обслуживания i -заявки могут быть выделены любые b_i свободных каналов, иными словами, при обслуживании заявок, каналы могут работать в любых сочетаниях. Вместе с тем, существует достаточно обширный класс *MRQ*, в которых вследствие некоторых физических и/или экономических особенностей каналы системы могут работать лишь в определенных сочетаниях. Так, например, в некоторых интегральных сетях при передаче разнородной информации используется *SAC* с размещением широкополосных заявок в заранее определенных местах кадра, а узкополосные заявки могут размещаться в любом месте кадра. Такие модели называются системами с частично коммутируемыми каналами [5].

В ряде работ модели *MRQ* изучаются с применением асимптотических методов, при этом их асимптотические свойства исследуются при $N \rightarrow \infty$, $K \rightarrow \infty$, при критических нагрузках входящих трафиков, а также при одновременном увеличении N, K и нагрузок [11, 12, 14, 19-23, 25, 27, 28, 30].

Отметим, что здесь приведены лишь некоторые результаты, полученные для *MRQ* с чистыми потерями. Достаточно подробное изложение этих результатов можно найти в работах [1-3, 17, 19-22, 30].

Литература к главе 1 части II

1. Башарин Г.П., Бочаров П.П., Коган Я.А. Анализ очередей в вычислительных сетях. Теория и методы расчета. — М.: Наука, 1989. — 336 с.
2. Лагутин В.С., Степанов С.Н. Телеграфик мультисервисных сетей связи. — М.: Радио и связь, 2000. — 320 с.

3. Меликов А.З. Методы расчета и оптимизации моделей мультисервисных систем обслуживания // Кибернетика и системный анализ. — 1996. — № 6. — С. 92 — 112.
4. Меликов А.З. Об одном алгоритме расчета мультисервисных систем обслуживания // Электронное моделирование. — 1992. — т. 14, № 5. — С. 52 — 56.
5. Меликов А.З., Молчанов А.А., Пономаренко Л.А. Мультисервисные системы массового обслуживания с частично коммутируемыми каналами // Электронное моделирование. — 1992. — т. 14, № 2. — С. 87 — 91.
6. Меликов А.З., Фаттахова М.И., Казиев Т.С. Численные методы исследования многоскоростных систем обслуживания типа Гимпельсона // System research & information technologies. — 2005. — № 2. — С. 83 — 96.
7. Меликов А.З., Фаттахова М.И., Казиев Т.С. Метод исследования характеристик многоскоростных сетей обслуживания с одним ресурсом и потерями // Электронное моделирование. — 2005. — т. 27, № 5. — С. 35 — 45.
8. Меликов А.З., Фаттахова М.И., Казиев Т.С. Анализ и оптимизация многоскоростной системы обслуживания со стратегией резервирования каналов // Электронное моделирование. — 2006. — т. 28, № 5. — С. 71 — 85.
9. Меликов А.З., Фаттахова М.И., Казиев Т.С. Многоскоростная система со специализированными каналами для обслуживания широкополосных заявок // Автоматика и вычислительная техника. — 2006. — № 2. — С. 15 — 25.
10. Burman D.Y., Lehoczky J.P., Lim Y. Insensitivity of blocking probabilities in a circuit-switching network // J. Appl. Prob. — 1984. — Vol. 21, No. 4. — pp. 850 — 859.
11. Choudhury G.L., Leung K.K., Whitt W. An inversion algorithm to compute blocking probabilities in loss networks with state-dependent rates // IEEE/ACM Transactions on Networking, October 1995. — Vol. 3, No. 5. — pp. 585 — 601.
12. Chung S., Ross K.W. Reduced load approximations for multirate loss networks // IEEE/ACM Transactions on Networking. — 1993. — Vol. 41, No. 8. — pp. 1222 — 1231.
13. Delaire M., Hebuterne G. Call blocking in multi-services systems on one transmission link // 5th Int. Workshop on Performance model and Evaluation of ATM Networks, July 1997, UK. — pp. 253 — 270.
14. Gazdzicki P., Lambadaris I., Mazumdar R.R. Blocking probabilities for large multirate Erlang loss system // Adv. Appl. Prob. — 25. — 1993. — pp. 997 — 1003.
15. Gimpelson L.A. Analysis of mixtures of wide- and narrow-band traffic // IEEE Trans. Commun. Technol. — 1965. — Vol. 13, No. 3. — pp. 258 — 266.

160
175

ИССЛЕДОВАНИЕ МОДЕЛЕЙ МНОГОСКОРОСТНЫХ СИСТЕМ С НЕЭЛАСТИЧНЫМИ И ЭЛАСТИЧНЫМИ ЗАЯВКАМИ

В настоящей главе исследуются модели многоскоростных систем, в которых обслуживаются неэластичные (n -э-заявки) и эластичные ($э$ -заявки) заявки. Здесь предполагается, что неэластичные заявки во всем периоде обслуживания занимают одновременно точно $b \geq 1$ каналов системы, т.е. как и в случаях, рассмотренных в главе 1, все b каналов начинают и завершают обслуживание одной неэластичной заявки одновременно. A эластичные заявки занимают одновременно m каналов, $\underline{m} \leq m \leq \bar{m}$, где границы изменения m (т.е. \underline{m} и \bar{m}) являются заданными величинами.

Вначале рассматриваются простые модели, в которых в любой момент времени на обслуживании может находиться не более одной эластичной заявки, и данная заявка занимает все свободные каналы системы; при этом интенсивность обслуживания пропорционально зависит от числа каналов, занятых обслуживанием $э$ -заявки. Подробно исследуются модели двух типов: с равноправным доступом для заявок обоих типов и с ограниченным доступом для неэластичных заявок.

Далее рассматриваются более сложные модели, в которых на обслуживании может находиться не более N эластичных заявок, где N общее число каналов (полос). При этом различаются модели двух типов: с непрерывной полосой и с дискретной полосой для эластичных заявок.

Для всех исследуемых моделей разрабатываются эффективные численные методы и алгоритмы расчета их характеристик, даны результаты вычислительных экспериментов и их содержательный анализ.

2.1. Модели с одной эластичной заявкой на обслуживании

Модели данного типа имеют такие основные свойства. На вход N -канальной ($N > 1$) системы поступают пуассоновские потоки заявок двух типов: неэластичные и эластичные заявки. Интенсивность поступления n -э-заявок ($э$ -заявок) равна $\lambda_1(\lambda_2)$.

Число каналов, требуемых для обслуживания одной n -э-заявки, является постоянной величиной и равно b , $1 \leq b \leq N$, т.е. для начала обслуживания поступившей n -э-заявки потребуются не менее b свободных каналов системы, причем все b каналов начинают и завершают обслуживание данной заявки одновременно. Время обслуживания n -э-заявки распределено экспоненциально с параметром μ .

Обслуживание одной $э$ -заявки происходит следующим образом. Поступившая $э$ -заявка получает отказ (блокируется), если в момент ее поступления либо с системе уже имеется одна $э$ -заявка, либо все каналы

16. Kaufman J.S. Blocking in shared resource environment // IEEE Trans. Commun. — 1981. — Vol. 10, No. 10. — pp. 1474 — 1481.
17. Kelly F.P. Loss networks // The Annals of Applied Probability. — 1991. — Vol. 1, No. 3. — pp. 319 — 378.
18. Melikov A.Z., Deniz D.Z. Non-exhaustive channel access strategy in multi-resource communication systems with non-homogeneous traffic // Proc. of 5th IEEE Symposium on Computers and Communications, July 3-6, 2000, France. — pp. 432 — 437.
19. Mitra D., Morrison J.A. Erlang capacity and uniform approximations for shared unbuffered resources // IEEE/ACM Transactions on Networking, December 1994. — Vol. 2, No. 6. — pp. 558 — 570.
20. Mitra D., Morrison J.A., Ramakrishnan K.G. ATM Network design and optimization: A multirate loss network framework // IEEE/ACM Transactions on Networking, August 1996. — Vol. 4, No. 4. — pp. 531 — 543.
21. Mitra D., Morrison J.A., Ramakrishnan K.G. Optimization and design of network routing refined asymptotic approximations // Performance Evaluation, 36-37. — 1999. — pp. 267 — 288.
22. Morrison J.A. Loss probabilities in a simple circuit-switched network // Adv. Appl. Prob. — 26. — 1994. — pp. 456 — 473.
23. Omahen K.J. Capacity bounds for multiresource queues // Journ. ACM. — 1977. — Vol. 24, No. 4. — pp. 646 — 663.
24. Piro M., Lubacs J., Korner U. Traffic engineering problems in multi-service circuit-switched networks // Comput. Networks and ISDN Syst. — 1990. — Vol. 1-5. — pp. 127 — 136.
25. Puhalskii A.A., Reiman M.I. A critically loaded multirate link with trunk reservation // Queueing Systems. — 28. — 1998. — pp. 157 — 190.
26. Roberts J.W. A service system with heterogeneous user requirements application to multi-service telecommunications systems // Performance of data communication system and their applications. (Pujolle G., ed.) North Holland, 1981.
27. Ross K.W., Tsang D.H. Teletraffic engineering for product-form circuit-switched networks // Adv. Appl. Prob. — 1990. — Vol. 22, No. 3. — pp. 657 — 675.
28. Simonian A., Roberts J.W., Theberge F., Mazumdar R. Asymptotic estimates for blocking probabilities in a large multi-rate loss network // Adv. Appl. Prob. — 29. — 1997. — pp. 806 — 829.
29. Taha S., Kavehrad M. Dynamic bandwidth allocation in multi-class connection-oriented networks // Computer Communications. — 27. — 2004. — pp. 13 — 26.
30. Whitt W. Blocking when service is required from several facilities simultaneously // AT&T Technical Journal. — 1985. — Vol. 64, No. 8. — pp. 1807 — 1856.

системы заняты обслуживанием n -э-заявок. Отсюда следует, что в произвольный момент времени в системе не может быть более одной э-заявки.

Если в момент поступления э-заявки в системе имеются m свободных каналов, $1 \leq m \leq N$, то данная заявка занимает все свободные каналы и начинается ее обслуживание всеми этими каналами одновременно. При этом время обслуживания э-заявки одним каналом, имеющее показательное распределение, равно μ_i^{-1} , и весь период обслуживания пропорционально зависит от числа занимаемых каналов.

Если в момент поступления n -э-заявки в системе имеется э-заявка, то необходимое число каналов переключается для обслуживания поступившей n -э-заявки, при этом уменьшается интенсивность обслуживания э-заявки. Здесь также возможны различные схемы переключения каналов. Ниже рассматриваются две наиболее важные схемы, и при этом для простоты изложения и не нарушая общности предполагается, что $b = 1$.

2.1.1. Модель с равноправным доступом для заявок обоих типов

Здесь предполагается, что заявки обоих типов (т.е. неэластичные и эластичные) имеют равноправный доступ к каналам системы, и при этом не допускается вытеснение заявок любого типа из каналов.

Исходя их вышеизложенного заключаем, что состояние данной системы в произвольный момент времени может задаваться двумерным вектором $\mathbf{k} = (k_1, k_2)$, где k_1 (k_2) означает число неэластичных (эластичных) заявок в системе, при этом $k_1 = 0, 1, 2, \dots, N$ и $k_2 = 0, 1$. Тогда ФПС системы определяется так:

$$S' := \{ \mathbf{k}: k_1 = 0, 1, \dots, N; k_2 = 0, 1; k_1 + k_2 \leq N \}. \quad (2.1)$$

Определим возможные переходы между состояниями ФПС (2.1) и их интенсивности. Поскольку переходы между состояниями возможны лишь в моменты поступления заявок и их уходов из системы, то рассмотрим эти события в отдельности.

При поступлении n -э-заявки происходит переход из состояния \mathbf{k} в состояние $\mathbf{k} + \mathbf{e}_1$, если $k_1 \leq N - 1$ и $k_2 = 0$ или $k_1 \leq N - 2$ и $k_2 = 1$; при поступлении э-заявки происходит переход из состояния \mathbf{k} в состояние $\mathbf{k} + \mathbf{e}_2$, если $k_1 \leq N - 1$ и $k_2 = 0$. Интенсивность перехода при поступлении n -э-заявки и э-заявки равна λ_1 и λ_2 , соответственно.

При уходе n -э-заявки из системы в состоянии \mathbf{k} происходит переход в состояние $\mathbf{k} - \mathbf{e}_1$, а уход э-заявки в состоянии \mathbf{k} порождает переход в состояние $\mathbf{k} - \mathbf{e}_2$. Интенсивность перехода при уходе n -э-заявки и э-заявки в состоянии \mathbf{k} равна $k_1 \mu_1$ и $(N - k_1) \mu_2$, соответственно.

Таким образом, элементы ПМ двумерной марковской цепи, описывающей функционирование данной системы, определяются так:

$$q(\mathbf{k}, \mathbf{k}') = \begin{cases} \lambda_1, & \text{если } k_1 \leq N - 1, k_2 = 0 \text{ или } k_1 \leq N - 2, k_2 = 1, \mathbf{k}' = \mathbf{k} + \mathbf{e}_1, \\ \lambda_2, & \text{если } k_1 \leq N - 1, k_2 = 0, \mathbf{k}' = \mathbf{k} + \mathbf{e}_2, \\ k_1 \mu_1, & \text{если } \mathbf{k}' = \mathbf{k} - \mathbf{e}_1, \\ (N - k_1) \mu_2, & \text{если } \mathbf{k}' = \mathbf{k} - \mathbf{e}_2, \\ 0 & \text{в остальных случаях.} \end{cases} \quad (2.2)$$

Принцип построения графа переходов между состояниями модели показан на рис.2.1.

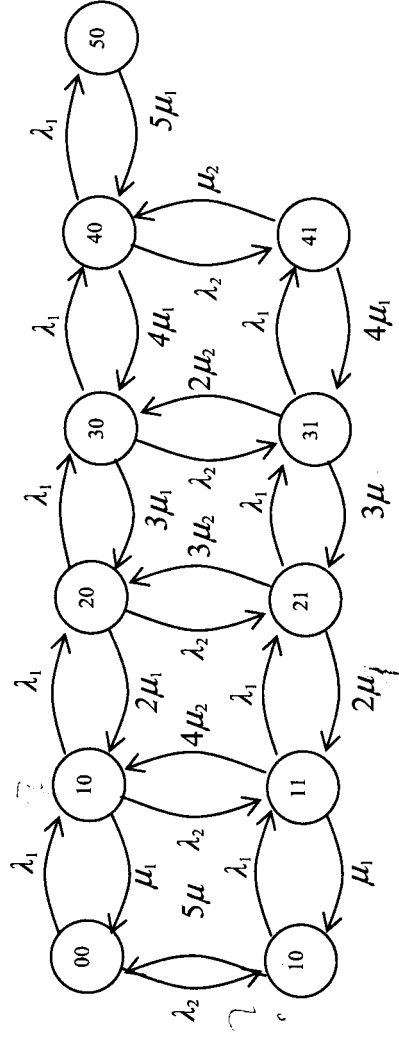


Рис.2.1. Фрагмент графа переходов между состояниями модели с равноправным доступом для заявок обоих типов, $N = 5$

Замечание 2.1. В данной главе для иллюстрации графа переходов между состояниями исследуемых моделей используется способ, отличный от того, который использовался в главе 1. Этот способ иллюстрации окажется полезным ниже при проверке наличия мультипликативного решения соответствующих СУР.

Стационарное распределение данной модели находится из стандартной СУР, которая составляется на основе соотношений (2.2). Она имеет следующий вид:

$$\begin{aligned} & \text{для состояний } \mathbf{k} \neq (N, 0) \text{ и } \mathbf{k} \neq (N - 1, 1): \\ & (\lambda_1 + \lambda_2) I(k_2 \neq 0) + k_1 \mu_1 + (N - k_1) \mu_2 p(\mathbf{k}) = \lambda_1 p(\mathbf{k} - \mathbf{e}_1) I(k_1 \neq 0) + \\ & + \lambda_2 p(\mathbf{k} - \mathbf{e}_2) \delta(k_2, 0) + (k_1 + 1) \mu_1 p(\mathbf{k} + \mathbf{e}_1) + (N - k_1) \mu_2 p(\mathbf{k} + \mathbf{e}_2) \delta(k_2, 0); \end{aligned} \quad (2.3)$$

для состояния $\mathbf{k} = (N, 0)$:

$$N \mu_1 p(\mathbf{k}) = \lambda_1 p(\mathbf{k} - \mathbf{e}_1); \quad (2.4)$$

для состояния $\mathbf{k} = (N - 1, 1)$:

$$((N - 1) \mu_1 + \mu_2) p(\mathbf{k}) = \lambda_1 p(\mathbf{k} - \mathbf{e}_1) + \lambda_2 p(\mathbf{k} - \mathbf{e}_2). \quad (2.5)$$

К этой системе уравнений добавляется ещё и нормирующее условие:

$$\sum_{\mathbf{k} \in S^1} p(\mathbf{k}) = 1. \quad (2.6)$$

Показатели качества обслуживания (QoS) разнотипных заявок в данной модели также оцениваются с помощью вероятностей их блокировки, обозначаемых через $PB_{н.э.}$ и PB_3 для н.э- и э-заявок, соответственно.

Из описания модели видно, что н.э-заявки любого типа теряются лишь тогда, когда все каналы системы заняты, т.е.

$$PB_{н.э.} = p(N, 0) + p(N-1, 1). \quad (2.7)$$

Эластичная заявка теряется, если в момент ее поступления либо все каналы системы заняты неэластичными заявками, либо в этот момент в системе уже обслуживается одна заявка данного типа. Отсюда получаем, что вероятность потери э-заявки определяется так:

$$PB_3 = p(N, 0) + PB_3 = \sum_{i=0}^{N-1} p(i, 1). \quad (2.8)$$

Другая характеристика данной модели – среднее число занятых каналов (\bar{N}), также определяется через ее стационарное распределение:

$$\bar{N} = N \cdot \sum_{i=0}^{N-1} p(i, 1) + \sum_{i=1}^N ip(i, 0). \quad (2.9)$$

Следовательно, для нахождения характеристик (2.7) – (2.9) необходимо определить стационарное распределение модели. Однако, наличие стационарного распределения модели с помощью СУР при больших значениях N связано с определенными вычислительными трудностями, т.к. указанная СУР не имеет мультипликативного решения. Для доказательства последнего факта можно использовать известный критерий Колмогорова для двумерных цепей Маркова (см. Часть 1).

Действительно, рассмотрим фрагмент графа, показанного на рис.2.1. Выберем четыре соседних состояния (см. рис.2.2).

В качестве начального состояния выберем одно из этих четырех состояний. Пусть для определенности выбирается состояние (1, 0). Тогда, выходя из этого состояния по часовой стрелке и прогив нее, вычисляются производения соответствующих интенсивностей переходов: по часовой стрелке – $\lambda_1 \cdot \lambda_2 \cdot 2\mu_1 \cdot 4\mu_2$; против часовой стрелки – $\lambda_2 \cdot \lambda_1 \cdot 3\mu_2 \cdot 2\mu_1$. Легко видеть, что эти величины не равны друг другу, т.е. в сети существует ненулевая циркуляция, что означает отсутствие мультипликативного решения СУР (2.3) – (2.6).

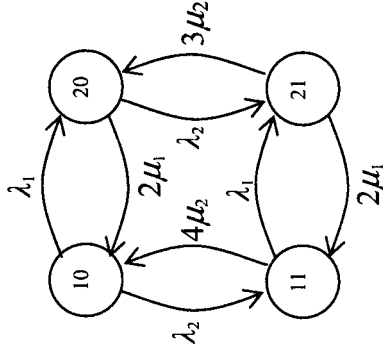


Рис.2.2. Фрагмент графа из рис.2.1

С целью корректного применения приближенного подхода предположим, что интенсивность трафика одного типа существенным образом превосходит соответствующий показатель другого трафика. Для определенности изложим предположим, что $\lambda_1 \gg \lambda_2$ (если выполняется условие $\lambda_1 \ll \lambda_2$, то описанная ниже процедура модифицируется очевидным способом).

Тогда рассмотрим следующее разбиение ФПС (2.1):

$$S^1 = S_0^1 \cup Y S_1^1, \quad S_0 \cap S_1 = \emptyset,$$

где $S_r^1 := \{k \in S, k_2 = r\}$, $r = 0, 1$, (т.е. осуществляется разбиение по строкам, см. рис.2.1).

Опуская детальное описание известных этапов указанной процедуры, отметим, что стационарные распределения внутри классов S_0^1 и S_1^1 полностью совпадают с аналогичными распределениями классических систем Эрланга $M|M|N|0$ и $M|M|N-1|0$ с нагрузкой ν_i эрл, соответственно, где $\nu_i := \lambda_1/\mu_i$. Для краткости изложения эти распределения обозначим $\theta_0(i)$, $i = 0, \dots, N$, и $\theta_1(j)$, $j = 0, N-1$, соответственно, т.е.

$$\theta_0(i) = \frac{\nu_i^i}{i!} \left(\sum_{r=0}^N \frac{\nu_r^r}{r!} \right)^{-1}, \quad i = 0, \dots, N; \quad (2.10)$$

$$\theta_1(j) = \frac{\nu_j^j}{j!} \left(\sum_{r=0}^{N-1} \frac{\nu_r^r}{r!} \right)^{-1}, \quad j = 0, \dots, N-1. \quad (2.11)$$

Следовательно, с учетом (2.2) и (2.10), (2.11) находим, что интенсивности переходов между двумя укрупненными состояниями $\langle 0 \rangle$ и $\langle 1 \rangle$, обозначаемые $q(\langle r \rangle, \langle r' \rangle)$, определяются так:

$$q(\langle r \rangle, \langle r' \rangle) = \begin{cases} \lambda_2(1 - E_B(v_1, N)), & \text{если } r = 0, r' = 1, \\ \mu_2 \sum_{i=0}^{N-1} (N-i)\theta(i), & \text{если } r = 1, r' = 0. \end{cases} \quad (2.12)$$

Тогда, из (2.12) находим, что стационарные вероятности укрупненных состояний $\langle 0 \rangle$ и $\langle 1 \rangle$, обозначаемые $\pi(\langle 0 \rangle)$ и $\pi(\langle 1 \rangle)$, вычисляются так:

$$\pi(\langle 0 \rangle) = \frac{q(\langle 1 \rangle, \langle 0 \rangle)}{q(\langle 1 \rangle, \langle 0 \rangle) + q(\langle 0 \rangle, \langle 1 \rangle)}; \quad (2.13)$$

$$\pi(\langle 1 \rangle) = \frac{q(\langle 0 \rangle, \langle 1 \rangle)}{q(\langle 1 \rangle, \langle 0 \rangle) + q(\langle 0 \rangle, \langle 1 \rangle)}. \quad (2.14)$$

Стационарное распределение исходной модели приближенно определяется с помощью (2.10), (2.11) и (2.13), (2.14).

Характеристики модели (2.7)–(2.9) приближенно определяются следующим образом:

$$PB_{n,3} \approx E_B(v_1, N)\pi(\langle 0 \rangle) + E_B(v_1, N-1)\pi(\langle 1 \rangle); \quad (2.15)$$

$$PB_3 \approx E_B(v_1, N)\pi(\langle 0 \rangle) + \pi(\langle 1 \rangle); \quad (2.16)$$

$$\tilde{N} \approx N\pi(\langle 1 \rangle) + \sum_{r=1}^N r\theta(r)\pi(\langle 0 \rangle). \quad (2.17)$$

Таким образом, вычисление характеристик модели (2.7)–(2.9) при любых значениях N осуществляется с помощью простых формул (2.15)–(2.17).

2.1.2. Модели с ограниченным доступом для неэластичных заявок

Теперь рассмотрим обобщение модели, исследованной выше. Здесь предполагается, что обслуживание э-заявки может начинаться лишь тогда, когда число свободных каналов не меньше, чем R , где $R \geq 1$. Обслуживание э-заявок со скоростью ниже $R\mu_2$ приводит к существенному ухудшению качества их передачи.

