

МЕТОДЫ И МОДЕЛИ ТЕОРИИ ТЕЛЕТРАФИКА

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«ОРЛОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ И.С. ТУРГЕНЕВА»

В.Т. Еременко, А.П. Фисун, И.А. Саитов,
А.Е. Миронов, А.Н. Орешин, А.В. Королев,
Р.Б. Трегубов, Д.Ю. Патронов

**МЕТОДЫ И МОДЕЛИ
ТЕОРИИ ТЕЛЕТРАФИКА**

Орёл
ОГУ имени И.С. Тургенева
2019

УДК 621.394.343
ББК 32.88
М54

Рецензенты:

доктор технических наук, профессор заведующий кафедрой
информационных технологий и систем безопасности
федерального государственного бюджетного
образовательного учреждения высшего образования
«Российский государственный гидрометеорологический университет»
В.Г. Бурлов,

доктор технических наук, профессор, заведующий кафедрой
компьютерных технологий и систем
федерального государственного бюджетного
образовательного учреждения высшего образования
«Брянский государственный технический университет»
А.В. Аверченков

Еременко, В.Т.

М54 Методы и модели теории телетрафика: учебное пособие / В.Т. Еременко [и др.]. – Орёл: ОГУ имени И.С. Тургенева, 2019. – 244 с.

ISBN 978-5-9929-0650-9

В учебном пособии рассмотрены базовые понятия и определения теории телетрафика, представлены основные свойства и характеристики случайных потоков вызовов. Отдельное внимание уделено вопросам измерения нагрузки. Описаны характеристики и показатели качества обслуживания.

Предназначено для студентов, обучающихся по направлениям: 10.03.01 «Информационная безопасность» (бакалавры), 11.03.02 и 11.04.02 «Инфокоммуникационные технологии и системы связи» (бакалавры и магистры), а также специалистов по направлению 10.05.01 «Компьютерная безопасность».

УДК 621.394.343
ББК 32.88

ISBN 978-5-9929-0650-9

© Коллектив авторов, 2019
© ОГУ имени И.С. Тургенева, 2019

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	6
1. ИСТОРИЯ ПОЯВЛЕНИЯ И РАЗВИТИЯ ДИСЦИПЛИНЫ	
«ТЕОРИЯ ТЕЛЕТРАФИКА»	9
1.1. Основоположник теории телетрафика А. К. Эрланг	10
1.2. Исторический обзор развития дисциплины	
«Теория телетрафика»	12
Вопросы для самоконтроля	15
2. ОСНОВНЫЕ ПОЛОЖЕНИЯ ТЕОРИИ ВЕРОЯТНОСТЕЙ, ИСПОЛЬЗУЕМЫЕ В РАМКАХ ДИСЦИПЛИНЫ	
«ТЕОРИЯ ТЕЛЕТРАФИКА»	16
2.1. Событие. Вероятность события	16
2.2. Законы распределения случайных величин	19
2.3. Числовые характеристики случайных величин	21
2.4. Использование преобразования Лапласа – Стильтеса для получения числовых характеристик случайных величин	26
Вопросы для самоконтроля	29
3. ТЕОРИЯ ТЕЛЕТРАФИКА. БАЗОВЫЙ ТЕОРЕТИЧЕСКИЙ КУРС	30
3.1. Объект, предмет и задачи теории телетрафика	30
3.2. Базовая модель систем распределения информации	34
Вопросы для самоконтроля	36
4. МОДЕЛИ ПОТОКОВ ВЫЗОВОВ	37
4.1. Определение потока вызовов. Свойства потоков вызовов	37
4.2. Характеристики потоков вызовов	43
4.3. Принципы классификации потоков вызовов	45
4.4. Математические модели потоков вызовов без последействия	
4.4.1. Модель простейшего потока вызовов. Свойства и характеристики	46
4.4.2. Модели нестационарных и неординарных пуассоновских потоков вызовов	55
4.4.2.1. Модель нестационарного пуассоновского потока вызовов	56
4.4.2.2. Модели стационарных неординарных пуассоновских потоков вызовов без последействия	57
4.5. Математические модели потоков вызовов с последействием..	58
4.5.1. Модель примитивного потока вызовов	60

4.5.2. Модель сглаженного пуассоновского потока.....	65
4.5.3. Модели потоков с повторными вызовами	66
4.5.4. Модель потока освобождений.....	68
4.6. Математические модели потоков с ограниченным последствием.....	70
Вопросы для самоконтроля	72
5. НАГРУЗКА И ЕЕ ВИДЫ	76
5.1. Нагрузка (трафик), интенсивность нагрузки, единицы измерения нагрузки	77
5.1.1. Нагрузка и ее виды.....	77
5.1.2. Соотношения между видами нагрузки.....	82
5.2. Колебания нагрузки и их причины.....	83
5.3. Измерение нагрузки... ..	86
5.3.1. Цели, задачи, методы измерения нагрузки	86
5.3.2. Принципы измерения нагрузки	92
5.4. Способы математического задания потоков нагрузки на межстанционных участках	96
5.5. Закономерности формирования потоков нагрузки.....	99
Вопросы для самоконтроля	105
6. КАЧЕСТВО ОБСЛУЖИВАНИЯ В ТЕЛЕФОННЫХ СЕТЯХ СВЯЗИ	108
6.1. Качество обслуживания (QoS), определяемое Рекомендациями Международного союза электросвязи E.800 (МСЭ-Т E.800)	108
6.2. Сущность понятия "дисциплина обслуживания»	113
6.3. Характеристики качества обслуживания голосовых сообщений	116
6.4. Особенности оценки качества обслуживания вызовов на коммутируемых сетях связи	120
6.5. Принципы нормирования потерь на телефонных сетях связи	124
6.5.1. Изменение подходов к определению блокировки трафика в телефонных сетях.....	124
6.5.2. Нормы потерь на местных сетях связи.....	126
Вопросы для самоконтроля	127
7. ЗАДАЧИ АНАЛИЗА И СИНТЕЗА ТЕЛЕКОММУНИКАЦИОННЫХ СИСТЕМ.....	129
7.1. Символическое обозначение исследуемых моделей.....	129
7.2. Основы теории расчета однозвенных систем распределения информации	132
7.2.1. Модель $M_r/M/V/L$	132

7.2.1.1. Постановка задачи по обслуживанию пуассоновского потока вызовов с условным параметром	132
7.2.1.2. Определение вероятностей состояния полностью доступного пучка линий	134
7.2.1.3. Характеристики качества обслуживания вызовов системами $M/M/V/L$ и $M_i/M/V/L$	146
7.2.2. Модель $M/M/V/L/PRA$	166
7.2.2.1. Постановка задачи	166
7.2.2.2. Характеристики качества обслуживания вызовов системой с абсолютным приоритетом	168
7.2.3. Модель системы с ожиданием $M/M/V/W, r/FF/R$	175
7.2.3.1. Постановка задачи по обслуживанию потока вызовов системой с ожиданием $M/M/V/W, r \rightarrow \infty/FF/R$	176
7.2.3.2. Вероятности состояний системы. Второе распределение Эрланга	177
7.2.3.3. Характеристики качества обслуживания вызовов системой $M/M/V/W, r/FF/R$	185
7.2.4. Модель системы с ожиданием $M_i/M/V/W, r/FF/R$	197
7.2.4.1. Постановка задачи по обслуживанию потока вызовов системой с ожиданием $M_i/M/V/W, r/FF/R$	197
7.2.4.2. Характеристики качества обслуживания вызовов системой $M_i/M/V/W, r/FF/R$	198
7.3. Основы теории построения и расчета потерь в многозвенных системах с явными потерями	201
7.3.1. Особенности построения двухзвенных и многозвенных коммутационных схем и влияние их структуры на характеристики качества обслуживания	203
7.3.2. Определение потерь в двухзвенных схемах. Комбинаторный метод расчета (метод Якобеуса)	210
7.3.3. Определение потерь в двухзвенных схемах. Метод эффективной доступности (метод Харкевича)	223
7.3.4. Определение потерь в многозвенных схемах. Метод комбинированной блокировки	230
Вопросы для самоконтроля	235
Заключение	240
Литература	241

ВВЕДЕНИЕ

Любую систему, в которой поток требований встречает ограниченные средства их удовлетворения, можно рассматривать как систему массового обслуживания. Один из классов систем массового обслуживания составляют системы распределения информации (системы коммутации). Моделированием, анализом и синтезом таких систем занимается теория телетрафика.

Основная цель преподавания дисциплины «Теория телетрафика» – подготовка специалистов, владеющих знаниями, навыками и умениями в области построения и расчета вероятностно-временных характеристик, структурных параметров инфокоммуникационных систем и сетей различной структуры и назначения.

Объектом изучения дисциплины «Теория телетрафика» являются системы распределения информации (системы коммутации), *предметом* – количественная сторона процесса обслуживания потока вызовов этими системами.

Как и любая другая математическая теория, теория телетрафика оперирует не с самими системами распределения информации (системами коммутации), а с их математическими моделями. *Математическая модель* системы распределения информации (системы коммутации) включает в себя три основных элемента:

- входящий поток вызовов;
- схему системы распределения информации (системы коммутации);
- дисциплину обслуживания.

Задачи дисциплины «Теория телетрафика»:

- установление зависимостей и значений величин, характеризующих качество обслуживания, при заданных параметрах и свойствах потока вызовов, схеме системы распределения информации (системы коммутации) и дисциплине обслуживания;

- определение структурных параметров системы распределения информации (системы коммутации) при заданных:

1) параметрах и свойствах потока вызовов, дисциплине и качестве обслуживания;

2) параметрах и свойствах потока вызовов, дисциплине и качестве обслуживания, учитывая, что стоимость и объем оборудования должны быть минимальными

3) параметрах и свойствах потока вызовов, дисциплине обслуживания и стоимости или объеме оборудования, учитывая, что качество обслуживания должно быть максимальным;

- разработка математических моделей краткосрочного и долгосрочного прогнозов параметров потоков вызовов и времени их обслуживания, исследование свойств потоков вызовов и законов распределения времени их обслуживания в реальных системах распределения информации (системах коммутации);

- нормирование и оптимальное распределение по участкам сети и этапам соединения показателей качества обслуживания.

В результате изучения дисциплины «Теория телетрафика» обучающийся должен:

а) з н а т ь :

- модели потоков вызовов в моносервисных и мультисервисных сетях связи;

- модели систем распределения информации моносервисных сетей связи;

- модели систем распределения информации мультисервисных сетей связи;

- принципы моделирования процессов обслуживания вызовов на ЭВМ;

- принципы распределения и измерения параметров нагрузки и потерь на сетях связи;

- методы расчета сетей связи с обходными путями.

б) у м е т ь :

- решать задачи анализа и синтеза систем распределения информации (систем коммутации);

- нормировать и распределять по участкам сети показатели качества обслуживания;

- измерять и прогнозировать параметры сетевого трафика.

в) в л а д е т ь :

- научно-методическим инструментарием оценки качества обслуживания существующих и перспективных систем распределения информации (систем коммутации).

Учебная дисциплина «Теория телетрафика» состоит из четырех тем:

- предметной области теории телетрафика;

- модели систем распределения информации моносервисных сетей связи;

- модели систем распределения информации мультисервисных сетей связи;
- методов анализа и оптимизации телекоммуникационных сетей связи.

Учебное пособие «Методы и модели теории телетрафика» включает в себя семь разделов.

Первый раздел посвящён истории появления и этапам развития теории телетрафика.

Во втором разделе рассматриваются основные положения теории вероятностей (случайная величина, законы распределения случайной величины, числовые характеристики случайной величины, преобразование Лапласа – Стилттьеса).

В третьем разделе особое внимание уделено предмету, объекту, задачам теории телетрафика и ее основным определениям (математической модели системы распределения информации, заявке, вызову, потоку вызовов, схеме системы распределения информации и дисциплине обслуживания).

В четвертом разделе приводятся классические модели потоков вызовов, характерные для звеньев моносервисных сетей связи (поток без последствия, поток с простым и ограниченным последствием).

Пятый раздел рассматривает виды нагрузки, единицы её измерения, методы и принципы измерения, закономерности формирования потоков нагрузки.

Шестой раздел знакомит с характеристиками качества обслуживания голосовых сообщений, особенностями оценки качества обслуживания вызовов на коммутируемых сетях связи и принципами нормирования потерь на телефонных сетях связи.

В седьмом разделе представлены классические модели звеньев моносервисных сетей связи, описывающие системы распределения информации: с потерями, ожиданием и приоритетной дисциплиной обслуживания.

Каждый раздел учебного пособия заканчивается вопросами для самоконтроля.

1. ИСТОРИЯ ПОЯВЛЕНИЯ И РАЗВИТИЯ ДИСЦИПЛИНЫ «ТЕОРИЯ ТЕЛЕТРАФИКА»

Базовые результаты теории массового обслуживания были сформулированы только в начале XX в. Основоположником ее прикладной ветви – теории телетрафика – считается датский математик Агнер Краруп Эрланг.

У датчан существует особое выражение для характеристики людей, отличающихся утонченным умом, интеллигентностью и добротой: «Эти люди высоки – до самого потолка». Безусловно, к таким людям принадлежал и А. Эрланг. Будучи человеком долга, в своей жизни он всегда следовал девизу: «Служить, но не властвовать».

Имя Агнера Эрланга известно не только тысячам инженеров на всех континентах. Его хорошо знали многие выдающиеся математики XX в.: А. Н. Колмогоров, А. Я. Хинчин, Т. Фрай, В. Феллер и др. Познакомившись с его идеями, они стали интенсивно разрабатывать открытое А. К. Эрлангом новое научное направление.

Теория телетрафика зародилась в связи с необходимостью проектирования автоматических телефонных станций (АТС), которые должны были обрабатывать поступающие от абонентов телефонной сети вызовы. Стоимость сети связи весьма внушительна и определяется количеством линий, соединяющих между собой АТС, находящиеся в разных районах. Поэтому при проектировании АТС количество этих линий должно выбираться минимально возможным, причем необходимо, чтобы вероятность потери вызова абонента или длительность ожидания им соединения была незначительной.

В работах А. К. Эрланга 1908–1918 гг. был заложен фундамент теории телетрафика, дана четкая математическая формулировка проблемы нахождения числа потерянных вызовов, возникающих в сетях связи из-за того, что по причинам экономического характера число телефонных каналов всегда существенно меньше количества установленных телефонов. Эрланг преследовал практическую цель – дать методику расчета числа каналов, достаточного для обслуживания всех подключенных к сети телефонных аппаратов с малой вероятностью потери вызова.

Именно с работ А. Эрланга началось проникновение в теорию связи широко используемых сегодня статистических методов исследования. Однако проблема, с которой он столкнулся, – очереди – оказалась более острой, нежели проектирование АТС.

Экономические потери, возникающие вследствие очередей, огромны. Определение надежности работы различных устройств радиоэлектроники и связи не менее актуальны. В настоящее время все эти проблемы решаются на основе теории массового обслуживания.

Таким образом, теория массового обслуживания, возникшая в связи с конкретной практической задачей и описанная в исследованиях и работах А.К. Эрланга, приобрела чрезвычайную популярность, найдя применение во многих областях человеческой деятельности.

Неудивительно, что теорию, имеющую столь важное практическое значение, стали развивать многие ученые, в том числе известные математики. В ее создании участвовали американцы Т. Фрай, выпустивший в 1928 г. первую книгу по теории вероятностей, в которой одна из глав посвящена теории телетрафика, В. Феллер, Д. Кендалл и др.

Значительный вклад в развитие данной теории внесли известные советские ученые: член-корреспондент Академии наук СССР А.Я. Хинчин и академик Академии наук БССР Б. В. Гнеденко.

1.1. Основоположник теории телетрафика А.К. Эрланг

Агнер Краруп Эрланг родился 1 января 1878 г. в Логборгене на юге Ютландии в Дании. Его отец Ганс Нилсен Эрланг учительствовал в сельской школе, мать Магдалена Краруп происходила из известной в Дании духовной семьи, среди предков которой был прославленный датский математик Томас Финк.

Агнер, его старший брат Фредерик и две младшие сестры Мэри и Ингеборг учились в школе своего отца. Одним из любимых предметов юноши стала астрономия, а излюбленным занятием – написание поэм на астрономические темы.

Завершив начальное школьное образование, в 14 лет он начал готовиться к сдаче экзамена на звание сельского учителя в Копенгагенском университете. Получив специальное разрешение и успешно сдав экзамены, Агнер в течение двух лет работал ассистентом учителя в школе своего отца, много занимался и изучал французский и латинский языки.

По совету отца, молодой человек решил поступать в Копенгагенский университет. Семья не имела достаточных средств, чтобы

оплатить учебу, поэтому он воспользовался своим правом готовиться к вступительным экзаменам, получая бесплатное питание и жилье в колледже Копенгагенского университета, где изучал математику и естественные науки.

Благополучно окончив университет в 1901 г., Агнер был удостоен магистерской степени учителя математики (в качестве главного предмета), а также астрономии, физики и химии.

После окончания университета в течение семи лет он работал в различных школах и, хотя его призванием была исследовательская работа, оказался превосходным преподавателем. По натуре А.К. Эрланг не был общительным человеком, не любил быть в центре внимания и в компаниях предпочитал оставаться в тени. За это друзья дали ему прозвище «частное лицо». Для него превыше всего была краткость, а не длинные речи.

Во время летних школьных каникул Агнер много путешествовал. Побывал во Франции, Швеции, Германии и Великобритании, где посещал художественные галереи и библиотеки. Будучи учителем, продолжал упорно изучать математику и естественные науки. Он был членом Датского математического общества. На одном из собраний А.К. Эрланг познакомился с Йохансенем, главным инженером Копенгагенской телефонной компании, который в 1908 г. пригласил его на работу. Вначале А. Эрланг работал в этой компании как научный сотрудник и, не имея достаточного количества помощников, сам выполнял все измерения для статистики телефонных потоков. Его часто можно было видеть на улицах Копенгагена в сопровождении рабочего с лестницей в руках, с помощью которой они спускались вниз в телефонные люки. Позже он стал руководителем лаборатории.

Научные статьи Эрланга публиковались регулярно. Его наиболее значимая публикация "Решение некоторых проблем теории вероятностей, важных для проектирования автоматических телефонных станций" вышла в 1917 г. В ней содержались формулы для определения вероятности потери вызова и среднего времени ожидания соединения, которые теперь в теории телефонного трафика являются общеизвестными. Эти формулы широко использовались во всем мире при проектировании автоматических телефонных станций.

Среди ученых интерес к научным работам Эрланга был очень велик. Известен случай, когда один из специалистов лаборатории Белла в США специально выучил датский язык для того, чтобы чи-

тать статьи Эрланга в оригинале. Наиболее серьезные из них переводились на английский, французский и немецкий языки. Написанные очень кратко, статьи были трудны для понимания. Положение усугублялось и тем, что Эрланг часто опускал доказательства своих результатов, имевших большое практическое значение.

Все свое время и энергию А. К. Эрланг отдавал исследованиям. Он никогда не был женат и имел обыкновение работать далеко за полночь. В его обширной библиотеке насчитывалось много книг по математике, астрономии и физике. Он также интересовался историей, философией и поэзией. Друзья считали Агнера весьма эрудированным человеком, знакомым со многими областями науки и культуры. Он щедро делился своими знаниями с окружающими и был очень добрым и благородным человеком. Часто, не афишируя этого, помогал нуждающимся людям.

В Копенгагенской телефонной компании Эрланг проработал почти 20 лет. Заниматься своим здоровьем было некогда, и когда в январе 1929 г. у него началось воспаление брюшной полости, ему пришлось лечь в больницу. К сожалению, медицина оказалась бессильной – через несколько дней Агнер скончался.

Работы Эрланга и после смерти оставались необычайно востребованы. В 1943 г. по предложению известного шведского ученого в области теории массового обслуживания К. Пальма был объявлен конкурс на название единицы телефонной нагрузки. За большие заслуги в создании развитии теории телетрафика имя Эрланга было увековечено в названии этой единицы.

1.2. Исторический обзор развития дисциплины «Теория телетрафика»

В каждом начинании именно великие люди – личности придают работе смысл, направление и цель. Агнер Краруп Эрланг был первым, кто приступил к исследованию проблем телефонных сетей и установил закономерности, которые оказались применимыми к огромному числу проблем, связанных с массовым обслуживанием.

Анализируя работу автоматических телефонных станций, он установил, что если λ – параметр потока, то вероятность поступления n вызовов за время t определяется законом Пуассона:

$$P_n(t) = \frac{(\lambda \cdot t)^n}{n!} \cdot e^{-\lambda \cdot t}. \quad (1.1)$$

Эрланг исследовал три вида обслуживания абонентов телефонной сети в системе коммутации V каналов, имеющихся на АТС:

1. Полнодоступный пучок линий (система с потерями) обеспечивает доступ любого из абонентов, число которых намного превышает значение V , к любому из свободных в данный момент каналов связи. Если в момент поступившего вызова свободных каналов нет, то происходит потеря вызова.

2. Полнодоступный пучок линий (система с ожиданиями) также обеспечивает доступ любого из абонентов к любому из свободных в данный момент каналов связи. Однако если в момент поступившего вызова свободных каналов нет, то потери вызова не происходит, а абонент ставится в очередь, через некоторое время происходит соединение и ему предоставляется канал связи.

3. Неполнодоступный пучок линий (система с потерями) обеспечивает доступ любого из абонентов только к ограниченному числу $D < V$ имеющихся линий в том случае, если хотя бы одна из этих D линий свободна в момент поступления вызова. В противном случае вызов теряется.

Ключевой вклад А. К. Эрланга в современную теорию стохастических процессов состоит в идее стохастического равновесия, которое означает, что вероятность пребывания системы в определенном состоянии (для телетрафика – число занятых в пучке линий) не зависит от момента, в который рассматривается система.

Используя эти предпосылки, А. К. Эрланг разработал теорию телетрафика и для трех указанных выше видов обслуживания получил формулы для расчета вероятностей потери вызова и распределения времени ожидания при равновесном состоянии системы. Такие формулы он получил для системы с ожиданием при постоянном времени обслуживания в случае использования одного, двух и трех каналов, а также при произвольном числе каналов и экспоненциальном времени обслуживания.

Исследования Эрланга по теории телетрафика имели международное признание. В литературе установленные Эрлангом формулы для трех видов обслуживания называют в его честь 1-й, 2-й и 3-й формулами Эрланга. Они были официально рекомендованы Британским министерством связи для использования при проектировании АТС. Эрланг был принят в число членов Британского института инженеров-электриков.

Несмотря на то, что модель Эрланга весьма проста, математическое описание работы современных сложных телефонных сетей все еще основывается именно на ней.

Теория телетрафика, с одной стороны – это один из разделов теории связи, а с другой – один из крупных, интенсивно развивающихся разделов теории массового обслуживания. Ее формирование как раздела теории связи обусловлено быстрыми темпами совершенствования сетей связи, что делает весьма важным повышение эффективности ее использования путем управления потоками сообщений, передаваемых по каналам сетей.

За рубежом значительный вклад в обобщение и развитие теории телетрафика внесли Г. О'Делл, К. Пальм, К. Якобеус, К. Ли, изучавшие случаи фиксированного числа абонентов, подключенных к АТС, ограниченного времени подключения абонентов к каналу связи и разрабатывавшие методы расчета многозвенных коммутационных систем и др.

В СССР первые работы по теории телетрафика были выполнены академиком М. С. Юрьевым в 1925 г. В последующие годы значительные научные исследования с привлечением метода статистического моделирования на ЭВМ для решения данных проблем были выполнены учеными Г. П. Башариным, Б. С. Лившицем, А. Д. Харкевичем, М. А. Шнепсом и др.

В области современной связи результаты А. К. Эрланга широко используются для проектирования не только АТС, но и сотовых, транкинговых и спутниковых систем связи с подвижными объектами, а также систем фиксированной спутниковой связи VSAT.

Кроме теории телетрафика, А. Эрланга интересовали вопросы составления таблиц математических функций. Он имел необыкновенно обширные познания по истории математических таблиц с древних времен до настоящего времени; установил новые принципы вычисления некоторых форм математических таблиц, в частности таблиц логарифмов.

С этой точки зрения, научные достижения Эрланга следует отнести к разряду крупнейших. Он открыл новый обширный научный материк, осваивать который начали сотни последователей. Развитие его идей привело к созданию теории массового обслуживания, которая дает специалистам инструмент для решения огромного числа со-

циальных проблем, связанных с обслуживанием людей, с поточным производством различной продукции, а также других прикладных задач.

Совсем недолгую жизнь прожил А. К. Эрланг. Однако его судьба показывает, что смерть не может быть и не является концом. Человек оставляет после себя то, с чем связывают его имя, и творения, которые делают из этого имени символ восторга, ненависти или безразличия.

Вопросы для самоконтроля

1. Кто является основоположником теории телетрафика?
2. В чем заключалась необходимость зарождения теории телетрафика?
3. Чему были посвящены основные работы А. К. Эрланга?
4. Раскройте краткую биографию Агнера Крауупа Эрланга.
5. Какие три вида обслуживания абонентов телефонной сети в системе коммутации каналов исследовал А. К. Эрланг?
6. Перечислите известных ученых, работы которых внесли значительный вклад в развитие теории телетрафика.

2. ОСНОВНЫЕ ПОЛОЖЕНИЯ ТЕОРИИ ВЕРОЯТНОСТЕЙ, ИСПОЛЬЗУЕМЫЕ В РАМКАХ ДИСЦИПЛИНЫ «ТЕОРИЯ ТЕЛЕТРАФИКА»

2.1. Событие. Вероятность события

Теория вероятностей – раздел математики, посвященный случайным величинам. Точное определение термина «случайная величина», отвечающее строгим математическим канонам, можно найти в изданиях, посвященных фундаментальным основам теории вероятностей. Для прикладных дисциплин можно использовать менее строгие определения.

Случайные величины обычно разделяют на дискретные и непрерывные. Дискретный или непрерывный характер случайной величины определяется объективными свойствами исследуемого процесса. Для проведения анализа некоторых систем массового обслуживания (СМО) целесообразно переходить от непрерывных случайных величин к дискретным и наоборот.

Допустим, что проведено N измерений числа разговаривающих абонентов АТС. При этом K раз их численность (событие « A ») была одной и той же. Тогда вероятность наступления интересующего события – $P(A)$ определяется следующим образом:

$$P(A) = \lim_{N \rightarrow \infty} \frac{K}{N}. \quad (2.1)$$

Рассмотрим пример, когда в результате проведения 1000 измерений 50 раз обнаружено 800 разговаривающих абонентов (событие « A »). Оценку 0,05, строго говоря, нельзя считать вероятностью $P(A)$, так как число проведенных измерений было конечной величиной. Эту оценку 0,05 называют частотой, или частостью.

В теории вероятностей важную роль играют аксиомы, которые сформулированы известным российским математиком А. Н. Колмогоровым. Четыре основные аксиомы приводятся ниже в следующей форме:

а) каждому событию « A » ставится в соответствие неотрицательное число – его вероятность $P(A) \geq 0$;

б) вероятность достоверного события " Ω " равна единице – $P(\Omega) = 1$;

в) если « A » и « B » – непересекающиеся события, то вероятность события « A » или « B » (оно обычно обозначается как « $A + B$ ») – $P(A + B)$ равна сумме $P(A) + P(B)$;

г) условная вероятность наступления события « B », если уже произошло событие « A », – $P(B|A)$ определяется как $P(AB) / P(A)$.

Для пояснения термина «условная вероятность» целесообразно использовать две простые геометрические фигуры. Они приведены на рис. 2.1. Подобный подход позволяет наглядно интерпретировать события и вероятность их наступления.

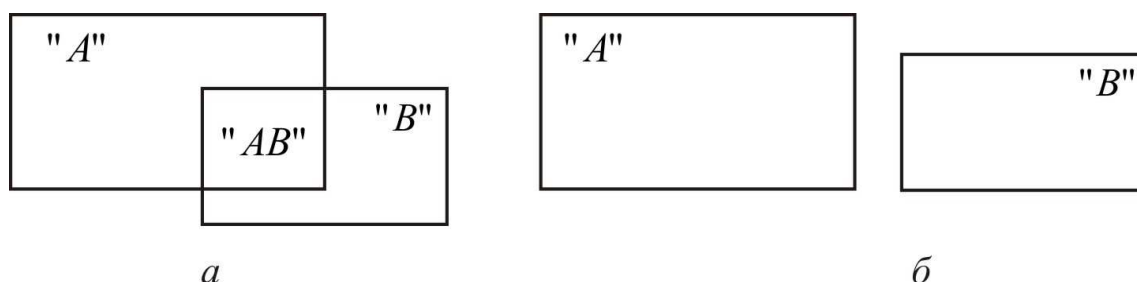


Рис. 2.1. Геометрическая интерпретация условной вероятности:

a – события « A » и « B » совместны;

$б$ – события « A » и « B » несовместны.

События « A » и « B » заключаются в попадании точки в одноименные области, которые показаны на рис. 2.1, a . Событие « AB » имеет одну особенность: оно происходит, когда события « A » и « B » наступают одновременно. Тогда вероятность события « B » при условии, что уже произошло событие « A », представима как отношение площадей « AB » и « A ». Если события « A » и « B » несовместны, то условная вероятность равна нулю. Это утверждение наглядно иллюстрирует рис. 2.1, $б$: площадь пересечения двух областей равна нулю.

Из аксиом теории вероятностей можно сделать ряд важных для теории и практики выводов. В частности если могут наступить только события « A » и « $\bar{A}$ », то справедливы такие соотношения:

$$P(A + \bar{A}) = 1 \text{ или } P(A) = 1 - P(\bar{A}). \quad (2.2)$$

Очевидно также, что $0 \leq P(A) \leq 1$. Из аксиомы (в) можно получить более общее соотношение для попарно непересекающихся событий:

$$P(A + B + C + \dots + Z) = P(A) + P(B) + P(C) + \dots + P(Z). \quad (2.3)$$

Полной характеристикой случайной величины служит закон ее распределения. Он устанавливает соответствие между возможными значениями случайной величины и соответствующими вероятностями. Знание закона распределения позволяет сравнительно простыми математическими методами получить оценки случайной величины, важные для практической работы.

Закон распределения может быть описан различными способами. Основным интерес представляет функция распределения (ФР) случайной величины – $F(t)$. Аргументом ФР служит буква « t », указывающая на время. Такой выбор объясняется тем, что во многих языках слово «время» начинается с буквы « t ». Далее будут рассматриваться случайные величины, определенные для $t \geq 0$. Пример ФР случайной величины показан на рис. 2.2. Эта функция определена на отрезке $[0, T_{\max}]$.

ФР представляет собой монотонно возрастающую функцию. Она определяется соотношением

$$F(t) = P(T < t). \quad (2.4)$$

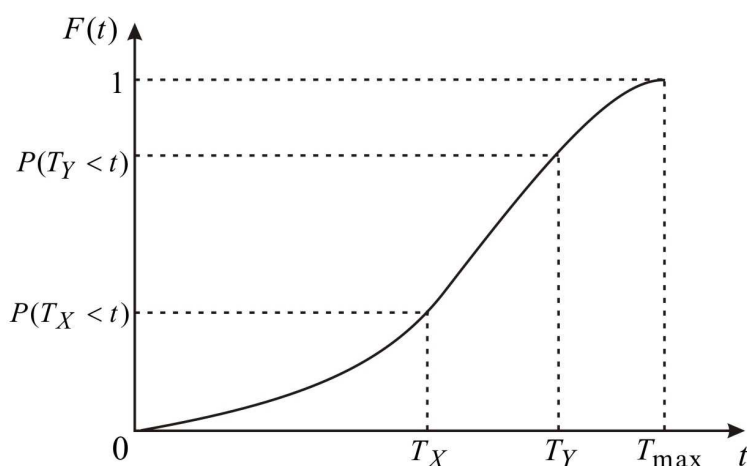


Рис. 2.2. Функция распределения случайной величины

Иными словами, ФР равна вероятности соблюдения неравенства $T < t$. На рис. 2.2 указаны две точки T_X и T_Y , которым соответствуют вероятности $P(T_X < t)$ и $P(T_Y < t)$. Вероятность того, что случайная величина попадет в интервал (T_X, T_Y) , равна разности

$F(T_Y) - F(T_X)$. В некоторых случаях практический интерес представляет дополнительная ФР – $P(T \geq t)$. Очевидно, что

$$P(T \geq t) = 1 - F(t). \quad (2.5)$$

Следует отметить, что в некоторых публикациях ФР $F(t)$ определяется иначе – нестрогим неравенством:

$$F(t) = P(T \leq t). \quad (2.6)$$

Различие в этих определениях существенно только для дискретных случайных величин.

2.2. Законы распределения случайных величин

В теории телетрафика часто используется предположение о пуассоновском законе распределения случайных величин. Такое предположение, например, было подтверждено экспериментально для потока вызовов, поступающих от абонентов телефонной станции. Говорят, что случайная величина X распределена по закону Пуассона, если вероятность того, что она примет определенное значение x , определяется следующим образом:

$$P_x = P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda} \quad (x = 0, 1, 2, \dots), \quad (2.7)$$

где λ – интенсивность потока вызовов.

Переменную « x » можно рассматривать как число вызовов, поступающих в течение интервала времени фиксированной длины. Функция распределения этого потока вызовов – $F(x)$ равна нулю для $x \leq 0$. Для $x > 0$ она определяется соотношением

$$F(x) = P(X \leq x) = e^{-\lambda} \sum_{i=0}^x \frac{\lambda^i}{i!} \quad (x = 1, 2, 3, \dots). \quad (2.8)$$

На рис. 2.3 показаны два примера функции P_x : в левой части $\lambda = 1$, а в правой – $\lambda = 3$.

Второй интересный пример – дискретное равномерное распределение. Буквой « b » обозначена левая граница области изменения случайной величины X . Допустим, что рассматриваемая случайная величина принимает n значений (рис. 2.4).

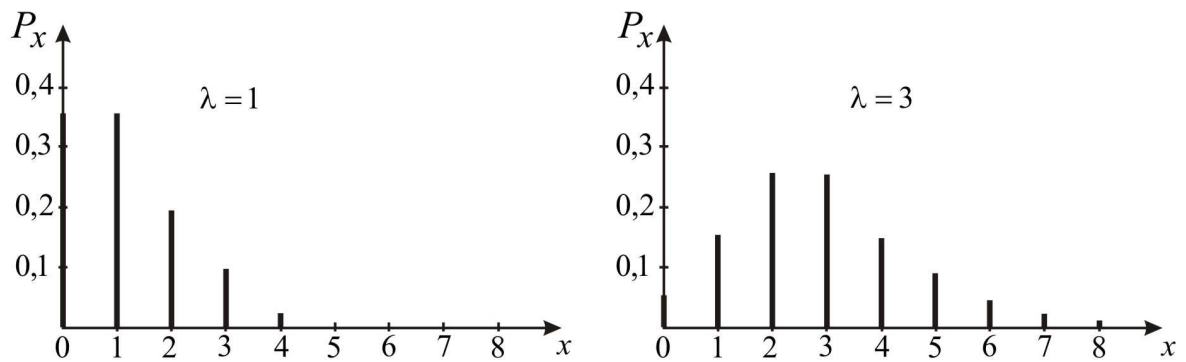


Рис. 2.3. Два примера распределения по закону Пуассона

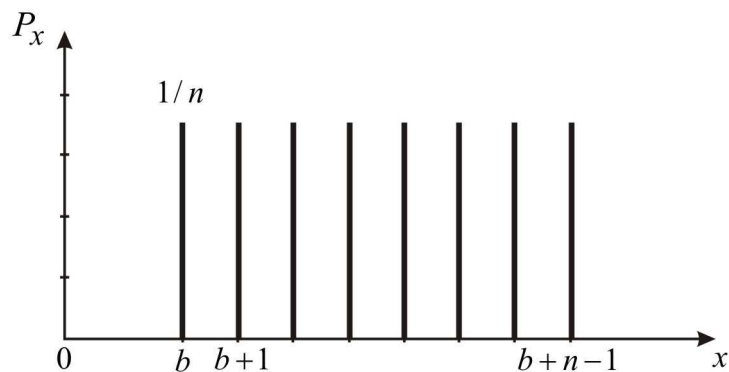


Рис. 2.4. Дискретное равномерное распределение

Для обоих примеров ФР будет ступенчатой. Это утверждение справедливо для всех законов распределения дискретных случайных величин.

В теории телетрафика часто используется экспоненциальное распределение, для которого функции $f(t)$ и $F(t)$ определяются формулами:

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t < T < t + \Delta t)}{\Delta t} = \begin{cases} \lambda e^{-\lambda t} & \text{при } t > 0; \\ 0 & \text{при } t \leq 0, \end{cases}$$

$$F(t) = P(T < t) = \begin{cases} 1 - e^{-\lambda t} & \text{при } t > 0; \\ 0 & \text{при } t \leq 0. \end{cases} \quad (2.9)$$

Еще один пример связан с распределением случайной величины на ограниченном интервале. Речь идет о равномерном распределении на отрезке времени $[a, b]$. В этом интервале функции $f(t)$ и $F(t)$ определяются следующим образом:

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t < T < t + \Delta t)}{\Delta t} = \frac{1}{b-a} \text{ при } a \leq t \leq b,$$

$$F(t) = P(T < t) = \frac{t-a}{b-a} \text{ при } a \leq t \leq b. \quad (2.10)$$

Функции $f(t)$ и $F(t)$ показаны на рис. 2.5.

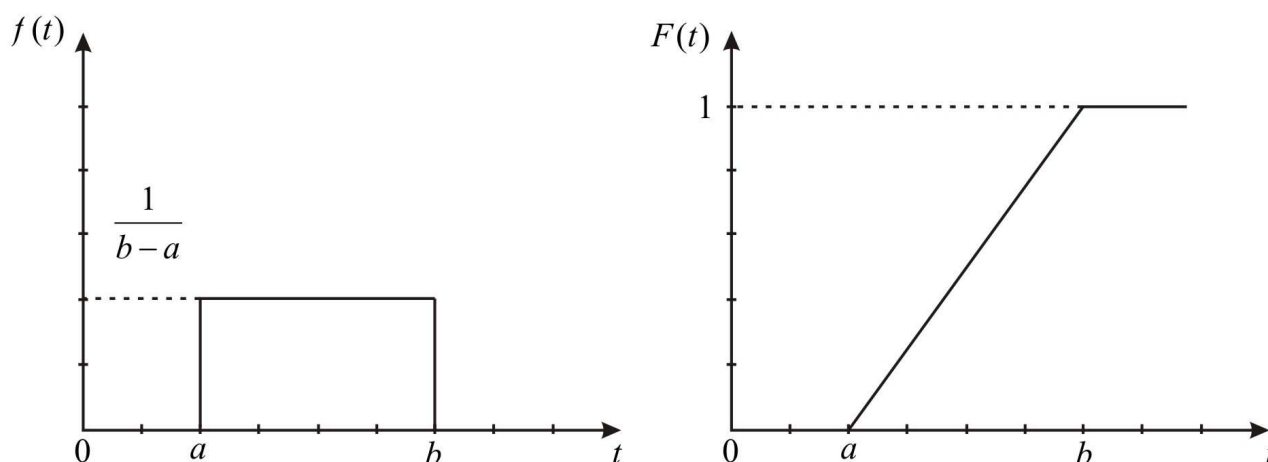


Рис. 2.5. Равномерное распределение на интервале $[a, b]$

2.3. Числовые характеристики случайных величин

Основные характеристики случайной величины могут быть получены из функции $F(t)$. Среди этих характеристик большое практическое значение имеют математическое ожидание, дисперсия, асимметрия и эксцесс. Математическое ожидание (или среднее значение) случайной величины – $M(T)$ может быть рассчитано по формуле

$$m_t = M(T) = \int_0^{\infty} t dF(t). \quad (2.11)$$

Математическое ожидание представляет собой начальный момент первого порядка. В общем случае начальный момент порядка R – $M(T^R)$ определяется соотношением

$$M(T^R) = \int_0^{\infty} t^R dF(t). \quad (2.12)$$

В некоторых изданиях используется иная формула для расчета математического ожидания случайной величины – через плотность вероятности $f(t)$. Последняя формула удобна тем, что она универсальна для непрерывных и дискретных случайных величин. Следует отметить, что случайная величина может не иметь математического ожидания. Для большинства задач такие ситуации не представляют практического интереса.

Математическое ожидание суммы K случайных величин определяется следующим образом:

$$M\left(\sum_{J=1}^K T_J\right) = \sum_{J=1}^K M(T_J). \quad (2.13)$$

Математическое ожидание произведения K независимых случайных величин может быть вычислено по формуле

$$M\left(\prod_{J=1}^K T_J\right) = \prod_{J=1}^K M(T_J). \quad (2.14)$$

Математическое ожидание определяет положение центра распределения случайной величины. Для функции $f(t)$ обычно определяют медиану и моду.

Медиана делит площадь под кривой $f(t)$ пополам. Мода непрерывной случайной величины – такое значение t , в котором функция $f(t)$ достигает локального максимума. Если функция $f(t)$ имеет один максимум, то распределение называется унимодальным. Мультимодальное распределение имеет несколько мод. На рис. 2.6 показаны два графика функции $f(t)$, которая является унимодальной.

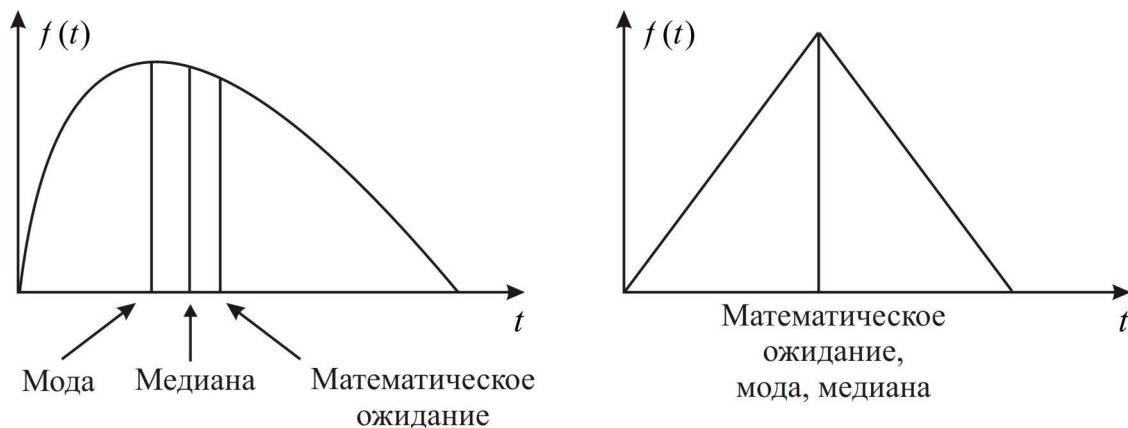


Рис. 2.6. Графики функции $f(t)$, поясняющие математическое ожидание, медиану и моду

В левой части этого рисунка дано распределение, для которого значения математического ожидания, медианы и моды различны. Для функции $f(t)$, изображенной в правой части рисунка, значения математического ожидания, медианы и моды совпадают.

Дисперсия случайной величины – $D(T)$ характеризует меру рассеяния случайной величины. Она рассчитывается как математическое ожидание квадрата отклонения случайной величины от своего среднего значения:

$$D_t = D(T) = M[(t - M(T))^2]. \quad (2.15)$$

Дисперсия является центральным моментом второго порядка. Центральный момент порядка $R - E(T^R)$ определяется соотношением

$$E(T^R) = M[(t - M(T))^R]. \quad (2.16)$$

Корень второй степени из дисперсии – σ_t называется среднеквадратическим (стандартным) отклонением. Отношение среднеквадратического отклонения к среднему значению (если оно больше нуля), именуется коэффициентом вариации случайной величины – k_t

$$k_t = \frac{\sigma_t}{M(T)}. \quad (2.17)$$

Начальные и центральные моменты порядка R для упрощения часто обозначают как a_R и μ_R соответственно. Для этих моментов справедливы следующие соотношения:

$$\begin{aligned} a_R &= \sum_{J=0}^R C_R^J \mu_{R-J} a_1^J, \\ \mu_R &= \sum_{J=0}^R (-1)^J a_{R-J} a_1^J. \end{aligned} \quad (2.18)$$

Коэффициент асимметрии распределения случайной величины – γ_t определяет степень неравномерности плотности вероятности $f(t)$ относительно своего центра. Он рассчитывается по формуле

$$\gamma_t = \frac{\mu_3}{(\sigma_t)^3}. \quad (2.19)$$

Коэффициент эксцесса распределения случайной величины – ν_t характеризует «островершинность» плотности вероятности $f(t)$. Обычно эту характеристику применяют к унимодальным распределениям:

$$\nu_t = \frac{\mu_4}{(\sigma_t)^4} - 3. \quad (2.20)$$

На рис. 2.7 приведены примеры непрерывных распределений.

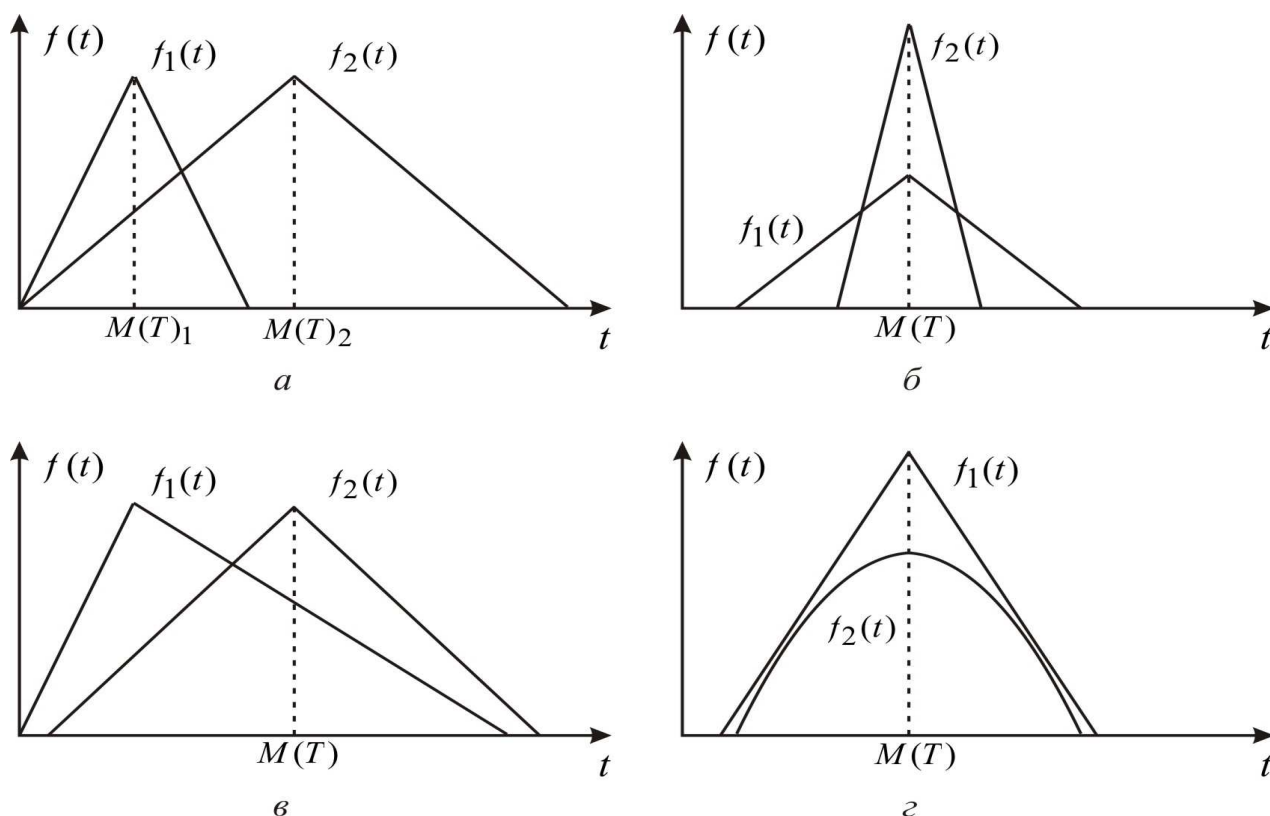


Рис. 2.7. Четыре примера непрерывных распределений:

а – сравнение математических ожиданий;

б – сравнение дисперсий;

в – сравнение коэффициентов асимметрии;

г – сравнение коэффициентов эксцесса.

Важной характеристикой ФР следует считать квантиль (t_X). На рис. 2.8 показаны два квантиля, для которых значения ФР составляют 0,5 и 0,95.

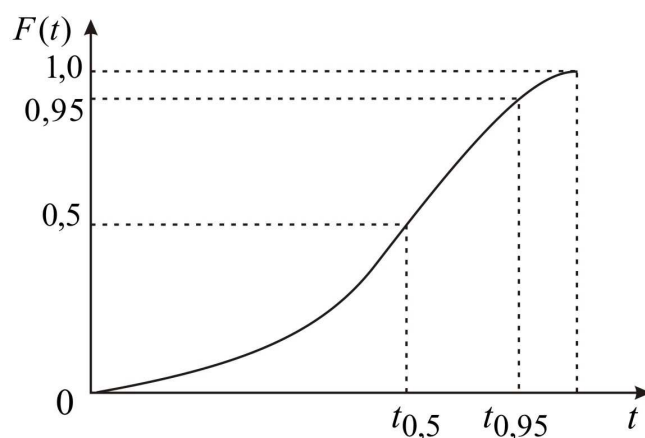


Рис. 2.8. Квантили функции распределения

Графически квантиль определяется очень просто. Аналитически значения квантиля t_X можно получить решением уравнения

$$F(t_X) = x. \quad (2.21)$$

Иногда функция распределения определяется в процентах – от 0 до 100%. Тогда квантиль также задается в процентах. В этом случае, вместо термина «квантиль», используется термин «процентиль».

Основные характеристики пуассоновского, дискретного равномерного, экспоненциального и непрерывного равномерного на отрезке $[a, b]$ распределений представлены в табл. 2.1 – 2.4.

Таблица 2.1

Характеристики пуассоновского распределения

Математическое ожидание	Дисперсия	Асимметрия	Эксцесс
λ	λ	$\frac{1}{\sqrt{\lambda}}$	$\frac{1}{\lambda}$

Таблица 2.2

Характеристики дискретного равномерного распределения

Математическое ожидание	Дисперсия	Асимметрия	Эксцесс
$b + \frac{n-1}{2}$	$\frac{n^2 - 1}{12}$	0	$-1,2 - \frac{2,4}{n^2 - 1}$

Таблица 2.3

Характеристики экспоненциального распределения

Математическое ожидание	Дисперсия	Асимметрия	Эксцесс
$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	2	9

Таблица 2.4

Характеристики непрерывного равномерного распределения на отрезке $[a, b]$

Математическое ожидание	Дисперсия	Асимметрия	Эксцесс
$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	0	-1,2

2.4. Использование преобразования Лапласа – Стилтеса для получения числовых характеристик случайных величин

Преобразование Лапласа устанавливает однозначную связь между функциями действительной – $B(t)$ и комплексной переменных – $\beta(s)$. Эти функции обычно называют оригиналом и изображением соответственно. Для оригиналов, существующих только в области неотрицательных значений аргумента (времени), целесообразно использовать одностороннее преобразование Лапласа:

$$\beta(s) = \int_0^{\infty} B(t)e^{-st} dt. \quad (2.22)$$

Функция $\beta(s)$ комплексной переменной $s = \sigma + i\omega$ позволяет найти оригинал по формуле

$$B(t) = \frac{1}{2\pi i} \int_{\sigma - i\infty}^{\sigma + i\infty} \beta(s)e^{st} ds. \quad (2.23)$$

Это соотношение известно в математике как формула обращения Римана – Моллина. Его также называют обратным преобразованием Лапласа.

В обеих формулах используются интегралы Римана. Во многих случаях приходится оперировать дискретными функциями плотности вероятности. Тогда предпочтительнее становится интеграл Стильтьеса, позволяющий унифицировать тип распределения случайной величины. Интеграл Стильтьеса определяется для двух функций – интегрируемой $\varphi(x)$ и интегрирующей $F(x)$:

$$\int \varphi(x) dF(x). \quad (2.24)$$

Очевидно, что интеграл Римана представляет собой частный случай интеграла Стильтьеса, когда $F(x) = x$. Преобразование Лапласа – Стильтьеса для функции $B(t)$ будет определяться следующим образом:

$$\phi(s) = \int_0^{\infty} e^{-st} dB(t). \quad (2.25)$$

Если изображения $\beta(s)$ и $\phi(s)$ существуют, то они связаны между собой очевидным соотношением

$$\beta(s) = s\phi(s). \quad (2.26)$$

Интерес к преобразованию Лапласа – Стильтьеса объясняется тем, что оно позволяет упростить исследование некоторых функций. Для вычисления параметров распределения следует выделить ряд свойств, которые характерны для преобразования Лапласа – Стильтьеса:

1. Преобразование Лапласа – Стильтьеса производной от функции $B(t)$ определяется по формуле

$$\frac{dB(t)}{dt} \rightarrow s\beta(s) - sB(+0). \quad (2.27)$$

2. Дифференцируя изображение, можно получить r -й начальный момент (b_r) распределения

$$b_r = (-1)^r \beta^{(r)}(s)|_{s=0}. \quad (2.28)$$

3. ФР двух независимых случайных величин – $\phi(s)$ определяются произведением их преобразований Лапласа – Стильеса – $\phi_1(s)$ и $\phi_2(s)$:

$$\phi(s) = \phi_1(s)\phi_2(s). \quad (2.29)$$

4. Сдвиг аргумента у оригинала на величину τ – первая теорема смещения – соответствует изменению изображения

$$B(t - \tau) \rightarrow e^{-\tau s} \phi(s). \quad (2.30)$$

Из всех свойств, присущих преобразованию Лапласа – Стильеса, следует выделить свертку оригиналов – третье свойство. Вычисление свертки функций с помощью подобных преобразований можно рассматривать как одну из самых эффективных операций среди возможностей, присущих преобразованию Лапласа – Стильеса.

Использование преобразования Лапласа – Стильеса в большинстве случаев можно представить в виде алгоритма, показанного на рис. 2.9. Этот алгоритм универсален с точки зрения решаемых задач.

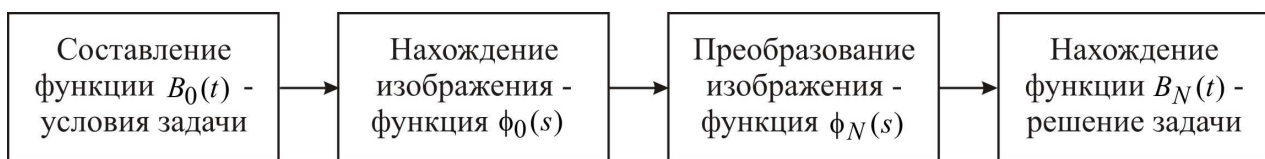


Рис. 2.9. Использование преобразования Лапласа – Стильеса

Функция $B_0(t)$ представляет собой выражение, которое необходимо преобразовать для искомого результата. Эту функцию можно считать своего рода условием задачи. Сначала находится изображение этой функции – $\phi_0(s)$. Изображение может быть получено с помощью таблиц преобразования Лапласа – Стильеса. Именно для функции $\phi_0(s)$ выполняются преобразования, позволяющие получить ответ в виде функции $\phi_N(s)$. Для решения поставленной задачи необходимо найти функцию $B_N(t)$. Эта процедура может быть выполнена с использованием таблиц или иным способом.

Для вычисления оригинала по известному изображению часто применяют теорему разложения, предложенную Хевисайдом (разло-

жение на элементарные дроби). Такой способ приемлем, если изображение $\phi(s)$ является рациональной дробью следующего вида:

$$\phi(s) = \frac{\phi_1(s)}{\phi_2(s)}. \quad (2.31)$$

Числителем и знаменателем этого выражения служат полиномы $\phi_1(s)$ и $\phi_2(s)$. Существенно то, что степень полинома $\phi_1(s)$ меньше степени полинома $\phi_2(s)$. Если уравнение $\phi_2(s) = 0$ имеет n различных корней, то функция $\phi(s)$ представима суммой:

$$\phi(s) = \sum_{K=1}^n \frac{\phi_1(s_K)}{\phi_2(s_K)} \cdot \frac{1}{s - s_K}. \quad (2.32)$$

Используя таблицы обратного преобразования Лапласа, можно найти оригинал в виде

$$B(t) = \sum_{K=1}^n \frac{\phi_1(s_K)}{\phi_2(s_K)} \cdot e^{s_K t}. \quad (2.33)$$

Вопросы для самоконтроля

1. Дайте определение вероятности события.
2. Какими способами может быть описан закон распределения случайной величины?
3. Перечислите и обоснуйте числовые характеристики случайных величин.
4. Раскройте основные характеристики пуассоновского, дискретного равномерного, экспоненциального и непрерывного равномерного распределений на фиксированном отрезке.
5. Между какими функциями устанавливает связь преобразование Лапласа – Стильтеса?

3. ТЕОРИЯ ТЕЛЕТРАФИКА. БАЗОВЫЙ ТЕОРЕТИЧЕСКИЙ КУРС

3.1. Объект, предмет и задачи теории телетрафика

Вначале попытаемся выяснить, чем же теория телетрафика (ТТТ) отличается от теории массового обслуживания (ТМО).

В ТМО все рассматриваемые объекты, в том числе системы распределения информации, объединяются под общим названием «системы массового обслуживания».

Объектом в ТТТ являются только системы распределения информации, или системы коммутации. При этом к системам распределения информации (коммутации) относят сеть связи, коммутационную станцию или узел и их отдельные части (коммутационные поля, управляющие устройства), обслуживающие сообщения по определенному алгоритму.

Сообщение – это совокупность информации, имеющей признаки начала и конца и предназначенной для передачи через сеть связи или коммутационную систему. Сообщения характеризуются видом, объемом, категоричностью, адресами источника и приемника сообщений, формой представления (аналоговой или дискретной).

Предметом изучения ТТТ является количественная сторона процессов обслуживания системами коммутации потоков сообщений.

Основная цель теории телетрафика заключается в разработке методов оценки качества функционирования систем распределения информации, т. е. составлении математических моделей систем коммутации, изучении соотношений между величиной, характером телефонной нагрузки (в настоящее время гетерогенной) и количеством коммутационного оборудования при заданном качестве обслуживания.

Примечание. Согласно Рекомендации МККТТ G.106 (Красная книга) качество обслуживания абонента на сети электросвязи (*Quality of service*) определяется как степень удовлетворенности абонента обслуживанием на сети.

Качество обслуживания абонента на сети электросвязи определяется совокупностью условий:

– обеспеченностью обслуживанием (возможностью предоставления абоненту различных видов обслуживания);

- эксплуатационной пригодностью обслуживания (возможностью абонента легко и успешно управлять процессом обслуживания);
- эффективностью обслуживания (способностью сети по требованию абонента предоставлять обслуживание на требуемое время в пределах заданных допусков на показатели качества передачи сигналов);
- целостностью обслуживания (способностью предоставлять обслуживание с заданным качеством передачи без его существенного ухудшения).

Из всех перечисленных условий основным, т. е. характеризующим качество обслуживания системой распределения информации как системой массового обслуживания (область исследования теории телетрафика), является эффективность обслуживания.

В свою очередь эффективность обслуживания определяется двумя основными свойствами:

- качеством обслуживания вызовов;
- пропускной способностью.

Качество обслуживания вызовов характеризуется вероятностью потерь вызовов (P), вероятностью того, что время ожидания будет превышать заданную величину ($P(t_{\text{ож}} > t_{\text{Д}})$), вероятностью наличия очереди ($P(t_{\text{ож}} > 0)$) и другими характеристиками.

Пропускная способность – это свойство системы коммутации обслуживать поступающую телефонную нагрузку с заданной интенсивностью и нормируемым качеством обслуживания вызовов.

Основные задачи, решаемые в рамках теории телетрафика.

На первом месте в ТТТ стоят задачи анализа, т. е. отыскание зависимостей и значений величин, характеризующих качество обслуживания требований, от характеристик и параметров входящего потока вызовов, схемы и дисциплины обслуживания. Символически это можно записать в виде зависимостей $P = f(V, d, Z, N)$, $P(t_{\text{ож}} > \tau) = f(V, d, Z, N)$, где P и $P(t_{\text{ож}} > \tau)$ – характеристики качества обслуживания (вероятность потерь вызовов, вероятность ожидания), V – канальный ресурс звена, d – доступность канального ресурса, Z – интенсивность поступающей нагрузки, N – число источников нагрузки.

Эти задачи в начальный период развития коммутационной техники были более актуальными и решались с помощью теории вероят-

ностей, поэтому наиболее значительные результаты на сегодняшний день получены при решении задач анализа. Но это не значит, что все они уже решены. Развитие сетей связи и, в частности, создание цифровых интегральных сетей или цифровых сетей с интеграцией служб (ЦСИС), внедрение цифровых систем коммутации (ЦСК), обслуживающих гетерогенный трафик, выдвигает перед исследователями необходимость решения задач анализа для новых условий.

Наряду с этим часто возникает необходимость в решении задач обратного вида (задачи синтеза) – нахождение параметров систем обслуживания при заданной их структуре в зависимости от параметров и свойств потока вызовов, времени обслуживания сообщений, дисциплины и качества обслуживания $V = f(Z, N, P, d)$; $V = f(Z, d, N, P(t_{\text{ож}} > \tau))$ либо определении пропускной способности системы обслуживания, производительности системы управления при заданной ее структуре и параметрах, нормируемом качестве обслуживания, известных дисциплине обслуживания, свойствах потока и законе распределения времени обслуживания сообщений $Y(P) = f(Z, N, d, V)$.

