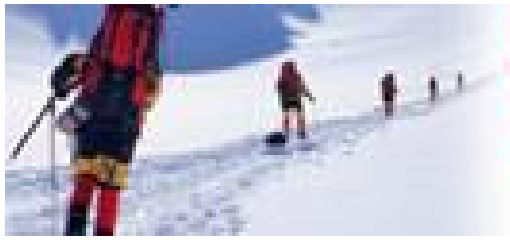


Data Communications and Networking Fourth Edition

Forouzan

第21章

地址映射、差错报告和多播

IP协议需要其他协议的帮助

- IP分组使用逻辑（主机到主机）地址，而分组需要封装成帧，帧需要物理地址（节点到节点），为此而设计地址解析协议ARP；
- 有时还需要逆映射（物理地址到逻辑地址），例如，为引导一个无盘网络或给主机租用IP地址的目的，设计了三个协议：RARP、BOOTP和DHCP；
- IP协议缺少流控制和差错控制，从而产生ICMP协议，提供差错告警，报告在网络上或目的端发生拥塞和差错的类型；
- IP最初为单播传送而设计，即一个源端和一个目的端，由于因特网对多播传送的极大需求，也就是一个源端和多个目的端，因此，IGMP提供了IP的一种多播能力。

21-1 地址映射

p将分组传递到主机或路由器需要两级地址：逻辑地址和物理地址. 我们需要将一个逻辑地址映射成为它对应的物理地址，反过来也一样，这可以通过静态或动态映射完成；

p问题：为什么不直接使用物理地址进行通信？

Ø物理地址在全局上可能不唯一；

Ø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，这几乎是不可能的事；

Ø 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，因为调用地址解析协议（比如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

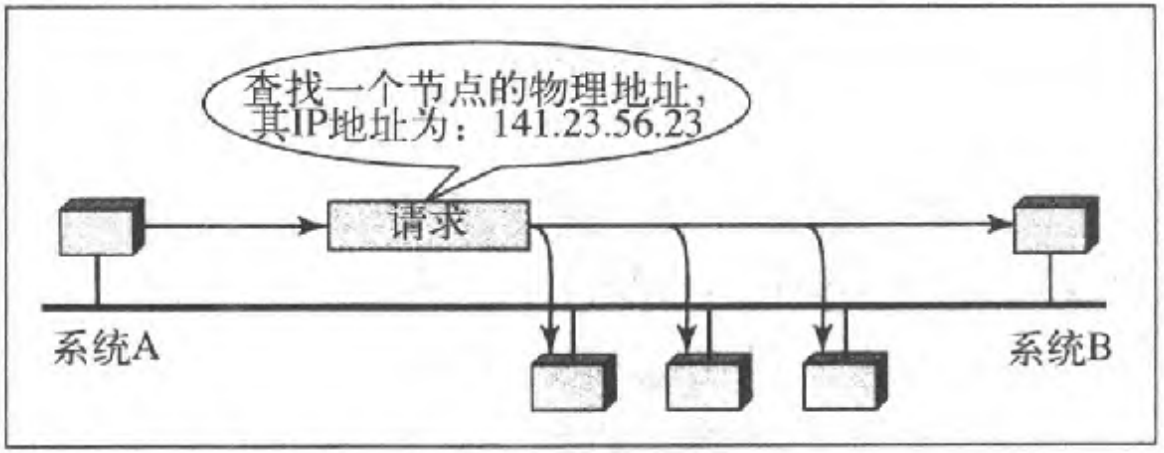
静态和动态映射

- 静态映射是创建一个表，它将一个逻辑地址与物理地址联系起来，这个表存储在网络上的每个机器上；
- 有某些局限性，因为物理地址可能发生变化：
  - Ø1. 一个机器可能会更换网卡，得到一个新的物理地址；
  - Ø2. 在某些局域网中，如LocalTalk，每当计算机加电时，其物理地址都要改变一次；
  - Ø3. 移动的计算机可以从一个物理网络转移到另一个物理网络，这就引起物理地址的改变；静态映射必须周期性地改变，增加了开销
- 所以通常使用动态映射，每当一个机器知道两个地址（逻辑地址和物理地址）中的一个时，就可使用协议将另一个求出。

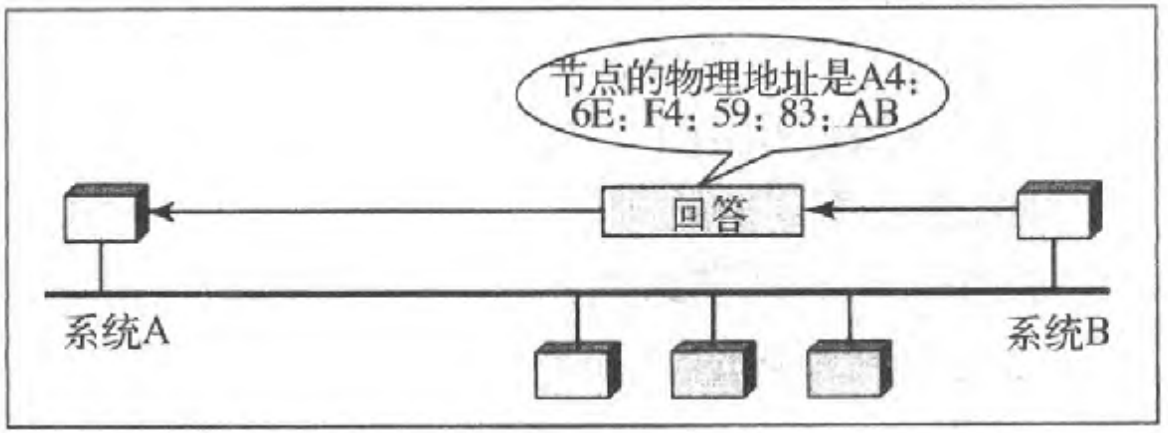
逻辑地址到物理地址的映射：ARP

- 任何时候，当主机或路由器有数据报要发送到另一个主机或路由器时，它必须有接收方的逻辑（IP）地址；如果发送方是主机，它可从DNS求得逻辑（IP）地址；如果发送方是路由器，它可从路由选择表求得；
- 但是，IP数据报必须封装成帧才能通过物理网络，所以发送方必须有接收方的物理地址；
- 主机或路由器发送一个ARP查询分组，该分组包括发送方的物理地址和IP地址以及接收方的IP地址；
- 由于发送方不知道接收方的物理地址，查询就在网络上广播（广播物理地址？）；
- 网络上的每个主机或路由器都接收和处理这个ARP查询分组，但只有预期的接收者才能发回ARP响应分组，这个分组使用接收到的查询分组中的物理地址直接用单播发送给查询者。

图21.1 ARP 操作



a) ARP请求用广播发送



b) ARP回答用单播发送

ARP高速缓存器

- 如果系统A对发送到系统B的每一个分组都需要广播一个ARP请求，那么对ARP协议的使用就是低效的；
- 为此，ARP协议可使用高速续存器，因为一个系统通常发送多个分组到同一目的地；
- 接收到ARP回答的系统将它的映射存储在高速缓存器中，保持20分钟到30分钟（除非高速缓存已满），在以后发送ARP请求之前，系统先在它的高速缓存器中检查是否可找到它的映射（如何查看自己计算机上的缓存表？）。

图21.2 ARP 分组

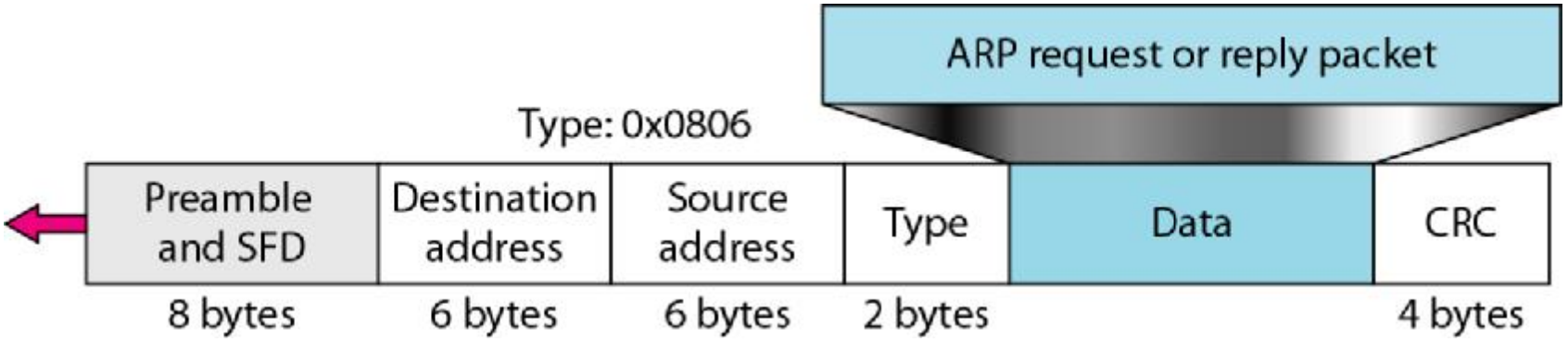

32位
8位
8位
16位
硬件类型
协议类型
硬件长度
协议长度
操作 请求1，回答2
发送方硬件地址
（例如，以太网是6字节）
发送方协议地址
（例如，IP是4字节）
目标硬件地址（例如，以太网是6个字节）
（在请求分组中，它没有被填满）
目标协议地址
（例如，IP是4字节）

ARP分组字段

- 硬件类型：16位，定义运行ARP的网络类型，例如以太网是类型1，ARP可用于任何物理网络；
- 协议类型：16位，定义协议的类型，例如IPv4协议这个字段的值是0x0800，ARP可用于任何高层协议；
- 硬件长度：8位，定义物理地址的字节长度，以太网是6；
- 协议长度：8位，定义逻辑地址的字节长度，IPv4协议是4；
- 操作：16位，定义分组的类型，已定义了两种类型：ARP请求（1）和ARP回答（2）；
- 发送方硬件地址：可变长，定义发送方的物理地址；
- 发送方协议地址：可变长，定义发送方的逻辑地址；
- 目标硬件地址：可变长，定义目标的物理地址，对ARP请求报文，这个字段是全0，因为发送方不知道目标的物理地址；
- 目标协议地址：可变长，定义目标的逻辑地址。

图21.3 ARP 分组的封装

- ARP分组直接封装在数据链路帧中，如图为ARP分组封装在以太网的帧中；
- 类型字段（0x0806）指出了此帧所携带的数据是ARP分组（IP分组类型字段值？RARP分组类型字段值？）



