高等学校计算机网络与通信技术教材

下一代因特网的移动支持技术

鲁士文　编著

清华大学出版社 · 北京交通大学出版社

高等学校计算机网络与通信技术教材

下一代因特网的移动支持技术

鲁士文　编著

清华大学出版社
北京交通大学出版社
·北京·

内 容 简 介

本书主要介绍 IPv6 和移动 IP 的基本概念、原理、关键技术和发展趋势，重点是 IPv6 支持无缝移动的技术。全书共分为 10 章，主要内容包括因特网和移动 IP 基本原理、IPv6、网络安全性基础技术、移动 IPv6 和路由优化、移动 IPv6 的安全机制、无线网络、移动 IPv6 切换过程、移动 IPv6 的快速切换、层次式移动 IPv6 和 IPv6 移动技术的未来发展。每一章都采用较为通俗易懂的描述和具有实际意义的例子及图表来说明相关原理、标准和核心技术。

本书融原理、技术和发展为一体，注重介绍新技术和系统设计方法，以提高读者从事研究工作和解决实际问题的能力为主要目标，可供高等学校和科研单位信息技术相关专业的研究生作为学习移动通信和计算机网络课程的参考书，也可作为开展相关研究课题的参考资料。另外，本书还可供从事通信网络研究和应用开发的人员作为了解通信网络的新发展或知识更新的一个媒介。

图书在版编目(CIP)数据

下一代因特网的移动支持技术/鲁士文编著. —北京：清华大学出版社；北京交通大学出版社，2007.5
（高等学校计算机网络与通信技术教材）
ISBN 978－7－81082－962－5

Ⅰ. 下…　Ⅱ. 鲁…　Ⅲ. 计算机网络－传输控制协议－高等学校－教材　Ⅳ. TN915.04

中国版本图书馆 CIP 数据核字（2007）第 031664 号

责任编辑：谭文芳
出版发行：清华大学出版社　邮编：100084　电话：010－62776969　http://www.tup.com.cn
　　北京交通大学出版社　邮编：100044　电话：010－51686414　http://press.bjtu.edu.cn
印 刷 者：北京东光印刷厂
经　　销：全国新华书店
开　　本：185×260　印张：13.5　字数：346 千字
版　　次：2007 年 5 月第 1 版　2007 年 5 月第 1 次印刷
书　　号：ISBN 978－7－81082－962－5/TN·51
印　　数：1～4000 册　定价：24.00 元

前　言

因特网给人们的工作和生活带来了重大的变革，它的发展已经成为国家信息化和现代化建设的一个重要内容。与此同时，随着无线接入技术的改进而发展起来的移动通信可为人们提供随时随地上网的便利。目前这两个最成功的领域正在融合，技术前景十分看好。特别是跟现在的 IPv4 相比，下一代因特网协议 IPv6 除了在地址空间、安全性和服务质量方面具有显著优势外，它也可为移动性提供更彻底、更全面的支持，相关研究工作的进展、成果、技术和标准受到了国内外学术界和工业界的广泛关注。

本书力图全面、系统、深入地向读者介绍 IPv6 和移动 IP 的基本概念、原理、关键技术和发展趋势，重点是 IPv6 支持无缝移动的技术。全书共分为 10 章。第 1 章简要介绍因特网和移动 IP 基本原理。第 2 章和第 3 章介绍 IPv6 和网络安全性基础技术。第 4 章和第 5 章重点阐述移动 IPv6 的路由优化和安全机制。第 6 章介绍无线网络。第 7～9 章分别描述移动 IPv6 的切换过程、快速切换和层次式移动。第 10 章讨论 IPv6 移动技术的未来发展。每一章都采用较为通俗易懂的描述和具有实际意义的例子及图表来说明相关原理、标准和核心技术。

本书是作者在多年计算机网络教学和科研工作的基础上编写的，融原理、技术和发展为一体，注重介绍实用技术和培养系统设计能力，以提高读者从事研究工作的能力和解决实际问题的能力为主要目标。

本书可供高等学校和科研单位信息技术相关专业的研究生作为学习移动通信和计算机网络课程的参考书，也可作为开展相关研究课题的参考资料。另外，本书还可供从事通信网络研究和应用开发的人员作为了解通信网络的新发展或知识更新的一个媒介。

