ФЕДЕРАЛЬНОЕ АГЕНТСТВО СВЯЗИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ
БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САНКТ-ПЕТЕРБУРГСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ТЕЛЕКОММУНИКАЦИЙ
им. проф. М. А. БОНЧ-БРУЕВИЧА»
(СПбГУТ)

Г. А. Фокин

СЕТИ РАДИОДОСТУПА

УЧЕБНОЕ ПОСОБИЕ

САНКТ-ПЕТЕРБУРГ
2019

УДК 621.396.93(075.8)
ББК 32.884.1я 73
Ф 75

Рецензенты:
доктор технических наук,
директор института физики, нанотехнологий и телекоммуникаций,
заведующий кафедрой «Радиоэлектронные средства защиты информации»
Санкт-Петербургского политехнического университета Петра Великого
С. Б. Макаров

доктор технических наук, почетный профессор СПбГУТ
М. А. Сиверс

Утверждено редакционно-издательским советом СПбГУТ
в качестве учебного пособия

Фокин, Г. А.
Ф 75 Сети радиодоступа : учебное пособие / Г. А. Фокин ; СПбГУТ. – СПб., 2019. – 314 с.

Даны методические рекомендации и материалы к лекционным, лабораторным и практическим занятиям по дисциплине «Сети радиодоступа».

Предназначено для студентов бакалавриата очной формы обучения по направлению 11.03.02 «Инфокоммуникационные технологии и системы связи».

УДК 621.396.93(075.8)
ББК 32.884.1я 73

СОДЕРЖАНИЕ

ВВЕДЕНИЕ 6

КОНСПЕКТ ЛЕКЦИЙ 7

1. ОБЩИЕ ПРИНЦИПЫ ПОСТРОЕНИЯ СЕТЕЙ 7

1.1. Эволюция компьютерных сетей 7

1.1.1. *Предпосылки компьютерных сетей* 7

1.1.2. *Первые компьютерные сети* 9

1.1.3. *Конвергенция сетей* 13

1.1.4. *Интернет как фактор развития сетевых технологий* 19

1.2. Организация связи двух компьютеров 22

1.2.1. *Сетевые интерфейсы* 22

1.2.2. *Сетевое программное обеспечение* 26

1.2.3. *Характеристики физических каналов* 32

1.3. Организация связи нескольких компьютеров 35

1.3.1. *Топология физических связей* 35

1.3.2. *Адресация узлов сети* 38

1.3.3. *Коммутация* 41

1.3.4. *Определение информационных потоков* 42

1.3.5. *Маршрутизация* 44

1.3.6. *Продвижение данных* 47

1.3.7. *Мультиплексирование/демультиплексирование* 49

1.3.8. *Разделяемая среда передачи* 50

2. КОММУТАЦИЯ КАНАЛОВ И ПАКЕТОВ 55

2.1. Коммутация каналов 55

2.1.1. *Элементарный канал* 55

2.1.2. *Составной канал* 57

2.1.3. *Неэффективность передачи пульсирующего трафика* 61

2.2. Коммутация пакетов 62

2.2.1. *Буферизация пакетов* 65

2.2.2. *Дейтаграммная передача* 67

2.2.3. *Передача с установлением логического соединения* 68

2.2.4. *Передача с установлением виртуального канала* 71

2.3. Сравнение сетей с коммутацией пакетов и каналов 74

2.3.1. *Транспортная аналогия сетей с коммутацией пакетов и каналов* 74

2.3.2. *Количественное сравнение задержек* 75

2.3.3. *Ethernet – пример технологии с коммутацией пакетов* 81

3. АРХИТЕКТУРА И СТАНДАРТИЗАЦИЯ СЕТЕЙ 85

3.1. Модель OSI 85

3.1.1. *Многоуровневый подход* 85

3.1.2. *Протокол и стек протоколов* 87

3.1.3. *Общая характеристика модели OSI* 88

3.1.4. *Уровни модели OSI* 91

3.1.5. *Модель OSI и сети с коммутацией каналов* 98

3.2. Стандартизация сетей 99

3.2.1. *Понятие открытой системы* 100

3.2.2. *Источники стандартов* 101

3.2.3. *Стеки коммуникационных протоколов* 103

3.2.4. *Информационные и транспортные услуги* 106

3.2.5. *Распределение протоколов по элементам сети* 107
3.2.6. *Классификация компьютерных сетей* 110

4. СЕТЕВЫЕ ХАРАКТЕРИСТИКИ И КАЧЕСТВО ОБСЛУЖИВАНИЯ 115
4.1. Сетевые характеристики 115
4.1.1. *Типы характеристик* 115
4.1.2. *Производительность* 117
4.1.3. *Надежность* 129
4.1.4 *Характеристики сети поставщика услуг* 129
4.2. Качество обслуживания 132
4.2.1. *Постановка задачи обеспечения качества обслуживания* 132
4.2.2. *Приложения и качество обслуживания* 134
4.2.3. *Управление очередями* 137
4.3. Методы обеспечения качества обслуживания 146
4.3.1. *Методы кондиционирования трафика* 146
4.3.2. *Методы обратной связи* 149
4.3.3. *Методы резервирования ресурсов* 151
4.3.4. *Методы инжиниринга трафика* 155
4.3.5. *Работа в недогруженном режиме* 159

5. БЕСПРОВОДНАЯ ПЕРЕДАЧА ДАННЫХ 162
5.1. Беспроводные линии связи 162
5.1.1. *Понятие беспроводной линии связи* 162
5.1.2. *Электромагнитные волны* 163
5.1.3. *Диапазоны радиоволн* 165
5.1.4. *Особенности распространения радиоволн* 169
5.1.5. *Помехи в беспроводной связи и лицензирование* 172
5.2. Беспроводные системы связи 173
5.2.1. *Беспроводные системы связи точка-точка* 173
5.2.2. *Беспроводные системы связи точка-многоточка* 174
5.2.3. *Типы спутниковых систем* 177
5.2.4. *Технологии широкополосного сигнала* 181

6. ТЕХНОЛОГИИ ЛОКАЛЬНЫХ СЕТЕЙ НА РАЗДЕЛЯЕМОЙ СРЕДЕ 186
6.1. Общая характеристика протоколов на разделяемой среде 186
6.1.1. *Стандартная топология и разделяемая среда* 186
6.1.2. *Стандартизация протоколов локальных сетей* 188
6.2. Ethernet 10 Мбит/с на разделяемой среде 191
6.2.1. *Mac-адреса* 191
6.2.2. *Форматы кадров технологии Ethernet* 192
6.2.3. *Доступ к среде и передача данных* 193
6.2.4. *Возникновение коллизии* 195
6.2.5. *Время оборота и распознавание коллизий* 196
6.2.6. *Физические стандарты 10m Ethernet* 198
6.2.7. *Производительность сети 10m Ethernet* 200
6.3. Беспроводные локальные сети IEEE 802.11 203
6.3.1. *Особенности беспроводных локальных сетей* 203
6.3.2. *Топологии локальных сетей стандарта IEEE 802.11* 205
6.3.3. *Стек протоколов IEEE 802.11* 206
6.3.4. *Распределенный режим доступа* 207
6.3.5. *Централизованный режим доступа* 210
6.3.6. *Физические уровни стандарта IEEE 802.11* 211

6.4. Персональные сети Bluetooth 216
6.4.1. *Особенности персональных сетей* 216
6.4.2. *Архитектура Bluetooth* 217
6.4.3. *Поиск и стыковка устройств Bluetooth* 220
6.4.4. *Развитие технологии Bluetooth* 221

ПРАКТИКУМ 223
1. МЕТОДЫ РАЗДЕЛЕНИЯ КАНАЛОВ 223
1.1. Практическое задянтие. Методы разделения каналов 223
1.1.1. *Уплотнение/множественный доступ с частотным разделением* 224
1.1.2. *Уплотнение/множественный доступ с временным разделением* 228
1.1.3. *Распределение ресурса связи в fdma и tdma* 231
1.1.4. *Сравнение производительности FDMA и TDMA* 232
1.1.5. *Организация кодового разделения каналов FHSS* 235
1.2. Лабораторная работа. Кодовое разделение каналов методом DSSS 238
1.2.1. *Организация кодового разделения каналов DSSS* 238
1.2.2. *Моделирование кодового разделения каналов DSSS* 245

2. МЕТОДЫ ДОСТУПА К СРЕДЕ ПЕРЕДАЧИ 248
2.1. Практическое занятие. Методы доступа к среде передачи 248
2.1.1. *Информационный поток в системах многостанционного доступа* 248
2.1.2. *Предоставление каналов по требованию* 249
2.1.3. *Классификация методов многостанционного доступа* 249
2.1.4. *Методы управляемого доступа* 250
2.2. Практическое занятие. Методы случайного доступа *ALOHA* 254
2.2.1. *Алгоритм доступа ALOHA* 254
2.2.2. *Алгоритм доступа SALOHA* 262
2.2.3. *Алгоритм доступа RALOHA* 265
2.2.4. *Производительность алгоритмов SALOHA, RALOHA* 266
2.3. Лабораторная работа. Оценка производительности алгоритма *ALOHA* 270
2.3.1. *Построение дискретно-событийной имитационной модели* 270
2.3.2. *Оценка производительности ALOHA/SALOHA* 274
2.4. Практическое занятие. Методы доступа с контролем несущей CSMA 279
2.4.1. *Особенности протоколов с контролем несущей* 279
2.4.2. *Стратегии настойчивости передачи в протоколах CSMA* 282
2.4.3. *Протокол CSMA с обнаружением коллизий (CSMA/CD)* 283
2.4.4. *Протокол CSMA с устранением коллизий (CSMA/CA)* 286
2.5. Лабораторная работа. Оценка производительности алгоритма CSMA 289
2.5.1. *Постоение дискретно-событийной имитационной модели* 289
2.5.2. *Оценка производительности протокола CSMA* 290

3. РАДИОКАНАЛ СЕТЕЙ РАДИОДОСТУПА 294
3.1. Практическое занятие. Потери распространения и замирания в СРД 294
3.1.1. *Модели распространения радиоволн в СРД* 294
3.1.2. *Оценка дальности связи в СРД* 298
3.2. Лабораторная работа. Эффект захвата в сетях радиодоступа 306
3.2.1. *Эффект захвата в сетях радиодоступа* 306
3.2.2. *Моделирование эффекта захвата в СРД ALOHA/SALOHA* 309

СПИСОК ИСТОЧНИКОВ 313

ВВЕДЕНИЕ

Настоящее учебное пособие предназначено для изучения дисциплины «Сети радиодоступа» в рамках профессионального цикла по направлению подготовки 11.03.02 «Инфокоммуникационные технологии и системы связи» и включает материалы к лекционным, лабораторным и практическим занятиям. Лекционный курс подготовлен с использованием материалов [1]. Практикум подготовлен с использованием материалов [2] и [3], проектов моделей [4] и авторских разработок.

Современные сети радиодоступа – это, в первую очередь, компьютерные сети, поэтому лекционный курс содержит следующие разделы, посвященные вопросам их построения и функционирования: общие принципы построения сетей, коммутация каналов и пакетов, архитектура и стандартизация сетей, сетевые характеристики и качество обслуживания, беспроводная передача данных, технологии локальных сетей на разделяемой среде. Практикум включает следующие разделы: методы разделения каналов, методы доступа к среде передачи, радиоканал сетей радиодоступа. Каждый раздел практикума включает лабораторную работу и практическое занятие, подготовленные в среде Matlab.

КОНСПЕКТ ЛЕКЦИЙ

1. ОБЩИЕ ПРИНЦИПЫ ПОСТРОЕНИЯ СЕТЕЙ

1.1. Эволюция компьютерных сетей

1.1.1. Предпосылки компьютерных сетей

Компьютерные сети являются логическим результатом эволюции двух важнейших научно-технических отраслей современной цивилизации – вычислительной техники и телекоммуникационных технологий.

С одной стороны, компьютерные сети представляют собой группу компьютеров, согласованно решающих набор взаимосвязанных задач, обмениваясь данными в автоматическом режиме. С другой стороны, компьютерные сети могут рассматриваться как средство передачи информации на большие расстояния, для чего в них применяются методы кодирования и мультиплексирования, получившие развитие в телекоммуникациях (рис. 1.1) [2].

Рис. 1.1. Эволюция компьютерных сетей на стыке вычислительной техники и телекоммуникационных технологий

Обратимся сначала к компьютерному корню вычислительных сетей. Первые компьютеры 1950-х гг. – большие, громоздкие и дорогие – предназначались для небольшого числа пользователей. Такие компьютеры применялись в **режиме пакетной обработки** и строились на базе мэйнфрейма – мощного и надежного компьютера универсального назначения. Пользователи подготавливали перфокарты, содержащие данные и команды программ, и передавали их в вычислительный центр (рис. 1.2). Задания нескольких пользователей группировались в пакет, который принимался на выполнение. Оператор мэйнфрейма вводил карты пакета в компьютер, который обрабатывал задание в многопрограммном режиме, оптимизируя распределение процессора и устройств ввода-вывода между заданиями.

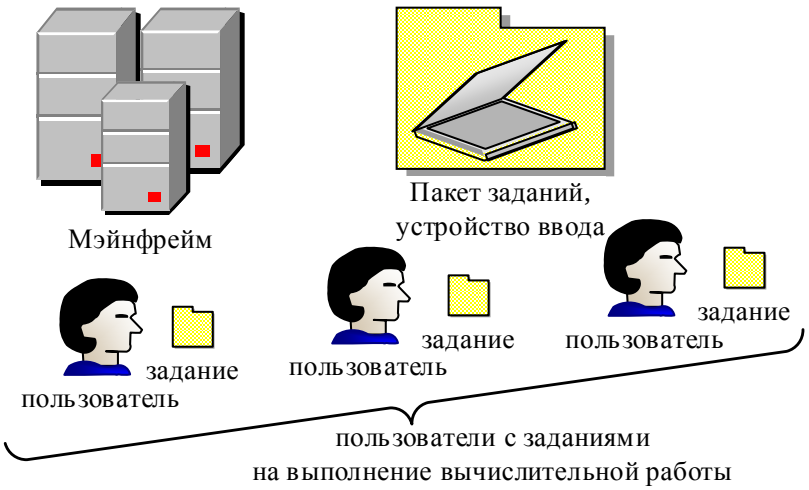


Рис. 1.2. Централизованная система
на базе мэйнфрейма

По мере удешевления процессоров в начале 60-х гг. появились новые способы организации вычислительного процесса, которые позволили учесть интересы пользователей. Начали развиваться интерактивные **многотерминальные системы разделения времени** (рис. 1.3). В таких системах каждый пользователь получал собственный терминал, с помощью которого он мог вести диалог с компьютером. Количество одновременно работающих с компьютером пользователей определялось его мощностью.

Многотерминальные системы, работающие в режиме разделения времени, стали прообразом локальных вычислительных сетей.

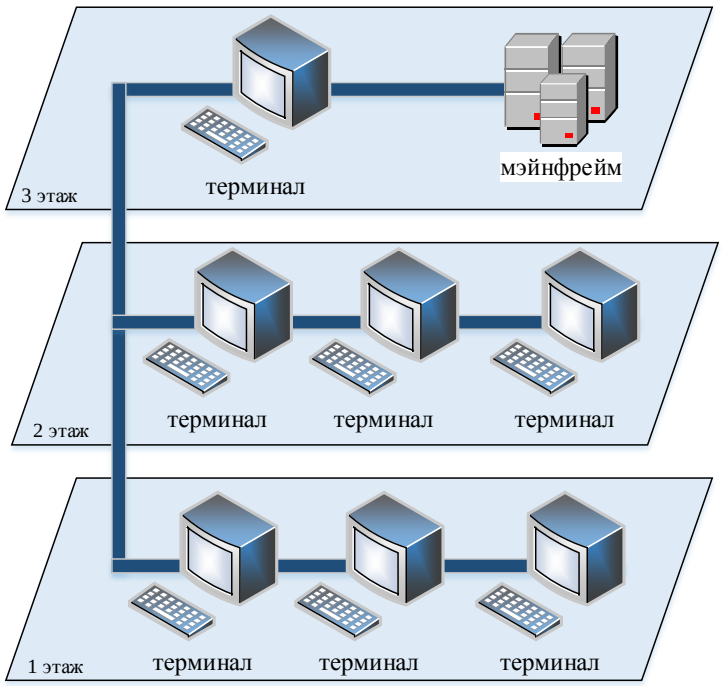


Рис. 1.3. Многотерминальная система –
прообраз вычислительной сети

Однако до появления локальных сетей нужно было пройти еще большой путь, так как многотерминальные системы хотя и имели внешние черты распределенных систем, все еще поддерживали централизованную обработку данных. К тому же потребность предприятий в создании локальных сетей в это время еще не созрела – в одном здании просто нечего было объединять в сеть, так как из-за высокой стоимости вычислительной техники предприятия не могли позволить себе приобретение нескольких компьютеров.

1.1.2. Первые компьютерные сети

Первые глобальные сети появились в результате соединении нескольких компьютеров, находящихся на большом расстоянии друг от друга. Началось все с доступа к отдельному компьютеру терминалов, удаленных от него на многие сотни и тысячи километров. Терминалы соединялись через телефонные сети с помощью модемов, позволив пользователям получать удаленный доступ к разделяемым ресурсам мощных суперкомпьютеров. Затем, наряду с удаленными соединениями типа *терминал – компьютер*, были реализованы и удаленные связи типа *компьютер – компьютер* [2].

Разнесенные территориально компьютеры получили возможность обмениваться данными в автоматическом режиме, что, собственно, и является базовым признаком любой вычислительной сети. В первых сетях были реализованы службы обмена файлами, синхронизации баз данных, электронной почты и другие, ставшие теперь традиционными сетевые службы.

Первыми появились глобальные сети (Wide Area Network, WAN), т. е. сети, объединяющие территориально рассредоточенные компьютеры, возможно находящиеся в различных городах и странах.

Именно при построении глобальных сетей были впервые предложены и отработаны основные идеи, лежащие в основе современных вычислительных сетей. Такие, например, как *многоуровневое построение коммуникационных протоколов, концепции коммутации и маршрутизации пакетов*.

Глобальные компьютерные сети очень многое унаследовали от других, гораздо более старых и распространенных глобальных сетей – *телефонных*. Главное технологическое новшество, которое привнесли с собой глобальные компьютерные сети, состояло в *отказе от принципа коммутации каналов*, на протяжении многих десятков лет использовавшегося в телефонных сетях.

Выделяемый на все время сеанса связи составной телефонный канал, передающий информацию с постоянной скоростью, не мог эффективно использоваться пульсирующим трафиком компьютерных сетей, у которого периоды интенсивного обмена чередуются с продолжительными паузами. Натурные эксперименты и математическое моделирование показали, что пульсирующий и в значительной степени не чувствительный к задержкам

компьютерный трафик гораздо эффективнее передается сетями, работающими по принципу коммутации пакетов, когда данные разделяются на небольшие порции – пакеты, которые самостоятельно перемещаются по сети благодаря наличию адреса конечного узла в заголовке пакета.

Так как прокладка высококачественных линий связи на большие расстояния обходится очень дорого, то в первых глобальных сетях часто использовались уже существующие каналы связи. Например, в течение многих лет глобальные сети строились на основе телефонных каналов тональной частоты, способных в каждый момент времени вести передачу только одного разговора в аналоговой форме. Поскольку скорость передачи дискретных компьютерных данных по таким каналам была очень низкой (десятки килобит в секунду), набор предоставляемых услуг в глобальных сетях подобного типа ограничивался передачей файлов (преимущественно в фоновом режиме) и электронной почтой. Помимо низкой скорости передачи такие каналы имеют и другой недостаток – они вносят значительные искажения в передаваемые сигналы. Поэтому протоколы глобальных сетей, построенных с использованием каналов связи низкого качества, отличались сложными процедурами контроля и восстановления данных. Примером таких сетей являются сети X.25, разработанные в начале 70-х гг.

В 1969 г. министерство обороны США инициировало работы по объединению в единую сеть суперкомпьютеров оборонных и научно-исследовательских центров. Эта сеть, получившая название ARPANET, стала отправной точкой для создания глобальной сети мирового масштаба – Internet.

Сеть ARPANET объединяла компьютеры разных типов, работающие под управлением различных операционных систем (ОС) с дополнительными модулями, реализующими коммуникационные протоколы. ОС этих компьютеров можно считать *первыми сетевыми операционными системами*.

Сетевые ОС позволили не только рассредоточить пользователей между несколькими компьютерами (как в многотерминальных системах), но и организовывать распределенное хранение и обработку данных. Любая сетевая операционная система, с одной стороны, выполняет все функции локальной операционной системы, а с другой – обладает некоторыми дополнительными средствами, позволяющими ей взаимодействовать через сеть с операционными системами других компьютеров.

Прогресс глобальных компьютерных сетей во многом определялся прогрессом телефонных сетей. С конца 60-х гг. в телефонных сетях все чаще стала применяться передача голоса в цифровой форме. Это привело к появлению высокоскоростных цифровых каналов, соединяющих автоматические телефонные станции (АТС) и позволяющих одновременно передавать десятки и сотни разговоров. К настоящему времени глобальные

сети по разнообразию и качеству предоставляемых услуг догнали локальные сети, которые долгое время лидировали в этом отношении, хотя и появились позже.

Первые локальные сети. Важное событие, повлиявшее на эволюцию компьютерных сетей, произошло в начале 70-х гг. В результате технологического прорыва в области производства компьютерных компонентов появились большие интегральные схемы (БИС). Их сравнительно невысокая стоимость и хорошие функциональные возможности привели к созданию *мини-компьютеров*, которые стали реальными конкурентами мэйнфреймов.

Даже небольшие подразделения предприятий получили возможность иметь собственные компьютеры, которые решали задачи управления технологическим оборудованием, другие задачи отдела предприятия. Таким образом, появилась концепция распределения компьютерных ресурсов по всему предприятию. Однако при этом все компьютеры одной организации по-прежнему продолжали работать автономно (рис. 1.4). Шло время и потребности пользователей росли, им уже хотелось в автоматическом режиме обмениваться данными с пользователями других подразделений. Ответом на эту потребность стало появление первых локальных вычислительных сетей (рис. 1.5).

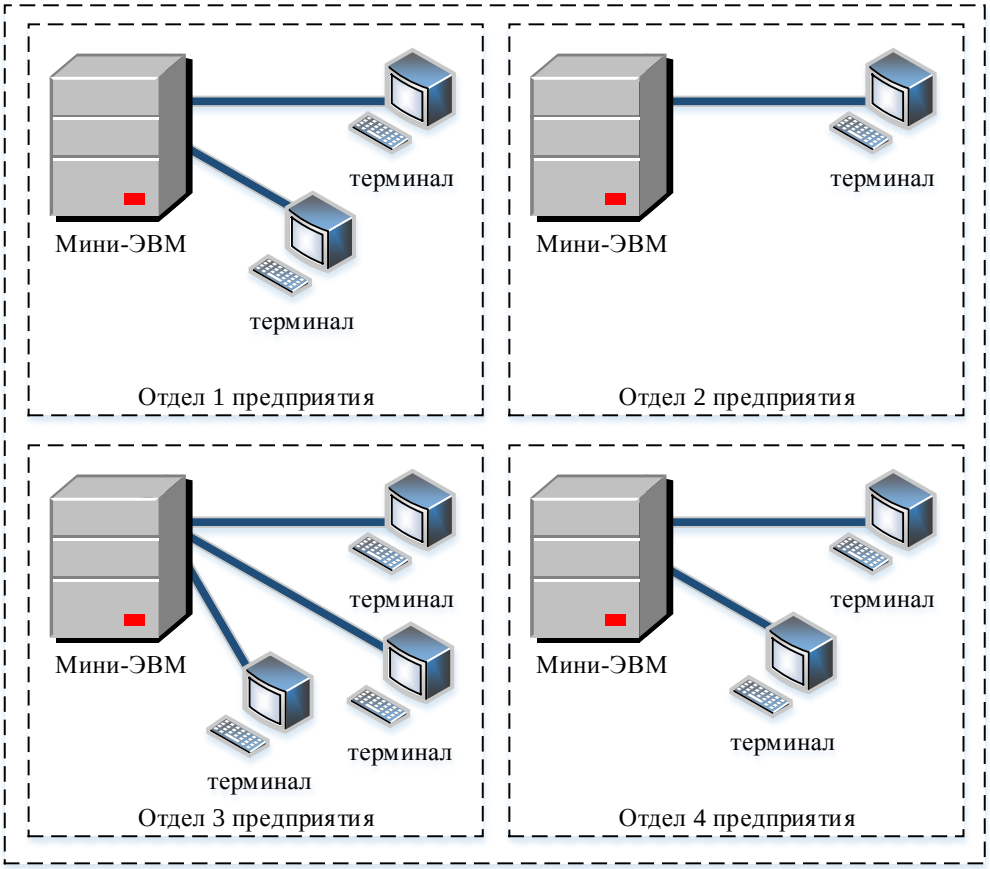


Рис. 1.4. Автономное использование
нескольких мини-компьютеров на одном предприятии

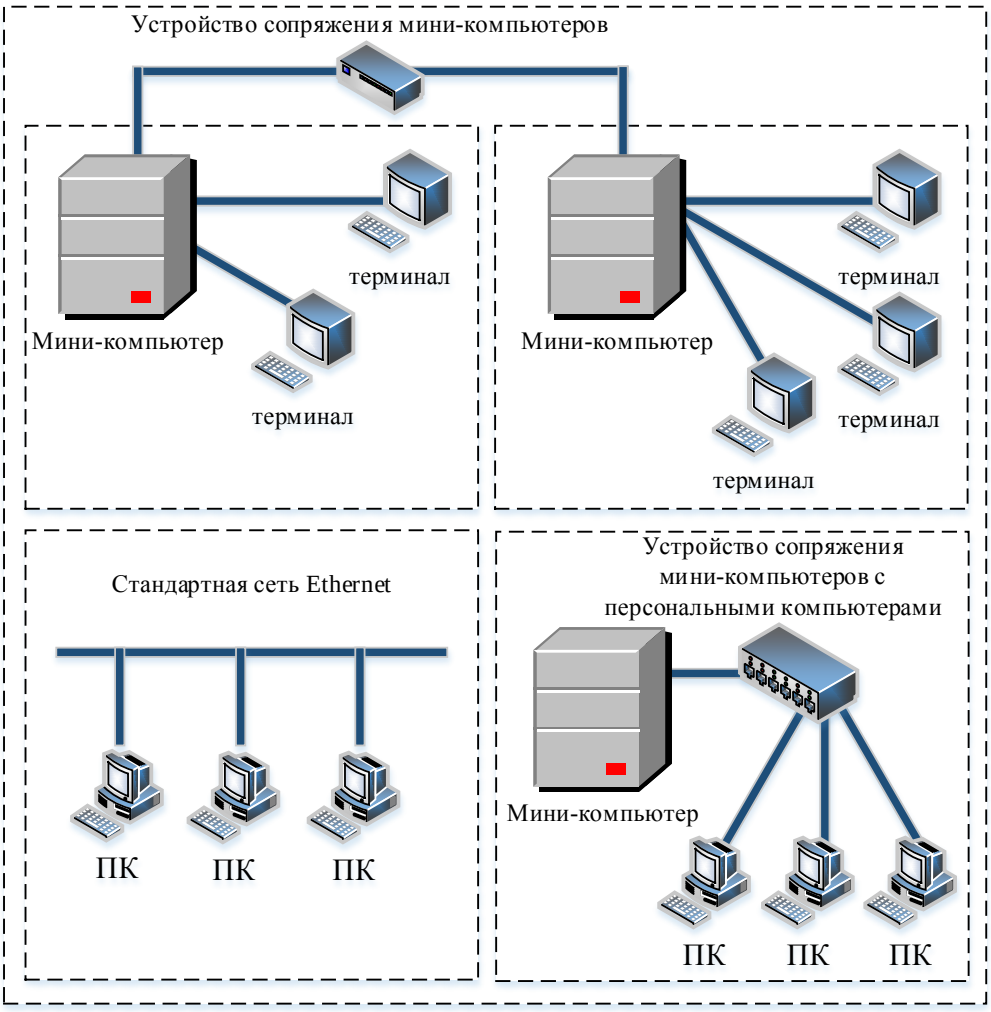


Рис. 1.5. Различные типы связей в первых локальных сетях

Локальные сети (Local Area Network, LAN) – это объединение компьютеров, сосредоточенных на небольшой территории, обычно в радиусе не более 1–2 км. В общем случае локальная сеть представляет собой коммуникационную систему, принадлежащую одной организации [2].

На первых порах для соединения компьютеров друг с другом использовались нестандартные технологии.

Сетевая технология – это согласованный набор программных и аппаратных средств (например, драйверов, сетевых адаптеров, кабелей и разъемов), а также механизмов передачи данных по линиям связи, достаточный для построения вычислительной сети.

Разнообразные устройства сопряжения, использующие собственные способы представления данных на линиях связи, свои типы кабелей и т. п., могли соединять только те конкретные модели компьютеров, для которых были разработаны. В середине 80-х гг. положение дел в локальных сетях кардинально изменилось. Утвердились стандартные сетевые технологии объединения компьютеров в сеть – *Ethernet, Arcnet, Token Ring, FDDI.*

Мощным стимулом для их появления послужили персональные компьютеры. Эти массовые продукты стали идеальными элементами построения сетей – с одной стороны они были достаточно мощными, чтобы обеспечивать работу сетевого программного обеспечения, а с другой – явно нуждались в объединении своей вычислительной мощности для решения сложных задач, а также разделения дорогих периферийных устройств и дисковых массивов. Поэтому персональные компьютеры стали преобладать в локальных сетях, причем не только в качестве клиентских компьютеров, но и в качестве центров хранения и обработки данных, т. е. сетевых серверов, потеснив с этих привычных ролей мини-компьютеры и мэйнфреймы.

Все стандартные технологии локальных сетей опирались на тот же принцип коммутации, который был с успехом опробован и доказал свои преимущества при передаче трафика данных в глобальных компьютерных сетях, – *принцип коммутации пакетов*.

Стандартные сетевые технологии превратили процесс построения локальной сети из решения нетривиальной технической проблемы в рутинную работу. Для создания сети достаточно было приобрести стандартный кабель, сетевые адаптеры соответствующего стандарта, например, Ethernet, вставить адаптеры в компьютеры, присоединить их к кабелю стандартными разъемами и установить на компьютеры одну из популярных сетевых ОС.

Разработчики локальных сетей привнесли много нового в организацию работы пользователей. Так стало намного проще и удобнее, чем в глобальных сетях, получать доступ к общим сетевым ресурсам. Последствием и одновременно движущей силой такого прогресса стало появление огромного числа непрофессиональных пользователей, освобожденных от необходимости изучать специальные команды для сетевой работы.

Конец 90-х гг. выявил явного лидера среди технологий локальных сетей – семейство Ethernet, в которое вошли классическая технология Ethernet со скоростью передачи 10 Мбит/с, а также Fast Ethernet со скоростью передачи 100 Мбит/с и Gigabit Ethernet со скоростью передачи 1000 Мбит/с [2].

Простые алгоритмы работы этой технологии предопределяют низкую стоимость оборудования Ethernet. Широкий диапазон иерархии скоростей позволяет рационально строить локальную сеть, выбирая ту технологию семейства, которая в наибольшей степени отвечает задачам предприятия. Важно также, что все технологии Ethernet очень близки друг другу по принципам работы, что упрощает обслуживание этих сетей.

1.1.3. Конвергенция сетей

Сближение локальных и глобальных сетей. В конце 1980-х гг. отличия между локальными и глобальными сетями проявились отчетливо.

Протяженность и качество линий связи. Локальные компьютерные сети по определению отличаются от глобальных сетей небольшими

расстояниями между узлами сети. Это делает возможным использование в локальных сетях более качественных линий связи. В глобальных сетях 80-х гг. преобладали низкоскоростные телефонные линии связи, передающие дискретную информацию компьютеров со сравнительно частыми искажениями.

Сложность методов передачи данных. В условиях низкой надежности физических каналов в глобальных сетях требуются более сложные, чем в локальных сетях, методы передачи данных и оборудование.

Скорость обмена данными в локальных сетях (10, 16 и 100 Мбит/с) тогда была существенно выше, чем в глобальных (от 2,4 Кбит/с до 2 Мбит/с).

Разнообразие услуг. Высокие скорости обмена данными позволили предоставлять в локальных сетях широкий спектр услуг: разнообразные механизмы использования файлов, хранящихся на дисках других компьютеров сети, совместное использование устройств печати, модемов, факсов, доступ к единой базе данных, электронная почта и другие. Глобальные же сети ограничивались почтовыми и файловыми услугами в их простейшем виде.

Постепенно различия между локальными и глобальными сетями стали сглаживаться. Изолированные ранее локальные сети начали объединять друг с другом, при этом в качестве связующей среды использовались глобальные сети. Тесная интеграция локальных и глобальных сетей привела к значительному взаимопроникновению соответствующих технологий.

Сближение в методах передачи данных происходит на платформе цифровой передачи данных по волоконно-оптическим линиям связи (ВОЛС). Эта среда передачи используется практически во всех технологиях локальных сетей для скоростного обмена информацией на расстояниях свыше 100 метров, на ней же стали строиться магистрали первичных сетей SDH и DWDM, предоставляющих свои цифровые каналы для объединения оборудования глобальных компьютерных сетей. Высокое качество цифровых каналов изменило требования к протоколам глобальных компьютерных сетей. На первый план вместо процедур обеспечения надежности вышли процедуры обеспечения гарантированной средней скорости доставки информации пользователям, а также механизмы приоритетной обработки пакетов особенно чувствительного к задержкам трафика, например, голосового. Эти изменения нашли отражение в таких технологиях глобальных сетей 90-х гг., как Frame Relay и ATM. В этих технологиях предполагается, что искажение битов происходит настолько редко, что ошибочный пакет выгоднее просто уничтожить, а все проблемы, связанные с его потерей, перепоручить программному обеспечению более высокого уровня, которое непосредственно не входит в состав сетей Frame Relay и ATM.

Большой вклад в сближение локальных и глобальных сетей внесло доминирование протокола IP. Этот протокол может работать поверх любых технологий локальных и глобальных сетей (Ethernet, MPLS, Token Ring, ATM, Frame Relay), объединяя различные подсети в единую составную сеть.

Начиная с 1990-х гг. компьютерные глобальные сети, работающие на основе скоростных цифровых каналов, существенно расширили спектр предоставляемых услуг и догнали в этом отношении локальные сети. Стало возможным создание служб доставки пользователю больших объемов информации в реальном времени – изображений, видеофильмов, голоса, т. е. мультимедийной информации. Наиболее яркий пример – гипертекстовая информационная служба *World Wide Web (веб-служба)*, ставшая основным поставщиком информации в Интернете. Ее интерактивные возможности превзошли возможности многих аналогичных служб локальных сетей, так что разработчикам приложений локальных сетей пришлось просто позаимствовать эту службу у глобальных сетей. Процесс переноса технологий из глобальной сети Интернет в локальные приобрел такой массовый характер, что появился специальный термин – *intranet-технологии* (*intra* – внутренний).

Возникли новые транспортные технологии, которые стали одинаково успешно работать как в локальных, так и в глобальных сетях. Первой такой технологией была ATM, которая могла эффективно объединять все существующие типы трафика в одной транспортной сети. Однако истинно универсальной транспортной технологией стала технология Ethernet. Долгие годы Ethernet был технологией только локальных сетей, однако дополненная новыми функциями и новыми уровнями скоростей, эта технология (называемая в этом варианте Carrier Ethernet, т. е. Ethernet операторского класса) сегодня преобладает на линиях связи и глобальных сетей. Следствием доминирования технологии Ethernet в первом десятилетии XXI в. стало упрощение структуры как локальных, так и глобальных сетей – в подавляющем большинстве подсетей сегодня работает протокол Ethernet, а объединяются подсети в составную сеть с помощью протокола IP.

Еще одним признаком сближения локальных и глобальных сетей является появление сетей, занимающих промежуточное положение. Городские сети, или сети мегаполисов (Metropolitan Area Network, MAN), предназначены для обслуживания территории крупного города [2].

Эти сети используют цифровые линии связи, часто оптоволоконные, со скоростями на магистрали 10 Гбит/с и выше. Они обеспечивают экономичное соединение локальных сетей между собой, а также выход в глобальные сети. Сети MAN первоначально были разработаны только для передачи данных, но сейчас перечень предоставляемых ими услуг расширился, в частности они поддерживают видеоконференции и интегральную передачу голоса и текста. Современные сети MAN отличаются разнообразием предоставляемых услуг, позволяя своим клиентам объединять коммуникационное оборудование различного типа, в том числе офисные АТС.

Новой вехой на пути конвергенции сетей обещают стать так называемые *облачные вычисления*, которые позволяют разгрузить пользовательский

компьютер и перенести выполнение приложений на некоторые удаленные компьютеры, связанные с пользовательским компьютером через сеть.

Конвергенция компьютерных и телекоммуникационных сетей. Начиная с 1980-х гг. предпринимаются попытки создания универсальной, так называемой *мультисервисной сети*, способной предоставлять услуги как компьютерных, так и телекоммуникационных сетей.

К телекоммуникационным сетям относятся радиосети, телефонные и телевизионные сети. Главное, что объединяет их с компьютерными сетями, – это то, что в качестве ресурса, предоставляемого клиентам, выступает информация. Однако имеется некоторая специфика, касающаяся вида, в котором предоставляют информацию компьютерные и телекоммуникационные сети. Компьютерные сети изначально разрабатывались для передачи алфавитно-цифровой информации, или просто *данных*, поэтому у компьютерных сетей имеется и другое название – *сети передачи данных*, в то время как телекоммуникационные сети были созданы для передачи только голосовой информации (и изображения в случае телевизионных сетей).

Сегодня мы являемся свидетелями конвергенции телекоммуникационных и компьютерных сетей. Наблюдается *сближение видов услуг*. Первая попытка создания мультисервисной сети, способной оказывать услуги телефонии и передачи данных, привела к появлению в 80-х гг. технологии *цифровых сетей с интегрированным обслуживанием (Integrated Service Digital Network, ISDN)*. Однако на практике ISDN предоставляет сегодня в основном телефонные услуги, а на роль глобальной мультисервисной сети нового поколения *Next Generation Network (NGN)* претендует *Интернет*.

Интернет уже сегодня превратился из сети, предназначенной для оказания небольшого набора услуг передачи данных, основными из которых были передача файлов и обмен текстовыми почтовыми сообщениями, в действительно мультисервисную сеть. Интернет может оказывать все виды телекоммуникационных услуг, в том числе услуг мгновенных сообщений, видеоконференций, IP-телефонии, IP-телевидения, а также услуг социальных сетей. Очевидно, что мультисервисность Интернета будет только возрастать.

Прорывом в процессе конвергенции сетей явилось появление *смартфонов* – терминальных устройств, которые объединили в себе функции мобильных телефонов и персональных компьютеров. Для поддержки новых функций современные сети мобильной связи также стали мультисервисной сетью и предоставляют полный набор как телефонных, так и компьютеризованных информационных услуг (просмотр вэб-страниц в такой же удобной форме, как и на экране компьютера, услуги электронной почты и видеоконференций, просмотр фильмов, публикация информации в социальных сетях и т. п.).

Технологическое сближение сетей происходит сегодня на основе цифровой передачи информации различного типа, *метода коммутации паке-*

тов и программирования услуг. Важным шагом телефонии навстречу компьютерным сетям было прежде всего представление голоса в цифровой форме, что сделало принципиально возможным передачу телефонного и компьютерного трафиков по одним и тем же цифровым каналам (телевидение сегодня также может передавать изображение в цифровой форме). Телефонные сети широко используют комбинацию методов *коммутации каналов* и пакетов. Так, для передачи служебных сообщений (называемых сообщениями сигнализации) применяются протоколы коммутации пакетов, а для передачи голоса между абонентами коммутируется традиционный составной канал.

Сегодня пакетные методы коммутации постепенно теснят традиционные для телефонных сетей методы коммутации каналов даже при передаче голоса. У этой тенденции есть достаточно очевидная причина – на основе метода коммутации пакетов можно более эффективно использовать пропускную способность каналов связи и коммутационного оборудования. Например, паузы в телефонном разговоре могут составлять до 40 % общего времени соединения, однако только пакетная коммутация позволяет «вырезать» паузы и использовать высвободившуюся пропускную способность канала для передачи трафика других абонентов. Другой веской причиной перехода к коммутации пакетов является популярность Интернета – сети, построенной на основе данной технологии.

Обращение к технологии коммутации пакетов для одновременной передачи через пакетные сети разнородного трафика – голоса, видео, текста – сделало актуальной разработку новых методов обеспечения требуемого *качества обслуживания (Quality of Service)*. Методы QoS призваны минимизировать уровень задержек для чувствительного к ним трафика, например, голосового, и одновременно гарантировать среднюю скорость и динамичную передачу пульсаций для трафика данных.

Однако неверно было бы говорить, что методы коммутации каналов устарели и у них нет будущего. На новом витке спирали развития они находят свое применение уже в новых технологиях первичных сетей, служащих основой как для компьютерных, так и телефонных сетей: Optical Transport Networks (OTN) и Dense Wavelength Division Multiplexing (DWDM).

Компьютерные сети многое позаимствовали у телефонных и телевизионных сетей. В частности, они взяли на вооружение методы обеспечения отказоустойчивости телефонных сетей, за счет которых последние демонстрируют высокую степень надежности, так недостающую порой Интернету.

Телефонные сети, в свою очередь, многое перенимают у компьютерных сетей. Особенно это заметно в мобильных телефонных сетях, которые стали использовать протокол IP в сетях 3-го и 4-го поколений (3G и 4G).

Сегодня становится все более очевидным, что мультисервисная сеть нового поколения не может быть создана в результате «победы» какой-

нибудь одной технологии. Ее может породить только процесс конвергенции, когда от каждой технологии берется все самое лучшее и соединяется в некоторый новый сплав, который и обеспечивает требуемое качество для поддержки существующих и создания новых услуг. Появившийся термин – *инфокоммуникационная сеть* – прямо говорит о двух составляющих современной сети – информационной (компьютерной) и телекоммуникационной.

1.1.4. Интернет как фактор развития сетевых технологий

Интернет является самой быстрорастущей технической системой в истории человечества. Интернет растет постоянно, начиная с 1980-х гг. и в соответствии с прогнозами специалистов будет продолжать расти. «Размеры» Интернета можно оценивать по-разному, чаще всего используют такие показатели, как число подключенных к Интернету терминальных устройств (компьютеров различных типов, планшетов, мобильных телефонов), количество пользователей, объем трафика, передаваемый в единицу времени.

На рис. 1.6 показан график роста числа пользователей Интернета за 40 лет существования этой сети. К 2016 г. их число достигло 3,5 миллиардов, что составляет 46 % населения земного шара [2].

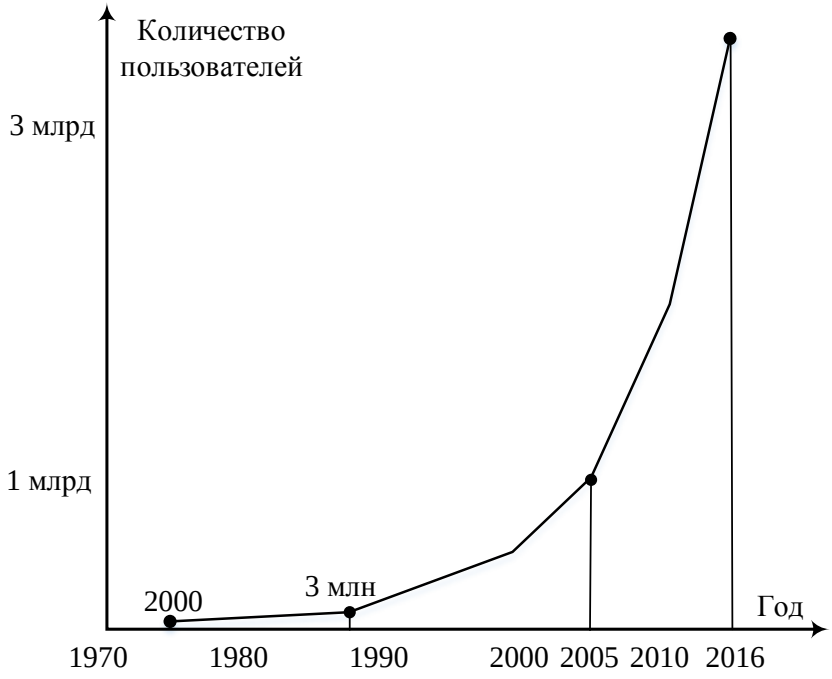


Рис. 1.6. Рост числа пользователей Интернета

Количество терминальных устройств, выполняющих функции серверов (без учета пользовательских устройств), росло примерно такими же темпами: в 1980 г. насчитывалось около 1000 хостов, подключенных к Ин-

тернету, в 1991 – более 1 000 000, в начале 2000-х – около 100 000 000 и, наконец, в 2013 – свыше 1 млрд. С учетом пользовательских устройств (настольных компьютеров, ноутбуков, планшетов и мобильных телефонов) общее количество терминальных устройств, подключенных к Интернету в 2013 г. составило 12 млрд. Абсолютно взрывным оказался рост объема трафика, переданного в месяц через магистрали Интернета):

- 1990 – 1 ТВ (1 терабайт = 10^{12} байт, или 1000 гигабайт);
- 1996 – 2000 ТВ;
- 2000 – 84 РВ (1 петабайт = 1000 терабайт);
- 2008 – 10 ЕВ (1 экзабайт = 100 петабайт);
- 2013 – 50 ЕВ.

В середине 1990-х трафик рос особенно быстро, удваиваясь каждый год, т. е. демонстрируя экспоненциальный рост. Трафик рос не только в количественном отношении – существенно менялся процентный состав приложений, генерирующих трафик. Так, если в 90-е гг. и начале 2000-х в общем объеме преобладал трафик приложений, передающих файлы (файлы электронной почты, веб-страниц, музыки, кинофильмов), то уже к 2010 г. он уступил лидерство трафику приложений, передающих видеопотоки в реальном масштабе времени (таких, как интернет-телевидение, просмотр кинофильмов в онлайновом режиме по требованию, видеоконференции). *Изменение характера трафика* породило новые вызовы разработчикам сетевых технологий, так как требования к характеристикам сети у этих приложений значительно отличаются от требований приложений передачи файлов.

Еще одним революционным изменением в области передаваемого трафика стало резкое увеличение его доли, генерируемой мобильными устройствами – планшетами и мобильными телефонами. И если пока еще большая часть трафика генерируется персональными компьютерами (67 % в 2013 г.), то к 2018 г. эта доля, по прогнозам, упадет до 43 %, остальное будут генерировать мобильные устройства.

Такой феноменальный рост и изменчивость Интернета (в различных аспектах) оказывали и оказывают сильнейшее влияние на технологии компьютерных сетей, заставляя их постоянно изменяться и совершенствоваться, приспосабливаясь к новым требованиям. Эту движущую силу нужно учитывать при изучении любых технологий компьютерных сетей, основные из которых рассматриваются далее. А пока для иллюстрации того, как технологии отвечали на вызов роста, ограничимся таким понятным показателем, как скорость передачи данных транспортными сетевыми технологиями, и посмотрим, как она изменялась в локальных и глобальных сетях (табл. 1.1.) [2].

Таблица 1.1

Хронология повышения скорости передачи данных сетевыми технологиями

Время	Локальные сети	Глобальные сети
1980-е годы	Большинство сетей используют Ethernet 10 Мбит/с, Token Ring 16 Мбит/с	Магистраль Интернета построена на цифровых телефонных каналах 56 Кбит/с; магистрали телефонных сетей используют цифровые линии 35–45 Мбит/с
1990-е годы	Переход на 100 Мбит/с (FDDI и Fast Ethernet)	Магистрали SDH 155 и 622 Мбит/с начинают применяться в Интернете
конец 1990-х – начало 2000-х	В 1998 г. появляется Gigabit Ethernet (1000 Мбит/с) и уже через четыре года, в 2002 г., – 10G Ethernet (10 Гбит/с)	Иерархия скоростей SDH повышается до 10 Гбит/с; технология DWDM позволяет мультиплексировать в одном оптическом волокне до 40–80 каналов по 10 Гбит/с (общая пропускная способность волокна составляет 400–800 Гбит/с)
начало 2010-х	40G и 100G Ethernet стандартизированы в 2012 г., версия 40G начинает применяться в серверах, а 100G – на магистральных сетях	

Как видно из краткой хронологии в табл. 1.1, разработчики сетевых транспортных технологий смогли за 35 лет повысить потолок скорости в 10 000 раз. Подводя итог, перечислим важнейшие события, ставшие историческими вехами на пути эволюции компьютерных сетей (табл. 1.2) [2].

Таблица 1.2

Хронология важнейших событий появления первых компьютерных сетей

Время	Этап
конец 1960-х	Первые глобальные связи компьютеров, первые эксперименты с пакетными сетями
конец 1960-х	Начало передач по телефонным сетям голоса в цифровой форме
начало 1970-х	Появление больших интегральных схем, первые мини-компьютеры, первые нестандартные локальные сети
1974	Стандартизация технологии X.25 для построения сети «удаленные терминалы – мэйнфрейм»
начало 1980-х	Появление персональных компьютеров, создание Интернета в современном виде, установка на всех узлах стека TCP/IP
середина 1980-х	Появление стандартных технологий локальных сетей (Ethernet – 1980 г., Token Ring, FDDI – 1985 г.)
конец 1980-х	Начало коммерческого использования Интернета
конец 1980-х	Появление первичных сетей SONET/SDH со скоростью передачи до 155 Мбит/с
1991	Изобретение Web
конец 1990-х	Доминирование Ethernet в локальных сетях, стандартизация Gigabit Ethernet
конец 1990-х	Появление технологии плотного мультиплексирования волн (DWDM) с возможностью передачи 40/80 волн в одном волокне
конец 1990-х	Появление первых смартфонов с ограниченными интернет-функциями
конец 1990-х – начало 2000-х	Интернет становится мультимедийным (IP-TV, IP-телефония)
начало 2000-х	Повышение скорости передачи данных до 10 Гбит/с (10G Ethernet и 10G SDH/OTN)
середина 2000-х	Смартфоны становятся полнофункциональными интернет-терминалами
начало 2010-х	Повышение скорости передачи данных до 100 Гбит/с (100G Ethernet и 100G OTN)

Выводы

Компьютерные сети стали логическим результатом эволюции вычислительной техники и телекоммуникационных технологий. Прообразом локальных вычислительных сетей являются многотерминальные системы, работающие в режиме разделения времени.

Хронологически первыми появились глобальные сети (Wide Area Network, WAN), т. е. сети, объединяющие территориально-рассредоточенные компьютеры, возможно, находящиеся в различных городах и странах. Для связывания компьютеров в сеть операционные системы, установленные на них, были дополнены модулями, которые реализовывали коммуникационные протоколы, общие для всех компьютеров сети. Такие ОС можно считать первыми сетевыми операционными системами. Сетевые ОС позволили не только рассредоточить пользователей между несколькими компьютерами (как в многотерминальных системах), но и организовать распределенное хранение и обработку данных.

В начале 1970-х гг. начались работы по созданию первой и самой известной ныне глобальной сети мирового масштаба – Internet. Важнейший этап в развитии сетей – появление стандартных сетевых технологий: Ethernet, FDDI, Token Ring, позволяющих быстро и эффективно объединять компьютеры различных типов.

Начиная с 1980-х гг. стала проявляться тенденция сближения технологий локальных и глобальных компьютерных сетей, а также технологий телекоммуникационных сетей разных типов: телефонных, радио, телевизионных. В настоящее время ведутся активные работы по созданию универсальных мультисервисных сетей, способных одинаково эффективно передавать информацию любого типа: данные, голос и видео.

Феноменальный рост количества узлов и трафика Интернета, появление мобильных терминальных устройств – планшетов и смартфонов – оказывали и оказывают сильнейшее влияние на технологии компьютерных сетей, заставляя их постоянно изменяться и совершенствоваться, приспосабливаясь к новым требованиям пользователей.

Контрольные вопросы

1. Что было унаследовано компьютерными сетями от вычислительной техники, а что от телефонных сетей?

2. Какие свойства многотерминальной системы отличают ее от компьютерной сети?

3. В чем технология коммутации пакетов превосходит технологию коммутации каналов?

4. Что такое ARPANET?

5. Какое из этих событий произошло позже других?

а) изобретение Web;

б) появление стандартных технологий LAN;

в) начало передачи голоса в цифровой форме по телефонным сетям.

6. Какое событие послужило стимулом к активизации работ по созданию LAN?

7. Поясните термины «мультисервисная сеть», «инфокоммуникационная сеть».

8. По каким направлением идет сближение компьютерных и телекоммуникационных сетей?

9. Каким образом развитие Интернета влияет на развитие сетевых технологий?

10. Поясните, почему глобальные компьютерные сети появились раньше локальных?

1.2. Организация связи двух компьютеров

1.2.1. Сетевые интерфейсы

Совместное использование ресурсов. Исторически главной целью объединения компьютеров в сеть было *разделение ресурсов*: пользователи компьютеров, подключенных к сети, или приложения, выполняемые на этих компьютерах, получают возможность автоматического доступа к разнообразным ресурсам остальных компьютеров сети, к числу которых относятся: а) периферийные устройства, такие как диски, принтеры, плоттеры, сканеры и др.; б) данные, хранящиеся в оперативной памяти или на внешних запоминающих устройствах; в) вычислительная мощность (за счет удаленного запуска «своих» программ на «чужих» компьютерах).

Чтобы обеспечить пользователей разных компьютеров возможностью совместного использования ресурсов сети, компьютеры необходимо оснастить некими дополнительными *сетевыми* средствами [2].

Рассмотрим простейшую сеть, состоящую из двух компьютеров, к одному из которых подключен принтер (рис. 1.7). Какие дополнительные средства должны быть предусмотрены в обоих компьютерах, чтобы с принтером мог работать не только пользователь компьютера В, к которому этот принтер непосредственно подключен, но и пользователь компьютера А?

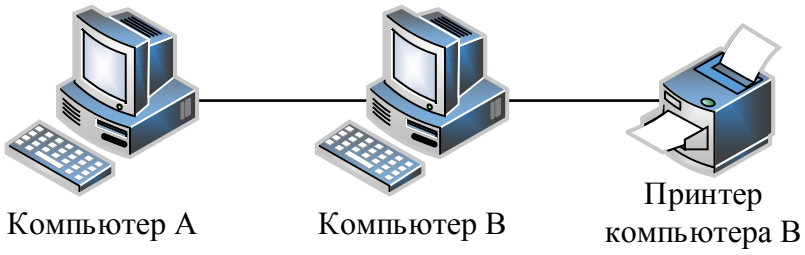


Рис. 1.7. Простейшая сеть

Сетевые интерфейсы. Для связи устройств в них прежде всего должны быть предусмотрены внешние интерфейсы. *Интерфейс* – в широком смысле – формально определенная логическая и/или физическая граница между взаимодействующими независимыми объектами. Интерфейс задает параметры, процедуры и характеристики взаимодействия объектов. Разделяют физический и логический интерфейсы.

Физический интерфейс (или *порт*) определяется набором электрических связей и характеристик сигналов. Обычно представляет собой разъем с набором контактов, каждый из которых имеет определенное назначение. Пара разъемов соединяется кабелем, состоящим из набора проводов, каждый из которых соединяет соответствующие контакты. В таких случаях говорят о создании *линии, или канала связи* между двумя устройствами.

Логический интерфейс (или *протокол*) – это набор сообщений определенного формата, которыми обмениваются два устройства или две программы, а также набор правил, определяющих логику обмена сообщениями.

На рис. 1.8 представлены интерфейсы двух типов: компьютер – компьютер и компьютер – периферийное устройство [2].

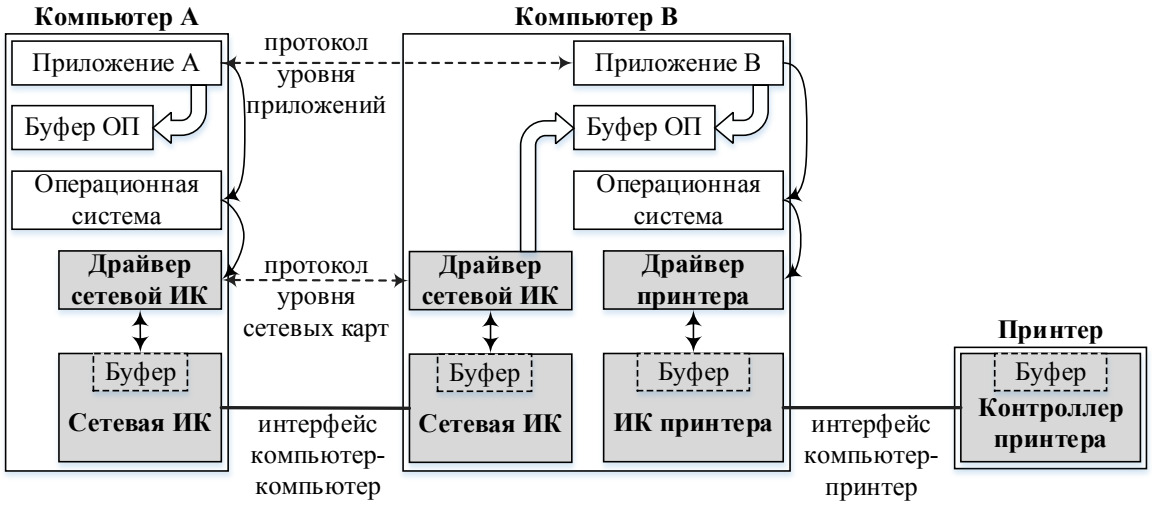


Рис. 1.8. Совместное использование принтера в компьютерной сети

Интерфейс компьютер – компьютер позволяет двум компьютерам обмениваться информацией и реализуется: а) аппаратным модулем, называемым *сетевым адаптером*, или *сетевой интерфейсной картой* (Network Interface Card, NIC); б) *драйвером сетевой интерфейсной карты* – специальной программой, управляющей работой сетевой интерфейсной карты.

Интерфейс компьютер – периферийное устройство (в данном случае интерфейс компьютер – принтер) позволяет компьютеру управлять работой периферийного устройства (ПУ) и реализуется: а) со стороны компьютера – *интерфейсной картой* и *драйвером ПУ* (принтера), подобным сетевой интерфейсной карте и ее драйверу; б) со стороны ПУ – *контроллером ПУ* (принтера), обычно представляющим собой аппаратное устройство, принимающее от компьютера как данные, например, байты информации, которую нужно распечатать, так и команды, которые он отрабатывает, управляя электромеханическими частями периферийного устройства.

Связь компьютера с периферийным устройством. Рассмотрим, как управляет периферийным устройством приложение, выполняемое на компьютере В, к которому данное ПУ подключено непосредственно (рис. 1.8).

1. Пусть приложению В в какой-то момент времени потребовалось вывести на печать некоторые данные. Для этого приложение обращается с запросом на выполнение операции ввода-вывода к операционной системе. В запросе указывается адрес данных, которые необходимо напечатать (адрес буфера оперативной памяти (ОП)), и информация о том, на каком периферийном устройстве эту операцию требуется выполнить.

2. Получив запрос, операционная система запускает программу – драйвер принтера. С этого момента все дальнейшие действия по выполнению операции ввода-вывода со стороны компьютера реализуются только драйвером принтера и работающим под его управлением аппаратным модулем – интерфейсной картой принтера без участия приложения и ОС.

3. Драйвер принтера оперирует командами, понятными контроллеру принтера, такими, например, как «Печать символа», «Перевод каретки». Драйвер в определенной последовательности загружает коды этих команд, а также данные, взятые из буфера ОП, в буфер интерфейсной карты принтера, которая побайтно передает их по сети контроллеру принтера.

4. Интерфейсная карта выполняет низкоуровневую работу, не вдаваясь в детали управления устройством, смысла данных и команд, передаваемых ей драйвером, считая их однородным потоком байтов. После получения от драйвера очередного байта интерфейсная карта просто последовательно передает биты в линию связи, представляя каждый бит электрическим сигналом. Чтобы контроллеру принтера стало понятно, что начинается передача байта, перед передачей первого бита информационная карта формирует стартовый сигнал специфической формы, а после передачи последнего информационного бита – стоповый сигнал. Эти сигналы синхронизируют передачу байта. Контроллер, опознав стартовый бит, начинает принимать информационные биты, формируя из них байт в своем приемном буфере.

5. Получив очередной байт, контроллер интерпретирует его и запускает заданную операцию принтера. Закончив работу по печати всех символов документа, драйвер принтера сообщает операционной системе о выполнении запроса, а та, в свою очередь, сигнализирует об этом событии приложению.

Обмен данными между двумя компьютерами. Механизмы взаимодействия компьютеров в сети многое позаимствовали у схемы взаимодействия компьютера с периферийными устройствами. В самом простом случае связь компьютеров может быть реализована с помощью тех же самых средств, которые используются для связи компьютера с периферией [2].

Приложения А и В (рис. 1.8) управляют процессом передачи данных путем обмена *сообщениями*. Чтобы приложения могли «понимать» получаемую друг от друга информацию, программисты, разрабатывающие эти приложения, должны *строго оговорить* форматы и последовательность сообщений, которыми приложения будут обмениваться во время выполнения

этой операции. Например, они могут договориться о том, что любая операция обмена данными начинается с передачи сообщения, запрашивающего информацию о готовности приложения В; что в следующем сообщении идут идентификаторы компьютера и пользователя, сделавшего запрос и т. п. Тем самым определяется *протокол взаимодействия приложений*.

Аналогично тому, как при выводе данных на печать необходимо передавать принтеру дополнительно некоторый объем служебной информации – в виде команд управления принтером, для передачи данных из одного компьютера в другой необходимо сопровождать эти данные дополнительной информацией в виде протокольных сообщений, которыми обмениваются приложения.

На стороне компьютера А приложение размещает в буфере ОП либо собственное очередное сообщение, либо данные и обращается к ОС с запросом на выполнение операции межкомпьютерного обмена данными. ОС запускает соответствующий драйвер сетевой карты, который загружает байт из буфера ОП в буфер интерфейсной карты и инициирует ее работу. Сетевая интерфейсная карта последовательно передает биты в линию связи, дополняя каждый новый байт стартовым и стоповым битами.

На стороне компьютера Б сетевая интерфейсная карта принимает биты, поступающие со стороны внешнего интерфейса, и помещает их в собственный буфер. После того как получен стоповый бит, интерфейсная карта устанавливает признак завершения приема байта и выполняет проверку корректности приема, например, путем контроля бита четности. Факт корректного приема байта фиксируется драйвером сетевой интерфейсной карты компьютера В. Драйвер переписывает принятый байт из буфера интерфейсной карты в заранее зарезервированный буфер ОП компьютера В. Приложение В извлекает данные из буфера и интерпретирует их в соответствии со своим протоколом либо как сообщение, либо как данные.

Таким образом, связав электрически и информационно два автономно работающих компьютера, мы получили простейшую *компьютерную сеть*.

Доступ к периферийным устройствам через сеть. Итак, мы имеем в своем распоряжении механизм обмена данными приложениями, выполняющимся на разных компьютерах. И хотя приложение А (рис. 1.8) по-прежнему не может управлять принтером, подключенным к компьютеру В, оно может теперь воспользоваться средствами межкомпьютерного обмена, чтобы передать приложению В «просьбу» выполнить для него требуемую операцию. Приложение А должно «объяснить» приложению В, какую операцию необходимо выполнить, с какими данными, на каком из имеющихся в его распоряжении устройств, в каком виде должен быть распечатан текст и т. п. В ходе печати могут возникнуть ситуации, о которых приложение В должно оповестить приложение А, например,

об отсутствии бумаги в принтере, т. е. для решения поставленной задачи – доступа к принтеру по сети – должен быть разработан специальный протокол взаимодействия приложений А и В.

А теперь рассмотрим, как работают вместе все элементы этой простейшей компьютерной сети при совместном использовании принтера.

1. В соответствии с принятым протоколом приложение А формирует сообщение-запрос к приложению В, помещает его в буфер ОП компьютера А и обращается к ОС, снабжая ее необходимой информацией.

2. ОС запускает драйвер сетевой интерфейсной карты, сообщая ему адрес буфера ОП, где хранится сообщение.

3. Драйвер и сетевая интерфейсная карта компьютера А, взаимодействуя с драйвером и интерфейсной картой компьютера В, передают сообщение байт за байтом в буфер ОП компьютера В.

4. Приложение В извлекает сообщение из буфера, интерпретирует его в соответствии с протоколом и выполняет необходимые действия. В число таких действий входит в том числе обращение к ОС с запросом на выполнение тех или иных операций с локальным принтером.

5. ОС запускает драйвер принтера, который вместе с интерфейсной картой и контроллером принтера выполняет требуемую операцию печати.

Уже на этом начальном этапе, рассматривая связь компьютера с периферийным устройством, мы столкнулись с важнейшими «сетевыми» понятиями: интерфейсом и протоколом, драйвером и интерфейсной картой, а также с проблемами, характерными для компьютерных сетей: согласованием интерфейсов, синхронизацией асинхронных процессов.

1.2.2. Сетевое программное обеспечение

Мы только что рассмотрели случай совместного использования принтера в простейшей сети из двух компьютеров и уже на этом начальном этапе можем ввести понятия сетевого программного обеспечения: сетевых служб, сетевой операционной системы и сетевых приложений.

Сетевые службы и сервисы. Потребность в доступе к удаленному принтеру может возникнуть у пользователей самых разных приложений: текстового редактора, графического редактора, системы управления базой данных (СУБД). Очевидно, что дублирование в каждом из приложений общих для всех них функций по организации удаленной печати является избыточным. Более эффективным представляется подход, при котором эти функции исключаются из приложений и оформляются в виде пары специализированных программных модулей – клиента и сервера печати (рис. 1.9), функции которых ранее выполнялись соответственно приложениями А и В. Теперь эта пара клиент – сервер может быть использована любым приложением, выполняемым на компьютере А [2].

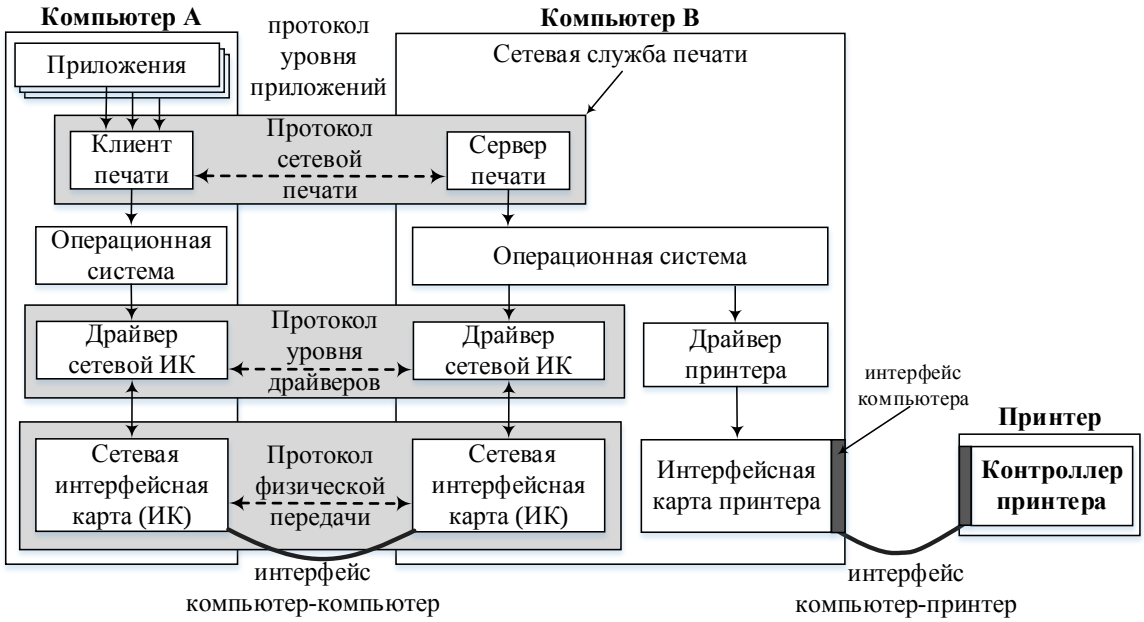


Рис. 1.9. Совместное использование принтера
с помощью сетевой службы печати

Клиент – это модуль, предназначенный для формирования и передачи сообщений-запросов к ресурсам удаленного компьютера от разных приложений с последующим приемом результатов из сети и передачей их соответствующим приложениям.

Сервер – это модуль, который постоянно ожидает прихода из сети запросов от клиентов и, приняв запрос, пытается его обслужить, как правило, с участием локальной ОС; один сервер может обслуживать запросы сразу нескольких клиентов (поочередно или одновременно).

Пара клиент-сервер, предоставляющая доступ к конкретному типу ресурса компьютера через сеть, образует сетевую службу.

Каждая служба связана с определенным типом сетевых ресурсов. Так, на рис. 1.9 модули клиента и сервера, реализующие удаленный доступ к принтеру, образуют сетевую службу печати.

Услуги, предоставляемые службой, называются сервисом. Служба может предоставить сервис как одного, так и нескольких типов.

Для поиска и просмотра информации в Интернете используется веб-служба, состоящая из веб-сервера и клиентской программы, называемой веб-браузером (web browser). На схеме веб-службы, показанной на рис. 1.10, два компьютера связаны не непосредственно, как это было во всех предыдущих примерах, а через множество промежуточных компьютеров и других сетевых устройств, входящих в состав Интернета. Для того чтобы отразить этот факт графически, мы поместили между двумя компьютерами так называемое *коммуникационное облако*, которое позволяет нам абстрагироваться от всех деталей среды передачи сообщений [2].

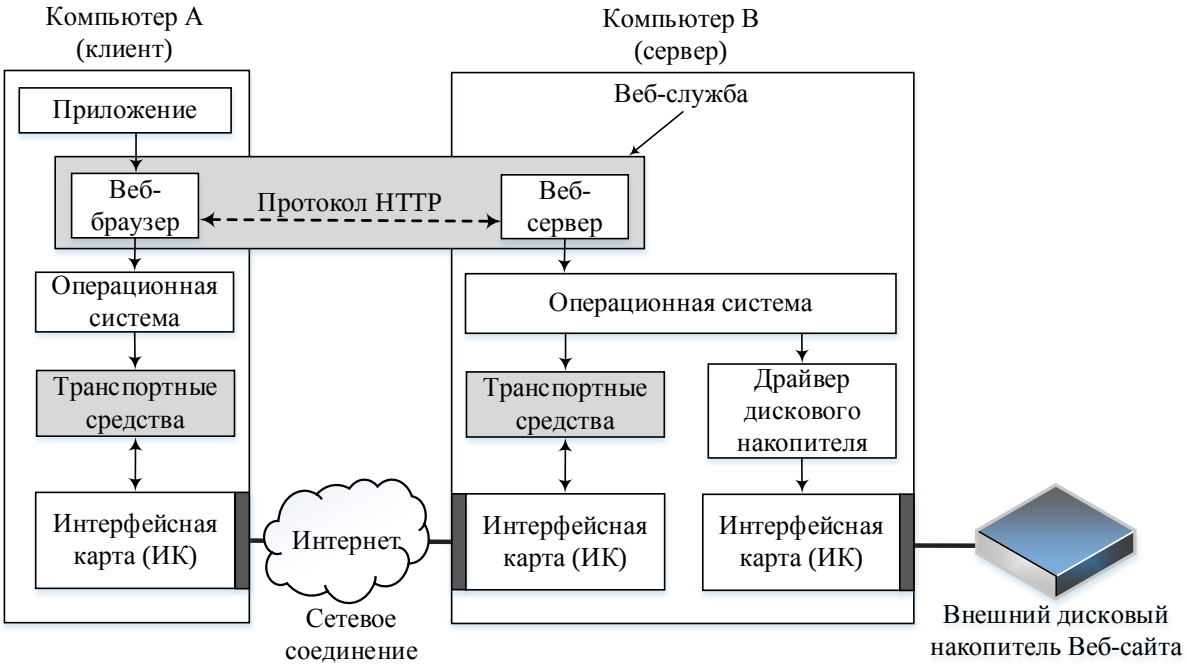


Рис. 1.10. Веб-служба

Обмен сообщениями между клиентской и серверной частями веб-службы выполняется по стандартному протоколу HTTP и никак не зависит от того, передаются ли эти сообщения «из рук в руки» (от интерфейса одного компьютера к интерфейсу другого) или через большое число посредников – транзитных коммуникационных устройств. Вместе с тем усложнение среды передачи сообщений приводит к возникновению новых дополнительных задач, на решение которых не был рассчитан упоминавшийся ранее драйвер сетевой информационной карты. Вместо него на компьютерах должны быть установлены более развитые программные *транспортные средства*.

Сетевая операционная система. *Операционную систему компьютера* часто определяют, как взаимосвязанный набор системных программ, который обеспечивает эффективное управление ресурсами компьютера (памятью, процессором, внешними устройствами, файлами и др.), а также предоставляет пользователю удобный интерфейс для работы с аппаратурой компьютера и разработки приложений.

Говоря о *сетевой ОС*, мы, очевидно, должны расширить границы управляемых ресурсов за пределы одного компьютера. *Сетевой операционной системой называют операционную систему компьютера, которая помимо управления локальными ресурсами предоставляет пользователям и приложениям возможность эффективного и удобного доступа к информационным и аппаратным ресурсам других компьютеров сети.* Сегодня практически все операционные системы являются сетевыми.

Из примеров, рассмотренных в предыдущих пунктах (рис. 1.9 и 1.10), мы видим, что удаленный доступ к сетевым ресурсам обеспечивается:

а) сетевыми службами; б) средствами транспортировки сообщений по сети (в простейшем случае – сетевыми интерфейсными картами и их драйверами).

Следовательно, именно эти функциональные модули должны быть добавлены к ОС, чтобы она могла называться сетевой (рис. 1.11).

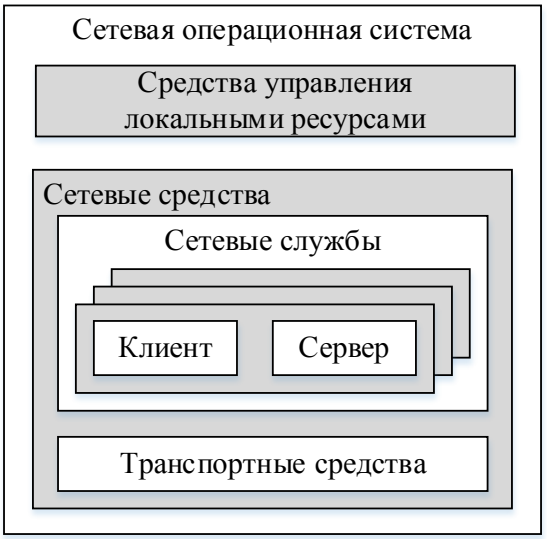


Рис. 1.11. Функциональные компоненты сетевой ОС

Помимо сетевых служб сетевая ОС должна включать *программные коммуникационные (транспортные) средства*, обеспечивающие совместно с аппаратными коммуникационными средствами передачу сообщений, которыми обмениваются клиентские и серверные части сетевых служб. Задачу коммуникации между компьютерами сети решают драйверы и протокольные модули. Они выполняют такие функции, как формирование сообщений, разбиение сообщения на части (пакеты, кадры), преобразование имен компьютеров в числовые адреса, дублирование сообщений в случае их потери, определение маршрута в сложной сети и т. д.

И сетевые службы, и транспортные средства могут являться неотъемлемыми (встроенными) компонентами ОС или существовать в виде отдельных программных продуктов. Например, сетевая файловая служба обычно встраивается в ОС, а вот веб-браузер чаще всего приобретается отдельно. Например, на основании антимонопольного закона США компании Microsoft было запрещено включать ее браузер Internet Explorer в состав ОС этой компании.

Сетевая служба может быть представлена в ОС либо обеими (клиентской и серверной) частями, либо только одной из них.

В первом случае операционная система, называемая *одноранговой*, не только позволяет обращаться к ресурсам других компьютеров, но и предоставляет собственные ресурсы в распоряжение пользователей других компьютеров. Например, если на всех компьютерах сети установлены и клиенты, и серверы файловой службы, то все пользователи сети могут совместно использовать файлы друг друга. Компьютеры, совмещающие функции клиента и сервера, называют *одноранговыми* узлами.

Операционная система, которая преимущественно содержит клиентские части сетевых служб, называется *клиентской*. Клиентские ОС устанавливаются на компьютеры, обращающиеся с запросами к ресурсам других компьютеров сети. За такими компьютерами, также называемыми клиентскими, работают рядовые пользователи.

К другому типу операционных систем относится *серверная ОС* – она ориентирована на обработку запросов из сети к ресурсам своего компьютера и включает в себя в основном серверные части сетевых служб. Компьютер с установленной на нем серверной ОС, занимающийся обслуживанием запросов других компьютеров, называют *выделенным сервером* сети.

На компьютере, подключенном к сети, могут запускаться **сетевые приложения** нескольких типов.

Локальное приложение целиком выполняется на данном компьютере и использует только локальные ресурсы (рис. 1.12, а). Для такого приложения не требуется никаких сетевых средств, оно может быть выполнено на автономно работающем компьютере.

Централизованное сетевое приложение целиком выполняется на данном компьютере, но обращается в процессе своей работы к ресурсам других компьютеров сети. В примере на рис. 1.12, б приложение, которое выполняется на клиентском компьютере, обрабатывает данные из файла, хранящегося на файл-сервере, а затем распечатывает результаты на принтере, подключенном к серверу печати. Работа такого типа приложений невозможна без участия сетевых служб и средств транспортировки сообщений.

Распределенное (сетевое) приложение состоит из нескольких взаимодействующих частей, каждая из которых выполняет какую-то определенную законченную работу по решению прикладной задачи, причем каждая часть может выполняться на отдельном компьютере сети (рис. 1.12, в). Части распределенного приложения взаимодействуют друг с другом, используя сетевые службы и транспортные средства ОС. Распределенное приложение в общем случае имеет доступ ко всем ресурсам компьютерной сети.

Очевидным преимуществом распределенных приложений является возможность распараллеливания вычислений, а также специализация компьютеров. Так, в приложении, предназначенном, скажем, для анализа климатических изменений, можно выделить три достаточно самостоятельные части (рис. 1.12, в), допускающие распараллеливание. Первая часть приложения, выполняющаяся на сравнительно маломощном персональном компьютере, могла бы поддерживать специализированный графический пользовательский интерфейс, вторая – заниматься статистической обработкой данных на высокопроизводительном мэйнфрейме, третья – генерировать отчеты на сервере с установленной стандартной СУБД. В общем случае каждая из частей распределенного приложения может быть представлена несколькими копиями, работающими на разных компьютерах.

Все сетевые службы, включая файловую службу, службу печати, службу электронной почты, службу удаленного доступа, интернет-телефонию и т. д., по определению относятся к классу распределенных приложений, так как любая сетевая служба включает в себя клиентскую и серверную части, которые могут выполняться на разных компьютерах.

Рис. 1.12. Типы приложений, выполняющихся в сети

На рис. 1.13, иллюстрирующем распределенный характер веб-службы, мы видим различные виды клиентских устройств – персональные компьютеры, ноутбуки и мобильные смартфоны – с установленными на них веб-браузерами, которые взаимодействуют по сети с веб-сервером. Таким образом, с одним и тем же веб-сайтом может одновременно работать множество – сотни и тысячи – сетевых пользователей [2].

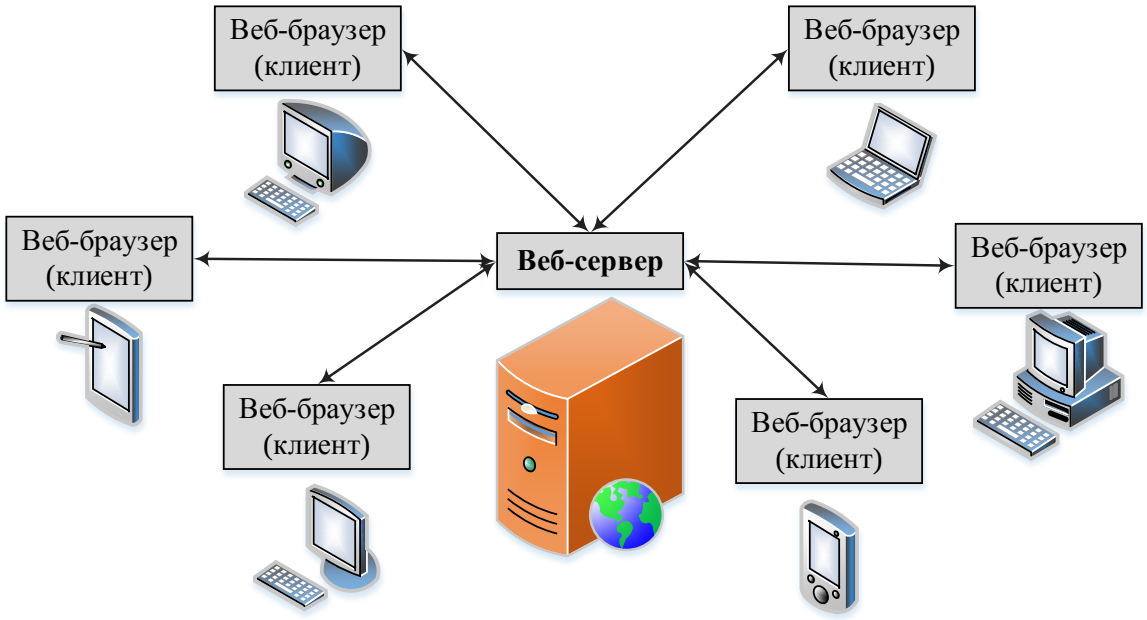


Рис. 1.13. Веб-служба как распределенное приложение

Многочисленные примеры распределенных приложений можно встретить и в такой области, как обработка данных научных экспериментов. Это неудивительно, так как многие эксперименты порождают такие большие объемы данных, генерируемых в реальном масштабе времени, которые просто невозможно обработать на одном, даже очень мощном, суперкомпьютере. Кроме того, алгоритмы обработки экспериментальных данных часто легко распараллеливаются. Одним из известных примеров распределенного научного приложения является программное обеспечение обработки данных большого андронного коллайдера (Large Hadron Collider, LHC), запущенного в 2008 г. в CERN, – это приложение работает более чем на 30 тысячах компьютеров, объединенных в сеть.

1.2.3. Характеристики физических каналов

В вычислительной технике для представления данных используется двоичный код. Внутри компьютера единицам и нулям данных соответствуют дискретные электрические сигналы.

Представление данных в виде электрических или оптических сигналов называется кодированием.

Линии связи между компьютерами отличаются от линий связи внутри компьютера гораздо большей протяженностью и подверженностью электромагнитных помех. Поэтому для надежного распознавания сигналов на приемном конце линии связи при передаче данных внутри и вне компьютера не всегда можно использовать одни и те же скорости и способы кодирования.

Еще одной проблемой, которую нужно решать при передаче сигналов, является проблема взаимной синхронизации передатчика одного компьютера с приемником другого. При организации взаимодействия модулей внутри компьютера эта проблема решается очень просто, так как в этом случае все модули синхронизируются от общего тактового генератора. Проблема синхронизации при связи компьютеров может решаться разными способами, как путем обмена специальными тактовыми синхроимпульсами по отдельной линии, так и путем периодической синхронизации заранее обусловленными кодами или импульсами характерной формы.

Несмотря на предпринимаемые меры (выбор соответствующей скорости обмена данными, линий связи с определенными характеристиками, способа синхронизации приемника и передатчика), существует вероятность искажения передаваемых данных. Для повышения надежности передачи данных между компьютерами, как правило, используется стандартный прием – подсчет *контрольной суммы* и передача полученного значения после некоторого блока байтов. Часто в протокол обмена данными включается как обязательный элемент *сигнал-квитанция*, который подтверждает правильность приема данных и посылается от получателя отправителю.

Приведем характеристики передачи данных через физические каналы.

Предложенная нагрузка – это поток данных, поступающий от пользователя на вход сети. Предложенную нагрузку можно характеризовать скоростью поступления данных в сеть в битах (килобитах и т. д.) в секунду.

Скорость передачи данных (*information rate, или throughput*) – это фактическая скорость потока данных, прошедшего через сеть. Эта скорость может быть меньше, чем скорость предложенной нагрузки, так как данные в сети могут искажаться или теряться.

Емкость канала связи (*capacity*), называемая также пропускной способностью, представляет собой максимальной возможную скорость передачи информации по каналу.

Полоса пропускания (*bandwidth*) – этот термин используется в двух значениях. Во-первых, с его помощью могут характеризовать *среду передачи*. В этом случае он означает ширину полосы частот, которую линия передает без существенных искажений. Во-вторых, термин «полоса пропускания» используется как синоним термина *емкость канала связи*. В первом случае полоса пропускания измеряется в герцах (Гц), во втором – в битах в секунду.

Еще одна группа характеристик канала связи связана с возможностью передачи информации по каналу в одну или обе стороны.

При взаимодействии двух компьютеров обычно требуется передавать информацию в обоих направлениях, от компьютера А к компьютеру В и обратно. Обычно существует основной поток данных, которые интересуют пользователя, и вспомогательный поток противоположного направления, который образуют квитанции о получении этих данных.

Физические каналы связи делятся на несколько типов в зависимости от того, могут они передавать информацию в обоих направлениях, или нет.

Дуплексный канал обеспечивает одновременную передачу информации в обоих направлениях. Дуплексный канал может состоять из двух физических сред, каждая из которых используется для передачи информации только в одном направлении. Возможен вариант, когда одна среда служит для одновременной передачи встречных потоков, в этом случае применяют дополнительные методы выделения каждого потока из суммарного сигнала.

Полудуплексный канал также обеспечивает передачу информации в обоих направлениях, но не одновременно, а по очереди. т. е. в течение определенного периода времени информация передается в одном направлении, а в течение следующего периода – в обратном.

Симплексный канал позволяет передавать информацию только в одном направлении. Часто дуплексный канал состоит из двух симплексных.

Выводы

Для того, чтобы пользователь сети получил возможность доступа к ресурсам «чужих» компьютеров, таким как диски, принтеры, плоттеры, необходимо дополнить все компьютеры сети специальными средствами.

В каждом компьютере функции передачи данных в линию связи совместно выполняют аппаратный модуль, называемый сетевым адаптером, или сетевой интерфейсной картой, и управляющая программа – драйвер. Задачи более высокого уровня – формирование запросов к ресурсам и их выполнение – решают собственно клиентские и серверные модули ОС.

Даже в простейшей сети, состоящей из двух компьютеров, возникают проблемы физической передачи сигналов по линиям связи: кодирование и модуляция, синхронизация передающего и принимающего устройств, контроль корректности данных.

Важными характеристиками, связанными с передачей трафика через физические каналы, являются: предложенная нагрузка, скорость передачи данных, пропускная способность, емкость канала связи, полоса пропускания.

Контрольные вопросы

1. Какие компоненты включает интерфейс устройства?
2. Какие задачи решает ОС при обмене с периферийным устройством?
3. Какие функции возлагаются на драйвер периферийного устройства?
4. Какие из перечисленных терминов в некотором контексте могут использоваться как синонимы? Варианты ответов:

а) емкость канала связи;

б) скорость передачи данных;

в) полоса пропускания канала связи;

г) пропускная способность канала.

5. Могут ли клиентская и серверная части приложений работать на одном и том же компьютере?

1.3. Организация связи нескольких компьютеров

Мы рассмотрели вырожденную сеть, состоящую всего из двух машин. При объединении в сеть большего числа компьютеров возникает целый комплекс новых проблем.

1.3.1. Топология физических связей

Объединяя в сеть несколько (больше двух) компьютеров, необходимо решить, каким образом соединить их друг с другом, другими словами, выбрать конфигурацию физических связей, или топологию.

Под топологией сети понимается конфигурация графа, вершинам которого соответствуют конечные узлы сети (например, компьютеры) и коммуникационное оборудование (например, маршрутизаторы), а ребрам – физические или информационные связи между вершинами [2].

Число возможных вариантов конфигурации резко возрастает при увеличении числа связываемых устройств. Так, если три компьютера мы можем связать двумя способами (рис. 1.14, а), то для четырех можно предложить уже шесть топологически разных конфигураций, что иллюстрирует рис. 1.14, б.

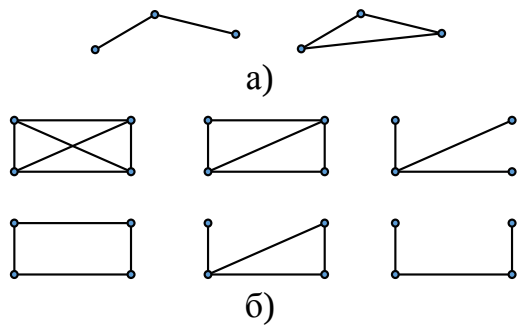


Рис. 1.14. Варианты связи компьютеров

Мы можем связывать каждый компьютер с каждым или же связывать их последовательно, предполагая, что они будут общаться, передавая сообщения друг другу «транзитом». Транзитные узлы должны быть оснащены специальными средствами, позволяющими им выполнять эту специфическую

посредническую операцию. В качестве транзитного узла может выступать как универсальный компьютер, так и специализированное устройство.

От выбора топологии связей существенно зависят характеристики сети. Например, наличие между узлами нескольких путей повышает надежность сети и делает возможным распределение загрузки между отдельными каналами. Простота присоединения новых узлов, свойственная некоторым топологиям, делает сеть легко расширяемой. Экономические соображения часто приводят к выбору топологий, для которых характерна минимальная суммарная длина линий связи.

Среди множества возможных конфигураций различают полносвязные и неполносвязные. *Полносвязная топология* соответствует сети, в которой каждый компьютер непосредственно связан со всеми остальными (рис. 1.15, а). Несмотря на логическую простоту, этот вариант оказывается громоздким и неэффективным. Действительно, в таком случае каждый компьютер в сети должен иметь большое количество коммуникационных портов, достаточное для связи с каждым из остальных компьютеров сети. Для каждой пары компьютеров должна быть выделена отдельная физическая линия связи. Полносвязные топологии в крупных сетях применяются редко, так как для связи N узлов требуется $N(N-1)/2$ физических дуплексных линий связи, т. е. имеет место квадратичная зависимость от числа узлов. Чаще этот вид топологии используется в многомашинных комплексах или сетях, объединяющих небольшое количество компьютеров [2].

Все другие варианты основаны на *неполносвязных топологиях*, когда для обмена данными между двумя компьютерами может потребоваться транзитная передача данных через другие узлы сети.

Ячеистая топология получается из полносвязной путем удаления некоторых связей (рис. 1.15, б). Ячеистая топология допускает соединение большого числа компьютеров и характерна, как правило, для крупных сетей.

В сетях с *кольцевой топологией* (рис. 1.15, в) данные передаются по кольцу от одного компьютера к другому. Главным достоинством кольца является то, что оно по своей природе обеспечивает резервирование связей. Действительно, любая пара узлов соединена здесь двумя путями – по часовой стрелке и против нее. Кроме того, кольцо представляет собой очень удобную конфигурацию для организации обратной связи – данные, сделав полный оборот, возвращаются к узлу-источнику. Поэтому источник может контролировать процесс доставки данных адресату. Часто это свойство кольца используется для тестирования связности сети и поиска узла, работающего некорректно. В то же время в сетях с кольцевой топологией необходимо принимать меры, чтобы в случае выхода из строя какого-либо компьютера не прерывался канал связи между остальными узлами кольца.

Звездообразная топология (рис. 1.15, г) образуется в случае, когда каждый компьютер подключается непосредственно к общему центральному

устройству, называемому *концентратором*[1]. В функции концентратора входит направление передаваемой компьютером информации одному или всем остальным компьютерам сети. В качестве концентратора может выступать как универсальный компьютер, так и специализированное устройство. К недостаткам звездообразной топологии относится более высокая стоимость сетевого оборудования из-за необходимости приобретения специализированного центрального устройства. Кроме того, возможности по наращиванию узлов в сети ограничиваются количеством портов концентратора.

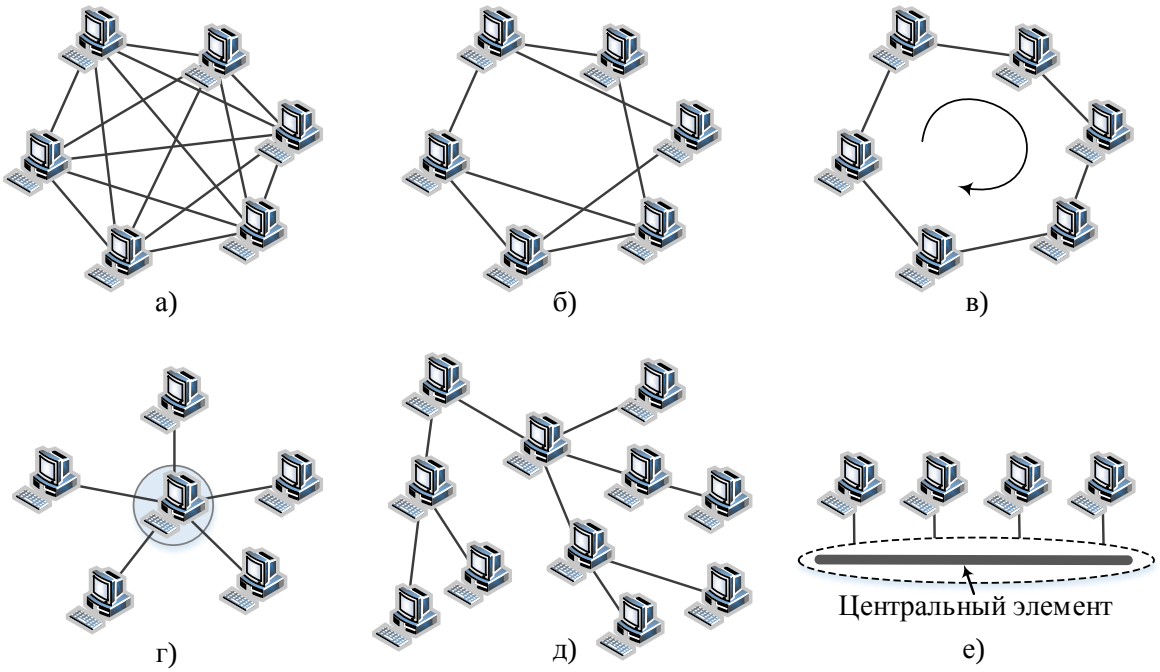


Рис. 1.15. Типовые топологии сетей

Иногда имеет смысл строить сеть с использованием нескольких концентраторов, иерархически соединенных между собой звездообразными связями (рис. 1.15, д). Получаемую в результате структуру называют *иерархической звездой* или *деревом*. В настоящее время дерево является самой распространенной топологией связей как в локальных, так и глобальных сетях.

Особым частным случаем звезды является *общая шина* (центральный элемент) (рис. 1.15, е). Здесь в качестве центрального элемента выступает пассивный кабель, к которому по схеме «монтажного ИЛИ» подключается несколько компьютеров (такую же топологию имеют многие сети, использующие беспроводную связь, – роль общей шины здесь играет общая радиосреда). Передаваемая информация распространяется по кабелю и доступна одновременно всем компьютерам, присоединенным к этому кабелю.

[1] Здесь термин «концентратор» используется в широком смысле, обозначая любое многовходовое устройство, способное служить центральным элементом, например, коммутатор или маршрутизатор.

Основными преимуществами такой схемы являются ее дешевизна и простота присоединения новых узлов к сети, а недостатками – низкая надежность (любой дефект кабеля полностью парализует всю сеть) и невысокая производительность (в каждый момент времени только один компьютер может передавать данные по сети, поэтому пропускная способность делится здесь между всеми узлами сети).

Если небольшие сети, как правило, имеют типовую топологию – звезда, кольцо или общая шина, для крупных сетей характерно наличие произвольных связей между компьютерами. В таких сетях можно выделить отдельные произвольно связанные фрагменты (подсети), имеющие типовую топологию, поэтому их называют сетями со смешанной топологией (рис. 1.16) [2].

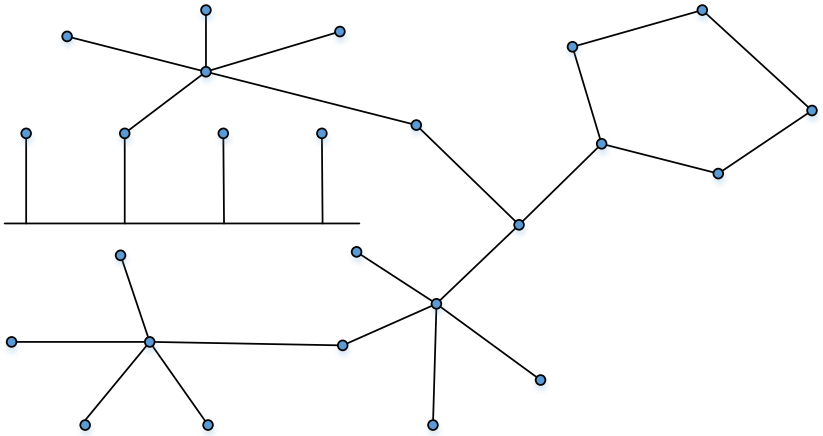


Рис. 1.16. Смешанная топология

1.3.2. Адресация узлов сети

Еще одной проблемой, которую нужно учитывать при объединении трех и более компьютеров, является проблема их адресации, точнее, адресации их сетевых интерфейсов. Один компьютер может иметь несколько сетевых интерфейсов. Например, для создания полносвязной структуры из N компьютеров необходимо, чтобы у каждого из них был $(N - 1)$ интерфейс.

По количеству адресуемых интерфейсов адреса можно классифицировать следующим образом: а) *уникальный адрес (unicast)* используется для идентификации отдельных интерфейсов; б) *групповой адрес (multicast)* идентифицирует сразу несколько интерфейсов, поэтому данные, помеченные групповым адресом, доставляются каждому из узлов, входящих в группу; в) данные, направленные по *широковещательному адресу (broadcast)*, должны быть доставлены всем узлам сети; г) *адрес произвольной рассылки (anycast)*, определенный в новой версии протокола IPv6, так же, как и групповой адрес, задает группу адресов, однако данные, посланные по этому

адресу, доставляются не всем узлам данной группы, а только одному из них; выбор этого узла осуществляется в соответствии с некоторыми правилами предпочтения.

Адреса могут быть числовыми (например, 129.26.255.255 или 81.1a.ff.ff) и символьными (например, site.domen.ru, willi-winki) [2].

Символьные адреса (имена) предназначены для запоминания людьми и поэтому обычно несут смысловую нагрузку. Для работы в больших сетях символьное имя может иметь иерархическую структуру, например, ftp-arch1.ucl.ac.uk. Этот адрес говорит о том, что данный компьютер поддерживает ftp-архив в сети одного из колледжей Лондонского университета (University College London) и эта сеть относится к академической ветви (ac) Интернета Великобритании (United Kingdom – uk). При работе в пределах сети Лондонского университета такое длинное символьное имя явно избыточно, и вместо него можно пользоваться кратким символьным именем ftp-arch1. Хотя символьные имена удобны для людей, из-за перемещения формата и потенциально большой длины их передача по сети не экономична.

Множество всех адресов, которые являются допустимыми в рамках некоторой схемы адресации, называются адресным пространством. Адресное пространство может иметь плоскую или иерархическую организацию.

При *плоской организации* множество адресов никак не структурировано. Примером плоского числового адреса является *MAC-адрес*, предназначенный для однозначной идентификации сетевых интерфейсов в локальных сетях. Такой адрес обычно используется только аппаратурой, поэтому его стараются сделать по возможности компактным и записывают в виде двоичного или шестнадцатеричного числа, например, 0081005e24a8. При задании MAC-адресов не требуется выполнять никакой ручной работы, так как они обычно встраиваются компанией-изготовителем, поэтому их называют также *аппаратными адресами (hardware address)*. Использование плоских адресов является жестким решением – при замене аппаратуры, например, сетевого адаптера, изменяется и адрес сетевого интерфейса компьютера.

При иерархической организации адресное пространство структурируется в виде вложенных друг в друга подгрупп, последовательно сужая адресную область, в конце концов, определяют отдельный сетевой интерфейс. Например, в трехуровневой структуре адресного пространства адрес конечного узла может задаваться тремя составляющими: а) идентификатором группы (K), в которую входит данный узел; б) идентификатором подгруппы (L); в) идентификатором узла (n), определяющим его в подгруппе.

Иерархическая адресация во многих случаях оказывается более рациональной, чем плоская. В больших сетях, состоящих из многих тысяч узлов, использование плоских адресов приводит к большим издержкам – конечным узлам и коммуникационному оборудованию приходится оперировать таблицами адресов, состоящими из тысяч записей. В противоположность этому иерархическая система адресации позволяет при перемещении данных

до определенного момента пользоваться только старшей составляющей адреса (например, идентификатором группы *K*), затем для дальнейшей локализации адресата задействовать следующую по старшинству часть (*L*) и в конечном счете – младшую часть (*n*).

Типичными представителями иерархических числовых адресов являются сетевые IP-адреса. В них поддерживается двухуровневая иерархия, адрес делится на старшую часть (номер сети) и младшую (номер узла). Такое деление позволяет передавать сообщения между сетями только на основании номера сети, а номер узла требуется уже после доставки сообщения в нужную сеть, точно так же, как название улицы используется почтальоном только после того, как письмо доставлено в нужный город.

На практике обычно применяют сразу несколько схем адресации, так что сетевой интерфейс компьютера может одновременно иметь несколько адресов-имен. Каждый адрес задействуется в той ситуации, когда соответствующий вид адресации наиболее удобен. А для преобразования адресов из одного вида в другой используются специальные вспомогательные протоколы, которые называются *протоколами разрешения адресов*.

Пользователи адресуют компьютеры иерархическими символьными именами, которые автоматически заменяются в сообщениях, передаваемых по сети, иерархическими числовыми адресами. С помощью этих числовых адресов сообщения доставляются из одной сети в другую, а после доставки сообщения в сеть назначения вместо иерархического числового адреса используется плоский аппаратный адрес компьютера. Проблема установления соответствия между адресами различных типов может решаться как централизованными, так и распределенными средствами.

При *централизованном подходе* в сети выделяется один или несколько компьютеров (серверов имен), в которых хранится таблица соответствия имен различных типов, например, символьных имен и числовых адресов. Все остальные компьютеры обращаются к серверу имен с запросами, чтобы по символьному имени найти числовой номер необходимого компьютера.

При *распределенном подходе* каждый компьютер сам хранит все назначенные ему адреса разного типа. Тогда компьютер, которому необходимо определить по известному иерархическому числовому адресу некоторого компьютера его плоский аппаратный адрес, посылает в сеть широковещательный запрос. Все компьютеры сети сравнивают содержащийся в запросе адрес с собственным. Тот компьютер, у которого обнаружилось совпадение, посылает ответ, содержащий искомый аппаратный адрес. Такая схема использована в протоколе разрешения адресов (Address Resolution Protocol, ARP) стека TCP/IP [2].

Достоинство распределенного подхода состоит в том, что он позволяет отказаться от выделения специального компьютера в качестве сервера имен, который к тому же часто требует ручного задания таблицы соответ-

ствия адресов. Недостатком его является необходимость широковещательных сообщений, перегружающих сеть. Именно поэтому распределенных подход используется в небольших сетях, а централизованный – в больших.

До сих пор мы говорили об адресах сетевых интерфейсов, компьютеров и коммуникационных устройств, однако конечной целью данных, пересылаемых по сети, являются не сетевые интерфейсы или компьютеры, а выполняемые на этих устройствах программы – процессы. Поэтому в адресе назначения наряду с информацией, идентифицирующей интерфейс устройства, должен указываться адрес процесса, которому предназначены посылаемые по сети данные. Очевидно, что достаточно обеспечить уникальность адреса процесса в пределах компьютера. Примером адресов процессов являются номера портов TCP и UDP, используемые в стеке TCP/IP.

1.3.3. Коммутация

Пусть компьютеры физически связаны между собой в соответствии с некоторой топологией и выбрана система адресации. Остается нерешенным вопрос: каким образом передавать данные между конечными узлами? Особую сложность приобретает эта задача для неполносвязной топологии сети, когда обмен данными между произвольной парой конечных узлов должен идти в общем случае через транзитные узлы.

Соединение конечных узлов через сеть транзитных узлов называют коммутацией. Последовательность узлов, лежащих на пути от отправителя к получателю, образует маршрут.

Например, в сети, показанной на рис. 1.17, узлы 2 и 4, непосредственно между собой не связанные, вынуждены передавать данные через транзитные узлы, в качестве которых могут выступить, например, узлы 1 и 5. Узел 1 должен выполнить передачу данных между своими интерфейсами *A* и *B*, а узел 5 – между интерфейсами *F* и *B*. В данном случае маршрутом является последовательность: 2-1-5-4, где 2 – узел-отправитель, 1 и 5 – транзитные узлы, 4 – узел получатель [2].

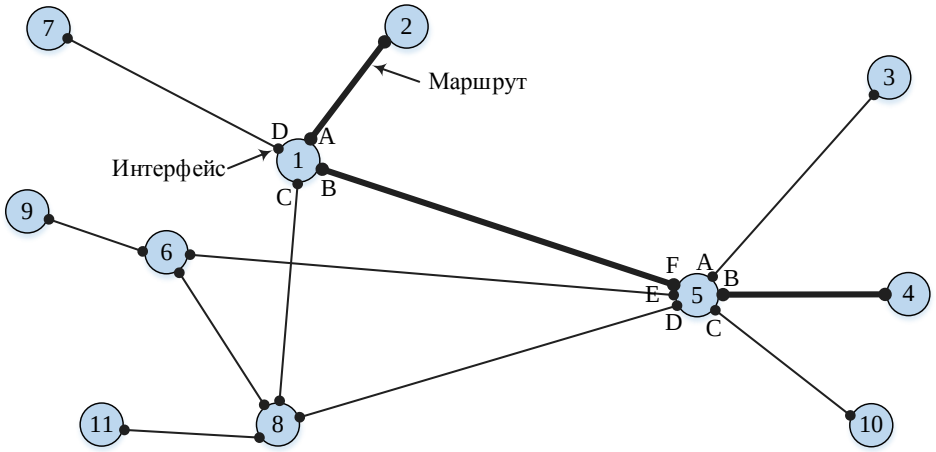


Рис. 1.17. Коммутация абонентов через сеть транзитных узлов

В самом общем виде **задача коммутации** может быть представлена в виде следующих взаимосвязанных задач.

1. *Определение информационных потоков,* для которых требуется прокладывать маршруты.

2. *Маршрутизация потоков*, т. е. фиксация маршрутов в конфигурационных параметрах и таблицах сетевых устройств).

3. *Продвижение потоков*, т. е. распознавание потоков и их локальная коммутация на каждом транзитном узле.

4. *Мультиплексирование и демультиплексирование потоков.*

5. *Разделение среды передачи.*

Среди множества возможных подходов к решению задачи коммутации абонентов в сетях выделяют два основополагающих, к которым относят коммутацию каналов и коммутацию пакетов.

Каждый из этих подходов имеет свои достоинства и недостатки. Существуют традиционные области применения каждой из техник коммутации, например, телефонные сети строились и продолжают строиться с использованием техники коммутации каналов, а компьютерные сети основаны на технике коммутации пакетов. Техника коммутации пакетов гораздо моложе своей конкурентки и пытается вытеснить ее из некоторых областей, например, из телефонии (в форме интернет- или IP-телефонии), но этот спор пока не решен, и скорее всего, две техники коммутации будут сосуществовать еще долгое время, дополняя друг друга. Тем не менее, по долгосрочным прогнозам, будущее принадлежит технике коммутации пакетов.

