

第5章

路由器和网络互连

目前多种类型的网络共存,并独立发展,每一种网络已经实现连接在该网络上的两个终端之间的通信过程,如以太网已经实现连接在以太网上的两个终端之间的 MAC 帧传输过程。问题是,任何一种网络都无法独立实现全球任何两个终端之间的通信过程,只有把多种有着不同的适用范围、不同的功能特性的网络互连在一起,才能实现全球任何两个终端之间的通信过程。这就需要一种新的独立于任何类型网络的端到端分组传输协议——网际协议(Internet Protocol,IP),一种新的用于互连不同类型网络的设备——路由器。

5.1 网络互连

异地信件投递过程与连接在不同类型网络上的终端之间的通信过程十分相似,可以通过分析异地信件投递过程得出实现连接在不同类型网络上的终端之间通信过程的思路。

5.1.1 不同类型网络互连需要解决的问题

1. 互连网络结构

互连网络结构如图 5.1 所示,由路由器实现以太网和公共交换电话网(Public Switched Telephone Network,PSTN)这两个不同类型的网络的互连。因此,路由器需要具备连接以太网的端口和连接 PSTN 的端口。以太网用 MAC 地址唯一标识结点,结点之间传输的数据需要封装成 MAC 帧,MAC 帧中需要给出源和目的结点的 MAC 地址,以太网能够实现 MAC 帧源结点至目的结点的传输过程。路由器连接以太网的端口同样具有 MAC 地址,因而能够实现终端 A 与路由器连接以太网的端口之间的 MAC 帧传输过程。

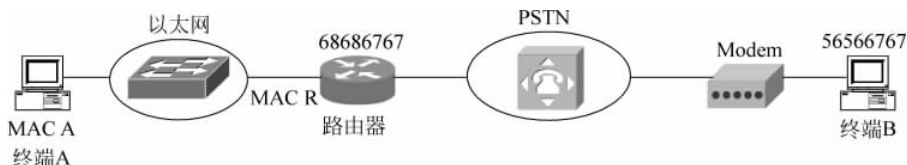


图 5.1 互连网络结构

PSTN 用电话号码唯一标识结点,结点之间开始数据传输过程前,必须通过呼叫连接建立过程建立两个结点之间的点对点语音信道,经过点对点语音信道传输的数据需要封装成

点对点协议(Point to Point Protocol, PPP)帧,由于 PPP 帧是沿着点对点语音信道传输,因此,PPP 帧中无须给出两个结点的地址信息。路由器连接 PSTN 的端口同样具有电话号码,因而终端 B 与路由器连接 PSTN 的端口之间能够建立点对点语音信道,因此,能够实现终端 B 与路由器连接 PSTN 的端口之间的 PPP 帧传输过程。

如图 5.1 所示,终端 A 与终端 B 之间的传输路径由终端 A 和终端 B、两个不同类型的传输网络与互连这两个不同类型传输网络的路由器组成。终端和路由器称为跳,对于终端 A 至终端 B 的传输路径,路由器称为终端 A 的下一跳,终端 B 称为路由器的下一跳。终端 A 只能通过以太网实现与下一跳路由器之间的 MAC 帧传输过程。同样,对于终端 B 至终端 A 的传输路径,路由器称为终端 B 的下一跳,终端 A 称为路由器的下一跳。终端 B 只能通过点对点语音信道实现与下一跳路由器之间的 PPP 帧传输过程。

实现两个不同类型网络互连,需要实现连接在不同类型网络上的两个终端之间的通信过程,如图 5.1 中终端 A 与终端 B 之间的通信过程。实现连接在不同类型网络上的两个终端之间的通信过程需要解决什么问题呢?

2. 需要解决的问题

(1) 地址标识不同。

由于以太网用 MAC 地址唯一标识连接在以太网上的终端,而 PSTN 用电话号码唯一标识连接在 PSTN 上的终端,因此,终端 A 无法识别终端 B 的电话号码,终端 B 无法识别终端 A 的 MAC 地址。双方无法用自己能够识别的标识符标识对方。

(2) 帧格式不同。

终端 A 只能发送、接收 MAC 帧,终端 A 发送的数据或者发送给终端 A 的数据必须封装成 MAC 帧。终端 B 只能通过点对点语音信道接收、发送 PPP 帧,终端 B 发送的数据或者发送给终端 B 的数据必须封装成 PPP 帧。MAC 帧和 PPP 帧是两种完全不同的帧格式。因此,双方无法用对方能够识别的帧格式封装传输给对方的数据。

(3) 无法建立传输路径。

以太网只能建立连接在同一以太网上的两个结点之间的交换路径,PSTN 只能建立连接在 PSTN 上的两个结点之间的点对点语音信道,因此,终端 A 无法建立与终端 B 之间的交换路径,终端 B 也无法建立与终端 A 之间的点对点语音信道。因此,双方无法建立可以直接将数据传输给对方的传输路径。

5.1.2 信件投递过程的启示

将信件从南京投递到长沙的过程如图 5.2 所示。首先,将寄信人用来传递信息的信纸封装为信件,信封上写上寄信人和收信人地址,然后将信件交给南京邮局。南京邮局根据收信人地址——长沙,确定信件的下一站——上海,由于南京至上海采用公路运输系统,因此,信件被封装为适合公路运输系统的形式——信袋,而且信袋上用车次 3536 表明运输该信袋的车辆及信袋的始站与终站。由于上海至长沙采用航空运输系统,因此,上海首先需要从信袋中提取出信件,然后将其封装成适合航空运输系统的形式——信箱,信箱上用航班号 AU765 表明运输该信箱的飞机及信箱的始站与终站。信件经过南京至上海的公路运输系统和上海至长沙的航空运输系统这两阶段运输服务到达目的地长沙。从图 5.2 所示的信件

传输过程可以得出以下启示：

- (1) 不同运输系统有着不同的封装信件的形式和标识始站与终站的方式。
- (2) 信件上收信人和寄信人地址是统一的,和实际提供运输服务的运输系统标识始站与终站的方式无关。
- (3) 信件是一种标准的封装形式,和实际提供运输服务的运输系统封装信件的形式无关。
- (4) 南京根据信件上的收信人地址确定下一站——上海,同样,上海也是根据信件上的收信人地址确定下一站——长沙,信件在南京至长沙的运输过程中是不变的。
- (5) 由实际的运输系统提供当前站至下一站的运输服务。

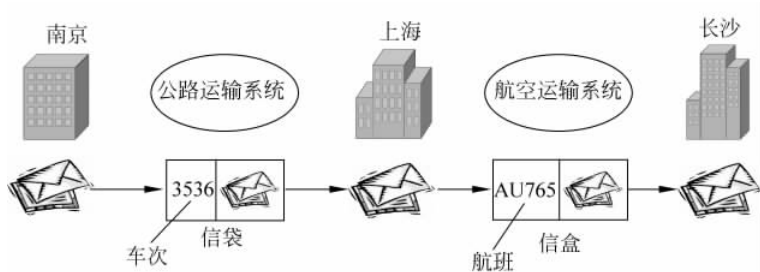


图 5.2 信件运输过程

5.1.3 端到端传输的思路

如果将图 5.1 中的终端 A 对应到图 5.2 中的南京,将以太网对应到公路运输系统,将路由器对应到上海,将航空运输系统对应到 PSTN,将终端 B 对应到长沙,可以通过仿真南京至长沙的信件运输过程,给出实现终端 A 至终端 B 数据传输过程的思路,如图 5.3 所示。

(1) 对应信件的收信人和寄信人地址,定义一种和具体传输网络无关的、统一的终端地址格式——IP 地址。

(2) 对应信件,定义一种和具体传输网络无关的、统一的数据封装格式——IP 分组,和信件相同,IP 分组用 IP 地址标识源和目的终端。

(3) 假定终端 A 的 IP 地址为 IP A,终端 B 的 IP 地址为 IP B,与南京发送给长沙的信件相同,终端 A 发送给终端 B 的数据封装成以 IP A 为源地址、以 IP B 为目的地址的 IP 分组。

(4) 与南京根据信件的收信人地址确定信件的下一站是上海,并将信件封装成适合连接南京和上海的公路运输系统运输的形式——信袋一样,终端 A 能够根据终端 B 的 IP 地址 IP B 确定 IP 分组的下一跳是路由器,并将 IP 分组封装成适合以太网传输的 MAC 帧格式。MAC 帧的源地址是终端 A 的 MAC 地址 MAC A,MAC 帧的目的地址是路由器连接以太网端口的 MAC 地址 MAC R。

(5) 与上海能够从信袋中取出信件,根据信件的收信人地址确定下一站是长沙,并将信件重新封装成适合连接上海与长沙的航空运输系统运输的形式——信箱一样,路由器能够从接收到的 MAC 帧中分离出 IP 分组,根据 IP 分组的目 IP 地址 IP B 确定下一跳是终端 B,路由器通过呼叫连接建立过程建立路由器连接 PSTN 的端口与终端 B 之间的点对点语

音信道,重新将 IP 分组封装成适合点对点语音信道传输的 PPP 帧,将 PPP 帧通过路由器与终端 B 之间的点对点语音信道传输给终端 B。

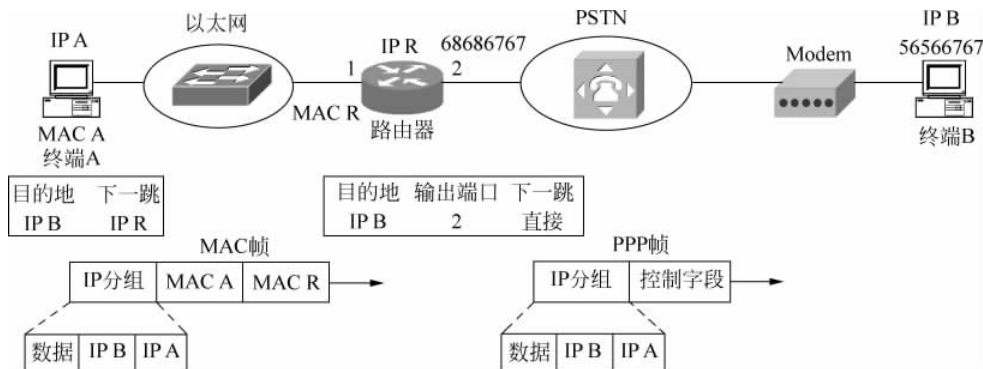


图 5.3 端到端数据传输过程

5.1.4 IP 实现网络互连机制

1. 路由器和 IP 实现端到端传输的过程

从图 5.3 所示的端到端数据传输过程,可以得出以下 IP 实现网络互连的机制。

(1) 规定了统一的且与传输网络地址标识方式无关的 IP 地址格式,所有接入互联网的终端必须分配一个唯一的 IP 地址。同时,由于每一个终端都和实际传输网络相连,具有实际传输网络相关的地址,如以太网的 MAC 地址,为了区分,将 IP 地址称为逻辑地址,将实际传输网络相关的地址称为物理地址。

(2) 规定了统一的且与传输网络数据封装格式无关的 IP 分组格式,端到端传输的数据必须封装成 IP 分组,IP 分组中给出源和目的终端的 IP 地址,每一跳通过 IP 分组携带的目的终端 IP 地址确定下一跳。

(3) 路由器的每一个端口分配一个 IP 地址,同时具有该端口连接的传输网络所对应的物理地址,如以太网对应的 MAC 地址、PSTN 对应的电话号码。路由器实现不同类型网络互连的关键是具有以下功能:一是能够从输入端口连接的传输网络所对应的帧格式中分离出 IP 分组;二是能够根据 IP 分组的目的 IP 地址确定 IP 分组的输出端口,并将 IP 分组重新封装成输出端口连接的传输网络所对应的帧格式。

(4) 对应每一个目的终端,每一跳必须建立用于确定通往该目的终端的传输路径上下一跳的信息,该信息称为路由项,它主要由 3 部分组成:目的终端 IP 地址、输出端口和通往该目的终端的传输路径上下一跳的 IP 地址。对应多个不同目的终端的路由项集合称为路由表。

(5) 必须由单个传输网络连接当前跳和下一跳,能够根据下一跳 IP 地址和输出端口确定连接当前跳和下一跳的传输网络,解析出下一跳与传输网络相关的地址,即物理地址。能够将 IP 分组封装成互连当前跳和下一跳的传输网络要求的帧格式。这个过程称为 IP over X,X 是指互连当前跳和下一跳的传输网络。通过互连当前跳和下一跳的传输网络实现封装 IP 分组的帧当前跳至下一跳的传输过程。对应如图 5.3 所示的端到端数据传输过程,分

别有 IP over 以太网、封装 IP 分组的 MAC 帧终端 A 至路由器的传输过程和 IP over PSTN、封装 IP 分组的 PPP 帧路由器至终端 B 的传输过程。

(6) IP 分组经过逐跳转发,实现源终端至目的终端的传输过程。

2. 路由表和路由项

实现互连网络端到端传输过程的关键有两点,一是源终端能够根据目的终端 IP 地址确定通往目的终端传输路径上的第一跳路由器。源终端至目的终端传输路径经过的其他路由器能够根据目的终端 IP 地址确定通往目的终端传输路径上的下一跳路由器。二是互连当前跳和下一跳的传输网络能够实现 IP 分组当前跳至下一跳的传输过程。如果互连当前跳和下一跳的传输网络是以太网,实现第二个关键点需要完成以下 3 个步骤:①根据下一跳的 IP 地址解析出下一跳连接以太网接口的 MAC 地址;②将 IP 分组封装成以当前跳连接以太网接口的 MAC 地址为源 MAC 地址、下一跳连接以太网接口的 MAC 地址为目的 MAC 地址的 MAC 帧;③经过以太网实现该 MAC 帧当前跳连接以太网接口至下一跳连接以太网接口的传输过程。前面有关以太网的章节已经详细讨论了完成步骤②、③的方法和过程。

实现第一个关键点的核心是路由表,对应每一个目的终端地址,需要路由表给出通往该目的终端传输路径上的下一跳结点的 IP 地址。因此,路由表中的路由项格式是<目的终端 IP 地址,下一跳结点 IP 地址>。如果当前跳和目的终端连接在同一个传输网络,则当前跳至目的终端传输路径上不存在其他路由器,下一跳结点 IP 地址用“直接”表示。如果当前跳通往一组目的终端的传输路径有着相同的下一跳结点,只需为这一组目的终端设置一项路由项<表示这一组目的终端的 IP 地址,下一跳结点 IP 地址>。

IP 分组端到端传输过程中,源终端将 IP 分组传输给第一跳路由器的过程,或者当前路由器根据目的 IP 地址和路由表确定下一跳路由器,并将 IP 分组传输给下一跳路由器的过程称为间接交付。源和目的终端位于同一个传输网络,或者路由器根据目的 IP 地址和路由表确定的下一跳是目的终端本身(目的 IP 地址匹配的路由项中的下一跳为“直接”),源终端或路由器通过直接连接的传输网络将 IP 分组传输给目的终端的过程称为直接交付。

5.1.5 数据报 IP 分组交换网络

1. 传输网络虚化为逻辑链路

为了简化互连网络端到端数据传输过程,在建立互连网络端到端传输路径时,可以将实现终端和路由器之间互连、路由器和路由器之间互连的传输网络虚化成逻辑链路,将如图 5.4(a)所示的互连网络结构简化为如图 5.4(b)所示的由路由器互连多条逻辑链路而成的数据报 IP 分组交换网络。在如图 5.4(b)所示的数据报 IP 分组交换网络中,路由器就是 IP 分组交换机,路由表就是转发表,每一个 IP 分组都是独立的数据报,路由器根据路由表和 IP 分组携带的目的终端 IP 地址实现 IP 分组的转发操作。由于可以将互连网络作为数据报 IP 分组交换网络进行分析、处理,因而常常用数据报 IP 分组交换网络(简称 IP 网络)来称呼互连网络。在 IP 网络中,传输网络的功能等同于逻辑链路,用于实现终端和 IP 分组交换机(路由器)之间及两个相邻 IP 分组交换机之间的 IP 分组传输过程,因而将传输网络

的分组格式称为帧,将传输网络中的分组交换设备,如以太网交换机,称为链路层设备(第二层设备)。

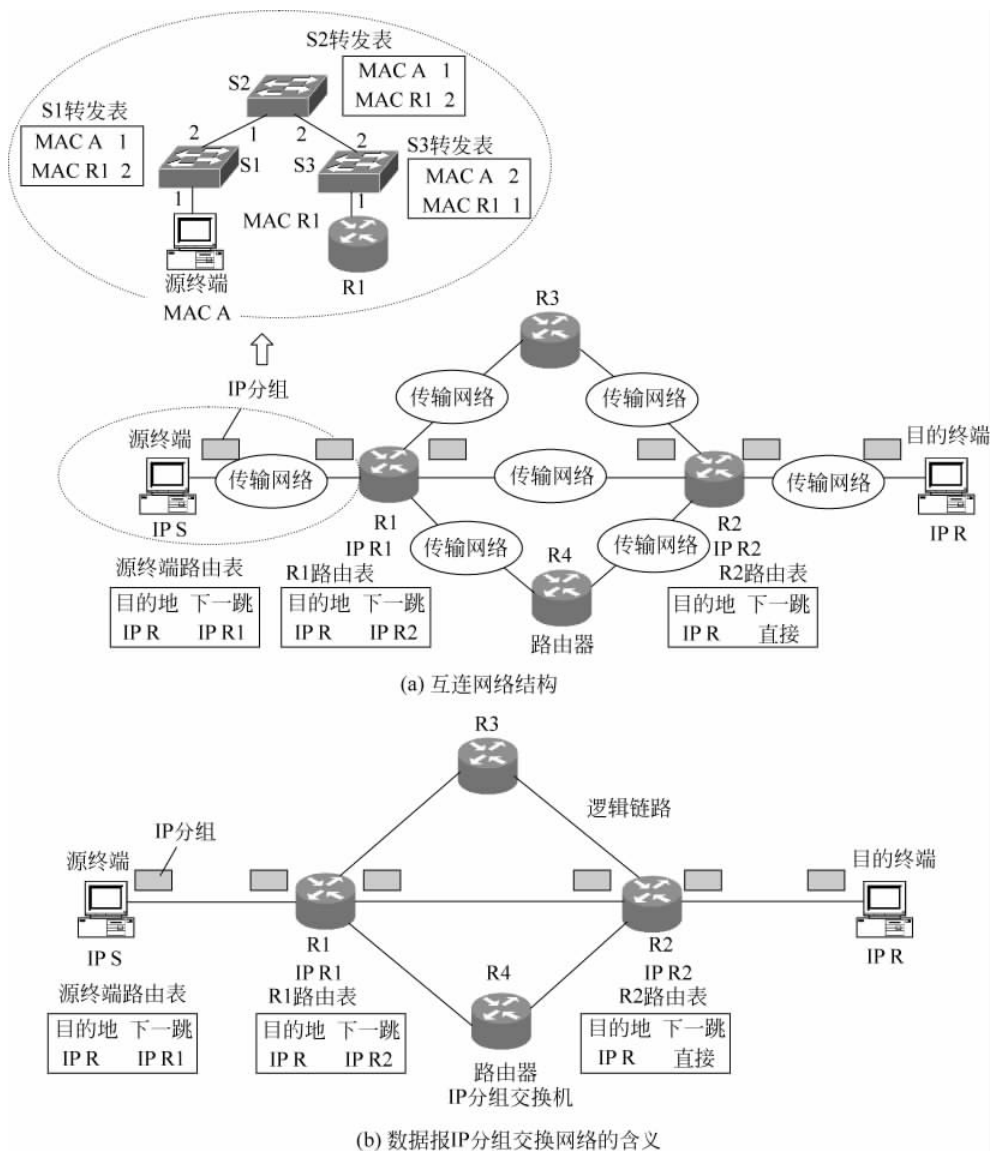


图 5.4 数据报 IP 分组交换网络

2. 两层传输路径

OSI 体系结构中定义的用于互连分组交换机的链路是点对点物理链路或广播物理链路(多点接入物理链路),因此,OSI 体系结构定义的网络应该是由终端、物理链路、分组交换机组成的,和如图 5.4(a)所示的由终端、传输网络和互连传输网络的路由器组成的互连网络是不同的。对于如图 5.4(a)所示的互连网络,端到端传输路径实际上由两层传输路径组成。一是 IP 层传输路径,由源终端、中间路由器和目的终端组成,如图 5.4(a)中源终端至目的终端 IP 层传输路径:源终端→路由器 R1→路由器 R2→目的终端。二是传输网络中当

前跳至下一跳的传输路径,如果连接源终端和路由器 R1 的传输网络是图 5.4(a)中展开的交换式以太网,则源终端至路由器 R1 传输路径就是由源终端、中间以太网交换机和路由器 R1 组成的交换路径:源终端→以太网交换机 S1→以太网交换机 S2→以太网交换机 S3→路由器 R1。因此,和互连网络有关的内容由 3 部分组成,它们分别是:IP、路由协议和 IP over X 技术。IP 详细规定了 IP 地址格式和 IP 分组格式。路由协议通过为每一个路由器建立路由表实现建立 IP 层传输路径的功能。IP over X(X 指特定的传输网络)技术实现解析出下一跳与传输网络相关的地址,即物理地址,并将 IP 分组封装成互连当前跳和下一跳的传输网络要求的帧格式的功能。X 传输网络实现封装 IP 分组的帧当前跳至下一跳的传输过程。

5.1.6 路由器结构

路由器从本质上是 IP 分组转发设备,根据 IP 分组首部携带的地址信息完成从输入端口至输出端口的转发过程。由于 IP 分组是终端之间传输的分组,因此,路由器以数据报交换方式转发 IP 分组。图 5.5 所示的是路由器功能结构,从功能上可以把路由器分成 3 部分,路由模块、线卡和交换模块。

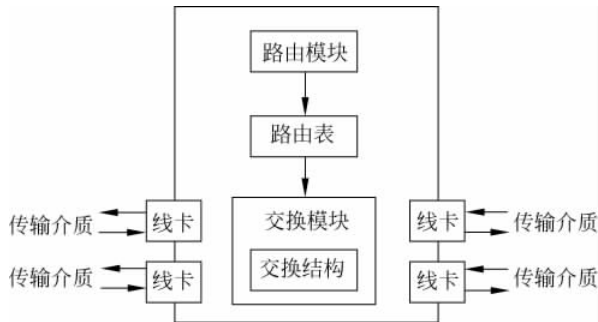


图 5.5 路由器结构

1. 路由模块

路由模块负责运行路由协议,生成路由表,在路由表中给出到达互连网络中任何一个终端的传输路径。当然,由于 IP 分组是逐跳转发,路由器的路由表中只需给出通往互连网络中某个终端的传输路径上下一跳路由器的地址。由于生成路由表的过程比较复杂,因此,路由模块的核心部件通常是 CPU,大部分功能由软件实现。除了生成路由表,路由模块也承担一些其他的管理功能。

