Учреждение образования «Витебский государственный
университет им. П.М. Машерова»
Кафедра инженерной физики

СИСТЕМЫ СВЯЗИ И СЕТИ ПЕРЕДАЧИ ИНФОРМАЦИИ

КУРС ЛЕКЦИЙ

Автор составитель – к.т.н. Краснобаев Е.А.

Витебск
УО «ВГУ им. П.М. Машерова»
2012

СОДЕРЖАНИЕ

Лекция 1. Элементы теории передачи информации

1.1. Информация, сообщение, сигнал

Понятие «информация» имеет много различных аспектов, и в связи с этим существует и несколько различных подходов к ее определению и оценкам (количественным, качественным и др.). Например, в философии принято считать, что информация есть отражение реального мира.

Исходя из специфики задач теории информации, академик А.А. Харкевич определил информацию как совокупность сведений о каком-либо событии, явлении, предмете и т.д., являющихся объектом хранения, передачи и преобразования. Для выполнения указанных действий используют условные символы (знаки) -буквы, математические знаки, рисунки, жесты, слова и.т.п., позволяющие выразить информацию в необходимой форме. Совокупность знаков, которые используются для хранения, передачи и преобразования информации, называют *сообщением*. Так, при телеграфной передаче информация представляется в виде букв и цифр. Соответственно сообщением является текст телеграммы, представляющий последовательность этих знаков. В телефонных системах сообщением является речь - определенный набор звуков, отображающих не только содержание, но и интонацию, тембр, ритм и иные свойства речи. В различных технических системах информация представляется в двоичной форме, т.е. только двумя условными символами, например 1 и 0. Соответственно сообщением служит последовательность конечного числа двоичных символов. При передаче движущихся изображений в телевизионных системах сообщение представляет собой изменение во времени яркости элементов изображения.

Различают дискретные и непрерывные сообщения. *Дискретные сообщения* формируются в результате последовательной выдачи источником сообщений отдельных знаков. Множество различных знаков называют *алфавитом источника сообщений,* а их количество - *объемом алфавита. Непрерывные сообщения* не разделимы на элементы. Они описываются функциями времени, принимающими непрерывное множество значений. Типичными примерами непрерывных сообщений могут служить речь, телевизионное изображение.

Передача сообщений (а следовательно, и информации) на расстояние осуществляется с помощью какого-либо материального носителя (бумаги, магнитной ленты и т. п.) или физического процесса (звуковых или электромагнитных волн, тока и т. п.). Физический процесс, посредством которого сообщение передается на расстояние, называется *сигналом.*

В качестве сигнала можно использовать любой физический процесс, изменяющийся в соответствии с переносимым сообщением. В современных системах управления и связи чаще всего используют электрические сигналы. Физической величиной, определяющей такой сигнал, являются ток или

напряжение. Сигналы формируются путем изменения тех или иных параметров физического носителя по закону передаваемых сообщений. Процесс изменения параметров носителя принято называть *модуляцией.*

Сообщения могут быть функциями времени, например речь при передаче телефонных разговоров, температура или давление воздуха при передаче телеметрических данных, спектакль при передаче по телевидению и т. п. В других случаях сообщение не является функцией времени (например, текст телеграммы, неподвижное изображение и т. д.). Сигнал является функцией времени, даже если сообщение таковым не является.

По своей природе сигналы могут быть электрическими, световыми, звуковыми и т.п. В системах телекоммуникаций (ТК) используются электрические и световые сигналы. Электрические сигналы получили наиболее широкое применение в системах ТК. Это связано с тем, что обработка электрических сигналов осуществляются гораздо проще с технической точки зрения, чем обработка сигналов другой физической природы. Поэтому при передаче сообщения неэлектрической природы предварительно преобразуются в электрические колебания с помощью преобразователей: микрофонов, передающих телевизионных трубок, датчиков температуры, давления т. п. Эти электрические колебания обычно называют *первичными сигналами.* Наряду с электрическими сигналами в системах ТК могут использоваться световые сигналы, передача которых на расстояние по волоконно-оптическим линиям связи может оказаться предпочтительнее, чем передача электрических сигналов.

Различают следующие виды сигналов:

- непрерывные по уровню и по времени (рис. 1.1, а);
- непрерывные по уровню и дискретные по времени (рис. 1.1, б);
- дискретные (квантованные) по уровню и непрерывные по времени (рис. 1.1, в);
- дискретные по уровню и по времени (рис. 1.1, г).

Сигналы первого вида (рис. 1.1, а), называемые непрерывными, задаются на конечном или бесконечном временном интервале и могут принимать любые значения в некотором диапазоне. Примером таких сигналов, называемых *аналоговыми*, являются сигналы на выходах микрофона, датчиков температуры, давления, положения и т. п.

Сигналы второго вида (рис. 1.1, б) задаются в определенные дискретные моменты времени и могут принимать любые значения из некоторого диапазона. Их можно получить из непрерывных сигналов путем взятия отсчетов в определенные моменты. Это преобразование называется *дискретизацией* во времени. *Шаг дискретизации* $T_Д$ (промежуток времени между двумя соседними отсчетами) может быть как постоянным, так и переменным. Обычно его значение выбирают, исходя из допустимой погрешности при восстановлении непрерывного сигнала по конечному числу его отсчетов. Устройство, осуществляющее формирование дискретных отсчетов сигналов, называется *дискретизатором.*

Сигналы третьего вида (рис. 1.1, в), называемые *квантованными по уровню,* задаются на некотором временном интервале и характеризуются тем, что принимают только вполне определенные дискретные значения. Их можно получить из непрерывных сигналов, применяя к ним операцию квантования по уровню. В результате этой операции непрерывный сигнал заменяется ступенчатой функцией. *Шаг квантования* Δa (расстояние между двумя соседними разрешенными уровнями) может быть как постоянным, так и переменным. Его обычно выбирают из условия обеспечения требуемой точности восстановления непрерывного сигнала из квантованного. Устройство, которое выполняет указанную операцию, носит название *квантователя*.

Сигналы четвертого вида (рис. 1.1, г), называемые *дискретными,* задаются в определенные дискретные моменты и принимают определенные дискретные значения. Их можно получить, например, из непрерывных сигналов, осуществляя операции дискретизации по времени и квантования по уровню. Такие сигналы легко представить в цифровой форме, т. е. в виде чисел с конечным числом разрядов. По этой причине их часто называют *цифровыми*. Аналогичная классификация возможна и для сообщений. Сообщения, подлежащие передаче, являются или случайной величиной, или случайной функцией. Детерминированные (заранее известные) сообщения не содержат информации, и нет смысла их передавать. Соответственно сигнал также следует рассматривать как случайный процесс. Детерминированные сигналы не несут информацию. В технике связи они используются для изучения свойств различных радиотехнических цепей

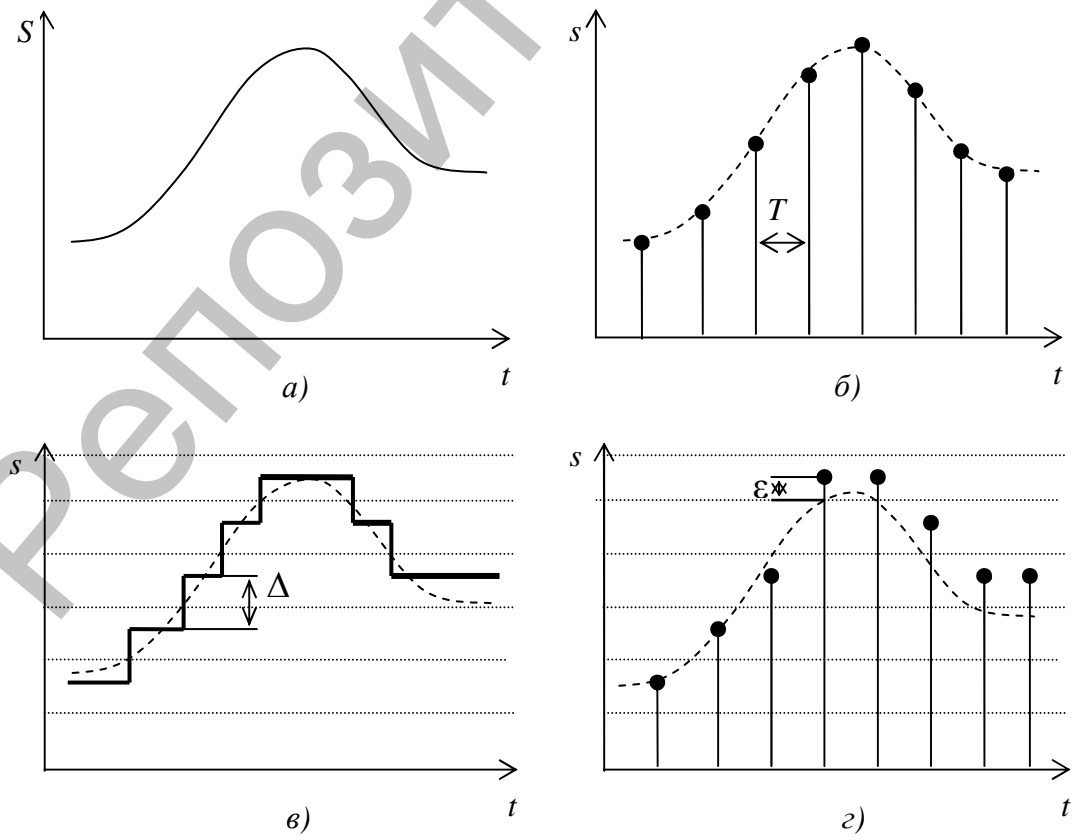


Рис. 1.1

1.2. Преобразование непрерывных сообщений в цифровую форму

Для сообщения с ограниченным спектром дискретизация осуществляется на основе теоремы Котельникова, в соответствии с которой любую непрерывную функцию со спектром $0 \le F \le F_{\max}$ можно однозначно определить последовательностью ее мгновенных значений, взятых через интервалы:

$$T_Д \le \frac{1}{2F_{\max}}. \qquad (1.1)$$

Восстановление непрерывной функции производится в соответствии с выражением:

$$\tilde{U}(t) = \sum_{i=-\infty}^{\infty} U(iT_Д)\frac{Sin2\pi(t-iT_Д)}{2\pi F_{\max}(t-iT_Д)}, \qquad (1.2)$$

называемым рядом Котельникова.

Условие ограничения спектра может выполняться не для всех сигналов. Но на практике вследствие ограниченности полосы пропускания канала, спектр сигнала можно считать ограниченным некоторой частотой $F_{\max}$. Обычно она определяется на основе частотного критерия. Спектр сигнала ограничивается областью частот от 0 до $F_{\max}$, в которой сосредоточена большая часть энергии сигнала (80-95%). Такое ограничение, естественно, приводит к некоторому искажению сигнала.

Восстановление непрерывного сигнала по дискретным отсчетам также сопровождается погрешностью. Основные причины:

1) Вместо δ-импульсов при формировании отсчетов используются импульсы конечной длительности.

2) Ограничение спектра сигнала (если оно есть).

3) Пренебрежение вкладом бесконечного числа функций отсчетов, соответствующих выборкам за пределами интервала T, так как обычно восстановление непрерывного сигнала проводится по конечному числу членов ряда Котельникова.

4) Для восстановления непрерывной функции последовательность дискретных значений необходимо подать на вход так называемого «идеального» фильтра нижних частот (ФНЧ). Поскольку «идеальный» ФНЧ является абстракцией и на практике используется реальный ФНЧ с характеристиками, отличающимися от характеристик «идеального» ФНЧ, то это также является причиной искажений при восстановлении непрерывного сигнала по дискретным отсчетам.

Таким образом, следует иметь в виду, что теорема Котельникова выражает предельные соотношения для идеализированных условий (ограниченность спектра и бесконечное время наблюдения). Однако в большинстве практических случаев ее применение для реальных сигналов

позволяет получить сравнительно небольшую погрешность, приемлемую в инженерной практике.

1.3. Обобщенная структурная схема системы связи. Канал связи.

Системой связи (телекоммуникационной системой) называют совокупность технических средств, предназначенных для передачи информации, включая источник информации и получателя информации. Структурная схема простейшей системы связи показана на рис. 1.2. Рассмотрим назначение основных ее элементов.

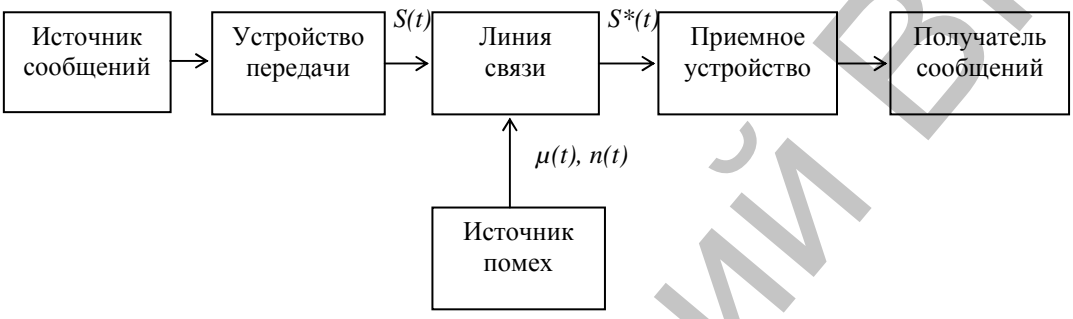


Рис. 1.2

Источником сообщений может быть человек или различного рода устройства (датчики, ЭВМ и т.п.). Источник сообщений осуществляет выбор сообщений из ансамбля сообщений. Если сообщение на выходе источника имеет неэлектрическую природу, то для его передачи в системе связи оно преобразуется в первичный электрический сигнал. Указанная операция производится в первичном преобразователе. В телефонии, например, в качестве первичного преобразователя применяется микрофон, превращающий перепады звукового давления в пропорционально изменяющийся электрический ток. В телевидении функции первичного преобразователя выполняет телевизионная камера.

Первичные сигналы обычно являются низкочастотными и не предназначены для передачи на расстояние. Для передачи на большие расстояния используют специальные электромагнитные колебания высокой частоты, называемые переносчиками, которые могут эффективно распространяться по *линии связи*. В *передающем устройстве* первичный сигнал превращается во вторичный (высокочастотный) сигнал $S(t)$. В качестве переносчика могут использоваться электромагнитные колебания, имеющие гармоническую $u(t) = A_0 \cdot \sin(\omega_0 t - \varphi_0)$ или импульсную форму.

Сами переносчики не содержат информации о передаваемом сообщении. Для того, чтобы заложить в них эту информацию, применяют операцию модуляции, которая заключается в изменении одного или нескольких параметров переносчика по закону передаваемого сообщения. Например, в гармоническом переносчике можно изменять амплитуду,

частоту или фазу колебания. При этом возможны три вида модуляции: амплитудная (АМ), фазовая (ФМ) и частотная (ЧМ), когда модулированные параметры могут быть представлены в следующем виде

$$A(t) = A_0 + \Delta A \cdot a(t),$$

$$\omega(t) = \omega_0 + \Delta\omega \cdot a(t),$$

$$\varphi(t) = \varphi_0 + \Delta\varphi \cdot a(t),$$

где $a(t)$ - закон изменения передаваемого сообщения (полагаем, что $-1 \le a(t) \le 1$),

ΔA, $\Delta\omega$, $\Delta\varphi$ - максимальные изменения соответственно амплитуды, частоты и фазы.

В импульсном переносчике можно изменять амплитуду, временное положение импульсов относительно выбранного начала отсчета, их длительность, период следования, параметры формы импульса и т.п.

Устройство, осуществляющее изменение одного или нескольких параметров переносчика, называется модулятором.

Линия связи – это среда, используемая для передачи сигналов. Линии связи могут быть проводные и беспроводные, например, радиолинии. В радиолиниях средой служит часть пространства, в котором распространяются электромагнитные волны от передатчика к приемнику.

Источник помех. В реальной системе сигнал передается при наличии помех, под которыми понимают любые случайные воздействия, накладывающиеся на сигнал и затрудняющие его прием. Поэтому сигнал $S^*(t)$, поступающий на вход приемного устройства (рис. 1.2), в общем случае отличается от сигнала $S(t)$, который был на выходе радиопередающего устройства. В некоторых случаях действие помех $n(t)$ можно описать соотношением

$$S^*(t) = S(t) + n(t),$$

где $n(t)$ не зависит от $S(t)$.

Помеха, удовлетворяющая такому условию, называется аддитивной. Если соотношение, связывающее сигналы на выходе радиопередатчика и приемника, имеет вид

$$S^*(t) = \mu(t) \cdot S(t),$$

где $\mu(t)$ - некоторая случайная функция, то помеха называется мультипликативной.

В реальных линиях связи действуют как аддитивная, так и мультипликативная помехи, поэтому

$$S^*(t) = \mu(t) \cdot S(t) + n(t).$$

В зависимости от характера изменения во времени различают флуктуационные, импульсные (сосредоточенные во времени) и узкополосные (сосредоточенные по частоте) помехи. Флуктуационная

помеха порождается различного рода флуктуациями, т.е. случайного рода отклонениями физических величин от их средних значений. Флуктуационная помеха может быть обусловлена дискретной природой носителей заряда в электронных приборах, тепловым движением носителей заряда и некоторыми другими причинами.

Импульсная помеха представляет собой случайную последовательность импульсов, следующих столь редко, что реакция приемника на текущий импульс успевает затухнуть к моменту появления очередного импульса. Типичным примером такой помехи могут служить атмосферная помеха, различного года индустриальные помехи и т.д.

Узкополосная помеха, как следует из ее названия, сосредоточена в сравнительно узкой полосе частот, существенно меньшей по сравнению с полосой частот сигнала.

Приемное устройство обрабатывает принятое колебание $S^*(t) = \mu(t) \cdot S(t) + n(t)$ и восстанавливает по нему сообщение (первичный сигнал) $b^*(t)$, которое с некоторой погрешностью отражает переданное сообщение $b(t)$. Другими словами, приемник должен на основе анализа колебания $S^*(t)$ определить, какое из возможных сообщений передавалось.

Получателем сообщения может быть человек, для которого оно предназначено, или различного рода устройства (автомат, ЭВМ, магнитофон и т.п.).

Совокупность технических средств передачи информации, включающая среду распространения и обеспечивающая передачу сигнала от некоторой точки А системы до точки В (рис. 1.3), называется *каналом*. Точки А и В могут быть выбраны произвольно в зависимости от решаемых задач. Так, в одних случаях канал может состоять только из линии связи, в других – из модулятора, линии связи и демодулятора и т.д. Часть системы связи, расположенная до точки А, является источником сигнала для этого канала. Если сигнал, поступающий на вход канала и снимаемый на его выходе, является дискретным по состояниям, то такой канал называется дискретным. Если входные и выходные сигналы канала непрерывные, то канал называется непрерывным. Встречаются также дискретно-непрерывные и непрерывно-дискретные каналы, на вход которых поступают дискретные сигналы, а на выходе снимаются непрерывные, и наоборот.

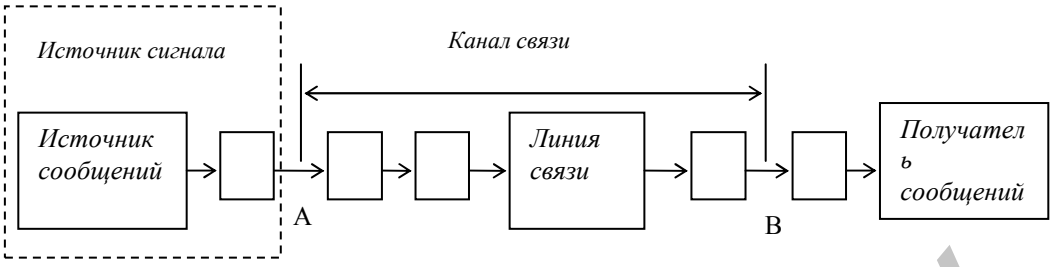


Рис. 1.3

Любая телекоммуникационная система характеризуется рядом показателей. Рассмотрим наиболее существенные из них с точки зрения передачи информации.

Достоверность передачи информации характеризует степень соответствия принятых сообщений переданным. Она зависит от параметров самой системы, степени ее технического совершенства и условий работы. Последние определяются типом и состоянием линий связи, видом и интенсивностью помех и т.д. Для различных телекоммуникационных систем критерии соответствия принятого сигнала переданному могут существенно отличаться. При передаче дискретных сообщений действие помех проявляется в том, что вместо переданного символа принимается другой. В этом случае достоверность передачи сообщений целесообразно характеризовать или вероятностью правильного приема символа $p_{ПР}$, или вероятностью ошибки $p_{ОШ} = 1 - p_{ПР}$.

При передаче непрерывных сообщений отличие принятого сообщения $b^*(t)$ от переданного $b(t)$ носит также непрерывный характер:

Для оценки достоверности передачи сообщений в данном случае обычно используют средний квадрат ошибки

$$\overline{\varepsilon^2} = \overline{[b^*(t) - b(t)]^2}.,$$

или относительный средний квадрат ошибки

$$\delta^2 = \overline{\varepsilon^2} / P_b = P_\varepsilon / P_b,$$

где усреднение производится по всем реализациям сообщений $b(t)$ и их оценкам $b^*(t)$,

$P_b = \frac{1}{T_c}\int_0^{Tc} b^2(t)dt$ - средняя мощность сообщения $b(t), T_c$ - его длительность, P_ε - мощность помехи на выходе приемника.

Возможны и другие показатели достоверности, как, например, показатель максимальной абсолютной ошибки $\varepsilon_{\max} = \max|\varepsilon(t)|$.

Под *помехоустойчивостью* понимают способность системы противостоять вредному действию помех на передачу сообщений. Количественно помехоустойчивость телекоммуникационных систем можно

характеризовать вероятностью ошибки $p_{ОШ}$ при заданном отношении средних мощностей сигнала и помехи в полосе частот занимаемой сигналом, или требуемым отношением средних мощностей сигнала и помехи на входе приемника системы, при котором обеспечивается заданная вероятность ошибки $p_{ОШ}$. Еще одна важная характеристика – скорость передачи информации – будет введена ниже.

1.4. Информационные характеристики источника дискретных сообщений и канала связи

Количественная оценка информации

Для сравнения различных систем телекоммуникаций необходимо ввести количественную меру, позволяющую оценивать объем информации, содержащейся в сообщении.

Строгие методы количественного определения информации были предложены К.Шенноном в 1948 г. и привели к построению теории информации, являющейся основой теории связи, информатики и ряда смежных отраслей науки и техники.

$$U = \begin{pmatrix} u_1; u\ ; ...u_N; \\ P(u_1); P(u_2); ...P(u_N) \end{pmatrix}$$

Рассмотрим основные идеи этой теории применительно к дискретному источнику сообщений, который в каждый момент времени случайным образом может принимать одно из конечного множества возможных состояний. Каждому состоянию источника сообщений ставится в соответствие условное обозначение в виде знака (в частности буквы) из алфавита данного источника u_1, u_2, u_3,..., u_N. Одни состояния выбираются источником сообщений чаще, другие реже. Поэтому наряду с множеством состояний целесообразно задать вероятность их появления:

Или:

$$U = \begin{pmatrix} u_1; u_2; ...; u_N \\ p_1; p_2; ... p_N \end{pmatrix},$$

$$\textit{причем} \sum_{i=1}^{N} p_i = 1.$$

Совокупность состояний и вероятностей их получения называется ансамблем U.

Перед тем как ввести определение количества информации, сформулируем условия, которым должна удовлетворять эта величина:

1) она должна быть аддитивной величиной, т.е. если рассматривать два последовательных события u_i и u_k, происходящих независимо друг от друга, как одно укрупненное, то количество информации в таком событии должно равняться сумме количества информации в каждом из них

$$I(u_i, u_k) = I(u_i) + I(u_k); \qquad (1.3)$$

2) количество информации в сообщении о достоверном событии (p=1) равно нулю (такое сообщение ничего не добавляет к нашим знаниям);

3) данная величина должна быть неотрицательной;

4) количество информации не должно зависеть от качественного содержания сообщения, в частности, от степени его важности для получателя, эмоциональной окраски и т.д.

Итак, для определения количества информации в некотором сообщении u_i из ансамбля U необходимо основываться только на таком параметре, который характеризует в самом общем виде это сообщение. Таким параметром, очевидно, является вероятность p_i появления данного сообщения на выходе источника.

Дальнейшее уточнение искомого определения не составит труда, если принять во внимание первые два из перечисленных выше условий. Пусть u_i и u_k - два независимых события. Вероятность того, что оба этих сообщения появятся на выходе источника одно за другим

$$P(u_i, u_k) = P(u_i) \cdot P(u_k), \qquad (1.4)$$

а количество информации в этих сообщениях должно удовлетворять условию (1.3). Следовательно, необходимо найти функцию, обладающую свойством, что при перемножении двух аргументов значения функции складываются. Единственная такая функция – это логарифмическая функция $I(u) = k \log P(u)$, где k - постоянный коэффициент. Заметим, что при таком определении количества информации выполняется и второе требование - при $P(u) = 1$, $I(u) = 0$. Основание логарифма не имеет принципиального значения и определяет только масштаб функции. Так как информационная техника широко использует элементы, имеющие два устойчивых состояния, то обычно основание логарифма выбирают равным 2. В дальнейшем обозначение $\log$, если основание не оговаривается особо, будет означать двоичный логарифм. Чтобы количество информации $I(u)$ было неотрицательной величиной, выбирают $k = -1$. Поэтому $I(u) = -\log P(u)$.

Если источник передает последовательность зависимых между собой сообщений, то получение предшествующих сообщений может изменить вероятность последующего, а следовательно, и количество информации в нем. Оно должно определяться по условной вероятности передачи данного сообщения u_k при известных предшествовавших сообщениях u_{k-1}, u_{k-2},…:

$$I(u_k | u_{k-1}, u_{k-2}, ...) = -\log P(u_k | u_{k-1}, u_{k-2}, ...). \qquad (1.5)$$

Введенное выше определение характеризует количество информации, содержащееся в одном сообщении из ансамбля U. При этом $I(u)$ является случайной величиной, зависящей от того, какое состояние источника в действительности реализуется. Для характеристики всего ансамбля (или источника) используется математическое ожидание количества информации, называемое энтропией и показывающее, какое количество информации *в среднем* содержится в одном сообщении данного источника

$$H(U) = M\{-\log P(u)\}. \qquad (1.6)$$

Для источника независимых сообщений выражение (1.6) можно представить в виде

$$H(U) = -\sum P(u_i)\log P(u_i),\ i = 1,\ \ldots,\ N. \qquad (1.7)$$

Пример 1.

Рассмотрим случай, когда алфавит состоит из двух знаков, появляющихся на выходе источника сообщений независимо друг от друга. Обозначим $P(u_1) = P$. Соответственно $P(u_2) = 1 - P$. Тогда на основании (1.7) имеем

$$H(U) = -P\log P - (1-P)\log(1-P).$$

Если события u_1 и u_2 являются равновероятными, то $P = 1/2$. Подставив это значение в данное уравнение, получим $H(U) = 1$. В общем виде зависимость $H(U)$ от P показана на рис. 1.4.

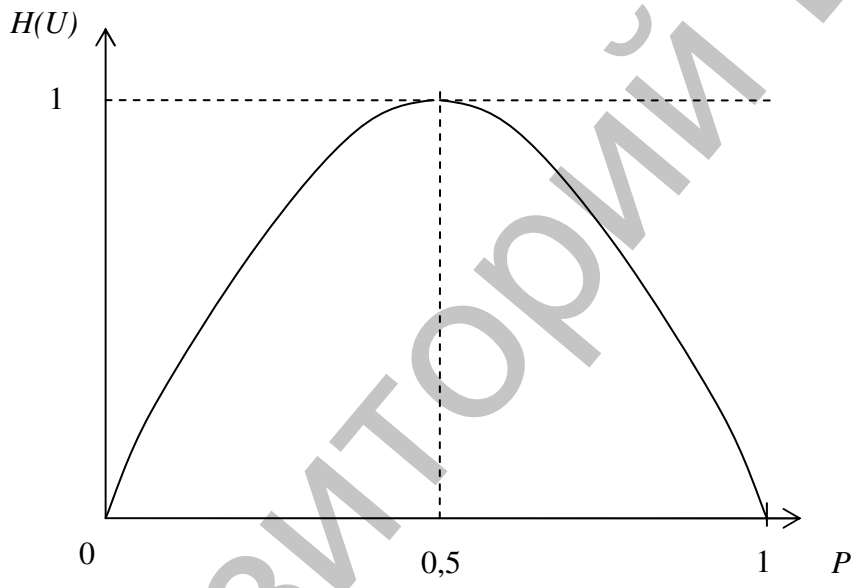


Рис.1.4

Количество информации, содержащееся в среднем в одном символе алфавита, состоящего из двух знаков, которые появляются на выходе источника сообщений независимо друг от друга и равновероятно, получило название "1 бит информации" или просто 1 бит.

Случай, когда алфавит источника сообщений состоит всего из двух знаков, широко распространен на практике. В качестве примера можно привести цифровые системы, использующие алфавит, состоящий из знаков 0 и 1.

Чем больше энтропия источника, тем больше степень неожиданности передаваемых им сообщений в среднем, т.е. тем более неопределенным является ожидаемое сообщение. Поэтому энтропию часто называют мерой неопределенности сообщений. При этом имеется в виду неопределенность, существующая до того, как сообщение передано. После приема сообщения, если оно передано верно, всякая неопределенность устраняется. Это

позволяет трактовать количество информации как меру уменьшения неопределенности.

Свойства энтропии источника дискретных сообщений

Перечислим основные свойства энтропии:

энтропия неотрицательна;

энтропия равна нулю для вырожденного ансамбля, когда одно сообщение передается с вероятностью 1, а остальные имеют нулевую вероятность;

энтропия аддитивна (смысл этого свойства рассмотрен выше);

если ансамбль содержит N различных сообщений, то

$$H(U) \le \log N, \qquad (1.8)$$

причем равенство имеет место только тогда, когда все сообщения передаются равновероятно и независимо ($P(u_i) = 1/N;\ i = 1,\ ...,\ N$).

Для доказательства последнего свойства воспользуемся неравенством

$$\ln w \le (w-1). \qquad (1.9)$$

Рассмотрим разность функций, стоящих в левой и правой частях неравенства (1.8). Путем несложных преобразований ее можно представить в виде:

$$H(U) - \log N = \sum P_i \log(1/P_i) - \sum P_i \log N = \\ = \sum P_i \log(1/NP_i), \qquad (1.10)$$

причем суммирование ведется по $i = 1,\ ...,\ N$.

Чтобы воспользоваться неравенством (1.9), перейдем в последнем выражении к натуральному основанию логарифма:

$$H(U) - \log N = \sum P_i \log(1/NP_i) = \log e \sum P_i \ln(1/NP_i). \qquad (1.11)$$

Используя (1.9), преобразуем выражение (1.11). Получим:

$$H(U) - \log N = \log e \sum P_i \ln(1/NP_i) \le \log e \sum P_i [1/NP_i - 1] = \\ = \log e \left(\sum 1/N - \sum P_i\right) = 0. \qquad (1.12)$$

Знак равенства в (1.8) имеет место тогда, когда в неравенстве (1.9) $w = 1$. Так как в нашем случае $w = 1/NP_i$, то условие $w = 1$ эквивалентно $P_i = 1/N;\ i = 1, ..., N$. То есть, $H(U) = \log N$ в случае, когда все элементы алфавита появляются на выходе источника равновероятно. Свойство (1.8) доказано.

Анализируя свойство (1.8), можно прийти к выводу, что чем больше объем алфавита источника дискретных сообщений, тем большее количество информации содержится в среднем в одном символе этого алфавита (при условии, что все элементы алфавита появляются на выходе источника равновероятно). Казалось бы, рассмотренный выше алфавит {0,1}, который широко применяется на практике, следует безоговорочно заменить алфавитами с *N»2*. Однако следует иметь ввиду, что в реальных условиях различение сигналов, которые могут принимать количество значений *N»2*, осуществить на фоне помех гораздо сложнее, чем в случае, когда сигналы могут принимать только два значения – 0 и 1.

*Пример 2.*В теории информации доказывается, что энтропия источника зависимых сообщений всегда меньше энтропии источника независимых сообщений при том же объеме алфавита и тех же безусловных вероятностях сообщений.

Пусть, например, источник выдает последовательность букв из алфавита объемом N=32. Если буквы выбираются равновероятно и независимо друг от друга, то энтропия источника

$H(U) = \log 32 = 5$ [бит].

Однако смысловое содержание такой последовательности букв вряд ли удовлетворит получателя сообщения. Если буквы передаются не хаотически, а составляют связный текст, например, на русском языке, то они оказываются неравновероятными и, главное, зависимыми (так, после гласных не может появиться «ь»; мала вероятность появления 3 гласных или согласных подряд и т.д.). В качестве примера приведем относительные частоты использования некоторых букв русского алфавита в текстах (в порядке убывания):

«о» - 0,090;
«е», «ё» - 0,072
«а», «и» - 0,062
«н», «т» - 0,053
……
«ц» - 0,004
«щ», «э» - 0,003
«ф» - 0,002.

Как видно из этих данных, различие в частоте появления букв в текстах достигает 45 раз!

Если рассматривать ансамбль текстов русской художественной прозы, то энтропия оказывается меньше 1,5 бит на букву. Еще меньше, около 1 бит на букву, энтропия ансамбля поэтических произведений, так как в них имеются дополнительные вероятностные связи, обусловленные ритмом и рифмами.

Избыточность сообщений

Рассмотрим ансамбль U, состоящий из N различных символов: $u_1, u_2, \ldots, u_N$. Энтропия такого дискретного источника достигает максимального значения $H_{\max}(U) = \log N$, если символы статистически независимы и равновероятны. На практике может оказаться так, что символы, образующие алфавит, нельзя рассматривать как независимые и равновероятные, поэтому для такого источника $H(U) < H_{\max}(U)$. Предположим, на выходе такого источника появилось сообщение, состоящее из n символов. Количество информации, содержащееся в нем

$$I = nH(U). \qquad (1.13)$$

При использовании алфавита с максимальной энтропией для передачи такого же количества информации потребовалось бы меньшее число символов

$$I = n_{\min} H_{\max}(U). \qquad (1.14)$$

Приравнивая (1.13) и (1.14), находим

$$\mathrm{n}_{\min} = \mathrm{n} \cdot \frac{\mathrm{H(U)}}{H_{\max}(U)} = \mu \cdot \mathrm{n}, \qquad (1.15)$$

где $\mu = H(U)/H_{\max}(U) < 1$ – коэффициент, характеризующий допустимую степень сжатия сообщений.

Величина $x = 1 - \mu$ называется избыточностью источника. Последствия от наличия избыточности неоднозначные. С одной стороны, избыточные сообщения требуют дополнительных затрат на передачу (например, увеличивается длительность передачи). С другой стороны, наличие избыточности способствует повышению помехоустойчивости сообщений, подчиняющихся априорно известным условиям (ограничениям), т.к. можно обнаружить и исправить ошибки, приводящие к нарушению этих ограничений.

Для сокращения избыточности на практике применяется *кодирование* источника дискретных сообщений, заключающееся в преобразовании исходного дискретного сообщения по определенному правилу в последовательность кодовых символов, удовлетворяющую требованиям равновероятности и статистической независимости.

Устройство, осуществляющее указанную операцию, называется *кодером* источника. В случае его использования, оно размещается в структурной схеме, приведенной на рис. 1.2, между источником сообщений и передающим устройством. Соответственно, на приемной стороне на выходе приемного устройства необходимо добавить устройство, которое будет осуществлять обратную операцию перед тем, как сообщение поступит получателю. Такое устройство называется *декодером*.

1.5. Скорость передачи информации по дискретному каналу. Пропускная способность

Наряду с введенными в разделе 1.3 такими характеристиками, как достоверность и помехоустойчивость, при оценке эффективности систем связи используется ряд других характеристик, рассмотренных ниже.

Технической скоростью V_T называется число элементарных сигналов (символов), передаваемых по каналу в единицу времени. Она зависит от свойств линии связи и быстродействия аппаратуры канала. С учетом возможных различий в длительностях символов

$V_T = 1/\tau_{ср}$,

$\tau_{ср}$ – средняя длительность символа.

Единицей измерения технической скорости служит Бод. 1Бод – скорость, при которой за одну секунду передается один символ.

Скорость передачи сигналов по дискретному каналу устанавливается с учетом ширины F_K полосы пропускания непрерывного канала. Ширина спектра сигнала $S^*(t)$ на выходе непрерывного канала не может превышать F_K. Учитывая, что длительность сигнала и ширина его спектра связаны обратно пропорциональной зависимостью, получим $(V_T)_{\max} \sim F_K$.

Информационная скорость или скорость передачи информации определяется средним количеством информации относительно заданного сообщения, которое передается по каналу в единицу времени. Она зависит как от характеристик данного канала связи, таких как объем алфавита используемых символов, техническая скорость их передачи, статистические свойства помех в линии, так и от вероятностей поступающих на вход символов и их статистической взаимосвязи.

При известной скорости V_T скорость передачи информации $W(V_T, U)$ относительно некоторого сообщения по каналу задается соотношением

$$W(V_T, U) = V_T \cdot H(U), \qquad (1.16)$$

где $H(U)$ – среднее количество информации, переносимое одним символом.

Для практических применений телекоммуникационных систем важно выяснить до какого предела и каким путем можно повысить скорость передачи информации по конкретному каналу. Предельные возможности канала по передаче сообщений характеризуются его пропускной способностью.

Пропускная способность канала $C_Д$ равна максимальной скорости передачи информации по данному каналу, которой можно достигнуть при самых совершенных способах передачи и приема:

$$C_Д = \max W(V_T, U) = \max\{V_T \cdot H(U)\}. \qquad (1.17)$$

Пропускная способность канала, как и скорость передачи информации измеряется числом двоичных единиц информации в секунду- бит/сек.

Пропускная способность канала является характеристикой его самого и не зависит от статистики сигнала.

Если передаваемые символы независимы, то для дискретного канала без помех соотношение (1.17) с учетом (1.8) может быть представлено в виде

$$C_Д = V_T \cdot \log. \qquad (1.18)$$

Каналы без помех можно рассматривать как некую идеализацию реальных каналов. Если мешающим действием помехи пренебречь нельзя, то можно показать, что для дискретных каналов с помехой пропускная способность будет тем меньше, чем больше энтропия источника помехи.

1.6. Согласование физических характеристик сигнала и канала связи

Канал можно охарактеризовать тремя основными параметрами:

1) Временем T_K, в течение которого он представлен для передачи сигнала.

2) Шириной полосы пропускания F_K.

3) Максимально допустимым превышением сигнала над помехой D_K.

Произведение указанных параметров называется объемом канала:

$V_K = T_K F_K D_K$

Аналогично объему канала вводится понятие объем сигнала:

$V_C = T_C F_C D_C$,

где T_C – длительность сигнала, F_C – ширина его спектра, D_C- превышение сигнала над помехой.

V_C и V_K можно представить в трехмерном пространстве.

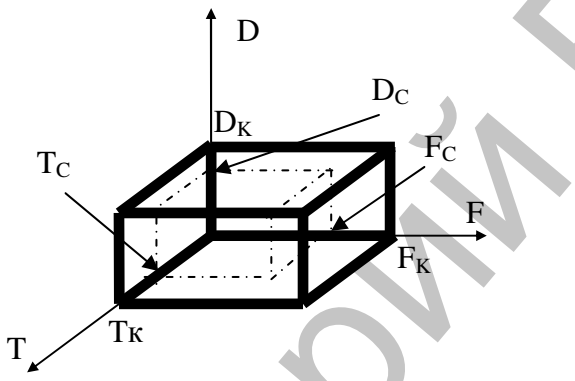


Рис. 1.5

Необходимым условием принципиальной возможности неискаженной передачи сигнала по данному каналу является условие:

$$V_C \le V_K. \qquad (1.20)$$

Для того, чтобы выполнялись достаточные условия передачи $T_C \le T_K$; $F_C \le F_K$; $D_C \le D_K$, могут потребоваться преобразования. Если канал имеет полосу пропускания

Лекция 2. Цифровые методы обработки и передачи звуковых сообщений

При реализации современных систем телекоммуникаций предпочтение отдают цифровым методам обработки и передачи сигналов. Цифровые системы по сравнению с аналоговыми имеют ряд существенных преимуществ при обработке, запоминании и передаче сигналов. Представление сообщений в цифровой форме обеспечивает высокую помехоустойчивость, возможность более полного использования пропускной способности каналов, стабильность параметров передачи и гибкость при построении телекоммуникационных сетей. Особое значение приобретает применение цифровых методов при передаче звуковых сообщений. В данном разделе рассмотрены методы кодирования источника звуковых сообщений. Устройство, осуществляющее преобразование первичного речевого сигнала в цифровую форму, называют *речевым кодером.*

2.1. Особенности представления звуковых сообщений в цифровой форме

Человеческий голос порождает первичный аналоговый сигнал, который занимает полосу частот примерно от 50 до 10000 Гц. Представление этого сигнала в цифровой форме осуществляется путем дискретизации во времени и квантования по уровням (рис. 2.1) и сопровождается неустранимой ошибкой, называемой шумом квантования. Шум квантования – один из факторов, определяющих верность передачи непрерывных сообщений по дискретному каналу (вторым фактором являются помехи в канале передачи, накладывающиеся на полезный сигнал и приводящие к ошибочному приему).

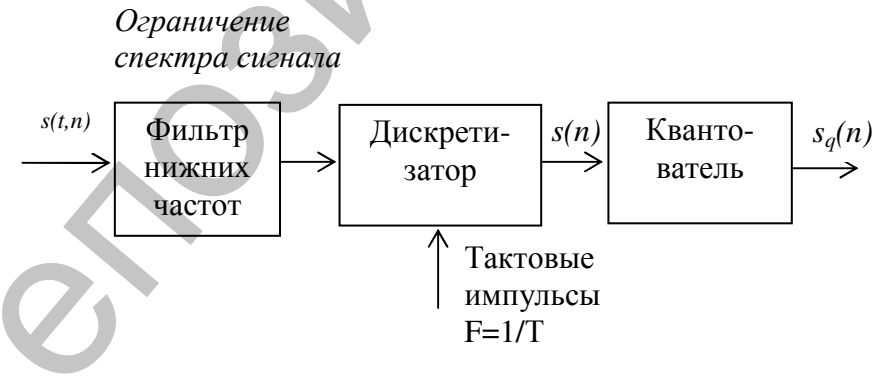


Рис. 2.1

При квантовании возникает ошибка квантования $\varepsilon(n)$, равная разнице между квантованным $s_q(n)$ и истинным значениями сигнала:

$$\varepsilon(n) = s_q(n) - s(n), \qquad (2.1)$$

где n - порядковый номер отсчета.

Искажения, вносимые квантователем, оценивают значением среднеквадратичной ошибки (СКО):

$$D = \sqrt{\frac{1}{M}\sum_{n=1}^{M}[s_q(n) - s(n)]^2}\ . \qquad (2.2)$$

Если значение квантуемого сигнала не выходит за пределы рабочего диапазона квантователя, то ошибка

$$|\varepsilon(n)| \le \frac{\Delta}{2}, \qquad (2.3)$$

где Δ – шаг квантования.

Как следует из выражения (2.3) для снижения ошибки (шума) квантования необходимо снижать шаг квантования и соответственно увеличивать число уровней квантования. Если полный размах непрерывного сигнала равен $2S_{\max}$, то число уровней квантования

$$L_{KB} = 2S_{\max}/\Delta + 1. \qquad (2.4)$$

Наиболее очевидный подход заключается в использовании квантователя с постоянным шагом квантования Δ. В этом случае для высококачественной передачи звуковых сообщений с малой ошибкой квантования, как показывает практика, требуется, чтобы $L_{KB} \ge 4000$. При цифровом кодировании такого сигнала с помощью двоичных символов на каждый дискретный отсчет потребуется не менее $n = 12$ разрядов, поскольку $L_{KB} = 2^n$.

Оценим скорость цифрового потока в телекоммуникационном канале при передаче звуковых сообщений.

Для передачи речи в аналоговой телефонии в 60-х годах 20 столетия была выбрана полоса частот 0,3-3,4 кГц. Решающими в выборе такой полосы были экономические соображения и нехватка телефонных каналов. Несмотря на определенное ухудшение восприятие ряда звуков (например, шипящих, существенная часть энергии которых сосредоточена в верхней части речевого спектра), такое ограничение незначительно повлияло на разборчивость речи.

При представлении речевых сигналов в цифровой форме верхнюю частоту в спектре дискретизируемого сигнала выбирают равной 4 кГц. Согласно теореме Котельникова при $F_{\max} = $ =4кГц, период дискретизации составляет $T_Д = 1/(2F_{\max}) = 125$ мкс. При этом частота дискретизации $F_Д = 1/T_Д = 8000$ Гц. Скорость цифрового потока соответственно равна

$$W = F_Д n. \qquad (2.5)$$

При передаче речевых сообщений, использующей 12-разрядное кодирование отсчетов, скорость цифрового потока, поступающего на вход телекоммуникационного канала, составит 96 кбит/сек. Еще более высокие требования будут предъявляться к пропускной способности канала при передаче высококачественных звуковых сообщений, например, музыки. Известно, что для высококачественного воспроизведения музыки на компакт-дисках частота дискретизации составляет $F_Д = 44{,}1$ кГц при 16-

разрядном кодировании отсчетов. Подставляя эти значения в (2.5), определим скорость цифрового потока, которая составит:

W =44100 отсчетов/сек · 16 бит/отсчет = 705,6 кбит/сек.

При использовании 2-х стереофонических каналов скорость цифрового потока превысит 1400 кбит/сек.

Необходимость эффективного использования телекоммуникационных каналов явилась причиной разработки специальных технических решений, позволяющих уменьшить скорость цифрового потока при передаче речевых сообщений. Процедуру преобразования речевых сигналов, при которой уменьшается скорость цифрового потока, назвали *компрессией* (сжатием). Практический эффект такого уменьшения скорости очевиден – появляется возможность обслужить большее количество абонентов на телекоммуникационном канале с заданной пропускной способностью или осуществлять передачу речевых сообщений по низкоскоростным каналам, по которым передача сигналов в некомпрессированном виде была бы невозможна.

2.2. Классификация методов и показатели качества компрессии сигналов при передаче звука

Речевые кодеры можно разделить на 3 основные группы: *кодеры формы*, *вокодеры* и *гибридные кодеры*.

В кодерах формы обработке подвергается каждый отсчет дискретизированной последовательности. Кодеры данного типа обеспечивают сохранение и передачу формы исходного аналогового сигнала. При этом, как правило, достигается достаточно высокое качество восстановленного сигнала, поскольку основным источником искажений формы выходного сигнала является квантование. Однако скорость цифрового потока на выходе кодера формы остается все-таки достаточно высокой. Так при передаче речевых сообщений кодеры формы формируют цифровой поток со скоростью от 24 до 64 кбит/с.

Работа вокодеров (от английских слов VOice – голос и CODER – кодировщик) основана на моделировании речевых сигналов с учетом их характерных особенностей. Это позволяет снизить скорость передачи до 0.5 - 16 кбит/с. Однако до середины 80-х годов 20-го столетия качество сигналов при вокодерном кодировании было плохим, и это ограничивало их практическое использование. Современные вокодеры обеспечивают качество, ненамного уступающее принятому в телефонной сети общего пользования, и их широко применяют, в частности, в системах подвижной радиосвязи.

В гибридных кодерах используется метод, объединяющий преимущества кодеров формы и вокодеров.

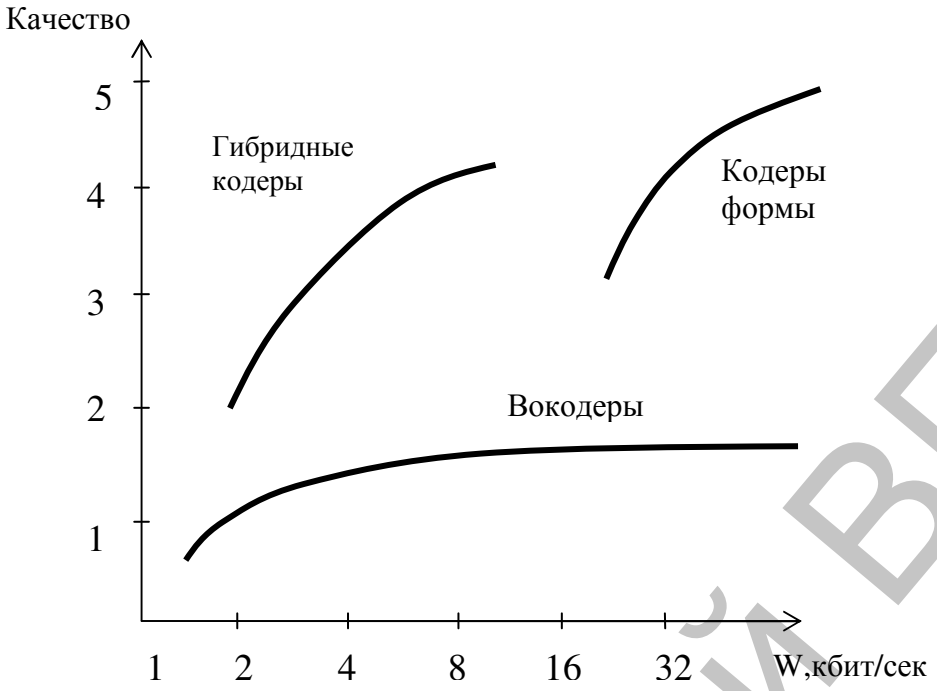


Рис.2.2

Качество кодирования и восстановления речевых сигналов измеряется часто по пятибалльной шкале MOS (mean opinion score - средняя субъективная оценка). Поскольку человек как получатель информации является ключевым элементом любой телекоммуникационной системы, качество сигнала оценивается по его субъективному восприятию речи. Оценка по шкале MOS определяется путем обработки оценок, даваемых группами слушателей нескольким речевым сигналам, воспроизводимым различными громкоговорителями. Каждый слушатель выносит оценку каждого сигнала: 1 - плохо, 2 - слабо, 3 - разборчиво, 4 - хорошо, 5 - отлично. Затем результаты усредняются. Соотношение качества и скорости для рассматриваемых методов приведено на рис.2.2. [2].

2.3. Кодеры формы

Нелинейное кодирование

Для уменьшения сравнительно большого количества уровней квантования, которое вытекает из соотношений, полученных в разделе 2.1 в предположении об использовании квантователя с равномерным шагом квантования, следует учесть особенности работы слухового аппарата человека.

Человеческое ухо воспринимает звук нелинейно: наиболее заметными оказываются искажения при слабом уровне звука, в то время как при большом уровне звука чувствительность к искажениям звукового сигнала снижается. Принимая во внимание указанные особенности, можно уменьшить количество уровней квантования и, соответственно, скорость

цифрового потока в телекоммуникационном канале, применив квантование с неравномерным шагом. Суть такого подхода состоит в изменении шага квантования пропорционально уровню входного сигнала. При этом малые уровни сигнала квантуются с меньшей ошибкой, чем большие. Закон изменения шага квантования определяют из условия, чтобы отношение сигнал-шум сохранялось постоянным при изменении уровня сигнала.

Условно неравномерное квантование можно представить как последовательное соединение устройства компрессии входного сигнала и равномерного квантователя (рис. 2.3.). При приеме нелинейные искажения сигнала, вносимые *компрессором*, устраняют *экспандером*, нелинейным устройством с амплитудной характеристикой, обратной характеристике компрессора.

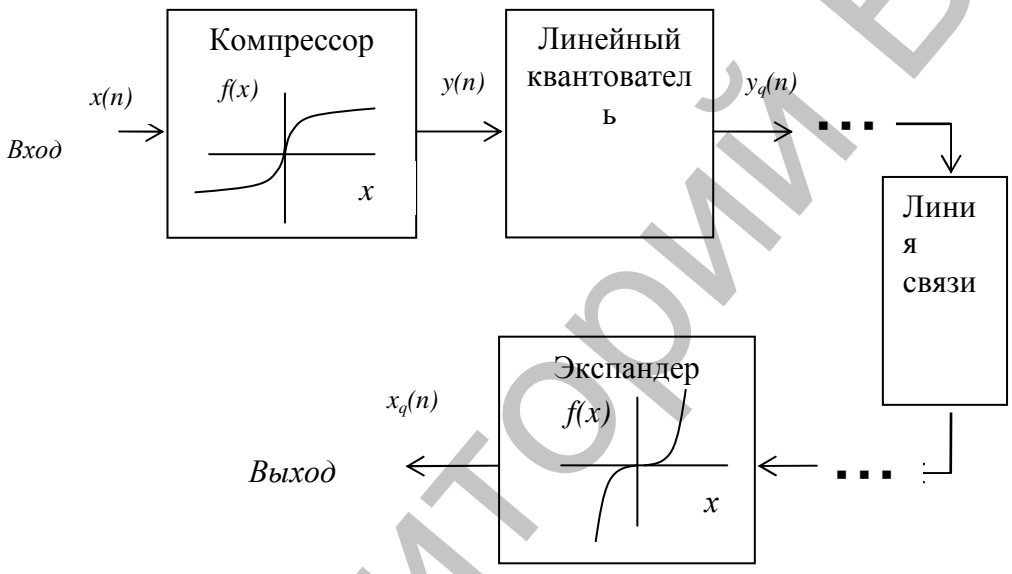


Рис.2.3

Для обозначения процессов компрессии и экспандирования для краткости пользуются одним термином – компандирование, а совокупность 2-х устройств - компрессора и экспандера называют компандером.

При передаче речи используют два типа компандирования: по μ-закону и по А-закону.

Первый метод используют в США и Японии. При μ-законе сигнал в *компрессоре* преобразуется следующим образом:

$$y_\mu = \frac{sign(x)}{\ln(1+\mu)} \ln\left(1+\mu\left|\frac{x}{x_{\max}}\right|\right), \qquad (2.6)$$

где x - сигнал на входе компрессора,

$x_{\max}$ - его максимальное значение,

μ - константа (обычно μ=255).

А-закон используется в Европе. В этом случае компрессор преобразует сигнал следующим образом:

$$y_A = \begin{cases} \dfrac{A}{1+\ln A}\left(\dfrac{x}{x_{\max}}\right), & \left|\dfrac{x}{x_{\max}}\right| \le \dfrac{1}{A} \\ \dfrac{sign\ (x)}{1+\ln A}\left(1+\ln A\left|\dfrac{x}{x_{\max}}\right|\right), & \dfrac{1}{A} \le \left|\dfrac{x}{x_{\max}}\right| \le 1 \end{cases}.$$

Наиболее часто используют значение параметра A=87.6.

Применение рассмотренных методов компадирования позволяет в одном и том же заданном диапазоне изменения речевого сигнала вместо 12-разрядных двоичных чисел использовать восьмиразрядные двоичные числа. Таким образом, скорость цифрового потока при передаче речевого сигнала уменьшится с 96 до 64 кбит/сек. Указанный способ компрессии речи закреплен в международной рекомендации G.711.

Учитывая, что сжатие и последующее восстановление к первоначальному виду непрерывных по величине отсчетов звуковых сигналов может сопровождается появлением погрешности из-за отклонений характеристик компрессора и экспандера от расчетных значений, вместо рассмотренного выше на практике обычно применяется другой способ нелинейного кодирования (рис.2.4). После равномерного квантования при числе уровней $L = 2^{12}$ и предварительного кодирования производится цифровая компрессия, в результате чего длина кодовой комбинации уменьшается до n=8 разрядов. Результатом преобразования является двоичная последовательность со скоростью 64 кбит/с.

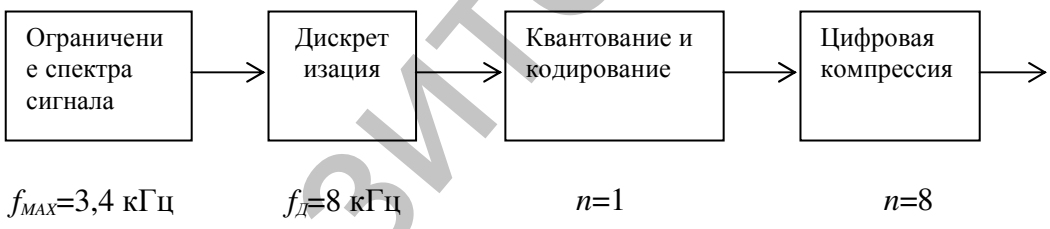


Рис. 2.4

Дифференциальная импульсно-кодовая модуляция (ДИКМ)

Обычно между двумя соседними отсчетами сигнала существует определенная взаимосвязь, которую в радиотехнике обозначают термином корреляция. Это справедливо для всех сигналов за исключением так называемого белого шума, отсчеты которого некоррелированы. Степень корреляции между отсчетами возрастает с ростом частоты дискретизации. Наличие корреляции указывает на наличие избыточности в сигнале.

Учитывая корреляцию между отсчетами, можно сжать сигнал по сравнению с обычной ИКМ. Самый распространенный метод кодирования, основанный на учете корреляции между отсчетами - *ДИКМ*-кодирование.

При ДИКМ кодируют и передают по каналу не сам отсчет (как в ИКМ), а разность (или ошибку) между текущим отсчетом и предварительной оценкой (предсказанным значением) этого отсчета, полученной из анализа предыдущих отсчетов

$$\varepsilon(n) = s(n) - s_{ПРЕДСК}(n). \qquad (2.7)$$

Чем точнее осуществляется предсказание очередного отсчета, тем меньше по величине разностный сигнал и, следовательно, тем меньшее количество разрядов потребуется для его кодирования в цифровом виде. В качестве сигнала предсказания можно использовать либо предыдущий отсчет *s(n-1),* либо *M* предшествующих отсчетов, что позволяет повысить точность предсказания:

$$s_{ПРЕДСК} = \sum_{i=1}^{M} c_i \cdot s(n-i), \qquad (2.8)$$

где c_i – коэффициенты.

Типовой вариант реализации метода ДИКМ приведен на

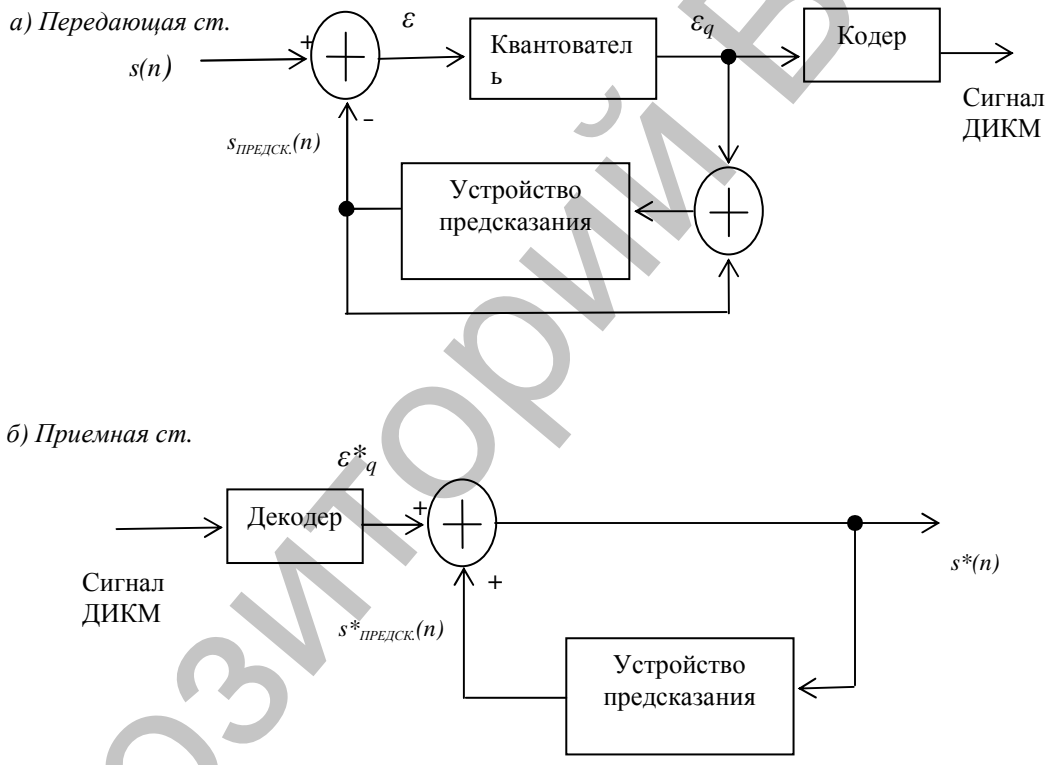


Рис.2.5

рис.2.5.

На приемной стороне (рис.2.5,б) восстановление исходного сообщения осуществляется на основе соотношения

$$s^*(n) = \varepsilon^*(n) + s^*_{ПРЕДСК}(n).$$

Эффективность метода ДИКМ может быть повышена путем перехода к адаптивной дифференциальной импульсно-кодовая модуляции (АДИКМ). При этом производится автоматическое регулирование величины шага квантования сигнала ошибки предсказания, а также автоматическая подстройка коэффициентов c_i в (2.8) в соответствии с изменением текущего спектра передаваемого сообщения. Для этого как в передающее, так и в приемное устройства вводятся дополнительные цепи автоматической регулировки усиления и подстройки параметров предсказателя на основе

статистического оценивания параметров передаваемого сообщения. Этот алгоритм дает практически такое же качество воспроизведения речи, как и ИКМ, однако для передачи информации при его использовании требуется полоса всего в 32 кбит/с. Алгоритм АДИКМ был принят в качестве международного стандарта G.726 в 1984 г.

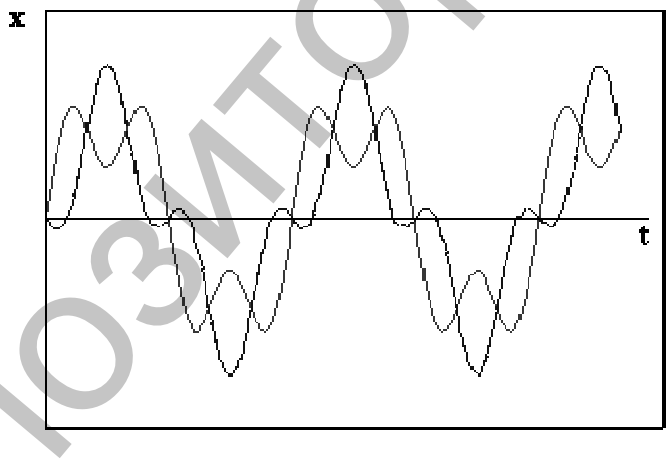
2.4. Вокодеры

Принципы вокодерного кодирования

Вокодеры предназначены для кодирования исключительно речевого сигнала. При их построении максимально учитывают особенности образования речи и ее восприятия человеком. Форма восстановленного сигнала при этом может радикально отличаться от формы исходного сигнала. В качестве примера на рис. 2.6 приведены осциллограммы двух сигналов, внешне весьма различных. Тем не менее, при их воспроизведении человек не заметит разницы. Дело в том, что спектральный состав обоих сигналов одинаков: они являются суммой синусоиды и ее третьей гармоники. Различны лишь значения начальной фазы третьей гармоники. Органы слуха человека не реагируют на фазовые соотношения.

