第 7 章　Internet 协议

网络层

本章要点：

1. 网络层协议
2. 传输层协议
3. 应用层协议
4. NAT
5. 浏览器
6. 搜索引擎
7. QoS 实现

Internet 主要协议

7.1.1 IPv4

IP 数据报、IP 包、IP 分组

节点：路由器

网际协议 IP 是 TCP/IP 体系中两个最主要的协议之一。与 IP 协议配套使用的还有四个协议：

地址解析协议 ARP

(Address Resolution Protocol)

逆地址解析协议 RARP

(Reverse Address Resolution Protocol)

网际控制报文协议 ICMP

(Internet Control Message Protocol)

网际组管理协议 IGMP

(Internet Group Management Protocol)

网际层的 IP 协议及配套协议

网络互连使用路由器

当中继系统是**转发器**或**网桥**时，一般并不称之为网络互连，这仅仅是把一个网络扩大了，而这仍然是一个网络。

网关由于比较复杂，目前使用得较少。

互联网都是指用路由器进行互连的网络。

由于历史的原因，许多有关 TCP/IP 的文献将网络层使用的路由器称为网关。

直接交付和间接交付

当主机 A 要向另一个主机 B 发送数据报时，先要检查目的主机 B 是否与源主机 A 连接在同一个网络上。

如果是，就将数据报直接交付给目的主机 B 而不需要通过路由器。

但如果目的主机与源主机 A 不是连接在同一个网络上，则应将数据报发送给本网络上的某个路由器，由该路由器按照转发表指出的路由将数据报转发给下一个路由器。这就叫作间接交付。

直接交付和间接交付

从网络层看 IP 数据报的传送

如果我们只从网络层考虑问题，那么 IP 数据报就可以想象是在网络层中传送。

7.1.1.1　分类的 IP 地址

IP 地址就是给每个连接在英特网上的**主机**(或路由器)分配一个在全世界范围是惟一的 32bit 的标识符。

IP 地址的编址方式分三个阶段：

(1)分类的 IP 地址。

(2)子网的划分。

(3)构成超网。即无分类编址方式。

1. IP 地址分类与表示

(1) IP 地址与表示

■ IP 地址：给每个主机（或路由器）分配的一个在全世界范围唯一的 32 位的标识符—逻辑地址。

■ IP 地址由因特网名字与号码分配公司 ICANN (Internet Corporation for Assigned Names and Numbers)进行分配点分十进制表示法：XX.XX.XX.XX（IPv4)

■ IP 地址表示一台主机与一个网络之间的连接，一台计算机与多个网络连接时，需要多个 IP 地址。

■ **两级的 IP 地址可以记为：**

■路由器转发分组的步骤

先按所要找的 IP 地址中的网络号 net-id 把**目的网络**找到。

当分组到达目的网络后，再利用主机号 host-id 将数据报直接交付给目的主机。

(2) IP 地址的分类：P201－黄

根据网络号可以把 IP 地址分为以下几种：各类地址范围

2. 常用三种类别的 IP 地址　P202－黄

A 类地址：

- A 类地址的前 8 位代表网络号，剩余的 24 位可由管理网络地址的管理用户来设置。
- 最大网络数 126(2^7-2)，网络号为 1~126，每个网络中最大主机数 16777214($2^{24}-2$)。
- A 类地址占有整个 IP 地址空间的 50%。

B 类地址：

- B 类地址的前 16 位代表网络号，剩余的 16 位可由管理网络地址的管理用户来设置。
- 最大网络数 16384(2^{14})，网络号为 128.0~191.255，每个网络中最大主机数 65534($2^{16}-2$)。
- B 类地址占有整个 IP 地址空间的 25%。

C 类地址：

- C 类地址的前 24 位代表网络号，剩余的 8 位可由管理网络地址的管理用户来设置。

- 最大网络数 2097152(2^{21})，网络号为 192.0.0~223.255.255，每个网络中最大主机数 254 (2^8-2)。
- C 类地址占有整个 IP 地址空间的 12.5%。

注意：网络号 net-id 和主机号 host-id 在全 1 和全 0 的含义。

各类地址范围

3. 特殊地址

4. IP 地址的分配

网络信息中心(NIC)统一负责全球地址的规划 。

全球性的网络信息中心有：

5. IP 地址的特点：

(1)从某种意义上来说，IP 地址是一种分等级的地址结构，但它不反映任何有关主机位置的地理信息。

(2) IP 地址是标志一个主机（或路由器）和一条链路的接口

- **一个主机同时连接到两个网络上时，必须同时具有两个相应的 IP 地址。这种主机叫作多接口主机。**
- **一个路由器至少连接到两个网络，因此一个路由器至少应当有两个不同的 IP 地址。**

(3)按照 Internet 的观点，用转发器或网桥连接起来的局域网仍为一个网络，因此这些局域网都具有同样的网络号。

- **若具有不同网络号的网络连接时，该使用什么互联设备呢？ ———路由器。**

(4)在 IP 地址中，所有分配到网络号的网络，不管是局域网还是广域网都是平等的。

(5)在 IP 地址中当主机号为全零时，可用来指明单个网络的地址。

如：10·0·0·0（A 类）

175·89·0·0（B 类）

201·123·56·0（C 类）

(6)当两个路由器直接相连时，在连线两端的接口处，可以指明也可以不指明 IP 地址。

互联网中的 IP 地址

互联网中的 IP 地址

互联网中的 IP 地址

互联网中的 IP 地址

6. IP 地址与硬件地址

硬件地址是数据链路层和物理层使用的地址，而 IP 地址是网络层和以上各层使用的地址。

IP 地址放在 IP 数据报首部，而硬件地址则放在 MAC 帧的首部。在网络层和网络层以上使用的是 IP 地址，而数据链路层及以下使用的是硬件地址。在数据链路层看不到 IP 地址。

IP 地址与硬件地址的区别：P181－182 谢(4)， P118－谢(5)

- **在 IP 数据报的首部既有源站的 IP 地址也有目的站的 IP 地址，但是在通信中路由器只根据目的站的 IP 地址进行路由选择。**
- **IP 数据报在通信过程中，首部的源站 IP 地址和目的站的 IP 地址在经过路由器时不会发生改变。**
- **一个路由器至少有两个 IP 地址和两个 MAC 地址。**
- **路由器的 IP 地址不会出现在 IP 数据报中。**
- **在具体的物理网络的数据链路层，看到的只是 MAC 帧。IP 数据报被封装在 MAC 帧中。MAC 帧在不同的网络上传送时，其 MAC 帧首部不同。**
- **MAC 帧(硬件地址)在经过路由器时，其首部中的源地址和目的地址会发生改变，路由器的 MAC 地址会出现在 MAC 帧中。**
- **IP 层抽象的互联网却屏蔽了下层的复杂的细节。**

问题：

(1)主机或路由器 MAC 帧的首部中硬件地址怎样得到？

(2)路由器中的路由表怎样得到？

7.1.1.2 IP 层转发分组的流程

1. 路由器和交换机相同点与不同点:

相似之处：

路由器和交换机都是采用查找表的方式，将分组转发到下一站。

区别：

- **路由器是用来连接不同的网络，交换机只是在一个特定的网络中工作。**
- **路由器是专门用来转发分组的，交换机还可以接上许多主机。**
- **路由器使用统一的 IP 协议，交换机使用所在广域网中特定协议。**
- **在查找路由表时，路由器根据目的站所在的网络号找出下一跳（下一个路由器），交换机则根据目的站所接入的交换机号找出下一跳（下一个交换机）。**

从表中可以看出，路由选择都是基于网络号，也就是说只用到了 IP 地址的网络号部分。

查找路由表

这样根据目的网络地址来确定下一跳路由器的方法结果：

(1) IP 数据报首先要设法找到目的主机所在目的网络上的路由器(间接交付) 。

(2) 只有到达最后一个路由器时，才试图向目的主机进行直接交付。

(3) 每个路由器都独立地进行路由选择，因此从主机 A 到主机 B 的路由选择与主机 B 到主机 A 的路由选择会不同。

(4) 可以使用默认路由以减少路由表所占用的空间和搜索路由表所用的时间。

默认路由(default route)

