问题

互联网边缘部分和核心部分的作用？
网络的体系结构是什么？
五层的体系结构是哪五层？TCP/IP 采用了几层？OSI 采用了几层
互联网采用的是什么协议？
常见的三种网络是哪三种？
什么是 ISP，它有什么作用
互联网的两种通信模式是什么?
什么是分组？什么是分组交换？存储转发技术是怎样的？
什么是接入网？
带宽的两种含义？

回答

边缘部分包括主机，作用是进行信息处理；核心部分包括路由器等，作用是按存储转发方式进行分组交换

网络的各层及其协议的集合

物理层、数据链路层、网络层、运输层、应用层。TCP/IP 采用了四层结构，将物理层

和数据链路层合并为了网络接口层。OSI 采用了七层，它将应用层进一步分为了 3 层。

TCP/IP 系列协议

有线电视网络、电信网络、计算机网络

ISP 是互联网服务提供商（国内三大运营商），它们掌握着很多 IP 地址与互联网基础设施。用户向 ISP 缴费以获得 IP 地址使用权。

服务器-客户端模式和对等模式（P2P）

分组时互联网中传送的基本单元，将分段的报文加上一个首部就构成了一个分组。存储转发技术即路由器收到分组后先暂时存储，然后检查首部，查找转发表，将分组发给下一个合适的路由器，一步步到达目的主机。

用户接入到互联网的网路

一种时频率上的范围，另一种是某一信道单位时间内可传输的最高数据率。

问题

网络中的时延的组成部分

高速链路的高速体现在哪部分时延上？

互联网有哪些应用层协议

运输层的两个主要协议。运输层的服务对象是什么？

网络层的主要协议是什么。网络层的服务对象是什么？

路由器会处理几层？

网络层和链路层的控制信息的影响范围差异

回答

发送时延、传播时延、处理时延、排队时延

高速链路的发送时延较短。光纤就是发送数据比较快，发送时延短。

DNS 协议、HTTP 协议、SMTP 协议、FTP 协议、P2P 协议

TCP 和 UDP 协议。运输层服务对象是主机中的进程

IP 协议。网络层服务对象是不同的主机

三层：最高到网络层

网络层的控制信息用于从源主机到目的主机的整个网络。链路层的控制信息仅用于两个相邻路由器之间。

第 1 章 概述

本章最重要的内容：

互联网边缘部分和核心部分的作用，什么是分组交换？

计算机网络的性能指标有哪些

计算机网络分层次的体系结构是怎样的？什么是协议和服务？

本章的重要概念

互联网采用**存储转发的分组交换技术**和**三层 ISP 结构**
互联网按工作方式划分为**边缘部分**和**核心部分**：

主机在边缘部分，**作用**是**进行信息处理**。

路由器在核心部分，**作用**是**按存储转发方式进行分组交换**。

计算机网络最常用的 **7 个性能指标：速率，带宽，吞吐量，时延，时延带宽积，往返时间，信道利用率**。
协议是为进行网络中的数据交换建立的**规则**。计算机网络的各层和其协议的集合称为**网络的体系结构**。
五层的体系结构包括：**应用层、运输层、网络层、数据链路层、物理层**。运输层最重要的协议是 **TCP** 和 **UDP** 协议，网络层最重要的协议是 **IP** 协议。
计算机网络（简称**网络**）把许多计算机连接在一起。**互连网**把许多网络连接在一起，是**网络的网络**。
internet 是互连网，通用名词，泛指网络。**Internet** 是**互联网**，专用名词，特指全球最大的互连网。
互联网采用 **TCP/IP 协议族**作为通信规则，前身是美国的 ARPANET。

计算机通信是计算机中的**进程**（即运行着的程序）之间的通信。网络的通信方式是**客户-服务器方式**和**对等连接方式**。
客户和服务器都是通信中所涉及的应用进程。客户是服务请求方，服务器是服务提供方。
计算机网络按作用范围分为**广域网（WAN）**，城域网，**局域网（LAN）**，个人区域网（PAN）

1.1 计算机网络在信息时代中的作用

常见的网络有三大类：**电信网络、有线电视网络、计算机网络**。其中计算机网络是核心。
互联网即因特网，是 Internet 的译名（注意不是 internet）
互联网是由数量极大的各种计算机网络互连起来的。
互联网的两个基本特点：连通性、共享
互联网+ 的含义：互联网 + 各个传统行业

1.2 互联网概述

1.2.1 网络的网络

计算机网络由多个结点和连接结点的链路组成。结点可以是计算机、集线器、交换机或路由器等。
网络之间可以通过路由器互连起来构成网络的网络，称为互连网。
网络把许多计算机连接在一起，互连网则把许多网络通过路由器连接在一起。与网络相连的计算机称为主机。

1.2.2 互联网基础结构发展的三个阶段

internet 是互连网，通用名词，泛指网络。互连网之间的通信协议可以任意选择，不一定是

TCP/IP
Internet 是**互联网**，专用名词，特指全球最大的互连网。互联网采用 **TCP/IP 协议族**作为通信规则，前身是美国的 ARPANET。

第三阶段的互联网

现在的互联网是**多层次 ISP 结构**：分为主干 ISP、地区 ISP、本地 ISP。
ISP 即互联网服务提供商，中国电信、中国联通、中国移动都是 ISP。
上网就是指接入到互联网。主机必须有 IP 地址才能上网。
ISP 从互联网管理机构申请到很多 IP 地址，同时拥有通信线路及路由器等联网设备。
用户向 ISP 交纳费用获得所需 IP 地址的使用权，然后就可以通过该 ISP 接入互联网。
互联网由全球无数的 ISP 所共同拥有。
万维网（WWW）是基于互联网开发的一种信息共享服务，浏览网址一般使用的就是万维网，而邮件等就没有用到万维网。

1.2.3 互联网的标准化工作

所有的互联网标准都以 RFC 文档的形式发表在互联网上。互联网上有很多 RFC 文档，但只有少部分是互联网标准

1.3 互联网的组成

互联网按工作方式划分为边缘部分和核心部分：

边缘部分：由连接到互联网的主机组成，作用是进行信息处理。

核心部分：由大量网络和连接网络的路由器组成，作用是按存储转发方式进行分组交换，为边缘部分提供通信服务。

1.3.1 互联网的边缘部分

主机又称端系统，个人电脑、摄像头、手机等都属于端系统。
边缘部分利用核心部分提供的服务进行通信，一般称为计算机之间通信。
计算机之间的通信实际上是计算机 A 上某个进程和计算机 B 上另一个进程之间的通信。
通信方式主要有两类：客户-服务器方式、对等方式（P2P）。

1.3.2 互联网的核心部分

核心部分最重要的功能是**分组交换**，主要组件是**路由器**。
路由器是实现分组交换的关键构件，用来转发分组。

分组交换

要发送的整块数据称为一个**报文**。将报文分为多个数据段，每个数据段加上一个**首部（包头）**构成一个**分组（包）**。
分组交换采用**存储转发技术**。路由器收到分组后，先暂时存储，检查首部，查找转发表，然后按照首部中的目的地址，找到合适的接口转发给下一个路由器，这样一步步交付给最终的目的主机。

首部中主要包含着目的地址、源地址等控制信息。
分组是在互联网中传送的数据单元。

三种交换方式

电路交换：有三个步骤：建立连接（占用通信资源，一条专用的物理通路）、通话（一直占用通信资源）、释放连接（归还通信资源）

缺点：在通话的全部时间内始终占用端到端的通信资源。

应用：电话使用的就是电路交换。而互联网数据因为其突发性，使用电路交换的话效率很低。

报文交换：整个报文先传送到相邻结点，全部存储下来后查找转发表，转发到下一个结点。

分组交换：单个分组传送到相邻结点，存储下来后查找转发表，转发到下一个结点。

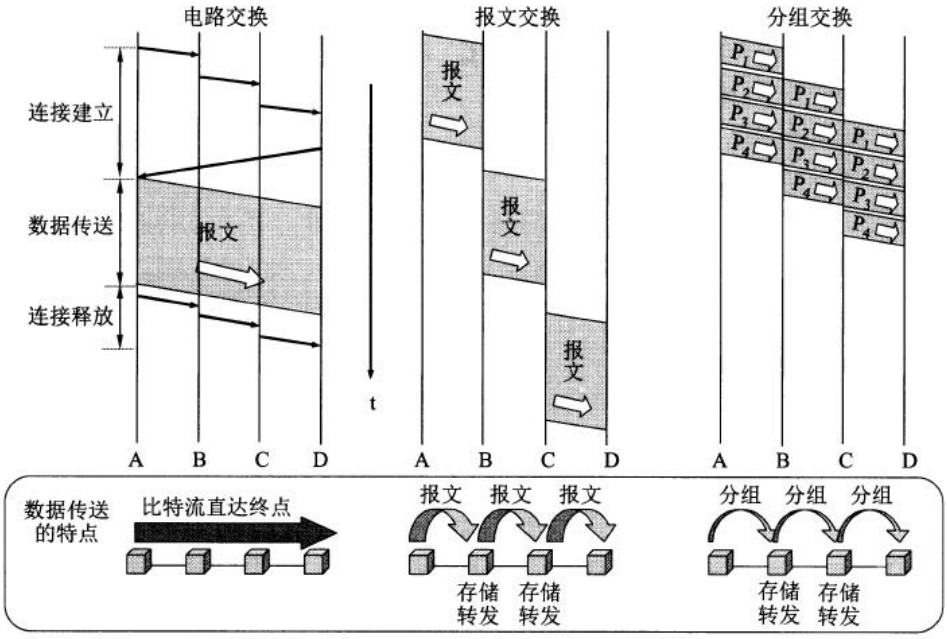


图 1-13　三种交换的比较。电路交换；报文交换；分组交换，P_1～P_4 表示 4 个分组

1.5 计算机网络的类别

1.5.1 计算机网络的定义

计算机网络并非专门用来传送数据，还支持很多其他应用。

1.5.2 几种不同类别的计算机网络

广域网 WAN：互联网的核心部分，连接广域网各结点的一般使高速链路。

城域网 MAN：很多城域网采用的是以太网技术，常并入局域网进行讨论

局域网 LAN：一般通过高速通信线路相连，地理范围较小。校园网、企业为一般为局域网

个人区域网 PAN：范围很小，一般在 10m 左右。

接入网

接入网 AN 是用来把用户接入到互联网的网络。
接入网既不属于核心部分，也不属于边缘部分，而是位于端系统到互联网中第一个路由器之间的一种网络。
宽带接入网就是一种接入网。

1.6 计算机网络的性能

1.6.2 计算机网络的性能指标

常用的 7 个性能指标：

速率
带宽
吞吐量
时延
时延带宽积
往返时间
利用率

速率

速率是计算机网络中最重要的性能指标。
速率是数据的传送速率，单位是 **bit/s**，有时也写为 **bps**。
1G 速率实际上是 1Gbit/s 而不是 1Gbyte/s。提到网络速率时，一般指的是**额定速率**或**标称速率**。

带宽

带宽有**两种意义**：

某个信号具有的频带宽度，是一个赫兹范围。表示某信道允许通过的信号频带范围就称为该信道的带宽。是频域称谓

计算机网络中，带宽表示网络中某通道传送数据的能力，网络带宽表示单位时间内网络中的某信道所能通过的最高数据率，单位是 **bit/s**。是时域称谓。

吞吐量

吞吐量表示单位时间内通过某个网络的**实际**的数据量。经常用于对现实世界中的网络的一种测量。
吞吐量受网络的带宽或额定速率的限制。

时延

时延是数据从网络的一端传送到另一端所需的时间。
网络中的时延由以下几个不同的部分组成：

发送时延：主机或路由器发送数据帧所需的时间，即从发送数据帧的第一个比特到最后一个比特发送完毕的时间，也叫传输时延。

计算公式：**发送时延 = 数据帧长度/发送速率**。

发送速率越快，发送时延越低。

传播时延：传播时延是电磁波在信道中传播一定的距离花费的时间。

计算公式：**传播时延=信道长度 / 电磁波在信道上的传输速率。**

电磁波在网络中的传输速率比真空中低，在铜线中为 2.3 亿米每秒，在光纤中是 2 亿米每秒。

距离越长，传输时延越长。

处理时延：主机或路由器在收到分组时要花费时间处理数据，如分析分组的首部，从分组中提取数据部分，差错检验等等。

排队时延：分组在经过网络传输时要经过许多路由器，要在路由器中排队等待转发。

网络的通信量越大，排队时延越长。

总时延 = 发送时延 + 传播时延 + 处理时延 + 排队时延。

高速链路（高带宽链路）相比其他提高的是数据的发送速率，减小发送时延。光纤传输速率高即向光纤信道发送数据的速率高。

时延带宽积

时延带宽积 = 传播时延 * 带宽。

时延带宽积表示的是一个管道的体积，表示这个链路可容纳的比特，或者说同一时刻正在链路上传送的数据。

只有链路中充满了比特时，链路才得到充分的利用。

往返时间

A 向 B 发数据后需要收到 B 发回的确认信息才会继续发送数据，A 发完数据后等待确认信息的时间就是**往返时间（RTT）**。

有效数据率 = 数据长度 / (发送时间 + RTT)。

在互联网中，往返时间还包括各中间结点的处理时延、排队时延和转发数据时的发送时延等。

使用卫星通信时，往返时间会比较长，是很重要的指标。

利用率

利用率有信道利用率和网络利用率两种：

信道利用率：信道有百分之多少的时间是被利用的，即有数据通过。

网络利用率：全网络的信道利用率的加权平均值。

信道利用率并非越大越好，因为根据排队论，当利用率增大时，信道引起的时延也会迅速增加。

D0 为网络空闲时的时延，则网络当前时延 D 和利用率 U 之间的关系为： **D = D0 / (1 - U）。**

当利用率 U 达到 0.5 时，时延就要加倍。利用率接近 1 时，时延会趋于无穷。

1.6.2 计算机网络的非性能特征

费用、质量、标准化、可靠性、可扩展性和可升级性、易于管理和维护

1.7 计算机网络体系结构

计算机的体系结构是分层次的。

1.7.1 计算机网络体系结构的形成

两个计算机之间通信要完的工作：

发起通信的计算机必须将数据通信的通路进行激活。

告诉网络如何识别接收数据的计算机

发起通信的计算机必须查明对方计算机是否已开机，并且与网络连接正常。

发起通信的计算机中的应用程序必须清楚，对方计算机中的文件管理程序是否已做好接收文件和存储文件的准备工作。

若计算机的文件格式不兼容，至少其中一台计算机可以完成格式转换。

如果出现差错或意外事故，应有可靠的措施保证对方计算机最后收到正确的文件。

相互通信的计算机系统必须高度协调工作才可以完成通信。

OSI 的七层协议的体系结构是法律上的国际标准。

应用最广泛的、事实上的国际标准是 **TCP/IP**。互联网即采用的 TCP/IP。

1.7.2 协议与划分层次

网络协议（简称**协议**）是为进行网络中的数据交换而建立的规则、标准或约定。

网络协议规定了网络中交换的数据的格式和有关的同步问题。

协议主要由**三个要素组成：**

语法：数据与控制信息的结构或格式

语义：需要发出何种控制信息，完成何种动作以及做出何种响应

同步：时间实现顺序的详细说明

网络协议是分层的，分层的好处：

各层之间互相独立。某一层并不知道另一层是如何实现的，只知道接口。

灵活性好。只要接口不变，一层发生变化不影响另一层。

结构上可分割开。

易于实现和维护。

能促进标准化工作。

各层需要完成的工作包括一下的一种或多种：

差错控制。使通信更加可靠。

流量控制。发送端的发送速率必须使接收端来得及接受，不能太快。

分段和重装。发送端将数据分块，接收端还原。

复用和分用。发送端几个高层会话复用一条低层的连接，在接收端再进行分用。

连接建立和释放。交换数据前先建立一条逻辑连接，数据传送结束后再释放。

计算机网络的各层及其协议的集合就是网络的体系结构。

1.7.3 具有五层协议的体系结构

TCP/IP 是个四层的体系结构，包含**应用层、运输层、网际层和网络接口层**。实质上 TCP/IP

只有上三层，最下面的网络接口层没什么内容。

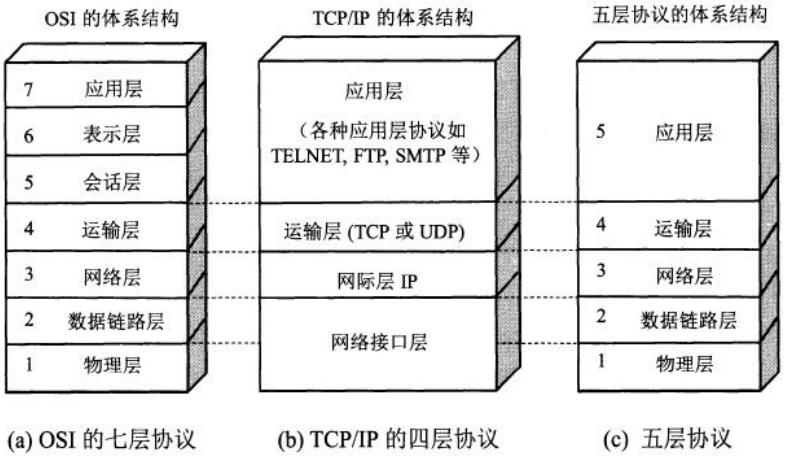


图 1-18　计算机网络体系结构

这里以五层协议为例来学习，五层协议中的物理层和数据链路层对于 TCP/IP 中的网络接口层。

应用层

任务：通过应用进程间的交互来完成特定网络应用。
应用层协议定义的是应用进程间通信和交互的规则。进程即主机中正在运行的程序。
互联网中的应用层协议有**域名系统 DNS**，支持万维网应用的 **HTTP 协议**，支持电子邮件的 **SMTP 协议**等。
应用层交互的数据单元称为**报文**。

运输层

任务：负责向两台主机中进程之间的通信提供通用的数据传输服务。应用层利用该服务传输应用层报文。
运输层有**复用**和**分用**的功能。因为一台主机有多个进程，复用就是多个应用层进程可同时使用下面运输层的功能。
运输层主要使用两种协议：

传输控制协议 TCP：提供面向连接的、可靠的数据传输服务。数据传输的单位是**报文段**。

用户数据报协议 UDP：提供无连接的、尽最大努力的数据传输服务，不保证数据传输的可靠性。数据传输的单位是**用户数据报**。

网络层

任务：为分组交换网上的不同主机提供通信服务；选择合适的路由。
发送数据时，网络层把运输层产生的报文段或用户数据报封装成分组或包进行传送。
TCP/IP 体系中，**网络层使用 IP 协议**，因此分组也叫 **IP 数据报**，或简称数据报，但注意这与用户数据报不同。
互联网使用无连接的网际协议 IP。
无论哪一层的数据单元，都可以笼统地用“分组”来表示。

数据链路层

任务：将网络层交下来的 IP 数据报组装成帧，在**两个相邻结点的链路**上传送帧。
每一帧包括数据和必要的控制信息（包括同步信息、地址信息、差错控制等）。

物理层

物理层传输的单位是比特。

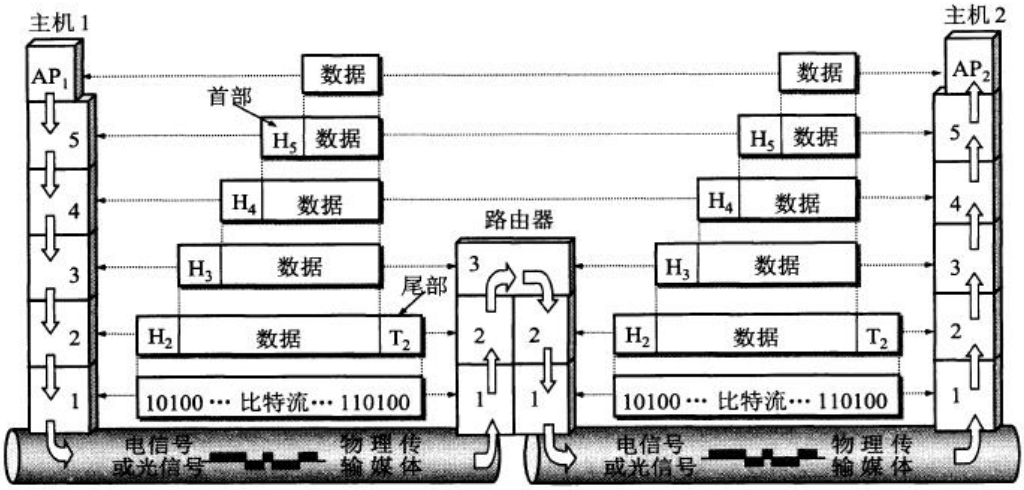


图 1-19　数据在各层之间的传递过程

上图中为主机 1 的 AP1 进程向主机 2 的 AP2 进程间传送数据的过程。
第 5 层、第 4 层、第 3 层分别为数据加上了属于本层的控制信息，都位于首部。第 2 层的控制信息分为两部分加到了首部和尾部。
对等层次间的数据单位称为该层的**协议数据单元**。
在路由器中，分组上升到第 3 层，过程中每一层都根据控制信息进行必要的操作并将该控制信息剥去。在第三层根据首部中的目的地址查找路由器中的转发表，找出转发分组的接口，然后依次加上新的控制信息下降到第 1 层，将数据发送出去。
理解：上图可以发现网际层的控制信息用于整个网络，而链路层的控制信息仅用于两个节点之间。路由器只到第 3 层。

1.7.4 实体、协议、服务和服务访问点

实体表示可发送或可接收信息的硬件或软件进程。
协议是控制两个对等实体间进行通信的规则的集合。
在协议的控制下，两个对等实体间的通信使得本层能够向上一层提供服务。要实现本层协议，也需要下一层提供的服务。
上层使用下层提供的服务必须通过与下层交换一些命令，即**服务原语**。
同一系统中相邻两层的实体交互的地方称为**服务访问点**，实际上就是一个逻辑接口。
协议必须把所有不利的条件都事先估计到，确保能应对所有异常情况。

1.7.5 TCP/IP 的体系结构

技术的发展不严格遵循分层概念，事实上现在某些应用程序可以直接使用 IP 层甚至网络接口层，如下图。

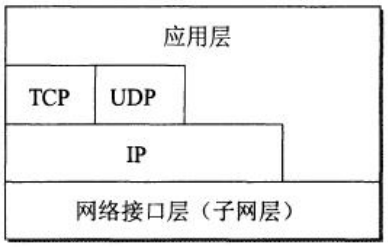


图 1-23　TCP/IP 体系结构的另一种表示方法

TCP/IP 协议族中的多种协议如下所示：

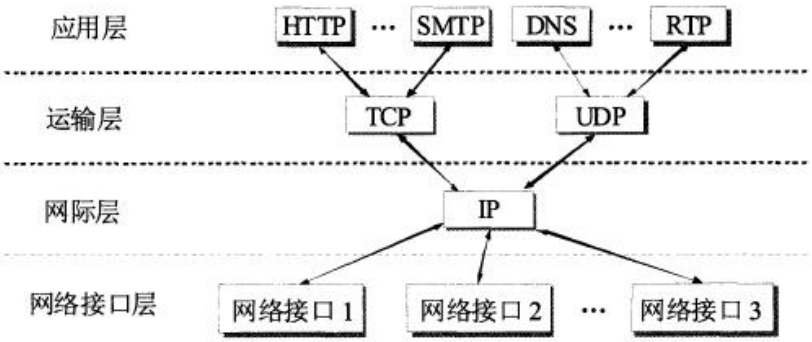


图 1-24　沙漏计时器形状的 TCP/IP 协议族示意

问题

物理层的任务是什么？
五种信道复用方式?
三种宽带接入方式？

回答

屏蔽掉不同传输媒体和通信手段间的差异，使链路层感受不到这种差异。
时分复用、频分复用、统计时分复用、码分复用、波分复用
非对称数字用户线（ADSL 技术，基于电话用户线改造）、光纤同轴混合网（HFC 网）、FTTx 技术（光纤到 x 技术）

第 2 章 物理层

本章最重要的内容：
物理层的任务
几种常用的信道复用技术
几种常用的宽带接入技术，主要是 ADSL 和 FTTx

2.1 物理层的基本概念

物理层关注的是如何在连接各种计算机的传输媒体上传输数据流。**物理层的任务**是尽可能屏蔽掉不同传输媒体和通信手段间的差异，使链路层感受不到这种差异。

物理层的主要任务：确定与传输媒体的接口有关的一些特性：

机械特性：指明接口所用接线器的形状、尺寸等机械特性。

电气特性：指明接口电缆的各条线上的电压的范围。

功能特性：指明电线上某一电平的电压的意义。

过程特性：指明不同功能的各种可能事件的出现顺序。

数据在计算机内部一般是并行传输，但在通信线路上是串行传输，所以物理层还要完成传输方式的转换。

物理层协议很多，因为物理连接的方式很多，传输媒体的种类也很多。

2.2 数据通信的基础知识

2.2.1 数据通信系统的模型

一个**数据通信系统**可划分为源系统、传输系统、目的系统，或称为发送端、传输网络、接收端。

源系统包括源点和发送器。典型的发送器是**调制器**。

目的系统包括接收器和终点。典型的接收器是**解调器**。

通信的目的是传送**消息**，**数据**是运送消息的实体，**信号**是数据的电气或电磁的表现。

信号可分为**模拟信号**和**数字信号**。

2.2.2 有关信道的几个基本概念

信道不等于电路，信道表示向某一方向传送信息的媒体，一条通信电路通常包含一条发送信道和一条接受信道。

通信方式

信息交互有以下三种基本方式：

单向通信，又称单工通信。如有线电广播等。需要一条信道。

双向交替通信，又称半双工通信。需要两条信道。

双向同时通信，又称全双工通信。需要两条信道。

调制

来自信源的信号成为**基带信号**，因为基带信号中包含较多低频成分，而许多信道不能传输低频分量和直流分量，所以需要对基带信号调制。

调制可分为两大类：

基带调制：将数字信号转换为另一种数字信号。又称编码。

带通调制：将基带信号的频率范围转换为另一频段，并化为模拟信号。

常用编码方式

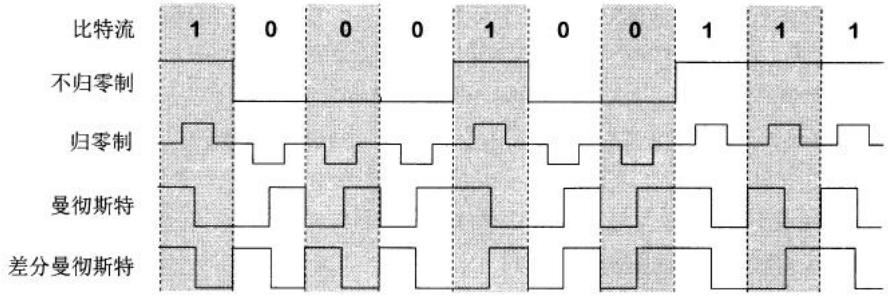


图 2-2　数字信号常用的编码方式

不归零制：正电平代表 1，负电平代表 0。
归零制：正脉冲代表 1，负脉冲代表 0。
曼彻斯特编码：位周期中心的向上跳变代表 0，向下跳变代表 1。
差分曼彻斯特编码：每一位的中心都有跳变。位开始的边界有跳变代表 0，没有代表 1。
曼彻斯特码的频率比不归零制高，但有自同步能力，即可以从信号波形自身中提取信号时钟频率。

