

第 1 章 概 述

1.1 引 言

随着信息网络的迅猛发展，对网络技术的掌握和应用越来越受到人们的重视。特别是在本世纪，信息网络无疑将成为信息社会的主要传播媒介和推动社会向前发展的巨大动力，发挥出越来越重要的作用。

目前，IP(Internet Protocol，网际协议)网络作为一种最有发展前景的网络技术，已广泛地应用于整个社会，并起着十分重要的作用。但对于大多数人来说，IP 究竟能做什么？它是如何工作的？它的各部分之间是怎样相互作用的？这些问题并不是每个人都能说清楚、讲明白的。虽然有很多书籍介绍了许多有关 IP 网络的知识，但它们大量讨论的还只是信息网络的使用和应用，以及如何创建时髦的网站等等。

本书为了使读者对 IP 网络有一个基本、全面的了解，在介绍 IP 网络的基本概念、特点、组成、结构及发展趋势等问题的基础上，主要讨论 IP 网络的核心部件路由器技术与原理、路由协议、IP 网络的编址与设计以及 IP 网络的应用等，从而使读者能够较为全面、深入地获得 IP 网络系统的有关知识，并能指导 IP 网络的工程建设、设计与研究工作。

1.2 IP 网络的基本概念

何谓 IP 呢？狭义地讲，IP 是计算机网络中的一种协议标准，对于在网络层采用 IP 协议的网络，我们称之为 IP 网络，其网络层的数据包格式和传输策略等都要符合 IP 协议的规范。那么，什么是 IP 包呢？IP 包实际上就是按照 IP 协议定义的格式所组成的数据包，通常由包头(Header)和随后的一些数据组成，这里的数据可以是文件、终端应用程序或一段 E-Mail 消息等。在 IP 网络上，所有的信息都将被划分为独立的 IP 包，每个 IP 包可以沿自己的路径通过网络传输。这就反映出两个问题：一个是 IP 网络是一种分组交换网络，另一个是 IP 提供面向无连接的数据报服务。下面就简要地介绍分组交换和数据报服务的概念。

分组交换 也称为包交换 是在 20 世纪 60 年代发展起来的，当时主要应用于

军事环境下的通信基础设施。不同于电路交换，在采用分组交换的网络中，用于通信的计算机把它们的数据划分为若干个分组，也就是包，然后将这些分组通过分组交换网传输。分组交换技术有很多，如 X.25、帧中继 FR，Frame Relay 和 IP 等。本书将重点讨论 IP 网络，即 IP 路由器原理与技术，因为 IP 路由器技术作为 Internet 的主要技术得到了广泛的应用，并且成为一种发展趋势。

计算机在通信时为什么要采用分组交换呢？为什么不在计算机之间建立一条临时的“电话呼叫”呢？这种差别主要来源于电话网和计算机网通信特性的不同。在电话网中，用户在通话之前首先要发一个呼叫，一旦呼叫成功，通话双方之间就建立了一条物理通路，这条通路直到通话结束才被释放，在整个通话过程中，它是一直被占用的。由于在电话网中很强调话音的质量，而且一次通话会持续较长的时间，其呼叫建立时间相对于通话的总时间来说是很短的，即呼叫建立的开销在整个通话中所占的比例不是很大，所以在电话网中采用电路交换还是比较合算的。计算机通信就不一样了，大多数计算机通信常常是短暂的，一般在秒的量级（在电话网中，一个呼叫建立的时间大约是 5s），而且经常是“突发性”的。计算机可能在很短的时间之内向多个对等体（层）发送一段较短的数据比特，强行在每秒内建立多个连接，以实现多个对等计算机之间的“会话”。很显然，如果在计算机通信中采用电路交换，那么通信的效率会大大降低。因此要采用分组交换的方式。

一般说来，分组交换有两种基本类型，即“虚电路”和“数据报”。

“虚电路”提供面向连接的服务，在每个虚呼叫的开头需要建立一条虚连接，以后的通信就在这条虚连接上进行。虚电路与传统的电路交换是有很大的区别的，在电路交换中，一个呼叫在其存在期间一直占用一条端到端的物理信道；而虚电路建立在分组交换的基础之上，尽管我们感觉是占用了一条端到端的物理信道，但实际上只是断续地占用一段又一段的链路。在虚电路中，没有必要让每个包都携带目的地址，只要建立连接时的数据包携带目的地址就可以了，其他后续的包使用短的虚电路号。虚电路的优点是可靠性比较高，但效率很低。实际上，网络层采用虚电路服务不是一件很明智的事情。

“数据报”是分组交换的另一种主要形式，它提供面向无连接的服务，IP 和许多其他网络层协议都采用这种服务。在数据报的模式中，数据发送之前并不建立一条可靠的连接，而是让每个包都携带完整的目的地址，即只管将包发送到网络中，至于如何将包转发到它所指明的目的地，完全依赖于中间的每个分组交换设备。数据报服务不提供可靠性，它只是尽最大能力交付，这样可以达到最大的效率，至于可靠性则由上层协议来负责。从目前互联网的发展趋势来看，在网络层提供数据报服务是非常成功的，IP 得到广泛的应用就是一个很好的例证。

在分组交换网中，分组交换设备是一个很关键的部件，它负责数据包的转发工作。IP 网也不例外。在 IP 网络的早期，分组交换设备通常称为“网关”，这可能是

因为它们常常作为本地校园网和广域网之间“网络关口”的缘故，现在，这些设备在更多的情况下被称为“路由器”如图 1.2.1 所示。路由器可以在其任何接口上接收数据包，并通过不同的接口把数据包转发出去，将其发送到目的地。图中的双向箭头用以指示数据包既可以从该接口进入路由器，也可以从该接口离开路由器，路由器是 IP 网络的基本和核心部件。



