第 1 章 计算机网络基础

计算机网络逐渐改变了人们的生活和工作方式，引起了世界范围内的产业革命，在各国的政治、经济、文化、军事、教育和社会生活等各个领域发挥着越来越重要的作用。本章首先介绍计算机网络的定义、产生和发展，然后讲解计算机网络的结构、类型和通信方法等基础概念。

1.1 计算机网络简介

计算机网络是计算机技术、通信技术和网络技术相结合的产物，是现代社会重要的基础设施，为人类获取和传播信息发挥了巨大的作用。下面首先讲解计算机网络的定义、产生和发展。

1.1.1 计算机网络的定义

计算机网络最简单的定义是：一些相互连接的、以共享资源为目的的、自治的计算机的集合。而最通用的定义是：计算机网络是指将地理位置不同的具有独立功能的多台计算机及其外部设备，通过通信线路连接起来，实现资源共享、信息传递和协同工作的计算机系统。

综上所述，计算机网络具备以下 3 个基本要素，三者缺一不可。

(1)不同地理位置、独立功能的计算机。在计算机网络中，每一台计算机都具有独立完成工作的能力，并且计算机之间可以不在同一个区域(如同一个校园、同一个城市或同一个国家等)。

(2)交互通信、资源共享及协同工作。资源共享是计算机网络的主要目的，而交互通信和协同工作则是计算机网络实现资源共享的重要前提。在计算机网络中，既可使用同轴电缆、光纤等有线传输介质，也可借助于微波、卫星等无线传输介质来实现多台计算机之间的

通信互连。例如，用户可以通过 Internet 为代表的计算机网络传递文件、发布信息、查阅资料、获取信息等。

(3)标准通信规则或协议。在计算机网络中，计算机需要互相通信时，它们之间必须使用相同的语言。而这种语言既是通信的规则，也是一种通信协议。

1.1.2 计算机网络的产生和发展

早在 1951 年，美国麻省理工学院林肯实验室就开始为美国空军设计称为 SAGE(semi-automatic ground environment)的半自动地面防空系统。该系统由 17 个防区组成，每个防区的指挥中心装有两台 IBM 公司的 AN/FSQ-7 计算机，通过通信线路连接防区内各雷达观测站、机场、防空导弹和高射炮阵地，由计算机程序辅助指挥员决策，自动引导飞机和导弹进行拦截。该系统最终于 1963 年建成，被认为是计算机和通信技术结合的先驱。

1966 年，被誉为“ARPANET 之父”的罗伯茨开始全面负责 ARPANET 的筹建。经过近一年的研究，罗伯茨选择了一种名为接口报文处理机(interface message processor，IMP，是路由器的前身)的技术来解决网络间计算机的兼容问题，并首次使用了“分组交换”(packet switching)作为网间数据传输的标准。这两项关键技术的结合为 ARPANET 奠定了重要的技术基础，创造了一种更高效、更安全的数据传递模式。

1968 年，一套完整的设计方案正式启用，同年，首套 ARPANET 的硬件设备问世。1969 年 10 月，罗伯茨完成了首个数据包通过 ARPANET 由加州大学洛杉矶分校出发，经过漫长的海岸线，完整无误地抵达斯坦福大学的实验。

在这之后，罗伯茨还不断完善 ARPANET 技术，从网络协议、操作系统再到电子邮件。1969 年 12 月，Internet 的前身——采用分组交换技术的 ARPANET 正式投入运行，它标志着计算机网络的兴起。分组交换技术使计算机网络的概念、结构和网络设计等方面都发生了根本性的变化，并为后来的互联网的发展打下了坚实的基础。ARPANET 启动了有 4 个节点的实验性网络，包括加州大学洛杉矶分校(UCLA)、加州大学圣塔芭芭拉分校(UCSB)、斯坦福研究院(SRI)和犹他大学。这些研究最终产生了现在的 TCP/IP(1974 年)模型和协议，随后加州大学伯克利分校把 TCP/IP 集成到了 Berkeley UNIX 操作系统(BSD UNIX 4.2)中，使得网络开发和网络互连变得更加容易，这一做法极大地推动了 TCP/IP 的推广和应用，也使得网络规模迅速扩大，最终形成今天的 Internet。

另一个对现代计算机网络影响至关重要的是以太网。施乐帕克研究中心(Xerox Palo Alto Research Center，PARC)的 Metcalfe 在 1972 年受命把施乐公司第一台名为 ALTO 的带图形用户界面的个人计算机(PC)连到 ARPANET，1972 年年底，Metcalfe 和 David Boggs 根据 ALOHA 的原理设计了一套网络，将不同的 ALTO 计算机连接起来，接着又把 NOVA 计算机连接到 EARS 激光打印机，产生了世界上第一个个人计算机局域网络。1973 年 5 月 22 日，该网络开始运行，Metcalfe 根据“电磁辐射是通过发光的以太来传播的”这一想法，将该网络改名为以太网(Ethernet)，网络的传输速率为 2.94 Mbit/s。1980 年 9 月 30 日，DEC、Intel 和 Xerox 公布了第三稿的“以太网，一种局域网：数据链路层和物理层规范，1.0 版”，完成著名的以太网蓝皮书，也称为 DIX(三家公司名称的第一个字母的组合)版以太网 1.0 规范。该网络数据传输速率为 10 Mbit/s。美国电气和电子工程师协会(IEEE)定义与促进局

域网标准的委员会(简称 802 委员会)在 1981 年 6 月成立了 802.3 分委员会,研究以太网的标准化问题,并在 1983 年通过了 802.3 标准,也就是以太网标准。1989 年 ISO 88023 采用 802.3 标准作为国际标准,从而使得以太网得到国际承认。

目前,广泛应用的 Internet 中的计算机大多使用 TCP/IP 和 IEEE 802.3 系列协议进行相互通信。

1.1.3　我国计算机网络的发展

我国计算机网络发展迅猛,大规模的网络建议是在 1989 年开始的,当时的国家计划委员会决定利用世界银行贷款筹建中国国家计算机与网络设施(National Computing and Network Facility of China,NCFC),该网络由北京大学、清华大学和中国科学院的 3 个子网互连构成。1994 年 5 月,它作为我国第一个互联网与 Internet 连通,使中国成为加入 Internet 的第 81 个国家,并在此基础上建立了由全国 100 所高校和科研单位参加的"中国教育和科研计算机网络(CERNET)"之后,几个全国范围的计算机信息网络相继建成,并开始提供 Internet 接入服务,Internet从此在我国得到了迅猛发展。

2012 年 1 月 16 日,中国互联网络信息中心(CNNIC)的《第 29 次中国互联网络发展状况统计报告》给出了如图 1-1 所示的国内互联网发展状况示意图。报告显示,2011 年年底,我国网民规模已突破 5 亿。与此同时,网站数量在 2011 年下半年实现止跌并快速回升。该报告显示,截至 2011 年 12 月底,我国网民规模达到 5.13 亿,全年新增网民 5 580 万;互联网普及率较上年底提升 4 个百分点,达到 38.3%。其中,我国手机网民数量达到 3.56 亿,同比增长 17.5%。手机网民在总体网民中的比例达 69.4%,成为中国网民的重要组成部分。

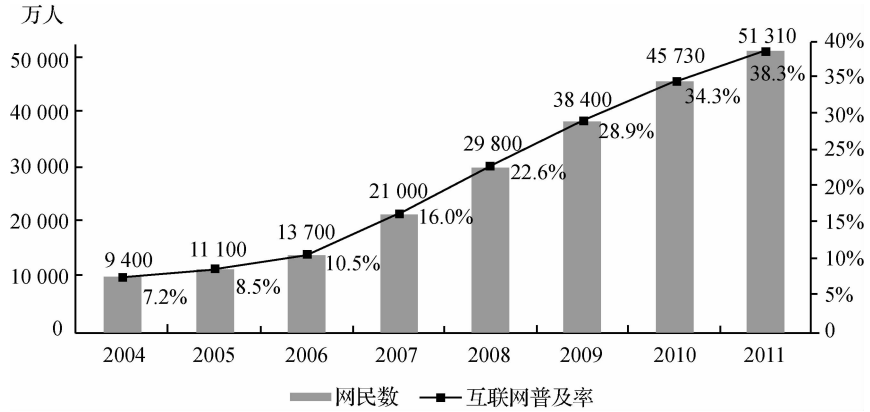


图 1-1　我国互联网发展状况示意图

1.1.4　计算机网络的应用

计算机网络已经成为现代生活必不可少的工作、学习和娱乐工具,我国互联网信息中心将网络应用划分为"信息获取""交流沟通""商务交易"和"网络娱乐"等几类,从计算机网络的角度来看,其主要应用包括如下几方面。

1)资源共享

设计和建立计算机网络的主要目的之一是实现不同计算机上的资源共享,这些资源包括硬件资源(如打印机、大容量驱动器以及扫描仪等贵重设备和不太常用的设备)、软件资源(如大型数据库系统、应用程序等)以及信息资源(如数据库中的共享数据、网页信息和数据文件等)。

2)数据通信

因为计算机网络是通过某种通信手段将不同地理位置的计算机连接起来形成的系统,所以网络中的计算机必然可以通过通信手段实现不同计算机之间的数据通信和数据传输,常用的文件下载、用聊天工具进行网络聊天等,都可以归结为不同计算机之间的数据通信。

3)远程教育

由于计算机覆盖的地理范围非常广,网络中可以传输各种各样的数据,利用这种功能可以开办远程教育,使本地的学生通过网络聆听专家学者的远程授课,从而使得教育资源能够得到更充分的利用。目前,通过网络学习已成为获得持续知识的主要方法之一。

4)电子商务

现代计算机网络的普及和高速发展使得通过计算机网络销售和购买商品成为计算机网络的重要功能之一。我国互联网络信息中心 2009 年的调查报告指出:“目前中国网民中,大约 4 个人中有 1 个人是网购用户,而在欧美和韩国等互联网普及率较高的国家和地区,每 3 个网民中就有 2 个人在网上购物。中国网络购物的潜力还远未被释放。此外,政府已相当重视电子商务对经济的拉动作用,出台了一系列政策规范和引导电子商务发展;业界电子商务的发展也如火如荼,不仅涌现出许多平台类电子商务网站,也有越来越多有远见的传统企业开始进军电子商务。在这种大形势下,预期未来几年电子商务会保持快速发展之势。”

5)休闲娱乐

网络也同样改变了人们的休闲娱乐方式,越来越多的人开始利用网络游戏、网络音乐和网络视频进行休闲娱乐,网络休闲娱乐已成为互联网最主要的功能之一。

6)人际交流

现在,计算机网络已经逐渐成为人际交流的重要工具。人们通过电子邮件服务和即时通信软件(如 QQ)交流工作安排、学习计划、出行要求等,利用即时通信软件还可以进行语音通信和视频通信。尤其是网络流行的“微博”、“博客”和“论坛”等,使得任何人都可以通过网络发布自己的观点和看法,也可以对别人的看法提出不同的意见。

1.2 计算机网络的结构与分类

计算机网络是一种极为复杂的计算机系统,其内部不但包含了数量众多的计算机和连接这些计算机的通信线路及设备,还需要通过软件对其进行合理的管理与控制,以实现不同计算机间的相互通信与资源共享。因此,为了更好地学习计算机网络知识,首先需要对计算

机网络的功能结构和拓扑结构有所了解。

1.2.1　计算机网络的功能结构

人们组建计算机网络的目的是实现不同位置计算机间的相互通信和资源共享，从计算机网络各组成部件所完成的功能来划分，可以将计算机网络划分为通信子网和资源子网两大部分，如图 1-2 所示。

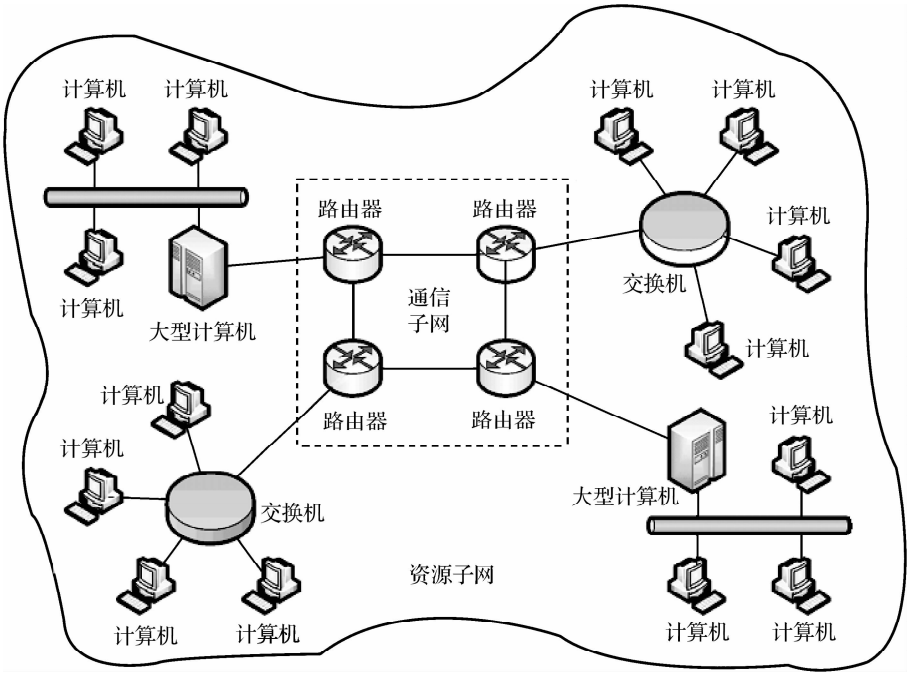


图 1-2　通信子网与资源子网

1)通信子网

多台计算机间的相互连通是组成计算机网络的前提，通信子网用于实现网络内多台计算机间的数据传输。通常情况下，通信子网由以下几部分组成。

(1)传输介质。传输介质是数据在传输过程中的载体，计算机网络内常见的传输介质分为有线传输介质和无线传输介质两种类型。

有线传输介质是指能够使两个通信设备实现互连的物理连接部分。计算机网络发展至今，共使用过同轴电缆、双绞线和光纤 3 种不同的有线传输介质。

无线传输介质是一种不使用任何物理连接，而是通过空间进行数据传输，以实现多个通信设备互连的技术，主要有红外线、激光以及微波等。

(2)中继器。中继器安装于传输介质之间，其作用是再生放大数字信号，以扩大网络的覆盖范围。

(3)集线器和交换机。集线器也叫集中器，在网络内主要用于连接多台计算机。随着网络技术的发展和应用需求的不断变化，具有更多功能及更高工作效率的交换机已经逐渐取

代了集线器。

(4)网络互连设备。随着计算机网络数量的增多,人们开始利用网桥、网关和路由器等网络互连设备来连接位于不同地理位置的计算机网络,以扩大计算机网络的规模,提高网络资源的利用率。

①网桥用于连接相同结构的局域网,以扩大网络的覆盖范围,并通过降低网络内冗余信息的通信流量来提高计算机网络的运行效率。

②网关通常位于不同类型的网络之间,以实现不同网络内计算机之间的相互通信。

③路由器一般用于连接较大范围的计算机网络,其作用是在复杂的网络环境中,为数据选择传输路径。

(5)调制解调器(Modem)。Modem 的功能是实现数字信号与模拟信号之间的相互转换,主要用于传统的拨号上网方式。

2)资源子网

对于计算机网络用户而言,资源子网实现了面向用户提供和管理共享资源,是计算机网络的重要组成部分,通常由以下几部分组成。

(1)服务器。服务器是计算机网络中向其他计算机或网络设备提供服务的计算机,通常会按照所提供服务的类型被冠以不同的名称,如数据库服务器、邮件服务器等。

(2)客户机。客户机是一个与服务器相对应的概念。在计算机网络中,享受其他计算机所提供服务的计算机就称为客户机。

(3)打印机、传真机等共享设备。共享设备是计算机网络共享硬件资源的一种常见方式,打印机、传真机等设备是较为常见的共享设备。

(4)网络软件。网络软件主要分为服务软件和网络操作系统两种类型。其中,网络操作系统管理网络内的软、硬件资源,并在服务软件的支持下为用户提供各种服务项目。

1.2.2 计算机网络的拓扑结构

拓扑(topology)是一种不考虑物体的大小、形状等物理属性,而仅仅使用点或线描述多个物体实际位置与关系的抽象表示方法。拓扑不关心事物的细节,也不在乎相互的比例关系,而只是以图的形式来表示一定范围内多个物体之间的相互关系。

网络拓扑结构形象地描述了网络的安排和配置方式,以及各种节点之间的相互关系,通俗地说,拓扑结构就是指这些计算机与通信设备是如何连接在一起的。了解网络的拓扑结构是认识网络的基础,也是设计、组建计算机网络时必须考虑的问题。

网络拓扑结构主要有星型结构、环型结构、总线型结构、树型结构和网状结构 5 种类型。

1)星型拓扑结构

星型拓扑结构以中央节点为中心,其他各节点与中央节点通过点与点的方式进行连接。例如,使用集线器组建而成的局域网便是一种典型的星型结构网络,如图 1-3 所示。

在星型拓扑结构中,由于任何两台计算机要进行通信都必须经过中央节点,因此中央节点需要执行集中式的通信控制策略,以保证网络的正常运行,这使得中央节点的负担往往较重。其优点是网络结构简单,便于集中控制与管理,组网较为容易;其缺点是网络的共享能

力较差，通信线路的利用率较低，且中央节点负担较重，一旦中央节点出现故障便会导致整个网络瘫痪。

2)环型拓扑结构

环型拓扑结构内的各节点通过环路接口连在一条首尾相连的闭合环型通信线路中，其结构如图 1-4 所示。在环型网络中，一个节点发出的信息会穿越环内的所有环路接口，并最终流回至发送该信息的环路接口。而在这一过程中，环型网内的各节点(信息发送节点除外)通过对比信息流内的目的地址来决定是否接收该信息。

环型拓扑结构的优点是由于信息在网络内沿固定方向流动，并且两个节点间仅有唯一的通路，简化了路径选择的控制。

环型拓扑结构的缺点是由于使用串行方式传递信息，因此当网络内的节点过多时，将严重影响数据传输效率，使网络响应时间变长。此外，环型网的扩展较为困难。

环型拓扑结构是局域网中较为常用的拓扑结构之一，较为适合信息处理系统和工厂自动化系统。在众多的环型网络中，由 IBM 公司于 1985 年推出的令牌环网(IBM token ring)是环型网络的典范。

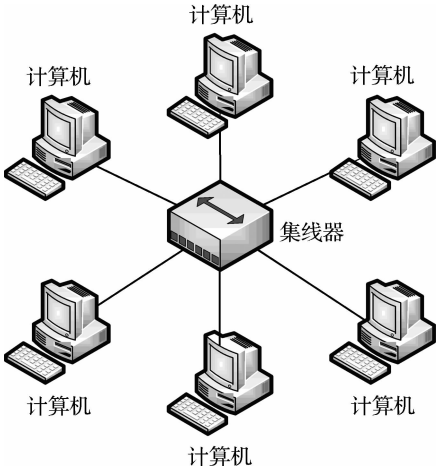


图 1-3　星型拓扑结构

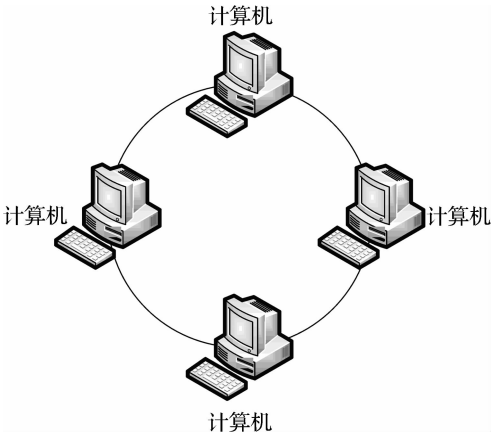


图 1-4　环型拓扑结构

3)总线型拓扑结构

使用一条中央主电缆将相互间无直接连接的多台计算机联系起来的布局方式，称为总线型拓扑结构，其中的中央主电缆便称为“总线”，其结构如图 1-5 所示。

在总线型网络中，所有计算机都必须使用专用的硬件接口直接连接在总线上，任何一个节点的信息都能沿着总线向两个方向传输，并且能被总线上的任何一个节点所接收。由于总线型网络内的信息向四周传播，类似于广播电台，因此总线型网络也被称为广播式网络。

总线型网络只能使用同轴电缆作为传输介质。并且，为了避免传输至总线两端的信号反射回总线产生不必要的干扰，总线两端还需要分别安装一个与总线阻抗相匹配的终结器(末端阻抗匹配器，或称终止器)，以最大限度地吸收传输至总线端部的能量。

4)树型拓扑结构

树型拓扑结构是一种层次结构，由最上层的根节点和多个分支组成，各节点按层次进行

连接，数据交换主要在上下节点之间进行，相邻节点或同层节点之间一般不进行数据交换，其结构如图 1-6 所示。

树型拓扑结构的优点是连接简单，维护方便。树型拓扑结构的缺点是资源共享能力较弱，可靠性比较差，任何一个节点或链路的故障都会影响整个网络的运行，并且对根节点的依赖过大。

在组建树型结构网络的过程中，分支与分支间不能相互连接，以避免因环路而产生的网络错误。

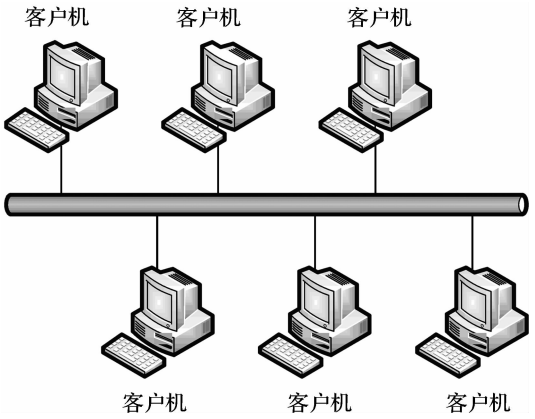


图 1-5　总线型拓扑结构

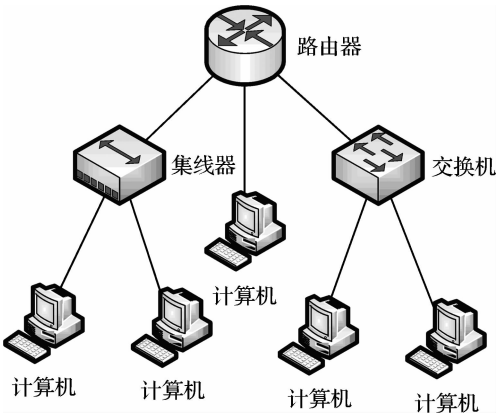


图 1-6　树型拓扑结构

5)网状拓扑结构

将多个子网或多个网络连接起来构成网状拓扑结构。在一个子网中，集线器、中继器将多个设备连接起来，而网桥、路由器和网关则将子网连接起来。

网状拓扑结构中的各节点通过传输线互连，并且每一个节点至少与其他两个节点相连，如图 7-1 所示。网状拓扑结构具有较高的可靠性，但其结构复杂，实现起来费用较高，不易管理和维护，因此不常用于局域网。

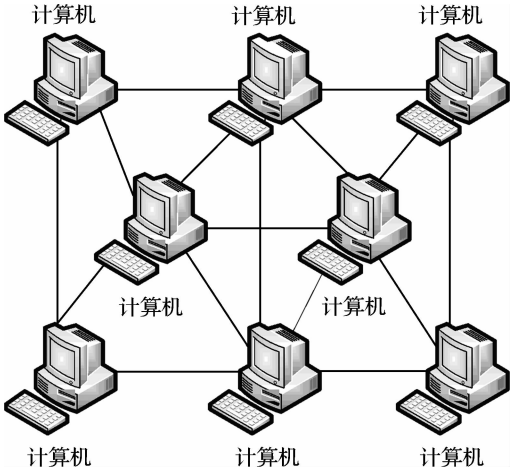


图 1-7　网状拓扑结构

1.2.3　计算机网络的分类

根据不同的划分方式，计算机网络可以分为不同的类型。

1)依据网络共享服务方式划分

根据网络共享服务方式的不同，网络可以划分为对等网络(peer-to-peer)、客户机/服务器网络和混合网络。

(1)对等网络是指网络中的计算机是平等的，网络中的每台计算机既提供网络服务也共享网络服务，没有主从之分。

(2)客户机/服务器网络是指网络中的主机划分为客户机和服务器，服务器提供所需要的网络服务，客户机不提供网络服务，只享受网络服务。需要注意的是，在这种网络中经常存在一台计算机既运行客户程序也运行服务器程序，看起来好像也是平等的，但实际上是不同的，因为同一个程序不能同时提供服务并享受服务，这一点和对等网络不同。

(3)混合网络是指网络中同时有两种服务存在，既有对等的网络服务，也有客户机/服务器服务方式。

2)依据节点分布的地理范围划分

根据网络节点分布的地理范围，可以将网络分为局域网、城域网、广域网和互联网。网络的规模可以以网上相距最远的两台计算机之间的距离来衡量。

(1)局域网。局域网(local area network，LAN)一般是指网络中的计算机分布在相对较小的区域，通常不超过 10 km，其应用场合如下。

①连接同一房间内的所有主机，覆盖范围为几十米。

②连接同一楼内的所有主机，覆盖范围为 100 m 左右。

③连接同一校园内、厂区内或院落内的所有主机，覆盖范围为 1 km 左右，也被称为校园网或者园区网。

(2)城域网。城域网(metropolitan area network，MAN)是指网络中的所有主机(工作站点)分布在同一城区内，覆盖范围在 10～100 km。

(3)广域网。广域网(wide area network，WAN)是指网络中所有主机与工作站点分布的地理范围能够覆盖10 km以上的范围，包括 10 km、100 km 与 1 000 km 及以上的数量级，如同一个城市、同一个国家或同一个洲甚至跨越几个洲等。

计算机网络的地理范围实际上也是采用不同的技术来划分的，通常来讲，局域网一般采用 IEEE 802.3、802.4、802.5 和 802.11(无线分组交换网)等局域网协议；城域网采用 802.6，即 DQDB 协议，也可以采用其他网络协议；广域网采用 TCP/IP。

3)依据网络的传输介质划分

根据计算机网络所采用的传输介质的不同，可以将计算机网络划分为有线网和无线网两种类型。

(1)有线网。有线网主要是指采用双绞线、同轴电缆和光纤来连接的计算机网络。

双绞线的价格便宜，安装方便且较为灵活，是目前局域网中最常见的传输介质。双绞线的缺点是容易受到干扰，且传输距离比同轴电缆要短。

光纤(光导纤维)传输距离长,传输速率高,抗干扰能力强,且不会受到电子监听设备的监听,是高安全性网络的理想选择。但由于光纤的成本相对较高,且需要较高的安装技术,因而常用于网络的主干部分。

(2)无线网。无线网是一种采用电磁波作为载体来实现数据传输的网络类型。无线网与有线网络的用途十分类似,最大的不同在于传输媒介的不同,利用无线电技术取代网线,可以和有线网络互为备份。无线网络已经成为互联网的有效补充和延伸。无线网能够将信号传输至很多有线传输介质无法到达的位置,且联网方式较为灵活,因此是一种很有发展前途的网络类型。

按作用范围的大小,无线网可分为无线个域网、无线局域网、无线城域网、无线广域网。

此外,根据网络的拓扑结构,计算机网络可以划分为总线型网络(如以太网)、环型网络(如令牌环网)、星型网络、树型网络、网状网络和混合网络。

1.3 计算机网络体系结构与协议封装

计算机网络是一个涉及计算机技术、通信技术等多个领域的复杂的系统,如此庞大而又复杂的系统要有效而且可靠地运行,网络中的各个部分就必须遵守一整套合理而严谨的结构化管理规则。计算机网络就是按照高度结构化的设计方法,采用功能分层原理来实现的,并利用协议支撑数据传输。

1.3.1 计算机网络体系结构

计算机网络的功能划分、分层划分和网络结构称为计算机网络体系结构。为了建立一个开放的、能为大多数机构和组织承认的网络互连标准,国际标准化组织(International Standard Organization,ISO)在充分考虑到当时所有的网络体系结构基础上,提出了开放系统互连参考模型(open system interconnection reference model),简称 OSI/RM 或 OSI 参考模型。OSI 参考模型定义了异种计算机连接标准的框架结构,为连接分布式应用处理的“开放”系统提供了基础和指导。所谓“开放”,是指任何两个系统只要遵守参考模型和有关标准,都能够进行互连。

OSI 参考模型采用层次化结构,将网络结构分为七层,但是在实际应用中很少有计算机网络是完全按照 OSI 参考模型建立的,而且 OSI 参考模型中的“会话层”和“表示层”也很少在实际中应用,因此,目前流行的计算机网络都只有五层协议体系,在 TCP/IP 模型中一般简化为四层,如图 1-8 所示。