Рис. 2.6.

Задача ИКМ, ДИКМ, АДИКМ и других аналогичных им методов -



максимального точно передать информацию о форме сигнала. Именно поэтому эти методы кодирования называют кодированием формы. Задача вокодерной обработки другая - обеспечить, чтобы восстановленный сигнал звучал как можно более сходно с исходным сигналом.

Принципиальное отличие вокодерного кодирования от кодирования формы состоит в том, что по каналу связи передают не сам сигнал, а параметры модели его образования. На приемном конце восстановленный сигнал синтезируют.

Существует большое число идей построения вокодеров. Например, в канальных или полосовых вокодерах спектр речи делят на 7 - 20 полос (каналов) аналоговыми или цифровыми полосовыми фильтрами. Большее

число каналов дает большую натуральность и разборчивость. С каждого полосового фильтра сигнал поступает на детектор и фильтр низких частот. На приемный конец раз в 20 мс передают информацию об уровне сигнала в каждом канале. Синтезатор речи представляет собой набор синусоидальных генераторов и регулируемых аттенюаторов, устанавливающих требуемые соотношения между амплитудами колебаний разных частот. Передача информации об уровне сигнала в каждом канале возможна в аналоговом или цифровом виде.

В фонемных вокодерах используют тот факт, что речь передается ограниченным числом слогов - фонем. Например, русский язык использует 42 фонемы. Выполняя фонемный анализ речи, можно периодически (например один раз в 20 мс) передавать на приемный конец номер соответствующей фонемы, закодированный 6 битами, а также информацию об уровне сигнала (еще 6 бит). Таким образом, скорость цифрового потока составит (6+6)/20=0,6 кбит/с. На приемном конце синтезатор воспроизводит соответствующую фонему, извлекая ее из памяти. Известны и другие принципы вокодерного кодирования. Хотя первые вокодеры были предложены в 30-е годы, до начала 80-х годов качество восстанавливаемой речи было крайне низким. Область применения вокодеров ограничивалась линиями командной связи, речевого управления и говорящими автоматами информационно-справочных служб. При этом достигалась низкая скорость передачи (порядка 0,6 - 4 кбит/с).

Прогресс вокодеров в 80-е и 90-е годы непосредственно связан с новыми возможностями цифровой обработки сигналов и микропроцессорной техники. С другой стороны, он явился ответом на потребности быстро развивающегося рынка массовых цифровых систем подвижной радиосвязи, в частности сотовых систем.

Вокодеры используют достаточно сложные алгоритмы обработки речевых сигналов и по этой причине выполняются на основе цифровых сигнальных процессоров (ЦСП). Производительность ЦСП обычно оценивают в миллионах операций в секунду. Вокодеры, использующие ЦСП, способные выполнять 15 млн. операций в секунду, относятся к низкопроизводительным, если указанный параметр превышает 30 млн. операций в секунду, то такие вокодеры считаются высокопроизводительными.

Некоторые особенности процесса речеобразования, учитываемые в вокодерах

Рассмотрим особенности процесса речеобразования. При разговоре грудная клетка сжимается и расширяется, поток воздуха проходит из легких через трахею и гортань в полости глотки, рта и носа. Голосовой тракт простирается от голосовой щели (отверстия между голосовыми складками в гортани) до губ. В процессе речеобразования его форма меняется.

Если произносятся звонкие звуки (гласные, носовые, звонкие согласные), голосовые складки в гортани смыкаются и размыкаются с той

или иной частотой, которая называется частотой основного тона. Получается последовательность импульсов воздушного потока, которые возбуждают полости голосового тракта. Говоря, человек меняет геометрические размеры этих полостей, соответственно меняются и их резонансные частоты, которые называют формантами. Звонкие звуки называются также вокализованными.

Частота основного тона обычно находится в интервале от 50 до 400 Гц. На рис. 2.7 приведены временная зависимость и спектр, соответствующие гласному звуку "и". Хорошо виден периодический характер сигнала; в спектре ярко выражены основной тон и форманты.

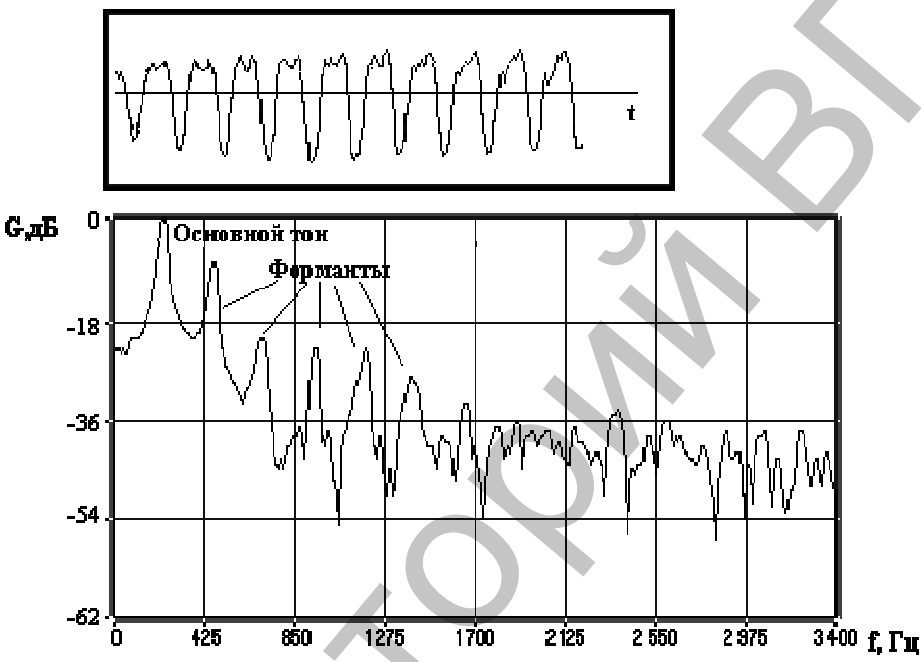


Рис. 2.7

При произнесении глухих (невокализованных) звуков голосовые складки расслаблены. Проходя по суженному голосовому тракту, воздух создает турбулентный поток. Полости рта и носа возбуждаются при этом шумоподобным сигналом. На рис. 2.8 показаны временная зависимость и спектр, соответствующие глухому согласному звуку "с".

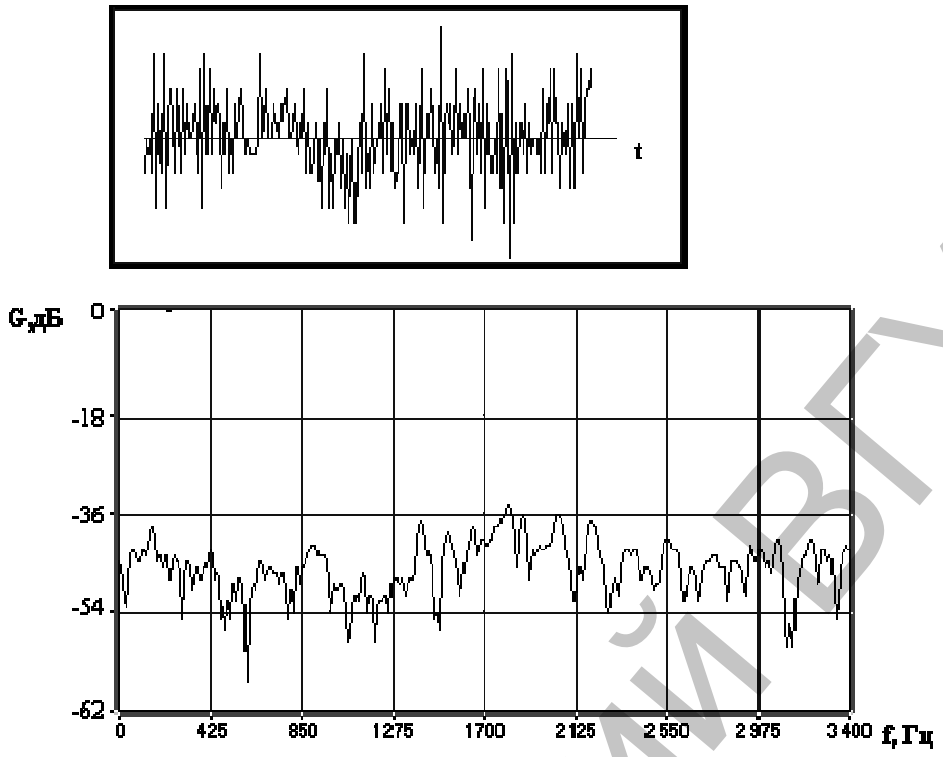


Рис. 2.8

Сигнал не содержит периодических составляющих и подобен шуму; в спектре отсутствуют форманты и основной тон.

Взрывные (смычные) звуки получаются путем кратковременного выхлопа - полного перекрытия речевого тракта, нагнетания давления и внезапного открытия тракта. Взрывные звуки бывают звонкие (б, д, г) и глухие (п, т, к), то есть могут образовываться с участием голосовых складок и без них.

Органы речи обладают инерционностью: на интервале 20 - 30 мс параметры речи можно считать постоянными.

Метод линейного кодирования с предсказанием

Многие из методов вокодерного кодирования берут свое начало от изобретенного довольно давно метода LPC (Linear Predictive Coding). В качестве входного сигнала в LPC используется та же последовательность цифровых значений амплитуды, однако этот метод применяется не к отдельным цифровым значениям, а к определенным их блокам. Для каждого такого блока значений вычисляются его характерные параметры: частота, амплитуда и ряд других. Именно эти значения и передаются по сети. При таком подходе к кодированию речи, во-первых, возрастают требования к вычислительным мощностям специализированных процессоров, используемых для обработки сигнала, а во-вторых, увеличивается задержка при передаче, поскольку кодирование применяется не к отдельным значениям, а к некоторому их набору, который перед началом преобразования следует накопить в определенном буфере. Подчеркнем, что задержка в передаче речи при использовании этого метода связана не только с необходимостью обработки цифрового сигнала (эту задержку можно

уменьшать, увеличивая мощность процессора), а непосредственно следует из характера метода сжатия. Этот метод позволяет, вообще говоря, достигать очень больших степеней сжатия, которым соответствует полоса пропускания 2,4 или 4,8 кбит/с, однако качество звука здесь сильно страдает. Поэтому в коммерческих приложениях он не используется, а применяется в основном для ведения служебных переговоров.

2.5. Гибридные кодеры

Более сложные методы сжатия речи основаны на применении LPC в сочетании с элементами кодирования формы сигнала. В этих алгоритмах используется кодирование с обратной связью, когда при передаче сигнала осуществляется оптимизация кода. Закодировав сигнал, процессор пытается восстановить его форму и сличает результат с исходным сигналом, после чего начинает варьировать параметры кодировки, добиваясь наилучшего совпадения. Достигнув такого совпадения, аппаратура передает полученный код по линиям связи; на противоположном конце происходит восстановление звукового сигнала. Ясно, что для использования такого метода требуются еще более серьезные вычислительные мощности.

Одной из наиболее распространенных разновидностей описанного метода кодирования является метод LD-CELP (Low-Delay Code-Excited Linear Prediction). Этот метод позволяет достичь удовлетворительного качества воспроизведения при пропускной способности 16 кбит/с; он был стандартизован Международным союзом электросвязи (International Telecommunications Union - ITU) в 1992 году как алгоритм кодирования речи G.728 [3]. Алгоритм применяется к последовательности цифр, получаемых в результате аналого-цифрового преобразования голосового сигнала с 16-разрядным разрешением.

Пять последовательных цифровых значений кодируются одним 10-битовным блоком - это и дает те самые 16 кбит/с. Для применения этого метода требуются очень большие вычислительные мощности, в частности, для прямолинейной реализации G.728 необходим процессор с быстродействием 44 млн. операций в секунду.

В марте 1995 года ITU принял новый стандарт G.723, который предполагается использовать при сжатии речи для организации видеоконференций по телефонным сетям. Этот стандарт является частью более общего стандарта H.324, описывающего подход к организации таких видеоконференций, при этом целью является обеспечение видеоконференций с использованием обычных модемов. Основой G.723 является метод сжатия речи MP-MLQ (Multipulse Maximum Likelihood Quantization). Он позволяет добиться весьма существенного сжатия речи при сохранении достаточно высокого качества звучания.

В основе метода лежит описанная выше процедура оптимизации; с помощью различных усовершенствований можно сжимать речь до уровня 4,

8; 6, 4; 7, 2 и 8,0 кбит/с. Структура алгоритма позволяет на основе программного обеспечения изменять степень сжатия голоса в ходе передачи. Вносимая кодированием задержка не превышает 20 мс.

Как показали испытания, проведенные ведущими американскими и европейскими телекоммуникационными компаниями, качество голоса, получаемое при сжатии методом MP-MLQ до уровня 6,4 кбит/с, не ниже того, что дает ADPCM при сжатии до 32 кбит/с.

Лекция 3. Цифровые методы передачи видеоизображений

3.1. Основные характеристики цифрового видео

Цифровое видео характеризуется четырьмя основными факторами [3]:

- частотой кадров
- экранным разрешением
- глубиной цвета
- качеством изображения.

Частота кадров. Вследствие инерционности процесса восприятия изображения человеческий глаз интерпретирует последовательность быстро сменяющих друг друга кадров не как отдельные изображения, а как непрерывно протекающий процесс. Это происходит при смене кадров с частотой 16-25 Гц. С учетом данной особенности стандартная скорость воспроизведения видеосигнала телевизионных системах - 25 кадров/с (в США и Японии – 30 кадров/с). Но даже при частоте 25 кадров/с может появляться эффект мелькания изображения. Для гарантированного устранения данного недостатка в телевизионных системах каждый кадр разбивается на 2 полукадра, каждый из которых прорисовывается либо по четным, либо по нечетным строкам изображения. Эти полукадры передаются поочередно с частотой 50 Гц каждый. И хотя на самом деле частота смены кадров полного изображения составляет 25 кадров/с, глаз воспринимает изображение при чересстрочной развертке как смену кадров 50 раз в секунду.

Монитор компьютера для формирования изображения на экране использует метод прогрессивного сканирования, при котором строки кадра формируются последовательно, сверху вниз, а полный кадр прорисовывается 30 раз каждую секунду или даже более часто. В этом заключается основное отличие между компьютерным и телевизионным методом формирования видеосигнала.

Глубина цвета. Этот показатель является комплексным и определяет количество цветов, одновременно отображаемых на экране. Компьютеры обрабатывают цвет в RGB-формате (красный-зеленый-синий), в то время как видео использует и другие методы. Одна из наиболее распространенных моделей цветности для видеоформатов – YUV. Здесь Y обозначает яркостный сигнал, а сигналы цветности определяются по следующим формулам:

U=B-Y,

V=R-Y.

Разностные сигналы U и V формируют вместе с сигналом Y полный видеосигнал.

Каждая из моделей RGB и YUV может быть представлена разными уровнями глубины цвета (максимального количества цветов).

Для цветовой модели RGB обычно характерны следующие режимы глубины цвета: 8 бит/пиксел (256 цветов), 16 бит/пиксел (65535 цветов) и 24 бит/пиксел (16,7 млн. цветов).

Экранное разрешение. Данным термином обозначают количество точек, из которых состоит изображение на экране. В современных мониторах может применяться разрешение 1024x768 пикселов, 1440x900(широкий формат) пикселов и выше.

Телевизионный стандарт NTSC предусматривает разрешение 768 на 484. Стандарт PAL распространенный в Европе, имеет несколько большее разрешение - 768 на 576 точек.

Качество изображения. Последняя, и наиболее важная характеристика - это качество видеоизображения. Требования к качеству зависят от конкретной задачи. Иногда достаточно, чтобы картинка была размером в четверть экрана с палитрой из 256-ти цветов (8 бит), при скорости воспроизведения 15 кадров/с. В других случаях требуется полноэкранное видео (768 на 576) с палитрой в 16,7 млн. цветов (24 бит) и полной кадровой разверткой (25 или 30 кадров/с).

3.2. Свойства системы зрения человека

Очень часто окончательную оценку изображения делает человек. Исследование системы зрения человека показывает, что она обладает нелинейной характеристикой, а ее отклик не является абсолютно верным. Рассмотрим указанные особенности подробнее.

Одной из характеристик системы зрения человека является способность восприятия яркости света. Эксперименты по определению восприятия людьми минимально различимых градаций яркости света, поступающего от калиброванного источника, показали, что яркость света воспринимается глазом нелинейно [1]. Если начертить график зависимости величины этой минимально различимой градации яркости от эталонной яркости, то при изменении яркости в пределах нескольких порядков этот график имеет логарифмический характер. Такие субъективные экспериментальные результаты согласуются с объективными данными, полученными в экспериментах на животных, в которых установлено, что светочувствительные клетки сетчатки и оптический нерв возбуждаются с частотой, пропорциональной логарифму интенсивности подводимого к ним света. По вполне понятным причинам подобные объективные измерения на людях не проводились.

Другой отличительной особенностью системы зрения человека является ее пространственно-частотный отклик. Точная форма частотной характеристики глаза исследовалась с помощью ряда психовизуальных экспериментов. Было показано, что глаз подавляет низкие и ослабляет высокие пространственные частоты. В определенном смысле пространственно-частотный отклик имеет полосовой характер.

Наконец, для системы зрения человека характерна способность к насыщению, т.е. к ограничению отклика при очень больших или очень малых интенсивностях наблюдаемого светового потока.Рассмотренные особенности

были учтены при разработке методов сокращения избыточности изображений, рассматриваемых ниже.

3.3. Межкадровая и внутрикадровая избыточность изображения

Избыточность изображения проявляется в высокой степени взаимной статистической прогнозируемости элементов изображения. В радиотехнике такую связь сигналов характеризуют понятием *корреляция*. Конечной целью операции сжатия видеоинформации является устранение этой статистической прогнозируемости, т.е. необходимо в максимально возможной степени уменьшить коррелированность отсчетов, полученных при дискретизации видеосигналов.

Типичное изображение содержит очень много избыточной информации. Межкадровая избыточность изображений связана с необходимостью передавать изображение с достаточно высокой частотой. При этом изменение либо целого изображения, либо отдельных его участков от кадра к кадру может быть или небольшим или даже отсутствовать. Такой эффект имеет место, например, когда в телевидении передается статическая заставка экрана. Поэтому одним из способов уменьшения избыточности передаваемых сигналов изображения является передача не абсолютных значений сигналов, соответствующих определенным элементам изображений в различных кадрах, а их изменения от кадру к кадру.

Внутрикадровая избыточность обусловлена высокой степенью однородности изображения на малых участках изображения в пределах одного кадра. Размер этого участка можно оценить, вычисляя коэффициент корреляции между яркостями точек изображения, находящихся на различном расстоянии от точки, выбранной в качестве опорной. Расстояние, при котором коэффициент корреляции становится меньше некоторой заданной величины (обычно 5-10% от максимального значения), и есть искомый размер. Анализ показал, что для большинства изображений размер участка, в пределах которого проявляется взаимосвязь яркостей точек изображения, составляет 16x16 точек [1].

3.4. Алгоритмы сжатия цифровых сигналов при передаче видеоизображений

Для того, чтобы оценить, насколько актуальной является проблема компрессии цифровых сигналов при передаче видеоизображений, определим скорость цифрового потока, которая получится при преобразовании изображения размером 800x600 пикселов, следующего с частотой 25 кадров/секунду (напомним, что это минимальная частота, которая требуется для устранения эффекта мелькания при смене кадров), при глубине цвета 24 бит/пиксел:

W=800x600x25x24=288 Мбит/с.

Еще сравнительно недавно в середине 90-х годов XX столетия не существовало телекоммуникационных технологий, с помощью которых такой цифровой поток в нескомрессированном виде можно было бы передать по линиям связи. Другая проблема заключается в том, что при сохранении нескомпрессированных сигналов видеоизображений в компьютере объем его дискового пространства может оказаться сравнительно быстро исчерпанным. Ну и конечно же рассмотренный выше пример показал, что проблема компрессии цифровых сигналов при передаче по телекоммуникационным линиям связи сигналов видеоизображений является намного более актуальной, чем при передаче аудио сигналов.

Эффективность процесса компрессии сигналов видеоизображений оценивают с помощью коэффициента сжатия. Коэффициент сжатия - это цифровое выражение соотношения между объемом сжатого и исходного видеоматериала. Для примера, коэффициент 200:1 означает, что если принять объем полученного после компрессии ролика за единицу, то исходный оригинал занимал объем в 200 раз больший. Обычно, чем выше коэффициент сжатия, тем хуже качество видео. Но многое, конечно, зависит от используемого алгоритма.

При определении необходимой степени сжатия сигналов видеоизображений следует исходить из разумной достаточности. При этом необходимо учитывать, как четыре характеристики (частота кадра, экранное разрешение, глубина цвета и качество изображения) влияют на объем и качество видео. Очень важно иметь представление, какую цену придется заплатить за качественное изображение. Чем больше глубина цвета, выше разрешение и лучше качество, тем более высокая производительность компьютера потребуется, не говоря уж о громадных объемах дискового пространства, необходимого под цифровое видео. Учитывая эти характеристики, можно выбрать оптимальный коэффициент сжатия. Надо отметить, что в профессиональном видео действует простое правило - чем ниже коэффициент сжатия, тем лучше.

Различают сжатие в режиме реального времени, симметричное или асимметричное, с потерей качества или без потери, сжатие видеопотока или покадровое сжатие.

Сжатие в режиме реального времени. Термин real-time (реальное время) имеет много толкований. Применительно к сжатию данных используется его прямое значение, т. е. работа в реальном времени. Многие системы оцифровывают видео и одновременно сжимают его, иногда параллельно совершая и обратный процесс декомпрессии и воспроизведения. Для качественного выполнения этих операций требуются очень высокопроизводительные специальные процессоры, поэтому некоторые платы ввода/вывода видео для персональных компьютеров не способны оперировать с полнометражным видео и часто пропускают кадры.

Симметричное и асимметричное сжатие. Этот признак классификации связан с соотношением способов сжатия и декомпрессии

видео. Симметричное сжатие предполагает возможность проиграть видеофрагмент с разрешением, например, 640 на 480 пикселов при скорости в 30 кадров/с, если оцифровка и запись его выполнялась с теми же параметрами. Асимметричное сжатие - это процесс обработки одной секунды видео за значительно большее время. Степень асимметричности сжатия обычно задается в виде отношения. Так цифры 150:1 означают, что сжатие одной минуты видео занимает примерно 150 минут реального времени. Асимметричное сжатие обычно более удобно и эффективно для достижения качественного видео и оптимизации скорости его воспроизведения. К сожалению, при этом кодирование полнометражного ролика может занять слишком много времени, вот почему подобный процесс выполняют специализированные компании, куда отсылают исходный материал на кодирование (что увеличивает материальные и временные расходы на проект).

Сжатие с потерей или без потери качества. Чем выше коэффициент сжатия, тем больше уменьшается качество видео. Почти все методы сжатия видно приводят к потере качества. Даже если это не заметно на глаз, всегда есть разница между исходным и сжатым материалом.

Сжатие видеопотока или покадровое сжатие. Покадровый метод подразумевает сжатие и хранение каждого видеокадра как отдельного изображения. Сжатие видеопотока основано на следующей идее: не смотря на то, что изображение все время претерпевает изменения, задний план в большинстве видеосцен остается постоянным -- отличный повод для соответствующей обработки и сжатия изображения. Создается исходный кадр, а каждый следующий сравнивается с предыдущим и последующим изображениями, а фиксируется лишь разница между ними. Этот метод позволяет существенно повысить коэффициент сжатия, практически сохранив при этом исходное качество.

3.5. Примеры форматов цифрового видео

AVI (Audio Video Interleave)

Разработанный фирмой Microsoft метод сжатия, записи и воспроизведения движущих изображений (Live Video) и звука на компьютере с использованием только программных средств. Файлы, созданные с использованием этого метода, имеют расширение AVI.

AVI может иметь или не иметь звуковые дорожки. При создании AVI файлов, включающих звуковое сопровождение, важным является правильная синхронизация звука с видеоизображением. Для этого используется технология чередования видеокадров и звука, которой, собственно, и определяется аббревиатура AVI (Audio Video Interleaved). Разные по типу видео и аудиоданные записываются в один файл на диске следующим образом: все информационные потоки разбиваются на множество равных частей (chunks) и затем записываются в один файл друг за другом по

очереди. Например, сначала записывается заголовок; затем - 1-я часть видео; затем - 1-я часть звука; затем - 2-я часть видео; затем - 2-я часть звука и т.д.

Motion-JPEG

Стандарт компрессии JPEG был разработан объединенной группой экспертов по фотографии (JPEG - Joint Photographic Expert Group) международной организации стандартов (ISO). Как ясно уже из названия, схема компрессии была разработана для неподвижных изображений. Так как телевидение, в сущности, и есть последовательность неподвижных изображений, то JPEG кодирование может применяться и для компрессии видеоизображений. Иногда этот стандарт называют "динамический" JPEG.

В основе схемы компрессии JPEG лежит дискретное косинусное преобразование (DCT). К преимуществам JPEG относится тот факт, что каждый кадр сжимается независимо от остальных и для восстановления исходного изображения не нужно задействовать информацию из соседних кадров. Такое построение сжатых данных позволяет осуществлять произвольный доступ, коммутацию и монтаж видеофрагментов проще, чем при использовании других методов кодирования. Недостатком данного формата является сравнительно невысокое значение коэффициента компрессии, а также высокие требования к производительности процессора, от которого требуется декодировать каждый кадр скомпрессированного изображения за 1/25 долю секунды. Различные варианты Motion-JPEG позволяют получить значение коэффициента компрессии от 5:1 до 100:1, однако следует подчеркнуть, что уже при значении коэффициента компрессии 20:1 качество изображения в большинстве случаев становится неудовлетворительным.

MPEG

В январе 1992 года группа экспертов в области движущихся изображений MPEG (Motion Picture Experts Group) представила первую часть стандарта для сжатия цифрового видео и звука - MPEG phase 1, или просто MPEG-1 (ISO 11172). Стандарт определяет методы компрессии и воспроизведения видео- и аудиоданных. Комитет MPEG также определил ряд других форматов для сжатого видео- и аудиоматериала. Форматы MPEG различаются по качеству результатов и скорости передачи данных:

MPEG-1: оригинальный формат для хранения и воспроизведения видео- и аудиоданных на мультимедиа носителях данных (компакт-дисках). Потенциально поддерживает телевизионное качество видео. Однако, при скорости передачи данных в диапазоне 150 - 255 Кбайт/сек. качество сопоставимо с видеозаписью VHS (разрешение 352 x 228 (PAL) или 320 x 240 (NTSC) при частоте 25 или 30 кадров в секунду соответственно).

MPEG-2: более новый стандарт (утвержден в ноябре 1994 г.). Разработан как дополнение к стандарту MPEG-1. Поддерживает передачу высококачественного видео по высокоскоростным цифровым каналам. Интенсивность потока данных от до 2 до 15 Мбит/сек. Разрешение 720x480 и 1280x720, частота 60 кадров в секунду со звуковыми данными CD-качества.

Подходит для всех стандартов телевидения и даже систем телевидения высокой точности (High Definition Television). Используется при записи DVD дисков.

MPEG-4: предназначен для передачи видео и аудиоданных по низкоскоростным линиям. Этот формат расчитан для применения в системах видеотелефонии, мультимедийной электронной почте, электронных информационных изданиях и т.п.. Базируется на формате файлов QuickTime. MPEG-4, версия 1 одобрен в октябре 1998 г. Стандарт ориентирован на разрешение 174x144 пиксела при 10 кадрах в секунду и позволяет передавать данные со скоростью от 4800 до 64000 бит/сек.

MPEG-1 и MPEG-2 признаны международными стандартами для сжатия видео.

Технология MPEG использует поточное сжатие видео, при котором обрабатывается не каждый кадр по отдельности (как это происходит при сжатии видео с помощью алгоритмов Motion-JPEG), а анализируется динамика изменений видеофрагментов и устраняются избыточные данные. Поскольку в большинстве моментов фон изображения остается достаточно стабильным, а действие происходит только на переднем плане, алгоритм MPEG начинает сжатие с создания исходного (ключевого) кадра. Играя роль опорных при восстановлении остальных изображений, они размещаются последовательно через каждые 10-15 кадров. Только некоторые фрагменты изображений, которые находятся между ними, претерпевают изменения, и именно эта разница сохраняется при сжатии. Таким образом, MPEG-последовательность содержит три типа кадров:

Intro frames (кадры типа "I") – опорные кадры, которые компрессируются без обращения к другим кадрам, кадры данного типа имеют самый большой размер;

Predicted frames (кадры типа "P") - кадры данного типа компрессируются на основе обращения к предшествующему кадру типа "I" или предыдущему кадру типа "P" с целью предсказания динамики изменений видеофрагментов и формирования разностного сигнала;

Bi-directional interpolated frames (кадры типа "B" -двунаправленные кадры), сжатие которых осуществляется при обращении к одному предшествующему и одному последующему кадру типа P, имеют минимальный размер из рассматриваемых типов кадров.

Последовательность кадров, которая передается по телекоммуникационному каналу, в соответствии с рекомендациями стандарта MPEG, имеет вид:

I B B P B B P B B P B B P B B I ...

Отдельные изображения состоят из структурных единиц - макроблоков, соответствующих участку изображения размером 16X16 пикселов. Компьютер анализирует изображения и ищет идентичные или похожие макроблоки, сравнивая базовые и последующие кадры. В результате

сохраняется только данные о различиях между кадрами, называемые вектором смещения (vector movement code) .

Макроблоки, которые не претерпевают изменений или претерпевают сравнительно небольшие изменения, при формировании разностного сигнала практически не вносят вклад, так что количество данных для реального сжатия и хранения существенно снижаются. В результате при использовании MPEG-технологии можно достигнуть рабочего коэффициента более чем 200:1, хотя это приводит к некоторой потере качества.

Лекция 4. Определение локальных сетей и их топология

Определение локальной сети

Чаще всего термин "локальные сети" или "локальные вычислительные сети" (**LAN,** Local Area Network) понимают буквально, то есть это такие сети, которые имеют небольшие, локальные размеры, соединяют близко расположенные компьютеры. Однако достаточно посмотреть на характеристики некоторых современных локальных сетей, чтобы понять, что такое определение не точно. Например, некоторые локальные сети легко обеспечивают связь на расстоянии нескольких десятков километров. Это уже размеры не комнаты, не здания, не близко расположенных зданий, а, может быть, даже целого города. С другой стороны, по глобальной сети (**WAN**, Wide Area Network или **GAN**, Global Area Network) вполне могут связываться компьютеры, находящиеся на соседних столах в одной комнате, но ее почему-то никто не называет локальной сетью. Близко расположенные компьютеры могут также связываться с помощью кабеля, соединяющего разъемы внешних интерфейсов (RS232-C, Centronics) или даже без кабеля по инфракрасному каналу (IrDA). Но такая связь тоже почему-то не называется локальной.

Из данного определения следует, что скорость передачи по локальной сети обязательно должна расти по мере роста быстродействия наиболее распространенных компьютеров. Именно это и наблюдается: если еще десять лет назад вполне приемлемой считалась скорость обмена в 10 Мбит/с, то сейчас уже среднескоростной считается сеть, имеющая пропускную способность 100 Мбит/с, активно разрабатываются, а кое-где используются средства для скорости 1000 Мбит/с и даже больше. Без этого уже нельзя, иначе связь станет слишком узким местом, будет чрезмерно замедлять работу объединенного сетью виртуального компьютера, снижать удобство доступа к сетевым ресурсам.

Таким образом, сформулировать отличительные признаки локальной сети можно следующим образом:

- Высокая скорость передачи информации, большая пропускная способность сети. Приемлемая скорость сейчас — не менее **10 Мбит/с**.
- Низкий **уровень ошибок передачи** (или, что тоже самое, высококачественные каналы связи). Допустимая вероятность ошибок передачи данных должна быть порядка 10^{-8} — 10^{-12}.

- Эффективный, быстродействующий **механизм управления обменом** по сети.
- Заранее четко ограниченное **количество компьютеров, подключаемых к сети.**

При таком определении понятно, что глобальные сети отличаются от локальных прежде всего тем, что они рассчитаны на неограниченное число абонентов. Кроме того, они используют (или могут использовать) не слишком качественные каналы связи и сравнительно низкую скорость передачи. А механизм управления обменом в них не может быть гарантированно быстрым. В глобальных сетях гораздо важнее не качество связи, а сам факт ее существования.

Нередко выделяют еще один класс компьютерных сетей — городские, региональные сети (**MAN**, Metropolitan Area Network), которые обычно по своим характеристикам ближе к глобальным сетям, хотя иногда все-таки имеют некоторые черты локальных сетей, например, высококачественные каналы связи и сравнительно высокие скорости передачи. В принципе городская сеть может быть локальной со всеми ее преимуществами.

Правда, сейчас уже нельзя провести четкую границу между локальными и глобальными сетями. Большинство локальных сетей имеет выход в глобальную. Но характер передаваемой информации, принципы организации обмена, режимы доступа к ресурсам внутри локальной сети, как правило, сильно отличаются от тех, что приняты в глобальной сети. И хотя все компьютеры локальной сети в данном случае включены также и в глобальную сеть, специфики локальной сети это не отменяет. Возможность выхода в глобальную сеть остается всего лишь одним из ресурсов, разделяемых пользователями локальной сети.

По локальной сети может передаваться самая разная цифровая информация: данные, изображения, телефонные разговоры, электронные письма и т.д. Кстати, именно задача передачи изображений, особенно полноцветных динамических, предъявляет самые высокие требования к быстродействию сети. Чаще всего локальные сети используются для разделения (совместного использования) таких ресурсов, как дисковое пространство, принтеры и выход в глобальную сеть, но это всего лишь незначительная часть тех возможностей, которые предоставляют средства локальных сетей. Например, они позволяют осуществлять обмен информацией между компьютерами разных типов. Полноценными абонентами (узлами) сети могут быть не только компьютеры, но и другие устройства, например, принтеры, плоттеры, сканеры. Локальные сети дают также возможность организовать систему параллельных вычислений на всех компьютерах сети, что многократно ускоряет решение сложных математических задач. С их помощью, как уже упоминалось, можно управлять работой технологической системы или исследовательской установки с нескольких компьютеров одновременно.

Абонент (узел, хост, станция) — это устройство, подключенное к сети и активно участвующее в информационном обмене. Чаще всего абонентом (узлом) сети является компьютер, но абонентом также может быть, например, сетевой принтер или другое периферийное устройство, имеющее возможность напрямую подключаться к сети. Далее в курсе вместо термина "абонент" для простоты будет использоваться термин "компьютер".

Сервером называется абонент (узел) сети, который предоставляет свои ресурсы другим абонентам, но сам не использует их ресурсы. Таким образом, он обслуживает сеть. Серверов в сети может быть несколько, и совсем не обязательно, что сервер — самый мощный компьютер. **Выделенный** (dedicated) сервер — это сервер, занимающийся только сетевыми задачами. **Невыделенный** сервер может помимо обслуживания сети выполнять и другие задачи. Специфический тип сервера — это сетевой принтер.

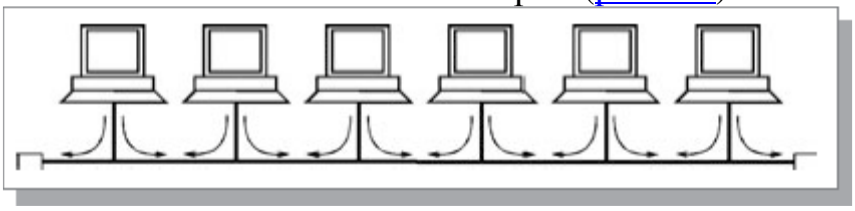
Клиентом называется абонент сети, который только использует сетевые ресурсы, но сам свои ресурсы в сеть не отдает, то есть сеть его обслуживает, а он ей только пользуется. Компьютер-клиент также часто называют **рабочей станцией**. В принципе каждый компьютер может быть одновременно как клиентом, так и сервером.

Под сервером и клиентом часто понимают также не сами компьютеры, а работающие на них программные приложения. В этом случае то приложение, которое только отдает ресурс в сеть, является сервером, а то приложение, которое только пользуется сетевыми ресурсами — клиентом.

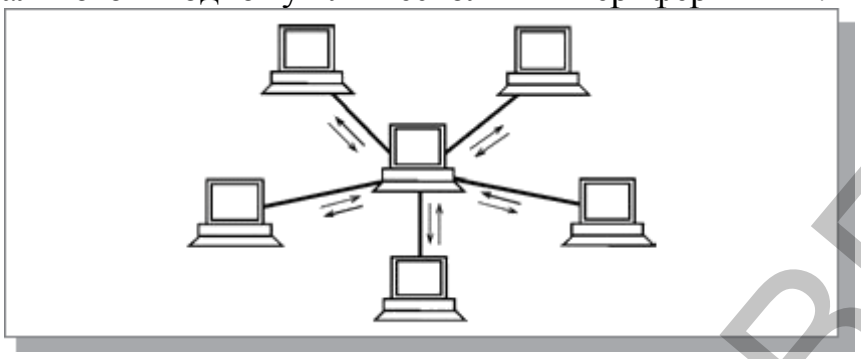
Топология локальных сетей

Под **топологией** (компоновкой, конфигурацией, структурой) компьютерной сети обычно понимается физическое расположение компьютеров сети друг относительно друга и способ соединения их линиями связи. Важно отметить, что понятие топологии относится, прежде всего, к локальным сетям, в которых структуру связей можно легко проследить. В глобальных сетях структура связей обычно скрыта от пользователей и не слишком важна, так как каждый сеанс связи может производиться по собственному пути. Топология определяет требования к оборудованию, тип используемого кабеля, допустимые и наиболее удобные методы управления обменом, надежность работы, возможности расширения сети. Существует три базовые топологии сети:

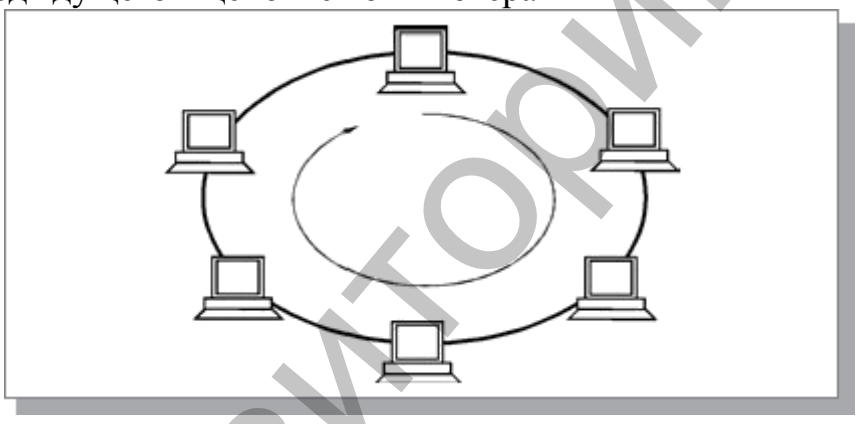
- **Шина** (bus) — все компьютеры параллельно подключаются к одной линии связи. Информация от каждого компьютера одновременно передается всем остальным компьютерам (рис. 1.5).



- **Звезда** (star) — к одному центральному компьютеру присоединяются остальные периферийные компьютеры, причем каждый из них использует отдельную линию связи (рис. 1.6). Информация от периферийного компьютера передается только центральному компьютеру, от центрального — одному или нескольким периферийным.



- **Кольцо** (ring) — компьютеры последовательно объединены в кольцо. Передача информации в кольце всегда производится только в одном направлении. Каждый из компьютеров передает информацию только одному компьютеру, следующему в цепочке за ним, а получает информацию только от предыдущего в цепочке компьютера



На практике нередко используют и другие топологии локальных сетей, однако большинство сетей ориентировано именно на три базовые топологии.

Прежде чем перейти к анализу особенностей базовых сетевых топологий, необходимо выделить некоторые важнейшие факторы, влияющие на физическую работоспособность сети и непосредственно связанные с понятием топология.

- Исправность компьютеров (абонентов), подключенных к сети. В некоторых случаях поломка абонента может заблокировать работу всей сети. Иногда неисправность абонента не влияет на работу сети в целом, не мешает остальным абонентам обмениваться информацией.
- Исправность сетевого оборудования, то есть технических средств, непосредственно подключенных к сети (адаптеры, трансиверы, разъемы и т.д.). Выход из строя сетевого оборудования одного из абонентов может сказаться на всей сети, но может нарушить обмен только с одним абонентом.

- Целостность кабеля сети. При обрыве кабеля сети (например, из-за механических воздействий) может нарушиться обмен информацией во всей сети или в одной из ее частей. Для электрических кабелей столь же критично короткое замыкание в кабеле.
- Ограничение длины кабеля, связанное с затуханием распространяющегося по нему сигнала. Как известно, в любой среде при распространении сигнал ослабляется (затухает). И чем большее расстояние проходит сигнал, тем больше он затухает (рис. 1.8). Необходимо следить, чтобы длина кабеля сети не была больше предельной длины $L_{пр}$, при превышении которой затухание становится уже неприемлемым (принимающий абонент не распознает ослабевший сигнал).

Топология шина

Топология шина (или, как ее еще называют, общая шина) самой своей структурой предполагает идентичность сетевого оборудования компьютеров, а также равноправие всех абонентов по доступу к сети. Компьютеры в шине могут передавать информацию только по очереди, так как линия связи в данном случае единственная. Если несколько компьютеров будут передавать информацию одновременно, она исказится в результате наложения (**конфликта**, **коллизии**). В шине всегда реализуется режим так называемого **полудуплексного** (half duplex) обмена (в обоих направлениях, но по очереди, а не одновременно).

В топологии шина отсутствует явно выраженный центральный абонент, через который передается вся информация, это увеличивает ее надежность (ведь при отказе центра перестает функционировать вся управляемая им система). Добавление новых абонентов в шину довольно просто и обычно возможно даже во время работы сети.

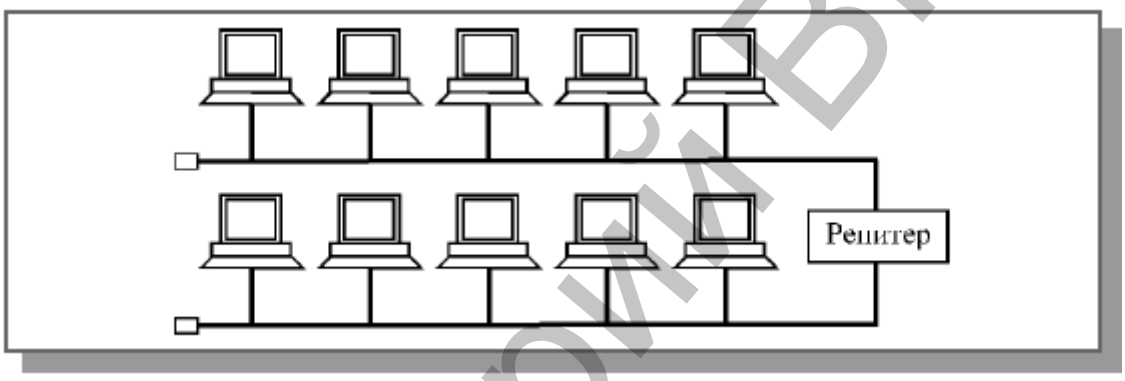
Поскольку центральный абонент отсутствует, разрешение возможных конфликтов в данном случае ложится на сетевое оборудование каждого отдельного абонента. В связи с этим сетевая аппаратура при топологии шина сложнее, чем при других топологиях. Тем не менее из-за широкого распространения сетей с топологией шина (прежде всего наиболее популярной сети Ethernet) стоимость сетевого оборудования не слишком высока.

Важное преимущество шины состоит в том, что при отказе любого из компьютеров сети, исправные машины смогут нормально продолжать обмен.

Однако надо учитывать, что из-за особенностей распространения электрических сигналов по длинным линиям связи необходимо предусматривать включение на концах шины специальных согласующих устройств, **терминаторов**, показанных на рис. 1.5 и 1.9 в виде прямоугольников. Без включения терминаторов сигнал отражается от конца линии и искажается так, что связь по сети становится невозможной. Отказ сетевого оборудования любого абонента в шине может вывести из строя всю

сеть. К тому же такой отказ довольно трудно локализовать, поскольку все абоненты включены параллельно, и понять, какой из них вышел из строя, невозможно.

Для увеличения длины сети с топологией шина часто используют несколько **сегментов** (частей сети, каждый из которых представляет собой шину), соединенных между собой с помощью специальных усилителей и восстановителей сигналов — **репитеров** или **повторителей** (на рис. 1.10 показано соединение двух сегментов, предельная длина сети в этом случае возрастает до 2 $L_{пр}$, так как каждый из сегментов может быть длиной $L_{пр}$). Однако такое наращивание длины сети не может продолжаться бесконечно. Ограничения на длину связаны с конечной скоростью распространения сигналов по линиям связи.



Топология звезда

Звезда — это единственная топология сети с явно выделенным центром, к которому подключаются все остальные абоненты. Обмен информацией идет исключительно через центральный компьютер, на который ложится большая нагрузка, поэтому ничем другим, кроме сети, он, как правило, заниматься не может. Понятно, что сетевое оборудование центрального абонента должно быть существенно более сложным, чем оборудование периферийных абонентов. О равноправии всех абонентов (как в шине) в данном случае говорить не приходится. Обычно центральный компьютер самый мощный, именно на него возлагаются все функции по управлению обменом. Никакие конфликты в сети с топологией звезда в принципе невозможны, так как управление полностью централизовано.

Если говорить об устойчивости звезды к отказам компьютеров, то выход из строя периферийного компьютера или его сетевого оборудования никак не отражается на функционировании оставшейся части сети, зато любой отказ центрального компьютера делает сеть полностью неработоспособной. В связи с этим должны приниматься специальные меры по повышению надежности центрального компьютера и его сетевой аппаратуры.

Обрыв кабеля или короткое замыкание в нем при топологии звезда нарушает обмен только с одним компьютером, а все остальные компьютеры могут нормально продолжать работу.

В отличие от шины, в звезде на каждой линии связи находятся только два абонента: центральный и один из периферийных. Чаще всего для их соединения используется две линии связи, каждая из которых передает информацию в одном направлении, то есть на каждой линии связи имеется только один приемник и один передатчик. Это так называемая передача **точка-точка**. Все это существенно упрощает сетевое оборудование по сравнению с шиной и избавляет от необходимости применения дополнительных, внешних терминаторов.

В центре сети с данной топологией помещается не компьютер, а специальное устройство — концентратор или, как его еще называют, хаб (hub), которое выполняет ту же функцию, что и репитер, то есть восстанавливает приходящие сигналы и пересылает их во все другие линии связи.

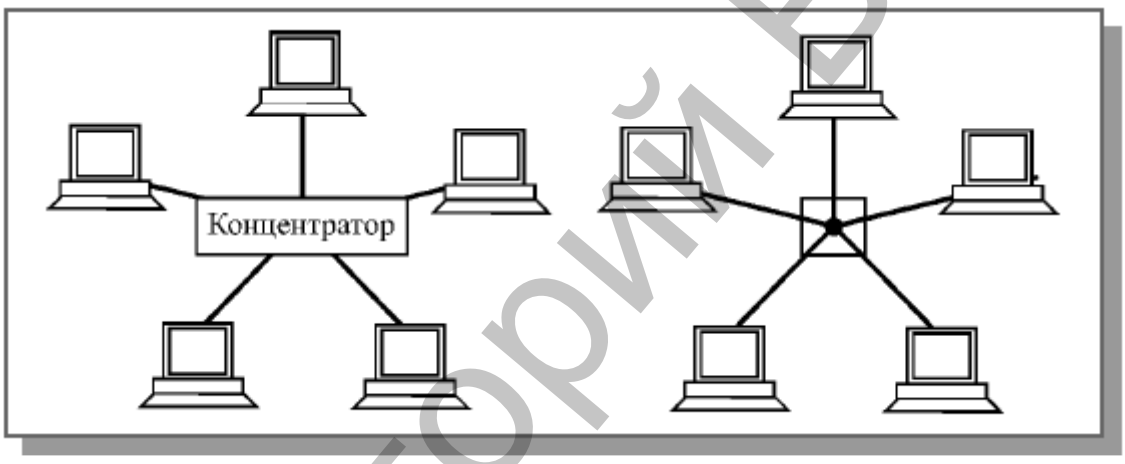


Рис. 1.11. Топология пассивная звезда и ее эквивалентная схема

Получается, что хотя схема прокладки кабелей подобна истинной или активной звезде, фактически речь идет о шинной топологии, так как информация от каждого компьютера одновременно передается ко всем остальным компьютерам, а никакого центрального абонента не существует. Безусловно, пассивная звезда дороже обычной шины, так как в этом случае требуется еще и концентратор. Однако она предоставляет целый ряд дополнительных возможностей, связанных с преимуществами звезды, в частности, упрощает обслуживание и ремонт сети. Именно поэтому в последнее время пассивная звезда все больше вытесняет истинную звезду, которая считается малоперспективной топологией.

Можно выделить также промежуточный тип топологии между активной и пассивной звездой. В этом случае концентратор не только ретранслирует поступающие на него сигналы, но и производит управление обменом, однако сам в обмене не участвует (так сделано в сети 100VG-AnyLAN).

Общим недостатком для всех топологий типа звезда (как активной, так и пассивной) является значительно больший, чем при других топологиях, расход кабеля. Например, если компьютеры расположены в одну линию (как на рис. 1.5), то при выборе топологии звезда понадобится в несколько раз

больше кабеля, чем при топологии шина. Это существенно влияет на стоимость сети в целом и заметно усложняет прокладку кабеля.

Топология кольцо

Кольцо — это топология, в которой каждый компьютер соединен линиями связи с двумя другими: от одного он получает информацию, а другому передает. На каждой линии связи, как и в случае звезды, работает только один передатчик и один приемник (связь типа точка-точка).

Четко выделенного центра при кольцевой топологии нет, все компьютеры могут быть одинаковыми и равноправными. Однако довольно часто в кольце выделяется специальный абонент, который управляет обменом или контролирует его. Понятно, что наличие такого единственного управляющего абонента снижает надежность сети, так как выход его из строя сразу же парализует весь обмен.

Строго говоря, компьютеры в кольце не являются полностью равноправными (в отличие, например, от шинной топологии). Ведь один из них обязательно получает информацию от компьютера, ведущего передачу в данный момент, раньше, а другие — позже. Именно на этой особенности топологии и строятся методы управления обменом по сети, специально рассчитанные на кольцо. В таких методах право на следующую передачу (или, как еще говорят, на захват сети) переходит последовательно к следующему по кругу компьютеру. Подключение новых абонентов в кольцо выполняется достаточно просто, хотя и требует обязательной остановки работы всей сети на время подключения. Как и в случае шины, максимальное количество абонентов в кольце может быть довольно велико (до тысячи и больше). Кольцевая топология обычно обладает высокой устойчивостью к перегрузкам, обеспечивает уверенную работу с большими потоками передаваемой по сети информации, так как в ней, как правило, нет конфликтов (в отличие от шины), а также отсутствует центральный абонент (в отличие от звезды), который может быть перегружен большими потоками информации.

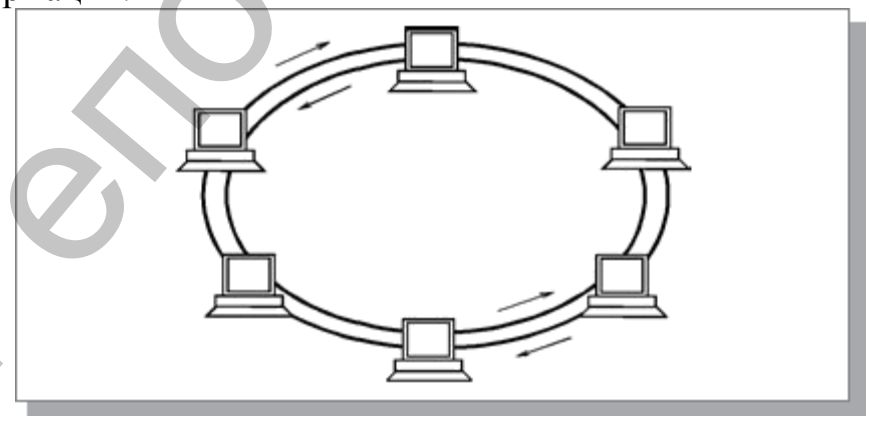


Рис. 1.12. Сеть с двумя кольцами

Сигнал в кольце проходит последовательно через все компьютеры сети, поэтому выход из строя хотя бы одного из них (или же его сетевого

оборудования) нарушает работу сети в целом. Это существенный недостаток кольца.

Точно так же обрыв или короткое замыкание в любом из кабелей кольца делает работу всей сети невозможной. Из трех рассмотренных топологий кольцо наиболее уязвимо к повреждениям кабеля, поэтому в случае топологии кольца обычно предусматривают прокладку двух (или более) параллельных линий связи, одна из которых находится в резерве.

Иногда сеть с топологией кольцо выполняется на основе двух параллельных кольцевых линий связи, передающих информацию в противоположных направлениях (рис. 1.12). Цель подобного решения — увеличение (в идеале — вдвое) скорости передачи информации по сети. К тому же при повреждении одного из кабелей сеть может работать с другим кабелем (правда, предельная скорость уменьшится).

Другие топологии

Кроме трех рассмотренных базовых топологий нередко применяется также сетевая топология дерево (tree), которую можно рассматривать как комбинацию нескольких звезд. Причем, как и в случае звезды, дерево может быть активным или истинным (рис. 1.13) и пассивным (рис. 1.14). При активном дереве в центрах объединения нескольких линий связи находятся центральные компьютеры, а при пассивном — концентраторы (хабы).

Лекция 5. Аппаратура локальных сетей

Аппаратура локальных сетей обеспечивает реальную связь между абонентами. Выбор аппаратуры имеет важнейшее значение на этапе проектирования сети, так как стоимость аппаратуры составляет наиболее существенную часть от стоимости сети в целом, а замена аппаратуры связана не только с дополнительными расходами, но зачастую и с трудоемкими работами. К аппаратуре локальных сетей относятся:

1. кабели для передачи информации;
2. разъемы для присоединения кабелей;
3. согласующие терминаторы;
4. сетевые адаптеры;
5. репитеры;
6. трансиверы;
7. концентраторы;
8. мосты;
9. маршрутизаторы;
10. шлюзы.

О первых трех компонентах сетевой аппаратуры уже говорилось в предыдущих главах. А сейчас следует остановиться на функциях остальных компонентов.

Средой передачи информации называются те линии связи (или каналы связи), по которым производится обмен информацией между компьютерами. В подавляющем большинстве компьютерных сетей (особенно локальных) используются проводные или кабельные каналы связи, хотя существуют и беспроводные сети, которые сейчас находят все более широкое применение, особенно в портативных компьютерах.

Информация в локальных сетях чаще всего передается в последовательном коде, то есть бит за битом. Такая передача медленнее и сложнее, чем при использовании параллельного кода. Однако надо учитывать то, что при более быстрой параллельной передаче (по нескольким кабелям одновременно) увеличивается количество соединительных кабелей в число раз, равное количеству разрядов параллельного кода (например, в 8 раз при 8-разрядном коде). Это совсем не мелочь, как может показаться на первый взгляд. При значительных расстояниях между абонентами сети стоимость кабеля вполне сравнима со стоимостью компьютеров и даже может превосходить ее. К тому же проложить один кабель (реже два разнонаправленных) гораздо проще, чем 8, 16 или 32. Значительно дешевле обойдется также поиск повреждений и ремонт кабеля.

Но это еще не все. Передача на большие расстояния при любом типе кабеля требует сложной передающей и приемной аппаратуры, так как при этом необходимо формировать мощный сигнал на передающем конце и детектировать слабый сигнал на приемном конце. При последовательной передаче для этого требуется всего один передатчик и один приемник. При

параллельной же количество требуемых передатчиков и приемников возрастает пропорционально разрядности используемого параллельного кода. В связи с этим, даже если разрабатывается сеть незначительной длины (порядка десятка метров) чаще всего выбирают последовательную передачу.

К тому же при параллельной передаче чрезвычайно важно, чтобы длины отдельных кабелей были точно равны друг другу. Иначе в результате прохождения по кабелям разной длины между сигналами на приемном конце образуется временной сдвиг, который может привести к сбоям в работе или даже к полной неработоспособности сети. Например, при скорости передачи 100 Мбит/с и длительности бита 10 нс этот временной сдвиг не должен превышать 5—10 нс. Такую величину сдвига дает разница в длинах кабелей в 1—2 метра. При длине кабеля 1000 метров это составляет 0,1—0,2%.

Надо отметить, что в некоторых высокоскоростных локальных сетях все-таки используют параллельную передачу по 2—4 кабелям, что позволяет при заданной скорости передачи применять более дешевые кабели с меньшей *полосой пропускания*. Но допустимая длина кабелей при этом не превышает сотни метров. Примером может служить сегмент 100BASE-T4 сети Fast Ethernet.

Промышленностью выпускается огромное количество типов кабелей, например, только одна крупнейшая кабельная компания Belden предлагает более 2000 их наименований. Но все кабели можно разделить на три большие группы:

- электрические (медные) кабели на основе *витых пар* проводов (twisted pair), которые делятся на экранированные (shielded twisted pair, STP) и неэкранированные (unshielded twisted pair, UTP);
- электрические (медные) *коаксиальные кабели* (coaxial cable);
- *оптоволоконные кабели* (fiber optic).

Каждый тип кабеля имеет свои преимущества и недостатки, так что при выборе надо учитывать как особенности решаемой задачи, так и особенности конкретной сети, в том числе и используемую топологию.

Можно выделить следующие основные параметры кабелей, принципиально важные для использования в локальных сетях:

- *Полоса пропускания* кабеля (частотный диапазон сигналов, пропускаемых кабелем) и *затухание сигнала* в кабеле. Два этих параметра тесно связаны между собой, так как с ростом частоты сигнала растет затухание сигнала. Надо выбирать кабель, который на заданной частоте сигнала имеет приемлемое *затухание*. Или же надо выбирать частоту сигнала, на которой *затухание* еще приемлемо. *Затухание* измеряется в децибелах и пропорционально длине кабеля.
- **Помехозащищенность** кабеля и обеспечиваемая им **секретность** передачи информации. Эти два взаимосвязанных параметра показывают, как кабель взаимодействует с окружающей средой, то есть, как он реагирует на внешние помехи, и насколько просто прослушать информацию, передаваемую по кабелю.

- **Скорость распространения сигнала** по кабелю или, обратный параметр – *задержка сигнала* на метр длины кабеля. Этот параметр имеет принципиальное значение при выборе длины сети. Типичные величины скорости распространения сигнала – от 0,6 до 0,8 от скорости распространения света в вакууме. Соответственно типичные величины *задержек* – от 4 до 5 нс/м.
- Для электрических кабелей очень важна величина **волнового сопротивления** кабеля. Волновое сопротивление важно учитывать при согласовании кабеля для предотвращения отражения сигнала от концов кабеля. Волновое сопротивление зависит от формы и взаиморасположения проводников, от технологии изготовления и материала диэлектрика кабеля. Типичные значения волнового сопротивления – от 50 до 150 Ом.

В настоящее время действует следующие стандарты на кабели:

- EIA/TIA 568 (Commercial Building Telecommunications Cabling Standard) – американский;
- ISO/IEC IS 11801 (Generic cabling for customer premises) – международный;
- CENELEC EN 50173 (Generic cabling systems) – европейский.

Эти стандарты описывают практически одинаковые кабельные системы, но отличаются терминологией и нормами на параметры. В данной работе предлагается придерживаться терминологии стандарта EIA/TIA 568.

Кабели на основе витых пар

Витые пары проводов используются в дешевых и сегодня, пожалуй, самых популярных кабелях. Кабель на основе ***витых пар*** представляет собой несколько пар скрученных попарно изолированных медных проводов в единой диэлектрической (пластиковой) оболочке. Он довольно гибкий и удобный для прокладки. Скручивание проводов позволяет свести к минимуму индуктивные наводки кабелей друг на друга и снизить влияние переходных процессов.

Обычно в кабель входит две (рис. 2.1) или четыре *витые пары*.

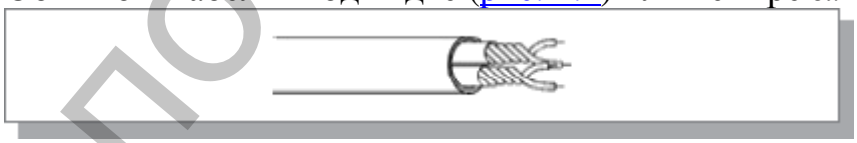


Рис. 2.1. Кабель с витыми парами

Неэкранированные *витые пары* характеризуются слабой защищенностью от внешних электромагнитных помех, а также от подслушивания, которое может осуществляться с целью, например, промышленного шпионажа. Причем перехват передаваемой по сети информации возможен как с помощью контактного метода (например, посредством двух иголок, воткнутых в кабель), так и с помощью бесконтактного метода, сводящегося к радиоперехвату излучаемых кабелем электромагнитных полей. Причем действие помех и величина излучения во вне увеличивается с ростом длины кабеля. Для устранения этих недостатков применяется экранирование кабелей.

В случае экранированной *витой пары* STP каждая из *витых пар* помещается в металлическую оплетку-экран для уменьшения излучений кабеля, защиты от внешних электромагнитных помех и снижения взаимного влияния пар проводов друг на друга (crosstalk – перекрестные наводки). Для того чтобы экран защищал от помех, он должен быть обязательно заземлен. Естественно, экранированная *витая пара* заметно дороже, чем неэкранированная. Ее использование требует специальных экранированных разъемов. Поэтому встречается она значительно реже, чем неэкранированная *витая пара*.

Основные достоинства неэкранированных *витых пар* – простота монтажа разъемов на концах кабеля, а также ремонта любых повреждений по сравнению с другими типами кабеля. Все остальные характеристики у них хуже, чем у других кабелей. Например, при заданной скорости передачи *затухание сигнала* (уменьшение его уровня по мере прохождения по кабелю) у них больше, чем у *коаксиальных кабелей*. Если учесть еще низкую помехозащищенность, то понятно, почему линии связи на основе *витых пар*, как правило, довольно короткие (обычно в пределах 100 метров). В настоящее время *витая пара* используется для передачи информации на скоростях до 1000 Мбит/с, хотя технические проблемы, возникающие при таких скоростях крайне сложны.

