# 课程 3

# IP 协议

# 目 录

课程说明 . . . . . 1
课程介绍 . . . . . 1
课程目标 . . . . . 1
相关资料 . . . . . 1
第一节 序言 . . . . . 1
1.1 Internet 的互连网协议IP . . . . . 2
第二节 IP地址及其转换 . . . . . 3
2.1 IP地址的表示方法 . . . . . 3
2.2 子网的划分 . . . . . 5
2.3 地址的转换 . . . . . 7
小 结 . . . . . 10
习 题 . . . . . 10
第三节 IP数据报的格式 . . . . . 11
3.1 IP数据报首部的固定部分 . . . . . 11
3.2 IP 首部的可变部分 . . . . . 14
小 结 . . . . . 16
习 题 . . . . . 16
第四节 路由段与路由表 . . . . . 17
4.1 IP 地址与物理地址 . . . . . 18
4.2 通过路由表进行选路 . . . . . 19
小 结 . . . . . 21
习 题 . . . . . 21
第五节 Internet控制报文协议ICMP . . . . . 22
小 结 . . . . . 25
习 题 . . . . . 25
习题答案 . . . . . 26
缩略词表 . . . . . 27

# 课程说明

## 课程介绍

本课程介绍 Internet 协议中 IP 协议的相关概念原理。主要包括IP地址及其转换，IP数据报的格式，路由技术及ICMP差错控制报文等内容。

## 课程目标

完成本课程学习，学员能够掌握：

- ✓ IP 协议的功能
- ✓ IP 地址及IP数据报的格式
- ✓ IP协议中的路由技术

## 相关资料

《TCP/IP Illstrated, Volume 1 - The Protocols》

《TCP/IP Illstrated, Volume 1 - The Implementation》

# 第一节 序言

## 1.1 Internet 的互连网协议 - IP

全球INTERNET网的广泛应用使IP 协议深入人心。IP 协议以其简单、有效、开放性成为事实上的工业标准。IP 协议使异种网互联方便可行，尤其值得一提的是它对下层通信技术的巨大包容性。

IP 协议作为通信子网的最高层，提供无连接的数据报传输机制。IP协议是点到点的，核心问题是寻径。它向上层提供统一的IP数据报，使得各种物理帧的差异性对上层协议不复存在。

互连网协议IP是TCP/IP体系中两个最重要的协议之一。与IP 协议配套使用的还有三个协议：

地址转换协议ARP（Address Resolution Protocol）

反向地址转换协议RARP（Reverse Address Resolution Protocol）

Internet控制报文协议ICMP（Internet Control Message Protcol）

图1-1 画出了这三个协议和IP协议的关系。在这一层中，ARP和RARP画在最下面，因为IP经常要使用着两个协议。ICMP画在这一层的上部，因为它要使用IP协议。这三个协议将在后面陆续介绍。

顺便指出，有时会听到一种不准确的说法：“我们用TCP/IP协议进行网络互连”。我们要请读者注意，TCP是与互连网协议IP配套使用的一个运输协议。TCP相当与OSI 的运输层协议而不是一个互连网协议。因此TCP和网络互连并没有直接的关系，只不过是TCP与IP经常配合起来使用而已。如图1所示：

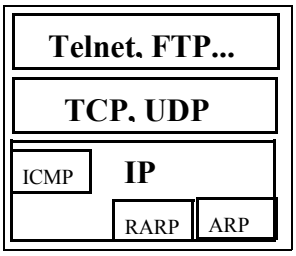



图1IP及配套协议

# 第二节 IP地址及其转换

在TCP/IP体系中，IP地址是一个很重要的概念。一定要把它弄清楚。

## 2.1 IP地址的表示方法

我们把 Internet 看成为一个网络。所谓IP地址就是给每一个连接在Internet上的主机分配一个唯一的32bit 地址。IP地址的结构使我们可以Internet上很方便地进行寻址，这就是：先按IP地址中的网络号码 net-id 把网络找到，再按主机号码 host-id 把主机找到。所以IP地址并不只是一个计算机的号码，而是指出了连接到某个网络上的某个计算机。IP地址有美国国防数据网DDN的网络信息中心NIC进行分配。

为了便于对IP地址进行管理，同时还考虑到网络的差异很大，有的网络拥有很多的主机，而有的网络上的主机则很少。因此Internet 的IP地址就分成为五类，即A类到E类。这样，IP地址（图2）由三个字段组成，即：

类别字段（又称为类别比特），用来区分IP地址的类型；

网络号码字段net-id;

主机号码字段host-id。

D类地址是一种组播地址，主要是留给Internet体系结构委员会IAB(Internet Architecture Board)使用。E类地址保留在今后使用。目前大量IP地址仅A至C类三种。

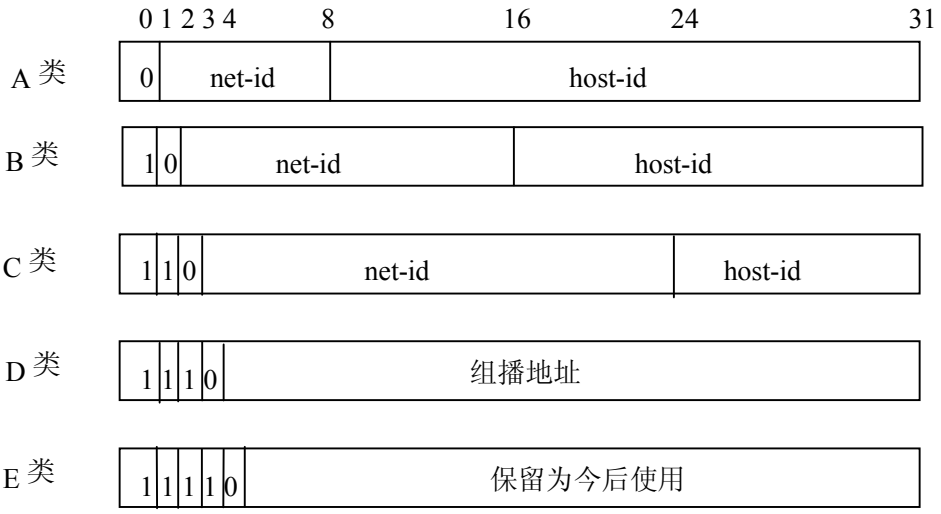



图 2 IP 地址的五种类型

net-id—网络号码，host-id—主机号码

A类IP地址的网络号码数不多。目前几乎没有多余的可供分配。现在能够申请到的IP地址只有B类和C类两种。当某个单位向IAB申请到IP地址时，实际上只是拿到了一个网络号码net-id。具体的各个主机号码host-id则由该单位自行分配，只要做到在该单位管辖的范围内无重复的主机号码即可。