2. 线卡

线卡负责连接外部传输介质,并通过传输介质连接传输网络,如连接以太网的线卡通过双绞线缆或光纤连接以太网交换机。线卡通过端口连接传输介质,不同类型的传输介质对应不同类型的端口,如连接 5 类双绞线缆的端口(俗称电端口)和连接光纤的端口(俗称光端口)。线卡除了实现和传输网络的物理连接,还需要按照所连接的传输网络的要求完成 IP 分组的封装和分离操作。封装操作将 IP 分组封装成适合通过传输网络传输的链路层帧格式,如以太网的 MAC 帧。分离操作和封装操作相反,从链路层帧中分离出 IP 分组。线卡进行接收操作时,从所连接的传输介质接收到的物理层信号(如曼彻斯特编码流)中分离出

链路层帧,如 MAC 帧,并从链路层帧中分离出 IP 分组。发送操作时,将 IP 分组封装成链路层帧(如 MAC 帧),将链路层帧通过物理层信号(如曼彻斯特编码流)发送到传输介质。

3. 交换模块

当网卡从某个端口接收到的物理层信号中分离出 IP 分组,就将该 IP 分组发送给交换模块。交换模块用 IP 分组的目的 IP 地址检索路由表,找到输出端口,并把 IP 分组转发给输出端口所在的线卡。随着端口的传输速率越来越高,如 10Gb/s 的以太网端口,端口每秒接收、发送的 IP 分组数量越来越大。对于 10Gb/s 的以太网端口,在极端情况下(假定 IP 分组的长度为 46B,MAC 帧的长度为 64B),端口每秒接收、发送的 IP 分组数量 $= 10 \times 10^9 / (64 \times 8) = 19.53 \text{Mpps}$ (pps 代表 packages per second),当路由器多个端口都线速接收、发送 IP 分组时,交换模块的处理压力将变得很大,因此,通常用称为交换结构的专用硬件来完成 IP 分组从输入端口到输出端口的转发处理。由于存在从多个输入端口输入的 IP 分组需要从同一个输出端口输出的情况,因此,即使交换结构能够支持所有端口线速接收、发送 IP 分组,输出端口也需要设置输出队列,用输出队列来临时存储那些无法及时输出的 IP 分组。

路由器是实现不同类型的传输网络互连的关键设备,它一方面通过路由模块建立到达任何终端的传输路径,另一方面,在确定互连下一跳的传输网络后,将 IP 分组封装成适合互连下一跳的传输网络所对应的链路层帧格式,并通过该传输网络实现 IP 分组当前跳至下一跳的传输过程。

5.2 IP

网际协议(IP)是实现连接在不同类型传输网络上的终端之间通信过程的基础,用于定义独立于传输网络的 IP 地址和 IP 分组格式。

5.2.1 IP 地址分类

1. IP 地址与接口

在深入讨论 IP 地址前,需要说明一下,IP 地址不是终端或路由器的标识符,而是终端或路由器接口的标识符,就像地址不是房子的标识符,而只是门牌号一样。一栋房子如果有多个门,则有多个不同的门牌号,也就有多个不同的地址,但以这些地址为收信人地址的信件都能投递给该房子的主人。同样,终端或路由器允许有多个接口,每一个接口都有独立的标识符——IP 地址,但以这些 IP 地址为目的地址的 IP 分组都到达该终端或路由器。接口是指终端或路由器连接网络的地方,多数情况下,终端或路由器的每一个端口都连接独立的网络,这种情况下,接口就是端口。但存在一个端口可能同时连接多个不同的网络或者多个端口连接同一个网络的情况,这种情况下,一个物理端口可能对应多个不同的接口,多个物理端口可能对应同一个接口。由于每一个 IP 地址指向唯一的终端或路由器,因此,从这一点上讲,IP 地址确实有终端或路由器标识符的作用。

2. IP 地址分类方法

图 5.6 给出了 IP 地址的分类方法。一般情况所指的 IP 地址是指 IPv4 所定义的 IP 地

址,它由 32 位二进制数组成,为了表示方便,采用点分十进制表示法。点分十进制表示 IP 地址的过程如下:将 32 位二进制数分成 4 个 8 位二进制数,每个 8 位二进制数单独用十进制表示(0~255),4 个用十进制表示的 8 位二进制数用点分隔。如 32 位二进制数 01011101 10100101 11011011 11001001 表示的 IP 地址对应的点分十进制表示是 93.165.219.201。

	1	2	3	4	
A	0	网络号	主机号		0.0.0.0~127.255.255.255
B	10	网络号	主机号		128.0.0.0~191.255.255.255
C	110	网络号	主机号		192.0.0.0~223.255.255.255
D	1110	多播地址			224.0.0.0~239.255.255.255
E	11110	保留			240.0.0.0~247.255.255.255

图 5.6 IP 地址分类方法

IP 是用来实现网络互连的协议,因此,用来标识互联网中终端设备的每一个 IP 地址由两部分组成:网络号和主机号。最高位为 0,表示是 A 类地址,用 7 位二进制数标识网络号,24 位二进制数标识主机号,A 类地址中网络号全 0 和网络号全 1 的 IP 地址有特别用途,不能作为普通地址使用。0.0.0.0 表示 IP 地址无法确定。终端没有分配 IP 地址前,可以用 0.0.0.0 作为 IP 分组的源地址。127.×.×.× 是回送测试地址。所有类型的 IP 地址中,主机号全 0 和主机号全 1 的 IP 地址也有特别用途,也不能作为普通地址使用。如网络号为 5 的 A 类 IP 地址的范围为 5.0.0.0~5.255.255.255,但 IP 地址 5.0.0.0 用于表示网络号为 5 的网络地址,而 IP 地址 5.255.255.255 作为在网络号为 5 的网络内广播的广播地址,这种类型的广播地址称为直接广播地址。A 类地址的范围是 0.0.0.0~127.255.255.255,但实际能用的网络号范围是 1~126,每一个网络号下允许使用的主机号= $2^{24}-2$,由此可以看出,A 类地址适用于大型网络。

最高位为 10,表示是 B 类地址,用 14 位二进制数标识网络号,用 16 位二进制数标识主机号,能够标识的网络号个数为 2^{14} ,每一个网络号下允许使用的主机号个数为 $2^{16}-2$ 。B 类地址的范围是 128.0.0.0~191.255.255.255,适用于大、中型网络。

最高位为 110,表示是 C 类地址,用 21 位二进制数表示网络号,8 位二进制数表示主机号,能够标识的网络号个数为 2^{21} ,每一个网络号下能够标识的主机号个数为 2^8-2 。很显然,C 类地址只适用于小型网络。

A、B 和 C 3 类地址中网络号全 0 的 IP 地址称为主机地址,用于标识本网络中的特定终端,如 0.0.0.37 表示本网络中主机号为 37 的终端。

A、B、C 3 类地址都称为单播地址,用于唯一标识 IP 网络中的某个终端,但任何网络都有一个主机号全 1 的地址作为该网络的广播地址,这种广播地址不能用于标识网络内的终端,只能在传输 IP 分组时作为目的地址,表明接收方是该网络内的所有终端。任何网络都有一个主机号全 0 的地址作为该网络的网络地址。根据单播 IP 地址求出对应的网络地址的过程如下:根据该 IP 地址的最高字节值确定该 IP 地址的类型,根据类型确定主机号位

数,将主机号字段清零得到的结果就是该 IP 地址对应的网络地址,如 IP 地址 193.1.2.7 对应的网络地址为 193.1.2.0。

最高位为 1110,表示是多播地址,用 28 位二进制数标识多播组,同一个多播组内的终端可以任意分布在 Internet 中,因此,多播组是不受网络范围影响的。有些多播地址有特殊用途,称为著名(well-known)多播地址,常用的著名多播地址见 7.1.2 节。

最高位为 11110,表示是 E 类地址,目前没有定义。

32 位全 1 的 IP 地址 255.255.255.255 称为受限广播地址,只能在传输 IP 分组时作为目的地址,表明接收方是本网络内的所有终端。

3. 互联网 IP 地址配置原则

互联网配置 IP 地址的原则如下:

(1) 连接在同一传输网络上的终端必须配置具有相同网络号、不同主机号的 IP 地址,如图 5.7 中连接在以太网上的终端 A 和 C。

(2) 每一个传输网络都有一个网络地址,如图 5.7 中以太网配置的网络地址 192.1.1.0 和 PSTN 配置的网络地址 192.1.2.0。

(3) 路由器的每一个接口都需配置 IP 地址,该 IP 地址对应的网络地址必须和分配给该接口连接的传输网络的网络地址相同。如图 5.7 中连接以太网接口配置的 IP 地址是 192.1.1.254,其网络地址为 192.1.1.0,和以太网配置的网络地址相同。

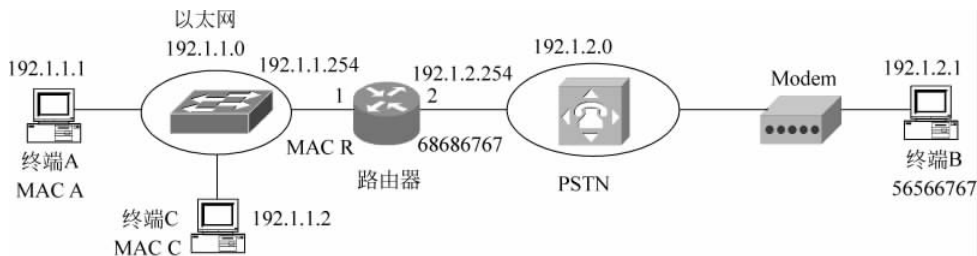


图 5.7 IP 地址配置

如果一个物理以太网被划分为多个 VLAN,则每一个 VLAN 就是一个独立的传输网络,不同 VLAN 须配置不同的网络地址,需要用路由器或其他具有路由功能的设备实现多个 VLAN 的互连。

5.2.2 IP 地址分层分类的原因

IP 地址分层指的是 32 位 IP 地址被分为网络号和主机号两部分,IP 地址分类指的是单播地址被分为 A、B 和 C 3 类。

1. IP 地址分层原因

1) 实现网络互连的需要

由于 IP 用于实现不同类型网络的互连,需要能够根据终端的 IP 地址确定终端连接的网络,并因此能够根据两个终端的 IP 地址区分出这两个终端是连接在同一个网络上的两个

不同的终端还是连接在不同网络上的两个不同的终端。所以，IP 地址需要包含网络标识符（网络号）和主机标识符（主机号）两部分，不同的网络有着不同的网络标识符，同一网络中的不同终端有着相同的网络标识符和不同的主机标识符。因此，如果两个终端的 IP 地址有着相同的网络标识符和不同的主机标识符，表明这两个终端是连接在同一个网络上的两个不同的终端，如果两个终端的 IP 地址有着不同的网络标识符，表明这两个终端是连接在不同网络上的两个不同的终端。

2) 减少路由项

(1) 网络地址与路由项。

网络地址 192.1.1.0 有着两重含义：一是标识网络，表示该网络的网络号是点分十进制数 192.1.1 表示的 24 位二进制数值；二是表示一组有着相同网络号的 IP 地址 192.1.1.0~192.1.1.255。由于该组 IP 地址只能分配给连接在网络地址为 192.1.1.0 的网络上的终端，因此，也可以用网络地址 192.1.1.0 表示连接在网络地址为 192.1.1.0 的网络上的终端集合。

路由器实现不同类型网络互连的一个关键因素是能够根据 IP 分组的目的 IP 地址确定通往该 IP 分组目的终端的传输路径。路由器中用路由项给出通往某个终端的传输路径。如图 5.8 所示，一是网络地址为 192.1.1.0 的网络上连接多个终端，二是对于路由器 R2 而言，通往所有连接在网络地址为 192.1.1.0 的网络上的终端的传输路径是相同的，即通往这些终端的传输路径上的下一跳是相同的，都是路由器 R1。这种情况下，不需要为每一个连接在网络地址为 192.1.1.0 的网络上的终端单独建立路由项，而是只需为连接在网络地址为 192.1.1.0 的网络上的所有终端建立一个路由项，该路由项用网络地址 192.1.1.0 表示 IP 地址为 192.1.1.0~192.1.1.255 的一组终端。IP 地址分层使得可以用网络地址表示一组终端，因而可以用一个路由项给出通往用网络地址表示的一组终端的传输路径。这种情况下，每一个路由项用于指明通往某个网络的传输路径，目的网络字段给出每一个网络的网络地址，下一跳给出当前路由器通往该网络的传输路径上的下一跳的 IP 地址。路由项信息结构如图 5.8 中 R1 和 R2 路由表中的路由项所示。

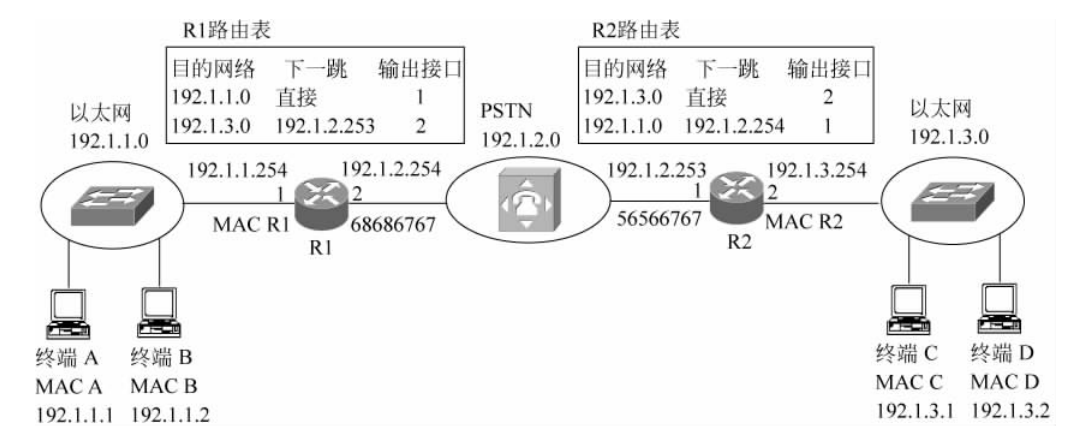


图 5.8 网络地址与路由项一

(2) 路由器确定通往指定终端的传输路径的过程。

路由器根据终端的 IP 地址确定通往该终端的传输路径。由于路由器用路由项给出通往每一个终端的传输路径，因此，路由器确定通往指定终端的传输路径的过程就是用该终端

的 IP 地址在路由表中查找对应的路由项的过程。

下面以路由器 R1 确定通往 IP 地址为 192.1.3.2 的终端的传输路径为例,讨论路由器确定通往指定终端的传输路径的过程。

① 根据终端的 IP 地址计算出该 IP 地址对应的网络地址。首先确定 IP 地址类型,然后根据 IP 地址类型得出 IP 地址中主机号对应的二进制数,将主机号对应的二进制数全部置 0。由于 IP 地址 192.1.3.2 是 C 类地址,得出该 IP 地址用最低 8 位二进制数表示主机号,将这 8 位二进制数全部置 0,得到 IP 地址 192.1.3.2 对应的网络地址是 192.1.3.0。

② 用 IP 地址对应的网络地址检索路由表。用 IP 地址对应的网络地址逐项比较路由项中的目的网络,找到目的网络与 IP 地址对应的网络地址相同的路由项。路由器 R1 用 192.1.3.0 逐项比较路由项中的目的网络,发现路由项 <192.1.3.0,192.1.2.253,2> 中的目的网络等于 192.1.3.0,确定该路由项指明的传输路径就是通往 IP 地址为 192.1.3.2 的终端的传输路径。该路由项表明,路由器 R1 通往 IP 地址为 192.1.3.2 的终端的传输路径上的下一跳是 IP 地址为 192.1.2.253 的路由器,路由器 R1 接口 2 连接的传输网络直接连接 IP 地址为 192.1.2.253 的路由器接口。

2. IP 地址分类原因

不同类型的单播地址的主要区别在于每一个网络地址表示的有效 IP 地址数,由于每一个终端需要分配唯一的 IP 地址,因此,网络地址表示的有效 IP 地址数也决定了连接在该网络地址指定的网络上的终端数。

不同类型的网络适合连接不同数量的终端。同一类型的网络在不同的应用环境下连接的终端数也不同。因此,可以根据网络需要连接的终端数量分配相应类型的 IP 地址。终端数量与 IP 地址类型之间的关系如下。

终端数量 $\leq 2^8 - 2$, 为 C 类地址。

$2^8 - 2 < \text{终端数量} \leq 2^{16} - 2$, 为 B 类地址。

$2^{16} - 2 < \text{终端数量} \leq 2^{24} - 2$, 为 A 类地址。

5.2.3 IP 地址分类的缺陷

IP 地址分类似乎解决了不同规模网络的网络地址分配问题,但事实不是如此,IP 地址分类有着严重的缺陷。

1. IP 地址浪费严重

由于单播 IP 地址分为 A、B 和 C 3 类,每一类 IP 地址有着固定位数的网络号和主机号,因此,会导致 IP 地址浪费。例如一个连接 4000 个终端的以太网,需要分配一个 B 类网络地址,但一个 B 类网络地址包含 $2^{16} - 2$ 个有效 IP 地址。但 4000 个终端只是使用了 $2^{16} - 2$ 个有效 IP 地址中的很小一部分,超过 90% 的有效 IP 地址被浪费了。造成这一问题的原因是固定分类只是将网络规模分为 3 种,而大多数网络规模介于两种网络规模之间,因此,不是网络地址包含的有效 IP 地址数不够,就是只使用了网络地址包含的有效 IP 地址中很小的一部分。如果能够随意确定 32 位 IP 地址中网络号和主机号的位数,适合分配给连接 4000 个终端的以太网的网络地址是网络号为 20 位二进制数,主机号为 12 位二进制数的网

络地址,该网络地址包含 $2^{12}-2=4094$ 个有效 IP 地址。但 A、B 和 C 3 类单播地址中不存在网络号为 20 位二进制数,主机号为 12 位二进制数的网络地址。

2. 不能更有效地减少路由项

网络地址减少路由项的原因是可以用一个网络地址表示一组 IP 地址,因而可以用一个网络地址表示一组终端。对于如图 5.9 所示的情况,网络地址 192.1.0.0 表示一组 IP 地址 192.1.0.0~192.1.0.255,网络地址 192.1.1.0 表示一组 IP 地址 192.1.1.0~192.1.1.255。如果将这两组 IP 地址对应的 32 位二进制数展开,如图 5.10 所示,可以发现,这两组 IP 地址对应的 32 位二进制数中的高 23 位维持不变,低 9 位二进制数从全 0 变化到全 1。如果能够随意确定 32 位 IP 地址中网络号和主机号的位数,可以将网络号的位数定为 23 位,主机号的位数定为 9 位,则 IP 地址集合 192.1.0.0~192.1.1.255 可以用一个网络号位数为 23 位的网络地址 192.1.0.0 表示。同样,网络地址 192.1.4.0 和网络地址 192.1.5.0 表示的两组 IP 地址(192.1.4.0~192.1.4.255 和 192.1.5.0~192.1.5.255)可以用一个网络号位数为 23 位的网络地址 192.1.4.0 表示。由此可以得出,如果 32 位 IP 地址中网络号的位数和主机号的位数可以随意改变,用多个分类的网络地址表示的几组 IP 地址可以用一个网络地址表示。当然,由于一个网络地址表示的一组 IP 地址是几个分类的网络地址表示的几组 IP 地址的组合,该网络地址中的网络号位数小于分类的网络地址中网络号的位数。

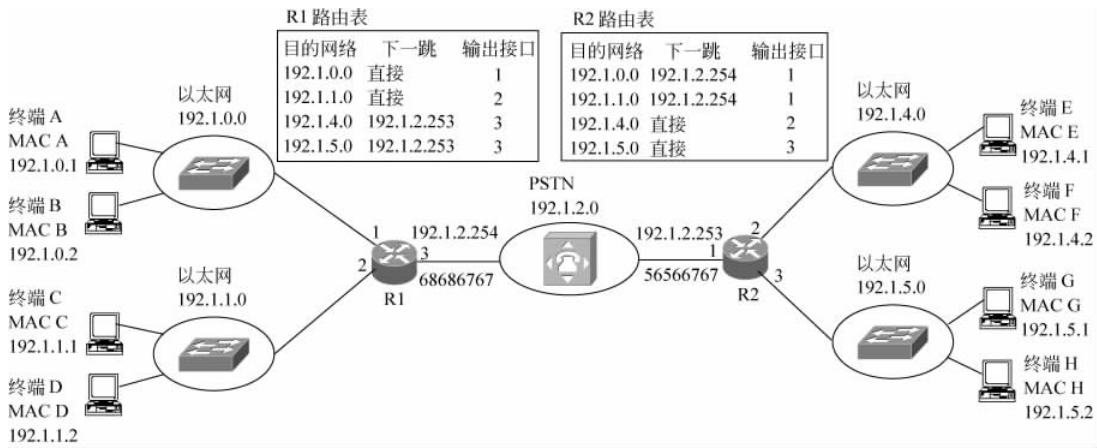


图 5.9 网络地址与路由项二

23位网络号			9位主机号	
11000000	00000001	00000000	0 00000000	192.1.0.0
11000000	00000001	00000000	0 11111111	192.1.0.255
11000000	00000001	00000000	1 00000000	192.1.1.0
11000000	00000001	00000000	1 11111111	192.1.1.255

图 5.10 两组 IP 地址合并成一组 IP 地址

3. C类地址使用率低

由于C类网络地址只能包含 2^8-2 个有效IP地址,因此,分配C类网络地址的网络所连接的终端数不能超过 2^8-2 个。由于大部分网络所连接的终端数大于 2^8-2 个,而且C类IP地址中网络号的位数是24位,使得C类网络地址是数量最多的网络地址,这两点导致C类网络地址成为使用率较低的网络地址,在A类和B类网络地址早已分配殆尽的情况下,仍然存在大量闲置的C类网络地址。实际上,两个高23位网络号相同的C类网络地址可以组合为网络号为23位,包含 2^9-2 个有效IP地址的网络地址,如图5.10所示。4个高22位网络号相同的C类网络地址可以组合为网络号为22位,包含 $2^{10}-2$ 个有效IP地址的网络地址,如4个C类网络地址192.1.0.0、192.1.1.0、192.1.2.0和192.1.3.0可以组合成包含IP地址集合192.1.0.0~192.1.3.255的网络地址192.1.0.0(22位网络号)。依此类推,高 $24-n$ 位网络号相同的 2^n 个C类网络地址,可以组合为网络号为 $24-n$ 位,包含 $2^{8+n}-2$ 个有效IP地址的网络地址。当然,这样做的前提是允许任意改变IP地址中网络号和主机号的位数。