这种转发方式在一个网络只有很少的对外连接时是很有用的。

默认路由在主机发送 IP 数据报时往往更能显示出它的好处。

如果一个主机连接在一个小网络上，而这个网络只用一个路由器和因特网连接，那么在这种情况下使用默认路由是非常合适的。

必须强调指出

IP 数据报的首部中没有地方可以用来指明“下一跳路由器的 IP 地址”。

当路由器收到待转发的数据报，不是将下一跳路由器的 IP 地址填入 IP 数据报，而是送交下层的网络接口软件。

网络接口软件使用 ARP 负责将下一跳路由器的 IP 地址转换成硬件地址，并将此硬件地址放在链路层的 MAC 帧的首部，然后根据这个硬件地址找到下一跳路由器。

3. IP 层转发流程 P215－黄

(1)目的地址→目的网络

(2)若相邻，则直接交付

(3)若不相邻，查路由表，找到下一个路由器（主机－>网络－>缺省），转发

(4)找不到，报告错误。

4. L1、L2 和 L3 交换机

- 我们把多端口中继器称为 L1 层的交换机，把网桥称为 L2 层的交换机,把路由器称为 L3。
- 在 L1 层进行交换的端口的集合有时称为冲突域。
- 通过 L1 或 L2 连接在一起的端口的集合有时称为广播域。也就是说，它们是处于一个局域网上。
- L3 交换的对象是分组，而 L2 交换的是帧。第三层交换就是第二层交换加上第三层路由，即把路由功能集成在交换机中。

7.1.1.3 子网及子网掩码

1. 两极 IP 地址存在的缺陷 P202－黄

(1) IP 地址空间的利用率有时很低；

(2) 网络中主机数越多，网络的吞吐量越低(会造成拥塞)；

(3) 物理网络数越多，则路由表越大，不利于路由的查找；

(4) 两极的 IP 地址不够灵活

3. 子网的划分

把主机号中的一部分划为子网号。(从高位划起)。这样，对外仍然表现为一个没有划分子网的网络。使两级 IP 地址变成三级 IP 地址。

IP 地址::= { <网络号>，<子网号>，<主机号>}

- 凡是从其他网络发送给本单位某个主机的 IP 数据报，仍然是根据 IP 数据报的目的网络号 net-id,先找到连接在本单位网络上的路由器。
- 然后此路由器在收到 IP 数据报后，再按目的网络号 net-id 和子网号 subnet-id 找到目的子网。
- 最后就将 IP 数据报直接交付给目的主机。

一个未划分子网的 B 类网络 145.13.0.0

划分为三个子网后对外仍是一个网络

4. 子网掩码的概念

- 从一个 IP 数据报的首部并无法判断源主机或目的主机所连接的网络是否进行了子网的划分。
- 子网掩码也是用 32 位表示。用于区分 IP 地址中的子网号和主机号。

6. 在一个给定的 IP 地址中如何求子网号、主机号？

例如：网络号为 130.51.0.0，子网掩码 255.255.248.0,求第 9 个子网上的第 258 个主机 的 IP 地址？

248 → 11111000

258 → 100000010

7. 采用三级寻址方式

- 用了子网掩码后，寻址采用三级寻址方式。

寻址网络号 子网络号 主机号

- 同一子网内：所有机器的网络地址、子网地址、子网掩码必须相同。
- 广义网络地址=(IP 地址) AND (子网掩码)
- 为了简化路由器的路由算法，对不划分子网的网络采用**默认子网掩码**。将 IP 地址与默认子网掩码进行“与”，即可得到该 IP 地址的网络地址。

A 类、B 类和 C 类 IP 地址的默认子网掩码

子网掩码是一个重要属性

路由器在和相邻路由器交换路由信息时，必须

把自己所在网络（或子网）的**子网掩码**告诉相邻路由器

路由器的路由表中的每一个项目，除了要给出**目的网络地址**外，还必须同时给出该**网络的子网掩码**。。

若一个路由器连接在**两个子网**上就拥有**两个网络地址**和**两个子网掩码**。

【例 1】已知 IP 地址是 141.14.72.24，子网掩码是 255.255.192.0。试求网络地址。

【例 2】在上例中，若子网掩码改为 255.255.224.0。试求网络地址，讨论所得结果。

子网例子：P203-黄

8. 使用子网掩码的分组转发过程

- **划分子网的情况下，从 IP 地址却不能唯一地得出网络地址来，因为网络地址取决于那个网络所采用的子网掩码，但数据报的首部并没有提供子网掩码的信息。**
- **因此分组转发的算法也必须做相应的改动。**
- **采用子网掩码后，路由表中的每一行包含的内容是：**

目的网络地址、子网掩码和下一跳地址

- **在划分子网的情况下路由器转发分组的算法 P134－谢(5)**

在划分子网的情况下路由器转发分组的算法

【例 3】已知互联网和路由器 R_1 中的路由表。主机 H_1 向 H_2 发送分组。试讨论 R_1 收到 H_1 向 H_2 发送的分组后查找路由表。

主机 H_1 首先将

本子网的子网掩码 255.255.255.128 与分组的目的 IP 地址 128.30.33.138 逐比特相“与”(AND 操作)

因此 H_1 必须把分组传送到路由器 R_1

然后逐项查找路由表

路由器 R_1 收到分组后就用路由表中第 1 个项目的

子网掩码和 128.30.33.138 逐比特 AND 操作

路由器 R_1 再用路由表中第 2 个项目的

子网掩码和 128.30.33.138 逐比特 AND 操作

7.1.1.4 无分类域间路由 CIDR（无分类编址 CIDR）

1. 网络前缀 P203－黄

在网络的分类编址中，虽然划分子网在一定程度上缓解了因特网在发展中的困难，但还是存在一些问题：

(1) 分类地址已基本分配完。

(2) 因特网主干网上的路由表的项目数急剧增长。

(3) 整个 IPv4 的地址空间最终将全部耗尽。

为了解决上述问题，采用了变长子网掩码 VLSM(Variable Length Subnet Mask)来提高 IP 地址资源的利用率。在 VLSM 的基础上进一步研究出无分类编址的方法，它的全称是无分类域间路由选择 CIDR(Classless Inter-Domain Routing)。

CIDR 最主要的特点：

CIDR 消除了传统的 A 类、B 类和 C 类地址以及划分子网的概念，因而可以更加有效地分配 IPv4 的地址空间。

CIDR 使用各种长度的“网络前缀”(network-prefix)来代替分类地址中的网络号和子网号。IP 地址从三级编址（使用子网掩码）又回到了两级编址。

- **CIDR 的划分方法：**

IP 地址::={<网络前缀>,<主机号>} 其中，网络前缀代替分类地址中的网络号和子网号。

表示： 起始地址/网络前缀位数。

如：128.14.46.34/20,说明前 20bit 表示网络前缀，主机号是 12bit。

- **CIDR 地址块：指网络前缀相同的连续的 IP 地址的组成**

如：128.14.32.0/20 则块的大小是 2^{12}。其中，全 1 全 0 的主机号一般不使用。

最小地址：128.14.32.0 10000000 00001110 00100000 00000000

最大地址：128.14.47.255 10000000 00001110 00101111 11111111

128.14.32.0/20 表示的地址（2^{12} 个地址）

- CIDR 虽然不使用子网了，但仍然使用“掩码”这一名词（但不叫子网掩码）。
- **掩码的构成：“1”对应网络前缀，“0”对应主机号。**如：/20 表示掩码有 20 个“1”。

CIDR 记法的其他形式

10.0.0.0/10 可简写为 10/10，低位连续的 0 省略

在网络前缀的后面加一个 * 。如：0000101000 *

10.0.0.0/10 隐含地指出 IP 地址 10.0.0.0 的掩码是 255.192.0.0。此掩码可表示为

11111111 11000000 00000000 00000000

- **分类 IP 地址设置成网络前缀应注意问题** P200－谢（4）第一段

在分配 IP 地址是，要注意软件是否支持 CIDR 编址方法。

- **使用 CIDR 的好处：**P137－谢(5)，P204－黄
  - 节约地址

需要 30 个地址：A 类，浪费 2^{24}-2-30，

C 类，浪费 254-30

CIDR /27 ，不浪费

- 地址分配灵活

  可以分配几个地址块

- 路由聚合，减少路由表项

路由聚合(route aggregation)