1.3.4. Определение информационных потоков

Через один транзитный узел может проходить несколько маршрутов, например, через узел 5 (рис. 1.17) проходят как минимум все данные, направляемые узлом 4 каждому из остальных узлов, а также все данные, поступающие в узлы 3, 4 и 10. Транзитный узел должен уметь *распознавать* поступающие на него потоки данных, чтобы обеспечить передачу каждого из них именно на тот интерфейс, который ведет к нужному узлу, и, возможно, чтобы выбрать специфический для данного потока способ его обработки.

Информационным потоком, или потоком данных, называют непрерывную последовательность данных, объединенных набором общих признаков, выделяющих эти данные из общего сетевого трафика [2].

Например, как поток можно определить все данные, поступающие от одного компьютера; объединяющим признаком в данном случае служит адрес источника. Эти же данные можно представить, как совокупность нескольких *подпотоков*, каждый из которых в качестве дифференцирующего признака имеет адрес назначения. Наконец, каждый из этих подпотоков,

в свою очередь, можно разделить на более мелкие подпотоки, порожденные разными сетевыми приложениями – электронной почтой, программой копирования файлов, веб-сервером. Данные, образующие поток, могут быть представлены в виде различных единиц данных – пакетов, кадров, ячеек.

При коммутации в качестве обязательного признака выступает адрес назначения данных. На основании этого признака весь поток входящих в транзитный узел данных разделяется на подпотоки, каждый из которых передается на интерфейс, соответствующий маршруту продвижения данных.

Адреса источника и назначения определяют поток для пары соответствующих конечных узлов. Однако часто бывает полезно представить этот поток в виде нескольких подпотоков, причем для каждого из них может быть проложен свой маршрут. Так, для приложений, которые предъявляют к сети свои особые требования, выбор маршрута должен осуществляться с учетом характера передаваемых данных, например, для файлового сервера важно, чтобы передаваемые им большие объемы данных направлялись по каналам, обладающим высокой пропускной способностью, а для программной системы управления, которая посылает в сеть короткие сообщения, требующие обязательной и немедленной отработки, при выборе маршрута более важна надежность линии связи и минимальный уровень задержек на маршруте.

Важная и обратная по отношению к выделению подпотоков операция – *агрегирование потоков*. Обычно она выполняется на магистралях сетей. Агрегирование потоков, имеющих общую часть маршрута, позволяет уменьшить количество хранимой промежуточными узлами информации, так как агрегированные потоки описываются как одно целое. В результате снижается нагрузка на промежуточные узлы и повышается их быстродействие.

Признаки потока могут иметь *глобальное* или *локальное* значение – в первом случае они однозначно определяют поток в пределах всей сети, а во втором – в пределах одного транзитного узла. Пара идентифицирующих поток адресов конечных узлов – это пример глобального признака. Примером признака, локально определяющего поток в пределах устройства, может служить номер (идентификатор) интерфейса данного устройства, на который поступили данные. Например, возвращаясь к рис. 1.17, узел 1 может быть настроен так, чтобы передавать на интерфейс *B* все данные, поступившие с интерфейса *A*, а на интерфейс *C* – данные, поступившие с интерфейса *D*. Такое правило позволяет отделить поток данных узла 2 от потока данных узла 7 и направлять их для транзитной передачи через разные узлы сети, в данном случае поток узла 2 – через узел 5, а поток узла 7 – через узел 8.

Метка потока – это особый тип признака. Она представляет собой некоторое число, которое несут все данные потока. *Глобальная метка*

назначается данным потока и не меняет своего значения на всем протяжении его пути следования от узла источника до узла назначения, таким образом, она уникально определяет поток в пределах сети [2].

Таким образом, распознавание потоков во время коммутации происходит на основании признаков, в качестве которых помимо обязательного адреса назначения данных могут выступать и другие признаки, такие, например, как идентификаторы приложений.

1.3.5. Маршрутизация

Задача маршрутизации в свою очередь включает в себя две подзадачи: а) определение маршрута; б) оповещение сети о выбранном маршруте.

Определить маршрут означает выбрать последовательность транзитных узлов и их интерфейсов, через которые надо передавать данные, чтобы доставить их адресату. Определение маршрута – сложная задача, особенно когда конфигурация сети такова, что между парой взаимодействующих сетевых интерфейсов существует множество путей. Чаще всего выбор останавливают на одном *оптимальном* по некоторому критерию маршруте. В качестве критериев оптимальности могут выступать, например, номинальная пропускная способность и загруженность каналов связи; задержки, вносимые каналами; количество промежуточных транзитных узлов; надежность каналов и транзитных узлов. Но даже в том случае, когда между конечными узлами существует только один путь, при сложной топологии сети его нахождение может представлять собой нетривиальную задачу.

Маршрут может определяться эмпирически («вручную») администратором сети на основании различных, часто не формализуемых соображений. Среди побудительных мотивов выбора пути могут быть: особые требования к сети со стороны различных типов приложений, решение передавать трафик через сеть определенного поставщика услуг, предположения о пиковых нагрузках на некоторые каналы сети, соображения безопасности.

Однако эмпирический подход к определению маршрута мало пригоден для большой сети со сложной топологией. В этом случае используются автоматические методы определения маршрутов. Для этого конечные узлы и другие устройства сети оснащаются специальными программными средствами, которые организуют взаимный обмен служебными сообщениями, позволяющими каждому узлу составить свое «представление» о сети. Затем на основе собранных данных определяются рациональные маршруты.

При выборе маршрута часто ограничиваются только информацией о топологии сети. Этот подход иллюстрирует рис. 1.18. Для передачи трафика между конечными узлами *A* и *C* существуют два альтернативных

маршрута: *A*-1-2-3-*C* и *A*-1-3-*C*. Если мы учитываем только топологию, то выбор очевиден – маршрут *A*-1-3-*C*, который имеет меньше транзитных узлов.

Решение было найдено путем минимизации критерия, в качестве которого в данном примере выступала длина маршрута, измеренная количеством транзитных узлов. Однако, возможно, наш выбор был не самым лучшим. На рис. 1.18 показано, что каналы 1-2 и 2-3 обладают пропускной способностью 100 Мбит/с, а канал 1-3 – только 10 Мбит/с. Если мы хотим, чтобы наша информация передавалась по сети с максимально возможной скоростью, то нам следовало бы выбрать маршрут *A*-1-2-3-*C*, хотя он и проходит через большее количество промежуточных узлов. т. е. можно сказать, что маршрут *A*-1-2-3-*C* в данном случае оказывается «более коротким».

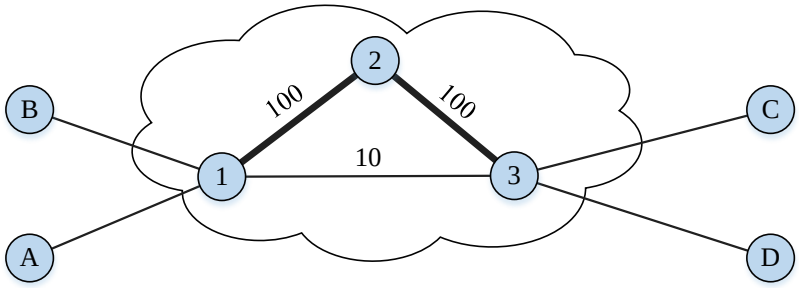


Рис. 1.18. Выбор маршрута

Абстрактная оценка условного «расстояния» между двумя узлами сети называется *метрикой*. Так, для измерения длины маршрута могут быть использованы разные метрики – количество транзитных узлов, как в предыдущем примере, линейная протяженность маршрута и даже его стоимость в денежном выражении. Для построения метрики, учитывающей пропускную способность, часто применяют следующий прием: длину каждого канала-участка характеризуют величиной, обратной его пропускной способности. Чтобы оперировать целыми числами, выбирают некоторую константу, заведомо большую, чем пропускные способности каналов в сети. Например, если мы в качестве такой константы выберем 100 Мбит/с, то метрика каждого из каналов 1-2 и 2-3 равна 1, а метрика канала 1-3 составляет 10. Метрика маршрута равна сумме метрик составляющих его каналов, поэтому часть пути 1-2-3 обладает метрикой 2, а альтернативная часть пути 1-3 – метрикой 10. Мы выбираем более «короткий» путь, т. е. путь *A*-1-2-3-*C*.

Описанные подходы не учитывают текущую загруженность каналов трафиком; методы, в которых используется информация о текущей загруженности, позволяют определять более рациональные маршруты, однако требуют интенсивного обмена служебной информацией между узлами сети. Используя аналогию с автомобильным трафиком, можно

сказать, что мы выбирали маршрут по карте, учитывая количество промежуточных городов и ширину дороги (аналог пропускной способности канала), отдавая предпочтение скоростным магистралям. Но мы не стали слушать радио- или телепрограмму, которая сообщает о текущих заторах. Так что наше решение оказывается отнюдь не лучшим, когда по маршруту *A*-1-2-3-*C* уже передается большое количество потоков, а маршрут *A*-1-3-*C* практически свободен.

После того как маршрут определен (вручную или автоматически), надо *оповестить* о нем все устройства сети. Сообщение о маршруте должно нести каждому транзитному устройству примерно такую информацию: «Каждый раз, когда в устройство поступят данные, относящиеся к потоку *n*, их следует передать для дальнейшего продвижения на интерфейс *if*1». Каждое подобное сообщение о маршруте обрабатывается транзитным устройством, в результате создается новая запись в *таблице коммутации* (называемой также *таблицей маршрутизации*). В этой таблице локальному или глобальному признаку (признакам) потока (например, метке, номеру входного интерфейса или адресу назначения) ставится в соответствие номер интерфейса, на который устройство должно передавать данные, относящиеся к этому потоку. Табл. 1.3 является фрагментом таблицы коммутации, содержащей запись, сделанную на основании сообщения о необходимости передачи потока *n* на интерфейс *F*.

В этой таблице в качестве признака потока использованы адрес назначения *DA*, адрес источника *SA* и тип приложения *A*, который генерирует пакеты потока. Детальное описание структуры сообщения о маршруте и содержимого таблицы коммутации зависит от конкретной технологии, однако эти особенности не меняют сущности рассматриваемых процессов. Чаще всего в качестве признака потока используется адрес назначения пакета.

Таблица 1.3

Фрагмент таблицы коммутации

Признаки потока	Направление передачи данных (номер интерфейса и/или адрес следующего узла)
...	...
$n = \{DA, SA, A\}$	*F*
...	...

Передача информации транзитным устройствам о выбранных маршрутах, так же, как и определение маршрута, может осуществляться вручную или автоматически. Администратор сети может зафиксировать маршрут, выполнив в ручном режиме конфигурирование устройства, например, жестко скоммутировав определенные пары входных и выходных интерфейсов

(как работали «телефонные барышни» на первых коммутаторах). Он может также внести запись о маршруте в таблицу коммутации.

Однако поскольку топология и состав информационных потоков могут меняться (отказы узлов или появление новых промежуточных узлов, изменение адресов или определение новых потоков), гибкое решение задач определения и задания маршрутов предполагает постоянный анализ состояния сети и обновление маршрутов и таблиц коммутации. В таких случаях задачи прокладки маршрутов, как правило, не могут быть решены без достаточно сложных программных и аппаратных средств.

1.3.6. Продвижение данных

Итак, пусть маршруты определены, записи о них сделаны в таблицах всех транзитных узлов, все готово к выполнению основной операции – передаче данных между абонентами (коммутации абонентов).

Для каждой пары абонентов эта операция может быть представлена несколькими (по числу транзитных узлов) *локальными* операциями коммутации. Прежде всего отправитель должен выставить данные на тот свой интерфейс, с которого начинается найденный маршрут, а все транзитные узлы должны соответствующим образом выполнить «переброску» данных с одного своего интерфейса на другой, другими словами, выполнить *коммутацию интерфейсов*. Устройство, функциональным назначением которого является коммутация, называется коммутатором. На рис. 1.19 показан коммутатор, который переключает потоки между четырьмя своими интерфейсами.

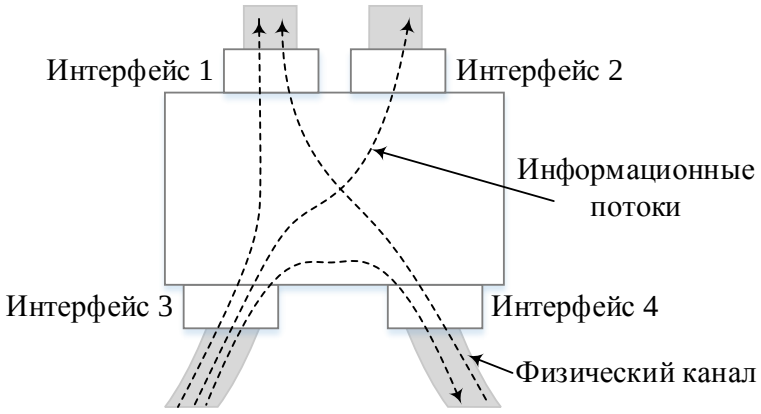


Рис. 1.19. Коммутатор

Прежде чем выполнить коммутацию, коммутатор должен распознать поток. Для этого в поступивших данных коммутатор пытается найти признак какого-либо из потоков, заданных в его таблице коммутации. Если произошло совпадение, то эти данные направляются на интерфейс, определенный для них в маршруте.

О терминах. Термины «коммутация», «таблица коммутации» и «коммутатор» в телекоммуникационных сетях могут трактоваться неоднозначно. Мы уже определили коммутацию как процесс соединения абонентов сети через транзитные узлы. Этим же термином мы обозначаем и соединение интерфейсов в пределах отдельного транзитного узла. Коммутатором в широком смысле слова называется устройство любого типа, способное выполнять операции переключения потока данных с одного интерфейса на другой. Операция коммутации может выполняться в соответствии с различными правилами и алгоритмами. Некоторые способы коммутации и соответствующие им таблицы и устройства получили специальные названия. Например, в технологии IP для обозначения аналогичных понятий используются термины «маршрутизация», «таблица маршрутизации», «маршрутизатор». В то же время за другими специальными типами коммутации и соответствующими устройствами закрепились те же самые названия «коммутация», «таблица коммутации» и «коммутатор», применяемые в узком смысле, например, как коммутация и коммутатор в локальной сети Ethernet. Для телефонных сетей, которые появились намного раньше компьютерных, также характерна аналогичная терминология, «коммутатор» является здесь синонимом «телефонной станции».

Коммутатором может быть как специализированное устройство, так и универсальный компьютер со встроенным программным механизмом коммутации. Компьютер может совмещать функции коммутации данных с выполнением своих обычных функций как конечного узла. Однако во многих случаях более рациональным является решение, в соответствии с которым некоторые узлы в сети выделяются *специально* для коммутации. Эти узлы образуют коммутационную сеть, к которой подключаются все остальные. На рис. 1.20 показана коммутационная сеть, образованная из узлов 1, 5, 6 и 8, к которой подключаются конечные узлы 2, 3, 4, 7, 9, 10 [2].

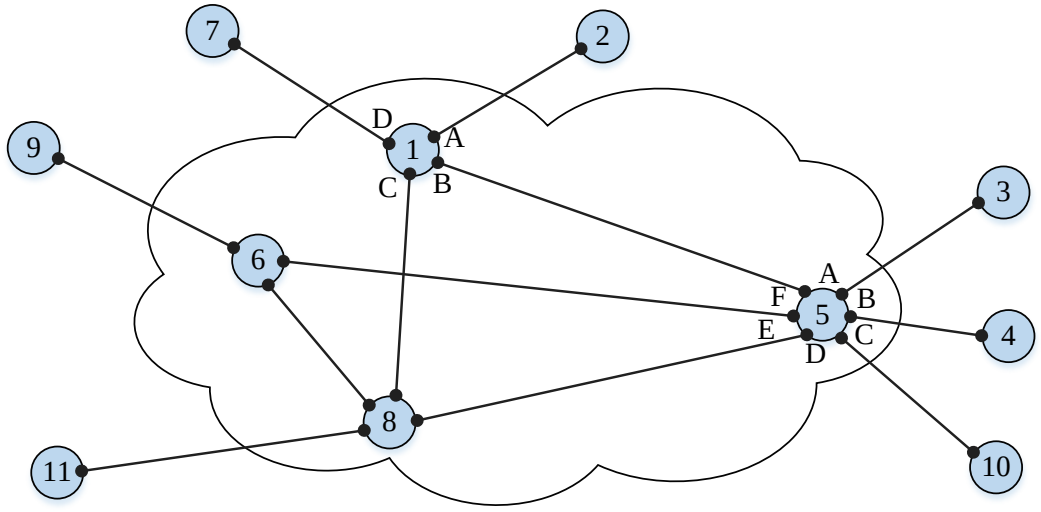


Рис. 1.20. Коммутационная сеть

1.3.7. Мультиплексирование / демультиплексирование

Чтобы определить, на какой интерфейс следует передать поступившие данные, коммутатор должен выяснить, к какому потоку они относятся. Эта задача должна решаться независимо от того, поступает на вход коммутатора только один «чистый» поток или «смешанный» поток, являющийся результатом агрегирования нескольких потоков. В последнем случае к задаче распознавания потоков добавляется задача демультиплексирования.

Демультиплексирование – разделение суммарного потока на несколько составляющих его потоков.

Мультиплексирование (агрегирование) – образование из нескольких отдельных потоков общего агрегированного потока, который передается по одному физическому каналу связи. Другими словами, мультиплексирование – это способ разделения одного имеющегося физического канала между несколькими одновременно протекающими сеансами связи абонентов сети.

Операции мультиплексирования/демультиплексирования имеют такое же важное значение в любой сети, как и операции коммутации, потому что без них пришлось бы для каждого потока предусматривать отдельный канал, что привело бы к большому количеству параллельных связей в сети и свело бы на нет все преимущества неполносвязной сети.

На рис. 1.21 показан фрагмент сети, состоящий из трех коммутаторов. Коммутатор 1 имеет четыре сетевых интерфейса. На интерфейс 1 поступают данные с двух интерфейсов – 3 и 4. Их надо передать в общий физический канал, т. е. выполнить операцию мультиплексирования [2].

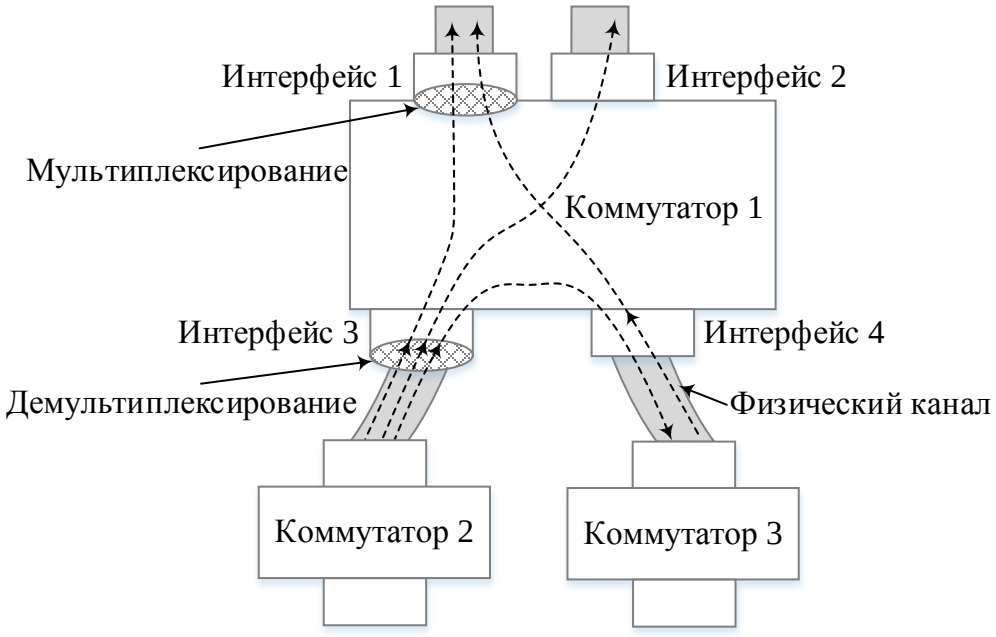


Рис. 1.21. Операции мультиплексирования
и демультиплексирования потоков при коммутации

Одним из основных способов мультиплексирования потоков является разделение времени. При этом способе каждый поток время от времени получает физический канал в полное свое распоряжение и передает по нему свои данные. Распространено также частотное разделение канала, когда каждый поток передает данные в выделенном ему частотном диапазоне.

Технология мультиплексирования должна позволять получателю такого суммарного потока выполнять обратную операцию – разделение (демультиплексирование) данных на слагаемые потоки. На интерфейсе 3 коммутатор выполняет демультиплексирование потока на три составляющих его подпотока. Один из них он передает на интерфейс 1, другой – на интерфейс 2, третий – на интерфейс 4.

Вообще говоря, на одном интерфейсе могут одновременно выполняться обе функции – мультиплексирование и демультиплексирование.

Частный случай коммутатора, у которого все входящие информационные потоки коммутируются на один выходной интерфейс, где они мультиплексируются в один агрегированный поток, называется мультиплексором. Коммутатор, который имеет один входной интерфейс и несколько выходных, называется демультиплексором (рис. 1.22)

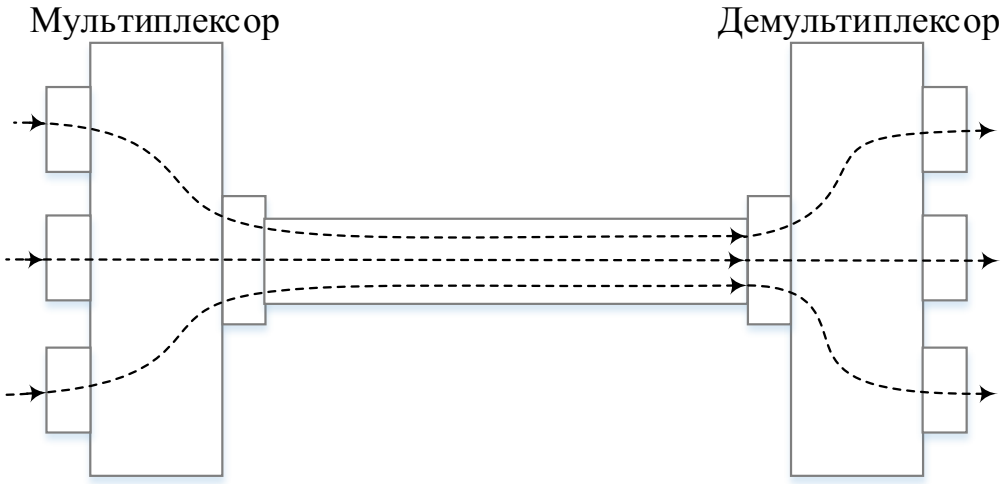


Рис. 1.22. Мультиплексор и демультиплексор

1.3.8. Разделяемая среда передачи

Во всех рассмотренных ранее примерах мультиплексирования потоков к каждой линии связи подключались только два интерфейса. В том случае, когда линия связи является дуплексным каналом связи, как это показано на рис. 1.23, каждый из интерфейсов монопольно использует канал связи в направлении «от себя». Это объясняется тем, что дуплексный канал состоит из двух независимых сред передачи данных (подканалов), и так как только передатчик интерфейса является активным устройством, а прием-

ник пассивно ожидает поступления сигналов от передатчика, то конкуренции подканалов не возникает. Такой режим использования среды передачи данных является в настоящее время основным в компьютерных сетях.

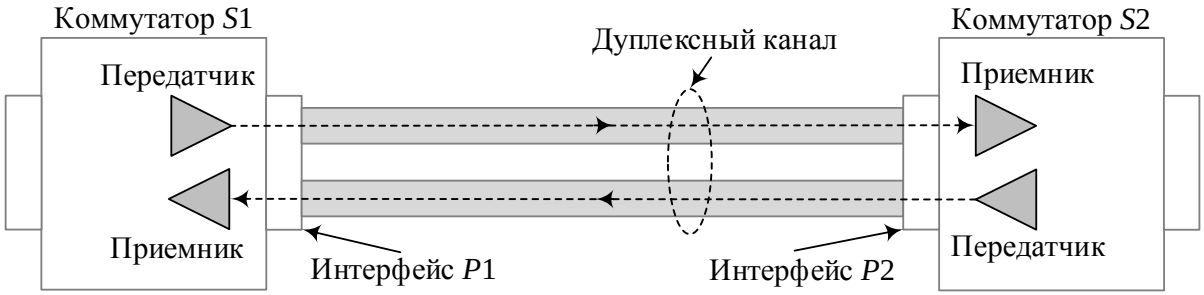


Рис. 1.23. Дуплексный канал – разделяемая среда отсутствует

Однако если в глобальных сетях такой режим использовался всегда, то в локальных сетях до середины 90-х гг. преобладал другой режим, основанный на разделяемой среде передачи данных. *Разделяемой средой (shared medium)* называется физическая среда передачи данных, к которой непосредственно подключено несколько передатчиков узлов сети. Причем в каждый момент времени только один из передатчиков какого-либо узла сети получает доступ к разделяемой среде и использует ее для передачи данных приемнику другого узла, подключенному к этой же среде.

В наиболее простом случае эффект разделения среды возникает при соединении двух интерфейсов с помощью полудуплексного канала связи, т. е. такого канала, который может передавать данные в любом направлении, но только попеременно (рис. 1.24). В этом случае к одной и той же среде передачи данных (например, к коаксиальному кабелю или общей радиосреде) подключены два приемника двух независимых узлов сети.

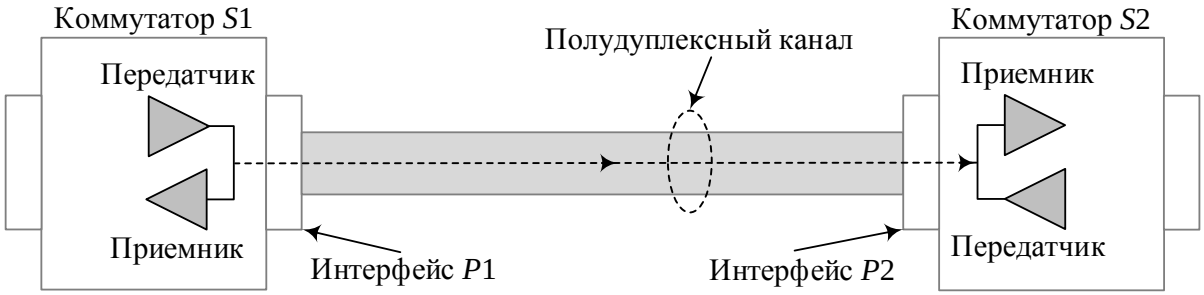


Рис. 1.24. Полудуплексный канал – разделяемая среда

При таком применении среды передачи данных возникает новая задача *совместного использования среды независимыми передатчиками* так, чтобы в каждый отдельный момент времени по среде передавались данные только одного передатчика. Другими словами, возникает *необходимость в механизме синхронизации доступа интерфейсов к разделяемой среде*.

Обобщением разделяемой среды является случай, показанный на рис. 1.25, когда к каналу связи подключаются более двух интерфейсов (в приведенном примере – три), при этом применяется топология общей шины.

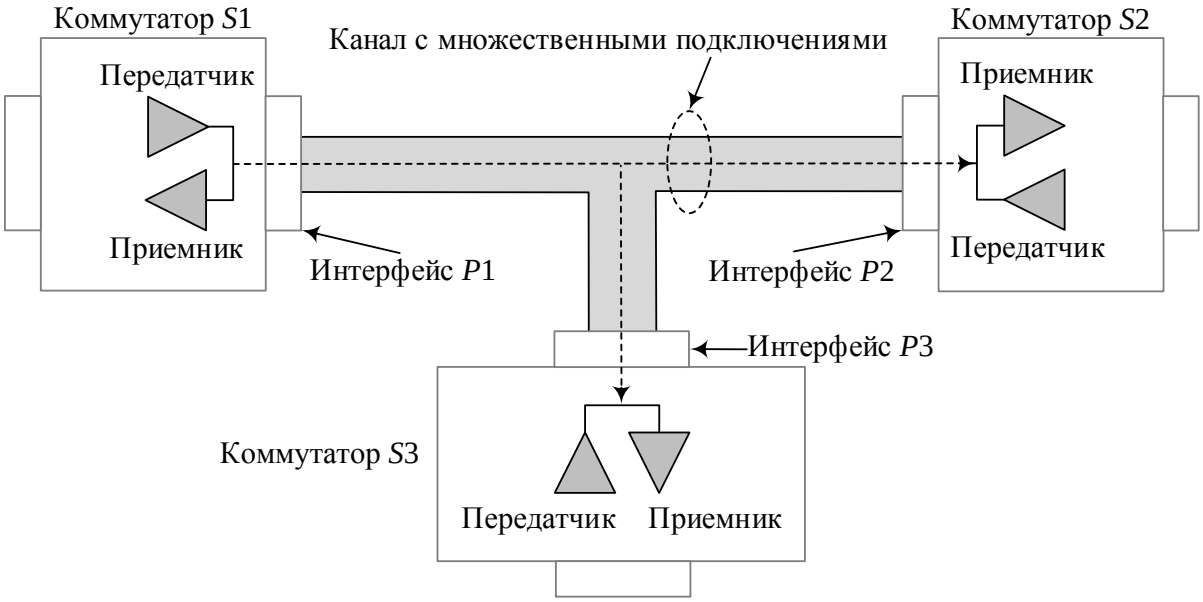


Рис. 1.25. Канал с множественными подключениями – разделяемая среда

Существуют различные способы организации совместного доступа к разделяемым линиям связи. Одни из них подразумевают *централизованный* подход, когда доступом к каналу управляет специальное устройство – *арбитр*, другие – *децентрализованный*. Если мы обратимся к организации работы компьютера, то увидим, что доступ к системной шине компьютера, которую совместно используют внутренние блоки компьютера, управляется централизованно – либо процессором, либо специальным арбитром шины.

В сетях организация совместного доступа к линиям связи имеет свою специфику из-за существенно большего времени распространения сигналов по линиям связи. Здесь процедуры согласования доступа к линиям связи могут занимать слишком большой промежуток времени и приводить к значительным потерям производительности сети. Именно по этой причине механизм разделения среды в глобальных сетях практически не используется [2].

На первый взгляд может показаться, что механизм разделения среды очень похож на механизм мультиплексирования потоков – в том и другом случае по линии связи передается несколько потоков данных. Однако здесь есть принципиальное различие, касающееся того, как контролируется (управляется) линия связи. При мультиплексировании дуплексная линия связи в каждом направлении находится под полным контролем одного коммутатора, который решает, какие потоки разделяют общий канал связи.

Для локальных сетей разделяемая среда сравнительно долго была основным механизмом использования каналов связи, который применялся

во всех технологиях локальных сетей – Ethernet, Token Ring, FDDI. При этом применялись децентрализованные методы доступа к среде, не требующие наличия арбитра в сети. Популярность техники разделения среды в локальных сетях объяснялась простотой и экономичностью аппаратных решений. Например, для создания сети Ethernet на коаксиальном кабеле никакого другого сетевого оборудования, кроме сетевых адаптеров компьютеров и самого кабеля, не требуется. Наращивание количества компьютеров в локальной сети Ethernet на коаксиальном кабеле выполняется также достаточно просто – путем присоединения нового отрезка кабеля к существующему.

Сегодня в проводных локальных сетях метод разделения среды практически перестал применяться. Основной причиной отказа от разделяемой среды явилась ее низкая плохо предсказуемая производительность, а также плохая масштабируемость[2]. Низкая производительность объясняется тем, что пропускная способность канала связи делится между всеми компьютерами сети. Например, если локальная сеть Ethernet состоит из 100 компьютеров, а для их связи используется коаксиальный кабель и сетевые адаптеры, работающие на скорости 10 Мбит/с, то в среднем на каждый компьютер приходится только 0,1 Мбит/с пропускной способности. Причина плохой масштабируемости в том, что чем больше компьютеров в такой сети, тем меньшая доля пропускной способности достается каждому компьютеру.

Описанные недостатки являются следствием самого принципа разделения среды, поэтому преодолеть их полностью невозможно. Появление в начале 90-х недорогих коммутаторов локальных сетей привело к настоящей революции в этой области, и постепенно коммутаторы вытеснили разделяемую среду полностью.

Сегодня механизм разделения среды используется только в беспроводных локальных сетях, где среда – *радиоэфир* – естественным образом соединяет все конечные узлы, находящиеся в зоне распространения сигнала.

Выводы

При связывании в сеть более двух компьютеров возникают проблемы выбора топологии (полносвязной, звезды, кольца, общей шины, иерархического дерева, произвольной); способа адресации (плоского или иерархического, числового или символьного); способа разделения линий связи и механизма коммутации.

В неполносвязных сетях соединение пользователей осуществляется путем коммутации через сеть транзитных узлов. При этом должны быть

[2] Масштабируемостью называют свойство сети допускать наращивание количества узлов и протяженность линий связи в очень широких пределах без снижения производительности.

решены следующие задачи: определение потоков данных и маршрутов для них, продвижение данных в каждом транзитном узле, мультиплексирование и демультиплексирование потоков.

Одной из основных проблем построения сетей является коммутация. Каждый узел, выполняющий транзитную передачу трафика, должен уметь его коммутировать, т. е. обеспечивать взаимодействие пользователей сети.

Среди множества возможных подходов к решению задачи коммутации выделяют два основополагающих – коммутацию каналов и пакетов.

Контрольные вопросы

1. Дайте определение понятию «топология».

2. К какому типу топологии можно отнести структуру, образованную тремя связанными друг с другом узлами в виде треугольника?

3. К какому типу топологии можно отнести структуру, образованную четырьмя связанными друг с другом узлами в виде квадрата?

4. К какому типу топологии можно отнести структуру, образованную тремя последовательно соединенными друг с другом узлами (последний не связан с первым)?

5. Частным случаем какой топологии является общая шина? Варианты ответов:

а) полносвязная;

б) кольцо;

в) звезда.

6. Какая из известных топологий обладает повышенной надежностью?

7. Какой тип топологии наиболее распространен сегодня в локальных сетях?

8. Какие требования предъявляются к системе адресации?

9. Каким типом адреса снабжают посылаемые данные, когда хотят, чтобы они были доставлены всем узлам сети? Варианты ответов:

а) multicast;

б) anycast;

в) broadcast;

г) unicast.

10. Какие признаки могут быть использованы для определения информационного потока? Варианты ответов:

а) адрес назначения;

б) адрес источника;

в) тип приложения;

г) номер интерфейса, на который поступит пакет;

д) все перечисленные.

11. Опишите основные подходы и критерии, используемые при выборе маршрута.

12. Что можно считать недостатком метода нахождения маршрута по критерию минимума промежуточных узлов? Варианты ответов:

а) не учитывается пропускная способность линий связи;

б) не учитывается загрузка линий связи;

в) не учитывается топология сети.

13. Какие методы используются при мультиплексировании?

14. Объясните различия между разделением среды передачи и мультиплексированием.

15. Опишите, какие основные задачи нужно решить, чтобы обеспечить информационное взаимодействие любой пары абонентов в коммуникационной сети любого типа.

2. КОММУТАЦИЯ КАНАЛОВ И ПАКЕТОВ

2.1. Коммутация каналов

Исторически коммутация каналов (КК) появилась раньше коммутации пакетов (КП) и ведет свое происхождение от первых телефонных сетей. Невозможность динамического перераспределения пропускной способности физического канала является принципиальным ограничением сети с КК.

Принцип КП был предложен для компьютерных сетей. При КП учитываются особенности компьютерного трафика, поэтому данный способ коммутации является более эффективным для компьютерных сетей по сравнению с традиционным методом КК, применяющимся в телефонных сетях. Однако достоинства и недостатки любой сетевой технологии относительны. Наличие буферной памяти в коммутаторах сетей с КП позволяет эффективно использовать пропускную способность каналов при передаче пульсирующего трафика, но приводит к случайным задержкам в доставке пакетов, что для трафика реального времени является серьезным недостатком.

Сети с КК, имеют богатую историю и до сих пор находят широкое применение в мире телекоммуникаций, являясь основой высокоскоростных магистральных каналов связи. Первые сеансы связи между компьютерами были осуществлены через телефонную сеть, т. е. также с применением техники КК, а пользователи, которые получают доступ в Интернет по модему, продолжают обслуживаться этими сетями, так как их данные доходят до оборудования провайдера по местной телефонной сети [2].

В сетях с КК решаются все те частные задачи коммутации, которые были сформулированы ранее. Так, в качестве информационных потоков в сетях с КК выступают данные, которыми обмениваются пары абонентов. Глобальным признаком потока является пара адресов (телефонных номеров) абонентов, связывающихся между собой. Для всех возможных потоков заранее определяются маршруты. Маршруты в сетях с КК либо задаются «вручную» администратором сети, либо находятся автоматически. Маршруты фиксируются в таблицах, в которых признаком потока ставятся в соответствие идентификаторы входных интерфейсов коммутаторов. На основании этих таблиц происходят движение и мультиплексирование данных.

2.1.1. Элементарный канал

Одной из особенностей сетей с КК является понятие элементарного канала. *Элементарный канал* (или просто канал) – это базовая техническая характеристика сети с коммутацией каналов, представляющая собой некоторое фиксированное в пределах данного типа сетей значение пропускной

способности. Любая линия связи в сети с коммутацией каналов имеет пропускную способность, кратную элементарному каналу, принятому для данного типа сети.

В предыдущих разделах мы использовали термин «канал» как синоним термина «линия связи». Говоря же о сетях с коммутацией каналов, мы придаем термину «канал» значение единицы пропускной способности.

Значение элементарного канала, или, другими словами, минимальная единица пропускной способности линии связи, выбирается с учетом разных факторов. Например, в телефонных сетях распространенным значением элементарного канала сегодня является скорость 64 Кбит/с – это минимально достаточная скорость для качественной цифровой передачи голоса.

Линии связи в сетях с коммутацией пакетов имеют разную пропускную способность, одни – бóльшую, другие – меньшую. Выбирая линии связи с разными скоростными качествами, специалисты, проектирующие сеть, стараются учесть разную интенсивность информационных потоков, которые могут возникнуть в разных фрагментах сети – чем ближе к центру сети, тем выше пропускная способность линии связи, так как магистральные линии агрегируют трафик большого количества периферийных линий связи.

Особенностью сетей с коммутацией каналов является то, что пропускная способность каждой линии связи должна быть равна целому числу элементарных каналов. Так, линии связи, подключающие абонентов к телефонной сети, могут содержать 2, 24 или 30 элементарных каналов, а линии, соединяющие коммутаторы, – 480 или 1920 каналов.

Обратимся к фрагменту сети, изображенному на рис. 2.1 [2].

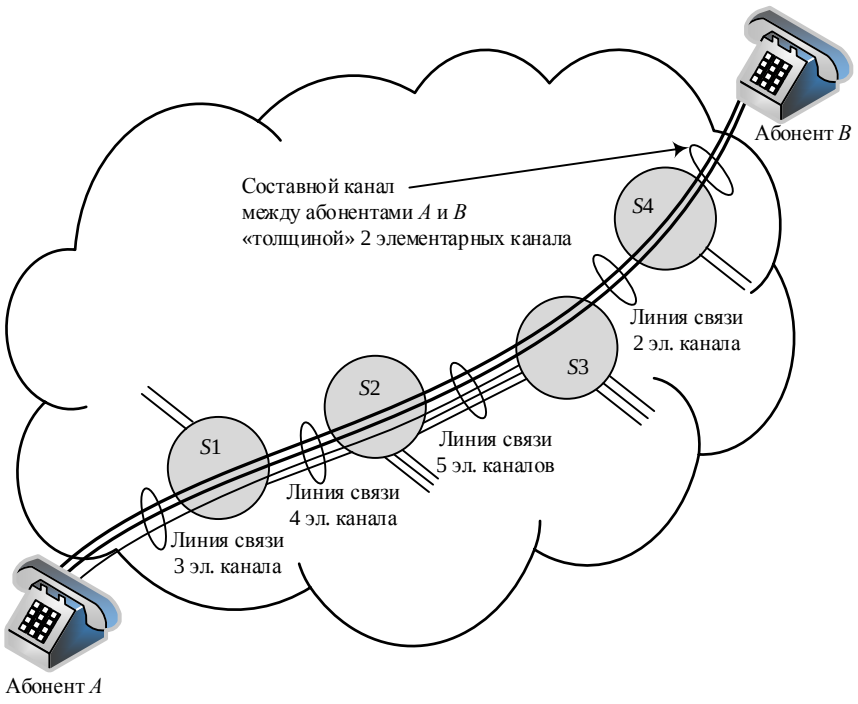


Рис. 2.1. Составной канал в сети с коммутацией каналов

Предположим, что сеть на рис. 2.1 характеризуется элементарным каналом *P* бит/с. В сети существуют линии связи разной пропускной способности, состоящие из 2, 3, 4 и 5 элементарных каналов. На рис. 2.1 показаны два абонента, *A* и *B*, генерирующие во время сеанса связи (телефонного разговора) *информационный поток*, для которого в сети был предусмотрен *маршрут*, проходящий через четыре коммутатора – *S*1, *S*2, *S*3 и *S*4. Предположим также, что интенсивность информационного потока между абонентами не превосходит 2P бит/с. Тогда для обмена данными этим двум абонентам достаточно иметь в своем распоряжении по паре элементарных каналов, «выделенных» из каждой линии связи, лежащей на маршруте следования данных от пункта *A* к пункту *B*. На рис. 2.1 эти элементарные каналы, необходимые абонентам *A* и *B*, обозначены толстыми линиями.

2.1.2. Составной канал

Канал, построенный путем коммутации (соединения) элементарных каналов, называют составным каналом. В рассматриваемом примере на рис. 2.1 для соединения абонентов *A* и *B* был создан составной канал «толщиной» в два элементарных канала. Подчеркнем следующие свойства составного канала: а) составной канал на всем протяжении состоит из *одинакового* количества элементарных каналов; б) составной канал имеет *постоянную и фиксированную пропускную способность* на всем своем протяжении; в) составной канал создается *временно* на период сеанса связи двух абонентов; г) на время сеанса связи все элементарные каналы, входящие в составной канал, поступают в *исключительное* пользование абонентов, для которых был создан этот составной канал; д) в течение всего сеанса связи абоненты могут посылать в сеть данные со скоростью, не превышающей пропускную способность составного канала; е) данные, поступившие в составной канал, гарантированно доставляются вызываемому абоненту *без задержек, потерь* и со скоростью источника вне зависимости от того, существуют ли в это время в сети другие соединения или нет; ж) после окончания сеанса связи элементарные каналы, входившие в соответствующий составной канал, *объявляются свободными* и возвращаются в пул распределяемых ресурсов для использования другими абонентами [2].

В сети может одновременно происходить несколько сеансов связи. Разделение сети между сеансами связи происходит на уровне элементарных каналов. Например, мы можем предположить, что после того, как в линии связи *S*2-*S*3 (рис. 2.1) было выделено два канала для связи абонентов *A* и *B*, оставшиеся три элементарных канала были распределены между тремя другими сеансами связи, проходившими в это же время и через эту же линию связи. Такое *мультиплексирование* позволяет одновременно передавать через каждый физический канал трафик нескольких логических соединений.

Мультиплексирование означает, что абоненты вынуждены конкурировать за ресурсы, в данном случае за элементарные каналы. Возможны ситуации, когда некоторая промежуточная линия связи уже исчерпала свободные элементарные каналы, тогда новый сеанс связи, маршрут которого пролегает через данную линию связи, не может состояться.

Для того чтобы распознать такие ситуации, обмен данными в сети с коммутацией каналов предваряется *процедурой установления соединения*. В соответствии с этой процедурой абонент, являющийся инициатором сеанса связи (например, абонент *A* в нашей сети), посылает в коммутационную сеть запрос, представляющий собой сообщение, в котором содержится адрес вызываемого абонента, например, абонента *B*; в телефонной сети посылке запроса соответствует набор телефонного номера. Цель запроса – проверить, можно ли образовать составной канал между вызывающим и вызываемым абонентами. А для этого требуется соблюдение двух условий: наличие свободных элементарных каналов в каждой линии связи, лежащей на пути от *A* к *B*, и незанятость вызываемого абонента в другом соединении.

Запрос перемещается по маршруту, определенному для информационного потока данной пары абонентов. При этом используются глобальные таблицы коммутации, ставящие в соответствие глобальному признаку потока (адресу вызываемого абонента) идентификатор выходного интерфейса коммутатора (такие таблицы часто называют таблицами маршрутизации).

Если в результате прохождения запроса от абонента *A* к абоненту *B* выяснилось, что ничто не препятствует установлению соединения, происходит *фиксация* составного канала. Для этого во всех коммутаторах вдоль пути от *A* до *B* создаются записи в *локальных таблицах коммутации*, в которых указывается соответствие между *локальными признаками потока* – номерами элементарных каналов, зарезервированных для этого сеанса связи. Только после этого составной канал считается установленным и абоненты *A* и *B* могут начать свой сеанс связи.

Таким образом, продвижение данных в сетях с коммутацией каналов происходит в два этапа [2].

1. В сеть поступает служебное сообщение – запрос, который несет адрес вызываемого абонента и инициирует создание составного канала.

2. По подготовленному составному каналу передается основной поток данных, для передачи которого уже не требуется никакой вспомогательной информации, в том числе адреса вызываемого абонента. Коммутация данных в коммутаторах выполняется на основе локальных признаков – номеров элементарных каналов.

Запросы на установление соединения не всегда завершаются успешно. Если на пути между вызывающим и вызываемым абонентами отсутствуют

свободные элементарные каналы или вызываемый узел занят, то происходит *отказ в установлении соединения*. Например, если во время сеанса связи абонентов *A* и *B* абонент *C* пошлет запрос в сеть на установление соединения с абонентом *D*, то он получит отказ, потому что оба необходимых ему элементарных канала, составляющих линию связи коммутатором *S*3 и *S*4, уже выделены соединению абонентов *A* и *B* (рис. 2.2). При отказе в установлении соединения сеть информирует вызывающего абонента специальным сообщением; телефонная сеть передает в этом случае короткие гудки – сигнал «занято». Чем больше нагрузка на сеть, т. е. чем больше соединений она в данный момент поддерживает, тем больше вероятность отказа в удовлетворении запроса на установление нового соединения.

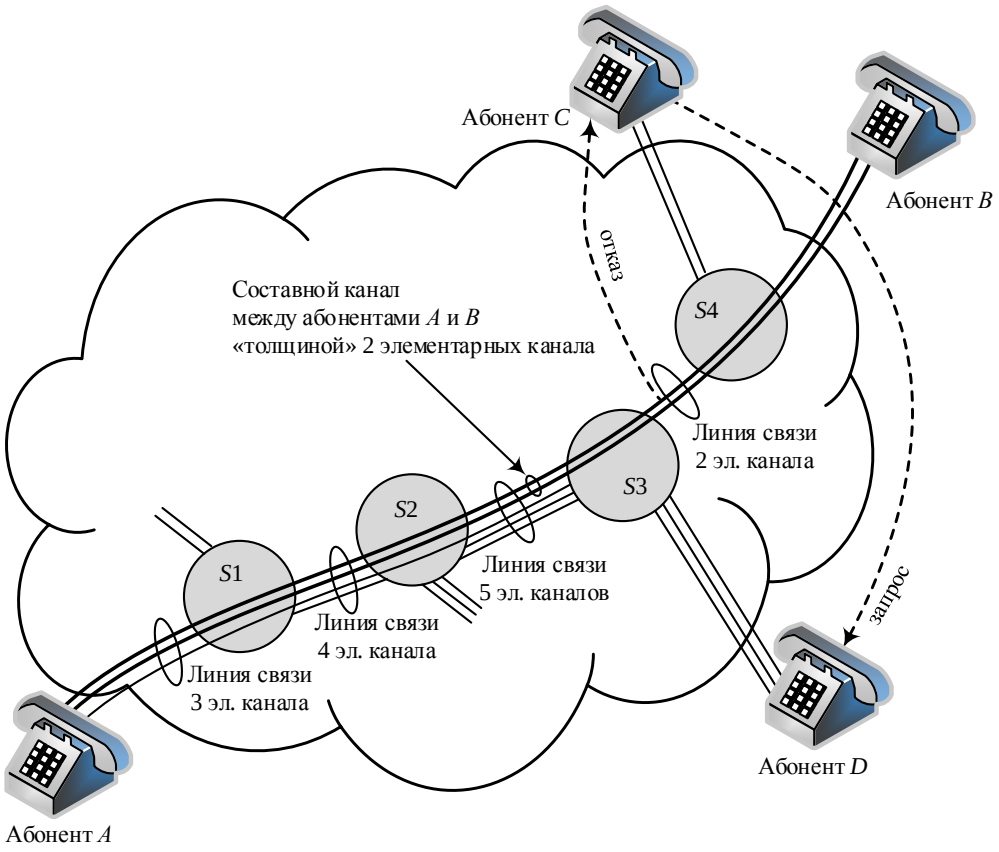


Рис. 2.2. Отказ в установлении соединения
в сети с коммутацией каналов

Мы описали процедуру установления соединения в автоматическом динамическом режиме, основанном на способности абонентов отправлять в сеть служебные сообщения – запросы на установление соединения – и способности узлов сети обрабатывать такие сообщения. Подобный режим используется телефонными сетями: телефонный аппарат генерирует запрос, посылая в сеть импульсы (или тоновые сигналы), кодирующие номер вызываемого абонента, а сеть либо устанавливает соединение, либо сообщает об отказе сигналами «занято».

Однако это не единственно возможный режим работы сети с коммутацией каналов, существует и другой статический ручной режим установления соединения. Этот режим характерен для случаев, когда необходимо установить составной канал не на время одного сеанса связи абонентов, а на более длительный срок. Создание такого долговременного канала не могут инициировать абоненты, он создается администратором сети. Очевидно, что статический ручной режим мало пригоден для традиционной телефонной сети с ее короткими сеансами связи, однако он вполне оправдан для создания высокоскоростных телекоммуникационных каналов между городами и странами на более или менее постоянной основе.

Технология коммутации каналов ориентирована на минимизацию случайных событий в сети, т. е. это технология, стремящаяся к детерминизму. Во избежание всяких возможных неопределенностей значительная часть работы по организации информационного обмена выполняется заранее, еще до того, как начнется собственно передача данных. Сначала по заданному адресу проверяется доступность необходимых элементарных каналов на всем пути от отправителя до адресата. Затем эти каналы закрепляются на все время сеанса для исключительного использования двумя абонентами и коммутируются в один непрерывный «трубопровод» (составной канал), имеющий «шлюзовые задвижки» на стороне каждого из абонентов. После этой исчерпывающей подготовительной работы остается сделать самое малое: «открыть шлюзы» и позволить информационному потоку свободно и без помех «перетекать» между заданными точками сети (рис. 2.3) [2].

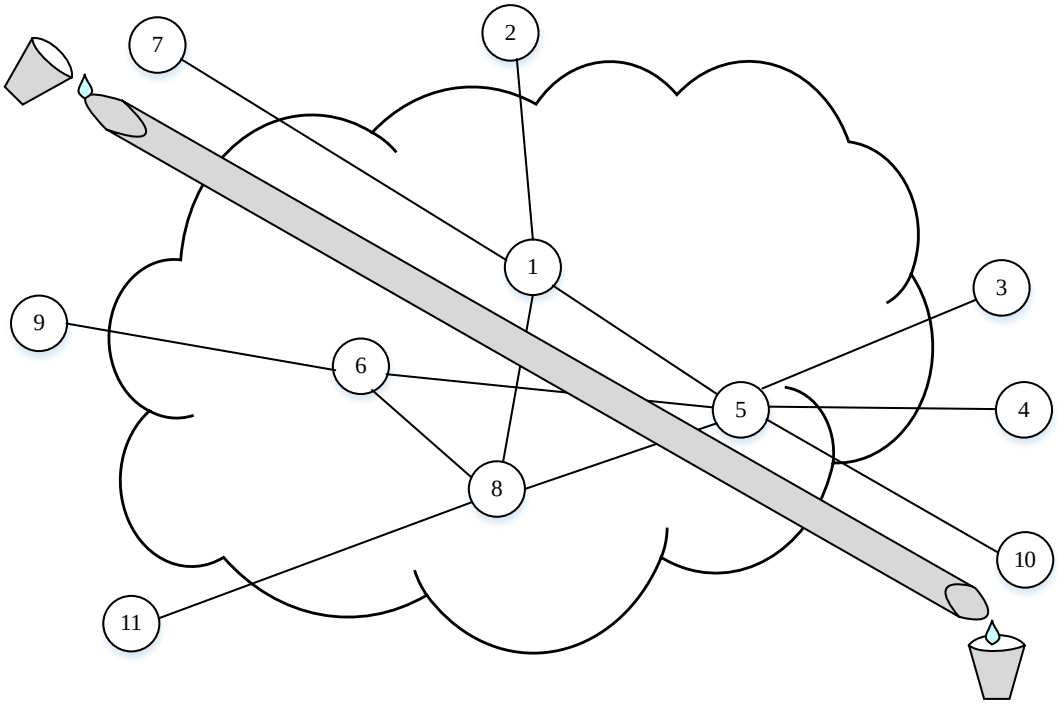


Рис. 2.3. Сеть с коммутацией каналов
как система трубопроводов

2.1.3. Неэффективность передачи пульсирующего трафика

Сети с коммутацией каналов (КК) наиболее эффективно передают пользовательский трафик в том случае, когда скорость его постоянна в течение всего сеанса связи и максимально соответствует *фиксированной* пропускной способности физических линий связи сети. Эффективность работы сети снижается, когда информационные потоки, генерируемые абонентами, приобретают *пульсирующий* характер [2].

Так, разговаривая по телефону, люди постоянно меняют темп речи, перемежая быстрые высказывания паузами. В результате соответствующие «голосовые» информационные потоки становятся неравномерными, а значит, снижается эффективность передачи данных. Правда, в случае телефонных разговоров это снижение оказывается вполне приемлемым и позволяет широко использовать сети с КК для передачи голосового трафика.

Гораздо сильнее снижает эффективность сети с коммутацией каналов передача так называемого компьютерного трафика, т. е. трафика, генерируемого приложениями, с которыми работает пользователь компьютера. Этот трафик практически всегда является пульсирующим. Например, когда вы загружаете из Интернета очередную страницу, скорость трафика резко возрастает, а после окончания загрузки падает практически до нуля. Если для описанного сеанса доступа в Интернет вы задействуете сеть с коммутацией каналов, то большую часть времени составной канал между вашим компьютером и вэб-сервером будет простаивать. В то же время производительность сети окажется закрепленной за вами и останется недоступной другим пользователям сети. Сеть в такие периоды похожа на пустой эскалатор метро, который движется, но полезную работу не выполняет, другими словами, «перевозит воздух». Для эффективной передачи неравномерного компьютерного трафика была разработана техника коммутации пакетов.

Выводы

В сетях с коммутацией каналов по запросу пользователя создается непрерывный информационный канал, который образуется путем резервирования «цепочки» линий связи, соединяющих абонентов на время передачи данных. На всем своем протяжении канал передает данные с одной и той же скоростью. Это означает, что через сеть с коммутацией каналов можно качественно передавать данные, чувствительные к задержкам (голос, видео). Однако невозможность динамического перераспределения пропускной способности физического канала является принципиальным недостатком сети с коммутацией каналов, который делает ее неэффективной для передачи пульсирующего трафика.

Контрольные вопросы

1. Какие типы мультиплексирования и коммутации используются в телефонных сетях?

2. Трафик какого типа сеть с коммутацией каналов передает неэффективно?

3. Может ли сеть с коммутацией каналов работать без буферизации данных?

4. Какие из сформулированных свойств составного канала всегда соответствуют действительности? Варианты ответов:

а) данные, поступившие в составной канал, доставляются вызываемому абоненту без задержек;

б) составной канал закрепляется за двумя абонентами на постоянной основе;

в) количество элементарных каналов, входящих в составной канал между двумя абонентами, равно количеству промежуточных узлов плюс 1;

г) составной канал имеет постоянную и фиксированную пропускную способность на всем своем протяжении.

5. Какой элемент сети с коммутацией каналов может отказать узлу в запросе на установление составного канала?

2.2. Коммутация пакетов

Сети с коммутацией пакетов (КП), так же, как и сети с коммутацией каналов, состоят из коммутаторов, связанных физическими линиями связи. Однако передача данных в этих сетях происходит совершенно по-другому. Образно говоря, по сравнению с сетью с КК сеть с КП ведет себя менее «ответственно». Например, она может принять данные для передачи, не заботясь о резервировании линий связи на пути следования этих данных и не гарантируя требуемую пропускную способность. Сеть с КП не создает заранее для своих абонентов отдельных каналов связи, выделенных исключительно для них. Данные могут задерживаться и даже теряться по пути следования.

Как же при таком хаосе и неопределенности сеть с коммутацией пакетов выполняет свои функции по передаче данных?

Важнейшим принципом функционирования сетей с КП является представление информации, передаваемой по сети, в виде структурно отделенных друг от друга порций данных, называемых пакетами; наряду с термином «пакет» используются также термины «кадр», «фрэйм», «ячейка» и др.

Каждый пакет снабжен *заголовком* (рис. 2.4), в котором содержатся адрес назначения и другая вспомогательная информация (длина поля данных, контрольная сумма и др.), используемая для доставки пакета адресату. Наличие адреса в каждом пакете является одной из важнейших особенностей техники коммутации пакетов, так как каждый пакет может быть обработан коммутатором *независимо*[3] от других пакетов, составляющих сетевой трафик. Помимо заголовка у пакета может иметься еще одно дополнительное

[3] В некоторых технологиях коммутации пакетов (например, в технологии виртуальных каналов) полная независимость обработки пакетов не обеспечивается.

поле, размещаемое в конце пакета и поэтому называемое *концевиком*. В концевике обычно помещается *контрольная сумма*, которая позволяет проверить, была ли искажена информация при передаче через сеть или нет [2].

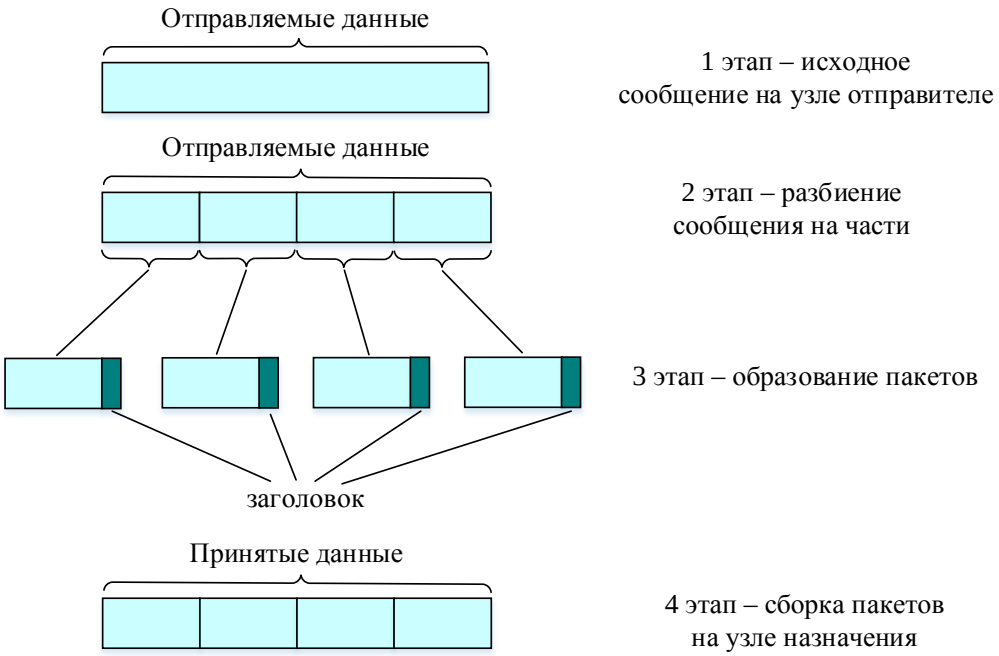


Рис. 2.4. Разбиение данных на пакеты

В зависимости от конкретной реализации технологии КП пакеты могут иметь фиксированную или переменную длину, кроме того может меняться состав информации, размещенной в заголовках пакетов. Например, в технологии ATM пакеты (называемые там ячейками) имеют фиксированную длину, а в технологии Ethernet установлены лишь минимально и максимально возможные размеры пакетов (кадров).

Пакеты поступают в сеть *без предварительного резервирования линий связи*[4] и *не с фиксированной заранее заданной скоростью*, как это делается в сетях с КК, а в том темпе, в котором их генерирует источник. Предполагается, что сеть с КП в отличие от сети с КК всегда готова принять пакет от конечного узла. Как и в сетях с КК, в сетях с КП для каждого из потоков определяется маршрут, фиксируемый в хранящихся на коммутаторах таблицах коммутации. Пакеты, попадая на коммутатор, обрабатываются

[4] Процедура резервирования пропускной способности может применятся и в сетях с коммутацией пакетов. Однако основная идея такого резервирования принципиально отличается от идеи резервирования пропускной способности в сетях с коммутацией каналов. Разница заключается в том, что пропускная способность канала сети с коммутацией пакетов может динамически перераспределяться между информационными потоками в зависимости от текущих потребностей каждого потока, чего не может обеспечить техника КК.

и направляются по тому или иному маршруту на основании информации, содержащейся в их заголовках, а также в таблице коммутации (рис. 2.5) [2].

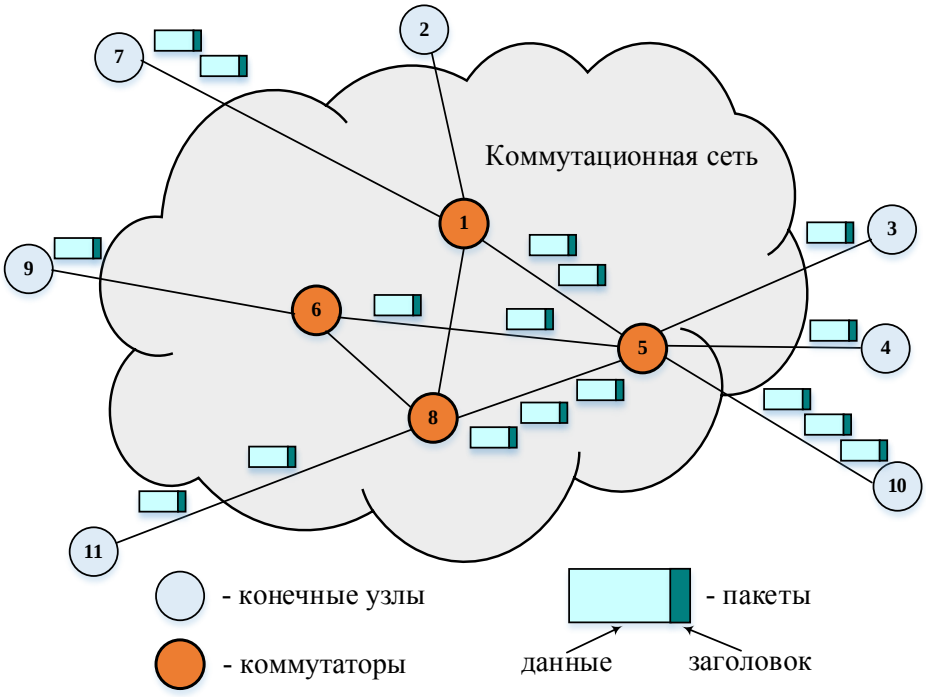


Рис. 2.5. Передача данных
в сетях с коммутацией пакетов

Пакеты, принадлежащие как одному и тому же, так и разным информационным потокам, при перемещении по сети могут «перемешиваться» между собой, образовывать очереди и «тормозить» друг друга. На пути пакетов могут встречаться линии связи, имеющие разную пропускную способность. В зависимости от времени суток может сильно меняться и степень загруженности линий связи. В таких условиях не исключены ситуации, когда пакеты, принадлежащие одному и тому же потоку, могут перемещаться по сети с разными скоростями и даже прийти к месту назначения не в том порядке, в котором они были отправлены.

Разделение данных на пакеты позволяет передавать неравномерный компьютерный трафик более эффективно, чем в сетях с коммутацией каналов. Это объясняется тем, что пульсации трафика от отдельных компьютеров носят случайный характер и распределяются во времени так, что их пики чаще всего не совпадают. Поэтому, когда линия связи передает трафик большого количества конечных узлов, в суммарном потоке пульсации сглаживаются и пропускная способность линии используется более рационально, без длительных простоев. Этот эффект иллюстрируется рис. 2.6, на котором показаны неравномерные потоки пакетов, поступающие от конечных узлов 3, 4 и 10 в сети, изображенной на рис. 2.5. Предположим, что

эти потоки передаются в направлении коммутатора 8, а следовательно, накладываются друг на друга при прохождении линии связи между коммутаторами 5 и 8. Получающийся в результате суммарный поток является более равномерным, чем каждый из образующих его отдельных потоков [2].

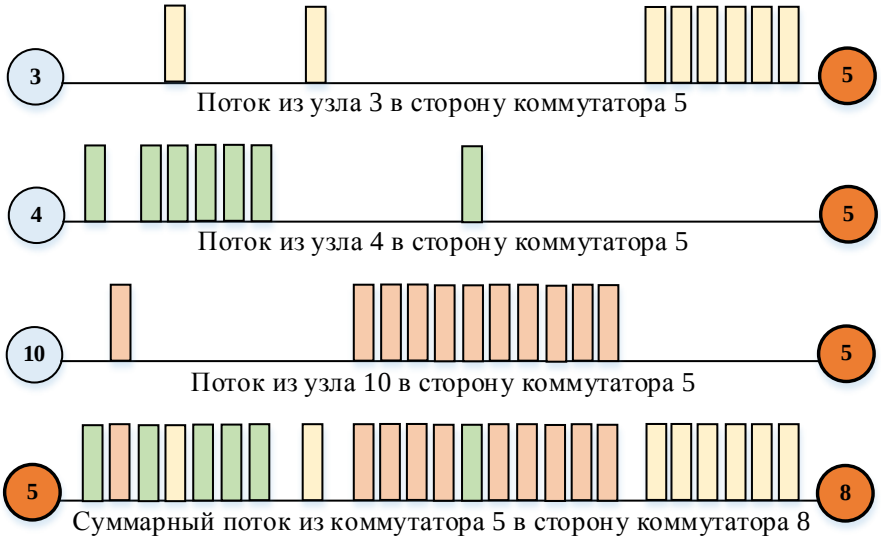


Рис. 2.6. Сглаживание трафика
в сетях с коммутацией пакетов

2.2.1. Буферизация пакетов

Неопределенность и асинхронность перемещения данных в сетях с коммутацией пакетов предъявляет особые требования к работе коммутаторов в таких сетях. *Главное отличие пакетных коммутаторов от коммутаторов в сетях с коммутацией каналов состоит в том, что они имеют внутреннюю буферную память для временного хранения пакетов.*

Действительно, пакетный коммутатор не может принять решение о продвижении пакета, не имея в своей памяти всего пакета. Коммутатор проверяет контрольную сумму, и только если она говорит о том, что данные пакета не искажены, начинает обрабатывать пакет и по адресу назначения определяет следующий коммутатор. Поэтому *каждый* пакет последовательно, бит за битом, помещается во *входной буфер*. Имея в виду это свойство, говорят, что сети с коммутацией пакетов используют технику *сохранения с продвижением* (store-and-forward). Заметим, что для этой цели достаточно иметь буфер размером в один пакет. Коммутатору нужны буферы для *согласования скоростей передачи данных в линиях связи*, подключенных к его интерфейсам. Действительно, если скорость поступления пакетов из одной линии связи в течение некоторого периода превышает пропускную способность той линии связи, в которую эти пакеты должны быть направлены, то во избежание потерь пакетов на целевом интерфейсе необходимо организовать выходную очередь (рис. 2.7) [2].

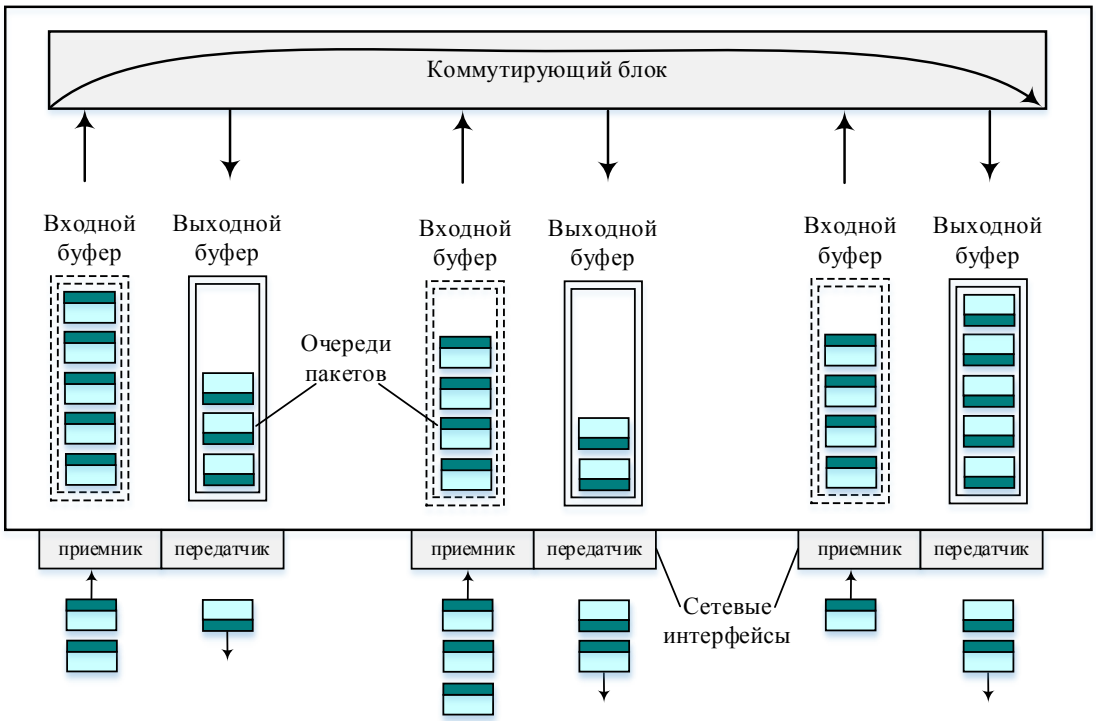


Рис. 2.7. Буферы и очереди пакетов в коммутаторе

Буферизация необходима пакетному коммутатору также для *согласования скорости поступления пакетов со скоростью их коммутации*. Если коммутирующий блок не успевает обрабатывать пакеты (анализировать заголовки и перебрасывать пакеты на нужный интерфейс), то на интерфейсах коммутатора возникают *входные очереди*. Очевидно, что для хранения входной очереди объем буфера должен превышать размер одного пакета. Современные способы построения коммутирующего блока основаны на многопроцессорном подходе, когда каждый интерфейс имеет свой встроенный процессор для обработки пакетов. Кроме того, существует центральный процессор, координирующий работу интерфейсных процессоров. Использование интерфейсных процессоров повышает производительность коммутатора и уменьшает очереди на входных интерфейсах.

Поскольку объем буферов в коммутаторах ограничен, иногда происходит потеря пакетов из-за переполнения буферов при временной перегрузке части сети, когда совпадают периоды пульсации нескольких информационных потоков. Для сетей с коммутацией пакетов потеря пакетов является обычным явлением, и для компенсации таких потерь в данной сетевой технологии предусмотрен ряд специальных механизмов.

Пакетный коммутатор может работать по одному из трех методов продвижения пакетов: а) дейтаграммная передача; б) передача с установлением логического соединения; в) передача с установлением виртуального канала.

2.2.2. Дейтаграммная передача

Дейтаграммный способ передачи данных основан на том, что все передаваемые пакеты продвигаются (передаются от одного узла сети другому) независимо друг от друга на основании одних и тех же правил.

Процедура обработки пакета определяется только значениями параметров, которые он несет в себе, и текущим состоянием сети (например, в зависимости от ее нагрузки пакет может стоять в очереди большее или меньшее время). Однако никакая информация об уже переданных пакетах сетью не хранится и в ходе обработки очередного пакета во внимание не принимается, т. е. каждый отдельный пакет рассматривается сетью как совершенно независимая единица передачи – *дейтаграмма*.

Решение о продвижении пакета принимается на основе таблицы коммутации, ставящей в соответствие адресам назначения пакетов информацию, однозначно определяющую следующий по маршруту транзитный (или конечный) узел. В качестве такой информации могут выступать идентификаторы интерфейсов данного коммутатора или адреса входных интерфейсов коммутаторов, следующих по маршруту. На рис. 2.8 показана сеть, в которой шесть конечных узлов (*N*1–*N*6) связаны семью коммутаторами (*S*1–*S*7).

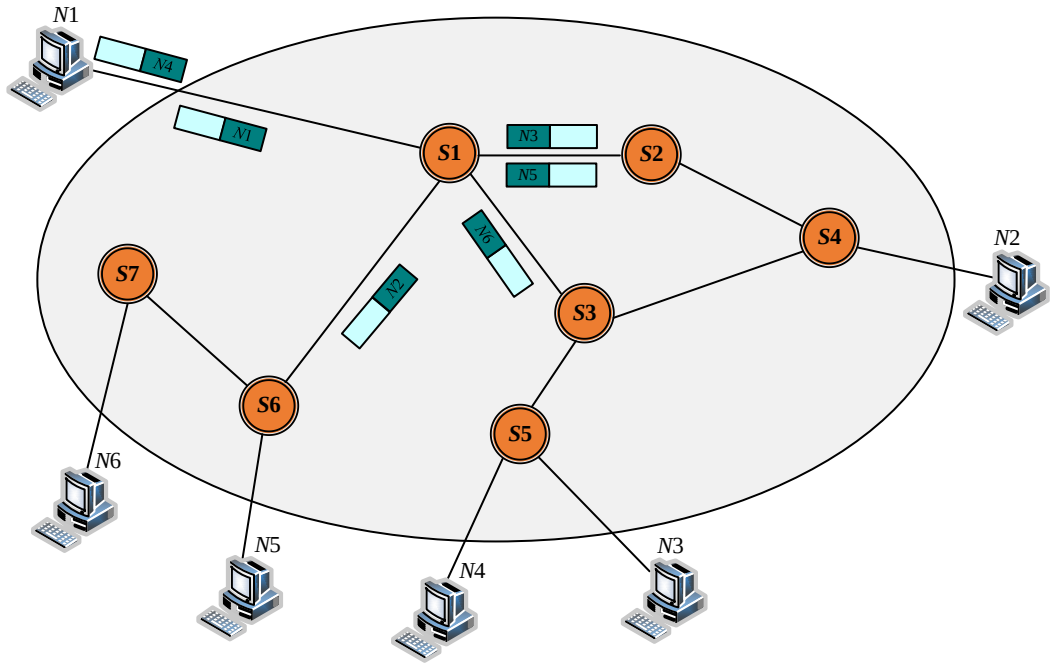


Рис. 2.8. Иллюстрация дейтаграммного принципа передачи пакетов

В табл. 2.1 представлена таблица коммутации коммутатора *S*1 при дейтаграммной передаче пакета.

Таблица 2.1

Таблица коммутации коммутатора S1 при дейтаграммной передаче пакета

Адрес назначения	Адрес следующего коммутатора
*N*1	пакет не требуется передавать через сеть
*N*2	*S*2
*N*3	*S*3
*N*4	*S*3
*N*5	*S*6
*N*6	*S*6

На рис. 2.8 показаны также несколько перемещающихся по разным маршрутам пакетов с разными адресами назначения (*N*1 – *N*6), на пути которых лежит коммутатор *S*1. При поступлении каждого из этих пакетов в коммутатор *S*1 выполняются просмотр соответствующей таблицы коммутации и выбор дальнейшего пути перемещения. Так, пакет с адресом *N*5 будет передан коммутатором *S*1 на интерфейс, ведущий к коммутатору *S*6, где в результате подобной процедуры этот пакет будет направлен конечному узлу получателю *N*5. В таблице коммутации для одного и того же адреса назначения может содержаться несколько записей, указывающих соответственно на различные адреса следующего коммутатора. Такой подход называется *балансом нагрузки* и используется для повышения производительности и надежности сети. В примере, показанном на рис. 2.8, пакеты, поступающие в коммутатор *S*1 для узла назначения с адресом *N*2, в целях баланса нагрузки распределяются между двумя следующими коммутаторами – *S*2 и *S*3, что снижает нагрузку на каждый из них, а значит сокращает очереди и ускоряет доставку. Некоторая «размытость» путей следования пакетов с одним и тем же адресом назначения через сеть является прямым следствием принципа независимой обработки каждого пакета, присущего дейтаграммному методу. Пакеты, следующие по одному и тому же адресу назначения, могут добираться до него разными путями также вследствие изменения состояния сети, например, отказа промежуточных коммутаторов.

Дейтаграммный метод работает быстро, так как никаких предварительных действий перед отправкой данных проводить не требуется. Однако при таком методе трудно проверить факт доставки пакета узлу назначения. В этом методе доставка пакета не гарантируется, а выполняется по мере возможности – для описания такого свойства используется термин *доставка по возможности* (best effort).

2.2.3. Передача с установлением логического соединения

Следующий рассматриваемый способ продвижения пакетов основын на знании устройствами сети «истории» обмена данными, например, на запоминании узлом-отправителем числа отправленных, а узлом-получателем – числа полученных пакетов. Такого рода информация фиксируется в рамках логического соединения.

Процедура согласования двумя конечными узлами сети некоторых параметров процесса обмена пакетами называется установлением логического соединения. Параметры, о которых договариваются два взаимодействующих узла, называются параметрами логического соединения.

Наличие логического соединения позволяет более рационально по сравнению с дейтаграммным способом обрабатывать пакеты. Например, при потере нескольких предыдущих пакетов может быть снижена скорость отправки последующих. Или, благодаря нумерации пакетов и отслеживанию номеров отправленных и принятых пакетов, можно повысить надежность путем отбрасывания дубликатов, упорядочивания поступивших и повторения передачи потерянных пакетов.

Параметры соединения могут быть: *постоянными*, т. е. не изменяющимися в течение всего соединения (например, идентификатор соединения, способ шифрования пакета или максимальный размер поля данных пакета), или *переменными*, т. е. динамически отражающими текущее состояние соединения (например, последовательные номера передаваемых пакетов).

Когда отправитель и получатель *фиксируют* начало нового соединения, они прежде всего «договариваются» о начальных значениях параметров процедуры обмена и только после этого начинают передачу собственно данных.

Передача с установлением соединения более надежна, но требует больше времени для передачи данных и вычислительных затрат от оконечных узлов, что иллюстрирует рис. 2.9 [2].

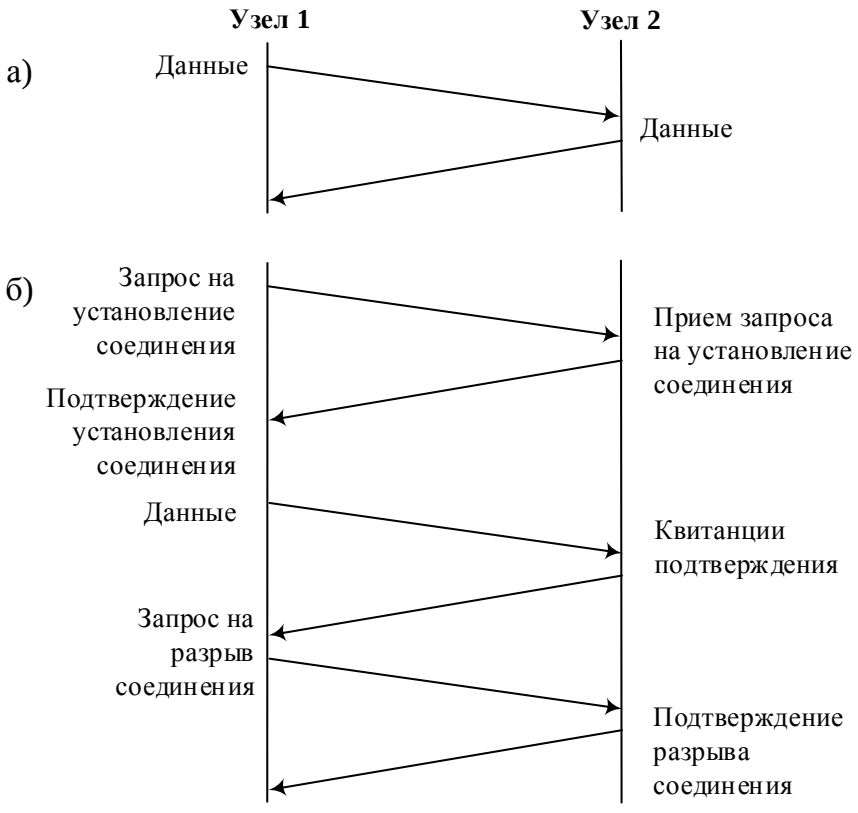


Рис. 2.9. Передача без установления соединения (а) и с установлением соединения (б)

Процедура установления соединения состоит обычно из трех шагов.

1. Узел-инициатор соединения отправляет узлу-получателю служебный пакет с предложением установить соединение.

2. Если узел-получатель согласен с этим, то он посылает в ответ другой служебный пакет, подтверждающий установление соединения и предлагающий некоторые параметры, которые должны использоваться в рамках данного логического соединения. Это могут быть, например, идентификатор соединения, количество кадров, которые можно отправить без получения подтверждения, и т. п.

3. Узел-инициатор соединения может закончить процесс установления соединения отправкой третьего служебного пакета, в котором сообщит, что предложенные параметры ему подходят.

Логическое соединение может быть рассчитано на передачу данных как в одном направлении – от инициатора соединения, – так и в обоих направлениях. После передачи некоторого законченного набора данных, например, определенного файла, узел-отправитель инициирует разрыв данного логического соединения, посылая соответствующий служебный кадр.

Заметим, что в отличие от передачи дейтаграммного типа, в которой поддерживается только один тип кадра – информационный, передача с установлением соединения должна поддерживать как минимум два типа кадров – информационные кадры переносят собственно пользовательские данные, а служебные предназначаются для установления (разрыва) соединения.

После того как соединение установлено и все параметры согласованы, конечные узлы начинают передачу собственно данных. Пакеты данных обрабатываются коммутаторами точно так же, как и при дейтаграммной передаче: из заголовка пакетов извлекаются адреса назначения и сравниваются с записями в таблицах коммутации, содержащих информацию о следующих шагах по маршруту. Так же как дейтаграммы, пакеты, относящиеся к одному логическому соединению, в некоторых случаях (например, при отказе линии связи) могут доставляться адресату по разным маршрутам.

Однако передача с установлением соединения имеет важное отличие от дейтаграммной передачи, поскольку в ней помимо обработки пакетов на коммутаторах имеет место *дополнительная обработка пакетов на конечных узлах*. Например, если при установлении соединения была оговорена передача данных в зашифрованном виде, то шифрование пакетов выполняется узлом-отправителем, а дешифрование – узлом-получателем. Аналогично для обеспечения в рамках логического соединения надежности всю работу по нумерации пакетов, отслеживанию номеров доставленных и недоставленных пакетов, посылке копий и отбрасыванию дубликатов берут на себя конечные узлы.

Механизм установления логических соединений позволяет реализовывать дифференцированное обслуживание информационных потоков. Разное обслуживание могут получить даже потоки, относящиеся к одной и той же паре конечных узлов. Например, пара конечных узлов может установить два параллельно работающих логических соединения, в одном из которых передавать данные в зашифрованном виде, а в другом – открытым текстом.

Как видим, передача с установлением соединения предоставляет больше возможностей в плане надежности и безопасности обмена данными, чем дейтаграммная передача. Однако этот способ более медленный, так как он подразумевает дополнительные вычислительные затраты на установление и поддержание логического соединения.

2.2.4. Передача с установлением виртуального канала

Следующий способ продвижения данных основан на частном случае логического соединения, в число параметров которого входит жестко определенный для всех пакетов *маршрут*. т. е. все пакеты в рамках данного соединения должны проходить по одному и тому же закрепленному за этим соединением пути.

Единственный заранее проложенный фиксированный маршрут, соединяющий конечные узлы в сети с коммутацией пакетов, называют *виртуальным каналом* (virtual circuit, или virtual channel).

Виртуальные каналы прокладываются для *устойчивых* информационных потоков. С целью выделения потока данных из общего трафика каждый пакет этого потока помечается признаком особого вида – *меткой*.

Так же как в сетях с установлением логических соединений, прокладка виртуального канала начинается с отправки узлом-источником специального пакета – запроса на установление соединения. В запросе указывается адрес назначения и метка потока, для которого прокладывается этот виртуальный канал. Запрос, проходя по сети, формирует новую запись в каждом из коммутаторов, расположенных на пути от отправителя до получателя. Запись говорит о том, каким образом коммутатор должен обслуживать пакет, имеющий заданную метку. Образованный виртуальный канал идентифицируется той же меткой[5].

После прокладки виртуального канала сеть может передавать по нему соответствующий поток данных. Во всех пакетах, которые переносят пользовательские данные, адрес назначения уже не указывается, его роль играет

[5] Эта метка в различных технологиях называется по-разному: номером логического канала (Logical Channel Number, LCN) в технологии X.25, идентификатором соединения уровня канала данных (Data Link Connection Identifier, DLCI) в технологии Frame Relay, идентификатором виртуального канала (Virtual Channel Identifier, VCI) в технологии ATM.

метка виртуального канала. При поступлении пакета на входной интерфейс коммутатор читает значение метки из заголовка пришедшего пакета и просматривает свою таблицу коммутации, по которой определяет, на какой выходной порт передать пришедший пакет.

На рис. 2.10 показана сеть, в которой проложено два виртуальных канала (*Virtual Channel, VC*), идентифицируемых метками *VC*1 и *VC*2. Первый проходит от конечного узла с адресом *N*1 до конечного узла с адресом *N*2 через промежуточные коммутаторы *S*1, *S*2 и *S*4. Второй виртуальный канал *VC*2 обеспечивает продвижение данных по пути *N*1-*S*1-*S*3-*S*5-*N*3. В общем случае между двумя конечными узлами может быть проложено несколько виртуальных каналов, например, еще один виртуальный канал между узлами *N*1 и *N*2 мог бы проходить через промежуточный коммутатор *S*3. На рис. 2.10 показаны два пакета, несущие в своих заголовках метки потоков *VC*1 и *VC*2, которые играют роли адресов назначения [2].

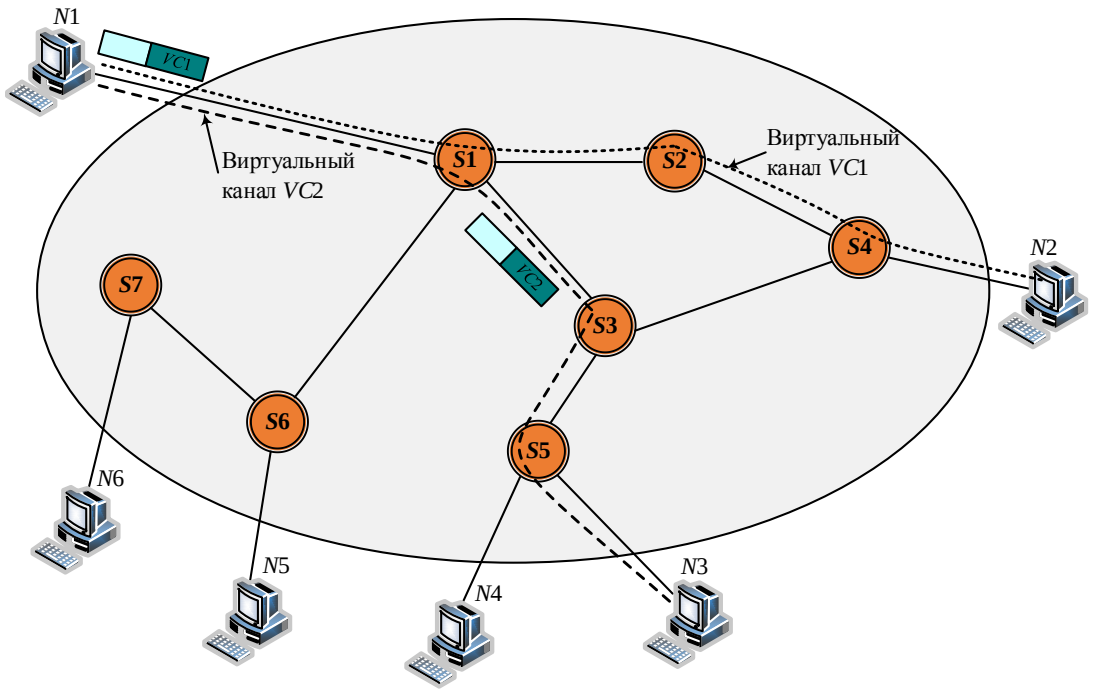


Рис. 2.10. Иллюстрация принципа работы виртуального канала

В табл. 2.2 представлена таблица коммутации коммутатора *S*1 при передаче с установлением виртуального канала.

Таблица 2.2

Таблица коммутации *S*1 при передаче с установлением виртуального канала

Адрес назначения	Адрес следующего коммутатора
*VC*1	*S*2
*VC*2	*S*3

Таблица коммутации в сетях, использующих виртуальные каналы, отличается от таблицы коммутации в дейтаграммных сетях. Она содержит записи *только о проходящих через коммутатор виртуальных каналах*, а не обо всех возможных адресах назначения, как это имеет место в сетях с дейтаграммным алгоритмом продвижения. Обычно в крупной сети количество проложенных через узел виртуальных каналов существенно меньше общего количества узлов, поэтому и таблицы коммутации в этом случае намного короче, а, следовательно, анализ такой таблицы занимает у коммутатора меньше времени. По той же причине метка короче адреса конечного узла и заголовок пакета в сетях с виртуальными каналами переносит по сети вместо длинного адреса компактный идентификатор потока.

Использование в сетях техники виртуальных каналов не делает их сетями с коммутацией каналов. Хотя в подобных сетях применяется процедура предварительного установления канала, этот канал является виртуальным, т. е. по нему передаются отдельные пакеты, а не потоки информации с постоянной скоростью, как в сетях с коммутацией каналов.

В одной и той же сетевой технологии могут быть задействованы разные способы продвижения данных. Так, дейтаграммный протокол IP используется для передачи данных между отдельными сетями, составляющими Интернет. В то же время обеспечением надежной доставки данных между конечными узлами этой сети занимается протокол TCP, устанавливающий логическое соединение без фиксации маршрута. И наконец, Интернет – это пример сети, применяющий виртуальные каналы, так как в состав Интернета входят сети ATM и Frame Relay, поддерживающие виртуальные каналы.

Выводы

При коммутации пакетов передаваемые данные разбиваются в исходном узле на небольшие части, пакеты. Пакет снабжается заголовком, в котором указывается адрес назначения, поэтому он может быть обработан коммутатором независимо от остальных данных. Коммутация пакетов повышает производительность сети при передаче пульсирующего трафика, так как при обслуживании большого числа независимых потоков периоды их активности не всегда совпадают во времени. Пакеты поступают в сеть без предварительного резервирования ресурсов в том темпе, в котором их генерирует источник. Однако этот способ коммутации имеет и недостатки: задержки передачи носят случайный характер, поэтому возникают проблемы при передаче трафика реального времени.

В сетях с коммутацией пакетов может использоваться один из трех алгоритмов продвижения пакетов: дейтаграммная передача, передача с установлением логического соединения и передача с установлением виртуального канала.

Контрольные вопросы

1. Из-за чего скорость передачи пользовательских данных в сетях с коммутацией пакетов всегда ниже пропускной способности каналов связи? Варианты ответов:

а) из-за наличия заголовков у пакетов;

б) из-за необходимости буферизовать пакеты перед обработкой;

в) из-за низкого быстродействия маршрутизаторов.

2. Какие свойства сетей с коммутацией пакетов негативно сказываются на передаче мультимедийной информации?

3. Учитывается ли в дейтаграммных сетях существование потоков данных?

4. Дайте определение логического соединения.

5. Какое логическое соединение может быть названо виртуальным каналом?

2.3. Сравнение сетей с коммутацией пакетов и каналов

Прежде чем проводить техническое сравнение сетей с коммутацией пакетов и сетей с коммутацией каналов, проведем их неформальное сравнение на основе весьма продуктивной транспортной аналогии.

2.3.1. Транспортная аналогия сетей с коммутацией пакетов и каналов

Для начала убедимся, что движение на дорогах имеет много общего с перемещением пакетов в сети с *коммутацией пакетов*.

Пусть автомобили в этой аналогии соответствуют пакетам, дороги – каналам связи, а перекрестки – коммутаторам. Подобно пакетам, автомобили перемещаются независимо друг от друга, разделяя пропускную способность дорог и создавая препятствия друг другу. Слишком интенсивный трафик, не соответствующий пропускной способности дороги, приводит к перегруженности дорог, в результате автомобили стоят в пробках, что соответствует очередям пакетов в коммутаторах.

На перекрестках происходит «коммутация» потоков автомобилей, каждый из автомобилей выбирает подходящее направление перекрестка, чтобы попасть в пункт назначения. Конечно, перекресток играет намного более пассивную роль по сравнению с коммутатором пакетов. Его активное участие в обработке трафика можно заметить только на регулируемых перекрестках, где светофор определяет очередность пересечения перекрестка потоками автомобилей.

Как и в сетях с коммутацией пакетов, к образованию заторов на дорогах приводит неравномерность движения автомобилей. Так, даже кратковременное снижение скорости одного автомобиля на узкой дороге может создать большую пробку, которой бы не было, если бы все автомобили всегда двигались с одной и той же скоростью и равными интервалами [2].

А теперь попробуем найти общее у автомобильного движения и сетей *с коммутацией каналов*.

Иногда на дороге возникает ситуация, когда нужно обеспечить особые условия для движения колонны автомобилей. Например, представим, что очень длинная колонна автобусов перевозит детей из города в летний лагерь по многополосному шоссе. Для того чтобы колонна двигалась без препятствий, для ее движения заранее разрабатывается маршрут.

Затем на протяжении всего этого маршрута, который пересекает несколько перекрестков, для колонны выделяется отдельная полоса на всех отрезках шоссе. При этом полоса освобождается от другого трафика еще за некоторое время до начала движения колонны, и это резервирование отменяется только после того, как колонна достигает пункта назначения.

Во время движения все автомобили колонны едут с одинаковой скоростью и приблизительно равными интервалами между собой, не создавая препятствий друг другу. Очевидно, что для колонны автомобилей создаются наиболее благоприятные условия движения, но дорога при такой организации движения используется нерационально, так как полоса простаивает значительную часть времени, как и полоса пропускания в сетях с коммутацией каналов.

2.3.2. Количественное сравнение задержек

Вернемся от автомобилей к сетевому трафику. Пусть пользователю сети необходимо передать достаточно неравномерный трафик, состоящий из периодов активности и пауз. Представим также, что он может выбрать, через какую сеть, с КК или КП, передавать свой трафик, причем в обеих сетях производительность каналов связи одинакова. Очевидно, что более эффективной с точки зрения временны́х затрат для нашего пользователя была бы работа в сети с коммутацией каналов, где ему в единоличное владение предоставляется зарезервированный канал связи. При этом способе все данные поступали бы адресату без задержки. Тот факт, что значительную часть времени зарезервированный канал будет простаивать (во время пауз), нашего пользователя не волнует – ему важно быстро решить собственную задачу.

Если бы пользователь обратился к услугам сети с коммутацией пакетов, то процесс передачи данных оказался бы более медленным, так как его пакеты, вероятно, не раз задерживались бы в очередях, ожидая освобождения необходимых сетевых ресурсов наравне с пакетами других абонентов.

Давайте рассмотрим более детально механизм возникновения задержек при передаче данных в сетях обоих типов. Пусть от конечного узла $N1$ отправляется сообщение к конечному узлу $N2$ (рис. 2.11). На пути передачи данных расположены два коммутатора [2].

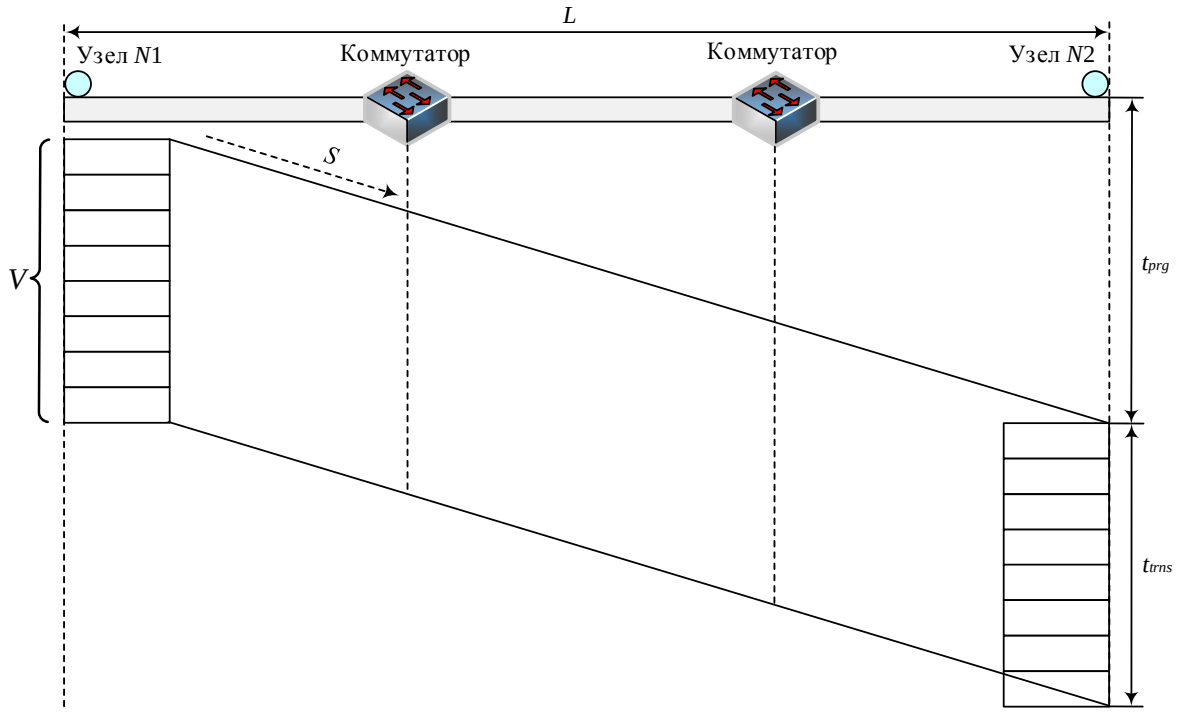


Рис. 2.11. Временная диаграмма передачи сообщения
в сети с коммутацией каналов

В сети с коммутацией каналов данные после задержки, связанной с установлением канала, начинают передаваться на стандартной для канала скорости. Время доставки данных *T* адресату равно сумме *времени распространения сигнала в канале* t_{prg} и *времени передачи сообщения в канал* (называемом также *временем сериализации*) t_{trns}.

Наличие коммутаторов в сети с коммутацией каналов никак не влияет на суммарное время прохождения данных через сеть.

Заметим, что время передачи сообщения в канал в точности совпадает со временем приема сообщения из канала в буфер узла назначения, т. е. временем буферизации.

Время распространения сигнала зависит от расстояния между абонентами *L* и скорости *S* распространения электромагнитных волн в конкретной физической среде, которая колеблется от 0,6 до 0,9 скорости света в вакууме:

$$t_{prg} = L/S\ .$$

Время передачи сообщения в канал (а значит, и время буферизации в узле назначения) равно отношению объема сообщения *V* в битах к пропускной способности канала *C* в битах в секунду:

$$t_{trns} = V/C\ .$$

В сети с коммутацией пакетов передача данных не требует обязательного установления соединения. Предположим, что в сеть, показанную на рис. 2.12, передается сообщение того же объема *V*, что и в предыдущем случае (рис. 2.11), однако оно разделено на пакеты, каждый из которых снабжен заголовком. Пакеты передаются от узла *N*1 узлу *N*2, между которыми расположены два коммутатора. На каждом коммутаторе каждый пакет изображен дважды: в момент прихода на входной интерфейс и в момент передачи в сеть с выходного интерфейса. Из рис. 2.12 видно, что коммутатор задерживает пакет на некоторое время. Здесь T_1 – время доставки адресату первого пакета сообщения, а T_{ps} – всего сообщения [2].

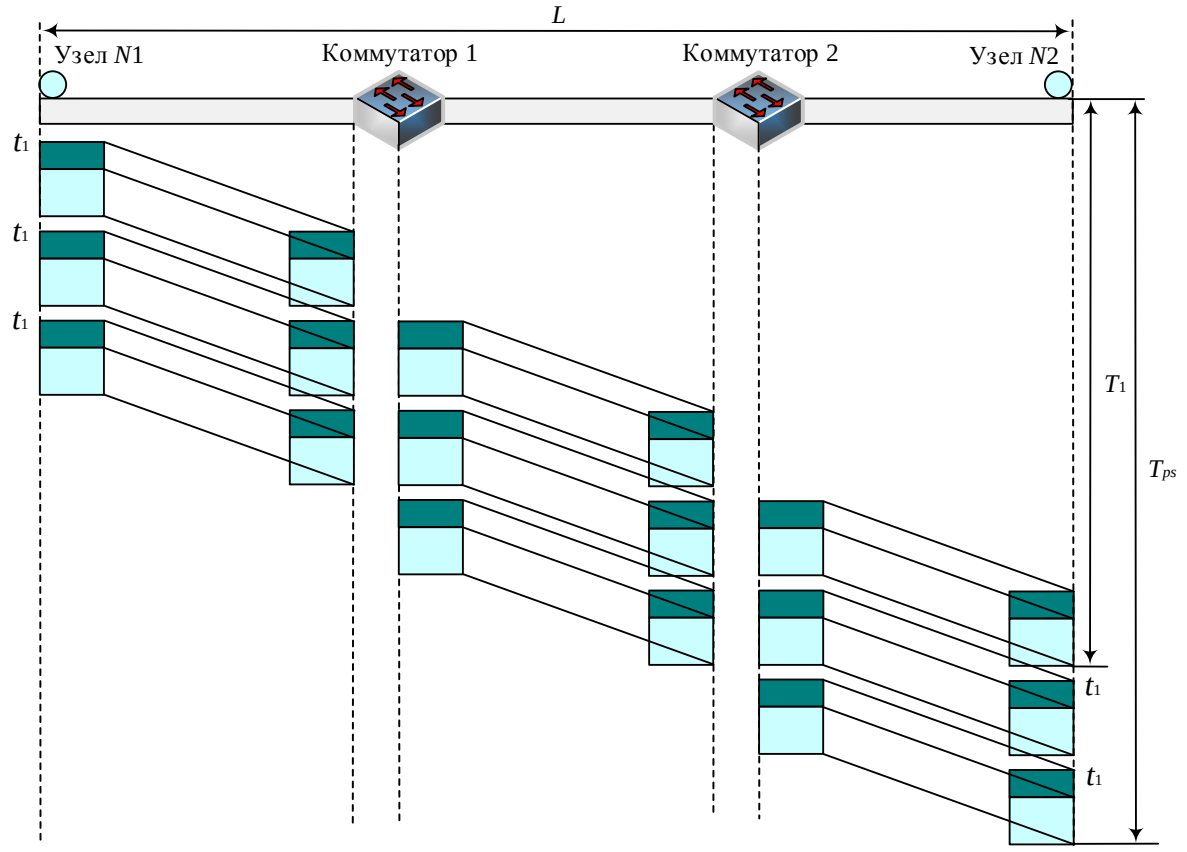


Рис. 2.12. Временная диаграмма передачи сообщения, разделенного на пакеты в сети с коммутацией пакетов

Сравнивая временны́е диаграммы передачи данных в сетях с коммутацией каналов и пакетов отметим два факта: а) значения времени распространения сигнала t_{prg} в одинаковой физической среде на одно и то же расстояние одинаковы; б) учитывая, что значения пропускной способности каналов в обеих сетях одинаковы, значения времени передачи сообщения в канал t_{trns} будут также равны.

Однако разбиение передаваемого сообщения на пакеты с последующей их передачей по сети с коммутацией пакетов приводит к дополнительным

задержкам. Проследим путь первого пакета и отметим, из каких составляющих складывается время его передачи в узел назначения и какие из них специфичны для сети с коммутацией пакетов (рис. 2.13) [2].

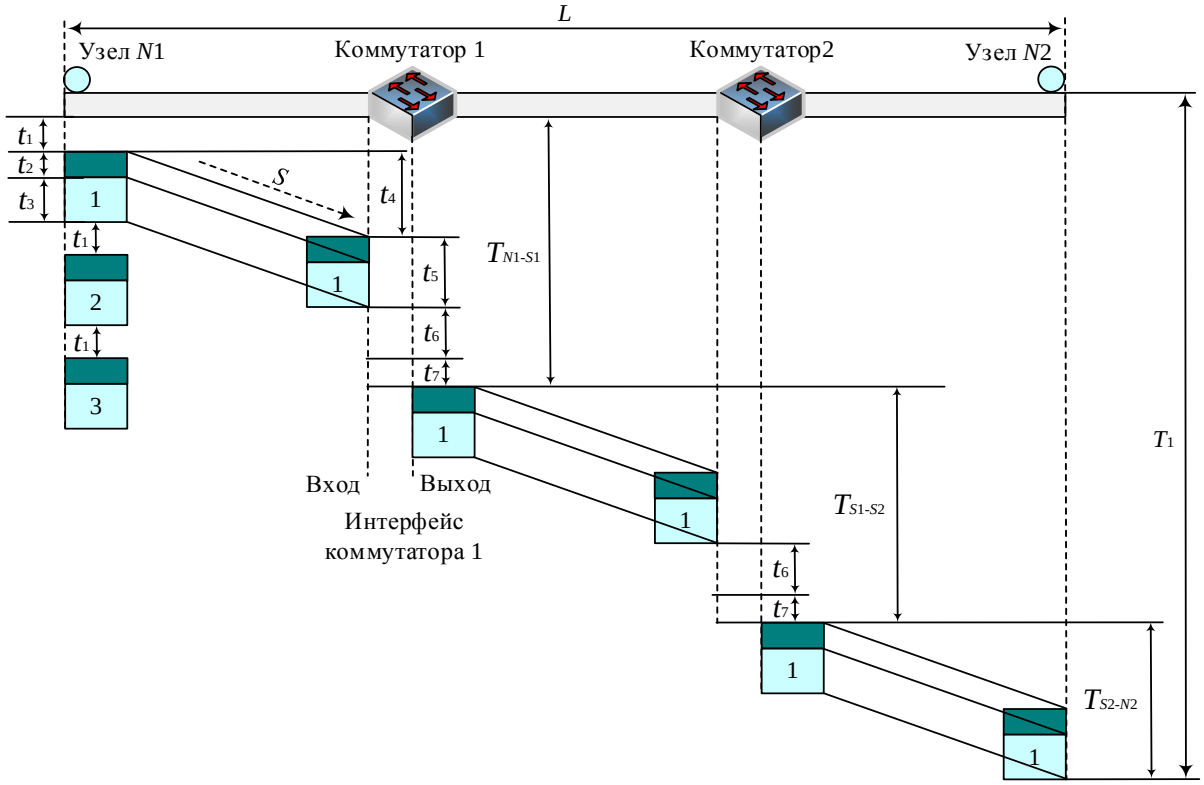


Рис. 2.13. Временная диаграмма передачи
одного пакета в сети с коммутацией пакетов

Время передачи одного пакета от узла $N1$ до коммутатора 1 можно представить в виде суммы нескольких слагаемых.

