下一代的IP网络技术保障

——多协议标签交换

吴 伟 编著

清 华 大 学 出 版 社

（京）新登字 158 号

内 容 简 介

本书主要介绍多协议标签交换技术（MPLS）及其相关应用如 MPLS 流量工程、MPLS 虚拟专用网、GMPLS 等。本书第一部分介绍 IP 网络的基本概念及技术；第二部分介绍 MPLS 构架，包括标签及其分配、标签交换过程、LSP 控制等；第三部分介绍 MPLS 的应用，包括 MPLS 的一般应用、MPLS 虚拟专用网以及 MPLS 流量工程；第四部分介绍最近提出的广义多协议标签交换（GMPLS）。

本书既适合于具备一定通信技术知识的电信网络专业技术人员，也可以作为通信或相关专业学生的参考用书。

书　　名：下一代的 IP 网络技术保障——多协议标签交换
作　　者：吴伟　编著
责任编辑：吴宏伟
出 版 者：清华大学出版社（北京清华大学学研大厦，邮编 100084）
http://www.tup.tsinghua.edu.cn
印 刷 者：北京牛山世兴印刷厂
发 行 者：新华书店总店北京发行所
开　　本：787×960 1/16　**印张**：9.5　**字数**：175 千字
版　　次：2002 年 4 月第 1 版　2002 年 4 月第 1 次印刷
书　　号：ISBN 7-302-05184-4/TP・3044
印　　数：0001～5000
定　　价：14.00 元

前　言

近年来，通信技术迅速发展，其中 IP 技术和光网络技术尤为显著。可以毫不夸张地说，IP 改变了传统通信技术领域以及应用市场的格局。从目前 IP 网络发展速度和趋势看，虽然还存在争议，但人们已经接受了 Everything Over IP 以及 IP Over Everything 这两个概念，各种基于 IP 的业务以及为 IP 服务的下层网络技术纷纷出现，IP 技术大有一统天下的趋势。很多通信标准制定组织、世界性的设备供应厂商都非常看好 IP 技术，将 IP 作为优先发展以及研究的重点。

IP 技术和 Internet 相互推动，迅速发展。Internet 正在戏剧性地改变着我们的生活方式，我们的商业运作模式，甚至整个社会的思维方式。随着新应用需求的诞生，人们把更多的目光投注在宽带 Internet 和无线 Internet 之上。从窄带发展到宽带和无线 Internet，这是 Internet 发展的必经之路。宽带 Internet 带来了性能和速度的提高，为用户提供了更愉悦的网上感受，同时带宽的提高也为高带宽的应用奠定基础，例如网上的视频点播、实时音频/视频交流、海量数据异地存储等。作为新业务的技术支撑，IP 网络必须满足这种宽带、实时的要求，传统的 IP 技术很难达到这样的要求，IPv6、资源预留协议、多协议标签交换等转发技术以及波分复用/密集波分复用等传输技术应运而生。

本书主要介绍多协议标签交换技术（MPLS）及其相关应用如 MPLS 流量工程、MPLS 虚拟专用网、GMPLS 等。本书第一部分介绍 IP 网络的基本概念及技术；第二部分介绍 MPLS 构架，包括标签及其分配、标签交换过程、LSP 控制等；第三部分介绍 MPLS 的应用，包括 MPLS 的一般应用、MPLS 虚拟专用网以及 MPLS 流量工程；第四部分介绍最近提出的广义多协议标签交换（GMPLS）。

本书既适合于具备一定通信技术知识的电信网络专业技术人员，也可以作为通信或相关专业学生的参考用书。

目　录

第一部分　概述

第二部分　MPLS 构架

第一部分　概　　述

Internet 正在成为人们生活中越来越重要的组成部分，回顾过去，以美国为例，Internet 几乎是普及最快的一种电子技术：电视达到 30%的普及率用了 38 年，电话达到同样的普及率用了 17 年，个人计算机用了 13 年，而 Internet 仅仅用了 7 年。