基本带通调制方法

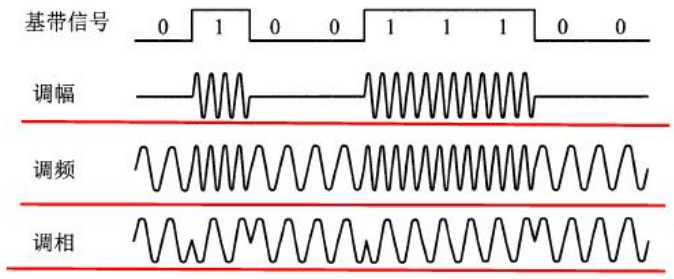


图 2-3　最基本的三种调制方法

2.2.3 信道的极限容量

数字通信的优点：信号在信道上传输时必然会失真，但只要能识别出原有信号，就没有影响。
传输速率越高，或距离越远，或噪声越大，失真就越严重。

信道能通过的频率范围

信道中码元传输的速率有上限，超过上限会出现严重的码间串扰问题，接收端无法识别编码。
信道的频带越宽，能通过的高频分量越多，最大速率越高。

信噪比

信号的平均功率与噪声的平均功率之比，写作 S/N，单位是分贝（dB）
信噪比 = 10 log10(S/N)。

香农公式

信道的极限传输速率 C = W log(2+S/N)。
香农公式表明带宽越大，信噪比越大，极限传输速率越高。还表明只要信息传输速率低于信道的极限速率，就一定可以实现无差错传输，但方法未知。
另一种提高传输速率的方法：通过编码让每个码元携带更多比特的信息。

2.3 物理层下面的传输媒体

传输媒体分为**导引型**和**非导引型**两大类。
导引型中电磁波沿着固体媒体传播，非导引型中传输媒体就是自由空间，又称无线传输。

2.3.1 导引型传输媒体

导引型传输媒体有架空明线，双绞线，同轴电缆，光纤等。
光纤的传输带宽远大于其他传输媒体的带宽。

2.3.2 非导引型传输媒体

利用无线信道进行传输是**运动中通信**的唯一手段。
短波通信质量较差，速率较低。无限电微波通信可传输电话、图像、数据等信息。紫外线及更高波段目前还不能用于通信。
卫星通信的优点是通信距离远，缺点是传播时延高，保密性差。

2.4 信道复用技术

信道复用：多个发送端使用同一条信道来传输信息。
发送端使用复用器将不同的信息合起来传输，接收端使用分用器将信息分开。

2.4.1 频分复用、时分复用和统计时分复用

三种复用：

频分复用 FDM：每个用户分配一个频带，通信中始终占用该频带。用户在同样时间占用不同的频带。

时分复用 TDM：将时间划分为等长的帧，每个用户在每个帧中占用其中一个固定序号的间隙。用户在不同时间占用同样的频带。

因为计算机数据的突发性，时分复用的信道利用率比较低。

统计时分复用 STDM：一种改进的时分复用，又称异步时分复用。STDM 不是固定分配时隙，而是按需动态地分配时隙。

2.4.2 波分复用

波分复用 WDM 就是光的频分复用。
一根光纤上可以复用几十路甚至更多的光载波信号。光信号传输一定距离后会衰减，因此需要使用**光纤放大器**放大后继续传输。

2.4.3 码分复用

码分复用 CDM：不同用户使用不同码型，在同样时间使用同样的频带通信。
如对某一个用户，序列 00011011 表示比特 1,11100100 表示比特 0。其他用户的码片序列必须与此用户的序列相互正交。
码分复用实际上是一种**扩频通信**。无线局域网中常用 CDM。

2.5 数字传输系统

数字通信相比模拟通信，在传输质量和经济上都更好。
光纤是长途干线最主要的传输媒体。
同步数字序列 SDH 和**同步光纤网 SONET** 是当前最主要的**数字传输国际标准**。简称 **SONET/SDH 标准**

2.6 宽带接入技术

用户连接到互联网，要先连接到某个 ISP，以便获得上网所需的 IP 地址。
宽带接入网是接入网的一种，即一种用来把用户接入到互联网的网络。
宽带接入可分为**有线宽带接入**和**无线宽带接入**。

2.6.1 ADSL 技术

非对称数字用户线 ADSL 技术是用数字技术对现有的模拟电话用户线进行改造，使其能够承载宽带数字业务。
标准模拟电话信号的频带在 300~3400Hz 范围，ADSL 技术将 4000Hz 以下的频带留给传统电话，4000Hz 以上用于上网。
因为用户一般都是下载，ADSL 的下行带宽（从 ISP 到用户）远大于上行带宽，所以叫做非对称。
ADSL 的好处是可以利用现有的电话线，缺点是传输距离有限，并且不能保证固定的数据率。ADSL 的速率依赖于用户线的质量、长度、线径等。
ADSL 在用户线（铜线）的两端各安装一个 ADSL 解调器。采用**基于频分复用的 DMT 调制技术**，将 4kHz 以上的频带划分为许多子信道，其中 25 个子信道用于上行，249 个子信道用于下行。
类似 ADSL 还有许多其他 xDSL 技术，速度更快，但在国内应用较少。

2.6.2 光纤同轴混合网(HFC 网)

光纤同轴混合网（HFC 网）是基于有线电视网开发的一种宽带接入网。
为提高传输的可靠性和质量，HFC 网将原有线电视网的同轴电缆主干部分改换为了光纤。
光纤从头端连接到光纤结点，在**光纤结点**处光信号转换为电信号，连接到一个光纤结点的典型用户数为 500。

光纤节点与头端的典型距离为 25km，到用户的距离不超过 3km。

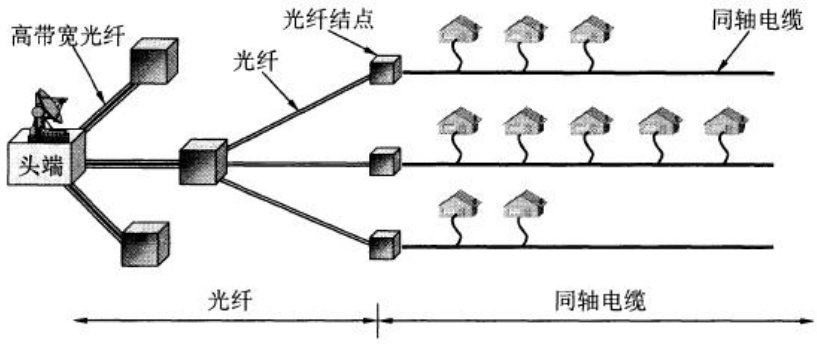


图 2-23　HFC 网的结构图

用户通过**电缆调制解调器**来使用 HFC 网，它比 ADSL 中的解调器复杂很多，因为要解决共享信道中的冲突问题。

使用 HFC 网的数据率大小不确定，它取决于这段电缆上有多少个用户正在传送数据，如果有很多人在用，每个人的速率会很慢。

2.6.3 FTTx 技术

光纤到户 FTTH(Fiber To The Home) 是把光纤一直铺设到用户家庭，在光纤进入用户家中后才把光信号转换为电信号，这样的上网速率最快。

现在信号在陆地上的长距离传输基本都是使用的光缆，在 ADSL 和 HFC 中长距离传输也是用的光缆。

多个用户通过**光配线网**共享一根光纤干线，光配线网使用波分复用，上行和下行使用不同的波长。

出光纤到户 FTTH 外，还有光纤到大楼 FTTB，光纤到楼层 FTTF 等，一般运行商所说的光纤到户并非真正的 FTTH。

第 3 章 数据链路层

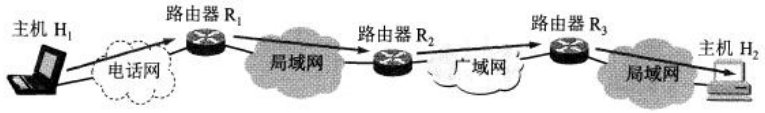
本章最重要的内容：

数据链路层的点对点信道和广播信道的特点。PPP 协议和 CSMA/CD 协议的特点。

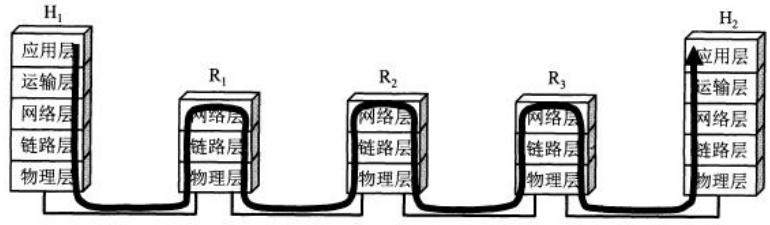
数据链路层的三个基本问题：封装成帧、透明传输、差错检测。

以太网 MAC 层的硬件地址。

适配器、转发器、集线器、网桥、以太网交换机的作用和使用场合。



(a) 主机H_1向H_2发送数据



(b) 从层次上看数据的流动

图 3-1　数据链路层的地位

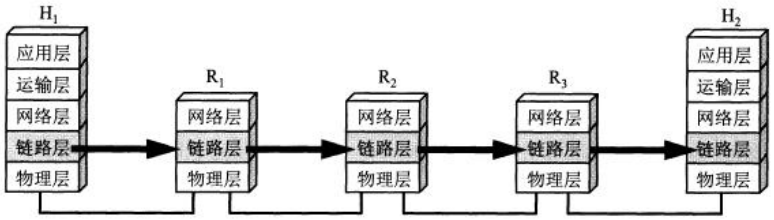


图 3-2　只考虑数据在数据链路层的流动

路由器转发分组时只涉及到下面三层。

3.1 使用点对点信道的数据链路层

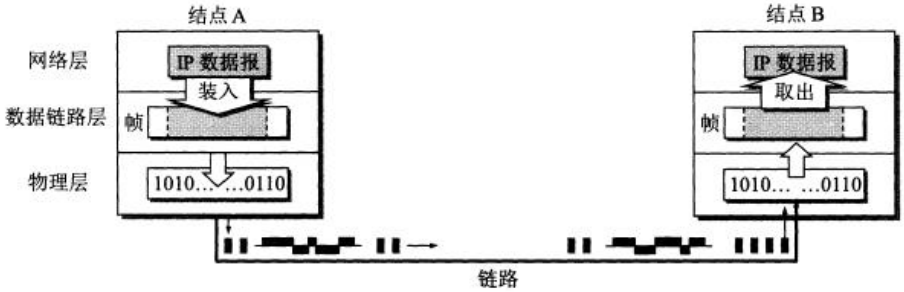
3.1.1 数据链路和帧

数据链路既包含物理线路也包含必要的通信协议，将实现协议的软件和硬件加到链路上就构成了数据链路。

常用**网络适配器**（既包括硬件也包括软件）来实现这些协议，一般适配器包括了数据链路层和物理层两层的功能。

帧是点对点信道的数据链路层的**协议数据单元**。网络层的协议数据单元是 **IP 数据报**，又称分组。

数据链路层将网络层交下来的数据构成帧发送到链路上，以及把接收到的帧里的数据取出并上交给网络层。



点对点信道的数据链路层在通信时的**主要步骤：**

结点 A 的数据链路层把网络层交下来的 IP 数据报加上首部和尾部封装成帧。

结点 A 把封装好的帧发送给结点 B。

结点 B 对接收到的帧进行差错检验，若无差错，从帧中提取出 IP 数据报上交给网络层，若有差错丢弃这个帧。

3.1.2 三个基本问题

数据链路层的三个基本问题：**封装成帧、透明传输、差错检验**

封装成帧

给 IP 数据报加上首部和尾部就构成了数据链路层的帧，IP 数据报成为帧的数据部分。
链路层协议规定了帧中数据部分的长度上限——**最大传送单元（MTU）**。
首部和尾部包括**帧定界符**（即确定帧的界限）和其他控制信息。
帧定界符包括开始符（SOH）和结束符（EOT），分别是一串 8 为二进制数字。
帧定界符的作用：确定帧的界限。当出现差错时可以根据帧定界符识别是否是一个完整的帧。

透明传输

透明传输是因为控制字符产生的。
透明传输即表示无论传送什么样的数据，都能按照原样无差错地通过数据链路层。
字节填充：因为存在帧定界符，如果传输的数据中出现了和 SOH、EOT 等控制字符一样的文本，就在文本前面插入一个**转义字符（ESC）**，接受端收到数据后在发送给网络层之前删除这个插入的转义字符。如果转义字符也出现在数据中，就在它前面再插入一个转义字符。

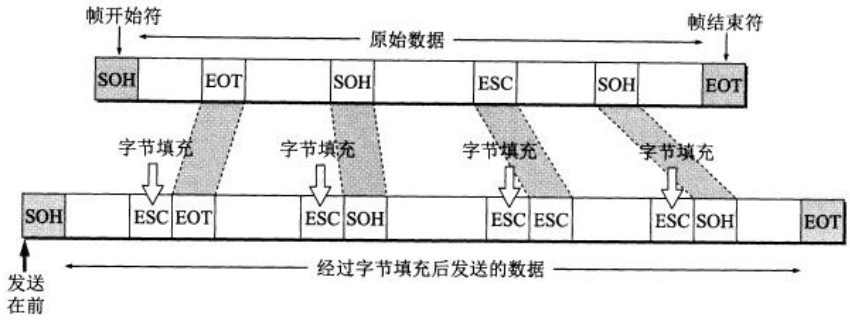


图 3-7　用字节填充法解决透明传输的问题

差错检测

比特差错：传输时产生差错，0 变成了 1 或 1 变成了 0.
误码率 BER：一段时间内，传输错误的比特占所传输比特总数的比率。提升信噪比可以减小误码率。
差错检验的方法：**循环冗余检验 CRC**
循环冗余检验的原理：
在要传输的数据后加上 n 位的冗余码（成为**帧检验序列 FCS**），如何得出冗余码：让数据乘以 2^n（相当于在后面加了 n 个 0），然后除以一个事先约定的 n+1 位的除数 P，得出 n 位的余数 R 就作为冗余码加到数据后面。接收端将收到的每一个帧除以同样的除数 P，若

余数为 0 就表明没有差错，否则就是有差错，丢弃这个帧。
循环冗余检验使用硬件完成，速度很快。
循环冗余检验只能识别比特差错，无法识别帧丢失、帧重复、帧失序，因此不是可靠传输。
对于通信质量较差的无线传输链路，数据链路层协议使用帧编号、确认和重传机制。即接收方收到正确的帧就向发送方发送确认，如果发送方没有收到确认就表明出现差错，就进行重传直到收到对方的确认。
对于通信质量较好的有线传输链路，只进行 CRC 检验，不使用确认和重传机制，即不需要数据链路层向上提供可靠传输，而是由上层协议来改正差错。
本章的 **PPP 协议**和 **CSMA/CD 协议**都不是可靠传输的协议。

3.2 点对点协议 PPP

点对点协议 PPP 是目前点对点链路中应用最广泛的数据链路层协议。

3.2.1 PPP 协议的特点

PPP 协议是用户和 ISP 通信时使用的数据链路层协议。

PPP 协议应满足的需求

简单。这是首要的需求。互联网体系结构中最复杂的部分在 TCP 协议中，网际协议 IP 和数据链路层协议都不是可靠传输。

封装成帧。PPP 协议规定使用特殊的字符作为帧定界符。

透明传输。

支持多种网络层协议。PPP 协议要能够在同一条物理链路上同时支持多种网络层协议。

支持多种类型链路。包括串行的或并行的、同步的或异步的等。例如 PPPoE（在以太网上运行的 PPP），用户通过以太网上网时使用的是 PPPoE 协议，它将 PPP 帧再封装到一个以太网帧中。

差错检测。如果收到有错的帧就丢弃。

最大传送单元。要为每一种类型的点对点链路设置最大传送单元 MTU。注意 MTU 是数据部分的最大长度。

网络层地址协商。PPP 协议要提供一种机制使通信的两个网络层的实体能通过协商知道彼此的网络层地址。

数据压缩协商。PPP 协议要提供一种方法来协商使用数据压缩算法。
TCP/IP 协议族中，可靠传输由 TCP 协议负责。PPP 不负责纠错等。PPP 只支持点对点的链路通信，且只支持全双工链路。

PPP 协议的组成

PPP 协议有**三个组成部分：**

一个将 IP 数据报封装到串行链路的方法。

一个用来建立、配置和测试数据链路连接的**链路控制协议 LCP**。

一套**网络控制协议 NCP**，其中的每一个协议支持不同的网络层协议。

3.2.2 PPP 协议的帧格式

各字段的含义

PPP 的首部和尾部分别为 4 个字段和 2 个字段。

首部的第一个字段和尾部的第二个字段都是**标志字段 F**，规定为 0x7E，它标志着一个帧的开始或结束。两个连续的帧之间只需要一个 F，如果连续出现两个标志字段，表示这是一个空帧，应该丢弃。

首部的第二个和第三个字段 目前都没有实际含义。第四个字段是 2 字节的协议字段，它表明了信息部分的数据类型（可能是 IP 数据报也可能是其他类型的数据）。**尾部的第一个字段是帧检验序列 FCS。**

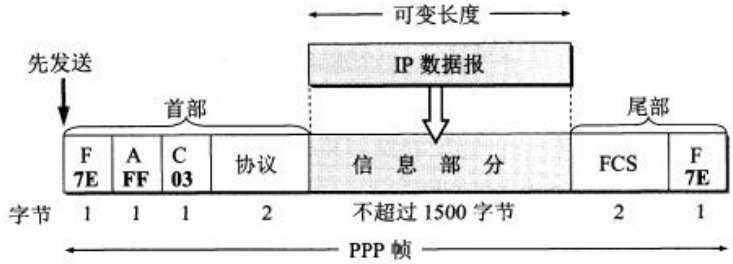


图 3-10 PPP 帧的格式

字节填充

PPP 使用异步传输时使用了**字节填充**，转义符为 0x7D。

零比特填充

PPP 协议用在 SONET/SDH 链路上时使用同步传输，此时采用**零比特填充**方法来实现透明传输。

零比特填充的方法：当信息字段中出现 5 个连续的 1，立即填入一个 0，这样信息字段中就不会出现 6 个连续的 1（PPP 的帧定界符中有 6 个连续的 1）。

3.2.3 PPP 协议的工作状态

PPP 链路从建立到释放的全过程：用户拨号接入 ISP 后，就建立了从用户到 ISP 的物理连接。这时用户向 ISP 发送一系列的链路控制协议 LCP 分组，以便建立 LCP 连接。然后网络控制协议 NCP 给新接入的用户电脑分配一个临时的 IP 地址。等用户通信完毕后，NCP 释放网络层连接，收回分配的 IP 地址，然后 LCP 释放数据链路层连接，最后释放物理层连接。

PPP 链路的状态变化：链路静止——链路建立——鉴别——网络层协议——链路打开——链路终止——链路静止。

链路静止：PPP 链路的其实和终止状态都是链路静止状态。

链路建立：当个人电脑当建立了到路由器的物理层连接后，PPP 进入链路建立状态，目的是建立链路层的 LCP 连接。

通过发送 LCP 的配置请求帧（是一个 PPP 帧，协议字段为 LCP 对应的代码，信息字段包括特定的配置请求）来协商配置选项，链路的另一端可以回复配置确认帧、配置否认帧

或配置拒绝帧。

鉴别：若鉴别身份失败就转到链路终止状态，若成功就进入网络层协议状态

网络层协议：这时 PPP 链路的两端的网络控制协议 NCP 根据网络层的不同协议互相交换网络层特定的网络控制分组。PPP 链路的两端可以运行不同的网络层协议。

链路打开：此时链路的两个 PPP 端点可以彼此发送分组。

数据传输结束后链路的一端发送终止请求 LCP 帧请求终止链路连接，收到终止确认后转到链路终止状态。

链路终止。终止后进入链路静止状态。

链路静止。

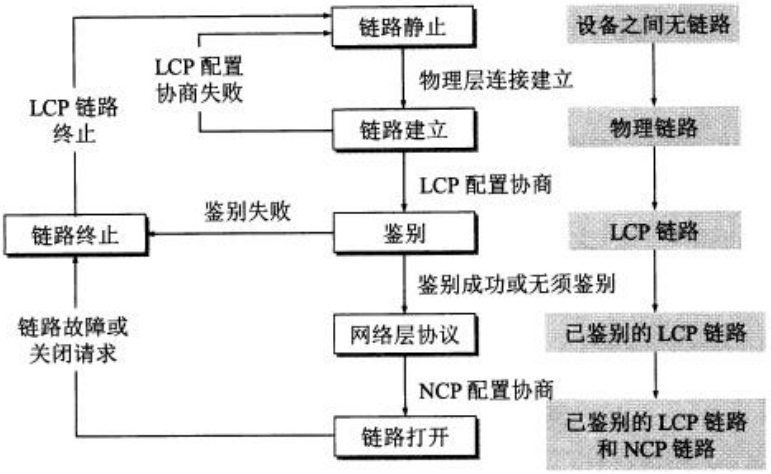


图 3-12　PPP 协议的状态图

可以看到，PPP 协议不是纯粹的数据链路层协议，还包含了物理层和网络层的内容。

3.3 使用广播信道的数据链路层

广播信道可以进行一对多的通信，**局域网用的就是广播信道。**

局域网可以按照拓扑结构进行分类。

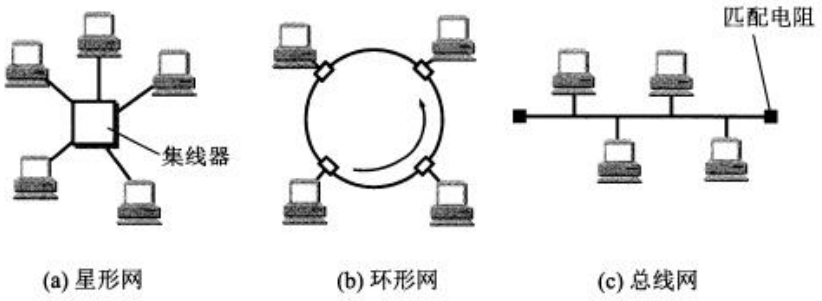


图 3-13　局域网的拓扑

3.3.1 局域网的数据链路层

局域网的最主要特点：网络为一个单位所拥有，且地理范围和站点数目都有限。

局域网的优点：

具有广播功能，可以从一个站点很方便地访问全网。局域网上的主机可以共享连接在局域网上的各种硬件和软件资源。

便于系统的扩张和逐渐演变

提高了系统的可靠性、可用性和生存性。

以太网是局域网的一种，绝大多数局域网都是以太网。双绞线是局域网中的主流传输媒体。

实现共享信道有两种方法：

静态划分信道，如频分复用、时分复用、码分复用等，但不适合局域网。

动态媒体接入控制，又称**多点接入**。特点是信道并非在用户通信时固定分配给用户。

随机接入：特点是用户可以随机地发送消息。如果有两个用户同时发送，在共享媒体上就会产生碰撞，是发送失败。这时就需要解决碰撞的网络协议，即 CSMA/CD 协议。

受控接入：特点是用户不能随机发送信息而必须服从一定的控制。

以太网应用的主要是随机接入。

由于历史原因以太网层被拆分为两个子层：逻辑链路控制 LLC 和**媒体接入控制 MAC**。现在 LLC 基本已经消失了，主要是 MAC 协议。

适配器的作用

计算机与外界局域网的连接是通过适配器进行的，适配器以前又称网卡。

适配器和局域网之间的通信通过电缆或双绞线以串行传输方式进行的，而适配器与计算机之间的通信是通过 I/O 总线并行传输的，因此适配器的一个重要功能就是进行数据串行传输和并行传输的转换。

适配器实现的功能包含了数据链路层和物理层两个层次的功能。

适配器收到正确的帧后，使用中断来通知计算机，并把数据交付给协议栈中的网络层。当计算机要发送 IP 数据报时，就由协议栈把分组交给适配器，适配器将其组装成帧后发送到局域网。（封装成帧、透明传输、差错检错等功能都是由适配器完成的）

计算机的**硬件地址存储在适配器中**，而**软件地址——IP 地址存储在计算机中。**

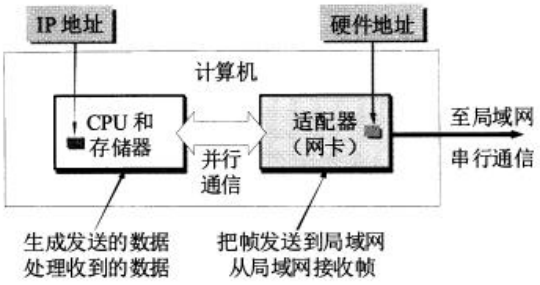


图 3-15　计算机通过适配器和局域网进行通信

3.3.2 CSMA/CD 协议

局域网上的计算机常被称为工作站、站点等。

为了通信的简便，以太网采取了以下两种措施：

采用较为灵活的无连接的工作方式，即不必建立连接就可以直接发送数据。适配器对发送的数据帧不编号，也不要求对方发回确认。它提供的是尽最大努力的交付，是不可靠的交付。对有差错帧是否进行重传由高层来决定。

同一时间只能有一台计算机发送数据，如果发生冲突，就使用 CSMA/CD 协议来协调。

以太网发送的数据使用的是曼彻斯特编码。

CSMA/CD 协议的要点

多点接入：多点接入说明是总线型网络，许多计算机以多点接入的方式连接在一根总线上。**协议的实质就是载波监听和碰撞检测**。

载波监听：使用电子技术检测信道上有没有其他计算机也在发送。不管是发送前还是发送中，每个站都要不停地检测信道。

碰撞检测：边发送边监听。如果几个站同时发送数据，总线上的信号电压变化会增大，就表明发生了碰撞。这时就立即停止发送。

在使用 CSMA/CD 协议时，不能同时发送和接收，因此使用 CSMA/CD 协议的以太网只能进行半双工通信（双向交替通信）。

发生碰撞是因为**传播时延**，A 发送了数据但是还没传到 B 处，B 就不知道有人发送了数据。

当 A 和 B 同时发送数据产生碰撞后，他们发送数据都失败，都要推迟一段时间重新发送。

因为不知道是否会发生碰撞，所以以太网存在**发送的不确定性**。

A 发送数据后最多 2θ 时间就知道是否碰撞，这 2θ 时间称为**争用期**。如果经过争用期还没碰撞，表明发送成功。

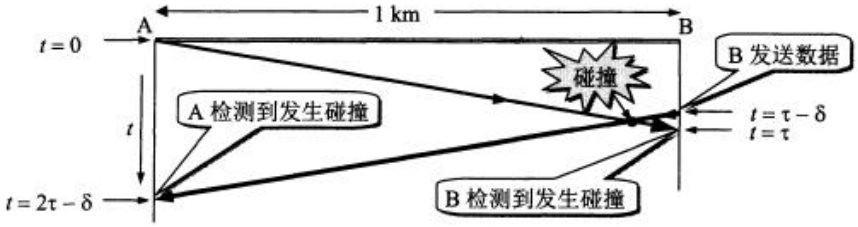
如果发生碰撞，以太网使用**截断二指数退避算法**来确定碰撞后重传的时机。

以太网规定了争用期的时长，这就约束了以太网的地理范围不能太大，不然传播时延会超过争用期限制。

以太网规定最短帧长为 64 字节（一个争用期可以发送的字节数），如果争用期发生碰撞就会停止发送，因此信道上长度小于 64 字节的帧就是无效帧。

碰撞后除了立即停止发送数据外还要继续发送一个人为干扰信号，通知所有用户现在发生了碰撞。

以太网还规定了**帧间最小间隔** 96 比特时间，这是为了使刚收到数据帧的栈清理缓存，准备接收下一帧。



CSMA/CD 协议的要点归纳

准备发送：适配器从网络层获得一个分组，加上首部和尾部组成以太网帧，放入适配器缓存中。在发送前先检测信道。

检测信道：若检测到信道忙，则不停地检测直到信道空闲。若检测到空闲，并在 96 比特时间内保持空闲（保证了帧间最小间隔），就发送这个帧。

在发送过程中仍不停地检测，即适配器要**边发送边监听**。这时有两种情况

发送成功：争用期内一直未检测到碰撞。发送成功后回到 1。

发送失败：争用期内检测到碰撞，立即停止发送，并按规定发送人为干扰信号。适配器接着执行**指数退避算法**，等待足够时间后回到 2。若重传 16 次仍不成功，就停止重传并