В обобщенной модели состояние системы в произвольный момент времени также описывается двумерным вектором $\mathbf{k} = (k_1, k_2)$, где $k_1(k_2)$ означает число неэластичных (эластичных) заявок в системе, при этом $k_1 = 0, 1, 2, \dots, N$ и $k_2 = 0, 1$. Фазовое пространство состояний модели (S^*) является подмножеством множества (2.1) и определяется следующим образом:

$$S^* := S \setminus \{\mathbf{k} \in S: k_1 \geq N - R + 1, k_2 = 1\}. \quad (2.18)$$

Равенство (2.18) означает, что ФПС обобщенной модели не содержит состояний типа $\mathbf{k} = (k_1, k_2)$, где $k_1 \geq N - R + 1, k_2 = 1$. Это означает, что если система находится в состоянии $\mathbf{k} = (k_1, k_2)$, в котором $k_2 = 1$ и $k_1 = N - R$, то поступившая неэластичная заявка не принимается на обслуживание, так как в противном случае скорость обслуживания эластичной заявки была бы меньше, чем $R\mu_2$, что не допускается из-за ухудшения качества обслуживания э-заявки. Иными словами, в данной модели вводится ограничение для доступа неэластичных заявок в каналы.

Замечание 2.2. В частном случае $R = 1$ из данной модели получается модель, описанная в § 2.1.1.

Интенсивности переходов между состояниями обобщенной модели определяются аналогично (2.2) с соответствующими изменениями. Из-за очевидности способа их нахождения они здесь не приводятся.

Граф переходов между состояниями данной модели показан на рис.2.3.

Система уравнений равновесия для стационарного распределения обобщенной модели ($p^*(\mathbf{k}): \mathbf{k} \in S^*$) составляется аналогично (2.3)–(2.6). Однако ввиду очевидности ее составления явный вид этой СУР здесь не приводится.

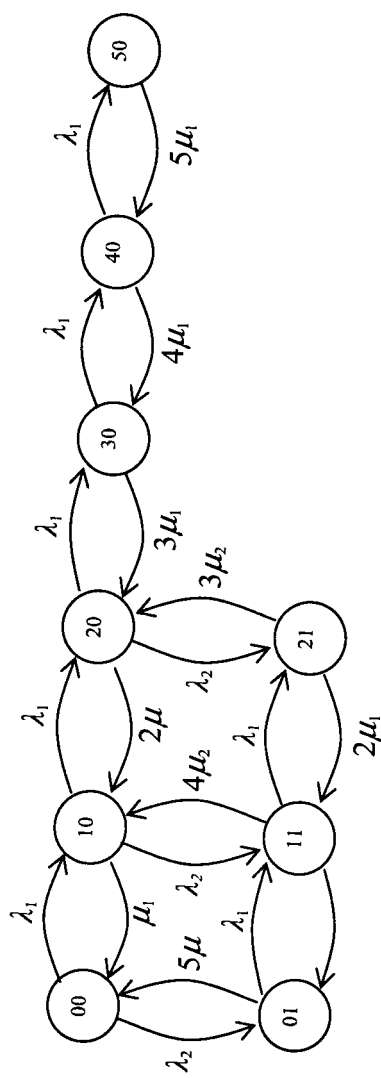


Рис.2.3. Граф переходов между состояниями модели с ограниченным доступом для неэластичных заявок, $N = 5, R = 3$

Вероятности потери неэластичных и эластичных заявок ($PB_{n,3}^*$, PB_3^*) и среднее число занятых каналов (\tilde{N}^*) определяется аналогично (2.7)–(2.9):

$$PB_{n,3}^* = p^*(N, 0) + p^*(N - R, 1); \quad (2.19)$$

$$PB_3^* = \sum_{i=0}^{N-R} p^*(i,1) + \sum_{i=N-R+1}^N p^*(i,0); \quad (2.20)$$

$$\tilde{N}^* = N \sum_{i=0}^{N-R} p^*(i,1) + \sum_{i=1}^N ip^*(i,0). \quad (2.21)$$

Здесь также с помощью использования известного критерия Колмогорова можно показать, что соответствующая СУР для данной модели не имеет стационарного распределения мультипликативного вида. Поэтому, необходимо разработать эффективную вычислительную процедуру для нахождения характеристик (2.19)–(2.21). Применяя описанный выше подход можно предложить следующий алгоритм для приближенного расчета указанных характеристик (ввиду очевидности промежуточные математические преобразования опускаются).

Шаг 1. Вычислить

$$\pi^*(\langle 0 \rangle) = \frac{q^*(\langle 1 \rangle, \langle 0 \rangle)}{q^*(\langle 1 \rangle, \langle 0 \rangle) + q^*(\langle 0 \rangle, \langle 1 \rangle)};$$

$$\pi^*(\langle 1 \rangle) = \frac{q^*(\langle 0 \rangle, \langle 1 \rangle)}{q^*(\langle 1 \rangle, \langle 0 \rangle) + q^*(\langle 0 \rangle, \langle 1 \rangle)};$$

где

$$q^*(\langle i \rangle, \langle i' \rangle) = \begin{cases} \lambda_2 \sum_{j=0}^{N-R} \theta_0(j), & \text{если } i=0, i'=1, \\ \mu \sum_{j=0}^{N-R} (N-j) \theta_1(j), & \text{если } i=1, i'=0. \end{cases}$$

Величины $\theta_0(j)$ вычисляются согласно (2.10), величины $\theta_1(j)$ также вычисляются согласно (2.11), но в этой формуле верхний предел суммирования заменен на $N-R$.

Шаг 2. Вычислить

$$PB_{n,3}^* \approx E_B(v_1, N) \pi^*(\langle 0 \rangle) + E_B(v_1, N-R) \pi^*(\langle 1 \rangle); \quad (2.22)$$

$$PB_3^* \approx \pi^*(\langle 1 \rangle) + \sum_{i=N-R+1}^N \theta_0(i) \pi^*(\langle 0 \rangle); \quad (2.23)$$

$$\tilde{N}^* \approx N \pi^*(\langle 1 \rangle) + \sum_{i=1}^N \theta_0(i) \pi^*(\langle 0 \rangle). \quad (2.24)$$

Замечание 2.3. Из данного алгоритма в частном случае $R=1$ получается алгоритм расчета модели с равноправным доступом для заявок обоих типов (см. замечание 2.2).

2.2. Модели со многими эластичными заявками на обслуживании

Теперь рассмотрим более сложные модели MRQ с неэластичными и эластичными заявками. В них допускается, что на обслуживании одновременно могут находиться более одной э-заявки.

Вообще говоря, в реальных мультимедийных сетях обслуживания э-заявки могут выбирать свои полосы m между \underline{m} и \bar{m} , либо непрерывно, либо дискретно, где \underline{m} и \bar{m} являются границами изменения полосы для э-заявок, $\underline{m} \leq m$, $\bar{m} \in [1, M]$. Исходя из этого, здесь рассматриваются модели обоих типов в отдельности, при этом все обозначения предыдущих параграфов данной главы сохраняются без изменения.

2.2.1. Модели с непрерывной полосой для эластичных заявок

Основные отличия рассматриваемых здесь моделей от моделей, исследуемых в §§ 2.1.1 и 2.1.2, состоят в схеме организации обслуживания разнотипных заявок.

Сначала рассмотрим схему организации обслуживания э-заявок. Прежде всего отметим, что э-заявки равноправным образом занимают все те каналы системы, которые не заняты обслуживанием n -э-заявок, т.е. если э-заявка поступает в момент, когда все каналы свободны, то она занимает все эти каналы, и обслуживание данной э-заявки осуществляется с максимальной скоростью. При поступлении в систему новых э-заявок находящиеся на обслуживании заявки данного типа равномерно распределяют между собой доступные полосы (т.е. те полосы, которые не заняты n -э-заявками). Эта процедура продолжается до тех пор, пока количество э-заявок на обслуживании не станет равным N . Иными словами, поступившая э-заявка получает отказ, если (1) либо в момент ее поступления все каналы системы заняты обслуживанием n -э-заявок (напомним, что каждая n -э-заявка требует для своего обслуживания лишь один канал), (2) либо число э-заявок в системе равно N .

Теперь рассмотрим схему организации обслуживания n -э-заявок. Если в момент поступления n -э-заявки в системе отсутствуют э-заявки и имеется хотя бы один свободный канал, то данная n -э-заявка принимает-ся на обслуживание. Если в момент поступления n -э-заявки количество таких заявок в системе меньше, чем $N-1$, и на обслуживании находятся э-заявки, то один канал, занятый обслуживанием э-заявок, выделяется для обслуживания поступившей n -э-заявки, и, таким образом, уменьшается скорость обслуживания э-заявок. Иными словами, поступившая n -э-заявка получает отказ, если либо в момент ее поступления число таких заявок в системе равно $N-1$ и обслуживаются э-заявки (с суммарной интенсивностью μ_2 , так как полосы являются непрерывными, и все э-заявки равноправным образом используют доступные полосы), либо в этот момент в системе уже имеются N таких заявок.

После завершения обслуживания заявки любого типа освобожденные полосы равномерным образом распределяются между обслуживаемыми э-заявками, и, таким образом, увеличивается суммарная скорость их обслуживания.

Очевидно, что предложенные схемы организации доступа разнотипных заявок и использования освобожденных каналов позволяют увеличить утилизацию каналов, а также уменьшить вероятность потери разнотипных заявок по сравнению с классическими схемами обслуживания.

Состояние данной системы, как и для случая, рассмотренного в § 2.1, в произвольный момент времени можно описать двумерным вектором $\mathbf{k} = (k_1, k_2)$, где k_1 (k_2) означает число неэластичных (эластичных) заявок в системе, при этом $k_i = 0, 1, 2, \dots, N$, $i = 1, 2$. Однако, в отличие от (2.1), здесь ФПС системы определяется так:

$$S := \{\mathbf{k} : k_1 = 0, \dots, N-1, k_2 = 0, \dots, N\} \cup \{(N, 0)\}. \quad (2.25)$$

Исходя из механизма обслуживания разнотипных заявок, в данной модели возможные переходы между состояниями из ФПС (2.25) и их интенсивности определяются следующим образом.

При поступлении n -э-заявки происходит переход из состояния \mathbf{k} в состояние $\mathbf{k} + \mathbf{e}_1$, если $k_1 \leq N-1$ и $k_2 = 0$ или $k_1 \leq N-2$; при поступлении э-заявки происходит переход из состояния \mathbf{k} в состояние $\mathbf{k} + \mathbf{e}_2$, если $k_1 \leq N-1$ и $k_2 \leq N-1$. Интенсивность перехода при поступлении n -э-заявки и э-заявки равна λ_1 и λ_2 , соответственно.

Переход из состояния \mathbf{k} в состояние $\mathbf{k} - \mathbf{e}_1$ происходит в момент ухода из системы n -э-заявки, и интенсивность этого перехода равна $k_1 \mu_1$; при уходе из системы э-заявки в состоянии \mathbf{k} происходит переход в состояние $\mathbf{k} - \mathbf{e}_2$, и интенсивность этого перехода равна $(N - k_1) \mu_2$.

Следовательно, элементы ПМ двумерной цепи Маркова, описывающей работу данной MRQ, определяются так:

$$q(\mathbf{k}, \mathbf{k}') = \begin{cases} \lambda_1, & \text{если } k_1 \leq N-1, k_2 = 0 \text{ или } k_1 \leq N-2, \mathbf{k}' = \mathbf{k} + \mathbf{e}_1, \\ \lambda_2, & \text{если } k_1 \leq N-1, \mathbf{k}' = \mathbf{k} + \mathbf{e}_2, \\ k_1 \mu_1, & \text{если } \mathbf{k}' = \mathbf{k} - \mathbf{e}_1, \\ (N - k_1) \mu_2, & \text{если } \mathbf{k}' = \mathbf{k} - \mathbf{e}_2, \\ 0 & \text{в остальных случаях.} \end{cases} \quad (2.26)$$

Граф переходов между состояниями модели показан на рис.2.4. Все состояния рассматриваемой конечномерной ЦМ являются сообщающимися (см. формулы (2.26) и рис. 2.4), и, следовательно, существует ее стационарное распределение. Стационарные вероятности состояний удовлетворяют следующей СУР, которая составляется на основе (2.26):

для состояния $\mathbf{k} = (N, 0)$:

$$N \mu_1 p(\mathbf{k}) = \lambda_1 p(\mathbf{k} - \mathbf{e}_1); \quad (2.27)$$

для состояния $\mathbf{k} \neq (N, 0)$:

$$\begin{aligned} (\lambda_1 + \lambda_2 + k_1 \mu_1 + (N - k_1) \mu_2) p(\mathbf{k}) &= \lambda_1 p(\mathbf{k} - \mathbf{e}_1) + \lambda_2 p(\mathbf{k} - \mathbf{e}_2) + \\ &+ (k_1 + 1) \mu_1 p(\mathbf{k} + \mathbf{e}_1) + (N - k_1) \mu_2 p(\mathbf{k} + \mathbf{e}_2). \end{aligned} \quad (2.28)$$

К системе уравнений (2.27), (2.28) добавляется также нормирующее условие:

$$\sum_{\mathbf{k} \in S} p(\mathbf{k}) = 1. \quad (2.29)$$

Нахождение стационарного распределения $p(\mathbf{k})$, $\mathbf{k} \in S$ из СУР (2.27)–(2.29) является достаточным для вычисления желаемых характеристик данной модели. Так, вероятность потери n -э-заявок ($P_{B_{n,э}}$) и э-заявок ($P_{B_э}$) определяется из следующих соотношений:

$$P_{B_{n,э}} = p(N, 0) + \sum_{i=1}^N p(N-1, i), \quad (2.30)$$

$$P_{B_э} = p(N, 0) + \sum_{i=0}^{N-1} p(i, N). \quad (2.31)$$

Среднее число занятых каналов (\tilde{N}) в данной модели определяется так:

$$\tilde{N} = \sum_{i=1}^N i p(i, 0) + N \left(1 - \sum_{i=0}^N p(i, 0) \right). \quad (2.32)$$

Кроме характеристик (2.30)–(2.32) для данной модели не менее важными являются вероятность того, что система находится в состоянии деградации, а также вероятность деградации э-заявки. С целью определения указанных характеристик введем следующие понятия.

Поскольку полосы (каналы) системы являются непрерывными и э-заявки равноправно используют эти полосы, то если система находится в состоянии $\mathbf{k} = (i, j)$, то доля полосы (интенсивности) на каждой э-заявке, находящейся на обслуживании, равна $(N - i) / j ((N - i) \mu_2 / j)$, $j \neq 0$. Однако, в реальных мультимедийных сетях качество обслуживания (передачи) заявки определяется именно этим параметром, т.е. в них имеются ограничения на значения данного параметра. Иными словами, если скорость (интенсивность) обслуживания э-заявки меньше некоторого порогового значения $\delta > 0$, то тогда качество передачи считается неудовлетворительным, и, следовательно, говорят, что система находится в состоянии деградации.

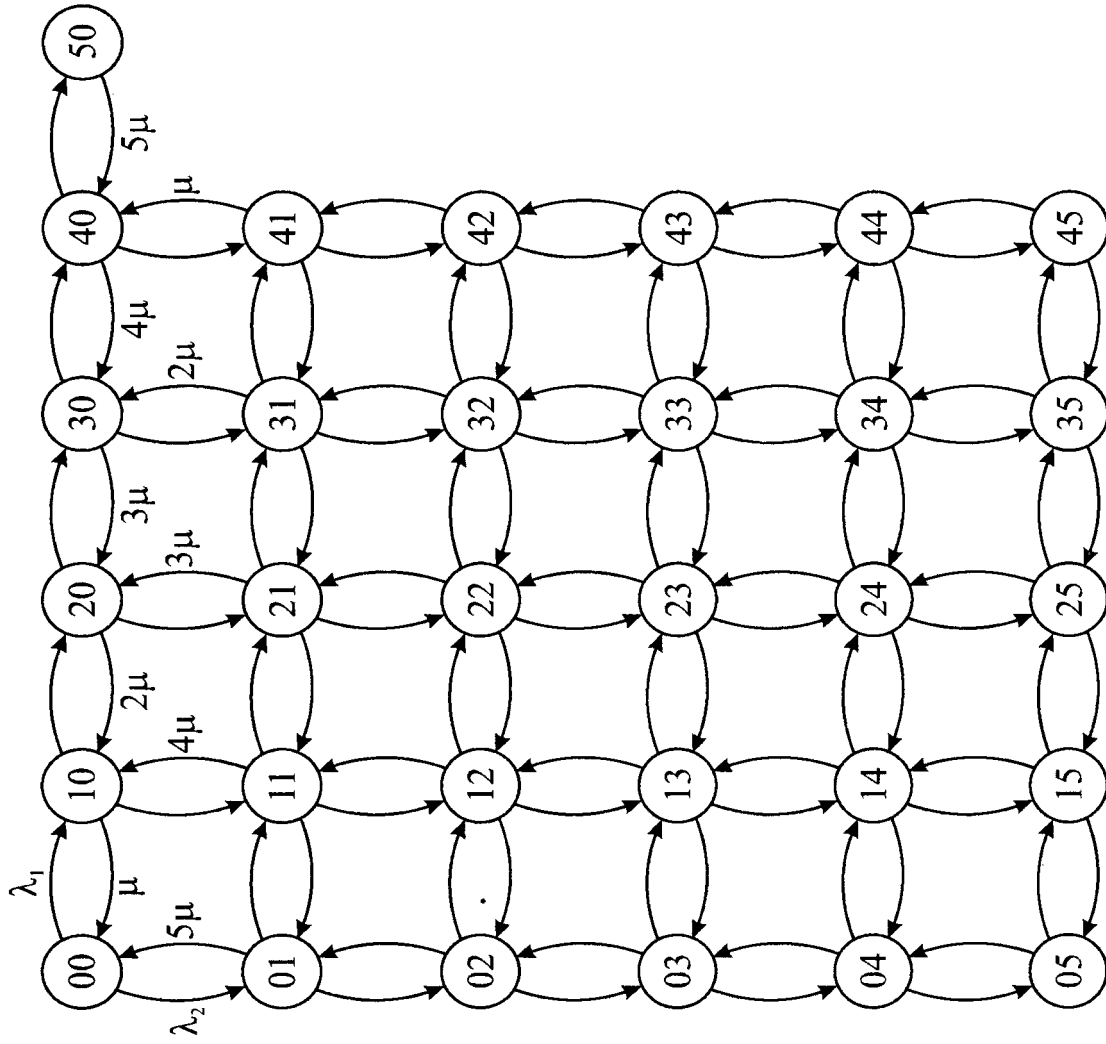


Рис. 2.4. Граф переходов между состояниями модели с непрерывной полосой для э-заявок при $N = 5$

Для оценки данной характеристики вводится вероятность деградационных состояний ($PDS - Probability of Degradation States$). С целью вычисления указанной величины в ФПС (2.25) определяется подмножество деградационных состояний (S_{deg}):

$$S_{deg} := \left\{ \mathbf{k} \in S : \frac{N - k_1}{k_2} < \delta \right\}. \quad (2.33)$$

Тогда, исходя из того, что данная система является марковской, для вычисления искомой вероятности PDS найдем следующие формулы:

$$PDS = \sum_{\mathbf{k} \in S_{deg}} p(\mathbf{k}). \quad (2.34)$$

С помощью (2.33) можно также определить альтернативную к (2.34) вероятность, т.е. вероятность того, что система не находится в состоянии деградации (иными словами, э-заявки обслуживаются с необходимым качеством). Обозначим эту вероятность через PHQ (*Probability of High Quality of Service*). Тогда имеем:

$$PHS = \sum_{\mathbf{k} \in S \setminus S_{deg}} p(\mathbf{k}). \quad (2.35)$$

либо

$$PHQ = 1 - PDS.$$

Другая характеристика – вероятность деградации э-заявок ($PDE - Probability of Elastic Call Degradation$) определяется так:

$$PDE = M_e^{deg} / M_e, \quad (2.36)$$

где M_e – среднее число э-заявок в системе, M_e^{deg} – среднее число э-заявок, обслуживаемых с деградационной интенсивностью.

Участвующие в (2.36) величины $M(k_2^{deg})$ и $M(k_2)$ определяются стандартным образом (т.е. как среднее значение некоторой дискретной случайной величины):

$$M_e = \sum_{i=1}^N \sum_{j=0}^{N-1} ip(j, i), \quad (2.37)$$

$$M_e^{deg} = \sum_{i=1}^N \sum_{(i, j) \in S_{deg}} ip(j, i). \quad (2.38)$$

Таким образом, для вычисления характеристик (2.30)–(2.32) и (2.34)–(2.36) потребуется предварительное нахождение стационарного распределения данной модели. Решение последней задачи не представляет никаких трудностей при умеренных размерах ФПС (2.25). Однако, как указывалось в § 2.1, решение данной задачи становится с огромными вычислительными трудностями при больших размерах ФПС (2.25), так как СУР (2.27)–(2.29) не имеет аналитического решения.

Для преодоления указанных выше трудностей можно использовать приближенный подход. С целью корректного применения этого подхода здесь также предполагается, что $\lambda_1 \gg \lambda_2$. Поскольку этот

подход был неоднократно использован в данной работе, то ниже его изложение применительно к рассмотренной модели максимально сокращается.

С учетом принятого выше допущения относительно соотношения нагрузок трафиков n -э- и $э$ -заявок фиксируется значение второй компоненты вектора соотношений и рассматривается соответствующее разбиение ФПС (2.25) по строкам.

При $k_2 = 0$ стационарное распределение внутри соответствующего класса совпадает с известным распределением классической системы Эрланга $M/M/0$ с нагрузкой ν_1 эрл., т.е. стационарное распределение внутри указанного класса ($\theta(i): i = 0, \dots, N$) определяется с помощью (2.10). А при $k_2 = j, j = 1, \dots, N$, стационарные распределения внутри всех классов являются одинаковыми, и совпадают с распределением стационарной системы $M/M/N-1|0$ с нагрузкой ν_1 эрл. Таким образом, стационарные распределения внутри всех указанных классов ($\theta(r): r = 0, \dots, N-1, j = 1, \dots, N$, определяются аналогично (2.11).

Следовательно, интенсивности переходов между укрупненными состояниями (их общее число равно $N+1 - q(<r>, <r'>)$, $r, r' = 0, \dots, N$, определяются из следующих соотношений:

$$q(<r>, <r'>) = \begin{cases} \lambda_2(1 - E_B(\nu_1, N)), & \text{если } r = 0, r' = r + 1, \\ \lambda_2, & \text{если } r > 0, r' = r + 1, \\ \mu_2 \sum_{i=0}^{N-1} (N-i)\theta_1(i), & \text{если } r' = r - 1, \end{cases} \quad (2.39)$$

{ 0 в остальных случаях.

Замечание 2.4. Поскольку $\theta(i) = \theta_1(i)$ для любого $j, j' = 0, \dots, N$, то в формуле (2.39) для определенности использована запись $\theta_1(i)$.

Тогда, стационарные вероятности состояния укрупненной модели, представляющей собой классический процесс размножения и гибели с переходными интенсивностями, которые определяются из соотношений (2.39), вычисляются очевидным образом:

$$\pi(<r>) = \left(\prod_{i=0}^{r-1} \frac{q(<i>, <i+1>)}{q(<i+1>, <i>)} \right) \pi(<0>), r = 1, \dots, N. \quad (2.40)$$

где

$$\pi(<0>) = \left(1 + \sum_{j=1}^N \prod_{i=0}^{j-1} \frac{q(<i>, <i+1>)}{q(<i+1>, <i>)} \right)^{-1} \quad (2.41)$$

Далее с использованием (2.10), (2.11) и (2.40), (2.41) определяется стационарное распределение исходной модели с ФСП (2.25).

