第 1 章

计算机网络基础

本章介绍在计算机网络组建与维护管理中所必须理解和掌握的一些计算机网络知识，并重点介绍 TCP/IP 模型和 TCP/IP 协议。

1.1 计算机网络基本概念

计算机网络是计算机技术和通信技术发展的必然产物，进入 20 世纪 90 年代以后，以因特网(Internet)为代表的计算机网络得到了飞速发展，加速了全球数字化、网络化和信息化革命的进程。计算机网络正日益影响和改变着人们的生活方式、工作方式和学习方式，现在人们的生活、工作、学习和交往都已离不开计算机网络了。

1.1.1 计算机网络的定义、分类与性能指标

1. 计算机网络的定义

计算机网络是指利用有线或无线的传输介质，将分布在不同地理位置、独立的计算机互联起来而构成的计算机集合。组建网络的目的是实现资源共享和通信。

目前，最庞大的计算机网络就是因特网(Internet)，它利用传输介质和网络互联设备将分布在全球范围内的计算机或计算机网络互联起来，从而形成一个全球性的计算机网络。

2. 计算机网络的分类

可以从不同的角度对计算机网络进行分类。

(1) 根据网络交换功能的不同，计算机网络可分为电路交换网、报文交换网、分组交换网和混合交换网。混合交换网就是在一个数据网络中同时采用了电路交换技术和分组交换技术的网络。

目前，计算机网络主要采用分组交换技术，电话网络采用电路交换技术。

(2) 根据网络覆盖地理范围的大小，计算机网络可分为局域网、城域网和广域网。

- 局域网(Local Area Network，LAN)。局域网是指网络覆盖范围在几百米至几千米的网络，网络覆盖的地理范围较小，如校园网、企事业单位内部网等。

  局域网可运行的协议主要有以太网协议(IEEE 802.3)、令牌总线(IEEE 802.4)、令牌环(IEEE 802.5)和光纤分布数据接口(FDDI)。目前，局域网最常用的是以太网协议，因此，在没有特别说明的情况下局域网通常是指以太网。以太网是指运行以太网协议的网络。

- 城域网(Metropolitan Area Network,MAN)。城域网是指网络覆盖范围在几千米至几十千米的网络,其作用范围为一个城市。城域网可采用局域网技术来组建,也可采用分布式队列双总线(Distributed Queue Dual Bus,DQDB)技术来组建,该协议已成为国际标准,编号为IEEE 802.6。
- 广域网(Wide Area Network,WAN)。广域网是指网络覆盖范围在几十至几千千米的网络,可以跨越不同的国家或洲。广域网通信所采用的技术与局域网有较大差别。

(3) 根据网络的使用者,计算机网络可划分为公用网络和专用网络。

3. 计算机网络的性能指标

计算机网络的主要性能指标有带宽和时延。

(1) 带宽

在模拟信号中,带宽是指通信线路允许通过的信号频率范围,其单位为赫兹(Hz)。

在数字通信中,带宽是指数字信道发送数字信号的速率,其单位为比特每秒(b/s或bps),因此带宽有时也称为吞吐量,常用每秒发送的比特数来表示。例如,通常说某条链路的带宽或吞吐量为100M,实际上是指该条链接的数据发送速率为100Mb/s或100Mbps,即每秒钟可传送100M比特的数据。

注意:在数字通信中,单位换算关系与计算机领域是不同的,其换算关系如下:

1kb/s=1000b/s

1Mb/s=1000kb/s

1Gb/s=1000Mb/s

(2) 时延

时延是指一个报文或分组从链路的一端传送到另一端所需的时间。时延由发送时延、传播时延和处理时延三部分构成。

发送时延是使数据块从发送节点进入到传输介质所需的时间,即从数据块的第一个比特数据开始发送算起,到最后一个比特发送完毕所需的时间,其值为数据块的长度除以信道带宽,因此,在发送的数据量一定的情况下,带宽越大,则发送时延越小、传输越快。发送时延又称传输时延。

传播时延是指电磁波在信道中传输一定的距离所花费的时间。一般情况下,这部分时延可忽略不计,但若通过卫星信道传输,则这部分时延较大。电磁波在铜线电缆中的传播速度约为2.3×10^5km/s,在光纤中的传播速度约为2.0×10^5km/s,1000km长的光纤线路产生的传播时延约为5ms。

处理时延是指数据在交换节点为存储转发而进行一些必要处理所花费的时间。在处理时延中,排队时延占的比重较大,通常可用排队时延来作为处理时延。

1.1.2 网络拓扑结构

网络拓扑结构是指用传输介质互联的各节点的物理布局。在网络拓扑结构图中,通常用点来表示联网的计算机,用线来表示通信链路。

在计算机网络中，网络拓扑结构主要有总线型、星型、环型、树型和网状型，最常用的是星型结构。在实际组网应用中，可能采取多种结构混合使用。

1. 总线型结构

总线型结构网络使用同轴电缆细缆或粗缆作为公用总线来连接其他节点，总线的两端安装一对 50Ω 的终端电阻，以吸收电磁波信号，避免产生有害的电磁波反射。采用细同轴电缆时，每一段总线的长度一般不超过 185m。其拓扑结构如图 1.1 所示。总线型结构网络可靠性差、速率慢(10Mb/s)，目前已很少使用。

主要优点：结构简单，所需电缆数量较少。

主要缺点：故障诊断和隔离较困难，可靠性差，传输距离有限，共享带宽，速度慢。

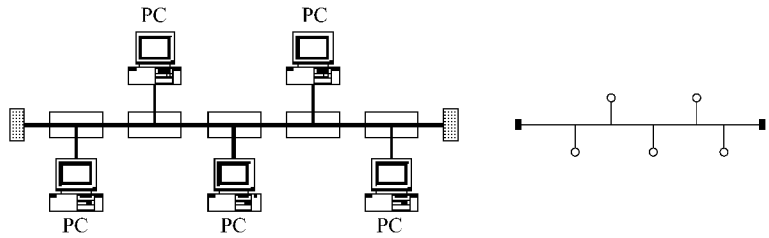

图 1.1　总线型网络结构

2. 星型结构

星型结构网络中各节点以星型方式连接到中心交换节点，从而实现各节点间的相互通信，是目前局域网的主要组网方式。中心交换节点可以采用集线器或交换机，目前主要采用交换机作为中心交换节点。其拓扑结构如图 1.2 所示。