限于时间与水平，不当之处欢迎批评指正。

作　者

2007 年 3 月

于中科院研究生院

目　录

第 1 章　因特网和移动 IP 基本原理

本章首先阐述因特网(Internet)设计和操作的基本原理,包括 TCP/IP 协议(Transmission Control Protocol/ Internet Protocol,传输控制协议/网际协议)体系、对在该协议体系内的不同层次协议的分析、名字服务,以及对因特网路由技术的综述。虽然所给出的概念是基于 IPv4(因特网协议,第 4 版),但就运行机制而言,大多数内容也适用于 IPv6(因特网协议,第 6 版)。本章的后一部分介绍在因特网上移动 IP 的一般概念和设计目标,以及它在当前 IPv4 网络环境下实现的技术途径。

1.1　因特网协议体系

根据目前流行的观点,可以把计算机网络定义为:按照网络协议,以共享资源为主要目的,将地理上分散且独立的计算机互相连接起来形成的集合体。基本的通信硬件包括在计算机之间传送位串序列的机制。但是,仅仅使用硬件来进行通信就好像用 0 和 1 二进制编程那样难以实现。为了方便网络程序设计,计算机通常都是连到使用复杂软件的网络上。这些软件为应用程序提供了方便的高层接口,自动处理大多数底层的通信细节和问题。因此,大多数应用程序依靠网络软件进行通信,并不直接与网络硬件打交道。

网络中的通信是指在不同系统中的实体之间的通信。所谓实体,是指能发送和接收信息的任何东西,包括终端、应用软件、通信进程等。跟在人与人之间交流一样,实体之间通信需要一些规则和约定,例如,传送的信息块采用何种编码和怎样的格式? 如何识别收发者的名称和地址? 传送过程中出现错误如何处理? 发送和接收速率不一致怎么办? 简单地讲,通信双方在通信时需要遵循的一些规则和约定就是协议。协议主要由语义、语法和定时三部分组成,语义规定通信双方准备“讲什么”,亦即确定协议元素的种类;语法规定通信双方“如何讲”,确定数据的信息格式、信号电平等;定时则包括速度匹配和排序等。

两个系统中实体间的通信是一个十分复杂的过程,为了减少协议设计和调试过程的复杂性,大多数网络的实现都按层次的方式来组织,每一层完成一定的功能,每一层又都建立在它的下层之上。不同的网络,其层的数量和各层的名字、内容和功能不尽相同。然而在所有的网络中,每一层都是通过层间接口向上一层提供一定的服务,而把这种服务是如何实现的细节对上层加以屏蔽。服务在形式上是由一组原语(Primitive)来描述的。这些原语供用户和其他实体访问该服务时调用。它们请求服务提供者采取某些行动或报告某个对等实体的活动。

更具体地讲,层次结构包括以下几个含义。

① 第 n 层的实体在实现自身定义的功能时,只使用 n - 1 层提供的服务。

② n 层向 n + 1 层提供服务,此服务不仅包括 n 层本身所执行的功能,还包括由下层服务提供的功能总和。

③ 最低层只提供服务,是提供服务的基础;最高层只是用户,是使用服务的最高层;中间

各层既是下一层的用户，又是上一层服务的提供者。

④ 仅在相邻层间有接口，而且下层提供的服务的具体实现细节对上层完全屏蔽。

应该指出，服务和协议是完全不同的概念。服务是各层向它的上层提供的一组原语。尽管服务定义了该层能够为它的上层完成的操作，但丝毫未涉及这些操作是如何完成的。服务定义了两层之间的接口，上层是服务用户，下层是服务提供者。

与之相对比，协议是定义在相同层次的对等实体之间交换的帧、分组和报文的格式及含义的一组规则。实体利用协议来实现它们的服务定义。只要不改变提供给用户的服务，实体可以任意地改变它们的协议。这样，服务和协议就被完全地分离开来。

TCP/IP 协议当初是为美国国防部研究计划局(DARPA)设计的，其目的在于让各种各样的计算机都可以在一个共同的网络环境中运行。TCP/IP 协议的形成有一个过程。1969 年初建的 ARPANET 主要是一项实验工程；70 年代初，在最初建网实践经验基础上，开始了第二代网络协议设计工作，称为网络控制协议(Network Control Protocol ，NCP)。70 年代中期，国际信息处理联合会进一步补充和完善了 NCP 的开发工作，从而导致了 TCP/IP 协议的出现。80 年代初，美国伯克利大学将 TCP/IP 设计在 UNIX 操作系统内核中，1983 年美国国防部 DOD 宣布，将 ARPANET 的 NCP 协议完全过渡到 TCP/IP 协议，成为正式的军事标准。与此同时，SUN 公司将 TCP/IP 协议引入了广泛的商业领域。