图 1.2.1 路由器的基本功能

在 IP 网络上流动的元素就是本节第一段提到的 IP 包。就 IP 包所携带的数据而言，可以是传统的计算机数据，也可以是在 IP 上使用的数字化音频和视频流。处理之后的语音和视频流其实也是数据，但是，与传统的数据传输（例如文件传输）不同，它们在传输上具有特殊的需求，因为语音和视频流对时间十分敏感，并且对延迟的容忍也十分有限。有趣的是，语音和视频流可以丢失一些数据而不会产生听觉或者视觉上的失真，但数据业务却必须保证传输的正确性，也就是说需要花费时间来重传丢失或者损坏的数据包以保证整个传输完好无损。

1.3 IP 网络的组成与结构

计算机网络中一个很基本的概念就是协议的分层。一个完整的 IP 网络可以分为 4 层 图 1.3.1 中给出了分层 IP 协议栈和 7 层 OSI (Open System Interconnect, 开放系统互联) 参考模型的结构比较。在计算机网络的分层模型中，4 层的 IP 参考模型和 7 层的 OSI 参考模型是两种最重要的类型。在这两种模型中，每层的功能都由一个以上的实体来执行。一个层中的实体与紧接着它的下一层中的实体直接相互作用，并为上一层提供服务。

在图 1.3.1 所示的 IP 协议栈分层结构中，子网层一般又称网络接口层，它大致相当于 OSI 参考模型中的物理层和数据链路层，负责从网络层接收 IP 报文并向物理网络发送，或从网络上接收物理帧，取出 IP 数据报并提交给网络层。子网层包含的协议主要有 ARP (Address Resolution Protocol, 地址解析协议) 和 RARP (Reverse Address Resolution Protocol, 地址反解析协议)。网络层负责处理分组在网络中的活动，例如分组的选路等。在 IP 网络中，一般常把网络层叫做 IP 层，该层的协议主要包括 IP 协议、ICMP 协议 (Internet Control Messages Protocol, 因特网控

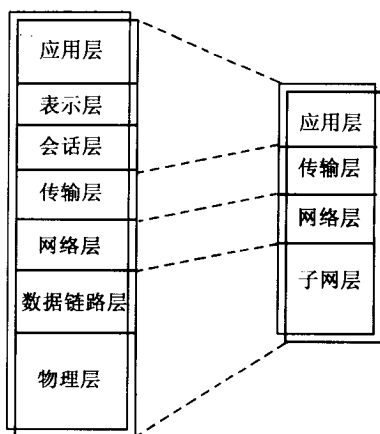


图 1.3.1 OSI和 IP模型比较

制报文协议)以及IGMP协议(Internet Group Management Protocol 因特网组管理协议)等。传输层主要为两台主机上的应用程序提供端到端的通信。在IP网络中,有两个互不相同的传输协议:TCP(Transmission Control Protocol 传输控制协议)和UDP(User Datagram Protocol, 用户数据报协议)。应用层负责处理特定的应用程序细节,几乎各种不同的TCP/IP实现都会提供下面这些通用的应用程序:Telnet(远程登录),FTP(File Transfer Protocol, 文件传输协议),SMTP(Simple Mail Transfer Protocol 简单邮件传送协议)和SNMP(Simple Network Management Protocol 简单网络管理协议)等。各种常见协议和IP网络各层次的对应关系如图1.3.2所示。

在IP协议栈的各个协议中,最重要的有两个:TCP和IP因此人们经常用“TCP/IP”来表示IP网络的体系结构(Internet就使用这个体系结构)或者整个TCP/IP协议族。

在进行IP通信时,通常使用低层“子网”技术来完成相应的数据传输。在这些子网技术中,有些属于LAN(Local Area Network, 局域网)子网,例如,Ethernet(以太网),Token Ring(令牌环),FDDI(Fiber Distributed Data Interface 光纤分部式数字接口)等;还有一些属于WAN(Wide Area Network, 广域网)子网,例如,静态和动态的点到点链路、X.25、帧中继、ATM(Asynchronous Transfer Mode 异步传输模式)、SMDS(Switched Multi-Megabit Data Service 交换型多兆位数据服务)等等。图1.3.3给出了可以运行IP的各种媒体介质使用路由器进行互联的示意图。这些子网中的每个子网都具有各自的内部地址格式和帧格式。有些子网技术既有头字段也有尾字段,而有些只用一个头来封装IP。每种技术都运行在单一的速度或者速度集之上,换句话说,它们之间可以完全不同。