主要优点：控制简单，故障诊断和隔离容易，易于扩展，可靠性好。

主要缺点：需要的电缆较多，中心交换节点负荷较重。

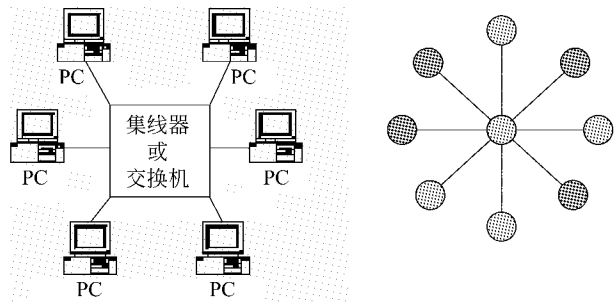

图 1.2　星型网络结构

3. 环型结构

环型结构由通信线路将各节点连接成一个闭合的环，数据在环上单向流动，网络中用令牌控制来协调各节点的发送，任意两节点都可通信。网络拓扑结构如图 1.3 所示。

主要优点：所需线缆较少，易于扩展。

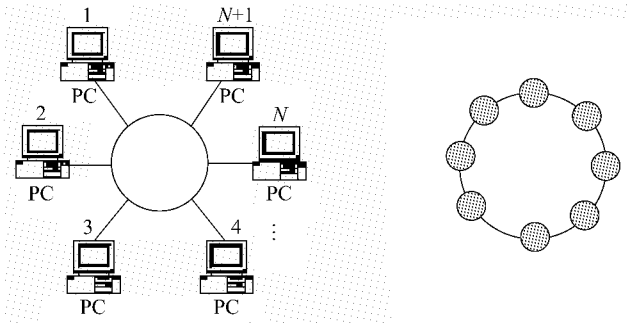

图 1.3 环型网络结构

主要缺点：可靠性差，一个节点的故障会引起全网故障；故障检测困难。

4. 网状型结构

网状型结构在网络的所有设备间实现点对点的互联，其拓扑结构如图 1.4 所示。

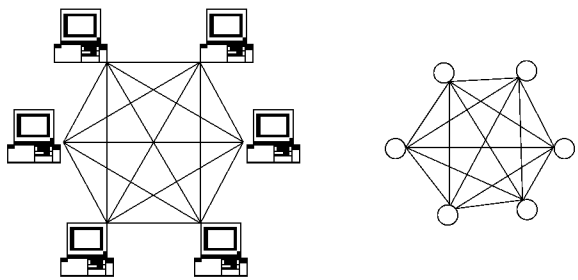

图 1.4 网状型网络拓扑结构

在局域网中，使用网状结构较少；在因特网中，骨干路由器彼此间的互联可采用该种网状结构，以提供到达目标网络的多种路径选择和冗余链路。

5. 树型结构

树型结构像一棵倒置的树，顶端是树根，树根以下带分支，每个分支还可以再进行分支。树型结构易于扩展，故障隔离较容易，其缺点是各个节点对根的依赖性太大。

1.1.3 网络通信协议

1. 网络通信协议的概念

在计算机网络中，要做到有条不紊地交换数据和通信，就必须共同遵守一些事先约定好的规则。这些为进行网络中的数据交换而建立的规则、标准或约定，就称为网络协议。

网络协议由语法、语义和同步三个要素组成。语法规定了数据与控制信息的结构或格式；语义定义了所要完成的操作，即完成何种动作或做出何种响应；同步定义了事件实现顺利的详细说明。

2. 常用的网络通信协议

在局域网中，常用的协议主要有 NetBEUI 协议和 TCP/IP 协议，用得最广泛的主要是 TCP/IP 协议。

(1) NetBEUI 协议

NetBEUI(NetBIOS Extended User Interface，NetBIOS 扩展用户接口)是 IBM 于 1985 年开发的一种体积小、效率高、速度快的通信协议，但不具备跨网段工作的能力，主要用于小型网络(小于 200 台主机)。

(2) TCP/IP 协议

TCP/IP 协议是因特网(Internet)的标准通信协议，支持路由和跨平台特性。在局域网中，也广泛采用 TCP/IP 协议来工作。

TCP/IP 协议是一个大的协议集，并不仅是 TCP 和 IP 这两个协议。有关 TCP/IP 协议的详细介绍将在后续部分讲解。

1.2　计算机网络体系结构

相互通信的两个计算机系统必须高度协调一致才能正常工作，而这种“协调”是相当复杂的，因此计算机网络实际上是一个非常复杂的系统。

对计算机网络体系结构进行分层，可将庞大而复杂的问题转化为若干较小的局部问题，这样就比较容易研究和处理。

对计算机网络体系结构的分层模型有 OSI 参考模型和 TCP/IP 模型两种。OSI 属于国际标准，由于分层较多，实现较复杂，主要用于理论研究；TCP/IP 模型分层较少，实现较容易，成为事实上的国际标准。

1.2.1　OSI 参考模型

OSI 参考模型(Open System Interconnection Reference Model，开放系统互联参考模型)是国际标准化组织 ISO 于 1983 年正式推出的，即著名的 ISO 7498 国际标准。

在 OSI 参考模型中，网络体系结构被分成了七层，由低层到高层依次是物理层、数据链路层、网络层、传输层、会话层、表示层和应用层。每一层均向相邻的上一层通过层间接口提供服务，上一层要在下一层所提供的服务的基础上实现本层的功能，因此服务是垂直的，而协议是水平的，协议是控制对等层实体之间通信的规则，即只有对等的层才能相互通信。

应用层为用户提供所需的各种应用服务，如 Web 服务、邮件服务、远程登录、文件传输服务、域名服务等。

表示层主要用于数据的表示、编码和解码，实现信息的语法语义表示和转换，如加密解密、转换翻译、压缩与解压缩等。

会话层用于为不同机器上的应用进程建立和管理会话。

传输层也称为运输层，主要用于解决数据在网络之间的传输质量问题，提高网络层服务质量，提供可靠的点对点的数据传输。它从会话层接收数据，并在必要时将数据分割成

适合在网络层传输的数据单元,然后将这些数据交给网络层,再由网络层负责将数据传送到目的主机。该层的数据传输单位为数据报。

网络层解决网络与网络之间的通信问题,主要功能有逻辑编址、分组传输、路由选择等。此层的数据传送单位为IP数据包。

数据链路层为网络层接供一条无差错的数据传输链路。在发送数据时,接收网络层传递来的数据包,封装成数据帧;在接收数据时,数据链路层将物理层传递来的二进制比特流还原为数据帧。数据链路层传送的基本单位为数据帧,使用物理地址(Media Access Control,MAC)进行寻址。

物理层负责传送原始比特流,并屏蔽传输介质的差异,使数据链路层不必考虑传输介质的差异,实现数据链路层的透明传输。另外,物理层还必须解决比特同步的问题。

1.2.2 TCP/IP模型

1. TCP/IP模型体系结构

OSI的七层体系结构仅是一个纯理论的分析模型,本身并不是一个具体协议的真实分层,既复杂又不实用,因此具有四层体系结构的TCP/IP模型得到了广泛应用,成为事实上的国际标准和工业生产标准。

在TCP/IP模型中,网络体系结构由低层到高层,依次分为网络接口层、网络层、传输层和应用层,其网络体系结构与OSI七层结构的对应关系如图1.5所示。

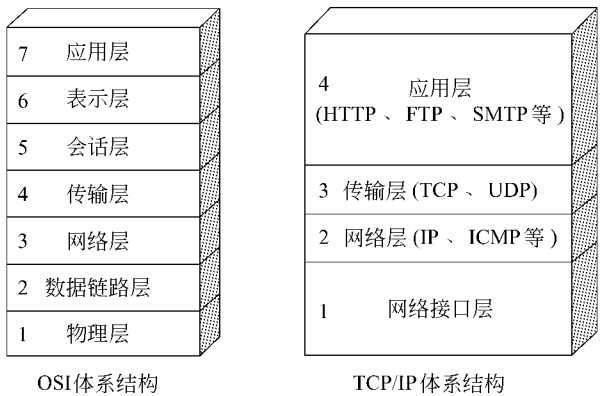

图1.5 OSI与TCP/IP体系结构

在TCP/IP体系结构中,网络接口层整合了OSI体系结构中的物理层和数据链路层的功能,因此,从协议的层次结构看,TCP/IP模型实际上是一个具有五层协议的体系结构。

在实际应用中,网络接口层主要由网络接口(网卡)来实现,它实现了数据链路层和物理层的功能。网络层主要由路由器或3层交换机来实现。传输层由用户主机中的应用进程来实现,它存在于分组交换网外面的主机之中。传输层的任务就是负责主机中两个进程之间的通信,传输层上层的应用层就不再关心信息的传输问题了。通常也将分组交换网称为通信子网,而将用户主机的集合称为资源子网。