Согласно стандарту EIA/TIA 568, существуют пять основных и две дополнительные категории кабелей на основе неэкранированной *витой пары* (UTP):

- Кабель категории 1 – это обычный телефонный кабель (пары проводов не витые), по которому можно передавать только речь. Этот тип кабеля имеет большой разброс параметров (волнового сопротивления, *полосы пропускания*, перекрестных наводок).
- Кабель категории 2 – это кабель из *витых пар* для передачи данных в полосе частот до 1 МГц. Кабель не тестируется на уровень перекрестных наводок. В настоящее время он используется очень редко. Стандарт EIA/TIA 568 не различает кабели категорий 1 и 2.
- Кабель категории 3 – это кабель для передачи данных в *полосе* частот до 16 МГц, состоящий из *витых пар* с девятью витками проводов на метр длины. Кабель тестируется на все параметры и имеет волновое сопротивление 100 Ом. Это самый простой тип кабелей, рекомендованный стандартом для локальных сетей. Еще недавно он был самым распространенным, но сейчас повсеместно вытесняется кабелем категории 5.
- Кабель категории 4 – это кабель, передающий данные в *полосе* частот до 20 МГц. Используется редко, так как не слишком заметно отличается от категории 3. Стандартом рекомендуется вместо кабеля категории 3 переходить сразу на кабель категории 5. Кабель категории 4 тестируется на все параметры и имеет волновое сопротивление 100 Ом. Кабель был создан для работы в сетях по стандарту IEEE 802.5.

- Кабель категории 5 – в настоящее время самый совершенный кабель, рассчитанный на передачу данных в *полосе* частот до 100 МГц. Состоит из *витых пар*, имеющих не менее 27 витков на метр длины (8 витков на фут). Кабель тестируется на все параметры и имеет волновое сопротивление 100 Ом. Рекомендуется применять его в современных высокоскоростных сетях типа Fast Ethernet и TPFDDI. Кабель категории 5 примерно на 30—50% дороже, чем кабель категории 3.
- Кабель категории 6 – перспективный тип кабеля для передачи данных в *полосе* частот до 200 (или 250) МГц.
- Кабель категории 7 – перспективный тип кабеля для передачи данных в *полосе* частот до 600 МГц.

Согласно стандарту EIA/TIA 568, полное волновое сопротивление наиболее совершенных кабелей категорий 3, 4 и 5 должно составлять 100 Ом ±15% в частотном диапазоне от 1 МГц до максимальной частоты кабеля. Требования не очень жесткие: величина волнового сопротивления может находиться в диапазоне от 85 до 115 Ом. Здесь же следует отметить, что волновое сопротивление экранированной *витой пары* STP по стандарту должно быть равным 150 Ом ±15%. Для согласования сопротивлений кабеля и оборудования в случае их несовпадения применяют согласующие трансформаторы (Balun). Существует также экранированная *витая пара* с волновым сопротивлением 100 Ом, но используется она довольно редко.

Второй важнейший параметр, задаваемый стандартом, – это максимальное *затухание сигнала*, передаваемого по кабелю, на разных частотах. В таблице 2.1 приведены предельные значения величины *затухания* в децибелах для кабелей категорий 3, 4 и 5 на расстояние 1000 футов (то есть 305 метров) при нормальной температуре окружающей среды 20°С.

Таблица 2.1. Максимальное затухание в кабелях

Частота, МГц	Максимальное затухание, дБ		
	Категория 3	Категория 4	Категория 5
0,064	2,8	2,3	2,2
0,256	4,0	3,4	3,2
0,512	5,6	4,6	4,5
0,772	6,8	5,7	5,5
1,0	7,8	6,5	6,3
4,0	17	13	13
8,0	26	19	18
10,0	30	22	20
16,0	40	27	25
20,0	—	31	28
25,0	—	—	32
31,25	—	—	36

62,5	—	—	52
100	—	—	67

Из таблицы видно, что величины *затухания* на частотах, близких к предельным, для всех кабелей очень значительны. Даже на небольших расстояниях сигнал ослабляется в десятки и сотни раз, что предъявляет высокие требования к приемникам сигнала.

Еще один специфический параметр, определяемый стандартом, это величина так называемой перекрестной наводки на ближнем конце (NEXT – Near End CrossTalk). Он характеризует влияние разных проводов в кабеле друг на друга. Суть данного параметра иллюстрируется на рис. 2.2. Сигнал, передаваемый по одной из *витых пар* кабеля (верхняя пара), наводит индуктивную помеху на другую (нижнюю) *витую пару* кабеля. Две *витые пары* в сети обычно передают информацию в разные стороны, поэтому наиболее важна наводка на ближнем конце воспринимающей пары (нижней на рисунке), так как именно там находится приемник информации. Перекрестная наводка на дальнем конце (FEXT – Far End CrossTalk) не имеет такого большого значения.

Таблица 2.2. Допустимые уровни перекрестных наводок NEXT

Частота, МГц	Перекрестная наводка на ближнем конце, дБ		
	Категория 3	Категория 4	Категория 5
0,150	- 54	-68	-74
0,772	-43	-58	-64
1,0	-41	-56	-62
4,0	-32	-47	-53
8,0	-28	-42	-48
10,0	-26	-41	-47
16,0	-23	-38	-44
20,0	—	-36	-42
25,0	—	—	-41
31,25	—	—	-40
62,5	—	—	-35
100,0	—	—	-32

В таблице 2.2 представлены значения допустимой перекрестной наводки на ближнем конце для кабелей категорий 3, 4 и 5 на различных частотах сигнала. Естественно, более качественные кабели обеспечивают меньшую величину перекрестной наводки.

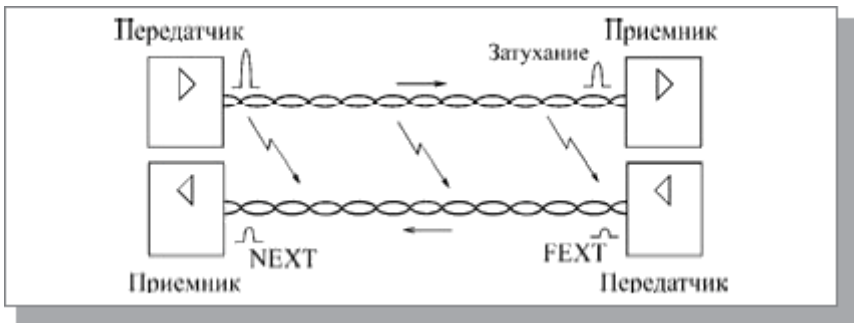


Рис. 2.2. Перекрестные помехи в кабелях на витых парах

Стандарт определяет также максимально допустимую величину рабочей емкости каждой из *витых пар* кабелей категории 4 и 5. Она должна составлять не более 17 нФ на 305 метров (1000 футов) при частоте сигнала 1 кГц и температуре окружающей среды 20°С.

Для присоединения *витых пар* используются разъемы (коннекторы) типа RJ-45, похожие на разъемы, используемые в телефонах (RJ-11), но несколько большие по размеру. Разъемы RJ-45 имеют восемь контактов вместо четырех в случае RJ-11. Присоединяются разъемы к кабелю с помощью специальных обжимных инструментов. При этом золоченые игольчатые контакты разъема прокалывают изоляцию каждого провода, входят между его жилами и обеспечивают надежное и качественное соединение. Надо учитывать, что при установке разъемов стандартом допускается расплетение *витой пары* кабеля на длину не более одного сантиметра.

Чаще всего *витые пары* используются для передачи данных в одном направлении (точка-точка), то есть в топологиях типа звезда или кольцо. Топология шина обычно ориентируется на *коаксиальный кабель*. Поэтому внешние терминаторы, согласующие неподключенные концы кабеля, для *витых пар* практически никогда не применяются.

Кабели выпускаются с двумя типами внешних оболочек:

- Кабель в поливинилхлоридной (ПВХ, PVC) оболочке дешевле и предназначен для работы в сравнительно комфортных условиях эксплуатации.
- Кабель в тефлоновой оболочке дороже и предназначен для более жестких условий эксплуатации.

Кабель в ПВХ оболочке называется еще non-plenum, а в тефлоновой – plenum. Термин plenum обозначает в данном случае пространство под фальшполом и над подвесным потолком, где удобно размещать кабели сети. Для прокладки в этих скрытых от глаз пространствах как раз удобнее кабель в тефлоновой оболочке, который, в частности, горит гораздо хуже, чем ПВХ – кабель, и не выделяет при этом ядовитых газов в большом количестве.

Еще один важный параметр любого кабеля, который жестко не определяется стандартом, но может существенно повлиять на работоспособность сети, – это скорость распространения сигнала в кабеле или, другими словами, *задержка распространения* сигнала в кабеле в расчете на единицу длины.

Производители кабелей иногда указывают величину ***задержки*** на метр

длины, а иногда – скорость распространения сигнала относительно скорости света (или NVP – Nominal Velocity of Propagation, как ее часто называют в документации). Связаны эти две величины простой формулой:

$t_з = 1/(3 \times 10^{10} \times NVP)$

где $t_з$ – величина *задержки* на метр длины кабеля в наносекундах. Например, если $NVP=0,65$ (65% от скорости света), то *задержка* $t_з$ будет равна 5,13 нс/м. Типичная величина *задержки* большинства современных кабелей составляет около 4—5 нс/м.

В таблице 2.3 приведены величины NVP и *задержек* на метр длины (в наносекундах) для некоторых типов кабеля двух самых известных компаний-производителей AT&T и Belden.

Таблица 2.3. Временные характеристики некоторых кабелей

Фирма	**Марка**	**Категория**	**Оболочка**	**NVP**	**Задержка**
AT&T	1010	3	non-plenum	0,67	4,98
AT&T	1041	4	non-plenum	0,70	4,76
AT&T	1061	5	non-plenum	0,70	4,76
AT&T	2010	3	plenum	0,70	4,76
AT&T	2041	4	plenum	0,75	4,44
AT&T	2061	5	plenum	0,75	4,44
Belden	1229A	3	non-plenum	0,69	4,83
Belden	1455A	4	non-plenum	0,72	4,63
Belden	1583A	5	non-plenum	0,72	4,63
Belden	1245A2	3	plenum	0,69	4,83
Belden	1457A	4	plenum	0,75	4,44
Belden	1585A	5	plenum	0,75	4,44

Стоит также отметить, что каждый из проводов, входящих в кабель на основе *витых пар*, как правило, имеет свой цвет изоляции, что существенно упрощает монтаж разъемов, особенно в том случае, когда концы кабеля находятся в разных комнатах, и контроль с помощью приборов затруднен.

Примером кабеля с экранированными *витыми парами* может служить кабель STP IBM типа 1, который включает в себя две экранированные *витые пары* AWG типа 22. Волновое сопротивление каждой пары составляет 150 Ом. Для этого кабеля применяются специальные разъемы, отличающиеся от разъемов для неэкранированной *витой пары* (например, DB9). Имеются и экранированные версии разъема RJ-45.

Коаксиальные кабели

Коаксиальный кабель представляет собой электрический кабель, состоящий из центрального медного провода и металлической оплетки (экрана), разделенных между собой слоем диэлектрика (внутренней изоляции) и помещенных в общую внешнюю оболочку (рис. 2.3).

Рис. 2.3. Коаксиальный кабель

Коаксиальный кабель до недавнего времени был очень популярен, что связано с его высокой помехозащищенностью (благодаря металлической оплетке), более широкими, чем в случае *витой пары*, *полосами пропускания* (свыше 1ГГц), а также большими допустимыми расстояниями передачи (до километра). К нему труднее механически подключиться для несанкционированного прослушивания сети, он дает также заметно меньше электромагнитных излучений вовне. Однако монтаж и ремонт *коаксиального кабеля* существенно сложнее, чем *витой пары*, а стоимость его выше (он дороже примерно в 1,5 – 3 раза). Сложнее и установка разъемов на концах кабеля. Сейчас его применяют реже, чем *витую пару*. Стандарт EIA/TIA-568 включает в себя только один тип *коаксиального кабеля*, применяемый в сети Ethernet.

Основное применение *коаксиальный кабель* находит в сетях с топологией типа шина. При этом на концах кабеля обязательно должны устанавливаться терминаторы для предотвращения внутренних отражений сигнала, причем один (и только один!) из терминаторов должен быть заземлен. Без заземления металлическая оплетка не защищает сеть от внешних электромагнитных помех и не снижает излучение передаваемой по сети информации во внешнюю среду. Но при заземлении оплетки в двух или более точках из строя может выйти не только сетевое оборудование, но и компьютеры, подключенные к сети. Терминаторы должны быть обязательно согласованы с кабелем, необходимо, чтобы их сопротивление равнялось волновому сопротивлению кабеля. Например, если используется 50-омный кабель, для него подходят только 50-омные терминаторы.

Реже *коаксиальные кабели* применяются в сетях с топологией звезда (например, пассивная звезда в сети Arcnet). В этом случае проблема согласования существенно упрощается, так как внешних терминаторов на свободных концах не требуется.

Волновое сопротивление кабеля указывается в сопроводительной документации. Чаще всего в локальных сетях применяются 50-омные (RG-58, RG-11, RG-8) и 93-омные кабели (RG-62). Распространенные в телевизионной технике 75-омные кабели в локальных сетях не используются. Марок *коаксиального кабеля* немного. Он не считается особо перспективным. Не случайно в сети Fast Ethernet не предусмотрено применение *коаксиальных кабелей*. Однако во многих случаях классическая шинная топология (а не пассивная звезда) очень удобна. Как уже отмечалось, она не требует применения дополнительных устройств – концентраторов.

Существует два основных типа *коаксиального кабеля*:

- тонкий (thin) кабель, имеющий диаметр около 0,5 см, более гибкий;
- толстый (thick) кабель, диаметром около 1 см, значительно более жесткий. Он представляет собой классический вариант *коаксиального кабеля*, который уже почти полностью вытеснен современным тонким кабелем.

Тонкий кабель используется для передачи на меньшие расстояния, чем толстый, поскольку сигнал в нем затухает сильнее. Зато с тонким кабелем гораздо удобнее работать: его можно оперативно проложить к каждому компьютеру, а толстый требует жесткой фиксации на стене помещения. Подключение к тонкому кабелю (с помощью разъемов BNC байонетного типа) проще и не требует дополнительного оборудования. А для подключения к толстому кабелю надо использовать специальные довольно дорогие устройства, прокалывающие его оболочки и устанавливающие контакт как с центральной жилой, так и с экраном. Толстый кабель примерно вдвое дороже, чем тонкий, поэтому тонкий кабель применяется гораздо чаще.

Как и в случае *витых пар*, важным параметром *коаксиального кабеля* является тип его внешней оболочки. Точно так же в данном случае применяются как non-plenum (PVC), так и plenum кабели. Естественно, тефлоновый кабель дороже поливинилхлоридного. Обычно тип оболочки можно отличить по окраске (например, для PVC кабеля фирма Belden использует желтый цвет, а для тефлонового – оранжевый).

Типичные величины *задержки распространения* сигнала в *коаксиальном кабеле* составляют для тонкого кабеля около 5 нс/м, а для толстого – около 4,5 нс/м.

Существуют варианты *коаксиального кабеля* с двойным экраном (один экран расположен внутри другого и отделен от него дополнительным слоем изоляции). Такие кабели имеют лучшую помехозащищенность и защиту от прослушивания, но они немного дороже обычных.

В настоящее время считается, что *коаксиальный кабель* устарел, в большинстве случаев его вполне может заменить *витая пара* или *оптоволоконный кабель*. И новые стандарты на кабельные системы уже не включают его в перечень типов кабелей.

Оптоволоконные кабели

Оптоволоконный (он же волоконно-оптический) кабель – это принципиально иной тип кабеля по сравнению с рассмотренными двумя типами электрического или медного кабеля. Информация по нему передается не электрическим сигналом, а световым. Главный его элемент – это прозрачное стекловолокно, по которому свет проходит на огромные расстояния (до десятков километров) с незначительным ослаблением.

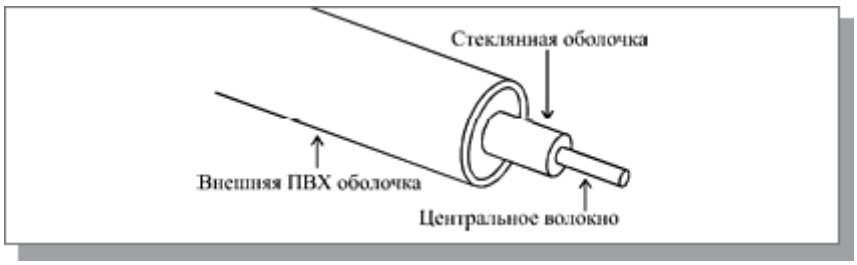


Рис. 2.4. Структура оптоволоконного кабеля

Структура *оптоволоконного кабеля* очень проста и похожа на структуру *коаксиального электрического кабеля* (рис. 2.4). Только вместо центрального медного провода здесь используется тонкое (диаметром около 1 – 10 мкм) стекловолокно, а вместо внутренней изоляции – стеклянная или пластиковая оболочка, не позволяющая свету выходить за пределы стекловолокна. В данном случае речь идет о режиме так называемого полного внутреннего отражения света от границы двух веществ с разными коэффициентами преломления (у стеклянной оболочки коэффициент преломления значительно ниже, чем у центрального волокна). Металлическая оплетка кабеля обычно отсутствует, так как экранирование от внешних электромагнитных помех здесь не требуется. Однако иногда ее все-таки применяют для механической защиты от окружающей среды (такой кабель иногда называют броневым, он может объединять под одной оболочкой несколько *оптоволоконных кабелей*).

Оптоволоконный кабель обладает исключительными характеристиками по помехозащищенности и секретности передаваемой информации. Никакие внешние электромагнитные помехи в принципе не способны исказить световой сигнал, а сам сигнал не порождает внешних электромагнитных излучений. Подключиться к этому типу кабеля для несанкционированного прослушивания сети практически невозможно, так как при этом нарушается целостность кабеля. Теоретически возможная *полоса пропускания* такого кабеля достигает величины 10^{12} Гц, то есть 1000 ГГц, что несравнимо выше, чем у электрических кабелей. Стоимость *оптоволоконного кабеля* постоянно снижается и сейчас примерно равна стоимости тонкого *коаксиального кабеля*.

Типичная величина *затухания сигнала* в *оптоволоконных кабелях* на частотах, используемых в локальных сетях, составляет от 5 до 20 дБ/км, что примерно соответствует показателям электрических кабелей на низких частотах. Но в случае *оптоволоконного кабеля* при росте частоты передаваемого сигнала *затухание* увеличивается очень незначительно, и на больших частотах (особенно свыше 200 МГц) его преимущества перед электрическим кабелем неоспоримы, у него просто нет конкурентов.

Однако *оптоволоконный кабель* имеет и некоторые недостатки.

Самый главный из них – высокая сложность монтажа (при установке разъемов необходима микронная точность, от точности скола стекловолокна и степени его полировки сильно зависит *затухание* в разъеме). Для установки разъемов применяют сварку или склеивание с помощью специального геля, имеющего такой же коэффициент преломления света, что и стекловолокно. В

любом случае для этого нужна высокая квалификация персонала и специальные инструменты. Поэтому чаще всего *оптоволоконный кабель* продается в виде заранее нарезанных кусков разной длины, на обоих концах которых уже установлены разъемы нужного типа. Следует помнить, что некачественная установка разъема резко снижает допустимую длину кабеля, определяемую *затуханием*.

Также надо помнить, что использование *оптоволоконного кабеля* требует специальных оптических приемников и передатчиков, преобразующих световые сигналы в электрические и обратно, что порой существенно увеличивает стоимость сети в целом.

Оптоволоконные кабели допускают разветвление сигналов (для этого производятся специальные пассивные **разветвители** (couplers) на 2—8 каналов), но, как правило, их используют для передачи данных только в одном направлении между одним передатчиком и одним приемником. Ведь любое разветвление неизбежно сильно ослабляет световой сигнал, и если разветвлений будет много, то свет может просто не дойти до конца сети. Кроме того, в разветвителе есть и внутренние потери, так что суммарная мощность сигнала на выходе меньше входной мощности.

Оптоволоконный кабель менее прочен и гибок, чем электрический. Типичная величина допустимого радиуса изгиба составляет около 10 – 20 см, при меньших радиусах изгиба центральное волокно может сломаться. Плохо переносит кабель и механическое растяжение, а также раздавливающие воздействия.

Чувствителен *оптоволоконный кабель* и к ионизирующим излучениям, из-за которых снижается прозрачность стекловолокна, то есть увеличивается *затухание сигнала*. Резкие перепады температуры также негативно сказываются на нем, стекловолокно может треснуть.

Применяют *оптоволоконный кабель* только в сетях с топологией звезда и кольцо. Никаких проблем согласования и заземления в данном случае не существует. Кабель обеспечивает идеальную гальваническую развязку компьютеров сети. В будущем этот тип кабеля, вероятно, вытеснит электрические кабели или, во всяком случае, сильно потеснит их. Запасы меди на планете истощаются, а сырья для производства стекла более чем достаточно.

Существуют два различных типа *оптоволоконного кабеля*:

- **многомодовый** или **мультимодовый** кабель, более дешевый, но менее качественный;
- **одномодовый** кабель, более дорогой, но имеет лучшие характеристики по сравнению с первым.

Суть различия между этими двумя типами сводится к разным режимам прохождения световых лучей в кабеле.

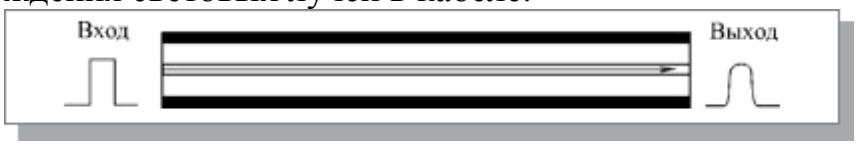


Рис. 2.5. Распространение света в одномодовом кабеле

В одномодовом кабеле практически все лучи проходят один и тот же путь, в результате чего они достигают приемника одновременно, и форма сигнала почти не искажается (рис. 2.5). Одномодовый кабель имеет диаметр центрального волокна около 1,3 мкм и передает свет только с такой же длиной волны (1,3 мкм). Дисперсия и потери сигнала при этом очень незначительны, что позволяет передавать сигналы на значительно большее расстояние, чем в случае применения многомодового кабеля. Для одномодового кабеля применяются лазерные приемопередатчики, использующие свет исключительно с требуемой длиной волны. Такие приемопередатчики пока еще сравнительно дороги и не долговечны. Однако в перспективе одномодовый кабель должен стать основным типом благодаря своим прекрасным характеристикам. К тому же лазеры имеют большее быстродействие, чем обычные светодиоды. *Затухание сигнала* в одномодовом кабеле составляет около 5 дБ/км и может быть даже снижено до 1 дБ/км.

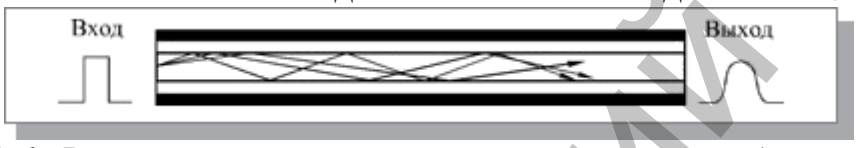


Рис. 2.6. Распространение света в многомодовом кабеле

В многомодовом кабеле траектории световых лучей имеют заметный разброс, в результате чего форма сигнала на приемном конце кабеля искажается (рис. 2.6). Центральное волокно имеет диаметр 62,5 мкм, а диаметр внешней оболочки 125 мкм (это иногда обозначается как 62,5/125). Для передачи используется обычный (не лазерный) светодиод, что снижает стоимость и увеличивает срок службы приемопередатчиков по сравнению с одномодовым кабелем. Длина волны света в многомодовом кабеле равна 0,85 мкм, при этом наблюдается разброс длин волн около 30 – 50 нм. Допустимая длина кабеля составляет 2 – 5 км. Многомодовый кабель – это основной тип *оптоволоконного кабеля* в настоящее время, так как он дешевле и доступнее. *Затухание* в многомодовом кабеле больше, чем в одномодовом и составляет 5 – 20 дБ/км.

Типичная величина *задержки* для наиболее распространенных кабелей составляет около 4—5 нс/м, что близко к величине *задержки* в электрических кабелях.

Оптоволоконные кабели, как и электрические, выпускаются в исполнении plenum и non-plenum.

Бескабельные каналы связи

Кроме кабельных каналов в компьютерных сетях иногда используются также бескабельные каналы. Их главное преимущество состоит в том, что не требуется никакой прокладки проводов (не надо делать отверстий в стенах, закреплять кабель в трубах и желобах, прокладывать его под фальшполами, над подвесными потолками или в вентиляционных шахтах, искать и устранять повреждения). К тому же компьютеры сети можно легко перемещать в пределах комнаты или здания, так как они ни к чему не

привязаны.

Радиоканал использует передачу информации по радиоволнам, поэтому теоретически он может обеспечить связь на многие десятки, сотни и даже тысячи километров. Скорость передачи достигает десятков мегабит в секунду (здесь многое зависит от выбранной длины волны и способа кодирования).

Особенность *радиоканала* состоит в том, что сигнал свободно излучается в эфир, он не замкнут в кабель, поэтому возникают проблемы совместимости с другими источниками радиоволн (радио- и телевещательными станциями, радарами, радиолюбительскими и профессиональными передатчиками и т.д.). В *радиоканале* используется передача в узком диапазоне частот и модуляция информационным сигналом сигнала несущей частоты.

Главным недостатком *радиоканала* является его плохая защита от прослушивания, так как радиоволны распространяются неконтролируемо. Другой большой недостаток *радиоканала* – слабая помехозащищенность.

Для локальных беспроводных сетей (WLAN – Wireless LAN) в настоящее время применяются подключения по *радиоканалу* на небольших расстояниях (обычно до 100 метров) и в пределах прямой видимости. Чаще всего используются два частотных диапазона – 2,4 ГГц и 5 ГГц. Скорость передачи – до 54 Мбит/с. Распространен вариант со скоростью 11 Мбит/с.

Сети WLAN позволяют устанавливать беспроводные сетевые соединения на ограниченной территории (обычно внутри офисного или университетского здания или в таких общественных местах, как аэропорты). Они могут использоваться во временных офисах или в других местах, где прокладка кабелей неосуществима, а также в качестве дополнения к имеющейся проводной локальной сети, призванного обеспечить пользователям возможность работать перемещаясь по зданию.

Популярная технология Wi-Fi (Wireless Fidelity) позволяет организовать связь между компьютерами числом от 2 до 15 с помощью концентратора (называемого точка доступа, Access Point, AP), или нескольких концентраторов, если компьютеров от 10 до 50. Кроме того, эта технология дает возможность связать две локальные сети на расстоянии до 25 километров с помощью мощных беспроводных мостов. Для примера на рис. 2.7 показано объединение компьютеров с помощью одной точки доступа. Важно, что многие мобильные компьютеры (ноутбуки) уже имеют встроенный контроллер Wi-Fi, что существенно упрощает их подключение к беспроводной сети.

Рис. 2.7. Объединение компьютеров с помощью технологии Wi-Fi

Радиоканал широко применяется в глобальных сетях как для наземной, так и для спутниковой связи. В этом применении у *радиоканала* нет конкурентов, так как радиоволны могут дойти до любой точки земного шара.

Инфракрасный канал также не требует соединительных проводов, так как использует для связи инфракрасное излучение (подобно пульту дистанционного управления домашнего телевизора). Главное его преимущество по сравнению с *радиоканалом* – нечувствительность к электромагнитным помехам, что позволяет применять его, например, в производственных условиях, где всегда много помех от силового оборудования. Правда, в данном случае требуется довольно высокая мощность передачи, чтобы не влияли никакие другие источники теплового (инфракрасного) излучения. Плохо работает инфракрасная связь и в условиях сильной запыленности воздуха.

Скорости передачи информации по инфракрасному каналу обычно не превышают 5—10 Мбит/с, но при использовании инфракрасных лазеров может быть достигнута скорость более 100 Мбит/с. Секретность передаваемой информации, как и в случае *радиоканала*, не достигается, также, требуются сравнительно дорогие приемники и передатчики. Все это приводит к тому, что применяют инфракрасные каналы в локальных сетях довольно редко. В основном они используются для связи компьютеров с периферией (интерфейс IrDA).

Инфракрасные каналы делятся на две группы:

- Каналы прямой видимости, в которых связь осуществляется на лучах, идущих непосредственно от передатчика к приемнику. При этом связь возможна только при отсутствии препятствий между компьютерами сети. Зато протяженность канала прямой видимости может достигать нескольких километров.
- Каналы на рассеянном излучении, которые работают на сигналах, отраженных от стен, потолка, пола и других препятствий. Препятствия в данном случае не помеха, но связь может осуществляться только в пределах одного помещения.

Если говорить о возможных топологиях, то наиболее естественно все беспроводные каналы связи подходят для топологии типа шина, в которой информация передается одновременно всем абонентам. Но при использовании узконаправленной передачи и/или частотного разделения по каналам можно реализовать любые топологии (кольцо, звезда, комбинированные топологии) как на *радиоканале*, так и на инфракрасном канале.

Сетевые адаптеры (они же контроллеры, карты, платы, интерфейсы, NIC – Network Interface Card) – это основная часть аппаратуры локальной сети. Назначение сетевого адаптера – сопряжение компьютера (или другого абонента) с сетью, то есть обеспечение обмена информацией между компьютером и каналом связи в соответствии с принятыми правилами

обмена. Именно они реализуют функции двух нижних уровней модели OSI. Как правило, сетевые адаптеры выполняются в виде платы (рис. 5.5), вставляемой в слоты расширения системной магистрали (шины) компьютера (чаще всего PCI, ISA или PC-Card). Плата сетевого адаптера обычно имеет также один или несколько внешних разъемов для подключения к ней кабеля сети.

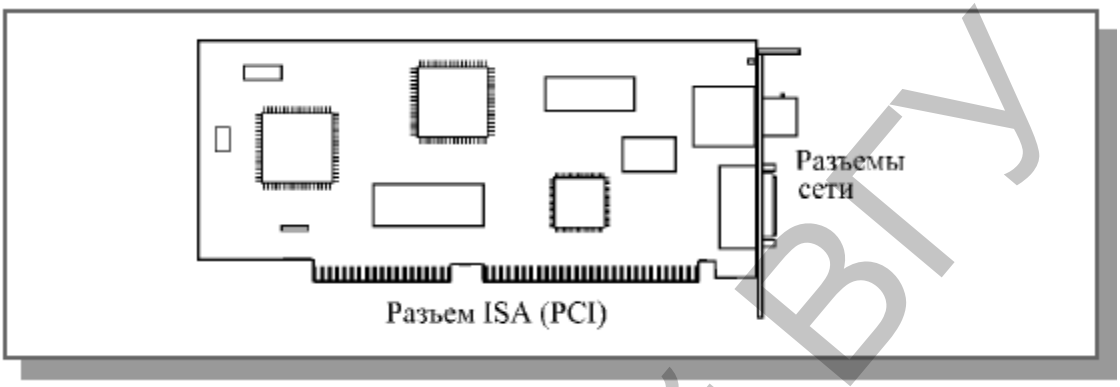


Рис. 5.5. Плата сетевого адаптера

Например, сетевые адаптеры Ethernet могут выпускаться со следующими наборами разъемов:

- TPO – разъем RJ-45 (для кабеля на витых парах по стандарту 10BASE-T).
- TPC – разъемы RJ-45 (для кабеля на витых парах 10BASE-T) и BNC (для коаксиального кабеля 10BASE2).
- FL – разъем ST (для оптоволоконного кабеля 10BASE-FL).

Функции сетевого адаптера делятся на магистральные и сетевые. К магистральным относятся те функции, которые осуществляют взаимодействие адаптера с магистралью (системной шиной) компьютера (то есть опознание своего магистрального адреса, пересылка данных в компьютер и из компьютера, выработка сигнала прерывания процессора и т.д.). Сетевые функции обеспечивают общение адаптера с сетью.

К основным сетевым функциям адаптеров относятся:

- гальваническая развязка компьютера и кабеля локальной сети (для этого обычно используется передача сигналов через импульсные трансформаторы);
- преобразование логических сигналов в сетевые (электрические или световые) и обратно;
- кодирование и декодирование сетевых сигналов, то есть прямое и обратное преобразование сетевых кодов передачи информации (например, манчестерский код);
- опознание принимаемых пакетов (выбор из всех приходящих пакетов тех, которые адресованы данному абоненту или всем абонентам сети одновременно);
- преобразование параллельного кода в последовательный при передаче и обратное преобразование при приеме;

• буферизация передаваемой и принимаемой информации в буферной памяти адаптера;

• организация доступа к сети в соответствии с принятым методом управления обменом;

• подсчет контрольной суммы пакетов при передаче и приеме.

Некоторые адаптеры позволяют реализовать функцию удаленной загрузки, то есть поддерживать работу в сети бездисковых компьютеров, загружающих свою операционную систему прямо из сети. Для этого в состав таких адаптеров включается постоянная память с соответствующей программой загрузки. Правда, не все сетевые программные средства поддерживают данный режим работы.

Все остальные аппаратные средства локальных сетей (кроме адаптеров) имеют вспомогательный характер, и без них часто можно обойтись. Это сетевые промежуточные устройства.

Трансиверы или приемопередатчики (от английского TRANsmitter + reCEIVER) служат для передачи информации между адаптером и кабелем сети или между двумя сегментами (частями) сети. Трансиверы усиливают сигналы, преобразуют их уровни или преобразуют сигналы в другую форму (например, из электрической в световую и обратно). Трансиверами также часто называют встроенные в адаптер приемопередатчики.

Репитеры или повторители (repeater) выполняют более простую функцию, чем трансиверы. Они не преобразуют ни уровни сигналов, ни их физическую природу, а только восстанавливают ослабленные сигналы (их амплитуду и форму), приводя их к исходному виду. Цель такой ретрансляции сигналов состоит исключительно в увеличении длины сети (рис. 5.7).

Концентраторы (хабы, hub), как следует из их названия, служат для объединения в сеть нескольких сегментов. Концентраторы (или репитерные концентраторы) представляют собой несколько собранных в едином конструктиве репитеров, они выполняют те же функции, что и репитеры (рис. 5.8).

Преимущество подобных концентраторов по сравнению с отдельными репитерами в том, что все точки подключения собраны в одном месте, это упрощает реконфигурацию сети, контроль и поиск неисправностей. К тому же все репитеры в данном случае питаются от единого качественного источника питания.

Концентраторы иногда вмешиваются в обмен, помогая устранять некоторые явные ошибки обмена. В любом случае они работают на первом уровне модели OSI, так как имеют дело только с физическими сигналами, с битами пакета и не анализируют содержимое пакета, рассматривая пакет как единое целое (рис. 5.9). На первом же уровне работают и трансиверы, и репитеры.

Выпускаются также совсем простые концентраторы, которые соединяют сегменты сети без восстановления формы сигналов. Они не увеличивают длину сети.

Коммутаторы (свичи, коммутирующие концентраторы, switch), как и концентраторы, служат для соединения сегментов в сеть. Они также выполняют более сложные функции, производя сортировку поступающих на них пакетов.

Коммутаторы передают из одного сегмента сети в другой не все поступающие на них пакеты, а только те, которые адресованы компьютерам из другого сегмента. Пакеты, передаваемые между абонентами одного сегмента, через коммутатор не проходят. При этом сам пакет коммутатором не принимается, а только пересылается. Интенсивность обмена в сети снижается вследствие разделения нагрузки, поскольку каждый сегмент работает не только со своими пакетами, но и с пакетами, пришедшими из других сегментов.

Коммутатор работает на втором уровне модели OSI (подуровень MAC), так как анализирует MAC-адреса внутри пакета (рис. 5.10). Естественно, он выполняет и функции первого уровня. В последнее время объем выпуска коммутаторов сильно вырос, цена на них упала, поэтому коммутаторы постепенно вытесняют концентраторы.

Мосты (bridge), маршрутизаторы (router) и шлюзы (gateway) служат для объединения в одну сеть несколько разнородных сетей с разными протоколами обмена нижнего уровня, в частности, с разными форматами пакетов, методами кодирования, скоростью передачи и т.д. В результате их применения сложная и неоднородная сеть, содержащая в себе различные сегменты, с точки зрения пользователя выглядит самой обычной сетью. Обеспечивается прозрачность сети для протоколов высокого уровня. Все они гораздо дороже, чем концентраторы, так как от них требуется довольно сложная обработка информации. Реализуются они обычно на базе компьютеров, подключенных к сети с помощью сетевых адаптеров. По сути, они представляют собой специализированные абоненты (узлы) сети.

Мосты – наиболее простые устройства, служащие для объединения сетей с разными стандартами обмена, например, Ethernet и Arcnet, или нескольких сегментов (частей) одной и той же сети, например, Ethernet (рис. 5.11). В последнем случае мост, как и коммутатор, только разделяет нагрузку сегментов, повышая тем самым производительность сети в целом. В отличие от коммутаторов мосты принимают поступающие пакеты целиком и в случае необходимости производят их простейшую обработку. Мосты, как и коммутаторы, работают на втором уровне модели OSI (рис. 5.10), но в отличие от них могут захватывать также и верхний подуровень LLC второго уровня (для связи разнородных сетей). В последнее время мосты быстро вытесняются коммутаторами, которые становятся более функциональными.

Маршрутизаторы осуществляют выбор оптимального маршрута для каждого пакета с целью избежания чрезмерной нагрузки отдельных участков сети и обхода поврежденных участков. Они применяются, как правило, в сложных разветвленных сетях, имеющих несколько маршрутов между отдельными абонентами. Маршрутизаторы не преобразуют протоколы

нижних уровней, поэтому они соединяют только сегменты одноименных сетей.

Маршрутизаторы работают на третьем уровне модели OSI, так как они анализируют не только MAC-адреса пакета, но и IP-адреса, то есть более глубоко проникают в инкапсулированный пакет

Существуют также гибридные маршрутизаторы (brouter), представляющие собой гибрид моста и маршрутизатора. Они выделяют пакеты, которым нужна маршрутизация и обрабатывают их как маршрутизатор, а для остальных пакетов служат обычным мостом.

Шлюзы – это устройства для соединения сетей с сильно отличающимися протоколами, например, для соединения локальных сетей с большими компьютерами или с глобальными сетями. Это самые дорогие и редко применяемые сетевые устройства. Шлюзы реализуют связь между абонентами на верхних уровнях модели OSI (с четвертого по седьмой). Соответственно, они должны выполнять и все функции нижестоящих уровней.

Лекция 6. Пакеты, протоколы, методы управления и обменом данных.

Информация в локальных сетях, как правило, передается отдельными порциями, кусками, называемыми в различных источниках пакетами (packets), кадрами (frames) или блоками. Причем предельная длина этих пакетов строго ограничена (обычно величиной в несколько килобайт). Ограничена длина пакета и снизу (как правило, несколькими десятками байт). Выбор пакетной передачи связан с несколькими важными соображениями.

Время доступа к сети (access time), которое определяется как временной интервал между моментом готовности абонента к передаче (когда ему есть, что передавать) и моментом начала этой передачи. Это время ожидания абонентом начала своей передачи. Естественно, оно не должно быть слишком большим, иначе величина реальной, интегральной скорости передачи информации между приложениями сильно уменьшится даже при высокоскоростной связи. Ожидание начала передачи связано с тем, что в сети не может происходить несколько передач одновременно (во всяком случае, при топологиях шина и кольцо). В связи с этим абоненты передают свою информацию по очереди. И каждому абоненту, прежде чем начать передачу, надо дождаться своей очереди. Вот это время ожидания своей очереди и есть время доступа.

Если бы вся требуемая информация передавалась каким-то абонентом сразу, непрерывно, без разделения на пакеты, то это привело бы к монопольному захвату сети этим абонентом на довольно продолжительное время. Все остальные абоненты вынуждены были бы ждать окончания передачи всей информации, что в ряде случаев могло бы потребовать десятков секунд и даже минут (например, при копировании содержимого целого жесткого диска). С тем чтобы уравнять в правах всех абонентов, а также сделать примерно одинаковыми для всех них величину времени доступа к сети и интегральную скорость передачи информации, как раз и применяются пакеты (кадры) ограниченной длины.

Дело в том, что каждый пакет помимо собственно данных, которые требуется передать, должен содержать некоторое количество служебной информации. Прежде всего, это адресная информация, которая определяет, от кого и кому передается данный пакет (как на почтовом конверте – адреса получателя и отправителя).

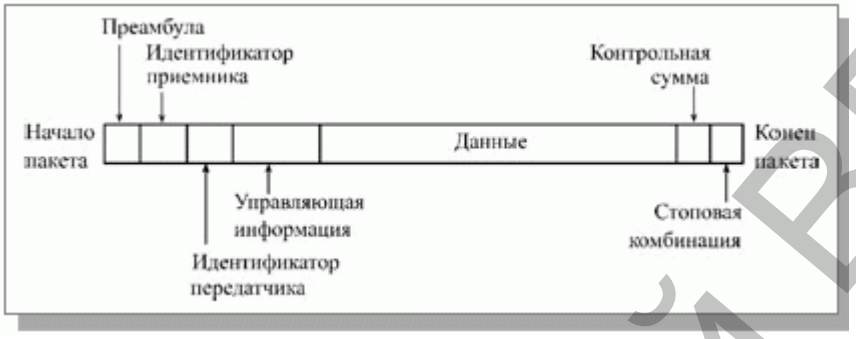
Структура и размеры пакета

в каждой сети жестко определены стандартом на данную сеть и связаны, прежде всего, с аппаратурными особенностями данной сети, выбранной топологией и типом среды передачи информации. Кроме того, эти параметры зависят от используемого протокола (порядка обмена информацией).

Но существуют некоторые общие принципы формирования структуры пакета, которые учитывают характерные особенности обмена информацией по любым локальным сетям.

Чаще всего пакет содержит в себе следующие основные поля или части (рис. 4.3):

Стартовая комбинация битов или преамбула, которая обеспечивает предварительную настройку аппаратуры адаптера или другого



• Сетевого устройства на прием и обработку пакета. Это поле может полностью отсутствовать или же сводиться к единственному стартовому биту.

• Сетевой адрес (идентификатор) принимающего абонента, то есть индивидуальный или групповой номер, присвоенный каждому принимающему абоненту в сети. Этот адрес позволяет приемнику распознать пакет, адресованный ему лично, группе, в которую он входит, или всем абонентам сети одновременно (при широком вещании).

• Сетевой адрес (идентификатор) передающего абонента, то есть индивидуальный номер, присвоенный каждому передающему абоненту. Этот адрес информирует принимающего абонента, откуда пришел данный пакет. Включение в пакет адреса передатчика необходимо в том случае, когда одному приемнику могут попеременно приходить пакеты от разных передатчиков.

• Служебная информация, которая может указывать на тип пакета, его номер, размер, формат, маршрут его доставки, на то, что с ним надо делать приемнику и т.д.

• Данные (поле данных) – это та информация, ради передачи которой используется пакет. В отличие от всех остальных полей пакета поле данных имеет переменную длину, которая, собственно, и определяет полную длину пакета. Существуют специальные управляющие пакеты, которые не имеют поля данных. Их можно рассматривать как сетевые команды. Пакеты, включающие поле данных, называются информационными пакетами. Управляющие пакеты могут выполнять функцию начала и конца сеанса связи, подтверждения приема информационного пакета, запроса информационного пакета и т.д.

• Контрольная сумма пакета – это числовой код, формируемый передатчиком по определенным правилам и содержащий в свернутом виде

информацию обо всем пакете. Приемник, повторяя вычисления, сделанные передатчиком, с принятым пакетом, сравнивает их результат с контрольной суммой и делает вывод о правильности или ошибочности передачи пакета. Если пакет ошибочен, то приемник запрашивает его повторную передачу. Обычно используется циклическая контрольная сумма (CRC). Подробнее об этом рассказано в главе 7.

• Стоповая комбинация служит для информирования аппаратуры принимающего абонента об окончании пакета, обеспечивает выход аппаратуры приемника из состояния приема. Это поле может отсутствовать, если используется самосинхронизирующийся код, позволяющий определять момент окончания передачи пакета.

Сеанс обмена

В процессе сеанса обмена информацией по сети между передающим и принимающим абонентами происходит обмен информационными и управляющими пакетами по установленным правилам, называемым протоколом обмена. Это позволяет обеспечить надежную передачу информации при любой интенсивности обмена по сети.

Сеанс обмена начинается с запроса передатчиком готовности приемника принять данные. Для этого используется управляющий пакет "Запрос". Если приемник не готов, он отказывается от сеанса специальным управляющим пакетом. В случае, когда приемник готов, он посылает в ответ управляющий пакет "Готовность". Затем начинается собственно передача данных. При этом на каждый полученный информационный пакет приемник отвечает управляющим пакетом "Подтверждение". В случае, когда пакет данных передан с ошибками, в ответ на него приемник запрашивает повторную передачу. Заканчивается сеанс управляющим пакетом "Конец", которым передатчик сообщает о разрыве связи. Существует множество стандартных протоколов, которые используют как передачу с подтверждением (с гарантированной доставкой пакета), так и передачу без подтверждения (без гарантии доставки пакета).

Адресация пакетов

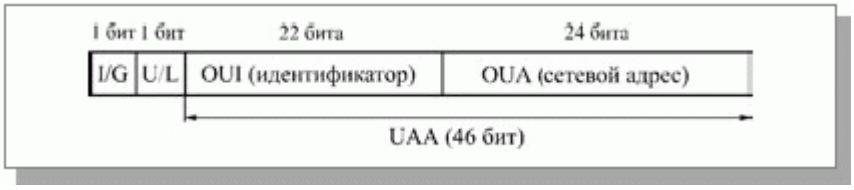
Каждый абонент (узел) локальной сети должен иметь свой уникальный адрес (идентификатор или MAC-адрес), для того чтобы ему можно было адресовать пакеты. Существуют две основные системы присвоения адресов абонентам сети (точнее, сетевым адаптерам этих абонентов).

Первая система сводится к тому, что при установке сети каждому абоненту пользователь присваивает индивидуальный адрес по порядку, к примеру, от 0 до 30 или от 0 до 254. Присваивание адресов производится программно или с помощью переключателей на плате адаптера. Например, восемь разрядов адреса достаточно для сети из 255 абонентов. Один адрес (обычно 1111....11) отводится для широковещательной передачи, то есть он используется для пакетов, адресованных всем абонентам одновременно.Именно такой подход применен в известной сети Arcnet.

Второй подход к адресации был разработан международной организацией IEEE, занимающейся стандартизацией сетей. Именно он используется в большинстве сетей и рекомендован для новых разработок. Идея этого подхода состоит в том, чтобы присваивать уникальный сетевой адрес каждому адаптеру сети еще на этапе его изготовления. Если количество возможных адресов будет достаточно большим, то можно быть уверенным, что в любой сети по всему миру никогда не будет абонентов с одинаковыми адресами. Поэтому был выбран 48-битный формат адреса, что соответствует примерно 280 триллионам различных адресов. Понятно, что столько сетевых адаптеров никогда не будет выпущено.

С тем чтобы распределить возможные диапазоны адресов между многочисленными изготовителями сетевых адаптеров, была предложена следующая структура адреса (рис. 4.7):

- Младшие 24 разряда кода адреса называются OUA (Organizationally Unique Address) – организационно уникальный адрес. Именно их присваивает каждый из зарегистрированных производителей сетевых адаптеров. Всего возможно свыше 16 миллионов комбинаций, то есть каждый изготовитель может выпустить 16 миллионов сетевых адаптеров.
- Следующие 22 разряда кода называются OUI (Organizationally Unique Identifier) – организационно уникальный идентификатор. IEEE присваивает один или несколько OUI каждому производителю сетевых адаптеров. Это позволяет исключить совпадения адресов адаптеров от разных производителей. Всего возможно свыше 4 миллионов разных OUI, это означает, что теоретически может быть зарегистрировано 4 миллиона производителей. Вместе OUA и OUI называются UAA (Universally Administered Address) – универсально управляемый адрес или IEEE-адрес.
- Два старших разряда адреса управляющие, они определяют тип адреса, способ интерпретации остальных 46 разрядов. Старший бит I/G (Individual/Group) указывает на тип адреса. Если он установлен в 0, то индивидуальный, если в 1, то групповой (многопунктовый или функциональный). Пакеты с групповым адресом получат все имеющие этот групповой адрес сетевые адаптеры. Причем групповой адрес определяется 46 младшими разрядами. Второй управляющий бит U/L (Universal/Local) называется флажком универсального/местного управления и определяет, как был присвоен адрес данному сетевому адаптеру. Обычно он установлен в 0. Установка бита U/L в 1 означает, что адрес задан не производителем сетевого адаптера, а организацией, использующей данную сеть. Это случается довольно редко.



Для широковещательной передачи (то есть передачи всем абонентам сети одновременно) применяется специально выделенный сетевой адрес, все

48 битов которого установлены в единицу. Его принимают все абоненты сети независимо от их индивидуальных и групповых адресов.

Данной системы адресов придерживаются такие популярные сети, как Ethernet, Fast Ethernet, Token-Ring, FDDI, 100VG-AnyLAN.

Методы управления обменом

Но, как уже отмечалось, по одному кабелю одновременно передавать два (или более) пакета нельзя, иначе может возникнуть конфликт (коллизия) который приведет к искажению либо потере обоих пакетов (или всех пакетов, участвующих в конфликте). Значит, надо каким-то образом установить очередность доступа к сети (захвата сети) всеми абонентами, желающими передавать.

В сети обязательно применяется тот или иной метод управления обменом (метод доступа, метод арбитража), разрешающий или предотвращающий конфликты между абонентами. От эффективности работы выбранного метода управления обменом зависит очень многое: скорость обмена информацией между компьютерами, нагрузочная способность сети (способность работать с различными интенсивностями обмена), время реакции сети на внешние события и т.д. Метод управления – это один из важнейших параметров сети.

Тип метода управления обменом во многом определяется особенностями топологии сети. Но в то же время он не привязан жестко к топологии, как нередко принято считать.

Методы управления обменом в локальных сетях делятся на две группы:

• Централизованные методы, в которых все управление обменом сосредоточенно в одном месте. Недостатки таких методов: неустойчивость к отказам центра, малая гибкость управления (центр обычно не может оперативно реагировать на все события в сети). Достоинство централизованных методов – отсутствие конфликтов, так как центр всегда предоставляет право на передачу только одному абоненту, и ему не с кем конфликтовать.

• Децентрализованные методы, в которых отсутствует центр управления. Всеми вопросами управления, в том числе предотвращением, обнаружением и разрешением конфликтов, занимаются все абоненты сети. Главные достоинства децентрализованных методов: высокая устойчивость к отказам и большая гибкость. Однако в данном случае возможны конфликты, которые надо разрешать.

Существует и другое деление методов управления обменом, относящееся, главным образом, к децентрализованным методам:

• Детерминированные методы определяют четкие правила, по которым чередуются захватывающие сеть абоненты. Абоненты имеют определенную систему приоритетов, причем приоритеты эти различны для всех абонентов. При этом, как правило, конфликты полностью исключены (или маловероятны), но некоторые абоненты могут дожидаться своей очереди на передачу слишком долго. К детерминированным методам относится,

например, маркерный доступ (сети Token-Ring, FDDI), при котором право передачи передается по эстафете от абонента к абоненту.

• Случайные методы подразумевают случайное чередование передающих абонентов. При этом возможность конфликтов подразумевается, но предлагаются способы их разрешения. Случайные методы значительно хуже, чем детерминированные, работают при больших информационных потоках в сети (при большом трафике сети) и не гарантируют абоненту величину времени доступа. В то же время они обычно более устойчивы к отказам сетевого оборудования и более эффективно используют сеть при малой интенсивности обмена. Пример случайного метода – CSMA/CD (сеть Ethernet).

Для трех основных топологий характерны три наиболее типичных метода управления обменом.

Управление обменом в сети с топологией звезда

Для топологии звезда лучше всего подходит централизованный метод управления. Это связано с тем, что все информационные потоки проходят через центр, и именно этому центру логично доверить управление обменом в сети. Причем не так важно, что находится в центре звезды: компьютер (центральный абонент), как на рис. 1.6, или же специальный концентратор, управляющий обменом, но сам не участвующий в нем. В данном случае речь идет уже не о пассивной звезде (рис. 1.11), а о некой промежуточной ситуации, когда центр не является полноценным абонентом, но управляет обменом. Это, к примеру, реализовано в сети 100VG-AnyLAN.

Самый простейший централизованный метод состоит в следующем.

Периферийные абоненты, желающие передать свой пакет (или, как еще говорят, имеющие заявки на передачу), посылают центру свои запросы (управляющие пакеты или специальные сигналы). Центр же предоставляет им право передачи пакета в порядке очередности, например, по их физическому расположению в звезде по часовой стрелке. После окончания передачи пакета каким-то абонентом право передавать получит следующий по порядку (по часовой стрелке) абонент, имеющий заявку на передачу (рис. 4.8). Например, если передает второй абонент, то после него имеет право на передачу третий. Если же третьему абоненту не надо передавать, то право на передачу переходит к четвертому и т.д.

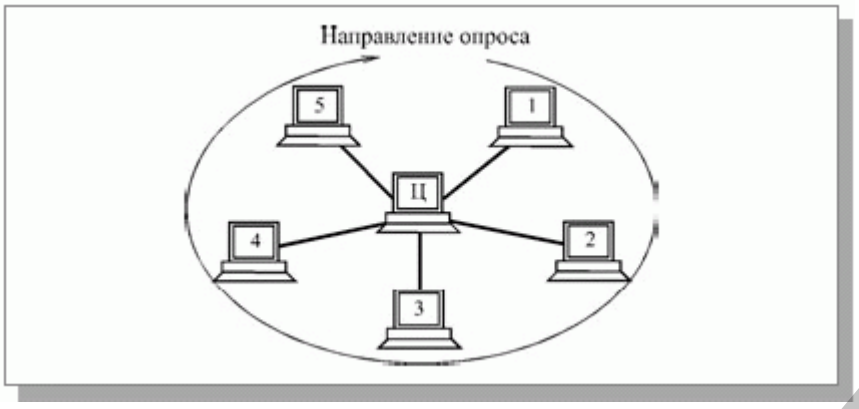


Рис. 4.8. Централизованный метод управления обменом в сети с топологией звезда

В этом случае говорят, что абоненты имеют географические приоритеты (по их физическому расположению). В каждый конкретный момент наивысшим приоритетом обладает следующий по порядку абонент, но в пределах полного цикла опроса ни один из абонентов не имеет никаких преимуществ перед другими. Никому не придется ждать своей очереди слишком долго. Максимальная величина времени доступа для любого абонента в этом случае будет равна суммарному времени передачи пакетов всех абонентов сети кроме данного. Для топологии, показанной на рис. 4.8, она составит четыре длительности пакета. Никаких столкновений пакетов при этом методе в принципе быть не может, так как все решения о доступе принимаются в одном месте.

Рассмотренный метод управления можно назвать методом с пассивным центром, так как центр пассивно прослушивает всех абонентов. Возможен и другой принцип реализации централизованного управления (его можно назвать методом с активным центром).

Управление обменом в сети с топологией шина

При топологии шина также возможно централизованное управление. При этом один из абонентов ("центральный") посылает по шине всем остальным ("периферийным") запросы (управляющие пакеты), выясняя, кто из них хочет передать, затем разрешает передачу одному из абонентов. Абонент, получивший право на передачу, по той же шине передает свой информационный пакет тому абоненту, которому хочет. А после окончания передачи передававший абонент все по той же шине сообщает "центру", что он закончил передачу (управляющим пакетом), и "центр" снова начинает опрос (рис. 4.9).

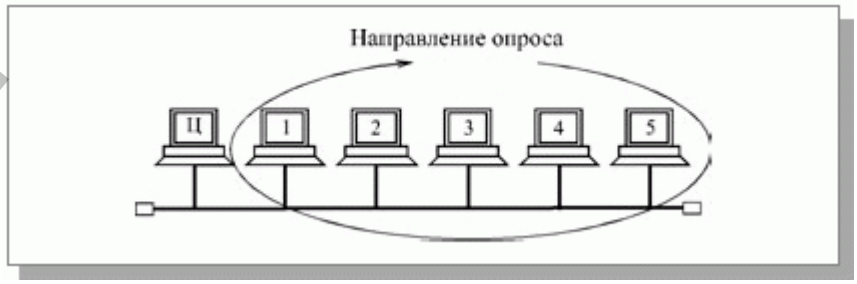


Рис. 4.9. Централизованное управление в сети с топологией шина

Преимущества и недостатки такого управления – те же самые, что и в случае централизованно управляемой звезды. Единственное отличие состоит в том, что центр здесь не пересылает информацию от одного абонента к другому, как в топологии активная звезда, а только управляет обменом.

Гораздо чаще в шине используется децентрализованное случайное управление, так как сетевые адаптеры всех абонентов в данном случае одинаковы, и именно этот метод наиболее органично подходит шине. При выборе децентрализованного управления все абоненты имеют равные права доступа к сети, то есть особенности топологии совпадают с особенностями метода управления. Решение о том, когда можно передавать свой пакет, принимается каждым абонентом на месте, исходя только из анализа состояния сети. В данном случае возникает конкуренция между абонентами за захват сети, и, следовательно, возможны конфликты между ними и искажения передаваемой информации из-за наложения пакетов.

Существует множество алгоритмов доступа или, как еще говорят, сценариев доступа, порой очень сложных. Их выбор зависит от скорости передачи в сети, длины шины, загруженности сети (интенсивности обмена или трафика сети), используемого кода передачи.

Иногда для управления доступом к шине применяется дополнительная линия связи, что позволяет упростить аппаратуру контроллеров и методы доступа, но заметно увеличивает стоимость сети за счет удвоения длины кабеля и количества приемопередатчиков. Поэтому данное решение не получило широкого распространения.

Суть всех случайных методов управления обменом довольно проста.

Если сеть свободна (то есть никто не передает своих пакетов), то абонент, желающий передавать, сразу начинает свою передачу. Время доступа в этом случае равно нулю.

Если же в момент возникновения у абонента заявки на передачу сеть занята, то абонент, желающий передавать, ждет освобождения сети. В противном случае исказятся и пропадут оба пакета. После освобождения сети абонент, желающий передавать, начинает свою передачу.

Возникновение конфликтных ситуаций (столкновений пакетов, коллизий), в результате которых передаваемая информация искажается, возможно в двух случаях.

- При одновременном начале передачи двумя или более абонентами, когда сеть свободна (рис. 4.10). Это ситуация довольно редкая, но все-таки вполне возможная.
- При одновременном начале передачи двумя или более абонентами сразу после освобождения сети (рис. 4.11). Это ситуация наиболее типична, так как за время передачи пакета одним абонентом вполне может возникнуть несколько новых заявок на передачу у других абонентов.

Существующие случайные методы управления обменом (арбитража) различаются тем, как они предотвращают возможные конфликты или же

разрешают уже возникшие. Ни один конфликт не должен нарушать обмен, все абоненты должны, в конце концов, передать свои пакеты.

В процессе развития локальных сетей было разработано несколько разновидностей случайных методов управления обменом.

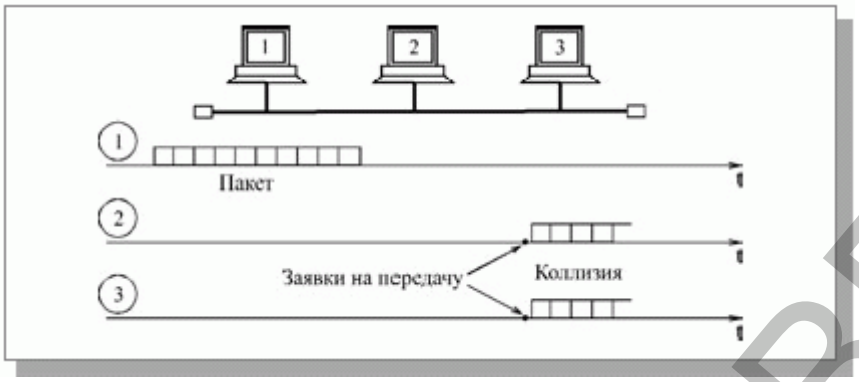


Рис. 4.10. Коллизии в случае начала передачи при свободной сети

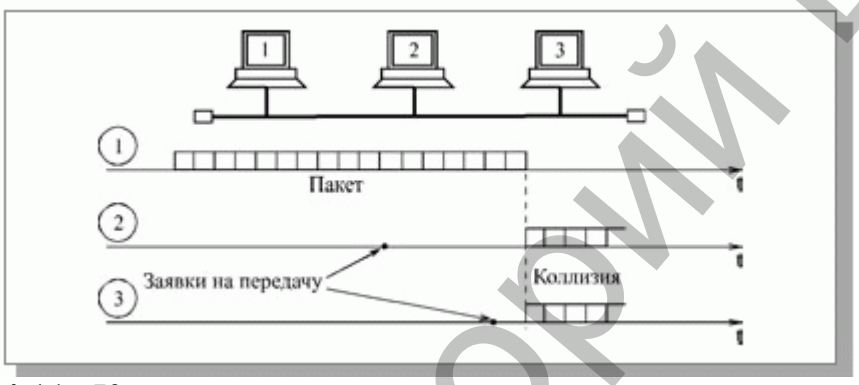


Рис. 4.11. Коллизии в случае начала передачи после освобождения сети

Например, был предложен метод, при котором не все передающие абоненты распознают коллизию, а только те, которые имеют меньшие приоритеты. Абонент с максимальным приоритетом из всех, начавших передачу, закончит передачу своего пакета без ошибок. Остальные, обнаружив коллизию, прекратят свою передачу и будут ждать освобождения сети для новой попытки. Для контроля коллизии каждый передающий абонент производит побитное сравнение передаваемой им в сеть информации и данных, присутствующих в сети. Побеждает тот абонент, заголовок пакета которого дольше других не искажается от коллизии. Этот метод, называемый децентрализованным кодовым приоритетным методом, отличается низким быстродействием и сложностью реализации.

При другом методе управления обменом каждый абонент начинает свою передачу после освобождения сети не сразу, а, выдержав свою, строго индивидуальную задержку, что предотвращает коллизии после освобождения сети и тем самым сводит к минимуму общее количество коллизий. Максимальным приоритетом в этом случае будет обладать абонент с минимальной задержкой. Столкновения пакетов возможны только тогда, когда два и более абонентов захотели передавать одновременно при свободной сети. Этот метод, называемый децентрализованным временным приоритетным методом, хорошо работает только в небольших сетях, так как каждому абоненту нужно обеспечить свою индивидуальную задержку.

В обоих случаях имеется система приоритетов, все же данные методы относятся к случайным, так как исход конкуренции невозможно предсказать. Случайные приоритетные методы ставят абонентов в неравные условия при большой интенсивности обмена по сети, так как высокоприоритетные абоненты могут надолго заблокировать сеть для низкоприоритетных абонентов.

