第 4 章 网络层

第 4 章网络层

- 4.1网络层提供的两种服务
- 4.2网际协议 IP
- 4.3划分子网和构造超网
- 4.4网际控制报文协议 ICMP
- 4.5互联网的路由选择协议
- 4.6 IP 多播

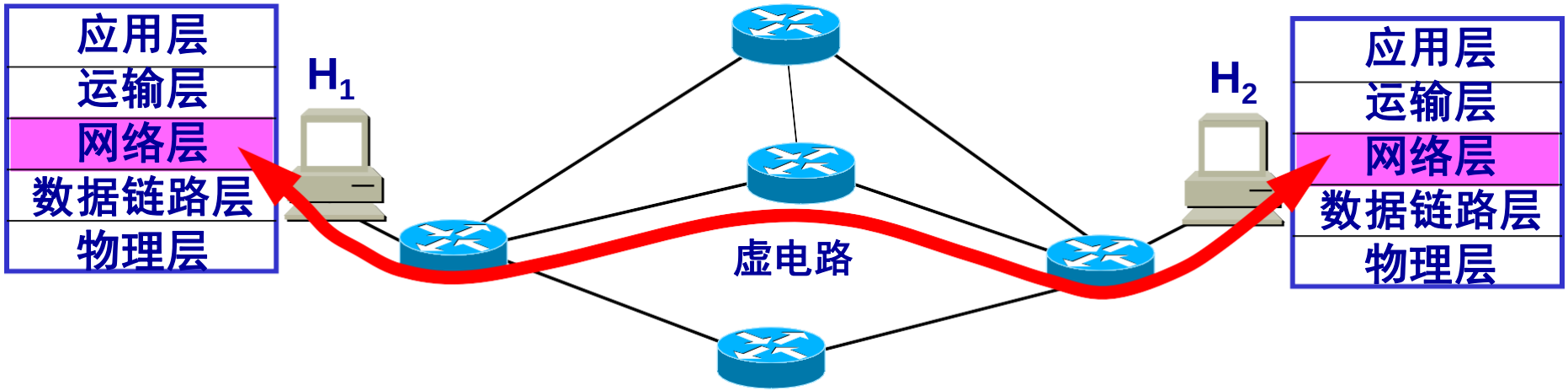
4.1 网络层提供的两种服务

- 在计算机网络领域，网络层应该向运输层提供怎样的服务（“面向连接”还是“无连接”）曾引 起了长期的争论。
- 争论焦点的实质就是：在计算机通信中，可靠交付应当由谁来负责？是网络还是端系统？

一种观点：让网络负责可靠交付

- 这种观点认为，应借助于电信网的成功经验，让 网络负责可靠交付，计算机网络应模仿电信网络， 使用**面向连接**的通信方式。
- 通信之前先建立**虚电路** (Virtual Circuit)，以保证双方通信所需的一切网络资源。
- 如果再使用**可靠传输**的网络协议，就可使所发送 的分组无差错按序到达终点，不丢失、不重复

虚电路服务



H_1 发送给 H_2 的所有分组都沿着同一条虚电路传送

虚电路是逻辑连接

- 虚电路表示这只是一条逻辑上的连接，分组都沿 着这条逻辑连接按照存储转发方式传送，而并不 是真正建立了一条物理连接。
- 请注意，电路交换的电话通信是先建立了一条真正的连接。
- 因此分组交换的虚连接和电路交换的连接只是类似，但并不完全一样。
- 数据链路：相邻两节点间的数据传输通道。
- 逻辑信道：相邻两点间的一条数据链路可支持多条逻辑信道，为多对通信服务。

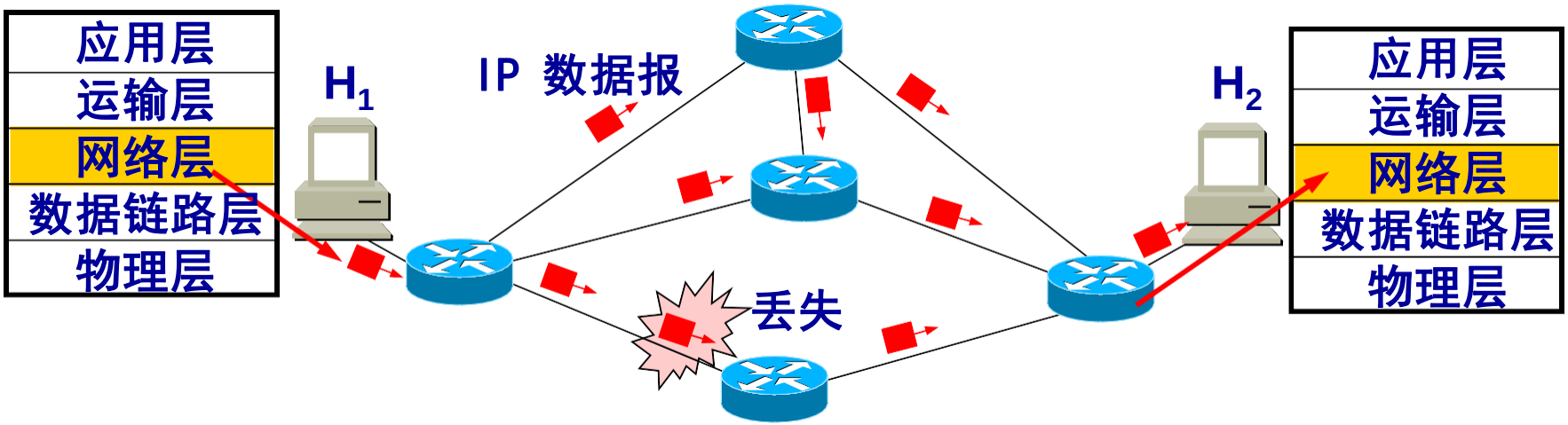
另一种观点：网络提供数据报服务

- 互联网的先驱者提出了一种崭新的网络设计思路。
- 网络层向上只提供简单灵活的、无连接的、尽最大努力交付(不可靠)的数据报服务，对源主机没有任何承诺。
- 网络在发送分组时不需要先建立连接。每一个分组（即 IP数据报）携带完整的地址信息，独立传输，独立寻址，彼此之间不需要保持任何的顺序关系（不进行编号）。
- 网络层不提供服务质量的承诺。即所传送的分组可能出错、丢失、重复和失序（不按序到达终点），当然也不保证分组传送的时限。

尽最大努力交付

- 由于传输网络不提供端到端的可靠传输服务，这就使网络中的路由器可以做得比较简单，而且价格低廉（与电信网的交换机相比较）。
- 如果主机（即端系统）中的进程之间的通信需要是可靠的，那么就由网络的主机中的运输层负责可靠交付（包括差错处理、流量控制等）。
- 采用这种设计思路的好处是：网络的造价大大降低，运行方式灵活，能够适应多种应用。
- 互联网能够发展到今日的规模，充分证明了当初采用这种设计思路的正确性。

数据报服务



H_1 发送给 H_2 的分组可能沿着不同路径传送

两种服务的思路来源不同

- 虚电路服务的思路来源于传统的电信网。
  - 电信网负责保证可靠通信的一切措施，因此电信网的结点交换机复杂而昂贵。
- 数据报服务 力求使 网络生存性好 和 对网络的控制功能分散，因而只能要求网络提供尽最大努力的服务。
  - 可靠通信由用户终端中的软件（即TCP）来保证。

虚电路服务与数据报服务的优缺点

- **传送代价方面**
  - 网络上传送的报文长度，在很多情况下都很短。
  - 用数据报既迅速又经济。
  - 若用虚电路，为了传送一个分组而建立虚电路和释放虚电路就显得太浪费网络资源了。

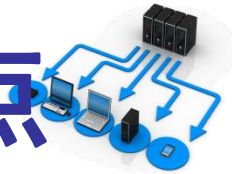
虚电路服务与数据报服务的优缺点

- **交换节点存储转发方面**
  - 在使用数据报时，每个分组必须携带完整的地址信息。
  - 在使用虚电路的情况下，每个分组不需要携带完整的目的地址，而仅需要有个很简单的虚电路号码的标志。
  - 这就使分组的控制信息部分的比特数减少，因而减少了额外开销。

虚电路服务与数据报服务的优缺点

- **差错和流量控制方面**
  - 在使用数据报时，主机承担端到端的差错控制和流量控制。
  - 在使用虚电路时，分组按顺序交付，网络可以负责差错控制和流量控制。

虚电路服务与数据报服务的优缺点



- 使用场合
  - 数据报服务对军事通信有其特殊的意义。当某个结点发生故障时，后续的分组就可另选路由，因而提高了可靠性。
  - 但在使用虚电路时，结点发生故障就必须重新建立另一条虚电路。
  - 数据报服务还很适合于将一个分组发送到多个地址(即广播或多播)。

虚电路服务与数据报服务的对比

对比的方面	虚电路服务	数据报服务
思路	可靠通信应当由网络来保证	可靠通信应当由用户主机来保证
连接的建立	必须有	不需要
终点地址	仅在连接建立阶段使用，每个分组使用短的虚电路号	每个分组都有终点的完整地址
分组的转发	属于同一条虚电路的分组均按照同一路由进行转发	每个分组独立选择路由进行转发
当结点出故障时	所有通过出故障的结点的虚电路均不能工作	出故障的结点可能会丢失分组，一些路由可能会发生变化
分组的顺序	总是按发送顺序到达终点	到达终点时不一定按发送顺序
端到端的差错处理和流量控制	可以由网络负责，也可以由用户主机负责	由用户主机负责

4.2 网际协议 IP

- 4.2.1 虚拟互连网络
- 4.2.2 分类的 IP 地址
- 4.2.3 IP 地址与硬件地址
- 4.2.4 地址解析协议 ARP
- 4.2.5 IP 数据报的格式
- 4.2.6 IP 层转发分组的流程

4.2 网际协议 IP (Internet Protocol)

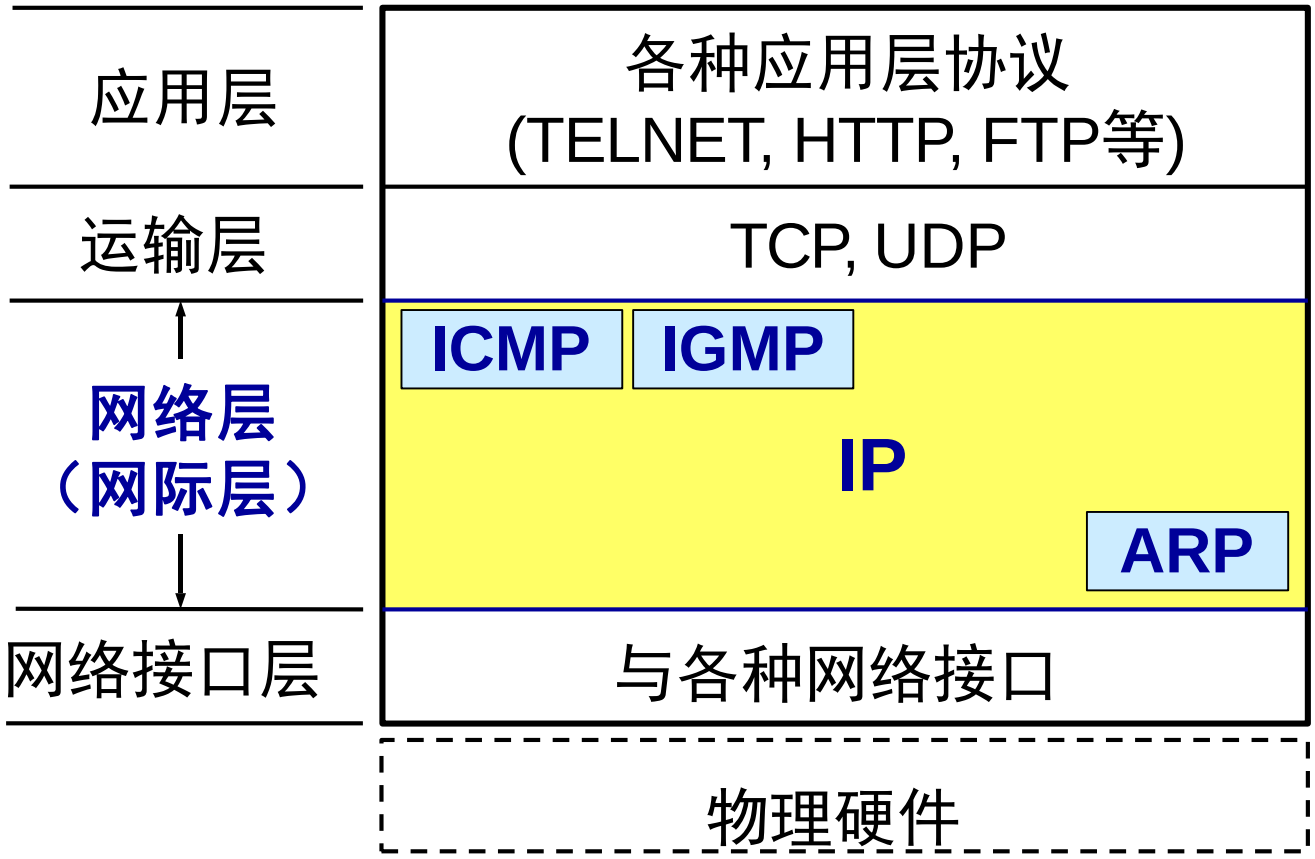
- 网际协议 IP 是 TCP/IP 体系中两个最主要的协议之一。
- 与 IP 协议配套使用的还有三个协议：
  - 地址解析协议 ARP

    (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



4.2.1 虚拟互连网络

- 将网络互连并能够互相通信，会遇到许多问题需要解决，如：
  - 不同的寻址方案
  - 不同的最大分组长度
  - 不同的网络接入机制
  - 不同的超时控制
  - 不同的差错恢复方法
  - 不同的状态报告方法
  - 不同的路由选择技术
  - 不同的用户接入控制
  - 不同的服务（面向连接服务和无连接服务）
  - 不同的管理与控制方式等

如何将异构的网络互相连接起来？

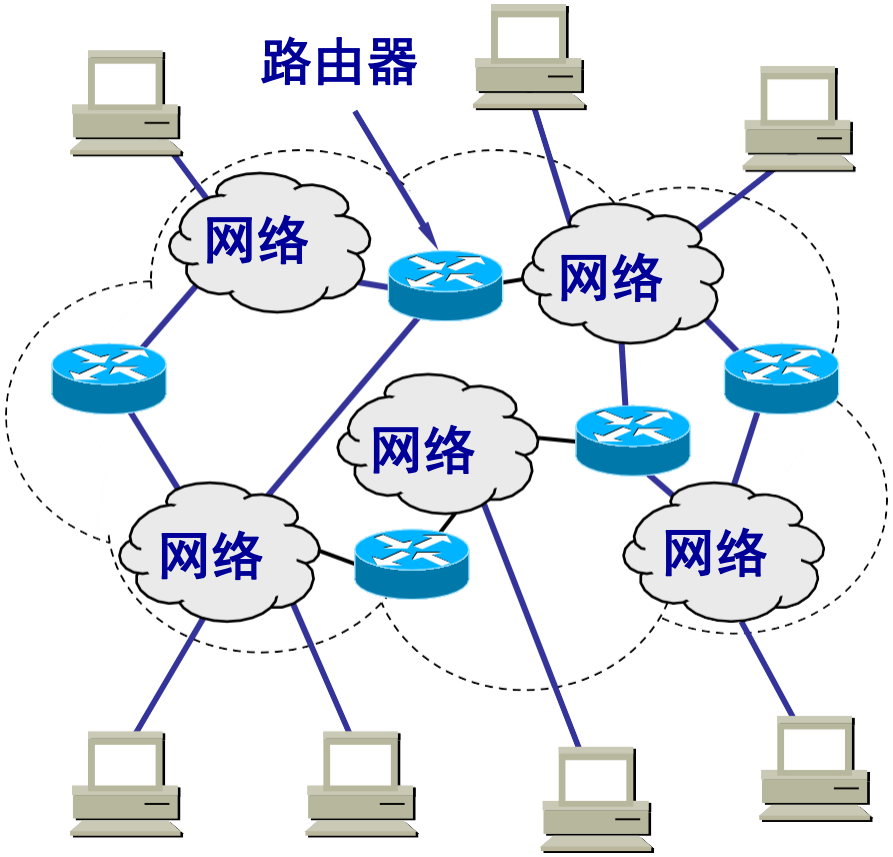
使用一些中间设备进行互连

- 将网络互相连接起来要使用一些中间设备。
- 中间设备又称为**中间系统**或**中继 (relay)系统**。
- 有以下五种不同的中间设备：
  - **物理层**中继系统：**转发器** (repeater)。
  - **数据链路层**中继系统：**网桥** 或 **桥接器** (bridge)。
  - **网络层**中继系统：**路由器** (router)。
  - 网桥和路由器的**混合物**：**桥路器** (brouter)。
  - **网络层以上**的中继系统：**网关** (gateway)。

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，因为这仅仅是把一个网络扩大了，而 这仍然是一个网络。
- 网络互连都是指用路由器进行网络互连和路由选择。(互连网：网络的网络)
- 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。

互连网络与虚拟互连网络



（a） 互连网络

（b） 虚拟互连网络

虚拟互连网络的意义

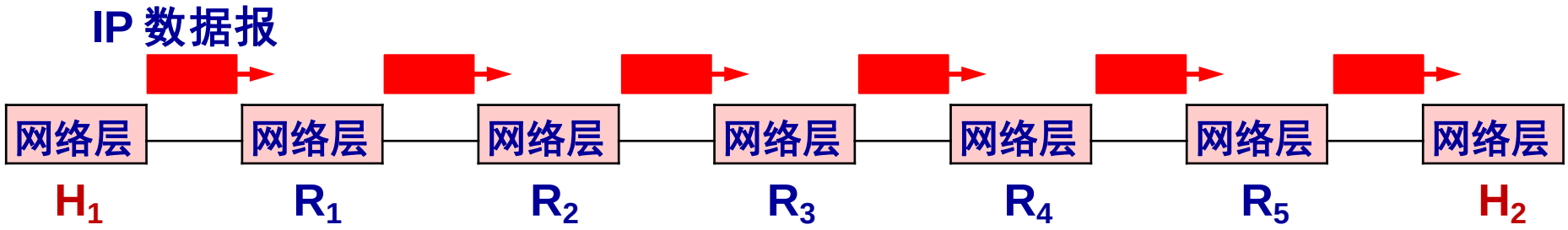
- 所谓虚拟互连网络也就是逻辑互连网络，它的意思就是 互连起来的各种物理网络的异构性本来是客观存在的，但是我们利用 IP 协议就可以使这些性能各异的网络在 网络层上看起来好像是一个统一的网络。
- 使用 IP 协议的虚拟互连网络可简称为 IP 网。
- 使用虚拟互连网络的好处是：当互联网上的主机进行通 信时，就好像在一个网络上通信一样，而看不见互连的 各具体的网络异构细节（如编址方案、路由选择协议等）。
- 如果在这种覆盖全球的 IP 网的上层使用 TCP 协议，那么就是现在的互联网 (Internet)。

分组在互联网中的传送

互联网可以由多种异构网络互连组成。

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。



4.2.2 分类的 IP 地址

- 在 TCP/IP 体系中，IP 地址是一个最基本的概念。
- 本部分重点学习：
  - 1. IP 地址及其表示方法
  - 2. 常用的三种类别的 IP 地址

1. IP 地址及其表示方法

- 我们把整个因特网看成为一个单一的、抽象的网络。
- IP 地址就是给每个连接在互联网上的主机（或路由器）分配一个在全世界范围是唯一的 32 位的标识符。
- IP 地址现在由互联网名字和数字分配机构 ICAN(Internet Corporatiofor Assigned Names and Numbers)进行分配。

IP 地址的编址方法

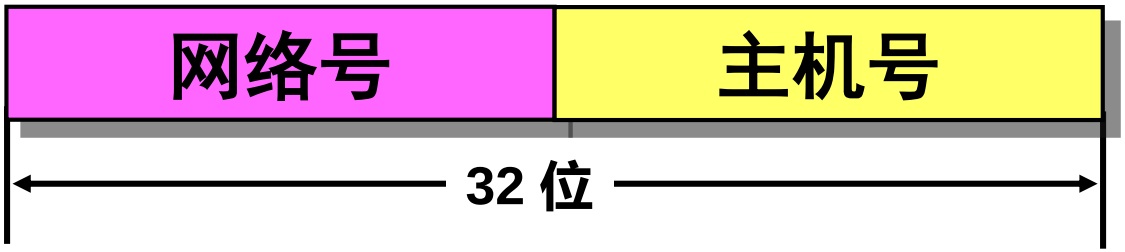
- IP地址的编址方法共经过了三个历史阶段：
- 分类的 IP 地址。这是最基本的编址方法，在 1981 年就通过了相应的标准协议。
- 子网的划分。这是对最基本的编址方法的改进，其标准 [RFC 950] 在 1985 年通过。
- 构成超网。这是比较新的无分类编址方法。1993 年提出后很快就得到推广应用。

分类 IP 地址

- 将IP地址划分为若干个固定类（A - E类）。
- 每一类地址都由两个固定长度的字段组成，其中 一个字段是网络号 net-id，它标志主机（或路由器）所连接到的网络，而另一个字段则是主机号 host-id，它标志该主机（或路由器）。
- 主机号在它前面的网络号所指明的网络范围内必须是唯一的。
- 由此可见，一个 IP 地址在整个互联网范围内是唯一的。