5.2.4 无分类编址

1. 无分类编址机制

单播IP地址分为A、B和C3类,每一类IP地址中网络号和主机号的位数是固定的。固定每一类IP地址中网络号和主机号的位数导致IP地址浪费严重、路由项增多和C类网络地址使用率较低等问题。解决上述问题的关键是允许随意改变IP地址中网络号和主机号的位数。无分类编址就是一种允许随意改变IP地址中网络号和主机号位数的编址方式。

无分类编址方式下,32位IP地址中网络号和主机号的位数是可变的,这样做,消除了IP地址的分类,也解决了因为分类带来的种种问题。但这种编址方式必须提出一种用于指明IP地址中作为网络号的二进制数的方法。

无分类编址通过子网掩码指明IP地址中作为网络号的二进制数。子网掩码也是一个32位的二进制数,和IP地址的表示方法一样,用4个用点分隔的十进制数表示,每个十进制数表示8位二进制数。如255.0.0.0,展开成二进制数表示为11111111 00000000 00000000 00000000。子网掩码中值为1的二进制数对应IP地址中作为网络号的二进制数。5.1.1.2/255.0.0.0表示IP地址是5.1.1.2,对应的子网掩码是255.0.0.0,如果将子网掩码展开成二进制数表示,只有高8位二进制数的值为1,其余为0,这就意味着IP地址的高8位为网络号,低24位为主机号。同样,5.1.1.2/255.255.255.0表示IP地址的高24位为网络号,低8位为主机号。

目前还有一种更直接的表示方式是直接给出IP地址中作为网络号的二进制数位数,如5.0.0.0/8、5.1.0.0/16、192.2.0.0/21等,其中,8、16和21分别是IP地址中网络号的位数。更简单的表示方式是省略IP地址中低位连续的0,如5.0.0.0/8可以表示成5/8,5.1.0.0/16可以表示成5.1/16。

2. CIDR地址块

网络地址是一组有着相同网络号的IP地址,该组IP地址只能分配给连接在同一网络

上的一组终端。许多情况下,只是需要指定有着相同高 N 位的一组 IP 地址,相同的高 N 位称为该组 IP 地址的 N 位网络前缀。该组 IP 地址可以是一组分配给连接在同一网络上的终端的 IP 地址,也可以是合并几组分配给连接在几个不同网络上的终端的 IP 地址后产生的 IP 地址集合,用 CIDR 地址块表示这样的 IP 地址集合。CIDR(Classless Inter-Domain Routing)是无分类域间路由的英文缩写,其意是可以通过为同一区域分配网络前缀相同的 IP 地址集合,有效减少域间路由项。

用 N 位网络前缀表示一组最高 N 位相同的连续的 IP 地址,网络前缀的表示方式和前面表示网络号的方式相同,可以用子网掩码或数字指定 32 位 IP 地址中网络前缀的位数,但网络前缀和网络号的含义不同,它只是用来表示具有相同网络前缀的一组 IP 地址。如 192.1.0.0/21 表示高 21 位相同的 CIDR 地址块,192.1.0.0 是该 CIDR 地址块的起始地址,称为网络前缀地址,该 CIDR 地址块对应的 IP 地址集合是 192.1.0.0~192.1.7.255。即维持高 21 位不变,低 11 位从全 0 到全 1 的 IP 地址范围。计算过程如图 5.11 所示。

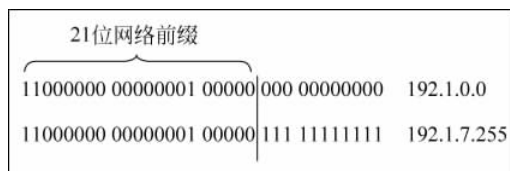


图 5.11 计算 CIDR 地址块表示的 IP 地址集合过程

<网络前缀,主机号>的 IP 地址结构完全取消了原先定义的 A、B、C 3 类 IP 地址的概念。 N 位网络前缀的 CIDR 地址块可以分配给单个网络。这种情况下, N 位网络前缀就是该网络的网络号。也可以分配给多个网络,这种情况下, N 位网络前缀只是用来确定 CIDR 地址块的 IP 地址范围。

值得指出的是,如果某个 CIDR 地址块分配给单个网络,和分类地址一样,该 CIDR 地址块中主机号全 0 的 IP 地址作为该网络的网络地址,主机号全 1 的 IP 地址作为该网络的直接广播地址。对于任何有效 IP 地址,网络前缀全 0 的 IP 地址作为该 IP 地址的主机地址。

3. CIDR 地址块的用途

1) 聚合路由项

(1) 路由项聚合过程。

在如图 5.9 所示的网络结构中,C 类网络地址 192.1.0.0 对应的 CIDR 地址块是 192.1.0.0/24(或者 192.1.0.0/255.255.255.0),对应的 IP 地址集合是 192.1.0.0~192.1.0.255。C 类网络地址 192.1.1.0 对应的 CIDR 地址块是 192.1.1.0/24,对应的 IP 地址集合是 192.1.1.0~192.1.1.255。这两组 IP 地址可以组合成高 23 位相同的 IP 地址集合 192.1.0.0~192.1.1.255,如图 5.10 所示。IP 地址集合 192.1.0.0~192.1.1.255 可以用 CIDR 地址块 192.1.0.0/23 表示。同样,C 类网络地址 192.1.4.0 和 192.1.5.0 对应的 CIDR 地址块分别是 192.1.4.0/24 和 192.1.5.0/24,这两组 IP 地址可以组合成高 23 位相同的 IP 地址集合 192.1.4.0~192.1.5.255,如图 5.12 所示,该 IP 地址集合可以用 CIDR 地址块 192.1.4.0/23 表示。因此,路由器 R1 路由表中的两项路由项<192.1.4.0,

192.1.2.253,3>和<192.1.5.0,192.1.2.253,3>由于下一跳和输出接口相同,可以聚合为一个路由项<192.1.4.0/23,192.1.2.253,3>,其中用 CIDR 地址块 192.1.4.0/23 表示组合 C 类网络地址 192.1.4.0 和 192.1.5.0 表示的两组 IP 地址后生成的 IP 地址集合。同样,路由器 R2 路由表中的两个路由项<192.1.0.0,192.1.2.254,1>和<192.1.1.0,192.1.2.254,1>可以聚合为一个路由项<192.1.0.0/23,192.1.2.254,1>。聚合后的路由器 R1 和 R2 路由表如图 5.13 所示。



图 5.12 两组 IP 地址合并成一组 IP 地址

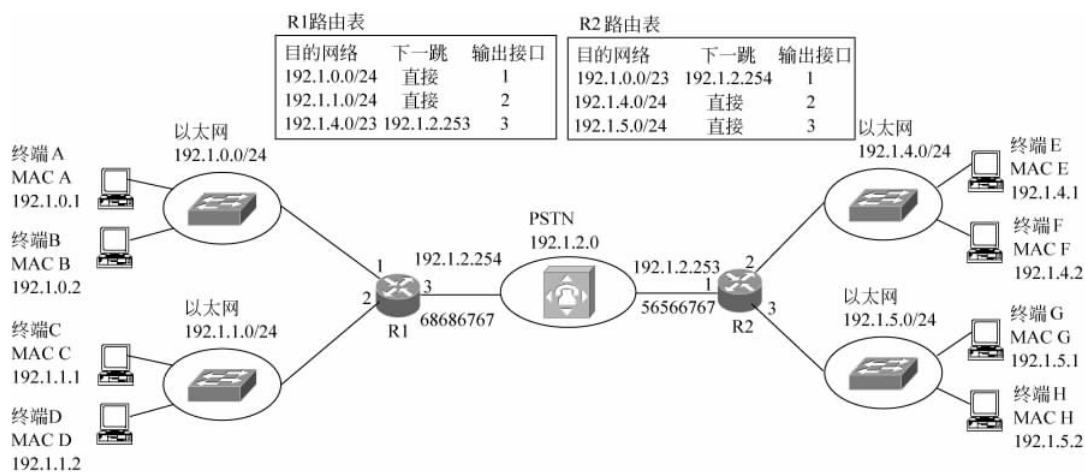


图 5.13 无分类编址与路由项聚合

(2) 路由器确定通往指定终端的传输路径的过程。

路由项中的目的网络地址是网络前缀地址,网络前缀地址和表示网络前缀位数的子网掩码(或数字表示的网络前缀位数)一起确定 CIDR 地址块,因此,如果某个路由项中目的网络地址指定的 CIDR 地址块包含该 IP 地址,表明该 IP 地址与该路由项匹配,该路由项中给出的传输路径就是通往地址为该 IP 地址的终端的传输路径。

下面以路由器 R1 确定通往 IP 地址为 192.1.5.2 的终端的传输路径为例,讨论路由器确定通往指定终端的传输路径的过程。

① 计算网络前缀地址。首先用目的网络地址指定的网络前缀位数求出该 IP 地址对应的网络前缀地址,根据网络前缀位数求出该 IP 地址对应的网络前缀地址的过程如下。32 位 IP 地址中网络前缀位数指定的二进制数维持不变,其他二进制数置 0。该过程也可以通过该 IP 地址和表示网络前缀位数的子网掩码之间的“与”运算完成。由于路由器 R1 路由表中第一个路由项中的目的网络地址指定的网络前缀位数为 24,对应的子网掩码为 255.255.255.0,将 IP 地址 192.1.5.2 与子网掩码 255.255.255.0 进行“与”运算,运算过程

如下:

11000000 00000001 00000101 00000010	192.1.5.2
& 11111111 11111111 11111111 00000000	255.255.255.0
11000000 00000001 00000101 00000000	192.1.5.0

② 逐项比较。用求出的该 IP 地址对应的网络前缀地址与路由项中的目的网络地址比较。如果相同,表明该路由项与该 IP 地址匹配;如果不同,比较下一个路由项。由于路由器 R1 路由表中第一个路由项中目的网络地址指定的网络前缀地址是 192.1.0.0,IP 地址 192.1.5.2 根据该路由项指定的 24 位网络前缀求出的网络前缀地址是 192.1.5.0,显然,第一个路由项和 IP 地址 192.1.5.2 不匹配。同样可以得出,第二个路由项和 IP 地址 192.1.5.2 不匹配。第三个路由项中目的网络地址指定的网络前缀位数是 23,对应的子网掩码为 255.255.254.0,将 IP 地址 192.1.5.2 与子网掩码 255.255.254.0 进行“与”运算,运算过程如下:

11000000 00000001 00000101 00000010	192.1.5.2
& 11111111 11111111 11111110 00000000	255.255.254.0
11000000 00000001 00000100 00000000	192.1.4.0

IP 地址 192.1.5.2 根据该路由项指定的 23 位网络前缀求出的网络前缀地址是 192.1.4.0,与该路由项中目的网络地址指定的网络前缀地址相同,因此,与 IP 地址 192.1.5.2 匹配的路由项是<192.1.4.0/23,192.1.2.253,3>。

2) 将 CIDR 地址块划分为多个网络地址

一个网络前缀位数为 n 的 CIDR 地址块由 2^{32-n} 个 IP 地址组成,该 CIDR 地址块可以划分为两个网络前缀位数为 $n+1$ 的 CIDR 地址块,每一个 CIDR 地址块由 2^{32-n-1} 个 IP 地址组成。如 CIDR 地址块 192.1.4.0/23 可以划分为两个 CIDR 地址块 192.1.4.0/24 和 192.1.5.0/24。CIDR 地址块 192.1.4.0/23 由 2^9 个 IP 地址组成,而 CIDR 地址块 192.1.4.0/24 和 192.1.5.0/24 分别由 2^8 个 IP 地址组成。依此类推,网络前缀位数为 n 的 CIDR 地址块可以划分为 2^i 个网络前缀位数为 $n+i$ 的 CIDR 地址块,每一个 CIDR 地址块由 2^{32-n-i} 个 IP 地址组成。

下面通过一个实例给出将 CIDR 地址块划分为多个网络地址的过程。

假定 CIDR 地址块是 192.1.2.0/24,需要将其分配给 6 个子网,每一个子网连接的终端数如下:子网 1 连接 20 台计算机,子网 2 连接 12 台计算机,子网 3 连接 45 台计算机,子网 4 连接 27 台计算机,子网 5 连接 5 台计算机,子网 6 连接 11 台计算机。CIDR 地址块 192.1.2.0/24 需要划分为 6 个 CIDR 地址块,每一个 CIDR 地址块包含的 IP 地址数量是不同的,与对应子网连接的终端数有关,如分配给子网 1 的 CIDR 地址块包含的 IP 地址数必须 $\geq 20+2$,以此保证至少有 20 个有效 IP 地址。划分 CIDR 地址块的过程是增加网络前缀位数,相应减少主机号位数的过程。CIDR 地址块 192.1.2.0/24 划分为 6 个 CIDR 地址块的过程如下:

00000000~00111111(0~63)分配给子网 3 中的 45 台计算机,网络地址为 192.1.2.0/26。

01000000~01011111(64~95)分配给子网 4 中的 27 台计算机,网络地址为 192.1.2.64/27。

01100000~01111111(96~127)分配给子网 1 中的 20 台计算机,网络地址为 192.1.2.96/27。

10000000~10001111(128~143)分配子网 2 中的 12 台计算机,网络地址为 192.1.2.128/28。

10010000~10011111(144~159)分配给子网 6 中的 11 台计算机,网络地址为 192.1.2.144/28。

10100000~10100111(160~167)分配给子网 5 中的 5 台计算机,网络地址为 192.1.2.160/29。

下面给出实现上述 CIDR 地址块划分过程的思路,终端数最多的子网 3 连接 45 台计算机(图 5.14(a)中用“子网 3(45)”表示),需要 6 位二进制数表示主机号,将 CIDR 地址块 192.1.2.0/24 等分为 4 个网络前缀位数为 26,主机号位数为 6 的 CIDR 地址块,每一个 CIDR 地址块包含 $64(2^6)$ 个 IP 地址。这 4 个 CIDR 地址块中的高 24 位是相同的,是 CIDR 地址块 192.1.2.0/24 中的 24 位网络前缀,3 段十进制数表示是 192.1.2。原来作为主机号的 8 位二进制数中的最高 2 位作为划分后的 4 个 CIDR 地址块的第 25 位和第 26 位网络前缀,4 个 CIDR 地址块的第 25 位和第 26 位网络前缀分别为 00、01、10 和 11,对应的 CIDR 地址块分别是 192.1.2.0/26、192.1.2.64/26、192.1.2.128/26 和 192.1.2.192/26。第 25 位和第 26 位网络前缀为 00 的 CIDR 地址块 192.1.2.0/26 分配给子网 3,如图 5.14(a)所示。

子网 4 和子网 1 连接的计算机数量分别是 27 和 20(图 5.14(a)中分别用“子网 4(27)”和“子网 1(20)”表示),需要 5 位二进制数表示主机号,第 25 位和第 26 位网络前缀为 01 的 CIDR 地址块可以分成两个网络前缀位数为 27,主机号位数为 5 的 CIDR 地址块,这两个 CIDR 地址块的高 26 位相同,第 27 位分别是 0 和 1,对应的 CIDR 地址块分别是 192.1.2.64/27 和 192.1.2.96/27,每一个 CIDR 地址块包含 $32(2^5)$ 个 IP 地址,这两个 CIDR 地址块分别分配给子网 4 和子网 1,如图 5.14(a)所示。

子网 2 和子网 6 连接的计算机数量分别是 12 和 11(图 5.14(a)中分别用“子网 2(12)”和“子网 6(11)”表示),需要 4 位二进制数表示主机号,第 25 位和第 26 位网络前缀为 10 的 CIDR 地址块可以分成 4 个网络前缀位数为 28,主机号位数为 4 的 CIDR 地址块,这 4 个 CIDR 地址块的高 26 位相同,第 27 位和第 28 位分别是 00、01、10 和 11,对应的 CIDR 地址块分别是 192.1.2.128/28、192.1.2.144/28、192.1.2.160/28 和 192.1.2.176/28,每一个 CIDR 地址块包含 $16(2^4)$ 个 IP 地址,前两个 CIDR 地址块分别分配给子网 2 和子网 6,如图 5.14(a)所示。

子网 5 连接的计算机数量是 5(图 5.14(a)中分别用“子网 5(5)”表示),需要 3 位二进制数表示主机号。第 27 位和第 28 位网络前缀为 10 的 CIDR 地址块又可以划分为两个网络前缀位数为 29,主机号位数为 3 的 CIDR 地址块 192.1.2.160/29 和 192.1.2.168/29,将其中一个 CIDR 地址块分配给子网 5,如图 5.14(a)所示。

由于 CIDR 地址块 192.1.2.0/24 涵盖了分配给 6 个子网的 IP 地址集合,因此,图 5.14(b)中的路由器 R1 只需给出一个路由项 $\langle 192.1.2.0/24, 192.1.1.1, 1 \rangle$,表明只要终端 IP 地址的高 24 位等于 192.1.2,通往该终端的传输路径上的下一跳是 IP 地址为 192.1.1.1 的路由器 R2。路由器 R2 对每一个子网均需给出一个路由项,目的网络字段值给出的 CIDR 地址块必须包含分配给该子网的全部 IP 地址。

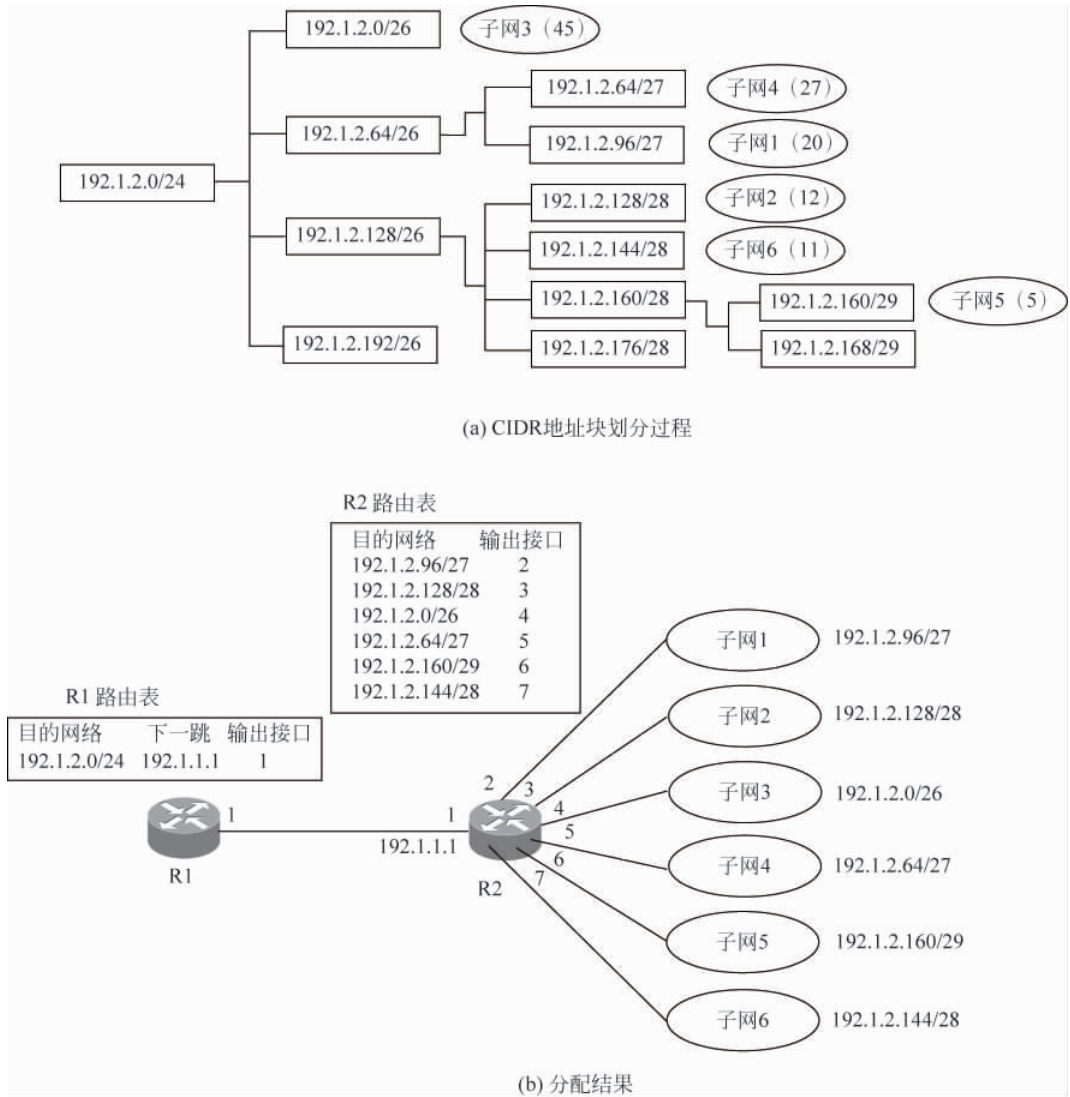


图 5.14 将 CIDR 地址块划分为多个网络地址的过程

4. 最长前缀匹配

由于 IP 地址 192.1.2.150 属于 CIDR 地址块 192.1.2.0/24, 路由器 R1 确定路由项 $\langle 192.1.2.0/24, 192.1.1.1, 1 \rangle$ 与 IP 地址 192.1.2.150 匹配。同样, 由于 IP 地址 192.1.2.150 属于 CIDR 地址块 192.1.2.144/28, 路由器 R2 确定路由项 $\langle 192.1.2.144/28, 7 \rangle$ 与 IP 地址 192.1.2.150 匹配。

如果图 5.14 中的子网 6 既要提高访问外部网络的速度, 但又不想改变自己的配置和访问其他子网的速度, 则采用同时连接路由器 R1 和 R2 的方式, 如图 5.15 所示。这种情况下, 路由器 R1 中的路由项变为两项, 分别指向路由器 R1 和子网 6。当路由器 R1 需要确定通往 IP 地址为 192.1.2.150 的终端的传输路径时, 发现该 IP 地址与目的网络字段值为

192.1.2.0/24 和 192.1.2.144/28 的两项路由项匹配,路由器 R1 如何确定最终匹配的路由项?