一个 CIDR 地址块可以表示很多地址，这种地址的聚合常称为路由聚合，它使得路由表中的一个项目可以表示很多个（例如上千个）原来传统分类地址的路由。

路由聚合也称为构成超网(supernetting)。

构成超网

前缀长度不超过 23 位的 CIDR 地址块都包含了多个 C 类地址

这些 C 类地址合起来就构成了超网

如：某大学需要 800 个 IP 地址。在使用分类 IP 地址时，则分配一个 B 类地址，此主机数最大 65534(2^{16}-2)，ISP 浪费了 65534－800＝64734 个 IP 地址。

利用 CRID，则 ISP 分配一个地址块 206.0.68.0/22，它包括 1024(2^{10})个 IP 地址。相当于 4 个连续的 C 类/24 地址块。

如果某 ISP 已拥有地址块 206.0.64.0/18(相当于有 64 个 C 类网络数)，则分配一个地址块 206.0.68.0/22 (相当于有 4 个 C 类网络数) ，只占有该 ISP 的地址空间的 4/64=1/16。这样**提高了地址空间的利用率。**

CIDR 地址块划分举例

CIDR 地址块划分举例

2.最长前缀匹配 P138－139 谢（5）

使用 CIDR 后路由表中的内容应相应改变，它由网络前缀和下一跳组成。这样在查找路由表时可能会得到不止一个匹配结果。

如：假设**目的**的 IP 地址为 D=206.0.71.130，路由表中有两个项目，即 206.0.68.0/22 和 206.0.71.128/25。则 D 与这两项的**掩码**逐比特“与”运算。

D 和 11111111 11111111 11111100 00000000 逐比特相“与”＝206.0.68.0/22 匹配

D 和 11111111 11111111 11111111 10000000 逐比特相“与”＝206.0.71.128/25 匹配

应当从匹配结果中选择具有最长网络前缀的路由。为什么？

网络前缀越长，其地址块就越小，因而路由就越具体。

3.使用二叉线索查找路由表

当路由表的项目数很大时，怎样设法减小路由表的查找时间就成为一个非常重要的问题。

如何使用更好的数据结构和更快的查找算法来查找路由表。

方法： (1) 循环查找：P139－谢（5）

(2) 二叉线索查找：P139－谢（5）

与惟一前缀相匹配。

为了提高二叉线索的查找速度，广泛使用了各种压缩技术。

例:一个自治系统有 4 个局域网，LAN_2 至 LAN_4 上的主机数分别为 91、150 和 15。该自治系统分配的 IP 地址块为 30.138.118/23。试给每个局域网分配地址块。

7.1.1.5 IP 数据报的格式 P122—123 谢(5)，P204－206 黄

1. IP 数据报首部固定部分中的各字段。IP 数据报分片的举例(MTU 为 1420 字节)

2. IP 首部的可变部分

可用来支持排错、测量以及安全等新的功能。目前已定义了 5 类选项，实现的功能分别是：

- **安全性：指明数据的安全程度（级别）。**
- **严格源路径选择：给出完整的路径，数据报需沿此路径传送。**
- **松散源路径选择：给出路由器的地址表，但是数据报还可能经过其他路由器。**
- **记录路由：使每个路由器都附上其 IP 地址，以使管理人**

员了解数据报执行的路径。

- **时间标记（时间戳）：使每个路由器都附上地址及时间**

标记。

7.1.3 地址解析协议 ARP 和逆地址解析协议 RARP

地址的转换:

1. 主机名与 IP 地址的转换（DNS 协议）

2. IP 地址与 MAC 地址的转换（ARP 协议）

(1) 不是一个简单的转换 P119－谢(5)

因为 IP 地址是 32bit，而硬件地址是 48bit。

(2) 每一个主机都有一个 ARP 高速缓存用来存放 IP 地址到 MAC 地址的映射表。主机怎样知道地址？ P119－120 谢(5) 可见下一页

(3) 当主机 A 欲向主机 B 发送 IP 数据报时，先查 ARP 缓存，取出 MAC 地址写入 MAC 帧

7.1.3.1 ARP P216-黄

1. 功能：从 IP 地址找到对应机器的 MAC 地址

2. 三种方法

①**对照表法**。建立一个 IP 地址与 MAC 地址的对照表，并将其存放于每台机器上

②**广播询问法**。广播包含 IP 地址的消息，收到该消息的每台计算机根据自己的 IP 地址确定是否应答。若是，则发送应答消息，将自己的 MAC 地址置于其中，否则不作应答。

③ARP 服务器法

实际使用②

高速缓存提高效率

- 当主机 A 欲向**本局域网**上的某个主机 B 发送 IP 数据报时，先在 ARP 高速缓冲中查看有无主机 B 的 IP 地址。**如有**，则将 B 主机的硬件地址写入 MAC 帧中，然后通过局域网将该 MAC 帧发往此硬件地址的主机。
- 映射表中如果**没有**目的主机的 IP 地址的项目怎么办？

发送站就自动运行 ARP，按以下步骤找出目的站的物理地址：

① ARP 进程在本局域网上广播发送一个 ARP 请求分组，上面有目的站的 IP 地址。

② 在**本局域网**的**所有主机**上运行的 ARP 进程都收到此请求分组。

③ 目的站(B)在 ARP 请求分组中见到自己的 IP 地址，就向发送 ARP 请求的发送主机(A)发回一个写有自己物理地址的 ARP 响应分组。而其它主机**都不理睬**这个 ARP 请求分组。ARP 请求分组是**广播发送**，ARP 响应分组是**单播方式**。

④ 发送主机(A)收到此响应分组后，就在它的 ARP 高速缓存中写入该目的主机(B)的 IP 地址到物理地址的映射。

⑤ 同时目的主机(B)写有发送主机(A)的 IP 地址到物理地址的映射 写入自己的高速缓存中，以备以后它向该发送主机发送分组时使用。 P120－谢(5)图 4-12

小结：

具有广播能力的网络（如，各种类型的局域网）

主机 A 与主机 B 进行数据通信：

(1) A 发 ARP 请求广播帧（带接收方 IP 地址、本机 IP 地址和物理地址）；

(2) B 收到 A 发来的 ARP 请求，予以响应，发 ARP 响应帧，返回自己的物理地址；

(3) 双方用物理地址在物理网中进行数据通信。

应当注意的问题：

- ARP 是解决**同一个局域网**上的主机或路由器的 IP 地址和硬件地址的映射问题。
- 如果所要找的**目的主机和源主机不在同一个局域网上**，通过 ARP **找**到一个位于**本局域网上的某个路由器的硬件地址**，然后把分组发送给这个路由器，让这个路由器把分组转发给下一个网络。剩下的工作就由下一个网络来做。
- 从 IP 地址到硬件地址的解析是**自动进行**的，主机的用户对这种地址解析过程是不知道的。
- 只要主机或路由器要和**本网络**上的另一个已知 IP 地址的主机或路由器进行通信，ARP 协议就会**自动**地将该 IP 地址解析为链路层所需要的硬件地址。

使用 ARP 的四种典型情况

发送方是主机，要把 IP 数据报发送到**本网络**上的另一个主机。这时用 ARP 找到目的主机的硬件地址。

发送方是主机，要把 IP 数据报发送到**另一个**网络上的一个主机。这时用 ARP 找到本网络上的一个**路由器**的硬件地址。剩下的工作由这个路由器来完成。

发送方是路由器，要把 IP 数据报转发到**本网络**上的一个主机。这时用 ARP 找到目的主机的硬件地址。

发送方是路由器，要把 IP 数据报转发到**另一个**网络上的一个主机。这时用 ARP 找到本网络上的一个路由器的硬件地址。剩下的工作由这个路由器来完成。

为什么我们不直接使用硬件地址进行通信？

由于全世界存在着各式各样的网络，它们使用**不同的硬件地址**。要使这些异构网络能够互相通信就必须进行非常复杂的硬件地址转换工作，因此几乎是不可能的事。

连接到因特网的主机都拥有统一的 IP 地址，它们之间的通信就像连接在同一个网络上那样简单方便，因为调用 ARP 来寻找某个路由器或主机的硬件地址都是由计算机软件自动进行的，对用户来说是看不见这种调用过程的。