Управление обменом в сети с топологией кольцо

Кольцевая топология имеет свои особенности при выборе метода управления обменом. В этом случае важно то, что любой пакет, посланный по кольцу, последовательно пройдя всех абонентов, через некоторое время возвратится в ту же точку, к тому же абоненту, который его передавал (так как топология замкнутая). Здесь нет одновременного распространения сигнала в две стороны, как в топологии шина. Как уже отмечалось, сети с топологией кольцо бывают однонаправленными и двунаправленными. Наиболее распространены однонаправленные.

В сети с топологией кольцо можно использовать различные централизованные методы управления (как в звезде), а также методы случайного доступа (как в шине), но чаще выбирают все-таки специфические методы управления, в наибольшей степени соответствующие особенностям кольца.

Самые популярные методы управления в кольцевых сетях маркерные (эстафетные), те, которые используют маркер (эстафету) – небольшой управляющий пакет специального вида. Именно эстафетная передача маркера по кольцу позволяет передавать право на захват сети от одного абонента к другому. Маркерные методы относятся к децентрализованным и детерминированным методам управления обменом в сети. В них нет явно выраженного центра, но существует четкая система приоритетов, и потому не бывает конфликтов.

Работа маркерного метода управления в сети с топологией кольцо представлена на рис. 4.15.

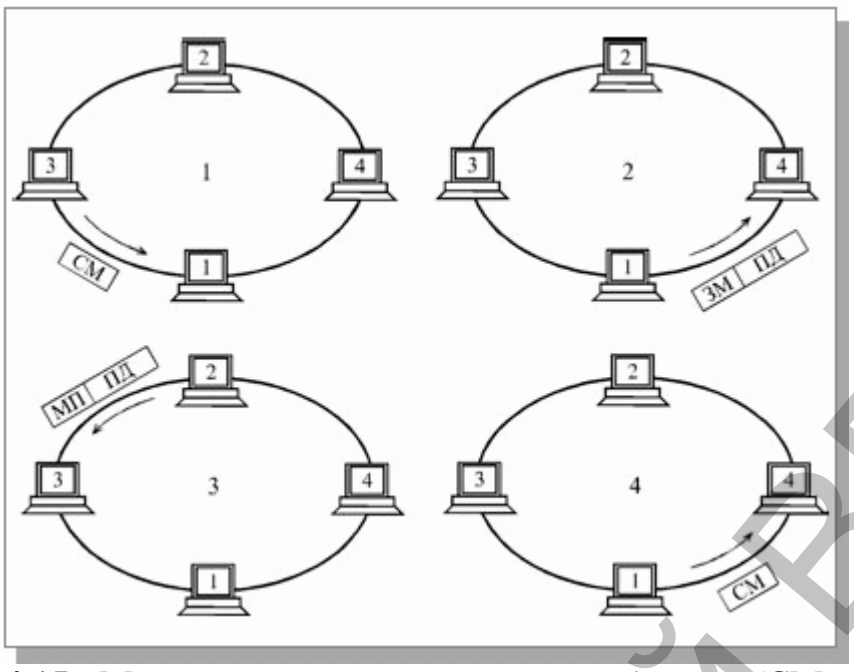


Рис. 4.15. Маркерный метод управления обменом (СМ—свободный маркер, ЗМ— занятый маркер, МП— занятый маркер с подтверждением, ПД—пакет данных)

По кольцу непрерывно ходит специальный управляющий пакет минимальной длины, маркер, предоставляющий абонентам право передавать свой пакет. Алгоритм действий абонентов:

1. Абонент 1, желающий передать свой пакет, должен дождаться прихода к нему свободного маркера. Затем он присоединяет к маркеру свой пакет, помечает маркер как занятый и отправляет эту посылку следующему по кольцу абоненту.

2. Все остальные абоненты (2, 3, 4), получив маркер с присоединенным пакетом, проверяют, им ли адресован пакет. Если пакет адресован не им, то они передают полученную посылку (маркер + пакет) дальше по кольцу.

3. Если какой-то абонент (в данном случае это абонент 3) распознает пакет как адресованный ему, то он его принимает, устанавливает в маркере бит подтверждения приема и передает посылку (маркер + пакет) дальше по кольцу.

4. Передававший абонент 1 получает свою посылку, прошедшую по всему кольцу, обратно, помечает маркер как свободный, удаляет из сети свой пакет и посылает свободный маркер дальше по кольцу. Абонент, желающий передавать, ждет этого маркера, и все повторяется снова.

Лекция 7. Модель OSI

Общая характеристика модели OSI

Из того, что протокол представляет собой соглашение, принятое двумя взаимодействующими объектами, в данном случае двумя работающими в сети компьютерами, совсем не следует, что он обязательно является стандартным. Но на практике при реализации сетей обычно используются стандартные протоколы. Это могут быть фирменные, национальные или международные стандарты.

В начале 80-х годов ряд международных организаций по стандартизации — ISO, ITU-T и некоторые другие — разработали модель, которая сыграла значительную роль в развитии сетей. Эта модель называется моделью ISO/OSI.

Модель взаимодействия открытых систем (Open System Interconnection, OSI) определяет различные уровни взаимодействия систем в сетях с коммутацией пакетов, дает им стандартные имена и указывает, какие функции должен выполнять каждый уровень.

Модель OSI была разработана на основании большого опыта, полученного при создании компьютерных сетей, в основном глобальных, в 70-е годы. Полное описание этой модели занимает более 1000 страниц текста.

В модели OSI (рис. 11.6) средства взаимодействия делятся на семь уровней: прикладной, представительный, сеансовый, транспортный, сетевой, канальный и физический . Каждый уровень имеет дело с определенным аспектом взаимодействия сетевых устройств.

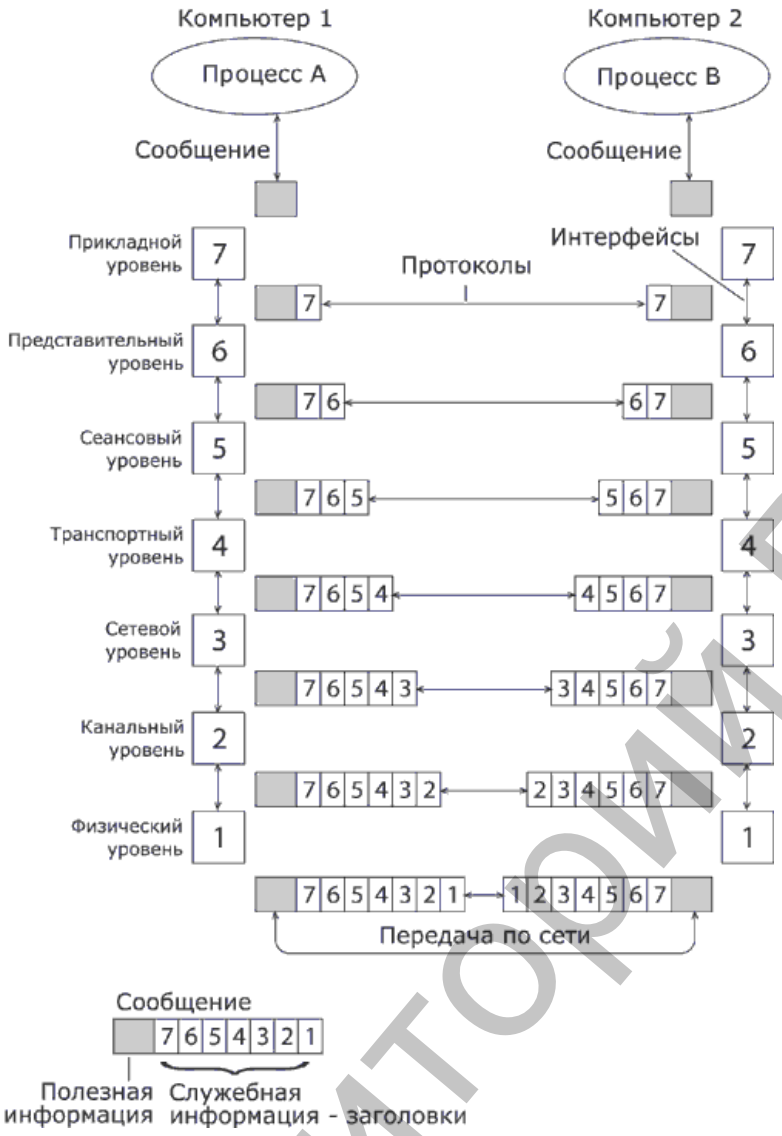


Рис. 11.6. Модель взаимодействия открытых систем ISO/OSI.

Модель OSI описывает только системные средства взаимодействия, реализуемые операционной системой, системными утилитами и аппаратными средствами. Модель не включает средства взаимодействия приложений конечных пользователей. Собственные протоколы взаимодействия приложения реализуют, обращаясь к системным средствам. Поэтому необходимо различать уровень взаимодействия приложений и прикладной уровень.

Следует также иметь в виду, что приложение может взять на себя функции некоторых верхних уровней модели OSI. Например, некоторые СУБД имеют встроенные средства удаленного доступа к файлам. В этом случае приложение, выполняя доступ к удаленным ресурсам, не использует системную файловую службу; оно обходит верхние уровни модели OSI и обращается напрямую к системным средствам, ответственным за транспортировку сообщений по сети, которые располагаются на нижних уровнях модели OSI.

Итак, пусть приложение обращается с запросом к прикладному уровню, например к файловой службе. На основании этого запроса программное

обеспечение прикладного уровня формирует сообщение стандартного формата. Обычное сообщение состоит из заголовка и поля данных. Заголовок содержит служебную информацию, которую необходимо передать через сеть прикладному уровню машины-адресата, чтобы сообщить ему, какую работу надо выполнить. В нашем случае заголовок, очевидно, должен содержать информацию о местонахождении файла и о типе операции, которую необходимо выполнить. Поле данных сообщения может быть пустым или содержать какие-либо данные, например те, которые необходимо записать в удаленный файл. Но для того чтобы доставить эту информацию по назначению, предстоит решить еще много задач, ответственность за которые несут нижележащие уровни.

После формирования сообщения прикладной уровень направляет его вниз по стеку представительному уровню. Протокол представительного уровня на основании информации, полученной из заголовка прикладного уровня, выполняет требуемые действия и добавляет к сообщению собственную служебную информацию — заголовок представительного уровня, в котором содержатся указания для протокола представительного уровня машины-адресата. Полученное в результате сообщение передается вниз сеансовому уровню, который в свою очередь добавляет свой заголовок, и т. д. (Некоторые протоколы 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. 11.7).

Когда сообщение по сети поступает на машину-адресат, оно принимается ее физическим уровнем и последовательно перемещается вверх с уровня на уровень. Каждый уровень анализирует и обрабатывает заголовок своего уровня, выполняя соответствующие данному уровню функции, а затем удаляет этот заголовок и передает сообщение вышележащему уровню.

Наряду с термином сообщение (message) существуют и другие термины, применяемые сетевыми специалистами для обозначения единиц данных в процедурах обмена. В стандартах ISO для обозначения единиц данных, с которыми имеют дело протоколы разных уровней, используется общее название протокольный блок данных (Protocol Data Unit, PDU>. Для обозначения блоков данных определенных уровней часто используются специальные названия: кадр (frame), пакет (packet), дейтаграмма (datagram), сегмент (segment).

Физический уровень

Физический уровень (Physical layer) имеет дело с передачей битов по физическим каналам связи, таким, как коаксиальный кабель, витая пара, оптоволоконный кабель или цифровой территориальный канал. К этому уровню имеют отношение характеристики физических сред передачи данных, такие как полоса пропускания, помехозащищенность, волновое сопротивление и другие. На этом же уровне определяются характеристики

электрических сигналов, передающих дискретную информацию, такую как крутизна фронтов импульсов, уровни напряжения или тока передаваемого сигнала, тип кодирования, скорость передачи сигналов. Кроме того, здесь стандартизируются типы разъемов и назначение каждого контакта.

Физический уровень :

- передача битов по физическим каналам;
- формирование электрических сигналов;
- кодирование информации;
- синхронизация;
- модуляция.

Реализуется аппаратно.

Функции физического уровня реализуются во всех устройствах, подключенных к сети. Со стороны компьютера функции физического уровня выполняются сетевым адаптером или последовательным портом.

Примером протокола физического уровня может служить спецификация 10Base-T технологии Ethernet, которая определяет в качестве используемого кабеля неэкранированную витую пару категории 3 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.

Канальный уровень

На физическом уровне просто пересылаются биты. При этом не учитывается, что в тех сетях, в которых линии связи используются (разделяются) попеременно несколькими парами взаимодействующих компьютеров, физическая среда передачи может быть занята. Поэтому одной из задач канального уровня (Data Link layer) является проверка доступности среды передачи. Другая задача канального уровня — реализация механизмов обнаружения и коррекции ошибок. Для этого на канальном уровне биты группируются в наборы, называемые кадрами (frames). Канальный уровень обеспечивает корректность передачи каждого кадра помещая специальную последовательность бит в начало и конец каждого кадра, для его выделения, а также вычисляет контрольную сумму, обрабатывая все байты кадра определенным способом, и добавляет контрольную сумму к кадру. Когда кадр приходит по сети, получатель снова вычисляет контрольную сумму полученных данных и сравнивает результат с контрольной суммой из кадра. Если они совпадают, кадр считается правильным и принимается. Если же контрольные суммы не совпадают, то фиксируется ошибка. Канальный уровень может не только обнаруживать ошибки, но и исправлять их за счет повторной передачи поврежденных кадров . Необходимо отметить, что функция исправления ошибок для канального уровня не является обязательной, поэтому в некоторых протоколах этого уровня она отсутствует, например в Ethernet и frame relay.

Функции канального уровня

Надежная доставка пакета:

1. Между двумя соседними станциями в сети с произвольной топологией.

2. Между любыми станциями в сети с типовой топологией:

- проверка доступности разделяемой среды;
- выделение кадров из потока данных, поступающих по сети; формирование кадров при отправке данных;
- подсчет и проверка контрольной суммы.

Реализуются программно-аппаратно.

В протоколах канального уровня, используемых в локальных сетях, заложена определенная структура связей между компьютерами и способы их адресации. Хотя канальный уровень и обеспечивает доставку кадра между любыми двумя узлами локальной сети, он это делает только в сети с определенной топологией связей, именно той топологией, для которой он был разработан. К таким типовым топологиям, поддерживаемым протоколами канального уровня локальных сетей, относятся "общая шина", "кольцо" и "звезда", а также структуры, полученные из них с помощью мостов и коммутаторов. Примерами протоколов канального уровня являются протоколы Ethernet, Token Ring, FDDI, 100VG-AnyLAN.

В локальных сетях протоколы канального уровня используются компьютерами, мостами, коммутаторами и маршрутизаторами. В компьютерах функции канального уровня реализуются совместными усилиями сетевых адаптеров и их драйверов.

В глобальных сетях, которые редко обладают регулярной топологией, канальный уровень часто обеспечивает обмен сообщениями только между двумя соседними компьютерами, соединенными индивидуальной линией связи. Примерами протоколов "точка-точка" (как часто называют такие протоколы) могут служить широко распространенные протоколы PPP и LAP-B. В таких случаях для доставки сообщений между конечными узлами через всю сеть используются средства сетевого уровня. Именно так организованы сети X.25. Иногда в глобальных сетях функции канального уровня в чистом виде выделить трудно, так как в одном и том же протоколе они объединяются с функциями сетевого уровня. Примерами такого подхода могут служить протоколы технологий ATM и frame relay.

В целом канальный уровень представляет собой весьма мощный набор функций по пересылке сообщений между узлами сети. В некоторых случаях протоколы канального уровня оказываются самодостаточными транспортными средствами, и тогда поверх них могут работать непосредственно протоколы прикладного уровня или приложения, без привлечения средств сетевого и транспортного уровней. Например, существует реализация протокола управления сетью SNMP непосредственно поверх Ethernet, хотя стандартно этот протокол работает поверх сетевого протокола IP и транспортного протокола UDP. Естественно, что применение такой реализации будет ограниченным — она не подходит для составных сетей разных технологий, например Ethernet и X.25, и даже для такой сети, в

которой во всех сегментах применяется Ethernet, но между сегментами существуют петлевидные связи. А вот в двухсегментной сети Ethernet, объединенной мостом, реализация SNMP над канальным уровнем будет вполне работоспособна.

Тем не менее, для обеспечения качественной транспортировки сообщений в сетях любых топологий и технологий функций канального уровня оказывается недостаточно, поэтому в модели OSI решение этой задачи возлагается на два следующих уровня — сетевой и транспортный.

Канальный уровень обеспечивает передачу пакетов данных, поступающих от протоколов верхних уровней, узлу назначения, адрес которого также указывает протокол верхнего уровня. Протоколы канального уровня оформляют переданные им пакеты в кадры собственного формата, помещая указанный адрес назначения в одно из полей такого кадра, а также сопровождая кадр контрольной суммой. Протокол канального уровня имеет локальный смысл, он предназначен для доставки кадров данных, как правило, в пределах сетей с простой топологией связей и однотипной или близкой технологией, например в односегментных сетях Ethernet или же в многосегментных сетях Ethernet и Token Ring иерархической топологии, разделенных только мостами и коммутаторами. Во всех этих конфигурациях адрес назначения имеет локальный смысл для данной сети и не изменяется при прохождении кадра от узла-источника к узлу назначения. Возможность передавать данные между локальными сетями разных технологий связана с тем, что в этих технологиях используются адреса одинакового формата, к тому же производители сетевых адаптеров обеспечивают уникальность адресов независимо от технологии.

Другой областью действия протоколов канального уровня являются связи типа "точка-точка" глобальных сетей, когда протокол канального уровня ответственен за доставку кадра непосредственному соседу. Адрес в этом случае не имеет принципиального значения, а на первый план выходит способность протокола восстанавливать искаженные и утерянные кадры , так как плохое качество территориальных каналов, особенно коммутируемых телефонных, часто требует выполнения подобных действий. Если же перечисленные выше условия не соблюдаются, например связи между сегментами Ethernet имеют петлевидную структуру, либо объединяемые сети используют различные способы адресации, как в сетях Ethernet и X.25, то протокол канального уровня не может в одиночку справиться с задачей передачи кадра между узлами и требует помощи протокола сетевого уровня.

Сетевой уровень

Сетевой уровень (Network layer) служит для образования единой транспортной системы, объединяющей несколько сетей, причем эти сети могут использовать различные принципы передачи сообщений между конечными узлами и обладать произвольной структурой связей. Функции сетевого уровня достаточно разнообразны. Рассмотрим их на примере объединения локальных сетей.

Протоколы канального уровня локальных сетей обеспечивают доставку данных между любыми узлами только в сети с соответствующей типовой топологией, например топологией иерархической звезды. Это жесткое ограничение, которое не позволяет строить сети с развитой структурой, например сети, объединяющие несколько сетей предприятия в единую сеть, или высоконадежные сети, в которых существуют избыточные связи между узлами. Можно было бы усложнять протоколы канального уровня для поддержания петлевидных избыточных связей, но принцип разделения обязанностей между уровнями приводит к другому решению. Чтобы, с одной стороны, сохранить простоту процедур передачи данных для типовых топологий, а с другой — допустить использование произвольных топологий, вводится дополнительный сетевой уровень.

На сетевом уровне сам термин "сеть" наделяют специфическим значением. В данном случае под сетью понимается совокупность компьютеров, соединенных между собой в соответствии с одной из стандартных типовых топологий и использующих для передачи данных один из протоколов канального уровня, определенный для этой топологии.

Внутри сети доставка данных обеспечивается соответствующим канальным уровнем, а вот доставкой данных между сетями занимается сетевой уровень, который и поддерживает возможность правильного выбора маршрута передачи сообщения даже в том случае, когда структура связей между составляющими сетями имеет характер, отличный от принятого в протоколах канального уровня.

Сети соединяются между собой специальными устройствами, называемыми маршрутизаторами. Маршрутизатор — это устройство, которое собирает информацию о топологии межсетевых соединений и пересылает пакеты сетевого уровня в сеть назначения. Чтобы передать сообщение от отправителя, находящегося в одной сети, получателю, находящемуся в другой сети, нужно совершить некоторое количество транзитных передач между сетями, или хопов (от слова hop — прыжок), каждый раз выбирая подходящий маршрут. Таким образом, маршрут представляет собой последовательность маршрутизаторов, через которые проходит пакет.

Сетевой уровень — доставка пакета:

- между любыми двумя узлами сети с произвольной топологией;
- между любыми двумя сетями в составной сети;
- сеть — совокупность компьютеров, использующих для обмена данными единую сетевую технологию;
- маршрут — последовательность прохождения пакетом маршрутизаторов в составной сети.

Проблема выбора наилучшего пути называется маршрутизацией, и ее решение является одной из главных задач сетевого уровня. Эта проблема осложняется тем, что самый короткий путь — не всегда самый лучший. Часто критерием при выборе маршрута является время передачи данных; оно

зависит от пропускной способности каналов связи и интенсивности трафика, которая может с течением времени изменяться. Некоторые алгоритмы маршрутизации пытаются приспособиться к изменению нагрузки, в то время как другие принимают решения на основе средних показателей за длительное время. Выбор маршрута может осуществляться и по другим критериям, таким как надежность передачи.

В общем случае функции сетевого уровня шире, чем функции передачи сообщений по связям с нестандартной структурой, которые мы рассмотрели на примере объединения нескольких локальных сетей. Сетевой уровень также решает задачи согласования разных технологий, упрощения адресации в крупных сетях и создания надежных и гибких барьеров на пути нежелательного трафика между сетями.

Сообщения сетевого уровня принято называть пакетами (packet). При организации доставки пакетов на сетевом уровне используется понятие "номер сети". В этом случае адрес получателя состоит из старшей части — номера сети и младшей — номера узла в этой сети. Все узлы одной сети должны иметь одну и ту же старшую часть адреса, поэтому термину "сеть" на сетевом уровне можно дать и другое, более формальное, определение: сеть — это совокупность узлов, сетевой адрес которых содержит один и тот же номер сети.

На сетевом уровне определяется два вида протоколов. Первый вид — сетевые протоколы (routed protocols) — 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 протоколами обмена маршрутной информацией или просто протоколами маршрутизации (routing protocols). С помощью этих протоколов маршрутизаторы собирают информацию о топологии межсетевых соединений. Протоколы сетевого уровня реализуются программными модулями операционной системы, а также программными и аппаратными средствами маршрутизаторов.

На сетевом уровне работают протоколы еще одного типа, которые отвечают за отображение адреса узла, используемого на сетевом уровне, в локальный адрес сети. Такие протоколы часто называют протоколами разрешения адресов — Address Resolution Protocol, ARP. Иногда их относят не к сетевому уровню, а к канальному, хотя тонкости классификации не изменяют сути.

Примерами протоколов сетевого уровня являются протокол межсетевого взаимодействия IP стека TCP/IP и протокол межсетевого обмена пакетами IPX стека Novell.

Транспортный уровень

На пути от отправителя к получателю пакеты могут быть искажены или утеряны. Хотя некоторые приложения имеют собственные средства обработки ошибок, существуют и такие, которые предпочитают сразу иметь дело с надежным соединением. Транспортный уровень (Transport layer)

обеспечивает приложениям или верхним уровням стека — прикладному и сеансовому — передачу данных с той степенью надежности, которая им требуется. Модель OSI определяет пять классов сервиса, предоставляемых транспортным уровнем. 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, таких как искажение, потеря и дублирование пакетов.

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, а с другой стороны, зависит от того, насколько надежной является система транспортировки данных в сети, обеспечиваемая уровнями, расположенными ниже транспортного — сетевым, канальным и физическим. Так, например, если качество каналов передачи связи очень высокое, и вероятность наличия ошибок, не обнаруженных протоколами более низких уровней, невелика, стоит воспользоваться одним из облегченных сервисов транспортного уровня, не обремененных многочисленными проверками, квитированием и другими приемами повышения надежности. Если же транспортные средства нижних уровней изначально очень ненадежны, то целесообразно обратиться к наиболее развитому сервису транспортного уровня, который работает, используя максимум средств для обнаружения и устранения ошибок, — с помощью предварительного установления логического соединения, отслеживания доставки сообщений по контрольным суммам и циклической нумерации пакетов, установления тайм-аутов доставки и т. п.

Транспортный уровень — обеспечение доставки информации с требуемым качеством между любыми узлами сети:

- разбивка сообщения сеансового уровня на пакеты , их нумерация;
- буферизация принимаемых пакетов;
- упорядочивание прибывающих пакетов;
- адресация прикладных процессов;
- управление потоком.

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

Протоколы четырех нижних 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.

Остальные три верхних уровня решают задачи предоставления прикладных сервисов на основании имеющейся транспортной подсистемы.

Сеансовый уровень

Сеансовый уровень (Session layer) обеспечивает управление диалогом: фиксирует, какая из сторон является активной в настоящий момент, предоставляет средства синхронизации. Последние позволяют вставлять контрольные точки в длинные передачи, 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, хотя функции этого уровня часто объединяют с функциями прикладного уровня и реализуют в одном протоколе.

Сеансовый уровень — управление диалогом объектов прикладного уровня:

- установление способа обмена сообщениями (дуплексный или полудуплексный);
- синхронизация обмена сообщениями;
- организация "контрольных точек" диалога.

Представительный уровень

Представительный уровень (Presentation layer) имеет дело с формой представления 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 в кодах ASCII и EBCDIC. На этом уровне может выполняться шифрование и дешифрование данных, благодаря которому секретность обмена данными обеспечивается сразу для всех прикладных служб. Примером такого протокола является протокол Secure Socket Layer (SSL), который обеспечивает секретный обмен сообщениями для протоколов прикладного уровня стека TCP/IP.

Уровень представления — согласовывает представление (синтаксис) данных при взаимодействии двух прикладных процессов:

- преобразование данных из внешнего формата во внутренний;
- шифрование и расшифровка данных.

Прикладной уровень

Прикладной уровень (Application layer) — это в действительности просто набор разнообразных протоколов, с помощью которых пользователи сети получают доступ к разделяемым ресурсам, таким как файлы, принтеры или гипертекстовые Web-страницы, а также организуют совместную работу, например с помощью протокола электронной почты. Единица данных, которой оперирует прикладной уровень, обычно называется сообщением (message).

Прикладной уровень — набор всех сетевых сервисов, которые предоставляет система конечному пользователю:

- идентификация, проверка прав доступа;
- принт- и файл-сервис, почта, удаленный доступ...

Существует очень много различных служб прикладного уровня. Приведем в качестве примера хотя бы несколько наиболее распространенных реализаций файловых служб: NCP в операционной системе Novell NetWare, SMB в Microsoft Windows NT, NFS, FTP и TFTP, входящие в стек TCP/IP.

Сетезависимые и сетенезависимые уровни

Функции всех уровней модели OSI могут быть отнесены к одной из двух групп: либо к функциям, зависящим от конкретной технической реализации сети, либо к функциям, ориентированным на работу с приложениями.

Три нижних уровня — физический, канальный и сетевой — являются сетезависимыми, то есть протоколы этих уровней тесно связаны с технической реализацией сети и используемым коммуникационным оборудованием. Например, переход на оборудование FDDI означает полную смену протоколов физического и канального уровней во всех узлах сети.

Три верхних уровня — прикладной, представительный и сеансовый — ориентированы на приложения и мало зависят от технических особенностей построения сети. На протоколы этих уровней не влияют какие бы то ни было изменения в топологии сети, замена оборудования или переход на другую сетевую технологию. Так, переход от Ethernet к высокоскоростной технологии 100VG-AnyLAN не потребует никаких изменений в программных средствах, реализующих функции прикладного, представительного и сеансового уровней.

Транспортный уровень является промежуточным, он скрывает все детали функционирования нижних уровней от верхних. Это позволяет разрабатывать приложения, не зависящие от технических средств непосредственной транспортировки сообщений.

Лекция 8. Сетевые программные средства

Функции верхних уровней эталонной модели OSI выполняют сетевые программные средства. Для установки сети достаточно иметь набор сетевого оборудования, его драйверы, а также сетевое программное обеспечение. От выбора программного обеспечения зависит очень многое: допустимый размер сети, удобство использования и контроля сети, режимы доступа к ресурсам, производительность сети в разных режимах и т.д. Правда, заменить одну программную систему на другую значительно проще, чем сменить оборудование.

С точки зрения распределения функций между компьютерами сети, все сети можно разделить на две группы:

- Одноранговые сети, состоящие из равноправных (с точки зрения доступа к сети) компьютеров.
- Сети на основе серверов, в которых существуют только выделенные (dedicated) серверы, занимающиеся исключительно сетевыми функциями. Выделенный сервер может быть единственным или их может быть несколько.

Согласно с этим, выделяют и типы программных средств, реализующих данные виды сетей.

Одноранговые сети

Одноранговые сети (Peer-to-Peer Network) и соответствующие программные средства, как правило, используются для объединения небольшого количества компьютеров (рис. 6.10). Каждый компьютер такой сети может одновременно являться и сервером и клиентом сети, хотя вполне допустимо назначение одного компьютера только сервером, а другого только клиентом. Принципиальна возможность совмещения функций клиента и сервера. Важно также и то, что в одноранговой сети любой сервер может быть невыделенным (non-dedicated), может не только обслуживать сеть, но и работать как автономный компьютер (правда, запросы к нему по сети сильно снижают скорость его работы). В одноранговой сети могут быть и выделенные серверы, только обслуживающие сеть.

Именно в данном случае наиболее правильно говорить о распределенных дисковых ресурсах, о виртуальном компьютере, а также о суммировании объемов дисков всех компьютеров сети. Если все компьютеры являются серверами, то любой файл, созданный на одном из них сразу же становится доступным всем остальным компьютерам, его не надо передавать на централизованный сервер.

Достоинством одноранговых сетей является их высокая гибкость: в зависимости от конкретной задачи сеть может использоваться очень активно либо совсем не использоваться. Из-за большой самостоятельности компьютеров в таких сетях редко бывает ситуация перегрузки (к тому же количество компьютеров обычно невелико). Установка одноранговых сетей довольно проста, к тому же не требуются дополнительные дорогостоящие

серверы. Кроме того, нет необходимости в системном администрировании, пользователи могут сами управлять своими ресурсами.

В одноранговых сетях допускается определение различных прав пользователей по доступу к сетевым ресурсам, но система разграничения прав не слишком развита. Если каждый ресурс защищен своим паролем, то пользователю приходится запоминать большое число паролей.

К недостаткам одноранговых сетей относятся также слабая система контроля и протоколирования работы сети, трудности с резервным копированием распределенной информации. К тому же выход из строя любого компьютера-сервера приводит к потере части общей информации, то есть все такие компьютеры должны быть по возможности высоконадежными. Эффективная скорость передачи информации по одноранговой сети часто оказывается недостаточной, поскольку трудно обеспечить быстродействие процессоров, большой объем оперативной памяти и высокие скорости обмена с жестким диском для всех компьютеров сети. К тому же компьютеры сети работают не только на сеть, но и решают другие задачи.

Несколько примеров одноранговых сетевых программных средств:

- NetWare Lite компании Novell (сейчас уже не производится);
- LANtastic компании Artisoft (выпуск практически прекращен);
- Windows for Workgroups компании Microsoft (первая версия ОС Windows со встроенной поддержкой сети, выпущенная в 1992 году);
- Windows NT Workstation компании Microsoft;
- Windows 95... Windows XP компании Microsoft.

Первые одноранговые сетевые программные средства представляли собой сетевые оболочки, работающие под управлением DOS (например, NetWare Lite). Они перехватывали все запросы DOS, те запросы, которые вызваны обращениями к сетевым устройствам, обрабатывались и выполнялись сетевой оболочкой, а те, которые вызваны обращениями к "местным", несетевым ресурсам, возвращались обратно в DOS и обрабатывались стандартным образом.

Более поздние одноранговые сетевые программные средства уже были встроены в операционную систему Windows. Это гораздо удобнее, так как исключается этап установки сетевых программ. Поэтому сетевые оболочки сейчас уже практически не используются, хотя многие их характеристики были заметно лучше, чем у сетевых средств Windows.

Сейчас считается, что одноранговая сеть наиболее эффективна в небольших сетях (около 10 компьютеров). При значительном количестве компьютеров сетевые операции сильно замедлят работу компьютеров и создадут множество других проблем. Тем не менее, для небольшого офиса одноранговая сеть – оптимальное решение.

Самая распространенная в настоящий момент одноранговая сеть – это сеть на основе Windows XP (или более ранних версий ОС Windows).

При этом пользователь, приобретая компьютер с установленной операционной системой, автоматически получает и возможность выхода в

сеть. Естественно, это во многих случаях гораздо удобнее, чем приобретать и устанавливать пусть даже и более совершенные продукты других фирм. К тому же пользователю не надо изучать интерфейс пользователя сетевой программы, так как он строится так же, как и интерфейс пользователя всех остальных частей операционной системы.

Если приобретаемый компьютер еще и имеет установленный сетевой адаптер, то построить сеть пользователю совсем просто. Надо только соединить компьютеры кабелем и настроить сетевые программы.

В Windows предусмотрена поддержка совместного использования дисков (в том числе гибких дисков и CD), а также принтеров. Имеется возможность объединения всех пользователей в рабочие группы для более удобного поиска требуемых ресурсов и организации доступа к ним. Пользователи имеют доступ к встроенной системе электронной почты. Это означает, что все пользователи сети получают возможность совместно применять многие ресурсы ОС своего компьютера.

При настройке сети пользователь должен выбрать тип сетевого протокола. По умолчанию используется протокол TCP/IP, но возможно применение IPX/SPX (NWLink), а также NetBEUI. При выборе TCP/IP можно задавать адреса IP вручную или с помощью автоматической настройки адресации (в этом случае компьютер сам присвоит себе адрес из диапазона, не используемого в Интернет).

Кроме того, надо задать индивидуальное имя компьютера и определить рабочую группу, к которой он относится.

После этого можно разрешить доступ по сети к ресурсам каждого компьютера сети, к его файлам, папкам, принтерам, сканерам, доступу в Интернет.

Сети на основе сервера

Сети на основе сервера (Server-based Network) применяются в тех случаях, когда в сеть должно быть объединено много пользователей. В этом случае возможностей одноранговой сети может не хватить. Поэтому в сеть включается специализированный компьютер – сервер, который обслуживает только сеть и не решает никаких других задач (рис. 6.11). Такой сервер называется выделенным. Сервер может быть и специализирован на решении одной задачи, например, сервер печати, но чаще всего серверами выступают именно компьютеры. В сети может быть и несколько серверов, каждый из которых решает свою задачу.

Серверы специально оптимизированы для быстрой обработки сетевых запросов на разделяемые ресурсы и для управления защитой файлов и каталогов. При больших размерах сети мощности одного сервера может оказаться недостаточно, и тогда в сеть включают несколько серверов. Серверы могут выполнять и некоторые другие задачи: сетевая печать, выход в глобальную сеть, связь с другой локальной сетью, обслуживание электронной почты и т.д. Количество пользователей сети на основе сервера может достигать нескольких тысяч. Одноранговой сетью такого размера

просто невозможно было бы управлять. Кроме того, в сети на основе серверов можно легко менять количество подключаемых компьютеров, такие сети называются масштабируемыми.

В любом случае в сети на основе сервера существует четкое разделение компьютеров на клиентов (или рабочие станции) и серверы. Клиенты не могут работать как серверы, а серверы – как клиенты и как автономные компьютеры. Очевидно, что все сетевые дисковые ресурсы могут располагаться только на сервере, а клиенты могут обращаться только к серверу, но не друг к другу. Однако это не значит, что они не могут общаться между собой, просто пересылка информации от одного клиента к другому возможна только через сервер, например, через файл, доступный всем клиентам. В данном случае реализуется некоторая "логическая звезда" с сервером в центре, хотя физическая топология сети может быть любой.

Достоинством сети на основе сервера часто называют надежность. Это верно, но только с одной оговоркой: если сервер действительно очень надежен. В противном случае любой отказ сервера приводит к полному параличу сети в отличие от ситуации с одноранговой сетью, где отказ одного из компьютеров не приводит к отказу всей сети. Бесспорное достоинство сети на основе сервера – высокая скорость обмена, так как сервер всегда оснащается быстрым процессором (или даже несколькими процессорами), оперативной памятью большого объема и быстрыми жесткими дисками. Так как все ресурсы сети собраны в одном месте, возможно применение гораздо более мощных средств управления доступом, защиты данных, протоколирования обмена, чем в одноранговых сетях.

К недостаткам сети на основе сервера относятся ее громоздкость в случае небольшого количества компьютеров, зависимость всех компьютеров-клиентов от сервера, более высокая стоимость сети вследствие использования дорогого сервера. Но, говоря о стоимости, надо также учитывать, что при одном и том же объеме сетевых дисков большой диск сервера получается дешевле, чем много дисков меньшего объема, входящих в состав всех компьютеров одноранговой сети.

Примеры некоторых сетевых программных средств на основе сервера:

- NetWare компании Novell (самая распространенная сетевая ОС);
- LAN Server компании IBM (почти не используется);
- LAN Manager компании Microsoft;
- Windows NT Server компании Microsoft;
- Windows Server 2003 компании Microsoft.

На файл-сервере в данном случае устанавливается специальная сетевая операционная система, рассчитанная на работу сервера. Эта сетевая ОС оптимизирована для эффективного выполнения специфических операций по организации сетевого обмена. На рабочих станциях (клиентах) может устанавливаться любая совместимая операционная система, поддерживающая сеть.

Для обеспечения надежной работы сети при авариях электропитания применяется бесперебойное электропитание сервера. В данном случае это гораздо проще, чем при одноранговой сети, где желательно оснащать источниками бесперебойного питания все компьютеры сети. Для администрирования сети (то есть управления распределением ресурсов, контроля прав доступа, защиты данных, файловой системы, резервирования файлов и т.д.) в случае сети на основе сервера необходимо выделять специального человека, имеющего соответствующую квалификацию. Централизованное администрирование облегчает обслуживание сети и позволяет оперативно решать все вопросы. Особенно это важно для надежной защиты данных от несанкционированного доступа. В случае же одноранговой сети можно обойтись и без специалиста-администратора, правда, при этом все пользователи сети должны иметь хоть какое-то представление об администрировании.

Процесс установки серверной сетевой операционной системы гораздо сложнее, чем в случае одноранговой сети. Так, он включает в себя следующие обязательные процедуры:

- форматирование и разбиение на разделы жесткого диска компьютера-сервера;
- присвоение индивидуального имени серверу;
- присвоение имени сети;
- установка и настройка сетевого протокола;
- выбор сетевых служб;
- ввод пароля администратора.

Сетевая операционная система на базе сервера Windows Server 2003 предоставляет пользователям гораздо больше возможностей, чем в случае одноранговой сети.

Она позволяет строить сложные иерархические структуры сети на основе логических групп компьютеров (доменов, domain), наборов доменов (деревьев, tree) и наборов деревьев (леса, forest).

Домен представляет собой группу компьютеров, управляемых контроллером домена, специальным сервером. Домен использует собственную базу данных, содержащую учетные записи пользователей, и управляет собственными ресурсами, такими как принтеры и общие файлы. Каждому домену присваивается свое имя (обычно домен рассматривается как отдельная сеть со своим номером). В каждый домен может входить несколько рабочих групп, которые формируются из пользователей, решающих общую или сходные задачи. В принципе домен может включать тысячи пользователей, однако обычно домены не слишком велики, и несколько доменов объединяются в дерево доменов. Это упрощает управление сетью. Точно так же несколько деревьев может объединяться в лес, самую крупную административную структуру, поддерживаемую данной ОС.

В процессе установки Windows Server 2003 необходимо задать тип протокола сети. По умолчанию используется TCP/IP, но возможно применение NWLink (IPX/SPX).

Каждому серверу необходимо назначить роль, которую он будет выполнять в сети:

- контроллер домена (управляет работой домена);
- файловый сервер (хранит совместно используемые файлы);
- сервер печати (управляет сетевым принтером);
- Web-сервер (содержит сайт, доступный по сети Интернет или по локальной сети);
- коммуникационный сервер (обеспечивает работу электронной почты и конференций);
- сервер удаленного доступа (обеспечивает удаленный доступ).

Каждому пользователю сети необходимо присвоить свое учетное имя и пароль, а также права доступа к ресурсам (полномочия). Права доступа могут задаваться как индивидуально, так и целой рабочей группе пользователей. Windows Server 2003 обеспечивает следующие виды полномочий для папок:

- полный контроль (просмотр, чтение, запись, удаление папки, подпапок, файлов, запуск на исполнение, установка прав доступа к папке);
- изменение (просмотр, чтение, запись, удаление подпапок и файлов, запуск на исполнение);
- чтение и исполнение (просмотр, чтение, запуск на исполнение);
- просмотр содержимого папки;
- запись нового содержимого в папку;
- чтение информации из папки.

Те же самые уровни полномочий (кроме просмотра содержимого) предусмотрены и для файлов, доступных по сети.

Сетевые операционные системы NetWare компании Novell сегодня очень популярны, что объясняется их высокой производительностью, совместимостью с разными аппаратными средствами и развитой системой средств защиты данных. Компания Novell выпускает сетевые программные средства с 1979 года: несколько версий сетевых ОС на базе файловых серверов (одна из последних версий – NetWare 6 и 6.5), клиентское программное обеспечение, а также средства диагностики работы сетей. Популярные до недавнего времени сетевые оболочки одноранговых сетей, такие как NetWare Lite и Personal NetWare сейчас уже не производятся.

Отличительной особенностью сетевых программных средств Novell всегда была их открытость, то есть совместимость с операционными системами различных фирм: Windows, UNIX, Macintosh, OS/2. Кроме того, они всегда обеспечивали возможность работы с аппаратными средствами практически всех известных производителей. Это позволяет строить на их основе сети из разнообразных абонентов – от самых простых до самых сложных.

Все сетевые продукты NetWare допускают подключение бездисковых рабочих станций (клиентов), что позволяет при необходимости значительно снизить стоимость сети. Во всех продуктах предусмотрена поддержка сетевых мостов.

Продуктам Novell NetWare присущи и недостатки, например, их стоимость для небольших сетей оказывается достаточно высокой по сравнению с ценой продуктов других производителей. Кроме того, их установка сравнительно сложна, но они уже стали фактическим стандартом, поэтому их позиции на рынке довольно прочны.

Рассмотрим кратко особенности сетевой ОС Novell NetWare 6.5.

Как и в случае Microsoft Windows Server 2003, Novell NetWare 6.5 требует создания древовидной иерархической структуры, включающей в себя сетевые деревья, серверы, пользователей, группы и прочие объекты.

Novell NetWare 6.5 предусматривает обязательное разбиение жестких дисков с использованием собственной системы хранения файлов NSS (Novell Storage Services), которое требует создания логических разделов (Volumes) на диске. Это позволяет серверу более эффективно решать сетевые задачи.

Для каждого сервера сети надо выбрать один из трех типов:

- Настраиваемый сервер (в частности, Web-сервер, FTP-сервер).
- Основной файловый сервер.
- Специальный сервер (например, DNS/DHCP-сервер, контролирующий сетевые адреса и имена, или сервер резервного копирования).

Кроме того, надо задать тип используемого протокола – TCP/IP или IPX/SPX.

На компьютеры-клиенты следует установить клиентское программное обеспечение. Это сравнительно простая процедура.

Каждому клиенту присваивается учетная запись, предоставляются свои права доступа к ресурсам. Клиенты могут быть объединены в рабочие группы, каждой из которых присваиваются имена и права доступа.

Предусмотрены следующие виды доступа к файлам и каталогам (папкам):

- Изменение прав доступа к каталогу или файлу;
- Просмотр каталога;
- Создание каталогов и файлов в данном каталоге;
- Удаление каталогов и файлов в данном каталоге;
- Изменение содержимого файлов;
- Любые операции над файлами каталога;
- Запись в файл.

Лекция 9. Технологии стандартных сетей.

За время, прошедшее с момента появления первых локальных сетей, было разработано несколько сот самых разных сетевых технологий, однако заметное распространение получили немногие. Это связано, прежде всего, с высоким уровнем стандартизации принципов организации сетей и с поддержкой их известными компаниями. Тем не менее, не всегда стандартные сети обладают рекордными характеристиками, обеспечивают наиболее оптимальные режимы обмена. Но большие объемы выпуска их аппаратуры и, следовательно, ее невысокая стоимость дают им огромные преимущества. Немаловажно и то, что производители программных средств также в первую очередь ориентируются на самые распространенные сети. Поэтому пользователь, выбирающий стандартные сети, имеет полную гарантию совместимости аппаратуры и программ.

В настоящее время уменьшение количества типов используемых сетей стало тенденцией. Дело в том, что увеличение скорости передачи в локальных сетях до 100 и даже до 1000 Мбит/с требует применения самых передовых технологий, проведения дорогих научных исследований. Естественно, это могут позволить себе только крупнейшие фирмы, которые поддерживают свои стандартные сети и их более совершенные разновидности. К тому же большинство потребителей уже установило у себя какие-то сети и не желает сразу и полностью заменять сетевое оборудование, В ближайшем будущем вряд ли стоит ожидать того, что будут приняты принципиально новые стандарты.

В табл. 7.1 приведены характеристики классических вариантов стандартных локальных сетей. Все стандартные сети имеют несколько вариантов, отличающихся типом используемого кабеля, скоростями передачи, допустимыми размерами сети. О них подробнее рассказано в разделах, посвященных конкретным типам сетей.

Таблица 7.1. Параметры базовых вариантов стандартных сетей

Параметр сети	**Ethernet**	**Token-Ring**	**Arcnet**	**FDDI**	**100VG-AnyLAN**
Стандарт	IEEE 802.3	IEEE 802.5	Datapoint	ISO 9314	IEEE 802.12
Топология	Шина	Кольцо	Шина	Кольцо	Звезда
Скорость передачи	10 (100) Мбит/с	(16) Мбит/с	2,5 Мбит/с	100 Мбит/с	100 Мбит/с
Длина	5 км	120 м	6 км	20 км	1 км
Среда	КК	ВП	КК	ОВ	ВП
Метод управления	CSMA/CD	Маркер	Маркер	Маркер	Центр
Код	Манчестер	Бифазны	Arcnet	4B/5B	5B/6B

		й			
Количество	До 1024	До 260	До 255	До 1000	До 1024
КК — коаксиальный кабель, ВП — кабель на витых парах, ОВ — оптоволоконный кабель					

Сети Ethernet и Fast Ethernet

Наибольшее распространение среди стандартных сетей получила сеть Ethernet. Впервые она появилась в 1972 году (разработчиком выступила известная фирма Xerox). Сеть оказалась довольно удачной, и вследствие этого ее в 1980 году поддержали такие крупнейшие компании, как DEC и Intel (объединение этих компаний назвали DIX по первым буквам их названий). Их стараниями в 1985 году сеть Ethernet стала международным стандартом, ее приняли крупнейшие международные организации по стандартам: комитет 802 IEEE (Institute of Electrical and Electronic Engineers) и ECMA (European Computer Manufacturers Association).

Стандарт получил название IEEE 802.3 (по-английски читается как "eight oh two dot three"). Он определяет множественный доступ к моноканалу типа шина с обнаружением конфликтов и контролем передачи, то есть с уже упоминавшимся методом доступа CSMA/CD. Этому стандарту удовлетворяли и некоторые другие сети, так как уровень его детализации невысок. В результате сети стандарта IEEE 802.3 нередко были несовместимы между собой как по конструктивным, так и по электрическим характеристикам. Однако в последнее время стандарт IEEE 802.3 считается стандартом именно сети Ethernet.

Основные характеристики первоначального стандарта IEEE 802.3:

- топология – шина;
- среда передачи – коаксиальный кабель;
- скорость передачи – 10 Мбит/с;
- максимальная длина сети – 5 км;
- максимальное количество абонентов – до 1024;
- длина сегмента сети – до 500 м;
- количество абонентов на одном сегменте – до 100;
- метод доступа – CSMA/CD;
- передача узкополосная, то есть без модуляции (моноканал).

Строго говоря, между стандартами IEEE 802.3 и Ethernet существуют незначительные отличия, но о них обычно предпочитают не вспоминать.

Сеть Ethernet сейчас наиболее популярна в мире (более 90% рынка), предположительно таковой она и останется в ближайшие годы. Этому в немалой степени способствовало то, что с самого начала характеристики, параметры, протоколы сети были открыты, в результате чего огромное число производителей во всем мире стали выпускать аппаратуру Ethernet, полностью совместимую между собой.

В классической сети Ethernet применялся 50-омный коаксиальный кабель двух видов (толстый и тонкий). Однако в последнее время (с начала 90-х годов) наибольшее распространение получила версия Ethernet,

использующая в качестве среды передачи витые пары. Определен также стандарт для применения в сети оптоволоконного кабеля. Для учета этих изменений в изначальный стандарт IEEE 802.3 были сделаны соответствующие добавления. В 1995 году появился дополнительный стандарт на более быструю версию Ethernet, работающую на скорости 100 Мбит/с (так называемый Fast Ethernet, стандарт IEEE 802.3u), использующую в качестве среды передачи витую пару или оптоволоконный кабель. В 1997 году появилась и версия на скорость 1000 Мбит/с (Gigabit Ethernet, стандарт IEEE 802.3z).

Помимо стандартной топологии шина все шире применяются топологии типа пассивная звезда и пассивное дерево. При этом предполагается использование репитеров и репитерных концентраторов, соединяющих между собой различные части (сегменты) сети. В результате может сформироваться древовидная структура на сегментах разных типов (рис. 7.1).

Рис. 7.1. Классическая топология сети Ethernet

В качестве сегмента (части сети) может выступать классическая шина или единичный абонент. Для шинных сегментов используется коаксиальный кабель, а для лучей пассивной звезды (для присоединения к концентратору одиночных компьютеров) – витая пара и оптоволоконный кабель. Главное требование к полученной в результате топологии – чтобы в ней не было замкнутых путей (петель). Фактически получается, что все абоненты соединены в физическую шину, так как сигнал от каждого из них распространяется сразу во все стороны и не возвращается назад (как в кольце).

Максимальная длина кабеля сети в целом (максимальный путь сигнала) теоретически может достигать 6,5 километров, но практически не превышает 3,5 километров.

В сети Fast Ethernet не предусмотрена физическая топология шина, используется только пассивная звезда или пассивное дерево. К тому же в Fast Ethernet гораздо более жесткие требования к предельной длине сети. Ведь при увеличении в 10 раз скорости передачи и сохранении формата пакета его минимальная длина становится в десять раз короче. Таким образом в 10 раз уменьшается допустимая величина двойного времени прохождения сигнала по сети (5,12 мкс против 51,2 мкс в Ethernet).

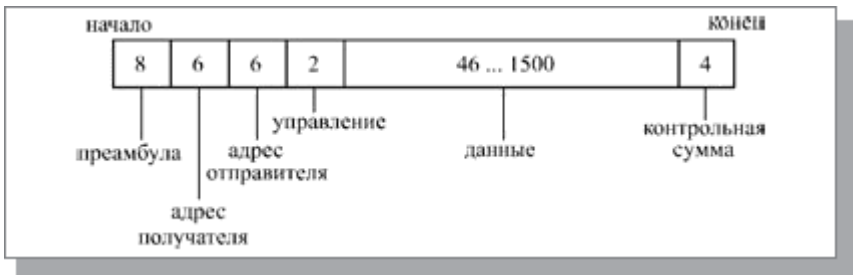


Рис. 7.2. Структура пакета сети Ethernet

Длина кадра Ethernet (то есть пакета без преамбулы) должна быть не менее 512 битовых интервалов или 51,2 мкс (именно такова предельная величина двойного времени прохождения в сети). Предусмотрена индивидуальная, групповая и широковещательная адресация.

В пакет Ethernet входят следующие поля:

• Преамбула состоит из 8 байт, первые семь представляют собой код 10101010, а последний байт – код 10101011. В стандарте IEEE 802.3 восьмой байт называется признаком начала кадра (SFD – Start of Frame Delimiter) и образует отдельное поле пакета.

• Адреса получателя (приемника) и отправителя (передатчика) включают по 6 байт и строятся по стандарту, описанному в разделе "Адресация пакетов" лекции 4. Эти адресные поля обрабатываются аппаратурой абонентов.

• Поле управления (L/T – Length/Type) содержит информацию о длине поля данных. Оно может также определять тип используемого протокола. Принято считать, что если значение этого поля не больше 1500, то оно указывает на длину поля данных. Если же его значение больше 1500, то оно определяет тип кадра. Поле управления обрабатывается программно.

• Поле данных должно включать в себя от 46 до 1500 байт данных. Если пакет должен содержать менее 46 байт данных, то поле данных дополняется байтами заполнения. Согласно стандарту IEEE 802.3, в структуре пакета выделяется специальное поле заполнения (pad data – незначащие данные), которое может иметь нулевую длину, когда данных достаточно (больше 46 байт).

• Поле контрольной суммы (FCS – Frame Check Sequence) содержит 32-разрядную циклическую контрольную сумму пакета (CRC) и служит для проверки правильности передачи пакета.

Таким образом, минимальная длина кадра (пакета без преамбулы) составляет 64 байта (512 бит). Именно эта величина определяет максимально допустимую двойную задержку распространения сигнала по сети в 512 битовых интервалов (51,2 мкс для Ethernet или 5,12 мкс для Fast Ethernet). Стандарт предполагает, что преамбула может уменьшаться при прохождении пакета через различные сетевые устройства, поэтому она не учитывается. Максимальная длина кадра равна 1518 байтам (12144 бита, то есть 1214,4 мкс для Ethernet, 121,44 мкс для Fast Ethernet). Это важно для выбора размера буферной памяти сетевого оборудования и для оценки общей загруженности сети.

Для сети Ethernet, работающей на скорости 10 Мбит/с, стандарт определяет четыре основных типа сегментов сети, ориентированных на различные среды передачи информации:

- 10BASE5 (толстый коаксиальный кабель);
- 10BASE2 (тонкий коаксиальный кабель);
- 10BASE-T (витая пара);
- 10BASE-FL (оптоволоконный кабель).

Наименование сегмента включает в себя три элемента: цифра "10" означает скорость передачи 10 Мбит/с, слово BASE – передачу в основной полосе частот (то есть без модуляции высокочастотного сигнала), а последний элемент – допустимую длину сегмента: "5" – 500 метров, "2" – 200 метров (точнее, 185 метров) или тип линии связи: "T" – витая пара (от английского "twisted-pair"), "F" – оптоволоконный кабель (от английского "fiber optic").

Точно так же для сети Ethernet, работающей на скорости 100 Мбит/с (Fast Ethernet) стандарт определяет три типа сегментов, отличающихся типами среды передачи:

- 100BASE-T4 (счетверенная витая пара);
- 100BASE-TX (сдвоенная витая пара);
- 100BASE-FX (оптоволоконный кабель).

Здесь цифра "100" означает скорость передачи 100 Мбит/с, буква "T" – витую пару, буква "F" – оптоволоконный кабель. Типы 100BASE-TX и 100BASE-FX иногда объединяют под именем 100BASE-X, а 100BASE-T4 и 100BASE-TX – под именем 100BASE-T.

Сеть Token-Ring

Сеть Token-Ring (маркерное кольцо) была предложена компанией IBM в 1985 году (первый вариант появился в 1980 году). Она предназначалась для объединения в сеть всех типов компьютеров, выпускаемых IBM. Уже тот факт, что ее поддерживает компания IBM, крупнейший производитель компьютерной техники, говорит о том, что ей необходимо уделить особое внимание. Но не менее важно и то, что Token-Ring является в настоящее время международным стандартом IEEE 802.5 (хотя между Token-Ring и IEEE 802.5 есть незначительные отличия). Это ставит данную сеть на один уровень по статусу с Ethernet.

Разрабатывалась Token-Ring как надежная альтернатива Ethernet. И хотя сейчас Ethernet вытесняет все остальные сети, Token-Ring нельзя считать безнадежно устаревшей. Более 10 миллионов компьютеров по всему миру объединены этой сетью.

Сеть Token-Ring имеет топологию кольцо, хотя внешне она больше напоминает звезду. Это связано с тем, что отдельные абоненты (компьютеры) присоединяются к сети не напрямую, а через специальные концентраторы или многостанционные устройства доступа (MSAU или MAU – Multistation Access Unit). Физически сеть образует звездно-кольцевую топологию (рис. 7.3). В действительности же абоненты объединяются все-таки в кольцо, то

есть каждый из них передает информацию одному соседнему абоненту, а принимает информацию от другого.

В качестве среды передачи в сети IBM Token-Ring сначала применялась витая пара, как неэкранированная (UTP), так и экранированная (STP), но затем появились варианты аппаратуры для коаксиального кабеля, а также для оптоволоконного кабеля в стандарте FDDI.

Основные технические характеристики классического варианта сети Token-Ring:

- максимальное количество концентраторов типа IBM 8228 MAU – 12;
- максимальное количество абонентов в сети – 96;
- максимальная длина кабеля между абонентом и концентратором – 45 метров;
- максимальная длина кабеля между концентраторами – 45 метров;
- максимальная длина кабеля, соединяющего все концентраторы – 120 метров;
- скорость передачи данных – 4 Мбит/с и 16 Мбит/с.

Все приведенные характеристики относятся к случаю использования неэкранированной витой пары. Если применяется другая среда передачи, характеристики сети могут отличаться. Например, при использовании экранированной витой пары (STP) количество абонентов может быть увеличено до 260 (вместо 96), длина кабеля – до 100 метров (вместо 45), количество концентраторов – до 33, а полная длина кольца, соединяющего концентраторы – до 200 метров. Оптоволоконный кабель позволяет увеличивать длину кабеля до двух километров.

Сеть Token-Ring в классическом варианте уступает сети Ethernet как по допустимому размеру, так и по максимальному количеству абонентов. Что касается скорости передачи, то в настоящее время имеются версии Token-Ring на скорость 100 Мбит/с (High Speed Token-Ring, HSTR) и на 1000 Мбит/с (Gigabit Token-Ring). Компании, поддерживающие Token-Ring (среди которых IBM, Olicom, Madge), не намерены отказываться от своей сети, рассматривая ее как достойного конкурента Ethernet.

По сравнению с аппаратурой Ethernet аппаратура Token-Ring заметно дороже, так как используется более сложный метод управления обменом, поэтому сеть Token-Ring не получила столь широкого распространения.

Однако в отличие от Ethernet сеть Token-Ring значительно лучше держит высокий уровень нагрузки (более 30—40%) и обеспечивает гарантированное время доступа. Это необходимо, например, в сетях производственного назначения, в которых задержка реакции на внешнее событие может привести к серьезным авариям.

В сети Token-Ring используется классический маркерный метод доступа, то есть по кольцу постоянно циркулирует маркер, к которому абоненты могут присоединять свои пакеты данных (см. рис. 4.15). Отсюда следует такое важное достоинство данной сети, как отсутствие конфликтов, но есть и недостатки, в частности необходимость контроля целостности

маркера и зависимость функционирования сети от каждого абонента (в случае неисправности абонент обязательно должен быть исключен из кольца).

Сеть Arcnet

Сеть Arcnet (или ARCnet от английского Attached Resource Computer Net, компьютерная сеть соединенных ресурсов) – это одна из старейших сетей. Она была разработана компанией Datapoint Corporation еще в 1977 году. Международные стандарты на эту сеть отсутствуют, хотя именно она считается родоначальницей метода маркерного доступа. Несмотря на отсутствие стандартов, сеть Arcnet до недавнего времени (в 1980 – 1990 г.г.) пользовалась популярностью, даже серьезно конкурировала с Ethernet. Большое количество компаний (например, Datapoint, Standard Microsystems, Xircom и др.) производили аппаратуру для сети этого типа. Но сейчас производство аппаратуры Arcnet практически прекращено.

Среди основных достоинств сети Arcnet по сравнению с Ethernet можно назвать ограниченную величину времени доступа, высокую надежность связи, простоту диагностики, а также сравнительно низкую стоимость адаптеров. К наиболее существенным недостаткам сети относятся низкая скорость передачи информации (2,5 Мбит/с), система адресации и формат пакета.

В качестве среды передачи в сети используется коаксиальный кабель с волновым сопротивлением 93 Ом, к примеру, марки RG-62A/U. Варианты с витой парой (экранированной и неэкранированной) не получили широкого распространения. Были предложены и варианты на оптоволоконном кабеле, но и они также не спасли Arcnet.

В качестве топологии сеть Arcnet использует классическую шину (Arcnet-BUS), а также пассивную звезду (Arcnet-STAR). В звезде применяются концентраторы (хабы). Возможно объединение с помощью концентраторов шинных и звездных сегментов в древовидную топологию (как и в Ethernet). Главное ограничение – в топологии не должно быть замкнутых путей (петель). Еще одно ограничение: количество сегментов, соединенных последовательной цепочкой с помощью концентраторов, не должно превышать трех.

Рис. 7.15. Топология сети Arcnet типа шина (B – адаптеры для работы в шине, S – адаптеры для работы в звезде)

Основные технические характеристики сети Arcnet следующие.

- Среда передачи – коаксиальный кабель, витая пара.
- Максимальная длина сети – 6 километров.
- Максимальная длина кабеля от абонента до пассивного концентратора – 30 метров.
- Максимальная длина кабеля от абонента до активного концентратора – 600 метров.
- Максимальная длина кабеля между активным и пассивным концентраторами – 30 метров.

• Максимальная длина кабеля между активными концентраторами – 600 метров.

• Максимальное количество абонентов в сети – 255.

• Максимальное количество абонентов на шинном сегменте – 8.

• Минимальное расстояние между абонентами в шине – 1 метр.

• Максимальная длина шинного сегмента – 300 метров.

• Скорость передачи данных – 2,5 Мбит/с.

В сети Arcnet используется маркерный метод доступа (метод передачи права), но он несколько отличается от аналогичного в сети Token-Ring. Ближе всего этот метод к тому, который предусмотрен в стандарте IEEE 802.4.

Сеть FDDI

Сеть FDDI (от английского Fiber Distributed Data Interface, оптоволоконный распределенный интерфейс данных) – это одна из новейших разработок стандартов локальных сетей. Стандарт FDDI был предложен Американским национальным институтом стандартов ANSI (спецификация ANSI X3T9.5). Затем был принят стандарт ISO 9314, соответствующий спецификациям ANSI. Уровень стандартизации сети достаточно высок.

За основу стандарта FDDI был взят метод маркерного доступа, предусмотренный международным стандартом IEEE 802.5 (Token-Ring). Несущественные отличия от этого стандарта определяются необходимостью обеспечить высокую скорость передачи информации на большие расстояния. Топология сети FDDI – это кольцо, наиболее подходящая топология для оптоволоконного кабеля. В сети применяется два разнонаправленных оптоволоконных кабеля, один из которых обычно находится в резерве, однако такое решение позволяет использовать и полнодуплексную передачу информации (одновременно в двух направлениях) с удвоенной эффективной скоростью в 200 Мбит/с (при этом каждый из двух каналов работает на скорости 100 Мбит/с). Применяется и звездно-кольцевая топология с концентраторами, включенными в кольцо (как в Token-Ring).

Основные технические характеристики сети FDDI.

• Максимальное количество абонентов сети – 1000.

