**06.11.2020 ИС 2к Основы архитектуры**

**Тема Ввод-вывод**

**Принтеры**

Рано или поздно пользователю потребуется напечатать созданный документ или
страницу, полученную из Интернета, поэтому компьютеры могут быть оснащены
принтером. В этом разделе мы опишем некоторые наиболее распространенные
типы монохромных (то есть черно-белых) и цветных принтеров.

Лазерные принтеры

Вероятно, самым удивительным изобретением в области печатных технологий со
времен Йоганна Гуттенберга ( Johann Gutenberg), который изобрел подвижную
литеру в XV веке, является **лазерный принтер**. Это устройство сочетает хоро-
шее качество печати, универсальность, высокую скорость работы и умеренную
стоимость. В лазерных принтерах используется почти та же технология, что
и в фотокопировальных устройствах. Многие компании производят устройства,
совмещающие свойства копировальной машины, принтера и иногда факса.

Схематически устройство принтера показано на рис. 2.32. Главной частью этого принтера является вращающийся барабан (в некоторых более дорогостоящих
системах вместо барабана используется лента). Перед печатью каждого листа
барабан получает напряжение около 1000 вольт и окружается фоточувствитель-
ным материалом. Свет лазера проходит вдоль барабана (по длине), почти как
пучок электронов в электронно-лучевой трубке, только вместо напряжения для
сканирования барабана используется вращающееся восьмиугольное зеркало. Луч
света модулируется, в результате получается набор темных и светлых участков.
Участки, на которые воздействует луч, теряют свой электрический заряд

 Вращающееся

**Рис. 2.32.** Схема работы лазерного принтера

После того как нарисована строка точек, барабан немного поворачивается для
создания следующей строки. В итоге первая строка точек достигает резервуара
с тонером (электростатическим черным порошком). Тонер притягивается к за-
ряженным точкам, и так формируется визуальное изображение строки. Через

некоторое время барабан с тонером прижимается к бумаге, оставляя на ней отпе-
чаток изображения. Затем лист проходит через нагретые валики, и изображение
закрепляется. После этого барабан разряжается и остатки тонера счищаются
с него. Теперь он готов к печати следующей страницы.

Едва ли нужно говорить, что этот процесс представляет собой чрезвычайно
сложную комбинацию приемов, требующих знания физики, химии, механики
и оптики. Впрочем, некоторые фирмы предлагают готовые модули, называемые
**блоками печати** (print engines). Производители лазерных принтеров допол-
няют блоки печати собственной электроникой и программным обеспечением.
Электроника лазерных принтеров состоит из быстродействующего процессора
и нескольких мегабайтов памяти для хранения полного изображения в битовой
форме и различных шрифтов, одни из которых встроены, а другие загружаются
из памяти. Большинство принтеров получают команды, описывающие печа-
таемую страницу (в противоположность принтерам, получающим изображения
в битовой форме от центрального процессора). Эти команды обычно даются на
языке PCL от HP или PostScript от Adobe - полноценных, хотя и специализи-
рованных, языках программирования.

Лазерные принтеры с разрешающей способностью 600 dpi и выше могу пе-
чатать черно-белые фотографии, но технология при этом гораздо сложнее, чем
может показаться на первый взгляд. Рассмотрим фотографию, отсканированную
с разрешающей способностью 600 dpi, которую нужно напечатать на принтере
с такой же разрешающей способностью (600 dpi). Сканированное изображение
содержит 600 X 600 пикселов на дюйм, каждый пиксел характеризуется опреде-
ленной градацией серого цвета от 0 (белый цвет) до 255 (черный цвет). Принтер
может печатать с разрешающей способностью 600 dpi, но каждый напечатанный
пиксел может быть либо черного цвета (когда есть тонер), либо белого цвета
(когда нет тонера). Градации серого печататься не могут.

При печати таких изображений имеет место так называемая **обработка полу-
тонов** (как при печати серийных плакатов). Изображение разбивается на ячейки,
каждая по 6 X 6 пикселов. Каждая ячейка может содержать от 0 до 36 черных
пикселов. Человеческому глазу ячейка с большим количеством черных пикселов
кажется темнее, чем ячейка с небольшим количеством черных пикселов. Серые
тона в диапазоне от 0 до 255 передаются следующим образом. Этот диапазон де-
лится на 37 зон. Серые тона от 0 до 6 расположены в зоне 0, от 7 до 13 - в зоне 1
и т. д. (зона 36 немного меньше, чем другие, потому что 256 на 37 без остатка не
делится). Когда встречаются тона зоны 0, ячейка оставляется белой, как показано
на рис. 2.33, *а.* Тона зоны 1 передаются одним черным пикселом в ячейке. Тона
зоны 2 - двумя пикселами в ячейке, как показано на рис. 2.33, *б.* Изображения
серых тонов других зон показаны на рис. 2.33, *в-е.* Если фотография отсканиро-

а б в г д е

**Рис. 2.33.** Изображение серых полутонов различных зон: 0-6 (а); 14-20 *(б);* 28-34 *(в);
56-62* (г); 105-111 (л); 161-167 (е)

вана с разрешающей способностью 600 dpi, после подобной обработки полутонов
фактическая разрешающая способность напечатанного изображения снижается
до 100 ячеек на дюйм. Данная величина называется **градацией полутонов** и из-
меряется в **lpi** (lines per inch - строк на дюйм).

Цветные принтеры

Хотя лазерные принтеры чаще всего являются монохромными, цветные лазер-
ные принтеры получают все более широкое распространение, поэтому о цветной
печати тоже стоит сказать пару слов (причем все сказанное также относится
к струйным и другим принтерам). Цветные изображения могут строиться дву-
мя способами: испусканием света и отражением света. Испускание света имеет
место, например, при создании изображений в электронно-лучевых мониторах.
В данном случае изображение строится путем аддитивного наложения трех
основных цветов: красного, зеленого и синего.

Отраженный свет используется при создании цветных фотографий и кар-
тинок в глянцевых журналах. В этом случае поглощается свет с определенной
длиной волны, а остальной свет отражается. Такие изображения создаются путем
субтрактивного наложения трех основных цветов: голубого (красный полностью
поглощен), желтого (синий полностью поглощен) и сиреневого (зеленый полно-
стью поглощен). Теоретически путем смешения голубых, желтых и сиреневых
чернил можно получить любой цвет. Но на практике очень сложно получить
такие чернила, которые полностью поглощали бы весь свет и в результате да-
вали черный цвет. По этой причине практически во всех цветных печатающих
устройствах используются чернила четырех цветов: голубого (Cyan), сиреневого
(Magenta), желтого (Yellow) и черного (blacK). Такая цветовая модель называется
**CMYK** (из слова «black» берется последняя буква, чтобы отличать его от слова
«blue» в модели RGB). Мониторы, напротив, для создания цветного изображения
используют испускаемый свет и наложение красного, зеленого и синего цветов.

Полный набор цветов, который может производить монитор или принтер, на-
зывается **цветовой шкалой**. Не существует такого устройства, которое полностью
передавало бы цвета окружающего нас мира. В лучшем случае устройство дает
всего 256 степеней интенсивности каждого цвета, и в итоге получается только
16 777 216 различных цветов. Несовершенство технологий еще больше сокра-
щает это число, а оставшиеся цвета не дают полного цветового спектра. Кроме
того, цветовосприятие связано не только с физическими свойствами света, но
и с работой «палочек» и «колбочек» в сетчатке глаза.