现在，TCP/IP 协议已成为一个完整的协议簇，一个网络体系结构。该协议簇除了传输控制协议 TCP 和因特网协议 IP 之外，还包括多种其他协议，其中有工具性协议，管理性协议及应用协议等。

TCP/IP 协议现在非常受重视，这是由于以下几方面原因。

第一，TCP/IP 协议最初是为美国 ARPANET 设计的，后来在 ARPANET 发展成为国际性的互联网(因特网)时，TCP/IP 仍是网际通信协议。几十年的开发与研究，TCP/IP 已充分显示出它的强大连网能力与对多种应用环境的适应能力。当前在以 ARPANET、MILNET 和美国国家科学基金会的 NSFNET 作为主干网的基础上，因特网已成了用 TCP/IP 协议连接世界各国、各部门、各机构计算机网络的最大的国际互联网。

第二，因特网在美国、欧洲的科学界、教育界、商业界及政府部门、军事部门等领域影响巨大。TCP/IP 协议已被各界公认为是异种计算机、异种网络彼此通信的重要协议，也是目前最为可行的协议。

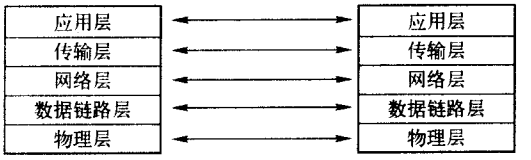

图 1-1 因特网的分层协议结构

第三，各主要计算机公司和一些软硬件厂商的计算机网络产品几乎都支持 TCP/IP 协议，TCP/IP 协议现在已成为事实上的国际标准和工业标准。

如图 1-1 所示，因特网协议集被划分成 5 个层次。

(1) 物理层

负责在有线或无线接口上发送和接收数位。在一些无线物理层中执行的功能包括调制/解调、功率测量和同步。一个设备可以有多个物理接口。

(2) 数据链路层

这一层负责链路控制和介质访问控制(Medium Access Control，MAC)。其他功能包括错

误检测和错误纠正。某些复杂的无线链路层还以在无线链路上加密的形式提供安全性。以太网是有线网络数据链路层的一个范例。

(3) 网络层

也称 IP 层,它向上层提供连接性,负责把分组从一个结点转发到另一个结点。IP 地址标识在因特网上的一台计算机。在因特网协议体系中,网络层包括 IP 和 ICMP(Internet Control Message Protocol,因特网控制报文协议)。IP 向上层提供非可靠的转发服务,也就是说,不能保证被发送的分组一定能够到达目的地,没有错误检测或纠正。另一方面,ICMP 用以在两个设备之间发送信令信息。术语信令指的是有关在两个设备之间连接的信息或者关于在这两个设备之间发送数据时所碰到的错误的信息。也就是说,信令包含不属于在两个设备之间转移的数据的信息。

(4) 传输层

负责控制在通信实体之间的数据流。传输层地址(传输协议和端口号)区分在同一台计算机上不同的应用层实体。在因特网协议簇中有 3 个不同的传输协议:传统 TCP 协议、用户数据报协议(User Datagram Protocol, UDP)和最近开发的流控制传输协议(Stream Control Transmission Protocol, SCTP)。TCP 向应用层提供可靠的传输服务,为此,它包括检测分组丢失、延迟和差错的功能,它使用重传机制保证可靠投递。相反,UDP 是非可靠的传输协议,它不提供投递保证,没有错误纠正。UDP 简单地在应用层和 IP 层之间起到一个接口的作用。对于一些实时应用,少量的信息丢失是可以容忍的,但不能容忍由重传而引起的延迟。采用 UDP 而非 TCP 的另一个可能的原因是一些应用程序为了满足特别应用的特别需求,想要执行自己的流控制机制。

与 TCP 相同,SCTP 向应用提供可靠的传输服务;所不同的是,它提供了更加适用于传统的电话信令应用的新特征;它还包含多流功能,允许它把同样的流有效地发送到多个接收方。

(5) 应用层

负责从用户或机器得到输入数据,把它们格式化,然后通过较低层发送。在因特网上最常用的应用包括文件传送协议(File Transfer Protocol,FTP)、执行远程上机的 TELNET(TELephone NETwork)、用于电子邮件的简单邮件传送协议(Simple Mail Transfer Protocol,SMTP)和用于万维网浏览的超文本传送协议(Hypertext Transfer Protocol,HTTP)。