Характеристики (2.30)–(2.32) данной модели приближенно определяются следующим образом:

$$PB_{н.э.} \approx E_B(\nu_1, N)\pi(<0>) + (N-1)E_B(\nu_1, N-1), \quad (2.42)$$

$$PB_3 \approx E_B(\nu_1, N)\pi(<0>) + \pi(<N>), \quad (2.43)$$

$$\tilde{N} \approx \sum_{i=1}^N i\theta(i)\pi(<0>) + N(1 - \pi(<0>)). \quad (2.44)$$

После достаточного сложных математических преобразований получим, что характеристики (2.34)–(2.36) могут быть приближенно вычислены из следующих формул:

$$PDS \approx 1 - PHQ, \quad (2.45)$$

$$PHQ \approx \sum_{i=0}^{N-1} \sum_{j=i}^{N-1} \pi(<j>)\theta_j(i); \quad (2.46)$$

$$PDE \approx M_e^{\text{deg}} / M_e, \quad (2.47)$$

где

$$M_e \approx \sum_{i=1}^N \sum_{j=0}^{N-1} i\pi(<i>)\theta_i(j), \quad (2.48)$$

$$M_e^{\text{deg}} \approx \sum_{i=1}^N \sum_{j=\lfloor \frac{N-i}{\delta} \rfloor + 1}^{N-1} i\pi(<i>)\theta_i(j). \quad (2.49)$$

Таким образом, для нахождения желаемых характеристик модели с непрерывной полосой для эластичных заявок предложены простые вычислительные процедуры, которые включают табулированные величины.

2.2.2. Модели с дискретной полосой для эластичных заявок

Теперь рассмотрим модель, в которой $э$ -заявки могут выбирать подосы между m и \underline{m} дискретным образом, при этом считается, что $\underline{m} = 1, m = N$.

Как и выше (см. §2.2.1), $э$ -заявки равноправным образом занимают все те каналы системы, которые не заняты обслуживанием n -э-заявок, т.е. если $э$ -заявка поступает в момент, когда все каналы свободны, то она занимает все эти каналы, и обслуживание данной $э$ -заявки осуществляется с максимальной скоростью; если $э$ -заявка

занимает более одного канала, то при поступлении новой э-заявки часть каналов, занятых обслуживанием э-заявки, переключается на обслуживание поступившей э-заявки, при этом может быть использована любая схема распределения каналов между обслуживаемыми э-заявками. В отличие от модели с непрерывной полосой, в данном случае не допускается выделение пропускной способности каналов для поступившей э-заявки, если в этот момент все каналы системы заняты, т.е. минимальное число полос, необходимое для обслуживания одной э-заявки, равно 1. Последнее означает, что здесь не разрешается дробление ресурса каждого канала, и, следовательно, максимальное число заявок каждого типа в системе равно N .

Теперь рассмотрим схему организации обслуживания n -э-заявок в данной модели. Если в момент поступления n -э-заявки в системе отсутствуют э-заявки и имеется хотя бы один свободный канал, то данная n -э-заявка принимается на обслуживание. При поступлении для обслуживания один канал, занятый обслуживанием э-заявок, выделяется для обслуживания поступившей n -э-заявки, если после принятия данной n -э-заявки общее число разнотипных заявок в системе не будет превышать N . Иными словами, в данной модели поступившая заявка любого типа получает отказ, если в момент ее поступления общее число разнотипных заявок в системе равно N .

Как и в случае модели с непрерывной полосой, здесь также после завершения обслуживания заявки любого типа, освобожденные полосы распределяются между обслуживаемыми э-заявками, и, таким образом, увеличивается суммарная скорость их обслуживания. При этом может быть использована любая схема распределения каналов между обслуживаемыми в данный момент заявками.

Фазовое пространство состояний данной модели также содержит двумерный вектор $\mathbf{k} = (k_1, k_2)$, где $k_i(k_2)$ означает число n -э-заявок (э-заявок) в системе, при этом $k_i = 0, \dots, N, i = 1, 2$. Однако, ФПС данной модели отличается от (2.25) и задается так:

$$S := \{\mathbf{k}; k_i = 0, \dots, N, i = 1, 2; k_1 + k_2 \leq N\}. \quad (2.50)$$

Граф переходов между состояниями ФПС (2.50) показан на рис.2.5.

С учетом механизма работы данной модели определим, что элементы ПМ соответствующей цепи Маркова определяются из следующих соотношений:

$$q(\mathbf{k}, \mathbf{k}') = \begin{cases} \lambda_i, & \text{если } \mathbf{k}' = \mathbf{k} + \mathbf{e}_i, i = 1, 2, \\ k_i \mu_i, & \text{если } \mathbf{k}' = \mathbf{k} - \mathbf{e}_i, \\ (N - k_1) \mu_2, & \text{если } \mathbf{k}' = \mathbf{k} - \mathbf{e}_2, \\ 0 & \text{в остальных случаях.} \end{cases} \quad (2.51)$$

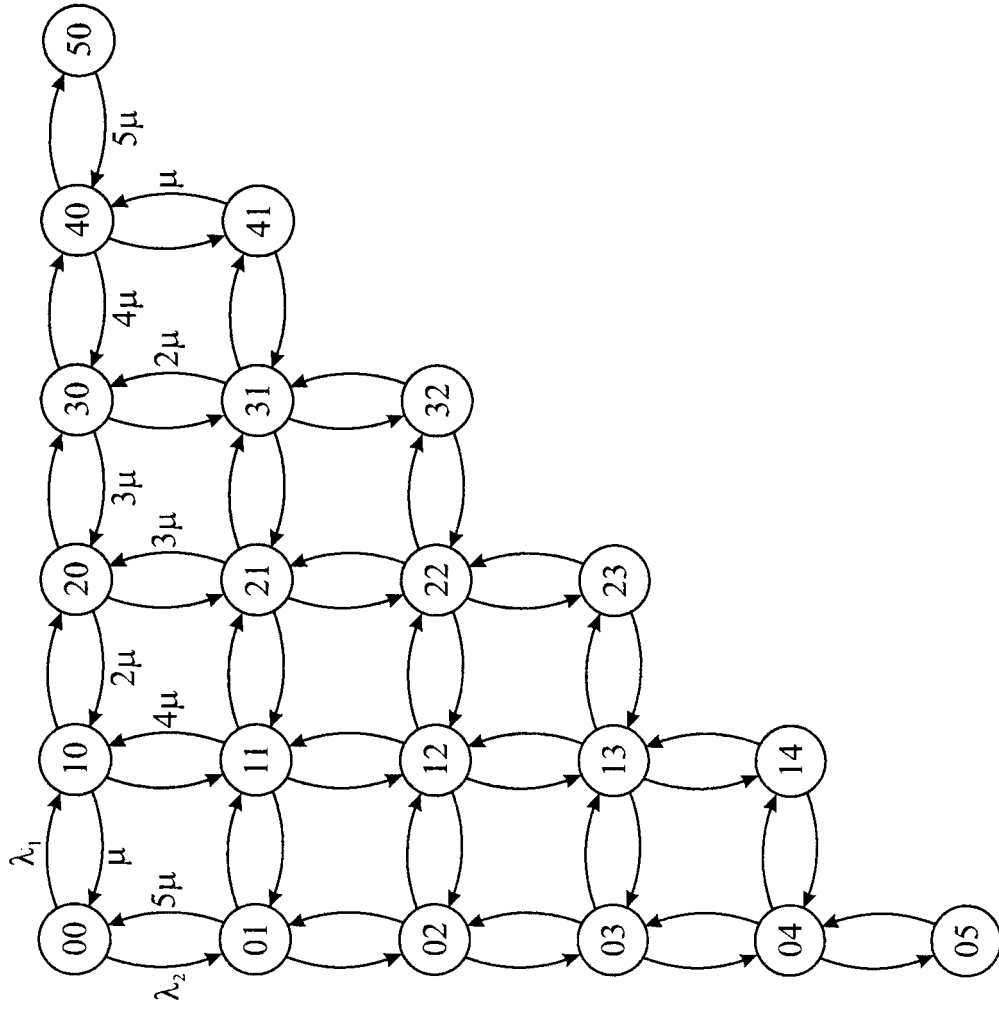


Рис.2.5. Граф переходов между состояниями модели с дискретной полосой для э-заявок при $N = 5$

На основе (2.51) составляется СУР для стационарных вероятностей состояний из ФПС (2.50). Она аналогична (2.27)–(2.29) и ввиду очевидности ее составления явный вид этой СУР здесь не приводится.

Характеристики данной модели определяются через ее стационарное распределение следующим образом.

Прежде всего, отметим, что в отличие от модели с непрерывной полосой для э-заявок, в данной модели вероятности потери разнотипных заявок являются одинаковыми:

$$PB_{n, \text{э}} = PB_{\text{э}} = \sum_{\mathbf{k} \in S_d} p(\mathbf{k}), \quad (2.52)$$

где $S_d := \{\mathbf{k} \in S; k_1 + k_2 = N\}$ – подмножество диагональных состояний.

Среднее число занятых каналов (\bar{N}) в данной модели также определяется согласно (2.32), но при этом следует иметь в виду, что в правой части указанной формулы участвует стационарное распределение рассматриваемой модели.

Для данной модели также можно ввести понятие «множество деградационных состояний» (S_{deg}). Однако, способ определения этого множества отличается от соответствующего способа для модели с непрерывной полосой. Поскольку в данной модели не разрешается дробление ресурса каждого канала, то считается, что система находится в состоянии деградации, если общее число каналов, занятых обслуживанием э-заявок, меньше определенного порогового значения $\delta > 0$. Иными словами, состояние \mathbf{k} называется деградационным, если для этого состояния $k_1 \geq N - \delta$ и $k_2 > 0$, где $1 \leq \delta \leq N - 1$.

Таким образом, вероятность деградационных состояний (PDS) для данной модели определяется так:

$$PDS = \sum_{\mathbf{k} \in S_{deg}} p(\mathbf{k}), \quad (2.53)$$

где $S_{deg} = \{\mathbf{k} \in S: k_1 \geq N - \delta, k_2 > 0\}$, $1 \leq \delta \leq N - 1$.

Вероятность деградации э-заявок (PDE) для данной модели определяется согласно формулам (2.36)–(2.38), где в формулах (2.37)–(2.38) следует иметь в виду стационарное распределение рассматриваемой модели.

Точные значения характеристик модели вычисляются через стационарное распределение модели, которое определяется из СУР, составленной на основе (2.51). Как было отмечено выше, при больших размерах ФПС модели определение точных значений стационарных вероятностей состояний представляет собой сложную вычислительную проблему. Поэтому зачастую приходится ограничиваться их приближенными значениями.

Приведем окончательный вид алгоритма расчета приближенных значений характеристик модели с дискретной полосой для э-заявок.

Шаг 1. Вычислить следующие величины $q(r, r')$; $r, r' = 0, \dots, N$:

$$q(r, r') = \begin{cases} \lambda_2(1 - E_B(v_1, N - r)), & \text{если } r' = r + 1, \\ \mu \sum_{j=0}^{N-r} (N - j)\theta_r(j), & \text{если } r' = r - 1, \\ 0 & \text{в остальных случаях,} \end{cases}$$

где $\theta(j)$, $j = 0, \dots, N$, представляет собой стационарное распределение классической системы Эрланга $M/M/N - i|0$ с нагрузкой v_1 эрл., $j = 0, \dots, N - i$:

$$\theta_j(j) = \frac{v_1^j}{j!} \left(\sum_{r=0}^{N-i} \frac{v_1^r}{r!} \right)^{-1}, \quad j = 0, \dots, N - i.$$

Шаг 2. Вычислить следующие величины $\pi(r)$, $r = 0, \dots, N$:

$$\pi(r) = \left(\prod_{i=0}^{r-1} \frac{q(i, i+1)}{q(i+1, i)} \right) \pi(0),$$

где

$$\pi(0) = \left(1 + \sum_{i=1}^N \prod_{j=0}^{i-1} \frac{q(i, i+1)}{q(i+1, i)} \right)^{-1}.$$

Шаг 3. Вычислить приближенные значения характеристик модели:

$$PB_{наз.} = PB_3 \approx \sum_{i=0}^N E_B(v_1, N - i)\pi(i);$$

$$\bar{N} \approx \pi(0) \sum_{i=1}^N i\theta(i) + N(1 - \pi(0));$$

$$PDS \approx \sum_{i=N-\delta}^{N-1} \sum_{j=1}^{N-i} \pi(j)\theta_j(i);$$

$$PHS \approx 1 - PDS;$$

$$PDE \approx M_e^{deg} / M_e,$$

$$M_e = \sum_{i=1}^N \sum_{j=0}^{N-i} i\pi(i)\theta_j(j),$$

$$M_e^{deg} = \sum_{i=1}^N \sum_{j=N-\delta}^{N-i} i\pi(i)\theta_j(j).$$

где

Таким образом, разработан эффективный алгоритм расчета приближенных значений желаемых характеристик модели MRQ с дискретной полосой для обслуживания эластичных заявок. Данный алгоритм, как и соответствующий алгоритм расчета характеристик модели MRQ с непрерывной полосой для обслуживания эластичных заявок, включает в себя табулированные величины распределения Эрланга.

2.3. Численные эксперименты

На основе предложенных выше алгоритмов расчета моделей MRQ с неэластичными и эластичными заявками разработаны соответствующие вычислительные программы. Здесь приводятся результаты численных экспериментов, выполненных с помощью указанных вычислительных программ.

Сначала рассмотрим результаты для модели с одной ε -заявкой на обслуживании. Поскольку модели с равноправным доступом являются частными случаями моделей с ограниченным доступом для n - ε -заявок, то здесь для краткости изложения показаны результаты численных экспериментов лишь для моделей последнего типа.

Относительно поведения функций вероятностей потерь разнотипных заявок и среднего числа занятых каналов при изменении общего числа каналов лишь отметим, что, как и следовало ожидать, вероятности потерь являются монотонно убывающими, а среднее число занятых каналов монотонно возрастающей функцией.

Поэтому, на рис. 2.6–2.8 показаны лишь графики зависимостей вероятностей потерь и среднее число занятых каналов от параметра R , соответственно. При этом общее число каналов и нагрузки трафиков считаются фиксированными.

Сделаем некоторые замечания относительно теоретических ожиданий относительно характера изменения изучаемых величин. Прежде всего отметим, что с ростом параметра R уменьшаются шансы доступа ε -заявок к каналам обслуживания, т.е. следует ожидать, что функция вероятности потери ε -заявок должна быть монотонно возрастающей относительно аргумента R . С другой стороны, с ростом параметра R уменьшаются также шансы n - ε -заявок быть принятыми на обслуживание, так как с ростом параметра R n - ε -заявки все реже могут использовать каналы, занятые обслуживанием ε -заявок. Иными словами, также следует ожидать, что функция вероятности потери n - ε -заявок является монотонно возрастающей относительно аргумента R .

Среднее число занятых каналов естественным образом зависит от нагрузочных параметров входящих трафиков. Вместе с тем, следует отметить, что с ростом аргумента R уменьшается число состояний типа $(j, 1)$ (так как их максимальное число равно $N - R$), и тем самым увеличивается вероятность того, что система находится в таких состояниях, и следовательно, с ростом аргумента R также увеличивается среднее число занятых каналов.

Анализ многочисленных экспериментов полностью подтвердили все теоретические ожидания. Они позволяют также сделать следующие выводы:

– вероятность потери неэластичных заявок является возрастающей функцией нагрузки эластичных заявок при фиксированных значениях всех остальных параметров модели (см. рис. 2.6). При этом скорость изменения этой функции естественным образом зависит от значения па-

раметра R . Так, например, при двух различных значениях нагрузки эластичных заявок, равных 0.1 и 0.01, значения вероятностей потерь являются почти одинаковыми при любых значениях R ($R \leq 8$); а именно, при $R = 8$ соответствующие значения указанной функции равны $6.3 \cdot 10^{-12}$ и $1.1 \cdot 10^{-12}$; при $9 \leq R \leq 39$ отношение соответствующих значений этой функции приблизительно равно 10 (например, при $R = 9$ указанные значения равны $1.8 \cdot 10^{-11}$ и $2.3 \cdot 10^{-12}$, а при $R = 39$ соответствующие значения равны $3.1 \cdot 10^{-1}$ и $4.4 \cdot 10^{-2}$). При максимально возможном значении $R = 40$ значения вероятностей потерь при указанных нагрузках эластичных заявок равны $8.4 \cdot 10^{-1}$ и $3.5 \cdot 10^{-1}$, т.е. их отношение приблизительно равно 2.4;

– характер поведения вероятности потери эластичных заявок относительно изменения нагрузки заявок данного типа схож с характером поведения аналогичной функции для неэластичных заявок (см. рис.2.7);

– среднее число занятых каналов несущественно зависит от значения нагрузки эластичных заявок (см. рис.2.8). Так, например, при десятикратном увеличении нагрузки эластичных заявок значение данной функции почти не изменяется, и при выбранных значениях параметров модели, указанных в подрисуночных подписях рис. 2.6, оно приблизительно равно 10 при $1 \leq R \leq 38$. Лишь при экстремальных значениях параметра R , $R = 39$ и $R = 40$, значения этой функции равны 20.16 и 35.39 при $\lambda_2 = 0.1$ и 11.46 и 20.65 при $\lambda_2 = 0.001$, соответственно.

Результаты численных экспериментов для моделей с непрерывной полосой показаны на рис.2.9 – 2.14. Анализ результатов многочисленных экспериментов позволяет сделать следующие выводы для моделей с непрерывной полосой передачи ε -заявок:

– функция вероятности потери заявок обоих типов является монотонно убывающей относительно числа каналов системы (см. рис.2.9 и 2.10), а среднее число занятых каналов, наоборот, монотонно возрастает с увеличением числа каналов (см. рис.2.11) при любых нагрузках системы. Эти результаты вполне соответствуют теоретическим ожиданиям;

– функция вероятности деградационных состояний является монотонно возрастающей относительно порогового параметра δ (см. рис. 2.12), а дополняющая функция, т.е. функция, оценивающая вероятность того, что ε -заявки обслуживаются с необходимой скоростью обслуживания, наоборот, монотонно убывает относительно указанного аргумента δ (см. рис.2.13). Оба этих результата имеют вполне логичный смысл;

– функция вероятности деградации ε -заявки при любых нагрузках и числе каналов, как и ожидалось, монотонно возрастает относительно параметра δ (см. рис.2.14).

Точно такими же видами зависимостей описываются аналогичные характеристики модели с дискретной полосой передачи э-заявок (см. рис.2.15–2.19).

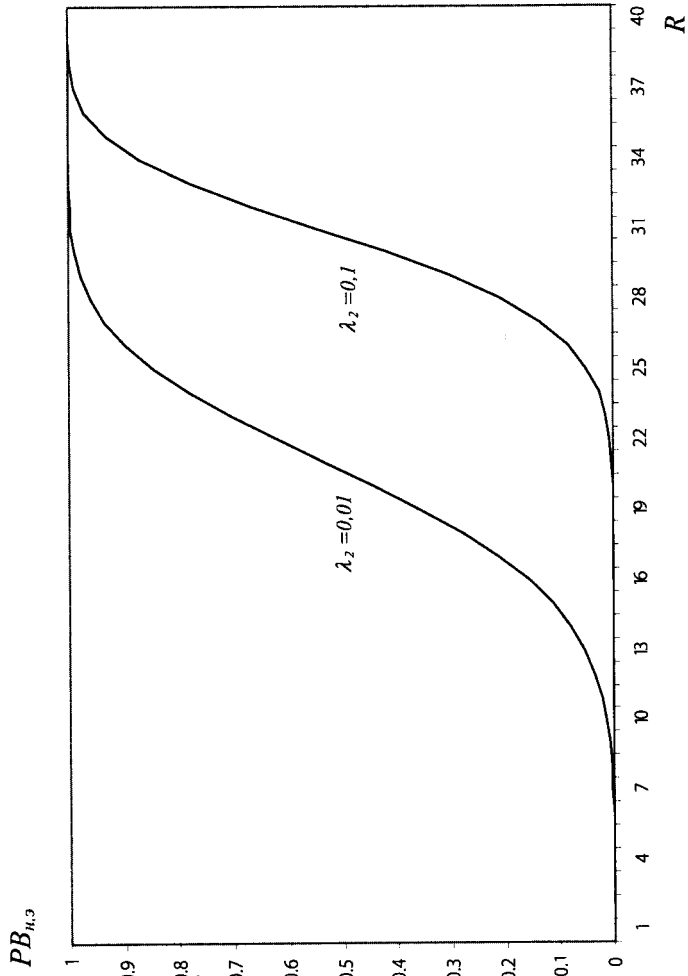


Рис. 2.6. Зависимость вероятности потери н.э-заявок от R при $N = 40$; $\nu_1 = 10$; $\mu_1 = 10$

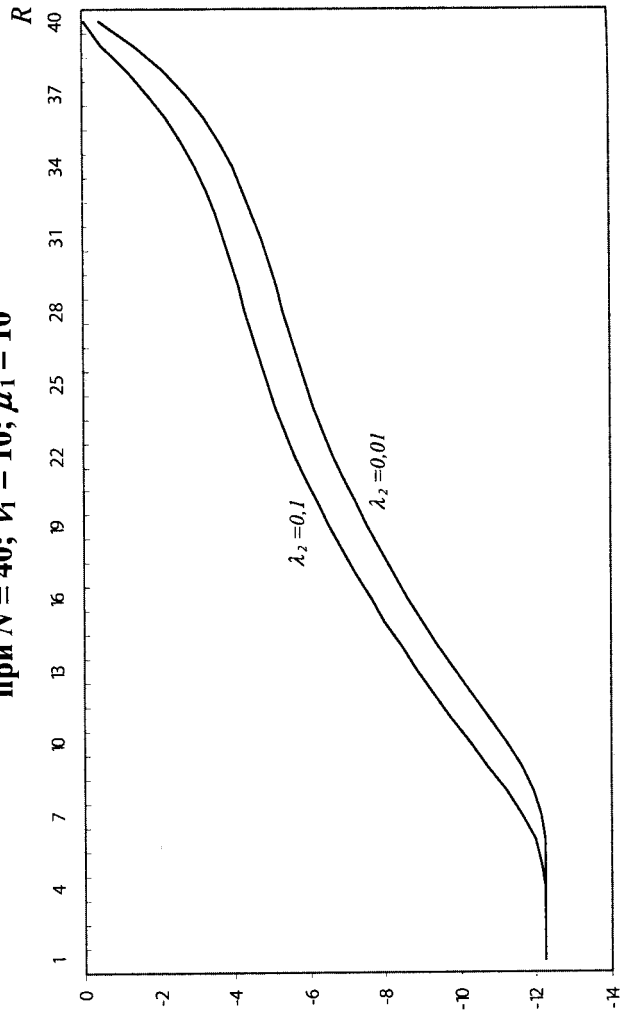


Рис. 2.7. Зависимость вероятности потери э-заявок от R при $N = 40$; $\nu_1 = 10$; $\mu_1 = 10$

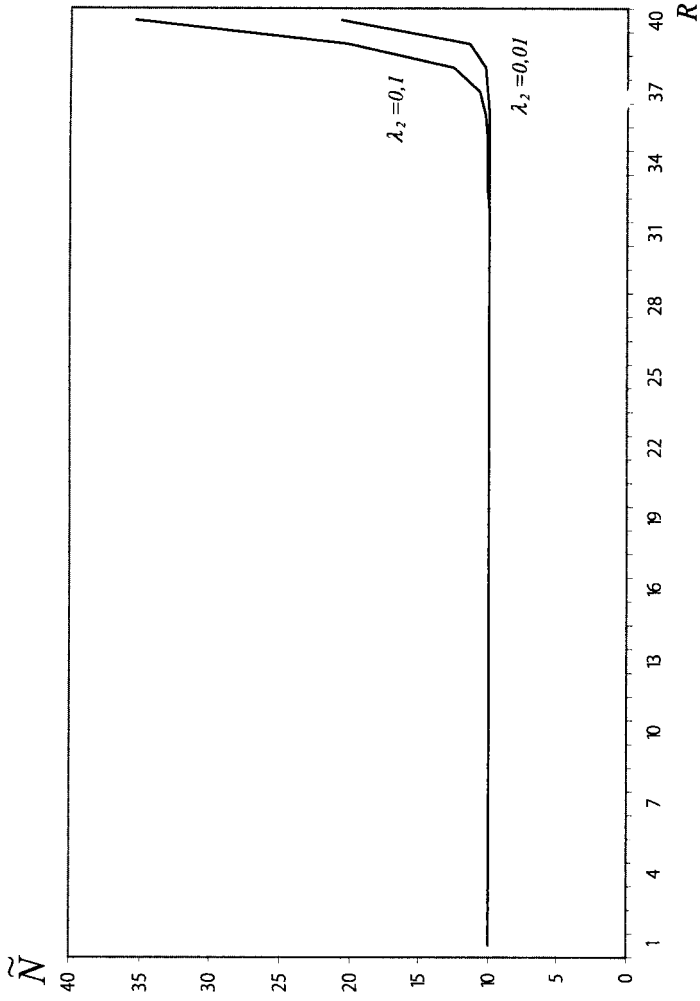


Рис. 2.8. Зависимость среднего числа занятых каналов от R при $N = 40$; $\nu_1 = 10$; $\mu_1 = 10$