根据 IDC（International Data Corporation）的统计和预测数据，Internet 上的 WWW 页面数量在 1998 年为 829 000 000，到 2000 年为 2 670 000 000，预计到 2002 年，WWW 页面数量将达到 7 700 000 000，增长态势如图 1 所示。

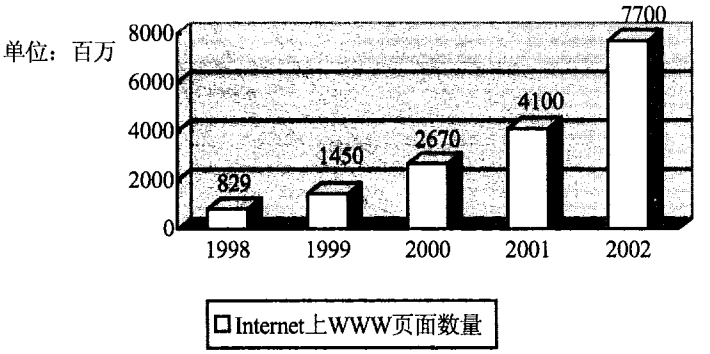

图 1　WWW 页面数量增长态势图

中国作为加入 Internet 的第 71 个国家，在 Internet 进入中国到现在短短的 14 年里，已经具有了相当的网络规模和用户数量。目前已经有上网计算机约 1 002 万台，其中专线上网计算机数为 163 万台，拨号上网计算机数为 839 万台。我国 Internet 用户人数也达到了约 2 650 万人，其中专线上网的用户人数为 454 万，拨号上网的用户人数为 1 793 万，同时使用专线与拨号的用户人数为 403 万，除计算机外同时使用其他设备（移动终端、信息家电）上网的用户人数为 107 万。

在 Internet 规模迅速膨胀的同时，它也已经从学术交流工具演变为商业工具，并且随着电子商务的迅速发展与完善，Internet 可能成为世界上最大的、效率最高的和

最安全的市场，也就是说，它将从目前的通信手段进一步演变成一个数字化的虚拟的商务平台。同样通过 IDC 的统计和预测数据，我们可以看到 Internet 商务的迅速发展，如图 2 所示。