Во-первых, время тратится в узле-отправителе $N1$: а) t_1 – время формирования пакета, также называемое временем пакетизации (зависит от различных параметров работы программного и аппаратного обеспечения узла-отправителя и не зависит от параметров сети); б) t_2 – время передачи в канал заголовка; в) t_3 – время передачи в канал поля данных пакета.

Во-вторых, дополнительное время тратится на распространение сигналов по каналам связи. Обозначим через t_4 время распространения сигнала, представляющего один бит информации, от узла $N1$ до коммутатора 1.

В-третьих, дополнительное время тратится в промежуточном коммутаторе: а) t_5 – время пакета с его заголовком из канала во входной буфер коммутатора; как уже было отмечено, это время равно $(t_2 + t_3)$, т. е. времени передачи пакета с заголовком в канал из узла источника; б) t_6 – время ожидания пакета в очереди колеблется в очень широких пределах и заранее не известно, так как зависит от текущей загрузки сети; в) t_7 – время комму-

тации пакета при его передаче в выходной порт фиксировано для конкретной модели и обычно невелико (от нескольких мкс до нескольких мс).

Обозначим через T_{N1-S1} время передачи пакета из узла $N1$ на выходной интерфейс коммутатора 1. Это время складывается из следующих составляющих: $T_{N1-S1} = t_1 + t_4 + t_5 + t_6 + t_7$.

Обратим внимание, что среди слагаемых отсутствуют составляющие t_2 и t_3. Из рис. 2.13 видно, что передача битов из передатчика в канал совмещается по времени с передачей битов по каналу связи.

Время, затрачиваемое на оставшиеся два отрезка пути, обозначим соответственно T_{S1-S2} и T_{S2-N2}. Эти величины имеют такую же структуру, что и T_{N1-S1}, за исключением того, что в них не входит время пакетизации, и кроме того, T_{S2-N2} не включает время коммутации (так как отрезок заканчивается конечным узлом). Итак, полное время передачи одного пакета по сети составляет: $T_1 = T_{N1-S1} + T_{S1-S2} + T_{S2-N2}$.

А чему же будет равно время передачи сообщения, состоящего из нескольких пакетов? Сумме времен передачи каждого пакета? Конечно, нет! Ведь сеть с коммутацией пакетов работает как конвейер (рис. 2.12): пакет обрабатывается в несколько этапов, и все устройства сети выполняют эти этапы параллельно. Поэтому время передачи такого сообщения будет значительно меньше, чем сумма значений времени передачи каждого пакета сообщения. Если предположить, что пакеты стоят в очереди примерно одинаковое время, то общее *время передачи сообщения, состоящего из n пакетов*, можно оценить следующим образом: $T_{PS} = T_1 + (n-1)(t_1 + t_5)$.

Пример. Используем для сравнения эффективности сетей с коммутацией каналов и пакетов пример на рис. 2.14. Два коммутатора объединены каналом связи с пропускной способностью 100 Мбит/с. Пользователи подключаются к сети с помощью каналов доступа (*access link*) с пропускной способностью 10 Мбит/с. Предположим, что все пользователи создают одинаковый пульсирующий трафик со средней скоростью 1 Мбит/с. При этом в течение непродолжительных периодов времени скорость данной предложенной нагрузки возрастает до максимальной скорости канала доступа, т. е. до 10 Мбит/с. Такие периоды длятся не более 1 секунды. Предположим также, что все пользователи, подключенные к коммутатору *S*1, передают информацию только пользователям, подключенным к коммутатору *S*2.

Пусть представленная на рис. 2.14 сеть является сетью с коммутацией каналов. Поскольку пики пользовательского трафика достигают 10 Мбит/с, каждому из пользователей необходимо установить соединение с пропускной способностью 10 Мбит/с. Таким образом, одновременно через сеть смогут передавать данные только 10 пользователей. Суммарная средняя скорость передачи данных через сеть будет равна только 10 Мбит/с (10 пользователей передают данные

со средней скоростью 1 Мбит/с). Следовательно, линия связи между коммутаторами хотя и имеет общую пропускную способность 100 Мбит/с, используется только на 10 %.

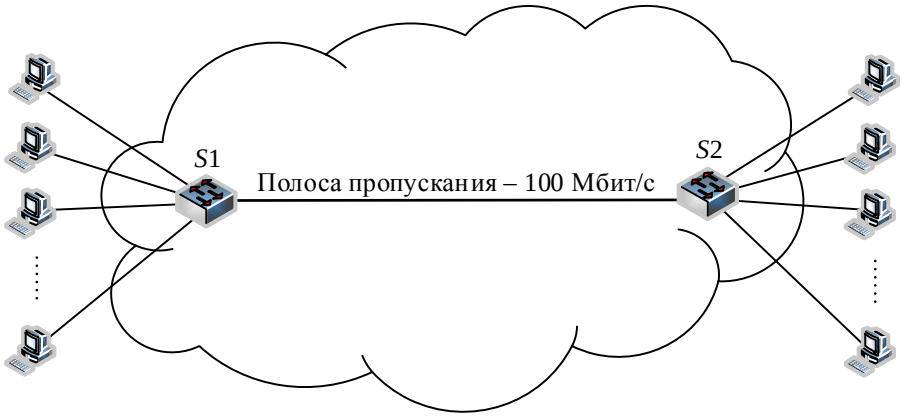


Рис. 2.14. Сравнение эффективности сетей
с коммутацией пакетов и каналов

Теперь рассмотрим вариант, когда та же сеть работает на основе техники коммутации пакетов. При средней скорости пользовательских потоков 1 Мбит/с сеть может передавать одновременно до 100/1 = 100 (!) информационных потоков пользователей, полностью расходуя пропускную способность канала между коммутаторами. Однако это справедливо, только если емкости буферов коммутаторов достаточно для хранения пакетов на периодах перегрузки, когда суммарная скорость потока данных превышает 100 Мбит/с. Оценим необходимый объем буфера коммутатора *S*1. За период перегрузки в коммутатор *S*1 от каждого потока поступит 10 Мбит/с × 1 с = 10 Мбит, а от 100 потоков – 1000 Мбит. Из этих данных за 1 с коммутатор успеет передать в выходной канал только 100 Мбит. Значит, чтобы ни один пакет не был потерян во время перегрузки сети, общий объем входных буферов коммутатора должен быть не меньше 1000 – 100 = 900 Мбит, или более 100 Мбайт. Сегодняшние коммутаторы обычно имеют меньшие объемы буферов (1–10 Мбайт). Однако не нужно забывать, что вероятность совпадения периодов пиковой нагрузки у всех потоков, поступающих на входы коммутатора, очень мала. Так что даже если коммутатор имеет меньший объем буферной памяти, в подавляющем большинстве случаев он будет справляться с предложенной нагрузкой.

При сравнении сетей с коммутацией каналов и пакетов уместна аналогия с мультипрограммными операционными системами. Каждая отдельная программа в такой системе выполняется дольше, чем в однопрограммной системе, когда программе выделяется все процессорное время, пока она не завершит свое выполнение. Однако общее число программ, выполняемых в единицу времени, в мультипрограммной системе больше, чем в однопрограммной. Аналогично однопрограммной системе, в которой время

от времени простаивают процессор или периферийные устройства, в сетях с коммутацией каналов при передаче пульсирующего трафика значительная часть зарезервированной пропускной способности каналов часто не используется.

Неопределенная пропускная способность сети с коммутацией пакетов – это плата за ее общую эффективность при некотором ущемлении интересов отдельных абонентов.

В заключении приведем табл. 2.3, в которой сведены свойства обоих видов сетей. На основании этих данных можно аргументированно утверждать, в каких случаях рациональнее использовать сети с коммутацией каналов, а в каких – с коммутацией пакетов.

Таблица 2.3

Сравнение сетей с коммутацией каналов и пакетов

Коммутация каналов	Коммутация пакетов
Необходимо предварительно устанавливать соединение	Отсутствует этап установления соединения (дейтаграммный способ)
Адрес требуется только на этапе установления соединения	Адрес и другая служебная информация передаются с каждым пакетом
Сеть может отказать абоненту в установлении соединения	Сеть всегда готова принять данные от абонента
Гарантированная пропускная способность (полоса пропускания) для взаимодействующих абонентов	Пропускная способность сети для абонентов неизвестна, задержки передачи носят случайный характер
Трафик реального времени передается без задержек	Ресурсы сети используются эффективно при передаче пульсирующего трафика
Высокая надежность передачи	Возможны потери данных из-за переполнения буферов
Нерациональное использование пропускной способности каналов, снижающее общую эффективность сети	Автоматическое динамическое распределение пропускной способности физического канала между абонентами

2.3.3. Ethernet – пример технологии с коммутацией пакетов

Рассмотрим, каким образом описанные ранее концепции воплощены в одной из первых стандартных сетевых технологий – технологии Ethernet, работающей с битовой скоростью 10 Мбит/с.

Топология. Существуют два варианта технологии Ethernet: Ethernet на разделяемой среде и коммутируемый вариант Ethernet. В первом случае все узлы сети разделяют общую среду передачи данных и сеть строится по топологии общей шины. На рис. 2.15 показан простейший вариант

Рис. 2.15. Сеть Ethernet на разделяемой среде

топологии – все компьютеры сети подключены к общей разделяемой среде, состоящей из одного сегмента коаксиального кабеля [2].

В том случае, когда сеть Ethernet не использует разделяемую среду, а строится на коммутаторах, объединенных дуплексными каналами связи, говорят о коммутируемом варианте Ethernet. Топология в этом случае является топологией дерева, т. е. такой, при которой между двумя любыми узлами сети существует ровно один путь. Пример топологии коммутируемой сети Ethernet показан на рис. 2.16.

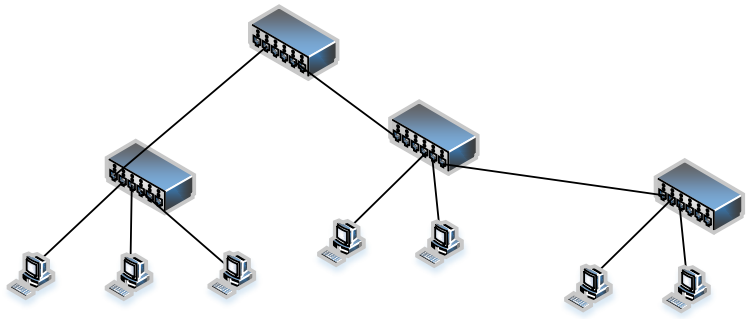


Рис. 2.16. Древовидная топология
коммутируемой сети Ethernet

Топологические ограничения (только древовидная структура связей коммутаторов) связаны со способом построения таблиц продвижения Ethernet-коммутаторами.

Способ коммутации. В технологии Ethernet используется дейтаграммная коммутация пакетов. Единицы данных, которыми обмениваются компьютеры в сети Ethernet, называются кадрами. Кадр имеет фиксированный формат и наряду с полем данных содержит различную служебную информацию. В том случае, когда сеть Ethernet построена на коммутаторах, каждый коммутатор продвигает кадры в соответствии с описанными ранее принципами коммутации пакетов. А в случае односегментной сети Ethernet возникает вопрос: где же выполняется коммутация? В Ethernet такой «коммутатор» состоит из сетевых адаптеров и разделяемой среды. Сетевые адаптеры представляют собой интерфейсы этого виртуального коммутатора, а разделяемая среда – коммутационный блок, который передает кадры между интерфейсами. Часть функций коммутационного блока выполняют адаптеры, так как они решают, какой кадр адресован их компьютеру, а какой – нет [2].

Адресация. Каждый компьютер, а точнее каждый сетевой адаптер, имеет уникальный аппаратный адрес (MAC-адрес). Ethernet-адрес является плоским числовым адресом, иерархия здесь не используется. Поддерживаются адреса для выборочной, широковещательной и групповой рассылки.

Разделение среды и мультиплексирование. В сети Ethernet на коммутаторах каждый канал является дуплексным каналом связи, и проблемы его разделения между интерфейсами узлов не возникает. Передатчики Ethernet-коммутаторов используют дуплексные каналы связи для мультиплексирования потоков кадров от разных оконечных узлов.

В случае Ethernet на разделяемой среде конечные узлы применяют специальный метод доступа с целью синхронизации использования единственного полудуплексного канала связи, объединяющего все компьютеры сети. Единого арбитра в сети Ethernet на разделяемой среде нет, вместо этого все узлы прибегают к распределенному случайному методу доступа. Информационные потоки, поступающие от конечных узлов сети Ethernet, мультиплексируются в единственном передающем канале в режиме разделения времени. т. е. кадрам разных потоков поочередно предоставляется канал.

Чтобы подчеркнуть не всегда очевидную разницу между понятиями мультиплексирования и разделения среды, рассмотрим ситуацию, когда из всех компьютеров сети Ethernet только одному нужно передавать данные, причем данные нескольких приложений. В этом случае проблема разделения среды между сетевыми интерфейсами не возникает, в то время как задача передачи нескольких информационных потоков по общей линии связи (т. е. мультиплексирование) остается.

Кодирование. Адаптеры в Ethernet работают с тактовой частотой 20 МГц, передавая в среду прямоугольные импульсы, соответствующие единицам и нулям данных компьютера. Когда начинается передача кадра, все его биты передаются в сеть с постоянной скоростью 10 Мбит/с. Эта скорость определяется пропускной способностью линии связи в сети Ethernet.

Надежность. Для повышения надежности передачи данных в Ethernet используется стандартный прием – подсчет контрольной суммы и передача ее в концевике кадра. Если принимающий адаптер путем подсчета контрольной суммы обнаруживает ошибку, то такой кадр отбрасывается.

Очереди. В коммутируемых сетях Ethernet очереди кадров, готовых к отправке, организуются обычным для сетей с коммутацией пакетов способом, т. е. с помощью буферной памяти интерфейсов коммутатора.

В сетях Ethernet на разделяемой среде коммутаторы отсутствуют. Однако отсутствие коммутатора с буферной памятью в сети Ethernet не означает, что очередей в ней нет. Просто здесь очереди переместились в буферную память сетевого адаптера. В те периоды времени, когда среда занята передачей кадров других сетевых адаптеров, данные (предложенная нагрузка) по-прежнему поступают в сетевой адаптер. Так как они не могут быть переданы в это время в сеть, они начинают накапливаться во внутреннем буфере Ethernet-адаптера, образуя очередь. Поэтому в сети Ethernet существуют переменные задержки доставки кадров, как и во всех сетях с коммутацией пакетов.

Выводы

В сети с коммутацией каналов данные после задержки, связанной с установлением канала, начинают передаваться на стандартной для канала скорости. В сети с коммутацией пакетов передача данных не требует обязательного установления соединения.

Разбиение передаваемого сообщения на пакеты с последующей их передачей по сети с коммутацией пакетов приводит к дополнительным задержкам. Неопределенная пропускная способность сети с коммутацией пакетов – это плата за ее общую эффективность при некотором ущемлении интересов отдельных абонентов.

Контрольные вопросы

1. Из чего складывается задержка передачи в сети с коммутацией каналов?
2. Из чего складывается задержка передачи в сети с коммутацией пакетов?
3. Приведите аналогии при сравнении сетей с коммутацией каналов и пакетов.
4. Какой способ мультиплексирования используется в сетях Ethernet?
5. Что является коммутатором в односегментной сети Ethernet на разделяемой среде? Варианты ответов:

а) разделяемая среда;

б) сетевые адаптеры;

в) разделяемая среда и сетевые адаптеры.

3. АРХИТЕКТУРА И СТАНДАРТИЗАЦИЯ СЕТЕЙ

3.1. Модель OSI

Сетевая архитектура – это концептуальная схема функционирования компьютерной сети, определяющая принципы работы сетевых компонентов, организацию их связей, протоколы взаимодействия и способы физической передачи данных. Архитектура сети отражает декомпозицию общей задачи сетевого взаимодействия на отдельные подзадачи конечных узлов (компьютеров) и промежуточных узлов (коммутаторов и маршрутизаторов).

3.1.1. Многоуровневый подход

Для сетевого взаимодействия используется известный универсальный прием – *декомпозиция,* которая состоит в четком определении функций каждого модуля, а также порядка их взаимодействия. При таком подходе каждый модуль можно рассматривать как «черный ящик», абстрагируясь от его внутренних механизмов и концентрируя внимание на способе взаимодействия модулей. В результате такого логического упрощения появляется возможность независимого тестирования, разработки и модификации модулей. Так, любой из показанных на рис. 3.1 модулей может быть переписан заново. Пусть, например, это будет модуль *A,* и если при этом разработчики сохранят без изменений межмодульные связи (в данном случае интерфейсы *A*-*B* и *A*-*C*), то это не потребует никаких изменений в остальных модулях.

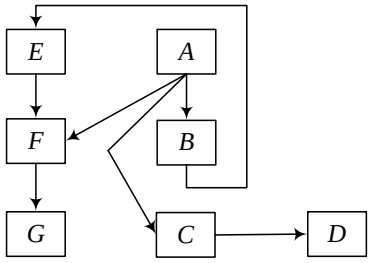


Рис. 3.1. Пример декомпозиции задачи

Еще более эффективной концепцией, развивающей идею декомпозиции, является *многоуровневый подход*. После представления исходной задачи в виде множества модулей эти модули группируют и упорядочивают по уровням, образующим иерархию. В соответствии с иерархией для каждого промежуточного уровня можно указать непосредственно примыкающие к нему соседние вышележащий и нижележащий уровни (рис. 3.2) [2].

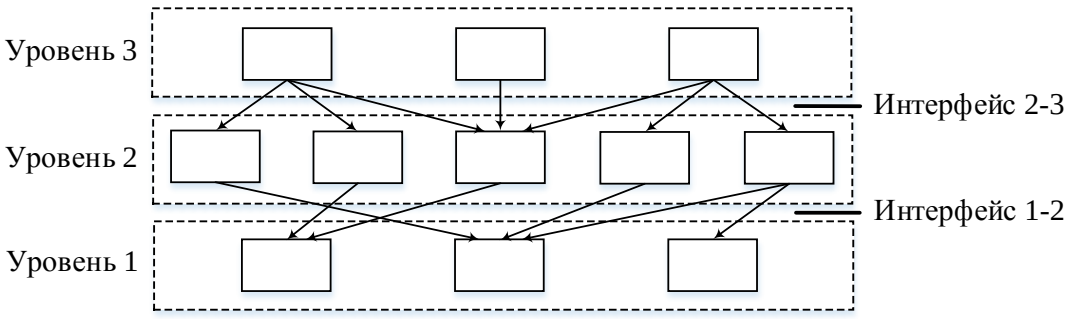


Рис. 3.2. Многоуровневый подход – создание иерархии задач

С одной стороны, группа модулей, составляющих каждый уровень для решения своих задач должна обращаться с запросами только к модулям соседнего нижележащего уровня. С другой стороны, результаты работы каждого из модулей, отнесенных к некоторому уровню, могут быть переданы только модулям соседнего вышележащего уровня. Такая иерархическая декомпозиция задачи предполагает четкое определение функций и интерфейсов не только отдельных модулей, но и каждого уровня.

Межуровневый интерфейс, называемый также *интерфейсом услуг*, определяет набор функций, которые нижележащий уровень предоставляет вышележащему (рис. 3.3) [2].

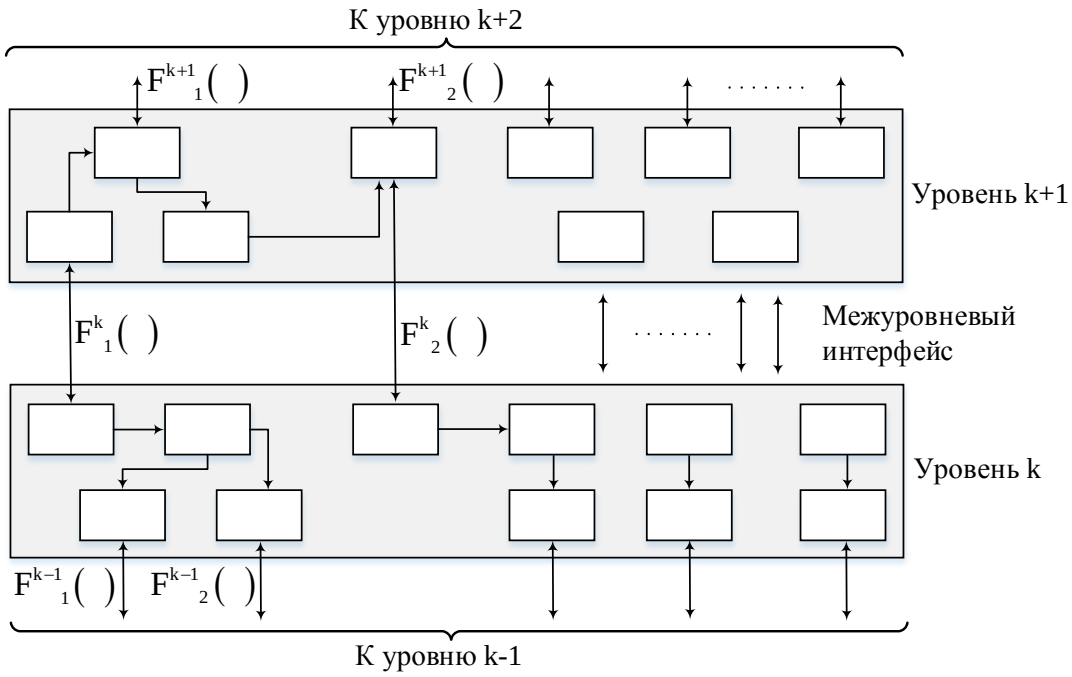


Рис. 3.3. Концепция многоуровневого взаимодействия

Взаимодействие компьютеров в сети тоже может быть представлено в виде иерархически организованного множества модулей. Например, модулям нижнего уровня можно поручить вопросы надежной передачи информации между двумя соседними узлами, а модулям следующего, более высокого уровня – транспортировку сообщений в пределах всей сети. Очевидно, что последняя задача является более общей, и ее решение может быть получено путем многократных обращений к нижележащему уровню. Так, организация взаимодействия узлов *A* и *B* может быть сведена к поочередному взаимодействию пар промежуточных смежных узлов (рис. 3.4).

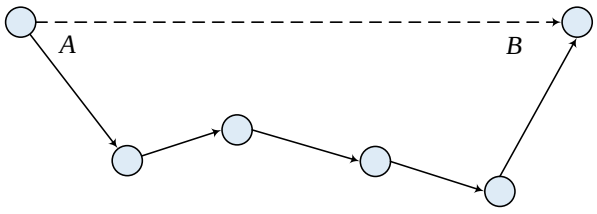


Рис. 3.4. Взаимодействие произвольной пары узлов

3.1.2. Протокол и стек протоколов

Многоуровневое представление средств сетевого взаимодействия имеет свою специфику, связанную с тем, что в процессе обмена сообщениями участвуют по меньшей мере две стороны, т. е. в данном случае необходимо организовать согласованную работу двух иерархий аппаратных и программных средств на разных компьютерах. Оба участника сетевого обмена должны принять множество соглашений. Например, они должны согласовать уровни и форму электрических сигналов, способ определения размера сообщений, договориться о методах контроля достоверности и т. п. Другими словами, соглашения должны быть приняты на всех уровнях, начиная от самого низкого – уровня передачи битов, и заканчивая самым высоким, реализующим обслуживание пользователей сети.

На рис. 3.5 показана модель взаимодействия двух узлов. С каждой стороны средства взаимодействия представлены четырьмя уровнями. Каждый уровень поддерживает интерфейсы двух типов. Во-первых, это интерфейсы услуг с выше- и нижележащим уровнями «своей» иерархии средств. Во-вторых, это интерфейс со средствами взаимодействия другой стороны, расположенными на том же уровне иерархии. Этот тип интерфейса называют *протоколом,* т. е. протокол всегда является одноранговым интерфейсом [2].

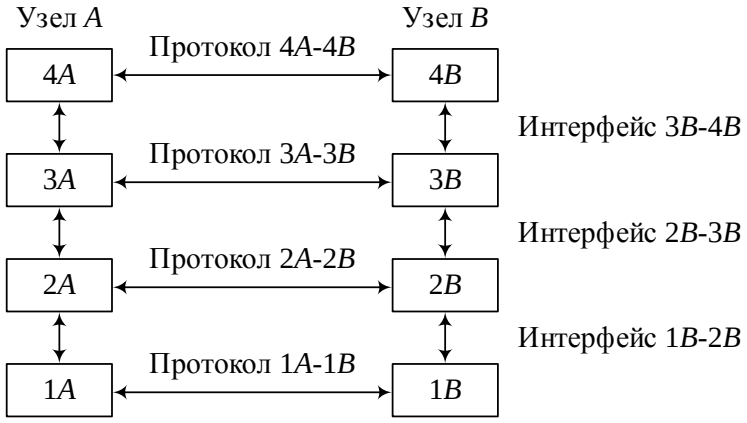


Рис. 3.5. Взаимодействие двух узлов

В сущности, термины «протокол» и «интерфейс» выражают одно и то же понятие – формализованное описание процедуры взаимодействия двух объектов, но традиционно в сетях за ними закрепили разные области действия: протоколы определяют правила взаимодействия модулей одного уровня в разных узлах, а интерфейсы – правила взаимодействия модулей соседних уровней в одном узле.

Иерархически организованный набор протоколов, достаточный для организации взаимодействия узлов в сети, называется *стеком протоколов.*

Протоколы нижних уровней часто реализуются комбинацией программных и аппаратных средств, а протоколы верхних уровней, как правило, программными средствами.

Программный модуль, реализующий некоторый протокол, называют протоколом. Один и тот же протокол может быть реализован с разной степенью эффективности, поэтому при сравнении протоколов следует учитывать не только логику их работы, но и качество программной реализации. Более того, на эффективность взаимодействия устройств в сети влияет качество всей совокупности протоколов, составляющих стек, в частности то, *насколько рационально распределены функции между протоколами* разных уровней и насколько хорошо определены интерфейсы между ними.

Протокольные модули одного уровня двух взаимодействующих сторон обмениваются сообщениями в соответствии с определенным для них протоколом. Сообщения состоят из заголовка и поля данных (иногда оно может отсутствовать). Обмен сообщениями является своеобразным языком общения, с помощью которого каждая из сторон «объясняет» другой стороне, что необходимо сделать на каждом этапе взаимодействия. Работа каждого протокольного модуля состоит в интерпретации заголовков, поступающих к нему сообщений и выполнении соответствующих действий. Заголовки сообщений имеют разную структуру: чем сложнее структура заголовка сообщения, тем более сложные функции возложены на соответствующий протокол.

3.1.3. Общая характеристика модели OSI

Из того что протокол является соглашением, принятым двумя взаимодействующими узлами сети, совсем не следует, что он обязательно является стандартным. Но на практике при реализации сетей стремятся использовать стандартные протоколы. Это могут быть фирменные, национальные или международные стандарты.

В начале 1980-х гг. ряд международных организаций по стандартизации, в частности, International Organization for Standartization (ISO), а также International Telecommunications Union (ITU) разработали стандартную модель взаимодействия открытых систем (Open System Interconnection, OSI). Эта модель сыграла значительную роль в развитии компьютерных сетей.

К концу 1970-х гг. в мире уже существовало большое количество фирменных стеков коммуникационных протоколов, среди которых можно назвать, например, DECnet, TCP/IP и IBM SNA. Подобное разнообразие средств межсетевого взаимодействия вывело на первый план проблему совместимости устройств, использующих разные протоколы. Одним из путей разрешения этой проблемы в то время виделся всеобщий переход на единый, общий для всех систем стек протоколов, созданный с учетом недостатков уже существующих стеков. Такой академический подход к созданию нового стека начался с разработки модели OSI и занял 7 лет (с 1977 по 1984 гг.). Назначение модели OSI состоит в обобщенном представлении средств сетевого взаимодействия. Она разрабатывалась в качестве универсального языка сетевых специалистов, поэтому ее называют справочной моделью.

Модель OSI определяет, во-первых, уровни взаимодействия систем в сетях с коммутацией пакетов, во-вторых, стандартные названия уровней, в-третьих, функции, которые должен выполнять каждый уровень. Модель OSI не содержит описаний реализаций конкретного набора протоколов.

В модели OSI средства взаимодействия делятся на семь уровней: прикладной, представления, сеансовый, транспортный, сетевой, канальный и физический (рис. 3.6). Каждый уровень связан с совершенно определенным аспектом взаимодействия сетевых устройств [2].

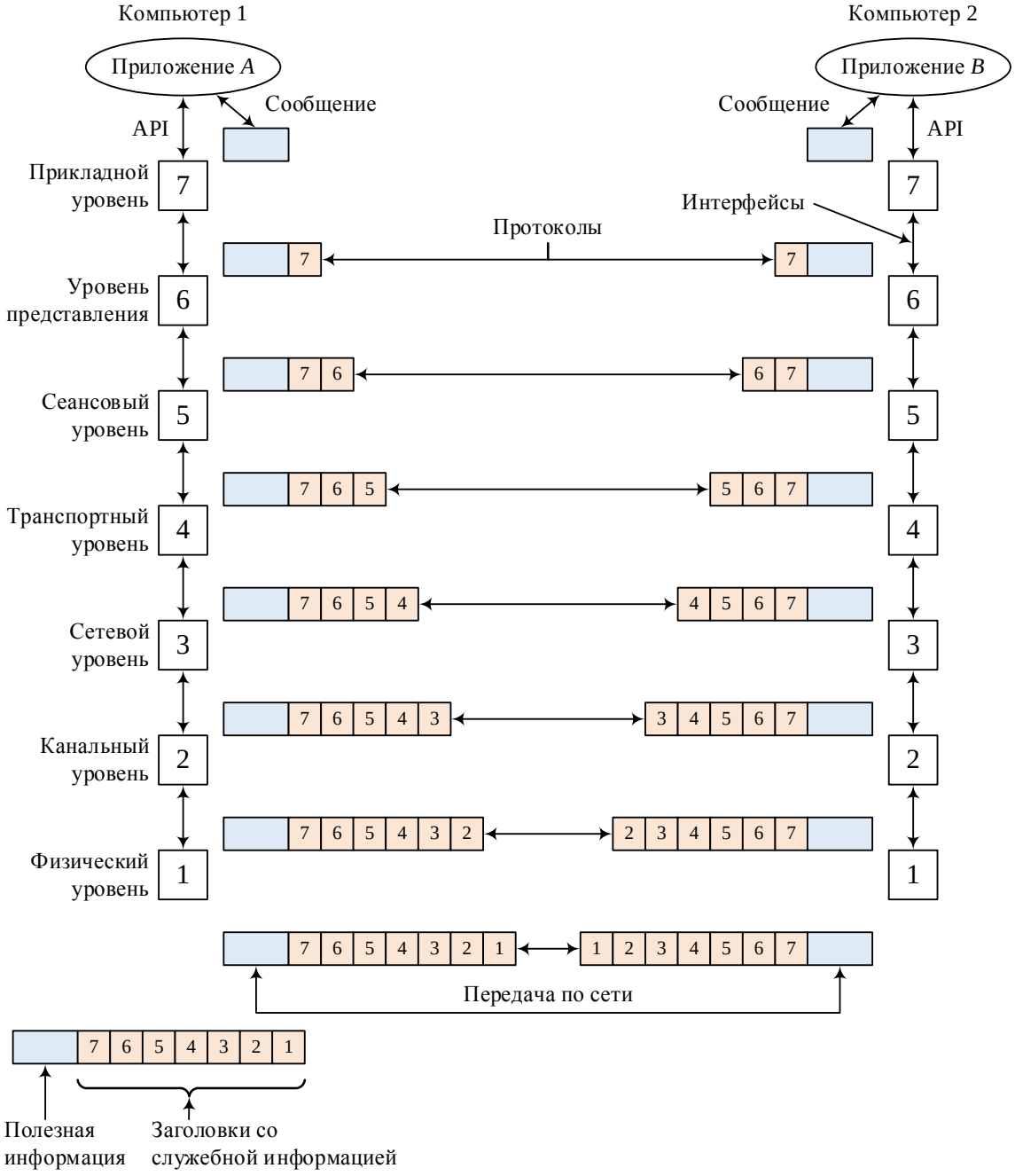


Рис. 3.6. Модель взаимодействия открытых систем ISO/OSI

Модель OSI описывает только системные средства взаимодействия, реализуемые операционной системой, системными утилитами, системными аппаратными средствами. Модель не включает средства взаимодействия приложений конечных пользователей. Важно различать уровень взаимодействия приложений и прикладной уровень семиуровневой модели.

Приложения могут реализовывать собственные протоколы взаимодействия, используя для этих целей многоуровневую совокупность системных средств. Именно для этого в распоряжение программистов предоставляется прикладной программный интерфейс (Application Program Interface, API). В соответствии с идеальной схемой модели OSI приложение может обращаться с запросами только к самому верхнему уровню – прикладному, однако на практике многие стеки коммуникационных протоколов предоставляют возможность программистам напрямую обращаться к сервисам или службам нижележащих уровней.

Допустим, приложение узла *A* хочет взаимодействовать с приложением узла *B*. Для этого приложение *A* обращается с запросом к прикладному уровню, например, к файловой службе. На основании этого запроса ПО прикладного уровня формирует сообщение стандартного формата. Но для того, чтобы доставить эту информацию по назначению, предстоит решить еще много задач, ответственность за которые несут нижележащие уровни [2].

После формирования сообщения прикладной уровень направляет его уровню представления вниз по стеку. Протокол уровня представления на основании информации, полученной из заголовка сообщения прикладного уровня, выполняет требуемые действия и добавляет к сообщению собственную служебную информацию – заголовок уровня представления, в котором содержатся указания для протокола уровня представления машины-адресата. Полученное в результате сообщение передается вниз сеансовому уровню, который, в свою очередь, добавляет свой заголовок и т. д. (некоторые реализации протоколов помещают служебную информацию не только в начале сообщения в виде заголовка, но и в конце в виде так называемого концевика). Наконец, сообщение достигает нижнего, физического, уровня, который, собственно, и передает его по линиям связи машине-адресату. К этому моменту сообщение «обрастает» заголовками всех уровней (рис. 3.7).

Физический уровень помещает сообщение на физический выходной интерфейс компьютера 1, и оно начинает свое «путешествие» по сети (до этого момента сообщение передавалось от одного уровня у другому в пределах компьютера 1).

Рис. 3.7. Вложенность сообщений различных уровней

Когда сообщение по сети поступает на входной интерфейс компьютера 2, оно принимается его физическим уровнем и последовательно перемещается вверх с уровня на уровень. Каждый уровень анализирует и обрабатывает заголовок своего уровня, выполняя соответствующие функции, а затем удаляет этот заголовок и передает сообщение вышележащему уровню.

Как видно из описания, протокольные сущности одного уровня не общаются между собой непосредственно, в этом общении всегда участвуют посредники – средства протоколов нижележащих уровней. И только физические уровни различных узлов взаимодействуют непосредственно.

В стандартах ISO для обозначения единиц обмена данными, с которыми имеют дело протоколы разных уровней, используется общее название *протокольная единица данных (Protocol Data Unit, PDU)*. Для обозначения единиц обмена данными конкретных уровней часто используются специальные названия, в частности: сообщение, кадр, пакет, дейтаграмма, сегмент.

3.1.4. Уровни модели OSI

Физический уровень *(physical layer)* имеет дело с передачей потока битов по физическим каналам связи, таким как коаксиальный кабель, витая пара, оптоволоконный кабель или радиоканал. Функции физического уровня реализуются на всех устройствах, подключенных к сети. Со стороны компьютера функции физического уровня выполняются сетевым адаптером или последовательным портом.

Примером протокола физического уровня может служить спецификация 1000Base-T технологии Ethernet, которая определяет в качестве используемого кабеля неэкранированную витую пару категории 5 с волновым сопротивлением 100 Ом, разъем RJ-45, максимальную длину физического сегмента 100 метров, манчестерский код для представления данных в кабеле, а также некоторые другие характеристики среды и электрических сигналов.

Физический уровень не вникает в смысл информации, которую он передает. Для него эта информация представляет собой однородный поток битов, который нужно доставить без искажений и в соответствии с заданной тактовой частотой (интервалом между соседними битами) [2].

Канальный уровень *(data link layer)* обеспечивает прозрачность соединения для сетевого уровня. Для этого он предлагает ему следующие услуги: а) установление логического соединения между взаимодействующими узлами; б) согласование в рамках соединения скоростей передатчика и приемника информации; в) обеспечение надежной передачи, обнаружение и коррекция ошибок.

Для решения этих задач канальный уровень формирует из пакетов собственные протокольные единицы – *кадры*, состоящие из поля данных и заголовка, затем помещает пакет в поле данных одного или нескольких кадров и заполняет собственной служебной информацией заголовок кадра.

В сетях, построенных на основе разделяемой среды, канальный уровень выполняет еще одну функцию – проверяет доступность разделяемой среды. Эту функцию иногда выделяют в отдельный подуровень *управления доступом к среде* (Medium Access Control, MAC). Протоколы канального уровня реализуются как на оконечных узлах (средствами сетевых адаптеров и их драйверов), так и на всех промежуточных сетевых устройствах.

Канальный уровень, получив пакет от сетевого уровня, создает кадр, который имеет поле данных и заголовок, затем помещает (*инкапсулирует*) пакет в поле данных кадра и заполняет служебной информацией заголовок кадра. Важнейшей информацией заголовка кадра является адрес назначения, на основании которого коммутаторы сети будут продвигать пакет.

Одной из задач канального уровня является *обнаружение и коррекция ошибок*. Канальный уровень может обеспечить надежность передачи, добавляя к кадру контрольную сумму. Контрольная сумма вычисляется по некоторому алгоритму как функция от всех байтов кадра. На стороне получателя канальный уровень снова вычисляет контрольную сумму и сравнивает результат с контрольной суммой, переданной в кадре. Если они совпадают, кадр считается правильным; в противном случае фиксируется ошибка.

В функции канального уровня входит не только обнаружение ошибок, но и их исправление за счет повторной передачи поврежденных кадров. Однако эта функция не является обязательной, и в некоторых реализациях канального уровня она отсутствует, например, в Ethernet.

Прежде чем переправить кадр физическому уровню для непосредственной передачи данных в сеть, канальному уровню может потребоваться решить еще одну важную задачу. Если в сети используется разделяемая среда, то прежде чем физический уровень начнет передавать данные, канальный уровень должен *проверить доступность среды*; функции проверки доступности разделяемой среды иногда выделяют в отдельный подуровень MAC.

Если разделяемая среда освободилась, кадр передается средствами физического уровня в сеть, проходит по каналу связи и поступает в виде последовательности битов на физический уровень узла назначения, где далее полученные биты поступают «наверх» канальному уровню своего узла.

Протокол канального уровня обычно работает в пределах сети, входящей в виде одной из составляющих в более крупную составную сеть, объединенную протоколами сетевого уровня. Адреса, с которыми работает протокол канального уровня, используются для доставки кадров только в пределах этой сети, а для перемещения пакетов между сетями применяются уже адреса следующего, сетевого, уровня.

В локальных сетях канальный уровень поддерживает весьма мощный и законченный набор функций по пересылке сообщений между узлами сети. В некоторых случаях протоколы канального уровня локальных сетей оказываются самодостаточными транспортными средствами и могут допускать работу протоколов прикладного уровня или приложений непосредственно поверх себя без привлечения средств сетевого и транспортного уровней. Тем не менее для качественной передачи сообщений в сетях с произвольной топологией функций канального уровня оказывается недостаточно [2].

Сетевой уровень *(network layer)* служит для образования единой транспортной системы, объединяющей несколько сетей и называемой *составной сетью*, или *интернетом*[6].

Технология, позволяющая соединить в единую сеть множество сетей, в общем случае построенных на основе разных технологий, называется *технологией межсетевого взаимодействия (internetworking).*

На рис. 3.8 показано несколько сетей, каждая из которых использует собственную технологию канального уровня: Ethernet, FDDI, Token Ring, ATM, Frame Relay. На базе этих технологий любая из указанных сетей может связывать между собой любых пользователей, но только *своей* сети, и не способна обеспечить передачу данных в другую сеть. Причина такого положения вещей кроется в существенных отличиях одной технологии от другой. Даже наиболее близкие технологии LAN – Ethernet, FDDI, Token Ring, имеющие одну и ту же систему адресации, отличаются друг от друга форматом используемых кадров и логикой работы протоколов. Еще больше отличий между технологиями LAN и WAN. Во многих технологиях WAN задействована техника предварительного устанавливаемых виртуальных каналов, идентификаторы которых применяются в качестве адресов. Все эти технологии имеют собственные форматы кадров (в технологии ATM кадр даже называется иначе – ячейкой) и, конечно, собственные стеки протоколов.

[6] Не следует путать интернет (со строчной буквы) с Интернетом (с прописной буквы). Интернет – это самая известная реализация составной сети, построенная на основе технологии TCP/IP.

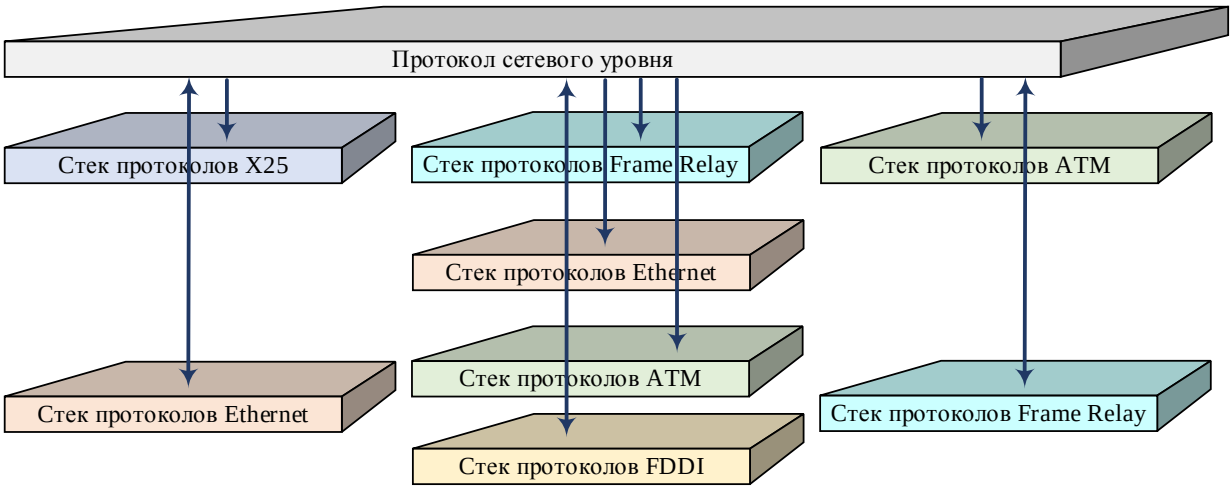


Рис. 3.8. Иллюстрация необходимости сетевого уровня

Чтобы связать между собой сети, построенные на основе столь отличающихся технологий, нужны *дополнительные средства*, и такие средства представляет сетевой уровень. Функции сетевого уровня реализуются группой протоколов и специальными устройствами – *маршрутизаторами*.

Одной из функций маршрутизатора является *физическое соединение сетей*. Маршрутизатор имеет несколько сетевых интерфейсов, к каждому из которых может быть подключена одна сеть. Обычно маршрутизаторы реализуются на базе специализированных аппаратных платформ. В состав ПО маршрутизатора входят протокольные модули сетевого уровня.

Итак, чтобы связать сети, показанные на рис. 3.8, необходимо соединить все эти сети маршрутизаторами и установить протокольные модули сетевого уровня на все конечные узлы пользователей, которые хотели бы связаться через составную сеть (рис. 3.9).

Данные, которые необходимо передать через составную сеть, поступают на сетевой уровень от вышележащего транспортного уровня. Эти данные снабжаются заголовком сетевого уровня. Данные вместе с заголовком образуют *пакет* – так называется PDU сетевого уровня. Заголовок пакета сетевого уровня имеет унифицированный формат, не зависящий от форматов кадров канального уровня тех сетей, которые могут входить в составную сеть, и несет данные об адресе назначения этого пакета.

Для того чтобы протоколы сетевого уровня могли доставлять пакеты любому узлу составной сети, эти узлы должны иметь адреса, уникальные в пределах данной составной сети. Такие адреса называют *сетевыми*, или *глобальными*. Каждый узел составной сети, который намерен обмениваться данными с другими узлами составной сети, наряду с адресом, назначенным ему на канальном уровне, должен иметь сетевой адрес. На рис. 3.9 компьютер в сети Ethernet, входящий в составную сеть, имеет адрес канального уровня MAC1 и адрес сетевого уровня NET-A1; аналогично в сети ATM

узел, адресуемый идентификаторами виртуальных каналов ID1 и ID2, имеет сетевой адрес NET-A2. В пакете в качестве адреса назначения должен быть указан адрес сетевого уровня, на основании которого определяется маршрут пакета. *Определение маршрута* является важной задачей сетевого уровня. Маршрут описывается последовательностью сетей (или маршрутизаторов), через которые должен пройти пакет, чтобы попасть к адресату. Например, на рис. 3.9 штриховой линией показаны три маршрута, по которым могут быть переданы данные от компьютера *A* к компьютеру *B*. Маршрутизатор собирает информацию о топологии связей между сетями и на основе этой информации строит таблицы коммутации, которые в данном случае носят специальное название *таблиц маршрутизации*.

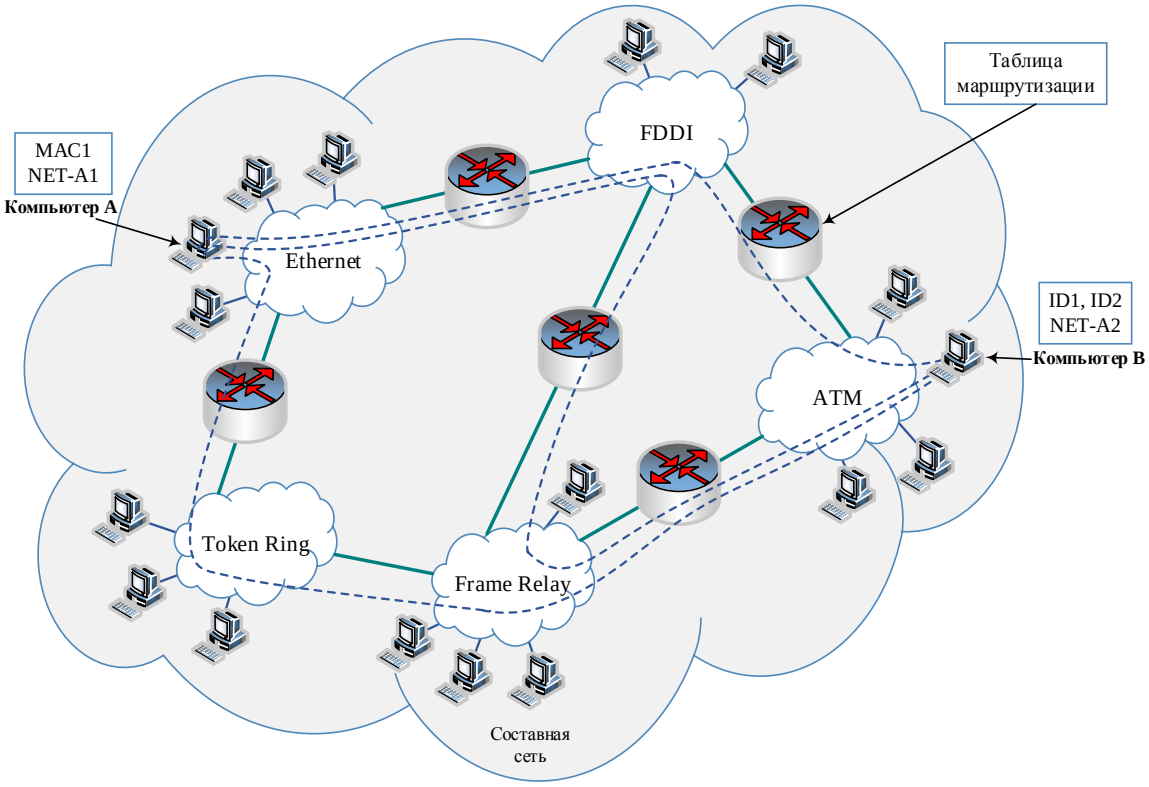


Рис. 3.9. Пример составной сети

В соответствии с многоуровневым подходом сетевой уровень для решения своей задачи обращается к нижележащему канальному уровню. Весь путь через составную сеть разбивается на участки от одного маршрутизатора до другого, причем каждый участок соответствует пути через отдельную сеть. Для того, чтобы передать пакет через очередную сеть, сетевой уровень помещает его в поле данных кадра соответствующей канальной технологии, указывая в заголовке кадра канальный адрес интерфейса следующего маршрутизатора. Сеть, используя свою канальную технологию, доставляет кадр с инкапсулированным в него пакетом по заданному адресу. Маршрутизатор извлекает пакет из прибывшего кадра и после необходимой

обработки передает пакет для дальнейшей транспортировки в следующую сеть, предварительно упаковав его в новый кадр канального уровня (в общем случае) другой технологии. Таким образом, сетевой уровень играет роль координатора совместной работы сетей, построенных на основе разных технологий.

Есть еще одна причина существования сетевого уровня помимо сглаживания различий технологий канального уровня. Сетевой уровень позволяет разбить большую сеть на подсети и управлять каждой из подсетей независимо. Составная сеть с иерархической двухуровневой структурой «канальный уровень» – «сетевой уровень» оказывается более масштабируемой, чем сеть с одноуровневой структурой, что и показала успешная история Интернета. Даже в условиях доминирования одной технологии канального уровня Ethernet построение всемирной сети с единой одноуровневой структурой оказалось практически невозможным. Поэтому сегодня Интернет представляет собой большое количество локальных и глобальных сетей Ethernet, объединенных общим сетевым уровнем, на котором работает протокол IP.

В общем случае функции сетевого уровня шире, чем обеспечение обмена в пределах составной сети. Так, сетевой уровень решает задачу создания надежных и гибких барьеров на пути нежелательного трафика между сетями. На сетевом уровне определяются два вида протоколов. Первый вид – маршрутизируемые протоколы – реализуют продвижение пакетов через сеть. Именно эти протоколы обычно имеют в виду, когда говорят о протоколах сетевого уровня. Однако часто к сетевому уровню относят и другой вид протоколов, называемых маршрутизирующими протоколами, или протоколами маршрутизации. С помощью этих протоколов маршрутизаторы собирают информацию о топологии межсетевых соединений, на основании которой осуществляется выбор маршрута продвижения пакетов [2].

Транспортный уровень *(transport layer)* обеспечивает приложениям и верхним уровням стека – прикладному, представления и сеансовому – передачу данных с той степенью надежности, которая им требуется. Модель OSI определяет пять классов транспортного сервиса: от низшего класса 0 до высшего класса 4. Эти виды сервиса отличаются качеством предоставляемых услуг: срочностью, возможностью восстановления прерванной связи, наличием средств мультиплексирования нескольких соединений между различными прикладными протоколами через общий транспортный протокол, а главное – способностью к обнаружению и исправлению ошибок передачи.

Выбор класса сервиса транспортного уровня определяется, с одной стороны, тем, в какой степени задача обеспечения надежности решается самими приложениями и протоколами более высоких, чем транспортный, уровней. С другой стороны, этот выбор зависит от того, насколько надежной является система транспортировки данных в сети, обеспечиваемая

уровнями, расположенными ниже транспортного: сетевым, канальным и физическим. Так, если качество каналов передачи очень высокое и вероятность возникновения ошибок, не обнаруженных протоколами более низких уровней, невелика, то разумно воспользоваться одним из облегченных сервисов транспортного уровня, не обремененных многочисленными проверками и квитированием. Если же транспортные средства нижних уровней очень ненадежны, то целесообразно обратиться к наиболее развитому сервису транспортного уровня, который работает, используя максимум средств для обнаружения и устранения ошибок, включая предварительное установление логического соединения, контроль доставки сообщений по контрольным суммам и т. п.

Все протоколы, начиная с транспортного уровня и выше, реализуются программными средствами конечных узлов сети – компонентами их операционных систем. В качестве примера транспортных протоколов можно привести протоколы TCP и UDP стека TCP/IP и протокол SPX стека Novell.

Протоколы нижних четырех уровней обобщенно называют *сетевым транспортом, или транспортной подсистемой,* так как они полностью решают задачу транспортировки сообщений с заданным уровнем качества в составных сетях с произвольной топологией и различными технологиями. Оставшиеся три верхних уровня решают задачи предоставления *прикладных сервисов*, используя нижележащую транспортную подсистему.

Сеансовый уровень *(session layer)* управляет взаимодействием сторон: фиксирует, какая из сторон является активной в настоящий момент, и предоставляет средства синхронизации сеанса. Эти средства позволяют в ходе длинных передач сохранять информацию о состоянии этих передач в виде контрольных точек, чтобы в случае отказа можно было вернуться назад к последней контрольной точке, а не начинать все сначала. На практике немногие приложения используют сеансовый уровень, и он редко реализуется в виде отдельных протоколов. Функции этого уровня часто объединяют с функциями прикладного уровня и реализуют в одном протоколе.

Уровень представления *(presentation layer)* обеспечивает представление передаваемой по сети информации, не меняя при этом ее содержания. За счет уровня представления информация, передаваемая прикладным уровнем одной системы, всегда понятна прикладному уровню другой системы. С помощью средств данного уровня протоколы прикладных уровней могут преодолеть синтаксические различия в представлении данных или же различия в кодах символов, например, кодов ASCII и EBCDIC. На этом уровне могут выполняться шифрование и дешифрование данных, благодаря которым секретность обмена данными обеспечивается сразу для всех прикладных служб. Примером такого протокола является протокол SSL (Secure Socket Layer – слой защищенных сокетов), который обеспечивает секретный обмен сообщениями протоколов прикладного уровня стека TCP/IP [2].

Прикладной уровень *(application layer)* – это в действительности просто набор разнообразных протоколов, с помощью которых пользователи сети получают доступ к общим ресурсам, таким как файлы, принтеры или гипертекстовые веб-страницы, а также организуют свою совместную работу, например, по протоколу электронной почты. Единица данных, которой оперирует прикладной уровень, обычно называется *сообщением*.

Существует большое разнообразие протоколов и соответствующих служб прикладного уровня, например: а) протоколы доступа к файлам NFS и FTP в стеке TCP/IP, SMB в Microsoft Windows, NCP в операционной системе Novell NetWare4; б) почтовые протоколы SMTP, IMAP, POP3; в) протокол передачи гипертекстовых сообщений HTTP.

3.1.5. Модель OSI и сети с коммутацией каналов

Модель OSI описывает процесс взаимодействия устройств в сети с *коммутацией пакетов*. А как же обстоит дело с сетями *коммутации каналов*? Существует ли для них собственная справочная модель? Можно ли сопоставить функции технологий коммутации каналов с уровнями модели OSI?

Да, для представления структуры средств межсетевого взаимодействия сетей с коммутацией каналов также используется многоуровневый подход, в соответствии с которым существуют протоколы нескольких уровней, образующих иерархию. Однако общей справочной модели, подобной модели OSI, для сетей с коммутацией каналов не существует. Например, различные типы телефонных сетей имеют собственные стеки протоколов, отличающиеся количеством уровней и распределением функций между уровнями. Первичные сети, такие как SDH или DWDM, также обладают собственной иерархией протоколов. Практически все типы современных сетей с коммутацией каналов задействуют эту технику только для передачи пользовательских данных, а для управления процессом установления соединений в сети и общего управления сетью применяют технику коммутации пакетов.

Для сетей с коммутацией пакетов сети с коммутацией каналов предоставляют сервис физического уровня, хотя сами они устроены достаточно сложно и поддерживают собственную иерархию протоколов. Например, когда несколько локальных пакетных сетей связываются между собой через цифровую телефонную сеть, маршрутизатор в каждой локальной сети должен быть оснащен интерфейсом для соединения через телефонную сеть с другой локальной сетью. После того как такое соединение установлено, в телефонной сети образуется поток битов, передаваемых с постоянной скоростью. Это соединение и предоставляет маршрутизаторам сервис физического уровня. Маршрутизаторы используют поверх этого физического канала какой-либо двухточечный протокол канального уровня.

Выводы

Эффективной моделью средств взаимодействия компьютеров в сети является многоуровневая структура, в которой модули вышележащего уровня при решении своих задач рассматривают средства нижележащего уровня как некий инструмент. Каждый уровень данной структуры поддерживает интерфейсы двух типов. Во-первых, это интерфейсы услуг с выше- и нижележащим уровнями «своей» иерархии средств. Во-вторых, это одноранговый интерфейс со средствами другой взаимодействующей стороны, расположенными на том же уровне иерархии. Этот интерфейс называют протоколом.

Иерархически организованный набор протоколов, достаточный для взаимодействия узлов в сети, называется стеком протоколов. Протоколы нижних уровней часто реализуются комбинацией программных и аппаратных средств, а протоколы верхних уровней – как правило, программными средствами. Программный модуль, реализующий некоторый протокол, называют протокольной сущностью, или собственно протоколом.

В начале 1980-х гг. ISO, ITU-T при участии некоторых других международных организаций по стандартизации разработали стандартную модель взаимодействия открытых систем (OSI). Модель OSI содержит описание обобщенного представления средств сетевого взаимодействия и используется в качестве своего рода универсального языка сетевых специалистов, именно поэтому ее называют справочной моделью. Модель OSI определяет 7 уровней взаимодействия и указывает функции каждого уровня.

Контрольные вопросы

1. Что стандартизирует модель OSI?

2. На каком уровне модели OSI работает прикладная программа?

3. Какие из приведенных утверждений не всегда справедливы? Варианты ответов:

а) протокол – это стандарт, описывающий правила взаимодействия двух систем;

б) протокол – это формализованное описание правил взаимодействия, включая последовательность обмена сообщениями и их форматы;

в) логический интерфейс – это формализованное описание правил взаимодействия, включая последовательность обмена сообщениями и их форматы.

4. Пусть на двух компьютерах установлено идентичное программное и аппаратное обеспечение, за исключением того, что драйверы сетевых адаптеров Ethernet поддерживают разные интерфейсы с протоколами сетевого уровня IP. Будут ли эти компьютеры нормально взаимодействовать, если их соединить в сеть?

5. Какое минимальное количество уровней протоколов (в терминах модели OSI) должны поддерживать маршрутизаторы сетей с коммутацией пакетов?

3.2. Стандартизация сетей

Универсальный тезис о пользе стандартизации, справедливый для всех отраслей, в компьютерных сетях приобретает особое значение. Суть сети –

объединение разного оборудования, а значит, проблема совместимости является здесь одной из наиболее острых. Без согласования всеми производителями общепринятых стандартов для оборудования и протоколов прогресс в деле «строительства» сетей был бы невозможен. Поэтому все развитие компьютерной отрасли в конечном счете отражено в стандартах – любая новая технология только тогда приобретает «законный» статус, когда ее содержание закрепляется в соответствующем стандарте. В компьютерных сетях идеологической основой стандартизации является модель OSI.

3.2.1. Понятие открытой системы

Открытой может быть названа любая система (компьютер, вычислительная сеть, ОС, программный пакет, другие аппаратные и программные продукты), которая построена в соответствии с открытыми спецификациями. Под термином «*спецификация*» в вычислительной технике понимают формализованное описание аппаратных или программных компонентов, способов их функционирования, взаимодействия с другими компонентами, условий эксплуатации; не всякая спецификация является *стандартом*.

Под *открытыми спецификациями* понимаются опубликованные, общедоступные спецификации, соответствующие стандартам и принятые в результате достижения согласия после всестороннего обсуждения всеми заинтересованными сторонами. Использование при разработке систем открытых спецификаций позволяет третьим сторонам создавать для этих систем различные аппаратные или программные средства расширения и модификации, а также программно-аппаратные комплексы разных производителей.

Открытый характер стандартов и спецификаций важен не только для телекоммуникационных протоколов, но и для разнообразных устройств и программ, выпускаемых для построения сети. Большинство стандартов, принимаемых сегодня, носят открытый характер. Все осознали, что возможность взаимодействия с продуктами конкурентов не снижает, а повышает ценность изделия, так как позволяет применять его в большем количестве работающих сетей, собранных из продуктов разных производителей.

Для реальных систем полная открытость является недостижимым идеалом. Модель OSI касается только одного аспекта открытости, а именно открытости средств взаимодействия устройств, связанных в компьютерную сеть. Здесь под открытой системой понимается сетевое устройство, готовое взаимодействовать с другими сетевыми устройствами по стандартным правилам. Если две сети построены с соблюдением открытости, это дает следующие преимущества: а) возможность построения сети из аппаратных и программных средств различных производителей, придерживающихся одного стандарта; б) безболезненная замена отдельных компонентов сети другими, более совершенными; в) легкость сопряжения одной сети с другой [2].

3.2.2. Источники стандартов

Закон РФ № 65-ФЗ «О техническом регулировании» определяет понятие «стандарт» следующим образом. *Стандарт* – это «документ, в котором в целях добровольного многократного использования устанавливаются характеристики продукции, правила осуществления и характеристики процессов проектирования (включая изыскания), производства, строительства, монтажа, наладки, эксплуатации, хранения, перевозки, реализации и утилизации, выполнения работ или оказания услуг.

По умолчанию *соблюдение стандарта не является обязательным*. Однако существует множество причин, по которым большинство компаний добровольно выбирают следование стандартам для обеспечения совместимости информационных технологий, продуктов и терминологии. Следование стандартам позволяет также создавать более качественные и конкурентоспособные технологии, системы и услуги, так как стандарты – это концентрированное выражение передовой технической мысли, они аккумулируют актуальные теоретические знания и так называемые «лучшие практики».

Как правило, стандарты разрабатываются рабочими группами, в состав которых на добровольной основе могут включаться представители органов исполнительной власти, научных, коммерческих и некоммерческих организаций, общественных объединений. Часто разработчиками стандартов являются компании и организации, успешно работающие в той области, для которой они предлагают стандарты. В зависимости от статуса организаций различают следующие виды стандартов.

Стандарты отдельных фирм, например, стек протоколов SNA компании IBM или графический интерфейс OPEN LOOK для Unix-систем от Sun.

Стандарты специальных комитетов и объединений создаются несколькими компаниями, например, стандарты ATM, разрабатываемые объединением ATM Forum, которое насчитывает около 100 участников, или стандарты союза Fast Ethernet Alliance технологии 100 Мбит Ethernet.

Национальные стандарты, например, стандарт FDDI, представляющий один из многочисленных стандартов института ANSI.

Международные стандарты, например, модель и стек коммуникационных протоколов Международной организации по стандартизации (ISO), стандарты Международного союза электросвязи (ITU), в том числе стандарты на сети с коммутацией пакетов X.25, сети Frame Relay, ISDN.

В нашей стране главную организационную роль в стандартизации играет Федеральное агентство по техническому регулированию и метрологии (Росстандарт). Росстандарт создает и координирует рабочие группы по разработке стандартов, организует общественное обсуждение и экспертизу новых стандартов, утверждает и публикует документы по стандартам, ведет учет и распространение национальных стандартов.

Некоторые стандарты, непрерывно развиваясь, могут переходить из одной категории в другую. В частности, фирменные стандарты на продукцию, получившую широкое распространение, обычно становятся международными стандартами де-факто, так как вынуждают производителей из разных стран следовать фирменным стандартам, чтобы обеспечить совместимость своих изделий с этими популярными продуктами. Например, из-за феноменального успеха персонального компьютера компании IBM фирменный стандарт на архитектуру IBM PC стал международным де-факто.

Более того, ввиду широкого распространения некоторые фирменные стандарты становятся основой для национальных и международных стандартов де-юре. Например, Ethernet, первоначально разработанный компаниями Digital Equipment, Intel и Xerox, через некоторое время и в несколько измененном виде был принят как национальный стандарт IEEE 802.3, а затем ISO утвердила его в качестве международного стандарта ISO 8802.3.

Ярким примером открытой системы является *Интернет*. Эта международная сеть развивалась в полном соответствии с требованиями, предъявляемыми к открытым системам. В разработке ее стандартов принимали участие тысячи специалистов – пользователей этой сети из различных университетов, научных организаций и фирм – производителей вычислительной аппаратуры и программного обеспечения, работающих в разных странах. Само название стандартов, определяющих работу Интернета, – *темы для обсуждения* (Requests For Comments, RFC) – показывает гласный и открытый характер принимаемых стандартов. В результате Интернет сумел объединить в себе разнообразное оборудование и программное обеспечение огромного числа сетей, разбросанных по всему миру. Ввиду постоянно растущей популярности Интернета RFC-документы становятся международными стандартами де-факто, многие из которых затем приобретают статус официальных международных стандартов в результате их утверждения какой-либо организацией по стандартизации, как правило, ISO и ITU-T.

Существует несколько организационных подразделений, отвечающих за развитие и, в частности, за стандартизацию архитектуры и протоколов Интернета. Основным из них является научно-административное *сообщество Интернета* (Internet Society, ISOC), которое занимается социальными, политическими и техническими проблемами эволюции Интернета. Под управлением ISOC работает *совет по архитектуре Интернета* (Internet Architecture Board, IAB). В IAB входят две основные группы: Internet Research Task Force (IRTF) и Internet Engineering Task Force (IETF). IRTF координирует долгосрочные исследовательские проекты по протоколам TCP/IP. IETF – это инженерная группа, которая занимается решением текущих технических проблем Интернета. Именно IETF определяет спецификации, которые затем становятся стандартами Интернета. Процесс раз-

работки и принятия стандарта состоит из обязательных этапов, включающих экспериментальную проверку.

В соответствии с принципом открытости Интернета все RFC-документы в отличие от стандартов ISO, находятся в свободном доступе. Список RFC-документов можно найти, в частности, на сайте www.rfc-editor.org [2].

3.2.3. Стеки коммуникационных протоколов

Важнейшим направлением стандартизации в области сетей является стандартизация коммуникационных протоколов. Наиболее известными стеками протоколов являются: OSI, TCP/IP, IPX/SPX, NetBIOS/SMB, DECnet, SNA (не все из них применяются сегодня на практике).

Стек OSI. Важно различать *модель* OSI и *стек протоколов* OSI. Модель OSI является концептуальной схемой взаимодействия открытых систем, а стек OSI представляет собой набор спецификаций конкретных протоколов.

В отличие от других стеков протоколов стек OSI полностью соответствует модели OSI, включая спецификации протоколов для всех семи уровней взаимодействия, определенных в этой модели (рис. 3.10).

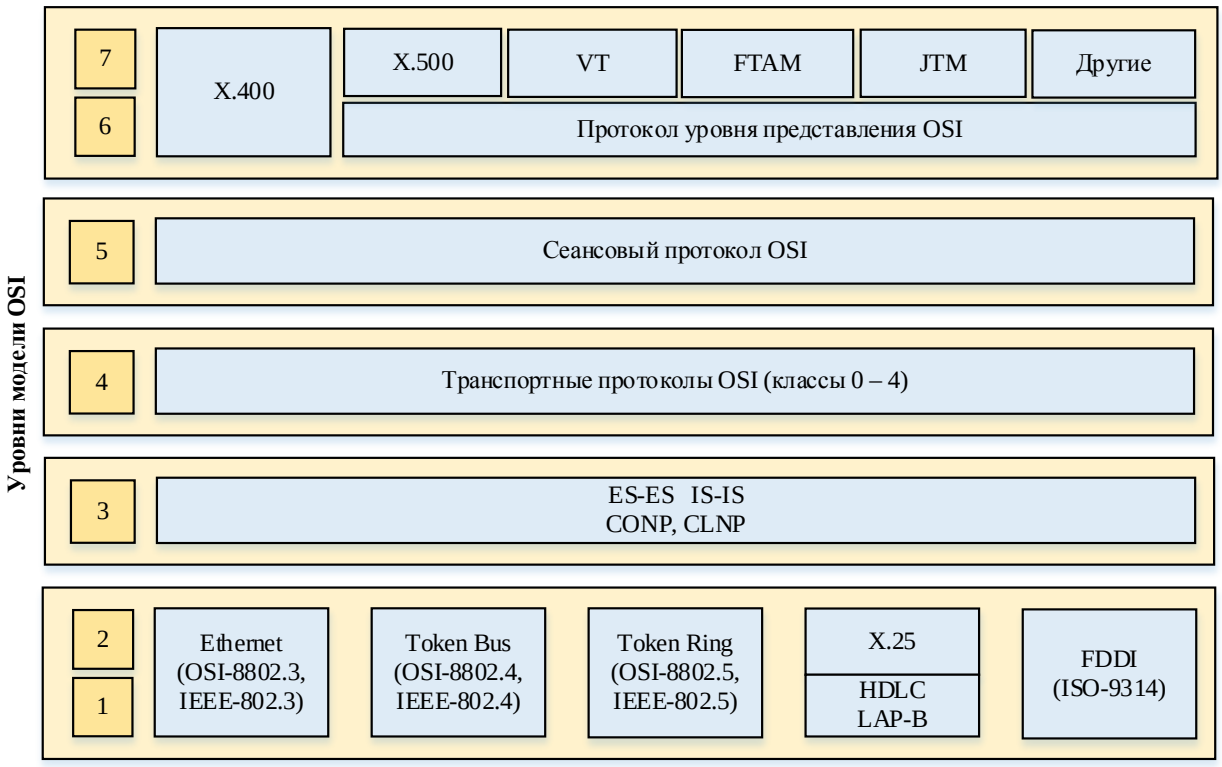


Рис. 3.10. Стек протоколов OSI

Разработчики стека OSI использовали модель OSI как прямое руководство к действию. Протоколы стека OSI отличает сложность спецификаций

в результате стремления разработчиков учесть в своих протоколах все многообразие уже существующих и появляющихся технологий.

На *физическом* и *канальном уровнях* стек OSI поддерживает протоколы Ethernet, Token Ring, FDDI, а также протоколы LLC, X.25 и ISDN, т. е., как и большинство других стеков, использует все разработанные вне стека популярные протоколы нижних уровней.

Сетевой уровень включает сравнительно редко используемые протоколы Connection-oriented Network Protocol (CONP) и Connectionless Network Protocol (CLNP). Как следует из названий, первый из них ориентирован на соединение (connection-oriented), второй – нет (connectionless).

Более популярны протоколы маршрутизации стека OSI: между конечной и промежуточной (End System – Intermediate System, ES-IS) и между промежуточными системами (Intermediate System – Intermediate System).

Транспортный уровень стека OSI в соответствии с функциями, определенными для него в модели OSI, скрывает различия между сетевыми сервисами с установлением соединения и без установления соединения, так что пользователи получают требуемое качество обслуживания независимо от нижележащего сетевого уровня. Чтобы обеспечить это, транспортный уровень требует, чтобы пользователь задал нужное качество обслуживания.

Службы *прикладного уровня* обеспечивают передачу файлов, эмуляцию терминала, сервис каталогов и почту. Из них наиболее популярными является сервис каталогов (стандарт X.500), электронная почта (X.400), протокол виртуального терминала (VTP), протокол передачи, доступа и управления файлами (FTAM), протокол пересылки и управления работами (JTM).

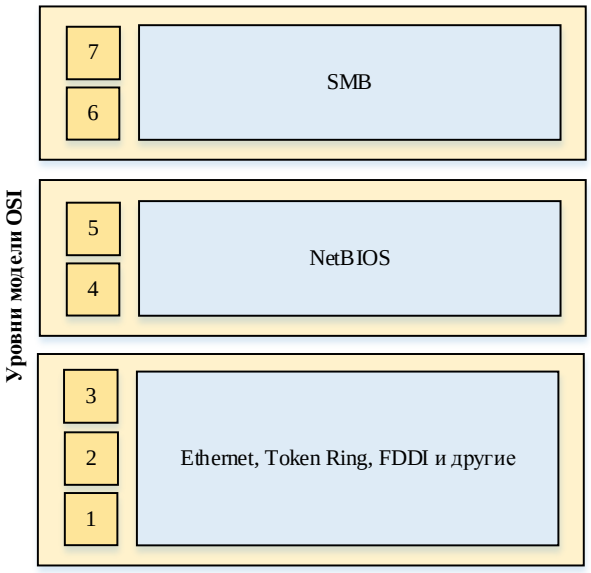


Рис. 3.11. Стек протоколов NetBIOS/SMB

Стек NetBIOS/SMB является совместной разработкой компаний IBM и Microsoft (рис. 3.11). На физическом и канальном уровнях этого стека задействованы получившие распространение протоколы Ethernet, Token Ring, FDDI, а на верхних уровнях – специфические протоколы NetBEUI и SMB.

Протокол Network Basic Input/Output System (NetBIOS) появился в 1984 г. как сетевое расширение стандартных функций базовой системы ввода-вывода (BIOS) IBM PC для сетевой программы PC Network фирмы IBM. В дальнейшем этот протокол был заменен так назы-

ваемым протоколом расширенного пользовательского интерфейса NetBEUI (NetBIOS Extended User Interface). NetBEUI разрабатывался как эффективный протокол, предназначенный для сетей, насчитывающих не более 200 рабочих станций. Этот протокол поддерживает много полезных сетевых функций, которые можно отнести к транспортному и сеансовому уровням модели OSI, однако с его помощью *невозможна маршрутизация* пакетов. Это ограничивает его локальными сетями, и делает невозможным его использование в составных сетях.

Протокол Server Message Block (SMB) поддерживает функции сеансового уровня, уровня представления и прикладного. На основе SMB реализуется файловая служба печати и передачи сообщений между приложениями.

Стек TCP/IP был разработан по инициативе Министерства обороны США для связи экспериментальной сети ARPAnet с другими сетями как набор общих протоколов для разнородной вычислительной среды. Большой вклад в развитие стека TCP/IP, который получил свое название по популярным протоколам IP и TCP, внес университет Беркли, реализовав протоколы стека в своей версии ОС UNIX. Популярность этой ОС привела к широкому распространению протоколов TCP, IP и других протоколов стека [2].

На рис. 3.12 показано, **в какой степени популярные стеки протоколов соответствуют рекомендациям модели OSI**. Как видно, часто это соответствие весьма условно. В большинстве случаев разработчики стеков отдавали предпочтение скорости работы сети в ущерб модульности – ни один стек, кроме стека OSI, не разбит на семь уровней. Чаще всего в стеке явно выделяются 3-4 уровня: уровень сетевых адаптеров, в котором реализуются протоколы физического и канального уровней, сетевой уровень, транспортный уровень и уровень служб, вбирающий в себя функции сеансового уровня, уровня представления и прикладного уровня.

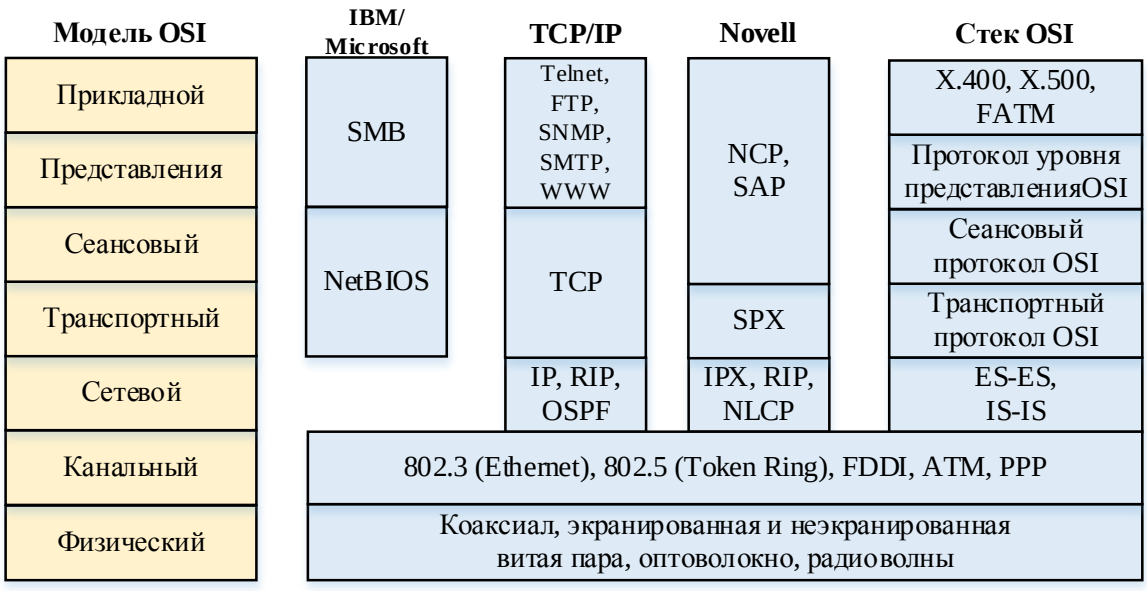


Рис. 3.12. Соответствие популярных стеков протоколов модели OSI

Структура стеков протоколов часто не соответствует рекомендуемому моделью OSI разбиению на уровни и по другим причинам. Давайте вспомним, чем характеризуется идеальная многоуровневая декомпозиция. С одной стороны, необходимо соблюсти принцип иерархии: каждый вышележащий уровень обращается с запросами только к нижележащему, а нижележащий предоставляет свои сервисы только непосредственно соседствующему с ним вышележащему. В стеках протоколов это приводит к тому, что PDU вышележащего уровня всегда инкапсулируется в PDU нижележащего.

С другой стороны, идеальная многоуровневая декомпозиция предполагает, что все модули, отнесенные к одному уровню, ответственны за решение общей для всех них задачи. Однако эти требования часто вступают в противоречие. Например, основной функцией протоколов сетевого уровня стека TCP/IP (так же, как и сетевого стека OSI) является передача пакетов через составную сеть. Для решения этой задачи в стеке TCP/IP предусмотрено несколько протоколов: протокол продвижения IP-пакетов, протоколы маршрутизации RIP, OSPF и др. Если считать признаком принадлежности к одному и тому же уровню общность решаемых задач, то очевидно, протокол IP и протоколы маршрутизации должны быть отнесены к одному уровню. Вместе с тем если принять во внимание, что сообщения протокола RIP инкапсулируются в UDP-дейтаграммы, а сообщения протокола OSPF – в IP-пакеты, то, следуя формально принципу иерархической организации стека, OSPF следовало бы отнести к транспортному, а RIP – к прикладному уровню. На практике же протоколы маршрутизации обычно включают в сетевой уровень.

3.2.4. Информационные и транспортные услуги

Услуги компьютерной сети можно разделить на две категории: *транспортные услуги и информационные услуги*.

Транспортные услуги состоят в передаче информации между пользователями сети в неизменном виде. При этом сеть принимает информацию от пользователя на одном из своих интерфейсов, передает ее через промежуточные коммутаторы и выдает другому пользователю через другой интерфейс. При оказании транспортных услуг сеть не вносит никаких изменений в передаваемую информацию, передавая ее получателю в том виде, в котором она поступила в сеть от отправителя. Примером транспортной услуги глобальных сетей является объединение локальных сетей клиентов.

Информационные услуги состоят в предоставлении пользователю некоторой новой информации. Информационная услуга всегда связана с операциями по обработке информации: хранению ее в некотором упорядоченном виде (файловая система, база данных, веб-сайт), поиску нужной информации. Информационные услуги существовали и до появления ком-

пьютерных сетей, например, справочные услуги телефонной сети. С появлением компьютеров информационные услуги пережили революцию.

В телекоммуникационных сетях «докомпьютерной» эры всегда преобладали транспортные услуги. Основной услугой телефонной сети была передача голосового трафика между абонентами, в то время как справочные услуги были дополнительными. В компьютерных сетях одинаково важны обе категории услуг. Эта особенность компьютерных сетей сегодня отражается в названии нового поколения телекоммуникационных сетей, которые появляются в результате конвергенции сетей различных типов. Такие сети все чаще стали называть *инфокоммуникационными*. Это название хорошо отражает новые тенденции, включая обе составляющие услуг.

3.2.5. Распределение протоколов по элементам сети

На рис. 3.13 показаны основные элементы компьютерной сети: конечные узлы – компьютеры, и промежуточные узлы – коммутаторы и маршрутизаторы.

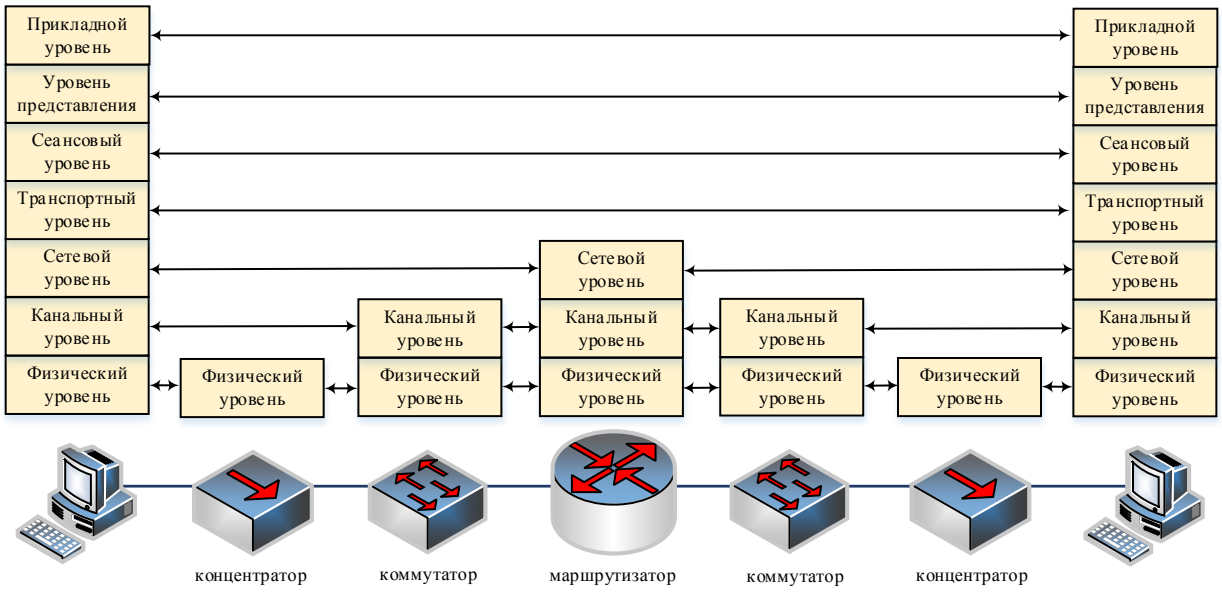


Рис. 3.13. Соответствие функций
различных устройств сети уровням модели OSI

Из рис. 3.13 видно, что полный стек протоколов реализован только на конечных узлах, а промежуточные узлы поддерживают протоколы всего трех нижних уровней. Это объясняется тем, что коммуникационным устройствам для продвижения пакетов достаточно функциональности трех нижних уровней. Более того, коммуникационное устройство может поддерживать только протоколы двух нижних уровней или даже одного физического уровня – это зависит от типа устройства. К устройствам, работающим на физическом уровне, относятся, например, сетевые повторители,

называемые также *концентраторами*, или *хабами*. Они повторяют электрические сигналы, поступающие на одни их интерфейсы, на других своих интерфейсах, улучшая их характеристики – мощность и форму сигналов.

Коммутаторы локальных сетей поддерживают протоколы двух нижних уровней, физического и канального, что дает им возможность работать в пределах стандартных топологий.

Маршрутизаторы должны поддерживать протоколы всех трех уровней, так как сетевой уровень нужен им для объединения сетей различных технологий, а протоколы нижних уровней – для взаимодействия с сетями, образующими составную сеть, например, Ethernet или Frame Relay.

Коммутаторы глобальных сетей (например, MPLS), работающие на основе технологии виртуальных каналов, могут поддерживать как два уровня протоколов, так и три. Протокол сетевого уровня нужен им в том случае, если они поддерживают процедуры автоматического установления виртуальных каналов. Так как топология глобальных сетей произвольная, без сетевого протокола обойтись нельзя.

Компьютеры, на которых работают сетевые приложения, поддерживают протоколы всех уровней. Протоколы прикладного уровня, пользуясь сервисами протоколов уровня представления и сеансового уровня, предоставляют приложениям набор сетевых услуг в виде сетевого прикладного программного интерфейса (API). Протокол транспортного уровня также работает на всех конечных узлах. При передаче данных через сеть два модуля транспортного протокола на узле-отправителе и узле-получателе взаимодействуют друг с другом для поддержания сервиса нужного качества.

В компьютерах коммуникационные протоколы всех уровней (кроме физического и части функций канального уровня) реализуются программно операционной системой или системными приложениями. Конечные узлы сети (компьютеры и компьютеризированные устройства, например, мобильные телефоны) всегда предоставляют как информационные, так и транспортные услуги, а промежуточные узлы сети – только транспортные.

Вспомогательные протоколы транспортной системы. На рис. 3.13 показан упрощенный вариант распределения протоколов между элементами сети. В реальных сетях некоторые из коммуникационных устройств поддерживают не только протоколы трех нижних уровней, но и протоколы верхних уровней. Так, маршрутизаторы реализуют протоколы маршрутизации, позволяющие автоматически строить таблицы маршрутизации, а концентраторы и коммутаторы часто поддерживают протоколы SNMP и telnet, которые не нужны для выполнения основных функций этих устройств, но позволяют конфигурировать их и управлять ими удаленно.

Большинство вспомогательных протоколов формально относится к прикладному уровню модели OSI, так как в своей работе они обращаются

к протоколам нижних уровней, таким как TCP, DP или SSL. Однако при этом вспомогательные протоколы не переносят пользовательские данные, т. е. они не выполняют непосредственно функций протокола прикладного уровня, описанного в модели OSI.

При рассмотрении вспомогательных протоколов мы сталкиваемся с ситуацией, когда деления протоколов на уровни иерархии (т. е. деления «по вертикали»), присущего модели OSI, оказывается недостаточно. Полезным оказывается деление протоколов на группы «по горизонтали». При горизонтальном делении все протоколы разделяют на три слоя (рис. 3.14).

Пользовательский слой (user plane) включает группу основных протоколов, т. е. протоколов, которые переносят пользовательский трафик.

Слой управления (control plane) составляют вспомогательные протоколы, необходимые для работы основных протоколов сети, например, протоколы маршрутизации, протоколы отображения имен на IP-адреса.

Слой менеджмента (management plane) объединяет вспомогательные протоколы, поддерживающие операции менеджмента (управления сетью администратором), такие как протокол SNMP для сбора информации об ошибках, протоколы удаленного конфигурирования устройств.

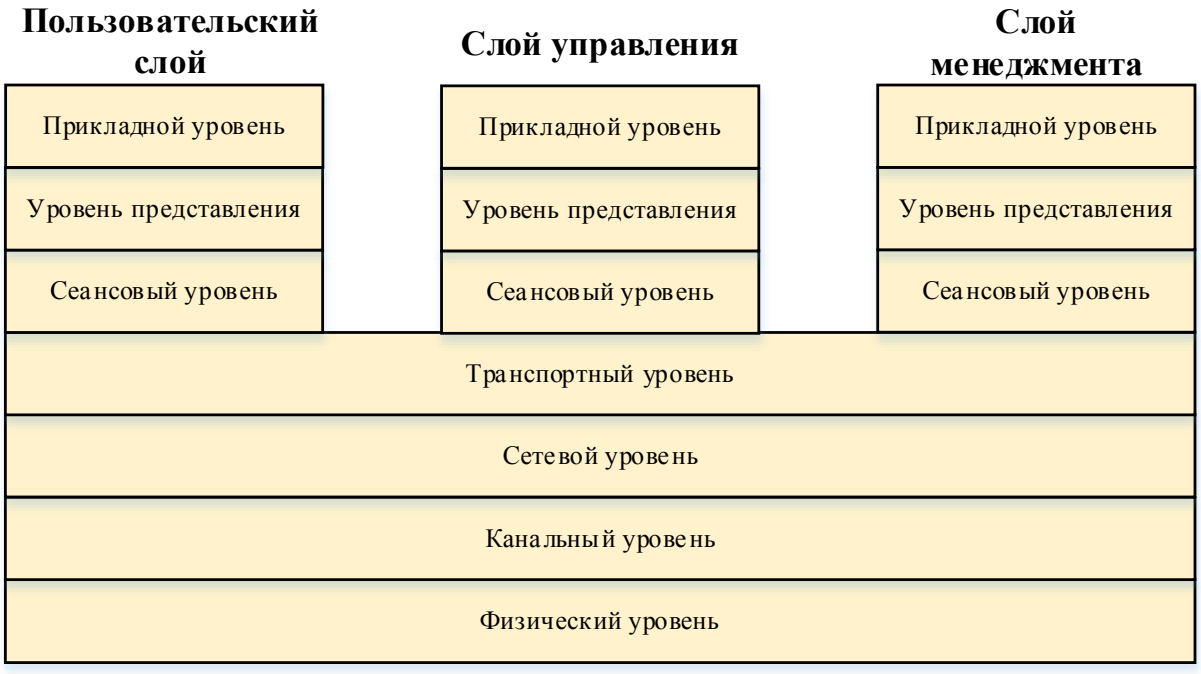


Рис. 3.14. Три группы протоколов

«Горизонтальное» деление протоколов снимает сложности, возникающие при соотнесении некоторых протоколов уровням модели OSI. Модель OSI хорошо подходит для стандартизации протоколов, которые переносят пользовательский трафик, т. е. протоколов пользовательского слоя. В то же время она в гораздо меньшей степени годится для определения места вспомогательных протоколов в общей модели функционирования сети.

3.2.6. Классификация компьютерных сетей

Классификация – процесс группирования (отнесения к тому или иному типу) объектов изучения в соответствии с их общими признаками.

Компьютерные сети сами собой являются элементом классификации телекоммуникационных сетей, а именно, телекоммуникационные сети *по виду передаваемого контента* делятся на радиосети, телефонные сети, телевизионные сети, компьютерные сети.

В зависимости от *территории покрытия* компьютерные сети можно разделить на три группы: локальные сети (Local Area Network, LAN), глобальные сети (Wide Area Network, WAN), городские сети, или сети мегаполиса (Metropolitan Area Network, MAN).

Используя в данном контексте термины «локальные сети» или «глобальные сети», имеют в виду прежде всего различия *технологий* локальных и глобальных сетей, а не тот факт, что эти сети имеют разный территориальный масштаб.

В локальных сетях качество линий связи между узлами обычно выше, чем в глобальных сетях. Это обусловлено рядом причин: а) существенно меньшей длиной линий связи (метры вместо сотен километров), а значит, и меньшими искажениями сигналов, вносимых неидеальной передающей средой; б) меньшим уровнем внешних помех, так как в локальной сети оборудование и кабели обычно размещаются в специальных защищенных экранированных помещениях, а линии связи глобальной сети могут проходить в сильно электромагнитно «зашумленной» среде, например, в туннелях подземных коммуникаций, рядом с силовыми кабелями, вдоль линий электропередач и т. п.; в) экономическими соображениями.

Сети MAN предназначены для обслуживания территории крупного города – мегаполиса – и сочетают в себе признаки как локальных, так и глобальных сетей. От первых они унаследовали большую плотность подключения конечных абонентов и высокоскоростные линии связи, а от вторых – большую протяженность линий связи.

В соответствии с технологическими признаками, обусловленными *средой передачи*, компьютерные сети подразделяются на два класса: а) *проводные сети* – сети, каналы связи которых построены с использованием медных или оптических кабелей; б) *беспроводные сети* – сети, в которых для связи используются беспроводные каналы связи, например, радио и СВЧ.

Любая беспроводная среда гораздо больше подвержена влиянию внешних помех, чем проводная, поэтому технологии беспроводных сетей должны обеспечивать работоспособность сети, несмотря на ухудшение внешних условий. Кроме того, существует ряд других специфических особенностей беспроводных сетей, которые служат основанием для выделения их в особый класс, например, *естественное разделение радиосреды всеми*

узлами сети, находящимися в радиусе действия всенаправленного передатчика; распределение диапазона радиочастот между сетями различного назначения, например, между телефонными и компьютерными.

В зависимости от *способа коммутации* сети подразделяются на два класса: а) *сети с коммутацией пакетов*; б) *сети с коммутацией каналов*.

Сейчас в компьютерных сетях преимущественно используется техника коммутации пакетов. Техника коммутации пакетов, в свою очередь, допускает несколько вариаций, отличающихся *способом продвижения пакетов*, в соответствии с чем сети делят: а) на *дейтаграммные сети*, например, Ethernet; б) *сети, основанные на логических соединениях,* например, IP-сети, использующие на транспортном уровне протокол TCP; в) *сети, основанные на виртуальных каналах,* например, MPLS-сети.

Сети могут быть классифицированы на основе *топологии*. Топологический тип сети весьма отчетливо характеризует сеть, он понятен как профессионалам, так и пользователям. Мы рассматривали базовые топологии сетей: *полносвязная топология, дерево, звезда, кольцо, смешанная топология*.

В зависимости от *типа пользователей услуг сети*, сети делятся на сети операторов связи, корпоративные и персональные сети.

Сети операторов связи предоставляют публичные услуги, т. е. клиентом сети может стать любой индивидуальный пользователь или любая организация, которая заключила соответствующий коммерческий договор на предоставление телекоммуникационной услуги. Традиционными услугами операторов связи являются услуги телефонии, а также предоставления каналов связи в аренду тем организациям, которые собираются строить на их основе собственные сети. С распространением компьютерных сетей операторы существенно расширили спектр своих услуг, добавив к ним услуги Интернета, виртуальных частных сетей, веб-хостинг, электронную почту и IP-телефонию, а также широковещательную рассылку аудио- и видеосигналов.

Корпоративные сети предоставляют услуги только сотрудникам предприятия, которое владеет этой сетью. Хотя формально корпоративная сеть может иметь любой размер, обычно под корпоративной понимают сеть крупного предприятия.

Персональные сети находятся в личном использовании. Для них характерно небольшое количество узлов, простая структура, а также небольшой (в пределах 30 метров) радиус действия. Узлами персональной сети наряду с настольными ПК могут быть телефоны, смартфоны, планшеты, ноутбуки.

В зависимости от *функциональной роли, которую играют некоторые части сети*, ее относят к сети доступа, магистральной сети или сети агрегирования трафика (рис. 3.15).

Сети доступа – это сети, предоставляющие доступ индивидуальным и корпоративным абонентам от их помещений (квартир, офисов) до первого помещения (пункта присутствия) оператора сети связи или оператора корпоративной сети. Другими словами, это сети, ответственные за расширение глобальной сети до помещений ее клиентов.

Магистральные сети – это сети, представляющие собой наиболее скоростную часть (ядро) глобальной сети, которая объединяет многочисленные сети доступа в единую сеть.

Сети агрегирования трафика – это сети, агрегирующие данные от многочисленных сетей доступа для компактной передачи их по небольшому числу каналов связи в магистраль. Сети агрегирования обычно используются только в крупных глобальных сетях, где они занимают промежуточную позицию, помогая магистральной сети обрабатывать трафик, поступающий от большого числа сетей доступа. В сетях среднего и небольшого размера сети агрегирования обычно отсутствуют.

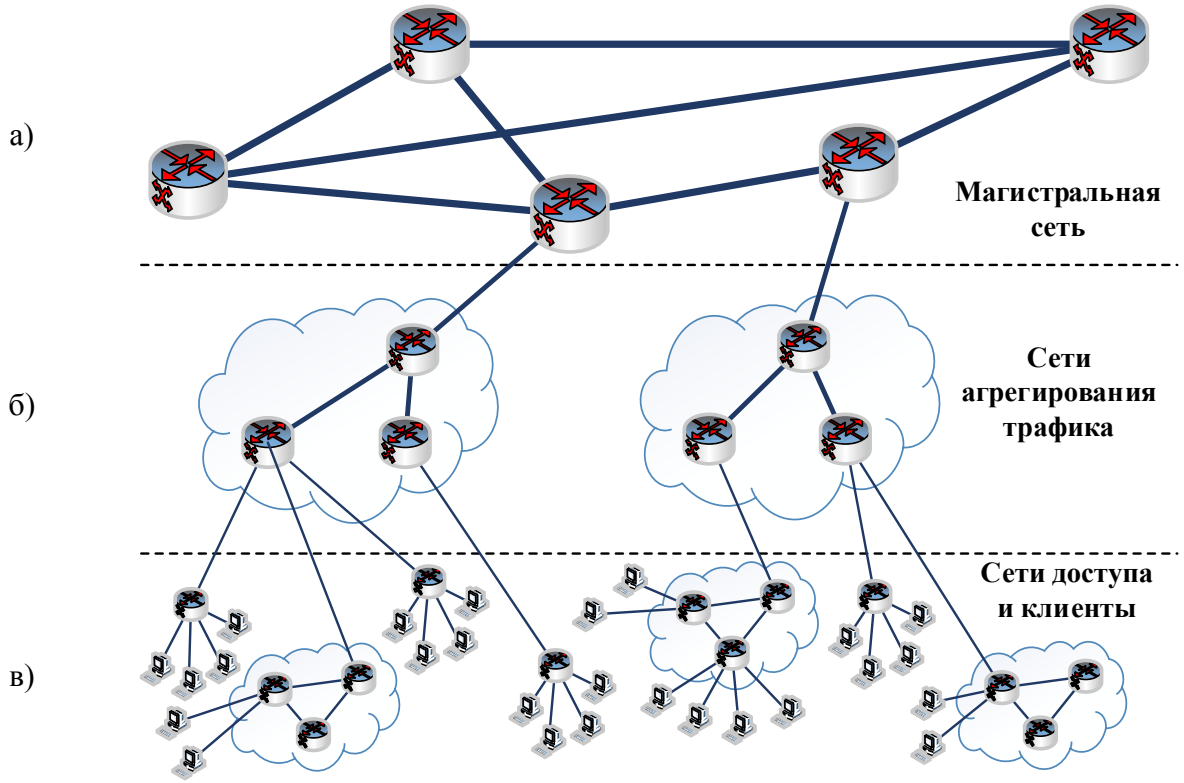


Рис. 3.15. Сети доступа (а), сети агрегирования трафика (б) и магистральная сеть (в)

Различают также *первичные и наложенные телекоммуникационные сети*. *Первичные сети* занимают особое положение в мире телекоммуникационных сетей; их можно рассматривать как *вспомогательные* сети, позволяющие гибко создавать постоянные физические двухточечные каналы для других компьютерных и телефонных сетей.

Наложенные сети в этой классификации – это все остальные сети, которые предоставляют услуги конечным пользователям и строятся на основе каналов первичных сетей – «накладываются» поверх этих сетей. И компьютерные, и телефонные, и телевизионные сети являются наложенными.

Оптоволоконные кабели обладают наилучшими на сегодняшний день характеристиками передачи данных, они используются как в локальных, так и в глобальных проводных сетях. Термин *оптические сети* часто трактуется в узком смысле: как синоним первичных сетей. Это объясняется тем, что оптические кабели являются для первичных сетей основным вариантом работы.

Интернет представляет собой уникальную сеть, объединяющую практически все компьютерные сети во всемирном масштабе. Если применить к Интернету признаки, описанные в классификации, можно сказать, что это: а) публичная сеть; б) сеть операторов связи, предоставляющая публичные услуги, как информационные, так и транспортные; в) сеть с коммутацией пакетов; г) сеть, состоящая из магистральных сетей, сетей агрегирования трафика и сетей доступа.

Выводы

В зависимости от области действия различают стандарты отдельных компаний, стандарты специальных комитетов и объединений, национальные стандарты, международные стандарты. Важнейшим направлением стандартизации в области сетей является стандартизация коммуникационных протоколов. Примерами стандартизированных стеков протоколов являются TCP/IP, IPX/SPX, NetBIOS/SMB, OSI, DECnet, SNA. Лидирующее положение занимает стек TCP/IP, он используется для связи десятков миллионов компьютеров всемирной информационной сети Интернет. Стек TCP/IP имеет 4 уровня: прикладной, транспортный, межсетевого взаимодействия и сетевых интерфейсов. Соответствие уровней стека TCP/IP уровням модели OSI достаточно условно.

Компьютерные сети предоставляют услуги двух типов: информационные и транспортные. Информационные услуги предоставляются конечными узлами сети – серверами, а транспортные – промежуточными узлами, которыми являются коммутаторы и маршрутизаторы сети.

Классификация компьютерных сетей может быть выполнена на основе различных критериев. Это могут быть технологические характеристики сетей, такие как топология, метод коммутации, метод продвижения пакетов, тип используемой среды передачи. Сети классифицируют и на основе других признаков, не являющихся технологическими, таких, например, как тип потребителей предоставляемых услуг (сети операторов и корпоративные сети) или функциональная роль (магистраль, сеть доступа).

Контрольные вопросы

1. Дайте определение открытой системы.

2. Какая организация разработала стандарты сетей Ethernet?

3. Какое из административных подразделений Интернета непосредственно занимается стандартизацией?

4. Какие из перечисленных терминов являются синонимами. Варианты ответов: а) стандарт; б) спецификация; в) RFC; г) никакие.

5. К какому типу стандартов могут относиться современные документы RFC? Варианты ответов:

а) к стандартам отдельных фирм;

б) к государственным стандартам;

в) к национальным стандартам;

г) к межгосударственным стандартам.

6. Определите основные особенности стека TCP/IP.

7. Сравните функции самых нижних уровней модели TCP/IP и OSI.

8. Дайте определение транспортных и информационных услуг.

9. Какие протоколы относятся к слою управления (control plane)? Какие протоколы относятся к слою менеджмента (management plane)?

10. Должны ли маршрутизатором поддерживаться протоколы транспортного уровня?

11. Пусть на двух компьютерах установлено идентичное программное и аппаратное обеспечение за исключением того, что драйверы сетевых адаптеров Ethernet поддерживают отличающиеся интерфейсы с протоколом сетевого уровня IP. Будут ли эти компьютеры нормально взаимодействовать, если их соединить в сеть?

12. Как организовать взаимодействие двух компьютеров, если у них отличаются протоколы физического и канального уровня?

13. Приведите критерии классификации компьютерных сетей.

14. Какие специфические особенности беспроводных сетей служат основанием для выделения их в особый класс?

15. К какому типу сети относится Интернет?

4. СЕТЕВЫЕ ХАРАКТЕРИСТИКИ И КАЧЕСТВО ОБСЛУЖИВАНИЯ

4.1. Сетевые характеристики

Компьютерная сеть представляет собой сложную и дорогую систему, обслуживающую большое количество пользователей, поэтому очень важно, чтобы сеть работала качественно. Для трактовки *качества обслуживания* существует ряд общепринятых характеристик. Основными характеристиками качества транспортных услуг сети, которые поддаются формализации, являются *производительность, надежность и безопасность*.

4.1.1. Типы характеристик

Субъективные оценки качества. Пользователи могут вкладывать в понятие качественных сетевых услуг следующие мнения: а) сеть работает быстро, без задержек; б) трафик передается надежно, данные не теряются; в) услуги предоставляются бесперебойно по схеме 24×7; г) служба поддержки работает хорошо; д) услуги предоставляются по гибкой схеме; е) поставщик не только передает трафик, но и защищает сеть от вирусов; ж) всегда можно проконтролировать, насколько быстро и без потерь сеть передает трафик; з) поставщик предоставляет широкий спектр услуг, помимо стандартного доступа в Интернет предлагает услуги IP-телефонии. Эти *субъективные* оценки отражают пожелания пользователей к качеству сетевых сервисов.

Количественные характеристики и требования. Пользователи сети – это хотя и важная, но только одна сторона бизнеса сетей передачи данных. Существует и другая сторона – *поставщик услуг*. Чтобы пользователи и поставщики услуг могли «найти общий язык», существуют *формализованные количественные характеристики* качества сетевых услуг.

Получая сетевые услуги, пользователь формулирует определенные *требования к характеристикам сети*. Все множество характеристик качества транспортных услуг сети можно отнести к одной из следующих групп: а) *производительность*; б) *надежность*; в) *безопасность*; г) характеристики, имеющие значение только для поставщика услуг. Первые три группы соответствуют трем наиболее важным для пользователя характеристикам транспортных услуг – возможности без потерь в обслуживании (*надежность*) передавать с заданной скоростью (*производительность*) защищенную от несанкционированного доступа и подмены информацию (*безопасность*). В то же время существует ряд характеристик, важных для поставщика сети, но не представляющих интереса для пользователей. Например, поставщика интересует производительность коммутатора (какое количество потоков пользователей он может обработать с помощью данного коммутатора); для

пользователя же производительность коммутатора значения не имеет, ему важен конечный результат обслуживания его потока [2].

Временная шкала. Еще один способ классификации характеристик – временная шкала, на которой эти характеристики определяются.

Долговременные характеристики (или *характеристики проектных решений*) определяются на промежутках времени от нескольких месяцев до нескольких лет. Примерами таких характеристик являются количество и схема соединения коммутаторов в сети, пропускная способность линий связи, конкретные модели и характеристики используемого оборудования. Одно проектное решение может оказаться удачным и сбалансированным, так что потоки трафика не будут испытывать перегрузок; другое может создавать узкие места для потоков, в результате задержки и потери пакетов превысят допустимые пределы. Глубокая модернизация сети связана с большими финансовыми и временными затратами и происходит редко, а значит, выбранные однажды параметры продолжают влиять на качество функционирования сети в течение продолжительного времени.

Среднесрочные характеристики определяются на интервалах времени от нескольких секунд до нескольких дней. Как правило, за это время происходит обслуживание большого количества пакетов. Например, к среднесрочным характеристикам может быть отнесено усредненное значение задержки пакетов по выборке, взятой в течение суток.

Краткосрочные характеристики относятся к темпу обработки отдельных пакетов и измеряются в микросекундном и миллисекундном диапазонах. Например, время буферизации, или время пребывания пакета в очереди коммутатора либо маршрутизатора, является характеристикой этой группы. Для анализа и обеспечения требуемого уровня краткосрочных характеристик разработано большое количество методов, получивших название методов *контроля и предотвращения перегрузок*.

Соглашение об уровне обслуживания. Основой нормального сотрудничества поставщика услуг и пользователей является договор. Такой договор заключается всегда, но далеко не всегда в нем указываются количественные требования к эффективности предоставляемых услуг. Однако существует и дугой тип договора, называемый соглашением об уровне обслуживания (Service Level Agreement, SLA). В таком соглашении поставщик услуг и клиент описывают качество предоставляемой услуги в количественных терминах. Например, в SLA может быть записано, что поставщик обязан передавать трафик клиента без потерь и с той средней скоростью, с которой пользователь направляет его в сеть. При этом оговорено, что это соглашение действует только в том случае, если средняя скорость трафика пользователя не превышает, например, 3 Мбит/с, в противном случае поставщик получает право просто не передавать избыточный трафик. Для того,

чтобы каждая сторона могла контролировать соблюдение этого соглашения, необходимо еще указать период времени, на котором будет измеряться средняя скорость, например, день, час или секунда [2].

4.1.2. Производительность

Наибольший интерес долговременные характеристики производительности сетевого оборудования (пропускная способность каналов или производительность коммутаторов и маршрутизаторов) представляют для поставщиков услуг; на их основе поставщик услуг может планировать свой бизнес, рассчитывая максимальное количество клиентов, которое он может обслужить, определяя рациональные маршруты прохождения трафика и т. п.

Однако клиента интересуют другие характеристики производительности, которые позволят ему количественно оценить, насколько быстро и качественно сеть передает его трафик. Для того чтобы определить эти характеристики, воспользуемся моделью идеальной сети.

Идеальная сеть. Составляющими задержек в сети с коммутацией пакетов являются показатели времени: а) передачи данных в канал (время сериализации); б) распространения сигнала; в) ожидания пакета в очереди; г) коммутации пакета. Две первые составляющие задержки полностью определяются свойствами каналов передачи данных (битовой скоростью и скоростью распространения сигнала в среде) и являются фиксированными для пакета фиксированной длины. Две последние составляющие зависят от характеристик сети коммутации пакетов (загрузки коммутаторов и их быстродействия) и для пакета фиксированной длины в общем случае являются переменными. Будем считать, что сеть с коммутацией пакетов работает идеально, если она передает каждый бит информации с постоянной скоростью, равной скорости распространения света в используемой физической среде. Другими словами, *идеальная сеть с КП* не вносит никаких дополнительных задержек в передачу данных помимо тех, которые вносятся каналами связи, т. е. две последние составляющие задержки равны нулю.