为方便起见，一般将32bit的IP地址中的每8个比特用它的等效十进制数字表示，并且在这些数字之间加上一个点。例如，有下面这样的IP地址：

10000000 00001011 00000011 00011111

这是一个B类IP地址，可记为128.11.3.31，这显然更方便得多。

在使用IP地址时，还要知道下列地址是保留作为特殊用途的，一般不使用。

- 全0的网络号码，这表示“本网络”或“我不知道号码的这个网络”。
- 全1的网络号码。
- 全0的主机号码，这表示该IP地址就是网络的地址。
- 全1的主机号码，表示广播地址，即对该网络上所有的主机进行广播。
- 全0的IP地址，即0.0.0.0。
- 网络号码为127.X.X.X.，这里X.X.X为任何数。这样的网络号码用作本地软件回送测试（Loopback test）之用。

- 全1地址255.255.255.255，这表示“向我的网络上的所有主机广播”。原先是使用0.0.0.0。

这样，我们就可得出表1所示的IP地址的使用范围。

表1 IP地址的使用范围

| 网络<br>类别 | 最大<br>网络数 | 第一个可用的<br>网络号码 | 最后一个可用<br>的<br>网络号码 | 每个网络中的<br>最大主机数 |
|---|---|---|---|---|
| A | 126 | 1 | 126 | 16.777.214 |
| B | 16.382 | 128.1 | 191.254 | 65.534 |
| C | 2.097.150 | 12.0.1 | 223.255.254 | 254 |

IP地址有一些重要的特点：

1. IP地址有一些是一种非等级的地址结构。这就是说，和电话号码的结构不一样，IP地址不能反映任何有关主机位置的地理信息。

2. 当一个主机同时连接到两个网络上时（作路由器用的主机即为这种情况），该主机就必须同时具有两个相应的IP地址，其网络号码net-id是不同的，这种主机成为多地址主机（multihomed host）.

3. 按照Internet的观点，用转发器或网桥连接起来的若干个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码net-id.

4. 在IP地址中，所有分配到网络号码net-id的网络（不管是小的局域网还是很大的广域网）都是平等的。

图3 画出了用路由器（用有R字的圆圈符号表示）和网桥（用有B字方框符号表示）连接起来的一个互连网。图中的小圆圈表示需要有一个不同的IP地址。可以看出，一个计算机若要和网络号码不同的计算机通信，就必须经过路由器。

## 2.2 子网的划分

IP地址的设计有不够合理的地方。例如，IP地址中的A至C类地址，可供分配的网络号码超过211万个，而这些网络上的主机号码的总数则超过37.2亿个，初看起来，似乎IP地址足够全世界来使用，（在70年代初期设计IP地址是就是这样认为的）。其实不然。第一，当初没有预计到微机会普及得如此之快。各种局域网和局域网上的主机数目急剧增长。第二，IP地址在使用时有很大的浪费。例如：某个单位申请到了一个B类地址。但该单位只有1万台主机。

于是，在一个B类地址中的其余5万5千多个主机号码就白白地浪费了。因为其他单位的主机无法使用这些号码。如图3所示：

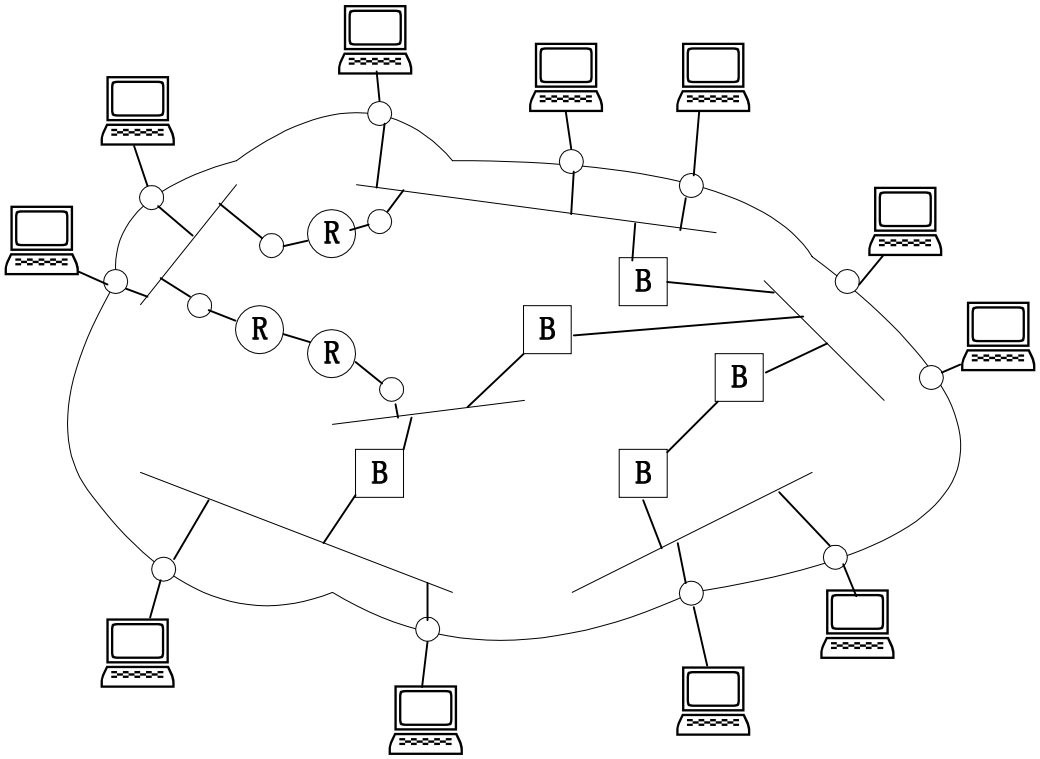



图3 什么地方需要一个IP地址

因此，目前正在研究如何将IP地址加以扩展[NETW93]，但这非常复杂。因为IP地址一旦改变，在各种主机上运行的大量软件就必须修改。这是一件耗费大量人力和财力的工作。有人也提出采用OSI的20个字节的网络层地址方案。读者应注意这一问题。

从1985年起，为了使IP地址的使用更加灵活，在IP地址的网络号码net-id，而后面的主机号码host-id则是受本单位控制，由本单位进行分配。本单位所有的主机都使用同一个网络号码。当一个单位的主机很多而且分布在很大的地理范围是，往往需要用一些网桥（而不是路由器，因为路由器连接的主机具有不同的网络号码）将这些主机互连起来。网桥的缺点较多。例如容易引起广播风暴，同时当网络出现故障时也不太容易隔离和管理。为了使本单位的各子网之间使用路由器来互连，因而便于管理。需要注意的是，子网的划分纯属本单位内部的是，在本单位以外是看不见这样的划分。从外部看，这个单位只有一个网络号码。只有当外面的分组进入到本单位范围后，本单位的路由器在根据子网号码进行选路，最后找到目的主机。若本单位按照主机所在的地理位置划分子网，那么在管理方面就会方便得多。

这里应注意，TCP/IP体系的“子网”（subnet）是本单位网络内的一个更小些的网络，和前面讲的OSI 体系中的子网（subnetwork）不同。它们的英文名字不同，但中文译名都是一样的。