ARP操作步骤

p1.发送方知道目标的IP地址，首先查找本地的ARP缓存表，如果存在对应的条目，使用它封装帧，发送之；否则转2；

p2. IP请求ARP协议产生一个ARP请求报文，填入发送方的物理地址、发送方的IP地址以及目标的IP地址，目标的物理地址字段则填入0；

p3. 将ARP请求报文发送给数据链路层，被封装成帧，使用发送方的物理地址为源地址，而将物理广播地址作为目的地址；

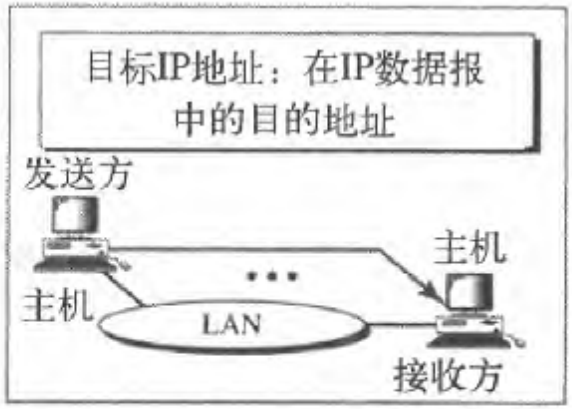
p4. （局域网中）每一个主机和路由器都接收到这个帧，所有站点都将此报文送交ARP；除了目标机器外，所有的机器都丢弃该分组，目标机器识别这个IP地址；

p5. 目标机器用ARP回答报文进行应答，此回答报文包含它的物理地址，报文使用单播发送；

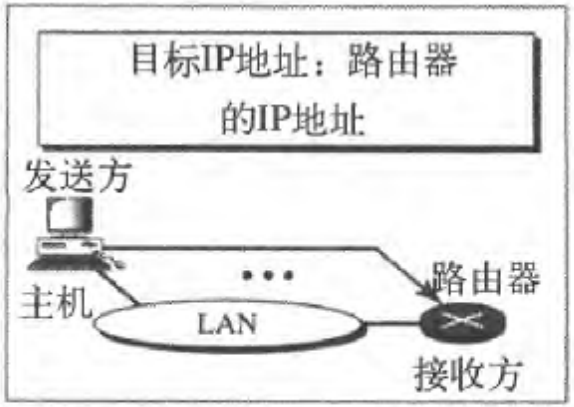
p6. 发送方接收到这个回答报文，得到了目标机器的物理地址；

p7. 携带发送给目标机器数据的IP数据报封装成帧，用单播发送给目的端。

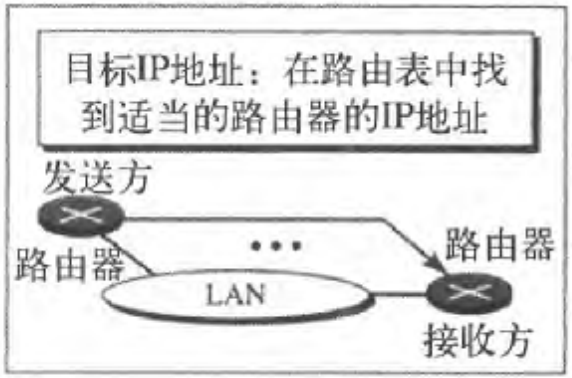
图21.4 使用ARP的四种情况



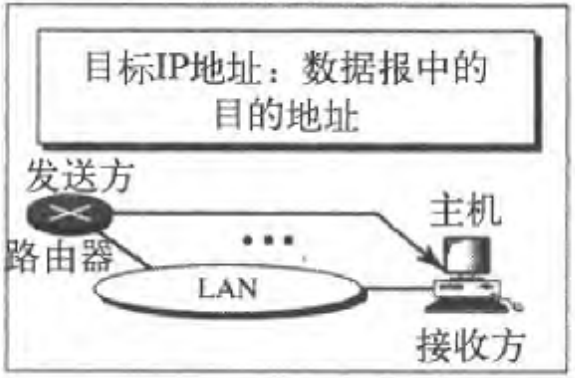
情况1.一个主机有分组要发送给在同一个网络上的另一个主机



情况2.一个主机有分组要发送给另一个网络上的另一主机，这个分组必须先传递给一个路由器



情况3.一个路由器接收到一个分组，要将该分组发送给在另一个网络上的主机，这个分组必须先传递给适当的路由器



情况4.一个路由器接收到一个分组，要将该分组发送给在同一个网络上的主机

ARP请求报文是广播发送；
ARP回答报文是单播发送。

例21.1

一个主机的IP地址为130.23.43.20，物理地址为B2:34:55:10:22:10，它有一个分组想要发送给另一个主机，其IP地址为130.23.43.25，物理地址为A4:6E:F4:59:83:AB（第一个主机并不知道该物理地址）。两个主机在同一个网络上，试说明ARP请求与回答分组如何封装在以太网帧中。

解：

图21.5显示了ARP请求与回答分组。注意：此时ARP数据字段是28个字节，而单个地址不适合用4字节表示界限，这就是我们为什么不以4字节界限表示这些地址

图21.5 例21.1的ARP 请求与回答分组

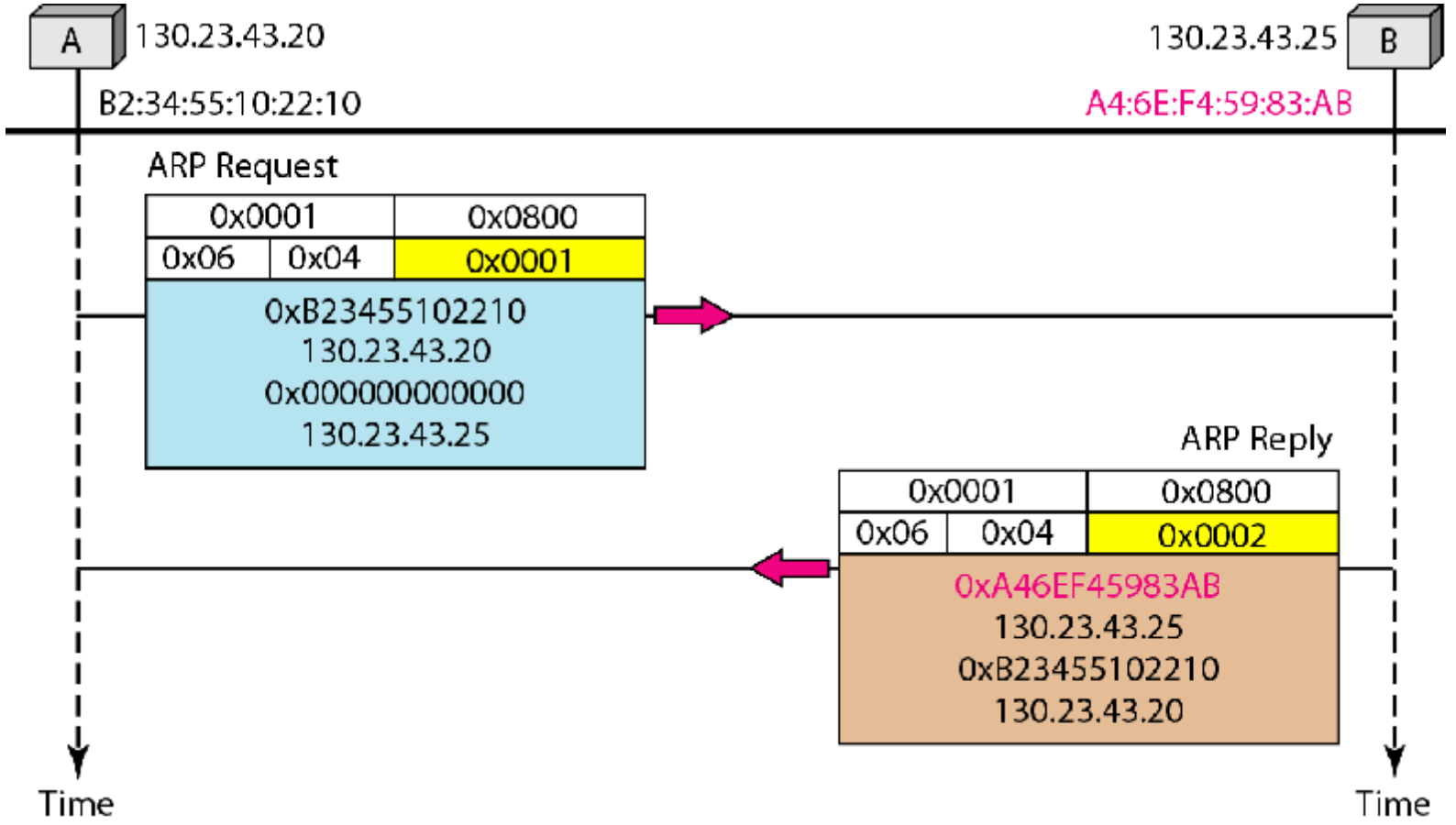
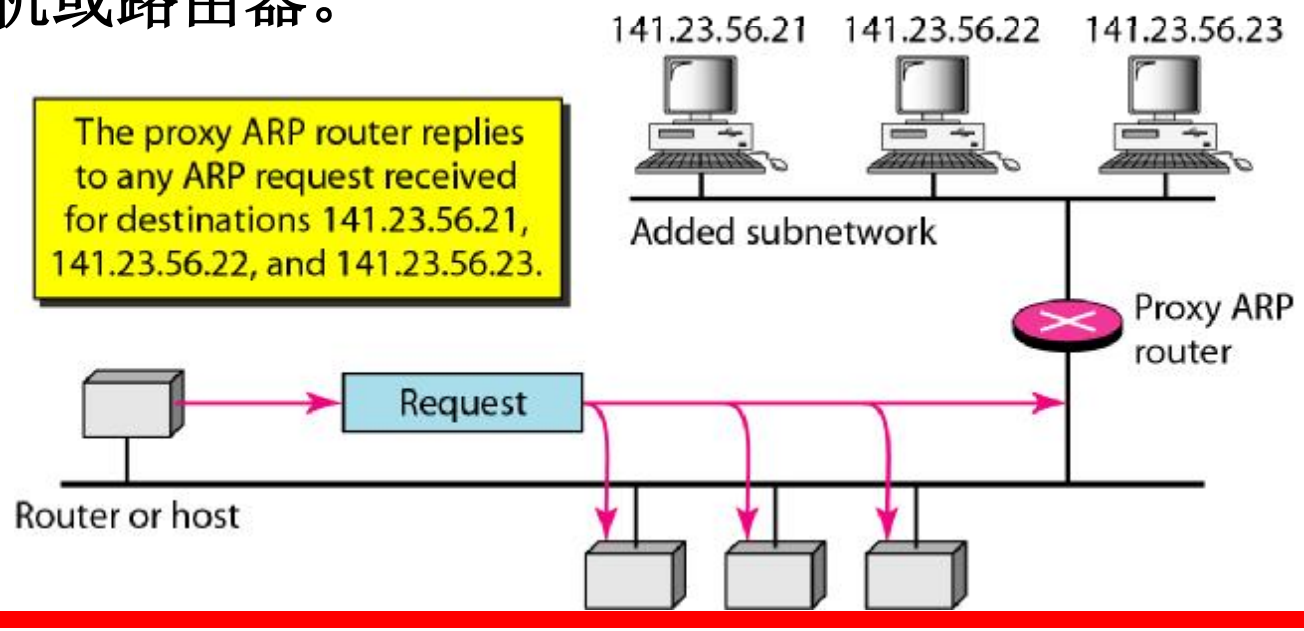


图21.6 代理ARP

- 代理ARP可产生子网化的效果，它可以代表一组主机的ARP；
- 每当运行代理ARP的路由器接收到一个寻找这些主机中的一个主机的IP地址的ARP请求时，路由器就发送一个ARP回答，宣布它自己的硬件（物理）地址；
- 当这个路由器收到真正的IP分组后，它就将这个分组发送给相应的主机或路由器。



物理地址映射到逻辑地址：RARP、BOOTP和DHCP

- 有两种可能的场合，一个主机知道它的物理地址但不知道其逻辑地址：
  - 1.无盘站点正在被引导，站点可以通过检查其接口得到它的物理地址，但不知道它的逻辑地址（使用RARP解决）；
  - 2. 一个组织机构没有足够的IP地址分配给每一个站点，只能按需分配，站点可发送它的物理地址并请求延续一个短暂的时间（使用BOOTP或DHCP解决）。