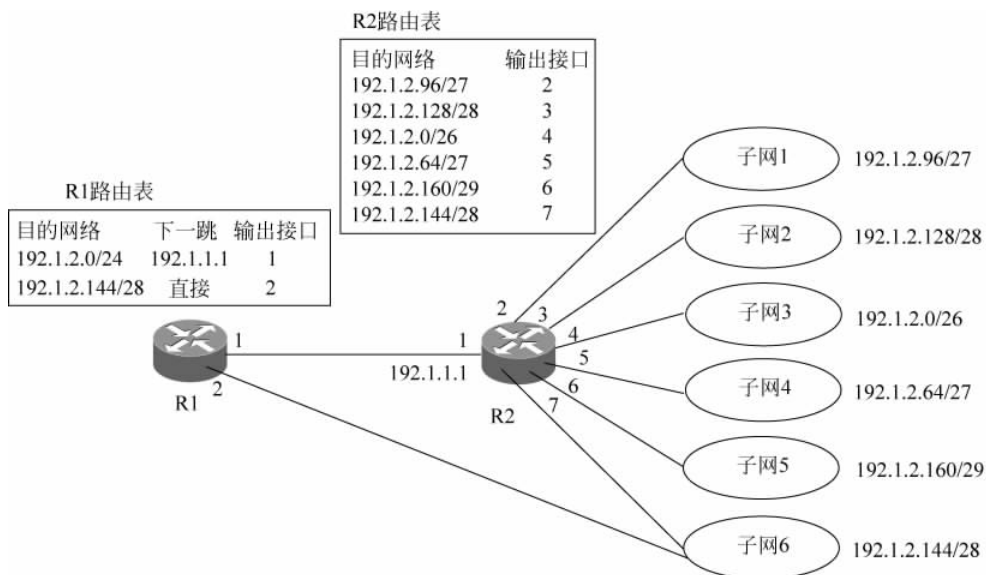


图 5.15 最长前缀匹配过程

显然,路由器 R1 应该选择直接连接子网 6 的传输路径,这也是子网 6 直接连接路由器 R1 的原因。路由器 R1 用最长前缀匹配来确定传输路径的优先级。最长前缀匹配是指,如果有多个路由项与某个 IP 地址匹配,比较这些路由项中目的网络指定的 CIDR 地址块的网络前缀位数,选择网络前缀位数最大的路由项作为最终与该 IP 地址匹配的路由项。路由器 R1 匹配的两个路由项中,一个路由项中的目的网络是 192.1.2.0/24,网络前缀位数是 24,另一个路由项中的目的网络是 192.1.2.144/28,网络前缀位数是 28,选择目的网络是 192.1.2.144/28 的路由项作为最终匹配的路由项。

5. 默认路由项

如果某个 IP 地址和路由表中所有路由项均不匹配,选择默认路由项指定的传输路径。默认路由项的目的网络为 0.0.0.0,对应的子网掩码为 0.0.0.0,由于所有 IP 地址与子网掩码 0.0.0.0“与”运算后的结果都是 0.0.0.0,因此,所有 IP 地址均与该路由项匹配。

当通往多个网络的传输路径具有相同的下一跳时,可用一项默认路由项指明通往这些网络的传输路径。如图 5.16 所示的互连网络中,内部网络通过路由器 R1 连接 Internet,由于 Internet 由无数个网络组成,如果在路由表中详细列出 Internet 中所有网络对应的路由项,路由项数目将十分庞大。根据图 5.16 所示的互连网络结构,路由器 R2 通往 Internet 的传输路径有唯一的下一跳——路由器 R1,因此,除了指明通往内部网络的传输路径的路由项外,可用一项默认路由项指明通往 Internet 的传输路径。如果某个 IP 地址和 3 个内部网络的网络地址都不匹配,意味着该 IP 地址标识的目的终端位于 Internet,路由器 R2 选择通往 Internet 的传输路径作为通往该目的终端的传输路径。

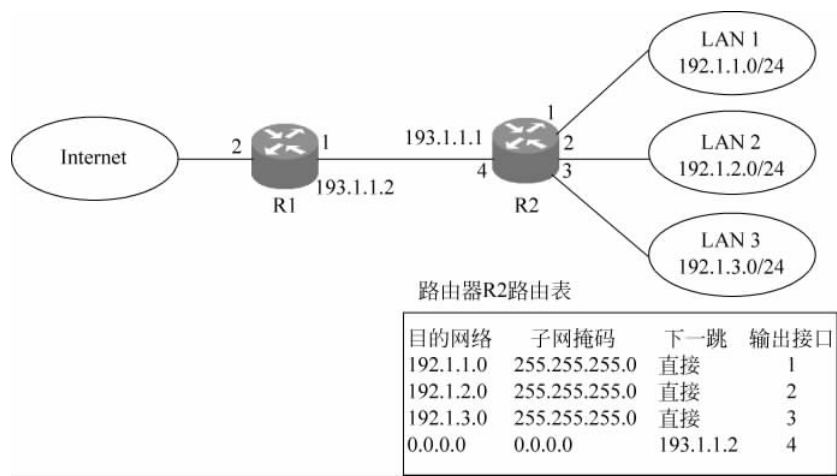


图 5.16 默认路由项功能

6. IP 地址分配机制

采用无分类编址后,可以最大程度地聚合路由项,减少路由表中路由项的数量。但路由表中只有满足以下两个条件的这些路由项才允许聚合:一是这些路由项的目的网络可以用单个 CIDR 地址块涵盖,二是这些路由项有相同的下一跳。因此,为了最大程度地减少路由项,分配 IP 地址时,尽可能使得相同区域的 IP 地址有相同的网络前缀。例如,对于如图 5.17 所示的 Internet 结构,位于欧洲的边界路由器 R1 将所有目的地是亚洲的 IP 分组转发给亚洲的边界路由器 R2。如果分配给亚洲的 IP 地址有相同的网络前缀,路由器 R1 可以用一个路由项指明通往亚洲的传输路径。

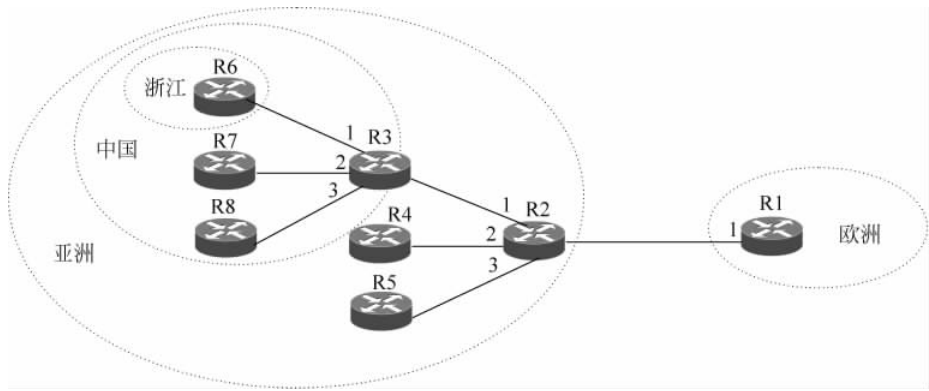


图 5.17 路由项聚合过程

路由器 R2 将所有目的地是中国的 IP 分组转发给中国的边界路由器 R3。如果分配给中国的 IP 地址有相同的网络前缀,路由器 R2 可以用一个路由项指明通往中国的传输路径。值得说明的是,由于中国属于亚洲,因此,标识中国的网络前缀中包含标识亚洲的网络前缀。

路由器 R3 将所有目的地是浙江的 IP 分组转发给浙江的边界路由器 R6。如果分配给浙

江的 IP 地址有相同的网络前缀,路由器 R3 可以用一个路由项指明通往浙江的传输路径。值得说明的是,由于浙江属于中国,因此,标识浙江的网络前缀中包含标识中国的网络前缀。

如表 5.1 所示,所有分配给亚洲的 IP 地址有相同的 4 位网络前缀 0001,因此,32 位 IP 地址中,最高 4 位(0~3 位)网络前缀 0001 成为亚洲标识符,欧洲边界路由器 R1 将所有目的 IP 地址中最高 4 位值等于 0001 的 IP 分组转发给亚洲边界路由器 R2。

表 5.1 各洲 IP 地址分配方案

区 域	网络前缀位数	网络前缀字段值	CIDR 地址块
亚洲	4	0001	16.0.0.0/4
欧洲	4	0010	32.0.0.0/4
美洲	4	0011	48.0.0.0/4
非洲	4	0100	64.0.0.0/4
大洋洲	4	0101	80.0.0.0/4

如表 5.2 所示,分配给亚洲的 IP 地址中,增加 4~7 位作为亚洲国家标识符,例如用 0001 标识中国,因此,可以用最高 8 位(0~7 位)网络前缀 00010001 唯一标识中国。其中最高 4 位(0~3 位)0001 用于唯一标识亚洲,4~7 位 0001 用于在分配给亚洲的 IP 地址中唯一标识中国。亚洲边界路由器 R2 将所有目的 IP 地址中最高 8 位值等于 00010001 的 IP 分组转发给中国边界路由器 R3。

表 5.2 亚洲各国 IP 地址分配方案

国 家	网络前缀位数	网络前缀字段值	CIDR 地址块
中国	8	00010001	17.0.0.0/8
越南	8	00010010	18.0.0.0/8
日本	8	00010011	19.0.0.0/8
⋮	⋮	⋮	⋮

如表 5.3 所示,分配给中国的 IP 地址中,增加 8~14 位作为省标识符,例如用 0000001 标识浙江,因此,可以用最高 15 位(0~14 位)网络前缀 000100010000001 唯一标识浙江。其中最高 8 位(0~7 位)00010001 用于唯一标识中国,8~14 位 0000001 用于在分配给中国的 IP 地址中唯一标识浙江。中国边界路由器 R3 将所有目的 IP 地址中最高 15 位值等于 000100010000001 的 IP 分组转发给浙江边界路由器 R6。

表 5.3 中国各省 IP 地址分配方案

省	网络前缀位数	网络前缀字段值	CIDR 地址块
浙江	15	00010001 0000001	17.2.0.0/15
江苏	15	00010001 0000010	17.4.0.0/15
湖南	15	00010001 0000011	17.6.0.0/15
⋮	⋮	⋮	⋮

因此,欧洲边界路由器可以用一个路由项指明通往亚洲的传输路径,如表 5.4 所示。亚洲边界路由器可以用一个路由项指明通往中国的传输路径,如表 5.5 所示。中国边界路由器可以用一个路由项指明通往浙江的传输路径,如表 5.6 所示。

表 5.4 R1 路由表

目的网络	子网掩码	下一跳	输出接口
16.0.0.0	240.0.0.0	R2	1

表 5.5 R2 路由表

目的网络	子网掩码	下一跳	输出接口
7.0.0.0	255.0.0.0	R3	1
18.0.0.0	255.0.0.0	R4	2
19.0.0.0	255.0.0.0	R5	3

表 5.6 R3 路由表

目的网络	子网掩码	下一跳	输出接口
17.2.0.0	255.254.0.0	R6	1
17.4.0.0	255.254.0.0	R7	2
17.6.0.0	255.254.0.0	R8	3

7. 例题解析

【例 5.1】 CIDR 地址块 59.67.159.0/27、59.67.159.32/27 和 59.67.159.64/26 聚合后可用的有效 IP 地址数为_____。

- A. 126 B. 186 C. 188 D. 254

【解析】 可以将 CIDR 地址块 59.67.159.0/27 和 59.67.159.32/27 聚合为 CIDR 地址块 59.67.159.0/26,聚合过程如下：

59.67.159.0/27

00111011 01000011 10011111 000

00111011 01000011 10011111 000

27位网络前缀

59.67.159.32/27

00111011 01000011 10011111 001

00111011 01000011 10011111 001

27位网络前缀

00000 59.67.159.0

11111 59.67.159.31

59.67.159.0/26

00111011 01000011 10011111 00

00111011 01000011 10011111 00

26位网络前缀

00000 59.67.159.0

11111 59.67.159.63

可以将 CIDR 地址块 59.67.159.0/26 和 59.67.159.64/26 聚合为 CIDR 地址块 59.67.159.0/25,聚合过程如下：

59.67.159.0/26

00111011 01000011 10011111 00

00111011 01000011 10011111 00

26位网络前缀

59.67.159.64/26

00111011 01000011 10011111 01

00111011 01000011 10011111 01

26位网络前缀

000000 59.67.159.0

111111 59.67.159.63

59.67.159.0/25

00111011 01000011 10011111 0

00111011 01000011 10011111 0

25位网络前缀

0000000 59.67.159.0

1111111 59.67.159.127

CIDR 地址块 59.67.159.0/25 的有效 IP 地址数为 $2^7 - 2 = 126$ 。因此,答案为 A。

【例 5.2】 计算并填写表 5.7。

表 5.7 例 5.2 用表

IP 地 址	191.173.21.9
子网掩码	255.240.0.0
网络地址	
直接广播地址	
主机地址	
CIDR 地址块中最后一个可用 IP 地址	

【解析】 将子网掩码 255.240.0.0 展开为 32 位二进制数: 11111111 11110000 00000000 00000000,32 位二进制数由高 12 位连续 1 和低 20 位连续 0 组成,表示 32 位 IP 地址 191.173.21.9 中,高 12 位是网络号,低 20 位是主机号。

将 32 位 IP 地址 191.173.21.9 展开为 32 位二进制数 **10111111 10101101 00010101 00001001**,其中高 12 位为网络号,低 20 位为主机号。

网络地址是将 32 位 IP 地址中的主机号全部置 0 后的结果。因此,网络地址如下: **10111111 10100000 00000000 00000000**,点分十进制表示是 191.160.0.0。

直接广播地址是将 32 位 IP 地址中的主机号全部置 1 后的结果。因此,直接广播地址如下: **10111111 10101111 11111111 11111111**,点分十进制表示是 191.175.255.255。

主机地址是将 32 位 IP 地址中的网络号全部置 0 后的结果。因此,主机地址如下: **00000000 00001101 00010101 00001001**,点分十进制表示是 0.13.21.9。

CIDR 地址块中最后一个可用 IP 地址是直接广播地址减 1 后的结果,为 191.175.255.254。

【例 5.3】 计算并填写表 5.8。

表 5.8 例 5.3 用表

IP 地 址	
子网掩码	
网络地址	
直接广播地址	
主机地址	0.24.37.9
CIDR 地址块中最后一个可用 IP 地址	191.159.255.254

【解析】 将 IP 地址中属于网络号的二进制数置 0 后的结果是主机地址,因此,主机地址中从最高位开始,自值不为 0 的二进制数以后的二进制数都是属于主机号的二进制数。但主机地址中高位为 0 的二进制数未必都是属于网络号的二进制数,有可能是主机号中值为 0 的高位。因此,可以根据主机地址得出以下结果:

00000000 00011000 00100101 00001001 (0.24.37.9)

xxxxxxx xxxhhhhh hhhhhhhh hhhhhhhh

其中,x 表示待定,h 表示是属于主机号的二进制数。

最后一个可用 IP 地址是直接广播地址减 1 后的结果,由于直接广播地址是将属于主机

号的二进制数置 1 后的结果,因此,直接广播地址中,从最低位开始,自值为 0 的二进制数以后的二进制数都是属于网络号的二进制数。但从最低位开始,值为 1 的二进制数未必都是属于主机号的二进制数,有可能是网络号中值为 1 的低位。因此,可以根据最后一个可用的 IP 地址得出以下结果:

```
10111111 10011111 11111111 11111110  (最后一个可用的 IP 地址 191.159.255.254)
10111111 10011111 11111111 11111111  (直接广播地址 191.159.255.255)
nnnnnnnnn nnnxxxxx xxxxxxxx xxxxxxxx
```

其中,x 表示待定,n 表示是属于网络号的二进制数。

比较根据主机地址得出的结果和根据最后一个可用的 IP 地址得出的结果,确定 32 位 IP 地址中最高 11 位二进制数 10111111 100 是属于网络号的二进制数,最低 21 位二进制数 11000 00100101 00001001 是属于主机号的二进制数。得出 IP 地址和子网掩码是 191.152.37.9/255.224.0.0。由此可以得出表 5.8 中其他空格处的值。

5.2.5 IP 分组格式

1. 首部字段

IP 分组由首部和数据两部分组成。首部由 20B 的固定项和可变长度的可选项组成。IP 分组首部格式如图 5.18 所示。

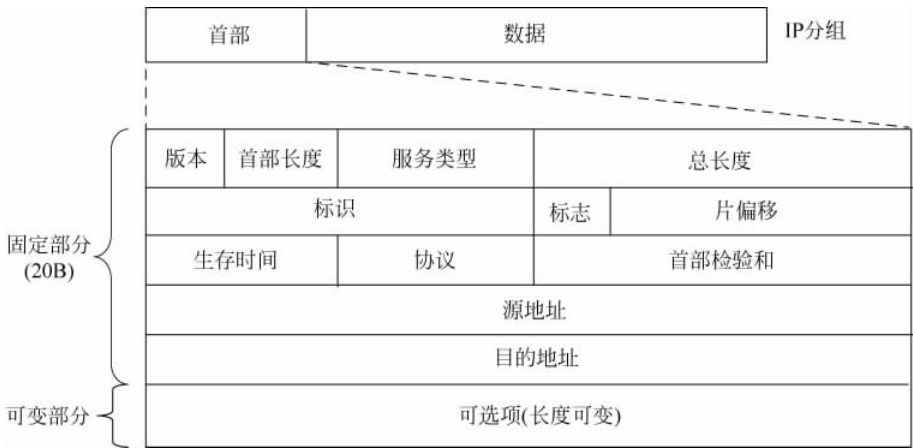


图 5.18 IP 分组首部格式

下面介绍 IP 分组首部中各字段的含义。

- (1) 版本: 4b 版本字段给出 IP 分组所属 IP 协议的版本。由于每一个 IP 分组都含有版本字段,因此允许不同版本的 IP 协议可同时在一个互连网络内运行。目前存在两种版本的 IP 协议: IPv4 和 IPv6,其版本号分别为 4 和 6。这里只讨论 IPv4,第 10 章讨论 IPv6。
- (2) 首部长度: 4b 首部长度字段以 32 位字(4B)为单位给出 IP 首部的实际长度。由于首部的长度不是固定的,需要用首部长度字段给出 IP 首部长度。字段最小值为 5,用于没有可选项的情况。由于 IP 首部长度的基本单位是 4B,意味着首部固定部分长度为 20B。最大值为 15,这就将首部长度限制在 60B 内,意味着可选项长度不能超过 40B。

(3) 服务类型: 8b 服务类型字段允许终端告诉网络它希望得到的服务, 可以通过服务类型字段指定 IP 分组的速度要求、可靠性要求及各种要求的组合。服务类型字段从左到右包括 3 位优先级位、3 位标志位(D、T、R)和目前没有使用的 2 位。3 位优先级位表示从 0(普通报文)到 7(网络控制报文)8 个分组优先级, 优先级高的 IP 分组优先得到服务。3 位标志位允许终端指定最希望得到的服务。允许指定的服务是: D—时延, T—吞吐率(实际测量到的瞬时传输速率, 或是一段时间内的平均传输速率), R—可靠性。D=1 表示该 IP 分组特别要求短的时延, T=1 表示该 IP 分组特别要求高的吞吐率, R=1 表示该 IP 分组要求尽可能不被损坏或丢弃。这些标志位可以帮助路由器选择对应的传输路径。实际上, 早先的路由器一般都不考虑这些标志位, 目前为支持多媒体应用, 路由器开始支持服务分类(COS)。1998 年, 该字段改为区分服务字段, 但只有支持区分服务(DiffServ)的网络使用该字段。

(4) 总长度: 16b 总长度字段以字节为单位给出包括首部和数据的 IP 分组的长度, 最大长度值为 65 535B。

(5) 标识: 16b 标识字段用于标识属于同一 IP 分组的数据片, 属于同一 IP 分组的数据片具有相同的标识字段值。发送端维持一个计数器, 每发送一个 IP 分组, 计数器加 1, 计数器的值就作为 IP 分组的标识字段值。

(6) 标志: 3b 标志字段包含 1 位保留位、1 位标志位 DF 和 1 位标志位 MF。DF 位置 1 要求不能对 IP 分组分片, 禁止路由器把 IP 分组分片成多个数据片。一旦 IP 分组中的 DF 位置 1, 表明该 IP 分组只能作为单个数据片传送, 这就要求路由器即使选择一条并不是最佳的传输路径, 也要避开只能传输长度很短的 IP 分组的传输网络。要求所有传输网络至少能传输小于 576B 的 IP 分组。MF 位置 0 表示是若干数据片中最后一个数据片, 除最后一个数据片外, IP 分组分片后所生成的所有其他数据片都必须将 MF 位置 1。MF 位用于让接收端判别某个 IP 分组分片后所生成的所有数据片是否已全部接收到。

(7) 片偏移: 13b 片偏移字段以 8B 为单位给出该数据片在分片前的原始数据中的起始位置。因此, 除最后数据片以外的所有其他数据片的长度必须是 8B 的倍数。

(8) 生存时间: 字段长度为 8b。此字段是用于限制 IP 分组存在时间的一个计数器, 假定该计数器以秒为单位计数, IP 分组允许存在的最长时间为 255s。目前, 该字段只是作为最大跳数使用, IP 分组每经过一跳路由器, 该字段值减 1, 当值减为 0 时, 丢弃该 IP 分组并发送一个警告消息给源终端。设置该字段的目的是为了避免 IP 分组在网络上无休止地漂荡。

(9) 协议: 字段长度为 16b。IP 分组中的数据是上层协议数据单元(Protocol Data Unit, PDU), 协议字段值给出了作为 IP 分组数据的 PDU 的协议类型, 例如, 协议字段值 6 表示 IP 分组中的数据是 TCP 报文, 协议字段值 17 表示 IP 分组中的数据是 UDP 报文。协议字段的作用是确定处理该 IP 分组中数据的进程, 即上层协议进程。TCP 进程和 UDP 进程是最有可能处理该 IP 分组中数据的进程。

(10) 首部检验和: 字段长度为 16b, 对首部用检验和算法计算出的检错码, 用于检测首部传输过程中发生的错误。每一跳路由器需要重新计算首部检验和, 因为每经过一跳路由器, 至少改变了一个首部字段值(生存时间字段)。