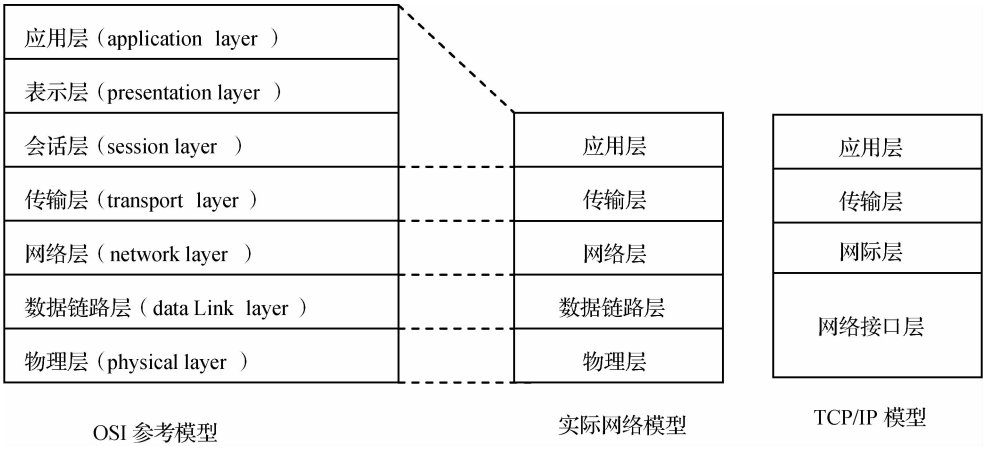


图 1-8　网络层次模型示意图

1)物理层

物理层定义信道上传输的原始比特流，如用多少伏特电压表示“1”，多少伏特电压表示“0”；一个比特持续多少微秒等，从而保证一方发出二进制“1”，另一方收到的也是“1”而不是“0”；还定义网络接插件有多少针以及各针的用途，采用什么样的接口等。

2)数据链路层

数据链路层的主要任务是加强物理层传输原始比特的功能，使之对网络层显现为一条无错线路。发送方把输入数据分装在数据帧(data frame)中，按顺序传送各帧，并处理接收方回送的确认帧(acknowledgement frame)。

3)网络层

网络层是网络核心(通信子网)的最高层，实现的功能最多。网络层协议的数据单元是分组(packet)。网络层的主要任务是对子网的运行进行控制，具体包括以下功能。

(1)确定分组从源端到目的端如何选择路由。路由既可以选用网络中固定的静态路由表，也可以在每一次会话开始时决定，还可以根据当前网络的负载状态高度灵活地为每一个分组决定路由。

(2)拥塞控制。如果在子网中同时出现过多的分组，它们将相互阻塞通路，形成瓶颈。对于此类拥塞的控制也属于网络层的范围。

(3)记账。网络层软件对每一个用户究竟发送了多少分组、多少字符或多少比特流记数，以便于生成账单。当分组跨越国界时，由于双方税率可能不同，记账则更加复杂。

(4)网络寻址。当分组跨越一个网络到达目的地时，中间网络的寻址方法可能和第一个网络完全不同，这时，网络层必须解决不同网络之间的寻址问题，以便异种网络能够互连。

4)传输层

传输层的主要功能是从上层接收数据，并且在必要时把它分成较小的单元，传递给网络层，并确保到达对方的各段信息正确无误；还要通过流量控制机制调节主机之间的通信量，使高速主机不会发生过快地向低速主机传输数据使低速主机来不及接收的情况，是真正的从源到目标“端到端”的层。也就是说，源端机上的某程序，利用报头和控制报文与目标机上的类似程序进行对话。

5)会话层

会话层在OSI参考模型中的第五层,主要解决面向用户的功能,如通信方式的选择,用户间对话的建立、拆除。会话层的主要任务是管理对话。会话层允许不同机器上的用户建立会话关系,允许信息同时双向传输。会话层提供的服务可建立应用和维持会话,并能使会话获得同步。会话层使用校验点,可使通信会话在通信失效时从校验点继续恢复通信,这种能力对于传送大的文件极为重要。

6)表示层

表示层主要用来定义所传输信息的语法和语义,如采用的数据结构和各个计算机采用的编码形式等。为了让采用不同表示法的计算机之间能进行通信,交换中使用的数据结构可以用抽象的方式来定义,并且使用标准的编码方式。表示层管理这些抽象数据结构,并且在计算机内部表示法和网络的标准表示法之间进行转换。

7)应用层

应用层是OSI参考模型中的最高层,为应用程序提供服务以保证通信,但不是进行通信的应用程序本身。应用层定义向最终用户提供的网络服务及实现这些服务的相关协议,包含大量人们普遍需要的应用协议,如DNS、HTTP、FTP等。

1.3.2 计算机网络的数据封装

数据封装是指将协议数据单元(PDU)封装在一组协议头和协议尾中的过程。在OSI参考模型的七层中,各层主要负责与其他机器上的对等层进行通信。该过程是在PDU中实现的,其中每层的PDU一般由本层的协议头、协议尾和数据封装构成。

每层可以添加协议头和协议尾到其对应的PDU中。协议头包括层到层之间的通信相关信息。协议头、协议尾和数据是3个相对的概念,这主要取决于进行信息单元分析的各个层。例如,传输头(TH)包含只有传输层可以看到的信息,而位于传输层以下的其他所有层将传输头作为各层的数据部分进行传送。在网络层,一个信息单元由层3协议头(NH)和数据构成;在数据链路层中,由网络层(层3协议头和数据)传送下去的所有信息均被视为数据。换句话说,特定OSI参考模型层中,信息单元的数据部分可能包含由上层传送下来的协议头、协议尾和数据。

图1-9给出了数据进入协议栈的封装过程。用户数据通过应用层协议封装应用层首部;封装了首部的应用数据作为整体,在传输层封装TCP首部,在网际层封装IP首部;封装后的IP数据包传输到以太网网卡后加入以太网首部,然后在线路上传输。接收方接到上述信包后依次解包,获得需要的应用数据。

1.4 计算机网络的通信基础

数据通信是计算机网络最基本的功能,数据通信用来快速传送计算机与终端或计算机与计算机之间的各种信息。

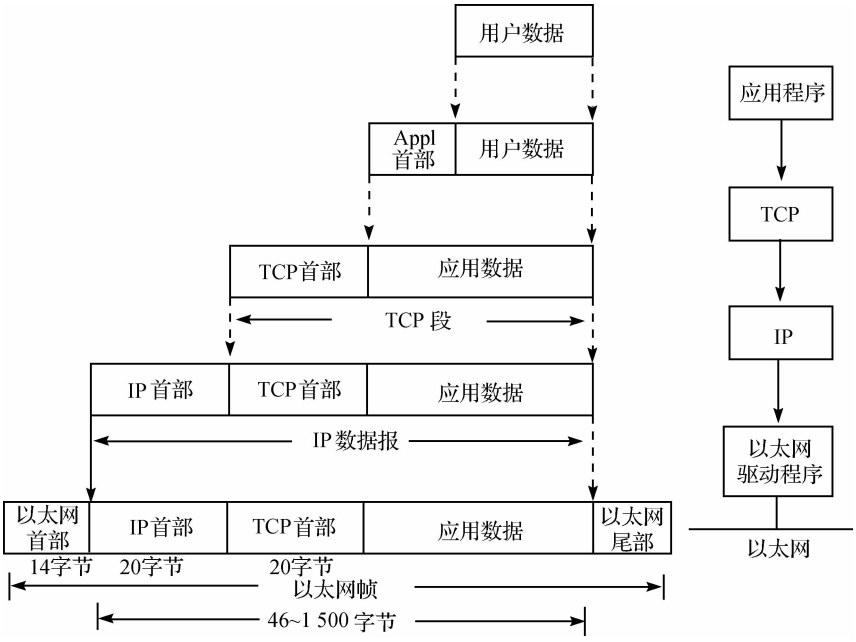


图 1-9　数据进入协议栈时的封装过程

1.4.1　信息、数据和信号

不同领域对信息有各种不同的定义，一般认为信息是人对现实世界事物存在方式或运动状态的某种认识。信息的表示形式可以是数值、文字、图形、声音、图像以及动画等，这些表示介质归根结底都是数据的一种形式。

数据是把事件的某些属性规范化后的表现形式，它能被识别，也可以被描述，如十进制数、二进制数、字符等。数据的概念包括两方面：其一，数据内容是事物特性的反映或描述；其二，数据以某种介质作为载体，即数据是存储在介质上的。

信号是数据具体的物理表现，具有确定的物理描述，如电压、磁场强度等。

信息、数据和信号这三者是紧密相关的。例如，当人们在教给小孩“苹果”这个词时，总是要向他(她)传递一些关于苹果的信息，如苹果是水果、苹果的外形和苹果的味道等。小孩便将苹果的信息和“苹果”这个词联系起来。当人们将“苹果”读出来或写出来时，就是用数据在传递苹果的信息。用数据传递信息总是依赖于一定的物理信号，如用嘴发出的声音是用声波信号传递信息，用笔写出字符是用文字来传递信息。当字符串数据“苹果”在计算机通信网络中传输时，通信线路上实际上传输的是一个二值的电压序列信号。从数据的角度来看，按照事先的约定，字符串“苹果”可以被识别，同时字符串“苹果”又可以用一个二值的电压信号序列来描述，如图 1-10 所示。

1.4.2　通信系统模型

通信的目的就是传递信息和交换信息。信息的载体可以是数字、文字、语音、图形或图

像。计算机产生的信息一般是字母、数字和符号的组合。为了传送这些信息，首先要将每一个字母、数字或符号用二进制代码表示。

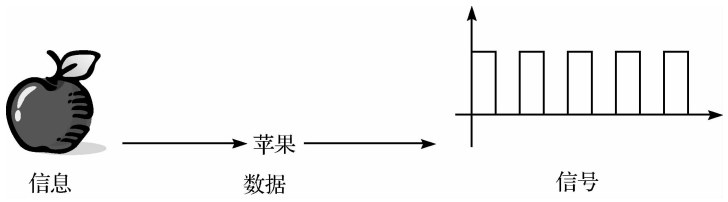


图 1-10　信息、数据和信号

数据通信是指在不同计算机之间传送表示字母、数字和符号的二进制代码 0、1 比特序列的过程。通信中产生和发送信息的一端叫信源，接收信息的一端叫信宿。信源和信宿之间要通过通信线路才能互相通信，通信线路通常称为信道。信源和信宿之间的信息交换是通过信道进行的。信道可以是有线传输介质，也可以是无线传输介质。抽象的通信系统模型如图 1-11 所示。

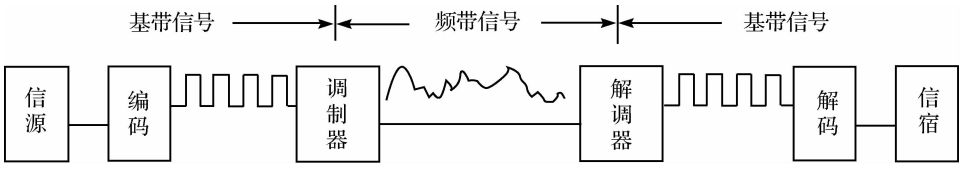


图 1-11　通信系统模型

信号传输方式分为基带传输和频带传输。计算机直接输出的数字信号称为基带信号，直接传送基带信号称为基带传输。基带传输不需要调制解调器，设备花费小，适合短距离的数据传输。由于在近距离范围内，基带信号的功率衰减不大，信道容量不会发生变化，因此，在局域网中通常使用基带传输技术。频带传输就是先将基带信号变换（调制）成便于在模拟信道中传输的、具有较高频率范围的模拟信号（频带信号），再将这种频带信号在模拟信道中传输。计算机网络的远距离通信通常采用频带传输。信号调制的目的是将数字信号变换成模拟信号，以便更好地适应信号传输通道的频率特性，提高线路的利用率。而接收部件则需要将模拟信号进行解调，变成数字信号，方便计算机处理。

1）数字信号编码

对于传输数字信号来说，最普通且最容易的方法是用两个不同的电压值来表示两个二进制值。用无电压（或负电压）表示 0，正电压表示 1。常用的数字信号编码有不归零（NRZ）编码、曼彻斯特编码（Manchester encoding）和差分曼彻斯特（differential manchester）编码。

（1）NRZ 编码。信号电平的一次反转代表 0，电平不变化表示 1，并且在表示完一个码元后，电压不需回到 0。不归零编码是效率最高的编码，其缺点是存在发送方和接收方的同步问题。NRZ 编码本身不能恢复同步信号（时钟），在进行多机通信时，同步是依靠发送和接收端使用相同频率的时钟发生器来保证的，因此 NRZ 编码更适于异步方式通信。要想使数据编码本身携带同步时钟信息，必须设法使数据与时钟一起编码发送，再由接收端借助锁相环电路恢复同步时钟，典型的编码方式是变形不归零（NRZI）、曼彻斯特编码等。

（2）曼彻斯特编码。曼彻斯特编码也叫做相位编码（PE），是一种同步时钟编码技术，被

物理层用来编码一个同步位流的时钟和数据。曼彻斯特编码被用在以太网媒介系统中。在曼彻斯特编码中，用电压跳变的相位不同来区分 1 和 0，即用正的电压跳变表示 0，用负的电压跳变表示 1，因此这种编码也称为相应编码。由于跳变都发生在每一个码元的中间，接收端可以方便地利用它作为位同步时钟，因此，这种编码也称为自同步编码。

(3)差分曼彻斯特编码。差分曼彻斯特编码是曼彻斯特编码的一种修改格式。其不同之处在于：每位的中间跳变只用于同步时钟信号；而 0 或 1 的取值判断是用位的起始处有无跳变来表示(若有跳变则为 0，若无跳变则为 1)。这种编码的特点是每一位均用不同电平的两个半位来表示，因而始终能保持直流的平衡。这种编码也是一种自同步编码。差分曼彻斯特编码主要用在环网和令牌网中。这种编码的调制速率要求是码元速率的 2 倍，对信道的带宽要求高，但有良好的抗噪性和自定时即自同步能力。

2)数字信号调制解调

调制就是用基带信号对载波波形的某些参数(频率、相位和幅度等)进行控制，使这些参数随基带信号的变化而变化，调制后的信号称为频带信号。常用的调制方式有 3 种：调幅、调频和调相。而解调正好与调制相反，就是从已调制的频带信号中恢复出原来的基带信号。由于已调的信号只占用一定的带宽，可以适应长距离的通信。

由于调制载波信号有 3 个特征：振幅(A)、频率(F)和相位(P)，相应地，把数字信号转换成模拟信号就有 3 种基本调制技术。

(1)振幅键控(ASK)：用数字调制信号控制载波的通断。如在二进制中，发 0 时不发送载波，发 1 时发送载波。有时也把代表多个符号的多电平振幅调制称为振幅键控。振幅键控实现简单，但抗干扰能力差。

(2)移频键控(FSK)：用数字调制信号的正负控制载波的频率。当数字信号的振幅为正时载波频率为 f_1，当数字信号的振幅为负时载波频率为 f_2。有时也把代表两个以上符号的多进制频率调制称为移频键控。移频键控能区分通路，但抗干扰能力不如移相键控和差分移相键控。

(3)移相键控(PSK)：用数字调制信号的正负控制载波的相位。当数字信号的振幅为正时，载波起始相位取 0；当数字信号的振幅为负时，载波起始相位取 180°。有时也把代表两个以上符号的多进制相位调制称为移相键控。移相键控抗干扰能力强，但在解调时需要有一个正确的参考相位，即需要相干解调。

3 种调制技术的波形如图 1-12 所示。也可以把 3 种调制技术混合使用，从而在波特率不变的情况下提高数据传输速率。

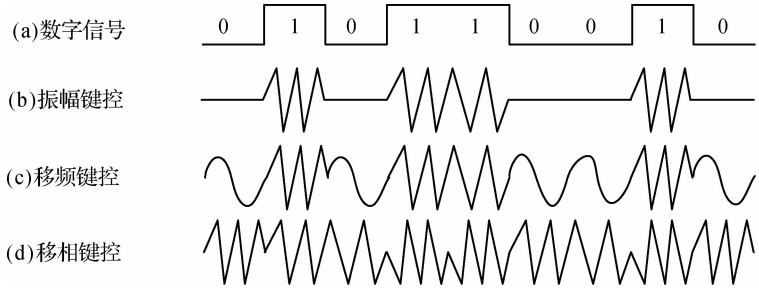


图 1-12　3 种调制技术的基本波形

1.4.3 通信系统的主要参数

数字通信系统的主要参数是信道带宽、波特率和数据传输速率。

1)信道带宽

信道带宽是指信道频率响应曲线上幅度取其频带中心处值的 $1/\sqrt{2}$倍的两个频率之间的宽度,如图 1-13 所示,带宽 $W=f_2-f_1$,其中 f_1 是信道能通过的最低频率,f_2 是信道能通过的最高频率,两者都是由信道的物理特性决定的。当组成信道的介质确定了,信道的带宽就决定了。为了降低信号传输中的失真,信道要有足够的带宽。

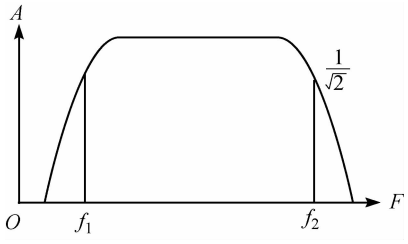


图 1-13 模拟信道的带宽示意图

2)波特率

数字信道是一种离散信道,它只能传送取离散值的数字信号。信道的带宽决定了信道中能不失真地传输脉冲序列的最高速率。通常把一个数字脉冲称为一个码元,并用码元速率表示单位时间内信号波形的变换次数,即单位时间内通过信道传输的码元个数。可知,若信号码元宽度为 T 秒,则码元速率 $B=1/T$。码元速率的单位叫波特(Baud),这是为纪念电报码的发明者法国人波特(Baudot)而命名的,所以码元速率也叫波特率。

1924 年,贝尔实验室的研究员亨利·尼奎斯特(Harry Nyquist)推导出了有限带宽无噪声信道的极限波特率:若信道带宽为 W,则最大码元速率为

$$B=2W \text{ (Baud)}$$

上式又称为尼奎斯特定理,它指定的信道容量也叫做尼奎斯特极限。尼奎斯特极限是由信道的物理特性决定的,超过尼奎斯特极限传送脉冲信号是不可能的,所以要进一步提高波特率必须改善信道带宽。

3)数据传输速率

单位时间内在信道上传送的信息量(比特数)称为数据传输速率。信号码元携带的信息量由码元取的离散值个数,即码元的种类数决定。若码元取两种离散值,则一个码元携带 1 比特(bit)信息。若码元取 4 种离散值,则一个码元携带 2 bit 信息。一个码元携带的信息量 n(bit)与码元的种类数 N 有如下关系。

$$n=\log_2 N$$

在一定的波特率下提高速率的途径是用一个码元表示更多的比特数。如果把 2 bit 编码为一个码元,则数据传输速率可成倍提高。数据传输速率和波特率之间有如下关系。

$$R=B\log_2 N=(1/T)\log_2 N=2W\log_2 N(\text{bit/s})$$

其中，R 表示数据传输速率，单位是 bit/s 或 b/s，故也称为比特率。

1.4.4　通信双方的控制方式

1)数据传送方式

根据数据传送方向的不同，数据传送方式可分为：单工方式、半双工方式、全双工方式 3 种，如图 1-14 所示。

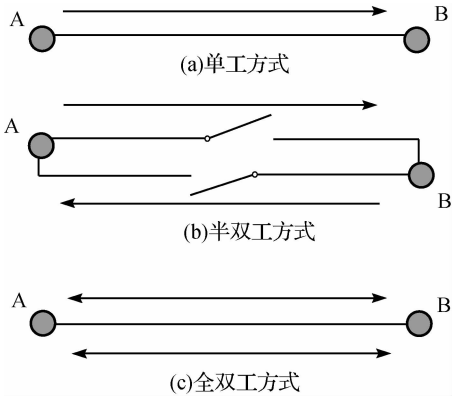


图 1-14　数据传送方式

单工方式是指信息传送只能有一个方向，即对于通信双方来说，一方只能发送信息，而另一方只能接收信息。

在半双工方式中，通信双方在任何时刻只能有一方发送信息，另一方接收信息，发送完一定信息后，双方再转换角色，继续进行信息传送。

在全双工方式中，通信双方同时发送和接收信息。

为了实现全双工方式，通常需要两套独立的通信线路，而半双工方式通常只有一套通信线路。

在计算机网络中主要使用半双工和全双工方式。

2)异步传输和同步传输

在计算机网络和通信网络中都采用串行传输，这是由于远距离传输要考虑技术实现上的要求和传输线路的费用。串行传输在具体应用中又分为异步传输和同步传输。异步传输和同步传输都要考虑收发双方的同步，即接收方能够正确地区分所收到数据的每一位。

(1)异步传输。异步传输是一次传输一个字符，字符内部的各个比特采用固定的时间模式，每个字符独立传输，字符之间间隔任意，用独特的起始位和终止位来限定每个字符，如图 1-15 所示。通常，每个字符由 1 位起始位标识，起始位的值为 0，字符由 5～8 位二进制数组成，后跟 1～2 位终止位，终止位的值为 1。异步传输方式的传输效率较低。

(2)同步传输。同步传输以一个帧(或称数据块)为传输单位，每个帧中包含多个字符。在通信过程中，每个字符间的时间间隔是相等的，而且每个字符中各相邻位代码间的时间间隔也是固定的。

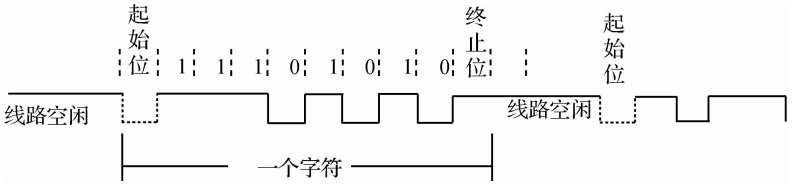


图 1-15　异步传输方式

同步传输方式又分为以下两种。

①面向比特的同步传输：以二进制位作为信息单位。现代计算机网络大多采用此类同步传输方式。最典型的是以太网中数据链路层使用的高级数据链路控制（HDLC）通信规程，其起始字节由 01111110（7EH）构成。

②面向字符的同步传输：以字符作为信息单位。字符是 EBCD 码或 ASCII 码。最典型的是 IBM 公司的二进制同步控制规程（BSC 规程）。在这种传输方式下，发送端与接收端采用交互应答的方式进行通信。为了便于接收部件识别一组信息的开始，需要在传送的信息之前插入若干同步字符 SYN，接收部件识别出同步字符后，就开始接收信息本身。

图 1-16 说明了这两种同步传输方式的帧格式。

图 1-16　两种同步传输方式的帧格式

1.4.5　多路复用技术

多路复用技术是把多个低速信道组合成一个高速信道的技术，它可以有效地提高数据链路的利用率，从而使得一条高速的主干链路同时为多条低速的接入链路提供服务，也就是使得网络干线可以同时运载大量的语音和数据传输。

复用可以实现的前提是信道的传输能力大于传输一路信号的需求，这体现在两方面：一是信道的带宽很宽，而传输一路信号所需的带宽很窄；另一方面是信道的数据传输率很高，而一路信号所需的数据传输率很低，这样在能力很强的信道上仅传输一路信号就很浪费。

常见的多路复用技术包括频分多路复用（FDM）、时分多路复用（TDM）、波分多路复用（WDM）和码分多路复用（CDMA）。

1)频分多路复用

频分多路复用的基本原理是在一条通信线路上设置多个信道，如图 1-17 所示。每路信道的信号以不同的载波频率进行调制，各路信道的载波频率互不重叠，这样一条通信线路就可以同时传输多路信号。

在信道之间要留有隔离频带，对每路信号以不同频率的载波进行调制，使其适应不同信道频段的要求。

图 1-17　频分多路复用

2)时分多路复用

时分多路复用是以信道传输时间作为分割对象，它使不同的信号在不同的时间内传送，将整个传输时间分为许多时序间隙(time slot，TS，简称时隙)。因此时分多路复用更适用于数字信号的传输。时分多路复用把传输时间分成时间片帧，每一时间片帧包含若干时隙，每个时隙对应一路信号的若干位，如图 1-18 所示。

时分多路复用又分为同步时分多路复用和统计时分多路复用。

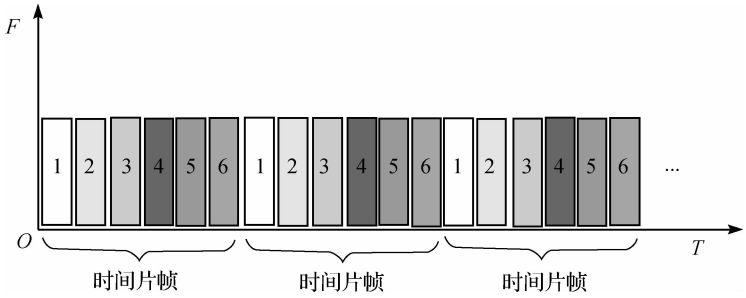


图 1-18　时分多路复用

3)波分多路复用

波分多路复用是将两种或多种不同波长的光载波信号(携带各种信息)在发送端经复用器(multiplexer，亦称合波器)汇合在一起，并耦合到光线路的同一根光纤中进行传输；在接收端，经分用器(demultiplexer，亦称分波器或去复用器)将各种波长的光载波分离，然后由光接收机作进一步处理以恢复原信号。这种在同一根光纤中同时传输两个或多个不同波长光信号的技术，称为波分多路复用。波分多路复用是指光的频分多路复用，它是在光学系统中利用衍射光栅来实现多路不同频率光波信号的合成与分解的，其原理如图 1-19 所示。

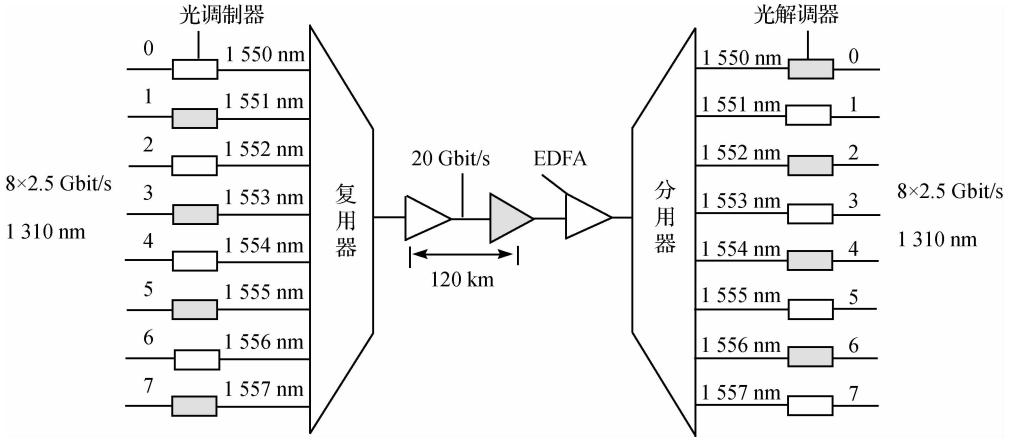


图 1-19　波分多路复用

WDM 是 FDM 的一个变种，频率和波长之间满足关系式 $f=c/\lambda$，这里 f 为频率，c 为光速，λ 为波长。

每根光纤上光信号的波长不同，两根光纤连接到一个棱柱或衍射光栅上，两束光通过棱柱或衍射光栅合成到一根共享的光纤上。传输到远方的目的地后，再用棱柱或衍射光栅将它们分解开交给接收方。

4)码分多路复用

码分多路复用也是一种共享信道的方法，每个用户可在同一时间使用同样的频带进行通信，但使用的是基于码型分割信道的方法，即每个用户分配一个地址码，各个码型互不重叠，通信各方之间不会相互干扰，且抗干扰能力强。码分多路复用技术主要用于无线通信系统，特别是移动通信系统。它不仅可以提高通信的话音质量和数据传输的可靠性以及减小干扰对通信的影响，而且增大了通信系统的容量。笔记本电脑、个人数字助理(personal data assistant，PDA)以及掌上电脑(handed personal computer，HPC)等移动性计算机的联网通信就是使用了这种技术。

1.5　计算机网络的标准化

所谓标准，即文档化的协议中包含推动某一特定产品或服务应如何设计或实施的技术规范。通过标准，不同的生产厂商可以确保产品、生产过程以及服务符合他们的要求。由于目前网络界所使用的硬件、软件种类繁多，标准尤其重要。如果没有标准，可能由于一种硬件不能与另一种兼容，或者一个软件应用程序不能与另一个通信而不能进行网络设计。例如，如果一个厂商设计的网络电缆插头为 1 in(1 in=0.0254 m)宽，而另一公司生产的槽口为0.9 in宽，则无法将电缆插入这种槽口。

主要的国际标准化组织有以下几个。

1)ANSI

美国国家标准协会(ANSI)负责制定电子工业的标准和其他行业标准。

2)EIA

电子工业联盟(EIA)是一个商业组织,其代表来自全美各电子制造公司。该组织不仅为自己的成员设定标准,还帮助制定 ANSI 标准,并建议通过立法促进计算机和电子工业发展。EIA 包括几个下属组织:电信工业协会(TIA),用户电子生产商协会(CEMA),电子部件、组装、设备与供应协会(ECA),联合电子设备工程委员会(JEDEC),固态技术协会,政府处以及电子信息组(EIG)。