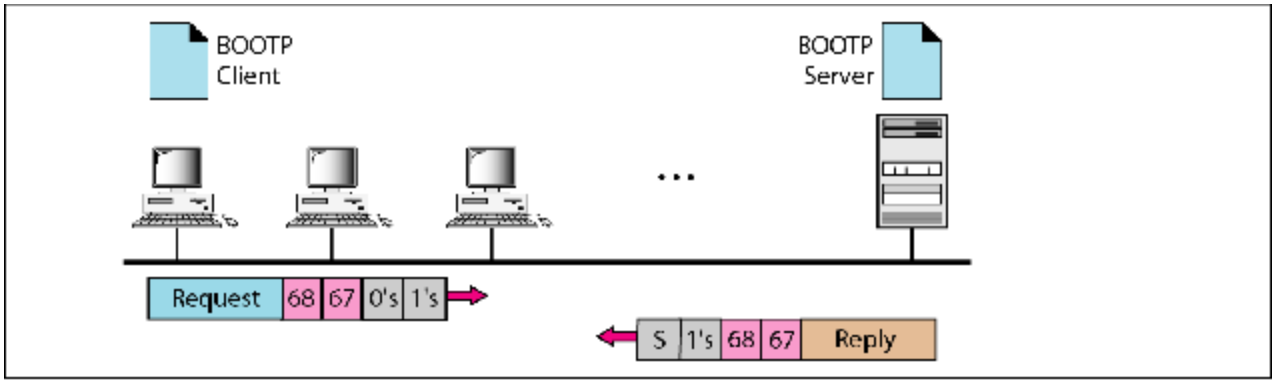
反向地址解析协议RARP

- RARP是为仅知道物理地址的机器寻找它的逻辑地址设计的；
- IP地址通常可从存储在磁盘文件中的配置文件中读出，但是无盘机器通常从ROM引导，ROM只有少量引导信息，不包括IP；
- 机器可以得到其物理地址（例如，读它的NIC），这在本地是唯一的，然后使用RARP协议从物理地址求得逻辑地址；
- 先创建一个RARP请求，并在本地网络上广播，在本地网络上知道所有IP地址的另一个机器就用RARP回答来响应；
- 请求的机器必须运行RARP客户程序，而响应的机器必须运行RARP服务器程序；
- RARP有个严重问题：在数据链路层进行广播，其广播地址在以太网中是全1，不能通过网络边界；如果网络管理员有多个网络或多个子网，它需要为每个网络或子网指定一个RARP服务器，这使得RARP几乎不再使用，而被BOOTP和DHCP取代。

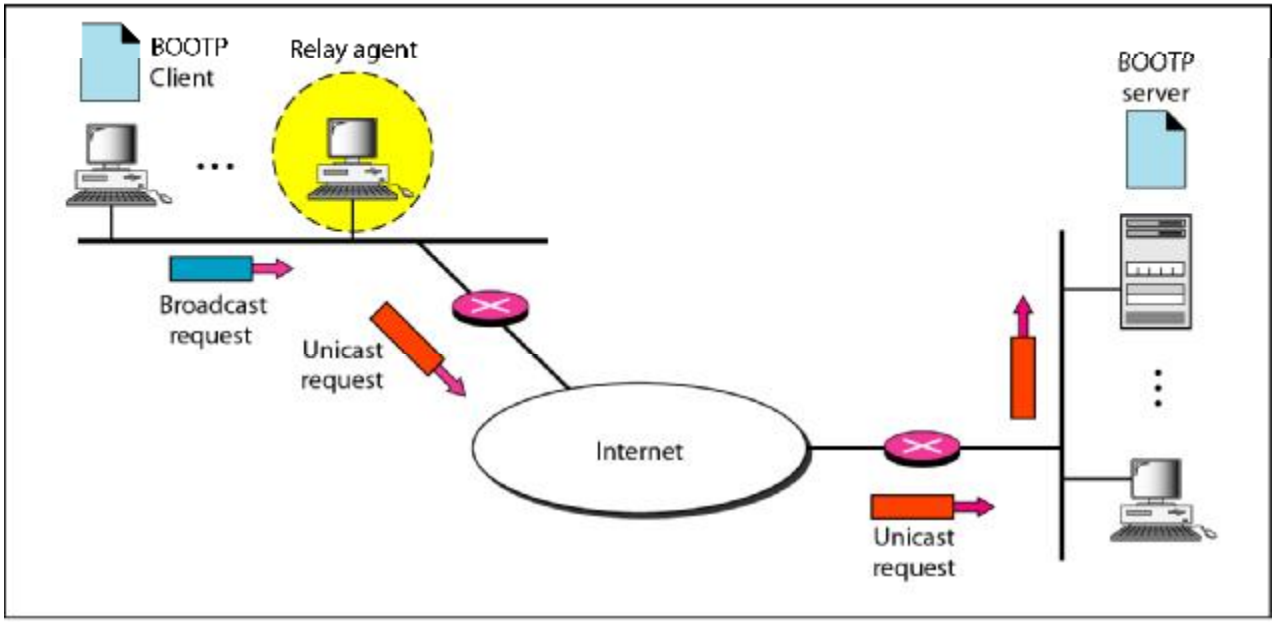
引导程序协议BOOTP（Bootstrap Protocol）

- BOOTP是一种客户机/服务器协议，提供物理地址到逻辑地址的映射，它是一个应用层协议；
- 客户机和服务器可以放置在同一个网络上或不同的网络上；
- BOOTP报文被封装在UDP分组中，而UDP分组本身被封装在IP分组中（使用两个保留的端口号，客户端68和服务器67）；
- 客户机使用全0作为源地址，全1作为目的地址发送请求；
- 优于RARP：客户机与服务器都是应用层的进程，客户机和服务器可以不在同一个网络上；
- 由于客户机不知道服务器的IP地址，所以BOOTP请求是广播的，而广播的数据报不能通过任何路由器，需要使用中继代理；
- 中继代理知道BOOTP服务器的单播地址，当它接收到这种类型的分组时，它在单播的数据报中封装报文并向BOOTP服务器发送请求；接收到回答之后，中继代理向客户机转发它。

图21.7 在同一 网络上和不同网络上的BOOTP客户和服务器



a. Client and server on the same network



b. Client and server on different networks

DHCP（Dynamic Host Configuration Protocol）

- BOOTP是一个静态配置协议，客户机的物理地址与IP地址的绑定必须已经存在，绑定是预先确定的；但是，如果一个主机从一个物理网络移动到另一个物理网络，或者主机想要一个临时的IP地址如何呢？BOOTP不能很好处理这些问题，因为表中物理地址和IP地址的绑定在网络管理员改动之前，是静态和固定的；
- DHCP能够提供静态的和动态的地址配置，可以是人工的或自动的；
- 静态地址配置：与BOOTP相同，DHCP服务器有一个数据库静态地绑定物理地址和IP地址；
- 动态地址配置：DHCP有第二个数据库，它拥有一个可用的IP地址池，使DHCP成为动态的；当DHCP客户机请求一个临时的IP地址时，DHCP服务器就查找可用（即未使用的）IP地址池，然后指定一个在可协商的期间内有效的IP地址。

DHCP（cont.）

- 当DHCP客户机向服务器发送请求时，服务器首先检查静态数据库，如果存在所请求的物理地址的项目，则返回该客户的永久IP地址；反之，服务器就从可用的IP地址池中选择一个IP地址，并将这个地址指定给该用户，然后将该项目加到动态数据库中；
- 当主机从一个网络移到另一个网络，或连接到一个网络后又断开连接时，DHCP需要是动态的，DHCP在有限的期间提供临时IP地址；
- 从可用IP地址池指定的地址是临时地址，DHCP服务器发出租用（lease）是对特定期间而言的；
- 租用期到时，客户机或停止使用这个IP地址，或更新其租用；服务器对这个更新可选择同意或不同意，如果不同意，客户机就停止使用该地址；
- DHCP使用UDP三个端口号：67、68和DHCPv6 546-failover

问题？

pQ：BOOTP和DHCP属于哪一层的协议？

21-2 ICMP

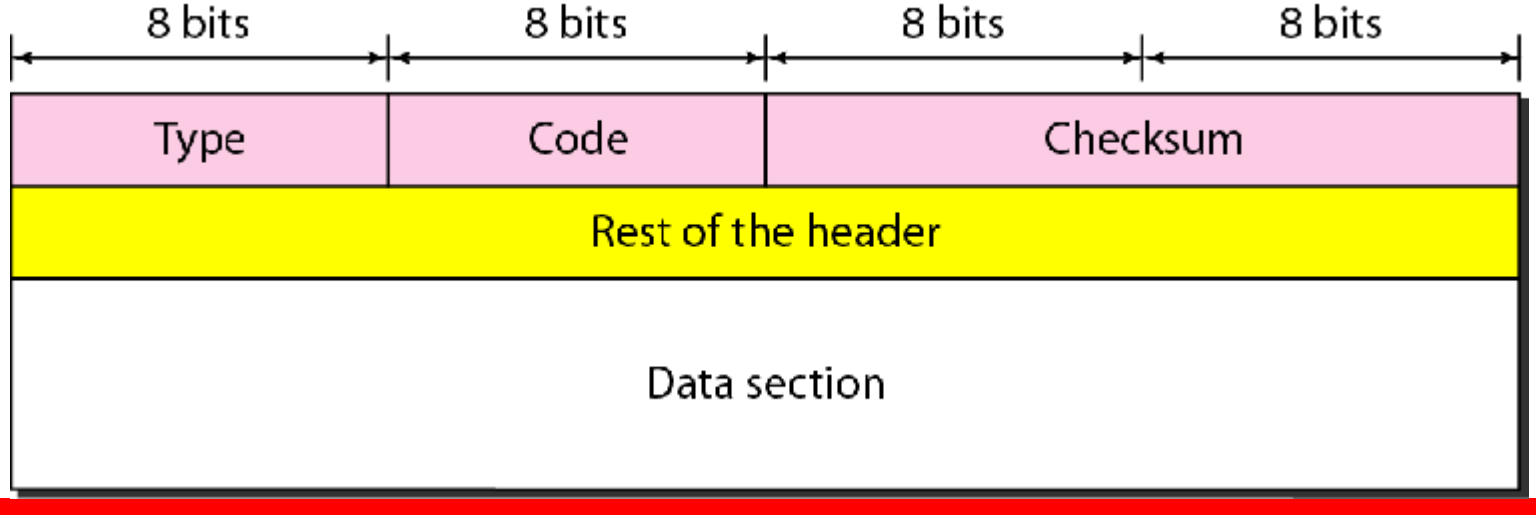
- IP协议没有差错报告或差错纠正机制；
- IP协议还缺少一种为主机和管理查询的机制，主机有时需要确定一个路由器或另一个主机是否是活跃的，有时网络管理员需要从另一个主机或路由器得到信息；
- 因特网控制报文协议（Internet Control Message Protocol，ICMP）就是为了弥补上述两个缺点而设计的，它是配合IP协议使用的。

ICMP报文类型

- 两大类：差错报告报文和查询报文；
- 差错报告报文向路由器或主机（目的端）报告在处理一个IP数据报时可能碰到的一些问题（或差错）；
- 查询报文是成对出现的，它帮助主机或网络管理员从一个路由器或另一个主机得到特定的信息；例如，节点能够发现它们的邻站，主机能够发现和知道在它们的网络上的一些路由器的情况，而一些路由器能帮助一个节点改变报文的路由等。

图21.8 ICMP报文一般格式

- 一个8字节的头部和一个可变长的数据部分；
- 不同类型报文的头部格式不同，但最前面4个字节相同；
- 第一个字段是ICMP的类型，定义报文的类型；代码字段指定了发送此特定报文类型的原因，最后一个共同的字段是校验和字段，头部其余部分对每种报文类型都是特定的；
- 差错报文数据部分所携带的信息可找出引起差错的原始分组，查询报文数据部分携带了基于查询类型的额外信息。