(11) 源地址和目的地址: 字段长度分别为 32b。源地址字段给出了源终端的 IP 地址,

目的地址字段给出了目的终端的 IP 地址。

(12) 可选项：设计该字段的目的如下。

- ① 允许以后协议版本提供原始设计中遗漏的信息；
- ② 允许经验丰富的人试验一些新的想法；
- ③ 避免在报文首部中固定分配一些并不常用的信息字段。

可选项长度可变。目前，定义了以下 5 种可选项。

① 保密：该选项给出如何保密 IP 分组。和军事应用有关的路由器可以用该选项来避开某些认为不安全的国家或地区。实际上，所有路由器都忽略该选项。

② 严格源站选路：该选项给出从源终端到目的终端完整传输路径的 IP 地址列表，IP 分组必须严格遵循给出的传输路径。系统管理员可以用这种功能在路由器路由表损坏的情况下发送紧急 IP 分组，或者用于发送测量时间参数的 IP 分组。

③ 不严格源站选路：该选项要求 IP 分组一定要经过列表中指定的路由器，并按指定的顺序经过。但允许通过传输路径上别的路由器。通常通过该选项指定少数几个路由器来强迫 IP 分组经过某一特殊传输路径。例如，强迫从伦敦到悉尼的 IP 分组经过美国西部而不是东部时，选项可指定 IP 分组必须经过纽约、洛杉矶、檀香山的路由器。当出于某种政治或经济考虑，需要 IP 分组经过或避开某些地区或国家时，可用该选项。

④ 记录路由：该选项要求所有经过的路由器把它们的 IP 地址添加到该选项字段中，通过记录路由，可以帮助系统管理员查出路由算法中的一些问题。由于 ARPA 网刚建立时，IP 分组经过的路由器最多不超过 9 个，因此用 40B 记录经过的路由器已经很充足了，但对现在的 Internet 来说，用 40B 记录经过的路由器则远远不够了。

⑤ 时间戳：该选项基本上与记录路由选项一样，不同的是，除记录 32 位 IP 地址外，还记录 32 位时间戳。该选项也主要用于诊断路由算法发生的错误。

IP 分组首部的可选项有很强的了解、管理网络的功能，常常被用来作为侦察网络的工具。为了网络的安全性，路由器需要关闭对一些可选项的支持功能。

2. 分片

传输网络链路层帧净荷字段允许的最大长度称为最大传送单元(Maximum Transfer Unit, MTU)，如以太网的 MTU 为 1500B，如果 IP 分组长度超过传输该 IP 分组的传输网络的 MTU，必须将 IP 分组分片。分片是将 IP 分组净荷字段中的数据分割为多个数据片的过程，除了最后一个数据片，其他数据片的长度必须是 8B 的整数倍。每一个数据片加上 IP 首部构成 IP 分组，必须保证分片后的数据片长度和 IP 首部长度和小于传输网络的 MTU。通常情况下，除最后一个数据片，其他数据片长度的分配原则是：应是 8 的倍数，且加上 IP 首部后尽量接近 MTU。为了标识这些由分片同一个 IP 分组净荷字段中的数据产生的 IP 分组序列，这些 IP 分组必须具有相同的标识字段值。为了在目的端将这些 IP 分组中净荷字段包含的数据片重新还原为原始数据，这些 IP 分组中的每一个 IP 分组必须在片偏移字段中给出该 IP 分组包含的数据片在原始数据中的起始位置。为了让目的端确定所有数据片对应的 IP 分组是否均已到达，必须标志最后一个数据片对应的 IP 分组。分片过程如图 5.19 所示，4000B 数据被分成 3 个数据片，长度分别是 1480B、1480B 和 1040B。这 3 个数据片在原始数据中的起始位置分别是 0、1480 和 2960，求出对应的片偏移分别是 $0/8 =$

0、 $1480/8=185$ 和 $2960/8=370$ 。

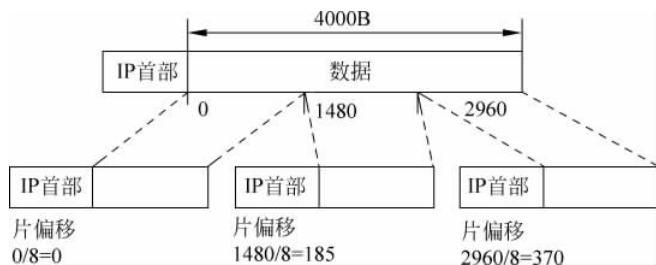


图 5.19 分片过程

【例 5.4】 终端 A 和终端 B 之间的传输路径由网络 1、网络 2 和网络 3 组成,其中网络 1 的 MTU=1500B,网络 2 的 MTU=800B,网络 3 的 MTU=420B,假定终端 A 传输给终端 B 的数据的长度为 1440B,给出终端 A 及传输路径经过的各个路由器分片数据的过程。

【解析】 终端 A 及传输路径经过的路由器分片数据的过程如图 5.20 所示。终端 A 生成的 IPv4 分组的总长度为 1460B(包括 20B 首部 and 1440B 净荷),由于终端 A 连接路由器 R1 的链路的 MTU=1500B,终端 A 可以直接将总长度为 1460B 的 IP 分组传输给路由器 R1。当路由器 R1 向路由器 R2 传输该 IP 分组时,发现输出链路的 MTU=800B,需要对 IP 分组进行分片操作。路由器 R1 将 IP 分组的净荷分片成两个数据片,两个数据片的长度分别为 776B 和 664B,加上 20B 的 IPv4 首部后,分别构成两个总长度分别为 796B(20B 首部+776B 净荷)和 684B 的 IPv4 分组。这两个 IPv4 分组的标识符字段值相同,第二个 IPv4 分组的片偏移=776/8=97。同样,当路由器 R2 向终端 B 传输这两个 IPv4 分组时,发现输出链路的 MTU=420B,路由器 R2 需要再一次对这两个 IPv4 分组进行分片操作,776B 的数据片被分片成长度分别为 400B 和 376B 的两个数据片,同样,664B 数据片被分片成长度分别为 400B 和 264B 的两个数据片,这 4 个数据片加上 IP 首部后构成 4 个 IP 分组,原来 M 标志位为 1 的 IPv4 分组分片后生成的 IPv4 分组序列的 M 标志位都为 1。原来 M 标志位为 0 的 IPv4 分组分片后生成的 IPv4 分组序列,除由最后一个数据片构成的 IPv4 分组外,其他 IPv4 分组的 M 标志位也都为 1。这些 IPv4 分组的标识字段值都相同,图 5.20 中每一个 IP 分组首部中的片偏移给出净荷中的数据片在原始净荷中的位置。

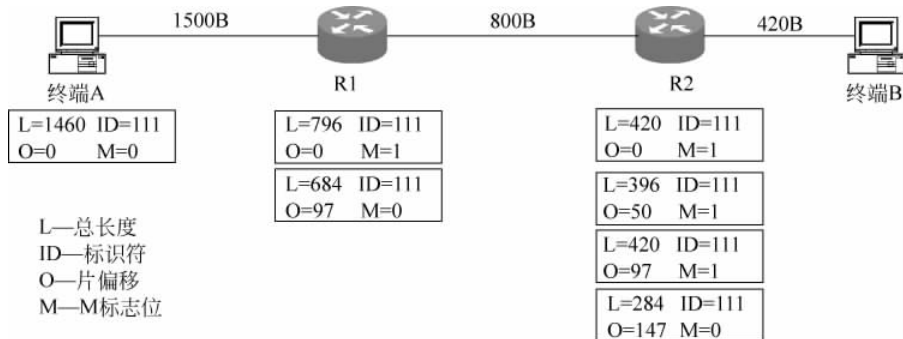


图 5.20 分片数据过程

5.3 路由表和 IP 分组传输过程

终端配置的默认网关地址和路由器中的路由表确定了连接在任意类型网络上的两个终端之间的 IP 传输路径,源终端发送给目的终端的 IP 分组沿着源终端至目的终端的 IP 传输路径,经过路由器逐跳转发,到达目的终端。

5.3.1 互联网结构与路由表

互联网是指由不同网络互连而成的网际网,网际层将任何两个分配不同网络地址的网络作为两个不同的网络,因此,两个类型不同且分配不同网络地址的传输网络,两个物理上独立且分配不同网络地址的以太网,两个共享同一物理以太网但逻辑上相互独立且分配不同网络地址的 VLAN 等,都是两个不同的网络。对于连接在相同物理网络上的两组终端,只要这两组终端分配了网络地址不同的 IP 地址,这两组终端就属于两个不同的网络。必须通过路由器或三层路由设备实现不同网络之间的互连,即如果两个终端分配了网络地址不同的 IP 地址,这两个终端之间的传输路径必须包含路由器或三层路由设备。

1. 互联网结构

互联网结构如图 5.21 所示,3 个路由器互连 4 个网络,为了着重讨论终端 A 与终端 B 之间的数据传输过程,图中只是详细给出 LAN1 和 LAN2 的网络地址。连接在 LAN1 上的终端和路由器接口分配的 IP 地址必须属于 CIDR 地址块 192.1.1.0/24。同样,连接在 LAN2 上的终端和路由器接口分配的 IP 地址必须属于 CIDR 地址块 192.1.2.0/24。路由器每一个接口分配的 IP 地址和子网掩码确定该路由器接口连接的网络的网络地址。由于路由器 R1 连接 LAN1 的接口分配的 IP 地址和子网掩码分别是 192.1.1.254 和 255.255.255.0,因此得出 LAN1 的网络地址为 192.1.1.0/24。同样,路由器 R3 连接 LAN2 的接口分配的 IP 地址和子网掩码分别是 192.1.2.254 和 255.255.255.0,因此得出 LAN2 的网络地址为 192.1.2.0/24。路由器连接不同网络的接口必须分配网络地址不同的 IP 地址。

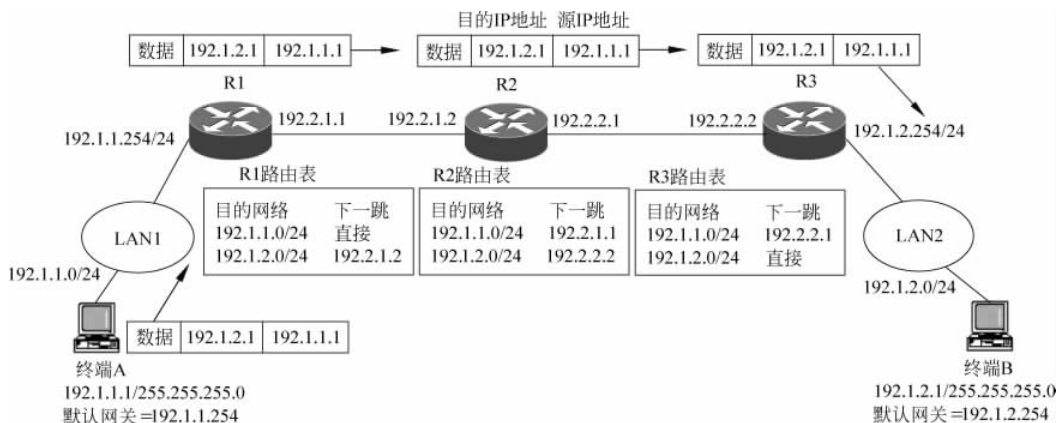


图 5.21 IP 分组传输过程

2. 路由表

路由器中的路由表给出连接在不同网络上的终端之间的 IP 传输路径,IP 传输路径由终端与路由器组成。例如终端 A 至终端 B 的 IP 传输路径为:终端 A→路由器 R1→路由器 R2→路由器 R3→终端 B。因为 192.1.1.254 是路由器 R1 连接 LAN1 的接口的 IP 地址,终端 A 根据配置的默认网关地址 192.1.1.254 确定终端 A 至终端 B 的 IP 传输路径上的下一跳是路由器 R1。路由器 R1 根据路由表确定通往终端 B 的 IP 传输路径上的下一跳是路由器 R2。路由器 R1 用终端 B 的 IP 地址检索路由表,确定路由表中与 IP 地址 192.1.2.1 匹配的路由项是<192.1.2.0/24,192.2.1.2>,下一跳 IP 地址 192.2.1.2 是路由器 R2 连接路由器 R1 的接口的 IP 地址。由于路由器 R3 路由表中与 IP 地址 192.1.2.1 匹配的路由项是<192.1.2.0/24,直接>,下一跳是直接表示终端 B 连接在路由器 R3 直接连接的网络上,因此,路由器 R3 确定通往终端 B 的 IP 传输路径上的下一跳是终端 B 自身。

5.3.2 IP 分组传输过程

为了更深刻地理解路由表的作用,详细讨论图 5.21 中终端 A 至终端 B 的 IP 分组传输过程。

1. 确定源和目的终端是否在同一个网络

终端 A 向终端 B 传输数据前,必须先获取终端 B 的 IP 地址,然后将数据封装成以终端 A 的 IP 地址 192.1.1.1 为源 IP 地址、以终端 B 的 IP 地址 192.1.2.1 为目的 IP 地址的 IP 分组,在进行 IP 分组传输前,先确定终端 B 是否和终端 A 位于同一个网络,步骤如下:

(1) 终端 A 根据自己的 IP 地址 192.1.1.1 和子网掩码 255.255.255.0,求出网络地址 192.1.1.0/24。

(2) 终端 A 根据终端 B 的 IP 地址 192.1.2.1 和自己的子网掩码 255.255.255.0,求出终端 B 的网络地址 192.1.2.0/24。

(3) 比较两个网络地址。

- 如果两个网络地址相同,说明终端 A 和终端 B 位于同一个网络,终端 A 至终端 B 的 IP 分组传输过程无须经过路由器。
- 如果两个网络地址不相同,说明终端 A 和终端 B 位于不同的网络,终端 A 将 IP 分组传输给终端 A 至终端 B 的 IP 传输路径上的第一跳路由器。

2. 根据默认网关地址找到第一跳路由器

一旦确定终端 B 和终端 A 不在同一个网络,终端 A 将 IP 分组转发给终端 A 至终端 B 的 IP 传输路径上的第一跳路由器,该路由器的 IP 地址通过终端 A 配置的默认网关地址获得。如果连接终端 A 和第一跳路由器的网络是以太网,必须将 IP 分组封装成以终端 A 的 MAC 地址为源 MAC 地址、以第一跳路由器连接以太网的端口的 MAC 地址为目的 MAC 地址的 MAC 帧,然后将 MAC 帧通过以太网传输给第一跳路由器,IP 分组经过以太网实现当前跳至下一跳的传输过程在 5.4 节中详细讨论。

3. 路由器逐跳转发

IP 分组到达路由器 R1 后,路由器 R1 根据 IP 分组的目的 IP 地址和路由表中的路由项确定该 IP 分组的下一跳路由器,步骤如下:

(1) 对应每一项路由项,根据路由项的子网掩码,求出目的 IP 地址对应的网络地址。由于路由器 R1 中每一个路由项中的目的网络字段给出的网络前缀位数都是 24,目的 IP 地址根据不同路由项的子网掩码求出的网络地址是相同的,都是 192.1.2.0。

(2) 用根据 IP 分组的目的 IP 地址和路由项子网掩码求出的网络地址比较每一个路由项的目的网络字段值,如果有若干路由项的目的网络字段值和目的 IP 地址对应该路由项求出的网络地址相同,则选择其中网络前缀最长的路由项作为最终匹配的路由项。

(3) 路由器 R1 的路由表中,只有路由项<192.1.2.0/24,192.2.1.2>和目的 IP 地址匹配,IP 分组被转发给 IP 地址为 192.2.1.2 的下一跳路由器。

(4) 传输路径上的路由器依次逐跳转发,IP 分组到达传输路径上最后一跳路由器 R3。

4. 直接交付

路由器 R3 中和 IP 分组目的 IP 地址匹配的路由项是<192.1.2.0/24,直接>,表明该路由器和终端 B 之间不再有其他路由器,即终端 B 和该路由器的一个接口连接在同一个网络上,路由器通过该网络将 IP 分组直接传输给终端 B。

5.3.3 实现 IP 分组传输过程的思路

从上述讨论的 IP 分组端到端传输过程可以得出以下实现 IP 分组端到端传输过程的基本思路:

(1) 建立一条以源终端为始点,以目的终端为终点,中间由若干路由器组成的 IP 分组端到端传输路径,IP 分组沿着端到端传输路径逐跳转发。源终端通过配置的默认网关地址获得第一跳路由器的 IP 地址。中间路由器根据路由表和 IP 分组的目的 IP 地址确定下一跳路由器地址。

(2) 在获取下一跳路由器的 IP 地址后,经过 IP over X 技术与 X 的物理层和链路层功能,实现 IP 分组当前跳至下一跳的传输过程,X 是连接当前跳和下一跳的传输网络,如以太网。

建立端到端传输路径的关键是每一个路由器建立路由表,路由表中每一个路由项指出通往特定网络的传输路径上的下一跳路由器,因此,解决 IP 分组端到端传输的第一步是为互联网中的每一个路由器建立路由表。

5.3.4 直连路由项和静态路由项

路由项分为直连路由项、静态路由项和动态路由项,直连路由项在完成路由器接口 IP 地址和子网掩码配置后由路由器自动生成。静态路由项通过人工配置。动态路由项由路由器通过运行路由协议生成。本节讨论直连路由项和静态路由项的建立过程。第 6 章讨论路由协议和动态路由项建立过程。

1. 互联网结构

互联网结构与路由器接口配置的 IP 地址和子网掩码如图 5.22 所示。完成网络设备之间、终端和网络设备之间的连接后,首先需要配置路由器接口,每一个连接网络的路由器接口需要配置 IP 地址和子网掩码,路由器接口配置的 IP 地址和子网掩码确定了该接口连接的网络的网络地址。两个路由器接口如果连接在同一个网络上,需要配置网络号相同、主机号不同的 IP 地址;两个路由器接口如果连接在不同的网络上,需要配置网络号不同的 IP 地址。一般情况下,同一路由器的不同接口不能配置网络号相同、主机号不同的 IP 地址,即不能将同一路由器的不同接口连接到同一个网络上。

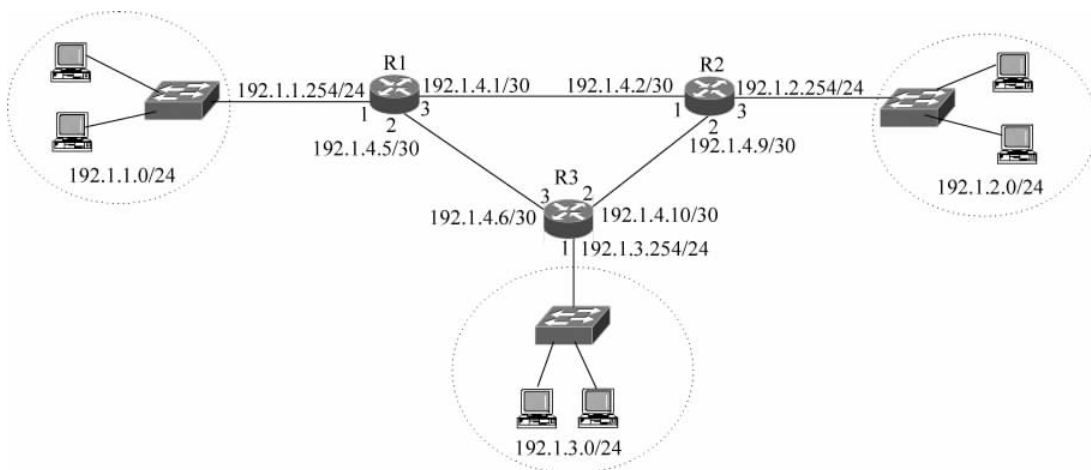


图 5.22 互联网结构

2. 直连路由项

完成每一个路由器接口 IP 地址和子网掩码配置,且将每一个路由器接口连接到某个网络后,路由器为该接口连接的网路自动创建一个路由项,将这种用于指明通往直接相连的网路的路由项称为直连路由项。例如路由器 R1 接口 2 配置 IP 地址 192.1.4.5 和子网掩码 255.255.255.252(30 位网络前缀)后,得出该接口连接的网路的网络地址是 192.1.4.4/30,因此,对应的直连路由项中的目的网络是 192.1.4.4,子网掩码是 255.255.255.252。当为图 5.22 所示的所有路由器接口配置 IP 地址和子网掩码,且将接口连接到网络后,每一个路由器自动生成只包含直连路由项的路由表,如表 5.9 至表 5.11 所示。

表 5.9 路由器 R1 路由表

目的网络	子网掩码	下一跳	输出接口
192.1.1.0	255.255.255.0	直接	1
192.1.4.0	255.255.255.252	直接	3
192.1.4.4	255.255.255.252	直接	2

表 5.10 路由器 R2 路由表

目的网络	子网掩码	下一跳	输出接口
192.1.2.0	255.255.255.0	直接	3
192.1.4.0	255.255.255.252	直接	1
192.1.4.8	255.255.255.252	直接	2

表 5.11 路由器 R3 路由表

目的网络	子网掩码	下一跳	输出接口
192.1.3.0	255.255.255.0	直接	1
192.1.4.4	255.255.255.252	直接	3
192.1.4.8	255.255.255.252	直接	2

3. 静态路由项

直连路由项只给出通往该路由器直接相连的网络的传输路径,由于路由表中没有用于指明通往未与该路由器直接相连的网络的传输路径的路由项,因此,路由器只能实现连接在直接相连的不同网络上的两个终端之间的 IP 分组传输过程。

对于没有直接相连的网络,必须通过分析,得出该路由器通往该网络的传输路径,并确定该传输路径上下一跳路由器的 IP 地址。对于路由器 R1,网络地址为 192.1.2.0/24 的网络没有与其直接相连,得出路由器 R1 通往该网络的传输路径是: R1→R2→网络 192.1.2.0/24,确定路由器 R1 通往网络 192.1.2.0/24 的传输路径上的下一跳是路由器 R2,下一跳 IP 地址是路由器 R2 接口 1 的 IP 地址 192.1.4.2。选择路由器 R2 接口 1 的 IP 地址作为下一跳的 IP 地址的原因是,路由器 R2 接口 1 与路由器 R1 接口 3 连接在互连路由器 R1 和 R2 的网络上,图 5.22 中互连路由器 R1 和 R2 的网络是点对点链路。

得出通往某个目的网络的传输路径,并确定该传输路径上下一跳的 IP 地址后,可以人工配置路由项,这些人工配置的路由项称为静态路由项。路由器 R1、R2 和 R3 包含静态路由项的完整路由表如表 5.12 至表 5.14 所示。

表 5.12 路由器 R1 完整路由表

目的网络	子网掩码	下一跳	输出接口
192.1.1.0	255.255.255.0	直接	1
192.1.2.0	255.255.255.0	192.1.4.2	3
192.1.3.0	255.255.255.0	192.1.4.6	2
192.1.4.0	255.255.255.252	直接	3
192.1.4.4	255.255.255.252	直接	2