分类 IP 地址

- 这种两级的 IP 地址结构如下：

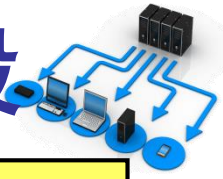


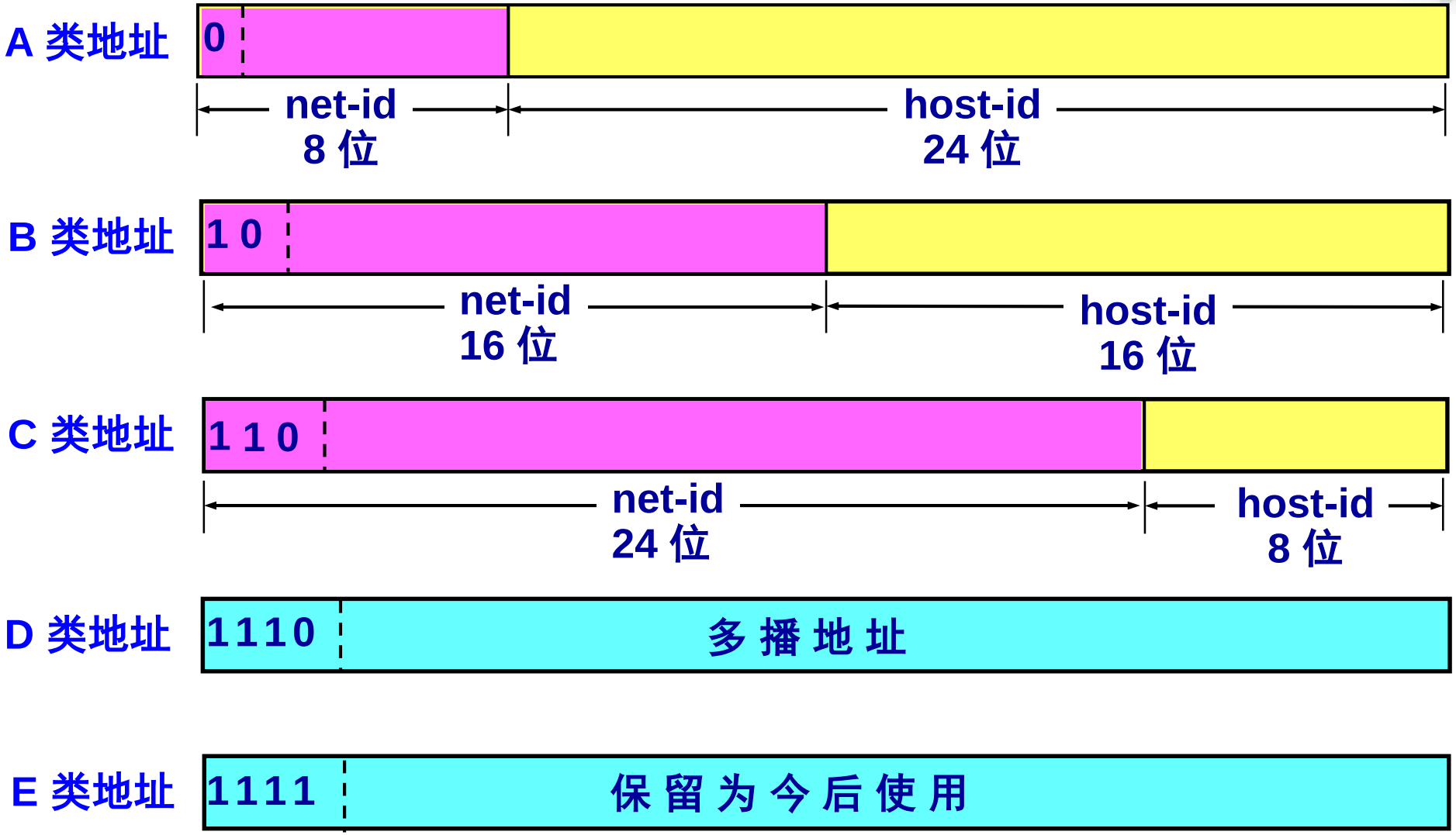
- 这种两级的 IP 地址可以记为：

IP 地址 ::= { <网络号>, <主机号>}　　(4-1)

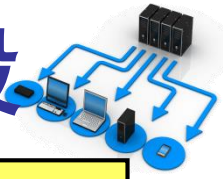
::= 代表“定义为”