Из всего этого следует, что превратить красивое цветное изображение, которое
замечательно смотрится на экране, в идентичное печатное изображение очень
сложно. Среди основных проблем можно назвать следующие:

+ цветные мониторы используют поглощенный свет; цветные принтеры -
отраженный;

+ электронно-лучевая трубка дает 256 оттенков каждого цвета, цветные
принтеры должны обеспечивать обработку полутонов;

+ мониторы имеют темный фон; бумага - светлый;

+ цветовая модель RGB монитора и модель CMYK принтера отличаются
друг от друга.

Чтобы цветные печатные изображения соответствовали реальной действи-
тельности (или хотя бы изображениям на экране), необходима калибровка обо-
рудования, сложное программное обеспечение и компетентность пользователя.

Струйные принтеры

Дома удобно использовать недорогие **струйные принтеры**. В таком принтере
подвижная печатающая головка содержит картридж с чернилами. Она двигается
горизонтально над бумагой, а чернила в это время выпрыскиваются из крошеч-
ных сопел. Объем одной порции чернил приблизительно равен один пиколитр.
Для наглядности уточним, что в одной капле воды может уместится около
100 миллионов таких порций.

Струйные принтеры бывают двух типов: пьезоэлектрические (производятся
Epson) и термографические (производятся Canon, HP и Lexmark). В пьезоэлек-
трических струйных принтерах рядом с чернильной камерой устанавливается
специальный кристалл. При подаче на этот кристалл напряжения он деформи-
руется, в результате из форсунки выпускаются чернила. Чем выше напряжение,
тем больше выходная порция чернил, причем управление этим процессом про-
изводится программно.

В термографических (пузырьковых) струйных принтерах в каждой форсунке
устанавливается небольшой резистор. При подаче напряжения резистор быстро
нагревается, доводит температуру чернил до точки кипения, в результате послед-
ние превращаются в пузырьки газа. Поскольку объем пузырька больше объема
чистых чернил, в форсунке создается повышенное давление, под влиянием кото-
рого чернила распыляются на бумагу. Затем форсунка охлаждается, и в результате
снижения давления внутри форсунки в нее из картриджа подается новая порция
чернил. Скорость работы принтера в рамках этой схемы ограничена временными
рамками цикла кипения/охлаждения. Размер всех формируемых чернильных
капель одинаков, причем, как правило, он уступает аналогичному показателю
пьезоэлектрических принтеров.

Струйные принтеры обычно имеют разрешающую способность от 1200 dpi
(dots per inch - точек на дюйм) до 4800 dpi. Они достаточно дешевы, работают
бесшумно, однако отличаются низкой скоростью печати и дороговизной кар-
триджей. Качество печати хорошее - если распечатать фотографию с высоким
разрешением на ведущей модели любой линейки струйных принтеров, результат
будет не отличить от обычной фотографии формата 8 X 10.

Для получения лучших результатов должны использоваться особые чернила
и особая бумага. Существует два вида чернил. **Чернила на основе красителя**состоят из красителей, растворенных в жидкой среде. Они дают яркие цвета
и легко вытекают из картриджа. Главным недостатком таких чернил является то,
что они быстро выгорают под воздействием ультрафиолетовых лучей, которые
содержатся в солнечном свете. **Чернила на основе пигмента** содержат твердые
частицы пигмента, погруженные в жидкость. Жидкость испаряется с бумаги,
а пигмент остается. Чернила не выгорают, но зато дают не такие яркие краски,
как чернила на основе красителя. Кроме того, частицы пигмента часто засоряют
выпускные отверстия картриджей, поэтому их нужно периодически чистить. Для
печати фотографий необходима мелованная или глянцевая бумага. Эти особые
виды бумаги созданы специально для того, чтобы удерживать капельки чернил
и не давать им растекаться.

Специализированные принтеры

Лазерные и струйные принтеры лидируют на рынке домашней и офисной печати.
Однако существуют и другие технологии печати, применяемые в других ситуаци-
ях с другими требованиями по качеству цвета, цене и другим характеристикам.

Следующий тип принтеров - **принтеры с твердыми чернилами**. В этих
принтерах содержится 4 твердых блока специальных восковых чернил, которые
затем расплавляются, для чего перед началом печати должно пройти 10 минут
(время, необходимое для того, чтобы расплавить чернила). Горячие чернила
выпрыскиваются на бумагу, где они затвердевают и закрепляются после про-
хождения листа между двумя валиками. В каком-то смысле эта технология
объединяет принципы работы струйных принтеров (напыление чернил) и ла-
зерных принтеров (закрепление изображения на бумаге при помощи твердых
резиновых роликов).

**Принтеры с восковыми чернилами** содержат широкую ленту из четырех-
цветного воска, которая разделяется на отрезки размером с лист бумаги. Тысячи
нагревательных элементов растапливают воск, когда бумага проходит под лентой.
Воск закрепляется на бумаге в форме пикселов в соответствие с цветовой моде-
лью CMYK. Такие принтеры когда-то были очень популярными, но сейчас их
вытеснили другие типы принтеров с более дешевыми расходными материалами.

Еще одна разновидность принтеров работает на основе технологии **субли-
мации**. Название вызывает фрейдистские ассоциации1, однако в науке под
сублимацией понимается переход твердых веществ в газообразное состояние
минуя стадию жидкости. Таким материалом является, например, сухой лед
(замороженный углекислый газ). В принтере, работающем на основе процесса
сублимации, контейнер с красителями CMYK двигается над термической печа-
тающей головкой, которая содержит тысячи программируемых нагревательных
элементов. Красители мгновенно испаряются и впитываются специальной
бумагой. Каждый нагревательный элемент может производить 256 различных
температур. Чем выше температура, тем больше красителя осаждается и тем
интенсивнее получается цвет. В отличие от всех других цветных принтеров,
данный принтер способен воспроизводить цвета практически сплошного спектра,
поэтому процедура обработки полутонов не нужна. Процесс сублимации часто
используется при изготовлении так называемых моментальных снимков на спе-
циальной дорогостоящей бумаге.

Последнюю разновидность составляют **термографические принтеры**. Они
содержат небольшую печатающую головку с множеством игольчатых элемен-
тов. При прохождении электрического тока иглы очень быстро нагреваются.
Над печатающей головкой проходит специальная термочувствительная бумага,
и в тех местах, где находятся нагретые иглы, появляются точки. В сущности,
термографический принтер работает по принципу старого матричного прин-
тера, в котором контакты через красящую ленту оставляли точки на бумаге.
Термографические принтеры широко применяются для печати чеков в магазинах,
банкоматах, автоматизированных заправках и т. д.

**Телекоммуникационное оборудование**

Большинство современных компьютеров подключаются к компьютерным сетям,
из которых наиболее распространен Интернет. Для доступа к подобного рода
сетям требуется специальное оборудование. В этом разделе рассматриваются
принципы работы такого оборудования.

Модемы

С ростом количества компьютеров в последние годы возникла необходимость
связать их между собой. Например, можно связать свой домашний компьютер
с компьютером на работе, с поставщиком услуг Интернета или банковской
системой. Для обеспечения такой связи часто используется телефонная линия.