图4 说明是在划分子网时要用到的子网掩码（subnet mask）的意义。图4（a）举了一个B类IP地址作为例子。图4（b）表示将本地控制部分再增加一个子网字段，子网号字段究竟选为多长，由本单位根据情况确定。TCP/IP体系规定用一个32bit的子网掩码来表示子网号字段的长度。具体的做法是：子网掩码由一连串的“1”和一连串的“0”组成。“1”对应于网络号码和子网号码字段，而“0”对应于主机号码字段（ 图4（c））

多划分出一个子网号码字段是要付出代价的。例如，对于图4的例子，本来一个B类IP地址可以容纳65534个主机号码。但划分出6bit长的子网字段后，最多可有62个子网（去掉全1和全0的子网号码）。每个子网有10bit的主机号码，即每个子网最多可有1022个主机号码。因此主机号码的总数是62*1022=63364个。比不划分子网时要少了一些。

若一个单位不进行子网的划分，则其子网掩码即为默认值，此时子网掩码中“1”的长度就是网络号码的长度。因此，对于A，B和C类IP地址，其对应的子网掩码默认值分别为255.0.0.0,255.255.0.0和255.255.255.0。

## 2.3 地址的转换

上面讲的IP地址还不能直接用来进行通信。这是因为：

1. IP地址中的主机地址只是主机在网络层中的地址，相当与前面讲过的NSAP。若要将网络层中传送的数据报交给目的主机，必须知道该主机的物理地址。因此必须在IP地址和主机的物理地址之间进行转换。
2. 用户平时不愿意使用难于记忆的主机号码，而是愿意使用易于记忆的 主机名字。因此也需要在主机名字和IP地址之间进行转换。

在TCP/IP体系中都有这两种转换的机制。

对于较小的网络，可以使用TCP/IP体系提供的叫做 hosts 的文件来进行从主机名字到IP地址的转换。文件hosts上有许多主机名字到IP地址的映射，供主叫主机使用。

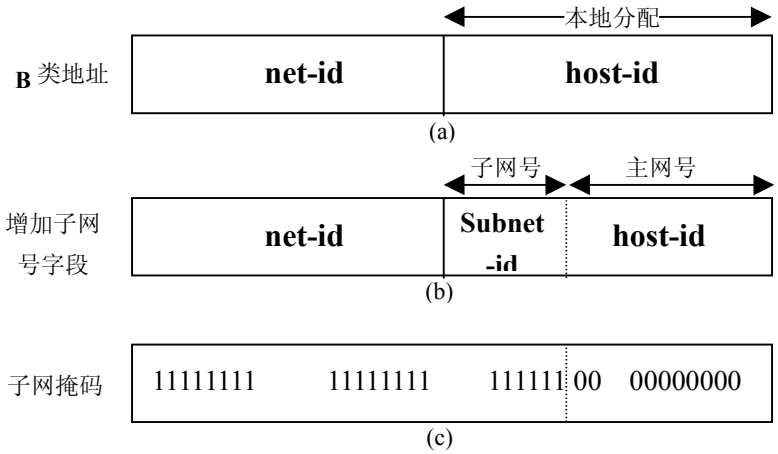



图 4 子网掩码的意义

对于较大的网络，则在网络中的几个地方放有域名系统DNS（Domain Name System）的名字服务器nameserver, 上面分层次放有许多主机名字到IP地址转换的映射表。主叫主机中的名字转换软件resolver 自动找到DNS的nameserver来完成这种转换。域名系统DNS属于应用层软件。

图5 中设名字为host-a的主机要与名字为host-b的主机通信，通过DNS从目的主机host-b得出其IP地址为209.0.0.6。

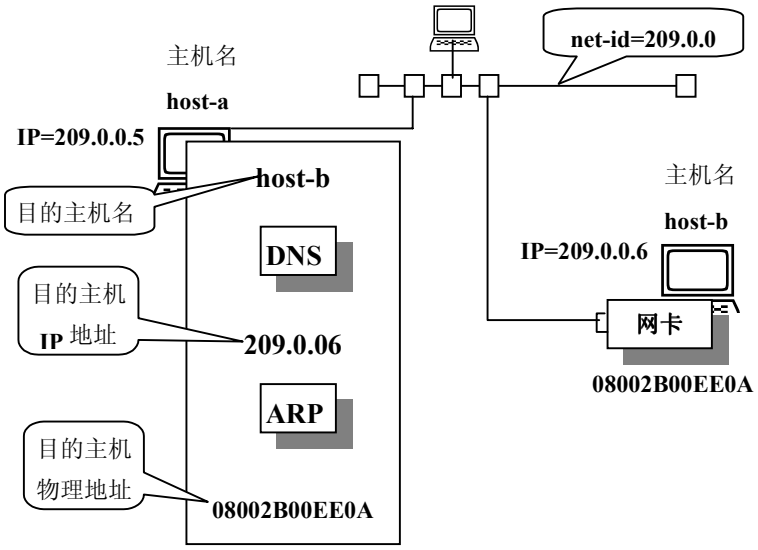



图 5 主机名字、主机物理地址与 IP 地址的转换

IP地址到物理地址的转换由地址转换协议ARP来完成。图5还表示出从IP地址209.0.0.6通过ARP得出了目的主机48bit的物理地址 08002B00EE0A（现在假

设此主机连接在某个局域网上。如网络是广域网，则转换出主机在广域网上的物理地址）。

由于IP地址有32bit,而局域网的物理地址（即MAC地址）是48bit，因此它们之间不是一个简单的转换关系。此外，在一个网络上可能经常会有新的计算机假如近来，或撤走一些计算机。更换计算机的网卡也会使其物理地址改变。可见在计算机中应当存放一个从IP地址到物理地址的转换表，并且能够经常动态更新。地址转换协议ARP很好地解决了这些问题。

每一个主机都有一个ARP高速缓存（ARP cache），里面有IP地址到物理地址的映射表，这些都是该主机目前知道的一些地址。当主机A 欲向本局域网上的主机B发送一个IP 数据报时，就先在其ARP高速缓存中查看有无主机B的IP地址。如有，就可查出其对应的物理地址，然后将该数据报发往此物理地址。

也有可能查不到主机B的IP地址的项目。这可能是主机B才入网，也可能是主机A 刚刚加电，其高速缓存还是空的。在这种情况下，主机A就自动运行ARP，按以下步骤找出主机B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，上面有主机B的IP地址。
2. 在本局域网上的所有主机上运行的ARP进程都收到此ARP请求分组。
3. 主机B在ARP请求分住中见到自己的IP地址，就向主机A发送一个ARP响应分组，上面写入自己的物理映射。
4. 主机A收到主机B的ARP响应分组后，就在其ARP高速缓存中写入主机B的IP地址到物理地址的映射。