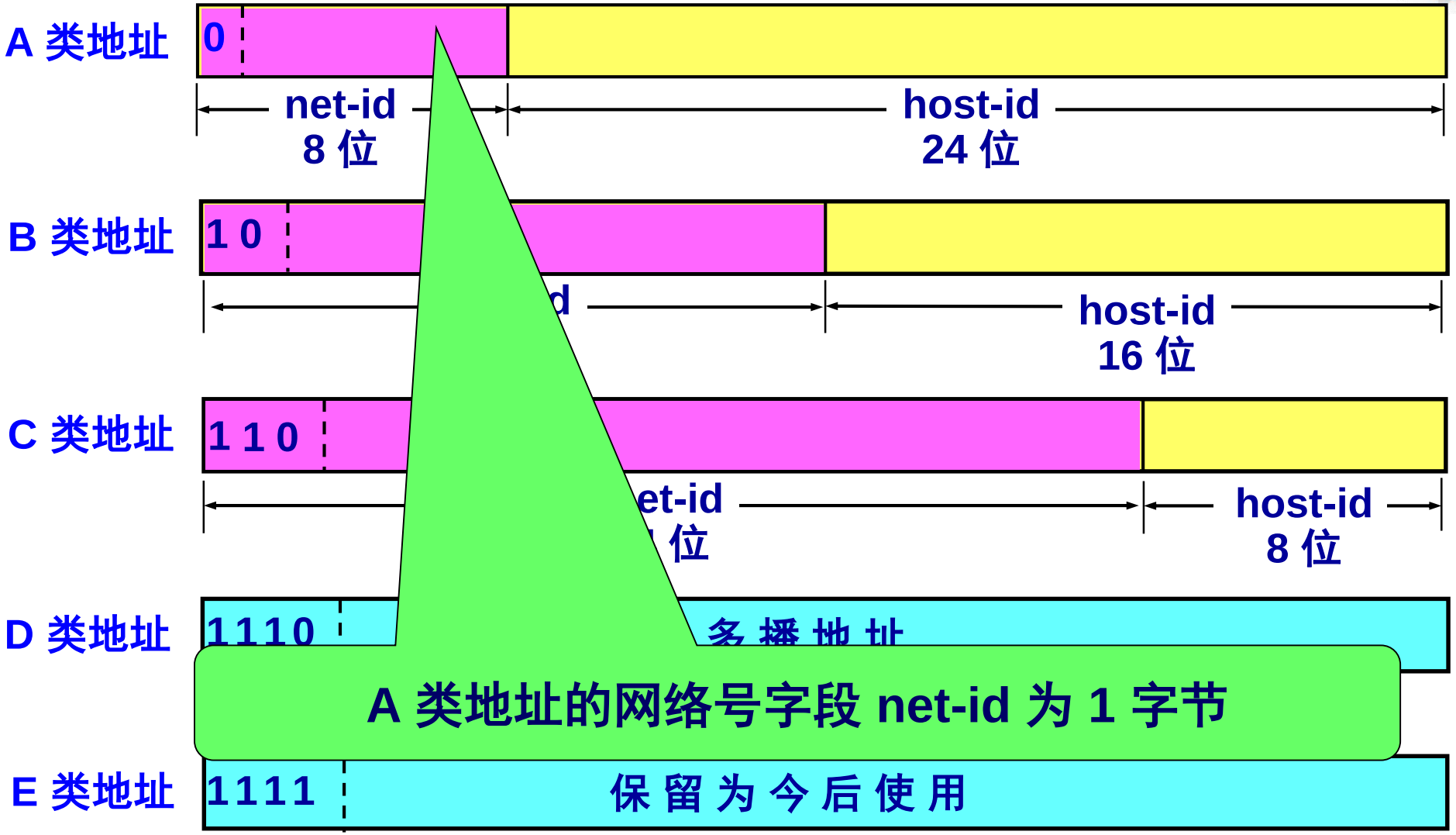
各类 IP 地址的网络号字段和主机号字段



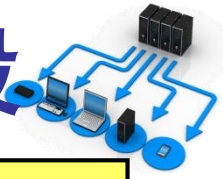


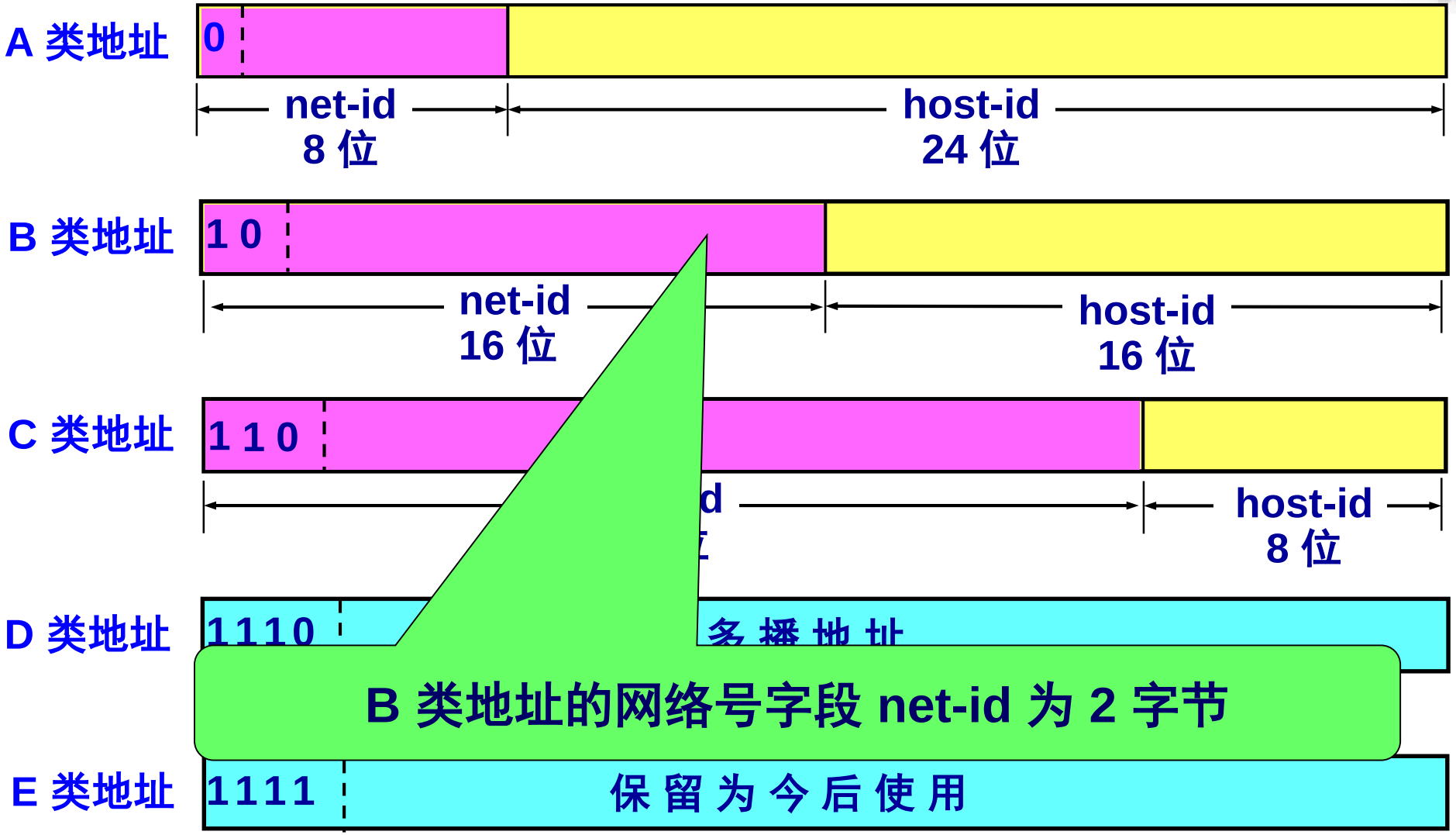
各类 IP 地址的网络号字段和主机号字段



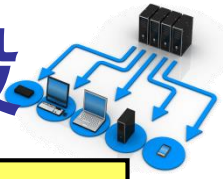


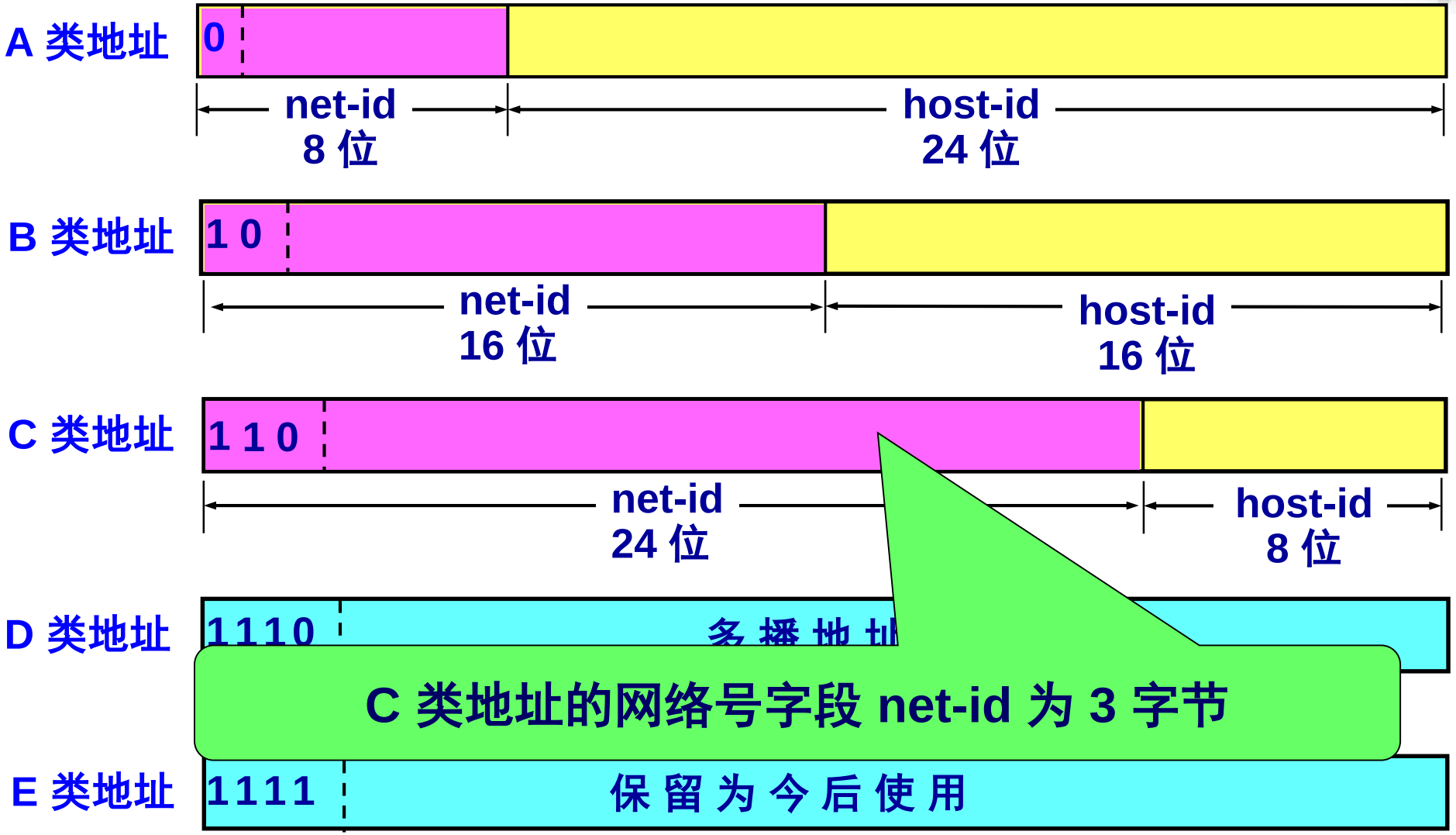
各类 IP 地址的网络号字段和主机号字段





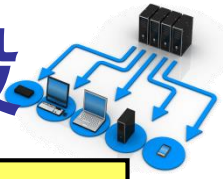
各类 IP 地址的网络号字段和主机号字段

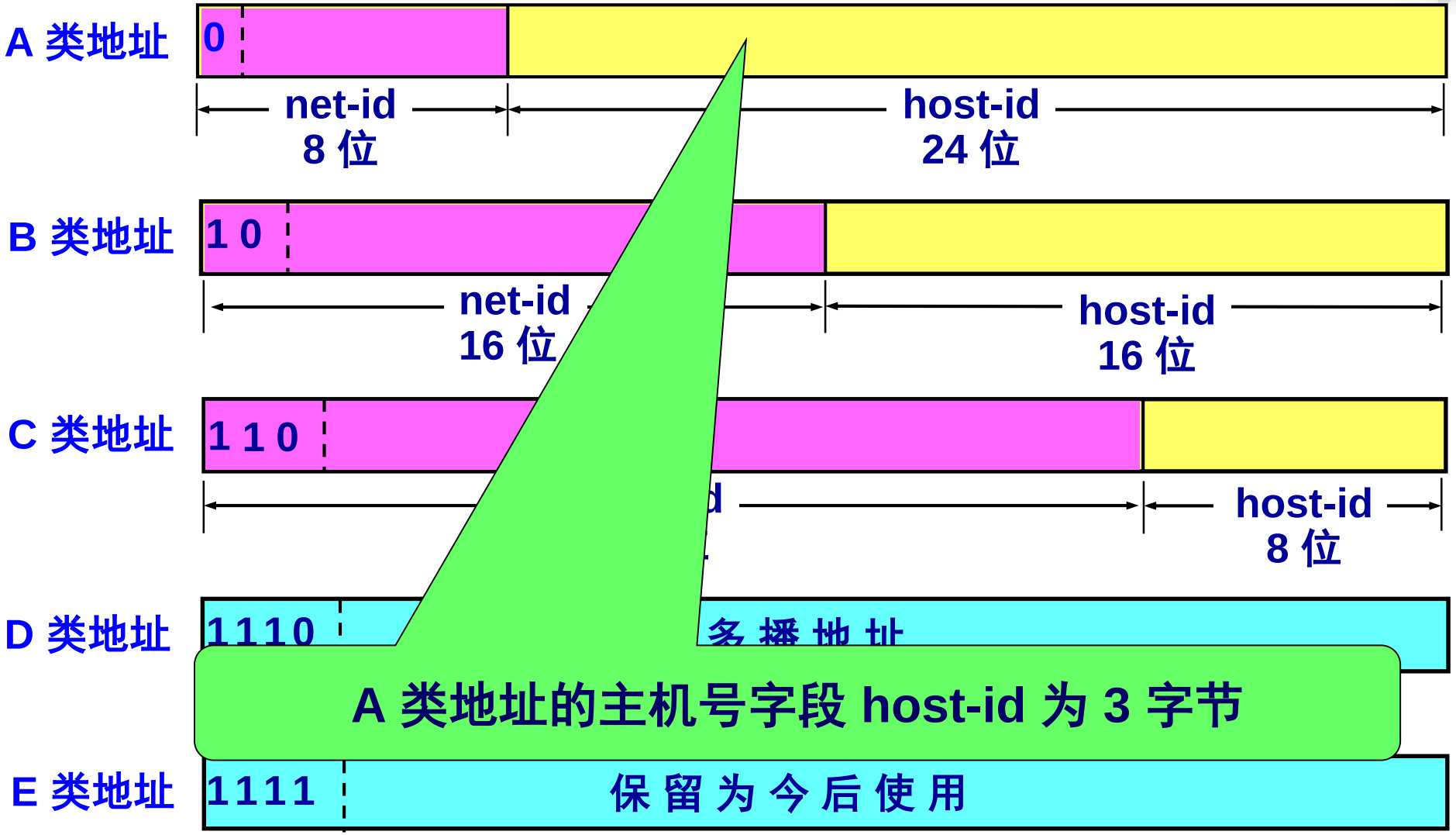




C 类地址的网络号字段 net-id 为 3 字节

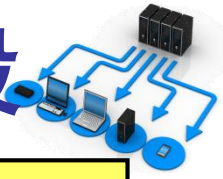
各类 IP 地址的网络号字段和主机号字段

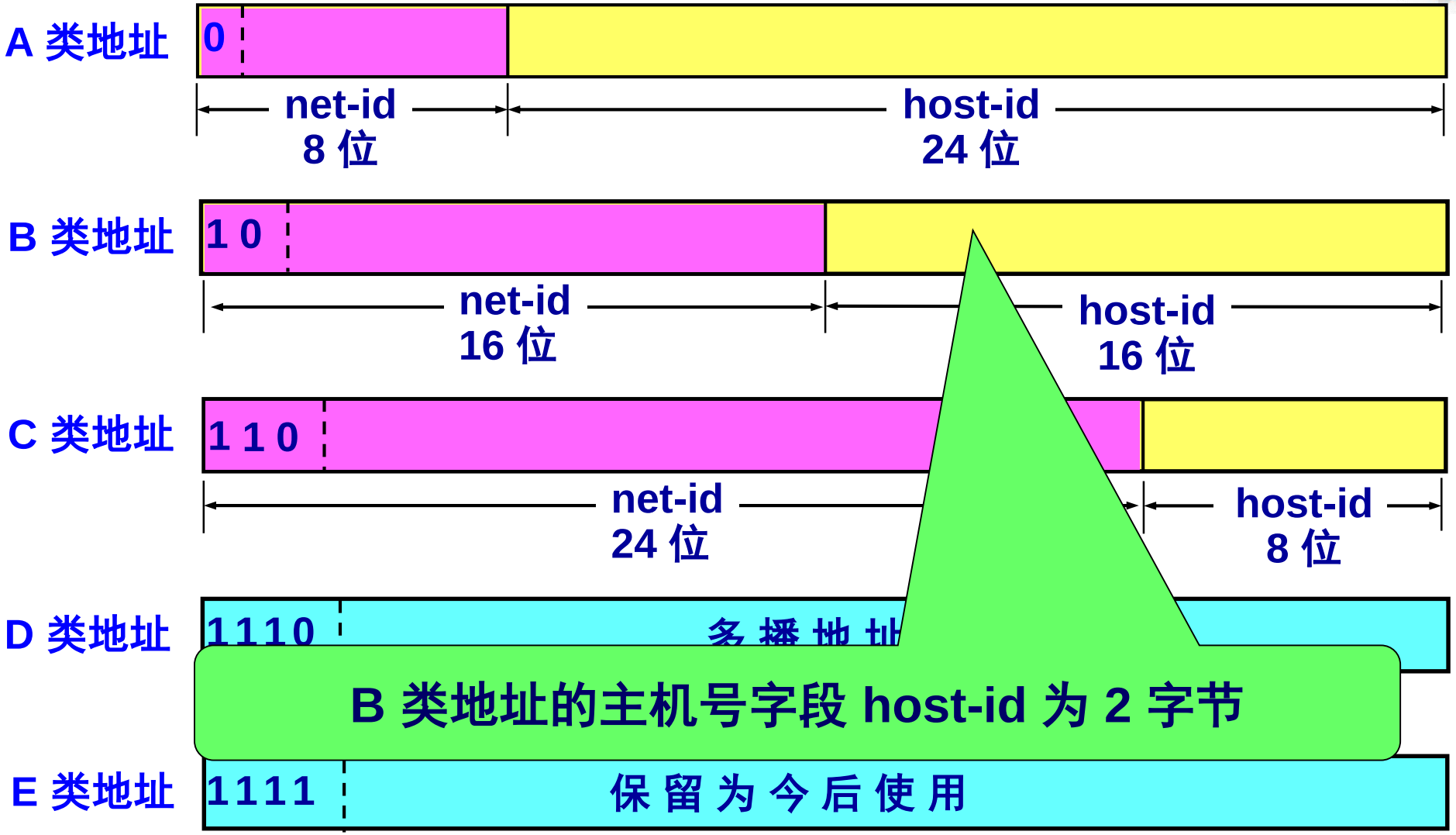




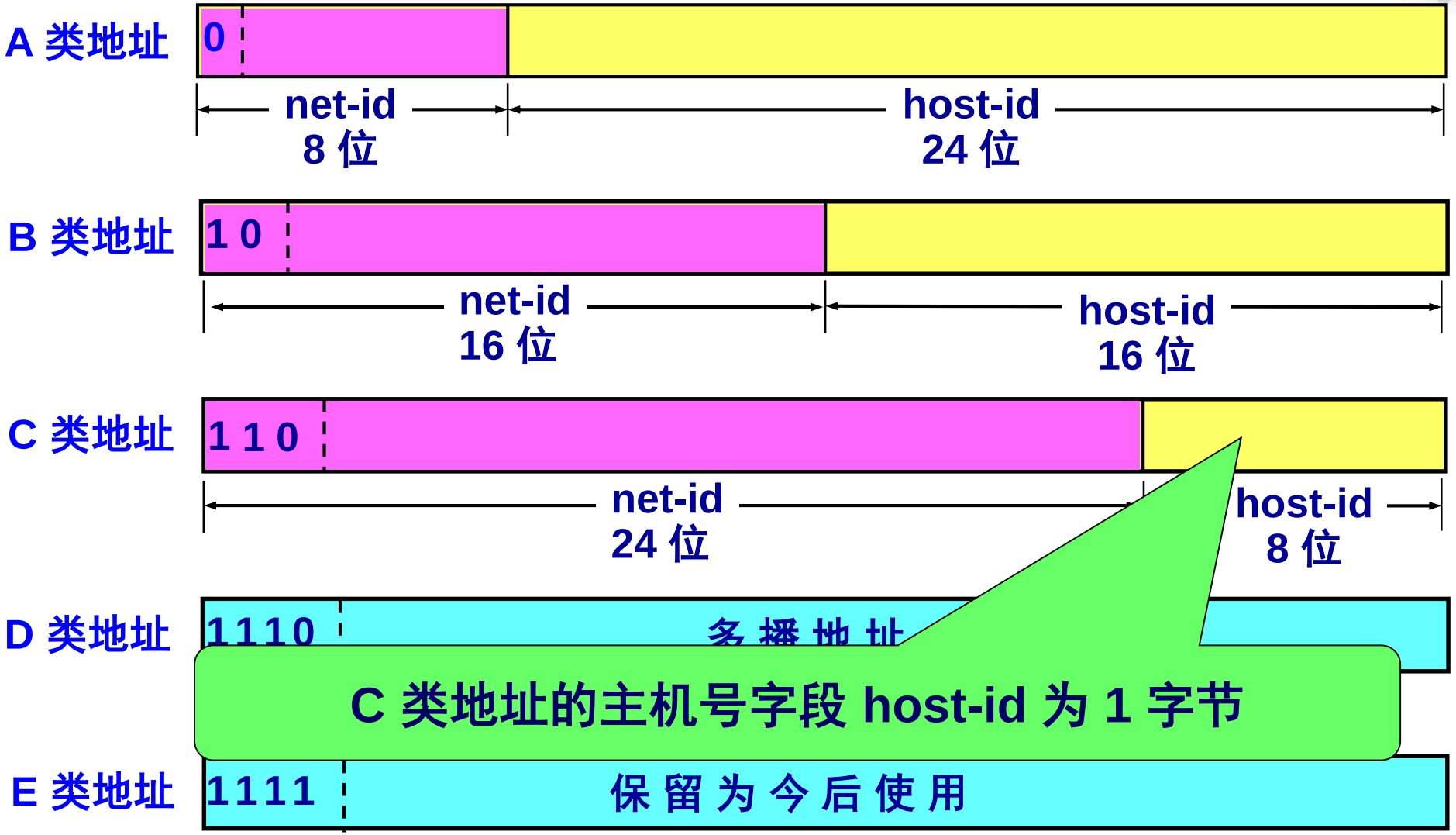
A 类地址的主机号字段 host-id 为 3 字节

各类 IP 地址的网络号字段和主机号字段



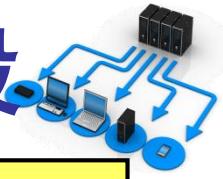


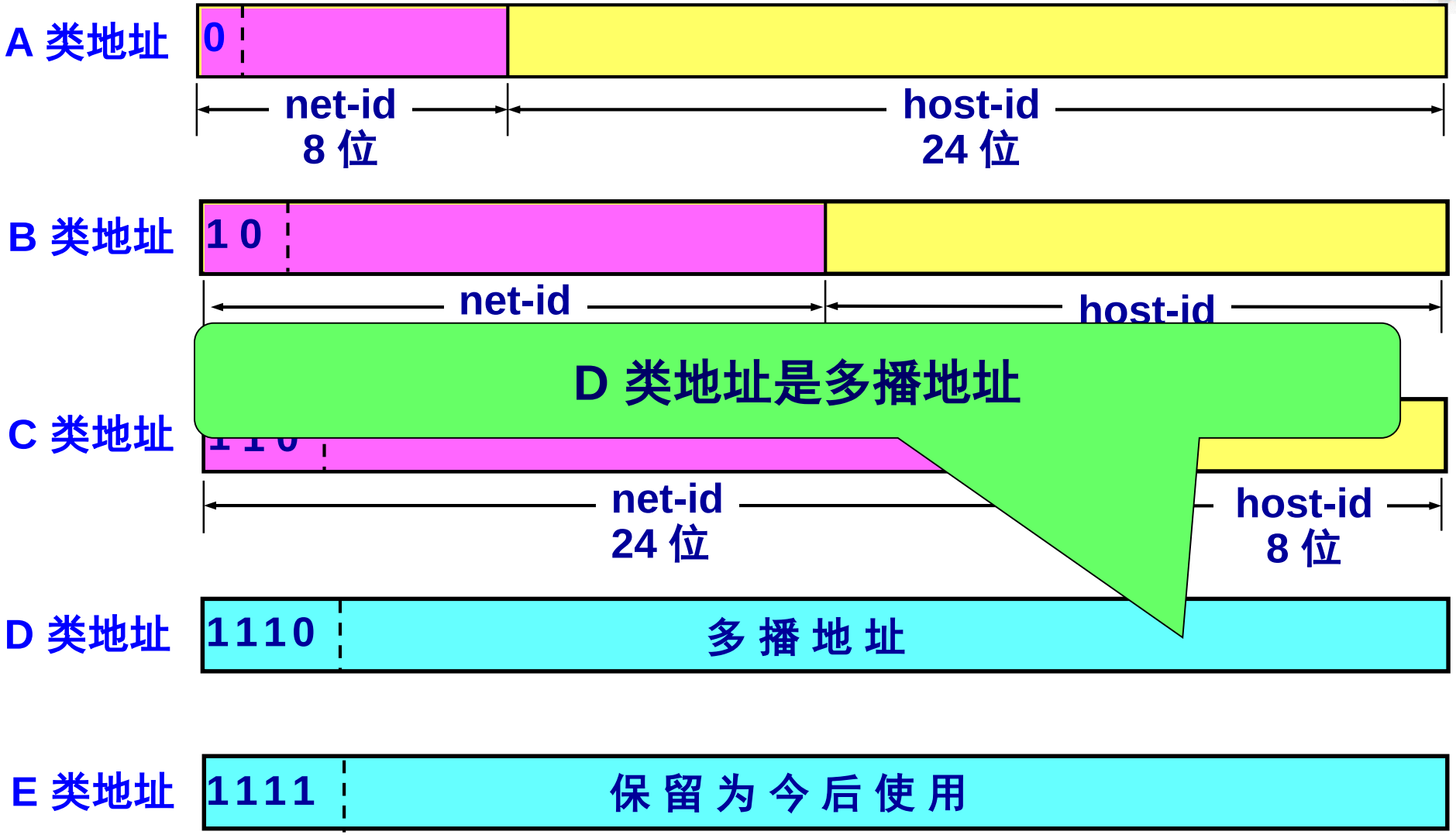
各类 IP 地址的网络号字段和主机号字段



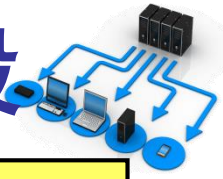
C 类地址的主机号字段 host-id 为 1 字节

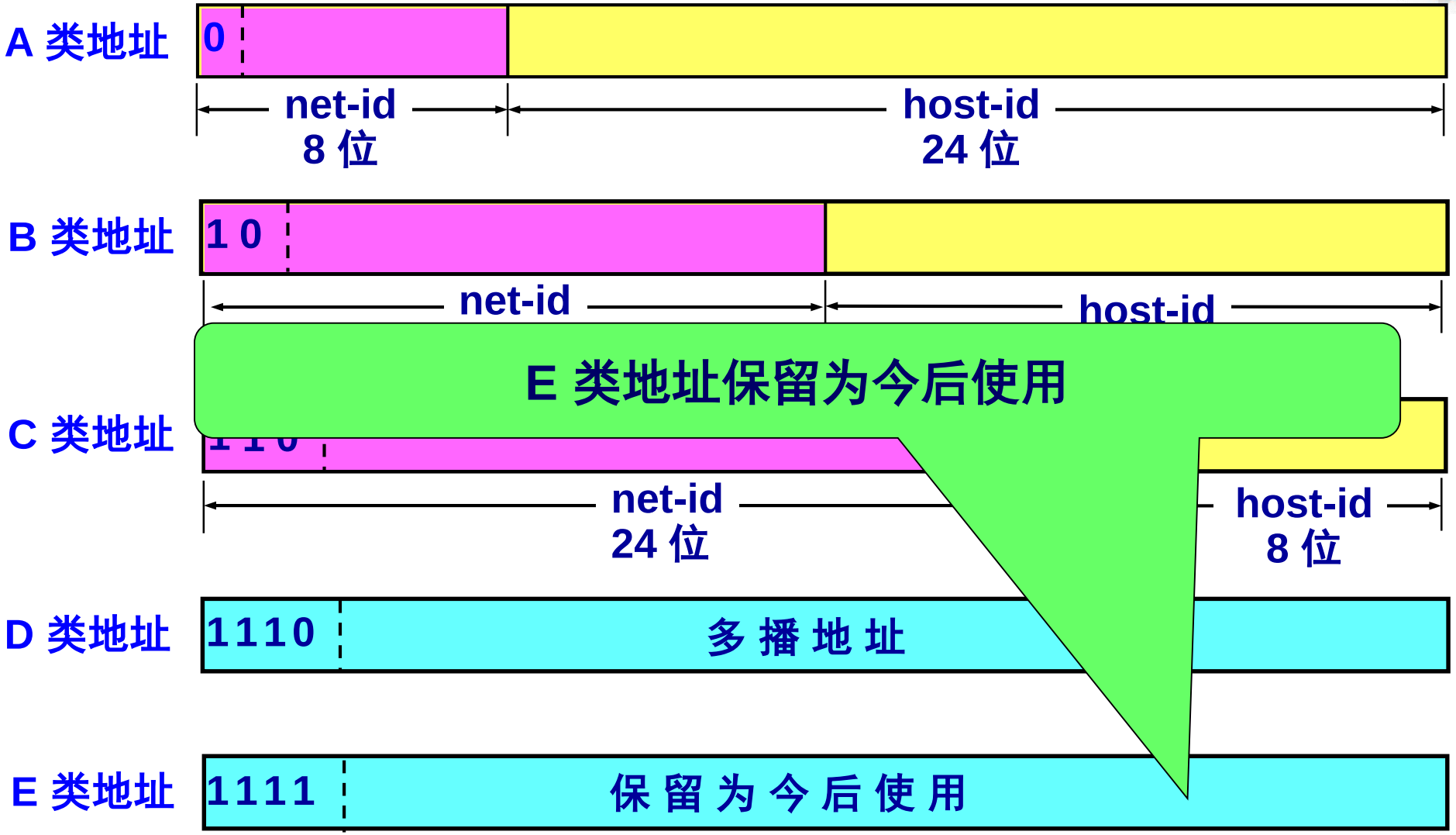
各类 IP 地址的网络号字段和主机号字段





各类 IP 地址的网络号字段和主机号字段





IP地址分类的好处

- 划分成不同类别的考虑
  - 各种网络差异很大，有的网络拥有很多主机，而有的网络拥有的主机数目很少。
  - 将IP地址划分成不同类别A、B、C可以满足不同用户的需求。
  - 当一个单位申请到一个IP地址时，只是申请了一个网络号Net-id，具体的主机号由各个单位自行分配。
- D类和E类使用较少
  - D类的多播地址主要留给IAB（因特网体系结构委员会）使用。

点分十进制记法

机器中存放的 IP 地址
是 32 位二进制代码

10000000000010110000001100011111

每 8 位为一组

10000000 00001011 00000011 00011111

将每 8 位的二进制数
转换为十进制数

128 11 3 31

采用点分十进制记法
则进一步提高可读性

128.11.3.31

点分十进制记法举例

32 位二进制数	等价的 点分十进制数
10000001 00110100 00000110 00000000	129.52.6.0
11000000 00000101 00110000 00000011	192.5.48.3
00001010 00000010 00000000 00100101	10.2.0.37
10000000 00001010 00000010 00000011	128.10.2.3
10000000 10000000 11111111 00000000	128.128.255.0

2. 常用的三种类别的 IP 地址

IP 地址的指派范围

网络类别	最大可指派的网络数	第一个可指派的网络号	最后一个可指派的网络号	每个网络中最大主机数
A	126 (2^7 – 2)	1	126	16777214 (2^{24}-2)
B	16383 (2^{14} – 1)	128.1	191.255	65534 (2^{16}-2)
C	2097151 (2^{21} – 1)	192.0.1	223.255.255	254 (2^8-2)

三种类别的 IP 地址

- 三个类别的IP地址中，2个特殊的Host-id含义：
  - 全“0”的Host-id表示该IP地址是“本主机”所连接 的单个网络地址。
    - 如IP地址为5.6.7.8，则网络地址为5.0.0.0；
  - 全“1”的Host-id表示“所有(all)”，即该网络上的所 有主机。

三种类别的 IP 地址

- A类地址
  - 网络号为126个（2^7-2）：减2的原因是什么？
  - (1) IP地址的全“0”表示“这个（This）”。
  - (2) Net-id为全“0”，是保留地址，表示“本网络”。
  - (3) Net-id为127，为本地软件的环路测试，本主机使用。
  - 全部的IP地址为2^{32}个，A类IP地址共有2^{31}占50%。
- B类地址：共有2^{30}个，占25%。
- C类地址：共有2^{29}个，占12.5%。

一般不使用的特殊的 IP 地址

网络号	主机号	源地址使用	目的地址使用	代表的意思
0	0	可以	不可	在本网络上的本主机（见 6.6 节 DHCP 协议）
0	host-id	可以	不可	在本网络上的某台主机 host-id
全 1	全 1	不可	可以	只在本网络上进行广播（各路由器均不转发）
net-id	全 1	不可	可以	对 net-id 上的所有主机进行广播
127	非全 0 或全 1 的任何数	可以	可以	用作本地软件环回测试之用

IP 地址的一些重要特点

- **(1) IP 地址是一种分等级的地址结构。**分两个等级的好处是：
  - **第一，**IP 地址管理机构在分配 IP 地址时只分配网络号，而剩下的主机号则由得到该网络号的单位自行分配。这样就方便了 IP 地址的管理
  - **第二，**路由器仅根据目的主机所连接的网络号来转发分组（而不考虑目的主机号），这样就可以使路由表中的项目数大幅度减少，从而减小了路由表所占的存储空间。

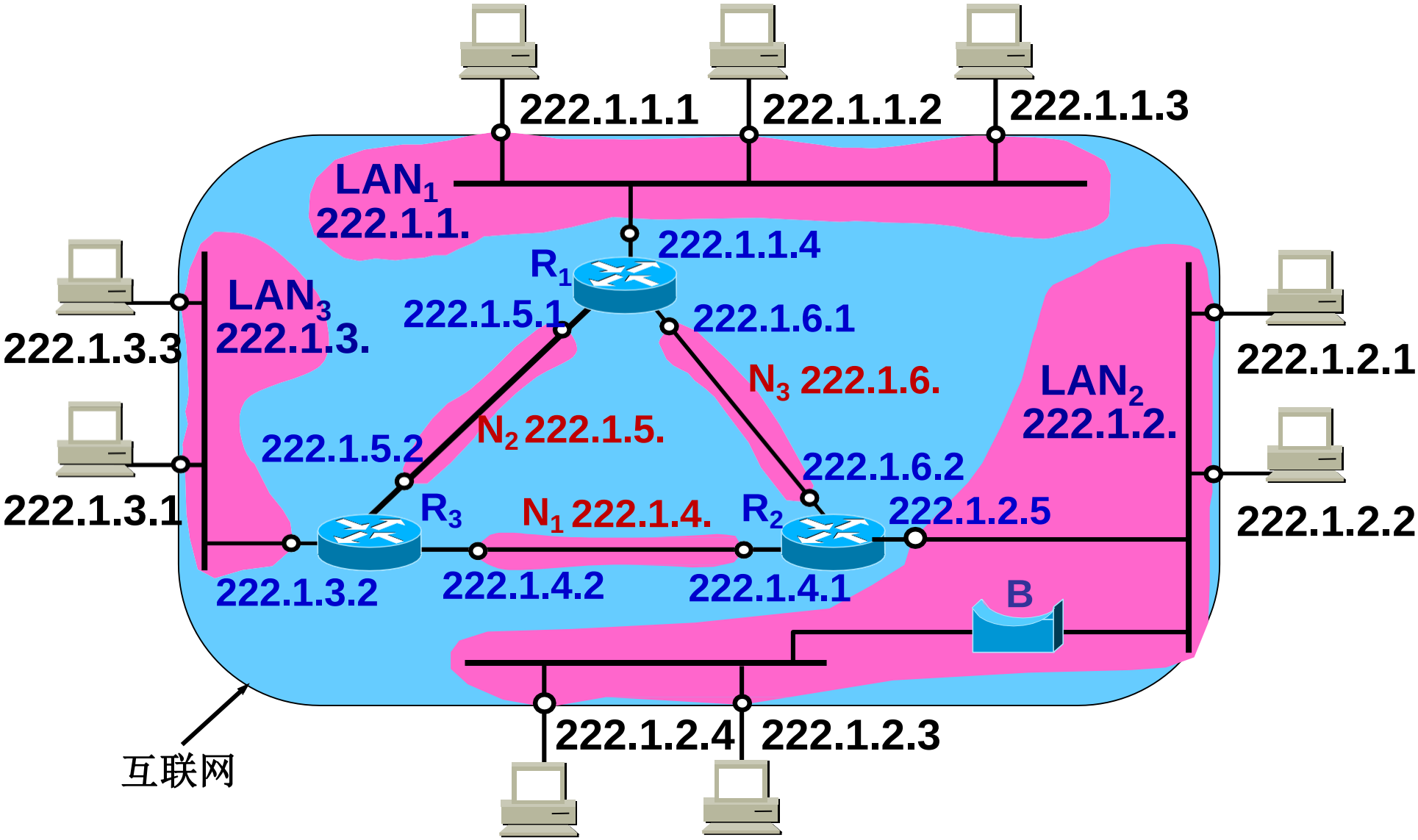
IP 地址的一些重要特点

- **(2) 实际上 IP 地址是标志一个主机（或路由器）和一条链路的接口。**
  - 当一个主机同时连接到两个网络上时，该主机就必须同时具有两个相应的 IP 地址，其网络号 net-id 必须是不同的。这种主机称为**多归属主机** (multihomed host)。
  - 由于一个路由器至少应当连接到两个网络（这样它才能将 IP 数据报从一个网络转发到另一个网络），因此**一个路由器至少应当有两个不同的 IP 地址。**