Однако обычная телефонная линия (равно как и кабель) плохо подходит для
передачи компьютерных сигналов, в которых 0 обычно соответствует нулевому
напряжению, а 1 - напряжению от 3 до 5 вольт (рис. 2.34, *а).* Двухуровневые
сигналы во время передачи по телефонной линии, которая предназначена для
передачи голоса, подвергаются сильным искажениям, ведущим к ошибкам
в передаче. Тем не менее синусоидальный сигнал с частотой от 1000 до 2000 Гц,
который называется **несущим**, может передаваться с относительно небольшими
искажениями, и это свойство используется при передаче данных в большинстве
телекоммуникационных систем.

Время

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| В2 °щ В1  | 1 | 0 0 | 1 | 0 | 1 | 1 | 0 0 0 | 1 |
|  |  I  |  |  | **I** | —I—>— |  |

к

Q.

п.

го

X

I I

I I

I. . . I

i i I

i i I

i i I

I I I

I. . I I.

11

Высокая

амплитуда i

Высокая

**Низкая
I амплитуда**

Низкая

i i

частота > > частота

i i

**ym/mvYv\MA**

**Фазовый**

**переход**

**Рис. 2.34.** Последовательная передача двоичного числа 01001011000100
по телефонной линии: двухуровневый сигнал (а); амплитудная модуляция *(б* );
частотная модуляция (в); фазовая модуляция (г )

Поскольку форма синусоидальной волны полностью предсказуема, она не не-
сет никакой информации. Однако изменяя амплитуду, частоту или фазу, можно



передавать последовательность нулей и единиц, как показано на рис. 2.34. Этот
процесс называется **модуляцией**. При **амплитудной модуляции** используются
2 уровня напряжения, соответственно для 0 и 1 (рис. 2.34, б). Если цифровые
данные передаются с очень низкой скоростью, то при передаче 1 слышен громкий
шум, а при передаче 0 шум отсутствует.

При **частотной модуляции** уровень напряжения не меняется, но частоты не-
сущего сигнала для 1 и 0 различаются (рис. 2.34, *в).* В этом случае при передаче
цифровых данных можно услышать два тона: один из них соответствует 0, а дру-
гой - 1. Частотную модуляцию иногда называют **частотной манипуляцией**.

При простой **фазовой модуляции** амплитуда и частота сохраняются на одном
уровне, а фаза несущего сигнала меняется на 180°, когда данные меняются с 0
на 1 или с 1 на 0 (рис. 2.34, г). В более сложных системах фазовой модуляции
в начале каждого неделимого временного отрезка фаза несущего сигнала резко
сдвигается на 45, 135, 225 или 315°, чтобы передавать 2 бита за один временной
отрезок. Это называется **дибитной фазовой кодировкой**. Например, сдвиг по
фазе на 45° представляет 00, на 135° - 01 и т. д. Существуют системы для пере-
дачи трех и более битов за один временной отрезок. Число таких временных
интервалов (то есть число потенциальных изменений сигнала в секунду) назы-
вается скоростью в **бодах**. При передаче двух или более битов за один временной
отрезок скорость передачи битов будет превышать скорость в **бодах**. Отметим,
что термины «бод» и «бит» в этом контексте часто путают. Еще раз: скорость
в бодах определятся количеством изменений сигнала за секунду, а скорость в би-
тах - количеством битов, переданных за секунду. Как правило, скорость в битах
кратна скорости в бодах, но теоретически может быть и ниже.

Если данные состоят из последовательности 8-разрядных символов, было бы
желательно иметь средство связи для передачи 8 бит одновременно, то есть 8 пар
проводов. Так как телефонные линии, предназначенные для передачи голоса, обе-
спечивают только один канал связи, биты должны пересылаться последовательно
один за другим (или в группах по два, если используется дибитная кодировка).
Устройство, которое получает символы из компьютера в форме двухуровневых
сигналов (по одному биту в каждый отрезок времени) и передает биты по одному
или по два в форме амплитудной, фазовой или частотной модуляции, называется
модемом. Для указания на начало и конец каждого символа в начале и конце
8-разрядной цепочки ставятся начальный и конечный биты, таким образом, всего
получается 10 бит.

Модем посылает отдельные биты каждого символа через равные временные
отрезки. Например, скорость 9600 бод означает, что сигнал меняется каждые
104 микросекунд. Второй модем, получающий информацию, преобразует моду-
лированный несущий сигнал в двоичное число. Биты поступают в модем через
равные промежутки времени. Если модем встречает начало символа, его часы
сообщают, когда нужно начать считывать поступающие биты.

Современные модемы передают данные со скоростью от 56 Кбит/с, что
обычно соответствует более низкой скорости в бодах. Они сочетают разные
технологий для передачи нескольких битов за один бод, модулируя амплитуду,
частоту и фазу. Почти все современные модемы являются **дуплексными**, то есть
могут передавать информацию в обоих направлениях одновременно, используя
различные частоты. Модемы и линии связи, которые не могут передавать ин-

формацию в обоих направлениях одновременно (как одноколейная железная
дорога), называются **полудуплексными**. Линии связи, которые могут передавать
информацию только в одном направлении, называются **симплексными**.

Цифровые абонентские линии

Взяв однажды планку в 56 Кбит/с, инженеры телефонных компаний с чувством
выполненного долга успокоились на достигнутом. Тем временем поставщи-
ки услуг кабельного телевидения стали предлагать абонентам подключение
к Интернету по общим кабелям на скорости до 10 Мбит/с. Поставщики услуг
спутниковой связи пошли еще дальше, обеспечив возможность подключения
на скорости свыше 50 Мбит/с. Чем большее значение приобретали услуги по
предоставлению доступа в Интернет для телефонных компаний, тем отчетливее
они понимали, что для сохранения конкурентоспособности нужно предложить
рынку какую-то более совершенную услугу, нежели подключение по обычному
модему. В результате этих раздумий на свет появилась новая цифровая услуга
доступа в Интернет. Услуги, в которых предлагается пропускная способность,
превышающая аналогичный показатель для стандартного модемного соединения,
иногда называют **широкополосными**, но, честно говоря, это скорее маркетинго-
вый, чем содержательный технический термин. Он подразумевает наличие не-
скольких сигнальных каналов, тогда как в узкополосном соединении такой канал
только один. Таким образом, теоретически 10-гигабитное соединение Ethernet,
работающее гораздо быстрее любого «широкополосного» соединения, широко-
полосным не является, так как в нем сигнальный канал только один.

Первоначально было предложено несколько технологий доступа под общим
именем x**DSL** (Digital Subscriber Line - **цифровая абонентская линия**) с пере-
менным значением х. Далее мы обсудим самую распространенную из них -
**ADSL** (Asymmetric DSL - **асимметричная цифровая абонентская линия**).
Работы над ADSL все еще продолжаются, и далеко не все стандарты на эту
технологию прописаны, поэтому некоторые детали со временем могут коррек-
тироваться. Впрочем, общая картина, скорее всего, останется неизменной. За до-
полнительными сведениями по ADSL обращайтесь к дополнительной литературе
[Summers, 1999, Vetter et al., 2000].

Почему обычные модемы работают так медленно? Да потому, что телефоны
были изначально предназначены для передачи голоса, и именно с учетом этой
цели сформировалась вся система телефонной связи. Передаче данных по теле-
фонным проводам уделялось слишком мало внимания. Пропускная способность
провода (он же - **абонентский канал**), связывающего абонентов с автоматиче-
ской телефонной станцией (АТС), традиционно ограничивалась специальным
фильтром. Фактическая пропускная способность абонентского канала во многом
зависит от его протяженности, но чаще всего (если протяженность не превышает
нескольких километров) достигает 1,1 МГц.