Решение задач анализа и синтеза применительно к системе управления программно-управляемой ЦАТС составило важное направление теоретических исследований.

К этим задачам относят определение:

- производительности отдельной ЭВМ или всей системы управления в целом при заданной структуре управления и известном объеме ресурсов;
- необходимого объема ресурсов для получения требуемой производительности отдельной ЭВМ или системы при заданной ее структуре;
- оптимального уровня децентрализации управления (выбор оптимальной структуры управления) в зависимости от конкретных характеристик разрабатываемой системы АТС и т. д.

Близкими к задачам анализа и синтеза являются задачи оптимизации, которые при проектировании систем распределения информации формулируются следующим образом – определить такие структурные параметры коммутационных систем (алгоритмы их функционирования), для которых:

- при заданных потоках, качестве и дисциплине обслуживания стоимость или объем оборудования системы распределения информации минимальны;

– заданных потоках, дисциплине обслуживания и стоимости качественные показатели функционирования системы оптимальны.

Поиск наиболее экономичных структур коммутационных схем, разработка принципов их построения и методов расчета пропускной способности становятся основными направлениями теории. При этом выполнение условий пропускной способности связывается не только со стоимостью коммутационной схемы, но и с такими факторами, как надежность функционирования, гибкость развития, модульность построения, время передачи сообщения через коммутационное поле.

К другим направлениям в решении этих задач относятся оптимизация структуры сети связи и расчет ее пропускной способности, а также других показателей качества обслуживания.

В постановке и методах решения этих задач, требованиях, предъявляемых к оптимальной структуре сети, много общего с рассмотренными выше задачами по анализу и синтезу коммутационных схем. Различие обусловлено, в основном, особенностями структур. Коммутационные схемы имеют, как правило, однородную симметричную структуру, топология их изменяется довольно редко, что позволяет с той или иной степенью точности рассчитывать и оптимизировать всю схему в целом.

Для сетей связи, наоборот, характерны неоднородность структуры и отсутствие симметрии между элементами, динамизм и частое изменение топологии – зарождение и развитие одних элементов и отмирание других. В этих условиях трудно найти решение сразу для всей сети. Расчет и оптимизацию приходится производить по частям для отдельных фрагментов и участков сети.

В последние десятилетия заметно усилился интерес к задачам иного рода – разработке математических моделей краткого и долгосрочного прогнозов ожидаемых параметров потоков сообщений и длительности обслуживания, исследованию свойств потоков вызовов и законов распределения времени обслуживания в реальных системах коммутации и управления.

Обусловлено это возросшими значениями стоимости и сложностью систем коммутации и сетей электросвязи и, соответственно, повышением требований к их экономичности и качеству работы. Опыт проектирования и эксплуатации систем коммутации и сетей электросвязи свидетельствуют о том, что наибольшие просчеты в определении объема необходимого станционного и линейного оборудования,

в расчете требуемой производительности управляющих устройств происходят именно из-за ошибок в прогнозе ожидаемой интенсивности потока сообщений и времени их обслуживания. При этом ошибки совершаются как на этапе создания оборудования, так и на стадии его проектирования и не всегда исправляются в процессе эксплуатации.

Нормирование и оптимальное распределение по участкам сети и этапам соединения показателей качества обслуживания – еще один класс задач, решаемых в теории телетрафика.

От установленной общей нормы качества обслуживания поступающих сообщений зависят, с одной стороны, объем и стоимость оборудования станции и всей сети, а с другой – время, затрачиваемое абонентом на установление требуемого соединения, и время передачи сообщения через сеть связи. Несвоевременная доставка или срыв передачи сообщений приводит к замедлению процессов управления в различных областях народного хозяйства, что наносит ущерб потребителям услуг связи.

3.2. Базовая модель систем распределения информации

Анализ особенностей исследования систем распределения информации на различных этапах их жизненного цикла (разработка, планирование, эксплуатация) показывает, что основными методами получения информации, необходимыми для принятия решений по их созданию и применению, являются теоретические методы. Их реализация сводится к моделированию объекта исследования. Причем применительно к системам распределения информации следует говорить лишь о математическом моделировании, под которым понимают исследование объекта, базирующееся на его подобии модели, в частности подобии математическом.

Таким образом, теория телетрафика, как и любая другая математическая теория, оперирует не с самими системами распределения информации, а с их математическими моделями.

Основная математическая модель системы распределения информации, исследуемая в теории телетрафика, содержит следующие компоненты:

- поток поступающих вызовов;
- время их обслуживания;
- систему обслуживания (систему коммутации);
- дисциплину;

– характеристики качества обслуживания поступающих сообщений.

Прежде чем приступить к рассмотрению потока вызовов, следует раскрыть понятия, используемые в теории телетрафика.

Заявка (телефонная) – требование на обслуживание, поступившее от источника телефонной нагрузки.

Вызов – телефонная заявка на установление соединения, поступившая от оконечного телефонного устройства на телефонную станцию.

Поток телефонных вызовов – совокупность последовательности телефонных вызовов, поступающих через фиксированные или случайные интервалы времени, или в какие-либо его моменты.

Понятие «поток вызовов» достаточно широкое. Оно включает в себя информацию о модели (свойствах и характеристиках) потока вызовов, виде передаваемых сообщений и форме их представления, множестве адресов источников и приемников информации.

Время обслуживания сообщений может быть случайным или детерминированным. В первом случае оно задается по вероятностному закону распределения, во втором – по длительности.

Система обслуживания характеризуется структурой построения (полно- или неполнодоступная, одно- или многозвенная, одно- или многоступенчатая, одно- или многофазная), режимом искания выходов коммутационной системы (свободный, групповой или индивидуальный и т. д.) и набором структурных параметров.

Дисциплина обслуживания потоков вызовов – это характеристика алгоритма взаимодействия поступающего потока телефонных вызовов с коммутационной станцией, включающая способ и порядок их обслуживания.

Способ обслуживания телефонных вызовов – характеристика процесса обслуживания телефонных вызовов, которая может осуществляться без потерь, с явными потерями, с ожиданием.

Порядок обслуживания телефонных вызовов – совокупность правил, устанавливающих последовательность обслуживания телефонных вызовов коммутационной станцией или узлом.

Обслуживание может осуществляться в случайном порядке, в порядке очередности поступления, с приоритетом.

К характеристикам качества обслуживания поступающего потока сообщений относятся вероятность явной или условной потерь со-

общений из-за отсутствия свободных приборов или путей установления соединения, среднее время задержки начала обслуживания сообщения, вероятность потери поступившего вызова, интенсивность обслуженной нагрузки, пропускная способность системы обслуживания, производительность системы управления.

Вопросы для самоконтроля

1. В чем состоят цель, объект и предмет теории телетрафика?
2. Каково отличие теории телетрафика от теории массового обслуживания?
3. Какими условиями определяется качество обслуживания абонента на сети?
4. Что понимают под пропускной способностью в теории телетрафика?
5. Перечислите основные задачи, решаемые в теории телетрафика.
6. Обоснуйте математическую модель системы распределения информации.
7. Раскройте понятия, используемые в теории телетрафика: заявка, вызов, поток телефонных вызовов, дисциплина обслуживания потоков вызовов, способ обслуживания телефонных вызовов, порядок обслуживания телефонных вызовов.

4. МОДЕЛИ ПОТОКОВ ВЫЗЫВОВ

4.1. Определение потока вызовов. Свойства потоков вызовов

Введем некоторые ограничения на рассматриваемые случайные потоки вызовов. Необходимость этого связана с тем, что потоки вызовов разделяются на неоднородные и однородные.

В неоднородном потоке вызовов каждый вызов имеет две и более характеристики. Например вызовы, поступающие от абонентов телефонной сети, определяются моментами их поступления, направлениями установления соединений, длительностью их обслуживания и другими характеристиками.

Однородный поток вызовов характеризуется последовательностью, определяющей только закономерность поступления вызовов, т. е. последовательностью моментов их поступления или промежутков времени между ними, либо иным способом задания потока вызовов. На практике потоки вызовов, как правило, являются неоднородными. Несмотря на это, целесообразно последовательность моментов их поступления изучить отдельно от других их характеристик, поэтому в дальнейшем под потоком вызовов будем понимать однородный поток вызовов.

Введем еще одно естественное ограничение: на любом конечном отрезке времени поступает конечное число вызовов. Такие потоки называются конечными (финитными), в отличие от взрывных потоков, т. е. таких, для которых на конечном интервале времени с положительной вероятностью может поступить бесконечное число вызовов (процесс деления атомных ядер).

Следует также различать *детерминированные* и *случайные* потоки вызовов.

Детерминированный поток вызовов – это последовательность вызовов, в которой они поступают в определенные, строго фиксированные, неслучайные промежутки времени. Детерминированные потоки, следующие через определенные, равные промежутки времени, называются регулярными.

Детерминированный поток можно задать (определить) следующими эквивалентными способами:

– последовательностью моментов поступления вызовов $[t_1, t_2, t_3, \dots, t_n]$, $0 \leq t_1 \leq t_2 \leq t_3 \leq \dots \leq t_{k-1} \leq t_k \leq \dots$, где k – порядковый

номер вызова. Равенство $t_{k-1} = t_k$ означает одновременное поступление $(k-1)$ и k -го вызова;

- последовательностью длин промежутков между вызовами $[\tau_1, \tau_2, \tau_3, \dots, \tau_n]$, $\tau_k = t_k - t_{k-1}$, $(k = 1, 2, 3, \dots, t_0 = 0)$; величину τ_k называют промежутком, предшествующим поступлению k -го вызова;

- последовательностью чисел $k_1, k_2, k_3, \dots, k_n$, определяющих количество вызовов, поступающих в течение заданных отрезков времени $[t_0, t_1], [t_0, t_2], \dots, [t_0, t_n]$;

- целочисленной неотрицательной и неубывающей функцией $x(t)$, характеризующей число вызовов (рис. 4.1), поступающих в течение заданных промежутков времени $[t_0, t_1], [t_0, t_2], \dots, [t_0, t_n]$.

Точки скачков функции $x(t)$ соответствуют вызывающим моментам, а величины скачков – числу вызовов, поступивших в вызывающий момент (рис. 4.1).

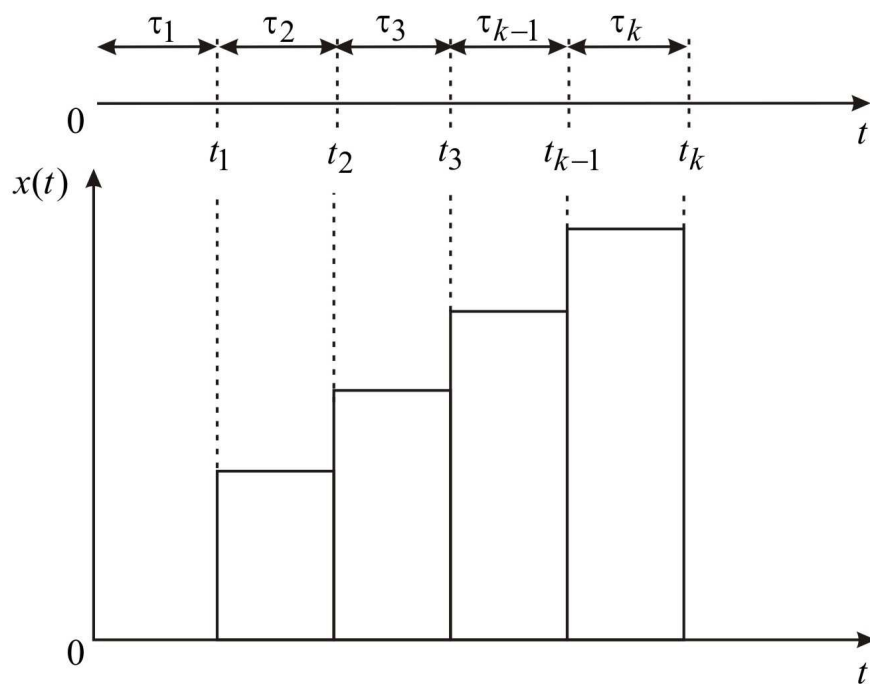


Рис. 4.1. Способы задания детерминированного потока

Вызывающим называется момент времени, в который поступает одновременно 1, 2, 3 и более вызовов. Для вызывающих моментов всегда $t_{k-1} \leq t_k$ и $\tau_k > 0$ (рис. 4.1).

Детерминированные потоки являются частным случаем случайных потоков и на практике встречаются редко. Примерами их могут служить поток сеансов связи с искусственными спутниками Земли

либо сеанс космической связи, телевизионные и радиовещательные программы, поток поступления деталей и выхода изделий ритмично работающего конвейерного производства и т. п. Однако и здесь возможны случайные, непредвиденные изменения расписания.

На практике, в основном, приходится иметь дело со случайными потоками вызовов, поэтому в дальнейшем будут рассмотрены только случайные потоки вызовов.

Случайный поток вызовов отличается от детерминированного тем и только тем, что моменты поступления вызовов и промежутки времени между ними являются не строго фиксированными, а случайными величинами.

Определение случайного потока вызовов отличается от определения детерминированного потока тем и только тем, что все величины, задающие детерминированный поток, являются случайными.

Для задания случайных потоков по аналогии используют вероятностные законы распределения случайных величин, в частности функции распределения моментов поступления вызовов t_k , промежутков между вызовами τ_k и числа вызовов в промежутке $(0, t)$.

Для задания случайного потока вызовов могут быть использованы следующие эквивалентные способы:

– совместный закон распределения n случайных вызывающих моментов:

$$P\{T_i < t_i, \quad i = 1, 2, \dots, n\} = P\{T_1 < t_1, T_2 < t_2, \dots, T_n < t_n\}, \quad (4.1)$$

где T_i – i -й вызывающий момент, n может принимать любое значение;

– совместный закон распределения n случайных промежутков времени между вызывающими моментами:

$$P\{T_i^* < \tau_i, \quad i = 1, 2, \dots, n\} = P\{T_1^* < \tau_1, T_2^* < \tau_2, \dots, T_n^* < \tau_n\}, \quad (4.2)$$

где τ_i – промежуток времени между $(i-1)$ и i -м вызывающими моментами, n может принимать любое значение;

– совместный закон распределения числа вызовов K на n отрезках времени $[t_0, t_1], [t_0, t_2], \dots, [t_0, t_n]$:

$$\begin{aligned} P\{K(t_0, t_i) = k_i, \quad i = 1, 2, \dots, n\} = \\ = P\{K(t_0, t_1) = k_1, K(t_0, t_2) = k_2, \dots, K(t_0, t_n) = k_n\}, \end{aligned} \quad (4.3)$$

где n может принимать любое значение;

$$k_1 < k_2 < \dots < k_n; t_1 < t_2 < \dots < t_n.$$

Случайные потоки вызовов классифицируются в зависимости от наличия или отсутствия следующих трех свойств:

- стационарности;
- ординарности;
- последствия.

Поток вызовов называется *стационарным*, если вероятность попадания того или иного числа вызовов на участок времени длиной τ зависит от длины этого промежутка и не зависит от того, где именно на оси времени $[0, t]$ расположен этот участок.

Стационарность означает, что с течением времени вероятностные характеристики потока не меняются. Стационарность потока равносильна постоянной плотности вероятности поступления вызовов в любой момент времени, иначе говоря, для стационарного потока вероятность поступления i вызовов за промежуток длиной t зависит от величины промежутка и не зависит от его расположения на оси времени.

Это не означает, что фактически число вызовов, появляющееся в единицу времени, постоянно. Поток неизбежно (если только он не регулярный), имеет какие-то случайные сгущения и разрежения. Важно, что для стационарного потока эти сгущения и разрежения не носят закономерного характера: на один участок длины l может попасть больше, а на другой – меньше событий, но среднее число событий, приходящееся на единицу времени, постоянно и от времени не зависит.

Любой стационарный поток можно задать семейством условных вероятностей $P_i(t)$ поступления i ($i = 0, 1, 2, 3, \dots$) вызовов в промежутке t , если в начальный его интервал произошел вызов.

Реально поступающий на АТС поток вызовов имеет нестационарный характер. Эти отклонения от стационарности имеют физические причины. Так, совершенно естественно, что интенсивность потока вызовов (число вызовов в единицу времени), поступающих на АТС ночью, меньше, чем их количество днем. Однако в течение суток (как правило, между 11 – 13 ч) можно выделить одночасовые промежутки времени, в течение которых поступающий поток вызовов близок к стационарному. Свойство стационарности потока удоб-

но для проведения исследования, поэтому основные модели систем распределения информации разработаны для случая стационарного потока.

Следующим свойством, используемым для классификации потока вызовов, является последствие.

Последствие означает зависимость вероятностных характеристик потока от предыдущих событий, иначе говоря, вероятность поступления вызовов в промежутке (t_1, t_2) зависит от числа, времени поступления и длительности обслуживания вызовов до момента t_1 .

Если исследуемый поток вызовов обладает данным свойством, то при моделировании СК приходится оперировать условными вероятностями, так как для определения числа вызовов на следующем временном интервале следует учитывать их число на предшествующем интервале времени.

Условие отсутствия последствия – наиболее существенное свойство. Отсутствие последствия означает, что заявки поступают в систему независимо друг от друга.

Примером потока без последствия может служить поток телефонных вызовов от большой группы источников. Е. Вентцель в своей работе «Исследование операций» поясняет, что поток заявок на городскую РАТС емкостью 10 тыс. номеров близок к потоку без последствия, так как причины, обусловившие приход отдельного вызова именно в тот, а не в другой момент времени, как правило, не связан с аналогичными причинами для других абонентов (поток пассажиров в метро).

Поток от малой группы, наоборот, обладает заметным последствием. Так, при емкости группы источников $N = 3$ вероятность поступления вызовов существенно зависит от числа свободных источников $N_{св}$ и будет заметно больше при $N_{св} = 3$, чем при $N_{св} = 2$.

В данном случае число свободных источников зависит от предыдущих событий, что и определяет последствие потока. С ростом емкости группы постепенно уменьшается доля занятых источников по отношению к общему их числу, соответственно ослабевает последствие потока и при каком-то предельном значении N его уже можно не учитывать.

Одновременно поток от ограниченной группы не является стационарным, так как вероятность поступления вызова зависит от рассматриваемого участка на оси времени через состояние системы, т. е. в разные моменты времени состояние системы различно.

Поток повторных вызовов также является примером потока с последствием, поскольку повторный вызов возникает как результат потери предыдущего вызова, т. е. налицо зависимость от предыдущих событий.

Для случайного потока без последствия условная вероятность поступления вызовов в промежутке (t_1, t_2) , вычисленная при любых предположениях о течении процесса обслуживания вызовов до момента t_1 , равна безусловной, что значительно упрощает процесс исследования данных потоков.

Стационарный поток без последствия соответственно можно задать семейством вероятностей $P_i(t)$ поступления i вызовов в любом промежутке длиной t .

Потоки с последствием подразделяются на два вида:

- с простым последствием;
- ограниченным последствием.

Их особенности будут рассмотрены ниже.

Третьим свойством, используемым для классификации потоков, является свойство ординарности.

Ординарность означает практическую невозможность группового поступления вызовов, иначе говоря, вероятность поступления двух и более вызовов за любой бесконечно малый промежуток времени Δt есть величина бесконечно малая более высокого порядка, чем Δt , т. е. $P_{i>2} = 0(\Delta t)$. Например поток поездов, подходящих к станции, как правило, ординарен, а поток вагонов – неординарен, поток клиентов к зубному врачу – ординарен, а поток в ЗАГС – неординарен. В сетях связи потоки вызовов, как правило, ординарны. Вместе с тем можно привести пример неординарных потоков. Так, во время ответов Президента РФ на вопросы граждан страны 23.12.2002 г. в центр обработки вопросов поступало 83 вызова в секунду, т. е. если в качестве единицы времени взять 1 с, то в вызывающий момент поступило 83 вызова.

При обработке телеграмм оператором возможна неординарность потока, если предполагается поступление телеграмм партиями.

В чем удобство ординарных потоков? Если поток ординарен, то вероятностью попадания на малый участок времени Δt двух или более событий можно пренебречь!

4.2. Характеристики потоков вызовов

К основным характеристикам случайных потоков относят ведущую функцию, параметр, интенсивность, вероятность $P_i(t, t + \Delta t)$ поступления ровно i заявок за интервал Δt , начиная с момента времени t , функцию распределения интервала времени между двумя соседними заявками.

Ведущая функция случайного потока $m_i(0, t)$ – есть математическое ожидание числа вызовов в промежутке $[0, t)$ (рис. 4.2). Функция $m_i(0, t)$ – неотрицательная, неубывающая, в практических задачах теории телетрафика непрерывна и принимает только конечные значения.

Потоки с непрерывной ведущей функцией называют регулярными, ступенчатыми, то есть сингулярными.

Параметр потока $\lambda(t)$ в момент времени t есть предел отношения вероятности поступления не менее одного вызова в промежутке $[t, t + \Delta t)$ к величине этого промежутка Δt при $\Delta t \rightarrow 0$:

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P_{i \geq 1}(t, t + \Delta t)}{\Delta t}. \quad (4.4)$$

Исходя из формулы (4.4), можно найти вероятность поступления одного и более вызовов за время $[t, t + \Delta t)$:

$$P_{i \geq 1}(t, t + \Delta t) = \lambda(t) \cdot \Delta t + o(\Delta t), \quad \Delta t \rightarrow 0. \quad (4.5)$$

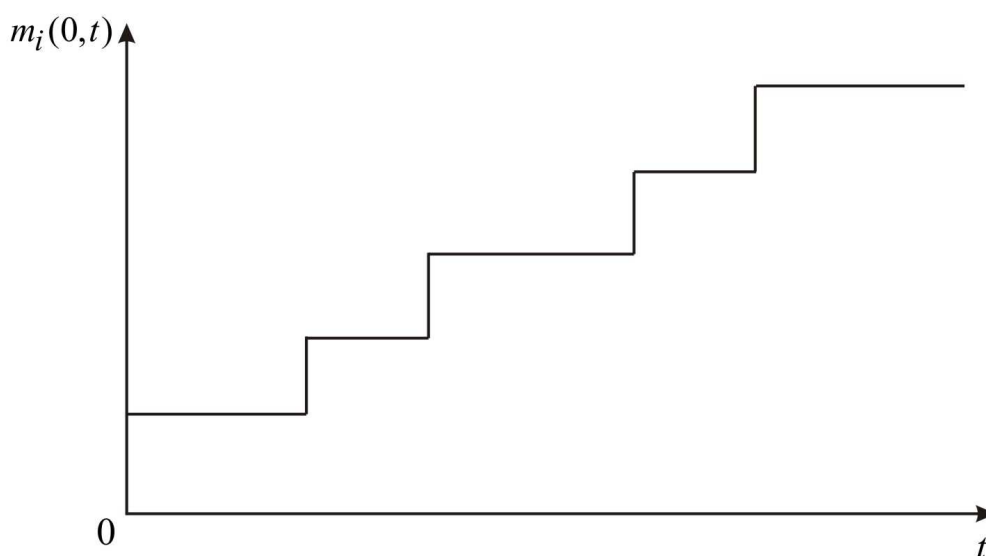


Рис. 4.2. Ведущая функция

Для стационарных потоков вероятность поступления определенного числа вызовов за промежутки времени Δt зависит только от длительности этого промежутка:

$$P_{i \geq 1}(t, t + \Delta t) = P_{i \geq 1}(\Delta t). \quad (4.6)$$

Следовательно, и плотность вероятности поступления вызовов стационарного потока (параметр) есть величина постоянная, не зависящая от момента t , т. е. $\lambda(t) = \lambda$. Соответственно получаем

$$P_{i \geq 1}(\Delta t) = \lambda \cdot \Delta t + o(\Delta t). \quad (4.7)$$

Ведущая функция дает среднее число вызовов на отрезке $[0, t)$, в то время как параметр $\lambda(t)$, в отличие от ведущей функции, во-первых, характеризует не столько поток вызовов, сколько поток вызывающих моментов, во-вторых, относится при фиксированном t не ко всему отрезку $[0, t)$, а лишь к моменту t .

Интенсивность стационарного потока c есть математическое ожидание числа вызовов в единицу времени. Единица времени может быть принята произвольно, однако в теории телетрафика в качестве такой единицы большей частью принимают среднюю длительность одного занятия.

Вследствие аддитивности математического ожидания для стационарного потока ведущая функция за промежуток $[0, t)$

$$m_i(0, t) = c \cdot t. \quad (4.8)$$

Для нестационарных потоков используются понятия средняя и мгновенная интенсивность.

Средняя интенсивность в промежутке $[t_1, t_2)$ есть математическое ожидание числа вызовов в единицу времени. Среднюю интенсивность можно выразить через ведущую функцию

$$c(t_1, t_2) = \frac{m_i(0, t_2) - m_i(0, t_1)}{t_2 - t_1}. \quad (4.9)$$

Мгновенная интенсивность потока $c(t)$ в момент t есть производная ведущей функции потока по t :

$$c(t) = \lim_{\Delta t \rightarrow 0} \frac{m_i(0, t + \Delta t) - m_i(0, t)}{\Delta t} = m_i'(0, t). \quad (4.10)$$

Из выражения (4.10) видно, что мгновенная интенсивность представляет собой производную ведущей функции потока. Как и параметр потока, мгновенная интенсивность потока относится не к отрезку времени поступления вызовов, а только к моменту t . В то же время, в отличие от параметра потока, характеризующего поток вызывающих моментов, мгновенная интенсивность потока характеризует поток поступления вызовов.

Поэтому всегда справедливо неравенство $c(t) \geq \lambda(t)$.

Для стационарных потоков $c(t) = c$, $\lambda(t) = \lambda$, следовательно, для любых стационарных потоков $c \geq \lambda$, а для стационарных и ординарных – $c = \lambda$.

4.3. Принципы классификации потоков вызовов

Все потоки вызовов можно классифицировать по наличию или отсутствию трех вышерассмотренных свойств: стационарности, ординарности, последействия. Однако на практике для классификации потоков используют лишь свойство последействия.

Использование признака последствия позволяет распределить все изучаемые модели по трем классам:

- потоки без последействия;
- с простым последействием;
- с ограниченным последействием.

В класс потоков без последействия входят:

- простейший;
- пуассоновский с переменным или случайным параметром;
- неординарный пуассоновский;
- пуассоновский с неординарными вызовами.

К потокам с простым последействием относятся:

- примитивный (поток Энгсета);
- сглаженный;
- с повторными вызовами;
- поток освобождений.

Ограниченным последействием обладают:

- рекуррентный поток;
- поток Пальма;
- потоки Эрланга.

4.4. Математические модели потоков вызовов без последствия

4.4.1. Модель простейшего потока вызовов. Свойства и характеристики

Простейшим [20] называется поток телефонных вызовов, который одновременно является стационарным, ординарным и без последствия. Простейший поток вызовов является наиболее распространенной моделью реального потока вызовов, применяемой в системах массового обслуживания, в том числе в теории телетрафика. Как правило, реальный поток вызовов от большой группы источников нагрузки аппроксимируют моделью простейшего потока вызовов.

Причина в том, что поток вызовов от числа источников $N > 100$ (300 – 500) и малой интенсивности потока вызовов от одного источника $\alpha \rightarrow 0$ характеризуется отсутствием последствия. Действительно, вероятность поступления вызова после момента времени t не зависит от того, сколько вызовов поступило до момента времени t , так как оставшаяся часть свободных источников нагрузки существенно больше занятых разговором абонентов.

Кроме того, реальные потоки, обычно, ординарны, так как за бесконечно малый промежуток времени вероятность поступления двух и более вызовов стремится к нулю. Случаи, когда потоки могут быть неординарны, имеют место, но они достаточно редки.

При ограничении исследуемого промежутка времени (1 – 3 ч) данный поток можно считать стационарным, так как вероятность поступления i вызовов за промежуток длиной t зависит только от длины промежутка и не зависит от его расположения на оси времени.

Название «простейший» означает, что процессы, связанные с данными потоками, имеют наиболее простое математическое описание. Самый простой, на первый взгляд, регулярный поток не является "простейшим", так как обладает последствием: моменты появления событий в таком потоке связаны жесткой функциональной зависимостью. Без специальных усилий по поддержанию его регулярности такой поток обычно не создается.

Как известно, математическое описание простейшего потока событий дано французским физиком и математиком С. Д. Пуассоном задолго до рождения основоположника теории телетрафика А.К. Эрланга. Заслуга Эрланга состояла в том, что он впервые начал применять теорию вероятностей для решения проблем телефонного трафика. С его легкой руки в настоящее время распределение Пуассона используется в качестве математической модели простейшего потока вызовов.

Как было уже отмечено, любой случайный стационарный поток, в том числе и простейший, задается семейством вероятностей $P_i(t)$ поступления i ($i = \overline{0, \infty}$) вызовов в промежутке t .

Для определения функции $P_i(t)$ исследуем процесс поступления i вызовов в течение двух соседних произвольно расположенных на оси времени промежутков, которые для простоты обозначим один через t , другой бесконечно малый промежуток времени — через Δt (рис. 4.3).

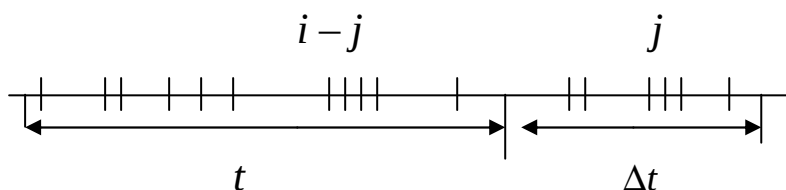


Рис. 4.3. Размещение промежутков на оси времени

Размещение i вызовов внутри промежутков $t + \Delta t$ возможно $i+1$ способами:

- в промежутке времени t поступает i вызовов, а в промежутке $\Delta t - 0$ вызовов (обозначим вероятность данного события как $P(i, t; 0, \Delta t)$);
- в промежутке времени t поступает $i-1$ вызовов, а в промежутке $\Delta t - 1$ вызовов ($P(i-1, t; 1, \Delta t)$);
- в промежутке времени t поступает $i-2$ вызова, а в промежутке $\Delta t - 2$ вызова ($P(i-2, t; 2, \Delta t)$) и т. д., вплоть до того, что в промежутке времени t не поступит ни одного вызова, а в промежутке $\Delta t -$ все i вызовов (вероятность — $P(0, t; i, \Delta t)$).

Поскольку указанные события несовместны, следовательно, вероятности их суммируются (формула полной вероятности):

$$P_i(t + \Delta t) = P(i, t; 0, \Delta t) + P(i-1, t; 1, \Delta t) + P(i-2, t; 2, \Delta t) + \dots + \\ + P(1, t; i-1, \Delta t) + P(0, t; i, \Delta t) = \sum_{j=0}^i P((i-j), t; j, \Delta t), \quad (4.11)$$

где $P(i-j, t; j, \Delta t)$ – вероятность следующего совместного события: в промежутке t поступает $(i-j)$ вызовов, а в промежутке Δt – j вызовов.

По причине отсутствия последействия в потоке вероятность поступления j вызовов в промежутке Δt не зависит от числа поступивших $(i-j)$ вызовов в промежутке t , поэтому вероятность $P((i-j), t; j, \Delta t)$ можно представить как произведение двух безусловных вероятностей $P_{i-j}(t)$ и $P_j(\Delta t)$, поскольку

$$P((i-j), t; j, \Delta t) = P_{i-j}(t) \cdot P_j(\Delta t).$$

Отсюда выражение (4.11) можно записать в виде

$$P_i(t + \Delta t) = \sum_{j=0}^i P_{i-j}(t) \cdot P_j(\Delta t). \quad (4.12)$$

Величина i при числе абонентов $N > 100$ может принимать большие значения, что усложняет решение уравнения (4.12). Однако его можно значительно упростить (уменьшить число слагаемых), если воспользоваться свойством ординарности простейшего потока вызовов, согласно которому вероятность поступления за бесконечно малый интервал времени Δt более одного вызова равна нулю, $P_{j>1}(\Delta t) = o(\Delta t)$. Отсюда выражение (4.12) можно записать в виде

$$P_i(t + \Delta t) = \sum_{j=0}^i P_{i-j}(t) \cdot P_j(\Delta t) = \\ = P_i(t) \cdot P_0(\Delta t) + P_{i-1}(t) \cdot P_1(\Delta t) + o(\Delta t), \quad (4.13)$$

где $o(\Delta t)$ – величина более высокого порядка малости по отношению к Δt .

Вероятность поступления одного вызова $P_1(\Delta t)$ за бесконечно малый интервал времени определим, используя формулу параметра потока для стационарного потока при условии, что $\lambda(t) = \lambda$:

$$\begin{aligned}\lambda(t) = \lambda &= \lim_{\Delta t \rightarrow 0} \frac{P_{i \geq 1}(t, t + \Delta t)}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{P_{i=1}(t, t + \Delta t)}{\Delta t} = \\ &= \lim_{\Delta t \rightarrow 0} \frac{P_{i=1}(\Delta t)}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{P_1(\Delta t)}{\Delta t}.\end{aligned}$$

Отсюда

$$P_1(\Delta t) = P_{i \geq 1}(\Delta t) + o(\Delta t) = \lambda \cdot (\Delta t) + o(\Delta t). \quad (4.14)$$

Вероятность непоступления ни одного вызова $P_0(\Delta t)$ за интервал Δt находим из условия нормировки (за бесконечно малый интервал времени в ординарном потоке может поступить либо один вызов, либо не поступит ни одного вызова, при этом сумма вероятностей этих двух событий равна единице):

$$P_0(\Delta t) = 1 - P_1(\Delta t) - P_{i \geq 2}(\Delta t) = 1 - \lambda \cdot \Delta t + o(\Delta t). \quad (4.15)$$

В формуле (4.15) вероятности поступления более одного вызова бесконечно малы: $P_{i \geq 2}(\Delta t) = o(\Delta t)$.

Подставляя выражения (4.14) и (4.15) в систему (4.13), получим

$$\begin{aligned}P_i(t + \Delta t) &= \sum_{j=0}^i P_{i-j}(t) \cdot P_j(\Delta t) = \\ &= P_i(t) \cdot (1 - \lambda \cdot \Delta t + o(\Delta t)) + P_{i-1}(t) \cdot \lambda \cdot \Delta t + o(\Delta t) = \\ &= P_i(t) - P_i(t) \cdot \lambda \cdot \Delta t + P_{i-1}(t) \cdot \lambda \cdot \Delta t + o(\Delta t).\end{aligned} \quad (4.16)$$

Перенесем величину $P_i(t)$ выражения (4.16) в левую его часть и разделим и левую, и правую части уравнения на Δt :

$$\frac{P_i(t + \Delta t) - P_i(t)}{\Delta t} = -\lambda \cdot P_i(t) + \lambda \cdot P_{i-1}(t) + \frac{o(\Delta t)}{\Delta t}. \quad (4.17)$$

Далее перейдем к пределу при $\Delta t \rightarrow 0$:

$$\lim_{\Delta t \rightarrow 0} \frac{P_i(t + \Delta t) - P_i(t)}{\Delta t} = P'_i(t) = -\lambda \cdot P_i(t) + \lambda \cdot P_{i-1}(t). \quad (4.18)$$

При этом $\lim_{\Delta t \rightarrow 0} \frac{o(\Delta t)}{\Delta t} = 0$.

Начальными условиями для системы (4.18) являются:

$$P_0(0) = 1; P_i(0) = 0, i = \overline{1, \infty}. \quad (4.19)$$

Решением системы (4.17) с учетом условий (4.19) служит выражение, которое носит название формулы Пуассона:

$$P_i(t) = \frac{(\lambda \cdot t)^i}{i!} \cdot e^{-\lambda \cdot t} \quad (i = 1, 2, 3, \dots). \quad (4.20)$$

Выражение (4.20) определяет вероятность поступления ровно i вызовов простейшего потока за отрезок времени t и является одним из возможных способов задания простейшего потока вызовов.

В зарубежной литературе простейший поток часто называют пуассоновским, хотя в отечественной литературе под пуассоновским потоком понимают более широкий класс, объединяющий все потоки без последствия.

Простейший поток также называют потоком чистой случайности, поскольку поступление вызовов не зависит ни от времени, ни от предыдущих событий и действительно определяется только случаем.

Функция $P_i(t)$ есть функция распределения случайной величины i . Из выражения (4.20) следует, что она зависит от $\lambda \cdot t$ и i . Проанализируем распределение Пуассона. Для этого рассмотрим отношение двух соседних вероятностей:

$$\frac{P_i(t)}{P_{i-1}(t)} = \frac{[(\lambda \cdot t)^i / i!] \cdot e^{-\lambda \cdot t}}{[(\lambda \cdot t)^{i-1} / (i-1)!] \cdot e^{-\lambda \cdot t}} = \frac{\lambda \cdot t}{i}.$$

При $i \leq \lambda \cdot t$ имеет место $P_{i-1}(t) \leq P_i(t)$, а при $i \geq \lambda \cdot t$ – обратное соотношение: $P_{i-1}(t) \geq P_i(t)$. Следовательно, с ростом i вероятность $P_i(t)$ растет, пока сохраняется условие $i < \lambda \cdot t$, и начинает уменьшаться при $i > \lambda \cdot t$. Максимум функции $P_i(t)$ достигает при целом значении $\lambda \cdot t$ в двух точках: $i = \lambda \cdot t$ и $i = \lambda \cdot t - 1$, а при дробном значении $\lambda \cdot t$ – в одной точке $i = \lfloor \lambda \cdot t \rfloor$, где действие в скобках $\lfloor \ \rfloor$ означает целую часть числа $\lambda \cdot t$.

С ростом $\lambda \cdot t$ огибающая кривая принимает все более симметричный вид, приближаясь к нормальному закону распределения не-

прерывной случайной величины. Уже при $\lambda \cdot t = 10$ совпадение огибающей с нормальным законом достаточно хорошее (рис. 4.4).

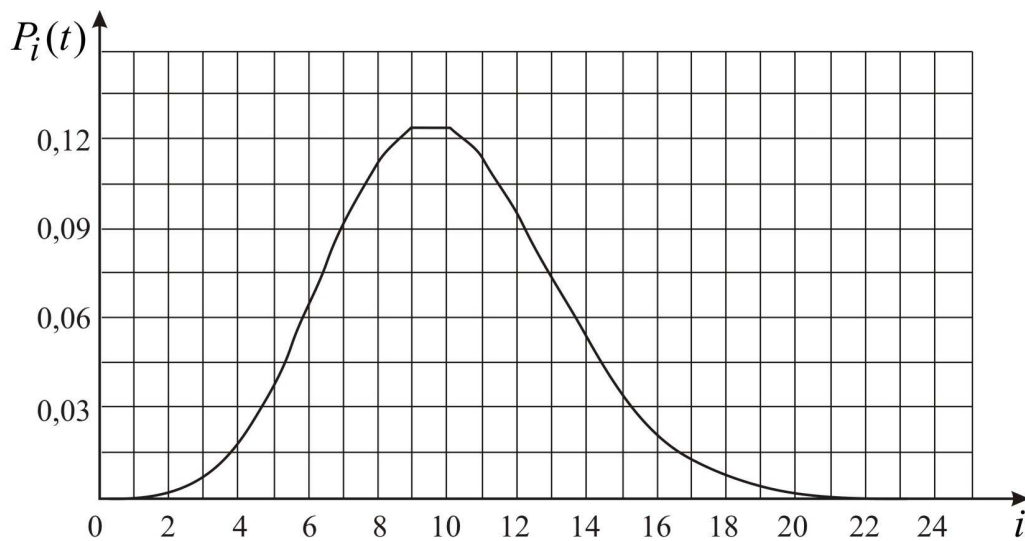


Рис. 4.4. Распределение $P_i(t)$ от i при значении $\lambda \cdot t = 10$

Наряду с распределением Пуассона $P_i(t)$ (4.20) при решении практических задач используют суммарные вероятности, определяющие поступление не менее i вызовов за промежуток t :

$$P_{k \geq i}(t) = \sum_{k=i}^{\infty} P_k(t). \quad (4.21)$$

По аналогии с формулой (4.21) вероятность поступления не более i вызовов за промежуток t определяется следующим образом:

$$P_{k \leq i}(t) = \sum_{k=0}^i P_k(t). \quad (4.22)$$

Кроме выражения (4.20), существует и другой способ задания простейшего потока в виде закона распределения промежутков между соседними вызовами:

$$P(\tau < t) = 1 - P(\tau \geq t). \quad (4.23)$$

В формуле (4.23) вероятность $P(\tau \geq t)$ трактуется как вероятность того, что промежуток между вызовами (τ) больше рассматриваемого промежутка времени (t), а это равносильно тому, что внутри

промежутка длиной t не поступит ни одного вызова, следовательно, можно записать, что

$$P(\tau \geq t) = P_0(t). \quad (4.24)$$

Тогда, используя формулу (4.20), выражение (4.23) можно записать в виде

$$\begin{aligned} P(\tau < t) &= 1 - P(\tau \geq t) = 1 - P_0(t) = \\ &= 1 - \frac{(\lambda \cdot t)^0}{0!} \cdot e^{-\lambda \cdot t} = 1 - e^{-\lambda \cdot t}. \end{aligned} \quad (4.25)$$

Дифференцируя выражение (4.25) по t , получаем плотность распределения

$$f(t) = \lambda \cdot e^{-\lambda \cdot t}. \quad (4.26)$$

Закон распределения с плотностью (4.26) называется экспоненциальным, или показательным, а величина λ – его параметром (рис. 4.4). Распределение (4.26) является не только необходимым, но и достаточным условием существования простейшего потока вызовов.

Характеристики простейшего потока вызовов

1. Для простейшего потока вызовов параметр потока есть величина постоянная, не зависящая от момента времени t :

$$\lambda(t) = \lambda. \quad (4.27)$$

2. Для простейшего потока характерно совпадение величин математического ожидания m_t и среднеквадратического отклонения σ_t промежутка времени между соседними вызовами.

Докажем это утверждение, определив математическое ожидание m_t , дисперсию D_t и среднеквадратическое отклонение σ_t промежутка времени между соседними вызовами:

$$m_t = \int_0^{\infty} t \cdot f(t) \cdot dt. \quad (4.28)$$

Подставляя в выражение (4.28) формулу (4.26) для $f(t)$ и используя правило интегрирования по частям [4]

$$(x \cdot e^x \cdot dx = x \cdot de^x; \quad x \cdot de^x = x \cdot e^x - \int e^x \cdot dx),$$

получим

$$\begin{aligned} m_t &= \int_0^{\infty} t \cdot f(t) \cdot dt = \int_0^{\infty} t \cdot \lambda \cdot e^{-\lambda \cdot t} \cdot dt = \int_0^{\infty} \lambda \cdot t \cdot d\left(-\frac{1}{\lambda} \cdot e^{-\lambda t}\right) = \\ &= \int_0^{\infty} \lambda \cdot \left(-\frac{1}{\lambda}\right) \cdot t \cdot d(e^{-\lambda t}) = \\ &= -t \cdot e^{-\lambda \cdot t} \Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda \cdot t} \cdot dt = \left[\left(-\frac{1}{\lambda}\right) \cdot e^{-\infty} - \left(-\frac{1}{\lambda}\right) \cdot e^0 \right] = \frac{1}{\lambda}. \end{aligned} \quad (4.29)$$

Аналогично находим дисперсию

$$D_t = \int_0^{\infty} t^2 \cdot \lambda \cdot e^{-\lambda \cdot t} \cdot dt - M_t^2 = \int_0^{\infty} t^2 \cdot \lambda \cdot e^{-\lambda \cdot t} \cdot dt - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}. \quad (4.30)$$

Среднеквадратическое отклонение составляет

$$\sigma_t = \sqrt{D_t} = \sqrt{\frac{1}{\lambda^2}} = \frac{1}{\lambda}. \quad (4.31)$$

Совпадение величин m_t и σ_t на практике используют как критерий первоначальной проверки соответствия гипотезы о показательном распределении длительности между соседними вызовами полученных статистических данных.

Показательное распределение широко применяется в теории телетрафика благодаря замечательному свойству: если известно, что случайный промежуток времени между соседними вызовами, распределенный по показательному закону, длился уже некоторое время t , то закон распределения оставшейся части промежутка также будет показательным (с тем же параметром) и не будет зависеть от t . Это свойство показательного распределения позволяет значительно упростить математические преобразования при анализе процессов поступления вызовов и их обслуживании.

3. Для простейшего потока характерно совпадение величин математического ожидания $m_i(t)$ и дисперсии $D_i(t)$ числа вызовов за промежуток времени t ($m_i(t) = D_i(t) = \lambda \cdot t$).

Докажем это утверждение, определив математическое ожидание $m_i(t)$ и дисперсию $D_i(t)$ дискретной случайной величины i :

$$\begin{aligned} m_i(t) &= \sum_{i=1}^{\infty} i \cdot P_i(t) = \sum_{i=1}^{\infty} i \cdot \frac{(\lambda \cdot t)^i}{i!} \cdot e^{-\lambda \cdot t} = e^{-\lambda \cdot t} \cdot (\lambda \cdot t) \cdot \sum_{i=1}^{\infty} \frac{(\lambda \cdot t)^{i-1}}{(i-1)!} = \\ &= e^{-\lambda \cdot t} \cdot (\lambda \cdot t) \cdot \sum_{k=0}^{\infty} \frac{(\lambda \cdot t)^k}{k!} = e^{-\lambda \cdot t} \cdot (\lambda \cdot t) \cdot e^{\lambda \cdot t} = \lambda \cdot t. \end{aligned} \quad (4.32)$$

$$\begin{aligned} D_i(t) &= \sum_{i=1}^{\infty} i^2 \cdot P_i(t) - M_i^2 = \sum_{i=1}^{\infty} i^2 \cdot \frac{(\lambda \cdot t)^i}{i!} \cdot e^{-\lambda \cdot t} - (\lambda \cdot t)^2 = \\ &= e^{-\lambda \cdot t} \cdot \lambda \cdot t \cdot \sum_{i=1}^{\infty} i \cdot \frac{(\lambda \cdot t)^{i-1}}{(i-1)!} - (\lambda \cdot t)^2 = \\ &= e^{-\lambda \cdot t} \cdot \lambda \cdot t \cdot \sum_{i=1}^{\infty} [1 + (i-1)] \cdot \frac{(\lambda \cdot t)^{i-1}}{(i-1)!} - (\lambda \cdot t)^2 = \\ &= e^{-\lambda \cdot t} \cdot \lambda \cdot t \cdot \left[\sum_{i=1}^{\infty} \frac{(\lambda \cdot t)^{i-1}}{(i-1)!} + \sum_{i=1}^{\infty} (i-1) \cdot \frac{(\lambda \cdot t)^{i-1}}{(i-1)!} \right] - (\lambda \cdot t)^2 = \\ &= e^{-\lambda \cdot t} \cdot \lambda \cdot t \cdot \sum_{k=0}^{\infty} \frac{(\lambda \cdot t)^k}{k!} + e^{-\lambda \cdot t} \cdot \lambda \cdot t \cdot \lambda \cdot t \cdot \sum_{i=1}^{\infty} \frac{(\lambda \cdot t)^{i-2}}{(i-2)!} - (\lambda \cdot t)^2 = \\ &= e^{-\lambda \cdot t} \cdot \lambda \cdot t \cdot e^{\lambda \cdot t} + e^{-\lambda \cdot t} \cdot (\lambda \cdot t)^2 \cdot \sum_{m=0}^{\infty} \frac{(\lambda \cdot t)^m}{m!} - (\lambda \cdot t)^2 = \\ &= \lambda \cdot t + (\lambda \cdot t)^2 - (\lambda \cdot t)^2 = \lambda \cdot t. \end{aligned} \quad (4.33)$$

Совпадение математического ожидания и дисперсии числа вызовов за промежуток t для простейшего потока удобно использовать на практике при проверке соответствия реального потока вызовов модели простейшему. Так, получив ряд чисел $x_1, x_2, x_3, x_4, \dots, x_n$, ха-

рактизирующих число вызовов, поступающих в n промежутков длиной t (как правило, за t принимают величину, равную 15 мин), далее рассчитывают среднее значение и несмещенную оценку дисперсии величины x :

$$\bar{x} = \sum_{j=1}^n x_j / n;$$

$$\tilde{D}_x = \frac{\sum_{j=1}^n (x_j - \bar{x})^2}{n-1}. \quad (4.34)$$

4. Объединение произвольного числа независимых простейших потоков с параметрами λ_i образует общий простейший поток с параметром $\lambda_{\Sigma} = \sum_i \lambda_i$. Вероятность поступления ровно k вызовов за отрезок времени t определяется по формуле Пуассона

$$P_k(t) = \frac{[(\lambda_1 + \lambda_2 + \dots + \lambda_n) \cdot t]^k}{k!} \cdot e^{-(\lambda_1 + \lambda_2 + \dots + \lambda_n) \cdot t}. \quad (4.35)$$

При разъединении поступающего простейшего потока с параметром λ на n направлений так, что каждый вызов исходного потока с вероятностью P_l ($\sum_{l=1}^n P_l = 1$) поступает на l -е направление, поток l -го направления также будет простейшим с параметром $\lambda \cdot P_l$.

5. Так как простейший поток является стационарным и ординарным, то параметр потока равен величине интенсивности потока $\lambda = \mu$.

4.4.2. Модели нестационарных и неординарных пуассоновских потоков вызовов

К потокам без последствия кроме простейшего потока относят также нестационарный пуассоновский поток и неординарные стационарные потоки без последствия.

4.4.2.1. Модель нестационарного пуассоновского потока вызовов

Нестационарный пуассоновский поток – это ординарный поток без последствия, для которого в любой момент времени существует конечный параметр $\lambda(t)$. Различают потоки с переменным и случайным параметрами. В первом случае функция $\lambda(t)$ заранее известна и значение ее зависит только от t , а во втором случае $\lambda(t)$ – случайная функция.

Функция $\lambda(t)$ может быть непрерывной или ступенчатой. В последнем случае параметр потока изменяется скачками в заранее известные или случайные моменты времени. Величины скачков также могут быть заранее известны или случайны, причем для пуассоновского потока со случайным параметром могут быть случайны только моменты или только величины изменения параметра, или одновременно то и другое.

По аналогии с простейшим потоком в качестве математической модели пуассоновского потока с переменным параметром (нестационарный пуассоновский поток) выбирается семейство вероятностей $P_i(t_0, t_0 + \Delta t)$ поступления точно i вызовов за заданный промежуток $[t_0, t_0 + \Delta t]$. В силу нестационарности потока эта вероятность зависит не только от длины промежутка времени Δt , но и от начального момента t_0 . Во многом повторяя ход рассуждений, можно показать, что

$$P_i(t_0, t_0 + \Delta t) = \frac{[\lambda(t_0, t_0 + \Delta t)]^i}{i!} \cdot e^{-\lambda(t_0, t_0 + \Delta t)}, \quad (4.36)$$

где $\lambda(t_0, t_0 + \Delta t)$ – математическое ожидание числа вызовов в промежутке $[t_0, t_0 + \Delta t]$:

$$\lambda(t_0, t_0 + \Delta t) = \int_{t_0}^{t_0 + \Delta t} \lambda(t) dt.$$

Модель пуассоновского потока с переменным параметром позволяет при соответствующем выборе зависимости $\lambda(t)$ достаточно хорошо описать реальный поток, например процесс поступления вызовов на телефонную станцию в течение суток.

4.4.2.2. Модели стационарных неординарных пуассоновских потоков вызовов без последствия

Неординарные пуассоновские потоки вызовов это стационарные потоки без последствия, у которых условие ординарности не выполняется. Различают неординарные (групповые) пуассоновские потоки и пуассоновские потоки с неординарными вызовами.

Стационарный поток без последствия называется неординарным (групповым) пуассоновским, если в вызывающий момент поступает более одного вызова.

Исходя из определения для данного потока вызовов, различают поток вызывающих моментов и поток вызовов.

Вызывающие моменты неординарного потока образуют простейший поток с параметром λ , поэтому вероятность поступления i вызывающих моментов в промежутке t определяется согласно распределению Пуассона.

В каждый вызывающий момент поступает l ($l = 1, \dots, r$) одинаковых вызовов. Величина l называется характеристикой неординарности потока, которая может быть постоянной или случайной. Если l является постоянной величиной, то с вероятностью $P_k(t)$ суммарное число вызовов, поступающих за отрезок t , составляет $k = l \cdot i$.

Для неординарного пуассоновского потока с переменной величиной l в каждый вызывающий момент с вероятностью P_l поступает l вызовов. Параметр такого потока для каждого значения l равен $\lambda \cdot P_l$.

Отсюда общий параметр потока $(\sum_{l=1}^r P_l \cdot \lambda = \lambda)$ такой же, как

и для потока вызывающих моментов, т. е. для простейшего потока.

Из-за громоздкости формулы для определения вероятности поступления ровно i вызовов она не приводится.

Интенсивность μ неординарного пуассоновского потока, как и любого стационарного неординарного потока, больше его параметра. Действительно,

$$\mu = \sum_{l=1}^r l \cdot \lambda \cdot P_l = \lambda \cdot \sum_{l=1}^r l \cdot P_l > \lambda. \quad (4.37)$$

К рассматриваемому потоку близко примыкает пуассоновский поток с неординарными вызовами.

Неординарным назовем вызов, требующий для своего обслуживания l ($l = 1, \dots, r$) приборов. Вероятность поступления i вызовов за промежуток t определяется выражением Пуассона, а вероятность поступления k требований на приборы за промежуток t – так же, как и для неординарного группового пуассоновского потока.

Рассматриваемые потоки, несмотря на ряд общих свойств, имеют существенные отличия. При неординарном пуассоновском потоке каждый вызов из группы одновременно поступивших может быть обслужен отдельно, время его обслуживания строго индивидуально и не зависит от времени обслуживания других. При пуассоновском потоке с неординарными вызовами все приборы, необходимые для обслуживания одного неординарного вызова, занимаются и освобождаются одновременно.

4.5. Математические модели потоков вызовов с последствием

Для описания реальных потоков вызовов, несмотря на удобную и простую в плане исследования модель простейшего потока, приходится иметь дело с нестационарными потоками и потоками с последствием. Действительно, абоненты учрежденческих АТС, а также ведомственных сетей связи, как правило, образуют ограниченную группу источников нагрузки. Ограниченное число занятых каналов направлений связи также формирует поток освобождений, имеющий определенное последствие.

Использование модели стационарного ординарного потока без последствия для описания процессов обслуживания вызовов в таких системах некорректно.

Действительно, потоки вызовов от ограниченной группы источников нагрузки, потоки освобождений ограниченного числа каналов связи, а также потоки вызовов, пропущенные через m ступеней искажения, обладают последствием.

В этой связи рассмотрение математических моделей потоков с последствием имеет важное значение для практики.

Ординарный поток, параметр которого $\lambda_{r(t)}$ полностью определяется состоянием $r(t)$ обслуживающей системы в рассматриваемый момент времени t , называется *пуассоновским с условным параметром*.

Под *состоянием системы* имеется в виду информация о числе и, возможно, номерах занятых входов, выходов и соединительных путей между ними, числе свободных, обслуживаемых, ждущих соединения или повторяющих вызовы источников. Все возможные состояния системы образуют дискретное множество. Так как КС всегда имеет конечное число входов, выходов и т. д., то конечным является и число возможных состояний системы обслуживания, которые называют микросостояниями.

Состояния системы, различающиеся только числом занятых входов (выходов), называются макросостояниями. Обычно исследуют макросостояния системы, так как на практике вероятности нахождения СК в том или ином состоянии в любой произвольный момент времени зависят только от числа занятых входов (источников нагрузки) или выходов (каналов, линий связи).

Поскольку состояние обслуживающей системы в момент времени t зависит от процесса поступления и обслуживания вызовов до момента t , то рассматриваемый поток обладает последствием.

Данное последствие называют *простым*, так как для определения параметра и других характеристик потока в момент t достаточно информации о состоянии системы только в этот момент времени.

Изменение параметра потока в соответствии с изменением системы приводит к нестационарности потока. Для любых двух произвольно расположенных, но одинаковых по величине промежутков времени вероятности поступления вызовов, попавших в данные интервалы, в общем случае не равны и зависят от конкретного набора состояний. Однако зависимость параметра от времени является косвенной, через состояние системы. Для каждого конкретного состояния системы параметр является величиной постоянной.

Понятие пуассоновского потока с условным параметром – одно из наиболее общих в теории потоков. Практически любой поток вызовов можно считать потоком с простым последствием, поскольку коммутационная система всегда влияет на процесс поступления вызовов.

К частным случаям потока с простым последствием относят:

- примитивный поток вызовов;
- сглаженный пуассоновский поток;
- поток с повторными вызовами;
- поток освобождений.

4.5.1. Модель примитивного потока вызовов

Модель примитивного потока вызовов, как правило, используется для описания реального потока от ограниченной группы источников нагрузки. Число источников нагрузки системы распределения информации не превышает 100 абонентов ($N < 100$). Данную модель можно применять и при числе источников до 300 абонентов ($N = 300$), но при этом должно выполняться условие $N < 15V$, где N – число источников нагрузки, а V – число обслуживающих приборов (каналов) системы коммутации.

При подключении к системе коммутации более 300 пользователей для теоретического описания потока вызовов обычно используют модель простейшего потока вызовов, так как последствие потока в этом случае проявляется незначительно и им, как правило, пренебрегают.

Необходимость учета кроме численности абонентов еще и обслуживающих приборов вызвана тем, что у потока с *простым последствием* вероятностные свойства зависят (определяются) от состояния системы (V).

В модели примитивного потока вызовов, в отличие от модели простейшего потока, учитывается параметр (интенсивность) потока от одного источника. При ограниченном числе абонентов параметр общего потока вызовов от всех источников будет величиной переменной, зависящей от состояния системы.

Согласно утверждениям А. Элдина и Г. Линда, требования на установление соединения от каждого источника поступают в случайные моменты времени, образуя случайный поток вызовов с интенсивностью α (которую можно интерпретировать как среднее число вызовов за единицу времени). Так как занятый источник не может посылать новые требования, то под α обычно понимают параметр (интенсивность) потока вызовов одного источника в свободном состоянии.

В силу того, что поток вызовов от одного источника случаен, то с достаточной для практических расчетов точностью можно утверждать, что промежуток между вызовами, а также промежуток свободности (см. предложение выше) подчиняются показательному закону с параметром α :

$$P(t_{\text{св}} < t) = 1 - e^{-\alpha \cdot t}. \quad (4.38)$$

Это равносильно предположению о том, что новые вызовы от источника поступают случайно, независимо от текущего времени и от моментов возникновения и окончания обслуживания предыдущих вызовов.

Эту вероятность также можно трактовать как вероятность поступления требования за промежуток времени, равный t :

$$P(t) = \begin{cases} 1 - e^{-\alpha \cdot t}, & \text{если абонент свободен;} \\ 0, & \text{если абонент занят.} \end{cases} \quad (4.39)$$

Разложим $e^{-\alpha \cdot t}$ в ряд:

$$e^{-\alpha \cdot t} = 1 - \alpha \cdot t + \frac{\alpha^2 \cdot t^2}{2!} - \frac{\alpha^3 \cdot t^3}{3!} + \frac{\alpha^4 \cdot t^4}{4!} - \dots$$

Очевидно, что если промежуток времени $t = \Delta t \rightarrow 0$ – бесконечно малая величина, то выражение (4.39) можно записать как

$$P(\Delta t) = \alpha \cdot \Delta t + o(\Delta t). \quad (4.40)$$

Далее перейдем к рассмотрению потока требований, поступающих на узел коммутации от всех $N < 100$ источников нагрузки.

При условии того, что вероятность поступления вызова от одного источника $P_1(\Delta t) = 1 - e^{-\alpha \cdot t}$ и, предположив, что вызовы от абонентов поступают независимо друг от друга, то, используя известную теорему о сложении вероятностей независимых случайных событий, получим

$$P_N(\Delta t) = \sum_{i=1}^N (1 - e^{-\alpha \cdot \Delta t}). \quad (4.41)$$

При условии, что все источники создают одинаковую нагрузку, формулу (4.41) можно записать в виде

$$P_N(\Delta t) = N \cdot (1 - e^{-\alpha \cdot \Delta t}). \quad (4.42)$$

Выполнив с выражением (4.42) действия, аналогичные произведенным для формул (4.39) – (4.40), получим

$$P_N(\Delta t) = N \cdot \alpha \cdot \Delta t. \quad (4.43)$$

В выражении (4.43) произведение $N \cdot \alpha = \lambda$ – есть интенсивность (параметр) потока требований от свободных источников нагрузки данного УК.

Следовательно, можно заключить, что интенсивность общего потока в рассматриваемый промежуток времени зависит от числа свободных источников на данный момент времени, поэтому примем, что в какой-то момент времени из N источников занято i , тогда общее число свободных источников нагрузки

$$N_{\text{св}} = N - i.$$

Отсюда общий параметр вызовов от ограниченной группы источников нагрузки

$$\lambda_i = \alpha \cdot N_{\text{св}} = \alpha \cdot (N - i). \quad (4.44)$$

Таким образом, примитивным потоком называется ординарный поток, параметр которого λ_i прямо пропорционален числу свободных источников N_i в состоянии обслуживающей системы i :

$$\lambda_i = \alpha \cdot N_i = \alpha \cdot (N - i),$$

где α – параметр потока вызовов одного источника в свободном состоянии (при этом имеет место естественное предположение: занятый источник не может производить вызовы), N – общее число источников вызова, i – число занятых источников.

В модели примитивного потока (4.44) параметр α является постоянной величиной, а параметр общего потока вызовов изменяется скачкообразно и убывает с увеличением числа занятых источников i .