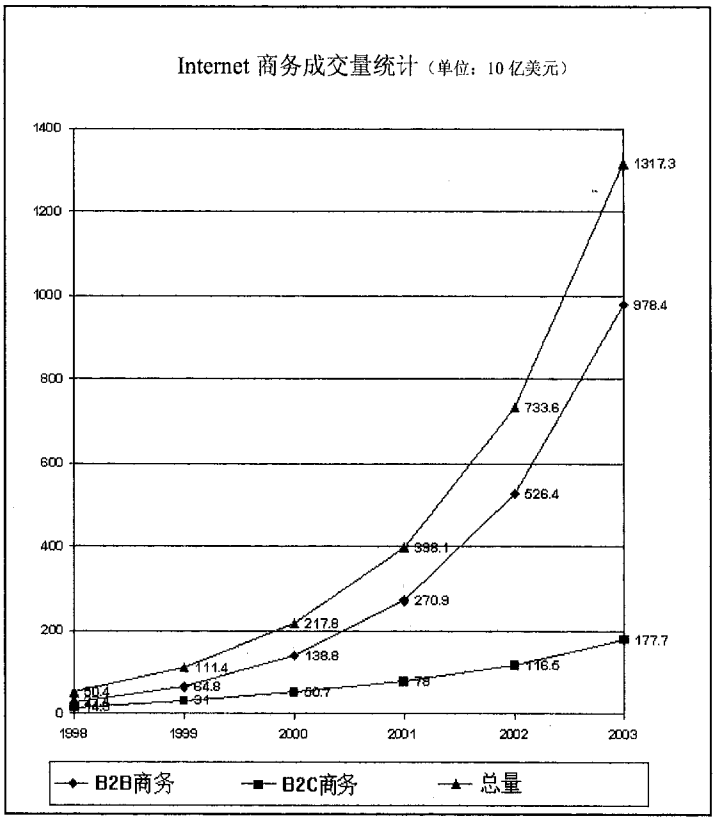

图 2 Internet 电子商务统计和预测数据

由上述数据可以看出，基于 Internet 上的传统数据业务几乎逐年成倍增长着，毫无疑问，网络通信的流量也将以同样的速度增长，这就对传统的 IP 网络技术提出了挑战。同时，传统通信业务向 IP 网络的靠拢和融合趋势越来越明显，并已经成为 IP 网络上的新兴业务，对 IP 网络形成了新的业务压力和发展动力。许多基于 IP 的新技术应运而生，在保证原有传统的数据业务的同时，适应新的数据业务，特别是宽带、实时业务的需求，大大提高了网络的传输性能、可靠性以及可扩展性。

第一章　IP 网络的基本概念及发展历程

IP 是英文 Internet Protocol 的缩写，意思是“网间互连协议”，是专门为计算机网络相互连接、进行通信而设计的协议。目前人们广泛使用的 Internet 就是基于 IP 协议构建的。IP 协议确定了计算机在网络上进行通信时应当遵守的规则，从而实现相互间的通信。任何厂家生产的计算机系统，只要遵守 IP 协议就可以与 Internet 互连互通。也正是因为有了 IP 协议，Internet 才得以迅速发展成为世界上最大的、开放的计算机通信网络。

要介绍 IP 网络，就不得不提到 OSI 参考模型和 TCP/IP 参考模型。

第一节　OSI 参考模型

ISO/IEC（国际标准化组织和国际电工委员会）是致力于国际标准的、自愿和非赢利的专门机构，其 OSI 模型（Reference Model of Open Systems Interconnection 开放系统互联参考模型）作为各种网络层上使用的国际标准化协议。凡是按照此模型建立的网络就可以实现互联，这就是“开放”的意思。OSI 模型有 7 层，根据不同层次的抽象功能分层，每层都可以实现一个明确的功能，每层功能的制定都有利于明确网络协议的国际标准，层次明确、避免各层的功能混乱。OSI 模型中的每一层使用下层提供的服务，并向其上一层提供服务。

OSI 参考模型的 7 层由低到高分别是物理层、数据链路层、网络层、传输层、会话层、表示层和应用层，如图 1-1 所示。

第 7 层	应用层（Application Layer）
第 6 层	表示层（Presentation Layer）
第 5 层	会话层（Session Layer）
第 4 层	传输层（Transport Layer）
第 3 层	网络层（Network Layer）
第 2 层	数据链路层（Data Link Layer）
第 1 层	物理层（Physical Layer）

图 1-1　OSI 参考模型的分层

• 物理层

物理层用以实现网络中两个实体之间的物理连接，涉及到的是在信道上传输的原始比特流。物理层的设计主要是处理机械、电气和过程的接口，以及物理层下的物理传输介质等，比如接插件的物理结构，针脚所代表的信号意义、传送方向、电压特性等。设计上必须保证一方发出“1”时，另一方接收到的同样是“1”而不是“0”。

• 数据链路层

数据链路层的主要任务是利用物理层传输原始比特流的功能，使之对网络层显示为一条无错（error free）的链路。由于物理层仅仅收发、传送比特流，确保物理比特的正确性，而并不关心比特流的意义和结构，所以只能依赖数据链路层将发送数据封装到帧（frame）里，按顺序发送各帧，处理接收方回送的确认（acknowledgement）。并且由此，数据链路层还涉及产生和识别帧边界的功能。数据链路层处理帧的破坏、丢失和重复等问题，以及流量调节控制、线路使用权竞争等。本层将不可靠的物理层处理为对上层而言的可靠信道。