2. 各层常用的协议

(1) 应用层协议

应用层常用协议主要有 HTTP(超文本传输协议)、S-HTTP(安全的超文本传输协议)、SMTP(简单邮件传输协议)、IMAP4(因特网信息访问协议第 4 版)、POP3(邮局协议第 3 版)、TELNET(终端仿真协议)、FTP(文件传输系统)、TFTP(简单文件传输协议)、DNS(域名系统)、DHCP(动态主机配置协议)、SNMP(简单网络管理协议)等。

(2) 传输层协议

传输层协议主要有 TCP(传输控制协议)和 UDP(用户数据报协议)。

(3) 网络层协议

网络层主要协议有 IP(网际协议)/IPv6 协议、ICMP(互联网控制信息协议)/ICMPv6、RIP2(路由信息协议第 2 版)、OSPF(开放最短路径优先协议)、IGRP(内部网关路由协议)、EGP(外部网关协议)等。

(4) 网络接口层协议

网络接口层完成了 OSI 中的数据链路层和物理层的功能,在进行数据分组传送时,负责建立无差错的数据传输链路。

数据链路层常用的协议主要有 MAC(媒体接入控制)、HDLC(高级数据链路控制协议)、PPP(点对点协议)、ARP(地址解析协议)、RARP(逆向地址解析协议)、MPLS(多协议标签交换协议)等。

ARP 协议用于将目的 IP 地址解析为数据链路层物理寻址所需的 MAC 地址,解析成功后,IP 地址与 MAC 地址的对应关系会保存在主机的 ARP 缓冲区中,建立起一个 ARP 列表,以供下次查询使用。RARP 则用于将 MAC 地址解析为对应的 IP 地址。

目前,使用得最多的是 TCP/IP 协议的第 4 版,即 IPv4。新的版本是 IPv6,已开始在骨干网络中应用,IPv4 与 IPv6 将共同存在较长的时间。

3. 数据在各层间的传递过程

为简化问题,假设计算机 1 和计算机 2 直接相连,现在计算机 1 的应用进程 AP_1 要向计算机 2 的应用进程 AP_2 发送数据。下面分析该数据在发送端和接收端的各层间的传递过程。

应用进程 AP_1 将要传送的数据交给应用层,应用层在数据首部加上必要的控制信息 H_5,然后将数据传递给下面的传输层,数据和控制信息就成为下一层的数据单元。

传输层接收到这个数据单元后,再在首部加上本层的控制信息 H_4,再交给下面的网络层,成为网络层的数据单元。

网络层接收到这个数据单元后,再在首部加上 IP 包头 H_3,并交给下面的数据链路层。

数据链路层接收到这个数据单元后,在首部和尾部分别加上控制信息 H_2 和 T_2,将数据单元封装成数据帧,然后交给物理层进行传送。

对于 HDLC 数据帧,在首部和尾部各加上 24bit 的控制信息;对于 Ethernet V2 格式的 MAC 帧,首部添加 14(6+6+2)字节,尾部添加 4 字节的帧校验序列(Frame Check

Sequence,FCS)。

物理层直接进行比特流的传送,不再加控制信息。当这一串比特流经网络传输介质到达目的主机时,就从第1层依次交付给上一层进行处理。每一层根据控制信息进行必要的操作,然后将本层的控制信息剥去,将剩下的数据单元再交付给上一层进行处理,最后应用进程 AP_2 就可收到来自 AP_1 应用进程传送的数据。

从中可见,在发送时,数据从高层向低层流动,每一层(物理层除外)都给收到的数据单元套上一个本层的"信封"(控制信息);接收时,数据从低层向高层流动,每一层(物理层除外)进行必要处理后,去掉本层的"信封",将"信封"中的数据单元再上交给上一层进行处理。整个传递过程如图1.6所示。

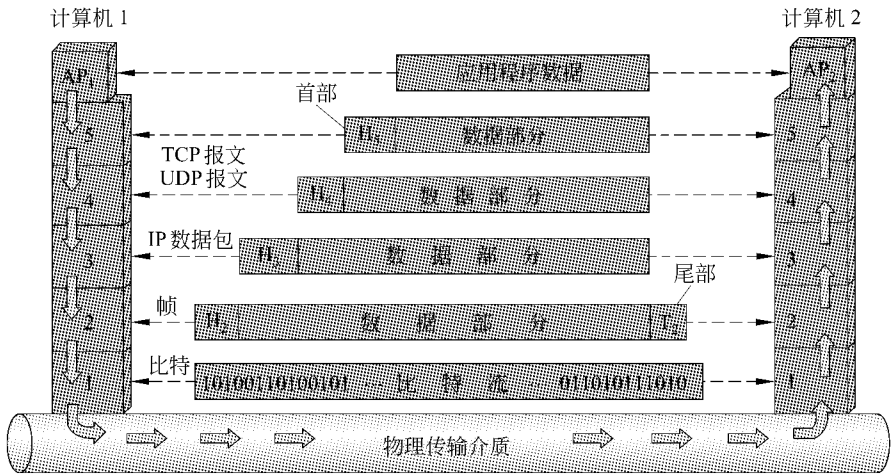

图1.6 数据在各层间的传递过程

1.3 以太网简介

1. 以太网标准

以太网是美国施乐(Xerox)公司于1975年研制成功的,是一种基于基带总线的局域网,采用无源电缆作为总线来传送数据帧,当时的数据速率为2.94Mb/s。

1980年,DEC、Intel和施乐公司联合提出了10Mb/s速率的以太网规范(DIX V1),1982年修改为第2版(DIX Ethernet V2),成为世界上第一个局域网规范。

在此基础上,IEEE 802委员会于1983年制定了第一个以太网标准,编号为802.3,数据速率为10Mb/s。该标准仅对帧格式做了很小的一点调整,允许基于这两种标准的硬件可以在同一个局域网上互操作。由于这两个标准差异很小,通常不严格区分,因此目前存在两个以太网协议标准,即国际标准的IEEE 802.3和DIX Ethernet V2标准。

由于商业竞争,IEEE 802委员会并未形成统一的局域网标准,除了以太网局域网标准(802.3)外,还有令牌总线(802.4)和令牌环网(802.5)的局域网标准。目前局域网主要采用以太网协议标准,称为以太局域网。

2. 以太网工作原理

以太网使用载波监听多点接入/碰撞检测协议，即 CSMA/CD（Carrier Sense Multiple Access with Collision Detection）协议进行工作。

载波监听是指每一个站点在发送数据之前，要先检测总线是否空闲，是否有其他计算机在发送数据。若有，则暂时不要发送数据，以免发生碰撞冲突。

碰撞检测是指站点应一边发送数据，一边检测信道上信号电压的大小，以判断当前是否有冲突产生。若有碰撞冲突产生，信号将产生严重失真，此时就必须立即停止发送，并等待一段随机时间后，再重新进行载波监听和发送。

从中可见，使用 CSMA/CD 协议工作时，一个站点不能同时发送数据和接收数据，属于半双工通信。连接在同一总线上的所有站点，均在同一个冲突域范围，站点越多，碰撞冲突的几率就越大，网络通信速率和效率就会大大降低。