3)IEEE

电气与电子工程师协会(IEEE)是一个由工程专业人士组成的国际团体,其目的在于促进电气工程和计算机科学领域的发展和教育。IEEE 主办大量研讨会或会议,出版有 IEEE 系列期刊。同时,IEEE 有自己的标准委员会,为电子和计算机工业制定自己的标准,并对其他标准制定组织(如ANSI)的工作提供帮助。IEEE 制定了十余个与计算机网络有关的标准。

(1)IEEE 802.1:通用网络概念及网桥等。

(2)IEEE 802.2:逻辑链路控制等。

(3)IEEE 802.3:CSMA/CD 访问方法及物理层规定。

(4)IEEE 802.4:token bus 访问方法及物理层规定。

(5)IEEE 802.5:token ring 访问方法及物理层规定。

(6)IEEE 802.6:城域网的访问方法及物理层规定。

(7)IEEE 802.7:宽带局域网。

(8)IEEE 802.8:光纤局域网(FDDI)。

(9)IEEE 802.9:ISDN 局域网。

(10)IEEE 802.10:网络的安全。

(11)IEEE 802.11:无线局域网。

(12)IEEE 802.15:无线个域网。

(13)IEEE 802.16:宽带无线接入。

4)ISO

国际标准化组织(ISO)是制作全世界工商业国际标准的各国国家标准机构代表的国际标准建立机构,它的总部设在瑞士的日内瓦。ISO 的目标是制定国际技术标准以促进全球信息交流和无障碍贸易。在 ISO 的大约 12 000 个标准中,仅有大约 500 个应用于计算机相关的产品和功能中。在 20 世纪 80 年代早期,ISO 即开始致力于制定一套普遍适用的规范集合,以使得全球范围的计算机平台可进行开放式通信,即 OSI 参考模型。OSI 参考模型将网络结构划分为物理层、数据链路层、网络层、传输层、会话层、表示层和应用层。每一层均有自己的一套功能集,并与紧邻的上层和下层交互作用。在顶层,应用层与用户使用的软件(如字处理程序或电子表格程序)进行交互。在 OSI 参考模型的底端是携带信号的网络电缆和连接器。

5)ITU

国际电信联盟(ITU)是管理国际电信的联合国机构,它管理无线电和电视频率、卫星和

电话的规范、网络基础设施、全球通信所使用的关税率。它为发展中国家提供技术专家和设备以提高其技术基础。ITU 过去常被称为 CCITT(国际电报电话咨询委员会),在一些手册和文档中可见对 CCITT 标准的引用。

习　题　1

1)选择题

(1)局域网的英文缩写为(　　)。

A. LAN　　B. WAN　　C. ISDN　　D. MAN

(2)计算机网络中广域网和局域网的分类是以(　　)来划分的。

A. 信息交换方式　　B. 网络使用者

C. 节点分布的地理范围　　D. 传输控制方法

(3)OSI 参考模型的最底层是(　　)。

A. 传输层　　B. 网络层　　C. 物理层　　D. 应用层

(4)OSI 参考模型描述(　　)层协议网络体系结构。

A. 四　　B. 五　　C. 六　　D. 七

(5)使用网络时,通信网络之间传输的介质,不可用(　　)。

A. 双绞线　　B. 无线电波　　C. 光缆　　D. 化纤

(6)(　　)是实现数字信号和模拟信号转换的设备。

A. 网卡　　B. 调制解调器　　C. 网络线　　D. 以上都不是

(7)在计算机网络中,为了使计算机或终端之间能够正确传送信息,必须按照(　　)来相互通信。

A. 信息交换方式　　B. 网卡　　C. 传输装置　　D. 网络协议

(8)目前,通过 Internet 进行通信的计算机间一般遵守(　　)协议。

A. IPX/SPX　　B. OSI　　C. SNMP　　D. TCP/IP

(9)数据通信中的信道传输速率的单位是 bit/s,被称为(　　),而单位时间内通过信道传输码元的个数称为(　　)。

A. 数率、比特率　　B. 频率、波特率

C. 比特率、波特率　　D. 波特率、比特率

(10)在数据通信的过程中,将模拟信号还原成数字信号的过程称为(　　)。

A. 调制　　B. 解调　　C. 流量控制　　D. 差错控制

(11)通过改变载波信号的相位值来表示数字信号 1 和 0 的方法叫做(　　)。

A. ASK　　B. ATM　　C. FSK　　D. PSK

(12)计算机网络的远程通信通常采用的是(　　)。

A. 基带传输　　B. 基带模拟传输　　C. 频带传输　　D. 频带数字传输

(13)最早出现的计算机互联网络是(　　)。

A. ARPANET　　B. Ethernet　　C. Internet　　D. Bitnet

(14)下列设备中,不属于通信子网的是(　　)。

A. 主机　　B. 交换机　　C. 路由器　　D. 中继器

(15)以太网的拓扑结构是(　　)。

A. 星型　　B. 环型　　C. 树型　　D. 总线型

2)填空题

(1)计算机网络的拓扑结构有_____、_____、_____、_____和_____。

(2)计算机网络采用的调制技术主要有_____、_____和_____ 3 种。

(3)根据节点分布的地理范围,计算机网络分为_____、_____和_____。

(4)OSI 参考模型将计算机网络分为_____、_____、_____、_____、_____、_____和_____七层。

(5)网络传输介质主要分为_____和_____两大类。

(6)有线传输介质分为_____、_____、_____和_____。

(7)按网络共享服务方式划分,计算机网络可分为_____、_____和_____。

(8)从计算机网络组成的角度看,计算机网络从逻辑功能上可分为_____子网和_____子网。

(9)调制就是用基带信号对载波波形的某些参数(频率、相位及幅度等)进行控制,使这些参数随基带信号的变化而变化,调制后的信号称为_____。

(10)数字数据的基本调制技术包括振幅键控、_____和_____。

3)简答题

(1)简述信号、数据和信息之间的区别和联系。

(2)计算机网络主要有哪些类型?

(3)简述计算机网络的拓扑结构。

(4)计算机网络的发展可以划分为哪几个阶段?每个阶段都有什么特点?

(5)局域网、城域网和广域网的主要特征是什么?简述 LAN 和 WAN 的区别与联系。

(6)简述 OSI 参考模型中各层的主要功能?

(7)数据通信系统由哪几部分组成?说明各部分的主要功能。

(8)数据传输方式有哪几种?简述每种方式的工作原理。

(9)什么是计算机网络体系结构?简述数据进入协议栈的封装过程。

(10)什么是信道的多路复用技术?简述 4 种信道复用技术的原理和特点。

4)计算题

(1)在数字传输系统中,码元速率为 600 Baud,数据传输速率为 1 200 bit/s,则信号取几种不同的状态?若要使得码元速率与数据传输速率相等,则信号取几种状态?

(2)已知某信道每秒最低可以传输 100 个 ASCII 字符,一个 ASCII 字符需要传输 7 位 ASCII 码、2 位校验位和 1 位终止位,共计 10 位。试计算该信道的最大比特率。

第 2 章 互联网服务与应用

计算机技术和通信技术的飞速发展，使得互联网服务无处不在，网络搜索、即时通信、网络社区和电子商务等 Internet 应用对人们的日常生活和学习产生了深远的影响。不同的网络应用和服务采用的模型和协议有所不同，本章将结合一些典型的互联网应用，介绍其各自所使用的协议及服务构建。

2.1 WWW 服务

WWW(world wide web，3W)又称万维网，是目前应用最广泛的一种互联网。通过 WWW 服务，人们可以很容易地访问 Internet 上相互链接在一起的所有可用信息和多媒体资源。

2.1.1 WWW 服务模型

在 Internet 网络应用中，客户机/服务器模型(client/server，C/S)是目前最常用的、不同计算机之间的通信模型。除此之外，还有两种典型的 WWW 服务模型，即浏览器/服务器(browser/server，B/S)模型和 P2P(peer-to-peer)模型，下面分别介绍这 3 种模型。

1)客户机/服务器模型

在互联网应用中，WWW、FTP、Telnet 和 E-mail 等许多典型应用都采用 C/S 模型，其基本工作原理如图 2-1 所示。在该模型中，服务器程序通常在一个众所周知(或者通信双方约定)的端口上监听客户程序发出的请求。服务进程一直保持休眠状态，直到有客户程序提出连接请求，服务器程序作出应答，并为客户程序提供相应的服务。

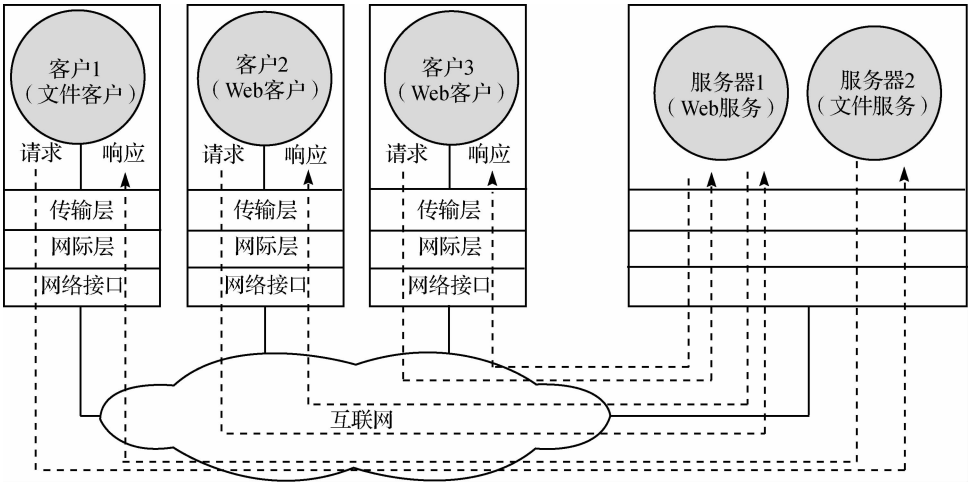


图 2-1　客户机/服务器模型

客户机和服务器分别是指参与一次通信的两个应用实体，客户机方主动发起通信请求，服务器方被动地等待通信的建立。由于服务器软件要支持多个客户机的同时访问，所以它必须具备并发性，服务器软件为每个新到的客户创建一个进程来处理和这台客户机的通信。服务器软件一般分为两部分：一部分用于接受请求并创建新的进程，另一部分用于处理实际的通信过程。

C/S 模型的应用能在一定程度上提高网络运行效率，减少客户机与服务器之间的数据传输量。同时，一个客户程序可与多个服务程序链接，用户能够根据需要访问多台主机。C/S模型最重要的特点是非对等的相互作用，这种模式适应于网络资源、运算能力和信息分布不均等现象，成为目前网络应用的主要模式之一。但这种模型对服务器的性能和可靠性要求较高，一旦服务器出现问题或因客户连接较多导致性能下降，整个网络将会瘫痪。

2）浏览器/服务器模型

B/S 是对客户机/服务器模型的一种改进，图 2-2 给出了一个最常见的 B/S 模型。在 B/S 模型中，客户主机上的用户访问接口是通过 WWW 浏览器实现的。当客户机有请求时，向 Web 服务器提出请求服务，Web 服务器通过某种机制请求数据库服务器的数据服务，然后再由 Web 服务器把查询结果返回浏览器，并显示出来。

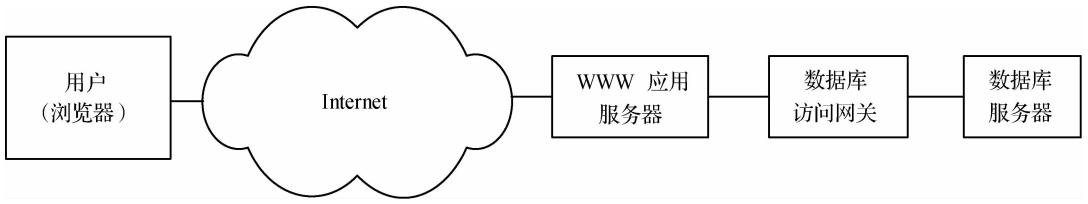


图 2-2　B/S 模型

B/S 模型也被称为“瘦”客户机/“胖”服务器模型。这是因为整个系统的复杂性都集中在服务器端，客户端不需要额外的开发，只需要采用标准的浏览器即可。采用 B/S 模型的系

统选择服务器非常灵活，完全不受客户端浏览器的制约，但是 B/S 模型同样也存在服务器负荷较重的问题。

3)P2P 模型

P2P 又称为“点对点”或“对等”技术。P2P 模型打破了传统的 C/S 模式，在网络中，每个节点的地位都是对等的，每个节点既充当服务器，为其他节点提供服务，同时也享受其他节点提供的服务。

2.1.2 HTTP

超文本传送协议(hypertext transport protocol，HTTP)是客户端和服务器端请求和应答的一个标准。通常由 HTTP 客户端发起一个请求，建立一个到服务器指定端口(默认是 80 端口)的连接。HTTP 服务器则在指定端口监听客户端发送过来的请求，一旦收到请求，服务器向客户端发回一个响应的消息。消息的消息体可能是请求的文件、错误消息或者其他信息。客户端接收服务器所返回的信息，通过浏览器显示在用户的显示屏上，然后客户机与服务器断开连接。

1)HTTP 简介

HTTP 的发展是万维网协会(World Wide Web Consortium，W3C)和 Internet 工程任务组(Internet Engineering Task Force，IETF)合作的结果，他们发布了一系列的标准，其中 RFC 2616 定义了 HTTP 中一个现今被广泛使用的版本，即 HTTP 1.1。HTTP 1.1 能很好地配合代理服务器的工作，支持以管道方式同时发送多个请求，能有效降低线路负载，提高传输速度，并且向下兼容较早的版本 HTTP 1.0。

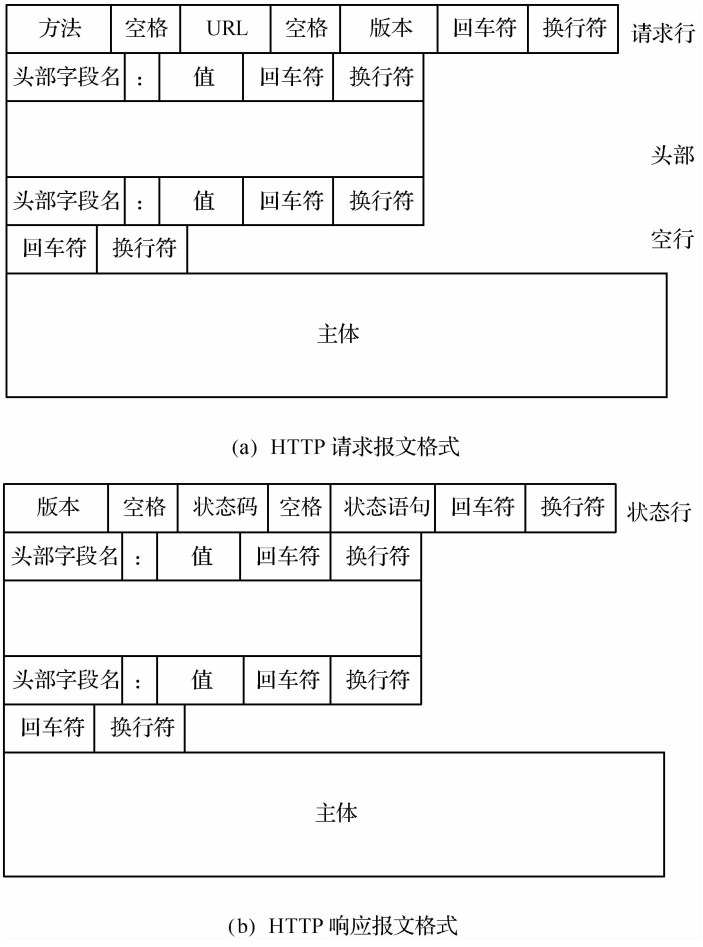
HTTP 1.0 使用非持久连接，客户端必须为每一个待请求的对象建立并维护一个新的连接。因为同一个页面可能存在多个对象，所以非持久连接可能使一个页面的下载变得很缓慢。HTTP 1.1 引入了持久连接，允许在同一个连接中存在多次数据请求和响应，即在持久连接情况下，服务器在发送完响应后并不关闭 TCP 连接，而客户端可以通过这个连接继续请求其他对象，这样有助于减轻网络传输的负担。

2)HTTP 报文格式

HTTP 消息分为客户端发往 Web 服务器的请求消息和 Web 服务器发往客户端的响应消息两类，HTTP 请求报文和响应报文的具体格式如图 2-3 所示。

(1)HTTP 请求报文格式。一个 HTTP 请求报文由请求行、请求头部、空行和请求数据 4 个部分组成。

①请求行由请求方法、统一资源定位器(URL)和 HTTP 版本号 3 个字段组成，它们用空格分隔。表 2-1 给出了 HTTP(HTTP 1.1)的常用方法。



(a) HTTP 请求报文格式

(b) HTTP 响应报文格式

图 2-3　HTTP 报文格式

表 2-1　HTTP 常用方法

方　法	描　述
GET	向 Web 服务器请求一个文件
POST	向 Web 服务器发送数据让 Web 服务器进行处理
PUT	向 Web 服务器发送数据并存储在 Web 服务器内部
HEAD	检查一个对象是否存在
DELETE	从 Web 服务器上删除一个文件
CONNECT	对通道提供支持
TRACE	跟踪到服务器的路径
OPTIONS	查询 Web 服务器的性能

URL:用于指明要下载的 Web 对象的位置。

HTTP 版本号:目前的版本号为 1.1,但 1.0 版本仍在使用。

②请求头部。请求头部由关键字/值对组成,每行一对,关键字和值用英文冒号":"分隔。请求头部通知服务器有关客户端请求的信息,典型的请求头部字段类型有以下几种。

Accept:指定客户端能够接收的内容类型,内容类型中的先后次序表示客户端接收的先后次序,客户端能够接收的类型有 gif、bitmap、jpeg 等。

Pragma:用来包含实现特定功能的指令,最常用的是 Pragma:no-cache。在 HTTP 1.1 中,用于指示请求或响应消息不能缓存。

User-Agent:包含 HTTP 客户端运行的浏览器类型。

Host:指定请求资源的 Internet 主机和端口号,必须表示请求 URL 的原始服务器或网关的位置。HTTP 1.1 请求必须包含主机头域,否则系统会以 400 状态码返回。

Connection:指定的连接类型为 Keep-Alive。

③空行。最后一个请求头部之后是一个空行,发送回车符和换行符,通知服务器以下不再有请求头部。

④请求数据。请求数据不在 GET 方法中使用,而是在 POST 方法中使用。POST 方法适用于需要客户填写表单的场合。与请求数据相关的最常使用的请求头是 Content-Type 和 Content-Length。

(2)HTTP 响应报文格式。HTTP 响应由状态行、消息报头和响应正文 3 个部分组成。

①状态行。HTTP 响应报文的状态行由版本号、响应状态代码、状态代码的文本描述等字段组成。

版本号表示服务器 HTTP 的版本;状态代码是一个 3 位数的结果代码,反映服务器端对客户端的请求。

其中第一位数字定义响应的类别,含义如下。

1xx:信息响应类,表示接收到请求并且继续处理。

2xx:处理成功响应类,表示动作被成功接收、理解和接受。

3xx:重定向响应类,为了完成指定的动作,必须接受进一步处理。

4xx:客户端错误,客户请求包含语法错误或者不能正确执行。

5xx:服务器端错误,服务器不能正确执行一个正确的请求。

常见的状态代码文本描述如下。

200 OK:客户端请求成功。

400 Bad Request:客户端请求有语法错误,不能被服务器所理解。

401 Unauthorized:请求未经授权,这个状态代码必须和 WWW-Authenticate 报头域一起使用。

403 Forbidden:服务器收到请求,但拒绝提供服务。

404 Not Found:请求资源不存在,如输入了错误的 URL。

500 Internal Server Error:服务器发生不可预期的错误。

②消息报头。其中包含如下头部类型。

Date:表示消息发送的时间,时间的描述格式由 rfc822 定义。

Server:表示服务器的类型。

Content-Type:用于向接收方指示对象的类型。

Last-modified:指明服务器对象的最后修订时间。

Content-length:用来指明正文中返回对象的长度。

3)Internet 中的域名服务

HTTP 中需要指明要下载的 Web 对象的位置,用 URL 表示。在网络中为了能够正确地定位到目的主机 Web 对象的位置,需要在发送的 IP 分组中携带目的 IP 地址。例如,要访问西安交通大学主页,需要在地址栏中输入 http://202.117.1.13。但这种 4 字节的 IP 地址很难记忆,为此 Internet 提供了域名系统(domain name system,DNS)。

域名服务可以有效地将 IP 地址映射到一组用“.”分隔的字符串,称为域名(domain name,DN),如 202.117.1.13 对应的域名是 www.xjtu.edu.cn。Internet 中的域名为树状层次结构,如图 2-4 所示。

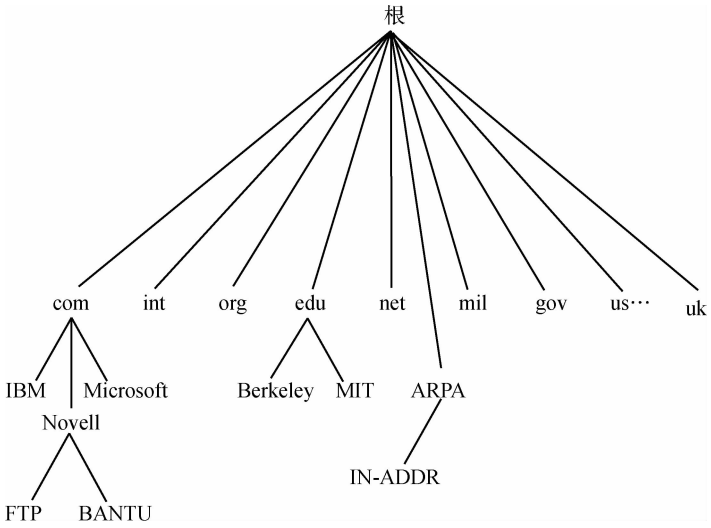


图 2-4　域名

最高级的节点称为“根”(root),根以下是顶层子域,再以下是第二层、第三层,以此类推,每个域对它下面的子域和主机进行管理。Internet 的顶级域名分为组织结构和地理结构两种。组织结构有 com(商业组织)、edu(大学等教育机构)、net(网络组织)、org(非商业组织)、gov(政府机构)、mil(军事单位)、int(国际组织)等;地理结构,除美国的顶层子域外,一般用国家名的两个字母缩写表示。

DNS 查询有递归和迭代两种方式,一般主机向本地域名服务器的查询采用递归查询,即当客户机向本地域名服务器发出请求后,若本地域名服务器不能解析,则会向它的上级域名服务器发出查询请求,以此类推,最后得到结果后转交给客户机。而本地域名服务器向根域名服务器的查询通常采用迭代查询,即当根域名服务器收到本地域名服务器的迭代查询请求报文时,如果本地域名服务器中存在映射,会直接给出所要查询的 IP 地址;否则,它仅告诉本地域名服务器下一级需要查找的 DNS 服务器,然后让本地域名服务器进行后续的查询。

2.1.3 HTML

WWW 服务的基础是将 Internet 上丰富的资源以超文本(hypertext)的形式组织起来。1963 年,Ted Nelson 提出了超文本的概念。超文本的基本特征是在文本信息之外还能提供超链接,即从一个网页指向另一个目标的连接关系,这个目标可以是另一个网页,也可以是图片、电子邮件地址或文件,甚至是一个应用程序。当浏览者单击已经链接的文字或图片后,链接目标将显示在浏览器上,并根据目标的类型来打开或运行。

超文本标记语言(hypertext markup language,HTML)就是通过各种各样的"标记"来描述 Web 对象的外观、格式、多媒体信息属性位置和超链接目标等内容,将各种超文本链接在一起的语言。HTML 是目前网络上应用最为广泛的语言,也是构成网页文档的主要语言。一个 HTML 文档是由一系列的元素(element)和标签(tag,或称标记)组成,用于组织文件的内容和指导文件的输出格式。

一个元素可以有多个属性,HTML 用标签来规定元素的属性和它在文件中的位置。浏览器只要读到 HTML 的标签,就会将其解释成网页或网页的某个组成部分。HTML 标签从使用内容上通常可分为两种:一种用来识别网页上的组件或描述组件的样式,如网页的标题<title>、网页的主体<body>、标题一<H1>、标题二<H2>、粗体<B>、斜体<I>、段落<P>等;另一种用来指向其他资源,如<img>用来插入图片,<applet>用来插入 JavaApplet,<a>用来识别网页内的位置或超链接等。

HTML 提供了数十种标签,可以构成丰富的网页内容和形式。通常标签由一对起始标签和结束标签组成,结束标签和起始标签的区别是结束标签要在"<"字符的后面加上一个斜杠。例如:

```
<html>标记网页的开始
    <head>标记头部的开始:头部元素描述,如文档标题等
    </head>标记头部的结束
    <body>标记正文开始
    页面实体部分
    </body>标记正文结束
</html>标记该网页的结束
```

2.1.4 Web 服务器的构建

互联网信息服务器(Internet information server,IIS)是 Microsoft 公司的一种集成了多种 Internet 服务(WWW 服务、FTP 服务等)的服务器软件,是当今流行的 Web 服务器之一。利用 IIS,可以很容易地构造 Web 站点。除此之外,Tomcat、Weblogic、Websphere 和 Jboss 等也是常用的 Web 服务器软件。下面以 Apache 的 Tomcat 6.0 服务器为例,简单介绍 Web 服务器的构建。