Наиболее распространенная схема предоставления услуг ADSL изображена
на рис. 2.35. Ее содержание сводится к удалению фильтра и разделению освобо-
дившегося спектра 1,1 МГц на 256 автономных каналов, по 4312,5 Гц каждый.
Канал 0 выделяется для голосовых данных. Каналы 1-5 не используются, за
счет чего устраняются взаимные помехи сигналов передачи голоса и данных. Из
оставшихся 250 каналов два выделяются для восходящей и нисходящей передачи

Напряжение

управляющих сигналов. По остальным каналам передаются пользовательские
данные. Таким образом, один ADSL-модем равноценен 250 обычным модемам.

**Голос Восходящий
поток**

**Нисходящий поток**

**Рис. 2.35.** Функционирование ADSL

В принципе, по каждому из оставшихся каналов можно пустить дуплекс-
ный поток данных, однако вспомогательные гармоники, перекрестные помехи
и другие физические эффекты не позволяют довести фактическую реализацию
технологии до теоретического уровня. Решение о том, в какой пропорции разде-
лить каналы на нисходящие и восходящие потоки, принимает поставщик услуги.
Технически возможно равное распределение этих каналов, но в большинстве
случаев 80-90 % выделяются на организацию нисходящего потока (обычно 32 ка-
нала выделяют для восходящего потока, а остальная часть - для нисходящего),
поскольку большинство пользователей принимают значительно больше данных,
чем отправляют. Именно по этой причине технология ADSL так успешна.

Качество передачи данных по каждому каналу постоянно отслеживается и при
необходимости корректируется, поэтому по скоростям каналы могут различаться.
Данные передаются в объеме до 15 бит/бод за счет сочетания амплитудной и фа-
зовой модуляции. Если, скажем, для передачи данных доступно 224 нисходящих
канала, а скорость передачи сигнала равна 4000 бод при 15 бит/бод, совокупная
пропускная способность нисходящего потока составляет 13,44 Мбит/с. На прак-
тике соотношение «сигнал/шум» не позволяет приблизиться к такому уровню,
но при небольшой удаленности от поставщика услуг и высоком качестве канала
скорость 4-8 Мбит/с вполне достижима.

Стандартная конфигурация оборудования ADSL изображена на рис. 2.36.
Согласно этой схеме, в помещении пользователя устанавливается **сетевое ин-
терфейсное устройство** (Network Interface Device, **NID**). Эта небольшая пласти-
ковая коробочка символизирует границу между собственностью пользователя
и собственностью телефонной компании. Рядом с NID (а иногда и в одном кор-
пусе с этим устройством) устанавливается **сплиттер** (разветвитель) - аналоговый
фильтр, разделяющий данные и сигналы на частоте 0-4000 Гц, применяемые
для передачи голоса. Поток данных направляется к ADSL-модему, а голосовые
сигналы - к телефону. ADSL-модем представляет собой процессор цифровых
сигналов, эмулирующий параллельную работу 250 обычных модемов на разных
частотах. Поскольку большинство ADSL-модемов выпускаются во внешнем ис-
полнении, их соединение с компьютером должно быть достаточно скоростным.
Обычно это требование удовлетворяется путем установки в компьютер платы
Ethernet и организации двухзвенного Ethernet-соединения с ADSL-модемом.
(Ethernet - распространенный и весьма доступный стандарт организации

локальных сетей.) Иногда ADSL-модем подключается к компьютеру через USB-
порт. В будущем следует ожидать появления специальных плат для соединения
с ADSL-модемом.

**Рис. 2.36.** Стандартная конфигурация оборудования ADSL

На противоположной стороне абонентского канала устанавливается другой
сплиттер, который отделяет голосовые сигналы и перенаправляет их на обыч-
ный телефонный коммутатор. Сигналы с частотой свыше 26 кГц передаются
**мультиплексору доступа к цифровой абонентской линии** (Digital Subscriber
Line Access Multiplexer, **DSLAM**). После преобразования цифровых сигналов
в поток битов происходит формирование пакетов, которые затем направляются
поставщику услуг.

Кабельный Интернет

Многие телевизионные компании предлагают пользователям возможность до-
ступа в Интернет по кабельным сетям. Эта технология несколько отличается от
ADSL, поэтому ее стоит рассмотреть отдельно. Во владении каждого оператора
кабельного телевидения, помимо центрального офиса, есть ряд головных узлов
(помещений с электронным оборудованием), рассредоточенных по территории
города. К центральному офису все они подключены широкополосным или опто-
волоконным кабелем.

От каждого головного узла к конечным потребителям отходит один или не-
сколько кабелей. Чтобы к такому кабелю можно было подключиться, он должен
проходить рядом с помещениями, в которых находятся пользователи. При этом
к одному и тому же кабелю подключаются сотни пользователей. Как правило,

пропускная способность такого кабеля составляет около 750 МГц. Как видно,
концептуальное отличие кабельного доступа от технологии ADSL заключается
в отсутствии индивидуального канала, подведенного к офису поставщика услуг.
Впрочем, на практике выгоды от наличия собственного канала пропускной
способностью 1,1 МГц, с одной стороны, и общего с еще четырьмя сотнями
пользователей канала совокупной пропускной способностью 200 МГц, с другой,
примерно равноценны (объясняется это тем, что в каждый отдельно взятый мо-
мент времени из 400 пользователей в сети находятся не более половины). Более
того - глубокой ночью кабельный Интернет работает значительно быстрее, чем
днем, в то время как скорость передачи данных по каналу ADSL в течение суток
одинакова. Логика такова: чтобы получить оптимальный доступ в Интернет по
кабельному каналу, нужно жить либо в очень богатом районе (где дома нахо-
дятся на большом расстоянии друг от друга, а, следовательно, к одному кабелю
подключено не так уж много пользователей), либо в очень бедном (где никто не
может себе позволить приобрести такую услугу).

Поскольку к одному кабелю подключаются многочисленные пользователи,
актуальной проблемой является временная и частотная регламентация потребле-
ния пропускной способности. Чтобы понять, как эта проблема решается, при-
дется сделать небольшой экскурс в технологию кабельного телевидения. В США
для вещания кабельных каналов выделен частотный диапазон 54-550 МГц (из
него, правда, исключается диапазон 88- 108 МГц, предназначенный для FM-
радиостанций). Каждый канал занимает 6 МГц (включая защитные полосы,
предотвращающие взаимные помехи смежных каналов). В Европе нижний по-
рог кабельного диапазона - 65 МГц, а каналы занимают по 6-8 МГц (за счет
этого обеспечивается повышенное разрешение по стандартам PAL/SECAM); во
всем остальном схема распределения частот аналогична американской. В обоих
случаях нижняя часть диапазона не используется для передачи телевизионных
сигналов.

Пытаясь реализовать технологию доступа в Интернет по кабелю, операторы
столкнулись с двумя проблемами:

1. Как предотвратить помехи при одновременной передаче данных и телевизионного сигнала?

2. Как организовать двунаправленный трафик при однонаправленных усилителях?

Выбранные решения таковы. Современные кабели работают на частоте значи-
тельно выше 550 МГц, достигая 750 МГц и более. Восходящие (то есть направ-
ленные от пользователя к головному узлу) каналы занимают диапазон 5-42 МГц
(в Европе он чуть выше), в то время как для передачи нисходящего (от головного
узла к пользователю) трафика используются высокие частоты (рис. 2.37).