3. 高速以太网

传统以太网的速率为 10Mb/s，且以半双工方式工作。速率达到和超过 100Mb/s 的以太网统称为高速以太网。

（1）快速以太网

快速以太网（Fast Ethernet）是指速率达到 100Mb/s 的以太网，采用星型拓扑结构，在双绞线（100Base-TX）或光纤（100Base-FX）上传送 100Mb/s 的基带信号。1995 年 IEEE 正式将快速以太网定为国际标准，编号为 802.3u。

快速以太网的 MAC 帧格式仍采用 IEEE 802.3 标准规定的帧格式，由于速率提高了 10 倍，为了保持最短帧长（64 字节）不变，采取了将网段的最大电缆长度减小到 100m，帧间时间间隔也从原来的 9.6μs 改为 0.96μs。

快速以太网是对 IEEE 802.3 标准的补充，能自动识别和适应 10Mb/s 和 100Mb/s 网速。快速以太网在半双工模式工作时，遵循 CSMA/CD 协议；但在全双工模式工作时，则不再遵循该协议。

（2）吉比特以太网

吉比特以太网又称为千兆以太网，IEEE 于 1997 年通过了吉比特以太网标准，编号为 IEEE 802.3z，1998 年成为正式标准。

吉比特以太网允许在 1Gb/s 速率下以全双工或半双工两种模式工作，向后兼容 10Base-T 和 100Base-T。使用 IEEE 802.3 协议规定的帧格式，在半双工模式工作时使用 CSMA/CD 协议。

吉比特以太网目前常用作主干网，对带宽要求较高的应用场合，也可采用吉比特以太网，千兆交换到桌面。

吉比特以太网的物理层可以使用基于光纤（1000Base-X）和双绞线（1000Base-T）的传输介质。可使用的光纤传输介质有以下两种。

- 1000Base-SX　SX 表示使用短波长（850nm）激光。使用纤芯直径为 62.5μm 和 50μm 的多模光纤时，传输距离分别为 275m 和 550m。
- 1000Base-LX　LX 表示使用长波长（1300nm）激光。使用纤芯直径为 62.5μm 和

50μm的多模光纤时,传输距离为550m;使用纤芯直径为10μm的单模光纤时,传输距离为5km。

1000Base-T使用4对5类UTP双绞线时,传输距离为100m。

(3) 10吉比特以太网

10吉比特以太网又称万兆以太网,由IEEE 802.3ae委员会制定,编号为IEEE 802.3ae,于2002年6月成为正式标准。

10吉比特以太网的帧格式、最小帧长和最大帧长均与IEEE 802.3标准规定相同,以利于以太网的升级。10吉比特以太网只能以全双工模式工作,因此不再使用CSMA/CD协议。其采用星型结构组网,由于数据速率很高,传输介质只能使用光纤,而不能使用铜线。

若使用多模光纤,传输距离可达65～300m;若使用长距离单模光纤,配合长距离单模光纤接口,传输距离可达40km。

1.4 数据链路层与以太网帧格式

1.4.1 数据链路层简介

为了使数据链路层能更好地适应多种局域网标准,802委员会将局域网的数据链路层分成了两个子层,分别是逻辑链路控制(Logical Link Control,LLC)子层和媒体接入控制(Medium Access Control,MAC)子层。与接入到传输媒体有关的内容均放在MAC子层中,这样LLC子层就与传输媒体无关,不管采用何种局域网协议标准,对LLC子层来说都是透明的。

LLC子层在MAC子层的基础上向网络层提供服务,MAC子层的存在屏蔽了不同物理链路种类的差异性。其主要功能包括数据帧的封装和拆封、帧的寻址和识别、帧的接收与发送、链路管理、帧差错控制等。

数据链路层传输的数据单位为数据帧,寻址时使用的地址为MAC地址(物理地址或硬件地址)。MAC地址采用6个字节共48bit二进制数编码表示,表达时采用十六进制数来表示。对于Windows系统,采用"xx-xx-xx-xx-xx-xx"格式表示,例如00-0F-EA-01-B9-4E。在华为和华三交换机或路由器中,采用"xxxx-xxxx-xxxx"格式表示,如000F-EA01-B94E;在Cisco交换机或路由器中,采用"xxxx. xxxx. xxxx"格式表示,如000F.EA01.B94E。

MAC地址是全球唯一的,不允许重复,前3个字节为厂商标识,后3个字节为该厂商所生产的网络设备的序号。

网络适配器(网卡)实现了数据链路层和物理层的功能。

1.4.2 以太网帧格式

目前以太网有四种不同标准的帧格式,分别是DIX Ethernet V2帧格式、IEEE 802.3 raw帧格式(Novell专用的以太网标准帧格式)、IEEE 802.3 SAP帧格式和IEEE 802.3 SNAP帧格式。目前最常用的是DIX Ethernet V2标准的帧格式,也是目前以太网的网

络设备默认采用的帧格式，该种帧格式比较简单。

以太网设备默认采用 Ethernet V2 帧格式，将网络层传输来的 IP 数据报文，通过添加帧头和帧尾，封装成数据帧，然后在物理层中传输。Ethernet V2 标准的 MAC 帧格式如图 1.7 所示。

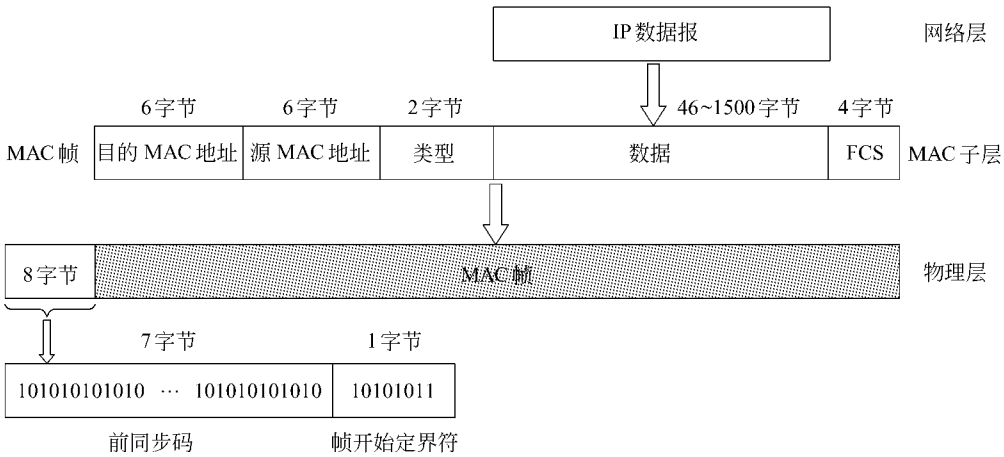

图 1.7　Ethernet V2 标准的 MAC 帧格式

Ethernet V2 帧格式的帧头由目的 MAC 地址、源 MAC 地址和 2 个字节的类型标识字段构成，帧头共 14 字节；接下来的是不定长的数据字段，该字段为帧的负荷(payload)，即所封装的 IP 数据报文，该部分数据的长度为 46～1500 字节，即 MTU(Maximum Transmission Unit，最大传输单元)值；帧尾由 4 个字节构成，代表帧校验序列(Frame Check Sequence，FCS)，采用 32 位的 CRC 循环冗余校验，对从目标 MAC 地址字段到数据字段的数据进行校验。

2 个字节的类型标识字段用于标识以太网帧中所携带的上层数据的协议类型，采用十六进制数表示。例如，0x0800 代表 IP 协议数据，0x86DD 代表 IPv6 协议数据，0x809B 代表 AppleTalk 协议数据，0x8137 代表 Novell IPX 协议数据。