在如图 1-1 所示的层次模型中,每一层都使用一个或多个协议,与位于同一层的远方对等实体通信。例如,TCP 允许在传输层上对等实体之间的通信,而 IP 允许在网络层上对等实体之间的通信。

层次模型的优点是每一层的修改或扩充不影响其他层,每一层都独立于其他层。例如,应用层不知道关于分组转发、链路层控制或信号调制方面的情况,它们是在不同的层次执行的。类似地,网络层和传输层不需要知道关于一个特别的应用的内容,但这不妨碍它们执行自己所承担的任务。

1.2　使用 IP 的网络互连

因特网的设计有两个基本的出发点:一是没有一个物理网络能够为所有用户服务,二是用户希望通用的互连。

第一个观点是技术方面的。提供最高速度通信的局域网在地理跨度上受限;广域网跨越大的距离,但不能提供高速的连接。任何单个网络技术都不能满足所有的需要,因此只能考虑包容多种基础网络的硬件技术。

第二个观点是显而易见的。说到底,希望在任何两点之间都能通信,特别是,希望一个通信系统不受物理网络边界的限制。

因特网的目标是要进行统一的合作网络的互连,支持一种通用的通信服务。在每一个网络内部,计算机使用依赖于技术的基础通信设施。在依赖于技术的通信机制和应用程序之间插入网际互连软件,它隐藏了低层的细节,使得集成网络看起来像是单个大的网络。这样一种互连方案就称为网际互连,所形成的网络称为因特网。

尽管位于源和目的地之间的中间网络与源发主机和目的计算机都没有直接的连接,但用户希望能够通过中间网络发送数据。为了建立一个有生命力的互联网,需要某些计算机能把报文自动以分组(IP 数据报)形式从一个网络转发到另一个网络。互连两个网络并且将报文分组从一个网络传递到另一个网络的计算机称为网关或路由器。

图 1-2 所示为一个由 3 个物理网络构成的互联网。图中,网关 1 将来自网络 1 的所有分组(其目的主机位于网络 2 或网络 3 上)传送到网络 2,并且将来自网络 2 或网络 3 的所有分组(其目的主机位于网络 1 上)从网络 2 传递到网络 1 上,网关 2 执行与网关 1 类似的操作。

图 1-2 三个网络用两个网关互连

在一个 TCP/IP 网络上,称为网关的计算机提供所有的物理网络之间的互连。网关负责为报文分组选择路由,送往目的地。网关进行的路由选择是基于目的网络,而不是基于目的主机。既然以网络为基础进行路由选择,那么网关需要保存的信息量就与 TCP/IP 网络中的网络数量成比例,而不是与互联网上的计算机数目成比例。IP 协议同等地看待所有网络,不论是以太网这样的局域网,X-25 这样的广域网,或者仅仅是在两台机器之间的一条点到点的链路,每一个都算作一个网络,本书以后会介绍,就 IP 地址而言它们都会被分配一个网络号。

在本质上,TCP/IP 定义了一种抽象"网络",隐藏了物理网络的细节。正是这种抽象概念,使得因特网协议非常强壮。

虽然概念性的协议分层结构图不能够表示所有的细节,但它有助于解释总体思想。图1-3 说明一个报文通过 3 个网络所用到的协议软件层次。在网关中只画了网络接口和 IP 协议层,因为网关在接收、路由选择和发送 IP 分组时仅需要这些层次。应该指出,任何一台连接到两个网络的机器都必须有两个网络接口模块。

如图 1-3 所示,源机器上的一个发送程序发出一个报文,IP 层把这个报文放在一个数据报中通过网络 1 发送。在中间机器上,该数据报(或称 IP 分组)向上到达 IP 层,在这里进行路由选择,IP 又把它在一个不同的网络上送出。仅当到达最后目的地机器时,IP 才抽出报文,并把报文传给协议软件的较高层。