表 5.13 路由器 R2 完整路由表

目的网络	子网掩码	下一跳	输出接口
192.1.1.0	255.255.255.0	192.1.4.1	1
192.1.2.0	255.255.255.0	直接	3
192.1.3.0	255.255.255.0	192.1.4.10	2

续表

目的网络	子网掩码	下一跳	输出接口
192.1.4.0	255.255.255.252	直接	1
192.1.4.8	255.255.255.252	直接	2

表 5.14 路由器 R3 完整路由表

目的网络	子网掩码	下一跳	输出接口
192.1.1.0	255.255.255.0	192.1.4.5	3
192.1.2.0	255.255.255.0	192.1.4.9	2
192.1.3.0	255.255.255.0	直接	1
192.1.4.4	255.255.255.252	直接	3
192.1.4.8	255.255.255.252	直接	2

5.3.5 例题解析

【例 5.5】 互联网结构如图 5.23 所示,路由表每一个路由项包含字段<目的网络,子网掩码,下一跳,输出接口>,回答下列问题:

(1) 将 CIDR 地址块 202.115.1.0/24 划分为两个子网地址,分别分配给 LAN 1 和 LAN 2,每个子网分配的 IP 地址数不少于 120,给出子网地址划分结果,说明理由或给出必要的计算过程。

(2) 给出 R1 的路由表,包含用于指明通往图 5.23 中所有网络和服务器的传输路径的路由项。

(3) 给出 R2 路由表中用于指明通往 LAN 1 和 LAN 2 的传输路径的路由项。

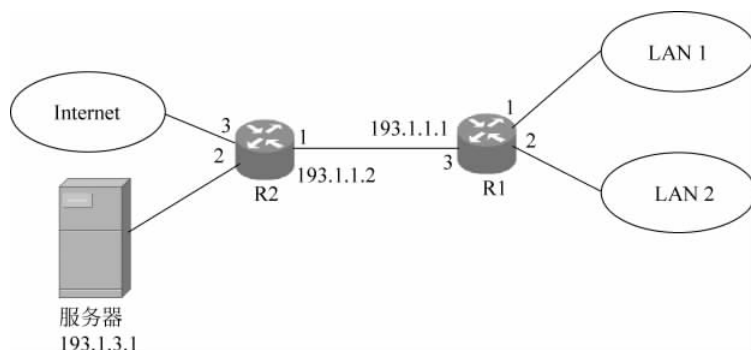


图 5.23 互联网结构

【解析】 (1) 由于要求每一个子网的 IP 地址数不小于 120,因此,每一个子网至少用 7 位二进制数作为主机号,这样,可以把给定的 CIDR 地址块分成两个等长的网络地址空间,把原来的主机号位数减少 1 位,由 8 位变为 7 位,并用这减下的 1 位标识两个不同的子网地址,这样,每一个网络地址空间中的网络号位数由 24 位变为 25 位。将 202.115.1.0/24 展开,用 25 位网络号重新划分子网后,可以得出如下两个子网的 IP 地址范围:

11001010 01110011 00000001 0 0000000

11001010 01110011 00000001 0 1111111

子网 0,地址范围 202.115.1.0~202.115.1.127
网络地址: 202.115.1.0
子网掩码: 255.255.255.128

11001010 01110011 00000001 1 0000000

11001010 01110011 00000001 1 1111111

子网 1,地址范围 202.115.1.128~202.115.1.255
网络地址: 202.115.1.128
子网掩码: 255.255.255.128

(2) 路由器 R1 的路由表如表 5.15 所示。由于两个子网和路由器 R1 直接相连,表中用于指明通往两个子网的传输路径的路由项的下一跳字段值为“直接”。用于指明通往服务器的传输路径的路由项中的目的网络和子网掩码字段值必须唯一匹配该服务器的 IP 地址。因此,用 32 位全 1 的子网掩码表示目的网络只包含单个 IP 地址 193.1.3.1。由于 Internet 是由无数个网络组成的,因此,只能以默认路由项指明通往 Internet 的传输路径。由于路由器 R1 至服务器和 Internet 传输路径上的下一跳是路由器 R2,且路由器 R2 接口 1 和路由器 R1 接口 3 连接在同一个网络上,用路由器 R2 接口 1 的 IP 地址作为这些路由项的下一跳 IP 地址。

表 5.15 路由器 R1 路由表

目的网络	子网掩码	下一跳	输出接口
202.115.1.0	255.255.255.128	直接	1
202.115.1.128	255.255.255.128	直接	2
193.1.3.1	255.255.255.255	193.1.1.2	3
0.0.0.0	0.0.0.0	193.1.1.2	3

(3) 路由器 R2 应该有两项路由项分别用于指明通往 LAN 1 和 LAN 2 的传输路径,但这两个路由项有着相同的输出接口和下一跳,这样的路由项可以尝试聚合为一个路由项,前提是聚合后的目的网络和子网掩码字段确定的 IP 地址空间等于聚合前两个路由项所包含的 IP 地址空间,这里,202.115.1.0~202.115.1.127 和 202.115.1.128~202.115.1.255 合并为 202.115.1.0~202.115.1.255,因此,可以完成如图 5.24 所示的合并过程。

目的网络	子网掩码	下一跳	接口	目的网络	子网掩码	下一跳	接口
202.115.1.0	255.255.255.128	193.1.1.1	1	202.115.1.0	255.255.255.0	193.1.1.1	1
202.115.1.128	255.255.255.128	193.1.1.1	1				

图 5.24 路由项合并过程

【例 5.6】 互联网结构如图 5.25 所示,两个路由器互连 3 个网络,网络旁边的数字表示该网络连接的终端数。假定 3 个网络共享 CIDR 地址块 192.1.1.64/26。将 CIDR 地址块 192.1.1.64/26 根据网络连接的终端数划分为 3 个 CIDR 地址块,并将其分配给 3 个网络,每一个网络从最大可用 IP 地址开始分配路由器连接该网络的接口的 IP 地址,由此确定路由器 R1、R2 的路由表与终端 A、终端 B 和终端 C 的网络配置信息。

【解析】 (1) 将 CIDR 地址块 192.1.1.64/26 分解为 3 个 CIDR 地址块,每一个 CIDR 地址块的网络地址相同,且使得一个 CIDR 地址块中的有效 IP 地址数 $\geq 27+1$,一个 CIDR

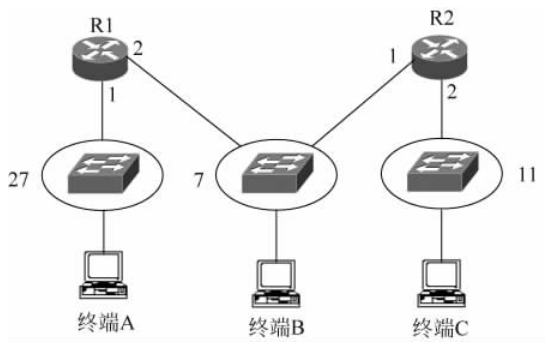


图 5.25 互联网络结构

地址块中的有效 IP 地址数 $\geq 7+2$,一个 CIDR 地址块中的有效 IP 地址数 $\geq 11+1$ 。由此确定其中一个 CIDR 地址块的网络号位数为 27,主机号位数为 5,另外两个 CIDR 地址块的网络号位数为 28,主机号位数为 4。将 CIDR 地址块 192.1.1.64/26 分解为 3 个 CIDR 地址块的过程如下:

- $\left. \begin{array}{l} 01\ 0\ 00000 \\ 01\ 0\ 11111 \end{array} \right\}$ 网络地址为 192.1.1.64/27,有效 IP 地址范围是 192.1.1.65~192.1.1.94
- $\left. \begin{array}{l} 01\ 1\ 0\ 0000 \\ 01\ 1\ 0\ 1111 \end{array} \right\}$ 网络地址为 192.1.1.96/28,有效 IP 地址范围是 192.1.1.97~192.1.1.110
- $\left. \begin{array}{l} 01\ 1\ 1\ 0000 \\ 01\ 1\ 1\ 1111 \end{array} \right\}$ 网络地址为 192.1.1.112/28,有效 IP 地址范围是 192.1.1.113~192.1.1.126

(2) 将网络地址 192.1.1.64/27 分配给路由器 R1 接口 1 连接的网络,路由器 R1 接口 1 的 IP 地址和子网掩码为 192.1.1.94/27(最大可用 IP 地址)。

将网络地址 192.1.1.96/28 分配给路由器 R1 接口 2 连接的网络,路由器 R1 接口 2 的 IP 地址和子网掩码为 192.1.1.110/28(最大可用 IP 地址),路由器 R2 接口 1 的 IP 地址和子网掩码为 192.1.1.109/28(次大可用 IP 地址)。

将网络地址 192.1.1.112/28 分配给路由器 R2 接口 2 连接的网络,路由器 R2 接口 2 的 IP 地址和子网掩码为 192.1.1.126/28(最大可用 IP 地址)。

(3) 确定路由器 R1 和 R2 的路由表如表 5.16 和表 5.17 所示。

表 5.16 路由器 R1 的路由表

目的网络	子网掩码	下一跳	输出接口
192.1.1.64	255.255.255.224	直接	1
192.1.1.96	255.255.255.240	直接	2
192.1.1.112	255.255.255.240	192.1.1.109	2

表 5.17 路由器 R2 的路由表

目的网络	子网掩码	下一跳	输出接口
192.1.1.64	255.255.255.224	192.1.1.110	1
192.1.1.96	255.255.255.240	直接	1
192.1.1.112	255.255.255.240	直接	2

(4) 得出终端 A 的一种正确的网络配置信息是：IP 地址和子网掩码为 192.1.1.65/27，默认网关地址为 192.1.1.94。

得出终端 B 的一种正确的网络配置信息是：IP 地址和子网掩码为 192.1.1.97/28，默认网关地址为 192.1.1.110。

得出终端 C 的一种正确的网络配置信息是：IP 地址和子网掩码为 192.1.1.113/28，默认网关地址为 192.1.1.126。

【例 5.7】 根据图 5.26 所示的互联网结构完成如表 5.18 所示的路由器 RG 路由表中的路由项，给出①～⑥的值。

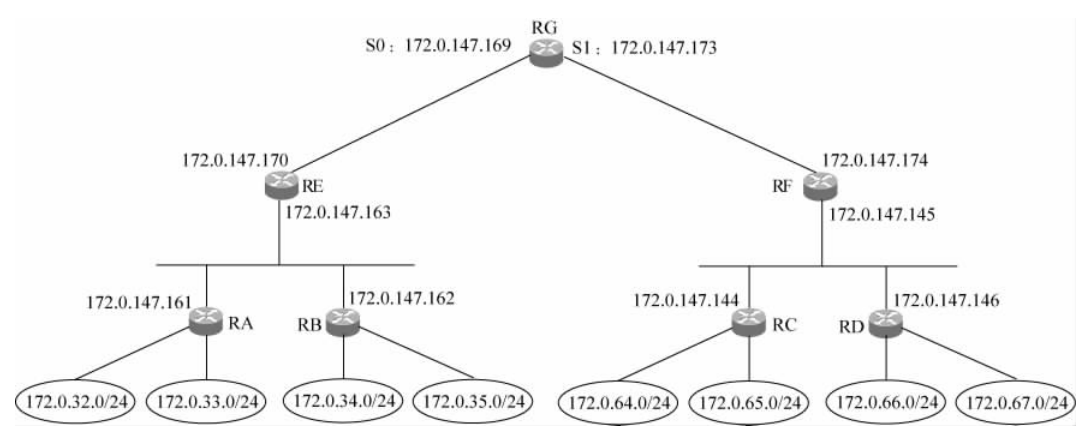


图 5.26 互联网结构

表 5.18 路由器 RG 路由表

目的网络	输出接口
①	S0(直接连接)
②	S1(直接连接)
③	S0
④	S1
⑤	S0
⑥	S1

【解析】 (1) 求出每一个网络的网络地址。
互连路由器 RG 和 RE 的网络的网络地址必须包括 IP 地址 172.0.147.169 和 172.0.147.170，因此网络地址可以是 172.0.147.168/30 或者 172.0.147.168/29 等，但网络地址 172.0.147.168/29 的 IP 地址范围是 172.0.147.168～172.0.147.175，已经包括互连路由器 RG 和 RF 的网络中的 IP 地址，因此，互连路由器 RG 和 RE 的网络的网络地址只能是 172.0.147.168/30。

互连路由器 RG 和 RF 的网络的网络地址必须包括 IP 地址 172.0.147.173 和 172.0.147.174，因此网络地址可以是 172.0.147.172/30 或者 172.0.147.168/29 等，显然，172.0.147.168/29 等网络地址包含其他网络中的 IP 地址，因此，互连路由器 RG 和 RE 的网络的网络地址只能是 172.0.147.172/30。

互连路由器 RA、RB 和 RE 的网络的网络地址必须包括 IP 地址 172.0.147.161、172.

0.147.162 和 172.0.147.163,163 的 8 位二进制数是 10100011,由于主机号全 1 不能作为有效 IP 地址,因此,IP 地址至少需要包含 3 位主机号,由此得出网络地址可以是 172.0.147.160/29 或者 172.0.147.160/28 等,但 172.0.147.160/28 的 IP 地址范围是 172.0.147.160~172.0.147.175,已经包含其他网络中的 IP 地址,因此,互连路由器 RA、RB 和 RE 的网络的网络地址只能是 172.0.147.160/29。

互连路由器 RC、RD 和 RF 的网络的网络地址必须包括 IP 地址 172.0.147.144、172.0.147.145 和 172.0.147.146,144 的 8 位二进制数是 10010000,由于主机号全 0 不能作为有效 IP 地址,因此,IP 地址至少需要包含 5 位主机号,由此得出网络地址可以是 172.0.147.128/27 或者 172.0.147.128/26 等,但网络地址 172.0.147.128/26 的 IP 地址范围是 172.0.147.128~172.0.147.191,已经包含其他网络中的 IP 地址,因此,互连路由器 RC、RD 和 RF 的网络的网络地址只能是 172.0.147.128/27。

(2) 聚合网络地址。

网络地址 172.0.32.0/24、172.0.33.0/24、172.0.34.0/24 和 172.0.35.0/24 可以聚合为 CIDR 地址块 172.0.32.0/22,网络地址 172.0.64.0/24、172.0.65.0/24、172.0.66.0/24 和 172.0.67.0/24 可以聚合为 CIDR 地址块 172.0.64.0/22。

(3) 完成路由器 RG 路由表。

根据上述分析结果,可以得到以下目的网络地址,依次填入表 5.18 中:

172.0.147.168/30

172.0.147.172/30

172.0.147.160/29

172.0.147.128/27

172.0.32.0/22

172.0.64.0/22

【例 5.8】 确定下述主机对是否连接在同一个网络上。

(1) 主机 1: 172.15.5.72/255.255.255.0; 主机 2: 172.15.5.79/255.255.255.0。

(2) 主机 1: 192.168.19.35/255.255.255.224; 主机 2: 192.168.19.48/255.255.255.224。

(3) 主机 1: 10.128.14.14/255.255.255.240; 主机 2: 10.128.14.19/255.255.255.240。

(4) 主机 1: 192.168.3.68/255.255.255.248; 主机 2: 192.168.3.74/255.255.255.248。

【解析】 判定两个主机是否连接在同一个网络上的依据是这两个主机的网络地址是否相同,主机的网络地址是主机的 IP 地址与主机的子网掩码“与”操作的结果。

(1) 由于子网掩码由 24 位 1 和 8 位 0 构成,因此,主机的网络地址是将主机 IP 地址最后 8 位清零后的结果,因此,主机 1 的网络地址为 172.15.5.0/24,主机 2 的网络地址为 172.15.5.0/24,主机 1 和主机 2 连接在同一个网络上。

(2) 由于子网掩码由 27 位 1 和 5 位 0 构成,且两个 IP 地址的前 24 位相同,因此,只需计算出 IP 地址最后 1 个字节的前 3 位值:

$00100011 (35) \& 11100000 (224) = 00100000 (32)$

$00110000 (48) \& 11100000 (224) = 00100000 (32)$

由此得出主机 1 和主机 2 的网络地址均是 192.168.19.32/27。主机 1 和主机 2 连接在同一个网络上。

(3) 由于子网掩码由 28 位 1 和 4 位 0 构成,且两个 IP 地址的前 24 位相同,因此,只需计算出 IP 地址最后 1 个字节的前 4 位值:

$00001110 (14) \& 11110000 (240) = 00000000 (0)$

$00010011 (19) \& 11110000 (240) = 00010000 (16)$

由此得出主机 1 的网络地址为 10.128.14.0/28,主机 2 的网络地址为 10.128.14.16/28。主机 1 和主机 2 连接在不同的网络上。

(4) 由于子网掩码由 29 位 1 和 3 位 0 构成,且两个 IP 地址的前 24 位相同,因此,只需计算出 IP 地址最后 1 个字节的前 5 位值:

$01000100 (68) \& 11111000 (248) = 01000000 (64)$

$01001010 (74) \& 11111000 (248) = 01001000 (72)$

由此得出主机 1 的网络地址为 192.168.3.64/29,主机 2 的网络地址为 192.168.3.72/29。主机 1 和主机 2 连接在不同的网络上。

5.4 IP over 以太网

IP over 以太网和以太网共同完成以下功能:①在确定互连当前跳与下一跳的传输网络为以太网和下一跳 IP 地址后,解析出下一跳连接以太网接口的 MAC 地址;②将 IP 分组封装成以当前跳连接以太网接口的 MAC 地址为源 MAC 地址、下一跳连接以太网接口的 MAC 地址为目的 MAC 地址的 MAC 帧;③经过互连当前跳与下一跳的以太网,完成该 MAC 帧当前跳至下一跳的传输过程。其中,功能①和②由 IP over 以太网实现,即由以太网对应的网络接口层实现,功能③由以太网的物理层和 MAC 层实现。

5.4.1 ARP 和地址解析过程

如图 5.27 所示,终端 A 和服务器 B 连接在以太网上,即使如此,终端 A 访问服务器 B 时所给出的也不会是服务器 B 的 MAC 地址,而往往是服务器 B 的域名,经过域名系统解析后得到的也是服务器 B 的 IP 地址。根据以太网交换机的工作原理,以太网交换机只能根据 MAC 帧的目的 MAC 地址和转发表来转发 MAC 帧,这就意味着:①不能在以太网上直接传输 IP 分组,必须将 IP 分组封装成 MAC 帧;②在将 IP 分组封装成 MAC 帧前,必须先获取连接在同一个网络上的源终端和目的终端的 MAC 地址。源终端的 MAC 地址可以直接从安装的网卡中读取,问题是如何根据目的终端的 IP 地址来获取目的终端的 MAC 地址。地址解析协议(Address Resolution Protocol,ARP)和地址解析过程就用于实现这一功能,ARP 请求帧格式如图 5.28 所示。

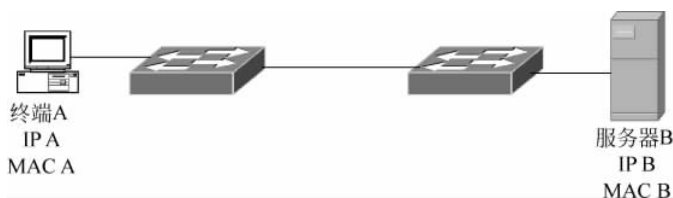


图 5.27 以太网内传送 IP 分组的过程

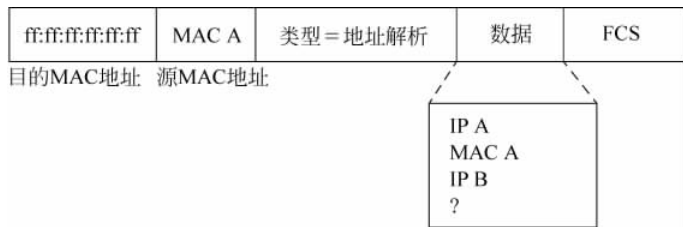


图 5.28 ARP 请求帧格式

图 5.29 中,终端 A 获知了服务器 B 的 IP 地址 IP B 后,广播一个 MAC 帧,该 MAC 帧的格式如图 5.28 所示,它的源 MAC 地址为终端 A 的 MAC 地址 MAC A,目的 MAC 地址为广播地址 ff: ff: ff: ff: ff: ff,MAC 帧中的数据字段包含终端 A 的 IP 地址 IP A 和 MAC 地址 MAC A,同时,包含服务器 B 的 IP 地址 IP B,IP B 是需要解析的 IP 地址,称为目标地址。该帧是 ARP 请求帧,要求 IP 地址为 IP B 的终端回复它的 MAC 地址。

由于该 MAC 帧的目的地址为广播地址,同一网络内的所有终端都能够接收到该 MAC 帧,每一个接收到该 MAC 帧的终端首先检测自己的 ARP 缓冲区,如果 ARP 缓冲区中没有发送终端的 IP 地址和 MAC 地址对,将发送终端的 IP 地址和 MAC 地址对(IP A 和 MAC A)记录在 ARP 缓冲区中,然后比较 MAC 帧中给出的目标 IP 地址是否和自己的 IP 地址相同,如果相同,回复自己的 MAC 地址。整个过程如图 5.29 所示。

ARP 地址解析过程只能发生在连接在同一个以太网上的源终端和目的终端之间,如果源终端和目的终端不在同一个网络内,则 IP 分组需要逐跳转发,源终端必须先将 IP 分组发送给由默认网关地址指定的第一跳路由器,当然,如果连接源终端和第一跳路由器的网络是以太网,源终端通过 ARP 地址解析过程获取第一跳路由器连接以太网的接口的 MAC 地址。同样,如果连接第一跳和下一跳路由器的网络也是以太网,如图 5.30 所示,第一跳路由器也需通过 ARP 地址解析过程获取下一跳路由器连接以太网的接口的 MAC 地址。总之,如果互连当前跳和下一跳的网络是以太网,IP 分组封装成 MAC 帧后才能经过以太网实现当前跳至下一跳的传输过程。在将 IP 分组封装成 MAC 帧前,必须获取下一跳连接以太网的接口的 MAC 地址,ARP 地址解析过程用于完成根据下一跳连接以太网的接口的 IP 地址求出该接口的 MAC 地址的过程。