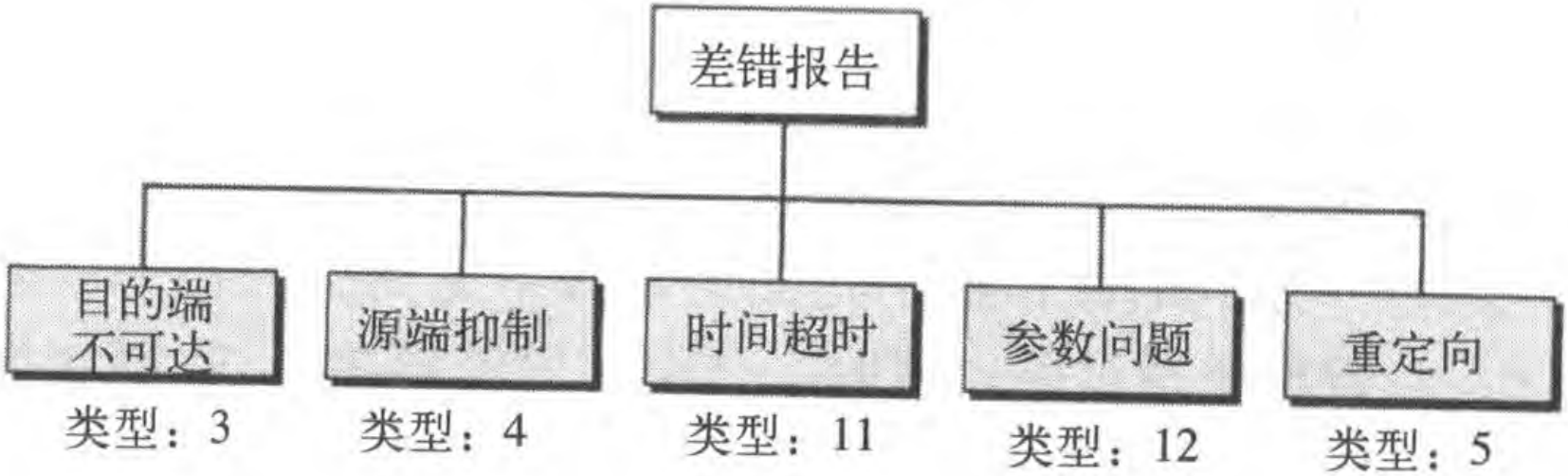


差错报告

- IP是不可靠的协议，不考虑差错校验和差错控制，ICMP就是为弥补这个缺点而设计的；
- 然而，ICMP不能纠正差错，它只是报告差错，差错纠正留给高层协议去完成；
- 差错报文总是发送给原始的源端，因为在数据报中关于路由唯一可用的信息就是源IP地址和目的IP地址，ICMP使用源IP地址将差错报文发送给数据报的源端（发送方）。

ICMP总是向原始的源方报告差错报文。

图21.9 差错报告报文类型

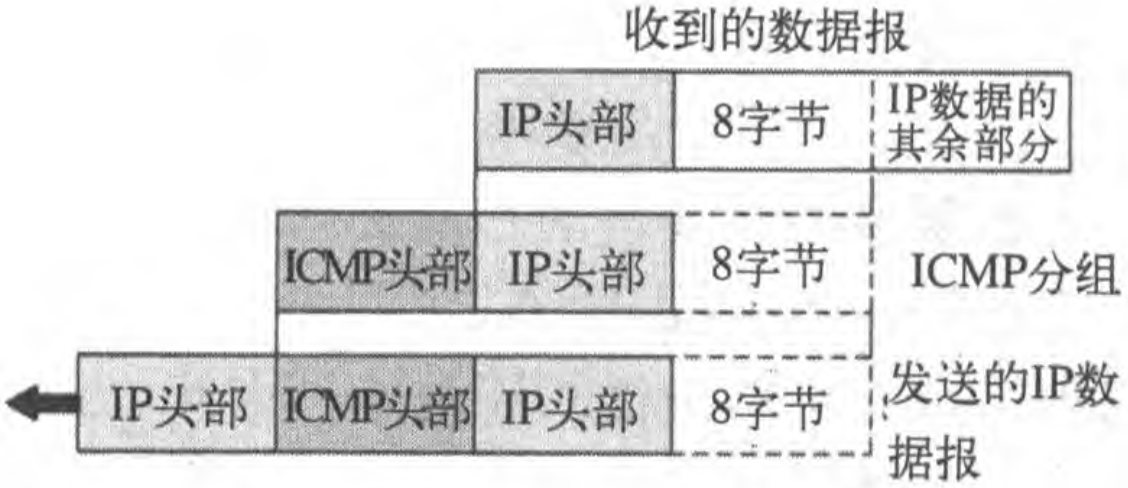


关于ICMP差错报文有下列要点：

- 对于携带ICMP差错报文的数据报，不再产生ICMP差错报文。
- 对于分段的数据报文，如果不是第一个分段则不产生ICMP差错报文。
- 对于多播地址的数据报文，不产生ICMP 差错报文。
- 具有特殊地址的数据报文，如127.0.0.0或者0.0.0.0，不产生ICMP差错报文。

图21.10 差错报文的数据字段的内容

- 差错报文都包括数据部分，数据部分包括原始数据报的IP头部加上数据报中前8个字节数据；
- 从原始数据报头部可得到原始的源端，它接收差错报文；
- 要包括8个字节数据是因为这前8个字节提供了关于端口号（UDP和TCP）和序列号（TCP）的信息；这个信息是需要的，这样源端可以将差错情况通知给这些协议；
- ICMP形成差错分组，然后再封装成IP数据报



目的端不可达和源端抑制

- 当路由器不能够给数据报找到路由或主机不能传递数据报时，就丢弃这个数据报；然后，这个路由器或主机就发回目的端不可达报文给发出该数据报的源主机；
- 缺乏流量控制可能会在路由器或目的主机中产生拥塞，若数据报接收速率比它们被转发或处理的速率快得多，缓存队列就会溢出，此时路由器或主机只能将某些数据报丢弃；
- **ICMP的源端抑制报文（source-quench message）就是为了给IP增加一种流量控制而设计的；**
- 当路由器或主机因拥塞而丢弃数据报时，它就向数据报的发送方发送源端抑制报文；
- 这个报文有两个目的：第一，它通知源端数据报已被丢弃；第二，它警告源端，在路径中的某处出现拥塞，因而源端必须放慢（抑制）发送过程。

时间超时和参数问题

- 时间超时报文在两种情况下产生：（1）收到数据报的路由器将其TTL减为0时，就丢弃它，并向源端发送时间超时报文；（2）当组成一个报文的所有分段未能在某一时限内到达主机时，也要产生时间超时报文；
- 参数问题：如果路由器或目的主机在数据报头部发现了二义性或在数据报的某个字段中缺少某个值，它就丢弃这个数据报，并向源端发送参数问题报文。

图21.11 重定向概念

p为了提高效率，主机都不参与路由选择更新过程；

p当主机开始联网工作时，通常只知道默认路由器的IP地址；

p当主机向另一个网络发送数据报时，就可能将数据报发给了错误的路由器，此时，收到此数据报的路由器会将该数据报转发给正确的路由器；

p但要更新主机中的路由表，它就要向主机发送重定向报文；

p图中R1在查找路由表后发现分组应当走R2，它把分组发送到R2，并向A发送重定向报文，主机A的路由表就更新了。

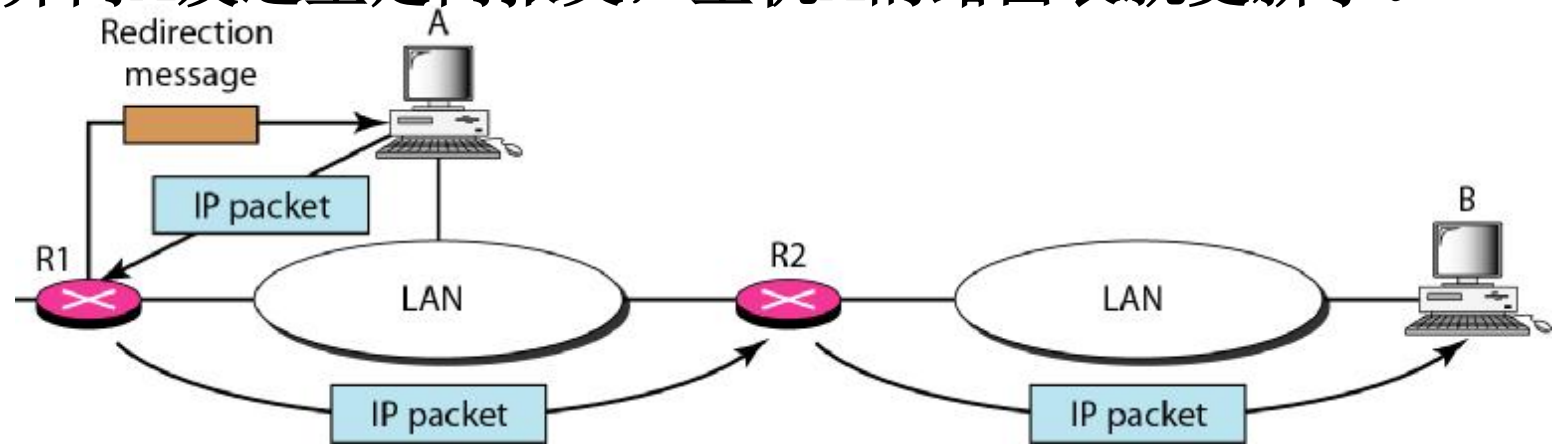


图21.12 查询报文

- ICMP能对某些网络问题进行诊断，通过使用由4对不同报文组成的查询报文来完成；
- 在ICMP查询报文中，一个节点发送出报文，然后由目的节点用特定的格式进行回答；
- 查询报文封装在IP分组中，然后再将IP分组封装在数据链路层的帧中；
- 原始IP字节没有包含在报文中

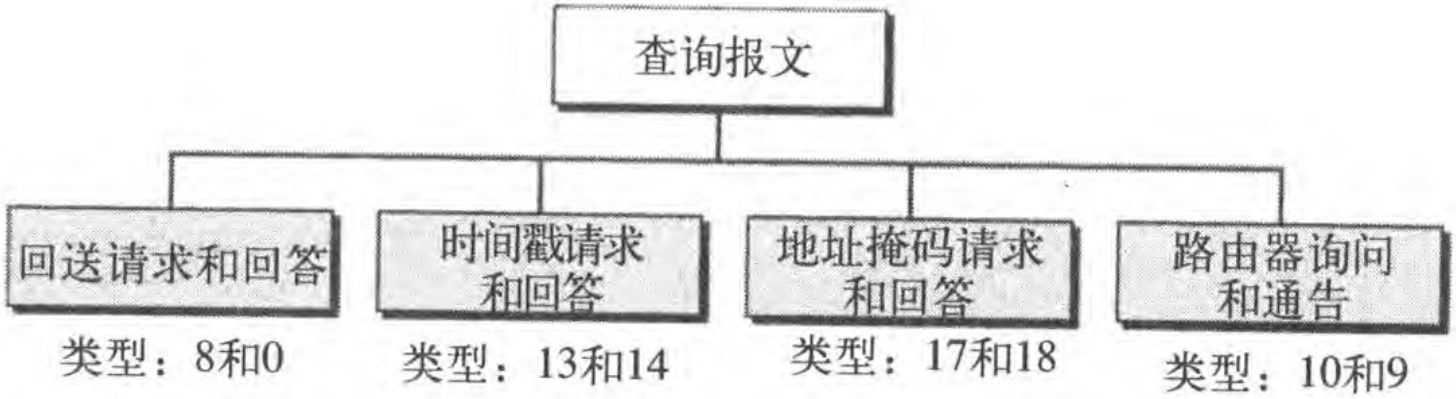
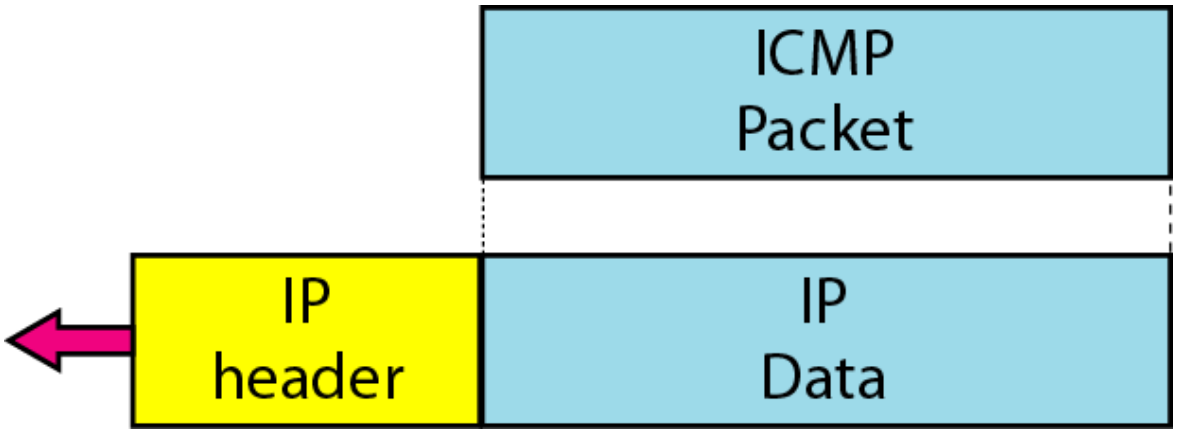