因特网的网络层协议是 IP 协议。与所有其他的网络层协议一样,IP 协议把信息分组从

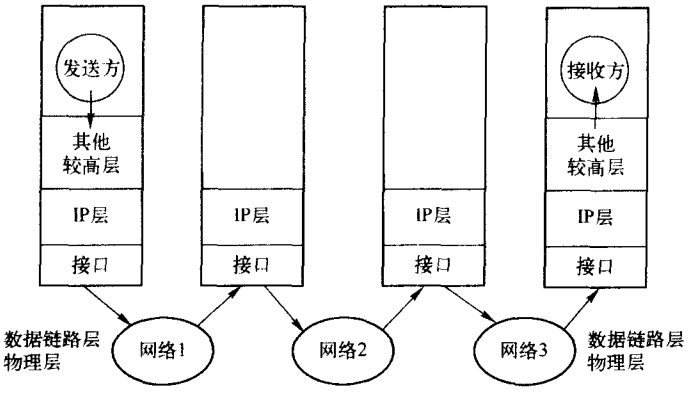

图 1-3　一个报文通过 3 个网络的结构

作为起点的源移动到最终的目的地。IP 协议最基本的服务是提供一个非可靠的尽最大努力去完成任务的、无连接的分组投递系统。非可靠，是因为所要求的投递不能保证成功，分组可能丢失，投递无序或重复投递，而 IP 协议并不检测这些情况，发生这些情况也不通知发送者或接收者。无连接，是因为每一个分组的处理都独立于其他分组，从一个机器发出的一串分组，可以经由不同的路径到达另一机器，也可能部分分组丢失了，而其余的仍被投递。说这种服务是尽最大努力完成的，是因为 IP 尽最大努力去投递分组，并不轻易地抛弃分组，仅在资源用尽或下面的物理网络失效时才会发生不可靠的现象。

IP 协议定义数据传送的基本单元——IP 分组及其确切的数据格式。IP 协议也包括一套规则，指明分组如何处理，错误怎样控制。特别是，IP 协议还包含非可靠投递及与此关联的分组路由选择的思想。

在一个物理网络上传递的单元是帧，它包含头和数据，头给出了源和目的地址类的信息，而因特网称基本传递的单元为 IP 分组。类似典型的物理网络帧，IP 分组也分头和数据区。分组的头包含源和目的地址。当然，不同点在于 IP 分组包含的是 IP 地址。IP 分组可以为任意长度(对于 IPv4，最大长度是 2^{16}字节)，当它们从一台机器移动到另一台机器时，必须被放在物理网络的帧中传输。

IP 模块是 TCP/IP 技术的核心，而 IP 模块的关键成分则是它的路由表。路由表放在内存储器中，IP 模块使用它为 IP 分组选择路由。

1.3　IP 地址

在当前使用的因特网上，IP 地址是分配给一个结点的每个网络接口的 32 位(4 字节)数字。一个结点理论上是通过一个接口连接到一条链路的。像路由器这样的多个网络接口的结点有多个 IP 地址，每个接口都有一个地址。

IP 地址通常书写成点分十进制的形式，4 个字节中的每一个都分别用一个十进制的整数表示，并用在英文中通常是表示句号的点分隔。例如，一个用十六进制数 C0 13 F1 12 表示的 IP 地址写成点分十进制的形式是 192.19.241.18，因为 C0(十六进制) = 192(十进制)，13(十六进制) = 19(十进制)，F1(十六进制) = 241(十进制)，12(十六进制) = 18(十进制)。

IP 地址由两部分构成，即网络前缀部分和主机部分。网络前缀是一个位串，连接到同一

条链路(即同一个物理网络,例如以太网链路、PPP(Point to Point Protocol,点对点)链路、X.25虚电路、帧中继虚电路、ATM 虚电路等)上的所有主机的地址都具有相同的前缀,而每台主机地址的其余部分即主机部分必须具有唯一性。因此可以用网络前缀标识一条链路,而用主机部分标识在那条链路上的特别的主机。前缀长度被定义成组成 IP 地址前缀的位的数目。显然,在 IPv4 中,组成 IP 地址主机部分的位的数目等于 32 减去前缀长度的差。

一个 IP 地址及其前缀的表示方法是"地址/前缀长度"。例如,可以把前缀长度是 24 的 IP 地址 129.61.18.26 表示成"129.61.18.26/24",在这里网络前缀是 129.61.18,主机部分是 26。

人们约定,在网络前缀部分的二进制编码为全零时,该网络号被解释为本地网;在主机部分的二进制编码为全 1 时,该主机号被解释为本地网络内的广播地址;主机部分等于零的 IP 地址从不分配给单个主机。取而代之的是用主机部分等于零的 IP 地址指称网络本身。