ARP 包格式

7.1.3.2 反向地址解析协议(RARP)

1. 功能：实现从 MAC 地址到 IP 地址的映射

这种情况往往是无盘工作站 P217－黄

无盘站要运行 ROM 中的 RARP 来获得其 IP 地址。

2. RARP 的工作方法：

- **为了使 RARP 能工作，在局域网上至少有一个主机要充当 RARP 服务器，该服务**

器有一个从无盘工作站的物理地址到IP地址的映射表。

- 无盘工作站先向局域网广播发出RARP请求分组（在格式上与ARP请求分组相似），并在此分组中给出自己的物理地址。
- 当RARP服务器收到请求分组后，它就从它的映射表中查出请求工作站的IP地址，而后就把该IP地址写入RARP响应分组，发回给无盘工作站。

7.1.4 Internet控制报文协议ICMP

1. ICMP协议和IP协议之间的关系

ICMP(Internet Control Message Protocol)是在网际层使用的协议，ICMP允许主机或路由器报告差错情况和提供有关异常情况的报告。ICMP报文是IP数据报的数据部分。P140－谢(5)，P217-黄

注意：ICMP不是高层协议，它仍是IP层的协议。

2. ICMP的功能：允许主机和路由器报告差错情况和提供有关异常情况的报告。网络内部使用，用于网络管理。

3. ICMP报文种类可分为：ICMP差错报文和ICMP询问报文。P141－谢(5)，P218－黄

4. ICMP差错报文种类：分为5种

(1) 终点不可达。

(2) 源站抑制；拥塞控制。

(3) 时间超过；TTL＝0。

(4) 参数问题；

(5) 改变路由；通知主机改变路由器的路径。

ICMP差错报告报文共有5种

在ICMP差错报文中，改变路由报文用的最多。

例如：可改变路由表和解决拥塞

现在A和C通信。但A的路由表中只有一个默认路由器R1。所以主机A发往主机C的分组就到了R1上。当R1查它的路由表可知，发往主机C的分组应经过R2。于是它把分组转发给R2，最后传到主机C。

很显然这样走不好，于是R1向主机A发送ICMP改变路由报文，指出主机A发往主机C的分组应先发给R2，这时它给出的是R2的IP地址。

主机A根据收到的信息更新其路由表。以后再有A发到C的分组就走新路由。

ICMP差错报告报文的数据字段的内容

5.不应发送ICMP差错报告报文的几种情况

对ICMP差错报告报文不再发送ICMP差错报告报文。

对第一个分片的数据报片的所有后续数据报片都不发送ICMP差错报告报文。

对具有多播地址的数据报都不发送ICMP差错报告报文。

对具有特殊地址（如127.0.0.0或0.0.0.0）的数据报不发送ICMP差错报告报文。

6. ICMP询问报文有两种

回送请求和回答报文

时间戳请求和回答报文

下面的几种ICMP报文不再使用

信息请求与回答报文

掩码地址请求和回答报文

路由器询问和通告报文

7. ICMP应用

目前，已经利用ICMP报文开发了许多网络诊断工具软件。

（1）Ping (Packet InterNet Groper)软件

- ICMP回送请求与回答报文测试两主机的连通性。
- PING是应用层直接使用网络层ICMP的例子，它没有通过运输层的TCP或UDP。
- 实现方法：

－测试方发送类型值为13的请求报文，被测试方回送类　　型值为14、且包含时间戳的应答报文。

－PING一般连续三次。

（2）跟踪IP数据报发送的路由

利用路由器对IP数据报中的生存期值作减1处理，一旦生存期值为0就丢弃该IP数据报，并返回主机不可达的ICMP报文的特点。

方法：源发端针对指定的目的结点，形成一系列收方结点无法处理的IP数据报。这些数据报除生存期值递增外，其它内容完全一样。第一个数据报的生存期为1；路由器对生存期值减一后，丢弃该IP数据报，并返回主机不可达ICMP报文；源发端继续发送生存期为2，3，4，…的数据报，由于主机和路由器中对路由信息的缓存能力，IP数据报将沿着原路径向目的结点前进。如果整个路径中包括了N个路由器，则通过返回N个主机不可达报文和一个端口不可达报文的信息，了解IP数据报的整个路由。Traceroute的应用举例

（3）测试整个路径的最大MTU(指帧格式的最大长度)

这种测试对于源宿端具有频繁的大量数据传输时，具有较高的实用价值。因为数据报长度越小，数据报传输的有效率越低；而传输较大的数据报时，路由器势必进行分段，既损耗了路由器的资源，更可能造成因某个数据

分段丢失，而导致宿主机在组装分段数据报时超时，丢弃整个数据报，造成带宽的浪费。

测试路径 MTU 的方法类似路由跟踪。**源发端发送一定长度、且不允许分段的 IP 数据报，并根据路由器返回的主机不可达 ICMP 报文，逐步缩短测试 IP 数据报的长度。**

（4）拓扑发现

- 目的：找出网络上有哪些主机 。
- 实现方法：

从第一个地址开始，依次向每个地址发送一个回送请求报文（类型=8），如果收到回答（类型=10），表明该地址当前连在网络上。

- 做成示意图，点亮。

7.1.2 Internet 路由协议　　P207－黄

- 按数据报工作方式传送，无连接。
- 分层次的路由选择协议。
- 网络层的**主要功能之一是将数据包从源节点送出经过一系列节点的存储转发最终送到目的节点。**
- 任意两个节点之间都可能**存在多条路径**，网络层在转发数据包的过程中必须确定转发路径。
- **不同的网络对路径选择算法的要求不一样，如军用网络要求可靠，普通商用网络要求经济，实时网络要求快速。**

功能：在通信子网内选择从源节点到目的节点的路径。

要求：　正确、简单、健壮、可靠、公平、最优。

优劣标准：最短（时间、成本、链路数…）。

7.1.2.1 有关路由器选择协议的几个基本概念

产生路由表的算法，称为路由算法。

1. 理想的路由算法　P144－谢(5)

(1)算法必须是正确的和完整的：正确性是指：路径选择算法应能使数据包迅速、正确地传送。

(2)算法在计算上应简单：算法应尽量简单，易实现，开销小，路由选择不应让网中的结点增加太多的开销。

(3)算法应能适应通信量和网络拓扑的变化：即要有**自适应性**或**稳健性**。算法能适应网络拓扑结构及流量地变化，在外部条件发生变化时能正确地完成要求的功能。

(4)算法应具有稳定性和可靠性：当通信量和网络拓扑稳定时，路由算法应**收敛于**一个可以**接受的解**。不管运行多长时间，均应保持正确。如计数器必须要有足够的位数等。

(5)算法应是公平的：算法对所有的用户都是公平的，不考虑优先条件。各节点具有均等的发送信息的机会。

(6)算法应是最佳的：所选路由费用最低。在某一特定条件下，求得的路由是最为合理的。

2. 关于“最佳路由”

不存在一种**绝对**的最佳路由算法。

所谓“最佳”只能是**相对**于某一种特定要求下得出的较为合理的选择而已。

实际的路由选择算法，应尽可能接近于理想的算法。

路由选择是个非常复杂的问题

它是网络中的所有结点共同协调工作的结果。

路由选择的环境往往是不断变化的，而这种变化有时无法事先知道。

路由选择与流量控制有一定的关系。好的流量控制可以使更多的通信量流入网络，而好的路由选择可以使网络的平均时延较低。网络的平均时延总是随网络的通信量增大。

吞吐量和平均时延是一个网络的最主要的性能指标。

3.非自适应路由选择(静态路径选择算法)