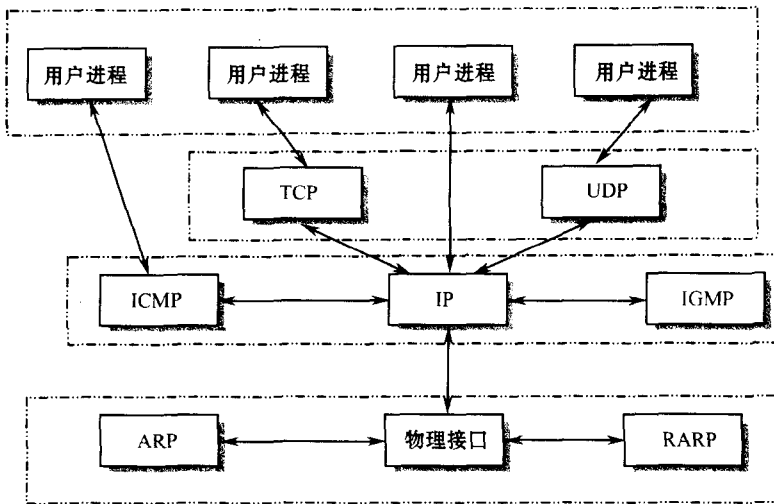


图 1.3.2 IP 网络四个层次及协议

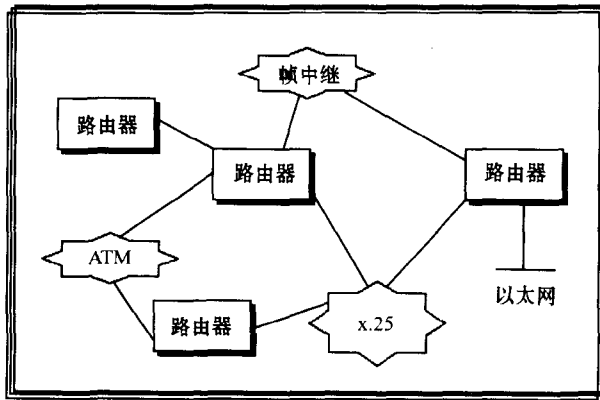


图 1.3.3 IP 运行于其他子网上的示意图

当一个路由器和另一个路由器或者端站进行通信时，需要以某种方式向其相邻节点发送数据包。一个包仅使用其 IP 地址是无法发送给其相邻节点的。IP 地址是高层地址，实际的操作是把 IP 包用特定的子网类型的帧格式封装起来，在该帧中包含有目标节点的地址，通常还包含有路由器的子网源地址。图 1.3.4 给出了 IP 协议栈中各层之间的关系以及封装的概念，这里可以用 TCP 代替 UDP 因为在分层的讨论中它们是相同的。

实际上，各种局域网和广域网都具有自己的编址方法，路由器在相互通信时必须使用相应子网的地址。因此，路由器除了需要知道其相邻节点的 IP 地址外，一

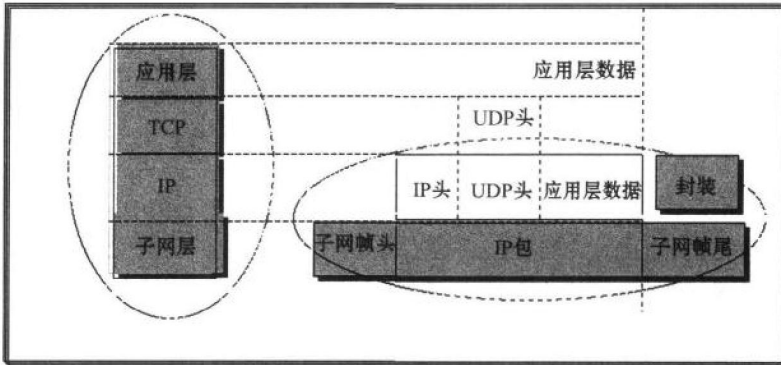


图 1.3.4 分层及封装

般还需要知道其相邻节点的子网层地址。对于每种子网介质，IP 将使用不同的技术来获取其相邻节点的子网地址。在现实生活中，具有多层地址的系统随处可见，如现实生活中的地区、街道和门牌等都是多层地址的表现。

1.4 IP 网络的包格式

1.4.1 IPv4 包格式

IP 包是由 IP 头和高层传输数据协议组成如图 1.4.1 所示。它使 IP 层具有一项很重要的功能，就是“子网无关性”，正是这项功能使得 IP 可以运行在几乎所有类型的子网之上。



图 1.4.1 IP 包格式

IP 向传输层屏蔽它所支持的众多子网和所有不同底层的特征。就像多个网络层协议可以共享同一个子网层一样，也可有多个传输层协议共享 IP 层，如图 1.4.2 所示。注意，图 1.4.2 中的 ICMP 不属于传输层协议，它属于网络层，是 IP 协议的附属协议，但其报文和 UDP、TCP 类似，是封装在 IP 分组中的。

目前使用的 IP 层协议是 IPv4 图 1.4.3 中给出了 IPv4 包头的格式。图中的数字是比特位，包头的宽度为 32 比特。下面分别介绍各个字段的含义。

(1) 版本 表示本 IP 报文的版本号 占 4 比特 对 IPv4 报文来说 该字段的值为 4。

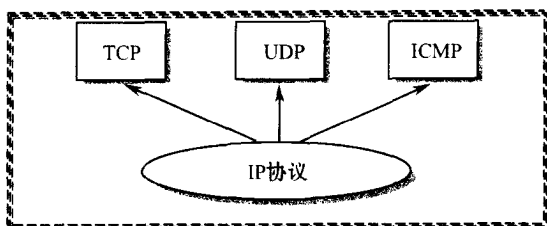


图 1.4.2 多个传输层协议共享 IP 层

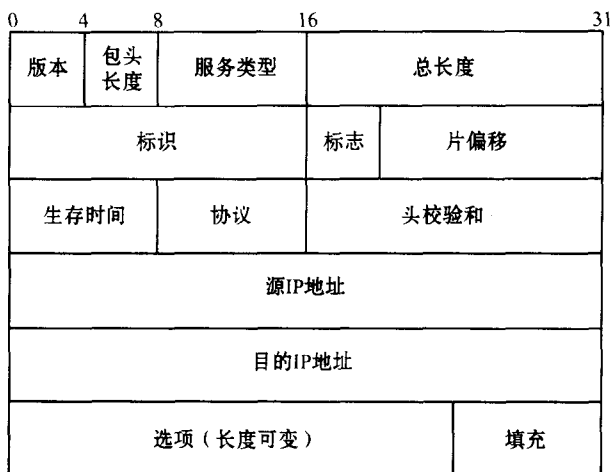


图 1.4.3 IP 包头的结构

(2) 包头长度：表示整个 IP 包头的长度，以 4 字节为 1 个单位。该字段占 4 比特。这样一个 IP 包头最多可以有 60(15×4) 字节。通常 IP 包头的长度是 20 字节，即没有头部选项，这时该字段的值为 5。如果有头部选项，该字段的值就要做相应的改变。