在很多情况下，当主机A向主机B发送数据报时，很可能以后不久主机B 还要向主机A 发送数据报，因而主机B也可能要向主机A发送ARP请求分组。为了减少网络上的通信量，主机A在发送其ARP请求分组时，就将自己的IP地址到物理地址的映射写入ARP 请求分组。当主机B收到主机A的ARP请求分组时，主机B就将主机A的这一地址映射写入主机B自己的ARP高速缓存中。这对主机B以后向主机A发送数据报时就更方便了。

在进行地址转换时，有时还要用到反向地址转换协议RARP。RARP使只知道自己物理地址的主机能够知道其IP地址。这种主机往往是无盘工作站。这种无盘工作站一般只要运行其ROM中的文件传送代码，就可用下行装载方

法，从局域网上其他主机得到所需的操作系统和TCP/IP通信软件，但这些软件中并没有IP地址。无盘工作站要运行ROM中的RARP来获得其IP地址。

RARP的工作过程大致如下。

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，无盘工作站先向局域网发出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。

RARP服务器有一个事先做好的从无盘工作站的物理地址到IP地址的映射表，当收到RARP请求分组后，RARP服务器就从这映射表查出该无盘工作站的IP地址。然后写入RARP响音分组，发回给无盘工作站。无盘工作站用这样的方法获得自己的IP地址。

## 小 结

IP地址就是给每一个连接在Internet上的主机分配一个唯一的32bit 地址。IP地址的结构使我们可以Internet上很方便地进行寻址。 Internet 的IP地址就分成为五类。为了提高IP地址的利用率，引进了子网的概念。 IP地址到物理地址的转换由地址转换协议ARP来完成。

## 习 题

2-1 A，B，C类IP地址各自的地址范围是多少？

2-2 为什么要划分子网？

2-3 简述ARP的工作流程？

# 第三节 IP数据报的格式

在TCP/IP的标准中，各种数据格式常常以32bit（即4字节）为单位来描述。图6是IP数据报的格式。

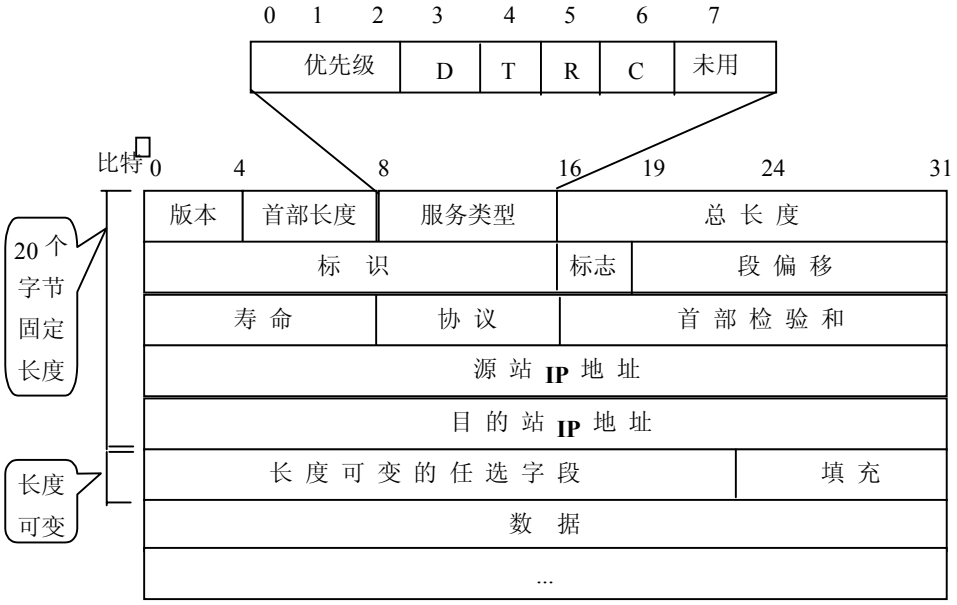



图 6 IP 数据报的格式

从图6可看出，一个IP数据报由首部和数据两部分组成。首部的前一部分长度是固定的20个字节，后面部分的长度则是可变长度。下面介绍首部各字段的意义。

## 3.1 IP数据报首部的固定部分

1、版本

版本字段站4 bit,指IP协议的版本。通信双方使用的IP协议的版本必须一致。目前使用的IP协议版本为4。

2、首部长度

首部长度字段占4bit，可表示的最大数值是15个单位（一个单位为4字节），因此IP的首部长度的最大值是60字节。当IP分组的首部长度不是4字节的整数倍时，必须利用最后一个填充字段加以填充。这样，数据部分永远在4字节的整数倍时开始，这样在实现起来会比较方便。首部长度限制为60字节的

缺点是有时（如采用源站选路时）不够用。但这样做的用意是要用户尽量减少额外的开销。

3、服务类型

服务类型字段共8bit长，用来获得更好的服务，其意义见图6的上面部分所示。

服务类型字段的前三个比特表示优先级，它可使数据报具有8个优先级中的一个。

第4个比特是D比特，表示要求有更低的时延。第5个比特是T比特，表示要求有更高的吞吐量。第6个比特是R比特，表示要求有更高的可靠性，即在数据报传的过程中，被结点交换机丢弃的概率要更小些。第7个比特是C比特，是新增加的，表示要求选择价格更低廉的路由。最后一个比特目前尚未使用。

4、总长度

总长度指首部和数据之和的长度，单位为字节。总长度字段为16bit，因此数据报的最大长度为65535字节。这在当前是够用的。

当很长的数据报要分段进行传送时，“总长度”不是指未分段前的数据报长度，而是指分段后每个段的首部长度与数据长度的总和。

5、标识

标识字段的意义和OSI的IPDU中的数据单元标识符的意义一样，是为了使分段后的各数据报段最后能准确地重装成为原来的数据报。请注意：这里的“标识”并没有顺序号的意思，因为IP是无连接服务，数据报不存在按序接收的问题。

6、标志

标志字段占3bit。目前只有前两个比特有意义。

标志字段中的最低位记为MF（More Fragment）。MF=1即表示后面还有分段的数据报。MF=0表示这已是若干数据报段中的最后一个。

标志字段中间的一位记为DF（Don’t Fragment）。只有当DF=0时才允许分段。

7、段偏移

段偏移字段的意义和OSI的IPDU中规定的相似，只是表示的单位不同。这里是以8个字节为偏移单位。可见IP数据报的段偏移字段（13bit长）和OSI的IPDU的段偏移字段（16bit长）是相当的。

8、寿命

寿命字段记为TTL（Time To Live）,其单位为秒。寿命的建议值是32秒。但也可设定为3-4秒，或甚至255秒。

9、协议

协议字段占8bit，它指出此数据携带的运输层数据是使用何种协议，以便目的主机的IP层知道应将此数据报上交给哪个进程。常用的一些协议和响应的协议字段值（写在协议后面的括弧中）是：UDP（17），TCP（6），ICMP（1），GGP（3），EGP（8），IGP（9），OSPF（89），以及ISO的TP4（29）。