(1)双击 Web 服务器的安装文件开始安装,进入安装向导欢迎界面,如图 2-5 所示。

图 2-5　安装向导欢迎界面

(2)单击 Next 按钮,选择安装的组件,如图 2-6 所示。

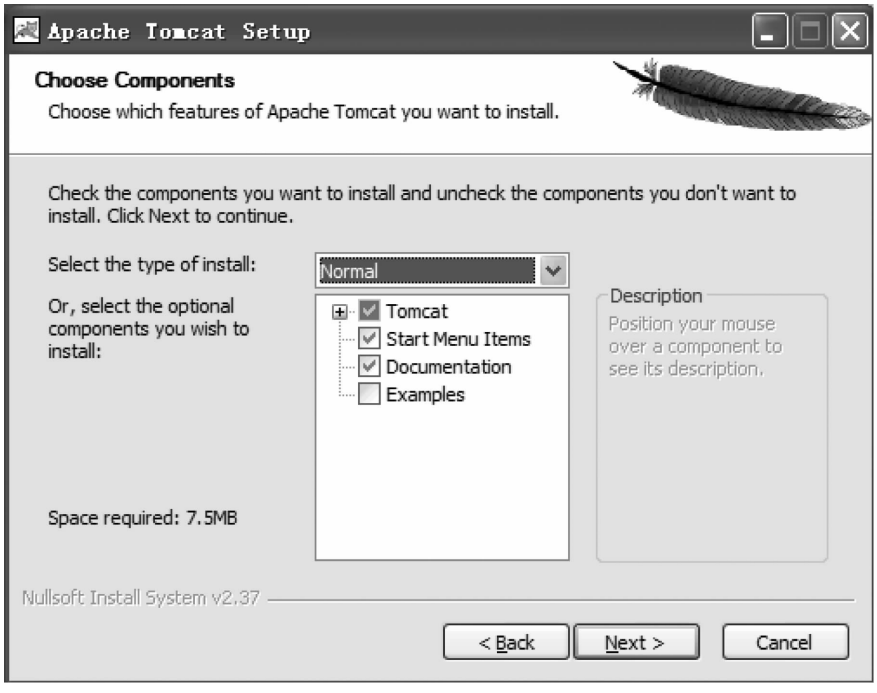


图 2-6　选择安装组件

(3)单击 Next 按钮,选择安装目录,如图 2-7 所示,单击 Next 按钮。

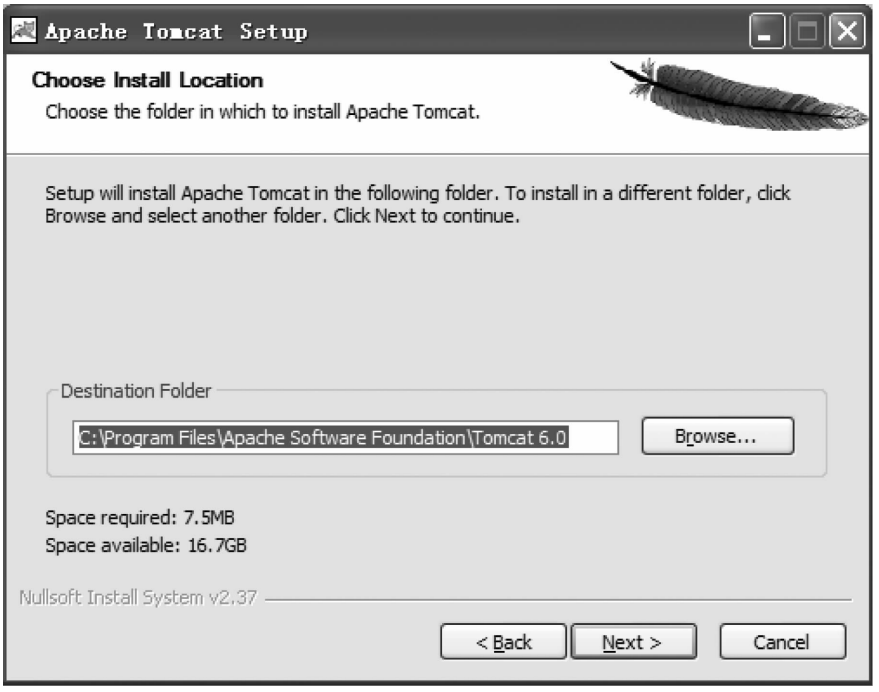


图 2-7　选择安装目录

(4)填写服务器端口,常用端口为 8080,此处选择 8086。填写服务器管理员的用户名和密码,如图 2-8 所示,单击 Next 按钮。

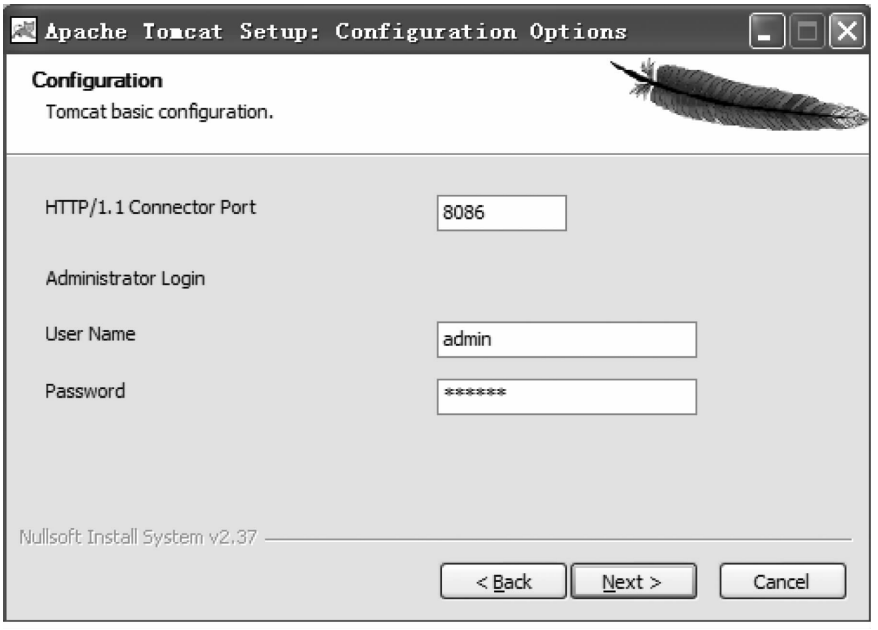


图 2-8　填写服务器端口及用户名和密码

(5)选择对应的 jre 安装地址。如果服务器上尚未安装过 jre,需要先安装 jre 再配置,如图 2-9 所示。

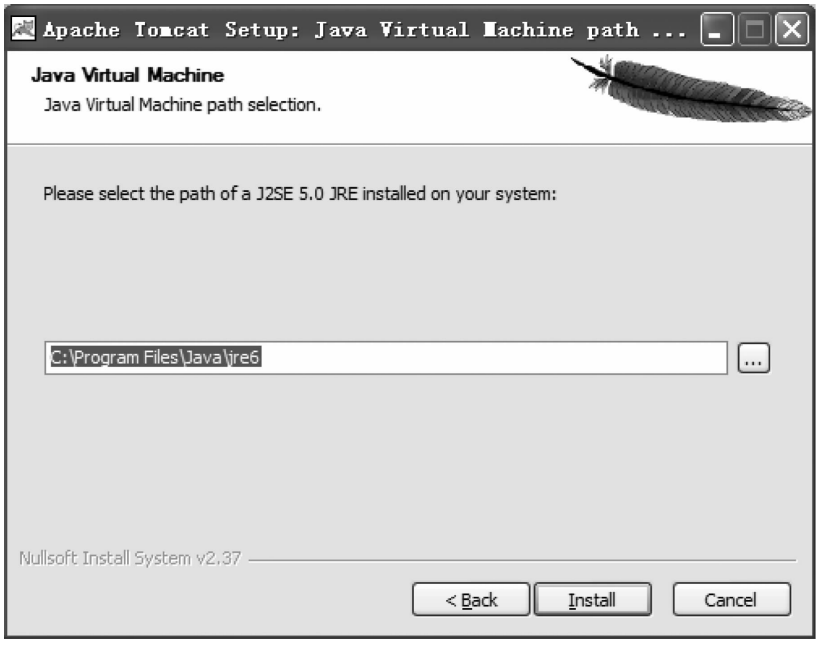
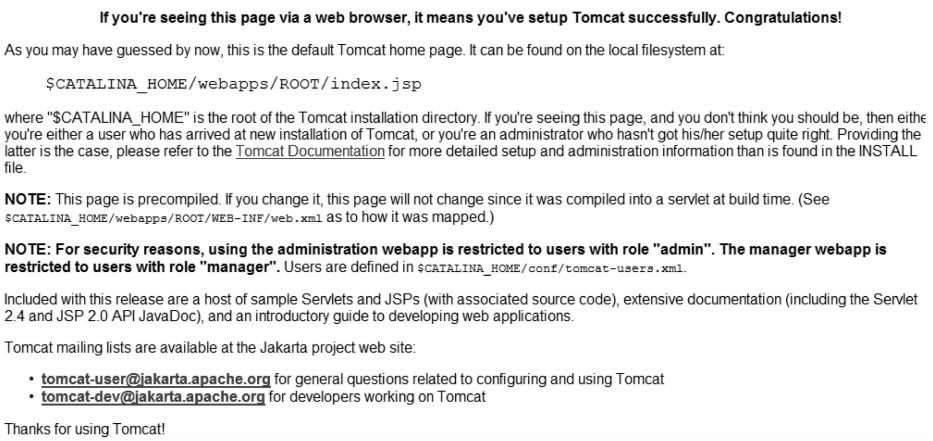


图 2-9　jre 的安装与配置

(6)单击 Install 按钮开始安装,安装完成后进入服务器启动界面,如图 2-10 所示,单击 Start 按钮启动 Tomcat 6。

图 2-10　服务器启动界面

(7)开启服务器后在浏览器中访问 http://localhost:8086,出现如图 2-11 所示的信息,至此安装配置成功。


If you're seeing this page via a web browser, it means you've setup Tomcat successfully. Congratulations!

As you may have guessed by now, this is the default Tomcat home page. It can be found on the local filesystem at:

$CATALINA_HOME/webapps/ROOT/index.jsp

where "$CATALINA_HOME" is the root of the Tomcat installation directory. If you're seeing this page, and you don't think you should be, then either you're either a user who has arrived at new installation of Tomcat, or you're an administrator who hasn't got his/her setup quite right. Providing the latter is the case, please refer to the Tomcat Documentation for more detailed setup and administration information than is found in the INSTALL file.

NOTE: This page is precompiled. If you change it, this page will not change since it was compiled into a servlet at build time. (See $CATALINA_HOME/webapps/ROOT/WEB-INF/web.xml as to how it was mapped.)

NOTE: For security reasons, using the administration webapp is restricted to users with role "admin". The manager webapp is restricted to users with role "manager". Users are defined in $CATALINA_HOME/conf/tomcat-users.xml.

Included with this release are a host of sample Servlets and JSPs (with associated source code), extensive documentation (including the Servlet 2.4 and JSP 2.0 API JavaDoc), and an introductory guide to developing web applications.

Tomcat mailing lists are available at the Jakarta project web site:

- tomcat-user@jakarta.apache.org for general questions related to configuring and using Tomcat
- tomcat-dev@jakarta.apache.org for developers working on Tomcat

Thanks for using Tomcat!

图 2-11　安装配置成功界面

2.2　文件传输服务

文件传送协议(file transfer protocol,FTP)是用于完成从 Internet 上一台主机到另一台主机文件传输的协议。

2.2.1　文件传输模型

同大多数 Internet 服务一样,FTP 也采用客户机/服务器模型,如图 2-12 所示。通常 FTP 应用程序被配置在某个服务器上,一旦客户发出文件传输请求,FTP 服务器会响应该请求,从而为客户端提供文件上传或下载的服务。

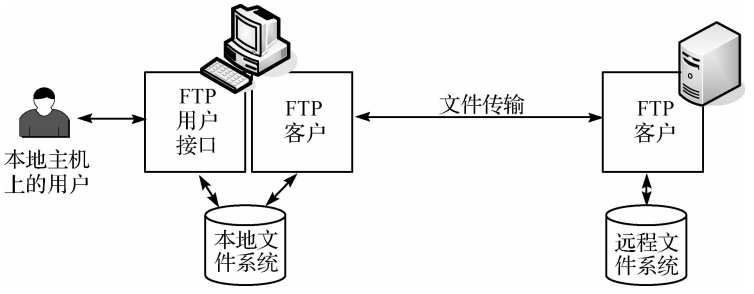


图 2-12　FTP 传输模型

一般情况下,用户通过某种 FTP 的客户端程序(如 CuteFTP、LeapFTP 等程序)完成文件传输过程。FTP 客户端程序为用户提供一个调用 FTP 的接口,用户可以通过这个接口方

便地调用 FTP 完成文件的上传、下载以及创建目录等操作。

2.2.2　FTP

1)FTP 连接

在 FTP 中，存在两种 TCP 连接：控制连接和数据连接，如图 2-13 所示。数据连接以及使用该连接的传输进程可以在需要时动态创建，但控制连接要在一次会话过程中保持。一旦控制连接撤销了，会话终止，也就不可能再有数据连接了。

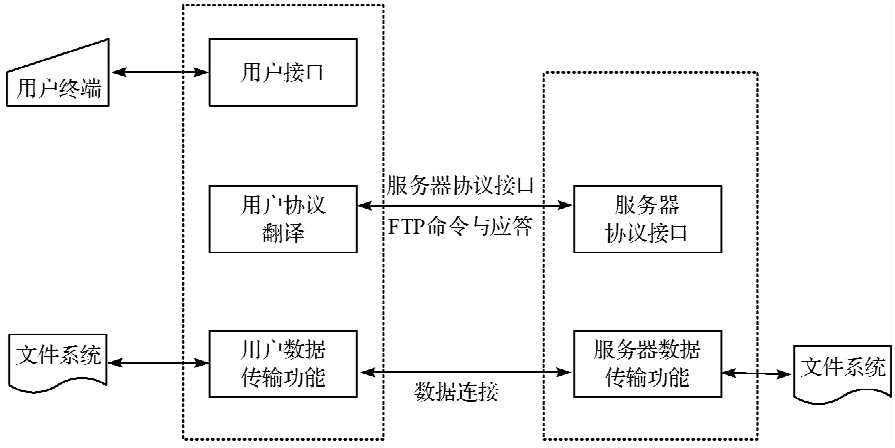


图 2-13　FTP 中的控制连接和数据连接

一般情况下，FTP 服务器以被动方式打开 FTP 端口（保留端口为 21），等待用户连接。一旦用户提出文件传输请求，在 21 端口建立用户和服务器的一条控制连接，然后经该控制连接把用户名和口令发送给服务器。用户通过服务器的验证后，由服务器发起建立一个从服务器端口 20 到用户指定端口之间的数据连接，进行数据传输。

2)FTP 命令

常用的 FTP 命令如表 2-2 所示。

表 2-2　常用的 FTP 命令

命　令	说　明
LIST filelist	以列表形式显示文件和目录
PASS password	输入用户口令
PORT n1，n2，n3，n4，n5，n6	客户端 IP 地址和端口
QUIT	从服务器注销
RETR filename	下载指定文件
STOR filename	上传指定文件
TYPE type	说明文件类型：A 表示 ASCII 码，I 表示图像
USER username	输入用户名

2.2.3 FTP 服务器的构建

构建 FTP 服务器，首先需要在一台联网的计算机上安装 FTP 服务器端软件。这类软件很多，可以使用微软的 IIS，也可以使用专业软件，如 Serv-U、WinFtp Server 或 File Zilla Server。不同的软件提供的功能不同，适应的需求和操作系统也不同。下面以在 Windows 7 中配置 FTP 服务器为例，简要说明 FTP 服务器的构建。

(1)安装 FTP 服务器。在控制面板中，单击“程序/打开或关闭 Windows 功能”选项，打开“Windows 功能”窗口，在“Internet 信息服务”树型列表中选中“FTP 服务器”复选框，单击“确定”按钮，安装 FTP 服务器，如图 2-14 所示。

图 2-14　安装 FTP 服务器

(2)建立 FTP 服务器。打开控制面板，选择“所有控制面板项”，单击“管理工具”，在打开的窗口中双击“Internet 信息服务(IIS)管理器”，打开“Internet 信息服务(IIS)管理器”窗口。右击“网站”选项，在弹出的快捷菜单中选择“添加 FTP 站点”命令，打开“添加 FTP 站点”对话框，设置 FTP 站点名称和内容目录物理路径，如图 2-15 所示。单击“下一步”按钮，打开“绑定和 SSL 设置”对话框，为 FTP 站点分配 IP 地址和端口号，如图 2-16 所示。单击“下一步”按钮，打开“身份验证和授权信息”对话框，确定 FTP 的身份验证方式及授权信息，如图 2-17 所示。然后单击“完成”按钮完成 FTP 服务器的建立过程。

图 2-15　设置 FTP 站点名称和内容目录物理路径

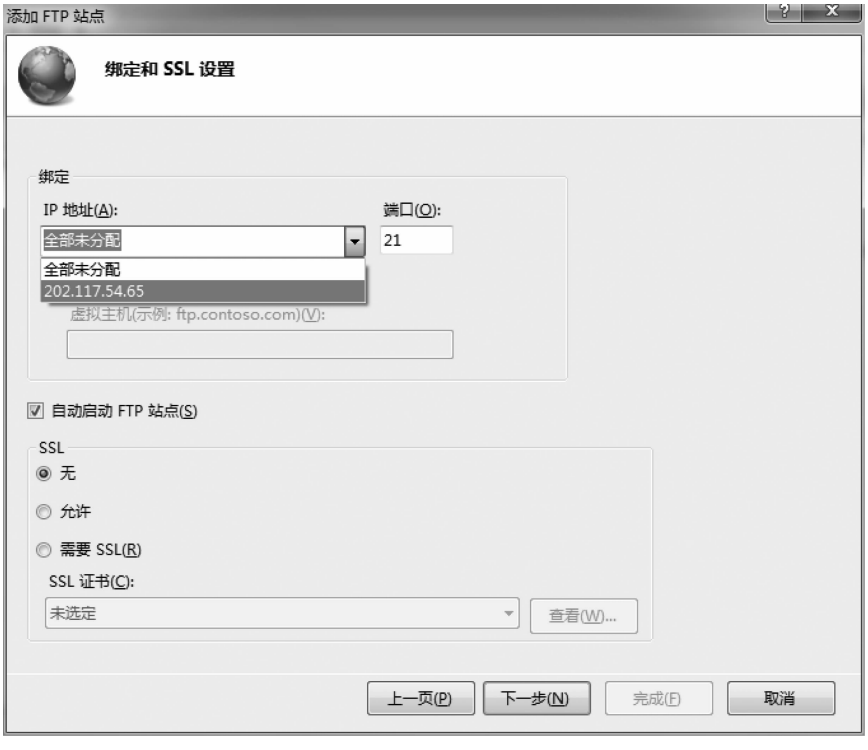


图 2-16　设置 FTP 的 IP 地址和端口

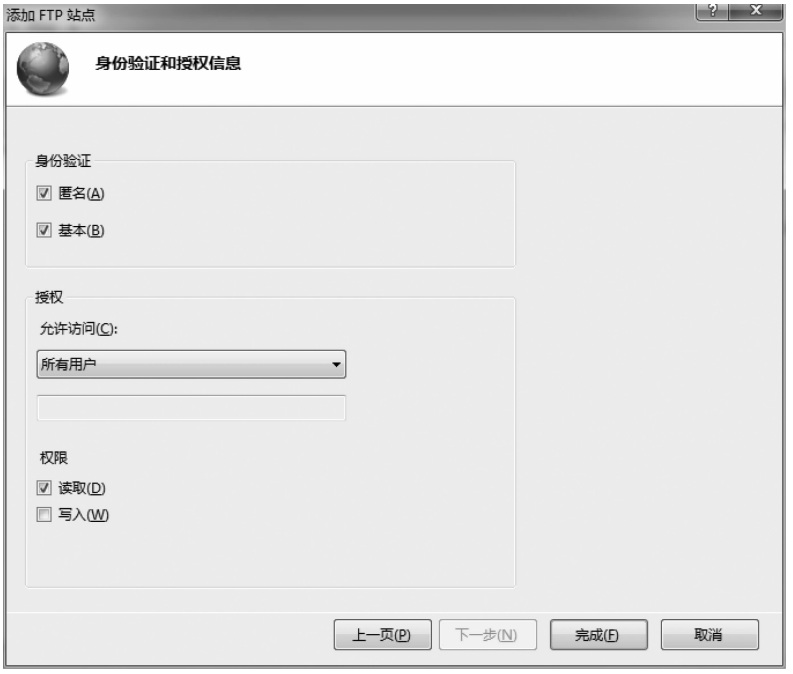


图 2-17　确定身份验证方式和授权信息

(3)使用 FTP 服务。为测试 FTP 服务器是否已经在正常工作,可以在另一个已经联网的 PC 上进行测试。在该 PC 上打开浏览器,输入“ftp://202.117.54.65/”进行登录。若能正确显示 FTP 服务器对应物理路径上的文件,则说明 FTP 服务器安装成功,如图 2-18 所示。

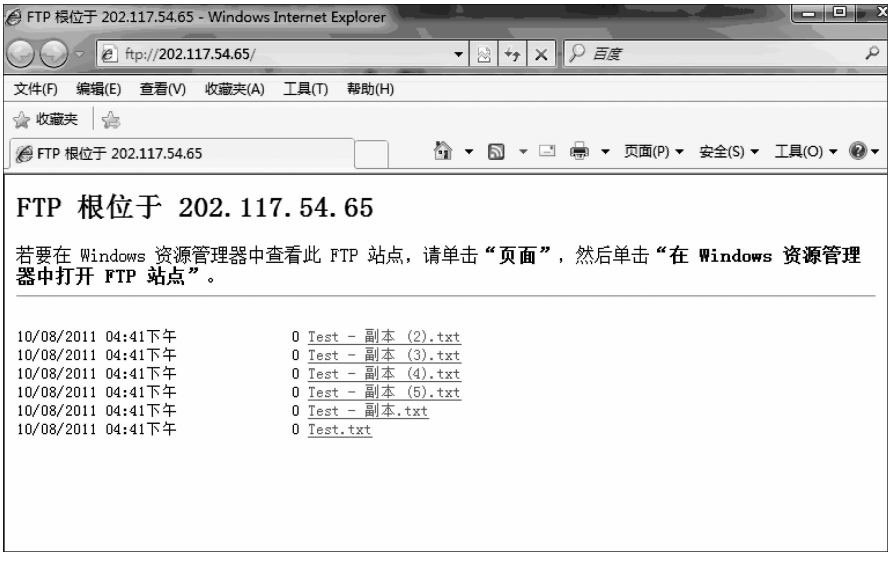


图 2-18　从浏览器登录 FTP

2.3　电子邮件服务

电子邮件(E-mail)是目前最常见、应用最广泛的互联网服务之一，是指用电子手段在互联网上传送文字、图片、音乐和应用程序等信息的通信方法。电子邮件具有传输速度快、内容形式多样、使用方便、费用低和安全性好等特点。

2.3.1　电子邮件服务模型

与 FTP 服务类似，电子邮件服务也是基于客户机/服务器模式的，如图 2-19 所示。电子邮件发送方和接收方作为客户端，一般都通过邮件代理(如 Hotmail、Foxmail，或称用户代理)来进行邮件的编辑、发送和接收。在图 2-19 中，与发送者用户代理相连的“邮件服务器 1”称为邮件发送服务器，与接收者用户代理相连的“邮件服务器 2”称为邮件接收服务器。邮件发送方和接收方分别通过各自的用户代理与邮件服务器相连，遵循一定的邮件接收和发送协议进行邮件的传输。

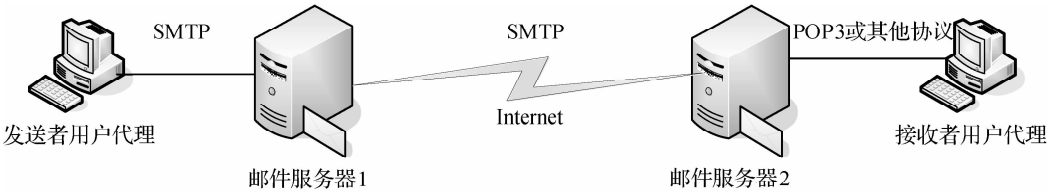


图 2-19　邮件传输模型

发送者将邮件发送到接收者的邮箱，一般需经历以下 3 个步骤。

(1)发送者用户代理通过邮件接收协议将邮件发送到自己的本地邮件服务器。

(2)当发送缓存队列达到一定的长度，或者等待了一定的时间后，发送者的邮件服务器将邮件发送到接收者的邮件服务器。

(3)接收者用户代理通过邮件接收协议访问自己邮件服务器上的邮箱。

2.3.2　邮件报文格式和 MIME

1)RFC822

RFC822(电子邮件的标准格式)定义了 Internet 文本邮件的标准消息格式，它规定邮件由首部(header)和邮件主体构成。邮件的主体部分用户可以自由撰写。邮件首部与传统邮件的信封类似，包含邮件地址、邮件主题和邮件发送人等信息。其中邮件地址的格式为“user@域名”，user 代表用户信箱的账号，“@”是分隔符，域名是指用户信箱的邮件接收服务器域名，用以标识其所在的位置。

下面列举 RFC822 中常用的一些信息字段。

(1)To:邮件的收信人地址。

(2)From:邮件的发信人地址。

(3)CC 抄送:另一个收信人地址。

(4)BCC 密送:收信人地址,但其他收信人看不到这个收信人的地址。

(5)Subject:邮件的主题。

(6)Comments:备注。

(7)Keywords:关键字,用来进一步搜索邮件。

(8)Date:发信日期。

随着电子邮件的发展,这种纯文本的消息格式已不再能满足用户的需要,主要表现在它对非文本消息(如多媒体消息)的格式没有规定。即使是文本消息,也只能处理 US-ASCII 字符集,对于非该字符集的字符不能进行处理。

2)MIME 格式

多用途互联网邮件扩充(multipurpose Internet mail extensions,MIME)标准对文本邮件格式进行了扩展,使其能够支持非 ASCII 字符、二进制格式附件等多种格式的邮件消息。MIME 在 RFC822 格式基础上增加了邮件主体的结构,并定义了传送非 ASCII 码的编码规则。

在一个符合 MIME 的信息中,邮件的各个部分叫做 MIME 段,每段前缀有一个 MIME 头。MIME 头根据在邮件包中的位置,大体上分为 MIME 信息头和 MIME 段头。MIME 信息头是指整个邮件的头,MIME 段头是指每个 MIME 段的头。MIME 头包括以下主要字段。

(1)MIME-Version:提供所用 MIME 的版本号。

(2)Content-Type:定义数据的类型,以便数据能被适当地处理。有效的类型有 text、image、audio、video、applications、multipart 和 message。

(3)Content-Transfer-Encoding:说明对数据所执行的编码方式,用于对附件进行解码。

(4)Content-Description:可选的头,它是任何信息段内容的自由文本描述。

(5)Content-Disposition:用于给客户程序提供提示,以决定是否在行内显示附件或作为单独的附件。

通过在标准的邮件格式上附加各种段头,使 MIME 能够满足人们对多媒体电子邮件的需求,弥补了原来信息格式的不足。实际上 MIME 不仅仅是邮件编码,现在已经成为HTTP 标准的一部分。

2.3.3 SMTP

简单邮件传送协议(simple mail transfer protocol,SMTP)是一种简单有效的、提供可靠电子邮件传输的协议。SMTP 独立于特定的传输子系统,且只需要可靠有序的数据流信道支持。使用 SMTP 可实现相同网络上处理机之间的邮件传输,也可通过中继器或网关实现某处理机与其他网络之间的邮件传输。

SMTP 工作在两种情况下:一是电子邮件从客户机传输到服务器;二是电子邮件从某一个服务器传输到另一个服务器。SMTP 也采用客户机/服务器模型实现,具体步骤如下。

(1)邮件接收服务器和邮件发送服务器在默认的 25 端口之间建立一条连接。

(2)连接建立好之后,邮件发送服务器等待邮件接收服务器传输信息。

(3)邮件接收服务器首先发出准备接收的 SMTP 消息,接着由发送服务器发出 HELO 消息,接收服务器回答以 HELO 消息,双方进入邮件传输状态。

(4)由邮件发送服务器首先发出邮件的发信人地址(MAIL FROM)和收信人的地址(RCPT TO),接收服务器确认收信人存在后,发出可以继续发送的指示,发送方再发送真正的消息(DATA)。

(5)邮件发送完成之后,释放 25 端口上建立的连接。

常见的 SMTP 命令如表 2-3 所示。

表 2-3　常见的 SMTP 命令

命　令	说　明
HELO	向服务器标识用户身份
MAIL	初始化邮件传输
RCPT	标识单个的邮件接收人,常在 MAIL 命令后面
DATA	在单个或多个 RCPT 命令后,表示所有的邮件接收人已标识
VRFY	用于验证指定的用户/邮箱是否存在
EXPN	验证给定的邮箱列表是否存在
HELP	查询服务器支持的命令
NOOP	无操作,要求服务器作 OK 应答
QUIT	结束会话
RSET	重置会话,当前传输被取消

2.3.4　POP3 协议

邮局协议(post office protocol,POP)用于电子邮件的接收,现在常用的是第三版,简称为 POP3。POP3 是第一个离线的电子邮件协议,允许用户从服务器上接收邮件并将其存储到本地主机,同时根据客户端的操作,删除或保存在邮件服务器上的邮件。

POP3 也采用客户机/服务器模型。初始时,客户通过用户代理向邮件服务器发出连接建立请求。当连接建立后,客户需要向邮件服务器发送用户名和密码信息以确认自己的身份;一旦认证成功,客户向 POP3 服务器发送命令来完成各种邮件操作,这个过程一直要持续到连接终止。

常用的 POP3 命令如下。

(1)USER username:认证用户名。

(2)PASS password:认证密码,认证通过则状态转换。
(3)APOP name,digest:一种安全传输口令的方法,执行成功则状态转换。
(4)STAT:请求服务器回送邮箱统计资料,如邮件数、邮件总字节数。
(5)UIDL:服务器返回用于该指定邮件的唯一标识,如果没有指定,则返回所有的。
(6)LIST:服务器返回指定邮件的大小等。
(7)RETR:服务器返回邮件的全部文本。
(8)DELE:服务器标记删除,QUIT 命令执行时才真正删除。
(9)RSET:撤销所有的 DELE 命令。
(10)TOP n,m:返回 n 号邮件的前 m 行内容,m 必须是自然数。
(11)NOOP:服务器返回一个肯定的响应。
(12)QUIT:结束会话。

2.3.5 邮件系统的构建

邮件系统也采用典型的客户机/服务器模型,创建邮件系统的核心就是配置邮件服务器端的 POP 和 SMTP。Windows Server 2003 自带的服务组件中包含电子邮件服务器的功能,可以通过简单的操作完成配置。另外,也有很多专门的软件提供邮件服务,如 Winmail Mail Server、Exchange、CmailServer 和 Foxmail 服务器程序等。

Winmail Mail Server 是一款安全易用的邮件服务器软件,既可以作为局域网邮件服务器和互联网邮件服务器,也可以作为拨号 ISDN、ADSL 和宽带有线通(cable modem)等接入方式的邮件服务器和邮件网关。下面以 Winmail Mail Server 为例,介绍邮件系统的构建与配置。

(1)双击 winmail.exe,打开如图 2-20 所示的 Winmail Mail Server 安装向导,单击“下一步”按钮,选择安装目录,单击“下一步”按钮。

图 2-20 开始安装

(2)在“选择组件”窗口中选择需要安装的组件,如图 2-21 所示。

图 2-21　选择安装组件

(3)单击“下一步”按钮,设置“开始”菜单的快捷方式名称后单击“下一步”按钮,设置要安装的附加任务,如图 2-22 所示。

图 2-22　选择附加任务

(4)单击“下一步”按钮,在打开的窗口中设置管理工具的密码后单击“安装”按钮,开始安装。安装完成后,启动 Winmail Mail Server,右击任务栏中的 Winmail Mail Server 图标,从弹出的快捷菜单中选择“邮件系统管理”命令,打开“连接服务器”对话框,配置端口、用户名及密码等信息,如图 2-23 所示。

在邮件服务器端配置成功后,可以选用任一邮件客户端(如 Foxmail),配置后即可实现电子邮件的收发。在第一次登录时,Foxmail 会自动提示用户输入邮箱账号,在正常情况下,执行“工具”→“账号管理”→“新建账号向导”命令,打开“新建账号向导”对话框进行配置,如图 2-24 所示。对于常见邮箱类型,Foxmail 可以自动识别接收和发送服务器,对于不常见的邮箱类型,可以手动设置服务器类型。配置完成后,就可以使用 Foxmail 收发邮件了。