(3) 服务类型 (ToS, Type of Service) 表示本报文所希望得到的服务，占 8 比特。该字段最近在业界引起了较多的关注，因为 Internet 提供商 (ISP, Internet Service Provider) 和顾客强烈要求能够用某种方式在他们的网络内部提供不同的“服务类”级别。当前“区分服务 (DS, Differentiated Service)”的模型定义工作正在进展之中。在该模型中，将 ToS 字段重新定义为 DS 字段。定义区分服务的两个 RFC 文档为 RFC-2474 和 RFC-2475。而 RFC-2430 仅仅是说明如何对区分服务进行部署的一个例子，但是，随着这项新技术实践经验的累积，将有可能出现其他的方案。

(4) 总长度：表示整个 IP 报文的长度，占 16 比特，单位为字节。这样，一份 IP

报文的最大长度是 65535 个字节 ,Internet 协议 RFC - 791 规定所有的 IP 端站必须能够发送和接收长度为 576 个字节的包。

(5) 标识、标志和片偏移 这 3 个字段和 IP 报文的分片有关 分别占 16 比特、3 比特和 13 比特。由于物理链路的限制 , IP 报文在传输过程中有可能被分片 , 分片时就要用到这 3 个字段。

(6) 生存时间 (TTL, Time - To - Live) 表示该报文所能经过的最大步跳数 , 占 8 比特 , 其初始值由源主机指定 , 一般为 32 或 64 最大可以是 255。IP 报文每经过一个路由器 , 该字段的值就要减 1 如果变为 0 报文就会被丢弃。

(7) 协议 表示 IP 包头后面的数据报的协议类型 , 占 8 比特 如 TCP 的协议类型值为 6 , UDP 的协议类型值为 17 , ICMP 的协议类型值为 1。

(8) 头校验和 计算 IP 包头的检验和 , 占 16 比特。该字段可以验证 IP 包头的完整性。

(9) 源 IP 地址和目的 IP 地址 : 这两个字段分别表示源主机的 IP 地址和目的主机的 IP 地址 都是占 32 比特。

(10) 选项和填充 : 选项字段长度可变 , 用于一些控制和测试的用途 , 如记录路径选项。为了保证 IP 头部长度字段的值为整数 , 在选项字段后还会有填充域 , 使 IP 包头长度始终是 32 比特的整数倍。

1.4.2 IPv6 包格式

IPv6 包头格式如图 1.4.4 所示 , 仔细查看 IPv6 的包头字符 , 就可以更好地理解 IPv6。

版本	业务流类别	流标签	
有效数据长度		下一个包头	路程段限制
源地址			
目的地址			

图 1.4.4 IPv6 包头格式

(1) 版本：这个字段的大小和 IPv4 中版本字段的大小是相同的，4 个比特位的版本字段放置数字 6。但是应该注意，使用这个字段时，IPv4 和 IPv6 数据包不是通过版本字段中的版本值来区分的，而是通过 2 层封装如 Ethernet 或者 PPP (Point to Point Protocol，点对点协议) 中的协议类型来区分的。

(2) 业务流类别：占 8 个比特，用来指明业务流的通信要求，用于区分不同的服务。

(3) 流标签：该字段占 20 比特，发送方可以用它来标识一系列属于同一个流的信息包。一个流可以唯一地标识为发送方地址和非零流标签的组合。多点活动流可能存在于发送方和目的地址之前，这时可以是相同的发送地址，但是可取不同的非零流标签。

由此可见，流是以某种方式相互关联的（如发送方、目的地址、QoS(Quality of Service 服务质量)记账、授权、身份验证以及安全这些参数)信息包的序列。流可以是单点传送，即从一个节点到另一个节点；也可以是多点传送，即从一个节点到一组节点。属于同一个流的信息包必须由 IPv6 路由器连续处理，处理属于给定流信息包的方法可以由信息包自己提供的信息来指定，也可以由一个控制协议传输过来的信息指定，如 RSVP(Resource Reservation Protocol 资源预留协议)同时，对于不能支持流标签字段功能的节点，在对信息包进行初始化时，应该把这个字段设置为 0；在转发信息包时，应该保持这个字段不被改变；在接收信息包时应该忽略这个字段。

流的流标签通常分配的是一个随机的数值，由发送节点在 $0x1 \sim 0xFFFFF$ 之间随机选择。这个数值必须不同于发送节点正在使用以及最近使用过的值。所有属于同一个流的信息包在发送时必须具有相同的发送方地址、目标地址、优先权和流标签。进一步说，如果有路径段或者路由器扩展包头，那么它们在属于同一个流的信息包里必须是相同的。

在路由器接收到新的流的第一个信息包时，它可以处理 IPv6 包头、路程段或者路由器扩展包头所携带的信息，记住缓存中的结果，然后把结果用于属于同一个流的其他信息包(具有相同的发送方地址和流标签)方法是直接从缓存中读取。

(4) 有效数据的长度：该字段占 16 比特，用以表示有效数据长度，即 IPv6 包头之后的数据字段的长度，单位为字节。因为它是 16 位的，所以 IPv6 信息包的最大有效数据长度为 65535 个字节。如果需要更大的数据字段，可以使用大有效数据扩展包头。如果有效数据长度字段的值为 0 表明使用了大有效数据。

(5) 下一个包头：占 8 比特用来说明 IPv6 包头之后的上层包头类型，上层包头位于 IPv6 信息包数据字段(有效数据)的开始处。最常使用的是 TCP(6) 和 UDP(17) 这两种，当然还有一些其他包头。这个字段采用的格式是 RFC-1700 为 IPv4 所建议的，并插入了一些为 IPv6 所做的集成，此包头总体上说和 IPv4 的协议