10、首部检验和

此字段只检验数据报的首部，不包括数据部分。不见眼数据部分是因为数据报每经过一个结点，结点处理机就要重新计算一下首部检验和（一些字段，如寿命、标志、段偏移等都可能发生变化）。如将数据部分一起检验，计算的工作量就太大了。

11、地址

源站IP地址字段和目的站IP地址字段都各占4字节。

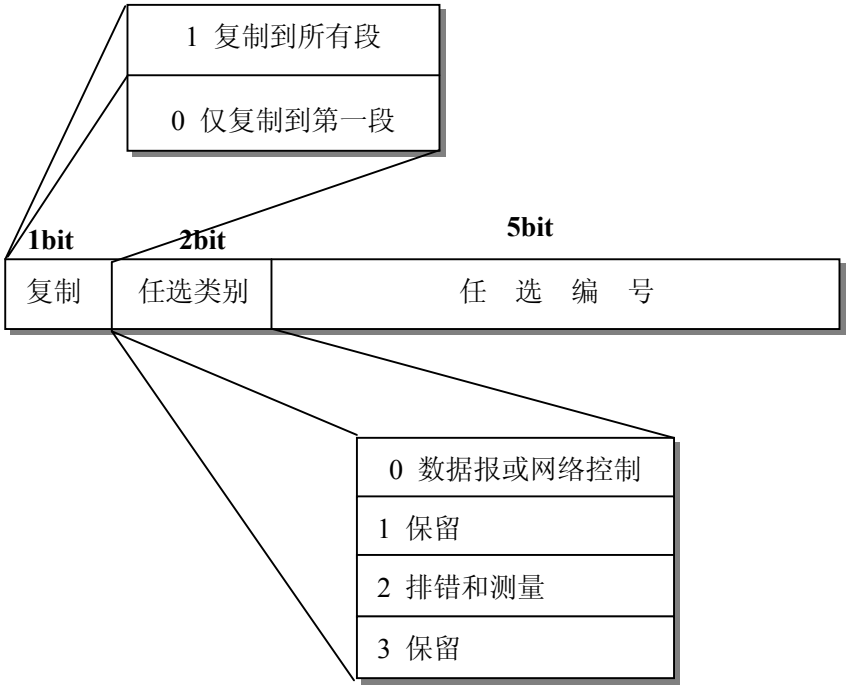



图7 Option 选项

## 3.2 IP 首部的可变部分

IP 首部的可变部分就是一个任选字段。任选字段用来支持排错、测量以及安全等措施，内容很丰富。此字段的长度可变，从一个字节到40个字节不等，取决于所选择的项目。某些任选项目只需要一个字节，它只包括一个字节的任选代码，图7画的是任选代码的格式。还有些任选项目需要多个字节，但其第一个字节的格式仍为图7所示的那样。这些任选项一个个拼接起来，中间不需要有分隔符，最后用全0的填充字段补齐成为4字节的整数倍。

可以看出，任选代码共有三个字段。

第一个字段是复制字段，占1bit，它的作用是控制网络中的路由器在将数据报进行分段时所作的选择。当复制字段为1时，必须将此任选字段复制到每一个数据报段。而当复制字段为0时，就只复制到第一个数据报段上。

第二个字段是任选类别字段，占2bit。但目前只有两种可供选用，如图8：

| 任选类别 | 意　义 |
|---|---|
| 0 | 数据报或网络控制(主要是这一类) |
| 1 | 保留今后使用 |
| 2 | 排错和测量,即 Internet 时间戳 |
| 3 | 保留今后使用 |

图 8 任选类别及意义

第三个字段是任选编号，占5个字节，它指出任选是做什么用的。

属于任选类别0的有下列一些任选编号：

任选编号为0：指出这是任选项目中的最后一个。

任选编号为1：无操作，用于需要按每4个字节对齐之用。和填充字段的功能是一样的。

以上两种都是只使用一个字节的任选代码。下面的几种则要使用若干个字节。

任选编号为2：为安全用的。只用在美国国防系统来传送机密文件。路由器在检测到这一安全任选项目时，就要使该数据报不要离开安全的环境。在商业上尚无此应用。

任选编号为7：为记录路由用的，其长度是可变的。图9是记录路由的任选项目的格式。

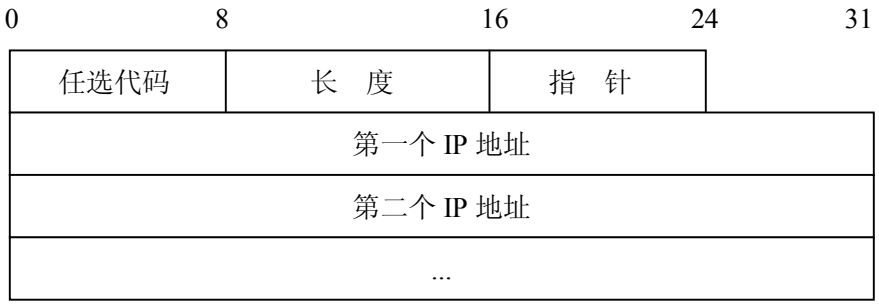



图 9 记录路由的任选项目的格式

这种数据报是用来监视和控制互连网中的路由器是如何转发数据报的。源站发出一个空白的表，让数据报所经过的个路由填上其IP地址，以获得路由信息。

前三个字节是：

1. 任选代码字段——其中的三个字段分别填入0，0，和7。

2. 长度字段——填入此任选项目的长度，包括这前三个字节。

3. 指针字段——指出下一个可填入IP地址的空白位置的偏移量。

在这之后，就是若干个4字节长的IP地址，让各个路由器填入。当一个路由器收到包含有记录路由任选项目的数据报时，先检查指针所指的位置是否超过了表的长度。如不超过，则填入自己的IP地址，并将指针值加4，然后转发出去。但如表已填满，则不填入自己的IP地址，而仅仅转发此数据报。

一般的计算机在受到这样的数据报是，并不会理睬该数据报中所记录的路由。因此，源站必须和有关的站主机在、协商好，请目的主机在收到记录的路由信息后，将路由信息提取出来，并发回源站。

下面两任选项目都是关于源站选路的。

- 任选编号为3：不严格的源站选路(loose source routing)，其长度是可变的。
- 任选编号为9：严格的源站选路(strict source routing)，其长度也是可变的。

源站选路本来是源站将数据报传送的路由事先规定好。严格的源站选路不允许改变源站规定好的路由。但不严格的源站选路允许在数据报传送的过程中，将路由表中源站已规定要经过的一些路由器，改换成别的路由器。

源站选路任选项目的格式与图记录路由的相似。前面也是三个固定的字节，但任选代码字节中的三个字段应分别填入1，0和3（不严格的源站选路）以及1，0和9（严格源站选路）。此外，这三个字节后的IP地址表不是空的，而是事先由源站写好的。数据报按源站指定的路由传送。当路由器收到此数据报后，若指针已超过表的范围，则转发此数据报，不写任何数据。若指针的指示是正确的，则填入自己的IP地址(覆盖掉原来的IP地址)，并按照表中指出的一下一个地址转发出去。这里要注意：一个路由器有两个或两个以上IP地址。原来在这个任选项目路由表中写入的是路由器的入口IP地址，而路由器写的IP地址则是路由器的出口IP地址。