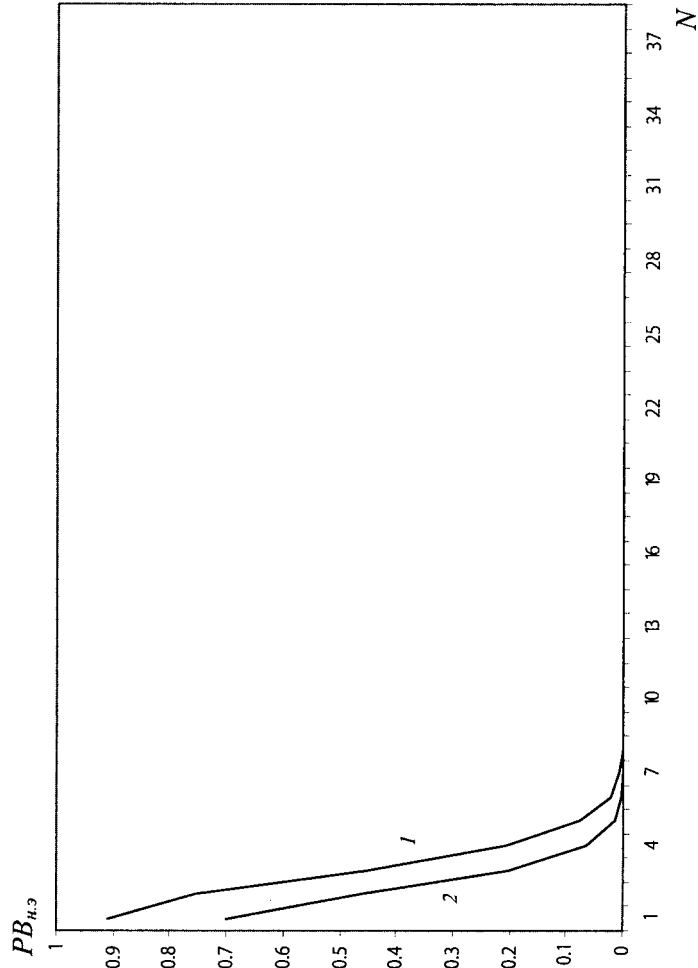


Рис. 2.9. Зависимость вероятности потерь н.э-заявок от N для моделей с непрерывной полосой передачи э-заявок при $\delta = 25$; $\mu = 10$; 1 - $\lambda_1 = 15$; $\lambda_2 = 0,01$; 2 - $\lambda_1 = 10$; $\lambda_2 = 0,1$

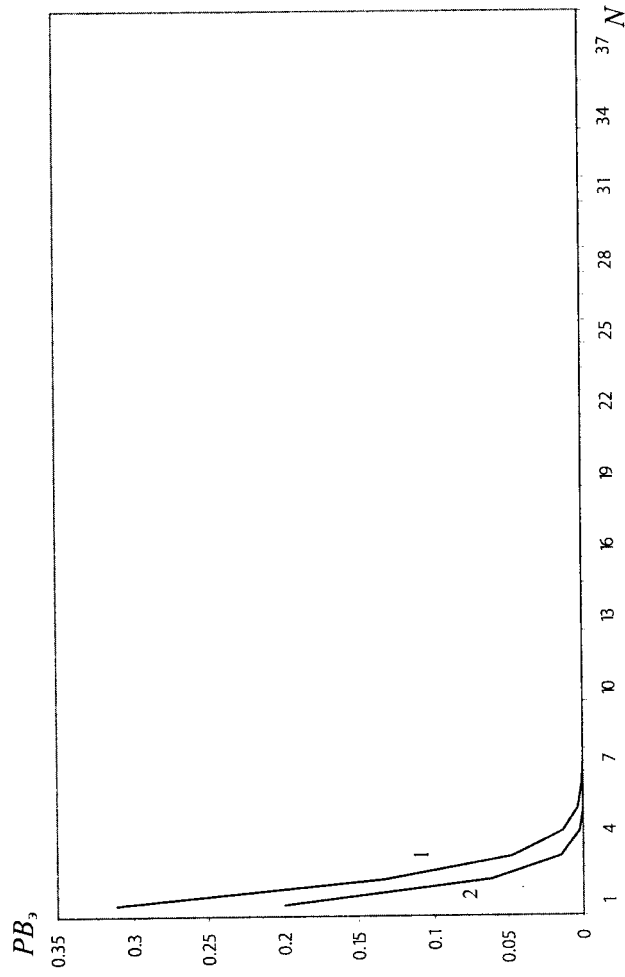


Рис. 2.10. Зависимость вероятности потерь э-заявок от N для моделей с непрерывной полосой передачи э-заявок

при $\delta=25, \mu=10$;

1 - $\lambda_1=15; \lambda_2=0,01$; 2 - $\lambda_1=10; \lambda_2=0,1$;

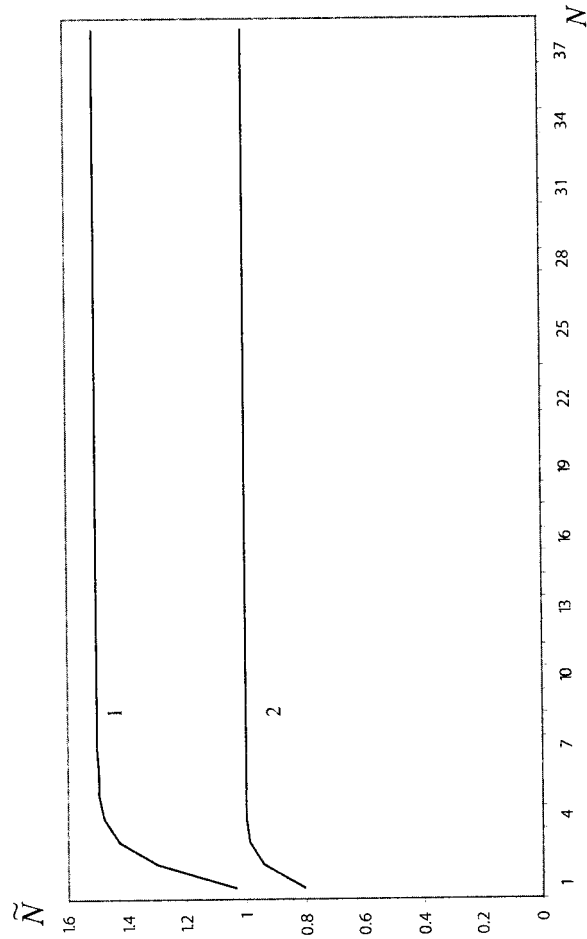
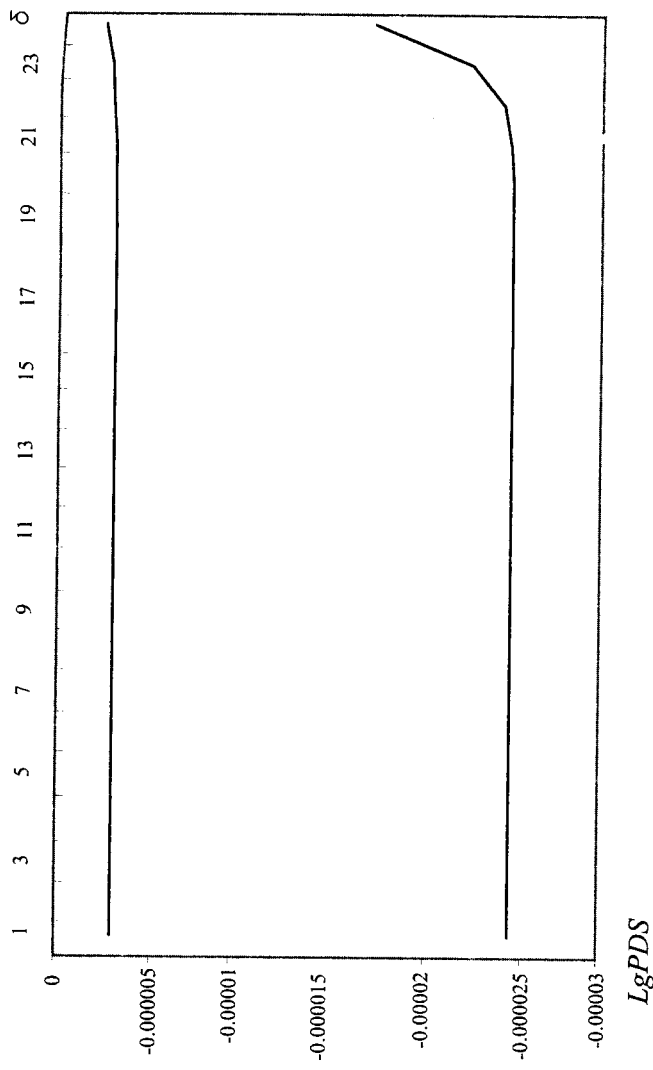


Рис. 2.11. Зависимость среднего числа занятых каналов от N для моделей с непрерывной полосой передачи э-заявок при $\delta=25; \mu=10$;

1 - $\lambda_1=15; \lambda_2=0,01$; 2 - $\lambda_1=10; \lambda_2=0,1$

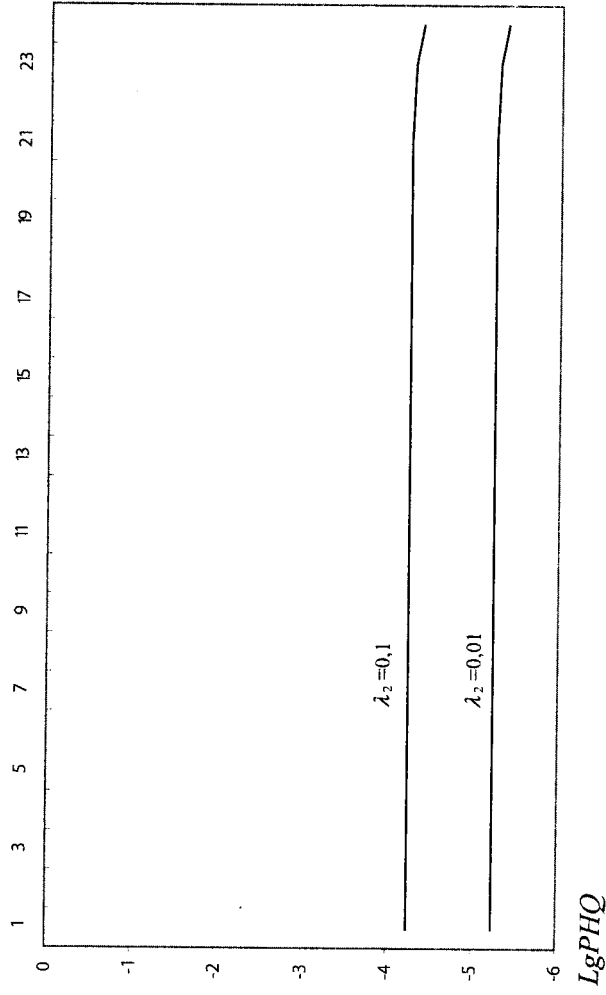


LgPDS

Рис. 2.12. Зависимость вероятности деградационных состояний от δ для моделей с непрерывной полосой передачи э-заявок

при $N=25; \mu=10$;

1 - $\lambda_1=10; \lambda_2=0,1$; 2 - $\lambda_1=10; \lambda_2=0,01$



LgPHQ

Рис. 2.13. Зависимость вероятности того, что э-заявки обслуживаются с необходимой скоростью обслуживания, от δ для моделей с непрерывной полосой передачи э-заявок при $N=25; \lambda_1=10; \mu=10$

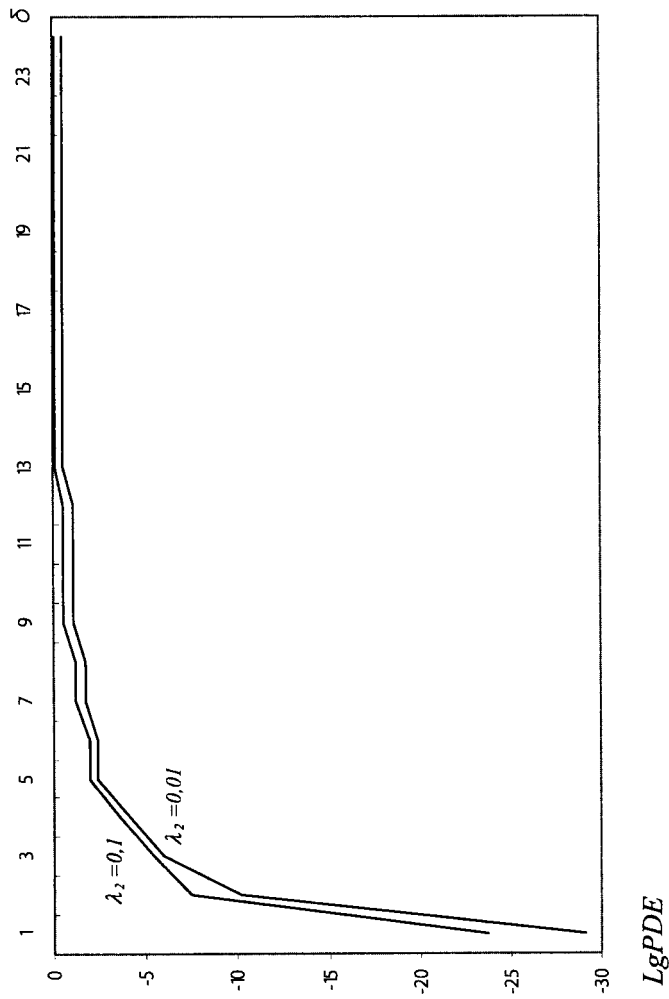


Рис. 2.14. Зависимость вероятности деградации э-заявок от δ для моделей с непрерывной полосой передачи э-заявок при $N = 25; \lambda_1 = 10; \mu = 10$

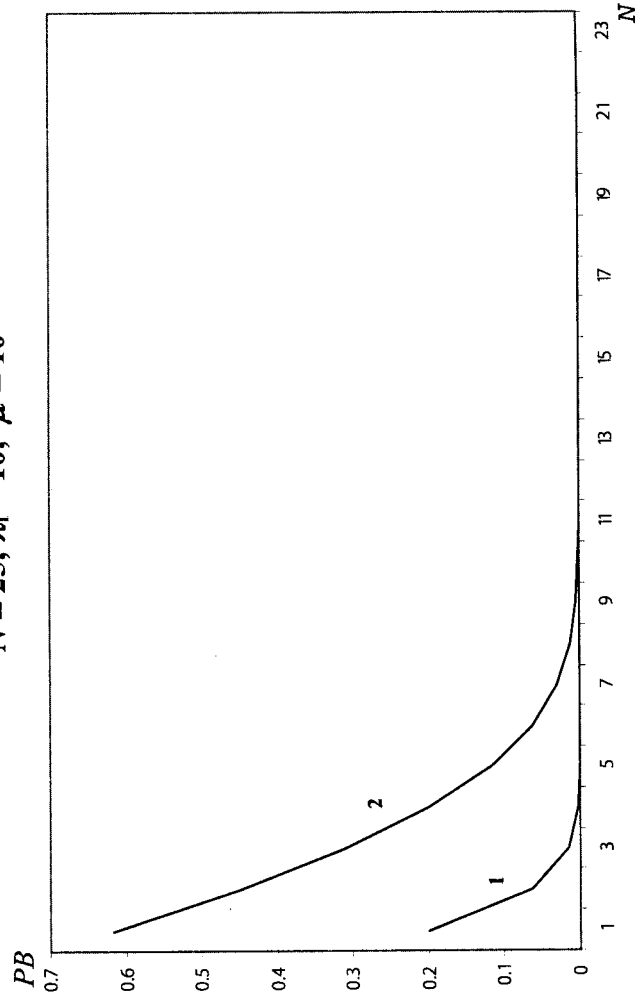


Рис. 2.15. Зависимость вероятности потерь заявок от N для моделей с дискретной полосой передачи э-заявок при $\delta = 25; \lambda_2 = 0,01$; $1 - \lambda_1 = 10; \mu = 10; 2 - \lambda_1 = 20; \mu = 5$

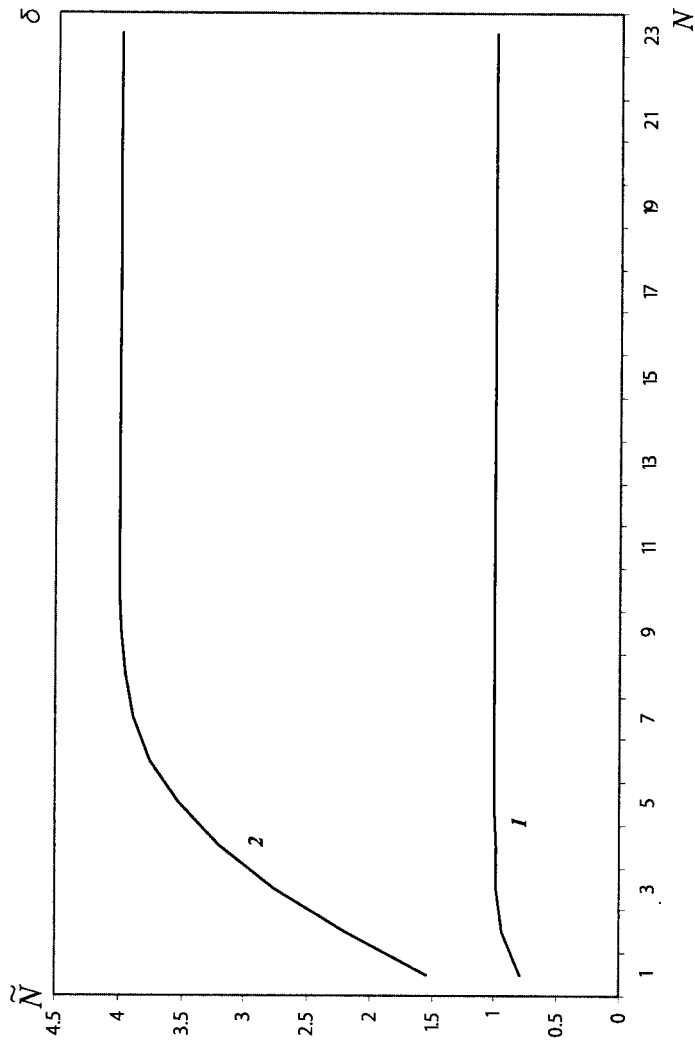


Рис. 2.16. Зависимость среднего числа занятых каналов от N для моделей с дискретной полосой передачи э-заявок при $\delta = 25; \lambda_2 = 0,01$;

$1 - \lambda_1 = 10; \mu = 10; 2 - \lambda_1 = 20; \mu = 5$

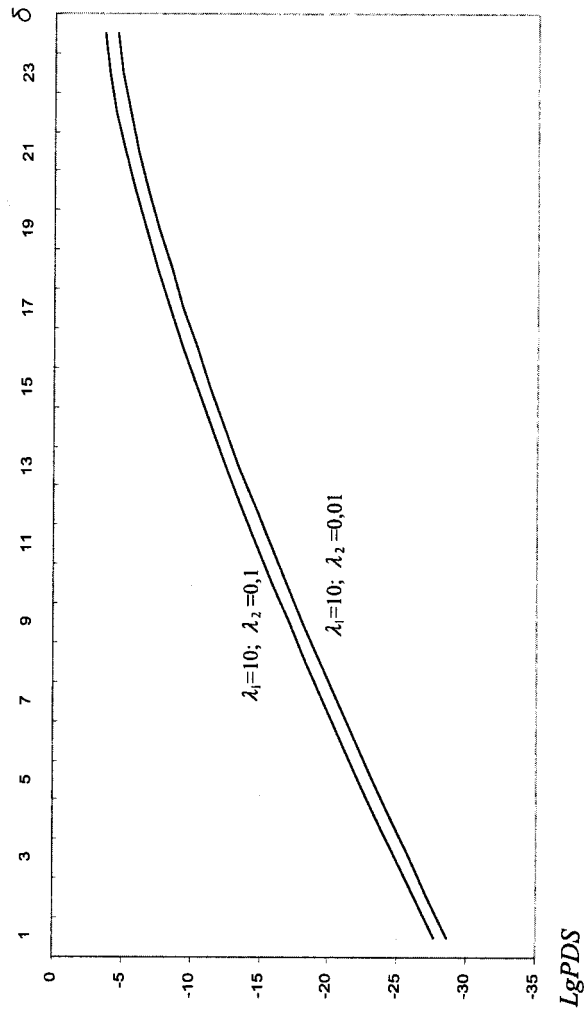


Рис. 2.17. Зависимость вероятности деградационных состояний от δ для моделей с дискретной полосой передачи э-заявок при $N = 25; \mu = 10$

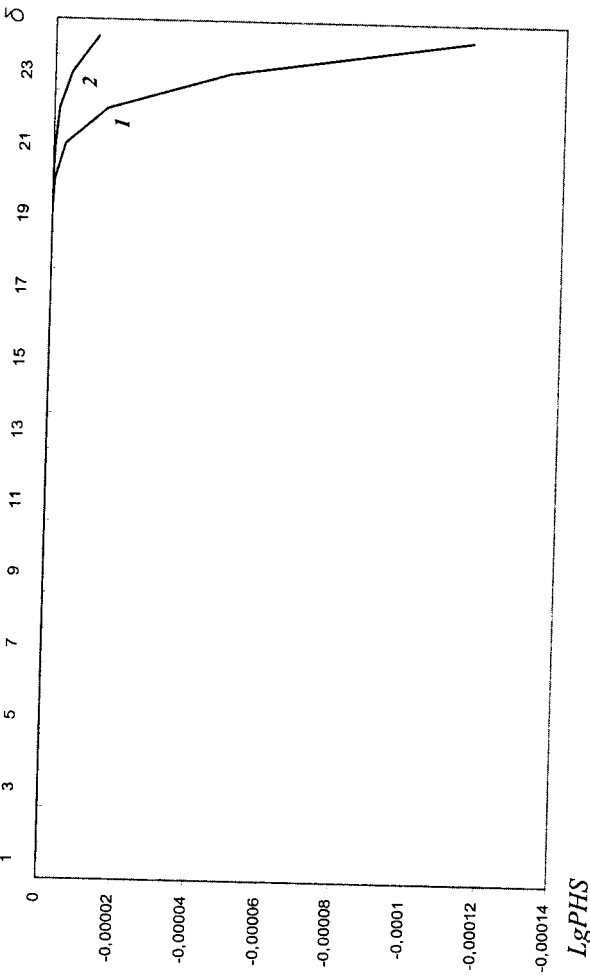


Рис. 2.18. Зависимость вероятности обслуживания э-заявки с необходимой скоростью от δ для моделей с дискретной полосой передачи э-заявок при $N = 25$; $\lambda_1 = 10$; $\mu = 10$;
 1 - $\lambda_2 = 0,1$; 2 - $\lambda_2 = 0,01$

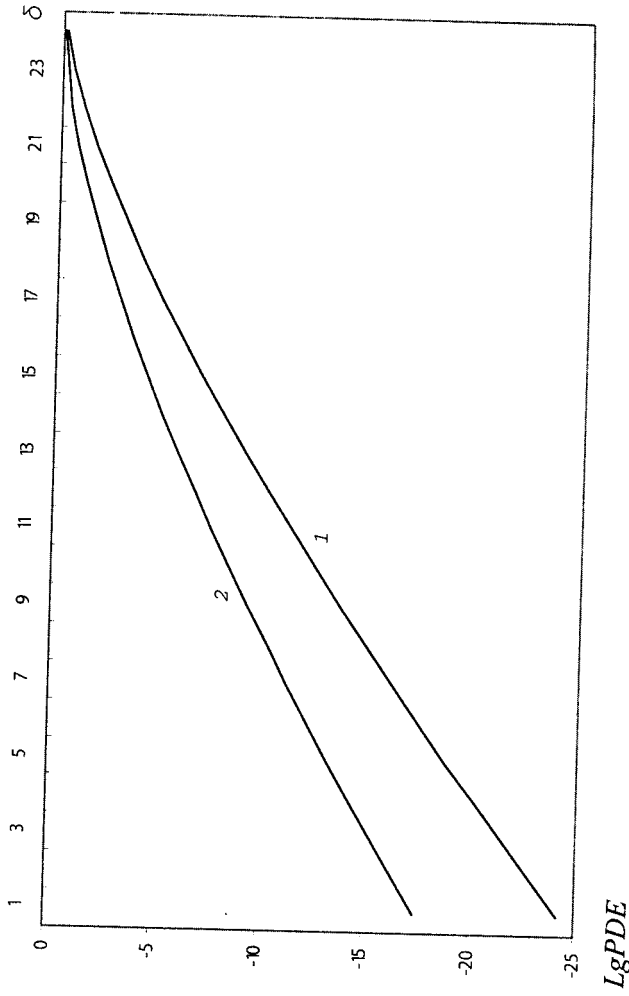


Рис. 2.19. Зависимость вероятности деградации э-заявок от δ для моделей с дискретной полосой передачи э-заявок при $N = 25$; $\lambda_1 = 10$;
 1 - $\lambda_2 = 0,1$; $\mu = 10$; 2 - $\lambda_2 = 0,01$; $\mu = 5$

Выводы и комментарии к главе 2 части II

Смешанные многокортные системы, в которых одновременно обслуживаются неэластичные и эластичные заявки, в доступной литературе мало исследованы. Известно лишь несколько работ в области исследования многокортных систем, в которых обслуживаются лишь эластичные заявки. Так, в главе 3 книги [1] изучены аналогичные модели (здесь же приводится список работ, в которых изучаются подобные модели). Там предполагается, что каждая вновь поступившая заявка при возможности выделении необходимых ресурсов сети использует полосы размера r_{max} , т.е. заявка обслуживается с наилучшим качеством. Если в этот момент сеть не может обеспечить такое качество, то заявка адаптируется к текущей ситуации и принимается на обслуживание с допустимой полосой r_d , где $r_{min} \leq r \leq r_{max}$. Если сеть является нагруженной и не может обеспечить полосу r_{min} , то одна или несколько заявок, которые обслуживаются с наилучшим качеством, могут передать часть своих полос для принятия поступившей заявки. Если сеть нагружена таким образом, что никакие переназначения каналов не позволяют принять поступившую заявку, то она теряется. После ухода из системы некоторой заявки, освобожденные полосы могут быть использованы другими заявками с целью улучшения качества их обслуживания. Показано, что работа этой системы может быть описана с помощью классической модели Эрланга $M/M/0$, т.е. одномерной цепью Маркова, где скалярная величина x указывает число заявок в системе, $0 \leq x \leq \lfloor \frac{N}{r_{min}} \rfloor$, N — общее число полос.