字段是相同的。

(6) 路程段限制：信息包每次经过一个路由器时，8 个比特位的路程段包头 (Hop Limit) 字段的数值就会减 1。如果路程段字段减少为 0 信息包就会被丢弃。这个字段的主要功能就是找出并且忽略那些由于错误的路由信息而循环的信息包。很明显，在两个 IPv6 节点之间，不能有多于 255 个路程段，这就意味着最多只有 254 个路由器。实际上该字段相当于 IPv4 包头中的 TTL 字段。

(7) 发送方地址：IPv6 地址的格式由 RFC - 2373 规定。128 位的源地址 (Source Address) 字段包含的是产生信息包的 IPv6 地址。

(8) 目标地址：128 位的目的地地址 (Destination Address) 字段包含的是接收信息包的 IPv6 地址。如果有路由报头，这个地址不是最终的接收地址。

(9) 扩展包头：一个 IPv6 信息包可以没有扩展包头，也可以具有一个或几个扩展包头。如图 1.4.5 所示的链结构 (以上层头是 TCP 为例) 扩展包头是基于这样一个原理：大多数信息包只需简单地处理，因此 IPv6 包头的基础字段就足够了，在网络层需要额外信息的信息包可以把这些信息编码到扩展包头，扩展包头可以位于 IPv6 包头和上层包头之间。包头之间由下一个包头进行连接，这样就组成了一个图 1.4.5 所示的结构。应该注意的是 IPv4 的包头为一些可选字段留出了空间，这些字段用来为信息包的特殊处理提出要求。这些可选字段并不经常使用，由于它们出现时必须对每个信息包进行检查，所以这个字段会严重地降低路由器的性能，而在 IPv6 中可使用扩展包头来代替可选字段，大大提高了路由器的处理效率。

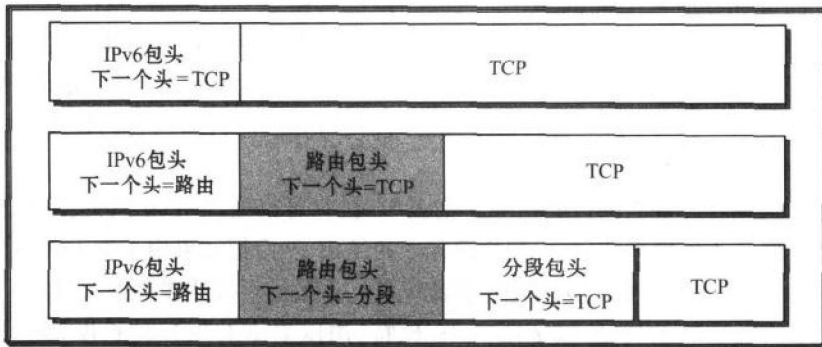


图 1.4.5 扩展包头示意图

1.5 IP 路由器技术发展趋势

随着科学技术的发展，信息已成为推动社会向前发展的巨大动力。在未来 21

世纪中，信息领域的竞争将是世界经济竞争的焦点，信息领域竞争的一个重点，则取决于高性能网络的建设和应用水平。而路由器技术作为该网络领域的主要和关键技术，将发挥出越来越重要的作用。

1.5.1 硬件体系结构的发展过程

作为 IP 网络最核心的设备 路由器在网络中处于至关重要的位置 通常它的主要功能可以概括为：

在网络间接收目的地为远方网段的网络数据报文，并转发出去；

为不同网络之间的用户提供最佳的通信途径；

子网隔离，抑制广播风暴；

维护路由表，并与其他路由器交换路由信息，这是网络层数据报文转发的基础；

实现对数据报的过滤和记账；

利用网际协议，可以为网络管理员提供整个网络的有关信息和工作情况，以便于对网络进行有效管理；

可进行数据包格式的转换，实现不同协议、不同体系结构网络的互联能力。

随着因特网的普及、网络带宽的迅速增加、数据业务的爆炸性增长以及用户对服务质量要求的不断提高，网络系统的规模、速度、种类和应用等都已发生巨大变化，这些网络系统本身的变化导致作为网络核心的路由器体系结构也发生了巨大变化。这种变化集中体现在从基于软件工作的单总线单 CPU 结构路由器发展到现在的基于硬件工作的交换式高性能路由器。这里对路由器硬件体系结构的发展过程给予简单的回顾。

1. 单总线单 CPU 结构路由器

最初的路由器采用了传统计算机体系结构，如图 1.5.1 所示，包括共享中央总线、CPU、内存及挂在共享总线上的多个网络物理接口。如 Cisco2501 路由器就是第 1 代路由器的典型代表，其中 CPU 是 Motorola 的 68302 处理器，具有一个以太网接口和两个广域网接口。

中央 CPU 完成除所有物理接口之外的其他所有功能，数据包从一个物理接口接收进来，经总线送到中央 CPU 中作出转发决定处理，然后又经总线送到另一个物理接口发送出去。这种单总线单 CPU 的主要局限是处理速度慢，一个 CPU 完成所有的任务，从而限制了系统的吞吐量。另外，系统容错性也不好，CPU 若出现故障容易导致系统完全瘫痪。但该结构的优点是系统价格低。目前的接入式路由器基本上都是这种结构。