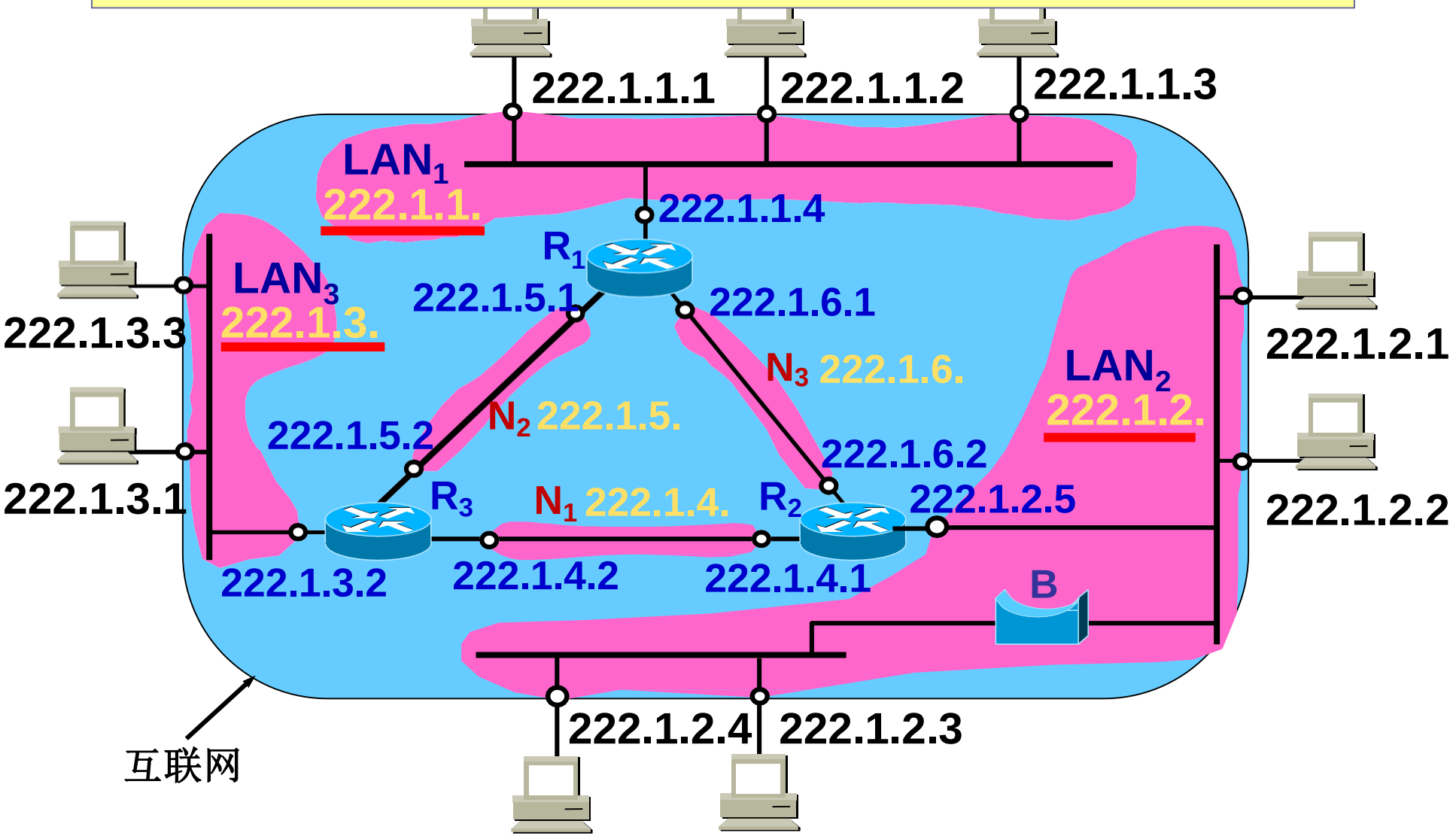
IP 地址的一些重要特点

- (3) 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 net-id。
- (4) 所有分配到网络号 net-id 的网络，无论是范围很小的局域网，还是可能覆盖很大地理范围的广域网，都是平等的。

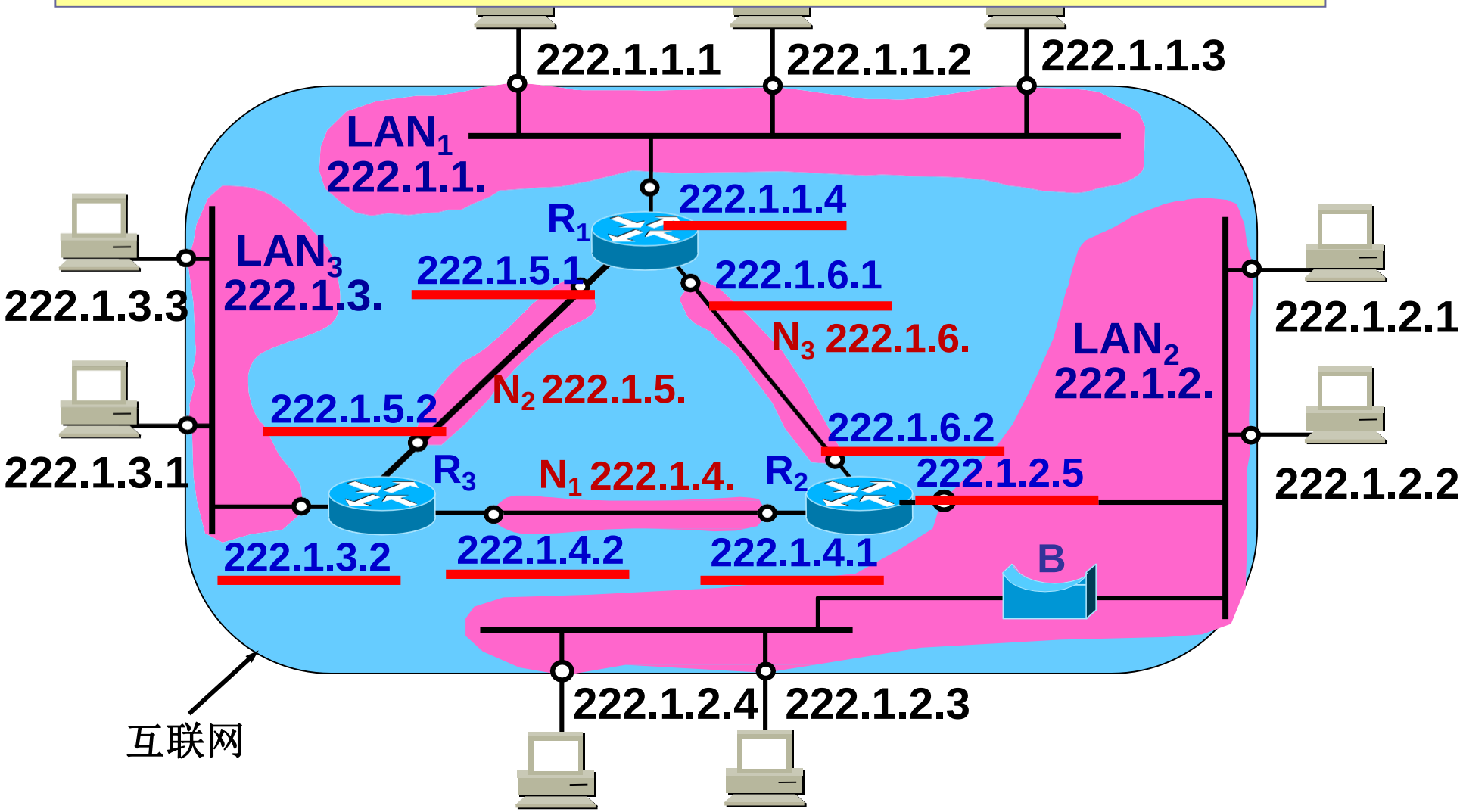
互联网中的 IP 地址


222.1.1.1
222.1.1.2
222.1.1.3
LAN1
222.1.1.
222.1.1.4
R1
LAN3
222.1.3.
222.1.5.1
222.1.6.1
222.1.3.3
222.1.2.1
N3 222.1.6.
LAN2
222.1.2.
N2 222.1.5.
222.1.5.2
222.1.6.2
222.1.3.1
R3
N1 222.1.4.
R2
222.1.2.5
222.1.2.2
222.1.3.2
222.1.4.2
222.1.4.1
B
222.1.2.4
222.1.2.3
互联网

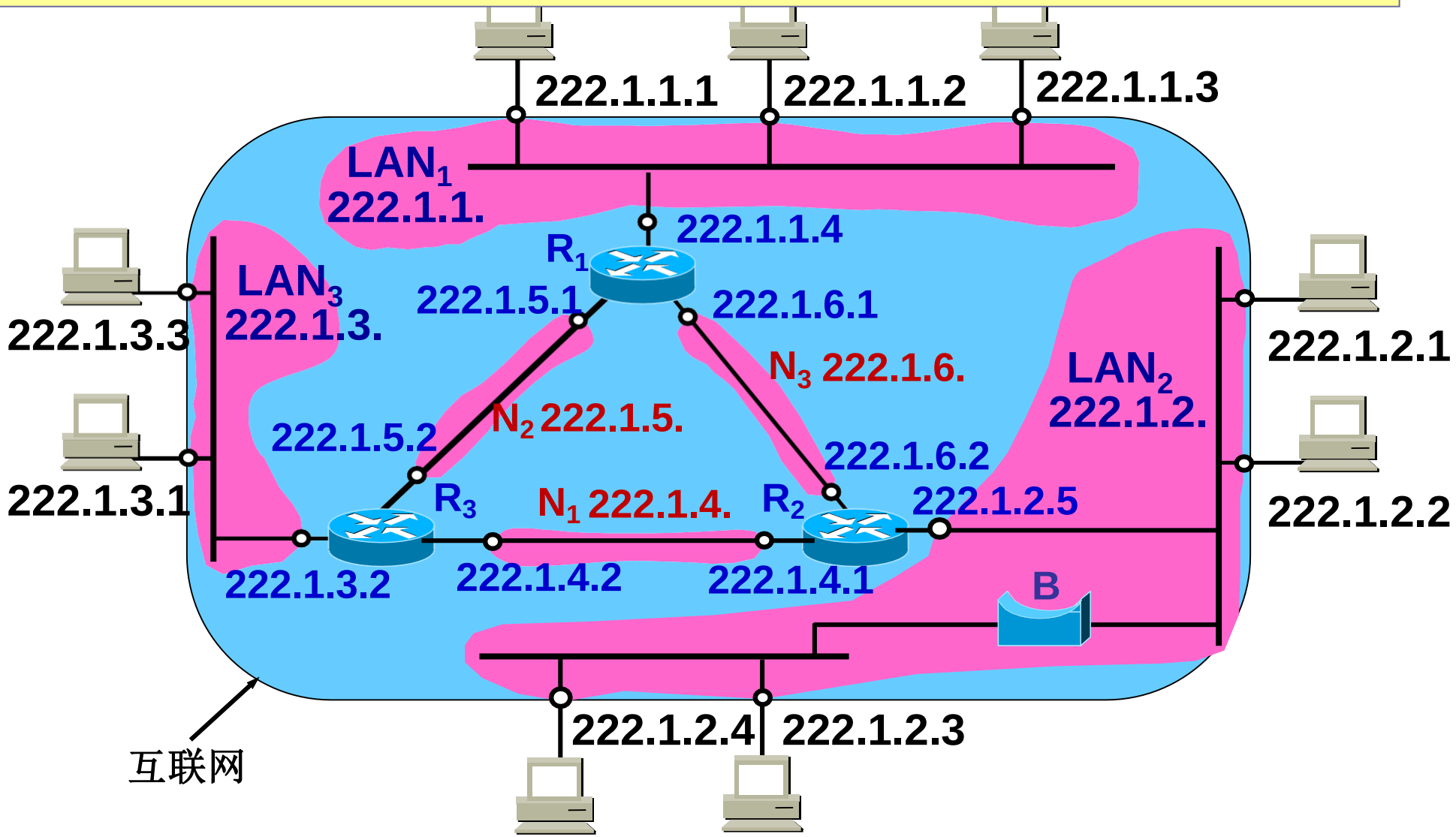
在同一个局域网上的主机或路由器的
IP 地址中的网络号必须是一样的。
图中的网络号就是 IP 地址中的 net-id



路由器总是具有两个或两个以上的 IP 地址。
路由器的每一个接口都有一个
不同网络号的 IP 地址。



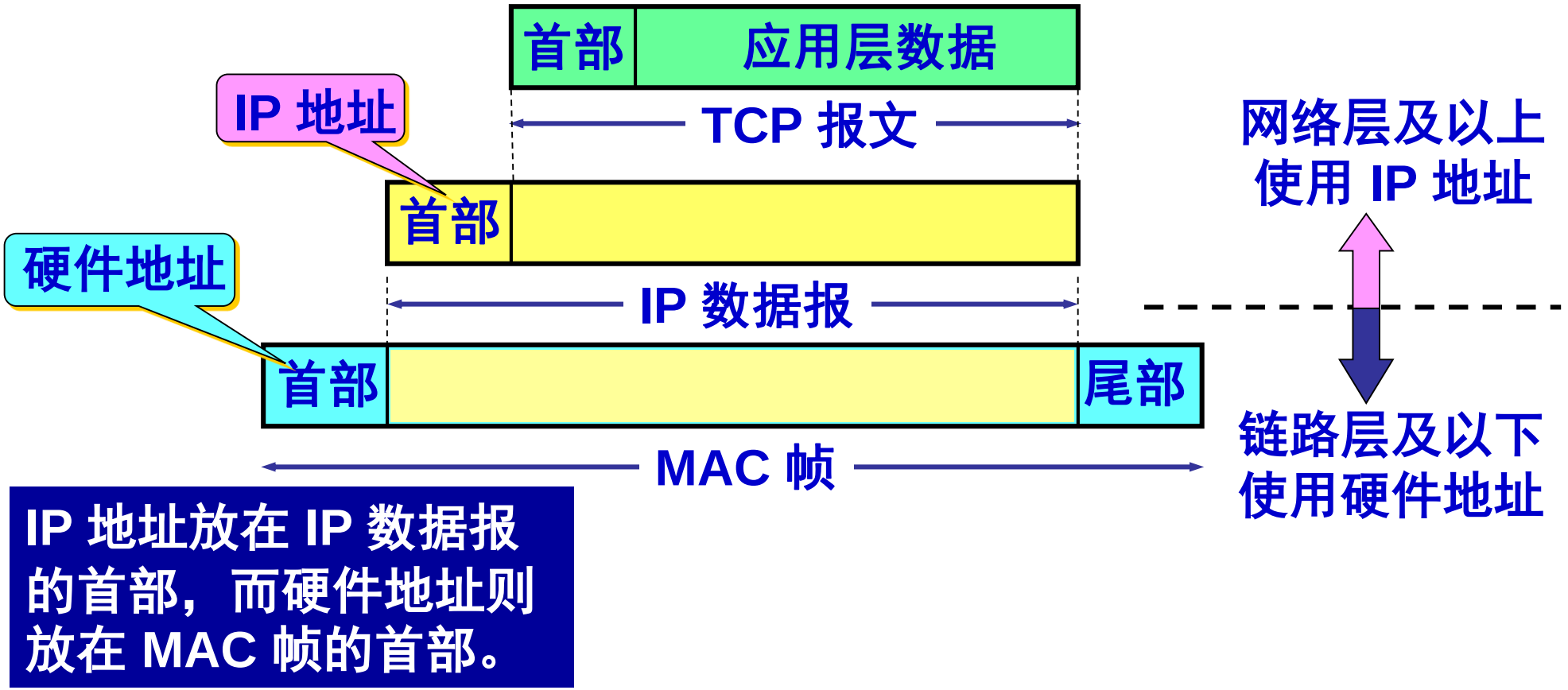
两个路由器直接相连的接口处，可指明也可不指明 IP 地址。如指明 IP 地址，则这一段连线就构成了一种只包含一段线路的特殊“网络”。现在常不指明 IP 地址。



4.2.3 IP 地址与硬件地址

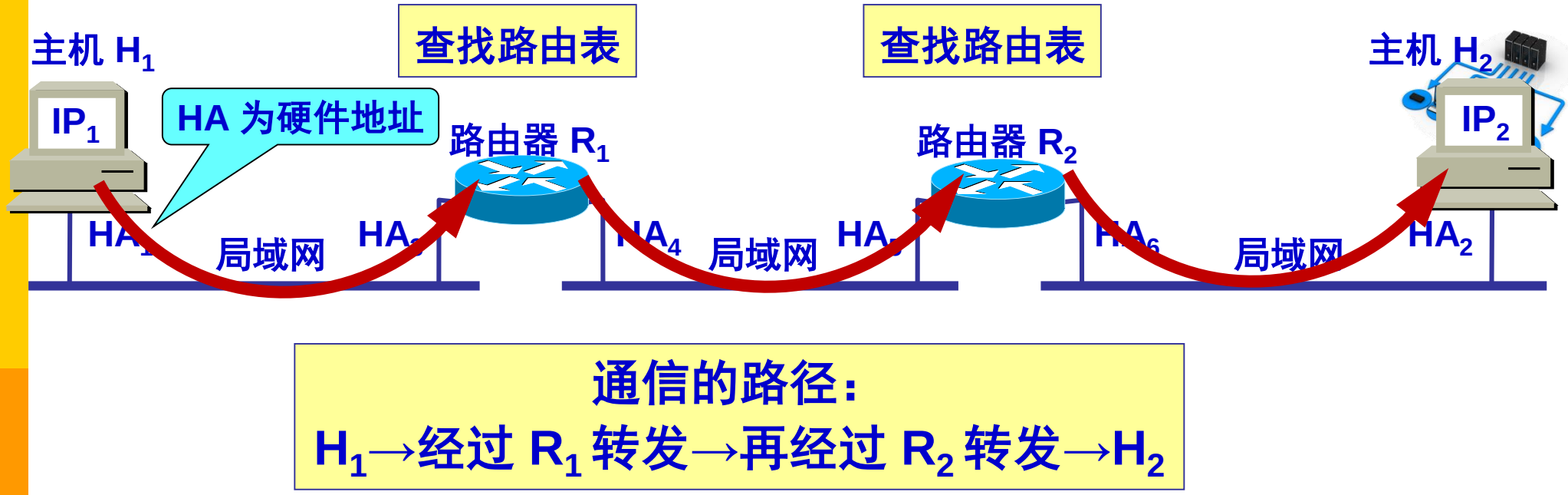
- IP 地址与硬件地址是不同的地址。
- 从层次的角度看，
  - 硬件地址（或物理地址）是数据链路层和物理层使用的地址。
  - IP 地址是网络层和以上各层使用的地址，是一种逻辑地址（称 IP 地址是逻辑地址是因为 IP 地址是用软件实现的）。