• Максимальная протяженность кольца сети – 20 километров.

• Максимальное расстояние между абонентами сети – 2 километра.

• Среда передачи – многомодовый оптоволоконный кабель (возможно применение электрической витой пары).

• Метод доступа – маркерный.

• Скорость передачи информации – 100 Мбит/с (200 Мбит/с для дуплексного режима передачи).

Стандарт FDDI имеет значительные преимущества по сравнению со всеми рассмотренными ранее сетями. Например, сеть Fast Ethernet, имеющая такую же пропускную способность 100 Мбит/с,

Лекция 10. Стек протоколов TCP/IP

Internet — это глобальная информационная система, которая:

• логически связана единым адресным пространством;

• может поддерживать соединения с коммутацией пакетов на основе семейства специализированных протоколов;

• предоставляет услуги высокого уровня.

После запуска Советским Союзом искусственного спутника Земли в 1957 году Министерство обороны США посчитало, что на случай войны Америке нужна надёжная система передачи информации. Агентство передовых оборонных исследовательских проектов США (DARPA) предложило разработать для этого компьютерную сеть. Разработка такой сети была поручена Калифорнийскому университету в Лос-Анджелесе, Стэнфордскому исследовательскому центру, Университету штата Юта и Университету штата Калифорния в Санта-Барбаре. Компьютерная сеть была названа ARPANET (англ. *Advanced Research Projects Agency Network*), и в 1969 году в рамках проекта сеть объединила четыре указанных научных учреждения, все работы финансировались за счёт Министерства обороны США. Затем сеть ARPANET начала активно расти и развиваться, её начали использовать учёные из разных областей науки.

Типичный на сегодня доступ абонента телефонной сети общего пользования (ТФОП) в Интернет показан на рис. 1.1. Абонент автоматической телефонной станции (АТС) должен купить у провайдера сети Интернет (Internet Service Provider — ISP) карту с предоплатой, в которой указан телефонный номер провайдера для доступа в Интернет, имя пользователя (User ID) и пароль (Password). Эти данные абонент должен ввести в персональный компьютер. При установлении модемного соединения с телефонным номером провайдера компьютер соединится с сервером сетевого доступа (Network Access Server, NAS), который запросит у компьютера имя и пароль. Компьютер автоматически перешлет ему запрошенную информацию. После этого NAS запросит те же данные (имя и пароль) у сервера аутентификации, авторизации и учета (Authentication, Authorization, Accounting) и сравнит данные имени и пароля, полученные от абонента и от AAA-сервера. В случае их совпадения NAS откроет домашнюю страничку провайдера и начнет обслуживание запросов абонента. Для реализации запроса может потребоваться соединение через магистральную сеть (Backbone Network-BN), которая использует высокоскоростные (от 622 Мбит/с до 1.28 Гбит/с) каналы связи и высокопроизводительные маршрутизаторы (R) для объединения зоновых сетей различных провайдеров.

Стек протоколов Интернета

Стек протоколов сети Интернет2 был разработан до модели OSI. Поэтому уровни в стеке протоколов Интернета не соответствуют

аналогичным уровням в модели OSI. Стек протоколов Интернета состоит из пяти уровней:

1. физического,
2. звена передачи данных,
3. сети,
4. транспортного и
5. прикладного.

Первые четыре уровня обеспечивают физические стандарты, сетевой интерфейс, межсетевое взаимодействие и транспортные функции, которые соответствуют первым четырем уровням модели OSI. Три самых верхних уровня в модели OSI представлены в стеке протоколов Интернета единственным уровнем, называемым прикладным уровнем

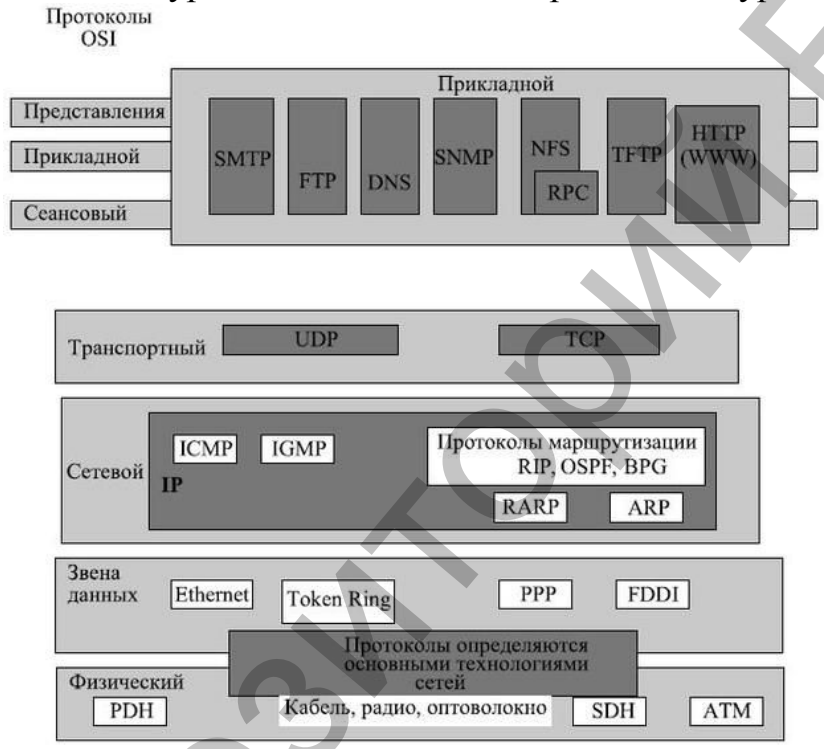


Рис. 1.3. Стек протоколов Интернета по сравнению с OSI

ARP	Address Resolution Protocol	Протокол нахождения адреса
BGP	Border Gateway Protocol	Протокол пограничной маршрутизации
DNS	Domain Name System	Система доменных имен
HTTP	Hyper Text Transfer Protocol	Протокол передачи гипертекста
FTP	File transfer Protocol	Протокол передачи файлов
ICMP	Internet Message Protocol	Протокол управляющих сообщений
IGMP	Internet Group Management Protocol	Протокол управления группами (пользователей) в Интернете
IP	Interworking Protocol	Межсетевой протокол

NFS	Network File System	Протокол сетевого доступа к файловым системам
OSPF	Open Shortest Path First	Открытый протокол предпочтения кратчайшего канала
RARP	Reverse Address Resolution Protocol	Протокол обратной конвертации адресов
RIP	Routing Information Protocol	Протокол обмена маршрутной информацией
RPC	Remote Procedure Call	Дистанционный вызов процедур
SMTP	Simple Transfer Protocol	Простой протокол электронной почты
SDH	Synchronous Digital Hierarchy	Синхронная цифровая иерархия
SNMP	Simple Network Management Protocol	Простой протокол электронной почты
TCP	Transmission Control Protocol	Протокол управления передачей
TFTP	Trivial File Protocol	Простейший протокол передачи данных
UDP	User Datagram Protocol	Дейтаграммный протокол пользователя
WWW	World Wide Web	Мировая паутина

Стек базовых протоколов Интернета — иерархический, составленный из диалоговых модулей, каждый из которых обеспечивает заданные функциональные возможности; но эти модули не обязательно взаимозависимые. В отличие от модели OSI, где определяется строго, какие функции принадлежат каждому из ее уровней, уровни набора протокола TCP/IP содержат относительно независимые протоколы, которые могут быть смешаны и согласованы в зависимости от потребностей системы. Термин иерархический означает, что каждый верхний протокол уровня поддерживается соответственно одним или более протоколами нижнего уровня.

На транспортном уровне стек определяет два протокола: протокол управления передачей (TCP) и протокол пользовательских дейтаграмм (UDP). На сетевом уровне — главный протокол межсетевого взаимодействия (IP), хотя на этом уровне используются некоторые другие протоколы, о которых будет сказано ниже.

Физический уровень

На физическом уровне и уровне звена передачи данных стек протоколов TCP/IP не отдает предпочтения ни одному протоколу. Он поддерживает все стандартные и частные протоколы передачи по кабелю, оптоволоконному кабелю и радиоканалам, которые определяются сетевыми технологиям на этом уровне (PDH — Plesiochronous Digital Hierarchy, SDH — Synchronous

Digital Hierarchy, ATM — Asynchronous Transfer Mode и другими).

Канальный уровень (звена передачи данных)

На этом уровне коммутационные устройства используют различные технологии: Ethernet, Token Ring, FDDI, PPP и другие, большинство из которых рассмотрено в предыдущих лекциях.

Интернет предназначен для транспортировки любого вида информации от источника к получателю. В транспортировке информации участвуют различные элементы сети (см. рис. 1.1) – оконечные устройства, коммутационные устройства и серверы. Группы узлов при помощи коммутационных устройств объединяются в локальную сеть, локальные сети соединяются между собой шлюзами (маршрутизаторами).

Узлы, с точки зрения сети, представляют собой источники и получатели информации. Четыре нижних уровня в совокупности независимы от вида передаваемой информации. Каждое приложение, связывающееся с четвертым уровнем, идентифицируется своим уникальным номером порта. Номера портов занимают диапазон от 0 до 65535. В этом диапазоне номера портов 0-1023 выделены под общесетевые приложения (well-known ports), номера портов 1024-49151 используются разработчиками специализированного программного обеспечения, номера портов 49152-65535 динамически закрепляются за приложениями пользователей на время сеанса связи. Численные значения номеров портов стека приведены в [38].

Транспортный уровень

На транспортном уровне TCP/IP определяет два протокола: протокол управления передачей (TCP) и протокол пользовательских дейтаграмм (UDP).

UDP и TCP — транспортные протоколы уровня, которые отвечают за доставку сообщения от процесса (функционирующей программы) к другому процессу.

Протокол пользовательских дейтаграмм (UDP – User Datagram Protocol) — наиболее простой из двух стандартных транспортных протоколов TCP/IP. Он выполняет функции передачи между прикладными уровнями разных рабочих станций, по адресу порта, контролирует ошибки по контрольной сумме и передает информацию верхним уровням.

Протокол управления передачей (TCP – Transmission Control Protocol) обеспечивает полные услуги транспортного уровня к приложениям. TCP — достоверный транспортный протокол потока, ориентированный на дуплексный режим связи с установлением логического соединения. Для этого каждый передаваемый пакет снабжается порядковым номером, и правильный его прием должен быть подтвержден приемной стороной. В этом контексте термин поток означает передачу данных, рассчитанную на то, что соединение должно быть установлено между обоими концами передачи прежде, чем начнется передача данных. Протокол TCP имеет код протокола 6 (в шестнадцатеричном коде – 0x06) и используется для гарантированной транспортировки информации.

Сетевой уровень

На сетевом уровне (или, более точно, межсетевом уровне) TCP/IP поддерживает протокол межсетевого взаимодействия (IP). IP, в свою очередь, содержит четыре протокола поддержки: протокол определения адреса (ARP — Address Resolution Protocol), протокол определения сетевого адреса по местоположению (RARP – Reserve Address Resolution Protocol), протокол управляющих сообщений Internet – (ICMP — Internet Control Message Protocol) и межсетевой протокол управления группами (IGMP – Internet Group Message Protocol). На этом же уровне применяются протоколы маршрутизации: протокол обмена маршрутной информацией (RIP — Routing Information Protocol), "открыть кратчайший путь первым" (OSPF — Open Shortest Path First), протокол пограничной маршрутизации (BPG — Border Gateway Protocol).

Протокол межсетевого взаимодействия (IP)

Протокол межсетевого взаимодействия (IP) — механизм передачи, используемый протоколами TCP/IP. Это ненадежная служба доставки дейтаграммы без установления соединения, но с "максимальными усилиями" (best-effort).

Термин с "максимальными усилиями" означает, что делается все возможное (максимальные усилия), чтобы передать информацию к ее пункту назначения, но IP не обеспечивает никакой проверки ошибок или их отслеживания. IP предполагает ненадежность основных уровней, без гарантий требуемого уровня сервиса.

IP транспортирует данные в пакетах, называемые дейтаграммами, каждая из которых транспортируется отдельно. Дейтаграммы могут перемещаться по различным маршрутам и могут прибыть не в исходной последовательности или быть дублированы. IP не сохраняет копию маршрутов и не имеет никаких средств для того, чтобы переупорядочить дейтаграммы, как только они достигают пункта назначения.

Ограниченные функциональные возможности IP, однако, нельзя считать слабостью. IP обеспечивает "чистые" функции передачи, которые освобождены от пользовательских особенностей, и предполагает, что на других уровнях будут добавлены те средства, которые необходимы для данного приложения, и таким образом будет достигнута максимальная эффективность.

Протокол определения адресов (ARP – Address Resolution Protocol) используется, чтобы связать адрес IP с физическим адресом. На типичной физической сети, типа локальной сети LAN (Local Area Network), каждое устройство на линии связи идентифицировано физическим адресом или адресом станции, обычно закрепленным в сетевой карте интерфейса (NIC – Network Interface Card). ARP используются, чтобы найти физический адрес узла, когда известен его адрес в сети Интернет.

Обратный протокол определения адресов (RARP — протокол определения сетевого адреса по местоположению) позволяет хосту

обнаруживать его адрес в сети Интернет, когда хост знает только свой физический адрес. Он используется, когда компьютер связывается с сетью впервые или когда компьютер загружается без диска.

Протокол управляющих сообщений Интернета (ICMP – Internet Control Message Protocol) — механизм, используемый хостами и шлюзами, чтобы передать извещение о дейтаграммных проблемах назад к передатчику.

Межсетевой протокол управления группами (IGMP – Internet Group Message Protocol) – обслуживает одновременную передачу сообщения к группе получателей.

Протокол пограничной маршрутизации (BGP — Border Gateway Protocol) — протокол маршрутизации между автономными системами, основанный на применении вектора пути.

Протокол обмена маршрутной информацией (RIP — Routing Information Protocol) — протокол маршрутизации, основанный на использовании алгоритма вектора расстояний.

"Открыть кратчайший путь первым" (OSPF — Open Shortest Path First) — внутрисетевой протокол маршрутизации, основанный на анализе состояния линий связи.

Прикладной уровень TCP/IP

Прикладной уровень в стеке протоколов Интернета эквивалентен объединению сеансового, представительского и прикладного уровня в модели OSI. На рис. 1.3 показаны следующие протоколы прикладного уровня:

SMTP (Simple Mail Protocol) – простой почтовый протокол. Он поддерживает передачу почтовых электронных сообщений по сети Интернет. Протокол называется простым, потому что обеспечивает передачу информации пользователям, готовым к немедленной доставке. Передача осуществляется в режиме 7-битовых слов. Он требует наличия программ перехода от принятого в большинстве программ формата с 8-разрядными словами к формату с 7-разрядными словами.

Система поддерживает:

• посылку одиночных сообщений одному или более получателям;

• посылку сообщений, включающих в себя текст, голосовые сообщения, видео или графические материалы.

Протокол передачи файлов (FTP — File Transfer Protocol) используется для передачи файлов от одного компьютера к другому. Обеспечивает просмотр каталогов удаленного компьютера, копирование, удаление и пересылку файлов. FTP отличается от других протоколов тем, что устанавливает два соединения между хостами. Одно используется для передачи информации, а другое — для управления передачей.

DNS (Domain Name System) – служба доменных имен. Она осуществляет присвоение уникальных имен всем пользователям и узлам сети Интернет и устанавливает логическую связь с их сетевыми адресами. Доменное имя представляется иерархической структурой, имеющей несколько уровней.

Типовые имена доменов верхнего уровня закреплены следующим образом:

.com – коммерческие организации;

.gov – правительственные учреждения;

.org – некоммерческие организации;

.net — центры поддержки сети;

.int – международные организации;

.mil – военные структуры.

SNMP (Simple Network Management Protocol) — простой протокол управления сетью. Он обеспечивает набор фундаментальных действий по наблюдению и обслуживанию Интернета.

Протокол разработан так, чтобы он мог контролировать устройства, созданные различными изготовителями и установленные на различных физических сетях. Другими словами, SNMP освобождает задачи управления от учета физических характеристик управляемых устройств и от основной технологии организации сети.

Сетевая файловая система (NFS — Network File System). Это один из многих протоколов (например, на рисунке показан еще один протокол RPC – Remote Procedure Call – вызов удаленной процедуры), который позволяет использование файлов, содержащих процедуры управления и периферии в другом компьютере.

Тривиальный (простейший) протокол передачи файлов TFTP (Trivial File Transfer Protocol). Используется в простых случаях при начальной загрузке рабочих станций или загрузке маршрутизаторов, не имеющих внешней памяти.

Протокол передачи гипертекста (HTTP — Hyper Text Transfer Protocol) — транспортный протокол, который применяется в Интернете при обмене документами, представленными на языке описания гипертекстовых документов.

Язык разметки гипертекста (HTML — Hyper Text Markup Language). Является одним из главных языков, используемых в сети WWW.

Мировая паутина (WWW — World Wide Web) – глобальная гипертекстовая информационная система. Она объединяет огромное количество документов, хранящихся во многих странах мира и доступных через сеть узлов в сети Интернет, которые связаны между собой каналами связи.

Связь между уровнями стека протоколов сети Интернет и адресацией

В сети Интернет используются три различных уровня адресов: физический адрес (линия связи), интернет-адрес (IP) и адрес порта (рис. 1.4).

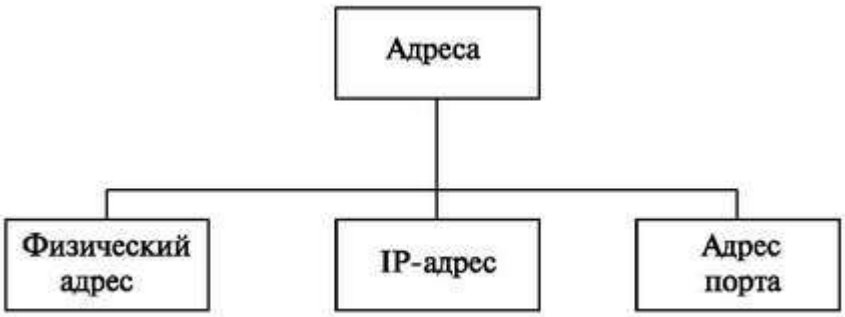
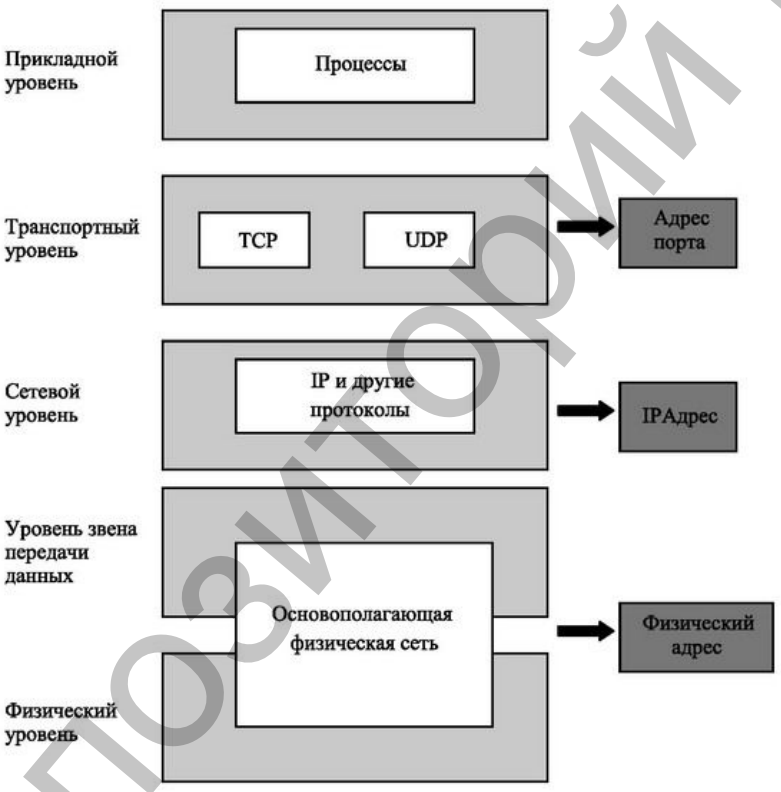


Рис. 1.4. Адреса TCP/IP

Каждый адрес принадлежит заданному уровню TCP/IP-архитектуры, как это показано на (рис. 1.5).



Физический адрес

Физический адрес (Media Access Control — MAC-адрес) используется для установления соединения в локальной сети (подсети). Этот адрес совпадает с номером сетевого адаптера (сетевой карты) компьютера и жестко устанавливается заводом-изготовителем из пула (диапазона) отведенных ему адресов. Записывается в виде шестнадцатеричных чисел, разделенных двоеточием, например, 08:00:06:3F:D4:E1, где первые три значения определяют фирму-производителя (00:10:5a:xx:xx:xx – 3Com, 00:03:ba:xx:xx:xx – Sun, 00:01:e3:xx:xx:xx – Siemens), а последующие – порядковый номер узла.

Компьютер может иметь несколько сетевых карт и, соответственно, несколько MAC-адресов. При замене аппаратуры изменяется и MAC-адрес, поэтому их использование в качестве сетевых адресов неудобно.

Физический адрес индивидуальной передачи, при групповой рассылке и при широковещательной передаче

Физические адреса могут быть либо индивидуальные (один единственный получатель) и групповые (группа получателей), либо широковещательные (для получения всеми системами в сети). Некоторые сети поддерживают все три типа адресов. Например, Локальная сеть Ethernet поддерживает однонаправленные физические адреса (6 байт), адреса групповой рассылки и широковещательные адреса. Некоторые сети не поддерживают групповую рассылку или широковещательно передают физические адреса. Если кадр нужно передать группе получателей или системе для всей системы, адрес групповой рассылки или широковещательный адрес должен моделироваться, используя однонаправленные адреса. Это означает, что множество пакетов рассылаются, используя однонаправленные адреса.

Интернет-адрес

Адреса Интернета необходимы для универсальных служб связи, которые не зависят от основных физических сетей. Физические адреса не адекватны в межсетевой среде, где различные сети могут иметь различные форматы адреса. Необходима универсальная система адресации, в которой каждый хост может быть идентифицирован уникально, независимо от основной физической сети.

Для этой цели применяются IP-адреса. Интернет(IP)-адрес в настоящее время состоит из 32 бит. Он может уникально определить хост, подключенный к сети Интернет. Никакие два хоста на сети Интернет не могут иметь один и тот же самый IP-адрес.

Адрес порта

Адрес IP и физический адрес необходимы для порции данных, перемещающихся от источника до хоста пункта назначения. Однако прибытие в хост пункта назначения — не конечная цель обмена сообщениями данных в Интернете. Система, которая передает только данные от одного компьютера до другого, не может считаться законченной. Сегодня компьютеры — устройства, которые могут выполнить множество процессов в одно и то же время. Конечная цель сети Интернет — коммутация процесса, работающего с другим процессом. Например, компьютер А общается с компьютером С, используя TELNET. В то же самое время компьютер А общается с компьютером с использованием протокола передачи файлов FTP. Для этих процессов, возникающих одновременно, нам надо иметь метод, позволяющий маркировать различные процессы.

Другими словами, процессы нуждаются в адресах. В архитектуре TCP/IP метка, назначаемая процессу, названа адресом порта. Адрес порта в TCP/IP — 16 битов длиной.

Символьные (доменные) адреса

Символьные (доменные) адреса предназначены для людей. Для работы в больших сетях символьные адреса имеют сложную иерархическую структуру, содержащую имя пользователя, имя подсети (поддомена), символьное имя страны или организации (домена). Например, адрес Ivan.Sidorov@sk.sut.ru обозначает, что адресат (Иван Сидоров) находится в подсети sk сети sut в России - ru (впрочем, возможно и не в России), а адрес www.protocols.com – адрес домена коммерческой организации (как правило коммерческой).

Лекция 11. Протокол IP. Типы адресации.

На сетевом уровне (или IP) мы должны уникально идентифицировать каждое устройство в Интернете, чтобы обеспечить глобальную связь между всеми устройствами.

Установление соединения между двумя и более узлами происходит на основе обработки адресной информации, которая по мере необходимости обрабатывается устройствами 3-го уровня в маршрутизаторах. К адресу предъявляются следующие требования:

- адрес должен быть универсальным;
- адрес должен иметь иерархическую структуру, удобную для обработки соответствующими узлами;
- адрес должен быть удобен для пользователя.

Идентификатор, используемый на уровне IP набора протокола TCP/IP, чтобы идентифицировать каждое устройство, подключенное к Интернету, назван адресом Интернета, или адресом IP. *Адрес IP — двоичный адрес на 32 бита, который уникально и универсально определяет подключение хоста или маршрутизатора к Интернету*.

Адреса IP уникальны. Они уникальны в том смысле, что каждый адрес определяет одно и только одно подключение к Интернету. Два устройства в Интернете никогда не могут иметь одного того же адреса. Если устройство имеет два подключения к Интернету, через две сети, оно имеет два адреса IP.

Адреса IP универсальны потому, что система адресации должна быть принята любым хостом, который хочет быть связанным с Интернетом.

Адресное пространство

Протокол, подобный IP, то есть определяющий адреса, имеет адресное пространство. Адресное пространство — общее количество адресов, применяемых в соответствии с протоколом. Если протокол использует N бит, чтобы определить адрес, адресное пространство – 2^N, потому что каждый бит может иметь два различных значения (0 и 1), а N бит могут иметь 2^N значений.

IPv4 использует адреса на 32 бита, то есть адресное пространство – 2^{32} или 4,294,967,296 (больше чем четыре миллиарда). Это означало бы, что, теоретически, если не было бы никаких ограничений, к Интернету могли бы быть подключены более чем 4 миллиарда устройств. Мы вскоре увидим, что фактически номеров намного меньше.

Система обозначений

Применяется, чтобы указать адрес IP. Есть три общих системы обозначений: двоичная, десятичная разделенная точками и шестнадцатеричная система обозначений.

Двоичная система обозначений

В двоичной системе обозначений адрес IP отображен как 32 бита. Чтобы сделать адрес читаемым, обычно вставляются один или более пробелов между каждым октетом (8 бит). Каждый октет часто упоминается как байт.

Поэтому адрес IP называется адресом с 4 октетами, или 4-байтовым адресом. Ниже показан пример адреса IP в двоичной системе обозначений:

10010001 11011101 01010101 10010100

Десятичная разделенная точками система обозначений

Чтобы сделать адрес IP более компактным и более простым для чтения, адреса Интернета обычно написаны в десятичной форме с точкой (точечное отделение байтов). показывает адрес IP в десятичной системе обозначений, разделенной точками. Обратите внимание на то, что поскольку каждый байт (октет) — только 8 битов, каждый номер в десятичной разделенной точками системе обозначений находится между 0 и 255.

Адресация по классам

В начале внедрения IP-адресации использовали концепцию классов. Эта архитектура названа адресацией по классам.

Имеется пять классов адресов, где жирным шрифтом выделена старшая часть IP-адреса, указывающая номер сети.

В версии 4 сетевые IP-адреса имеют двухуровневую иерархию, старшая часть которых отображает номер сети (netid), а младшая – номер узла (компьютера) в сети (hostid). Общая длина адреса имеет длину 4 байта и записывается в виде десятичных чисел, разделенных точками. Первые биты сетевого адреса задают класс адреса, по которому определяется, какая его часть относится к номеру сети, а какая – к номеру узла. Если сеть является частью Интернета, то номер сети назначается централизованно по рекомендации специального органа Интернета – Internet Information Center. Номер узла в IP-адресе назначается из разрешенного для этого класса диапазона независимо от физического адреса. Маршрутизатор объединяет несколько сетей, поэтому каждый порт (интерфейс) маршрутизатора имеет свой IP-адрес.

Класс	**Первые биты IP-адреса**	**Наименьший номер сети**	**Наибольший номер сети**	**Максимальное число сетей**	**Макс. число узлов в каждой сети**
A	0	0.0.0.0	127.0.0.0	$2^7 - 2$	$2^{24} - 2$
B	10	128.0.0.0	191.255.0.0	$2^{14} - 2$	$2^{16} - 2$
C	110	192.0.0.0	223.255.255.0	$2^{21} - 2$	$2^8 - 2$
D	1110	224.0.0.0	239.255.255.255	15×2^{24}	Групповые адреса
E	11110	240.0.0.0	255.255.255.255	7×2^{24}	Резерв

Большие сети используют адреса класса A, средние – класса B, маленькие – класса C.

В IPv4 существуют определенные соглашения об использовании адресов:

1. В каждом классе имеется диапазон адресов для локального использования, которые сетевые маршрутизаторы не обрабатывают ни при каких условиях, — они применяются для маршрутизации в локальных сетях. В классе A – это сеть 10.0.0.0, в классе B – диапазоны сетей от 172.16.0.0 до 172.31.0.0, в классе C – диапазон сетей от 192.168.0.0. до 192.168.255.255.

2. Если в поле номера сети установлены все двоичные "0", то пакет адресован соответствующему узлу той же сети.

3. Если в полях номера сети и номера узла установлены все двоичные "1", то пакет адресован всем узлам той же сети (широковещательная рассылка, limited broadcast).

4. Если в поле номера узла установлены все двоичные "1", то пакет адресован всем узлам соответствующей сети.

5. Основное назначение групповых (Multicast) адресов – распространение информации по схеме "один-ко-многим" для групповой рассылки в Интернете аудио- и видеоинформации.

6. Адреса класса E зарезервированы для будущих применений.

Сетевой (Netid) и локальный (Hostid) адреса

При адресации по классам адрес IP в классах A, B и C разделен на сетевой (Netid) и локальный (Hostid) адреса. Длина адреса зависит от класса объекта. Обращаем внимание на то, что классы D и E не разделены на эти части.

В классе A 1 байт определяет сетевой адрес и 3 байта определяют локальный адрес. В классе B 2 байта определяют сетевой адрес и 2 байта — локальный. В классе C 3 байта определяют сетевой адрес и 1 байт — локальный.

Классы и блоки

При адресации по классам каждый класс разделен на фиксированное число блоков, и каждый блок имеет фиксированный размер. Давайте рассмотрим каждый класс.

Класс A разделяется на 128 блоков, где каждый блок имеет различный netid. Первый блок охватывает адреса от 0.0.0.0 до 0.255.255.255 (netid 0), второй блок — адреса от 1.0.0.0 до 1.255.255.255 (netid 1), последний блок — адреса от 127.0.0.0 до 127.255.255.255 (netid 127). Обратите внимание, что для каждого блока адресов первый байт (netid) является тем же самым, но другие 3 байта (hostid) могут принимать любое значение в данном диапазоне.

Пример 9

Дан сетевой адрес 17.0.0.0, найдите класс, блок и диапазон адресов.

Решение

Класс — A, потому что первый байт — между 0 и 127. Блок имеет сетевой номер 17. Адреса располагаются от 17.0.0.0 до 17.255.255.255.

Пример 11

Дан сетевой адрес 220.34.76.0, найдите класс, блок и диапазон адресов.

Решение

Класс — С, потому что первый байт — между 192 и 223. Блок имеет сетевой номер 220.34.76. Адреса располагаются от 220.34.76.0 до 220.34.76.255.

Маска

Уже давно наблюдается дефицит IP-адресов, который обусловлен не только ростом числа пользователей, но и необходимостью выделения IP-адресов на каждый порт маршрутизатора. Имеется несколько подходов смягчения этой проблемы, в том числе за счет использования масок.

Традиционно номер сети и узла определяется в зависимости от класса адреса. Однако наличие только четырех классов адресов часто бывает неудобно. Например, администратор получил от поставщика услуг номер сети 135.38.0.0 (адрес класса В, двоичный код сети – 10000111 00100110 00000000 00000000). В такой сети потенциально можно иметь 65 534 узла, но такое количество узлов администратору не нужно, ему достаточно иметь 32 000. Проблема решается с помощью масок. Количество "единиц" в маске показывает число старших разрядов, которые определяют номер сети. Для нашего случая следует выбрать маску со значением 255.255.192.0 (двоичный код 11111111 11111111 11000000 00000000). В результате наложения маски на сетевой адрес получается четыре подсети: 135.38.0.0; 135.38.64.0; 135.38.128.0; 135.38.192.0. (табл. 2.2).

Таблица 2.2. Результат наложения маски

Номер сети	**Число узлов в подсети**
10000111 00100110 00000000 00000000 135.38.00.0	16382
10000111 00100110 01000000 00000000 135.38.64.0	16382
10000111 00100110 10000000 00000000 135.38.128.0	16382
10000111 00100110 11000000 00000000 135.38.192.0	16382

Две полученные подсети с общим количеством узлов 32764 = 2 х 16382 администратор использует для своих нужд, а остальные может отдать другому администратору.

Для адресации по классам есть три маски. Для класса А маска – из восьми "единиц" и двадцати четырех "нулей" (255.0.0.0). Для класса В маска – шестнадцать "единиц" и шестнадцать "нулей" (255.255.0.0). На класс С маска — двадцать четыре "единицы" и восемь "нулей" (255.255.255.0). "Единицы" сохраняют сетевой адрес (netid); "нули" устанавливают локальный адрес (hostid) на "0".

Пример 12

Дан адрес 23.56.7.91 и заданный по умолчанию класс маски А; найдите начальный адрес (сетевой адрес).

Решение

Заданная по умолчанию маска — 255.0.0.0, что означает, что только первый байт сохраняется, а другие 3 байта устанавливаются на "нуль". Сетевой адрес — 23.0.0.0.

Исключительные номера

Прямой широковещательный адрес

В классах A, B и C, если сетевой номер (hostid) состоит только из единиц, адрес называется прямым широковещательным адресом. Такой адрес используется маршрутизатором для того, чтобы передать пакеты для всех хостов в заданной сети. Все хосты примут пакет, имеющий этот тип адреса пункта назначения. Заметим, что введение такого адреса уменьшает число номеров локальных адресов в каждом из классов A, B, C.

x.255.255.255 A

Этот хост на этой сети

Если адрес IP составлен из всех нулей, это означает, что это хост расположенный на этой сети. Он используется хостом во время начальной загрузки, когда передающий хост не знает свой адрес IP. Хост передает пакет IP серверу начальной загрузки, используя этот адрес как исходный адрес пункта назначения, который может определить собственный адрес передающего хоста. Его можно использовать только как исходящий адрес. Этот адрес – всегда адрес класса A независимо от сети, что уменьшает число сетей в классе A на одну сеть.

0.0.0.0 A

Заданный хост на этой сети

Адрес IP с сетевым номером (netid) из всех нулей означает адрес заданного хоста на этой сети. Он используется хостом, чтобы передать сообщение другому хосту на той же самой сети. Заметим, что он может использоваться только как входящий адрес пункта назначения. Это адрес класса A независимо от размера сети

Адрес кольцевой проверки (Loopback address)

Адрес IP с первым байтом, равным 127, применяется для адреса кольцевой проверки, который является адресом, используемым для проверки программного обеспечения компьютера. Когда этот адрес задействуется, пакет никогда не покидает хост; он просто возвращается хосту и обрабатывается согласно протоколу и установленному на управляющем компьютере программному обеспечению. Он может использоваться, чтобы проверить программное обеспечение IP.

Частные адреса

Некоторое количество блоков в каждом классе предназначено для частного использования. Эти блоки изображены в таблице 2.4 Эти адреса используются при подключении к сетям с другими способами адресации.

Таблица 2.4. Специальные адреса

Класс	Сетевой адрес	Величина блока
A	10.0.0	1

B	172.16 до 172.31	16
C	192.168.0 до 192.168.255	256

IP-адресация в версии 6

Принципиальным решением устранения дефицита сетевых адресов является расширение адресного поля с 4-байтного (текущей версии IPv4) на 16-байтное (в новой версии IPv6) [56, 58, 66, 67, 70]. Адреса присваиваются интерфейсам; если к интерфейсу подключен один узел, то его адрес совпадает с адресом узла. Адресация использует бесклассовую систему междоменной маршрутизации (Classless Inter Domain Routing, CIDR).

Предложено три формы представления адреса:

- Предпочтительная форма (шестнадцатеричная система счисления с двоеточием) – восемь 16-битовых шестнадцатеричных чисел, разделенных двоеточием. Например, FEDC:BA98:7654:3210:FEDC:BA98:7654:3210.
- Сжатая форма – запись длительной последовательности "0" путем введения двойного двоеточия. Двоеточие допускается использовать только в одном месте адреса. Например, адрес 1080:0:0:0:8:800:200C:417A может быть представлен в виде 1080::8:800:200C:417A.
- Смешанная форма – шесть старших чисел (96 бит) записываются в предпочтительной форме, а младшие числа (32 бита) представляются в виде, принятом в IPv4. Например, 0:0:0:0:0:0:13.1.68.3, или в сжатой форме ::13.1.68.3.

Длина префикса записывается в конце адреса и отделяется от него косой линией. Например, адрес 12AB:0000:0000:CD30:0000:0000:0000:0000/60 содержит 60 битовый префикс 12AB:0000:0000:CD3, или в сжатой форме 12AB:0:0:CD30::/60.

В версии IPv6 имеется три типа адресов.

Индивидуальный адрес (Unicast address). Индивидуальные адреса определяют единственный интерфейс. Пакет, посланный по такому адресу, доставляется на указанный интерфейс. Для облегчения функции маршрутизации индивидуальный адрес формируются в виде составного адреса (AGUA — Aggregatable Global Unicast Address).

Альтернативный адрес (Anycast address). Адрес такого типа относится к различным узлам, но имеет один и тот же адресный префикс. Например, все компьютеры соединены одной и той же физической сетью, использующей один и тот же префикс. Пакет, посланный по альтернативному адресу, должен быть доставлен точно к одному из участников группы ближайшим или наиболее доступным маршрутом, который имеет наименьшую метрику. Транспортировка пакета предполагается по фиксированному пути путем создания стека выборочных адресов.

Широковещательный адрес (Multicast address). Широковещательный адрес определяет группу интерфейсов. Каждый участник этой группы может

иметь или не иметь один префикс. Участники могут быть или не быть подключены к одной и той же физической сети. Пакет, посланный по широковещательному адресу, доставляется каждому участнику этой группы.

Имеются специальные адреса. Не специфицированный адрес 0:0:0:0:0:0:0:0 используется при запросе назначения адреса, адрес шлейфа (loopback) 0:0:0:0:0:0:0:1 — для посылки пакета самому себе. Эти и еще некоторые типы адресов уже упоминались в предыдущем разделе.

1. Найти сетевой адрес (netid) и локальный адрес (hostid) следующих адресов IP:
   - 114.34.2.8;
   - 19.34.21.5;
   - 23.67.12.1;
   - 127.23.4.0.
2. Найти сетевой адрес (netid) и локальный адрес (hostid) следующих адресов IP:
   - 129.14.6.8;
   - 132.56.8.6;
   - 171.34.14.8;
   - 190.12.67.9.
3. Дан сетевой адрес 220.34.76.0, найдите класс, блок и диапазон адресов.
4. Дан сетевой адрес 132.21.0.0, найдите класс, блок и диапазон адресов
5. Дан адрес 201.180.56.5 и маска класса C, заданная по умолчанию; найдите начальный адрес (сетевой адрес).
6. Дан адрес 132.6.17.85 и задана по умолчанию маска класса B; найдите начальный адрес (сетевой адрес).

Лекция 12. Система доменных имен

Для идентификации объекта протоколы используют IP-адреса, которые уникально идентифицирует соединения хоста с Интернетом. Однако люди предпочитают имена адресам. Поэтому нам необходима система, которая сопоставляет имя с адресом или адрес к имени.

Соответствие между доменными именами и IP-адресами может устанавливаться как средствами локального хоста, так и средствами централизованной службы. На раннем этапе развития Internet на каждом хосте вручную создавался текстовый файл с известным именем hosts. Этот файл состоял из некоторого количества строк, каждая из которых содержала одну пару «IP-адрес - доменное имя», например 102.54.94.97 - rhino.acme.com.

По мере роста Internet файлы hosts также росли, и создание масштабируемого решения для разрешения имен стало необходимостью.

- До появления DNS соответствие между символьными именами и IP-адресами можно было установить в специальном файле. Этот способ можно использовать и сейчас.
  - Windows:
    WinDir\system32\drivers\etc\hosts
  - UNIX
    /etc/hosts
- Файл hosts содержит строки, Каждая из которых определяет одно соответствие между именем и IP-адресом
  - 127.0.0.1 localhost
  - 192.168.0.1 mygate

Таким решением стала специальная служба - система доменных имен (Domain Name System, DNS). *DNS - это централизованная служба, основанная на распределенной базе отображений «доменное имя - IP-адрес».* Служба DNS использует в своей работе протокол типа «клиент-сервер». В нем определены DNS-серверы и DNS-клиенты. DNS-серверы поддерживают распределенную базу отображений, а DNS-клиенты обращаются к серверам с запросами о разрешении доменного имени в IP-адрес.

Пространство доменных имен

Имеет структуру инвертированного дерева с корнем в вершине. Дерево может иметь 128 уровней: от уровня 0 (корень) до уровня 127. Принимая во внимание, что корень скрепляет целое дерево вместе, каждый уровень дерева определяет иерархический уровень, и называется меткой. Рисунок – дерево.

Доменное имя

Каждый узел дерева имеет доменное имя. Полное доменное имя — последовательность строк, отделенных точками (.). Доменные имена всегда читают от узла к корню.

Последняя строка — это метка-корень (нуль). Это означает, что полное доменное имя всегда оканчивается нулевой отметкой, которую означает последний символ – точка, потому нулевая строка ничего не обозначает.

Полностью определенное доменное имя

Если метка завершается нулевой строкой, это называется "полностью определенное доменное имя" – имя хоста, которое содержит полное имя хоста. Оно включает в себя все метки, от наиболее специфичной до наиболее общей, которые уникально определяют имя хоста. Например, доменное имя

`kafedra.gut.edu.`

Это FQDN компьютера, названного kafedra и установленного в Государственном университете телекоммуникаций. Заметим, что имя должно заканчиваться нулевым ярлыком, но поскольку он ничего не обозначает, метка заканчивается точкой (.).

Частично определенное имя домена

Если метка не заканчивается нулевой строкой, это называется "частично определенным доменными именем" - начинается от узла, но не достигает корня. Оно используется, если в компьютере будет отмечено, что имя принадлежит тому же самому сайту, что и клиент. Здесь компьютер может заменить отсутствующую часть так называемым суффиксом. Это применяется тогда, когда пользователь сайта sut.edu. хочет иметь IP-адрес компьютера "kafedra", он может определить частичное имя

`kafedra`

DNS клиент добавляет суффикс sut.edu перед тем, как передать адрес к DNS-серверу.

Домен

Домен — это фрагмент дерева в пространстве доменных имен. Имя домена – это доменное имя узла на вершине поддерева.

Распределение имен

Для хранения доменной информации используются сервера. Но и один сервер не может в одиночку решать эту проблему.

Иерархия серверов имен

Решение этих проблем – распределить информацию по компьютерам, называемым DNS-серверы. Один из путей сделать это – разделить полное пространство на много доменов, базирующихся на первом уровне. Другими словами, считать корень автономным и создавать и предоставить полномочия, создавать столько доменов (поддеревьев), сколько имеется узлов. Поскольку домен, создаваемый таким способом, очень большой, DNS позволяет разделить домен на более мелкие домены (поддомены). Каждый сервер может обслуживать (уполномочен) любой большой или маленький домен. Другими словами, мы имеем *иерархию серверов в соответствии с иерархией имен*.

Зона

Zone (зона) –непрерывное пространство имен (домен, возможно, за исключением некоторых или всех поддоменов.

Если сервер назначен отвечать за домен и домен не разделен на поддомены, "домен" и "зона" относятся к одним и тем же понятиям. Сервер создает базу данных, называемую файлом зоны, и сохраняет всю информацию для всех узлов под этим доменом. Однако если сервер разделяет свои домены на поддомены и делегирует часть своих полномочий другому серверу, "домен" и "зона" относятся к различным понятиям.

Корневой сервер

Корневой сервер – это сервер, зона которого состоит из полного дерева. Корневой сервер обычно не накапливает информацию о домене, но делегирует свои полномочия другому серверу, сохраняя ссылки на полное пространство имен. Серверы распределены по всему миру.

Первичные и вторичные серверы

DNS определяет два типа серверов: первичные и вторичные. **Первичный сервер** — это сервер, накапливающий файл о зоне, на которую он имеет полномочия. Он несет ответственность за создание, эксплуатацию и изменения зонового файла. Зоновый файл накапливается на локальном диске.

Вторичный сервер – это сервер, который передает полную информацию о зоне для других серверов (первичных или вторичных) и накапливает файл на своем локальном диске. Вторичный сервер не создает и не изменяет зоновый файл. Если изменение требуется, он должен сделать это с помощью первичного сервера, который посылает измененную версию на вторичный.

Вторичный сервер – это сервер, который передает полную информацию о зоне для других серверов (первичных или вторичных) и накапливает файл на своем локальном диске. Вторичный сервер не создает и не изменяет зоновый файл. Если изменение требуется, он должен сделать это с помощью первичного сервера, который посылает измененную версию на вторичный.

DNS в Интернете

DNS – это протокол, который может быть использован в различных платформах. В Интернете пространство доменных имен (дерево) разделяется на три различных секции: родовой домен, домен страны и инверсный домен.

Родовой домен

Родовой домен определяет регистрацию хоста (generic domain) в соответствии с его родовой природой. Эти уровни связаны с типами организаций, как это, например, приведено для США.

Каждый узел дерева — домен, который является частью базы пространства доменных имен.

Домены страны

Секция **домены страны** придерживается того же формата, что и родовые домены, но использует двухсимвольные сокращения страны (например, ru для России) вместо трехсимвольной организационной

структуры первого уровня. Аббревиатуры второго уровня могут быть организационными или могут более детально определять национальную принадлежность. Россия (ru), например, использует аббревиатуры отдельных городов (например, spb.ru). Адрес gut.spb.ru может быть расшифрован как Государственный университет телекоммуникаций, Санкт-Петербург, Россия.

Разрешение имен

Отображение имени в адрес или адреса в имя называется "Разрешение имен".

DNS-resolver (ресолвер DNS) – программное обеспечение, обеспечивающее разрешение адресов посредством выполнения запросов к DNS-серверам.

Обычно реализуется в виде библиотечных функций, но может выполняться в служебной программе. Может обращаться к локальному (выполняющемуся на том же узле, что и ресолвер) или удаленному серверу DNS

Протоколы DNS разработаны как приложение сервер-клиент. Хост, который нуждается в отображении адреса в имя или имени в адрес, вызванного DNS клиента, называется ресолвер DNS. Ресолвер DNS получает доступ к ближайшему серверу DNS с запросом на отображение. Если сервер имеет информацию, он выполняет запрос распознавателя; в противном случае он либо отсылает распознаватель к другим серверам, либо сам запрашивает другие сервера для того, чтобы обеспечить эту информацию.

После того как ресолвер получит это отображение, он анализирует отклик для того, чтобы посмотреть, является ли это реальным распознаванием или ошибкой. В конечном итоге результат доставляется процессу, который запросил его.

Отображение имен в адреса

Приложение может запросить у ресолвера разрешение имени. Ресолвер имеет локальный кеш, содержащий результаты обработки предыдущих запросов

1. Если кеш ресолвера не содержит ответа, ресолвер посылает запрос DNS-серверу

- Одним из параметров настройки узла TCP/IP является IP-адрес DNS-сервера, обеспечивающего разрешение имен
- Если настройки узла содержат IP-адреса нескольких DNS-серверов, ресолвер обращается к ним в том порядке, в котором они перечислены в настройках, до получения положительного или отрицательного ответа о разрешении имени
- Если параметры настройки узла не содержат адресов DNS-серверов, ресолвер возвращает приложению ошибку разрешения адреса

2. DNS-сервер при обработке запроса

❑ Если запрос относится к зоне, для которой он является авторизованным DNS-сервером – формирует ответ на основании содержимого файла соответствующей зоны

❑ В противном случае, если кеш обработанных запросов содержит ответ, возвращается результат из кеша

❑ В противном случае выполняется запрос (или последовательность запросов) к другим DNS-серверам

3. Запросы к другим DNS-серверам (предположим, мы выполняем запрос на разрешении имени www.ya.ru.)

❑ Сначала выполняется запрос одному из DNS-серверов, отвечающих за корневой домен

- Корневые сервера, скорее всего не являются авторизованными для зоны ya.ru., соответственно, запись с данным именем не хранится в их базах данных
- Если корневой сервер содержит ответ в своем кеше разрешенных запросов – отправляется ответ из кеша
- В противном случае, если адрес в запросе принадлежит одному из поддоменов корневого домена – возвращается IP-адрес DNS-сервера, отвечающего за данный домен. Если такого поддомена нет – возвращается ответ об отсутствии запрошенного имени

4. Далее выполняется запрос к DNS-серверу, отвечающему за домен ru.

❑ Если данный DNS-сервер является авторизованным для зоны ya.ru., он сформирует ответ в соответствии с содержимым своей базы данных.

Ответ от сервера, авторизованного для зоны, содержащей запрашиваемое имя, называется авторизованным. Ответ от любого другого DNS-сервера является неавторизованным

❑ Если сервер не является авторизованным, может быть возвращен ответ из кеша

❑ В противном случае возвращается IP-адрес DNS-сервера, отвечающего за поддомен ya.ru.

5. Далее выполняется запрос к DNS-серверу, отвечающему за домен ya.ru.

❑ Данный DNS-сервер обязательно является авторизованным для зоны ya.ru., поэтому он формирует авторизованный ответ о данном имени

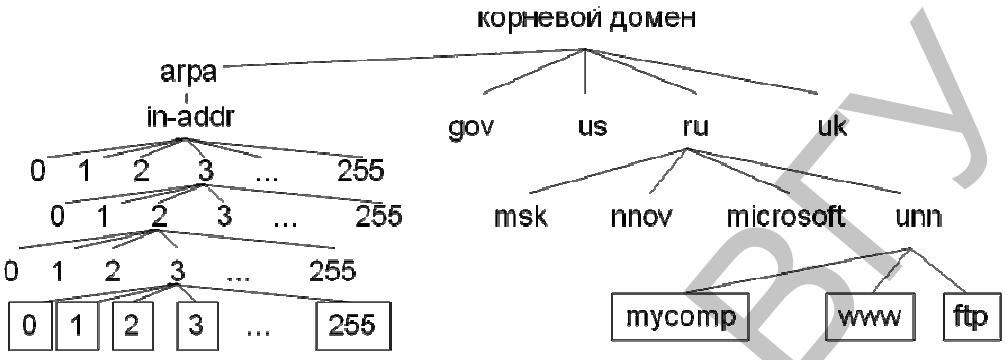
6. После получения ответа DNS-сервер, обслуживавший запрос, возвращает результат ресолверу

Отображение адресов в имена

Клиент может послать IP-адрес на сервер для того, чтобы отобразить доменное имя. Как уже было упомянуто прежде, это называется PTR-запрос. Для подобного запроса DNS использует инверсный домен. Однако IP-адрес запроса должен быть реверсирован, и должны быть прикреплены две метки,

in-addr или arpa, чтобы создать доступный домен с помощью инверсной доменной секции.

Domain Name System
Зоны обратного просмотра



- Reverse lookup zones – зоны обратного просмотра, предназначены для определения доменных имен по IP-адресам
  - Домен in-addr.arpa. содержит 256 поддоменов с именами 0, 1, 2, ..., 255
  - Домены d.in-addr.arpa. содержат по 256 поддоменов с именами 0, 1, 2, ..., 255
  - Домены c.d.in-addr.arpa. содержат по 256 поддоменов с именами 0, 1, 2, ..., 255
  - Домены b.c.d.in-addr.arpa. содержат по 256 записей с именами 0, 1, 2, ..., 255
  - Записи a.b.c.d.in-addr.arpa. содержат канонические имена узлов

Нижний Новгород 2007 | Компьютерные сети Domain Name System | 21 из 37

При разрешении имени могут использоваться два типа запросов

- Рекурсивный – клиент требует, чтобы ему вернули запрашиваемую запись ресурса, либо установили, что такой записи не существует
  - Такой запрос направляет ресолвер DNS-серверу
- Нерекурсивный (итеративный) – клиент просит вернуть ему либо запрашиваемую запись, либо IP-адрес DNS-сервера, от которого можно получить более конкретную информацию
  - В нашем примере – запросы от обслуживающего DNS-сервера к DNS-серверам, авторизованным для корневой зоны и зон ru. и unn.ru.

DDNS

Когда DNS был разработан, не предполагалось делать так много изменений адресов. В DNS, когда происходят перемены, такие как дополнение новых хостов, перемещение хоста, изменение IP-адреса, изменения может делать DNS-мастер-файл. Этот тип изменений включает множество ручных коррекций. Масштаб сегодняшнего Интернета не позволяет такого рода ручные операции.

DNS-мастер-файл должен быть скорректирован динамически. Поэтому была изобретена динамическая система доменных имен (DDNS – Dynamic Domain Name System). В DDNS, когда связь между именем и адресом определена, информация посылается обычно с помощью действий по протоколу динамической реконфигурации хостов (DHCP — Dynamic Host

Configuration Protocol) к первичному DNS-серверу. Первичный сервер модернизирует зону. Вторичные серверы уведомляются либо активно, либо пассивно. При активном уведомлении первичный сервер посылает сообщение вторичным серверам об изменениях в зоне, в то время как при пассивном уведомлении вторичные серверы периодически проверяются на любые изменения. В любом случае, после осуществления уведомления об изменении вторичные серверы запрашивают информацию об изменениях во всей зоне (зоновая передача), чтобы обеспечить безопасность. Для предотвращения неполномочных изменений DNS-записей DDNS может использовать полномочный механизм.

Инкапсуляция

DNS может использовать UDP или TCP. В обоих случаях сервером задействуется закрепленный порт – 53. UDP применяется, когда размер сообщения ответа меньше чем 512 байт, потому что большинство пакетов UDP имеют ограничение на размер пакета 512 байт. Если размер сообщения ответа больше чем 512 байт, должно использоваться соединение TCP. В этом случае может возникнуть один из двух сценариев:

• Если распознаватель заранее подтвердил, что размер ответного сообщения больше чем 512 байт, оно должно использовать TCP-соединение. Например, если передается вторичное имя сервера (активизированного как клиент) и нужна зоновая передача от первичного сервера, он должен применить TCP-соединение, потому что размер передаваемой информации обычно превышает 512 байт.

• Если распознаватель не знает размера ответного сообщения, он может задействовать UDP-порт. Однако если размер ответного сообщения больше чем 512 байт, сервер усекает сообщение и возвращает TC-бит. Распознаватель теперь открывает TCP-соединение, повторяет запрос и получает полный ответ от сервера.

- Данная программа работает из командной строки и предназначена для проверки работоспособности DNS-серверов и устранения неполадок в их работе. Средство Nslookup.exe устанавливается при установке протокола TCP/IP (например при установке из панели управления).

Nslookup.exe

Данная программа работает из командной строки и предназначена для проверки работоспособности DNS-серверов и устранения неполадок в их работе. Средство Nslookup.exe устанавливается при установке протокола TCP/IP (например при установке из панели управления).

При использовании Nslookup.exe необходимо учитывать следующее.

- На компьютере, на котором запускается Nslookup.exe, должен быть установлен протокол TCP/IP.
- В параметрах протокола TCP/IP должен быть указан хотя бы один сервер DNS (чтобы просмотреть данные параметры, выполните в командной строке команду IPCONFIG /ALL).

Если в свойствах протокола TCP/IP на вкладке DNS определен список DNS-суффиксов, используемых для разрешения неполных имен, то рассмотренный выше порядок регрессирования не применяется. В этом случае имя, заданное пользователем в запросе, будет добавляться к доменным суффиксам из указанного списка. Чтобы средство nslookup не выполняло лишних запросов, всегда используйте полностью определенные доменные имена (добавляйте к имени завершающую точку).

Чтобы запустить средство Nslookup.exe в интерактивном режим, выполните в командной строке команду nslookup (пример см. ниже).

```
C:\> nslookup
Default Server:  nameserver1.domain.com
Address:  10.0.0.1
>
```

Чтобы ознакомиться со списком доступных команд, выполните в командной строке средства nslookup команду help или «?». Если данные, введенные в командной строке, не являются правильной командой средства nslookup, то эти данные рассматриваются как имя узла, и делается попытка разрешить это имя с помощью сервера по умолчанию. Чтобы прервать выполнение команды в интерактивном режиме, нажмите клавиши CTRL+C. Чтобы завершить работу средства nslookup, находясь в интерактивном режиме, введите в командной строке команду exit.

Получение данных с удаленного сервера имен напрямую

```
C:\> nslookup

   Default Server:  nameserver1.domain.com
   Address:  10.0.0.1

   > server 10.0.0.2

   Default Server:  nameserver2.domain.com
   Address:  10.0.0.2
   >
```

Использование средства Nslookup.exe для передачи зоны

```
>ls domain.com
```

Лекция 13. Протоколы сетевого и транспортного уровней

Протокол определения адресов (ARP) и протокол определения сетевого адреса по местоположению (RARP)

Коммутация в локальной сети происходит на основе MAC-адресов, поэтому IP-модуль пользуется таблицей соответствия вида IP-адрес – MAC-адрес, которую заполняет протокол нахождения адреса (ARP – Address Resolution Protocol). Чтобы найти оптимальный маршрут, IP-модуль использует таблицу маршрутизации, которую составляет протокол маршрутизации. О возникших проблемах маршрутизаторы извещают друг друга при помощи протокола управляющих сообщений (ICMP). Рассылка одного и того же пакета множеству получателей производится протоколом управления группами в Интернете (IGMP – Internet Group Management Protocol).

Это означает, что доставка пакета хосту или маршрутизатору требует двух уровней адресации: логического и физического. Поэтому необходимы средства для того, чтобы отображать логический адрес в соответствующий ему физический адрес и наоборот. Они могут использовать либо статическое, либо динамическое отображение.

Статическое отображение (static mapping) означает создание таблицы, которая объединяет логический адрес с физическим адресом. Эта таблица сохраняется в каждом устройстве на сети. Каждое устройство, которое знает, например, IP-адрес другого устройства, но не его физический адрес, может отыскать его в таблице. Такой метод имеет некоторые ограничения, потому что физические адреса могут изменяться

Чтобы осуществлять эти изменения, статическая таблица отображения должна быть периодически модифицирована. Эта перезагрузка могла бы затронуть работу сети

При динамическом отображении (dynamic mapping) каждый раз, когда машина знает один из двух адресов (логический или физический), она может использовать протокол, чтобы найти другой из них.

Чтобы выполнять динамическое отображение, были разработаны два протокола: протокол определения адресов (ARP – Address Resolution Protocol) и протокол определения сетевого адреса по местоположению (RARP – Reserve Address Resolution Protocol). Первый отображает логический адрес в физический адрес; второй отображает физический адрес в логический адрес.

Протокол определения адреса (ARP)

Протокол ARP (Address Resolution Protocol) разработан для запроса передатчиком, когда необходимо, объявление приемником своего физического адреса.

ARP связывает адрес IP с его физическим адресом. На типичной физической сети, такой как LAN, каждое устройство на линии связи

идентифицировано с физическим адресом или адресом станции, который обычно закрепляется в центре сетевой информации.

Рассмотрим функции ARP для типичного Интернета. Рассмотрим шаги. Затем обсудим и четыре случая, в которых хост или маршрутизатор должны использовать ARP.

Шаги, составляющие процесс ARP

1. Передатчик знает IP-адрес получателя.
2. IP запрашивает, чтобы ARP создал сообщение запроса ARP, заполняющее в передатчике физический адрес, адрес IP-передатчика и целевой адрес IP. Целевое физическое поле адреса заполняется нулями.
3. Сообщение передают уровню звена передачи данных, где оно инкапсулируется в кадр, используя физический адрес передатчика как исходный адрес и физический широковещательный адрес как адрес пункта назначения.
4. Каждый хост или маршрутизатор получают кадр. Поскольку кадр содержит широковещательный адрес пункта назначения, все станции удаляют сообщение и передают его ARP. Все устройства, кроме того, изымают один целевой адрес протокола. Целевое устройство опознает свой адрес IP.
5. Ответ целевого устройства с сообщением ответа ARP, которое содержат свой физический адрес. Сообщение идет от одного узла только к узлу, запросившему адрес.
6. Передатчик получает сообщение ответа. Он теперь знает физический адрес целевого устройства.
7. Дейтаграмма IP, которая переносит данные для целевой машины, теперь инкапсулирована в кадре и направляется к пункту назначения.

Четыре различных случая использования ARP

1. Передатчик – это хост, и он хочет передать пакет другому хосту на той же самой сети. В этом случае логический адрес должен быть отображен в физический адрес в адресе IP пункта назначения в дейтаграммном заголовке.
2. Передатчик – это хост, и он хочет передать пакет другому хосту на другой сети. В этом случае хост просматривает свою таблицу маршрутизации и находит адрес IP следующего переприемного участка (маршрутизатора) для этого пункта назначения. Если он не имеет таблицы маршрутизации, он ищет адрес IP заданного по умолчанию маршрутизатора. Адрес IP маршрутизатора становится логическим адресом, который должен быть отображен в физический адрес.
3. Передатчик – маршрутизатор, который получил дейтаграмму, предназначенную для хоста на другой сети. Он проверяет свою таблицу маршрутизации и находит адрес IP следующего маршрутизатора. Адрес IP следующего маршрутизатора становится логическим адресом, который должен быть отображен в физический адрес

4. Передатчик — это маршрутизатор, который получил дейтаграмму, предназначенную для хоста в той же самой сети. Адрес IP пункта назначения дейтаграммы становится логическим адресом, который должен быть отображен в физический адрес.

ICMP

Протокол IP – это служба доставки с максимальными усилиями (best-effort), которая доставляет дейтаграмму от ее первоначального источника до ее конечного пункта назначения. Однако он имеет два дефекта: отсутствие контроля ошибок и отсутствие механизмов помощи в доставке.

Протокол IP не имеет механизма, сообщающего об ошибке или исправляющего ее. Протоколу IP также недостает ICMP-механизма запросов управления от хоста. Хост иногда должен определять, исправен ли маршрутизатор или другой хост. И иногда сетевой менеджер нуждается в информации от другого хоста или маршрутизатора. Сообщение об ошибке переносит данные о проблемах, возникающих при обмене сообщениями, с которыми маршрутизатор или хост (пункт назначения) могут столкнуться, когда они обрабатывают пакет IP.

IGMP — один из необходимых (но не достаточный, как мы увидим) протоколов, которые включаются в групповую передачу. IGMP взаимодействует с протоколом IP. показывает позиции протокола IGMP относительно других протоколов на сетевом уровне.

Для того чтобы рассылать сообщения по многим адресам в Интернете, мы нуждаемся в маршрутизаторах, которые способны направлять пакеты, рассылаемые по многим адресам. Таблицы маршрутизации этих маршрутизаторов должны быть модифицированы с использованием одного из протоколов маршрутизации групповой передачи.