数据链路层上的帧传输控制通常分为媒质访问控制（MAC）子层和逻辑链路控制（LLC）子层。MAC 主要处理网络中用户对信道资源的竞争，比如共享媒质以太网中的冲突检测/避让，以及如 APON 中上行信道对各个用户的分配问题。LLC 主要实现数据成帧、差错控制、流量控制和链路控制等功能。

• 网络层

网络层关系到网络的运行和控制，尤其是确定分组从源端到目的端的路由。路由选择可以通过网络中预先由网管配置的静态路由表，也可以在每次建立会话时确定，还可以根据网络当前的负载状况，高度灵活地为每一个数据包进行路由选择。

• 传输层

传输层的基本功能是从会话层接收数据，并且在必要的时候将它分成较小的单元，传输给网络层，并确保到达对方的各段信息正确无误。传输层向上层提供一个标准的、通用的界面或接口，使面向应用的上层细节和下层通信网络的细节相互隔离。传输层是真正的从源到目标“端到端”层，提供端到端的差错控制、顺序控制和流控制。也就是说，源端机上的程序，利用报文头和控制报文与目标机上的类似程序进行对话。

• 会话层

会话层允许不同计算机上的用户建立会话关系。会话层允许进行类似传输层的普通数据的传输，并提供了对某些应用有用的增强服务会话，也可以被用于远程登录到分时系统或在两台机器间传递文件。

• 表示层

表示层完成某些特定的功能：数据的表示（格式和代码）、转换，消除网内各个实体间的语义差异，比如用一种大家一致同意的标准方法对数据编码，数据的加密/解密、压缩/解压缩，终端的字符集、显示特性等格式转换。总之，表示层提供应用程序和它所需要的服务之间的标准接口。因为在异构网络中的数据格式必须要在各个平台之间相互转换，所以表示层在异构网络互联方面尤为关键。

• 应用层

应用层包含大量人们普遍需要的协议、应用，为用户提供各种基于OSI结构网络的应用，如电子邮件、远程登录、文件传输等。为了使所有的应用程序都可以无缝地与网络结合，方法之一是定义一个抽象的网络虚拟终端（network virtual terminal），编辑程序和其他所有的程序都面向该虚拟终端。而对每一种终端类型都写一软件把网络虚拟终端映射到实际终端，所有虚拟终端软件都位于应用层。

OSI七层网络模型是一种理想的、抽象的网络分层概念，目前并没有一个完全的实现。实际应用中各层的功能往往相互交织，并且很难清晰地分开。

第二节 TCP/IP参考模型

TCP/IP参考模型是ARPANET及其“后继——Internet——使用”的参考模型。ARPANET是由美国国防部赞助的研究网络。ARPA计划于1969年正式联网实验，1970年的ARPANET已初具雏形，并且开始向非军用部门开放。逐渐地它通过租用的电话线连结了数百所大学和政府部门。到1976年，ARPANET已经发展到有60多个结点，连接了100多台主机，跨越整个美国大陆，并通过卫星连至夏威夷，触角伸至欧洲，形成了覆盖世界范围的通信网络。

新的传输手段和网络技术出现后，比如无线网络和卫星，当时的协议在和它们相连的时候出现了问题，所以需要一种新的参考体系结构。这个体系结构在它的两个主要协议出现以后，被称为TCP/IP参考模型。1973年9月，“国际网络工作小组”（INWG即现在IFIPWG6）特别会议上，Vinton Cerf和Bob Kabn提交了第一份协议草稿，提出Internet最初设想，TCP/IP的开发正式跨出了第一步。1974年，Vinton Cerf发表题为《分组网络互连的一个协议》的论文，提出TCP协议。后来分成TCP和IP两个协议，合称TCP/IP。直到1983年，陈旧的“网络控制协议”（NCP）停止使用，完全被TCP/IP所替代。1986年，NSFnet的建立是Internet历史上的一个里程碑。它