4.2.3 IP 地址与硬件地址



IP 地址放在 IP 数据报的首部，而硬件地址则放在 MAC 帧的首部。

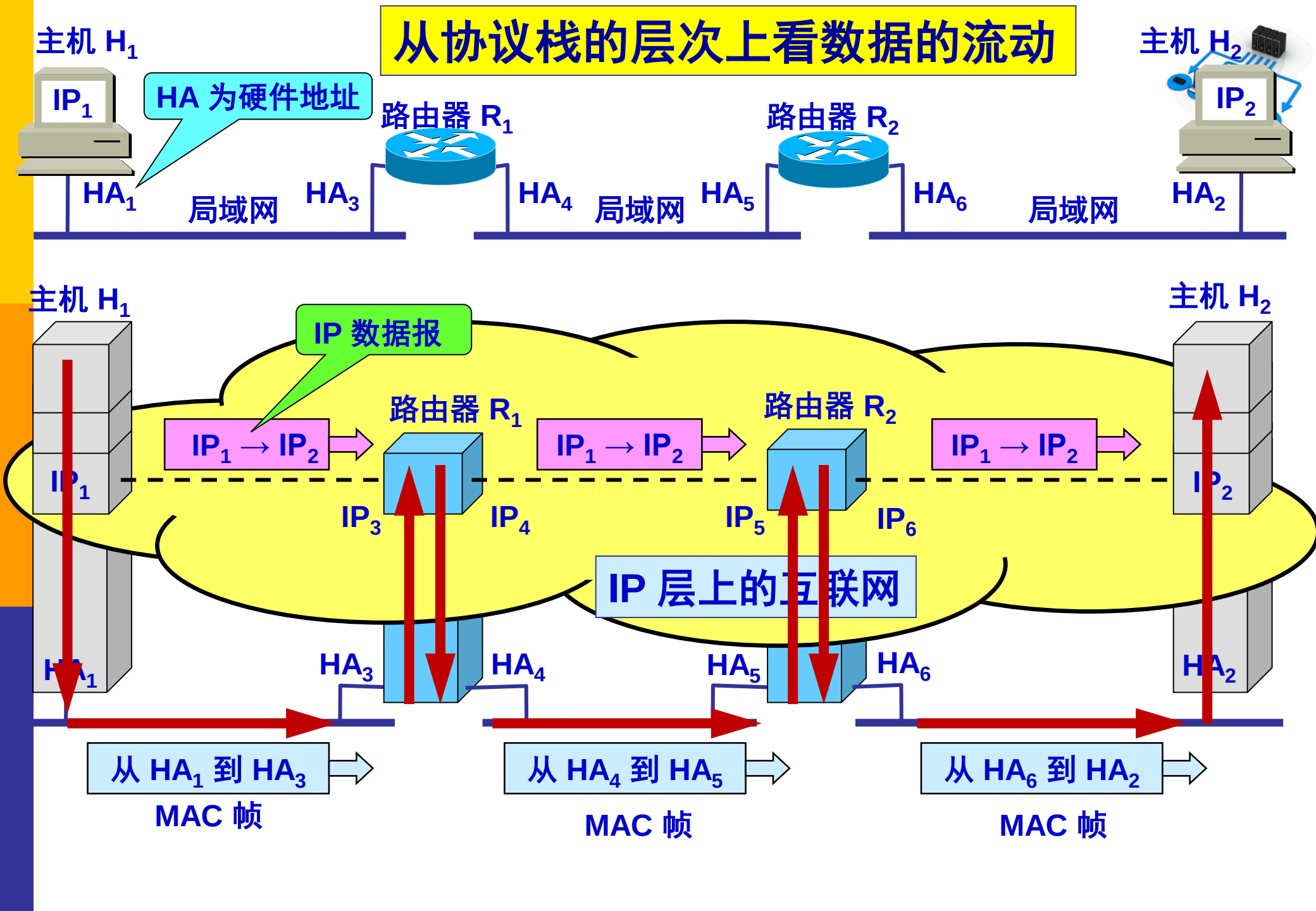
IP 地址与硬件地址的区别

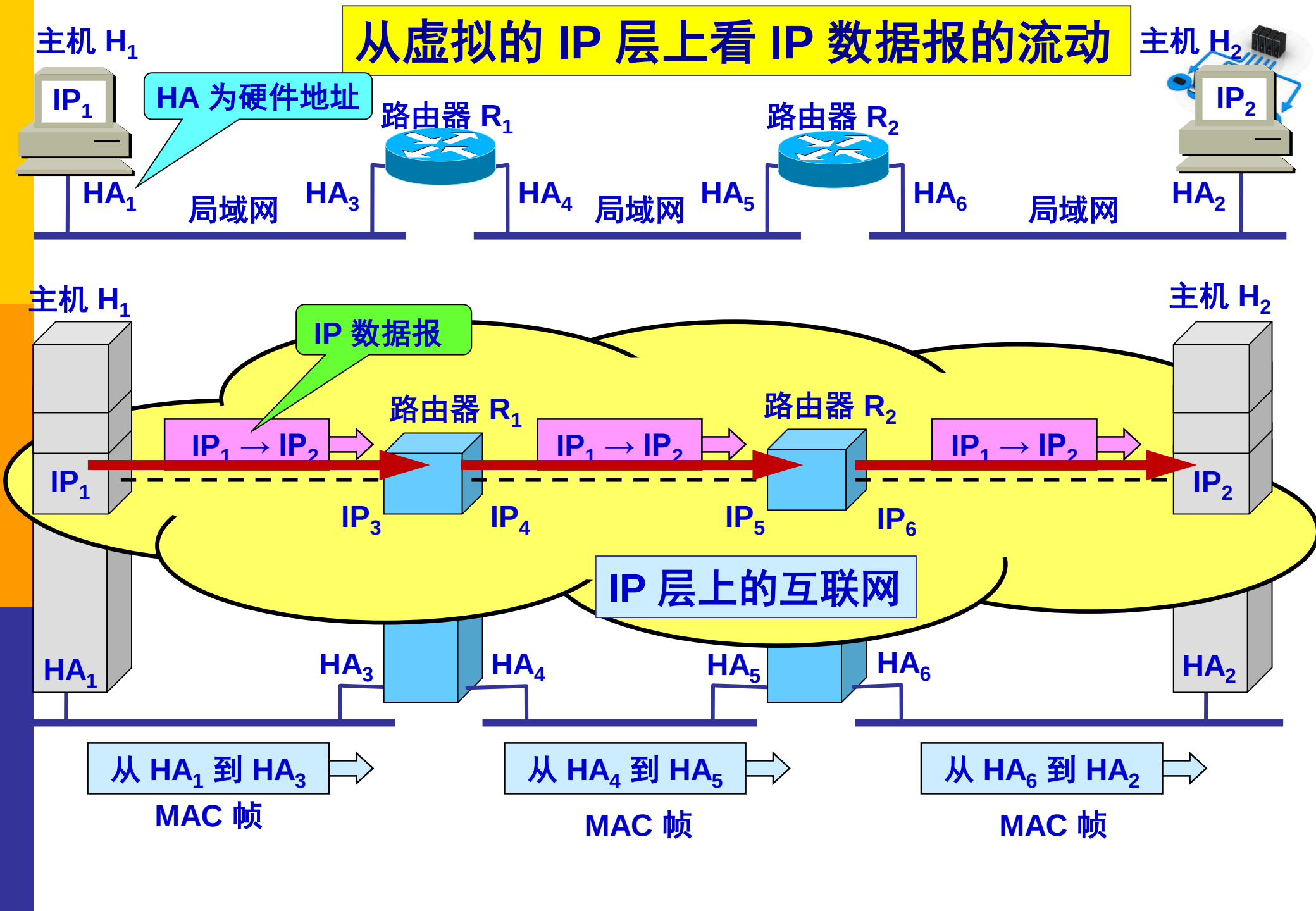


通信的路径：

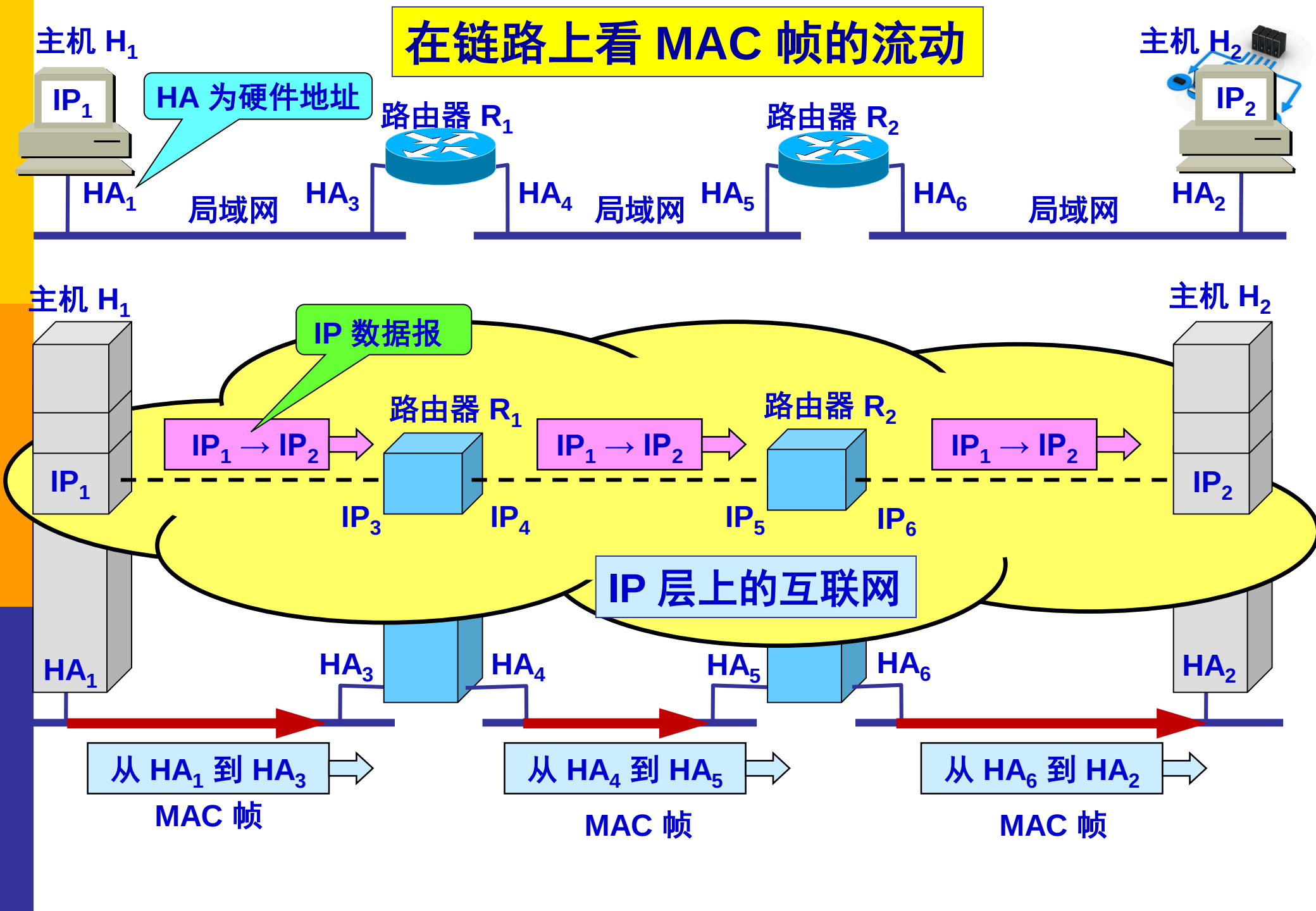
H_1→经过 R_1 转发→再经过 R_2 转发→H_2

从协议栈的层次上看数据的流动




从虚拟的 IP 层上看 IP 数据报的流动
主机 H1
IP1
HA 为硬件地址
HA1
局域网
路由器 R1
HA3
HA4
局域网
路由器 R2
HA5
HA6
局域网
主机 H2
IP2
HA2
主机 H1
IP 数据报
IP1 → IP2
IP1
IP3
路由器 R1
IP4
IP1 → IP2
IP5
路由器 R2
IP6
IP1 → IP2
IP2
主机 H2
IP 层上的互联网
HA1
HA3
HA4
HA5
HA6
HA2
从 HA1 到 HA3
MAC 帧
从 HA4 到 HA5
MAC 帧
从 HA6 到 HA2
MAC 帧