在数据报中加入源站选路任选项目，可以使网络的管理者了解沿网络中的某一条通路的通信状况是否正常。一般的用户并不使用这一功能。

最后一个任选项目是Internet 的时间戳。

- 任选编号为4：作时间戳用，其长度是可变的。格式和图类似，但一开始除了原来的任选代码字段(填入0，2和4)、长度字段和指针字段这三个字节外，再加上一个字节的溢出和标志两个字段。标志字段区分几种情况：(1)只写入时间戳；（2）写入IP地址和时间戳；（3）IP地址由源站规定好，路由器只写入时间戳。溢出字段写入一个数，此数值即数据报所经过的路由器的最大数目（考虑到太多的时间戳可能会写不下）。

时间戳记录了路由器收到数据报的日期和时间，占用了4个字节。时间的单位是毫秒，是从午夜算起的通用时间(Universal Timer)，也就是以前的格林尼治时间。当网络中的主机的本地时间和时钟不一致时，记录的时间戳会有一些误差。时间戳可用来统计数据报经路由器产生的时延和时延的变化。

## 小 结

一个IP数据报由首部和数据两部分组成。首部的前一部分长度是固定的20个字节，后面部分的长度则是可变长度。 IP 首部的可变部分就是一个任选字段。任选字段用来支持排错、测量以及安全等措施。

## 习 题

3-1 IP 数据头中那些域与数据报分组相关？

3-2 请简述源路径选项？

# 第四节 路由段与路由表

在互连网中进行路由选择要使用路由器，它平等地看待每一个网络。不论是较大的广域网还是较小的局域网，在路由器看来都只是一个网络。因此在图中将每一个网络画成为一片云，表示路由器产不知道在每一个网络中一个分组是如何选择具体的路由。路由器只是根据所收到的数据报上的目的主机地址选择一个合适的路由器(通过某一个网络)，将数据报传送到下一个路由器。通路上最后的路由器负责将数据报送交目的主机。

路由器将分组在某一个网络中走过的通路（从进入网络算起到离开网络为止），从逻辑上看成是一个路由单位，并将此路由单位称为一个路由段(hop)，或简称为段。例如，在图10中，主机A到主机C共经过了3个网络和2个路由器，因此共经过3个路由器，布从主机A到主机B 则经过了5个网络和4个路由器，即经过5个路由段。由此可见，若一结点通过一个网络与另一结点相连接，则此二结点相隔一个路由段，因而在互连网中是相邻的。同理，相邻的路由器是指这两个路由器都连接在同一个网络上。一个路由器到本网络中的某个主机制路由段数算作零。在图中用粗的箭头表示这些路由段。至于每一具体路由段又由哪几条链路构成，路由器并不关心。

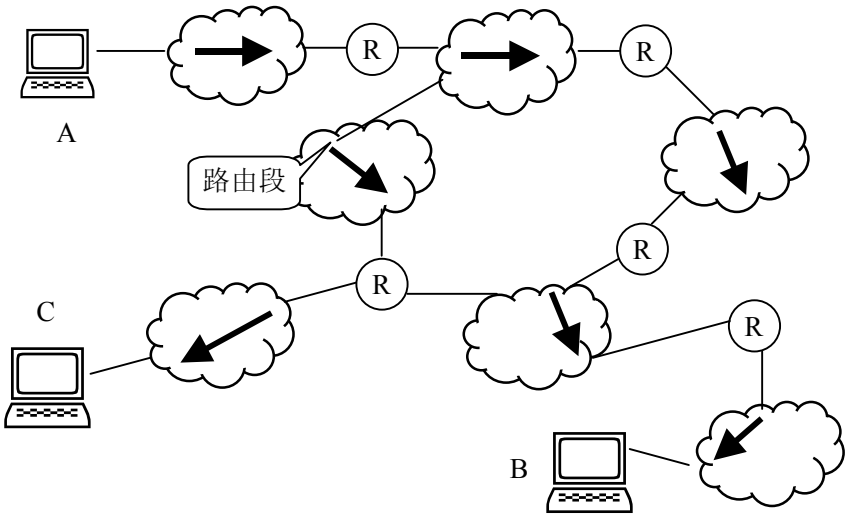



图10 路由段的概念

在互连网的情况下，只能计算各条通路所包含的路由段数。由于网络大小可能相差很大，而每个路由段的实际长度并不相同。因此对不同的网络，可以将其路由段乘以一个加权系数，用加权后的路由段数来衡量通路的长短。

因此，如果把互连网中的路由器看成是网络中的结点，把互连网中的一个路由段看成是网络中的一条链路，那么互连网中的路由选择就与简单网络中的路由选择相似了。

采用路由段数最小的路由有时也产不一定是理想的。例如，经过三个局域网路由段的路由可能比经过两个广域网络路由段的路由快得多。

## 4.1 IP 地址与物理地址

下面通过一个最简单的例子IP地址和物理地址在选路过程中的作用。

设主机A 要向主机B 发送一个数据报。两个主机分别连接在两个网络上，这两个网络通过一个路由器相连。

主机A 的IP层收到欲发送的数据报后，就比较目的主机的源主机的网络号码是否相同（这就是从数据报首部的IP地址中抽出网络号码 net-id 部分进行比较）。如相同，则表明这两个主机在同一个网络内，这样就只需要用目的主机的物理地址进行通信。如果不知道目的主机的物理地址，则可向ARP进行查询。但当主机A和B的网络号码不一样时，就表明它们连接在不同的网络上，因此必须将数据报发给路由器进行转发。

源主机从配置中读出路由器的IP地址。然后从ARP得到路由器的物理地址。随后将数据报发送给这个路由器。

这里要强调指出，在数据报的首部写上的源IP地址和目的IP地址是指正在通信的两个主机的IP地址。路由器的IP地址并没有出现在数据报的首部中。当然，路由器的IP地址是很有用的，但它是用来使源主机得知路由器的物理地址。总之，数据报在一个路由段上传送时，要用物理地址才能找到路由器。