IGMP не протокол маршрутизации групповой передачи; это — протокол, который управляет членством группы. В любой сети есть один или более маршрутизаторов групповой рассылки пакетов, которые распределяют пакеты, рассылаемые по многим адресам хостов или других маршрутизаторов. Протокол IGMP дает информацию маршрутизаторам групповой рассылки о состоянии членства хостов (маршрутизаторов), подключенных к сети.

Маршрутизация

Интернет – это комбинация сетей, соединяемых с помощью маршрутизаторов. Когда дейтаграмма идет от источника к пункту назначения, она, вероятнее всего, проходит много маршрутизаторов, пока достигает маршрутизатора, закрепленного за сетью пункта назначения. Маршрутизатор получает пакет от сети и передает его другой сети. Маршрутизатор обычно закрепляется за несколькими сетями. Когда он получает пакет, он должен решить две задачи:

1. к какой сети он должен его передать;
2. по какому пути.

Последнее решение основано на выборе оптимального пути. Какой доступный путь является оптимальным путем? Это обычно определяется метрикой. Метрика – это условная стоимость передачи по сети. Полное измерение конкретного маршрута равно сумме метрик сетей, которые включают в себя маршрут. Маршрутизатор выбирает маршрут с наименьшей метрикой. Метрика назначается для каждой сети в зависимости от типа протокола. Некоторые простые протоколы, подобно протоколу маршрутной информации (RIP – Routing Information Protocol), рассматривают все сети как одинаковые. Тогда стоимость прохождения через каждую сеть — одна и та же, и для определения метрики подсчитываются участки. Так, если пакет, чтобы достигнуть конечного пункта, проходит через 10 сетей, полная стоимость составляет 10 участков.

Другие протоколы, такие как "первоочередное открытие наикратчайших путей" (OSPF — Open Shortest Path First), позволяют администратору назначить стоимость для передачи через сеть, основанную на типе требуемого обслуживания. Маршрут через сеть может иметь различную стоимость (метрику). Например, если для типа сервиса желательна максимальная производительность, спутниковый канал имеет меньшую метрику, чем оптическая линия. С другой стороны, если типу сервера желательна минимальная задержка, оптическая линия имеет меньшую метрику, чем спутниковый канал. OSPF позволяет каждому маршрутизатору иметь таблицу последовательностей маршрутов, основанную на требуемом типе сервиса.

Другие протоколы определяют метрику различно. В протоколе пограничной маршрутизации (BGP — Border Gateway Protocol) критерий — это политика, которую может устанавливать администратор. Политика — это принцип, по которому определяется путь.

В любой метрике маршрутизатор должен иметь таблицы маршрутизации, чтобы консультироваться при дальнейшей передаче пакета. Таблица маршрутизации задает оптимальный путь для пакета. Таблица может быть либо статическая, либо динамическая. Статическая таблица — одна из тех, которые часто не меняются. Динамическая таблица — одна из тех, которая обновляется автоматически, когда имеются изменения где-либо в Интернете. Сегодня Интернет нуждается в динамических таблицах. Таблицы нужно обновлять по мере появления изменений в Интернете. Например, их нужно обновить, когда маршрут вышел из строя, или они должны быть обновлены всякий раз, когда создается лучший маршрут.

Протоколы маршрутизации созданы для отображения требований таблиц динамической маршрутизации. Протокол маршрутизации — комбинация правил и процедур, которые позволяют в Интернете маршрутизаторам информировать друг друга об изменениях. Протоколы маршрутизации также включают процедуры для комбинирования информации, полученной от других маршрутизаторов.

Внутренняя и внешняя маршрутизация

Сегодня Интернет — громадная сеть, так что один протокол маршрутизации не может обрабатывать задачу обновления таблиц всех маршрутизаторов. По этой причине Интернет разделяется на автономные системы. Автономная система (Autonomous System – AS) — группа сетей и маршрутизаторов под управлением одного администратора. Маршрутизация внутри автономной системы отнесена к внутренней маршрутизации. Маршрутизация между автономными системами отнесена к внешней маршрутизации. Каждая автономная система может выбрать протокол внутренней маршрутизации для того, чтобы обрабатывать маршрутизацию внутри автономной системы. Однако для обработки маршрутизации между автономными системами выбирается только один протокол маршрутизации.

Разработано несколько внутренних и внешних протоколов. В этой лекции мы коснемся только наиболее популярных из них — внутренних протоколов RIP и OSPF и одного внешнего протокола BGP. RIP и OSPF используются для обновления таблиц маршрутизации внутри автономной системы. Протокол BGP применяется в обновлении таблиц маршрутизации для маршрутизаторов, которые объединяют вместе автономные системы.

Протокол маршрутной информации (RIP)

Протокол маршрутной информации (RIP – Routing Information Protocol) — внутренний протокол маршрутизации, используется внутри автономной системы. **Таблицы маршрутизации**

Каждый маршрутизатор хранит таблицы маршрутизации, имеющие один вход для каждой сети назначения, которую маршрутизатор зарегистрировал. Вход содержит:

- адрес сети пункта назначения,
- кратчайший путь для того, чтобы достичь пункта назначения, отсчитываемый в участках,
- следующий участок (следующий маршрутизатор), к которому должен быть доставлен пакет по пути к своему конечному пункту назначения,
- счетчик участков – это число сетей, которые пакет пересечет для достижения своего конечного пункта назначения.

Таблица может содержать другую информацию, такую как маску подсети (или префикс) или время, когда этот вход был обновлен.

Таблица 8.1. Таблица вектора расстояния маршрутизации

Номер входа в таблицу участков	Пункт назначения	Счет участков	Следующий участок	Другая информация
0	163.5.0.0	7	172.6.23.4	
1	197.5.13.0	5	176.3.6.17	
2	189.45.0.0	4	200.5.1.6	
3	115.0.0.0	6	131.4.7.19	

Протоколы транспортного уровня

UDP (User Datagram Protocol)

Транспортный уровень обеспечивает соединение между прикладными программами и программами сетевого уровня. Прикладная программа посылает поток данных транспортному уровню. На передающей станции транспортный уровень разбивает поток на транспортабельные единицы, нумерует их и посылает их один за другим.

На приемном конце транспортный уровень собирает все различные блоки, принадлежащие к одной и той же прикладной программе, проверяет их и те, которые свободны от ошибок, передает дальше или доставляет к прикладной программе в виде потока. После того как будет передан весь поток, транспортный уровень завершает соединение. Набор программ для транспортного уровня задается двумя протоколами: UDP и TCP. В этом разделе мы изучим простейший из них — пользовательский протокол дейтаграмм (UDP — User Datagram Protocol).

UDP, как и TCP, обслуживает взаимодействие между прикладным уровнем и уровнями IP, а также обслуживает взаимодействие между прикладными программами и сетевыми операциями.

Протоколы транспортного уровня имеют несколько задач.

1. создать связь "процесс-процесс" (процесс – это работающая прикладная программа); чтобы это выполнить, UDP использует номер порта.
2. обеспечить механизм управления транспортным уровнем.

UDP занимается этой задачей в очень малой степени: он не обеспечивает механизма управления потоками, а следовательно, не обрабатывает подтверждения полученных пакетов. Однако UDP обеспечивает в некоторой степени контроль ошибок. Если UDP обнаруживает ошибку в принятом пакете, он, ничего не оповещая, удаляет его. Он может только получать блоки данных от процесса и доставлять их недостоверно приемнику. Блоки данных должны быть достаточно малы, чтобы подогнать их в пакет передачи. UDP называют **не ориентированным на соединение, недостоверным** транспортным протоколом.

Наряду с недостатками он имеет некоторые преимущества. UDP — очень простой протокол, использующий минимальные дополнительные затраты. Если процесс может посылать маленькие сообщения и не заботится о достоверности, он вполне может применять UDP. UDP, передающий маленькие сообщения, требует меньшего взаимодействия (обмена сигналами) между передатчиком и приемником, чем при использовании TCP.

Пользовательская дейтаграмма

UDP-пакеты называются пользовательскими дейтаграммами, имеют фиксированный размер 8 байт.

Поля заголовков.

• Адрес порта источника. Это номер порта, котоый используется процессом, выполняющимся в хосте сервера. Он равен 16 битам длины; это означает, что номер порта может быть в пределах от 0 до 65 535.

• Адрес порта пункта назначения. Это номер порта, используемый процессом в хосте пункта назначения. Он также имеет 16 бит длины.

• Длина. Это поле длиной 16 бит, которое определяет полную длину дейтаграммы пользователя, плюс заголовок данных. Эти 16 бит могут определять полную длину от 0 до 65 535 байт.

• Контрольная сумма. Это поле используется для обнаружения всей пользовательской дейтаграммы (заголовок плюс данные).

Протокол управления передачей (Transmission Control Protocol — TCP)

TCP называют надежный, ориентированный на соединение транспортный протокол сетевого уровня со свойствами, ориентированными на управление соединением и обеспечение надежности для обслуживания.

TCP не похож на UDP – это протокол, ориентированный на поток. В UDP процесс (прикладная программа) посылает большую порцию байт UDP для доставки, UDP добавляет свой собственный адрес к этой порции данных, которая теперь называется дейтаграммой, и доставляет ее IP для передачи. Процесс может доставлять несколько порций данных к UDP, но UDP обрабатывает каждую порцию независимо, "не глядя" на связь между ними.

TCP позволяет создать процесс, передающий информацию, доставлять данные как поток байт, создать процесс приема и получать данные как поток байтов. TCP создает среду, где кажется, что два процесса соединены воображаемой "трубой", которая переносит их данные по сети Интернет.

В связи с тем что процессы передачи и приема могут производить и потреблять данные на разных скоростях, TCP нуждается в буферной памяти для накопления. Имеются два буфера, передачи и приема, для каждого направления. (Заметим, что эти буферы также применяются в TCP-механизме управления потоком и контролем над ошибками.)

На передающей стороне буфер имеет три типа участков.

1. участок, который может заполняться с помощью процесса передачи (производитель).
2. байты, которые переданы, но на них еще не получено подтверждение. TCP сохраняет эти байты в буфере, пока не примет подтверждение.
3. участок содержит байты для передачи TCP.

Сегменты

На транспортном уровне TCP группирует несколько байтов в пакет, называемый сегментом. TCP добавляет заголовок к каждому сегменту (с целью контроля) и доставляет сегмент на IP-уровень для передачи. Сегмент инкапсулируется и передается в IP-дейтаграмме. Сегменты могут быть получены в беспорядке, потеряны или искажены и разрушены. Все эти сегменты обрабатываются TCP и передаются процессу на стороне приема, не подозревающему об этих действиях.

Заметим, что сегменты могут отличаться размером. Реально сегменты переносят сотни, если не тысячи байт.

TCP предлагает полное дуплексное обслуживание, где данные могут двигаться в обоих направлениях одновременно. Каждый TCP поэтому имеет буфер приема и передачи и посылает сегменты в оба направления.

TCP, в отличие от UDP, — протокол, ориентированный на соединение. Когда процесс на стороне A посылает и принимает данные от другого процесса на стороне B, необходимо провести следующие действия:

1. TCP на стороне A информирует TCP на стороне B и получает подтверждение от стороны B.
2. TCP стороны A и TCP стороны B обмениваются данными в обоих направлениях.
3. После того как у обоих процессов не остается больше данных для передачи и буферы пустые, оба TCP уничтожают буферы.

Заметим, что это не физическое, а виртуальное соединение. TCP-сегмент инкапсулируется в IP-дейтаграмму и может посылать данные в любом порядке или потерять их, либо исказить, либо передать повторно. Каждая дейтаграмма может использовать различный путь для достижения пункта назначения. Физического соединения не происходит. TCP создает среду, ориентированную на поток в каждом направлении, в которой он принимает ответственность за доставку байтов в заданном порядке на другую сторону.

TCP — достоверный транспортный протокол. Он использует механизм подтверждения для проверки сохранности и нормальности пребывающих данных.

Контроль ошибок

TCP — достоверный протокол транспортного уровня. Это означает, что прикладная программа доставляет поток данных к TCP, к прикладной программе на другом конце в порядке, без ошибок и без потери любой части или дублирования.

TCP обеспечивает достоверность, используя контроль ошибок. Контроль ошибок включает в себя механизмы обнаружения:

- искаженных сегментов;
- потери сегментов, нарушения порядка следования сегментов;
- дублирования сегментов.

Контроль ошибок также включает механизм для коррекции ошибок, после того как они обнаружены.

Лекция 14. Протоколы прикладного уровня: TELNET, FTP (TFTP). TELNET

Главная задача Интернета и его набора протоколов TCP/IP — это обеспечить сервис для пользователя. Например, пользователь хочет иметь возможность выполнять различные прикладные программы на удаленном сайте и создать результат, который может быть передан к его местному сайту. Один из путей удовлетворения такой потребности — создать различные прикладные программы клиент-сервер для каждой услуги. Уже доступны программы передачи файлов (FTP и TFTP), электронной почты (SMTP) и так далее. Однако все конкретные программы клиентсервер для каждого применения описать невозможно.

Лучшее решение — общецелевая программа клиент-сервер, которая позволяет пользователю иметь доступ к любой прикладной программе на удаленном компьютере. После входа в систему пользователь может использовать услуги, доступные на удаленном компьютере, и принимать результаты на местном компьютере.

TELNET — это сокращение от Terminals NETwork. Это стандартный протокол TCP/IP для услуг виртуального терминала. TELNET дает возможность устанавливать соединение с удаленным компьютером таким образом, что создается впечатление, как будто местный терминал – это терминал удаленной системы.

Внешняя среда с разделением времени

TELNET был разработан в эпоху, когда большие операционные системы, такие как UNIX, работали с внешней средой по принципу разделения времени. Согласно этому принципу, большой компьютер поддерживал множество пользователей, предоставляя им часть общего времени. Взаимодействие между пользователем и компьютером осуществляется с помощью терминала, который обычно состоит из комбинации клавиатуры, монитора и мышки. В среде с разделением времени вся обработка информации проводится в центральном компьютере. Когда пользователь печатает символ на клавиатуре, символ обычно посылается компьютеру и отражается на мониторе. Разделение по времени создается средой, в которой для каждого пользователя создается иллюзия специализированного компьютера.

Логин

В среде с разделением времени пользователь — это часть системы с некоторыми правами и, вероятно, с паролем. Каждый полномочный пользователь имеет идентификатор и пароль. Пользовательская идентификация определяет пользователя как часть системы. Для доступа к системе пользователь начинает сеанс с пользовательского идентификатора (id) или с регистрационного имени (login name). Система помогает проверке пароля, чтобы предотвратить доступ к ресурсу неполномочного пользователя.

Местный логин

Когда пользователь входит в местную систему с разделением времени, это называется местный логин. Как только пользователь напечатает некое слово на терминале или рабочей станции, выполняющей эмуляцию терминала, сразу начинает работать терминальная программа (драйвер), которая распознает значение введенных символов. Терминальный драйвер передает символы операционной системе, в рамках этой системы комбинация символов интерпретируется и вызывает желаемую прикладную программу или утилиту

Удаленный логин

Когда пользователь хочет иметь доступ к прикладной программе или утилите, размещенным на удаленном компьютере, он выполняет дистанционный вход в систему (логин). Здесь TELNET берет на себя функции клиента и сервера. Пользователь посылает сигнал нажатия кнопки терминальному драйверу, где местная операционная система принимает символы и интерпретирует их. Эти символы посылает TELNET-клиент, который преобразует символы к универсальному набору, называемому символы виртуального сетевого терминала (Network Virtual Terminal Characters), и доставляет их к местному стеку протоколов TCP/IP (рис. 12.2).

Команды или текст в форме сетевого виртуального терминала (NTV) перемещаются через Интернет и прибывают на стек протоколов TCP/IP в удаленной машине. Здесь символы доставляются операционной системе и проходят к TELNET-серверу, который преобразует их в символы, понятные удаленному компьютеру. Однако символы не могут пройти прямо на операционную систему, потому что удаленная операционная система не разработана для получения трактовки этих символов от TELNET. Она спроектирована так, чтобы принимать символы от драйвера терминала. Решение, добавляющее необходимое программное обеспечение, называется псевдотерминальным драйвером, который преобразовывает поступившие символы как символы, поступающие от местного терминала. Операционная система затем предает символы к соответствующей прикладной программе.

Режим работы

Большинство реализаций TELNET работает в одном из трех режимов: заданный по умолчанию режим, символьный режим и режим линии.

Режим, заданный по умолчанию

Режим, заданный по умолчанию, используется, когда с помощью опции переговоров не запрошены никакие другие режимы. В этом режиме возвращение символов делается клиентом. Пользователь печатает символ, а клиент отображает символ на экране (или принтере), но не посылает его, пока не закончится вся строка. После посылки полной строки на сервер клиент ждет команду GA (go ahead) от сервера, перед принятием новой строки — от пользователя. Эта работа — полудуплексная. Полудуплексная работа не эффективна, когда связь в самом TCP является дуплексной, так что этот режим устаревает.

Символьный режим

В символьном режиме каждый напечатанный клиентом символ посылается серверу. Сервер обычно обрабатывает символ, чтобы отобразить на экране клиента. В этом режиме отражение символа может быть отсрочено, если передача происходит длительное время (такое, как при спутниковой связи). Оно также создает перегрузку (трафика) для сети, потому что для каждого символа данных нужно послать три сегмента TCP:

1. пользователь вводит символ, который посылает серверу;
2. сервер признает полученный символ и повторяет символ назад (в одном сегменте);
3. клиент подтверждает получение отображенного на экране символа.

Режим строки

Новый режим был предложен, чтобы компенсировать недостатки режима по умолчанию и символьного режима. В этом режиме, названном режимом строки, редактирование строки (повторение, стирание символа, стирание строки и так далее) делается клиентом. Затем клиент посылает целую строку серверу.

Хотя режим строки напоминает режим, заданный по умолчанию, это только внешнее сходство. Режим, заданный по умолчанию, работает в полудуплексном режиме; режим строки является дуплексным, с клиентом, посылающим одну строку за другой, без потребности во вмешательстве символа GA (иди дальше — go ahead) от сервера.

Пользовательский интерфейс

Обычно пользователь не использует команды TELNET так, как это определено выше. Как правило, операционная система (например, UNIX) определяет интерфейс с командами, дружественными пользователю. Пример из такого набора команд может быть найден в таблице 12.6. Заметим, что интерфейс отвечает за перевод команд, дружественных пользователю, к командам, определенным ранее в протоколе.

Таблица 12.6. Пример команд интерфейса

Команда	**Смысловое значение**
Open	Связь к удаленному компьютеру
Close	Завершение связи
Display	Показ рабочих параметров
Set	Установка рабочих параметров
Status	Отображение информации о состоянии
Send	Посылка специальных символов
quit	Выход из TELNET

Протокол FTP

Протокол передачи файлов (File Transfer Protocol – FTP) – это стандартный механизм для копирования файла от одного хоста другим. Передача файлов от одного компьютера к другому – это одна из большого числа общих задач, выполнение которой ожидается от организованной сети и взаимодействия между сетями.

Хотя передача файлов от одной системы к другой кажется простой и прямолинейной задачей, вначале должны быть решены некоторые проблемы. Например, две системы могут использовать различные соглашения об именах файлов. Две системы могут иметь различные пути для представления текстов и данных. Две системы могут иметь различные структуры директорий. Все эти проблемы решает FTP очень простым и элегантным методом.

FTP отличается от других приложений типа клиент-сервер тем, что он устанавливает два соединения между хостами. Одно соединение применяется для передачи данных, другое — для управления информацией (команды и отклики). Разделение команд и передачи управляющих данных делает FTP более эффективным. Управление соединением использует очень простые правила для связи. Нам нужна для передачи только линия команд или линия откликов. С другой стороны, соединение для данных нуждается в более сложных правилах из-за разнообразия типов данных.

FTP использует два заданных порта: порт 21 для управления и порт 20 для передачи данных.

Соединение передачи сигналов управления остается открытым в течение всей интерактивной сессии FTP. Соединение передачи данных каждый раз открывается командой, чтобы вызвать передаваемый файл, и затем закрывается, когда файл передан. Другими словами, когда пользователь начинает FTP-сессию, соединение для передачи сигналов управления открывается. Пока оно открывается, соединение для передачи данных может быть открыто и закрыто много раз, если передается несколько файлов.

Соединения

Два FTP-соединения – для передачи команд управления и передачи данных — используют различные стратегии и различные номера портов.

Соединение для передачи команд управления

Соединение для передачи команд управления создается тем же самым методом, что и другие соединения, рассмотренные далее. Имеется два шага:

1. сервер пассивно открывается, подключается к заданному порту и ждет клиента;
2. клиент использует временный порт, и сессия активно открывается.

Соединение для передачи команд управления остается открытым в течение всего процесса. Тип услуги, используемый в соответствии с IP-протоколом, – это минимизация задержки, потому что это диалоговая связь между пользователем (человеком) и сервером. Пользователи различного типа посылают команды и ожидают получение откликов без существенной

задержки. Рис. 13.2 показывает начальное соединение между сервером и клиентом. Конечно, после начального соединения процесс сервера порождает "дочерние" процессы и назначает свободное обслуживание клиента "дочерним" процессом, использующим кратковременный порт.

Соединение для передачи данных

Соединение для передачи данных использует заданный порт 20. Однако создание соединения для передачи данных отличается от предыдущего. FTP создает соединение для передачи данных следующим образом:

1. Клиент (не сервер) вызывает пассивное открытие кратковременного порта. Это может быть сделано клиентом, потому что клиент вызывает команды для передачи файлов.
2. Клиент посылает номер этого порта серверу, используя команду PORT (ниже эта команда будет рассмотрена).
3. Сервер получает номер порта, вызывает активное открытие заданного порта 20 и получает номер временного порта.

Шаги для создания начального соединения для передачи данных показаны на рис. 13.3. Позднее мы увидим, что эти шаги меняются, если используется команда PASV.

Установление соединения

Процессы FTP клиента и сервера, которые выполняются на различных компьютерах, могут устанавливать соединение друг с другом. Эти два компьютера могут использовать различные операционные системы, различные наборы символов, различные структуры и различные форматы файлов. FTP должен сделать совместимой всю эту неоднородность.

FTP обладает двумя различными подходами для управления соединением: одним для соединения для передачи команд управления и одним для передачи данных. Рассмотрим отдельно каждый метод.

Тип файла

FTP может передавать через соединение для передачи данных следующие типы файлов:

- ASCII-файл. Это формат, используемый по умолчанию для трансляции текстовых файлов. Каждый символ закодирован с использованием NVT ASCII-символов. Передатчик преобразует файл из собственного представления в NVT ASCII, и приемник преобразует символы NVT ASCII в собственное представление.
- EBCDIC-файл. Если оба конца соединения используют кодирование EBCDIC, файл может быть передан с использованием EBCDIC-кодирования.
- Image-файл. Этот файл по умолчанию — формат для передачи двоичных файлов. Файл посылается как непрерывный поток бит без всякой интерпретации и кодирования. Он в большинстве случаев используется для передачи двоичных файлов, таких как компилированная программа.

Если файл закодирован в ASCII или EBCDIC, другие атрибуты должны дополнятся, чтобы определить возможность печати файла:

1. Запрещенный для печати. Это формат по умолчанию для передачи текстовых файлов. Файл не содержит "вертикальных" спецификаций для печати. Это означает, что файл не может быть напечатан без предварительной обработки, потому что он не содержит символов, интерпретируемых для вертикального передвижения печатающей головки. Этот формат используется для файлов, которые будут накоплены и обработаны позднее.

2. TELNET. В этом формате файл содержит NVT ASCII вертикальные символы, такие, как CR (перевод каретки), LN (перевод строки), NL (новая строка) и VT (вертикальное табулирование). Эти файлы могут быть напечатаны после передачи

Структура данных

FTP может передавать файл по соединению для передачи данных, используя одну из следующих интерпретаций структуры данных:

- Файловая структура (по умолчанию). Этот файл не имеет структуры. Это непрерывный поток данных.
- Структура записи. Этот файл, разделенный внутри записи Он может быть использован только с текстовым файлом.
- Страничная структура. Это файл, разделенный на страницы, каждая страница имеет номер и заголовок страницы. Страницы могут быть накоплены или достигнуты с помощью произвольного или последовательного доступа.

Режимы передачи

FTP может передавать файл по соединению для передачи данных, используя один из трех следующих режимов передачи:

- Поточный режим. Это режим по умолчанию. Данные доставляются от FTP к TCP как непрерывный поток данных. TCP отвечает за разбиение данных на сегменты соответствующего размера. Если данные — просто поток байтов (файловая структура), то не нужно никакого признака окончания файла. Окончание файла в этом случае — это разъединение соединения данных отправителем. Если данные разделены на записи (структура по записи), каждая запись будет иметь однобайтный символ окончания записи (EOR — end of record).
- Блочный режим. Данные могут быть доставлены от FTP и TCP в блоках. В этом случае блоку предшествует трехбайтный заголовок. Первый байт называется дескриптор блока, следующие два байта определяют размер блока в байтах.
- Сжатый режим. Если файл большой, данные могут быть сжаты. Метод сжатия использует нормальное кодирование длины. В этом методе последовательное повторное появление блока данных заменяется одним вхождением и числом повторений. В тексте файла это обычно пробел (пустоты). В двоичном файле нулевые символы обычно сжимаются.

Команды

Команды, которые посылаются от FTP-процесса управления клиента в форме заглавных букв ASCII, могут сопровождаться или не сопровождаться аргументом. Мы можем грубо разделить команды на шесть групп: команды доступа, команды управления файлами, команды форматирования данных, команды определения порта, команды передачи файла и прочие команды.

- Команды доступа. Эти команды позволяют пользователю обращаться к удаленной системе. Табл. 13.1 перечисляет общие команды в этой группе.

Таблица 13.1. Команды доступа

Команды	**Аргументы**	**Описание**
USER	ID пользователя	Пользовательская информация
PASS	Пароль пользователя	Пароль
ACCT	Загруженная учетная запись	Учетная информация
REIN		Перезапуск
QUIT		Выход из системы
ABOR		Прерывание предыдущей команды

- Команды управления файлом. Эти команды дают пользователю возможность обращаться к удаленному компьютеру, передвигаться по структуре директории, создавать новые директории, удалять файлы и так далее. В Табл. 13.2 даны общие команды этой группы.

Таблица 13.2. Команды управления файлом

Команды	**Аргументы**	**Описание**
CWD	Имя директории	Изменение другой директории
CDUP		Изменение вышестоящей директории
DELE	Имя файла	Удаление файла
LIST	Имя директории	Список поддиректорий и файлов
NLIST	Имя директории	Список имен поддиректорий или файлов, не имеющих атрибутов
MKD	Имя директории	Создать новую директорию
PWD		Имя текущей директории на дисплее
RMD	Имя директории	Удалить директорию
RNER	Имя файла (старое имя)	Идентификатор файла, который будет переименован
RNTO	Имя файла (новое имя	Переименование файла

	файла)	
SMNT	Системное имя файла	Вершина системы

- Команды форматирования данных. Эти команды дают пользователю возможность определить данные о структуре, типе файла и режиме передачи. Определенный формат затем используется командами передачи файлов. Табл. 13.3 показывает общие команды этой группы.

Таблица 13.3. Команды форматирования данных

Команды	**Аргументы**	**Описание**
TYPE	A (ASCII), E (EBCDIC), I (IMAGE), N (Nonprint), T(Telnet)	Определяет тип файла, если необходим формат для печати
STRU	F (File), R (Record), P (Page)	Определяет организацию данных
MODE	S (Stream), B (Block), C (Compressed)	Определяет режим передачи

- Команды определения порта. Эти команды определяют номер порта для соединения передачи данных на стороне клиента. Имеется два метода, чтобы сделать это. Первый метод применяет команду PORT, чтобы клиент мог выбрать кратковременный номер порта и послать серверу для использования при пассивном открытии. Сервер задействует номер порта и порождает активное открытие этого порта. Во втором методе используется команда PASV, клиент сначала запрашивает сервер о выборе номера порта. Сервер производит пассивное открытие этого порта и посылает в отклике номер порта (см. отклик, пронумерованный 227 в таблице 13.7). Клиент делает активное открытие, используя номер порта. Таблица 13.4 показывает команды определения порта.

Таблица 13.4. Команды определения порта

Команды	**Аргументы**	**Описание**
PORT	6-цифровой идентификатор	Клиент выбирает порт
PASV		Сервер выбирает порт

- Команды передачи файла. Эти команды позволяют передачу файлов. В таблице 13.5 перечислены общие команды этой группы.

Таблица 13.5. Команды передачи файла

Команды	**Аргументы**	**Описание**
RETR	Имя файла (ов)	Извлечение файла: файл(ы) передан(ы) от сервера к клиенту

STOR	Имя файла (ов)	Накопление файла: файл(ы) передан(ы) от клиента к серверу
APPE	Имя файла (ов)	Совпадает с STOR за исключением того, что если файл существует, то данные могут быть прикреплены к нему
STOU	Имя файла (ов)	То же самое, что STORE, за исключением того, что имя файла будет уникальным в этой директории; однако существующий файл не должен быть переписан
AALLO	Имя файла (ов)	Распределение места для накопления файлов в сервере
REST	Имя файла (ов)	Установка отметки в определенной точке данных
STAT	Имя файла (ов)	Возврат состояния файла

- Различные команды. Эти команды доставляют информацию к пользователю FTP на стороне клиента. Табл. 13.6 показывает общие команды этой группы.

Таблица 13.6. Прочие команды

Команды	**Аргументы**	**Описание**
HELP		Запрос информации
NOOP		Проверка, является ли сервер действующим
SITE	Команды	Определение сайта заданных команд
SYST		Запрос об операционной системе, используемой сервером

Анонимный FTP

Чтобы использовать FTP, пользователю необходимо передать учетную запись (имя пользователя) и пароль на удаленный сервер. Некоторые сайты имеют набор файлов, доступных для общего пользования. Чтобы иметь доступ к этим файлам, пользователю не нужна учетная запись или пароль. Вместо этого пользователь может использовать анонимность (anonymous) как пользовательское bvz-имя и гостевой (guest) пароль.

Доступ к системе пользователя очень ограничен. Некоторые сайты разрешают анонимным пользователям только поднабор команд. Например, большинство сайтов дают пользователю возможность копировать некоторые файлы, но запрещают осуществлять навигацию по директориям.

Лекция 15. Протоколы электронной почты: SMTP, POP, IMAP

Простой протокол электронной почты (SMTP — Simple Mail Transfer Protocol)

Одна из наиболее популярных сетевых услуг – это электронная почта (e-mail). TCP/IP протокол, который поддерживает сообщения электронной почты в Интернете — это простой протокол электронной почты (SMTP — Simple Mail Transfer Protocol). Он описывает систему команд и соглашений для посылки сообщений к другим компьютерным пользователям, основанную на адресах электронной почты. SMTP обеспечивает обмен почтовыми сообщениями между пользователями одной и той же или различных компьютерных сетей. Система поддерживает:

- посылку одиночных сообщений одному или более получателям;
- посылку сообщений, включающих в себя текст, голосовые сообщения, видео или графические материалы;
- посылку сообщений для пользователей сетей, не входящих в Интернет.

Агент пользователя (User Agent – UA)

Начнем с того, что надо отделить работу сервера клиента от почтовой сети, так чтобы он мог иметь режим работы, независимый от почтовой сети, и наоборот. Чтобы его повседневная работа не влияла на почтовую сеть, введем агента пользователя (User Agent – UA). Аналогичный компонент требуется для почтового сервера – агент почтовой передачи (Mail Transfer Agent). Эти компоненты будут представлять соответственно клиента и почтовый сервер в сети.

Агент пользователя подготавливает сообщение, адрес и вкладывает сообщение в конверт (см. например, систему Microsoft Outlook).

Теперь рассмотрим основные компоненты.

Агент пользователя без деталей реализации определяется в SMTP. АП – это обычная программа для передачи и получения почты. Одна из наиболее популярных программ — агент пользователя Outlook Express. Большинство агентов пользователя используют специальный интерфейс (типы окон) для взаимодействия клиент-система.

Услуги, обеспечиваемые агентом пользователя

Почтовый агент пользователя обеспечивает оформление письма, чтение письма, создание ответного сообщения, пересылку полученного письма одному или нескольким адресатам, работу с любыми типами писем (входящие, исходящие, отправленные, черновики). Рассмотрим кратко работу этих прикладных программ.

Оформление письма

Большинство агентов пользователя предоставляют специальную форму для заполнения письма, которая выводится на экран. Она содержит поля адресов (исходящий адрес, тема письма, адреса отправки копий). В скрытом виде добавляется адрес отправителя. Как правило, агент пользователя

предоставляет возможность просмотра имеющихся адресов, поиска в них нужного и установки его в письмо.

Кроме этого, современные пользовательские агенты позволяют проводить грамматическую проверку текста письма на нескольких языках.

Чтение письма

Агент пользователя позволяет читать входящие или накопленные письма. Для этого он предоставляет список входящих писем, с отметкой — от кого (имя), адрес отправителя и дата поступления. Дополнительно может быть указан объем письма. Из дополнительных услуг обычно предоставляется отметка нового или непрочитанного письма.

Ответ на письмо

Агент предоставляет возможность пользователю сформировать ответ на письмо. По запросу пользователя он автоматически выводит на экран форму ответа, в которой занесен адрес приславшего письмо и текст этого письма. Пользователь может сохранить этот текст, добавить в его начало, конец или середину в любом месте свой текст или стереть поступивший текст и создать новый.

Пересылка входящего письма

Пересылка предоставляет возможность переслать текст третьему лицу. При этом пользователю выводится форма, где указано, что это пересылка (знак "Fw:"); пользователь должен вставить адрес пересылки и может дополнить или исправить пересылаемый текст.

Работа с почтовым ящиком

Агент позволяет хранить все письма рассортированными по группам, пока они не будут удалены пользователем. При этом обеспечивается упорядочивание по различным критериям – по алфавиту, по фамилии приславшего, по дате и т. п.

Посылка почты

На рис. 14.3. показано прохождение почтового сообщения от абонента с именем Alex@mail2.spez.com к почтовому серверу абонента с именем Bob@mail3.param.ru по протоколу SMTP. Для посылки почты пользователь с помощью программы-агента порождает отправление, очень похожее на почтовое отправление. Оно содержит конверт и сообщение.

Конверт

Конверт обычно содержит адрес отправителя и получателя и другую информацию.

Сообщение

Сообщение содержит заголовок и само содержание (тело сообщения). Заголовок определяет отправителя, получателя и субъект (название) передачи и некоторую другую информацию. Тело сообщения содержит информацию, которую надо прочесть получателю.

Получение письма

Программа-агент пользователя при подключении к почтовому серверу получает все письма для данного пользователя и информирует его о наличии

писем. Обычно информация состоит из указания для каждого письма отправителя, субъекта письма и времени, когда письмо было получено или послано. Если пользователь готов читать письма, он выбирает одно из них и дает команду на раскрытие его содержания на экране.

Адресация

Адресация почтовых сообщений аналогична той, что принята в системе доменных имен (Domain Name System, DNS). Почтовый адрес имеет вид Alex@spez.com, где Alex – символическое имя, spez.com – почтовый домен.

На рисунке локальная (местная) часть адреса определяет имя специального файла, названного по имени почтового ящика, где накапливается вся почта пользователя с целью обработки ее программой-агентом отправителя.

Вторая часть адреса – доменное имя. Дается для выбора одной или более главных машин (хост) для посылки и получения электронной почты; иногда эти машины называют почтовыми станциями. Доменное имя присваивается каждой почтовой станции согласно базе доменных имен или по логическому имени (например, по имени организации).

Электронный адрес становится более сложным, когда используется почтовый шлюз. В этом случае электронный адрес должен определять оба адреса – адрес шлюза и адрес реального получателя. Доменное имя должно определять имя почтового шлюза, в базе данных доменных имен и локальной части должен быть определен локальный физический адрес, компьютер присоединяет номера сетевого и пользовательского почтового ящика. Большинство почтовых систем не использует почтовую адресацию, определяемую SMTP, — это может породить проблемы и ошибки.

Задержка в доставке

SMTP отличается от других прикладных программ тем, что вносит задержки. Это означает, что этот протокол не обеспечивает немедленную доставку, а задерживает ее на стороне отправителя, стороне получателя или на серверах межсетевого общения.

Задержка на стороне отправителя

Отправляемое сообщение может быть задержано стороной передачи. SMTP предусматривает, что отправитель должен располагать системой размещения в очереди, в которой сообщение накапливается, перед тем как быть переданным. После того как агент пользователя создаст сообщение, оно доставляется для постановки в очередь, которая является накапливающей структурой. Система почтовой передачи периодически проверяет почтовый накопитель и рассматривает возможность передачи. Это зависит от того, может ли получить заданный адрес сервера доступ через DNS. Если сообщение не доставлено в определенный период (обычно от пяти до трех дней), почта возвращается отправителю.

Задержка на приемной стороне

После того как сообщение получено, SMTP не передает его непосредственно. Почта может быть накоплена в почтовом ящике приемника.

Задержка среды передачи

Как уже упоминалось, SMTP позволяет обслуживать непосредственно почтовых агентов как клиентов и как серверы. Они также могут получать почту, сохранять почту в их собственных почтовых ящиках, записывать в буфер и отсылать сообщения в соответствующий момент.

Псевдоним (групповое имя)

SMTP позволяет использовать только одно имя, псевдоним, позволяющий последовательную различную адресацию — это называют расширением "один ко многим". Также отдельный пользователь может обладать различными адресами — это называется расширение "много к одному". Такие операции должны опираться на включение средств расширения на обоих концах (рис. 14.4.), в частности средств буферизации – накопления (spooling).

Расширение "один ко многим"

В ситуации, где одно и то же сообщение может посылаться различным получателям, пользователь может создать псевдоним, который отображает список получателей. В момент посылки сообщения система проверяет имя получателя в базе данных псевдонимов; если оно имеется, сообщение разделяется на отдельные сообщения, и в каждый из адресов должно быть отправлено и обработано почтовой системой отдельное сообщение. Если это имя отсутствует, то сообщение передается как одиночное по адресу получателя.

Расширение "много к одному"

Пользователь может иметь много сетевых адресов, но пользовательский агент при этом может иметь одно имя. Обычно в таком случае различается локальная часть адреса. Когда система получает почту, она проверяет базу данных типа "много к одному". Если имя связано с локальной частью полученного адреса, почта посылается в этот почтовый ящик; в противном случае почта удаляется.

Агент почтовой передачи

Реальная передача происходит через почтового агента передачи. Чтобы передать сообщение, система должна иметь клиентского почтового агента, а приемная сторона должна иметь агента почтового сервера. Протокол SMTP не предъявляет специальных требований к почтовым агентам. Он определяет команды и отклики, которые должны посылаться назад и далее. Каждая сеть свободна в выборе пакета программ для реализации. Далее в этой лекции еще будет обсуждаться механизм передачи SMTP. Однако вначале представим полную картину двусторонней передачи электронной почты, как это определено SMTP. Рис. 14.5. иллюстрирует процесс посылки и получения электронной почты, как это было рассмотрено ранее. Для получения и передачи сообщения в соответствии с протоколом SMTP пользовательский

интерфейс не является необходимым, но создает дружественный диалог с компьютером, как это, например, сделано в Outlook Express.

Фазы передачи почты

Процесс передачи почтовых сообщений осуществляется в три фазы: установление соединения, передача почты и подключение оконечного устройства.

Установление соединения

После того как клиент установит соединение TCP к заранее известному порту 25, сервер SMTP начинает фазу соединения.

1. Сервер посылает код 220 (Готов к обслуживанию), чтобы сказать клиенту, что он готов принять почту. Если сервер не готов, то он посылает код 421 (Обслуживание не готово).
2. Клиент посылает сообщение HELLO, чтобы идентифицировать себя, используя доменное имя адреса. Этот шаг необходим, чтобы информировать сервер доменного имени клиента. Напомним, что во время установления TCP отправитель и получатель знает друг друга только по IP-адресам.
3. Сервер отвечает кодом 250 – "Требуемая команда завершена" или другим кодом в зависимости от ситуации.

Передача сообщения

После того как соединение будет установлено между SMTP-клиентами и сервером, можно обменяться одиночным сообщением между отправителем и одним или более получателями. Эта фаза включает восемь шагов.

Клиент посылает сообщение MAIL FROM, чтобы представить отправителю почтовый адрес отправителя (имя почтового ящика и доменное имя). Этот шаг необходим, чтобы дать серверу адрес для возврата ошибок или для доклада о продвижении сообщений.

1. Сервер отвечает кодом 250 или другим соответствующим кодом.
2. Клиент посылает сообщение RCPT TO (получатель), который включает почтовый адрес получателя.
3. Сервер отвечает кодом 250 или другим соответствующим сообщением.
4. Клиент посылает сообщение DATA, чтобы инициализировать передачу сообщений.
5. Сервер отвечает кодом 354 (Начало ввода почты) или другим подходящим сообщением.
6. Клиент посылает содержание сообщения в виде последовательности строк. Каждая строка завершается двумя символами конец строки (возврат каретки и продвижение на другую линию). Конец сообщения содержит только метку окончания строки.
7. Сервер отвечает кодом 250 или соответствующим кодом.

Окончание соединения

После того как сообщение будет успешно передано, клиент заканчивает соединение. Клиент посылает команду QUIT.

1. Сервер отвечает кодом 221 или соответствующим другим кодом.

2. После фазы окончания соединения TCP-соединение должно быть завершено.

Многоцелевое расширение интернет-почты

Многоцелевое расширение интернет-почты (Multipurpose Internet Mail Extensions – MIME) улучшает возможности протокола SMTP [33, 34, 35, 38, 39, 40]. Этот протокол может посылать в терминалы только 7-битовые форматы в коде ASCII. Другими словами, он имеет ограничения. Например, не могут быть использованы языки, которые не поддерживают 7-битовые символы (французский, немецкий, иврит, русский, китайский, японский). Также нельзя использовать его для посылки двоичных файлов или посылки видео- или аудиоинформации.

Многоцелевое расширение интернет-почты (MIME) — дополняющий протокол, позволяющий передавать сообщения, используя SMTP-данные, которые не имеют вид ASCII. MIME — не почтовый протокол и не отменяет SMTP; он только его расширяет.

MIME преобразовывает данные, отличающиеся от ASCII, к виду ASCII и доставляет их клиенту SMTP через Интернет. Сервер SMTP на приемной стороне получает данные в виде ASCII и доставляет к MIME, чтобы преобразовать данные в первоначальный вид.

Можно упрощенно сказать, что MIME — это набор программного обеспечения, который преобразует данные, не представленные в ASCII, в данные ASCII и, соответственно, наоборот.

MIME определяет пять заголовков, которые могут быть дополнены к исходной секции заголовков SMTP для определения параметров преобразования:

- MIME – Version (MIME – Версия).
- Content – Type (Содержание – Тип).
- Content — Transfer – Encoding (Содержание – Передача – Кодирование).
- Content – Id (Содержание – Идентификатор).
- Content – Description (Содержание — Описание).

MIME – Version

Доставка почты

Доставка почты от отправителя к получателю проходит через три стадии

Первая стадия

На первой стадии электронная почта проходит через пользовательского агента в локальный сервер. Почта, возможно, сразу не посылается на удаленный сервер, поскольку он может быть недоступен к этому моменту. Поэтому почта накапливается в локальном сервере, пока ее не удастся отправить. Пользовательский агент использует программное обеспечение SMTP-клиента, локальный сервер использует программное обеспечение SMTP-сервера.

Вторая стадия

На втором шаге электронная почта идет с помощью локального сервера, который теперь действует как клиент SMTP. Электронная почта доставляется удаленному серверу, но не к удаленному агенту пользователя. Если бы SMTP был принятым сервером, всегда можно было бы обработать прибывшую почту в любой момент времени. Однако люди часто выключают свой компьютер до конца дня, а мини-компьютер или переносные компьютеры зачастую нормально не работают. Обычно организации предназначают свой компьютер для принятия электронной почты и постоянной работы в качестве программного сервера. Электронная почта получается с помощью такого сервера и накапливается в почтовом ящике для дальнейшего использования.

Третья стадия

На третьей ступени удаленный агент пользователя применяет протокол POP3 или IMAP4 (оба протокола обсуждаются в следующих секциях), чтобы запустить почтовый ящик и получить почту.

Протоколы почтового доступа

Первая и вторая стадия доставки почты используют SMTP. Однако SMTP не включен в третью стадию, потому что STMP "проталкивает" сообщение от отправителя к получателю, даже если получатель этого не желает. Работу SMTP начинает отправитель, а не получатель. С другой стороны, третья стадия нуждается в протоколе, который "притягивает" сообщение, и эта операция должна начинаться у получателя. Третья ступень использует протокол почтового доступа.

В настоящее время применяются два протокола: Post Office Protocol, Version 3 — POP3 и Internet Mail Access Protocol, Version 4 — IMAP4.

POP3

Post Office Protocol, Version 3 (POP3) — протокол простой, но ограниченный функционально. Программное обеспечение клиента POP3 устанавливается в компьютере получателя; программное обеспечение POP3-сервера устанавливается в почтовом сервере.

Почтовый доступ стартует от клиента, когда пользователю надо загрузить его электронную почту из почтового ящика в почтовый сервер. Клиент (агент пользователя) устанавливает с сервером порт 110 и далее посылает ему имя и пароль для доступа к почтовому ящику. Пользователь может затем перечислить и отыскать почтовые сообщения одно за другим.

POP3 имеет два режима: режим удаления и режим сохранения. В режиме удаления почта удаляется из почтового ящика после каждого запроса. В режиме сохранения почта остается в почтовом ящике после запроса. Режим удаления обычно используют, когда пользователь постоянно работает с компьютером и может сохранить и упорядочить почту после чтения и ответа. Режим сохранения применяют, когда пользователь имеет доступ к своей почте через первичный компьютер (например, переносной компьютер). Почта читается, но сохраняется в системе для дальнейшего запроса и упорядочения.

IMAP4

Другой протокол почтового доступа к сообщениям Интернета — Internet Mail Access Protocol, Version 4 (IMAP4). IMAP4 похож на POP3, но имеет некоторые особенности: IMAP4 более мощный и более сложный.

POP3 является несовершенным по нескольким причинам. Он не позволяет пользователю организовать почту на сервере; пользователь не может иметь различные "папки". (Конечно, пользователь может организовать папки на собственном компьютере.) В дополнение к этому, POP3 не позволяет пользователю частично проверить содержание почты перед загрузкой.

IMAP4 обеспечивает следующие дополнительные функции:

- Пользователь может проверить заголовки электронной почты перед загрузкой.
- Пользователь может искать содержимое электронной почты для специальных строк-символов перед загрузкой.
- Пользователь может частично загружать электронную почту. Это полезно в специальных случаях, если ресурсы ограничены и электронная почта содержит сообщения различного типа, требующие больших ресурсов.
- Пользователь может создавать, удалять или переименовывать почтовый ящик почтового сервера.
- Пользователь может создавать иерархию почтовых ящиков в папке для накопления электронной почты.

Почта на основе WEB

Услуги электронной почты сегодня могут быть обеспечены для пользователей WEB-сайтов (Yahoo,Yandex и.т. д).

Идея такой связи проста: передача письма проходит с помощью протокола HTTP (см. следующие лекции). Передача сообщения от передающего сервера к входящему почтовому серверу проходит с помощью протокола SMTP. В конечном итоге, сообщение от входящего сервера (он же WEB-сервер) достигает сервера пользователя Б, используя протокол HTTP. Если пользователь Б хочет получить это сообщение, он посылает запрос на свой WEB-сайт (например, YANDEX). WEB-сайт высылает форму, которая содержит запрос логина (зарегистрированного имени пользователя) и пароля, затем передает сообщение на компьютер пользователя Б в формате HTML.

Лекция 16. Протоколы HTTP и WWW

World Wide Web (WWW)

– это хранилище информации, размещенной во всем мире и соединенной воедино. WWW – уникальная комбинация гибкости, мобильности дружественных пользователю свойств, что отличает ее от других служб, обеспечиваемых с помощью Интернета.

WWW-проект был инициирован CERN (Center European Laboratory for Practice Physics), чтобы создать систему для обработки распределенных ресурсов, необходимых для научных исследований.

WWW сегодня — распределенная система клиент-сервер обслуживания, в которой клиент, использующий браузер1, может иметь доступ к этой службе с применением сервера. Однако обеспечиваемая служба распределяется по многим местам, называемым вместе websites.

Архитектура

Эта услуга может обеспечиваться во многих местах, которые называются сайтами, как это показано на рис. 16.1.

Каждый сайт содержит одну или более ссылок на веб-документы. Каждая веб-страница может содержать линк (связь) с другими страницами на том же самом сайте. Страницы могут быть вызваны для работы с браузерами. Рассмотрим сценарий, показанный рис. 16.1. Клиенту нужна информация, которая принадлежит сайту A. Он посылает запрос через браузер, программа которого доставляет веб-документ. Запрос, который включает в себя адрес веб-сайта и веб-страницы (web-page), называется универсальным идентификатором ресурса — URL (Uniform Resource Locator), который будет рассмотрен далее. Сервер находит и посылает документ клиенту. Когда пользователь смотрит документ, он может найти ссылки на другие документы, включая веб-страницы сайта B. Ссылка содержит URL для нового сайта. Пользователь может рассмотреть другой интересующий его документ. Клиент посылает другой запрос к новому сайту и вызывает другую страницу.

Архитектура браузеров

Ве́б-обозрева́тель, или **бра́узер** (от англ. *Web browser*) — это программное обеспечение для поиска, просмотра веб-сайтов, то есть для запроса веб-страниц для их обработки, вывода и перехода от одной страницы к другой.

. По данным компании Net Applications в мае 2008 года рыночная доля браузеров со строкой User Agent MSIE составляла 73,75 %, Firefox — 18,41 %, Safari — 6,37 %, Netscape — 0,62 %, Opera — 0,71 %, Mozilla Suite — 0.08 %, Opera Mini — 0.05 %, Microsoft Pocket Internet Explorer — 0.02 %, PlayStation — 0.03 %, Konqueror — 0.02 %, Blazer — 0.02 %, WebTV — 0.01 %, ACCESS NetFront — 0.01 %, Danger Web Browser, BlackBerry, ANT Galio, Lotus Notes, iCab — около 0 % [3].

Каждый браузер обычно содержит три части: контроллер, клиентский протокол и интерпретатор. Контроллер получает входную информацию от клавиатуры или "мыши" и использует клиентскую программу для доступа к документам. После того как документ доступен, контроллер применяет один из интерпретаторов, чтобы отобразить документ на экране. Клиентские программы могут использовать один из методов (протоколов) – такие как HTTP, FTP, TELNET. Интерпретатор может быть HTML или Java, JavaScript, в зависимости от типа документа (рис. 16.2.).

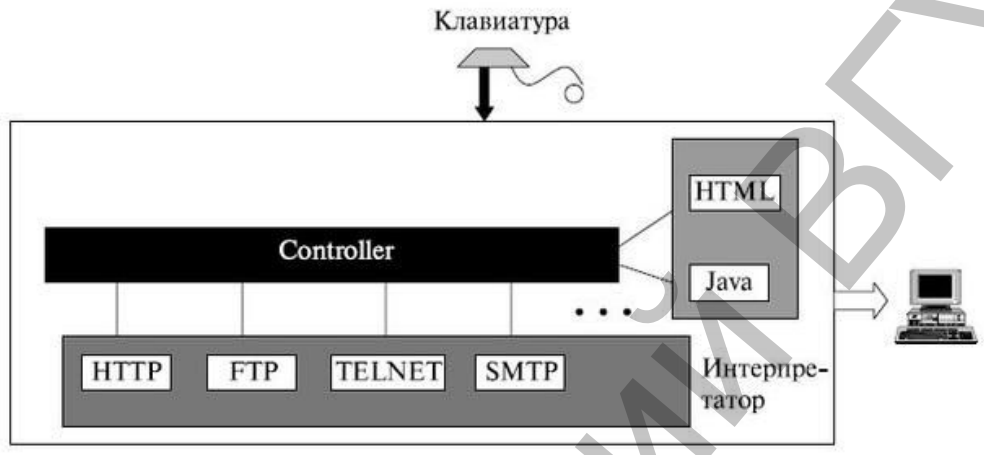


Рис. 16.2. Архитектура браузера

Сервер

Веб-страница хранится в веб-сервере. Каждый раз, когда поступает запрос клиента, ему посылается соответствующий документ. Для повышения эффективности работы обычно сервер хранит запрашиваемые файлы в кэш-памяти, это ускоряет поиск при запросе. Сервер более эффективен, если он может выполнять параллельные процессы или является многопроцессорным. В этом случае он способен отвечать одновременно на несколько запросов.

Унифицированный локатор ресурса — URL (Uniform Resource Locator) Единый указатель ресурсов

Клиент, который хочет вызвать веб-страницу, должен располагать ее адресом. Чтобы обеспечить доступ к документам, разбросанным во всем мире, существует протокол передачи гипертекста (HTTP – Hypertext Transfer Protocol). Индивидуальный индикатор ресурса — URL (Uniform Resource Locator) — стандарт для любой заданной информации в Интернете. URL определяет четыре элемента: протокол, хост, порт и путь (рис. 16.3.).

Протокол (метод) – программа клиент-сервер, используемая для доставки документа. Несколько различных протоколов могут доставлять документ; среди них Gopher, FTP, HTTP, News и TELNET. На сегодня наиболее общий протокол — HTTP.

Хост – компьютер, где находится информация, хотя имя компьютера может быть псевдонимом. Веб-страницы обычно накапливаются в компьютерах, и компьютеры дают псевдонимы именам, которые обычно начинаются с символов "www". Однако это не обязательно, поскольку хост

может быть с любым именем, данным компьютеру, который является хостом веб-страницы.

Рис. 16.3. Универсальный идентификатор ресурса — URL

URL иногда может содержать номер порта сервера. Если порт включен, он должен быть вставлен между хостом и путем и должен быть отделен от хоста двоеточием.

Путь — имя пути к файлу, где находится информация. Заметим, что путь сам может содержать "слеши" (наклонные черточки), которые в операционной системе UNIX отделяют директории от поддиректорий и файлов.

Cookies

«Куки» (от англ. Cookies — печенье) — короткий фрагмент текста в протоколе HTTP, присылаемый сервером веб-клиенту (обычно браузеру). Применяется для сохранения данных на стороне пользователя, на практике обычно используется для:

- авторизации
- отслеживания состояния сессии
- ведения статистики о пользователях

Создание и хранение cookies

Создание и сохранение cookies (на сленге — "плюшки") зависит от клиента; однако имеются общие принципы.

1. Когда сервер получает запрос от клиента, он сохраняет информацию о клиенте в файле или строке. Информация может включать в себя доменное имя клиента, содержание cookie (собранная информация сервера о клиенте, такая как имя, регистрационный номер и т. п.), метку времени и другую информацию, которая зависит от применения.
2. Сервер включает cookie в ответ, который посылается клиенту.
3. Когда клиент получает ответ, браузер накапливает cookie в директории "cookie", которая сохраняет эту информацию под доменным именем сервера.

Использование cookies

Когда клиент посылает запрос серверу, браузер просматривает директорию "cookie". Если cookie найдены, они посылаются серверу. Они включаются в запрос. Когда сервер получает запрос, он узнает, что этот клиент "старый". Заметим, что содержание cookie никогда не читается браузером и не раскрывается пользователем. Они создаются сервером и

"потребляются" сервером. Рассмотрим теперь cookies, используемые для четырех целей, которые были упомянуты ранее.

• Сайту, имеющему строгий доступ, посылается cookie, только когда клиент регистрируется первый раз. Для повторного доступа только этого клиента допускается посылать соответствующий cookie.

• Электронный магазин (e-commerce) может использовать cookie. Когда клиент выбирает предмет и помещает его в корзину, он посылает браузеру cookie, которое содержит информацию об этом предмете, такую как номер и цену за единицу. Если клиент выбирает второй предмет, cookie обновляется новой информацией и так далее. Когда клиент заканчивает покупки и хочет рассчитаться, вызывается последний cookie и подсчитывается вся сумма стоимости покупок.

• Веб-портал использует cookie следующим образом. Когда пользователь выбирает свою страничку, порождается и посылается cookie. Если сайт посещается вновь, cookie посылается серверу, чтобы показать, что это ожидаемый клиент.

• Cookie также применяется рекламными агентствами. Рекламные агентства могут предоставлять полосу-баннер (banner) в дополнение к главному веб-сайту, который часто посещают пользователи. Рекламное агентство поддерживает только URL, который дает адрес, вместо самого баннера. Когда пользователь посещает главный веб-сайт и вызывает икону (изображение), размещенное рекламной корпорацией, посылается запрос в рекламное агентство. Рекламное агентство посылает баннер, например, графический файл (GIF — Graphic Interchange Format), но также и cookie с ID пользователя. Любое будущее использование баннера добавляется к базе данных, к профайлам веб-окружения пользователя. Коммерческое агентство собирает интересующую информацию пользователя и может продать другому пользователю. Это весьма сомнительное использование cookie. Следует надеяться только на законы, защищающие личную и коммерческую тайну.

Протокол передачи гипертекстовых файлов (HTTP)

HTTP (англ. HyperText Transfer Protocol — «протокол передачи гипертекста») — протокол прикладного уровня передачи данных в первую очередь в виде текстовых сообщений. Основой HTTP является технология «клиент-сервер», то есть предполагается существование потребителей (клиентов), которые инициируют соединение и посылают запрос, и поставщиков (серверов), которые ожидают соединения для получения запроса, производят необходимые действия и возвращают обратно сообщение с результатом.

HTTP-функции подобны комбинации FTP и SMTP. Они похожи на FTP, потому что он передает файлы и использует услуги TCP/IP. Однако он гораздо проще, чем FTP, потому что использует только один TCP (хорошо известный порт 80). Нет никакого отдельного соединения управления; между клиентом и сервером передаются только данные.

HTTP похож на SMTP, потому что данные, передаваемые между клиентом и сервером, выглядят точно так же, как SMTP-сообщения. В дополнение, формат сообщений управляется с помощью MIME-подобного заголовка. Однако HTTP отличается от SMTP способом, которым сообщения посылаются от клиента к серверу и от сервера к клиенту. В отличие от SMTP, сообщения HTTP не предназначены для чтения людьми; они читаются и интерпретируются HTTP-сервером и HTTP-клиентом (браузером). SMTP-сообщения сохраняются и передаются, а HTTP-сообщения доставляются непосредственно.

Идея HTTP очень проста. Клиент посылает запрос, который выглядит как почтовое сообщение к серверу. Север посылает ответ, который выглядит как почтовый ответ к клиенту. Сообщения запроса и ответа переносят данные в виде формата, подобного MIME.

Команды от клиента к серверу вставляются в сообщение запроса, похожее на письмо. Содержание затребованного файла или другая информация вставляется в ответное сообщение, подобное письму.

HTTP-переходы

HTTP использует услуги TCP, потому что сам HTTP – протокол, не основанный на смене состояния. Клиент инициализирует переход посылкой сообщения запроса. Сервер отвечает посылкой ответа.

Сообщение запроса

Линейка запроса содержит тип запроса, URL, и версию HTTP.

• Тип запроса (Request type). В версии HTTP 1.1 определены несколько типов запросов. Типы запроса разделяются по категориям несколькими методами, которые мы обсудим позднее.

• Унифицированный локатор информационного ресурса (URL — Uniform Resource Locator). Клиент, который хочет иметь доступ к странице, нуждается в адресе, чтобы осуществить доступ к документам, распределенным по всему миру. HTTP пользуется концепцией локаторов. URL — стандарт для определения любого вида информации в Интернете. URL определяет четыре элемента: метод, хост, компьютер, порт и путь.

Методы

Поле типа запроса в сообщении запроса определяет несколько видов сообщений, называемых методы. Метод запроса – реальная команда или запрос, с которым клиент выходит к серверу. Мы здесь коротко обсудим цели некоторых методов.

GET

Метод GET используют, когда клиент хочет доставить документ от сервера. Адрес документа определяется в URL; это главный метод для доставки документа. Сервер обычно отвечает содержанием документа в "теле" ответного сообщения, если нет ошибки.

HEAD

Метод HEAD используется, когда клиент хочет получить некоторую информацию о документе, но не сам документ. Он подобен GET, но ответ от

сервера не содержит "тело".

POST

Метод POST используется, когда клиент обеспечивает информацией сервер. Например, это может быть нужно для посылки информации ввода к серверу.

PUT

Метод PUT используется клиентом, чтобы обеспечить накопление нового или обновленного документа на сервере. Этот документ включает в "тело" запрос и будет сохранен в месте, определенном URL.

PATCH

PATCH похож на PUT, за исключением того что запрос содержит только список отличий, которые нужно внести в существующий файл.

COPY

Метод COPY используется, чтобы скопировать файл в другое место. Дается место исходного файла в линейке запроса (URL); место пункта назначения дается в заголовке (обсуждается в разделе "Заголовок").

MOVE

Метод MOVE используется для переноса файла в другое место. Место файла источника дается в линейке запроса (URL); место пункта назначения дается в заголовке.

DELETE

Метод DELETE используется для удаления документа из сервера.

LINK

Метод LINK используется для создания ссылки (линка) или ссылок (линков) от одного документа к документу, расположенному в другом месте. Расположение файла дано в линейке запроса (URL); место пункта назначения дано в заголовке.

UNLINK

Метод UNLINK используется для удаления ссылок (линков), созданных методом LINK.

OPTION

Метод OPTION используется клиентом для запроса сервера о доступности опции.

Сообщение ответа

Сообщение ответа содержит линейку состояния, заголовок и иногда "тело"

Продолжительное соединение в сравнении с непродолжительным

HTTP-версия 1.0 — непродолжительное соединение, в то время как продолжительное соединение есть по умолчанию в версии 1.1.

Непродолжительное соединение

В непродолжительном соединении одно TCP-соединение делается по каждому запросу/ответу.

Ниже перечислены шаги в этой стратегии.

1. Клиент открывает TCP-соединение и посылает запрос.

2. Сервер посылает ответ и заканчивает соединение.

Клиент читает данные, пока не наталкивается на метку конца файла, затем он закрывает соединение. В этой стратегии для N различных картинок в различных файлах соединение должно быть открыто и закрыто N раз. Непродолжительное соединение вызывает перегрузку сервера, потому что сервер нуждается в N различных буферах и требует процедуру медленного старта каждый раз, когда открывает соединение.

Продолжительное соединение

HTTP задает продолжительное соединение по умолчанию. В продолжительном соединении сервер оставляет соединение открытым для большого числа запросов после посылки ответа. Сервер заканчивает соединение по запросу клиента или если достигается заданное время (таймаут). Передатчик обычно посылает длину данных при каждом ответе. Однако имеется несколько случаев, когда передатчик не знает длины данных — когда документ создается динамически или активно. В этом случае сервер информирует клиента, что длина не известна, и закрывает соединение после посылки данных, так как клиент знает, что был достигнут конец передачи данных.