图21.13 封装ICMP 查询报文



回送请求和回答/时间戳请求和回答

- **Echo-request和echo-reply：为诊断目的而设计；**
- 回送请求和回送回答组合起来确定了两个系统（主机或路由器）是否彼此能通信；
- 发送回送请求的机器在收到回送回答报文时，就证明了在发送方和接收方之间能够使用**IP**数据报通信；
- 此外，还证明了中间的一些路由器能够接收、处理和转发**IP**数据报；
- 大多数系统都提供**ping**命令，它可以创建一系列（不仅是一个）回送请求或回送回答报文，提供统计信息；
- 两个机器（主机或路由器）可使用时间戳请求和时间戳回答报文来确定**IP**数据报在两个机器之间往返所需的时间，它也可用做两个机器之间的时钟同步。

地址掩码请求和回答

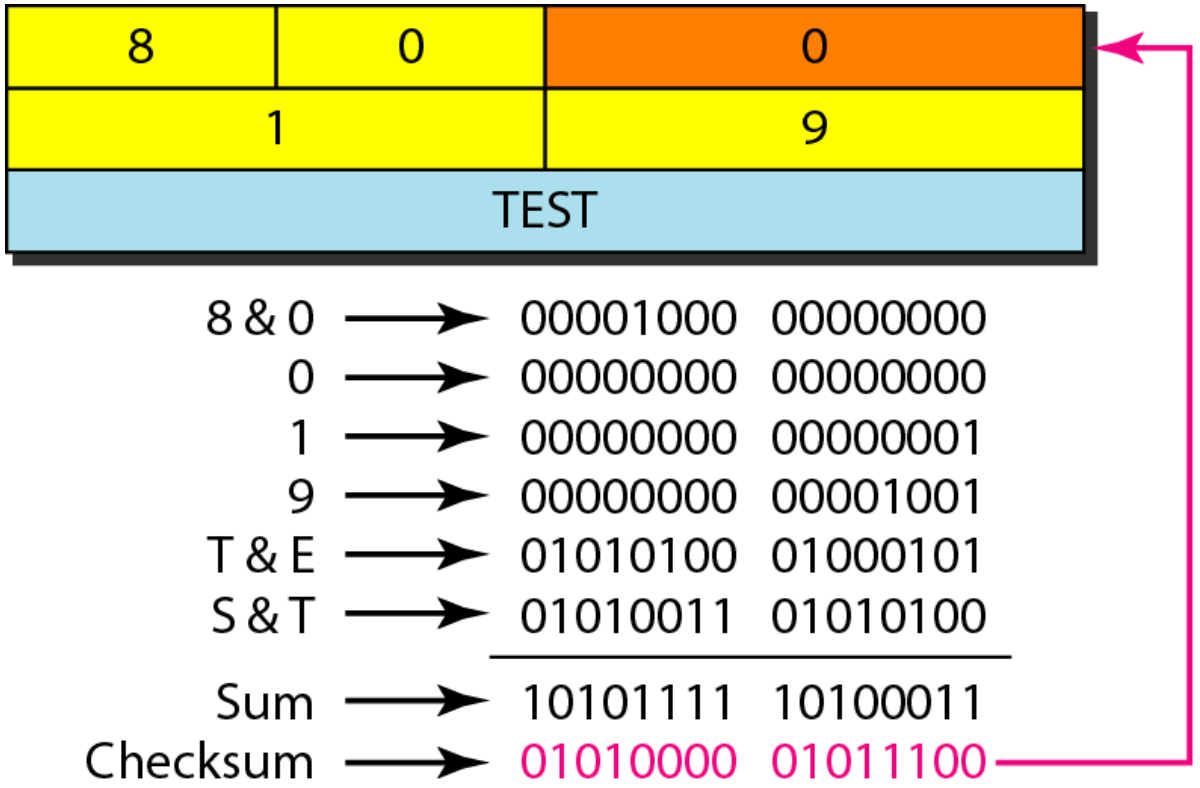
- 主机可能知道它的IP地址但不知道相应的掩码，为此主机应向局域网上的路由器发送地址掩码请求报文；
- 如果主机知道该路由器的地址，它就直接向路由器发送该请求；如果它不知道地址，就广播该报文；
- 路由器接收到地址掩码请求报文后，用地址掩码回答报文进行响应，向主机提供所需要的掩码；将这个掩码应用到完整的IP地址上，就可求得子网的地址（P413）。

路由器询问和通告

- 主机如果想将数据发送给另一网络上的主机，就需要知道连接到它自己的网络上的路由器的地址；另外，该主机还需要知道这些路由器是否正常工作，路由器询问报文和路由器通知报文就用于这种情况；
- 主机可将路由器询问报文进行广播（或多播），收到询问报文的一个或多个路由器就使用路由器通告报文广播其路由选择信息；
- 即使在没有主机询问时，路由器也可周期性地发送路由器通告报文；
- 注意：路由器在发送通告报文时，它不仅通告自己，而且通告它所知道的所有在这个网络上的路由器。

例21.2

图21.14表示了对一个简单的回送请求报文的校验和计算的实验。我们随机选定标识符为1，而序列号为9。将报文划分为16位（2字节）的字；将这些字相加，然后将其和取反；现在发送方就可将此值放置在校验和字段。



例21.3

我们用ping程序测试服务器fhda.edu。结果如下：ping程序发送报文的序列号从0开始，每次探测给出这次的往返时间。封装了ICMP报文的IP数据报的TTL（生存时间）字段已被为置为62，这就是说该分组的传输不能超过62跳。一开始ping程序定义数据部分为56个字节，IP数据报总长度为84字节。这是显然的，这是因为我们要增加ICMP头部8字节和IP头部20字节，所以其结果是84字节。但是注意：每次探测ping程序定义的字节个数为64，这是ICMP分组的总长度（56+8）。

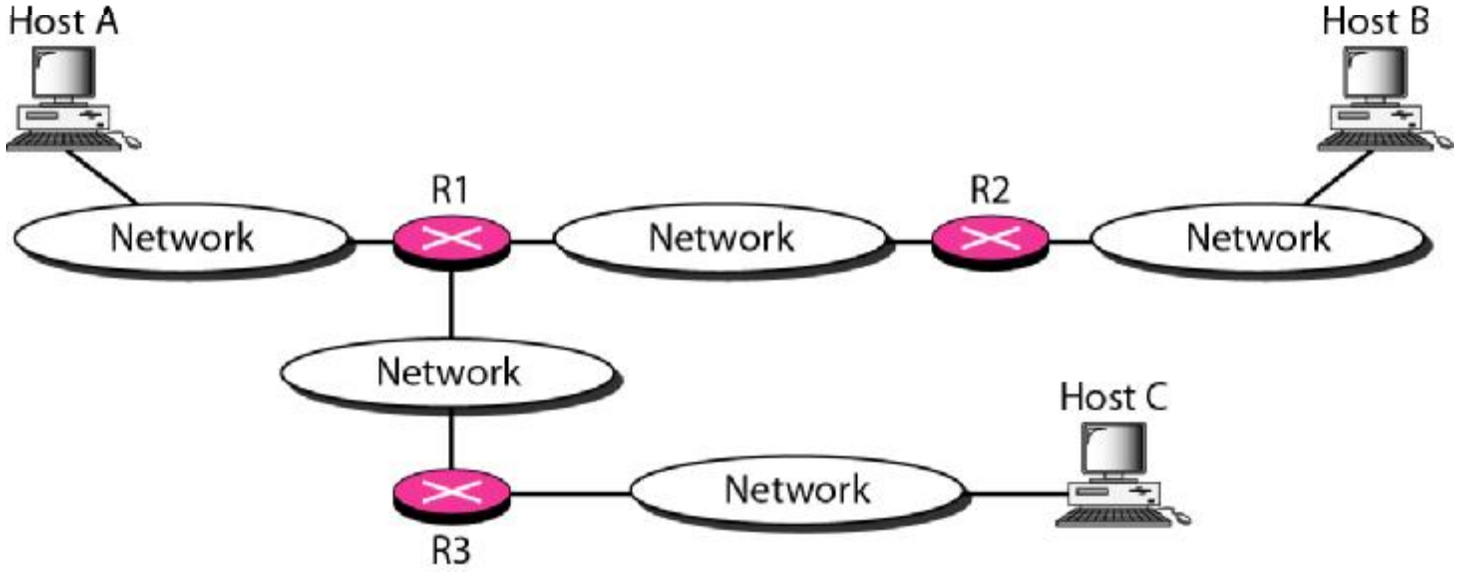
例21.3（cont.）

```
$ ping fhda.edu
PING fhda.edu (153.18.8.1)   56 (84)  bytes of data.
64 bytes from tiptoe.fhda.edu (153.18.8.1): icmp_seq=0     ttl=62     time=1.91 ms
64 bytes from tiptoe.fhda.edu (153.18.8.1): icmp_seq=1     ttl=62     time=2.04 ms
64 bytes from tiptoe.fhda.edu (153.18.8.1): icmp_seq=2     ttl=62     time=1.90 ms
64 bytes from tiptoe.fhda.edu (153.18.8.1): icmp_seq=3     ttl=62     time=1.97 ms
64 bytes from tiptoe.fhda.edu (153.18.8.1): icmp_seq=4     ttl=62     time=1.93 ms
64 bytes from tiptoe.fhda.edu (153.18.8.1): icmp_seq=5     ttl=62     time=2.00 ms
64 bytes from tiptoe.fhda.edu (153.18.8.1): icmp_seq=6     ttl=62     time=1.94 ms
64 bytes from tiptoe.fhda.edu (153.18.8.1): icmp_seq=7     ttl=62     time=1.94 ms
64 bytes from tiptoe.fhda.edu (153.18.8.1): icmp_seq=8     ttl=62     time=1.97 ms
64 bytes from tiptoe.fhda.edu (153.18.8.1): icmp_seq=9     ttl=62     time=1.89 ms
64 bytes from tiptoe.fhda.edu (153.18.8.1): icmp_seq=10    ttl=62     time=1.98 ms

--- fhda.edu ping statistics ---
11 packets transmitted, 11 received, 0% packet loss, time 10103ms
      rtt min/avg/max = 1.899/1.955/2.041 ms
```

Traceroute程序

- UNIX中的traceroute程序或Windows中的tracert程序可用于追踪一个分组从源端到目的端所经过的路由；
- 将这个程序和ICMP数据报一道使用，程序巧妙地利用时间超时与目的端不可达两种ICMP报文找出分组经过的路由；这是一个UDP提供服务的应用程序
- traceroute程序利用ICMP报文和IP分组的TTL字段找出路由



Traceroute程序步骤

- a.A上的traceroute程序利用UDP协议发送一个分组到目的主机B，该报文被封装在TTL为1的IP分组中；
- b.路由器R1接收到分组并将TTL的值减1，成为0，然后R1丢弃该分组并发送一个时间超时ICMP报文给源端；
- c.traceroute程序接收到该ICMP报文，利用封装了ICMP报文的IP分组的目的地址求得路由器R1的IP地址；
- traceroute程序重复上面的步骤求得R2的地址，但是，需将TTL设为2；增加TTL的值，以此类推，一直到达目的主机B；
- 当主机B接收到分组时，TTL值减1，但主机B不丢弃该报文，因为它已到达了它的最终的目的地；

Traceroute程序步骤（cont.）

- ICMP报文如何发送回到主机A?traceroute程序在此处采用一种不寻常的策略，UDP分组的目的端口被设置为UDP协议不提供服务的端口；
- 当主机B接收到该分组，它找不到接收该传递的应用程序，它丢弃该分组并发送一个ICMP目的端不可达报文到主机A；
- traceroute记录到达的IP数据报的目的地址（接收到目的端不可达的ICMP报文就表示已求得全部路由，不需要再发送分组）