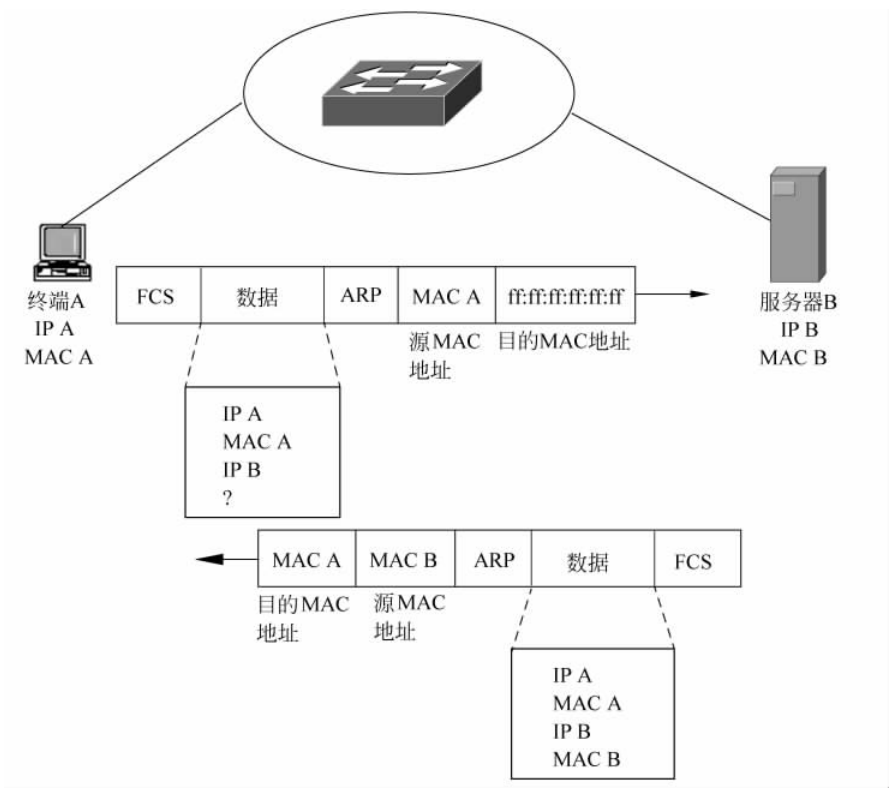


图 5.29 ARP 解析地址过程

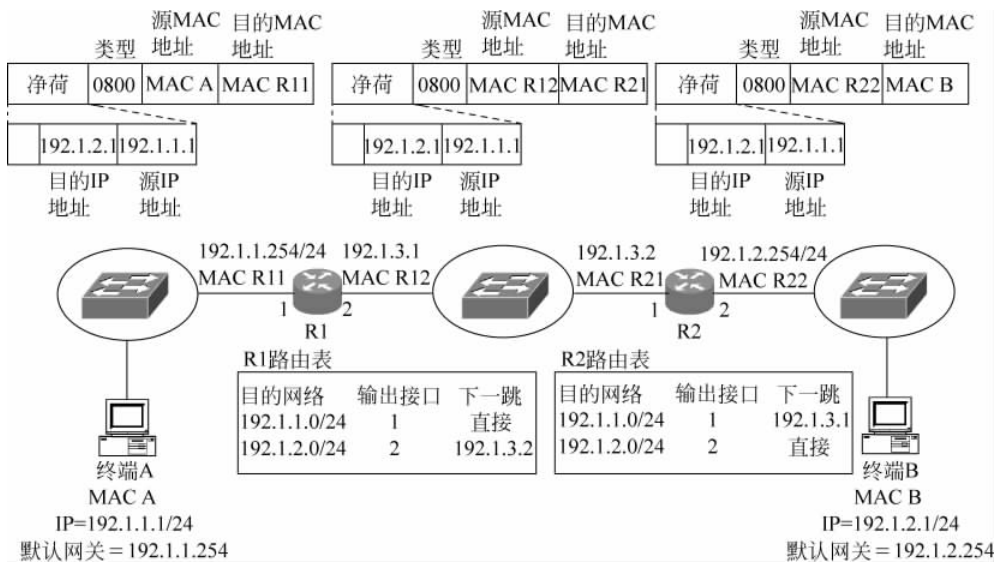


图 5.30 由多个以太网互连而成的互联网

5.4.2 逐跳封装

图 5.30 给出了终端 A 传输给终端 B 的 IP 分组经过各个以太网时的封装过程,IP 分组在终端 A 至终端 B 的传输过程中是不变的,但 IP 分组经过互连终端 A 和路由器 R1 的以太网时要封装成以路由器 R1 接口 1 的 MAC 地址 MAC R11 为目的 MAC 地址、以终端 A 的 MAC 地址 MAC A 为源 MAC 地址的 MAC 帧,类型字段 0800 表示净荷是 IP 分组。终端 A 通过解析默认网关地址 192.1.1.254 获得路由器 R1 接口 1 的 MAC 地址。IP 分组经过互连路由器 R1 和路由器 R2 的以太网时,封装成以路由器 R2 接口 1 的 MAC 地址 MAC R21 为目的 MAC 地址、以路由器 R1 接口 2 的 MAC 地址 MAC R12 为源 MAC 地址的 MAC 帧。路由器 R1 通过检索路由表获取路由器 R2 接口 1 的 IP 地址 192.1.3.2,通过解析 IP 地址 192.1.3.2 获得路由器 R2 接口 1 的 MAC 地址。IP 分组经过互连路由器 R2 和终端 B 的以太网时,封装成以终端 B 的 MAC 地址 MAC B 为目的 MAC 地址、以路由器 R2 接口 2 的 MAC 地址 MAC R22 为源 MAC 地址的 MAC 帧。路由器 R2 通过检索路由表得知终端 B 直接连接在接口 2 连接的以太网上,因此通过解析终端 B 的 IP 地址 192.1.2.1 获得终端 B 的 MAC 地址。

5.5 虚拟路由器冗余协议

默认网关是终端通往其他网络的传输路径上的第一跳路由器,每一个终端通常只能配置单个默认网关地址。容错网络结构使得可以在只为终端配置单个默认网关地址的前提下,为终端配置多个默认网关,且其中一个默认网关失效不会影响该终端与连接在其他网络上的终端之间的通信过程。虚拟路由器冗余协议(Virtual Router Redundancy Protocol, VRRP)就是这样一种实现具有上述功能的容错网络的协议。

5.5.1 容错网络结构

在如图 5.31 所示的网络结构中,每一个以太网内部通过链路冗余和生成树协议保证在发生单条链路故障的情况下仍然保持连接在同一以太网上的终端之间的连通性。同时,路由器 R1 和 R2 分别有接口连接到两个以太网,保证在其中一个路由器发生故障的情况下仍然保持连接在不同以太网上的终端之间的连通性。因此,如图 5.31 所示的互联网结构是一种不会因为单点故障导致网络连通性发生问题的容错网络结构。

由于每一个以太网同时连接两个路由器接口,因此,连接在每一个以太网上的终端可以在分配给这两个路由器接口的两个 IP 地址中选择一个作为默认网关地址。例如终端 A 可以选择 192.1.1.254 或 192.1.1.253 作为默认网关地址,但由于目前终端一般只能配置一个默认网关地址,因此,即使对于如图 5.31 所示的容错网络结构,终端在只能配置单个默认网关地址的情况下,一旦默认网关地址指定的路由器失效,必须通过手工配置新的默认网关地址来保持该终端和连接在其他网络上的终端之间的连通性。如果终端 A 配置了默认网关地址 192.1.1.254,一旦路由器 R1 失效,必须通过手工配置方式为终端 A 配置新的默认网关地址 192.1.1.253,否则,终端 A 无法和连接在其他网络上的终端通信。

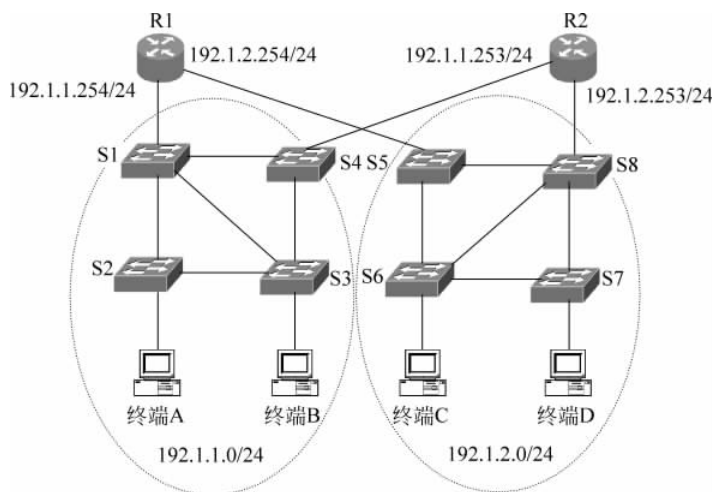


图 5.31 容错网络结构

对于如图 5.31 所示的容错网络结构,希望有一种和生成树协议相似的协议,该协议能够根据优先级在多个可以作为默认网关的路由器中选择一个路由器作为其默认网关,一旦该路由器发生故障,就能够自动选择另一个路由器作为默认网关,并自动完成两个路由器之间的功能切换。虚拟路由器冗余协议(VRRP)就是这样一种协议。

5.5.2 VRRP 工作原理

1. VRRP 工作环境

VRRP 工作环境如图 5.32 所示,支持 VRRP 的路由器称为 VRRP 路由器,多个由接口连接在同一个网络上的 VRRP 路由器(如图 5.32 中路由器 R1 和 R2)构成一个虚拟路由器,这些 VRRP 路由器中只有一个 VRRP 路由器是主路由器,其他路由器作为备份路由器。VRRP 作用的网络可以是任意支持广播的网络,如以太网、令牌环网和 FDDI,连接在这些网络上的终端和路由器接口有唯一的 MAC 地址。这里以以太网为例来讨论 VRRP 的工作原理。

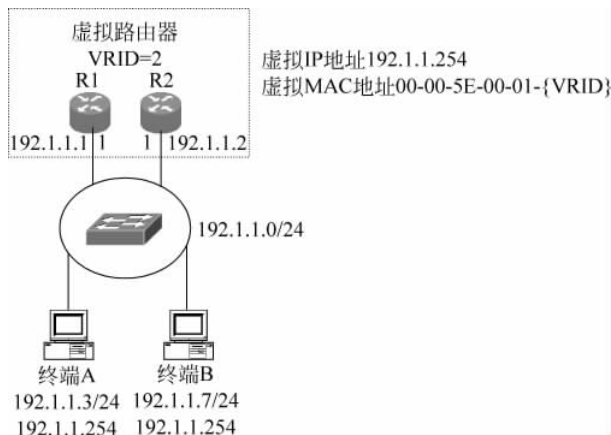


图 5.32 VRRP 工作环境

每一个 VRRP 路由器连接以太网的接口可以分配多个 IP 地址,从这些 IP 地址中选择一个 IP 地址作为接口的基本 IP 地址,接口发送的 VRRP 报文以接口的基本 IP 地址作为 IP 分组的源 IP 地址。可以对虚拟路由器配置多个 IP 地址,这些 IP 地址称为虚拟 IP 地址,虚拟 IP 地址可以与为 VRRP 路由器接口配置的 IP 地址相同,如果某个 VRRP 路由器为某个接口配置的 IP 地址与为该接口所属的虚拟路由器配置的虚拟 IP 地址相同,该路由器称为 IP 地址拥有者。每一个虚拟路由器分配唯一的 8 位二进制数的虚拟路由器标识符 (Virtual Router Identifier,VRID),属于同一个虚拟路由器的多个 VRRP 路由器有着相同的虚拟路由器标识符。虚拟路由器对外有唯一的 MAC 地址 00-00-5E-00-01-{VRID},对于 VRID 为 2 的虚拟路由器,虚拟 MAC 地址为 00-00-5E-00-01-02。终端配置的默认网关地址必须是虚拟 IP 地址,对虚拟 IP 地址进行地址解析得到的结果必须是虚拟 MAC 地址,以虚拟 MAC 地址为目的 MAC 地址的 MAC 帧一定能够到达主路由器,只有主路由器转发封装在以虚拟 MAC 地址为目的 MAC 地址的 MAC 帧中的 IP 分组。

VRRP 需要解决的问题主要有以下 3 个:

- 在属于同一个虚拟路由器的多个 VRRP 路由器中产生主路由器。
- 一旦接收到终端发送的请求解析虚拟 IP 地址的 ARP 请求报文,虚拟路由器将虚拟 MAC 地址作为与虚拟 IP 地址绑定的 MAC 地址回送给终端。
- 以太网(严格地讲是所有支持广播的局域网)一定能够将以虚拟 MAC 地址为目的 MAC 地址的 MAC 帧送达主路由器。

2. 路由器初始配置

对于如图 5.32 所示的 VRRP 工作环境,路由器 R1 和 R2 需要完成以下基本配置:

- 分别在路由器 R1 和 R2 创建 VRID 为 2 的虚拟路由器,分别将路由器 R1 和 R2 的接口 1 配置给 VRID 为 2 的虚拟路由器,使得路由器 R1 和 R2 成为 VRID 为 2 的虚拟路由器的 VRRP 路由器。
- 分别为路由器 R1 和 R2 的接口 1 分配 IP 地址 192.1.1.1/24 和 192.1.1.2/24,这两个接口的 IP 地址必须与它们所连接的以太网的网络地址 192.1.1.0/24 一致。由于路由器 R1 和 R2 的接口 1 只分配了一个 IP 地址,该 IP 地址作为接口的基本 IP 地址。
- 为路由器 R1 和 R2 的接口 1 分配优先级,优先级的范围为 1~254。主路由器用优先级 0 表示愿意主动放弃主路由器地位,IP 地址拥有者的优先级为 255。优先级值越大,VRRP 路由器在竞争主路由器时的优先级越高。
- 为 VRID 为 2 的虚拟路由器分配虚拟 IP 地址 192.1.1.254。该 IP 地址成为连接在网络 192.1.1.0/24 上的终端的默认网关地址。
- 虚拟路由器根据 VRID=2 生成虚拟 MAC 地址 00-00-5E-00-01-02。

3. VRRP 报文格式

VRRP 报文封装成 IP 分组的格式如图 5.33 所示,不直接将 VRRP 报文封装成 MAC 帧格式的主要原因是,VRRP 作用的网络可以是支持广播的任意网络,不一定是以太网。IP 分组的源 IP 地址是发送 VRRP 报文的接口的基本 IP 地址,对于路由器 R1 接口 1 发送

的 VRRP 报文,源 IP 地址为 192.1.1.1。目的 IP 地址是多播地址 224.0.0.18。所有 VRRP 路由器将以该多播地址为目的地址的 IP 分组提交给 VRRP 实体。VRRP 报文对应的协议字段值是 112。VRRP 报文中给出发送该 VRRP 报文的接口所属的虚拟路由器的 VRID、该接口的优先级、分配给虚拟路由器的虚拟 IP 地址等。VRRP 只有一种类型报文——通告报文。



图 5.33 VRRP 报文格式

如果 VRRP 作用的网络是以太网,如图 5.33 所示的 IP 分组将封装成 MAC 帧,该 MAC 帧的源 MAC 地址是发送接口所属虚拟路由器对应的虚拟 MAC 地址,对于路由器 R1 接口 1,源 MAC 地址是 00-00-5E-00-01-02,目的 MAC 地址是多播地址 224.0.0.18 对应的 MAC 组地址。根据多播地址 224.0.0.18 求出对应的 MAC 组地址的过程如图 5.34 所示。

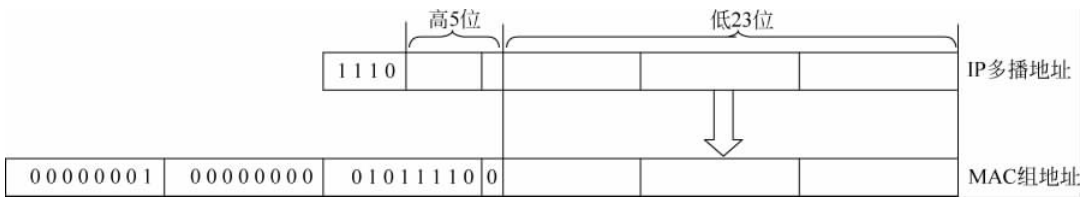


图 5.34 IP 多播地址映射到 MAC 组地址过程

从图 5.34 中可以看出,映射后的 MAC 组地址的高 25 位固定为 00000001、00000000、01011110 和 0,低 23 位等于 IP 多播地址的低 23 位。因此,多播地址 224.0.0.18 对应的 MAC 组地址为 01-00-5E-00-00-12。由于 IP 多播地址中用于标识多播组的地址有 28 位,因此,标识多播组的 IP 多播地址中的高 5 位在映射过程中没有使用,这就使得 IP 多播地址和 MAC 组地址之间的映射不是唯一的,32 个不同的 IP 多播地址有可能映射为同一个 MAC 组地址。

4. 主路由器产生过程

路由器状态转换过程如图 5.35 所示,每一个 VRRP 路由器启动后,处于初始化状态,如果该 VRRP 路由器是 IP 地址拥有者,该 VRRP 路由器立即成为主路由器,并立即发送如图 5.33 所示的 VRRP 报文,然后周期性地发送 VRRP 报文。如果某个 VRRP 路由器不是 IP 地址拥有者,该 VRRP 路由器立即成为备份路由

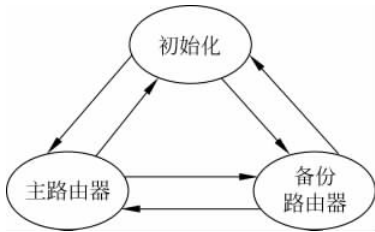


图 5.35 路由器状态转换过程

器,启动 Master_Down_Timer,等待接收主路由器发送的 VRRP 报文。

任何路由器接收到 VRRP 报文后,都依序进行下列检查:

- 判别接收该 VRRP 报文的接口是否属于 VRRP 报文中 VRID 指定的虚拟路由器。
- 根据 VRRP 报文中的 VRID 确定虚拟路由器,判别路由器为该虚拟路由器配置的虚拟 IP 地址是否与 VRRP 报文中给出的虚拟 IP 地址相同。

上述检查中只要有一项不匹配,路由器就丢弃该 VRRP 报文。

如果主路由器接收到 VRRP 报文,而且 VRRP 报文中的优先级大于主路由器为接收该 VRRP 报文的接口配置的优先级,或者虽然 VRRP 报文中的优先级等于主路由器为接收该 VRRP 报文的接口配置的优先级,但 VRRP 报文的源 IP 地址大于主路由器接收该 VRRP 报文的接口的基本 IP 地址,该主路由器立即转换为备份路由器,停止发送 VRRP 报文,启动 Master_Down_Timer,等待新的主路由器发送的 VRRP 报文。

备份路由器接收到主路由器发送的 VRRP 报文后,根据备份路由器的工作方式对 VRRP 报文进行处理,如果备份路由器配置为允许抢占方式,且发现 VRRP 报文中的优先级小于备份路由器为接收该 VRRP 报文的接口配置的优先级,备份路由器立即转换为主路由器,并立即发送 VRRP 报文,然后周期性地发送 VRRP 报文。如果备份路由器配置为不允许抢占方式,或者发现 VRRP 报文中的优先级大于或等于备份路由器为接收该 VRRP 报文的接口配置的优先级,则刷新 Master_Down_Timer。

如果某个备份路由器的 Master_Down_Timer 溢出,表示主路由器已经失效,该备份路由器立即转换为主路由器,并立即发送 VRRP 报文,然后周期性地发送 VRRP 报文。有可能因为网络拥塞导致主路由器发送的 VRRP 报文不能及时到达备份路由器,因而使备份路由器误认为主路由器失效而重新开始主路由器选择过程。为了避免发生这种情况,Master_Down_Timer 溢出时间大于 3 倍的主路由器 VRRP 报文发送间隔。

5. 主路由器和备份路由器功能

主路由器功能如下:

- 必须对请求解析虚拟 IP 地址的 ARP 请求报文做出响应。
- 必须对封装在以虚拟 MAC 地址为目的 MAC 地址的 MAC 帧中的 IP 分组进行转发操作。
- 在成为主路由器时,立即发送将所有虚拟 IP 地址绑定到虚拟 MAC 地址的 ARP 报文,使得网络内的所有终端将默认网关地址与虚拟 MAC 地址绑定在一起。

备份路由器功能如下:

- 不对请求解析虚拟 IP 地址的 ARP 请求报文做出响应。
- 丢弃接收到的以虚拟 MAC 地址为目的地址的 MAC 帧。
- 丢弃接收到的以虚拟 IP 地址为目的地址的 IP 分组。

6. 虚拟 IP 地址解析过程

如果终端在 ARP 缓冲区中找不到与默认网关地址绑定的 MAC 地址,会发送一个请求解析该默认网关地址的 ARP 请求报文,该 ARP 请求报文在终端所连接的网络中广播,连接在该网络上的所有 VRRP 路由器都接收到该 ARP 请求报文,但只有主路由器对该 ARP

请求报文做出响应,并在 ARP 响应报文中将虚拟 MAC 地址与默认网关地址绑定在一起。终端发送给默认网关的 IP 分组封装成以终端 MAC 地址为源 MAC 地址、以虚拟 MAC 地址为目的 MAC 地址的 MAC 帧,只有主路由器对封装在这样 MAC 帧中的 IP 分组进行转发操作,其他 VRRP 路由器即使接收到该 MAC 帧,也将丢弃该 MAC 帧。

7. 交换机转发表更新过程

如果将图 5.32 中的以太网扩展为如图 5.36 所示的以太网结构,在路由器 R2 成为主路由器后,以太网中各个交换机的转发表需要生成如表 5.19 所示的转发项,否则,可能导致发生终端发送给默认网关的 MAC 帧在以太网中广播的情况。为了在各个交换机中生成如表 5.19 所示的转发项,当路由器 R2 成为主路由器时,立即发送一个 VRRP 报文,该 VRRP 报文最终被封装成以虚拟 MAC 地址 00-00-5E-00-01-02 为源 MAC 地址、以组地址 01-00-5E-00-00-12 为目的 MAC 地址的 MAC 帧,该 MAC 帧在以太网中广播,如图 5.36 所示,以太网中所有交换机都接收到该 MAC 帧,通过地址学习,在转发表中建立如表 5.19 所示的转发项。路由器 R2 定期发送的 VRRP 报文定期刷新各个交换机中虚拟 MAC 地址对应的转发项,使得各个交换机将一直在转发表中维持该转发项。