Математическое ожидание параметра данного потока вызовов от ограниченного числа источников нагрузки λ :

$$\lambda = \sum_{i=0}^n P_i \cdot \lambda_i, \quad (4.45)$$

где P_i – вероятность того, что в системе занято i источников.

Величина λ , отнесенная к одному источнику $\nu = \frac{\lambda}{N}$, определяет среднюю интенсивность потока вызовов от источника.

Таким образом, в модели примитивного потока вызовов для характеристики одного источника нагрузки имеются два показателя: α и ν .

В чем отличие величин α и ν ?

Рис. 4.5 позволяет пояснить различие между этими параметрами. Пусть на достаточно большом промежутке времени от источника поступило n вызовов.

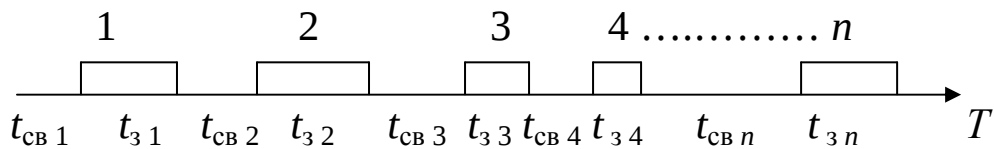


Рис. 4.5. Поток вызовов от одного источника

Как известно, интенсивность – это величина, обратная времени. Поэтому интенсивность потока вызовов от свободного источника α найдем как величину, обратную среднему времени, в течение которого источник свободен: $\alpha = \frac{1}{t_{св}}$.

Физически эту формулу можно пояснить следующим образом: чем меньше время между занятиями источника, тем интенсивнее абонент создает нагрузку.

Из рис. 4.5 видно также, что среднее время свободности источ-

ника можно определить как $\bar{t}_{св} = \frac{\sum_{j=1}^n t_{св j}}{n}$, где $t_{св j}$ – время свободности между $(j - 1)$ и j занятии абонента, n – число занятии абонента (длительность времени, в течение которого абонент занимал для разговора обслуживающий прибор: канал, линию связи).

Следовательно, интенсивность источника в свободном состоянии можно записать как

$$\alpha = \frac{n}{\sum_{j=1}^n t_{св j}}. \quad (4.46)$$

Средняя интенсивность потока от одного источника есть отношение числа поступивших вызовов ко всему рассматриваемому промежутку времени:

$$\nu = \frac{n}{T}. \quad (4.47)$$

При этом фиксируемый промежуток (T) состоит из длительности времени, свободности и занятости источника нагрузки

$$T = \bar{t}_{\text{св}} + \bar{t}_{\text{зан}}, \text{ где } \bar{t}_{\text{зан}} = \bar{t}_{\text{зан}} = \frac{\sum_{j=1}^n t_{\text{зан } j}}{n}.$$

$$\text{Так как } \alpha = \frac{1}{\bar{t}_{\text{св}}}, \quad \nu = \frac{1}{(\bar{t}_{\text{св}} + \bar{t}_{\text{зан}})}, \text{ то очевидно } \alpha > \nu.$$

Таким образом, интенсивность потока от свободного источника есть величина, обратная среднему свободному времени, а средняя интенсивность потока от источника – величина, обратная среднему промежутку между вызовами.

Примитивный поток, как было отмечено ранее, является более общим понятием по сравнению с простейшим потоком. С ростом числа источников N и соответствующим уменьшением величины α последствие примитивного потока сокращается. В пределе при $N \rightarrow \infty$ и $\alpha \rightarrow 0$, но так, что $N \cdot \alpha = \text{const}$ и i ограничено сверху, примитивный поток переходит в простейший с параметром $\lambda = N \cdot \alpha$. Практически уже при $N = 300 \dots 500$ источников нагрузки (в зависимости от величины α и максимального значения i) можно пользоваться моделью простейшего потока. Вносимая при этом погрешность невелика.

В развитие рассмотренной модели примитивного потока можно предположить, что каждый из N источников в свободном состоянии имеет свою, отличную от других источников интенсивность потока α_j ($j = 1, \dots, N$). В этом случае интенсивность (параметр) общего потока вызовов будет определяться не только числом источников в данном состоянии системы, но и их конкретным составом. Данная модель рассмотрена в работах Б. С. Лившица, В. Я. Фидлина, Ю. Н. Корнышева, где доказано «экстремальное свойство» примитивного потока.

Это свойство заключается в том, что если источники нагрузки создают одинаковую нагрузку на систему коммутации, то суммарный параметр нагрузки будет больше, чем для случая, когда все источники различны. Эта особенность показана на рис. 4.6.

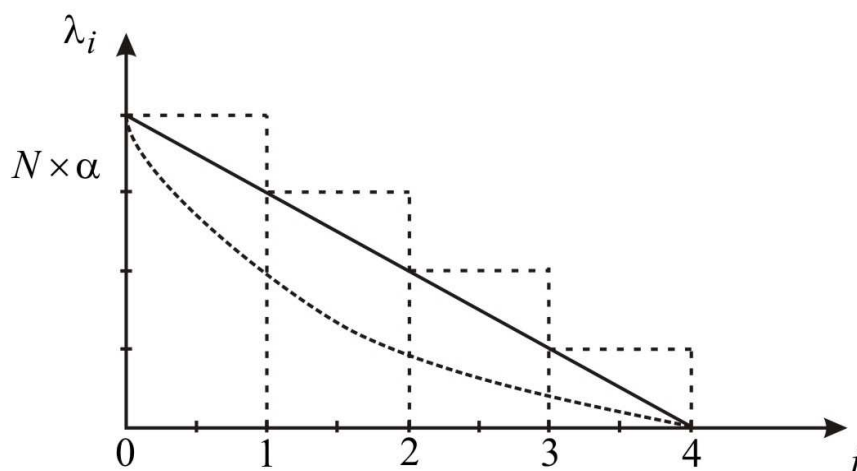


Рис. 4.6. Зависимости интенсивности нагрузки λ_i от числа занятых источников в случаях $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4$ и $\alpha_1 > \alpha_2 > \alpha_3 > \alpha_4$

4.5.2. Модель сглаженного пуассоновского потока

Поток, пропущенный через m ($m = 1, 2, 3, \dots$) ступеней искания (этапов соединения) и поступающий на $(m + 1)$ -ю ступень коммутационной системы, называется *сглаженным*.

Если же на первую ступень поступает простейший поток, то он называется *сглаженным пуассоновским*.

Параметр сглаженного пуассоновского потока полностью определяется состоянием предыдущих ступеней искания.

Любой пучок линий (группа приборов) ступени искания по своему действию подобен клапану, который пропускает на следующую ступень все поступающие на него вызовы до тех пор, пока хотя бы одна линия пучка свободна, и закрывается при занятости всех линий.

Поэтому если на пучок поступает простейший поток вызовов с параметром λ , параметр потока на выходе пучка будет иметь одно из двух значений: λ (при наличии хотя бы одной свободной линии) или 0 (при занятости всех линий).

Отсюда можно сделать вывод, что сглаженный пуассоновский поток относится также к классу потоков с простым последствием, поскольку параметр его зависит от состояния системы обслуживания.

Предположим теперь, что ступень искания состоит из N одинаковых пучков емкостью V и на каждый пучок поступает простейший поток с параметром λ , тогда сглаженный поток на выходе ступени будет иметь параметр

$$\lambda_i = \lambda \cdot (N - i), \quad (4.48)$$

где i – число пучков, в которых заняты все V линий.

Следует отметить, что характерное последствие для сглаженного пуассоновского потока, как и для примитивного, правда в меньшей степени, способствует увеличению пропускной способности следующей ступени искания.

4.5.3. Модели потоков с повторными вызовами

Реальная система коммутации, на которую поступает поток вызовов, обслуживает не все поступающие вызовы. Часть из них не обслуживается, а теряется по различным причинам, к которым следует отнести занятость или неотчет абонента, ошибки в процессе набора номера, занятость всех соединительных устройств, способных обслужить поступивший вызов. Все или часть источников необслуженных вызовов осуществляют повторение вызова. Следовательно, повторный вызов возникает как результат потери предыдущего первичного или повторного вызова и является реакцией источника на работу системы обслуживания.

Поток с повторными вызовами состоит из потока первичных вызовов (по своим свойствам простейший, примитивный или нестационарный пуассоновский) и повторных вызовов.

Если поток первичных вызовов является простейшим, то его параметр равен λ .

Параметр потока повторных вызовов равен произведению числа источников, повторяющих вызовы j , на интенсивность потока от одного источника, повторяющего вызовы β .

В этом случае суммарный параметр потока с повторными вызовами (с учетом потоков первичных и повторных вызовов) будет определяться по формуле

$$\lambda_{\Sigma} = \lambda + j \cdot \beta, \quad (4.49)$$

где β – интенсивность источника, повторяющего вызовы.

Интенсивность β обратно пропорциональна среднему времени повторения вызовов и определяется по формуле $\beta = \frac{1}{t_{\text{ПВ}}}$. Очевидно, что чем меньше время между повторными вызовами, тем больше β . Для абонентов местных телефонных сетей $\beta \geq \alpha$, причем величина β превышает значение интенсивности источника в свободном состоянии α примерно в 50 – 100 раз.

Если поток первичных вызовов аппроксимируется моделью примитивного потока вызовов с параметром $\lambda_i = (N - i) \cdot \alpha$, то в этом случае параметр суммарного потока

$$\lambda_{\Sigma} = (N - i - j) \cdot \alpha + j \cdot \beta, \quad (4.50)$$

где i – число источников, занятых обслуживанием;

j – число источников, повторяющих вызов;

α – интенсивность вызовов от одного свободного источника нагрузки;

N – число источников нагрузки ($N \leq 100$).

При равных значениях интенсивности α и β потоки первичных и повторных вызовов становятся в вероятностном смысле неразличимы, а суммарный поток определяется выражением

$$\lambda_{\Sigma} = (N - i) \cdot \alpha.$$

Следует отметить, что повторные вызовы распределены во временном интервале весьма неравномерно и появляются, как правило, в интервалах с большой концентрацией первичных вызовов. При этом, поступая через короткий промежуток времени ($t_{\text{ПВТ}} = 1/\beta$) после получения отказа, повторный вызов "успевает" в этот же интервал поступить в систему, увеличивая концентрацию вызовов.

При проведении исследований по нагрузке и, в частности, при анализе реального потока вызовов возникает необходимость отделения повторных вызовов от первичных и оценки их величин α и β . В качестве признака повторного вызова обычно используют величину промежутка τ_i , предшествующего поступлению i -го вызова. Задавшись граничным значением τ_{Γ} , можно отделить вызовы с промежутками $\tau_i \leq \tau_{\Gamma}$, полагая, что они повторные, а оставшиеся вызовы отнести к первичным. Данная процедура дает хорошие результаты на практике.

4.5.4. Модель потока освобождений

Потоком освобождений называется последовательность моментов окончания обслуживания вызовов.

Свойства потока освобождений в общем случае зависят:

- от свойств поступающего потока вызовов;
- дисциплины обслуживания;
- качества работы коммутационной системы;
- закона распределения времени обслуживания.

При обслуживании поступающего потока без потерь при постоянной длительности обслуживания $t_{\text{обсл}}$ свойства потока освобождений совпадают со свойствами поступающего потока вызовов. При этом происходит только сдвиг во времени на величину $t_{\text{обсл}}$ между моментами поступления вызова и моментом его обслуживания.

Длительность обслуживания реального вызова, определяемая в основном длительностью разговора, является величиной случайной. Случайное время обслуживания задается законом и плотностью распределения. Наиболее простым и самым распространенным законом распределения случайного времени обслуживания является показательный:

$$P(T_{\text{обсл}} < t) = 1 - e^{-\frac{t}{\bar{t}_{\text{обсл}}}} = 1 - e^{-\mu t}, \quad (4.51)$$

где $\bar{t}_{\text{обсл}}$ – среднее время обслуживания, $\mu = \frac{1}{\bar{t}_{\text{обсл}}}$ – интенсивность освобождения обслуживающего прибора.

В силу свойства показательного закона моменты окончания обслуживания не зависят от моментов поступления вызовов.

В этом случае свойства потока освобождений не зависят от свойств поступающего потока вызовов, качества работы коммутационной системы и полностью определяются числом обслуживаемых вызовов. Докажем это.

Если в коммутационной системе занято k линий, то вероятность освобождения i линий за промежуток t можно рассматривать как i успешных испытаний среди общего числа k независимых испытаний и по теореме повторения опытов, используя распределение Бернулли:

$$P(i, k, t) = C_k^i \cdot p^i \cdot (1 - p)^{k-i}, \quad (4.52)$$

где p – вероятность освобождения одной конкретной линии за время t .

Так как в данном случае используется показательный закон распределения времени освобождения, то вероятность освобождения линии p определяется из выражения

$$p = P(T_{\text{обсл}} < t) = 1 - e^{-t/t_{\text{обсл}}}.$$

Тогда выражение (4.52) можно записать в виде

$$P(i, k, t) = C_k^i \cdot \left(1 - e^{-t/t_{\text{обсл}}}\right)^i \cdot e^{-(k-i) \cdot t/t_{\text{обсл}}}. \quad (4.53)$$

Формулу (4.53) можно использовать для определения вероятности неосвобождения ни одной занятой линии пучка за время t :

$$\begin{aligned} P(0, k, t) &= C_k^0 \cdot \left(1 - e^{-t/t_{\text{обсл}}}\right)^0 \cdot e^{-(k-0) \cdot t/t_{\text{обсл}}} = \\ &= e^{-k \cdot t/t_{\text{обсл}}} = e^{-k \cdot t \cdot \mu}, \end{aligned} \quad (4.54)$$

где $\mu = \frac{1}{t_{\text{обсл}}}$ — интенсивность освобождения занятых линий.

Учитывая, что сумма вероятностей, образующих полную группу событий, равна единице, можно рассчитать вероятность освобождения хотя бы одной линии:

$$P(i \geq 1, k, t) = 1 - P(0, k, t) = 1 - e^{-k \cdot t \cdot \mu}. \quad (4.55)$$

В этом случае параметр потока освобождений

$$\lambda_{k \text{ осв}} = \lim_{\Delta t \rightarrow 0} \frac{P(i \geq 1, k, t)}{\Delta t}. \quad (4.56)$$

С учетом разложения функции e^{-x} в ряд Маклорена $\sum_{j=0}^{\infty} (-1)^j \cdot \frac{x^j}{j!}$ выражение (4.56) примет вид:

$$\lambda_{k \text{ осв}} = \lim_{\Delta t \rightarrow 0} \frac{1 - \sum_{j=0}^{\infty} (-1)^j \cdot (k \cdot \Delta t \cdot \mu)^j / j!}{\Delta t} =$$

$$= \lim_{\Delta t \rightarrow 0} \left[k \cdot \mu + \frac{o(\Delta t)}{\Delta t} \right] = k \cdot \mu. \quad (4.57)$$

Таким образом, из формулы (4.57) видно, что параметр потока освобождений зависит от числа занятых линий k и интенсивности освобождения одной линии μ , что и требовалось доказать.

4.6. Математические модели потоков с ограниченным последствием

Под потоком с ограниченным последствием понимается поток вызовов, у которого последовательность промежутков между вызовами $\tau_1, \tau_2, \tau_3, \dots, \tau_n$ представляет собой последовательность взаимно независимых случайных величин, имеющих любые функции распределения. Следовательно, такой поток вызовов можно задать последовательностью функций распределения промежутков между вызовами τ_n :

$$F_n(\tau) = P(T_n^* < \tau), \quad n = 1, 2, \dots \quad (4.58)$$

У потока с ограниченным последствием вероятность поступления нового вызова в промежуток τ зависит только от расположения этого промежутка по отношению к моменту поступления последнего вызова и не зависит от времени поступления всех остальных, и все последствие ограничивается величиной промежутка между вызовами.

В качестве примера потока с ограниченным последствием может служить радиолампа, которая работает до своего отказа (выхода из строя). Отказавшая лампа тут же заменяется новой. Если каждая лампа выходит из строя независимо от других, то поток отказов будет являться потоком с ограниченным последствием.

Частным случаем потока с ограниченным последствием является рекуррентный поток, который характеризуется одинаково распределенными промежутками времени между вызовами:

$$F_1(\tau) = F_2(\tau) = F_3(\tau) = \dots = F(\tau). \quad (4.59)$$

Можно показать, что стационарный (следовательно, все промежутки между вызовами распределены по экспоненте) рекуррентный поток будет простейшим.

Рекуррентный поток, у которого только первый промежуток имеет отличное от всех других промежутков распределение, называется рекуррентным потоком с запаздыванием:

$$F_2(\tau) = F_3(\tau) = \dots = F(\tau); \quad F_1(\tau) \neq F(\tau). \quad (4.60)$$

Таким образом, данный поток задается двумя функциями распределения. Функция $F_1(\tau)$ характеризует распределение промежутка времени от произвольно выбранного начала отсчета t_0 до момента поступления первого вызова.

Стационарный рекуррентный поток с запаздыванием называют потоком Пальма.

Модель потока Пальма задается условной вероятностью $\phi_0(\tau)$ отсутствия вызовов в промежутке длиной τ :

$$F_1(\tau) = P(T_\tau^* < \tau) = \lambda \cdot \int_0^\tau \phi_0(U) \cdot dU;$$

$$F_n(\tau) = P(T_n^* < \tau) = 1 - \phi_0(\tau), \quad n \geq 2. \quad (4.61)$$

В выражении (4.61) величина λ – параметр потока Пальма ($\lambda = \frac{1}{\tau}$). При $\phi_0(\tau) = e^{-\lambda \cdot \tau}$ получается простейший поток, поскольку все промежутки, включая первый, распределены по показательному закону.

Модель потока Пальма хорошо описывает поток не обслуженных коммутационной системой вызовов. В частности, если на коммутационную систему с явными потерями поступает простейший поток вызовов, то поток потерянных вызовов обладает ограниченным последствием и по своим свойствам будет потоком Пальма. Модель потока Пальма используется для расчета недоступных включе-

ний сетей с обходами. Определенную сложность при этом представляет то обстоятельство, что при объединении двух и более независимых потоков Пальма суммарный поток по своим свойствам будет отличаться от потока Пальма. Однако при разъединении потока Пальма на несколько направлений так, что с вероятностью P_i вызов поступит на i -е направление, поток в i -м направлении также будет потоком Пальма.

Понятие "поток Пальма" объединяет широкий класс потоков, отличающихся друг от друга используемым законом распределения промежутков между вызовами.

Одним из примеров потока Пальма является поток Эрланга, образуемый просеиванием простейшего потока вызовов. Если из последнего выбросить каждый второй вызов, то оставшиеся вызовы образуют поток Эрланга второго порядка. Поток Эрланга третьего порядка получится, если сохранить в простейшем потоке каждый третий вызов. Поток Эрланга n -го порядка называется поток, получаемый из простейшего, если сохранить каждый n -й вызов, а остальные выбросить. Очевидно, что простейший поток можно рассматривать как поток Эрланга первого порядка.

Промежутки между вызовами в потоке Эрланга независимы между собой и одинаково распределены, поскольку они получаются при суммировании одинакового числа независимых промежутков исходного простейшего потока.

Кроме рассмотренных потоков с ограниченным последствием, разработаны модели нормированного потока Эрланга n -го порядка, обобщенный поток Эрланга n -го порядка, прерывистый пуассоновский поток n -го порядка и другие.

Вопросы для самоконтроля

1. Что подразумевается под потоком вызовов?
2. Что такое неоднородный поток вызовов?
3. Что такое однородный поток вызовов?
4. Приведите примеры детерминированных потоков вызовов.
5. Что имеется в виду под финитным потоком вызовов?
6. Какими способами можно задать детерминированные потоки вызовов?
7. Как задаются случайные потоки вызовов?
8. Назовите свойства потоков вызовов.

9. Что понимают под стационарностью потока вызовов?
10. Что понимают под ординарностью потока вызовов?
11. Что понимают под отсутствием последствия в потоке вызовов?
12. Является ли поток вызовов от районной АТС 43 г. Орла потоком с последствием?
13. Является ли поток вызовов от абонентов на коммутатор СДС-4 емкостью 10 абонентов и 3 каналов стационарным?
14. Поток телеграмм является ординарным или неординарным?
15. Перечислите характеристики потоков вызовов.
16. Что понимают под ведущей функцией потока вызовов?
17. Что понимают под параметром потока вызовов?
18. Что означает интенсивность потока вызовов?
19. Какое свойство используется для классификации потоков вызовов?
20. Какие потоки относятся к классу потоков без последствия?
21. Какие потоки относятся к классу потоков с простым последствием?
22. Какие потоки относятся к классу потоков с ограниченным последствием?
23. Какие потоки входят в класс потоков без последствия?
24. Дайте определение простейшего потока вызовов.
25. Какое число источников нагрузки (абонентов) может создать поток вызовов, аппроксимируемый моделью простейшего потока?
26. Почему простейший поток вызовов получил такое название?
27. Простейший поток является детерминированным потоком или потоком чистой случайности?
28. Модель простейшего потока вызовов предложена Пуассоном или Эрлангом?
29. "Простейший поток" – это модель потока вызовов или название реального потока, поступающего на телефонную станцию?
30. Способы задания простейшего потока вызовов.
31. Подходы для определения семейства вероятностей $P_i(t)$.
32. Какие свойства потоков вызовов были использованы при выводе формулы Пуассона?
33. Что позволяет определить вероятность $P_i(t)$?
34. Какое распределение имеет $P_i(t)$ с ростом $\lambda \cdot t$?
35. При каких значениях i вероятность $P_i(t)$ будет иметь два экстремума?

36. При каких значениях i вероятность $P_i(t)$ будет иметь один экстремум?
37. Какое число вызовов может поступить с наибольшей вероятностью за $t = 2$ с при $\lambda \cdot t = 10$?
38. Приведите характеристики простейшего потока.
39. Какой тип потока получается в результате объединения 20 стационарных потоков, поступающих на систему?
40. Какой тип раздельных потоков получается при разъединении простейшего потока на несколько потоков с определенной вероятностью?
41. Что означает нестационарный ординарный поток без последствия с переменным параметром?
42. Что означает нестационарный ординарный поток без последствия со случайным параметром?
43. Какой из потоков: простейший или нестационарный ординарный поток без последствия – точнее описывает реальный поток вызовов на телефонную станцию?
44. Какую модель потока вызовов чаще используют для описания реального потока, поступающего на телефонную станцию?
45. Дайте определение стационарному, неординарному, групповому потокам без последствия.
46. Что такое вызывающий момент?
47. Какой моделью потока описываются вызывающие моменты стационарного, неординарного, группового потоков без последствия?
48. Что поясняет характеристика неординарности потока?
49. В чем отличие неоднородного потока от неординарного?
50. Дайте определение пуассоновского потока с неординарными вызовами.
51. Приведите примеры пуассоновских потоков с неординарными вызовами.
52. Что понимают под потоком с условным параметром?
53. Почему поток с условным параметром обладает последствием?
54. Что такое простое последствие?
55. Поток с простым последствием обладает стационарностью?
56. Назовите модели потоков, обладающих простым последствием.

57. Для описания каких систем распределения информации можно использовать модель примитивного потока?
58. Как вычисляется интенсивность источника в свободном состоянии α и что она определяет?
59. Как соотносятся между собой интенсивность источника в свободном состоянии и средняя интенсивность источника?
60. В чем сущность "экстремального свойства" примитивного потока?
61. Дайте определение сглаженного потока и сглаженного пуассоновского потока.
62. Чему равна интенсивность сглаженного потока на выходе ступени, если на ее вход поступает поток с параметром λ ?
63. Чем обусловлено появление повторных вызовов?
64. Что такое интенсивность повторения от одного источника и как она определяется?
65. Чему равен параметр суммарного потока с повторными вызовами, если первичный поток является простейшим?
66. Чему равен параметр суммарного потока с повторными вызовами, если первичный поток является примитивным?
67. Как соотносятся между собой интенсивность свободного источника и интенсивность источника, повторяющего вызовы?
68. В каком интервале времени чаще появляются повторные вызовы?
69. Что понимают под потоком освобождений?
70. От каких факторов зависят свойства потоков освобождений?
71. Чем определяются свойства потока освобождений, если длительность обслуживания есть величина детерминированная, а способ обслуживания осуществляется без потерь?
72. Чем определяются свойства потока освобождений, если длительность обслуживания является величиной случайной, распределённой по экспоненциальному закону?
73. Дайте определение потоку с ограниченным последствием.
74. В чем проявляется последствие в данных потоках?
75. Что такое рекуррентный поток?
76. Что такое рекуррентный поток с запаздыванием?
77. Что такое поток Пальма?
78. Что такое потоки Эрланга n -го порядка?
79. Как получить поток Эрланга 5-го порядка?

5. НАГРУЗКА И ЕЕ ВИДЫ

При рассмотрении модели процесса обслуживания один из составляющих ее компонентов – поток вызовов – характеризует лишь внешнее воздействие на систему обслуживания. Умение классифицировать поток вызовов и использовать соответствующую математическую модель для оценки его вероятностных характеристик еще недостаточно для оценки качества функционирования системы коммутации.

В этой связи важным является рассмотрение таких понятий, как дисциплина обслуживания, нагрузка, характеристики качества обслуживания.

При рассмотрении основных понятий теории телетрафика следует придерживаться определений, приведенных в отечественных ГОСТах, а также Рекомендациях Международного союза электросвязи E.600 (МСЭ E.600).

Основными понятиями, используемыми в теории телетрафика, являются связь, или коммуникация (*communication*), которая предусматривает процесс передачи информации в соответствии с некоторыми правилами, а также соединение (*connection*) – некоторая ассоциация двух и более устройств внутри сети или посредством сети для осуществления связи между ними.

При рассмотрении процессов обслуживания часто встречаются такие понятия, как пользователь (*user*), являющийся общим термином для всех внешних по отношению к сети сущностей, а также ресурс (*resource*), под которым понимают технические средства и сооружения электросвязи, используемые для предоставления услуг [Постановление Правительства РФ от 17 октября 1997 г. № 1331 "Об утверждении Основных положений ведения взаиморасчетов между операторами сетей электросвязи, образующих сеть электросвязи общего пользования, за предоставляемые сетевые ресурсы и участие в передаче нагрузки этих сетей", п. 2 Основных положений].

В процессе коммуникации пользователей в сети возникает поток сообщений. В отечественной литературе по телефонии под потоком сообщений понимают нагрузку, в других источниках используют такое понятие, как трафик или телетрафик (греч. *tele* – далеко и лат. *traff* – движение).

5.1. Нагрузка (трафик), интенсивность нагрузки, единицы измерения нагрузки

Нагрузка (трафик) может быть охарактеризована количественно. Очевидным параметром трафика является его объем (*traffic volume*). Для цифровых систем эта величина естественным образом ассоциируется с числом бит, переданных за данное время. Однако для аналоговых систем, доля которых на сети специальной связи еще достаточно существенна, такой подход оказывается неприемлемым. Кроме того, при использовании цифровых систем коммутации со сложными способами модуляции и кодирования сигналов определение объема трафика становится неоднозначным. Сравним данную ситуацию автомобильного движения на дороге с большим трафиком, когда автомобили очень плотно заполняют полосы движения, а количество груза в них и даже число автомобилей не имеют при этом большого значения. Приведенная аналогия позволяет охарактеризовать объем трафика как меру занятости пространства для движения. Роль такого пространства для телетрафика играет время, в пределах которого доступен тот или иной ресурс сети. Поэтому под объемом трафика, "пропущенным" тем или иным ресурсом, следует понимать величину суммарного, интегрального интервала времени, в течение которого данный ресурс был занят за анализируемый период времени.

5.1.1. Нагрузка и ее виды

В статье 225 ГОСТ 19472–88 для телефонных систем введено понятие телефонной нагрузки, под которой понимается суммарное время телефонного занятия линий, каналов телефонной сети или групп коммутационных приборов связи за интервал времени. При этом в статье 215 под телефонным занятием подразумевается процесс использования линий или канала телефонной сети, или коммутационного прибора связи для обслуживания телефонного вызова независимо от того, закончился ли он разговором.

В качестве единицы измерения нагрузки (трафика) принято часозанятие, реже минутозанятие и секундозанятие.

Одно часозанятие – это такая нагрузка, которая может быть обслужена одним выходом в течение часа при непрерывном занятии этого выхода.

Однако рекомендации Международного союза электросвязи (ITU) дают размерность объема трафика в эрлангочасах.

Это объясняется тем, что специалисты часто используют понятие "интенсивность нагрузки" или «интенсивность трафика» (*traffic intensity*). Под интенсивностью нагрузки понимается нагрузка за единицу времени. В этом случае речь идет о "средней интенсивности нагрузки". При этом интервал времени усреднения обычно задается, и в качестве единицы времени, как правило, используют отрезок времени, равный одному часу. В этом случае интенсивность нагрузки будет измеряться в часозанятиях.

Если в каждый момент времени t из заданного интервала (t_1, t_2) число занятых обслуживанием трафика ресурсов из данного набора равно $Y(t)$, то средняя интенсивность трафика может быть оценена как

$$\overline{Y}(t_1, t_2) = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} Y(t) dt.$$

В качестве исторической справки следует отметить следующее. В 1943 г. по предложению известного шведского ученого в области теории массового обслуживания К. Пальма был объявлен конкурс на название единицы телефонной нагрузки. За большие заслуги в создании и развитии теории телетрафика имя Агнера Крапуа Эрланга было увековечено в названии этой единицы.

С 1944 г. этот термин использовался в скандинавских странах, а в конце второй мировой войны на собрании Международного союза электросвязи единица измерения "Эрланг" была принята в качестве международной.

Применительно к линии связи (ресурсу) единица измерения интенсивности нагрузки 1 эрланг (1 Эрл, 1 Е) – это непрерывное занятие данной линии в течение одного часа.

В американской литературе можно встретить другую единицу измерения интенсивности трафика, называемую CCS – *Centrum (or hundred) Calls Second* (гектосекундозанятие).

Число CCS отражает время занятия каналов связи в 100 секундных интервалах за один час. Нетрудно видеть, что интенсивность, измеренную в CCS, всегда можно пересчитать в Эрлангах по формуле

$$36 \text{ CCS} = 1 \text{ Эрл.}$$

Практика показывает, что на входы и выходы систем коммутации поступает разное число вызовов, при этом длительность занятия входов и выходов различна, поэтому специалисты при исследовании систем коммутации вынуждены были ввести классификацию нагрузки.

В настоящее время различают пять видов нагрузки: потенциальную, поступающую, обслуженную, потерянную, избыточную.

Под потенциальной нагрузкой $A(t_1, t_2)$, поступающей на коммутационную систему за промежуток времени $[t_1, t_2)$, понимается такая нагрузка, которая была бы обслужена коммутационной системой за рассматриваемый промежуток времени, если бы каждому поступающему вызову тотчас было предоставлено соединение со свободным выходом.

Из данного определения можно сделать вывод, что потенциальная нагрузка – это та нагрузка, которую хотел создать пользователь, не взаимодействуя с телефонной станцией. Поэтому данную нагрузку в различных источниках называют абстрактной, так как практически измерить ее на АТС невозможно.

В отличие от потенциальной нагрузки *поступающая нагрузка* есть результат взаимодействия абонента (ов) с телефонной станцией. Ввиду недостаточного числа обслуживающих приборов и каналов в системе коммутации, а также по другим причинам (занятость вызываемого абонента, неправильный набор номера или его искажение при передаче и т. д.), не связанных с данной станцией, для передачи одного сообщения требуется осуществить несколько вызовов, занимая на различное время ее ресурсы (обслуживающие приборы, каналы, соединительные линии). Отсюда понятно, что в общем случае потенциальная нагрузка меньше поступающей. Данную нагрузку, в отличие от потенциальной, можно измерить на телефонной станции.

При теоретических расчетах обычно считают, что поступающая нагрузка равна произведению мгновенной интенсивности потока в рассматриваемый момент времени на среднее время одного занятия:

$$Z(t_1, t_2) = \mu \cdot \bar{t}_{\text{зан}}. \quad (5.1)$$

Если на телефонной станции осуществлялся мониторинг числа вызовов в течение заданного интервала времени, то практически поступающую нагрузку вычисляют по формуле

$$Z(t_1, t_2) = C \cdot \bar{t}_{\text{зан}}, \quad (5.2)$$

где C – общее число вызовов, поступающих на станцию в течение интервала времени $[t_1, t_2)$, $\bar{t}_{\text{зан}}$ – среднее время одного занятия.

Важное место среди отмеченных выше видов нагрузок занимает *обслуженная нагрузка*, так как весь персонал телефонных сетей связи занимается обеспечением заданных объемов именно обслуженного трафика.

Обслуженная коммутационной системой за промежуток времени $[t_1, t_2)$ нагрузка $Y(t_1, t_2)$ есть суммарное время занятия линий, каналов телефонной сети или коммутационных приборов связи телефонными вызовами за рассматриваемый интервал времени.

В теории измерения обслуженной нагрузки широкое применение находит следующая теорема о количественной оценке обслуженной нагрузки.

Обслуживаемая системой коммутации в момент времени t нагрузка равна числу i занятых линий, приборов, входов или выходов системы в этот момент времени.

Поскольку число занятых линий в момент времени t есть величина случайная (СВ), то и нагрузка также является случайной, поэтому при теоретических исследованиях и практических расчетах используют ее математическое ожидание, дисперсию и моменты более высокого порядка.

Для определения математического ожидания интенсивности обслуженной нагрузки используют выражение

$$Y(t) = m_i(t) = \sum_{i=1}^V i \cdot P_i(t), \quad (5.3)$$

где $P_i(t)$ – вероятность занятия i ($i = 0, \dots, V$) линий из V возможных в момент времени t .

Если вероятность $P_i(t)$ постоянна в некотором промежутке времени, то интенсивность нагрузки Y будет одинаковой для всех точек этого промежутка.

Дисперсия нагрузки в момент времени t вычисляется по формуле

$$D_i(t) = \sum_{i=1}^V i^2 \cdot P_i(t) - m_i(t)^2. \quad (5.4)$$

Докажем правильность теоремы о количественной оценке обслуженной нагрузки.

Пусть на входы коммутационной системы, имеющей V выходов, поступает поток вызовов. Будем наблюдать за каждым из выходов в течение промежутка времени $[t_1, t_2)$. В результате наблюдения увидим, что в течение суммарного времени $\tau_1^\Sigma(t_1, t_2)$ был занят один выход, в течение времени $\tau_2^\Sigma(t_1, t_2)$ – два выхода, а в течение времени $\tau_i^\Sigma(t_1, t_2)$ – ровно i выходов.

За промежуток $[t_1, t_2)$ суммарное время занятия всех выходов будет равно $\sum_{i=1}^V i \cdot \tau_i^\Sigma(t_1, t_2)$.

Эта сумма, по определению, является нагрузкой, обслуженной всеми выходами коммутационной системы за интервал времени $[t_1, t_2)$. Интенсивность данной нагрузки (нагрузка за единицу времени)

$$Y(t_1, t_2) = \frac{\sum_{i=1}^V i \cdot \tau_i^\Sigma(t_1, t_2)}{t_2 - t_1}. \quad (5.5)$$

В то же время среднее число одновременно занятых выходов может быть рассчитано по формуле

$$m_i(t_1, t_2) = \sum_{i=1}^V i \cdot P_i(t_1, t_2), \quad (5.6)$$

где $P_i(t_1, t_2)$ – вероятность занятости i каналов (приборов) связи на интервале времени $[t_1, t_2)$. Вероятность $P_i(t_1, t_2)$ можно рассматривать как долю времени, в течение которого занято ровно i выходов. Тогда

$$P_i(t_1, t_2) = \frac{\tau_i^\Sigma(t_1, t_2)}{t_2 - t_1}. \quad (5.7)$$

Отсюда $Y(t_1, t_2) = m_i(t_1, t_2)$, что и требовалось доказать.

Следующим видом нагрузки является *потерянная нагрузка* $R(t_1, t_2)$. Она равна разности между потенциальной и обслуженной нагрузками:

$$R(t_1, t_2) = A(t_1, t_2) - Y(t_1, t_2). \quad (5.8)$$

Разность между поступающей и обслуженной нагрузками определяет *избыточную нагрузку*

$$I(t_1, t_2) = Z(t_1, t_2) - Y(t_1, t_2). \quad (5.9)$$

5.1.2. Соотношения между видами нагрузки

Наличие такого числа видов нагрузки обуславливает возникновение резонного вопроса: как они соотносятся между собой? Оказывается, эти соотношения различны для разных типов потоков вызовов.

Так, для пуассоновских потоков без последействия и потоков с ограниченным последействием интенсивность потенциальной и поступающей нагрузок численно совпадают:

$$A = Z. \quad (5.10)$$

Это объясняется тем, что при большой численности источников нагрузки доля нагрузки одного источника с учетом его активности или пассивности при условии обслуживания с отказами очень мала. Поэтому для простейшего потока поступающая нагрузка равна потенциальной нагрузке, следовательно, в этом случае упрощаются расчеты и достаточно лишь вести разговор о трех видах нагрузки: *поступающей, обслуженной и потерянной*.

В отличие от простейшего потока вызовов при использовании модели примитивного потока в системе с потерями и ограниченном числе источников нагрузки интенсивность поступающей нагрузки всегда выше потенциальной:

$$Z > A. \quad (5.11)$$

Дело в том, что для потока с простым последействием интенсивность зависит от состояния системы обслуживания и различна в системах с потерями и без потерь. Так, при одинаковой интенсивности вызовов источника в свободном состоянии, интенсивность по-

тока в системе с потерями всегда выше, чем в системе без потерь. Обусловлено это прежде всего тем, что источник, получив отказ в соединении, становится свободным и наряду с остальными свободными источниками может посылать новые вызовы (чем меньше источников, занятых по причине отказа в обслуживании, тем больше λ_i).

В системе с повторением различают три вида поступающей нагрузки – *первичную, повторную и общую*. Интенсивность первично поступающей нагрузки совпадает с интенсивностью потенциальной нагрузки, если первичный поток вызовов является простейшим или с ограниченным последствием: $A = Z_{\text{перв. пост.}}$.

В случае аппроксимации первичного потока вызовов моделью примитивного потока интенсивность первично поступающей нагрузки меньше потенциальной нагрузки: $A > Z_{\text{перв. пост.}}$.

Объясняется это тем, что часть источников переходит в состояние повторения и новые первичные вызовы в это время ими не создаются.

Кроме того, в системе с повторением при отсутствии явных потерь сообщений обслуженная нагрузка может быть выше потенциальной за счет частичного обслуживания вызовов, потерянных на последующих этапах соединения.

В системе с ожиданием, наоборот, находясь в очереди, источник лишен возможности посылать новые вызовы, следовательно, интенсивность потока вызовов снижается по сравнению с системой без потерь. Поэтому интенсивность поступающей нагрузки становится меньше интенсивности потенциальной: $Z < A$.

5.2. Колебания нагрузки и их причины

Интенсивность нагрузки в общем случае различна. В разные или в одни и те же часы суток, но в разные дни наблюдениями установлено, что наряду со случайными колебаниями интенсивности нагрузки по часам суток, дням недели и месяцам года существуют и периодические, относительно регулярные колебания, которые необходимо учитывать при прогнозировании нагрузки. Из регулярных колебаний нагрузки наиболее значительными являются колебания по часам суток (рис. 5.1).

В значительной степени они зависят от следующих факторов:

- распорядка жизни в городе (начало и конец рабочего дня, время обеденного перерыва);
- структурного состава абонентов (доля телефонных аппаратов квартирного, народно-хозяйственного и административного секторов), включенных в АТС;
- программы телевидения (во время концертов, кинофильмов и спортивных программ нагрузка снижается);
- дня недели (наибольшая интенсивность утренней нагрузки обычно наблюдается в понедельник и вторник, для вечерней нагрузки лидером становится пятница);
- число месяца (в начале месяца интенсивность нагрузки несколько ниже, в конце – выше);
- месяца года (в зависимости от местности в летние месяцы нагрузка может снижаться или, наоборот, повышаться, в другие месяцы года интенсивность нагрузки также меняется, обычно повышаясь к концу квартала, года).

У, Эрл

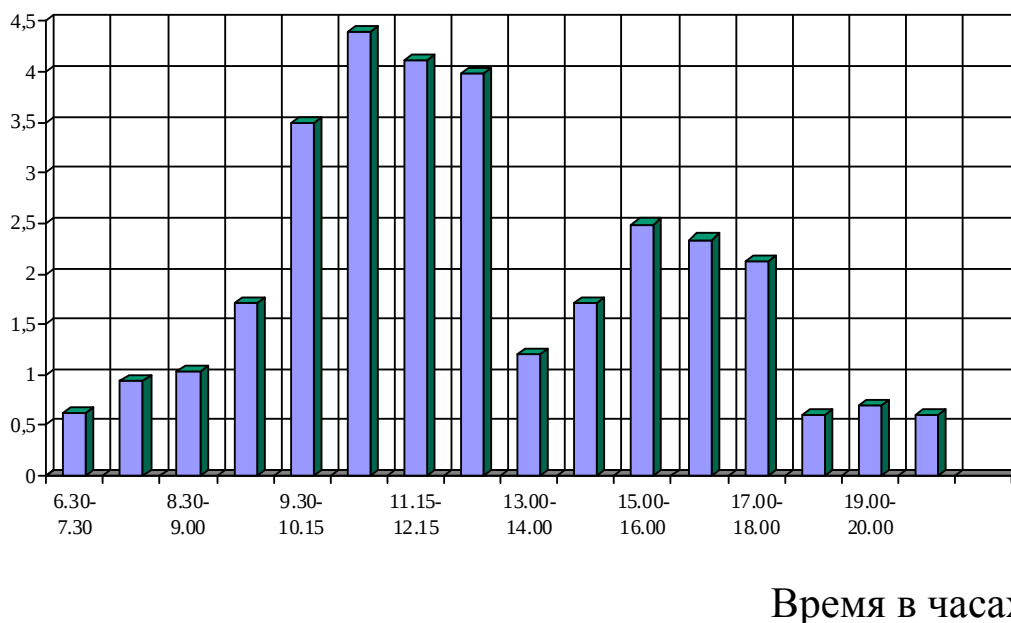


Рис. 5.1. Изменение величины нагрузки в течение суток

Для удовлетворения качества обслуживания абонентов в любое время суток расчет объема оборудования необходимо выполнять из

значения интенсивности нагрузки в тот час, когда она является наибольшей. Этот час называется часом наибольшей нагрузки (ЧНН).

Час наибольшей нагрузки – это непрерывный интервал времени в 60 мин, в течение которого средняя интенсивность нагрузки является наибольшей.

Степень концентрации нагрузки в ЧНН оценивается коэффициентом концентрации нагрузки:

$$K_{\text{ЧНН}} = \frac{Y_{\text{ЧНН}}}{\sum_{i=1}^{24} Y_i}, \quad (5.12)$$

где $Y_{\text{ЧНН}}$ – величина нагрузки за ЧНН;

Y_i – величина нагрузки в i -й час.

Величина коэффициента концентрации в основном зависит от структурного состава абонентов АТС и лежит в пределах 0,09 – 0,15.

Для систем военного назначения [28] $K_{\text{ЧНН}}$ колеблется от 0,08 до 0,12 со средним значением 0,1.

Чтобы объем оборудования был минимальным и загрузка его равномерной, величина коэффициента концентрации должна быть как можно меньше.

Уменьшить $K_{\text{ЧНН}}$ можно только путем правильного объединения абонентов квартирного и народно-хозяйственного секторов по абонентским блокам АТС.

Так, по данным Московской ГТС, коэффициент концентрации телефонной нагрузки имеет минимальную величину при доле абонентов народно-хозяйственного сектора в емкости АТС до 20 – 30 %.

При исследовании суточного распределения интенсивности нагрузки, кроме ЧНН, выделяют непрерывные промежутки времени длиной 3 ч с наибольшей интенсивностью нагрузки (ПНН):

$$K_{\text{ПНН}} = \frac{\sum_{r=1}^3 Y_r}{\sum_{i=1}^{24} Y_i}. \quad (5.13)$$

Для большинства станций местных и междугородных сетей характерно наличие трех периодов наибольшей интенсивности нагруз-

ки: утреннего (9.00 – 12.00), послеобеденного (14.00 – 17.00) и вечернего (19.00 – 22.00). По опыту эксплуатации АТС, коэффициент $K_{ПНН} = 0,2 - 0,4$.

Представляет также интерес отношение интенсивности нагрузки в ЧНН и ПНН: $\frac{Y_{ЧНН}}{Y_{ПНН}}$. По данным измерений, оно составляет 1,05 – 1,3.

5.3. Измерение нагрузки

Эффективность капитальных вложений в развитие сетей связи и качество их функционирования существенным образом зависят от регулярности измерений, правильности обработки и толкования статистических данных параметров нагрузки.

5.3.1. Цели, задачи, методы измерения нагрузки

В соответствии с Рекомендацией Е.502 МККТТ наблюдение (измерение) за нагрузкой осуществляется в целях:

- контроля качества работы станции (сети), т. е. регистрации каждого установленного и неустановленного соединений и автоматического ведения журнала учета соединений;
- управления функционированием станции (сети);
- технического обслуживания станции (выявления неисправности отдельных узлов станции);
- прогнозирования нагрузки для определения числа каналов на различных направлениях связи при проектировании станций и сетей.

Одной из целей измерения нагрузки является практическая проверка основных положений теории телетрафика.

При организации измерений параметров нагрузки в соответствии с поставленной целью необходимо решить следующие вопросы:

- установить объекты измерений;
- определить период измерений;
- установить продолжительность и величину допустимой ошибки измерений.

Измерения должны включать следующие действия:

- регистрацию данных (по запросу персонала) в соответствующих счетчиках;

- формирование отчетов, базирующихся на информации, извлеченной из счетчиков;
- формирование отчетов, содержащих данные по структуре потока вызовов, нагрузке и среднему времени занятия в сравнении с пороговыми значениями.

Измерения должны выполняться и записываться автоматически во время нормальной работы станции и быть обеспечены по объектам следующих типов:

- абонентским линиям;
- группам абонентских линий;
- каналам связи;
- направлениям связи;
- управляющим устройствам;
- вспомогательным устройствам.

Измерению должны быть подвержены вызовы:

- внутристанционные;
- исходящие, поступающие к абонентам других АМТС;
- входящие, поступающие от абонентов других АМТС;
- транзитные;
- входящие и исходящие к (от) РМО МК.

Важным элементом любого измерения нагрузки, потерь и т.п. является выбор метода измерения.

При выборе периода измерений необходимо учитывать, что в основе расчета объема оборудования станций и сети в целом лежит нагрузка в ЧНН. Определяют ЧНН с точностью до 15 мин. С этой целью интенсивность нагрузки регистрируют каждые 15 мин и ее значения за каждые четыре соседних промежутка заносят в специальную таблицу (при ручном способе обработки измерений).

В качестве оценки нагрузки в ЧНН можно брать данные, полученные в результате использования различных методов измерения или определенные на основе статистики.

Первый метод основывается на измерении нагрузки в течение предполагаемого часа наибольшей нагрузки (в англоязычной литературе – *busy hour*), выбранного на основе апостериорных данных, полученных в ходе эксплуатации станции или ранее проведенных испытаний [29]. Ориентировочно интервал времени, соответствующий ЧНН, повторяется каждые сутки с 11 до 12 ч.

Во время предполагаемого ЧНН производятся измерения нагрузки (с помощью счетчиков числа занятий, минут занятий) в течение определенного времени. Период измерения, как правило, не задается. Полученные данные по нагрузке Z_i за период наблюдения (n дней измерения) усредняются по формуле

$$\bar{Z} = \frac{1}{n} \cdot \sum_{j=1}^n Z_j. \quad (5.14)$$

При этом полученное среднее арифметическое считается несмещенной, эффективной и состоятельной оценкой исследуемого параметра только при использовании достаточно большой выборки. Данная интенсивность поступающей нагрузки обычно выражена в Эрлангах.

Недостатки данного метода заключаются в следующем.

Во-первых, по причине случайного характера потоков вызовов и нагрузки, обусловленного рядом объективных и субъективных причин, возможны случаи, когда предполагаемый и реальный ЧНН не совпадают по времени, в результате чего измеренное значение нагрузки (к примеру, с 10.00 до 11.00) не будет иметь максимальное значение.

Во-вторых, не определен период измерения.

Таким образом, при усреднении значений нагрузки по формуле (5.14) отдельные ее члены не будут отражать нагрузку в ЧНН. В этом случае число требуемых каналов в направлениях связи, определяемое как функция, аргументами которой являются интенсивность нагрузки (5.14) и требуемое качество обслуживания (вероятность потерь), в определенные периоды времени будет недостаточным для обслуживания нагрузки с заданным качеством.

Следующий метод получил название "классический" [29].

Он основан на концепции статистического часа наибольшей нагрузки. В отличие от первого метода, созданного на предполагаемом ЧНН, в данном методе ЧНН определяется, исходя из полученных и обработанных данных статистики.

Согласно данному методу, необходимо выполнить следующие процедуры:

– для определения нагрузки в ЧНН в течение периода наибольшей нагрузки (ПНН) делаются пять часовых измерений со сдви-

гом в 15 мин и в каждом из них рассчитывается интенсивность нагрузки;

- выбираются максимальные значения интенсивности нагрузки среди пяти измеренных значений Z_i , $i = 1, \dots, 5$, и считается, что данный астрономический час является ЧНН;

- все найденные максимальные значения интенсивности нагрузки усредняются за все дни наблюдений (n) путем скользящего суммирования средних значений нагрузки за четыре последовательных 15-минутных интервала:

$$Z_1 = \frac{1}{n} \cdot \sum_{j=1}^n \max_{1 \leq i \leq 5} Z_{i,j}. \quad (5.15)$$

Таким образом, полученные суточные значения нагрузки будут всегда иметь наибольшие значения, однако выбранный час может не совпадать с реальным ЧНН.

В этом случае значение интенсивности нагрузки будет превышать полученное на основе первого метода, что в результате расчета потребует большего числа каналов связи. Следовательно, во время ЧНН в данном направлении будет меньше потерянных либо ожидающих вызовов, но с учетом того, что коэффициент концентрации для военных сетей составляет $K = 0,08 - 0,15$, эффективность использования оборудования в оставшееся время ($0,85 - 0,92$ %) будет значительно ниже.

Третий метод основан на классическом подходе, но порядок снятия измерений уточнен Рекомендацией Е.500 МККТТ.

При данном методе определение ЧНН осуществляется так же, как и по классическому методу, посредством анализа пяти часовых измерений, сдвинутых относительно друг друга на 15 мин.

Однако данный метод регламентирует количество дней, необходимых для проведения измерений интенсивности нагрузки. При этом сама нагрузка делится на два уровня:

- A – нормальная интенсивность нагрузки в ЧНН;
- B – повышенная интенсивность нагрузки в ЧНН.

Для междугородных станций:

- интенсивность нагрузки уровня A определяют как среднее значение за 10 наиболее нагруженных дней 12-месячного периода;
- интенсивность нагрузки уровня B получают, усредняя данные за 5 самых нагруженных дней 10-дневного периода нормальной нагрузки.

Для АТС местных сетей:

- при определении значения интенсивности нагрузки уровня A выбираются 30 наиболее нагруженных дней 12-месячного периода;
- при определении интенсивности нагрузки уровня B выбираются 5 самых нагруженных дней 30-дневного периода нормальной нагрузки.

Кроме этого, отличие данного метода заключается в порядке усреднения полученной статистики.

Вначале вычисляются средние арифметические значения нагрузки по первому часу, например с 9.00 до 10.00 за весь период измерений, затем по второму часу – 9.15 – 10.15, затем по третьему – 9.30 – 10.30, далее по четвертому – 9.45 – 10.45, по пятому – 10.00 – 11.00. В качестве оценки нагрузки в ЧНН берется максимальное значение интенсивности нагрузки среди пяти значений рассчитанных арифметических средних:

$$Z_1 = \frac{1}{n} \cdot \max_{1 < i < 5} \left(\sum_{j=1}^n Z_{i,j} \right). \quad (5.16)$$

Рассмотренные способы обработки априори предполагают статистическую однородность зарегистрированных данных. Поэтому результаты регистрации нагрузки в праздничные, выходные дни и во время стихийных бедствий рекомендуют исключать и обрабатывать отдельно.

Характерной особенностью ведомственных сетей, в том числе сетей специальной связи, являются значительные колебания нагрузки во времени. Поэтому регистрация нагрузки, полученная в разные дни, статистически однородной заведомо считаться не может. В связи с этим возникает необходимость определения такой совокупности дней, в которые распределение во времени нагрузки можно считать статистически однородными, а затем с учетом полученной типологии определять число каналов связи, необходимых для обслуживания нагрузки с заданном качеством.

При проведении измерений также следует помнить, что их продолжительность зависит от числа измерений, которые, в свою очередь, определяются допустимой величиной ошибки измерений. Из математической статистики известно, что при измерении случайных величин число измерений находится в обратной зависимости от квадрата величины допустимой ошибки.

Например для увеличения точности результатов в два раза, число измерений необходимо увеличить в четыре раза. Поэтому не следует требовать от измерений большой точности, чем это нужно для решения поставленной задачи.

Следует также остановиться на измерении трафика в сетях, когда сообщения представлены пакетами символов, которые могут иметь различную длину и передаваться друг за другом по каналам связи, соединяющим между собой узлы сети – устройства объединения, ответвления и перенаправления пакетов.

При этом можно измерять исходящий трафик из некоторого сетевого элемента независимо от получателя или входящий трафик в выбранный сетевой элемент, откуда бы он ни исходил.

Измерение трафика при этом сводится к подсчету числа пакетов, проходящих по выбранному пути, и измерению длины каждого из пакетов. Обозначим число пакетов, прошедших по выбранному пути, буквой n , длину i -го пакета – l_i (бит), скорость передачи информации по каналу – R (бит/с), а время измерения – буквой T (секунд).

Тогда среднюю за время T интенсивность трафика на данном интерфейсе можно определить как

$$Y = \frac{\sum_i \frac{l_i}{R}}{T} = \frac{1}{R} \cdot \frac{\sum_i l_i}{T}. \quad (5.17)$$

Однако в практике пакетных сетей часто оценивают интенсивность трафика величиной $I = \frac{\sum_i l_i}{T} = Y \cdot R$ (бит или байт в секунду).

При постоянной длине пакета интенсивность трафика характеризуют просто количеством пакетов в секунду. Иногда даже при переменной длине пакета вначале вычисляют его среднюю длину, а затем пользуются описанием интенсивности трафика как числа пакетов в секунду.

Таким образом, в пакетных сетях интенсивность трафика измеряется не тем, насколько быстро перемещаются пакеты, а насколько "плотно" они движутся.

Например, если интенсивность трафика равна 30 пакетам в секунду при средней длине пакета 1000 бит, то для сети *Ethernet*, где

скорости передачи 10 Мбит/с, величина нагрузки за время $T = 1$ с $Y = \frac{30\,000}{10\,000\,000} = \frac{3}{1000} = 0,003$ Эрл.

Для телефонного модема, работающего на скорости 32 кбит/с ($R = 32\,000$ бит/с), $Y = \frac{30\,000}{32\,000} = \frac{30}{32} = 0,9375$ Эрл.

В последнем случае модем просто перегружен и задержки, возникающие при передаче, будут недопустимыми.

Как измерять интенсивность трафика в сетях ПД? Очевидно, что для этого необходимо регистрировать время появления каждого пакета и его длину.

Для проведения измерений подобного рода существуют специальные приборы – анализаторы пакетов, позволяющие выполнять гораздо более сложный анализ трафика: определять длину пакета, выделять ошибочные пакеты, систематизировать полученную информацию в виде графиков, таблиц и диаграмм.

5.3.2. Принципы измерения нагрузки

Измерения параметров нагрузки и потерь можно классифицировать:

- по способу получения исходных данных (автоматические и ручные);
- способу регистрации измеряемой величины (прямые и косвенные);
- способу организации процесса измерений (непрерывные, периодические и эпизодические (спорадические));
- охвату объектов изучаемой совокупности (сплошные и выборочные).

Ручные измерения допускается применять только при отсутствии автоматической аппаратуры, так как они являются слишком дорогостоящими и не обеспечивают необходимой точности.

Примером прямых измерений может служить регистрация числа вызовов с помощью счетчиков при измерении числа поступающих вызовов, а косвенных – измерение интенсивности обслуженной нагрузки путем регистрации через некоторые интервалы времени числа одновременно занятых приборов.

Непрерывные измерения являются дорогостоящими и обычно организуются:

- при измерении коэффициентов концентрации нагрузки;
- определении характера распределения нагрузки по часам суток, дням недели, месяцам года.

В большинстве случаев параметры нагрузки измеряются периодически, так как это значительно дешевле проведения непрерывных измерений, в то же время при выборе соответствующей длительности измерений обеспечивают требуемую точность.

Эпизодические (спорадические) измерения могут начинаться, например, при появлении симптомов неудовлетворительной работы оборудования.

Для экономии средств и затрат труда при наблюдениях за параметрами нагрузки применяются, как правило, не сплошные, а выборочные измерения.

В математической статистике вся подлежащая изучению совокупность однородных элементов называется генеральной, а часть генеральной совокупности, отобранная для измерения, – выборочной совокупностью.

При измерении нагрузки наибольшее распространение получили следующие три принципа:

- непрерывного измерения;
- сканирования;
- подсчета числа случайных событий.

Принцип непрерывного измерения нагрузки предполагает непрерывное измерение числа занятых каналов связи. В основе данного принципа лежит так называемый метод усреднения по времени.

В качестве оценки обслуженной нагрузки используется выражение

$$Y = \frac{1}{T} \cdot \int_0^T X(t) \cdot dt, \quad (5.18)$$

где $X(t)$ – число линий, занятых в момент t в V линейном пучке с потерями. $X(t)$ выражается функцией, совершающей единичные скачки между 0 и V .

Поясним реализацию данного принципа на примере схемы, представленной на рис. 5.2.

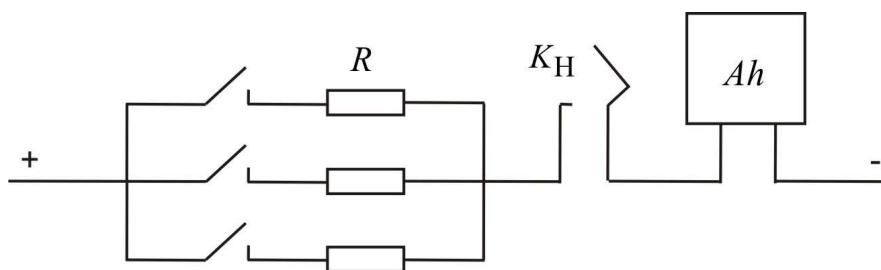


Рис. 5.2. Схема непрерывного измерения обслуженной нагрузки

Каждый прибор имеет измерительный резистор R , через который протекает ток, если данный прибор занят, а если он свободен, то ток через резистор не протекает. При этом величина тока определяется выражением

$$I(t) = \frac{U}{R} \cdot V(t), \quad (5.19)$$

где $V(t)$ – число занятых устройств в момент времени t ;

U – напряжение питания.

В приборе, измеряющем ампер-часы, ток интегрируется за весь период измерения T , при этом интенсивность обслуженной нагрузки составит величину

$$Y = k \cdot \int_0^T I(t) \cdot dt, \quad (5.20)$$

где k – градуировочная постоянная.

Принцип сканирования, или метод периодических отсчетов заключается в производстве измерений числа занятых линий в интервалы времени, отстоящие друг от друга на некотором расстоянии Δ (Δ – интервал сканирования).

Одним из часто употребляемых способов изображения процесса обслуживания заявок при использовании принципа сканирования является диаграмма Ганта. Она представляет собой прямоугольную систему координат, ось абсцисс которой изображает время, а на оси ординат помечают дискретные точки, соответствующие линиям или каналам связи. Далее каждый интервал времени, когда линия занята обслуживанием заявки входного потока, отмечается отрезком жирной горизонтальной прямой. На рис. 5.3 представлена диаграмма Ганта для системы с четырьмя линиями.

В первые два интервала времени Δt в системе заняты только первая и четвертая линии. В следующий интервал времени занята только одна, четвертая линия. В следующий интервал времени Δt заняты вторая и третья линии, в последующий интервал – все четыре линии, затем – первая, третья и четвертая линии, а в крайнем правом интервале Δt заняты первая и четвертая линии.

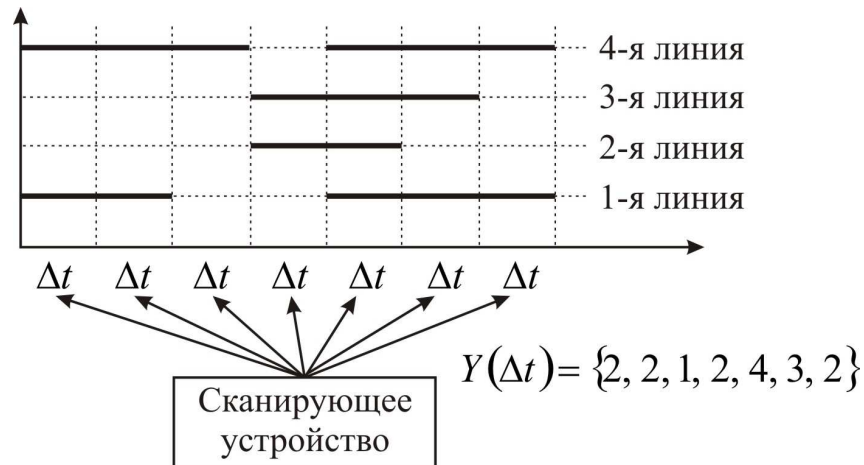


Рис. 5.3. Диаграмма Ганта для системы с четырьмя линиями

Сразу заметим, что построенная диаграмма Ганта отражает только обслуженный трафик (*traffic carried*), поскольку ничего не говорит о том, поступали ли в систему заявки, которые не смогли быть обслужены выделенными каналами связи. Объем трафика, т. е. интенсивность обслуженной нагрузки вычисляется как суммарная длина всех отрезков диаграммы Ганта.