例21.4

我们用traceroute程序求计算机voyager.deanza.edu到服务器fuda.edu的路由，下面表示其结果：

```
$ traceroute fhda.edu
traceroute to fhda.edu    (153.18.8.1), 30 hops max, 38 byte packets
 1  Dcore.fhda.edu        (153.18.31.254)    0.995 ms    0.899 ms    0.878 ms
 2  Dbackup.fhda.edu      (153.18.251.4)     1.039 ms    1.064 ms    1.083 ms
 3  tiptoe.fhda.edu       (153.18.8.1)       1.797 ms    1.642 ms    1.757 ms
```

命令后无序列的行表示了目的端是153.18.8.1，TTL值是30跳，该分组包含38个字节：IP头部20个字节、UDP头部8个字节和应用数据10个字节。traceroute程序使用应用数据存储分组的轨迹。

例21.4（cont.）

第一行表示访问第一个路由器，该路由器名称是Dcore.fuda.edu，其IP地址为135.18.31.254。第一个往返时间是0.995ms，第二个往返时间是0.899ms，第三个往返时间是0.878ms。

第二行表示访问第二个路由器，该路由器名称是Dbackup.fhda.edu，其IP地址为153.18.251.40，也显示了三个往返时间。

第三行表示目的主机。我们知道这是目的主机，因为没有更多的行。目的主机是服务器fhda.edu，但它的名称是tiptoe.fuda.edu，其IP地址为153.18.8.1，也显示了三个往返时间。

例21.5

下页的例子我们追踪了一个较长的路由，到xerox.com的路由。此处从源端到目的端有17跳。注意：有些往返时间看起来非常异常，这可能是路由器太忙不能立即处理分组。

```
$ traceroute xerox.com
traceroute to xerox.com (13.1.64.93), 30 hops max, 38 byte packets
 1  Dcore.fhda.edu          (153.18.31.254)     0.622 ms    0.891 ms    0.875 ms
 2  Ddmz.fhda.edu           (153.18.251.40)     2.132 ms    2.266 ms    2.094 ms
 3  Cinic.fhda.edu          (153.18.253.126)    2.110 ms    2.145 ms    1.763 ms
 4  cenic.net               (137.164.32.140)    3.069 ms    2.875 ms    2.930 ms
 5  cenic.net               (137.164.22.31)     4.205 ms    4.870 ms    4.197 ms
      . . . .                 . . . .             . . .       . . . .     . . .
14  snfc21.pbi.net          (151.164.191.49)    7.656 ms    7.129 ms    6.866 ms
15  sbcglobal.net           (151.164.243.58)    7.844 ms    7.545 ms    7.353 ms
16  pacbell.net             (209.232.138.114)   9.857 ms    9.535 ms    9.603 ms
17  209.233.48.223          (209.233.48.223)    10.634 ms   10.771 ms   10.592 ms
18  alpha.Xerox.COM         (13.1.64.93)        11.172 ms   11.048 ms   10.922 ms
```

21-3 IGMP

- IP协议可用到两种类型的通信：单播和多播；
- 单播是一个发送方和一个接收方之间的通信，它是一对一的通信；
- 但是，有些过程有时需要将同一个报文同时发送给许多的接收方，这称为多播或组播（multicast），即一到多的通信；
- 多播有许多应用，例如，股票价格的变动要通知多个证券经纪人，或视频会议、远程学习等；
- 因特网组管理协议（Internet Group Management Protocol，IGMP）是其中一个必要的、但不是充分的协议，多播也包含其他的协议（多播路由协议）；
- 在IP协议中，IGMP是一个辅助协议。

组管理

- 对于因特网上的多播，需要具有路由多播分组能力的路由器，这些路由器的路由表必须由某些多播路由协议更新；
- IGMP不是一个多播路由协议，而是一个管理组成员的协议；
- 在任何网络中，都存在一个或多个多播路由器把多播分组分发给主机或其他的路由器，IGMP协议为多播路由器提供关于连接到网络上的主机（路由器）成员状态的信息；
- 一个多播路由器每天可以接收到数以千计的属于不同组的多播分组，如果路由器不了解主机的成员状态，它就必须广播所有的分组，这造成通信量大量的增加并浪费带宽；
- 一个较好的解决办法是在网络中保存一份组列表，该列表中至少有一个忠实的成员，IGMP帮助多播路由器创建和更新这个表；
- IGMP当前版本是v3

图21.16 IGMPv2报文类型

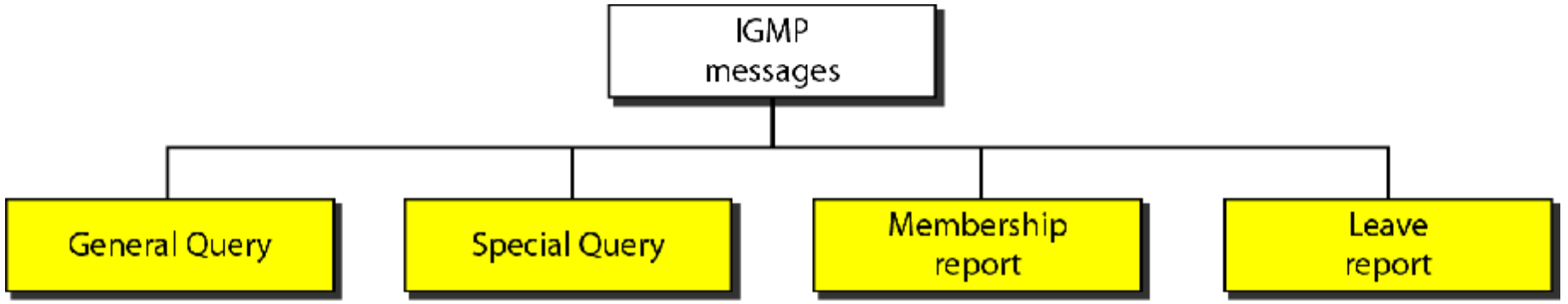


图21.17 IGMPv2报文格式

- 类型：8位，如表21.1所示；
- 最大响应时间：8位，定义回答一个查询所需的时间，以十分之一秒为单位，在查询报文中，该值是非0的，在其余两个类型的报文其值为0；
- 校验和：16位，以8字节的报文计算；
- 组地址：对于普通的查询报文，字段值为0；在特殊的查询、组成员报告和离开报告报文中，定义为组标识符（组的多播地址）

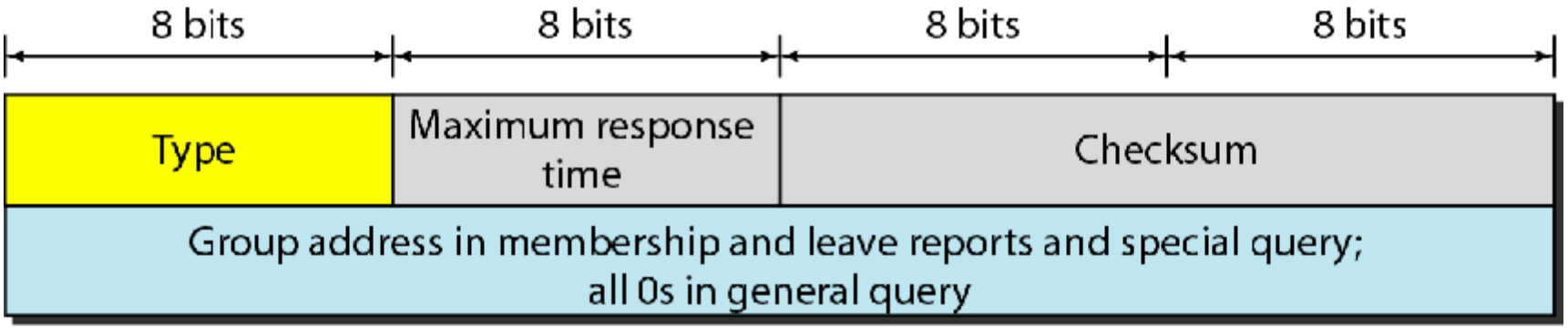
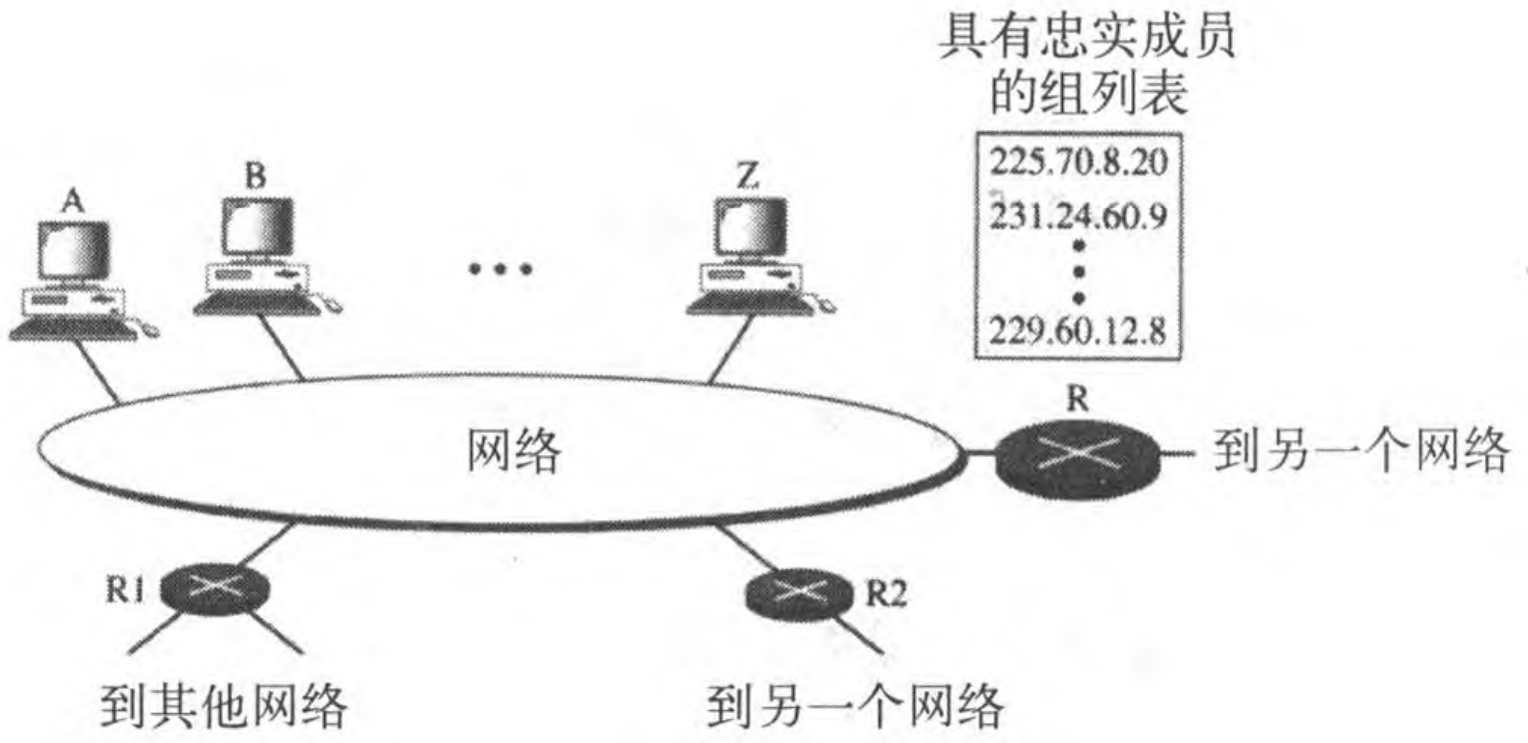


表21.1 IGMP 类型字段

Type	*Value*
General or special query	0x11 or 00010001
Membership report	0x16 or 00010110
Leave report	0x17 or 00010111