Сервер-посредник (Proxy server)

HTTP поддерживается прокси-серверами. Это компьютеры, которые содержат программные средства, предназначенные для защиты локальной и корпоративной сети от несанкционированного доступа или опасных приложений. Proxy(промежуточный)-сервер – это компьютер, который сохраняет копии ответов на прежние запросы. При наличии прокси-сервера клиент HTTP посылает запрос к именно к нему. Прокси-сервер проверяет свою кэш-память. Если ответ в кэше не сохранен, прокси-сервер посылает запрос к соответствующему серверу. Входящий ответ посылается к прокси-серверу и запоминается для будущих запросов от клиентов.

Прокси-сервер уменьшает загрузку исходного сервера, уменьшает нагрузку, улучшает реакцию. Однако чтобы использовать прокси-сервер, клиент должен быть конфигурирован для доступа к proxy вместо целевого сервера.

Пример диалога HTTP

Запрос:

```
GET /wiki/HTTP HTTP/1.1
Host: ru.wikipedia.org
User-Agent: Mozilla/5.0 (X11; U; Linux i686; ru;
rv:1.9b5) Gecko/2008050509 Firefox/3.0b5
Accept: text/html
Connection: close
```

Ответ:

```
HTTP/1.0 200 OK
Server: nginx/0.6.31
Content-Language: ru
Content-Type: text/html; charset=utf-8
Content-Length: 1234
Connection: close
```

Лекция 17. Сотовые системы связи

История GSM

В начале 1980-х в Европе, особенно в Скандинавии и Великобритании, а также и во Франции и Германии наблюдался быстрый рост мобильной телефонной связи. Каждая страна разработала свою собственную систему, которая была несовместима со всеми другими в части оборудования и функционирования. Такая ситуация была нежелательна, потому что подвижная аппаратура не только была ограничена функционированием в пределах национальных границ, которых в объединенной Европе было все больше и больше, но очень ограничивало рынок для каждого типа оборудования. Страдали и сбыт, и окупаемость расходов на мобильную связь.

В 1982 г. Конференция европейских почт и телекоммуникаций (CEPT — Conference of European Post and Telecommunication) сформировала группу GSM (Group Special Mobile) для изучения и разработки европейской мобильной наземной системы. Предложенная система должна была соответствовать некоторым критериям:

- хорошее субъективное качество речи;
- низкая стоимость оконечных устройств и обслуживания;
- поддержка международной подвижной связи;
- способность обслуживать малогабаритные терминалы;
- обеспечение диапазона новых услуг и средств;
- эффективное использование радиодиапазона;
- совместимость с ISDN.

В 1989 г. ответственность за разработку GSM была передана Европейскому институту стандартов в области телекоммуникаций (ETSI — European Telecommunication Standards Institute). Первые спецификации GSM были изданы в 1990 г. Коммерческая эксплуатация была начата в середине 1991 г., и к 1993 г. существовало 36 сетей GSM в 22 странах.

Хотя GSM стандартизировано в Европе, это не только европейский стандарт. Сегодня работают более чем 200 сетей GSM (включая DCS-1800 и PCS-1900) в 110 странах во всем мире. В начале 1994 г. во всем мире было 1,3 миллиона абонентов. Сегодня эта цифра выросла до 70 миллионов. Северная Америка имеет разновидность GSM, названную PCS-1900.

Системы GSM существуют теперь на каждом континенте, и сокращение "GSM" (Global System for Mobile Communications) теперь обозначает "Глобальная система для мобильной связи".

Создатели GSM выбрали не опробованную (в то время) цифровую систему, в противоположность применявшимся тогда аналоговым сотовым системам, подобным усовершенствованной системе мобильной связи в Соединенных Штатах (AMPS — Advances Mobile Phone System) и системе с полным доступом (TACS — Total Access Communications System) в Великобритании. Разработчики верили, что усовершенствованные алгоритмы

сжатия информации и применение цифровых сигнальных процессоров позволят выполнить поставленные выше задачи и непрерывно совершенствовать систему в смысле качества и стоимости. Более чем 8000 страниц рекомендаций GSM дали возможность построить гибкую и конкурентоспособную систему и обеспечить достаточную стандартизацию, чтобы гарантировать надлежащее межсетевое взаимодействие между компонентами системы, — для этого были созданы описания интерфейсов каждого из функциональных объектов, определенных в системе.

1.2. Услуги, обеспечиваемые GSM

С самого начала разработчики GSM хотели гарантировать совместимость ее с цифровой сетью интегрального обслуживания ISDN в части услуг и передачи сигналов управления. Однако ограничения радиопередачи по пропускной способности и стоимости не позволяли достигнуть стандартной для ISDN скорости передачи информации в битах B-канала 64 Кбит/с.

В соответствии с определением ITU-T (Международного союза электросвязи), телекоммуникационные услуги могут быть разделены на основные и дополнительные услуги. Основная услуга, поддерживаемая GSM, — телефонная связь. Речь закодирована в цифровой форме и передается через сеть GSM как цифровой поток. Существуют также экстренные службы, где, набирая три цифры (например 911), можно получить связь с ближайшим пунктом этой службы.

GSM предоставляет следующие услуги:

1. телефонная связь (совмещается со службой сигнализации: охрана квартир, сигналы бедствия и пр.);
2. передача коротких сообщений;
3. доступ к службам "Видеотекст", "Телетекст";
4. служба "Телефакс" (группа 3).

Пользователи GSM могут обмениваться данными со скоростью свыше 9600 битов в сек.:

- с пользователями обычной телефонной сети (POTS — Plain Ordinary Service);
- с пользователями цифровой сети интегрального обслуживания (ISDN);
- с пользователями сети передачи данных общего пользования с пакетной коммутацией (PSPDN — Packet Switched Public Data Networks);
- с пользователями сети передачи данных общего пользования с коммутацией каналов (CSPDN — Circuit Switched Public Data Networks).

Стандарт GSM предусматривает передачу данных:

- асинхронно в дуплексном режиме со скоростями 300, 600, 1200, 2400, 4800 и 9600 бит/с через телефонные сети общего пользования;
- синхронно в дуплексном режиме со скоростями 1200, 2400, 4800 и 9600 бит/с через телефонные сети общего пользования, коммутируемые сети передачи данных общего пользования (CSPDN) и ISDN;

- в режиме доступа с помощью адаптера к пакетной асинхронной передаче данных со стандартными скоростями 300–9600 бит/с через коммутируемые сети пакетной передачи данных общего пользования (PSPDN);
- в режиме синхронного дуплексного доступа к сети пакетной передачи данных со стандартными скоростями 2400–9600 бит/с.

При передаче данных со скоростью 9,6 Кбит/с всегда задействуется канал связи с полной скоростью передачи. В случае передачи на скоростях ниже 9,6 Кбит/с могут использоваться полускоростные каналы связи.

При этом применяются разнообразные методы доступа и протоколов, таких как X.25 или X.32. Так как GSM — цифровая сеть, между пользователем и сетью GSM не требуется модем, хотя аудиомодем необходим в сети GSM для взаимодействия с обычной телефонной сетью.

Уникальная особенность GSM, которая отсутствует в старых аналоговых системах, — Служба передачи коротких сообщений (SMS — Short Message Service). SMS — двунаправленное обслуживание коротких алфавитно-цифровых (не свыше 160 байтов) сообщений. Сообщения транспортируются способом с промежуточным накоплением (store-and-forward fashion). При соединении между двумя абонентами SMS-сообщение можно передать третьему абоненту и получить подтверждение. SMS может также использоваться в широковещательном режиме, чтобы послать такие сообщения, как модификации трафика или модификации новостей. Сообщения могут также быть сохранены в SIM-карте абонента (SIM — Subscriber Identification Module) для использования в дальнейшем.

Дополнительно стандартизован широкий спектр особых услуг (включение в закрытую группу пользователей, передача вызова, оповещения о тарифных расходах).

Они включают несколько вариантов переадресации вызова и запрет на вызов при входящей и исходящей связи, например, при роуминге (изменении местоположения) в другой стране. Осуществляются такие услуги, как идентификация вызывающего абонента, режим "ждущий вызов", многосторонняя (конференц-) связь.

Важной услугой признается стандарт "закрытой группы". Закрытая группа пользователей (CUG — Closed User Group) — это группа абонентов, в которой устанавливается соединение и происходит обмен информацией преимущественно в пределах этой группы. Возможно предоставление входящей и исходящей связи вне этой группы. При этом абонентам при связи внутри группы предоставляются льготы. Примером такой группы является связь внутри членов одной семьи. Обычно компании сотовой связи предоставляют при такой связи абонентам пониженный тариф или бесплатную связь.

Следующая услуга: сопровождающий вызов обеспечивает переадресацию входящего вызова на номер абонента стационарной сети.

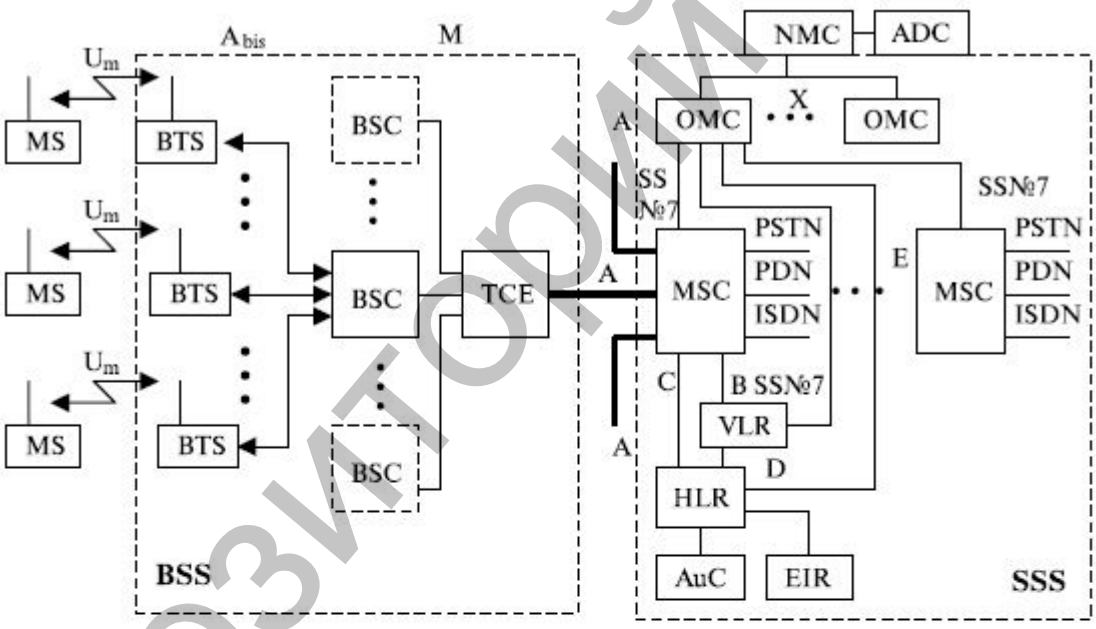
В режиме "ждущий вызов" при занятости абонента входящий вызов ставится в режим ожидания освобождения предыдущего соединения. Абонент, к которому адресован вызов, получает предупреждающий сигнал. Абонент может:

- завершить предыдущий вызов;
- кратковременным нажатием рычага трубки перейти на новое соединение;
- после разговора по новому соединению вернуться к старому и повторить это многократно.

Все перечисленные соединения относятся к группе дополнительных видов обслуживания, которые реализуются в сетях ISDN и современных сетях PSTN.

1.3. Архитектура сети GSM

Сеть GSM состоит из нескольких функциональных объектов, функции и интерфейсы которых показаны на рис. 1.1.



ADC – Administration Center	Административный центр
AuC – Authentication	Центр аутентификации
BTS – Base Telephone Station	Базовая приемо-передающая станция
BSC – Base Station Controller	Контроллер базововй станции
BSS – Base Station System	Подсистема базовой станции
EIR – Equipment Identification Register	Регистр идентификации оборудования
HLR – Home Location Register	Домашний регист местоположения
ISDN – Integrated Service Digital Network	Цифровая сеть с интеграцией служб
MS – Mobile Station	Мобильная станция
MSC – Mobile Switching Center	Цент коммутации мобильной связи
NMC – Network Management Center	Центр управления сетью
OMC– Operation and Maintenance Center	Центр эксплуатации и технического обслуживания
PDN – Packet Data Networks	Сеть пакетной коммутации
PSTN – Public Switched Telephone Network	Телефонная сеть общего пользования
SSS – Switching Subsystem	Коммутационная подсистема
TCE – Transcoder Equipment	Транскодер
VLR – Visit Location Register	Визитный регистр местоположения

Рис. 1.1. Архитектура сети и интерфейсы GSM

Сеть GSM включает три основные части:

- мобильные станции (MS), которые перемещаются с абонентом;
- подсистему базовых станций (BSS), которая управляет радиолинией связи с мобильной станцией;
- подсистему сети (NSS), главная часть которой — центр коммутации мобильной связи (MSC) — выполняет коммутацию между мобильными станциями и между мобильными или стационарными сетевыми пользователями. MSC также управляет работой, связанной с передвижением абонента.

На рис. 1.1 не показан центр обслуживания, который наблюдает за надежным функционированием и изменениями на сети. Мобильная станция (MS) и подсистема базовых станций (BSS) связываются по Um-интерфейсу, также известному как "воздушный интерфейс" или радиолиния связи. Подсистема базовых станций взаимодействует с центром коммутации мобильной связи по A интерфейсу.

1.3.1. Мобильная станция

Мобильная станция (MS) состоит из подвижной аппаратуры (терминал) и карты с интегральной схемой, включающей микропроцессор, которая называется модулем абонентской идентификации (SIM — Subscriber Identification Module). SIM-карта обеспечивает при перемещении пользователя доступ к оплаченным услугам независимо от используемого терминала. Вставляя SIM-карту в другой терминал GSM, пользователь может принимать вызовы, делать вызовы с этого терминала и получать другие услуги.

Подвижная аппаратура однозначно определяется с помощью международного опознавательного кода мобильного оборудования (IMEI — International Mobile Equipment Identity). SIM-карта содержит международный опознавательный код мобильного абонента (IMSI — International Mobile Subscriber Identity), используемый для идентификации абонента, секретный код для удостоверения подлинности и другую информацию. IMEI и IMSI независимы — это дает возможность обеспечить наиболее вероятное опознавание личности при передвижении абонента. SIM-карта может быть защищена против неправомочного использования паролем или личным номером.

Применяются три типа оконечного оборудования подвижной станции:

- MT0 (Mobile Termination 0) — многофункциональная подвижная станция, в состав которой входит терминал данных с возможностью передачи и приема данных и речи;
- MT1 (Mobile Termination 1) — подвижная станция с возможностью связи через терминал с ISDN;

- MT2 (Mobile Termination 2) — подвижная станция с возможностью подключения терминала для связи по протоколу МККТТ V- или X-серий.

Терминальное оборудование может состоять из оборудования одного или нескольких типов, такого как телефонная трубка с номеронабирателем, аппаратура передачи данных (DTE), телекс и т. д.

Различают следующие типы терминалов: TE1 (Terminal Equipment 1) — терминальное оборудование, обеспечивающее связь с ISDN; TE2 (Terminal Equipment 2) — терминальное оборудование, обеспечивающее связь с любым оборудованием через протоколы МККТТ V- или X-серий (связь с ISDN не обеспечивает). Терминал TE2 может быть подключен как нагрузка к MT1 (подвижной станции с возможностью связи с ISDN) через адаптер TA.

1.3.2. Подсистема базовых станций

Подсистема базовых станций содержит два вида оборудования: базовая приемопередающая станция (BTS — Base Transceiver Station) и контроллер базовой станции (BSC — Base Station Controller). Они взаимодействуют через стандартизированный интерфейс A_{bis} (см. рис. 1.1).

На базовой приемопередающей станции размещается приемопередатчик, который для одной определенной соты реализует протоколы радиолинии с передвижной станцией. В большом городе обычно размещено большое количество BTS. Поэтому основные требования к BTS — прочность, надежность, портативность и минимальная стоимость.

Контроллер базовой станции управляет радиоресурсами для одного или более BTS: выбором и установлением соединения по радиоканалу, скачком частоты и хэндовером (переключением), как это будет показано ниже. BSC подключается между базовой приемопередающей станцией (BTS) и центром коммутации мобильной связи (MSC).

1.3.3. Коммутационная подсистема сети

Центр коммутации мобильный связи (MSC)

Центральный компонент подсистемы сети — центр коммутации мобильной связи (MSC). Он работает как обычный узел коммутации общедоступной телефонной сети (PSTN — Public Switched Telephone Network) или цифровой сети интегрального обслуживания (ISDN — Integrated Service Digital Network). Дополнительно он обеспечивает все функциональные возможности мобильного абонента, такие как регистрация, аутентификация, обновление местоположения, передача соединения (хэндовер) и маршрутизация вызова при передвижении абонента. Эти функции обеспечиваются совместно несколькими функциональными объектами, которые вместе формируют подсистему сети. MSC обеспечивает подключение к фиксированным сетям (таким как общедоступная телефонная сеть PSTN или цифровая сеть интегрального обслуживания ISDN). Передача сигналов между функциональными объектами в подсистеме сети использует ОКС № 7 (SS7) — отдельный канал сигнализации, такой же, как применяется для обмена в ISDN и в сетях общего пользования.

Центр коммутации подвижной связи обслуживает группу сот и обеспечивает все виды соединений, в которых нуждается в процессе работы подвижная станция. MSC аналогичен ISDN коммутационной станции и реализует интерфейс между фиксированными сетями (PSTN, PDN, ISDN и т. д.) и сетью подвижной связи. Он обеспечивает маршрутизацию вызовов и функции управления вызовами. Кроме выполнения функций обычной ISDN коммутационной станции на MSC возлагаются функции коммутации радиоканалов. К ним относятся "эстафетная передача", в процессе которой достигается непрерывность связи при перемещении подвижной станции из соты в соту, и переключение рабочих каналов в соте при появлении помех или неисправностях.

Каждый MSC обеспечивает обслуживание подвижных абонентов, расположенных в пределах определенной географической зоны (например, Москва и область). MSC управляет процедурами установления вызова и маршрутизации. Для телефонной сети общего пользования (PSTN) MSC обеспечивает функции сигнализации по протоколу ОКС №7, передачи вызова или поддержки других видов интерфейсов в соответствии с требованиями конкретного проекта.

MSC формирует данные, необходимые для выписки счетов за предоставленные сетью услуги связи, накапливает данные по состоявшимся разговорам и передает их в центр расчетов (биллинг-центр). MSC составляет также статистические данные, необходимые для контроля работы и оптимизации сети. Он же поддерживает процедуры безопасности, применяемые для управления доступами к радиоканалам.

MSC не только участвует в управлении вызовами, но также управляет процедурами регистрации местоположения и передачи управления, кроме передачи управления в подсистеме базовых станций (BSS). Регистрация местоположения подвижных станций необходима для обеспечения доставки вызова перемещающимся подвижным абонентам от абонентов телефонной сети общего пользования или других подвижных абонентов. Процедура передачи вызова позволяет сохранять соединения и обеспечивать ведение разговора, когда подвижная станция перемещается из одной зоны обслуживания в другую. Передача вызовов в сотах, управляемых одним контроллером базовых станций (BSC), осуществляется этим BSC. Когда передача вызовов происходит между двумя сетями, управляемыми разными BSC, то первичное управление осуществляется в MSC. В стандарте GSM также предусмотрены процедуры передачи вызова между сетями (контроллерами), относящимися к разным MSC. Центр коммутации осуществляет постоянное слежение за подвижными станциями, используя домашний регистр местоположения (HLR) и визитный регистр местоположения (VLR).

Домашний регистр местоположения (HLR — Home Location Register)

В HLR хранится та часть информации о местоположении какой-либо подвижной станции, которая позволяет центру коммутации доставить вызов определенной мобильной станции. Практически HLR представляет собой справочную базу данных о постоянно зарегистрированных в сети абонентах. В ней содержатся опознавательные номера и адреса, а также параметры подлинности абонентов, состав услуг связи, специальная информация о маршрутизации. Ведется регистрация данных об изменении местоположения и роуминге ("блуждании") абонента, включая данные о временном идентификационном номере подвижного абонента (TMSI — Temporary Mobile Subscriber Identity) и соответствующем визитном регистре местоположения (VLR). Регистр HLR содержит международный идентификационный номер подвижного абонента (IMSI — International Mobile Subscriber Identity), состав услуг связи, специальную информацию о маршрутизации. Он используется для опознавания подвижной станции в центре аутентификации (AUC — Authentication Center).

Домашний регистр местоположения (HLR) вместе с MSC обеспечивает маршрутизацию вызова и изменения местоположения (роуминг) мобильной станции и содержит всю административную информацию каждого абонента, зарегистрированного в соответствующей сети GSM, наряду с текущим местоположением мобильных станций. Местоположение мобильных станций находится обычно в форме адреса данной мобильной станции в VLR. Фактическая процедура маршрутизации будет описана позже. Логически существует только один HLR в сети GSM, хотя он может быть реализован как распределенная база данных. К данным, содержащимся в HLR, имеют дистанционный доступ все MSC и VLR сети, и, если в сети имеются несколько HLR, в базе данных содержится только одна запись об абоненте, поэтому каждый HLR представляет собой определенную часть общей базы данных сети об абонентах. Доступ к базе данных об абонентах осуществляется по номеру IMSI (IMSI — International Mobile Station Identity) или по MSISDN-номеру подвижной станции в сети ISDN (MSISDN — Mobile Station ISDN Number). К базе данных могут получить доступ MSC или VLR, относящиеся к другим сетям, в рамках обеспечения межсетевого роуминга абонентов.

Визитный регистр местоположения (VLR — Visit Location Register)

Второе основное устройство, обеспечивающее контроль над передвижением подвижной станции из зоны в зону, — визитный регистр местоположения VLR. С его помощью достигается функционирование подвижной станции за пределами зоны, контролируемой HLR. Когда в процессе перемещения подвижная станция переходит из зоны действия одного контроллера базовой станции BSC, объединяющего группу базовых станций, в зону действия другого BSC, она регистрируется новым BSC, и в VLR заносится информация о номере области связи, которая обеспечит доставку вызовов подвижной станции. Для сохранности данных,

находящихся в HLR и VLR, в случае сбоев предусмотрена защита устройств памяти этих регистров.

VLR включает в себя такие же данные, как и HLR, однако эти данные содержатся в VLR только до тех пор, пока абонент находится в зоне, контролируемой VLR.

В сети подвижной связи GSM соты группируются в географические зоны (LA — Location Area), которым присваивается свой идентификационный номер (LAC — Location Area Code). Каждый VLR содержит данные об абонентах в нескольких LA. Когда подвижный абонент перемещается из одной LA в другую, данные о его местоположении автоматически обновляются в VLR. Если старая и новая LA находятся под управлением различных VLR, то данные на старом VLR стираются после их копирования в новый VLR. Текущий адрес VLR абонента, содержащийся в HLR, также обновляется.

VLR обеспечивает также присвоение номера для услуг роуминга мобильной станции (MSRN — Mobile Station Roaming Number). Когда подвижная станция принимает входящий вызов, VLR выбирает его MSRN и передает его на MSC, который осуществляет маршрутизацию этого вызова к базовым станциям, находящимся рядом с подвижным абонентом.

Во время движения подвижная станция может покинуть зону, обслуживаемую одним MSC/VLR, и переместиться в зону, которую обслуживает другой MSC/VLR. В этом случае MSC/VLR участвует в передаче управления от одного MSC/VLR к другому. Он также присваивает новый временный мобильный опознавательный код станции TMSI (Temporary Mobile Subscriber Identity) и передает его в HLR. Новый MSC/VLR инициирует процедуру установления подлинности абонента и его оборудования. Кроме случая, когда подвижный абонент меняет зону местоположения, временный номер может периодически изменяться по решению оператора с целью защиты от злонамеренного перехвата номеров участников разговора. В этом случае процедура изменения идет также с использованием VLR, для доступа к VLR могут использоваться идентификационные номера IMSI, TMSI и MSRN.

В заключение отметим, что VLR — это локальная база данных в данной зоне, которая содержит информацию о подвижном абоненте. Применение VLR позволяет сократить число запросов HLR, и это снижает сетевой трафик и уменьшает время обслуживания.

В табл. 1.1, 1.2 и 1.3 приведены примеры состава долговременных и временных данных, хранящихся в HLR и VLR.

Таблица 1.1. Состав долговременных данных, хранящихся в HLR и VLR

1	Международный идентификационный номер подвижного абонента (IMSI- International Mobile Subscriber Identity)
2	Номер подвижной станции в сети ISDN (MSISDN)
3	Категория подвижной станции
4	Ключ засекречивания

5	Используемые пароли
6	Класс приоритетного доступа
7	Список разрешенных дополнительных видов обслуживания
8	Оповещение вызываемого абонента о номере вызывающего абонента
9	Разрешение/запрещение идентификации номера вызываемого абонента
10	График работы мобильной станции
11	Индекс закрытой группы пользователей
12	Код блокировки закрытой группы пользователей
13	Состав вызовов, которые могут быть переданы
14	Свойства закрытой группы пользователей
15	Льготы закрытой группы пользователей
16	Запрещенные исходящие вызовы в закрытой группе пользователей
17	Максимальное количество абонентов закрытой группы

Таблица 1.2. Состав временных данных находящихся в HLR

1	Временный номер подвижного абонента в VLR (TMSI- Temporary Mobile Subscriber Identity)
2	Параметры аутентификации и шифрования
3	Адрес регистра местоположения VLR
4	Код Зоны Местоположения
5	Номер соты при хэндовере
6	Регистрационные данные
7	Таймер ожидания ответа
8	Состав текущих паролей
9	Активность (есть/нет соединения)

Таблица 1.3. Состав временных данных находящихся в VLR

Временный идентификационный номер подвижного абонента (TMSI – Temporary Mobile Subscriber Identity)
Идентификаторы зоны расположения
Код Зоны Местоположения
Номер соты при хэндовере
Параметры аутентификации и шифрования

Регистры защиты и аутентификации

Для защиты и аутентификации используются два устройства: регистр идентификации оборудования (EIR — Equipment Identity Register) и центр аутентификации (AUC — Authentication Center). Регистр идентификации оборудования — база данных, которая содержит список всей допустимой к обслуживанию подвижной аппаратуры на сети, где каждая мобильная станция идентифицирована ее международным опознавательным кодом мобильного оборудования (IMEI). IMEI может быть маркирован как запрещенный к обслуживанию, если станция украдена или такого типа, который не обслуживается. Центр аутентификации — защищенная база данных, которая накапливает копии ключей засекречивания, хранящихся в

SIM-карте каждого абонента, и используется для аутентификации абонента и его оборудования, а также и шифрования для передачи по радиоканалу.

Каждый подвижный абонент имеет стандартный модуль подлинности абонента (SIM), который содержит: международный идентификационный номер (IMSI), свой индивидуальный ключ аутентификации (Ki), алгоритм аутентификации (A3).

С помощью записанной в SIM информации в результате взаимного обмена данными между подвижной станцией и сетью осуществляется полный цикл аутентификации и разрешается доступ абонента к сети.

EIR — регистр идентификации оборудования, содержит централизованную базу данных для подтверждения подлинности международного идентификационного номера оборудования подвижной станции (IMEI). Эта база данных относится исключительно к оборудованию подвижной станции. Она состоит из списков номеров IMEI, организованных следующим образом.

БЕЛЫЙ СПИСОК содержит номера IMEI, о которых есть сведения, что они закреплены за санкционированными подвижными станциями. Терминалу позволяют соединиться с сетью.

ЧЕРНЫЙ СПИСОК содержит номера IMEI подвижных станций, которые украдены, имеют некорректный тип мобильной станции для сети GSM или им отказано в обслуживании по другой причине. Терминалу не позволяют соединиться с сетью.

СЕРЫЙ СПИСОК содержит номера IMEI подвижных станций, у которых существуют проблемы, выявленные по данным программного обеспечения, но не являющиеся основанием для внесения в "черный список". Терминал находится под наблюдением сети ввиду возможных проблем.

К базе данных EIR получают дистанционный доступ MSC данной сети, а также MSC других подвижных сетей.

Как и в случае с HLR, сеть может иметь более одного EIR, при этом каждый EIR управляет определенной группой оборудования, имеющей свой идентификационный номер IMEI. В состав MSC входит транслятор, который при получении номера IMEI выбирает адрес EIR — он содержит данные о части оборудования, имеющей этот номер.

Оборудование эксплуатации и технического обслуживания

OMC (Operations and Maintenance Center) — центр эксплуатации и технического обслуживания, является центральным элементом сети GSM, который обеспечивает контроль и управление другими компонентами сети, а также контроль качества ее работы. OMC соединяется с другими компонентами сети GSM по каналам пакетной передачи протокола X.25. Он обеспечивает функции обработки аварийных сигналов, предназначенных для оповещения обслуживающего персонала, и регистрирует сведения об аварийных ситуациях в других компонентах сети. В зависимости от характера неисправности OMC позволяет обеспечить ее устранение автоматически или при активном вмешательстве персонала. Центр может

провести проверку состояния оборудования сети и прохождения вызова подвижной станции. OMC позволяет производить управление нагрузкой в сети. Функция эффективного управления включает сбор статистических данных о нагрузке от компонентов сети GSM, запись их в дисковые файлы и вывод на дисплей для визуального анализа. OMC обеспечивает управление изменениями программного обеспечения и базами данных о конфигурации элементов сети. Загрузка программного обеспечения в память может производиться из OMC в другие элементы сети или из них в OMC.

NMC (Network Management Center) — центр управления сетью, дает возможность рационального иерархического управления сетью GSM. Он обеспечивает эксплуатацию и техническое обслуживание на уровне всей сети, поддерживаемой центрами OMC, которые отвечают за управление региональными сетями. NMC отвечает за управление трафиком во всей сети и обеспечивает диспетчерское управление сетью при сложных аварийных ситуациях, как, например, выход из строя или перегрузка узлов. Кроме того, он контролирует состояние устройств автоматического управления, задействованных в оборудовании сети, и отражает на дисплее состояние сети для операторов NMC. Это позволяет операторам контролировать региональные проблемы и при необходимости оказывать помощь OMC, обслуживающему конкретный регион. Таким образом, персонал NMC знает состояние всей сети и может дать указание персоналу OMC изменить стратегию решения региональной проблемы.

NMC следит за состоянием маршрутов сигнализации и соединений между узлами, чтобы не допускать условий для возникновения перегрузки в сети. Контролируются также маршруты соединений между сетью GSM и PSTN во избежание распространения условий перегрузки между сетями. При этом персонал NMC координирует вопросы управления сетью с персоналом других NMC. NMC обеспечивает также возможность управления трафиком для сетевого оборудования подсистемы базовых станций (BSS). Операторы NMC в экстремальных ситуациях могут задействовать такие процедуры управления, как "приоритетный доступ", когда только абоненты с высоким приоритетом (экстренные службы) могут получить доступ к системе.

NMC может брать на себя ответственность в каком-либо регионе, когда местный OMC не способен обслуживать нагрузку, при этом OMC действует в качестве транзитного пункта между NMC и оборудованием сети. NMC обеспечивает операторов функциями, аналогичными функциям OMC.

NMC является также важным инструментом планирования сети, так как контролирует сеть и ее работу на сетевом уровне, а следовательно, снабжает планировщиков сети данными, определяющими нагрузочные параметры сети.

ADC (Administration Center) — административный центр — сетевая служба, ответственная за организацию связи, административное управление сетью и соблюдение установленных правил доступа.

TCE (Transcoder Equipment) — транскодер, обеспечивает преобразование выходных сигналов передачи речи и данных MSC (64 Кбит/с ИКМ) к виду, соответствующему рекомендациям GSM по радиоинтерфейсу. В соответствии с этими требованиями скорость передачи речи, представленной в цифровой форме, составляет 13 Кбит/с. Этот канал передачи цифровых речевых сигналов называется "полноскоростным". Стандартом предусматривается в перспективе использование полускоростного речевого канала (скорость передачи 6,5 Кбит/с).

Снижение скорости передачи обеспечивается применением специального речепреобразующего устройства, использующего кодирование с линейным предсказанием (LPC — Linear Predictive Coding), долговременное предсказание (LTP — Long Term Predicting), возбуждение регулярной импульсной последовательностью (RPE — иногда называется RELP).

Транскодер обычно располагается вместе с MSC. Передача цифровых сообщений по направлению к контроллеру базовых станций (BSC) ведется с добавлением к потоку со скоростью передачи 13 Кбит/с дополнительных битов (stuffing). Таким образом, скорость передачи данных становится 16 Кбит/с. Затем осуществляется уплотнение с кратностью 4 в стандартный канал 64 Кбит/с. Так формируется определенная рекомендациями GSM

30-канальная ИКМ линия, обеспечивающая передачу 120 речевых каналов. Шестнадцатый канал (64 Кбит/с) (slot) выделяется отдельно для передачи информации сигнализации и часто содержит сигналы ОКС № 7 или процедуры доступа к звену передачи данных для канала "D" — LAPD (Link Access Procedure for the D - channel).

В других каналах (64 Кбит/с) могут передаваться также пакеты данных, согласующиеся с протоколом X.25 МККТТ.

1.4. Основные принципы организации сети GSM

1.4.1. Внутренние интерфейсы GSM

Внутренние интерфейсы показаны на рис. 1.1 и перечислены в табл. 1.4.

Таблица 1.4. Типы внутренних интерфейсов сети GSM

Тип	Связь между устройствами
A	MSC-BSS
A_{bis}	BSC-BTS
B	MSC-VLR
C	MSC-HLR
D	HLR-VLR
E	MSC-MSC
O	BSC-OMC
M	BSC-TCE
U_m	MS-BTS

X	OMC-OMC

Примечание: X-интерфейс предназначен для связи OMC различных GSM

A-интерфейс. Интерфейс между MSC и BSS (подсистема базовых станций –BSC- BTS) обеспечивает передачу сообщений для управления BSS, передачи вызова (хэндовер), управления при изменении местоположения. A-интерфейс объединяет каналы связи и линии сигнализации. Последние используют протокол ОКС № 7 МККТТ. Полная спецификация A-интерфейса соответствует требованиям серии 08 Рекомендации ETSI/GSM.

Abis-интерфейс. Интерфейс служит для связи между BSC с BTS и определен Рекомендациями ETSI/GSM для процессов установления соединений и управления оборудованием, передача осуществляется цифровыми потоками со скоростью 2,048 Мбит/с. Возможно использование физического интерфейса 64 Кбит/с.

B-интерфейс. Интерфейс между MSC и VLR. Когда MSC необходимо определить местоположение подвижной станции, он обращается к VLR. Если подвижная станция инициирует процедуру изменения местоположения, то MSC информирует свой VLR, который заносит всю изменяющуюся информацию в свои регистры. Эта процедура происходит всегда, когда MS переходит из одной области в другую. В случае если абонент запрашивает специальные дополнительные услуги или изменяет некоторые свои данные, MSC также информирует VLR, который регистрирует изменения и при необходимости сообщает о них HLR.

C-интерфейс. Интерфейс используется для обеспечения взаимодействия между MSC и HLR. MSC может послать сообщение HLR в конце сеанса связи для того, чтобы абонент мог оплатить разговор. Когда сеть фиксированной телефонной связи не способна выполнить процедуру установления соединения подвижного абонента, MSC может запросить HLR с целью определения местоположения абонента, чтобы послать вызов MS.

D-интерфейс. Интерфейс между HLR и VLR используется для расширения обмена данными о положении подвижной станции и управления процессом связи. Основные услуги, предоставляемые подвижному абоненту, заключаются в возможности передавать или принимать сообщения независимо от местоположения. Для этого HLR должен пополнять свои данные. VLR сообщает HLR о положении MS, управляя ею и изменяя информацию в процессе обновления местоположения, а также посылает все необходимые данные для обеспечения обслуживания подвижной станции.

E-интерфейс. Интерфейс обеспечивает взаимодействие между разными MSC при осуществлении процедуры handover — "передачи" абонента из зоны в зону при его движении в процессе сеанса связи без ее перерыва.

O-интерфейс. Интерфейс между BSC и OMC предназначен для связи BSC с OMC, используется в сетях с пакетной коммутацией МККТТ X.25.

M-интерфейс. Внутренний BSC-интерфейс контроллера базовой станции обеспечивает связь между различным оборудованием BSC и оборудованием транскодирования (TCE); использует стандарт ИКМ-передачи 2,048 Мбит/с и позволяет организовать из четырех каналов со скоростью 16 Кбит/с один канал на скорости 64 Кбит/с.

Um-радиоинтерфейс. Интерфейс между MS и BTS определен в сериях 04 и 08 Рекомендаций ETSI/GSM.

X-интерфейс. Сетевой интерфейс между OMC разных сетей и так называемый управляющий интерфейс между OMC и элементами сети, определен ETSI/GSM Рекомендациями 12.01 и является аналогом интерфейса Q.3, который определен в многоуровневой модели открытых сетей ISO OSI.

Соединение сети с OMC могут обеспечиваться системой сигнализации МККТТ ОКС №7 или сетевым протоколом X.25. Сеть X.25 может соединяться с объединенными сетями или с PSDN в открытом или замкнутом режимах.

GSM-протокол управления сетью и обслуживанием также должен удовлетворять требованиям Q.3 интерфейса, который определен в ETSI/GSM Рекомендациях 12.01.

1.4.2. Интерфейсы с внешними сетями

Соединение с PSTN

Соединение с телефонной сетью общего пользования осуществляется MSC по линии связи 2 Мбит/с в соответствии с системой сигнализации ОКС № 7. Электрические характеристики 2 Мбит/с интерфейса соответствуют Рекомендациям МККТТ G.732.

Соединение с ISDN

Для соединения с создаваемыми сетями ISDN предусматриваются четыре линии связи 2 Мбит/с, поддерживаемые системой сигнализации ОКС №7. Система сигнализации ОКС № 7 будет рассмотрена в дальнейшем.

Соединения с международными сетями GSM

В настоящее время обеспечивается подключение сети российской сети GSM к общеевропейским сетям GSM. Эти соединения осуществляются на основе протоколов систем сигнализации ОКС№7 четвертого уровня (SCCP — Signaling Connection Control Part) и межсетевого коммутационного центра мобильной связи (GMSC — Gateway MSC). Центр представляет узловую станцию, осуществляющую объединение сети GSM с одной или более наземными сетями. В ее функции входит преобразование форматов сигналов, конвертирование сетевых протоколов, а также взаимодействие с ТфОП.

1.4.3. Географические зоны сети GSM

Сеть GSM составлена из географических областей. Как показано на рис. 1.2, эти области включают ячейки, зоны местоположения (LA's — Location Areas), зоны обслуживания MSC/VLR и мобильную наземную сеть общего пользования (PLMN — Public Land Mobile Network).

Сота — область радиоохвата одного приемопередатчика одной BTS. Сеть GSM определяет каждую соту с помощью опознавательного кода

глобального идентификатора соты (CGI — Cell Global Identity), номера, который назначается каждой соте.

Зона местоположения (LA — Location Area) — группа сот. Это область, в которой вероятнее всего может в данный момент перемещаться абонент.

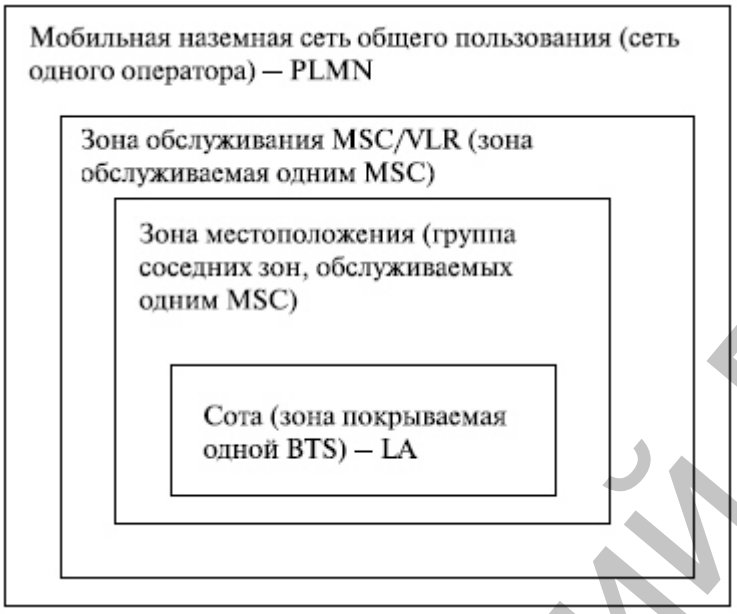


Рис. 1.2. Географические зоны системы GSM

Каждая зона местоположения обслуживается одним или более контроллерами базовых станций и только единственным центром коммутации мобильной связи — MSC (см. рис. 1.2). Каждой зоне местоположения (LA) назначен идентификатор зоны нахождения абонента (LAI — Location Area Identification).

Зона обслуживания MSC/VLR представляет собой часть сети GSM, которая обслуживается одним MSC и зарегистрирована в VLR данного MSC (рис. 1.3).

Мобильная наземная сеть общего пользования (PLMN — Public Land Mobile Network) — это совокупность зон обслуживания, принадлежащих одному сетевому оператору.

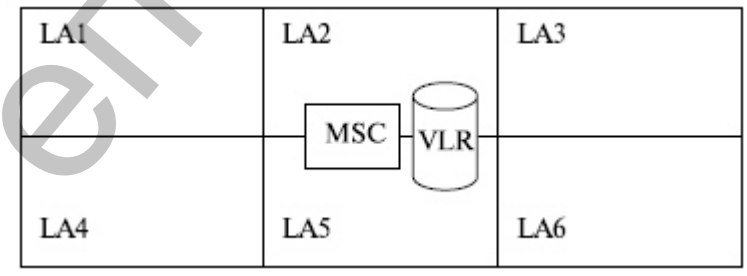


Рис. 1.3. Зона местоположения (LA)

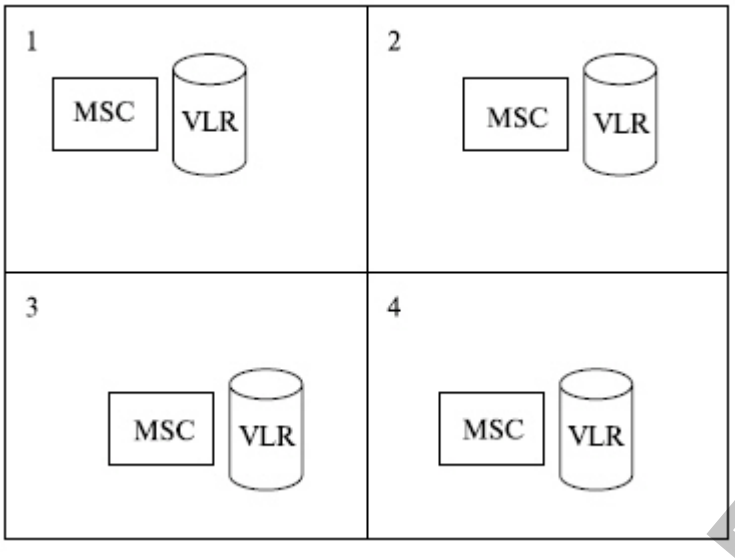


Рис. 1.4. Мобильная наземная сеть (PLMN)

1.4.4. Повторное использование частот (Frequency reuse)

Повторное использование частот — способ организации связи, при котором одни и те же частоты многократно используются в разных зонах обслуживания. Применение частотно-территориального планирования с повторным использованием частот позволяет увеличить пропускную способность при ограниченном количестве частотных каналов.

Расстояние повторного использования частот (Frequency reuses distance) — расстояние между центрами двух удаленных сот, начиная с которого допускается повторное использование. В общем случае оно определяется по формуле , где — число ячеек в кластере, — радиус ячейки (радиус окружности, описанной вокруг гексагональной ячейки).

Кластер (cluster). Кластер — это группа из близко расположенных сот, в пределах которых недопустимо повторное использование из-за опасности превышения уровня взаимных помех. Размер кластера определяется по формуле:

Из этой формулы видно, что кластер может содержать только определенное число сот.

При:

```
i=0, j=1, N=1;
  i=1, j=1, N=3;
  i=1, j=2, N=4; и т. д.
```

Приведенное соотношение для показывает, что чем меньше радиус ячейки , тем выше коэффициент повторяемости частот ( ), а следовательно, и эффективность использования выделенного диапазона частот. Отношение называется коэффициентом снижения внутриканальных помех и характеризует степень взаимного влияния удаленных сот, в которых используются одни и те же частотные каналы. Для

приведенных выше значений значение равно:

.

Пример распределений частот при повторном использовании показан на рис. 1.5 (4-элементный кластер) и рис. 1.6 (7-элементный кластер).

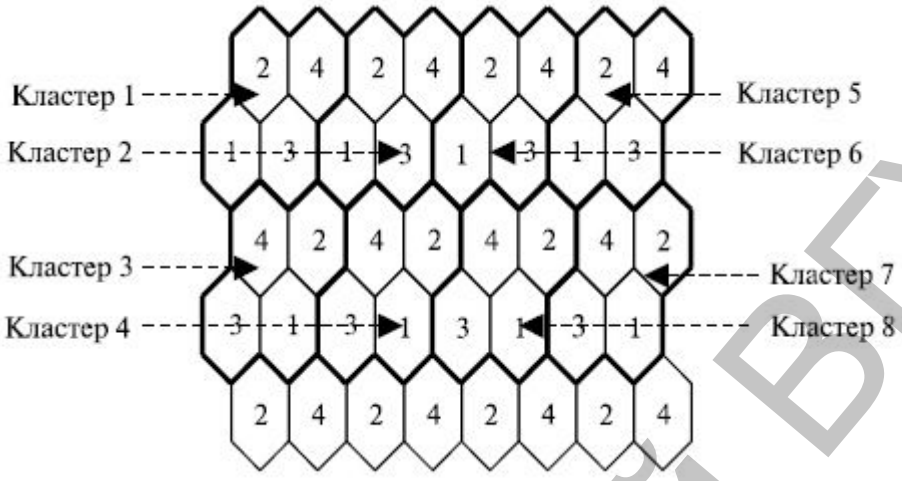


Рис. 1.5. Повторное использование частот при 4-элементном кластере

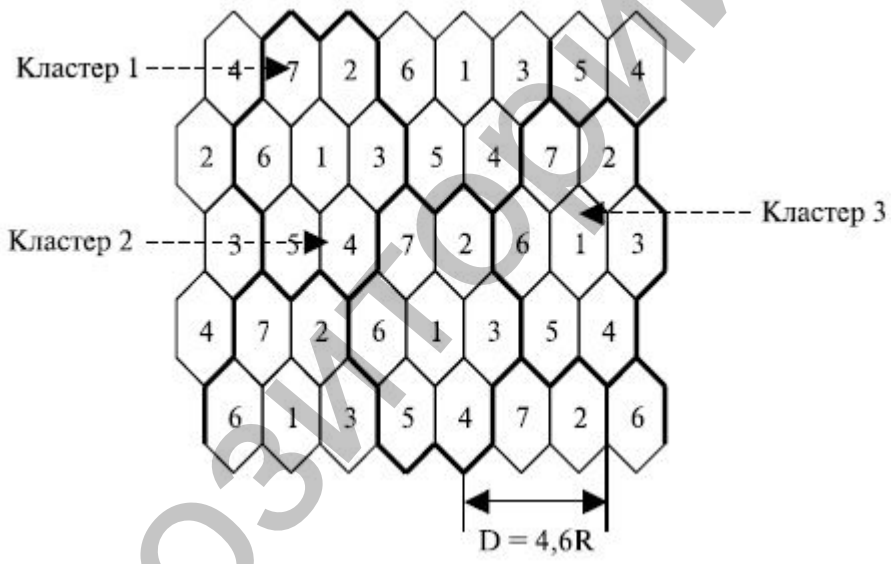


Рис. 1.6. Повторное использование частот при 7-элементном кластере

1.4.5. Секторизованная сота

Сота, в которой обслуживание абонентов осуществляется базовой станцией с секторной антенной, называется секторизованной сотой. При этом зона покрытия антенны разделяется на секторы. Секторизация позволяет повысить пропускную способность системы сотовой связи без уменьшения размеров зоны покрытия или снижения мощности, излучаемой базовой станцией. Ширина направленности секторной антенны соответствует угловому размеру сектора. В системах сотовой связи обычно используют антенны с шириной диаграммы направленности 120° (трехсекторная антенна). Обычно применяются кластеры размерностью 3/9, 4/12, 7/21, где первая цифра обозначает число сот в кластере, а вторая — число секторов. На рис. 1.7-а показан пример применения 3-секторной антенны для кластера

3/9. В этом примере распределяются 9 групп частот и применяются шестисекторные антенны — с шириной диаграммы направленности 60°. На рис. 1.7-б показана разработанная корпорацией Motorola сотовая сеть с шириной одного из вариантов диаграммы направленности 60° и 12 группами несущих частот. Этот кластер содержит 4 элемента и 6-секторную антенну (размер кластера 4/24).

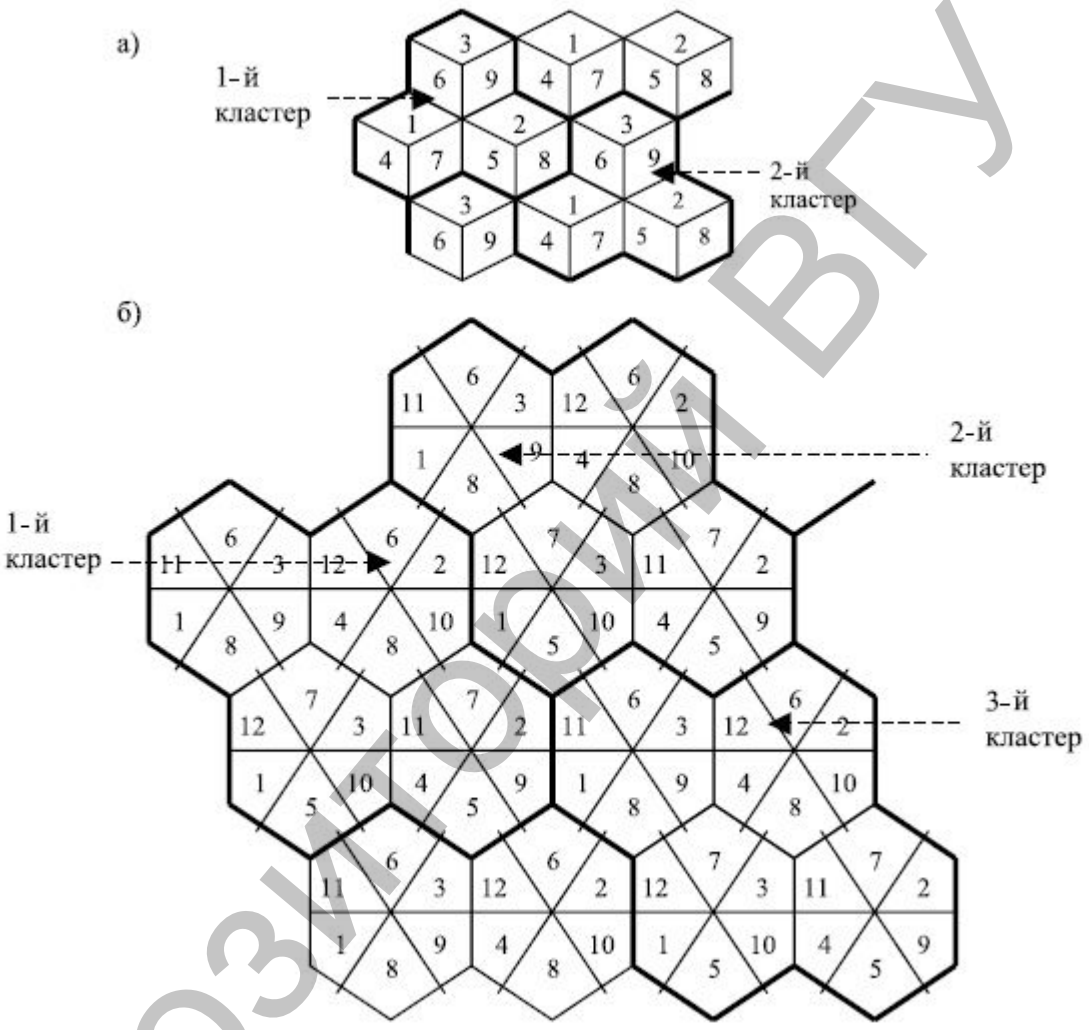


Рис. 1.7. Повторное использование частот в: а) 3-секторной соте; б) 6-секторной соте

1.4.6. Задачи каналов в системе GSM

Очевидно, что использование радиоканалов в мобильной сети GSM отличается от их применения в стационарной сети. Принцип использования каналов в системе GSM показан на рис. 1.8.

В стационарной сети абонентские линии (абонентские каналы трафика) закреплены за телефонным аппаратом. Когда известен номер абонента, то при исходящей или входящей связи не требуется выбор абонентской линии.

В сети GSM определены два типа каналов трафика: полноскоростные речевые каналы, работающие на полной скорости (TCH/F — Traffic Channel/Full) — 22,8 Кбит/с, и полускоростные речевые каналы, работающие на половинной скорости (TCH/H — Traffic Channel/Half) — 11,4 Кбит/с.

Половинная скорость позволяет вдвое увеличить число каналов в одном и том же частотном диапазоне.

В мобильной связи каналы трафика доступны любому абоненту. Поэтому в процессе установления соединения может быть выбран любой канал, к которому может быть подключена станция. Поскольку в свободном состоянии абонентская линия не имеет связи с каналами трафика, она нуждается в канале управления, например, для передачи сигнала "вызов", "setup", номера вызывающего абонента и т. п.

Поэтому для передачи запроса сети на установление соединения применяется канал, направленный от MS к сети. Это канал случайного доступа (RACH — Random Control Channel).

Поскольку запрос на установление соединения передается только в начале соединения и в дальнейшем выделяется канал для обмена управляющей информацией, этот канал является общим для всех станций зоны местонахождения.

Общему каналу всегда требуется процедура доступа для избежания и разрешения конфликтов. В данном случае чаще всего применяется процедура случайного многостанционного доступа с временным разделением типа ALOHA (TDMA — Time Division Multiple Access ALOHA). Принцип такого доступа основан на том, что все станции используют один канал связи, контролируя его работу, а передача осуществляется в случайные моменты времени, что уменьшает вероятность конфликтов. Такой доступ подробно описан в курсе лекций "Абонентские оконечные устройства и сети доступа".

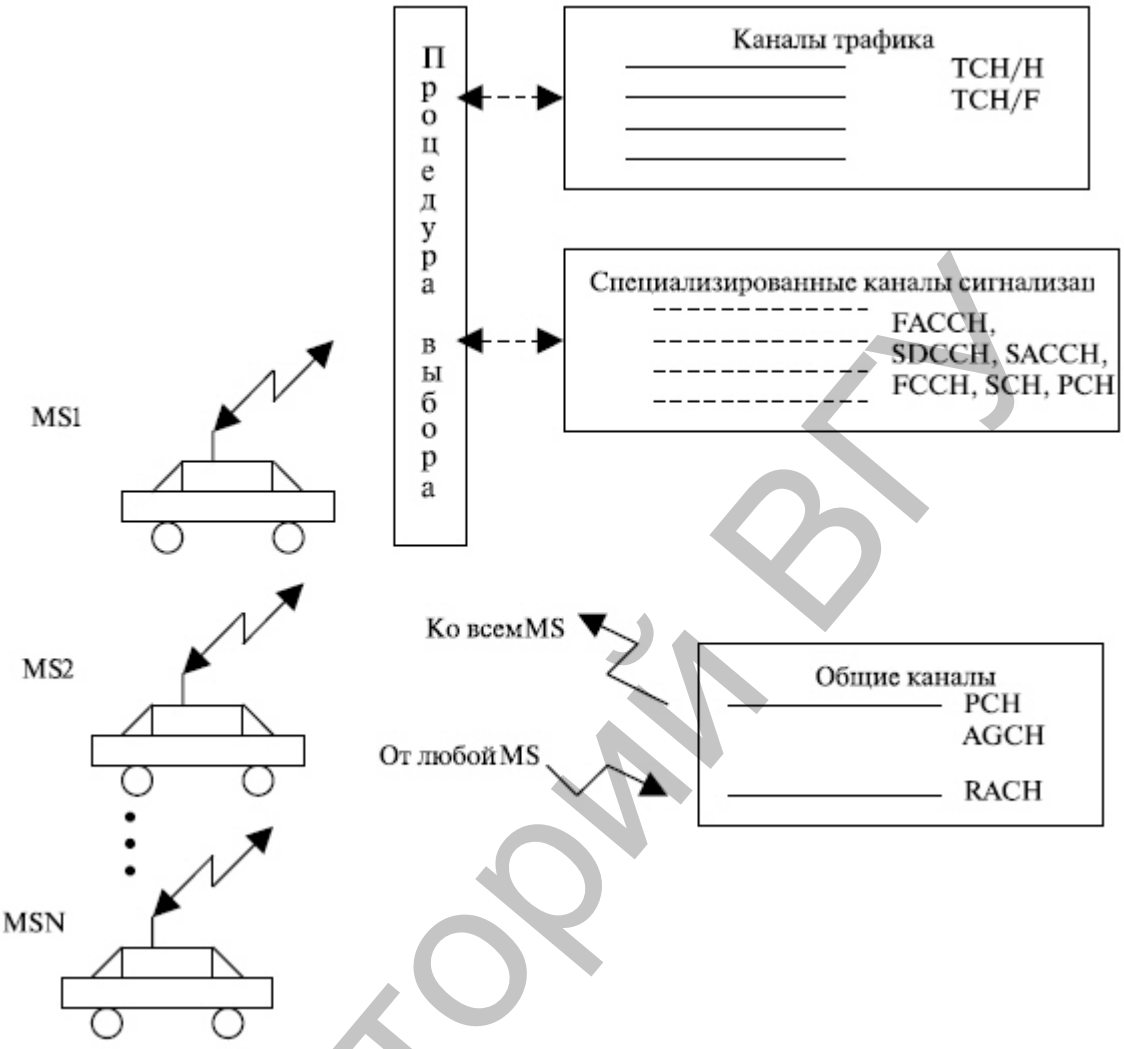


Рис. 1.8. Принцип использования каналов трафика и сигнальных каналов в системе GSM

В ответ на сигнал вызова выбирается автономный специализированный канал управления (SDCCH — Stand-alone dedicated Control Channel), по которому в дальнейшем передается служебная информация от MS в течение установления вызова прежде, чем будет найден канал трафика (TCH).

Для входящей связи передача сигнала "занятие" к MS реализуется по широковещательному каналу коротких сообщений (канал вызова) (PCH — Paging Channel), общему для всей соты. Это широковещательный канал коротких сообщений, который передает сигнал "вызов" всем станциям зоны местоположения (LA). Получив такой сигнал, станция MS определяет свой номер и отвечает на широковещательный сигнал так же, как при исходящем вызове, — сигналом запроса по каналу случайного доступа (RACH — Random Control Channel).

Далее сигналы установления соединения проходят как и при исходящей связи.

Порядок обмена сигналами для входящего и исходящего соединения приведен на рис. 1.9.

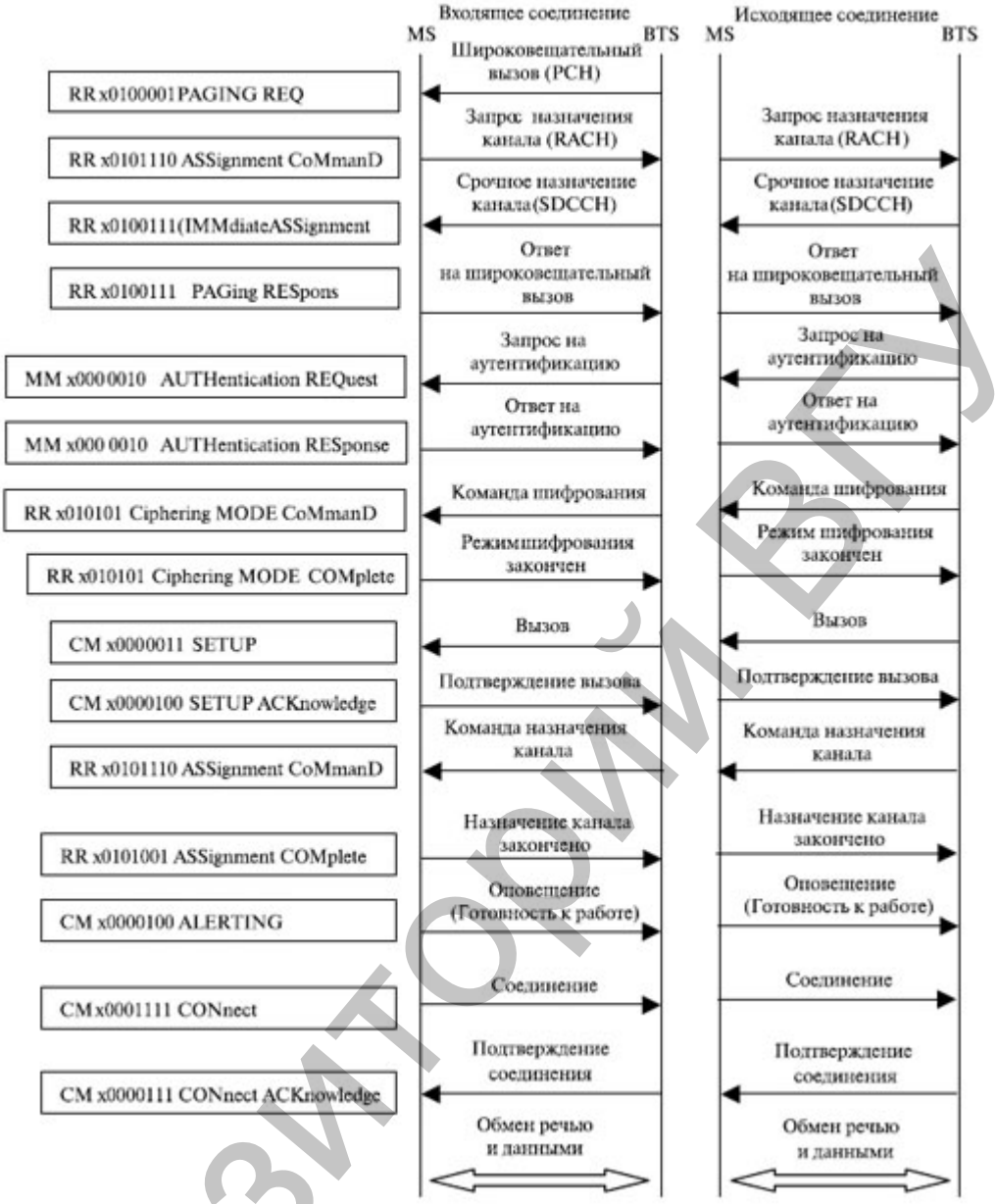


Рис. 1.9. Порядок обмена сигналами для входящего и исходящего соединения

На рисунке показаны некоторые особенности передачи сигналов. Ниже даны некоторые пояснения. Нарисованные слева коды сигналов будут рассмотрены далее.