Обратите внимание: поскольку телевизионные сигналы передаются исключи-
тельно в нисходящем направлении, восходящие усилители могу работать только
в диапазоне 5-42 МГц, а нисходящие - в диапазоне от 54 МГц и выше. Таким
образом, пропускная способность двух направлений оказывается асимметричной,
поскольку восходящий диапазон значительно меньше нисходящего. Впрочем, это
обстоятельство не сильно беспокоит операторов кабельного телевидения, так как
и трафик по большей части передается *к* пользователю, а не *от* него. В конце

5 4254 88

|  |  |  |  |
| --- | --- | --- | --- |
| о/ |  |  | Г |
| Восходящаяпередачаданных |  | ТВ | FM | ТВ | Нисходящая передачаданных |

Частоты для нисходящей передачи

**Рис. 2.37.** Распределение частот в стандартной системе кабельного телевидения
с возможностью доступа в Интернет

концов, телефонные компании тоже успешно предоставляют услугу DSL с асим-
метричным доступом, хотя никаких технических ограничений на восходящий
трафик здесь не существует.

Для подключения к Интернету конечного пользователя применяются ка-
бельные модемы. Это устройства с двумя интерфейсами - для подключе-
ния к компьютеру, с одной стороны, и к кабельной сети, с другой. Интерфейс
«компьютер- кабельный модем» несложен - как и в случае с ADSL, для передачи
данных организуется миниатюрная сеть Ethernet. В будущем кабельные модемы,
скорее всего, будут производиться в виде плат, устанавливаемых в системный
блок, - точно так же, как это произошло со старыми телефонными модемами.

На противоположной стороне устанавливается более сложное оборудование.
Стандарты кабельных соединений - это тема из области радиотехники, поэтому
в нашем контексте ее подробное изложение неуместно. Единственное, что стоит
отметить, так это беспрерывную работу кабельных модемов - в этом отношении
они напоминают ADSL-модемы. Соединение устанавливается и поддерживается
постоянно и прерывается только при отключении источника питания - связано
это с тем, что операторы кабельных сетей не взимают повременную плату за
свои услуги.

Чтобы лучше понять, как работают кабельные модемы, рассмотрим последо-
вательность операций при подсоединении и включении модема. Сначала модем
просматривает содержимое нисходящих каналов в поисках специального пакета,
который с определенной регулярностью отправляется с головного узла и со-
держит системные параметры для недавно подключенных модемов. Обнаружив
таковой, модем объявляет о своем присутствии в одном из восходящих каналов.
Далее головной узел назначает модему определенные восходящие и нисходящие
каналы. Впоследствии, если головной узел сочтет необходимым сбалансировать
нагрузку, модему могут быть назначены другие каналы.

Затем модем определяет расстояние до головного узла путем отправки ему
специального пакета и вычисления времени ответа. Этот процесс называется
**калибровкой** (ranging). Зная расстояние до головного узла, модем может соот-
ветствующим образом скорректировать работу восходящих каналов. Дело в том,
что восходящий поток данных подразделяется на временные интервалы, или
**мини-слоты** (minislots). Каждый восходящий пакет должен уместиться в рамках

одного или нескольких последовательных мини-слотов. Головной узел регулярно
высылает оповещения о начале новых циклов мини-слотов, но, так как модемы
находятся на разных расстояниях от головного узла, они получают эти оповеще-
ния в разное время. В то же время, зная, на каком расстоянии от головного узла
он находится, модем может вычислить фактическое время начала мини-слота.
Длина мини-слота определяется характеристиками конкретной сети. Полезная
нагрузка одного мини-слота обычно составляет 8 байт.

В ходе инициализации головной узел привязывает каждый модем к опреде-
ленному мини-слоту, в результате тот получает возможность отравлять запросы
на предоставление пропускной способности. Обычно к одному и тому же мини-
слоту привязываются несколько модемов, за счет чего формируется механизм
конкуренции. Перед отправкой пакета с компьютера в сеть он передается мо-
дему, который затем запрашивает соответствующее количество мини-слотов.
Если запрос удовлетворяется, головной узел отсылает по нисходящему каналу
подтверждение, в котором указывает зарезервированные для передачи пакета
мини-слоты. Далее, начиная с первого зарезервированного мини-слота, начи-
нается отправка. Запросы на передачу дополнительных пакетов встраиваются
в специальное поле заголовка.

Если в условиях конкуренции за запрошенный мини-слот модем не полу-
чает подтверждение, он ждет случайный интервал времени и повторяет запрос.
С каждой неуспешной попыткой время ожидания удваивается, что способствует
разряжению интенсивного трафика.

Нисходящие каналы управляются по-другому. Во-первых, при нисходящей
передаче отправитель всего один - головной узел. Следовательно, состязатель-
ность отсутствует, равно как и необходимость выделения мини-слотов, которые,
по существу, есть не что иное, как средство статистического мультиплексирования
с разделением времени. Во-вторых, нисходящий трафик обычно значительно ин-
тенсивнее, чем восходящий, поэтому он передается в пакетах по 204 байта. В состав
пакета, помимо полезной нагрузки в 184 байта, входит код исправления ошибок
Рида- Соломона и некоторые другие служебные поля. Этот размер пакета выбран
в целях совместимости с цифровым телевидением формата MPEG-2 - в итоге
каналы нисходящей передачи телевизионного сигнала и данных форматируются
единообразно. Логическая схема этих соединений изображена на рис. 2.38.

Впрочем, вернемся к процедуре инициализации модема. После калибровки,
получения восходящего и нисходящего каналов и назначения минислотов модем
может приступать к передаче пакетов. Пакеты отправляются на головной узел,
с которого они по выделенному каналу уходят в центральный офис оператора
кабельного телевидения, а от него - к поставщику услуг Интернета (Internet
Service Provider, ISP), в качестве которого может выступать и сам оператор.
Первый пакет, отправляемый поставщику услуг, содержит запрос на предоставле-
ние в динамическом режиме сетевого адреса (IP-адреса). Другой запрос в составе
этого пакета касается точного времени дня.

На следующем этапе решаются вопросы безопасности. По одному кабелю
свои данные передают множество пользователей, поэтому при большом желании
пользователь может организовать перехват всего проходящего трафика. Чтобы
не допустить со стороны соседей коллективного слежения друг за другом, весь
трафик, в каком бы направлении он ни отправлялся, в обязательном порядке

Пакет Восходящий канал с состязательностью:

9 Мбит/с по технологии QPSK при
передаче данных в рамках 8- байтных
мини-слотов

**Рис. 2.38.** Стандартная организация восходящих и нисходящих каналов в США.
Технология QAM-64 (квадратурная амплитудная модуляция) допускает передачу
со скоростью 6 бит/Г ц, но работает только на высоких частотах. Технология QPSK
(квадратурная фазовая модуляция) действует на низких частотах, зато максимальная
скорость передачи составляет 2 бит/Гц

шифруется. Поэтому в процессе инициализации, помимо прочего, определяют-
ся ключи шифрования. Казалось бы, провести согласование секретного ключа
между модемом и головным узлом под бдительным оком тысяч пользователей
невозможно. На самом деле, это не так - для определения ключа шифрования
задействуется алгоритм Диффи-Хелмана [Kaufman et al., 2002].