向上报错。
以太网发送完一帧后要把已发送的帧保留一下。如果争用期检测到碰撞，推迟一段时间后还要重传。

3.3.3 使用集线器的星形拓扑

现在的以太网采用星形拓扑，在星形中心使用可靠性非常高的集线器。每个站用两对双绞线，分别用于发送和接收。
集线器的特点：

表面上使用集线器的局域网在物理上是一个星形网，但是在**逻辑上仍是一个总线网**，各站共享逻辑上的总线，还是使用 CSMA/CD 协议。

一个集线器有很多接口，像是一个多接口的转发器。

集线器工作在物理层，每个接口仅负责转发比特，不进行碰撞检测。

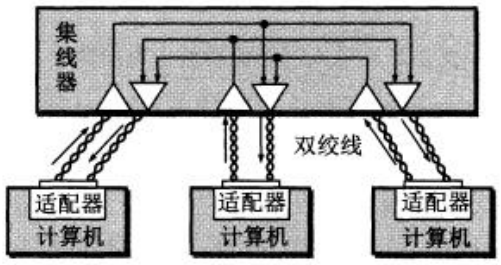


图 3-20　具有三个接口的集线器

3.3.4 以太网的信道利用率

因为发生碰撞会浪费信道资源，所以以太网的信道利用率达不到 100%。
减少端到端的传播时延、可以提高信道利用率，因此以太网的连线的长度不能太长，同时以太网的帧长不能太短。

3.3.5 以太网的 MAC 层

MAC 层的硬件地址

局域网中，**硬件地址又叫 MAC 地址。**
IEEE 为局域网规定了一种 **6 字节的全球地址**，是局域网上的每一台计算机中**固化在适配器中的地址**。因此如果更换了新的适配器，硬件地址也就变了。
适配器上的标识符 **EUI-48** 就是计算机的硬件地址。
路由器通过适配器连接到局域网时，适配器上的硬件地址标志路由器的一个接口。如果路由器同时连到多个网络上，就需要多个适配器有多个硬件地址。
局域网中适配器收到的帧有三种：

单播帧（一对一）：即收到的帧的 MAC 地址与本站的地址相同。

广播帧（一对全体）。

多播帧（一对多）。

适配器至少能够识别前两种帧。

以太网适配器有一种特殊的工作方式：**混杂方式**。混杂方式的适配器只要”听到“有帧再传输就悄悄接收下来。
混杂方式可以用来监视和分析以太网上的流量，黑客也常用混杂方式非法获取信息。

MAC 层的帧格式

最常用的 MAC 帧格式是”**以太网 V2 标准**“，此外还有 IEEE 的 **802.3 标准**。
MAC 帧的首部共有源地址字段、目的地址字段、用来标识上层使用什么协议的类型字段这 3 个字段，尾部有一个帧检验序列 FCS。
MAC 帧没有帧定界符也没有帧长度字段。因为它用的是曼彻斯特码，曼彻斯特码的码元中间有一个电压跳变。当发送方发完一个帧后就不发送码元了，这是接收方发现没有跳变了就知道帧结束了。
MAC 帧在向下传送到物理层时要在帧前面插入 8 字节，包括一个前同步码和一个帧开始定界符。前同步码用来通知接收端调整时钟频率以与发送端的时钟同步。
MAC 帧的最小长度是 64 字节，其中数据字段最小长度是 46 字节。如果不够就要进行填充。IP 数据报的首部有一个”总长度“字段，网络层根据它来识别填充字段的长度并丢弃掉。

3.4 扩展的以太网

有时会对以太网的范围进行扩展。这种扩展的以太网在网络层看来仍然是一个网络。

3.4.1 在物理层扩展以太网

由于 CSMA/CD 协议的限制，以太网的主机之间距离不能太远。
可以使用**光纤**来扩展主机和集线器之间的距离，因为光纤的时延小带宽宽，所以可以很轻松地将主机和集线器的距离扩展到几千米。
可以将多个以太网通过**主干集线器**连接起来形成一个更大的以太网。

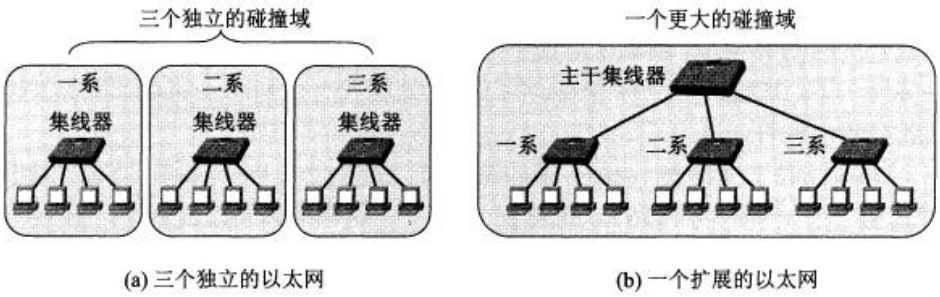


图 3-24　用多个集线器连成更大的以太网

多个以太网通过连接进行扩展后的优点是可以实现不同以太网间的通信，且扩大了地理范围。缺点是碰撞域会增大，发送数据产生碰撞的概率增加。

3.4.2 在数据链路层扩展以太网

扩展以太网更多的是在数据链路层扩展。

以太网通过**以太网交换机（又称第二层交换机）**来在数据链路层进行扩展。

以太网交换机的特点

以太网交换机实际上是一个多接口的网桥，每个接口直接与一台主机或另一个交换机相连，一般工作在全双工方式。
以太网交换机具有并行性，可以同时连接多对接口，使多对主机同时通信。相互通信的主机都独占传输媒体，无碰撞地传输数据。
以太网交换机是一种即插即用设备，它内部的帧交换表（又称地址表）是通过**自学习算法**自动地逐渐建立起来的。

以太网交换机的最大优点

交换机的最大优点：它的并行性。多对主机同时通信并不会平分总带宽，因为每对主机独占其传输媒体的带宽，所以每对主机的带宽还是原带宽。
传统的 10Mbit/s 的共享式以太网，如果有 10 个用户，则每个用户的平均带宽为 1Mbit/s，而用以太网交换机来连接这些主机，10 个用户的带宽都是 10Mbit/s，相当于总带宽 100Mbit/s。

以太网交换机的自学习功能

实现自学习的两个关键点：

若有主机发送数据，就把该主机的 MAC 地址与对应接口存入交换表。
若交换表中找不到数据接收方的对应接口，就对所有接口进行广播。

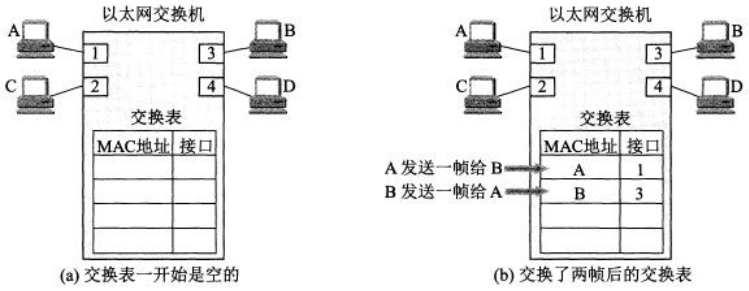


图 3-25　以太网交换机中的交换表

A 先向 B 发送一帧，从接口 1 进入到交换机。交换机收到帧后，先查找交换表，没有查到应从哪个接口转发这个帧（在 MAC 地址这一列中，找不到目的地址为 B 的项目）。接着，交换机把这个帧的源地址 A 和接口 1 写入交换表中，并向除接口 1 以外的所有接口广播这个帧（这个帧就是从接口 1 进来的，当然不应当把它再从接口 1 转发出去）。

C 和 D 将丢弃这个帧，因为目的地址不对。只 B 才收下这个目的地址正确的帧。这也称为过滤。

因对接口连接的主机可能会改变，主机的网络适配器也可能改变，所以交换表中的每个项目都有有效时间，时间过了就会删除。

从总线以太网到星形以太网

总线以太网使用 **CSMA/CD** 协议，以**半双工方式**工作。
而以太网交换机**不使用共享总线，没有碰撞问题**，因此不使用 CSMA/CD 协议，而是以全双工方式工作。

3.4.3 虚拟局域网

使用以太网交换机可以方便地实现**虚拟局域网 VLAN**。

虚拟局域网 VLAN：它是由一些局域网网段构成的与物理位置无关的逻辑组，这些网段具有某些共同的需求。每一个 VLAN 的帧都有一个明确的标识符，指明发送这个帧的计算机属于哪一个 VLAN。

虚拟局域网是局域网给用户提供的一种服务，不是一种新型局域网。

下图中每一个 VLAN 的计算机可以处于不同的局域网中。

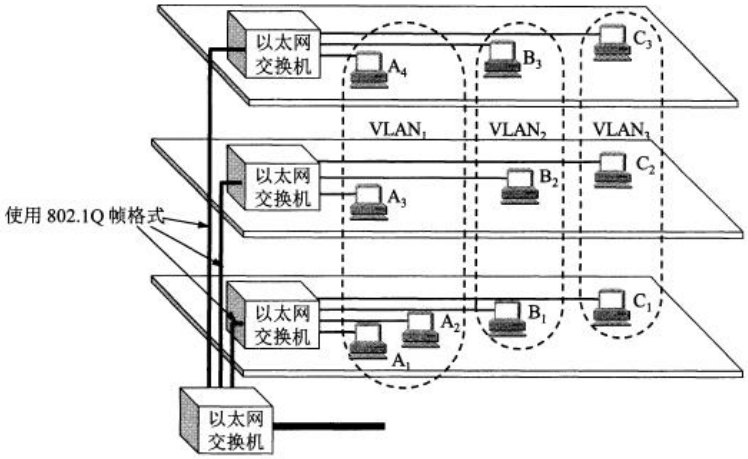


图 3-27　三个虚拟局域网 $VLAN_1$，$VLAN_2$ 和 $VLAN_3$ 的构成

3.5 高速以太网

现在的以太网的速率已经从传统的 10Mbits/s 发展到了 1Gbit/s 的吉比特以太网。

3.5.1 100BASE-T 以太网

100BASE-T 是在双绞线上传送 100Mbit/s 基带信号的星形拓扑以太网，仍使用 CSMA/CD 协议，又称快速以太网。

100BASE-T 可以使用以太网交换机，当使用以太网交换机时工作在全双工状态，且不使用 CSMA/CD 协议。

快速以太网使用的 MAC 帧格式仍然是 IEEE802.3 标准规定的帧格式。

3.5.2 吉比特以太网

吉比特以太网有以下几个特点：

允许在 1Gbit/s 下以全双工和半双工两种方式工作。

使用 IEEE 802.3 协议规定的帧格式。

在半双工方式下使用 CSMA/CD 协议，在全双工方式不使用。

与 10BASE-T 和 100BASE-T 向后兼容。

吉比特以太网在半双工时，采用了**“载波延伸”**方法，延长争用期与发送的 MAC 帧最小长度到 512 字节。这在发送短帧时需要进行大量填充造成了额外开销。

还增加了**“分组突发”**的功能，当很多短帧要发送时，第一个短帧采用载波延伸的方法进行填充，后面的则一个接一个地发送而不需填充。

3.5.3 10 吉比特以太网和更快的以太网

10GE 的帧格式与 10Mbit/s, 100Mbit/s 和 1Gbit/s 的帧格式完全相同，最小帧长和最大帧长也相同。

10GE 只工作在全双工方式，**不使用 CSMA/CD 协议**，这使它的传输距离极大地提高。

以太网技术发展很快，10GE 后面又制定了 40GE 和 100GE 的标准，他们都只工作在全双工方式。传输距离可达几十千米。

现在以太网的工作范围已经扩大到城域网和广域网，它的优点是：

技术成熟。

互操作性好。

价格便宜。广域网中使用以太网时价格比同步光纤网 SONET 便宜很多。

端到端的以太网连接使帧的格式全都是以太网的格式，不需要进行帧格式转换。

以太网的发展证明了**以太网的优点：**

可扩展（从 10Mbit/s 到 10 Gbit/s）。

灵活（多种媒体、全/半双工，共享/交换）。

易于安装。

稳健性好。

3.5.4 使用以太网进行宽带接入

现在也使用以太网进行宽带接入互联网。

以太网接入可以提供双向的宽带通信，且可以根据需要灵活地升级（如从 10M 到 10G）。

但是以太网的帧格式中没有用户名字段和让用户键入密码来鉴别用户身份的过程。于是诞生了 PPPoE（在以太网上运行 PPP），它把 PPP 协议中的 PPP 帧封装到以太网中来传输。

现在的光纤宽带接入 FTTx 都是用 PPPoE。

问题 4.1~4.2

互联网的 IP 协议提供的是什么样的服务？传统电信网提供的是什么样的服务
与 IP 协议配套的三个协议是什么？

知识点：数据链路层使用网桥将以太网连接成桥接以太网，网络层使用路由器将不同网络连接起来。

什么是 IP 地址

IP 地址由哪两部分组成

IP 地址可以分为哪两类

IP 地址和硬件地址有哪些区别？

地址解析协议的用处？

地址解析协议的作用范围？

地址解析协议的作用原理

IP 数据报的首部包含哪些部分？

路由器的路由表中存储什么内容

IP 层的分组转发是怎么实现的？

回答 4.1~4.2

IP 协议提供简单灵活的、无连接的、尽最大努力交付的服务。传统电信网是面向连接的通信方式。

地址解析协议 ARP，网际控制报文协议 ICMP，网际组管理协议 IGMP。

知识点：数据链路层使用网桥将以太网连接成桥接以太网，网络层使用路由器将不同网络连接起来。

分配给每一台主机或路由器的每一个接口的一个 32 位的全球唯一的标识符。

IP 地址由网络号部分和主机号部分组成，路由器在转发分组时只根据网络号转发。

IP 地址可以分为单播地址和多播地址。

从存储位置分：IP 地址存储在主机内存中，硬件地址存储在网络适配器中。IP 地址用于网络层以上的传输，硬件地址用于物理层和数据链路层的传输。

从 IP 地址中解析出硬件地址。

ARP 作用范围是一个局域网内。

每台主机上都有 ARP 高速缓存，其中存储本网络中每台主机与路由器的 IP 地址到硬件地址的映射。如果路由器或主机 A 想要向主机 B 发送消息而它没有存储主机 B 的硬件地址，就通过广播方式询问获取主机 B 的地址。

源地址、目的地址、版本号、首部长度、总长度、片偏移、用于区分分片的标识、用于表示是否是最后一个分片的标志、携带的数据采用的协议、区分服务，首部检验和。

路由表中存储目的网络地址和对应的下一跳地址。

根据路由表中的存储的下一跳地址一跳一跳地到达目的网络。

问题 4.3~4.8

什么是无分类编址 CIDR

CIDR 中路由器如何选择下一跳地址

使用 CIDR 的路由表采用了什么算法来快速查找

网际控制报文协议 ICMP 的作用是什么

ICMP 报文分为几类？

ICMP 的两个典型应用

回答 4.3~4.8

它不再对地址进行分类，而是采用子网掩码来识别网络号。子网掩码也是 32 位，它的前 n 位是 1，通过前 n 位等于 1 的子网掩码与一个 IP 地址相结合就可以得到它的 n 位的网络地址。

路由器存储着每个目的网络地址及对应的子网掩码，将子网掩码与目的主机 IP 地址相与，得到的结果如果与目的网络地址相同，就把分组转发给该条目对应的下一跳地址。

它使用了二叉线索树来存储路由表，二叉树的每一层对应 IP 地址中的一位，最多 32 层。

用于更有效地转发 IP 数据报和提高交付的机会

分为差错报告报文和查询报文两大类，差错报告报文包括终点不可达报文、时间超过报文、参数问题报文、改变路由报文等，查询报文包括回送请求或回答报文、时间戳请求或回答报文。

分组网间探测 PING：采用了回送请求和回送回答报文。tracert：用于跟踪分组经过的路径。通过发送一连串具有不同生存时间的无法交付的 UDP 数据报，采用了差错报文报文。

第 4 章 网络层

4.1 网络层提供的两种服务

网络层提供的服务可以是”面向连接“的或是”无连接“的服务。
用于打电话的传统电信网使用面向连接的通信方式，它先建立连接预留出网络资源（建立一条虚电路），然后再传送信息，提供的是可靠传输的服务。
互联网采用的是**无连接**的方式，发送分组时不需要建立连接。
网络层向上只提供简单灵活的、无连接的、尽最大努力交付的数据报服务。这样使造价降低，运行灵活。

4.2 网际协议 IP

这里的 IP 是 IP 协议的第 4 个版本，实际叫做 IPv4。较新的还有 IPv6。
与 IP 协议配套的还有三个协议：

地址解析协议 ARP
网际控制报文协议 ICMP
网际组管理协议 IGMP

4.2.1 虚拟互连网络

不同网络的区别很大，因为没有一种单一的网络能够适应所有用户的需求。因此需要通过一些中间设备将网络连接起来。

根据所在层次，可以将中间设备分为以下四种：

物理层使用的叫**转发器**。

数据链路层使用的叫**网桥或桥接器**。

网络层使用的叫**路由器**。

网络层以上使用的叫网关。

在物理层使用转发器或在数据链路层使用网桥时，仅是把一个网络扩大了。从网络层看，这还是一个网络，不是网络互连。

网络互连是在网络层通过路由器实现的。

都使用 IP 协议的网络互连以后叫虚拟互连网络，含义是这些在物理层面不同的网络在网络层看起来好像是一个统一的网络，又叫 **IP 网**。

现在的互联网就是使用了 IP 协议和 TCP 协议。

4.2.2 分类的 IP 地址

IP 地址及其表示方法

整个互联网就是一个单一的、抽象的网络。

IP 地址就是给互联网上每一台主机或路由器的**每一个接口**分配一个全世界唯一的 32 位的标识符。

IP 地址的编址方法经历了三个阶段：

分类的 IP 地址。

子网的划分。

构成超网。

分类的 IP 地址就是讲 IP 地址划分为多个固定类，每一类地址由两个固定长度的字段组成。

第一个字段是网络号，标志主机所连接到的网络，一个网络号在整个互联网范围内是唯一的。

第二个字段是主机号，标志该主机（或路由器），一个主机号在该网络号所指明的网络范围内是唯一的。

IP 地址 = {<网络号>，<主机号>}，它既指明了主机接口，也指明了所在网络。

分类的 IP 地址分为以下 5 类：

A 类、B 类、C 类都是单播地址，是最常用的。

D 类是用于多播（一对多通信）。

E 类地址保留为以后用。

分类是考虑到了不同网络间的差异性，有的网络主机很多，有的则很少。

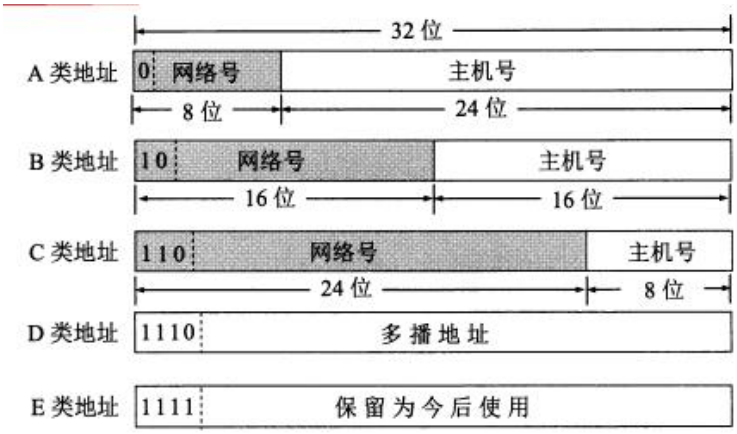


图 4-5　IP 地址中的网络号字段和主机号字段

现在广泛使用**无分类 IP 地址**进行路由选择，分类的地址已经成为历史。

IP 地址是 4 字节共 32 位字符，平常电脑上显示的是每个字节按转化为 10 进制后的结果，称为**点分十进制法**。

图 4-6　采用点分十进制记法能够提高可读性

IP 地址有以下几个特点：

每一个 IP 地址都是由网络号和主机号两部分组成，是一种分等级的地址结构。这种结构有几个优点

IP 地址管理机构在分配 IP 地址时只分配网络号，而主机号由得到网络号的单位内部自行分配。

路由器仅根据目的主机所连接的网络号来转发分组而不考虑主机号。这使路由表中的项目数大幅减少，减小了路由表的存储空间和查找时间。

IP 地址是标志一台主机（或路由器）和一条链路的接口。如果一台主机同时连接到两个网络，它就有两个 IP 地址。

每个路由器至少连接到两个网络，所以一个路由器至少有两个不同的 IP 地址。

互联网中，一个网络指的是具有相同网络号的主机的集合。所以用转发器或网桥连接起来的若干局域网仍是一个网络。

IP 地址中，所有分配到网络号的网络都是平等的，不管它的范围多大或多小。

4.2.3 IP 地址与硬件地址

硬件地址（又称物理地址、MAC 地址）是数据链路层和物理层使用的地址。MAC 帧传送时使用的源地址和目的地址都属于硬件地址，放在 MAC 帧的首部。

IP 地址是网络层和以上各层使用的地址，是一种**逻辑地址**。放在 IP 数据报的首部。

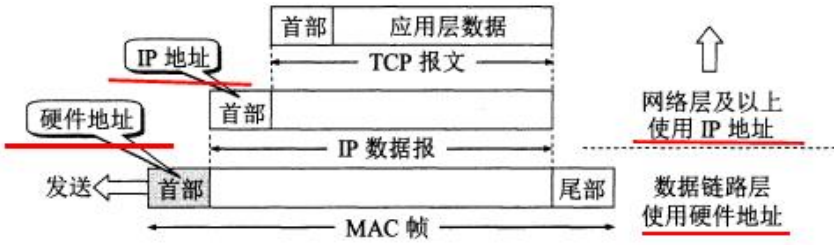


图 4-8　IP 地址与硬件地址的区别

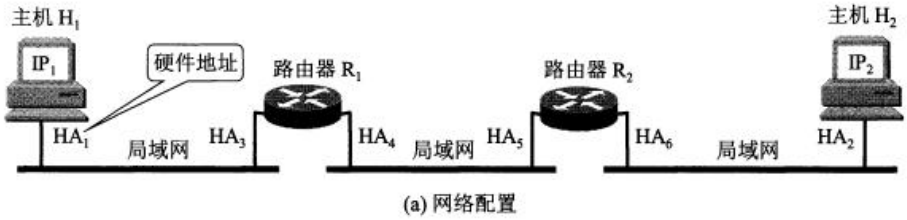
下面是三个局域网通过两个路由器连接在一起，主机 H1 要与主机 H2 通信。
路由器因为同时连在两个局域网上，所以有两个硬件地址。
注意：

在 IP 层抽象的互联网上只能看到 IP 数据报。虽然信息要经过路由器 R1 和 R2 的转发，但是 **IP 报首部中的源地址和目的地址始终是 IP1 和 IP2。**

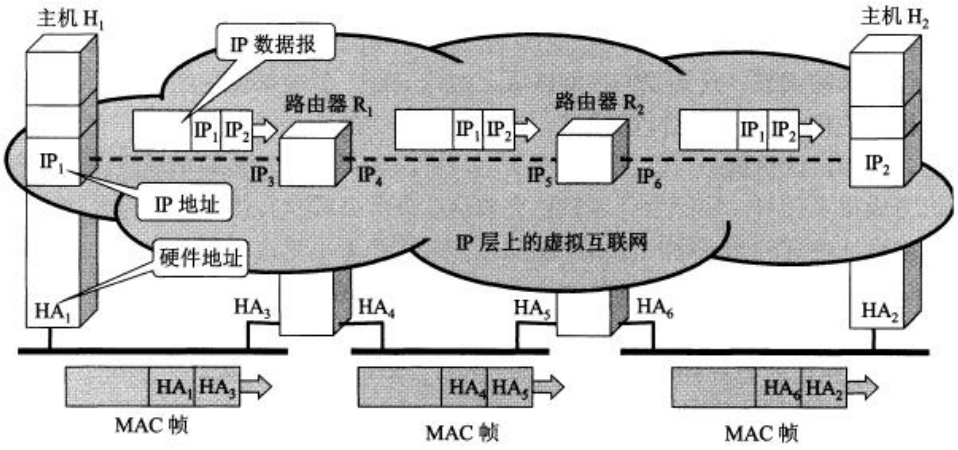
虽然 IP 数据报首部有源地址，路由器只根据目的站的 IP 地址的网络号进行选择。

在局域网的链路层，只能看见 MAC 帧。**MAC 帧在不同网络上传送时，MAC 帧首部中的源地址和目的地址要发生变化。**

IP 层抽象的互联网屏蔽了下层的复杂细节。只要在网络层上，就可以使用统一的、抽象的 IP 地址来研究主机和主机间的通信。



(a) 网络配置



(b) 不同层次、不同区间的源地址和目的地址

4.2.4 地址解析协议 ARP

网络层用的是 IP 地址，但实际网络的链路上传送数据帧时还是要用硬件地址。当数据传到不同网络时，MAC 帧中的硬件地址还会发生改变，主机或路由器怎么知道该在 MAC 帧的首部中填入什么硬件地址呢？

ARP 协议的用途是从网络层使用的 IP 地址解析出数据链路层使用的硬件地址。

根据地址解析协议 ARP，每台主机都有一个 **ARP 高速缓存**，里面有本局域网上的各主机和路由器的 IP 地址到硬件地址的映射表。

主机的硬件地址可能会发生改变，因此该映射表会时常更新，映射表中的每个项目都有生存时间，超过生存时间的项目会被删掉。

当主机 A 向本局域网上的主机 B 发送 IP 数据报时有两种情况：

主机 A 的 ARP 高速缓存的映射表中有主机 B 的 IP 地址，就把对应的硬件地址写入 MAC 帧，然后通过局域网把该 MAC 帧发给此硬件地址。

主机 A 的 ARP 高速缓存中没有 B 的 IP 地址。此时按以下步骤找出 B 的硬件地址：

主机 A 自动运行 ARP 进程，ARP 进程在本局域网上广播发送一个 **ARP 请求分组**。分组中指明了**自己的 IP 地址与硬件地址，和主机 B 的 IP 地址。**