标志着从当时起，美国国家科学基金（National Science Foundation，NSF）成为促进 Internet 发展的主要角色，作为军事用途的 ARPAnet 则开始逐渐退出舞台。

TCP/IP 参考模型共有五层：应用层、传输层、网络层、网络接口层和物理层。与 OSI 参考模型相比，TCP/IP 参考模型没有表示层和会话层。网络层相当于 OSI 模型的网络层，网络接口层相当于 OSI 模型中的数据链路层。其结构如图 1-2 所示。

第 5 层	应用层（Application Layer）
第 4 层	传输层（Transport Layer）
第 3 层	网络层（Network Layer）
第 2 层	网络接口层（Data Link Layer）
第 1 层	物理层（Physical Layer）

图 1-2 TCP/IP 参考模型

网络层是整个体系结构的关键部分，使主机可以把分组发往任何网络并使分组独立地向目标转发（可能经由不同的路由）。这些分组到达的顺序和发送的顺序可能不同，因此如果需要按顺序发送和接收时，上层必须对分组进行排序。

网络层定义了正式的分组格式和协议，即 IP（Internet Protocol）协议。网络层的功能就是把 IP 分组发送到应该去的地方。分组路由和避免阻塞是这里主要的设计问题。TCP/IP 网络层和 OSI 网络层在功能上非常相似。

传输层的功能是使源端和目标主机上的对等实体可以进行会话。在这一层定义了两个端到端的协议。一个是传输控制协议 TCP（Transmission Control Protocol），它是一个面向连结的协议，允许从一台机器发出的字节流无差错地发往另一台机器。它将输入的字节流分成报文段并传给网络层。TCP 还要处理流量控制，以避免快速发送方向低速接收方发送过多的报文而使接收方无法处理。该层的另一个协议是用户数据报协议 UDP（User Datagram Protocol），它是一个不可靠的、无连结的协议，用于不需要 TCP 排序和流量控制能力而是本身具备这些功能的应用。

应用层包含所有的高层的协议。主要有：Telnet、FTP、SMTP、DNS、NNTP、HTTP 协议等。

主机至网络层，TCP/IP 参考模型没有真正描述这一部分，只是指出主机必须使用某种协议与网络相连。

TCP/IP 参考模型协议分层如图 1-3 所示。

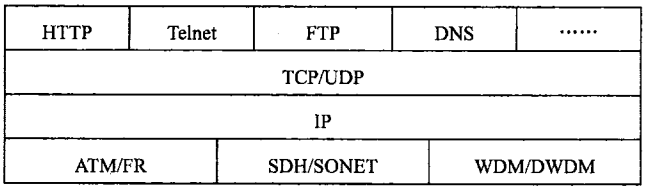

图 1-3　TCP/IP 参考模型协议分层

由于数据传输以及当时对网络军用目的（网络不受子网硬件损失的影响）的要求，基于 TCP/IP 模型的网络应当具有相当的灵活性和生存性，从而形成了基于无连结的分组交换网络。

在基于 IP 协议的网络中，位于 IP 上层的应用将数据递交给 IP 层，由 IP 层对数据进行封装并交由下层传输，从网络层的角度看，网络上统一以 IP 数据包的形式封装用户数据，从而使 IP 网络成为一个具有普适性的平台，各种计算机或其他具有 IP 能力的设备都能在网络上实现互通，具有“开放性”的特点。

IP 网络节点之间的数据通信通过数据包（数据包或分组）的形式实现，每个数据包都由头部和有效荷载（Payload）这两个部分组成，数据包头部中有目的地址等必要内容，需要传送的数据则被分割并封装到数据包的有效荷载中。网络节点检查数据包头部信息，根据目的地址等决定这个数据应该被包发给哪一个下游节点。封装在数据包有效荷载中的数据在最终的目的节点被取出并重新组合，恢复成发送前的形式。即便是拥有同一个目的地址的数据包也可能通过不同的网络路径达到同一个目的节点。在实际传送过程中，数据包还要能根据所经过网络规定的分组大小来改变数据包的长度，IP 数据包的最大长度可达 65 535 个字节。