图 2-23 “连接服务器”对话框

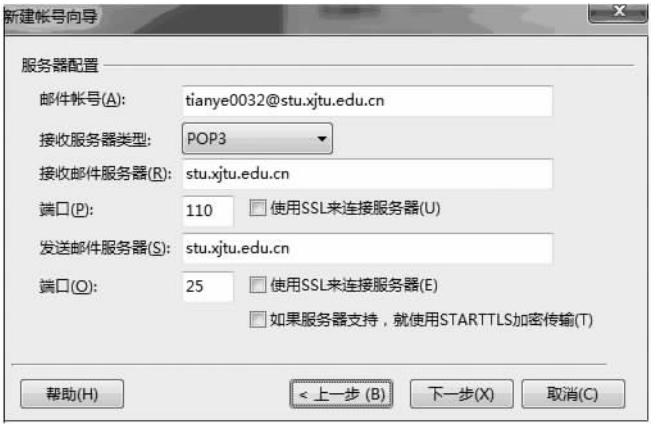


图 2-24 新建账号向导

2.4 搜索引擎服务

搜索引擎(search engine)是指根据一定的策略,运用搜索程序从互联网上获取信息,为用户提供信息服务的系统。搜索引擎源于 1990 年由蒙特利尔大学 Alan Emtage 发明的 Archie,Archie 是一个自动索引互联网上匿名 FTP 网站文件的程序。1995 年推出的 AltaVista 被认为是第一个支持自然语言搜索的搜索引擎。1998 年 10 月,Google(谷歌)诞生,在这之后 Sohu、Baidu 等搜索工具不断涌现,搜索技术也不断推陈出新。

2.4.1　搜索引擎的原理和分类

1)搜索引擎的原理

概括而言,搜索引擎的实现就是一个搜集信息、整理信息、接收用户查询和反馈查询的过程,分为以下几个步骤。

(1)爬行和抓取。搜索引擎的信息搜集基本都是自动的,它通过名为“网络蜘蛛(web spider)”或“爬虫”的程序在网上发现新网页并抓取文件。爬虫从网站某一个页面开始,找到网页中的其他链接地址,然后通过这些链接地址寻找下一个网页,直到把这个网站所有的网页都抓取完为止,这个过程称为爬行。

(2)建立索引(index)数据库。搜索引擎整理信息的过程称为“创建索引”。由分析索引系统对抓取回来的页面进行分析,提取页面的相关信息(如网页文字内容、关键词出现的位置、页面所在的 URL 以及与其他页面的链接关系等),并利用这些信息建立网页索引数据库。

(3)接受查询和返回查询结果。用户在搜索引擎界面输入关键词,由搜索引擎程序根据所有包含搜索词的网页对搜索词的关联关系,将排序后的搜索结果返回给用户。

2)搜索引擎的分类

按工作方式来分,搜索引擎主要分为全文搜索引擎、目录搜索引擎和元搜索引擎 3 种。

全文搜索引擎主要依靠网络爬虫自动获取网页信息。用户查询时,检索程序根据事先建立的索引进行查找,并将查找的结果反馈给用户。常用的全文搜索引擎有百度、Google 等。

目录搜索引擎以人工方式搜集信息,为用户提供目录浏览和检索服务。用户可以不用关键词进行查询,仅靠分类目录也可找到需要的信息。目录索引中最具代表性的有Yahoo、新浪分类目录搜索。

元搜索引擎(meta search engine)最大的特点是共享多个搜索引擎的资源库。它没有自己的索引库,而是将用户的查询请求提交至多个搜索引擎,对返回的结果进行排序等处理后返回给用户。著名的元搜索引擎有 InfoSpace、Dogpile、Vivisimo 和国内的搜星搜索引擎等。

2.4.2　搜索引擎的结构

搜索引擎一般由搜集器、分析器、索引器、检索器和用户接口等五大模块构成。

1)搜集器

搜集器(即爬虫)的功能是搜集互联网中各种类型的信息。搜索器常采用分布式和并行处理技术,以提高信息发现和更新的效率。由于互联网上的信息太多,搜集器需采用一定的搜索策略对互联网进行遍历并下载文档,常见的搜索策略包括线性搜索、深度优先搜索和宽度优先搜索等。

(1)线性搜索策略。线性搜索策略是指从一个起始的 IP 地址出发,按 IP 地址递增的方式搜索后续每一个 IP 地址中的信息,而不考虑各站点的 HTML 文件中指向其他 Web 站点

的超链接地址。线性搜索策略用于小范围内的信息搜索，可以发现被引用较少或者还没有被其他 HTML 文件引用的新文件信息源。

(2)深度优先搜集策略。深度优先搜集策略是沿着 HTML 文件上的超链接层层深入，一直到达被搜索结构的叶节点。当不再有其他超链接可选择时，说明搜索已经结束。深度优先搜索策略适宜遍历一个指定的站点或者深层嵌套的 HTML 文件集，不太适合 Web 结构太复杂的大规模搜索。

(3)宽度优先搜集策略。宽度优先搜索(又称广度优先搜索)策略是从根节点开始，沿着树的宽度遍历树的节点，一旦所有节点均被访问，则搜索终止。宽度优先搜集策略往往需要花费比较长的时间才能到达深层的 HTML 文件。

2)分析器

分析器用来对搜集器搜集来的网页信息或者下载的文档进行分析处理，包括分词、过滤、词缀去除、同义词转换等，以便建立索引文档。

3)索引器

索引器的功能是从分析器处理后的信息中抽取索引项，生成索引表。倒排索引库(inversion list)是一种适合大规模数据检索的数据结构，能有效且快速查询出相关的文档。

4)检索器

检索器的功能是根据用户的查询在索引库中快速检索出文档，进行文档与查询的相关度评价，并对将要输出的结果进行排序。检索器常用的信息检索模型有集合理论模型、代数模型、概率模型和混合模型等多种，可以查询到文本信息中出现在标题或是正文中的任意字词。

5)用户接口

用户接口的作用是为用户提供可视化的查询输入和结果输出界面，方便用户输入查询条件，获取查询结果。

2.4.3 搜索引擎的应用

互联网上每天都有无数个页面被更新、创建，随之就产生了无数新的信息。面对互联网上的海量信息，用户想要获得自己所需的信息如同大海捞针。这使得用户对搜索引擎的依赖度不断增加，搜索的领域也从原来单一的页面信息扩大到更多专门的信息领域。

搜索引擎被广泛应用于各行各业中，如政府、金融、电信、航空航天、教育、出版及零售等。从搜索内容上来看，搜索引擎可以被用于各个领域的信息搜索。以 Google 为例，它提供的主要搜索服务有：网页搜索、图片搜索、视频搜索、地图搜索、新闻搜索、论坛搜索、学术搜索和财经搜索等，能够同时满足不同人群对不同领域信息的搜索。

随着搜索引擎各项技术的日趋成熟，业界逐渐将目光从搜索技术本身转移到搜索产品的多样化上，新兴的搜索产品主要分为多媒体搜索、网页级搜索和对象级搜索。目前的图片搜索、音乐搜索和视频搜索都属于多媒体搜索，技术相对成熟。而网页级搜索和对象级搜索是更细粒度的搜索，通常以页面或者页面中某些具体内容(称为对象)的形式返回用户。典型的对象级搜索，如微软的 Libra 论文搜索系统，能根据用户的需求提取出论文的相关

属性。

从用户查询角度而言，如何更准确地理解用户的查询需求，也是有效保证查准、查全的一个关键问题。为了克服关键词检索和目录查询的缺点，现在已经出现了自然语言智能查询。用户可以输入简单的疑问句，搜索引擎在对提问进行结构和内容的分析之后，或直接给出提问的答案，或引导用户从几个可选择的问题中进行选择。自然语言的优势在于，使网络交流更加人性化，使查询变得更加方便、直接和有效。

应该说搜索引擎还有很大的发展空间，无论从技术上还是应用上而言，未来的搜索引擎还将面临很多新的挑战。

2.5　P2P 服务

P2P 技术是指网络中的所有节点都动态参与到路由、信息处理和带宽增强等工作中，而不是单纯依靠服务器来完成这些工作。与传统的客户机/服务器模型相比，P2P 模型的优势在于降低了节点对服务器的依赖以及它的分布式控制能力强。

2.5.1　P2P 服务的模型

从结构上划分，P2P 模型分为以下几种类型。

1）集中式 P2P

集中式 P2P 是最早出现的 P2P 模式，其结构如图 2-25 所示。在这种结构中，一台或多台有特殊用途的中心服务器为对等节点提供目录服务。对等节点向目录服务注册关于自身的信息（如名称、地址、资源和元数据等），并通过目录服务中信息的查询，使用目录服务来定位其他对等节点。

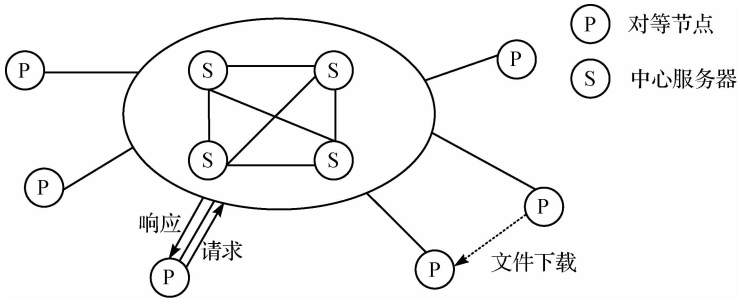


图 2-25　集中式 P2P 模型的结构

集中式 P2P 模式最大的优点是维护简单，资源发现效率高并能够实现复杂查询。缺点是与传统客户机/服务器模式类似，存在诸如对中心服务器依赖过高、可靠性和安全性较低等问题。集中式 P2P 模式的典型代表是 MP3 共享软件 Napster。

2)分布式非结构化 P2P

分布式非结构化 P2P 又称纯 P2P 或广播式 P2P,它的特点是取消了中心服务器,每个用户随机接入网络,并与自己相邻的一组邻居节点通过端到端连接构成一个逻辑覆盖的网络,节点度数服从 power-law 规律(幂次法则)。Gnutella 模型是应用最广泛的广播式 P2P 模型,其结构如图 2-26 所示。在这种结构中,每个节点 P 既是客户机也是服务器,既可以接收请求,也可以响应请求。

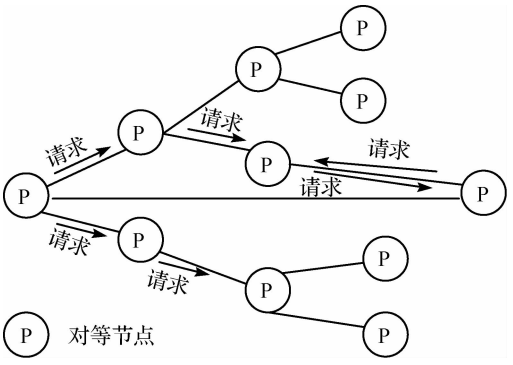


图 2-26　广播式 P2P 模型的结构

广播式 P2P 的优点在于能够较快发现目的节点,具有较好的容错能力和可用性。缺点是广播带来的控制信息的泛滥会消耗大量的网络带宽,很容易造成网络拥塞,导致网络中部分节点因网络资源过载而失效。

3)分布式结构化 P2P

分布式结构化 P2P 主要采用分布式散列表(distributed hash table,DHT)技术来组织网络中的节点。DHT 是一个由广域范围大量节点共同维护的巨大散列表,散列表被分割成不连续的块,每个节点被分配给一个散列块,并成为这个散列块的管理者。DHT 类结构能够自适应节点的动态加入和退出,有良好的可扩展性和鲁棒性,典型的应用包括 Tapestry、Pastry、Chord 和 CAN。

4)混合式 P2P

混合式 P2P 又称半分布式 P2P,这种结构同时具备集中式结构和分布式非结构化结构的一些特点。Kazaa 模型是混合式 P2P 模型的典型代表,如图 2-27 所示。在 Kazaa 模型中,选择在处理、存储、带宽等各方面性能较高的节点作为超级节点,在各个超级节点上存储了系统中其他部分节点的信息,由超级节点负责将查询请求转发给适当的叶子节点。混合式结构的优点是性能、可扩展性较好,较容易管理,但对超级节点依赖性大,易于受到攻击,容错性也受到影响。

2.5.2　P2P 协议

Internet 工程任务组(IETF)、中国通信标准化协会(CCSA)、美国分布式计算产业协会(DCIA)以及中国 P2P 标准化工作组都制定了相关的 P2P 协议标准,覆盖 P2P 结构、P2P 流

量优化、P2P 流媒体直播和 P2P 传播内容等各个方面。与传统的 C/S 模式相比，P2P 协议中涉及一些特有的属性和内容，主要体现在以下几方面。

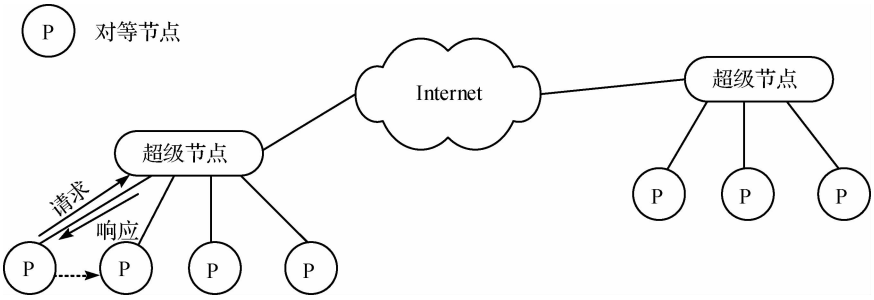


图 2-27　混合式 P2P 模型的结构

1)P2P 的资源请求

P2P 在拓扑结构上与传统的客户机/服务器模式完全不同，这决定了 P2P 在资源传输和对服务的请求方式等方面也存在很大区别。若客户机要请求的另一对等客户机的资源在 C/S 模式下，首先需要客户端 A 与 Web 服务器建立连接，将创建的资源发布到 Web 服务器上，再由客户端 B 与 Web 服务器建立连接，向 Web 服务器发出查看资源的请求；最后由 Web 服务器响应 B 的请求，将 A 发布的资源提供给 B。而当客户端 B 以典型的 P2P 方式查看客户端 A 上的资源时，由客户端 B 和客户端 A 直接建立连接，发出请求，A 可以直接将资源传给 B。

2)P2P 的资源标识和存储

在 P2P 结构中，资源分散存储在各个节点上。用户要找到某个资源除了需要用到一定的搜索策略外，还需要给每个资源一个特定的标识。根据不同的应用场合和特性，P2P 系统有不同的资源标识方法。在以文件共享为主的应用中，资源主要以文件名称、关键字和源数据等进行标识。在即时通信系统中通常采用类似电子邮件的命名方式，即为每个节点分配一个全局统一的地址标识，形如 node@domain/resource 的形式。其中，node 是用户标识；domain 是主要的 ID 标识，是与多个用户连接进行消息转发的标识；resource 属于一个 node，用于标识属于一个用户的多个资源。

3)P2P 的内容传输

要实现 P2P 资源请求和内容查询，除了解决 P2P 资源的标识和存储之外，还要解决的一个主要问题就是 P2P 的内容传输，即 P2P 结构中任何两个建立连接的节点之间，信息（或资源）如何传递以及内容传输过程中存在的防火墙、数据质量以及数据描述统一性等问题。

为了保证节点之间的互操作性，P2P 需要借助一个通用的平台和必要的协议来保证发布出去的信息能够被所有节点接收，这些协议有 XML、SOAP、UDDI 等。P2P 结构具有节点多、网络带宽占用多、数据流量大等特点，因此对于实时性要求比较高的业务类型（如多媒体应用），基于 P2P 进行内容传输时还需要考虑数据丢包率、时延等网络服务质量问题。

2.5.3 P2P 信息检索与共享

P2P 信息检索是指从分布在各个独立节点上的 P2P 网络信息资源中高效地索引、查找和检索出用户所需信息的过程。为了提供信息检索，P2P 网络中的每个节点在加入网络时，会对存储在本节点上的内容进行索引，以满足本地内容检索的要求，然后按某种预定的规则选择一些节点作为自己的邻居，加入 P2P 网络中。

按检索对象来分，P2P 信息检索可分为文本信息检索、图像信息检索、声音信息检索、视频信息检索以及其他文件的信息检索等。按检索方式来分，P2P 信息检索又分为集中式检索、泛洪式检索与分布式哈希表检索 3 种。

1)集中式检索

P2P 集中式信息检索是指构建一个中央服务器，在中央服务器中存储注册的文件，由中央服务器完成其他节点提交的资源查询请求。集中式检索一般基于混合式 P2P 结构，各节点间的通信采用点对点方式直接进行。显然，中央服务器具有单点失效、可扩展性不强等缺陷。

2)泛洪式检索

泛洪式检索的基本思想是把查询信息传递给节点的所有相邻节点，如果所有相邻节点含有这个资源，就返回信息给请求者。如果所有相邻的节点都没有找到被查询文件，就把这条消息继续转发给自己的相邻节点，这种方式叫做泛洪式搜索。泛洪式检索模型通常基于集中式 P2P 结构。

3)分布式散列表检索

在分布式散列表(distributed hash table，DHT)检索中，每个节点不仅存储本身的内容索引，而且维护其他节点上部分特定内容的索引，维护该内容的索引节点 ID 可通过散列计算得到。因此，为了获得相关内容节点的 ID，仅对该关键字进行散列计算即可，常见的算法有 Chord、CAN、Tapestry、Pastry 等，它们都是基于 DHT 提供的可扩展的分布 peer ID 查找机制。

2.5.4 P2P 的应用实例

随着国内外机构对 P2P 技术的关注和研究，P2P 技术正不断应用到军事、商业、政府信息和通信等领域。下面介绍几种典型的基于 P2P 技术的应用软件。

1)Napster

Napster 是一款可以在网络中下载 MP3 文件的软件，是一个大型的 P2P 应用软件。Napster 本身并不提供 MP3 文件的下载，它实际上提供的是整个网络的 MP3 文件目录，而 MP3 文件分布在网络中的每一台机器中供用户选择。Napster 具有强大的搜索功能，可以将在线用户的 MP3 音乐信息进行自动搜寻并分类整理，以备其他用户查询。

2)JXTA

JXTA 是以 Java 技术为背景开发的一种标准组件平台，它提供了用于开发分布式服务

和应用程序的基本组件。JXTA 能简化文件共享，自动侦测到新的网站目录，实现 P2P 中对等节点的远程监控以及访问深层网络的数据。

3）Skype

Skype 是一种网络即时语音通信工具，具有视频聊天、多人语音会议、传送文件和文字聊天等功能。Skype 是 P2P 技术演进到混合式后的典型应用，在网络的边缘节点采用集中式的网络结构，而在超级节点之间采用分布式的网络结构。2011 年 5 月微软收购了 Skype。

4）PPLive

PPLive 是一款全球安装量很大的 P2P 网络电视软件，支持对海量高清影视内容的“直播＋点播”功能。与传统的网络电视相比，PPLive 的特点是用户越多，速度越快，彻底改变了用户量和网络带宽之间的问题。同时，PPLive 有效地解决了内网穿透问题，解除了 Windows XP 对 TCP 连接数的限制。

5）Maze

Maze 是北京大学网络实验室开发的一个中心控制与对等连接相融合的 P2P 文件共享系统，在结构上类似于 Napster，对等计算搜索方法类似于 Gnutella。每个节点可以将自己的一个或多个目录下的文件共享给系统的其他成员，也可以分享其他成员的资源。

2.6　博客与社交网络

随着 Internet 的不断发展，人们对互联网的要求已不单单是进行数据传输和内容查询，而是使之成为人们之间互动和联系的一个纽带。博客、微博等新型网络应用的出现，使得用户可以通过 Web、无线应用协议（wireless application protocol，WAP）以及各种客户端方便地发表言论、抒发情感，进而通过互联网来建立社交关系，构筑社交网络。社交网站作为兼容博客、微博、论坛、群组和即时通信等功能的社会化媒体，在人们的网络交流中扮演着越来越重要的角色。

2.6.1　博客与微博

“博客”一词从英文单词 blog 音译而来。blog 是 weblog 的简称，中文意思即网络日志。博客是近些年一种新生的网络应用，为人们提供了一种简易迅速发布自己的心得、有效轻松地与他人进行交流的手段。

一个 blog 就是一个网页，它通常是由简短且经常更新的帖子（post）构成。按功能划分，博客分为基本博客和微型博客（microblog，又称微博）两大类。基本博客是 blog 中最简单的形式，是指单个作者对于特定的话题提供相关的资源或发表简短的评论。微博与基本博客最大的区别在于它对博客作者发表文字的字数通常限制在 150 字左右，这样人们可以方便地通过手机等移动设备随时随地与他人交互信息。

国外最早也是最著名的微博 Twitter 于 2006 年 3 月由 Blog 的创始人 Evan Williams 推出，Twitter 英文原意为小鸟的叽叽喳喳声。用户可以由短消息业务（short message service，SMS）、即时通信、电子邮件、Twitter 网站或 Twitter 客户端软件（如 Twitterrific）等多种工具更新信息。截止到 2010 年 11 月，Twitter 拥有约 1.75 亿注册用户。

与此同时，国内各大门户网站（如新浪、网易、搜狐和腾讯等）也纷纷推出自己的博客和微博系统。中国互联网络信息中心（CNNIC）发布的《第 28 次中国互联网络发展状况统计报告》显示，2011 年上半年，中国微博用户从 6 331 万增至 1.95 亿。

2.6.2 社交网络

1）社交网络定义

社会性网络服务（social network service，SNS，简称社交网络）一词于 1954 年由 J. A. Barnes 首先使用，旨在帮助人们建立社会性网络的互联网应用服务，前面介绍的微博或博客就构成了典型的社交网络。

一般而言，社交网络是由个人或社区组成的点状网络拓扑结构，如图 2-28 所示。其中每个点（node）代表个体，可以是个人，也可以是一个团队或一个社区。个体与个体之间可能存在各种相互依赖的社会关系，在拓扑网络中以点与点之间的边（tie）表示。图中虚线圈起来的部分表示某些具有紧密社会关系的节点集合。

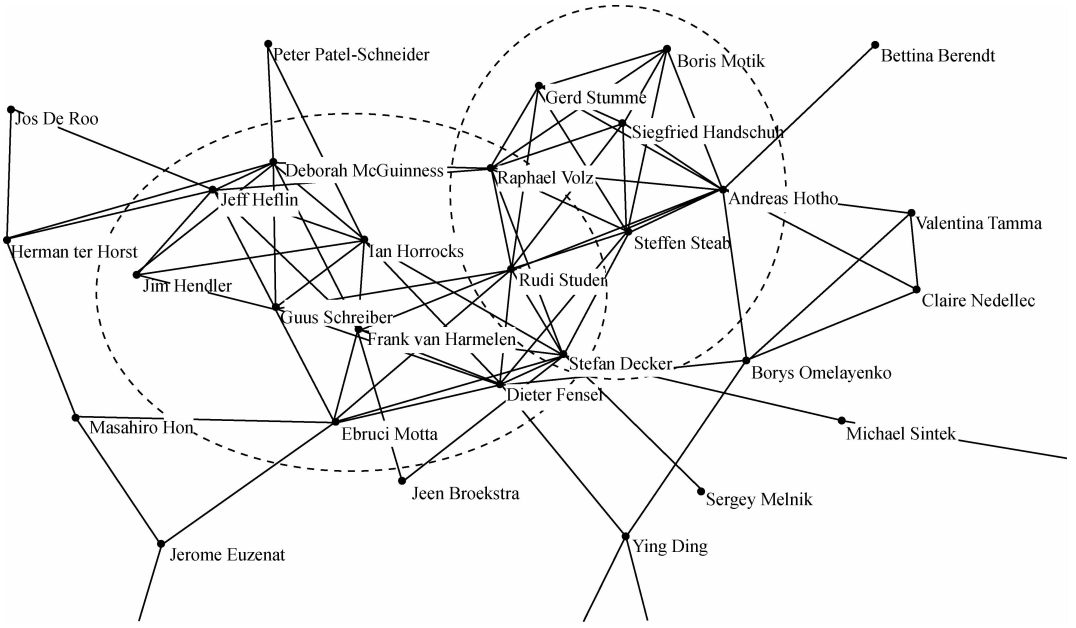


图 2-28　社交网络图例

社交网络的理论模型是哈佛大学著名心理学教授 Stanley Milgram 所创立的“六度分隔理论”。该理论的核心思想是你和任何一个陌生人之间所间隔的人不会超过 6 个，也就是说，最多通过 6 个人你就能够认识任何一个陌生人。按照六度分隔理论，每个个体的社交圈

都不断放大，最后成为一个大型网络，这就是社会性网络的早期理解。后来人们根据这种理论创立了面向社会性网络的互联网服务，通过人与人之间的交往来进行网络社交拓展，从而形成现在的社交网络。

目前，社交网络几乎涵盖了以人类社交为核心的所有网络服务形式，使得互联网从研究部门、学校、政府、商业应用平台扩展成人类社会交流的一种工具。

2）社交网络的发展阶段

早期社交网络服务呈现为在线社区的形式，用户多通过 BBS、新闻组或聊天室进行交流。随着 blog 等新的网络交际工具的出现，用户可以通过网站上创建的个人主页来分享信息。2002 年至 2004 年，三大最受欢迎的社交网络服务类网站分别是 Friendster、MySpace 和 Bebo。2006 年 Facebook 一跃成为全球用户量增长最快的社交网站。

总结起来，社交网络经历了下面几个发展阶段。

（1）现实工具的网络映像。在这个阶段，用户主要通过一些现实工具的网络映像来进行社交活动，如电子邮件、FTP 和个人 Web 网站。它们的共同特点是信息不规范且孤立，只有极少数人能将其发展成拓宽交往的有效工具。

（2）即时通信软件。这个阶段以即时通信（instant messenge，IM）软件和 BBS（bulletin board system）为代表，如利用 QQ 或 MSN 进行网络聊天。但这些工具依然以某一项软件服务和网站服务为中心设计，缺乏对用户潜在心理需求的挖掘分析和主动发现。

（3）基于 SNS 理论的社交网络服务。SNS 理论是对人群交往的心理动机、行为特点和相互影响因素进行分析的社会学分支，通过 SNS 服务，交友变得有目的性和易于操作。

（4）高级阶段——社会性网络。严格意义上来说，互联网提供的 SNS 服务，是帮助人们更有效地交流，更便捷地维护这些交往关系。社会性网络是现实社交网络的一种补充，它通过网络把用户的需求获取和发布延伸到无限广的范围。

2.6.3　社交网络平台

目前有许多提供社交网络服务的网站，国外比较著名的有 Facebook、Quazza. com、MySpace、Orkut 和 Twitter 等，国内的社交网络有人人网、开心网等。多数社交网络会提供多种用户交互的方式，包括聊天、发送邮件、文件分享、博客和讨论组群等。下面简单介绍几个社交网络平台。

1）Facebook

Facebook 又称脸谱、面书、面簿，是目前注册人数最多、影响力最大的社交网站之一。Facebook 的创始人马克·扎克伯格是哈佛大学的学生。最初，网站的注册仅限于哈佛大学的学生，但很快很多其他学校的学生也加入进来。

Faccbook 提供涂鸦墙（the wall）、戳（poke）、礼物（gift）、市场（marketplace）、活动（events）和视频（video）等基本功能，可以很方便地进行用户之间的信息交流。2007 年 5 月，Facebook 推出开放平台应用程序接口。利用这个接口，第三方软件开发者可开发与 Facebook核心功能集成的应用程序。目前已有超过 5 000 个应用程序被开发出来，包括小游戏、社会化音乐发现、分享服务和数据统计等，如可以很方便地将 Twitter 发布的信息同

步到 Facebook 的应用程序，也可以在 Facebook 发布信息到 Twitter 上。

2）MySpace

MySpace（聚友）也是一个著名的社交网络服务网站，提供人际互动、用户自定的朋友网络、个人档案页面、博客、音乐和视讯影片分享与存放等功能。MySpace 在 2003 年 7 月开始，由汤姆·安德森、克里斯·德沃菲以及一个小型程序设计师团队创立，巅峰时期全球用户数量超过 1 亿。直到 2007 年之前，MySpace 一直占据着全球最大社交网站的宝座。但近几年随着 Facebook 和 Twitter 的快速崛起，MySpace 已滑落至 Windows Live 之后。MySpace 于 2011 年 6 月被 Specific Media 收购。

3）iWebSNS

iWebSNS 是一套基于开源软件平台 LAMP（Linux＋Apache＋Mysql＋Perl/PHP/Python）开发构建的社交网络软件。iWebSNS 内置日志、相册、分享、群组、心情、投票等十多个功能模块。借助 iWebSNS 平台，不仅可以轻松构建起一个以人际关系为核心的网站，还可以获得支持热插拔及快速增加新节点的集群计算与处理能力，以方便管理 Web 2.0 类站点持续增长的数据量。

4）人人网

人人网（原称校内网）是一个类似 Facebook 的 SNS 网站，同时也是国内最早的校园社交关系网络平台之一。校内网于 2005 年 12 月由清华大学和天津大学的几名大学生创建，并于 2009 年 8 月 14 日改称人人网。人人网搭建了一个功能丰富且高效的用户交流互动平台，推出的人人派对、人人农场、人人爱听等应用受到用户的欢迎。

5）开心网

开心网是国内第一家以办公室白领用户群体为主的社交网站，由北京开心人信息技术有限公司于 2008 年创办。开心网成立几年来，网站注册用户突破 1 亿。开心网组件主要分为基础工具、社交游戏和其他应用三大类。基础工具类别主要提供信息分享方面的服务，包括照片、日记、记录、转帖等丰富的应用；社交游戏类别包括开心城市、开心庄园和开心餐厅等众多热门游戏；其他应用类别包括天气预报、在线购票、模拟炒股等众多实用工具，已被用户广泛使用。

习　题　2

1）选择题

（1）在 Internet 中，用来进行文件传输控制的协议是（　　）。

A. IP　　B. TCP　　C. HTTP　　D. FTP

（2）Internet 的域名中，顶级域名 gov 代表（　　）。

A. 教育机构　　B. 商业机构　　C. 政府部门　　D. 军事部门

(3)在 http://www.sina.com 中,http 代表(　　)。

A. 主机　　B. 地址　　C. 协议　　D. TCP/IP

(4)HTML 的含义是(　　)。

A. 主页制作语言　　B. 超文本标记语言

C. WWW 编程语言　　D. Internet

(5)超文本的含义是(　　)。

A. 文本中可含有图像

B. 文本中可含有声音

C. 文本中有超级链接

D. 文本中有二进制字符

(6)用 Internet 访问某主机可以通过(　　)。

A. 地理位置　　B. IP 地址

C. 域名　　D. 从属单位名

(7)如果一个电子邮件的地址为××××@0451.com,则××××代表(　　)。

A. 用户地址　　B. 用户名

C. 用户口令　　D. 主机域名

(8)在 Internet 电子邮件系统中,(　　)。

A. 发送邮件和接收邮件都使用 SMTP

B. 发送邮件使用 POP3 协议,接收邮件使用 SMTP

C. 接收邮件使用 POP3 协议,发送邮件使用 SMTP

D. 发送邮件和接收邮件都使用 POP3

(9)下面不是 P2P 常用模型结构的是(　　)。

A. 集中式　　B. 广播式　　C. 点对点式　　D. 混合式

(10)P2P 的资源标识 node@domain/resource 中,domain 表示的是(　　)。

A. 用户标识　　B. ID 标识　　C. 消息转发标识　　D. 都不是

2)填空题

(1)在客户机/服务器模型中,由_____主动发起通信请求。

(2)在电子邮件系统中,SMTP 用于实现_____。

(3)传统的 Web 服务、E-mail 服务和 FTP 服务都采用_____模型来实现。

(4)Internet 的主要服务有_____、_____、_____和_____。

(5)用于电子邮件接收的常用协议是_____。

(6)搜索引擎一般由_____、_____、_____、_____和_____ 5 个模块组成。

(7)P2P 常用的 3 种信息搜索方式为_____、_____和_____。

(8)按工作方式来分,搜索引擎主要分为_____、_____和_____。

(9)常用的全文搜索引擎有_____和_____。

(10)列举 3 个应用广泛的社交网络,包括_____、_____和_____。

3)简答题

(1)为什么绝大多数的 Internet 应用层协议采用客户机/服务器模型?