从中可见，以太网中 MAC 帧最短有效帧长为 64 字节，凡是长度小于 64 字节的帧都是无效帧，并直接丢弃。最大帧长为 1518 字节。

从 MAC 子层将 MAC 帧交给物理层进行传输时，还要在帧的前面插入 8 个字节(由硬件自动生成和插入)。这 8 个字节由两部分构成，第一部分为 7 个字节构成的前同步码(1 和 0 交替出现)，其作用是使接收端在接收 MAC 帧时能迅速实现比特同步；第二部分为 1 个字节的帧开始定界符，定义为 10101011，表示在这后面的信息就是 MAC 帧了。因此在物理层传输的数据要比 MAC 帧多 8 个字节，添加的目的是实现比特同步。这是因为在刚开始接收 MAC 帧时，由于尚未与到达的比特流达到同步，MAC 帧最前面的若干比特就无法接收到，会使整个 MAC 帧成为无效的帧。

1.5 TCP/IP 协议

1.5.1 TCP 协议

1. TCP 协议简介

TCP(Transmission Control Protocol,传输控制协议)和 UDP(User Datagram Protocol,用户数据报协议)是传输层所使用的协议。TCP 提供面向连接的可靠传输服务,利用 TCP 协议传送数据时,有建立连接→传送数据→释放连接的过程。UDP 是无连接的协议,提供尽最大努力交付的传输服务,属于不可靠服务,常用于传输语音、视频等数据量大且对可靠性要求不高的应用。

2. TCP 协议的功能

TCP 协议主要是建立连接,然后从应用层的应用进程中接收数据并进行传输。TCP 采用虚电路连接方式进行工作,在发送数据前,它需要在发送方和接收方之间建立起一个连接,数据在发送出去后,发送方会等待接收方给出一个收到数据的确认性应答。否则,发送方将认为此数据报丢失,并将重新发送此数据报,以保证数据传输的可靠性。

3. TCP 报头

传输层使用 TCP 协议时,在数据单元首部所添加的控制信息,就是 TCP 报头。TCP 报文首部(报头)的前 20 个字节是固定的,后面有 4N(N 为整数)字节是根据需要而增加的选项,因此 TCP 报头的总长度最小为 20 个字节,报头结构如图 1.8 所示。

<table>
<tr><td colspan="2">比特 0</td><td colspan="2">比特 15 比特 16</td><td>比特 31</td></tr>
<tr><td colspan="3">源端口 (16)</td><td colspan="2">目的端口 (16)</td></tr>
<tr><td colspan="5">序列号 (32)</td></tr>
<tr><td colspan="5">确认号 (32)</td></tr>
<tr><td>TCP 偏移量 (4)</td><td>保留 (6)</td><td>标志 (6)</td><td colspan="2">窗口 (16)</td></tr>
<tr><td colspan="3">校验和 (16)</td><td colspan="2">紧急 (16)</td></tr>
<tr><td colspan="5">选项 (0 或 32)</td></tr>
<tr><td colspan="5">数据 (可变)</td></tr>
</table>

图 1.8 TCP 报头结构

源端口:指定了发送端所使用的端口号。端口采用 2 个字节共 16 个比特编码表示,因此 TCP 端口最多可有 65536 个端口。端口是传输层向应用层提供服务的层间接口。

目的端口:指定了接收端所使用的端口号。

序列号:占 4 个字节,32bit。TCP 给在一个 TCP 连接中传送的数据流中的每一个字节都编上一个序号,整个数据的起始序列号在连接建立时设置。TCP 报头中的序列号字段的值代表本报文段所发送数据的第一个字节的序号。例如,若当前 TCP 报头中的序列号值为 101,本报文所携带的数据为 100 个字节,则下一个 TCP 报文的报头序列号值就应为 201。

确认号：占 4 个字节，代表期望收到的下一个报文段数据的第一个字节序号，即期望收到的下一个报文段首部的序列号字段的值。

TCP 在传输的过程中，使用序列号和确认号来跟踪数据的接收情况。

TCP 偏移量：占 4bit，它指定了段头的长度。即 TCP 报文段的数据起始处距离 TCP 报文段的起始处有多远。段头的长度与段头选项字段的设置有关。

保留：占 6bit，指定了一个保留字段，以备将来使用，目前应置为 0。

标志：占 6bit，从左到右依次是 URG、ACK、PSH、RST、SYN、FIN 标志位，含义如下。

- URG：表示紧急指针。当 URG 位为 1 时，TCP 报头的紧急字段才有效，它相当于告诉系统该报文段有紧急数据，需要尽快传送，而不是按原来的排队顺序传送。
- ACK：表示确认。只有当 ACK 标志位为 1 时，TCP 报头的确认字段才有效。
- PSH：表示尽快地将数据送往接收进程处理，而不再等到缓冲区填满后才向上交付给应用进程处理。
- RST：表示复位连接。当 RST 位被置为 1 时，表明 TCP 连接中出现严重差错，必须释放连接，然后再重新建立连接。利用复位比特可实现异常终止一个连接。
- SYN：表示同步，在连接建立时用来同步序号。当 SYN＝1 而 ACK＝0 时，表明这是一个连接请求报文。若对方同意建立连接，则在响应报文中，应使 SYN＝1，ACK＝1。因此，同步比特 SYN 置为 1，就表明这是一个连接请求报文或连接接受的响应报文。
- FIN：用于释放一个连接。当 FIN 位为 1 时，表明此报文段的发送端数据已发送完毕，并要求释放连接。

窗口：占 2 个字节，16bit，用于指定发送端允许传输的下一报文段数据的大小，单位为字节。发送方与接收方之间的流量控制是通过调整发送方的窗口大小来实现的，是用接收方的数据接收能力来控制发送方的窗口大小，从而控制发送端的数据发送量。

校验和：校验和包含 TCP 报头和数据部分，用来校验报头和数据部分在传输过程中的完整性。

紧急：指明报文中包含紧急信息，只有当 URG 标志位置 1 时，紧急指针才有效。

选项：长度可变。目前，TCP 只规定了一个选项，即 MSS(Maximum Segment Size，最大报文段长度)，它代表了 TCP 报文中数据字段的最大长度。在连接建立过程中，双方应将自己能够支持的 MSS 填写在这一字段中，在以后的数据传送阶段，MSS 取双方的较小值来决定 TCP 报文负荷的大小。若选项字段未填(0 值)，则 MSS 默认值为 536 个字节，此时 TCP 报文的大小为 536＋20＝556 字节。

对 TCP 报文的解码视图如图 1.9 所示。

4. TCP 协议的工作原理

(1) TCP 连接的建立过程

TCP 传送数据应首先建立起 TCP 连接，其连接的建立过程又称为 TCP 的三次握手。

第一次握手：首先发送方主机向接收方主机发起一个建立连接的同步(SYN)请求

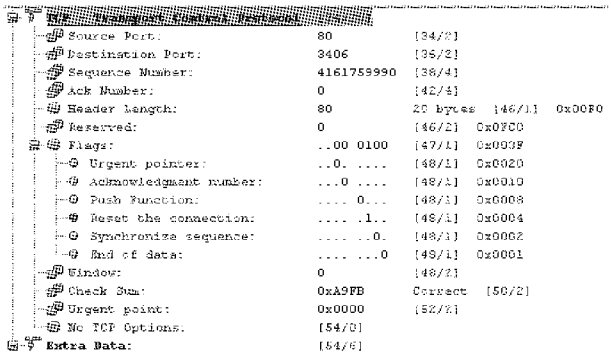

图 1.9 TCP 报文解码视图

SYN(X),并进入 SYN_SEND 状态,等待接收方主机确认。

第二次握手:接收方主机在收到这个请求后,如果同意建立连接,则发送确认 ACK,确认序号为收到的序号加 1,并且报文中的 SYN 也要置为 1,即向发送方主机回复一个同步/确认(SYN/ACK)应答报文,并进入 SYN_RCVD 状态。