2. 单总线主从 CPU 结构路由器

采用主从两个 CPU 的结构代替了原来仅一个 CPU 的结构，从而较大幅度地

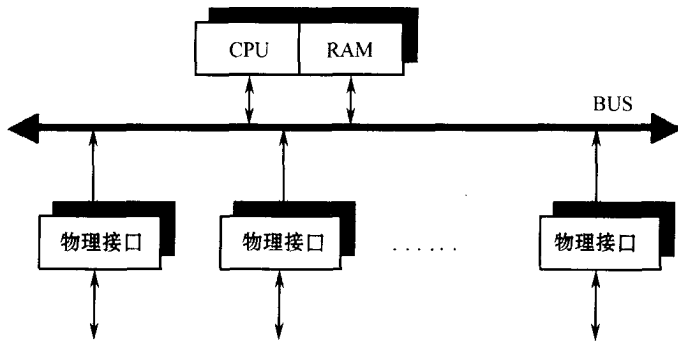


图 1.5.1 第 1 类路由器体系结构

降低了 CPU 的负荷 提高了处理速度 如图 1.5.2 所示。

两个 CPU 为非对称主从式关系结构，其中一个 CPU 负责通信链路层的协议处理 另一个 CPU 则作为主 CPU 负责网络层以上的处理，主要包括转发决定、路由算法和配置控制等计算工作。

第 2 类体系结构实际上是第 1 代体系结构的简单延伸，对系统的容错性能没有多大提高，速度的提高也非常有限。像这种单总线主从 CPU 结构的典型设备有 3Com 公司的 Net Builder2 路由器等。

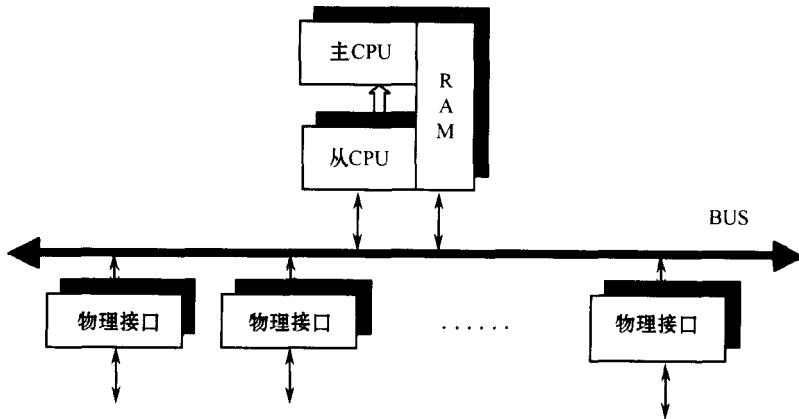


图 1.5.2 第 2 类路由器体系结构

3. 单总线对称式多 CPU 结构路由器

第 3 类路由器可以说克服了第 2 类体系结构中的主要限制，因为它开始采用简单的并行处理技术，即做到在每个接口处都有一个独立 CPU 专门负责接收和发送本接口的数据包，管理接收发送队列、查询路由表做出转发决定等，而主控 CPU 仅完成路由器的配置、控制和管理等非实时功能，如图 1.5.3 所示。

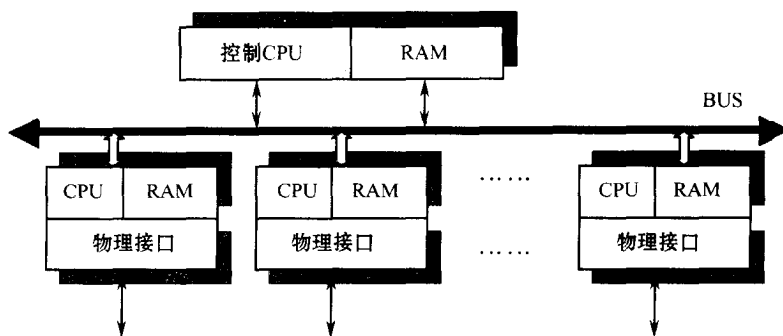


图 1.5.3 第 3 类路由器体系结构

这种体系结构的优点是本地转发 / 过滤数据包的决定由每个接口的专用 CPU 来完成，对数据包的处理被分散到每块接口卡上。第 3 类路由器的主要代表有 Bay BCN 系列，其中大部分接口 CPU 采用的是性能并不算高的 Motorola 60MHz 的 MC68060 或 33MHz 的 MC68040。

4. 多总线多 CPU 结构路由器

第 4 类路由器至少包括 3 类以上总线和 3 类以上的 CPU 如图 1.5.4 所示。显然，这种路由器的结构非常复杂，性能和功能也非常强大。这完全可以从该类路由器的典型之作 Cisco7000 系列中看出。在 Cisco7000 中共有 3 类 CPU 和 3 条总线 分别是 接口 CPU、交换 CPU、路由 CPU、CxBUS、dBUS、SxBUS。