IGMP操作

- IGMP是本地运行的，连接到网络的多播路由器有一个组多播地址表，该表内至少有一个忠实的成员



IGMP操作（cont.）

- 对每一组，都有一个路由器负责将多播分组分发给指定的那个组，就是说如果与网络相连的多播路由器有三个，它们的组标识符都是相互排斥的；例如，图21.18中只有路由器R分发其多播地址为225.70.8.20的分组；
- 主机或多播路由器在一个组中可能存在成员关系；
- 当一个主机有成员关系时，就意味着它的一个进程（一个应用程序）接收到来自某一组的多播分组；当一个路由器具有成员关系时，就意味着一个连接到它的其他接口的网络接收到这些多播分组；我们说该主机或路由器对该组有加入请求；
- 在这两种情况下，主机和路由器保存一个组标识符表，并将它们的关系中继给分发路由器；
- 例如图21.18中，R是分发路由器，其余两个多播路由器R1和R2依赖于R维护的组列表，它们可能是这个网络中R的接收者

加入一组

- 一个主机或路由器可加入一个组；
- 每一个主机维护一个组内成员进程表，当一个进程要加入到一个新组时，它就向主机发送请求，该主机就在它的表中增加该进程的名字和所请求的组的名字；
- 如果这是在该特殊组中的第一个成员关系的请求，则主机就向多播路由器发送一个成员关系报告报文；如果不是第一个成员关系，则不需要发送成员关系报告，因为主机已是组的成员，它已经接收了这个组的多播分组；
- 协议要求发送两次成员关系报告，在几分钟内一个接着一个地发送，这样，如果第一个丢失或损坏了，第二个可代替它。

在IGMP中，成员关系报告一个接着一个地发送两次。

监视成员关系

- 为了防止主机关闭或从系统中删除，路由器周期性地（默认125 秒）发送一个普通的查询报文，组地址设置为0.0.0.0，代表查询成员关系的延续需要考虑主机所加入的所有的组，而不是仅一个组；
- 路由器期待从多播地址表中的每个组得到一个回答，即使一个新组也不例外；
- 当主机或路由器接收到普通查询报文时，如果它对一个组有加入请求，它就用一个成员关系报告来响应；如果这是一个公用的加入请求（例如，两个主机对同一组有加入请求）时，只有一个响应发送到那个组，以防止不必要的通信量，这称为一个延迟响应；
- 注意：必须仅有一个路由器发送查询报文（通常称为查询路由器），这也是为了防止不必要的通信量

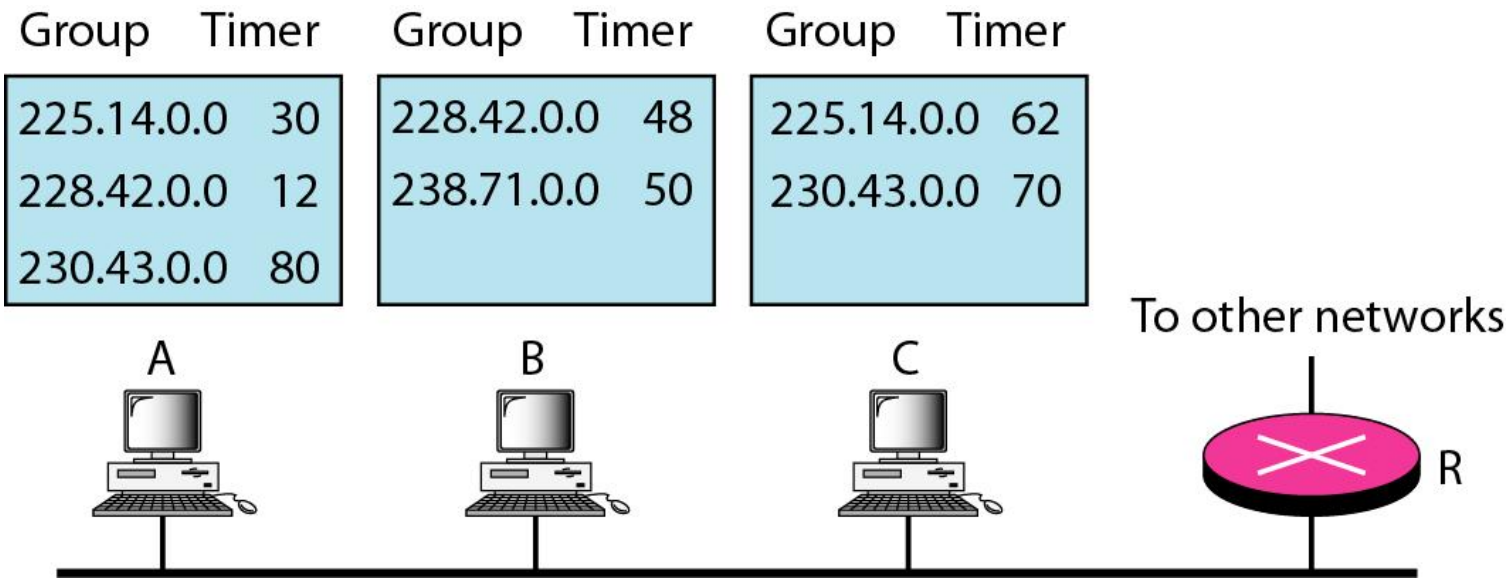
普通查询报文没有定义一个特殊的组。

延迟响应

- 为了防止不必要的通信，IGMP使用延迟响应策略；
- 当一个主机或路由器接收到查询报文后，它并不立即响应而是经过一定的时间间隔后才发出响应；
- 每一个主机或路由器使用一个随机数建立一个计时器，计时器在1秒到10秒之间，对多播组地址表中的每一个组设置一个计时器；
- 每个主机或路由器在它的计时器截止之前一直等待，直到它的计时器截止时间到了才发送一个成员关系报告报文；
- 在等待过程中，如果另一个主机或路由器的计时器对同一个组截止得较早，该主机或路由器就发送一个成员关系报告；
- 报告是广播的，正等待的主机或路由器接收到报告而且知道不需要为同一个组发送一个重复报告，取消它相应的计时器。

例21.6

设想网络有三个主机，如图21.19所示。在时刻0，接收到查询报文，每一个组的随机延迟时间（以十分之一秒为单位）在该组地址后面表出，如图所示。试说明其报告序列。



例21.6（cont.）

解：

事件发生的序列如下：

a. 时刻12：此时主机A中的组地址228.42.0.0的计时器截止，发送一个成员关系报告。路由器和每个主机（包括主机B）接收到这个报告，这样主机B删除组地址228.42.0.0的计时器。

b. 时刻30：此时主机A中的组地址225.14.0.0的计时器截止，发送一个成员关系报告。路由器和每个主机（包括主机C）接收到这个报告，这样主机C删除组地址225.14.0.0的计时器。

c. 时刻50：此时主机B中的组地址238.71.0.0的计时器截止，发送一个成员关系报告。路由器和每个主机接收到这个报告。

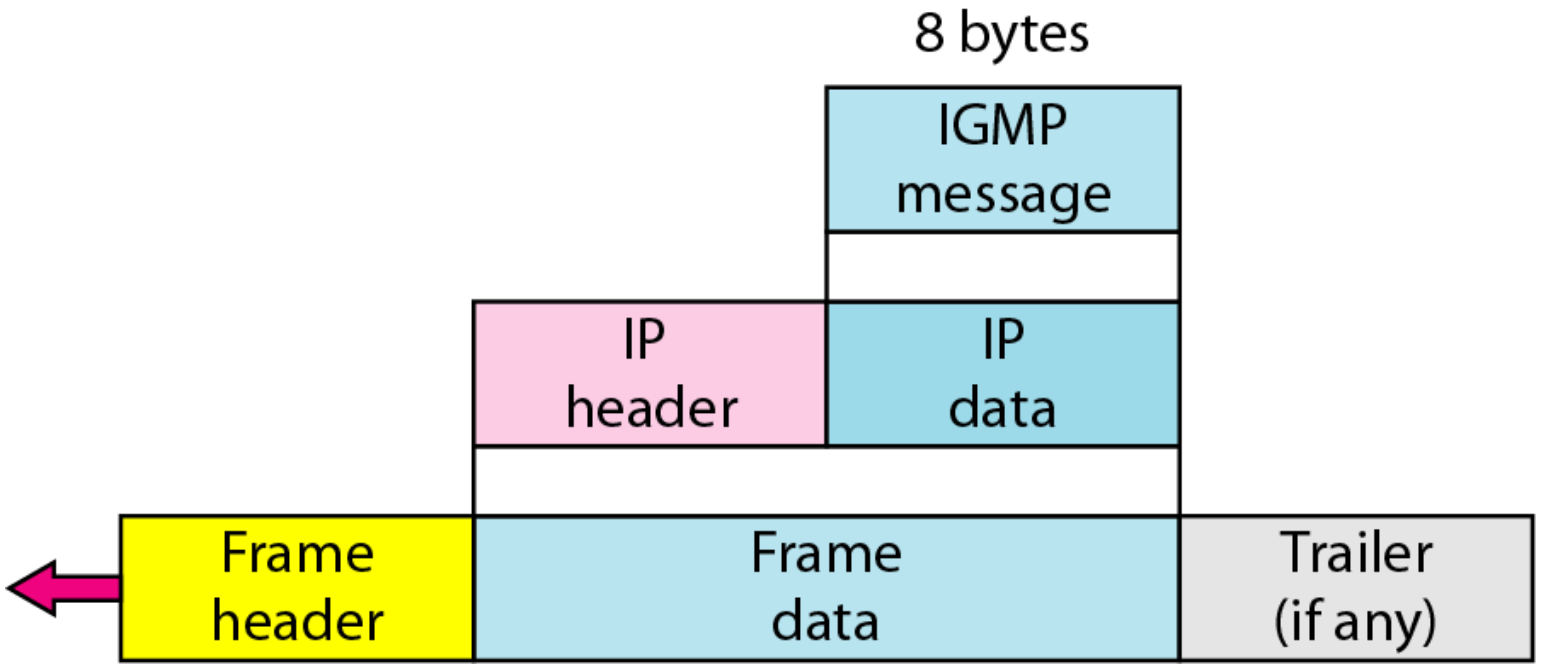
d. 时刻70：此时主机C中组地址230.43.0.0的计时器截止，发送一个成员关系报告。路由器和每个主机（包括主机A）接收到这个报告，主机A删除组地址230.43.0.0的计时器。

查询路由器

- 查询报文可以生成许多响应，为了防止不必要的通信，IGMP为每一个网络设计了一个路由器作为查询路由器；
- 只有这个路由器发送查询报文，其他路由器都是被动的（它们接收响应并更新它们的列表）

图21.20 IGMP分组的封装

- IGMP报文封装在IP数据报中，而IP数据报本身又封装在帧中



在网络层的封装

- 对于IGMP协议，IP分组的协议类型字段的值是2；
- 携带IGMP分组的IP分组的TTL字段的值为1，因为IGMP的域是局域网；
- 查询报文是用多播地址224.0.0.1（作为目的地址）进行广播，所有（支持多播的）主机和路由器都将接收到该报文；
- 类型成员关系使用与被报告的多播地址（组标识符）相同的目的地址进行广播，接收到该分组的每个站点（主机或路由器）可立即确定（从头部）哪一个组报告已经被发送；
- 离开报告报文用多播地址224.0.0.2 （作为目的地址）广播（这个子网上所有的路由器），因此路由器都接收到这类报文，主机也都接收到这类报文，但不理它。