条件：不考虑网络的状态。

(1) 固定路由法

算法思想：当节点收到数据包后，检查目的地，然后在输出线**选择表**中查找到该目的节点的主路径输出线并从该输出线上转发数据包。**表中可以规定多条输出线**。算法速度快、开销小。

缺点：固定路径由**网管人员指定**，一旦网络本身出现故障或其它原因导致拓扑结构发生变化，则原来指定的路径就可能走不通，数据包无法到达目的地，必须重新指定路径。

用途：永久虚电路。

(2) 分散通信量法

算法思想：是事先在每个结点的**内存中设置**一个路由表，路由表中给出**几个可供采用**的输出链路，并且对每条链路赋予一个**概率**。当一个分组到达该结点时，此结点即产生一个从 0.00 到 0.99 的随机数，然后按此随机数的大小，查表找出相应的输出链路。　　**例如：分组到 K 站，目的站为 B。**

目的站	经过	概率	经过	概率	经过	概率
A	M	0.50	L	0.40	N	0.10
B	M	0.35	N	0.35	L	0.30
C	N	0.65	M	0.25	P	0.10
E	N	0.55	P	0.30	M	0.15
D	P	0.45	N	0.30	M	0.25
…	…		…		…	

这种方法与固定路由相比，可使网内的通信量更加平衡和得到较小的平均分组时延。

(3) 洪泛法(扩散路径选择算法)

算法思想：当某个结点收到一个不是发给它的分组时，就**向所有与此结点相连的链路转发出去**。当然，**不能再把这个分组发到它刚刚离开的那个结点**，否则就永远有一些分组来回不停地在各条链路上“振荡”。

缺点：会产生大量的重复包，包的数目可能会呈指数规律增加。结果导致网络出现拥塞。

(3) 洪泛法(扩散路径选择算法)

解决办法：① 采用站计数法，即在每个包中增加一个站点**计数字段**，初值设为从源节点到目的节点的路径长度(最多节点数)。数据包每经过一个节点，站计数器减 1。当该值变为 0 时，若还未到达目的节点，就丢弃该数据包。

② **首次登录法**，即在每个包中增加一个序号字段，在每个节点设置一张表记录首次到达本节点的包的次序。**当收到一个包时**，检查相应源节点发送的**该包是否首次到达节点**。若是，则登录序号，并扩散转发；否则，丢弃该包。

(4) 随机走动法（随机路径选择算法）

算法思想：当数据包到达一个节点后，**随机选择一条输出线**转发该数据包。有两种方法，一是**完全随机法**；二是**轮选法**。

缺点：可能将所收到的数据包又从输入线上转发出去，即将**数据包原路返回**。

解决办法：采用**计程法**，即在数据包中增加一个字段，记录包所经过的节点。

4.**动态路径选择算法(自适应路由选择算法)**

条件：需要考虑网络当前的状态

工作过程包括如下四部分：

- **测量：**测量并感知网络状态，主要包括拓扑结构、流量及通信延迟。
- **报告：**向有关进程或结点报告测量结果。
- **更新：**根据测量结果更新路由表。
- **决策：**根据新路由表重新选择合适路径转发数据分组。

(1) 孤立自适应路径选择算法

这类算法只根据本结点获知网络信息确定数据分组的输出线，**结点之间不交换路径信息**。

① 热土豆算法

算法思想：在网络中，每条输出线都有若干缓冲区，供等待输出的数据分组排队之用。每收到一个数据分组，总是选择**队列最短**的输出线转发数据分组，以求**最快输出**。

缺点：只**考虑队列的长度即分组的数量，没有考虑网络的带宽及全网的负载状况**。当网络每部分的带宽不一样时，该算法不能保证转发的路径是最优路径。

②反向探知算法

算法思想：通过本结点先前转发过的分组中所记录的目的结点到源结点的数据分组，则本结点的可利用该信息，试探着沿原路径的反向路径转发数据分组。**即根据反向路由推测正向路由**。

缺点：①路径信息是间接的，不可靠的；②当没有反向路径信息时，正常的路径选择就难以完成；③存在来回传送即振荡的可能。

静态路由选择策略**特点**—— 简单和开销较小，但不能及时适应网络状态的变化。

动态路由选择策略**特点**—— 能较好地适应网络状态的变化，但实现起来较为复杂，开销也比较大。

4.分层次路由选择协议

为什么采用分层的路由选择协议？ P145 一谢(5)

- 因特网规模非常大，不利于路由器进行路由选择。
- 有些单位不愿意外界了解自己单位的网络布局和路由选择协议。

采取方法：

- 将 Internet 划分为较小的**自治系统 AS** 。
- 在 AS 内部选择路由采用**内部网关协议** IGP（RIP、OSPF），在不同的 AS 之间选择路由采用**外部网关协议** EGP （BGP）

(1) 自治系统的概念

同类型的路由器（实现同样的路由算法）互联的，由同一机构控制的互联网络部分称为自治系统（Autonomous System），简称为 AS。

一个 AS 是一个互联网，其最重要的特点就是它有权自主地决定在本系统内应采用何种路由选择协议。

一个自治系统内的所有网络都属于一个行政单位来管辖。

(2) 内部网关协议 IGP（Interior Gateway Protocol）

在一个 AS 内部使用的路由选择协议，主要设法使数据报从源到目的站传送的尽可能有效，而不考虑其它方面。这与互联网中的其它 AS 选用什么路由选择协议无关。

用的最多的协议 RIP（Routing Information Protocol），OSPF(Open Shortest Path First)

(3) 外部网关协议 EGP (External Gateway Protocol)

外部邻站使用的向其他自治系统通知可达信息的协议称为外部网关协议(EGP)。使用 EGP 的路由器称为外部路由器。

若在两个 AS 之间传输数据时，当数据报传到一个 AS 的边界时，就需要使用一种协议将路由选择信息传到另一个 AS 中。这种协议就是外部网关协议。主要考虑的是路由选择策略（政治、经济、安全），目前使用的最多的是 BGP-4 协议。

这里要指出两点

因特网的早期 RFC 文档中未使用“路由器”而是使用“网关”这一名词。但是在新的 RFC 文档中又使用了“路由器”这一名词。应当把这两个属于当作同义词。

IGP 和 EGP 是协议类别的名称。但 RFC 在使用 EGP 这个名词时出现了一点混乱，因为最早的一个外部网关协议的协议名字正好也是 EGP。因此在遇到名词 EGP 时，应弄清它是指旧的协议 EGP 还是指外部网关协议 EGP 这个类别。

7.1.2.2 RIP（路由信息协议）

1. 距离向量路由算法 [分布式算法：分布式 Bellman－Ford 算法]。

①原理

- 每个节点保存一张距离向量表。
- 表中每行表示一个目的网络，代表从本节点到该节点的**最短距离**及其对应的**输出链路**。
- 通过与**相邻节点**交换距离信息更新距离向量表。
- 转发数据包时根据包中的目的地址查找到该节点的输出节点并转发。

②距离信息传播

- 每个节点定时测量到**相邻节点**的距离，并把结果**广播**到邻节点（Internet：30 秒，超过 180 秒为不可达）。
- 每个节点收到距离信息包后**更新距离向量表**。

③更新方法

节点 J 经相邻节点到达目的节点 Y，J 的**邻节点**为 X_1，X_2，…，X_n。则 J 需要选择输出线（X_1，X_2，…，X_n）之一进行转发。选择前先计算延迟时间：

$$T_{JYmin} = \min\{t_{JX1}+T_{X1Y}, t_{JX2}+T_{X2Y}, \cdots, t_{JXn}+T_{XnY}\}$$

t_{JX1}，t_{JX2}，…，t_{JXn} 是当前已知数值；T_{X1Y}，T_{X2Y}，…，T_{XnY} 是各邻节点到目的节点的延迟，通过交换信息后得到的值。找出 T_{JY} 最小的一条路径。

2. 原理

- 分布式算法
- 基于距离向量路由算法
- 距离定义：跳（hop）数（链路数）
- 最大值（TTL）：15，即 16 段链路

3. RIP 路由信息的传送

- 获取路由信息：与**相邻节点**交换信息来获得邻居节点信息
- 交换路由信息：将本节点的**路由表传送给邻居节点**
- 交换信息的时机：周期性，一般 30 秒。超过 180 秒未收到邻居的路由表时，将邻居标为不可达。