Результат передачи пакетов такой идеальной сетью иллюстрирует рис. 4.1. На верхней оси показаны значения времени поступления пакетов в *сеть* от узла отправителя, а на нижней – значения времени поступления пакетов в узел назначения. Говорят, что верхняя ось показывает *предложенную нагрузку* сети, а нижняя – результат передачи этой нагрузки через сеть [2].

Пусть задержка передачи пакета определяется как интервал времени между моментом отправления первого бита пакета в канал связи узлом отправления и моментом поступления первого бита пакета в узел назначения соответственно (на рис. 4.1 обозначены задержки d_1, d_2 и d_3 пакетов 1, 2 и 3 соответственно). Как видно из рисунка, идеальная сеть доставляет все

пакеты узлу назначения: а) не потеряв ни одного из них (и не исказив информацию ни в одном из них); б) в том порядке, в котором они были отправлены; в) с одной и той же минимально возможной задержкой $d_1 = d_2 = d_3$ и т. д.

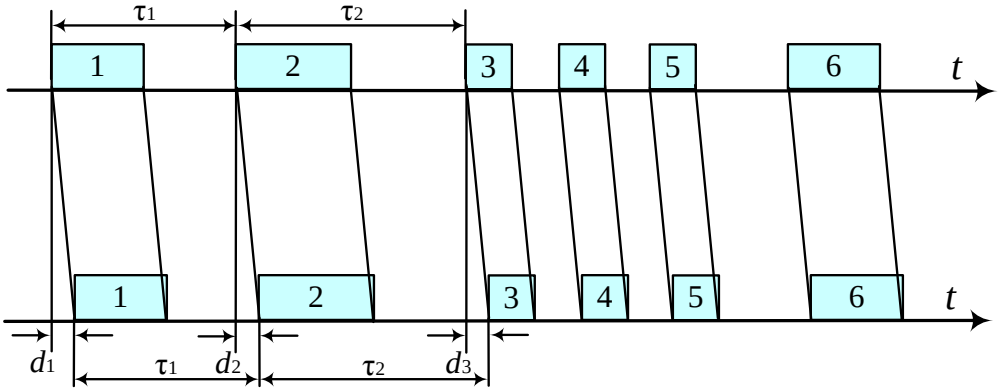


Рис. 4.1. Передача пакетов идеальной сетью

Важно, что все интервалы между соседними пакетами сеть сохраняет в неизменном виде. Например, если интервал между первым и вторым пакетами составляет при отправлении τ_1 секунд, а между вторым и третьим – τ_2, то такими же интервалы останутся в узле назначения.

Надежная доставка всех пакетов с минимально возможной задержкой и сохранением временны́х интервалов между ними удовлетворит любого пользователя сети независимо от того, трафик какого приложения он передает по сети – веб-сервиса или IP-телефонии.

Теперь рассмотрим, какие отклонения от идеала могут встречаться в реальной сети и как эти отклонения можно описывать (рис. 4.2).

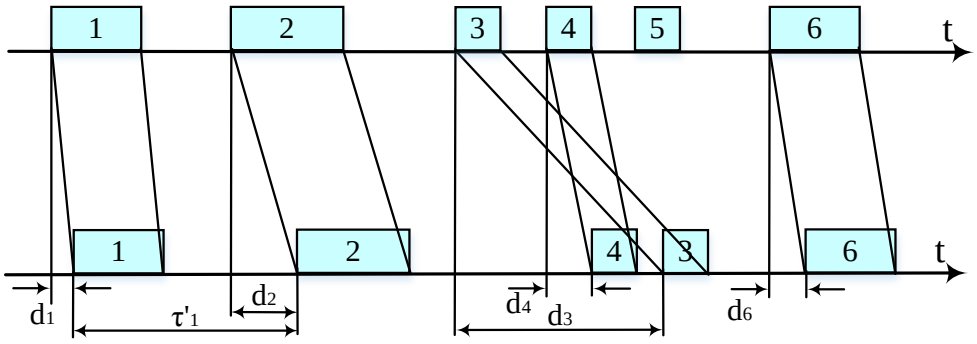


Рис. 4.2. Передача пакетов реальной сетью

Пакеты доставляются в сеть узлу назначения с *различными задержками*. Как мы уже знаем, это неотъемлемое свойство сетей с коммутацией пакетов. Случайный характер процесса образования очередей приводит к случайным задержкам, при этом задержки отдельных пакетов могут быть

значительными, в десятки раз превосходя среднюю величину задержек ($d_1 \neq d_2 \neq d_3$ и т. д.). Неравномерность задержек изменяет относительное положение пакетов в выходном потоке, а это может катастрофически сказаться на качестве работы некоторых приложений. Например, при цифровой передаче речи исходный поток представляет собой равномерно отстоящие друг от друга пакеты, несущие замеры голоса. Неравномерность интервалов между пакетами выходного потока приводит к существенным искажениям речи.

Пакеты могут доставляться узлу назначения *не в том порядке*, в котором они были отправлены, например, на рис. 4.2 пакет 4 поступил в узел назначения раньше, чем пакет 3. Такие ситуации встречаются в дейтаграммных сетях, когда различные пакеты одного потока передаются через сеть различными маршрутами, а следовательно, ожидают обслуживания в разных очередях с разным уровнем задержек. Очевидно, что пакет 3 проходил через перегруженный узел или узлы, так что его суммарная задержка оказалась настолько большой, что пакет 4 прибыл раньше него.

Пакеты *могут теряться* в сети или же приходить в узел назначения с искаженными данными, что равносильно потере пакета, так как большинство протоколов не способно восстанавливать искаженные данные, а только определяет этот факт по значению контрольной суммы в заголовке кадра.

Пакеты также могут *дублироваться* по разным причинам, например, из-за ошибочных повторных передач пакета, предпринятых протоколом, в котором таким образом обеспечивается надежный обмен данными.

В реальной сети средняя скорость информационного потока на входе узла назначения может отличаться от средней скорости потока, направленного в сеть узлом-отправителем. Виной этому являются не задержки пакетов, а их потери[7]. Так, в примере на рис. 4.2, *средняя скорость исходящего потока снижается* из-за потери пакета 5. Чем больше потерь и искажений пакетов происходит в сети, тем ниже скорость информационного потока.

Как видно из приведенного описания, существуют различные *характеристики производительности сети* (называемые также *метриками производительности сети*). Относительная важность характеристик зависит от типа приложения, трафик которого переносит сеть. Так, существуют приложения, которые очень чувствительны к задержкам пакетов, но в то же время весьма терпимы к потере отдельного пакета – примером может служить передача голоса через пакетную сеть. Примером приложения, которое, напротив, мало чувствительно к задержкам пакетов, но очень чувствительно к их потерям, является загрузка файлов.

[7] Скорость передачи данных определяется как частное от деления объема передаваемых данных на время их передачи (задержку). Из определения следует, что эта характеристика всегда является усредненной.

Статистические оценки характеристик сети. Для оценки характеристик *случайных процессов* служат статистические методы. Сами характеристики производительности сети, такие как, например, задержка пакета, являются *случайными величинами*. Статистические характеристики выявляют закономерности в поведении сети, которые устойчиво проявляются только в длительных периодах времени. Под длительным понимается интервал, в миллионы раз больший, чем время передачи одного пакета, которое в современной сети измеряется микросекундами. Так, время передачи пакета Fast Ethernet составляет около 100 мкс, Gigabit Ethernet – 10 мкс, ячейки ATM – от долей микросекунды до 3 мкс (в зависимости от скорости передачи). Поэтому для получения устойчивых результатов нужно наблюдать поведение сети по крайней мере в течение минут, а лучше – нескольких часов.

Основным инструментом статистики является так называемая *гистограмма* распределения оцениваемой случайной величины. Рассмотрим, например, гистограмму задержки пакета. Будем считать, что нам удалось измерить задержку доставки каждого из 2600 пакетов, переданных между двумя узлами сети, и сохранить полученные результаты. Эти результаты называются *выборкой* случайной величины. Чтобы получить гистограмму распределения, мы должны разбить весь диапазон измеренных значений задержек на несколько интервалов и подсчитать, сколько пакетов из нашей выборки попало в каждый интервал. Пусть все значения задержек укладываются в диапазон 20–90 мс. Разобьем его на семь интервалов по 10 мс. В каждый из этих интервалов, начиная с интервала 20–30 мс, попало 100 ($n1$), 200 ($n2$), 300 ($n3$), 300 ($n4$), 400 ($n5$), 800 ($n6$) и 500 ($n7$) пакетов соответственно. Отобразив эти числа в виде горизонтальных уровней для каждого интервала мы получим гистограмму, показанную на рис. 4.3 [2].

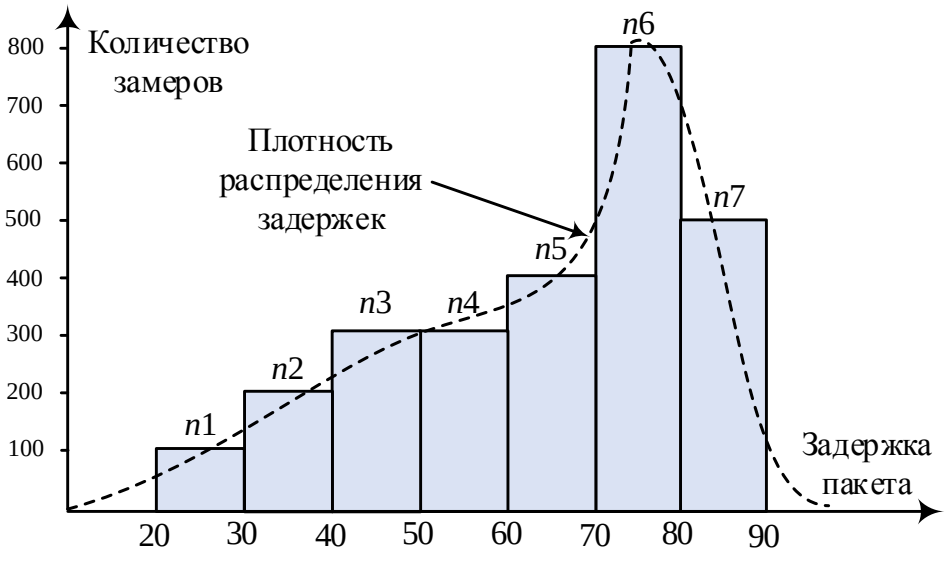


Рис. 4.3. Гистограмма распределения задержек сетью

Гистограмма задержек дает хорошее представление о производительности сети. По ней можно судить, какие уровни задержек более вероятны, а какие – менее. Чем больше период времени, в течение которого собираются данные для построения гистограммы, тем с более высокой степенью достоверности можно предсказать поведение сети в будущем. Например, пользуясь гистограммой на рис. 4.3, можно сказать, что и в будущем при измерениях задержек пакетов у 65 % пакетов задержка превысит 60 мс. Для получения такой оценки мы сложили общее количество пакетов, задержки которых попали во все интервалы, превышающие 60 мс (1700 замеров), и разделили эту величину на общее количество пакетов (2600 замеров).

При увеличении количества интервалов и времени наблюдения гистограмма в пределе переходит в непрерывную функцию, которая называется *плотностью распределения* задержки доставки пакета (показана пунктиром). Вероятность того, что значение случайной величины окажется в определенном диапазоне, равна интегралу плотности распределения случайной величины от нижней до верхней границы данного диапазона. Таким образом, может быть вычислено вероятностное значение задержки пакета.

Гистограмма дает хорошее графическое описание соответствующей характеристики, но чаще используются более компактные *статистические оценки* характеристик, которые позволяют представить характеристику *одним числом* на основе некоторой математической обработки имеющейся выборки. Для описания характеристик производительности сети используются следующие статистические оценки.

Среднее значение D вычисляется как сумма всех значений оцениваемой величины d_i, деленная на количество всех измерений N:

$$D = \frac{\sum d_i}{N}.$$

Для примера, приведенного на рис. 4.3, среднее значение равно:

(100×25+200×35+300×45+300×55+400×65+800×75+500×85)/2600 = 64,6 мс

(при вычислениях использованы средние значения интервалов) [2].

Медиана представляет такое значение оцениваемой величины, которое делит ранжированную (упорядоченную) выборку пополам, т. е. таким образом, чтобы количество замеров, значения которых меньше или равны значению медианы, равнялось количеству замеров, значения которых больше или равны значению медианы. В примере на рис. 4.3 медиана равна 70 мс, так как число замеров, значения которых меньше или равны 70 мс, составляет 1300, как и число замеров, значения которых больше или равны 70 мс.

Стандартное отклонение J есть среднее отклонение каждого отдельного замера от среднего значения оцениваемой величины:

$$J = \sqrt{\frac{\sum (d_i - D)^2}{N - 1}}.$$

Если все задержки d_i равны между собой, то вариация отсутствует, что подтверждают приведенные формулы, – в этом случае $D = d_i$ и $J = 0$.

Коэффициент вариации CV – безразмерная величина, равная отношению стандартного отклонения к среднему значению оцениваемой величины:

$$CV = \frac{J}{D}.$$

Коэффициент вариации характеризует оцениваемую величину без привязки к ее абсолютным значениям. Так, идеальный равномерный поток пакетов всегда будет обладать нулевым значением коэффициента вариации задержки пакета. Коэффициент вариации задержки пакета, равный 1, означает достаточно пульсирующий трафик, так как средние отклонения интервалов от некоторого среднего периода следования пакетов равны этому периоду.

Квантиль (процентиль) – это значение оцениваемой величины, делящее ранжированную выборку на две части так, что процент замеров, значения которых меньше или равны значению квантиля, равен некоторому заданному уровню. В этом определении фигурируют два числа: заранее заданный процент и найденное по нему и замерам выборки значение квантиля. Например, для выборки на рис. 4.3 значение 50-процентного квантиля будет 70 мс, так как 50 % замеров выборки (т. е. 1300 замеров из 2600) имеют значения, меньшие или равные 70 мс. Значение 80-процентного квантиля равно 80 мс, так как именно 80 % всех замеров имеют значения задержки менее 80 мс. Медиана является частным случаем квантиля – это 50-процентный квантиль. Для оценки характеристик сети обычно используют квантили с большим значением процента, например, 90-, 95- или 99-процентные квантили[8].

Рассмотренные статистические методы помимо задержек применяются и ко всем другим характеристикам – времени ожидания в буфере, времени коммутации, количеству потерянных пакетов и др.

Активные и пассивные измерения в сети. Чтобы оценить некоторую характеристику производительности, необходимо провести измерения последовательности пакетов, поступающих на некоторый интерфейс сетевого устройства. Существуют активные и пассивные измерения.

[8] Если пользователю скажут, что сеть будет обеспечивать уровень задержек в 100 мс с вероятность 0,5, то это его не очень обрадует, так как он ничего не будет знать об уровне задержек половины своих пакетов.

Активные измерения основаны на генерации в узле-источнике специальных «измерительных» пакетов. Эти пакеты должны пройти через сеть тот же путь, что и пакеты, характеристики которых подлежат оценке. Измерения в узле назначения проводятся на последовательности «измерительных» пакетов. Для измерения задержки пакетов некоторого приложения *A*, которые передаются от компьютера-клиента приложения *A* компьютеру-серверу приложения *A* через сеть, в сети устанавливается два дополнительных компьютера: сервер-генератор и сервер-измеритель (рис. 4.4). Чтобы измеряемые значения были близки к значениям пакетов приложения *A*, нужно, чтобы измерительные пакеты проходили через сеть по тому же пути, что и пакеты приложения *A*. В нашем примере эта цель достигается за счет подключения дополнительных узлов к портам тех же коммутаторов *S*1 и *S*2, к которым подключены оригинальные узлы. Кроме того, нужно, чтобы измерительные пакеты как можно больше «походили» на оригинальные пакеты – размерами, признаками, помещенными в заголовки пакетов. Это требуется для того, чтобы сеть обслуживала их так же, как оригинальные пакеты [2].

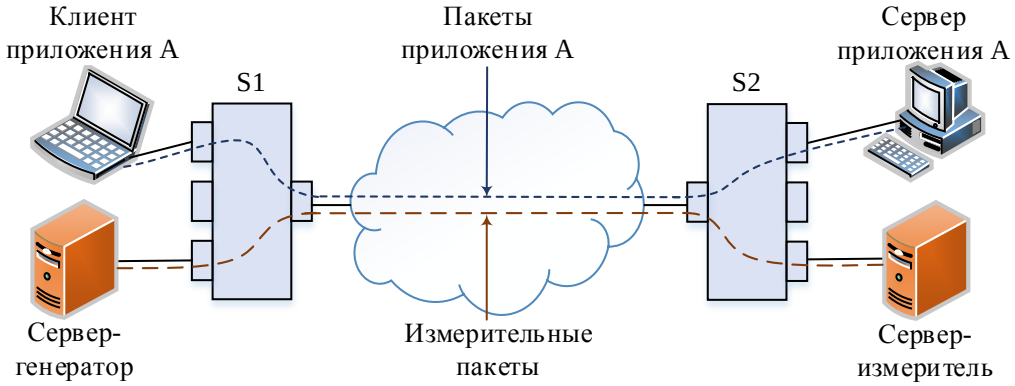


Рис. 4.4. Схема активных измерений

Измерительные пакеты не должны генерироваться слишком часто, иначе нагрузка сети может измениться и результаты замеров будут отличаться от тех, которые были бы получены в отсутствие измерительных пакетов, т. е. измерения не должны менять условий работы сети. Возникает естественный вопрос: зачем нужно решать столько лишних проблем: размещать дополнительное оборудование, создавать условия для измерительных пакетов, близкие к условиям обработки оригинальных пакетов, и в то же время стараться не изменить нагрузку сети? Не проще ли измерять параметры реальных пакетов? Ответ заключается в том, что *активная схема упрощает процесс проведения измерений и позволяет добиться их высокой точности*.

Во-первых, так как сервер-генератор создает измерительные пакеты, то он легко может использовать специальный формат пакетов для того,

чтобы поместить в них необходимую для измерения информацию, например, временну́ю метку (time-stamp) отправки пакета. Затем сервер-измеритель использует эту временну́ю метку для вычисления времени задержки.

Во-вторых, для того, чтобы измерения задержки были точными, нужна хорошая синхронизация сервера-генератора и сервера-измерителя. Так как в схеме активных измерений они представляют собой выделенные узлы, такой синхронизации добиться проще, чем в случае синхронизации клиентской и серверной частей приложения *A* на обычных компьютерах.

В-третьих, иногда у инженеров, проводящих измерения, просто нет доступа к компьютерам, на которых работают приложения, чтобы установить там программное обеспечение для требуемых измерений.

В-четвертых, если такой доступ и существует, то ОС клиента и сервера и их аппаратная платформа, скорее всего, не оптимизированы для измерений временных интервалов, а значит, вносят большие искажения в результаты.

Однако преимущества активной схемы не являются абсолютными. Иногда предпочтительной является схема пассивных измерений.

Пассивные измерения основаны на измерениях характеристик реального трафика. Эту схему иллюстрирует рис. 4.5 [2].

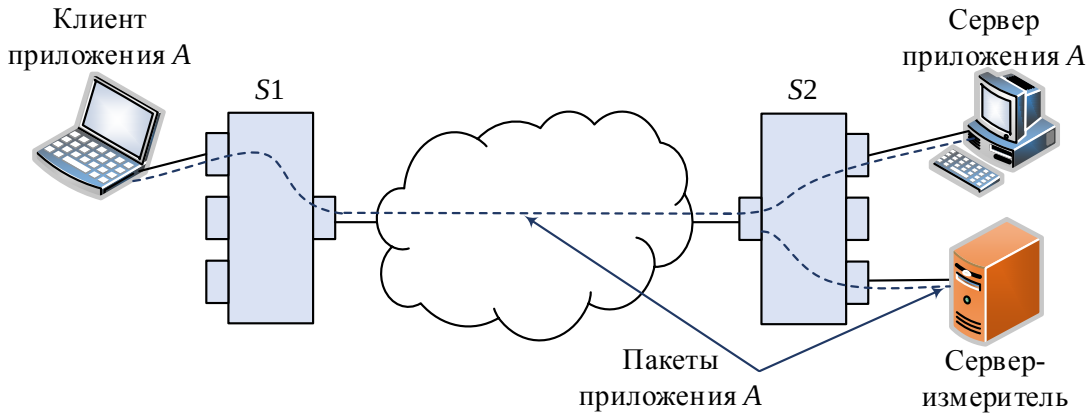


Рис. 4.5. Схема пассивных измерений

Приводя аргументы в пользу схемы активных измерений, мы, в сущности, описали проблемы, которые приходится решать при использовании схемы пассивных измерений: сложности синхронизации клиента и сервера, дополнительные и неопределенные задержки, вносимые операционными системами этих компьютеров, отсутствие в заголовке используемых приложением пакетов поля для переноса по сети временно́й метки. Частично эти проблемы решаются за счет применения отдельного сервера измерителя. Этот сервер принимает тот же входной поток пакетов, что и один из узлов, участвующий в обмене пакетами, характеристики которых нужно измерить (на рис. 4.5 показан случай, когда сервер-измеритель ставится в параллель

с сервером приложения *A*). Для того чтобы сервер-измеритель получал тот же входной поток пакетов, что и оригинальный узел, обычно прибегают к дублированию измеряемого трафика на порт, к которому подключен сервер-измеритель. Такую функцию, называемую *зеркализацией портов*, поддерживают многие коммутаторы локальных сетей. Сервер-измеритель может работать под управлением специализированной операционной системы, оптимизированной для выполнения точных измерений временны́х интервалов.

Сложнее реши́ть проблему синхронизации. Некоторые протоколы переносят временны́е отметки в своих служебных полях, так что если, например, приложение *A* использует такой протокол, то часть проблемы решается. Однако и в этом случае точность системного времени в компьютере клиента приложения *A*, скорее всего, невысока. Поэтому в пассивном режиме измеряют те характеристики, которые не требуют синхронизации передатчика и приемника, например, оценивают долю потерянных пакетов.

Возможным вариантом пассивной схемы измерений является отсутствие выделенного сервера-измерителя. Некоторые приложения сами выполняют измерения задержек поступающих пакетов, например, такими функциями обладают многие приложения IP-телефонии и видеоконференций, так как информация о задержках пакетов помогает определить возможную причину неудовлетворительного качества работы приложения.

Характеристики задержек и потерь пакетов. Для оценки производительности сети используются различные характеристики задержек и потерь пакетов, в том числе: а) односторонняя задержка пакетов; б) вариация задержки пакета; в) время реакции сети; г) время оборота пакета.

Единичное значение односторонней задержки пакетов (One-Way Delay Metric, OWD) определенного типа определяется как интервал времени между моментом помещения в исходящую линию связи первого бита пакета узлом-отправителем и моментом приема последнего бита пакета с входящей линии связи узла-получателя.

Под определенным типом пакета понимается пакет, который имеет некоторый заранее определенный набор признаков, например, размер пакета, тип приложения, сгенерировавшего пакет, тип протокола транспортного уровня, который доставил пакет. Цель использования набора признаков состоит в том, чтобы выделить из общего потока пакетов, приходящего в узел назначения, те пакеты, характеристики которых измеряются.

Так как в этом определении учитывается время буферизации пакета узлом-получателем, то задержка зависит от размера пакета и для получения сопоставимых результатов желательно в определении типа пакетов задавать определенный размер пакета. Определение времени задержки с учетом буферизации упрощает измерение времени прихода пакета, так как программно его можно измерить только после завершения записи всего пакета

в буфер ОС. Кроме того, при получении только одного первого бита пакета невозможно понять, относится ли пакет к интересующему типу. Если пакет не прибыл в узел назначения за достаточно большое время, пакет считается утерянным, а его задержка неопределенной (равной бесконечности).

Последовательность замеров рекомендуется выполнять в случайные моменты времени, подчиняющиеся распределению Пуассона. Такой порядок выбора времени замеров позволяет избежать возможной синхронизации измерений с любыми периодическими флюктуациями в поведении сети.

Для одностороннего времени задержки используются следующие *статистические оценки*: а) квантиль для некоторого процента; б) среднее значение задержки; в) минимальное значение задержки (в выборке).

Квантили удобны для оценки задержек в тех случаях, когда процент потерь пакетов высок. Для вычисления квантиля потерянные пакеты можно рассматривать как пакеты, пришедшие с бесконечно большой задержкой.

Вариация задержки пакета (IP Packet Delay Variation, IPDV), которую также называют *джиттером* (*jitter*), очень важна для некоторых приложений. Так, при воспроизведении видеоклипа сама по себе задержка не очень существенна. Например, если все пакеты задерживаются ровно на десять секунд, то качество воспроизведения не пострадает, а тот факт, что картинка появляется чуть позже, чем ее отослал сервер, пользователь даже не заметит (однако в интерактивных видеоприложениях, таких как видеоконференции, подобная задержка будет, конечно, уже ощутимо раздражать). А вот если задержки постоянно изменяются в пределах от нуля до 10 секунд, то качество воспроизведения клипа заметно ухудшится, для компенсации таких переменных задержек нужна предварительная буферизация поступающих пакетов в течение времени, превышающего вариацию задержки.

Единичное значение оценки *вариации задержки* определяется стандартом как разность односторонних задержек для пары пакетов определенного типа, полученных на интервале измерений *T*.

Как и для односторонней задержки, тип пакета может задаваться любыми признаками, при этом размеры обоих пакетов должны быть одинаковыми. Выбор пары пакетов на интервале измерения *T* должен осуществляться в соответствии с некоторым заранее принятым правилом, например, пары могут образовываться из всех последовательных пакетов, полученных на интервале.

Время реакции сети представляет собой интегральную характеристику производительности сети с точки зрения пользователя. Именно эту характеристику имеет в виду пользователь, когда говорит: «Сегодня сеть работает медленно». *Время реакции сети определяется как интервал времени между отправкой запроса пользователя к какой-либо сетевой службе и получением*

ответа на этот запрос. Время реакции сети можно представить в виде нескольких слагаемых (рис. 4.6): времени подготовки запросов на клиентском компьютере ($t_{клиент1}$), времени передачи запросов между клиентом и сервером через сеть ($t_{сеть}$), времени обработки запросов на сервере ($t_{сервер}$), времени передачи ответов от сервера клиенту через сеть (снова $t_{сеть}$) и времени обработки получаемых от сервера ответов на клиентском компьютере ($t_{клиент2}$) [2].

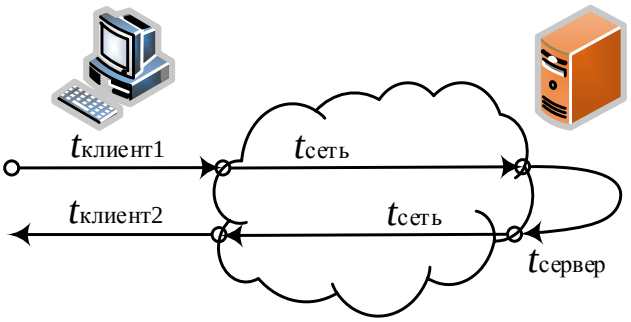


Рис. 4.6. Время реакции
и время оборота

Время реакции сети характеризует сеть в целом, в том числе качество работы аппаратного и программного обеспечения серверов. Для того, чтобы отдельно оценить транспортные возможности сети, используется другая характеристика – время оборота данных по сети.

Время оборота (Round Trip Time, RTT) пакета является составляющей времени реакции сети – это «чистое» время транспортировки данных от узла отправителя до узла назначения и обратно без учета времени, затраченного узлом назначения на подготовку ответа: $RTT = 2 \times t_{сеть}$.

Единичное значение *времени оборота* определяется как интервал времени между отправкой первого бита пакета определенного типа узлом-отправителем узлу-получателю и получением последнего бита этого пакета узлом-отправителем после того, как пакет был получен узлом-получателем и отправлен обратно. При этом узел-получатель должен отправить пакет узлу-отправителю как можно быстрее, чтобы не вносить искажения за счет времени обработки. *Последовательность замеров* RTT выполняется через случайные интервалы, подчиняющиеся распределению Пуассона.

RTT является удобной для измерений характеристикой, так как для ее получения не требуется синхронизации узла-отправителя и узла-получателя: узел-отправитель ставит временную отметку на отправляемый пакет, а затем по прибытии его от узла-получателя сравнивает эту отметку со своим текущим системным временем. Однако информативность RTT меньше, чем информативность односторонней задержки, так как информация о задержке в каждом направлении теряется, а это может затруднить поиск проблемного пути в сети.

Характеристики скорости передачи. *Скорость передачи данных* (information rate) измеряется на каком-либо промежутке времени как частное от деления объема переданных данных за этот период на продолжительность периода. Таким образом, данная характеристика всегда по определению является средней скоростью передачи данных. В зависимости от величины интервала, на котором измеряется скорость, для этой характеристики используется одно из двух наименований: средняя или пиковая скорость.

Средняя скорость передачи данных (Sustained Information Rate, SIR) – это среднесрочная характеристика. Она определяется на достаточно большом периоде времени, достаточном, чтобы можно было говорить об устойчивом поведении такой случайной величины, которой является скорость.

Средняя скорость должна использоваться в паре с параметром, оговаривающим период контроля этой величины, например, 10 секунд. Это означает, что скорость информационного потока вычисляется каждые 10 секунд.

Пиковая скорость передачи данных (Peak Information Rate, PIR) – это наибольшая скорость, которую разрешается достигать пользовательскому потоку в течение оговоренного небольшого периода времени *T*.

Этот период обычно называют *периодом пульсации*. Как правило, он выбирается существенно меньшим, чем период измерения средней скорости передачи. Пиковая скорость является краткосрочной характеристикой и позволяет оценить способность сети справляться с пиковыми нагрузками, характерными для пульсирующего трафика и приводящими к перегрузке. Если в SLA оговорены обе скорости (SIR и PIR), очевидно, что периоды пульсации должны сопровождаться периодами относительного «затишья», когда скорость падает ниже средней.

Величина пульсации (обычно обозначаемая *B*) служит для оценки емкости буфера коммутатора, необходимого для хранения данных во время перегрузки. Величина пульсации равна общему объему данных, поступающих на коммутатор в течение разрешенного интервала *T* (периода пульсации) передачи данных с пиковой скоростью (PIR): $B = PIR \times T$.

Еще одной характеристикой скорости передачи является *коэффициент пульсации трафика* – это отношение максимальной скорости на какомлибо небольшом периоде времени к средней скорости трафика, измеренной на длительном периоде времени. Неопределенность временных периодов делает коэффициент пульсации качественной характеристикой трафика.

Скорость передачи данных можно измерять между любыми двумя узлами. Из-за последовательного характера передачи данных различными элементами сети общая пропускная способность любого составного пути в сети будет равна минимальной из пропускных способностей составляющих элементов маршрута. Для повышения пропускной способности составного пути необходимо в первую очередь обратить внимание на самые медленные элементы, называемые *узкими местами* (*bottleneck*).

4.1.3. Надежность

Характеристикой потерь пакетов является доля потерянных пакетов (обозначим ее L), равная отношению количества потерянных пакетов (N_L) к общему количеству переданных пакетов (N): $L = N_L/N$. Может также использоваться характеристика, оперирующая не количеством потерянных и переданных пакетов, а объемами данных, содержащихся в этих пакетах.

Доступность и отказоустойчивость. Для описания надежности отдельных устройств служат такие показатели надежности, как *среднее время наработки на отказ, вероятность отказа, интенсивность отказов*. Однако эти показатели пригодны только для оценки надежности простых элементов и устройств, которые при отказе любого своего компонента переходят в неработоспособное состояние. Сложные системы, состоящие из многих компонентов, могут при отказе одного из компонентов сохранять свою работоспособность. В связи с этим для оценки надежности сложных систем применяется другой набор характеристик.

Доступность (availability) означает долю времени, в течение которого система или служба находится в работоспособном состоянии.

Доступность является долговременной статистической характеристикой и измеряется на большом промежутке времени (день, месяц, или год). Примером высокого уровня доступности является коммуникационное оборудование телефонных сетей, лучшие представители которого обладают так называемой доступностью «пять девяток» – это означает, что доступность равна 0,99999, что соответствует чуть более чем пяти минутам простоя в год. Оборудование и услуги передачи данных только стремятся к такому рубежу, но рубеж трех девяток уже достигнут. Доступность услуги важна как пользователям, так и поставщикам услуг [2].

Еще одной характеристикой надежности сложных систем является *отказоустойчивость* (*fault tolerance*), т. е. способность системы скрывать от пользователя отказ ее отдельных элементов. Например, если коммутатор оснащен двумя коммутационными центрами, работающими параллельно, то отказ одного из них не приведет к полному останову коммутатора, однако производительность коммутатора снизится вдвое. В отказоустойчивой системе отказ одного из ее элементов приводит к снижению качества ее работы – деградации, а не к полному останову.

4.1.4 Характеристики сети поставщика услуг

Рассмотрим основные характеристики, которыми оперирует поставщик услуг, оценивая эффективность всей сети. Эти характеристики – *расширяемость, масштабируемость, управляемость* и *совместимость* – являются качественными, т. е. не могут быть выражены числами и соотношениями.

Термины «расширяемость» и «масштабируемость» иногда неверно используют как синонимы.

Расширяемость означает возможность сравнительно простого добавления отдельных компонентов сети (пользователей, компьютеров, приложений, служб), наращивания длины сегментов кабелей и замены существующей аппаратуры более мощной.

При этом принципиально важно, что простота расширения системы иногда может обеспечиваться в *определенных пределах*. Например, локальная сеть Ethernet, построенная на основе одного разделяемого сегмента коаксиального кабеля, обладает хорошей расширяемостью в том смысле, что позволяет легко подключать новые станции. Однако такая сеть имеет ограничение на число станций – оно не должно превышать 30–40. При подключении к сегменту большего числа станций (до 100) резко снижается производительность сети. Наличие такого ограничения и является признаком плохой масштабируемости системы при ее хорошей расширяемости.

Масштабируемость означает, что сеть позволяет наращивать количество узлов и протяженность связей в очень широких пределах, при этом производительность сети не снижается.

Для обеспечения масштабируемости сети приходится применять дополнительное коммуникационное оборудование и специальным образом структурировать сеть. Обычно масштабируемое решение обладает многоуровневой иерархической структурой, которая позволяет добавлять элементы на каждом уровне иерархии без изменения главной идеи проекта. Примером хорошо масштабируемой является Интернет, технология которого (TCP/IP) оказалась способной поддерживать сеть в масштабах земного шара.

Не только сама сеть должна быть масштабируемой, но и устройства, работающие на магистрали сети, также должны обладать этим свойством, так как рост сети не должен приводить к необходимости постоянной смены оборудования. Поэтому магистральные коммутаторы и маршрутизаторы строятся обычно по модульному принципу, позволяя наращивать количество интерфейсов и производительность обработки пакетов в широких пределах.

Управляемость сети подразумевает возможность централизованно контролировать состояние основных элементов сети, выявлять и разрешать проблемы, анализировать производительность и планировать развитие сети.

Управляемость предполагает наличие в сети некоторых автоматизированных средств администрирования, которые взаимодействуют с программным и аппаратным обеспечением сети с помощью коммуникационных протоколов. В идеале средства администрирования сети обеспечивают наблюдение и контроль за каждым элементом сети и, обнаружив проблему, активизируют определенное действие, например, исправляют ситуацию и уведомляют администратора о том, что произошло и какие шаги пред-

приняты. Одновременно с этим система администрирования должна накапливать данные, на основании которых можно планировать развитие сети.

Решая тактические задачи, администраторы и технический персонал сталкиваются с ежедневными проблемами поддержания работоспособности сети. Эти задачи требуют быстрого решения, обслуживающий сеть персонал должен оперативно реагировать на сообщения о неисправностях, поступающие от пользователей или автоматических средств администрирования сети. Постепенно становятся заметными более общие проблемы производительности, конфигурирования сети, обработки сбоев и безопасности данных, требующие стратегического подхода, т. е. *планирования* сети. Планирование, кроме того, подразумевает умение прогнозировать изменения в требованиях пользователей к сети, решение вопросов о применении новых приложений, новых сетевых технологий и т. п.

Совместимость, или *интегрируемость* сети означает, что сеть способна включать в себя самое разнообразное программное и аппаратное обеспечение, т. е. в ней могут сосуществовать различные ОС, стеки коммуникационных протоколов, а также аппаратные средства и приложения разных производителей. Сеть, состоящая из разнотипных элементов, называется неоднородной, или гетерогенной, а если гетерогенная сеть работает без проблем, то она является интегрированной. Основной путь построения интегрированных сетей – использование открытых стандартов и спецификаций.

Выводы

Главным требованием, предъявляемым к компьютерной сети, является обеспечение высокого качества предоставляемых сетью услуг. В широком понимании в понятие «качество обслуживания» включают все возможные характеристики услуг и сети, желательные для пользователя. Наиболее важные формализованные характеристики сети относятся к ее производительности и надежности.

Производительность сети оценивается с помощью статистических характеристик двух типов: характеристик скорости передачи информации и характеристик задержек передачи пакетов. В первую группу входят средняя и максимальная скорости на периоде пульсации, а также длительность этого периода. Во вторую группу входят: средняя величина задержки, средняя вариация задержки (джиттер), коэффициент вариации, а также максимальные значения задержки и вариации задержки.

Для оценки надежности сетей применяются различные характеристики, в том числе: доля потерь пакетов, коэффициент доступности, означающий долю времени, в течение которого система может быть использована, отказоустойчивость – способность системы работать в условиях отказа некоторых ее элементов.

Надежность транспортных услуг сети обеспечивается надежностью ее компонентов (каналов и коммуникационного оборудования), наличием альтернативных маршрутов, а также повторной передачей потерянных или искаженных пакетов.

Особую важность для поставщика услуг представляют такие качественные характеристики сети, как ее масштабируемость, расширяемость и управляемость.

Контрольные вопросы

1. Назовите характеристики качества обслуживания, которые интересуют: а) только пользователя; б) только поставщика услуг; в) и пользователя, и поставщика.
2. Между какими сторонами заключается соглашение об уровне обслуживания?
3. Зависит ли средняя скорость потока от величины задержек пакетов?
4. Предложите набор характеристик, которые вы хотели бы видеть в SLA, если вы планируете передавать через сеть трафик IP-телефонии.
5. Какой вид представления информации используется для результатов измерения задержек пакетов?
6. Могут ли различаться краткосрочные и долгосрочные значения одной и той же характеристики, например, средней скорости потока?
7. Какие составляющие задержки пакета являются фиксированными для пакета фиксированной длины?
8. Может ли трафик передаваться с большими задержками, но без джиттера?
9. Чем расширяемость сети отличается от масштабируемости?
10. Является ли коэффициент пульсации трафика количественной характеристикой?

4.2. Качество обслуживания

4.2.1. Постановка задачи обеспечения качества обслуживания

Методы обеспечения качества обслуживания (Quality of Service, QoS) занимают сегодня важное место в арсенале технологий сетей с коммутацией пакетов и обеспечивают устойчивую работу современных мультимедийных приложений, таких как IP-телефония, видео- и радиовещание, интерактивное дистанционное обучение и т. п. В методах обеспечения QoS используются различные алгоритмы управления очередями, резервирования и обратной связи, позволяющие снизить негативные последствия временных перегрузок, возникающих в сетях с коммутацией пакетов.

Очереди являются неотъемлемым атрибутом сетей с коммутацией пакетов. Сам принцип работы таких сетей подразумевает наличие буфера у каждых входного и выходного интерфейсов коммутатора пакетов. Буферизация пакетов во время перегрузок представляет собой основной механизм поддержания пульсирующего трафика, обеспечивающий *высокую производительность* сетей этого типа (в сетях с коммутацией каналов про-

межуточная буферизация данных не поддерживается). В то же время очереди означают неопределенную задержку при передаче пакетов через сеть, а в некоторых случаях и потери пакетов из-за переполнения буфера коммутатора или маршрутизатора, отведенного под очередь [2].

Операторам пакетных сетей, заинтересованных в передаче пульсирующего трафика, необходимо достижение *компромисса* между предельной загрузкой своей сети и требуемым клиентами качеством обслуживания их трафика. К характеристикам QoS относят: а) одностороннюю (от отправителя к получателю) и двустороннюю (от отправителя к получателю и обратно) задержку пакетов; б) вариацию задержек пакетов; в) потери пакетов.

В методах QoS используются механизмы снижения негативных последствий пребывания пакетов в очередях с сохранением положительной роли очередей. Большинство из них учитывает существования в сети трафика различного типа, а именно то, что каждый тип трафика предъявляет разные требования к характеристикам производительности и надежности сети.

Определяющим фактором качества обслуживания является уровень загрузки сети трафиком, т. е. *уровень использования пропускной способности линий связи* сети. Напомним, что пропускная способность – это характеристика физического канала, которая представляет собой максимально возможную скорость передачи информации по этому каналу.

Пропускную способность сети изменить непросто, так как она определяется быстродействием интерфейсов коммуникационного оборудования и качеством линий связи, их соединяющих. Повышение пропускной способности сети – это дорогостоящая операция, связанная с заменой оборудования, которую операторы сетей проводят не очень часто, раз в несколько лет.

Если уровень использования пропускной способности постоянно является достаточно низким, то трафик всех приложений обслуживается с высоким качеством бóльшую часть времени. Такое состояние сети называют «недогруженным» либо используют термин *сеть с избыточной пропускной способностью*. Постоянно поддерживать все части сети в недогруженном состоянии весьма дорого, но для наиболее ответственной части сети, такой как магистраль, подобный подход применяется.

Методы QoS основаны на тонком *перераспределении* имеющейся пропускной способности между трафиком различного типа в соответствии с требованиями приложений. Эти методы усложняют сетевые устройства, которые теперь должны «знать» требования всех классов трафика, уметь их классифицировать и распределять пропускную способность сети между ними. Последнее свойство обычно достигается за счет использования для каждого выходного интерфейса коммуникационного оборудования нескольких очередей пакетов вместо одной очереди; при этом в очередях

применяют различные алгоритмы обслуживания пакетов, чем и достигается дифференцированное обслуживание трафика различных классов. Поэтому методы QoS часто ассоциируются с *техникой управления очередями*.

Помимо собственно техники организации очередей к методам QoS относят методы контроля параметров потока трафика, так как для гарантированного качественного обслуживания нужно быть уверенным, что обслуживаемые потоки соответствуют определенному профилю. Эта группа методов QoS получила название *методов кондиционирования трафика*.

Особое место занимают *методы обратной связи*, которые предназначены для уведомления источника трафика о перегрузке сети. Эти методы рассчитаны на то, что при получении уведомления источник снизит скорость выдачи пакетов в сеть и тем самым ликвидирует причину перегрузки.

К методам QoS тесно примыкают *методы инжиниринга трафика*, согласно которым маршруты передачи данных управляются таким образом, чтобы обеспечить сбалансированную загрузку всех ресурсов сети и исключить за счет этого перегрузку коммуникационных устройств и образование длинных очередей.

4.2.2. Приложения и качество обслуживания

Существующие приложения предъявляют разные требования к QoS. Основными критериями классификации приложений являются три характеристики порождаемого ими трафика: а) относительная предсказуемость скорости передачи данных; б) чувствительность трафика к задержкам пакетов; в) чувствительность трафика к потерям и искажениям пакетов.

В отношении **предсказуемости скорости передачи данных** приложения делятся на приложения с потоковым и пульсирующим трафиком.

Приложения с потоковым трафиком (stream) порождают равномерный поток данных, который поступает в сеть с *постоянной битовой скоростью* (Constant Bit Rate, CBR). При КП трафик таких приложений представляет собой последовательность пакетов одинакового размера (B бит), следующих друг за другом через один и тот же интервал времени T (рис. 4.7). CBR потокового трафика может быть вычислена путем усреднения на интервале: $CBR = B/T$ бит/с. В общем случае постоянная битовая скорость потокового трафика меньше *номинальной битовой скорости протокола*[9].

Приложения с пульсирующим трафиком (*burst*) отличаются высокой степенью непредсказуемости; в этих приложениях периоды молчания сменяются пульсациями, в течение которых пакеты «плотно» следуют друг за другом. В результате трафик характеризуется *переменной битовой ско-*

[9] Например, номинальная скорость протокола Ethernet равна 10 Мбит/с.

ростью (Variable Bit Rate, VBR), что иллюстрирует рис. 4.7. Так, при работе приложений файлового сервиса интенсивность трафика, генерируемого приложением, может падать до нуля, когда файлы не передаются, и повышаться до максимально доступной, когда файловый сервер передает файл [2].

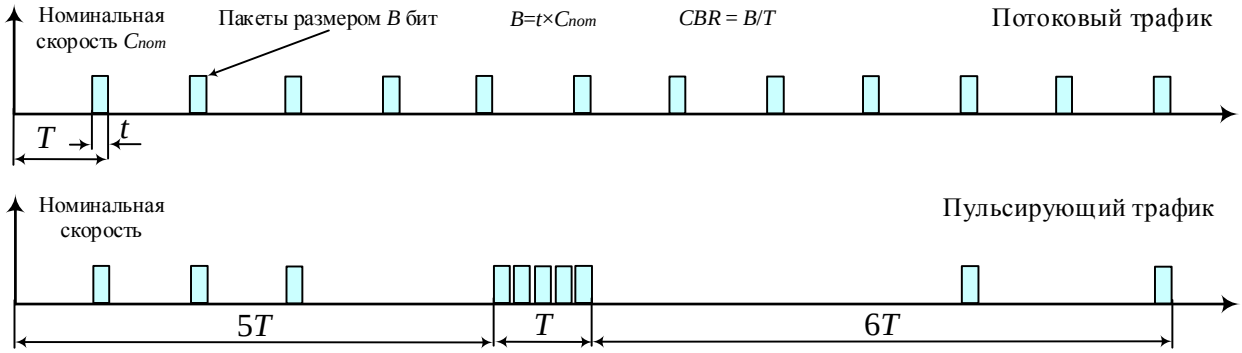


Рис. 4.7. Потоковый и пульсирующий трафики

На рис. 4.7 для пульсирующего трафика показана ситуация, когда на периоде длительностью $5T$ было передано три пакета, затем на периоде длительностью T было передано 5 пакетов, а на периоде длительностью $6T$ – 2 пакета. Пиковой скоростью трафика является скорость на втором периоде, когда за время T было передано 5 пакетов, поэтому PIR = $5B/T$. В то же время средняя скорость передачи данных (Sustained Information Rate, SIR) на всех периодах наблюдений составила $10B/12T = 5B/6T$.

Для приведенного примера можно подсчитать *коэффициент пульсации*. По определению он равен отношению пиковой скорости на каком-либо небольшом периоде времени к средней скорости трафика, измеренной на длительном периоде времени: PIR/SIR = $(5B/T)/(5B/6T) = 6$.

Значения коэффициентов пульсации у потокового и пульсирующего трафиков существенно различаются. У приложений с пульсирующим трафиком он обычно находится в пределах от 2 до 100, а у потоковых приложений он близок к 1. В локальных сетях коэффициент пульсации обычно выше, чем в глобальных, поскольку на магистралях глобальных сетей трафик представляет собой сумму трафиков многих источников, что приводит к сглаживанию результирующего трафика.

Еще один критерий классификации приложений по типу трафика – их **чувствительность к задержкам пакетов**. Далее перечислены типы приложений в порядке повышения чувствительности к задержкам пакетов.

Асинхронные приложения практически не имеют ограничений на время задержки (эластичный трафик). Пример приложения – электронная почта.

Интерактивные приложения. Задержки могут быть замечены пользователями, но они не сказываются негативно на функциональности приложений. Пример – текстовый редактор, работающий с удаленным файлом.

Изохронные приложения имеют порог чувствительности к вариациям задержек, при превышении которого резко снижается функциональность приложений. Пример – передача голоса, когда при превышении порога вариации задержек в 100–150 мс резко снижается качество голоса.

Сверхчувствительные к задержкам приложения. Задержка доставки данных сводит функциональность приложения к нулю. Пример – приложения, управляющие техническим объектом в реальном времени. При запаздывании управляющего сигнала на объекте может произойти авария.

Вообще говоря, интерактивность приложения всегда повышает его чувствительность к задержкам. Например, широковещательная рассылка аудиоинформации может выдерживать значительные задержки в передаче пакетов (оставаясь чувствительным к вариациям задержек), а интерактивный телефонный или телевизионный разговор их не терпит, что хорошо заметно при трансляции разговора через спутник.

Наряду с приведенной дифференциацией чувствительности приложений к задержкам и их вариациям существует и более грубое деление приложений по этому признаку на два класса: асинхронные и синхронные. К *асинхронным* относят те приложения, которые нечувствительны к задержкам передачи данных в очень широком диапазоне, вплоть до нескольких секунд, а все остальные приложения, на функциональность которых задержки влияют существенно, относят к *синхронным* приложениям. *Интерактивные* приложения могут относиться как к асинхронным (например, текстовый редактор), так и к синхронным (например, видеоконференция).

Последним критерием классификации приложений является их **чувствительность к потерям пакетов**. Здесь делят приложения на две группы.

Приложения, чувствительные к потере данных. Практически все приложения, передающие алфавитно-цифровые данные (к которым относятся текстовые документы, коды программ, числовые массивы и т. п.), обладают высокой чувствительностью к потере даже небольших фрагментов данных. Такие потери часто ведут к полному обесцениванию остальной успешно принятой информации. Все традиционные сетевые приложения (файловый сервис, электронная почта и т. д.) относятся к этому типу приложений.

Приложения, устойчивые к потере данных. К этому типу относятся многие приложения, передающие трафик с информацией об инерционных физических процессах. Устойчивость к потерям объясняется тем, что небольшое количество отсутствующих данных можно определить на основе принятых. Так, при потере одного пакета, несущего несколько последовательных замеров голоса, отсутствующие замеры при воспроизведении голоса могут быть заменены аппроксимацией на основе соседних значений. Однако устойчивость к потерям имеет свои пределы, поэтому процент потерянных пакетов не может быть большим (например, не более 1 %).

4.2.3. Управление очередями

Для понимания механизмов поддержки QoS полезно исследовать процесс образования очередей на сетевых устройствах. Здесь оказывается полезной дисциплина теорией очередей, которая рассматривает временны́е процессы образования очередей в буфере устройства, в который поступает случайный поток заявок на обслуживание. Модели теории очередей позволяют оценить среднюю длину очереди в буфере и среднее время ожидания заявки в очереди в зависимости от характеристик входного потока и времени обслуживания. При анализе очередей в компьютерных сетях заявками на обслуживание являются пакеты данных, а обслуживающими устройствами – интерфейсы коммутаторов и маршрутизаторов (рис. 4.8) [2].

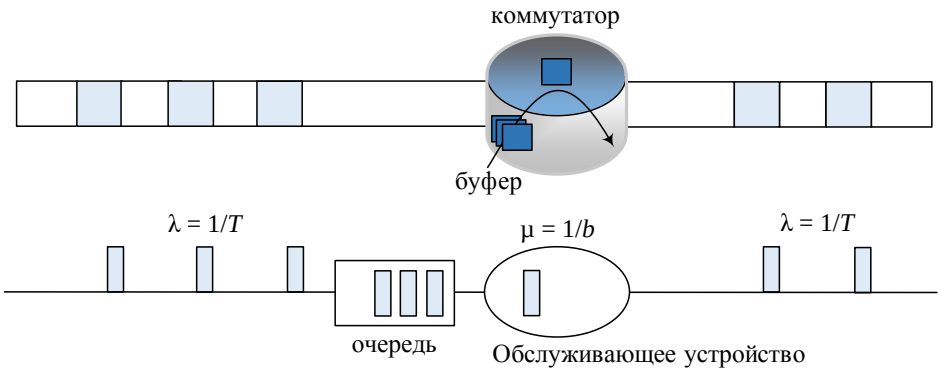


Рис. 4.8. Выходной интерфейс коммутатора
как разделяемый ресурс

Среднее время обслуживания заявки μ соответствует среднему времени продвижения пакета коммутатором из входного буфера в выходной канал.

Модели теории очередей упрощенно описывают процессы в коммутаторе/маршрутизаторе и полезны для понимания основных факторов, влияющих на величину очереди в буфере сетевого устройства и среднее время пребывания пакета в буфере. Одним из таких факторов является коэффициент загрузки (использования) интерфейса, равный отношению средней интенсивности потока поступления пакетов λ в интерфейс к среднему времени обработки пакета μ (это время включает все стадии продвижения пакета в выходной канал). В теории очередей этот коэффициент обозначается ρ.

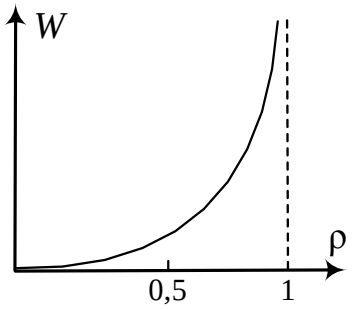


Рис. 4.9. Зависимость среднего времени ожидания заявки от коэффициента использования ресурса

На рис. 4.9 показана зависимость среднего времени ожидания пакета в буфере *W* от ρ. Как видно из поведения кривой, коэффициент ρ играет ключевую роль в образовании очереди. Если значение ρ близко к нулю, то среднее время ожидания

тоже очень близко к нулю. А это означает, что пакеты почти никогда не ожидают обслуживания в буфере (в момент их прихода он оказывается пустым), а сразу передаются на выход. И наоборот, если ρ приближается к 1, то время ожидания растет очень быстро и нелинейно (и в пределе равно бесконечности). Такое поведение очереди интуитивно понятно, ведь чем ближе средние значения интервалов между пакетами (λ) к среднему времени их обслуживания (μ), тем сложнее обслуживающему устройству (интерфейсу) справляться с нагрузкой.

Из графика, представленного на рис. 4.9, следует резкое ухудшение качества обслуживания при достижении коэффициента использования пропускной способности интерфейсов сети некоторого порогового значения.

Также из приведенного графика не очень понятна причина существования очередей при значениях ρ в окрестностях 0,5. Интенсивность обработки трафика вдвое превышает интенсивность нагрузки, а очереди существуют!

Такой парадоксальный на первый взгляд результат характерен для систем, в которых протекают случайные процессы. Так как во внимание принимаются средние значения интенсивностей потоков на больших промежутках времени, то на небольших промежутках времени они могут существенно отклоняться от этих значений. Очередь создается на тех промежутках, на которых интенсивность поступления пакетов намного превосходит интенсивность обслуживания.

Перегрузка ресурсов может привести к полной деградации сети, когда, несмотря на то, что сеть передает пакеты, полезная скорость передачи данных оказывается равной нулю. Эта ситуация имеет место, если задержки доставки всех пакетов превосходят некоторый порог и пакеты по тайм-ауту отбрасываются узлом назначения как устаревшие. Если же протоколы используют процедуры квитирования и повторной передачи утерянных пакетов, то перегрузка будет нарастать лавинообразно [2].

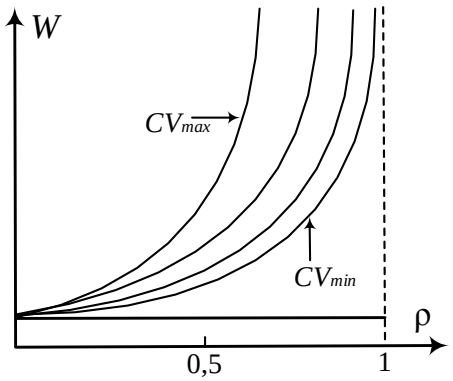


Рис. 4.10. Влияние степени пульсации потока на задержки

Существует еще один важный параметр, оказывающий непосредственное влияние на образование очередей в сетях с КП. Это вариация интервалов входного потока пакетов, т. е. пульсация входного трафика. На рис. 4.10 показано семейство зависимостей *W* от ρ, полученных для разных значений коэффициента вариации *CV* входного потока пакетов. Из рис. 4.10 видно, что чем меньше пульсирует входной поток (*CV* приближается к нулю), тем меньше проявляется эффект лавинообразного образования оче-

реди при приближении коэффициента загрузки ресурса к 1. И наоборот, чем больше *CV*, тем раньше (при меньших значениях ρ) начинает проявляться этот эффект.

Из поведения графиков на рис. 4.10 можно сделать два вывода: во-первых, для оценки значений задержек в очередях на коммутаторах сети недостаточно информации о коэффициенте загрузки ρ, необходимо также знать параметры пульсации трафика. Во-вторых, для снижения уровня задержек нужно снижать значение ρ и уменьшать пульсацию трафика.

Очереди и различные классы трафика. Посмотрим, как можно применить знания о зависимости поведения очередей от коэффициента загрузки для реализации основной идеи методов QoS, а именно – дифференцированного обслуживания классов трафика с различными требованиями к характеристикам производительности и надежности сети. Будем пока делить все потоки на два класса – чувствительный к задержкам и эластичный, допускающий большие задержки, но чувствительный к потерям данных.

Мы знаем, что если обеспечить для чувствительного к задержкам трафика коэффициент загрузки каждого ресурса не более 0,2, то задержки в каждой очереди будут небольшими и, скорее всего, приемлемыми для приложений этого класса. Для эластичного трафика, слабо чувствительного к задержкам, можно допустить более высокий коэффициент загрузки, но не более 0,9. Чтобы пакеты этого класса не терялись, нужно предусмотреть для них буферную память, достаточную для хранения всех пакетов периода пульсации. Эффект от такого распределения загрузки иллюстрирует рис. 4.11.

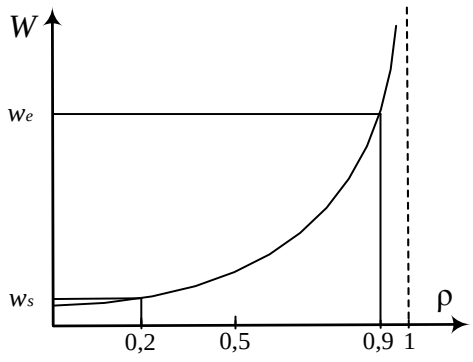


Рис. 4.11. Обслуживание эластичного и чувствительного к задержкам трафика

Задержки чувствительного к задержкам трафика равны w_s, а задержки эластичного трафика – w_e. Можно ввести более чем два класса обслуживания и стараться, чтобы каждый класс работал на своей части кривой задержек.

Чтобы добиться различных коэффициентов использования ресурсов для разных классов трафика, нужно в каждом коммутаторе для каждого ресурса поддерживать две разные очереди. Алгоритм выборки пакетов из очередей должен отдавать предпочтение очереди чувствительных к задержкам пакетов. Если бы все пакеты первой очереди обслуживались приоритетно, а пакеты второй очереди – только тогда, когда первая очередь пуста, то для трафика первой очереди трафик второй очереди фактически перестал бы существовать. Поэтому если отношение средней интенсивности приоритетного трафика λ_1 к производительности ресурса μ равно 0,2,

то и коэффициент загрузки для него равен 0,2. А вот для эластичного трафика, пакеты которого всегда ждут обслуживания приоритетных пакетов, коэффициент загрузки подсчитывается по-другому. Если средняя интенсивность эластичного трафика равна λ_2, то для него ресурс будет загружен на $(\lambda_1 + \lambda_2)/\mu$. Так что если мы хотим, чтобы для эластичного трафика коэффициент загрузки составлял 0,9, то его интенсивность должна вычисляться из соотношения $\lambda_2/\mu = 0{,}7$.

Идея, лежащая в основе всех методов поддержания характеристик *QoS*, заключается в следующем: общая производительность каждого ресурса должна делиться между разными классами трафика неравномерно.

Техника управления очередями нужна для работы в периоды перегрузок, когда сетевое устройство не справляется с передачей пакетов на выходной интерфейс в том темпе, в котором они поступают. Если причиной перегрузки является недостаточная производительность процессорного блока сетевого устройства, то необработанные пакеты временно накапливаются во входной очереди соответствующего входного интерфейса. В том же случае, когда причина перегрузки заключается в ограниченной пропускной способности выходного интерфейса, пакеты временно сохраняются в выходной очереди (или очередях) этого интерфейса.

В **очереди FIFO** в случае перегрузки все пакеты помещаются в общую очередь и выбираются из нее в том порядке, в котором поступили. Во всех устройствах с коммутацией пакетов алгоритм FIFO используется по умолчанию. Достоинствами этого подхода являются простота реализации и отсутствие потребности в конфигурировании. Однако ему присущ и коренной недостаток – *невозможность дифференцированной обработки пакетов различных потоков*. Все пакеты стоят в общей очереди на равных основаниях. Вместе оказываются как пакеты чувствительного к задержкам голосового трафика, так и нечувствительного к задержкам, но очень интенсивного трафика резервного копирования, длительные пульсации которого могут надолго задержать голосовой пакет.

Очереди с приоритетным обслуживанием очень популярны во многих областях вычислительной техники, в частности в ОС, когда одним приложениям нужно отдать предпочтение перед другими при обработке их в мультипрограммной смеси. Применяются эти очереди и для преимущественной по сравнению с другими обработки одного класса трафика.

Механизм приоритетного обслуживания основан на разделении всего сетевого трафика на небольшое количество классов и последующем назначении каждому классу некоторого числового признака – *приоритета*.

Классификация трафика представляет собой отдельную задачу. Пакеты могут разбиваться на приоритетные классы на основании различных признаков: адреса назначения, адреса источника, идентификатора приложения,

генерирующего этот трафик, любых других комбинаций признаков, которые содержатся в заголовках пакетов. Правила классификации пакетов представляют собой часть политики администрирования сети.

Точка классификации трафика может размещаться в каждом коммуникационном устройстве. Более масштабируемое решение – размещение механизмов классификации трафика в одном или нескольких устройствах, расположенных на границе сети (например, в коммутаторах корпоративной сети, к которым подключаются компьютеры пользователей, или во входных маршрутизаторах сети поставщика услуг). В этом случае необходимо специальное поле в пакете, в котором можно запомнить назначенное значение приоритета, чтобы им могли воспользоваться остальные сетевые устройства, обрабатывающие трафик после классифицирующего устройства. Такое поле имеется в заголовке многих протоколов.

В сетевом устройстве, поддерживающем приоритетное обслуживание, имеется *несколько* очередей (буферов) – по одной для каждого приоритетного класса. Пакет, поступивший в период перегрузок, помещается в очередь, соответствующую его приоритетному классу. На рис. 4.12 приведен пример использования четырех приоритетных очередей с высоким, средним, нормальным и низким приоритетами. До тех пор, пока из более приоритетной очереди не будут выбраны все имеющиеся в ней пакеты, устройство не переходит к обработке следующей, менее приоритетной очереди. Пакеты с низким приоритетом обрабатываются только тогда, когда пустеют все вышестоящие очереди: с высоким, средним и нормальным приоритетами [2].

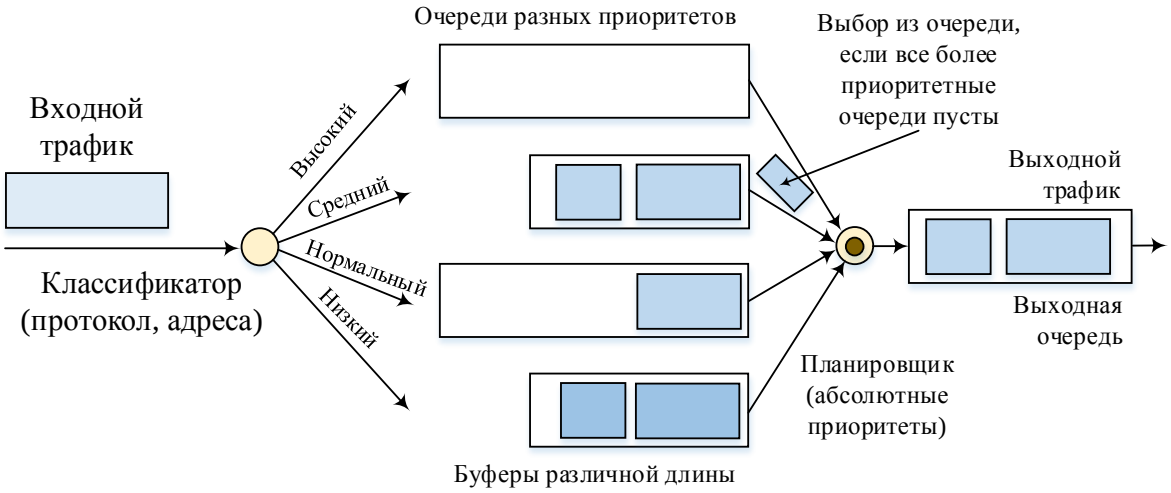


Рис. 4.12. Приоритетные очереди

Размер буфера сетевого устройства определяет максимальную длину очереди ожидающих обслуживания пакетов; если пакет поступает при заполненном буфере, то он просто отбрасывается. Размер буфера определяется в идеальном случае таким образом, чтобы его хватало с некоторым запасом

для хранения очереди среднестатистической длины. Однако установить это значение достаточно сложно, так как оно изменяется в зависимости от нагрузки сети, поэтому требуется постоянное и длительное наблюдение за работой сети. В общем случае, чем выше значимость трафика для предприятия, чем больше его интенсивность и пульсации, тем больший размер буфера требуется этому трафику. В примере, приведенном на рис. 4.12, для трафика высшего и нормального приоритетов выбраны большие размеры буферов, а для остальных двух классов – меньшие. Мотивы принятия решения для высшего приоритета очевидны, а трафик нормального приоритета имеет высокую интенсивность и значительный коэффициент пульсаций.

Приоритетное обслуживание очередей обеспечивает высокое качество обслуживания для пакетов из самой приоритетной очереди. Если средняя интенсивность их поступления в устройство не превосходит пропускной способности выходного интерфейса, то пакеты высшего приоритета всегда получают ту пропускную способность, которая им нужна. Уровень задержек высокоприоритетных пакетов также минимален. Однако он не нулевой и зависит в основном от характеристик потока этих пакетов – чем выше пульсации потока и его интенсивность, тем вероятнее возникновения очереди, образованной пакетами данного высокоприоритетного потока. Трафик всех остальных приоритетных классов почти прозрачен для пакетов высшего приоритета. Слово «почти» относится к ситуации, когда высокоприоритетный пакет вынужден ждать завершения обслуживания низкоприоритетного пакета, если его приход совпадает по времени с началом продвижения низкоприоритетного пакета на выходной интерфейс. Этот эффект иллюстрирует рис. 4.13, на котором показано, что после разделения всего трафика на приоритетный и обычный (здесь имеются две очереди) коэффициент использования приоритетного трафика снизился с 50 до 15 %, так как нагрузка обычного трафика перестала влиять на использование выходного интерфейса приоритетным трафиком [2].

Что же касается остальных приоритетных классов, то качество их обслуживания будет ниже, чем у пакетов самого высокого приоритета. Если коэффициент нагрузки выходного интерфейса, определяемый только трафиком высшего приоритетного класса, приближается к единице, то трафик остальных классов на это время просто замораживается. Поэтому приоритетное обслуживание обычно применяется для чувствительного к задержкам класса трафика, имеющего небольшую интенсивность. Например, голосовой трафик чувствителен к задержкам, но его интенсивность обычно не превышает 8–16 Кбит/с, так что при назначении ему высшего приоритета ущерб остальным классам трафика оказывается не очень значительным.

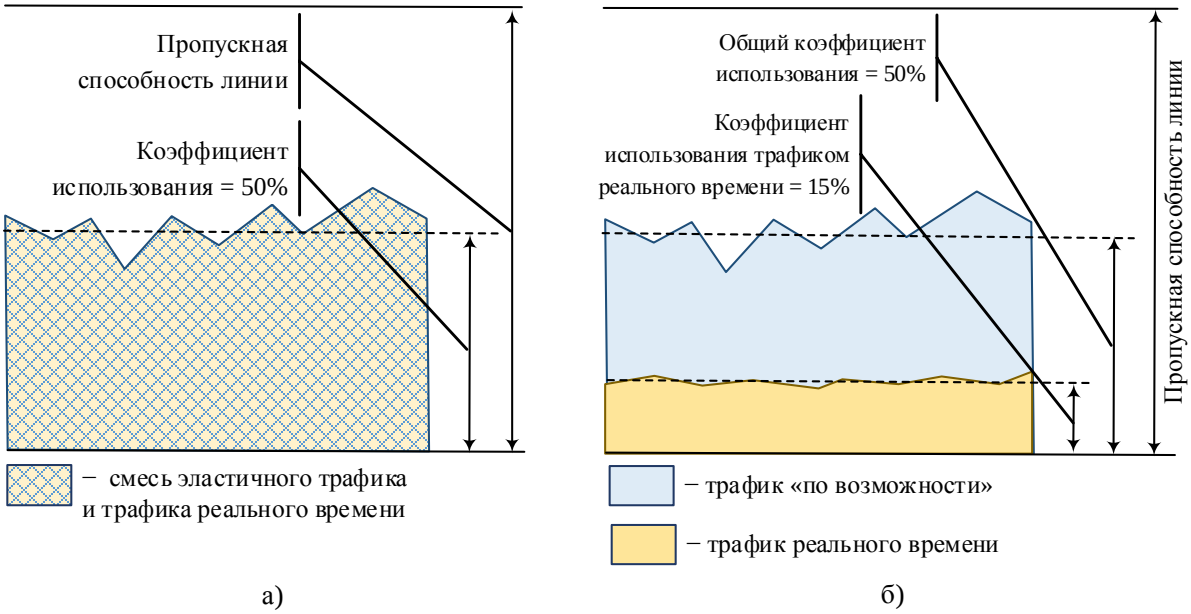


Рис. 4.13. Снижение коэффициента использования линии для приоритетного трафика:
а) весь трафик обслуживается одной очередью;
б) трафик реального времени обслуживается приоритетной очередью,
а остальной трафик – очередью по умолчанию

Механизм взвешенных очередей разработан для того, чтобы можно было предоставить всем классам трафика определенный минимум пропускной способности. Под *весом* данного класса понимается процент предоставляемой классу трафика пропускной способности от полной пропускной способности выходного интерфейса. При взвешенном обслуживании, так же как и при приоритетном, трафик делится на несколько классов и для каждого класса ведется отдельная очередь пакетов. Но с каждой очередью связывается *не приоритет, а процент пропускной способности* ресурса, гарантируемый данному классу трафика при перегрузках этого ресурса. Например, на рис. 4.14 устройство поддерживает для пяти классов трафика пять очередей к выходному интерфейсу коммутатора. Этим очередям при перегрузках выделяется соответственно 10, 10, 30, 20 и 30 % пропускной способности выходного интерфейса [2].

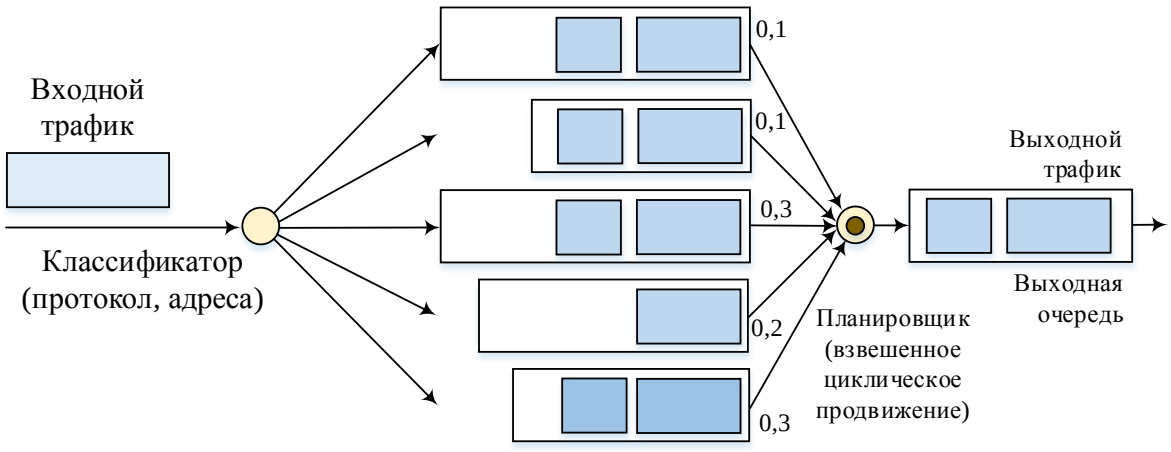


Рис. 4.14. Взвешенные очереди

Достигается поставленная цель за счет того, что очереди обслуживаются последовательно и циклически и в каждом цикле обслуживания из каждой очереди выбирается такое число байтов, которое соответствует весу данной очереди. Так, если цикл просмотра очередей в рассматриваемом примере равен одной секунде, а скорость выходного интерфейса составляет 100 Мбит/с, то при перегрузках в каждом цикле первой очереди уделяется 10 % времени, т. е. 100 мс, и выбирается 10 Мбит данных, из второй – тоже 10 Мбит, из третьей – 30 Мбит, из четвертой – 20 Мбит, из пятой – 30 Мбит. В результате каждому классу трафика достается гарантированный минимум пропускной способности, что является более желательным результатом, чем подавление низкоприоритетных классов высокоприоритетными.

Так как данные выбираются из очереди пакетами, а не битами, то реальное распределение пропускной способности между классами трафика отличается от планируемого. Так, в примере вместо 10 % первый класс трафика мог бы получать при перегрузках 9 или 12 %. Чем больше время цикла, тем точнее соблюдаются требуемые пропорции между классами трафика, так как из каждой очереди выбирается большее число пакетов и влияние размера каждого пакета усредняется. В то же время длительный цикл приводит к большим задержкам передачи пакетов. Так, при выбранном цикле в одну секунду задержка может составить одну и более секунд – ведь арбитр возвращается к каждой очереди не чаще чем раз в секунду. Поэтому при выборе времени цикла нужно обеспечить баланс между точностью соблюдения пропорций пропускной способности и стремлением к снижению задержки.

Для нашего примера более сбалансированным выглядит время цикла в 1000 мкс. С одной стороны, такое время гарантирует более низкий уровень задержек, так как очереди просматриваются намного чаще, чем при секундном цикле. С другой стороны, этого времени достаточно, чтобы выбрать из каждой очереди в среднем по несколько пакетов (первой очереди в нашем примере будет отводиться 100 мкс, что достаточно, например, для передачи в выходной канал одного пакета Fast Ethernet или десяти пакетов GBE).

На уровень задержек и вариации задержек пакетов для некоторого класса трафика при взвешенном обслуживании в значительной степени влияет *относительный коэффициент использования*. В этом случае коэффициент подсчитывается как отношение интенсивности входного трафика класса к пропускной способности, выделенной этому классу в соответствии с его весом. Например, если мы выделили первой очереди 10 % от общей пропускной способности выходного интерфейса, т. е. 10 Мбит/с, а средняя интенсивность потока, который попадает в эту очередь, равна 3 Мбит/с,

то коэффициент использования для этого потока составит 3/10 = 0,3. Качественное поведение очереди и соответственно задержек здесь выглядит примерно так же, как и в случае очереди FIFO – чем меньше коэффициент загрузки, тем меньше средняя длина очереди и тем меньше задержки.

Еще одним вариантом взвешенного обслуживания является *взвешенное справедливое обслуживание* (Weighted Fair Queuing, WFQ). В случае подобного обслуживания пропускная способность ресурса делится между всеми потоками поровну, т. е. «справедливо».

Взвешенное обслуживание обеспечивает требуемые соотношения между интенсивностями трафика различных очередей только в *периоды перегрузок*, когда каждая очередь постоянно заполнена. Если же какая-нибудь из очередей пуста (т. е. для трафика данного класса текущий период не является периодом перегрузки), то при просмотре очередей она игнорируется, а ее время обслуживания распределяется между остальными очередями в соответствии с их весом. Поэтому в отдельные периоды трафик определенного класса может обладать большей интенсивностью, чем соответствующий процент от пропускной способности выходного интерфейса.

Каждый из описанных подходов имеет свои достоинства и недостатки. Приоритетное обслуживание, обеспечивая минимальный уровень задержек для очереди наивысшего приоритета, не дает никаких гарантий в отношении средней пропускной способности для трафика очередей более низких приоритетов. Взвешенное обслуживание обеспечивает заданное распределение средней пропускной способности, но не учитывает требований к задержкам.

Существуют **комбинированные алгоритмы обслуживания очередей.** В наиболее популярном алгоритме подобного рода поддерживается одна приоритетная очередь, а остальные очереди обслуживаются в соответствии со взвешенным алгоритмом. Обычно приоритетная очередь используется для чувствительного к задержкам трафика, а остальные – для эластичного трафика нескольких классов. Каждый класс эластичного трафика получает некоторый минимум пропускной способности при перегрузках. Этот минимум вычисляется как процент от пропускной способности, оставшейся от приоритетного трафика. Очевидно, что нужно как-то ограничить приоритетный трафик, чтобы он не поглощал всю пропускную способность ресурса. Обычно для этого применяются механизмы кондиционирования трафика.

Выводы

Методы обеспечения качества обслуживания занимают сегодня важное место в семействе технологий сетей с коммутацией пакетов, так как без их применения сложно обеспечить качественную работу современных

мультимедийных приложений, таких как IP-телефония, видео- и радиовещание, интерактивное дистанционное обучение и т. п.

Характеристики QoS отражают отрицательные последствия пребывания пакетов в очередях, которые проявляются в снижении скорости передачи, задержках пакетов и их потерях. Существуют различные типы трафика, отличающиеся чувствительностью к задержкам и потерям пакетов. Наиболее грубая классификация трафика разделяет его на два класса: трафик реального времени (чувствительный к задержкам) и эластичный трафик (нечувствительный к задержкам в широких пределах).

Методы QoS основаны на перераспределении имеющейся пропускной способности линий связи между трафиком различного типа в соответствии с требованиями приложений. Приоритетные и взвешенные очереди являются основным инструментом выделения пропускной способности определенным потокам пакетов.

Контрольные вопросы

1. Возникают ли очереди в сетях с коммутацией каналов?
2. В чем состоят позитивные и негативные эффекты применения очередей в коммутаторах пакетов?
3. Какие параметры влияют на размер очереди?
4. Какие типы трафика передает сеть с коммутацией пакетов? Какие требования эти типы трафика предъявляют к сети?
5. К каким нежелательным последствиям может привести приоритетное обслуживание?
6. Может ли отсутствовать очередь в системе, коэффициент использования которой близок к единице?
7. Объясните причину возможного возникновения очередей даже при невысокой загрузке коммутаторов или маршрутизаторов сети с коммутацией пакетов.
8. Для трафика какого типа наиболее подходит взвешенное обслуживание?
9. Можно ли комбинировать приоритетное и взвешенное обслуживание?
10. Какой тип обслуживания целесообразно применить, если нужно обеспечить различную минимальную гарантированную способность трем классам трафика?

4.3. Методы обеспечения качества обслуживания

4.3.1. Методы кондиционирования трафика

Основной идеей методов QoS является выделение определенной доли пропускной способности определенным потокам трафика, при этом величина полученной потоком доли должна быть достаточной для того, чтобы качество обслуживания потока было удовлетворительным. Очереди с различными алгоритмами обслуживания позволяют реализовать только одну часть этой идеи – они выделяют определенную долю пропускной способности некоторому потоку пакетов. Вторая же часть задачи – обеспечение

требуемого качества обслуживания потока – решается ограничением его скорости. Скорость, а также связанный с ней относительный коэффициент использования пропускной способности не должны превышать значений, предельных для поддержания требуемого качества обслуживания. Эту задачу решают *механизмы кондиционирования трафика*, включающие классификацию, профилирование и формирование трафика.

Мы уже имели дело с *классификацией трафика*, когда при изучении приоритетных и взвешенных очередей предполагали наличие некоего механизма, решающего, какие пакеты нужно отправить в ту или иную очередь. Такого рода классификация обычно выполняется в коммутаторах и маршрутизаторах пакетных сетей по различным признакам пакетов, например, адреса назначения и источника, типа протокола транспортного уровня.

Профилирование (policing) представляет собой меру принудительного воздействия на трафик, направленную на ограничение скорости потока пакетов. Профилирование обеспечивает соответствие потока пакетов заданному скоростному *профилю* – набору заданных параметров потока. В качестве основного параметра обычно выступает средняя скорость потока пакетов, измеренная на определенном интервале времени. Пакеты, которые не укладываются в заданный профиль, либо отбрасываются, либо помещаются в класс обслуживания с более низким приоритетом.

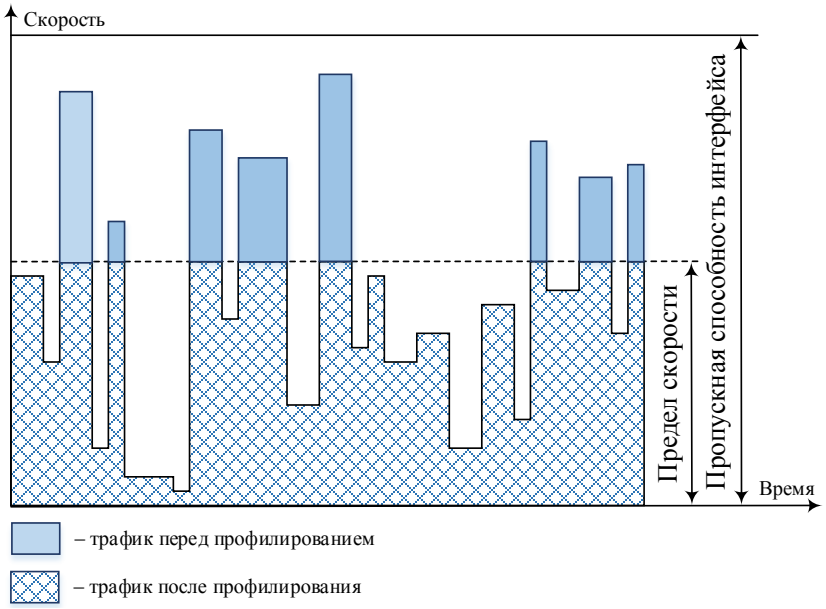


Рис. 4.15. Эффект профилирования – отбрасывание избыточного трафика

Профилирование чаще всего применяют для ограничения трафика, поступающего в приоритетную очередь, так как этот механизм является единственно возможным средством предотвращения вытеснения всего остального трафика приоритетным трафиком. Рис. 4.15 иллюстрирует действие механизма профилирования, показывая значения скорости трафика, изме-

ренные на достаточно малых интервалах времени до и после профилирования. Как видно из рис. 4.15, отбрасывание пакетов при профилировании приводит к удержанию скорости потока на заданном уровне в те интервалы времени, когда скорость входящего потока превосходит этот предел, и к сохранению исходной скорости в остальные периоды.

Формирование трафика (*shaping*) в каком-то смысле подобно профилированию, так как имеет схожую цель – ограничение скорости трафика, или более точно – приведение параметров потока к заданному профилю. Однако достигается эта цель другим способом. Вместо того чтобы отбрасывать избыточные пакеты, т. е. те, передача которых могла бы привести к превышению лимита скорости, механизм формирования трафика задерживает пакеты-нарушители так, что результирующая скорость оказывается в заданных пределах. Эффект формирования трафика иллюстрирует рис. 4.16 [2].

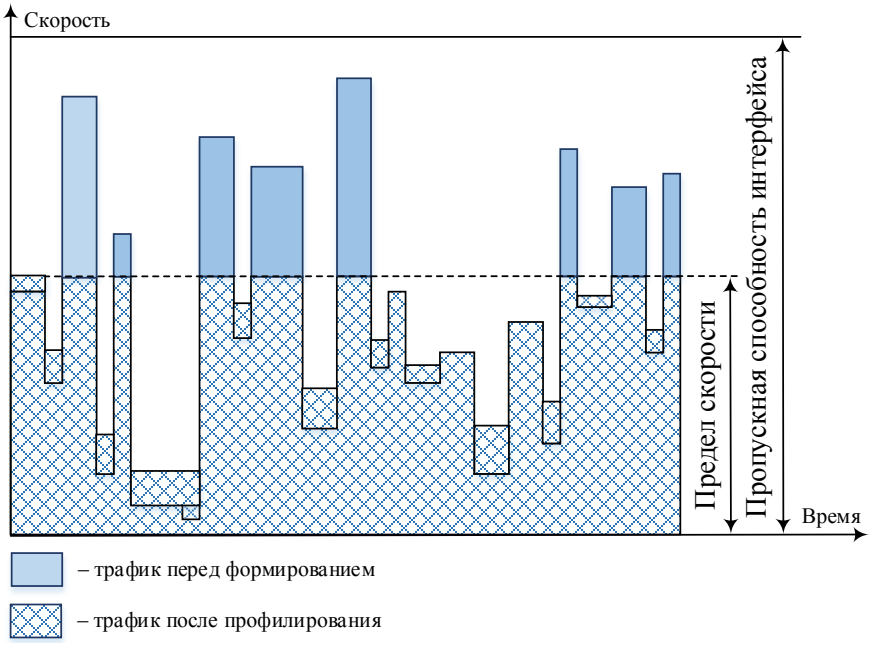


Рис. 4.16. Эффект формирования трафика – сглаживание

График скорости трафика сглаживается за счет «срезания» выступов (задержки пакетов, выходящих за уровень предельной скорости) и заполнения впадин (перемещения их в другие интервалы времени, в которых скорость оказывается меньше установленного предела).

Обычно формирование трафика применяется к потокам, исходящим от коммутатора или маршрутизатора, когда известно, что далее по маршруту следования потока некоторое коммуникационное устройство применяет профилирование. Если профиль, задаваемый для формирования трафика, совпадает с профилем последующего профилирования, то это гарантирует отсутствие потерь трафика из-за отбрасывания избыточных пакетов.

Механизмы кондиционирования трафика могут поддерживаться каждым узлом сети, либо реализовываться только в пограничных устройствах. Последнее используют поставщики услуг, кондиционируя трафик клиентов.

4.3.2. Методы обратной связи

Алгоритмы управления очередями и кондиционирования трафика не предотвращают перегрузок, а лишь перераспределяют ресурсы между различными потоками или классами трафика. Алгоритмы управления очередями относятся к механизмам *управления перегрузкой* (*congestion management*), которые начинают работать, когда сеть уже перегружена. Существует другой класс средств, которые носят название механизмов *предотвращения перегрузок* (congestion avoidance). Этот механизм основан на использовании обратной связи, с помощью которой перегруженный узел сети просит предыдущие узлы, расположенные вдоль маршрута следования потока, *временно* снизить скорость трафика. После того как перегрузка в данном узле исчезнет, можно разрешить повысить скорость передачи данных.

Существует несколько механизмов обратной связи (рис. 4.17). Они отличаются информацией, которая передается по обратной связи, а также тем, кто генерирует информацию, и кто реагирует на эту информацию – конечный узел (компьютер) или промежуточный (коммутатор/маршрутизатор).

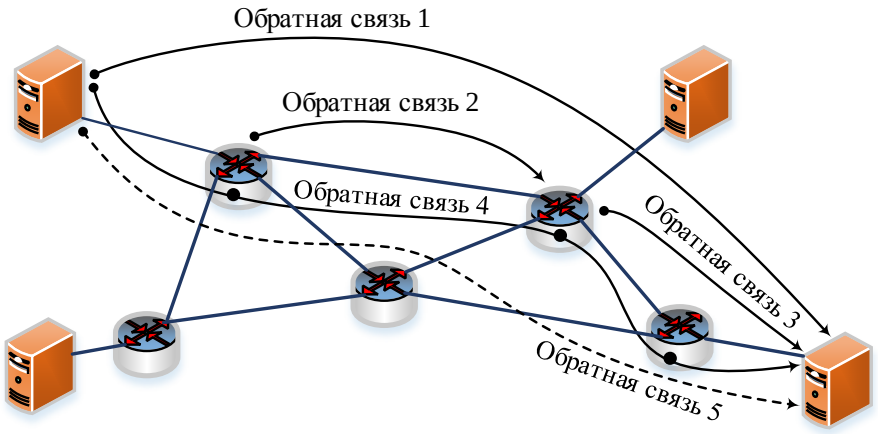


Рис. 4.17. Участники обратной связи

Обратная связь 1 организована между двумя конечными узлами сети. Этот вариант обеспечивает наиболее радикальное снижение нагрузки на сеть, так как только конечный узел может снизить скорость поступления информации в сеть. Назначение этого вида обратной связи – борьба с перегрузками узла назначения, а не с перегрузками промежуточных сетевых устройств, поэтому за ним закрепилось собственное название – *контроль потока*. Устройства сети не принимают участие в работе этой обратной связи, а только передают соответствующие сообщения между конечными узлами.

При организации обратной связи важно учитывать влияние, которое вносит задержка передачи информации по сети. Так, в высокоскоростных глобальных сетях за время, которое тратится на передачу сообщения о перегрузке узла назначения, узел-источник может успеть направить в сеть тысячи пакетов, так что перегрузка не будет ликвидирована вовремя. Также возможны ситуации, когда узел-источник начинает снижать скорость передачи информации, хотя в действительности очереди в узле-получателе уже нет, и наоборот, повышать скорость передачи информации в тот момент, когда узел-получатель начал испытывать перегрузку. Для борьбы с такими явлениями в контур обратной связи вводится интегрирующий элемент, который на каждом шаге обрабатывает не только текущее сообщение обратной связи, но и несколько предыдущих, что позволяет учесть динамику.

Обратная связь 2 организована между двумя соседними коммутаторами. Коммутатор сообщает соседу, находящемуся выше по течению потока, что он испытывает перегрузку и его буфер заполнился до критической величины. Получив такое сообщение, сосед, расположенный выше по течению, должен снизить на некоторое время скорость передачи в направлении перегруженного коммутатора и тем самым решить проблему перегрузки.

Обратная связь 3 организована между некоторым промежуточным коммутатором и узлом-источником; все остальные промежуточные коммутаторы, лежащие между этими двумя узлами, только передают сообщения обратной связи в направлении к узлу-источнику, никак на них не реагируя.

В *обратной связи 4*, как и в обратной связи 1, сообщение о перегрузке порождается узлом получателем и передается узлу-источнику. Однако теперь каждый промежуточный коммутатор реагирует на это сообщение. Во-первых, он снижает скорость передачи в направлении узла назначения, во-вторых, он может изменить содержание сообщения. Например, если узел назначения просит снизить скорость до 300 Мбит/с, то промежуточный коммутатор может снизить ее до 200 Мбит/с, оценив состояние своего буфера.

При описании обратной связи мы полагали, что сообщение о перегрузке идет в направлении, обратном направлению передачи пользовательской информации. Однако некоторые коммуникационные протоколы не предусматривают возможности генерации подобных сообщений промежуточными узлами. В таких условиях применяют искусственный прием – передачу сообщения о перегрузке узлу назначения, который преобразует его в сообщение обратной связи и отправляет в направлении источника (*обратная связь 5*).

В применяемых сегодня методах обратной связи используются следующие основные типы сообщений: а) признак перегрузки; б) максимальная скорость передачи; в) максимальный объем данных; г) косвенные признаки.

Признак перегрузки не говорит о степени перегруженности сети или узла, он только фиксирует факт наличия перегрузки. Реакция узла, получившего такое сообщение, может быть разной. В некоторых протоколах узел обязан прекратить передачу информации в определенном направлении до тех пор, пока не будет получено другое сообщение обратной связи, разрешающее продолжение передачи. В других протоколах узел ведет себя адаптивно, он снижает скорость на некоторую величину и ожидает реакции сети.

Во втором типе сообщений указывается *максимальная скорость передачи*, т. е. порог скорости, который должен соблюдать источник или промежуточный узел, расположенный выше по течению потока. В этом случае обязательно нужно учитывать время передачи сообщения по сети, чтобы исключить колебательные процессы в сети и обеспечить нужную скорость реакции на перегрузку. Поэтому в территориальных сетях такой способ обычно реализуется силами всех коммутаторов сети (обратная связь 4).

Сообщение о *максимальном объеме данных* используется в широко применяемом в пакетных сетях алгоритме скользящего окна. Параметром, несущим информацию обратной связи, является «окно» – число, связанное с текущим размером свободного пространства в буфере принимающего узла. Передающий узел может с любой скоростью передать объем информации, равный определенному для него окну. Но если этот лимит исчерпан, то передающий узел не имеет права передавать информацию, пока не получит следующее окно. При перегрузках принимающий узел уменьшает размер окна, тем самым снижая нагрузку. Если эффект перегрузки исчезает, то принимающий узел увеличивает размер окна.

В некоторых случаях передающий узел определяет, что принимающий узел испытывает перегрузку, по некоторым косвенным признакам, без обратной связи. Такими косвенными признаками могут быть факты потери пакетов. Примером протокола, использующего неявную информацию о перегрузках, является протокол TCP. Этот протокол с помощью явной информации обратной связи (о размере окна) осуществляет контроль потока, а с помощью неявной (потери пакетов) управляет перегрузкой.

4.3.3. Методы резервирования ресурсов

Рассмотренные методы поддержания QoS ориентированы в основном на борьбу с перегрузками или предотвращение их в пределах отдельного узла сети. Вместе с тем понятно, что для поддержания гарантированного уровня QoS некоторого потока пакетов необходимо скоординированное применение этих методов на всем пути следования потока через сеть.

Резервирование ресурсов – это координирующая процедура, которая настраивает все механизмы поддержания качества обслуживания вдоль следования потока таким образом, чтобы поток с некоторыми заданными характеристиками скорости был обслужен с заданными характеристиками QoS.

Основная идея процедуры резервирования ресурсов состоит в следующем. *Перед тем как* реальный поток будет направлен в сеть, каждому узлу сети вдоль маршрута его следования задается вопрос, может ли этот узел обслужить некоторый новый поток с заданными характеристиками QoS, если известны предельные характеристики скорости потока, такие как средняя и пиковая скорости? Каждый узел при ответе на этот вопрос должен оценить свои возможности, т. е. проверить, достаточно ли у него свободных ресурсов, чтобы принять на обслуживание новый поток и обслужить его качественно. При положительном ответе узел должен некоторым образом зарезервировать часть своих ресурсов для данного потока, чтобы при поступлении пакетов потока на входные интерфейсы использовать эти ресурсы для обслуживания поступающих пакетов с гарантированным уровнем качества.

В общем случае каждый узел самостоятельно решает, какие ресурсы он должен зарезервировать для обслуживания некоторого потока. Основным ресурсом для качественного обслуживания пакетов является пропускная способность интерфейса, через который пакеты потока покидают узел. Поэтому в дальнейшем будем употреблять формулировку «резервирование пропускной способности» вместо «резервирование ресурсов».

Концепция резервирования рассматривалась в сетях с коммутацией каналов. Для сетей с коммутацией пакетов механизм резервирования пропускной способности не является принципиально необходимым, он имеет вспомогательное значение и используется только в тех случаях, когда требуется гарантированное обеспечение заданного качества обслуживания пакетов. Процедура резервирования подобна аналогичной процедуре в сетях с коммутацией каналов: определенному потоку данных назначается определенная часть пропускной способности линии связи. Однако в сетях с коммутацией пакетов эта процедура является более гибкой, а именно – если отведенная пропускная способность в какой-то период времени недоиспользуется потоком, то она может быть передана другим потокам. Еще одним отличием резервирования в пакетных сетях является то обстоятельство, что резервирование может выполняться не только «из конца в конец», но и для каких-то отдельных узлов по маршруту потока.

Контроль допуска. Резервирование пропускной способности в пакетной сети «из конца в конец» начинается с операции, называемой *контролем допуска в сеть* (admission control) потока, который просит зарезервировать для своего обслуживания некоторую пропускную способность сети между ее двумя конечными узлами. Эта операция состоит в проверке наличия доступной пропускной способности на каждом из узлов сети на протяжении всего маршрута следования потока. Максимальная средняя скорость потока должна быть меньше, чем запрашиваемая пропускная способность,

иначе поток будет обслужен с плохим качеством, даже несмотря на то, что ему была зарезервирована некоторая пропускная способность.

Если результат контроля допуска положителен в каждом узле (рис. 4.18), то сетевые устройства запоминают факт резервирования, чтобы при появлении пакетов данного потока распознать их и выделить им зарезервированную пропускную способность. Кроме того, при успешном резервировании доступная для резервирования (в будущем) пропускная способность уменьшается на величину, зарезервированную за данным потоком.

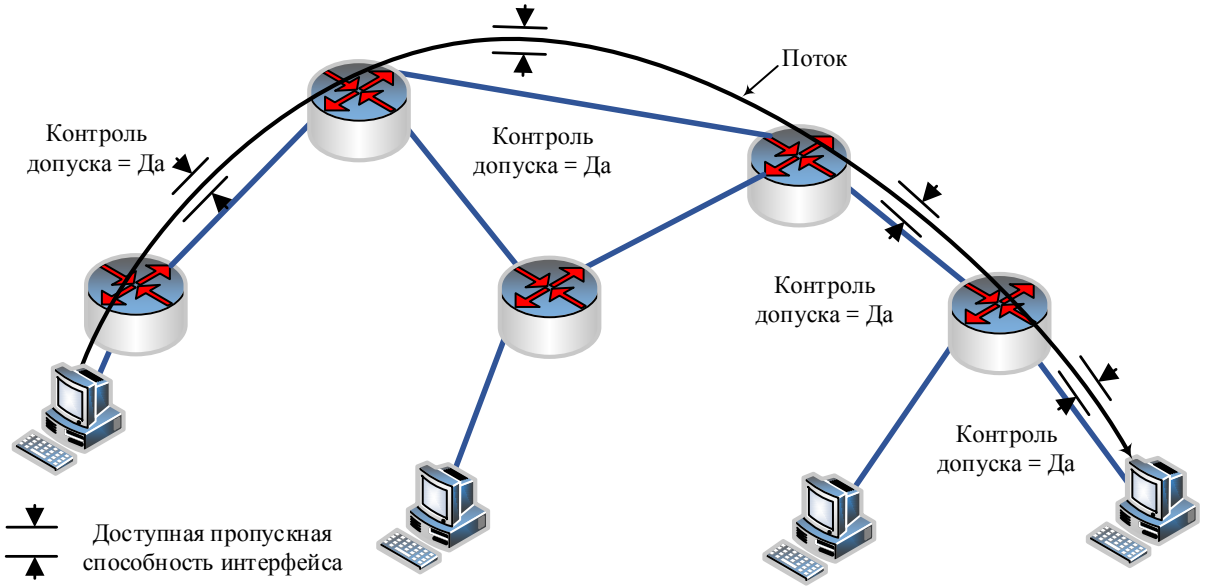


Рис. 4.18. Контроль допуска потока

Рассмотрим выделение пропускной способности потоку в моменты времени, когда его пакеты поступают на вход коммуникационного устройства *S*2, которое запомнило факт резервирования пропускной способности для потока *F*1 на выходном интерфейсе *P*2 (рис. 4.19) [2].

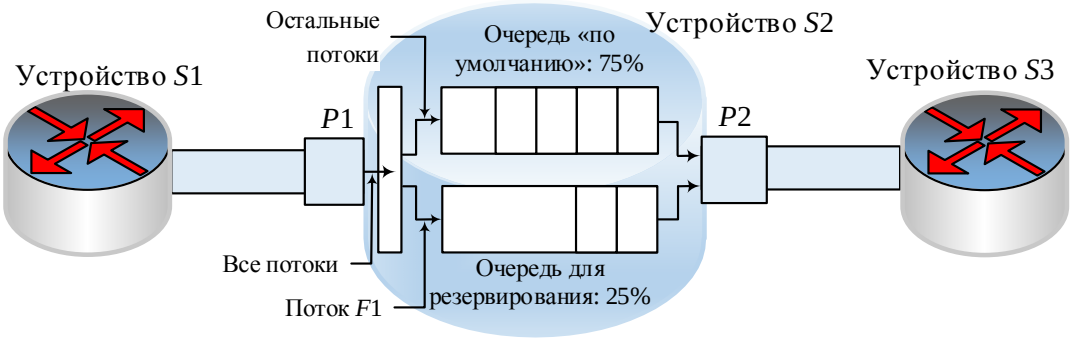


Рис. 4.19. Выделение зарезервированной пропускной способности

Такое выделение можно обеспечить разными способами, в нашем примере это будет сделано с использованием взвешенных очередей. Пусть

потоку *F*1 при резервировании было выделено 25 % пропускной способности интерфейса *P*2. Для простоты будем считать, что резервирование было выполнено только для потока *F*1, а для всех других потоков, которые проходят через выходной интерфейс *P*2, резервирование не производилось.

Для того чтобы добиться желаемого результата, достаточно организовать для выходного интерфейса две взвешенные очереди – очередь для потока *F*1 с весом 25 % и очередь «по умолчанию» для всех остальных потоков. Кроме того, необходимо активизировать *классификатор*, который будет проверять пакеты на всех входных интерфейсах устройства *S*2 (на рис. 4.19 показан только один входной интерфейс *P*1), отбирать пакеты потока *F*1 по заданным при резервировании признакам и направлять их в очередь для потока *F*1. В те периоды времени, когда скорость потока *F*1 окажется меньше зарезервированной пропускной способности в 25 %, неиспользованная ее часть будет потребляться потоками из очереди «по умолчанию» – в силу алгоритма работы взвешенных очередей. Зато в периоды, когда скорость потока *F*1 достигнет заявленного максимума потребления пропускной способности в 25 %, остальные потоки будут довольствоваться оставшимися 75 %.

В описанном примере не задействован механизм профилирования трафика. При наличии отдельной взвешенной очереди для потока, зарезервировавшего пропускную способность, этот механизм не является обязательным, так как сам механизм взвешенных очередей ограничит пропускную способность потока в нужных пределах в периоды перегрузок, когда все взвешенные очереди заполняются полностью.

Использование взвешенных очередей – не единственный вариант резервирования пропускной способности в пакетных сетях. Для той же цели можно задействовать приоритетные очереди. Применение приоритетной очереди может быть необходимым, если потоку помимо пропускной способности требуется обеспечить минимально возможный уровень задержек.

При использовании приоритетной очереди профилирование необходимо всегда, так как приоритетный механизм не обеспечивает ограничения скорости потока, как это делает механизм взвешенного обслуживания.

Резервирование приводит к ожидаемым результатам только тогда, когда реальная скорость потоков, для которых было выполнено резервирование, оказывается не выше, чем пропускная способность, запрошенная при резервировании. В противном случае результаты могут оказаться даже хуже, чем при наличии единственной очереди «по умолчанию» и обслуживании «по возможности». Так, если скорость потока окажется выше, чем предел, учитываемый механизмом профилирования, то часть пакетов будет отброшена даже в том случае, если устройство не перегружено и могло бы отлично справиться с предложенным трафиком без применения механизмов QoS.

Обеспечение заданного уровня задержек. При описании процедуры резервирования пропускной способности мы сфокусировались на механизмах выделения пропускной способности некоторому потоку и оставили без внимания одну важную деталь: какую пропускную способность должен запрашивать поток, для того чтобы задержки его пакетов не превышали некоторой величины? Единственное соображение заключалось в том, что запрашиваемая пропускная способность должна быть выше, чем максимальная скорость потока, иначе некоторая часть пакетов просто может постоянно отбрасываться сетью. Однако мы не можем, например, сконфигурировать очередь приоритетного или взвешенного обслуживания так, чтобы она строго обеспечила какой-либо заранее заданный порог задержек и их вариации. Направление пакетов в приоритетную очередь только позволяет гарантировать, что задержки будут достаточно низкими – существенно ниже, чем у пакетов, которые обрабатываются в очереди по умолчанию. Мы также знаем, что при наличии взвешенных очередей задержки будут снижаться со снижением относительного коэффициента использования пропускной способности отведенной очереди. Но это все качественные рассуждения, а вот количественно оценить значения задержек очень сложно.

Каким же образом поставщик услуг может выполнить свои обязательства перед клиентами? Он должен постоянно *измерять фактические значения характеристик трафика в сети* и гарантировать пользователям сети величины задержек в соответствии с наблюдаемыми результатами. На практике обеспечить постоянный мониторинг задержек и потерь пакетов в сети оказывается совсем не просто – это требует установки в сети большого количества агентов-измерителей, хорошо синхронизированных друг с другом, а также программной системы регистрации и анализа измерительной информации. Поэтому операторы часто предпочитают давать качественное описание различных классов услуг, говоря, например, о минимальных задержках наивысшего класса обслуживания.

4.3.4. Методы инжиниринга трафика

При рассмотрении резервирования мы не стали затрагивать вопрос выбора маршрутов следования потоков через сеть, а старались обеспечить соблюдение требований QoS в условиях заданности маршрутов. Однако задачу обеспечения требований QoS можно решить более эффективно, если считать, что маршруты следования трафика не фиксированы, а также подлежат выбору. Это позволило бы сети обслуживать больше потоков с гарантиями QoS при тех же характеристиках самой сети, т. е. пропускной способности каналов и производительности коммутаторов и маршрутизаторов.

Задачу выбора маршрутов для потоков (или классов) трафика с учетом соблюдения требований QoS решают методы *инжиниринга трафика* (Traffic Engineering). С помощью этих методов стремятся по возможности максимально и сбалансированно загрузить все ресурсы сети, чтобы сеть при заданном уровне QoS обладала бы максимальной производительностью.

Методы инжиниринга трафика основаны на резервировании ресурсов; они не только позволяют найти рациональный маршрут для потока, но и резервируют пропускную способность ресурсов сети вдоль этого маршрута.

Недостатком традиционных методов маршрутизации является принцип работы протоколов, когда выбор маршрута осуществляется на основе только топологии сети без учета информации о ее текущей загрузке.

Для каждой пары «адрес источника – адрес назначения» такие протоколы выбирают единственный маршрут, не принимая во внимание информационные потоки, протекающие через сеть. В результате все потоки между парами конечных узлов сети идут по *кратчайшему* (в соответствии с некоторой метрикой) маршруту. Выбранный маршрут может быть более рациональным, если учитывается пропускная способность каналов или вносимые ими задержки. Примером неэффективности такого подхода является так называемая «рыба» – сеть с топологией, приведенной на рис. 4.20. Несмотря на то, что между коммутаторами *A* и *E* существуют два пути (верхний – через коммутатор *B*, и нижний – через коммутаторы *C* и *D*), весь трафик от коммутатора *A* к коммутатору *E* в соответствии с традиционными принципами маршрутизации направляется по верхнему пути. Только потому, что нижний путь немного (на один ретрансляционный участок) длиннее, чем верхний, он игнорируется, хотя мог бы работать «параллельно» с верхним.

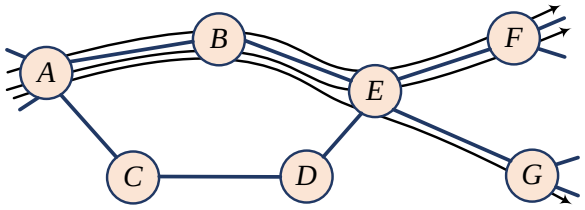


Рис. 4.20. Неэффективность кратчайших путей

Такой подход приводит к тому, что даже если кратчайший путь перегружен, пакеты все равно посылаются по этому пути. Налицо явная ущербность методов распределения ресурсов сети – одни ресурсы работают с перегрузкой, а другие не используются вовсе.

Исходными данными **методов инжиниринга трафика** являются: а) характеристики передающей сети; б) сведения о предложенной нагрузке сети.

К *характеристикам передающей сети* относится ее топология, а также производительность составляющих ее коммутаторов и линий связи. Предполагается, что производительность процессора каждого коммутатора достаточна для обслуживания трафика всех его входных интерфейсов. При таких условиях в качестве резервируемых ресурсов выступает пропускная способность линий связи между коммутаторами (рис. 4.21).