(2)P2P 模型与客户机/服务器模型的主要区别是什么?
(3)在 FTP 中,21 和 20 端口各实现什么功能?
(4)在电子邮件协议中为什么要引入 MIME?
(5)简单地说 SMTP 和 FTP 都是完成文件传输的,它们在本质上有何区别?
(6)简述 POP3 协议的工作原理。
(7)搜索引擎的基本工作原理是什么?
(8)P2P 模型与 C/S 和 B/S 模型的主要区别是什么?
(9)分布式散列表检索是如何实现的?
(10)集中式检索、泛洪式检索与 DHT 检索各自的特点是什么?

第 3 章 互联网传输协议

在互联网中主要采用 TCP/IP 体系结构作为其支撑协议，TCP/IP 体系结构由一组以 TCP 和 IP 为代表的协议组成。按照第 1 章介绍的计算机网络的层次化结构，这组协议中的网络层协议解决了将源节点产生的数据送到目的节点的问题，而传输层的协议则保证了数据不仅能正确送到目的地，还能保证数据传输满足一定的服务质量要求，如可靠性、时延等。本章介绍 TCP/IP 体系结构中的传输层和网络层的核心协议，包括 UDP、TCP、IP 和多播协议等。

3.1 TCP/IP 体系结构概述

TCP/IP 体系结构起源于美国远景研究规划局（Advanced Research Project Agency，ARPA）提出的 ARPANET，由于其具有简单性、灵活性、易于实现并能充分满足各层次用户的需求等优点，自从在 20 世纪 70 年代诞生以来已经赢得了大量的用户和投资。TCP/IP体系结构的成功促进了 Internet 的发展，Internet 的发展又进一步扩大了 TCP/IP 体系结构的影响。

TCP/IP 体系结构的层次结构以及与 OSI/RM 的对应关系如图 3-1 所示。从图 3-1 中可以看出 TCP/IP 体系结构分为 4 个层次，自下而上分别是网络接口层、网际层、传输层和应用层。

1）网络接口层

网络接口层（network interface layer）对应于 OSI/RM 中的物理层和数据链路层，TCP/IP体系结构中将物理层和数据链路层作为一层处理的原因是这两层解决的是物理网络（或接口）的点到点传输问题。简单地说，网络接口层负责将网际层的 IP 分组通过物理网络发送出去，或者从物理网络中接收数据帧，抽取出 IP 分组后递交给网际层。TCP/IP 体系结构

中的网络接口层严格说并不是一个独立的层次，只是一个接口，TCP/IP 并没有定义任何特定的协议，而是支持现有物理网络的标准和协议。这样设计的目的是提供灵活性，以适用于不同的物理网络。

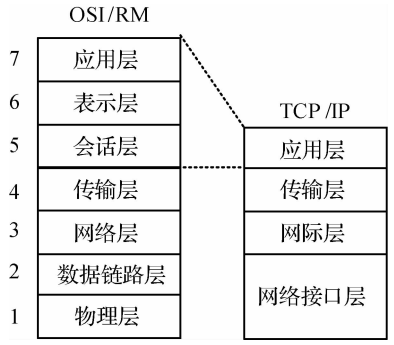


图 3-1　TCP/IP 体系结构

在 Internet 中，可以使用的物理网络种类很多，如各种 LAN、MAN、WAN，甚至点对点链路。各种物理网络的差异可能很大，这种差异体现在数据格式、传输速率和地址格式等方面。但在 TCP/IP 看来，各种物理网络都是 Internet 的构件，在 IP 分组的传输过程中，它们都作为两个相邻交换节点之间的一条物理链路。物理网络不同，只体现在接口上的不同。网络接口层使得上层的 TCP/IP 和底层的物理网络无关。

2)网际层

网际层(Internet layer)也被称为互连网络层，对应于 OSI/RM 中的网络层。网际层提供的是一种无连接、不可靠但尽力而为(best effort)的数据报传输服务，该服务负责将数据分组从源主机传送到目的主机。从一台主机传送到另一台主机的分组可能会通过不同的路由，且分组可能出现丢失和乱序等问题。为了达到较高的分组传输速度，网际层放弃了可靠性等服务质量的保障。

TCP/IP 体系结构的网际层中最核心的协议是网际协议(Internet protocol，IP)。网际层传送的数据单位是 IP 数据报(IP datagram)，通常被称为 IP 分组。在网际层还有很多协议来协助 IP 完成更复杂的功能，如完成 IP 地址到 MAC 地址转换的 ARP，提供差错和状态报告机制的 ICMP，实现互联网路由信息交换的 OSPF 协议等。网际层的主要协议将在 3.5 节中介绍。

3)传输层

传输层(transport layer)也称为运输层，对应于 OSI/RM 中的传输层。传输层为应用进程提供端到端的传输服务。简单地说，传输层负责将源主机产生的完整报文送到目的主机。传输层的端到端只涉及源节点和目的节点的两个传输层实体(通常是两个进程)，不涉及网络中的路由器等中间节点。

TCP/IP 在传输层主要包括两个协议，即传输控制协议(transmission control protocol，TCP)和用户数据报协议(user datagram protocol，UDP)。一个新制定的传输层协议是流控制传输协议(stream control transmission protocol，SCTP)，它最初是为 IP 电话制定的。

一般情况下，TCP 和 UDP 共存于互联网中，前者提供高可靠性的面向连接服务，后者提供高效率的无连接服务。高可靠性的 TCP 常用于一次传输要交换大量报文的情形（如文件运输、远程登录等）；高效率的 UDP 常用于一次传输交换少量报文（尤其是交易型应用，如数据库查询）的情况，其可靠性由上层应用程序提供；因为交换次数不多，即便发生传输错误，必须重传，应用程序也不会为此付出太大的代价。

SCTP 用于提供基于不可靠传输业务的协议之上的可靠数据报传输协议。它在两个端点之间提供稳定、有序的数据传递服务，并且可以保护数据消息边界。与 TCP 和 UDP 最大的不同是 SCTP 是以多宿主和多流为基础的。多宿主为应用程序提供了比 TCP 更高的可用性，多宿主主机就是一台具有多个网络接口的主机，可以通过多个 IP 地址来访问这台主机。这样连接的每个端点的地址＝ 一组 IP 地址＋ 端口号。SCTP 将这种连接的方式称为联合，它可以使用每台主机上的多个接口进行通信。SCTP 中的流是指联合中需要按照顺序提交到高层协议的用户消息序列，在同一个流中消息需要按照其顺序进行递交。可以将流理解为联合中从一个端点到另一个端点的单向逻辑通道，通常一个联合由多个这样单向的流组成，这些流相互独立，并通过流 ID 进行区分。

传输层与网际层在功能上的最大区别是前者提供进程通信能力，后者则不提供。在进程通信的意义上，网络通信的最终地址就不仅仅是主机地址了，还包括可以描述进程的某种标识符。

为此，TCP 和 UDP 提出了协议端口（简称端口）的概念，用于标识通信的进程。应用进程通过系统调用与某个端口建立连接后，传输层送给该端口的数据都被该进程所接收。每个端口都拥有一个被称为端口号的整数标识符，以区分不同的端口。TCP 和 UDP 各自的端口号是相互独立的，端口号长度为 16 位。

TCP 和 UDP 中端口的分配方法是将端口号分为两部分，一部分是保留端口，一部分是自由端口。其中，保留端口只占很小的部分（256 以内），以全局方式进行分配，保留端口一般分配给常用的服务器进程，如 WWW、FTP 等；自由端口占据了端口号的绝大部分，以本地方式进行分配。进程与远地进程通信之前，首先需要申请一个自由端口。

表 3-1 和表 3-2 列出了一些常用的 UDP 和 TCP 保留端口。

表 3-1　常用的 UDP 保留端口

端 口 号	关 键 字	描　　述
42	NAMESERVER	主机名称服务
53	DOMAIN	域名服务
67	BOOT PS	启动协议服务
69	TFTP	简单文件传输协议
111	SUN RPC	微系统公司的 RPC（远程过程调用协议）

表 3-2　常用的 TCP 保留端口

端 口 号	关 键 字	描　　述
20	FTP-DATA	文件传输服务(数据连接)
21	FTP	文件传输服务(控制连接)
23	TELNET	远程终端服务
25	SMTP	简单电子邮件传输服务
42	NAMESERVER	主机名服务
53	DOMAIN	域名服务

UDP 和 TCP 的详细内容将在 3.2 节和 3.3 节中介绍。

4)应用层

TCP/IP 体系结构的应用层(application layer)对应 OSI 参考模型的高三层,提供面向用户的网络服务。TCP/IP 体系结构的应用层已经存在许多面向应用的著名协议,如第 2 章介绍的文件传送协议 FTP、简单邮件传送协议 SMTP 和超文本传送协议 HTTP 等。

3.2　UDP

用户数据报协议(user datagram protocol,UDP)是 TCP/IP 体系结构提供的一种无连接的传输层协议,提供面向事务的简单不可靠信息传送服务。

3.2.1　UDP 简介

UDP 的标准文档是 1980 年颁布的 RFC768。到目前为止,经历了 30 多年的时间,UDP 的内容几乎没有变化。这充分说明了 UDP 以“简单”为核心的设计原则的吸引力和生命力。

UDP 建立在 IP 之上,提供无连接的数据报传输服务。相对于 IP,它最主要的功能是提供协议端口,以保证进程通信。因此,UDP 在时间和空间上的开销都比较小,主要应用于多媒体通信、多播通信等对实时性要求比较高或者具有重复性行为的场合。

UDP 提供的是一种无连接的服务,它并不保证可靠的数据传输。它和对等的 UDP 实体在传输时不建立端到端的连接,而只是简单地向网络上发送数据报文或从网络上接收数据报文。UDP 保留应用程序产生的报文边界,即它不会对报文进行合并或分段处理,这样,接收方收到的报文与发送时的报文大小完全一致。

另外,UDP 可以通过校验和来提供差错的检测功能。校验和是可选的,即一个应用程序在使用 UDP 发送数据时,可以自由地选择是否要求产生校验和(缺省时要求产生校验和)。当一个 IP 模块收到一个 IP 分组时,它就将其中的 UDP 数据报传递给 UDP 模块。UDP 模块在收到由 IP 模块传来的 UDP 数据报后,首先检验 UDP 校验和。如果校验和为

0，表示发送方没有计算校验和；如果校验和非 0，并且校验和不正确，则 UDP 将丢弃这个数据报；如果校验和非 0，并且正确，则 UDP 根据数据报中的目的端口号，将其送给指定应用程序等待队列。

图 3-2 给出了 UDP 报文的封装过程。从图中可以看出 UDP 报文是封装到 IP 分组中进行传输的，即 UDP 报文作为 IP 分组的数据部分加上 IP 分组头后构成了 IP 分组。当然，为了将 IP 分组通过物理接口送到物理介质中进行传输，还需要经过数据链路层的封装，即将 IP 分组作为帧的数据部分加上帧头后构成数据帧进行传输。

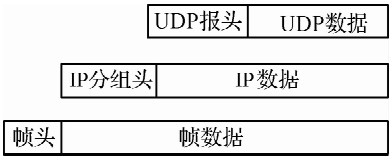


图 3-2　UDP 报文的封装

3.2.2　UDP 报文结构

UDP 报文由报头和数据两大部分组成，如图 3-3 所示。

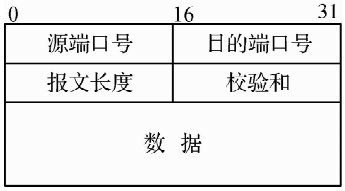


图 3-3　UDP 报文格式

其中，UDP 报头包括以下 4 个字段。

1）源端口号

源端口号指示发送方的 UDP 端口号。当不需要返回数据时，可将这个字段的值置为 0。

2）目的端口号

目的端口号指示接收方的 UDP 端口号。UDP 将根据这个字段内容将报文送给指定的应用进程。

3）报文长度

报文长度指示数据报的总长度，包括报头及数据区的总长度。单位为字节，最小值为 8，即 UDP 报头部分的长度。

4）校验和

校验和用于对 UDP 报文进行差错检测。UDP 的校验和为可选字段，当校验和的值为 0 时，表示发送方未计算校验和；为全 1（负 0 的反码）表示校验和的值为 0。UDP 校验和的可

选性是提高 UDP 效率的另一举措，因为计算校验和是一个非常耗时的工作，如果应用程序对效率要求非常高，可以不进行校验和的计算。

UDP 校验和还有一个重要内容——伪报头。伪报头用于验证 UDP 数据报是否被送到了正确的目的节点。一个目的节点的地址由两部分组成：目的 IP 地址和目的端口号，由于 UDP 报文中只包含了目的端口号，因此在伪报头中添加了目的节点的 IP 地址。伪报头不是 UDP 数据报的有效成分，而是根据 IP 分组头中的信息产生的。伪报头参与校验和的计算，但不进行实际的传输。计算校验和时，伪报头的位置在 UDP 报头的前面。伪报头格式如图 3-4 所示。

图 3-4　UDP 伪报头格式

伪报头包括以下 5 个字段。

(1)发送方 IP 地址和接收方 IP 地址：与 IP 分组中的源 IP 地址和目的 IP 地址一致。

(2)填充字段：目的是使伪报头的长度为 32 位的整数倍。

(3)协议类型：说明 IP 上层协议的协议类型，UDP 的协议类型为 17。

(4)UDP 长度：UDP 报文(不包含伪报头)的长度。

3.2.3　UDP 的传输控制

UDP 虽然不进行流量控制和拥塞控制等传输控制工作，但却要完成针对不同用户进程的多路复用和分解操作。图 3-5 给出了一个 UDP 根据目的端口号进行多路分解的例子。

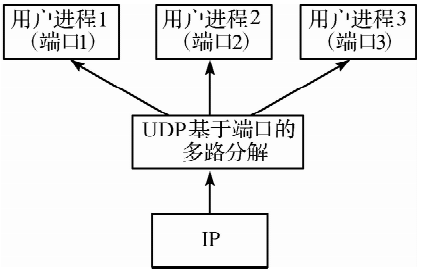


图 3-5　UDP 中的多路分解

从图 3-5 可以看出，在通信网络中进行传输时，网络中的路由节点(如路由器)根据 IP 地址将 IP 分组转发到目的主机；目的主机的 IP 模块将处理完后的 IP 分组提交给 UDP 模块，UDP 模块需要确定到底应该由哪个应用进程来接收和处理这个数据包，即进行多路分解。而进行多路分解的依据就是数据包中携带的目的端口号。

多路复用是一个和图 3-5 相反的过程。假如图 3-5 中的 3 个进程发送数据包的目的 IP

地址相同，那么它们在网络中通常会按照相同的路径发送（当网络拓扑结构和流量等发生变化，导致路由表更新时除外）。通常，把这种同一主机上的不同进程将数据经传输层交付给网络层进行传输的过程称为多路复用。

3.3　TCP

传输控制协议（transmission control protocol，TCP）是 TCP/IP 体系结构提供的一种面向连接的传输层协议，提供基于字节流的可靠信息传送服务。

3.3.1　TCP 简介

TCP 用于提供有序可靠的面向连接的数据传输服务。与 UDP 相比，TCP 最大的特点是以牺牲效率为代价换取高可靠的服务。为了达到这种高可靠性，TCP 必须处理分组丢失、分组乱序以及由于时延而产生的重复数据报等问题。

TCP 与 UDP 不同的另一个特点是：为了能够独立于特定的网络，TCP 对报文长度有一个限定，即 TCP 传送的数据报长度要小于 64 KB。这样，长报文需要进行分段处理后才能传输。

在对上层数据进行处理时，TCP 与 UDP 是截然不同的。UDP 是面向报文流的，而 TCP 是面向字节流的，即 TCP 以字节作为最小处理单位，所有的控制都是基于字节进行的。例如，为了保证数据传输的可靠性，TCP 为字节流中的每一字节分配一个顺序号，并以此为基础，采用确认加超时重发的机制来保证可靠的数据传输。由于 TCP 是基于字节流的，因此对上层应用进程提交数据的边界无法保留，上层应用进程必须能够从 TCP 提交的字节流中确定不同报文间的边界。

TCP 不支持多播，即 TCP 只能解决两个进程间的通信问题。但是，TCP 支持同时建立多条连接。TCP 的连接服务采用全双工方式。在数据传输之前，TCP 必须在两个不同主机的传输端口之间建立一条连接，一旦连接建立成功，在两个进程间就建立了两条相反方向的数据传输通道，可同时在两个相反方向传输字节流。TCP 建立的端到端的连接是面向应用进程的，对中间节点（如路由器）是透明的。

图 3-6 给出了两个进程建立 TCP 连接时，数据的传输情况。图 3-6 中给出的只是其中一个方向的数据传输。在图 3-6 中需要特别注意的是，由于 TCP 是基于字节流的，在上层发送进程的应用数据到达 TCP 发送缓冲后，原始数据的边界将淹没在字节流中。TCP 在发送时，从发送缓冲中取一定数量的字节加上报头后组织成 TCP 报文进行发送。在到达接收方的接收缓冲时，TCP 报文携带的数据也将被作为字节流处理，在提交给应用进程时也是以字节流的形式提供。这时，接收进程必须能从这些字节流中划分出原始的数据边界。

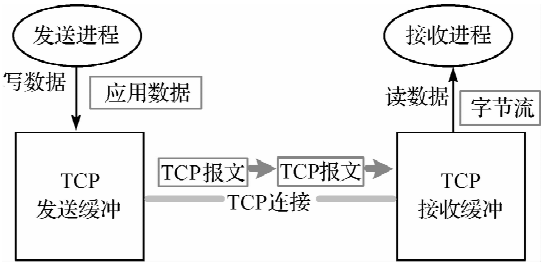


图 3-6　使用 TCP 连接进行数据传输

3.3.2　TCP 报文格式

与 UDP 报文一样，TCP 报文也是封装在 IP 分组中进行传输的。TCP 报头固定部分的长度为 20 字节，其具体格式如图 3-7 所示。

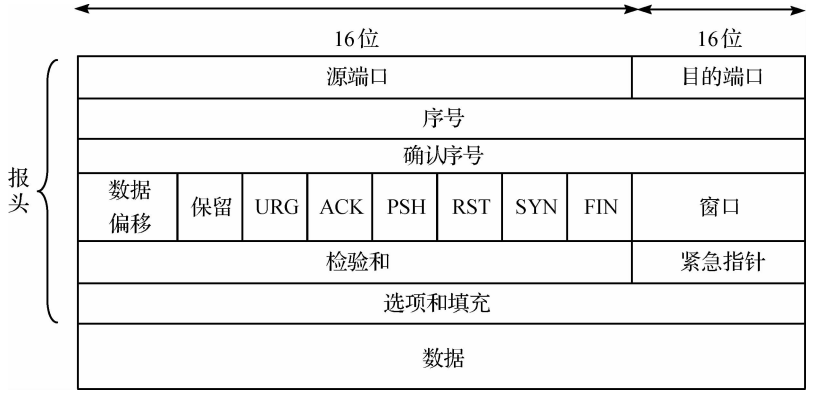


图 3-7　TCP 报文格式

TCP 报头包括以下字段。

1)源端口和目的端口

源端口和目的端口字段各占 16 位，分别标识连接两端的应用进程。

2)序号

序号字段占 32 位。TCP 的序号不是对每个 TCP 报文的编号，而是对每字节的编号。这样，序号字段是指该 TCP 报文中数据起始字节的序号。由于序号长度为 32 位，它可对 2^{32}(4G)字节进行编号。因此，序号重复时，旧序号数据早已在网络中消失。TCP 在连接建立时还采用了三次握手协议，确保不会把旧的序号当成新的序号。

3)确认序号

确认序号字段占 32 位，采用附载应答方式，指出下一个期望接收的字节序号，也就是告诉对方，这个序号以前的字节都已经正确收到。例如，确认序号 1024 表示序号为 1023 及其之前的字节都已经正确收到，期望收到的下一字节的序号为 1024。

4)数据偏移

数据偏移字段占 4 位，单位为 32 位(4 字节)，用以指明报文头部的总长度。这个字段的

出现是由于在报文头部中选项字段的长度是可变的。TCP 报头的最大长度为 60 字节。

5)保留

保留字段占 6 位,未使用。

6)标志位

标志位字段由 6 位组成,用于说明 TCP 字段的目的与内容。其中:

(1)URG:表示紧急指针字段有效。

(2)ACK:表示确认字段是否有效。当 ACK 为 0 时,表示确认字段无效。

(3)PSH:表示本 TCP 段请求一次推进(push),要求立即交予应用。

(4)RST:表示要求重建传输连接。

(5)SYN:同步位,表示要求建立连接,具体过程将在下面详细介绍。

(6)FIN:表示已经发送完所有字节,要求释放连接。

上述标志字段都是为 1 时有效。

7)窗口

窗口字段用于控制对方所能发送的数据量,单位为字节。在 TCP 流量控制中的作用参见 3.3.4 中的介绍。

8)校验和

校验和字段用于对 TCP 报文的首部和数据部分进行校验,与 UDP 类似的是校验和计算时也需要包含伪报头,TCP 伪报头的格式与 UDP 伪报头一样。

9)紧急指针

紧急指针字段用于指出窗口中紧急数据的位置,这些紧急数据应优先于其他数据进行传送。

10)选项

选项字段用于处理其他情况。目前被正式使用的选项字段有最大报文长度值(maximum segment size,MSS),它只能在连接建立时使用。

11)填充

因为选项字段的长度是不确定的,所以填充字段用于保证 TCP 报头的长度为 32 位的整数倍。

3.3.3　TCP 的连接与释放

TCP 的连接管理包括建立连接和释放连接。在 TCP 中,为了提高连接的可靠性,在连接的建立阶段采用三次握手协议;在连接的释放阶段采用对称释放方式,即连接的每端只能释放以自己为起点的那个方向的连接。

1)连接的建立

TCP 使用三次握手协议建立连接的过程如图 3-8 所示。在图 3-8 中,主机 A 是连接的发起方(一般为客户端),主机 B 为连接的响应方(一般为服务器端)。

首先,主机 A 发送一个 SYN 位被置 1 的连接请求报文,当主机 B 收到这个报文后,发送 SYN 和 ACK 位均被置 1 的报文来对第一个 SYN 报文段进行确认。最后,A 再发送一个

ACK 位置 1 的确认报文，用来通知主机 B 双方已完成连接的建立。

TCP 是建立在不可靠的 IP 分组传输服务上的，报文可能丢失、延迟、重复和乱序；并且，如果一个连接已经建立之后，某个延迟的连接请求才到达，就会出现问题。因此 TCP 建立连接所使用的三次握手协议还必须使用超时和重传机制。

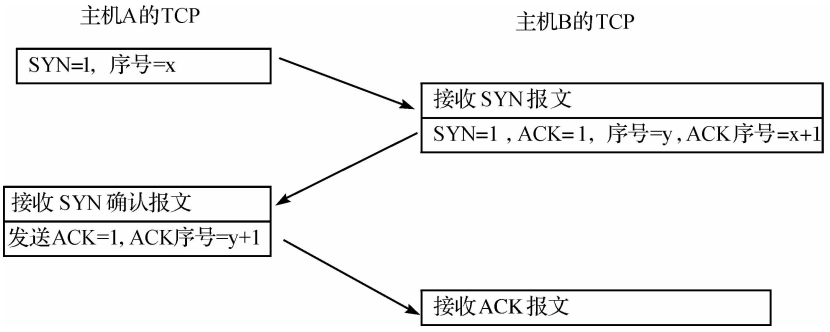


图 3-8　三次握手协议建立 TCP 连接

三次握手协议除了完成可靠连接的建立外，还使双方确认了各自的初始序号。从图 3-8 中可以看出，主机 A 在发送连接建立请求报文时，同时携带了序号 x；在主机 B 对连接请求进行响应时，一方面对主机 A 的起始序号 x 进行了确认（ACK 序号＝x＋1），另一方面也发送了自己的起始序号 y。最后，主机 A 在确认中携带了对主机 B 的起始序号 y 的确认（ACK 序号＝y＋1）。需要注意的是，第一次和第二次握手信号（SYN 报文）并不携带任何数据，但是需要消耗一个序号，下面介绍的 FIN 报文也是如此。

此外，为了达到最佳传输性能，TCP 在建立连接时还需要协商双方可接受的最大数据段长度（MSS）。当一个连接建立时，连接的双方都要通告各自的 MSS。当然，双方都希望 MSS 越大越好，报文段越大意味着允许每个报文段传送的数据越多，这样有效数据相对于 IP 首部和 TCP 首部就有更高的网络利用率。当 TCP 发送一个 SYN 报文时，它能将 MSS 值设置为外出接口的网络最大传送单元（maximum transmission unit，MTU）长度减去 IP 首部和 TCP 首部长度。对于以太网，MSS 值可达 1 460 字节。

如果目的地址为非本地的，MSS 通常采用默认值 536 字节。MSS 限制了接收方发送数据报的长度，由于主机也能控制它发送数据报的长度，所以可以使以较小 MTU 连接到网络上的主机避免分段。

2）连接的释放

TCP 连接是全双工的，可以看做是两个不同方向的独立数据流的传输。因此，TCP 采用对称的连接释放方式，即对每个方向的连接单独释放。如果一个应用程序通知 TCP 数据已经发送完毕，TCP 将单独关闭这个方向的连接。在关闭一个方向的连接时，连接释放的发起方在数据发送完毕后首先等待最后报文段的确认，然后发送一个 FIN 标志位置 1 的 TCP 报文，如图 3-9 所示。响应方的 TCP 进程对 FIN 报文段进行确认，并通知应用程序，整个通信会话已结束。一旦在某一个方向上的连接已关闭，TCP 将拒绝该方向上的数据。但是，在相反方向上还可以继续发送数据，直到这个方向的连接也被释放。尽管连接已经释放，确认信息还是会反馈给发送方。当连接的两个方向都已关闭，该连接的两个端点的 TCP 进程将删除这个连接记录。