第三节　IP 网络的应用

从历史上来看，IP 网络主要用于非延时敏感的数据业务。这主要是因为传统 IP 网络的固有特点造成的。

数据包在 IP 网络中的传送通常以逐跳路由方式进行，在这种无连接的数据传输过程中，由于带宽、线路质量等网络资源的使用情况随时发生变化，而各个节点对数据包的路由选择运算受到网络资源分布的影响，因此对同一业务流中不同的数据包所作出的路由转发决定可能不同，各个数据包到达目的地所经过的路径未必相同，因此，它们经历的延时大小也未必相同。

另外，由于传统的IP协议在QoS上没有强有力的支持，对业务类型的支持实际上也只停留在理论上，因而实际应用中对业务流几乎没有什么带宽等质量保证，碰到链路带宽不足而导致的拥塞，通常采用另选路由或者丢弃数据包（然后重发该数据包）的方式，更造成了较大的延时和延时抖动。

因此，在传统的IP网络中常见的应用都是些对网络延时要求不高，没有确定带宽要求的业务。比如E-mail、传真、文件传输、WWW浏览、远程登录等业务。

随着Internet这个世界性的IP网络规模不断扩大，用户数量和业务量也飞速增长，但Internet迅速壮大的初期并没有在业务构成上有太大的变化，仍然以非延时敏感业务为主。主要是WWW浏览的数据流量（HTTP业务和FTP业务）迅速增长。

当Internet显现出无与伦比的发展速度和发展前景时，人们开始尝试将一些传统的通信业务转移到这个基于IP的网络上，试图实现一个综合业务的IP网络。其中最明显的就是话音业务和视频业务。

话音业务和视频业务在传统的电路交换网络上能够很好地传输，是因为电路交换网络中业务传输是面向连接的，电路一旦建立，在业务流存在期间一直存在，而且电路上的资源都归该业务流所拥有，从而很容易就可以满足话音、视频等业务的带宽和延时要求。但是传统的IP网络只是为原先所谓的数据业务（没有延时和带宽质量保证要求）设计的，难以保证话音和视频流的传输质量。就像我们现在通过计算机连接到Internet上传送话音一样，质量远无法和固定电话相比。因此，人们又针对这些IP网络上新出现的具有QoS要求业务类型设计了一些补救措施，比如资源预留协议（RSVP）等，在IP网络中适当地引入面向连接的概念，采用在特定路径上保留足够的网络资源（带宽）给特定的业务流使用，相当于在IP网络中建立一条供某业务流使用的虚电路，通过保证带宽来保证其延时质量要求。

MPLS也是采用了类似的方法，在基于IP的网络中引入面向连接的概念（LSP），提高IP网络对各种业务的承载能力，真正实现综合业务的IP网络。

第二章　IP 网络和协议

第一节　IPv4 协议

经过近 20 年时间的发展，IP 协议已经相当成熟。特别是 Internet 的迅猛发展以及传统通信业务和数据通信业务的融合趋势更加推进了 IP 协议的发展。

目前广泛使用的 IP 协议为版本 4，被称为 IPv4，其数据包结构如图 2-1 所示，由头部和有效荷载两部分组成，头部由一个 20 字节的固定长度部分和一个任意长度的可选部分组成。任何 IP 头部必须是 32 位的倍数，最小 IP 头部长度为 20 字节。