本局域网上所有主机上运行的 ARP 进程都受到这个 ARP 请求分组。

主机 B 的 IP 地址与 ARP 请求分组中要查询的 IP 地址一致，收下分组，并向主机 A 发送响应分组，在其中写入自己的硬件地址。其他主机则不作响应。

主机 A 收到主机 B 的响应分组后，把主机 B 的 IP 地址到硬件地址的映射写入 ARP 高速缓存中。

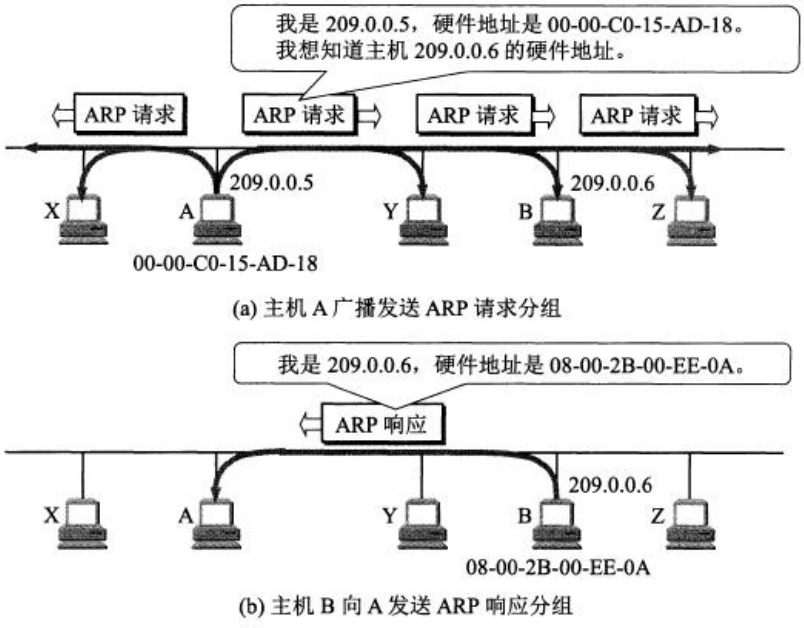


图 4-11　地址解析协议 ARP 的工作原理

ARP 解决的是**同一个局域网上**主机或路由器的 IP 地址到硬件地址的映射问题。它无法解析另一个局域网上主机的硬件地址，实际上也不用。

使用 ARP 的四种典型情况

发送方是主机，要把 IP 报发到同一个网络上的另一台主机，如从 H1 到 H2，这时 H1 发送 ARP 请求分组（在网 1 上广播）找到 H2 的硬件地址。

发送方是主机，要把 IP 报发到另一个网络上的某一台主机，如从 H1 到 H3，这时 H1 发送 ARP 请求分组（在网 1 上广播）找到网 1 上的一个路由器 R1 的硬件地址。剩下的事情由 R1 完成。

发送方是路由器，要把 IP 报转发到与它连接在同一个网络上的主机，如 R1 到 H2，这时 R1 发送 ARP 请求分组（在网 2 上广播）找到主机 H3 的硬件地址。

发送方是路由器，要把 IP 报转发到另一个网络上的一台主机，如 R1 到 H3，这时这时 R1 发送 ARP 请求分组（在网 2 上广播）找到本网络上另一个路由器的硬件地址。
总的来说，当发送方与接收方不在同一个网络时，要通过同时位于两个或更多个网络上的路由器来中转，而 ARP 协议则用于每个局域网内部的地址解析。

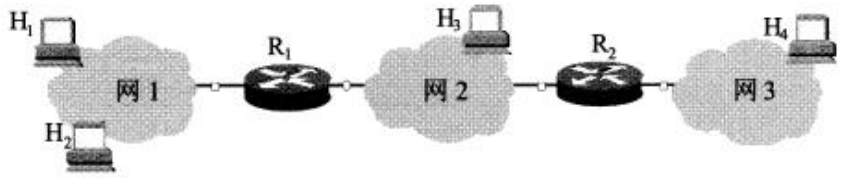


图 4-12　使用 ARP 的四种典型情况

4.2.5 IP 数据报的格式

一个 IP 数据报的**首部包括两部分**，前一部分是固定长度，共 20 字节。后面是一些可选字段，长度可变。

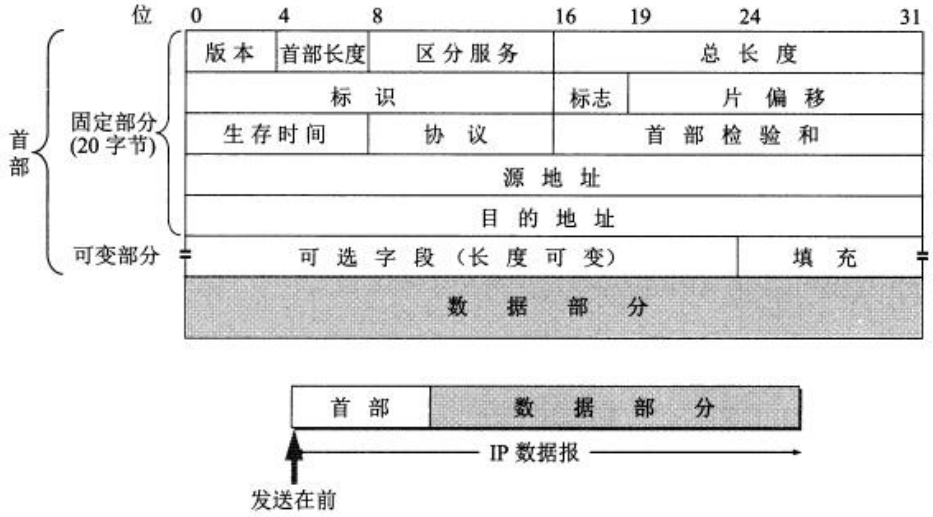


图 4-13　IP 数据报的格式

IP 数据报首部的固定部分的各字段

版本。占 4 位，通信双方使用的 IP 协议的版本必须一致。

首部长度。 占 4 位，最大值是 15，注意其单位是 4 字节，也就是首部最大长度为 15*4=60 字节。首部长度必须是 4 字节的整数倍。因此可选字段后面还有一个填充字段。

区分服务。占 8 位，一般不使用，只有使用**区分服务 DiffServ** 时此字段才有意义，根据字段的数值为提供不同等级的服务质量。

总长度。占 16 位，最大值是 65535，是首部和数据部分的长度和，单位是字节。IP 数据报的长度还受到 MAC 帧最大长度的限制，因此不能太大。如果长度过长需要进行分片。分片后总长度指的是该分片的首部长度和数据长度之和。

标识。 占 16 位。同一个数据报的不同分片标识相同。因此接收方能根据标识将不同分片重装为原本的数据报。

标志。占 3 位。最低位为 1 表示后面还有分片，为 0 表示这是最后一个分片。中间位为 1 表示不能分片，为 0 表示可以分片。首位没有含义。

片偏移。占 13 位。片偏移指出：较长的分组分片后，某片在原分组的相对位置。单位是 8 字节，故每个分片的长度是 8 字节的整数倍。

生存时间。占 8 位。表明数据报在网络中的寿命。单位是跳数，指明了数据报在互联网中至多可经过多少个路由器。

协议。占 8 位。指明了数据报携带的数据使用了哪种协议。

首部检验和。占 16 位。这个字段只检验首部，不包括数据部分。数据报每经过一个路由器，路由器就要重新计算一下首部检验和。

源地址。占 32 位。

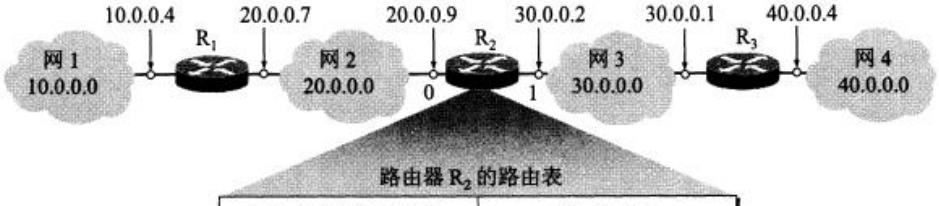
目的地址。占 32 位。

IP 数据报首部的可变部分

长度可变，具有多种功能，但很少使用。IPv6 已经把这部分做成固定长度的了。

4.2.6 IP 层转发分组的流程

路由器的路由表中不直接存储主机地址，而是**存储目的网络的地址和对应下一跳的地址。**路由表中并没有指明完整的网络路径，仅指出要想到达某个网络，需要先到哪个路由器，即仅指出下一步该怎么走。这样一跳一跳直到最后到达目的网络。



路由器 R_2 的路由表

目的主机所在的网络	下一跳地址
20.0.0.0	直接交付，接口 0
30.0.0.0	直接交付，接口 1
10.0.0.0	20.0.0.7
40.0.0.0	30.0.0.1

(a) 路由器 R_2 的路由表

分组转发算法如下：

从数据报的首部提取出目的主机的 IP 地址 D，得出目的网络地址为 N。

若 N 就是与此路由器直接相连的某个网络地址，就直接交付，即直接把数据报交付目的主机；否则就是间接交付，执行 3。

若路由表中有目的地址为 D 的特定主机路由，则把数据报传送给路由表中所指明的下一跳路由器。否则执行 4。

若路由表中有到达网络 N 的路由，则把数据报传送给路由表所指明的下一跳路由器，否则执行 5。

若路由表有一个默认路由，则把数据报传送给路由表所指明的默认路由器，否则执行 6。

报告转发分组出错。

4.3 划分子网和构造超网

4.3.1 划分子网

之前是两级 IP 地址，缺点很多。后在 IP 地址中又增加了一个子网号字段，将 IP 地址分为了三级。
划分子网是把 IP 地址的主机号再划分，未改变网络号。

子网掩码

划分子网后，IP 数据报的首部无法体现是否进行了划分。需要使用**子网掩码**。
现在的互联网规定所有的网络都必须使用子网掩码，**路由器的路由表中也必须有子网掩码这一栏。**
路由器在和相邻路由器交换信息时，必须把自己所在子网的子网掩码告诉对方。

4.3.2 使用子网时分组的转发

使用子网划分后，路由表中必须包含目的网络地址、子网掩码和下一跳地址三项内容。
此时的分组转发算法如下：

从数据报的首部提取出目的主机的 IP 地址 D。

判断是否为直接交付。对路由器直接相连的网络逐个检查：用各网络的子网掩码逐个与 D 按位相与，看结果是否和相应的网络地址匹配。若匹配，就直接交付；否则就是间接交付，执行 3。

若路由表中有目的地址为 D 的特定主机路由，则把数据报传送给路由表中所指明的下一跳路由器。否则执行 4。

对路由表中的每一行，用其中的子网掩码和 D 逐位相与，若结果与该行的**目的网络地址**匹配，则把数据报传送给该行指明的下一跳路由器，否则执行 5。

若路由表有一个默认路由，则把数据报传送给路由表所指明的默认路由器，否则执行 6。

报告转发分组出错。

4.3.3 无分类编址 CIDR(构造超网)

无分类编址全名**无分类域间路由选择 CIDR。**
CIDR 有两个主要特点：

CIDR 消除了传统的 A,B,C 类地址和划分子网的概念，它把 32 位的 IP 地址分为前后两部分。“前缀”用来指明网络，后面部分用来指明主机。因此它使用的是两级编址，但是是**无分类的两级编址。**

CIDR 把前缀都相同的连续的 IP 地址组成一个 CIDR 地址块，只要知道该地址块中的任意一个地址，就可以知道地址块的起始地址、最大地址和地址块中的地址数。
CIDR 使用 32 位的地址掩码，地址掩码中 1 的个数对应的就是前缀的长度。前缀越短，其地址块包含的地址数越多。
使用 CIDR 可以更有效地分配地址空间。

最长前缀匹配

CIDR 中，路由表的每个项目由网络前缀和下一跳地址组成，查找时可能得到不止一个匹配结果。这是从匹配结果中选择具有最长网络前缀的路由，因为它对应的地址块最小。

使用二叉线索从查找路由表

无分类编址的路由表通常存放在一个**二叉线索树**中。
下图的二叉线索树表示了一个有 5 个 IP 地址的路由表。树的每一层对应 IP 地址中的一位，树最多有 32 层。
给定一个 IP 地址，查找它是否在该项目表中，只需在二叉线索树中一层层对应向下寻找，若中间无法在二叉树中找到对应分支，表明这个地址不在这个二叉线索中。

32 位的 IP 地址	唯一前缀
01000110 00000000 00000000 00000000	0100
01010110 00000000 00000000 00000000	0101
01100001 00000000 00000000 00000000	011
10110000 00000010 00000000 00000000	10110
10111011 00001010 00000000 00000000	10111

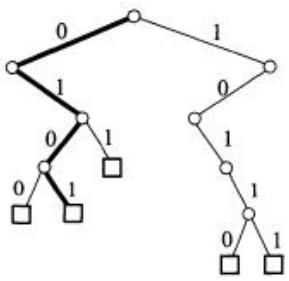


图 4-26　用 5 个前缀构成的二叉线索

4.4 网际控制报文协议 ICMP

网际控制报文协议 ICMP 用于更有效地转发 IP 数据报和提高交付成功的机会。
ICMP 报文装在 IP 数据报中，作为其中的数据部分。
ICMP 报文的首部共 8 个字节，具体如下图。
其中检验和字段用来检验整个 ICMP 报文。

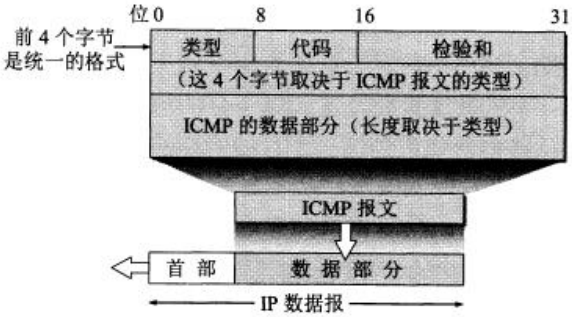


图 4-27　ICMP 报文的格式

4.4.1 ICMP 报文的种类

ICMP 报文包括 **ICMP 差错报告报文**和 **ICMP 询问报文**两类，每类下细分为几种不同的类型。

表 4-8　几种常用的 ICMP 报文类型

ICMP 报文种类	类型的值	ICMP 报文的类型
差错报告报文	3	终点不可达
	11	时间超过
	12	参数问题
	5	改变路由(Redirect)
询问报文	8 或 0	回送(Echo)请求或回答
	13 或 14	时间戳(Timestamp)请求或回答

表中给出了 4 种常用的 ICMP 差错报告报文：

终点不可达：当路由器或主机不能交付数据报时就向源点发送此报文。

时间超过：路由器收到生存时间为 0 的报文时，除丢弃该数据报外，还要向源点发送此报文。当终点在约定时间内未收到一个数据报的全部分片时，就丢弃已收到的所有分片，并向源点发送此报文。

参数问题：当路由器或目的主机收到的数据报的首部中有的字段值不正确时，就丢弃该数据报并发送此报文。

改变路由（重定向）：路由器把此报文发送给主机，以告诉主机下次将数据报发给另外的路由器。

另外 2 种常用的 ICMP 询问报文：

回送请求报文和回答报文：回送请求报文是由主机或路由器向一个特定目的主机发出的询问。收到此报文的主机必须给源主机发送 ICMP 回送回答报文。

时间戳请求报文和回答报文：时间戳请求报文是请某台主机或路由器回答当前的日期和时间。通过它可以进行时钟同步和时间测量。

ICMP 差错报告报文的**数据字段是固定格式**的：把收到的需要进行差错报告的 IP 数据报的首部和数据字段的前 8 个字节（为了得到运输层的端口号和运输层报文的发送序号）提取出来作为 ICMP 报文的数据部分。

4.4.2 ICMP 的应用举例

PING

ICMP 的一个重要应用是进行**分组网间探测 PING（Packet InterNet Groper）**，以测试两台主机之间的连通性。

PING 使用了 ICMP 回送请求和回送回答报文。它会连续发送 4 条回送请求报文。

PING 是应用层直接使用 ICMP 的例子，未经过运输层。

使用方法：在 Windows 的 Dos 窗口中键入 **ping hostname** 即可测试本机与主机 hostname 之间的连通性，hostname 应该是某个主机的 IP 地址或域名

ping www.baidu.com;//测试与百度之间的连通性

ping 192.168.100.5;//测试与 IP 地址为 192.168.100.5 的之间的连通性

tracert

tracert 可以用来跟踪一个分组从源点到终点的路径。

tracert 从源主机向目的主机发送一连串的 IP 数据报。数据报中封装的是**无法交付的 UDP 用户数据报**。

这些数据报中，第一个数据报的生存时间 TTL 设为 1，后面依次增长。当第 i 个数据报到达了路径上的第 i 个路由器，其 TTL 也减到了 0，此时该路由器就会发送 ICMP 时间超过差错报告报文给源主机。由此就可以获得到达目的主机所经过的所有路由器的 IP 地址，以及到达每一个路由器的往返时间。

使用方法：在 Windows 的 Dos 窗口中键入 **tracert hostname** 即可测试本机到主机 hostname 所经过的路由器。

4.5 互联网的路由选择协议

4.5.1 有关路由选择协议的几个基本概念

路由选择协议的核心是采用何种算法来获得路由表中的各项目。

路由选择算法可以分为静态路由选择策略和动态路由选择策略。

其中动态的可以较好地使用网络状态的变化，但实现起来较复杂，适用于大网络。互联网采用的主要是**动态的、分层次的路由选择协议**。

分层次的路由选择协议

互联网被划分为许多小的自治系统（AS），一个 AS 是在单一技术管理下的一组路由器，一个 AS 对另一个 AS 表现出的是一个单一的和一致的路由选择策略。目前互联网中，一个大的 ISP 就是一个 AS。也可以进一步划分。

这样互联网就把路由选择协议分为了两类：

内部网关协议 IGP：如 **RIP** 和 **OSPF** 协议，是在一个 AS 内部使用的路由选择协议。每个 AS 自己决定在自己内部使用哪一种 IGP。

外部网关协议 EGP：如 **BGP-4** 协议，用在两个不同的 AS 之间的路由选择协议。每个 AS 中位于与其他 AS 交界处的路由器除了使用 IGP 外还要使用 EGP。

4.5.2 内部网关协议 RIP

RIP 是一种**分布式的基于距离向量的路由选择协议**，最大优点是简单。

RIP 协议要求每一个路由器都要维护从它自己到其他每一个目的网络的距离记录，距离的单位是跳数。**RIP 选择一条具有最少路由器的路径。**

RIP 允许一条路径最多有 15 个路由器，因此 RIP 只适用于小型互联网。

RIP 和 OSPF 同为分布式路由选择协议，特点是**每一个路由器都要不断地和其他路由器交换路由信息。**

RIP 的特点是：

仅和相邻路由器交换信息。

交换的信息是当前本路由器知道的所有信息，即自己的路由表。

按固定的时间间隔交换路由信息，比如 30s。

路由器刚开始的路由表是空的，通过不断地和与它直接相连的路由器交换并更新信息，经过多次更新后，所有的路由表就都会知道到达本自治系统中任何一个网络的最短距离和下一跳路由器的地址了。

路由表中最主要的信息就是**到某个网络的最短距离和下一跳地址**。

4.5.3 内部网关协议 OSPF

OSPF 是分布式的链路状态协议，适用于大型互联网。OSPF 只**在链路状态发生变化时**，才向本自治系统中的所有路由器**用洪泛法**发送**与本路由器相邻的所有路由器的链路状态信息**。

链路状态指明本路由器和哪些路由器相邻，以及该链路的度量（度量可表示费用、距离、时延、带宽等），所有的路由器最终都能建立一个全网的拓扑结构图。

4.5.4 外部网关协议 BGP

BGP 是 **BGP-4** 的简写。BGP 是**不同 AS 的路由器之间**交换路由信息的协议，是一种路径向量路由选择协议。BGP 力求寻找一条能够到达目的网络（**可达**）且比较好的路由（**不兜圈子**），而非寻找最佳路由。

4.5.5 路由器的构成

4.6 IPv6

IPv4 的地址已经耗尽。

IPv6 目前尚未推出标准协议。

4.6.1 IPv6 的基本首部

IPv6 将**协议数据单元 PDU** 称为**分组**，而非 IP 数据报。

IPv6 的主要变化：

更大的地址空间：地址位数增大到了 128 位。

扩展的地址层次结构：因为地址空间很大，所以可以划分更多层。

灵活的首部格式：IPv6 的首部和 IPv4 的首部不兼容。

改进的选项。IPv6 的首部长度是固定的，把选项放在了有效载荷中。

允许协议继续扩充。

支持自动配置，也就是不需要 DHCP 协议。

支持资源的预分配。

首部改为 8 字节对齐：即首部长度应为 8 字节的整数倍。

IPv6 数据报分为基本首部和有效载荷。有效载荷中允许有 0 个或多个**扩展首部**。注意扩展首部不属于首部。

IPv6 的首部包括：

版本。

通信量类。

流标号。

有效载荷长度。

下一个首部。

跳数限制。等同于 TTL。

源地址。
目的地址。

4.6.2 IPv6 的地址

IPv6 数据报的目的地址可以是以下三种之一：

单播：点对点通信。

多播：一点对多点的通信。

任播：任播的终点是一组计算机，但是数据报只交付其中一个，一般是最近的一个。

IPv6 地址有 128 位，采用**冒号十六进制计法**：每 16 位用 16 进制表示并用冒号隔开，因此共分为了 8 段，每段是一个不超过 4 位的 16 进制数。

4.6.3 从 IPv4 向 IPv6 过渡

两种从 IPv4 向 IPv6 过渡的策略：**双协议栈**和**隧道技术**。

双协议栈

双协议栈即使一部分主机或路由器装有**双协议栈：一个 IPv4 和一个 IPv6**，当它与 IPv6 主机通信时使用 IPv6 地址，与 IPv4 主机通信时使用 IPv4 地址。

双协议栈**使用域名系统 DNS 来查询目的主机使用哪一种地址**。

隧道技术

当源主机和目的主机都采用 IPv6 时，中间经过的网络有可能是 IPv4 网络。

在 IPv6 数据报要进入 IPv4 网络时，把 IPv6 数据报作为数据部分封装到 IPv4 数据报中，等离开 IPv4 网络后在把数据部分取出来。

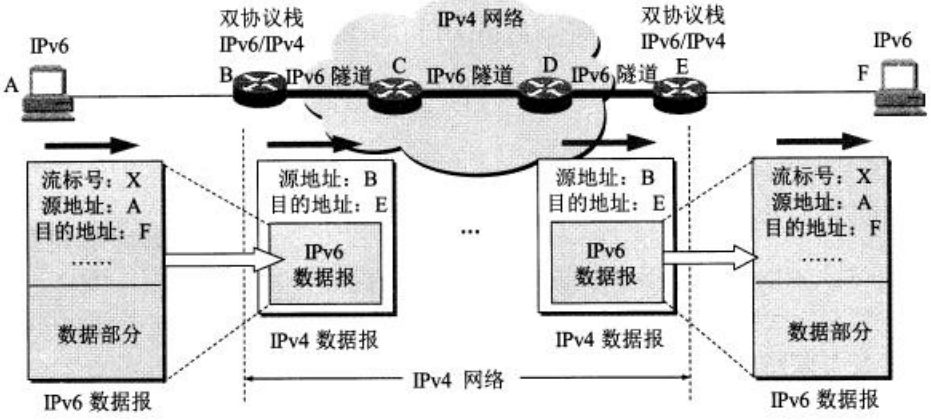
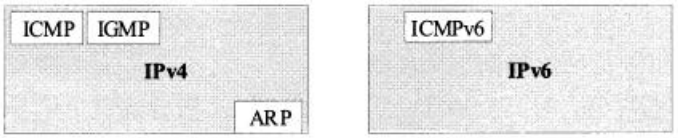


图 4-50　使用隧道技术进行从 IPv4 到 IPv6 的过渡

4.6.4 ICMPv6

ICMPv6 是应用于 IPv6 的 ICMP 协议版本，比 ICMPv4 复杂很多，**地址解析协议 ARP 和网际组管理协议 IGMP 的功能都合并到了 ICMPv6 中**。

ICMPv6 是面向报文的协议，利用报文来报告差错、获取信息。



版本 4 中的网络层　　版本 6 中的网络层

图 4-51　新旧版本中的网络层的比较

4.7 IP 多播

4.7.1 IP 多播的基本概念

在一对多的通信中，多播可以比单播节省很多资源。

局域网具有硬件多播功能，所以不需要复制分组就能使所有的多播组成员收到分组。

IP 多播所传送的分组需要使用多播 IP 地址。在传统的 IP 地址中的 D 类地址就是多播地址，每个 D 类地址可以标识一个**多播组**。

多台主机可以**加入到一个多播组中共享一个多播地址**。不同网络的主机可以加入到同一个多播组中。**每一台主机可以随时加入或离开一个多播组。**

能够运行多播协议的路由器为多播路由器。

多播地址只能用于目的地址，不能用于源地址。

IP 多播有两种

在本局域网上硬件多播。

在互联网范围内进行多播。

4.7.2 在局域网上进行硬件多播

多播 IP 地址与以太网硬件地址之间有映射关系，但不是一一对应的。收到多播数据报的主机需要在 IP 层进行过滤。

4.7.3 网际组管理协议 IGMP 和多播路由选择协议

IGMP 协议用于让连接在本地局域网上的多播路由器知道本局域网上是否有主机参加或退出了某个多播组。IGMP 工作在单个本地局域网内部。

因为主机随时可能加入或退出某个多播组。并且发送多播数据报的主机可以位于多播组内，也可以不位于多播组内。所以 IP 多播很复杂。

IGMP 协议的工作内容：

当某台主机加入某个多播组时，该主机向多播组的多播地址发送一个 IGMP 报文，声明自己成为了该组的成员。本地的多播路由器收到 IGMP 报文后还要转发给其他多播路由器。

组成员关系是动态的。本地多播路由器要周期性地探询本地局域网上的主机是否还是组的成员。

多播路由选择协议

多播路由选择协议有多种，尚未进行标准化。

多播路由选择协议在转发多播数据报时有以下三种方法：

洪泛与剪除。

隧道技术。

基于核心的发现技术。

目前的多播路由选择协议：

距离向量多播路由选择协议 DVMRP

基于核心的转发树 CBT

开放最短通路优先的多播扩展 MOSPF

协议无关多播-稀疏方式 PIM-SM

协议无关多播-密集方式 PIM-DM

4.8 虚拟专用网 VPN 和网络地址转换 NAT

4.8.1 虚拟专用网 VPN

因为 IP 地址的紧缺。所以现在使用了一种**“本地地址”**。**本地地址仅在本机构内部有效**，不是全球唯一的地址，又称**可重用地址**。

本地地址只能用于一个机构的内部通信，不能和互联网行的主机通信。互联网中的所有路由器对目的地址是**专用地址**的数据报一律不转发。

IPv4 标准指明了以下地址为**专用地址**，他们只能作为本地地址用于机构内部的通信：

10.0.0.0-10.255.255.255

172.16.0.0-172.31.255.255

192.168.0.0-192.168.255.255

采用专用地址的网络称为专用互联网或**本地互联网**。

有时一个机构的分布范围很广，就需要用公共的互联网作为**本机构各专用网之间的通信载体**，这样的称为**虚拟专用网 VPN**。

VPN 依然只用于机构内的通信，但是要经过公用的互联网，通过互联网传送的数据都要加密。这里使用了隧道技术。

VPN 中每个不同的场所必须至少有一个合法的全球 IP 地址。

VPN 代理就是依托 VPN 技术进行的。

4.8.2 网络地址转换 NAT

网络地址转换 NAT 用于实现专用网中的主机到互联网上的主机的通信。

它需要在专用网连接到互联网的路由器上安装 NAT 软件，这种路由器称为 NAT 路由器，**NAT 路由器至少有一个全球地址。**

使用本地地址的主机和外界通信时要在 NAT 路由器上将本地地址转换为全球地址。

NAT 路由器中有一个地址转换表，存储本地地址与转换后的全球地址的对应关系。

通过 NAT 路由器的通信必须由专用网内的主机发起，因此**专用网内的主机不能作为服务器**。

现在的 NAT 转换表把端口号也利用上了。这样 NAT 路由器只需要有一个全球地址，通过给具有不同本地地址的主机分配不同的端口号就可以实现内部多个主机与外界互联网的通信。

4.9 多协议标记交换 MPLS

第 5 章 运输层

问题

运输层服务对象
端口和套接字的意义
UDP 的特点
TCP 的特点
可靠传输的工作原理
停止等待协议如何工作
停止等待协议的信道利用率很低，采用什么方法提高？
什么是 ARQ 协议