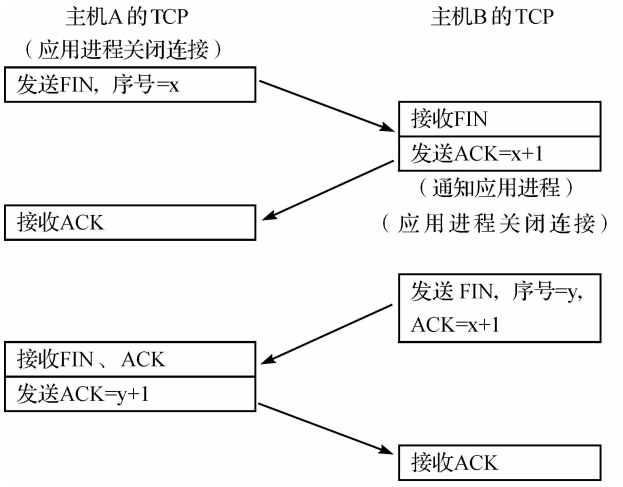


图 3-9　改进的三次握手完成 TCP 连接的释放

图 3-10 为 TCP 的状态变迁示意图。图 3-10 中的方框为 TCP 进程可能处在的状态，状态间的箭头表示可能发生的状态间的变迁。其中，粗实线箭头表示客户进程的变迁，粗虚线箭头表示服务器进程的变迁，细线箭头表示非典型的变迁。

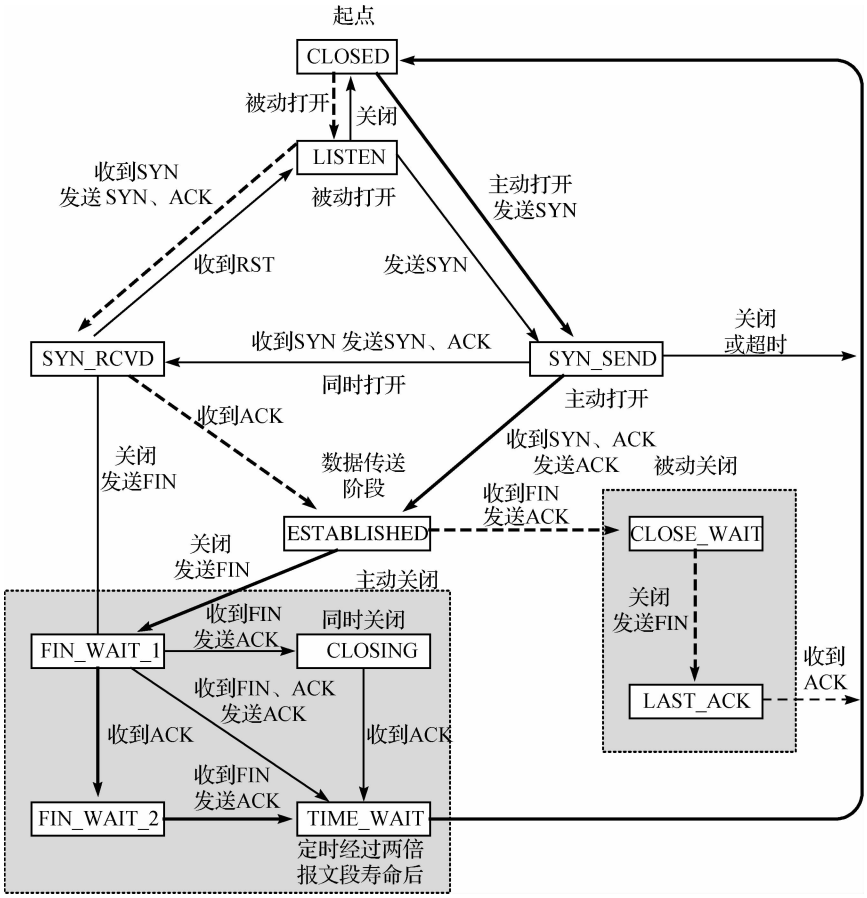


图 3-10　TCP 的状态变迁图

图 3-10 中的主要状态说明如下。

(1)CLOSED:连接关闭,即无可用连接。

(2)LISTEN:收到 SYN 报文。

(3)SYN_SEND:已经发送 SYN 报文,收到 ACK 报文。

(4)SYN_RCVD:已经发送 SYN+ACK 报文,收到 ACK 报文。

(5)ESTABLISHED:连接建立,数据传输进行中。

(6)FIN_WAIT_1:第一个 FIN 报文已经发送,等待 ACK 报文。

(7)FIN_WAIT_2:第一个 FIN 报文的 ACK 已经收到,等待第二个 FIN 报文。

(8)CLOSE_WAIT:收到第一个 FIN 报文,已经发送 ACK 报文,等待应用程序关闭。

(9)TIME_WAIT:收到第二个 FIN 报文,已经发送 ACK 报文,等待 2MSL 时间后可进入 CLOSED 状态。

(10)LAST_ACK:已经发送第二个 FIN 报文,等待 ACK 报文。

(11)CLOSING:双方决定同时关闭连接。

在图 3-10 中需要特别说明以下几点。

(1)TCP 的半关闭。TCP 提供了连接的一端在结束它的发送后还能接收来自另一端数据的能力,这就是 TCP 的半关闭状态。一端进程发送 FIN,并且另一端进程发送了对这个 FIN 的确认 ACK 报文段后进入半关闭状态。此时,连接的另一个方向还可以进行传输,并且在完成数据传送后,再发送 FIN 关闭这个方向的连接,从而将这个连接彻底关闭。

(2)2MSL 连接。TIME_WAIT 状态也称为 2MSL 等待状态。每个 TCP 必须选择一个报文段最大生存时间(maximum segment lifetime,MSL)。它是任何报文段被丢弃前在网络中的最长时间。当 TCP 进程执行了一个主动关闭,并发回最后一个 ACK 后,该连接在 TIME_WAIT 状态停留的时间必须为 2 倍的 MSL(记作 2MSL)。这样做的目的是让 TCP 可以再次发送最后的 ACK,以避免这个 ACK 丢失导致对方无法释放连接的情况(另一端超时并重发最后的 FIN)。

TCP 在重启的 MSL 时间内不能建立任何连接,即平静时间。

(3)FIN_WAIT_2 状态。在 FIN_WAIT_2 状态中一端 A 已经发出了 FIN,并且另一端 B 也对它进行了确认。只有 B 端的进程完成了这个关闭,A 端才会从 FIN_WAIT_2 状态进入 TIME_WAIT 状态。这意味着 A 端可能永远保持这个状态,B 端也将处于 CLOSE_WAIT 状态,并一直保持这个状态直到应用层决定关闭。

(4)同时打开。对于两端同时发出的连接建立请求,仅建立一条连接而不是两条连接。这时,两端几乎同时发送 SYN,并进入 SYN_SEND 状态。当每一端收到 SYN 时,状态变为 SYN_RCVD,并且它们都将再次发送 SYN 报文并对收到的 SYN 报文进行确认。当双方都收到 SYN 报文及相应的 ACK 报文时,双方状态都变为 ESTABLISHED。一个同时打开的连接需要交换 4 个报文段,比正常的三次握手多了一次。

(5)同时关闭。当应用层发出关闭命令,两端均从 ESTABLISHED 变为 FIN_WAIT _1。这将导致双方各发送一个 FIN,两个 FIN 经过网络传送后分别到达另一端。收到 FIN 后,状态由 FIN_WAIT_1 变为 CLOSING,并发送最后的 ACK。当收到最后的 ACK,状态变为 TIME_WAIT。

3.3.4　TCP 的传输控制

TCP 采用了基于字节流的传送方式，其基本特征是以字节为基本处理单位，不保留上层提交数据的边界。在前面的介绍中，我们已经知道 TCP 报文是按照字节编号、按照字节确认的。下面将介绍 TCP 传输控制的主要技术，包括字节流的封装及发送策略、流量控制及拥塞控制。

1)TCP 传输控制概述

在发送方，上层应用进程按照自己产生数据的规律，陆续将大小不等的数据块送到 TCP 的发送缓冲区中。在以下条件之一满足时，TCP 从缓存中取一定长度的字节流，封装成一个 TCP 报文段后发送。

(1)当缓冲区中数据的长度达到最大数据段长度 MSS 时，从缓冲区中取 MSS 长度的数据封装成 TCP 报文后发送。

(2)发送方应用进程要求立即发送报文，即要求 TCP 执行“推(push)”操作。

(3)当发送方的定时器超时时，也需要将缓冲区中数据封装成 TCP 报文，立即发送。

实际上，TCP 字节流的发送还需要遵循其他规则，如 Nagle 算法、流量控制和拥塞控制的策略等。下面先简单介绍一下在 TCP 实现中被广泛采用的 Nagle 算法。其基本思想是：当应用程序向传输实体传输数据时，传输实体封装并发出第一字节，对其后的所有字节进行缓存，直到收到对第一字节的确认；然后将已存入缓存的所有字节封装成数据报文发出，并继续将后续收到的字节存入缓存，直至收到下一个确认。这样，当数据到达速度较快而网络速度较慢时，可以明显减轻对网络带宽的消耗。Nagle 算法还规定，当到达的数据达到窗口大小的一半或者报文段的最大长度时，需要立即封装并发送一个报文段。

TCP 在发送中，还有可能会面临“傻窗口综合症(silly window syndrome)”，即当应用程序一次从传输层实体读出一字节时，传输层实体立即产生一个一字节的窗口更新段，使得发送方每次只能发送一字节，这大大降低了系统的性能。这个问题的解决办法是限制接收方只有在具备一半以上的空缓存或最大段长的空缓存时，才能产生一个窗口更新段。

对于每个 TCP 连接，TCP 管理着以下 4 个主要的定时器。

(1)重传定时器。TCP 为每一个发送的数据报文启动一个重传定时器，用于完成出错后的恢复。该定时器的时间间隔被称为重发超时(RTO)。TCP 超时和重传中最重要的就是对一个特定连接的往返时间(RTT)的测量。由于路由器和网络流量均会发生变化，因此 TCP 需要跟踪这些变化并相应地改变超时时间。TCP 必须测量在发送一个带有特别序号的字节和接收到包含该字节的确认之间的 RTT。

(2)坚持(persist)定时器。ACK 的传输并不可靠，也就是说，TCP 不对 ACK 报文段进行确认，TCP 只确认那些包含数据的 ACK 报文段。为了防止因为 ACK 报文段丢失而导致的双方进行等待的问题，发送方用一个坚持定时器来周期性地向接收方查询。

(3)保活(keepalive)定时器。用于检测一个空闲连接的另一端是否已经崩溃或重启。如果一个连接在指定的时间内没有任何动作，那么服务器就向客户发送一个探查报文段。

(4)2MSL 定时器。测量一个连接处于 TIME_WAIT 状态的时间。

2)TCP 的流量控制

TCP 的流量控制主要用于解决收发双方处理能力方面的不匹配问题,简单地说就是解决低处理能力(如慢速、小缓存等)的接收方无法处理过快到达的报文的问题。最简单的流量控制解决策略是接收方通知发送方自己的处理能力,然后发送方按照接收方的处理能力来发送。由于接收方的处理能力是在动态变化的,因此这种交互过程也是个动态的过程。

在 TCP 中采用动态缓存分配和可变大小的滑动窗口协议来实现流量控制。在 TCP 报文中的窗口字段就是用于双方交换接收窗口的大小。该窗口大小说明了接收方的接收能力(以字节为单位的缓冲区大小),发送方允许连续发送未应答的字节数量不能超过该窗口大小。图 3-11 给出了一个简单的 TCP 流量控制的例子。

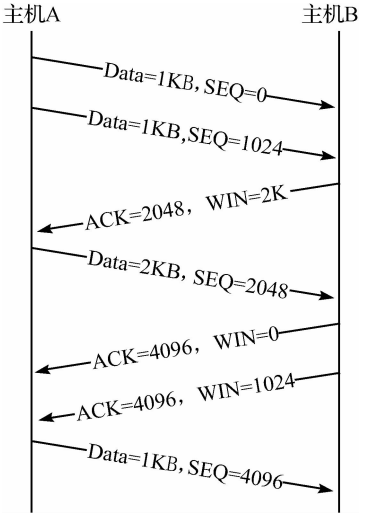


图 3-11　TCP 流量控制

在图 3-11 中,假设在初始时,主机 A 的发送窗口大小为 2 KB。这样,主机 A 在连续发送了两个 1 KB(1 024 B)的 TCP 报文后,必须停止发送等待主机 B 的应答。当主机 B 收到主机 A 发来的 2 KB 数据,将其处理完提交给上层实体后,发送应答;主机 B 在应答中再次声明了窗口大小为 2 KB。主机 A 按照主机 B 的应答更新自己的发送窗口大小,并继续发送。这次数据到达后,由于某种原因主机 B 的接收缓冲区并没有很快腾空,因此主机 B 给主机 A 发送的应答中将窗口大小声明为 0;收到这个应答后,主机 A 必须停止发送,直到收到主机 B 对窗口大小重新声明的应答。假如这个重新声明窗口大小的应答丢失了,就会造成灾难性的后果,主机 A 将再也不能发送数据了。为了避免进入这样一种死锁状态,TCP 规定在发送窗口大小为 0 时发送方仍可发送 1 字节的 TCP 段,这样接收方就可以重新声明确认号和窗口大小。

3)TCP 的拥塞控制

通信子网中传输的分组过多导致网络传输性能明显下降的现象称为拥塞。图 3-12 给出了拥塞引起网络性能下降的现象。当各主机输入通信子网的分组数量未超过网络能承受的最大能力时,所有分组都能正常传送,并且子网传送的分组数量与主机注入通信子网的分组数量成正比。但是当主机输入通信子网的分组数继续增大时,由于通信子网资源的限制,中间

节点会丢掉一些分组；如果通信子网传送的分组数继续增大，性能会变得更差，如递交给主机的分组数反而大大减少，响应时间急剧增加，网络反应迟钝，严重时还会导致死锁。为了最大限度地利用资源，网络工作在轻度拥塞状态时应该是较为理想的，但这也增加了导致拥塞崩溃的可能性，因此需要一定的拥塞控制机制来加以约束和限制。

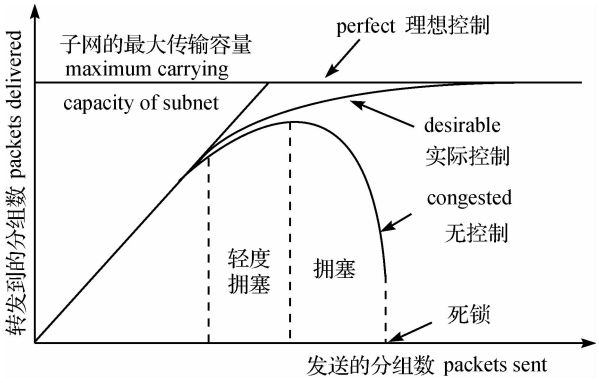


图 3-12　拥塞引起的性能下降

在网络中，通常使用缺乏缓冲区造成的丢包率、平均队列长度、超时重传包的数目、平均包延迟、包延迟变化(jitter)来衡量网络是否出现拥塞。在这些参数中，前两个参数是中间节点(路由器)用来监测拥塞的指标，后三个参数是源节点用来监测拥塞的指标。在 TCP 中通常选取丢包率作为判定拥塞的指标。

拥塞产生的本质是用户需求大于网络的传输能力，因此，解决拥塞主要有两类方法：增加网络资源和降低用户需求。增加网络资源一般通过动态配置网络资源来提高系统容量；降低用户需求通过拒绝服务、降低服务质量和调度来实现。由于拥塞的发生是随机的，网络很难做到在拥塞发生时增加资源，因此网络中主要采用降低用户需求的方式。

最初的 TCP 只有基于滑动窗口的流量控制机制而没有拥塞控制机制；1986 年初，Van Jacobson 提出了“慢启动”算法，后来这个算法与拥塞避免算法、快速重传和快速恢复算法共同用于解决 TCP 中的拥塞控制问题。TCP 拥塞控制的具体过程如图 3-13 所示。

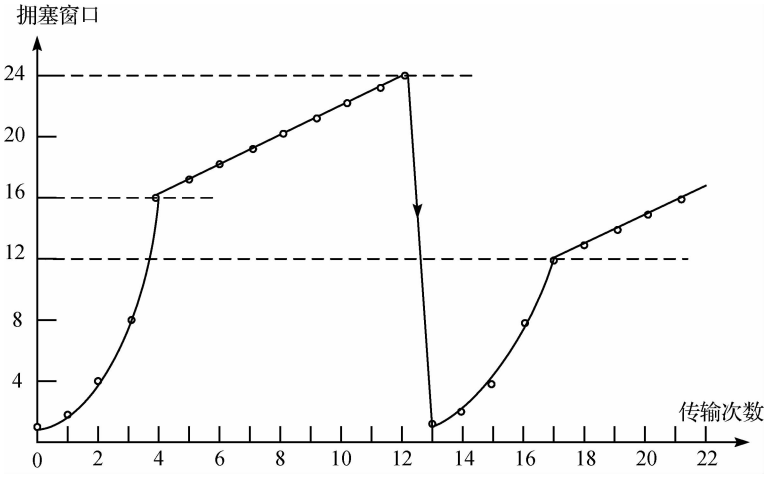


图 3-13　TCP 中的拥塞控制

图 3-13 中的 TCP 拥塞控制主要包括以下几个阶段。

(1)慢启动阶段(slow start)。发送方维护着两个窗口:接收方窗口(rwnd,最近从接收方收到的窗口大小)和拥塞窗口(cwnd),发送方按两个窗口的最小值发送。为了描述方便还引入了如下参数。

①MSS:MSS 是通信双方进行通信时的最大数据段。这个值以物理网络最大传送单元(MTU)等因素为基础,在连接开始时双方在 MSS 选项中说明。该大小不包括 TCP 头部、IP 头部以及选项字段。

②慢启动阈值(ssthresh):用来确定是使用慢启动还是拥塞避免算法来控制数据传送。ssthresh 的初始值可以任意大(如一些 TCP 实现中经常使用接收方窗口大小),但是作为对拥塞的响应,其大小可能会被减小。慢启动算法在 cwnd＜ssthresh 时使用;拥塞避免算法在 cwnd＞ssthresh 时使用。当 cwnd 和 ssthresh 相等时,发送端既可以使用慢启动算法也可以使用拥塞避免算法。

慢启动算法的基本思想是:由于在刚开始发送时并不了解网络当前状态,如果一个节点简单地按照发送窗口向网络中发送较多的报文,很容易引起网络拥塞;为此,发送过程采用由少到多的发送策略。当与另一个主机建立 TCP 连接时,拥塞窗口 cwnd 被初始化为一个 MSS。每收到一个应答报文 ACK,拥塞窗口就增加一个 MSS(注意:拥塞窗口以字节为单位,但慢启动算法以报文段大小为单位进行增加)。发送方取拥塞窗口和发送窗口中的最小值作为发送上限。简单地说,拥塞窗口是发送方使用的流量控制,而发送窗口则是接收方使用的流量控制。

发送方开始时发送一个报文段,然后等待 ACK。当收到该报文段的 ACK 后,拥塞窗口从 1 增加为 2,即可以发送两个报文段。当收到这两个报文段的 ACK 时,拥塞窗口就增加为 4,这是一种指数增加的关系,如图 3-13 所示。这样的指数增长一直持续到 cwnd≤ssthresh,随后进入拥塞避免阶段。

(2)拥塞避免阶段(congestion avoidance)。在拥塞避免期间,发送方的拥塞窗口 cwnd 以 1 个往返时间(round trip time,RTT)为单位线性增长,此阶段的增长与 ACK 的数量无关。拥塞避免算法继续保持拥塞窗口的增长直到检测到拥塞。

无论是在慢启动阶段还是在拥塞避免阶段,只要感知到网络中拥塞的出现,就需要立即停止增长。TCP 中的拥塞避免算法假定由于分组受到损坏引起的丢失是非常小的(远小于 1%),因此分组丢失就意味着在源主机和目的主机之间的某处网络上发生了拥塞。有两种情况表明分组丢失:发生超时和接收到重复的确认。例如,当发现超时或收到 3 个相同 ACK 确认帧时,表示发生了分组丢失,说明网络已发生拥塞现象,此时要进行相应的拥塞控制。首先,将 ssthresh 设置为发生拥塞时窗口值的一半,这个值取 rwnd 和 cwnd 的较小值,但不能小于 2。然后将拥塞窗口设置为 l,并重新进入慢启动阶段。

在图 3-13 的例子中,假定 rwnd 始终大于 cwnd,即发送方始终按照拥塞窗口的大小发送。为了简化研究,拥塞窗口采用 MSS 作为单位而不是实际中使用的字节。图 3-13 中的拥塞控制主要包括以下 3 个步骤。

①系统初始时 ssthresh＝16,并且 cwnd＝1。

②在开始时,发送方采用慢启动算法控制发送,因此 cwnd 按照指数增长。当 cwnd 增长到阈值 ssthresh(16)时,转而执行拥塞避免算法,拥塞窗口按照线性增长。

③当拥塞窗口增长到 24 时,网络开始丢包,表示出现了拥塞;这时,系统取值出现如下

变化：阈值 ssthresh＝12（即当前拥塞窗口 24 的一半），cwnd＝1（拥塞窗口被重新设置为 1），随后重新进入慢启动阶段。

（3）快速重传（fast retransmit）阶段和快速恢复（fast recovery）阶段。上述的慢启动算法和拥塞避免算法是 TCP 拥塞控制的基本策略，后来为了解决等待重传计时器超时而引起的信道空闲问题，又引入了快速重传和快速恢复两个拥塞控制算法。

当一个次序紊乱的数据报文到达 TCP 接收方时，接收方应该立即发送一个重复的 ACK 应答。这个 ACK 的目的是通知发送方收到了一个次序紊乱的数据报文，以及重申接收方期望的序号。从发送方的观点来看，重复的 ACK 可能是由许多网络问题引起的。首先，可能是数据报文的丢失引起的。在这种情况下，所有在丢失的数据报文之后发送的数据报文都将引起重复的 ACK。其次，可能是由网络对数据的重新排序引起的。最后，重复的 ACK 可能是由网络对 ACK 或数据报文的复制引起的。另外，当接收到的数据报文填补了全部或部分序列号间隔时，TCP 接收方应该立即发送一个新的 ACK，这将避免发送方的重传定时器超时。

当收到重复的 ACK 时，TCP 发送方使用快速重传算法来探测或者修复数据的丢失。快速重传算法以 3 个重复的 ACK 到达（即总共收到 4 个相同的 ACK，其间没有任何其他 ACK 报文到达）为一个数据段已经丢失的标志。在收到 3 个重复 ACK 之后，TCP 不等重传定时器超时就重传看来已经丢失的数据段。

在快速重传算法发送了看来已经丢失的数据报文后，快速恢复算法用来控制新数据报文的传送，直到一个非重复的 ACK 到达。不进行慢启动的原因是收到重复的 ACK 不仅意味着一个数据报文已经丢失，而且意味着接收方收到了一个后续的数据报文。

下面简单说明快速重传和快速恢复算法的工作过程。

①当收到第三个重复 ACK 时，将 ssthresh 设置为当前拥塞窗口的一半。

②重传丢失的数据段并设置 cwnd 的值为 ssthresh。也有的快速重传方案是将 cwnd 的值设置为 ssthresh＋3，这是人为地按已经离开网络的报文段数目（3 个）和接收端缓冲数据量来扩充拥塞窗口。

采用了快速重传和快速恢复机制后，图 3-13 就变为图 3-14。但是，同一个数据段在传送期间多次丢失时，这个算法就不能有效地恢复了。

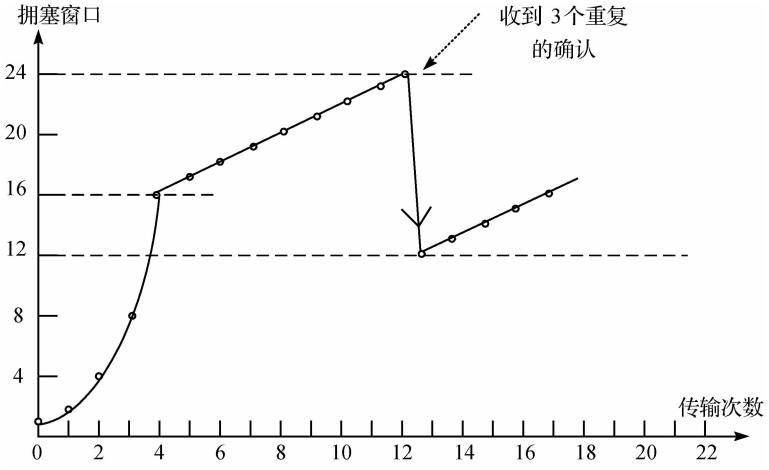


图 3-14　快速重传和快速恢复

3.4 套接字编程

应用层实体通过传输层进行数据传输时，TCP 和 UDP 会遇到同时为多个应用程序进程提供并发服务的问题。例如，多个 TCP 连接或多个应用程序进程可能需要通过同一个 TCP 端口传输数据。为了区别不同的应用程序进程和连接，操作系统通常会采用被称为套接字(Socket)的接口方式。

3.4.1 套接字概述

Socket 是从 UNIX 系统中的 I/O 命令集发展起来的，其基本模式是打开—读/写—关闭(open write/read close)。即在一个用户进程进行 I/O 操作时，它首先调用“打开”操作以获得对指定文件或设备的使用权，并返回称为文件描述符的整型数，然后这个用户进程多次调用“读/写”操作来完成数据的传输。当所有的传输操作完成后，用户进程关闭调用，通知操作系统已经完成了对某对象的使用。

Socket 取自英文原意中的“插座”，作为不同系统进程间的通信机制，Socket 的基本思想是为上层实体提供一种透明地访问网络的能力。从本质上说，Socket 是一组传输层的服务原语。Socket 在网络体系结构中的位置如图 3-15 所示。

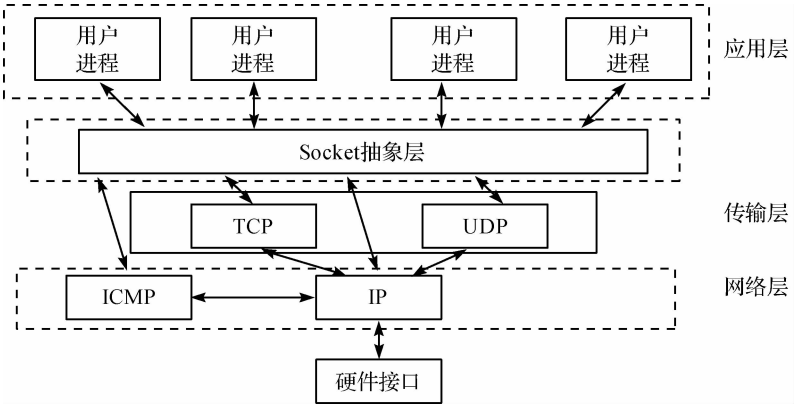


图 3-15 Socket 在网络体系结构中的位置

1)通信端口的标识

在网络通信中，参与通信的两个进程通常位于不同的机器上，为了能够找到对方，需要进行统一的编址。在 Internet 中，两台机器还可能位于不同的网络中，这些网络通过网络互连设备(如路由器)连接。因此需要如下的三级寻址结构。

(1)某一主机可与多个网络相连，主机必须指定网络地址(在 Internet 中为 IP 地址)。

(2)网络上每一台主机具有唯一的地址(物理地址/MAC 地址)。

(3)每一主机上的每一进程具有该主机上的唯一标识符(端口号)。

在 Internet 中，当 IP 分组到达目的节点所在的子网后，可通过 ARP 完成目的 IP 地址到目的物理地址的转换，因此主机地址可以简单地由 IP 地址和端口号来标识。由于在 Internet中存在两种服务方式不同的协议——TCP 和 UDP，因此在两个网络进程的通信过程中，还需要指明所采用的协议类型。综上所述，网络中用一个如下形式的三元组就可以全局唯一地标识一个进程。

（协议，本地地址，本地端口号）

这样的三元组被称为一个半相关（half-association），它指定一个连接中的一个进程；而全相关是指一个完整的进程通信需要由两个进程组成，并且这两个进程只能使用同一种高层协议。这样一个相关可以表示成如下的五元组形式。

（协议，本地地址，本地端口号，远程地址，远程端口号）

2）套接字的类型

在采用 TCP/IP 体系结构的 Internet 中，提供以下 3 种类型的套接字。

（1）流式套接字（SOCK_STREAM）。提供了一个面向连接、可靠的数据传输服务，数据无差错、无重复地发送，且按发送顺序接收。这种套接字主要是针对 TCP 设计的。

（2）数据报式套接字（SOCK_DGRAM）。提供了一个无连接服务。数据包以独立包的形式被发送，不提供无错保证，数据可能丢失或重复，并且接收顺序混乱。这种数据报式套接字主要是针对 UDP 设计的。

（3）原始式套接字（SOCK_RAW）。原始式套接字主要针对那些直接使用 IP 层服务的协议而设计的，如 ICMP、OSPF 等协议。

3）套接字提供的系统调用

在 Socket 实现中主要采用客户机/服务器模型。下面是 Socket 中提供的主要系统调用，需要注意的是，这些系统调用有些是专门针对面向连接的 TCP 而设计的。

（1）socket()。应用程序在使用套接字前，必须先建立一个套接字；系统调用 socket()向应用程序提供创建套接字的手段，这个系统调用的执行结果是返回一个套接字，并为其分配相应的资源，同时返回一个整型套接字号。