Приведенная выше диаграмма позволяет найти объем обслуженного трафика за семь интервалов Δt :

$$Y(7 \cdot \Delta t) = 2 \cdot 2\Delta t + 1\Delta t + 2\Delta t + 4\Delta t + 3\Delta t + 2\Delta t. \quad (5.21)$$

Если интервал времени $\Delta t = 1$ с, то величина обслуженной нагрузки в эрлангосекундах или секундозанятиях

$$Y(t) = 2 \cdot 2 \cdot 1 + 1 \cdot 1 + 2 \cdot 1 + 4 \cdot 1 + 3 \cdot 1 + 2 \cdot 1 = 16 \text{ [Es]}. \quad (5.22)$$

В каждый интервал времени Δt имеется, так называемая мгновенная интенсивность обслуженной нагрузки $Y(\Delta t)$. Для примера, где дискретность измерения составляет $\Delta t = 1$ с, можно представить мгновенную интенсивность следующей последовательностью:

$$Y(\Delta t) = \{2, 2, 1, 2, 4, 3, 2\}. \quad (5.23)$$

Средняя за период в $7 \cdot \Delta t = 7$ с интенсивность обслуженной нагрузки

$$\bar{Y} = \frac{1}{7} \cdot \sum_{i=1}^7 Y(\Delta t) = \frac{2+2+1+2+4+3+2}{7} \approx 2,3 \text{ (Эрл)}. \quad (5.24)$$

Принцип подсчета числа случайных событий основан на приеме импульса при появлении каждого события.

Таким способом на современных АТС регистрируется число поступивших, обслуженных и потерянных вызовов при измерении потерь по вызовам, число случаев занятости всех линий пучка при измерении потерь по времени и др.

Этот способ может быть использован также при измерении нагрузки, обслуженной устройствами, время занятия которых постоянно (измерение нагрузки, обслуженной управляющими устройствами).

5.4. Способы математического задания потоков нагрузки на межстанционных участках

Величины потоков телефонной нагрузки полностью определяются взаимной заинтересованностью в телефонной связи абонентов разных станций, поэтому при проектировании АТС точно установить величины межстанционных потоков нагрузки невозможно. Это можно сделать только после введения АТС в действие путем специальных наблюдений за потоками нагрузки.

Как известно, наблюдения за интенсивностями потоков нагрузки осуществляются периодами по 15 мин. Временные положения промежутков продолжительностью в 1 ч, начинающихся со сдвигом в 15 мин, обозначим моментами времени $t_1, t_2, t_3, \dots, t_r$, расположенными в середине каждого из промежутков (рис. 5.4).

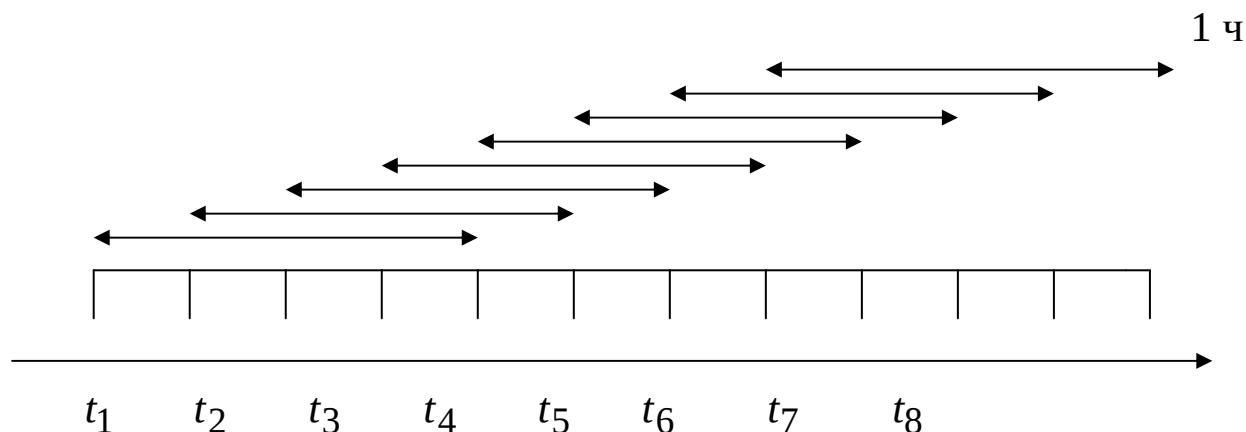


Рис. 5.4. Временные положения промежутков продолжительностью в 1 ч

Средние значения интенсивностей нагрузки от АТС_{*i*} к АТС_{*j*} обозначим вектором

$$Z_{i,j} = \{Z_{i,j}(t_1), Z_{i,j}(t_2), \dots, Z_{i,j}(t_r)\} \quad (5.25)$$

Если на направлении *ij* ЧНН приходится на *k*-й промежуток времени, то интенсивность нагрузки в ЧНН обозначим через $Z_{ij}^*(t_k)$.

При *m* АТС на сети интенсивности межстанционных потоков нагрузки полностью характеризуются квадратной матрицей векторов:

$$\|Z_{ij}\| = \begin{pmatrix} Z_{11} & Z_{12} & Z_{13} & \dots & Z_{1m} \\ Z_{21} & Z_{22} & Z_{23} & \dots & Z_{2m} \\ \dots & \dots & \dots & \dots & \dots \\ Z_{m1} & Z_{m2} & Z_{m3} & \dots & Z_{mm} \end{pmatrix}. \quad (5.26)$$

На телефонных сетях имеют место объединение и разделение потоков нагрузки. При этом пользуются понятиями исходящей и входящей нагрузок.

Под исходящей от АТС_{*i*} нагрузкой понимают сумму потоков нагрузки, исходящих от АТС_{*i*} ко всем АТС сети. Вектор интенсивностей исходящей от АТС_{*i*} нагрузки определяется как сумма элементов *i*-й строки матрицы (5.26):

$$Z_{\text{исх}_i} = \sum_{j=1}^m Z_{ij}. \quad (5.27)$$

По правилу сложения векторов выражение (5.27) с учетом вектора (5.25) можно записать в виде

$$Z_{\text{исх}_i} = \left\{ \sum_{j=1}^m Z_{i,j}(t_1), \sum_{j=1}^m Z_{i,j}(t_2), \dots, \sum_{j=1}^m Z_{i,j}(t_r) \right\}. \quad (5.28)$$

Интенсивность исходящей нагрузки в ЧНН определяется из выражения

$$Z_{\text{исх}_i}^* = \max_k \left\{ \sum_{j=1}^m Z_{ij}(t_k), \quad k = 1, 2, \dots, r \right\}. \quad (2.29)$$

В частном случае при совпадении временных положений ЧНН на направлениях межстанционной связи

$$Z^*_{исх_i} = \sum_{j=1}^m Z^*_{ij}(t_k). \quad (5.30)$$

При поступлении нагрузки в j -й узел мы имеем дело с объединением нагрузки. Под входящей на АТС j нагрузкой понимают сумму потоков нагрузки, входящих на АТС j от всех АТС сети.

Вектор интенсивностей входящей на АТС j нагрузки определяется как сумма элементов j -го столбца матрицы:

$$Z_{вх_j} = \sum_{i=1}^m Z_{ij}. \quad (5.31)$$

Используя правило сложения векторов, по аналогии с выражением (5.28) можно записать, что

$$Z_{вх_j} = \left\{ \sum_{i=1}^m Z_{i,j}(t_1), \sum_{i=1}^m Z_{i,j}(t_2), \dots, \sum_{i=1}^m Z_{i,j}(t_r) \right\}. \quad (5.32)$$

Интенсивность входящей нагрузки в ЧНН определяется из выражения

$$Z^*_{вх_j} = \max_k \left\{ \sum_{i=1}^m Z_{ij}(t_k), \quad k = 1, 2, \dots, r \right\}. \quad (5.33)$$

Из выражений (5.30) и (5.33) следует, что при объединении и разделении потоков нагрузки в условиях совпадения временных положений ЧНН на всех направлениях межстанционной связи ij ($i = 1, 2, 3, \dots, m$) интенсивность результирующих потоков в ЧНН однозначно определяются через интенсивность в ЧНН составляющих их потоков, т. е. вместо матрицы векторов (5.26) достаточно иметь матрицу скаляров.

При условии совпадения временных положений ЧНН на всех направлениях межстанционной связи ij ($i = 1, 2, 3, \dots, m$) интенсив-

ность суммарной нагрузки в ЧНН, исходящей от всех АТС сети, определяется из выражения

$$Z^*_{\text{исх}} = \sum_{i=1}^m Z^*_{\text{исх}i} = \sum_{i=1}^m \sum_{j=1}^m Z^*_{ij}, \quad (5.34)$$

а интенсивность суммарной нагрузки в ЧНН, входящей на все АТС сети, – из выражения

$$Z^*_{\text{вх}} = \sum_{j=1}^m Z^*_{\text{вх}j} = \sum_{j=1}^m \sum_{i=1}^m Z^*_{ij}. \quad (5.35)$$

Из сравнения формул (5.34) и (5.35) следует, что интенсивность суммарной нагрузки, исходящей от всех АТС сети, равна интенсивности суммарной нагрузки, входящей на все АТС сети.

5.5. Закономерности формирования потоков нагрузки

Закономерности формирования потоков нагрузки могут быть выявлены только путем наблюдений на действующих сетях. Наблюдениями установлено, что временное положение ЧНН на направлении ij существенным образом зависит от структурного состава абонентов АТС $_i$ и АТС $_j$.

Если эти АТС обслуживают преимущественно абонентов квартирного сектора, то имеет место вечерний ЧНН, если народно-хозяйственного сектора – то утренний ЧНН.

При прочих равных условиях величины интенсивности потоков нагрузки в ЧНН Z^*_{ij} ($i, j = 1, 2, \dots, m$) тем больше, чем территориально ближе расположены абоненты АТС $_i$ к абонентам АТС $_j$.

Так, величина интенсивности внутристанционной нагрузки АТС $_i$ при прочих равных условиях обычно бывает больше интенсивности потоков нагрузки к другим АТС.

Анализ закономерностей формирования абсолютных значений потоков нагрузки обычно выполнить достаточно сложно. Особенно это касается проектирования вновь вводимых АТС. Трудность прогнозирования различных коэффициентов состоит в том, что они зави-

сят от целого ряда факторов, которые определяют взаимное телефонное тяготение абонентов АТС_{*i*} к абонентам АТС_{*j*}.

Тем не менее при проектировании часто используют отношения интенсивности нагрузки в конкретном направлении межстанционной связи к интенсивности нагрузки, исходящей от АТС:

$$k_{ij} = \frac{Z_{ij}}{Z_{\text{исх}_i}} \quad (i, j = 1, 2, \dots, m). \quad (5.36)$$

Данное отношение называют коэффициентами распределения нагрузки.

Очевидно, что

$$\sum_{j=1}^m k_{ij} = \sum_{j=1}^m \frac{Z_{ij}}{Z_{\text{исх}_i}} = 1. \quad (5.37)$$

Зная значения коэффициентов распределения нагрузки k_{ij} , можно определить интенсивность потоков нагрузки в направлениях связи от АТС_{*i*}:

$$Z_{ij} = k_{ij} \cdot Z_{\text{исх}_i} \quad (i, j = 1, 2, \dots, m). \quad (5.38)$$

Коэффициент распределения нагрузки от АТС_{*i*} к АТС_{*i*} k_{ii} также называют коэффициентом внутривыделенного сообщения.

Кроме коэффициента распределения нагрузки в современной литературе описаны различные коэффициенты для учета тяготения между абонентами. Рассмотрим только два из них:

- коэффициент тяготения;
- нормированный коэффициент тяготения.

Количественной оценкой телефонного тяготения между абонентами является коэффициент тяготения.

Для его определения предположим, что на сети реализовано равномерное телефонное тяготение между абонентами.

Пусть на телефонной сети емкостью N номеров имеется три станции: A , B и C . Емкость каждой из них равна N_a , N_b , N_c . Станции связаны друг с другом по принципу "каждая с каждой" односторонними линиями связи. Сообщение от станции A к станции C проходит по одному пучку соединительных линий, а от станции C к A – по другому.

Пусть среднее число занятий от одного абонента станции A равно c_a , на станции B — c_b , на станции C — c_c , а средняя длительность занятия соответственно — t_a , t_b , t_c . Тогда нагрузка в ЧНН по станциям A , B и C будет составлять:

$$\begin{aligned} Z_a &= N_a \cdot c_a \cdot t_a; \\ Z_b &= N_b \cdot c_b \cdot t_b; \\ Z_c &= N_c \cdot c_c \cdot t_c. \end{aligned} \quad (5.39)$$

Нагрузка всей сети: $Z = Z_a + Z_b + Z_c$.

Для определения нагрузки, поступающей со станции A на каждую из трех станций, можно воспользоваться следующими геометрическими построениями. Представим величину нагрузки станции A в некотором масштабе отрезком Z_a (рис. 5.5).

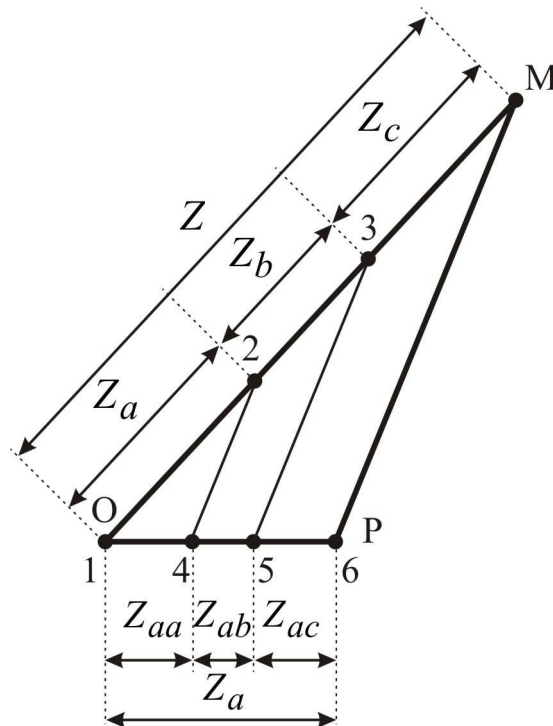


Рис. 5.5. График к определению потоков телефонных сообщений

Под произвольным углом к этому отрезку проведем прямую линию, выражающую в том же масштабе общую нагрузку сети Z . Соединим концы обеих линий прямой MP , в результате чего получим треугольник MOP . Отрезок OM , выражающий общую нагрузку сети Z , разделим на части, соответствующие нагрузкам трех отдельных станций $Z_a; Z_b; Z_c$. Из точек 2 и 3 проведем линии, параллельные

МР, тогда отрезок ОР, выражающий нагрузку Z_a станции A , разделится на три отрезка, пропорциональные нагрузкам $Z_a; Z_b; Z_c$ станций A, B, C . Нагрузка, поступающая со станции A к абонентам своей станции и к абонентам станций B и C , будет пропорциональна соответственно отрезкам 1 – 4, 4 – 5, 5 – 6.

Из треугольника МОР имеем

$$\frac{Z}{Z_a} = \frac{Z_a}{Z_{aa}}, \quad (5.40)$$

откуда

$$Z_{aa} = \frac{Z_a \cdot Z_a}{Z}. \quad (5.41)$$

Из соотношения $\frac{Z}{Z_b} = \frac{Z_a}{Z_{ab}}$ получим

$$Z_{ab} = \frac{Z_a \cdot Z_b}{Z}. \quad (5.42)$$

Из соотношения $\frac{Z}{Z_c} = \frac{Z_a}{Z_{ac}}$ получим

$$Z_{ac} = \frac{Z_a \cdot Z_c}{Z}. \quad (5.43)$$

Из соотношений (5.41) – (5.43) можно записать формулу нагрузки, поступающей со станции i на станцию j в общем виде:

$$Z_{ij} = \frac{Z_i \cdot Z_j}{Z}, \quad (5.44)$$

где Z_i – нагрузка станции i ; Z_j – нагрузка станции j ; Z – нагрузка всей сети.

В различных источниках можно увидеть формулу (5.44) в несколько ином виде, при этом ее сущность не изменяется:

$$Z'_{ij} = Z_{\text{исх}i} \cdot \frac{Z_{\text{исх}j}}{\sum_{j=1}^m Z_{\text{исх}j}}. \quad (5.45)$$

Как видно из формулы (5.45), интенсивность нагрузки от АТС_{*i*} к АТС_{*j*} Z'_{ij} пропорциональна доле интенсивности нагрузки, исходящей от АТС_{*j*}, в суммарной интенсивности нагрузки, исходящей от всех АТС сети (в случае использования односторонних линий связи).

Просуммировав выражения (5.41) – (5.45), получим

$$Z_a = \frac{Z_a \cdot Z_a}{Z} + \frac{Z_a \cdot Z_b}{Z} + \frac{Z_a \cdot Z_c}{Z}. \quad (5.46)$$

Расчет потоков трафика на практике усложняется тем, что в действительности тяготение станций друг к другу не бывает равномерным, поэтому выведенные формулы нуждаются в поправках, учитывающих эту неравномерность.

Неравномерность в телефонном сообщении возникает вследствие различной заинтересованности или различного тяготения одних абонентов к другим. К АТС, расположенной в центре города, как правило, тяготение больше, нежели к АТС, расположенной на окраине города. Это создает неравномерность в распределении телефонной нагрузки по сети.

Неравномерность в тяготении можно учесть, введя в уравнение (5.46) некоторые множители f_i , которые носят название коэффициентов (факторов) тяготения:

$$Z_a = f_{aa} \cdot \frac{Z_a \cdot Z_a}{Z} + f_{ab} \cdot \frac{Z_a \cdot Z_b}{Z} + f_{ac} \cdot \frac{Z_a \cdot Z_c}{Z}, \quad (5.47)$$

где f_{aa} – коэффициент внутреннего тяготения;

f_{ab} – коэффициент тяготения станции *A* к станции *B*;

f_{ac} – коэффициент тяготения станции *A* к станции *C*.

С учетом коэффициентов тяготения выражения (5.41) – (5.44) можно представить в следующем виде:

$$\begin{aligned} Z_{aa} &= f_{aa} \cdot \frac{Z_a \cdot Z_a}{Z}; \\ Z_{ab} &= f_{ab} \cdot \frac{Z_a \cdot Z_b}{Z}; \\ Z_{ac} &= f_{ac} \cdot \frac{Z_a \cdot Z_c}{Z}. \end{aligned} \quad (5.48)$$

В общем виде выражение (5.48) можно представить как

$$Z_{ij} = f_{ij} \cdot Z_{исх\ i} \cdot \frac{Z_{исх\ j}}{\sum_{j=1}^m Z_{исх\ j}}, \quad (5.49)$$

а затем из него выразить коэффициенты тяготения следующим образом:

$$\begin{aligned} f_{aa} &= \frac{Z_{aa} \cdot Z}{Z_a \cdot Z_a}; \\ f_{ab} &= \frac{Z_{ab} \cdot Z}{Z_a \cdot Z_b}; \\ f_{ac} &= \frac{Z_{ac} \cdot Z}{Z_a \cdot Z_c}. \end{aligned} \quad (5.50)$$

В общем случае для АТС с односторонними линиями коэффициенты тяготения можно выразить как

$$f_{ij} = \frac{Z_{ij}}{Z'_{ij}} = \frac{Z_{ij}}{Z_{исх\ i} \cdot Z_{исх\ j}} \cdot \sum_{j=1}^m Z_{исх\ j}. \quad (5.51)$$

По формуле (5.51) можно рассчитать коэффициенты тяготения между станциями i и j . Однако сделать это можно только для действующей сети, если в результате измерений определена входящая в формулу интенсивность нагрузок, но это невозможно для проектируемой сети, так как неизвестна нагрузка Z_{ij} .

На ГТС коэффициенты тяготения могут принимать различные значения, изменяясь в широких пределах от самых малых величин (0,4) до довольно больших (около 10 – 20).

На основании опытных данных можно отметить некоторые особенности коэффициентов тяготения:

- при неизменной емкости станции и росте емкости сети внутренний коэффициент тяготения станции повышается;
- с увеличением емкости станции внутренний коэффициент тяготения снижается;
- тяготение между абонентами периферийных станций сети, находящимися в разных административных районах города, невелико, а коэффициенты тяготения между такими станциями имеют малую величину;

– тяготение окраинных станций к станциям центра города сильнее, чем станций центра к окраинным, но обычно не превышает единицы;

– тяготение между смежными станциями больше, чем между несмежными;

– коэффициенты различаются тем больше, чем больше общее число станций в городе, т. е. чем крупнее сеть.

Кроме коэффициентов тяготения, часто используются и другие так называемые нормированные коэффициенты тяготения, которые определяются следующим образом:

$$n_{ij} = \frac{f_{ij}}{f_{ii}} = \frac{Z_{ij} \cdot Z_{исх_i}}{Z_{ii} \cdot Z_{исх_j}}. \quad (5.52)$$

Значение нормированного коэффициента тяготения существенным образом зависят от расстояния между абонентами (рис. 5.6).

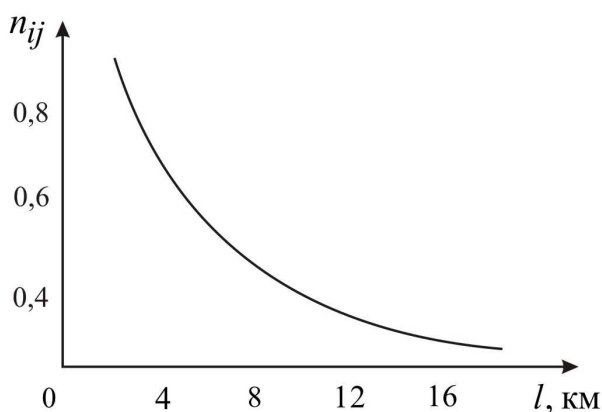


Рис. 5.6. Зависимость нормированных коэффициентов тяготения от кратчайшего расстояния между АТС

В заключение следует отметить, что какие бы коэффициенты не применялись, прогнозирование значений коэффициентов может осуществляться только на основе наблюдений за их значениями на действующих сетях связи.

Вопросы для самоконтроля

1. Что такое нагрузка и каковы ее виды?
2. Дайте определение обслуженной нагрузки и сформулируйте теорему о ее количественной оценке.
3. Дайте определение потенциальной нагрузки.

4. Дайте определение поступающей нагрузки.
5. Дайте определение потерянной и избыточной нагрузок.
6. В каких единицах измеряются нагрузка и интенсивность нагрузки?
7. Перечислите причины изменения (колебания) нагрузки в течение дня, недели, месяца.
8. В какой период времени осуществляется измерение нагрузки для определения необходимого объема оборудования?
9. Что понимают под часом наибольшей нагрузки (ЧНН)?
10. Что такое коэффициент концентрации нагрузки, как он вычисляется?
11. Пределы изменения коэффициента концентрации нагрузки на сетях связи (в том числе военных).
12. Что такое период наибольшей интенсивности нагрузки и в каких пределах он варьирует?
13. Каковы основные цели измерения параметров нагрузки?
14. Назовите основные задачи, решаемые в процессе измерения трафика.
15. Перечислите основные вопросы, которые необходимо учесть при организации измерений.
16. Назовите методы измерения (виды) нагрузки.
17. Что подразумевается под интенсивностью нагрузок уровней А и В?
18. Как классифицируются принципы измерения?
19. В чем сущность принципа непрерывного измерения?
20. В чем сущность принципа сканирования?
21. В чем сущность принципа подсчета числа случайных событий?
22. Что понимают под исходящей от АТС_і нагрузкой?
23. Каков порядок расчета интенсивности исходящей нагрузки в ЧНН?
24. Каков порядок расчета интенсивности исходящей нагрузки при совпадении временных положений ЧНН на направлениях межстанционной связи?
25. Что понимают под входящей на АТС_ј нагрузкой?
26. Каков порядок расчета интенсивности входящей нагрузки в ЧНН?

27. Каков порядок расчета интенсивности входящей нагрузки при совпадении временных положений ЧНН на направлениях межстанционной связи?

28. Что понимают под коэффициентами распределения нагрузки?

29. Что такое коэффициент внутристанционного сообщения?

30. Что такое коэффициент тяготения?

31. Что такое нормированный коэффициент тяготения?

6. КАЧЕСТВО ОБСЛУЖИВАНИЯ В ТЕЛЕФОННЫХ СЕТЯХ СВЯЗИ

6.1. Качество обслуживания (QoS), определяемое Рекомендациями Международного союза электросвязи E.800 (МСЭ-Т E.800)

Словосочетание "качество обслуживания" часто встречается не только в отечественной технической литературе, но и в публикациях на английском языке (*Quality of Service – QoS*). Авторы книг, статей и официальных документов на обоих языках употребляют эти слова при описании различных аспектов функционирования телефонных сетей. Ряд специалистов руководствуется Рекомендациями Международного союза электросвязи E.800 (МСЭ-Т E.800), которые посвящены терминологии в области качества обслуживания.

Рекомендации МСЭ-Т E.800 разработаны более десяти лет назад и утверждены 23 сентября 2008 года. Это означает, что предлагаемые термины многими организациями, участвующими в работе МСЭ, считаются устоявшимися и применимы ко всем телекоммуникационным услугам и всем видам оборудования, которое используется в сетях электросвязи.

Основным понятием считается качество обслуживания. Оно рассматривается МСЭ как результат совместного проявления характеристик обслуживания. Этот результат определяет степень удовлетворенности пользователя предоставленной ему услугой. Вместе с тем МСЭ не рекомендует использовать термин "качество обслуживания" ни для сравнительной оценки, ни для каких-либо количественных соотношений.

На рис. 6.1, заимствованном из Рекомендаций МСЭ-Т E.800, показана модель, которая определяет компоненты качества обслуживания и их взаимные связи. Пунктирная линия делит рисунок на две части: в верхней приведены основные характеристики качества обслуживания, в нижней – качество работы сети. Во всех блоках указаны только названия на русском языке.

Большинство пользователей не представляют себе ни принципов операторской деятельности, ни работу инфокоммуникационной

системы. Они, используя терминалы, считают, что обращаются к некому поставщику инфокоммуникационных услуг. Степень удовлетворенности уровнем обслуживания может оцениваться такими характеристиками:

- обеспечение обслуживания (*service support*);
- управляемость обслуживания (*service operability*);
- возможности обслуживания (*serviceability*);
- безопасность обслуживания (*service security*).

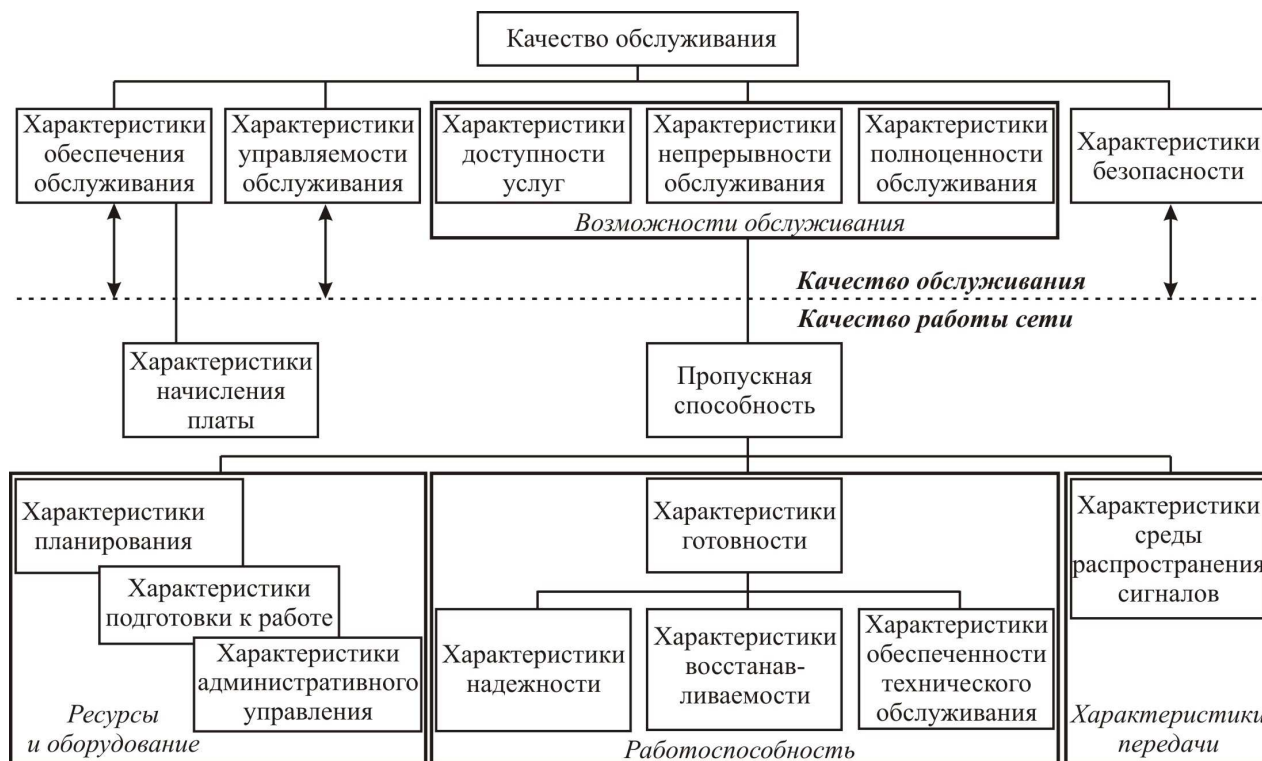


Рис. 6.1. Модель МСЭ по терминам в области качества обслуживания

Характеристики обеспечения обслуживания отражают способность оператора предоставлять услуги и способствовать их использованию.

Характеристики управляемости обслуживания оценивают удобство и простоту пользования услугами.

Возможности обслуживания, в свою очередь, делятся на три группы, для которых в Рекомендациях МСЭ-Т Е.800 предлагаются такие характеристики:

- доступность услуг (*service accessibility*);
- непрерывность обслуживания (*service retainability*);
- полноценность обслуживания (*service integrity*).

Характеристики доступности услуг оценивают возможность их получения (с заранее специфицированными допусками и с соблюдением других заданных условий) по запросу пользователя. Характеристики непрерывности обслуживания определяют возможность пользования полученной услугой с заданными атрибутами в течение запрошенного интервала времени. Характеристики полноценности обслуживания – это меры того, что обслуживание, будучи полученным, происходит без значительного ухудшения.

В прежней редакции Рекомендаций МСЭ характеристики безопасности отсутствовали. Они связаны со следующими аспектами функционирования сети связи: несанкционированный мониторинг, жульническое использование, злонамеренное повреждение, неправильное применение, ошибка человека, стихийное бедствие.

Все перечисленные характеристики обслуживания зависят от качества работы сети, а также от ее функциональных возможностей. Соответствующие связи также показаны на рис. 6.1.

Характеристики начисления платы (*Charging Performance*) оцениваются проще, чем в ряде других международных документов. Они определяются как вероятность корректного начисления платы с точки зрения вида связи, пункта назначения, времени суток и длительности соединения.

Характеристики пропускной способности (*Trafficability Performance*) определяют способность технических средств с известными свойствами обслуживать трафик с определенными параметрами. Эти характеристики разделены на три группы. Термины для первой группы – "Ресурсы и оборудование" – еще не определены. МСЭ считает, что соответствующая работа должна быть выполнена в ближайшее время.

Во вторую группу, названную "Работоспособность" (*Dependability*), входят следующие характеристики:

- готовность (*availability*) – способность технического средства выполнять требуемые функции в данный момент времени или в любой момент внутри заданного интервала времени (при наличии соответствующих внешних ресурсов, если они необходимы);

- надежность (*reliability*) – способность технического средства выполнять требуемые функции при заданных условиях в течение определенного интервала времени;

– восстанавливаемость (*maintainability*) – пригодность технического средства к тому, чтобы в установленных условиях его использования техническое обслуживание с применением установленных процедур и ресурсов обеспечивало поддержание или восстановление такого состояния этого средства, в котором оно может выполнять требуемые функции;

– обеспеченность техобслуживанием (*maintenance support*) – способность оператора при заданных правилах технического обслуживания предусмотреть и, если нужно, задействовать ресурсы, необходимые для поддержания работоспособности определенного технического средства.

К третьей группе относятся характеристики передачи (*Transmission Performance*). Они определяются как уровень воспроизведения сигнала, переданного через систему связи, которая находится в работоспособном состоянии, при заданных условиях. В Рекомендации МСЭ Е.800 выделены характеристики среды распространения сигналов (*propagation performance*). Они определяются как способность этой среды обеспечивать прохождение сигнала с заданными допусками (в отношении шума, помех, колебаний уровня и прочих) без искусственного регулирования данного процесса.

Количественные показатели, соответствующие рассмотренным выше характеристикам, могут либо относиться к некоторому моменту времени (мгновенные значения), либо выражаться как среднее значение за какой-то временной интервал. Эти показатели могут быть связаны с событиями (повреждением, восстановлением), состоянием (хорошим, плохим, нерабочим) или действиями (операциями по техническому обслуживанию). Ниже приводятся некоторые примеры количественных показателей.

Первая группа примеров относится к характеристикам временных перерывов в обслуживании. Основным понятием для этих характеристик можно считать простой в обслуживании (*interruption; break of service*) – временная невозможность обеспечить обслуживание в течение периода, превышающего допустимый, обусловленная тем, что, по крайней мере, один из параметров, необходимых для нормального обслуживания, вышел за установленные для него пределы.

Возможные причины – повреждение технических средств, ухудшение характеристик передачи, слишком высокий спрос на обслуживание. К количественным показателям относятся:

- время между простоями (*time between interruptions*) – промежуток времени между окончанием одного простоя и началом следующего;
- длительность простоя (*interruption duration*) – продолжительность времени простоя;
- среднее время между простоями (*mean time between interruptions*) – математическое ожидание времени между простоями;
- средняя длительность простоя (*mean interruption duration*) – математическое ожидание времени простоя.

Вторая группа примеров связана с количественными показателями доступности услуг (*service accessibility*):

- вероятностью успешного доступа к услуге (*service access probability*) – вероятностью того, что нужная услуга будет получена пользователем (с заданными допусками и с соблюдением других условий) по его запросу;
- средней длительностью задержки получения доступа к услуге (*mean service access delay*) – математическим ожиданием длительности интервала времени между первым запросом пользователя на предоставление услуги и моментом, когда он получил к ней доступ и при этом обслуживание оказалось соответствующим заданным допускам и условиям;
- вероятностью получения неверного соединения (*misrouting probability*) – вероятностью того, что при правильно набранном номере пользователь получит соединение не с тем адресатом.

Ранее речь шла о необходимости соответствия (подобия) характеристик для взаимодействующих сетей. С точки зрения качества обслуживания, такая задача рассматривается в Рекомендации МСЭ Е.801. Она посвящена соглашению о качестве обслуживания (*service quality agreement* – SQA). Далее будет рассматриваться соглашение об уровне обслуживания, известное по аббревиатуре SLA. Его можно рассматривать как адаптацию SQA для сетей с коммутацией пакетов. Для разработки соглашений SQA в Рекомендации МСЭ Е.801 предлагается процедура, которая показана на рис. 6.2.

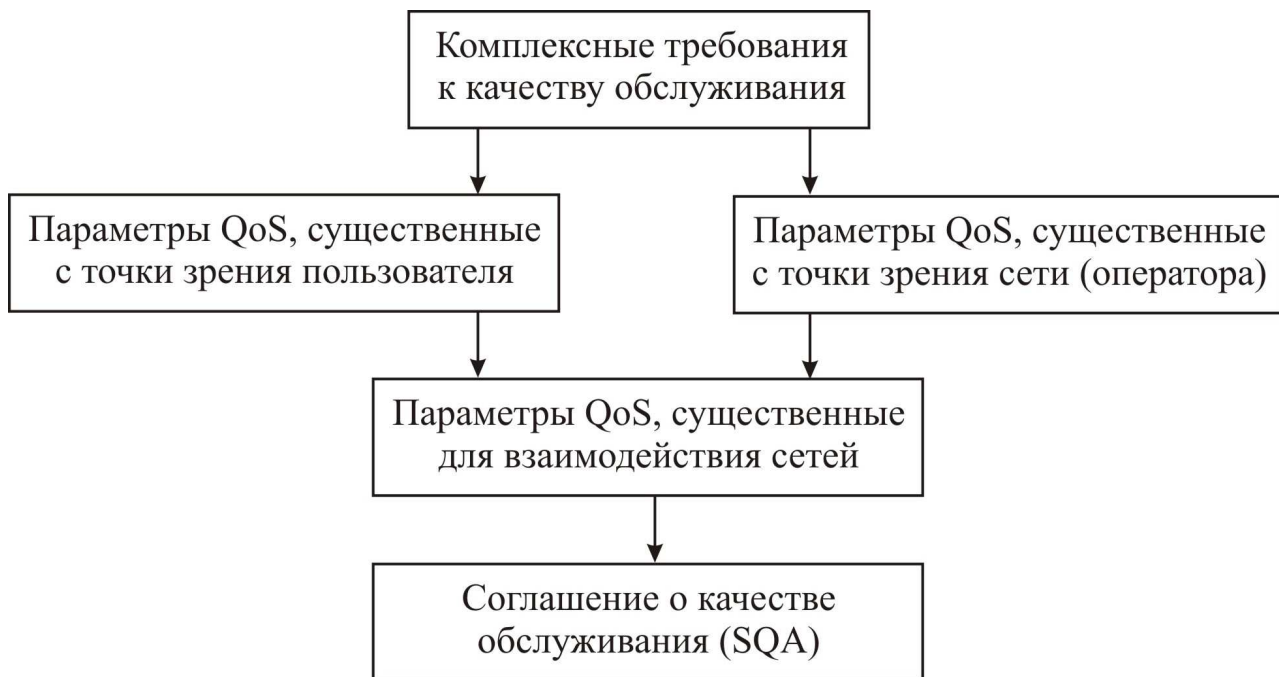


Рис. 6.2. Процедура разработки соглашения об уровне обслуживания

6.2. Сущность понятия «дисциплина обслуживания»

Под дисциплиной обслуживания [12] потоков телефонных вызовов имеется в виду характеристика алгоритма взаимодействия поступающего потока телефонных вызовов с коммутационной станцией, включающего в себя способ и порядок обслуживания телефонных вызовов.

Способ обслуживания телефонных вызовов – это характеристика процесса обслуживания телефонных вызовов коммутационной станцией или узлом.

При этом способ обслуживания телефонных вызовов может быть:

- без потерь;
- с потерями;
- с ожиданием;
- комбинированный.

Обслуживание без потерь – способ, при котором для передачи каждого сообщения немедленно предоставляется линия или канал. По экономическим соображениям реальные системы коммутации обычно проектируются с потерями.

Обслуживание с потерями – способ, при котором часть сообщений получает отказ в обслуживании или обслуживание их задер-

живается на некоторое время. Этим обусловлено наличие двух видов потерь: явные и условные потери сообщений.

Обслуживание с явными потерями предполагает, что сообщение и соответствующий ему вызов при получении отказа в немедленном соединении полностью теряются и на обслуживание в данную систему не поступают.

В качестве примера реализации данного способа можно привести обслуживание вызовов на сетях с обходными направлениями. Так, вызов, получивший отказ в обслуживании по причине занятости всех каналов направления первого выбора, поступает на пучок каналов направления второго выбора. При этом для направления первого выбора все сообщения, получившие отказ в соединении, будут явно потеряны.

Однако чаще модель обслуживания с явными потерями сообщений используется для упрощенного описания реальных систем с возможностью повторения вызовов. При малых потерях вызовов ($P < 0,01$) реализация этой модели дает вполне удовлетворительные для практики результаты.

Условные потери выражаются в задержке передачи сообщения сверх требуемого времени передачи и имеют место в системах с ожиданием.

Обслуживание с ожиданием означает, что в системе коммутации задержанные источники становятся в очередь и обслуживаются по мере освобождения соединительных устройств (каналов, линий). Очередь может быть общая или индивидуальная к каждому прибору, группе приборов или направлению связи.

Порядок обслуживания – это совокупность правил, устанавливающих последовательность обслуживания телефонных вызовов коммутационной станцией или узлом.

Обслуживание вызовов может осуществляться в случайном порядке, в порядке очередности поступления, с приоритетом.

Например: выбор источников из очереди, может происходить в порядке поступления вызовов (упорядоченная очередь), в обратном порядке (реверсивная очередь), в случайном порядке и с приоритетом.

Приоритетное обслуживание вызовов широко применяется в ЭВМ, системах управления АТС, междугородных и международных телефонных сетях, сетях передачи данных, интегральных сетях.

В военных системах коммутации обслуживание абонентов также осуществляется в приоритетном порядке.

В системах с потерями используют так называемый абсолютный приоритет. Сущность его заключается в том, что при поступлении приоритетного вызова в момент времени, когда в системе заняты все обслуживающие приборы, система коммутации прерывает обслуживание вызова более низкого приоритета и предоставляет освободившееся устройство приоритетному вызову (междугородный вызов, например, может прервать установленное местное соединение).

В системах с ожиданием для обслуживания отдельных категорий используется относительный приоритет. Его сущность заключается в том, что при поступлении приоритетного вызова во время занятости всех обслуживающих приборов система коммутации данный вызов ставит впереди очереди ожидающих вызовов низшего приоритета. Приоритетное обслуживание на сетях специальной связи находит широкое применение для обслуживания отдельных категорий пользователей как при способе с потерями, так и с ожиданием.

Кроме обслуживания с явными потерями на единой сети электросвязи Российской Федерации, а также сетях специальной связи находит широкое применение обслуживание с повторением вызовов.

Данный способ является определенным обобщением способа обслуживания с ожиданием. Необслуженные источники образуют активную очередь и повторяют вызовы через случайные или детерминированные промежутки времени вплоть до установления соединения или отказа от очередной попытки сформировать повторный вызов. Следует иметь в виду, что при данном способе величина вероятности потерь вызовов будет значительно больше значения вероятности потерь сообщений, поскольку потеря отдельных вызовов (попыток) не ведет к явной потере сообщения. Происходит только задержка обслуживания, т. е. при регламентированном времени можно говорить об условных потерях. Данный способ, с математической точки зрения, является более общим по сравнению со способами с явными потерями и ожиданием. С ростом интенсивности повторения промежутки между соседними попытками уменьшаются и при бесконечной интенсивности все попытки сливаются в непрерывное ожидание. С уменьшением интенсивности повторения сокращается различие между источниками первичных и повторных вызовов.

Наряду с основными способами обслуживания – с явными и условными потерями – возможен ряд **комбинированных способов**.

В системе с ожиданием часто ограничивается время пребывания источника в очереди или длина очереди. Тогда часть поступивших сообщений обслуживается с ожиданием, а часть – с повторением или явными потерями.

При обслуживании с повторением возможны ограничения на число повторных попыток или источников, повторяющих вызовы. При этом наряду с условными потерями возникают явные потери сообщений. Необходимо отметить, что характер накладываемых ограничений может быть детерминированным или случайным. В последнем случае задается соответствующий закон распределения случайной величины.

6.3. Характеристики качества обслуживания голосовых сообщений

Нормирование показателей качества обслуживания (QoS) для современных моносервисных сетей осуществляется следующим образом.

Для каждой дисциплины обслуживания вызовов характерны свои основные и вспомогательные характеристики качества обслуживания вызовов.

В системах коммутации с явными потерями сообщений основной характеристикой качества обслуживания является вероятность потери поступившего вызова, а вспомогательными – вероятности потерь по времени и по нагрузке.

Кроме этих характеристик, для сравнения и анализа коммутационных систем и отдельных пучков каналов или линий связи вводится понятие "пропускная способность системы" (пучка). Наряду с пропускной способностью всей системы коммутации часто рассматривают пропускную способность отдельных линий и среднюю пропускную способность одной линии. Эти величины называют использованием и средним использованием линии.

Более подробно порядок определения каждой из перечисленных характеристик будет изложен при рассмотрении конкретных математических моделей. Ниже представлены наиболее простые вероятностные характеристики качества обслуживания вызовов, которые используются для описания систем коммутации с потерями.

Так, вероятность потери вызова $P_B(t_1, t_2)$, поступившего в промежутке $[t_1, t_2)$, есть отношение средней интенсивности потерянных вызовов к общей средней интенсивности всех поступивших вызовов в этом промежутке:

$$P_B(t_1, t_2) = \frac{\mu_{\text{пот}}(t_1, t_2)}{\mu_{\text{пост}}(t_1, t_2)}, \quad (6.1)$$

где $\mu_{\text{пот}}(t_1, t_2)$ и $\mu_{\text{пост}}(t_1, t_2)$ – математические ожидания числа вызовов в промежутке $[t_1, t_2)$.

Для стационарных потоков вероятность P_B – величина постоянная и не зависит от длины и расположения рассматриваемого промежутка времени.

Статистически вероятность потерь вызовов $P_B(t_1, t_2)$ обычно определяют следующим образом:

$$P_B = \frac{C_{\text{пот}}(t)}{C_{\text{пост}}(t)}, \quad (6.2)$$

где $C_{\text{пот}}(t)$ и $C_{\text{пост}}(t)$ – соответственно число потерянных и число поступивших вызовов в промежутке времени t .

Вероятность потерь по времени $P_t(t_1, t_2)$ в промежутке $[t_1, t_2)$ есть вероятность занятости в этом промежутке всех доступных источнику соединительных путей в требуемом направлении. Подобное состояние, когда заняты все доступные линии (каналы), называют состоянием насыщения или "опасным временем" системы.

Статистически вероятность потерь по времени определяют как

$$P_t(t_1, t_2) = \frac{t_3}{T}, \quad (6.3)$$

где t_3 – доля времени, в течение которого заняты все обслуживающие приборы, доступные группе источников заявок из общего периода $T = (t_2 - t_1)$.

Вероятность потерь по времени $P_t(t_1, t_2)$ характеризует потенциальную возможность потери вызова в промежутке $[t_1, t_2)$. Для стационарных потоков вероятность P_t – величина постоянная для любого промежутка времени.

Вероятность потерь по нагрузке $P_H(t_1, t_2)$ в промежутке времени $[t_1, t_2)$ есть отношение интенсивности потерянной $R(t_1, t_2)$ и потенциальной $A(t_1, t_2)$ нагрузок:

$$P_H(t_1, t_2) = \frac{R(t_1, t_2)}{A(t_1, t_2)}. \quad (6.4)$$

Для стационарных потоков вероятность потерь по нагрузке $P_H(t_1, t_2)$ в промежутке времени $[t_1, t_2)$ есть отношение интенсивности потерянной $R(t_1, t_2)$ и поступающей $Z(t_1, t_2)$ нагрузок:

$$P_H(t_1, t_2) = \frac{R(t_1, t_2)}{Z(t_1, t_2)}. \quad (6.5)$$

Так как потери по нагрузке являются абстрактной характеристикой, то на практике она не измеряется. Однако при решении задач, связанных со стационарными потоками, часто используют следующее выражение:

$$P_H = \frac{Z - Y}{Z}. \quad (6.6)$$

Доля потерянной нагрузки может быть выражена в промилле или процентах (1 ‰ промилле = 0,001 доля числа = 0,1 %).

Под пропускной способностью коммутационной системы с потерями понимают интенсивность обслуженной этой системой нагрузки в рассматриваемом промежутке времени при заданном качестве обслуживания:

$$Y(P_{\text{треб}}) = Z \cdot (1 - P_{\text{треб}}), \quad (6.7)$$

где $P_{\text{треб}}$ – нормируемое значение вероятности потерь.

Пропускная способность зависит от величины потерь, емкости пучков линий, включенных в выходы коммутационной системы, способа объединения этих выходов, распределения длительности обслуживания и дисциплины обслуживания. Чем больше допустимая норма потерь (P), тем больше пропускная способность системы и хуже качество обслуживания вызовов.

Наряду с пропускной способностью всей системы часто рассматривают пропускную способность отдельных линий и среднюю

пропускную способность одной линии. Эти величины называют использованием и средним использованием линии.

Для характеристики систем управления используют понятие производительности системы. Она определяет предельное, статистически усредненное число вызовов, которое может обработать система управления или отдельные ее элементы за час без снижения заданных качественных показателей. Данное усреднение необходимо, поскольку для разных видов вызовов требуется разная процедура обработки и, соответственно, разные ресурсы системы управления.

В системах с ожиданием основными характеристиками качества обслуживания (ХКО) являются:

- вероятность ожидания для поступившего вызова $P(t_{\text{ож}} > 0)$;
- функция распределения длительности ожидания вызовом начала обслуживания $P(t_{\text{ож}} > t_{\text{доп}})$.

Вспомогательными ХКО являются:

- средняя длительность ожидания начала обслуживания для ждущих вызовов $\bar{t}_{\text{ожжд}}$;
- средняя длительность ожидания начала обслуживания для всех вызовов $\bar{t}_{\text{ож}}$;
- средняя длина очереди $\bar{r}$.

В системах с повторными вызовами используют несколько иные ХКО:

- вероятность потери первичного, повторного и любого вызовов;
- вероятность явной потери поступившего сообщения;
- среднее число источников ($\bar{S}$) повторных вызовов;
- среднее число повторных попыток на одно установленное соединение (Q_1) или на один первичный вызов;
- вероятность повторения любой попытки (H) и ряд других.

Как правило, нормирование показателей качества обслуживания (QoS) для современных инфокоммуникационных сетей осуществляется следующим образом:

- для систем с потерями устанавливается максимальный порог вероятности потери заявки (вызова или требования иной природы);

– систем с ожиданием определяются две допустимые величины – среднее значение времени нахождения в очереди или задержки (т. е. вместе с обслуживанием) и квантиль соответствующей функции распределения (ФР).

Такие принципы нормирования используются МСЭ и ЕТСИ в документах, прямо или косвенно касающихся показателей качества обслуживания.

6.4. Особенности оценки качества обслуживания вызовов на коммутируемых сетях связи

В настоящее время основной дисциплиной обслуживания потоков вызовов на сетях связи является обслуживание с явными потерями. Практика показывает, что почти 90 % трафика обслуживается по скорой системе автоматическим оборудованием цифровых АТС или АМТС; 10 % – полуавтоматическим оборудованием ручных коммутаторов по заказной или немедленной системе обслуживания.

В этом случае основным показателем качества обслуживания выступает сквозная вероятность потерь P между двумя точками сети. При этом соединение может устанавливаться по нескольким путям (основным и обходным), входящим в данное направление связи.

Как правило, соединительный тракт между двумя АМТС содержит n последовательно включенных ступеней искания (рис. 6.3). В качестве ступени искания может выступать транзитная станция.

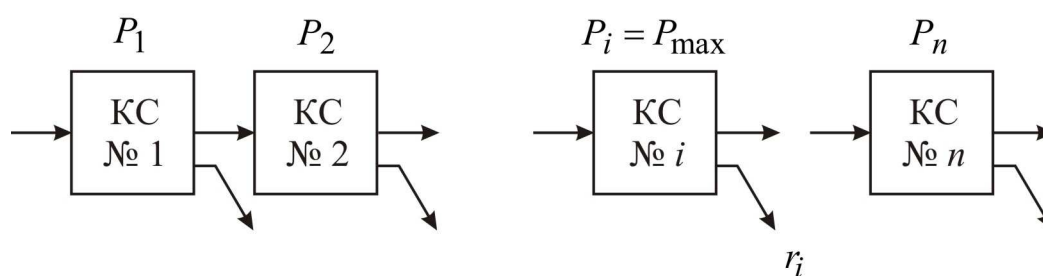


Рис. 6.3. Схема соединительного тракта

Поступающий на i -ю ступень искания трафик делится на r_i потоков в соответствии с числом направлений, организуемых на ступени искания. При этом число соединительных устройств на каждом направлении i -й ступени рассчитывается [19] при величине потерь P_i ($i = 1, 2, 3, \dots, n$).

Обычно предполагают, что на каждую ступень искания поступает либо простейший, либо примитивный поток вызовов. Строго говоря, на вторую и последующие ступени искания поступает сглаженный поток, отличающийся от простейшего или примитивного тем, что часть вызовов во время ЧНН теряется, а сам поток сглаживается. Однако при малых значениях вероятности потерь этим обстоятельством можно пренебречь [19].

Для определения качества обслуживания вызовов в большинстве случаев требуется рассчитать результирующую (сквозную) величину вероятности потерь (P) при установлении соединения между двумя телефонными аппаратами, проходящего через n последовательно включенных ступеней искания.

Исследование сложных коммутационных систем и сетей показало, что величина P зависит от величины потерь на отдельных ступенях искания P_i , числа ступеней искания n , числа направлений r_i , включаемых в каждую ступень искания. Эту зависимость можно выразить функционалом:

$$P = f(P_1, P_2, P_3, \dots, P_n, n, r_1, r_2, \dots, r_n). \quad (6.8)$$

При этом величина вероятности потерь P будет находиться в пределах

$$P_{\max} \leq P \leq \sum_{i=1}^n P_i, \quad (6.9)$$

где $P_{\max} = \max\{P_i, i = 1, 2, \dots, n\}$ – максимальное из значений потерь на ступенях искания сложной коммутационной схемы.

Если при любом числе ступеней искания n на каждой ступени число направлений $r = 1$ (рис. 6.4), то суммарные потери

$$P \approx P_{\max}. \quad (6.10)$$

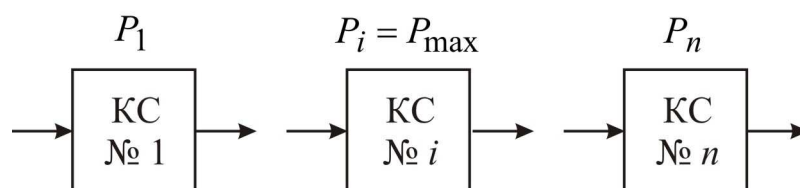


Рис. 6.4. Схема соединительного тракта, $r = 1$

Физически это можно объяснить, представив пропускную способность каждой КС в виде трубы соответствующего диаметра. Следовательно, пропускная способность соединительного тракта сложной коммутационной системы будет определяться тем звеном, в котором меньшая пропускная способность (больше вероятность потери вызова $P_{\max}$).

Если же на n ступенях искания, образующих коммутационную систему, величина $r \rightarrow \infty$ очень велика (большое число направлений связи), а доля рассматриваемого потока вызовов в каждом направлении стремится к нулю (рис. 6.5), то в этом предельном случае процессы обслуживания потоков, поступающих на направления разных ступеней искания, можно считать независимыми.

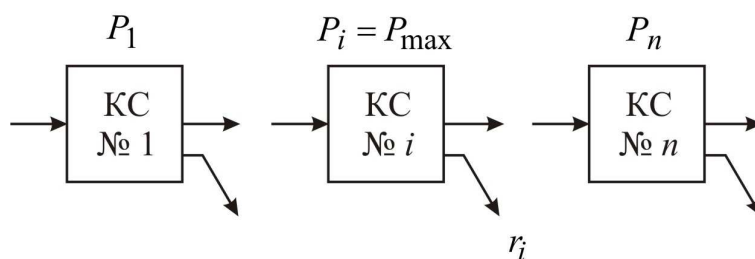


Рис. 6.5. Схема соединительного тракта при числе направлений на каждом звене, $r_i \rightarrow \infty$, $r_i \ll \infty$

В этом случае суммарные потери можно определять по формуле

$$P = 1 - \prod_{i=1}^n (1 - P_i). \quad (6.11)$$

Для случая, когда $n = 2$, вероятность того, что поступающий вызов будет обслужен на первой ступени, равна $(1 - P_1)$, на второй ступени — $(1 - P_2)$. Тогда вероятность того, что поступающий вызов будет обслужен (q) двумя ступенями, при условии независимости обслуживания вызова на каждой ступени, будет равна $q = (1 - P_1) \cdot (1 - P_2)$. Вероятность потери вызова

$$P = 1 - (1 - P_1) \cdot (1 - P_2). \quad (6.12)$$

Выражение (6.11) следует использовать, когда заведомо известно, что $P_i \gg 0,01$.

В реальных системах коммутации, как правило, величина r конечна ($r_i \ll \infty$). В этом случае допущение о независимости соединительных устройств нескольких ступеней искания использовать нельзя.

Кроме того, из-за потерь на ступенях искания меняется характер потока вызовов (в этих случаях формула (6.11) описывает процессы обслуживания вызовов достаточно приближенно).

Поэтому в большинстве случаев, когда $P_i \leq 0,01$ ($i = 1, 2, \dots, n$), без большой погрешности формулу (6.11) можно заменить еще более простым выражением:

$$P \approx \sum_{i=1}^n P_i, \quad (6.13)$$

так как в этом случае члены, содержащиеся в произведении (6.11) пренебрежительно малы.

В соответствии с формулой (6.13) результирующие потери равны сумме потерь на всех ступенях искания, поэтому потери P называют *суммарными*.

Из предыдущих разделов известно, что с уменьшением вероятности потерь (повышением качества обслуживания) увеличивается объем станционных и линейных сооружений, а это приводит к увеличению затрат на построение сети. При этом емкости пучков каналов, их протяженность и стоимость существенно различаются для междугородной, внутризонавой, городских и сельских телефонных сетей. В этом случае затраты на построение отдельных участков соединительного тракта могут значительно различаться.

Для снижения суммарных затрат на построение сетей связи потери на отдельных участках нормируются. При этом чем больше протяженность направления связи и меньше его емкость, тем большее значение имеет величина вероятности потерь.

По этим соображениям на единой сети электросвязи (ЕСЭ) РФ минимальная норма потерь принята для городских телефонных сетей и большая норма – для сельских телефонных сетей, в которых преобладают пучки соединительных линий большой протяженности и малой емкости.

В пределах одной сети суммарные потери распределяются по ступеням искания таким образом, чтобы на более дорогих участках соединительного тракта была большая величина вероятности потерь.

Полностью такой подход реализовать на сетях специальной связи (ССС) невозможно. Это объясняется спецификой задач, решаемых СССР. Однако по аналогии с порядком нормирования потерь на ЕСЭ РФ на сетях специальной связи минимальную норму потерь можно принять на направлениях связи, ведущих к вышестоящим узлам, и несколько большую – на направлениях к взаимодействующим узлам специальной связи.

6.5. Принципы нормирования потерь на телефонных сетях связи

6.5.1. *Изменение подходов к определению блокировки трафика в телефонных сетях*

В Рекомендациях МСЭ-Т E.520 (1988 г.) "Определение необходимого количества каналов при автоматическом и полуавтоматическом обслуживании" указывалось, что "при автоматическом обслуживании вероятность потерь P на пучке не должна превышать 1 % (0,01) в средний час наибольшей нагрузки".

Та же норма потерь рекомендована для совмещенного полуавтоматического и автоматического обслуживания трафика. Эти нормы были разработаны для сетей связи без обходов.

По мере создания сетей связи с обходами такая норма (0,01) Рекомендациями МСЭ-Т E.521 (1988 г.) была перенесена на блокировку одного звена на пути "последнего выбора" (т. е. нормировалось значение сетевого трафика).

В цифровых системах коммутации с помощью записанных в памяти процессора таблиц маршрутизация вызовов производится по нескольким путям, очередность выбора которых может меняться в зависимости от загрузки сети (динамическая маршрутизация). Вследствие этого понятие "путь последнего выбора" потеряло смысл, так как с ним стало невозможно связывать определенный путь в маршруте, поэтому в Рекомендациях МСЭ-Т E.529 (1997 г.) "Определение параметров сети по критерию качества обслуживания (GoS – *grade of service*) из конца в конец" были установлены параметры коммутируемых сетей, позволяющие оператору обеспечить требуемое качество с учетом интересов пользователей. Согласно этому подходу, в телефонных сетях нормируется не сетевое качество обслужи-

вания, как было ранее, а качество обслуживания "из конца в конец" для отдельных направлений.

В Рекомендациях МСЭ-Т E.721 (1999 г.) было определено понятие блокировки "из конца в конец" для различных участков телефонной сети. Согласно этому документу, вероятностью блокировки "из конца в конец" или между оконечными пунктами сети является вероятность того, что попытка вызова была неудачной из-за недостатка технических средств на всех путях маршрута. Стандартные значения вероятностей блокировки из "конца в конец" для местных, междугородных и международных соединений представлены в Рекомендациях МСЭ-Т E.721.

Поскольку эти результаты, как показывает практика, недостаточно известны, они приведены в табл. 6.1 и 6.2.

Таблица 6.1

Маршрутизация вызовов в зависимости от типа сети

Тип сети	Определения
Региональная (местная)	От абонентского узла до абонентского узла
Национальная транзитная (междугородная)	От транзитного узла до транзитного узла
Международная	От международного узла до международного узла

Таблица 6.2

Нормы потерь в зависимости от типа соединения

Тип соединения	Норма потерь (‰)	
	Нормальный уровень	Повышенный уровень
Региональная	20 (0,02)	30 (0,03)
Национальная транзитная	30 (0,03)	45 (0,045)
Международная	50 (0,05)	75 (0,075)

Эти нормы взяты из следующих соображений. На основании исследований, проведенных с точки зрения удобства пользования телефонной связью [24], установлено, что при потерях более 50 ‰ (0,05) абоненты субъективно ощущают дискомфорт в обслуживании.

6.5.2. Нормы потерь на местных сетях связи

В отечественной практике опыт обслуживания абонентов на различных сетях ЕСЭ РФ обобщен в руководящем документе РД 45.120–2000, принятом Министерством связи и информатизации РФ в качестве основного нормативного документа при проектировании сетей городской и сельской связи.