Сведения о предложенной нагрузке сети представляют собой информацию о потоках трафика, которые сеть должна передавать между своими пограничными коммутаторами. Каждый поток характеризуется точкой входа в сеть, точкой выхода из сети и профилем трафика. Для получения оптимальных решений можно использовать детальное описание каждого потока, например, учитывать пульсации трафика. Однако поскольку количественно оценить их влияние на работу сети достаточно сложно, для нахождения субоптимального распределения путей прохождения потоков через сеть, как правило, учитываются их средние скорости передачи данных (рис. 4.22) [2].

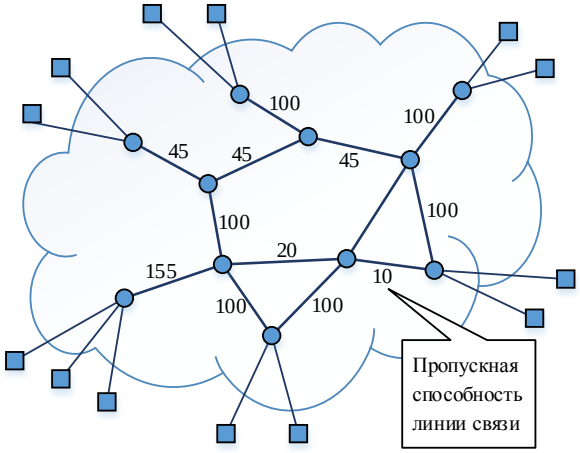


Рис. 4.21. Топология сети и производительность ее ресурсов

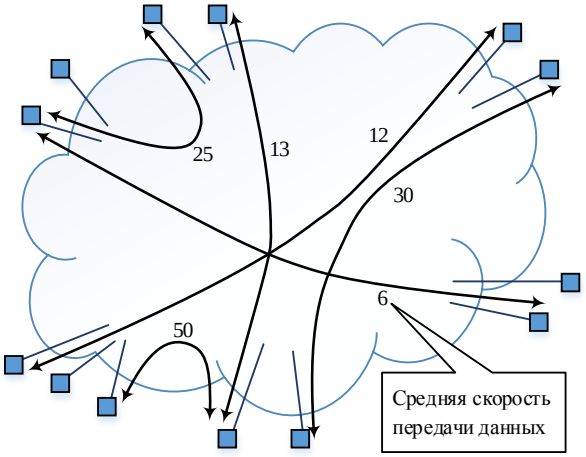


Рис. 4.22. Предложенная нагрузка

Методы инжиниринга трафика чаще применяют не к отдельным, а к *агрегированным* потокам, которые являются объединением нескольких потоков. Так как мы ищем общий маршрут для нескольких потоков, то агрегировать можно только потоки, имеющие общие точки входа в сеть и выхода из сети. Агрегированное задание потоков позволяет упростить задачу выбора путей, поскольку индивидуальных потоков может быть очень много. Необходимо, однако, подчеркнуть, что агрегирование отдельных потоков в один возможно только в том случае, когда все потоки, составляющие агрегированный поток, предъявляют одни и те же требования к качеству обслуживания. Далее в этом разделе мы будем для краткости пользоваться термином «поток» как для индивидуального потока, так и для агрегированного, поскольку принципы инжиниринга трафика от этого не меняются.

Задача инжиниринга трафика состоит в определении маршрутов прохождения потоков через сеть, т. е. для каждого потока требуется найти последовательность промежуточных коммутаторов и их интерфейсов. При

этом маршруты должны быть такими, чтобы все ресурсы сети были нагружены максимально, а каждый поток получал требуемое QoS.

Максимальный уровень использования ресурсов выбирается таким образом, чтобы механизмы управления перегрузкой могли обеспечить требуемое качество обслуживания. Это означает, что для эластичного трафика максимальное значение выбирается не больше чем 0,9, а для чувствительного к задержкам трафика – не больше чем 0,5. Так как обычно резервирование производится не для всех потоков, нужно оставить часть пропускной способности для свободного использования. Поэтому приведенные максимальные значения обычно уменьшают до 0,75 и 0,25 соответственно.

Существуют различные формальные математические постановки задачи инжиниринга трафика. На практике решением задачи инжиниринга трафика является такой набор маршрутов для заданного множества потоков, для которого все значения коэффициентов использования ресурсов вдоль маршрута следования каждого потока не превышают заданного порога $K_{\max}$.

Решение задачи инжиниринга трафика можно искать по-разному. Во-первых, можно искать его заблаговременно, в *фоновом режиме*. Для этого нужно знать исходные данные: топологию и производительность сети, а также предложенную нагрузку. После этого задачу рационального распределения путей следования трафика при фиксированных точках входа и выхода, а также заданном уровне максимального значения коэффициента использования ресурса можно передать некоторой программе, которая, например, путем направленного перебора вариантов найдет точные маршруты для каждого потока с указанием всех промежуточных коммутаторов.

Во-вторых, задачу инжиниринга трафика можно решать в *оперативном* режиме, поручив ее самим коммутаторам сети. Для этого используются модифицированные стандартные протоколы маршрутизации. Модификация протоколов маршрутизации состоит в том, что они сообщают друг другу не только топологическую информацию, но и текущее значение свободной пропускной способности для каждого ресурса.

После того как решение найдено, нужно его реализовать, т. е. отразить в таблицах маршрутизации. На этом этапе может возникнуть проблема – в том случае, если мы хотим проложить эти маршруты в дейтаграммной сети. Дело в том, что таблицы маршрутизации в них учитывают только адреса назначения пакетов. Коммутаторы и маршрутизаторы таких сетей (например, IP-сетей) не работают с потоками, для них поток в явном виде не существует, каждый пакет при его продвижении является независимой единицей коммутации. Таблицы продвижения этих сетей отражают только топологию сети (направления продвижения к узлам назначения).

Поэтому привнесение методов резервирования в дейтаграммные сети происходит с большими трудностями. В протоколах резервирования, чтобы определить поток для дейтаграммного маршрутизатора, помимо адреса назначения используется некоторый дополнительный набор признаков. При этом понятие потока привлекается только на этапе резервирования, а при продвижении пакетов по-прежнему работает традиционная для этого типа сетей схема, учитывающая лишь адрес назначения.

Методы инжиниринга трафика сегодня используются только в сетях с виртуальными каналами, для которых не составляет труда реализовать найденное решение для группы потоков. Каждому потоку (или группе потоков с одинаковыми маршрутами) выделяется виртуальный канал, который прокладывается в соответствие с выбранным маршрутом. Методы инжиниринга трафика успешно применялись в сетях ATM и Frame Relay до тех пор, пока эти технологии не прекратили свое существование. Сегодня задачи инжиниринга трафика решаются в сетях IP поверх MPLS, так как MPLS использует технику виртуальных каналов для продвижения пакетов.

4.3.5. Работа в недогруженном режиме

Самым простым способом обеспечения требований QoS для всех потоков является работа сети в недогруженном режиме, или с избыточной пропускной способностью. Говорят, что сеть имеет избыточную пропускную способность, когда все части сети в любой момент времени обладают такой пропускной способностью, которой достаточно, чтобы обслужить все потоки трафика с удовлетворительными характеристиками производительности и надежности. Другими словами, ни одно из сетевых устройств такой сети никогда не подвергается перегрузкам, которые могли бы привести к значительным задержкам или потерям пакетов из-за переполнения очередей.

Простота обеспечения требований QoS за счет работы сети в недогруженном режиме является главным достоинством этого подхода – он требует только увеличения пропускной способности линий связи и соответственно производительности коммуникационных устройств сети. Никаких дополнительных усилий по исследованию характеристик потоков, как в случае применения методов QoS, здесь не требуется.

Чтобы быть уверенным, что сеть обладает достаточной пропускной способностью для качественной передачи трафика, необходим постоянный мониторинг временны́х характеристик (задержек и из вариаций) процессов передачи пакетов сетью. А в том случае, когда результаты мониторинга начинают стабильно показывать ухудшение характеристик качества обслуживания, необходимо проводить очередную модернизацию сети и увеличивать пропускную способность линий связи и коммуникационных устройств.

Однако мониторинг задержек и их вариаций является тонкой и трудоемкой работой. Обычно операторы, которые хотят поддерживать свою сеть в недогруженном состоянии и за счет этого обеспечивать высокое качество обслуживания, решают более простую задачу – они осуществляют мониторинг уровня трафика в линиях связи сети, т. е. *измеряют коэффициент использования пропускной способности линий связи*. При этом линия связи считается недогруженной, если ее коэффициент использования постоянно не превосходит некоторый достаточно низкий уровень, например, 20–30 %.

Выводы

Механизм профилирования позволяет контролировать скорость потока пакетов и ограничивать ее в соответствии с заранее заданным уровнем.

Обратная связь является одним из механизмов QoS; она позволяет временно снизить скорость поступления пакетов в сеть для ликвидации перегрузки в узле сети. Резервирование пропускной способности «из конца в конец» позволяет добиться гарантированного качества обслуживания потока пакетов. Резервирование основано на процедуре контроля допуска потока в сеть, в ходе которой проверяется наличие доступной пропускной способности для обслуживания потока вдоль маршрута его следования.

Методы инжиниринга трафика состоят в выборе рациональных маршрутов прохождения потоков через сеть. Выбор маршрутов обеспечивает максимизацию загрузки ресурсов сети при одновременном соблюдении необходимых гарантий в отношении параметров качества обслуживания трафика.

Недогруженная сеть может обеспечить качественное обслуживание трафика всех типов без применения методов QoS. Однако для того, чтобы убедиться, что сеть действительно недогружена, требуется постоянно проводить мониторинг уровней загрузки линий связи сети, выполняя измерения с достаточно высокой частотой.

Контрольные вопросы

1. Назовите отличия между резервированием пропускной способности в сетях с коммутацией каналов и пакетов.

2. Какой механизм нужно применять для того, чтобы высокоприоритетный трафик не подавил низкоприоритетный?

3. Верно ли утверждение, что резервирование ресурсов в сети с коммутацией пакетов лишает ее возможности динамического перераспределения пропускной способности между потоками?

4. Какую задачу решают методы инжиниринга трафика?

5. Какой параметр трафика меняется при инжиниринге трафика?

6. Какой из 5 потоков будет меньше в среднем задерживаться в очереди к выходному интерфейсу 100 Мбит/с, если потоки обслуживаются взвешенными очередями,

при этом потокам отведено 40, 15, 10, 30 и 5 % пропускной способности интерфейса. Потоки имеют средние скорости: 35, 2, 8, 3 и 4 Мбит/с соответственно.

7. Что является причиной того, что поток, который обслуживается в очереди самого высокого приоритета, все равно сталкивается с необходимостью ожидания в очереди? Варианты ответов:

а) очереди более низких приоритетов;

б) собственная пульсация;

в) пульсации низкоприоритетного трафика.

8. Поясните суть работы сети в недогруженном режиме.

9. В чем заключается сложность инжиниринга трафика в дейтаграммных сетях?

10. Как оценить коэффициент пульсации трафика?

5. БЕСПРОВОДНАЯ ПЕРЕДАЧА ДАННЫХ

5.1. Беспроводные линии связи

5.1.1. Понятие беспроводной линии связи

Преимущества беспроводной передачи данных заключаются в возможности передавать информацию без проводов, привязывающих (в буквальном смысле этого слова) абонентов к определенной точке пространства.

Технологии радиодоступа достигли определенной степени зрелости в конце 1970-х гг. и обеспечили производство сравнительно компактных и недорогих радиотелефонов. С этого времени начался бум мобильной телефонии, который продолжается до настоящего времени.

Беспроводная связь не обязательно означает мобильность. Существует так называемая *фиксированная беспроводная связь*, когда взаимодействующие узлы постоянно располагаются в пределах небольшой территории, например, в определенном здании. Фиксированная беспроводная связь применяется вместо проводной, когда по какой-то причине невозможно или невыгодно использовать кабельные линии связи. Например, малонаселенная или труднодоступная местность – болотистые районы и джунгли, пустыни, Крайний Север или Антарктида еще не скоро дождутся своих кабельных систем. Другой пример – здания, имеющие историческую ценность, стены которых непозволительно подвергать испытанию прокладкой кабеля. Наконец, организация временной связи, например, при проведении конференции в здании, в котором отсутствует проводной канал.

Беспроводная связь используется для передачи данных уже достаточно давно. До недавнего времени бóльшая часть применений беспроводной связи в компьютерных сетях была связана с ее фиксированным вариантом. Начиная с середины 1990-х гг. достигла необходимой зрелости и технология *мобильных компьютерных сетей*. С появлением стандарта IEEE 802.11 в 1997 г. стало возможным строить мобильные сети Ethernet, обеспечивающие взаимодействие пользователей независимо от того, в какой стране они находятся и оборудование какого производителя применяют.

Развитие технологии мобильных телефонных сетей привело к тому, что эти сети стали очень широко использоваться для доступа в Интернет. Третье поколение мобильных сетей, известное как сети 3G, обеспечивает передачу данных со скоростью 2–10 Мбит/с. В мобильных сетях четвертого поколения 4G предел скорости возрос до 100 Мбит/с (в теории, на практике пока средняя скорость загрузки данных находится в пределах 10–20 Мбит/с).

Беспроводные сети часто связывают с радиосигналами, однако это не всегда верно. В беспроводной связи используется широкий диапазон электромагнитного спектра, от радиоволн низкой частоты в несколько килогерц до видимого света, частота которого составляет примерно 8×10^{14} Гц.

Беспроводная линия связи строится по достаточно простой схеме (рис. 5.1). Каждый узел оснащается антенной, которая одновременно является передатчиком и приемником электромагнитных волн. Электромагнитные волны распространяются в атмосфере или вакууме со скоростью 3×10^8 м/с во всех направлениях или же в пределах определенного сектора.

Рис. 5.1. Беспроводная линия связи

Направленность или ненаправленность распространения зависит от типа антенны. На рис. 5.1 показана *параболическая антенна*, которая является *направленной*. Другой тип антенн – *изотропная антенна*, представляющая собой вертикальный проводник длиной в четверть волны излучения. Изотропные антенны являются ненаправленными, они широко используются в автомобилях и портативных устройствах.

Так как при ненаправленном распространении электромагнитные волны заполняют все пространство (в пределах определенного радиуса, определяемого затуханием мощности сигнала), то это пространство может служить *разделяемой средой*. Разделение среды передачи порождает те же проблемы, что и в локальных сетях, однако здесь они усугубляются тем, что пространство в отличие от кабеля является общедоступным, а не принадлежит одной организации. Кроме того, проводная среда строго определяет направление распространения сигнала в пространстве, а беспроводная среда является, в общем случае, ненаправленной.

Для передачи дискретной информации с помощью беспроводной линии связи необходимо модулировать электромагнитные колебания передатчика в соответствии с потоком передаваемых битов.

5.1.2. Электромагнитные волны

Характеристики беспроводной линии связи – расстояние между узлами, территория охвата, скорость передачи информации и т. п. – во многом зависят от частоты используемого электромагнитного спектра (частота f и длина волны λ связаны соотношением $c = f \times \lambda$).

Радиоволны – электромагнитные волны, частоты которых условно ограничены частотами ниже 3000 ГГц, распространяющиеся в пространстве без искусственного волновода. Радиоволны в электромагнитном спектре располагаются от крайне низких частот вплоть до инфракрасного диапазона.

С учетом классификации Международным союзом электросвязи радиоволн по диапазонам, к радиоволнам относят электромагнитные волны с частотами от 0,03 Гц до 3 ТГц, что соответствует длине волны от 10 млн километров до 0,1 миллиметра.

В широком смысле радиоволнами являются всевозможные волновые процессы электромагнитного поля в аппаратуре (например, в волноводных устройствах, схемах СВЧ и др.), в линиях передачи и, наконец, в природных условиях, в среде, разделяющей передающую и приемную антенны.

Естественными источниками радиоволн являются вспышки молний и астрономические объекты. Искусственно созданные радиоволны используются для стационарной и мобильной радиосвязи, радиовещания, радиолокации, радионавигации, спутниковой связи и в других приложениях.

Электромагнитное излучение принято делить по частотным диапазонам (табл. 5.1). Между диапазонами нет резких переходов, они иногда перекрываются, а границы между ними условны.

Таблица 5.1

Диапазоны электромагнитного излучения

Название диапазона		Длины волн, λ	Частоты, *f*	Источники
Радиоволны	Сверхдлинные	более 10 км	менее 30 кГц	Атмосферные и магнитосферные явления. Радиосвязь
	Длинные	10–1 км	30–300 кГц	
	Средние	1 км –100 м	300 кГц – 3 МГц	
	Короткие	100–10 м	3–30 МГц	
	Ультракороткие	10 м – 0,1 мм	30 МГц – 3000 ГГц	
Инфракрасное излучение		1 мм – 780 нм	300 ГГц – 429 ТГц	Излучение молекул и атомов при тепловых и электрических воздействиях
Видимое излучение		780–380 нм	429–750 ТГц	
Ультрафиолетовое		380–10 нм	$7{,}5 \cdot 10^{14}$ – $3 \cdot 10^{16}$ Гц	Излучение атомов под воздействием ускоренных электронов
Рентгеновское		10 нм – 5 пм	$3 \cdot 10^{16}$– $6 \cdot 10^{19}$ Гц	Атомные процессы при воздействии ускоренных заряженных частиц
Гамма		Менее 5 пм	более $6 \cdot 10^{19}$ Гц	Ядерные и космические процессы, радиоактивный распад

Радиочастоты – частоты или полосы частот в диапазоне от 3 Гц до 3000 ГГц, которым присвоены условные наименования. Закон РФ «О связи» устанавливает следующие понятия, относящиеся к радиочастотам.

Радиочастотный спектр – совокупность радиочастот в установленных Международным союзом электросвязи пределах, которые могут быть использованы для функционирования радиоэлектронных средств.

Радиочастота – частота электромагнитных колебаний, устанавливаемая для обозначения единичной составляющей радиочастотного спектра.

Распределение полос радиочастот – определение предназначения полос радиочастот посредством записей в Таблице распределения полос радиочастот между радиослужбами Российской Федерации, на основании которых выдается разрешение на использование конкретной полосы радиочастот, а также устанавливаются условия такого использования.

5.1.3. Диапазоны радиоволн

Использование диапазонов по радиослужбам устанавливается Регламентом радиосвязи Российской Федерации и международными соглашениями. По регламенту МСЭ радиоволны разделены на диапазоны от $0{,}3\times10^N$ Гц до 3×10^N Гц, где N – номер диапазона (табл. 5.2). Российский ГОСТ 24375-80 почти полностью повторяет эту классификацию.

Таблица 5.2

Диапазоны радиоволн согласно классификации ITU

Обозн. МСЭ	Длины волн	Название волн		Диапазон	Название частот	Применение
ELF	100–10 Мм	Декамегметровые	СДВ	3–30 Гц	Крайне низкие (КНЧ)	Связь с подводными лодками, геофизические исследования
SLF	10–1 Мм	Мегаметровые		30–300 Гц	Сверхнизкие (СНЧ)	Связь с подводными лодками, геофизические исследования
ULF	1000–100 км	Гектокилометровые		300–3000 Гц	Инфранизкие (ИНЧ)	Связь с подводными лодками
VLF	100–10 км	Мириаметровые		3–30 кГц	Очень низкие (ОНЧ)	Служба точного времени, радиосвязь с подводными лодками
LF	10–1 км	Километровые	ДВ	30–300 кГц	Низкие (НЧ)	Радиовещание, радиосвязь земной волной, навигация
MF	1000–100 м	Гектометровые	СВ	300–3000 кГц	Средние (СЧ)	Радиовещание и радиосвязь земной волной и ионосферная
HF	100–10 м	Декаметровые	КВ	3–30 МГц	Высокие (ВЧ)	Радиовещание и радиосвязь ионосферная, загоризонтная радиолокация, рации
VHF	10–1 м	Метровые волны	УКВ	30–300 МГц	Очень высокие (ОВЧ)	Телевидение, радиовещание, радиосвязь тропосферная и прямой волной, рации
UHF	1000–100 мм	Дециметровые		300–3000 МГц	Ультравысокие (УВЧ)	Телевидение, радиосвязь тропосферная и прямой волной, мобильные телефоны, рации, УВЧ-терапия, микроволновые печи, спутниковая навигация

Окончание табл. 5.2

Обозн. МСЭ	Длины волн	Название волн		Диапазон	Название частот	Применение
SHF	100–10 мм	Сантиметровые		3–30 ГГц	Сверхвысокие (СВЧ)	Радиолокация, интернет, спутниковое телевещание, спутниковая- и радиосвязь прямой волной, беспроводные компьютерные сети
EHF	10–1 мм	Миллиметровые	УКВ	30–300 ГГц	Крайне высокие (КВЧ)	Радиоастрономия, высокоскоростная радиорелейная связь, радиолокация (метеорологическая, управление вооружением), медицина, спутниковая радиосвязь
THF	1–0,1 мм	Децимиллиметровые		300–3000 ГГц	Гипервысокие частоты (ГВЧ), длинноволновая область ИК излучения	Экспериментальная «терагерцовая камера», регистрирующая изображение в длинноволновом ИК (которое излучается теплокровными организмами)

На практике под низкочастотным диапазоном часто подразумевают диапазон звуковых частот, под высокочастотным – весь радиодиапазон, от 30 кГц и выше. В отечественной литературе диапазоном СВЧ в широком смысле иногда называют диапазоны УВЧ, СВЧ и КВЧ (от 0,3 до 300 ГГц), на Западе этому соответствует распространенный термин *микроволны*.

Также в отечественной литературе сложилась классификация, согласно которой мириаметровые волны называют сверхдлинными волнами (СДВ), километровые – длинными волнами (ДВ), гектометровые – средними волнами (СВ), декаметровые – короткими волнами (КВ), а все остальные, с длинами волн короче 10 м, относят к ультракоротким волнам (УКВ).

Сверхдлинные волны – радиоволны с длиной волны свыше 10 км. Они легко огибают Землю, слабо поглощаются земной поверхностью, проникают вглубь морской воды, хорошо отражаются от ионосферы. Сверхдлинные радиоволны пока имеют ограниченное применение, прежде всего из-за сложностей с сооружением огромных антенн, пригодных для работы с СДВ. Сверхдлинные волны способны обогнуть земной шар, что ценно

для исследования состояния разных слоев атмосферы. Их способность частично проникать в морскую воду и грунт позволяет использовать их для зондирования.

Длинные волны – диапазон радиоволн с частотой от 30 кГц (длина волны 10 км) до 300 кГц (длина волны 1 км). Длинные волны распространяются на расстояния до 1–2 тысяч км за счет дифракции на сферической поверхности Земли. Затем их распространение происходит за счет направляющего действия сферического волновода, не отражаясь. Диапазон используется для радиовещания (148,5–283,5 кГц), радиотелеграфной связи, радионавигационных служб и для связи с подводными лодками (9–148,5 кГц). Участок 135,7–137,8 кГц используется для любительской радиосвязи. В этом диапазоне используется сверхузкополосная (полоса до единиц Гц) телеграфная связь с медленной амплитудной манипуляцией (длина точек и тире может составлять, в зависимости от ширины полосы, десятки секунд и даже минуты). Длинные волны способны обогнуть Земной шар.

Средние волны – диапазон радиоволн с частотой от 300 кГц (длина волны 1000 м) до 3 МГц (длина волны 100 м). Средние волны (наряду с короткими) – наиболее используемый диапазон для радиовещания (526,5–1606,5 кГц) с амплитудной модуляцией. В бытовых радиоприемниках называется СВ, MW или АМ (по названию модуляции). Сетка частот вещательных станций в Европе составляет 9 кГц, в Северной и Южной Америке – преимущественно 10 кГц. Диапазон 160 м (1,8…2,0 МГц) выделен для любительской радиосвязи. Частота 500 кГц – стандартная частота для подачи сигналов бедствия. Средние волны способны распространяться на довольно большие расстояния – сотни и тысячи километров – огибая земную поверхность, а также (преимущественно в ночное время) отражаясь от ионосферы.

Короткие волны – диапазон радиоволн с частотой от 3 МГц (длина волны 100 м) до 30 МГц (длина волны 10 м). Короткие волны отражаются от ионосферы с малыми потерями. Поэтому, путем многократных отражений от ионосферы и поверхности Земли, они могут распространяться на большие расстояния. Короткие волны используются для радиовещания, а также для любительской и профессиональной радиосвязи. Качество приема при этом зависит от различных процессов в ионосфере, связанных с уровнем солнечной активности, временем года и временем суток. Так днем лучше распространяются волны меньшей длины, а ночью – большей. Для связи между наземными станциями и космическими аппаратами они непригодны, так как не проходят сквозь ионосферу. На коротких волнах наблюдаются *замирания* – изменение уровня принимаемого сигнала. Они проявляются как кратковременное снижение амплитуды несущей частоты или вовсе пропадание последней. Замирания возникают из-за того, что

радиоволны от передатчика идут к приемнику разными путями, и приходят с разной фазой и, интерферируя в приемнике, могут ослаблять друг друга.

Ультракороткие волны – традиционное название диапазона радиоволн, объединяющего метровые, дециметровые, сантиметровые, миллиметровые и децимиллиметровые волны (или диапазоны очень высоких частот – ОВЧ, ультравысоких частот – УВЧ, сверхвысоких частот – СВЧ, крайне высоких частот – КВЧ и гипервысоких частот – ГВЧ). т. е. это все радиоволны с длиной от 10 м до 0,1 мм, что соответствует частотам от 30 МГц до 3000 ГГц; – такая классификация сложилась в отечественной учебной и технической литературе. В отличие от более длинных волн распространение УКВ происходит в основном в пределах прямой видимости. Существенная особенность УКВ – это отсутствие регулярного зеркального отражения от ионосферы Земли. Вместе с тем значительное влияние на распространение УКВ оказывает тропосфера. В тропосфере происходит рефракция луча радиоволны, а также возникают другие механизмы, способствующие загоризонтному распространению УКВ. Диапазон УКВ используется в радиовещании, телевидении, мобильной радиосвязи, радиорелейной и спутниковой связи, радиолокации и для множества других применений.

Метровые волны – диапазон радиоволн с длиной волны от 10 до 1 м, что соответствует частоте от 30 до 300 МГц (очень высокие частоты, ОВЧ (Very high frequency, VHF)). Метровые волны распространяются в пределах прямой видимости на расстояния до нескольких десятков километров. Характеристики распространения метровых волн существенно зависят от рельефа местности и типа подстилающей поверхности. Влияние атмосферы Земли выражается в рассеянии метровых волн слабыми неоднородностями ионосферы и тропосферы, отражении метровых волн от ионизированных следов метеоров и искусственно ионизированных областей в атмосфере, что приводит к дальнему (до 2 тыс. км) распространению метровых волн.

Дециметровые волны – диапазон радиоволн с длиной волны от 1 м до 10 см, что соответствует частоте от 300 МГц до 3 ГГц (ультравысокие частоты, УВЧ (Ultra high frequency, UHF)). При распространении вдоль земной поверхности дециметровые волны распространяются только в пределах прямой видимости и передача более чем на 100 километров затруднена. Дальность приема сигнала может быть увеличена за счет способности дециметровых волн рассеиваться на неоднородностях тропосферы.

Сантиметровые волны – диапазон радиоволн с длиной волны от 10 до 1 см, что соответствует частоте от 3 до 30 ГГц (сверхвысокие частоты, СВЧ (Super high frequency, SHF)). Излучение этого диапазона находит разнообразное применение в современной технике. Например, стандартом

частоты для микроволновых печей и промышленных плазменных СВЧ-установок является частота 2,45 ГГц. Это частота резонансного поглощения для молекул воды, а поскольку во все продукты питания входит вода, то в СВЧ-печи с этой частотой можно эффективно нагревать любой продукт. Кроме того, для излучения на этой частоте атмосфера непрозрачна, из-за его поглощения парами воды. Связь с космическими аппаратами на орбите Земли производится преимущественно в диапазонах *C* и *Ku*.

Микроволновое излучение, сверхвысокочастотное излучение (СВЧ-излучение) – электромагнитное излучение, включающее в себя дециметровый, сантиметровый и миллиметровый диапазоны радиоволн (длина волны от 1 м (частота 300 МГц) до 1 мм (300 ГГц)). Микроволновое излучение малой интенсивности используется в портативных средствах связи – рациях, сотовых телефонах, устройствах Bluetooth, Wi-Fi и WiMAX. Поддиапазоны СВЧ в различных системах обозначений различаются; используемые в спутниковой связи, согласно классификации IEEE, приведены в табл. 5.3.

Таблица 5.3

Диапазоны спутниковой связи согласно классификации IEEE

Название диапазона	Частотный диапазон, ГГц	
	Диапазон частот РЛС	Диапазон частот в спутниковой связи
L	1,0–2,0	
S	2,0–4,0	
C	4,0–8,0	4,0–7,0
X	8,0–12,0	7,0–10,7
Ku	12,0–18,0	10,7–18,0
K	18,0–26,5	18,3–20,2; 27,5–31,5
Ka	26,5–40,0	

5.1.4. Особенности распространения радиоволн

Приведем классификацию радиоволн по способу распространения.

Прямые волны – радиоволны, распространяющиеся в свободном пространстве, например, от одного космического аппарата к другому, или от земной станции к космическому аппарату. Для этих волн влиянием атмосферы, посторонних объектов и Земли можно пренебречь.

Земные или поверхностные волны – радиоволны, распространяющиеся вдоль сферической поверхности Земли и частично огибающие ее вследствие явления *дифракции*. Способность волны огибать встречаемые препятствия и *дифрагировать* вокруг них, как известно, определяется соотношением между длиной волны и размерами препятствий. *Чем короче длина волны, тем слабее проявляется дифракция.* По этой причине волны диапазонов УВЧ и выше очень слабо дифрагируют вокруг поверхности земного

шара и дальность их распространения в первом приближении определяется расстоянием прямой видимости (прямые волны).

Тропосферные – радиоволны диапазонов ОВЧ и УВЧ, распространяющиеся за счет рассеяния на неоднородностях тропосферы на расстояние до 1000 км. Ионосферные или пространственные – радиоволны длиннее 10 м, распространяющиеся вокруг земного шара на сколь угодно большие расстояния за счет однократного или многократного отражения от ионосферы и поверхности Земли. Направляемые – радиоволны, распространяющиеся в направляющих системах (радиоволноводах).

Перечислим некоторые **общие закономерности распространения радиоволн (РРВ)**, связанные с частотой излучения.

Чем выше несущая частота, тем выше возможная скорость передачи.

Чем выше частота, тем хуже проникает сигнал через препятствия. Низкочастотные радиосигналы гектометровых волн (1000–100 м) легко проникают в дома, позволяя обходиться комнатной антенной. Более высокочастотный сигнал телевидения метровых (10–1 м) и дециметровых (1000–100 мм) волн требует, как правило, внешней антенны. И, наконец, инфракрасный и видимый свет не проходят через стены, ограничивая передачу прямой видимостью (Line of Sight, LOS).

Чем выше частота, тем быстрее убывает энергия сигнала с расстоянием от источника. При распространении радиоволн в свободном пространстве (без отражений) затухание мощности сигнала пропорционально произведению квадрата расстояния от источника сигнала на квадрат частоты сигнала.

Низкие частоты до 2 МГц распространяются вдоль поверхности Земли; именно поэтому сигналы радиовещания на ДВ и СВ могут передаваться на расстояния в сотни километров. Сигналы от 2 до 30 МГц отражаются ионосферой Земли, поэтому они могут распространятся даже на более значительные расстояния – в несколько тысяч километров (при достаточной мощности передатчика). Сигналы в диапазоне выше 30 МГц распространяются только по прямой, т. е. являются сигналами прямой видимости. При частоте свыше 4 ГГц их подстерегает неприятность – они начинают поглощаться водой, т. е. и дождь, и туман могут стать причиной резкого ухудшения качества.

Потребность в скоростной передаче информации является превалирующей, поэтому все современные системы беспроводной передачи информации работают в высокочастотных диапазонах, начиная с 800 МГц. Для успешного использования микроволнового диапазона необходимо учитывать проблемы, связанные с поведением сигналов, распространяющихся в режиме прямой видимости и встречающих на своем пути препятствия.

На рис. 5.2 показано, что сигнал, встретившись с препятствием, может распространяться в соответствии с тремя механизмами: *отражением, дифракцией* и *рассеиванием*.

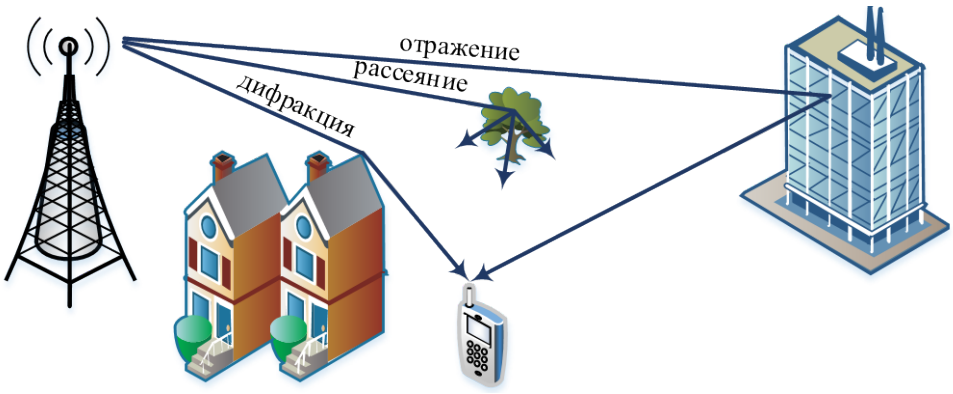


Рис. 5.2. Механизмы РРВ: отражение, дифракция, рассеяние

Когда сигнал встречается с препятствием, которое частично прозрачно для данной длины волны и в то же время размеры которого намного превышают длину волны, то часть энергии сигнала *отражается* от такого препятствия. Волны микроволнового диапазона имеют длину несколько сантиметров, поэтому они частично отражаются от стен домов при передаче сигналов в городе. Если сигнал встречает непроницаемое для него препятствие (например, металлическую пластину) большего размера, чем длина волны, то происходит *дифракция* – сигнал как бы огибает препятствие, так что такой сигнал можно получить, даже не находясь в зоне прямой видимости. И, наконец, при встрече с препятствием, размеры которого соизмеримы с длиной волны, сигнал *рассеивается*, распространяясь под различными углами. В результате подобных явлений, которые повсеместно встречаются при беспроводной связи в городе, приемник может получить несколько копий одного и того же сигнала. Такой эффект называется *многолучевым распространением сигнала*. Результат многолучевого распространения сигнала часто оказывается отрицательным, поскольку один из сигналов может прийти в противофазе и подавить основной сигнал.

Так как время распространения сигнала вдоль различных путей является в общем случае различным, то может также наблюдаться *межсимвольная интерференция* – ситуация, когда в результате задержки сигналы, кодирующие соседние биты данных, доходят до приемника одновременно.

Искажения из-за многолучевого распространения приводят к эффекту *многолучевого замирания*. В городах многолучевое замирание приводит к тому, что ослабление сигнала становится пропорциональным не квадрату расстояния, а его кубу или даже четвертой степени!

5.1.5. Помехи в беспроводной связи и лицензирование

Отказ от проводов и обретение мобильности приводят к высокому уровню помех в беспроводных линиях связи. Если интенсивность битовых ошибок (*BER*) в проводных линиях связи равна 10^{-9}–10^{-10}, то в беспроводных линиях связи она достигает величины 10^{-3}. Проблема высокого уровня помех беспроводных каналов решается различными способами. Важную роль играют специальные *технологии широкополосного сигнала*. Кроме того, передатчики сигнала (и приемники, если это возможно) стараются разместить на высоких башнях, чтобы избежать многократных отражений. Еще одним примером является применение протоколов с установлением соединения и повторными передачами кадров на *канальном* уровне стека протоколов.

Понятие лицензирования. Радиоволны могут распространяться во всех направлениях на значительные расстояния и проходить через препятствия, такие как стены домов. Поэтому проблема разделения электромагнитного спектра является весьма острой и требует *централизованного* регулирования. В каждой стране есть государственный орган, который выдает *лицензии* операторам связи на использование определенной части спектра. Лицензия выдается на определенную территорию, в пределах которой оператор задействует закрепленный за ним диапазон частот монопольно.

Существует также три частотных диапазона, 900 МГц, 2,4 ГГц и 5 ГГц, которые рекомендованы ITU как диапазоны для международного использования *без лицензирования*. Эти диапазоны выделены промышленным товарам беспроводной связи общего назначения, например, устройствам блокирования дверей автомобилей, научным и медицинским приборам. В соответствии с назначением эти диапазоны получили название ISM-диапазонов (Industrial, Scientific, Medical – промышленность, наука, медицина). Активно осваивается диапазон 2,4 ГГц, например, в технологиях IEEE 802.11 и Bluetooth. Обязательным условием использования этих диапазонов на совместной основе является ограничение максимальной мощности передаваемых сигналов. Это условие сокращает радиус действия устройств, чтобы их сигналы не стали помехами для других пользователей.

Выводы

Беспроводная связь делится на фиксированную и мобильную. Для организации мобильной связи беспроводная среда является единственной альтернативой. Фиксированная беспроводная связь обеспечивает доступ к узлам сети, расположенным в пределах небольшой территории, например, здания. Каждый узел беспроводной линии связи оснащается антенной,

которая одновременно является передатчиком и приемником электромагнитных волн. Электромагнитные волны могут распространяться во всех направлениях или же в пределах определенного сектора. Тип распространения зависит от типа антенны. Из-за отражения, дифракции и рассеяния электромагнитных волн возникает многолучевое распространение одного и того же сигнала, что приводит к замиранию.

Передача данных в ISM-диапазонах 900 МГц, 2,4 ГГц и 5 ГГц не требует лицензирования, если мощность передатчика не превышает заданную величину.

Контрольные вопросы

1. Назовите основные области применения беспроводных линий связи.
2. Приведите основные диапазоны электромагнитных и радиоволн.
3. В чем достоинства и недостатки беспроводной передачи информации по сравнению с проводной?
4. За счет чего радиоволны с частотами от 2 до 30 МГц могут распространяться на сотни километров?
5. Поясните суть лицензирования спектра.
6. Поясните происхождение и особенности ISM-диапазонов.
7. Поясните происхождение помех в беспроводной связи.
8. Сформулируйте основные закономерности PPB, связанные с частотой излучения.
9. Какие атмосферные явления мешают распространению микроволн?
10. Какие препятствия вызывают дифракцию? Варианты ответов:

а) непроницаемые препятствия, размер которых соизмерим с длиной волны;

б) непроницаемые препятствия, размер которых намного больше длины волны;

в) непроницаемые препятствия, размер которых намного меньше длины волны.

5.2. Беспроводные системы связи

5.2.1. Беспроводные системы связи точка-точка

По **двухточечной схеме** могут работать беспроводные каналы различного назначения, использующие различные диапазоны частот. В телекоммуникационных сетях эта схема уже долгое время применяется для создания радиорелейных линий связи. Такую линию образуют несколько башен, на которых установлены параболические направленные антенны (рис. 5.3).

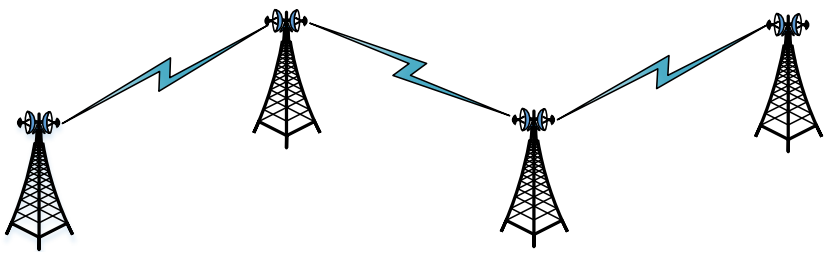


Рис. 5.3. Радиорелейная линия связи

Каждая линия работает в микроволновом диапазоне на частотах в несколько гигагерц. Направленная антенна концентрирует энергию в узком пучке, что позволяет передавать информацию на значительные расстояния, обычно до 50 км. Высокие башни обеспечивают прямую видимость антенн [2].

Пропускная способность линии может быть достаточно высокой, обычно она находится в пределах от нескольких до сотен мегабит в секунду. Это могут быть как магистральные линии, так и линии доступа. Операторы связи часто используют подобные линии, когда прокладка оптического волокна либо невозможна, либо экономически невыгодна.

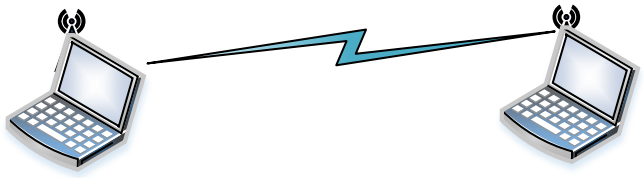


Рис. 5.4. Беспроводная связь двух компьютеров

Другой пример беспроводной двухточечной линии связи показан на рис. 5.4. Здесь она служит для соединения двух компьютеров. Такая линия образует простейший сегмент локальной сети, поэтому расстояния и мощности сигнала здесь принципиально иные.

Для расстояний в пределах одного помещения может использоваться микроволновый диапазон. Микроволновый диапазон работает в пределах нескольких десятков или сотен метров – предельное расстояние предсказать сложно, так как при РРВ в помещении происходят многочисленные отражения, дифракции и рассеивания, к которым добавляются эффекты проникновения волн через стены и межэтажные перекрытия.

5.2.2. Беспроводные системы связи точка-многоточка

Схема беспроводного канала **с одним источником и несколькими приемниками** характерна для организации доступа, когда многочисленные абонентские терминалы соединяются с базовой станцией (Base Station, BS).

Беспроводные линии связи в схеме с одним источником и несколькими приемниками служат как для фиксированного, так и для мобильного доступа. На рис. 5.5 показан вариант фиксированного доступа с помощью микроволновых линий связи. Оператор связи использует высокую башню (возможно, телевизионную), чтобы обеспечить прямую видимость с антеннами, установленными на крышах зданий своих клиентов. Фактически такой вариант может представлять собой набор двухточечных линий связи – по количеству зданий, которые необходимо соединить с базовой станцией. Однако это достаточно расточительный вариант, так как для каждого нового клиента нужно устанавливать новую антенну на башне. Поэтому для экономии обычно применяют антенны, захватывающие определенный сектор, например, в 45°. Тогда за счет нескольких антенн оператор может обеспечить связь в пределах полного сектора в 360°, конечно, на ограниченном расстоянии (обычно несколько километров).

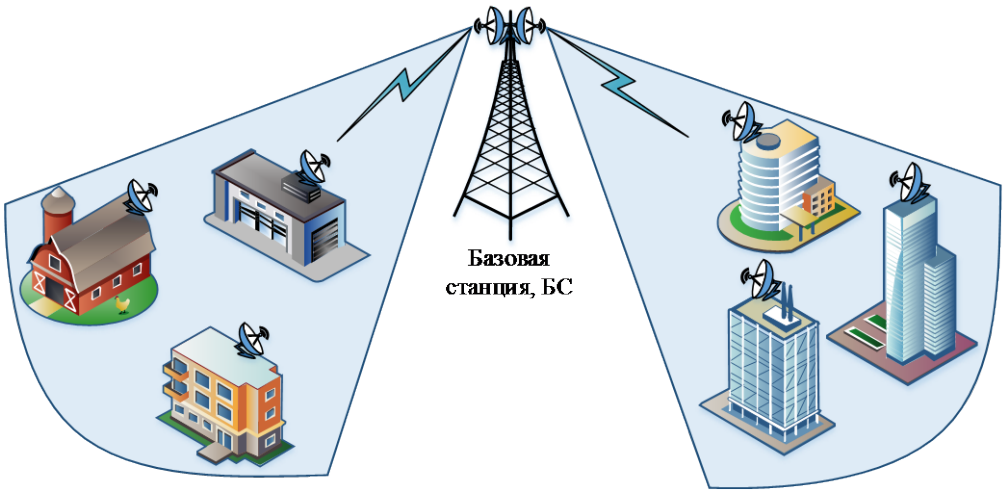


Рис. 5.5. Фиксированный беспроводный доступ

Пользователи линий доступа могут обмениваться информацией только с базовой станцией, а она, в свою очередь, транзитом обеспечивает взаимодействие между отдельными пользователями.

Базовая станция обычно соединяется проводной связью с проводной частью сети, обеспечивая взаимодействие с пользователями других базовых станций или пользователями проводных сетей. Поэтому базовая станция также называется точкой доступа (Access Point, AP). Точка доступа включает не только оборудование, необходимое для образования линии связи, но и чаще всего является коммутатором сети, доступ к которой она обеспечивает, – телефонным коммутатором или коммутатором пакетов.

В большинстве схем мобильного доступа используется сегодня принцип *com*, которые представляют собой небольшие по площади территории, обслуживаемые одной базовой станцией. Идея сот родилась не сразу, первые мобильные телефоны работали по другому принципу, обращаясь к одной базовой станции, покрывающей большую территорию. Идея небольших сот была впервые сформулирована еще в 1945 г. С тех пор прошло довольно много времени, пока заработали первые коммерческие сотовые телефонные сети – пробные участки появились в конце 1960-х гг., а широкое коммерческое применение началось в начале 1980-х гг.

Принцип разбиения всей области охвата сети на небольшие соты дополняется идеей многократного использования частоты. На рис. 5.6 показан вариант организации сот при наличии всего трех частот, при этом ни одна из соседних пар сот не задействует одну и ту же частоту. Многократное использование частот позволяет оператору экономно расходовать выделенный ему частотный диапазон, при этом абоненты и базовые станции соседних сот не испытывают проблем из-за интерференции сигналов. Конечно, базовая станция должна контролировать мощность излучаемого сигнала, чтобы две соты (несмежные), работающие на одной и той же

частоте, не создавали друг другу помех. При гексагональной форме сот количество повторяемых частот может быть больше, чем 3, например, 4, 7, 9, 12, 13 и т. д.

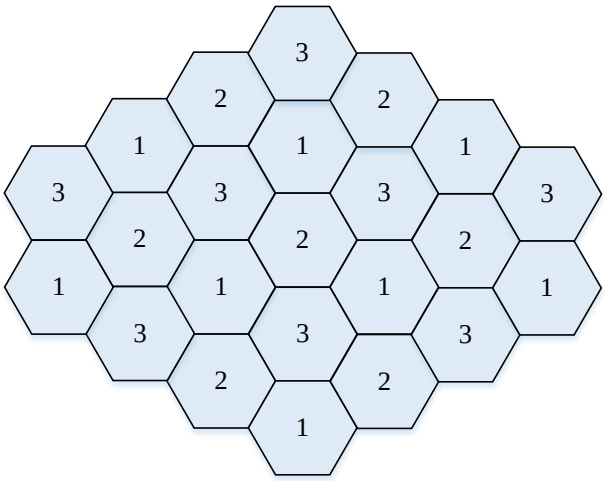


Рис. 5.6. Многократное использование
частот в сотовой сети

Важной проблемой мобильной линии связи является переход терминального устройства из одной соты в другую. Эта процедура, которая называется *эстафетной передачей*, отсутствует при фиксированном доступе и относится к протоколам более высоких уровней, нежели физический.

Поддержка передачи компьютерных данных стала обязательной в мобильных телефонных сетях третьего и четвертого поколений (3G и 4G).

В **схеме с несколькими источниками и несколькими приемниками** беспроводная линия связи представляет собой общую электромагнитную среду, разделяемую несколькими узлами. Каждый узел может использовать эту среду для взаимодействия с любым другим узлом без обращения к базовой станции. Так как базовая станция отсутствует, то необходим децентрализованный алгоритм доступа к среде. Чаще всего такой вариант беспроводного канала применяется для соединения компьютеров (рис. 5.7).

Рис. 5.7. Беспроводная
многоточечная линия связи

Собственно, первая локальная сеть Aloha, созданная в 1970-е гг. на Гавайях, в точности соответствовала схеме, приведенной на рис. 5.7. Ее отличие от современных беспроводных локальных сетей состояло в низкой скорости передачи данных (9600 бит/с), а также в весьма неэффективном способе доступа, позволяющем использовать только 18 % полосы пропускания.

Децентрализованные многоточечные схемы беспроводного доступа не являются широко распространенными, но в некоторых ситуациях, когда обычная связь с центральной точкой доступа оказывается нерабочей (например, в результате стихийного бедствия, технического отказа сети провайдера или же ее отключения по политической причине), такие схемы оказываются очень востребованными и эффективными. Яркий пример – использование участниками протестов в Гонконге осенью 2014 г. приложения для смартфонов FireChat, обеспечивающего децентрализованную маршрутизацию сообщений между телефонами, находящимися в пределах прямой доступности по протоколу Bluetooth или WiFi.

5.2.3. Типы спутниковых систем

Спутниковая связь служит для организации высокоскоростных микроволновых линий. Для таких линий нужна прямая видимость и спутник, как отражатель сигнала оказывается естественным решением (рис. 5.8) [2].

Рис. 5.8. Спутник как отражатель сигнала

Первый спутник, запущенный Советским Союзом, обладал очень ограниченными телекоммуникационными возможностями, – он только передавал радиосигнал «бип-бип», извещая мир о своем присутствии в космосе. Однако успех России в космосе подхлестнул усилия Америки, и в 1962 г. она запустила первый телекоммуникационный спутник Telstar-1, который поддерживал 600 голосовых каналов. Со времени запуска первого телекоммуникационного спутника прошло уже более 50 лет, и функции спутника как телекоммуникационного узла, естественно, усложнились. Сегодня спутник может играть роль узла первичной сети, а также телефонного коммутатора и коммутатора/маршрутизатора компьютерной сети. Для этого аппаратура спутников взаимодействует не только с наземными станциями, но и между собой, образуя прямые космические беспроводные линии связи. Принципиально техника передачи микроволновых сигналов в космосе и на Земле не отличается, однако у спутниковых линий связи есть и очевидная специфика – один из узлов такой линии постоянно находится в полете, причем на большом расстоянии от других узлов.

Для спутниковой связи ITU выделил несколько *частотных диапазонов*.

Исторически первым использовался диапазон **C**, в котором для каждого из дуплексных потоков Земля-спутник (восходящая частота 4 ГГц) и спутник-Земля (нисходящая частота 6 ГГц) выделяется по 500 МГц – этого достаточно для большого числа каналов. Диапазоны **L** и **S** занимают более

низкие частоты и предназначаются для организации мобильных услуг с помощью спутников. Они также часто используются наземными системами.

Спутники отличаются высотой орбиты над Землей. Существует три группы орбит (рис. 5.9): а) геостационарная орбита (Geostationary Orbit, GEO) – 35863 км; б) средневысотная орбита (Medium Earth Orbit, MEO) – 5000–15000 км; в) маловысотная орбита (Low Earth Orbit, LEO) – 100–1000 км.

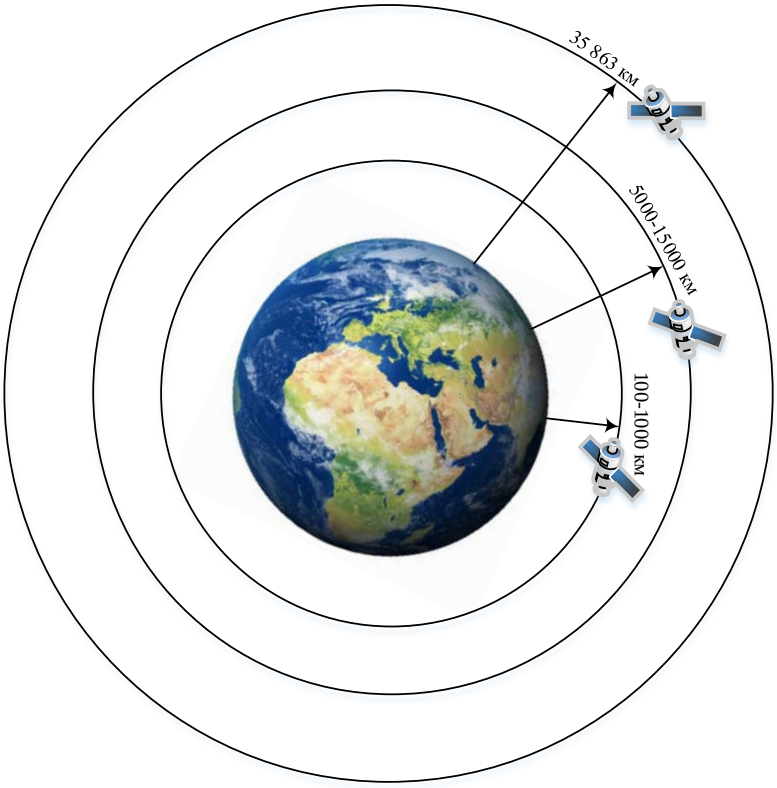


Рис. 5.9. Типы орбит спутников

Геостационарный спутник «висит» над определенной точкой экватора, в точности следуя скорости вращения Земли. Такое положение выгодно по нескольким обстоятельствам.

Геостационарный спутник «висит» над определенной точкой экватора, в точности следуя скорости вращения Земли. Такое положение выгодно по нескольким обстоятельствам. Во-первых, с такой высоты четверть поверхности Земли оказывается в зоне прямой видимости, поэтому с помощью геостационарных спутников *просто организовать широковещание в пределах страны или даже континента*. Во-вторых, геостационарный спутник находится за пределами земной атмосферы и *меньше* «*изнашивается*», чем низкоорбитальные и средневысотные спутники; низкоорбитальные спутники из-за трения о воздух постоянно теряют высоту, и им приходится восстанавливать ее с помощью двигателей. Путем применения нескольких антенн геостационарные спутники *поддерживают большое количество каналов*.

Наряду с достоинствами у геостационарных спутников есть и недостатки. Наиболее очевидные связаны с *большим удалением спутника от поверхности Земли*. Это приводит к большим задержкам распространения сигнала (от 230 до 280 мс), при передаче разговора или телевизионного диалога возникают неудобные паузы, мешающие нормальному общению.

Принципиальным недостатком геостационарного спутника с его круговой орбитой является также *плохая связь для районов, близких к Северному и Южному полюсам*. Сигналы в таких районах проходят бóльшие расстояния, чем в районах, расположенных в экваториальных и умеренных широтах, и, естественно, больше ослабляются.

Среднеорбитальные спутники обеспечивают диаметр покрытия от 10 000 до 15 000 км и задержку распространения сигнала 50 мс. Наиболее известной услугой, предоставляемой спутниками этого класса, является глобальная система навигации (Global Positioning System, GPS), известная также под названием NAVigation Satellites providing Time And Range (NAVSTAR). GPS – это всеобщая система определения текущих координат пользователя на поверхности Земли или в околоземном пространстве. GPS состоит из 24 спутников – это то минимальное число спутников, которое необходимо для стопроцентного покрытия территории Земли. Первый тестовый спутник GPS был запущен в 1974 г., первый промышленный спутник – в 1978 г., а 24-й промышленный – в 1993 г. Спутники GPS летают на орбите высотой около 20 000 км. Помимо спутников в систему GPS входит сеть наземных станций слежения за ними и неограниченное количество пользовательских приемников-вычислителей, среди которых и популярные приемники автомобильных систем навигации.

По радиосигналам спутников GPS-приемники пользователей устойчиво и точно определяют координаты; для этого на поверхности Земли приемнику необходимо принять сигналы как минимум от трех спутников. Погрешности не превышают десятков метров. Этого достаточно для решения задач навигации подвижных объектов (самолеты, корабли, автомобили и т. д.).

В СССР была разработана и реализована система аналогичного назначения под названием ГЛОНАСС (ГЛОбальная Навигационная Спутниковая Система). Первый спутник ГЛОНАСС был запущен в 1982 г., а в сентябре 1993 г. система была официально введена в эксплуатацию. В 1995 г. количество спутников достигло плановой цифры 24 (такое количество необходимо для глобального покрытия Земли, для покрытия территории России достаточно 18 спутников), но затем из-за проблем с финансированием не все выходившие из строя спутники заменялись новыми, поэтому к 2001 г. число работающих спутников сократилось до 6. В 2001 г. была принята новая федеральная программа «Глобальная навигационная система», предусматривающая модернизацию системы спутников к 2008 г. и ее

полноценную эксплуатацию в 2010 г. С некоторыми задержками эта программа была выполнена, глобальное покрытие было обеспечено в конце 2011 г. Система ГЛОНАСС совместима с GPS, существует навигационное оборудование, которое может принимать сигналы от спутников обоих типов.

Достоинства и недостатки *низкоорбитальных спутников* противоположны соответствующим качествам геостационарных спутников. Главное их достоинство – близость к Земле, а значит, пониженная мощность передатчиков, малые размеры антенн и небольшое время распространения сигнала (около 20–25 мс). Кроме того, их легче запускать. Основной недостаток – малая площадь покрытия, диаметр которой составляет всего около 8000 км. Период оборота такого спутника вокруг Земли составляет 1,5–2 часа, а время видимости спутника наземной станцией – всего 20 минут. Это значит, что постоянная связь с помощью низкоорбитальных спутников может быть обеспечена, только когда на орбите находится достаточно большое их количество. Кроме того, атмосферное трение снижает срок службы таких спутников до 8–10 лет. Если основным назначением геостационарных спутников является широковещание и дальняя связь, то низкоорбитальные спутники рассматриваются как важное средство поддержания мобильной связи.

В начале 1990-х гг. достоинства компактных терминальных устройств для низкоорбитальных спутников показались руководителям компании *Motorola* более важными, чем их недостатки. Вместе с несколькими крупными партнерами эта компания начала проект *Iridium*, который имел весьма амбициозную цель – создать всемирную спутниковую сеть, обеспечивающую мобильную связь в любой точке земного шара. В конце 1980-х гг. еще не существовало такой плотной системы сот мобильной телефонии, как сегодня, так что коммерческий успех казался обеспеченным.

В 1997 г. группа из 66 спутников была запущена, а в 1998 г. началась коммерческая эксплуатация системы *Iridium.* Спутники *Iridium* действительно покрывают всю поверхность земного шара, вращаясь по 6 орбитам, проходящим через полюсы Земли. На каждой орбите находится по 11 спутников, передатчики которых работают на частоте 1,6 ГГц с полосой пропускания 10 МГц. Эта полоса расходуется 240 каналами по 41 кГц каждый. За счет многократного использования частот система *Iridium* поддерживает 253 440 каналов, организуя системы скользящих по поверхности Земли сот. Для пользователей системы *Iridium* основным видом услуги является телефонная связь и передача данных со скоростью 2,4 Кбит/с.

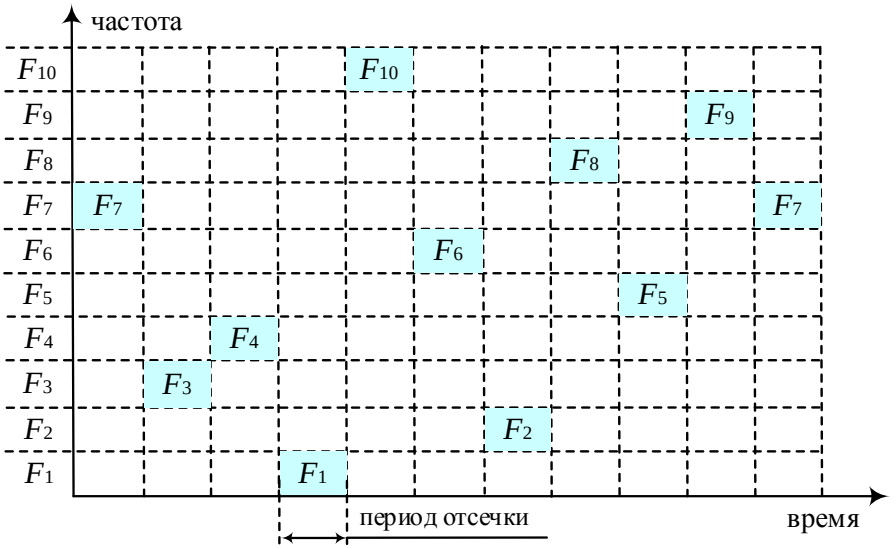
К сожалению, коммерческие успехи *Iridium* оказались очень скромными, и через два года своего существования компания обанкротилась. Расчет на мобильных телефонных абонентов оказался неверным – к моменту начала работы наземная сеть сотовой связи уже покрывала бóльшую часть территории развитых стран. А услуги по передаче данных со скоростью 2,4 Кбит/с не соответствовали потребностям пользователей конца XX в.

Сегодня система *Iridium* снова работает, теперь уже с новым владельцем и именем – *Iridium Communications.* У нее теперь более скромные планы, связанные с созданием местных систем связи там, где другая связь практически отсутствует, например, на научных станциях Антарктиды.

5.2.4. Технологии широкополосного сигнала

Техника расширенного спектра разработана специально для беспроводной передачи. Она позволяет повысить помехоустойчивость сигналов малой мощности за счет увеличения спектра передаваемого сигнала, что очень важно в мобильных приложениях. Существует несколько методов расширения спектра.

Идея метода **расширения спектра скачкообразной перестройкой рабочей частоты** (ППРЧ, Frequency Hopping Spread Spectrum, FHSS) возникла во время Второй мировой войны, когда радио широко использовалось для секретных переговоров и управления военными объектами, например, торпедами. Для того, чтобы радиообмен нельзя было перехватить или подавить узкополосным шумом, было предложено вести передачу с постоянной сменой несущей в пределах широкого диапазона частот. В результате мощность сигнала распределялась по всему диапазону и прослушивание какой-то определенной частоты давало только небольшой шум. Последовательность несущих частот выбиралась псевдослучайной, известной только передатчику и приемнику. Попытка подавления сигнала в каком-то узком диапазоне также не слишком ухудшала сигнал, так как подавлялась только небольшая часть информации. Идею этого метода иллюстрирует рис. 5.10 [2].



Последовательность перестройки частот: F_7–F_3–F_4–F_1–F_{10}–F_6–F_2–F_8–F_5–F_9

Рис. 5.10. Расширение спектра
скачкообразной перестройкой частоты

В течение определенного фиксированного интервала времени передача ведется на неизменной несущей частоте. На каждой несущей частоте для передачи дискретной информации применяются стандартные методы модуляции, такие как FSK и PSK. Чтобы приемник синхронизировался с передатчиком, передаются синхробиты, поэтому полезная скорость этого метода оказывается меньше из-за накладных расходов на синхронизацию.

Несущая частота меняется в соответствии с номерами частотных подканалов, вырабатываемых псевдослучайно. Псевдослучайная последовательность зависит от некоторого параметра, называемого начальным числом. Если приемнику и передатчику известны алгоритм и значение начального числа, то они меняют частоты в одинаковой последовательности, называемой *последовательностью псевдослучайной перестройки частоты*.

Если частота смены подканалов ниже, чем скорость передачи данных в канале, то такой режим называют медленным расширением спектра (рис. 5.11, а); в противном случае – быстрым расширением спектра (рис. 5.11, б).

Метод быстрого расширения спектра более устойчив к помехам, поскольку узкополосная помеха, подавляя сигнал в определенном подканале, не приводит к потере бита, так как его значение передается несколько раз в различных частотных подканалах. В этом режиме не проявляется эффект межсимвольной интерференции, так как ко времени прихода, задержанного вдоль одного из путей сигнала, система успевает перейти на другую частоту.

Метод медленного расширения спектра таким свойством не обладает, но зато он проще в реализации и имеет меньшие накладные расходы.

Методы FHSS применяют в технологиях IEEE 802.11 (Wi-Fi) и Bluetooth. В методах FHSS подход к использованию частотного диапазона не такой, как в других методах передачи, – вместо экономного расходования узкой полосы делается попытка занять весь доступный диапазон. На первый взгляд это кажется не очень эффективным – ведь в каждый момент времени в диапазоне работает только один канал. Однако последнее утверждение не всегда справедливо, поскольку коды расширенного спектра можно задействовать также и для мультиплексирования *нескольких* каналов в широком диапазоне. В частности, методы FHSS позволяют организовать одновременную работу нескольких каналов путем выбора для каждого канала таких псевдослучайных последовательностей, которые в каждый момент времени дают каждому каналу возможность работать на собственной частоте.

В методе **прямого последовательного расширения спектра** (Direct Sequence Spread Spectrum, DSSS) частотный диапазон расширяется не за счет постоянных переключений с частоты на частоту, как в методе FHSS, а за счет того, что каждый бит информации заменяется N битами, поэтому

тактовая скорость передачи сигналов увеличивается в N раз. А это, в свою очередь, означает, что спектр сигнала также расширяется в N раз.

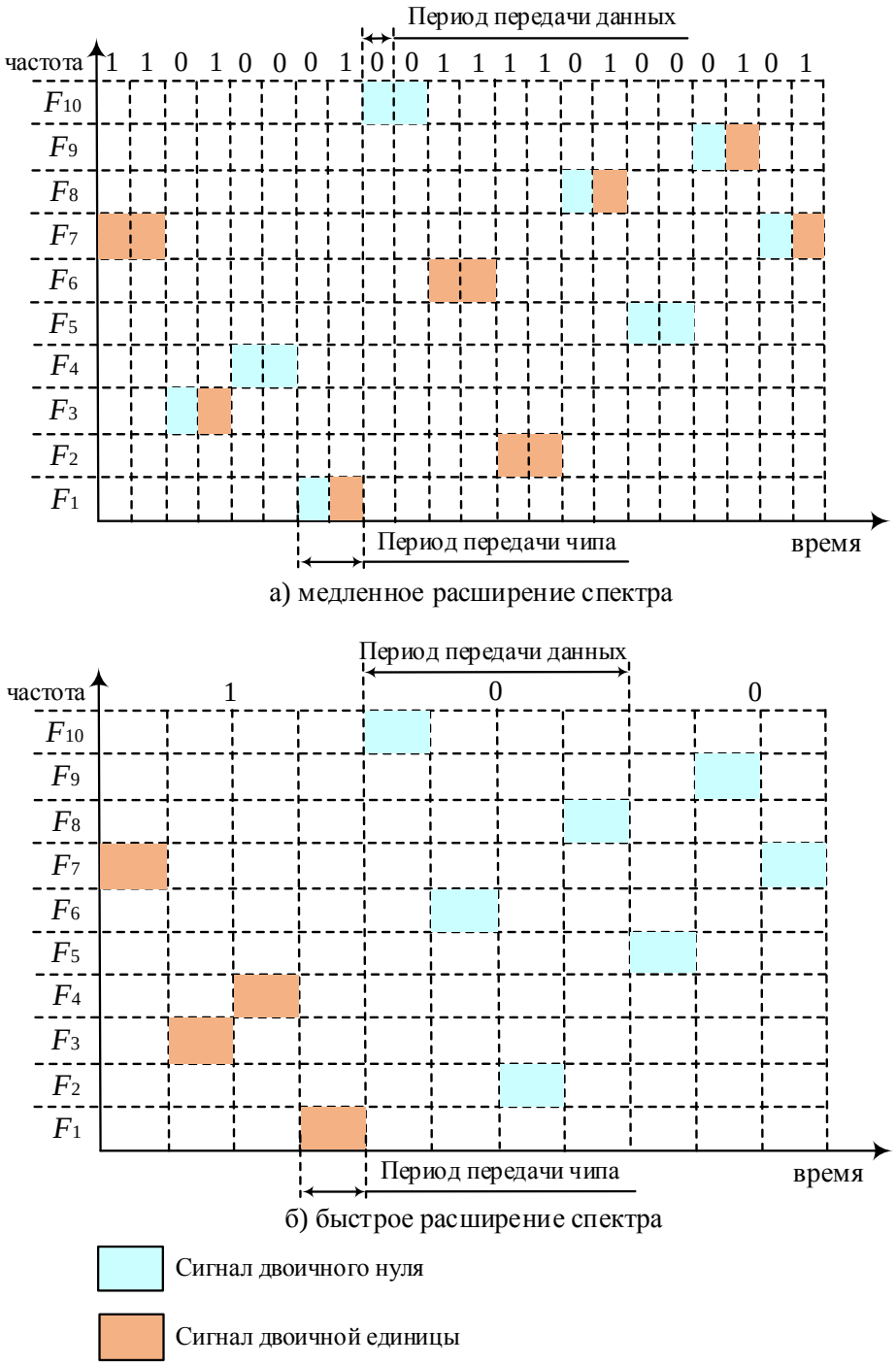


Рис. 5.11. Соотношение
между скоростью передачи данных и частотой смены подканалов

Цель кодирования методом DSSS та же, что и методом FHSS, – *повышение помехоустойчивости*. Узкополосная помеха будет искажать только определенные частоты спектра сигнала, так что приемник с большой степенью вероятности сможет правильно распознать передаваемую информацию.

Код, которым заменяется двоичная единица исходной информации, называется *расширяющей последовательностью*, а каждый бит такой последовательности – *чипом*. Соответственно скорость передачи результирующего сигнала называют чиповой скоростью. Двоичный нуль кодируется инверсным значением расширяющей последовательности. Приемники должны знать расширяющую последовательность, которую использует передатчик, чтобы понять передаваемую информацию. Количество чипов в расширяющей последовательности определяет *коэффициент расширения* исходного кода. Как и в случае FHSS, для кодирования битов результирующего кода может использоваться любой вид модуляции. Чем больше коэффициент расширения, тем шире спектр результирующего сигнала и тем больше степень подавления помех. Но при этом растет занимаемый каналом диапазон спектра. Обычно коэффициент расширения имеет значения от 10 до 100.

Примером расширяющей последовательности является последовательность Баркера (Barker), которая состоит из 11 бит: 10110111000. Если передатчик использует эту последовательность, то передача трех битов 110 ведет к отправке следующих битов: 10110111000 10110111000 01001000111.

Последовательность Баркера позволяет приемнику быстро синхронизироваться с передатчиком, т. е. надежно выявлять начало последовательности. Приемник определяет такое событие, поочередно сравнивая получаемые биты с образцом последовательности. Действительно, если сравнить последовательность Баркера с такой же последовательностью, но сдвинутой на один бит влево или вправо, то мы получим меньше половины совпадений значений битов. Значит, даже при искажении нескольких битов с большой долей вероятности приемник правильно определит начало последовательности, а значит, сможет правильно интерпретировать принимаемые данные.

Метод DSSS в меньшей степени защищен от помех, чем метод быстрого расширения спектра, так как мощная узкополосная помеха влияет на часть спектра, а значит, и на результат распознавания единиц и нулей.

Множественный доступ с кодовым разделением. Как и в случае FHSS, кодирование методом DSSS позволяет мультиплексировать несколько каналов в одном диапазоне. Техника такого мультиплексирования называется *множественным доступом с кодовым разделением* (Code Division Multiple Access, CDMA). Идея CDMA заключается в том, что каждый узел сети задействует собственное значение расширяющей последовательности. Эти значения выбираются так, чтобы принимающий узел, который знает значение расширяющей последовательности передающего узла, мог выделить данные передающего узла из суммарного сигнала, образующегося в результате одновременной передачи несколькими узлами.

Выводы

Беспроводные двухточечные линии связи служат для создания радиорелейных линий, соединения зданий, а также пары компьютеров. Беспроводные линии связи с одним источником и несколькими приемниками строятся на основе базовой станции. Такие линии используются в мобильных сотовых сетях, а также в системах фиксированного доступа. Топология с несколькими источниками и несколькими приемниками характерна для беспроводных локальных сетей.

В системах спутниковой связи используются три группы спутников: геостационарные, среднеорбитальные и низкоорбитальные.

Для кодирования дискретной информации в беспроводных системах прибегают к методам расширения спектра (FHSS и DSSS). В методах расширения спектра для представления информации используется широкий диапазон частот, это уменьшает влияние на сигналы узкополосных шумов. На основе методов FHSS и DSSS можно мультиплексировать несколько каналов в одном диапазоне частот. Такая техника мультиплексирования называется множественным доступом с кодовым разделением (CDMA).

Контрольные вопросы

1. Приведите пример беспроводной системы связи точка-точка.
2. Приведите пример беспроводной системы связи точка-многоточка.
3. В чем суть многократного использования частот в сотовой сети?
4. Приведите существующие группы орбит для организации спутниковой связи.
5. Сформулируйте достоинства и недостатки геостационарных спутников связи.
6. Сформулируйте достоинства и недостатки низкоорбитальных спутников.
7. Чем отличается быстрое расширение спектра от медленного в методе FHSS?
8. Сформулируйте особенности функционирования метода DSSS.
9. Каковы недостатки геостационарного спутника? Варианты ответов:

а) велики задержки сигнала;

б) велико затухание сигнала, что приводит к необходимости использования антенн большого диаметра;

в) мало покрытие территории;

г) хорошая связь обеспечивается лишь в районах, близких к Северному и Южному полюсам.

10. В чем заключается причина неудачи проекта Iridium?

6. ТЕХНОЛОГИИ ЛОКАЛЬНЫХ СЕТЕЙ НА РАЗДЕЛЯЕМОЙ СРЕДЕ

6.1. Общая характеристика протоколов на разделяемой среде

Сегодня технологии локальных сетей на разделяемой среде применяются только в беспроводных локальных сетях Wi-Fi. В проводных же локальных сетях с середины 1990-х гг. разделяемая среда не используется из-за плохой масштабируемости. И хотя в стандартах единственной выжившей технологии локальных проводных сетей – Ethernet – вариант работы на разделяемой среде все еще описан, он разрешен только для низко- и среднескоростных версий Ethernet, но не для скоростей 10 и 100 Гбит/с.

Тем не менее описание основных идей и характеристик Ethernet целесообразно, так как это помогает понять особенности техники применения разделяемой среды, что полезно при разработке новых технологий беспроводных сетей, где эта техника является естественной. Кроме того, знание истории развития технологии помогает лучше понять некоторые ее унаследованные черты, такие как, например, размер и формат кадра Ethernet, сохранившиеся и в современных коммутируемых версиях Ethernet.

6.1.1. Стандартная топология и разделяемая среда

Основная цель, которую ставили перед собой разработчики первых локальных сетей во второй половине 1970-х гг., заключалась в нахождении простого и дешевого решения для объединения в вычислительную сеть нескольких десятков компьютеров, находящихся в пределах одного здания. Решение должно было быть недорогим, поскольку компьютеры, объединявшиеся в сеть, были недороги. Количество их в одной организации было небольшим, поэтому предел в несколько десятков компьютеров представлялся вполне достаточным для практически любой локальной сети. Задача связи локальных сетей в глобальные не была первоочередной, поэтому практически все технологии локальных сетей ее игнорировали.

Для упрощения и соответственно удешевления аппаратных и программных решений разработчики первых локальных сетей остановились на совместном использовании общей среды передачи данных.

Этот метод связи компьютеров впервые был апробирован при создании радиосети ALOHA Гавайского университета в начале 1970-х гг. под руководством Нормана Абрамсона. Радиоканал определенного диапазона частот естественным образом является общей средой для всех передатчиков, использующих частоты этого диапазона для кодирования данных. Сеть ALOHA работала по методу случайного доступа, когда каждый узел мог начать передачу пакета в любой момент времени. Если после этого он

не дожидался подтверждения приема в течение определенного тайм-аута, он посылал этот пакет снова. Общим был радиоканал с несущей частотой 400 МГц и полосой 40 кГц, что обеспечивало передачу данных со скоростью 9600 кбит/с [2].

Немного позже Роберт Метклаф повторил идею разделяемой среды уже для проводного варианта технологии LAN. Непрерывный сегмент коаксиального кабеля стал аналогом общей радиосреды. Все компьютеры присоединялись к этому сегменту кабеля по схеме монтажного ИЛИ, поэтому при передаче сигналов одним из передатчиков все приемники получали один и тот же сигнал, как и при использовании радиоволн.

В технологиях Token Ring и FDDI компьютеры также используют разделяемую среду. Физическая топология этих сетей – кольцо, в котором каждый узел соединяется кабелем с двумя соседними узлами (рис. 6.1). Однако эти отрезки кабеля также являются разделяемыми, так как в каждый момент времени только один компьютер может задействовать кольцо для передачи.

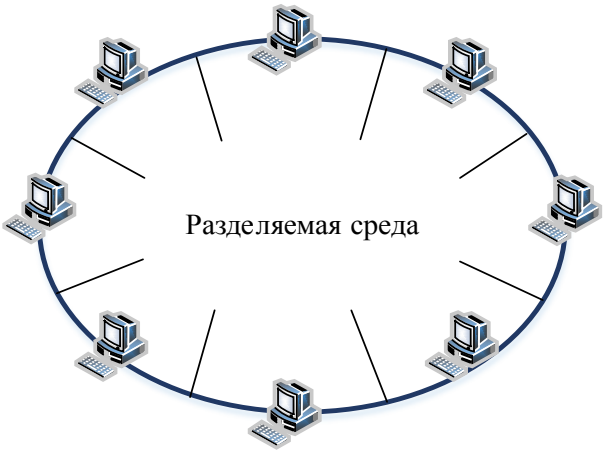


Рис. 6.1. Разделяемая среда
в кольцевых топологиях

Простые стандартные топологии физических связей (звезда у коаксиального кабеля Ethernet и кольцо у Token Ring и FDDI) обеспечивают простоту разделения кабельной среды.

Использование разделяемых сред позволяет *упростить* логику работы узлов сети. Поскольку в каждый момент времени выполняется только одна передача, отпадает необходимость в буферизации кадров в транзитных узлах, и, как следствие, – в самих транзитных узлах. Соответственно отпадает необходимость в сложных процедурах управления потоком и перегрузками.

Основной недостаток разделяемой среды – плохая масштабируемость. Этот недостаток является принципиальным, так как независимо от метода доступа к среде ее пропускная способность делится между всеми узлами

сети. Здесь применимо положение теории очередей: как только коэффициент использования общей среды превышает определенный порог, очереди к среде начинают расти нелинейно и сеть становится практически неработоспособной. Значение порога зависит от метода доступа. Так, в сетях ALOHA это значение является крайне низким – всего около 18 %, в сетях Ethernet – около 30 %, а в сетях Token Ring и FDDI оно возросло до 60–70 %.

Локальные сети, являясь пакетными сетями, используют принцип временно́го мультиплексирования, т. е. разделяют передающую среду во времени. Алгоритм *управления доступом к среде* является одной из важнейших характеристик любой технологии LAN на разделяемой среде, в значительно большей степени определяя ее облик, чем метод кодирования сигнала или формат кадра. В технологии Ethernet в качестве алгоритма разделения среды применяется *метод случайного доступа*. И хотя его трудно назвать совершенным – при росте нагрузки полезная пропускная способность сети резко падает – он благодаря своей простоте стал основой успеха технологии Ethernet в 1980-е гг. Технологии Token Ring и FDDI используют метод маркерного доступа, основанный на передаче от узла к узлу особого кадра – маркера (токена) доступа. При этом только узел, владеющий маркером доступа, имеет право доступа к разделяемому кольцу. Более детерминированный характер доступа технологий Token Ring и FDDI предопределил более эффективное использование разделяемой среды, чем у технологии Ethernet, но одновременно и усложнил оборудование.

Отказ от разделяемой среды привел к исчезновению такого важного компонента технологии локальных сетей, как метод доступа. В принципе, коммутатор локальной сети работает так же, как и обобщенный коммутатор сети с коммутацией пакетов. Поэтому с распространением коммутаторов стали стираться различия между технологиями локальных сетей, так как в сети, где все связи между узлами являются индивидуальными, и коммутируемая версия Ethernet, и коммутируемая версия Token Ring работают весьма схоже, различаются только форматы кадров этих технологий. Работа коммутируемых локальных сетей Ethernet существенно отличается от работы Ethernet на разделяемой среде, так что ее можно считать новой технологией со старым названием. Хотя, с другой стороны, формат кадра Ethernet сохранился, так что это дает формальный повод считать ее той же технологией.

6.1.2. Стандартизация протоколов локальных сетей

Каждая из технологий локальных сетей первоначально появляется как фирменная технология. Так, например, технология Ethernet «появилась на свет» в компании Xerox, а за технологией Token Ring стояла фирма IBM. Первые стандарты технологий локальных сетей также были фирменными,

что было, естественно, не очень удобно как для пользователей, так и для компаний-производителей сетевого оборудования.

Для исправления ситуации в 1980 г. в институте IEEE был организован комитет 802 по стандартизации технологий LAN. Результатом работы комитета IEEE 802 стало принятие семейства стандартов IEEE 802.x, содержащих рекомендации по проектированию нижних уровней локальных сетей. Эти стандарты базировались на обобщении популярных фирменных стандартов, в частности Ethernet и Token Ring.

Комитет IEEE 802 и сегодня является основным международным органом, разрабатывающим стандарты технологий локальных сетей, в том числе стандарты беспроводных локальных сетей на разделяемой среде. Структуру стандартов IEEE 802 иллюстрирует рис. 6.2 [2].

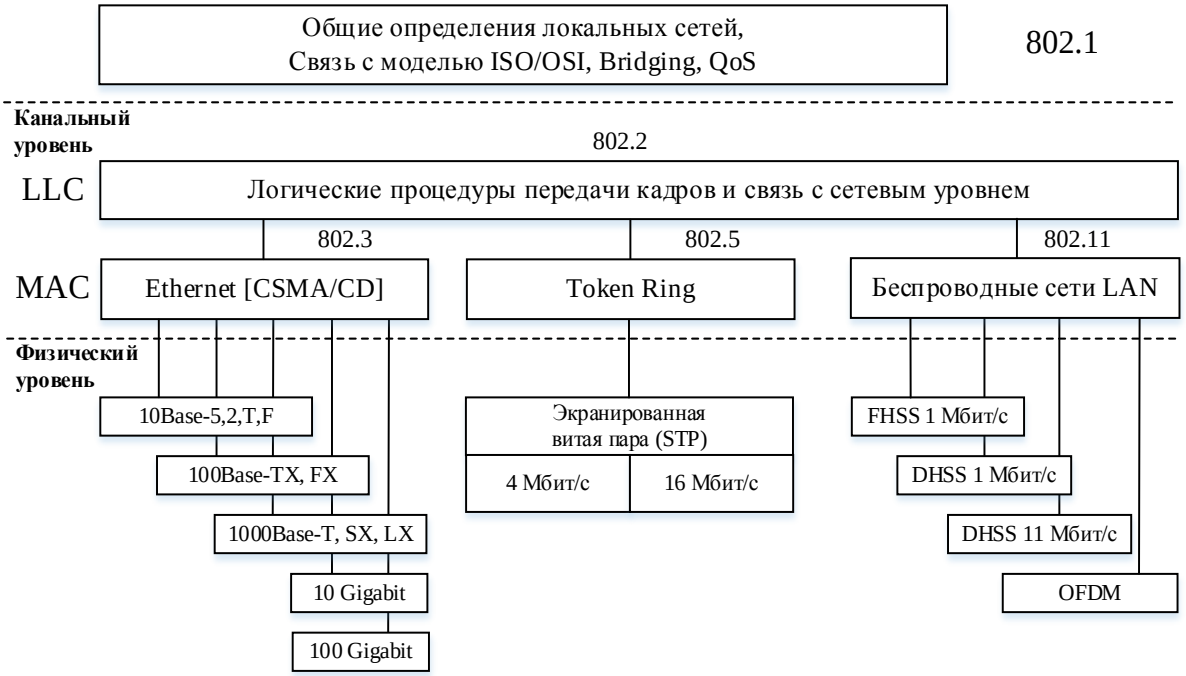


Рис. 6.2. Структура стандартов IEEE 802.x

Стандарты IEEE 802 описывают функции физического и канального уровней модели OSI. Как видно из рис. 6.2, эти стандарты имеют и общие, и индивидуальные для всех технологий части.

Общую группу составляют *стандарты рабочей группы 802.1*. Эти стандарты описывают наиболее высокоуровневые функции локальных сетей. Так, в документах 802.1 даются общие определения локальных сетей и их свойств, показана связь трех уровней модели IEEE 802 с моделью OSI. Набор стандартов, разработанных рабочей группой 802.1, продолжает расти. Например, этот комитет стандартизировал технологию виртуальных локальных сетей, также он занимается стандартизацией Carrier Ethernet.

Каждая из рабочих групп 802.3, 802.4, 802.5 и т. д. была ответственна за стандартизацию конкретной технологии, например, группа 802.3

занималась технологией Ethernet, группа 802.4 – технологией ArcNet, группа 802.5 – технологией Token Ring, группа 802.11 – технологией беспроводных локальных сетей. Сегодня из этих групп активными остались только 802.3 и 802.11 (существуют также и другие активные группы комитета IEEE 802, но они не занимаются технологиями локальных сетей).

Стандарты этих рабочих групп описывают как физический уровень (или несколько возможных физических уровней), так и канальный уровень конкретной технологии (последний включает описание метода доступа, используемого технологией).

Основу стандарта 802.3 составила технология экспериментальной сети Ethernet Network, которую фирма Xerox разработала и реализовала в 1975 г. В 1980 г. фирмы DEC, Intel и Xerox (сокращенно DIX) совместно разработали и опубликовали стандарт Ethernet версии II для сети, построенной на основе коаксиального кабеля. Эту последнюю версию фирменного стандарта Ethernet называют стандартом Ethernet DIX, или Ethernet II. На базе стандарта Ethernet DIX был разработан стандарт IEEE 802.3.

Однако, как видно из рис. 6.2, помимо индивидуальных для каждой технологии уровней существует общий уровень, который был стандартизован рабочей группой 802.2. Появление этого уровня связано с тем, что комитет 802 разделил функции канального уровня модели OSI на два уровня: а) управления логическим каналом (Logical Link Control, LLC); б) управления доступом к среде (Media Access Control, MAC).

Основными функциями уровня MAC являются: обеспечение доступа к разделяемой среде и передача кадров между оконечными узлами посредством физического уровня. Если уровень MAC специфичен для каждой технологии и отражает различия в методах доступа к разделяемой среде, то уровень LLC представляет собой обобщение функций разных технологий по обеспечению передачи кадра с различными требованиями к надежности. Логика образования общего для всех технологий уровня LLC заключается в следующем: после того как узел сети получил доступ к среде в соответствии с алгоритмом, специфическим для конкретной технологии, дальнейшие действия узла по обеспечению передачи кадров не зависят от этой технологии.

Так как в зависимости от требований приложения может понадобиться разная степень надежности передачи, то рабочая группа 802.2 определила три типа услуг:

а) услуга LLC1 – это услуга без установления соединения и без подтверждения получения данных;

б) услуга LLC2 дает пользователю возможность установить логическое соединение перед началом передачи любого блока данных, и если это требуется, выполнить процедуры восстановления после ошибок и упорядочивание потока блоков в рамках установленного соединения;

в) услуга LLC3 – это услуга без установления соединения, но с подтверждением получения данных. Какой из трех режимов уровня LLC будет использован, зависит от требований протокола верхнего уровня.

На практике идея обобщения функций обеспечения надежной передачи кадров в общем уровне LLC не оправдала себя. Технология Ethernet DIX изначально функционировала в наиболее простом дейтаграммном режиме – в результате оборудование Ethernet и после опубликования стандарта IEEE 802.2 продолжало поддерживать только этот режим работы, т. е. LLC1. В то же время оборудование сетей Token Ring, которое изначально поддерживало режимы LLC2 и LLC3, также продолжало поддерживать эти режимы и никогда не поддерживало LLC1. В настоящее время задача обеспечения надежной передачи данных в наибольшей мере возлагается на протокол TCP.

Выводы

Локальные сети на разделяемой среде представляют собой наиболее простой и дешевый в реализации тип локальных сетей. Основной недостаток разделяемых локальных сетей состоит в плохой масштабируемости, так как при увеличении числа узлов сети снижается доля пропускной способности, приходящаяся на каждый узел.

Уровень MAC отвечает за доступ к разделяемой среде и отправку через нее кадров.

Протокол LLC обеспечивает для протоколов верхних уровней нужное качество транспортных услуг, передавая кадры либо дейтаграммным способом, либо с помощью процедур с установлением соединения и восстановлением кадров.

Контрольные вопросы

1. Какую цель и почему ставили перед собой разработчики первых локальных сетей?
2. Приведите достоинства и недостатки разделяемой среды передачи.
3. Опишите структуру стандартов IEEE 802.x. Какие общие и индивидуальные для всех технологий части имеют эти стандарты?
4. Сформулируйте основные функции подуровня MAC.
5. Сформулируйте основные функции подуровня LLC.

6.2. Ethernet 10 Мбит/с на разделяемой среде

6.2.1. Mac-адреса

На уровне MAC, который обеспечивает доступ к среде и передачу кадра, для идентификации узлов сети используются регламентированные стандартом IEEE 802.3 уникальные 6-байтовые адреса, называемые *MAC-адресами*.

Обычно MAC-адрес записывают в виде шести пар шестнадцатеричных цифр, разделенных дефисами или двоеточиями, например, 11-A0-17-3D-BC-01. Каждый сетевой адаптер имеет по крайней мере один MAC-адрес.

Помимо отдельных интерфейсов MAC-адрес может определять группу интерфейсов или даже все интерфейсы сети. Первый (младший[10]) бит старшего байта адреса назначения – это признак того, является адрес индивидуальным или групповым. Если он равен 0, то адрес является *индивидуальным*, т. е. идентифицирует один сетевой интерфейс, а если 1, то *групповым*. Групповой адрес связан только с интерфейсами, сконфигурированными (автоматически или по запросу вышележащего уровня) как член группы, номер которой указан в групповом адресе. Если сетевой интерфейс включен в группу, то наряду с уникальным MAC-адресом с ним ассоциируется еще один адрес – групповой. В частности, если групповой адрес состоит из всех единиц, т. е. имеет шестнадцатеричное представление 0xFFFFFFFFFFFF, он идентифицирует все узлы сети и называется *широковещательным*.

Второй бит старшего байта адреса определяет способ назначения адреса – *централизованный* или *локальный*. Если этот бит равен 0 (что бывает почти всегда в стандартной аппаратуре Ethernet), это говорит о том, что адрес назначен централизованно по правилам IEEE 802. Комитет IEEE распределяет между производителями так называемые *организационно уникальные идентификаторы* (Organizationally Unique Identifier, OUI). Каждый производитель помещает выделенный ему идентификатор в три старших байта адреса (например, идентификатор 0x00000C определяет компанию Cisco) [2].

6.2.2. Форматы кадров технологии Ethernet

Формат кадра Ethernet DIX, который иногда называют кадром Ethernet II по номеру последнего стандарта DIX, представлен на рис. 6.3. Первые два поля заголовка отведены под адреса: а) *DA (Destination Address)* – MAC-адрес узла назначения; б) *SA (Source Adress)* – MAC-адрес узла отправителя.

---

[10] *В стандартах IEEE Ethernet младший бит байта изображается в самой левой* позиции поля, а старший бит – в самой правой. Этот нестандартный способ отображения порядка следования битов в байте соответствует порядку передачи битов в линию связи передатчиков Ethernet (первым передается младший бит). В стандартах других организаций, например, RFC IETF, ITU-T, ISO, используется традиционное представление байта, когда младший бит считается самым правым битом байта, а старший – самым левым. При этом порядок следования байтов остается традиционным. Например, групповой адрес имеющий в нотации IEEE вид 80-00-A7-F0-00-00 будет, скорее всего, отображен анализатором протоколов в традиционном виде как 01-00-5E-0F-00-00.

6 байт	6 байт	2 байта	46–1500 байт	4 байта
DA	*SA*	*T*	Данные	*FCS*

Рис. 6.3. Формат кадра Ethernet DIX (II)

Для доставки кадра достаточно одного адреса – адреса назначения; адрес источника помещается в кадр для того, чтобы узел, получивший кадр, знал, от кого пришел кадр и кому нужно на него ответить. Принятие решения об ответе не входит в компетенцию протокола Ethernet, это дело протоколов верхних уровней, Ethernet лишь выполнит такое действие, если с сетевого уровня поступит соответствующее указание.

Поле T (Type, или *EtherType)* содержит условный код протокола верхнего уровня, данные которого находятся в поле данных кадра, например, шестнадцатеричное значение 08-00 соответствует протоколу IP. Это поле требуется для поддержки функций мультиплексирования и демультиплексирования кадров при взаимодействии с протоколами верхних уровней.

Поле данных может содержать от 46 до 1500 байт. Если длина пользовательских данных меньше 46 байт, то это поле дополняется до минимального размера байтами заполнения.

Поле контрольной последовательности (Frame Check Sequence, FCS) состоит из 4 байт контрольной суммы и вычисляется по алгоритму CRC-32.

Кадр Ethernet DIX (II) не отражает разделения канального уровня Ethernet на уровни MAC и LLC: его поля поддерживают функции обоих уровней, например, интерфейсные функции поля T относятся к функциям LLC, в том время как все остальные поля поддерживают функции MAC.

6.2.3. Доступ к среде и передача данных

Метод доступа, используемый в сетях Ethernet на разделяемой проводной среде, носит название CSMA/CD (Carrier Sense Multiple Access with Collision Detection – прослушивание несущей частоты с множественным доступом и распознаванием коллизий).

Все компьютеры в сети на разделяемой среде имеют возможность немедленно (с учетом задержки распространения сигнала в физической среде) получить данные, которые любой из компьютеров сети начал передавать в общую среду. Говорят, что среда, к которой подключены все станции, работает в режиме *коллективного доступа* (Multiple Access, MA) [2].

Чтобы получить возможность передавать кадр, интерфейс-отправитель должен убедиться, что разделяемая среда свободна. Это достигается прослушиванием основной гармоники сигнала, которая еще называется *несущей частотой* (Carrier Sense, CS). Признаком «незанятости» среды является

отсутствие в ней несущей частоты. Если среда свободна, то узел имеет право начать передачу. В примере, показанном на рис. 6.4, узел 1 обнаружил, что среда свободна, и начал передавать свой кадр. В коаксиальном кабеле сигналы передатчика узла 1 распространяются в обе стороны, так что их получают все узлы сети. Кадр данных всегда сопровождается *преамбулой,* которая нужна для вхождения приемника в побитовую и побайтовую синхронизацию с передатчиком. Все станции, подключенные к кабелю, начинают записывать байты передаваемого кадра в свои внутренние буферы. Первые 6 байт кадра содержат адрес назначения. Та станция, которая узнает собственный адрес в заголовке кадра, продолжает записывать его содержание в свой внутренний буфер, а остальные станции на этом прием кадра прекращают. Станция назначения обрабатывает полученные данные и передает их вверх по своему стеку. Кадр Ethernet содержит не только адрес назначения, но и адрес источника, поэтому станция-получатель знает, кому посылать ответ.

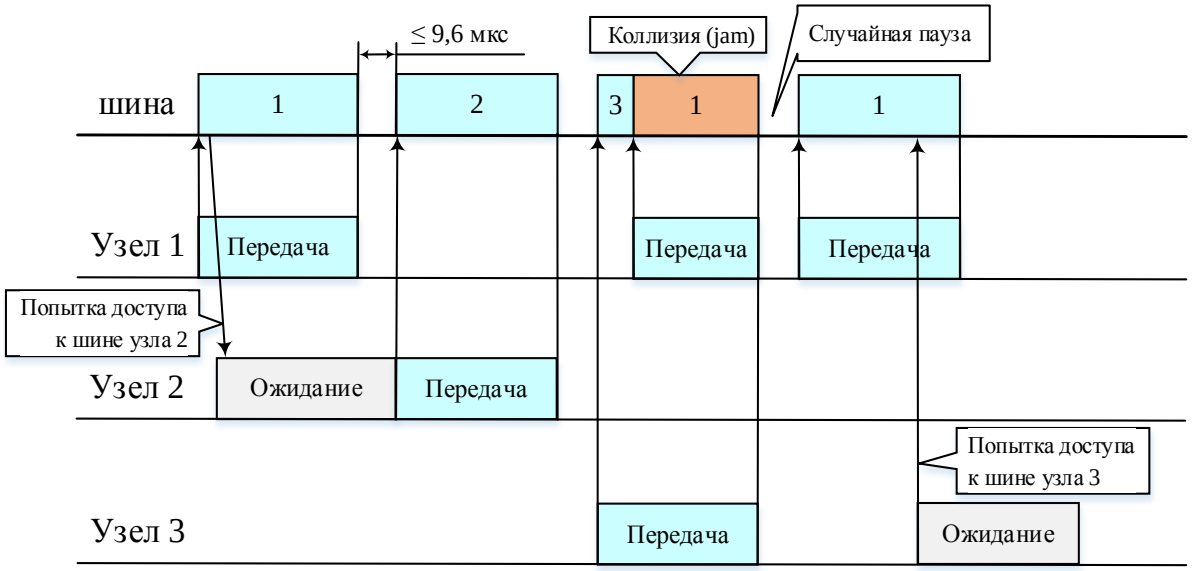


Рис. 6.4. Метод случайного доступа CSMA/CD

Узел 2 во время передачи кадра узлом 1 также пытался начать передачу своего кадра, однако, обнаружив, что среда занята – на ней присутствует несущая частота, – вынужден ждать, пока узел 1 не прекратит передачу кадра.

После окончания передачи кадра все узлы сети обязаны выдержать технологическую паузу, равную *межпакетному интервалу* (Inter Packet Gap, IPG) в 9,6 мкс. Эта пауза нужна для приведения сетевых адаптеров в исходное состояние, а также для предотвращения монопольного захвата среды одной станцией. После окончания технологической паузы узлы имеют право начать передачу своего кадра, так как среда свободна. В приведенном примере узел 2 дождался окончания передачи кадра узлом 1, сделал паузу в 9,6 мкс и начал передачу своего кадра.

6.2.4. Возникновение коллизии

Механизм прослушивания среды и пауза между кадрами не гарантируют исключения ситуации, когда две или более станции одновременно решают, что среда свободна, и начинают передавать свои кадры. Говорят, что при этом происходит *коллизия*, так как содержимое обоих кадров сталкивается в общем кабеле и происходит искажение информации.

Коллизия – это нормальная ситуация в работе сетей Ethernet на разделяемой среде. В примере на рис. 6.5 коллизию породила одновременная передача данных узлами 3 и 1. Для возникновения коллизии не обязательно, чтобы несколько станций начали передачу абсолютно одновременно, такая ситуация маловероятна. Более вероятна ситуация, когда один узел начинает передачу, а через некоторое (короткое) время другой узел, проверив среду и не обнаружив несущую (сигналы первого узла еще не успели до него дойти), начинает передачу своего кадра. Таким образом, возникновение коллизии является следствием распределения узлов сети в пространстве [2].

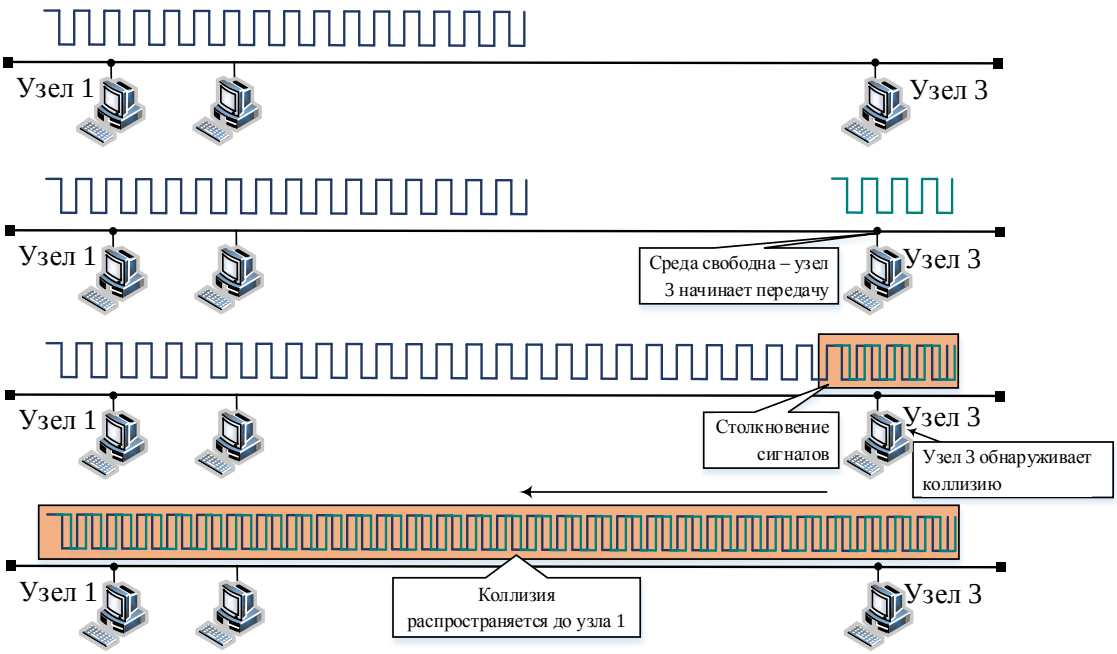


Рис. 6.5. Схема возникновения и распространения коллизии

Чтобы корректно обработать коллизию, все станции одновременно наблюдают за возникающими на кабеле сигналами. Если передаваемые и наблюдаемые сигналы отличаются, то фиксируется факт *обнаружения коллизии* (*Collision Detection*, CD). Для повышения вероятности скорейшего обнаружения коллизии всеми станциями сети станция, которая обнаружила коллизию, прерывает передачу своего кадра и усугубляет коллизию посылкой в сеть специальной *jam-последовательностью* из 32 бит.

После этого обнаружившая коллизию станция обязана прекратить передачу и сделать паузу в течение короткого случайного интервала времени. Затем она может снова предпринять попытку захвата среды и передачи кадра. Случайная пауза выбирается по следующему алгоритму:

$$\text{Пауза} = L \times (\text{интервал отсрочки}).$$

В технологии Ethernet интервал отсрочки выбран равным значению 512 битовых интервалов. Битовый интервал соответствует времени между появлением двух последовательных битов данных на кабеле; для скорости 10 Мбит/с величина битового интервала равна 0,1 мкс, или 100 нс. L представляет собой целое число, выбранное с равной вероятностью из диапазона $[0, 2^N]$, где N – номер повторной попытки передачи кадра: 1, 2, …, 10. После 10-й попытки интервал, из которого выбирается пауза, не увеличивается. *Таким образом, случайная пауза в Ethernet может быть от 0 до 52,4 мс (51,2 мкс* $\times 2^{10}$*).*

Если 16 попыток передачи кадра вызывают коллизию, то передатчик должен прекратить попытки и отбросить этот кадр. Этот алгоритм носит название *усеченного экспоненциального двоичного алгоритма отсрочки*.

Поведение сети Ethernet при значительной нагрузке, когда коэффициент использования среды растет и начинает приближаться к 1, в целом соответствует графикам, которые были приведены при анализе модели теории очередей. Администраторы сетей Ethernet на разделяемой среде руководствовались простым эмпирическим правилом – коэффициент использования среды не должен превышать 30 %. Для поддержки чувствительного к задержкам трафика сети Ethernet (на разделяемой среде) могут применять только один метод поддержания QoS – *недогруженный режим работы*.

6.2.5. Время оборота и распознавание коллизий

Надежное распознавание коллизий всеми станциями является необходимым условием корректной работы сети Ethernet. Если какая-либо передающая станция не распознает коллизию и решит, что кадр данных передан ею верно, этот кадр будет утерян. Из-за наложения сигналов при коллизии информация кадра исказится, и он будет отбракован принимающей станцией. Скорее всего, недошедшие до получателя данные будут повторно переданы протоколом верхнего уровня, либо протоколом LLC, если он работает в режиме LLC2. Однако повторная передача сообщения протоколами верхних уровней произойдет гораздо позже (иногда по прошествии нескольких секунд), чем повторная передача средствами сети Ethernet, работающей с микросекундными интервалами. Поэтому если коллизии не будут надежно распознаваться узлами сети Ethernet, то это приведет к заметному сни-

жению полезной пропускной способности сети. Для надежного распознавания коллизий должно выполняться следующее соотношение:

$$T_{\min} \geq \text{RTT},$$

где $T_{\min}$ – время передачи кадра минимальной длины, а RTT – время оборота, т. е. время, за которое сигнал, посланный некоторой станцией сети, доходит до точки коллизии, а затем возвращается к станции-отправителю в уже искаженной коллизией форме. В худшем случае сигнал должен пройти дважды между наиболее удаленными друг от друга станциями сети.

При выполнении этого условия передающая станция должна успеть обнаружить коллизию, которую вызвал переданный ею кадр, еще до того, как она закончит передачу этого кадра. Выполнение этого условия зависит, с одной стороны, от минимальной длины кадра и скорости передачи данных протокола, с другой – от длины кабельной системы сети и скорости распространения сигнала в кабеле. *Все параметры протокола Ethernet, в том числе минимальный размер кадра, подобраны таким образом, чтобы при нормальной работе сети коллизии четко распознавались*.

Так, стандарт Ethernet определяет минимальную длину поля данных кадра в 46 байт (что вместе со служебными полями дает минимальную длину кадра 64 байт, а вместе с преамбулой – 72 байт, или 576 бит). Отсюда может быть вычислено ограничение на расстояние между станциями. В стандарте Ethernet 10 Мбит/с время передачи кадра минимальной длины равно 575 битовых интервалов, следовательно, время оборота должно быть меньше 57,5 мкс. Расстояние, которое сигнал может пройти за это время, зависит от типа кабеля и для толстого коаксиального кабеля равно примерно 13 280 метров. Учитывая, что за время 57,5 мкс сигнал должен пройти по линии связи дважды, расстояние между двумя узлами не должно быть выше 6635 метров. В стандарте величина этого расстояния выбрана равной 2500 м, что существенно меньше. Это объясняется тем, что повторители, которые нужны для соединения 5 сегментов кабеля, вносят задержки в распространение сигнала.

Описанные соображения объясняют выбор минимальной длины поля данных кадра в 46 байт. Уменьшение этого значения до 0 привело бы к значительному сокращению максимальной длины сети.

Из требования $T_{\min} \geq \text{RTT}$ следует: чем выше скорость протокола, тем меньше должна быть максимальная длина сети. Поэтому для Ethernet на разделяемой среде при скорости в 100 Мбит/с максимальная длина сети пропорционально уменьшается до 250 м, а при скорости 1 Гбит/с – до 25 м. Эта зависимость, наряду с резким ростом задержек при повышении загрузки, говорит еще об одном недостатке метода CSMA/CD.

6.2.6. Физические стандарты 10m Ethernet

При первоначальной стандартизации Ethernet рабочей группой IEEE 802.3 был выбран вариант Ethernet на «толстом» коаксиальном кабеле 10Base-5. Число 10 в этом названии обозначает номинальную битовую скорость передачи 10 Мбит/с, а слово «Base» – метод передачи на одной базовой частоте 10 МГц. Последний символ обозначает тип кабеля, в данном случае 5 отражает то, что диаметр «толстого» коаксиала равен 0,5 дюйма. Принятый подход к обозначению типа физического уровня Ethernet сохранился до настоящего времени, только вместо диаметра коаксиального кабеля в современных стандартах кодируется тип кабеля (например, 1000Base-T определяет спецификацию для витой пары) или же способ кодирования.

Затем сети Ethernet на «толстом» коаксиальном кабеле были вытеснены сетями на более «тонком» коаксиале (диаметром 0,25 дюйма, что отражает название 10Base-2 этого стандарта). Однако сети Ethernet на коаксиальном кабеле обладали одним существенным недостатком, а именно отсутствием оперативной информации о состоянии кабеля и сложностью нахождения места его повреждения. Альтернатива появилась в середине 1980-х гг., когда благодаря использованию витой пары и повторителей сети Ethernet стали гораздо более ремонтопригодными.

К этому времени телефонные компании уже достаточно давно применяли многопарный кабель на основе неэкранированной витой пары для подключения телефонных аппаратов внутри зданий. Идея приспособить этот популярный вид кабеля для локальных сетей оказалась очень плодотворной, так как многие здания уже были оснащены нужной кабельной системой. Оставалось разработать способ подключения сетевых адаптеров и прочего коммуникационного оборудования к витой паре таким образом, чтобы изменения в сетевых адаптерах и программном обеспечении сетевых операционных систем были минимальными по сравнению с сетями Ethernet на коаксиале. Эта попытка оказалась успешной – переход на витую пару требует только замены приемника и передатчика сетевого адаптера, а метод доступа и все протоколы канального уровня остаются теми же, что и в сетях Ethernet на коаксиале. Результатом стал стандарт 10Base-T (T – от Twisted Pair). Правда, для соединения узлов в сеть теперь обязательно требуется устройство – *многопортовый повторитель* Ethernet на витой паре.