TCP 的滑动窗口是什么？有什么作用？

什么是流量控制？

回答

不同主机的应用进程。

端口用于区分同一主机上应用层的不同应用进程。IP 地址加上端口就是套接字，套接字是运输层传输的端点。

UDP 的特点：无连接、尽最大努力交付、可以一对一或一对多或多对一或多对多、首部简单、面向报文。

TCP 的特点：面向连接、可靠传输、仅支持点对点通信、全双工通信、面向字节流、有流量控制、拥塞控制等功能。

可靠传输依靠停止等待协议来实现。

停止等待协议的实现方式是超时重传。超时未收到确认信息就重传分组。

使用连续 ARQ 协议和滑动窗口协议。

自动重传请求 ARQ 是发送端每发送一个报文就维持一个超时计时器，它的时长比平均往返时间长一些，到时就重传。

包括发送端的发送窗口和接收端的接收窗口，发送窗口内的信息可发送，**发送窗口后沿收到确认号就前移至确认号位置，前沿受后沿位置及接收方发来 TCP 报文首部中的窗口字段影响**。滑动窗口可以提高传输效率（减少确认的发送频率）、帮助实现流量控制的功能。

当发送方发送的太快使接收方来不及接收，需要控制发送方的发送速率。

问题

TCP 的流量控制如何实现？

什么是拥塞？拥塞控制的原理是什么？

判断出现拥塞的依据是什么？

流量控制和拥塞控制的区别？

TCP 的拥塞控制如何实现？

TCP 的连接过程是怎样的？

TCP 的释放过程是怎样的？

回答

通过 TCP 报文首部中的窗口字段和滑动窗口来实现。发送方的发送窗口不能超过确认报文中的窗口字段值。当接收方剩余的缓存不足时，就把确认报文中的窗口字段调小，使发送方下次发送的报文长度减小。

拥塞就是当前对网络中某一资源的需求超过了该资源所能提供的可用部分，网络性能变坏。**拥塞控制就是防止过多的数据注入到网络中**。

判断网络拥塞的依据是出现超时。这里的超时和重传不一样。

流量控制是端到端的问题，是接收端抑制发送端发送数据的速率。**拥塞控制**是全局性的问题，涉及到所有的主机、路由及相关因素。

TCP 的拥塞控制是**基于窗口的拥塞控制**。在控制拥塞窗口的大小时，采取了**慢开始、**

拥塞避免、快重传、快恢复四种算法。慢开始的慢是起点低，但是增长快，指数增长。拥塞避免是窗口超过界限值后就增长变慢，等差增长。快重传目的是快速重传避免超时。快恢复是超时后快速恢复。

三次握手建立连接：连接请求报文（SYN=1），连接响应报文（SYN=1，ACK=1），第三个报文（ACK=1，此报文没有数据部分时不占序列号）

四次挥手释放连接：A 首先发出**连接释放报文**、B 发回**确认报文**，然后 B 继续发送未发送完的数据，发完后 B 再一个**连接释放报文（ACK =1）**，A 返回一个**确认报文**，B 收到后就关闭，A 等待一段时间后也关闭（避免超时重传情况）。

第 5 章 运输层

5.1 运输层协议概述

5.1.1 进程之间的通信

网络层为主机之间提供通信，运输层为**应用进程**提供**端到端**的逻辑通信。
通信的真正端点是**主机中的进程**，即应用进程之间的通信是端到端的通信。

运输层的复用和分用

即发送方的**不同进程通过不同的端口号使用同一个运输层协议**，接收方的运输层则把收到的报文根据端口号分发给不同的进程。

5.1.2 运输层的两个主要协议

运输层的两个主要协议是 **传输控制协议 TCP** 和 **用户数据报协议 UDP**，他们都有复用和分用，和检错的功能。
UDP 的特点：无连接、尽最大努力交付、面向报文、无拥塞控制、支持一对一、一对多、多对一、多对多，**首部开销小**。
TCP 的特点：面向连接的、点对点通信、提供可靠传输、全双工通信、**面向字节流**。

UDP

接收方的主机收到 UDP 后不需要发出确认。

TCP

TCP 传送数居前要建立连接，传送完成后要释放连接。
TCP 不提供广播或多播服务。
因为 TCP 的功能较多，所以首部很长，且占用处理器资源。

表 5-1　使用 UDP 和 TCP 协议的各种应用和应用层协议

应用	应用层协议	运输层协议
名字转换	DNS（域名系统）	UDP
文件传送	TFTP（简单文件传送协议）	UDP
路由选择协议	RIP（路由信息协议）	UDP
IP 地址配置	DHCP（动态主机配置协议）	UDP
网络管理	SNMP（简单网络管理协议）	UDP
远程文件服务器	NFS（网络文件系统）	UDP
IP 电话	专用协议	UDP
流式多媒体通信	专用协议	UDP
多播	IGMP（网际组管理协议）	UDP
电子邮件	SMTP（简单邮件传送协议）	TCP
远程终端接入	TELNET（远程终端协议）	TCP
万维网	HTTP（超文本传送协议）	TCP
文件传送	FTP（文件传送协议）	TCP

5.1.3 运输层的端口

运输层使用 16 位（即两字节）端口号来标志一个端口。端口号用来标志本计算机应用层中的不同进程。不同计算机间的端口没有关联。
这里的端口是软件端口，作为交互的地址使用，不同于路由器上的硬件端口。

端口号的分配

运输层的端口号分为服务器端使用的端口号和客户端使用的端口号。
服务器端的端口号包括 0~49151，其中 0~1023 是**熟知端口号**（又称系统端口号），剩下的是登记端口号。
客户端使用的端口号包括 49152~65535。这些端口号是给某个客户进程暂时使用的，通信结束后端口号就要恢复未分配状态。

表 5-2　常用的熟知端口号

应用程序	FTP	TELNET	SMTP	DNS	TFTP	HTTP	SNMP	SNMP (trap)	HTTPS
熟知端口号	21	23	25	53	69	80	161	162	443

5.2 用户数据报协议 UDP

5.2.1 UDP 概述

UDP 的特点：**无连接、尽最大努力交付、面向报文、无拥塞控制**、支持一对一、一对多、多对一、多对多，**首部开销小**。

5.2.2 UDP 的首部格式

UDP 的首部总共 8 个字节，只有四个字段：源端口、目的端口、长度、检验和。
如果接收方发现报文中的目的端口号不对，就丢弃报文，并使用 ICMP 发送“端口不可达”差错报文给发送方。ICMP 的应用 tracert 就是使用了 UDP 报文。
因为 UDP 的通信之间是无连接的，所以虽然**要用到端口号**，但是**不用套接字**（TCP 必须要在套接字之间建立连接）。
UDP 的检验和用来检验整个 UDP 报文的差错。
UDP 的差错检验方法是各个 4 字节段的反码求和，和作为检验和序列放入检验和字段。检验时对数据报各个 4 字节段反码求和，若每一位都是 1 则无错。
这种差错检验方法的检错能力不强，但是处理起来快。

5.3 传输控制协议 TCP 概述

5.3.1 TCP 最主要的特点

TCP 的特点：**面向连接**的、**点对点通信**、**提供可靠传输**、全双工通信、**面向字节流**。

面向连接：使用 TCP 前要先建立连接，通信完后要释放连接。

点对点通信。

可靠传输：无差错、不丢失、无重复、按序到达。

全双工通信：TCP 的两端都设有发送缓存和接收缓存。发送时，应用程序把数据放到 TCP 的发送缓存后，TCP 在合适的时候把数据发送出去。接收时，TCP 把收到的数据放入接收缓存，应用进程合适时读取缓存中的数据。

面向字节流：流是流入到进程或从进程流出的字节序列。TCP 把应用进程交下来的数据看成一串无结构的字节流。

TCP 的**报文长度**是根据接收窗口和网络拥塞程度决定。如果应用进程一次往发送缓存中放了很长的数据，那 TCP 可能会把它划分为多个短的数据块发送，如果应用进程一次只发来一个字节，TCP 也可以等积累足够多的字节后再把它们构成报文段发出去。

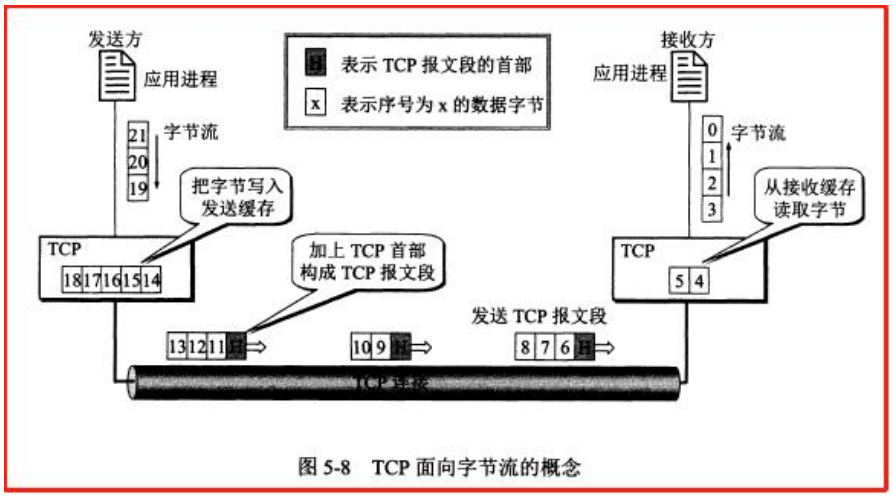


图 5-8　TCP 面向字节流的概念

5.3.2 TCP 的连接

IP 地址加上端口号称为**套接字**，套接字就是 TCP 连接的端点。
套接字不是应用进程，也不是端口。
套接字的格式： IP 地址：端口号（如 192.168.100.2：80）

每一条 TCP 连接唯一地被它地两个端点的套接字所确定。

5.4 可靠传输的工作原理

5.4.1 停止等待协议

停止等待协议用来**在不可靠的传输网络上**实现可靠通信。
原理：每发送完一个分组就停止发送，等待对方的确认，收到确认后再发送下一个分组。
分组需要进行编号。

出现差错

如果发送方发送的数据在传输过程丢失了，或者到达了接收方但是报文内容出了差错，那么接收方都不会发送任何信息。这时发送方超时没有收到确认，就会进行重传。
超时重传是超过一定时间没收到确认就要重传刚发送过的分组。实现方式是每发送完一个分组就设置一个超时计时器，重传时间比平均往返时间长一些，这又称**自动重传请求 ARQ**。
这里要注意：

发送方在发送完一个分组后，必须暂时保留已发送的分组的副本，以在超时重传时使用，只有收到相应的确认后才能清除保留的副本（对照发送缓存和发送窗口的后沿来理解）

分组和确认分组都需要进行编号，以明确是哪个分组收到了确认，哪个没有收到（对照 TCP 报文首部中的序号和确认号来理解）。

超时计时器的重传时间要比平均往返时间长一些，具体重传时间设为多少是一个很复杂的问题。

确认丢失和确认迟到

如果接收方发送给发送方的确认丢失或迟到了，那么发送方超时未收到确认，也会进行重传。而接收方收到重传的报文后，会丢弃这个重复的报文，并向发送方发送确认。发送方收到了重复的确认会直接丢弃。
若对方收到重复分组，就丢弃该分组，同时还要发送确认。接收方收到重复的确认后不做任何操作。

提高信道利用率

停止等待协议的信道利用率很低，为了提高效率，采用了流水线传输方式，这用到了**连续 ARQ 协议**和**滑动窗口协议**
流水线传输就是发送方可以连续发送多个分组，而不必每发完一个分组都要停下来等待对方的确认。

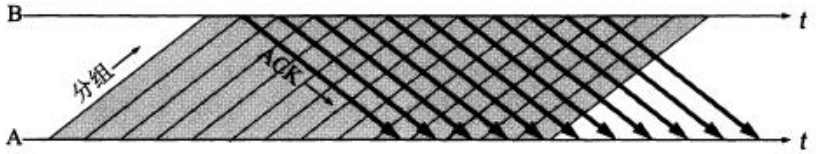


图 5-12　流水线传输可提高信道利用率

5.4.2 连续 ARQ 协议

连续 ARQ 协议用来提高利用率，它规定：

发送方维持一个**发送窗口**，凡位于发送窗口内的分组都可以连续发送出去，而不需要等待对方确认。发送方每收到一个确认，就根据确认号将发送窗口向前滑动一定距离。

接收方采用**累积确认**：不必对收到的分组逐个发送确认，而是只需对按序到达的最后一个分组发送确认，表明这个分组以前的所有分组都正确收到。

由上可见，连续 ARQ 协议是在滑动窗口上实现的。滑动窗口协议是 TCP 协议的精髓。

5.5 TCP 报文段的首部格式

TCP 传输的数据单元是报文段，一个 TCP 报文段分为首部和数据两部分。

TCP 报文段首部的前 20 个字节是固定的，后面有 4N 个字节是按需增加的选项。

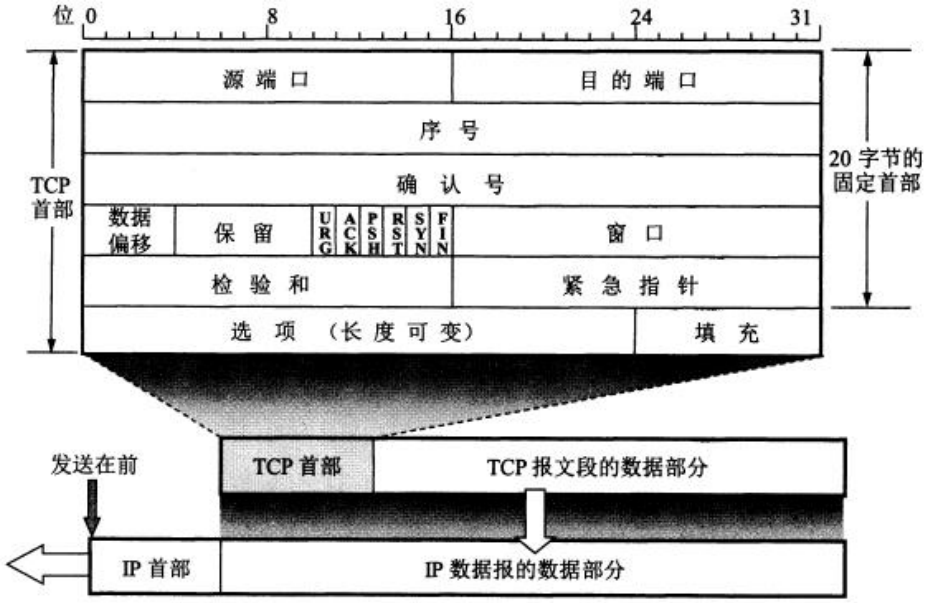


图 5-14　TCP 报文段的首部格式

首部各字段的作用：

序号：在一个 TCP 连接中传送的字节流中的每一个字节都要按顺序编号。首部中的序号字段存储本报文段发送的**数据**（是数据，不包含首部）的第一个字节的序号。序号字段只有 32 位，序号值不能超过 2^32。

理解：数据部分的每个字节都会占用一个序号。

确认号：首部中的确认号是期望收到对方下一个报文端的第一个数据字节的序号。**若确认号为 N，表明到 N-1 为止的数据都已正确收到。**

数据偏移：数据部分距报文起始点的偏移，实际等于首部长度。首部长度在 20~60 字节之间。

保留。保留为今后使用。

6 个控制位：

紧急 URG：当 URG=1，此报文段需要尽快传送，优先级高。

确认 ACK：当 ACK=1，确认号字段有效。**连接建立后的所有报文段都必须令 ACK=1**。

推送 PSH：当 PSH=1，接收方收到报文后尽快交付应用进程，而非等缓存满了再交付。

复位 RST：当 RST=1，表明 TCP 连接中出现严重差错，需要释放连接再重新建立连接。

同步 SYN：**当 SYN=1，表明这个一个连接报文**。如果 ACK=0 则是连接请求报文，如果 ACK=1 表明这是连接接受报文。

终止 FIN：当 **FIN=1，表明发送方已发送完数据，并要求释放连接**。

窗口：窗口指的是发送本报文段的一方的接收窗口。首部中的窗口字段指出了从本报文段中的确认号算起，**当前允许对方发送的数据量**（以字节为单位）。

检验和：检验整个数据报。

紧急指针，当 URG=1 时才有意义，指出本报文段中的紧急数据的字节数。即使在窗口为 0 时也可以发送紧急数据。

选项：长度可变，最长 40 字节。选项有**最大报文段长度 MSS**、窗口扩大选项、时间戳选项、选择确认选项等。

最大报文段长度 MSS

最大报文段长度 MSS 是每一个 TCP 报文段中的数据字段的最大长度，而不是整个 TCP 报文段的最大长度。

MSS 并不是一个标准固定值，而是可以由连接双方各自确定的值，且两个传送方向可以有不同的 MSS 值。MSS 的值可能达到几千字节。

连接建立时，双方都把自己支持的 MSS 写入这个选项字段中，以后就按照这个值传送数据。

如果未填写这一选项，那么 MSS 的默认值是 536 字节长。

窗口扩大选项

TCP 中窗口字段长度是 16 位，因此最大的窗口大小是 64K 字节。但是对于卫星网络，因为传播时延和带宽都很大，为了获得高吞吐率就需要更大的窗口。

时间戳选项

时间戳选项字段中包括时间戳值字段和时间戳回送回答字段。

时间戳选项用来计算往返时间 RTT。发送方把发送报文时的时间放入时间戳字段，接收方在确认该报文时把时间戳字段值复制到它的时间戳回送回答字段中。这样发送方收到确认报文后就可以计算出 RTT 来。

另一方面时间戳选项还可以用来防止序号绕回，因为 TCP 序号字段只有 32 位，序号值不能超过 2^32，所以可能出现具有相同序号的报文段，时间戳可以用来区分这样的报文段。

5.6 TCP 可靠传输的实现

5.6.1 以字节为单位的滑动窗口

滑动窗口是以字节为单位的，每个字节都有序号。

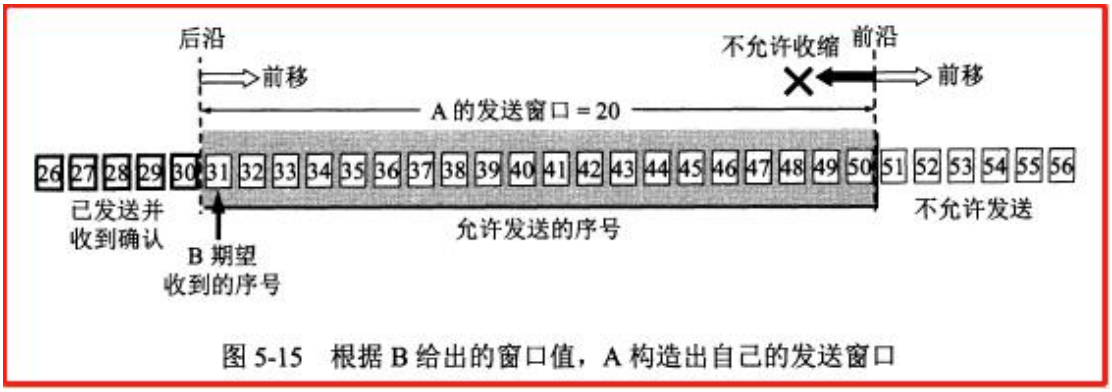


图 5-15　根据 B 给出的窗口值，A 构造出自己的发送窗口

TCP 使用**滑动窗口机制**。**发送窗口里的序号表示允许发送的序号，发送窗口后沿的后面部分表示已发送且已收到了确认**，发送窗口前沿的前面部分表示不允许发送。发送窗口的前沿会不断向前移动（也可能不动或后移），发送窗口的后沿可能不动（没有收到新的确认）也可能**前移（收到了新的确认）**，不可能后移。

接收方会**把接收窗口的值放到窗口字段中**发给发送方，**发送窗口的大小不能超过接收方传来的报文首部中的窗口字段值。**

接收方**发回的确认号是自己按序收到的数据的最高序号加 1**。

发送方会根据接收方发来的确认号和窗口字段来构造自己的发送窗口，确认号决定了发送窗口的后沿，窗口字段值和拥塞窗口共同决定发送窗口的大小。

发送窗口中的数据是可以直接连续发送出去的，所以发送窗口越大，可能获得的传输效率越高。

一个使用滑动窗口进行可靠传输的例子

A 为发送方，B 为接收方。

例子的开始时，A 根据 B 发来的窗口值（20）和确认号（31）构建了自己的发送窗口，如上图所示。此时发送窗口内的数据都是允许发送但尚未发送的数据。

现在 A 发送了 11 个字节的数据（序号 31-41），如下图所示。此时发送窗口内的数据包含已发送但未收到确认（P1~P2）的数据和允许发送但尚未发送的数据（P2~P3）两部分。

这时发送窗口的状态需要三个指针来描述：P1 指向发送窗口的后沿，P2 指向允许发送但尚未发送的第一个字节，P3 指向发送窗口的前沿外即将进入发送窗口的字节。P2~P3 之间的部分又称可用窗口。

然后 B 收到了序号为 32，33 的数据，但是没有收到序号为 31 的数据，因为 B 只对按序收到的数据中的最高序号进行确认，所以此时 B 发送给 A 的确认报文段中确认号仍为 31。

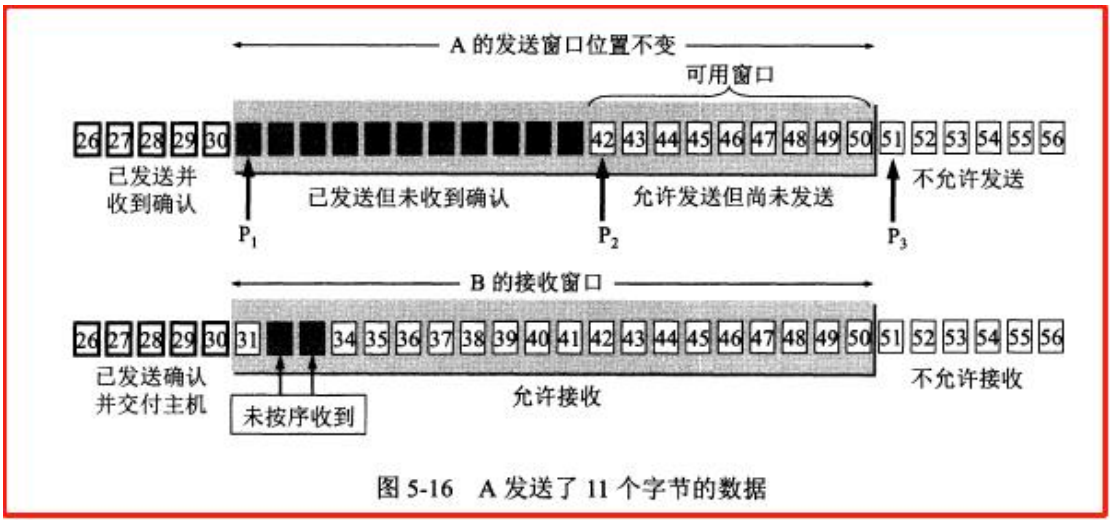
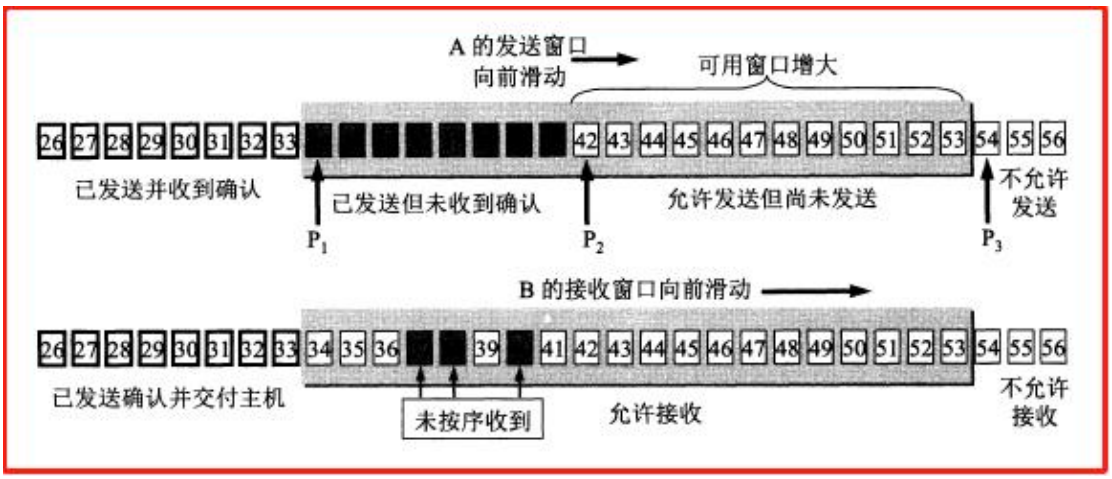


图 5-16　A 发送了 11 个字节的数据

接下来 B 收到了序号为 31 的数据，并把 31~33 的数据交付给应用进程，然后删除了这些数据。B 的接收窗口也向前移动了 3 个序号，同时给 A 发送确认号为 34 的确认报文。
A 收到确认号为 33 的报文后，也将发送窗口向前滑动了 3 个序号，此时发送窗口大小没变，但是可用窗口变大了。



接下来 A 继续把可用窗口中的数据发送完后，P2 指针向前移动和 P3 重合，此时 A 的可用窗口已减小到 0，要暂时停止发送，等待收到确认。
如果 A 超时未收到确认报文，就重传这部分数据，直到收到 B 的确认报文为止。

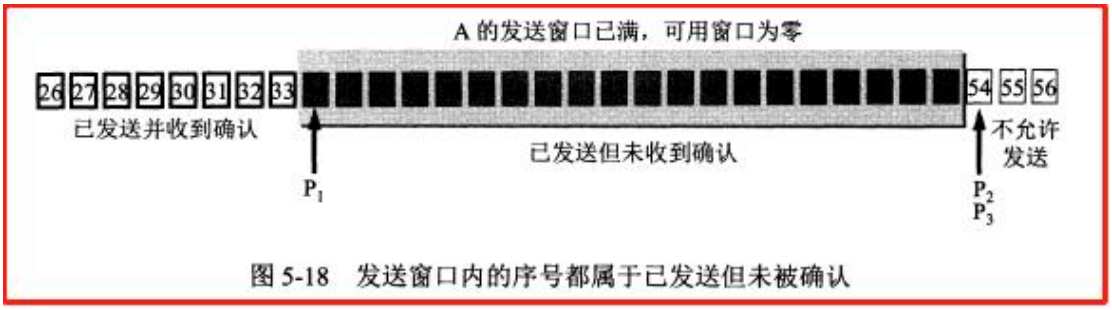


图 5-18　发送窗口内的序号都属于已发送但未被确认

发送缓存

发送方维持一个**发送缓存**，其中存放

准备发送的数据。

已经发送但尚未收到确认的数据。

发送窗口是发送缓存的一部分。已被确认的的数据会被从发送缓存中删除，因此发送缓存和发送窗口的后沿是重合的。

应用进程向发送缓存写入数据时不能太快，否则填满发送缓冲后就没有存放数据的空间了。

实际发送/接收缓存和窗口中的字节数是非常大的。

接收缓存

接收方维持一个**接收缓存**，用来存放：

按序到达的、尚未被接收应用程序读取的数据。

未按序到达的数据。

接收窗口是接收缓存的一部分。如果应用程序来不及读取收到的数据，接收缓存就会被填满，使接收窗口减小到 0，反之接收窗口会增大，但最大不能超过接收缓存的大小。

TCP 没有明确规定如何处理未按序到达的数据，但通常是先临时存放在接收窗口中，等字节流中缺少的字节到达后，再交付给上层的应用进程。

累积确认

TCP 要求接收方必须有累积确认的功能，这样可以减小传输开销。

接收方可以在合适的时候发送确认，也可以在自己有数据要发送时把确认信息捎带上。

但是注意接收方不能过分推迟发送确认，以避免发送方产生不必要的重传。如果收到一连串具有最大长度的报文段，则必须每隔一个报文段就发送一个确认。

TCP 的通信是全双工通信。通信中每一方都在发送和接收报文段，因此每一方都有自己的发送窗口和接收窗口。

5.6.2 超时重传时间的选择

重传时间的选择是 TCP 最复杂的问题之一。

因为 TCP 的报文可能只在一个高速局域网中传送，也可能要经过多个低速率的网络，所以重传时间不能设为固定值。

TCP 采用了一种自适应算法，它依据**往返时间 RTT** 来设置超时重传时间。

TCP 保留了一个 **历史 RTT 的加权平均结果** RTTs。

RTTs 的计算方式：新的 RTTs = (1-a)*旧的 RTTs + a*新的 RTT。a 建议取 0.125。

超时重传时间 RTO 比 RTTs 略大：RTO = RTTs + 4*RTTd。

RTTd 是 RTT 的偏差的加权平均值：新的 RTTd = (1-b)*旧的 RTTd + b*|RTTs - 新的 RTT|。b 建议取 0.25。

具体实现

要解决当报文重传的特殊情况。

实现方式：报文段每重传一次，就把 RTO 增加一倍，当不重传了，就继续使用上述公式计算 RTO。

5.6.3 选择确认 SACK

当报文未按序到达（到达的字节不连续，一段一段的），发送方需要重传收到的确认号之后的所有报文，而有些确认号之后的不连续的报文实际上已经到了，全部重传会浪费资源。**选择确认**用来解决这种情况。
首部的选项中可以有选择确认 SACK 字段。SACK 使用两个字节块分别指明一个连续字段的开始位置和长度。最多可以指出 4 个连续字节块的边界情况。
SACK 应用不多。