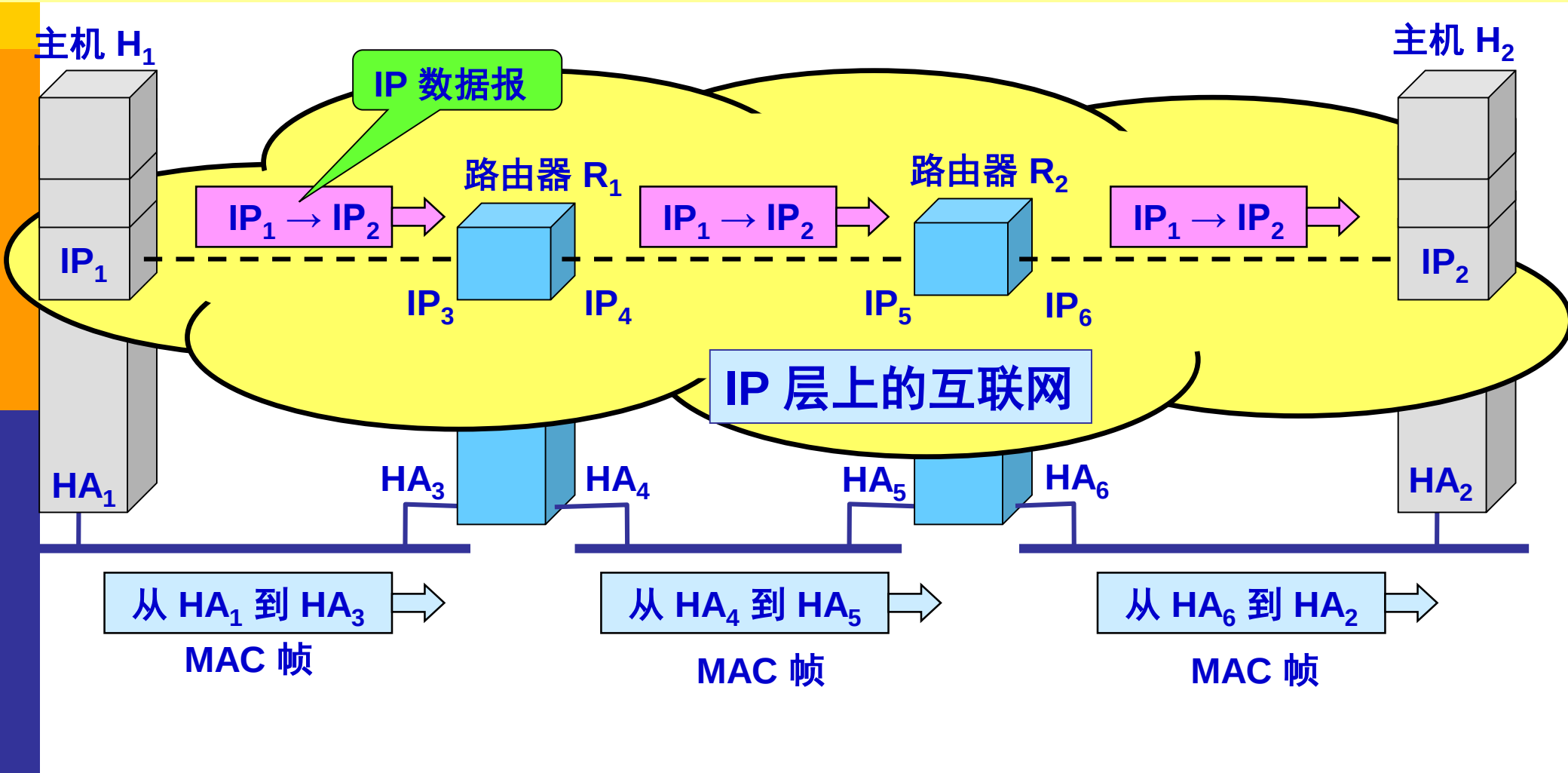
在链路上看 MAC 帧的流动



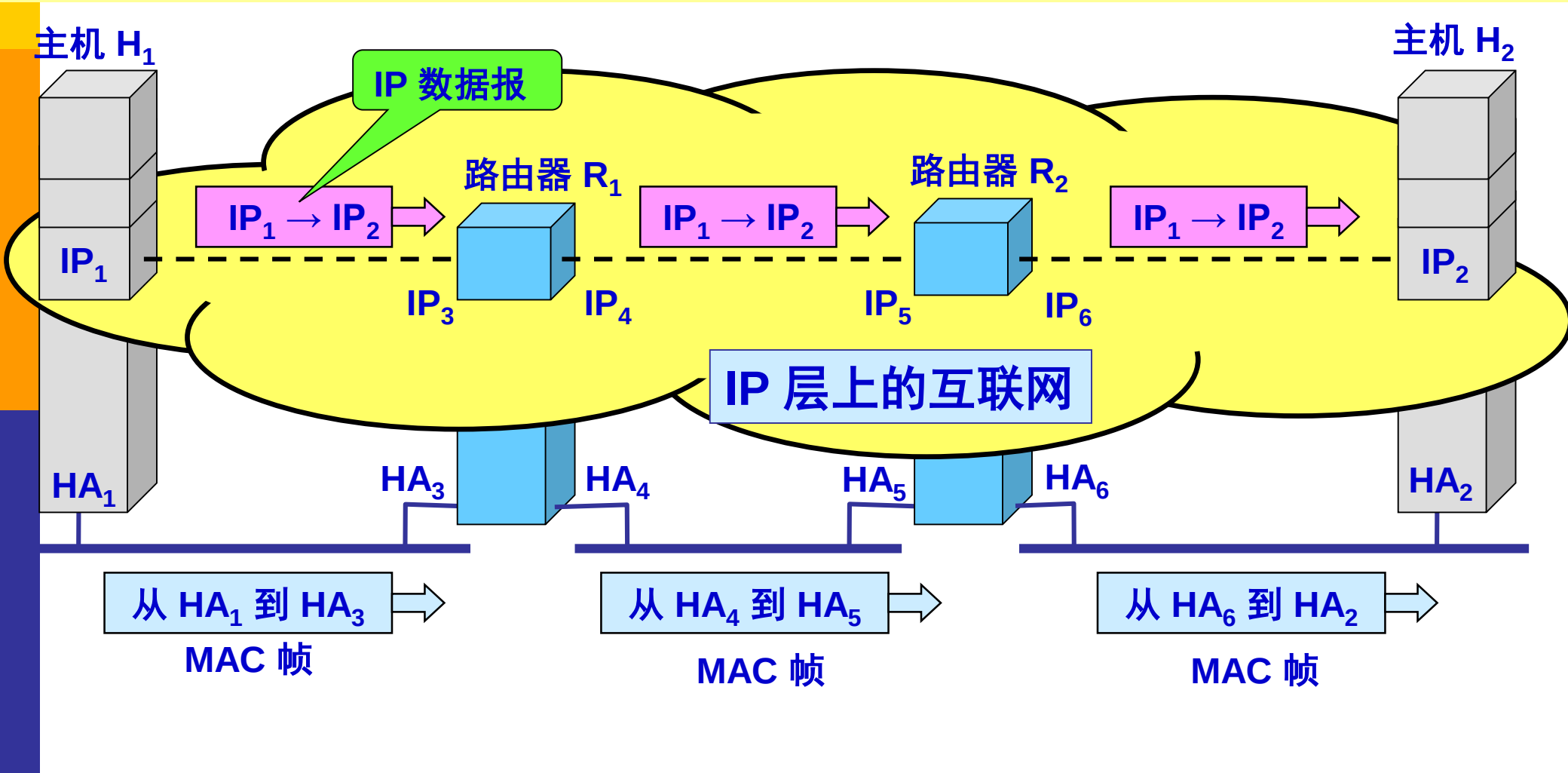
在 IP 层抽象的互联网上只能看到 IP 数据报。

图中的 $IP_1 \rightarrow IP_2$ 表示从源地址 IP_1 到目的地址 IP_2 。

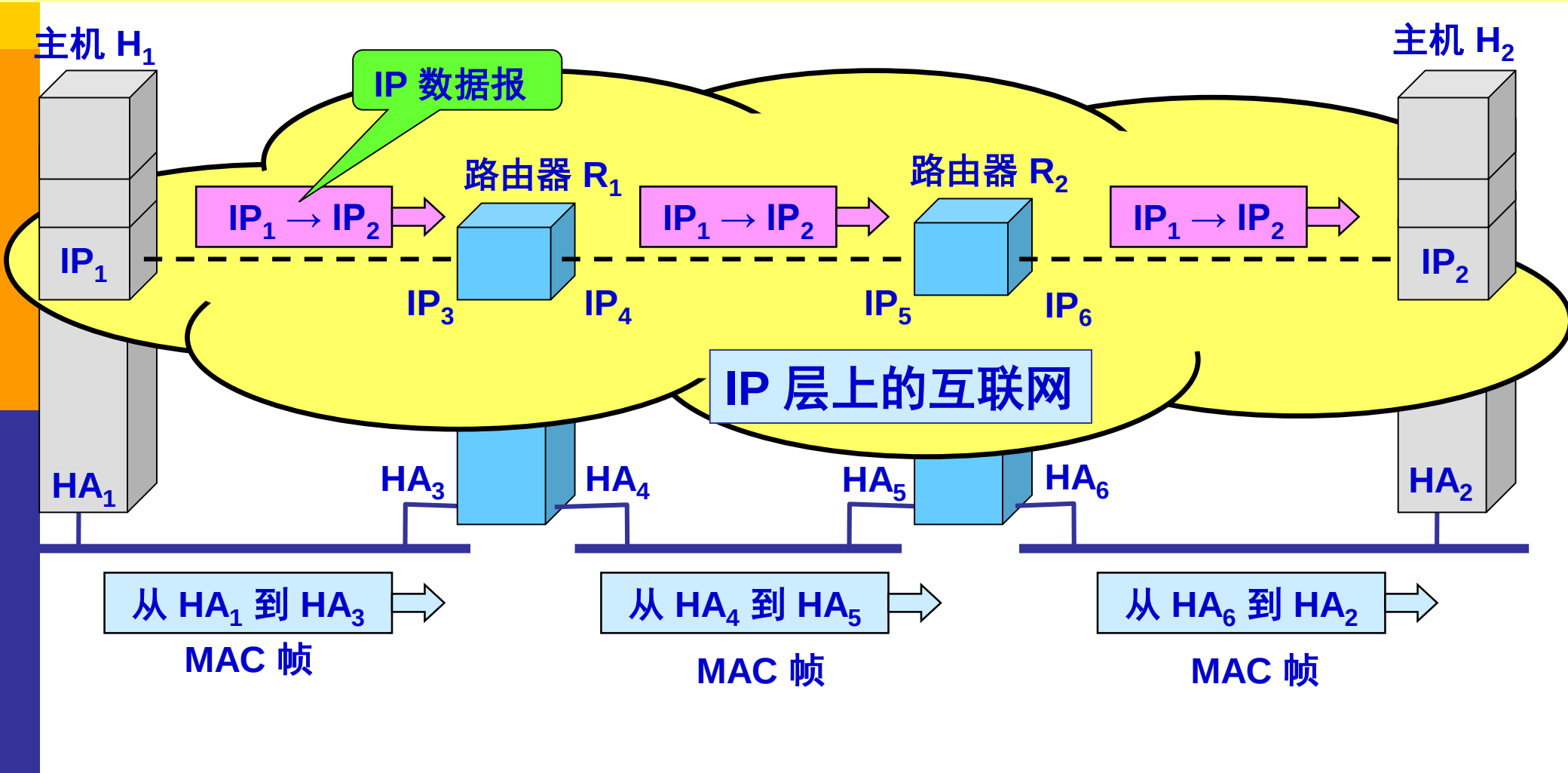
两个路由器的 IP 地址并不出现在 IP 数据报的首部中。



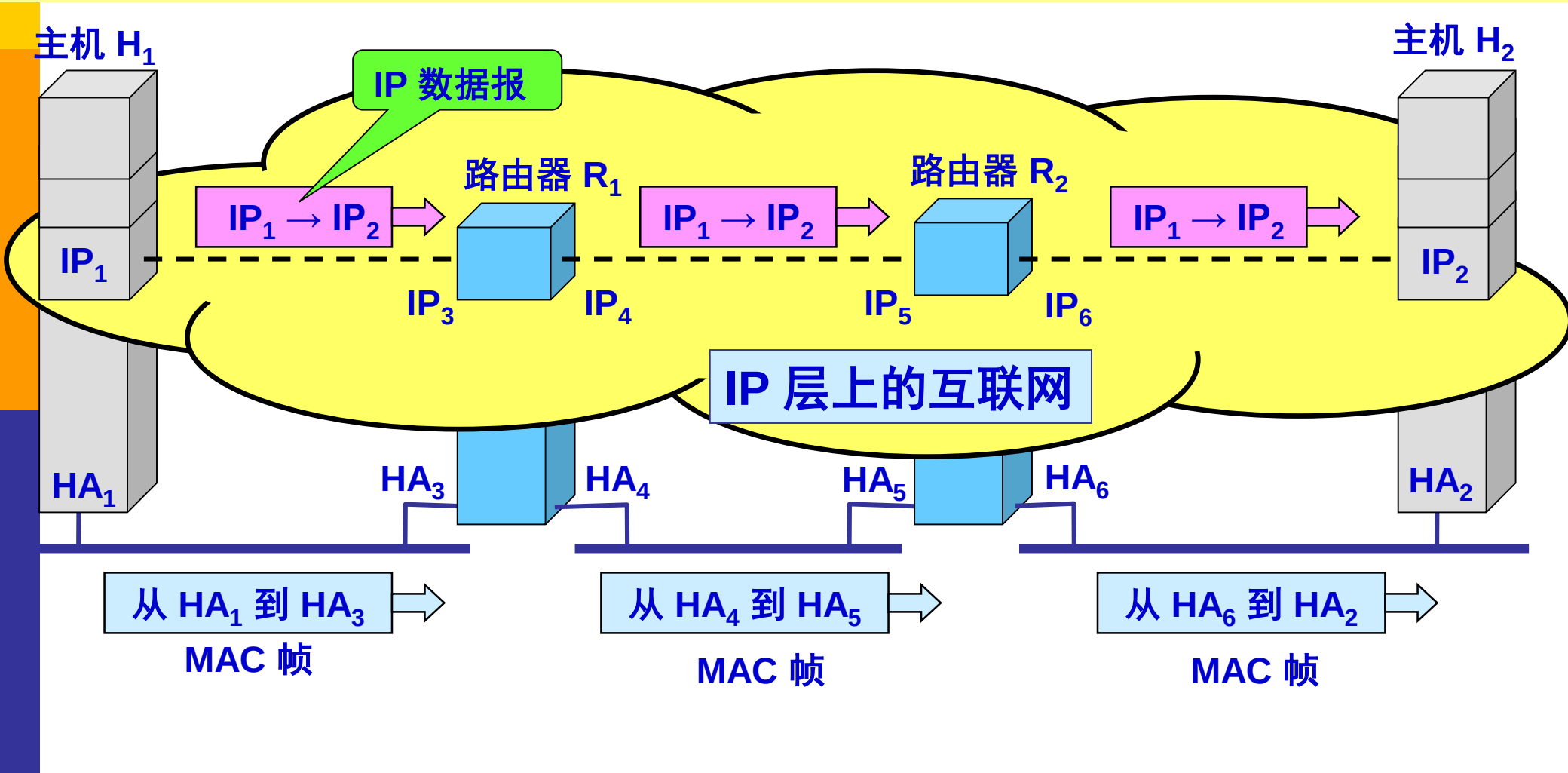
路由器只根据目的站的 IP 地址的网络号进行路由选择。



在具体的物理网络的链路层
只能看见 MAC 帧而看不见 IP 数据报



IP 层抽象的互联网屏蔽了下层很复杂的细节。
在抽象的网络层上讨论问题，就能够使用
统一的、抽象的 IP 地址
研究主机和主机 或 主机和路由器 之间的通信 。

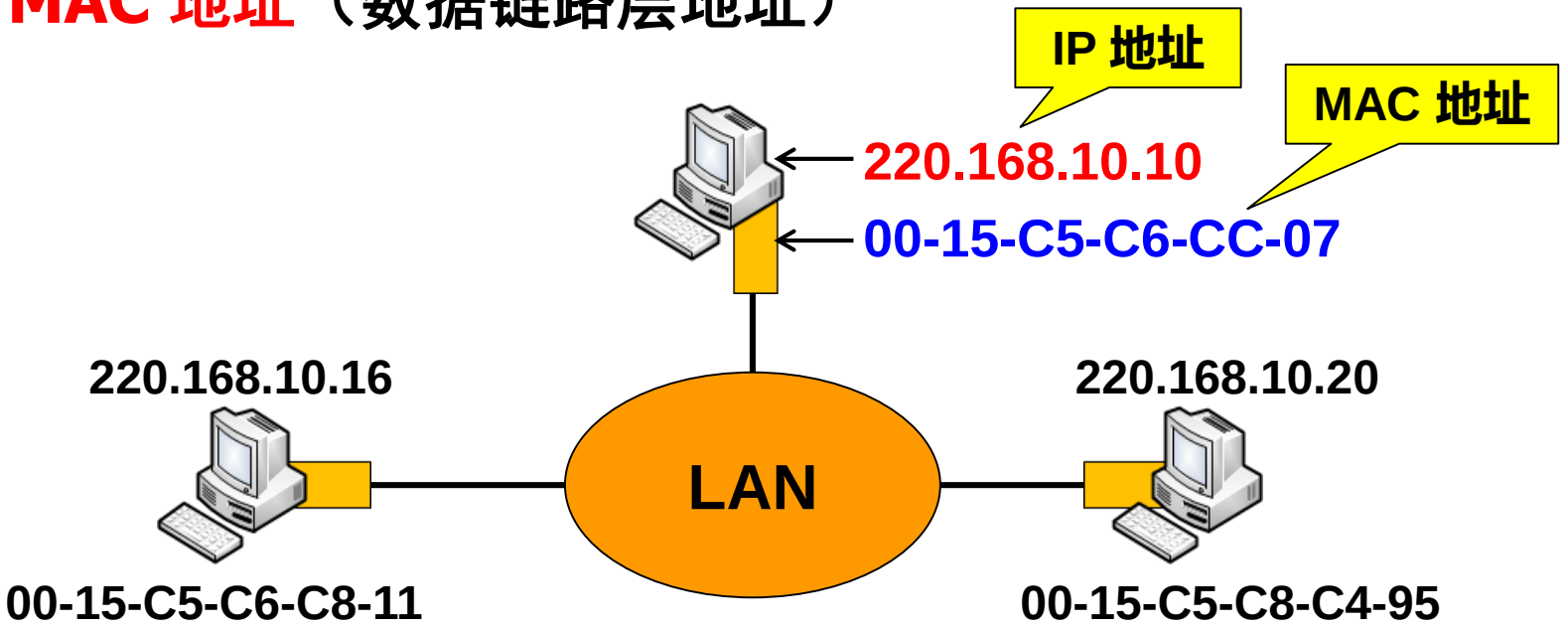


主机 H_1 与 H_2 通信中使用的 IP地址 与 硬件地址 HA

	在网络层 写入 IP 数据报首部的地址		在数据链路层 写入 MAC 帧首部的地址	
	源地址	目的地址	源地址	目的地址
从 H_1 到 R_1	IP_1	IP_2	HA_1	HA_3
从 R_1 到 R_2	IP_1	IP_2	HA_4	HA_5
从 R_2 到 H_2	IP_1	IP_2	HA_6	HA_2

4.2.4 地址解析协议 ARP (Address Resolution Protocol)

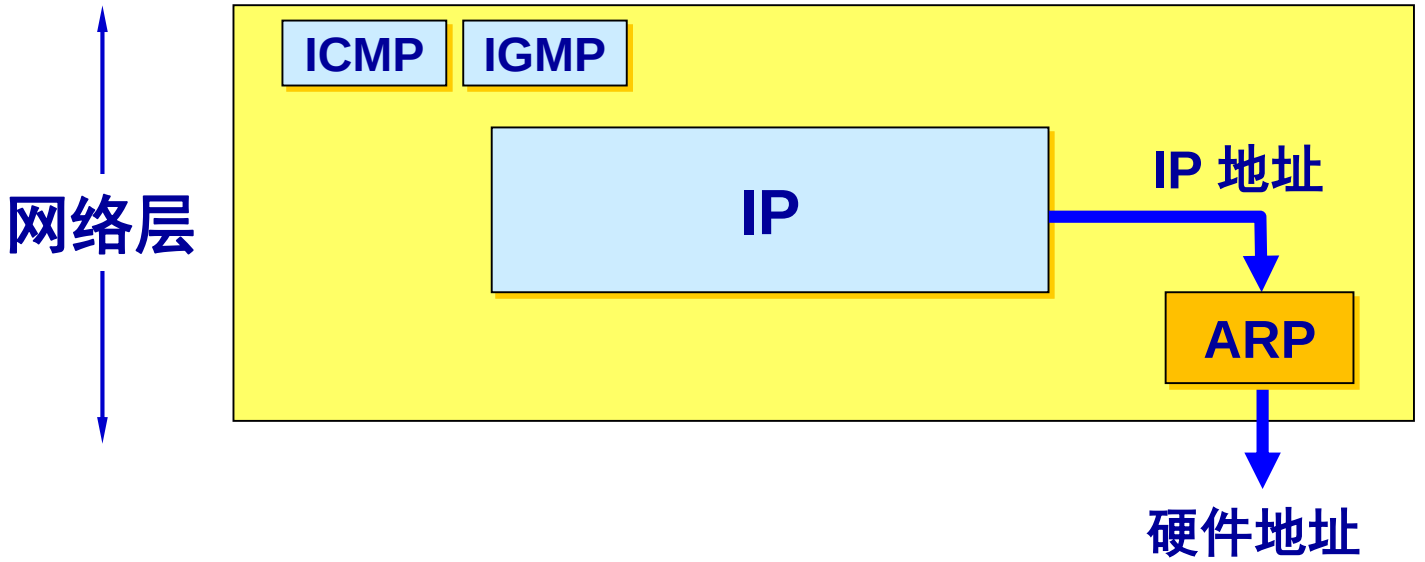
- 通信时使用了两个地址：
  - IP 地址（网络层地址）
  - MAC 地址（数据链路层地址）



每个接口都有两个地址

地址解析协议 ARP 的作用

- 已经知道了一个机器（主机或路由器）的IP地址，如何找出其相应的硬件地址？
- 地址解析协议 ARP 就是用来解决这样的问题的。



ARP 协议的作用

ARP 作用：
从网络层使用的 IP 地址，解析出在数据链路层使用的硬件地址。

地址解析协议 ARP 要点

- 不管网络层使用的是什么协议，在实际网络的链路上传送数据帧时，最终还是必须使用硬件地址。
- 每一个主机都设有一个 ARP 高速缓存 (ARP cache)，里面有所在的局域网上的各主机和路由器的 IP 地址到硬件地址的映射表，动态更新（新增或超时删除）。

< IP address；MAC address；TTL >

TTL (Time To Live)：地址映射有效时间 。

地址解析协议 ARP 要点

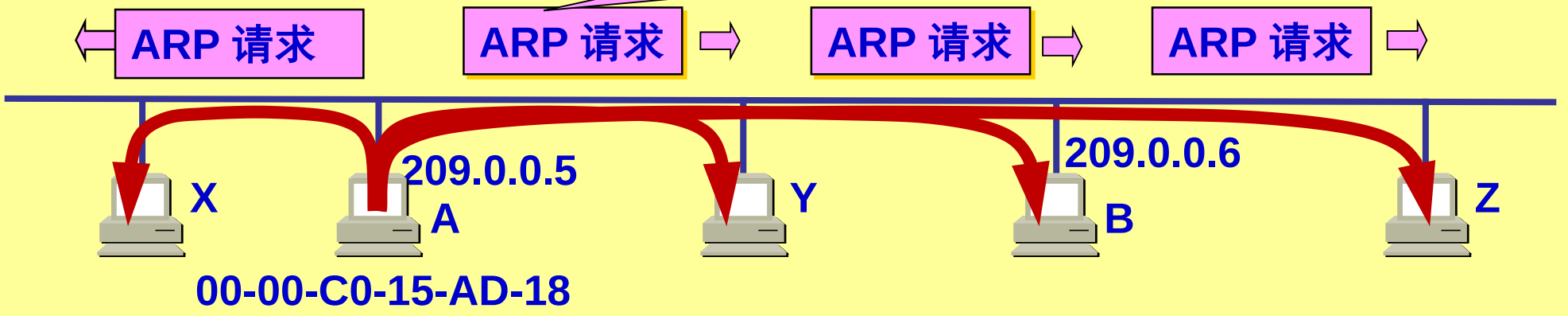
- 当主机 A 欲向本局域网上的某个主机 B 发送 IP 数据报时，就先在其 ARP 高速缓存中查看有无主机 B 的 IP 地址。
  - 如有，就可查出其对应的硬件地址，再将此硬件地址写入 MAC 帧，然后通过局域网将该 MAC 帧发往此硬件地址。
  - 如没有， ARP 进程在本局域网上广播发送一个 ARP 请求分组。当主机B收到请求分组后，向主机A发送 ARP 响应分组，写入自己的硬件地址。A收到响应分组后，将得到的 IP 地址到硬件地址的映射写入 ARP 高速缓存。

地址解析协议 ARP 要点

- ARP请求分组的内容是发送方硬件地址 / 发送方 IP 地址 / 目标方硬件地址(未知时填 0) / 目标方 IP 地址。
- 本地广播 ARP 请求（路由器不转发ARP请求），本局域网上所有主机都收到此ARP请求分组。
- ARP 响应分组的内容是发送方硬件地址 / 发送方 IP地址 / 目标方硬件地址 / 目标方 IP 地址。
- ARP 分组封装在帧中传输。
- ARP请求分组是 广播，ARP响应分组是单播

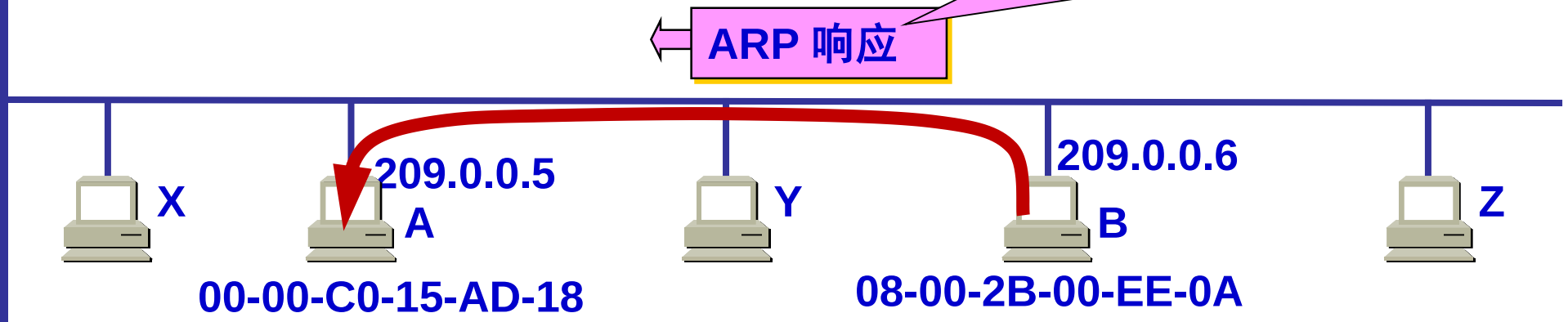
主机 A 广播发送
ARP 请求分组

我是 209.0.0.5，硬件地址是 00-00-C0-15-AD-18
我想知道主机 209.0.0.6 的硬件地址



主机 B 向 A 发送
ARP 响应分组

我是 209.0.0.6
硬件地址是 08-00-2B-00-EE-0A



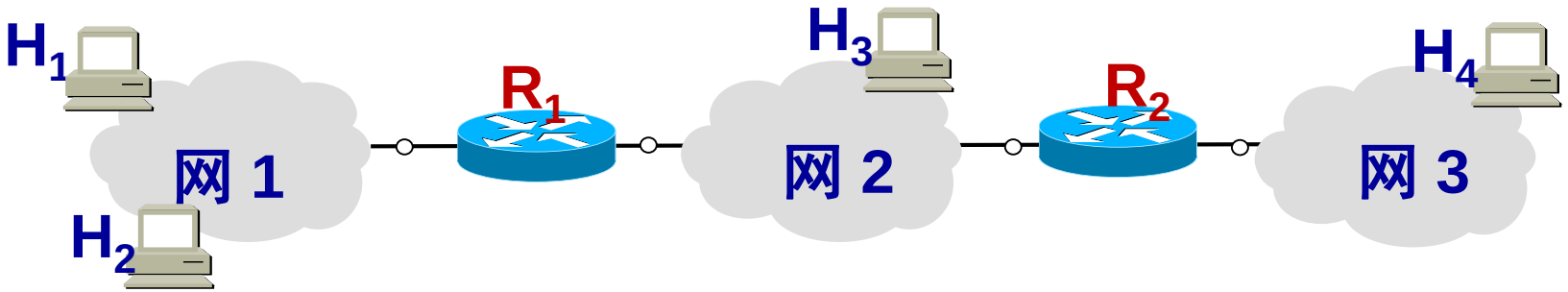
ARP 高速缓存的作用

- 存放最近获得的 IP 地址到 MAC 地址的绑定，以减少 ARP 广播的数量。
- 对每一个映射地址项目都设置生存时间。
- 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 地址到硬件地址的映射写入 ARP 请求分组。
- 当主机 B 收到 A 的 ARP 请求分组时，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 A 发送数据报时就更方便了。

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。
- 如果所要找的主机和源主机不在同一个局域网上，那么就要通过 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



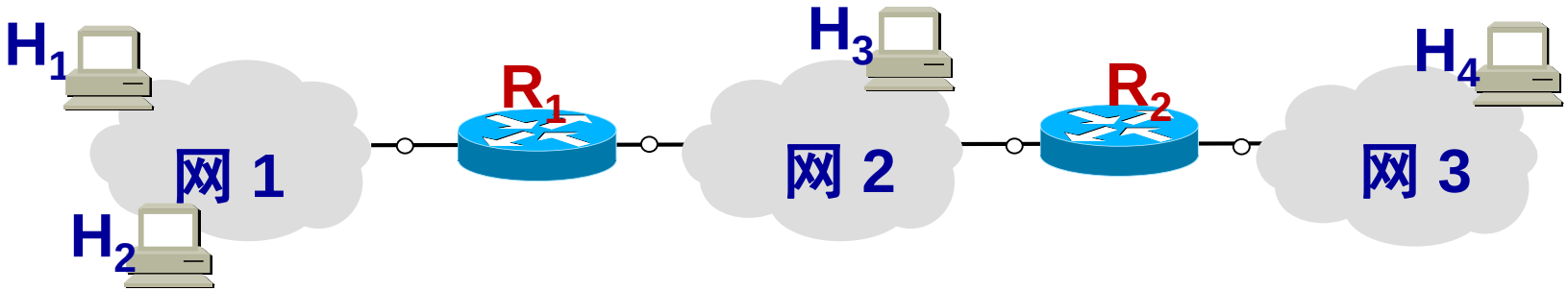
使用 ARP 的四种典型情况

① 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。

② 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。

③ 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。

④ 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 找到本网络上另一个路由器的硬件地址。剩下的工作由这个路由器来完成。