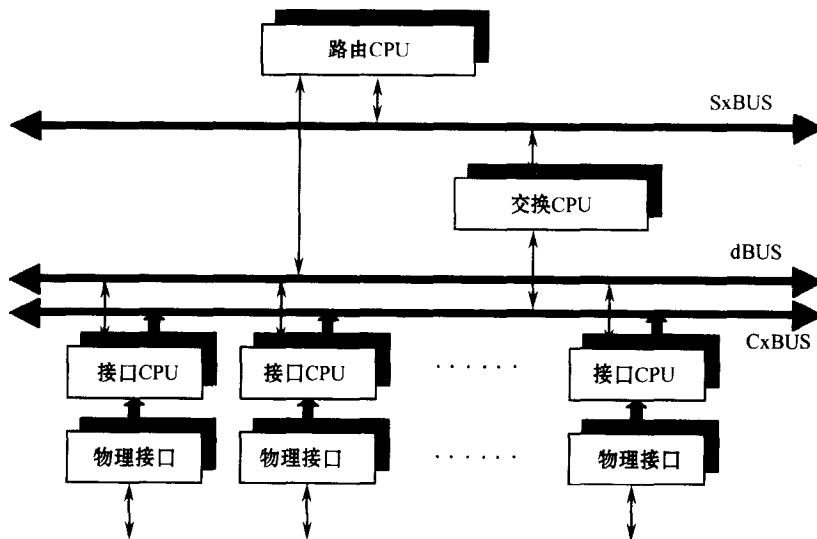


图 1.5.4 第 4 类路由器体系结构

5. 交叉开关 / 交换式体系结构路由器

基于交叉开关设计的路由器有很好的可扩展性能，并且省去了控制大量存储模块的复杂性和高成本，如图 1.5.5 所示。

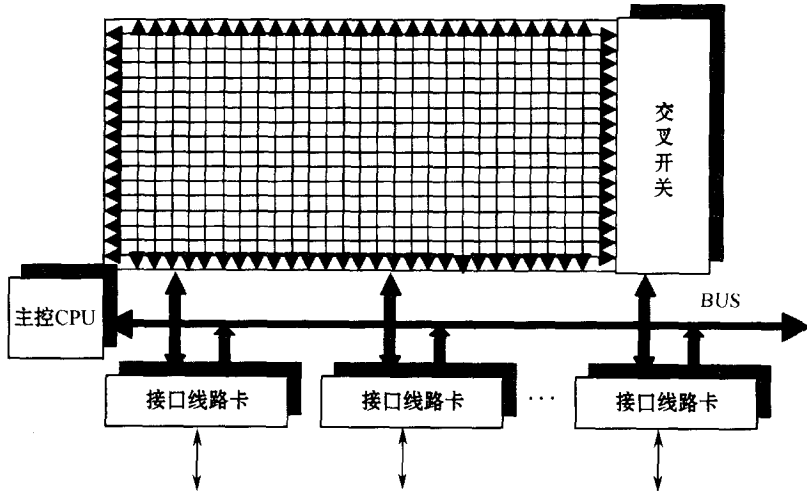


图 1.5.5 第 5 类路由器体系结构

在交叉开关体系结构路由器中，数据直接从输入端经过交叉开关流向输出端。它采用交叉开关结构替代共享总线，这样就允许多个数据包同时通过不同的线路进行传送，从而极大地提高了系统的吞吐量，使得系统性能得到了显著提高。系统的最终交换带宽仅取决于中央交叉阵列和各模块的能力，而不是取决于互连线自身。

就目前来看，这是高速高性能路由器比较好的一个解决方案。实际上，这种路由器多借鉴 ATM 方法，实现各端口之间的线速无阻塞互连。高速交叉开关技术已经十分成熟，在 ATM 交换机和高速并行计算机中广泛应用，市场上可直接买到的高速交叉开关速率就高达 50Gb/s。

目前，市场上已经出现了成熟的交换式路由器产品，其功能也比较完善。例如 Cabletron 的智能交换式路由器产品，它实现的功能超过了传统路由器和第 3 层交换机的功能，能提供在所有端口上以每秒千兆位速率进行第 2、3、4 层交换，而且符合标准的解决方案。高速的专用 ASIC (Application Specific Integrated Circuit 特定用途集成电路) 芯片通过对数据包第 2、3、4 层报头的查找实现数据包的转发。除了能完成第 2、3 层交换功能外，Cabletron 公司的智能交换式路由器可通过在第 4 层交换数据包实现带宽分配、故障诊断和对 TCP/IP 应用程序数据流进行访问控制的功能。

6. 共享并行处理器和全光空分交换结构

尽管第 5 类路由器的结构使得路由器性能大幅度提高，但它的问题是：一方面专用的 ASIC 芯片使系统的成本增加，并且路由器的生产厂家还需要一支开发高性能 ASIC 的队伍；另一方面当今的 Internet 仍处于不断发展和变化的阶段，协议的标准和用户的需求可能随时发生变化，而 ASIC 不能很好而及时跟上并适应这种变化，从而使产品的生命周期减小。

针对这种情况，新一代的半导体生产商提出网络处理器（NP, Network Processor）的概念，采用多微处理器 Multi-Microprocessors）并行工作和专业分工的方法把转发数据包的性能提高几个数量级。这样使网络处理器具有与 ASIC 相当的包转发性能，同时又具有很好的可编程能力，可以适应未来发展的不确定性。另外网络处理器厂家和交换设备生产厂家提出了通用交换接口（CSIX, Common Switch Interface）的接口标准，进一步使网络处理器成为一种标准的部件。

第 6 类路由器就采用了并行处理、光交叉连接和网络处理器等技术，使路由器的包转发速率和吞吐量进一步提高，图 1.5.6 给出了一种示意实现结构。

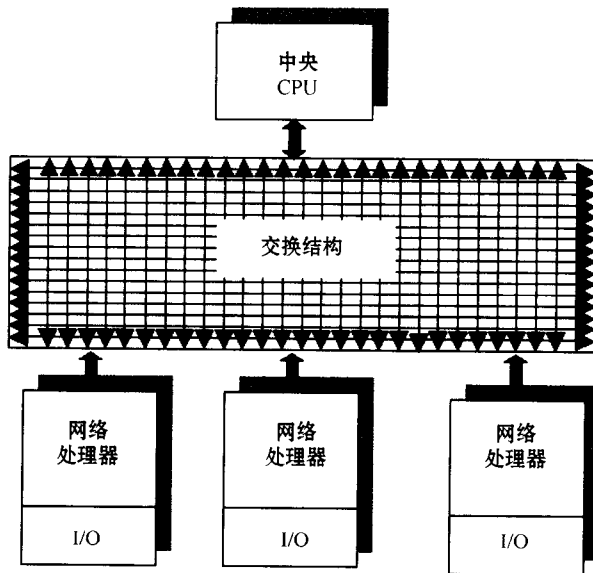


图 1.5.6 第 6 类路由器体系结构

1.5.2 软件体系结构的发展过程

路由器通过路由来决定数据的转发。转发策略可以是人为指定，但人为指定工作量大而且灵活性差，于是动态路由协议应运而生。通过传播、分析、计算、挑选路由，来实现路由选择、路由切换和负载分担等功能。Internet 上现在大量运行的