3. RIP 路由表的更新工作过程：P148－谢(5)，P208－黄

(1) 收到相邻路由器 X 的 RIP 报文，将 RIP 报文中“下一跳”字段都改为 X，将所有“距离”＋1

(2) 对修改后的 RIP 报文中的每一行，重复做：

若不在路由表中，则添加到路由表中；

否则， 若下一跳的内容与路由表中的相同，则替换路由表中的对应行；

否则，若收到的距离小于路由表中的距离，则更新路由表；

(3) 若 3 分钟还未收到，邻居路由器的路由表，则将到邻居路由器的距离置为 16。

4. RIP 协议的三个要点

仅和相邻路由器交换信息。

交换的信息是当前本路由器所知道的全部信息，即自己的路由表。

按固定的时间间隔交换路由信息，例如，每隔 30 秒。

5. 进一步说明 P209-210 谢（4），P147-谢（5）

- **RIP 是一种分布式的基于距离向量的路由选择协议。从一路由器到直接的网络的距离定义为 1。从一路由器到非直接的网络的距离定义为所经过的路由器数加 1。这里“距离”称为“跳数”。**
- **RIP 认为一个好的路由就是它通过的路由器的数目少，即“距离短”。**
- **IP 不能在两个网络之间同时使用多条路由。这是因为 RIP 选择距离最短(路由器)的路由，而不管是否还有另一条路由器较多的路由。**
- **RIP 进程使用 UDP 的 520 端口来进行发送和接收。RIP 协议的位置在应用层，但转发 IP 数据报的过程在网络层完成。**

6. RIP 存在的缺点 P213－214 谢（4）

- 当网络出现故障时，要经过较长的时间才能将信息传送到所有的路由器(**坏消息反应慢**)。
- RIP 协议限制了网络的**规模**，因为最大有效距离为 15。

- **开销大**。每个结点不仅要记录大量数据，而且还要周期性地与邻结点交换信息，增加大量地通信开销。
- 使更新路由表的过程较长，不适合规模较大的网络。
- 可能造成阻塞。

7. RIP 解决无穷计算的方案

8.RIP 协议的报文格式 P150－谢(5),P208－黄

(1) 报文格式

课堂练习

7.1.2.3 OSPF（开放最短路径优先）路由协议

1. OSPF 协议简介

■ “开放”表明 OSPF 协议不是受某一家厂商控制，而是公开发表的。

■ “最短路径优先”是因为使用了 Dijkstra 提出的最短路径算法 SPF。

■ OSPF 只是一个协议的名字，它并不表示其他的路由选择协议不是“最短路径优先”。

■ 是分布式的链路状态协议。

2. 链路状态路由算法

■ 每个节点保存一张链路状态表(矩阵)；

节点	1	2	3	…
1				
2				
…				

■ 矩阵表中每个元素(i,j)表示节点 i 与节点 j 之间的状态（距离、成本、带宽…)；

■ 转发数据包时利用 Dijkstra 算法计算到目的节点的最短路径并转发。

Dijstra 算法

3. OSPF 特点

■ 向 AS 内的所有路由器扩散路由信息；

■ 扩散的路由信息只有通向邻居节点的链路状态；

■ “链路状态”就是说明本路由器都和哪些路由器相邻，以及该链路的“度量”(metric)。

■ 只有链路状态发生变化时才扩散；

■ 不同链路可使用不同的成本度量值。

4. 链路状态数据库

由于各路由器之间频繁地交换链路状态信息，因此**所有的路由器**最终都能建立一个链路状态数据库。

这个数据库实际上就是全网的拓扑结构图，它在全网范围内是一致的（这称为链路状态数据库的同步）。

OSPF 的链路状态数据库能**较快地进行更新**，使各个路由器能及时更新其路由表。OSPF 的更新过程收敛得快是其重要优点。

4. OSPF 协议与 RIP 协议不同特点：

(1) 采用洪泛法向本自治系统中所有路由器发送信息。而 RIP 仅仅向自己相邻的几个路由器发送信息。

(2) 发送的信息就是与本路由器相邻的所有路由器的链路状态信息，它只是路由器所知道的部分信息。而 RIP 协议发送的信息是“到所有网络的距离和下一跳路由器”。

(3) 只有当链路状态发生变化时，路由器才用洪泛法向所有路由器发送此信息。而 RIP 不管网络拓扑结构有无变化，路由器都要定期交换路由表的信息。

5. OSPF 的工作原理： P214－215 谢(4)，P152－153 谢(5)，

■ OSPF 通过路由器之间采用**洪泛法**通告网络接口的状态来建立整个网络的链路状态数据库，**利用最短路径算法**生成最短路径树，每个 OSPF 路由器使用这些最短路径构造路由表。

6.支持的网络连接形式

■ 两个路由器之间的点到点连接；

■ 具有广播功能的局域网；

■ 无广播功能的广域网。

7.路由过程

(1)获取链路状态信息（邻居信息)；

(2)扩散链路状态信息 ；

(3)Dijkstra 算法计算最短路径。

区域划分好处：就是利用洪泛法交换链路状态信息的范围局限于每一个区域而不是整个自治系统，减少整个网络上的通信量。

什么是主干区？什么是主干路由器？什么是区域边界路由器？什么是自治系统边界路由器？ P153-谢(5)

OSPF 直接用 IP 数据报传送

OSPF 不用 UDP 而是直接用 IP 数据报传送。OSPF 构成的数据报很短。这样做可减少路由信息的通信量。

数据报很短的另一好处是可以不必将长的数据报分片传送。分片传送的数据报只要丢失一个，就无法组装成原来的数据报，而整个数据报就必须重传。

8. OSPF 路由信息的传送

(1) 格式说明：

■ **类型：5 种分组类型** P154－155 谢(5)，P211－黄

a.类型 1，问候(HELLO)分组，用来发现和维护邻站的可达性。

（每 10 秒一次，三次收

不到时断开）

b.类型 2，数据库描述分组，向邻站给出自己的链路状态数据库中的所有链路状态项目的摘要信息。

c.类型 3，链路状态请求分组，向对方请求发送某些链路状态项目的详细信息。

d.类型 4，链路状态更新分组，用洪泛法对全网更新链路状态。

e.类型 5，链路状态确认分组，对链路更新分组确认。

- **分组长度：包含 OSPF 首部的总长度**
- **路由器标识符：发送该分组的路由器的 IP 地址**
- **区域标识符：该分组所属的区域**
- **鉴别类型：0－不需要，1－口令**
- **鉴别：鉴别类型＝1 时填入 8 个字符口令**

(2) 各种分组的工作过程

9. OSPF 链路状态表的更新　P211－黄

7.1.2.4 外部网关协议(边界网关协议 BGP)

外部网关协议是不同自治系统的路由器之间交换路由信息的协议。

1.内部网关协议的作用是尽可能将数据报在一个自治系统中有效地从源站转送到目的站。

2.边界网关协议产生的原因　P220－谢（4），P157－谢（5）

- 因特网的**规模太大**。
- 对于自治系统之间的路由选择,要**寻找最佳路由**是很不现实的。
- 自治系统之间的路由选择必须考虑有关策略。

边界网关协议的作用是力求寻找一条能够到达目的网络且比较好的路由(不存在回路)，而并非要寻找一条最佳路由。

3. BGP 的工作原理　见 P220－221 谢（4）

通过 BGP 发言人(路由器)进行会话的建立连接和撤销连接。

- 每个 AS 指定一个 BGP 发言人（边界路由器）；
- BGP 发言人之间使用 TCP (端口号为 179)通信交换路由信息(构造路由表)；
- 计算 BGP 发言人之间的最短路径（包含具体路径）