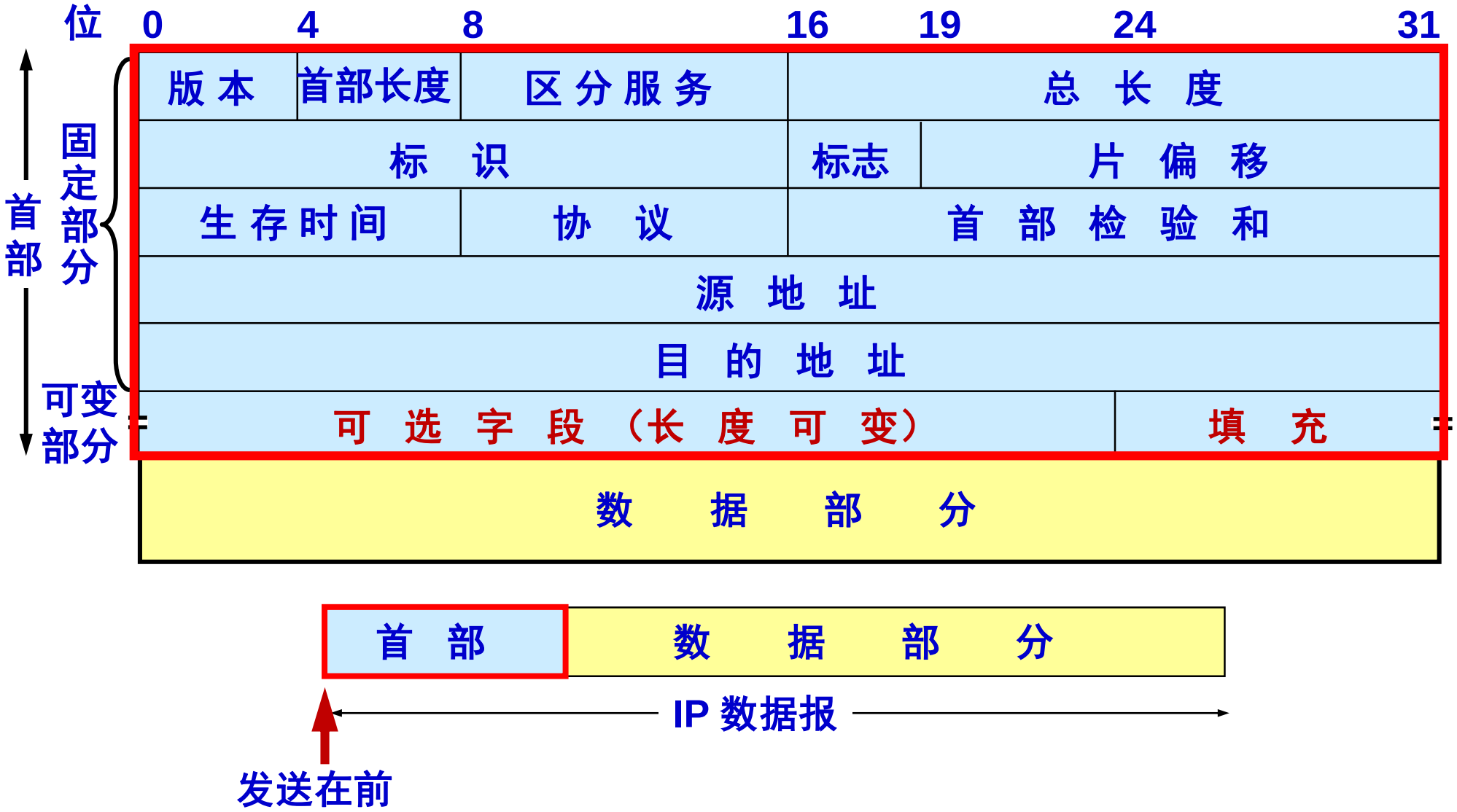
什么？不直接使用硬件地址进行通信？

- 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。
- IP 编址把这个复杂问题解决了。连接到互联网的主机只需各自拥有一个唯一的 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，因为上述的调用 ARP 的复杂过程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。
- 因此，在虚拟的 IP 网络上用 IP 地址进行通信给广大的计算机用户带来了很大的方便。

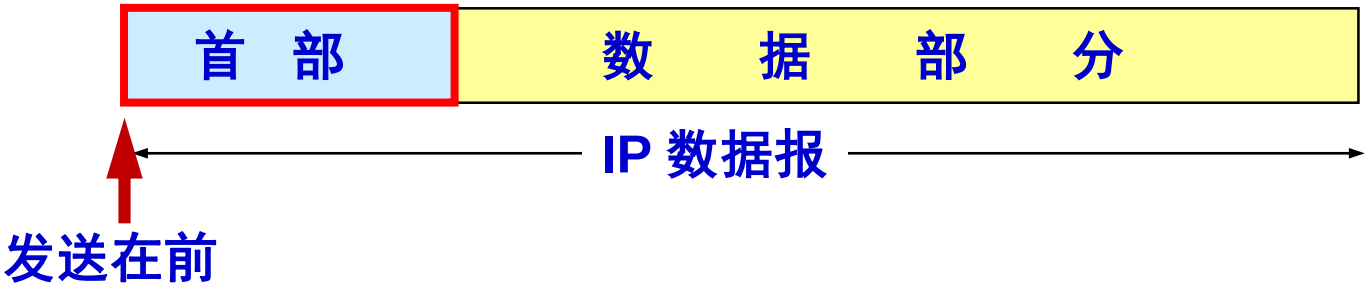
4.2.5 IP 数据报的格式

- 一个 IP 数据报由首部和数据两部分组成。
- 首部的前一部分是固定长度，共 20 字节，是所有 IP 数据报必须具有的。
- 在首部的固定部分的后面是一些可选字段，其长度是可变的。

IP 数据报由首部和数据两部分组成

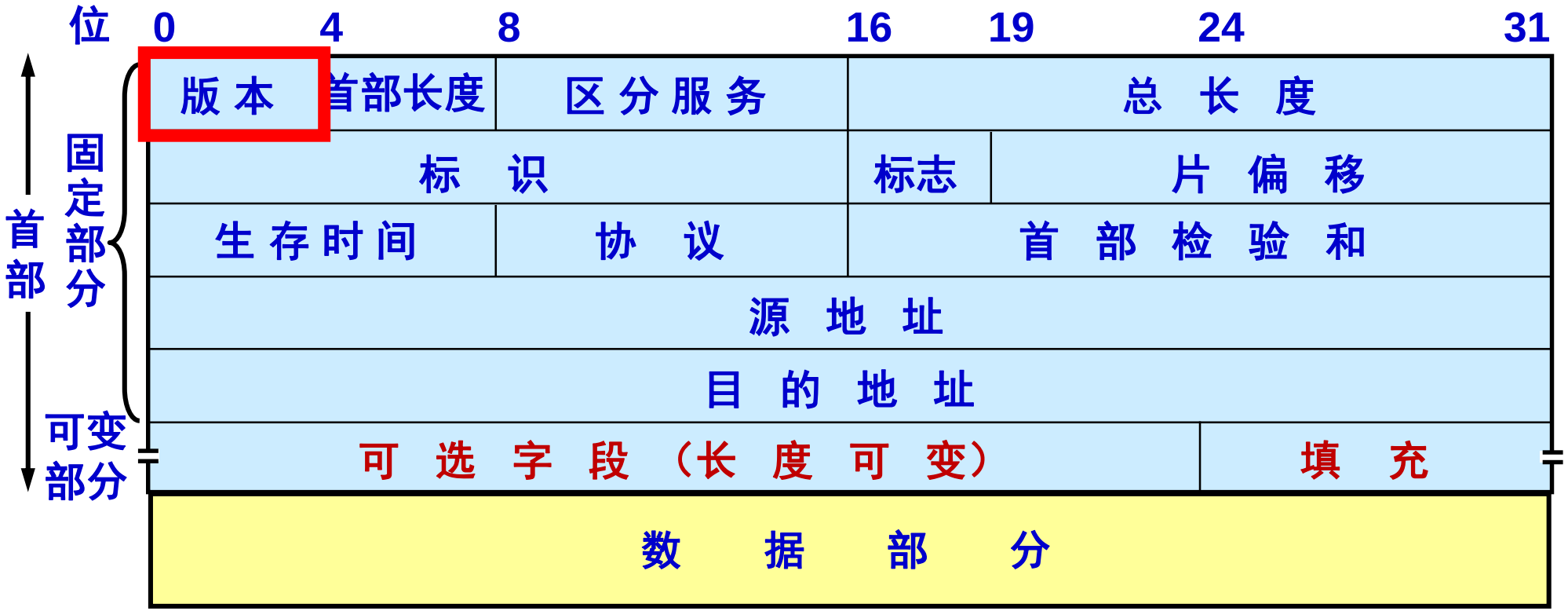


首部的前一部分是固定长度，共 20 字节，是所有 IP 数据报必须具有的。



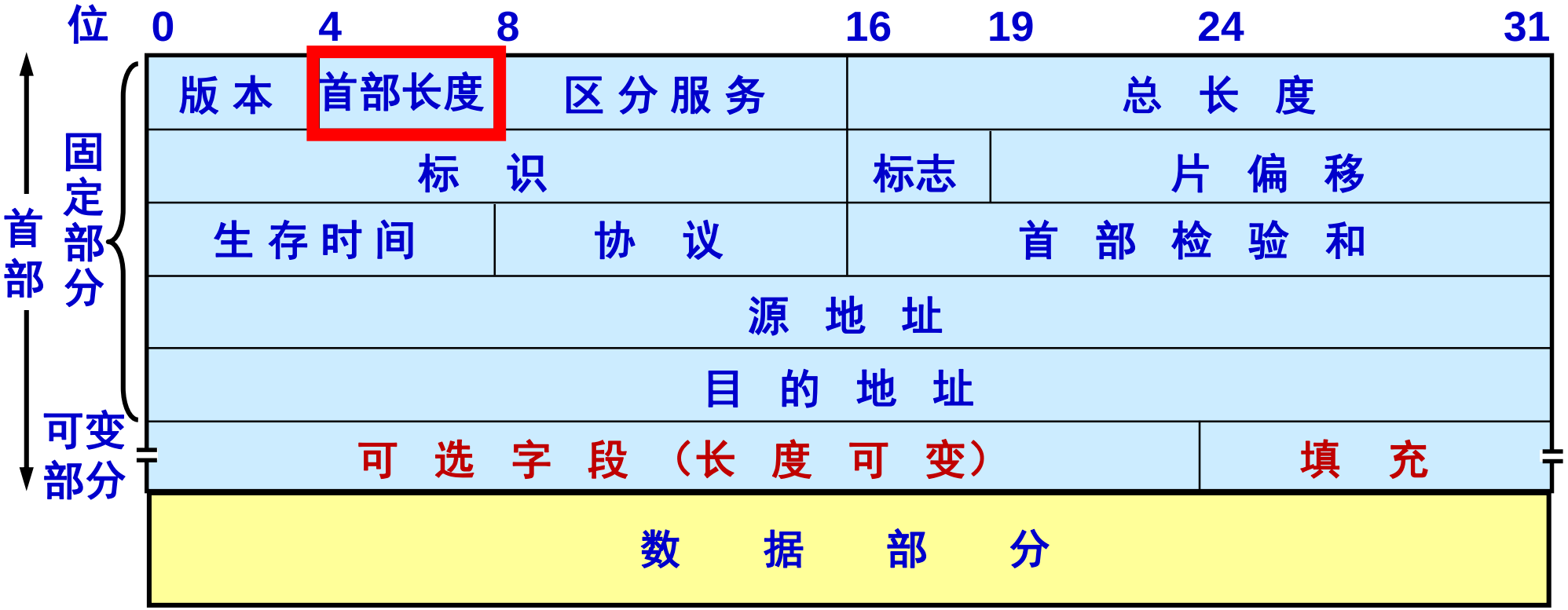
可选字段，其长度是可变的

1. IP 数据报首部的固定部分中的各字段



版本——占 4 位，指 IP 协议的版本。
目前的 IP 协议版本号为 4 (即 IPv4)。

1. IP 数据报首部的固定部分中的各字段

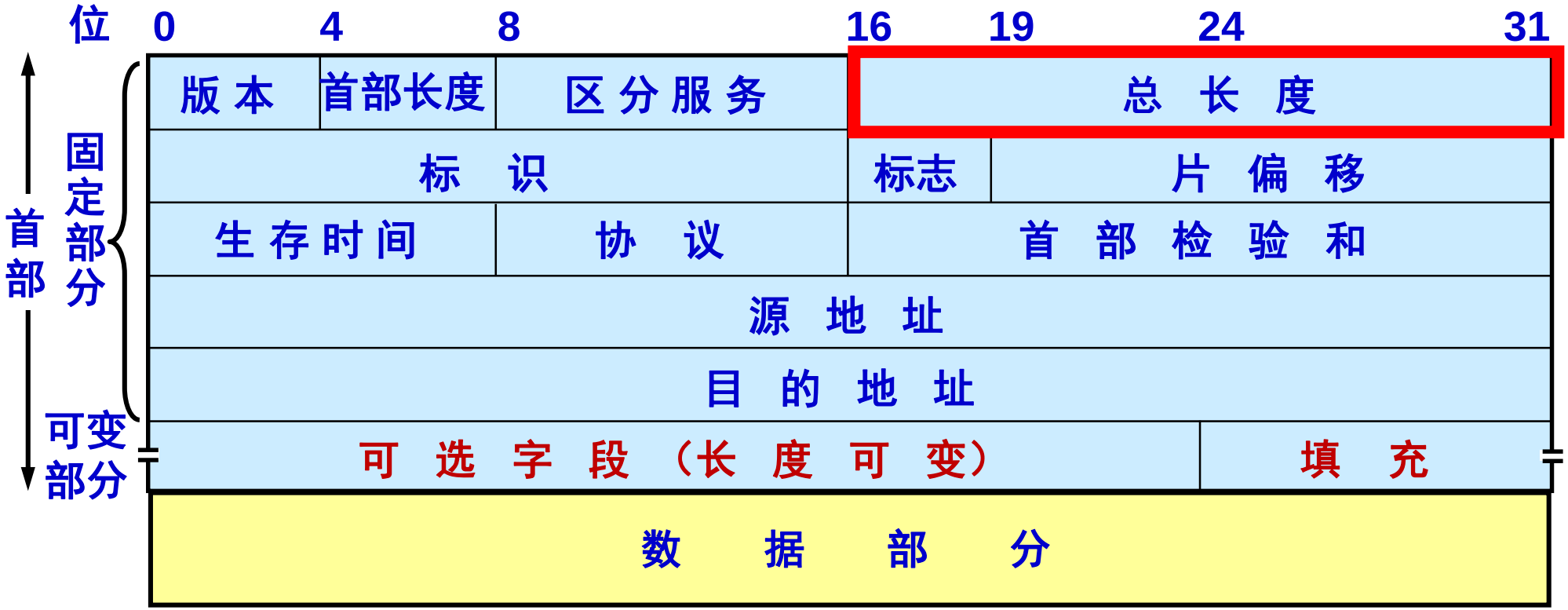


首部长度——占 4 位，可表示的最大数值是 15 个单位(一个单位为 4 字节)，
因此 IP 的首部长度的最大值是 60 字节。

1. IP 数据报首部的固定部分中的各字段

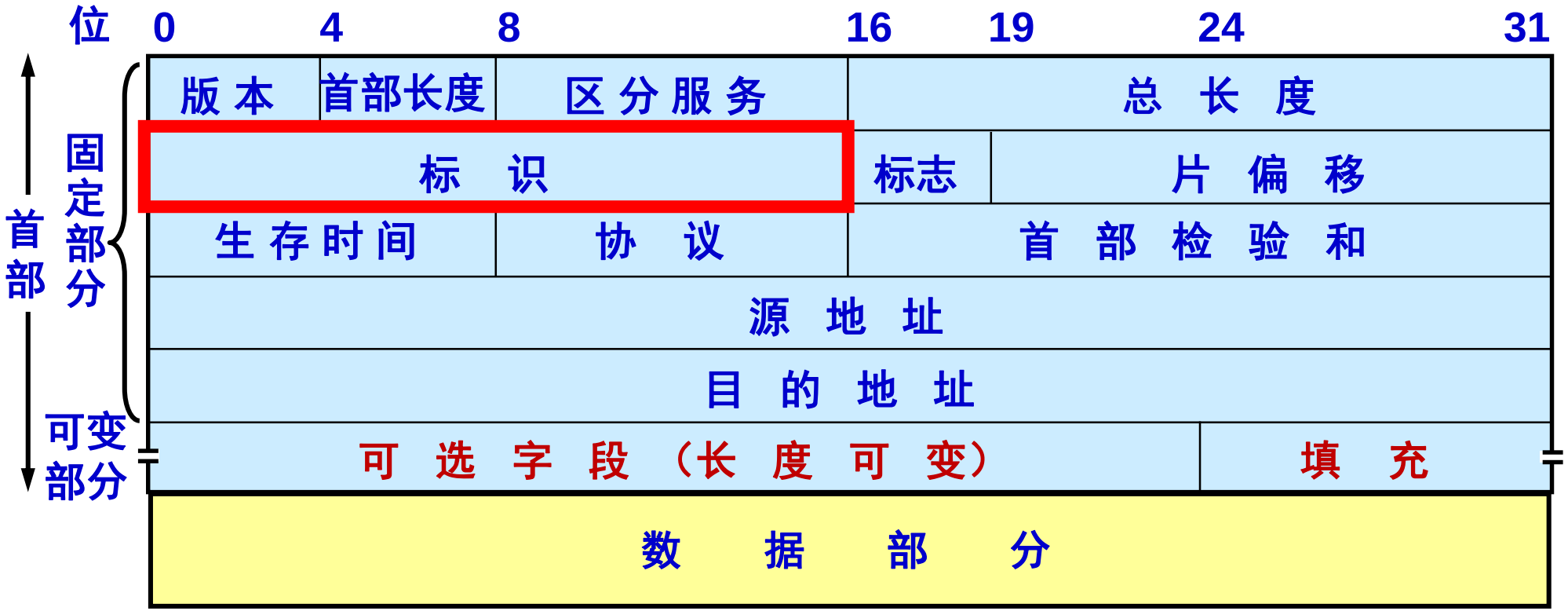
区分服务——占 8 位，用来获得更好的服务。
在旧标准中叫做服务类型，但实际上一直未被使用过。
1998 年这个字段改名为区分服务。
只有在使用区分服务（DiffServ）时，这个字段才起作用。
在一般的情况下都不使用这个字段

1. IP 数据报首部的固定部分中的各字段



总长度——占 16 位，指首部和数据之和的长度，
单位为字节，因此数据报的最大长度为 65535 字节。
总长度必须不超过最大传送单元 MTU。

1. IP 数据报首部的固定部分中的各字段



标识(identification) ——占 16 位，
它是一个计数器，用来产生 IP 数据报的标识。

1. IP 数据报首部的固定部分中的各字段

标志(flag) ——占 3 位，目前只有前两位有意义。

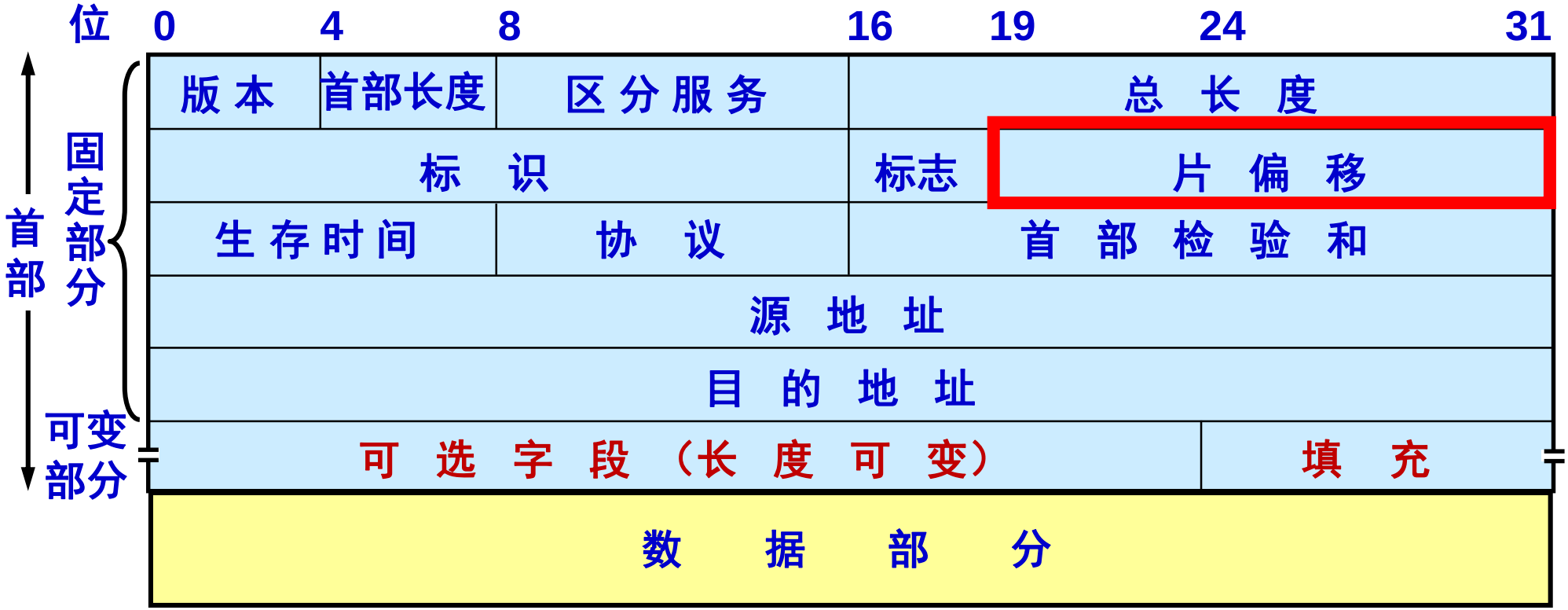
标志字段的最低位是 MF (More Fragment)。

MF = 1 表示后面“还有分片”。MF = 0 表示最后一个分片。

标志字段中间的一位是 DF (Don't Fragment)。

只有当 DF = 0 时才允许分片。

1. IP 数据报首部的固定部分中的各字段



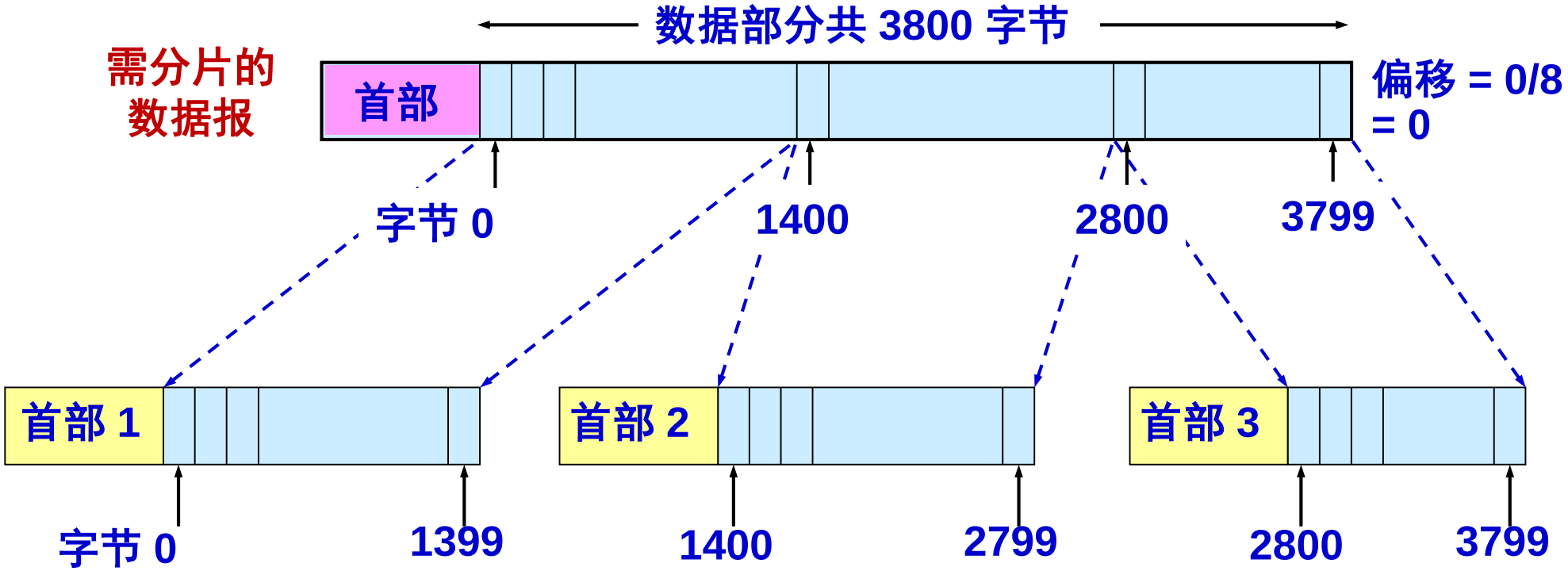
片偏移——占13 位，指出：较长的分组在分片后某片在原分组中的相对位置。

片偏移以 8 个字节为偏移单位。

【例4-1】 IP 数据报分片

- 一数据报的总长度为 3820 字节，其数据部分的长度为 3800 字节（使用固定首部），需要分片为长度不超过 1420 字节的数据报片。
- 因固定首部长度为 20 字节，因此每个数据报片的数据部分长度不能超过 1400 字节。
- 于是分为 3 个数据报片，其数据部分的长度分别为 1400、1400 和 1000 字节。
- 原始数据报首部被复制为各数据报片的首部，但必须修改有关字段的值。