（2）bind()。当一个套接字用 socket()创建后，使用 bind()将套接字所使用的地址（包括 IP 地址和端口地址）与所创建的套接字号联系起来，即指定本地半相关。

（3）listen()。这个系统调用是针对面向连接的 TCP 设计的，表示服务器进程已经处于就绪状态，随时可以响应来自客户端的连接建立请求。

（4）connect()。这个系统调用是针对面向连接的 TCP 设计的，用于向远程进程（通常为服务器进程）发出连接建立请求，这个调用通常是由客户端发起的。

（5）accept()。accept()适用于面向连接的 TCP，用于响应请求连接队列上的第一个客户。accept()调用执行后创建一个与原套接字有相同特性的新套接字，新的套接字可用于处理服务器并发请求。

通过 socket()、bind()、connect()、accept()4 个套接字系统调用，可以完成一个五元组的建立。socket()指定五元组中的协议元，它的用法与是否为客户或服务器、是否面向连接无关。bind()指定五元组中的本地二元，即本地主机地址和端口号，其用法与是否面向连接有关：在服务器方，无论是否面向连接，均要调用 bind()；在客户方，若采用面向连接，则可以不调用 bind()，而通过 connect()自动完成。若采用无连接，客户方必须使用 bind()以获得

一个唯一的地址。

(6)read()与 write()。当一个连接建立以后,就可以传输数据了。常用的系统调用有 read()和 write()。read()用于在指定的数据报或流套接字上发送输出数据,write()用于从指定的数据报或流套接字上接收数据。

(7)close()。close()用于关闭指定的套接字,并释放分配给该套接字的资源;如果涉及一个打开的 TCP 连接,则该连接被释放。

3.4.2 TCP 套接字编程

TCP 采用面向连接的客户机/服务器模式,因此其套接字编程分为客户端和服务器端两部分。TCP 使用套接字建立连接和完成数据传输的过程如图 3-16 所示。

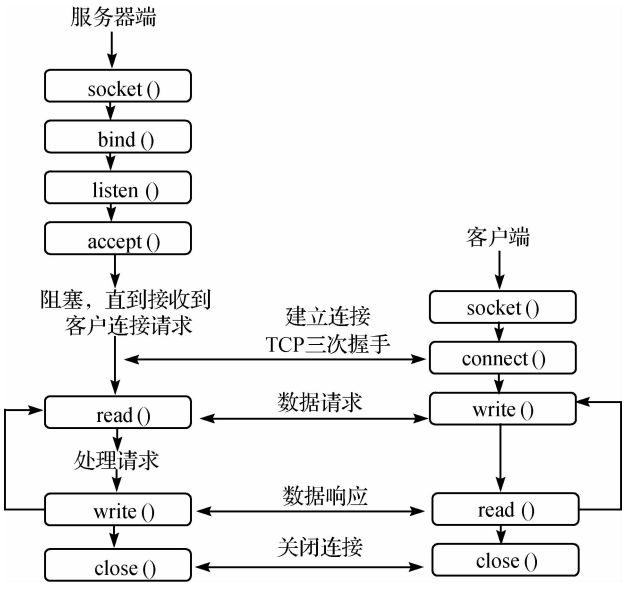


图 3-16 TCP 使用套接字建立连接和完成数据传输的过程

服务器端的主要工作步骤如下。

(1)使用 socket()创建套接字,并确定所使用的协议。

(2)使用 bind()实现与本地 IP 地址和端口号的绑定。

(3)使用 listen()使服务器套接字做好接收连接请求的准备。

(4)使用 accept()等待接收来自客户端由 connect()发出的连接请求。

(5)根据连接请求建立连接后,使用 write()发送数据,或者使用 read()接收数据。

(6)数据传输结束后,使用 close()关闭套接字。

客户端的主要工作步骤如下。

(1)使用 socket()创建套接字,并确定所使用的协议。

(2)使用 bind()实现与本地 IP 和端口号的绑定。

(3)使用 connect()发出与服务器建立连接的请求。

(4)连接建立后使用 write()发送数据,或使用 read()接收数据。

(5)使用 closes()关闭套接字。

TCP 服务器端的 socket 编程模板示例如下。

```
int main(void)
{
    int sockfd,connect_sock;
    if((sockfd=socket(AF_INET,SOCK_STREAM,0))==-1) {
        perror("create socket failed.");
        exit(-1);
    }
    /* 绑定套接字的地址 */
    /* 进入监听状态 */
    …
    loop{

    if((connect_sock=accept(sockfd,NULL,NULL))==-1){
        perror("Accept error.");
        exit(-1);
    }
    /* 读并处理请求 */
        close(connect_sock);
    }
    close(sockfd);
}
```

TCP 客户端的 socket 编程模板示例如下。

```
/* include some header files */
int main(void)
{
    int sockfd;
    if((sockfd=socket(AF_INET,SOCK_STREAM,0))==-1)
    {
        perror("Create socket failed.");
        exit(-1);
    }
    /* 连接到服务器 */
    …
    /* 发送请求并接收响应 */
    …
    close(sockfd);
}
```

3.4.3 UDP 套接字编程

UDP 采用无连接的客户机/服务器模式，其套接字编程中客户端与服务器端的工作过程类似。图 3-17 为 UDP 使用套接字进行数据传输的过程。

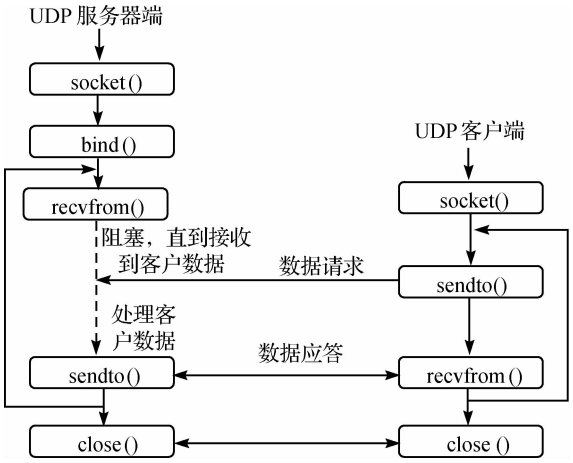


图 3-17 UDP 使用套接字进行数据传输的过程

服务器(客户)端的工作步骤如下。

(1)使用 socket()创建套接字，并确定所使用的协议。

(2)使用 bind()实现与本地 IP 地址和端口号的绑定。

(3)使用 sendto()发送数据，或者使用 recvfrom()接收数据。

(4)数据传输结束后，使用 close()关闭套接字。

UDP 套接字编程虽然也采用客户机/服务器模式，但与 TCP 套接字编程有以下主要不同。

(1)UDP 是无连接的，所以可以发送或接收长度为 0 的数据(仅包含 IP 头部和 UDP 头部)，这与 TCP 不同，TCP 的 read()返回 0 表示对方已关闭了连接。

(2)大多数 TCP 服务器是并发的，而大多数 UDP 服务器是迭代的。

(3)TCP 中有多少个客户连接，就有多少个接收缓冲区；而 UDP 只有一个缓冲区来存放所有接收到的数据。

(4)对于 TCP 客户，调用 connect()时分配端口，而对 UDP 套接字，如果进程调用 sendto()时还没有绑定本地端口，内核就选择临时端口。

3.5 IP

IP 是 TCP/IP 体系结构中网络层的核心协议，它提供无连接的数据报传送机制。IP 实现上非常简单，它对数据提供“尽力而为服务”，即它不能保证传输的可靠性，只负责将分组

送到目的节点，至于传输是否正确，不进行验证，不发确认，也不保证分组的正确顺序，而是将可靠性工作交给传输层处理。例如，如果应用层要求较高的可靠性，可在传输层使用 TCP。

简单地说，IP 主要完成无连接的数据报传输、数据报路由（IP 路由）、分组的分段和重组工作。

3.5.1　IP 地址

由于不同的物理网络有不同的编址方式，不同物理网络中的主机有不同的物理地址。因此，为了做到不同物理结构网络的互连和互通，要解决的首要问题就是统一编址，即在互联网上采用全局统一的地址格式，为每一个子网、每一个主机分配一个全网唯一的地址。IP 地址就是 IP 为此而制定的。同时，IP 地址是按照网络的位置分配的，因此通过 IP 地址很容易定位到该 IP 地址所绑定的主机所在的物理网络。

由网络信息中心（Network Information Center，NIC）统一负责全球 IP 地址的规划、管理，由其下属机构 Inter NIC、APNIC、RIPE 等网络信息中心具体负责美国及全球其他地区的 IP 地址分配。我国申请 IP 地址要通过亚太互联网络信息中心（APNIC）。

1）IP 地址的结构

IP 地址由一个 4 字节（32 位）的数字组成，包括网络号和主机号两部分。其中网络号的长度决定了整个网络中可包含的子网数，主机号的长度决定了每个子网能容纳多少台主机。4 字节的 IP 地址通常以小圆点（.）分隔字节，并且每字节都用十进制表示，如 106. 106. 71. 1。

2）IP 地址的分类

IP 地址分为 A 类、B 类、C 类、D 类和 E 类共 5 类。由于 D 类地址分配给多播，E 类地址保留，所以实际可分配的 IP 地址只有 A 类、B 类和 C 类，如图 3-18 所示。

图 3-18　3 类可分配的 IP 地址

A 类地址由最高位的 0 标志、7 位的网络号和 24 位的网内主机号组成。这样，在一个互联网中最多有 126 个 A 类网络（网络号 1～126，号码 0 和 127 保留）。而每一个 A 类网络允许有约 1 600 万台主机。A 类网络一般用于网络规模非常大的地区网。

B 类地址由最高两位的 10 标志、14 位的网络号和 16 位的网内主机号组成。这样，在互联网中大约有 16 000 个 B 类网络，而每一个 B 类网络可以有 65 000 多台主机。B 类网络一般用于较大规模的单位和公司。

C 类地址由最高 3 位的 110 标志、21 位的网络号和 8 位的网内主机号组成。一个互联

网中允许包含约 200 万个 C 类网络，而每一个 C 类网络中最多可有 254 台主机（主机号全 0 和全 1 有特殊含义，不能分配给主机）。C 类网络一般用于较小的单位和公司。

NIC 对 IP 地址还有如下规定。

(1)主机号全为 1：用做指定网络的广播地址。

(2)主机号全为 0：表示网络本身。

(3)网络号全为 0：表示本网络。

(4)32 位全为 1：用于本网广播，该地址又称为有限广播地址。

(5)第一字节为 127：表示回送地址（loopback address），用于网络软件测试以及本地机进程间通信。无论什么程序，一旦使用回送地址发送数据，协议软件立即将其回送，不进行任何网络传输。最常用的回送地址是 127.0.0.1。

此外，NIC 还为每类地址都保留了一个地址段用做私有地址（private address），私有地址属于非注册地址，专门为组织机构内部使用。

3 类地址保留的私有地址范围如下。

A 类：10.0.0.0～10.255.255.255。

B 类：172.16.0.0～172.31.255.255。

C 类：192.168.0.0～192.168.255.255。

这些私有地址主要用于企业内部网络中。私有网络由于不与外部互连，因此可以使用任意的 IP 地址。保留这样的地址是为了避免以后接入 Internet 时引起地址混乱。使用私有地址的私有网络在接入 Internet 时，要使用地址翻译协议将私有地址翻译成公用合法 IP 地址。在 Internet 中，这类私有地址是不能出现的。

3)IP 地址的分配

如果用户希望使用 TCP/IP 体系结构来组建企业内部网络，需要首先为网络上的所有主机分配 IP 地址。在分配 IP 地址时按网络规模的大小，使用上述各类私有地址，并且在分配 IP 地址时还必须满足下面两个条件。

(1)每个主机 IP 地址的网络号部分必须相同。

(2)网络上每个主机的 IP 地址必须唯一。

如果用户希望将主机或网络加入到 Internet 中，则必须向 NIC 或其下属机构申请 IP 地址。

使用 A 类或 B 类 IP 地址的网络，由于网络规模比较大，为了方便管理还可以把网络划分成若干子网，并分配网络内部的子网号和主机号。这时，4 字节的 IP 地址就被分成了以下 3 个部分。

IP 地址＝网络地址＋子网地址＋主机地址

例如，一个拥有 B 类地址的网络，NIC 为其分配了 2 字节的网络号。另外 2 字节的主机号用户可自由分配。这时可根据子网的数目和子网中的最大主机数进行再分配，若最多有 6 个子网，则可将后两字节的前 3 位作为子网地址，后 13 位作为主机地址。

4)子网掩码

划分子网后，网络地址和子网地址构成了真正的网络地址部分，每个子网看起来就像一个独立的网络，而对于远程主机，这种子网的划分是透明的。在进行子网划分后，为了能确

定 IP 地址中的网络号部分，引入了子网掩码的概念。子网掩码不能单独存在，它必须与 IP 地址一起使用，并采用和 IP 地址相同的格式。简单地说，子网掩码的作用就是说明与其相关的 IP 地址的哪部分是网络地址。

子网掩码由 n 位连续的 1 和 32－n 位连续的 0 共 32 位组成，用于说明该子网掩码所说明的 IP 地址前 n 位为网络地址，后 32－n 位为主机地址。例如，标准 A、B、C 三类地址的子网掩码如图 3-19 所示。

图 3-19　标准 A、B、C 三类地址的子网掩码

例如，一个网络分配到了一个 B 类地址：130.1.0.0，该网络的管理员为了方便管理把整个网络划分成 12 个子网。这样，由于 $2^3<12<2^4$，需要用 B 类地址后两字节中的前 4 位表示子网号，最后 12 位表示主机号。这样划分后实际上获得了 $2^4=16$ 个子网，这些子网的地址分别为 130.1.0.0、130.1.16.0、130.1.32.0、130.1.48.0、…、130.1.240.0。并且该网络所有子网的子网掩码由原来的 255.255.0.0 变为 255.255.240.0（11111111.11111111.11110000.00000000）。前面已经介绍了 IP 网络中主机号部分全 0 和全 1 是有特殊含义的，因此不能分配给主机，这样每个子网可分配的 IP 地址为 $2^{12}-2$ 个。

主机之间要想能够直接通信，它们必须在同一子网内，否则需要通过路由器（或者网关）进行转发。因此，每台主机在发送数据之前，必须计算自己的 IP 地址与目的 IP 地址的网络号是否相同。通过计算“子网掩码∧IP 地址”可获得 IP 地址的网络号。例如，当 IP 地址为 202.117.1.207，子网掩码为 255.255.255.224 时，通过下式计算，可得子网地址为 202.117.1.192。

```
          11001010   01110101   00000001   110 01111
∧         11111111   11111111   11111111   111 00000
-------------------------------------------------------
          11001010   01110101   00000001   110 00000
```

3.5.2　IP 分组及其转发

使用 TCP/IP 体系结构的 Internet 中传输的基本数据单元是 IP 分组，通过 IP 分组来进行不可靠、无连接的数据传输，是 IP 的具体体现。

1）IP 分组的格式

IP 分组由分组头和有效数据两部分组成。其中，分组头用来存放 IP 的具体控制信息，数据区包含了上层协议（如 TCP）提交给 IP 传送的数据。整个 IP 分组的长度是 4 字节的整数倍，如图 3-20 所示。

图 3-20　IP 分组格式

其中，IP 分组头部分由以下字段组成。

(1)版本字段：长度为 4 比特，表示与 IP 分组对应的 IP 版本号。在处理 IP 分组前，IP 模块都要首先检查 IP 分组的版本字段，以保证 IP 分组格式与 IP 模块的一致。目前 Internet 上广泛采用的是版本 4 的 IP 协议，即 IPv4。

(2)分组头长字段：长度为 4 比特，用于指明 IP 分组头的长度，以 4 字节为 1 个单位。由于包含任选项字段，IP 分组头长度是可变的。

(3)服务类型字段：长度为 8 比特，用于指明 IP 分组所希望得到的有关优先级、可靠性、吞吐量及时延等方面的服务质量要求。由于大多数路由器不处理这个字段，因此这个字段形同虚设。

(4)总长度字段：长度为 16 比特，用于指明 IP 分组的总长度，单位是字节，包括分组头和数据区的长度。由于总长度字段为 16 比特，因此 IP 分组最多允许有 2^{16}(65 535)字节。

(5)标识符字段：长度为 16 比特，用于唯一标识一个 IP 分组。标识符字段是 IP 分组在传输中进行分段和重组所必需的。

(6)标志字段：长度为 3 比特，其中 1 位保留，另两位为 DF 和 MF，DF 用于指明 IP 分组是否允许分段，MF 用于表明是否有后续分段。

(7)偏移量字段：长度为 13 比特，以 8 字节为 1 个单位，用于指明当前 IP 分组中的数据在原始 IP 分组中的位置，这是分段和重组所必需的。

(8)生存期字段：长度为 8 比特，用于指明 IP 分组在网络中可以传输的最长“距离”，每经过一个路由器时 TTL 字段减 1，当减到 0 值时，该 IP 分组将被丢弃。这个字段用于保证 IP 分组不会在网络出错时无休止地传输。

(9)协议类型字段：长度为 8 比特，用于指明调用 IP 进行传输的高层协议，高层协议的编码由 TCP/IP 体系结构的管理机构统一分配。例如，ICMP 的值为 1(十进制，以下同)，TCP 的值为 6，UDP 的值为 17。

(10)校验和字段：长度为 16 比特，用于保证 IP 分组头的完整性。其算法的基本思想为：校验和字段的初值为 0，对 IP 分组头以每 16 位为 1 个单位进行求异或，再将结果求反，得到校验和。

(11)源地址字段：长度为 32 比特，用于指明发送 IP 分组的源主机的 IP 地址。

(12)目的地址字段:长度为 32 比特,用于指明接收 IP 分组的目标主机的 IP 地址。

(13)选项字段:长度可变,该字段主要用于以后对 IP 的扩展。该字段的使用有一些特殊的规定,表 3-3 给出了一些常用的选项。

表 3-3　IP 分组选项

选项类型	描　　述
安全选项	表示该分组的保密级别
严格源路由选项	由源给出完整的路由列表
宽松源路由选项	由源给出必须经历的路由列表
路由记录	让每个路由器在 IP 分组中记录其 IP 地址
时间戳	让每个路由器在 IP 分组中记录其 IP 地址和经历的时间

(14)填充字段:长度不定,由于 IP 分组头必须是 4 字节的整数倍,因此当使用选项的 IP 分组头长度不足 4 字节的整数倍时,必须用 0 填入填充字段来满足这一要求。

2)IP 分组的分段及重组

下面简单说明 IP 分组在转发过程中分段和重组的过程。IP 分组的分段和重组主要涉及标识符字段、标志字段、偏移量字段。当 IP 分组所经历的物理网络的最大传输单元(MTU)比分组长度小时,IP 需要把该 IP 分组分割成若干满足 MTU 长度要求的更小的 IP 分组(称为分段)后,再进行发送。由于偏移量字段的单位为 8 字节,因此除最后一个分段外,前面所有分段的长度必须为 8 字节的整数倍,且一般都取相同的长度。分段后的每个分段都是一个完整的 IP 分组,其 IP 分组头除片偏移、标志字段中的 MF 标志位、总长度和校验和字段外,其他与原始 IP 分组头相同。重组是分段的反过程,根据片偏移和 MF 标志判断是否进行了分段。如果 MF=0 并且 Offset =0则为一个完整分组;如果 MF=1 并且 Offset≠0 则表示进行了分段,需要在目的节点进行重组。

例如,在以太网中 MTU 为 1 500 字节,一个长度为 4 000 字节的 IP 分组进入以太网后,按照如图 3-21 所示进行分段。

	长度	标识符	分段标志	偏移量	…
	=4 000	=X	=0	=0	

上述IP分组被分解为以下3个IP分组

	长度	标识符	分段标志	偏移量	…
	=1 500	=X	=1	=0	
	长度	标识符	分段标志	偏移量	…
	=1 500	=X	=1	=185	
	长度	标识符	分段标志	偏移量	…
	=1 040	=X	=0	=370	

图 3-21　IP 分组分段过程

3)IP 分组的转发

在 Internet 中,IP 分组的转发具有如下特点。

(1)每个 IP 分组包含目的主机的 IP 地址。

(2)IP 地址中的网络地址唯一标识 Internet 中的一个物理网络。

(3)所有连接到相同物理网络的主机和路由器共享其地址中的网络地址,它们在该物理网络内可以直接通信。

(4)Internet 中的每个物理网络至少有一个与之相连的路由器,通过路由器和其他物理网络相连,路由器负责在该物理网络和 Internet 上其他物理网络间转发分组。

(5)路由器根据分组携带的目的 IP 地址进行路由转发。路由器中的路由表格式如表 3-4 所示。在路由表中,目的网络通常使用 IP 地址和子网掩码的形式来描述。

表 3-4　路由表示例

目的网络	下一路由器	距　离
20.0.0.0	直接投递	0
30.0.0.0	直接投递	0
10.0.0.0	20.0.0.5	1
40.0.0.0	30.0.0.7	1

一个简单的两个子网(子网 1:202.1.64.0 和子网 2:202.1.61.0)通过路由器连接的例子如图 3-22 所示。在图 3-22 中,路由器的两个接口 202.1.64.5 和 202.1.61.6 分别与子网 1 和子网 2 相连,路由器在两个子网间完成数据转发。

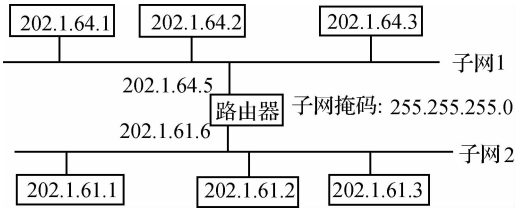


图 3-22　子网互连示例

3.5.3　与 IP 地址相关的一组协议

前面已经介绍了 Internet 上的所有主机都有两个地址:一个是主机所在网络的固有物理地址,另一个是 IP 为全网范围内每个主机分配的唯一的 IP 地址。由于最终通信总是由物理网络实现的,即每个主机的物理接口(通常为网卡)并不能识别 IP 地址,而源主机在发送 IP 分组时携带的是目的主机的 IP 地址,因此必须实现 IP 地址与物理地址间的转换。此外,对于存在无盘工作站的网络,IP 地址的获取也是必须解决的问题。为此,TCP/IP 体系结构在网络层提供了一组协议来解决地址转换等问题。

1)ARP

地址转换协议(address resolution protocol,ARP)用于将 IP 地址转换成物理地址。虽然 ARP 包含在 TCP/IP 体系结构的网际层,但它实际上是一个低层协议,它使网际层与硬件及数据链路层隔离,数据链路层可直接使用 ARP。目前,ARP 已支持很多硬件和协议类

型(如 IEEE 802 系列局域网等)。

ARP 的工作原理是:在每台使用 ARP 的主机中都存放一张 ARP 表。表 3-5 给出了一张 ARP 表的样例。

表 3-5　ARP 表样例

IP 地 址	以太网 MAC 地址
138.129.32.3	08-60-8c-42-29-93
138.129.32.4	08-60-8c-2f-d2-2b
138.129.32.5	08-60-2d-2-91-e3

当主机收到一个 IP 分组时,只要根据 IP 分组携带的目的 IP 地址去查找内存中的 ARP 表,即可将 IP 地址转换为物理地址。现在的问题是这张表是如何产生的呢? 首先应明确的是,这张 ARP 表不是一张静态的表.它是动态生成的。为提高查找的速度,ARP 表保存在内存中,因此在主机初启时,这张 ARP 表是空的,以后每当有一个不在 ARP 表中的 IP 地址到来时,ARP 将负责自动填入一条记录。ARP 表中的每条记录都关联着一个计时器,计时器超时后,该记录将被删除,这样做的目的是保证 IP 地址和物理地址的映射关系始终是最新的。

下面以以太网为例来说明 ARP 表的生成过程。当在 ARP 表中找不到某一目的 IP 地址时,ARP 就使用以太网广播地址,发送一个 ARP 请求报文给子网中的每一台计算机。子网中每个计算机的以太网接口(通常为网卡)都会收到这个以太网广播帧,并判断请求报文中携带的目的 IP 地址是否与自己的 IP 地址相同,如果相同则将自己的物理地址填入一个响应报文中,并传送给发送请求报文的主机,该主机将得到的目的主机的 IP 地址和以太网地址加入 ARP 表中。若没收到响应,则表示目的主机不存在,则本地 IP 模块将抛弃发送到这个目的地址的 IP 分组。

将 IP 地址转换到以太网地址的 ARP 报文格式如图 3-23 所示。其中,硬件类型字段指明了发送方想知道的硬件接口类型,以太网的值为 1。协议类型字段指明了发送方提供的高层协议类型,IP 为 0806(十六进制)。硬件地址长度和协议地址长度指明了硬件地址和高层协议地址的长度,这样 ARP 报文就可以在任意硬件和任意协议的网络中使用。操作字段用来表示发送这个报文的目的,ARP 请求为 1,ARP 响应为 2,RARP 请求为 3,RARP 响应为 4。当发出 ARP 请求时,发送方填好发送方首部和发送方的 IP 地址,还要填写目的 IP 地址。当目的主机收到这个 ARP 广播包时,就会在响应报文中填上自己的 48 位主机地址。

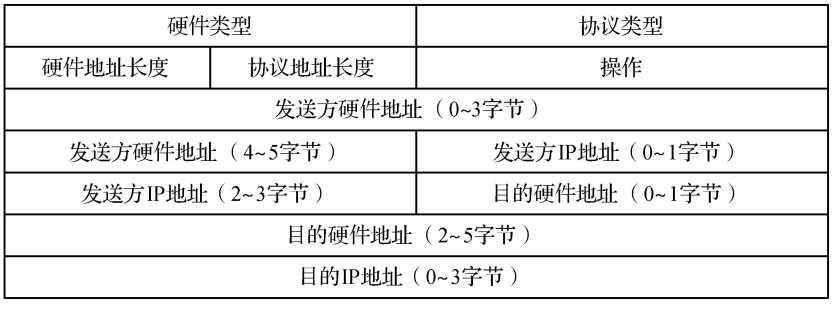


图 3-23　ARP 报文格式

ARP 又被称为 Ethernet ARP，最初是为以太网制定的，但是它现在可以在具有类似机制的其他物理网络上使用。

2）RARP

反向地址转换协议（reverse ARP，RARP）的功能与 ARP 相反，它负责为主机查找其物理地址对应的 IP 地址，即将物理地址转换为 IP 地址。RARP 主要是为无盘工作站设计的。主机的 IP 地址通常保存在硬盘中，主机启动时从硬盘中读出 IP 地址并存放在内存中。这样，对于没有硬盘的无盘工作站来说，如何获悉自己的 IP 地址就变得非常困难。RARP 正是为解决这一问题而设计的。

主机为了获得自己对应的 IP 地址，必须向网络上广播发送一个 RARP 请求报文，并在 RARP 请求报文中指明自己的物理地址。在网络上被授权提供 RARP 服务的 RARP 服务器在收到一个 RARP 请求报文后，将根据 RARP 报文携带的物理地址，查找其对应的 IP 地址及其他 IP 地址相关参数，并发送应答报文给请求的主机。这个过程只在无盘工作站启动时执行一次。

在网络中通常会有多台 RARP 服务器同时工作，如果一台服务器发生故障，还会有其他 RARP 服务器响应 RARP 请求。

3）BOOTP

引导协议（BOOTP）在实际网络中应用的最大问题在于：RARP 采用广播方式发送请求，而路由器并不支持广播，因此必须在每个网络中部署一个 RARP 服务器。因此，在网络中通常使用 BOOTP 来替代 RARP。

BOOTP 是一种基于 UDP 的协议，因此可以通过路由器转发。它的主要工作也是帮助无盘工作站获得自己的 IP 地址、服务器的 IP 地址、启动映像文件名、网关 IP 等。其工作原理与 RARP 类似，具体过程如下。

（1）由 BOOT ROM 芯片中的 BOOTP 启动代码启动客户机，此时客户机还没有 IP 地址，它就用广播形式以 IP 地址 0.0.0.0 向网络中发出 IP 地址查询的请求，这个请求中包含了客户机网卡的 MAC 地址。

（2）网络中运行 BOOTP 服务的服务器接收到这个请求，根据请求中的 MAC 地址在 BOOTPTAB 启动数据库中查找这个 MAC 的记录，如果没有此 MAC 的记录则不响应这个请求；如果有就将 FOUND 帧发送回客户机。FOUND 帧中包含的主要信息有客户机的 IP 地址、服务器的 IP 地址、硬件类型、网关 IP 地址、客户机 MAC 地址和启动映像文件名。

（3）客户机根据 FOUND 帧中的信息通过 TFTP 服务器下载启动映像文件，并将此文件模拟成磁盘，从这个模拟磁盘启动。

4）DHCP

动态地址配置协议（dynamic host configuration protocol，DHCP）是 BOOTP 的扩展，采用客户机/服务器模式，提供了一种动态指定 IP 地址及其配置参数的机制。DHCP 主要用于大型网络环境或者其他配置比较困难的情况。DHCP 服务器自动为客户机指定 IP 地址和相关的一组配置参数。

在网络中至少有一台 DHCP 服务器，它监听网络中的 DHCP 请求，并与客户端协商