Устройство такого повторителя схематично изображено на рис. 6.6. Каждый сетевой адаптер соединяется с повторителем двумя витыми парами. Одна витая пара требуется для передачи данных от станции к повторителю (выход *Tx* сетевого адаптера), другая – для передачи данных от повторителя к станции (вход *Rx* сетевого адаптера). Повторитель побитно принимает сигналы от одного из конечных узлов и синхронно передает их на все свои

остальные порты, исключая тот, с которого поступили сигналы, одновременно улучшая их электрические характеристики [2].

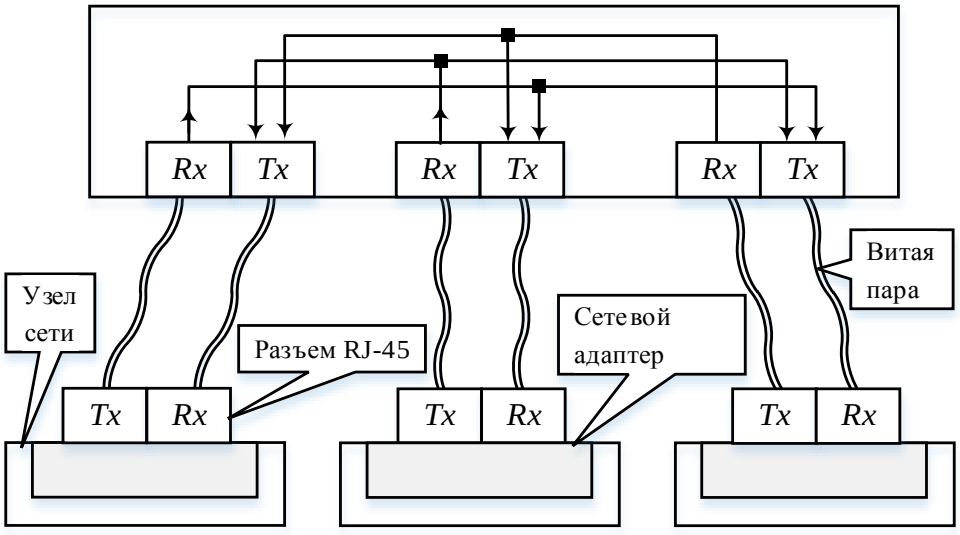


Рис. 6.6. Повторитель Ethernet на витой паре

Многопортовый повторитель часто называют *концентратором*, или *хабом* (от английского *hub* – центр, ступица колеса), так как в нем сконцентрированы соединения со всеми конечными узлами сети. Фактически хаб имитирует сеть на коаксиальном кабеле в том отношении, что физически отдельные отрезки кабеля на витой паре логически все равно представляют единую разделяемую среду. Правила доступа по алгоритму CSMA/CD сохраняются.

При создании сети Ethernet на витой паре с большим числом конечных узлов хабы можно соединять друг с другом иерархическим способом, образуя древовидную структуру (рис. 6.7). Добавление каждого хаба изменяет физическую структуру, но оставляет без изменения логическую структуру сети. То есть независимо от числа хабов в сети сохраняется одна общая для всех интерфейсов разделяемая среда, так что передача кадра с любого интерфейса блокирует передатчики всех остальных интерфейсов.

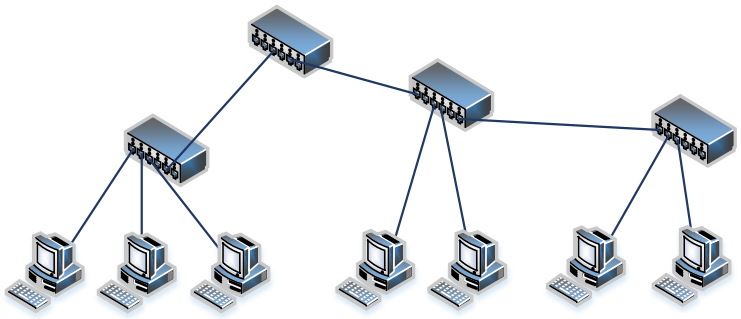


Рис. 6.7. Иерархическое соединение хабов

Физическая структуризация сетей на витой паре повышает надежность и упрощает обслуживание сети, поскольку в этом случае появляется возможность контролировать состояние и локализовать отказы отдельных кабельных отрезков, подключающих конечные узлы к концентраторам.

Для контроля целостности физического соединения между двумя портами 10Base-T введен тест целостности соединения (Link Integrity Test, LIT). Эта процедура заключается в том, что в те периоды, когда порт не посылает или не получает кадры, он посылает своему соседу импульсы 100 нс каждые 16 мс. Если порт принимает такие импульсы от своего соседа, то он считает соединение работоспособным и индицирует это зеленым светом светодиода.

6.2.7. Производительность сети 10m Ethernet

Производительность сети зависит от скорости передачи кадров по линиям связи и скорости обработки этих кадров коммуникационными устройствами, передающими кадры между своими портами, к которым эти линии связи подключены. Скорость передачи кадров по линиям связи зависит от используемых протоколов физического и канального уровней, например, Ethernet на 10 Мбит/с, Ethernet на 100 Мбит/с, Token Ring или FDDI.

Скорость, с которой протокол передает биты по линии связи, называется *номинальной скоростью протокола*.

Скорость обработки кадров коммуникационным устройством зависит от производительности его процессоров, внутренней архитектуры и других параметров. Скорость коммуникационного устройства должна соответствовать скорости работы линии. Если она меньше скорости работы линии, то кадры будут стоять в очередях и отбрасываться при переполнении последних. В то же время нет смысла применять устройство, которое в сотни раз производительнее, чем того требует скорость подключаемых к нему линий.

Для оценки производительности коммуникационных устройств Ethernet необходимо оценить производительность сегмента Ethernet, но не в битах в секунду (ее мы знаем – это 10 Мбит/с), а в кадрах в секунду, так как именно этот показатель помогает оценить требования к производительности коммуникационных устройств. Это объясняется тем, что на обработку каждого кадра, независимо от его длины мост, коммутатор или маршрутизатор тратят примерно равное время, которое уходит на просмотр таблицы продвижения пакета, формирование нового кадра (для маршрутизатора) и т. п.

При постоянной битовой скорости количество кадров, поступающих на коммуникационное устройство в единицу времени, является максимальным при их минимальной длине. Поэтому для коммуникационного оборудования наиболее тяжелым режимом является обработка кадров минимальной длины.

Рассчитаем максимальную производительность сегмента Ethernet в числе переданных кадров (пакетов) минимальной длины в секунду. Для расчета максимального количества кадров минимальной длины, проходящих по сегменту Ethernet, вспомним, что подсчитанное нами ранее время, затрачиваемое на передачу кадра минимальной длины (576 бит), составляет 57,5 мкс. Прибавив межкадровый интервал в 9,6 мкс, получаем, что период следования кадров минимальной длины составляет 67,1 мкс. Отсюда *максимально возможная пропускная способность сегмента Ethernet составляет 14880 кадр/с* (рис. 6.8). Наличие в сегменте нескольких узлов снижает эту величину за счет ожидания доступа к среде, а также из-за коллизий [2].

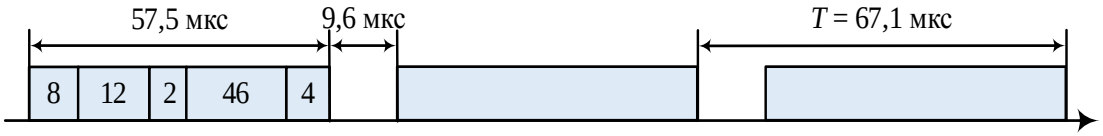


Рис. 6.8. К расчету пропускной способности протокола Ethernet

Кадры максимальной длины в Ethernet имеют поле данных в 1500 байт, что вместе со служебной информацией дает 1518 байт, а с преамбулой составляет 1526 байт, или 12208 бит. *Максимально возможная пропускная способность сегмента Ethernet для кадров максимальной длины составляет 813 кадр/с*. Очевидно, что при работе с большими кадрами нагрузка на мосты, коммутаторы и маршрутизаторы довольно ощутимо снижается.

Теперь рассчитаем, какой максимально полезной пропускной способностью, измеряемой в битах в секунду, обладают сегменты Ethernet при использовании кадров разного размера.

Полезной пропускной способностью протокола называется максимальная скорость передачи пользовательских данных, которые переносятся полем данных кадра. Эта пропускная способность всегда меньше номинальной битовой скорости протокола Ethernet за счет нескольких факторов: а) служебной информации кадра; б) межкадровых интервалов (IPG); в) ожидания доступа к среде.

Для кадров минимальной длины полезная пропускная способность равна B = 14880 × 46 × 8 = 5,48 Мбит/с. Это несколько меньше, чем 10 Мбит/с, но следует учесть, что кадры минимальной длины используются в основном для передачи квитанций, так что к передаче собственно данных файлов эта скорость имеет небольшое отношение.

Для кадров максимальной длины полезная пропускная способность равна $B = 813 \times 1500 \times 8 = 9{,}76$ Мбит/с. При использовании кадров среднего размера с полем данных в 512 байт пропускная способность составляет 9,29 Мбит/с. В двух последних случаях пропускная способность протокола

оказалась достаточно близкой к предельной пропускной способности в 10 Мбит/с, однако следует учесть, что при расчете мы предполагали, что двум взаимодействующим станциям «не мешают» никакие другие станции сети, т. е. отсутствуют коллизии и ожидание доступа.

Таким образом, при отсутствии коллизий коэффициент использования сети зависит от размера поля данных кадра и имеет максимальное значение 0,976 при передаче кадров максимальной длины. Эти расчеты нетрудно скорректировать и для других битовых скоростей Ethernet, учитывая соответствующий масштабный коэффициент $n \times 10$ (он справедлив и для межкадрового интервала).

Выводы

В технологии Ethernet на разделяемой среде применяется случайный метод доступа CSMA/CD, который очень прост в реализации.

Коллизия – это ситуация, когда две станции одновременно пытаются передать кадр данных через общую среду. Наличие коллизий – неотъемлемое свойство сетей Ethernet, являющееся следствием принятого случайного метода доступа.

В зависимости от типа физической среды стандарт IEEE 802.3 определяет различные спецификации Ethernet со скоростью 10 Мбит/с: 10Base-5, 10Base-2, 10Base-T, 10Base-FL, 10Base-FB.

Контрольные вопросы

1. Выберите утверждения, корректно описывающие особенности метода доступа технологии Ethernet? Варианты ответов:

а) узел обязан «прослушать» разделяемую среду;

б) узел может передать свой кадр в разделяемую среду в любой момент времени независимо от того, занята среда или нет;

в) узел ожидает подтверждения приема переданного кадра от узла назначения в течение некоторого времени, а по истечении этого времени повторяет передачу.

2. В чем состоят функции преамбулы кадра в стандарте Ethernet?

3. Что такое коллизия в сетях Ethernet?

4. Зачем в технологии Ethernet введен межкадровый интервал?

5. Почему стандарт 10Base-T вытеснил стандарты Ethernet на коаксиальном кабеле?

6. Как длина кадра влияет на работу сети? Какие проблемы связаны со слишком длинными кадрами? В чем состоит неэффективность коротких кадров?

7. Как коэффициент использования влияет на производительность сети Ethernet?

8. Чем объясняется, то что минимальный размер поля данных кадра Ethernet выбран равным 46 байт? Варианты ответов:

а) для предотвращения монопольного захвата среды узлом;

б) для устойчивого распознавания коллизий;

в) для сокращения накладных расходов.

9. К какому типу относится MAC-адрес 01:80:C2:00:00:08? Варианты ответов:

а) групповой;

б) индивидуальный;

в) локальный;

г) централизованный.

10. Как скорость передачи данных технологии Ethernet на разделяемой среде влияет на максимальный диаметр сети? Варианты ответов:

а) чем выше скорость передачи, тем меньше максимальный диаметр сети;

б) чем выше скорость передачи, тем больше максимальный диаметр сети;

в) не влияет.

6.3. Беспроводные локальные сети IEEE 802.11

6.3.1. Особенности беспроводных локальных сетей

Беспроводные локальные сети (БЛС, Wireless Local Area Network, WLAN) в некоторых случаях являются предпочтительным по сравнению с проводными сетями решением, а иногда и просто единственно возможным.

Преимущество беспроводных локальных сетей очевидно – их проще и дешевле разворачивать и модифицировать, так как вся громоздкая кабельная инфраструктура оказывается излишней. Еще одно преимущество – обеспечение мобильности пользователей. Сегодня беспроводные локальные технологии успешно применяются во многих типах сетей: в домашних сетях, в сетях аэропортов, вокзалов, кафе и других публичных местах, для организации временных сетей на различных конференциях и совещаниях, в сетях исторических зданий с уникальной архитектурой, исключающей возможность прокладки кабелей, а также как городские сети доступа в Интернет.

Однако за эти преимущества беспроводные сети расплачиваются перечнем проблем, которые несет с собой неустойчивая беспроводная среда. Помехи от бытовых приборов и других телекоммуникационных систем, атмосферные помехи и отражения сигнала создают серьезные трудности для надежного приема информации. Локальные сети – это прежде всего сети зданий, а распространение сигнала внутри здания еще сложнее, чем вне его.

Методы *расширения спектра* помогают снизить влияние помех на полезный сигнал, кроме того, в беспроводных сетях широко используется *прямая коррекция ошибок* (Forward Error Correction, FEC) и протоколы с повторной передачей потерянных кадров.

Неравномерное распределение интенсивности сигнала приводит не только к битовым ошибкам передаваемой информации, но и к неопределенности зоны покрытия беспроводной локальной сети. В проводных локальных сетях такой проблемы нет; те и только те устройства, которые подключены к кабельной системе здания или кампуса, получают сигналы и участвуют в работе LAN. Беспроводная локальная сеть не имеет точной области покрытия. Покрытие в форме шестиугольника или круга является абстракцией. В действительности сигнал может быть ослаблен так, что устройства, находящиеся в предполагаемой зоне покрытия, вообще не могут его принимать.

С течением времени ситуация с распределением сигнала может измениться вместе с изменением состава WLAN. По этой причине даже технологии, рассчитанные на фиксированные (не мобильные) узлы сети, должны учитывать, что беспроводная локальная сеть является неполносвязной. Даже если считать, что сигнал распространяется идеально во все стороны, образованию полносвязной топологии может мешать то, что радиосигнал затухает пропорционально квадрату расстояния от источника (в случае свободного пространства). Поэтому при отсутствии базовой станции некоторые пары узлов не смогут взаимодействовать из-за того, что расположены за пределами зоны покрытия передатчиков партнера.

В примере на рис. 6.9, а показана такая фрагментированная локальная сеть. Неполносвязность беспроводной сети порождает проблему доступа к разделяемой среде, известную под названием *скрытого терминала*. Проблема возникает в том случае, когда два узла находятся вне зон досягаемости друг друга (узлы *A* и *C* на рис. 6.9, а), но существует третий узел *B*, который принимает сигналы как от *A*, так и от *C*. Предположим, что в радиосети используется традиционный метод доступа, основанный на прослушивании несущей, например, CSMA/CD. В данном случае коллизии будут возникать значительно чаще, чем в проводных сетях. Пусть, например, узел *B* занят обменом с узлом *A*. Узлу *C* сложно определить, что среда занята, он может посчитать ее свободной и начать передавать свой кадр. В результате сигналы в районе узла *B* исказятся, т. е. произойдет коллизия, вероятность возникновения которой в проводной сети была бы неизмеримо ниже [2].

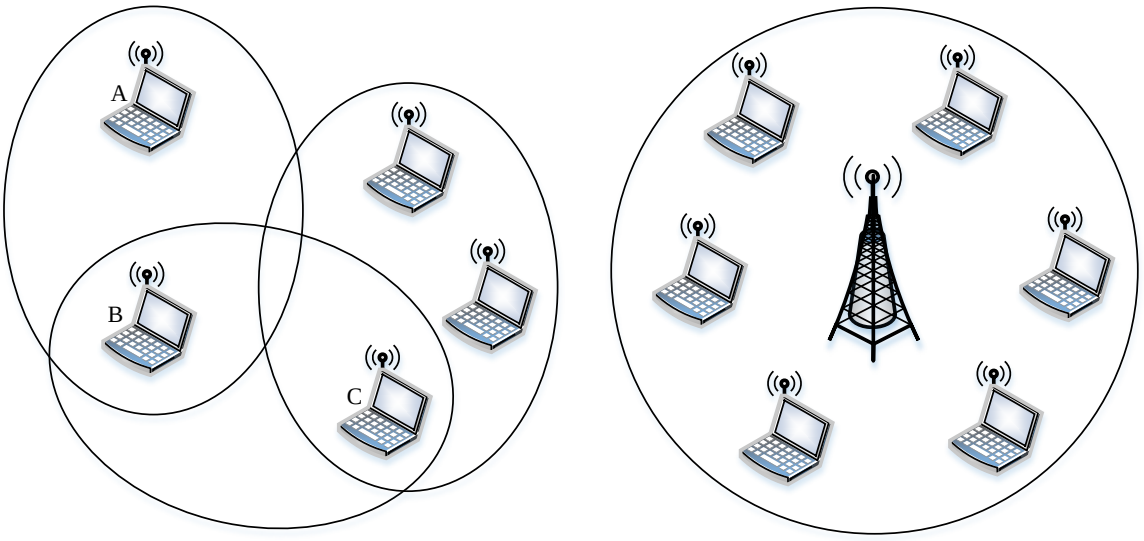


Рис. 6.9. Связность беспроводной локальной сети:
а) специализированная беспроводная сеть;
б) беспроводная сеть с базовой станцией

Распознавание коллизий затруднено в радиосети еще и потому, что сигнал собственного передатчика существенно подавляет сигнал удаленного передатчика и распознать искажение сигнала чаще всего невозможно.

В методах доступа, применяемых в радиосетях, отказываются не только от прослушивания несущей, но и от распознавания коллизий. Вместо этого используют предотвращение коллизий, включая *методы опроса*.

Применение базовой станции (БС) может улучшить связность сети (рис. 6.9, б). БС обычно обладает большей мощностью, а ее антенна устанавливается так, чтобы более равномерно покрывать нужную территорию. В результате все узлы беспроводной локальной сети получают возможность обмениваться данными с БС, которая транзитом передает данные между узлами.

Далее будет рассмотрен стандарт беспроводных локальных сетей – IEEE 802.11. Сети IEEE 802.11 также известны под названием WiFi[11] – по имени консорциума Wi-Fi Alliance, который занимается вопросами совместимости и сертификации оборудования стандартов IEEE 802.11.

6.3.2. Топологии локальных сетей стандарта IEEE 802.11

Стандарт 802.11 поддерживает два типа топологий локальных сетей: с базовым и расширенным наборами услуг.

Сеть с базовым набором услуг (Basic Service Set, BSS) образуется отдельными станциями, базовая станция отсутствует, узлы взаимодействуют друг с другом непосредственно (рис. 6.10). Для того чтобы войти в сеть BSS, станция должна выполнить процедуру присоединения.

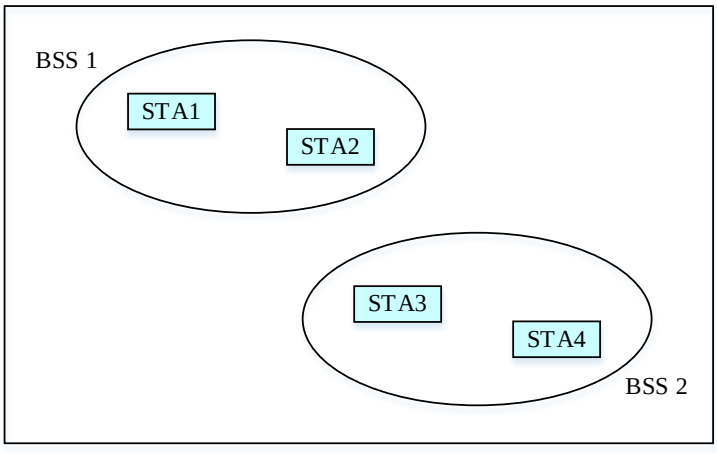


Рис. 6.10. Сети с базовым набором услуг

Станции могут передавать данные: а) непосредственно друг другу в пределах одной сети BSS; б) в пределах одной сети BSS транзитом через

[11] Wi-Fi является сокращением от Wireless Fidelity – «беспроводная точность».

точку доступа; в) между разными сетями BSS через две точки доступа и распределенную систему; г) между сетью BSS и проводной локальной сетью через точку доступа, распределенную систему и портал [2].

В сетях с *расширенным набором услуг* некоторые станции сети являются базовыми, или, в терминологии 802.11, *точками доступа* (Access Point, AP). Станция, которая выполняет функции AP, является членом какой-нибудь сети BSS (рис. 6.11). Все базовые станции сети связаны между собой с помощью распределенной системы (Distribution System, DS), в качестве которой может использоваться та же среда (т. е. радио), что и среда взаимодействия между станциями, или же отличная от нее, например, проводная. Точки доступа вместе с распределенной системой поддерживают *службу распределенной системы* (Distribution System Service, DSS). Задачей DSS является передача пакетов между станциями, которые по каким-то причинам не могут или не хотят взаимодействовать между собой непосредственно.

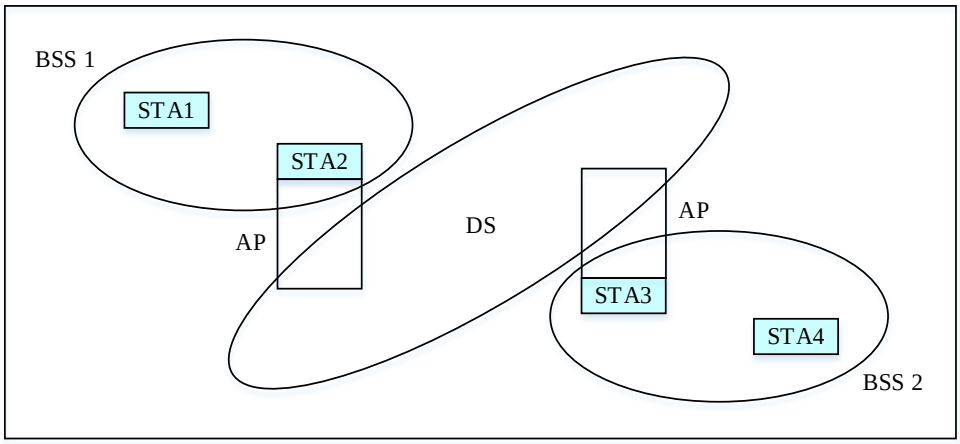


Рис. 6.11. Сеть с расширенным набором услуг

Сеть с расширенным набором услуг (Extended Service Set, ESS) состоит из нескольких сетей BSS, объединенных распределенной средой.

Сеть ESS обеспечивает станциям мобильность – они могут переходить из одной сети BSS в другую. Эти перемещения обеспечиваются функциями уровня MAC рабочих и базовых станций, поэтому они совершенно прозрачны для уровня LLC. Сеть ESS может также взаимодействовать с проводной локальной сетью.

6.3.3. Стек протоколов IEEE 802.11

Стек протоколов стандарта IEEE 802.11 соответствует общей структуре стандартов комитета 802, т. е. состоит из физического уровня и уровня MAC, поверх которых работает уровень LLC. Как и у всех технологий семейства 802, технология 802.11 определяется нижними двумя уровнями,

т. е. физическим уровнем и уровнем MAC, а уровень LLC выполняет свои стандартные функции, общие для всех технологий LAN. Структура стека протоколов IEEE 802.11 показана на рис. 6.12 [2].

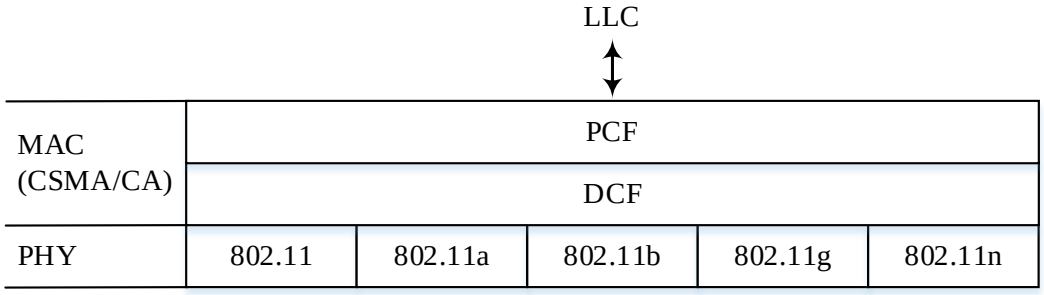


Рис. 6.12. Стек протоколов IEEE 802.11

Уровень MAC выполняет в беспроводных сетях больше функций, чем в проводных. Функции уровня MAC в стандарте IEEE 802.11 включают: а) доступ к разделяемой среде; б) обеспечение мобильности станций при наличии нескольких базовых станций; в) обеспечение безопасности, эквивалентной безопасности проводных локальных сетей.

В сетях 802.11 уровень MAC поддерживает два режима доступа к разделяемой среде: *распределенный режим* (Dsitributed Coordination Function, *DCF*) и *централизованный режим* (Point Coord fination Function, *PCF*). Режим PCF применяется в тех случаях, когда необходимо приоритезировать чувствительный к задержкам трафик.

На физическом уровне существует несколько вариантов спецификаций, которые отличаются используемым частотным диапазоном, методом кодирования и скоростью передачи данных. Все варианты физического уровня работают с одним и тем же алгоритмом уровня MAC, но некоторые временны́е параметры уровня MAC зависят от используемого физического уровня.

6.3.4. Распределенный режим доступа

Рассмотрим сначала, как обеспечивается доступ в распределенном режиме DCF. В этом режиме реализуется метод CSMA/CA (Carrier Sense Multiple Access with Collision Avoidance – метод прослушивания несущей частоты с множественным доступом и предотвращением коллизий). Вместо неэффективного в беспроводных сетях прямого распознавания коллизий по методу CSMA/CD здесь они выявляются косвенно. Для этого каждый переданный кадр должен подтверждаться *кадром положительной квитанции*, посылаемым станцией назначения. Если же по истечении оговоренного тайм-аута квитанция не поступает, станция-отправитель считает, что произошла коллизия.

Режим DCF требует синхронизации станций. Она достигается за счет того, что временны́е интервалы начинают отсчитываться от момента окончания передачи очередного кадра (рис. 6.13). Это не требует каких-либо специальных синхронизирующих сигналов [2].

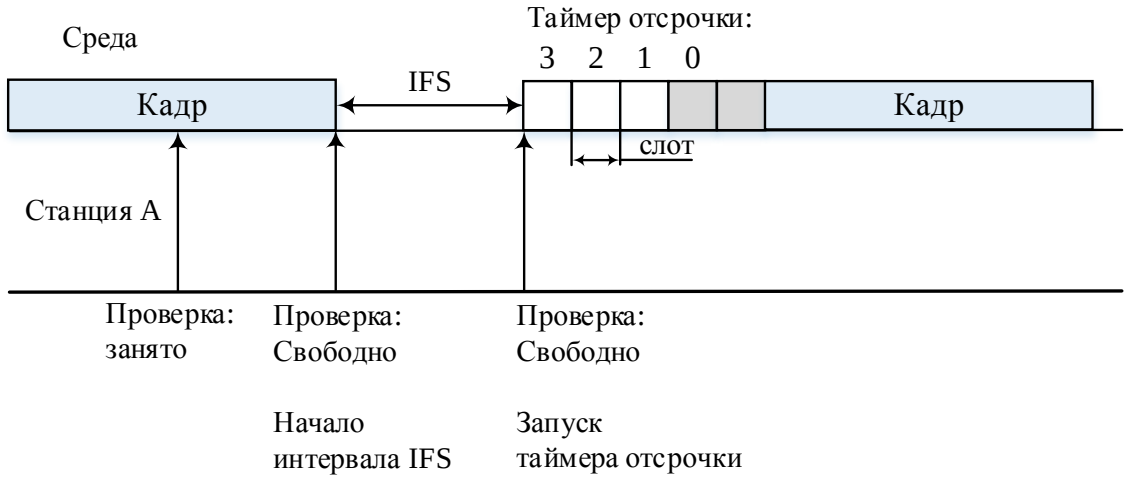


Рис. 6.13. Распределенный режим доступа (DCF)

Станция, которая хочет передать кадр, обязана предварительно прослушать среду. Как только она фиксирует окончание передачи кадра, она обязана отсчитать интервал времени, равный межкадровому интервалу (Inter-Frame Space, IFS). Если после истечения IFS среда все еще свободна, то начинается отсчет слотов фиксированной длительности. Кадр можно начать передавать только в начале какого-либо из слотов при условии, что среда свободна. Станция выбирает для передачи слот на основании *усеченного экспоненциального двоичного алгоритма отсрочки*, аналогичного используемому в методе CSMA/CD. Номер слота выбирается как случайное целое число, равномерно распределенное в интервале [0, CW], где CW означает Contention Window (*конкурентное окно*).

Рассмотрим этот довольно простой метод доступа на примере рис. 6.13. Пусть станция *A* на основании усеченного экспоненциального двоичного алгоритма отсрочки выбрала для передачи слот 3. При этом она присваивает *таймеру отсрочки* значение 3 и начинает проверять состояние среды в начале каждого слота. Если среда свободна, то из значения таймера отсрочки вычитается 1, и если результат равен нулю, то начинается передача кадра. *Таким образом обеспечивается условие незанятости всех слотов, включая выбранный. Это условие является необходимым для начала передачи.*

Если же в начале какого-нибудь слота среда оказывается занятой, то вычитания единицы не происходит и таймер «замораживается». В этом случае станция начинает новый цикл доступа к среде. Как и в предыдущем цикле, станция следит за средой и при ее освобождении делает паузу в те-

чение межкадрового интервала. Если среда осталась свободной, то станция *использует значение «замороженного» таймера в качестве номера слота* и выполняет описанную процедуру проверки свободных слотов с вычитанием единиц, начиная с замороженного значения таймера отсрочки.

Размер слота выбирается так, чтобы он превосходил время распространения сигнала между любыми двумя станциями сети плюс время, затрачиваемое станцией на распознавание занятости среды. Размер слота зависит от метода кодирования сигнала. Так, например, для метода FHSS размер слота равен 28 мкс, а для DSSS – 1 мкс. Если такое условие соблюдается, то каждая станция сумеет правильно распознать начало передачи кадра при прослушивании слотов, предшествующих слоту, выбранному ею для передачи. Это, в свою очередь, означает, что *коллизия может случиться только в том случае, когда несколько станций выбирают один и тот же слот для передачи*.

В этом случае кадры искажаются и квитанции подтверждения приема от станций назначения не приходят. Не получив в течение определенного времени квитанцию, отправители фиксируют факт коллизии и пытаются передать свои кадры снова. При каждой повторной неудачной попытке передачи кадра интервал [0, CW], из которого выбирается номер слота, удваивается. Если, например, начальный размер окна выбран равным 8 (т. е. CW = 7), то после первой коллизии размер окна должен быть 16 (CW = 15), после второй последовательной коллизии – 32, и т. д. Начальное значение CW в соответствии со стандартом 802.11 должно выбираться в зависимости от типа физического уровня, используемого в беспроводной локальной сети.

Как и в методе CSMA/CD, в данном методе количество неудачных попыток передачи одного кадра ограничено, но стандарт 802.11 не дает точного значения верхнего предела. Когда верхний предел в N попыток достигнут, то кадр отбрасывается, а счетчик последовательных коллизий устанавливается в нуль. Этот счетчик также устанавливается в нуль, если кадр после некоторого количества неудачных попыток все же передается успешно.

В режиме DCF применяются меры для *устранения эффекта скрытого терминала*. Для этого станция, которая начинает передачу, вместо кадра данных сначала посылает станции назначения короткий служебный кадр RTS (Request To Send – запрос на передачу). На этот запрос станция назначения должна ответить служебным кадром CTS (Clear To Send – свободна для передачи), после чего станция-отправитель посылает кадр данных. Кадр CTS должен оповестить о захвате среды те станции, которые находятся вне зоны сигнала станции-отправителя, но в зоне досягаемости станции-получателя, т. е. являются скрытыми терминалами для станции-отправителя.

6.3.5. Централизованный режим доступа

В том случае, когда в сети BSS имеется станция, выполняющая функции точки доступа, может применяться также централизованный режим доступа (PCF), обеспечивающий приоритетное обслуживание трафика. В этом случае говорят, что точка доступа играет роль арбитра среды. Режим PCF в сетях 802.11 сосуществует с режимом DCF. Оба режима координируются с помощью трех типов межкадровых интервалов (рис. 6.14) [2].

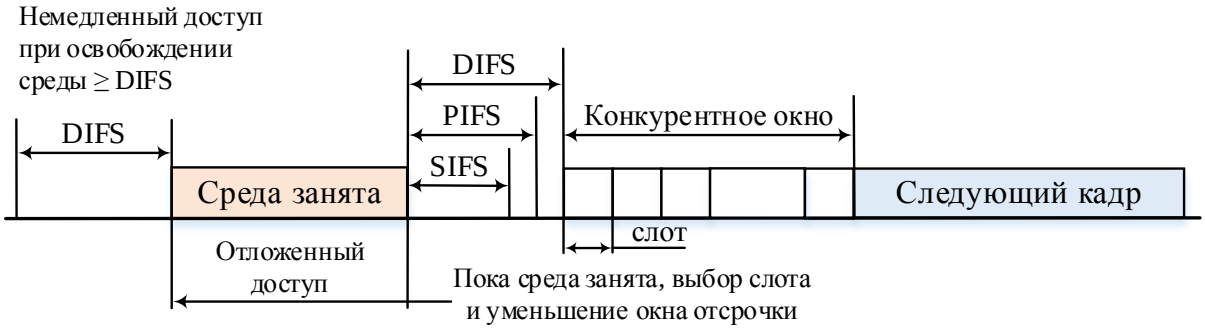


Рис. 6.14. Сосуществование режимов PCF и DCF

После освобождения среды каждая станция отсчитывает время простоя среды, сравнивая его с тремя значениями: а) короткий межкадровый интервал (Short IFS, *SIFS*); б) межкадровый интервал режима PCF (*PIFS*); в) межкадровый интервал режима DCF (*DIFS*).

Захват среды с помощью распределенной процедуры режима DCF возможен только в том случае, когда среда свободна в течение времени, равного или большего, чем DIFS. Т. е.в качестве IFS в режиме DCF нужно использовать интервал DIFS – самый длительный период из трех возможных, что дает этому режиму самый низкий приоритет.

Межкадровый интервал SIFS имеет наименьшее значение, он служит для первоочередного захвата среды ответными кадрами CTS или квитанциями, которые продолжают или завершают уже начавшуюся передачу кадра.

Значение межкадрового интервала PIFS больше, чем SIFS, но меньше, чем DIFS. Промежутком времени между завершением PIFS и DIFS пользуется арбитр среды. В этом промежутке он может передать специальный кадр, который говорит всем станциям, что начинается *контролируемый период*. Получив этот кадр, станции, которые хотели бы воспользоваться алгоритмом DCF для захвата среды, уже не могут этого сделать, они должны дожидаться окончания контролируемого периода. Длительность этого периода объявляется в специальном кадре, но этот период может закончиться и раньше, если у станций нет чувствительного к задержкам трафика. В этом случае арбитр передает служебный кадр, после которого по истечении интервала DIFS начинает работать режим DCF.

На управляемом интервале реализуется централизованный метод доступа (PCF). Арбитр выполняет процедуру опроса, чтобы по очереди предоставить каждой такой станции право на использование среды, направляя ей специальный кадр. Станция, получив такой кадр, может ответить другим кадром, который подтверждает прием специального кадра и одновременно передает данные (либо по адресу арбитра для транзитной передачи, либо непосредственно станции). Каждая станция может работать в режиме PCF, для этого она должна подписаться на эту услугу при присоединении к сети.

Чтобы какая-то доля среды всегда доставалась асинхронному трафику, контролируемый период ограничен; после его окончания арбитр передает соответствующий кадр и начинается неконтролируемый период.

6.3.6. Физические уровни стандарта IEEE 802.11

С момента принятия первой версии стандарта 802.11 в 1997 г. одной из главных была проблема повышения скорости передачи данных, чтобы приложения, хорошо работающие в проводных сетях, при переходе на беспроводную связь значительно не деградировали. Другой немаловажной проблемой был выбранный частотный диапазон радиоспектра. В соответствии с рекомендациями ITU диапазоны 2,4; 3,6 и 5 ГГц отведены для беспроводной передачи. В разных странах существуют различные правила для этих диапазонов, от свободного использования до обычного лицензирования.

Физические уровни стандарта 802.11 1997 г. В 1997 г. комитетом 802.11 был принят стандарт, который определял функции уровня MAC вместе с тремя вариантами физического уровня со скоростями 1 и 2 Мбит/с.

В первом варианте средой являются инфракрасные волны диапазона 850 нм, которые генерируются либо полупроводниковым лазерным диодом, либо светодиодом (LED). Так как инфракрасные волны не проникают через стены, область покрытия LAN ограничивается зоной прямой видимости.

Во втором варианте в качестве передающей среды используется *микроволновый диапазон 2,4 ГГц*. Этот вариант основан на методе FHSS. В методе FHSS каждый узкий канал имеет ширину 1 МГц. Частотная манипуляция (FSK) с двумя состояниями сигнала (частотами) дает скорость 1 Мбит/с, с четырьмя состояниями – 2 Мбит/с.

Третий вариант, в котором используется тот же *микроволновый диапазон*, основан на методе DSSS, где в качестве последовательности чипов применяется 11-битный код 10110111000. Каждый бит кодируется путем двоичной фазовой (1 Мбит/с) или квадратурной фазовой (2 Мбит/с) манипуляции.

Физические уровни стандартов 802.11a и 802.11b. В 1999 г. были приняты два варианта стандарта физического уровня: *802.11a* и *802.11b*, заменяющие спецификации физического уровня 802.11 редакции 1997 г.

В *спецификации 802.11b* института IEEE по-прежнему используется диапазон 2,4 ГГц. Для повышения скорости до 11 Мбит/с применяется более техника Complementary Code Keying (CCK), заменившая коды Баркера.

Однако диапазон 2,4 ГГц с шириной полосы примерно в 80 МГц используется стандартом 802.11b, отличным от стандарта 1997 г. способом. Этот диапазон разбит на 14 каналов, каждый из которых, кроме последнего, отстоит от соседей на 5 МГц (рис. 6.15) [2].

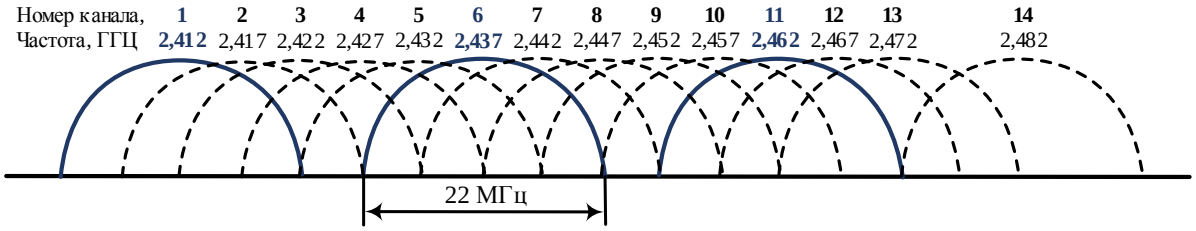


Рис. 6.15. Разбиение диапазона 2,4 ГГц на каналы

Для передачи данных в стандарте 802.11b используется полоса частот шириной в 22 МГц. Для того чтобы гарантировать некоторый минимум взаимных помех, возникающих от передатчиков, работающих в диапазоне 2,4 ГГц, комитет определил так называемую спектральную маску, определяющую разрешенный спектр мощности передатчика, работающего в каком-либо из каналов. Этот спектр мощности должен затухать не меньше чем на 30 дБ на расстоянии 11 МГц от центра канала, что и создает укрупненную полосу шириной в 22 МГц с центром в одном из 14 каналов.

В результате одновременно в одной и той же области покрытия могут работать несколько независимых беспроводных сетей стандарта 802.11b. На рис. 6.15 показан вариант для трех сетей, использующих каналы 1, 6 и 11. Такое использование каналов типично для США, где частотные каналы 12, 13 и 14 для сетей стандарта 802.11 не разрешены. В странах Европы максимальное количество независимых сетей, работающих в одной области покрытия, достигает четырех; обычно они занимают каналы 1, 5, 9 и 13. Оборудование 802.11b может конфигурироваться для любого из 14 каналов, так что при возникновении помех можно перейти на другой канал.

Спецификация 802.11a обеспечивает повышение скорости передачи данных за счет использования полосы шириной 300 МГц из диапазона 5 ГГц. Для передачи данных задействована полоса шириной 20 МГц, что дает возможность иметь 12 и более независимых сетей в одной области покрытия.

Для передачи данных в стандарте 802.11a используется техника ортогонального частотного мультиплексирования (OFDM). Данные первоначально кодируются на 52 первичных несущих частотах методом BPSK, QPSK, 16-QAM или 64-QAM, а затем сворачиваются в общий сигнал с ши-

риной спектра в 20 МГц. Скорость передачи в зависимости от метода кодирования первичной несущей частоты составляет 6, 9, 12, 18, 24, 36, 48 или 54 Мбит/с.

Диапазон 5 ГГц в спецификации 802.11a меньше «населен» и предоставляет больше частотных каналов. Однако его использование связано с проблемами: во-первых, оборудование для этих частот дороже; во-вторых, в некоторых странах частоты этого диапазона подлежат лицензированию, в-третьих, радиоволны этого диапазона хуже проходят через препятствия.

Физический уровень стандарта 802.11g. *Стандарт 802.11g* для физического уровня разработан рабочей группой института IEEE в 2003 г. Он быстро завоевал популярность, так как обеспечивал те же скорости, что и стандарт 802.11a, т. е. до 54 Мбит/с, но в диапазоне 2,4 ГГц, т. е. в том диапазоне, где до этого удавалось достигать максимальной скорости в 11 Мбит/с на оборудовании стандарта 802.11b. В то же время стоимость оборудования стандарта 802.11g достаточно быстро стала соизмеримой со стоимостью оборудования стандарта 802.11b, что и стало причиной роста популярности новой спецификации. В ней, так же, как и в спецификации 802.11a, используется ортогональное частотное мультиплексирование (OFDM).

Дальность работы сети стандарта 802.11 зависит от многих параметров, в том числе от используемого диапазона частот. Обычно дальность беспроводной локальной сети находится в пределах от 100 до 300 м вне помещений и от 30 до 40 м внутри помещений.

Физический уровень стандарта 802.11n. Стандарт 802.11n был принят в октябре 2009 г. Основной его особенностью является дальнейшее повышение скорости передачи данных (до 600 Мбит/с). Оборудование стандарта 802.11n может работать как в диапазоне 5 ГГц, так и в диапазоне 2,4 ГГц, хотя рекомендуемым является диапазон 5 ГГц благодаря большему числу доступных каналов и меньшей интерференции с многочисленным оборудованием, работающим сегодня в диапазоне 2,4 ГГц. Для достижения высоких скоростей в 802.11n применено несколько новых механизмов.

Улучшенное кодирование OFDM и сдвоенные частотные каналы. Вместо каналов с полосой 20 МГц, которые использовались в технологиях 802.11a и 802.11g, в технологии 802.11n применены каналы с полосой 40 МГц (для обратной совместимости допускается также работать с каналами 20 МГц). Само по себе расширение полосы в два раза должно приводить к повышению битовой скорости в два раза, но выигрыш здесь больше за счет усовершенствованной в кодировании OFDM: вместо 52 первичных несущих частот на полосу в 20 МГц здесь используется 57 таких частот, а на полосу в 40 МГц соответственно 114. Это приводит к повышению битовой скорости с 54 до 65 Мбит/с для каналов 20 МГц и до 135 Мбит/с для каналов 40 МГц.

Уменьшение межсимвольного интервала. Для надежного распознавания символов в технологиях 802.11a/g используется межсимвольный интервал в 800 нс. Технология 802.11n позволяет передавать данные с таким же межсимвольным интервалом, а также с межсимвольным интервалом в 400 нм, что повышает битовую скорость для каналов 40 МГц до 150 Мбит/с.

Применение техники MIMO (Multiple Input Multiple Output – множественные входы и выходы) основано на использовании одним сетевым адаптером нескольких антенн с целью лучшего распознавания сигнала, пришедшего к приемнику разными путями. Обычно из-за таких эффектов PPB, как отражение, дифракция и рассеивание, приемник получает несколько сигналов, дошедших от передатчика по разным путям и имеющим, следовательно, сдвиг по фазе. До введения MIMO такие явления считались негативными и с ними боролись путем применения нескольких антенн, из которых в каждый момент времени использовалась только одна – та, которая принимала сигнал лучшего качества. Техника MIMO принципиально изменила отношение к сигналам, пришедшим разными путями, – эти сигналы комбинируются и путем обработки из них восстанавливается исходный сигнал.

Техника MIMO не только способствует улучшению соотношения сигнал/помеха. Благодаря возможности обрабатывать сигналы, пришедшие разными путями, для каждого потока с целью создания избыточного сигнала можно передавать с помощью нескольких антенн несколько независимых потоков данных (обычно их число меньше, чем число антенн). Эта способность систем MIMO называется *пространственным мультиплексированием* (spatial multiplexing). Для систем MIMO принято использовать обозначение:

$$T \times R : S,$$

где T – количество передающих антенн узла, R – количество приемных антенн узла, а S – количество потоков данных, которые пространственно мультиплексируются. Типичной системой MIMO стандарта 802.11n является система 3×3:2, т. е. система с тремя передающими и тремя принимающими антеннами, которая позволяет передавать два независимых потока данных. Система MIMO 3×3:2 обеспечивает повышение битовой скорости в два раза, т. е. до 300 Мбит/с для каналов 40 МГц. Стандарт 802.11n предусматривает различные варианты системы MIMO вплоть до 4×4:4, что позволяет повысить битовую скорость до 600 Мбит/с.

Физический уровень стандарта 802.11ac. Спецификация 802.11ac является развитием спецификации 802.11n, она обеспечивает скорости передачи данных до 1 Гбит/с за счет: а) расширения полосы индивидуального канала до 80 МГц (обязательная опция) или 160 МГц (возможная опция);

б) поддержки до 8 каналов MIMO; в) применения модуляции с большим числом состояний: 256QAM вместо 64QAM у стандарта 802.11n.

Физический уровень стандарта 802.11ad. Эта версия стандарта 802.11 была создана беспроводным гигабитным альянсом (Wireless Gigabit Alliance, WiGig) в 2009–2011 гг. Стандарт отличается тем, что в нем используется частотный диапазон 60 ГГц (а также диапазоны 2,4 и 5 ГГц). В диапазоне 60 ГГц может существовать четыре канала шириной 2,16 ГГц каждый. Широкая полоса канала позволяет передавать данные с гигабитными скоростями – до 4,6 Гбит/с при наличии одного канала и до 7 Гбит/с при мультиплексированной передаче OFDM одновременно по четырем каналам.

Однако передача данных с несущей частотой 60 ГГц сталкивается с проблемой распространения сигнала – он не проходит через стены, как сигнал частот 2,4 или 5 ГГц (хотя и отражается от препятствий, что используют приемопередатчики 802.11ad). Поэтому область покрытия сети 802.11ad ограничена одной комнатой, в которой находятся устройства, требующие обмена данными с гигабитными скоростями, например, приемник и телевизор стандарта HD. Приёмопередатчики стандарта 802.11ad могут также работать и в диапазонах 2,4 и 5 ГГц, чтобы обеспечить связь, когда невозможно использовать частоты 60 ГГц (из-за размещения узлов сети в разных помещениях).

Выводы

Стандарты 802.11 являются основными стандартами беспроводных локальных сетей. Существует несколько вариантов спецификаций физического уровня 802.11, отличающихся диапазоном используемых частот (2,4 и 5 ГГц), а также методом кодирования и мультиплексирования (FHSS, DSSS, OFDM).

Метод доступа 802.11 является комбинацией случайного метода доступа с предотвращением коллизий (DCF) и централизованного детерминированного метода доступа с опросом (PCF). Гибкое применение режимов DCF и PCF позволяет обеспечить поддержку показателей QoS для синхронного и асинхронного трафиков.

Контрольные вопросы

1. Приведите основные особенности функционирования БЛС.
2. К чему приводит наличие скрытого терминала в сети IEEE 802.11? Варианты ответов:

а) к нарушению связности сети;

б) к повышению уровня помех в радиосреде;

г) к более частому возникновению коллизий.

3. Опишите основные топологии БЛС стандарта 802.11.
4. Охарактеризуйте стек протоколов IEEE 802.11.

5. Сформулируйте принцип работы распределенного режима доступа DCF.
6. Сформулируйте принцип работы централизованного режима доступа PCF.
7. Как межкадровый интервал влияет на возможность захвата среды передачи?
8. Опишите физический уровень стандарта 802.11 1997 г.
9. Опишите физический уровень стандартов 802.11, 802.11b, 802.11g.
10. Опишите физический уровень стандарта 802.11n, 802.11ac, 802.11ad.

6.4. Персональные сети Bluetooth

6.4.1. Особенности персональных сетей

Персональные сети (Personal Area Network, PAN) предназначены для взаимодействия устройств, принадлежащих одному владельцу, на небольшом расстоянии, обычно в радиусе 10 м. Такими устройствами могут быть ноутбук, мобильный телефон, принтер, карманный компьютер (Personal Digital Assistant, PDA), а также многочисленные бытовые приборы.

Типичным примером PAN является беспроводное соединение компьютера с периферийными устройствами, такими как принтер, наушники, мышь, клавиатура и т. п. Мобильные телефоны также используют технологию PAN для соединения со своей периферией (чаще всего это наушники), а также с компьютером своего владельца. Некоторые наручные часы стали поддерживать PAN, превращаясь в универсальные устройства с функциями PDA.

Персональные сети во многом похожи на локальные, но у них есть и свои особенности.

Многие из устройств, которые могут входить в персональную сеть, *гораздо проще*, чем традиционный узел LAN – компьютер. Кроме того, такие устройства обычно имеют небольшие габариты и стоимость. Поэтому в стандартах PAN требуется учитывать, что их реализация должна приводить к недорогим решениям с низким энергопотреблением.

Область покрытия PAN меньше области покрытия LAN, узлы PAN часто находятся на расстоянии нескольких метров друг от друга.

Высокие требования к безопасности. Персональные устройства, путешествуя вместе со своим владельцем, попадают в различное окружение. Иногда они должны взаимодействовать с устройствами других персональных сетей, например, если их владелец встретил на улице своего знакомого и решил переписать из его устройства PDA в свое несколько адресов общих знакомых. В других случаях такое взаимодействие явно нежелательно, так как может привести к утечке конфиденциальной информации. Поэтому протоколы PAN должны обеспечивать разнообразные методы аутентификации устройств и шифрования данных в мобильной обстановке.

При соединении малогабаритных устройств между собой желание избавиться от кабелей проявляется сильнее, чем при соединении компьютера

с принтером или концентратором. Из-за этого персональные сети в гораздо большей степени, чем локальные, *тяготеют к беспроводным решениям*.

Если человек постоянно носит устройство PAN с собой и на себе, то оно не должно причинять вреда его здоровью. Поэтому такое устройство должно *излучать сигналы небольшой мощности*, желательно не более 100 мВт.

Сегодня самой популярной технологией PAN является *Bluetooth*, которая обеспечивает взаимодействие восьми устройств в разделяемой среде диапазона 2,4 ГГц с битовой скоростью передачи данных до 3 Мбит/с.

6.4.2. Архитектура Bluetooth

Стандарт Bluetooth разработан группой Bluetooth SIG (Bluetooth Special Interest Group), которая была организована по инициативе компании Ericsson. Стандарт Bluetooth также адаптирован рабочей группой IEEE 802.15.1 в соответствии с общей структурой стандартов IEEE 802.

В технологии Bluetooth используется концепция *пикосети*. Название подчеркивает небольшую область покрытия, от 10 до 100 м, в зависимости от мощности излучения передатчика устройства. В пикосеть может входить до 255 устройств, но только 8 из них могут в каждый момент времени быть активными и обмениваться данными. Одно из устройств в пикосети является *главным*, остальные – *подчиненными* (рис. 6.16) [0].

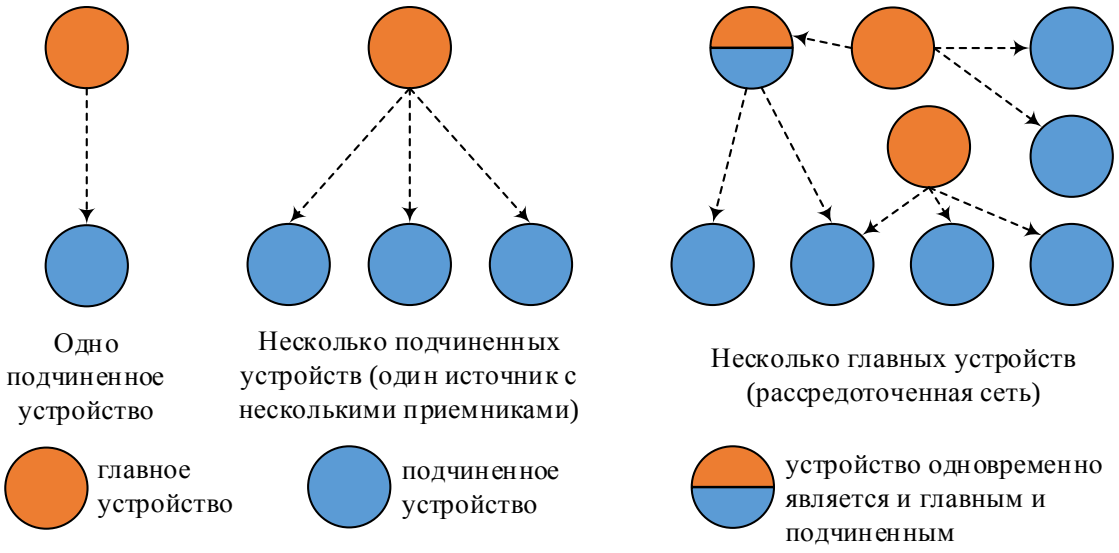


Рис. 6.16. Разбиение диапазона 2,4 ГГц на каналы

Активное подчиненное устройство может обмениваться данными только с главным устройством, прямой обмен между подчиненными устройствами невозможен. Все подчиненные устройства данной пикосети, кроме семи активных, должны находиться в режиме пониженного энергопотребления и периодически прослушивать команду главного устройства

для перехода в активное состояние. Главное устройство отвечает за доступ к разделяемой среде пикосети, которая представляет собой нелицензируемые частоты диапазона 2,4 ГГц. Разделяемая среда передает данные со скоростью до 3 Мбит/с, но из-за накладных расходов на заголовки пакетов и смену частот полезная скорость передачи данных в среде не превышает 2,1 Мбит/с. Пропускная способность среды делится главным устройством между семью подчиненными устройствами на основе техники TDM.

Такая архитектура позволяет применять более простые протоколы в устройствах, выполняющих функции подчиненных (например, в радионаушниках), и отдает более сложные функции управления пикосетью компьютеру, который, скорее всего, и будет главным устройством этой сети.

Присоединение к пикосети происходит динамически. Главное устройство пикосети, используя процедуру опроса, собирает информацию об устройствах, которые попадают в зону его пикосети. После обнаружения нового устройства главное устройство проводит с ним переговоры. Если желание подчиненного устройства присоединиться к пикосети совпадает с решением главного устройства (подчиненное устройство прошло проверку аутентичности и оказалось в списке разрешенных устройств), то новое подчиненное устройство присоединяется к сети.

Несколько пикосетей, которые обмениваются между собой данными, образуют рассредоточенную сеть. Взаимодействие в пределах рассредоточенной сети осуществляется за счет того, что один узел (называемый мостом) одновременно является членом нескольких пикосетей, причем этот узел может исполнять роль главного устройства одной пикосети и подчиненного устройства другой.

Сеть Bluetooth использует технику расширения спектра FHSS. Для того чтобы сигналы разных пикосетей не интерферировали, каждое главное устройство задействует собственную последовательность псевдослучайной перестройки рабочей частоты (ППРЧ). Наличие различающихся последовательностей ППРЧ затрудняет общение пикосетей между собой. Устройство, играющее роль моста, должно при подключении к каждой из пикосетей соответствующим образом менять последовательность.

Коллизии, хотя и с очень небольшой вероятностью, все же могут происходить, когда два или более устройства из разных пикосетей выберут для работы один и тот же частотный канал. Для надежной передачи данных в Bluetooth может выполняться прямая коррекция ошибок (FEC), а получение кадра подтверждается с помощью квитанций.

В сетях Bluetooth для передачи информации двух типов используются разные методы.

Для *чувствительного к задержкам трафика* (например, голоса) сеть поддерживает синхронный канал, ориентированный на соединение (Syn-

chronous Connection-Oriented link, SCO). Этот канал работает на скорости 64 Кбит/с. Для канала SCO пропускная способность резервируется на все время соединения.

Для *эластичного трафика* (например, компьютерных данных) используется работающий с переменной скоростью асинхронный канал, не ориентированный на соединение (Asynchronous Connection-Less link, ACL). Для канала ACL пропускная способность выделяется по запросу подчиненного устройства или по потребности главного устройства.

Bluetooth является законченной оригинальной технологией, рассчитанной на самостоятельное применение в электронных персональных устройствах. Поэтому эта технология поддерживает полный стек протоколов, включая собственные прикладные протоколы. В этом заключается ее отличие от рассмотренных ранее технологий, таких как Ethernet или IEEE 802.11, которые лишь выполняют функции физического и канального уровней.

Создание для технологии Bluetooth собственных прикладных протоколов объясняется стремлением разработчиков реализовать ее в разнообразных простых устройствах, которым не под силу, да и ни к чему поддерживать стек протоколов TCP/IP. Технология Bluetooth появилась в результате попыток разработать стандарт для взаимодействия мобильного телефона с беспроводными наушниками. Понятно, что для решения такой простой задачи не нужен ни протокол передачи файлов (FTP), ни протокол передачи гипертекста (HTTP). В результате для технологии Bluetooth был создан оригинальный стек протоколов и большое количество профилей.

Стек протоколов Bluetooth постоянно совершенствуется. Версия 1.0 стандартов стека была принята в 1999 г., версия 1.2 – в 2003, версия 2.0 – в 2004, версия 2.1 – в 2007, версия 3.0 – в 2009, версия 4.0 – в 2010, версия 4.1 – в 2013, а версия 4.2 – в декабре 2014 г.

Профили определяют конкретный набор протоколов для решения той или иной задачи. Например, существует профиль для взаимодействия компьютера или мобильного телефона с беспроводными наушниками. Имеется также профиль для устройств, которые могут передавать файлы и т. д.

Разделяемая среда представляет собой последовательность частотных каналов технологии FHSS в диапазоне 2,4 ГГц. Каждый частотный канал имеет ширину 1 МГц, количество каналов равно 79.

Чиповая скорость равна 1600 Гц, поэтому период чипа составляет 625 мкс. Главное устройство разделяет общую среду на основе временно́го мультиплексирования (TDM), используя в качестве тайм-слота время пребывания системы на одном частотном канале, т. е. 625 мкс [2].

В версии протокола 1.0 информация кодируется с тактовой частотой 1 МГц путем двоичной частотной манипуляции (BFSK), в результате битовая скорость составляет 1 Мбит/с. В течение одного тайм-слота пикосеть

Bluetooth передает 625 бит, но не все они служат для передачи полезной информации. При смене частоты устройствам сети требуется некоторое время для синхронизации, поэтому из 625 бит только 366 передают кадр данных.

В версии 2.0 был введен режим *улучшенной скорости передачи данных* (Enhanced Data Rate, EDR), в котором для кодирования данных используется комбинация методов частотной (BFSK) и фазовой (PSK) модуляции; за счет этого удалось повысить битовую скорость до 3 Мбит/с, а полезную скорость передачи данных – до 2,1 Мбит/с. Режим EDR дополняет основной режим передачи данных со скоростью 1 Мбит/с.

Кадр данных может занимать 1, 3 или 5 слотов. Когда кадр занимает больше одного слота, частота канала остается неизменной в течение всего времени передачи кадра. В этом случае накладные расходы на синхронизацию меньше, так что размер кадра, состоящего, например, из пяти последовательных слотов, равен 2870 бит (с полем данных до 2744 бит).

6.4.3. Поиск и стыковка устройств Bluetooth

Устройство Bluetooth, обычно посылает периодические запросы на предмет обнаружения других устройств Bluetooth в зоне досягаемости. Если устройство Bluetooth получает такой запрос, и оно сконфигурировано так, чтобы отвечать на запросы, то в ответ устройство передает сведения о себе: имя и тип устройства, имя производителя, поддерживаемые сервисы.

Имя устройства конфигурируется в отличие от его уникального MAC-адреса, который дается производителем. Часто устройства выпускаются со сконфигурированными по умолчанию именами, соответствующими названию модели устройства, поэтому в сфере досягаемости вашего мобильного телефона может оказаться несколько других телефоном с одинаковыми именами Bluetooth, если их владельцы не дали им собственные имена.

После предварительного обмена информацией устройства Bluetooth могут начать так называемую процедуру стыковки (pairing), если конфигурация устройств ее требует. Стыковка подразумевает установление защищенного канала между устройствами; безопасность в данном случае означает, что устройства доверяют друг другу, а данные между ними передаются в зашифрованном виде. Стыковка устройств Bluetooth требует введения в каждое из них одного и того же пароля, называемого также PIN-кодом Bluetooth. Обычно устройство, получившее запрос на стыковку, просит пользователя ввести PIN-код. Устройства, успешно прошедшие процедуру стыковки, запоминают этот факт и устанавливают безопасное соединение автоматически всякий раз, когда оказываются в зоне досягаемости, при этом повторное введение PIN-кода пользователем не требуется.

6.4.4. Развитие технологии Bluetooth

В последних версиях стандартов Bluetooth были анонсированы некоторые нововведения, одно из которых – повышение скорости передачи данных в режиме EDR до 3 Мбит/с – мы уже упомянули. Далее перечислены другие наиболее важные новые свойства этой технологии.

Пониженная скорость обмена в ждущем режиме. Это свойство заключается в снижении частоты обмена служебными сообщениями keepalive («работоспособен»), которыми узлы поддерживают соединение в открытом состоянии при отсутствии пользовательских данных для передачи, с нескольких сообщений в секунду до одного сообщения раз в 5 или 10 секунд. Такой режим позволяет увеличить время работы батарей портативных устройств в 3–10 раз. Свойство введено в версии 2.1.

Безопасная простая стыковка (secure simple pairing) позволяет ускорить процедуру стыковки и в то же время предлагает более высокую степень защиты соединений. Свойство введено в версии 2.1.

Использование технологии NFC (Near Field Communication – связь ближнего радиуса действия) для автоматической стыковки устройств. NFC – это новая технология, разработанная для беспроводного взаимодействия устройств на расстояниях 10–20 см. При обнаружении сигналов устройства с интерфейсами NFC автоматически устанавливают соединение. Устройства Bluetooth могут использовать технологию NFC для автоматического обнаружения при приближении их друг к другу в ходе стыковки и обмена информацией. Это свойство является частью процедуры безопасной простой стыковки, оно также введено в версии Bluetooth 2.1.

Альтернативные MAC-уровень и физический уровень. При необходимости передачи большого объема данных устройство Bluetooth может переключиться на соединение, использующее отличную от Bluetooth технологию передачи данных. В версии 3.0 протоколов Bluetooth как возможная альтернатива определены пока только технологии 802.11.

Bluetooth с низким энергопотреблением (Bluetooth low energy). В апреле 2009 г. группа Bluetooth SIG объявила о совершенно новом дополнительном стеке протоколов под названием Bluetooth Low Energy (Bluetooth LE). Протоколы Bluetooth LE предназначены для устройств, батареи которых должны иметь примерно годичный срок действия; это могут быть, например, наручные часы или медицинские приборы.

Технология Bluetooth LE получила маркетинговое название Bluetooth Smart, сегодня она реализована в большинстве смарт-телефонов и планшетов. Существуют реализации протоколов Bluetooth Smart в качестве единственного стека протоколов Bluetooth некоторого устройства, а также в варианте второго стека протоколов, работающего наряду с классическим стеком, – в этом случае такое устройство маркируется как Bluetooth Smart Ready.

Для передачи данных Bluetooth Smart использует тот же частотный диапазон 2,4 ГГц, что и классический вариант Bluetooth, но в нем организуется не 79 каналов с полосой 1 МГц, а 40 каналов с полосой 2 МГц каждый. Битовая скорость передачи данных составляет, как и у классической версии, 1 Мбит/с. Передача голоса по Bluetooth Smart не предусмотрена.

Спецификация Bluetooth Smart описана в версии Bluetooth SIG 4.0 стандарта, а в версии 4.2 вводятся некоторые ее усовершенствования, например, расширенный размер пакета, за счет чего повышается эффективная пропускная способность приложений, работающих поверх Bluetooth Smart.

Выводы

Персональные сети (PAN) предназначены для взаимодействия принадлежащих одному владельцу устройств на небольшом расстоянии, обычно в радиусе от 1 до 10 м. Персональные сети должны обеспечивать как фиксированный доступ, например, в пределах дома, так и мобильный, когда владелец устройств перемещается вместе с ними между помещениями или городами.

Сегодня самой популярной технологией PAN является Bluetooth, в которой используется концепция пикосети, объединяющей до 255 устройств, но только восемь из них могут в каждый момент времени быть активными.

Для чувствительного к задержкам трафика сеть Bluetooth поддерживает синхронные каналы, ориентированные на соединение (SCO), а для эластичного – асинхронные каналы, не ориентированные на соединение (ACL).

Контрольные вопросы

1. Дайте определение персональных сетей.
2. Приведите особенности персональных сетей, отличающие их от локальных сетей.
3. Сформулируйте концепцию пикосети Bluetooth.
4. Приведите особенности функционирования активных и подчиненных устройств в пикосети Bluetooth.
5. Как несколько пикосетей могут быть объединены в рассредоточенную сеть? Как при этом работает узел, называемый мостом?
6. Сформулируйте особенности работы метода псевдослучайной перестройки рабочей частоты в Bluetooth.
7. Какие типы каналов в Bluetooth используются для передачи чувствительного к задержкам и эластичного трафика?
8. Поясните происхождение профилей в технологии Bluetooth.
9. Как осуществляется стыковка двух устройств Bluetooth?
10. В каких направлениях идет развитие технологии Bluetooth?

ПРАКТИКУМ

1. МЕТОДЫ РАЗДЕЛЕНИЯ КАНАЛОВ

1.1. Практическое задянтие. Методы разделения каналов

Цель занятия: изучить принципы уплотнения сигналов с частотным, временным и кодовым разделением каналов; оценить производительность систем с TDMA и FDMA.

Ресурс связи представляет время и ширину полосы, доступные для передачи сигнала в определенной системе. Графически ресурс связи можно изобразить на двумерном графике, где ось абсцисс представляет время, а ось ординат – частоту. Для создания эффективной системы связи необходимо спланировать распределение ресурса между пользователями системы так, чтобы время/частота использовались максимально эффективно. С проблемой совместного использования ресурса связи связаны термины «уплотнение» и «множественный доступ». При использовании термина *уплотнение,* требования пользователя к совместному использованию ресурса связи постоянны либо изменяются незначительно, а распределение ресурса выполняется априорно.

Существуют следующие основные способы распределения ресурса связи [3].

1. *Частотное разделение* (frequency division – FD). Распределяются определенные поддиапазоны используемой полосы частот.

2. *Временное разделение* (time division – TD). Пользователям выделяются периодические временны́е интервалы. В некоторых системах пользователям предоставляется ограниченное время для связи. В других случаях время доступа пользователей к ресурсу определяется динамически.

3. *Кодовое разделение* (code division – CD). Выделяются определенные элементы набора ортогонально (либо почти ортогонально) распределенных спектральных кодов, каждый из которых использует весь диапазон частот.

4. *Пространственное разделение* (space division – SD), или многолучевое многократное использование частоты. С помощью лучевых антенн радиосигналы разделяются и направляются в разные стороны. Данный метод допускает многократное использование одного частотного диапазона.

5. *Поляризационное разделение* (polarization division – PD), или *двойное поляризационное использование частоты*. Для разделения сигналов применяется ортогональная поляризация в одном частотном канале.

Ключевым моментом во всех схемах уплотнения является то, что при использовании ресурса различными сигналами избежать взаимных помех между разными пользователями позволяет использование в разных каналах ортогональных сигналов.

Сигналы $x_i(t)$, где $i = 1, 2, ...,$ являются ортогональными, если во временной области выполняется условие:

$$\int_{-\infty}^{\infty} x_i(t) x_j(t) dt = \begin{cases} K \text{ при } i = j; \\ 0 \text{ при } i \neq j, \end{cases} \quad (1.1)$$

где K – ненулевая константа. Подобным образом сигналы ортогональны, если в частотной области выполняется условие:

$$\int_{-\infty}^{\infty} X_i(f) X_j(f) df = \begin{cases} K \text{ при } i = j; \\ 0 \text{ при } i \neq j, \end{cases} \quad (1.2)$$

где функции $X_i(f)$ являются Фурье-образами сигналов $x_i(t)$. Распределение по каналам, характеризующееся ортогональными спектрами, для которых выполняется (1.1), называют уплотнением с временным разделением (*time-division multiplexing* – TDM) или множественным доступом с временным разделением (time-division multiple access – TDMA). Распределение по каналам, характеризующееся ортогональными волнами, для которых выполняется условие (1.2), называют уплотнением с частотным разделением (frequency-division multiplexing – FDM) или множественным доступом с частотным разделением (frequency-division multiple access – FDMA).

1.1.1. Уплотнение/множественный доступ с частотным разделением

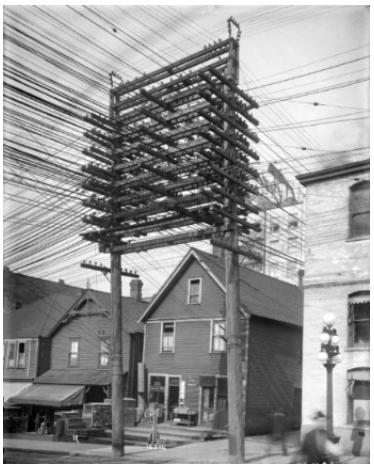


Рис. 1.1. Два провода для каждой магистральной телефонной линии

На заре создания телефонной связи для каждой телефонной линии, соединяющей телефонные центры, было необходимо устанавливать два провода. Как видно из рис. 1.1, небо над городами становилось все темнее по мере развития телефонной связи. Важное открытие в области телефонной связи начала XX в. – уплотнение с частотным разделением (FDM) – позволило передавать несколько телефонных сигналов по одному проводу.

Ресурс связи представлен на рис. 1.2 в виде частотно-временной зависимости. Спектральное распределение по каналам является примером технологии FDM или FDMA. Здесь распределе-

ние сигналов или пользователей по диапазону частот является *долгосрочным* или *постоянным*. Ресурс связи может одновременно содержать несколько сигналов, разнесенных в спектре. Первый частотный диапазон содержит сигналы, которые используют промежуток между f_0 и f_1, второй – между f_2 и f_3 и т. д. Области спектра, находящиеся между используемыми диапазонами, называют защитными полосами частот. Защитные полосы выполняют роль буфера, что позволяет снизить интерференцию между соседними (по частоте) каналами. Может возникнуть вопрос: как преобразовать сигнал так, чтобы он использовал более высокий диапазон частот? Ответ: при помощи смешивания (модуляции) информационного сигнала и синусоидального сигнала фиксированной частоты.

Рис. 1.2. Уплотнение с частотным разделением

Если два модулируемых входных сигнала описываются синусоидами с частотами f_A и f_B, их смешение/перемножение дает две частоты: f_{A+B} и f_{A-B}. Процесс модуляции описывается тригонометрическим равенством:

$$\cos A \cos B = \frac{1}{2}\left[\cos(A+B) + \cos(A-B)\right]. \tag{1.3}$$

На рис. 1.3, а показано модулирование типичного голосового телефонного сигнала $x(t)$ (частоты немодулированного сигнала принадлежат диапазону 300–3400 Гц) синусоидальным сигналом с частотой 20 кГц. Двусторонний спектр немодулированного сигнала $|X(f)|$ показан на рис. 1.3, а. На рис. 1.3, б представлен односторонний спектр $|X(f - f_0)|$ на выходе смесителя. В результате смешивания, описанного в уравнении (1.3), спектр смещается в сторону более высоких частот, по сравнению с немодулированным спектром, и центрирован теперь на частоте 20 кГц. Данный спектр называют двухполосным (double-sideband – DSB), поскольку информация находится в двух различных диапазонах частот.

На рис. 1.3, в показана нижняя боковая полоса (lower sideband – LSB), которой принадлежат частоты 16600–19700 Гц. Иногда нижнюю боковую полосу называют инвертированной боковой полосой, поскольку частотные составляющие этой полосы расположены в обратном порядке, по сравнению с немодулированным сигналом. Подобным образом фильтрование может использоваться для выделения верхней боковой полосы (upper sideband – USB), которой, как показано на рис. 1.3, г, принадлежат частоты 20300–23400 Гц. Данную боковую полосу иногда называют прямой, поскольку частотные составляющие этой полосы расположены в том же порядке, что и в немодулированном сигнале. Для восстановления исходных данных немодулированного сигнала необходима лишь одна боковая полоса – верхняя или нижняя.

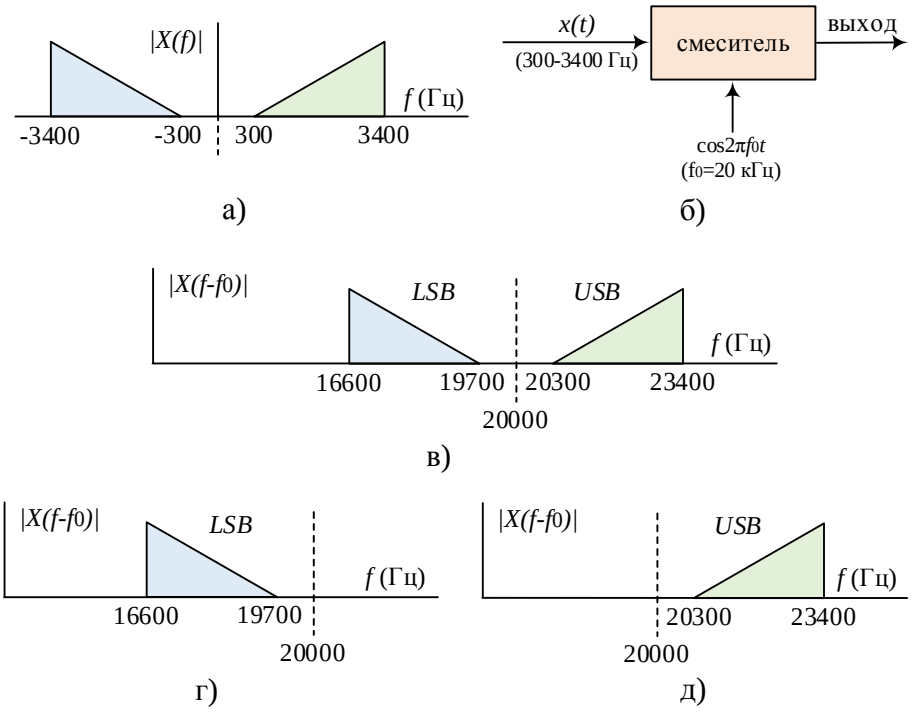


Рис. 1.3. Смешивание сигналов:
а) входной немодулированный голосовой спектр;
б) процесс смешивания; в) выходной двухполосный спектр;
г) нижняя боковая полоса; д) верхняя боковая полоса

На рис. 1.4 приведен простейший пример технологии FDM. В данном случае реализована схема с тремя каналами передачи речи. В канале 1 голосовой сигнал из диапазона 300–3400 Гц модулируется сигналом с частотой 20 кГц. В каналах 2 и 3 аналогичный голосовой сигнал модулируется сигналами с частотами 16 и 12 кГц. В приведенном примере сохраняются лишь нижние боковые полосы. Суммарный выходной сигнал есть сумма трех сигналов и принадлежит диапазону 8,6–19,7 кГц [3].

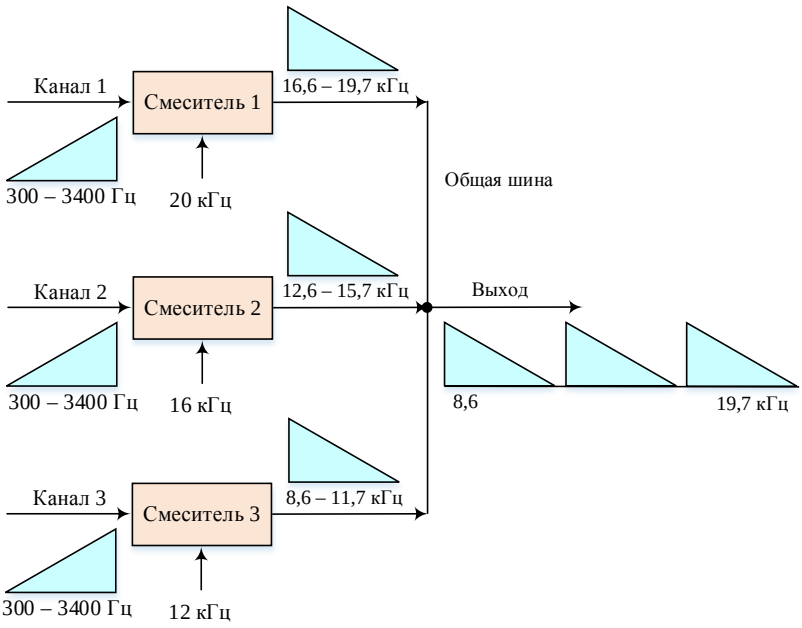


Рис. 1.4. Простейший пример FDM.
Три сдвинутых по частоте канала передачи речи

На рис. 1.5 представлены два наиболее низких уровня иерархии уплотнения телефонных каналов с использованием FDM. Первый уровень состоит из *группы* 12 каналов, модулируемых поднесущими с частотами из диапазона 60–108 кГц. Второй уровень, состоящий из 5 групп (60 каналов), называют *супергруппой*. Супергруппа модулируется поднесущими с частотами из диапазона 312–552 кГц. Уплотненные каналы теперь рассматриваются как составной сигнал, который может передаваться по кабелю или модулироваться несущей с целью последующей радиопередачи [3].

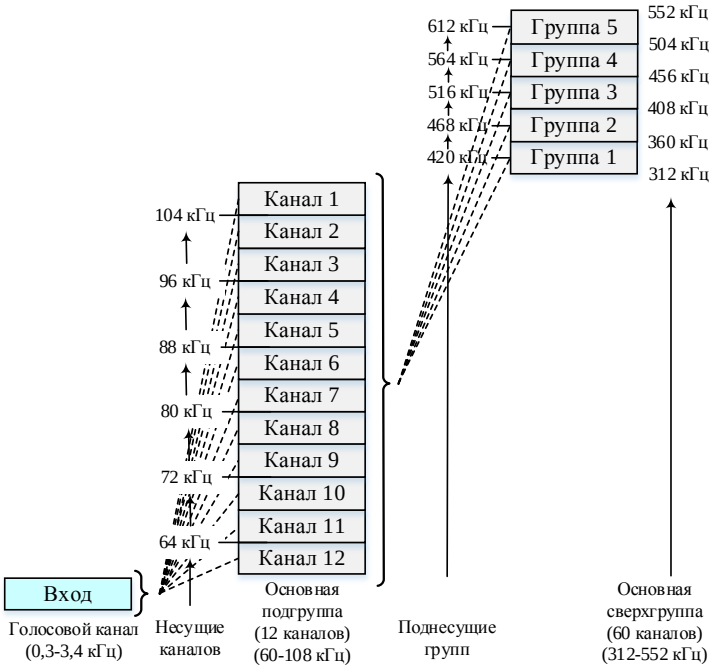


Рис. 1.5. Схема модулирования
типичной системы уплотнения с частотным разделением

Множественный доступ с частотным разделением в спутниковых системах. Большинство спутников связи расположено на *геостационарной* или *геосинхронной* орбите. Это означает, что спутник находится на круговой орбите, лежащей в плоскости земного экватора. При этом спутник находится на такой высоте над уровнем моря (приблизительно 35830 км), на которой период его обращения вокруг Земли равен периоду вращения самой Земли. Поскольку при наблюдении с Земли такие объекты кажутся неподвижными, три спутника, расположенных через 120° друг от друга, позволяют охватить территорию всего земного шара (за исключением полярных областей). Большинство спутниковых систем связи используют нерегенеративные ретрансляторы или транспондеры. *Нерегенеративный* означает, что сигналы «Земля – спутник» усиливаются, сдвигаются по частоте и ретранслируются на Землю без обработки сигнала, демодуляции или повторной модуляции. Наиболее распространенные транспондеры работают в режиме FDM-FM-FDMA (уплотнение с частотным разделением, частотная модуляция, множественный доступ с частотным разделением). Рассмотрим составляющие указанного режима: 1) FDM. Сигналы, подобные телефонным, имеющие одиночную боковую полосу шириной 4 кГц, обрабатываются с использованием FDM, в результате чего формируется составной многоканальный сигнал; 2) FM. Составной сигнал модулируется несущей и передается на спутник; 3) FDMA. Поддиапазоны полосы транспондера могут распределяться между различными пользователями; каждому пользователю выделяется определенная полоса, на которой он получает доступ к транспондеру. Преимуществом технологии FDMA, в сравнении с TDMA, является простота. Каналы FDMA не требуют синхронизации или централизованного распределения времени.

1.1.2. Уплотнение/множественный доступ с временным разделением

На рис. 1.6 показано совместное использование ресурса связи путем предоставления каждому из *M* сигналов (или пользователей) всего спектра в течение небольшого отрезка времени, называемого *временным интервалом* (time slot). Промежутки времени, разделяющие используемые интервалы, называются *защитными интервалами* (guard time). Защитный интервал создает некоторую временную неопределенность между соседними сигналами и выступает в роли буфера, снижая тем самым интерференцию [3].

На рис. 1.7 приведен пример использования технологии TDMA в спутниковой связи. Время разбито на интервалы, называемые *кадрами* (frame). Каждый *кадр* делится на *временные интервалы*, которые могут быть распределены между пользователями. Общая структура кадров периодически повторяется, так что передача данных по схеме TDMA – это один или бо-

лее временных интервалов, которые периодически повторяются на протяжении каждого кадра. Каждая наземная передающая станция транслирует информацию в виде пакетов таким образом, чтобы они поступали на спутник в соответствии с установленным расписанием. После принятия транспондером такие пакеты ретранслируются на Землю вместе с информацией от других передающих станций. Принимающая станция детектирует и разуплотняет уплотненные данные соответствующего пакета.

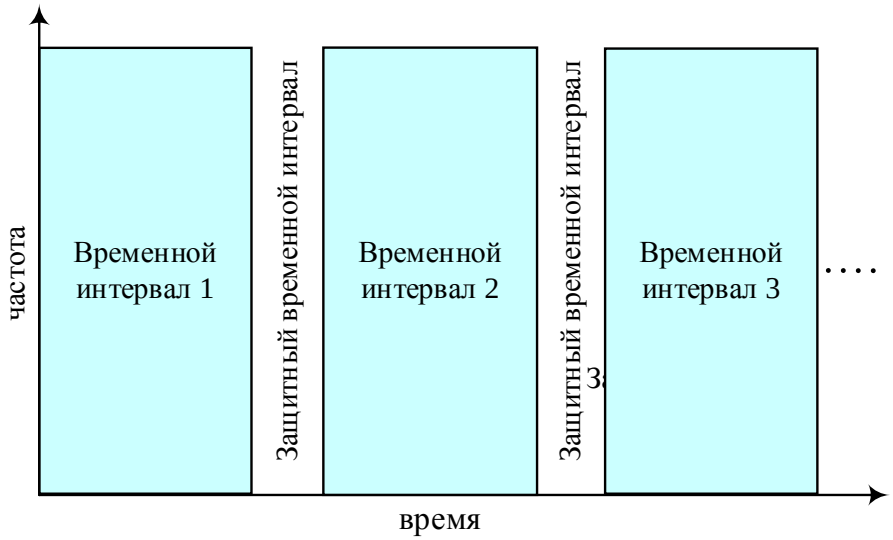


Рис. 1.6. Уплотнение с временным разделением

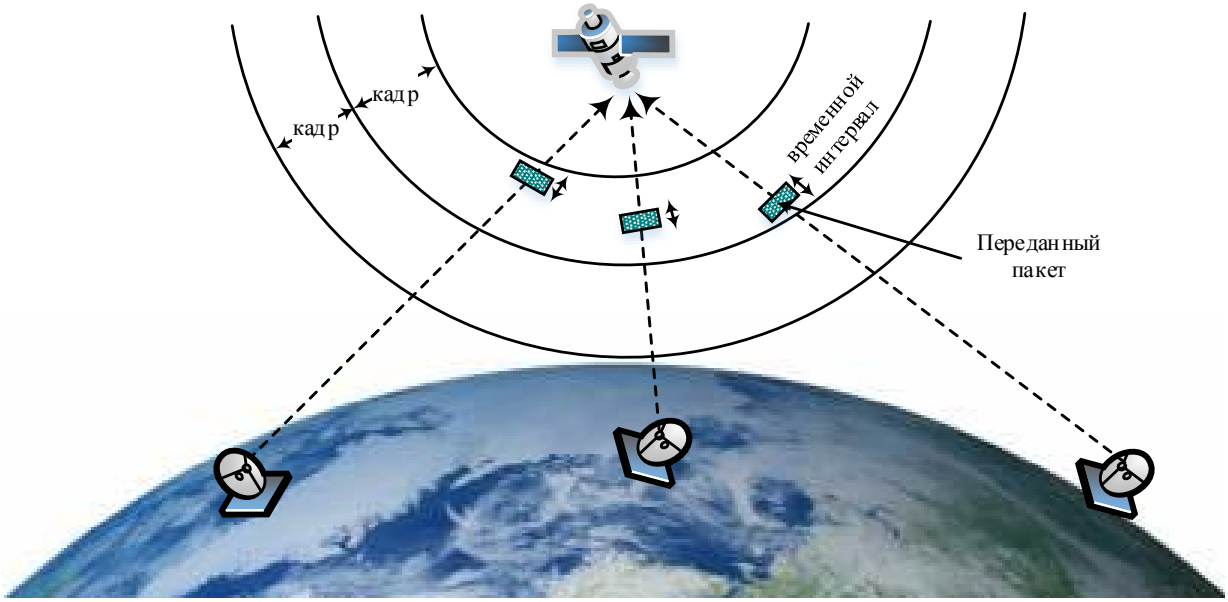


Рис. 1.7. Типичная конфигурация TDMA

Простейшая схема TDM/TDMA именуется *TDM/TDMA с фиксированным распределением*. При использовании такой схемы *M* временных интервалов,

составляющих кадр, заранее распределены между источниками сигнала на достаточно длительный промежуток времени. На рис. 1.8 в виде блок-схемы показана работа такой системы. Операция уплотнения состоит в предоставлении каждому источнику возможности использовать один или более интервалов. Разуплотнение – это распознавание интервалов с последующим распределением данных между пользователями [3].

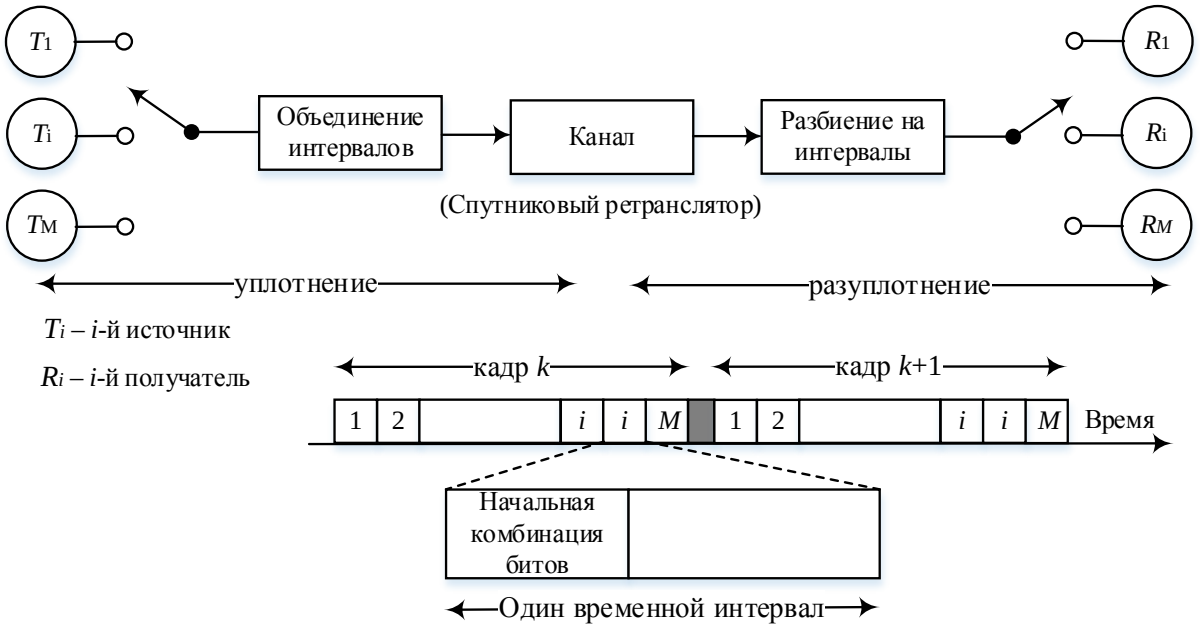


Рис. 1.8. TDM с фиксированным распределением

Два коммутирующих ключа на рис. 1.8 должны быть синхронизированы таким образом, чтобы сообщение, соответствующее источнику 1, попадало на выход канала 1 и т. д. Само по себе сообщение в общем случае состоит из начальной комбинации битов (preamble) и собственно информационной части. Начальная комбинация обычно состоит из элементов, которые отвечают за синхронизацию, адресацию и защиту от ошибок.

Схема TDM/TDMA с фиксированным распределением является чрезвычайно эффективной, когда требования пользователя можно предвидеть, а поток данных значителен (т. е. временные интервалы практически всегда заполнены). В случае же пульсирующего или случайного потока данных указанный метод себя не оправдывает. Рассмотрим простой пример, представленный на рис. 1.9. Здесь кадр составляют четыре интервала, каждый из которых закреплен за пользователями *A*, *B*, *C* и *D*. На рис. 1.9, а изображены схемы активности четырех пользователей.

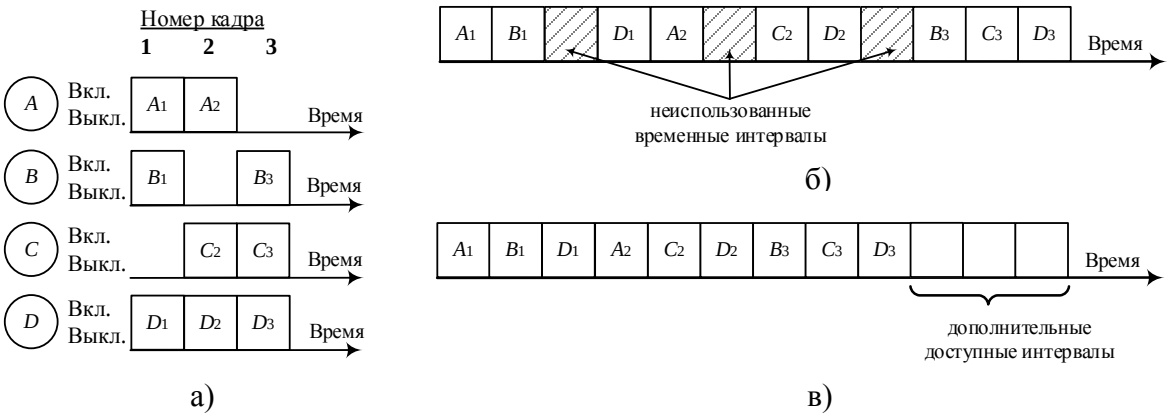


Рис. 1.9. TDM с фиксированным распределением
и система с коммутацией пакетов:
а) схема активности пользователей; б) TDM с фиксированным распределением;
в) коммутация пакетов с временным разделением (концентрация)

На протяжении первого интервала передачи кадра пользователь *C* не отправляет данных, пользователь *B* не передает данных в течение второго интервала, а пользователь *A* – в течение третьего. В случае использования TDMA с фиксированным распределением все интервалы кадра распределены заранее. Если «владелец» интервала не передает данных в течение указанного промежутка времени, данный интервал не используется. На рис. 1.9, б показан поток данных и неиспользованные интервалы. Если требования пользователей непредсказуемы, как в приведенном примере, то должны применяться более эффективные методы с динамическим распределением интервалов. Таких методов существует несколько – применение систем с коммутацией пакетов, статистических мультиплексоров или концентраторов. Данные системы позволяют достигнуть результата, изображенного на рис. 1.9, в, где пропускная способность системы остается постоянной благодаря использованию всех доступных временны́х интервалов.

1.1.3. Распределение ресурса связи в FDMA и TDMA

На рис. 1.10 представлена комбинированная схема распределения частотно-временного ресурса связи – *комбинированный FDMA/TDMA*. Рассмотрим случай равномерного пропорционального распределения полосы шириной *W* между *M* группами пользователей. Частотный диапазон будем считать разбитым на полосы шириной *W*/*M* Гц, которые будут постоянно доступны соответствующим группам. Аналогично для распределения временны́х интервалов ось времени разбивается на кадры продолжительностью *T*. Каждый из кадров разбивается на *N* интервалов длительностью *T*/*N* [3].

Предположим, что активность пользователей синхронизирована по времени и распределенные интервалы периодично расположены в кадрах. Каждый пользователь может передавать данные, когда начинается его

интервал времени, а также на протяжении данного интервала пользователь может использовать выделенную полосу частот. Временной интервал (*n*, *m*) однозначно задается как *m*-й интервал кадра *n* (рис. 1.10):

$$\text{временной интервал}\,(n,m) = nT + \frac{(m-1)T}{N} \le t \le nT + \frac{mT}{N},\ m = 1, 2, ..., N\,. \quad (1.4)$$

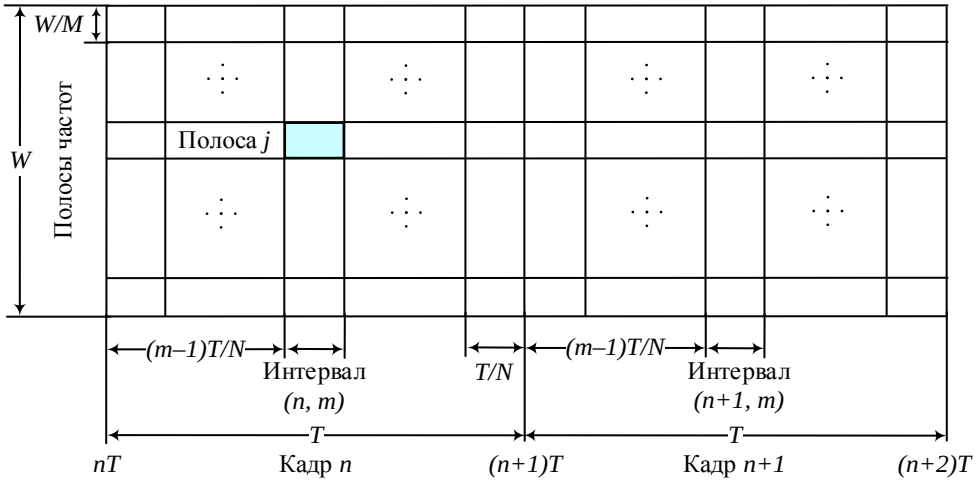


Рис. 1.10. Ресурс связи: частотно-временное распределение по каналам

Длительность *n*-го кадра, *T* – это интервал $\left[nT,(n+1)T\right]$. Область сигнала (рис. 1.10) является пересечением временного интервала (*m*, *n*) и частотного диапазона (*j*). Предположим, что система модуляции/кодирования выбрана так, что полная полоса *W* ресурса связи может поддерживать скорость передачи данных *R* бит/с. Для любого частотного диапазона, содержащего полосу *W*/*M* Гц, соответствующая скорость передачи данных будет составлять *R*/*M* бит/с. Технология FDMA позволяет использовать M диапазонов с шириной полосы 1/*M* полной ширины полосы ресурса связи, а TDMA – полный диапазон частот для каждого из *N* интервалов времени, при этом длительность каждого интервала составит 1/*N* длительности кадра.

1.1.4. Сравнение производительности FDMA и TDMA

Скорость передачи данных FDMA и TDMA. Пусть ресурс связи поддерживает скорость передачи данных *R* бит/с. На рис. 1.11, а полоса системы разделена на *M* ортогональных полос частот; следовательно, все *M* источников T_i ($1 \le i \le M$) могут одновременно передавать со скоростью *R*/*M* бит/с каждый. На рис. 1.11, б кадр разделен на *M* ортогональных временны́х интервалов; таким образом, каждый из *M* источников может передавать со скоростью *R* бит/с, что *в M раз больше скорости передачи пользователя FDMA за время (1/M)*. В обоих случаях T_i передает со средней скоростью *R*/*M* бит/с.

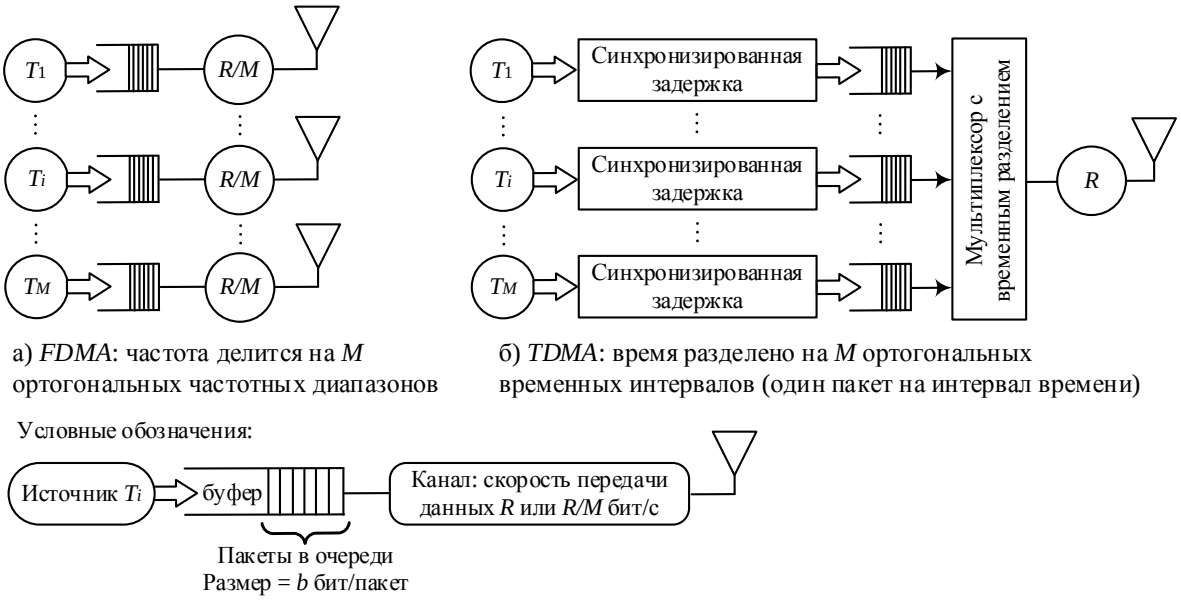


Рис. 1.11. Сравнительное представление технологий *FDMA*/*TDMA*

Пусть информация, передаваемая каждым источником, собирается в b-битовые пакеты. В случае FDMA b-битовые пакеты передаются за T секунд по каждому из M непересекающихся каналов. Таким образом, полная скорость передачи данных может быть представлена в следующем виде:

$$R_{FD} = M\frac{b}{T} \text{ бит/с}. \qquad (1.5)$$

При использовании TDMA каждым источником за T/M секунд передается b бит. Следовательно, скорость передачи данных равна

$$R_{TD} = \frac{b}{T/M} \text{ бит/с}. \qquad (1.6)$$

Поскольку (1.5) и (1.6) идентичны, можно сделать вывод о том, что обе системы обладают одинаковой скоростью передачи данных – R бит/с:

$$R_{FD} = R_{TD} = R = Mb/T \text{ бит/с}. \qquad (1.7)$$

Задержка сообщений в системах FDMA и TDMA. Несмотря на некоторые различия, FDMA и TDMA не отличаются по скорости передачи. Однако среднее время задержки при TDMA меньше, чем при FDMA.

Предположим, что при FDMA диапазон частот разбит на M ортогональных полос; при TDMA кадр разделен на M ортогональных временных интервалов. Для анализа времени задержки рассмотрим случай детерминистических источников данных. Предположим, что ресурс связи используется на 100 %. Тогда все частотные диапазоны при FDMA и все временны́е

интервалы при TDMA будут заполнены пакетами данных. Будем считать, что отсутствуют дополнительные задержки, связанные с защитными полосами или интервалами. Тогда время задержки сообщения можно выразить:

$$D = w + \tau, \tag{1.8}$$

где w – среднее время ожидания пакета (до передачи), τ – время передачи пакета. При FDMA каждый пакет пересылается в течение T секунд; передача пакета для технологии FDMA будет следующей:

$$\tau_{FD} = T. \tag{1.9}$$

При использовании TDMA каждый пакет пересылается в течение временного интервала T/M секунд. С помощью уравнения (1.7) время передачи пакета можно выразить следующим образом:

$$\tau_{TD} = \frac{T}{M} = \frac{b}{R}. \tag{1.10}$$

Поскольку каналы FDMA доступны постоянно, а пакеты пересылаются непосредственно после создания, время ожидания w_{FD} составляет

$$w_{FD} = 0. \tag{1.11}$$

На рис. 1.12 сравниваются потоки данных для схем FDMA и TDMA. Как показано на рис. 1.12, а, при использовании TDMA временные интервалы пользователей начинаются в разных точках кадра протяженностью T секунд. Пакет S_{mk} отправляется по прошествии $(m-1)T/M$, $1 \le m \le M$ секунд после создания пакета. Таким образом, для TDMA среднее время ожидания пакета перед отправкой составит

$$w_{TD} = \frac{1}{M}\sum_{m=1}^{M}(m-1)\frac{T}{M} = \frac{T}{M^2}\sum_{n=0}^{M-1} n = \frac{T}{M^2}\frac{(M-1)M}{2} = \frac{T}{2}\left(1-\frac{1}{M}\right). \tag{1.12}$$

Для сравнения среднего времени задержки D_{FD} и D_{TD} при использовании FDMA и TDMA, соответственно, подставим уравнения (1.9) и (1.11) в (1.8) и уравнения (1.10) и (1.12) в (1.8). В результате получим:

$$D_{FD} = T; \tag{1.13}$$

$$D_{TD} = \frac{T}{2}\left(1-\frac{1}{M}\right) + \frac{T}{M} = D_{FD} - \frac{T}{2}\left(1-\frac{1}{M}\right). \tag{1.14}$$

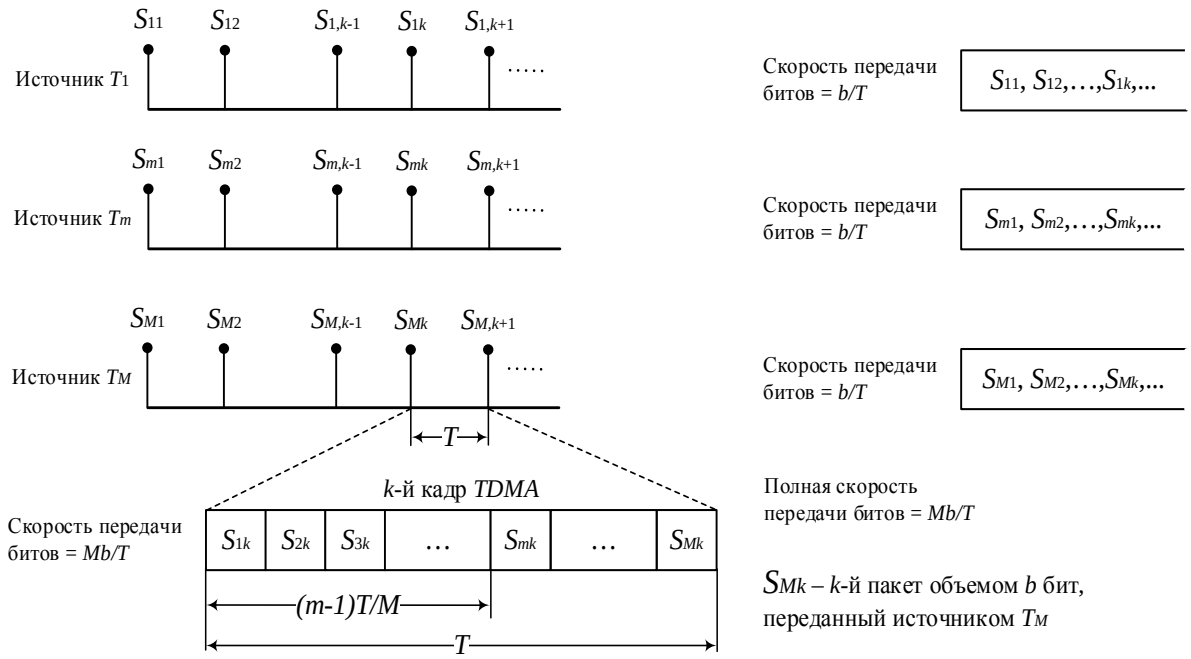


Рис. 1.12. Распределение по каналам:
а) TDMA; б) FDMA

С помощью (1.7) (1.14) можно записать в следующем виде:

$$D_{TD} = D_{FD} - \frac{b}{2R}(M-1). \qquad (1.15)$$

Результат (1.15) свидетельствует о том, что FDMA значительно уступает TDMA по времени задержки сообщения.

1.1.5. Организация кодового разделения каналов FHSS

На рис. 1.13 приводится иллюстрация метода множественного доступа, являющегося результатом совмещения FDMA и TDMA. Этот метод называется *множественным доступом с кодовым разделением* (code-division multiple access – CDMA). CDMA является практическим приложением методов *расширения спектра* (spread spectrum – SS), которые можно разделить на две основные категории: расширение спектра методом *прямой последовательности* (direct sequence – DS) и расширение спектра методом *скачкообразной перестройки частоты* (frequency hopping – FH) [3].

Простейший пример CDMA с перестройкой частоты изображен на рис. 1.13. В каждом из коротких временных интервалов происходит перераспределение частотных диапазонов: в течение интервала 1 сигнал 1 использует диапазон 1, сигналы 2 и 3 – диапазоны 2 и 3. Во время интервала 2 сигнал 1 «перескакивает» в диапазон 3, сигнал 2 – в диапазон 1, сигнал 3 – в диапазон 2 и т. д. Таким образом, ресурс связи используется полностью, причем диапазоны пользователей перераспределяются в каждый после-

дующий момент времени. Каждому пользователю присваивается псевдошумовой (pseudonoise – PN) код, который указывает последовательность перестройки частоты. Псевдошумовые коды ортогональны друг другу.