Наконец, модем регистрируется в сети и сообщает по защищенному каналу
свой уникальный идентификатор. На этом процесс инициализации заканчива-
ется - пользователь может начинать работу.

Наш обзор кабельных модемов получился довольно сжатым. За подроб-
ностями обратитесь к дополнительной литературе [Adams and Dulchinos, 2001;
Donaldson and Jones, 2001; Dutta-Roy, 2001].

**Цифровые фотокамеры**

Все больше компьютерные технологии внедряются в сферу цифровой фото-
графии - уже сейчас цифровые фотокамеры вполне правомерно рассматривать
как один из видов компьютерных периферийных устройств. Давайте вкратце
рассмотрим принцип их работы. Все камеры снабжены объективом, с помощью
которого в задней части камеры формируется изображение объекта. В традици-
онной камере в качестве носителя скрытых изображений, которые формируются
в момент проникновения света, выступает фотопленка. Изображения проявля-
ются в лаборатории за счет воздействия определенных химических реактивов.
Принцип действия цифровой камеры аналогичен за одним исключением -
вместо пленки носителем изображения становится прямоугольная матрица
светочувствительных **устройств с зарядовой связью** (Charge-Coupled Devices,
**CCD**). (Некоторые цифровые камеры действуют на основе технологии КМОП,
но вариант с CCD более распространен.)

При попадании на устройство CCD света устройство получает электрический
заряд. Чем больше света, тем существеннее изменение заряда. Заряд считывается

аналогово-цифровым преобразователем в виде целого числа от 0 до 255 (в каме-
рах низкой ценовой категории) или от 0 до 4095 (на цифровых однообъективных
зеркальных фотоаппаратах). Соответствующая схема изображена на рис. 2.39.

Каждое устройство CCD, независимо от падающего на него света, на выходе
генерирует единственное значение. Для формирования цветных изображений
устройства CCD объединяются в группы из четырех элементов. Поверх группы
размещается **фильтр Байера** (Bayer filter), который делает одно устройство CCD
чувствительным к красному цвету, другое - к синему, а два оставшихся - к зеле-
ному. Наличие двух зеленых элементов объясняется двумя факторами: во-первых,
это удобнее, а во-вторых, человеческий глаз воспринимает зеленый цвет лучше,
чем синий и красный. Если производитель цифровой камеры заявляет, что ее
разрешение равно 6 млн пикселов, знайте - это неправда. В ней 6 млн устройств
CCD, которые в совокупности формируют 1,5 млн пикселов. При таком раз-
решении изображение считывается в виде матрицы 2828 X 2121 (в недорогих
камерах) или 3000 X 2000 (в однообъективных зеркальных фотоаппаратах) пик-
селов. Дополнительные пикселы генерируются путем программной интерполяции.

При нажатии кнопки открытия затвора объектива программное обеспечение
камеры выполняет три операции: устанавливает фокус, определяет экспозицию
и проводит балансировку белого. Автоматическая фокусировка осуществляется
путем анализа высокочастотных данных изображения и выдвижения объектива
на предельную позицию в целях максимальной детализации. При определении
экспозиции сначала вычисляется интенсивность света, падающего на CCD, по-
сле диафрагма и выдержка корректируются таким образом, чтобы полученное

значение интенсивности пришлось на середину диапазона CCD. Балансировка
белого сводится к измерению спектра падающего света с целью последующей
цветокоррекции.

Далее изображение считывается с CCD и сохраняется в виде матрицы пик-
селов во встроенной оперативной памяти камеры. Профессиональные однообъ-
ективные зеркальные фотоаппараты, с которыми работают фотокорреспонденты,
могут в течение пяти секунд снимать по восемь кадров с высоким разрешением
в секунду; при этом объем встроенной оперативной памяти, в которой изобра-
жения размещаются перед последующей обработкой и постоянным хранением,
составляет около 1 Гбайт. В недорогих камерах оперативной памяти меньше, но
все равно вполне достаточно.

После создания снимка программное обеспечение проводит цветокоррекцию
на основе баланса белого, тем самым нейтрализуя избыток красного или синего
света (что имеет место, например, при фотографировании объекта, находящегося
в тени, а также при использовании вспышки). Затем выполняются алгоритмы
шумоподавления и корректировки дефектных устройств CCD. После этого (если
соответствующая функция включена) производится попытка повысить резкость
изображения - выполняется поиск краев и увеличение интенсивности градиента
вокруг них.

Наконец, изображение сжимается с целью уменьшения объема занимаемой им
памяти. Самый распространенный формат, применяемый для этих целей, - **JPEG**( Joint Photographic Experts Group - объединенная группа экспертов в области
фотографии). Он предусматривает двухмерное пространственное преобразова-
ние Фурье и удаление высокочастотных составляющих. Конечное изображение
оказывается весьма компактным, но мелкие детали утрачиваются.

По окончании обработки изображение записывается на постоянный носи-
тель, в качестве которого обычно выступает карта флэш-памяти или небольшой
съемный жесткий диск - так называемый **микродиск**. На обработку и запись
каждого изображения уходит несколько секунд.

Затем пользователь может подключить камеру к компьютеру - посредством,
например, кабеля USB или FireWire. Это позволяет перенести изображения из
памяти камеры на жесткий диск компьютера. При помощи специального про-
граммного обеспечения (например, редактора Adobe Photoshop) пользователь
может обрезать изображение, настроить яркость, контраст и баланс, увеличить
резкость или, наоборот, частично размыть изображение, удалить ненужные
элементы и наложить в произвольном сочетании фильтры. Удовлетворившись
результатом, пользователь волен распечатать изображения на цветном принте-
ре, разместить их в Интернете, а также записать на компакт-диск или DVD для
архивации или последующей печати.

По вычислительным мощностям, объему оперативной памяти и дискового
пространства, равно как и по сложности программного обеспечения, цифровые
однообъективные зеркальные фотоаппараты (Single-Lens Reflex, SLR) сопоставимы
с настольными системами двух- трехлетней давности. Помимо вышеперечисленных
операций, компьютер такого фотоаппарата должен обеспечивать взаимодействие
с процессором объектива и вспышки, обновлять изображение на жидкокристал-
лическом экране, не говоря уже о координации действий всех кнопок, колесиков,
индикаторов, дисплеев и прочих приспособлений в реальном времени.

**Коды символов**

У каждого компьютера есть набор символов, который он использует. Как мини-
мум, этот набор включает 26 прописных и 26 строчных букв1, цифры от 0 до 9,
а также некоторые специальные символы, в том числе пробел, точку, запятую,
минус и символ возврата каретки и т. д.

Для того чтобы передавать эти символы в компьютер, каждому из них при-
писывается номер, например, a = 1, b = 2, ..., z = 26, + = 27, - = 28. Представление
символа в виде целого числа называется **кодом символа**. Важно отметить, что
связанные между собой компьютеры должны поддерживать одни и те же коды
символов, иначе они не смогут обмениваться информацией. По этой причине
были разработаны стандарты. Здесь мы рассмотрим два самых важных из них.

ASCII

Один из двух широко распространенных кодов называется **ASCII** (American
Standard Code for Information Interchange - **американский стандартный код для
обмена информацией**). Каждый ASCII-символ содержит 7 бит, таким образом,
всего можно закодировать 128 символов. Коды от 0 до 1F (в шестнадцатеричной
системе счисления) соответствуют управляющим символам, которые не печата-
ются (табл. 2.4). Коды от 0 до 1F (в шестнадцатеричной записи) соответствуют
управляющим символам, которые не выводятся на печать. Коды от 128 до 255
не входят в кодировку ASCII; на IBM PC за ними были закреплены специаль-
ные символы (улыбающиеся лица и т. д.), которые до сих пор поддерживаются
большинством компьютеров.