在 IPv4 中,第 1 字节的值等于 127(十进制)的地址是为回送保留的,用于本地机器上的测试和进程间通信,当任何程序使用回送地址 127.X.X.X 发送数据时,计算机中的协议软件就将数据返回,不在任何网络上传输。

另外,开头 4 个二进制位等于 1110 的 IPv4 地址用于表示多播中的组地址,后随的 28 位是具体的组的标识。例如,地址 224.0.0.1 永久地分配给"所有主机组",该组包括参与 IP 多播递交的所有主机和路由器,一般地讲,"所有主机组"地址用于在一个本地网上到达参与 IP 多播递交的所有机器。然而,没有任何 IP 多播地址可以指称在因特网上的所有主机。另外,IP 多播地址只能用作目的地址,它们绝不能出现在一个 IP 分组的源地址段中。

1.4 IP 分组

IP 协议的目的是提供必要的功能,使 IP 分组逐个地从源发主机通过网络互连系统传递到目的主机。IP 分组也称 IP 数据报,它是以无连接方式通过网络传输的。无连接的意思就是指在数据传输之前源结点与目的结点并不建立连接。

在 IP 分组的传递过程中,不管行走多长的距离,或跨越多少个物理网络,IP 模块的寻址机制和路由选择功能都能把数据送到正确的目的地。所经过的各个物理网络可能采用不同的链路协议和帧格式,但是,无论是在源主机和目的主机中,还是在路过的每个路由器中,网络层都使用始终如一的协议(IP 协议)和不变的分组格式(IP 分组)。

如图 1-4 所示,IP 分组头的长度是 32 位(4 字节)的整数倍。从任选项往后是可变长部分,这一部分也可以没有,以下是对分组头中各个段的解释。

4 位的版本号段表示协议支持的 IP 版本号。

4 位的 IP 分组头长表示 IP 分组头的长度,以 32 个二进制位(4 个字节)为单位,取值的范围是 5~15(默认值是 5)。由于 IP 分组头的长度是可变的,所以这个段是必不可少的。

8 位的服务类型段说明分组所希望得到的服务质量。它允许主机指定在网络上传输分组的服务种类,也允许选择分组的优先级,以及希望得到的可靠性和资源消耗。该段的目的是请求网络提供所希望的服务。

16 位的总长度段给出 IP 分组的总长度,单位是字节,包括分组头和数据的长度。数据段的长度可以从总长度减去分组头长度计算出来。由于总长度段有 16 位,所以最大 IP 分组允许有65 535 个字节。IP 规范规定,所有主机和路由器至少能支持 576 字节的分组长度。如果

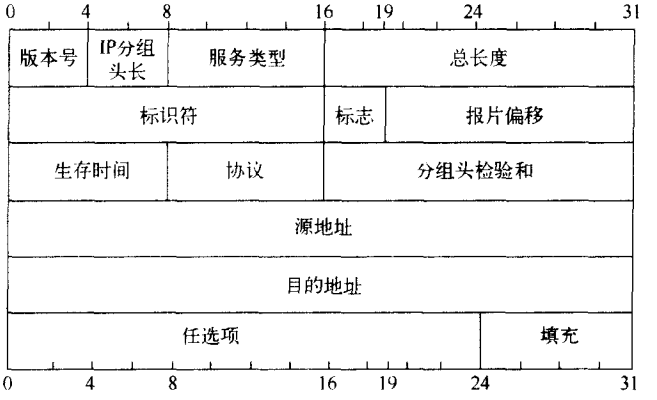

图 1-4　IP 分组头

IP 分组在网络传送过程中被分成报片，那么分片后形成的 IP 分组中的总长度段指的是单个报片的总长度，而不是原先 IP 分组的总长度。

16 位的标识符段包含一个整数，唯一地标识 IP 分组。IP 分组在传输时，其间可能会通过一些子网。这些子网允许的最大协议数据单元(PDU)长度可能小于该 IP 分组的长度。为了处理这种情况，IP 为以数据报方式传送的 IP 分组提供了分片和重组的功能。这也正是 IP 模块的主要功能之一。当一个路由器分割一个 IP 分组时，要把 IP 分组头中的大多数段的值复制到每个报片中，标识符段必须复制。它的基本目的是使目的主机知道到达的报片属于哪个 IP 分组，源主机必须为发送的每个 IP 分组分别产生一个唯一的标识符段值。为此，IP 软件在计算机存储器保持一个全局计数器，每建立一个 IP 分组就加 1，再把结果放到 IP 分组标识符段中。