5.7 TCP 的流量控制

流量控制是为了让发送方的发送速率不要太快，要让接收方来得及接收。

5.7.1 利用滑动窗口实现流量控制

流量控制是通过滑动窗口实现的。接收方会把接收窗口的大小放到给发送方的报文的窗口字段中。
发送方的发送窗口不能超过接收方给出的窗口字段的数值。

死锁

当接收方的接收窗口减小到 0，发送方停止发送数据后。过了一段时间 B 的接收窗口恢复了一些，但是它发给发送方的报文丢失了，然后 A 就会一直等待 B 发送的非零窗口的通知，而 B 也一直等待 A 发送的数据，这时就进入了死锁状态。
死锁的解决：TCP 每一个连接都设有一个持续计时器。只要 TCP 连接的一方收到对方的零窗口通知，就启动持续计时器。若持续计时器的时间到期，就发送一个零窗口探测报文段，对方则在返回这个探测报文段的确认报文时给出窗口值。

5.7.2 TCP 的传输效率

应用进程把数据传送到 TCP 的发送缓存后，剩下的发送任务就交给 TCP 来完成了。

Nagle 算法

TCP 的实现中广泛使用了 **Nagle 算法：**

发送方先把到达发送缓存的第一个数据字节发送出去，收到确认后，再把发送缓存中剩下的数据组装成报文段发送。

收到前一个报文段的确认后再发送下一个报文段。

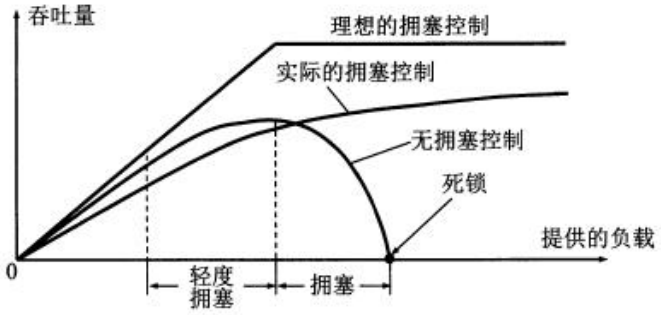
当缓存中的数据达到发送窗口的一半大小或报文段最大长度后，就立即发送一个报文段。

Nagle 算法用来避免发送方发送很小的报文段。

5.8 TCP 的拥塞控制

5.8.1 拥塞控制的一般原理

拥塞就是当前对网络中某一资源的需求超过了该资源所能提供的可用部分，网络性能变坏。
概括而言拥塞的条件就是：对资源的需求 > 可用资源。
网络拥塞的成因往往十分复杂，并拥塞常常趋于恶化。比如因为网络拥塞有报文未能按时到达，那么发送方就会超时重传使更多的分组进入网络，加剧网络拥塞。
拥塞控制就是防止过多的数据注入到网络中，使网络中的路由器或链路不致过载。
流量控制是**端到端的问题**，是接收端抑制发送端发送数据的速率。拥塞控制是**全局性的问题**，涉及到所有的主机、路由器及相关因素。
TCP 连接的端点只要迟迟收不到对方的确认信息，就猜想网络中某处可能出现了拥塞。



上图中横坐标代表单位时间内输入到网络中的分组数，纵坐标代表单位时间内从网络输出的分组数。
拥塞控制很复杂。网络拥塞的指标有：被丢弃的分组数，平均队列长度、超时重传的分组数、平均分组时延等。

5.8.2 TCP 的拥塞控制方法

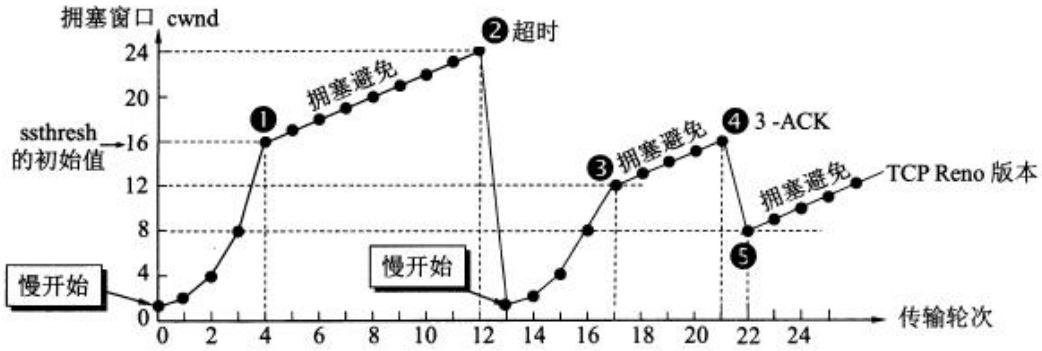
TCP 的拥塞控制采取了**慢开始、拥塞避免、快重传、快恢复**四种算法。
这种方法是**基于窗口的拥塞控制**。发送方维持一个**拥塞窗口**，并让自己的发送窗口等于拥塞窗口（实际上发送窗口取拥塞窗口和接收窗口中的较小者）。
控制拥塞窗口的原则是：只要网络中没有出现拥塞，就把拥塞窗口增大一些；但只要网络出现拥塞或可能出现了拥塞，就把拥塞窗口减小一些。
判断网络拥塞的依据是**出现超时**。当出现拥塞就使拥塞窗口减小，反之增大。

慢开始

初始拥塞窗口很小，然后由小到大逐渐增大发送窗口。
初始拥塞窗口一般不超过 2-4 个 SMSS（发送方最大报文段）长度。每收到一个新的确认后，就增加一次拥塞窗口。
使用慢开始算法，每经过一个传输轮次，拥塞窗口 cwnd 就会加倍。

拥塞避免

拥塞避免算法是让拥塞窗口缓慢地增大，不像慢开始那样加倍增长。
当 cwnd 大于一个界限值时，就使用拥塞避免算法，小于时就使用慢开始算法。
当出现超时，拥塞窗口就恢复初始值重新进行慢开始，且界限值减半。



快重传

快重传要求当接收方收到报文段后立即发送确认。
当接收方收到的报文段出现丢失，它后面不论收到什么报文段，发回的确认号都是对失序之前的那个报文段的确认。
当发送方**连续收到 3 个对同一报文段的重复确认**（表明下一个报文段未收到），就立即重传下一个报文段，这样就可以避免超时。
这是为了避免出现超时，使发送方误判为网络拥塞。

快恢复

对于一般的超时，界限值减半，拥塞窗口直接置为初始值；对于快重传情况下，界限值减半，拥塞窗口设置为和界限值一样，以实现快恢复。

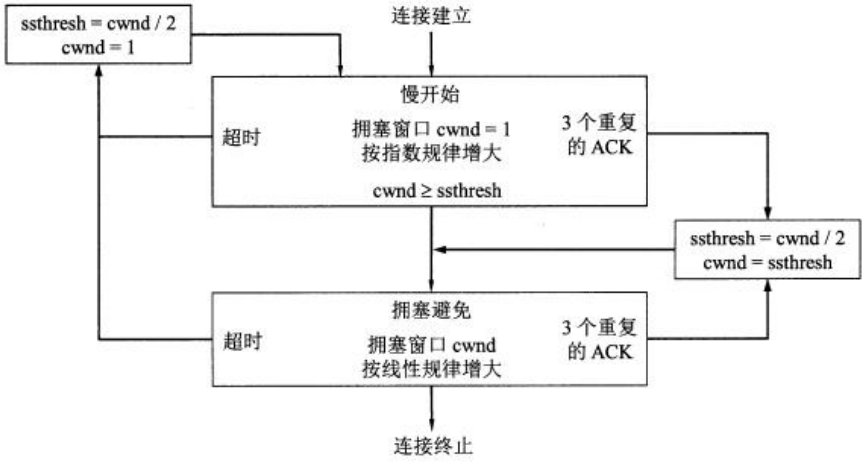


图 5-27　TCP 的拥塞控制的流程图

发送方的发送窗口实际设置为**接收方窗口 rwnd 和拥塞窗口 cwnd 中较小的一个**。

5.8.3 主动队列管理 AQM

在**网络层**，路由器可以通过采用**主动队列管理 AQM** 来减少网络拥塞的发生。
主动队列管理就是路由器采取的一种分组丢弃策略。
路由器的队列中的分组是先进先出，在队列满了之后到达的分组就直接丢弃。
主动队列管理就是在队列长度还未满时就丢弃后面的分组。但是具体丢弃策略尚未标准化。

5.9 TCP 的运输连接管理

TCP 连接有三个阶段：连接建立、数据传送、连接释放。
主动发起 TCP 连接的应用进程是客户，另一方是服务器。
TCP 的连接开始前客户和服务器都会创建一个**传输控制块 TCB**，其中存储如 TCP 连接表、指向发送/接收缓存的指针、指向重传队列的指针等。连接释放后删除。

5.9.1 TCP 的连接建立

TCP 的连接采用**三次握手机制：服务器要确认客户的连接请求，客户要对服务器的确认进行确认。**

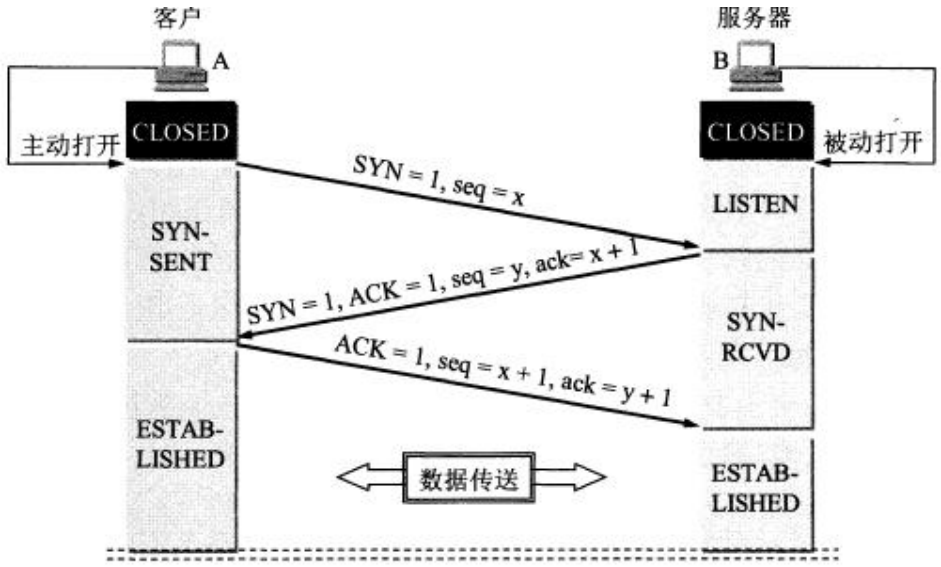


图 5-28 用三报文握手建立 TCP 连接

连接请求报文和**连接接受报文**段都不能携带数据，但是都消耗一个序号。
第三个 **ACK 报文**段可以携带数据也可以不携带，若不携带则不消耗序号（即下一个报文序号和此报文序号相同）。
SYN-SENT 表示同步已发送状态，SYN-RCVD 表示同步收到状态，ESTABLISHED 表示已连接状态。
TCP 有如下规定：

SYN = 1 的报文段都不能携带数据，但要消耗掉一个序号。所以三次握手中前两个报文都不能携带数据。

ACK = 1 的报文段如果不携带数据则不消耗序号。

三报文握手中的第二个报文，也就是服务器发给客户的 SYN 报文也可以拆分成两个报文段，一个确认报文段（ACK = 1, ack = x+1）和一个同步报文段（SYN = 1, seq = y），那样就是四报文握手了。
采用三报文握手是为了解决客户发送的连接请求报文中途滞留发生重传的情况。当发生重传情况，客户可能连续发送了两个连接请求，而服务器也会回复两个连接接受，此时发送端通过最后一个确认报文保证只建立一个连接。
如果不采用三报文握手，那么只要服务器发出确认，新的连接就建立了。

5.9.2 TCP 的连接释放

TCP 的连接释放采用**四次挥手机制**。**任何一方**都可以在数据传送结束后发出连接释放的通知，等待对方确认后进入半关闭状态。当另一方也没有数据发送后，则发送连接释放通知，对方确认后完全关闭 TCP 连接。

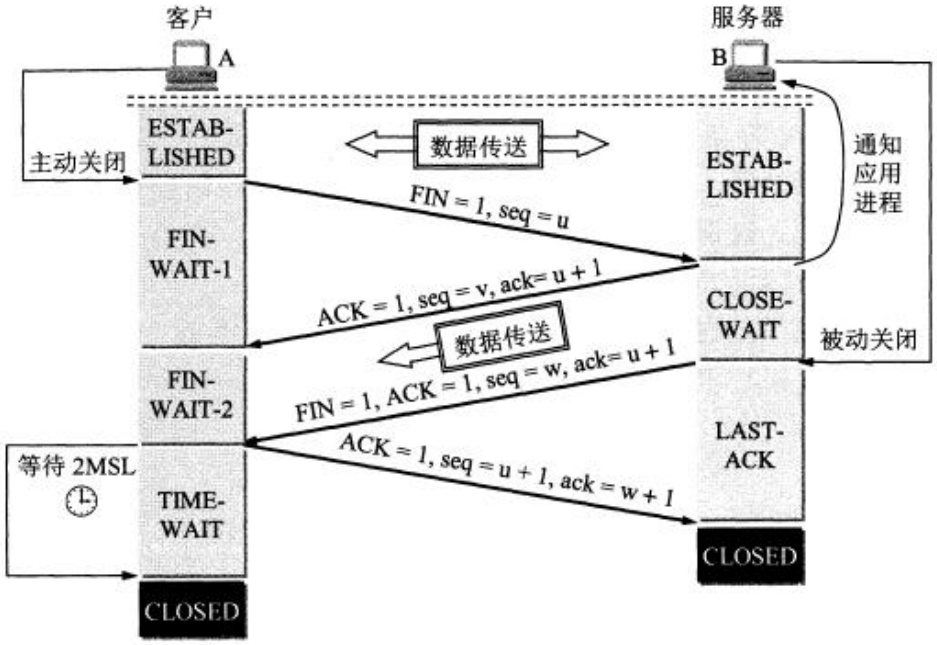


图 5-29　TCP 连接释放的过程

上图中客户发起了**连接释放报文(FIN=1)**，此时不再传送数据。但是服务器可能还要传送数据，因此在发送一个 **ACK 报文**后，如果有数据传送还要继续发送完。

当服务器发送完数据，就发送连接释放报文，接着客户端对此确认，服务器收到确认后彻底关闭连接。

TCP 规定：FIN 报文即使不携带数据，也要消耗一个序号。

四次挥手的详细流程

A 向 B 发送连接释放报文（FIN 报文）后进入 FIN-WAIT-1（终止等待 1）状态。

B 收到连接释放报文后立即发出确认，然后 B 就进入了 CLOSE-WAIT（关闭等待）状态。TCP 服务器进程这时要通知高层应用进程，从 A 到 B 方向的连接已经释放了，TCP 连接实际上已经是半关闭状态。

A 收到 B 的确认后，就进入 FIN-WAIT-2（终止等待 2）状态，等待 B 发出的连接释放报文段。

B 在发送确认报文后，如果没有数据要发送，应用进程就通知 TCP 释放连接，这时 B 会发送连接释放报文（FIN 报文）。如果 B 还要发送数据，就等发送完数据后再发送连接释放报文。这之后 B 进入 LAST-ACK（最后确认）状态

A 收到 B 的连接释放报文后，要再发送一个确认报文段（ACK 报文），然后进入 TIME-WAIT（时间等待）状态。A 会在这一状态保持 2MSL 的时间（这里有一个时间等待计时器），之后进入 CLOSED 状态。

B 收到 A 的确认报文后也会进入 CLOSED 状态。B 会比 A 更早地结束连接。

MSL 含义是最长报文段寿命，RFC 标注建议是 2 分钟，实际一般小于等于 2 分钟，因此 TIME-WAIT 的时长一般小于等于 4 分钟。

TIME-WAIT 状态的意义

保证 A 发送的最后一个 ACK 报文段能够到达 B。这个 ACK 报文段可能丢失，当 B 没有收到 A 的 ACK 报文，会超时重传它之前发送的连接释放报文，这样 A 就能在 2MSL 时间内收到 B 重传的 FIN 报文，然后 A 重传 ACK 报文，并重新启动 2MSL 计时器。

如果没有 TIME-WAIT 状态，而是发完 ACK 报文后就立即释放连接，就无法收到 B 重传的 FIN 报文段，这样 B 就无法正常进入 CLOSED 状态。

防止已失效的报文遗留到下一个连接中，经过 2MSL 在关闭连接可以使本连接产生的所有报文都从网络中消失。

保活计时器

保活计时器：TCP 还会设置一个保活计时器。如果客户发生故障，服务器不再收到客户发来地数据，可以通过保活计时器来避免服务器白白等待。

保活计时器的工作方式：服务器每收到一次客户的数据，就重新设置保活计时器，通常是 2 小时。如果 2h 每收到客户数据，服务器就发送一个探测报文段，之后每隔 75s 发送一次。如果连续发送 10 个探测报文后客户仍没有响应，服务器就认为客户端出了故障，关闭当前的 TCP 连接。

TIME_WAIT

客户端在发送最后一个确认报文后不能直接进入 CLOSED 状态，要等待 **2MSL** 的时间（一般小于 2 分钟）。这是为了**保证它发送的确认报文能够到达服务器**。如果未能及时到达，服务器会超时重传连接释放报文，客户就会重新发送确认，并重新计时。另外也是保证本次连接产生的所有报文都从网络中消失。

5.9.3 TCP 的有限状态机

下面的状态机中虚线表示的是服务器进程的状态变迁，实线表示的是客户进程的状态变迁。

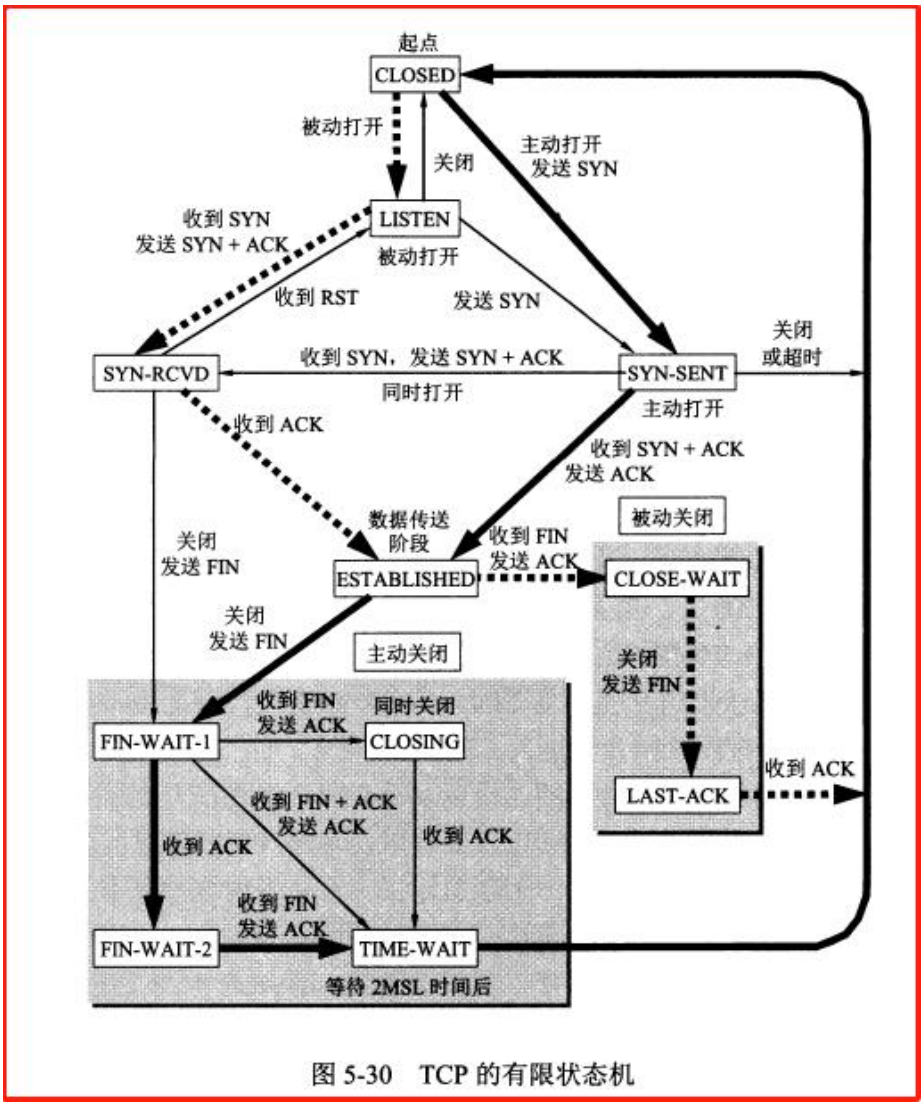


图 5-30　TCP 的有限状态机

第 6 章 应用层

6.1 域名系统 DNS

6.1.1 域名系统概述

6.1.2 互联网的域名结构

6.1.3 域名服务器

6.2 文件传送协议

6.2.1 FTP 概述

6.2.2 FTP 的基本工作原理

6.2.3 简单文件传送协议 TFTP

6.3 远程终端协议 TELNET

6.4 万维网 www

6.4.1 万维网概述

6.4.2 统一资源定位符 URL

6.4.3 超文本传送协议 HTTP

6.4.4 万维网的文档

6.4.5 万维网的信息检索系统

6.5 电子邮件

6.5.1 电子邮件概述

6.5.2 简单邮件传送协议 SMTP

6.5.3 电子邮件的信息格式

6.5.4 邮件读取协议 POP3 和 IMAP

6.5.5 基于万维网的电子邮件

6.5.6 通用互联网邮件扩充 MIME

6.6 动态主机配置协议 DHCP

6.7 简单网络管理协议 SNMP

6.7.1 网络管理的基本概念

6.7.2 管理信息结构 SMI

6.7.3 管理信息库 MIB

6.7.4 SNMP 的协议数据单元和报文

6.8 应用进程跨越网络的通信

6.8.1 系统调用和应用编程接口

6.8.2 几种常用的系统调用

6.9 P2P 应用

6.9.1 具有集中目录服务器的 P2P 工作方式

6.9.2 具有全分布式结构的 P2P 文件共享程序

6.9.3 P2P 文件分发的分析

6.9.4 在 P2P 对等方中搜索对象

问题

什么是域名？DNS 如何实现域名解析？
万维网采用了哪些技术？
HTTP 协议有哪些特点？
电子邮件系统用到了哪些协议？
动态主机配置协议 DHCP 的作用是什么？有什么特点
网络管理的三个组成部分是什么？
什么是系统调用接口？

回答

域名是互联网采用的命名方式，**域名可以解析为 IP 地址**。DNS 通过**分布式域名服务器**来解析域名。

HTTP 协议，统一资源定位符 URL，超文本标记语言 HTML 等。
无连接、无状态的。
邮件发送协议 SMTP 和**邮件读取协议 POP3、IMAP**。
动态配置 IP 地址，实现即插即用。
SNMP 本身、管理信息结构 SMI、管理信息库 MIB。
应用进程和操作系统之间转让控制权的接口。

第 6 章 应用层

应用层的协议多是基于客户-服务器方式。这里的客户和服务器都是**应用进程**。
应用层协议规定了**应用进程**通信时遵循的协议。

6.1 域名系统 DNS

6.1.1 域名系统概述

域名系统 DNS 是互联网使用的**命名系统**，用来**把便于识别的名字转换为 IP 地址**。
DNS 是一个**联机分布数据库系统**，采用客户-服务器方式。
DNS 使大多数名字在本地进行解析，只有少量解析要在互联网上通信。
域名到 IP 地址的解析是由互联网上的**许多域名服务器共同完成的**。
域名到 IP 地址的解析过程：

当某一应用进程需要解析域名，就调用解析程序，成为 DNS 的一个客户，**把待解析的域名放到 DNS 请求报文中，以 UDP 用户数据报**方式发给**本地域名服务器**。本地域名服务器查找域名后把对应的 IP 地址发给该应用进程。应用进程**获得 IP 地址后即可进行通信**。

如果本地域名服务器不能回答该请求，就向其他域名服务器请求，此时它就成为了客

户。

6.1.2 互联网的域名结构

互联网采用层次树状结构的命名方法，任何一台连接在互联网上的**主机**或**路由器**都有唯一一个域名。

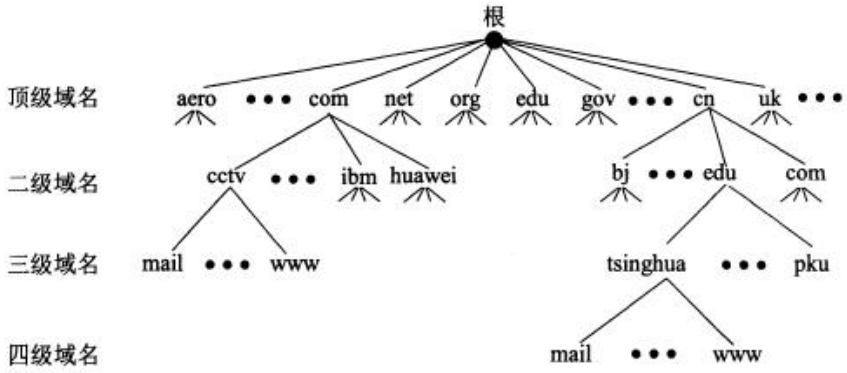


图 6-1　互联网的域名空间

如以下两个网站采用了不同的四级域名：

www.buaa.edu.cn ：北航主页

yzb.buaa.edu.cn ：北航研招办

域名由标号序列组成，**最右边的标号是顶级域名**，往左依次降低。

域名中的标号只能由**英文字母、数字和 '-'** 组成，**不区分大小写字母**。完整域名不超过 255 个字符。

各级域名由其上一级的域名管理机构管理，顶级域名由 ICANN 管理。

顶级域名 TLD 包括：

国家顶级域名：cn 为中国，us 为美国，uk 为英国。

通用顶级域名：com(公司企业), net(网络服务机构), org(非营利性组织), int(国际组织), edu(美国教育机构), gov(美国政府部门), mil(美国军事部门)。