**Таблица 2.4.** Таблица кодов для управляющих ASCII-символов

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Чис-****ло** | **Назва-****ние** | **Значение** | **Чис-****ло** | **Назва-****ние** | **Значение** |  |
| 0 | NUL | Нуль | 8 | BS | BackSpace (Отступ назад) |  |
| 1 | SOH | Start Of Heading (Началозаголовка) | 9 | HT | Horizontal Tab (Горизон-тальная табуляция) |  |
| 2 | STX | Start Of Text (Началотекста) | A | LF | Line Feed (Перевод строки) |  |
| 3 | ETX | End of Text (Конец тек-ста) | B | VT | Vertical Tab (Вертикальнаятабуляция) |  |
| 4 | EOT | End Of Transmission (Ко-нец передачи) | C | FF | From Feed (Перевод стра-ницы) |  |
| 5 | ENQ | ENQuiry (Запрос) | D | CR | Carriage Return (Возвраткаретки) |  |
| 6 | ACK | ACKnoligement (Под-тверждение приема) | E | SO | Shift Out (Переключе-ние на дополнительныйрегистр) |  |
| 7 | BEL | Bell (Звуковой сигнал) | F | SI | Shift In (Переключение настандартный регистр) |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Чис-****ло** | **Назва-****ние** | **Значение** | **Чис-****ло** | **Назва-****ние** | **Значение** |
| 10 | DLE | Data Link Escape (Сменаканала данных) | 18 | CAN | CANcel (Отмена) |
| 11 | DC1 | Device Control 1 (Управ-ление устройством 1) | 19 | EM | End of Medium (Конецносителя) |
| 12 | DC2 | Device Control 2 (Управ-ление устройством 2) | 1A | SUB | SUBstitute (Подстрочныйиндекс) |
| 13 | DC3 | Device Control 3 (Управ-ление устройством 3) | 1B | ESC | ESCape (Выход) |
| 14 | DC4 | Device Control 4 (Управ-ление устройством 4) | 1C | FS | File Separator (Разделительфайлов) |
| 15 | NAK | Negative AcKnolidgement(Неподтверждение приема) | 1D | GS | Group Separator (Раздели-тель группы) |
| 16 | SYN | SYNcronous idle (Пауза) | 1E | RS | Record Separator (Раздели-тель записи) |
| 17 | ETB | End of Transmission Block(Конец блока передачи) | 1F | US | Unit Separator (Раздели-тель модуля) |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Чис-****ло** | **Сим-****вол** | **Чис-****ло** | **Сим-****вол** | **Чис-****ло** | **Сим-****вол** | **Чис-****ло** | **Сим-****вол** | **Чис-****ло** | **Сим-****вол** | **Чис-****ло** | **Сим-****вол** |
| 20 | (про-бел) | 30 | 0 | 40 | @ | 50 | P | 60 |  | 70 | P |
| 21 | ! | 31 | 1 | 41 | A | 51 | Q | 61 | a | 71 | Q |
| 22 | “ | 32 | 2 | 42 | B | 52 | R | 62 | b | 72 | R |
| 23 | # | 33 | 3 | 43 | C | 53 | S | 63 | c | 73 | S |
| 24 | $ | 34 | 4 | 44 | D | 54 | T | 64 | d | 74 | T |
| 25 | % | 35 | 5 | 45 | E | 55 | U | 65 | e | 75 | U |
| 26 | & | 36 | 6 | 46 | F | 56 | V | 66 | f | 76 | V |
| 27 | ’ | 37 | 7 | 47 | G | 57 | W | 67 | g | 77 | W |
| 28 | ( | 38 | 8 | 48 | H | 58 | X | 68 | h | 78 | X |
| 29 | ) | 39 | 9 | 49 | I | 59 | Y | 69 | i | 79 | Y |
| 2A | \* | 3A |  | 4A | J | 5A | Z | 6A |  | 7A | Z |
| 2B | + | 3B | ; | 4B | K | 5B | [ | 6B | k | 7B | { |
| 2C | , | 3C | < | 4C | L | 5C | \ | 6C | l | 7C | 1 |
| 2D | - | 3D | = | 4D | M | 5D | ] | 6D | m | 7D | } |
| 2E |  | 3E | > | 4E | N | 5E | л | 6E | n | 7E | ~ |
| 2F | / | 3F | ? | 4F | O | 5F |  | 6F | o | 7F | DEL |

Многие управляющие ASCII-символы предназначены для передачи данных.
Например, послание может состоять из символа начала заголовка SOH (Start
of Header), самого заголовка, символа начала текста STX (Start of Text), самого
текста, символа конца текста ETX (End of Text) и, наконец, символа конца пере-
дачи EOT (End of Transmission). Однако на практике послания, отправляемые по
телефонным линиям и сетям, форматируются по-другому, так что управляющие
ASCII-символы для передачи практически не используются.

Печатные ASCII-символы включают буквы верхнего и нижнего регистров,
цифры, знаки пунктуации и некоторые математические символы.

Unicode

Компьютерная промышленность развивалась преимущественно в США, что
привело к появлению кода ASCII, более подходящего для английского языка,
чем для других языков. Во французском языке есть надстрочные знаки, в не-
мецком - умляуты и т. д. В некоторых европейских языках есть несколько букв,
которых нет в наборе ASCII-символов. Некоторые языки имеют совершенно
другой алфавит (например, русский или арабский), а у некоторых вообще нет
алфавита (например, китайский). Компьютеры распространились по всему свету,
и поставщики программного обеспечения хотят реализовывать свою продукцию
не только в англоязычных, но и в тех странах, где большинство пользователей
не говорят по-английски и нужен другой набор символов.

Первой попыткой расширения кода ASCII стал стандарт IS 646, который
добавлял к набору ASCII-символов еще 128 символов, в результате чего полу-
чился 8-разрядный набор под названием **Latin-1**. Добавлены были в основном
латинские буквы со штрихами и диакритическими знаками. Следующей попыт-
кой был стандарт IS 8859, который ввел понятие **кодовой страницы**. Кодовая
страница - набор из 256 символов для определенного языка или группы языков,
в IS 8859-1 это набор Latin-1. Стандарт IS 8859-2 включает славянские языки
с латинским алфавитом (например, чешский, польский и венгерский), стандарт
IS 8859-3 описывает символы турецкого, мальтийского и галисийского языков,
эсперанто и т. д. Главным недостатком такого подхода является то, что программ-
ное обеспечение должно контролировать, с какой именно кодовой страницей оно
имеет дело, при этом смешивать языки недопустимо. К тому же эта система не
охватывает японский и китайский языки.

Группа компьютерных компаний разрешила эту проблему, создав новую
систему кодирования под названием **Unicode**, и объявила эту систему междуна-
родным стандартом (IS 10646). Unicode поддерживается некоторыми языками
программирования (например, Java), некоторыми операционными системами
(например, Windows NT) и многими приложениями. Вероятно, эта система будет
распространяться по всему миру.

Основная идея Unicode - приписать каждому символу единственное постоян-
ное 16-разрядное значение, которое называется **кодовым пунктом**. Многобайтные
символы и символы-заменители не используются. Поскольку каждый символ
состоит из 16 бит, писать программное обеспечение гораздо проще.