3 位的标志段含有控制标志。3 位中的低序 2 位控制 IP 分组的分片，这 2 位分别称为不可分片位和还有报片位；高序位没有被使用。当不可分片位置 1 时，规定不要将 IP 分组分片。仅在完整的 IP 分组有用的情况下，应用程序才可选择禁止分片。标志段的“还有报片”位标明这个报片包含的数据是取自原始 IP 分组的中间，还是取自原始 IP 分组的最后。一旦报宿收到一个报片，如果它的“还有报片”位置 0，就知道这个报片中的数据取自原始分组的尾部。

13 位的报片偏移段标明当前报片在初始 IP 分组中的位置。为了重组 IP 分组，报宿必须得到从偏移 0 开始，直到最高偏移值之间的所有报片。这些报片不需要按顺序到达，接收报片的报宿与分割 IP 分组的路由器之间不进行通信，报宿也能重新组合 IP 分组。报片偏移以 64 位(8 个字节)为单位，取值范围 0～8191，默认值是 0。

8 位的生存时间段指定 IP 分组能在互联网中停留的最长时间，以秒为单位。当该值降为 0 时，IP 分组就应被舍弃。该段的值在 IP 分组每通过一个路由器时都减去 1。该段决定了源发 IP 分组在网上存活时间的最大值，它保证 IP 分组不会在一个互联网中无休止地往返传输。

8 位的协议段表示哪一个高层协议将用于接收 IP 分组中的数据。高层协议的号码由 TCP/IP 中央权威管理机构予以分配。例如，该段值的十进制表示对应 ICMP 协议是 1，对应 TCP 协议是 6，对应 UDP 协议是 17，对应 ISO 传输层协议第 4 类(ISO-TP4)是 29。

16 位的分组头检验和段保证 IP 分组头值的完整性，当 IP 分组头通过路由器时，分组头发生变化(例如生存时间段值减 1)，检验和必须重新计算。检验和的计算十分简单：首先，在计算前将检验和段的所有 16 位均置成 0；然后 IP 分组头从头开始每两个字节为一个单位相加，若相加的结果有进位，那么将和加 1；如此反复，直到所有分组头的信息都相加完为止，将最后的值对 1 求补，即得出 16 位的检验和。

32 位的源地址段包含发送 IP 分组的源主机的 IP 地址。32 位的目的地址段包含 IP 分组的目的主机的 IP 地址。

可变长的任选段提供了一种策略，允许今后的版本包含在当前设计的头中尚未出现的信息，也避免使用固定的保留长度，从而可以根据实际需要选用某些头部登记项。

填充段是为了使有任选项的 IP 分组满足 4 个字节长度的整数倍而设计的，通常用 0 填入填充段来满足这一要求。填充段的有无或所需要的长度取决于选择项的使用情况。

1.5 IP 路由选择

当主机上的一个应用程序要进行通信时，TCP/IP 协议产生一个或多个 IP 分组。人们把位于同一 IP 网上(其 IP 地址具有相同的标识链路的网络前缀)的两台计算机之间的通信称为直接路由通信，把位于不同 IP 网上的两台计算机之间的通信称为间接路由通信。

IP 模块从高层接收数据，形成 IP 分组后，必须决定是直接还是间接发送该分组，并且选择一个低层网络接口，这些选择在访问路由表后作出。对于存在多个网关情况下的间接路由通信，主机还必须决定把 IP 分组送给本地网络上的哪个网关，因为没有一个网关能对所有目的地提供最好路径。当然，网关是要做路由选择决定的，这是它们的主要任务，这也是它们同时被称为路由器的原因。

对于一个从底层接口收到的 IP 分组，IP 模块必须决定是否要将该分组传给上层模块；如果该分组需要转发给其他计算机，则与在本计算机建立的分组进行同样的发送处理。但是，对于从网上收到的 IP 分组，无论何时都不能沿原网络接口转发回去。

通常，因特网路由算法在每个主机和路由器上都采用一张路由选择表，该表包含可能的目的地信息。当一个 IP 分组到达一个网关时，IP 软件就找到目的 IP 地址，抽出网络号，然后网关使用该网络标识决定路由。前面已经叙述过，IP 选择路由是基于目的网络号，而不是目的主机号。