图11 是上述概念的示意图。这就是：MAC地址（设物理地址就是局域网的MAC地址）用于主机到路由器之间的通信（即在一个路由段上通信），而

IP地址则用于两个主机之间的通信，并用来决定找哪一个路由器。符号(1)到(8)表示数据报传送的先后顺序。

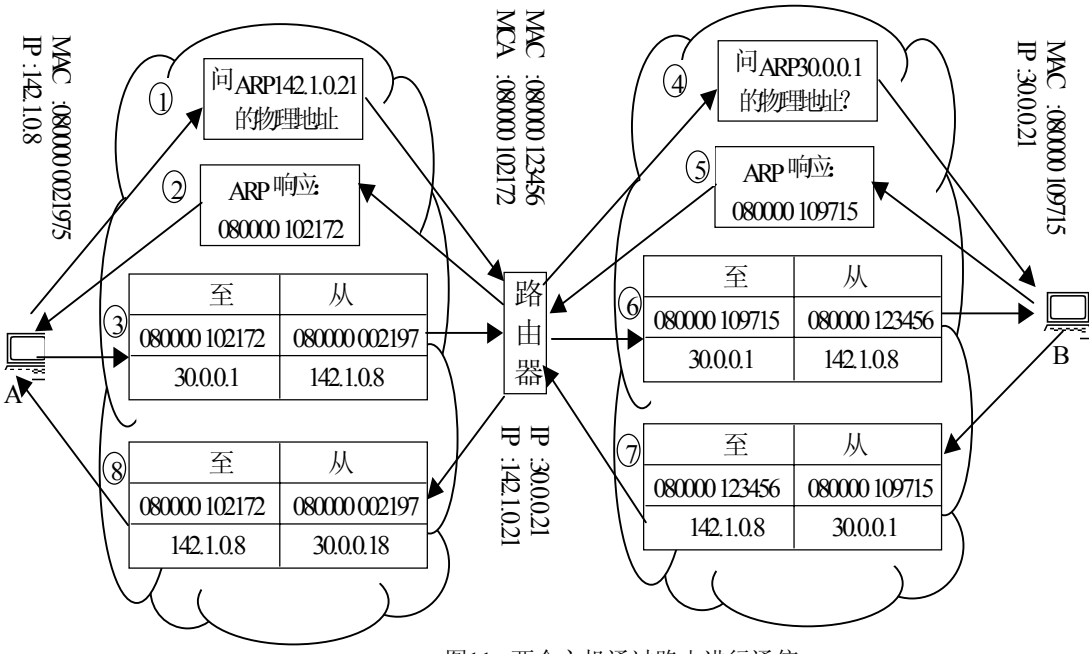



图11 两个主机通过路由进行通信

我们应当注意到，路由器由于连接在两个网络上，因此具有两个IP地址和两个物理地址(MAC地址)。主机A发送的数据报经过路由器后，数据报中的两个IP地址都没有发生变化，但数据帧中的MAC地址（源地址和目的地址）却都改变了。

最后发回来的信息是主机B向主机A的应答（(7)和(8)）。

上面的简单例子只有一个路由器。在更加复杂的例子中，两个通信的主机要经过多个网络和路由器。这时，通信的通路上紧后的路由器负责将数据报交付给目的主机。

## 4.2 通过路由表进行选路

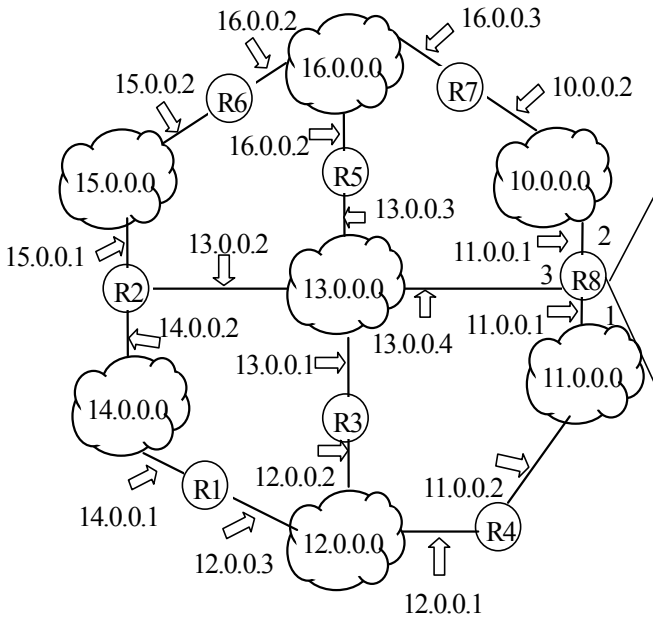
当源主机发送数据报时，IP层先检查目的主机IP地址中的网络号码。如发现与源主机处在同一个网络内，则不经过路由器，只要按照目的主机的物理地址传送即可。

如目的主机不是和源主机在同一个网络中，那么就查一下是否对此特定的目的主机规定了一个特定的路由。如有，则按此路由进行传送。这种情况有时很有用，因为在某些情况下，需要对到达某一个目的主机的特定路由进行性能测试。

如不属于以上情况，则应查找路由表。路由表中写明，找某某网络上的主机，应通过路由器的哪个物理端口，然后就可找到某某路由器（再查找这个路由器的路由表），或者不再经过别的路由器而只要在同一个网络中直接传送这个数据报。

为了不使路由表过于庞大，可以在网络中设置一个默认路由器(default router)。凡遇到在路由表中查不到要找的网络，就将此数据报交给网络中的默认路由器。默认路由器继续负责下一步的选路。这对只用一个路由器与Internet相连的的小网特别方便，因为只要不是发送给本网络的主机的数据报，统统送交给默认路由器。

图12 的例子说明其中一个路由器（路由器80的路由表的主要内容。这里有7个网络通过8个路由器互连在一起。我们应注意到，每一个路由器具有不止一个IP地址。图中各网络中的数字是该网络的网络地址（前面讲过，主机号码为全零的IP地址就是网络地址）。路由器8由于与三个网络相连，因此有三个IP地址和三个物理端口。




路由器 R8 的路由表

| 目的主机所在网络 | 从哪个路由器转发 | 经过哪个端口 |
|---|---|---|
| 10.0.0.0 | 直接 | 2 |
| 11.0.0.0 | 直接 | 1 |
| 12.0.0.0 | 11.0.0.2 | 1 |
| 13.0.0.0 | 直接 | 3 |
| 14.0.0.0 | 13.0.0.2 | 3 |
| 15.0.0.0 | 10.0.0.2 | 2 |
| 16.0.0.0 | 10.0.0.2 | 2 |

图12 路由表举例

各路由表的数据可以是人工输入，也可能通过各种路由选择协议来生成。

## 小 结

在互连网中进行路由选择要使用路由器。选择路由是IP协议的重要内容。 路由器只是根据所收到的数据报上的目的主机地址选择一个合适的路由器(通过某一个网络)，将数据报传送到下一个路由器。通路上最后的路由器负责将数据报送交目的主机。路由选择主要通过路由表进行。

## 习 题

4-1 路由器的IP地址有没有出现在数据报的首部中？ 为什么？

4-2 请简述源主机发送数据报的流程？

# 第五节 Internet控制报文协议ICMP

IP数据报的传送不保证不丢失。但互连网层对数据报的传送还有一定的质量保证功能，这就是使用Internet控制报文协议ICMP(Internet Control Message Protocol)。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。但ICMP不是高层协议，它仍是互连网层中的协议。ICMP报文作为互连网层数据报的数据，加上数据报的首部，组成IP数据报发送出去。ICMP报文的格式如图13所示。