Так, на городской телефонной сети руководящим документом РД 45.120–2000 предусмотрены следующие *нормы потерь*:

- при местной связи (абонент – абонент одной АТС) – 20 ‰;
- местной связи абонента РАТС с УПАТС или пригородной зоной – 25 ‰;
- от абонента охранно-пожарной сигнализации до узла спецслужбы – 1 ‰;
- узла спецслужбы до экстренной службы №№ 01, 02, 03) – 1 ‰;
- узла спецслужбы (УСС) до справочно-информационных и заказных служб (муниципальных и других операторов) – не более 30 ‰;
- абонента до службы АМТС по ЗСЛ – 20 ‰;
- от УСС до служб доступа к федеральным сетям персонального радиовызова – по согласованию с оператором.

Величина потерь на АТС между ступенью ГИ и ЛИ (АИ) не должна превышать 0,005.

Для небольших сетей, в которых преобладают пучки малой емкости, с целью повышения эффективности их использования допустимое значение суммарных потерь существенно увеличивается: $P_i = 10\%$ ($P_i = 0,1$).

Нормы потерь на сельских телефонных сетях:

- суммарные потери при установлении соединения между двумя телефонными аппаратами одной СТС (местная связь) не должны превышать 30 ‰ (при проектировании станции (70 ‰ на существующей сети);
- при внутризонавой связи 40/80 ‰;
- от абонента ОС до ЦС – 10 ‰;
- ЦС (УСС) до экстренных служб – 1 ‰;
- ЦС (УСС) до справочно-информационных и заказных служб (муниципальных и других операторов) не более 30 ‰;

- ЦС до службы АМТС по ЗСЛ – 10 ‰;
- ЦС (УСС) до служб доступа к федеральным сетям персонального радиовызова общего пользования – по согласованию с оператором.

Нормы потерь на зоновой телефонной сети.

Суммарные потери при установлении соединения между двумя телефонными аппаратами разных местных сетей одной зоновой телефонной сети не должны превышать 0,03 – 0,13 (30 ‰ – 130 ‰).

Нормы потерь на междугородной телефонной сети.

Суммарные потери при установлении соединения между двумя телефонными аппаратами ГТС, расположенных в разных зонах семи-значной нумерации, не должны превышать 0,1 (100 ‰). Число каналов на одном участке пути последнего выбора должно рассчитываться при величине потерь 0,01 (10 ‰) (при наличии обходных направлений).

Доля потерь в коммутационной системе проектируемой АМТС и УАК не должна превышать 0,003 (3 ‰).

Нормы, регламентирующие длительность установления соединения.

Среднее время установления соединения на междугородной телефонной сети при оконечной связи между двумя АМТС зависит от числа транзитов и не должно превышать:

- для различных типов коммутационного оборудования в соединительном тракте – 14 – 17 с;
- при многочастотном способе передачи номерной информации на всех участках сети – 4 – 11 с;
- при передаче номерной информации на всех участках сети по общим каналам сигнализации – 3 – 5 с;
- для ожидания вызовов при установлении соединения через четыре транзитные станции – 30 с.

Вопросы для самоконтроля

1. Дайте определение понятию "дисциплина обслуживания".
2. Что понимают под порядком и способом обслуживания?
3. Что значат явные и условные потери?
4. В чем заключаются способы обслуживания с потерями, с повторными вызовами и с ожиданием?

5. Что такое нагрузка и каковы ее виды?
6. Дайте определение обслуженной, потенциальной, поступающей, потерянной и избыточной нагрузкам.
7. Обоснуйте соотношение между указанными видами нагрузки в условиях стационарного потока, потоков с условным параметром и ограниченным последствием.
8. В каких единицах измеряются нагрузка и интенсивность нагрузки?
9. Какие показатели качества обслуживания вызовов используются для количественной оценки обслуживания?
10. Какие нормы потерь предусмотрены на сетях связи?

7. ЗАДАЧИ АНАЛИЗА И СИНТЕЗА ТЕЛЕКОММУНИКАЦИОННЫХ СИСТЕМ

7.1. Символическое обозначение исследуемых моделей

В теории телетрафика (ТТТ) для быстрого описания моделей различных систем распределения информации используется специальная символика, получившая название символики Кендалла – Башарина.

Символика включает в свой состав до шести разрядов, каждый из которых поясняет один из компонентов модели системы распределения информации:

- первый разряд определяет поступающий поток вызовов;
- второй – закон распределения времени обслуживания;
- третий – структуру системы обслуживания;
- четвертый – способ обслуживания вызовов;
- пятый – дисциплину очереди;
- шестой – порядок занятия свободных приборов.

В первом разряде символами соответственно обозначают:

M – экспоненциальное (марковское) распределение промежутков между вызовами (данный символ получил название в честь русского математика А. А. Маркова.

D – регулярное (детерминированное);

E^n – эрланговское n -го порядка;

HM^n – гиперэкспоненциальное n -го порядка (промежутки между вызовами распределены по смешанному показательному закону – с вероятностью q_1 промежутков между вызовами распределен по показательному закону с параметром λ_1 , а с вероятностью q_2 – по показательному закону с параметром λ_2);

HE^n – гиперэрланговское n -го порядка (с вероятностью q_1 промежутков между вызовами распределен по распределению Эрланга первого порядка (показательному закону), а с вероятностью q_2 – по распределению Эрланга второго порядка);

CM^n – комбинированное n -го порядка;

GI – произвольное (*General*) распределение независимых (*Independent*) промежутков между вызовами рекуррентного потока.

Кроме этого, в первом разряде совместно с символом M могут использоваться индексы t, r, i , например M_t , M_r и M_i , которые обозначают соответственно пуассоновский поток с переменным параметром, условным параметром и примитивный поток. Для обозначения неординарного потока после соответствующего символа в первом разряде дополнительно указывается символ неординарности l , а для обозначения потока с неординарными вызовами символ l дополнительно указывается во втором разряде.

Стрелка над символом $\bar{X}$ соответствует многомерному случаю. Например символ $\bar{M}$ в первом разряде, означает обслуживание n простейших потоков вызовов с различными параметрами и разной приоритетностью обслуживания или с другими дополнительными признаками, что оговаривается отдельно.

Во втором разряде для определения соответствующих распределений длительностей обслуживания используются те же символы (только вместо $GI - G$), что и в первом. При этом символ $\bar{M}_n$ во втором разряде определяет показательный закон распределения времени обслуживания с параметром, зависящим либо от номера входящего потока вызовов (если поступает несколько потоков), либо от номера обслуживающего прибора или пути соединения.

Третий разряд обозначает систему обслуживания. Так, систему обслуживания произвольной структуры обозначают символом S .

Для полnodоступных систем можно использовать символ FM (*Full Matrix*) или указать только число приборов V .

Для неполnodоступных систем указываются либо общий символ G (*Grading*), либо отдельные символы для равномерных и ступенчатых схем: HG (*Homogeneous Grading*) и PG (*Progressive Grading*).

Вместо этих символов можно указать число приборов V , доступность D и число нагрузочных групп g .

Многозвенную систему определяют символом LS (*Link System*).

Четвертый разряд соответственно обозначает способ обслуживания:

- без потерь – LL (*Lossles*);
- с явными потерями – L (*Loss*);
- ожиданием – W (*Wait*);
- повторением – R (*Reattempt*).

Комбинированный способ обслуживания обозначается комбинацией соответствующих символов. Например символы WL и WR , оз-

начают, что часть поступающих вызовов обслуживаются с ожиданием, ограниченным по времени пребывания вызова в очереди и/или по длине очереди, а часть из-за этого – соответственно с явными потерями или с повторением (неограниченным). Символы *RL* соответствуют обслуживанию с повторением, ограниченным по числу попыток и/или по числу повторяющих вызовы источников. Из-за наложенных ограничений возможны явные потери сообщений.

Символы *WRL* обозначают обслуживание с ограниченным ожиданием, ограниченным повторением и явными потерями.

При необходимости более детального описания применяемых ограничений вместо символов *WL* указываются *W*, *r* или просто *r* при ограниченной длине очереди, *W*, *v* или *W*, *t* при ограниченном времени пребывания вызова в очереди, *W*, *r*, *v* или *W*, *r*, *t* при обоих видах ограничений. Здесь *r* – число мест для ожидания, *v* – интенсивность ухода вызовов из очереди, *t* – предельное время пребывания вызова в очереди. Запись *r* = 0 или *v* = 8, или *t* = 0 эквивалентна символу *L*, а запись *r* = 8 и *v* = 0, и *t* = 8 – символу *W*. Аналогично вместо символов *WR* применяют одну из следующих уточняющих записей: *W*, *r*, *R*; *W*, *v*, *R*; *W*, *t*, *R*; *W*, *r*, *v*, *R*; *W*, *r*, *t*, *R* или *W*, *r*, *v*, *t*, *R*.

Ограничения в системе с повторением отмечаются дополнительными символами *H* и τ , где *H* – настойчивость абонента, а τ – предельное число источников повторных вызовов. Например запись *W*, *r*, *R*, *H*, τ , означает, что в системе имеется конечное число мест для ожидания *r*, при их занятости вызовы повторяются с настойчивостью $H < 1$ и число источников повторных вызовов не превышает $m < \infty$, поэтому часть поступивших вызовов обслуживается с явными потерями.

В пятом разряде символики указывается дисциплина очереди в системе с ожиданием. Общая очередь к обслуживающим приборам подразумевается по умолчанию, а при индивидуальной очереди к каждому прибору или к каждой группе приборов прибавляется первый символ *I* (*Individual*). Порядок выбора вызовов из очереди можно обозначить следующими символами: *SP* (*Same Probability*) – равновероятный, *FF* (*First-in-First-out*) – упорядоченный (первый пришел – первый ушел), *LF* (*Last-in-First-out*) – стековый или магазинный (последний пришел – первый ушел), *PR* (*Priority*) – приоритетный, *PRR* (*Relative*) – с относительным приоритетом, *PRA* (*Absolute*) – с абсолютным приоритетом. В пятом разряде могут также отражаться особенности очереди в системе с повторением.

Шестой разряд символики используется в тех моделях, где требуется указать порядок занятия свободных приборов или путей. Так, при последовательном занятии применяется символ S (*Sequential*), при случайном – символ R (*Random*).

7.2. Основы теории расчета однозвенных систем распределения информации

7.2.1. Модель $M_r/M/V/L$

7.2.1.1. Постановка задачи по обслуживанию пуассоновского потока вызовов с условным параметром

Введем некоторые обозначения.

Общее число каналов системы коммутации обозначим символом V . Число занятых каналов или линий обозначим через i , при этом $i = 0, \dots, V$. Состояния системы коммутации при i занятых линиях обозначим через S_i .

Под τ будем понимать любой промежуток времени, в течение которого возможны переходы системы в различные состояния, а под $\Delta t \rightarrow 0$ бесконечно малый интервал времени, в течение которого также могут происходить изменения системы.

В соответствии с символикой Кендалла – Башарина, которая описывает данную модель, следует обратить внимание на следующие моменты:

1. На вход системы коммутации поступает пуассоновский поток с условным параметром (M_r). Данный поток обладает простым последствием.

Что дает исследователю данное допущение?

Во-первых, у потока с простым последствием параметр зависит от состояния системы обслуживания – $\lambda_i(t)$.

Во-вторых, поток является нестационарным, но зависимость вероятностных характеристик потока от времени – косвенная, через состояние системы, поэтому параметр потока обозначим символом λ_i (выз/у. е. в.).

В-третьих, поток ординарный (за $\Delta t \rightarrow 0$ $P_{i>1}(\Delta t) = o(\Delta t)$).

В-четвертых, моменты появления новых вызовов не зависят от рассматриваемого случайного процесса до момента t .

2. Время обслуживания одного вызова является случайной величиной, распределенной по показательному закону (M).

Введение данного допущения позволяет получить следующие утверждения:

- моменты окончания занятия линий не зависят от течения процесса до момента t (при показательном распределении время оставшейся части занятия не зависит от того, сколько времени оно продолжалось до момента t);

- поток освобождения линий (каналов) является потоком с условным параметром $\mu_i = i \cdot \mu$;

- поток освобождений также является ординарным, следовательно, за бесконечно малый интервал не может произойти более одного освобождения канала связи.

Предположим, что среднее значение длительности обслуживания равно одной условной единице времени ($\bar{t}_{\text{обсл}} = 1$ у. е. в.). В этом

случае интенсивность обслуживания вызова $\mu = \frac{1}{\bar{t}_{\text{обсл}}} = 1$, следова-

тельно, $\mu_i = i \cdot \mu = i$. Кроме того, параметр потока можно также трактовать как интенсивность поступающей нагрузки в течение условной единицы времени в состоянии системы i . Действительно,

при $\mu = 1$ интенсивность поступающей нагрузки $Z = \frac{\lambda}{\mu} = \frac{\lambda}{1} = \lambda$.

3. Схема системы распределения информации – однозвенная (неблокируемая – без внутренних блокировок), полнодоступная (V).

Это значит, что все выходы доступны каждому входу, отсутствуют внутренние блокировки в коммутационном поле (потери вызова по причине занятости промежуточных линий равны нулю), потери вызовов будут обусловлены единственной причиной – занятостью выходов из схемы станции.

4. Способ обслуживания – с явными потерями сообщений (L).

Это значит, что вызов, поступивший в состоянии системы $i < V$, обслуживается, а при $i = V$ – теряется и более в эту систему не поступает.

Требуется определить характеристики качества обслуживания вызовов однозвенной полнодоступной системой распределения информации с потерями.

В качестве характеристик качества обслуживания вызовов (ХКОВ) в системе с потерями выступают вероятности потерь по вызовам, времени и нагрузке.

При этом в качестве основной характеристики выступает вероятность потерь по вызовам.

Для определения ХКОВ вначале следует определить распределение вероятности состояний системы коммутации. Получив формулу для определения $P_i(t)$ – вероятности нахождения системы в i -м состоянии, можно определить вероятность состояния системы в случае занятости всех имеющихся каналов связи. При занятости всех каналов обслуживание вновь поступающих вызовов прекращается, и наступают потери.

Тем самым вывод уравнения для нахождения $P_i(t)$ позволит получить формулы для расчета ХКОВ систем коммутации (СК) в случае использования различных моделей поступающих потоков вызовов (простейший и примитивный).

7.2.1.2. Определение вероятностей состояния полnodоступного пучка линий

Для определения вероятностей нахождения системы в состояниях i следует обратиться к функциональной схеме однозвенной СК с потерями. Так как причины потерь обусловлены занятостью каналов связи, то в изученной ранее схеме следует оставить только канальные комплекты или каналы связи, но разместить их горизонтально в виде овалов.

Таким образом, переходим от реальной схемы станции к графу состояний, представленному на рис. 7.1.

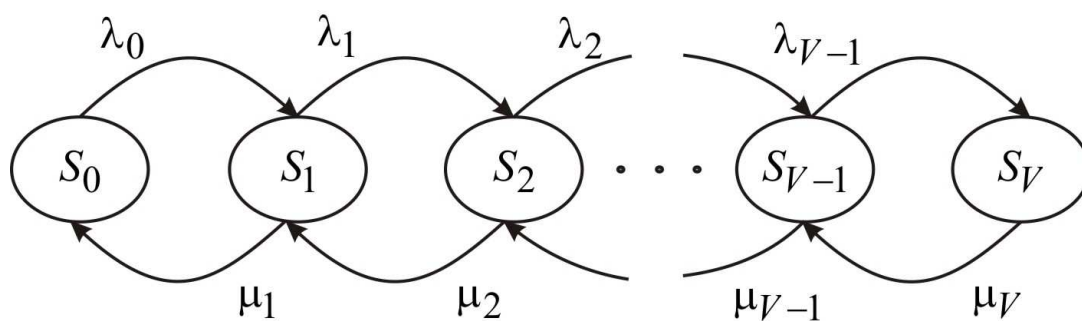


Рис. 7.1. Граф состояний

В данном графе состояния системы будем обозначать символом S_i , где индекс i показывает число занятых выходов в СК.

Сами состояния можно обозначать в виде овалов, квадратов, прямоугольников, точек. Так как в реальных системах коммутации число V конечно, то все возможные состояния системы коммутации можно перечислить:

S_0 – все каналы свободны;

S_1 – один канал занят;

S_2 – два канала заняты;

S_3 – три канала заняты;

S_V – в системе заняты все V каналов связи.

Следовательно, при наличии в системе V обслуживающих приборов имеем конечное множество из $(V + 1)$ состояний.

Возможные переходы из состояния в состояние обозначим стрелками, соединяющими состояния (рис. 7.2).

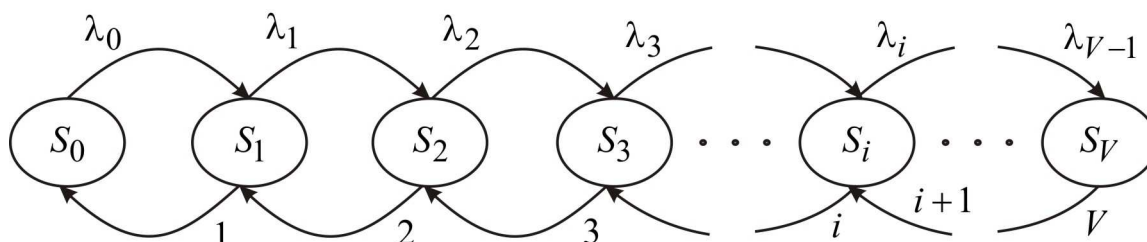


Рис. 7.2. Граф состояний

Так как переход из состояния S_i (в системе занято i выходов) в S_{i+1} (в системе занято $(i + 1)$ выходов) возможен в результате поступления вызова, то верхние стрелки переходов в графе будем обозначать символом λ_i . Индекс i говорит о том, что параметр потока определяется состоянием системы обслуживания.

Переход из состояния S_i в S_{i-1} обусловлен освобождением одного любого выхода. Поток освобождения относится к потокам с условным параметром, и его параметр в момент времени t пропорционален числу занятых каналов связи. Поэтому нижние стрелки имеют обозначение μ_i , где индекс i – число занятых обслуживающих приборов.

Так как при постановке задачи предположили, что среднее время обслуживания $\bar{t}_{\text{обсл}} = 1$ у. е. в. и каждый канал освобождается

с интенсивностью, равной единице ($\mu = \frac{1}{t_{\text{обсл}}} = 1$), то на графе состояний переходы (стрелки), связанные с освобождением каналов связи, следует записать цифрами. При этом интенсивность освобождения одного канала будет равна 1, двух каналов – 2, трех каналов – 3 и т. д. (см. рис. 7.2).

Рассмотрим порядок составления уравнений состояний $P_i(t)$.

Для этого воспользуемся уравнением Колмогорова – Чепмена. Прежде чем записать это уравнение, введем дополнительные обозначения.

Для этого возьмем отрезок времени, на котором отметим точку и обозначим ее буквой t . Вероятность того, что система в момент времени (t) будет находиться в состоянии i , обозначим $P_i(t)$. Интервал времени, в течение которого возможно изменение состояния системы по причине поступления вызова или освобождения, обозначим через τ . Тогда $P_k(t - \tau)$ будет вероятность того, что в момент времени ($t - \tau$) система находилась в любом состоянии k .

Если не ограничивать величину этого промежутка времени τ бесконечно малой величиной $\Delta t \rightarrow 0$, то в состояние i система может перейти из любого другого состояния, например из состояния k .

Вероятность перехода системы из состояния k в состояние i за интервал времени τ обозначим через $P_{k,i}(\tau)$.

В этом случае уравнение Колмогорова – Чепмена будет иметь следующий вид $P_i(t)$:

$$P_i(t) = \sum_k P_k(t - \tau) \cdot P_{k,i}(\tau). \quad (7.1)$$

Физический смысл уравнения (7.1) поясняет рис. 7.3.

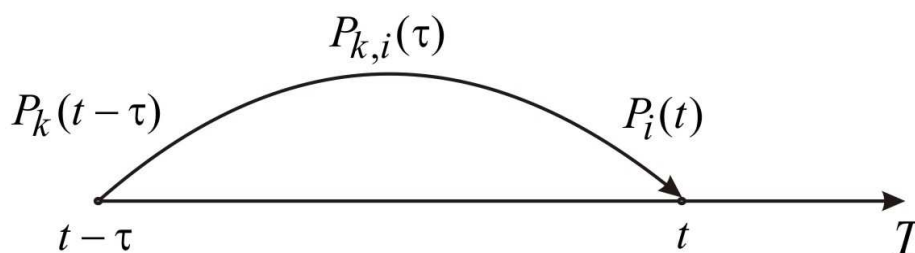


Рис. 7.3. Пояснение уравнения (7.1)

Выражение (7.1) означает, что вероятность нахождения системы в момент времени t в состоянии i равна сумме произведений вероятности нахождения системы в предшествующий момент времени $(t - \tau)$ в любом состоянии k на вероятность перехода системы за промежуток τ из состояния k в состояние i .

Для РАТС, а тем более для крупных АМТС, УИС или УВС число k может принимать достаточно большие значения, но в силу того, что суммирование в формуле (7.1) производится по всем возможным состояниям системы k , процесс вычисления $P_i(t)$ достаточно неоднозначен.

Как же уменьшить число возможных слагаемых? Для этого необходимо использовать одно из свойств потока с условным параметром. Известно, что поток с условным параметром обладает свойством ординарности.

Для преобразования уравнения (7.1) в качестве промежутка τ возьмем бесконечно малую величину Δt ($\Delta t \rightarrow 0$). В этом случае все вероятности перехода из состояния k в состояния $k \pm j$ при $j \geq 2$ есть бесконечно малые величины более высокого порядка малости ($o(\Delta t)$) по сравнению с $\Delta t \rightarrow 0$, т. е. $P_{k, k \pm j}(\Delta t) = o(\Delta t)$. Такое предположение позволит уменьшить число возможных значений k .

В этом случае в состояние i к моменту времени t за бесконечно малый промежуток времени Δt система может перейти либо из состояния $(i - 1)$ для момента времени $(t - \Delta t)$, либо из состояния $(i + 1)$, либо остаться в исходном состоянии i . При этом величина i (следовательно, и k) может принимать любые значения в диапазоне $0 - V$.

Используя уравнения Колмогорова – Чепмена, составим уравнение для нулевого состояния системы – S_0 (состояние, при котором нет занятых каналов, выходов). Обозначим вероятность нахождения системы в момент времени t в состоянии S_0 через $P_0(t)$.

Для определения данной вероятности уточним причины, обусловившие нахождение системы в момент времени t в состоянии S_0 .

Очевидно, что причин этому всего три:

- либо в момент $(t - \Delta t)$ система находилась в состоянии S_0 с вероятностью – $P_0(t - \Delta t)$ и за время Δt не вышла из него с вероятностью $P_{0,0}(\Delta t)$ (за интервал Δt не поступил вызов);

– либо в момент времени $(t - \Delta t)$ система находилась в состоянии S_1 с вероятностью $P_1(t - \Delta t)$ и за время Δt с вероятностью $P_{1,0}(\Delta t)$ перешла из состояния S_1 в S_0 (произошло освобождение занятого канала);

– за время Δt произошло два и более переходов (освобождение двух и более каналов). Так как потоки вызовов и освобождений являются ординарными, то вероятность такого события мала и равна $o(\Delta t)$, где $o(\Delta t)$ – малая величина более высокого порядка, чем $\Delta t \rightarrow 0$, т. е. $\lim_{\Delta t \rightarrow 0} \frac{o(\Delta t)}{\Delta t} = 0$.

Далее, используя формулу Колмогорова для вероятности $P_0(t)$, можно записать:

$$P_0(t) = P_0(t - \Delta t) \cdot P_{0,0}(\Delta t) + P_1(t - \Delta t) \cdot P_{1,0}(\Delta t) + o(\Delta t). \quad (7.2)$$

Найдем вероятности $P_{0,0}(\Delta t)$ и $P_{1,0}(\Delta t)$.

Для нахождения вероятности $P_{0,0}(\Delta t)$ вначале найдем вероятность перехода из состояния S_0 в состояние S_1 – $P_{0,1}(\Delta t)$. Причина перехода системы из состояния S_0 в состояние S_1 обусловлена поступлением вызова. Вероятность поступления в промежутке Δt точно одного вызова $P_{i=1}$ рассчитаем, используя определение параметра ординарного потока $\lambda(\Delta t) = \lim_{\Delta t \rightarrow 0} \left(\frac{P_{i \geq 1}(\Delta t)}{\Delta t} \right)$. Получим выражение для вероятности поступления ровно одного вызова:

$$P_{i=1}(\Delta t) = \lambda(\Delta t) \cdot \Delta t.$$

Так как у потока с условным параметром связь параметра со временем косвенная через состояние системы (при одном и том же состоянии параметр – величина постоянная, зависящая от состояния системы), поэтому в дальнейшем будем считать, что $\lambda(\Delta t) = \lambda_i$.

Следовательно, для нулевого состояния системы вероятность перехода системы из S_0 в S_1 равносильна вероятности поступления вызова и она определяется по формуле $\lambda_0 \cdot \Delta t$. Вероятности поступления или непоступления вызова в состоянии S_0 за бесконечно малый интервал времени Δt образуют полную группу событий, отсюда

вероятность того, что система за Δt не выйдет из состояния S_0 , будет вычисляться по формуле

$$P_{0,0}(\Delta t) = 1 - \lambda_0 \cdot \Delta t + o(\Delta t). \quad (7.3)$$

Пояснение уравнения (7.3) представлено на рис. 7.4.

Подставив формулу (7.3) в выражение (7.2), получим первое слагаемое в виде

$$P_0(t - \Delta t) \cdot (1 - \lambda_0 \cdot \Delta t) + o(\Delta t). \quad (7.4)$$

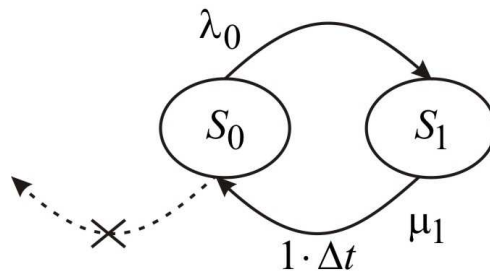


Рис. 7.4. Пояснение уравнения (7.3)

Далее определим вероятность перехода системы из состояния S_1 в состояние S_0 (освобождение занятого канала) за промежуток времени Δt . Вероятность освобождения пропорциональна параметру освобождения $\mu_i = i$. Так как в состоянии S_1 занят всего один канал, а его интенсивность освобождения равна $\mu_1 = 1$, то вероятность перехода системы из состояния S_1 в состояние S_0 за время Δt запишем следующим образом:

$$P_{1,0}(\Delta t) = 1 \cdot \Delta t + o(\Delta t). \quad (7.5)$$

Следовательно, математическое описание второй причины выражения (7.2) можно представить в виде

$$P_1(t - \Delta t) \cdot (1 \cdot \Delta t). \quad (7.6)$$

С учетом того, что рассмотренные события несовместны, используя теорему о сложении вероятностей несовместных событий (вероятность появления одного из двух несовместных событий, безразлично какого, равна сумме вероятностей этих событий $P(A + B) = P(A) + P(B)$) первое уравнение для состояния системы S_0 можно записать как

$$P_0(t) = (1 - \lambda_0 \cdot \Delta t) \cdot P_0(t - \Delta t) + 1 \cdot \Delta t \cdot P_1(t - \Delta t) + o(\Delta t). \quad (7.7)$$

В аналогичных уравнениях для других состояний системы отметим лишь, что для состояний $i = 1, 2, \dots, V - 1$ в правой части уравнения (7.7) будут четыре слагаемых (включая $o(\Delta t)$), а для крайних состояний при $i = V$, а также $i = 0$ – три слагаемых.

Уравнения для состояний $i \geq 1$ можно получить, рассуждая аналогично и учитывая, что

$$\begin{aligned} P_{i-1,i}(\Delta t) &= \lambda_{i-1} \cdot \Delta t + o(\Delta t), \\ P_{i+1,i}(\Delta t) &= (i+1) \cdot \Delta t + o(\Delta t). \end{aligned} \quad (7.8)$$

Вероятность $P_{i,i}(\Delta t)$ неизменности состояния системы за промежуток Δt определяется из очевидного равенства (сумма вероятностей нахождения в данном состоянии i и выхода из него за промежуток Δt равна единице):

$$\begin{aligned} P_{i,i}(\Delta t) + P_{i,i+1}(\Delta t) + P_{i,i-1}(\Delta t) + o(\Delta t) &= 1, \\ P_{i,i}(\Delta t) &= 1 - \lambda_i \cdot \Delta t - i \cdot \Delta t + o(\Delta t). \end{aligned} \quad (7.9)$$

Тогда получим систему уравнений:

$$\left\{ \begin{aligned} P_0(t) &= (1 - \lambda_0 \cdot \Delta t) \cdot P_0(t - \Delta t) + 1 \cdot \Delta t \cdot P_1(t - \Delta t) + o(\Delta t), \\ P_1(t) &= \lambda_0 \cdot \Delta t \cdot P_0(t - \Delta t) + (1 - \lambda_1 \cdot \Delta t - 1 \cdot \Delta t) \cdot P_1(t - \Delta t) + \\ &+ 2 \cdot \Delta t \cdot P_2(t - \Delta t) + o(\Delta t), \\ P_2(t) &= \lambda_1 \cdot \Delta t \cdot P_1(t - \Delta t) + (1 - \lambda_2 \cdot \Delta t - 2 \cdot \Delta t) \cdot P_2(t - \Delta t) + \\ &+ 3 \cdot \Delta t \cdot P_3(t - \Delta t) + o(\Delta t), \\ &\dots \dots \dots \\ P_i(t) &= \lambda_{i-1} \cdot \Delta t \cdot P_{i-1}(t - \Delta t) + (1 - \lambda_i \cdot \Delta t - i \cdot \Delta t) \cdot P_i(t - \Delta t) + \\ &+ (i+1) \cdot \Delta t \cdot P_{i+1}(t - \Delta t) + o(\Delta t), \\ &\dots \dots \dots \\ P_V(t) &= \lambda_{V-1} \cdot \Delta t \cdot P_{V-1}(t - \Delta t) + (1 - \lambda_V \cdot \Delta t - V \cdot \Delta t) \times \\ &\times P_V(t - \Delta t) + o(\Delta t). \end{aligned} \right. \quad (7.10)$$

Преобразуем эту систему уравнений, перенеся $P_i(t - \Delta t)$ в левую часть и разделив обе части на Δt :

$$\begin{aligned}
\frac{P_0(t) - P_0(t - \Delta t)}{\Delta t} &= -\lambda_0 \cdot P_0(t - \Delta t) + P_1(t - \Delta t) + \frac{o(\Delta t)}{\Delta t}; \\
\frac{P_1(t) - P_1(t - \Delta t)}{\Delta t} &= \lambda_0 \cdot P_0(t - \Delta t) - (\lambda_1 + 1) \cdot P_1(t - \Delta t) + \\
&+ 2 \cdot P_2(t - \Delta t) + \frac{o(\Delta t)}{\Delta t}, \\
\frac{P_2(t) - P_2(t - \Delta t)}{\Delta t} &= \lambda_1 \cdot P_1(t - \Delta t) - (\lambda_2 + 2) \cdot P_2(t - \Delta t) + \\
&+ 3 \cdot P_3(t - \Delta t) + \frac{o(\Delta t)}{\Delta t}, \\
&\dots \dots \dots \\
\frac{P_i(t) - P_i(t - \Delta t)}{\Delta t} &= \lambda_{i-1} \cdot P_{i-1}(t - \Delta t) - (\lambda_i + i) \cdot P_i(t - \Delta t) + \\
&+ (i + 1) \cdot P_{i+1}(t - \Delta t) + \frac{o(\Delta t)}{\Delta t}, \\
&\dots \dots \dots \\
\frac{P_V(t) - P_V(t - \Delta t)}{\Delta t} &= \lambda_{V-1} \cdot P_{V-1}(t - \Delta t) - (\lambda_V + V) \times \\
&\times P_V(t - \Delta t) + \frac{o(\Delta t)}{\Delta t}.
\end{aligned}
\tag{7.11}$$

При переходе к пределу при $\Delta t \rightarrow 0$ из уравнения (7.11) получаем систему дифференциально-разностных уравнений. При этом учитываем, что $\lim_{\Delta t \rightarrow 0} \frac{o(\Delta t)}{\Delta t} = 0$.

$$\left\{ \begin{array}{l} 1 \cdot P_1 = \lambda_0 \cdot P_0, \\ 2 \cdot P_2 = -[\lambda_0 \cdot P_0 - (\lambda_1 + 1) \cdot P_1] = \lambda_1 \cdot P_1, \\ 3 \cdot P_3 = -[\lambda_1 \cdot P_1 - (\lambda_2 + 2) \cdot P_2] = \lambda_2 \cdot P_2, \\ \dots\dots\dots \\ i \cdot P_i = -[\lambda_{i-2} \cdot P_{i-2} - (\lambda_{i-1} + (i-1)) \cdot P_{i-1}] = \lambda_{i-1} \cdot P_{i-1}, \\ \dots\dots\dots \\ V \cdot P_V = \lambda_{V-1} \cdot P_{V-1}. \end{array} \right. \quad (7.13)$$

Предельное распределение вероятностей P_i характеризует работу системы обслуживания в состоянии статистического равновесия. В этих условиях система подвержена изменениям, однако вероятности, описывающие ее поведение, не меняются со временем.

Данная система уравнений была математически выведена А. Н. Колмогоровым значительно позже того времени, когда она была получена А. К. Эрлангом. Эрланг же получил эту систему, введя важное для практики правило статистического равновесия.

Согласно этому правилу, сумма интенсивностей, выводящих систему из состояния i , взвешенная вероятностью P_i , равна сумме интенсивностей, вводящих систему в данное состояние, взвешенных вероятностями соответствующих состояний.

Это правило позволит в дальнейшем получить требуемые уравнения состояний, предварительно рассмотрев все возможные переходы системы:

$$(\lambda_j + i)P_j = \lambda_{j-1} \cdot P_{j-1} + (i+1) \cdot P_{j+1}. \quad (7.14)$$

Из уравнения (7.13), задавая значениями i , последовательно равными $0, 1, 2, \dots, V$, получаем

$$\sum_{i=0}^V P_i = 1. \quad (7.17)$$

Тогда

$$\sum_{i=0}^V P_i = \sum_{i=0}^V \frac{\prod_{k=0}^{i-1} \lambda_k}{i!} \cdot P_0 = 1. \quad (7.18)$$

Отсюда

$$P_0 = \left[\sum_{i=0}^V \frac{\prod_{k=0}^{i-1} \lambda_k}{i!} \right]^{-1}. \quad (7.19)$$

Подставляя выражение (7.19) в формулу (7.16), окончательно получаем

$$P_i = \frac{\prod_{k=0}^{i-1} \lambda_k}{i!} \cdot P_0 = \frac{\prod_{k=0}^{i-1} \lambda_k}{i!} \cdot \frac{1}{\sum_{j=0}^V \frac{\prod_{n=0}^{j-1} \lambda_n}{j!}}. \quad (7.20)$$

Выражение (7.20) есть искомое распределение вероятностей установившегося режима. Оно позволяет определить вероятности занятости точно i любых выходов из V полнодоступной неблокируемой системы в произвольный момент времени t при обслуживании с явными потерями пуассоновского потока вызовов с условным параметром λ_i .

Рассмотрим следующие логические трактовки распределения вероятностей, полученных в формуле (7.20).

1. По ансамблю коммутационных систем. Пусть имеется n ($n \rightarrow \infty$) одинаковых независимо функционирующих полнодоступных КС с V выходами и каждая система обслуживает с явными потерями пуассоновский поток вызовов с условным параметром. Тогда вероятность P_i есть доля систем, в которых в произвольный момент времени t занято точно i выходов, т. е. $P_i = \lim_{n \rightarrow \infty} \frac{n_i}{n}$, где n_i – число систем с i занятыми выходами в момент времени t .

2. По ансамблю моментов времени. Если фиксировать состояния определенного полнодоступного пучка в m ($m \rightarrow \infty$) произвольных моментов времени, то P_i есть доля моментов, в которых пучок находится в состоянии i , т. е.

$$P_i = \lim_{m \rightarrow \infty} \left(\frac{m_i}{m} \right),$$

где m_i – число моментов времени, когда зафиксировано состояние системы i .

3. По времени. Вероятность P_i есть доля времени (на бесконечном интервале времени), в течение которого в исследуемом полнодоступном пучке занято точно i линий.

7.2.1.3. Характеристики качества обслуживания вызовов системами $M/M/V/L$ и $M_i/M/V/L$

Рассматривается система $M/M/V/L$. Согласно данной символике, число источников $N > 300 - 500$ ($N > 15V$), модель потока вызова – простейший, параметр потока вызовов $\lambda = \text{const}$. Длительность обслуживания является случайной величиной, распределенной по экспоненциальному закону $P(t_{\text{зан}} < t) = 1 - e^{-\mu t}$, где μ – интенсивность освобождения одного канала, равная $\mu = \frac{1}{t_{\text{обсл}}}$. Параметр потока освобождений каналов прямо пропорционален числу занятых каналов связи $\mu_i = i \cdot \mu$, где i – число занятых каналов связи в системе обслуживания ($i = \overline{0, V}$). Схема станции полнодоступная, однозвен-

ная, что обуславливает потери только по причине занятости обслуживающих приборов (каналов связи). Внутренние блокировки в коммутационном поле системы отсутствуют. Способ обслуживания – с явными потерями – означает, что при поступлении вызова в момент, когда все приборы (каналы) заняты, он получает отказ и из дальнейшего обслуживания выбывает ("теряется"). Данную модель можно также использовать и для описания систем с повторными вызовами, но при этом вероятность потерь не должна превышать 0,01 ($P_B < 0,01$).

Требуется определить вероятности потерь по вызовам, времени и нагрузке (P_B , P_t , P_H).

Согласно условию задачи, поток вызовов описывается моделью простейшего потока, следовательно, $\lambda_k = \lambda$. В этом случае выражение (7.20) при условии $\bar{t}_{зан} = 1$ примет вид

$$P_i = \frac{\frac{\lambda^i}{i!}}{\sum_{j=0}^V \frac{\lambda^j}{j!}}. \quad (7.21)$$

В случае $\bar{t}_{зан} \neq 1$ у. е. в. и $\mu_i = i \cdot \mu$ выражение (7.21) следует записать в виде

$$P_i = \frac{\frac{\lambda^i}{\mu^i \cdot i!}}{\sum_{j=0}^V \frac{\lambda^j}{\mu^j \cdot j!}} = \frac{\frac{(\lambda \cdot \bar{t}_{зан})^i}{i!}}{\sum_{j=0}^V \frac{(\lambda \cdot \bar{t}_{зан})^j}{j!}}. \quad (7.22)$$

Выражение (7.22) – это так называемое *первое распределение Эрланга*, или усеченное распределение Пуассона. Такое название данная формула получила по следующей причине. Если в формуле

Пуассона $P_i(t) = \frac{(\lambda \cdot t)^i}{i!} \cdot e^{-\lambda t}$ функцию $e^{-\lambda t}$ разложить в ряд Макло-

рена, то получим $e^{-\lambda t} = \frac{1}{\sum_{j=0}^{\infty} \frac{(\lambda \cdot t)^j}{j!}}.$

Поэтому, данная формула будет представлена в виде

$$P_i = \frac{\frac{(\lambda \cdot t)^i}{i!}}{\sum_{j=0}^{\infty} \frac{(\lambda \cdot t)^j}{j!}}. \quad (7.23)$$

Как видим, в знаменателе формулы (7.22) сумма ограничена числом V , а в формуле Пуассона сумма в знаменателе не ограничена (∞).

Из выражения (7.22) вероятность P_i можно выразить через поступающую нагрузку (Z). Действительно, поступающую нагрузку на СК можно определить, используя выражение

$$Z = \frac{\lambda}{\mu} = \lambda \cdot \bar{t}_{\text{зан}},$$

где λ – интенсивность потока вызовов, $\bar{t}_{\text{зан}}$ – средняя длительность обслуживания (в данном примере она равна условной единице времени (у. е. в.), а в общем случае – положительное число).

Здесь распределение Эрланга (7.22) можно записать в виде

$$P_i = \frac{\frac{Z^i}{i!}}{\sum_{j=0}^V \frac{Z^j}{j!}}. \quad (7.24)$$

Вероятность P_i есть вероятность занятости точно i любых выходов из V имеющихся в системе коммутации при поступлении на нее простейшего потока вызовов с интенсивностью нагрузки, равной Z .

На рис. 7.5 в качестве примера показаны вероятности состояний трехлинейной системы P_i ($i = 0, \dots, 3$) в зависимости от интенсивности поступающей нагрузки.

На рис. 7.6 представлена зависимость функции P_i от числа занятых каналов i при фиксированном значении $Z = 8$ Эрл и числе каналов $V = 15$. Анализ поведения графика зависимости показывает, что так же, как и в случае распределения Пуассона, в распределении Эрланга $P_i = f(i)$, при $Z = \text{const}$ (величина Z – целое число) имеют место два экстремума функции в точках при $i = Z$ и $i = Z - 1$.

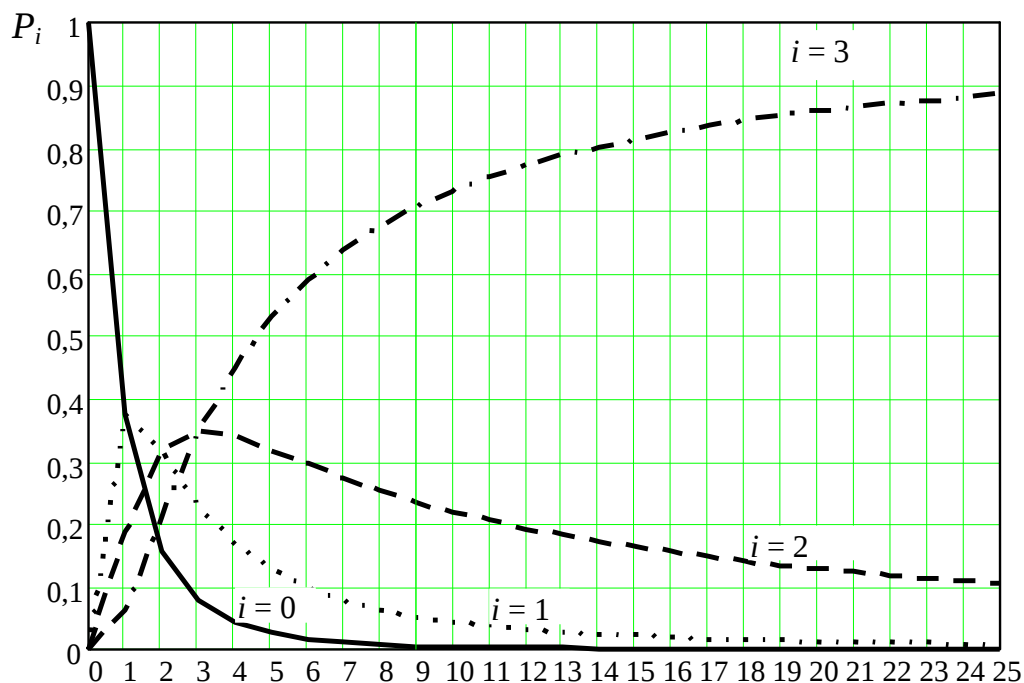


Рис. 7.5. Зависимости вероятностей состояний P_i от интенсивности поступающей нагрузки для трехлинейной системы

При дробном значении Z экстремум функции будет только в одной точке при $i = \lfloor Z \rfloor$, где $\lfloor x \rfloor$ означает целую часть числа x . Например при $Z = 9,5$, функция $P_i = f(i)$ будет иметь максимум в точке $i = 9$.

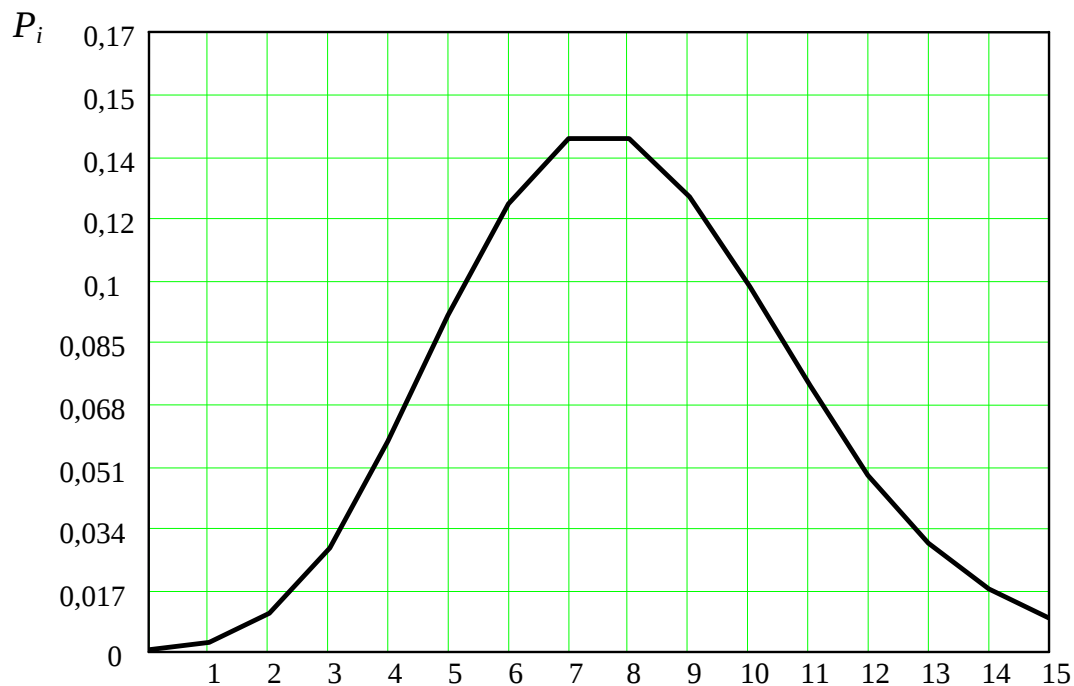


Рис. 7.6. Распределение вероятностей P_i (первое распределение Эрланга) при фиксированном значении $Z = 8$ Эрл и $V = 15$

Огибающие распределения Эрланга и Пуассона $P_i = f(i)$ по своему характеру близки к друг другу (рис. 7.7). Отличие заключается лишь в том, что для эрланговского распределения i конечно и не превышает V , а для пуассоновского – значение i сверху не ограничено. Поэтому вероятность P_i , рассчитанная по формуле (7.21), всегда больше аналогичной вероятности, рассчитанной по формуле (7.20).

При $Z > V$ с ростом i вероятность P_i будет увеличиваться, достигая наибольшего значения при $i = V$.

В системе с явными потерями сообщений характеристиками качества обслуживания являются вероятность потерь по времени, вероятность потерь по вызовам и вероятность потерь по нагрузке. При этом основной характеристикой качества обслуживания является вероятность потери поступившего вызова [24], а вспомогательными – вероятность потерь по времени и нагрузке.

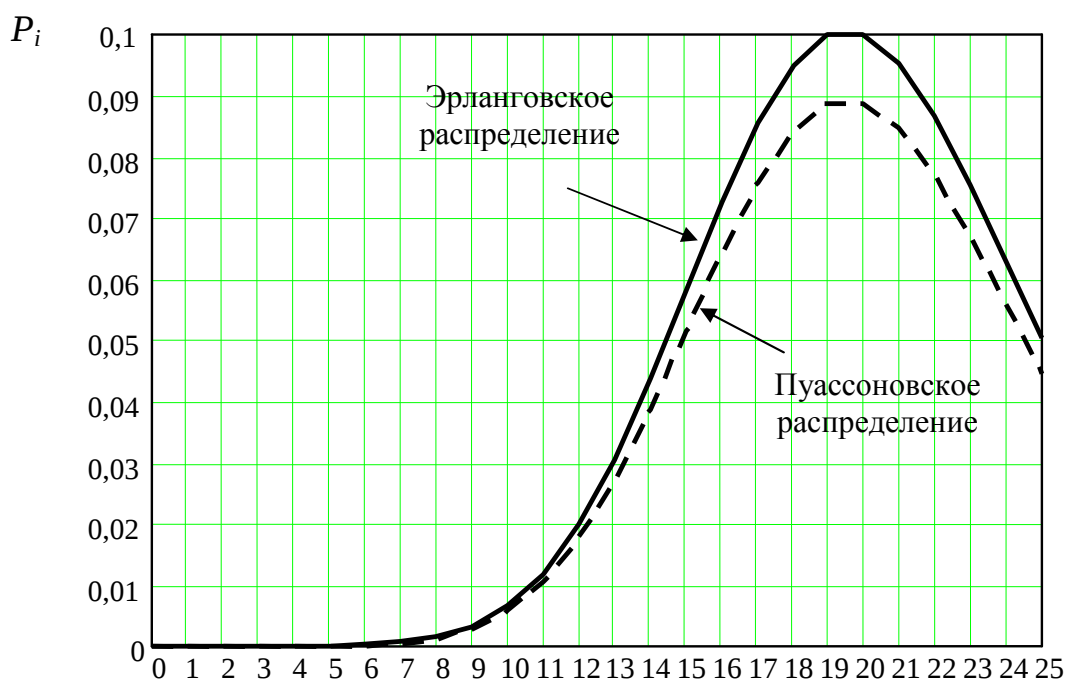


Рис. 7.7. Распределение Пуассона и Эрланга при фиксированном значении $Z = 20$ Эрл и $V = 25$

Вероятность потерь по времени в рассматриваемом интервале времени есть вероятность занятости в этом промежутке всех доступных источнику соединительных путей в требуемом направлении. Следовательно, в соответствии с четвертым толкованием вероятности P_t вероятность потерь по времени можно определить как

$$P_t = P_V = \frac{Z^V / V!}{\sum_{j=0}^V Z^j / j!} = E_V(Z), \quad (7.25)$$

где $E_V(Z)$ – символическая запись эрланговых потерь. Формулу (7.25) называют первой формулой Эрланга.

Вероятность потерь по вызовам (P_B) в общем случае определяется как отношение интенсивности избыточной (I) и поступающей (Z) нагрузок:

$$P_B = \frac{I}{Z}. \quad (7.26)$$

Данная вероятность определяет долю потерянных заявок, которую можно рассчитать по формуле

$$P_B = \frac{\lambda_{\text{пот}}}{\lambda}, \quad (7.27)$$

где $\lambda_{\text{пот}}$ есть математическое ожидание потерянных заявок.

Потери вызовов при данном способе обслуживания и схеме станции возможны только во время занятости всех V приборов. Так как параметр простейшего потока равен его интенсивности и является случайной величиной (СВ), то и параметр потока отказов также будет дискретной СВ. Для определения математического ожидания параметра $\lambda_{\text{пот}}$ воспользуемся формулой

$$\lambda_{\text{пот}} = \lambda_V \cdot P_V. \quad (7.28)$$

Интенсивность λ (в выражении (7.27)) есть математическое ожидание числа вызовов за единицу времени и для ее определения используем выражение

$$\lambda = \sum_{k=0}^V \lambda_k \cdot P_k. \quad (7.29)$$

Учитывая то, что $\lambda_k = \lambda$, а сумма всех возможных состояний системы (условие нормировки) $\sum_{k=0}^V P_k = 1$, получим

$$P_B = \frac{\lambda_{\text{пот}}}{\lambda} = \frac{\lambda_V \cdot P_V}{\sum_{k=0}^{\infty} \lambda_k \cdot P_k} = \frac{\lambda \cdot P_V}{\lambda \cdot \sum_{k=0}^{\infty} P_k} = P_V = P_t. \quad (7.30)$$

Таким образом, при обслуживании простейшего потока вызовов при способе с потерями вероятность потерь по времени равна вероятности потерь по вызовам.

Вероятность потерь по нагрузке есть отношение потерянной нагрузки R к потенциальной нагрузке A :

$$P_H = \frac{R}{A}. \quad (7.31)$$

Интенсивность потенциальной нагрузки при условии, что поток вызовов простейший ($\lambda_k = \lambda$), каждый вызов имеет среднюю длительность занятия, равную одной условной единице времени, а также учитывая рекуррентное соотношение $[(i+1) \cdot P_{i+1} = \lambda_i \cdot P_i]$ для вычисления вероятности P_i , можно определить из следующего выражения:

$$A = \sum_{i=1}^{\infty} i \cdot P_i = \lambda \cdot \sum_{i=1}^{\infty} P_{i-1} = \lambda \cdot \sum_{i=0}^{\infty} P_i = \lambda = Z. \quad (7.32)$$

В выражении (7.32) вероятность P_i определяется в соответствии с распределением Пуассона, а при $t_{\text{зан}} = 1$ у. е. в. величина поступающей нагрузки будет равна $Z = \lambda \cdot 1$ (у. е. в.) $= \lambda$.

Таким образом, исходя из выражения (7.32), можно сделать вывод о том, что поступающая и потенциальные нагрузки при поступлении на СК простейшего потока вызовов равны между собой ($Z = A$), а следовательно, равны (по определению) и интенсивность избыточной и потерянной нагрузок ($I = R$).

Примечание. Разность между потенциальной и обслуженной нагрузками носит название потерянной нагрузки R .

Разность между поступающей и обслуженной нагрузками называется избыточной нагрузкой I .

Отсюда

$$P_H = \frac{R}{A} = \frac{I}{Z} = P_B. \quad (7.33)$$

Таким образом, для СК, обслуживающей простейший поток вызовов, вероятности потерь по вызовам, времени и нагрузке равны между собой ($P_B = P_t = P_H$).

Кроме вероятностей потерь важными характеристиками являются интенсивность обслуженной нагрузки, пропускная способность системы и степень использования одного канала (линии) связи (пропускная способность одного канала связи).

При определении величины обслуженной нагрузки воспользуемся теоремой о количественной ее оценке: интенсивность обслуженной нагрузки (Y), выраженная в эрлангах, количественно равна среднему числу одновременно занятых выходов, обслуживающих эту нагрузку. Так, число занятых линий – величина случайная, для ее определения необходимо использовать выражение для математического ожидания:

$$Y = \sum_{i=1}^V i \cdot P_i = \lambda \cdot \sum_{i=1}^V P_{i-1} = \lambda \cdot \sum_{i=0}^{V-1} P_i. \quad (7.34)$$

Так как $\sum_{i=0}^{V-1} P_i + P_V = 1$, то выражение (7.31) можно записать в виде

$$Y = \lambda \cdot \sum_{i=0}^{V-1} P_i = \lambda \cdot (1 - P_V). \quad (7.35)$$

Интенсивность потока вызовов при $\bar{t}_{\text{зан}} = 1$ у. е. в. представляет собой интенсивность поступающей нагрузки Z , значит, выражение (7.35) можно записать в виде

$$Y = Z \cdot (1 - P_V) = Z \cdot (1 - P_B). \quad (7.36)$$

Пропускная способность пучка каналов связи

$$Y(P_{\text{треб}}) = Z \cdot (1 - P_{B_{\text{треб}}}), \quad (7.37)$$

где $P_{B_{\text{треб}}}$ – требуемое значение вероятности потерь вызовов.

Интенсивность нагрузки, обслуженной одним выходом, характеризует степень использования отдельных выходов системы.

Занятие выходов в системе бывает последовательным и случайным (равновероятным). При последовательном занятии всегда первым занимается свободный выход с меньшим номером, при случайном – с одинаковой вероятностью занимается любой из свободных выходов. При случайном занятии среднее использование всех выходов одинаково:

$$\eta = \frac{Y}{V}. \quad (7.38)$$

Часто вместо символа η используют символ y .

С ростом числа каналов V вначале среднее использование каналов быстро растет, затем наступает насыщение, и величина η практически не увеличивается. Порог насыщения зависит от нормы потерь (рис. 7.8).

На практике часто возникает необходимость в определении вероятности занятия i определенных, фиксированных выходов (H_i). При этом не существенно, заняты только отмеченные i ($i = 0, V$) выходы или одновременно с ними заняты еще некоторые j ($j = 0, V - i$).

Если в системе с вероятностью P_{i+j} занято точно $i + j$ выходов, то вероятность занятия одной конкретной комбинации выходов равна

отношению $\frac{P_{i+j}}{C_V^{i+j}}$.

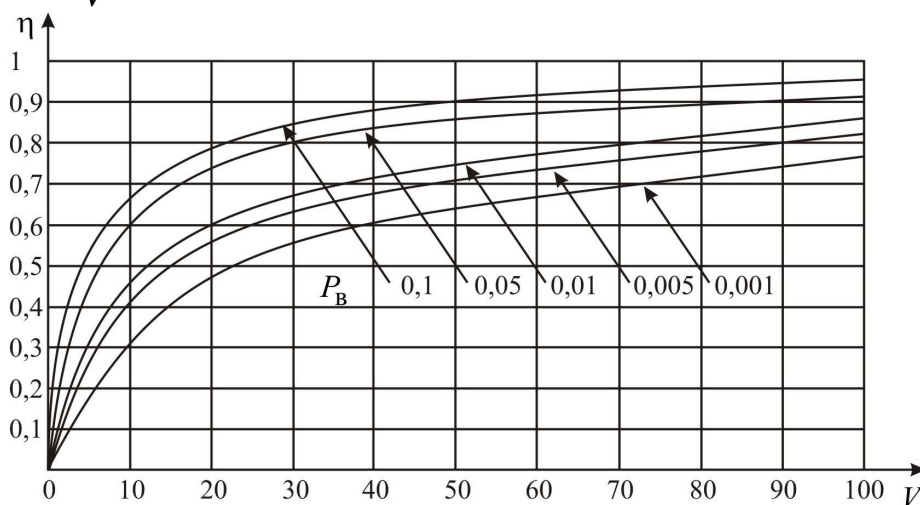


Рис. 7.8. Зависимость среднего использования выходов η от их числа V при различных вероятностях потерь P_B

Отмеченные i выходов могут быть заняты совместно с любыми j выходами в C_{V-i}^j комбинациях. С учетом того, что j – любое число в пределах $j = 0, \dots, V-i$, получаем вероятность H_i занятия фиксированных выходов:

$$H_i = \sum_{j=0}^{V-i} C_{V-i}^j \cdot \frac{P_{i+j}}{C_V^{i+j}} = \frac{E_V(Z)}{E_{V-i}(Z)}. \quad (7.39)$$

Таким образом, для расчета вероятностных характеристик систем распределения информации с потерями при абонентской ёмкости $N > 100$ или $N > 15V$ применяют первую формулу Эрланга. При этом основной характеристикой качества обслуживания, используемой в дальнейших расчетах, является вероятность потерь по вызовам. Удобство данной модели заключается в том, что другие характеристики обслуживания: вероятность потерь по времени и нагрузке – дополнительно рассчитывать не требуется, так как $P_B = P_t = P_H$.

Порядок расчета систем распределения информации с потерями при использовании таблиц и номограмм

Ввиду важности выражения (7.25) для расчета систем распределения информации, реализующих способ обслуживания с потерями, формулы Эрланга табулированы для различных значений V , Z , p . Составлены таблицы, получившие название "таблицы Пальма", а также построены номограммы.

Использование номограмм и таблиц позволяет, не прибегая к расчетам по вышеприведенным формулам, найти один из неизвестных параметров V , Z или P по двум известным. Исходными данными для расчета чаще всего являются результаты статистической обработки реальной нагрузки, функционирующей на действующих сетях связи, либо результаты моделирования при проектировании сетей связи на ЭВМ.

Таблицы приведены в [22, 23]. Они предназначены для расчета систем коммутации с неограниченным числом источников нагрузки и полнодоступным включением обслуживающих приборов и служат для определения величины потерь.

В верхней части табл. 7.1 по горизонтали представлены значения каналов связи (V), а в левой части по вертикали – значения ин-

тенсивности поступающей нагрузки (Z), выраженные в эрлангах. На пересечении вертикалей и горизонталей указаны значения вероятности потерь. Точность определения вероятности потерь в таблицах доведена до четырех десятичных знаков. Так как вероятность потерь меньше единицы, то нуль и запятая для сокращения опущены. Таблицы приведены для числа каналов (V) от 1 до 110 в диапазоне нагрузок от 0,05 до 110 Эрл.

В таблице приведены только те значения потерь, которые превосходят $0,5 \cdot 10^{-4}$ – в противном случае соответствующие места таблицы оставлены пустыми.

Таблица 7.1

Фрагмент таблицы Пальма

$Z, \text{Эрл} \backslash V$	1	2	3	4	5	6	7	8	9	10	11
4,8	8276	6651	5156	3822	2684	1768	1081	0609	0315	0149	0065
4,9	8305	6705	5227	3904	2767	1843	1143	0654	0344	0166	0073
5,0	8333	6757	5297	3983	2849	1918	1205	0700	0375	0184	0083
5,1	8361	6807	5364	4062	2929	1994	1268	0748	0407	0203	0093
5,2	8387	6856	5430	4138	3009	2068	1332	0797	0440	0224	0105
5,3	8413	6903	5495	4213	3087	2143	1396	0846	0475	0245	0117
5,4	8438	6949	5557	4287	3164	2217	1460	0897	0511	0268	0130
5,5	8462	6994	5618	4358	3241	2290	1525	0949	0548	0293	0144
5,6	8485	7038	5678	4429	3316	2363	1590	1002	0587	0318	0159
5,7	8507	7080	5736	4498	3389	2436	1655	1055	0626	0345	0175
5,8	8529	7121	5793	4565	3462	2507	1720	1109	0667	0372	0193
5,9	8551	7161	5848	4631	3534	2579	1785	1164	0709	0401	0211
6,0	8571	7200	5902	4696	3604	2649	1851	1219	0751	0431	0230

Указанная интенсивность нагрузки и выделенное число приборов на пересечении соответствующей строки и столбца определяют величину вероятности потерь по вызовам.