第三次握手:发送方主机收到此应答报文后,再向接收方主机发送一个确认(ACK)报文,然后发送方和接收方均进入 ESTABLISHED 状态,完成三次握手。至此,TCP 连接建立成功,发送方和接收方就可开始传送数据了。

TCP 建立连接的三次握手过程如图 1.10 所示。

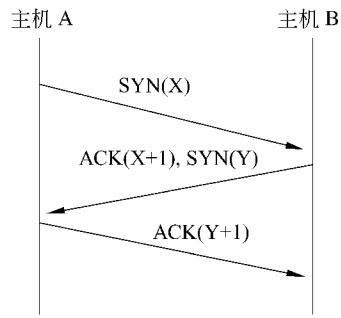

图 1.10 TCP 建立连接的三次握手过程

(2) TCP 连接的关闭

当应用进程结束数据传送后,就要释放已建立的连接。TCP 连接是双向的,每个方向都必须单独进行关闭。首先进行关闭的一方执行主动关闭,而另一方则执行被动关闭,关闭连接的过程如下。

① 当客户端的数据传输完后,可主动发送出 FIN 置 1 的报文给服务端(客户端主动关闭),以关闭客户端至服务端方向的数据传送,并等待服务端的 ACK 确认应答,同时进入 FIN_WAIT_1 状态。

② 服务端收到 FIN 置 1 的报文后,进入被动关闭,回复一个 ACK 确认报文,并进入 CLOSE_WAIT 状态;客户端收到该 ACK 确认报文后,进入 FIN_WAIT_2 状态。

③ 至此完成了 TCP 连接的半关闭,即完成了客户端至服务端方向的数据发送。此时,客户端虽然不能发送数据,但还仍能接收服务端发给客户端的数据,即服务端至客户端方向的连接还未被关闭。

服务端发送一个 FIN 置 1 的报文给客户端,关闭服务端至客户端方向的数据传送,并等待客户端的 ACK 确认应答,同时进入 LAST_ACK 状态。

客户端收到 FIN 置 1 的报文后,回复 ACK 确认报文,并进入 TIME_WAIT 状态,经过 2 倍报文最大生存时间(MSL)后,TCP 删除原来建立的连接记录,返回到初始的 CLOSED 状态。服务端收到 ACK 确认报文后,进入 CLOSED 状态,完成连接的双向

关闭。

整个 TCP 连接和关闭过程中的状态变迁如图 1.11 所示。图中的“收”代表当收到什么 TCP 报文时所发生的状态变迁；“发”代表为了进行某个状态变迁需要发送的 TCP 报文；“应用进程”说明当应用进程进行某种操作时发生的状态变迁。

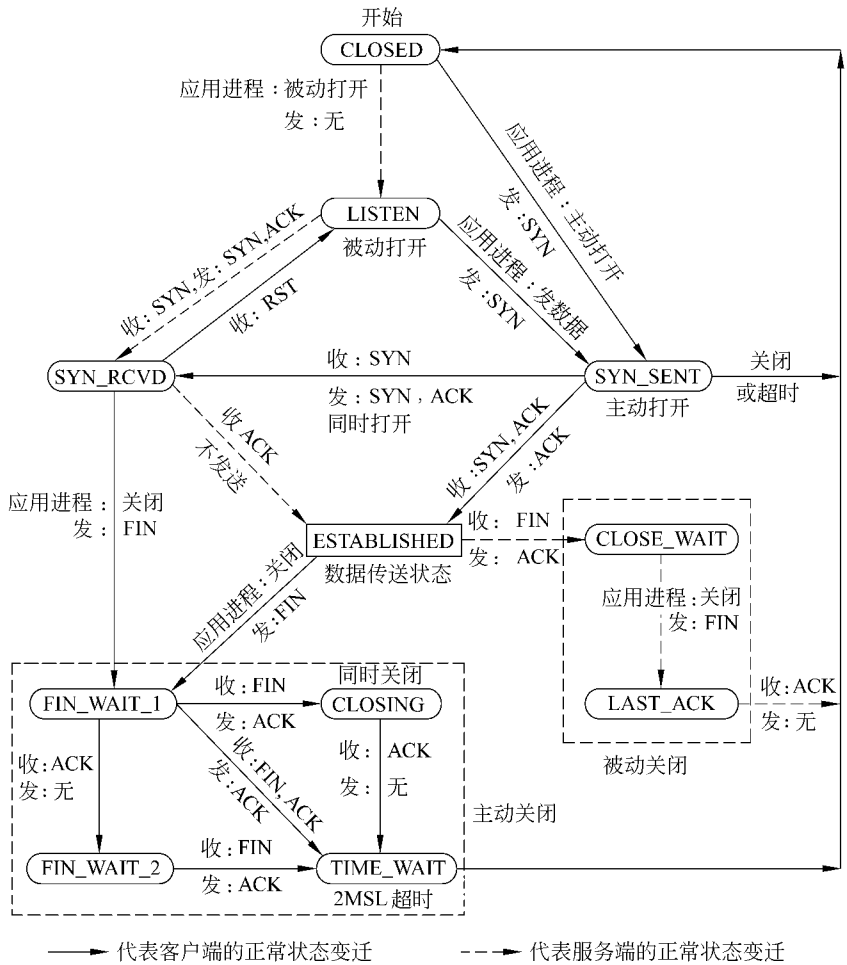

图 1.11　TCP 连接状态变迁

若要观察 TCP 连接状态的变迁，依次单击 Windows“开始”菜单中的“运行”菜单项，然后执行“cmd”命令进入 MS-DOS 状态。接着打开浏览器访问某一个网站，然后快速在 MS-DOS 状态执行查看连接状态的“netstat -an”命令，即可查看到当前的网络连接和各连接的状态。通过不断地执行“netstat -an”命令，进行状态刷新显示，即可观察到 TCP 连接的不同状态。

(3) TCP 重传

在 TCP 的传输过程中，如果在重传超时时间内，没有收到接收方主机对某数据报文的确认回复，发送方主机就认为此数据报文丢失，并再次发送这个数据报文给接收方，这称为 TCP 重传。

5. 端口的概念

端口(Port)是传输层的服务访问点,传输层使用端口与位于上层的应用进程进行通信,应用层的各种进程也通过相应的端口与传输层进行交互。

在发送数据时,应用层的应用进程通过相应的端口将数据传递给传输层,传输层就会在数据的首部添加一个报文头,并在报文头中写入源端口和目的端口号,然后将封装后的数据交给下层的网络层进行传输。

在接收数据时,网络层将收到的IP包的包头去掉,将数据上交给传输层,传输层再从报文头部取出该数据要送达的目的端口号,然后将报文头部去掉,将剩下的数据通过目的端口上交给相应的应用进程接收和处理。

从中可见,端口的作用就是让应用层的各种应用进程能将其数据通过端口向下交付给传输层;在接收数据时,让传输层知道应该将数据通过哪一个端口向上交付给目的应用进程接收处理。因此,端口可用来标志应用进程,或者说端口代表了某一种服务。