Так как Unicode-символы состоят из 16 бит, всего получается 65 536 кодо-
вых пунктов. Поскольку во всех языках мира в общей сложности около 200 000
символов, кодовые указатели являются очень дефицитным ресурсом, который

нужно распределять с большой осторожностью. Около половины кодов уже
распределено, и консорциум, разработавший Unicode, постоянно рассматривает
предложения по распределению оставшейся части. Чтобы ускорить принятие
системы Unicode, консорциум использовал набор Latin-1 для кодов от 0 до 255,
обеспечивающий простое преобразование ASCII- в Unicode-символы. Во избе-
жание излишней растраты кодов каждый диакритический знак имеет собствен-
ный код, а объединение диакритического знака с той или иной буквой должно
осуществляться программным обеспечением.

Вся совокупность кодов разделена на блоки, каждый блок содержит 16 кодов.
Каждый алфавит в Unicode имеет ряд последовательных зон. Приведем некото-
рые примеры (в скобках указано число задействованных кодов): латынь (336),
греческий (144), русский (256), армянский (96), иврит (112), деванагари (128),
гурмукхи (128), ория (128), телугу (128) и каннада (128). Отметим, что каждому
из этих языков приписано больше кодов, чем в нем есть букв. Это сделано от-
части потому, что во многих языках у каждой буквы есть несколько вариантов
написания. Например, каждая буква в английском языке представлена в двух
вариантах: там есть строчные и ПРОПИСНЫЕ буквы. В некоторых языках бук-
вы имеют три или более формы написания, выбор конкретного варианта зависит
от того, где находится буква: в начале, конце или середине слова.

Кроме того, некоторые коды были приписаны диакритическим знакам (112),
знакам пунктуации (112), подстрочным и надстрочным знакам (48), знакам
валют (48), математическим символам (256), геометрическим фигурам (96),
рисункам (192).

Нужны также символы для китайского, японского и корейского языков.
Сначала идут 1024 фонетических символа (например, катакана и бопомофо),
затем иероглифы, используемые в китайском и японском языках (20 992), потом
слоги корейской азбуки хангыль (11 156).

Чтобы пользователи могли создавать новые символы для особых целей, су-
ществуют еще 6400 кодов.

Хотя система Unicode разрешила многие проблемы, связанные с интерна-
ционализацией, она все же не позволила разрешить абсолютно все проблемы.
Например, латинский алфавит упорядочен, а иероглифы - нет, поэтому програм-
ма для английского языка может расположить слова «cat» и «dog» по алфавиту,
сравнив значение кодов первых букв, а программе для японского языка нужны
дополнительные таблицы, чтобы можно было вычислять, в каком порядке рас-
положены символы в словаре.

Еще одна проблема состоит в том, что постоянно появляются новые слова.
50 лет назад никто не говорил об апплетах, киберпространстве, гигабайтах,
лазерах, модемах, смайликах или видеопленках. Появление новых слов в ан-
глийском языке новых кодов не требует, а вот в японском они нужны. Помимо
новых терминов, необходимо также добавить по крайней мере 20 000 новых
имен собственных и географических названий (в основном китайских). Шрифт
Брайля, вероятно, тоже должен быть задействован. В наличии тех или иных осо-
бых символов заинтересованы и представители различных профессиональных
сообществ. Консорциум Unicode рассматривает все новые предложения и вы-
носит по ним решения.

В системе Unicode используется один и тот же код для символов, которые
выглядят почти одинаково, но имеют несколько значений или пишутся немного
по-разному в китайском и японском языках (как если бы английские текстовые
процессоры всегда писали слово «blue» как «blew», потому что они произносятся
одинаково). Одни считают такой подход оптимальным для экономии скудного
запаса кодов, другие рассматривают его как англосаксонский культурный импе-
риализм (а вы думали, что назначение символам 16-разрядных кодов не носит
политического характера?). Дело усложняется тем, что полный японский сло-
варь содержит 50 000 иероглифических знаков (не считая знаков, используемых
только в именах собственных), поэтому при наличии 20 992 кодов приходится
делать выбор и чем-то жертвовать. Далеко не все японцы считают, что консор-
циум компьютерных компаний, даже если некоторые из них японские, идеально
подходит для принятия решений о том, чем именно нужно жертвовать.

В общем, оказалось, что 65 536 кодовых пунктов недостаточно для всех по-
требностей, поэтому в 1996 году были определены дополнительные шестнадцать
16-разрядных **плоскостей**, в результате чего общее количество символов увели-
чилось до 1 114 112.

UTF-8

Стандарт Unicode был лучше ASCII, но со временем в нем тоже возникла не-
хватка кодовых пунктов. Кроме того, он требовал 16 бит для представления
«чистого» ASCII-текста, что было расточительно. Для решения этих проблем
была разработана новая схема кодирования **UTF-8 UCS Transformation Format**.
Сокращение UCS в названии означает «Universal Character Set» - по сути это
Unicode. Коды UTF-8 имеют переменную длину от 1 до 4 байт, и позволяют
кодировать до двух миллиардов символов. Этот способ кодировки сейчас до-
минирует в World Wide Web.

Одно из преимуществ UTF-8 заключается в том, что коды от 0 до 127 ис-
пользуются для представления ASCII-символов. Таким образом, одному сим-
волу соответствует один байт (вместо двух байтов в Unicode). Для символов,
не входящих в набор ASCII, старший бит первого байта устанавливается в 1;
это означает, что за ним следуют один и более дополнительных байтов. Всего
используются шесть разных форматов, представленных в табл. 2.5. Биты с по-
меткой «d» являются битами данных.

**Таблица 2.5.** Схема кодирования UTF-8

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Биты** | **Байт 1** | **Байт 2** | **Байт 3** | **Байт 4** | **Байт 5** | **Байт 6** |
| 7 | 0ddddddd |  |  |  |  |  |
| 11 | 110ddddd | 10dddddd |  |  |  |  |
| 16 | 1110dddd | 10dddddd | 10dddddd |  |  |  |
| 21 | 11110ddd | 10dddddd | 10dddddd | 10dddddd |  |  |
| 26 | 111110dd | 10dddddd | 10dddddd | 10dddddd | 10dddddd |  |
| 31 | 1111110x | 10dddddd | 10dddddd | 10dddddd | 10dddddd | 10dddddd |

UTF-8 обладает рядом преимуществ перед Unicode и другими схемами.
Прежде всего, если в программе или документе используются только символы

из набора ASCII, каждый символ может быть представлен 8 битами. Во-вторых,
первый байт каждого символа UTF-8 однозначно определяет порядок байтов
в символе. В-третьих, дополнительные байты символа UTF-8 всегда начинаются
с префикса 10, который никогда не встречается в начальном байте; соответствен-
но код становится самосинхронизирующимся. В частности, при возникновении
ошибки передачи данных или памяти всегда возможно перейти к началу следую-
щего (неповрежденного) символа.

Обычно UTF-8 используется для кодирования только 17 плоскостей Unicode,
хотя схема позволяет представить много более 1 114 112 кодовых пунктов. Но
если антропологи найдут где-нибудь в Новой Гвинее племя с неизвестным
языком (или мы вступим в контакт с инопланетными цивилизациями), UTF-8
успешно справится с представлением их алфавитов или идеограмм.