BGP 发言人交换路径向量

BGP 发言人交换路径向量

4. BGP 协议的特点

BGP 协议交换路由信息的结点数量级是自治系统数的量级，这要比这些自治系统中的网络数少很多。

每一个自治系统中 BGP 发言人（或边界路由器）的数目是很少的。这样就使得自治系统之间的路由选择不致过分复杂。

BGP 协议的特点

BGP 支持 CIDR，因此 BGP 的路由表也就应当包括目的网络前缀、下一跳路由器，以及到达该目的网络所要经过的各个自治系统序列。

在 BGP 刚刚运行时，BGP 的邻站是交换整个的 BGP 路由表。但以后只需要在发生变化时更新有变化的部分。这样做对节省网络带宽和减少路由器的处理开销方面都有好处。

5.BGP-4 使用的四种报文来进行 BGP 路由表的建立和维护　P159－谢（5），P213－黄

- 打开(Open)报文，用来与相邻的另一个 BGP 发言人建立关系。
- 更新(Update)报文，用来发送某一路由信息，以及列出要撤销的多条路由。
- 保活(Keepalive)报文，用来确认打开报文和周期性地证实邻站关系。
- 通知(Notification)报文，用来发送检测到的差错。

6. BGP 报文 P159－160 谢（5），P213－黄

BGP 报文的格式

6. BGP-4 工作过程

(1) BGP 发言人向邻居 BGP 发言人发送 Open 报文，对方确认后，双方即称为邻居（只有邻居之间才能发送路由信息）

(2) BGP 发言人可向邻居发送 Update 报文，告知对方路由信息，或撤销先前的路由

(2) 收到 Update 报文的 BGP 发言人，记录或更新路由信息，并将其发往其它邻居

(4) 每个 BGP 发言人根据自己保存的路由信息，为两个不同的 AS 确定一条可行的路由。

7.1.2.5　IP 组播协议 PIM 与 MOSPF

组播可明显地减少网络中资源的消耗

组播可明显地减少网络中资源的消耗

多播可明显地减少网络中资源的消耗。

局域网的多播的实现是用硬件。见 P224－谢（4）

因特网的多播的实现是靠路由器。见 P224－谢（4）

2. IP 多播所具有的特点：见 P224－225 谢(4)，P164－165 谢(5)

(1) 多播使用组地址

用 D 类地址表示，地址范围 224.0.0.0~239.255.255.255。主机组共有 28bit，有 2^{28} 个多播组。

(2) 永久地址

由因特网号码指派管理局所分配。

(3) 动态的组成员

主机组中的成员是动态的，通过多播路由器周期性地向本地网络上的主机进行轮询，以确定主机组中的成员。

(4) 使用硬件进行多播

通过以太网本身具有硬件多播能力进行多播。

D 类 IP 地址与以太网多播地址的映射关系

3. 因特网组管理协议 IGMP

IGMP(Internet Group Management Protocol) 位于网际层。其目的帮助多播路由器识别加入到一个多播组的成员主机。IGMP 与 ICMP 一样，使用 IP 数据报传递其报文。

IGMP 的工作分两个阶段：见 P226－谢(4), P167－谢（5）

第一阶段：当某个主机加入新的多播组时，该主机应向多播地址发送一个 IGMP 报文，声明自己要成为该组的成员。

第二阶段：

●由于组的成员是动态的，因此本地多播路由器要周期性地探询本地局域网上的主机，以便知道这些主机是否还继续是组的成员。

●只要对某个组有一个主机响应，那么多播路由器就认为这个组是活跃的。

●但一个组在经过几次的探询后仍然没有一个主机响应，则不再将该组的成员关系转发给其他的多播路由器。

IGMP 使多播路由器
知道多播组成员信息
IGMP 的本地使用范围
IGMP 并非在因特网范围内对所有多播组成员进行管理的协议。
IGMP 不知道 IP 多播组包含的成员数，也不知道这些成员都分布在哪些网络上。
IGMP 协议是让连接在本地局域网上的多播路由器知道本局域网上是否有主机（严格讲，是主机上的某个进程）参加或退出了某个多播组。

IGMP 采用的具体措施：见 P226－谢（4），P168－谢（5）

成员加入和退出多播组时 IGMP 的工作情况：见 P227－谢（4）

IGMP 的报文格式：见 P227－228 谢（4）

当多播数据报在传输的过程中，若遇到有不运行多播软件的路由器或网络时，就要采用一种隧道技术。

4.多播路由协议
PIM
(1)PIM-SM
约会集中点 RP
类似 CBT (Core Based Tree)
(2)PIM-DM
先扩散再剪枝
MOSPF
依赖 OSPF
计算一棵树
7.1.5 IPv6 (IPng)
IPv4 问题
IP 地址不够用
路由表爆炸
流量激增（Tbps）
QoS
安全问题
7.1.5.1 IPv6 特点
① 更大的地址空间:128 位
每平方米大约可分得 6.6×10^{23} 个 IP 地址
② 有效的分级寻址和路由结构
③ 简洁的首部格式
减少了协议的开销，处理效率更高。
④ 增强的安全性
⑤ QoS 支持
可区分流和优先级，提供 QoS 支持
⑥ 自动地址配置
简化网络配置和管理。
IPv6 数据报的首部
IPv6 将首部长度变为固定的 40 字节，称为基本首部(base header)。
将不必要的功能取消了，首部的字段数减少到只有 8 个。
取消了首部的检验和字段，加快了路由器处理数据报的速度。
在基本首部的后面允许有零个或多个扩展首部。
所有的扩展首部和数据合起来叫做数据报的有效载荷(payload)或净负荷。
7.1.5.2 IPv6 的地址
1. 地址结构
IPv6 将 128 位地址空间分为两大部分。
可变长度的类型前缀，定义地址的目的
其余部分，长度可变
7.1.5.2 IPv6 的地址
站点本地（Site Local）：相当于 IPv4 中的 192.168。
链路本地（Link Local）：单一链路上的双方通信用的地址，自动配置。
2. IPv6 地址表示
32 个十六进制位，四个一组，各值之间用冒号分隔

68E6:8C64:FFFF:FFFF:0:1180:960A:FFFF
零压缩，一连串连续的零可以用一对冒号取代
FF05:0:0:0:0:0:0:B3　　可以写成：FF05::B3
点分十进制记法的后缀
0:0:0:0:0:0:128.10.2.1
　　再使用零压缩即可得出：　::128.10.2.1
CIDR 的斜线表示法仍然可用。
60 位的前缀 12AB00000000CD3 可记为：

12AB:0000:0000:CD30:0000:0000:0000:0000/60
或 12AB::CD30:0:0:0:0/60
或 12AB:0:0:CD30::/60
3.　IPv6 的地址类型
三种地址类型：
(1) 单播(unicast) 地址：点到点
(2) 组播(multicast) 地址：　一点对多点的通信
(3) 任播(anycast) 地址（泛播）：　任播的目的站是一组计算机，但数据报在交付时只交付其中的一个，通常是距离最近的一个。
（1）单播地址
两种：
A.**可聚集全球单播地址**，n 一般为 48
全球路由前缀：由 IANA 管理
子网 ID：ISP 在自己的网络中建立多级结构所使用的区分 ID
接口 ID：节点与子网的接口地址，即 EUI-64 地址（将 48 位的 MAC 地址映射为 EUI-64 地址，映射方法：首字节第 7 位取反，第 3、4 字节之间插入 FF-FE。见下面）。
可实现三层路由结构：第一层为公共拓扑，表示多个 ISP。第二层为站点拓扑，表示一个机构内部子网的层次结构。第三层唯一标识一个接口。
接口 ID：EUI-64
IEEE 定义了一个标准的 64 位全球唯一地址格式 EUI-64。
EUI-64 前三个字节(24 位)仍为公司标识符，但后面的扩展标识符是五个字节（40 位）。
把以太网地址转换为 IPv6 接口 ID
（1）单播地址
B.本地单播地址
自动配置
记为 FE80::/64
（2）组播地址
标记：ORPT，R：是否为内嵌汇聚点地址的组播，P：是否为基于单播网络前缀的组播地址，T：暂时态标记，0 为 IANA 分配的永久性地址，1 为临时型地址
范围：IPv6 组播分组传送的范围，0-保留，1-节点本地范围，2-链路本地范围，5-站点本地范围，8-机构本地范围，E-全球范围，F-保留
组 ID：前 80 位为 0，后 32 位为 ID
（3）任播地址
（4） 特殊地址
未指明地址：全 0 地址，记为::，为还没有配置到一个标准的 IP 地址的主机当作源地址使用，类似于 IPv4 的 0.0.0.0。
回送地址： 0:0:0:0:0:0:0:1（记为 ::1）
基于 IPv4 的地址：与 IPv4 兼容
（5）主机 IPv6 地址
一个网卡至少三个地址：
链路本地地址
全球单播地址
回送地址
主机侦听的地址类型有：
节点本地范围内组播地址（FF01::1）
链路本地范围内组播地址（FF02::1）
本节点为目的地址的单播地址
同组组播地址
（6）路由器 IPv6 地址
路由器要侦听的地址类型：
节点本地范围内所有节点的组播地址（FF01::1）
节点本地范围内所有路由器的组播地址（FF01::2）
链路本地范围内所有节点的组播地址（FF02::1）
链路本地范围内所有路由器的组播地址（FF02::2）
站点本地范围内所有路由器的组播地址（FF05::2）
全球单播地址
同组组播地址
7.1.5.3　IPv6 分组格式
IPv6 的基本首部
16 字节（128 位）地址
简化包头（8 个字段），取消了检验和字段，加快路由器处理速度
支持更多选项
增强安全性（如身份认证）
对 QoS 的支持
➔
保留了 IPv4 的优点，抛弃了 IPv4 的缺点
与 IPv4 不兼容，与 Internet 其它协议兼容
通信量类
Priority:0~7 表示可以进行流量控制（减速）的包，8~15 表示恒速传送（实时通信）
0:无特征通信量