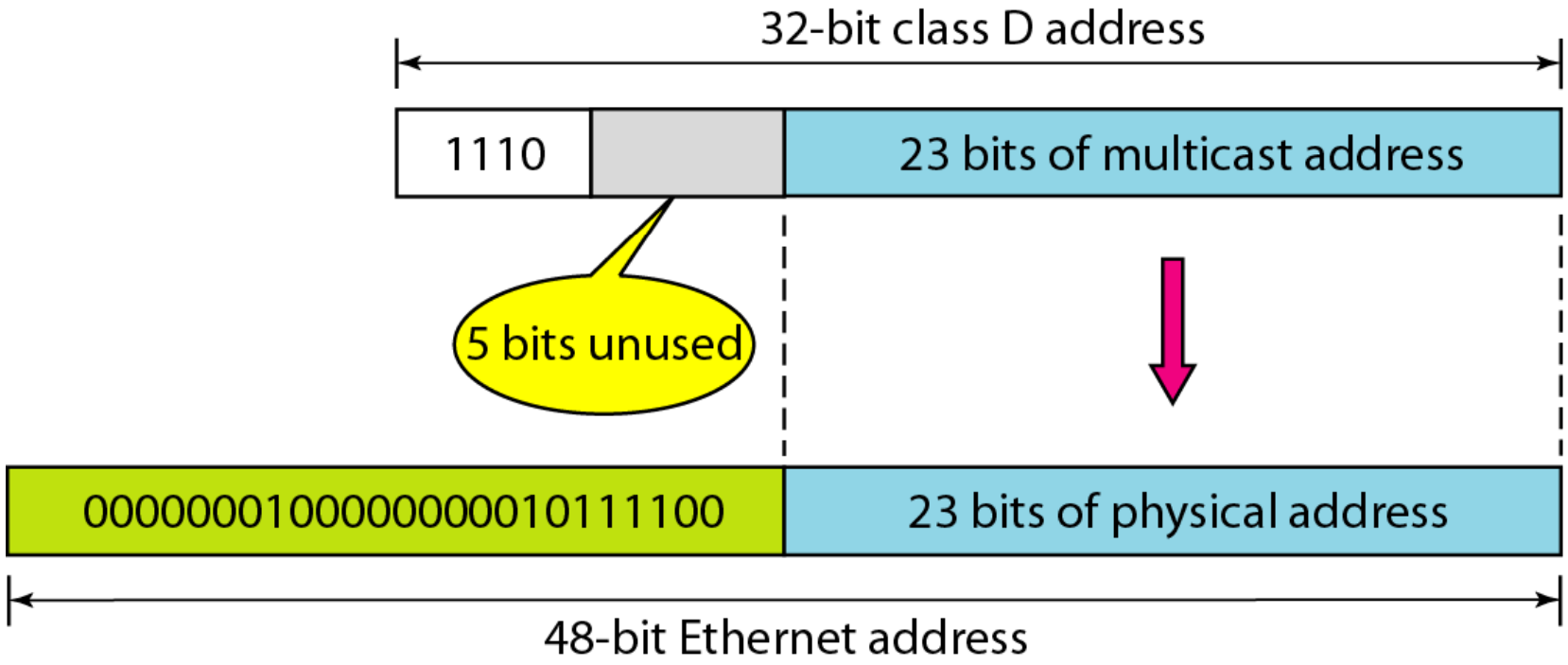
表21.2 目的IP地址

Type	*IP Destination Address*
Query	224.0.0.1 All systems on this subnet
Membership report	The multicast address of the group
Leave report	224.0.0.2 All routers on this subnet

在数据链路层的封装

- 由于封装IGMP的IP分组有一个多播IP地址，ARP协议不能找到在数据链路层转发该分组的对应的MAC（物理）地址；
- 接着发生的事情依赖于下面的数据链路层是否支持物理多播地址；
- 大多数局域网支持物理多播寻址，以太网就是其中的一种；
- 以太网的物理地址（MAC地址）是6字节（48位），如果以太网地址的前25位是00000001000000000010111100，它定义TCP/IP协议的物理多播地址，余下的23位可用来定一个组；
- 要将IP多播地址转换为以太网地址，多播路由器提取D类IP地址最低23位，并将其插入到以太网物理地址中（见下页）；
- 但D类IP地址的组标识符是28位长，这表示32个IP级的多播地址映射到一个多播地址，映射是多对一，为此，主机必须检查IP地址，并丢弃任何不属于它的分组。

图21.21 将D类地址映射到以太网物理地址



以太网的多播物理地址范围：
01:00:5E:00:00:00-01:00:5E:7F:FF:FF

例21.7

将多播IP地址230.43.14.7转换成以太网多播物理地址。

解：

这可用2个步骤来完成：

a. 用十六进制写出IP地址的最右23位。这可将最右边3个字节变换成十六进制，然后如果最左边的数字大于或等于8，则将该数减去8。在本例中，其结果是2B:0E:07。

b. 将a步所得的结果加到开始的以太网多播地址01:00:5E:00:00:00，其结果是01:00:5E:2B:0E:07。

例21.8

将多播IP地址238.212.24.9转换成以太网多播地址。

解：

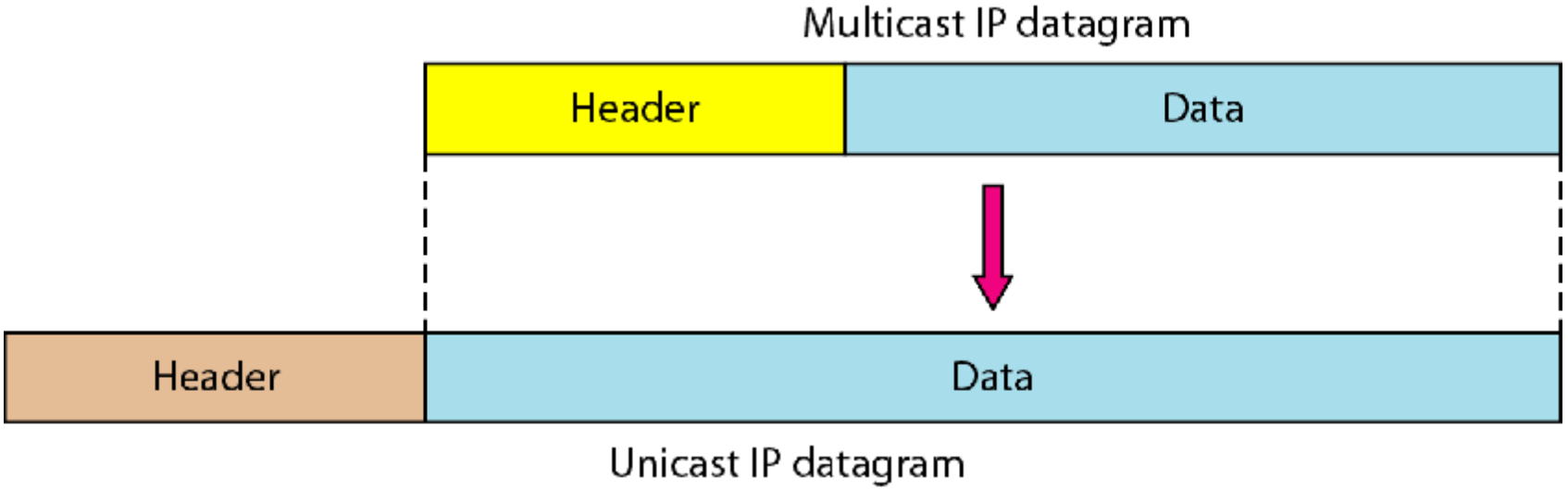
a.用十六进制表示的最右边3个字节是D4:18:09。我们需将最左边的数字减去8，其结果是54:18:09。

b.将a步所得的结果加上以太网多播的开始地址，其结果是：01:00:5E:54:18:09。

图21.22 隧道技术

p大多数广域网不支持物理多播地址，要通过这样的网络发送多播分组，就要使用称为隧道技术的过程；

p在隧道技术（**tunneling**）中，多播分组封装成单播分组并通过网络发送，然后在另一端，这个分组又转换成多播分组



例21.9

我们使用带有三个选项的netstat命令。
选项-n是以数字形式显示IP地址、选项-r显示路由表、选项-a显示所有的地址（单播和多播地址）。注意：这里仅给出与我们讨论有关的项目。“Destination”定义目的地址，“Gateway”定义路由器，“Iface”定义接口，“Mask”定义掩码，“Flag”定义标记。标记U表示该路由可使用，标志G表示该路由是一个网关（路由器）。

注意：多播地址用彩色表示。具有从240.0.0.0到239.255.255.255多播地址的任何分组都被masked，并传递给以太网接口。

例21.9（续）

```
$ netstat -nra
Kernel IP routing table
```

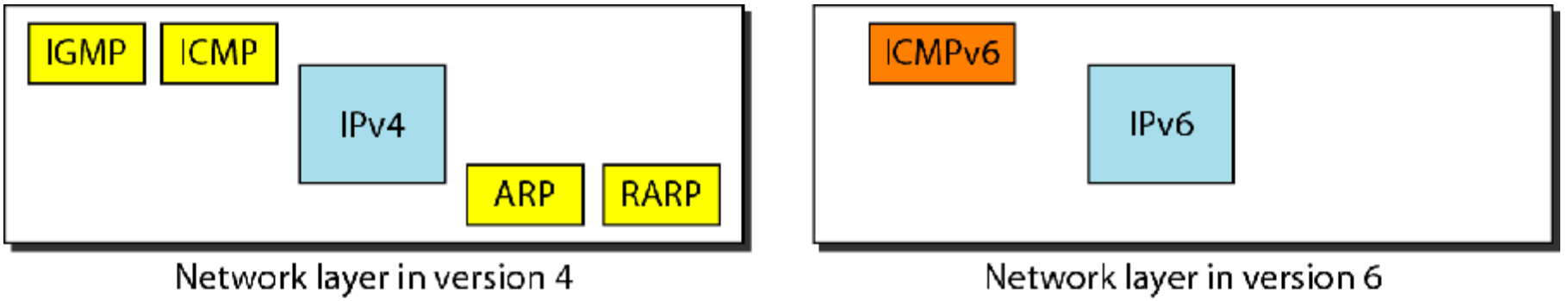
Destination	Gateway	Mask	Flags	Iface
153.18.16.0	0.0.0.0	255.255.240.0	U	eth0
169.254.0.0	0.0.0.0	255.255.0.0	U	eth0
127.0.0.0	0.0.0.0	255.0.0.0	U	lo
224.0.0.0	0.0.0.0	224.0.0.0	U	eth0
0.0.0.0	153.18.31.254	0.0.0.0	UG	eth0

21-4 ICMPv6

- 在TCP/IP协议族的版本6中被修改的另一个协议是ICMP（ICMPv6），这个新版本与版本4的策略和目的一样；
- 修改ICMPv4使得它更加适合IPv6；
- 此外，在版本4中的一些独立的协议现在成为网际控制报文协议ICMPv6的一部分。

图21.23 版本4和版本6的网络层比较

- 版本4中的ARP和IGMP协议合并到ICMPv6；
- RARP协议从这个协议族中取消了，因为它很少使用，而且BOOTP具有与RARP相同的功能；
- 正如ICMPv4那样，将ICMP报文分成两大类，但每一类都比前面版本有更多的报文类型。



ICMPv6差错报告

- 在版本6中源端抑制报文取消了，因为优先级和流标号字段允许路由器控制拥塞，并将最不重要的报文丢弃，不需要通知发送方将速率放慢；
- 增加了分组太大报文，因为在IPv6中分段是发送方的责任；如果发送方没有对分组长度做出正确的判断，路由器就没有任何选择，而只能丢弃分组，并发送差错报告报文给发送方。

表21.3 ICMPv4和ICMPv6的差错报告的比较

Type of Message	*Version 4*	*Version 6*
Destination unreachable	Yes	Yes
Source quench	Yes	No
Packet too big	No	Yes
Time exceeded	Yes	Yes
Parameter problem	Yes	Yes
Redirection	是	Yes

ICMPv6查询报文

- 已定义了4组查询报文：回送请求和回答、路由器询问和通知、邻站询问和通知（ARP）以及组成员关系（IGMP）；
- ICMPv6取消了两组查询报文：时间戳请求和回答以及地址掩码请求和回答；
- 取消时间戳请求和回答是因为它们在其他协议如TCP中已实现了，同时也因为过去没有使用过它们；
- 在IPv6中取消地址掩码请求和回答报文是因为地址的子网部分允许用户使用多达$2^{32}-1$个子网，因此，定义在IPv4中的子网掩码在这里是不需要的。

表21.4 ICMPv4和ICMPv6中查询报文的比较

Type of Message	*Version 4*	*Version 6*
Echo request and reply	Yes	Yes
Timestamp request and reply	Yes	No
Address-mask request and reply	Yes	No
Router solicitation and advertisement	Yes	Yes
Neighbor solicitation and advertisement	ARP	Yes
Group membership	IGMP	Yes