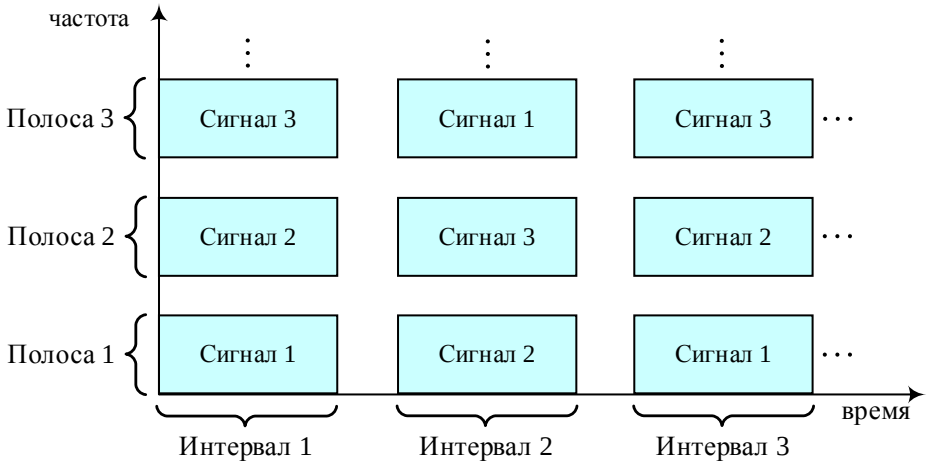
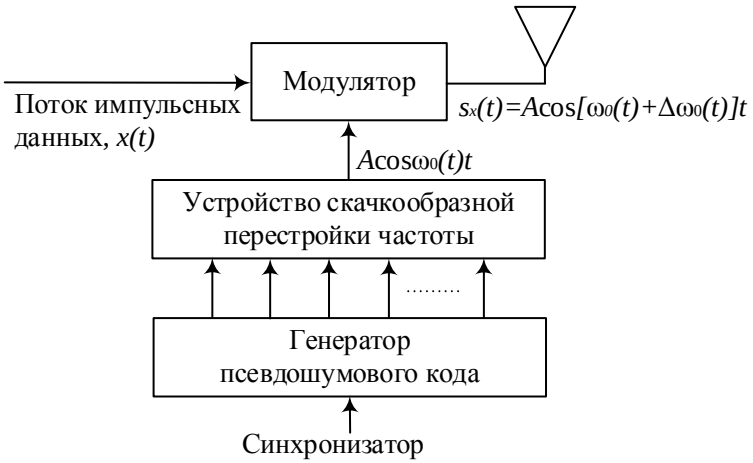


Рис. 1.13. Уплотнение с кодовым разделением (FHSS)

На рис. 1.14, показан процесс модуляции с использованием перестройки частоты. Во время каждого изменения частоты генератор псевдошумовой последовательности направляет кодовую последовательность на *устройство скачкообразной перестройки частоты*. Данное устройство выдает одну из допустимых для скачка частоту. Допустим, что используется *M*-арная частотная манипуляция (*M-ary frequency shift keying* – MFSK). При обычной системе MFSK данные модулируют несущую волну с *фиксированной* частотой. В случае MFSK с перестройкой частоты (FH-MFSK) частота несущей скачет по всему диапазону частот. FH-модуляцию на рис. 1.14 можно рассматривать как процесс, состоящий из двух этапов: модуляции данных и модуляции перестройки частоты. Указанные действия могут быть совмещены – в этом случае модулятор на основе псевдошумового кода и собственно данных генерирует тон передачи.



Приведем уник Рис. 1.14. Процесс модуляции FH-CDMA

Конфиденциальность. Если код группы пользователей известен лишь разрешенным членам этой группы, CDMA обеспечивает конфиденциальность связи, поскольку несанкционированные лица, не имеющие кода, не могут получить доступ к передаваемой информации.

Каналы с замираниями. Если для определенной части используемого спектра характерно замирание, сигналы в данной части будут ослабленными. При FDMA пользователь данной части спектра может испытывать постоянные затруднения со связью. При схеме FH-CDMA пользователь будет испытывать аналогичные проблемы только при изменении частоты в соответствующую часть спектра. Таким образом, возможные проблемы со связью равномерно распределятся между всеми пользователями.

Сопротивляемость подавлению. В течение времени между изменениями частоты полоса сигнала идентична полосе обычной схемы MFSK, т. е. обычно равна минимальной ширине полосы символов MFSK. В то же время в течение нескольких временных интервалов система совершает скачки в диапазоне частот, ширина которого намного превышает ширину полосы данных. Такое использование полосы называется расширением спектра.

Гибкость. Наиболее важным преимуществом CDMA, по сравнению с TDMA, является отсутствие необходимости синхронизации одновременно передающих устройств. Разные передачи не влияют на ортогональность процессов передачи с различными кодами.

Контрольные вопросы

1. Поясните понятия разделения каналов и множественного доступа.
2. Приведите основные способы распределения ресурсов связи.
3. Приведите условие ортогональности сигналов во временной/частотной области.
4. Приведите особенности функционирования уплотнения с частотным разделением.
5. Зачем нужны защитные полосы при уплотнении с частотным разделением?
6. Проиллюстрируйте пример смешивания сигналов при частотном уплотнении.
7. Проиллюстрируйте организацию группы, супергруппы и сверхгруппы при частотном уплотнении.
8. Приведите пример реализации метода FDMA в системах спутниковой связи.
9. Как функционирует нерегенеративный ретранслятор?
10. Приведите особенности функционирования уплотнения с временным разделением.
11. Зачем нужны защитные интервалы при уплотнении с временным разделением?
12. Приведите пример реализации метода TDMA в системах спутниковой связи.
13. Проиллюстрируйте принцип работы метода TDM/TDMA с фиксированным распределением.
14. Проиллюстрируйте принцип коммутации пакетов с временным разделением при пульсирующем трафике.
15. При каких обстоятельствах схема TDM/TDMA с фиксированным распределением оказывается эффективной, а при каких – нет?
16. Приведите особенности функционирования комбинированной схемы FDMA/TDMA.
17. Приведите и поясните выражение для скорости передачи данных в схеме FDMA.

18. Приведите и поясните выражение для скорости передачи данных в схеме TDMA.

19. Приведите и поясните выражение для задержки передачи данных в схеме FDMA.

20. Приведите и поясните выражение для задержки передачи данных в схеме TDMA.

21. Как организуется кодовое разделение каналов при расширении спектра методом скачкообразной перестройки частоты?

22. Приведите и поясните порядок формирования сигнала в системе FH-CDMA.

23. За счет чего схема FH-CDMA обеспечивает конфиденциальность?

24. За счет чего схема FH-CDMA оказывается устойчивой к замиранию?

25. За счет чего схема FH-CDMA оказывается устойчивой к радиоподавлению?

Задачи

1. Разработайте набор сигналов FDM, состоящий из 5 каналов передачи речи, каждый в диапазоне 300–3400 Гц. Уплотненный набор сигналов должен состоять из инвертированных боковых полос и занимать спектральную область от 30 до 50 кГц. а) Изобразите составной спектр, указав отдельные спектры и положение защитных полос. б) Изобразите блок-схему, показывающую процессы смешивания частот и фильтрования, а также параметры местного гетеродина приемника.

2. Средняя величина задержки сообщения в схеме TDMA меньше, чем в схеме FDMA. Какими будут практические результаты уменьшения времени задержки в схеме TDMA (как функции времени передачи кадра) для спутникового канала с односторонним радиусом действия 36000 км? Для каких значений времени передачи кадра схема TDMA будет иметь значительное преимущество перед FDMA?

1.2. Лабораторная работа. Кодовое разделение каналов методом DSSS

Цель работы: реализовать метод кодового разделения каналов DSSS и визуализировать процедуры мультиплексирования и демультиплексирования.

1.2.1. Организация кодового разделения каналов DSSS

Наиболее известны три метода многостанционного доступа с разделением каналов: а) частотное разделение каналов – каналы разделяются по частоте (на физическом уровне); б) временное разделение каналов – каналы разделяются по времени (на физическом уровне); в) кодовое разделение каналов – каналы разделяются кодами (на канальном уровне).

Частотное разделение каналов (рис. 1.15): а) доступная полоса частот распределяется между всеми станциями; б) каждая станция использует для передачи выделенную полосу частот.

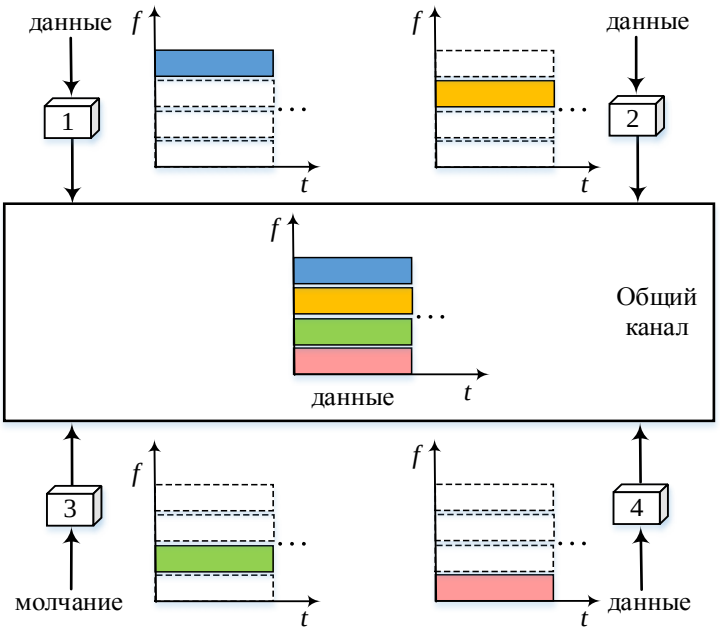


Рис. 1.15. Многостанционный доступ
с частотным разделением каналов

Временное разделение каналов (рис. 1.16): а) доступная полоса частот используется полностью всеми станциями с разделение по времени; б) каждая станция использует для передачи выделенный временной интервал.

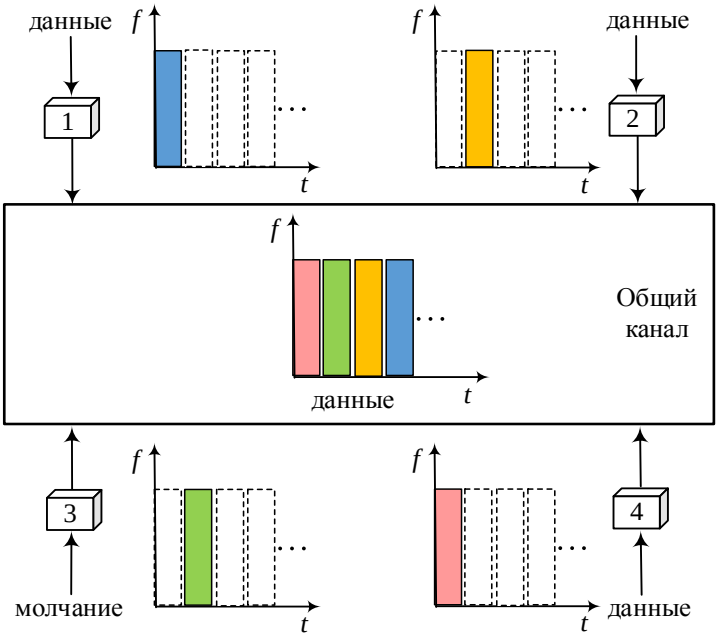


Рис. 1.16. Многостанционный доступ
с временным разделением каналов

Особенности кодового разделения каналов. Все станции занимают общую полосу частот и могут передавать данные без разделения по времени. Каждой станции назначается код, состоящий из последовательности

элементов, называемых чипами (Chips). *Чип* – это элементарный символ, длительность которого во много раз меньше информационного символа.

Каждый передаваемый информационный символ d на станции умножается на соответствующий этой станции код c и кодированные данные $d \cdot c$ от каждой станции суммируются в общем канале (рис. 1.17).

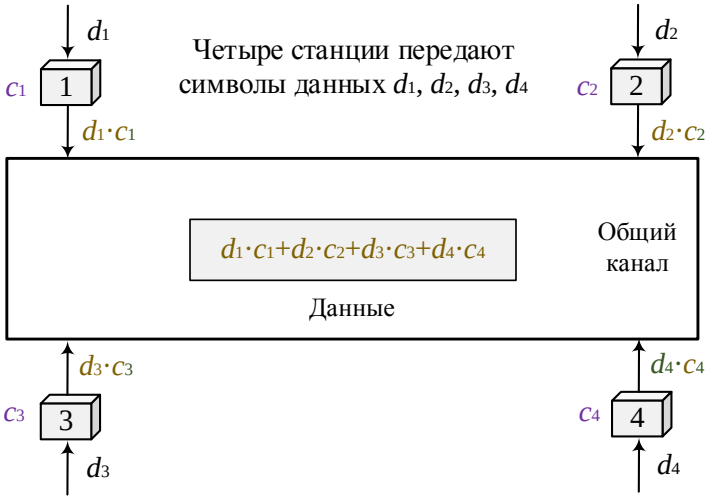


Рис. 1.17. Идея кодового разделения

Бит данных 0 → символ –1

Бит данных 1 → символ +1

Тишина → символ 0

Рис. 1.18. Биполярное представление битов в CDMA

Используется биполярное кодирование информационных символов (рис. 1.18): а) если станции необходимо передать бит 0, то она передает символ «–1»; б) если станции необходимо передать бит 1, то она передает символ «+1»; в) когда станция не передает сигнала, т. е. формируется символ «0».

Предположим, что каждой станции поставлены следующие соответствующие последовательности чипов (коды c). Рассмотрим пример (рис. 1.19), как 4 станции разделяют среду передачи в течение интервала 1 бит (станции 1 и 2 посылают 0, станция 4 посылает 1, и станция 3 молчит).

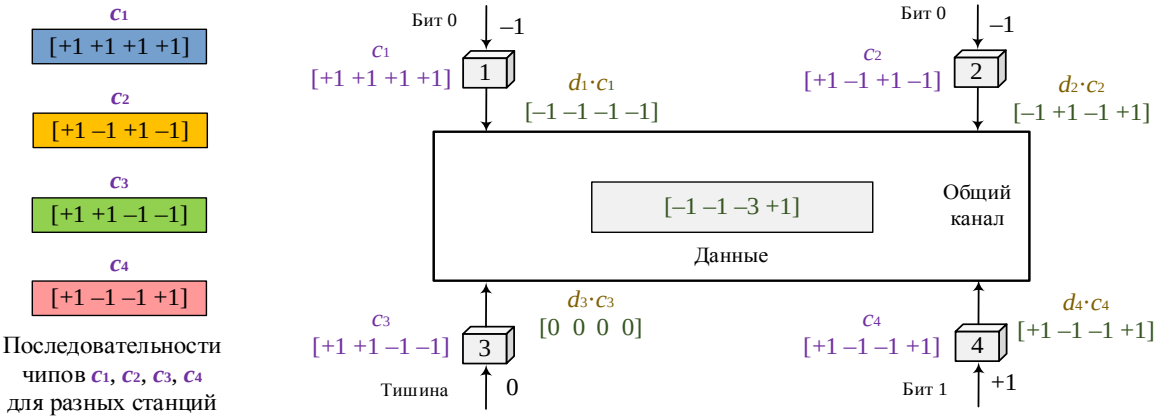


Рис. 1.19. Пример кодового разделения

Алгоритм муль

:

• мультиплексор параллельно принимает по одному кодированному символу информационных данных от каждой станции: –1, –1, 0 и +1;

• символ «–1», переданный станцией 1, умножается на каждый чип кода С1, результат – (–1, –1, –1, –1); символ «–1», переданный станцией 2, умножается на каждый чип кода С2, результат – (–1, +1, –1, +1); такое же правило умножения применяется для станций 3 и 4;

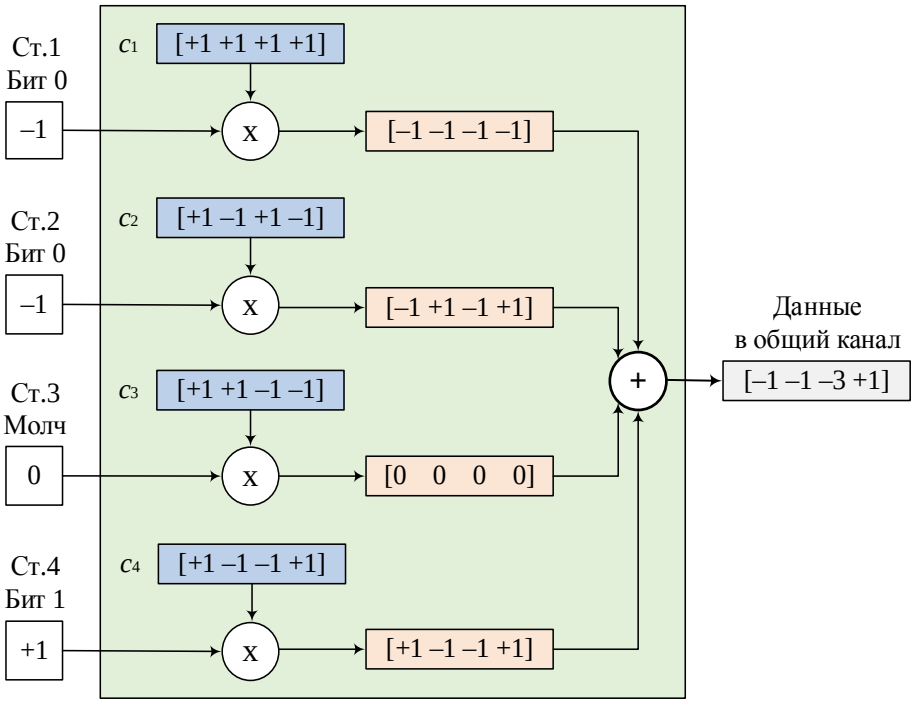


Рис. 1.20. Алгоритм мультиплексирования в CDMA

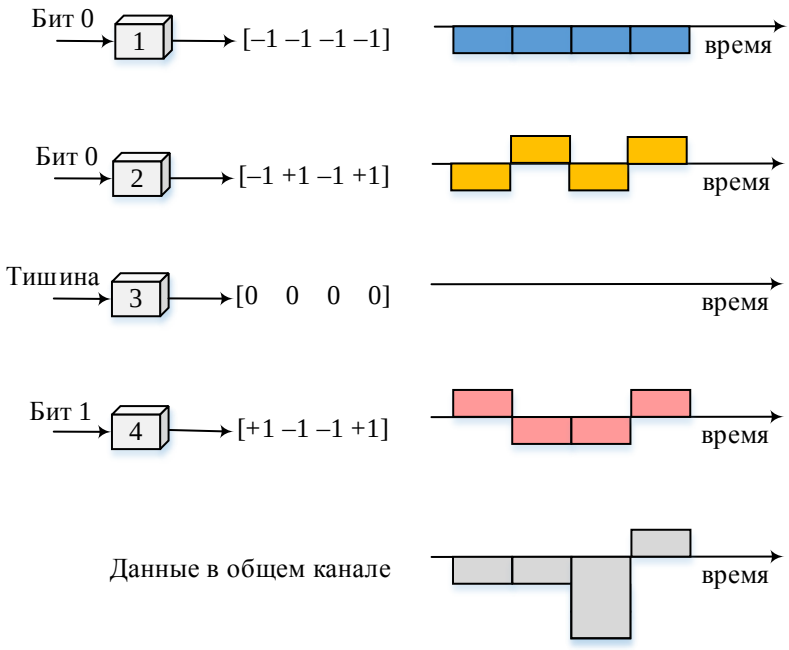


Рис. 1.21. Формирование цифрового сигнала в CDMA

• сложение чипов одного порядка каждой из последовательности приводит к новой последовательности (–1, –1, –3, +1), которая передается в общую среду; сложение последовательности чипов каждой станции производится по правилу арифметического сложения;

• аналогично обрабатывается каждый бит из информационных последовательностей от всех станций;

• результирующая последовательность может дополнительно кодироваться и модулироваться перед передачей в линию.

Алгоритм демультиплексирования (рис. 1.22, 1.23):

• демультиплексоры всех приемников принимают одну и ту же последовательность чипов, переданную по общей среде. Каждый приемник поэлементно умножает эту последовательность на свой код;

• результирующая последовательность чипов в каждом приемнике поэлементно складывается Возможные варианты результатов в примере: «+4», «–4» или «0»;

• результат сложения делится на количество станций и затем приемник выделяет исходный бит. Сложение элементов последовательности на каждой станции производится по правилу арифметического сложения. Из последовательности каждым приемником извлекаются свои биты, процесс извлечения побитный;

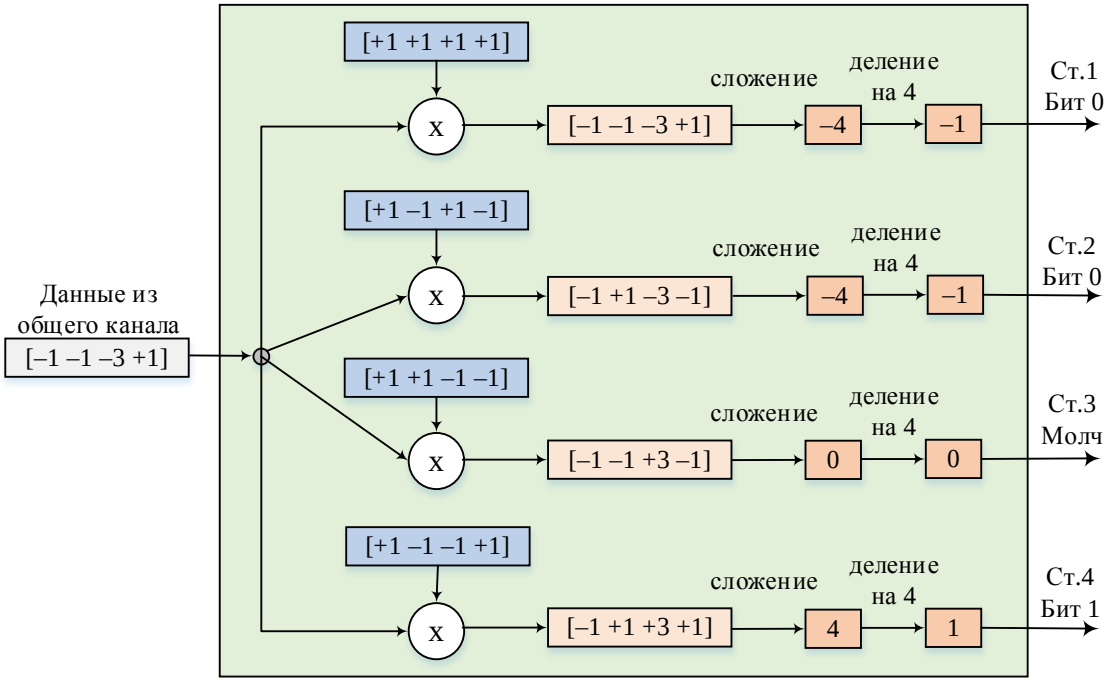


Рис. 1.22. Алгоритм демультиплексирования в CDMA

• станция 2 (рис. 1.23) принимает мультиплексированные данные и умножает на свой код (+1, –1, +1, –1). Промежуточный результат (–1, +1, –3, –1) арифметически складывается поэлементно. Результат сложения –4 делится на 4 и извлекается исходный бит 0.

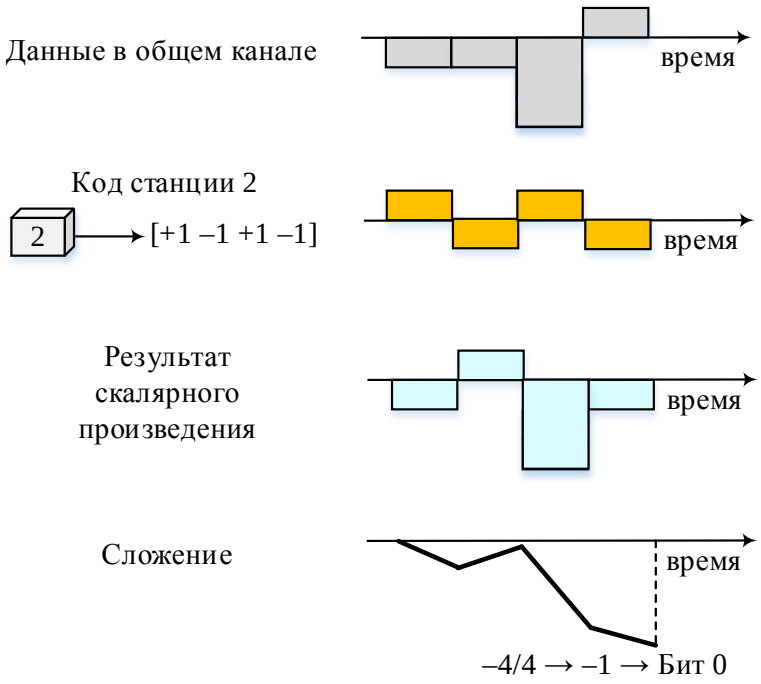


Рис. 1.23. Демультиплексирования составного сигнала
в исходный сигнал на станции 2

Результат кодового разделения. В разобранном примере каждая станция принимает то, что ей передано. Станция 3 не принимает данный, так как соответствующий отправитель молчит. По общему каналу передается только одна последовательность, сформированная на передаче как сумма исходных последовательностей. Каждый приемник извлекает свои собственные данные из этой суммы.

Ортогональные коды: коды С (последовательность чипов) для каждой станции должны выбираться не случайно, а по определенному правилу. Такие коды называются ортогональными кодами.

Формирование кодов:

- для формирования кодов используются таблицы Уолша (Walsh table) и такие коды называются кодами Уолша;
- таблицы Уолша являются двумерными квадратными таблицами; последовательность чипов одного ряда является кодом одной станции;
- таблица Уолша W_1 для одноэлементной последовательности имеет один ряд и один столбец; элементом таблицы может быть «–1» или «+1»;
- если задана/известна таблица W_N порядка N, то можно сформировать таблицу W_{2N} порядка $2N$

$$W_{2N} = \begin{bmatrix} W_N & W_N \\ W_N & \overline{W_N} \end{bmatrix},$$

где $\overline{W_N}$ – инверсия элемента (каждый «+1» заменяется на «–1» и наоборот).

Рассмотрим формирование таблиц W_1, W_2 и W_4 из таблицы W_1. Таблица W_2 формируется из четырех таблиц W_1, при этом последний чип является инверсией таблицы W_1. Таблица W_4 формируется из четырех таблиц W_2, при этом элементы новой таблицы являются предыдущими таблицами, но все элементы последней таблицы инвертируются. Аналогично, таблицу W_8 можно сформировать из четырех таблиц W_4 и т. д.:

$$W_1 = [+1]; \qquad W_2 = \begin{bmatrix} +1 & +1 \\ +1 & -1 \end{bmatrix}; \qquad W_4 = \begin{bmatrix} +1 & +1 & +1 & +1 \\ +1 & -1 & +1 & -1 \\ +1 & +1 & -1 & -1 \\ +1 & -1 & -1 & +1 \end{bmatrix}.$$

Элементы строк в таблице Уолша образуют коды Уолша, скалярное произведение любой пары строк должно быть равно 0.

Строки в таблицах образуют коды для станций (W_2 формирует для двух станций следующие коды: «+1, +1» и «+1, –1», таблица W_4 формирует коды для четырех станций, и т. д.).

Свойства ортогональных кодов:

- умножение последовательности на –1 инвертирует ее элементы (+1 переходит в –1, а –1 переходит в +1);
- умножение двух последовательностей (кодов) со сложением результирующих элементов является операцией скалярного произведения и обозначается точкой «·». Если две исходные последовательности одинаковые, то результат скалярного произведения равен N, где N – общее количество последовательностей, например, $c_1 \cdot c_1 = N$, если они различны, то результат равен нулю, например, $c_1 \cdot c_2 = 0$;
- скалярное произведение инверсных последовательностей равно $-N$, например, $c_1 \cdot (-c_1) = -N$.

Пример. Подтвердим второе свойство ортогональных кодов для примера с W_4.

Решение:

- скалярное произведение кода на самого себя должно быть равно N. Покажем на примере кода c_3:

$$c_3 \cdot c_3 = [+1,+1,-1,-1] \cdot [+1,+1,-1,-1] = 1+1+1+1 = 4;$$

- скалярное произведение двух разных кодов таблицы Уолша должно быть равно 0. Покажем на примере c_2 и c_3:

$$c_2 \cdot c_3 = [+1,-1,+1,-1] \cdot [+1,+1,-1,-1] = 1-1-1+1 = 0.$$

Пример кодирования в мультиплексоре: в мультиплексоре каждая станция передает в общий канал соответствующую кодированную последовательность $d \cdot c$:

- станция 1 передает последовательность $-c_1$ (символ –1 умножается на код c_1), станция 2 передает последовательность $-c_2$, станция 3 передает пустую последовательность (все нули) и станция 4 передает последовательность c_4;
- смешивание всех последовательностей в общем канале производится арифметическим суммированием элементов (чипов) этих последовательностей $s = -c_1 - c_2 + c_4$.

Пример декодирования в демультиплексоре: в демультиплексоре все станции принимают последовательность s:

- станция 1 вычисляет скалярное произведение

$$s \cdot c_1 = \left(-c_1 - c_2 + c_4\right) \cdot c_1 = -c_1 \cdot c_1 - c_2 \cdot c_1 + c_4 \cdot c_1 = -4 + 0 + 0 = -4.$$

Деление результата на 4 дает –1, т. е. принимается бит 0;

- для станции 2: $s \cdot c_2 = -4$, деление на 4 дает –4, значит принимается бит 0;
- для станции 3: $s \cdot c_3 = 0$, деление на 4 дает 0, значит нет приема;
- для станции 4: $s \cdot c_4 = 4$, деление на 4 дает +1, значит принимается бит 1.

1.2.2. Моделирование кодового разделения каналов DSSS

Реализуем алгоритмы мультиплексирования и демультиплексирования в среде Matlab. На рис. 1.24 представлен пример мультиплексирования чипов четырех абонентов (скрипт 1.1).

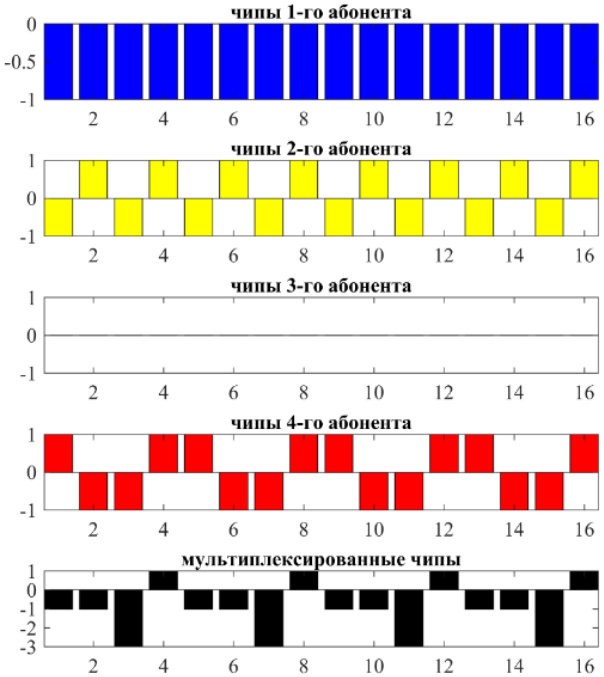


Рис. 1.24. Пример кодирования символов данных четырех абонентов

Скрипт 1.1. Алгоритмы мультиплексирования и демультиплексирования CDMA

```
% Мультиплексирование и демультиплексирование в системах связи
% с кодовым разделение каналов (см. пример из лекции)
clc; clear all;
% передаваемые символы данных для 4-х абонентов
D = [ -1 -1  -1 -1  ;
      -1 -1  -1 -1;
       0  0   0  0;
       1  1   1  1];
 % уникальные коды Уолша для 4-х абонентов
C = [ 1  1  1  1 ;
      1 -1  1 -1 ;
      1  1 -1 -1 ;
      1 -1 -1  1];
 % параметры
M = length(C); % длина кода Уолша в чипах
Y = size(D);   % размер матрицы D
N = Y(1);      % число абонентов (символьных потоков) / число строк матр.D
I = Y(2);      % число символов в потоке / число столбцов матрицы D
T = [];        % кодированный поток символов абонентов (длиной I*M)
R = [];        % кодированный поток символов абонентов (длиной I)
MUX = [];      % мультиплексированный поток символов
DEMUX = [];    % демультиплексированный поток символов

% передаваемые символы данных для 4-х абонентов
'передаваемые символы данных абонентов:', D
'уникальные коды Уолша абонентов:', C
% G - матрица кодированных и мультиплексированных символов (в каждой
% строке G записаны кодированные и мультиплексированные чипы от всех
% абонентов в данный интервал времени, соответствующий одному символу
G = zeros(I,M);

% мультиплексирование символов данных
for n = 1:N         % цикл по абонентам
    % Z - матрица кодированных символов (в каждой строке
    % Z записаны кодированные чипы от одного абонента
    % в данный интервал времени, соответствующий одному символу
    Z = zeros(I,M);
    Tn=[]; % вектор кодированных символов одного абонента
    for i = 1:I         % цикл по символам данных
        for m = 1:M         % цикл по чипам кода (Уолша)
            Z(i,m) = D(n,i)*C(n,m);    % почиповое кодирование символов
        end
        Tn=[Tn Z(i,:)];
    end
    T(n,:)=Tn; % матрица кодированных символов всех абонентов
    G = G + Z;
end

% мультиплексированный поток символов
for i = 1:I
    MUX = [ MUX G(i,:) ];
end
'мультиплексированный поток символов:', MUX
% 'мультиплексированный поток символов:', sum(T,1)

% демультиплексирование символов данных
for n = 1:N         % цикл по абонентам
    QSUM = zeros(1,I);    % вектор декодированных и демультиплексированных
                          % символов одного абонента (без масштабирования)
```

```
    % Q - матрица декодированных и мультиплексированных символов (в каждой
    % строке Q записаны декодированные и мультиплексированные чипы от всех
    % абонентов в данный интервал времени, соответствующий одному символу
    Q   = zeros(I,M);
    Rn=[]; % вектор декодированных
           % и демультиплексированных символов одного абонента
    for i = 1:I          % цикл по символам данных
        for m = 1:M          % цикл по чипам кода (Уолша)
            Q(i,m) = G(i,m) * C (n,m);
            QSUM(i) = QSUM(i) + Q (i,m);
        end
        Rn=[Rn QSUM(i)/M];
    end
    R(n,:)=Rn; % матрица декодированных символов всех абонентов
    DEMUX = [DEMUX; QSUM/M];
end
'демультиплексированный поток символов:', DEMUX
% 'демультиплексированный поток символов:', R

% визуализация процесса мультиплексирования чипов всех абонентов
figure(2);
subplot(5,1,1); bar(T(1,:),'b');title('чипы 1-го абонента'); axis('tight');
subplot(5,1,2); bar(T(2,:),'y');title('чипы 2-го абонента'); axis('tight');
subplot(5,1,3); bar(T(3,:),'g');title('чипы 3-го абонента'); axis('tight');
subplot(5,1,4); bar(T(4,:),'r');title('чипы 4-го абонента'); axis('tight');
subplot(5,1,5); bar(MUX,'k');
title('мультиплексированные чипы'); axis('tight');
```

Контрольные вопросы

1. Сформулируйте особенности метода частотного разделения каналов.
2. Сформулируйте особенности метода временного разделения каналов.
3. Сформулируйте особенности метода кодового разделения каналов.
4. Изобразите пример алгоритма мультиплексирования в CDMA.
5. Изобразите пример алгоритма демультиплексирования в CDMA.
6. Сформулируйте свойства ортогональных кодов применительно к реализации метода CDMA.
7. Приведите правило формирования кодов Уолша.
8. Сформируйте код Уолша длиной 8 чипов.
9. Сформируйте коды Уолша длиной 4 чипа и проиллюстрируйте их свойство ортогональности.
10. Визуализируйте символы и чипы для 2/2/4-го абонента в рассмотренном примере.

Задачи

1. Усовершенствовать файл-функцию в части задания матрицы Уолша встроенной функцией в Matlab.
2. Усовершенствовать файл-функцию в части исходного формирования битов с их последующим преобразованием в биполярные символы.

Содержание отчета

1. Цель, задачи. Результаты и выводы по каждому пункту.
2. Листинг файл-функции с комментариями.
3. Алгоритм функционирования ИМ.

2. МЕТОДЫ ДОСТУПА К СРЕДЕ ПЕРЕДАЧИ

2.1. Практическое занятие. Методы доступа к среде передачи

Цель занятия: изучить понятие многостанционного доступа, метод предоставления каналов по требованию, классификацию методов многостанционного доступа, методы управляемого доступа.

Информация об использовании времени, частоты и кодовых комбинация содержится в протоколе/алгоритме множественного доступа (multiple access algorithm – MAA). Основная задача такой системы – своевременное, упорядоченное и эффективное предоставление пользователю услуг связи.

2.1.1. Информационный поток в системах многостанционного доступа

На рис. 2.1 представлена обобщенная блок-схема потока данных между алгоритмом множественного доступа и станцией связи.

Контролировать ММА может центральная (базовая) станция; также контроль может быть распределен между всеми станциями [3].

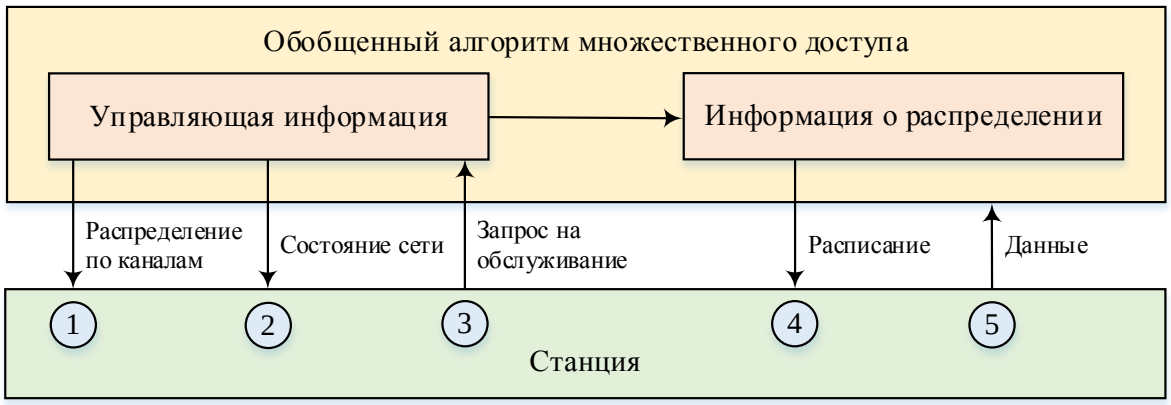


Рис. 2.1. Информационный поток в системах множественного доступа

Передача в общем случае происходит в следующем порядке (рис. 2.1):

1. *Распределение по каналам*. Данный термин относится к распределению информации (например, каналы 1 – *N* могут быть предоставлены пользователю *X*, а каналы (*N* + 1)-*M* – пользователю *Y*).

2. *Состояние сети* (Network State – NS). Этот термин связан с состоянием ресурса связи. Станция получает указания относительно доступности ресурса связи, а также о том, как следует использовать частоту, время, кодовые позиции ресурса для передачи запроса на обслуживание.

3. *Запрос на обслуживание.* Станция передает запрос на обслуживание (например, на выделение ресурса для передачи *m* сегментов сообщения).

4. По получении запроса (запросов) на обслуживание *контроллер* передает станции *расписание*, в соответствии с которым данные должны распределяться в ресурсе связи.

5. Станция передает данные в соответствии с указанным расписанием.

2.1.2. Предоставление каналов по требованию

Системы множественного доступа, позволяющие передающей станции периодически получать доступ к каналу независимо от реальных потребностей, называются системами с *фиксированным распределением*. Существуют также системы с *динамическим распределением*, которые предоставляют *доступ к каналу только при соответствующем запросе*. Их именуют системами множественного доступа с предоставлением каналов по требованию (demand-assignment multiple access – DAMA). Если передача данных станцией связи ведется нерегулярно или скачкообразно, схема DAMA может быть значительно эффективнее схемы фиксированного распределения. На рис. 2.2 обобщаются различия между системой с фиксированным распределением, требуемая пропускная способность которой равна сумме требований всех пользователей, и динамической системой, требуемая пропускная способность которой определяется средними требованиями пользователей.

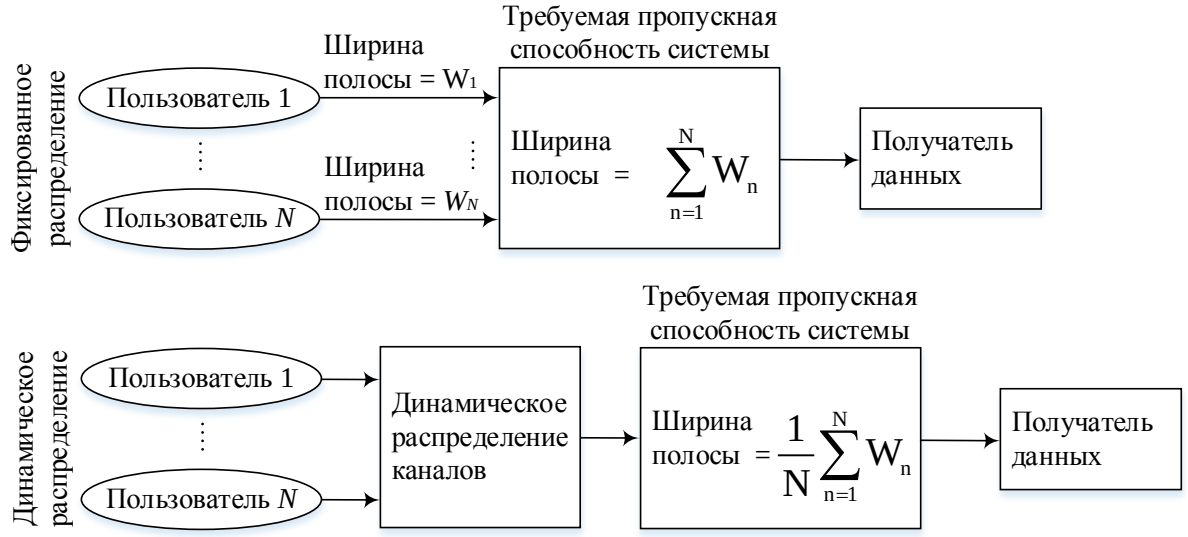


Рис. 2.2. Уменьшение ширины полосы
для систем связи с динамическим распределением каналов

2.1.3. Классификация методов многостанционного доступа

Для соединения нескольких узлов часто используется общая среда передачи. Основными предпосылками функционирования протоколов многостанционного доступа являются следующие: а) в общей среде необходим

протокол многостанционного доступа для координации доступа узлов к общей среде передачи; б) этикет общения определяет правила: несколько человек не говорят одновременно, не перебивают друг друга и т. п., так и протокол МА определяет порядок доступа к общей среде передачи.

Место многостанционного доступа в канальном уровне (рис. 2.3): а) протоколы организации многостанционного доступа относятся к подуровню канального уровня, называемому Подуровнем управления доступом к среде (Medium Access Control – MAC); б) подуровень управления каналом отвечает за организацию канала.

Рис. 2.3. Место многостанционного доступа в канальном уровне

Методы многостанционного доступа можно классифицировать по трем основным группам (рис. 2.4): а) методы произвольного доступа; б) методы управляемого доступа; в) методы разделения каналов.

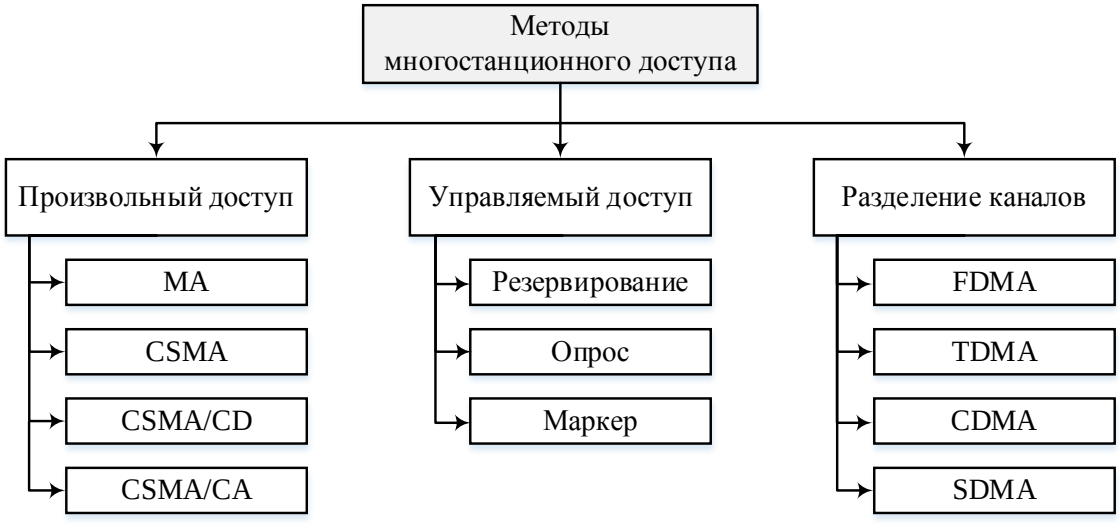


Рис. 2.4. Классификация протоколов многостанционного доступа

2.1.4. Методы управляемого доступа

В методах управляемого доступа станции заранее договариваются друг с другом о праве доступа к среде. Известны три основных метода управляемого доступа: *метод резервирования; метод опроса и маркерный метод*.

Метод управляемого доступа с резервированием представлен на рис. 2.5 и работает следующим образом: а) время делится на интервалы; кадр резервирования в интервале определяет очередность передачи данных; б) номер мини-слота в кадре резервирования соответствует номеру станций в сети; в) перед тем как передать кадр станция помечает свой мини-слот и, тем самым резервирует «место» для передачи в соответствующем интервале.

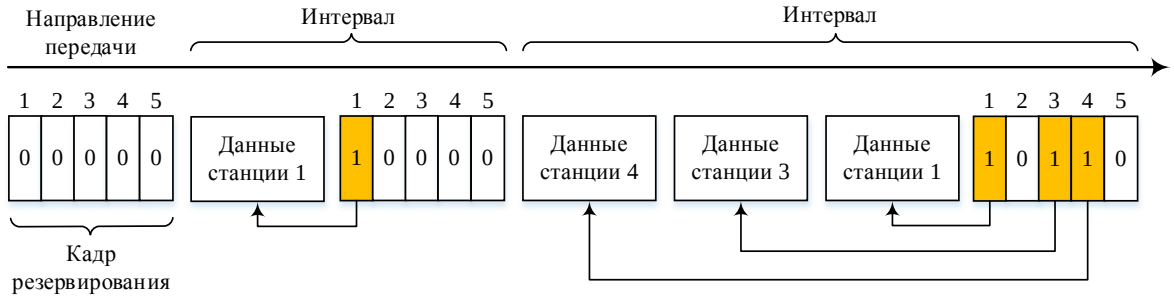


Рис. 2.5. Метод доступа с резервированием

Метод доступа с опросом пригоден в сетях с несбалансированным (ассиметричным) режимом, в которых одна станция является ведущей (primary), а остальные – ведомые (secondary). Имеет следующие особенности функционирования: а) передача данных между любыми станциями осуществляется через ведущую станцию; б) ведущая станция управляет линией, а ведомые следуют ее инструкциям; она определяет, какой станции разрешено использовать канал в заданное время; в) ведущая станция всегда является инициатором сессии (начала передачи) и в зависимости от действий она выполняет две процедуры: *выбор и опрос*; г) процедура выбора – если ведущая станция собирается передать данные, она обращается к соответствующей ведомой станции с просьбой приготовиться к приему; д) процедура опроса – если ведущая станция собирается принять данные, она опрашивает все ведущие станции о намерении их передачи.

Процедура выбора (ведущий передает) представлена на рис. 2.6; используется, когда ведущая станция собирается передать данные; процедура может быть сформулирована следующим образом: а) так как ведущая станция полностью контролирует среду, то если она не передает и не принимает никаких данных, значит среда свободна; б) до передачи данных ведущая станция должна удостовериться о готовности их приема ведомой станцией; в) для этого ведущая станция оповещает о предстоящей передаче специальным кадром выбора (SEL),

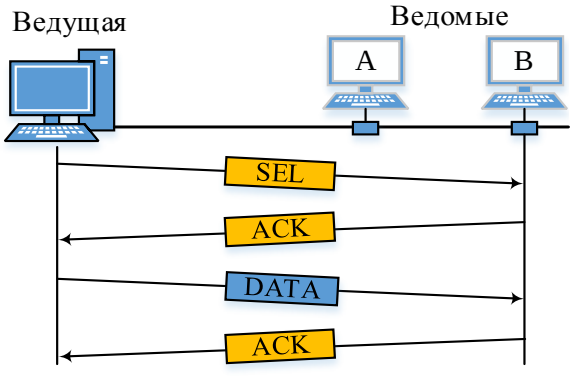


Рис. 2.6. Процедура выбора

содержащим адрес требуемой ведомой станции, и ждет подтверждения (ACK) о статусе готовности; г) после приема ACK производится передача данных (Data) с подтверждением.

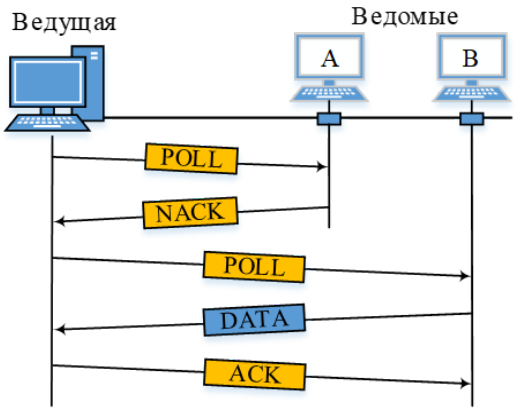


Рис. 2.7. Процедура опроса

Процедура опроса (ведущий принимает) представлена на рис. 2.7; используется ведущей станцией для того, чтобы определить, собирается ли какая-то ведомая станция передавать данные; процедура может быть сформулирована следующим образом: а) процедура опроса используется ведущей станцией для того, чтобы определить, собирается ли какая-то ведомая станция передавать данные; б) опрос производится циклически и все ведомые станции опрашиваются по порядку; в) когда ведущая станция готова к приему данных (т. е. если не передает), она передает кадры опроса (Poll) всем станциям по очереди; г) если ведомая станция не собирается передавать данные, то передает кадр отрицательного подтверждения (NAK); д) когда ведомая станция имеет данные для передачи, то передает кадр данных; е) ведущая станция подтверждает прием данных кадром (ACK); ж) в случае приема кадра NAK ведущая станция опрашивает следующую ведомую станцию.

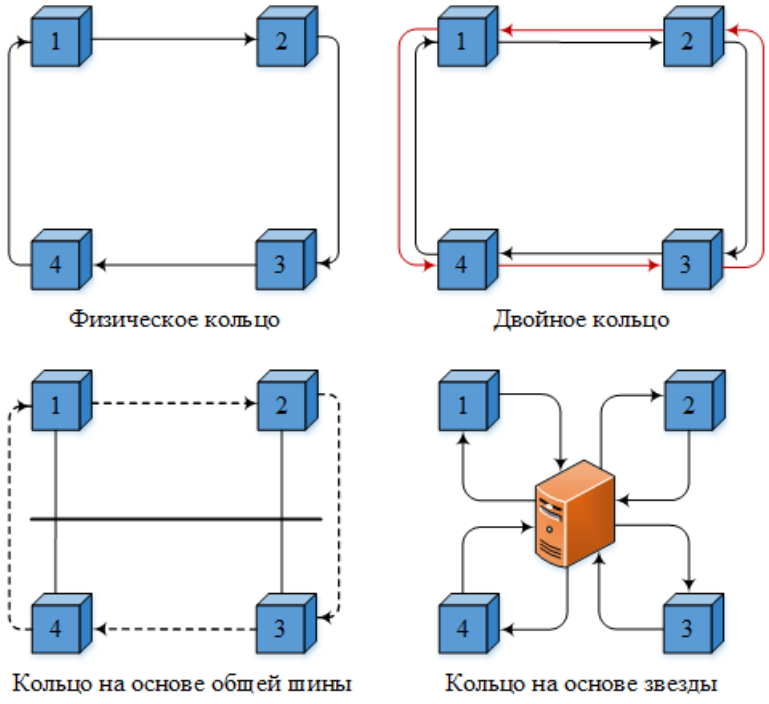


Рис. 2.8. Логическое кольцо и физическая топология в сетях с маркерным доступом

Особенности метода маркерного доступа: а) любая станция имеет возможность передавать данные по приему специального кадра – Маркера; б) метод подразумевает передачу данных по кольцу, т. е. все соединения станций образуют логическое кольцо (хотя физическое соединение может быть другим, рис. 2.8); в) каждая станция принимает кадры от «предшественника» и передает их «последователю».

Процедура маркерного доступа представлена на рис. 2.9 и может быть сформулирована следующим образом: а) когда ни одна из станций не передает данные, маркер циркулирует по кольцу (т. е. кадр маркера передается от одной станции к другой по эстафете); б) любая из станций приняв маркер, «захватывает» его, и, если передавать нечего, «освобождает» маркер, т. е. передает его дальше; в) при необходимости передачи данных, станция ожидает маркер, захватывает его и передает один или более кадров, но не более заданного времени; г) после передачи данных маркер освобождается для следующей станции по кольцу; д) в реальных протоколах добавлены процедуры приоритета и резервирования.

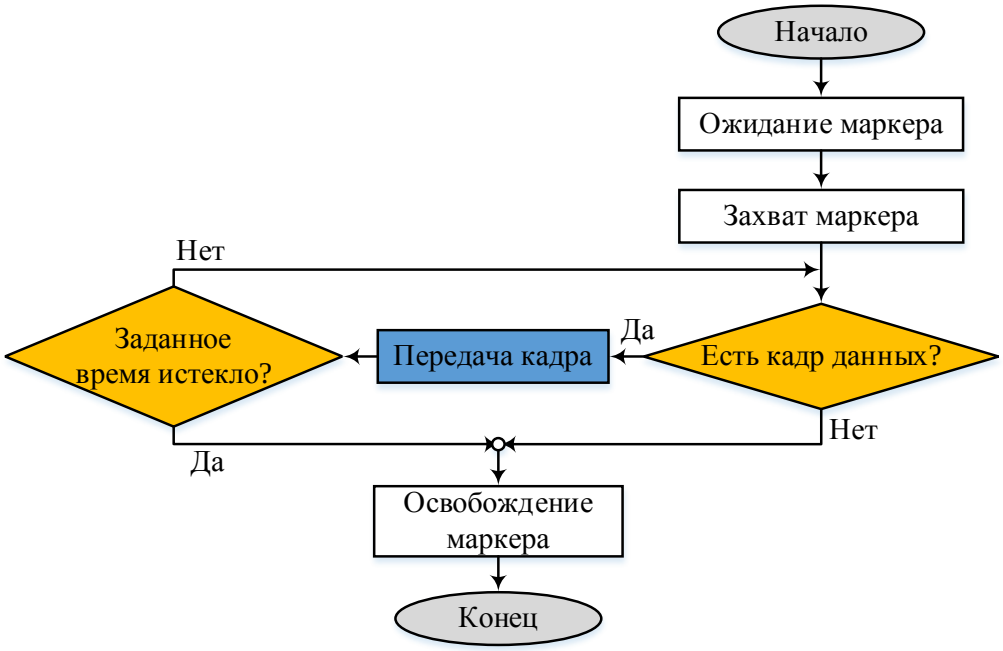


Рис. 2.9. Упрощенная процедура маркерного доступа

Контрольные вопросы

1. Дайте определение понятию алгоритм множественного доступа.
2. Приведите основные процедуры организации информационного потока в системах множественного доступа.
3. Сформулируйте особенности функционирования систем множественного доступа с предоставлением каналов по требованию.
4. За счет чего обеспечивается преимущество систем множественного доступа с предоставлением каналов по требованию по сравнению с системами с фиксированным распределением?
5. Сформулируйте предпосылки работы протоколов многостанционного доступа.
6. Определите место многостанционного доступа в модели OSI.
7. По каким группам классифицируются протоколы многостанционного доступа?
8. Приведите основные методы управляемого доступа.
9. Сформулируйте особенности функционирования доступа с резервированием.
10. Сформулируйте особенности функционирования метода доступа с опросом.

11. Опишите процедуру выбора в методе доступа с опросом.
12. Опишите процедуру опроса в методе доступа с опросом.
13. Сформулируйте особенности функционирования метода маркерного доступа.
14. Приведите примеры организации соединения станций по логическому кольцу.
15. Опишите процедуру функционирования маркерного доступа.

2.2. Практическое занятие. Методы случайного доступа ALOHA

Цель занятия: изучить методы случайного доступа ALOHA, SALOHA, RALOHA.

2.2.1. Алгоритм доступа ALOHA

В 1971 г. Гавайский университет разработал и начал использовать систему ALOHA. Спутник применялся для связи нескольких университетских компьютеров посредством протокола произвольного доступа. Принцип работы системы чрезвычайно прост и включает в себя следующие режимы.

1. *Режим передачи*. Пользователи передают данные в любой момент времени, кодируя свои сообщения с помощью кода обнаружения ошибок.

2. *Режим ожидания*. После передачи пользователь ожидает от приемника подтверждения (acknowledgement – ACK) приема. Иногда передачи пользователей перекрываются во времени, что приводит к возникновению ошибок (коллизий). Ошибки обнаруживаются, и пользователи получают отрицательное подтверждение приема (negative acknowledgement – NAK).

3. *Режим повторной передачи*. После получения сообщения NAK информация передается повторно. Естественно, если пользователи попытаются осуществить повторную передачу непосредственно после возникновения ошибки, конфликтная ситуация может повториться. Поэтому повторная передача производится после *случайной* задержки.

4. *Режим истечения времени ожидания*. Если после передачи пользователь в течение определенного времени не получил сообщения ACK или NAK, производится повторная передача.

Статистика получения сообщений. Предположим, что для работы некоторой системы необходима определенная средняя частота успешного поступления сообщений (пакетов) λ. Вследствие конфликтных ситуаций (коллизий) некоторые из сообщений не будут получены либо будут отклонены. Следовательно, общую частоту поступления сообщений λ_t можно определить как сумму частоты успешного поступления сообщений λ и частоты отклонения данных λ_r:

$$\lambda_t = \lambda + \lambda_r . \qquad (2.1)$$

Обозначим размер сообщения или пакета через b бит. Тогда средний объем успешно переданных данных, иначе говоря пропускную способность канала, ρ', можно представить следующим образом:

$$\rho' = b\lambda \text{ бит/с}. \tag{2.2}$$

Также можно определить полный информационный обмен канала, G':

$$G' = b\lambda_t \text{ бит/с}. \tag{2.3}$$

Если максимальная скорость передачи битов (емкость канала) равна R *бит/с*, *нормированную пропускную способность* можно записать как:

$$\rho = \frac{b\lambda}{R}. \tag{2.4}$$

Также запишем *нормированный полный информационный обмен*:

$$G = \frac{b\lambda_t}{R}. \tag{2.5}$$

Нормированная пропускная способность ρ выражает пропускную способность как часть $(0 \le \rho \le 1)$ емкости канала. Нормированный полный информационный обмен G выражает полный информационный обмен как часть $(0 \le G \le \infty)$ емкости канала. Следует отметить, что G может иметь значения, превышающие 1.

Время передачи пакета может быть выражено в следующем виде:

$$\tau = \frac{b}{R} \text{ секунд/пакет}. \tag{2.6}$$

Подставляя (2.6) в (2.4) и (2.5), можем записать следующее:

$$\rho = \lambda\tau, \tag{2.7}$$

$$G = \lambda_t\tau. \tag{2.8}$$

Пользователь может успешно передавать данные, если ни один из пользователей не начал передачу в течение предыдущих τ секунд. В противном случае возникнет конфликт. Поэтому для успешной передачи каждого сообщения требуется время 2τ.

Статистика получения сообщений пользователями моделируется пуассоновским процессом, согласно которому вероятность поступления k новых сообщений в течение τ секунд описывается распределением:

$$P(k) = \frac{(\lambda\tau)^k e^{-\lambda\tau}}{k!}, k \ge 0, \tag{2.9}$$

где λ – средняя частота поступления сообщений. Поскольку в системе ALOHA пользователи передают данные независимо друг от друга, приве-

денное выше выражение может быть использовано для вычисления вероятности события, когда в течение временного интервала 2τ будет получено точно $k = 0$ других сообщений. Таким образом, получаем P_s – вероятность успешной (бесконфликтной) передачи пользовательского сообщения. Для вычисления P_s подставим в уравнение (2.9) значения λ_t и 2τ:

$$P_s = P(k=0) = \frac{(2\tau\lambda_t)^0 e^{-2\tau\lambda_t}}{0!} = e^{-2\tau\lambda_t}. \quad (2.10)$$

В уравнении (2.1) общая частота поступления сообщений λ_t определялась как сумма частоты успешного поступления сообщения λ и частоты отклонения данных λ_r. Тогда, по определению, вероятность успешного получения пакета может быть выражена в следующем виде:

$$P_s = \frac{\lambda}{\lambda_t}. \quad (2.11)$$

Объединив уравнения (2.10) и (2.11), получаем следующее:

$$\lambda = \lambda_t e^{-2\tau\lambda_t}. \quad (2.12)$$

Подставив в формулу (2.12) выражения (2.7) и (2.8), можно записать

$$\rho = Ge^{-2G}. \quad (2.13)$$

Уравнение (2.13) связывает нормированную пропускную способность ρ и нормированный полный информационный обмен G при использовании канала системы ALOHA. График данной зависимости отмечен на рис. 2.10 как ALOHA (алгоритм «чистая ALOHA»). По мере роста G увеличивается и ρ до тех пор, пока большое количество конфликтных ситуаций не приведет к снижению пропускной способности. Максимум ρ, равный $1/2e = 0{,}18$, достигается при $G = 0{,}5$. Таким образом, в канале ALOHA может быть использовано лишь 18 % ресурса связи.

ALOHA был первым протоколом многостанционного доступа, предложенный Н. Абрамсоном из Гавайского университета (США) в 1971 г. (название протокола происходит от гавайского приветствия Aloha) и использовался в беспроводной сети на скорости 9600 бит/с. Использовалась широковещательная радиосвязь со стационарными передатчиками, однако сама идея применима к любой системе, в которой независимые пользователи соревнуются за право использования одного общего канала. **Два вида протокола ALOHA:** а) «Чистая ALOHA» (Pure ALOHA) – все станции работают в одном канале связи и передают данные в случайные моменты времени; такая система является непрерывной; б) «Дискретная ALOHA»

(Slotted ALOHA) предложенная в 1972 г. – время делится на дискретные интервалы (сегменты), соответствующие времени одного кадра и станция начинает передачу в начале такта; такая система является дискретной, позволяет значительно повысить пропускную способность канала.

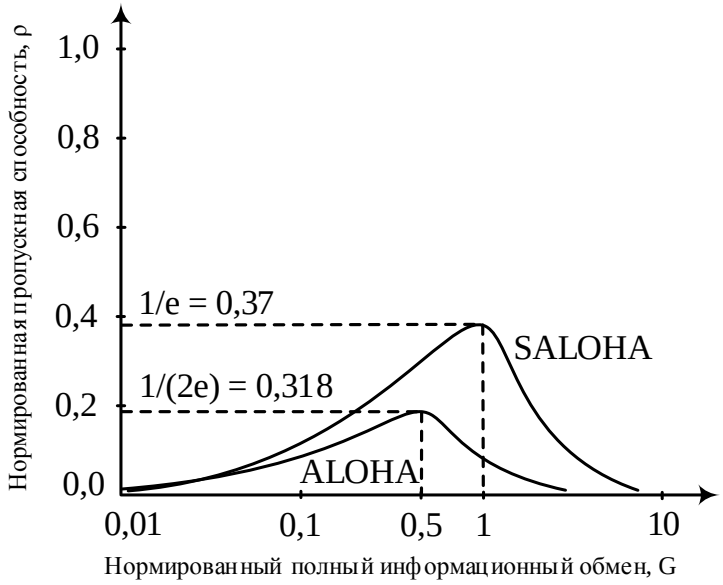


Рис. 2.10. Пропускная способность каналов ALOHA

Порядок функционирования первой сети «Чистая ALOHA» представлен на рис. 2.11 и может быть сформулирован следующим образом: а) базовая станция является центральным контроллером, который действует как промежуточный узел передачи кадров по одному каналу с частотной модуляцией; б) частота несущей для передачи данных от пользовательской станции к базовой (вверх) – 407 МГц, от базовой станции к пользовательской (вниз) – 413 МГц; в) среда передачи является общей и разделяется между станциями; г) при попытке станций занять канал одновременно, кадры сталкиваются и уничтожаются и с помощью обратной связи (подтверждения) станция может установить, передан ли кадр получателю.

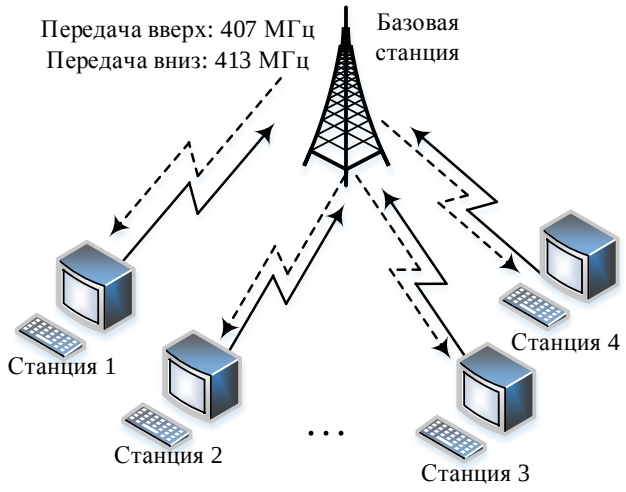


Рис. 2.11. Первая сеть ALOHA

Проблема возникновения коллизий в протоколе «Чистая ALOHA» представлена на рис. 2.12 и может быть сформулирована следующим образом: а) если передача кадров разными станциями перекрывается по времени, то происходят столкновения кадров (коллизии) и кадры уничтожаются; б) даже если только один первый бит второго кадра перекрывается последним битом первого кадра, оба кадра уничтожаются полностью, но позднее могут быть переданы повторно.

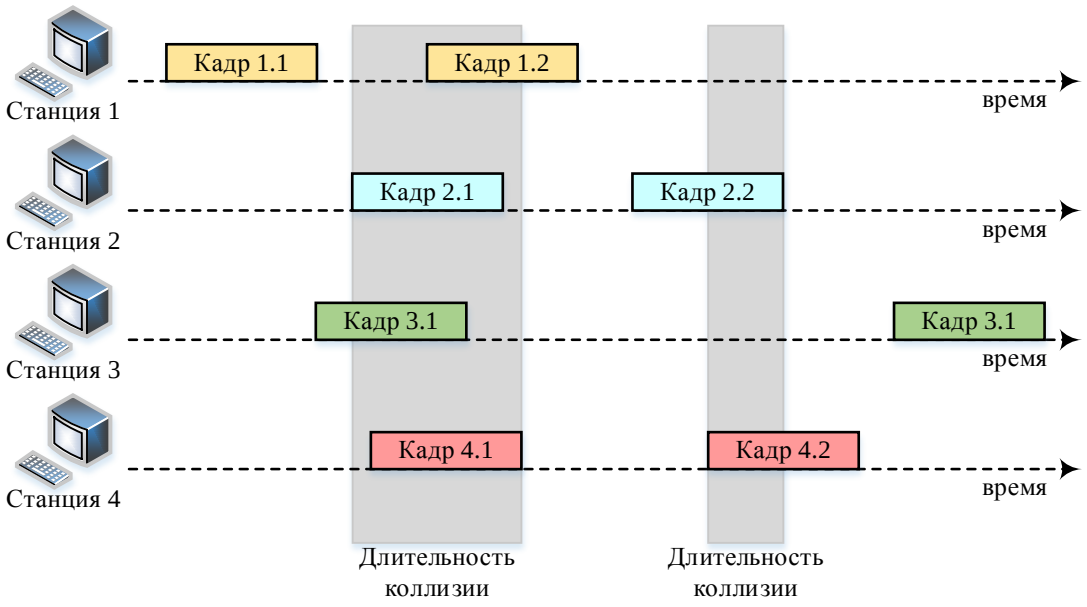


Рис. 2.12. Примеры возникновения коллизий в системе «Чистая ALOHA»

Два базовых принципа для протокола «Чистая ALOHA»: а) многостанционный доступ – любая станция начинает передачу, как только она имеет данные для отправки; б) обратная связь – после отправки кадра станция ожидает подтверждение; если станция не принимает подтверждение в течение времени, равному удвоенной максимальной задержки распространения сигнала, значит кадр потерян и станция снова передает этот кадр через случайный промежуток времени.

Алгоритм функционирования протокола «Чистая ALOHA» представлена на рис. 2.13 и может быть сформулирован следующим образом: а) как только у станции появляется кадр для отправки, она начинает его передачу и затем ожидает подтверждение в течение удвоенного максимального времени распространения сигнала ($2 \times T_p$); б) передача кадра считается успешной, если принято подтверждение в течение этого времени, в противном случае применяется стратегия отсрочки передачи (backoff), при которой станция пытается передавать кадр через случайные промежутки времени до тех пор, пока передача не будет успешной; в) после определенного количества неуспешных попыток (максимум $K_{\max}$), при которых нет под-

тверждения, передача кадра прекращается; г) время отсрочки повторной передачи кадра T_B определяется произведением максимального времени распространения сигнала T_p либо времени передачи всех бит одного кадра T_{fr} на сгенерированное случайное число; д) случайное число R для задания времени отсрочки повторной передачи генерируется в диапазоне от 0 до $2^K - 1$.

Рис. 2.13. Алгоритм функционирования
для протокола «Чистая ALOHA»

Пример 1. Найти возможное время отсрочки повторной передачи кадра T_B для различных значений K, если максимальное расстояние на пути прохождения сигнала между станциями в радиосети ALOHA равно 600 км.

Решение: скорость распространения сигнала 3×10^8 м/с. Время распространения сигнала: $T_p = (600 \times 10^3)/(3 \times 10^8) = 2$ мс.

а) $K = 1$. Диапазон генерируемых возможных случайных чисел от 0 до $2^K - 1$, т. е. $R = \{0, 1\}$. Следовательно, время отсрочки повторной передачи может быть $T_B = R \times T_p$=0 мс (0 × 2) либо 2 мс (1 × 2);

б) $K = 2$. Диапазон генерируемых возможных случайных чисел $R = \{0, 1, 2, 3\}$. Следовательно, T_B может быть 0, 2, 4 или 6 мс.

Для оценки **производительности канала для протокола «Чистая ALOHA»** следует учесть следующие факторы: **а)** основным параметром оценки эффективности канала является его пропускная способность, определяющая, какая часть из всех передаваемых кадров способна избежать коллизии; **б)** время кадра T_{fr} – интервал времени, требующийся для передачи

стандартного кадра фиксированной длины, определяется отношением длины кадра (бит) к скорости передачи данных (бит/с); **в)** пусть количество пользователей в сети достаточно велико, и они порождают новые кадры с Пуассоновским распределением (N кадров за время кадра); допущение большом количестве пользователей гарантирует, что величина N не уменьшается по мере блокировки кадров; **г)** при $N > 1$ кадры формируются с большей скоростью, чем может быть передано по каналу и почти каждый кадр будет испытывать столкновение, поэтому предположим, что $0 < N < 1$; **д)** кадры, испытавшие столкновение, передаются повторно; с учетом повторов вероятность k попыток передачи первичных и повторных кадров за время кадра также имеет Пуассоновское распределение со средним значением G кадров за время кадра (G – это интенсивность поступления кадров); **е)** в общем случае $G \geq N$; при малой загрузке ($N \approx 0$) столкновений мало и также мало будет повторных передач, т. е. $G \approx N$; однако при увеличении загрузки увеличивается количество столкновений и, следовательно, количество повторных передач $G > N$; **ж)** при любой нагрузке производительность канала S определяется произведением интенсивности поступления кадров G на вероятность успешной передачи: $S = G \times P$, где P – вероятность того, что кадр не пострадает от коллизий; **з)** если станция A начинает передачу кадра в момент t, а станция B при этом начнет передавать кадр в промежутке от $t - T_{fr}$ до t, то конец кадра B столкнется с началом кадра A (рис. 2.14); **и)** при формировании кадра станцией C в промежутке от t до $t + T_{fr}$, кадр C обязательно столкнется с кадром A; **к)** поэтому при передаче кадра A уязвимый период для других кадров составляет $2T_{fr}$.

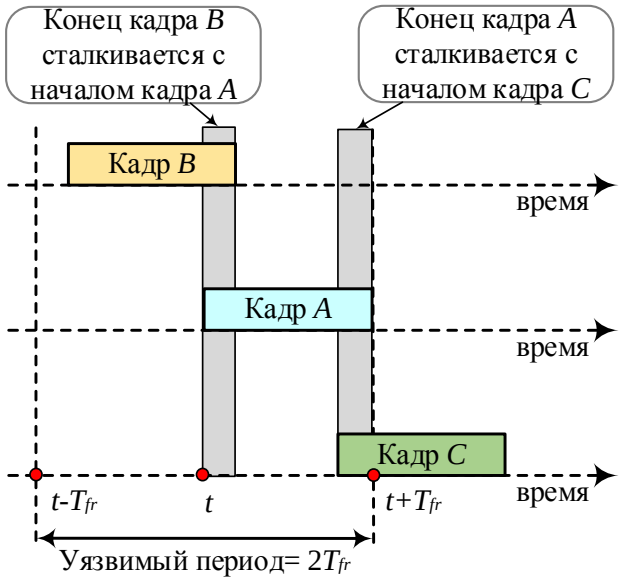


Рис. 2.14. Уязвимое время передачи кадра
для протокола «Чистая ALOHA»

Пример 2. В сети с протоколом «Чистая ALOHA» по разделяемой среде передачи на скорости 200 кбит/с передаются кадры длиной 200 бит каждый. При каком условии передаваемые кадры не подвергаются коллизиям.

Решение. Время кадра T_{fr} = (200 бит)/(200 кбит/с) = 1 мс. Уязвимый период $2T_{fr}$ = 2 мс. Следовательно, при передаче кадра *A*, любой другой кадр *B* должен начинаться не позже чем за 1 мс до начала кадра *A* либо через 2 мс и более после начала кадра *A*.

Оценку **производительности канала для протокола «Чистая ALOHA»** можно выполнить следующим образом. Вероятность того, что в течение времени кадра поступит *k* кадров определяется формулой Пуассона:

$$\Pr(k) = \frac{G^k e^{-G}}{k!},\ k \geq 0.$$

Вероятность поступления 0 кадров за время T_{fr} равна e^{-G}. Интенсивность поступления кадров за интервал времени двух кадров $2T_{fr}$ равна $2G$. Вероятность того, что кадр не поступит в течение уязвимого периода, т. е. не пострадает от коллизий, равна $P = e^{-2G}$. С учетом $S = G \times P$, получаем $S = Ge^{-2G}$. Наилучший показатель $S = 1/(2e) = 0{,}184$. **Производительность канала для протокола «Чистая ALOHA»:** $S = Ge^{-2G}$**. Максимально возможная производительность:** $S_{\max} = 0{,}184$ при $G = 0{,}5$.

Пример 3. В сети с протоколом «Чистая ALOHA» по разделяемой среде передачи на скорости 200 кбит/с передаются кадры длиной 200 бит каждый. Какова производительность канала для случаев, когда от всех станций поступает: а) 1000 кадров в секунду, б) 500 кадров в секунду, в) 250 кадров в секунду.

Решение: время кадра T_{fr} = (200 бит)/(200 кбит/с) = 1 мс.

а) по каналу передается 1 кадр в миллисекунду, т. е. поступающая нагрузка равна 1 (если единица времени – 1 мс). В этом случае $S = Ge^{-2G} =$ = 0,135 (13,5 %). Это означает, что в среднем 135 кадров из 1000 не подвергаются коллизиям;

б) на 1 мс приходится 0,5 кадра, т. е. поступающая нагрузка равна 0,5. В этом случае $S = 0{,}184$ (18,4 %). В среднем 18,4 % от всех передаваемых кадров не подвергаются коллизиям (это максимальная производительность);

в) на 1 мс приходится 0,25 кадра, т. е. поступающая нагрузка равна 0,25. В этом случае $S = 0{,}152$ (15,2 %). В среднем 15,2 % от всех передаваемых кадров не подвергаются коллизиям.

2.2.2. Алгоритм доступа SALOHA

Чистый алгоритм ALOHA можно улучшить, если ввести небольшую координацию между станциями. Примером является система ALOHA с выделением временных интервалов (slotted ALOHA – SALOHA). Всем станциям передается последовательность синхронизирующих импульсов. Сообщения могут передаваться только в течение временного интервала между синхронизирующими импульсами, а начало передачи пакета обязательно должно совпадать с *началом* интервала. Внесение таких дополнений позволяет вдвое снизить число конфликтных ситуаций, поскольку теперь конфликтовать могут только сообщения, передаваемые в течение одного временного интервала. Можно показать, что при использовании алгоритма SALOHA сокращение конфликтного промежутка с 2τ до τ дает следующее соотношение между нормированной пропускной способностью ρ и нормированным полным информационным обменом G:

$$\rho = Ge^{-G}. \qquad (2.14)$$

График зависимости (2.14) приведен на рис. 2.10, где он отмечен как SALOHA («система ALOHA с выделением временных интервалов»). В данном случае максимальное значение ρ равно $1/e = 0{,}37$, что в два раза больше аналогичного показателя чистого алгоритма ALOHA.

Режим повторной передачи системы SALOHA отличается от соответствующего режима чистого алгоритма тем, что при получении пользователем отрицательного подтверждения (NAK) следующая попытка производится после *случайной* паузы. Работа алгоритма S-ALOHA представлена на рис. 2.15.

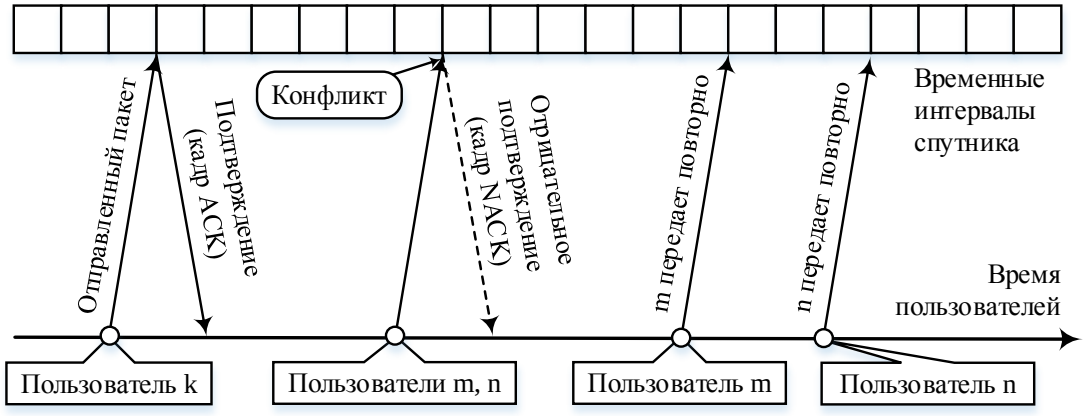


Рис. 2.15. Работа алгоритма ALOHA
с выделением временных интервалов (SALOHA)

После успешной передачи пакета данных пользователь *k* получает со спутника подтверждение о получении. Также показаны пользователи *m* и *n*,

которые одновременно начинают передачу пакетов, что приводит к конфликту, и спутник передает сигнал NAK обоим пользователям. Для определения времени повторной передачи обе станции используют генератор случайных чисел. Далее на рисунке показано возможное продолжение: повторная передача пользователями *m* и *n* после случайно выбранной паузы.

Пример. Процесс Пуассона. Пусть передачу и повторную передачу пакетов можно описать как пуассоновский процесс. Полная частота передачи пакетов равна $\lambda_t = 10$ пакетов в секунду; длительность пакета $\tau = 10$ мс. Вероятность возникновения в процессе передачи пакета конфликта с еще одним пользователем (используется алгоритм SALOHA):

$$P(k=1) = \left.\frac{(\tau\lambda_t)^k e^{-\tau\lambda_t}}{k!}\right|_{K=1} = (0{,}01\cdot 10)^1 \cdot e^{-0{,}1} = 0{,}1\cdot e^{-0{,}1} = 0{,}09.$$

Протокол «Дискретная ALOHA» предложен в 1972 г. Робертсоном и позволяет удвоить производительность по сравнению с протоколом «Чистая ALOHA». Время делится на интервалы (слоты), соответствующие длительности одного кадра, и станции могут отправлять кадры не в любой момент, а дожидаются начала тактового интервала (необходима специальная станция для синхронизации).

Проблема возникновения коллизий в протоколе «Дискретная ALOHA» представлена на рис. 2.16.

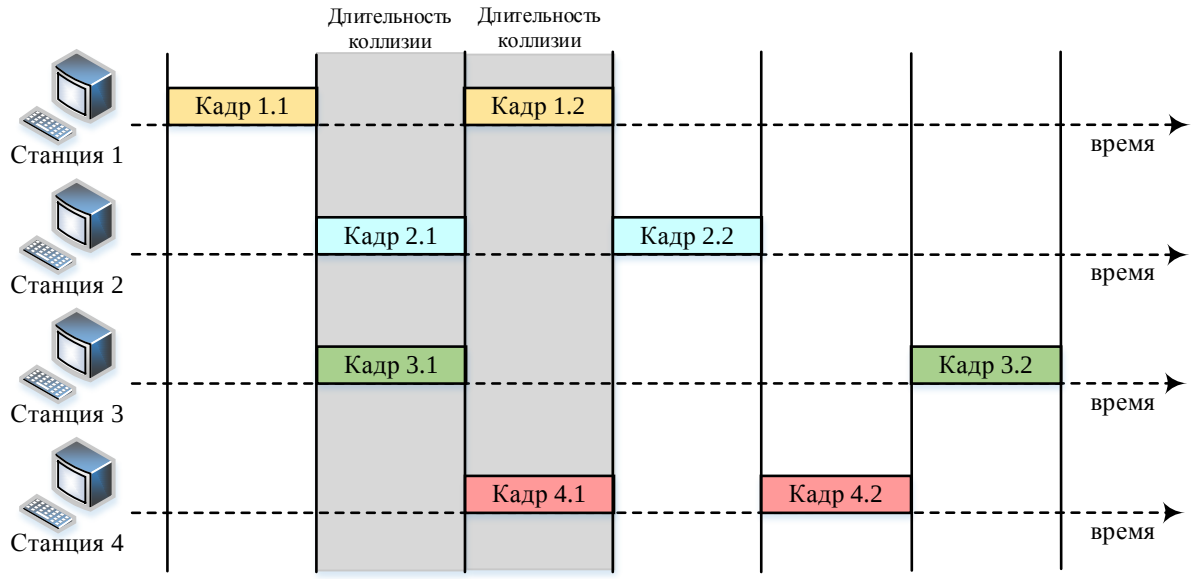


Рис. 2.16. Примеры возникновения коллизий
в системе «Дискретная ALOHA»

Для оценки **производительности канала для протокола «Дискретная ALOHA»** следует учесть следующие факторы: а) если станция *A* начинает передачу кадра в начале такта *t*, а станция *B* при этом начнет передавать кадр в начале предыдущего такта $t - T_{fr}$, то конец кадра *B* не столкнется с началом кадра *A*; б) при передаче кадра станцией *C* в интервале вместе с кадром *A*, они столкнутся; в) уязвимый временной интервал равен времени кадра T_{fr} (рис. 2.17); г) вероятность отсутствия передачи за время кадра равна e^{-G}; д) производительность канала $S = Ge^{-G}$. Наилучший показатель производительности достигается при $G = 1$: $S = 1/e = 0{,}368$.

Производительность канала для протокола «Дискретная ALOHA»: $S = Ge^{-G}$. **Максимально возможная производительность:** $S_{max} = 0{,}368$ при $G = 1$**.**

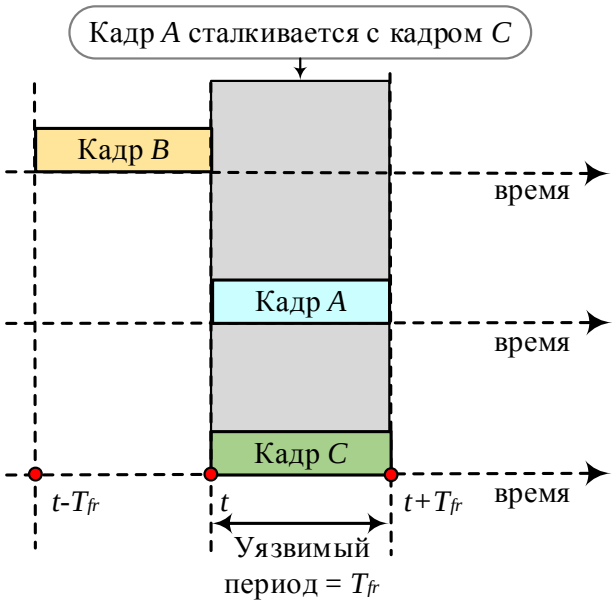


Рис. 2.17. Уязвимое время передачи кадра
для протокола «Дискретная ALOHA»

Пример 4. В сети с протоколом «Дискретная ALOHA» по разделяемой среде передачи на скорости 200 кбит/с передаются кадры длиной 200 бит каждый. Какова производительность канала для случаев, когда от всех станций поступает: а) 1000 кадров в секунду; б) 500 кадров в секунду; в) 250 кадров в секунду.

Решение: время кадра T_{fr} = (200 бит)/(200 кбит/с) = 1 мс.

а) По каналу передается 1 кадр в миллисекунду, т. е. поступающая нагрузка равна 1 (если единица времени – 1 мс). В этом случае $S = Ge^{-G}$ = = 0,368 (36,8 %). Это означает, что в среднем 368 кадров из 1000 не подвергаются коллизиям (это максимальная производительность).

б) На 1 мс приходится 0,5 кадра, т. е. поступающая нагрузка равна 0,5. В этом случае $S = 0{,}303$ (30,3 %). В среднем 30,3 % от всех передаваемых кадров не подвергаются коллизиям.

в) На 1 мс приходится 0,25 кадра, т. е. поступающая нагрузка равна 0,25. В этом случае $S = 0{,}195$ (19,5 %). В среднем 19,5 % от всех передаваемых кадров не подвергаются коллизиям.

2.2.3. Алгоритм доступа RALOHA

Работа систем ALOHA была значительно улучшена в результате резервирования (reservation-ALOHA – RALOHA). Системы RALOHA могут использоваться в двух основных режимах.

Режим без резервирования (состояние покоя): а) выделенный интервал времени разбивается на небольшие подынтервалы резервирования; б) эти подынтервалы используются для резервирования интервалов передачи сообщений; в) после запроса резервирования пользователь ожидает подтверждения и распределения интервалов.

Режим с резервированием: а) если не выполняется резервирование, временной интервал разбивается на $M + 1$ интервалов; б) первые M интервалов используются для передачи сообщений; в) последний интервал разбивается на подынтервалы, которые используются для резервирования или передачи запросов; г) пользователи передают пакеты данных только в выделенных им элементах M интервалов.

Рассмотрим пример использования схемы RALOHA, представленный на рис. 2.18.

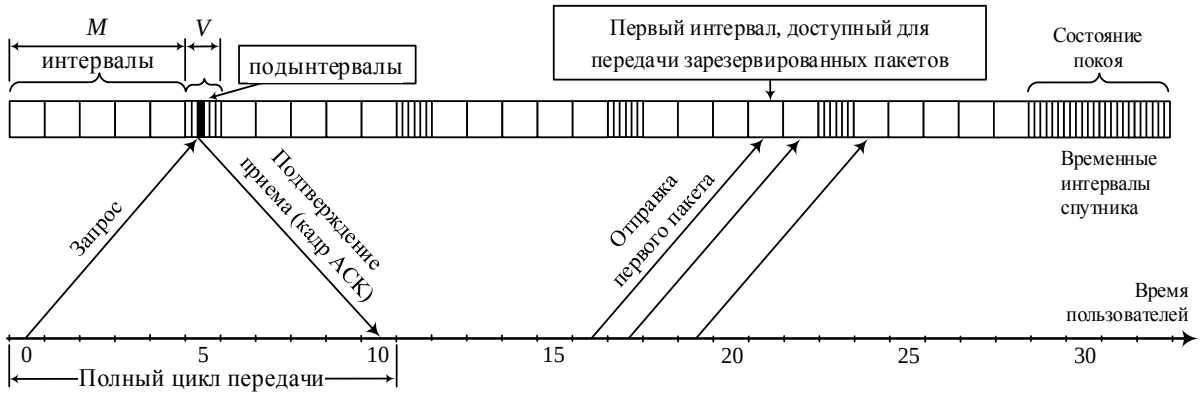


Рис. 2.18. Пример алгоритма ALOHA с использованием резервирования. Передающая станция резервирует три интервала ($M = 5$ интервалов, $V = 6$ подынтервалов)

В состоянии покоя время (с целью резервирования) разбивается на небольшие подынтервалы. После резервирования система конфигурируется так, что после $M = 5$ интервалов передачи следуют $V = 6$ подынтервалов резервирования; далее эта структура повторяется. В данном примере пере-

дающей станции необходимо зарезервировать три интервала времени. В подтверждении спутника содержатся инструкции относительно размещения первого пакета данных. Управление распределено, поэтому все пользователи получают сигнал со спутника и, соответственно, информацию о резервировании и распределении времени. В сигнале-подтверждении спутника находится *вся* необходимая информация, которая заключается в сообщении о выделении первого временного интервала. Как показано на рис. 2.18, в течение следующего интервала времени станция передает второй пакет. Далее пользователь знает, что следующий интервал состоит из шести подынтервалов, предназначенных для резервирования, поэтому передача информационных пакетов в течение этого времени *не производится*. Третий (последний) пакет отсылается в течение четвертого интервала. Если резервирование не производится, система возвращается в состояние покоя [3].

2.2.4. Производительность алгоритмов SALOHA, RALOHA

Для анализа систем множественного доступа используется нормированный показатель – зависимость средней задержки от нормированной пропускной способности. На рис. 2.19 представлена *идеальная зависимость задержки от пропускной способности*: для нормированных значений пропускной способности, $0 \le \rho < 1$, время задержки равно нулю, при $\rho = 1$ оно неограниченно возрастает.

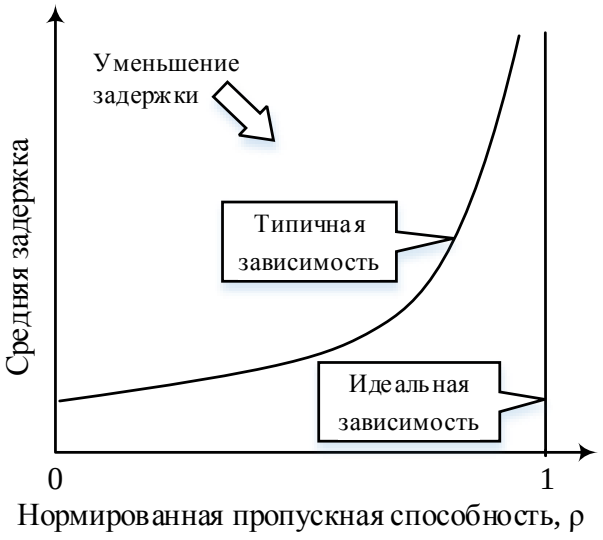


Рис. 2.19. Зависимость времени задержки
от пропускной способности

На рис. 2.20 сравниваются зависимости времени задержки от пропускной способности для алгоритмов SALOHA и RALOHA (формат сообщений: два интервала передачи данных и шесть подынтервалов резервирова-

ния). Время задержки этих двух систем сравнивают с помощью *идеальной* кривой. Для пропускной способности ρ < 0,2 среднее время задержки для системы SALOHA меньше, чем для системы RALOHA. В то же время для ρ, принадлежащего диапазону 0,2 < ρ < 0,67, RALOHA превосходит SALOHA, поскольку у первой среднее время задержки существенно меньше. В чем причина превосходства схемы SALOHA при малоинтенсивном обмене данными? Данный алгоритм не требует служебных издержек для резервирования подынтервалов, как в случае RALOHA. Таким образом, при небольших значениях ρ производительность RALOHA ниже из-за более высоких расходов. При ρ > 0,2 конфликтные ситуации и повторная передача данных в системе SALOHA приводят к тому, что время задержки растет быстрее, чем в RALOHA (и неограниченно возрастает при ρ = 0,37). При более высоких значениях пропускной способности 0,2 < ρ < 0,67 служебные издержки схемы RALOHA полностью окупаются и обеспечивают менее резкое возрастание времени задержки при росте ρ. При использовании RALOHA время задержки возрастает до бесконечности при ρ = 0,67.

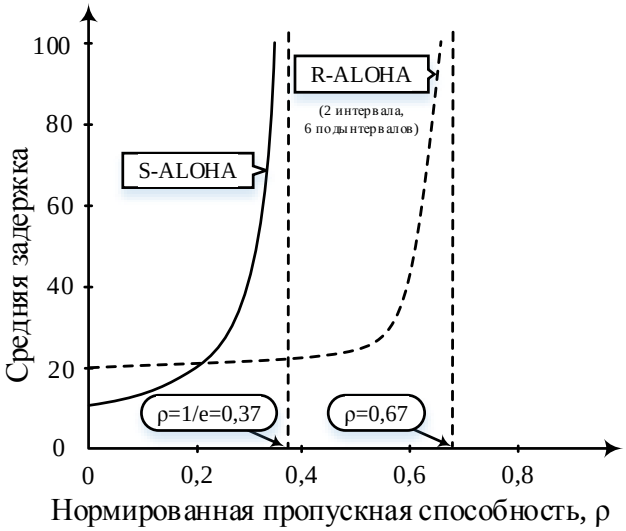


Рис. 2.20. Зависимость времени задержки
от пропускной способности для схем
SALOHA и RALOHA

Пример. Использование канала связи. а) В качестве меры использования канала выбрана нормированная пропускная способность ρ. Ее можно найти как отношение успешно переданных данных к полному объему данных (включая отклоненные данные). Найдите нормированную пропускную способность канала связи с максимальной скоростью передачи данных R = 50 кбит/с, который используется M = 10 станциями связи, каждая из которых передает данные со средней частотой λ = 2 пакета в секунду. Формат системы предусматривает пакеты по b = 1350 бит. б) Применение какой из систем ALOHA будет оптимальным в данном случае?

Решение. а) Обобщая уравнение (2.4) для информационного потока от нескольких станций, получаем следующее:

$$\rho = \frac{Mb\lambda}{R} = \frac{10 \cdot 1350 \cdot 2}{50000} = 0,54.$$

б) в данной системе может использоваться только схема RALOHA, поскольку два других алгоритма не позволяют использовать 54 % ресурса.

Контрольные вопросы

1. Приведите основные режимы работы алгоритма ALOHA.
2. Чем определяется общая интенсивность поступления сообщений (пакетов)?
3. Чем определяется пропускная (средний объем успешно переданных данных) способность канала?
4. Чем определяется полный информационный обмен канала?
5. Чем определяется нормированная пропускная способность канала?
6. Чем определяется нормированный полный информационный обмен канала?
7. Чем определяется время передачи пакета?
8. Приведите и поясните выражение нормированной пропускной способности канала через время передачи пакета.
9. Приведите и поясните выражение нормированного полного информационного обмена канала через время передачи пакета.
10. Сформулируйте условие успешной передачи в протоколе «Чистая ALOHA».
11. Приведите и поясните формулу распределения Пуассона применительно к описанию статистике передачи сообщений.
12. Приведите и поясните выражение вероятности успешной передачи сообщения при допущении о пуассоновском характере поступления сообщений в протоколе чистая ALOHA.
13. Приведите и поясните выражение вероятности успешной передачи сообщения через общую частоту поступления сообщений и частоту успешного поступления сообщений в протоколе «Чистая ALOHA».
14. Как связана полная частота поступления сообщений, частоту успешного поступления сообщений и время передачи пакета в протоколе «Чистая ALOHA»?
15. Как связана нормированная пропускная способность канала с нормированным полным информационным обменом в протоколе «Чистая ALOHA»?
16. Чему равен коэффициент использования канала в протоколе «Чистая ALOHA»?
17. Сформулируйте особенности функционирования протокола SALOHA с выделением временных интервалов.
18. Чему равен коэффициент использования канала в протоколе SALOHA с выделением временных интервалов?
19. Сформулируйте условие успешной передачи в протоколе SALOHA.
20. Приведите и поясните выражение вероятности успешной передачи сообщения при допущении о пуассоновском характере поступления сообщений в протоколе SALOHA.
21. Сформулируйте особенности функционирования протокола RALOHA в режиме без резервирования и с резервированием.
22. Какими характеристиками оценивается производительность протоколов множественного доступа?

23. Как в общем случае средняя задержка передачи сообщений связана с нормированной пропускной способностью канала для протоколов множественного доступа?

24. Почему при нормированной пропускной способности $\rho<0,2$ схема SALOHA превосходит схему RALOHA? Почему при нормированной пропускной способности $0,2 < \rho < 0,67$ схема RALOHA превосходит схему SALOHA?

25. Сформулируйте порядок функционирования первой сети «Чистая ALOHA».

26. Проиллюстрируйте проблему возникновения коллизий в системе «Чистая ALOHA».

27. Сформулируйте базовые принципы функционирования протокола «Чистая ALOHA».

28. Приведите и поясните алгоритм функционирования протокола «Чистая ALOHA».

29. Сформулируйте предпосылки для оценки производительности канала для протокола «Чистая ALOHA».

30. Проиллюстрируйте уязвимое время передачи кадра для протокола «Чистая ALOHA».

31. Чему равна максимальная производительность протокола «Чистая ALOHA».

32. Проиллюстрируйте проблему возникновения коллизий в системе «Дискретная ALOHA».

33. Сформулируйте предпосылки для оценки производительности канала для протокола «Дискретная ALOHA».

34. Проиллюстрируйте уязвимое время передачи кадра для протокола «Дискретная ALOHA».

35. Чему равна максимальная производительность протокола «Дискретная ALOHA»?

Задачи

1. Группа станций совместно использует канал по протоколу «Чистая ALOHA», поддерживающий скорость 56 Кбит/с. В среднем каждые 10 с любая станция передает пакет данных, даже если на данный момент предыдущий пакет еще не отправлен (т. е. станция заносит пакеты в буфер). Размер каждого пакета равен 3000 бит. Найдите максимальное число станций, которые могут одновременно использовать данный канал. Процесс прибытия пакетов считать пуассоновским.

2. Группа из трех станций совместно использует канал по протоколу «Чистая ALOHA», поддерживающий скорость 56 Кбит/с. Средняя скорость передачи данных станциями равна следующему: R_1 = 7,5 Кбит/с, R_2 = 10 кбит/с, R_3 = 20 Кбит/с. Размер каждого пакета составляет 100 бит. Вычислите нормированный объем информации, которой обмениваются через канал, нормированную пропускную способность, вероятность успешной передачи и скорость поступления успешно переданных пакетов. Процесс поступления пакетов считать пуассоновским.

3. Докажите, что при использовании чистой схемы ALOHA нормированная пропускная способность не превышает $1/(2e)$, а максимум наблюдается при нормированном объеме переданной информации, равном 0,5.

2.3. Лабораторная работа. Оценка производительности алгоритма ALOHA

Цель работы: оценить производительность протоколов ALOHA и SALOHA средствами имитационного моделирования.

2.3.1. Построение дискретно-событийной имитационной модели

Для оценки производительности протоколов случайного доступа ALOHA и SALOHA в настоящей лабораторной работе используется дискретно-событийное имитационное моделирование.

Дискретно-событийное моделирование (англ. *discrete-event simulation,* DES) – это вид имитационного моделирования, в котором функционирование системы представляется как хронологическая последовательность событий; событие происходит в определенный момент времени и знаменует собой изменение состояния системы. Компоненты системы дискретно-событийного моделирования включают переменные, определяющие состояние системы, и логику, определяющую, что произойдет в ответ на какое-то событие. Для построения модели *введем следующие переменные*.

Нормированная предложенная нагрузка. Обозначим через T_t (бит) объем в битах заново сгенерированных пакетов трафика и пакетов трафика повторной передачи от всех станций сети, а через R (бит/с) – скорость передачи данных в канале. Тогда нормированная предложенная нагрузка G (с) может быть выражена как $G = T_t/R$. Если пакеты не генерируются, $G = 0$.

Нормированная пропускная способность. Обозначим через T (бит) – размер пакета в битах, а n – общее число успешно переданных пакетов всеми станциями сети за единицу времени. Тогда объем успешно переданных данных в битах $T{\cdot}n$ (бит), нормированный скоростью передачи данных в канале R (бит/с), представляет нормированную пропускную способность S (с), определяемую выражением $S = (T \cdot n)/R$. Если пакеты не генерируются, или все передачи подвергаются коллизиям и оказываются безуспешными, то $S = 0$. Если же, напротив, все генерируемые пакеты передаются успешно, то $S = 1$.

Нормированная средняя задержка передачи. Интервал времени между моментом генерирования пакета станцией-отправителем и моментом доставки его станции-получателю определяет задержку передачи. Средняя задержка передачи зависит, в том числе, и от размера пакета. Обозначим через D – нормированную размером пакета T среднюю задержку передачи от всех станций сети. Измерение задержки передачи выполняется для успешно переданных пакетов.

Для построения имитационной модели *сделаем следующие допущения*.

Пусть интервал генерации пакетов подчиняется экспоненциальному распределению со средним значением $1/\lambda$, а вероятность того, что за время *t* станция сети сгенерирует *n* пакетов определяется формулой Пуассона:

$$P_n(t) = \frac{e^{-\lambda t}(\lambda t)^n}{n!}. \quad (2.15)$$

Из формулы Пуассона следует, что вероятность того, что за интервал времени от 0 до *t* не будет сгенерировано ни одного пакета определяется выражением $P_0(t) = e^{-\lambda t}$; следовательно вероятность того, что первый пакет будет сгенерирован за время *t* определяется выражением

$$p(t) = 1 - e^{-\lambda t}. \quad (2.16)$$

Пусть *G* – нормированная предложенная нагрузка сети, M_{num} – число станций сети, T_{time} – время передачи пакета, T_{int} – средний интервал генерации пакета станцией, тогда интенсивность генерации пакетов λ равна

$$\lambda = 1/T_{\text{int}}. \quad (2.17)$$

Вероятность *p*(*t*) можно выразить через нормированную предложенную нагрузку сети *G* и число станций сети M_{num}

$$p(t) = \frac{G}{M_{\text{num}}}. \quad (2.18)$$

Примем за рассматриваемый интервал *t* время передачи пакета T_{time}

$$t = T_{\text{time}}. \quad (2.19)$$

Тогда, подставив (2.17) и (2.19) в (2.16) и приравняв (2.18), получим вероятность генерации пакета за время T_{time}

$$p(t) = \frac{G}{M_{\text{num}}} = 1 - e^{\frac{T_{\text{time}}}{T_{\text{int}}}}. \quad (2.20)$$

Из (2.20) получим средний интервал генерации пакета станцией

$$T_{\text{int}} = -\frac{T_{\text{time}}}{\log\left(1 - \frac{G}{M_{\text{num}}}\right)}. \quad (2.21)$$

Далее, введем режимы (состояния), в которых могут находиться станции при функционировании по протоколу ALOHA/SALOHA: а) режим ожидания `STANDBY`; б) режим передачи `TRANSMIT`; в) режим коллизии `COLLISION`.

Вероятность генерации пакета $p(t)$ можно выразить через средний интервал генерации пакета станцией T_{int}

$$p(t) = 1 - e^{-\frac{t}{T_{\text{int}}}}. \tag{2.22}$$

При дискретно-событийном имитационном моделировании нас интересует момент времени генерации пакета, который можно выразить из (2.22)

$$t = -T_{\text{int}} \cdot \ln\left[1 - p(t)\right], \tag{2.23}$$

где вероятность $p(t)$ равномерна распределена на интервале от 0 до 1.

Параметры имитационной модели сведены в табл. 2.1.

Таблица 2.1

Параметры моделирования

Параметр	Ед. измерения	Содержание параметра
brate	бит/с	Битовая скорость передачи данных в канале
plen	бит	Размер пакета
Ttime	с	Время передачи пакета
Dtime		Нормированная задержка распространения сигнала
Mnum	шт.	Число станций
Mplen	бит	Вектор размера пакетов станций сети
Mstate		Вектор состояний станций сети
mgtime	с	Вектор времен генерации новых пакетов станцией
mtime	с	Вектор времен изменения состояния станций
Mstime	с	Вектор времен передачи пакетов станциями
G	с	Нормированная предложенная нагрузка
Tint	с	Средний интервал генерации пакета станцией
Rint	с	Средний интервал повторной передачи пакета станцией
Spnum		Число успешно переданных пакетов
spend		Число пакетов для имитационного моделирования
Splen	бит	Объем данных успешно переданных пакетов в битах
Tplen	бит	Объем данных предложенных к передаче пакетов
Wtime	с	Задержка передачи пакетов

Дискретно-событийное имитационное моделирование для оценки производительности протоколов доступа ALOHA и SALOHA выполняется в цикле для каждого отдельно взятого значения нормированной предложенной нагрузки G из заданного диапазона, например `G=[0.1:0.1:2]`. Это значение G определяет интервал генерации новых пакетов станцией

по формуле (2.21) `Tint=-Ttime/log(1-G(indx)/Mnum)`. Интервал поступления пакетов повторной передачи принимается равным интервалу поступления новых пакетов `Rint=Tint`.

Исходными данными для оценки времени передачи пакета являются битовая скорость передачи `Brate` и размер пакета `Plen`. Оценка времени передачи пакета выполняется по формуле `Ttime=Plen/Brate`.

Для каждого отдельно взятого значения нормированной предложенной нагрузки *G* в цикле моделирования производится инициализация следующих параметров: а) объем успешно переданных пакетов в битах `Splen`; б) объем данных предложенных к передаче пакетов в битах `Tplen`; в) задержка передачи пакетов в секундах `Wtime`:

```
% объем данных успешно переданных пакетов, бит
Splen=0;
% объем данных предложенных к передаче пакетов, бит
Tplen=0;
% задержка передачи пакетов, с
Wtime=0;
```

По ходу имитационного моделирования (скрипт 2.1, 2.2) значения данных параметров инкрементируются. Условием завершения дискретно-событийного имитационного моделирования для отдельно взятого значения нормированной предложенной нагрузки *G* является достижение числом успешно переданных пакетов `Spnum` значения `TOTAL`; обработка пакетов для отдельно взятого значения нормированной предложенной нагрузки *G* осуществляется в цикле `while (Spnum<TOTAL)`. По достижению числа успешно переданных пакетов `Spnum` значения `TOTAL` заполняются вектора для индекса `indx` отдельно взятого значения нормированной предложенной нагрузки *G*, соответствующие ранее инициализированным параметрам: а) вектор нормированной предложенной нагрузки `Traffic(indx)`; б) вектор нормированной пропускной способности `S(indx)`; в) вектор нормированной средней задержки передачи `Delay(indx)`:

```
 % вектор нормированной предложенной нагрузки
Traffic(indx)=Tplen/Brate/now_time;
% вектор нормированной пропускной способности
S(indx)=Splen/Brate/now_time;
% вектор нормированной средней задержки передачи
Delay(indx)=Wtime/TOTAL*Brate/Plen;
```

Вектор времен генерации новых пакетов станцией моделируется выражением (2.23) :

```
% вектор времен поступления нового пакета, с
mgtime=-Tint*log(1-rand(1,Mnum));
```

Вектор времен изменения состояния станций `mtime` инициализируется значением `mgtime`:

```
% вектор времен изменения состояния станций, с
mtime=mgtime;
```

Далее производится инициализация векторов состояния станций сети, размера пакетов станций сети:

```
% вектор состояний станций сети
Mstate=zeros(1,Mnum);
% вектор размера пакетов станций сети, бит
Mplen(1:Mnum)=Plen;
```

Текущее время инициализируется минимальным временем наступления события вектора `mtime`:

```
% текущее время, с
now_time=min(mtime);
```

Оценим аналитически нормированную пропускную способность для протокола ALOHA. Нормированная предложенная нагрузка за время передачи пакета τ с интенсивностью генерации пакетов λ определяется выражением $G = \lambda \cdot \tau$. Для успешной передачи станцией пакета, сгенерированного в момент времени t_1, необходимо, чтобы другие станции сети не генерировали пакетов в интервале времени от $t_1 - \tau$ до $t_1 + \tau$. Исходя из Пуассоновского закона генерации пакетов станциями (2.15), вероятность успешной передачи определяется выражением $P_{succ} = P_0(2\tau) = e^{-2\lambda t} = e^{-2G}$. Нормированная пропускная способность для протокола ALOHA тогда может быть вычислена по формуле $S = G \cdot P_{succ} = G \cdot e^{-2G}$:

```
% теоретическая пропускная способность ALOHA
S=Traffic.*exp(-2*Traffic);
```

По ходу имитационного моделирования (скрипт 2.1, 2.2) значения данных параметров инкрементируются. Далее выполним оценку производительности протоколов ALOHA/SALOHA средствами дискретно-событийного имитационного моделирования.