Кроме таблиц для определения вероятности потерь используют номограммы [23] (рис. 7.9). Однако в этом случае точность расчетов несколько ниже, чем при использовании таблиц Пальма.

Модель $M/M/V/L$. Условия применения данной модели:

– реальный поток вызовов аппроксимируется моделью примитивного потока. Это означает, что число источников нагрузки ограничено ($N < 100$ либо при $N = 300 - 500$ выполняется условие $N < 15V$). При этом параметр примитивного потока вызовов прямо

пропорционален числу свободных источников нагрузки
 $\lambda_i = \alpha \cdot (N - i);$

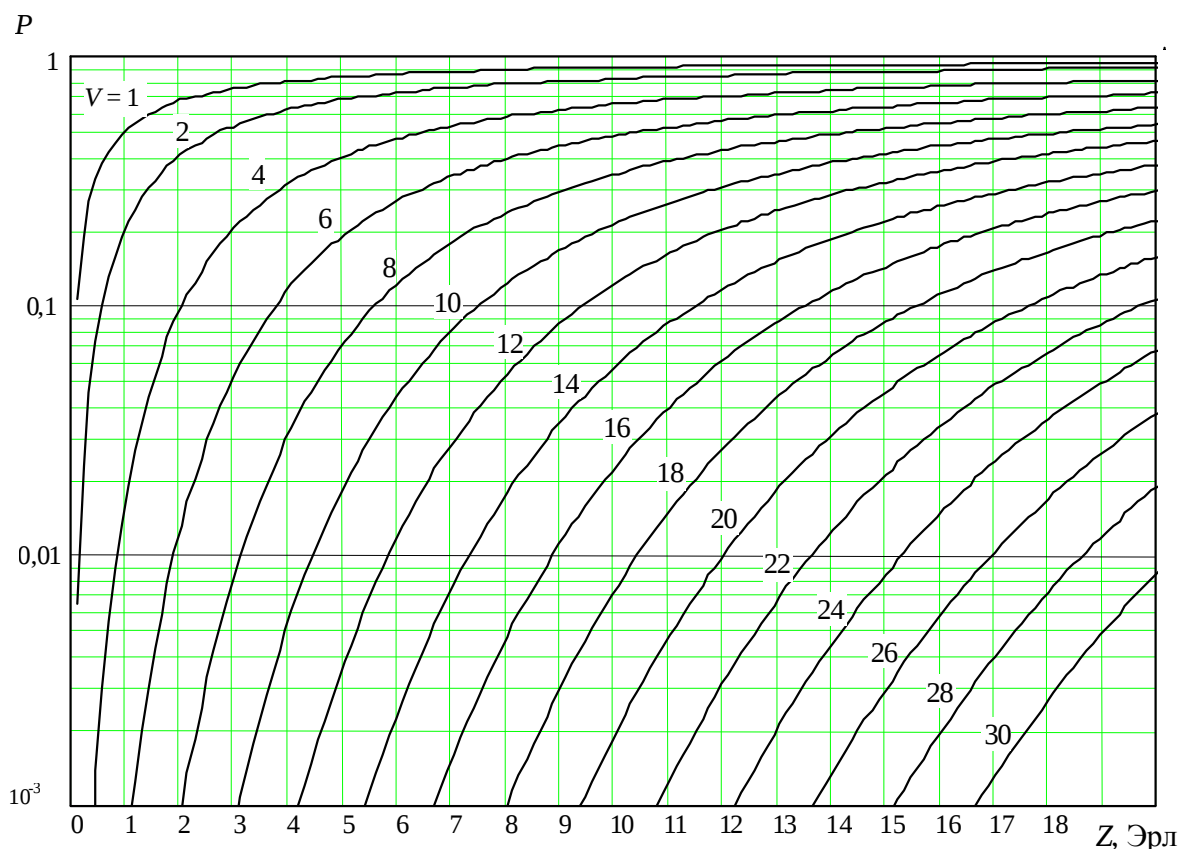


Рис. 7.9. Номограмма для расчета параметров систем распределения информации с явными потерями при числе источников $N > 100$ и полnodоступным включением обслуживающих приборов

– длительность обслуживания является случайной величиной, распределенной по экспоненциальному (произвольному) закону. Следовательно, параметр потока освобождения каналов при выбранном способе обслуживания прямо пропорционален числу занятых обслуживанием каналов связи;

– схема полnodоступная, однозвенная. Единственной причиной потерь является занятость всех каналов связи системы коммутации;

– способ обслуживания с явными потерями.

Требуется определить вероятностные характеристики качества обслуживания вызовов:

- вероятность потерь по вызовам (P_B);
- вероятность потерь по времени (P_t);
- вероятность потерь по нагрузке (P_H).

Воспользуемся формулой (7.22) при условии $\bar{t}_{\text{зан}} = 1$ у.е. в:

$$P_i = \frac{\frac{\lambda_0 \cdot \lambda_1 \cdot \lambda_2 \cdot \dots \cdot \lambda_{i-1}}{i!}}{\sum_{j=0}^V \frac{\lambda_0 \cdot \lambda_1 \cdot \lambda_2 \cdot \dots \cdot \lambda_{j-1}}{j!}}. \quad (7.40)$$

В данном уравнении параметр потока определяется по формуле

$$\lambda_i = \alpha(N - i). \quad (7.41)$$

Подставим формулу (7.41) в выражение (7.40) и преобразуем его числитель:

$$\frac{\prod_{k=0}^{i-1} \lambda_k}{i!} = \frac{\alpha \cdot (N - 0) \cdot \alpha \cdot (N - 1) \cdot \alpha \cdot (N - 2) \cdot \dots \cdot \alpha \cdot (N + 1 - i)}{i!}. \quad (7.42)$$

Умножив числитель и знаменатель формулы (7.42) на $(N - i)!$, получим

$$\begin{aligned} & \frac{\prod_{k=0}^{i-1} \lambda_k}{i!} = \\ & = \frac{\alpha \cdot (N - 0) \cdot \alpha \cdot (N - 1) \cdot \alpha \cdot (N - 2) \cdot \dots \cdot \alpha \cdot (N + 1 - i) \cdot (N - i)!}{i! \cdot (N - i)!} = \\ & = C_N^i \cdot \alpha^i. \end{aligned} \quad (7.43)$$

Проделав аналогичную процедуру со знаменателем формулы (7.40), получим выражение, которое в теории телетрафика получило название формулы Энгсета:

$$P_i = \frac{C_N^i \cdot \alpha^i}{\sum_{j=0}^V C_N^j \cdot \alpha^j}. \quad (7.44)$$

Выражение (7.44) позволяет определить зависимости вероятности нахождения системы коммутации с ограниченным числом источ-

ников нагрузки в любом i -ом состоянии от интенсивности источника в свободном состоянии ($\alpha = \frac{1}{t_{св}}$).

Однако в практических расчетах используют зависимость P_i от потенциальной либо поступающей удельной нагрузки на абонентскую линию (a, z , Эрл).

Для получения зависимости вероятности P_i от потенциальной удельной нагрузки на абонентскую линию (a) рассмотрим систему без потерь, в которой число источников (N) равно числу обслуживающих приборов (V), причем $N = V = 1$. Используя выражение (7.44), найдем вероятности незанятости (P_0) и занятости (P_1) обслуживающего прибора системы коммутации:

$$P_0 = \frac{C_N^i \cdot \alpha^i}{\sum_{j=0}^1 C_N^j \cdot \alpha^j} = \frac{C_1^0 \cdot \alpha^0}{C_1^0 \cdot \alpha^0 + C_1^1 \cdot \alpha^1} = \frac{1}{1 + \alpha};$$

$$P_1 = \frac{C_N^i \cdot \alpha^i}{\sum_{j=0}^1 C_N^j \cdot \alpha^j} = \frac{C_1^1 \cdot \alpha^1}{C_1^0 \cdot \alpha^0 + C_1^1 \cdot \alpha^1} = \frac{\alpha}{1 + \alpha}. \quad (7.45)$$

Вероятность P_1 определяет вероятность занятости одного обслуживающего прибора (линии, канала связи). При условии $N = V$ эту вероятность можно рассматривать как вероятность занятости одного источника, так как в данных условиях за источником как бы закрепляется определенный выход. В этом случае вероятность P_1 будет определять долю времени, в течение которого источник занят в системе без потерь. Следовательно, вероятность P_1 численно равна интенсивности потенциальной нагрузки от одного источника. Обозначив интенсивность потенциальной нагрузки от одного источника через a , получим выражение

$$a = \frac{\alpha}{1 + \alpha}. \quad (7.46)$$

Из формулы (7.46) можно получить выражение для определения интенсивности источника в свободном состоянии:

$$(1 + \alpha) \cdot a = \alpha;$$

$$a = (\alpha - \alpha \cdot a);$$

$$\alpha = \frac{a}{(1 - a)}. \quad (7.47)$$

Подставив формулу (7.47) в выражение (7.44), получим

$$P_i = \frac{C_N^i \cdot a^i \cdot (1 - a)^{-i}}{\sum_{j=0}^V C_N^j \cdot a^j \cdot (1 - a)^{-j}}. \quad (7.48)$$

Умножив и разделив выражение (7.48) на $(1 - a)^N$, получим

$$P_i = \frac{C_N^i \cdot a^i \cdot (1 - a)^{-i}}{\sum_{j=0}^V C_N^j \cdot a^j \cdot (1 - a)^{-j}} \cdot \frac{(1 - a)^N}{(1 - a)^N} = \frac{C_N^i \cdot a^i \cdot (1 - a)^{N-i}}{\sum_{j=0}^V C_N^j \cdot a^j \cdot (1 - a)^{N-j}}. \quad (7.49)$$

Если в качестве исходных данных используют удельную поступающую нагрузку z , то выражение (7.49) можно записать в виде

$$P_i = \frac{C_N^i \cdot z^i \cdot (1 - z)^{N-i}}{\sum_{j=0}^V C_N^j \cdot z^j \cdot (1 - z)^{N-j}}. \quad (7.50)$$

Основные характеристики качества обслуживания.

В соответствии с третьим толкованием вероятности P_i вероятность потерь по времени P_t равна вероятности занятости всех V линий полнодоступного пучка:

$$P_t = P_V = \frac{C_N^V \cdot \alpha^V}{\sum_{j=0}^V C_N^j \alpha^j} = \xi(N, V, \alpha), \quad (7.51)$$

где $\xi(N, V, \alpha)$ – условная запись потерь в системе коммутации малой ёмкости.

Выражение (7.51) можно представить и в другом виде:

$$P_t = \frac{C_N^V \cdot a^V \cdot (1-a)^{N-V}}{\sum_{j=0}^V C_N^j \cdot a^j \cdot (1-a)^{N-j}}, \quad (7.52)$$

где a – интенсивность потенциальной нагрузки одного источника. Выражение (7.52) также называют формулой Энгсета.

При известной удельной поступающей нагрузке формулу (7.51) можно представить в виде

$$P_t = \frac{C_N^V \cdot z^V \cdot (1-z)^{N-V}}{\sum_{j=0}^V C_N^j \cdot z^j \cdot (1-z)^{N-j}}. \quad (7.53)$$

Вероятность потерь по вызовам определим как отношение избыточной нагрузки (I) и поступающей (Z):

$$P_B = \frac{I}{Z}, \quad (7.54)$$

где $I = \lambda_{\text{пот}} \cdot \bar{t}_{\text{обсл}} = \lambda_{\text{пот}}$, а $Z = \lambda \cdot \bar{t}_{\text{обсл}}$ ($\bar{t}_{\text{обсл}} = 1$ у.е. в.).

Отсюда

$$P_B = \frac{I}{Z} = \frac{\lambda_{\text{пот}}}{\lambda} = \frac{C_{N-1}^V \cdot \alpha^V}{\sum_{j=0}^V C_{N-1}^j \cdot \alpha^j}. \quad (7.55)$$

Если задана величина удельной потенциальной нагрузки источника, то выражение (7.55) примет вид

$$P_B = \frac{C_{N-1}^V \cdot a^V \cdot (1-a)^{N-V-1}}{\sum_{j=0}^V C_{N-1}^j \cdot a^j \cdot (1-a)^{N-j-1}}. \quad (7.56)$$

Если в условиях задачи задана удельная поступающая нагрузка, то

$$P_B = \frac{C_{N-1}^V \cdot z^V \cdot (1-z)^{N-V-1}}{\sum_{j=0}^V C_{N-1}^j \cdot z^j \cdot (1-z)^{N-j-1}}. \quad (7.57)$$

Сравнение формул (7.51) и (7.55) показывает, что всегда имеет место неравенство $P_t > P_B$.

Равенство наступает только в предельном случае при $N \rightarrow \infty$, когда примитивный поток переходит в простейший. Выражение (7.57) можно логически получить из формулы (7.55). Вызов, поступивший от конкретного свободного источника, будет потерян, если в этот момент заняты все V выходов. Эта занятость должна быть обеспечена остальными $(N-1)$ источниками.

Вероятность потерь по нагрузке, как известно, есть отношение интенсивности потерянной (R) к потенциальной нагрузке (A):

$$P_H = \frac{R}{A}.$$

$$P_H = \frac{N-V}{N} \cdot \frac{C_N^V \cdot \alpha^V}{\sum_{j=0}^V C_N^j \cdot \alpha^j} = \left(1 - \frac{V}{N}\right) \cdot P_t. \quad (7.58)$$

Интенсивность обслуженной нагрузки

$$Y = \frac{\alpha \cdot N \cdot (1 - P_B)}{1 + \alpha(1 - P_B)}, \quad (7.59)$$

где α – интенсивность источника в свободном состоянии.

Интенсивность потенциальной нагрузки (A) можно определить, зная величину интенсивности потенциальной нагрузки от одного источника a :

$$A = N \cdot a. \quad (7.60)$$

Если задана интенсивность источника в свободном состоянии α , то выражение (7.60) примет вид

$$A = N \cdot a = \frac{N \cdot \alpha}{1 + \alpha}. \quad (7.61)$$

С учетом ранее полученного соотношения $P_B \leq P_t$ имеет место общее неравенство $P_H \leq P_B \leq P_t$.

Интенсивность поступающей нагрузки (Z) определяется не только потребностями абонента в передаче сообщения, но и возможностями системы в передаче сообщений (количеством каналов V):

$$Z = \frac{\alpha \cdot N}{1 + \alpha \cdot (1 - P_B)}, \quad (7.62)$$

где α – интенсивность источника в свободном состоянии.

Следовательно, поступающая нагрузка от одного источника z будет составлять

$$z = \frac{\alpha}{1 + \alpha \cdot (1 - P_B)}. \quad (7.63)$$

Таким образом, при обслуживании примитивного потока взаимодействие источника с системой обслуживания характеризуется четырьмя параметрами нагрузки, физический смысл которых необходимо четко различать, поскольку численные значения их неодинаковы. Неправильное применение их при расчетах может привести к заметным погрешностям, особенно в области средних и больших потерь. Рассмотрим эти параметры:

– интенсивность потока вызовов от свободного источника в одну условную единицу времени α ($\overline{0, \infty}$), выз/у. е. в. (Эрл);

– средняя интенсивность потока вызовов от источника в одну условную единицу времени или интенсивность поступающей от источника нагрузки $z(\overline{0, \infty})$, выз/у. е. в. (Эрл). Из формулы (7.63) получим:

$$z = \lambda / N = \alpha / [1 + \alpha(1 - P_B)]; \quad (7.64)$$

– интенсивность потенциальной нагрузки от источника $z(\overline{0, 1})$, Эрл:

$$a = \alpha / (1 + \alpha); \quad (7.65)$$

– интенсивность обслуженной нагрузки, отнесенная к одному источнику $y(\overline{0, V / N})$ (Эрл):

$$y = Y / N = \alpha(1 - P_B) / [1 + \alpha(1 - P_B)]; \quad (7.66)$$

Из сравнения формул (7.63) – (7.65) следует, что

$$y \leq a \leq z \leq \alpha. \quad (7.67)$$

Выражения (7.51) и (7.55) используются для определения вероятности занятости любых i из V выходов коммутационной системы малой емкости. В случае необходимости определения вероятности (H_i) занятости i конкретных (фиксированных) выходов, если при этом заняты только эти отмеченные i ($i = 0, \dots, V$) выходы или одновременно с ними заняты еще некоторые j ($j = 0, \dots, V - i$) выходы, (при условии $N > V$), следует использовать следующее выражение:

$$H_i = \frac{\frac{C_N^V \cdot \alpha^V}{\sum_{j=0}^V C_N^j \cdot \alpha^j}}{\frac{C_{N-i}^{V-i} \cdot \alpha^{V-i}}{\sum_{k=0}^V C_{N-i}^{k-i} \cdot \alpha^{k-i}}} = \frac{\xi(N, V, \alpha)}{\xi(N-i, V-i, \alpha)}. \quad (7.68)$$

При распределении Бернулли ($N = V$) вероятность занятости конкретных i выходов определяется выражением

$$H_i = a^i. \quad (7.69)$$

Для расчета показателей качества обслуживания вызовов систем с ограниченным числом источников нагрузки используются специальные таблицы, с помощью которых можно оценить вероятности потерь по вызовам.

Особенностью данных таблиц является то, что над таблицей всегда указывается численность источников нагрузки $N < 100$, а вместо общей поступающей нагрузки задается удельная нагрузка на одну абонентскую линию.

С помощью таблиц при заданном числе источников нагрузки (N) и известных значениях двух параметров, к примеру числа каналов (V) и удельной поступающей нагрузки (z , Эрл), можно определить вероятность потерь P_B (табл. 7.2).

Порядок определения вероятности потерь по вызовам P_B в зависимости от интенсивности удельной поступающей нагрузки z , числа источников N и ёмкости группы выходов V такой же, как и при использовании таблиц Пальма, за исключением того, что сначала определяется соответствующая таблица с требуемой численностью источников нагрузки. Например при $N = 10$, $z = 0,19$ Эрл, $V = 5$, вероятность потерь по вызовам (табл. 7.2) $P_B = 0,0135$.

Используя соотношение $P_t(N, V, z) = P_B(N + 1, V, z)$, по этим таблицам можно также определить вероятность потерь по времени.

Таблица 7.2

Фрагмент таблицы Энгсета при числе источников $N = 10$

$\begin{matrix} V \\ z, \text{Эрл} \end{matrix}$	1	2	3	4	5	6	7	8	9
0,01	0833	0034	0001	0000	0000	0000	0000	0000	0000
0,02	1552	0125	0006	0000	0000	0000	0000	0000	0000
0,05	3214	0634	0077	0006	0000	0000	0000	0000	0000
0,07	4038	1084	0187	0021	0002	0000	0000	0000	0000
0,10	5000	1818	0450	0074	0008	0001	0000	0000	0000
0,11	5266	2066	0562	0103	0013	0001	0000	0000	0000
0,12	5510	2311	0685	0138	0019	0002	0000	0000	0000
0,13	5735	2553	0817	0180	0027	0003	0000	0000	0000
0,15	6136	3022	1107	0285	0050	0006	0000	0000	0000
0,17	6483	3469	1422	0419	0085	0012	0001	0000	0000
0,18	6639	3683	1587	0497	0108	0016	0001	0000	0000
0,19	6786	3890	1755	0582	0135	0021	0002	0000	0000
0,20	6923	4091	1927	0674	0166	0028	0003	0000	0000

Кроме таблиц, находят применение номограммы. С их помощью также можно определить любую из трех величин при двух известных: поступающую нагрузку ($Z = N \cdot z$), число каналов (обслуживающих приборов (V)) и вероятность потерь по вызовам P_B при заданном числе источников вызовов ($N = 10, 20, 30, \dots, 90$).

Номограмма для расчета СК при $N < 100$ представлена на рис. 7.10.

Порядок пользования номограммами аналогичен вышеописанному для случая простейшего потока вызовов.

Пересечение величины нагрузки Z ($Z = N \cdot z$) и кривой, соответствующей числу обслуживающих приборов V , на ординате соответствует потерям P_B .

Следует иметь в виду, что при использовании номограмм результат будет менее точен, чем при использовании таблиц.

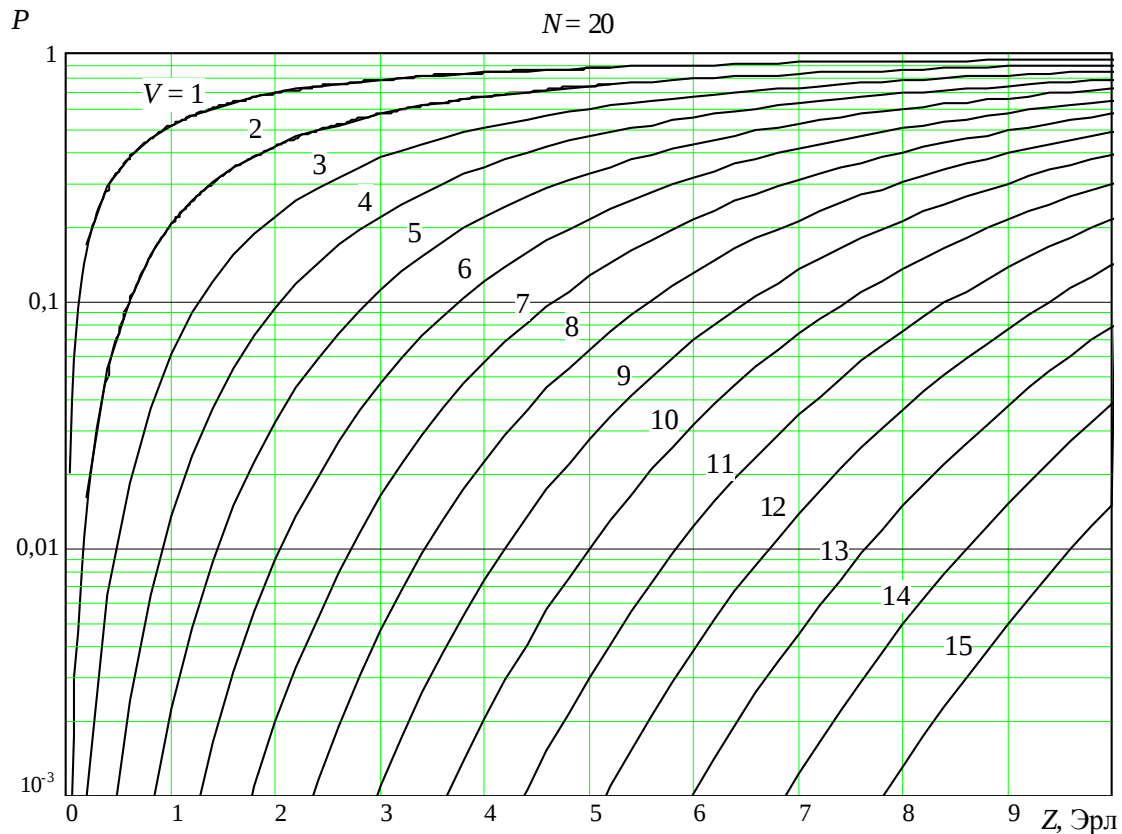


Рис. 7.10. Номограмма расчета параметров полноступенчатых систем распределения информации с явными потерями при числе источников $N = 20$

7.2.2. Модель $M/M/V/L/PRA$

7.2.2.1. Постановка задачи

Рассматриваемая модель системы распределения информации с приоритетной дисциплиной обслуживания в символике Кендалла – Башарина имеет обозначение – $M/M/V/L/PRA$.

Разберём основные особенности данной модели.

1. Первый символ $\bar{M}$ поясняет, что на систему распределения информации поступают несколько потоков вызовов, при этом каждый из них является простейшим.

Каждый поток требований формирует отдельная группа ($i = 1, \dots, Q$) источников нагрузки численностью ($N_Q > 15 \cdot V$).

Интенсивность поступающей нагрузки на систему коммутации (СК) от каждой группы источников нагрузки равна $Z_1, Z_2, Z_3, \dots, Z_Q$.

Используя свойство аддитивности модели простейшего потока, можно утверждать, что суммарный поток также будет аппроксимироваться моделью простейшего потока вызовов с общей интенсив-

ностью нагрузки $Z_\Sigma = \sum_{i=1}^Q Z_i$.

2. Второй символ M показывает, что время обслуживания является случайной величиной, распределенной по экспоненциальному закону.

3. Третий символ определяет структуру системы коммутации.

При этом символ V обозначает, что система однозвенная, полностью доступная, имеющая V каналов связи. Следовательно, одной из причин потерь вызовов каждой группы источников нагрузки является занятость каналов связи. Вторая причина потерь определяется выбранной дисциплиной обслуживания.

4. Четвертый символ определяет способ обслуживания.

Символ L означает, что система обслуживает с явными потерями.

Символ PRA , стоящий на пятом месте, символизирует порядок обслуживания с абсолютным приоритетом.

Обслуживание с явными потерями и абсолютным приоритетом означает, что при поступлении очередной заявки во время занятости всех обслуживающих приборов заявками равной или вышестоящих категорий поступающий вызов получает отказ в обслуживании и повторно в данную систему не поступает. Если же во время поступления приоритетного вызова в системе на обслуживании находится заявка нижестоящей категории, то система прерывает ее обслуживание, а освободившийся ресурс предоставляет приоритетному вызову. Заявки, обслуживание которых было прервано вызовами абонентов вышестоящих категорий, повторно не дообслуживаются и считаются потерянными из-за приоритетного прерывания.

Источники первой группы N_1 обладают абсолютным приоритетом по отношению к источникам $N_2 - N_Q$; источники второй катего-

рии являются приоритетными по отношению к $N_3 - N_Q$ и т. д. Неприоритетными являются абоненты Q -й категории численностью N_Q .

Таким образом, при данной дисциплине обслуживания одной из причин потерь вызовов для источников категорий $N_2 - N_Q$ будет прерывание обслуживания более высокой категории источников нагрузки в случае занятости каналов связи.

Кроме занятости каналов, второй причиной потерь вызовов для нижестоящих категорий абонентов является прерывание обслуживания вызова.

Требуется определить характеристики качества обслуживания вызовов абонентов разных приоритетных категорий: $(P_1, P_2, \dots, P_Q)$.

При рассмотрении данной модели с целью удобства получения конечных формул будем оперировать вероятностью потерь по нагрузке $-P_H$, что не снижает важности и не ограничивает область использования данной модели, так как для моделей с простейшими потоками вызовов ХКОВ равны между собой: $P_H = P_B = P_t$.

7.2.2.2. Характеристики качества обслуживания вызовов системой с абсолютным приоритетом

Формально приоритетную систему с потерями можно представить в виде пирамиды (рис. 7.11), на вершине которой размещаются абоненты более высокого приоритета.

В данной модели потери заявок из-за занятости обслуживающих приборов определяются, исходя из следующих очевидных положений.

Так, абоненты первой категории обладают абсолютным приоритетом. Следовательно, на обслуживание их нагрузки абоненты низших приоритетов никакого влияния не оказывают. Поясним это положение. Если вызов поступит от абонента первой категории в момент времени, когда в системе нет ни одного свободного канала, то система коммутации прерывает обслуживание вызова любой нижестоящей категории (заявка низшего приоритета, которая обслуживается на данное время) и предоставляет освободившийся канал данному абоненту. Поэтому вероятность потери вызовов (вероятность по-

терь по нагрузке) абонентов высшего приоритета (P_1) будет обусловлена лишь имеющимся канальным ресурсом и их занятостью обслуживанием абонентов только первой категории. Очевидно, что вероятность потерь вызовов в этом случае можно записать в виде функции $P_1 = P_V(1)$.

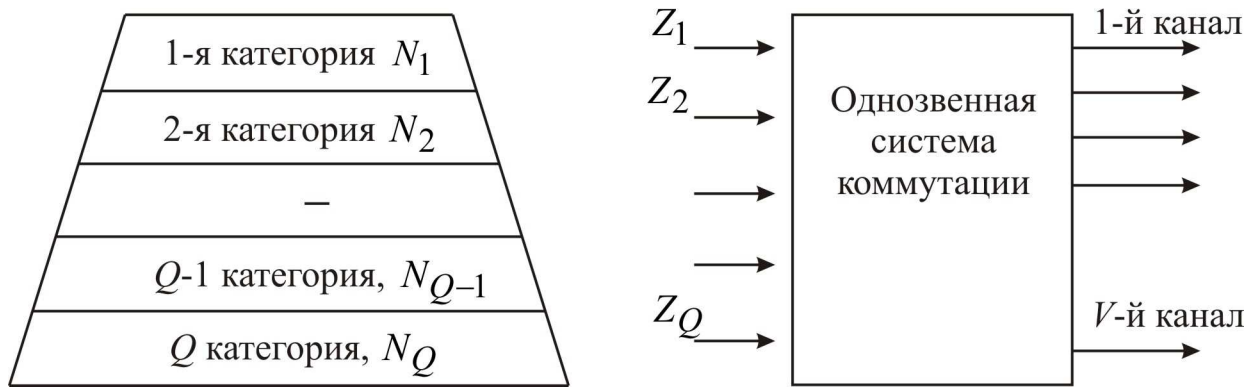


Рис. 7.11. Формальная модель приоритетной СК

Так как число источников первой категории $N_1 > 100$, что предполагает аппроксимацию потока вызовов моделью "простейшего", а система однозвенная полнодоступная с потерями, следовательно, для определения вероятности потерь $P_V(1)$ используем первую формулу Эрланга:

$$P_V(1) = \frac{\frac{Z_1^V}{V!}}{\sum_{i=0}^V \frac{Z_1^i}{i!}} = E_V(Z_1). \quad (7.70)$$

Рассмотрим порядок определения вероятности потерь для источников $i = 2, \dots, Q$ категорий, начиная с процесса обслуживания абонентов самой низшей Q -й категории.

Для простоты уяснения материала рассмотрим с позиции теории вероятности наступление события потери вызова для абонентов низшей категории ($i = Q$). В дальнейшем, получив необходимые формулы, принятый подход распространим на абонентов любой категории с учетом того, что все нижестоящие категории на обслуживание абонентов более высокого приоритета влияния не оказывают.

Как было отмечено выше, заявки низшей категории ($i = Q$) приоритетом по отношению к абонентам высших категорий не обладают.

Вызовы абонентов Q -й категории могут теряться как в случае поступления вызова в момент времени, когда все приборы (каналы) заняты обслуживанием заявок всех групп пользователей, в том числе Q категории, так и при взятии на обслуживание заявки Q -й категории и поступления вызова абонента любой вышестоящей категории ($1, Q-1$) в момент занятости всех каналов. При этом взятая ранее на обслуживание заявка низшего приоритета в результате прерывания обслуживания приоритетным вызовом теряется и не дообслуживается.

Первое событие, обусловленное потерей вызовов по причине занятости каналов обслуживанием нагрузки как Q , так и всех вышестоящих $1, \dots, Q-1$ категорий, обозначим как событие A .

Второе событие, обусловленное прерыванием взятой на обслуживание нагрузки абонентов Q -й категории по причине поступления вызова от абонентов приоритетных ($1, \dots, Q-1$) категорий во время занятости всех каналов связи, обозначим как событие B .

Эти два события являются несовместными, так как нельзя прервать нагрузку, ранее уже потерянную по причине занятости каналов связи.

В этом случае вероятность потери вызова для заявки Q категории $P(Q)$ можно определить по формуле сложения вероятностей несовместных событий

$$P(Q) = P(A + B) = P(A) + P(B), \quad (7.71)$$

где $P(A)$ – вероятность потерь вызовов абонентов Q категории по причине занятости каналов во время их поступления, а $P(B)$ – вероятность потерь по причине прерывания уже взятой на обслуживание заявки Q категории.

Вначале определим вероятность первого события $P(A)$. Обозначим данную вероятность через $P_V(Q)$. В этом обозначении индекс "V" подчеркивает причины потерь в результате занятости всех V каналов связи. Следовательно, физическая интерпретация данной вероятности – потери трафика источников Q -й категории по причине занятости всех V каналов обслуживанием всех категорий.

Используя характеристики простейших потоков вызовов, заключающиеся в том, что интенсивность суммируется при объединении нескольких простейших потоков, данную вероятность можно определить как функцию $P_V(Q) = f(Z_\Sigma; V)$, где Z_Σ – сумма нагрузок всех категорий источников нагрузки.

Интенсивность суммарного простейшего потока вызовов определим по формуле $Z_\Sigma = Z_1 + Z_2 + \dots + Z_Q = \sum_{i=1}^Q Z_i$.

Далее воспользуемся первой формулой Эрланга для определения вероятности потерь заявок источников нагрузки Q -й категории по причине занятости всех приборов $P_V(Q)$ и получим:

$$P_V(Q) = \frac{\frac{(Z_1 + Z_2 + \dots + Z_Q)^V}{V!}}{\sum_{k=0}^V \frac{(Z_1 + Z_2 + \dots + Z_Q)^k}{k!}} = \frac{\frac{\left(\sum_{i=1}^Q Z_i\right)^V}{V!}}{\sum_{k=0}^V \frac{\left(\sum_{i=1}^Q Z_i\right)^k}{k!}} = E_V\left(\sum_{i=1}^Q Z_i\right). \quad (7.72)$$

Очевидно, что для вызовов любой $j = 2, \dots, Q-1$ категорий вероятность потери вызова по причине занятости каналов (событие A) запишется в виде

$$P_V(j) = \frac{\frac{\left(\sum_{k=1}^j Z_k\right)^V}{V!}}{\sum_{i=0}^V \frac{\left(\sum_{k=1}^j Z_k\right)^i}{i!}} = E_V\left(\sum_{k=1}^j Z_k\right). \quad (7.73)$$

Особенность вычисления вероятности потерь по причине занятости всех каналов связи для абонентов высших категорий определяется без учета нагрузки низших категории, так как более приоритет-

ные источники в случае такого конфликта могут прервать уже взятые на обслуживание заявки нижестоящих категорий.

Рассмотрим событие B , поясняющее вторую причину потерь заявок Q категории, связанную с приоритетным прерыванием уже взятых на обслуживание вызовов Q -й категории.

Обозначим эту вероятность потерь символом $P_{\text{пр}}(Q)$. Данная вероятность есть вероятность прерывания обслуживания заявки Q -й категории требованиями вышестоящих категорий. Индекс "пр" обозначает прерывание обслуживания заявки Q -й категории.

Для определения этой вероятности воспользуемся выражением для определения вероятности потерь по нагрузке. В общем случае вероятность потерь по нагрузке равна отношению интенсивности потерянной нагрузки к потенциальной нагрузке за определенный промежуток времени (как правило, за ЧНН):

$$P_{\text{н}} = \frac{R}{A}, \quad (7.74)$$

где R – интенсивность потерянной нагрузки, A – интенсивность потенциальной нагрузки. Для источников нагрузки Q -й категории выражение (7.74) записывается в следующем виде:

$$P_{\text{пр}}(Q) = P_{\text{н}} = \frac{R_Q}{A_Q}. \quad (7.75)$$

Ввиду того, что поток вызовов от абонентов Q -й категории является простейшим, то потенциальная нагрузка равна поступающей ($A_Q = Z_Q$). Поэтому выражение (7.75) можно переписать в виде

$$P_{\text{пр}}(Q) = \frac{R_{\text{пр}}(Q)}{Z_Q}, \quad (7.76)$$

где $R_{\text{пр}}(Q)$ – величина потерянной в результате прерывания нагрузки Q категории;

Z_Q – поступающая нагрузка от рассматриваемой категории.

Рассмотрим физическую интерпретацию величины нагрузки Z_Q . По причине того, что прервать можно лишь ту нагрузку источников Q -й категории, которая уже была принята на обслуживание

(потерянную в результате занятости каналов нагрузку нельзя прерывать, так как эти заявки уже покинули систему), то нагрузка Z_Q по своей физической сущности является обслуженной нагрузкой источников Q -й категории (Y_Q) при вероятности потерь, равной $P_V(Q)$. Следовательно, выражение (7.76) можно еще раз переписать в виде

$$P_{\text{пр}}(Q) = \frac{R_{\text{пр}}(Q)}{Y_Q}, \quad (7.77)$$

где интенсивность взятой на обслуживание нагрузки Y_Q определяется как

$$Y_Q = Z_Q \cdot (1 - P_V(Q)), \quad (7.78)$$

где $P_V(Q)$ – вероятность потерь вызовов Q -й категории по причине занятости всех каналов.

Далее определим $R_{\text{пр}}(Q)$. Величина $R_{\text{пр}}(Q)$ в формуле (7.77) совпадает с той дополнительной нагрузкой источников высших категорий 1, 2, 3, ..., ($Q-1$), которая обслуживается за счет прерывания заявок Q -й категории, т. е.

$$R_{\text{пр}}(Q) = -\Delta Y_{\text{пр}}(Q-1), \quad (7.79)$$

где знак "–" означает прирост обслуженной нагрузки для источников высших категорий, $\Delta Y_{\text{пр}}(Q-1)$ – величина дополнительно обслуженной нагрузки высших категорий пользователей ($i = 1, \dots, Q-1$).

Величину $\Delta Y_{\text{пр}}(Q-1)$ определим следующим образом. Для этого вначале рассмотрим потерянную нагрузку высших категорий 1, 2, 3, ..., ($Q-1$) в беспriorитетной системе коммутации (когда все источники любых категорий равны между собой):

$$R_{1,Q-1(\text{б/п})} = \sum_{i=1}^{Q-1} Z_i \cdot P_V(Q). \quad (7.80)$$

В рассматриваемой приоритетной системе нагрузка абонентов Q -й категории не оказывает влияния на качество обслуживания вы-

шестоящих категорий ($i = 1, \dots, Q-1$), поэтому фактически потерянная нагрузка для источников $i = 1, \dots, Q-1$ будет составлять

$$R_{1,Q-1} = \sum_{i=1}^{Q-1} Z_i \cdot R_V(i). \quad (7.81)$$

Следовательно, величина $\Delta Y_{\text{пр}}(Q-1)$ для источников $(Q-1)$ категории будет равна разности

$$\Delta Y_{\text{пр}}(Q-1) = \sum_{i=1}^{Q-1} Z_i \cdot [R_V(Q) - R_V(i)]. \quad (7.82)$$

Подставляя формулы (7.78) и (7.82) в выражение (7.77), получим

$$P_{\text{пр}}(Q) = \frac{\sum_{i=1}^{Q-1} Z_i \cdot [R_V(Q) - R_V(i)]}{Z_Q \cdot [1 - R_V(Q)]}. \quad (7.83)$$

Так как приоритетное прерывание возможно лишь при условии, что нагрузка низшего приоритета уже взята на обслуживание, то эти события совместны и для определения вероятности прерывания необходимо использовать условные вероятности.

Согласно [4], условной вероятностью называют вероятность события "B" (в данном конкретном случае – событие, заключающееся в прерывании обслуживания заявок низших категорий), вычисленную в предположении о том, что событие "A" произошло (т. е. нагрузка низших категорий взята на обслуживание). Совместная вероятность появления двух событий равна произведению вероятности одного из них на условную вероятность другого в предположении, что первое событие уже наступило.

Вероятность первого события – это есть вероятность обслуживания заявки низшей (Q) категории, которая определяется по формуле $(1 - R_V(Q))$, а условная вероятность второго события при условии, что первое уже произошло, – $P_{\text{пр}}(Q)$.

Отсюда вероятность совместного появления двух событий (взятия на обслуживание заявки Q категории с вероятностью $1 - R_V(Q)$, затем ее прерывание с вероятностью $P_{\text{пр}}(Q)$) будет составлять

$$P_{\text{пр}}(Q) \cdot [1 - P_V(Q)]. \quad (7.84)$$

Таким образом, суммарные потери заявок абонентов i -й категории ($i = \overline{2-Q}$) для рассмотренных выше событий можно записать в виде

$$P_i = P_V(i) + P_{\text{пр}}(i) \cdot [1 - P_V(i)]. \quad (7.85)$$

7.2.3. Модель системы с ожиданием $M/M/V/W, r/FF/R$

Автоматизация процессов обслуживания вызовов, наблюдаемая не только на телефонных сетях общего пользования, но также и на сетях специальной связи, не исключает способ обслуживания с ожиданием, особенно при ограниченном числе каналов в направлениях связи. Дело в том, что данный способ обслуживания имеет очень важное для междугородной связи свойство, заключающееся в том, что при обслуживании с ожиданием (по заказной системе обслуживания) у оператора существует возможность перевести часть трафика с часа наибольшей нагрузки (ЧНН) на другое время, когда интенсивность требований на установление соединений уменьшается. Это приводит к более рациональному и равномерному использованию междугородных каналов связи в перегруженных направлениях и к возможности обслуживания всей нагрузки, хотя со сдвигом во времени.

Часто сравнивают два способа обслуживания (с потерями и ожиданием) с целью рационального их выбора для обеспечения требуемого качества обслуживания. Сделать это достаточно сложно.

Причина в том, что в условиях резкого возрастания нагрузки (перегрузки сети) более предпочтительным выглядит способ обслуживания с потерями, так как обслуживание вызовов с ожиданием в этом случае ведет к большому количеству ожидающих вызовов и снижению пропускной способности по сравнению с системами с потерями.

Совершенно иные выводы следуют из рассмотрения целесообразности использования дисциплины обслуживания с ожиданием в управляющих устройствах. Постоянная длительность обслуживания и малое ее значение позволяют устанавливать большие значения условной вероятности потерь: $P(t_{\text{ож}} > t_{\text{доп}})$.

Это объясняется тем, что в данном случае вероятность $P(t_{\text{ож}} > t_{\text{доп}})$ при относительно больших t оказывается очень малой величиной и одновременно допустимое время ожидания вызовом начала обслуживания имеет небольшое значение, не ощущаемое источником вызова.

7.2.3.1. Постановка задачи по обслуживанию потока вызовов системой с ожиданием $M/M/V/W, r \rightarrow \infty / FF/R$

В символике Кендалла – Башарина рассматриваемая модель обозначается следующими символами: $M/M/V/W, r \rightarrow \infty / FF/R$.

Исходя из данной символики, можно отметить:

1. Предполагается, что на входы коммутационной системы поступает простейший поток вызовов. При этом параметр потока есть величина постоянная – $\lambda_r = \text{const}$.

2. Длительность занятия линии при обслуживании вызова t_3 является случайной величиной, распределенной по показательному закону:

$$F(t) = P(t_3 < t) = 1 - e^{-\mu t},$$

где $\mu = \frac{1}{t_3}$ – параметр показательного закона распределения потока освобождения занятой линии.

Необходимо отметить, что число вызовов, поступающих за среднюю длительность занятия, определяет поступающую нагрузку, равную $Z = \frac{\lambda}{\mu}$.

Одним из ограничений данной модели является то, что интенсивность поступающей нагрузки $Z < V$.

3. Символ в третьем разряде V ($1 \leq V \leq \infty$) обозначает, что коммутационное поле (КП) системы распределения информации не блокируемое, имеет одно звено, в выходы которого включен полнодоступный пучок линий емкостью V . Следовательно, ожидание в системе будет обусловлено только занятостью выделенных для обслуживания каналов связи.

4. Символы в четвертом разряде ($W, r \rightarrow \infty$) обозначают способ обслуживания с ожиданием, при этом число мест ожидания не ограничено.

5. Символы FF , стоящие в пятом разряде, свидетельствуют о том, что вызовы из очереди обслуживаются в порядке поступления (*First-in-First-out* (первым пришел – первым ушел)).

6. Порядок занятия обслуживающих приборов определяется символом R (*Random*), что означает случайный порядок занятия линий.

Совокупность всех ограничений и допущений создает так называемое концептуальное представление объекта исследования.

Для решения задачи при данных допущениях и ограничениях требуется определить:

1. Основные характеристики качества обслуживания (ХКО):

- функцию распределения длительности ожидания вызовом начала обслуживания $P(t_{ож} > t_{доп})$;

- вероятность ожидания для поступившего вызова $P(t_{ож} > 0)$.

2. Вспомогательные ХКО:

- среднюю длительность ожидания начала обслуживания для ждущих вызовов $\bar{t}_{ож_{жв}}$;

- среднюю длительность ожидания начала обслуживания для всех вызовов $\bar{t}_{ож}$;

- среднюю длину очереди $\bar{r}$.

Задача в такой постановке впервые сформулирована и решена А. К. Эрлангом. Полученное им выражение для определения $P(t_{ож} > 0)$ получило название второй формулы Эрланга.

7.2.3.2. Вероятности состояний системы. Второе распределение Эрланга

В данной модели, как и в модели полнодоступного пучка с отказами, для определения характеристик качества обслуживания достаточно рассмотреть только макросостояния системы (состояния, определяемые одной переменной). Эти состояния можно представить в виде графа состояний, изображенного на рис. 7.12.

Общее число вызовов, находящихся в системе на обслуживании и в очереди в момент времени t , обозначим через i ($i = 0, \dots, V + r$).

При этом если в момент времени t величина $i = 0, \dots, V - 1$, то i характеризует число занятых выходов в системе или число обслуживаемых вызовов.

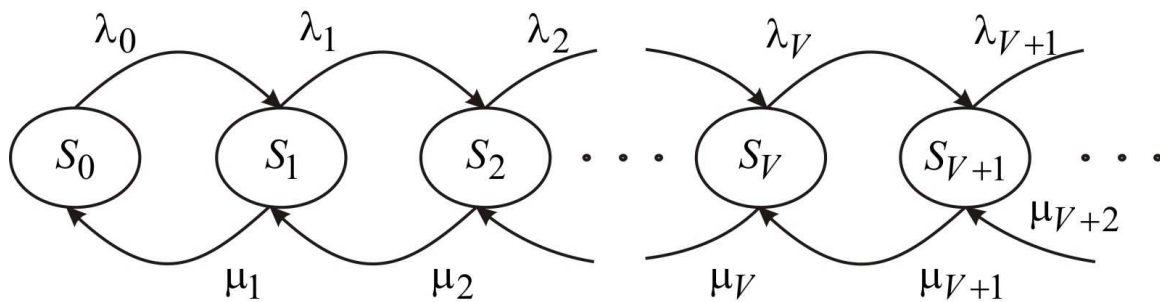


Рис. 7.12. Диаграмма состояний и переходов в системе обслуживания, состоящей из полнодоступного пучка ёмкостью V линий и неограниченного числа комплектов ожидания.

При $i = V + r \rightarrow \infty$ величина i показывает, что на обслуживании находятся V вызовов (число занятых выходов равно V) и в очереди ожидают обслуживания еще $r \rightarrow \infty$ вызовов (при этом разность $i - V$ равна числу вызовов, находящихся на ожидании).

В общем случае, как представлено на схеме графа состояний, параметр потока вызовов для каждого состояния различен и обозначается символом λ_i . Однако в условии задачи сказано, что поток вызовов аппроксимируется моделью простейшего потока, следовательно, во всех состояниях системы параметр потока $\lambda_r = \text{const}$, что упрощает решение поставленной задачи.

Поток освобождений до момента занятости всех каналов связи (включая и это состояние $i \leq V$) является потоком с условным параметром, у которого параметр потока зависит от состояния системы или, другими словами, прямо пропорционален числу занятых каналов и интенсивности освобождения одного канала.

Действительно, если в системе занята одна линия (канал связи), то интенсивность ее освобождения будет равна μ (где $\mu = \frac{1}{t_3}$ – интенсивность освобождения канала). При занятости двух линий интенсивность освобождения будет равна 2μ , при трех – 3μ , при занятости всех линий – $V\mu$.

В состояниях системы $i = V, \dots, V + r$, если предположить, что после освобождения канала он тут же занимается ожидающим вызовом, можно считать, что интенсивность потока освобождений будет величиной постоянной и равной $V \cdot \mu$ (следует напомнить, что постоянный параметр характерен для модели простейшего потока вызо-

вов). Исходя из этих рассуждений, можно записать формулу для интенсивности освобождения каналов связи системы коммутации:

$$\mu_i = \begin{cases} i \cdot \mu, & 0 \leq i \leq V; \\ V \cdot \mu & i \geq V. \end{cases} \quad (7.86)$$

Данный процесс обслуживания при рассмотренных ограничениях и допущениях является марковским процессом рождения и гибели со счетным множеством состояний. В соответствии с ранее полученной формулой для определения вероятностей состояний полнодоступного пучка с потерями ($0 \leq i \leq V$) для условия стационарного режима можно записать формулу для определения вероятностей состояний:

$$P_i = \frac{\lambda_0 \cdot \lambda_1 \cdot \lambda_2 \cdot \dots \cdot \lambda_{i-1}}{\mu_1 \cdot \mu_2 \cdot \mu_3 \cdot \dots \cdot \mu_i} \cdot P_0. \quad (7.87)$$

В формуле (7. 87), учитывая, что входящий поток вызовов является простейшим ($\lambda_i = \lambda = \text{const}$), а интенсивность освобождения линий $\mu_i = i \cdot \mu$, где $\mu = \frac{1}{t_3} \left(\frac{\lambda}{\mu} = Z \right)$, то для условия $0 \leq i \leq V$ вероятность состояния системы можно записать в виде

$$P_i = \frac{\lambda^i}{\mu^i \cdot 1 \cdot 2 \cdot 3 \cdot \dots \cdot (i-1) \cdot i} \cdot P_0 = \frac{\left(\frac{\lambda}{\mu} \right)^i}{i!} \cdot P_0 = \frac{Z^i}{i!} \cdot P_0. \quad (7.88)$$

Для условий $i > V$ вероятности состояний можно определить следующим образом:

$$\begin{aligned} P_i &= \frac{\lambda_0 \cdot \lambda_1 \cdot \lambda_2 \cdot \dots \cdot \lambda_{V-1}}{\mu_1 \cdot \mu_2 \cdot \mu_3 \cdot \dots \cdot \mu_V} \cdot \frac{\lambda_V \cdot \lambda_{V+1} \cdot \lambda_{V+2} \cdot \dots \cdot \lambda_{i-1}}{\mu_{V+1} \cdot \mu_{V+2} \cdot \mu_{V+3} \cdot \dots \cdot \mu_i} \cdot P_0 = \\ &= \frac{\lambda^V}{\mu^V \cdot V!} \cdot \frac{\lambda^{i-V}}{\mu^{i-V} \cdot V^{i-V}} \cdot P_0 = \\ &= \frac{Z^V}{V!} \cdot \left(\frac{\left(\frac{\lambda}{\mu} \right)^{i-V}}{V^{i-V}} \right) \cdot P_0 = \frac{Z^V}{V!} \cdot \left(\frac{Z}{V} \right)^{i-V} \cdot P_0. \end{aligned} \quad (7.89)$$

Таким образом, выражение (7.89) позволяет определить вероятности состояний системы для условий, когда в системе заняты все ОП и в очереди находятся r ждущих абонентов. Однако для вычисления P_i необходимо уметь вычислять вероятность нахождения системы в свободном состоянии, когда ни один канал не занят P_0 .

Данную вероятность P_0 , полученную ранее для системы с потерями, использовать для системы с ожиданием нельзя по причине того, что состояний в системе с ожиданием значительно больше ($i = 0, \dots, V + r$), чем в системе с потерями.

Для нахождения P_0 воспользуемся тем обстоятельством, что сумма вероятности всех состояний равна единице (условие нормировки $\sum_{j=0}^{\infty} P_j = 1$).

Для этого запишем уравнение для состояний системы при $i \leq V$ и $i > V$:

$$\sum_{j=0}^V \frac{Z^j}{j!} \cdot P_0 + \sum_{j=V+1}^{\infty} \frac{Z^V}{V!} \cdot \left(\frac{Z}{V}\right)^{j-V} \cdot P_0 = 1. \quad (7.90)$$

Из уравнения (7.90) получим

$$P_0 = \frac{1}{\sum_{j=0}^V \frac{Z^j}{j!} + \sum_{j=V+1}^{\infty} \frac{Z^V}{V!} \cdot \left(\frac{Z}{V}\right)^{j-V}}. \quad (7.91)$$

Несмотря на простоту выражения (7.91), получить численный результат для значения P_0 затруднительно, так как в знаменателе (второе слагаемое) присутствует бесконечная сумма.

Вынесем из суммы второго слагаемого знаменателя сомножитель, не зависящий от j . В этом случае $\sum_{j=V+1}^{\infty} \left(\frac{Z}{V}\right)^{j-V}$ представляет собой сумму бесконечной геометрической прогрессии. Из курса математики известно, что геометрическая прогрессия – это последовательность чисел, в которой отношение между последующим и предыдущим членами остается неизменными. Это отношение называется знаменателем прогрессии [20].

При $Z > V$ ряд $\left(\frac{Z}{V}\right)^{j-V}$ расходится и сумма его стремится к бесконечности. Соответственно, $P_0 \rightarrow 0$ и все вероятности $P_i \rightarrow 0$.

Отсюда можно сделать вывод, что при интенсивности поступающей нагрузки Z , равной или большей числа выходов V системы коммутации, с вероятностью, равной единице, постоянно будут заняты все выходы, и длина очереди будет бесконечной.

Поэтому чтобы система могла функционировать нормально и очередь не росла безгранично, необходимо выполнить условие $Z < V$. В этом случае прогрессия $\sum_{j=V+1}^{\infty} \left(\frac{Z}{V}\right)^{j-V}$ будет убывающей.

Как известно, сумма (S) бесконечно убывающей прогрессии [20] выражается формулой $S = \frac{a_1}{1-q}$, где $a_1 = \left(\frac{Z}{V}\right)^1$ – первый член геометри-

ческой прогрессии; $q = \frac{\left(\frac{Z}{V}\right)^i}{\left(\frac{Z}{V}\right)^{i-1}} = \frac{Z}{V}$ – знаменатель прогрессии. Отсю-

да сумму геометрической прогрессии можно представить в виде

$$\sum_{j=V+1}^{\infty} \left(\frac{Z}{V}\right)^{j-V} = \frac{Z}{V-Z}. \quad (7.92)$$

С учетом (7.92) выражение (7.91) записывается в виде

$$P_0 = \frac{1}{\sum_{j=0}^V \frac{Z^j}{j!} + \sum_{j=V+1}^{\infty} \frac{Z^V}{V!} \cdot \left(\frac{Z}{V}\right)^{j-V}} = \frac{1}{\sum_{j=0}^V \frac{Z^j}{j!} + \frac{Z}{V-Z} \cdot \frac{Z^V}{V!}}. \quad (7.93)$$

Полученное выражение (7.93) позволяет достаточно просто определить значение вероятности P_0 , так как обычно V является величиной конечной.

Используя формулу (7.93), стационарные вероятности состояний (7.88) и (7.89) можно записать в виде

$$P_i = \frac{\frac{Z^i}{i!}}{\sum_{j=0}^V \frac{Z^j}{j!} + \frac{Z}{V-Z} \cdot \frac{Z^V}{V!}}, \quad 0 \leq i \leq V;$$

$$P_i = \frac{\frac{Z^V}{V!} \cdot \left(\frac{Z}{V}\right)^{i-V}}{\sum_{j=0}^V \frac{Z^j}{j!} + \frac{Z}{V-Z} \cdot \frac{Z^V}{V!}}, \quad i > V. \quad (7.94)$$

Уравнения (7.94) получили название второго распределения Эрланга. Данные выражения можно преобразовать в более удобный для расчета вид.

Для этого разделим числитель и знаменатель выражений (7.94) на величину, равную $\sum_{i=0}^V \frac{Z^i}{i!}$:

$$P_i = \frac{\frac{\frac{Z^i}{i!}}{\sum_{j=0}^V \frac{Z^j}{j!}}}{\frac{\sum_{j=0}^V \frac{Z^j}{j!}}{\sum_{j=0}^V \frac{Z^j}{j!}} + \frac{\frac{Z^V}{V!}}{\sum_{j=0}^V \frac{Z^j}{j!}} \cdot \frac{Z}{V-Z}}, \quad 0 \leq i \leq V. \quad (7.95)$$

$$P_i = \frac{\frac{Z^V}{V!} \cdot \left(\frac{Z}{V}\right)^{i-V}}{\sum_{j=0}^V \frac{Z^j}{j!}} = \frac{E_V(Z) \cdot \left(\frac{Z}{V}\right)^{i-V}}{1 + E_V(Z) \cdot \frac{Z}{V-Z}}, \quad i > V. \quad (7.96)$$

$$\frac{\sum_{j=0}^V \frac{Z^j}{j!} + \frac{\frac{Z^V}{V!} \cdot \frac{Z}{V-Z}}{\sum_{j=0}^V \frac{Z^j}{j!}}}{\sum_{j=0}^V \frac{Z^j}{j!}}$$

Обозначив $E_{i,V}(Z) = \frac{\frac{Z^i}{i!}}{\sum_{j=0}^V \frac{Z^j}{j!}}$, а $E_V(Z) = \frac{\frac{Z^V}{V!}}{\sum_{j=0}^V \frac{Z^j}{j!}}$,

формула (7.95) для случая $0 \leq i \leq V$ записывается в виде

$$P_i = \frac{E_{i,V}(Z)}{1 + E_V(Z) \cdot \frac{Z}{V-Z}}, \quad (0 \leq i \leq V). \quad (7.97)$$

Выражение (7.96) для условия $i > V$ преобразуется к следующему виду:

$$P_i = \frac{\frac{Z^V}{V!} \cdot \left(\frac{Z}{V}\right)^{i-V}}{\sum_{j=0}^V \frac{Z^j}{j!}} = \frac{E_V(Z) \cdot \left(\frac{Z}{V}\right)^{i-V}}{1 + E_V(Z) \cdot \frac{Z}{V-Z}}. \quad (7.98)$$

$$\frac{\sum_{j=0}^V \frac{Z^j}{j!} + \frac{\frac{Z^V}{V!} \cdot \frac{Z}{V-Z}}{\sum_{j=0}^V \frac{Z^j}{j!}}}{\sum_{j=0}^V \frac{Z^j}{j!}}$$

Так как часто сравнивают между собой системы с потерями и ожиданием, рассмотрим значения величин их вероятностей состояний.

Сравнивать системы следует для случая $0 \leq i \leq V$, при этом:

– вероятности состояний системы с потерями определяются по формуле

$$P_i = E_{i,V}(Z) = \frac{\frac{Z^i}{i!}}{\sum_{j=0}^V \frac{Z^j}{j!}}; \quad (7.99)$$

– вероятности состояний системы с ожиданием рассчитываются по формуле

$$P_i = \frac{E_{i,V}(Z)}{1 + E_V(Z) \cdot \frac{Z}{V-Z}}. \quad (7.100)$$

Из формулы (7.100) видно, что какое бы значение ни принимала величина $0 \leq E_V(Z) \leq 1$ (за исключением крайнего, равного 0), знаменатель всегда больше единицы. Поэтому значения вероятностей P_i для системы с ожиданием всегда меньше, чем аналогичные вероятности для систем с потерями.

Сравнение двух систем с использованием первого ($E_i(Z)$) и второго распределений Эрланга (P_i) для случая $0 \leq i \leq V$ приведено на рис. 7.13.

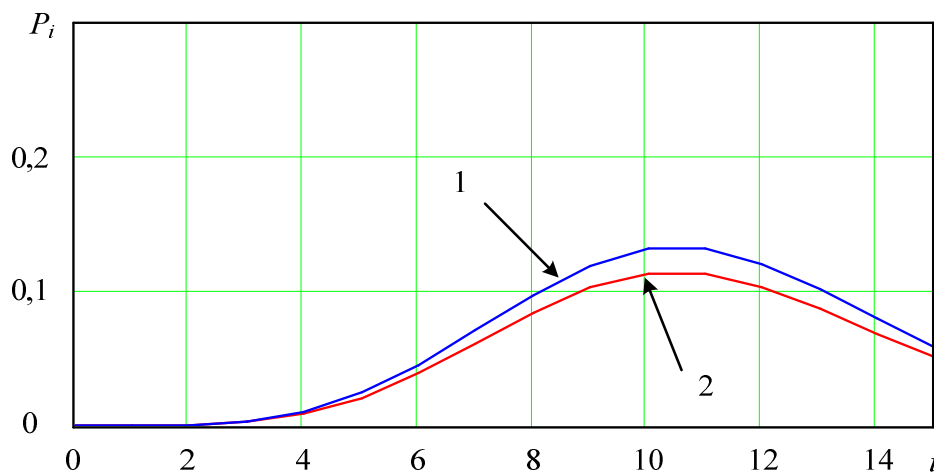


Рис. 7.13. Зависимости вероятностей P_i от состояний i для систем с ожиданием (2) и потерями (1) при $V = 15$ и $Z = 11$ Эрл

Физически это можно интерпретировать следующим образом. Время, в течение которого коммутационная система находится в состояниях, когда вызовы немедленно обслуживаются ($i < V$), больше для систем с потерями, чем для систем с ожиданием.

Отсюда нагрузка, обслуженная системой с потерями (теорема о количественной оценке обслуженной нагрузки), больше нагрузки, которая обслуживается немедленно системой с ожиданием (т. е. без учета обслуженной нагрузки, создаваемой ожидающими вызовами).

7.2.3.3. Характеристики качества обслуживания вызовов системой $M/M/V/W, r/FF/R$

Для модели $M/M/V/W, r \rightarrow \infty/FF/R$ основными характеристиками качества обслуживания (ХКО) являются:

- функция распределения длительности ожидания вызовом начала обслуживания $P(t_{ож} > t_{доп})$;
- вероятность ожидания для поступившего вызова $P(t_{ож} > 0)$.

Вспомогательными ХКО являются:

- средняя длительность ожидания начала обслуживания для ждущих вызовов $\bar{t}_{ож_{жв}}$;
- средняя длительность ожидания начала обслуживания для всех вызовов $\bar{t}_{ож}$;
- средняя длина очереди $\bar{r}$.

Начнем с рассмотрения вероятности ожидания для поступившего вызова $P(t_{ож} > 0)$.