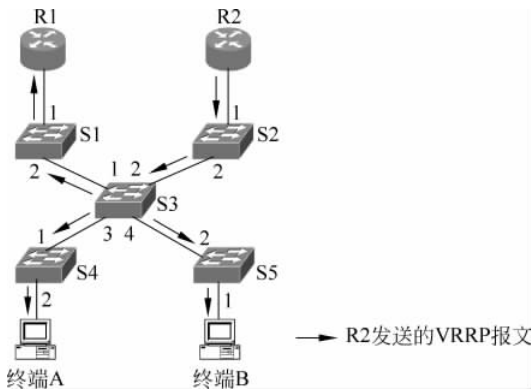


图 5.36 以太网结构

表 5.19 交换机转发表

交 换 机	MAC 地址	转 发 端 口
交换机 S1	00-00-5E-00-01-02	端口 2
交换机 S2	00-00-5E-00-01-02	端口 1
交换机 S3	00-00-5E-00-01-02	端口 2
交换机 S4	00-00-5E-00-01-02	端口 1
交换机 S5	00-00-5E-00-01-02	端口 2

8. 负载均衡

如图 5.32 所示的 VRRP 工作环境能够解决容错问题,但无法实现负载均衡。为了实现负载均衡,采用如图 5.37 所示的 VRRP 工作环境。创建两个 VRID 分别为 2 和 3 的虚拟路由器,同时将路由器 R1 和 R2 连接以太网的接口分配给两个虚拟路由器,为 VRID 为

2 的虚拟路由器分配虚拟 IP 地址 192.1.1.1,使得路由器 R1 因为是 IP 地址拥有者而自然成为 VRID 为 2 的虚拟路由器中的主路由器。为 VRID 为 3 的虚拟路由器分配虚拟 IP 地址 192.1.1.2,使得路由器 R2 因为是 IP 地址拥有者而自然成为 VRID 为 3 的虚拟路由器中的主路由器。将一半连接在网络 192.1.1.0/24 上的终端(图 5.37 中的终端 A)的默认网关地址配置成 VRID 为 2 的虚拟路由器对应的虚拟 IP 地址 192.1.1.1,将另一半连接在网络 192.1.1.0/24 上的终端(图 5.37 中的终端 B)的默认网关地址配置成 VRID 为 3 的虚拟路由器对应的虚拟 IP 地址 192.1.1.2,这样,连接在网络 192.1.1.0/24 上的终端,一半将路由器 R1 作为默认网关,另一半将路由器 R2 作为默认网关,一旦某个路由器发生故障,另一个路由器将自动作为所有终端的默认网关,既实现了容错,又实现了负载均衡。

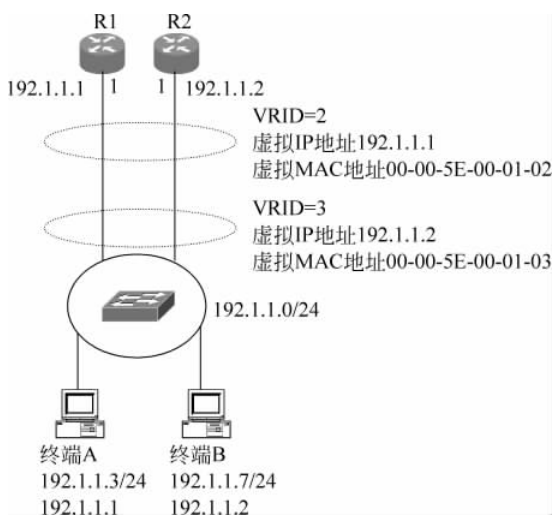


图 5.37 均衡负载的 VRRP 工作环境

5.5.3 VRRP 应用实例

1. 互联网结构与基本配置

如图 5.38 所示的网络结构是如图 5.31 所示的网络结构简化版,为了实现容错和负载均衡,对互联网进行如下配置:

- 根据如图 5.38 所示的配置信息分别为路由器 R1 和 R2 的两个接口配置 IP 地址和子网掩码,完成路由器接口 IP 地址和子网掩码配置后,路由器 R1 和 R2 自动生成如图 5.38 所示的路由表,路由表中给出用于指明通往路由器直接连接的网络的传输路径的路由项。
- 创建 VRID 分别为 2 和 3 的两个虚拟路由器,将路由器 R1 接口 1 和路由器 R2 接口 1 分配给 VRID 为 2 的虚拟路由器,并将路由器 R1 接口 2 和路由器 R2 接口 2 分配给 VRID 为 3 的虚拟路由器,VRID 为 2 的虚拟路由器对应的虚拟 MAC 地址为 00-00-5E-00-01-02,VRID 为 3 的虚拟路由器对应的虚拟 MAC 地址为 00-00-5E-00-01-03。
- 为 VRID 为 2 的虚拟路由器分配虚拟 IP 地址 192.1.1.254,使得路由器 R1 成为

VRID 为 2 的虚拟路由器的主路由器。为 VRID 为 3 的虚拟路由器分配虚拟 IP 地址 192.1.2.253,使得路由器 R2 成为 VRID 为 3 的虚拟路由器的主路由器。

- 连接在网络 192.1.1.0/24 上的终端配置默认网关地址 192.1.1.254,连接在网络 192.1.2.0/24 上的终端配置默认网关地址 192.1.2.253。

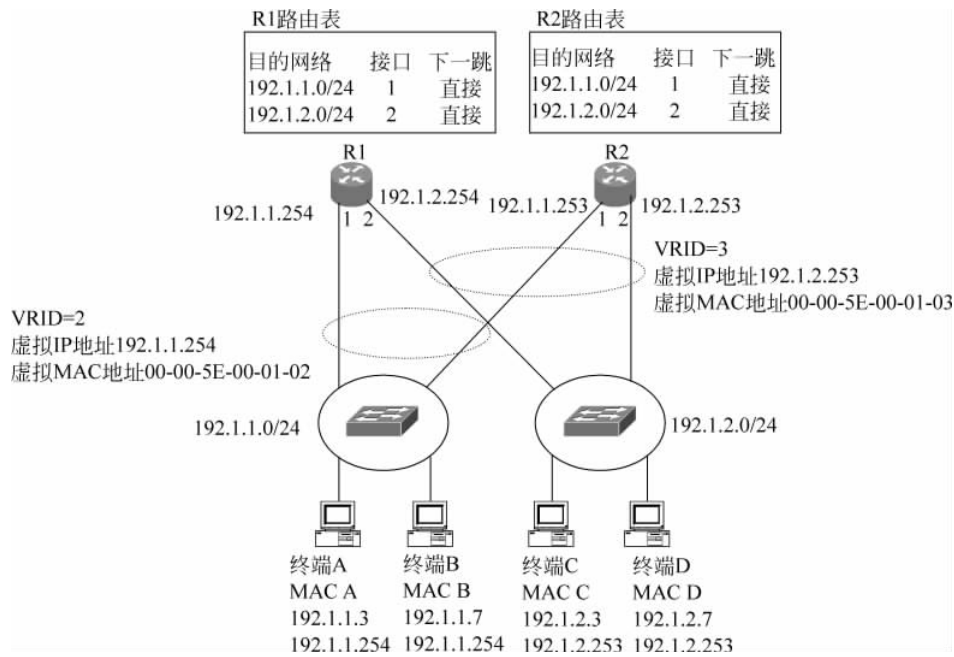


图 5.38 互联网结构与基本配置

2. 生成主路由器和转发项

路由器 R1 因为是虚拟 IP 地址 192.1.1.254 的 IP 地址拥有者,自然成为 VRID 为 2 的虚拟路由器的主路由器。R1 在成为主路由器后,通过发送 VRRP 报文,在网络 192.1.1.0/24 的各个交换机中建立将目的 MAC 地址为 00-00-5E-00-01-02 的 MAC 帧转发给路由器 R1 接口 1 的转发项;同时,通过在网络 192.1.1.0/24 中广播将虚拟 IP 地址 192.1.1.254 与虚拟 MAC 地址 00-00-5E-00-01-02 绑定的 ARP 报文,在连接在网络 192.1.1.0/24 上的所有终端的 ARP 缓冲区中建立 IP 地址 192.1.1.254 与 MAC 地址 00-00-5E-00-01-02 之间的绑定。基于同样的原因,在路由器 R2 成为 VRID 为 3 的虚拟路由器的主路由器后,在网络 192.1.2.0/24 的各个交换机中建立将目的 MAC 地址为 00-00-5E-00-01-03 的 MAC 帧转发给路由器 R2 接口 2 的转发项,并在连接在网络 192.1.2.0/24 上的所有终端的 ARP 缓冲区中建立 IP 地址 192.1.2.253 与 MAC 地址 00-00-5E-00-01-03 之间的绑定。

3. IP 分组传输过程

如果终端 A 需要向终端 D 发送 IP 分组,首先获取终端 D 的 IP 地址 192.1.2.7,构建源 IP 地址为 192.1.1.3、目的 IP 地址为 192.1.2.7 的 IP 分组。通过判别终端 A 和终端 D 所在网络的网络地址(192.1.1.0/24 和 192.1.2.0/24)发现终端 A 和终端 D 不在同一个网

络,终端 A 需要将 IP 分组发送给默认网关。终端 A 从 ARP 缓冲区中获取与默认网关地址 192.1.1.254 绑定的 MAC 地址 00-00-5E-00-01-02,构建以终端 A 的 MAC 地址 MAC A 为源 MAC 地址、以 MAC 地址 00-00-5E-00-01-02 为目的 MAC 地址的 MAC 帧,网络 192.1.1.0/24 保证将该 MAC 帧转发给路由器 R1。路由器 R1 由于是 VRID 为 2 的虚拟路由器的主路由器,必须对封装在以虚拟 MAC 地址 00-00-5E-00-01-02 为目的 MAC 地址的 MAC 帧中的 IP 分组进行转发操作。路由器 R1 从 MAC 帧中分离出 IP 分组,用 IP 分组的目的 IP 地址 192.1.2.7 匹配 R1 路由表中的路由项,发现和路由项 <192.1.2.0/24,2,直接>匹配,下一跳为“直接”,表明目的终端连接在接口 2 连接的网络上。通过 ARP 地址解析过程获取与目的 IP 地址 192.1.2.7 绑定的 MAC 地址 MAC D,构建以接口 2 所属的虚拟路由器对应的虚拟 MAC 地址 00-00-5E-00-01-03 为源 MAC 地址、以终端 D 的 MAC 地址 MAC D 为目的 MAC 地址的 MAC 帧,通过网络 192.1.2.0/24 将该 MAC 帧转发给终端 D,终端 D 从 MAC 帧中分离出 IP 分组,完成 IP 分组终端 A 至终端 D 的传输过程。

当终端 D 向终端 A 发送 IP 分组时,先将 IP 分组转发给默认网关——路由器 R2,以实现路由器 R1 和 R2 的负载均衡。当其中一个路由器发生故障时,另一个路由器将作为连接在两个网络上的终端的默认网关。

本章小结

- 网络互连需要实现连接在不同类型的传输网络上的两个终端之间的通信过程。
- IP 地址是独立于任何传输网络的统一的地址格式。
- IP 分组是独立于任何传输网络的统一的分组格式。
- IP 将互联网简化为 IP 分组交换网络,IP over X 和 X 传输网络实现由 X 传输网络互连的两个结点之间的 IP 分组传输过程。
- 路由器是实现网络互连的专业设备,它的核心功能如下:一是建立路由表,实现 IP 分组转发;二是针对连接的多个不同类型的传输网络,实现 IP over X 和 X 传输网络对应的物理层和链路层功能,X 分别指该路由器连接的多个不同类型的传输网络。
- 路由项分为直连路由项、静态路由项和动态路由项。
- IP over 以太网和以太网在确定下一跳 IP 地址和互连当前跳与下一跳的网络是以太网的前提下,完成 IP 分组当前跳至下一跳的传输过程。
- ARP 用于完成根据下一跳的 IP 地址获取下一跳的 MAC 地址的过程。
- VRRP 使得可以在只为终端配置单个默认网关地址的前提下,为终端配置多个默认网关,且其中一个默认网关失效不会影响该终端与连接在其他网络上的终端之间的通信过程。

习题

5.1 为什么说 IP 是一种网际协议? IP 实现连接在不同传输网络上的终端之间通信的技术基础是什么?

- 5.2 为什么要为每一个路由器接口分配 IP 地址?
- 5.3 作为中继系统,集线器、网桥和路由器有何区别?
- 5.4 解释不能用网桥实现两个分别连接在以太网和 ATM 网络的终端之间通信的原因。
- 5.5 解释路由器和网桥的主要区别。
- 5.6 何为默认网关? 终端配置默认网关地址的作用是什么?
- 5.7 路由器实现不同类型的传输网络互连的技术基础是什么?
- 5.8 路由器主要由几部分组成? 如何实现 IP 分组的转发过程?
- 5.9 IP 地址分为几类? 各如何表示? 它们的主要特点是什么?
- 5.10 简述 IP 地址和 MAC 地址的不同及各自的作用。
- 5.11 为什么需要无分类编址? 它对路由项聚合和子网划分带来什么好处?
- 5.12 什么是最长前缀匹配算法? 在什么条件下需要使用最长前缀匹配算法?
- 5.13 子网掩码 255.255.255.0 代表什么意思? 如果某一网络的子网掩码为 255.255.255.248,该网络能够连接多少主机?
- 5.14 以下地址中的哪一个地址和网络前缀 85.32/12 匹配,说明理由。
A. 85.33.224.123 B. 85.79.65.216
C. 85.58.119.74 D. 85.68.205.154
- 5.15 以下网络前缀中的哪一个和地址 2.52.90.140 匹配,说明理由。
A. 0/4 B. 32/4 C. 4/6 D. 80/4
- 5.16 请辨认以下 IP 地址的网络类型。
(1) 128.35.199.3
(2) 21.12.240.17
(3) 183.194.75.253
(4) 192.12.69.248
(5) 89.3.0.1
(6) 200.3.5.2
- 5.17 一个 3200b 的 TCP 报文传到 IP 层,加上 160b 的首部后成为 IP 分组,下面的互联网由两个局域网通过路由器连接起来,但第二个局域网的 MTU=1200b,因此,IP 分组必须在路由器进行分片。第二个局域网实际需要为上层传输多少比特的数据?
- 5.18 假定传输层将包含 20B 首部和 2048B 数据的 TCP 报文递交给 IP 层,源终端至目的终端传输路径需要经过两个网络,其中第一个网络的 MTU=1024B,第二个网络的 MTU=512B,IP 首部是 20B。给出到达目的终端时分片后的 IP 分组序列,并计算出每一片的净荷字节数和片偏移。
- 5.19 路径 MTU 是端到端传输路径所经过的网络中最小的 MTU,假定源终端能够发现路径 MTU,并以路径 MTU 作为源终端封装 IP 分组的依据。根据题 5.18 的参数,给出到达目的终端时分片后的 IP 分组序列,并计算出每一片的净荷字节数和片偏移。
- 5.20 有人说“ARP 向网络层提供了转换地址的服务,应该属于数据链路层”。为什么这种说法是错误的?
- 5.21 ARP 缓冲器中每一项的寿命是 10~15min,简述寿命太长或者太短可能出现的

问题。

5.22 如果重新设计 IP 地址时,将 IP 地址设计为 48 位,能否通过 IP 地址和 MAC 地址之间的一一对应关系消除 ARP 地址解析过程?

5.23 设某路由器建立了如下路由表(这 3 列分别是目的网络、子网掩码和下一跳路由器,若直接交付,则最后一列给出输出接口)。

128.96.39.0	255.255.255.128	接口 0
128.96.39.128	255.255.255.128	接口 1
128.96.40.0	255.255.255.128	R2
192.4.153.0	255.255.255.192	R3
默认		R4

现收到 5 个 IP 分组,其目的 IP 地址如下:

- (1) 128.96.39.10
- (2) 128.96.40.12
- (3) 128.96.40.151
- (4) 192.4.153.17
- (5) 192.4.153.90

试分别计算出下一跳路由器或输出接口。

5.24 某单位分配到一个 B 类 IP 地址,其网络地址为 124.250.0.0,该单位有 4000 多台机器,分布在 16 个不同的地点,如果选用的子网掩码为 255.255.255.0,试分别给每个地点分配一个网络地址,并根据网络地址计算出每个地点可分配的 IP 地址范围。

5.25 一个 IP 分组的数据长度为 4000B(固定长度首部),需要经过一个 MTU 为 1500B 的网络。应当划分为几个数据片? 每一个数据片的数据字段长度、片偏移字段和 MF 标志为何值?

5.26 IP 分组中的首部检验和只检验 IP 分组首部,这样做的好处是什么? 坏处是什么? IP 分组首部检错码为什么不采用 CRC?

5.27 一个自治系统有 5 个局域网,其连接如图 5.39 所示,LAN 2 至 LAN 5 上的主机数分别为 91、150、3 和 15,该自治系统分配到的 IP 地址块为 30.138.118.0/23。试给出每一个局域网的地址块(包括前缀)。

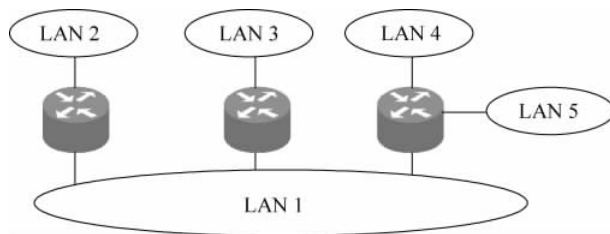


图 5.39 题 5.27 图

5.28 对如下 4 个地址块进行最大可能的聚合:

212.55.132.0/24

212.55.133.0/24

212.55.134.0/24

212.55.135.0/24

5.29 根据图 5.40 所示的网络地址配置,给出路由器 R1、R2 和 R3 的路由表。如果要求路由器 R2 中的路由项最少,如何调整网络地址配置? 根据调整后的网络地址配置,给出路由器 R1、R2 和 R3 的路由表。

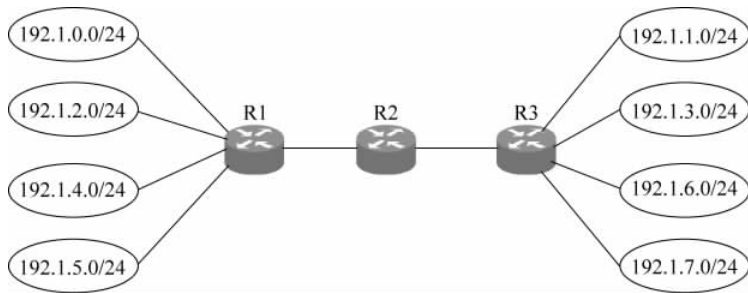


图 5.40 题 5.29 图

5.30 假定分配的 CIDR 地址块为 192.77.33.0/24, 根据图 5.41 所示的互连网络结构,为每一个局域网分配合适的网络前缀地址(图中每一个局域网旁边标明的数字是该局域网的主机数)。

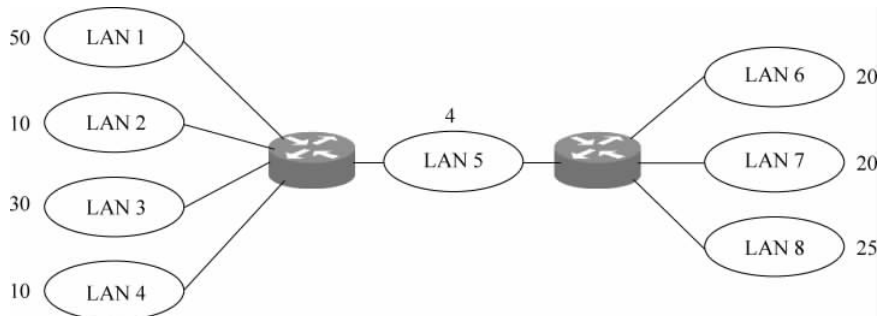


图 5.41 题 5.30 图

5.31 某单位分配到一个地址块 136.23.12.64/26, 现在需要进一步划分为 4 个一样大的子网。

- (1) 每个子网的网络前缀有多长?
- (2) 每个子网有多少地址?
- (3) 每个子网的地址块是什么?
- (4) 每个子网可分配给主机的最小和最大地址是什么?

5.32 网络结构如图 5.42 所示,给出的 CIDR 地址块是 192.1.1.64/26。确定每一个子网的网络地址,将最大可用地址分配给路由器连接对应子网的接口,给出路由器 R1、R2 的路由表。

5.33 互连网络结构如图 5.43 所示。

(1) 补齐图中终端和路由器的配置信息,包括路由表,使其能够实现终端 A 和终端 B 之间的 IP 分组传输。

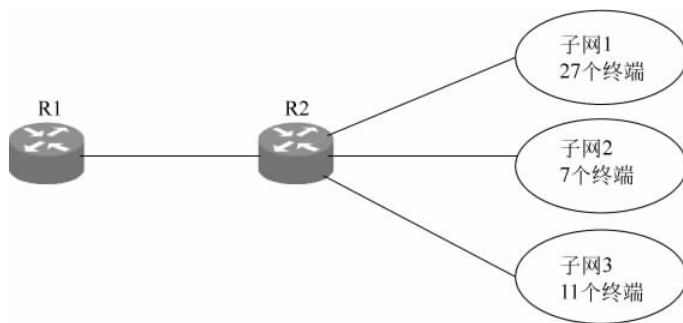


图 5.42 题 5.32 图

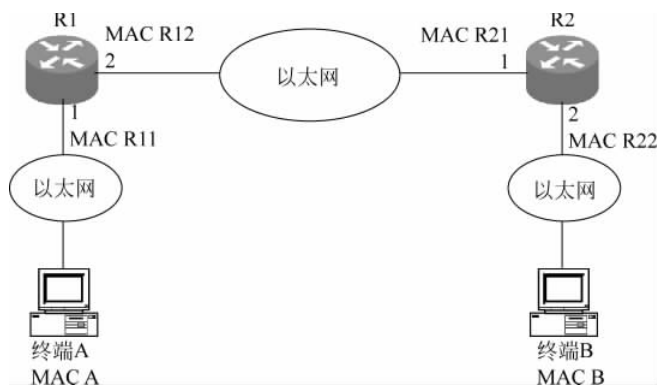


图 5.43 题 5.33 图

(2) 以(1)补齐的配置信息为基础,给出终端 A 至终端 B 的 IP 分组传输过程中涉及的所有 MAC 帧,并给出这些 MAC 帧的源和目的 MAC 地址(假定终端和路由器的 ARP 缓冲器为空)。

5.34 VRRP 的作用是什么?

5.35 简述主路由器转换为备份路由器的条件。

5.36 简述备份路由器转换为主路由器的条件。

5.37 对于如图 5.38 所示的网络结构,如果要求连接在网络 192.1.1.0/24 和 192.1.2.0/24 上的终端各有一半以路由器 R1、R2 为默认网关,给出实现这一功能所需的 VRRP 配置。