在单个物理网络的两台主机之间，发送 IP 分组无须网关的转发功能，发送方将分组放在物理帧内直接传给目的主机。判断一个目的地址是否在与本计算机直接连接的一个物理网络上的方法是，发送者抽出目的 IP 地址的网络部分，与自己 IP 地址的网络部分比较，如果相同，就直接投递。

严格地讲，路由选择功能包括两个成分：路由选择智能和转发机制。所有的路由器都要实现这两部分功能。路由选择智能部分涉及在各个路由器之间的信息交流，每个路由器都把自己所掌握的路由信息告知与其交流的其他路由器，使得它们都可以通过最好的路由转发 IP 分组。在因特网上，这部分的功能是通过使用路由协议实现的。路由协议与应用程序一样也在用户空间运行。路由选择智能产生路由表，路由表为指定的目的地址列出下一跳段的链路或下一个路由器。

转发机制基于由路由表得到的信息。在这里，转发的含义是把分组发送给它为了到达最

终目的地而应该前往的下一跳段。每个路由器都把它的转发信息存储在一个路由表中。路由表确定前往每个目的地的下一跳段。如果一个结点在其路由表中对于一个给定的目的地址有多个路由，那么按照规则，必须使用具有最大前缀长度的路由。

作为例子，考虑一个结点要为一个目的地址是 7.7.7.1 的 IP 分组做转发决定。该结点的路由表包含如表 1-1 所示的登记项。为了说明方便，目的地址的网络前缀部分（由各自的前缀长度决定）用**粗体**表示。

表 1-1　示例路由表

目的地址/前缀长度	下 一 跳 段	接　口
7.7.7.99 /32	路由器 1	a
7.7.7. 0/24	路由器 2	a
0.0.0.0/0	路由器 3	a

为了在路由表中查找匹配的登记项，必须把表中列出的每个目的地址最左边长度等于前缀长度的位串（即前缀）跟分组的目的 IP 地址的最左边相同长度的位串比较。由于 7.7.7.99 不等于 7.7.7.1，所以第一个路由表项不匹配该 IP 分组。第二个路由表项仅需要分组的目的 IP 地址跟在该表项中列出的目的地址的最左边 24 位相同，由于二者开头 24 位都是 7.7.7，所以第二个路由表项匹配该 IP 分组。第三个路由表项表示它匹配所有的 IP 分组，当然也匹配该 IP 分组。在本例中有两个匹配路由，根据最大匹配长度原则，应该使用表中的第二个登记项转发分组，因此该结点把分组通过接口 a 转发给路由器 2。

在路由器之间使用路由协议时，每个路由器都可以通知其他路由器它所连接的链路的状态和对应的邻居结点，或者它到达一些目的结点的代价。代价是跟一个特别目的地相关的参数。该参数取决于多个因素，例如距离（以跳段数或延迟时间计算）、带宽等。这就允许路由器为到达一个目的地选择最好的通路，也允许它们在前往某个目的地的通路上时，在路由器失效的情况下能够动态地找到另一条替代路由。

因特网是在 ARPANET 的基础上发展起来的。当因特网实验开始时，它以 ARPANET 作为其主干。所谓核心网关（Core Gateway）的思想就是将局部网络连接到 ARPANET。对于路由选择来说，由单个管理当局控制的一组网络和网关称为自治系统（Autonomous System，AS），一个自治系统内的网关自由地选择它们自己的发现、传播、验证和检查路由一致性的机制。根据这一定义，核心网关也形成一个自治系统。

交换路由信息的两个网关如果属于两个不同的自治系统，那么就称它们是外部相邻；如果它们属于同一个自治系统，就称它们是内部相邻。外部网关用以通告可达性信息给其他自治系统的协议称为外部网关协议（Exterior Gateway Protocol，EGP）。内部网关用以在一个自治系统内部（也称域内）交换网络可达性和路由选择信息的任何算法都称为内部网关协议（Interior Gateway Protocol，IGP）。

边界网关协议（Border Gateway Protocol，BGP）是一个广泛接受的 EGP 标准，它是一个自治系统之间（也称域间）的路由协议，被用来在 BGP 路由器中间交换网络可达性信息。BGP 是一个通路向量协议，它通告前往目的地的一系列自治系统号，例如“路由 10.10.1.0/24 可以通过 AS1、AS2、AS6 和 AS7 到达”就是一个通路向量。在 BGP 路由器中间交换的信息允许一个路由器建立描述自治系统连接性的图。起初，BGP 对等通信交换整个 BGP 路由表。随后仅发