Модели смешанных многокортных систем были исследованы в недавних работах [2–8]. Так, в работах [2–4] изучены частные модели таких систем, в которых э-заявки (пакеты данных типа GPRS) могут требовать 1, 2 или 3 канала, а при поступлении n -э-заявки (речевые вызовы) один канал, занятый обслуживанием э-заявки, может переклучаться для обслуживания поступившей n -э-заявки, если в этот момент все каналы системы заняты. В этих работах предполагалось, что освобожденные каналы остаются свободными и не используются для обслуживания э-заявок.

В работе [5] указанные выше ограничения устранены и рассмотрены общие модели. В ней предполагается, что э-заявки могут требовать 1, 2, ..., M каналов, а переназначение канала для обслуживания вновь поступившей n -э-заявки осуществляется за счет э-заявки с максимальной шириной; если все э-заявки обслуживаются одним каналом, то переназначение не происходит и поступившая n -э-заявка теряется. Здесь освобожденные каналы используются для обслуживания э-заявок, при этом принимается следующая схема: освобожденные каналы в первую очередь выделяются для заявок с наихудшей деградацией. Эти результаты обобщены в работе [6] для моделей с ограниченной очередью n -э-заявок.

В работах [7, 8] изучены модели, аналогичные моделям, исследуемым в п. 2.2.2, но с ограниченной очередью э-заявок и абсолютными приоритетами *n*-э-заявок.

Отметим, что во всех указанных работах, посвященных смешанным моделям многоскоростных систем, для расчета их характеристик используется стандартный метод, основанный на решении соответствующих СУР. Как известно, такой подход является эффективным лишь для моделей малой размерности.

Литература к главе 2 части II

1. **Chen H., Huang L., Kumar S., Kuo C.C.** Radio resource management for multimedia QoS support in wireless networks. – Boston: Kluwer Academic Publishers, 2004. – 256 p.
2. **Chen W.Y., Wu J.L.C., Lu L.L.** Performance comparisons of dynamic resource allocation with/without channel de-allocation in GSM/GPRS networks // IEEE Commun. Letters. – 2003, Vol.7, no.1.
3. **Lin P., Lin Y.B.** Channel allocation for GPRS // IEEE Transaction on Vehicular Technology. – 2001. – Vol.50, no.2. – pp.375 – 387.
4. **Lin P.** Channel allocation for GPRS with buffering mechanisms // Wireless Networks. – 2003. – Vol.9. – pp.431 – 441.
5. **Zhang Y., Soong B.H.** Performance evaluation of GSM/GPRS networks with channel re-allocation scheme // IEEE Commun. Letters. – 2004. – Vol.8, no.5. – pp.280 – 282.
6. **Zhang Y., Soong B.H., Ma M.** A dynamic channel assignment scheme for voice/data integration in GPRS networks // Computer Communications – 2006. – Vol.29, no.9. – pp.1163 – 1173.
7. **Zhuang W., Bensaou B., Chua K.C.** Adaptive quality of service handoff priority scheme for wireless mobile multimedia networks // IEEE Trans. on Vehicular Technology. – 1999. – Vol.40, no.2. – pp. 494 – 505.
8. **Zhuang W., Bensaou B., Chua K.C.** Handoff priority scheme with preemptive, finite queueing and renegeing in mobile multiservice networks // Telecommunication Systems. – 2000. – Vol.15. – pp.37 – 51.

ЗАДАЧИ НАХОЖДЕНИЯ ОПТИМАЛЬНЫХ ЗНАЧЕНИЙ ПАРАМЕТРОВ СТРАТЕГИЙ ДОСТУПА

Некоторые стратегии доступа, исследованные в предыдущих главах этой части, содержат параметры, поддающиеся регулированию. Иными словами, за счет выбора подходящих значений этих параметров, можно добиться желаемых значений интересующих (выбираемых) характеристик исследуемой системы.

В настоящей главе рассматриваются некоторые задачи данного типа и предлагаются алгоритмы их решения. Эти задачи представляют собой научный и, особенно, практический интерес. Реализация результатов рассматриваемых задач в реальных мультимедийных системах не представляет никаких трудностей.

Поскольку рассматриваемые задачи для общей модели *MRQ* с произвольным числом входящих трафиков представляются собой сложные математические проблемы, то здесь подробно исследуются аналогичные задачи для моделей типа Гимпельсона. Приводятся результаты численных экспериментов для рассматриваемых задач и дан их подробный содержательный анализ.

3.1. Задачи для моделей типа Гимпельсона

3.1.1. Решение задачи эквивалентной производительности при CS-стратегии доступа

Результаты исследований, приведенные в § 2.1.1 части I, показывают, что при использовании неконтролируемой CS-стратегии доступа в каналы в наличии имеется лишь очень ограниченное число управляемых параметров. Зачастую в реальных системах оказывается достаточно трудно управлять входящим трафиком, т.е. в зависимости от текущей ситуации динамически управлять интенсивностями входящих трафиков, хотя в некоторых работах исследуются и такие системы. Поэтому реальными являются проблемы улучшения желаемых показателей *QoS* системы за счет выбора соответствующих значений скоростей обслуживания заявок (особенно с помощью изменения значения параметра *m*, который указывает число каналов, требуемых для обслуживания *m*, который ла каналов обслуживания).

Анализ вычислительных экспериментов по расчету показателей *QoS* системы при CS-стратегии доступа показывает их неординарное поведение, особенно это относится к $PB_n(N, m)$, во всей допустимой области изменения как структурных, так и нагрузочных параметров модели, что затрудняет формулировку и решения задач оптимизации модели при использовании данной стратегии доступа.

Вместе с тем, для данной системы интересной, как теоретической, так и практической точек зрения, является решение задачи эквивалентной производительности. Эта задача для данной системы формулируется следующим образом: при заданном числе каналов и при известных интенсивностях поступления n - и w -заявок требуется найти такие минимальные значения скоростей их обслуживания, чтобы вероятности потери не превышали заданных пределов.

Теперь рассмотрим математическую постановку этой задачи. Она записывается так:

$$\mu_n \rightarrow \min \quad (3.1)$$

при ограничениях

$$PB_w^{CS}(N, m) \leq \varepsilon_w, \quad (3.2)$$

$$PB_n^{CS}(N, m) \leq \varepsilon_n, \quad (3.3)$$

где ε_n и ε_w являются заданными величинами.

Сначала предположим, что скорость канала обслуживания является непрерывной величиной и пусть $\mu_w = m\mu_n$. Последнее допущение отражает реальную ситуацию, т.к. поскольку w -заявки обслуживаются одновременно m каналами, то естественно предполагать, что их скорость обслуживания в m раз превосходит скорость обслуживания n -заявок.

Обозначим оптимальное решение задачи (3.1)–(3.3) через μ^* . Для решения этой задачи с учетом монотонности функций в левых частях ограничений (3.2) и (3.3) можно использовать алгоритм деления пополам. С этой целью требуется предварительно определить границы интервала $[\underline{\mu}, \bar{\mu}]$, в котором находится μ^* . Эти границы могут быть определены, в частности, как решение параллельных задач эквивалентной производительности для двух вспомогательных «минорантной» и «мажорантной» (относительно исходной) систем обслуживания, т.е. $\underline{\mu}(\underline{\mu})$ является решением указанной задачи для «минорантной» («мажорантной») системы $M/M/0 (M/M/[N/m]/0)$ с интенсивностью поступления однотипных заявок $\lambda_n(\lambda_n + \lambda_w)$ и с требуемым порогом вероятности потери ε , где $\varepsilon = \min\{\varepsilon_n, \varepsilon_w\}$. Задача эквивалентной производительности для однополюсных систем обслуживания легко решается и потому решения задач нахождения $\underline{\mu}$ и $\bar{\mu}$ не представляют принципиальных трудностей.

Некоторые результаты решения задачи (3.1)–(3.3) с использованием указанного подхода показаны в табл. 3.1, при этом погрешность нахождения μ^* не превышает 0.1%.

Таблица 3.1
Результаты решения задачи (3.1)–(3.3)

N	60	60	60	50	50	50	40	40	40	30
m	6	6	12	6	12	12	12	12	6	6
λ_n	50	40	30	30	30	50	45	60	60	50
λ_w	50	30	40	40	30	60	40	40	60	40
ε_n	10^{-4}	10^{-5}	10^{-5}	10^{-6}	10^{-6}	10^{-7}	10^{-5}	10^{-4}	10^{-5}	10^{-5}
ε_w	10^{-5}	10^{-5}	10^{-5}	10^{-6}	10^{-6}	10^{-7}	10^{-4}	10^{-4}	10^{-6}	10^{-6}
μ^*	7,69	5,04	23,86	11,91	46,52	141,95	40,83	44,07	34,46	77,93

Анализ результатов решения задачи (3.1)–(3.3) показывает, что с ростом нагрузки входящих трафиков и ширины w -заявок оптимальное значение μ^* растет, а с увеличением ε_n и/или ε_w оно уменьшается.

В случае, когда скорость канала обслуживания может принимать лишь дискретные и конечные значения, задача (3.1)–(3.3) еще больше упрощается, и в таких случаях оптимальное решение данной задачи (если оно существует) определяется точно.

В заключение данного параграфа отметим, что решение аналогичной задачи для системы, в которой принята CSE-стратегия доступа, еще больше упрощается, так как при использовании этой стратегии вероятности потери n - и w -заявок равны друг другу. В связи с этим здесь не рассматривается аналогичная задача оптимизации для CSE-стратегии доступа.

3.1.2. Задачи нахождения оптимальных значений параметров TR-стратегии доступа

Вначале рассмотрим частную модель Гимпельсона с N каналами, в

которой $\mathbf{b} = \left(\underbrace{1, \dots, 1}_k, \underbrace{N, \dots, N}_{k_2} \right)$. Очевидно, что для данной модели допустимыми значениями параметров резервирования каналов являются $r_N = 0$ и $1 \leq r_1 \leq N - 1$ (при $r_1 = 0$ получаются результаты для CS-стратегии).

С использованием описанного в § 2.1.3 алгоритма находим, что

$$PB_1(r_1) = \hat{v}_1 \pi(0), \quad PB_N(r_1) = 1 - \pi(0), \quad (3.4)$$

$$\text{где } \hat{v}_1 := \sum_{i=1}^{k_1} v_i, \quad \hat{v}_2 := \hat{v}_3 := \dots := \hat{v}_{N-1} := 0, \quad \hat{v}_N := \sum_{i=k_1+1}^{k_1+k_2} v_i, \quad \pi(0) = \left(\hat{v}_N + \sum_{i=0}^{N-1} \frac{\hat{v}_1^i}{i!} \right)^{-1}.$$

Здесь $PB_1(r_1)$ означает вероятность блокировки заявок любого типа, требующих одновременно i каналов, $i = 1$ и N .

Из формулы (3.4) можно сделать следующие выводы:

- 1) при фиксированных \hat{v}_1, \hat{v}_N и r_1 с ростом N функция $PB_N(r_1)$ растет, а функция $PB_1(r_1)$ убывает;
- 2) при фиксированных \hat{v}_1, \hat{v}_N и N с ростом r_1 функция $PB_1(r_1)$ растет, а функция $PB_N(r_1)$ убывает;
- 3) $PB_N(r_1) > PB_1(r_1)$ при любых \hat{v}_1, \hat{v}_N, N и r_1 , так как это неравенство

$$\text{эквивалентно неравенству } \sum_{i=0}^{N-1} \frac{\hat{v}_1^i}{i!} > 1;$$

$$4) \min_{r_1} \{PB_N(r_1) - PB_1(r_1)\} = PB_N(N-1) - PB_1(N-1) = \hat{v}_1 / (1 + \hat{v}_1 + \hat{v}_N). \quad (3.5)$$

Ряд задач оптимизации для данной модели Гимпельсона решается даже аналитически с использованием разработанного алгоритма. Так, рассмотрим задачу минимизации взвешенной суммы вероятностей блокировки $PB_1(r_1)$ и $PB_N(r_1)$:

$$WPB(r_1) := \frac{\hat{v}_N}{\hat{v}_1 + \hat{v}_N} \cdot PB_N(r_1) + \frac{\hat{v}_1}{\hat{v}_1 + \hat{v}_N} \cdot PB_1(r_1) \xrightarrow{r_1} \min. \quad (3.6)$$

С использованием формулы (3.4) можно показать, что задача (3.6) эквивалентна следующей задаче:

$$\sum_{i=0}^{N-1} \frac{\hat{v}_1^i}{i!} + \hat{v}_N \xrightarrow{r_1} \max. \quad (3.7)$$

Отсюда заключаем, что при любых возможных значениях N, \hat{v}_1 и \hat{v}_N решением задачи (3.7), и следовательно, задачи (3.6), является $r_1^* = 1$.

Аналогичным путем можно доказать, что максимальное значение среднего числа занятых каналов также достигается при $r_1^* = 1$. Таким образом, для данного класса моделей Гимпельсона оптимальное значение параметра резервирования для узкополосных заявок в обоих безусловных задачах минимизации взвешенной суммы вероятностей блокировок и максимизации утилизации каналов является $r_1^* = 1$.

Указанные выше факты делают актуальными проблемы организации справедливого обслуживания разнотипных заявок. Одна из таких задач — организация ϵ -справедливого обслуживания состоит в следующем: при

фиксированных значениях N, \hat{v}_1 и \hat{v}_N , требуется найти такое минимальное значение параметра r_1 , чтобы выполнялось условие:

$$PB_N(r_1) - PB_1(r_1) \leq \epsilon, \quad (3.8)$$

где $\frac{\hat{v}_1}{1 + \hat{v}_1 + \hat{v}_N} < \epsilon < 1$ (согласно равенству (3.5), при остальных значениях

параметра ϵ данная задача не имеет решения).

Легко можно показать, что задача (3.8) эквивалентна нахождению максимального целочисленного решения относительно x следующего неравенства:

$$\sum_{i=0}^x \frac{\hat{v}_1^i}{i!} \leq \frac{1 + \hat{v}_N}{1 - \hat{v}_N} - \hat{v}_N, \quad (3.9)$$

где $x := N - r_1$.

С учетом того, что левая часть неравенства (3.9) является монотонно возрастающей функцией от x , то для нахождения его максимального значения (т.е. соответственно минимального значения r_1), может быть использован метод деления пополам.

Общая схема данного подхода состоит в следующем. Если в средней точке интервала $[1, N-1]$ (т.е. в $[N/2]$) условие (3.9) удовлетворяется, то рассматривается новый интервал $[[N/2], N-1]$, в противном случае рассматривается интервал $[1, [N/2]]$. Условием проверки окончания алгоритма является нахождение такого интервала единичной длины, для которого левая граница удовлетворяет условию (3.9), а правая — нет. Тогда, именно эта левая граница является искомым величиной x^* , т.е. $r_1^* = N - x^*$. Алгоритм является конечным, так как левая часть неравенства (3.9) является монотонно возрастающей функцией от x .

Теперь рассмотрим общую модель Гимпельсона с двумя типами трафиков. Для простоты изложения, не нарушая при этом общности, предположим, что $b_1 = 1$ и $b_2 = m > 1$.

В отличие от рассмотренной выше частной модели Гимпельсона, в данной модели удается достичь абсолютного справедливого обслуживания в смысле равенства вероятностей потерь разнотипных заявок. Действительно, из разработанных в § 2.1.3 формул для расчета указанных показателей QoS легко можно получить, что при любых значениях N, v_1 и v_2 имеет место следующее соотношение:

$$PB_1(r_1, r_2) = PB(r_1, r_2), \quad \text{если } r_1 - r_2 = m - 1. \quad (3.10)$$

Отметим, что этот результат имеет простое интуитивное объяснение. Действительно, при выполнении условия $r_1 - r_2 = m - 1$ заявки обоих типов имеют одинаковую «ширину», т.е. заявки обоих типов принимаются для обслуживания при наличии $m + r_2 = 1 + r_1$ свободных каналов, и поэтому вероятности их потерь являются одинаковыми.

С использованием разработанных в § 2.1.3 формул также можно показать, что имеют место следующие соотношения:

С использованием результатов решения задачи (3.14) можно также решить задачу максимизации каналов ($\tilde{N}(r_m)$):

$$\tilde{N}(r_2) \rightarrow \max_{r_2} \quad (3.17)$$

Решение данной задачи не представляет никаких трудностей, так как функция $\tilde{N}(r_2)$ является убывающей, т.е. максимальное значение данной функции достигается в точке r_2^* . Некоторые результаты решения рассмотренных задач (5.13) и (5.16) показаны в табл. 3.3.

Таблица 3.3

Результаты решения задач (3.14), (3.17) при $N=50, m=5, r_1=0$

V_1	V_m	ϵ_1	ϵ_2	$[r_2^*, r_2^{**}]$
5	2	10^{-3}	10^{-1}	[1,22]
5	10	10^{-3}	10^{-1}	\emptyset
5	5	10^{-2}	10^{-1}	[1,5]
5	5	$2 \cdot 10^{-3}$	10^{-1}	[1,5]
5	5	10^{-3}	$2 \cdot 10^{-1}$	[1,13]
5	15	10^{-4}	$7 \cdot 10^{-4}$	[4,18]
15	15	10^{-4}	$7 \cdot 10^{-1}$	[7,9]
15	10	10^{-1}	$5 \cdot 10^{-1}$	[1,4]
15	5	10^{-1}	$5 \cdot 10^{-1}$	[1,17]
15	5	10^{-3}	$3 \cdot 10^{-1}$	[3,8]
10	10	10^{-3}	$3 \cdot 10^{-1}$	\emptyset
10	10	10^{-1}	$5 \cdot 10^{-1}$	[1,9]
10	5	10^{-1}	$5 \cdot 10^{-1}$	[1,22]
10	5	10^{-4}	$5 \cdot 10^{-1}$	[4,22]
4	5	10^{-4}	10^{-1}	[2,6]
4	5	10^{-6}	10^{-1}	[4,6]

Анализ результатов решения задач (3.14) и (3.17) позволяет сделать следующие выводы:

- с ростом нагрузок входящих графиков оптимальный интервал уменьшается, а с увеличением ϵ_1 и/или ϵ_2 оптимальный интервал также увеличивается;
- оптимальное решение задачи не является сверх чувствительным относительно ϵ_1 .

3.1.3. Задачи нахождения оптимальных значений параметров стратегии доступа со специальными каналами для обслуживания широкополосных заявок

При использовании указанной стратегии доступа в наличии имеется два управляемых параметра (т.е. A и k_w) и, следовательно, за счет выбора соответствующих значений этих параметров можно достичь желаемого уровня потери разнотипных заявок. При этом с учетом монотонности исследуемых функций по обоим параметрам (см. § 2.2.2) здесь для конкретности изложения рассматривается лишь одна задача оптимизации модели, где в качестве управляемого параметра выбирается параметр, ограничивающий максимальное число w -заявок в общей зоне (т.е. k_w). Иными словами, предполагается, что общее число каналов (N) и размер индивидуальной зоны для обслуживания w -заявок (A) являются фиксированными, а нагрузки графиков (V_n и V_w) также считаются известными. Для подчеркивания зависимости вероятностей потерь от параметра k_w и для удобства дальнейшего изложения указанные вероятности обозначаются $PB_n(\bar{k}_w)$ и $PB_w(\bar{k}_w)$.

Пусть эффективность работы системы оценивается через вероятности потерь разнотипных заявок и заданы ограничения на указанные показатели качества обслуживания:

$$PB_n(\bar{k}_w) \leq \epsilon_n, \quad (3.18)$$

$$PB_w(\bar{k}_w) \leq \epsilon_w, \quad (3.19)$$

где ϵ_n и ϵ_w являются заданными величинами.

Тогда задача оптимизации ставится следующим образом. При фиксированных N, m, A, V_n и V_w требуется найти такой интервал изменения значений \bar{k}_w максимальной длины $[\bar{k}_w^*, \bar{k}_w^{**}]$, внутри которого удовлетворяются условия (3.18) и (3.19).

При разработке алгоритма решения данной задачи используются свойства монотонности функций $PB_n(\bar{k}_w)$ и $PB_w(\bar{k}_w)$, т.е. имеют место следующие соотношения:

$$PB_n(1) \leq PB_n(\bar{k}_w) \leq PB_n\left(\left[\frac{N}{m} - A\right]\right), \quad (3.20)$$

$$PB_w\left(\left[\frac{N}{m} - A\right]\right) \leq PB_w(\bar{k}_w) \leq PB_w(1). \quad (3.21)$$

Тогда с учетом (3.20) и (3.21) можно предложить следующий алгоритм решения рассматриваемой задачи.

Шаг 1. Если $PB_n(1) > \varepsilon_n$ и/или $PB_w \left(\left[\frac{N}{m} - A \right] \right) > \varepsilon_w$, то задача не имеет

решения.

Шаг 2. Параллельно решаются следующие задачи:

$$k_n^* := \arg \max \{ PB_n(\bar{k}_w) \leq \varepsilon_n \}, \quad (3.22)$$

$$k_w^* := \arg \min \{ PB_w(\bar{k}_w) \leq \varepsilon_w \}. \quad (3.23)$$

Шаг 3. Если $k_n^* < k_w^*$, то задача не имеет решения; в противном случае решением задачи является $\bar{k}_w^* := k_w^*$, $\bar{k}_n^{**} := k_n^*$.