2.3.2. Оценка производительности ALOHA/SALOHA

На рис. 2.21 представлен пример оценки производительности протокол ALOHA/SALOHA (скрипт 2.1).

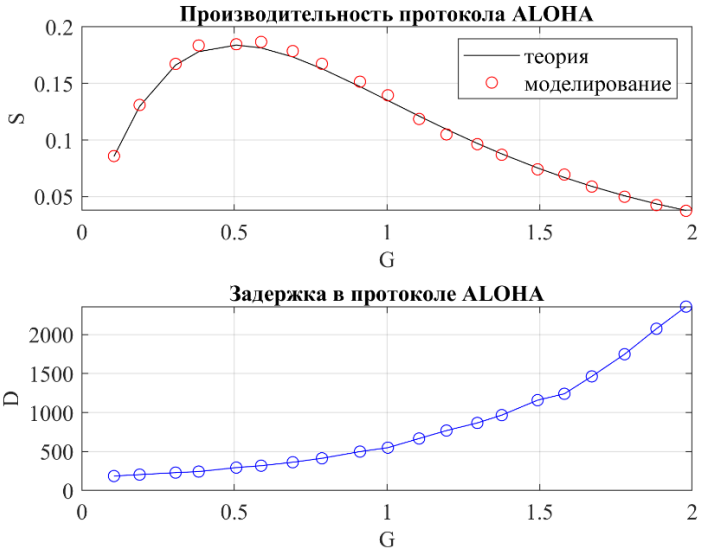


Рис. 2.21. Производительность протокола ALOHA

Скрипт 2.1. Оценка производительности протокола ALOHA

```
clear all; clc;
% ALOHA
STANDBY=0;          % режим ожидания
TRANSMIT=1;         % режим передачи
COLLISION=2;        % состояние коллизии
TOTAL=1000;         % число пакетов для моделирования

Brate=0.25e6;       % битовая скорость передачи, бит/с
Plen=500;           % размер пакета, бит
Ttime=Plen/Brate;   % время передачи пакета
Dtime=0.01;         % нормированная задержка передачи
Mnum=100;           % число станций
G=[0.1:0.1:2];      % нагрузка

for indx=1:length(G)    % цикл по нагрузке

    % интервал поступления новых пакетов, с
    Tint=-Ttime/log(1-G(indx)/Mnum);
    % интервал поступления пакетов повторной передачи, с
    Rint=Tint;

    Spnum=0;        % число успешно переданных пакетов, шт
    Tplen=0;        % объем данных предложенных к передаче пакетов, бит
    Splen=0;        % объем данных успешно переданных пакетов, бит
    Wtime=0;        % задержка передачи пакетов, с

    % вектор времен генерации новых пакетов станцией , с
    mgtime=-Tint*log(1-rand(1,Mnum));
    % вектор времен изменения состояния станций, с
    mtime=mgtime;
    % вектор состояний станций сети
    Mstate=zeros(1,Mnum);
    % вектор размера пакетов станций сети, бит
    Mplen(1:Mnum)=Plen;
    % текущее время, с
    now_time=min(mtime);

    while (Spnum<TOTAL)
```

```
        % при успешной передаче пакета для станций с индексом idx:
        idx=find(mtime==now_time & Mstate==TRANSMIT);
        if length(idx)>0
            Spnum=Spnum+1;                      % инкремент Spnum
            Splen=Splen+Mplen(idx);             % инкремент Splen
            Wtime=Wtime+now_time-mgtime(idx);   % инкремент Wtime
            Mstate(idx)=STANDBY;                % переход в режим ожидания
            % время генерации нового пакета для станций с индексом idx
            mgtime(idx)=now_time-Tint*log(1-rand);
            % инициализация времени смены состояния станций с индексом idx
            mtime(idx)=mgtime(idx);
        end

        % при коллизии для станций с индексом idx:
        idx=find(mtime==now_time & Mstate==COLLISION);
        if length(idx)>0
            Mstate(idx)=STANDBY;                % переход в режим ожидания
            % время повторной передачи пакета станциями
            mtime(idx)=now_time-Rint*log(1-rand(1,length(idx)));
        end

        % нахождение станций, в режиме передачи
        idx=find(mtime==now_time);
        if length(idx)>0
            Mstate(idx)=TRANSMIT;   % переход в режим передачи
            % время завершения передачи пакета станциями
            mtime(idx)=now_time+Mplen(idx)/Brate;
            Tplen=Tplen+sum(Mplen(idx));        % инкремент Tplen
        end
        idx=find(Mstate==TRANSMIT | Mstate==COLLISION);
            if length(idx)>1
                Mstate(idx)=COLLISION;
            end
        now_time=min(mtime);
    end
    % вектор нормированной предложенной нагрузки
    Traffic(indx)=Tplen/Brate/now_time;
    % вектор нормированной пропускной способности
    S(indx)=Splen/Brate/now_time;
    % вектор нормированной средней задержки передачи
    Delay(indx)=Wtime/TOTAL*Brate/Plen;
end
% теоретическая пропускная способность для протокола ALOHA
Stheory=Traffic.*exp(-2*Traffic);
subplot(2,1,1); plot(Traffic,Stheory,'-k',Traffic,S,'ro'); grid on;
xlabel('G'); ylabel('S'); legend('теория', 'моделирование');
title('Производительность протокола ALOHA');
subplot(2,1,2); plot(Traffic,Delay,'-bo'); grid on;
xlabel('G'); ylabel('D'); title('Задержка в протоколе ALOHA');
```

Оценим аналитически нормированную пропускную способность для протокола SALOHA. В протоколе SALOHA станции сети синхронизированы и могут начинать передачу пакета только в течение заданного временного интервала (кадра), поэтому для успешной передачи станцией пакета, сгенерированного в момент времени t_1 необходимо, чтобы другие станции сети не генерировали пакетов в интервале времени от t_1 до $t_1 + \tau$. Исходя из Пуассоновского закона генерации пакетов станциями (2.15) , вероятность ус-

пешной передачи определяется выражением $P_{succ} = P_0(\tau) = e^{-\lambda t} = e^{-G}$. Нормированная пропускная способность для протокола SALOHA тогда может быть вычислена по формуле $S = G \cdot P_{succ} = G \cdot e^{-G}$ и равна вероятности того, что за интервал времени τ будет сгенерирован ровно 1 пакет

$$P_1(t) = \frac{e^{-\lambda\tau}(\lambda\tau)^1}{1!} = \lambda\tau \cdot e^{-\lambda\tau} = G \cdot e^{-G} = S. \qquad (2.24)$$

На рис. 2.22 представлен пример оценки производительности протокола SALOHA (скрипт 2.2).

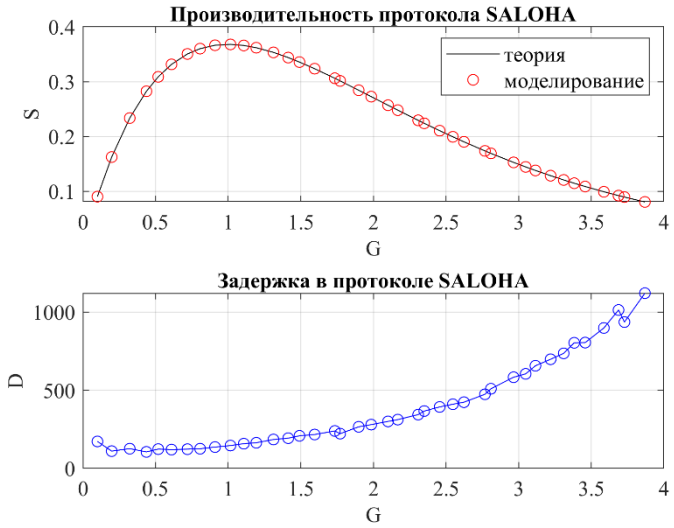


Рис. 2.22. Производительность протокола SALOHA

Скрипт 2.2. Оценка производительности протокола SALOHA

```
clear all; clc;
% SALOHA
STANDBY=0;          % режим ожидания
TRANSMIT=1;         % режим передачи
COLLISION=2;        % состояние коллизии
TOTAL=1000;         % число пакетов для моделирования

Brate=0.25e6;       % битовая скорость передачи, бит/с
Plen=500;           % размер пакета, бит
Ttime=Plen/Brate;   % время передачи пакета
Dtime=0.01;         % нормированная задержка передачи
Mnum=100;           % число станций
G=[0.1:0.1:4];      % нагрузка

for indx=1:length(G)    % цикл по нагрузке
        % интервал поступления новых пакетов, с
    Tint=-Ttime/log(1-G(indx)/Mnum);
    % интервал поступления пакетов повторной передачи, с
    Rint=Tint;

    Spnum=0;        % число успешно переданных пакетов, шт
    Splen=0;        % число успешно переданных бит, бит
    Tplen=0;        % число бит нагрузки (от поступивших пакетов), бит
    Wtime=0;        % задержка передачи, с
```

```
    % длительность слота, с
    slot=Plen/Brate;
    % вектор времен поступления нового пакета, с
    mgtime=-Tint*log(1-rand(1,Mnum));
    % вектор времен смены состояния станции, с
    mtime=(fix(mgtime));
    % вектор состояния станций
    Mstate=zeros(1,Mnum);
    % вектор размера пакетов станций, бит
    Mplen(1:Mnum)=Plen;
    % текущее время, с
    now_time=min(mtime);

    while (Spnum<TOTAL)

        % при успешной передаче пакета для станций с индексом idx:
        idx=find(mtime==now_time & Mstate==TRANSMIT);
        if length(idx)>0
            Spnum=Spnum+1;                      % инкремент Spnum
            Splen=Splen+Mplen(idx);             % инкремент Splen
            Wtime=Wtime+now_time-mgtime(idx);   % инкремент Wtime
            Mstate(idx)=STANDBY;                % переход в режим ожидания
            % время генерации нового пакета для станций с индексом idx
            mgtime(idx)=now_time-Tint*log(1-rand);
            % инициализация времени смены состояния станций с индексом idx
            mtime(idx)=(fix(mgtime(idx)/slot)+1)*slot;
        end

        % при коллизии для станций с индексом idx:
        idx=find(mtime==now_time & Mstate==COLLISION);
        if length(idx)>0
            Mstate(idx)=STANDBY;                % переход в режим ожидания
            % время повторной передачи пакета станциями
            mtime(idx)=now_time-Rint*log(1-rand(1,length(idx)));
            mtime(idx)=(fix(mtime(idx)/slot)+1)*slot;
        end

        % нахождение станций, в режиме передачи
        idx=find(mtime==now_time);
        if length(idx)>0
            Mstate(idx)=TRANSMIT;           % переход в режим передачи
            % время завершения передачи пакета станциями
            mtime(idx)=now_time+Mplen(idx)/Brate;
            mtime(idx)=round(mtime(idx)/slot)*slot;
            Tplen=Tplen+sum(Mplen(idx));    % инкремент Tplen
        end
        idx=find(Mstate==TRANSMIT | Mstate==COLLISION);
            if length(idx)>1
                Mstate(idx)=COLLISION;
            end
        now_time=min(mtime);
    end
    % вектор нормированной предложенной нагрузки
    Traffic(indx)=Tplen/Brate/now_time;
    % вектор нормированной пропускной способности
    S(indx)=Splen/Brate/now_time;
    % вектор нормированной средней задержки передачи
    Delay(indx)=Wtime/TOTAL*Brate/Plen;
end
% теоретическая пропускная способность для протокола SALOHA
```

```
S=Traffic.*exp(-Traffic);
subplot(2,1,1); plot(Traffic,S,'-k',Traffic,S,'ro'); grid on;
xlabel('G'); ylabel('S'); legend('теория', 'моделирование');
title('Производительность протокола SALOHA');
subplot(2,1,2); plot(Traffic,Delay,'-bo'); grid on;
xlabel('G'); ylabel('D'); title('Задержка в протоколе SALOHA');
```

Контрольные вопросы

1. Дайте определение дискретно-событийному имитационному моделированию.
2. Дайте определение нормированной предложенной нагрузке.
3. Дайте определение нормированной пропускной способности.
4. Дайте определение нормированной задержке передачи.
5. Почему при построении модели используются нормированные характеристики?
6. Какие допущения и ограничения вводятся при построении имитационной модели для оценки производительности протоколов ALOHA/SALOHA?
7. Чем определяется средний интервал генерации пакета станцией при сделанных ранее допущениях?
8. Приведите и поясните состояния, в которых может находиться станция при сделанных ранее допущениях.
9. Какие параметры имитационной модели описываются скалярными, а какие векторными величинами и почему?
10. Чем определяется выбор диапазона значений нормированной предложенной нагрузки для оценки производительности протоколов ALOHA/SALOHA?
11. По какому параметру организуется цикл при построении имитационной модели для оценки производительности протоколов ALOHA/SALOHA?
12. По каким исходным данным можно оценить время передачи пакета?
13. Какие параметры инициализируются в каждом отдельном цикле моделирования?
14. При каком условии имитационное моделирование завершается?
15. Чем можно объяснить некоторое расхождение производительности протоколов ALOHA/SALOHA, полученной теоретически и средствами имитационного моделирования?

2.4. Практическое занятие. Методы доступа с контролем несущей CSMA

Цель занятия: изучить методы доступа с контролем несущей CSMA, CSMA/CD, CSMA/CA.

2.4.1. Особенности протоколов с контролем несущей

Особенности метода произвольного доступа: а) каждая станция имеет право независимого доступа к среде передачи без управления какой-либо другой станции; б) при попытке передать данные несколькими станциями одновременно возникает конфликт доступа (коллизия), данные могут быть потеряны или искажены.

Для решения проблемы произвольного доступа необходима процедура, отвечающая на вопросы: а) когда станция может получить доступ к среде передачи? б) как ведет себя станция в случае, если среда занята? в) как станция определяет, успешно ли осуществлена передача данных? г) как ведет себя станция в случае конфликта доступа?

В протоколах ALOHA станции передают данные, когда хотят, без учета поведения других станций. Возникает большое количество коллизий и коэффициент использования канала очень низкий. Эволюция протоколов многостанционного доступа представлена на рис. 2.23.

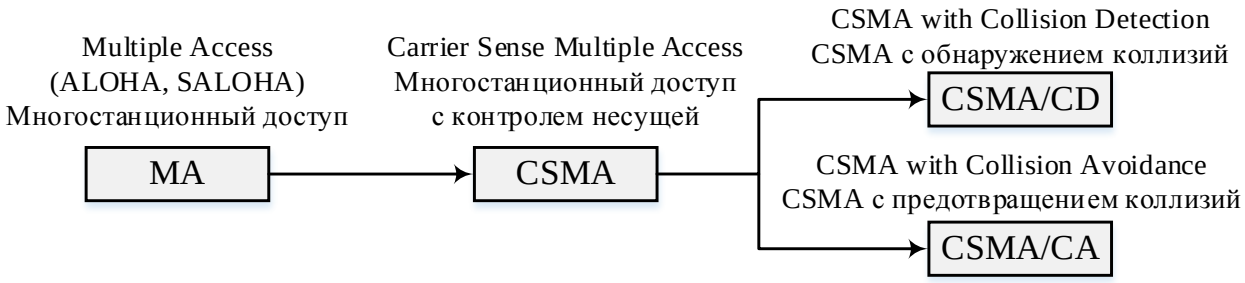


Рис. 2.23. Эволюция протоколов многостанционного доступа

Для уменьшения вероятности возникновения коллизий разработан метод многостанционного доступа с опросом несущей (Carrier Sense Multiple Access – CSMA), при котором станция «прослушивает» среду (совместный канал) перед попыткой ее использования.

Используется принцип «проверь перед тем, как передавать», аналогичный принципу «выслушай перед тем как говорить». Метод CSMA уменьшает вероятность возникновения коллизий, но не устраняет.

Причины возникновения коллизий в протоколах CSMA обусловлены следующими факторами: а) возможность коллизий существует из-за времени распространения сигнала, т. е. до перемещения кадра от одной станции к другой проходит определенное время; б) станция может «прослушать» канал и обнаружить его свободным, хотя другая станция уже собирается или начала передавать данные, которые еще не дошли.

Проблема возникновения коллизий в протоколах с опросом несущей CSMA представлена на рис. 2.24, 2.25 и обусловлена следующими обстоятельствами: а) после контроля несущей стация A передает кадр в момент t_1; станция Z в момент t_2 «слушает» канал и определяет, что он свободен (кадр станции A еще не дошел) и начинает передачу кадра; б) в момент t_3 кадры сталкиваются; искаженный сигнал распространяется в оба направления и достигает станции Z в момент t_4 и станции A в момент t_5.

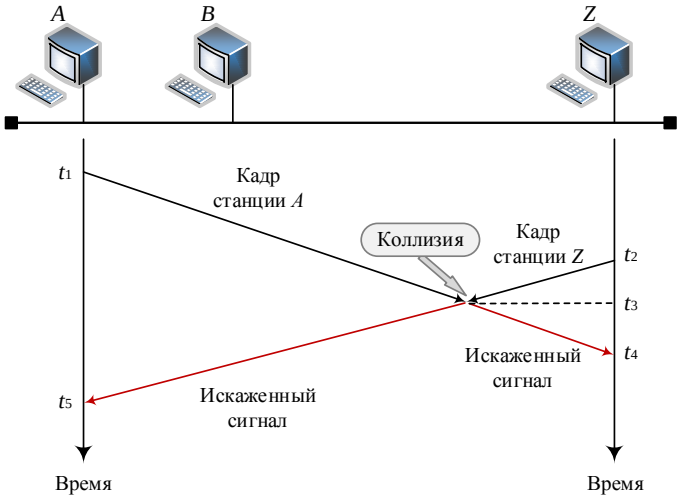


Рис. 2.24. Коллизии в протоколах CSMA

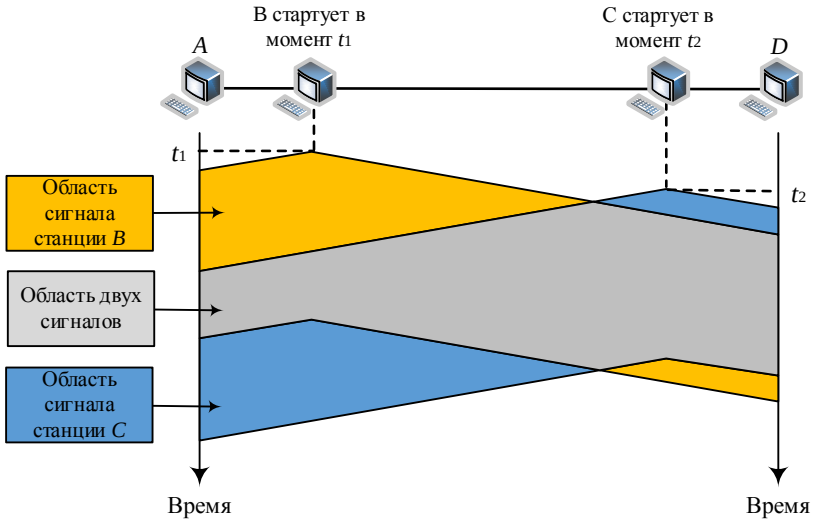


Рис. 2.25. Наложение сигналов в протоколах CSMA

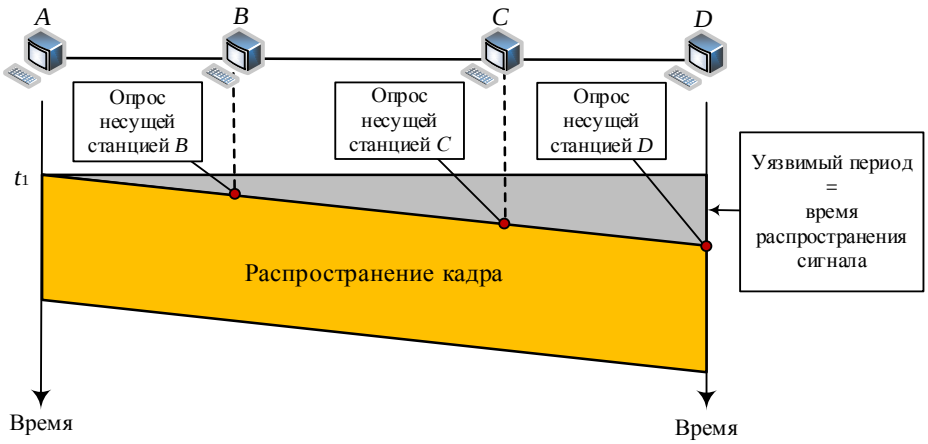


Рис. 2.26. Уязвимый период для возникновения коллизий
в протоколах CSMA

Уязвимый период в протоколах с опросом несущей CSMA представлен на рис. 2.26 и обусловлен следующими обстоятельствами: а) веро-

ятность возникновения коллизий зависит от продолжительности распространения сигнала от одной станции к другой; б) чем больше расстояние между станциями, тем больше время распространения сигала и, следовательно, уязвимый период для коллизий.

2.4.2. Стратегии настойчивости передачи в протоколах CSMA

Производительность канала зависит от того, каким способом осуществляется опрос несущей и как ведет себя станция при передаче данных после освобождения канала. Известны три стратегии настойчивости: стратегия с настойчивостью 1; ненастойчивая стратегия и стратегия с настойчивостью *p*.

Стратегия передачи с настойчивостью 1 представлена на рис. 2.27 и формулируется следующим образом: а) при появлении у станции данных для передачи она прослушивает канал и, если канал занят – ждет до тех пор, пока он освободится; б) станция передает кадр с вероятностью 1 как только обнаружит канал свободным; в) причины коллизий: задержки распространения сигнала и ожидание передачи несколькими станциями.

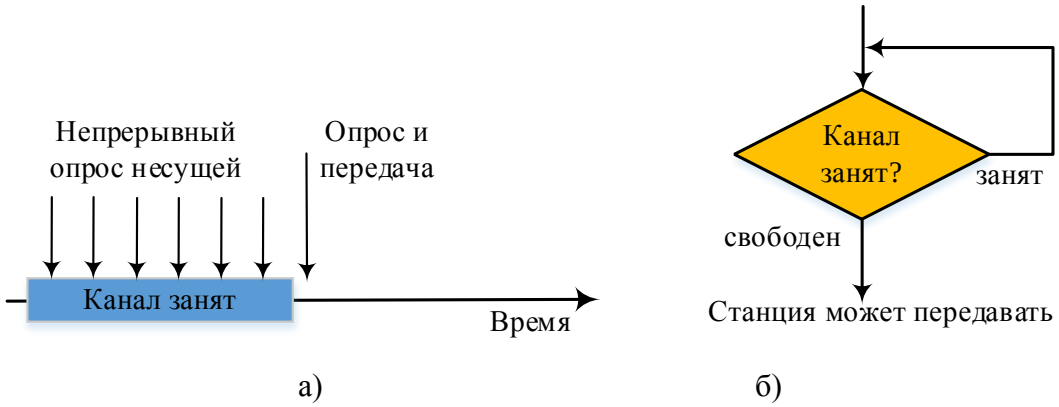


Рис. 2.27. Стратегия передачи с настойчивостью 1:
а) поведение станции; б) диаграмма событий

Ненастойчивая стратегия передачи представлена на рис. 2.28 и формулируется следующим образом: а) сдерживается стремление станции начинать передачу, как только освобождается канал; б) прежде чем начать передачу, станция опрашивает канал и, если он свободен, данные передаются немедленно; если канал занят, то станция ожидает в течение случайного интервала времени и затем снова прослушивает канал; в) данные начинают передаваться если при опросе канал оказывается свободным; г) при использовании ненастойчивой стратегии уменьшается вероятность коллизий, однако увеличивается общее время ожиданий.

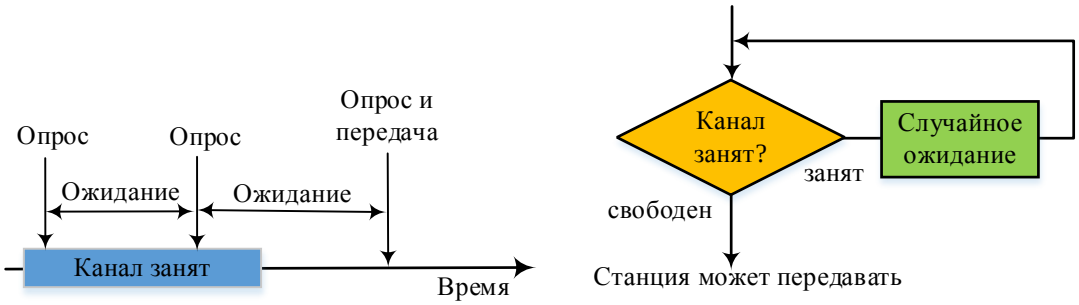


Рис. 2.28. Ненастойчивая стратегия передачи

Стратегия передачи с настойчивостью *p* представлена на рис. 2.29 и формулируется следующим образом: а) прежде чем начать передачу, станция опрашивает канал и, если он свободен, она с вероятностью p начинает передачу и с вероятностью $q = 1 - p$ ожидает начало следующего такта в течение фиксированного интервала; б) процесс продолжается до тех пор, пока кадр не будет передан или какая-либо другая станция не начнет передачу; в) если при ожидании другая станция начинает передачу, то станция ведет себя так же, как в случае столкновения, т. е. станция ждет в течение случайного интервала времени (отсрочка передачи) и начинает новый опрос несущей для передачи; г) если при первом прослушивании канал занят, то станция ожидает следующий интервал и т. д.

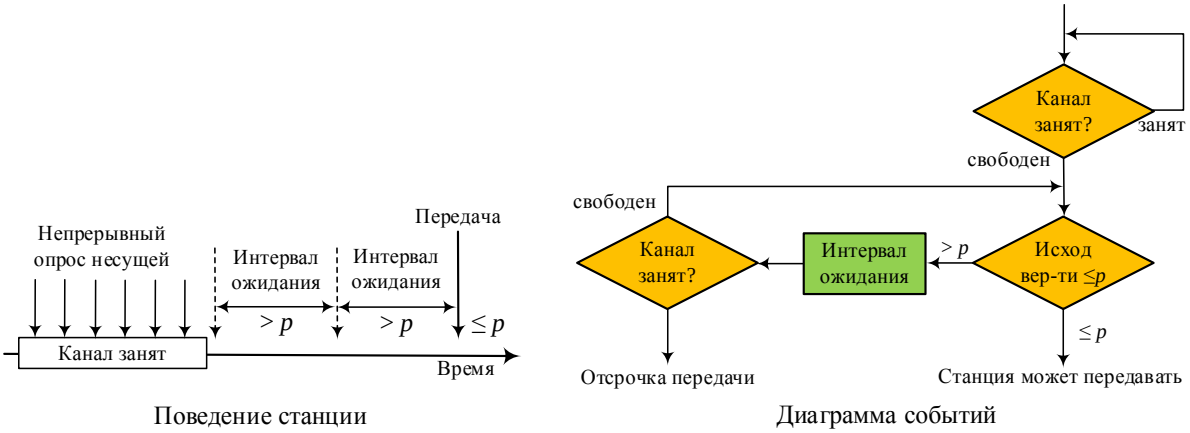


Рис. 2.29. Стратегия передачи с настойчивостью p

2.4.3. Протокол CSMA с обнаружением коллизий (CSMA/CD)

Особенности протокола CSMA/CD: а) метод доступа с опросом несущей CSMA со стратегиями настойчивости не рассматривает обнаружение коллизий, поэтому в чистом виде не нашел применения; б) при обнаружении коллизии станции разумно прервать передачу кадров, так как они уже не могут быть приняты получателем; в) процедура обнаружения

и обработки коллизий реализована в протоколе CSMA/CD (CSMA with Collision Detection); г) протокол CSMA/CD применяется в локальных сетях Ethernet в подуровне MAC.

Опрос несущей в протоколе CSMA/CD производится следующим образом: а) перед передачей кадра станция должна убедиться, что среда свободна; б) станция прослушивает основную гармонику сигнала (несущую), наличие частоты основной гармоники сигнала есть признак занятости канала. Например, в сетях Ethernet 10 Мбит/с с манчестерским кодом, частота основной гармоники сигнала меняется в пределах 5–10 МГц.

Обнаружение коллизий в протоколе CSMA/CD представлено на рис. 2.30 и происходит следующим образом: а) возможна ситуация, когда две и более станции одновременно попытаются передать кадр по общей среде, либо одна станция начнет передачу раньше второй, но до второй станции сигнал первой еще не дошел к моменту необходимости передачи кадра второй станцией; столкновение неизбежно; б) для корректной обработки коллизии все станции во время передачи кадра одновременно наблюдают за сигналами в кабеле; в) коллизия обнаруживается по мощности или длительности импульса принимаемого сигнала по сравнению с передаваемым; г) обнаружение коллизий – аналоговый процесс; применение манчестерского кода позволяет обнаружить столкновение символов в течение первого бита.

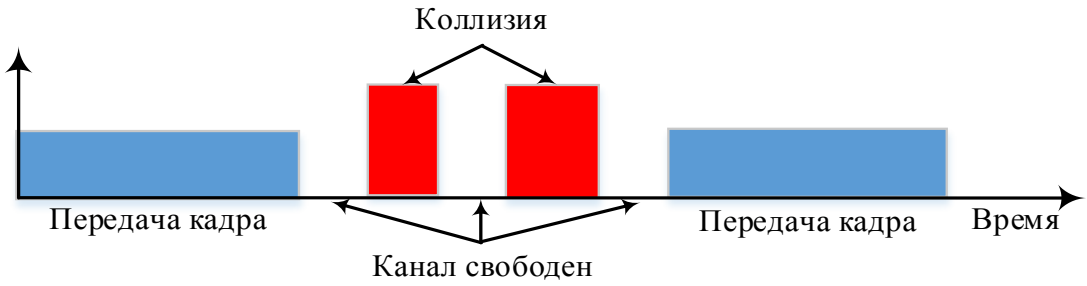


Рис. 2.30. Энергия сигнала в периоды передачи кадра,
свободного канала и коллизии

Прерывание передачи после обнаружения коллизий в протоколе CSMA/CD представлено на рис. 2.31 и происходит сразу, как только станция обнаружила коллизию. Для увеличения вероятности обнаружения коллизии всеми станциями сети, станция, обнаружившая коллизию, после прекращения передачи кадра передает в сеть специальную последовательность из 32 бит, называемой *jam*-сигналом, ситуация коллизии усиливается.

Условие обнаружения коллизий в протоколе CSMA/CD можно сформулировать следующим образом: а) необходимым условием корректной работы сети является распознавание коллизий всеми станциями, в противном случае кадр, переданный какой-либо станцией, будет отбракован

на приеме; б) в случае не обнаружения коллизии информация может быть повторно передана протоколом верхнего уровня, но с большими затратами по времени; в) в худшем случае сигнал пройдет дважды между наиболее удаленными друг от друга станциями сети (в одну сторону неискаженный сигнал и на обратном пути – сигнал, искаженный коллизией); затраченное на это время называется *временем двойного оборота* (Path Delay Value, PDV); г) передающая станция прослушивает канал, выявляя всплески шума, и коллизия кадра должна быть обнаружена в течение передачи этого кадра; д) при выполнении такого условия передающая станция успеет обнаружить коллизию, до того, как она закончит передачу кадра; е) время передачи кадра минимальной длины должно быть не менее времени двойного оборота T_{fr}: $T_{fr} \geq$ PDV.

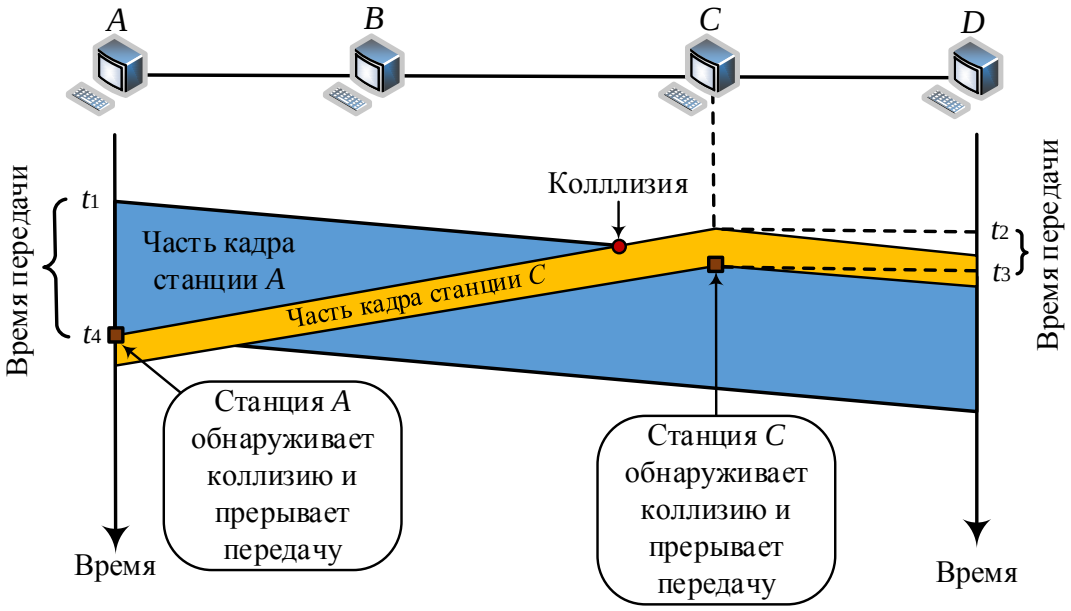


Рис. 2.31. Прерывание передачи
после обнаружения коллизий в протоколе CSMA/CD

Алгоритм функционирования протокола CSMA/CD представлен на рис. 2.32 и работает следующим образом: а) если в течение передачи кадра коллизия не обнаружена, то кадр передан успешно; б) при обнаружении коллизии в линию передается *jam*-сигнал для оповещения других станций, все станции отклоняют часть принятого кадра; в) счетчик попыток влияет на случайное время отсрочки; г) случайное число R для задания времени отсрочки повторной передачи генерируется в диапазоне от 0 до $2^K - 1$.

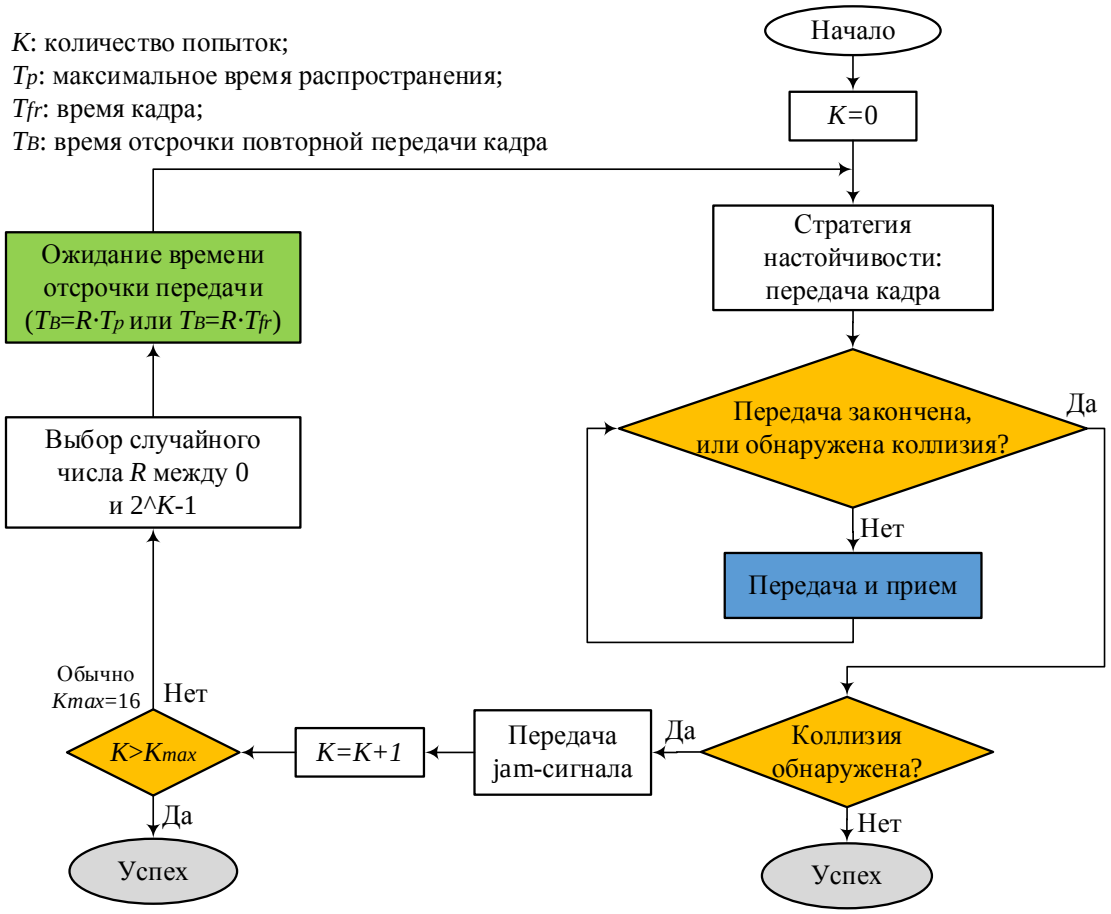


Рис. 2.32. Алгоритм функционирования для протокола CSMA/CD

Пример 5. В сети с протоколом CSMA/CD по разделяемой среде передачи на скорости 10 Мбит/с передаются кадры. Максимальное время распространения сигнала T_p между наиболее удаленными станциями (включая задержки на станциях и не учитывая время передачи *jam*-сигнала) равно 25,6 мс. Определите минимальный размер кадра.

Решение: время кадра минимальной длины $T_{fr} = 2 \times T_p = 51{,}2$ мс (т. е. в худшем случае для обнаружения коллизий время передачи кадра должно быть не менее 51,2 мс). Минимальный размер кадра равен 10 Мбит/с × 51,2 мс= 512 бит (64 байт). Это минимальный размер кадра для протокола Ethernet.

2.4.4. Протокол CSMA с устранением коллизий (CSMA/CA)

Алгоритм функционирования CSMA/CD представлен на рис. 2.33.

Особенности протокола CSMA/CA (CSMA/CA with Collision Avoidance) отличается от CSMA/CD процедурой устранения коллизий. В протоколе CSMA/CA используются подтверждения приема кадров ACK.

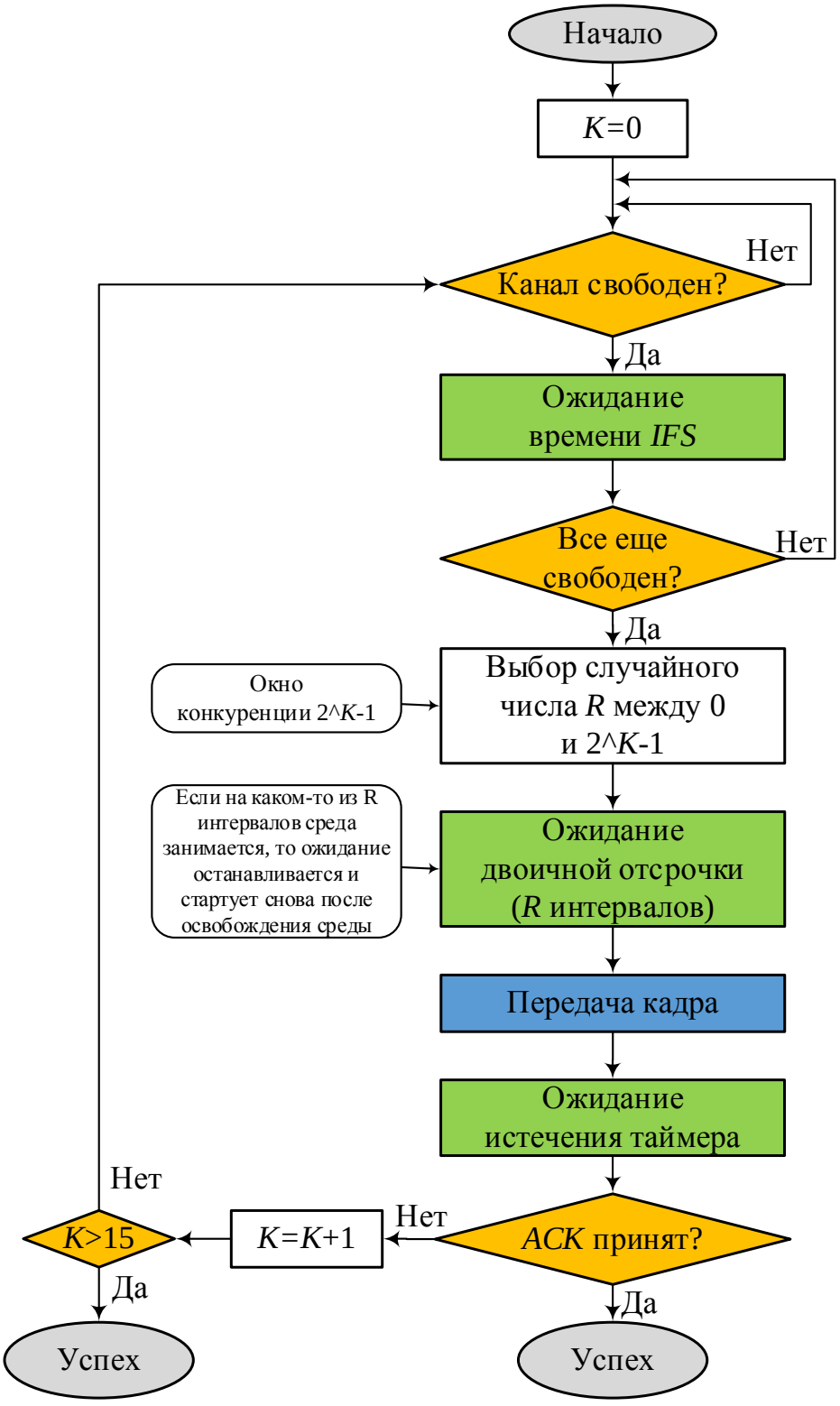


Рис. 2.33. Алгоритм функционирования
для протокола CSMA/CA

Алгоритм функционирования протокола CSMA/CA может быть сформулирован следующим образом: а) после обнаружения канала свободным станция ожидает Межкадровый интервал (Inter Frame Space – IFS); если

после интервала IFS среда снова занята, то станция продолжает наблюдение за средой (рис. 2.34), но если среда все еще свободна, то станция ждет случайное время *R* интервалов (окно конкуренции), передает кадр и устанавливает таймер; б) станция ожидает подтверждение и если оно получено до истечения таймера, значит кадр передан успешно, если нет, то при свободном канале станция ожидает время IFS и время двоичной экспоненциальной отсрочки (с каждой попыткой *K* диапазон случайных значений *R* удваивается).

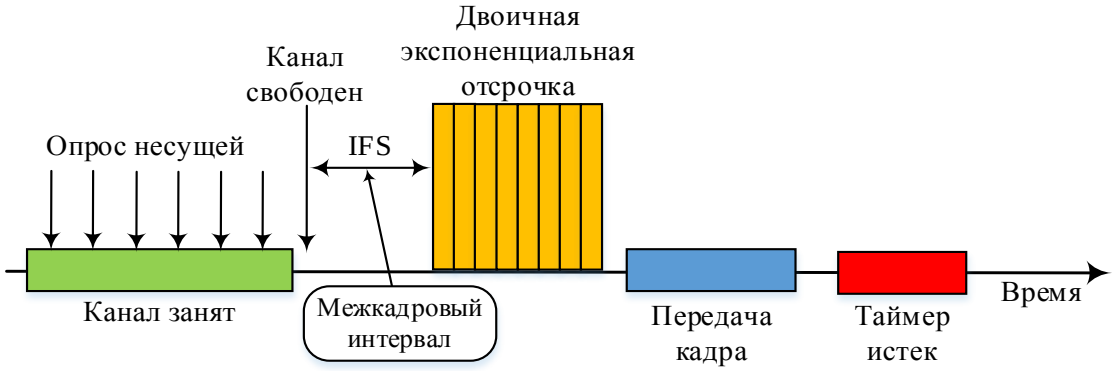


Рис. 2.34. Пример временной диаграммы для протокола CSMA/CA

Контрольные вопросы

1. Сформулируйте особенности метода произвольного доступа.
2. Какие процедуры необходимы для решения проблем произвольного доступа?
3. Сформулируйте принципы работы метода доступа с контролем несущей CSMA.
4. За счет чего вероятность возникновения коллизий в протоколе CSMA уменьшается по сравнению с протоколами ALOHA?
5. Каковы причины возникновения коллизий в протоколах CSMA?
6. Чем обусловлен уязвимый период в протоколах CSMA?
7. Перечислите стратегии настойчивости передачи в протоколах CSMA.
8. Проиллюстрируйте и сформулируйте стратегию передачи с настойчивостью 1.
9. Проиллюстрируйте и сформулируйте ненастойчивую стратегию передачи.
10. Проиллюстрируйте и сформулируйте стратегию передачи с настойчивостью *p*.
11. Сформулируйте принципы работы метода доступа CSMA/CD.
12. Поясните процедуру обнаружения коллизий в протоколе CSMA/CD.
13. Проиллюстрируйте и сформулируйте процедуру прерывания передачи после обнаружения коллизий в протоколе CSMA/CD.
14. Сформулируйте условие обнаружения коллизий в протоколе CSMA/CD.
15. Приведите и поясните алгоритм функционирования протокола CSMA/CD.
16. Приведите и поясните алгоритм функционирования протокола CSMA/CA.
17. Почему протокол CSMA/CD не используется в беспроводных сетях?
18. В чем отличие протокола CSMA/CA от CSMA/CD?
19. В чем назначение межкадрового интервала IFS в протоколе CSMA/CA?
20. Как передающая станция узнает о том, что переданный ею кадр принят успешно?

2.5. Лабораторная работа. Оценка производительности алгоритма CSMA

2.5.1. Постоение дискретно-событийной имитационной модели

В протоколе ненастойчивый (nonpersistent) np-CSMA пакет, сгенерированный станцией, передается тогда, когда среда освобождается, т. е. в состоянии IDLE. Временная диаграмма функционирования протокола np-CSMA представлена на рис. 2.35.

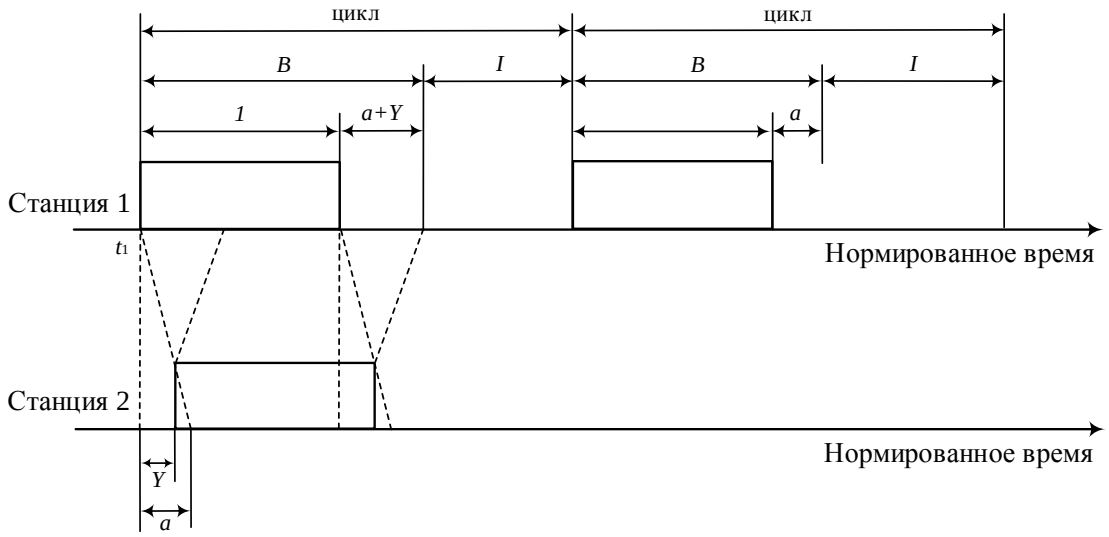


Рис. 2.35. Временная диаграмма функционирования протокола np-CSMA

Обозначим через B нормированную продолжительность состояния BUSY, через I – нормированную продолжительность состояния IDLE и через U – нормированную продолжительность состояния успешной передачи кадров без коллизий. Тогда пропускную способность протокола np-CSMA можно определить по формуле

$$S = \frac{U}{B + I}. \qquad (2.25)$$

Допустим, что длительности кадров и задержка распространения нормированы так, что длительность кадра равна 1, а задержка распространения равна α. Тогда вероятность успешной передачи кадра, сгенерированного станцией в момент времени t_1 определяется вероятностью того, что другие станции не сгенерируют кадр в интервале времени от t_1 до $t_1 + \alpha$.

Пусть G – нормированная нагрузка, тогда нормированная продолжительность состояния передачи кадров без коллизий U может быть определена выражением

$$U = G \cdot e^{-\alpha G}. \qquad (2.26)$$

Нормированная продолжительность состояния IDLE I, когда кадры не генерируются, подчиняется экспоненциальному распределению и может быть представлена выражением

$$I = \frac{1}{G}. \tag{2.27}$$

Допустим, другая станция сгенерирует кадр в интервале от t_1 до $t_1+\alpha$ в момент времени $t_1 + Y$, где $\overline{Y}$ – среднее значение Y (рис. 2.35), тогда нормированную продолжительность состояния BUSY B можно представить следующим образом

$$B = 1 + \alpha + \overline{Y}. \tag{2.28}$$

Распределение величины Y определяется вероятностью того, что в интервале длительностью α-y не будет сгенерировано ни одного кадра

$$F_Y(y) \cong \Pr\{Y \le y\} = e^{-(\alpha - y)G},\ y \le \alpha. \tag{2.29}$$

Среднее значение величины Y определяется выражением

$$\overline{Y} = \alpha - \frac{1}{G}\left(1 - e^{-\alpha G}\right). \tag{2.30}$$

Подставив (2.30) в (2.28) получим

$$B = 1 + 2\alpha - \frac{1}{G}\left(1 - e^{-\alpha G}\right). \tag{2.31}$$

Подставив (2.26), (2.27) и (2.31) в (2.25), получим

$$S = \frac{G \cdot e^{-\alpha G}}{G(1 + 2\alpha) + e^{-\alpha G}}. \tag{2.32}$$

Величина в (2.32) вычисляется следующим образом

```
% теоретическая пропускная способность для протокола CSMA
S=Traffic.*exp(-0.1*Traffic)./(Traffic*(1+2*0.1)+exp(-0.1*Traffic));
```

2.5.2. Оценка производительности протокола CSMA

На рис. 2.36 представлен пример оценки производительности протокола CSMA (скрипт 2.3).

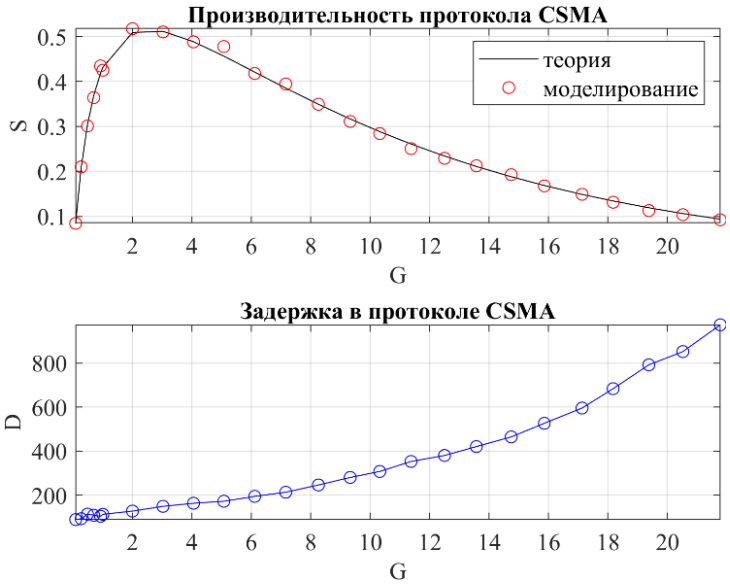


Рис. 2.36. Производительность протокола CSMA

Скрипт 2.3. Оценка производительности протокола CSMA

```
clear all; clc;
% np-CSMA
STANDBY=0;   % состояние дежурного приема
TRANSMIT=1;  % состояние передачи
COLLISION=2; % состояние коллизии
TOTAL=1000;  % общее количество кадров для передачи

Brate=0.25e6;       % символьная скорость (симв/с)
Plen=500;           % длина кадра (симв)
Ttime=Plen/Brate;   % время передачи кадра (с)
Dtime=0.1;          % задержка распространения кадра (% от Ttime)
delay=Dtime*Ttime; % задержка распространения кадра (с)
Mnum=100;           % число станций

% сгенерированный траффик/нагрука (кадров за время передачи кадра)
G=[0.1:0.2:1,1:1:20];
for indx=1:length(G) % цикл по нагрузке

    Tint=-Ttime/log(1-G(indx)/Mnum); % интервал генерирования кадров
    % (чем больше нагрузка G, тем меньше Tint,
    % т.е. тем чаще генерируются новые кадры)
    Rint=Tint;
    Spnum=0;     % число успешно переданных кадров
    Splen=0;     % объем успешно переданных кадров (симв)
    Tplen=0;     % объем сгенерированных кадров (симв)
    Wtime=0;     % задержка передачи сгенерированных кадров (с)

    mgtime=-Tint*log(1-rand(1,Mnum)); % вектор времен генерирования кадров
    mtime=mgtime;                     % время текущего события  (с)
    Mstate=zeros(1,Mnum);             % вектор состояний станций
    Mplen(1:Mnum)=Plen;               % вектор длина кадра (симв)
    now_time=min(mtime);              % текущее время (с)
    Mstime=zeros(1,Mnum);             % вектор времен передачи кадров (с)

    while Spnum<TOTAL
```

```
        % нахождение станций, начинающих передачу
        idx=find(mtime==now_time & Mstate==TRANSMIT);
        if length(idx)>0
            Spnum=Spnum+1;
            Splen=Splen+Mplen(idx);
            Wtime=Wtime+now_time-mgtime(idx);
            Mstate(idx)=STANDBY;
            mgtime(idx)=now_time-Tint*log(1-rand);
            mtime(idx)=mgtime(idx);
        end

        % нахождение станций в состоянии коллизии
        idx=find(mtime==now_time & Mstate==COLLISION);
        if length(idx)>0
            Mstate(idx)=STANDBY;
            mtime(idx)=now_time-Rint*log(1-rand(1,length(idx)));
        end

        % нахождение станций в состоянии ожидания
        idx=find(mtime==now_time & Mstate==STANDBY);
        if length(idx)>0
            Tplen=Tplen+sum(Mplen(idx));
            for ii=1:length(idx)
                jj=idx(ii);
                idx1=find((Mstime+delay)<=...
                    now_time & now_time<=(Mstime+delay+Ttime));
                if length(idx1)==0
                    Mstate(jj)=TRANSMIT;
                    Mstime(jj)=now_time;
                    mtime(jj)=now_time+Mplen(jj)/Brate;
                else
                    mtime(jj)=now_time-Rint*log(1-rand);
                end
            end
        end

        % нахождение станций в состоянии передачи или коллизии
        idx=find(Mstate==TRANSMIT | Mstate==COLLISION);
            if length(idx)>1
                Mstate(idx)=COLLISION;
            end
        now_time=min(mtime);
    end
    Traffic(indx)=Tplen/Brate/now_time;
    S(indx)=Splen/Brate/now_time;
    Delay(indx)=Wtime/TOTAL*Brate/Plen;
end

% теоретическая пропускная способность для протокола CSMA
Stheory=Traffic.*exp(-0.1*Traffic)./(Traffic*(1+2*0.1)+exp(-0.1*Traffic));
subplot(2,1,1); plot(Traffic,Stheory,'-k',Traffic,S,'ro'); grid on;
axis('tight');
xlabel('G'); ylabel('S'); legend('теория', 'моделирование');
title('Производительность протокола CSMA');
subplot(2,1,2); plot(Traffic,Delay,'-bo'); grid on;
xlabel('G'); ylabel('D'); axis('tight'); title('Задержка в протоколе CSMA');
```

Контрольные вопросы

1. Сформулируйте особенности работы протокола np-CSMA.
2. Приведите и поясните временную диаграмму функционирования протокола np-CSMA.
3. Как нормированная продолжительность состояния BUSY, IDLE и нормированная продолжительность состояния передачи кадров без коллизий связаны с пропускной способностью протокола np-CSMA?
4. Как нормированная нагрузка связана с нормированной продолжительностью состояния передачи кадров без коллизий в протоколе np-CSMA?
5. Как нормированная продолжительность состояния IDLE связана с нормированной нагрузкой в протоколе np-CSMA?
6. Чем определяется вероятностный характер параметра нормированной продолжительности состояния BUSY в протоколе np-CSMA?
7. Сравните пропускную способность протокола CSMA и ALOHA/SALOHA. Поясните, почему в первом случае она выше, используя предпосылки построения дискретно-событийной имитационной модели.
8. Сравните задержку передачи протоколов CSMA и ALOHA/SALOHA. Поясните, почему в первом случае она ниже, используя предпосылки построения дискретно-событийной имитационной модели.
9. Какой из протоколов CSMA или ALOHA/SALOHA больше подходит для передачи трафика реального времени и почему?
10. В каком из протоколов CSMA или ALOHA/SALOHA больше коллизий и почему?

3. РАДИОКАНАЛ СЕТЕЙ РАДИОДОСТУПА

3.1. Практическое занятие. Потери распространения и замирания в СРД

Цель занятия: изучить физические явления и математические модели в радиоканале сетей радиодоступа.

3.1.1. Модели распространения радиоволн в СРД

Модель распространения радиоволн (РРВ) в сетях радиодоступа (СРД) и способ отображения реальных условий распространения в выбранной модели являются основными факторами, определяющими выбор типа модуляции, мощности передатчика и структуры приемника. Условия РРВ в СРД могут варьироваться от простейшей ситуации однолучевого распространения радиоволн между приемником и передатчиком в условиях прямой видимости до многолучевого распространения при многократных отражениях от искусственных сооружений и особенностей рельефа местности в условиях допплеровского изменения частоты при движении объекта или препятствий. В отличие от проводных линий связи радиоканал является статистической системой, свойства которой определяются только с некоторой вероятностью. Результаты расчета параметров радиоканала в значительной степени зависят от выбранной модели канала. Модели, основанные на одних и тех же принципах, могут различаться способом отображения реальной ситуации. Не существует единой общепринятой модели расчета электромагнитного поля в городских условиях. Рекомендации различных национальных и международных организаций в значительной степени отличаются друг от друга. Сложность выбора адекватной модели распространения радиоволн усугубляется трудностями практического определения параметров модели и сравнения качества работы систем связи, основанных на различных моделях расчета электромагнитного поля (ЭМП). Измерение параметров модели может быть только вероятностным и требует проведения огромного числа испытаний в самых различных условиях (крупный город или небольшой населенный пункт, вид подстилающей поверхности, особенности рельефа местности и т. д.). Известные формулы и рекомендации носят эмпирический характер и являются прямым обобщением экспериментальных данных. Существует три основных механизма, воздействующих на распространение радиоволн в сетях радиодоступа [4].

Отражение (*reflection*) происходит тогда, когда распространяющаяся электромагнитная волна сталкивается с гладкой поверхностью, размер которой гораздо больше длины волны λ радиочастотного сигнала.

Дифракция (*diffraction*) встречается тогда, когда путь распространения между передатчиком и приемником преграждается плотным телом,

размеры которого велики по сравнению с λ, что вызывает появление вторичных волн, образующихся позади преграждающего тела. Дифракция – это явление, которое является причиной того, что распространение радиочастотной энергии от передатчика к приемнику происходит в обход пути прямой видимости между ними. Ее часто называют затенением (*shadowing*), поскольку дифрагированное поле может достичь приемника, даже если оно затенено непроницаемой преградой.

Рассеяние (*scattering*) встречается тогда, когда радиоволна сталкивается с любой неровной поверхностью или с поверхностью, размеры которой порядка λ или меньше, что приводит к распространению (рассеянию) или отражению энергии во всех направлениях. В городской местности обычные препятствия, вызывающие рассеивание сигнала, – это фонарные столбы, уличные знаки и листья. Название *рассеивающий элемент* (*scatterer*) применимо к любым препятствиям на пути распространения, которые являются причиной отражения или рассеяния сигнала.

Далее рассматриваются наиболее распространенные методы расчета дальности радиосвязи, а также типы искажений и замираний сигналов при распространении в условиях многолучевого канала.

Целью анализа РРВ является расчет дальности радиосвязи и определение реальных характеристик принимаемого сигнала. Классический подход к расчету распределения ЭМП в присутствии отражающих и поглощающих объектов заключается в расчете напряженности поля в однородном изотропном пространстве на основе законов отражения, дифракции и рассеяния. Однако специфические условия города исключают возможность непосредственного применения такой методики. Непостоянство расположения приемников и передатчиков в сетях радиодоступа, перемещение приемников, передатчиков и препятствий, огромное количество фиксированных препятствий сложной формы делают невозможным точный расчет распределения электромагнитного поля. Возникающие при таких расчетах трудности описания реального расположения и передвижения препятствий, а также требуемый объем вычислений превосходят существующие технические возможности. Поэтому точный аналитический расчет распределения поля используется только в исключительных, простейших случаях, например, при расчете теневой зоны за очень большим зданием при точно известном положении передатчика базовой станции. Реальный расчет распределения электромагнитного поля осуществляется на основе двух моделей – «большого расстояния» (*large scale model*) и «малого расстояния» (*small scale model*) [4].

В модели «большого расстояния» рассматривается влияние на электромагнитное поле макроэффектов, обусловленных препятствиями большого размера (по сравнению с длиной волны). Согласно этой модели, электромагнитное поле в городских условиях описывается теми же самыми

уравнениями, что и для свободного пространства, но с иными параметрами распространения, а также некоторой вероятностью отклонения реальных значений распределения электромагнитного поля от расчетных. Предполагается, что наличие препятствий «в среднем» не влияет на структуру электромагнитного поля, которое остается таким же, как и в свободном пространстве, а именно, стационарным, монотонным и гладким. Стационарность означает неизменность структуры поля во времени, монотонность – непрерывное убывание величины поля с увеличением расстояния от передатчика до приемника, гладкость – соответствие малых изменений расстояния малым изменениям напряженности поля [4].

Параметры распространения радиоволн в городе отличаются от параметров распространения в свободном пространстве. Напряженность ЭМП в городских условиях уменьшается с расстоянием значительно быстрее, чем вторая степень расстояния, из-за рассеяния электромагнитных волн на многочисленных препятствиях. В результате взаимодействия с препятствием только некоторая часть мощности передатчика дойдет до приемника, остальная часть либо будет поглощена препятствием, либо отразится под произвольным углом и пройдёт мимо приемника. Кроме того, уменьшающаяся «в среднем» напряженность поля реально испытывает флуктуации, вызванные экранирующим действием отдельных зданий, сооружений и складок местности. Распределение теневых и освещенных областей в сложной, нерегулярной городской застройке и пересеченной местности с большой долей вероятности можно считать случайным. В результате напряженность поля в каждой точке пространства лишь с некоторой вероятностью равна средней, реально испытывая случайные флуктуации около среднего значения, монотонно уменьшающегося по мере удаления от передатчика. Принято говорить, что флуктуации напряженности поля при перемещении в городе вызывают «медленные замирания» сигнала. Практически глубина медленных замираний, зависящая от величины дисперсии случайного распределения напряженности поля, определяет процент территории, на которой величина напряженности поля превышает заданную пороговую величину [4].

Модель «большого расстояния» лежит в основе всех методик расчета дальности радиосвязи, отличающихся друг от друга только способом введения коэффициентов коррекции, отражающих реальные условия распространения, в формулы распространения поля в свободном пространстве. Все варианты определения поправочных коэффициентов к скорости уменьшения поля с расстоянием, а также дисперсии случайного отклонения напряженности поля от среднего значения опираются на экспериментальные данные, полученные в различных городах, на разных частотах, в различных географических условиях, в разное время суток и т. д. Результатом расчета по модели «большого расстояния» является вероятное значение напряженности поля на некотором расстоянии от передатчика. Например, расчет

может показать, что при удалении от передатчика на расстояние, не превышающее R, заданная напряженность поля E достигается с вероятностью $p_1\%$ на $p_2\%$ территории. В диапазоне УКВ (от 30 МГц (длина волны 10 м) до 3000 МГц (длина волны 0,1 м)), где дальность связи часто определяется расстоянием до горизонта с высоты подъема антенны базовой радиостанции, рассчитывается необходимая мощность передатчика P, которая на всей территории от передатчика до горизонта обеспечивает заданную напряженность поля E с вероятностью не менее $p_1\%$ на $p_2\%$ территории. Расчет усредненного поля в приближении «большого расстояния» применяется при проектировании сетей связи, для оптимизации расположения и величины мощности передатчиков путем определения размеров зоны уверенного приема, зон взаимного перекрытия передатчиков, теневых и освещенных зон и т. д. [4].

Модель «малого расстояния» отражает интерференционную структуру электромагнитного поля, возникающую вследствие взаимодействия когерентных волн, излученных передатчиком. Суммарная величина электромагнитного поля в каждой точке пространства определяется амплитудами и фазами нескольких когерентных волн, которые за счет многократных отражений прошли путь различной длины от передатчика до данной точки приема. Очевидно, что на значительном расстоянии от передатчика амплитуды и фазы волн статистически независимы и в результате получается интерференционная картина поля в виде случайного чередования максимумов (сложение в фазе) и минимумов (сложение в противофазе) поля. Поскольку расстояние между минимумами и максимумами в интерференционной картине поля равно половине длины волны, то и существенные изменения величины напряженности поля также происходят на очень малых расстояниях, порядка нескольких сантиметров в диапазоне УКВ. Очевидно, что структура поля на малых расстояниях в небольшой области пространства уже не гладкая, не монотонная и не стационарная. Увеличение или уменьшение напряженности поля в локальной области приема не связано с расстоянием до передатчика, так как определяется случайным состоянием многолучевого радиоканала (взаимным расположением и передвижением приемника, передатчика и препятствий) в текущий момент времени. В результате возможны очень сильные изменения величины электромагнитного поля на небольших расстояниях и в короткие промежутки времени. С точки зрения теории сигналов нестационарная интерференционная структура поля соответствует приему нескольких копий одного и того же сигнала. Сигнал передатчика достигает приемника несколькими путями различной длины, что и приводит к появлению в приемнике нескольких копий сигнала, каждая из которых имеет собственное время распространения. Накладывающиеся друг на друга копии сигнала вызывают искажения формы принимаемого

сигнала. Эти искажения характеризуются как «быстрые замирания» величины принимаемого сигнала. Величина быстрых замираний принимаемого сигнала определяется мгновенным состоянием многолучевого канала распространения радиоволн, т. е. перемещением передатчика, приемника и препятствий между ними, а также скоростью этих перемещений. Практический запас на быстрые замирания определяет процент времени, в течение которого величина напряженности поля превышает заданную величину [4].

3.1.2. Оценка дальности связи в СРД

Основой расчета дальности связи в модели «большого расстояния» является формула потерь РРВ в свободном пространстве

$$L(d) = \left(\frac{4\pi d}{\lambda}\right)^2, \tag{3.1}$$

где d – это расстояние между передатчиком и приемником; λ – длина волны распространяемого сигнала.

При таком идеальном распространении мощность принятого сигнала на заданном расстоянии d от передатчика определяется выражением

$$P_{rx}(d) = \frac{P_t G_{tx} G_r \lambda^2}{(4\pi)^2 d^2}, \tag{3.2}$$

где P_t – это мощность передатчика; G_t – усиление антенны передатчика; G_r – усиление антенны приемника.

Для практических расчетов потери РРВ в свободном пространстве $L_0(d)$ оцениваются в логарифмическом виде (в дБ) из (3.2)

$$L_0(d) = 10\lg\left(\frac{P_{tx}}{P_{rx}}\right) = -10\lg\left(\frac{G_t G_r \lambda^2}{(4\pi d)^2}\right). \tag{3.3}$$

Без учета коэффициентов усиления передатчика и приемника выражение (3.3) преобразуется к виду

$$L_0(d) = 10\lg\left(\frac{P_{tx}}{P_{rx}}\right) = 20\lg\left(\frac{4\pi d}{\lambda}\right). \tag{3.4}$$

Вообще, модели распространения как для комнатных, так и для наружных каналов показывают, что средние потери в тракте $L_{mean}(d)$, как функция расстояния между передатчиком и приемником d, пропорциональны n-й степени d, выраженного в единицах эталонного расстояния d_0

$$L_{mean}(d) = L_0(d_0) + 10n\lg\left(\frac{d}{d_0}\right), \qquad (3.5)$$

где $n = 2,...,5$ – коэффициент затухания радиоволн. Превышение величины этого коэффициента над теоретическим значением n = 2 для свободного пространства отражает величину дополнительных потерь вследствие поглощения и отражения радиоволн естественными и искусственными препятствиями. На величину n влияет плотность городской застройки, преобладающий тип зданий (бетон, кирпич, дерево), характер подстилающей поверхности (земля, вода, лес).

Эталонное расстояние d_0 соответствует точке, размещенной в дальнем поле передающей антенны (параметр d_0 можно определить как «начало» дальней зоны, $d_0 > \lambda$). Обычно значение d_0 берется равным 1 км для крупных ячеек, 100 м – для микроячеек и 1 м – для комнатных каналов. Начальные потери $L_0(d_0)$ рассчитываются с помощью уравнения (3.4) или измеряются. Коэффициент потерь мощности $L_0(d_0)$ от выхода передатчика до точки, находящейся в непосредственной близости d_0 от передатчика в (3.5), может также включать параметры антенно-фидерного тракта. $L_{\text{mean}}(d)$ – это средние (по множеству местоположений) потери для данного значения d.

Таким образом, из (3.2) и (3.5) среднюю мощность принятого сигнала без учета коэффициентов усиления передатчика и приемника в логарифмическом виде (в дБ) можно определить следующим выражением

$$P_{\text{mean}}(d) = P_t - L_0(d_0) - 10n\lg\left(\frac{d}{d_0}\right). \qquad (3.6)$$

Формула (3.6) для расчета мощности принимаемого сигнала в городских условиях определяет среднюю (наиболее вероятную) величину мощности на заданном расстоянии от передатчика. В присутствии поглощающих и отражающих радиоволны объектов искусственного или естественного происхождения реальная величина принимаемого сигнала отличается от средней величины. Вследствие нерегулярного расположения препятствий в области распространения радиоволн можно предположить, что распределение зон, где мощность больше или меньше среднего значения (так называемых теневых и освещенных зон), будет чисто случайным. Экспериментальные данные показывают, что вероятность случайного отклонения мощности принимаемого сигнала в произвольном месте от среднего значения, предсказанного формулой (3.6), определяется нормальным логарифмическим законом распределения с нулевым средним значением:

$$p\left(P_\sigma\right)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{P_\sigma^2}{2\sigma^2}\right), \qquad (3.7)$$

где P_σ – случайное относительное отклонение мощности от среднего значения, дБ; σ – среднеквадратическое отклонение (СКО) распределения P_σ, дБ; σ^2 – дисперсия распределения P_σ.

Дисперсия σ нормального логарифмического (логнормального) распределения определяет взаимосвязь между величиной случайного отклонения мощности принимаемого сигнала P_σ от среднего значения $P_{\text{r\_mean}}$ и вероятностью этого отклонения. Одновременно величина дисперсии является численной мерой случайных медленных замираний (флуктуаций) уровня принимаемого сигнала при перемещениях радиостанции по освещенным и теневым участкам в зоне радиосвязи. При расчете дальности радиосвязи возможная глубина медленных замираний фактически определяет запас по мощности передатчика, необходимый для обеспечения минимально допустимого уровня принимаемого сигнала.

Независимость величины случайного отклонения мощности сигнала P_σ от расстояния d между передатчиком и приемником совершенно очевидна, так как относительная величина отклонения мощности принимаемого сигнала от среднего значения (уменьшение или увеличение мощности принимаемого сигнала по сравнению с ожидаемым средним значением) определяется только геометрией препятствий и их взаимным расположением. Экспериментальные измерения показывают, что в большинстве случаев отклонения реальной мощности от средней в типовых городских условиях в диапазоне УКВ не превышают величины 10–15 дБ. Другими словами, почти в любой точке зоны действия передатчика мощность принимаемого сигнала может отличаться от среднего значения на несколько децибел, эта величина не зависит от расстояния до передатчика и определяется только плотностью застройки и величиной перепада высот зданий или естественных складок местности. С учетом (3.7) формула (3.6) для средней мощности принимаемого сигнала на расстоянии d от передатчика преобразуется в уравнение для вероятного реального значения мощности принимаемого сигнала при наличии случайно распределенных препятствий:

$$P\left(d\right)=P_t-L_0\left(d_0\right)-10n\lg\left(\frac{d}{d_0}\right)-P_\sigma. \qquad (3.8)$$

На рис. 3.1 (скрипт 3.1) представлен график зависимости средней мощности принимаемого сигнала P_{mean} и вероятной реальной мощности принимаемого сигнала P, рассчитанных по формуле (3.8) для мощности пе-

редатчика $P_t = 30$ дБм, начальных потерь $L_0(d_0) = 30$ дБ, коэффициента затухания $n = 3$ и СКО логарифмического нормального распределения $\sigma = 6$ дБ в зависимости от нормированного расстояния d/d_0.

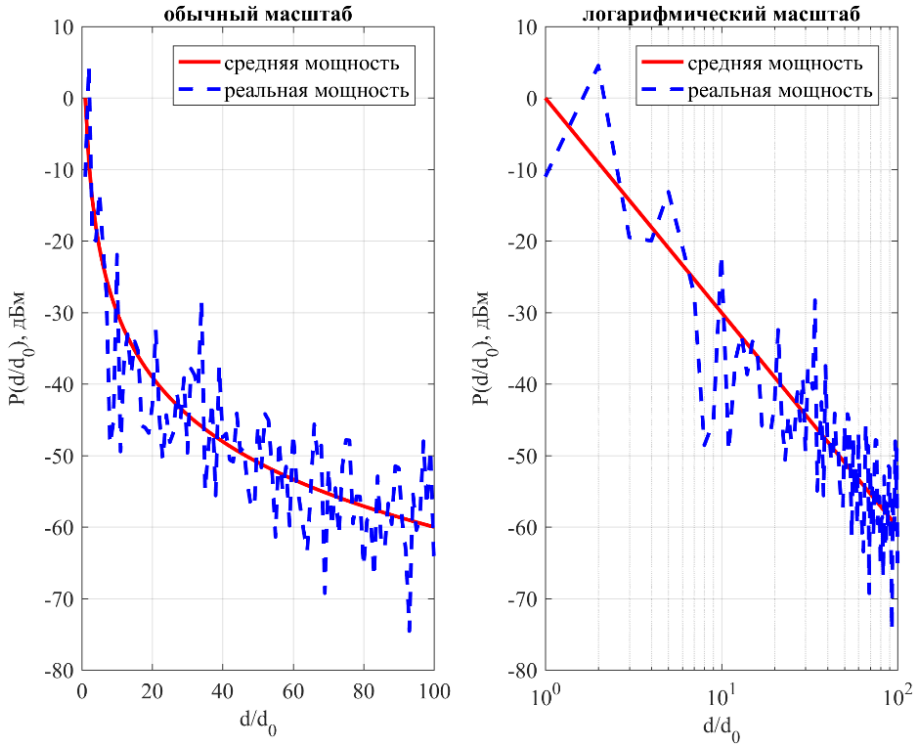


Рис. 3.1. Зависимость мощности принимаемого сигнала от нормированного расстояния

Скрипт 3.1. График мощности принимаемого сигнала

```
Pt=30;          % дБм
PL0=30;         % дБ
n=3;
sigma=6;        % дБ
d_d0=0:1:100;
PL_mean=Pt-PL0-10*n.*log10(d_d0);
PL=Pt-PL0-10*n.*log10(d_d0)-randn(1,length(d_d0))*sigma;
subplot(1,2,1);
plot(d_d0,PL_mean,'r-', d_d0,PL,'b--','linewidth',2);
legend('средняя мощность','реальная мощность');
xlabel('d/d_{0}'); ylabel('P(d/d_{0}), дБм'); grid on;
title('обычный масштаб');
subplot(1,2,2);
semilogx(d_d0,PL_mean,'r-', d_d0,PL,'b--','linewidth',2);
legend('средняя мощность','реальная мощность');
xlabel('d/d_{0}'); ylabel('P(d/d_{0}), дБм'); grid on;
title('логарифмический масштаб');
```

Из графика зависимости средней мощности принимаемого сигнала от нормированного расстояния в логарифмическом масштабе видно, что получилась прямая линия с наклоном, равным $10n$.

Таким образом, мощность принимаемого сигнала является случайной величиной, отклоняющейся от среднего значения (3.6) на величину P_σ,

а соответствующие потери в тракте $L(d)$ можно выразить через средние потери $L_{\text{mean}}(d)$, введя в уравнение (3.5) случайную переменную P_σ:

$$PL(d) = L_0(d_0) + 10n\lg(d/d_0) + P_\sigma. \qquad (3.9)$$

Здесь, как и в (3.8), P_σ обозначает случайную гауссову переменную с нулевым средним (в децибелах) со среднеквадратическим отклонением σ (также в децибелах). Поскольку P_σ и $L(d)$ – это случайные переменные, то, если для вычисления потерь в тракте использовать уравнение (3.9), предварительно нужно выбрать какое-то определенное значение P_σ. Часто выбор этого значения основывается на измерениях (сделанных для большого числа взаимных расположений передатчика и приемника). Обычные значения P_σ – это 6–10 дБ или даже выше. Таким образом, для статистического описания потерь в тракте при произвольном расположении с определенным расстоянием между передатчиком и приемником будут необходимы следующие параметры: 1) эталонное расстояние d_0; 2) показатель степени потерь в тракте n; 3) среднеквадратическое отклонение σ.

Уравнение мощности принимаемого сигнала (3.8) в зависимости от расстояния, мощности передатчика и условий распространения радиоволн является основой для расчета дальности радиосвязи. Различные методики расчета по (3.8) различаются способом определения регулярной и случайных составляющих мощности принимаемого сигнала в зависимости от конкретных условий местности. Другими словами, основной задачей при расчете дальности радиосвязи является определение соотношений параметров модели d_0, n, σ с реальными параметрами среды распространения радиоволн. Наибольшая точность расчетов обеспечивается при непосредственном применении уравнения, описывающего модель «большого расстояния» (3.8), с использованием экспериментально измеренных параметров модели d_0, n, σ для конкретного региона. Для приближенного расчета вероятностной дальности радиосвязи используются методики на основе рекомендаций МККР (Международный Консультативный Комитет по Радиосвязи), рекомендаций EUROCOST (Европейское Объединение для Научных и Технических Исследований), а также методики, рекомендованные различными стандартами радиосвязи. Все методики основаны на результатах статистической обработки большого количества экспериментальных данных, которые определяют параметры модели d_0, n, σ для типовых условий распространения, параметров радиооборудования и диапазона частот.

Основные методики расчета дальности радиосвязи (МККР, EUROCOST, модель Хата и т. д.) ориентированы в основном на сотовые и транкинговые сети связи. Эти сети характеризуются относительно большой

дальностью радиосвязи (до нескольких десятков километров) и относительно большой высотой подъема передающих антенн. При этих условиях предположения о «среднем» значении мощности оправдывается в наибольшей степени. Для расчета дальности связи в сотах небольшого размера эти методики применимы уже с большими оговорками, так как предположение о «средней» равномерности поля плохо оправдывается на расстояниях в несколько сот метров и тем более не применимо для расчета затухания поля в зданиях. Методики, рекомендованные стандартами радиосвязи, учитывают конкретные особенности распределения поля на относительно небольших расстояниях в рабочем диапазоне частот сети связи. Оценка радиуса действия радиосвязи в помещениях, как правило, производится по известным экспериментальным данным, обязательно указывающим место и условия измерений. В любом случае результат расчета дальности радиосвязи может быть только вероятностным и достоверным настолько, насколько реальные условия распространения совпадают со «средними», а введенные в модель поправочные коэффициенты отражают конкретные условия РРВ.

На рис. 3.2 (скрипт 3.2–3.4) представлен график зависимости потерь распространения от расстояния.

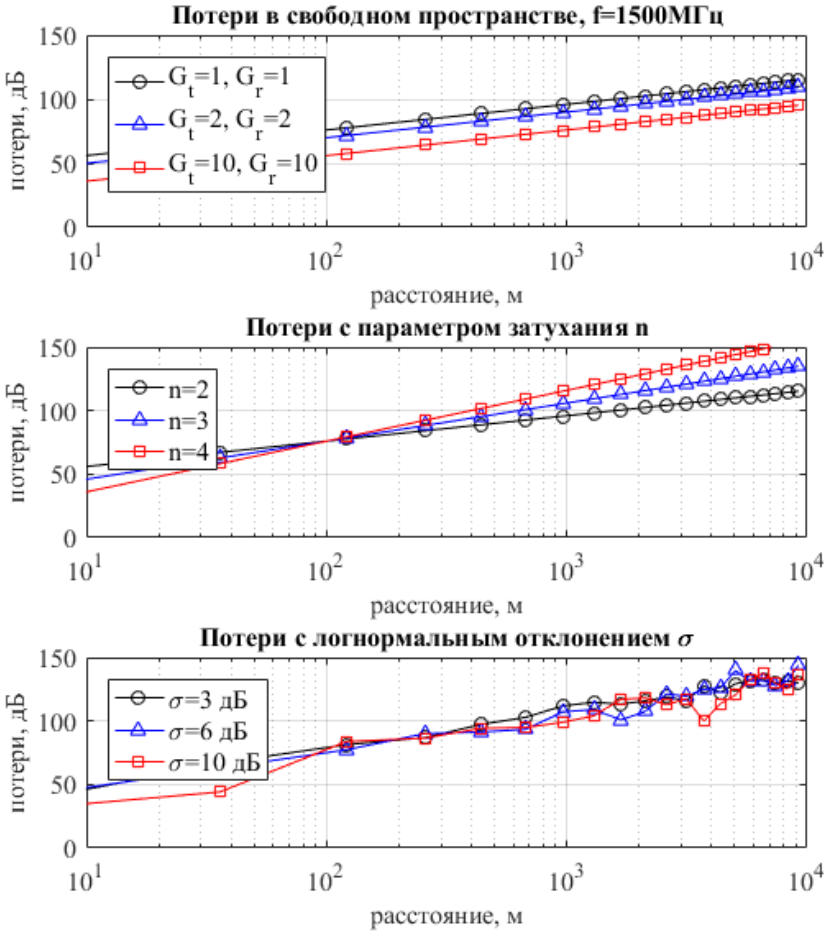


Рис. 3.2. Зависимость потерь распространения от расстояния

Скрипт-функция 3.2. Потери в свободном пространстве

```
function L=L_free(f,d,Gt,Gr)
% Потери распространения радиоволн в свободном пространстве
% Входные данные
%       f : несущая частота, Гц
%       d : расстояние между передатчиком и приемником, м
%       Gt: коэффициент усиления антенны передатчика
%       Gr: коэффициент усиления антенны приемника
% Выходные данные
%       L : потери распространения, дБ
lamda = 3e8/f;
tmp = lamda./(4*pi*d);
if nargin>2,  tmp = tmp*sqrt(Gt);  end
if nargin>3,  tmp = tmp*sqrt(Gr);  end
L = -20*log10(tmp);
```

Скрипт-функция 3.3. Потери с параметром затухания n и логарифмически нормальным отклонением σ

```
function L=L_logdist_or_norm(f,d,d0,n,sigma)
% Потери распространения радиоволн с параметром затухания n и учетом
% логарифмически нормального отклонения мощности принимаемого сигнала
% Входные данные
%       f     : несущая частота, Гц
%       d     : расстояние между передатчиком и приемником, м
%       d0    : эталонное расстояние, м
%       n     : показатель затухания
%       sigma: СКО мощности принимаемого сигнала, дБ
% Выходные данные
%       L     : потери распространения, дБ
lamda = 3e8/f;
L = -20*log10(lamda/(4*pi*d0)) + 10*n*log10(d/d0);
if nargin>4,  L = L + sigma*randn(size(d));  end
```

Скрипт 3.4. График зависимости потерь распространения

```
% Построение графиков потерь распространения
clear all; clf; clc;
f=1.5e9;              % несущая частота, Гц
d0=100;               % эталонное расстояние, м
sigma=[3, 6, 10];     % СКО мощности принимаемого сигнала, дБ
d=[1:5:100].^2;       % расстояние между передатчиком и приемником, м
Gt=[1, 2, 10];        % коэффициент усиления антенны передатчика
Gr=[1, 2, 10];        % коэффициент усиления антенны приемника
n=[2, 3, 4];          % показатель затухания
for k=1:3
   y_Free(k,:)=    L_free(f,d,Gt(k),Gr(k));
   y_logd(k,:)= L_logdist_or_norm(f,d,d0,n(k));
   y_logn(k,:)= L_logdist_or_norm(f,d,d0,n(2),sigma(k));
end
subplot(311);
semilogx(d,y_Free(1,:),'k-o',d,y_Free(2,:),'b-^',d,y_Free(3,:),'r-s')
grid on; axis([10 10000 0 150]);
title(['Потери в свободном пространстве, f=',num2str(f/1e6),'МГц'])
xlabel('расстояние, м'); ylabel('потери, дБ');
legend('G_{t}=1, G_{r}=1','G_{t}=2, G_{r}=2',...
    'G_{t}=10, G_{r}=10','Location','northwest');
subplot(312);
semilogx(d,y_logd(1,:),'k-o',d,y_logd(2,:),'b-^',d,y_logd(3,:),'r-s')
grid on; axis([10 10000 0 150]);
```

```
title('Потери с параметром затухания n');
xlabel('расстояние, м'); ylabel('потери, дБ');
legend(['n=', num2str(n(1))],['n=', num2str(n(2))],...
    ['n=', num2str(n(3))],'Location','northwest');
subplot(313);
semilogx(d,y_logn(1,:),'k-o',d,y_logn(2,:),'b-^',d,y_logn(3,:),'r-s')
grid on; axis([10 10000 0 150]);
title('Потери с логнормальным отклонением \sigma');
xlabel('расстояние, м'); ylabel('потери, дБ');
legend(['\sigma=', num2str(sigma(1)), ' дБ'],...
    ['\sigma=',num2str(sigma(2)), ' дБ'],...
    ['\sigma=', num2str(sigma(3)), ' дБ'],'Location','northwest');
```

Контрольные вопросы

1. Чем радиоканал отличается от проводных линий связи с точки зрения прогноза условий РРВ?
2. Почему не существует единой общепринятой модели расчета ЭМП в сетях радиодоступа в городских условиях?
3. Поясните термин «эмпирическая модель» применительно к прогнозу условий РРВ в СРД.
4. Поясните суть отражения радиоволн.
5. Поясните суть дифракции радиоволн.
6. Поясните суть рассеяния радиоволн.
7. Сформулируйте суть классического подхода к расчету распределения ЭМП.
8. Почему классический подход к расчету распределения ЭМП слабо применим на практике СРД?
9. Поясните суть прогноза РРВ по модели «большого расстояния».
10. Чем отличается РРВ в городских условиях от РРВ в свободном пространстве?
11. Каково происхождение флуктуации напряженности поля в СРД?
12. Что такое глубина медленных замираний?
13. Что является результатом расчета напряженности поля по модели «большого расстояния»?
14. Почему расчет напряженности поля по модели «большого расстояния» носит вероятностный характер?
15. Что отражает модель «малого расстояния» применительно к прогнозу ЭМП?
16. Почему существенные изменения величины ЭМП могут происходить на очень малых расстояниях и в короткие промежутки времени?
17. Что такое быстрые замирания в СРД?
18. Поясните назначение запаса на быстрые замирания в СРД?
19. Приведите и поясните формулу оценки потерь РРВ в свободном пространстве.
20. Приведите и поясните формулу оценки мощности принятого сигнала на заданном расстоянии от передатчика при потерях РРВ в свободном пространстве.
21. Приведите и поясните понятие коэффициента затухания радиоволн. Почему и насколько он отличается от аналогичного коэффициента в свободном пространстве?
22. Приведите и поясните выражение для средней мощности принятого сигнала в логарифмическом виде (в дБ).
23. Приведите и поясните формулу вероятности случайного отклонения мощности принимаемого сигнала в произвольном месте от среднего значения.
24. Поясните физический смысл понятия «среднеквадратическое отклонение» (СКО) мощности принимаемого сигнала.

25. Почему величины случайного отклонения мощности принимаемого сигнала не зависят от расстояния между передатчиком и приемником?

26. В каких пределах находится величина отклонения реальной мощности от средней в типовых городских условиях в диапазоне УКВ?

27. Приведите и поясните график зависимости мощности принимаемого сигнала от нормированного расстояния в линейном масштабе.

28. Приведите и поясните график зависимости мощности принимаемого сигнала от нормированного расстояния в логарифмическом масштабе.

29. Приведите и поясните уравнение мощности принимаемого сигнала в зависимости от расстояния, мощности передатчика и условий распространения радиоволн.

30. Какие методики используются для приближенного расчета ЭМП?

3.2. Лабораторная работа. Эффект захвата в сетях радиодоступа

Цель работы: построить имитационную модель для изучения эффекта захвата в сетях радиодоступа ALOHA/SALOHA.

3.2.1. Эффект захвата в сетях радиодоступа

В сетях радиодоступа существует так называемый эффект захвата, при котором успешная передача возможна даже в том случае, когда несколько радиостанций ведут передачу одновременно.

Рассмотрим сценарий на рис. 3.3, когда несколько АС распределены по территории покрытия БС с радиусом *R*, а БС находится в начале локальной системы координат.

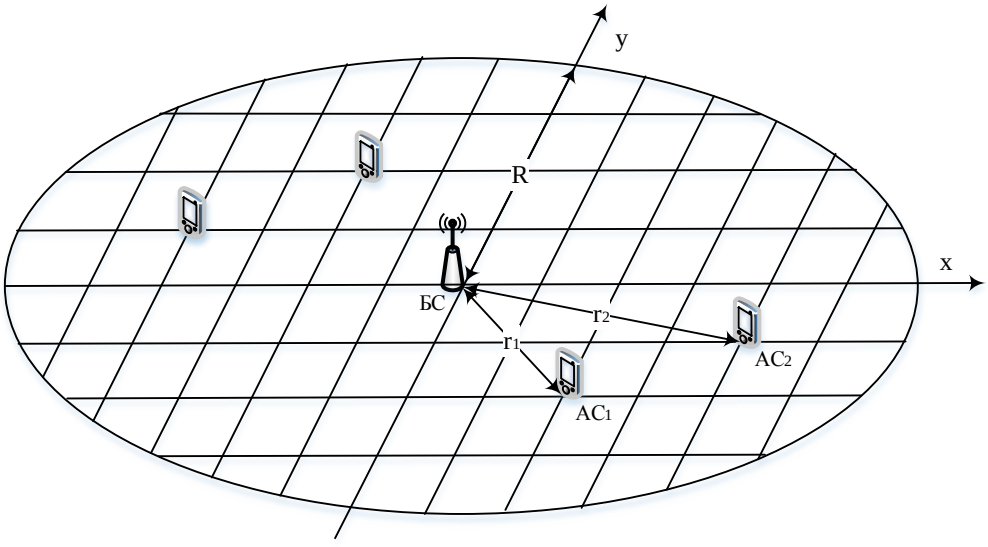


Рис. 3.3. Распределение АС по территории покрытия БС радиусом *R*

Функция для распределения АС по территории покрытия БС радиусом *r* представлена в скрипте 3.5.

Скрипт 3.5. Функция для распределения АС на территории покрытия БС радиусом r

```
% Расположение АС внутри круга радиусом r
% Input arguments
%   r - радиус зоны покрытия БС (БС в начале координат)
%   n - число АС
% Output argument
%   posxy : (x,y)
function [posxy] = position(r, n)
ms = 4 * r; % максимальное число позиций
ms = ms + 4 * sum(fix(sqrt(r^2-[1:r-1].^2)));
if n > ms
    error('n превышает число позиций');
end
posxy = zeros(n,2);
for ii=1:n
    while 1
        xx  = round(r*rand) * sign(sin(2*pi*rand));
        yy  = round(r*rand) * sign(cos(2*pi*rand));
        if xx^2+yy^2 <= r^2 & (xx~=0 | yy~=0)
            if length(find(posxy(:,1)==xx & posxy(:,2)==yy)) == 0
                break
            end
        end
    end
    posxy(ii,[1 2]) = [xx yy];
end
```

Успешный радиоприем определяется затуханием радиосигнала на трассе от АС к БС и оказывается возможным даже при одновременной передаче несколькими АС за счет того, что сигнал от одной АС при приеме на БС оказался по мощности сильнее суммы сигналов других АС, ведущих одновременную передачу.

Введем понятие показателя потерь распространения радиоволн n. Пусть две АС находятся на расстояниях r_1 и r_2 ($r_2 > r_1$) от БС (рис. 3.3). Тогда при одновременной передаче (в сети без затухания возникла бы коллизия) БС успешно воспримет сигнал от $АС_2$ как шум на фоне сигнала от $АС_1$, если отношение этих сигналов больше некоторой пороговой величины SIR [дБ] (Signal to Interference Ratio):

$$10\lg\left(\frac{r_2}{r_1}\right)^n > SIR. \qquad (3.10)$$

Таким образом, $АС_i$ с радиус-вектором r_i будет конкурировать только с теми АС, которые расположены не дальше, чем на расстоянии $r_i \cdot 10^{SIR/10 \cdot n}$ от БС. Именно эти станции определяют вероятность коллизии и вероятность передачи для $АС_i$.

Рассмотрим структуру скрипта для оценки влияния эффекта захвата в радиосетях ALOHA/SALOHA (скрипт 3.5).

Инициализация параметров радиосети и РРВ:

```
% параметры радиосети и РРВ
n=3;                  % параметр потреь РРВ
sigma=6;              % параметр СКО медленных замираний, дБ
r=100;                % радиус радиосети, м
bxy=[0,0];            % координаты БС, м
sir = 10;             % требуемое отношение сигнал/помеха, дБ
prx = 20;             % мощность сигнала от АС на расстоянии r от БС, дБ
capture = 1;          % учитывать эффект захвата (capture = 1), или нет (=0)
mxy=position(r,mnum);% распределение АС по территории радиосети
randn('state',sum(100*clock));
mrnd=randn(1,mnum);  % случайный параметр медленных замираний
ptx = 10^(prx/10)*r^n;    % оценка мощности передатчика АС
```

Требуемое отношение сигнал/помеха `sir` определяет пороговое значение, при котором возможен успешный прием на БС, когда несколько АС ведут одновременную передачу. Параметр `prx` определяет одновременно мощность передачи АС и потери на трассе РРВ от АС к БС и представляет собой отношение сигнал/шум на БС при условии, что АС ведет передачу на границе зоны покрытия на удалении R от БС. Используя индикатор `capture` можно включать/отключать эффект захвата: при отключении эффекта захвата, когда `capture = 0`, моделируется проводная сеть, как в пп. 2.3; при включении эффекта захвата, когда `capture = 1`, моделируется радиосеть, учитывающая распределение АС по территории покрытия БС радиусом R, потери и медленные замирания при РРВ от АС к БС, а также вероятность успешного радиоприема при одновременной передаче нескольких АС при выполнении (3.9). Если `prx` как $P_{rx} = P_{tx} / r^n$, тогда мощность передатчика АС `ptx` можно оценить как $P_{tx} = P_{rx} \cdot r^n$.

Далее в цикле по значениям нагрузки G производится передача необходимого для сбора статистики числа пакетов `TOTAL`, обработка текущих состояний АС и инкремент счетчиков производительности и задержки.

Оценка возможности успешного радиоприема с эффектом захвата осуществляется при нахождении АС в состоянии передачи или коллизии:

```
% нахождение станций в состоянии передачи или коллизии
idx=find(Mstate==TRANSMIT | Mstate==COLLISION);
if capture==0   % без эффекта захвата
    if length(idx)>1
        Mstate(idx)=COLLISION;
    end
else            % с учетом эффекта захвата
    if length(idx)>1
        dxy=sqrt(sum((bxy-mxy(idx,:)).^2,2))'; % расст. м/д БС и АС
        pow=ptx*dxy.^-n.*10.^(sigma/10*mrnd(idx));
        [maxp no]=max(pow);
        if Mstate(idx(no))==TRANSMIT
```

```
                if length(idx)==1
                    cn=10*log10(maxp);
                else
                    cn=10*log10(maxp/(sum(pow)-maxp+1));
                end
                Mstate(idx)=COLLISION;
                if cn >= sir % мощность принятого сигнала выше порога
                    Mstate(idx(no))==TRANSMIT;
                end
            else
                Mstate(idx)=COLLISION;
            end
        end
    end
```

Успешно принятыми на БС могут пакеты тех АС, у которых отношение мощности принятого сигнала к помехам окажется выше требуемого порогового отношения SIR. Помехи при этом создаются другими АС, которые ведут одновременную передачу. Для оценки фактического отношение сигнал/помеха выполняют следующие процедуры. Сначала определяют расстояние `dxy` между БС и активными АС, ведущими передачу в данный момент времени. Затем вычисляют мощности `pow` принятых на БС сигналов от активных АС. Далее из рассчитанных мощностей выбирают сигнал той АС, мощность которой на приемнике БС оказывается максимальной. Именно этот самый сильный сигнал и может быть принят успешно или «захвачен». Когда активных станций оказывается больше одной, вычисляется отношения сигнал/помеха `cn=10*log10(maxp/(sum(pow)-maxp))`, в котором помехи образуются сигналами других активных АС, ведущими передачу в данный момент времени. Если вычисленное отношение сигнал/помеха выше порога, `cn >= sir`, передача считается успешной, в противном случае фиксируется коллизия.

3.2.2. Моделирование эффекта захвата в СРД ALOHA/SALOHA

Ниже представлен скрипт для оценки влияния эффекта захвата в радиосетях ALOHA/SALOHA (скрипт 3.6).

Скрипт 3.6. Оценка влияния эффекта захвата в радиосетях ALOHA/SALOHA

```
clear all; clc;
% состояния
STANDBY=0;          % режим ожидания
TRANSMIT=1;         % режим передачи
COLLISION=2;        % состояние коллизии

% параметры протокола
Brate=0.25e6;       % битовая скорость передачи, бит/с
Plen=500;           % размер пакета, бит
Ttime=Plen/Brate;   % время передачи пакета
```

```
Dtime=0.01;          % задержка распространения кадра (% от Ttime)
G=[0.1:0.1:4];       % траффик/нагрузка (кадров за время передачи кадра)
TOTAL=1000;          % число пакетов для моделирования
mnum=100;            % число станций
protocol = 'saloha';% paloha

% параметры радиосети и РРВ
n=3;                  % параметр потреь РРВ
sigma=6;              % параметр СКО медленных замираний, дБ
r=100;                % радиус радиосети, м
bxy=[0,0];            % координаты БС, м
sir = 10;             % требуемое отношение сигнал/помеха, дБ
prx = 20;             % мощность сигнала от АС на расстоянии r от БС, дБ
capture = 1;          % учитывать эффект захвата (capture = 1), или нет (=0)
mxy=position(r,mnum);% распределение АС по территории радиосети
randn('state',sum(100*clock));
mrnd=randn(1,mnum);  % случайный параметр медленных замираний
ptx = 10^(prx/10)*r^n;     % оценка мощности передатчика АС

for indx=1:length(G)  % нагрузка

    Tint=-Ttime/log(1-G(indx)/mnum);% интервал поступления новых пакетов, с
    Rint=Tint;       % интервал поступления пакетов повторной передачи, с
    Spnum=0;         % число успешно переданных пакетов, шт
    Tplen=0;         % объем данных предложенных к передаче пакетов, бит
    Splen=0;         % объем данных успешно переданных пакетов, бит
    Wtime=0;         % задержка передачи пакетов, с

    % Инициализация состояний и переменных
    mgtime=-Tint*log(1-rand(1,mnum)); % вектор времен генерирования кадров
    mtime=mgtime;                     % время текущего события  (с)
    if protocol=='saloha'
        mtime=(fix(mgtime));
    end
    Mstate=zeros(1,mnum);             % вектор состояний станций
    Mplen(1:mnum)=Plen;               % вектор длина кадра (бит)
    now_time=min(mtime);              % текущее время (с)
    slot=Plen/Brate;                  % длительность слота, с (для saloha)

    while (Spnum<TOTAL)

        % нахождение станций, начинающих передачу
        idx=find(mtime==now_time & Mstate==TRANSMIT);
        if length(idx)>0
            Spnum=Spnum+1;                      % инкремент Spnum
            Splen=Splen+Mplen(idx);             % инкремент Splen
            Wtime=Wtime+now_time-mgtime(idx);   % инкремент Wtime
            Mstate(idx)=STANDBY;                % переход в режим ожидания
            % время генерации нового пакета для станций с индексом idx
            mgtime(idx)=now_time-Tint*log(1-rand);
            % инициализация времени смены состояния станций с индексом idx
            mtime(idx)=mgtime(idx);
            if protocol=='saloha'
                mtime(idx)=(fix(mgtime(idx)/slot)+1)*slot;
            end
        end
```

```
    % нахождение станций в состоянии коллизии
    idx=find(mtime==now_time & Mstate==COLLISION);
    if length(idx)>0
        Mstate(idx)=STANDBY;                % переход в режим ожидания
        % время повторной передачи пакета станциями
        mtime(idx)=now_time-Rint*log(1-rand(1,length(idx)));
        if protocol=='saloha'
            mtime(idx)=(fix(mtime(idx)/slot)+1)*slot;
        end
    end

    % нахождение станций в режиме передачи
    idx=find(mtime==now_time);
    if length(idx)>0
        Mstate(idx)=TRANSMIT;   % переход в режим передачи
        % время завершения передачи пакета станциями
        mtime(idx)=now_time+Mplen(idx)/Brate;
        if protocol=='saloha'
            mtime(idx)=round(mtime(idx)/slot)*slot;
        end
        Tplen=Tplen+sum(Mplen(idx));        % инкремент Tplen
    end

    % нахождение станций в состоянии передачи или коллизии
    idx=find(Mstate==TRANSMIT | Mstate==COLLISION);

    if capture==0   % без эффекта захвата
        if length(idx)>1
            Mstate(idx)=COLLISION;
        end
    else            % с учетом эффекта захвата
        if length(idx)>1
            dxy=sqrt(sum((bxy-mxy(idx,:)).^2,2))'; % расст. м/д БС и АС
            pow=ptx*dxy.^-n.*10.^(sigma/10*mrnd(idx));
            [maxp no]=max(pow);
            if Mstate(idx(no))==TRANSMIT
                if length(idx)==1
                    cn=10*log10(maxp);
                else
                    cn=10*log10(maxp/(sum(pow)-maxp+1));
                end
                Mstate(idx)=COLLISION;
                if cn >= sir % мощность принятого сигнала выше порога
                    Mstate(idx(no))==TRANSMIT;
                end
            else
                Mstate(idx)=COLLISION;
            end
        end
    end
    now_time=min(mtime);
end
% вектор нормированной предложенной нагрузки
Traffic(indx)=Tplen/Brate/now_time;
% вектор нормированной пропускной способности
S(indx)=Splen/Brate/now_time;
% вектор нормированной средней задержки передачи
Delay(indx)=Wtime/TOTAL*Brate/Plen;
```

```
end
subplot(2,1,1); plot(Traffic,S,'ro'); grid on;
xlabel('G'); ylabel('S');
title(strcat({'Производительность протокола'}, {' '}, protocol));
subplot(2,1,2); plot(Traffic,Delay,'-ko'); grid on;
xlabel('G'); ylabel('D');
```

Контрольные вопросы

1. В чем суть эффекта захвата в сетях радиодоступа?
2. Почему в сетях радиодоступа пакет от данной АС может быть принят успешно, даже если одновременно передачу ведут другие АС?
3. Сформулируйте последовательность процедур для вычисления отношения сигнал/помеха для данной АС.
4. Проведите имитационное моделирование и визуализируйте территориальное распределение АС.
5. Проведите имитационное моделирование и проследите последовательность процедур для вычисления отношения сигнал/помеха для АС с максимальным сигналом.

СПИСОК ИСТОЧНИКОВ

1. Сети радиодоступа: основная профессиональная образовательная программа: 11.03.02 Инфокоммуникационные технологии и системы связи для бакалавров / Федеральное агентство связи, Федеральное государственное образовательное бюджетное учреждение высшего профессионального образования «Санкт-Петербургский государственный университет телекоммуникаций им. проф. М. А. Бонч-Бруевича», каф. РС и В ; сост. Г. А. Фокин. – СПб. : СПбГУТ, 2015. – 15 с.

2. *Олифер, В. Г.* Компьютерные сети. Принципы, технологии, протоколы : учебник для вузов / В. Г. Олифер, Н. А. Олифер ; рец. Ю. А. Григорьев, Б. Ф. Прижуков. – 5-е изд. – СПб. : Питер, 2016. – 992 с.

3. *Скляр, Б.* Цифровая связь. Теоретические основы и практическое применение / Б. Скляр ; пер. с англ. Е. Г. Грозы [и др.] ; под ред. А. В. Назаренко; науч. консультант Е. В. Гусева. – 2-е изд., испр. – М. : Вильямс, 2004. – 1104 с.

4. *Галкин, В. А.* Цифровая мобильная радиосвязь : учеб. пособие / В. *А.* Галкин. – Горячая линия-Телеком, 2012.

5. *Harada, H.* Simulation and software radio for mobile communications / H. Harada, R. Prasad. – Artech House, 2002.

Фокин Григорий Алексеевич

СЕТИ РАДИОДОСТУПА

Учебное пособие

Редактор *Л. К. Паршина*
Компьютерная верстка *Н. А. Ефремовой*

План издания 2019 г., п. 2

Подписано к печати 18.06.2019
Объем 19,75 печ. л. Тираж 26 экз. Заказ 950

Редакционно-издательский отдел СПбГУТ
193232 СПб., пр. Большевиков, 22

Отпечатано в СПбГУТ

314