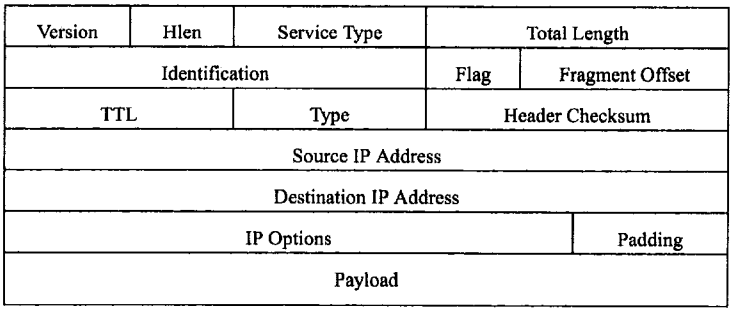

图 2-1　IPv4 包结构

IPv4 包头部结构：

• 版本（Version）域，长度 4 比特，表示 IP 协议的版本，值为 4。

• IP 包头部长度（Header Length）域，长度 4 比特，以 32 比特（4 字节）为单位表示 IP 包头部长度。

• 业务类型（Service Type）域，长度 8 比特，其中 3 比特优先级，4 比特业务类型，1 比特未使用，业务类型的每一位对应于特殊的包传输请求，分别为最小时延（Minimize Delay）、最大吞吐量（Maximize Throughput）、最高可靠性（Maximize Reliability）、最小费用（Minimize Monetary Cost），4 位全 0 表示一般业务。通常的应用中，最高可靠性被指明给网络管理（SNMP）和路由选择协议所使用。用户网络新

闻（Usenet news，NNTP）是惟一要求最小费用的应用。

• IP包总长度（Total Length）域，长度16比特，以8比特为单位表示整个IP包的长度（头部加上数据）。

• IP包标识（Identification）域，长度16比特，每个IP包惟一的标识。

• 标志（Flag）域，长度3比特，用于IP包的分片控制。

• 分片偏移（Fragment Offset）域，长度13比特，以8比特为单位表示本IP包携带的数据作为分片在原数据中的偏移。

• 生存时间（Time to Live）域，长度8比特，表示本IP包在网络中传送时的生存时间（最大跳数），逐跳递减，为0时该IP包被丢弃。

• 协议类型（Protocol Type）域，长度8比特，表示有效荷载中携带数据的类型是TCP还是UDP包。

• IP包头部校验（Header Checksum）域，长16比特，对IP包整个头部的校验和。这种校验和用来检测由路由器中的内存坏字节带来的错误。当数据传送时，该算法将头部所有16位半字数据的补累加起来，写入头部校验和域。值得注意的是，头部校验和在每个节点都要重新计算，因为至少有一个域（生存时间域）总是在变。

• 源IP地址（Source IP Address）域，长度32比特，表示发出本IP包的源地址。

• 目的IP地址（Destination IP Address）域，长度32比特，表示本IP包的目的地址。

• IP可选项（IP Options）域，长度不定，用于携带一些关于源路由、安全和流控制等方面的数据。可选项域的长度以4字节计，现在已经定义了5个可选项：安全性选项、严格源路由选择选项、宽松源路由选择选项、记录路由选项、时间标记选项。但并不是所有的路由器都支持这5个可选项。

安全性选项说明信息的安全程度。事实上，目前几乎所有的路由器都忽略该字段。

源路由选择选项以IP地址序列方式，指定了从源到目的地所经过的节点。严格源路由选择选项指定了数据包途经的完整路径。数据包必须严格地从这条路径传送。当路由选择表崩溃时，系统管理员发送紧急分组，或者作时间测量时，该字段很有用。

宽松源路由选择选项只要求数据包以所指定的次序经过所列出的路由器，而不指定在IP地址序列相邻节点之间经过的路径。

记录路由选项让沿途的路由器都将其IP地址加到该可选字段后，这为系统管理员以后分析这个数据包的来源提供了方便，可用来为路由选择算法查错。

最后，时间标记选项像记录路由选项一样，除了记录32位的IP地址以外，每个路由器还要记录一个32位的时间标记。同样地，这一选项可用来为路由选择算法查错。