Для решения задач (3.22), (3.23), в частности, может быть применен метод дихотомии. Отметим, что аналогичным образом может быть сформулирована и решена задача выбора оптимального значения параметра A .

Часть результатов приведенных вычислительных экспериментов показаны в табл. 3.4, где символ \emptyset означает, что задача не имеет решения.

Таблица 3.4

Результаты решения задачи (3.18), (3.19) при $N = 20$, $m = 3$

V_n	10	10	10	10	20	25	15	10	10	10	10	10
V_w	10	10	10	10	10	10	30	10	10	10	10	10
A	4	7	7	7	5	5	5	5	5	10	10	10
ε_n	10^{-2}	10^{-2}	10^{-1}	10^{-1}	10^{-3}	10^{-3}	10^{-3}	10^{-1}	10^{-2}	10^{-1}	10^{-1}	10^{-1}
ε_w	10^{-3}	10^{-3}	10^{-3}	10^{-4}	10^{-2}	10^{-2}	10^{-2}	10^{-2}	10^{-1}	10^{-1}	10^{-1}	10^{-2}
$[\bar{k}_w^*, \bar{k}_w^{**}]$	$[17,22]$	$[14,19]$	$[14,19]$	\emptyset	$[13,16]$	\emptyset	\emptyset	$[13,21]$	$[3,16]$	$[3,16]$	$[8,16]$	

Анализ результатов вычислительных экспериментов позволяет сделать ряд выводов:

- с ростом значения A оптимальный интервал движется влево;
- с одновременным уменьшением величин ε_n и ε_w длина оптимального интервала сокращается, и ниже их определенных значений, задавае-

мых с помощью соотношений (3.20) и (3.21), задача вообще не имеет решения;

- с ростом нагрузок длина оптимального интервала сокращается и выше их определенных критических значений задача не имеет решения.

3.2. Задачи нахождения оптимальных значений параметров стратегии доступа для моделей с неэластичными и эластичными заявками

Теперь рассмотрим задачи оптимизации характеристик моделей с неэластичными и эластичными заявками. При этом соответствующие задачи для моделей с одной и многими эластичными заявками рассматриваются отдельно.

3.2.1. Задачи нахождения оптимальных значений параметра стратегии доступа для моделей с одной эластичной заявкой на обслуживании

В моделях данного типа с равноправным доступом для заявок обоих типов отсутствуют параметры, которые поддаются регулированию, и поэтому для данного класса моделей не удается сформулировать задачи оптимизации их характеристик. Исходя из этого, здесь рассматривается задача оптимизации моделей данного класса, в которых ограничивается доступ неэластичных заявок к каналам системы.

Напомним, что в моделях с ограниченным доступом для неэластичных заявок предполагается, что n -я заявка не принимается на обслуживание, если в момент ее поступления число свободных каналов, занятых обслуживанием ε -заявок равно R , где $1 \leq R \leq N$ (в частном случае при $R = 1$ из данной модели получается модель с равноправным доступом).

Наличие в данной модели порогового параметра R позволяет ставить задачу о выборе его подходящих значений (если они существуют) с целью достижения желаемых значений характеристик модели, т.е. можно потребовать, чтобы вероятности потери разнотипных заявок были ниже заранее заданных величин, а также чтобы утилизация каналов была не ниже определенного значения.

Одна из возможных постановок задач подобного типа более точно формулируется следующим образом.

Пусть общее число каналов системы и нагрузки разнотипных потоков являются фиксированными. Пусть также заданы ограничения на вероятность потери разнотипных заявок, а также на утилизацию каналов, т.е. потребуются выполнение следующих условий:

$$PB_\varepsilon(R) \leq \varepsilon_\varepsilon, \quad (3.24)$$

$$PB_{n,\varepsilon}(R) \leq \varepsilon_{n,\varepsilon}, \quad (3.25)$$

$$\bar{N}(R) \geq N_0, \quad (3.26)$$

где ε_s , $\varepsilon_{н.э}$ и N_0 являются заданными величинами (здесь с целью подчеркивания зависимости функций PB_s , $PB_{н.э}$ и \tilde{N} лишь от аргумента R этот параметр явным образом указан в скобках).

Требуется найти такое максимальное (\bar{R}) и минимальное (\underline{R}) значения параметра R , чтобы удовлетворялись ограничения (5.32)–(5.34), т.е. необходимо найти такие $R \in [\underline{R}, \bar{R}]$, $\underline{R}, \bar{R} \in [1, N]$, чтобы выполнялись условия (3.24)–(3.26).

С учетом монотонности функций PB_s , $PB_{н.э}$ и \tilde{N} для решения данной задачи можно предложить следующий алгоритм.

На первом этапе параллельно решаются следующие задачи:

$$R_0 := \arg \min_R \{ \tilde{N}(R) \geq N_0 \}, \quad (3.27)$$

$$R_1 := \arg \max_R \{ PB_s(R) \leq \varepsilon_s, PB_{н.э}(R) \leq \varepsilon_{н.э} \}. \quad (3.28)$$

Далее решение исходной задачи определяется так:

$$[\underline{R}, \bar{R}] = \begin{cases} [R_0, R_1] & \text{если } R_1 \geq R_0, \\ \emptyset, & \text{если } R_1 < R_0. \end{cases} \quad (3.29)$$

Результаты решения данной задачи при двух существенно различных соотношениях нагрузок разнотипных заявок показаны в табл. 3.5 и 3.6.

Их анализ позволяет сделать ряд выводов:

- оптимальное решение задачи является очень чувствительным относительно значения N_0 при $\nu_1/\nu_2 = 100$. Так, из табл. 3.5 видно, что при увеличении значения данного параметра всего на 0,01 оптимальное решение резко изменяется, т.е. если, например, при $N = 50$, $\varepsilon_s = 0,01$, $\varepsilon_{н.э} = 0,001$ и $N_0 = 10,02$ оптимальное решение определяется как [30; 32], то уже при тех же исходных данных, но при $N_0 = 10,03$ решение вообще не существует;

- оптимальное решение задачи не является сверхчувствительным относительно N_0 при $\nu_1/\nu_2 = 10$ (см. табл. 3.6);

- уменьшение ε_s и/или $\varepsilon_{н.э}$ приводят так же к уменьшению длины интервала $\bar{R} - \underline{R}$ (см. табл. 3.5 и 3.6).

Таблица 3.5

Результаты решения задачи (3.24) – (3.26): $\lambda_1 = 10$, $\lambda_2 = 0,1$, $\mu = 5$
($\nu_1 = 2$, $\nu_2 = 0,02$)

N	ε_s	$\varepsilon_{н.э}$	N_0	$[\underline{R}, \bar{R}]$
40	0,1	0,1	2	[1;26]
40	0,3	0,4	10	\emptyset
50	0,1	0,001	10,02	[30;36]
50	0,01	0,001	10,02	[30;32]
50	0,01	0,001	10,03	\emptyset
50	0,0005	0,0001	10	[1;20]
70	0,0005	0,0001	10	[1;47]
70	0,01	0,01	10,02	[49;52]
70	0,001	0,01	10,02	\emptyset
70	0,1	0,01	10,02	[49;56]
70	0,1	0,01	10,03	\emptyset
70	0,6	0,2	10,03	[60;61]
80	0,6	0,2	10,03	[70;71]
80	0,6	0,001	10,03	[70;71]
80	0,6	0,001	10,02	[59;71]

Таблица 3.6

Результаты решения задачи (3.24) – (3.26): $\lambda_1 = 40$, $\lambda_2 = 4$, $\mu = 2$
($\nu_1 = 20$, $\nu_2 = 2$)

N	40	40	40	40	40	50
ε_s	0,1	0,1	0,1	0,1	0,1	0,1
$\varepsilon_{н.э}$	0,1	0,1	0,001	0,001	0,001	0,001
N_0	22	21	21	15	15	10
$[\underline{R}, \bar{R}]$	\emptyset	[1;9]	[1;9]	[1;9]	[1;21]	[1;21]

3.2.2. Задачи нахождения оптимальных значений параметров стратегии доступа для моделей со многими эластичными заявками на обслуживание

Ввиду идентичности постановки задач и методов их решения, здесь рассматриваются лишь задачи оптимизации моделей MRQ с дискретной полосой передачи эластичных заявок.

Для конкретности изложения рассмотрим следующую задачу.

Пусть заданы ограничения на верхние границы вероятности потерь различного типа заявок и вероятности деградационных состояний эластичных заявок (напомним, что для моделей с дискретной полосой передачи вероятности потерь эластичных и неэластичных заявок являются равным друг другу). Требуется найти такое минимальное число каналов системы и максимальное значение порогового параметра (δ) для ε -заявок, чтобы удовлетворялись заданные ограничения на вероятности потерь заявок и вероятности деградационных состояний ε -заявок.

Указанная задача формально записывается так:

$$N \rightarrow \min, \quad (3.30)$$

$$\delta \rightarrow \max \quad (3.31)$$

при ограничениях

$$PB(N) \leq \varepsilon, \quad (3.32)$$

$$PDS(N, \delta) \leq \gamma, \quad (3.33)$$

где ε и γ являются заданными величинами.

Поскольку вероятность потери не зависит от порогового параметра δ и обе функции PB и PDS являются монотонно убывающими относительно аргумента N (для функции PDS при фиксированном значении δ), то для решения задачи (3.30)–(3.33) можно предложить следующий алгоритм.

Вначале находится такое минимальное значение N , обозначаемое через N_1 , чтобы удовлетворить ограничения (3.32). Для нахождения N_1 , в частности, может быть применен такой способ. Границы изменения N_1 определяются исходя из пределов изменения функции PB :

$$E_B(V_1, N) \leq PB(N) \leq E_B(V_1 + V_2, N). \quad (3.34)$$

Левая граница в неравенстве (3.34) непосредственно вытекает из формулы для $PB(N)$ (см. шаг 3 алгоритма в § 2.2.2), правая часть границы является очевидной.

После нахождения N_1 проверяется условие: $PDS(N_1, 1) \leq \gamma$. Если это условие выполняется, то в интервале $[1, N_1]$ находится такое максимальное значение δ , чтобы выполнялось условие $PDS(N_1, \delta) \leq \gamma$. Если $PDS(N_1, 1) > \gamma$, то необходимо увеличить число каналов до тех пор, пока не выполнится условие $PDS(N_1, 1) \leq \gamma$ и после этого находится соответствующее значение δ .

Таким образом, резюмируя можно предложить следующий алгоритм решения задачи (3.30)–(3.33). Для общности рассмотрим k -ю итерацию.

Шаг 1. Найти такое минимальное N_k , чтобы выполнялось условие (3.32).

Шаг 2. Если $PDS(N_k, 1) \leq \gamma$, то найти такое максимальное значение $\delta_k \in [1, N_k]$, чтобы выполнялось условие (3.33).

Шаг 3. Полагается $N_{k+1} := N_k + 1$ и осуществляется переход к шагу 1.

Алгоритм является убывающей относительно N (число каналов) при фиксированном значении δ .

Некоторые результаты решения этой задачи показаны в табл. 3.7.

Таблица 3.7

Результаты решения задачи (3.30)–(3.33)

ε	10^{-2}	10^{-2}	10^{-2}	10^{-2}	10^{-2}	10^{-2}	10^{-2}	10^{-2}	10^{-2}	10^{-3}	10^{-3}	10^{-3}	10^{-3}	10^{-3}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-5}	10^{-5}	10^{-6}	10^{-7}
γ	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-4}	10^{-5}	10^{-4}	10^{-4}	10^{-4}	10^{-4}
N_{\min}	18	19	22	25	21	21	21	22	25	24	24	24	25	24	24	25	27	29	31			
δ_{\max}	4	2	2	2	7	3	2	2	2	11	6	3	2	7	9	11						

Анализ результатов задачи (3.30)–(3.33) позволяют сделать следующие выводы:

- при фиксированных требованиях к вероятностям потери заявок (неэластичных или эластичных) ужесточение требований к вероятности деградационных состояний приводит к увеличению общего числа необходимых каналов, при этом соответствующее максимальное значение порогового параметра для ε -заявок, вообще говоря, не увеличивается;
- при фиксированных требованиях к вероятности деградационных состояний ужесточение требований к вероятности потери заявок (неэластичных или эластичных) также приводит к увеличению общего числа необходимых каналов, но при этом соответствующее максимальное значение порогового параметра для ε -заявок также увеличивается.

Заключение и комментарий к главе 3 части II

Задачи улучшения характеристик многократных систем за счет выбора соответствующих значений параметров используемых стратегий доступа мало исследованы, хотя они являются весьма актуальными видами следующих причин. Во-первых, применение соответствующих результатов позволяет без особых усилий существенно образом улучшить желаемые характеристики системы. Во-вторых, имеется возможность изучить все необходимые характеристики системы при найденных значениях параметров используемой стратегии доступа и адаптировать

их к текущим трафикам. В-третьих, найденные стратегии доступа легко реализуются в реальных системах.

Рассматриваемые в данной главе задачи, а также другие задачи подобного типа подробно исследованы в указанных ниже работах.

Литература к главе 3 части II

1. **Лохмотко В.В., Пирогов К.И.** Анализ и оптимизация цифровых сетей интегрального обслуживания. – Минск: Наука и техника, 1991. – 192 с.
2. **Меликов А.З., Фаттахова М.И., Казиев Т.С.** Численные методы исследования многократных систем обслуживания типа Гимпельсона // System research & information technologies. – 2005. – № 2. – С. 83 – 96.
3. **Меликов А.З., Фаттахова М.И., Казиев Т.С.** Метод исследования характеристик многократных сетей обслуживания с одним ресурсом и потерями // Электронное моделирование. – 2005. – Т. 27, № 5. – С. 35 – 45.
4. **Меликов А.З., Фаттахова М.И., Казиев Т.С.** Анализ и оптимизация многократной системы обслуживания со стратегией резервирования каналов // Электронное моделирование. – 2006. – Т. 28, № 5. – С. 71 – 85.
5. **Меликов А.З., Фаттахова М.И., Казиев Т.С.** Многократная система со специализированными каналами для обслуживания широкополосных заявок // Автоматика и вычислительная техника. – 2006. – № 2. – С.15 – 25.
6. **Melikov A.Z., Deniz D.Z.** Non-exhaustive channel access strategy in multi-resource communication systems with non-homogeneous traffic // Proc. of 5th IEEE Symposium on Computers and Communications, July 3-6, 2000, France, pp.432 – 437.

ЧАСТЬ III

МАРКОВСКИЕ ПРОЦЕССЫ ПРИНЯТИЯ РЕШЕНИЙ В ЗАДАЧАХ ОПТИМИЗАЦИИ МОДЕЛЕЙ СИСТЕМ ТЕЛЕГРАФИКА

В предыдущих частях данной книги были рассмотрены различные задачи улучшения характеристик традиционных и мультимедийных беспроводных сетей связи. При их решении улучшение желаемых характеристик достигается, в основном, за счёт выбора соответствующих значений параметров используемой стратегии доступа. Иными словами, класс допустимых стратегий доступа определяется заранее, и задача состоит в определении оптимальных (в известном смысле) значений параметров стратегии доступа.

Такой подход к решению указанных задач, естественно, сужает множество стратегий, в которых определяется оптимальная стратегия доступа. Для решения этих задач в широком классе стратегий целесообразно использовать подход теории марковских процессов принятия решений (МППР).

В данной части книги рассматриваются как классические, так и новые методы теории МППР, которые позволяют определить оптимальные стратегии доступа в системах телеграфика.

Глава 1

МЕТОДЫ ИССЛЕДОВАНИЯ ЗАДАЧ МППР

1.1. Классическая постановка задач МППР и методы их исследования

В системах телеграфика, как правило, вызовы (заявки) отличаются друг от друга как по нагрузочным характеристикам, так и по важности (срочности). Это обстоятельство делает актуальным исследование моделей с различными дисциплинами обслуживания, в том числе и с различными видами приоритетов.

Любая приоритетная дисциплина обслуживания в системах телеграфика задаёт правила для (А) приёма вызова в систему, (В) выбора типа вызова, направляемого на обслуживание при освобождении канала, и (С) назначения канала обслуживания. При этом необходимость в определении правила С возникает в тех системах, где каналы не идентичны (например, отличаются друг от друга по скорости обслуживания и/или по стоимости включения и стоимости работы в единицу времени и т.д.).

Указанные выше правила А-С реализуются с помощью приоритетов двух типов: внесистемных (экзогенных) и внутрисистемных (эндогенных).

При использовании внесистемных приоритетных дисциплин правила А-С определяются на основе некоторого заранее разработанного регламента и не учитывают текущего состояния системы, причём состояние системы может оцениваться и учитываться различными способами. С точки зрения удобства практического использования более предпочтительными являются приоритеты этого типа, не требующие при их реализации особых затрат на математическое обеспечение. Однако реальные системы телеграфика функционируют в условиях неопределённости относительно параметров входных трафиков, что делает невозможным во многих случаях жесткое предварительное назначение приоритетов. Эти обстоятельства сделали актуальными вопросы изучения моделей систем телеграфика с внутрисистемными приоритетами.

Внутрисистемные приоритеты в системах телеграфика могут быть определены по следующей схеме. Для каждого состояния выделяется конечное множество допустимых решений (управлений), и каждому решению ставится в соответствие некоторый параметр, оценивающий вероятность принятия данного решения в текущем состоянии. Критерии качества в конкретных системах телеграфика определяются по-разному, исходя из предназначения системы. Вместе с тем, актуальным является исследование моделей с критерием, представляющим собой функционал (линейный или нелинейный), оценивающий суммарные издержки (экономические, технические и т. д.), связанные с пребыванием систем в тех или иных состояниях. В таких случаях в марковских моделях имеется универсальный метод их вычисления, заключающийся в выражении этого функционала через стационарные вероятности состояний. Это удаётся благодаря уникальному свойству марковских систем, для которых стационарная вероятность состояния представляет собой долю времени пребывания системы в соответствующем состоянии за достаточно большой интервал времени наблюдения.

Тогда задача определения оптимальных внутрисистемных приоритетов состоит в выборе соответствующих решений в конфликтных состояниях, т.е. в состояниях, в которых имеется необходимость выбрать решение из некоторого конечного или бесконечного множества решений. Таким образом, исследование моделей марковских систем телеграфика с внутрисистемными приоритетами эквивалентно исследованию некоторой задачи МППР.

Сформулируем один из возможных способов определения МППР.

МППР с бесконечным горизонтом планирования и без дисконтирования считается заданным, если определены следующие объекты:

1. Задана ЦМ с конечным фазовым пространством состояний $X := \{x_1, x_2, \dots, x_N\}$ (будем использовать далее в тексте символ $x_k, k = 1, \dots, N$, для обозначения состояний, а в формулах для обозначения состояний употреблять символ k).

2. Задано конечное множество решений $D, D := \bigcup_{k \in X} D_k$, где D_k — мно-

жество допустимых управлений (решений) в состоянии $x_k \in X$.

3. Заданы вероятностный закон принятия решений $\alpha^d := Pr \{управление d | состояние x_k \}$ и соответствующая ему переходная матрица $P = \|p^d(k, k')\|$, где $p^d(k, k')$ — вероятность перехода из состояния x_k в состояние $x_{k'}$ при выборе в состоянии x_k управления $d \in D_k$. При этом

$$\alpha_k^d \geq 0, \sum_{d \in D_k} \alpha_k^d = 1, \forall x_k \in X;$$

$$p^d(k, k') \geq 0, \sum_{k' \in X} p^d(k, k') = 1, \forall x_k \in X, \forall d \in D_k.$$

4. Задан вектор средних потерь за один шаг $C = (C_1, C_2, \dots, C_N)$, где

$$C_k := \sum_{d \in D_k} c_k^d \alpha_k^d, \quad x_k \in X.$$

Здесь c_k^d — потери за один шаг, если в состоянии x_k принято управление $d \in D_k$, при этом c_k^d являются равномерно ограниченными величинами;

C_k — математическое ожидание потерь, связанных с выходом из состояния x_k за один шаг, $k = 1, \dots, N$.

Задача оптимизации ЦМ формулируется так: необходимо найти такую стратегию управления, чтобы минимизировать средние потери за один шаг:

$$W = \sum_{k \in X} C_k \rightarrow \min. \quad (1.1)$$

Задача (1.1) решается, как правило, при ограничениях, задаваемых траекториями соответствующих марковских процессов и системами уравнений равновесия для определения стационарных вероятностей $\rho(k)$, $x_k \in X$, с учётом управления $d \in D$. Соответствующая ей задача линейного программирования (ЛП) имеет вид

$$W := \sum_{k,d} c_k^d x_k^d \rightarrow \min; \quad (1.2)$$

$$\sum_d x_k^d - \sum_{k,d} p^d(k, k') x_k^d = 0, \quad k' = 1, 2, \dots, N; \quad (1.3)$$

$$\sum_{k,d} x_k^d = 1, \quad x_k^d \geq 0, \quad k = 1, 2, \dots, N, \quad d = 1, 2, \dots, M; \quad (1.4)$$

где M — мощность множества D .

Любое оптимальное решение задачи (1.2)–(1.4) имеет следующее свойство: для каждого состояния x_k существует единственное $d = d(k)$, для которого $x_k^{d(k)} > 0$ и все $x_k^d = 0$ при $d \neq d(k)$. Это означает, что оптимальная стратегия является нерандомизированной и не зависит от начального распределения цепи.

Задачи МППР решаются обычно методами динамического (ДП) или линейного программирования (ЛП), при этом в вычислительном отношении оба подхода почти равнозначны и приводят к аналогичным результатам. Вместе с тем, для решения задач оптимизации реальных систем телеграфика целесообразно использовать методы ЛП. Это объясняется двумя обстоятельствами. Во-первых, нагрузки систем телеграфика определяются на практике с некоторой погрешностью, и поэтому для исследователей и проектировщиков этих систем большой интерес представляет вопрос о том, в каких диапазонах изменения значений нагрузок трафиков оптимальная стратегия сохраняет свой вид. Как известно, современные пакеты прикладных программ ЛП позволяют произвести достаточно глубокий последоптимизационный анализ, в частности, ответить и на указанный выше вопрос. Во-вторых, при использовании ЛП имеется возможность учитывать и некоторые дополнительные нелинейные ограничения (хотя при этом не гарантируется нерандомизированность найденной стратегии). В тоже время, применение ДП не позволяет разрешить эти проблемы.

1.2. Иерархический алгоритм типа фазового укрупнения для исследования МППР

Описанный выше подход к исследованию задач МППР назовём точным. Этот подход оказывается эффективным при исследовании задач МППР, для которых ФПС исходной управляемой ЦМ содержит небольшое число микросостояний (МС).

Очевидно, что ФПС моделей систем телеграфика при большом числе трафиков и при больших значениях структурных параметров системы (число каналов и мест для ожидания) содержит огромное количество МС. Поэтому для больших систем телеграфика применение точного подхода к решению задач их оптимизации налагивается на существенные вычислительные трудности, слабо преодолеваемые даже с использованием современных компьютеров. Отсюда возникает проблема упрощения описания таких систем.

Здесь предложен новый иерархический алгоритм (приближенный) типа фазового укрупнения, который одновременно использует принцип декомпозиции и не накладывает никаких ограничений на структуру переходных матриц соответствующих управляемых цепей Маркова.

Разработанный здесь приближенный иерархический алгоритм позволяет решать задачи оптимизации управляемых ЦМ практически любой размерности. Поскольку на каждой ступени иерархии алгоритм состоит из идентичных шагов, то с целью упрощения обозначений будем описывать работу алгоритма лишь на первой ступени.

Пусть конечномерный МППР с бесконечным горизонтом планирования и без дисконтирования определён с помощью объектов 1–4, как это было указано в § 1.1.

Шаг 1. Рассматривается некоторое расщепление ФПС X , схематическое изображение которого дано на рис.1.1:

$$X = \bigcup_{v=1}^V X_v, \quad X_v \cap X_{v'} = \emptyset, \quad v \neq v'. \quad (1.5)$$

Шаг 2. Все микросостояния, входящие в подмножество X_v , объединяются в одно укрупнённое состояние (УС), обозначаемое $\omega_v, v = 1, 2, \dots, V$. Все полученные таким путём УС составляют некоторое пространство $\Omega := \{\omega_1, \omega_2, \dots, \omega_V\}$.