$P(t_{ож} > 0)$ есть вероятность того, что поступивший в произвольный момент времени вызов не будет немедленно обслужен, а ожидает начала обслуживания в течение времени $t_{ож} > 0$. Данную вероятность часто называют вероятностью наличия очереди. Для ее определения необходимо рассмотреть события, способствующие наличию очереди.

Очевидно, что ожидание начала обслуживания вызова возможно лишь при условии занятости всех линий в системе коммутации.

Для простейшего потока независимо от того, сколько вызовов находятся на обслуживании, вероятность поступления вызова не изменяется, следовательно, ожидание возможно, когда в системе заняты все V каналы связи, а также при занятости всех каналов и наличии

в очереди $r = 0, 1, 2, \dots, \infty$ ожидающих требований. Сумма вероятностей этих событий будет определять вероятность наличия очереди:

$$P(t_{\text{ож}} > 0) = \sum_{i=V}^{\infty} P_i. \quad (7.101)$$

В формуле (7.101) для определения $P(t_{\text{ож}} > 0)$ необходимо подставить $P_i = \frac{Z^V}{V!} \cdot \left(\frac{Z}{V}\right)^{i-V} \cdot P_0$ для случая $i > V$.

В этом случае можно записать следующее:

$$\begin{aligned} P(t_{\text{ож}} > 0) &= \sum_{i=V}^{\infty} \frac{Z^V}{V!} \cdot \left(\frac{Z}{V}\right)^{i-V} \cdot P_0 = \frac{Z^V}{V!} \cdot P_0 + \frac{Z^V}{V!} \cdot P_0 \cdot \sum_{i=V+1}^{\infty} \left(\frac{Z}{V}\right)^{i-V} = \\ &= \frac{Z^V}{V!} \cdot P_0 \cdot \left(1 + \sum_{i=V+1}^{\infty} \left(\frac{Z}{V}\right)^{i-V}\right). \end{aligned} \quad (7.102)$$

Так как в ограничениях для данной модели системы коммутации с ожиданием было указано, что $Z < V$, то сумма в выражении (7.102) есть сумма убывающей геометрической прогрессии:

$$\sum_{i=V+1}^{\infty} \left(\frac{Z}{V}\right)^{i-V} = \frac{Z}{V-Z}. \quad (7.103)$$

Подставив формулу (7.103) в выражение (7.102), получим

$$\begin{aligned} P(t_{\text{ож}} > 0) &= \frac{Z^V}{V!} \cdot P_0 \cdot \left(1 + \sum_{i=V+1}^{\infty} \left(\frac{Z}{V}\right)^{i-V}\right) = \frac{Z^V}{V!} \cdot P_0 \cdot \left(1 + \frac{Z}{V-Z}\right) = \\ &= \frac{Z^V}{V!} \cdot P_0 \cdot \left(\frac{V-Z+Z}{V-Z}\right) = \frac{Z^V}{V!} \cdot \frac{P_0}{\frac{V-Z}{V}} = \frac{Z^V}{V!} \cdot \frac{P_0}{\left(1 - \frac{Z}{V}\right)}. \end{aligned} \quad (7.104)$$

Разделим числитель и знаменатель (7.91) на $\sum_{i=0}^V \frac{Z^i}{i!}$, а также

учтем, что вероятность P_0 (вероятность незанятости ни одного канала связи) определяется следующим выражением:

$$P_0 = \frac{1}{\sum_{i=0}^V \frac{Z^i}{i!} + \frac{Z^V}{V!} \cdot \frac{Z}{V-Z}}. \quad (7.105)$$

В этом случае выражение (7.91) можно записать в виде

$$\begin{aligned} P(t_{\text{ож}} > 0) &= \frac{Z^V}{V!} \cdot \frac{P_0}{\left(1 - \frac{Z}{V}\right)} = \frac{\frac{Z^V}{V!}}{\sum_{i=0}^V \frac{Z^i}{i!} + \frac{Z^V}{V!} \cdot \frac{Z}{V-Z}} \cdot \frac{1}{1 - \frac{Z}{V}} = \\ &= \frac{E_V(Z)}{[1 + E_V(Z) \cdot \frac{Z}{V-Z}] \cdot \frac{V-Z}{V}} = \frac{E_V(Z)}{\frac{V-Z}{V} + E_V(Z) \cdot \frac{Z}{V-Z} \cdot \frac{V-Z}{V}} = \\ &= \frac{E_V(Z)}{1 - \frac{Z}{V} + E_V(Z) \cdot \frac{Z}{V}} = \frac{E_V(Z)}{1 - \frac{Z}{V} \cdot (1 - E_V(Z))}. \end{aligned} \quad (7.106)$$

Таким образом, вероятность того, что поступивший в произвольный момент вызов будет ожидать начала обслуживания в течение времени $t_{\text{ож}} > 0$:

$$P(t_{\text{ож}} > 0) = \frac{E_V(Z)}{1 - \frac{Z}{V} \cdot (1 - E_V(Z))}, \quad (7.107)$$

где $E_V(Z)$ – условная запись эрланговых потерь (первая формула Эрланга), Z – интенсивность поступающей нагрузки, V – канальный ресурс.

Выражение (7.107) в технической литературе часто называют второй формулой Эрланга. Ввиду важности данной формулы для оценки качества обслуживания вызовов (КОВ) на сетях связи она табулирована, т. е. созданы расчетные таблицы и номограммы. Номограмма, отражающая зависимость вероятности ожидания $P(t_{\text{ож}} > 0)$ от нагрузки и выделенного числа каналов, представлена на рис. 7.14.

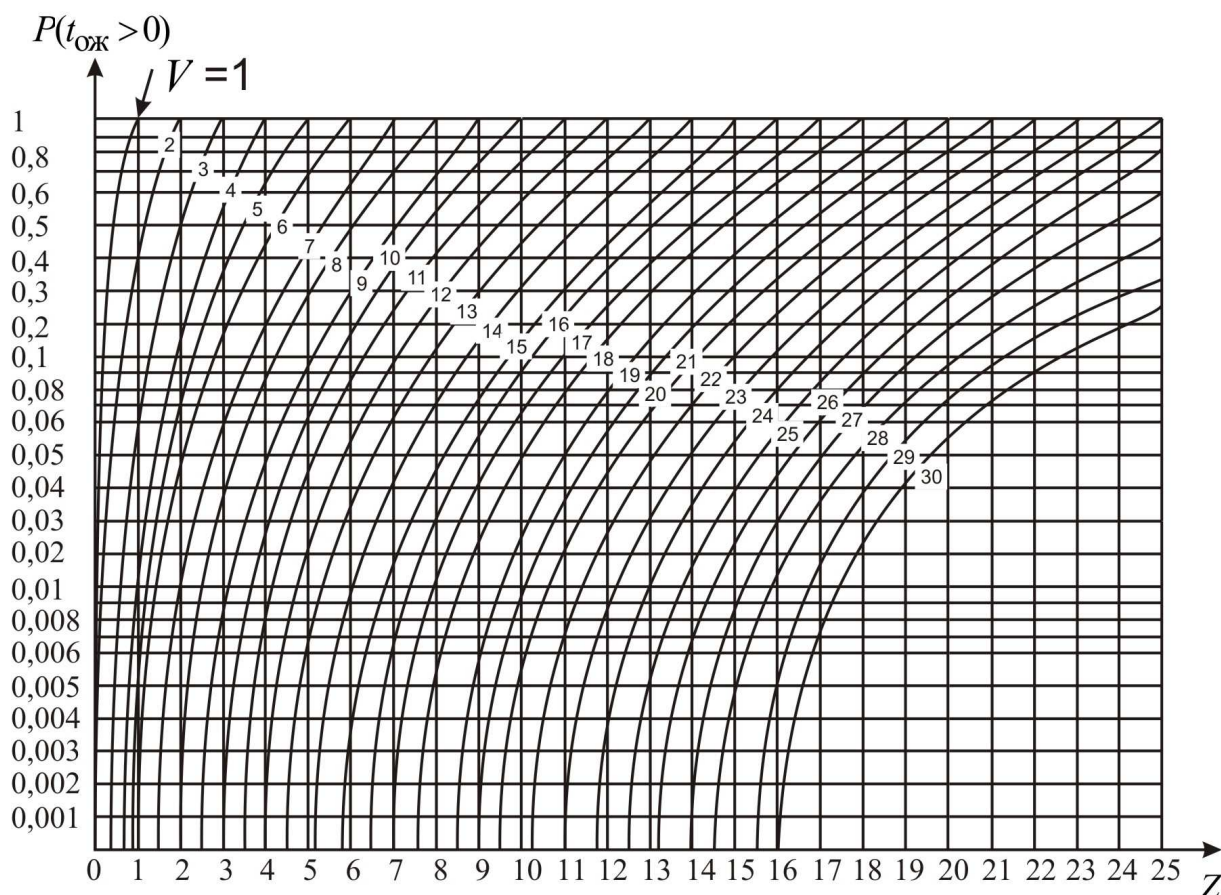


Рис. 7.14. Номограмма расчета вероятности ожидания ($P(t_{\text{ож}} > 0)$) в системе распределения информации

Сравнение формулы (7.107) с первой формулой Эрланга ($P_B = E_V(Z)$) позволяет сделать вывод о том, что при одинаковых значениях интенсивности нагрузки Z и числе линий V вероятность ожидания в системе с ожиданием $P(t_{\text{ож}} > 0)$ больше, чем вероятность потерь по вызовам P_B в системе с явными потерями, так как независимо от значения $E_V(Z)$ знаменатель всегда меньше единицы.

Физически данный факт объясняется существующим различием в дисциплинах обслуживания сравниваемых систем. Так, в системе с явными потерями поступивший вызов в момент занятости всех

линий теряется и никакого влияния на систему в последующем не оказывает, а в системе с ожиданием вызов становится на ожидание. При освобождении линии в системе с явными потерями она предоставляется новому поступающему вызову, а в системе с ожиданием при наличии очереди – ожидающему вызову. Вновь поступившему вызову в данную систему приходится становиться в очередь.

Отсюда можно сделать еще один вывод, заключающийся в том, что нагрузка, обслуженная системой с потерями, больше нагрузки, которая обслуживается немедленно системой с ожиданием (т. е. без учета обслуженной нагрузки, создаваемой ожидающими вызовами).

В связи с этим в области малых потерь, исчисляемых единицами промилле и даже несколькими процентами, когда абонент не ощущает неудобств, предпочтительнее система с потерями, обладающая более высокой пропускной способностью. При больших значениях потерь система с потерями не обеспечивает должного качества обслуживания абонентов и не пригодна для применения.

В случае больших потерь нельзя однозначно дать оценку и системам с ожиданием. Все зависит от допустимого времени ожидания $t_{\text{доп}}$. Поэтому требуется рассмотреть следующую важную характеристику качества обслуживания вызовов системой с ожиданием, а именно вероятность ожидания свыше допустимого времени $P(t_{\text{ож}} > t_{\text{доп}})$.

Величину $P(t_{\text{ож}} > t_{\text{доп}})$ часто называют условными потерями, так как задержка в предоставлении соединения абоненту свыше нормированного допустимого времени, естественно, ведет к снижению актуальности передаваемой информации.

В отличие от предыдущей характеристики, вероятность $P(t_{\text{ож}} > t_{\text{доп}})$ зависит от вида очереди. Наиболее простое выражение для вероятности $P(t_{\text{ож}} > t_{\text{доп}})$ получается при упорядоченной очереди – FF . Рассмотрим этот случай.

Для определения вероятности $P(t_{\text{ож}} > t_{\text{доп}})$ следует рассмотреть вначале события, обуславливающие задержку в предоставлении соединения абоненту свыше нормированного допустимого времени.

Время ожидания может превысить требуемое в следующих случаях:

– при занятости всех каналов связи и отсутствии в очереди ожидающих требований ($r = 0$). При этом очередной поступающий

вызов может ждать начало обслуживания большее время, чем регламентированное время $t_{\text{доп}}$ (время окончания обслуживания любого из занятых каналов больше, чем время $t_{\text{доп}}$);

– при занятости всех каналов связи и наличии в очереди ожидающих заявок источников нагрузки ($r = 1, \dots, \infty$).

Иными словами, во всех состояниях ($i \geq V$) возможно выполнение условия $t_{\text{ож}} > t_{\text{доп}}$.

Обозначим через $P_i(t_{\text{ож}} > t_{\text{доп}})$ условную вероятность превышения времени ожидания очередным вызовом требуемой величины $t_{\text{доп}}$ для каждого возможного состояния i ($i \geq V$), а через P_i – вероятность того, что система находится именно в этом конкретном i -м состоянии.

Величина $P_i(t_{\text{ож}} > t_{\text{доп}})$ есть вероятность того, что в любом состоянии i ($i = V, \dots, V + r$) системы распределения информации время ожидания превысит требуемую величину.

Все события, обуславливающие превышение времени ожидания вызовом требуемого времени, несовместны, так как при осуществлении одного из них другое событие невозможно и образуется полная группа событий.

Поэтому для определения $P(t_{\text{ож}} > t_{\text{доп}})$ используем формулу полной вероятности:

$$P(t_{\text{ож}} > t_{\text{доп}}) = \sum_{i=V}^{\infty} P_i \cdot P_i(t_{\text{ож}} > t_{\text{доп}}). \quad (7.108)$$

В выражении (7.108) вероятность нахождения системы в состоянии P_i рассчитывается по ранее приведенной формуле для $i > V$:

$$P_i = \frac{\frac{Z^V}{V!} \cdot \left(\frac{Z}{V}\right)^{i-V}}{\sum_{j=0}^V \frac{Z^j}{j!} + \frac{Z}{V-Z} \cdot \frac{Z^V}{V!}} = \frac{E_V(Z) \cdot \left(\frac{Z}{V}\right)^{i-V}}{1 + E_V(Z) \cdot \frac{Z}{V-Z}}, \quad i > V. \quad (7.109)$$

Рассмотрим порядок определения вероятности $P_i(t_{\text{ож}} > t_{\text{доп}})$.

Вероятность $P_i(t_{\text{ож}} > t_{\text{доп}})$ можно рассматривать как вероятность того, что в i -м состоянии в системе за промежуток времени

длиной $t_{\text{доп}}$ после момента поступления рассматриваемого вызова произойдет не более $(i - V)$ освобождений, в то время как данный вызов стоит в очереди на $(i - V + 1)$ месте.

Эти вероятности равносильны, так как и в первом, и во втором случаях время ожидания $t_{\text{ож}}$ превысит допустимое время $t_{\text{доп}}$.

По условию задачи длительность обслуживания одного вызова без учета времени ожидания распределена по показательному закону:

$$F(t) = P(T < t) = 1 - e^{-\mu t}. \quad (7.110)$$

Величина $F(t)$ в вероятностном смысле характеризует событие, что случайная величина времени освобождения T будет меньше рассматриваемого промежутка времени $t_{\text{доп}}$. Иными словами, в рассматриваемом промежутке времени $t_{\text{доп}}$ произойдет освобождение (рис. 7.15).

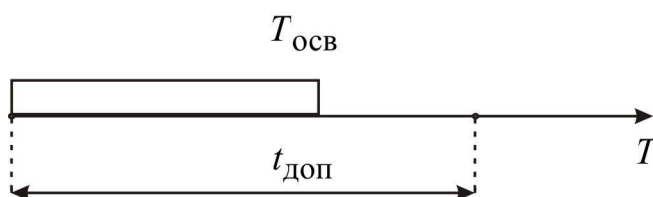


Рис. 7.15. Условия освобождения в рассматриваемом промежутке времени

В этом случае вероятность (функция распределения промежутков между моментами освобождения линий пучка) того, что произойдет освобождение за время $t_{\text{доп}}$ при условии занятости в пучке всех V каналов (линий) будет составлять

$$F_{\text{осв}}(t) = P(T_{\text{осв}} < t) = 1 - e^{-V\mu t}. \quad (7.111)$$

Из формулы (7.111) следует, что промежутки между освобождениями в рассматриваемом потоке распределены по экспоненциальному закону. Следовательно, поток освобождений за время ожидания начала обслуживания является простейшим по следующим причинам:

– параметр потока освобождений во время ожидания обслуживания $V \cdot \mu = \text{const}$ (во время ожидания обслуживания в очереди каждый освободившийся канал тут же занимается очередным вызовом, находящимся в очереди, следовательно, параметр потока освобождения

ния является постоянной величиной и пропорционален числу выделенных каналов (V) в данном направлении связи и интенсивности освобождения одного канала связи $\left(\mu = \frac{1}{t_{\text{зан}}}\right)$;

– промежутки между моментами освобождений в рассматриваемом потоке распределены по экспоненциальному закону: $F_{\text{осв}}(t) = P(T_{\text{осв}} < t) = 1 - e^{-V\mu t}$.

Так как поток освобождений во время ожидания обслуживания источника в очереди является простейшим, то для определения вероятности освобождения точно j линий за время $t_{\text{доп}}$ можно использовать формулу Пуассона

$$P_j = \frac{(V \cdot \mu \cdot t_{\text{доп}})^j}{j!} \cdot e^{-V\mu t_{\text{доп}}}. \quad (7.112)$$

Для упрощения $t_{\text{доп}}$ заменим символом t . Тогда вероятность освобождения не более $(i - V)$ линий за время t (рис. 7.16), что равносильно вероятности $P_i(t_{\text{ож}} > t_{\text{доп}})$, можно определить по формуле

$$P_i(t_{\text{ож}} > t) = \sum_{j=0}^{i-V} P_j = \sum_{j=0}^{i-V} \frac{(V \cdot \mu \cdot t)^j}{j!} \cdot e^{-V\mu t}. \quad (7.113)$$

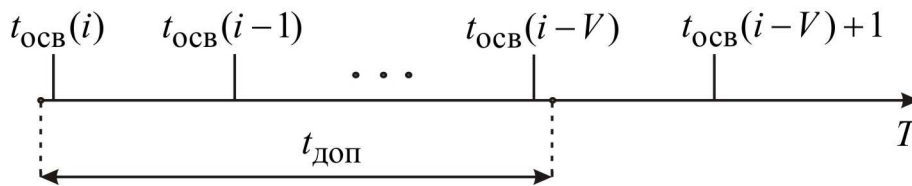


Рис. 7.16. Условия освобождения в рассматриваемом промежутке времени

Подставляя формулы (7.109) и (7.113) в выражение (7.108), получим

$$P(t_{\text{ож}} > t_{\text{доп}}) = \sum_{i=V}^{\infty} P_i \cdot \sum_{j=0}^{i-V} \frac{(V \cdot \mu \cdot t)^j}{j!} \cdot e^{-V\mu t} =$$

$$\begin{aligned}
&= \sum_{i=V}^{\infty} \frac{Z^V}{V!} \cdot \left(\frac{Z}{V}\right)^{i-V} \cdot P_0 \cdot \sum_{j=0}^{i-V} \frac{(V \cdot \mu \cdot t)^j}{j!} \cdot e^{-V\mu t} = \\
&= \frac{Z^V}{V!} \cdot e^{-V\mu t} \cdot P_0 \cdot \sum_{i=V}^{\infty} \left(\frac{Z}{V}\right)^{i-V} \cdot \sum_{j=0}^{i-V} \frac{(V \cdot \mu \cdot t)^j}{j!}. \quad (7.114)
\end{aligned}$$

Меняя порядок суммирования, выражение (7.114) можно записать в виде

$$\begin{aligned}
P(t_{\text{ож}} > t_{\text{доп}}) &= \frac{Z^V}{V!} \cdot e^{-V\mu t} \cdot P_0 \cdot \sum_{i=V}^{\infty} \left(\frac{Z}{V}\right)^{i-V} \cdot \sum_{j=0}^{i-V} \frac{(V \cdot \mu \cdot t)^j}{j!} = \\
&= \frac{Z^V}{V!} \cdot e^{-V\mu t} \cdot P_0 \cdot \sum_{j=0}^{\infty} \frac{(V \cdot \mu \cdot t)^j}{j!} \cdot \frac{Z^j}{V^j} \cdot \sum_{i=V+j}^{\infty} \left(\frac{Z}{V}\right)^{i-(V+j)}. \quad (7.115)
\end{aligned}$$

Так как в выражении (7.115) $\sum_{i=V+j}^{\infty} \left(\frac{Z}{V}\right)^{i-(V+j)}$ есть сумма бесконечно убывающей геометрической прогрессии, то по аналогии с преобразованием в формуле (7.114) получим

$$\sum_{i=V+j}^{\infty} \left(\frac{Z}{V}\right)^{i-(V+j)} = \frac{1}{1 - \frac{Z}{V}}. \quad (7.116)$$

В этом случае первый член геометрической прогрессии

$$a_1 = \left(\frac{Z}{V}\right)^{V+j-(V+j)} = 1,$$

а знаменатель (q) прогрессии

$$q = \frac{\left(\frac{Z}{V}\right)}{1} = \frac{Z}{V}.$$

Подставляя формулу (7.116) в выражение (7.115) и учитывая, что $Z = \frac{\lambda}{\mu}$, получим следующее выражение:

$$\begin{aligned}
 P(t_{\text{ож}} > t_{\text{доп}}) &= \frac{Z^V}{V!} \cdot e^{-V\mu t} \cdot P_0 \cdot \sum_{j=0}^{\infty} \frac{(V \cdot \mu \cdot t)^j}{j!} \cdot \frac{Z^j}{V^j} \cdot \sum_{i=V+j}^{\infty} \left(\frac{Z}{V}\right)^{i-(V+j)} = \\
 &= \frac{\frac{Z^V}{V!} \cdot e^{-V\mu t} \cdot P_0}{1 - \frac{Z}{V}} \cdot \sum_{j=0}^{\infty} \frac{(V \cdot \mu \cdot t)^j}{j!} \cdot \frac{Z^j}{V^j} = \frac{\frac{Z^V}{V!} \cdot e^{-V\mu t} \cdot P_0}{1 - \frac{Z}{V}} \cdot \sum_{j=0}^{\infty} \frac{t^j}{j!} \cdot \frac{V^j \cdot \mu^j \cdot Z^j}{V^j} = \\
 &= \frac{\frac{Z^V}{V!} \cdot e^{-V\mu t} \cdot P_0}{1 - \frac{Z}{V}} \cdot \sum_{j=0}^{\infty} \frac{(\lambda \cdot t)^j}{j!}, \tag{7.117}
 \end{aligned}$$

где $\mu \cdot Z = \mu \cdot \frac{\lambda}{\mu} = \lambda$, $t = t_{\text{доп}}$. Так как $\sum_{j=0}^{\infty} \frac{(\lambda \cdot t)^j}{j!} = e^{\lambda t}$, то выражение (7.117) можно привести к следующему виду:

$$\begin{aligned}
 P(t_{\text{ож}} > t_{\text{доп}}) &= \frac{\frac{Z^V}{V!} \cdot e^{-V\mu t} \cdot P_0}{1 - \frac{Z}{V}} \cdot \sum_{j=0}^{\infty} \frac{(\lambda \cdot t)^j}{j!} = \\
 &= \frac{\frac{Z^V}{V!} \cdot P_0}{1 - \frac{Z}{V}} \cdot e^{-V\mu t} \cdot e^{\lambda t} = \frac{\frac{Z^V}{V!} \cdot P_0}{1 - \frac{Z}{V}} \cdot e^{-(V\mu - \lambda)t}. \tag{7.118}
 \end{aligned}$$

Таким образом, в результате преобразования получаем выражение

$$P(t_{\text{ож}} > t_{\text{доп}}) = \frac{\frac{Z^V}{V!}}{1 - \frac{Z}{V}} \cdot P_0 \cdot e^{-(V\mu - \lambda)t}. \tag{7.119}$$

Учитывая, что в формуле (7.119) $\frac{Z^V}{1 - \frac{Z}{V}} \cdot P_0 = P(t_{\text{ож}} > 0)$, выраже-

ние (7.118) можно представить в более удобном для расчета виде:

$$\begin{aligned} P(t_{\text{ож}} > t_{\text{доп}}) &= P(t_{\text{ож}} > 0) \cdot e^{-(V\mu - \lambda)t_{\text{доп}}} = \\ &= P(t_{\text{ож}} > 0) \cdot e^{-\mu(V - Z)t_{\text{доп}}}. \end{aligned} \quad (7.120)$$

Если в качестве времени обслуживания брать условную единицу времени $\bar{t}_{\text{обсл}} = 1$ у.е.в., при этом интенсивность обслуживания $\mu = 1$, то вероятность $P_i(t_{\text{ож}} > t_{\text{доп}})$ будет рассчитываться по формуле

$$P(t_{\text{ож}} > t_{\text{доп}}) = P(t_{\text{ож}} > 0) \cdot e^{-(V - Z)t_{\text{доп}}}. \quad (7.121)$$

Анализ формулы (7.121) показывает, что для систем коммутации, реализующих дисциплину обслуживания с ожиданием, необходимо стремиться к тому, чтобы время ожидания начала обслуживания для подавляющего большинства вызовов не вызывало неудобств у абонентов. Иными словами, системы с ожиданием следует применять в тех случаях, когда для повышения эффективности использования коммутационных устройств целесообразно допускать большие условные потери (исчисляемые десятками процентов), если возможно при этом обеспечить малоощутимое для абонентов время задержки начала обслуживания. Одним из путей решения данной проблемы является предупреждение абонента о допустимом времени разговора и его ограничение, что ведет к возрастанию интенсивности освобождения занятого канала связи, а следовательно, к уменьшению вероятности $P(t_{\text{ож}} > t_{\text{доп}})$.

Ввиду важности данной характеристики качества обслуживания вызовов $P(t_{\text{ож}} > t_{\text{доп}})$ по формуле (7.120) составлены номограммы, которые при фиксированном значении $P(t_{\text{ож}} > t_{\text{доп}}) = 0,1$ и $P(t_{\text{ож}} > t_{\text{доп}}) = 0,2$ позволяют определять любые неизвестные величины: каналный ресурс (V), допустимое время ожидания ($t_{\text{доп}}$), среднее время занятия канала связи $t_{\text{зан}}$, допустимая величина интенсивности поступающей нагрузки Z .

Данные номограммы представлены в [23].

Рассмотрим остальные (второстепенные) характеристики качества обслуживания вызовов системой с ожиданием.

Для любого поступающего вызова средняя длительность ожидания начала обслуживания $\bar{t}_{\text{ож}}$ может рассматриваться как отношение суммарной продолжительности ожидания всех вызовов к числу поступивших вызовов (в данном случае $\bar{t}_{\text{ож}}$ рассматривается по отношению ко всем поступающим вызовам).

Для определения $\bar{t}_{\text{ож}}$ необходимо учитывать, что данная величина является случайной величиной (СВ), непрерывно распределенной на интервале времени от 0 до ∞ , поэтому воспользуемся выражением для математического ожидания.

По правилу определения математического ожидания непрерывной случайной величины имеем:

$$\bar{t}_{\text{ож}} = M[t_{\text{ож}}] = \int_0^{\infty} t_{\text{ож}} \cdot W(t_{\text{ож}}) \cdot dt = \frac{P(t_{\text{ож}} > 0)}{V - Z} \text{ при } \mu \neq 1; \quad (7.122)$$

$$\bar{t}_{\text{ож}} = \frac{P(t_{\text{ож}} > 0)}{V - Z} \text{ при } \mu = 1. \quad (7.123)$$

Следующей характеристикой является среднее время ожидания начала обслуживания $\bar{t}_{\text{з.ож}}$.

Данное время относится только к тем вызовам, которые попадали на ожидание (в данном случае время усредняется не по всем поступившим вызовам, а только по тем, которые находились в очереди задержанных в обслуживании:

$$\bar{t}_{\text{з.ож}} = \frac{1}{\mu \cdot (V - Z)} = \frac{t_{\text{обсл}}}{(V - Z)}. \quad (7.124)$$

Сравнивая выражения (7.124) и (7.122), можно заметить, что величина $\bar{t}_{\text{з.ож}}$ будет иметь значение несколько большее, чем $\bar{t}_{\text{ож}}$, что не противоречит физическому смыслу.

В случае $\mu = 1$ выражение (7.124) примет вид

$$\bar{t}_{\text{з.ож}} = \frac{1}{V - Z}. \quad (7.125)$$

Среднее число вызовов, ожидающих обслуживания, или средняя длина очереди $\bar{r}$ есть также случайная дискретная величина, распределенная на интервале от V до ∞ , поэтому для ее определения используем математическое ожидание дискретной СВ. В этом случае величину $\bar{r}$ можно определить, используя выражение для математического ожидания дискретной СВ:

$$\bar{r} = \sum_{i=V}^{\infty} (i - V) \cdot P_i = P(t_{\text{ож}} > 0) \cdot \frac{Z}{V - Z}, \quad (7.126)$$

где P_i определяется из выражения (7.109).

Из выражений (7.122 и 7.126) следует, что средняя длина очереди определяется как среднее время ожидания начала обслуживания вызова, отнесенное ко всем поступающим вызовам и умноженное на интенсивность поступающей нагрузки. С учетом формулы (7.124) выражение (7.126) можно представить следующим образом:

$$\bar{r} = \bar{t}_{\text{ож}} \cdot Z. \quad (7.127)$$

7.2.4. Модель системы с ожиданием $M_i/M/V/W, r/FF/R$

7.2.4.1. Постановка задачи по обслуживанию потока вызовов системой с ожиданием $M_i/M/V/W, r/FF/R$

По аналогии с предыдущей моделью системы с ожиданием, обслуживающей простейший поток вызовов, для модели системы распределения информации с ограниченным числом источников нагрузки следует рассмотреть область ее применения с учетом основных ограничений и допущений. Для данной модели характерно следующее.

На входы коммутационной системы поступает примитивный поток вызовов ($N < 100$ или $N < 15V$). Следовательно, в первом разряде символики Кендалла – Башарина будет стоять буква M_i . Особенностью модели примитивного потока является то, что параметр потока вызовов изменяется скачкообразно и зависит от состояния системы, т. е. от числа занятых выходов системы или источников нагрузки. Параметр потока определяется по ранее приведенной формуле $\lambda_i = \alpha \cdot (N - i)$.

Длительность занятия линии при обслуживании вызова $t_3 = t_{\text{обсл}}$ является случайной величиной, распределенной по показательному закону $F(t) = P(t_3 < t) = 1 - e^{-\mu t}$, где $\mu = 1/\bar{t}_3$ – параметр показательного закона распределения потока освобождения. Экспоненциальное распределение является обязательным условием данной модели. Во втором разряде символики будет стоять буква M .

Коммутационное поле (КП) системы обслуживания однозвенное, неблокируемое. В выходы КП включен полнодоступный пучок линий емкостью V . Следовательно, вероятностные характеристики качества обслуживания будут зависеть только от вероятности занятости выходов системы. В третьем разряде символики будет стоять символ V .

В четвертом разряде символики Кендалла – Башарина будет стоять символ W (*Wait*).

Вызовы из очереди обслуживаются в порядке поступления. Для описания данного порядка обслуживания в пятом разряде будут стоять символы FF (*first-in-First-out* – упорядоченный, т. е. первым пришел – первым ушел).

Порядок занятия свободных линий – случайный, следовательно, в шестом разряде используется символ R (*Random*).

Таким образом, в обозначении Кендалла – Башарина данная модель будет записана в виде $M_i/M/V/W, r/FF/R$.

Требуется определить характеристики качества обслуживания вызовов полнодоступной однозвенной системой с ожиданием и ограниченным числом источников нагрузки.

7.2.4.2. Характеристики качества обслуживания вызовов системой $M_i/M/V/W, r/FF/R$

При выводе основных выражений для указанной модели используются те же подходы, которые были применены для случая обслуживания системой с ожиданием простейшего потока телефонных вызовов. Однако получение требуемых характеристик качества обслуживания существенно усложняется по многим причинам.

Во-первых, следует учитывать влияние состояния системы обслуживания на параметр примитивного потока, который в данном случае будет определяться по формуле $\lambda_i = \alpha \cdot (N - i)$.

Во-вторых, авторы, разработавшие данную математическую модель, стремились для простоты вычисления характеристик качества обслуживания привести математические выражения к широко известной первой формуле Эрланга $E_V(Z)$, по которой были составлены таблицы и номограммы.

Поэтому, не загружая дополнительно материал математикой, в данном учебном пособии авторы приводят лишь конечные выражения, позволяющие рассчитывать вероятностно-временные характеристики качества обслуживания требований систем коммутации с ожиданием и ограниченным числом источников нагрузки.

Для расчета вероятности наличия очереди $P(t_{\text{ож}} > 0)$ при условии, что время занятия обслуживающих приборов $t_3 \neq 1$ условной единице времени (у. е. в.), а время свободности источника нагрузки $\bar{t}_{\text{св}} = \frac{1}{\alpha}$, следует использовать формулу

$$P(t_{\text{ож}} > 0) = \frac{C_{N-1}^V}{C_{N-1}^V + E_{N-V-1}\left(\frac{V \cdot t_{\text{св}}}{t_3}\right) \cdot \sum_{i=0}^{V-1} C_{N-1}^i \cdot \frac{(t_{\text{св}}/t_3)^V}{(t_{\text{св}}/t_3)^i}}. \quad (7.128)$$

В случае если $t_3 = 1$ у. е. в. выражение (7.128) записывается в виде

$$P(t_{\text{ож}} > 0) = \frac{C_{N-1}^V \cdot \alpha^V}{\left(C_{N-1}^V \cdot \alpha^V + E_{N-V-1}\left(\frac{V}{\alpha}\right) \cdot \sum_{i=0}^{V-1} C_{N-1}^i \cdot \alpha^i \right)}, \quad (7.129)$$

где C_{N-1}^V – число сочетаний из $(N-1)$ по V ;

$\alpha = \frac{1}{t_{\text{св}}}$ – интенсивность источника в свободном состоянии;

$E_{N-V-1}\left(\frac{V}{\alpha}\right)$ – условная запись Эрланговых потерь (первая формула Эрланга).

Вероятность ожидания свыше допустимого времени $P(t_{\text{ож}} > t_{\text{доп}})$ при $t_3 \neq 1$ у. е. в. вычисляется по формуле

$$\begin{aligned}
& P(t_{\text{ож}} > t_{\text{доп}}) = \\
& = P(t_{\text{ож}} > 0) \cdot \frac{\left(\frac{t_{\text{св}} + t_{\text{доп}}}{t_3} \right)^{N-V-1} \cdot E_{N-V-1}(V \cdot t_{\text{св}} / t_3) \cdot e^{-\frac{V t_{\text{доп}}}{t_3}}}{E_{N-V-1}\left(\frac{V \cdot (t_{\text{св}} + t_{\text{доп}})}{t_3} \right)} \cdot e^{-\frac{V t_{\text{доп}}}{t_3}}, \quad (7.130)
\end{aligned}$$

где $t_{\text{доп}}$ – допустимое время ожидания.

В случае $t_3 = 1$ у. е. в. выражение (7.130) приводится к виду

$$\begin{aligned}
& P(t_{\text{ож}} > t_{\text{доп}}) = \\
& = P(t_{\text{ож}} > 0) \cdot \frac{E_{N-V-1}\left(\frac{V}{\alpha} \right)}{E_{N-V-1}\left(\frac{V}{\alpha} + V \cdot t_{\text{доп}} \right)} \cdot (1 + \alpha \cdot t_{\text{доп}})^{N-V-1} \cdot e^{-V t_{\text{доп}}}. \quad (7.131)
\end{aligned}$$

Средняя длительность ожидания начала обслуживания для ждущих в очереди требований ($\bar{t}_{\text{жд. ож}}$) вычисляется по формуле

$$\bar{t}_{\text{жд. ож}} = \frac{(N-V) \cdot t_3}{V} - t_{\text{св}} \cdot \left[1 - E_{N-V-1}\left(V \cdot \frac{t_{\text{св}}}{t_3} \right) \right]. \quad (7.132)$$

Средняя продолжительность ожидания относительно всех поступающих требований определяется по формуле

$$\bar{t}_{\text{ож}} = P(t_{\text{ож}} > 0) \cdot \bar{t}_{\text{жд. ож}}. \quad (7.133)$$

При решении практических задач по определению характеристик качества обслуживания вызовов системой с ожиданием и ограниченным числом источников нагрузки следует иметь в виду, что расчет вероятностных характеристик качества обслуживания имеет особенности. Методика расчета включает итерационную схему вычисления.

Вначале определяют среднюю продолжительность свободы источника $t_{\text{св}}$, пренебрегая величиной среднего времени ожидания $\bar{t}_{\text{ож}}$ (так как она неизвестна), используя выражение

$$t_{\text{св}} = t_3 \cdot \frac{(N - Z)}{Z} - \bar{t}_{\text{ож}}. \quad (7.134)$$

Затем по формуле (7.128) вычисляется вероятность наличия очереди $P(t_{\text{ож}} > 0)$.

Далее вычисляется средняя продолжительность ожидания для ждущих вызовов по формуле

$$\bar{t}_{\text{жд. ож}} = \frac{(N - V) \cdot t_3}{V} - t_{\text{св}} \cdot \left[1 - E_{N-V-1} \left(V \cdot \frac{t_{\text{св}}}{t_3} \right) \right]. \quad (7.135)$$

Определив $P(t_{\text{ож}} > 0)$ и $\bar{t}_{\text{жд. ож}}$, рассчитывают значение средней продолжительности ожидания относительно всех поступающих в систему вызовов по формуле $\bar{t}_{\text{ож}} = P(t_{\text{ож}} > 0) \cdot \bar{t}_{\text{жд. ож}}$.

Далее переходят к итерационной схеме расчета.

Для этого:

- рассчитывают интенсивность входящей нагрузки по найденному значению $\bar{t}_{\text{ож}}$:

$$Z = \frac{N \cdot t_3}{t_{\text{св}} + t_3 + \bar{t}_{\text{ож}}}; \quad (7.136)$$

- вычисляют среднюю продолжительность свободности источника $t_{\text{св}}$ при известной общей поступающей нагрузке Z .

Если полученная в результате вычисления величина $t_{\text{св}}$ мало отличается от первоначального своего значения, от дальнейшего расчета характеристик можно отказаться, так как это уже не повлияет на полученные результаты.

7.3. Основы теории построения и расчета потерь в многозвенных системах с явными потерями

Рассмотренные выше методы расчета однозвенных систем распределения информации базировались на том, что причины потерь вызовов обуславливались занятостью выходов схемы.

В случае рассмотрения многозвенных схем потери вызовов обуславливаются не только занятостью всех V выходов схемы. Они могут иметь место даже тогда, когда еще существуют свободные выхо-

ды ($i < V$), но промежуточные линии между звеньями, ведущие к свободным выходам, являются занятыми и недоступными для обслуживания данного соединения. Кроме того, в многозвенных схемах значение вероятности потерь зависит от способа включения промежуточных линий (ПЛ), режима искания и способа включения выходов в коммутаторы последнего звена.

Следует остановиться на исторических аспектах появления методов расчета многозвенных (звеньевых) систем распределения информации.

Теория расчета потерь в звеньевых схемах впервые была разработана шведским ученым К. Якобеусом в начале 50-х гг. XX в.

В основу его метода положена формула, которую К. Пальм в 1943 г. предложил для определения вероятности занятости точно i конкретных линий в однозвенном полнодоступном пучке линий. В 1947 г. К. Якобеус применил эту формулу для определения вероятности потерь в двухзвенных коммутационных полях.

На основе выведенных им теоретических соотношений шведская фирма "Эрикссон" составила рекомендации по расчету пропускной способности различного типа звеньевых схем, применяемых в координатных АТС того времени.

С усложнением оборудования АТС развивалась и теория расчета многозвенных систем распределения информации. В частности, в 1957 г. советским ученым А. Д. Харкевичем был предложен метод эффективной доступности для расчета потерь в двухзвенных схемах в режиме группового искания (ГИ). В 60-е гг. XX в. под руководством Бориса Самойловича Лившица в Ленинградском отделении научно-исследовательского института связи был разработан так называемый метод ЛОНИИС.

Следует также отметить, что ввиду сложности строго математического описания процесса обслуживания вызовов в многозвенных системах коммутации точные формулы для расчета пропускной способности и потерь, пригодные для инженерной практики, удастся получить лишь для простейших случаев, поэтому рассматриваемые ниже методы основываются на допущениях, ограничениях и приближениях, которые следует учитывать при оценке функционирования рассматриваемых систем.

7.3.1. Особенности построения двухзвенных и многозвенных коммутационных схем и влияние их структуры на характеристики качества обслуживания

Понимание сущности методов расчета двух- и многозвенных коммутационных схем необходимо начать с рассмотрения особенностей их построения, причин, обуславливающих потери вызовов в многозвенных системах распределения информации.

В рамках дисциплины "Системы коммутации" было установлено, что использование многозвенных схем позволяет получить значительную экономию в объеме коммутационного оборудования, особенно при создании АТС средней и большой емкости.

Однако такая экономия привела к возрастанию потерь в коммутационных полях многозвенных систем распределения информации за счет влияния внутренних блокировок, чего в однозвенных схемах не наблюдалось.

Учет влияния внутренних блокировок на процесс обслуживания вызовов в многозвенных системах потребовал разработки нового математического аппарата, так как использование существующих методов расчета однозвенных схем приводило к существенным ошибкам при оценке функционирования многозвенных систем коммутации.

Как известно, в звеньевых схемах общее число входов и выходов КП делится на соответствующие группы. Для каждой группы входов создается своя коммутационная матрица, называемая коммутатором. Аналогично образуются коммутаторы для включения всех групп выходов. Совокупность коммутаторов одного типа, обеспечивающих включение входов КП или его выходов, образует звено.

Каждый коммутатор одного звена (A) соединяется с коммутатором другого звена (B) определенным количеством промежуточных линий (рис. 7.17).

Методы расчета многозвенных схем следует рассматривать в режиме группового искания. Понятие "искание" исходит от коммутаторов Алмана Строуджера (первый автоматический коммутатор он изобрел в 1891 г.).

С тех пор по настоящее время под исканием понимают процесс выбора требуемой линии.

К настоящему времени известны три режима искания: линейный, свободный, групповой. Именно введение группового режима

искания, предложенного М. Ф. Фрейденбергом в 1981 г., сделало возможным неограниченное расширение ёмкости современных телефонных станций.

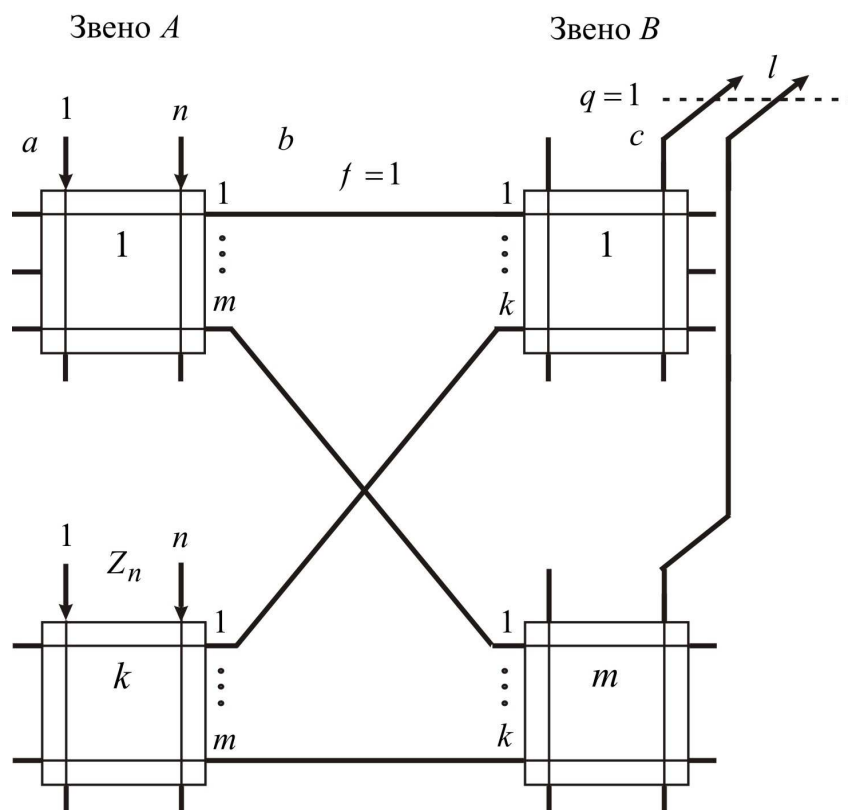


Рис. 7.17. Односвязная двухзвенная схема

Как известно, для описания звеньевых коммутационных схем используют коммутационные параметры. В частности, для двухзвенных схем такими параметрами являются:

k — число коммутаторов первого звена;

n — количество входов в каждый коммутатор звена A ;

m — число коммутаторов звена B ;

q — количество выходов из каждого коммутатора последнего звена в направлении искания;

l — число направлений искания;

f — связность схемы, т. е. количество промежуточных линий между каждым коммутатором звена A и каждым коммутатором звена B .

Кроме приведенных основных коммутационных параметров двухзвенной схемы можно получить производные параметры:

mf — число выходов из каждого коммутатора звена A ;

kf – количество входов в каждый коммутатор звена B ;

ml – число выходов из каждого коммутатора звена B ;

$N = nk$ – общее количество входов в звено A ;

$V = mql$ – общее число выходов в звене B .

Для выяснения сути потерь в двухзвенных схемах рассмотрим самую элементарную для понимания происходящих в ней процессов, так называемую простую двухзвенную схему. Простой односвязной схемой называется схема, в которой связность f и число выходов q из одного коммутатора второго звена (звено B) в направлении иска-ния равны единице, т. е. $f = q = 1$ (см. рис. 7.17).

В данной схеме причины в отказах в установлении соединения могут наступать по следующим причинам:

- занятости всех V выходов требуемого направления иска-ния;
- занятости всех промежуточных линий (ПЛ) между звеньями;
- занятости тех ПЛ, которые ведут к свободным выходам тре-буемого направления иска-ния.

Такую комбинацию причин рассматривал К. Якобеус при иссле-довании многозвенных схем, что и обусловило второе название его метода – комбинаторный метод расчета.

Очевидно, что внутренние блокировки, увеличивая число отка-зов в соединении, снижают пропускную способность коммутацион-ной схемы и увеличивают потери. При этом чем больше нагрузка по-ступает на двухзвенную схему, тем больше вероятность внутренних блокировок и потерь вызовов.

Кроме внутренних блокировок, на вероятность потерь в много-звенных схемах оказывают влияние следующие факторы:

- режим иска-ния выходов (свободный, групповой, линейный);
- способы включения промежуточных линий (способ меж-звеньевых соединений);
- методы иска-ния в многозвенных схемах.

Если рассматривать простую двухзвенную схему в режиме ли-нейного иска-ния (ЛИ) (рис.7.18), когда заданному входу требуется конкретный выход, то в этом случае для соединения можно исполь-зовать всего одну ПЛ. Очевидно, что даже в случае занятости только одной этой линии одним из абонентов (одним входом из n), вклю-ченным в первый коммутатор и занимающим любой другой выход

в коммутаторе звена B , возникнут потери в обслуживании других вызовов, поступающих на оставшиеся входы первого коммутатора звена A .

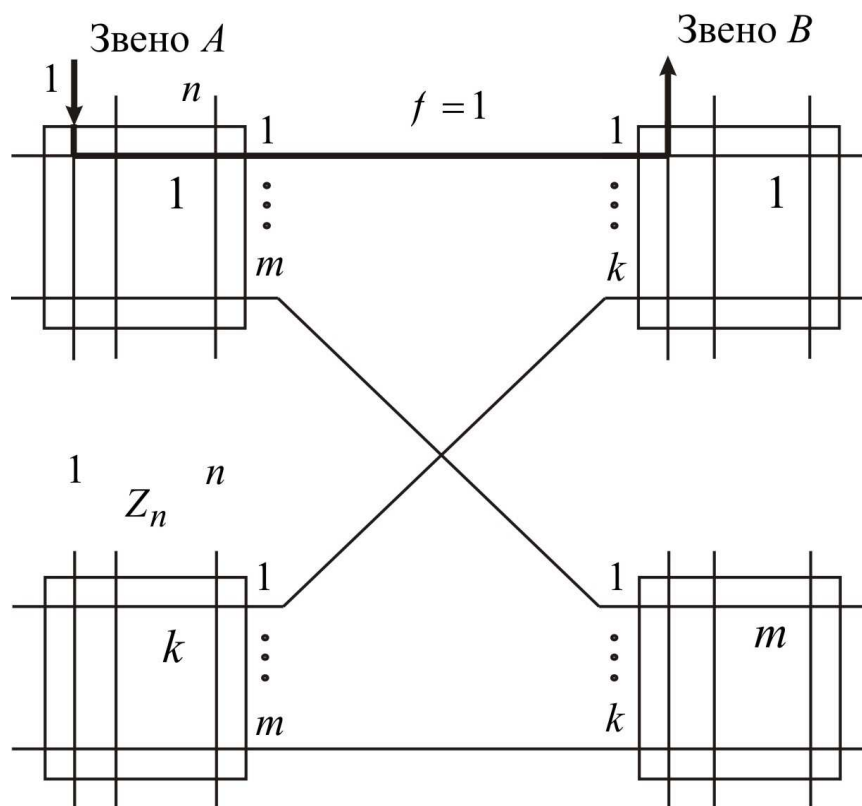


Рис. 7.18. Режим линейного искания

В режиме свободного искания (как правило, он используется в ступени регистрового искания для выбора любого свободного регистра) для установления соединения могут использоваться любые m ПЛ, доступных данному входу (рис. 7.19). Если вызов поступил на первый вход, то через ПЛ № 1 возможно соединение с одним из выходов 1-го коммутатора звена B , через ПЛ № 2 – с одним из выходов 2-го коммутатора звена B и т. д.

Интуитивно можно предположить, что в этом случае потери будут существенно меньше, чем в режиме ЛИ, так как отказ в обслуживании возможен лишь тогда, когда все ПЛ будут заняты. Следовательно, вероятность этого события значительно меньше, чем в предыдущем случае.

Рассматривая режим группового искания, следует обратить внимание на влияние способа разбиения выходов по направлениям искания на вероятность потерь.

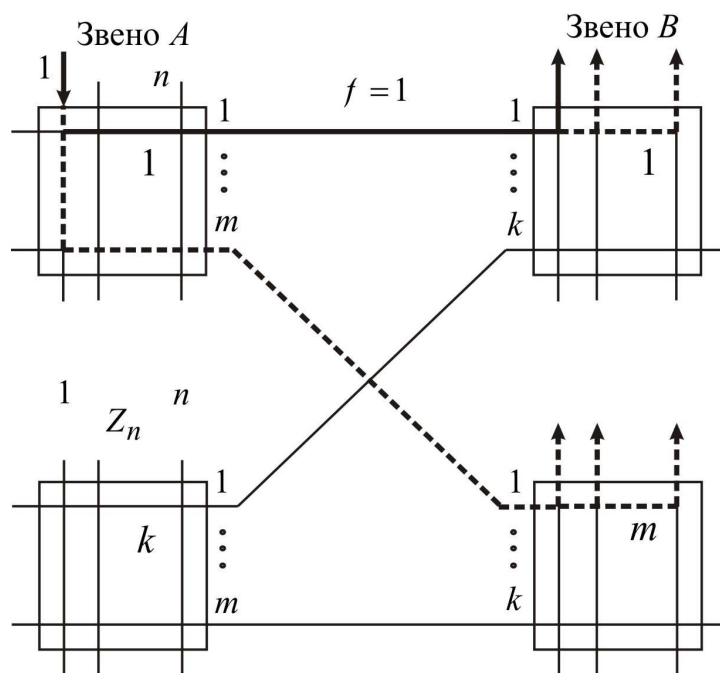


Рис. 7.19. Режим свободного искания

Предположим, что для некоторого направления в режиме группового искания выделяется группа из l линий. Если все эти l линий подключить к одному коммутатору звена B , то в процессе искания можно использовать всего одну из m ПЛ, соединяющую коммутатор звена A с коммутатором звена B (рис.7.20). Следовательно, наблюдается рост вероятности потерь.

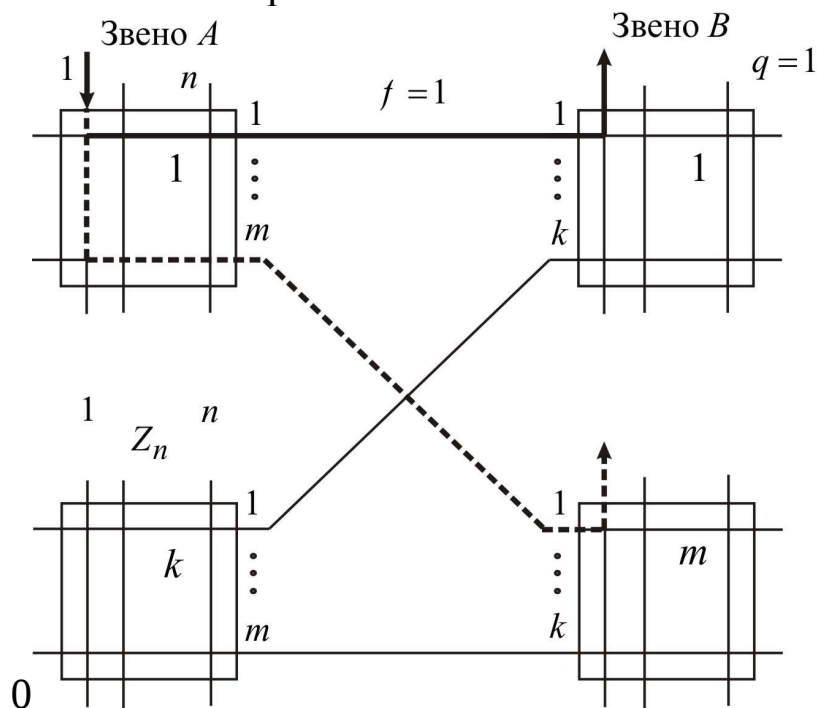


Рис. 7.20. Режим группового искания с выделением $q=1$ выходов из коммутаторов звена B для каждого направления

Если для организации направления выделять q выходов из каждого коммутатора звена B , то в этом случае, как и в режиме свободного искания, для установления соединения в данном направлении могут быть использованы все m промежуточных линий, доступных входу. Интуитивно можно заключить, что последний способ обеспечивает лучшее качество работы схемы, чем предыдущий.

Кроме режима искания, вероятность потерь в звеньевых блокируемых коммутационных схемах зависит от размеров коммутаторов, наличия или отсутствия расширения, сжатия (концентрации) в каждом звене, способа межзвеньевых соединений и метода искания свободных промежуточных линий.

В звеньевых коммутационных схемах наиболее часто используются два способа межзвеньевых соединений, показанных на рис. 7.21 и 7.22.

На рис. 7.21 изображен способ последовательных межзвеньевых соединений, при котором номера занимаемых входов в коммутаторах звена B соответствуют номеру коммутатора в звене A .

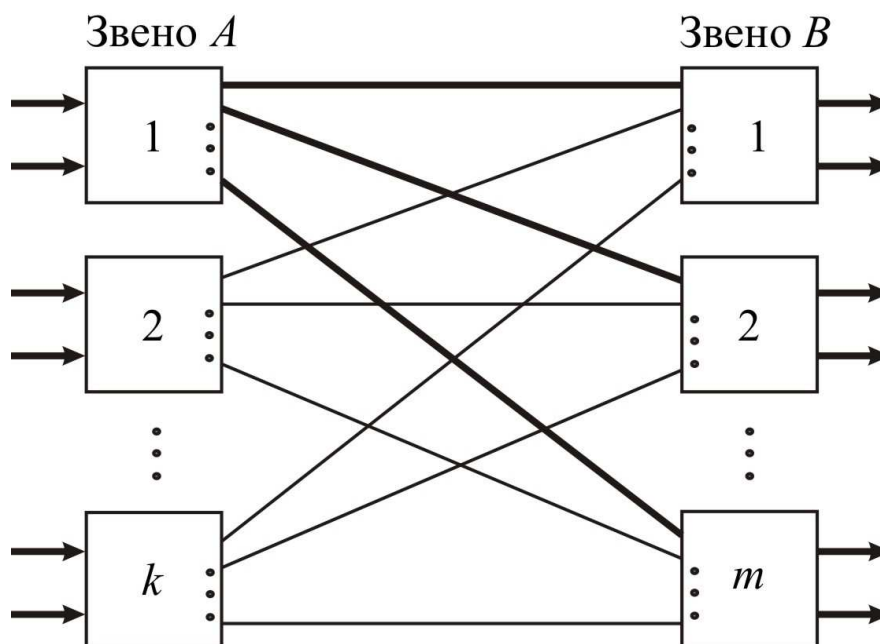


Рис. 7.21. Схема последовательных межзвеньевых соединений

На рис. 7.22 приведен способ циклических межзвеньевых соединений. При этом способе осуществляется циклический сдвиг номера занимаемого входа в коммутаторах звена B при переходе от одного выхода из коммутатора звена A к другому.

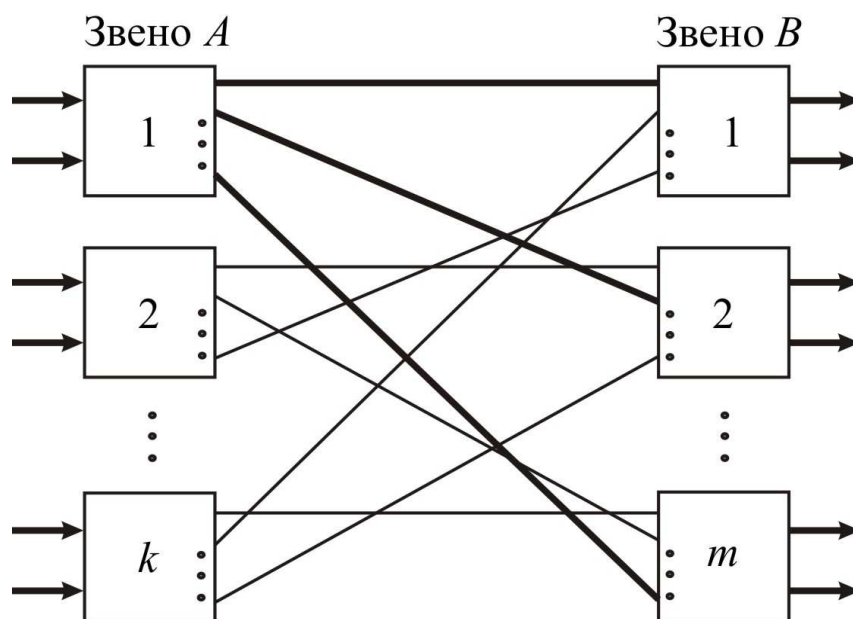


Рис. 7.22. Схема циклических межзвеньевых соединений

Методами искания, наиболее часто применяемыми в коммутационной технике, считаются:

- последовательное искание из фиксированного исходного положения;
- последовательное искание из случайного положения (равновероятный выбор любой свободной промежуточной линии).

Каждый из указанных методов искания может независимо от другого использоваться между любой парой звеньев как внутри блоков, так и на междублочных ПЛ.

Методы искания оказывают влияние на величину обслуженной нагрузки линии. Действительно, при первом методе наибольшую нагрузку будет обслуживать первая линия.

При использовании второго метода занятость всех линий будет равновероятной (одинаковой).

Помимо двухзвенных коммутационных схем в коммутационной технике широко применяются многозвенные схемы. Как правило, в квазиэлектронных АТС средней и большой емкости соединение может проходить через три, четыре, пять и более звеньев. При рассмотрении данного вопроса, не вдаваясь в частности, следует обратить внимание на важные с позиции теории телетрафика особенности данных схем. Одной из них является то, что способы построения многозвенных схем оказывают различное влияние на вероятность потерь в них.

В настоящее время по способам построения различают веерные и связанные многозвенные схемы.

В веерной схеме между любыми из входов и любым коммутатором последнего звена имеется не более одного соединительного пути, что обуславливает значительные потери вызовов. Естественно, такую многозвенную схему не рекомендуется использовать на ступени ЛИ, так как наличие одного возможного пути установления соединения ведет к большим значениям вероятности потерь.

В связанных схемах между заданным входом и требуемым выходом существует более одного пути установления соединения.

По этой причине связанная трехзвенная схема может быть использована в любом режиме искания.

При рассмотрении многозвенных схем с позиции вероятности потерь следует обратить внимание на такие понятия, как концентрация и расширение (рис. 7.23).

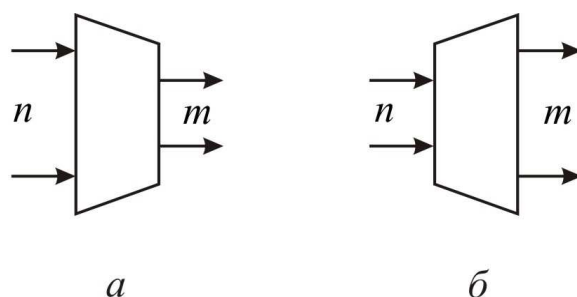


Рис. 7.23. Многозвенная схема с позиции вероятности потерь:
а – с концентрацией; б – с расширением.

Для характеристики этих понятий вводят отношение числа выходов (m) к числу входов (n) на каждом звене. Если это отношение больше единицы ($\frac{m}{n} > 1$), то оно называется коэффициентом расширения (схема с расширением), а схемы – схемами с расширением. Если отношение $\frac{m}{n} < 1$, то оно называется коэффициентом концентрации, а схемы, соответственно, с концентрацией.

7.3.2. Определение потерь в двухзвенных схемах.

Комбинаторный метод расчета (метод Якобеуса)

Данный метод был разработан К. Якобеусом для режима группового искания двухзвенной схемы. Особенностью данного метода

является то, что он включает два вида искания: выбор требуемого направления (вынужденное линейное искание) и выбор любого свободного выхода в данном направлении (свободное искание). Требуемое направление определяется по индексу набранного номера, а выбор свободного выхода может осуществляться либо упорядоченно, либо случайно (равновероятно). В данном методе предполагается случайное занятие линий.