上面是最初的 7 个顶级域名，后面还增加了很多其他的。现在已经有了中文的顶级域名。

基础结构域名：arpa，用于反向域名解析。

我国的二级域名分为两类：

类别域名：**com(企业)**, ac(科研机构), **edu**, **gov**, mil, net, org 等。

行政区域名：适用于各省。如 bj 为北京。

6.1.3 域名服务器

域名服务器分为**根域名服务器、顶级域名服务器、权限域名服务器**和**本地域名服务器**。

所有的根域名服务器都**包含所有的顶级域名服务器的域名和 IP 地址。**

根域名服务器是最重要的服务器，如果本地域名服务器无法解析域名，**首先求助于根域名服务器**。

根域名服务器在全球有成百上千个，但是分布是很不均衡的。

根域名服务器使用了**任播技术**。

顶级域名服务器：管理在该服务器注册的二级域名。

权限域名服务器：负责一个区的应服务器

本地域名服务器：主机查询域名时首先询问本地域名服务器，**计算机属性中的 DNS 服务器就是本地域名服务器**。

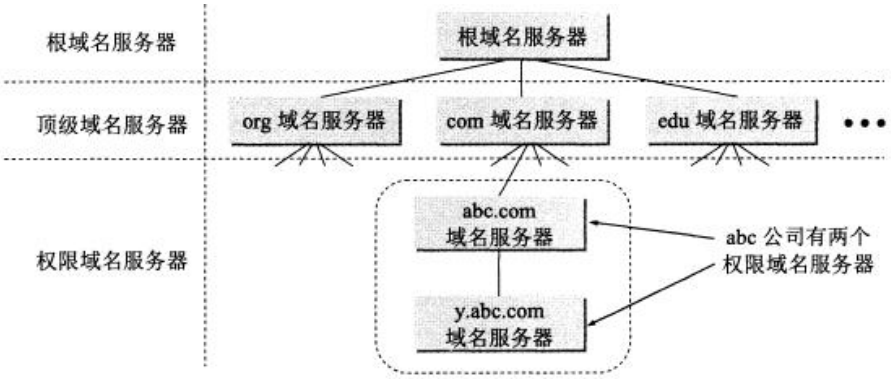
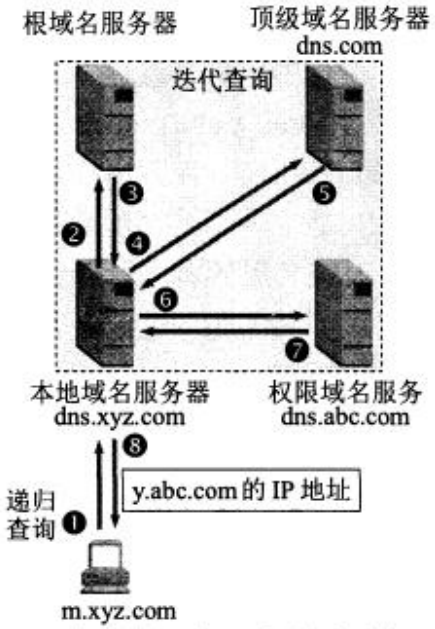


图 6-3　树状结构的 DNS 域名服务器

域名解析过程中主机向本地域名服务器的查询是递归查询。

本地域名服务器向根域名服务器的查询是迭代查询，即当根域名服务器无法完成解析时，就把下一步应该查询的域名服务器告诉本地域名服务器。



(a) 本地域名服务器采用迭代查询

为提高查询效率，域名服务器中采用了**高速缓存**，存放最近查询过的域名信息。

6.2 文件传送协议

6.2.1 FTP 概述

文件传送协议 FTP 使用 **TCP** 的可靠运输服务，为客户-服务器模式。

简单文件传送协议 TFTP 使用 **UDP** 协议。

FTP 和 TFTP 都属于文件共享协议中的一大类：**复制整个文件**。特点是要存取一个文件，就

必须先获得一个本地的文件副本。要修改文件，只能对文件副本进行修改，然后将修改后的文件副本传送回到原节点。

6.2.2 FTP 的基本工作原理

FTP 主要功能是减少或消除在不同操作系统下处理文件的不兼容性。

FTP 的服务器进程

一个 FTP 服务器进程可以同时为多个客户进程提供服务。
FTP 的服务器进程包括**一个主进程和若干个从属进程：**

主进程：负责接受新的请求。当主进程收到客户进程发来的连接请求后，将其交给从属进程进行处理。然后回到等待状态继续接受其他客户的请求。

从属进程：负责处理单个请求。

服务器进程中的主进程和从属进程是并发执行的。

文件传输功能的实现

文件传输时，FTP 的客户和服务器之间会建立**两个并行的 TCP 连接：控制连接和数据连接**，其中数据连接用于传输文件。因此 **FTP 要使用两个端口号**。
两个连接对应服务器端的两个从属进程：控制进程和数据传送进程。
FTP 客户发送的传送请求通过控制连接发送给控制进程。然后**控制进程创建数据传送进程和数据连接**。

6.2.3 简单文件传送协议 TFTP

简单文件传送协议 TFTP 使用 UDP 协议。
TFTP 的主要优点是可用于 UDP 环境，代码简便。
TFTP 采用了类似**停止等待协议**的**重传机制**。即发送后就等待确认，没有确认就重传。

6.3 远程终端协议 TELNET

远程终端协议 TELNET 采用客户服务器模型。能**将客户端的操作传到服务器端，然后将服务器端的输出返回到客户端屏幕**。
TELNET 采用 TCP 协议。

TELNET 的服务器进程

TELNET 的服务器进程类似 FTP，由主进程等待新的请求，并产生从属进程来处理每一个连接。

6.4 万维网 www

6.4.1 万维网概述

万维网 WWW 是**一个大规模的、联机式的信息储藏所**。万维网的简称是 **Web**。
超文本指的是**包含指向其他文档的链接的文本**，一个超文本由多个信息源链接组成。超文本仅包含文本信息，超媒体扩充为包含图形、声音、视频等。
万维网是一个分布式的超媒体系统。
Web 的客户程序向互联网中的服务器程序发出请求，服务器程序向客户程序送回客户所要的万维网文档。
页面就是在客户程序主窗口显示出的万维网窗口。
Web 要处理的几个问题及解决方式：

如何标志分布在整个互联网上的**文档**：采用**统一资源定位符 URL**。
用什么协议来实现万维网上的链接：采用**超文本传送协议 HTTP**。
怎么实现创作不同风格的万维网文档：使用**超文本标记语言 HTML**。
怎样使用户很方便地找到所需信息：通过搜索引擎实现。

6.4.2 统一资源定位符 URL

万维网使用统一资源定位符 URL 来标志万维网上的各种文档，**每个文档有在互联网内唯一的 URL**。
URL 相当于指向互联网上任何可访问对象的一个指针。
URL 的一般形式：**<协议>://<主机>:<端口>/<路径>**

协议：指出采用何种协议来获取该万维网文档，一般为 http，其次为 ftp
主机：即该主机的域名。
端口：通常都省略掉，HTTP 的默认端口号是 **80**。
路径：有时可省略。

输入 URL 时协议和 www 都可以省略，浏览器会自动补上。

使用 HTTP 的 URL

HTTP 的默认端口号是 80，通常都省略掉了。
当路径也省略掉，则 URL 指向互联网上的某个主页，比如 http://www.buaa.edu.cn。
主页可以是：

一个 WWW 服务器的最高级别的页面。
某一个组织的一个定制的页面，从这个页面可以链接到本组织的其他站点。
个人设计的 WWW 页面。

https://www.buaa.edu.cn/jgsz1/yxsz.htm 左边是一个 URL 链接，cn/ 右侧的为路径，最后的 .htm 表明这是一个 htlm 文档

6.4.3 超文本传送协议 HTTP

HTTP 是**面向事务**的应用层协议。**事务**指的是一系列的不可分割的信息交换（即这些信息交换是一个整体）。
万维网客户与服务器程序之间交互使用的协议是 HTTP 协议。万维网的客户就是浏览器
HTTP 本身是**无连接、无状态**的，使用可靠传输的 **TCP 协议**。

无连接：通信双方在交换 http 报文前不需要先建立 http 连接。
无状态：HTTP 服务器不记得曾经访问过的客户。

HTTP 连接的建立与释放

每个万维网网点都有一个服务器进程，它不断地监听 TCP 的端口 **80**，当发现有浏览器向它发来 TCP 的连接建立请求并建立连接后，浏览器就会发出浏览某个页面的请求，服务器接着返回所请求的页面作为响应。最后 TCP 连接释放。
注意这个过程：**首先建立 TCP 连接**，且该连接的**端口为 80**。客户会**把 HTTP 请求报文作为 TCP 连接三次握手中的第三个报文的数据**。然后服务器直接返回文档作为响应。

HTTP/1.1

HTTP/1.1 协议使用了**持续连接**。就是万维网服务器在发送响应后一段时间内仍保持这条连接，当客户继续访问时不需要重新建立 TCP 连接。
HTTP/1.1 协议的持续连接有两种工作方式：

非流水线方式：客户收到前一个响应后才能发下一个请求。

流水线方式：客户收到上一个响应前就可以接着发新的请求。

代理服务器

代理服务器又称**万维网高速缓存**，它把最近的一些请求和响应暂存在本地磁盘中。当新请求与暂存的请求相同，就返回暂存的响应。
代理服务器可以在客户端或服务端工作，也可以在中间系统上工作。
比如某个校园网使用了代理服务器，当校园网中某个主机的浏览器请求服务时，先和代理服务器建立 TCP 连接并发出 HTTP 请求报文，如果代理服务器有所请求对象就返回这个对象，如果没有，代理服务器就代表用户与源点服务器建立连接并发送 HTTP 请求报文。

HTTP 的报文结构

HTTP 报文的每一个字段都是 **ASCII 码串**。
HTTP 有请求报文和响应报文两类报文。
HTTP 报文由三部分组成：

开始行：用于区分是请求报文还是响应报文。

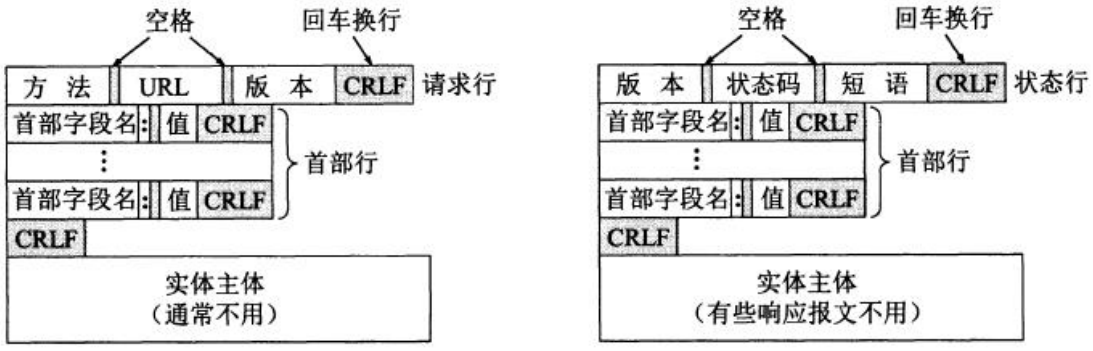
请求行：包括方法（表明了请求报文的类型），**请求的 URL**，HTTP 的版本。

状态行：响应报文的第一行叫状态行，包括：HTTP 的版本，**状态码**，解释状态码的短语。

首部行：用来说明浏览器、服务器或报文主体的一些信息。

实体主体：请求报文一般不使用这个字段，响应报文可能会使用。

注意：HTTP 报文中，**请求行最后有一个 CRLF，其他所有首部之间也都有一个 CRLF，首部与实体主体之间有两个 CRLF**。



(a) 请求报文　　(b) 响应报文

一个 HTTP 请求报文的例子：

GET /dir/index.htm HTTP/1.1 //请求行，这里使用了相对 URL
Host:www.buaa.edu.cn //第一个首部行，给出了主机的域名
Connection:close //告诉服务器发送完请求的文档后就可以释放连接。
User-Agent:Mozilla/5.0 //表明用户代理使用了火狐浏览器
Accept-Launguage:cn //表明用户希望优先收到中文版本的文档
//请求报文的最后还有一个空行

HTTP 请求报文的一些方法

表 6-1 HTTP 请求报文的一些方法

方法（操作）	意义
OPTION	请求一些选项的信息
GET	请求读取由 URL 所标志的信息
HEAD	请求读取由 URL 所标志的信息的首部
POST	给服务器添加信息（例如，注释）
PUT	在指明的 URL 下存储一个文档
DELETE	删除指明的 URL 所标志的资源
TRACE	用来进行环回测试的请求报文
CONNECT	用于代理服务器

HTTP 响应报文的 5 大类状态码：

1xx。表示通知信息，如请求收到了或正在处理。

2xx。表示成功。

3xx。表示重定向，如要完成请求还需要采取进一步的行动。如 HTTP/1.1 301 Moved Permanently。这时后面还会跟一个首部行表明新的 URL 地址：Location:http://www.buaa.edu.cn/dd/index.html

4xx。表示客户的差错。如 Http/1.1 404 Not Found

5xx。表示服务器的差错。

在服务器上存放用户的信息

HTTP 是无状态的。当有时想要保存一些信息，比如保存某网站的账号与密码，就需要采用 **Cookie 技术**。**万维网站点可以使用 Cookie 来跟踪用户**。

Cookie 的工作原理：当用户 A 浏览某网站，该网站的服务器就为 A 产生一个唯一的识别码，并存储在数据库中，接着在给 A 的响应报文中添加一个字段名为 Set-cookie，值为识别码的首部行，A 收到此响应报文后把它存储在自己的 Cookie 文件中。这样服务器就能够知道用户 A 什么时候访问了哪些页面。当 A 之后再次访问时，服务器可以识别出 A，这样就不需要用户 A 再次输入姓名，密码等。

6.4.4 万维网的文档

万维网使用超文本标记语言 HTML 来显示各种万维网页面。

HTML 文档是一种可以使用任何文本编辑器创建的 **ASCII 码文件**。

HTML 文档一般以 .html 或 .htm 为后缀。

HTML 规定了链接的设置方法。链接一般显示为蓝色字且加上下划线，链接的终点是其他页面。
当链接指向的是本计算机中的文件，就是**本地链接**。

其他语言

可扩展标记语言 XML 和 HTML 很相似。但从设计宗旨而言，**XML 用于传输数据**，而 **HTML 用于显示数据**。
可扩展超文本标记语言 XHTML 是作为一种 XML 应用被重新定义的 HTML，将逐步取代 HTML。
层叠样式表 CSS 是一种样式表语言，用于为 HTML 文档定义布局。如规定在浏览器上显示的字体、颜色、边距等。

静态文档

万维网**静态文档**：文档创作完毕后就存放在万维网服务器中，在用户浏览过程中，内容不会变。

动态文档

动态文档：文档的内容是在浏览器访问万维网服务器时由应用程序动态创建的。每次访问用户看到的内容都是不一样的。
比如天气预报、股市行情等都要用动态文档。
动态文档和静态文档的差别主要在服务器一端。
通用网关接口 CGI 是一种标准，定义了动态文档如何创建。服务器端使用 **CGI 程序**来创建动态文档。
脚本指的是**被另一个程序而非处理器来解释或执行的程序**。JavaScript 等就是脚本语言。

活动文档

活动文档：可以使浏览器屏幕连续更新。
当浏览器请求一个活动文档，服务器直接返回**一段活动文档程序的副本**，使该程序副本在浏览器端运行。活动文档程序可以与用户直接交互，并可以连续地改变屏幕的显示。

6.4.5 万维网的信息检索系统

搜索引擎是在万维网中进行搜索的工具，可以分为全文搜索引擎（百度、谷歌等）和分类目录搜索引擎（搜狐、新浪等门户网站）两大类。
全文检索搜索引擎的工作原理：通过搜索软件（如爬虫程序）到各网站搜集信息，像蜘蛛爬行一样从一个网站连接到另一个网站，然后建立一个在线索引数据库。当用户查询时，就从已经建立的索引数据库中进行查询。

Google 搜索技术的特点

Google 的核心技术是 PageRank，即**网页排名**。
将搜索结构根据重要性排名。关键字频率、是否知名网站等都会影响重要性。

6.5 电子邮件

6.5.1 电子邮件概述

电子邮件系统包括三个主要构件：**用户代理**、**邮件服务器**、**邮件协议**（包括**邮件发送协议**和**邮件读取协议**）。
用户代理就是电脑上的邮件客户端。
从用户代理把邮件发送到邮件服务器，以及邮件服务器之间的传送都要使用 **SMTP 协议**。
用户代理从邮件服务器读取邮件时则使用 **POP3 或 IMAP 协议**。

发送邮件的过程

用户代理使用 SMTP 协议把邮件发给“发送方邮件服务器”，然后“发送方邮件服务器”与“接收方邮件服务器”建立 TCP 连接并把邮件发送过去，邮件不会在某个中间服务器落地。收件人收信时，使用 POP3 协议从“接收方邮件服务器”读取邮件。
一个邮箱地址的格式是：用户名@邮件服务器的域名，如 dhb@buaa.edu.cn。

6.5.2 简单邮件传送协议 SMTP

SMTP 采用客户-服务器模式。
发件人的邮件会存在发送方邮件服务器的邮件缓存中，发送方邮件服务器（此时它是 SMTP 客户）定期扫描邮件缓存，如果有邮件就与接收方邮件服务器建立连接并发送过去。
SMTP 的熟知端口是 **25**。
SMTP 发送的是明文，不利于保密；发送邮件不需要鉴别，方便了垃圾文件的泛滥。新出的扩展的 SMTP 即 **ESMTP** 对这些进行了改进。

6.5.3 电子邮件的信息格式

略。

6.5.4 邮件读取协议 POP3 和 IMAP

常用的邮件读取协议有 POP3 和 IMAP。

POP3

POP3 采用客户-服务器模式，它非常简单、但功能有限。POP3 的特点是只要用户从 POP3 服务器读取了邮件，服务器就把该邮件删除。

IMAP

IMAP4 也采用客户-服务器模式，但是复杂地多。IMAP4 是一个联机协议，用户在自己计算机上就可以操纵邮件服务器的邮箱。
用户打开邮件时，邮件才传到用户的计算机上。用户未主动删除邮件前，IMAP 服务器邮箱中的邮件就一直保存着。
IMAP 的缺点是如果用户没有将邮件复制到自己计算机上，每次查阅邮件都必须上网。

网易邮箱大师中的服务器设置

6.5.5 基于万维网的电子邮件

基于万维网的电子邮件即用户使用浏览器收发电子邮件，这种情况用户浏览器和邮件服务器之间的传送使用 HTTP 协议，邮件服务器之间的传送仍使用 SMTP 协议。
万维网电子邮件不需要在计算机中安装用户代理软件。

6.5.6 通用互联网邮件扩充 MIME

SMTP 只能传送 **ASCII 码**，不能传送非英语文字，也不能传送可执行文件等。
通用互联网邮件扩充 MIME 对 SMTP 进行了扩充，它定义了传送非 ASCII 码的编码规则。
网络中传送的还是 ASCII 码，MIME **采用一些编码方式来用 ASCII 码表示其他字符**。
MIME 指定了几百上千种可传送的文件类型，这些类型涵盖了常用的各种文件类型。

6.6 动态主机配置协议 DHCP

因为 IP 地址中包含了网络号，而计算机第一次使用前不知道它会连到哪个网络，所以无法在出厂前就设置好 IP 地址。

当计算机的 IP 地址发生变化，比如计算机到了一个新的网络中，就要**使用动态主机配置协议 DHCP 来配置 IP 地址**，通过 DHCP 可以实现即插即用联网，而不需要人工配置 IP 地址。

配置 IP 地址的方法

DHCP 采用了客户-服务器模式。

需要配置 IP 地址的主机启动时就**向 DHCP 服务器广播发送发现报文。DHCP 收到后会给该计算机发送一个提供报文来提供分配的 IP 地址。**

响应 DHCP 客户的 DHCP 服务器可能有多个，客户机会从中选择一个给其发送请求报文。

每个网络至少有一个 **DHCP 中继代理**（一般是一个路由器），用来做主机与 DHCP 服务器之间的中转。

DHCP 服务器分配的地址有一个**租用期限制**。可能是几小时也可能是几年。当接近租用期了 DHCP 会请求更新租用期。

DHCP 客户的熟知端口是 **68**，DHCP 服务器的熟知端口是 **67**。

当一个手机从连接到一个新的 wifi 时，就要通过 DHCP 来获取新的 IP 地址。

6.7 简单网络管理协议 SNMP

6.7.1 网络管理的基本概念

网络管理包括对硬件、软件和人力的使用、综合和协调。

在一个网络管理系统中会有一个管理者和许多被管设备。被管设备可能是主机、路由器、集线器等。

每个被管设备中都要运行一个网络管理代理程序。代理程序在管理程序的命令和控制下，在被管设备上采取本地的行动。

网络管理采用的协议就是 **SNMP 协议**。

SNMP 协议

SNMP 协议中，管理程序运行 SNMP 客户程序，代理程序运行 SNMP 服务器程序。被管对象上的 SNMP 服务器程序不停监听 SNMP 客户程序的请求和命令，一旦发现就执行对应动作。

网络管理有一个基本原理：要管理某个对象，就必然要给这个对象添加一些软件或硬件，但是这种添加的影响应该尽量小一些。SNMP 最重要的思想是尽量简单。

简单网络管理协议 SNMP 包括三部分：

SNMP 本身：SNMP 定义了管理站和代理间交换的分组格式，分组中包含各代理中的变量名和状态值。SNMP 负责读取和改变这些值。

管理信息结构 SMI：定义了一套通用的规则，包括如何定义命名对象、如何定义对象类型、如何对对象编码的规则。

管理信息库 MIB：用来在被管实体中创建命名对象。

6.7.2 管理信息结构 SMI

SMI 的功能有三个：

被管对象怎样命名。

用来存储被管对象的数据类型有哪些。

在网络上传送的管理数据如何编码。

被管对象的命名

SMI 规定所有的被管对象的名字都必须在一颗对象命名树上。即类似于 URl 的命名方式。

被管对象的数据类型

SMI 把数据类型分为两大类：简单类型和结构化类型，简单类型有 Interger32 等，结构化类型有 sequence（类似结构体） 和 sequence of（类似数组） 两种

SMI 采用了抽象语法记法来定义数据类型。**抽象语法**只描述数据的结构形式，不考虑具体的编码格式，也不考虑数据结构在内存中如何存放。

编码方法

SMI 使用**基本编码规则 BER** 来进行数据编码，BER 指明了数据类型和值。它把所有的数据元素都组织为一个 T-L-V 三字段序列，T 定义数据类型，L 定义 V 字段的长度，V 定义数据的值。

6.7.3 管理信息库 MIB

管理信息就是被管对象的集合。被管对象必须维持供管理程序读写的若干控制和状态信息，这些被管对象就构成了一个虚拟的信息存储器，称为管理信息库 MIB。

只有 MIB 中的对象才是 SNMP 可以管理的。

6.7.4 SNMP 的协议数据单元和报文

实际上 SNMP 的操作只有两种基本的管理功能：

读操作：用 Get 报文来检测被管对象的状况。

写操作：用 Set 报文来改变被管对象的状况。

SNMP 使用无连接的 UDP。

SNMP 实现管理功能的方式：

使用探询操作：定期向被管设备发送探询信息，以了解其状况。

被管对象的代理检测到严重异常事件时主动向管理者发送报告。

6.8 应用进程跨越网络的通信

6.8.1 系统调用和应用编程接口

系统调用接口是应用进程的控制权和操作系统的控制权进行转换的接口，又称为**应用编程接口 API**。

API 就是应用程序和操作系统之间的接口。
API 和一般的函数调用很相似，应用程序调用 API 来将控制器传递给操作系统。
现在的 TCP/IP 协议软件是驻留在操作系统中的。**套接字接口**就是一种供应用程序使用 TCP/IP 服务的 API，Windows 系统就采用了套接字接口。
套接字是应用进程和运输层协议之间的接口，是应用进程为了获得网络通信服务而与操作系统进行交互时使用的一种机制。
理解：套接字实际上就是一套 API 接口，应用进程（应用层）通过套接字来使用位于操作系统内核的 TCP/IP 服务（运输层）。

套接字描述符

应用进程需要使用网络时，就要请求系统为其创建一个套接字，这个请求实际上是请求操作系统把网络通信所需的一些系统资源（如存储器时间、CPU 时间、网络带宽等）分配给它，操作系统使用一个**套接字描述符**来表示这些资源的总和，并将这个套接字描述符返回给应用进程。此后，应用进程的所有网络操作都要使用这个套接字描述符。
套接字描述符是套接字接口中的第一个参数。
通信完毕后，系统要回收该套接字描述符相关的所有资源。

套接字的数据结构

在机器中有一个**套接字描述符表**，其中存储了多个套接字描述符，每个进程对应一个套接字描述符，每个描述符有一个指针指向存放套接字的地址。
在套接字的数据结构中有很多参数要填写，如**协议族（PF_INET 表示 TCP/IP 协议族）、服务(SOCK_STREAM 表示 TCP 服务)、本地和远地 IP 地址、本地和远地端口**等。

6.8.2 几种常用的系统调用

下面以使用 TCP 服务为例介绍了几种常用的系统调用

并发方式工作的服务的工作模式

一个服务器要能够同时处理多个连接，即以并发方式工作。
采用一个主服务器进程 + 多个从属服务器进程是并发方式工作的一种实现方法。
主服务器进程 M 用来不停地接受新的连接请求，M 原本就有一个套接字，但是每收到一个新的请求就为它创建一个新的套接字，并把这个新的套接字的标识符返回给客户。然后它会创建一个新的从属服务器进程使用刚才创建的新的套接字来和客户建立连接。而主服务器进程 M 则使用原来的套接字继续接受下一个连接请求。

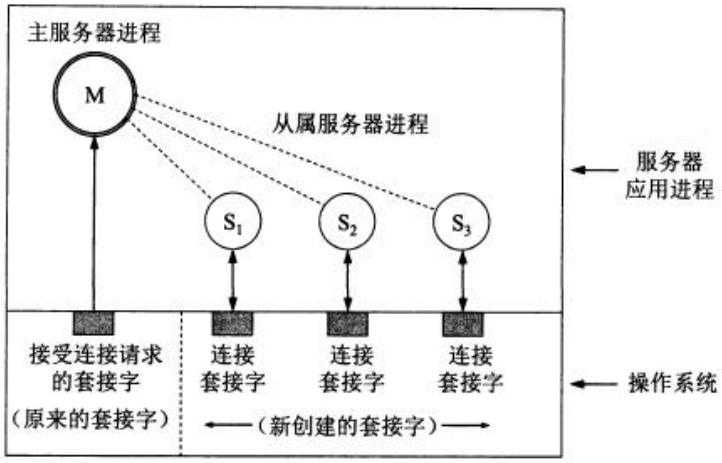


图 6-30　并发方式工作的服务器

连接建立阶段

刚创建的套接字的端口号和 IP 地址都是空的，此时服务器端的应用进程要调用 **bind** 来**指明套接字的本地端口号和本地 iP 地址**。在客户端可以调用 bind 也可以不调用而**由操作系统自动分配一个动态端口号**。