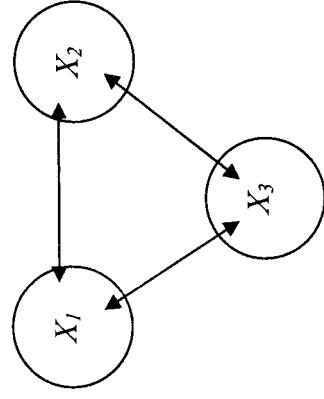


Рис. 1.1. Схема разбиения ФПС X при $V = 3$

Шаг 3. На основе расщепления (1.5) строятся функции укрупнения $U_v: X \rightarrow \hat{X}_v$, где $\hat{X}_v := X \cap \Omega \setminus \{\omega_v\}, v = 1, 2, \dots, V$, которые определяются следующим образом:

$$U_v = \begin{cases} x, & \text{если } x \in X_v \\ \omega_{v'}, & \text{если } x \in X_{v'}, v' \neq v \end{cases} \quad (1.6)$$

Функции укрупнения (1.6) определяют V укрупнённых (относительно исходной) моделей, при этом v -я укрупнённая модель (УМ) имеет ФПС $\hat{X}_v, v = 1, 2, \dots, V$.

Различные варианты построения укрупнённых моделей изображены на рис.1.2.

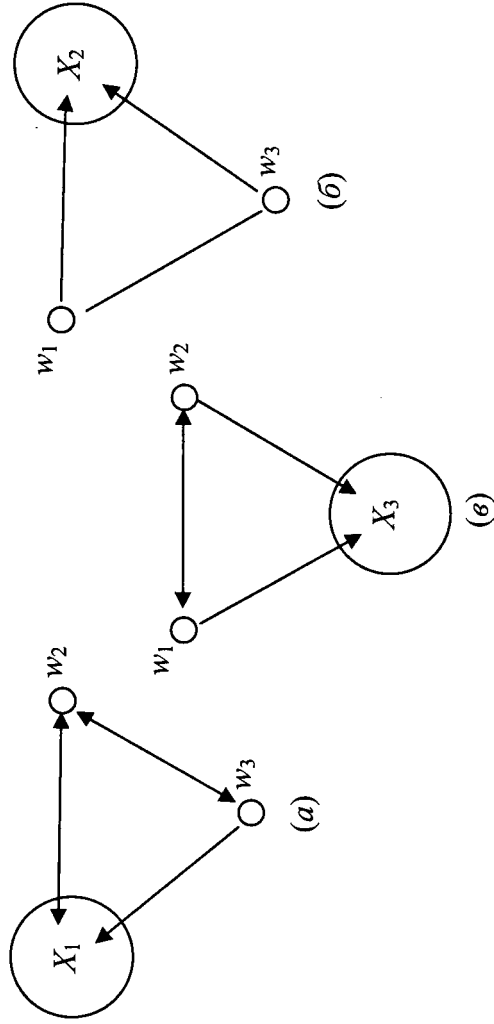


Рис. 1.2. Структура ФПС \hat{X}_v при $V=3$; (а) $v=1$, (б) $v=2$, (в) $v=3$

Шаг 4. Определяется множество допустимых решений D^v для v -й УМ. Оно определяется как проекция множества D на множество \hat{X}_v .

$$D_v := \bigcup_{x \in \hat{X}_v} D_x$$

Шаг 5. Определяются элементы переходной матрицы v -й укрупнённой модели $P_v = \|P_v^d(x, x')\|$, $x, x' \in \hat{X}_v$:

$$P_v^d(x, x') = \begin{cases} p^d(k, k') \rho(k), & \text{если } x = x_k, x' = x'_k, \\ \sum_{k \in \hat{X}_v} p^d(k, k') \rho(k), & \text{если } x = x_k, x' = \omega_k, \\ \sum_{k \in \hat{X}_v, d \in D_k} p^d(k, k') \rho(k), & \text{если } x = \omega_v, x' = x'_k, \\ \sum_{k \in \hat{X}_v, d \in D_k} p^d(k, k') \rho(k), & \text{если } x = \omega_v, x' = \omega_v, \end{cases} \quad (1.8)$$

Поскольку стационарное распределение исходной модели неизвестно, то при исследовании v -й укрупнённой модели невозможно воспользоваться значениями $P_v^d(x, x')$, $d \in D^v$, определяемыми с помощью формулы (1.8). Следовательно, необходимо, используя различные схемы ап-

проксимации, оценить значения неизвестных элементов переходной матрицы сверху или снизу.

Замечание 1.1. С практической точки зрения такую оценку целесообразно производить сверху, так как при этом конечные результаты будут более надёжными. В частности, для такой оценки может быть использован такой факт:

$$\text{если } a_i \geq 0, \quad i=1, 2, \dots, S, \text{ и } 0 < \sum_{i=1}^S b_i \leq 1, \text{ то } \sum_{i=1}^S a_i b_i \leq \max \{ a_i \}.$$

Шаг 6. Элементы вектора средних потерь (для v -й укрупнённой модели ${}^v C_x^d$ — потери за один шаг, если в состоянии $x \in \hat{X}_v$, принято управление $d \in D^v$ в соответствии с формулой (1.8)) определяются через неизвестное стационарное распределение исходной модели. Поэтому здесь также возникает необходимость аппроксимации значения этих величин, которые для случаев оценки сверху и снизу определяются соответственно формулами (1.9) и (1.10).

$${}^v \bar{C}_x^d = \begin{cases} C_k^d, & \text{если } x = x_k \in \hat{X}_v, \\ \max_{d \in D_k, k \in \hat{X}_v} \max(C_k^d), & \text{если } x = \omega_v \in \hat{X}_v, v' \neq v \end{cases} \quad (1.9)$$

$${}^v \underline{C}_x^d = \begin{cases} C_k^d, & \text{если } x = x_k \in \hat{X}_v, \\ \min_{d \in D_k, k \in \hat{X}_v} \min(C_k^d), & \text{если } x = \omega_v \in \hat{X}_v, v' \neq v \end{cases} \quad (1.10)$$

Шаг 7. Для аппроксимации значений критерия качества (1.1) используется оценка сверху и снизу, определяемые соответственно формулами (1.11) и (1.12).

$${}^v \bar{W} := \sum_{x \in \hat{X}_v} {}^v \bar{C}(x), \quad (1.11)$$

$${}^v \underline{W} := \sum_{x \in \hat{X}_v} {}^v \underline{C}(x), \quad (1.12)$$

$$\text{где } {}^v \bar{C}(x) := \sum_{x \in D^v} {}^v \bar{C}_x^d \alpha_x^d; \quad {}^v \underline{C}(x) := \sum_{x \in D^v} {}^v \underline{C}_x^d \alpha_x^d.$$

Шаг 8. Если во всех УМ все параметры (элементы переходной матрицы и критерии качества) оцениваются сверху, то получим V “мажорантных” задач относительно исходной задачи оптимизации. В противном случае, т.е., если во всех УМ указанные параметры оцениваются снизу, то получится V “минорантных” задач относительно исходной за-

дачи оптимизации. В результате решения задачи оптимизации v -й УМ ("мажорантной" или "минорантной") найдутся приближенные оптимальные значения α_k^d , $x_k \in X$, и, таким образом, после параллельного решения всех V аналогичных задач найдутся приближённо оптимальные значения всех α_k^d , где $x_k \in X$.

Точность предложенного метода оценивается так:

$$\max_v \underline{W}^* \leq W^* \leq \min_v \overline{W}^* \quad (1.13)$$

где \overline{W}^* , \underline{W}^* , \overline{W}^* — соответственно оптимальные значения критерия качества в исходной, v -й "мажорантной" и v -й "минорантной" задаче, $v = 1, 2, \dots, V$.

Важное достоинство разработанного приближенного алгоритма состоит в том, что в отличие от известных алгоритмов он не накладывает никаких ограничений на структуру переходной матрицы исходной модели, а также предусматривает возможность многократного применения для построения УМ, и, таким образом, получения иерархии задач оптимизации управляемых ЦМ. Последнее означает, что если для сверхбольшой ЦМ в результате однократного применения разработанного алгоритма не удаётся в достаточной степени снизить размерность задачи оптимизации исходной модели, то следует повторно применить данный алгоритм теперь уже к укрупнённым моделям, и тогда получается иерархия укрупнённых моделей. При этом на каждой ступени иерархии действия по описанному выше алгоритму повторяются.

На погрешность предложенного приближения влияют два фактора: схема расщепления ФПС исходной модели и схема аппроксимации неизвестных параметров в укрупнённых моделях.

Ясно, что укрупнённые модели определяются выбором конкретной схемы расщепления ФПС исходной модели, и, следовательно, можно использовать различные схемы расщепления. Относительно оценки неизвестных параметров в укрупнённых моделях следует отметить, что если использовать более точные схемы аппроксимации, то получается более близкие к исходному решению. При этом оценка (1.13) остаётся верной во всех случаях. Эти факты подтверждают результаты численных экспериментов, проведённых с применением иерархического варианта данного алгоритма для оптимизации марковских систем обслуживания с приоритетами [6].

Следует отметить, что вопрос о том, совпадут ли оптимальные значения марковских управлений при точном и приближенном подходах остаётся открытым.

Заключение и комментарии к главе 1 части III

Монографическое изложение основ классической теории МППР можно найти в книгах [1–3]. Принципы фазового укрупнения были использованы в книге [4] (глава 3). Полученные там результаты базируются на следующей гипотезе: считается, что ФПС исходной цепи распадается на такие подмножества, что при любых управлениях вероятности переходов внутри подмножеств существенно превосходят вероятности переходов между МС, входящими в различные подмножества. На основе использования этой гипотезы разработан приближенный иерархический алгоритм оптимизации управлений марковской цепи с доходами. Он позволяет существенно уменьшить размерность исходной оптимизационной задачи.

Декомпозиционный подход также использован в работе [5] для задач МППР, в которых производящая матрица исходной управляемой ЦМ имеет специальную структуру.

Предложенный в § 1.2 подход для приближенного анализа МППР в различных модификациях был ранее описан в работах [6, 7]. Он в идейном плане близок к работе [8]. Обзор работ по данной тематике можно найти в [9–11].

Литература к главе 1 части III

1. **Howard R.A.** Dynamic programming and Markov processes. — N.Y.: MIT Press, 1960.
2. **Derman C.** Finite state Markovian decision processes. — N.Y.: Academic Press, 1970.
3. **Kallenberg L.C.M.** Linear programming and finite Markovian control problems. — Math. Centre Tracts, Amsterdam, vol. 148, 1983.
4. **Первозванский А.А., Гайцори В.Г.** Декомпозиция, агрегирование и приближенная оптимизация. — М.: Наука, 1979.
5. **Ross K.W., Varadarajan R.** Multichain Markov decision processes with a simple path constraint: A decomposition approach // Math. of Oper. Res. — 1991. — vol.16, no.1. — pp.195–207.
6. **Меликов А.З., Пономаренко Л.А., Рюшин Н.А.** Математические модели многопоточковых систем обслуживания. — Киев: Техника, 1991. — 244 с.
7. **Меликов А.З., Пономаренко Л.А.** Иерархический алгоритм типа фазового укрупнения для оптимизации управляемых цепей Маркова // Автоматизация производственных процессов (Киев). — 1996. — №1. — С. 86 — 90.
8. **Hahnwald-Busch A.** Verfahren zur aggregation in zustandstraum bei Markovschen Entscheidungsproblemen // Wiss. Berlin Tech. Hochsch. Leipzig. — 1986. — № 7. — pp.10 — 12.

9. **Stidham S., Weber R.A.** Survey of Markov decision models for control of networks of queues // Queueing Systems. — 1993. — vol.13. — pp.291–314.
10. **White D.J.** Real applications of Markov decision processes // Interfaces. — 1985. — vol.15, no.6. — pp. 73–83.
11. **Hordijk A., Kallenberg L.G.** Linear programming methods for solving finite markovian decision problems // Oper.Res.:Proc. — 1980. — DGOR. — Pap.Ann.Meet., Raumen, 1980. — Berlin etc., 1981. — pp.468 — 482.

Глава 2

ПРИМЕНЕНИЕ МПНР В ЗАДАЧАХ ОПТИМИЗАЦИИ МНОГОСКОРОСТНЫХ СИСТЕМ ОБСЛУЖИВАНИЯ

2.1. Нахождение оптимальной стратегии доступа в многоскоростной системе обслуживания

В предыдущей части книги подробно исследованы модели многоскоростных систем обслуживания (MRQ) при различных стратегиях доступа. Поэтому здесь не будем останавливаться на подробном описании их работы. Отметим, что все обозначения предыдущей части сохраняются без изменения.

Как было отмечено в предыдущей главе, выбор критерия качества системы зависит от её конкретного назначения. Для определённости изложения здесь предполагается, что качество функционирования общей модели многоскоростной системы оценивается средним числом занятых каналов системы. Задача состоит в следующем: требуется найти такую стратегию доступа, при которой было бы возможным максимизировать загрузку каналов указанной системы. Иными словами, целью исследования является решение задачи достижения

$$\max \tilde{N}_{av}, \quad (2.1)$$

где $\tilde{N}_{av} := \sum_{n \in S} (n \bullet b) p(n)$ — среднее число занятых каналов, S — фазовое

пространство состояний модели, определённое в главе 1 части II.

Стратегия доступа, являющаяся решением задачи (2.1), называется оптимальной. Сразу следует отметить, что здесь оптимальная САС ищется в классе стратегий, не допускающих прерывания уже начатого процесса обслуживания заявок любых типов.

Предполагается, что заявка любого типа, поступившая в момент, когда в системе отсутствует необходимое количество свободных каналов, с вероятностью 1 получает отказ. Следовательно, оптимальная САС должна определить судьбу поступающих заявок в тех моментах, когда имеется необходимое количество свободных каналов, ведь именно в этих моментах возможно принятие двух альтернативных решений:

- 1) поступившая заявка принимается на обслуживание;
- 2) ей отказывают в обслуживании.

Последнее решение связано с необходимостью резервирования некоторого количества каналов для их использования при обслуживании широкополосных заявок.

С целью определения управляющих решений рассмотрим момент поступления заявки i -го типа (i -заявка), $i = 1, 2, \dots, K$, где K — общее число типов разнотипных заявок.

Пусть в этот момент система находится в состоянии $n \in S$, в котором $f(n) \geq b_i$, где $f(n) := N - (n \bullet b)$ означает число свободных каналов в состоянии $n \in S$ (в противном случае, т.е. когда $f(n) < b_i$, как было указано выше, i -заявка получает отказ). Тогда возможно принятие одного из двух решений: либо (а) i -заявка принимается на обслуживание, либо (б) она получает отказ.

Вероятности принятия указанных решений обозначаются соответственно через $\alpha_i^+(n)$ и $\alpha_i^-(n)$. Определенные таким образом вероятности называются управляющими решениями. Они удовлетворяют следующим условиям:

$$0 \leq \alpha_i^+(n) \leq 1 \quad (2.2)$$

$$\alpha_i^+(n) + \alpha_i^-(n) = 1, \quad \forall I \in F(n), \quad (2.3)$$

где $F(n) = \{i \in Z_k^+ : f(n) \geq b_i\}$, $Z_k^+ := \{1, \dots, K\}$.

Замечание 2.1. При определенных значениях введенных управляющих решений получаем известные стратегии. Так, если $\alpha_i^+(n) = 1$ для любого $I \in F(n)$, то получается CS-стратегия доступа; если $\alpha_i^+(n) = 1$ для любого $i \in Z_k^+$ при выполнении условия $f(n) \geq b_i$, где $b := \max\{b_i\}$, то получается CSE-стратегия доступа; если $\alpha_i^+(n) = 1$ для любого $i \in Z_k^+$ при выполнении условия $f(n) \geq b_i + r_i$, $0 \leq r_i \leq N - b_i$, то получается TR-стратегия доступа. Другие возможные SAC также могут быть получены из описанной выше схемы доступа при определенных значениях управляющих решений. Следовательно, описанная схема определяет достаточно общий класс стратегий доступа, в которых не допускается вытеснение заявок из канала.

При использовании введенных управляющих решений элементы производящей матрицы $q(n, n')$, $n, n' \in S$, определяются так:

$$q(n, n') = \begin{cases} \lambda_i \alpha_i^+(n), & \text{если } n' = n + e_i \\ n_i \mu_i, & \text{если } n' = n - e_i \\ 0 & \text{в остальных случаях} \end{cases} \quad (2.4)$$

Следовательно, в явном виде задача (2.1) записывается следующим образом:

$$\sum_{n \in S} (n \bullet b) p(n) \rightarrow \max \quad (2.5)$$

при ограничениях

$$\sum_{i=1}^K (\lambda_i \alpha_i^+(n) I(f(n) \geq b_i) + \sum_{i=1}^K n_i \mu_i) p(n) = \sum_{i=1}^K \lambda_i p(n - e_i) \alpha_i^+(n - e_i) \chi(n_i) +$$

$$+ \sum_{i=1}^K (n_i + 1) \mu_i p(n + e_i) I(f(n) \geq b_i), \quad n \in S; \quad (2.6)$$

$$\sum_{n \in S} p(n) = 1. \quad (2.7)$$

$$\alpha_i^+(n) + \alpha_i^-(n) = 1, \quad \forall i \in F(n), \quad n \in S. \quad (2.8)$$

В задаче (2.5)–(2.8) ограничения (2.6), (2.7) задают СУР данной модели при использовании управляющих решений, они составляются с учетом (2.4). Эта задача является задачей МППР, и, как было отмечено в главе 1 данной части книги, имеет оптимальное решение, согласно которому значения $\alpha_i^+(n)$ равны либо 0, либо 1 для каждого $i \in F(n)$. Последнее обстоятельство позволяет построить в исследуемой системе простой алгоритм, реализующий найденную оптимальную нерандомизированную SAC.

Описанный выше точный подход (ТП) к решению задачи нахождения оптимальной SAC эффективен при небольших значениях N и K , но с их ростом, как было отмечено во второй части книги (глава 1), размерность ФПС S растёт экспоненциально. В связи с этим использование точного подхода становится практически невозможным. Ниже для решения этой задачи для больших MRQ предлагается использовать приближенный подход (ПП), предложенный в § 1.2 предыдущей главы.

Рассмотрим следующее расщепление ФПС S :

$$S := \bigcup_{r=0}^N S_r, \quad S_r \cap S_{r'} = \emptyset, \quad r = r', \quad (2.9)$$

где $S_r := \{n \in S : n \bullet b = r\}$, т.е. класс $MC S_r$ объединяет те состояния $n \in S$, в которых общее число занятых каналов равно r . В дальнейшем каждый класс S_r описывается одним укрупненным состоянием, обозначаемым членом $\langle r \rangle$, $r = 0, \dots, N$.

На основе расщепления (2.9) строятся следующие функции укрупнения:

$$U_r S \rightarrow \hat{S}_r, \quad (2.10)$$

где $\hat{S}_r := S_r \cup \{Z_N^0 \setminus \langle r \rangle\}$, $Z_N^0 = \{0, 1, \dots, N\}$.

Функции укрупнения (2.10) определяются так:

$$U_r(n) = \begin{cases} n, & \text{если } n \in S_r \\ \langle r' \rangle, & \text{если } n \in S_{r'}, r \neq r'. \end{cases}$$

Следовательно, функции укрупнения (2.10) определяют $N+1$ укрупнённых (относительно исходной) моделей, при этом r -я укрупнённая модель (УМ) имеет ФПС \hat{S}_r . Поскольку все УМ имеют аналогичную структуру, то в дальнейшем зафиксируем значение $r \in Z_N^0$ и рассмотрим r -ю модель.

Перейдём к определению элементов ПМ r -ой УМ, обозначаемых $q_r(x, y)$; $x, y \in \hat{S}_r$. Согласно (2.4) имеем:

$$q_r(n, n') = 0, \forall n, n' \in \hat{S}_r, n \neq n'; \quad (2.11)$$

$$q_r(n, \langle r' \rangle) = \begin{cases} \sum_{j=1}^k \lambda_j \alpha_j^+(n) p(n) \delta(b_j, b_i), & \text{если } r' = r + b_i, i = \overline{1, K} \\ \sum_{j=1}^k n_j \mu_j p(n) \delta(b_j, b_i), & \text{если } r' = r - b_i, b_i \leq r, i = \overline{1, K} \\ 0 & \text{в остальных случаях;} \end{cases} \quad (2.12)$$

$$q_r(\langle r' \rangle, n) = \begin{cases} \sum_{j=1}^k \lambda_j \alpha_j^+(n - e_j) p(n - e_j) \chi(n_j) \delta(b_j, b_i), & \text{если } r' = r - b_i, i = \overline{1, K} \\ \sum_{j=1}^k (n_j + 1) \mu_j p(n + e_j) \delta(b_j, b_i), & \text{если } r' = r + b_i, i = \overline{1, K}, \\ 0, & \text{в остальных случаях;} \end{cases} \quad (2.13)$$

$$q_r(\langle r' \rangle, \langle r'' \rangle) = \begin{cases} \sum_{j=1}^k \sum_{n \in S_r} \lambda_j \alpha_j^+(n) p(n) \chi(n_j) I(N - r' \geq b_j) \delta(b_j, b_i), & \text{если } r' = r' + b_i \\ \sum_{j=1}^k \sum_{n \in S_r} n_j \mu_j p(n) \delta(b_j, b_i), & \text{если } r' = r' - b_i, b_i \leq r, i = \overline{1, K} \\ 0, & \text{в остальных случаях} \end{cases} \quad (2.14)$$

где $n' = (n'_1, n'_2, \dots, n'_k)$.

Справедливость соотношений (2.11)–(2.14) может быть доказана

следующим образом. Выберем некоторое состояние $n \in \hat{S}_r$. Поскольку изменения состояний возможны лишь при поступлении заявок и в моменты окончания их обслуживания, то рассмотрим эти моменты в отдельности. После окончания обслуживания i -заявки в состоянии $n \in \hat{S}_r$

происходит переход $n \rightarrow n - e_i$, где $n - e_i \in S_{r - b_i}$ (т.е. $n - e_i \in S_r$). Если в мо-

мент поступления некоторой i -заявки система находится в состоянии $n \in \hat{S}_r$, в котором $f(n) < b_i$ (т.е. $i \in \overline{F(n)}$), то поступившая заявка с вероятностью 1 теряется; в противном случае (т.е. когда $f(n) \geq b_i$) возможны два решения: (1) если i -заявка принимается на обслуживание (с вероятностью $\alpha_i^+(n)$), то происходит переход $n \rightarrow n + e_i$, где $n + e_i \in S_{r + b_i}$ (т.е.

$n + e_i \in S_r$); (2) если i -заявка получает отказ (с вероятностью $\alpha_i^-(n)$), то происходит виртуальный переход $n \rightarrow n$.

Следовательно, в любом состоянии $n \in \hat{S}_r, r = 0, 1, \dots, N$, после окончания обслуживания i -заявки система переходит в состояние $n - e_i \in S_{r - b_i}$, а в моменты поступления i -заявки при принятии различных решений в системе происходит либо виртуальный переход (т.е. она остаётся в этом же состоянии), либо система переходит в состояние $n + e_i \in S_{r + b_i}$, иными словами, равенство (2.11) верно.

Определим $q_r(n, \langle r' \rangle), n, \langle r' \rangle \in \hat{S}_r$. Из МС $n \in \hat{S}_r$ можно попасть в УС $\langle r' \rangle$ только тогда, когда $r' = r + b_i$, если $r \leq N - b_i$, или $r' = r - b_i$, если $n_i > 0$. Переход $n \rightarrow r + b_i$ может произойти в моменты поступления в систему i -заявки в состоянии n , где $f(n) \geq b_i$ (т.е. $i \in F(n)$), так как в этом случае поступившая i -заявка с вероятностью $\alpha_i^+(n)$ принимается на обслуживание и, следовательно, происходит переход $n \rightarrow n + e_i$, где

$n + e_i \in S_{r + b_i}$ (т.е. происходит переход в УС $\langle r + b_i \rangle$). После окончания

обслуживания i -заявки в состоянии $n \in \hat{S}_r$ происходит переход $n \rightarrow n - e_i$, где

$n - e_i \in S_{r - b_i}$ (т.е. происходит переход в УС $\langle r - b_i \rangle$). Таким образом, для

вычисления $q_r(n, \langle r' \rangle), n, \langle r' \rangle \in \hat{S}_r$, с учётом формулы полной вероятности, получим равенство (2.12).