1:填充通信量（如新闻）
2:不需等待的数据（如邮件）
3:保留
4:要等待的块传输（如 FTP、HTTP）
5:保留
6:交互式通信量（如 TELNET）
7:Internet 控制通信量（如路由协议、SNMP）
8:可能丢弃的
…
15:最不能丢弃的
IPv6 的扩展首部
扩展首部由源站和目的站主机来处理
分组途中经过的路由器都不处理扩展首部（只有一个首部例外，即逐跳选项扩展首部）。
大大提高了路由器的处理效率。
六种扩展首部
在 RFC 2460 中定义了六种扩展首部：
逐跳选项
路由选择
分片
鉴别
封装安全有效载荷
目的站选项
IPv6 的扩展首部
扩展首部举例
IPv6 把分片限制为由源站来完成。源站可以采用保证的最小 MTU（1280 字节），或者在发送数据前完成路径最大传送单元发现(Path MTU Discovery)，以确定沿着该路径到目的站的最小 MTU。
分片扩展首部的格式如下：
扩展首部举例
IPv6 数据报的有效载荷长度为 3000 字节。下层的以太网的最大传送单元 MTU 是 1500 字节。
分成三个数据报片，两个 1400 字节长，最后一个是 200 字节长。
7.1.5.4 IPv6 地址自动配置
分配地址、指定前缀
两种：有状态、无状态
有状态地址自动配置 P226-黄
主机从专用地址分配服务器获得接口地址，或从服务器上获得配置信息和参数
服务器中维护着一个数据库，记录着主机和地址分配的列表
自动配置协议 DHCPv6
无状态地址自动配置 P227-黄
IPv6 地址由 64 位前缀和 64 位 EUI-64 接口 ID 组成
(1)根据链路本地前缀 FE80::/64 与 EUI-64 接口标识符生成临时链路本地地址。
(2)通过发送“邻节点请求”报文，进行重复地址检测。如果接收到 “邻节点公告”报文，表明已经有节点在使用该临时地址，停止；否则，可以使用该链路本地地址。
(3)发送“路由器请求”报文，要求本链路的路由器发送带有各种路由器信息的“路由器公告报文”，路由器周期性地发送“路由器公告”报文。
(4)如果接收到“路由器公告”报文，主机根据报文内容来设置跳数限制、可到达时间、重发定时器和 MTU。
无状态地址自动配置（con’t）
(5)如果有地址前缀选项，则：
①若链路标识为 1，将前缀添加到前缀队列。
②若自治标识为 1，用前缀和 EUI-64 接口 ID 生成一个临时地址，并检测其唯一性。
(6)如果“路由器公告”报文中的管理地址配置标识为 1，则使用有状态地址自动配置方式获取其它地址。
(7)如果“路由器公告”报文中的有状态配置标识为 1，则使用有状态地址自动配置方式获取其它配置参数。
7.1.5.5 ICMPv6
具备 ICMPv4 的所有基本功能
删除了所有不再使用的报文类型，定义了新的报文类型
合并了 ICMP、IGMP、ARP 等多个协议的功能
在 IPv6 下，没有了 ARP 与 RARP
ICMPv6 需要通过 IPv6 传送，是 IPv6 的一部分，二者配合完成规定的功能
ICMPv6 报文格式
7.1.5.6 邻节点发现
① 路由器发现
② 前缀发现
③ 参数发现
④ 地址自动配置
⑤ 地址解析
⑥ 下一跳选择
⑦ 邻节点不可达检测
⑧ 重复地址检测
⑨ 重定向
7.1.5.7 IPv6 分组转发过程
(1)检验版本字段的值，确定所使用的协议。
(2)递减跳数字段的值，如果=0，丢弃该分组，并向源节点发送超时 ICMPv6 报文。
(3)检查下一个首部的值，如果为 0，则处理逐跳选项首部。

(4)进行路由选择，使用目的地址字段中的值和本地路由表中的内容进行比较，确定转发接口和下一跳 IPv6 地址。如果没有找到路由，则向源地址发送“目的不可达”ICMPv6 报文，丢弃该分组。
(5)处理有效载荷长度字段，如果转发接口链路的 MTU 小于有效载荷长度字段值加上 40 之和，则向源地址发送“包过大”ICMPv6 报文，丢弃该分组。
(6)根据路由选择结果转发分组。

7.1.6 从 IPv4 向 IPv6 过渡

（1）双协议栈

（4）网络协议转换

设置类似于 NAT 的转换设备，实现 IPv4 和 IPv6 之间的转换

7.1.7 移动 IPv4 P232－黄

移动 IP 就是指在 IP 网络上的多个区域均可使用同一 IP 地址的技术。一个移动节点可以在**不改变其 IP 地址**的情况下改变其驻留位置。

移动 IP 中的功能实体包括：

（1）**移动节点：**具有永久 IP 地址的移动节点。

（2）**本地代理（家乡代理）：**连接到移动节点本地网络的路由器，它保存有移动节点的位置信息，当移动节点离开本地网络时，能根据移动用户的**转交地址**，采用隧道技术将发往移动节点的数据包传给移动节点。

（3）**外部代理：**是指移动节点当前的所在的外地网络上的一个路由器，它能够把由本地代理送来的数据包转发给移动节点。

移动 IP 的通信过程大致如下：

（1）家乡代理和外地代理周期性发送组播或广播报文，通告它们的存在，移动节点据此获知外地代理的地址。

（2）移动节点漫游到一个外地网络时，仍然使用固定的 IP 地址进行通信。为了能够收到通信对端发给它的 IP 分组，移动节点需要向外地代理请求外地代理信息，外地代理会返回给移动节点一个转交地址。

（3）移动节点利用转交地址向家乡代理注册当前的位置地址。

（4）家乡代理接收来自转交地址的注册、认证后，会构建一条通向转交地址的隧道，家乡代理将截获的发给移动节点的 IP 分组通过隧道送到转交地址处。

（5）在转交地址处解除隧道封装，恢复出原始的 IP 分组，最后送到移动节点，这样移动节点在外网就能够收到这些发送给它的 IP 分组。

（6）移动节点在外网通过外网路由器或者外部代理向通信对端发送 IP 数据包，此时，通信对端地址作为目的地址，转交地址作为源地址。

（7）当移动节点来到另一个外网时，只需要向本地代理更新注册的转交地址，就可以继续通信。

（8）当移动节点回到本地网时，移动节点向本地代理注销转交地址，这时移动节点又将使用传统的 TCP/IP 方式进行通信。

移动节点的通信过程

作业 1：