При входящей связи BTS и MS пункта назначения (работа других элементов сети на данном рисунке не изображается):

1. передает широковещательный сигнал всем станциям в зоне обслуживания данного MSC. Сигнал передается по отдельному каналу управления - широковещательному каналу коротких сообщений — PCH (Paging Channel);

2. после чего MS по каналу управления (канал со случайным доступом — RACH, Random Access Channel) посылает запрос на срочное назначение индивидуального канала управления на время обмена сигналами.

Слова "произвольный доступ" означают применение методов случайного доступа, наиболее распространенным из которых является ALOHA (см. курс лекций "Абонентские оконечные устройства и сети доступа"). Принцип работы при таком методе заключается в том, что все станции работают по одному каналу связи, контролируя его работу, передача осуществляется в случайный момент времени. BTS выбирает канал для обмена управляющими сигналами (SDCCH — Stand-alone dedicated Control Channel);

3. BTS запрашивает данные аутентификации. Проводится аутентификация с помощью данных, полученных ранее при реализации процедуры аутентификации и защиты пользователя. В ответ на запрос MS передает накопленный в SIM-карте зашифрованный отклик (SRES — Signed Response), что позволяет BTS установить подлинность MS;

4. после чего BTS передает запрос ключа шифрования;

5. и получает ответный ключ шифрования. Если ключ правильный, то далее проводится процедура установления соединения, которая совпадает с процедурой исходящего соединения.

Теперь можно рассмотреть подробнее весь состав сигнальных каналов.

1.4.7. Каналы сигнализации радиоинтерфейса

Сигнальные каналы радиоинтерфейса используются для установления вызова, широковещательной рассылки коротких сообщений (paging), технического обслуживания вызова, синхронизации и т. д. (рис. 1.10).

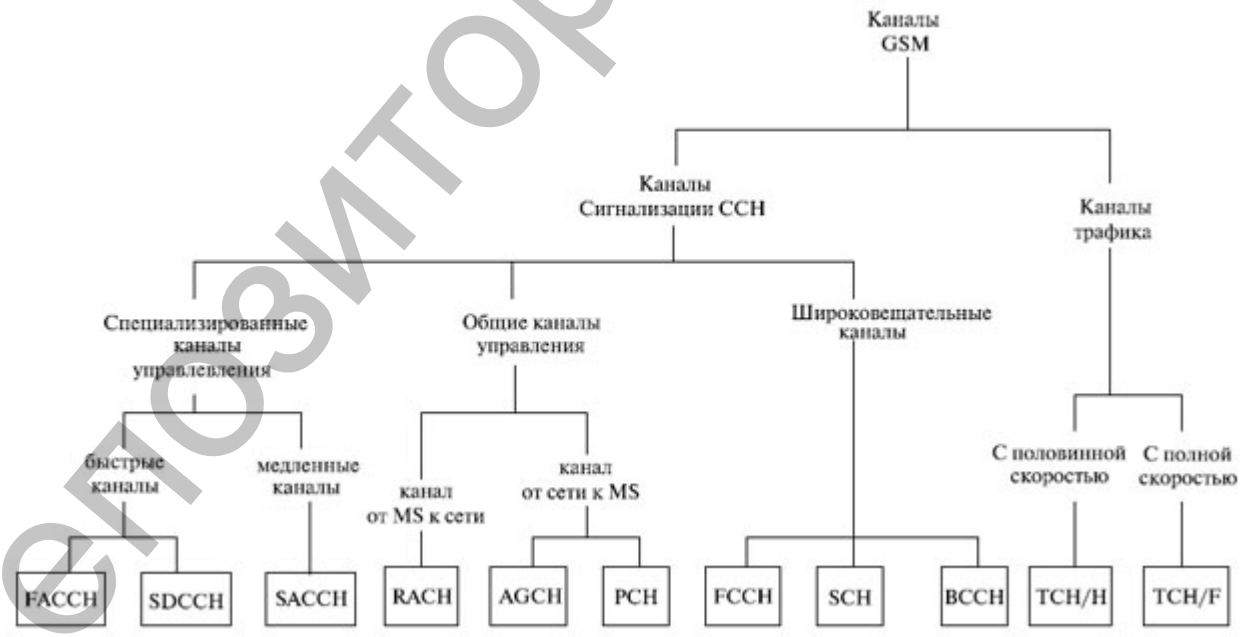


Рис. 1.10. Состав каналов радио интерфейса системы GSM

Имеется 3 группы сигнальных каналов.

Широковещательные каналы (BCH — Broadcast Channel). Доставляют информацию от станции к абоненту (downstream) и предназначены главным образом для коррекции частоты и синхронизации. Это единственный тип канала, допускающий связь "от точки — ко многим

точкам", при которой короткие сообщения могут быть переданы одновременно нескольким мобильным телефонам.

BCH включают следующие каналы:

- широковещательный канал управления (BCCH — Broadcast Control Channel). Общая информация, касающаяся сот; например, код зоны местоположения (LAC — Location Area Code), сетевой оператор, доступ, параметры, список соседних ячеек и т. д. MS получают сигналы через BCCH от многих BTS в пределах той же самой сети или различных сетей;
- канал подстройки частоты (FCCH — Frequency Correction Channel). Канал связи от сети к MS, предназначенный только для коррекции частот MS и передачи частоты к MS. Он также используется для вхождения в синхронизм, обеспечивая соблюдение заданной дистанции между временными интервалами и позицией первого временного интервала кадра TDMA (множественного доступа с временным уплотнением);
- канал синхронизации (SCH — Synchronizing Channel). Исходящий канал от MS к сети; отвечает за синхронизацию кадра TDMA и идентификацию базовой станции. SCH обеспечивает MS всей информацией, необходимой для синхронизации с BTS.

Общие каналы управления (CCCH — Common Control Channels): группа канала связи от абонента к станции и каналы связи от сети к MS. Эти каналы используются, чтобы передать информацию между сетью и MS. Общие каналы управления CCCH включают следующие каналы:

- широковещательный канал коротких сообщений (канал вызова) (PCH — Paging Channel): исходящий канал только от сети к MS; BTS информирует MS о входящих вызовах через PCH;
- канал предоставления доступа (AGCH): исходящий канал только от сети к MS. BTS распределяет TCH или SDCCH к MS, таким образом разрешая MS доступ к сети;
- канал с произвольным доступом (RACH): канал связи только от MS к сети; позволяет MS запрашивать SDCCH. Это делается в ответ на широковещательный запрос или на вызов. MS для передачи на этом канале работает по принципу случайного доступа.

PCH и AGCH передают информацию в одном канале, называемую широковещательным сообщением, и каналом предоставления доступа, как это будет показано далее.

Специализированные каналы управления (DCCH — Dedicated Control Channel). Предназначены, например, для обслуживания: роуминга, изменения местоположения, передачи соединения (хэндовер), шифрования и т. д.

DCCH включают следующие каналы:

- автономный выделенный канал управления (SDCCH — Stand-Alone Dedicated Control Channel): канал, соединяющий MS и BTS, для передачи сигналов в течение установления вызова прежде, чем будет найден канал трафика (TCH);

- низкоскоростной совмещенный канал управления (SACCH — Slow Associated Control Channel): передает непрерывные сообщения об измерениях (например, напряженность поля). Параллельно с ним могут работать TCH или SDCCH необходим, например, для решений хэндовера; применяется подобно TCH или SDCCH для несрочных процедур, например, для измерения радиосигналов, управления мощностью (только исходящий канал от сети к MS);
- быстродействующий объединенный канал управления (FACCH — fast associated control channel): его функции сродни SDCCH, но он может использоваться временно для работы как TCH в режиме перераспределения каналов (borrowing mode) совместно с SDCCH, если скорость данных SDCCH (низкоскоростного выделенного канала управления) недостаточна.

Дополнительная пропускная способность применяется, например, для процедур, связанных с установлением подлинности (аутентификацией), установлением соединения, хэндовером и т. д.

Почти все сигнальные каналы используют формат нормального пакета, кроме RACH (пакет произвольного доступа), FCCH (пакет коррекции частоты) и SCH (пакет синхронизации).

1.4.8. Некоторые примеры работы сети GSM

Обслуживание вызова от абонента стационарной сети к абоненту мобильной сети GSM

Следующий пример описывает обслуживание вызова от абонента стационарной сети к абоненту мобильной сети GSM (рис. 1.11):

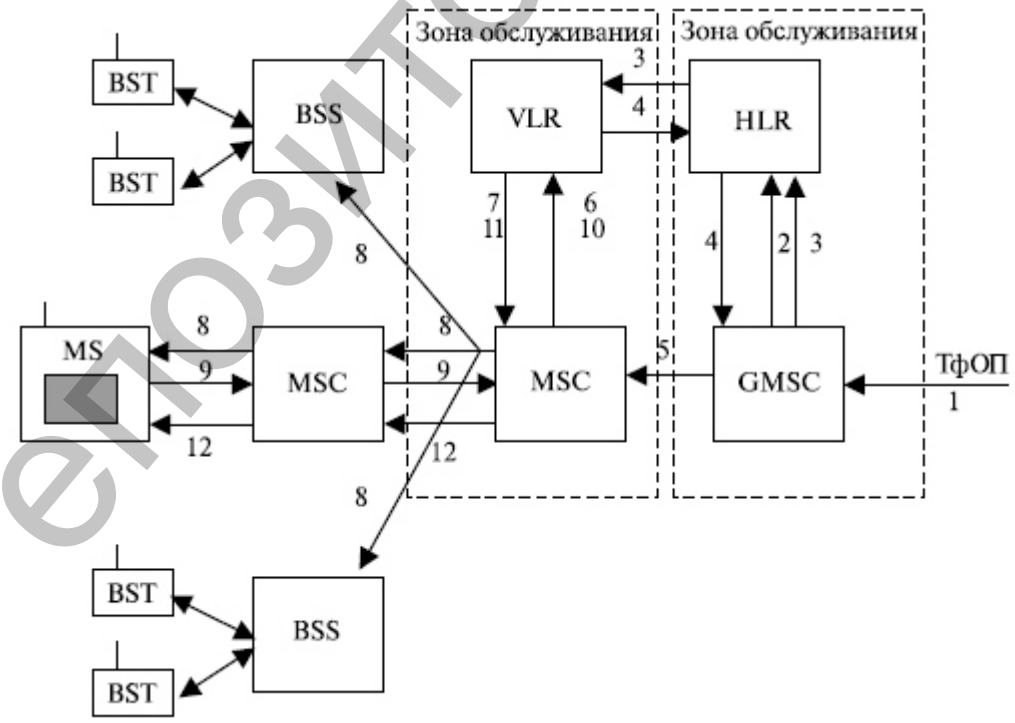


Рис. 1.11. Обслуживание вызова от абонента стационарной сети к абоненту мобильной сети GSM

В рассматриваемом примере порядок действий следующий:

1. Входящий вызов поступает от стационарной сети ТфОП на вход шлюза MSC (GMSC — Gateway MSC).

2. На основе международного мобильного идентификационного номера станции (IMSI — International Mobile Station Identity) вызываемого абонента определяется домашний регистр местоположения (HLR).

3. Затем запрашивают соответствующий визитный регистр местоположения (VLR) для того, чтобы определить для мобильной станции номер для услуг роуминга (рис. 1.12) — MSRN (Mobile Station Roaming Number).

4. Он передается назад в HLR GMSC.

5. Затем соединение переключается к соответствующему MSC.

6. MSC вырабатывает запрос VLR.

7. Теперь визитный регистр местоположения (VLR) делает запрос зоны местоположения (LA — Location Area) и о состоянии (доступности) мобильного абонента. Если MS отмечена как доступная, то выполняется п. 8.

8. Передается широковещательный вызов по всей зоне нахождения, записанной в визитном регистре местоположения (VLR).

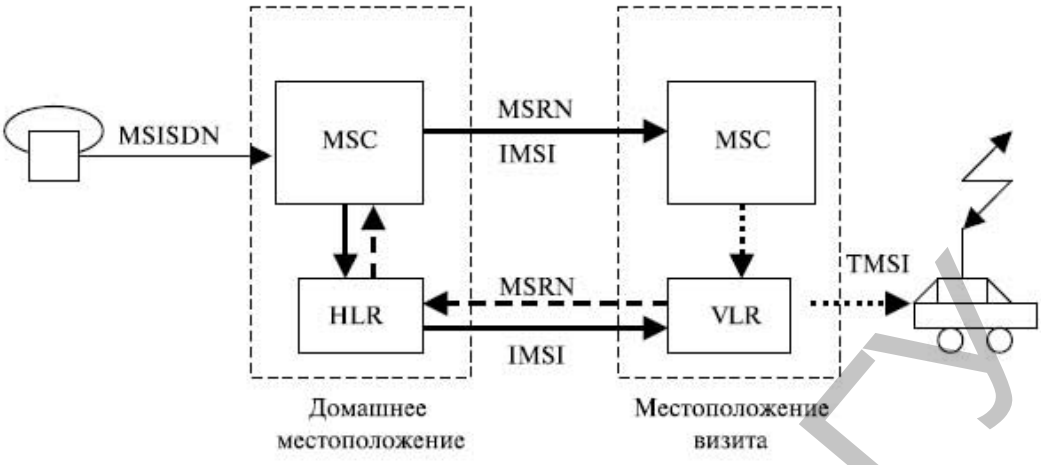
9. Мобильный абонентский телефон отвечает на широковещательный запрос из текущей радиосоты.

10. После этого выполняются все необходимые процедуры безопасности (аутентификация и обмен шифровальными ключами). Если они выполнены успешно, то выполняется п. 11.

11. Визитный регистр местоположения (VLR) указывает для MSC, что вызов закончен, и передает MSC временный мобильный опознавательный код станции (TMSI — Temporary Mobile Station Identity).

12. MSC передает MS TMSI и информирует его о начале работы.

На рис. 1.12 отдельно отображен процесс изменения номеров в процессе установления входящего вызова.



MSISDN	Mobile Station international ISDN Number	Международный ISDN номер мобильной станции
MSRN	Mobile Station Roaming Number	Временный роуминговый номер мобильной станции
IMSI	International Mobile Station Identity	Международный Мобильный мобильной Станции
TMSI	Temporary Mobile Station Identity	Временный Мобильный Опознавательный код мобильной станции
HLR	Home Location Register	Домашний регистр местоположения
VLR	Visit Location Register	Визитный регистр местоположения

Рис. 1.12. Принцип изменения номера

Регистрация в сети

При каждом включении телефона после выбора сети начинается процедура регистрации (рис. 1.13). Рассмотрим наиболее общий случай — регистрацию не в домашней, а в чужой, так называемой гостевой, сети (будем предполагать, что услуга роуминга абоненту разрешена).

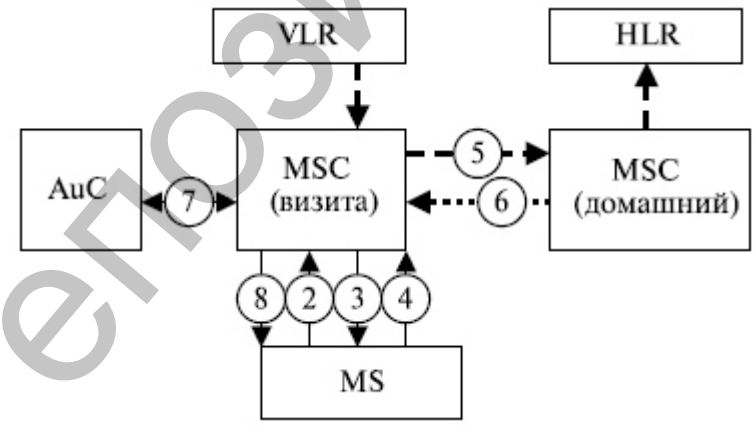


Рис. 1.13. Процесс регистрации

1. MS по широковещательному каналу управления (BCCH) проводит сканирование свыше 16 соседних сот, и формируется список шести лучших кандидатов на возможную передачу соединения, основанную на полученной напряженности поля сигналов.

2. MS находит канал BCCH с наиболее высоким уровнем сигнала, проводит синхронизацию, расшифровывает идентификатор BTS и передает эту информацию к BSC и MSC.

3. По запросу MSC производит запрос MS с номером IMSI.

4. MS передает IMSI абонента. IMSI начинается с кода страны "приписки" его владельца, далее следуют цифры, определяющие домашнюю сеть, а уже потом — уникальный номер конкретного подписчика. Начало IMSI соответствует коду страны и оператору (например, 250 — Россия, 99 — Билайн).

5. По номеру IMSI VLR гостевой сети определяет домашнюю сеть и запрашивает ее HLR.

6. Домашний регистр мобильного центра коммутации (MSC/HLR) передает всю необходимую информацию об абоненте в VLR, который сделал запрос, а у себя размещает ссылку на этот VLR, чтобы в случае необходимости знать, где "искать" абонента.

7. MSC совместно с VLR проводит проверку полномочий.

8. В положительном случае MSC включает MS в обслуживание.

После процедуры идентификации и взаимодействия гостевого VLR с домашним HLR запускается счетчик времени, задающий момент перерегистрации в случае отсутствия каких-либо сеансов связи. Обычно период обязательной регистрации составляет несколько часов. Перерегистрация необходима для того, чтобы сеть получила подтверждение, что телефон по-прежнему находится в зоне ее действия. Дело в том, что в режиме ожидания "трубка" только отслеживает сигналы, передаваемые сетью, но сама ничего не излучает. Процесс передачи начинается только в случае установления соединения, а также при значительных перемещениях относительно сети (ниже это будет рассмотрено подробно), — в таких случаях таймер, отсчитывающий время до следующей перерегистрации, запускается заново. Поэтому при "выпадении" телефона из сети (например, был отсоединен аккумулятор, или владелец аппарата зашел в метро, не выключив телефон) система об этом не узнает.

При первой установке абонента в сети выполняется операция закрепления международного идентификационного номера мобильной станции (IMSI — International Mobile Station Identity). Обратная закреплению процедура — открепление — позволяет сети знать, что передвижная станция недостижима, и устраняет необходимость напрасно распределять каналы и передавать широковещательные сообщения. Процедура закрепления похожа на обновление местоположения и сообщает, что мобильная станция доступна снова.

Все пользователи случайным образом разбиваются на 10 равноправных классов доступа (с номерами от 0 до 9). Абоненту присваивается класс доступа. Существует несколько специальных классов с номерами с 11 по 15 (разного рода аварийные и экстренные службы, служебный персонал сети). Информация о классе доступа хранится в SIM-карте. Особый, 10-й класс

доступа позволяет совершать экстренные звонки (по номеру 112), если пользователь не принадлежит к какому-либо разрешенному классу или вообще не имеет IMSI (SIM). В случае чрезвычайных ситуаций или перегрузки сети некоторым классам может быть на время закрыт доступ в сеть.

Активация IMSI и закрепление/открепление осуществляется оператором на базе определенной соты.

Обновление местоположения

При подвижной связи в случае включенной мобильной станции осуществляется постоянное слежение за местоположением даже в случае отсутствия соединения. В частности, это необходимо для установления входящей связи. Включенная мобильная станция информируется о входящем вызове широковещательным сообщением, передаваемым по широковещательному каналу коротких сообщений (PCH — Paging Channel).

Один из вариантов определения местоположения — периодически сообщать о расположении объектов в каждой соте. При этом, если объект редко меняет свое местоположение (соту), такая процедура лишь понапрасну расходовала бы пропускную способность радиосети. Другой крайний случай — уведомлять систему при изменении местоположения мобильной станции широковещательным сообщением. но и это очень расточительно из-за большого количества мобильных станций, обновляющих свое местоположение. Компромиссное решение, используемое в GSM, — оповещение о местоположении при смене группы сот в зоне местоположения, приводящей к ухудшению связи. Обновляющие сообщения требуются при перемещении между областями местоположения, и передвижные станции просматриваются в сотах их текущей области.

Процедуры обновления местоположения и соответствующая последующая маршрутизация используют центр коммутации мобильной связи (MSC) и два регистра местоположения: домашний регистр местоположения (HLR) и визитный регистр местоположения (VLR). Когда передвижная станция:

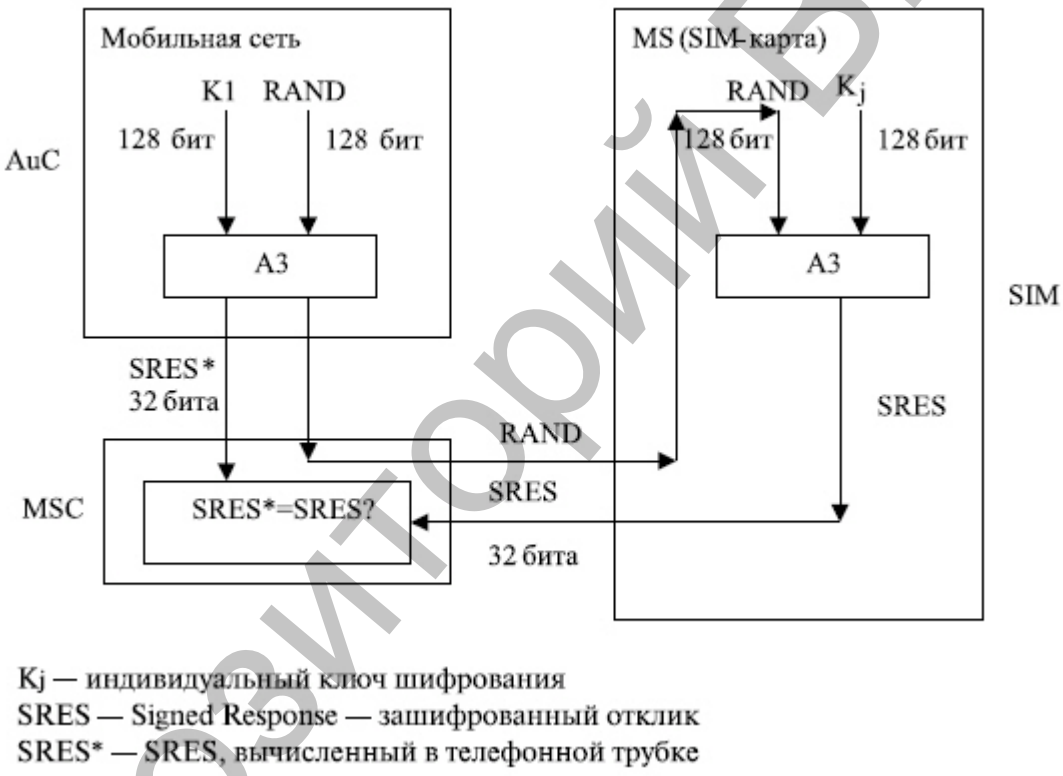
- переключается к другой BTS и BSC в области местоположения,
- перемещается в новую область местоположения,
- перемещается к другому оператору общедоступной телефонной сети (для наземных объектов) (PLMN — Public Land Mobile Telephone Network)

Тогда это перемещение должно регистрироваться сетью, чтобы отметить текущее местоположение. В нормальном случае сообщение обновления местоположения передают новому центру коммутации мобильной связи — MSC (визитному регистру местоположения VLR), который записывает информацию в области памяти местоположения и затем передает ее домашнему регистру местоположения — HLR абонента. Информация, передаваемая HLR, — обычно через ОКС № 7, — это адрес нового VLR, хотя это может быть номер направления. Если абонент имеет

право на обслуживание в новой области местоположения, HLR передает набор абонентской информации, необходимой для управления вызовом, новому центру коммутации мобильной связи (MSC/VLR) и посылает сообщение старому MSC/VLR об отмене старой регистрации.

Аутентификация и защита

Так как к радиосреде имеют доступ много устройств и абонентов, требуется аутентифицировать пользователей. Эта процедура (рис. 1.14) устанавливает подлинность и принадлежность к сети абонента и оборудования, определяет права и полномочия абонента и право доступа к сетевым ресурсам. Аутентификация проводится с помощью двух функциональных объектов: SIM-карты в мобильной станции и центра аутентификации (AuC — Authentication Center).



Kj — индивидуальный ключ шифрования
SRES — Signed Response — зашифрованный отклик
SRES* — SRES, вычисленный в телефонной трубке

Рис. 1.14. Обеспечение аутентификации абонента и защиты информации

При регистрации AuC в домашней сети генерирует 128-битовое случайное число — RAND, пересылаемое телефону. Внутри SIM с помощью ключа Ki (ключ идентификации — так же, как и IMSI, он содержится в SIM) и алгоритма идентификации A3 вычисляется 32-битовый ответ — SRES (Signed respons) по формуле . Точно такие же вычисления проделываются одновременно и в AUC (по выбранному из HLR Ki пользователя). Если SRES, вычисленный в телефоне, совпадет со SRES, рассчитанным AuC, то процесс авторизации считается успешным и абоненту присваивается TMSI (Temporary Mobile Subscriber Identity — временный номер мобильного абонента). TMSI служит исключительно для повышения безопасности взаимодействия подписчика с сетью и может периодически меняться (в том числе при смене VLR).

То же самое случайное начальное число и абонентский ключ засекречивания также используются, чтобы вычислить ключ шифрования, который применяет алгоритм шифрования речи. Этот ключ шифрования, вместе с номером кадра TDMA, алгоритму нужен, чтобы создать последовательность на 114 битов, применяя операцию "исключающее ИЛИ" (XOR) с 114 битами пакета (два блока на 57 битов).

Другой уровень защиты выполняется в MS непосредственно для защиты оборудования от несанкционированного использования. Как упомянуто ранее, каждый терминал GSM идентифицирован уникальным международным опознавательным кодом мобильного оборудования (IMEI — Mobile Equipment Identity). Список IMEI в сети сохраняется в регистре идентификации оборудования (EIR — Equipment Identity Register), и в ответ на запрос IMEI к EIR ему возвращается одно из следующих состояний, в соответствии с тем, в каком списке находится номер абонента:

- белый список — терминалу позволяют соединиться с сетью;
- серый список — терминал находится под наблюдением сети ввиду возможных проблем;
- черный список — терминал заявлен как украденный или некорректный тип для сети GSM. Терминалу не позволяют соединиться с сетью.

Более детально вопросы безопасности сетей связи рассмотрены в курсе лекций "Криптография и безопасность сетей"

Передача соединения (хэндовер)

В сотовой сети радиоресурсы и фиксированные линии связи в течение вызова не остаются занятыми постоянно. Хэндовер (передача соединения), или хэндофф (handoff), как его называют в Северной Америке, — это переключение каналов и линий по мере перемещения подвижного объекта по различным каналам или ячейкам сотовой сети. Обнаружение и измерение уровня радиосигналов для хэндовера составляют одну из основных функций уровня RRM (Radio Resources Management).

Хэндоверы принято разделять на четыре типа, указанных цифрами на рис. 1.15:

1. Смена каналов в пределах одной базовой станции.
2. Смена канала одной базовой станции на канал другой станции, но находящейся под управлением того же BSC.
3. Переключение каналов между базовыми станциями, контролируемыми разными BSC, но одним MSC.
4. Переключение каналов между базовыми станциями, за которые отвечают не только разные BSC, но и разные MSC.

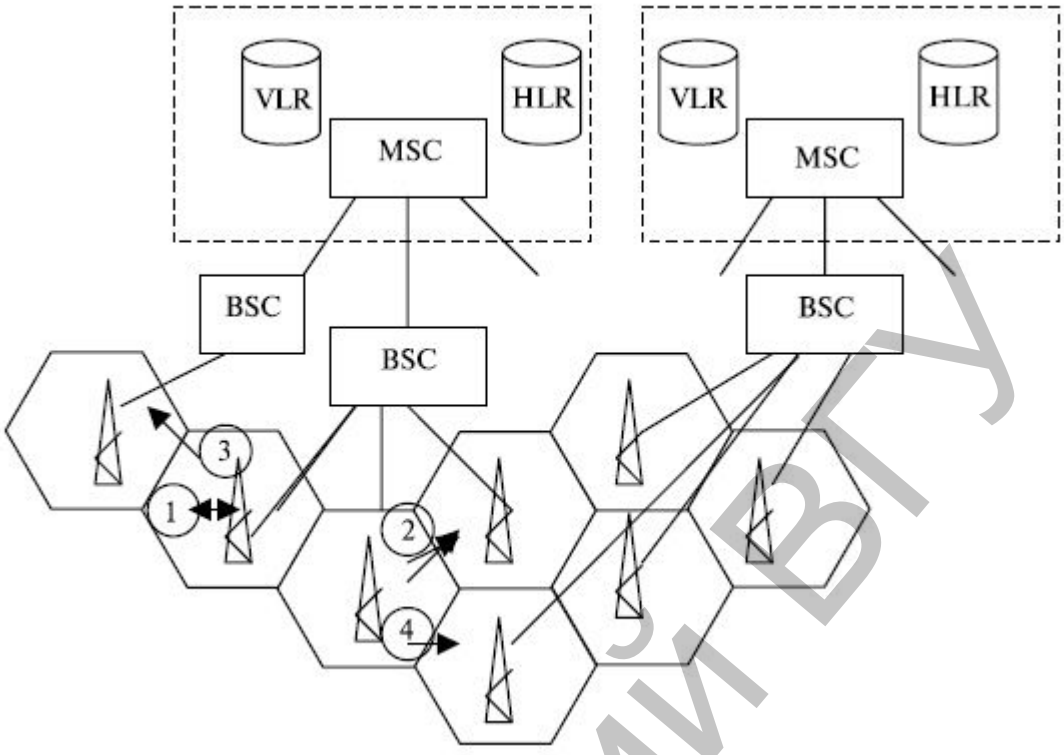


Рис. 1.15. Варианты хэндовера

В общем случае проведение хэндовера — задача MSC. Но в двух первых случаях, называемых внутренними хэндоверами, чтобы снизить нагрузку на коммутатор и служебные линии связи, процесс смены каналов управляется BSC, а MSC лишь информируется о происшедшем.

Первые два типа передачи соединения называются внутренними передачами соединения и включают только один контроллер базовой станции (BSC). Чтобы сохранять способность обмена сигналами, достаточно взаимодействия базовых станций (BSC), без использования управления центра коммутации мобильной связи (MSC). После окончания передачи соединения (хэндовера) необходимо уведомить об этом событии коммутации мобильной связи (MSC).

Последние два типа передачи соединения называются внешними передачами соединения и обрабатываются центрами коммутации мобильной связи (MSC), участвующими в соединении. Важный аспект — то, что первоначальный MSC (anchor MSC — анкерный центр), который обеспечивает доступ к сети, остается ответственным за большинство переключений.

Передачи соединения (хэндовер) могут быть инициализированы или мобильной станцией, или центром коммутации мобильной связи (MSC). MS по широковещательному каналу управления (BCCH) проводит сканирование не менее 16 соседних сот, и формируется список шести лучших кандидатов на возможную передачу соединения, основанную на полученной напряженности поля сигналов. Эта информация передается к BSC и MSC не

менее одного раза в секунду для использования алгоритмом передачи соединения (хэндовера).

Алгоритм момента времени, когда должно быть принято решение передачи соединения (хэндовер), не определен в рекомендациях GSM. Есть два основных используемых алгоритма, оба тесно связаны с управлением мощностью. Это объясняется тем, что базовая станция (BSC) обычно не знает, является ли плохое качество сигнала следствием замирания из-за многолучевости или следствием перемещения мобильной станции к другой ячейке. Особенно часто это происходит при маленьких городских ячейках.

Алгоритм "минимально допустимая характеристика" дает приоритет управлению мощностью, а не передаче соединения (хэндоверу). Когда сигнал ухудшился до некоторой заданной точки, уровень мощности мобильной станции увеличивается с помощью управления. Если дальнейшее увеличение мощности не улучшает сигнал, то начинают передачу соединения (хэндовер). Это наиболее простой и наиболее общий метод, но он создает эффект "расплывчатой границы" соты, когда мобильная станция передает сигналы, используя пиковую мощность, проходя некоторое расстояние вне границы ячейки исходной соты в другую соту.

"Метод бюджета мощности" предоставляет приоритет передаче соединения (хэндоверу). Целью является поддержание или улучшение качества сигнала при том же самом или более низком уровне мощности. В этом случае отсутствует проблема "расплывчатой границы" соты и уменьшаются межканальные помехи, но весьма усложняется алгоритм.

Рассмотрим процесс обмена сигналами, показанный на рис. 1.16, как хэндовер 4-го типа (см.рис. 1.15. Ниже приводится его описание.

1. Когда MS включена, она периодически извещает о качестве сигналов BTS1 с помощью сообщения об измерении. Эти сообщения передаются в каждом SACHH (низкоскоростной выделенный канал управления) с периодичностью 480 мсек. Сообщение об измерении содержит измерения качества сигналов соседних ячеек.
2. Если качество сигнала хорошее, то MS не предпринимает никаких действий. Когда MS достигает границы между зонами обслуживания MSC2 и MSC1, она извещает BTS1, что получает слабый сигнал.
3. BTS1 принимает решение об инициализации процесса хэндовер для того, чтобы улучшить качество обслуживания MS, и передает результаты измерений BSC1, включая измерения качества сигналов соседних ячеек BSC1.
4. BSC1 проводит анализ результатов измерения, чтобы определить зону обслуживания с лучшим качеством.
5. Если BSC1 решает запросить хэндовер, то он передает MSC1 номер используемой соты и список целевых сот с лучшими показателями, чем у используемой соты. При этом станция BTS2 включена в список целевых сот. На BSC1 включается таймер, чтобы ограничить время ожидания

начала хэндовера (поступления сигнала от MSC1 о начале процесса хэндовера).

6. MSC1 передает запрос на хэндовер к MSC2. При этом из регистра MSC1 (это может быть VLR или HLR) передаются данные для маршрутизации и аутентификации. На MSC1 включается таймер, чтобы ограничить время ожидания начала хэндовера в зоне обслуживания MSC2 (время ответа от BCS2).

7. На MSC2 запрос на передачу соединения обрабатывается как новый исходящий вызов и выбирается канал для нового вызова. Новые данные записываются в VLR MSC2. VLR MSC2 обеспечивает присвоение номера "блуждающей" подвижной станции (MSRN — Mobile Station Roaming Number). Процедурами установления подлинности во время обработки вызова управляет VLR MSC2.

8. Передача подтверждения запроса хэндовера от MSC2 (начало хэндовера) к MSC1. На MSC1 отключается таймер, ограничивающий время ожидания начала хэндовера (см. п. 7), так как получена команда о начале хэндовера. Если MSC1 был центром визита, то данные на VLR MSC1 стираются. Если он был домашним центром, то текущий адрес VLR абонента, содержащийся в HLR, также обновляется.

9. MSC1 передает сообщение BSC1, что соединение закончено.

10. BSC1 освобождает канал и информирует MSC1, что разъединение закончено.

11. MSC1 освобождает оборудование и передает MSC2 сигнал окончания процедуры.

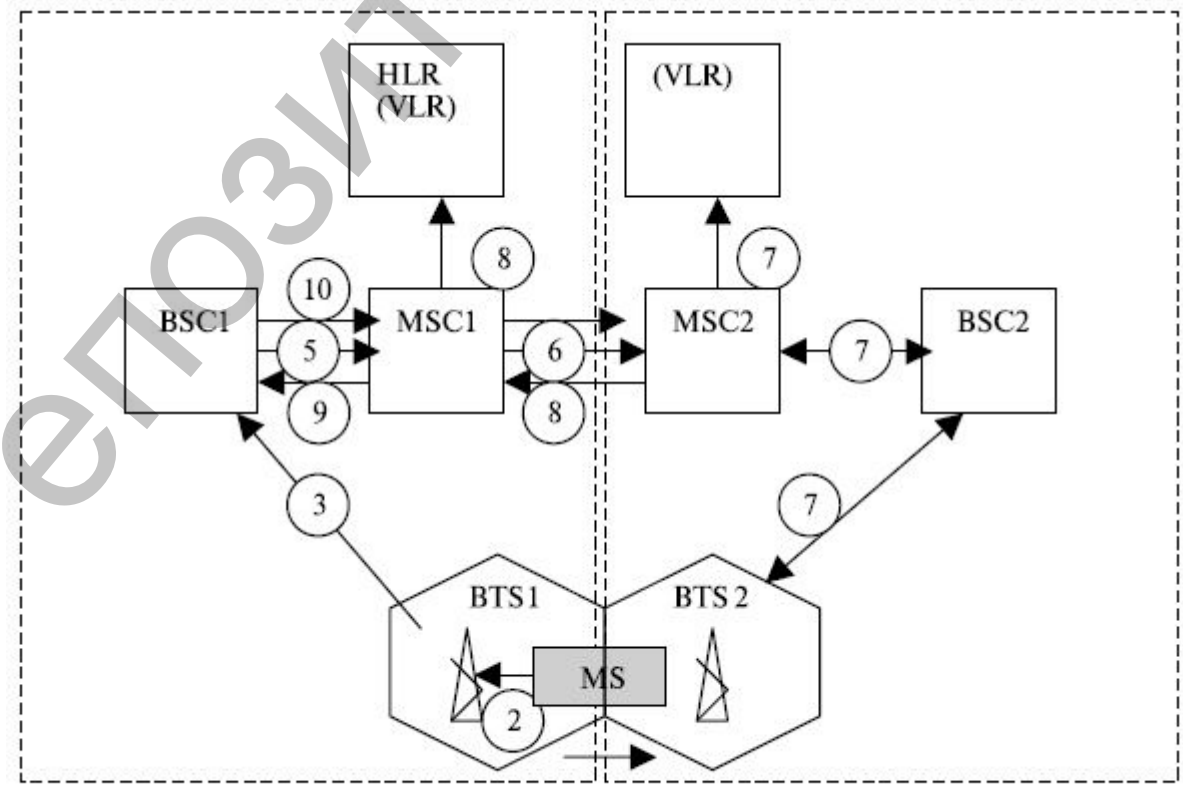


Рис. 1.16. Обмен сигналами при хэндовере

Роуминг

Роуминг — одна из самых важных функций сотовой связи. Необходимость в роуминге возникает каждый раз, когда абонент изменяет свое местоположение и перемещается в сеть, принадлежащую другому оператору. Роуминг бывает локальный (переезд внутри города или в пригород), национальный (в другой город или область) и международный (переезд в другую страну).

При перемещении абонента в другую сеть ее центр коммутации **мобильной связи** (MSC/VLR) запрашивает информацию в первоначальной сети (MSC/HLR) и при наличии подтверждения полномочий абонента регистрирует его. Данные о местоположении абонента постоянно обновляются в центре коммутации первоначальной сети (MSC/HLR), и все поступающие туда вызовы автоматически переадресовываются в ту сеть, где в данный момент находится абонент.

По способу регистрации различают следующие виды роуминга:

- автоматический, т. е. с возможностью провести процесс хэндовера;
- полуавтоматический, когда предварительно следует оповестить оператора о намерении посетить соответствующий регион;
- ручной, при котором абоненту вручается радиотелефон, включенный в сеть визита.

Для обеспечения роуминга необходимо выполнение следующих условий:

- наличие в требуемых регионах сотовых систем стандарта, совместимого со стандартом компании, у которой был приобретен радиотелефон;
- наличие соответствующих организационных и экономических соглашений о роуминговом обслуживании абонентов;
- наличие каналов связи между системами, обеспечивающих передачу звуковой, сигнальной и другой информации для роуминговых абонентов.

При организации роуминга недостаточно провести только технические мероприятия по соединению различных сетей сотовой связи. Очень важно еще решить проблему взаиморасчетов между операторами этих сетей.

Кроме того, для организации передачи сигнальных сообщений при автоматическом роуминге нужно создать соответствующие сигнальные каналы и программное обеспечение. Это требует определенных затрат. Поэтому между областями обслуживания различных операторов должна быть большая потребность в обслуживании роуминговой связи— больше, чем просто трафик от "случайно заехавших" абонентов.

Лекция 18. Системы персональной спутниковой связи

Спу́тниковая свя́зь— один из видов космической радиосвязи, основанный на использовании искусственных спутников земли в качестве ретрансляторов. Спутниковая связь осуществляется между земными станциями, которые могут быть как стационарными, так и подвижными.

Спутниковая связь является развитием традиционной радиорелейной связи путем вынесения ретранслятора на очень большую высоту (от десятков до сотен тысяч км). Так как зона его видимости в этом случае — почти половина Земного шара, то необходимость в цепочке ретрансляторов отпадает — в большинстве случаев достаточно и одного.

История

В 1945 году в статье «Внеземные ретрансляторы» («Extra-terrestrial Relays»), опубликованной в октябрьском номере журнала «Wireless World», английский учёный, писатель и изобретатель Артур Кларк предложил идею создания системы спутников связи на геостационарных орбитах, которые позволили бы организовать глобальную систему связи.

Впоследствии Кларк на вопрос, почему он не запатентовал изобретение (что было вполне возможно), отвечал, что не верил в возможность реализации подобной системы при своей жизни, а также считал, что подобная идея должна приносить пользу всему человечеству.

Первые исследования в области гражданской спутниковой связи в западных странах начали появляться во второй половине 50-х годов XX века. В США толчком к ним

12 августа 1960 года специалистами США был выведен на орбиту высотой 1500 км надувной шар. Этот космический аппарат назывался «Эхо-1». Его металлизированная оболочка диаметром 30 м выполняла функции пассивного ретранслятора.

Инженеры работают над первым в мире коммерческим спутником связи Early Bird 20 августа 1964 года 11 стран (СССР в их число не вошёл) подписали соглашение о создании международной организации спутниковой связи Intelsat (International Telecommunications Satellite organization). В СССР к тому времени была собственная развитая программа спутниковой связи, увенчавшаяся 23 апреля 1965 года успешным запуском связного советского спутника Молния-1. В рамках программы Intelsat первый коммерческий спутник связи Early Bird (англ.) («ранняя пташка»), произведенный корпорацией COMSAT, был запущен 6 апреля 1965 год.

По сегодняшним меркам спутник Early Bird (INTELSAT I) обладал более чем скромными возможностями: обладая полосой пропускания 50 МГц, он мог обеспечивать до 240 телефонных каналов связи[4]. В каждый конкретный момент времени связь могла осуществляться между земной станцией в США и только одной из трёх земных станций в Европе (в Великобритании, Франции или Германии), которые были соединены между собой кабельными линиями связи.

В дальнейшем технология шагнула вперед, и спутник INTELSAT IX уже обладал полосой пропускания 3456 МГц.

В СССР долгое время спутниковая связь развивались только в интересах Министерства Обороны СССР. В силу большей закрытости космической программы развитие спутниковой связи в социалистических странах шло иначе чем в западных странах. Развитие гражданской спутниковой связи началось соглашением между 9 странами социалистического блока о создании системы связи «Интерспутник» которое было подписано только в 1971 году

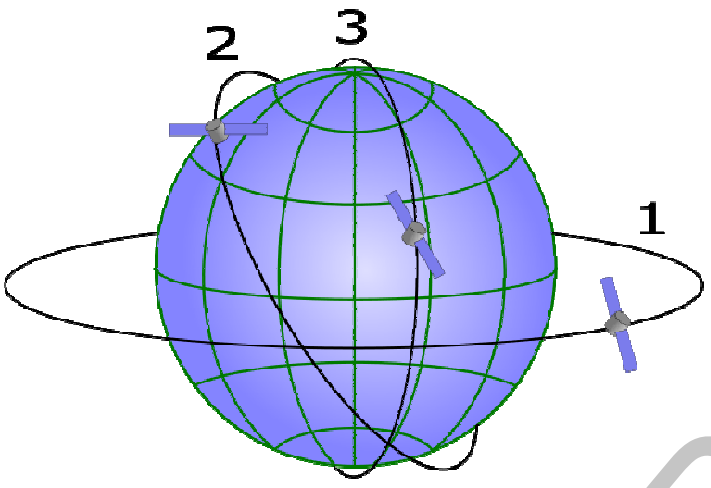
Спутниковые ретрансляторы

Пассивный спутник связи Echo-2. Металлизированная надувная сфера выполняла функции пассивного ретранслятора

В первые годы исследований использовались пассивные спутниковые ретрансляторы (примеры — спутники «Эхо» и «Эхо-2»), которые представляли собой простой отражатель радиосигнала (часто — металлическая или полимерная сфера с металлическим напылением), не несущий на борту какого-либо приёмопередающего оборудования. Такие спутники не получили распространения. Все современные спутники связи являются активными. Активные ретрансляторы оборудованы электронной аппаратурой для приема, обработки, усиления и ретрансляции сигнала. Спутниковые ретрансляторы могут быть нерегенеративными и регенеративными. Нерегенеративный спутник, приняв сигнал от одной земной станции, переносит его на другую частоту, усиливает и передает другой земной станции. Спутник может использовать несколько независимых каналов, осуществляющих эти операции, каждый из которых работает с определенной частью спектра (эти каналы обработки называются транспондерами).

Регенеративный спутник производит демодуляцию принятого сигнала и заново модулирует его. Благодаря этому исправление ошибок производится дважды: на спутнике и на принимающей земной станции. Недостаток этого метода — сложность (а значит, гораздо более высокая цена спутника), а также увеличенная задержка передачи сигнала.

Орбиты спутниковых ретрансляторов



Орбиты: 1 — экваториальная, 2 — наклонная, 3 — полярная

Орбиты, на которых размещаются спутниковые ретрансляторы, подразделяют на три класса[9]:

- экваториальные,
- наклонные,
- полярные.

Важной разновидностью экваториальной орбиты является геостационарная орбита, на которой спутник вращается с угловой скоростью, равной угловой скорости Земли, в направлении, совпадающем с направлением вращения Земли. Очевидным преимуществом геостационарной орбиты является то, что приемник в зоне обслуживания «видит» спутник постоянно.

Однако геостационарная орбита одна, и все спутники вывести на неё невозможно. Другим её недостатком является бо́льшая высота, а значит, и бо́льшая цена вывода спутника на орбиту. Кроме того, спутник на геостационарной орбите неспособен обслуживать земные станции в приполярной области.

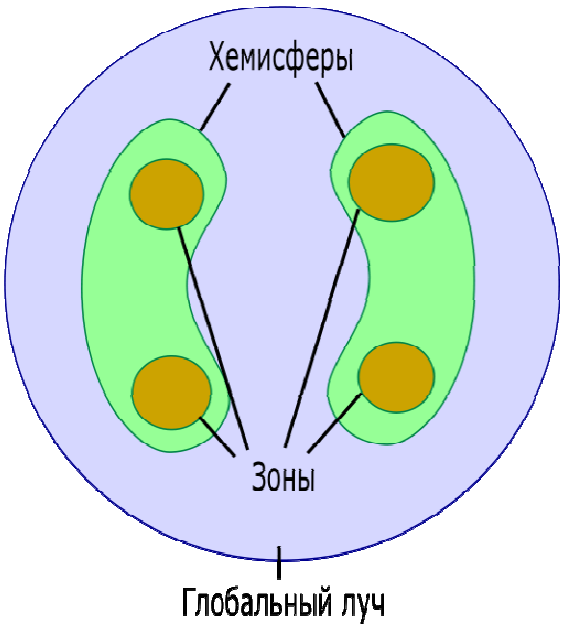
Наклонная орбита позволяет решить эти проблемы, однако, из-за перемещения спутника относительно наземного наблюдателя необходимо запускать не меньше трех спутников на одну орбиту, чтобы обеспечить круглосуточный доступ к связи.

Полярная орбита — предельный случай наклонной (с наклонением 90º).

При использовании наклонных орбит земные станции оборудуются системами слежения, осуществляющими наведение антенны на спутник. Станции, работающие со спутниками, находящимися на геостационарной орбите, как правило, также оборудуются такими системами, чтобы компенсировать отклонение от идеальной геостационарной орбиты. Исключение составляют небольшие антенны, используемые для приема спутникового телевидения: их диаграмма направленности достаточно

широкая, поэтому они не чувствуют колебаний спутника возле идеальной точки.

Многократное использование частот. Зоны покрытия



Типичная карта покрытия спутника, находящегося на геостационарной орбите

Поскольку радиочастоты являются ограниченным ресурсом, необходимо обеспечить возможность использования одних и тех же частот разными земными станциями. Сделать это можно двумя способами:

- пространственное разделение — каждая антенна спутника принимает сигнал только с определенного района, при этом разные районы могут использовать одни и те же частоты,
- поляризационное разделение — различные антенны принимают и передают сигнал во взаимно перпендикулярных плоскостях поляризации, при этом одни и те же частоты могут применяться два раза (для каждой из плоскостей).

Типичная карта покрытия для спутника, находящегося на геостационарной орбите, включает следующие компоненты:

- глобальный луч — производит связь с земными станциями по всей зоне покрытия, ему выделены частоты, не пересекающиеся с другими лучами этого спутника.
- лучи западной и восточной полусфер — эти лучи поляризованы в плоскости А, причем в западной и восточной полусферах используется один и тот же диапазон частот.
- зонные лучи — поляризованы в плоскости В (перпендикулярной А) и используют те же частоты, что и лучи полусфер. Таким образом, земная станция, расположенная в одной из зон, может использовать также лучи полусфер и глобальный луч.

При этом все частоты (за исключением зарезервированных за глобальным лучом) используются многократно: в западной и восточной полусферах и в каждой из зон.

Частотные диапазоны

Выбор частоты для передачи данных от земной станции к спутнику и от спутника к земной станции не является произвольным. От частоты зависит, например, поглощение радиоволн в атмосфере, а также необходимые размеры передающей и приемной антенн. Частоты, на которых происходит передача от земной станции к спутнику, отличаются от частот, используемых для передачи от спутника к земной станции (как правило, первые выше).

Частоты, используемые в спутниковой связи, разделяют на диапазоны, обозначаемые буквами. К сожалению, в различной литературе точные границы диапазонов могут не совпадать. Ориентировочные значения даны в рекомендации ITU-R V.431-6:

Название диапазона	**Частоты (согласно ITU-R V.431-6)**	**Применение**
L	1,5 ГГц	Подвижная спутниковая связь
S	2,5 ГГц	Подвижная спутниковая связь
C	4 ГГц, 6 ГГц	Фиксированная спутниковая связь
X	Для спутниковой связи рекомендациями ITU-R частоты не определены. Для приложений радиолокации указан диапазон 8-12 ГГц.	Фиксированная спутниковая связь (для военных целей)
Ku	11 ГГц, 12 ГГц, 14 ГГц	Фиксированная спутниковая связь, спутниковое вещание
K	20 ГГц	Фиксированная спутниковая связь, спутниковое вещание
Ka	30 ГГц	Фиксированная спутниковая связь, межспутниковая связь

Используются и более высокие частоты, но повышение их затруднено высоким поглощением радиоволн этих частот атмосферой. Ku-диапазон позволяет производить прием сравнительно небольшими антеннами, и поэтому используется в спутниковом телевидении (DVB), несмотря на то, что в этом диапазоне погодные условия оказывают существенное влияние на качество передачи.

Для передачи данных крупными пользователями (организациями) часто применяется C-диапазон. Это обеспечивает более высокое качество приема, но требует довольно больших размеров антенны.

Модуляция и помехоустойчивое кодирование

Особенностью спутниковых систем связи является необходимость работать в условиях сравнительно низкого отношения сигнал/шум, вызванного несколькими факторами:

- значительной удаленностью приемника от передатчика,
- ограниченной мощностью спутника (невозможностью вести передачу на большой мощности).

В связи с этим спутниковая связь плохо подходит для передачи аналоговых сигналов. Поэтому для передачи речи её предварительно оцифровывают, используя, например, импульсно-кодовую модуляцию (ИКМ.

Для передачи цифровых данных по спутниковому каналу связи они должны быть сначала преобразованы в радиосигнал, занимающий определенный частотный диапазон. Для этого применяется модуляция (цифровая модуляция называется также манипуляцией). Наиболее распространенными видами цифровой модуляции для приложений спутниковой связи являются фазовая манипуляция и квадратурная амплитудная модуляция[15]. Например, в системах стандарта DVB-S2 применяются QPSK, 8-PSK, 16-APSK и 32-APSK.

Модуляция производится на земной станции. Модулированный сигнал усиливается, переносится на нужную частоту и поступает на передающую антенну. Спутник принимает сигнал, усиливает, иногда регенерирует, переносит на другую частоту и с помощью определённой передающей антенны транслирует на землю.

Из-за низкой мощности сигнала возникает необходимость в системах исправления ошибок. Для этого применяются различные схемы помехоустойчивого кодирования, чаще всего различные варианты свёрточных кодов (иногда в сочетании с кодами Рида-Соломона), а также турбо-коды и LDPC-коды.

Множественный доступ

Для обеспечения возможности одновременного использования спутникового ретранслятора несколькими пользователями применяют системы множественного доступа:

- Множественный доступ с частотным разделением — при этом каждому пользователю предоставляется отдельный диапазон частот.
- множественный доступ с временны́м разделением — каждому пользователю предоставляется определенный временной интервал (таймслот), в течение которого он производит передачу и прием данных.
- множественный доступ с кодовым разделением — при этом каждому пользователю выдается кодовая последовательность, ортогональная кодовым последовательностям других пользователей. Данные пользователя накладываются на кодовую последовательность таким образом, что передаваемые сигналы различных пользователей не мешают друг другу, хотя и передаются на одних и тех же частотах.

Кроме того, многим пользователям не требуется постоянный доступ к спутниковой связи. Этим пользователям канал связи (таймслот) выделяется

по требованию с помощью технологии DAMA (Demand Assigned Multiple Access — множественный доступ с предоставлением каналов по требованию).

Применение спутниковой связи

Магистральная спутниковая связь

Изначально возникновение спутниковой связи было продиктовано потребностями передачи больших объёмов информации. Первой системой спутниковой связи стала система Intelsat, затем были созданы аналогичные региональные организации (Eutelsat, Arabsat и другие). С течением времени доля передачи речи в общем объёме магистрального трафика постоянно снижалась, уступая место передаче данных.

С развитием волоконно-оптических сетей последние начали вытеснять спутниковую связь с рынка магистральной связи.

Системы VSAT

Системы VSAT (Very Small Aperture Terminal — терминал с очень маленькой апертурой) предоставляют услуги спутниковой связи клиентам (как правило, небольшим организациям), которым не требуется высокая пропускная способность канала. Скорость передачи данных для VSAT-терминала обычно не превышает 2048 кбит/с.

Слова «очень маленькая апертура» относятся к размерам антенн терминалов по сравнению с размерами более старых антенн магистральных систем связи. VSAT-терминалы, работающие в C-диапазоне, обычно используют антенны диаметром 1,8-2,4 м, в Ku-диапазоне — 0,75-1,8 м.

В системах VSAT применяется технология предоставления каналов по требованию.

Системы подвижной спутниковой связи

Особенностью большинства систем подвижной спутниковой связи является маленький размер антенны терминала, что затрудняет прием сигнала. Для того, чтобы мощность сигнала, достигающего приемника, была достаточной, применяют одно из двух решений:

- Спутники располагаются на геостационарной орбите. Поскольку эта орбита удалена от Земли на расстояние 35786 км, на спутник требуется установить мощный передатчик. Этот подход используется системой Inmarsat (основной задачей которой является предоставление услуг связи морским судам) и некоторыми региональными операторами персональной спутниковой связи (например, Thuraya).
- Множество спутников располагается на наклонных или полярных орбитах. При этом требуемая мощность передатчика не так высока, и стоимость вывода спутника на орбиту ниже. Однако такой подход требует не только большого числа спутников, но и разветвленной сети наземных коммутаторов. Подобный метод используется операторами Iridium и Globalstar.

С операторами персональной спутниковой связи конкурируют операторы сотовой связи. Характерно, что как Globalstar, так и Iridium

испытывали серьёзные финансовые затруднения, которые довели Iridium до реорганизационного банкротства в 1999 г.

В декабре 2006 года был запущен экспериментальный геостационарный спутник Кику-8 с рекордно большой площадью антенны, который предполагается использовать для отработки технологии работы спутниковой связи с мобильными устройствами, не превышающими по размерам сотовые телефоны.

Спутниковый Интернет

Спутниковая связь находит применение в организации «последней мили» (канала связи между интернет-провайдером и клиентом), особенно в местах со слабо развитой инфраструктурой

Особенностями такого вида доступа являются:

- Разделение входящего и исходящего трафика и привлечение дополнительных технологий для их совмещения. Поэтому такие соединения называют асимметричными.
- Одновременное использование входящего спутникового канала несколькими (например 200-ми) пользователями: через спутник одновременно передаются данные для всех клиентов «вперемешку», фильтрацией ненужных данных занимается клиентский терминал (по этой причине возможна «Рыбалка со спутника»).

По типу исходящего канала различают:

- Терминалы, работающие только на прием сигнала (наиболее дешевый вариант подключения). В этом случае для исходящего трафика необходимо иметь другое подключение к Интернету, поставщика которого называют наземным провайдером. Для работы в такой схеме привлекается туннелирующее программное обеспечение, обычно входящее в поставку терминала. Несмотря на сложность (в том числе сложность в настройке), такая технология привлекательна большой скоростью по сравнению с dial-up за сравнительно небольшую цену.
- Приемо-передающие терминалы. Исходящий канал организуется узким (по сравнению со входящим). Оба направления обеспечивает одно и то же устройство, и поэтому такая система значительно проще в настройке (особенно если терминал внешний и подключается к компьютеру через интерфейс Ethernet). Такая схема требует установки на антенну более сложного (приемо-передающего) конвертера.

И в том, и в другом случае данные от провайдера к клиенту передаются, как правило, в соответствии со стандартом цифрового вещания DVB, что позволяет использовать одно и то же оборудование как для доступа в сеть, так и для приема спутникового телевидения.

Недостатки спутниковой связи

Слабая помехозащищённость

Огромные расстояния между земными станциями и спутником являются причиной того, что отношение сигнал/шум на приемнике очень невелико (гораздо меньше, чем для большинства радиорелейных линий

связи). Для того, чтобы в этих условиях обеспечить приемлемую вероятность ошибки, приходится использовать большие антенны, малошумящие элементы и сложные помехоустойчивые коды. Особенно остро эта проблема стоит в системах подвижной связи, так как в них есть ограничение на размер антенны и, как правило, на мощность передатчика.

Влияние атмосферы

На качество спутниковой связи оказывают сильное влияние эффекты в тропосфере и ионосфере.

Поглощение в тропосфере

Поглощение сигнала атмосферой находится в зависимости от его частоты. Максимумы поглощения приходятся на 22,3 ГГц (резонанс водяных паров) и 60 ГГц (резонанс кислорода). В целом, поглощение существенно сказывается на распространении сигналов с частотой выше 10 ГГц (то есть, начиная с Ku-диапазона). Кроме поглощения, при распространении радиоволн в атмосфере присутствует эффект замирания, причиной которому является разница в коэффициентах преломления различных слоев атмосферы.

Ионосферные эффекты

Эффекты в ионосфере обусловлены флуктуациями распределения свободных электронов. К ионосферным эффектам, влияющим на распространение радиоволн, относят мерцание, поглощение, задержку распространения, дисперсию, изменение частоты, вращение плоскости поляризации[27]. Все эти эффекты ослабляются с увеличением частоты. Для сигналов с частотами, большими 10 ГГц, их влияние невелико.

Эффект	100 МГц	300 МГц	1 ГГц	3 ГГц	10 ГГц
Вращение плоскости поляризации	30 оборотов	3,3 оборота	108°	12°	1,1°
Дополнительная задержка сигнала	25 мс	2,8 мс	0,25 мс	28 нс	2,5 нс
Поглощение в ионосфере (на полюсе)	5 дБ	1,1 дБ	0,05 дБ	0,006 дБ	0,0005 дБ
Поглощение в ионосфере (в средних широтах)	<1 дБ	0,1 дБ	<0,01 дБ	<0,001 дБ	<0,0001 дБ

Сигналы с относительно низкой частотой (L-диапазон и частично C-диапазон) страдают от ионосферного мерцания, возникающего из-за неоднородностей в ионосфере. Результатом этого мерцания является постоянно меняющаяся мощность сигнала.

Задержка распространения сигнала

Проблема задержки распространения сигнала так или иначе затрагивает все спутниковые системы связи. Наибольшей задержкой обладают системы, использующие спутниковый ретранслятор на геостационарной орбите. В этом случае задержка, обусловленная

конечностью скорости распространения радиоволн, составляет примерно 250 мс, а с учетом мультиплексирования, коммутации и задержек обработки сигнала общая задержка может составлять до 400 мс.

Задержка распространения наиболее нежелательна в приложениях реального времени, например, в телефонной связи. При этом, если время распространения сигнала по спутниковому каналу связи составляет 250 мс, разница во времени между репликами абонентов не может быть меньше 500 мс.

В некоторых системах (например, в системах VSAT, использующих топологию «звезда») сигнал дважды передается через спутниковый канал связи (от терминала к центральному узлу, и от центрального узла к другому терминалу). В этом случае общая задержка удваивается.

Влияние солнечной интерференции

При приближении Солнца к оси спутника-наземная станция радиосигнал, принимаемый со спутника наземной станцией, искажается в результате интерференции.

Литература

1. Олифер Н.А., Олифер В.Г., Компьютерные сети. Принципы, технологии, протоколы. СПб: Питер, 2006. – 672с.
2. Таненбаум Компьютерные сети. СПб: Питер, 2007 г. - 992 стр.
3. Уилсон, Э. Мониторинг и анализ сетей. Методы выявления неисправностей / Э. Уилсон. – М. : ЛОРИ, 2002.
4. Филимонов, А. Протоколы Интернета / А. Филимонов. – СПб. : BHV-Санкт-Петербург, 2003.
5. Золотов, С. Протоколы Internet / С. Золотов. – СПб. : BHV-Санкт-Петербург, 1998.
6. Кожанов Ю.Ф Интерфейсы и протоколы сетей следующего поколения: научно-популярное издание СПб., 2006. –218 с
7. Емельянов Г.А., Шварцман В.О.Передача дискретной информации: Учебник для вузовМ.: Радио и связь, 1982. — 240 с.
8. Прангишвили И.В., Подлазов В.С., Стецюра Г.Г. Локальные микропроцессорные вычислительные сети. М.: Наука, 1984. — 176 с.
9. Флинт Д Локальные сети ЭВМ: Пер. с англ. М.: Финансы и статистика, 1986. — 357 с.
10. Семенов Ю.А Алгоритмы телекоммуникационных сетей. Часть 2.Протоколы и алгоритмы маршрутизации INTERNETM:ИНТУИТ.РУ, БИНОМ. Лаборатория знаний, 2007 – 832 с
11. А.А. Мячев, В.Н. Степанов, В.К, Щербо Интерфейсы систем обработки данных: Справочник/ Под ред. А.А. Мячева. М.: Радио и связь, 1989. — 416 с.
12. Овчинников В.В., Рыбкин И.И Техническая база интерфейсов локальных вычислительных сетей. М.: Радио и связь, 1989. — 272 с.
13. Дженнингс Ф. Практическая передача данных: Модемы, сети и протоколы: Пер. с англ М.: Мир, 1989. — 272 с.