传输层的协议有TCP和UDP,因此TCP和UDP均有端口的概念。在报文头中,端口采用一个16位的二进制数编码表示,故TCP和UDP的端口总数均有65536个。

0～1023号端口分配给一些常用的标准服务固定使用,用户自行开发的应用进程应使用1024及以上的端口。常用的标准服务所使用的端口号如表1.1所示。

表1.1 常用的标准服务所使用的端口

服务程序	协议及端口	服务程序	协议及端口	服务程序	协议及端口
HTTP	TCP 80	SMTP	TCP 25	RPC	TCP/UDP 135
S-HTTP	TCP 443	POP3	TCP 110	NetBIOS-ns	TCP/UDP 137
FTP	TCP 21 和 TCP 20	IMAP4	TCP 143	NetBIOS-dgm	UDP 138
TFTP	UDP 69	SNMP	UDP 161	NetBIOS-ssn	TCP 139
DNS	UDP 53	SNMPTRAP	UDP 162	CIFS	TCP/UDP 445
Telnet	TCP 23	SSH	TCP 22		

CIFS(Common Internet File System)是从Windows 2000和Windows XP系统开始新增的,使用TCP/UDP 445号端口。

在Windows系统中,SMB(Server Message Block)用于实现文件和打印共享服务。NBT(NetBIOS over TCP/IP)使用TCP 137、UDP 138和TCP 139端口,来实现基于TCP/IP的NetBIOS网际互联。在Windows NT 4.0中,SMB基于NBT实现,从Windows 2000开始,SMB除了基于NBT实现外,还可直接通过TCP/UDP 445号端口来实现。

当Windows 2000/Windows XP系统未禁用NBT时(UDP 137、138和445以及TCP 139和445号端口将开放),当连接SMB服务时,会尝试连接TCP 139和TCP 445号端口。若TCP 445号端口有响应,则会发送RST标志位置1的报文给TCP 139号端口,以断开TCP 139号端口的TCP连接,然后改用TCP 445号端口建立连接;当TCP

445 号端口无响应时，才使用 TCP 139 号端口。

当系统禁用 NBT（只会开放 TCP/UDP 445 号端口）的情况下访问 SMB 服务时，只会尝试连接 TCP 445 号端口，若无响应，则连接失败。

RPC（Remote Procedure Call，远程过程调用）协议使用 TCP/UDP 135 端口工作，用于提供 DCOM（分布式组件对象模型）服务，利用 RPC 可以实现远程调用执行另一台计算机中的程序代码。

冲击波病毒、磁碟机病毒等很多病毒都是通过 135 端口传播的。为防止病毒在局域网中的大范围传播，通常可在楼宇汇聚层交换机、核心层交换机和防火墙上，通过配置 IP 包过滤规则，禁止访问 TCP 的 135、137、139 和 445 号端口以及 UDP 的 135、137、138 和 445 号端口，以达到阻断病毒传播的目的。

若要使用网管软件（SNMP 协议），在防火墙上应注意打开 UDP 161 和 UDP 162 端口。

6. TCP 协议的缺陷

TCP 协议在三次握手过程中存在一定的缺陷，利用这个缺陷，可发动 SYN 泛洪（SYN Flood）攻击，最终导致系统拒绝服务（Denial of Service，DoS）。

如果一个客户端向服务器发送了 SYN 连接请求报文后突然死机或掉线，则服务器在发出 SYN＋ACK 应答报文后是无法收到客户端的 ACK 应答报文的，第三次握手无法完成，此时的连接状态称为半连接状态。这种情况下，服务器端一般会重试（再次发送 SYN＋ACK 给客户端）并等待一段时间，若仍接收不到 ACK 确认报文，则丢弃这个未完成的连接，这段时间的长度称为 SYN 超时（Timeout），一般为 30s～2min。

一个客户端出现异常导致服务器的一个线程等待并不会造成大的问题，但如果一个恶意的攻击者或者大量的攻击者大量模拟这种情况，就会消耗掉系统的大量资源，导致系统运行缓慢甚至发生崩溃，出现无法响应正常用户访问请求的现象，从用户的角度看来，服务器没有响应，拒绝提供服务。

目前，由于服务器的运行速度和性能非常高，一对一地进行 SYN 泛洪攻击一般不会成功，可采取大量的攻击者同时对一台服务器发起 SYN 泛洪攻击来实现，这种攻击方式称为分布式拒绝服务攻击（Distributed Denial of Service），一般容易成功，危害性较大。

对 SYN 泛洪攻击可通过在路由器上配置 TCP 拦截来预防。

7. 利用 TCP 协议分析网络连接故障

对于网络连接访问故障，可利用 TCP 协议的相关知识来分析，以发现和解决问题。

在故障分析中，要注意从 TCP 建立连接的三次握手过程来进行分析，并注意连接是双向的，数据包（访问请求包）有去就必有回（响应包），否则就无法建立连接。可沿着数据包出去的路径和回来的路径进行分析，并注意请求包的源端口与目的端口以及响应包的源端口和目的端口的变化，以及注意检查沿途的网络设备（三层交换机、路由器或防火墙）是否开放了对这些目的端口的访问和允许来自某个源端口的响应包的通过。

例如在实际应用中，出现了局域网用户可正常访问网页和其他互联网服务但就是无法登录网上银行的现象。对于该网络故障，其分析思路如下。

既然用户能访问网页,说明网络连接和互联网出口没有问题,而出现某一项服务无法访问,则说明对该服务的访问请求包或者其响应包被防火墙、路由器或三层交换机给阻挡了;接下来就可沿着请求包出去和回来的路径,检查路径中的三层设备中的 ACL 包过滤规则,看是否对该项服务(端口)的访问请求包或响应包给去掉了。网上银行出于安全考虑,使用的是 S-HTTP 协议(https://),其服务端口为 TCP 443。

另外也要结合网络拓扑图,考虑和检查三层网络设备上相关的路由是否已设置,以及设置是否正确,特别要注意响应包的返回路由是否已添加。很多时候是添加了数据包出去的路由,而忘了添加响应包回来时的路由,使响应包回到局域网边界设备后,由于缺乏返回方向的路由,导致 TCP 连接无法建立而造成访问失败。

为了在通信时不致发生混乱,就必须把端口号和主机的 IP 地址结合在一起使用。一个 TCP 连接由它的两个端点来标志,而每一个端点又是由 IP 地址和端口号来决定的,因此在分析 TCP 连接建立过程时,常常要结合 IP 地址与端口号。

[**例 1.1**] 现有某单位组建了局域网,网络边界设备为一台路由器,连接因特网的接口地址为 222.177.59.47。通过在路由器上配置网络地址转换(NAT),可实现代理服务器的功能,从而解决局域网内用户访问因特网的应用需求。该单位还要组建一台 Web 服务器,由于没有多余的公网 IP 地址给 Web 服务器使用,因此 Web 服务器使用内网的私有地址 192.168.252.10,然后在路由器上配置端口地址转换,将对 222.177.59.47 地址的 TCP 80 端口的访问,映射为对 192.168.252.10 主机的 TCP 80 端口的访问。这样,位于因特网中的用户,就可通过访问 222.177.59.47 地址的 TCP 80 端口(http://222.177.59.47)实现对该单位 Web 服务器的访问。事实证明,该解决方案是完美可行的。现在的问题是,该单位的内网用户使用"http://222.177.59.47"访问本单位的 Web 服务器时,却发现无法访问,而因特网用户访问正常,试分析内网用户无法访问的原因。

分析:外网用户访问 Web 服务器正常,排除服务器本身的故障。下面从 TCP 建立连接的过程进行分析查找。

为便于分析和表达,假设访问请求的客户机 IP 地址为 192.168.2.10。为简化示意图,路径中的中间设备忽略,示意图中使用"S:"代表源主机,"D:"代表要访问的目标主机。