路由协议有 RIP(Routing Information Protocol 路由信息协议)、OSPF(Open Shortest Path First, 开放最短路径优先)等内部网关协议(IGP, Interior Gateway Protocol)和 BGP(Border Gateway Protocol 边界网关协议 筹外部网关协议 EGP, Exterior Gateway Protocol)。

1. 基于 IPv4 的传统路由协议

目前大量使用的 IP 网络,主要是基于 IPv4 协议栈。经过多年的研究与开发,IPv4 协议栈已基本成熟,基于 IPv4 的路由协议如 RIPv1、RIPv2、OSPFv1、OSPFv2 以及 BGPv4 等也大量应用于全球的路由器中,成为网络沟通的“灵魂”。

RIP 是推出时间最长的路由协议,也是很简单的路由协议。它主要通过传递路由信息来广播路由:每隔 30s 广播一次路由表,维护与相邻路由器的关系,同时根据收到的路由表更新自己的路由表。RIP 运行简单,适用于小型网络,Internet 上还在部分使用着 RIP。

OSPF 协议中的“开放”是针对当时某些厂家的“私有”路由协议而言,而正是因为协议开放性,才造成 OSPF 今天强大的生命力和广泛的用途。它通过传递链路状态(连接信息)来得到网络信息,维护一张网络有向拓扑图,利用最小生成树算法得到路由表。OSPF 是一种相对复杂的路由协议。

BGP 处理各 IGP 之间的路由传递。其特点是有丰富的路由策略,这是 RIP、OSPF 等协议无法做到的。BGP 通过 IGP 边界的路由器加上一定的策略,选择过滤路由把 RIP、OSPF、BGP 等的路由发送到对方。BGP 的出现,引起了 Internet 的重大变革,它把多个 IGP 有机地连接起来,真正成为全球范围内的网络。带来的副作用是 Internet 的路由爆炸,现在 Internet 网的路由由大概是 60000 条,这还是经过‘聚合’后的数字。

2. 基于 IPv6 的新一代路由协议

虽然 IPv4 网络取得了巨大的成功,但它与生俱来的缺陷导致了必然被革新的命运。这些缺陷主要体现在:地址不够使用(地址字段为 32 位)安全性差、移动性差、服务质量不高等。随着科学技术的发展和社会及市场需求的不断提高,IPv4 网络本身所存在的这些瓶颈决定了 IPv6 协议的必然出现。IPv6 协议吸收了 IPv4 的优点而且解决了 IPv4 协议的许多缺陷和不足之处。下面分别介绍 IPv6 在几个方面的优势。

1) IPv6 的地址

IPv6 的主要改变就是地址的长度变为 128 位,这样就可以有 2^{128} 个地址,大约是 10^{38} 个。地球的表面积是 $511\,263\,971\,197\,990\text{m}^2$,那么在每平方米的地球表面上将会有 $655\,570\,793\,348\,866\,943\,898\,599$ 个 IPv6 地址,即使考虑到以后到其他星球的移民,这个数值也足够了。

我们知道 IPv4 的地址长度为 32 位(即 4 个字节)每个字节代表一个无符号

整数,4个字节写成由3个点(·)分开的4个十进制数,例如:224.148.27.24。对于IPv6地址来说,定义相似的语法是必要的,但要考虑到IPv6地址的长度是原来的四倍。由RFC2373规定的标准语法推荐把IPv6地址的128位(即16个字节)写成8个16位无符号整数,每个整数用4位十六进制数来表示,这些数之间用冒号(:)分开,例如:

AEDC:FA20:7484:32B0:AEFC:BC91:2645:3214

从这个例子可以清楚地看到手工管理IPv6地址的难度,同时也说明了动态主机配置协议(DHCP,Dynamic Host Configuration Protocol)和域名服务器(DNS,Domain Name System)的重要性。某些IPv6的设计者看到了用户记忆和书写IPv6地址的困难,越来越多的用户被迫使用名称,然后地址就渐渐地变成了一个网络内部的问题,以及信息包的路由方面的问题。需要说明的是,前面这个例子并不很现实,更为现实的地址的例子是:

ABCD:0000:0000:0000:0008:0800:800C:417C

0000:0000:0000:0000:0000:0000:0B00:00001

显然,这些地址很容易用更为紧凑的格式来表示。一个简化方式就是没有必要书写每一组数值前面的0;例如,可以用0来代替0000,用1来代替0001,用20来代替0020,用300来代替0300。如果使用了这种简化方式,上面的两个地址就会变成下面的形式:

ABCD:0:0:0:8:800:800C:417C

0:0:0:0:0:0:B00:1

可以使用符号(:)做进一步的简化,它代表一系列的0。使用了这种简化之后,上面的两个地址将会变成下面的形式:

ABCD::8:800:800C:417C

::B00:1

注意上述简化对每个地址只能使用一次,我们假设IPv6地址的长度是一定的,因此可以计算出省略了多少个0。这种简化可以用在地址的中间,也可以用在地址的开始或者地址的结尾。如果考虑多点传送、回送或者没有使用的地址,就会认识到这种简化是多么有用,实际上,这些地址的扩展形式如下所示:

2000:0:0:0:0:0:0:43 一个多点传送地址

0:0:0:0:0:0:0:1 回送地址

0:0:0:0:0:0:0:0 没有使用的地址

它们可以使用下面的压缩格式代替:

2000::43 一个多点传送地址

::1 回送地址

: : 没有使用的地址