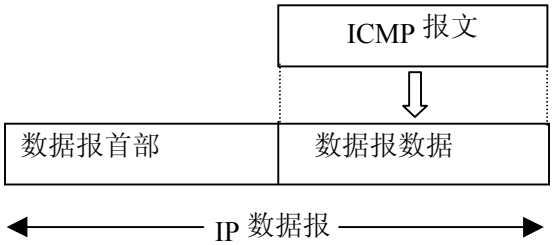



图13 ICMP报文与IP数据报的关系


图14 ICMP报文的格式

ICMP报文的前四个字节是统一的格式，共有三个字段。但后面是和长度可变部分，其长度取决于ICMP的类型。

ICMP报文的类型字段占一个字节。类型字段的值与ICMP报文的类型关系如下：

| 类型字段的值 | ICMP报文的类型 |
|---|---|
| 0 | Ech0(回送)回答 |
| 3 | 目的站不可达 |
| 4 | 源站抑制(Source Quench) |

| | |
|---|---|
| 5 | 改变路由(Redirect) |
| 8 | Echo请求 |
| 11 | 数据报的时间超过 |
| 12 | 数据报的参数有问题 |
| 13 | 时间戳(Timestamp)请求 |
| 14 | 时间戳回答 |
| 17 | 地址掩码(Address Mask)请求 |
| 18 | 地址掩码回答 |

ICMP报文的代码字段也占有一个字节。为的是进一步区分某种类型中的几种不同的情况。后面的检验和占两个字节，它检验整个ICMP报文。数据报首部的检验和不检验数据报的内容，因此不能保证ICMP报文是正确的。

ICMP报文的类型很多，但可分为两种类型，即ICMP差错报文和ICMP询问报文。

在ICMP差错报文中，改变路由报文用得最多。我们以图15为例来说明改变路由报文的用法。

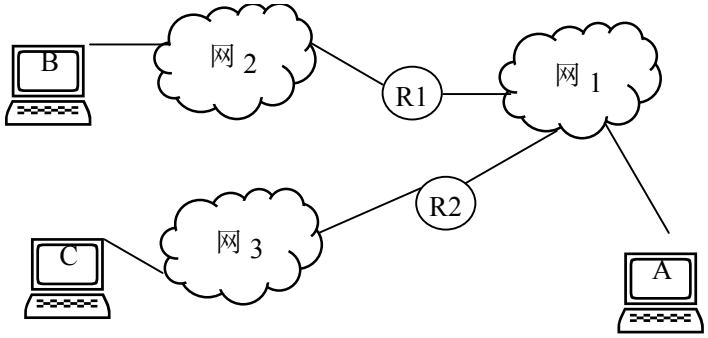



图15 ICMP改变路由报文的使用举例

从图15可看出，主机A向主机B发送IP数据报应经过路由器R1，而向主机C发送数据报则应经过路由R2。现在假定主机A启动后，其路由表中只有一个默认路由器R1。当主机A向主机C发送数据报时，数据报就被送到路由器R1。从路由器R1的路由表可查出：发往主机C的数据报应经过路由器R2。于是数据报从路由器R1再转到路由器R2，最后传到主机C。显然，这个路由不好，应改变。于是，路由器R1向主机A发送一个ICMP改变路由报文，指出此数据报应经过的下一个路由器R2的IP地址。主机A根据收到的信息更新其路由

表。以后主机A再向主 C发送数据报时，根据路由表就知道应将数据报传到路由器R2，而不再传到默认路由器R1了。

图16 是ICMP改变路由报文的格式。在第5-8字节了写入数据报应经过的路由器的IP地址。再后面就是说明是哪一个数据报。数据报的首部都要写入，而数据部分则写入其前8个字节，这里面有运输层数据单元首部的一些数据（端口号），有时要用到。

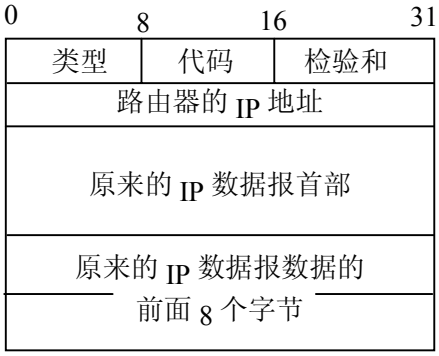



图16 ICMP改变路由报文的格式

当某个速率较高的源主机向另一个速率较慢的目的主机（或路由器）发送一连串的数据报时，就有可能使速率较慢的目的主机产生拥塞，因而不得不丢弃一些数据报。通过高层协议，源主机得知丢失了一些数据报，就不断地重发这些数据报。这就使得本来就已经拥塞的目的主机更加拥塞。在这种情况下，目的主机就要向源主机发送ICMP源站抑制报文，使源站暂停发送数据报，过一段时间再逐渐恢复正常。

下面介绍几个常用的ICMP询问报文。

- ICMP Echo请求报文是由主机或路由器向一个特定的目的主机发出的询问。收到此报文的机器必须给主机发送ICMP Echo回答报文。这种询问报文用来测试目的站是否可达以及了解其有关状态。在应用层有一个服务叫做PING(Packet InterNet Groper)，用来测试两个主机之间的连通性。PING使用了ICMP Echo请求与Echo回答报文。

- ICMP时间戳请求报文是请某个主机或路由器回答当前的日期和时间。在ICMP时间戳回答报文中有一个32bit的字段，其中写入的整数代表从1900年1月1日起到当前时刻一共有多少秒。时间戳请求与回答可用来进行时钟同步和测量时间。

- ICMP地址掩码请求与回答可使主机向子网掩码服务器得到某个接口的地址掩码。

## 小 结

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。但ICMP不是高层协议，它仍是互连网层中的协议。ICMP报文作为互连网层数据报的数据，加上数据报的首部，组成IP数据报发送出去。 ICMP报文的类型很多，但可分为两种类型，即ICMP差错报文和ICMP询问报文。

## 习 题

5-1 为什么ICMP不作为高层协议？

5-2 请简述Ping 的流程？

# 习题答案

2-1 A类：1 - 126

B类：128 - 191

C类：192 -223

2-2 提高IP地址的利用率

2-3 详见第1课3节

3-1 ID, 标志域，Offset

3-2 详见正文

4-1 没有。 IP协议只是设法尽量将数据报向前传递，而不关心之间经过的路由器。

4-2 详见正文

5-1 因为 ICMP只是辅助IP进行控制和差错通知的，并不是作为一个独立的协议层。

5-2 PING使用了ICMP Echo请求与Echo回答报文。

# 缩略词表

| | | |
|---|---|---|
| ARP | Address Resolution Protocol | IP地址 -> 物理地址 |
| RARP | Reverse Address Resolution Protocol | 物理地址 -> IP地址 |
| ICMP | Internet Control Message Protcol | 差错控制协议 |