TCP 建立连接时,客户端会随机选择一个未用的较低的端口(≥1024),与服务器的目的端口(TCP 80)建立连接。假设客户端使用 TCP 1025 号端口。

① 客户端发出访问请求包,包中的源地址和源端口为 192.168.2.10:1025,目标地址和目标端口为 222.177.59.47:80。

② 访问请求包通过路由和转发,最终到达网络的边界路由器。在路由器中,由于定义了端口地址转换,将对 222.177.59.47 地址的 80 端口的访问替换为对 192.168.252.10 主机的 80 端口的访问,经过端口地址转换后的访问请求包的目标主机就发生了变化。

③ 通过路由和转发,访问请求包最终到达目标主机 192.168.252.10。目标主机对访问请求进行响应,生成响应包。该响应包中的源主机为 192.168.252.10,响应端口为 80,即 192.168.252.10:80,目标主机和端口为 192.168.2.10:1025。

④ 通过网络内部三层交换机的路由转发,响应包到达客户端,但此时的响应包不会

通过路由器转发，且此时收到的响应包并不是客户机期望收到的响应包。这是因为访问请求的目标主机是 222.177.59.47，正确的响应包应是来自 222.177.59.47 主机的响应包，即响应包中的源主机地址应是 222.177.59.47，而现在收到的响应包中的源地址是 192.168.252.10，且客户机又没有发出对 192.168.252.10 主机的访问请求，因此会被认为是无效的包而被丢弃，而期望收到的回应包是永远无法收到的，因此 TCP 连接无法建立，访问失败。从中也可以看出，响应包未按原路径返回，被重新定向了。

整个访问过程中，数据包中的源主机和源端口以及目标主机和目标端口的变化如图 1.12 所示。

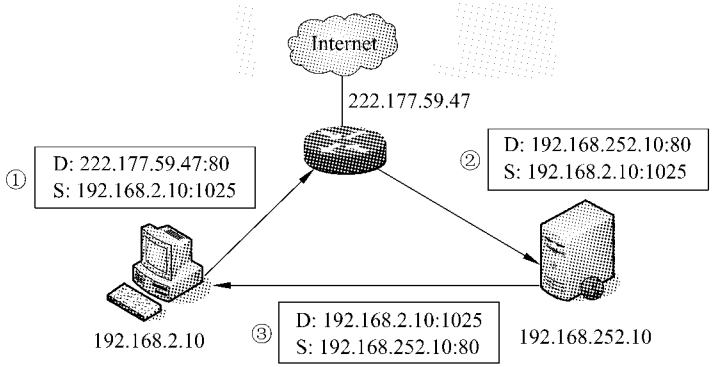

图 1.12　访问过程中源与目的 IP 地址和端口变化

因此，内网用户要访问自己的 Web 服务器，不能采用公网地址访问，只能采用 Web 服务器的内网地址访问，即采用 http://192.168.252.10 形式访问。

Web 服务器通常会注册域名的。采用域名方式访问，域名注册时，域名解析为的 IP 地址肯定是 222.177.59.47 这个公网地址，否则因特网用户将无法访问。这样，内网用户就不能采用域名来访问自己的 Web 服务器。除了采用内网 IP 地址方式访问外，有没有什么办法使内网用户也能采用域名方式访问自己的 Web 服务器呢？

解决办法是，在局域网内部配置一台 DNS 服务器，假设 DNS 服务器的 IP 地址为 192.168.252.254。在该 DNS 服务器中，将自己单位 Web 服务器的域名解析为 192.168.252.10，然后设置 DNS 服务器的转发器地址，将转发器设置为当地的、位于因特网中合法的 DNS 服务器的 IP 地址。然后，对于局域网中的所有用户主机配置首选 DNS 服务器的地址为 192.168.252.254，即使用自己的 DNS 服务器进行域名解析，这样就可完美解决该问题。

1.5.2　IP 协议

1. IP 协议简介

IP(Internet Protocol，因特网协议)协议是负责网络互联的网络层的核心协议，也是 TCP/IP 体系中最主要的协议之一。IP 协议具有分组与重新组装、寻址和路由的功能。

IP 协议提供一种无连接的传输机制，在发送数据时，IP 协议将数据进行分割，封装成 IP 数据包在网络中传输。每个 IP 数据包都作为独立的单元来对待，根据 IP 数据包中的

目标网络地址进行路由转发,以将IP数据包送达到目标主机。IP数据包全部到达目标主机后,再进行重新组装还原。

无论传输层使用何种协议,都要依赖IP协议来发送和接收数据。IP协议不保证数据传输的可靠性,其可靠性由传输层的TCP协议负责。

2. IP数据包的格式

IP数据包由首部和数据两部分构成。IP首部由固定部分和可变部分组成,固定部分总共为20个字节,可变部分最多为40个字节。最常用的首部长度为20个字节(不使用任何可选项)。IP数据包的格式如图1.13所示。

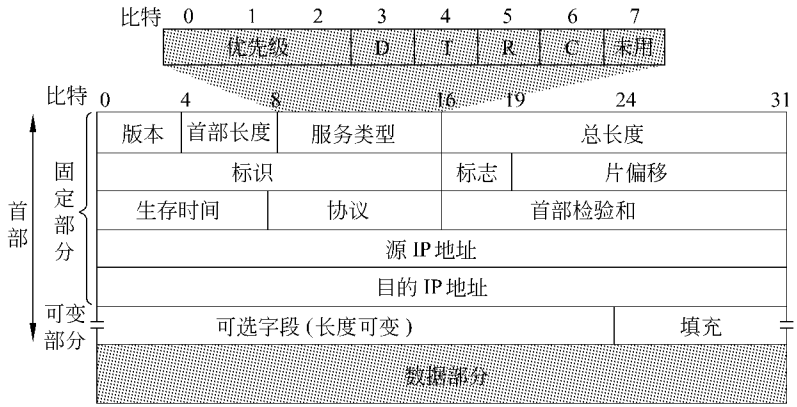

图1.13 IP数据包的格式

- 版本 占用4bit,指定IP协议的版本,目前常用的是IPv4版,IPv6是发展方向。
- 首部长度 占用4bit,可表示的最大值为15个单位,每个单位代表4个字节,因此IP的首部长度最大值为60字节。数据部分在4字节的整数倍时开始。
- 服务类型 占用8bit,前3个比特代表优先级,因此IP数据包具有8个优先级。D比特表示要求有更低的时延,T比特表示要求有更高的吞吐量,R比特表示要求有更高的可靠性,C比特表示要求选择代价更小的路由。
- 总长度 代表首部和数据之和的长度,单位为字节。由于总长度占用16bit,因此IP数据包的最大长度为65536字节,即64KB。
- 标识 占16bit,是一个计数器,用来产生数据包的标识。当数据包的长度超过网络允许的MTU值时,就必须对IP数据包进行分片传输。分片时,这个标识字段的值就会被复制到所有IP数据包片的标识字段中。在接收端对各分片的IP数据包进行重装还原时,就根据该标识字段的值来识别,具有相同标识字段值的IP分片包组装在一起。
- 标志 占3bit,目前只有低2位的比特有意义。最低位代表MF(More Fragment),当该位为1时,表示后面还有分片;该位为0时,表示这是若干个数据包片中的最后一个。中间一位为DF(Don't Fragment)标志位,代表不允许分片。只有DF位为0时,才允许分片传输。