Аналогично рассуждая, находим, что переход $\langle r' \rangle \rightarrow n$, $\langle r' \rangle, n \in \hat{S}_r$, возможен только в случаях $r' = r - b_i$ или $r' = r + b_i$, а переход, $\langle r' \rangle \rightarrow \langle r'' \rangle$, $\langle r' \rangle, \langle r'' \rangle \in \hat{S}_r$, возможен только в случаях $r'' = r + b_i$ или $r' = r - b_i$. Интенсивности переходов между этими состояниями определяются соответственно из соотношений (2.13) и (2.14). Таким образом, справедливость соотношений (2.11)–(2.14) доказана.

Для построения СУР данной упрощенной модели следует аппроксимировать точные значения $q(x, y)$, $x, y \in \hat{S}_r$, в выражениях (2.12)–(2.14). Необходимость в этом объясняется тем, что в эти выражения входит стационарное распределение исходной модели, а также управляющие решения, которые не определены для данной УМ.

С этой целью будем использовать *Замечание 1.1* из предыдущей главы. Тогда величины $q_r(x, y)$, $x, y \in \hat{S}_r$ оцениваются сверху следующим образом:

$$\sum_{i=1}^k \lambda_i \alpha_i^+(n) p(n) \delta(r + b_i, r') \leq \sum_{i=1}^k \lambda_i \alpha_i^+(n) \delta(r + b_i, r'), \quad (2.15)$$

$$\sum_{i=1}^k n_i \mu_i p(n) \delta(r - b_i, r') \leq \max_i \{n_i \mu_i : r - b_i = r'\}, \quad (2.16)$$

$$\sum_{i=1}^k \sum_{n \in \hat{S}_r} \lambda_i \alpha_i^+(n - e_i) p(n - e_i) \chi(n_i) \leq \max_i \{\lambda_i\}, \quad (2.17)$$

$$\sum_{i=1}^k \sum_{n \in \hat{S}_r} (n_i + 1) \mu_i p(n + e_i) \delta(r + b_i, r') \leq \max_i \{(n_i + 1) \mu_i : r + b_i = r'\}, \quad (2.18)$$

$$\sum_{i=1}^k \sum_{n \in \hat{S}_r} \lambda_i \alpha_i^+(n) p(n) I(N - r' \geq b_i) \leq \max_i \{\lambda_i : b_i \leq N - r'\}, \quad (2.19)$$

$$\sum_{i=1}^k \sum_{n \in \hat{S}_r} n_i \mu_i p(n) \leq \max_i \{n_i \mu_i : n_i b_i = r', r' = r - b_i\}. \quad (2.20)$$

Использование неравенства (2.15) для оценки $q_r(n, r')$ в случае $r' = r + b_i$ в таком виде объясняется тем, что в левой части этого неравенства участвуют неизвестные управляющие решения $\alpha_i^+(n)$, которые подлежат определению именно в результате решения задачи оптимизации r -ой укрупнённой модели. Следует также отметить, что в правых частях неравенств (2.15)–(2.20) могут быть использованы более грубые (и в то же время более простые) оценки.

Таким образом, в качестве приближенных значений $q_r(x, y)$, $x, y \in \hat{S}_r$ в выражениях (2.15)–(2.20) используются следующие величины:

$$\left. \begin{aligned} & \approx \sum_{i=1}^k \lambda_i \alpha_i^+(n) \delta(r + b_i, r'), & \text{если} & & r' = r + b_i \\ & \approx \max_i \{n_i \mu_i : r - b_i = r'\}, & \text{если} & & r' = r - b_i \\ & = 0, & & & \text{в остальных случаях;} \end{aligned} \right\} q_r(n, r') \quad (2.21)$$

$$\left. \begin{aligned} & \approx \max_i \{\lambda_i : n_i > 0\}, & \text{если} & & r' = r - b_i \\ & \approx \max_i \{(n_i + 1) \mu_i : r + b_i = r'\}, & \text{если} & & r' = r + b_i \\ & = 0, & & & \text{в остальных случаях;} \end{aligned} \right\} q_r(r', n) \quad (2.22)$$

$$\left. \begin{aligned} & \approx \max_i \{\lambda_i : b_i \leq N - r'\}, & \text{если} & & r' = r' + b_i \\ & \approx \max_i \{n_i \mu_i : n_i \bullet b_i = r', r' = r' - b_i\}, & \text{если} & & r' = r' - b_i \\ & = 0, & & & \text{в остальных случаях.} \end{aligned} \right\} q_r(r', r'') \quad (2.23)$$

Система уравнений равновесия для стационарных вероятностей состояний $\pi(x)$, $x \in \hat{S}_r$, r -ой УМ составляется на основе выражений (2.11), (2.21)–(2.23). Ввиду очевидности её составления явный вид этой СУР здесь не приводится.

Значение функционала в \tilde{N}_{av} в r -й УМ аппроксимируется сверху следующим образом:

$$\tilde{N}_{av}(\hat{S}_r) = \sum_{n \in \hat{S}_r} (n \bullet b) \pi(n) + \sum_{r' \in \hat{S}_r} r' \pi(r'). \quad (2.24)$$

Следовательно, задача оптимизации r -й УМ состоит в максимизации функционала (2.24). При этом ограничениями данной задачи являются СУР для $\pi(x)$, $x \in \hat{S}_r$, и ограничения (2.8), записанные лишь для МС $n \in S_r$.

Данная задача, так же как и задача оптимизации исходной модели при точном подходе, относится к классу МППР. Следовательно, в ре-

ФПС модели при точном подходе содержит 139 микросостояний. Задача максимизация загрузки каналов данной MRQ решалась при следующих значениях нагрузочных параметров:

- 1) $(\lambda_1, \dots, \lambda_{10}) = (0.25, \dots, 0.25)$;
- 2) $(\lambda_1, \dots, \lambda_{10}) = \underbrace{(0.25, \dots, 0.25, 0.5, \dots, 0.5)}_5$;
- 3) $(\lambda_1, \dots, \lambda_{10}) = \underbrace{(0.25, \dots, 0.25, 1, \dots, 1)}_5$.

Во всех вариантах принимается, что $(\mu_1, \dots, \mu_{10}) = (1, \dots, 1)$. Результаты численных экспериментов приведены в табл.2.1. В графе CS указаны значения загрузки каналов при CS-стратегии доступа (т.е. в случае $\alpha^+(n) = 1, \forall i \in F(n), n \in S$); в графе ТП указаны оптимальные (максимальные) значения загрузки каналов, полученные с помощью точного подхода; в графе ИП верхние строки (нижние строки) соответствуют оптимальным значениям загрузки каналов, полученные с помощью приближенного подхода при использовании расщепления (2.9) ((2.25)).

Таблица 2.1

Результаты решения задачи оптимизации MRQ при различных подходах

№ варианта	Загрузка каналов		
	CS	ТП	ИП
1	1.176	3.214	3.856 3.452
2	3.029	5.738	6.043 5.988
3	7.829	8.945	9.056 9.112

Основная цель решения задачи оптимизации САС заключается в нахождении оптимальных значений управляющих решений $\alpha_i^+(n)$. В этой связи следует отметить, что для исследуемой модели MRQ оптимальные значения этих параметров совпадают при точном и приближенном подходе. Вместе с тем, как было отмечено в § 1.2 предыдущей главы, в общем случае нельзя утверждать, что оптимальные САС, найденные с помощью различных подходов, будут совпадать.

Многочисленные эксперименты показали, что введение оптимальной САС позволяет иногда существенным образом улучшить утилизацию

зультате её решения находятся приближённо оптимальные значения управляющих решений $\alpha^+(n)$, где $n \in S, i \in F(n)$. При этом, как было отмечено в предыдущей главе (см.(1.13)),

$$\tilde{N}_{av}^*(S) \leq \min_{r \in Z_N^0} \tilde{N}_{av}^*(\hat{S}_r),$$

где $\tilde{N}_{av}^*(S)$ – оптимальное (максимальное) значение критерия в исходной задаче (2.5)–(2.8); $\tilde{N}_{av}^*(\hat{S}_r)$ – оптимальное значение критерия (2.24) в r -й задаче, $r \in Z_N^0$.

Рассмотрим некоторые замечания относительно размерностей задач оптимизации исходной и укрупнённых моделей. Известно, что размерность оптимизационной задачи соответствует размерности ФПС модели. При использовании схемы расщепления (3.22) ФПС исходной модели размерность ФПС r -й УМ равна $|\hat{S}_r| = |S_r| + N$, где запись $|X|$ означает размерность множества X . Отсюда видно, что размерность ФПС \hat{S}_r будет намного меньше, чем размерность ФПС исходной модели S . Так, например, для модели MRQ с параметрами $N = 10, K = 10, b = (1, 2, \dots, 10)$ размерность ФПС S равна 139, а ФПС \hat{S}_9 , имеющее максимальную размерность среди всех УМ, содержит 35 элементов, т.е. применение приближенного подхода позволяет снизить размерность исходной задачи почти в четыре раза.

Здесь же отметим, что если для сверхбольшой MRQ в результате однократного применения процедуры укрупнения не удаётся в достаточной степени снизить размерность задачи оптимизации исходной модели, то следует повторно применить процедуру укрупнения теперь уже к укрупнённым моделям, и тогда образуется иерархия укрупнённых моделей (см. §1.2 предыдущей главы).

Важно также отметить, что при построении УМ схема расщепления ФПС исходной модели (2.9) является не единственно возможной. Так, например, можно также использовать следующее расщепление ФПС исходной модели:

$$S = \bigcup_{v=0}^V S^v, S^v \cap S^{v'} = \emptyset, V \neq V', \quad (2.25)$$

$$\text{где } S^v := \{n \in S: \sum_{i=1}^K n_i = v\}, V := [N/b].$$

Построение УМ при использовании расщепления (2.25) представляется читателю.

Рассмотрим результаты численных экспериментов для модели MRQ с параметрами $N = 10, K = 10, b = (1, 2, \dots, 10)$. Как было указано выше,

каналов, при этом эффект улучшения зависит от нагрузочных параметров разнотипных трафиков.

Численные эксперименты также показали, что в общем случае оптимальная САС не принадлежит к классу стратегий порогового типа, согласно которой узкополосные заявки принимаются на обслуживание лишь тогда, когда число свободных каналов меньше определённого (порогового) значения.

2.2. Нахождение субоптимальной стратегии доступа к многоскоростной системе обслуживания

Предложенный в предыдущем параграфе подход к нахождению оптимальной САС в многоскоростной системе предполагает, что состояние системы наблюдается полностью. Однако на практике, особенно при исследовании моделей МRQ с большим числом типов заявок, состояние системы не наблюдается полностью, т.е. имеется частичная информация о её состоянии, а именно, наблюдается лишь общее число занятых (свободных) каналов.

Поэтому искомая САС должна принимать решение на основе такой частичной информации. Оптимальная (в определённом смысле) САС, основанная лишь на информации о числе занятых (свободных) каналов, называется субоптимальной. Ниже предложен подход к нахождению субоптимальной САС в многоскоростной системе, исследованной в предыдущем параграфе.

Рассматривается расщепление (2.9), но, в отличие от (2.9), здесь функция укрупнения определяется следующим образом:

$$U(\mathbf{n}) = \langle r \rangle, \text{ если } \mathbf{n} \in S_r, r \in Z_N^0$$

С целью определения элементов производящей матрицы, рассмотрим произвольный класс микросостояний $S_r, r \in Z_N^0$, и некоторое микросостояние $\mathbf{n} \in S_r$. Поскольку изменения состояния возможны лишь при поступлении заявок и в моменты окончания их обслуживания, то рассмотрим эти моменты в отдельности.

Если в момент поступления i -заявки система находится в состоянии $\mathbf{n} \in S_r$, в котором $i \in F(\mathbf{n})$ (т.е. $f(\mathbf{n}) < b_i$), то поступившая заявка с вероятностью 1 теряется; в противном случае (т.е. когда $i \in F(\mathbf{n})$) возможны два решения: (1) если i -заявка принимается на обслуживание (с вероятностью $\alpha^+(\mathbf{n})$), то происходит переход $\mathbf{n} \rightarrow \mathbf{n} + \mathbf{e}_i$ где $\mathbf{n} + \mathbf{e}_i \in S_{r+b_i}$ (т.е. $\mathbf{n} + \mathbf{e}_i \in S_r$); (2) если i -заявка получает отказ (с вероятностью $\alpha^-(\mathbf{n})$), то происходит виртуальный переход $\mathbf{n} \rightarrow \mathbf{n}$. Следовательно, в моменты поступления i -заявок при принятии альтернативных решений в системе происходит либо виртуальный переход (т.е. она остаётся в том же состоянии), либо система переходит в состояние $\mathbf{n} + \mathbf{e}_i \in S_{r+b_i}$. После окон-

чания обслуживания i -заявки в состоянии $\mathbf{n} \in S_r$ с интенсивностью $n_i \mu_i$ происходит переход $\mathbf{n} \rightarrow \mathbf{n} - \mathbf{e}_i$, где $\mathbf{n} - \mathbf{e}_i \in S_{r-b_i}$.

Таким образом, в данном случае элементы производящей матрицы укрупнённой модели $q(\langle r \rangle, \langle r'' \rangle), r', r'' \in Z_N^0$, определяются так:

$$q(\langle r \rangle, \langle r'' \rangle) = \begin{cases} \lambda_i \sum_{\mathbf{n} \in S_r} p(\mathbf{n}) \alpha_i^+(\mathbf{n}), & \text{если } r'' = r' + b_i, i \in Z_K^+, \\ n_i \mu_i \sum_{\mathbf{n} \in S_r} p(\mathbf{n}), & \text{если } r'' = r' - b_i, i \in Z_K^+, \\ 0 & \text{в остальных случаях,} \end{cases} \quad (2.26)$$

где $Z_K^+ = \{1, 2, \dots, K\}$.

Стационарная вероятность состояния $\langle r \rangle$, обозначаемая $\pi(\langle r \rangle)$, определяется так:

$$\pi(\langle r \rangle) = \sum_{\mathbf{n} \in S_r} p(\mathbf{n}), r \in Z_N^0. \quad (2.27)$$

С учётом (2.27) находим, что среднее число занятых каналов (\tilde{N}_{av}) системы выражается через стационарное распределение укрупнённой модели следующим образом:

$$\tilde{N}_{av} = \sum_{r=1}^N r \pi(r). \quad (2.28)$$

Для построения СУР для укрупнённой модели следует аппроксимировать точные значения $q(\langle r \rangle, \langle r'' \rangle), r', r'' \in Z_N^0$, в выражениях (2.26). Необходимость в этом объясняется тем, что в эти выражения входит неизвестное стационарное распределение исходной модели, а также управляющие решения $\alpha_i^\pm(\mathbf{n}), \mathbf{n} \in S_i \in F(\mathbf{n})$, которые для укрупнённой модели не определены.

Поскольку для любого микросостояния $\mathbf{n} \in S_r$ при принятии решения о доступе поступившей i -заявки на обслуживание, $i \in F(\mathbf{n})$, происходит переход в укрупнённое состояние S_{r+b_i} , а при отказе ей в обслуживании происходит виртуальный переход, то для укрупнённой модели управляющие решения могут быть определены следующим образом:

$$\alpha_i^+(\langle r \rangle) + \alpha_i^-(\langle r \rangle) = 1 \quad \forall i \in \tilde{F}(r), \quad (2.29)$$

где $\alpha_i^-(\langle r \rangle) = P(\text{поступившая } i\text{-заявка принимается} / \text{система находится в УС } \langle r \rangle)$;

$\alpha_i^+(\langle r \rangle) = P(\text{поступившая } i\text{-заявка теряется} / \text{система находится в УС } \langle r \rangle)$;

$$\bar{F}(r) = \{i \in Z_K^+ : b_i \leq N-r\}.$$

Замечание 2.2. Стационарная вероятность блокировки i -заявок (PV_i) также может быть определена через стационарное распределение укрупнённой модели:

$$PV_i = \sum_{j=0}^{b_i-1} \pi(\langle N-j \rangle) + \sum_{j=b_i}^N \pi(\langle N-j \rangle) \alpha_j^-(\langle N-j \rangle).$$

Тогда с учётом (2.26) и (2.29) находим, что в качестве приближённых значений величин $q(\langle r' \rangle, \langle r'' \rangle)$, $r', r'' \in Z_N^0$, могут быть использованы следующие соотношения, которые оценивают их сверху:

$$\tilde{q}(\langle r' \rangle, \langle r'' \rangle) = \begin{cases} \approx \lambda_i \alpha_i^+(\langle r' \rangle), & \text{если } r'' = r' + b_i, i \in Z_K^+, \\ \approx \left[\frac{r}{b_i} \right] \mu_i, & \text{если } r'' = r' - b_i, i \in Z_K^+, \\ 0 & \text{в остальных случаях,} \end{cases} \quad (2.30)$$

здесь $\left[\frac{r}{b_i} \right]$ означает максимальное число i -заявок в микросостояниях $i \in E_r$.

С учётом (2.30) СУР для укрупнённой модели имеет следующий вид:

$$\sum_{r \in Z_N^0} \pi(\langle r \rangle) = 1; \quad (2.31)$$

$$\left(\sum_{i \in \bar{F}(r)} \lambda_i \alpha_i^+(\langle r \rangle) + \sum_{i=1}^K \left[\frac{r}{b_i} \right] \mu_i \pi(\langle r \rangle) = \sum_{i=1}^K \lambda_i \pi(\langle r - b_i \rangle) \alpha_i^+(\langle r - b_i \rangle) \chi(r - b_i) + \right. \quad (2.32)$$

$$\left. \sum_{i \in \bar{F}(r)} \left[\frac{r}{b_i} + 1 \right] \mu_i \pi(\langle r + b_i \rangle) \right), \quad r \in Z_N^0.$$

Таким образом, задача нахождения субоптимальной САС в исследуемой многократной системе состоит в максимизации (2.28) при ограничениях (2.29), (2.31), (2.32). Она так же, как и задача нахождения оптимальной САС, относится к классу МППР, и в результате её решения

находится субоптимальная нерандомизированная САС т.е. определяются оптимальные значения $\alpha_i^+(\langle r \rangle)$, где $r \in Z_N^0$.

Рассмотрим некоторые замечания относительно размерностей задач нахождения оптимальной и субоптимальной САС. Очевидно, что размерность задачи нахождения субоптимальной САС будет намного меньше, чем размерность аналогичной задачи нахождения оптимальной САС. Так, например, для модели MRQ с параметрами $N = 30, K = 30, b = (1, 2, \dots, 30)$ размерность ФПС исходной модели равна 28628, а размерность ФПС укрупнённой модели равна 31, причём она не зависит от числа типов заявок, т.е. задача нахождения субоптимальной САС имеет размерность почти в 10^3 раз меньше, чем задача нахождения оптимальной САС.

В заключение отметим, что в частных случаях оптимальная и субоптимальная стратегии могут совпадать. Вместе с тем, в общем случае нельзя утверждать, что эти стратегии будут совпадать.

Заключение и комментарии к главе 2 части III

Результаты данной главы базируются на идеях работ [1–6]. Здесь были рассмотрены лишь задачи нахождения оптимальной САС в классе стратегий, не позволяющих вытеснение заявок из каналов. Однако предложенный подход позволяет исследовать и класс САС, в которых допускается вытеснение заявок из каналов. Эти стратегии также представляют определённый научный и практический интерес. Задачи нахождения оптимальной САС в моделях MRQ, в которых каналы системы могут работать лишь в определённых сочетаниях, исследованы в работах [8, 9]. Задача нахождения субоптимальной САС была рассмотрена в [10], где она ищется в узком классе стратегий доступа без вытеснения, а именно, в классе стратегий доступа с резервированием каналов. Задача нахождения оптимальной САС для модели MRQ с двумя типами заявок решена в [11].

В работе [12] также используется подход МППР для нахождения оптимальной САС в одномерных мультимедийных сетевых связях. Он в идеальном плане очень близок к точному подходу, изучаемому в данной главе.

Достаточно полный обзор работ по применению МППР в задачах оптимизации многопоточковых систем обслуживания можно найти в работах [7, 13–15].

Литература к главе 2 части III

1. **Меликов А.З.** Приближенная оптимизация мультиресурсных систем обслуживания // Кибернетика. — 1990. — №6. — С.120 — 122.
2. **Меликов А.З.** Методы анализа и оптимизации многопоточковых систем обслуживания // Автоматика и вычислительная техника. — 1992. — №3. — С.13 — 21.
3. **Меликов А.З.** Метод приближенной оптимизации мультиресурсных систем массового обслуживания с очередями // Электронное моделирование. — 1991. — Т.13, № 6. — С.92 — 98.
4. **Меликов А.З., Пономаренко Л.А.** Нахождение оптимальной стратегии доступа в цифровых сетях интегрального обслуживания // Автоматика и вычислительная техника. — 1992. — № 2. — С. 28 — 36.
5. **Меликов А.З., Пономаренко Л.А.** Оптимизация цифровой сети интегрального обслуживания с конечным числом пользователей и обслуживания // Автоматика и телемеханика. — 1992. — № 6. — С. 86 — 92.
6. **Меликов А.З., Пономаренко Л.А.** Замкнутые мультиресурсные системы массового обслуживания с управляемыми приоритетами // Кибернетика. — 1989. — № 5. — С. 90 — 93.
7. **Меликов А.З.** Методы расчета и оптимизации моделей мультиресурсных систем обслуживания // Кибернетика и системный анализ. — 1996. — № 3. — С. 92 — 112.
8. **Меликов А.З., Молчанов А.А., Пономаренко Л.А.** Мультиресурсные системы массового обслуживания с частично коммутируемыми каналами // Электронное моделирование. — 1992. — Т.14, № 2. — С.87 — 91.
9. **Ross K.M., Tsang D.H.** Optimal circuit access policies in an ISDN environment: A Markov decision approach // IEEE Trans. on Commun. — 1989. — vol 37, № 9. — pp.934 — 939.
10. **Oda T., Watanabe Y.** Optimal trunk reservation for a group with multislot traffic streams // IEEE Trans. on Commun. — 1990. — vol 38, № 7. — pp.1078 — 1084.
11. **Courcoubetis C.A., Reiman M.I.** Optimal control of queuing system with simulations service requirements // IEEE Trans. Autom. Control. — 1987. — vol.32, № 8. — pp.717 — 727.
12. **Choi S., Kwon T., Choi Y., Naghshineh M.** Call admission control for multimedia services in mobile cellular networks: A markov decision approach // Proc. 5th IEEE Comp. Commun. Symposium, France, 2000. — pp. 594 — 599.
13. **Alidrisi M.** Linear programming model for the optimal control of queuing network // Int. J. Syst. Sci. — 1987. — vol. 18, № 6. — pp.1079 — 1089.
14. **Stidham S.** Scheduling, routing and flow control of stochastic networks // Stochastic. Contr. Theory and Appl. Proc. Workshop Southend on Sea, June 9-19, 1986. — № 4. — 1988. — pp.529 — 561.
15. **Stidham S., Weber R.A.** Survey of Markov decision models for control of networks of queues // Queueing Systems. — 1993. — vol.13. — pp.291 — 314.

Наукове видання

Меліков Агасі Зарбали огли

Пономаренко Леонід Анатолійович

Паладюк Володимир Васильович

Телеграфік:

Моделі, методи, оптимізація

Монографія

Російською мовою

Видання підготовлено в авторській редакції

Комп'ютерний набір *Ю.Л.Пономаренко*

Комп'ютерна верстка *Г.О.Онищенко*

Коректура *І.В.Голованова, Л.В.Замно, Н.В.Малик*

Дизайн обкладинки *Н.В.Синеуцька*

Підписано до друку 09.08.2007 р.

Формат 70x100/16. Папір офсетний.

Гарнітура "Times". Друк офсетний.

Ум. друк. арк. 12,32. Обл.-вид. арк. 13,55

Тираж 300 прим. Зам. № 0217099.

Віддруковано у видавництві

«Преса України», 03047, Київ-47, пр. Перемоги, 50.