服务器调用 bind 后，还要调用 listen 把套接字设置为被动模式，来随时接受用户的服务请求。**UDP 服务器采用无连接方式，所以不使用 listen**

然后服务器（主服务器进程）要调用 **accept**，来完成给发出请求的远地客户分配从属服务器进程与新的套接字。

在客户端创建了套接字后，客户进程要**调用 connect 来向服务器发出连接请求**。在 connect 调用中，客户需指明远地服务器的 IP 地址和端口号。

数据传送阶段

客户和服务器都调用 **send** 来传送数据，调用 **recv** 来接收数据：

调用 send 需要三个变量：数据要发往的套接字的描述符，要发送的数据的地址、数据的长度。

调用 recv 也需要三个变量：要使用的套接字的描述符、缓存的地址、缓存空间的长度。

连接释放阶段

调用 **close** 来释放连接和撤销套接字。

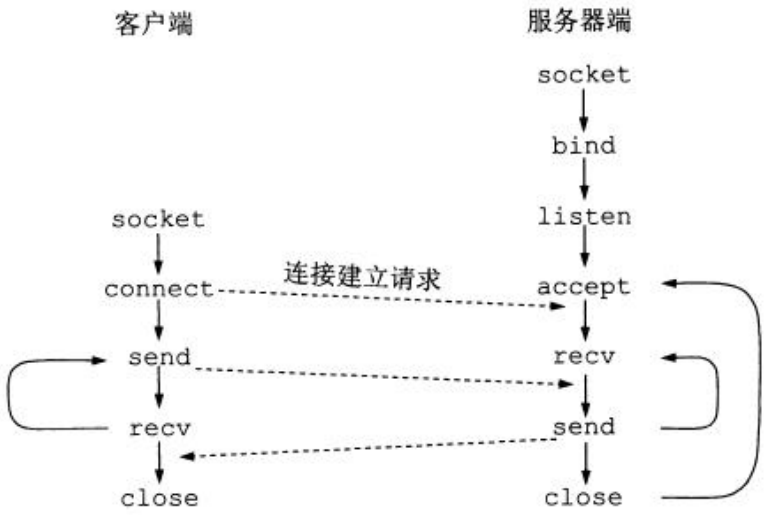


图 6-31　系统调用使用顺序的例子

6.9 P2P 应用

P2P 应用采用了 P2P 体系结构：没有固定的服务器，绝大多数交互都使用**对等方式**。
目前 P2P 工作方式下的文件共享占据了互联网流量中最大的份额，比万维网所占比例大得多。

6.9.1 具有集中目录服务器的 P2P 工作方式

第一代的 P2P 文件共享网站采用了这种方式。
这种方式下，所有用户机上文件的地址存放在一台服务器上，某个用户要下载资源时首先向该服务器询问资源地址，再从存储资源的计算机下载资源。

6.9.2 具有全分布式结构的 P2P 文件共享程序

BT 是一种很流行的 P2P 应用，采用**“最稀有的优先”**的技术，尽快把最稀有的文件块收集到。
BT 中参与某个文件分发的所有对等方构成了一个“洪流”，每一个洪流都有一个“追踪器”，当有对等方加入洪流时，要向追踪器登记。
如果有当前以最高数据率向某个对等方传送文件块的相邻对等方，该对等方就要优先把所请求的文件块传送给这些相邻对等方。这样使对等方彼此都能以较高的速率交换文件块。

6.9.3 P2P 文件分发的分析

当对等方的数量很大时，采用 P2P 下载文件比传统的客户-服务器模型快很多。

6.9.4 在 P2P 对等方中搜索对象

P2P 中广泛使用的索引和查找技术是**分布式散列表 DHT**，它实际上是一个分布式数据库。数据库中仅包含**两部分信息：关键字是资源名，值是存放对象的节点的 IP 地址**，只要给出资源名就能查到 IP 地址。但是数据库是分布式的，资源名保存在哪一台主机中呢？这就要用到基于 DHT 技术的算法。

Chord 算法是一种基于 DHT 的算法，它采用了散列函数来**将资源名映射为了一个均匀分布的数字（标识符），然后将其放到 Chord 环上**。保存资源名的主机也通过散列函数映射为一个标识符放到 Chord 环上作为环上的结点（显然结点数目远少于资源名数目），然后每个资源名就保存到 Chord 环上离他最近的结点所对应的主机中。

第 9 章 无线网络和移动网络

9.1 无线局域网 WLAN

无线局域网：Wireless Local Area Netwrok (**WLAN**)。

9.1.1 无线局域网的组成

无线局域网可以分为两大类:

有固定基础设施的无线局域网：使用了预先建立起来的基站覆盖一定范围的固定地址，比如蜂窝移动通信网使用了电信公司建立的固定基站。

有固定基础设施的无线局域网采用的是 **802.11 系列协议**。

无固定基础设施的无线局域网：移动自组网络，比如蓝牙。

802.11

无线以太网的标准是 802.11 系列协议，**使用 802.11 系列协议的局域网又称 Wi-Fi。**

理解：WLAN 表示无线局域网，Wi-Fi 表示采用 802.11 系列协议的局域网，因此 Wi-Fi 是一种局域网，且属于无线局域网。Wi-Fi 实际上已经成了 WLAN 的代名词。

802.11 无线以太网标准是用星形拓扑，其中心叫做**接入点 AP**，在 MAC 层使用 **CSMA/CA 协议**和**停止等待协议**。

802.11 规定无线局域网的最小构件是**基本服务集 BSS**。一个 BSS 包括一个基站和若干个移动站，接入点 AP 就是 BSS 内的基站。

所有的站在本 BSS 以内都可以直接通信，与其他站通信则要通过接入点 AP。

每个 AP 都有一个分配的名字，称为**服务集标识符 SSID**，它其实就是使用该 AP 的无线局域

网的名字（**也就是 wifi 名字**）。

理解：日常说的 wifi 路由器其实就是接入点 AP，而接入点 AP 本身就是一种用于 wifi 的路由器。

一个 BSS 所覆盖的地理范围叫做一个基本服务区 BSA，直径一般不超过 100m。

一个 BSS 可以是孤立的，也可以通过接入点 AP 连到分配系统（DS）。

AP 与 AP 之间的连接是有线的。

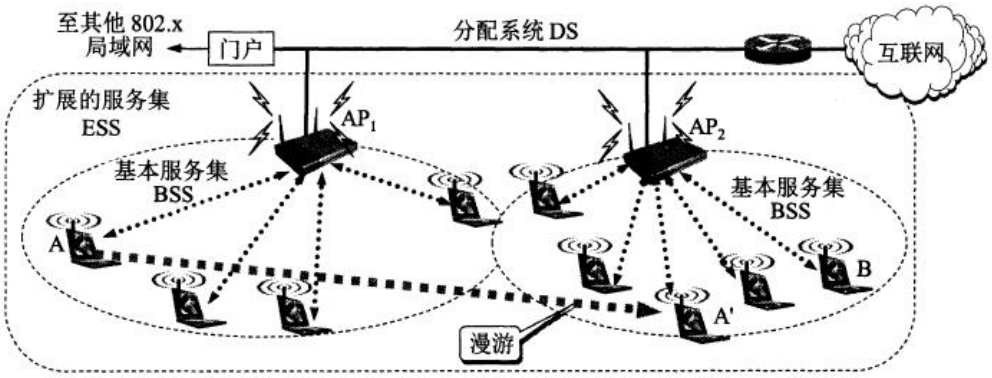


图 9-1　IEEE 802.11 的基本服务集 BSS 和扩展服务集 ESS

当一个移动站（如上图中的 A）从一个服务集漫游到了其他服务集的范围，就要选择一个接入点 AP 与之建立关联，建立关联后加入该 BSS。

移动站关联 AP 后，要通过该 AP 向该子网**发送 DHCP 发现报文以获取 IP 地址**。这之后，移动站就作为该 AP 子网的成员加入到了网络中

移动站（手机、平板电脑等）通常选择信号最强的 AP 来连接，但是一个 AP 提供的信道是有限的，如果已经耗尽了，就只能连接其他 AP。

移动站与 AP 间通信采用的协议就是 802.11 协议。

现在的手机和电脑上都有内置的无线局域网适配器，它能够实现 802.11 的物理层和 MAC 层的功能。

现在的无线局域网一般采用了**加密方案 WPA 或 WPA2**，这时要加入该无线局域网就要输入密码。

移动自组网络

无固定基础设施的无线局域网叫做**移动自组网络**。蓝牙就是一种自组网络。

移动自组网络没有基站，而是由一些处于平等状态的移动站相互通信组成的**临时网络**。

自组网络一般不和外界的其他网络相连接。

无线传感器网络 WSN 是一种近年来发展很快的移动自组网络。它由大量传感器结点通过无线通信技术构成。物联网 IoT 就是 WSN 的应用领域。

9.1.2 802.11 局域网的物理层

802.11 无线以太网标准是用星形拓扑，其中心叫做**接入点 AP**，接入点 AP 就是基本服务集内的基站。

wifi 的历史版本

2020 年 6 月正式发布的 **wifi6** 的标准是 **IEEE 802.11ax**，又称 HEW（High Effieciency WLAN），wifi6 标准支持 1GHz 到 6GHz 的所有 ISM 频段，包括 6GHz 和目前使用的 2.4GHz

和 5GHz，向下兼容 a/b/g/n/ac。包含多种带宽，其中最高带宽为 **160MHz**,数据速率为**单条流最高 1201Mbit/s**

第 4 代 wifi 是 802.11n，第 5 代 wifi 是 802.11ac

9.1.3 802.11 局域网的 MAC 层协议

区分 CSMA/CA 协议和 CSMA/CD 协议：

CSMA/CA：载波监听多点接入/碰撞避免

CSMA/CD：载波监听多点接入/碰撞检测

802.11 无线以太网在 MAC 层使用 **CSMA/CA 协议和停止等待协议**。

使用停止等待协议是因为无线信道的通信质量远不如有限信道，要使用停止等待来保证**可靠传输**。

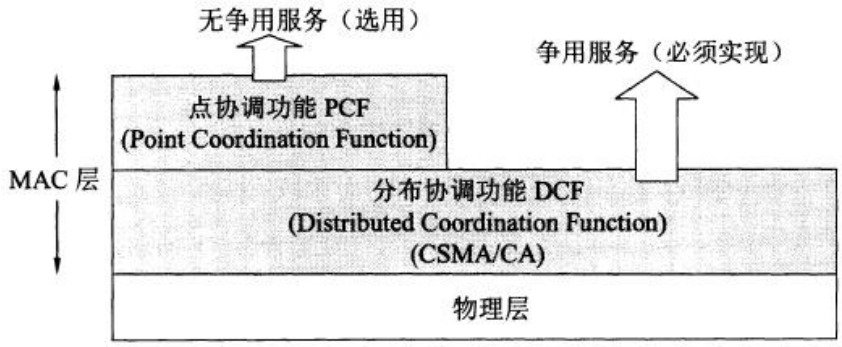
无线局域网中不能使用 CSMA/CD 协议，因为：无线局域网中不是所有的站点都能听见对方，因此无法实现碰撞检测。使用 CSMA/CA 协议是为了减小碰撞发生的概率。

802.11 的 MAC 层协议通过**协调功能**来确定基本服务集 BSS 中的各移动站**在什么时间**发送和接收数据。

它包括两个子层

分布协调功能 DCF：DCF 不采用中心控制，它在每一个结点使用 CSMA 机制的分布式接入算法，让各个移动站通过**争用信道**来获取发送权。

点协调功能 PCF：PCF 是选项，它用接入点 AP 集中控制整个 BSS 内各移动站的活动，使用类似探询的方法将发送数据权轮流交给各个站，以避免碰撞发生。**对时间敏感的业务应该采用 PCF**。



802.11 无线局域网中的 MAC 帧首部有一个持续期字段，它指出在本帧结束后还要占用信道多少时间。

802.11 标准允许**要发送数据的站对信道进行预约**，即**在发送数据帧之前先发送 RTS 帧请求发送**，收到响应允许发送的 CTS 帧后，就可发送数据帧。

802.11 采用了几种机制：

帧间间隔时间：所有的站在完成发送后，必须再等待一段**帧间间隔时间**才能发送下一帧。不同优先级的帧具有不同的帧间间隔 IFS。这可以降低碰撞概率

虚拟载波监听：源站把要占用信道的时间通知给其他所有站，其他站在这段时间就停止发送（其他站实际上没有监听，因此叫做虚拟载波监听），这可以降低碰撞概率。

随机退避算法：当某个站发现信道变为空闲时，要等待一个 DIFS 的间隔，再执行退避算法，维护一个退避计时器，计时器归零后就立即发送。这样也可以降低碰撞概率。

预约机制：要发送数据的站对信道进行预约，即**在发送数据帧之前先发送 RTS 帧请求发送**，收到响应允许发送的 CTS 帧后，就可发送数据帧。预约帧中会指明预约时间，期间其他站不会再发送数据。用户可以选择性使用预约机制。

9.1.4 802.11 局域网的 MAC 帧

802.11 的 MAC 帧有**三种类型：控制帧、数据帧、管理帧**。

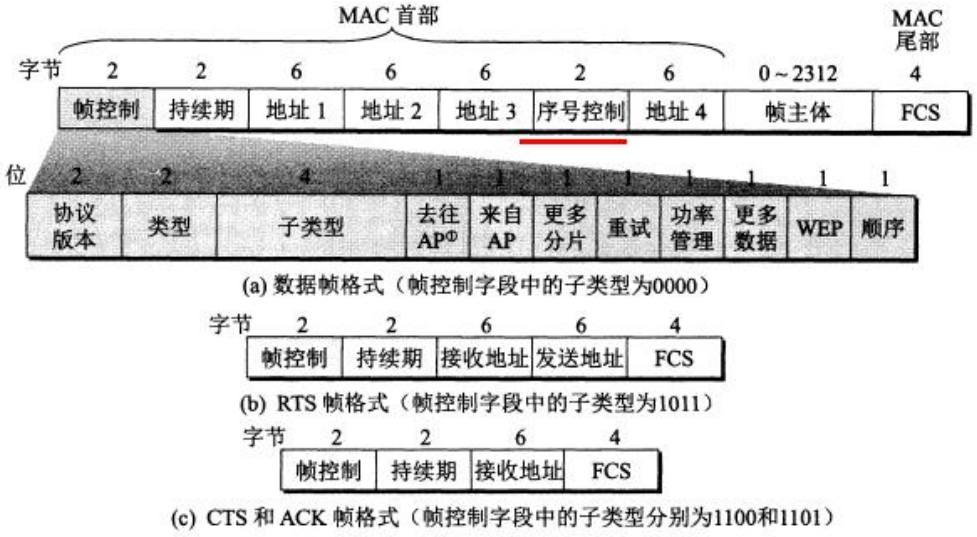


图 9-10　802.11 局域网的帧格式

MAC 帧的首部有 30 字节，尾部是 4 字节的帧检验序列。

802.11 数据帧的地址

802.11 数据帧有 4 个地址字段。地址 4 用于自组网络。
前三个地址的内容取决于帧控制字段中的“去往 AP”和“来自 AP”。

表 9-2　802.11 帧的地址字段最常用的两种情况

去往 AP	来自 AP	地址 1	地址 2	地址 3	地址 4
0	1	目的地址	AP 地址	源地址	——
1	0	AP 地址	源地址	目的地址	——

9.2 无线个人区域网 WPAN

无线个人区域网 WPAN 就是把个人设备用无线技术连起来的自组网络。
WPAN 都工作在 **2.4GHz 频段**。
无线个人区域网包括：**蓝牙系统**、ZigBee、超高速 WPAN 等。

9.3 无线城域网 WMAN

无线城域网 WMAN 可提供最后一英里的宽带无线接入，可以用来代替现在的有线宽带接入。
无线城域网的标准是 802.16 系列协议，它可以覆盖一个城市的部分区域。

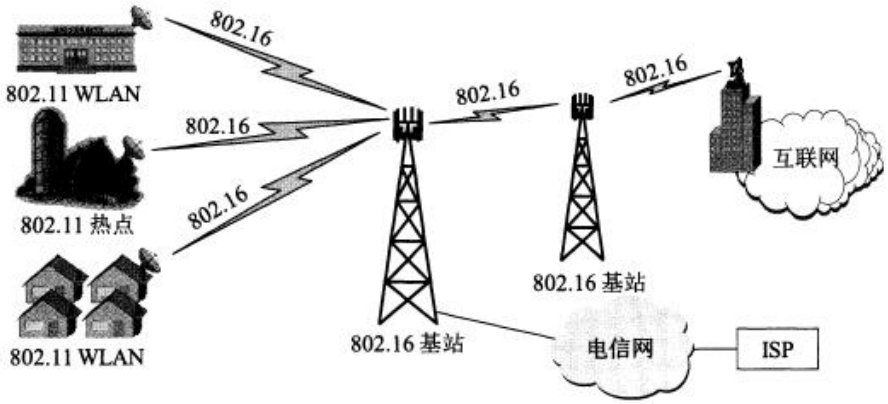


图 9-16　802.16 无线城域网服务范围示意图

9.4 蜂窝移动通信网

9.4.1 蜂窝无线通信技术简介

移动通信有多种，如蜂窝移动通信、卫星移动通信等。
蜂窝移动通信网原来属于通信领域，但是现在的蜂窝移动通信网采用了许多 iP 技术，可以支持手机、电脑上网。
蜂窝移动通信是小区制的移动通信，它把整个网络划分成许多小区（也就是蜂窝），每个小区设置一个基站。移动站的通信都必须通过基站完成。

2G 即第二代蜂窝移动通信，带宽为 200kHz，它的代表是 GSM 系统，GPRS 也是 2G 的一种技术。2G 基本只能提供电话和短信服务。
3G 的带宽为 5MHz，并使用了 IP 的体系结构和混合的交换体制（电路交换和分组交换），3G 以后的蜂窝移动通信就是以传输业务为主的通信系统了。
3G/4G 时代诞生了上网卡，上网卡像一个 U 盘，可以插到电脑的 USB 接口上，然后电脑就可以通过 3G/4G 蜂窝移动网络接入互联网。
使用蜂窝移动通信是与同一个蜂窝小区的其他用户**共享带宽**的，每个用户实际分配到的带宽是不确定的。
小区的组成像蜂窝一样，每个基站的发射功能要能够覆盖本小区，又不会太大以至干扰邻近小区。采用蜂窝结构可以最大化频分复用，每个基站使用不同的频率。

3G 通信网络的构件

无线网络控制器 RNC 控制一组基站，基站通过 RNC 连接到移动交换中心 MSC，MSC 控制所有 RNC 的话音业务，MSC 可以通过网关移动交换中心连接到公用电话网。
RNC 还连接到 GPRS 核心网络，当移动站要上网，就通过 GPRS 来进行。
RNC 处于无线接入网的边缘，它进行无线通信和有线通信的转换。在有线通信这边，RNC 把电路交换的话音通信传送到 MSC，把分组交换的数据传送到 GPRS。

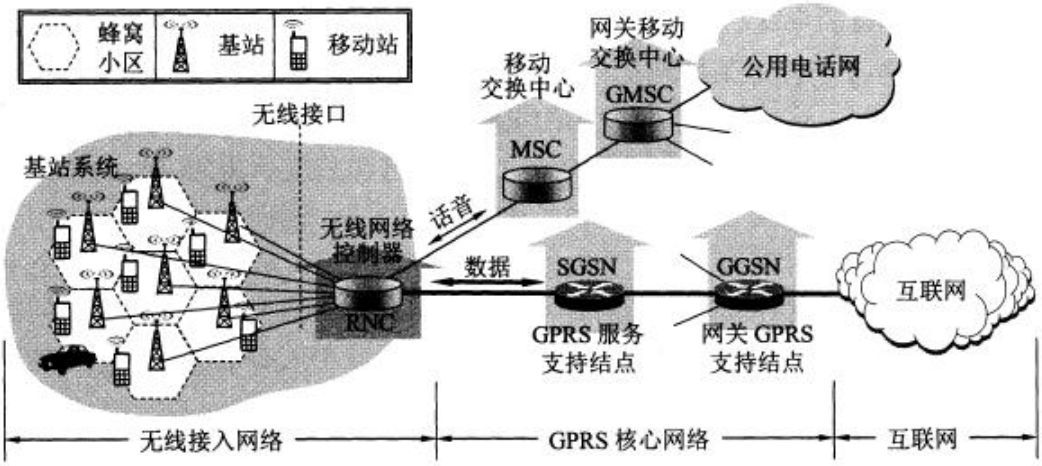


图 9-17　3G 蜂窝通信系统的重要组成构件

9.4.2 移动 IP

概念

当计算机移动到外地，移动 IP 技术允许该计算机**仍保留原来的 IP 地址**。
移动 iP 与 DHCP 的区别：当用户带着电脑换了个位置，离开了原来的网络，通过 DHCP 协议就可以自动获取所需的 IP 地址。而移动 IP 用于**在移动中上网**
移动 IP 使用了**一些新概念：永久地址、归属地址、归属网络、被访网络、归属代理、外地代理、转交地址**、同址转交地址等。
移动 IP 使用了**几种协议：移动站到外地代理的协议、外地代理到归属代理的登记协议、归属代理数据报封装协议、外地代理拆封协议**等。

实现

一个移动站必须有一个原始地址，即**归属地址**（又称**永久地址**），移动站原始连接到的网络叫做**归属网络**。

移动 IP 通过使用代理来让地址的改变对互联网的其他部分是透明的。归属代理通常就是连接在归属网络上的路由器。

当移动站 A 移动到另一个地点，所接入的网络叫做**外地网络**（又叫**被访网络**），被访网络中中使用的代理叫**外地代理**，通常是连接在被访网络上的路由器。

外地代理负责两件事：

为移动站 A 创建一个临时的地址：**转交地址**。转交地址的网络号与外地网络一致。

及时把移动站 A 的转交地址通知 A 的归属代理。

注意：转交地址供移动站、外地代理、归属代理使用，各种应用程序都不使用转交地址。转交地址也不具有唯一性，外地代理向移动站 A 发送数据时不使用地址解析协议 ARP，而是用移动站 A 的 MAC 地址。

有时移动站本身也可以作为外地代理。

一个通信的例子

一个通信者 B 要与移动站 A 进行通信，其步骤如下：

B 发送给 A 的数据报目的地址是 A 的永久地址，它被 A 的归属代理所截获。

归属代理把 B 发来的数据报再封装（使用了隧道技术），新的数据报的目的地址是 A 现在的转交地址，它被发送给被访网络中的外地代理。

外地代理把收到的数据报拆封，取出 B 发送的原始数据发送给 A。A 收到 B 的数据报后也就知道了 B 的 IP 地址。

如果 A 要回复 B，就使用自己的永久地址作为源地址，用 B 的 IP 地址为目的地址发送数据报。这时不再需要通过 A 的归属代理。

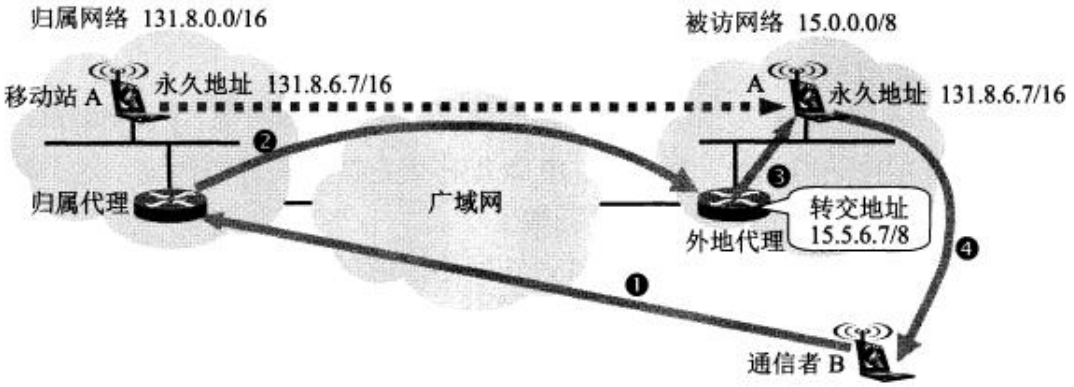


图 9-18　永久地址与转交地址的作用

移动 IP 使用了几种协议：

移动站到外地代理的登记协议：移动站接入到被访网络时，要**向外地代理登记，获得临时的转交地址**。离开该被访网络时要注销。

外地代理到归属代理的登记协议：外地代理向归属代理登记移动站的转交地址。

归属代理数据报封装协议：将收到的发给移动站的数据报进行封装，以转交地址为新的目的地址。

外地代理拆封协议：收到数据报后拆封并发给移动站。

9.4.3 蜂窝移动通信网中对移动用户的路由选择

移动 IP 的路由选择有间接路由选择和直接路由选择。上述方法就是间接路由选择。直接路由选择需要使用通信者代理和锚外地代理。
移动交换中心维护了两个数据库：

归属位置寄存器 HLR：类似归属代理的功能。

来访用户位置寄存器 VLR：类似外地代理的功能。

9.4.4 GSM 中的切换

略。

9.4.5 无线网络对高层协议的影响

移动站在不同无线网络间漫游时，网络的连接会发生中断。TCP 报文段会频繁丢失，TCP 的拥塞控制会受到影响，缩小拥塞窗口，而实际上网络中并不拥塞。
处理的方法：

本地恢复。

让 TCP 发送方知道什么地方使用了无线链路。

让含有移动用户的端到端 TCP 连接**拆成两个互相串接的 TCP 连接**：从移动用户到无线接入点一个 TCP 连接，剩下的有线网络使用另一个 TCP 连接。

9.5 两种不同的无线上网

蜂窝移动网络的收费采用的是按流量计费
wifi 是通过宽带上网的，宽带入网的收费是根据用户使用的带宽和使用时间收费的。

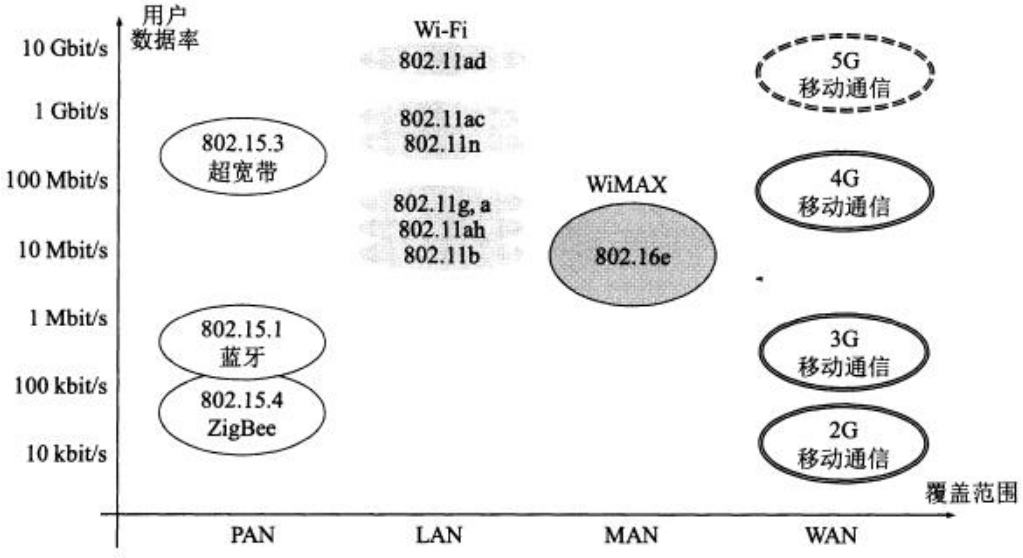


图 9-22　几种无线网络的比较