Сущность метода рассмотрим на примере односвязной двухзвенной схемы, приведенной на рис. 7.24.

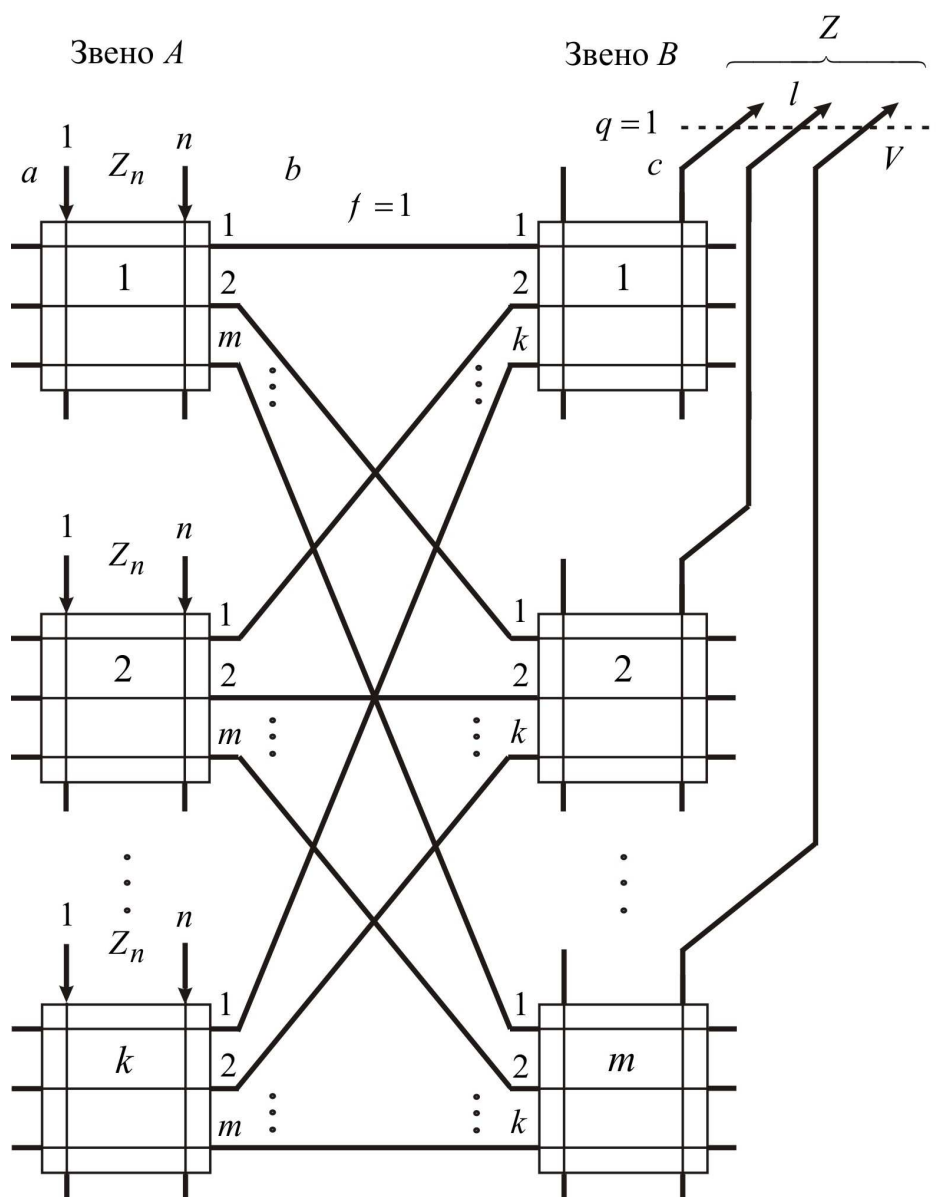


Рис. 7.24. Простая двухзвенная схема

Данная модель в символике Кендалла – Башарина имеет следующий вид:

$M/M/LS/L/R$ или $M_i/M/LS/L/R$.

Рассмотрим подробнее допущения и ограничения, принятые в этой модели.

1. На двухзвенную схему поступает простейший или примитивный поток вызовов $[M(M_i)]$.

2. Длительность обслуживания является случайной величиной (G) , имеющей любое распределение вероятностей.

3. Схема системы распределения информации – двухзвенная (LS) , поэтому при определении потерь следует учитывать их причины, обусловленные занятием выходов схемы, промежуточных линий между звеньями (внутренние блокировки) и неудачными комбинациями.

4. Способ обслуживания с потерями (L) .

5. Занятие выходов и промежуточных линий осуществляется случайным образом (R) .

Введем следующие обозначения. Обозначим через Z_n (Эрл) – интенсивность поступающей нагрузки на n входов одного коммутатора звена A . Интенсивность поступающей нагрузки на m выходов в направлении связи, подключенных к коммутаторам звена B , обозначим символом Z (Эрл). Интенсивность нагрузок: удельной поступающей на вход коммутатора звена A обозначим символом a (Эрл), удельной обслуженной на выходы звена A (промежуточными линиями) – b (Эрл), обслуженной на выходы двухзвенной схемы в направлении связи – c (Эрл).

Требуется определить потери в простой односвязной двухзвенной схеме.

Для решения данной задачи К. Якобеус использовал следующие допущения:

– независимость событий, описываемых вероятностями занятости конкретных линий на звене A от занятости любых i линий на звене B (предположение является условным, так как выходы и ПЛ занимают парами);

– случайное (равновероятное) занятие ПЛ и выходов. При этом вероятности занятия любой ПЛ из m считаются равными между собой и не зависят от того, какие i из m линий заняты. Аналогичное допущение распространяется и на выходы простой двухзвенной схемы.

Предположим, что к рассматриваемому моменту времени вызов поступил на первый вход первого коммутатора звена A коммутационной схемы.

Установление соединения между определенным входом и одним из выходов рассматриваемого направления заключается в использовании одной из свободных промежуточных линий и одного из свободных выходов требуемого направления, взаимно доступных друг другу.

Для обслуживания поступившего вызова в рассматриваемом случае могут быть использованы m промежуточных линий и m выходов требуемого направления.

Соединение может быть установлено, если имеется пара свободных и взаимно доступных звеньев. Если такой пары нет, то наступают потери соединений.

Для вычисления вероятности потерь в данной схеме необходимо рассмотреть все возможные причины, обуславливающие потерю вызова. Так, для рассматриваемого вызова, поступившего на первый вход первого коммутатора звена A , потери будут обусловлены тремя причинами:

- занятостью всех промежуточных линий, которые могут быть использованы для обслуживания поступившего вызова (в нашем примере для простой двухзвенной схемы $n = m$, следовательно, вероятность этого события равна нулю);
- занятостью всех выходов, подключенных к коммутаторам звена B требуемого направления связи;
- неудачными комбинациями свободных ПЛ и свободных выходов, т. е. при наличии в схеме $(m - i)$ свободных выходов на звене B окажутся занятыми конкретные $(m - i)$ ПЛ, подключенные к выходам первого коммутатора звена A и ведущие к этим свободным выходам звена B .

Если предположить, что рассматриваемый вызов поступил на отмеченный вход первого коммутатора в момент, когда i промежуточных линий из m , подключенных к выходам данного коммутатора, заняты, то для подключения входа к одному из выходов требуемого направления связи могут быть использованы только оставшиеся $(m - i)$ ПЛ.

Если же выходы данного направления, соответствующие этим свободным $(m - i)$ ПЛ, заняты, то наступают потери. Это утверждение справедливо для любого i , лежащего в пределах $0 \leq i \leq m$, и охватывает два случая занятости: всех промежуточных линий $i = m$ и всех выходов в направлении ($i = 0$).

Если вероятность занятия любых i из m ПЛ, принадлежащих одному коммутатору первого звена, обозначить через $W_A(i)$, а вероятность занятия определенных $(m-i)$ выходов на звене B – через условную вероятность $H_B[(m-i)/i]$, то, используя теорему умножения вероятностей совместного появления двух событий, можно записать следующую формулу для определения вероятности потерь:

$$P = \sum_{i=0}^m W_A(i) \cdot H_B[(m-i)/i]. \quad (7.137)$$

По аналогии можно записать, что

$$P = \sum_{i=0}^m W_A[(m-i)/i] \cdot H_B(i). \quad (7.138)$$

Используя приведенные в постановке задачи допущения, касающиеся независимости событий, описываемых вероятностями занятости $(m-i)$ конкретных линий на одном звене от занятости любых i линий на другом звене и случайного (равновероятного) занятия ПЛ и выходов из схемы, можно преобразовать уравнения (7.137) и (7.138):

$$P = \sum_{i=0}^m W_A(i) \cdot H_B(m-i). \quad (7.139)$$

Для вычисления вероятности потерь P в формуле (7.139) необходимо знать значения вероятностей занятости i любых линий на звене A – $W_A(i)$ и занятости конкретных $(m-i)$ линий на звене B – $H_B(m-i)$ (функции распределения вероятностей занятия ПЛ и выходов).

Для определения этих вероятностей ($W_A(i)$ и $H_B(m-i)$) К. Якобеусом было предложено использовать два известных в то время распределения случайной величины:

- Эрланга;
- Бернулли.

Условием применения распределения Эрланга на любом из звеньев схемы является неограниченное число источников нагрузки, создающих вызовы на входы одного коммутатора звена A ($n > 100$, $N \geq V$) и на входы коммутатора звена B данной двухзвенной схемы.

В этом случае для определения вероятности занятости выходов коммутатора звена A следует использовать первое распределение Эрланга:

$$W_A(i) = E_{i, m}(Z) = \frac{\frac{Z_n^i}{i!}}{\sum_{j=0}^m \frac{Z_n^j}{j!}}. \quad (7.140)$$

Вероятность занятости i любых соединительных устройств в пучке из m таких устройств на выходе коммутаторов звена B при интенсивности Z (Эрл) на пучок принимается равной

$$H_B(i) = E_{i, m}(Z) = \frac{\frac{Z^i}{i!}}{\sum_{j=0}^m \frac{Z^j}{j!}}. \quad (7.141)$$

Для случая нахождения вероятности занятости $(m-i)$ фиксированных выходов в пучке из m для коммутатора звена A следует использовать выражение

$$W_A(i) = \frac{E_m(Z_n)}{E_{m-(m-i)}(Z_n)} = \frac{E_m(Z_n)}{E_i(Z_n)}. \quad (7.142)$$

Вероятность занятия $(m-i)$ фиксированных выходов в пучке из m для звена B будет определяться по формуле

$$H_B(m-i, m) = H_B(m-i) = \frac{E_m(Z)}{E_i(Z)}, \quad (7.143)$$

где $E_m(Z)$ – условная запись эрланговых потерь в полнодоступном пучке из m соединительных устройств при интенсивности поступающей нагрузки Z (Z_n) (Эрл) на данный пучок, которые вычисляются на примере второго звена по первой формуле Эрланга

$$E_m(Z) = \frac{\frac{Z^m}{m!}}{\sum_{j=0}^m \frac{Z^j}{j!}}. \quad (7.144)$$

$E_i(Z)$ – формула, позволяющая определить в пучке из m линий вероятность занятости i любых линий:

$$E_i(Z) = \frac{\frac{Z^i}{i!}}{\sum_{j=0}^m \frac{Z^j}{j!}}. \quad (7.145)$$

Условием применения распределения Бернулли является ограниченное число источников нагрузки, включенных во входы коммутатора звена A ($n < 100$), не превышающее число соединительных устройств, и ограниченное число входов в коммутаторы звена B .

Например если имеется m линий, каждая из которых в любой момент времени независимо от других занята с вероятностью p и свободна с вероятностью $(1-p)$, тогда вероятность того, что i линий будут заняты, равна p^i , а вероятность того, что остальные линии $(m-i)$ будут свободны – $(1-p)^{m-i}$. Всего i определенных линий из имеющихся m можно выбрать C_m^i различными способами. Если учесть все возможные комбинации, при которых точно i линий из m заняты, а $(m-i)$ линий будут свободны, то получаем известную формулу Бернулли

$$W_A(i) = C_m^i \cdot p^i \cdot (1-p)^{m-i}. \quad (7.146)$$

Так как вероятность занятия линии равна величине интенсивности обслуженной этой линией нагрузки ($p=a$, $p=b$, $p=c$ или $p=y$), то для звена A формулу (7.146) можно записать в следующем виде:

$$W_A(i) = C_m^i \cdot b^i \cdot (1-b)^{m-i}. \quad (7.147)$$

Для выходов звена B вероятность занятия любых линий будет определяться по формуле

$$H_B(i) = C_m^i \cdot c^i \cdot (1-c)^{m-i}. \quad (7.148)$$

В случае определения вероятности занятости конкретных линий $(m-i)$ на звене A или B при использовании распределения Бернулли следует применять следующие выражения:

- $W_A(m-i) = b^{m-i}$ для звена A ;
- $H_B(m-i) = c^{m-i}$ для звена B .

Подводя итоги, следует еще раз подчеркнуть, что сущность метода Якобеуса заключается в том, что причиной потерь вызовов в двухзвенной схеме является занятость линий. При этом занятыми могут быть как линии на выходе коммутатора звена A , так и выходы коммутаторов звена B . Кроме того, в данной схеме могут возникнуть неудачные комбинации, когда свободные выходы звена A ведут к занятым выходам звена B и наоборот.

Используя допущения о независимости занятости линий на звеньях, К. Якобеус предложил выражение $P = \sum_{i=0}^m W_A(i) \cdot H_B(m-i)$,

учитывающее все возможные комбинации занятости выходов на звене A и B , при которых поступивший вызов теряется. Для вычисления ее сомножителей он решил использовать известные к моменту разработки комбинаторного метода распределения Эрланга и Бернулли.

Отсутствие сжатия и расширения в двухзвенной схеме означает, что коммутаторы звена A имеют следующие параметры: $n = m$. Для упрощения получения необходимых уравнений для вероятности потерь предположим, что $f = 1$ и $q = 1$.

Распределение Бернулли (Б) на звеньях A и B используют в следующих случаях:

- ограниченное число входов в один коммутатор на звене A ($n < 100$, $n \approx m$);
- ограниченное число коммутаторов (k) на звене A , что обуславливает ограниченное число входов в коммутаторы звена B ;
- ограниченное число входов в коммутаторы звена B .

Условие отсутствие сжатия (концентрации) и расширения предполагает равенство нагрузок, исполняемых одним входом и выходом коммутатора звена A ($a = b$), что вытекает из равенства числа входов и выходов каждого коммутатора первого звена ($n = m$).

Следует отметить, что распределение Бернулли на практике применяют достаточно часто по причине того, что в известных АТС коммутаторы обычно имеют ограниченное число входов и выходов.

Причины, повлекшие за собой потери в данной схеме, следующие:

- занятость всех выходов в требуемом направлении связи на звене B ;
- занятость всех m ПЛ, выходящих из одного (любого) коммутатора звена A ;
- неудачные комбинации, обусловленные тем, что свободные ПЛ, подключенные к выходам коммутатора звена A , ведут к занятым выходам коммутаторов звена B , к которым подключены каналы направления связи.

При применении распределения Бернулли используют значения удельных нагрузок на линии – a , b , c , где a – нагрузка, обслуженная одним входом коммутатора звена A , b – одним выходом коммутатора звена A (или ПЛ), c – одним выходом (или каналом) коммутатора звена B .

В этом случае, относя вероятность занятости i любых промежуточных линий к выходам коммутаторов звена A ($W_A(i)$), а вероятность занятости конкретных $(m-i)$ линий – к выходам звена B ($H_B(m-i)$), вероятность потерь вызовов запишем в виде

$$P = \sum_{i=0}^m W_A(i) \cdot H_B(m-i). \quad (7.149)$$

В выражении (7.149) вероятность занятости любых i линий на выходе коммутатора звена A – $W_A(i)$ определяется по формуле Бернулли:

$$W_A(i) = C_m^i \cdot b^i \cdot (1-b)^{m-i}, \quad (7.150)$$

где C_m^i ($C_m^i = \frac{m!}{i!(m-i)!}$) – число сочетаний из m по i ; b – средняя интенсивность нагрузки (Эрл), обслуженная одной промежуточной линией (выходом коммутатора звена A). Величина b равна значению вероятности занятости одного выхода из коммутатора звена A , b^i – вероятность занятости i любых линий, $(1-b)^{m-i}$ – вероятность незанятости оставшихся $(m-i)$ линий.

Если на звене A определяется вероятность занятости любых линий, то на звене B следует рассматривать вероятность занятия конкретных линий, при занятии которых возможны отказы.

При использовании распределения Бернулли вероятность занятости конкретных выходов $H_B(m-i)$ на звене B определяется на основе выражения

$$H_B(m-i) = c^{m-i}, \quad (7.151)$$

где c – средняя интенсивность нагрузки (Эрл), обслуженной одним выходом двухзвенной схемы рассматриваемого направления.

Подставляя полученные выражения для вероятности $W_A(i)$ и $H_B(m-i)$ в выражение (7.149), получим формулу

$$P = \sum_{i=0}^m W_A(i) \cdot H_B(m-i) = \sum_{i=0}^m C_m^i \cdot b^i \cdot (1-b)^{m-i} \cdot c^{m-i}. \quad (7.152)$$

Для решения выражения (7.152) воспользуемся формулой бинома Ньютона

$$\sum_{j=0}^m C_m^j \cdot a^j \cdot b^{m-j} = (a+b)^m. \quad (7.153)$$

С учетом данной формулы и предположения, что $a=b$, а $b=(1-b) \cdot c$, формула (7.152) преобразуется в более удобный для решения вид:

$$\begin{aligned} P &= \sum_{i=0}^m C_m^i \cdot b^i \cdot (1-b)^{m-i} \cdot c^{m-i} = \sum_{i=0}^m C_m^i \cdot b^i \cdot [(1-b) \cdot c]^{m-i} = \\ &= [b + (1-b) \cdot c]^m = (b + c - b \cdot c)^m, \end{aligned} \quad (7.154)$$

где b и c – величины исполненных нагрузок выходами коммутаторов звена A и B двухзвенной схемы соответственно.

Если для образования пучка каналов связи направления связи отводится в каждом коммутаторе второго звена B более одного выхода $q > 1$, то для данного случая (занятие выходов и занятие промежуточных линий можно описать распределением Бернулли) для выходов звена A выражение не изменится:

$$W_A(i) = C_m^i \cdot b^i \cdot (1-b)^{m-i}, \quad (7.155)$$

а для выходов звена B вероятность занятия конкретных линий будет определяться по формуле

$$H_B(m-i) = c^{(m-i)q}. \quad (7.156)$$

Подставляя эти выражения в (7.149) и используя формулу бинома Ньютона, получим

$$\begin{aligned} P &= \sum_{i=0}^m C_m^i \cdot b^i \cdot (1-b)^{m-i} \cdot c^{(m-i)q} = \sum_{i=0}^m C_m^i \cdot b^i \cdot [(1-b) \cdot c^q]^{m-i} = \\ &= [b + (1-b) \cdot c^q]^m = (b + c^q - b \cdot c^q)^m. \end{aligned} \quad (7.157)$$

Аналогично рассуждая, можно получить конечную формулу для вычисления потерь для случая $q > 1$ и $f > 1$:

$$P = (b^f + c^q - b^f \cdot c^q)^m. \quad (7.158)$$

Распределение Бернулли на звене A и распределение Эрланга на звене B (Б – Э) можно применить в случаях:

- построения звена A на коммутаторах с ограниченным числом входов и выходов (так же, как и в первом случае);
- число коммутаторов (k) на звене A должно быть большим ($k > 100$) либо число входов в коммутаторы звена B больше 100.

Для данной схемы при описании занятости выходов из одного коммутатора звена A также можно использовать распределение Бернулли ($n \approx m < 100$).

Но так как число таких коммутаторов k на звене A велико, следовательно, число входов в коммутатор звена B также будет большим, поэтому для выходов звена B целесообразно применить распределение Эрланга.

Отнесем вероятность $W_B(i)$ занятости i любых линий к выходам двухзвенной схемы (звено B), а вероятность $H_A(m-i)$ – к промежуточным линиям (выходам коммутатора звена A). В этом случае вероятность $W_B(i)$ найдем, используя следующее выражение:

$$W_B(i) = \frac{\frac{Z^i}{i!}}{\sum_{j=0}^m \frac{Z^j}{j!}}. \quad (7.159)$$

Вероятность занятости $(m-i)$ конкретных выходов из коммутатора звена $A - H_A(m-i)$ при условии, что $m = n$, $n < 100$, $f = 1$ будет определяться по формуле

$$H_A(m-i) = b^{m-i}. \quad (7.160)$$

Подставляя выражения (7.159) и (7.160) в формулу (7.149), получим

$$P = \sum_{i=0}^m W_A(i) \cdot H_B(m-i) = \sum_{i=0}^m \frac{\frac{Z^i}{i!}}{\sum_{j=0}^m \frac{Z^j}{j!}} \cdot b^{m-i}. \quad (7.161)$$

Преобразуем выражение (7.161) в удобную для расчета форму, чтобы можно было воспользоваться таблицами Пальма либо соответствующими номограммами. Для этого умножим и числитель, и знаменатель выражения (7.161) на одну и ту же величину $\frac{Z^m}{m!}$, тогда получим

$$\begin{aligned} P &= \sum_{i=0}^m \frac{\frac{Z^i}{i!}}{\sum_{j=0}^m \frac{Z^j}{j!}} \cdot b^{m-i} = \sum_{i=0}^m \frac{\frac{Z^i}{i!} \cdot \frac{Z^m}{m!}}{\sum_{j=0}^m \frac{Z^j}{j!} \cdot \frac{Z^m}{m!}} \cdot \frac{b^m}{b^i} = \\ &= \sum_{i=0}^m \frac{\frac{Z^m}{m!}}{\sum_{j=0}^m \frac{Z^j}{j!} \cdot \frac{Z^m}{m!}} \cdot \frac{\frac{Z^i}{i!}}{\frac{Z^m}{m!}} \cdot \frac{b^m}{b^i}. \end{aligned} \quad (7.162)$$

Следует обратить внимание на то, что в формуле (7.162) выра-

жение $\frac{\frac{Z^m}{m!}}{\sum_{j=0}^m \frac{Z^j}{j!} \cdot \frac{Z^m}{m!}}$ представляет первую формулу Эрланга, которую можно

выразить через условную запись эрланговых потерь – $E_m(Z)$.

Так как $E_m(Z)$ не зависит от i , то ее можно вынести за знак суммы. Отсюда выражение (7.162) можно представить в следующем виде:

$$\begin{aligned}
 P &= \sum_{i=0}^m \frac{\frac{Z^m}{m!} \cdot \frac{Z^i}{i!} \cdot \frac{b^m}{b^i}}{\sum_{j=0}^m \frac{Z^j}{j!} \cdot \frac{1}{m!}} = E_m(Z) \cdot \frac{\sum_{i=0}^m \frac{Z^i}{i! \cdot b^i}}{\frac{Z^m}{m! \cdot b^m}} = \\
 &= E_m(Z) \cdot \frac{\sum_{i=0}^m \frac{\left(\frac{Z}{b}\right)^i}{i!}}{\frac{\left(\frac{Z}{b}\right)^m}{m!}} = E_m(Z) \cdot \frac{1}{E_m\left(\frac{Z}{b}\right)} = \frac{E_m(Z)}{E_m\left(\frac{Z}{b}\right)}. \quad (7.163)
 \end{aligned}$$

Таким образом, выражение (7.163) на основе использования таблиц Пальма позволяет достаточно просто произвести расчет потерь в двухзвенной схеме:

$$P = \frac{E_m(Z)}{E_m\left(\frac{Z}{b}\right)}. \quad (7.164)$$

Физическая интерпретация формулы (7.163) заключается в следующем:

- числитель $E_m(Z)$ представляет собой потери в однозвенной схеме, имеющей m выходов, на которые поступает нагрузка Z ;
- знаменатель представляет собой коэффициент $0 \leq E_m(Z/b) \leq 1$, учитывающий увеличение потерь в двухзвенной схеме (по сравнению с однозвенной) за счет внутренних блокировок.

Рассмотрим также случай, когда занятие выходов (звено B) подчиняется распределению Эрланга, а занятие промежуточных линий (выходы звена A) – распределению Бернулли, но количество линий в направлении искания, подключаемых к выходам коммутатора звена B , $q > 1$. В этом случае выражение для потерь (7.163) при некоторых дополнительных ограничениях может быть преобразовано к виду

$$P = \frac{E_{mq}(Z)}{E_{mq}\left(\frac{Z}{b}\right)}. \quad (7.165)$$

В соответствии с рассматриваемым методом данная формула может применяться и для дробных значений q .

Отсюда следует, что выражения (7.164) и (7.165) являются более общими. От них можно перейти к формулам (7.154) и (7.163), полагая $q = 1$.

Если связность схемы не равна единице ($f > 1$), т. е. при наличии нескольких соединительных путей между парой коммутаторов первого и второго звеньев, считают справедливым во всех полученных выше формулах заменить b на b^f .

Таким образом, для условий $n = m$; $a = b$; $f \geq 1$; $q \geq 1$

$$P = \frac{E_{mq}(Z)}{E_{mq}\left(\frac{Z}{b^f}\right)} = \frac{E_{mq}(Z)}{E_{mq}\left(\frac{Z}{a^f}\right)}. \quad (7.166)$$

В заключение следует отметить, что вывод уравнений для определения вероятности потерь в двухзвенной схеме без сжатия и расширения показывает, что использование известных распределений Эрланга и Бернулли позволило применить известные таблицы Пальма.

7.3.3. Определение потерь в двухзвенных схемах.

Метод эффективной доступности (метод Харкевича)

Метод эффективной доступности предназначен для расчета вероятности потерь как для полноступных, так и неполноступных двухзвенных схем в режиме группового искания (ГИ). Он основан на понятии эффективной доступности.

Сущность метода заключалась в том, что А. Д. Харкевич представил работу двухзвенной схемы в режиме ГИ подобно работе однозвенной неполноступной схемы с доступностью, изменяющейся во времени. Сложность заключалась в расчете этой доступности. Однако он предложил формулу для расчета доступности D , названной им

эффективной, и дальнейшая процедура вычисления вероятности потерь двухзвенной схемы при известных значениях V , D состояла в использовании уже известных методов расчета однозвенных неподоступных (НПДС) с потерями, равными потерям в рассматриваемой двухзвенной схеме.

В режиме группового искания в выходы двухзвенной схемы включаются соединительные устройства нескольких направлений. Для подключения соединительных устройств последующей ступени, принадлежавших одному направлению, в каждом коммутаторе второго звена в общем случае может отводиться q выходов.

А. Д. Харкевич рассматривал ее с той позиции, что в двухзвенной полнодоступной схеме каждому входу, подключенному, к примеру, к первому коммутатору звена A , доступен любой выход из $q \cdot t$ имеющихся выходов требуемого направления только тогда, когда нет занятых соединительных путей (промежуточных линий).

При отсутствии поступления вызовов данная схема является полнодоступной двухзвенной схемой, в которой любому из $k \cdot n$ входов доступен любой из $t \cdot q$ выходов. В этом случае доступность выходов направления будет максимальной и $D_{\max} = t \cdot q$ (все выходы доступны).

Однако в процессе обслуживания вызовов доступность выходов в конкретном направлении связи изменяется по мере занятости промежуточных линий в ходе установления соединений в других направлениях связи.

Так, в случае занятости одним входом одной промежуточной линии из t (необязательно в рассматриваемом направлении), подключенных к выходам первого коммутатора звена A , для всех остальных входов этого коммутатора доступность выходов в указанном направлении уменьшится для конкретного случая на единицу, а в общем случае – на величину q . При связности $f = 1$ только одна занятая ПЛ заблокирует все выходы рассматриваемого направления, подключенные, например к коммутатору № 1 звена B .

При наличии i занятых промежуточных линий (выходов) первого коммутатора звена A для всех входов доступность уменьшается на $i \cdot q$ линий и составит величину

$$D_i = (t - i) \cdot q. \quad (7.167)$$

При наличии на первом звене A концентрации (сжатия – $n > m$) минимальная доступность выходов рассматриваемого направления будет равна нулю $D_{\min} = 0$, так как для $(m+1)$ -го занятия, поступающего на вход коммутатора звена A , из-за занятости всех m промежуточных линий невозможно подключение ни к одному из выходов, даже и свободных в данном направлении связи $D_{\min} = 0$.

Для схем с расширением или без расширения и сжатия ($n \leq m$) минимальная доступность больше нуля и составляет величину

$$D_{\min} = (m - n + 1) \cdot q, \quad (7.168)$$

так как к моменту поступления последнего вызова (в коммутаторе первого звена к этому времени $(n-1)$ входов уже заняты) число свободных ПЛ будет равно $(m - n + 1)$, а доступность выходов из схемы $D = (m - n + 1) \cdot q$.

Таким образом, в процессе обслуживания вызовов доступность D_i двухзвенной схемы в режиме группового искания выходов может меняться в пределах между наибольшим и наименьшими значениями:

$$D_{\min} \leq D_i \leq D_{\max}. \quad (7.169)$$

Каждое из значений доступности D_i появляется с вероятностью W_i , где W_i – вероятность занятия i промежуточных линий из m линий, подключенных к выходам одного коммутатора звена A .

При этом вероятность W_i можно трактовать как долю времени в течение рассматриваемого интервала, в пределах которого занято i промежуточных линий, т. е. в течение которого существует доступность D_i .

Из этого следует, что в процессе работы двухзвенной схемы доступность меняется от 0 до $m \cdot q$. При этом число каналов, подключенных к рассматриваемой схеме, равно $V = m \cdot q$. Очевидно, что работа схемы аналогична работе схемы НПДС с параметрами V и D , причем $V \geq D$. Следовательно, в интервалах времени, в пределах которых в системе занято i промежуточных линий, работу двухзвенной полнодоступной схемы можно уподобить работе однозвенной схемы неполнодоступного включения с доступностью D_i . При этом потери P_i , возникающие в это время, будут определяться значением доступности D_i .

Так как значения D_i определяются соотношением (7.169), то потери P в случае включения соединительных устройств в выходы двухзвенной схемы будут меньше, чем при неполнодоступном включении с доступностью $D_{\min}$ (т. е. меньше $P_{\max}$), и больше, чем при неполнодоступном включении с доступностью $D_{\max}$ (т. е. меньше $P_{\min}$):

$$P_{\min} < P < P_{\max}. \quad (7.170)$$

Соотношение (7.170) означает, что рассматриваемая двухзвенная схема с точки зрения потерь будет хуже однозвенного неполнодоступного включения с доступностью $D_{\max}$ и лучше неполнодоступного включения с доступностью $D_{\min}$.

А. Д. Харкевич предположил, что для данной двухзвенной схемы существует эквивалентное ей (с точки зрения потерь) однозвенное неполнодоступное включение с определенной доступностью.

Доступность такого однозвенного неполнодоступного включения с потерями, равными потерям в полнодоступной двухзвенной схеме, он назвал эффективной доступностью и обозначил символом $D_{\mathcal{E}}$.

Из соотношения (7.170) вытекает, что

$$D_{\min} < D_{\mathcal{E}} < D_{\max}. \quad (7.171)$$

Если зависимость вероятности потерь P от доступности D при определенных значениях интенсивности нагрузки Y и числа соединительных устройств последующей ступени искания V обозначить функцией

$$P = P_{Y,V}(D), \quad (7.172)$$

то вероятность потерь в двухзвенной схеме при случайном изменении доступности, эквивалентной ей с позиции потерь однозвенной НПДС в пределах $D_{\min} < D_{\mathcal{E}} < D_{\max}$, можно определить, используя выражение для математического ожидания СВ:

$$P = \sum_{i=0}^r W_i \cdot P_{Y,V}(D_i), \quad (7.173)$$

где $r = n - 1$ при $n \leq m$ или $r = m$ при $n > m$.

Значение вероятности P из формулы (7.173) будет удовлетворять неравенству (7.170).

Так как потери в двухзвенной схеме равны потерям в однозвенной неполнодоступной схеме с доступностью, равной эффективной, то выражение (7.173) можно записать в виде

$$P_{Y,V}(D_{\mathcal{D}}) = \sum_{i=0}^r W_i \cdot P_{Y,V}(D_i). \quad (7.174)$$

Если функциональные зависимости (7.172, 7.173, 7.174) являются линейными, т. е. $P = a \cdot D + b$, где a и b – коэффициенты, то из соотношения (7.174) следует, что для определения потерь необходимо рассчитать эффективную доступность $D_{\mathcal{D}}$, которая в данном случае будет равна средней доступности (математическому ожиданию доступности) – $\bar{D}$:

$$D_{\mathcal{D}} = \bar{D}. \quad (7.175)$$

Однако исследования показали, что в общем случае справедливо соотношение для эффективной доступности:

$$D_{\mathcal{D}} \leq \bar{D}. \quad (7.176)$$

Соотношение (7.176) по сравнению с равенством (7.172) сужает диапазон, в котором может лежать значение $D_{\mathcal{D}}$. Этот диапазон определяется неравенствами

$$D_{\min} < D_{\mathcal{D}} \leq \bar{D}. \quad (7.177)$$

Для нахождения потерь было предложено неравенство (7.177) записать в более удобном для вычислений виде:

$$D_{\mathcal{D}} = D_{\min} + \Theta \cdot (\bar{D} - D_{\min}), \quad (7.178)$$

где Θ – коэффициент, определяемый зависимостью потерь от доступности и распределением W_i (вероятностей доступности). Выбор величины Θ должен производиться в зависимости от характера применяемой расчетной формулы неполнодоступного включения. Обычно величину коэффициента Θ можно принимать в пределах 0,75–0,85.

Для вычисления $D_{\exists}$ необходимо определить минимальную доступность:

- в случае концентрации на первом звене минимальная доступность $D_{\min} = 0$;
- в случае отсутствия концентрации или наличия в схеме расширения минимальная доступность рассчитывается по вышеприведенной формуле (7.168).

Кроме минимальной доступности $D_{\min}$ следует рассчитать среднюю доступность по формуле

$$\begin{aligned}\bar{D} &= \sum_{i=0}^r D_i \cdot W_i = \sum_{i=0}^r W_i \cdot (m - i) \cdot q = \\ &= q \cdot \sum_{i=0}^r (m \cdot W_i - i \cdot W_i) \approx q \cdot (m - Y_m),\end{aligned}\quad (7.179)$$

где $Y_m = i \cdot W_i$ – интенсивность нагрузки, обслуженной m промежуточными линиями, подключенными к выходам коммутатора первого звена (теорема о количественной оценке обслуженной нагрузки),

а $\sum_{i=0}^r m \cdot W_i = m \cdot \sum_{i=0}^r W_i = m$, так как сумма вероятностей занятости ПЛ

$$\sum_{i=0}^r W_i = 1.$$

Таким образом, для определения средней доступности $\bar{D}$ можно использовать формулу

$$\bar{D} = q \cdot (m - Y_m). \quad (7.180)$$

Средняя нагрузка характеризует среднее количество свободных соединительных путей между заданным входом коммутационной системы и выходами требуемого направления. Физическое толкование формулы (7.180) заключается в том, что Y_m – нагрузка, исполненная m выходами, численно равна среднему числу одновременно занятых линий i из m , поэтому формула (7.180) аналогична выражению (7.167).

Для двухзвенной схемы со связностью, равной $f = 2$ математическое ожидание доступности можно определить из следующего выражения:

$$D = q \cdot \left[m - \frac{Y_{mf}^2}{2} \cdot (m \cdot f - 1) \right]. \quad (7.181)$$

После определения эффективной доступности НПДС, потери в которой равны потерям в двухзвенной схеме, дальнейший расчет потерь сводится к использованию одного из известных методов расчета НПДС: Эрланга, Бабицкого, Лотце (метод МПЯ) либо О'Делла.

Когда требуется определить число каналов в направлении связи двухзвенной полnodоступной схемы в режиме ГИ, необходимо использовать формулу О'Делла. Для этого по найденному значению эффективной доступности $D_{\text{Э}}$ и заданной величине потерь P , используя таблицы, следует определить коэффициенты α и β , а затем интенсивность обслуженной нагрузки $Y = Z \cdot (1 - P)$:

$$V = \alpha \cdot Y + \beta. \quad (7.182)$$

Если требуется определить вероятность потерь (для потерь, не превышающих 0,02) при заданной структуре и нагрузке, поступающей на двухзвенную схему при эквивалентной замене реальной двухзвенной схемы на однозвенную НПДС с доступностью $D_{\text{Э}}$, то вероятность потерь в данном случае можно определить, используя формулы Пальма или Бабицкого для расчета неполнодоступных схем:

$$P = \frac{E_V(Z)}{E_{V-D_{\text{Э}}}(Z)}. \quad (7.183)$$

При больших значениях вероятности потерь необходимо применять предложенную А. Лотце так называемую модифицированную формулу Пальма – Якобеуса, в которой вместо реально поступающей нагрузки Z НПДС используется меньшая интенсивность фиктивной поступающей нагрузки $Z_{\text{ф}}$, которая равна обслуженной нагрузке Y при потерях, характерных для полnodоступного включения:

$$P = \frac{E_V(Z_{\text{ф}})}{E_{V-D_{\text{Э}}}(Z_{\text{ф}})}. \quad (7.184)$$

Методика расчета числа необходимых выходов в направлении связи из двухзвенной схемы методом эффективной доступности заключается в следующем:

1. Определяют значения минимальной доступности $D_{\min}$ для случая, когда она не равна нулю:

$$D_{\min} = (m - n + 1) \cdot q. \quad (7.185)$$

2. Находят величину математического ожидания доступности $\bar{D}$:

$$\bar{D} = (m - Y_m) \cdot q. \quad (7.186)$$

3. Определяют эффективную доступность $D_{\mathcal{D}}$:

$$D_{\mathcal{D}} = D_{\min} + \Theta \cdot (\bar{D} - D_{\min}). \quad (7.187)$$

4. Если $D_{\mathcal{D}}$ по условиям точности расчетов можно заменить ближайшим целым числом, то определяют коэффициенты α и β в зависимости от полученной величины $D_{\mathcal{D}}$ и требуемой величины потерь.

5. По известному значению поступающей нагрузки при заданном качестве обслуживания определяют интенсивность обслуженной нагрузки $Y = Z \cdot (1 - P)$.

Число соединительных устройств, которое должно быть включено в выходы рассматриваемой двухзвенной схемы для обслуживания интенсивности поступающей на них нагрузки при заданных потерях, рассчитывается по формуле

$$V = \alpha \cdot Y + \beta. \quad (7.188)$$

7.3.4. Определение потерь в многозвенных схемах.

Метод комбинированной блокировки

Метод комбинированной блокировки основан на том, что потери в многозвенных коммутационных схемах (КС) возникают по двум причинам: во-первых, из-за отсутствия свободных выходов в коммутаторе первого звена (звено A), на который поступил вызов. Эта причина обусловлена тем, что число входов в коммутатор n больше чис-

ла выходов из данного коммутатора m . Вероятность потери вызова из-за ограниченного числа выходов коммутатора первого звена обозначим через P_0 .

Во-вторых, если соединение прошло через коммутатор первого звена, то потери могут возникнуть по причине внутренних блокировок в коммутационном поле и отсутствия свободных выходов в заданном направлении на последнем звене. Вторую составляющую общих потерь обозначим через $P_{\text{бл}}$.

Если предположить, что события, вызывающие первую и вторую причины, независимы, то общая вероятность потерь в многозвенной КС

$$P = P_0 + P_{\text{бл}}. \quad (7.189)$$

Таким образом, для определения вероятности потерь в многозвенной схеме необходимо определить вероятности P_0 и $P_{\text{бл}}$. Рассмотрим порядок их вычисления.

Вероятность потерь P_0 в коммутаторе первого звена представляет вероятность потерь в полnodоступном пучке из m_1 линий, где m_1 – число выходов коммутатора звена A .

В зависимости от числа входов n_1 в коммутатор первого звена и соотношения между числом входов n_1 и его выходов m_1 для определения P_0 могут быть использованы следующие соотношения.

Если $n_1 > m_1$ и $n_1 > 100$, то можно считать, что вероятность поступления вызова не зависит от числа обслуженных входов и значение вероятности P_0 следует определять по формуле Эрланга для однозвенной схемы

$$P_0 = \frac{Z_0^{m_1} / m_1!}{\sum_{i=0}^{m_1} Z_0^i / i!} = E_{m_1}(Z_0), \quad (7.190)$$

где $E_{m_1}(Z_0)$ – условная запись эрланговых потерь, m_1 – число выходов коммутатора звена A ; Z_0 – нагрузка, поступающая на m_1 промежуточных линий.

Таким образом, для нахождения потерь на коммутаторе первого звена схемы достаточно рассчитать нагрузку Z_0 , определить число m_1 . Затем по таблицам Пальма определяется вероятность потерь P_0 .

Ввиду того, что в современных АТС коммутаторы, как правило, имеют ограниченное число входов и выходов (либо 4, либо 8), то вероятность поступления вызова в момент t будет зависеть от числа обслуживаемых в этот момент сообщений, следовательно, поток вызовов можно аппроксимировать моделью примитивного потока, а для определения вероятности потерь в коммутаторе первого звена P_0 следует использовать формулу Энгсета

$$P_0 = \frac{C_{n_1}^{m_1} \cdot z_1^{m_1} \cdot (1 - z_1)^{n_1 - m_1}}{\sum_{i=0}^{m_1} C_{n_1}^i \cdot z_1^i \cdot (1 - z_1)^{n_1 - i}}, \quad (7.191)$$

где z_1 – поступающая нагрузка на один вход звена A , $C_{n_1}^{m_1}$ – число сочетаний из n_1 по m_1 .

При $n_1 < 100$ и $n_1 \approx m_1$ (но $n_1 \geq m_1$) формулу Энгсета можно заменить выражением более простого вида, причем результат получится достаточно точным:

$$P_0 = \left(\frac{Z_0}{m_1} \right)^{m_1}, \quad (7.192)$$

где Z_0 – суммарная нагрузка, поступающая на m_1 выходов коммутатора первого звена.

При числе входов в коммутаторе первого звена, равных числу выходов ($n_1 = m_1$), вероятность потерь на данном звене P_0 будет равна 0 и, следовательно, вероятность потерь в многозвенной схеме будет определяться лишь второй составляющей – $P_{0Л}$.

Рассмотрим порядок определения вероятности потерь $P_{0Л}$ для режима группового искания в многозвенной схеме.

Для этого по аналогии с методом эффективной доступности предполагают, что пучок линий одного направления в многозвенной

схеме можно рассматривать как неполнодоступный пучок в однозвенной схеме из V линий с доступностью $\bar{D}$. Следовательно, вероятность потерь в многозвенной схеме равна вероятности потерь в однозвенной схеме с доступностью $\bar{D}$.

Для случая представления многозвенной схемы в виде однозвенной НПДС вероятность потерь $P_{\text{бл}}$ будет определяться по модифицированной формуле Пальма – Якобеуса:

$$P_{\text{бл}} = \frac{E_V(Z_{\text{ф}})}{E_{V-\bar{D}}(Z_{\text{ф}})}. \quad (7.193)$$

В основе этой формулы лежит замена действующей поступающей нагрузки Z некоторой фиктивной нагрузкой $Z_{\text{ф}}$, поступающей на полнодоступную схему, в которой вероятность занятия $\bar{D}$ конкретных выходов $W_{D_{\text{ср}}}$ равна вероятности потерь $P_{\text{бл}}$ в данной неполнодоступной схеме.

В выражении (7.193) числитель представляет собой условную запись эрланговых потерь в полнодоступном пучке, содержащем V линий, на который поступает фиктивная нагрузка $Z_{\text{ф}}$.

В знаменателе выражения (7.193) содержится условная запись эрланговых потерь, выражающая потери в полнодоступном пучке из $(V - D_{\text{ср}})$ линий при той же нагрузке.

Фиктивная нагрузка определяется путем подбора из следующего соотношения:

$$Z_{\text{ф}} = \frac{Z_{\text{напр}}}{1 - Z_V(Z_{\text{ф}})}, \quad (7.194)$$

где $Z_{\text{напр}}$ – нагрузка, поступающая в рассматриваемое направление многозвенной схемы.

Доступность в многозвенных коммутационных схемах является величиной переменной, зависящей от величины пропущенной нагрузки, а также от числа промежуточных линий между звеньями.

Пусть каждый вход коммутатора первого звена звеньевой КС обслуживает удельную нагрузку a_1 , которая представляет собой пропущенную нагрузку, отнесенную к одному входу. Общая нагрузка,

поступающая на входы одного коммутатора первого звена, равна $a_1 \cdot n_1$. При условии малых потерь на первом звене можно записать равенство $a_1 \cdot n_1 = b_1 \cdot m_1$, где b_1 – нагрузка, исполненная одним выходом коммутатора первого звена.

Согласно определению нагрузки, величина исполненной нагрузки, прошедшей через m_1 выходов, численно равна среднему количеству одновременно занятых этих выходов. Следовательно, зная величину исполненной нагрузки всеми выходами коммутатора $b_1 \cdot m_1$, можно определить число занятых выходов из коммутатора первого звена $i = b_1 \cdot m_1$. Отсюда среднее число свободных и доступных выходов в коммутаторе первого звена $(m_1 - b_1 \cdot m_1)$.

Следовательно, очередному вызову, поступившему на вход коммутатора, в состоянии $b_1 \cdot m_1$ будут доступны в среднем лишь $(m_1 - b_1 \cdot m_1)$ свободных выходов.

Каждый свободный выход коммутатора первого звена соединен в односвязной схеме с одним коммутатором второго звена. В каждом коммутаторе второго звена в среднем имеется $m_2 \cdot (1 - b_2)$ свободных выходов, где b_2 – нагрузка, исполненная одним выходом коммутатора второго звена. Поэтому поступившему вызову с выхода первого звена КС в среднем доступно $m_1 \cdot (1 - b_1) \cdot m_2 \cdot (1 - b_2)$ выходов из коммутатора второго звена.

Среднее число свободных соединительных путей в S -звенной КС, доступных вызывающей линии, будет составлять

$$m_1 \cdot (1 - b_1) \cdot m_2 \cdot (1 - b_2) \cdot \dots \cdot m_{S-1} \cdot (1 - b_{S-1}), \quad (7.195)$$

а средняя доступность при условии того, что в коммутаторе последнего звена для данного направления выделяется q_H линий,

$$D_{cp} = q_H \cdot m_1 \cdot (1 - b_1) \cdot m_2 \cdot (1 - b_2) \cdot \dots \cdot m_{j-1} \cdot (1 - b_{j-1}). \quad (7.196)$$

Ввиду введенных допущений и преобразований формула (7.196) может быть скорректирована:

$$D_{cp} = \left[\prod_{i=1}^{S-1} (m_i - Y_{m_i}) \cdot q_H \right] + v_k \cdot Y_{m_1}, \quad (7.197)$$

где v_k – доля нагрузки в рассматриваемом направлении;

$Y_{m_1} = b \cdot m_1$ – обслуженная нагрузка m_1 выходами (промежуточными линиями) коммутатора первого звена.

При использовании S -звеньевых блокирующих коммутационных схем в режиме индивидуального (линейного) искания (см. рис. 7.18) пучок, состоящий из n_z ПЛ между предпоследним и последним звеньями, входящими в тот коммутатор последнего звена, в который включен требуемый выход, можно рассматривать как некоторое направление j коммутационной схемы с числом звеньев на единицу меньше.

В этом случае работа S -звеньевой схемы в режиме индивидуального искания эквивалентна работе $(S - 1)$ -звеньевой схеме в режиме группового искания.

Вопросы для самоконтроля

1. Опишите модель $M/M/HG/W, r, R/PRR/S$.
2. Опишите модель $M/G/FM/W, r, v, t, R/LF/R$.
3. Символическое обозначение моделей. Что обозначает каждый из символов в символике Кендалла – Башарина?
4. Составьте граф состояний коммутационной системы при $V = 5$.
5. Запишите уравнения состояний для полнодоступного пучка $V = 2$.
6. Что понимают под переходным и установившимся режимами?
7. Перечислите логические трактовки вероятности (по времени, по ансамблю моментов времени, по ансамблю коммутационных систем).
8. Какие свойства потока были использованы при выводе дифференциально-разностных уравнений?
9. Каким образом решается система дифференциально-разностных уравнений?
10. Почему первую формулу Эрланга называют "усеченным распределением Пуассона"?
11. Как соотносятся все виды потерь в модели Эрланга?
12. Как соотносятся все виды потерь в модели Энгсета?
13. Как соотносятся нагрузки в модели СК с примитивным потоком вызовов?

14. Чем отличаются таблицы Пальма от таблиц для расчета СК по формуле Энгсета?
15. Чем отличаются расчеты по таблицам Энгсета и по номограммам?
16. Почему модель системы с ожиданием в (ТТТ) представляется в виде графа состояния?
17. Назовите условия (ограничения и допущения) применения первой и второй формул Эрланга.
18. Каковы характеристики потока вызовов в системе с ожиданием при $i < V$, $i > V$?
19. Как известно, поток освобождений является потоком с условным параметром. Почему же в модели Эрланга (2-я формула) во время нахождения вызовов в очереди поток освобождений является простейшим?
20. В чем отличие аналитического выражения вероятности P_0 для систем с потерями и ожиданием?
21. Для каких значений Z получено выражение P_i для системы с ожиданием: $Z < V$; $Z > V$; $Z = V$?
22. Чем важен для специалистов тот факт, что график первого распределения Эрланга (распределения вероятностей состояния системы с потерями) проходит выше второго распределения (P_i) для систем с ожиданием при $0 \leq i \leq V$?
23. Назовите события, обуславливающие вероятность ожидания в СК.
24. Почему вероятность потерь по вызовам $E_V(Z)$ в СК с потерями меньше вероятности наличия очереди в СК с ожиданием при одинаковых значениях Z и V ?
25. Чем объяснить, что нагрузка, обслуженная системой с потерями, больше нагрузки, которая обслуживается немедленно системой с ожиданием?
26. Как соотносятся поступающая и обслуженная нагрузки в системе с ожиданием?
27. Каков порядок расчета систем распределения с ожиданием с использованием номограмм?
28. Используя номограммы, определите вероятность $P(t_{ож} > 0)$, если $Z = 6$ Эрл, а $V = 4$ каналам.

29. Используя номограммы, определите вероятность $P(t_{\text{ож}} > 0)$, если $Z = 6$ Эрл, а $V=5$ каналам.
30. Используя номограммы, определите вероятность $P(t_{\text{ож}} > 0)$, если $Z = 6$ Эрл, а $V = 10$ каналам.
31. Обоснуйте влияние режимов искания на вероятность потерь.
32. Можно ли использовать двухзвенную односвязную схему в режимах ГИ, ЛИ, СИ?
33. Объясните с позиции потерь, как рационально распределять выходы в направлении связи на звене B двухзвенной схемы.
34. В чем сущность метода Якобеуса?
35. Какими причинами обусловлены потери в простой двухзвенной схеме?
36. Какие ограничения ввел К. Якобеус при разработке метода?
37. Поясните формулу $P = \sum_{i=0}^m W_A(i) \cdot H_B(m-i)$. Что означают составляющие ее вероятности и для чего необходима в формуле сумма?
38. Какие законы распределения использовал К. Якобеус для описания результатов воздействия поступающего потока вызовов на отдельные звенья соединения?
39. Можно ли для определения любых занятых линий использовать распределение Пуассона?
40. Назовите условия применения формулы Пуассона.
41. Можно ли для определения любых и конкретных занятых линий использовать распределение Бернулли? Назовите условия применения формулы Бернулли.
42. В каком случае можно использовать распределение Эрланга для определения вероятности занятости i линий на звене A двухзвенной схемы?
43. В каком случае можно использовать распределение Эрланга для определения вероятности занятости i линий на звене B двухзвенной схемы?
44. В каком случае можно использовать распределение Бернулли для определения вероятности занятости i линий на звене A двухзвенной схемы?
45. В каком случае можно использовать распределение Бернулли для определения вероятности занятости i линий на звене B двухзвенной схемы?

46. Перечислите причины возникновения потерь в двухзвенной схеме при наличии сжатия на звене A .
47. Назовите причины возникновения потерь в двухзвенной схеме при наличии расширения на звене A .
48. По какой формуле определяется вероятность потерь в двухзвенных схемах при отсутствии сжатия и расширения при $n=30$, $k=20$, $f=1$ и $m=30$?
49. По какой формуле определяется вероятность потерь в двухзвенных схемах при отсутствии сжатия и расширения при $n=120$, $k=10$, $f=4$ и $m=30$?
50. По какой формуле определяется вероятность потерь в двухзвенных схемах при $n=30$, $k=10$, $f=1$, $m=20$?
51. По какой формуле определяется вероятность потерь в двухзвенных схемах при $n=10$, $k=10$, $f=2$, $m=20$?
52. Для каких схем пригоден расчёт методом эффективной доступности?
53. В чем сущность метода эффективной доступности?
54. Поясните причины того, что в процессе обслуживания вызовов свободные каналы направления связи становятся недоступными для использования. Поясните причины того, что двухзвенная полностью доступная схема аналогична работе схемы неполнодоступного включения.
55. Чему равна максимальная доступность в двухзвенной односвязной схеме?
56. Чему равна минимальная доступность в двухзвенной односвязной схеме при сжатии на первом звене?
57. Чему равна минимальная доступность в двухзвенной односвязной схеме при смешивании на первом звене?
58. Что такое эффективная доступность?
59. В каких пределах изменяется эффективная доступность?
60. Назовите порядок расчета двухзвенной схемы методом эффективной доступности.
61. Что лежит в основе метода комбинированной блокировки?
62. Какова область применения метода комбинированной блокировки?
63. Перечислите причины отказов в обслуживании в многозвенных схемах.

64. Назовите причины отказов в обслуживании на первом звене P_0 , порядок определения.

65. Перечислите причины отказов в обслуживании на первом звене $P_{0л}$, порядок определения.

66. Чем обусловлена возможность представления полнодоступной звеньевой схемы в виде однозвенной НПДС?

67. Запишите выражение для модифицированной формулы Пальма – Якобеуса, дайте пояснения её составляющих.

68. Что такое фиктивное значение нагрузки, каков порядок ее определения?

69. От каких параметров зависит средняя доступность?

70. Каков порядок определения средней доступности?

71. Какие преобразования необходимо осуществить для использования данного метода при расчете звеньевых схем в режиме линейного искажения?

ЗАКЛЮЧЕНИЕ

Проблема повышения эффективности использования канального ресурса сети связи имеет первостепенное значение как на этапе ее проектирования, так и на этапе ее эксплуатации. Решение данной проблемы стало возможным благодаря разработке методов, которые в значительной степени обобщены в учебном пособии.

За последние годы происходило очень быстрое развитие теории телетрафика мультисервисных сетей связи. Многие задачи, которые ранее казались сложными, успешно решены.

Были исследованы уравнения процесса размножения и гибели, описывающие процедуру обслуживания произвольного числа потоков заявок, различающихся интенсивностью поступления, количеством канального ресурса, выделяемого для обслуживания одной заявки, временем его занятия и др.

На сегодняшний день исследуются модели систем распределения информации, учитывающие зависимость распределения входящего потока и распределения времени обслуживания; явления группового поступления и группового обслуживания заявок, наличия повторных заявок; динамически назначаемых приоритетов; ненадежности обслуживающих приборов и т.п.

Применяя для этих целей метод псевдосостояний, удастся описать немарковские процессы с помощью марковских. Однако при большом числе псевдосостояний решение линейных уравнений не только в аналитическом, но и в численном виде становится затруднительным. В таких случаях для исследования процессов размножения и гибели используют метод статистических испытаний (имитационное моделирование). Нередко такой подход оказывается проще, чем попытки построить аналитическую модель явления и исследовать зависимость между его параметрами на этой модели.

ЛИТЕРАТУРА

1. Алексеев, И. В. Использование сетевых моделей массового обслуживания при анализе телефонных сетей связи с ненадежными каналами / И. В. Алексеев // Техника средств связи. Сер. СС. – 1988. – Вып. 4. – С. 9–13.
2. Арутюнян, А. В. Изучение идеально-симметричных схем с потерями, ожиданием и уходом из очереди / А. В. Арутюнян // Информационные сети и их структура : кн. – М. : Наука, 1976. – С. 106–114.
3. Бронштейн, О. И. Модели приоритетного обслуживания в информационно-вычислительных системах / О. И. Бронштейн, И. М. Духовный. – М. : Наука, 1976. – 220 с.
4. Вентцель, Е. С. Теория вероятностей : учебное пособие для вузов / Е. С. Вентцель. – М. : Высшая школа, 2001. – 575 с.
5. Гнеденко, Б. В. Введение в теорию массового обслуживания / Б. В. Гнеденко, И. Н. Коваленко. – М. : Наука, 1987. – 336 с.
6. Денисьева, О. М. Системы массового обслуживания с ограниченным ожиданием / О. М. Денисьева. – М. : Радио и связь, 1986. – 112 с.
7. Джейсуол, Н. Очереди с приоритетами / Н. Джейсуол. – М. : Мир, 1973. – 279 с.
8. Еонг, С. К. Анализ производительности мультиплексора речи и данных на основе моделирования с помощью марковского процесса восстановления / С. К. Еонг, С. К. Ун // ТИИЭР. – 1988. – Т. 76. – № 10. – С. 148–151.
9. Ершова, Э. Б. Цифровые системы распределения информации / Э. Б. Ершова, В. А. Ершов. – М. : Радио и связь, 1983. – С. 97–107.
10. Захаров, Г. П. Методы исследования сетей передачи данных / Г. П. Захаров. – М. : Радио и связь, 1982. – 320 с.
11. Игнатьев, В. О. Современные методы проектирования автоматических систем коммутации : учебное пособие / В. О. Игнатьев. – Ленинград : ЛЭИС, 1987. – 50 с.
12. Клейнрок, Л. Теория массового обслуживания : пер. с англ. / Л. Клейнрок. – М. : Машиностроение, 1979. – 432 с.
13. Концептуальные положения по построению мультисервисных сетей на ВСС России, 2002.
14. Концепция "Глобальная информационная инфраструктура", 1993.

15. Концепция развития в России сетей кабельного телевидения и систем широкополосного беспроводного доступа типа MMDS, LMDS и MWS (MVDS).

16. Концепция развития национальной системы спутниковой связи и вещания на период до 2010 года.

17. Концепция развития рынка услуг связи в области телерадиовещания в Российской Федерации на период до 2010 года и в перспективе до 2015 года.

18. Концепция развития телекоммуникационных сетей в сельской местности Российской Федерации в 2002 – 2010 гг.

19. Корнышев, Ю. Н. Теория распределения информации : учебное пособие для вузов / Ю. Н. Корнышев, Г. Л. Фань. – М. : Радио и связь, 1985. – 285 с.

20. Корнышев, Ю. Н. Теория телетрафика : учебник для вузов / Ю. Н. Корнышев, А. П. Пшеничников, А. Д. Харкевич. – М. : Радио и связь, 1996. – 272 с.

21. Мартин, Дж. Системный анализ передачи данных. Т. 2. Проектирование систем передачи данных : пер. с англ. / Дж. Мартин. М. : Мир, 1975. – 431 с.

22. Миронов, А. Е. Основные соотношения и табулированные данные, используемые в рамках дисциплины "Теория телетрафика" : пособие / А. Е. Миронов, В. Ю. Головачев, А. И. Макашенко. – Орёл : Академия ФАПСИ, 2000. – 165 с.

23. Миронов, А. Е. Расчетно-справочный материал по теории телетрафика : пособие / А. Е. Миронов, И. А. Саитов, А. И. Макашенко. – Орёл : Академия ФСО России, 2007. – 99 с.

24. Миронов, А. Е. Теория телетрафика. Вып. 1. Основные понятия и определения теории телетрафика : пособие / А. Е. Миронов, В. Ю. Головачев. – Орёл : Академия ФАПСИ, 2002. – 96 с.

25. Программа Евросоюза "Информационное общество для всех" (Information Society for All) e-Europe'2005.

26. Рекомендации МСЭ-Т. Сер. У.100, У. 1000, У. 2000.

27. Седол, Я. Я. Определение вероятностных характеристик неполнодоступной схемы с потерями и ожиданием / Я. Я. Седол // Проблемы распределения информации : кн. – М. : Наука, 1973. – С. 177–180.

28. Степанов, С. Н. Телетрафик мультисервисных сетей связи : учебник для вузов / С. Н. Степанов, В. С. Лагутин. – М. : Радио и связь, 2000. – 320 с.

29. Шнепс, М. А. Системы распределения информации. Методы расчета / М. А. Шнепс. – М. : Связь, 1979. – 342 с.

30. Штермер, Х. Теория телетрафика / Х. Штермер ; под ред. Г. Б. Башарина. – М. : Связь, 1971. – 319 с.

Учебное издание

Еременко Владимир Тарасович
Фисун Александр Павлович
Саитов Игорь Акрамович
Миронов Александр Ефимович
Орешин Андрей Николаевич
Королев Александр Васильевич
Трегубов Роман Борисович
Патронов Дмитрий Юрьевич

Учебное пособие

Редактор Т.С. Новикова
Технический редактор Т.П. Прокудина

Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Орловский государственный университет имени И.С. Тургенева»

Подписано к печати 01.04.2019 г. Формат 60×90 1/16.

Усл. печ. л. 15,2. Тираж 100 экз.

Заказ №_____

Отпечатано с готового оригинал-макета
на полиграфической базе ОГУ имени И.С. Тургенева
302026, г. Орел, ул. Комсомольская, 95.