【例4-1】 IP 数据报分片



数据报片 1
偏移 = 0/8 = 0

数据报片 2
偏移 = 1400/8 = 175

数据报片 3
偏移 = 2800/8 = 350

【例4-1】 IP 数据报分片

IP 数据报首部中与分片有关的字段中的数值

	总长度	标识	MF	DF	片偏移
原始数据报	3820	12345	0	0	0
数据报片1	1420	12345	1	0	0
数据报片2	1420	12345	1	0	175
数据报片3	1020	12345	0	0	350

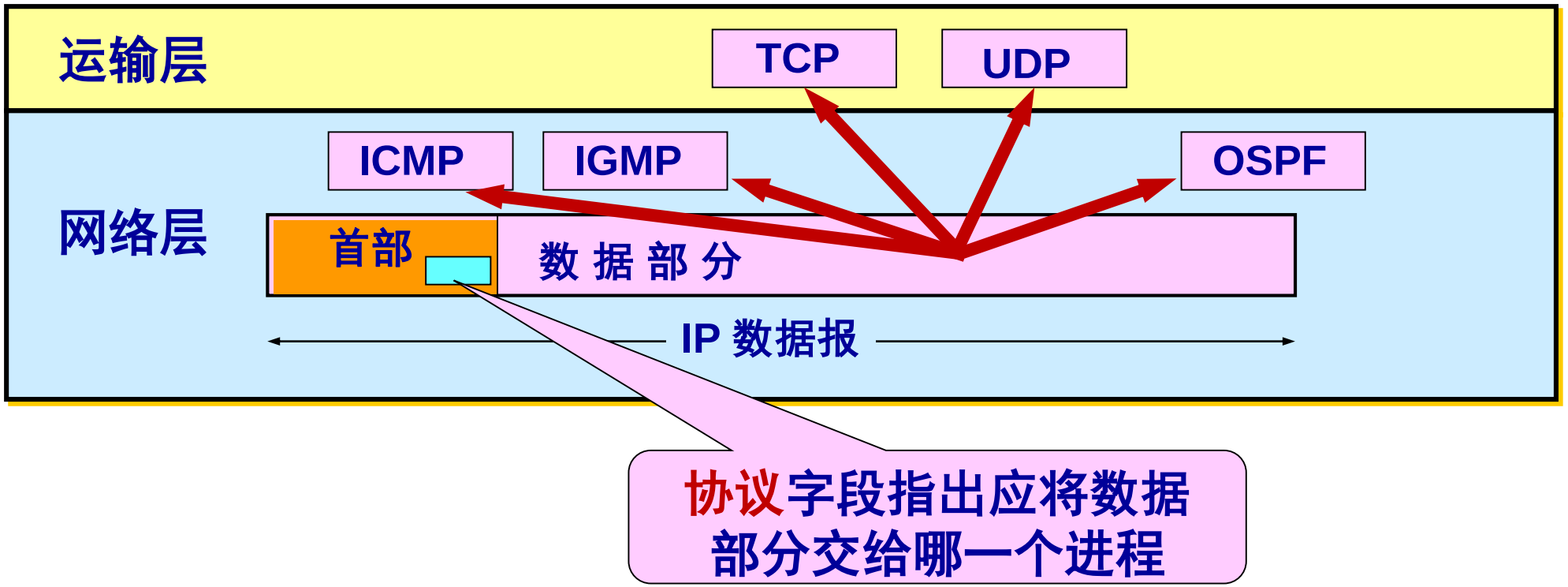
1. IP 数据报首部的固定部分中的各字段

生存时间——占8 位，记为 TTL (Time To Live)，
指示数据报在网络中可通过的路由器数的最大值。

1. IP 数据报首部的固定部分中的各字段

协议——占8 位，指出此数据报携带的数据使用何种协议，
以便目的主机的 IP 层将数据部分
上交给那个处理过程

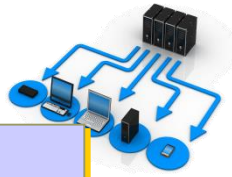
IP 协议支持多种协议，IP 数据报可以封装多种协议 PDU。



1. IP 数据报首部的固定部分中的各字段

首部检验和——占16 位，只检验数据报的首部，不检验数据部分。这里不采用 CRC 检验码而采用简单的计算方法。

IP 数据报首部检验和的计算采用 16 位二进制反码求和算法



发送端

数据报首部

字 1 — 16 位

字 2 — 16 位

…

检验和 — 置为全 0

…

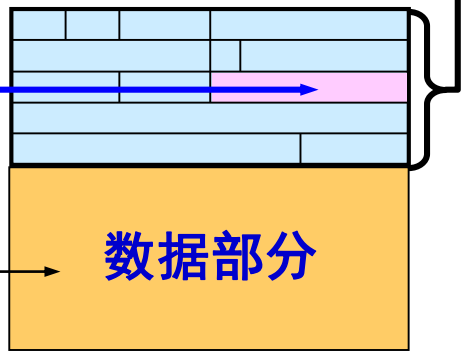
字 n — 16 位

反码算术运算求和 — 16 位

取反码

检验和 — 16 位

IP 数据报



数据部分

不参与检验和的计算

数据部分

接收端

字 1 — 16 位

字 2 — 16 位

…

检验和 — 16 位

…

字 n — 16 位

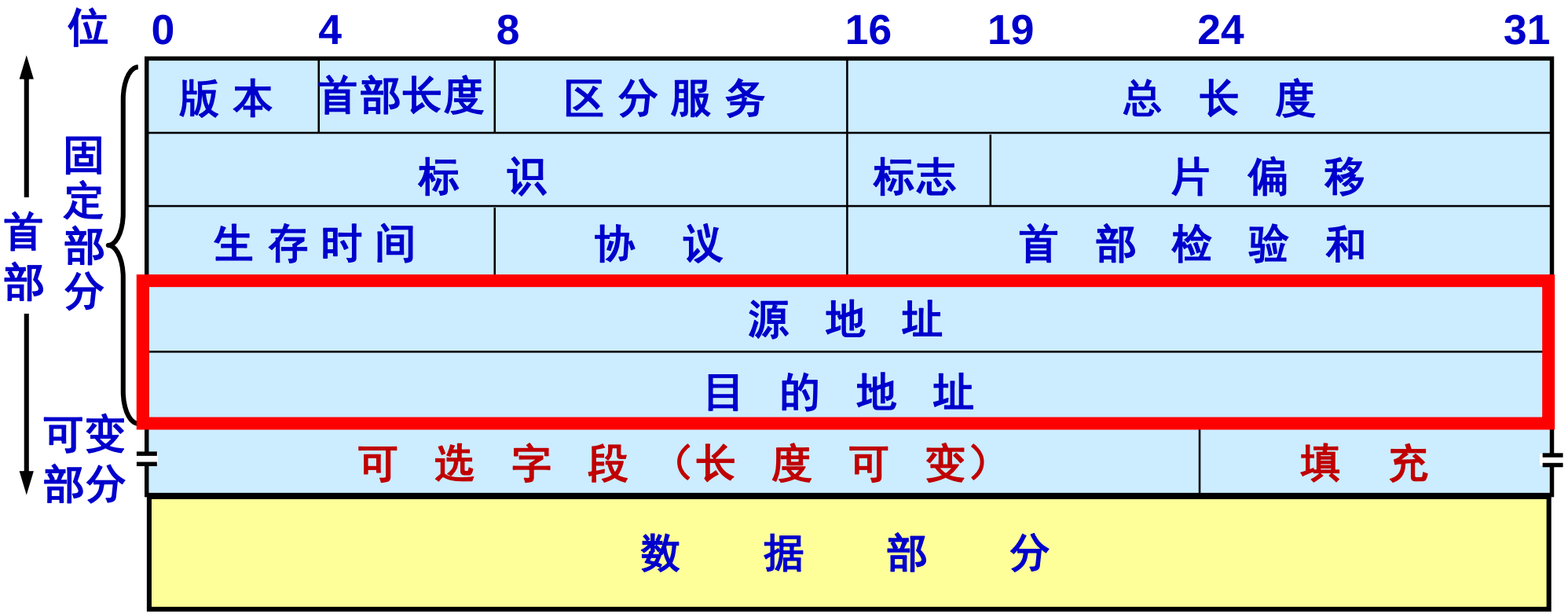
反码算术运算求和 — 16 位

取反码

结果 — 16 位

若结果为 0, 则保留；
否则，丢弃该数据报

1. IP 数据报首部的固定部分中的各字段



源地址和目的地址都各占 4 字节

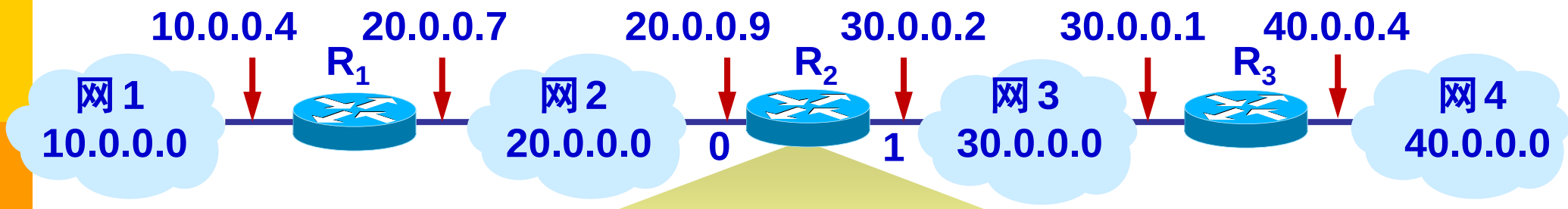
2. IP 数据报首部的可变部分

- IP 首部的可变部分就是一个选项字段，用来支持排错、测量以及安全等措施，内容很丰富。
- 选项字段的长度可变，从 1 个字节到 40 个字节不等，取决于所选择的项目。
- 增加首部的可变部分是为了增加 IP 数据报的功能，但这同时也使得 IP 数据报的首部长度成为可变的。这就增加了每一个路由器处理数据报的开销。
- 实际上这些选项很少被使用。

4.2.6 IP 层转发分组的流程

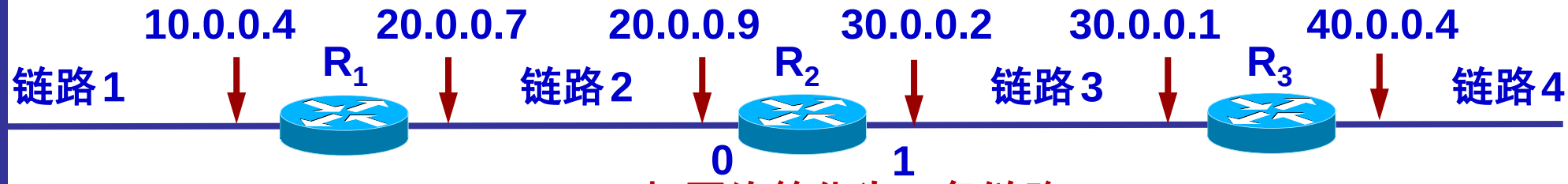
- 假设：有四个 A 类网络通过三个路由器连接在一起。每一个网络上都可能有成千上万个主机。
- 可以想像，若按目的主机号来制作路由表，每一个路由表就有 4 万个项目，即 4 万行（每一行对应于一台主机），则所得出的路由表就会过于庞大。
- 但若按主机所在的网络地址来制作路由表，那么每一个路由器中的路由表就只包含 4 个项目（每一行对应于一个网络），这样就可使路由表大大简化。

在路由表中，对每一条路由，最主要的是（目的网络地址，下一跳地址）



路由器 R_2 的路由表

目的主机所在的网络	下一跳地址
20.0.0.0	直接交付，接口 0
30.0.0.0	直接交付，接口 1
10.0.0.0	20.0.0.7
40.0.0.0	30.0.0.1



把网络简化为一条链路

查找路由表

根据目的网络地址就能确定下一跳路由器，这样做的结果是：

- IP 数据报最终一定可以找到目的主机所在目的网络上的路由器（可能要通过多次的间接交付）。
- 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。

关于路由表

- 路由表没有给分组指明到某个网络的完整路径。
- 路由表指出，到某个网络应当先到某个路由器（即下一跳路由器）。
- 在到达下一跳路由器后，再继续查找其路由表，知道再下一步应当到哪一个路由器。
- 这样一步一步地查找下去，直到最后到达目的网络。

特定主机路由

- 分组转发大都是基于目的主机所在的网络，但也有特例。
- 特定主机路由，这种路由是为特定的目的主机指明一个路由。
- 采用特定主机路由可使网络管理人员能更方便地控制网络和测试网络，同时也可在需要考虑某种安全问题时采用这种特定主机路由。

默认路由 (default route)

- 路由器还可采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。
- 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互联网连接，那么在这种情况下使用默认路由是非常合适的。
- 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。

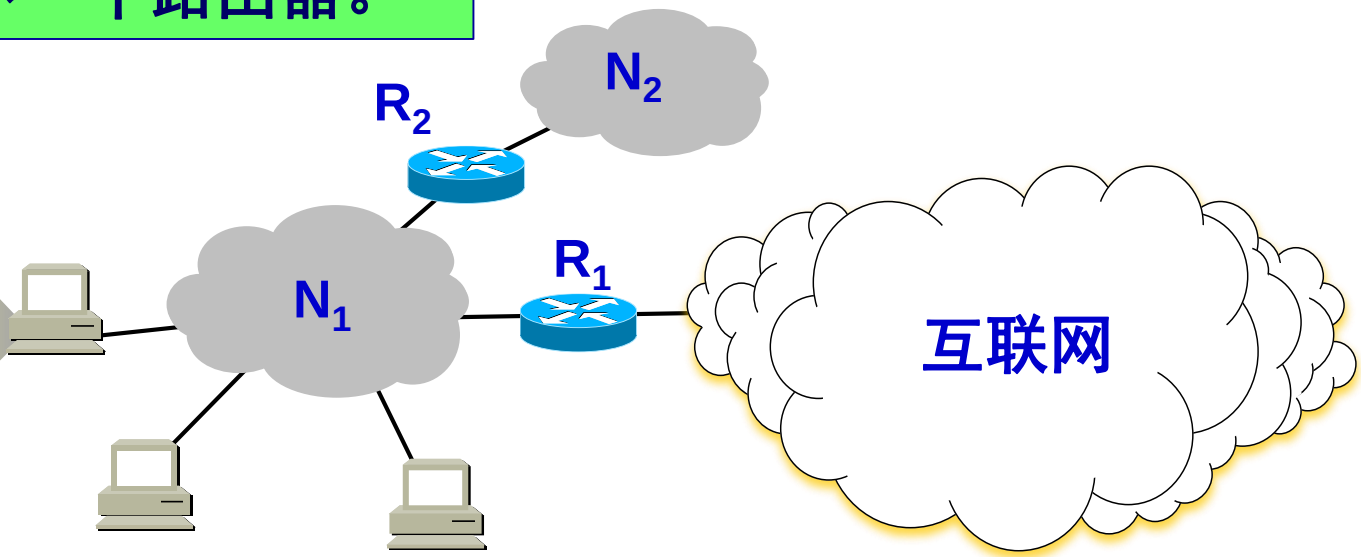
默认路由举例

只要目的网络不是 N_1 和 N_2，
就一律选择默认路由，
把数据报先间接交付路由器 R_1，
让 R_1 再转发给下一个路由器。

路由表

目的网络	下一跳
N_1	直接
N_2	R_2
默认	R_1

“直接”、“默认”
被记为0.0.0.0



路由器 R_1 充当网络 N_1 的默认路由器

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。

路由器分组转发算法

(1) 从数据报的首部提取目的主机的 IP 地址 D, 得出目的网络地址为 N。

(2) 若网络 *N* 与此路由器直接相连，则把数据报直接交付目的主机 *D*；否则是间接交付，执行 (3)。

(3) 若路由表中有目的地址为 *D* 的特定主机路由，则把数据报传送给路由表中所指明的下一跳路由器；否则，执行 (4)。

(4) 若路由表中有到达网络 *N* 的路由，则把数据报传送给路由表指明的下一跳路由器；否则，执行 (5)。

(5) 若路由表中有一个默认路由，则把数据报传送给路由表中所指明的默认路由器；否则，执行 (6)。

(6) 报告转发分组出错。

4.3 划分子网和构造超网

- 4.3.1 划分子网
- 4.3.2 使用子网时分组的转发
- 4.3.3 无分类编址 CIDR（构造超网）

4.3.1 划分子网

1. 从两级 IP 地址到三级 IP 地址

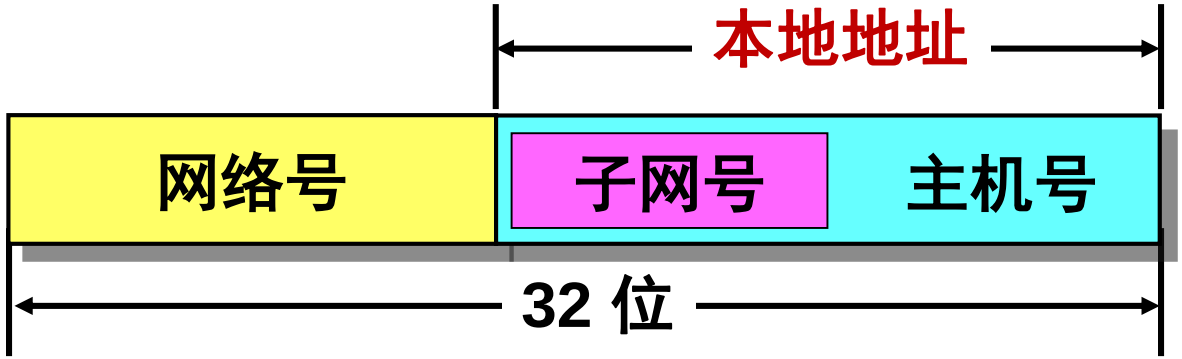
- 在 ARPANET 的早期，IP 地址的设计确实不够合理：
  - (1) IP 地址空间的利用率有时很低。
  - (2) 给每一个物理网络分配一个网络号会使路由表变得太大因而使网络性能变坏。
  - (3) 两级的 IP 地址不够灵活。

三级 IP 地址

- 从 1985 年起在 IP 地址中又增加了一个“子网号字段”，使两级的 IP 地址变成为三级的 IP 地址。
- 这种做法叫做划分子网 (subnetting) 或 子网寻址 或 子网路由选择。
- 划分子网已成为互联网的正式标准协议。

划分子网的基本思路

- 划分子网纯属一个**单位内部的事情**。单位对外仍然表现为一个网络。
- 从主机号**借用**若干个位作为**子网号** subnet-id，而主机号 host-id 也就相应减少了若干个位。
- 当没有划分子网时，IP 地址是两级结构。划分子网后 IP 地址就变成了三级结构。

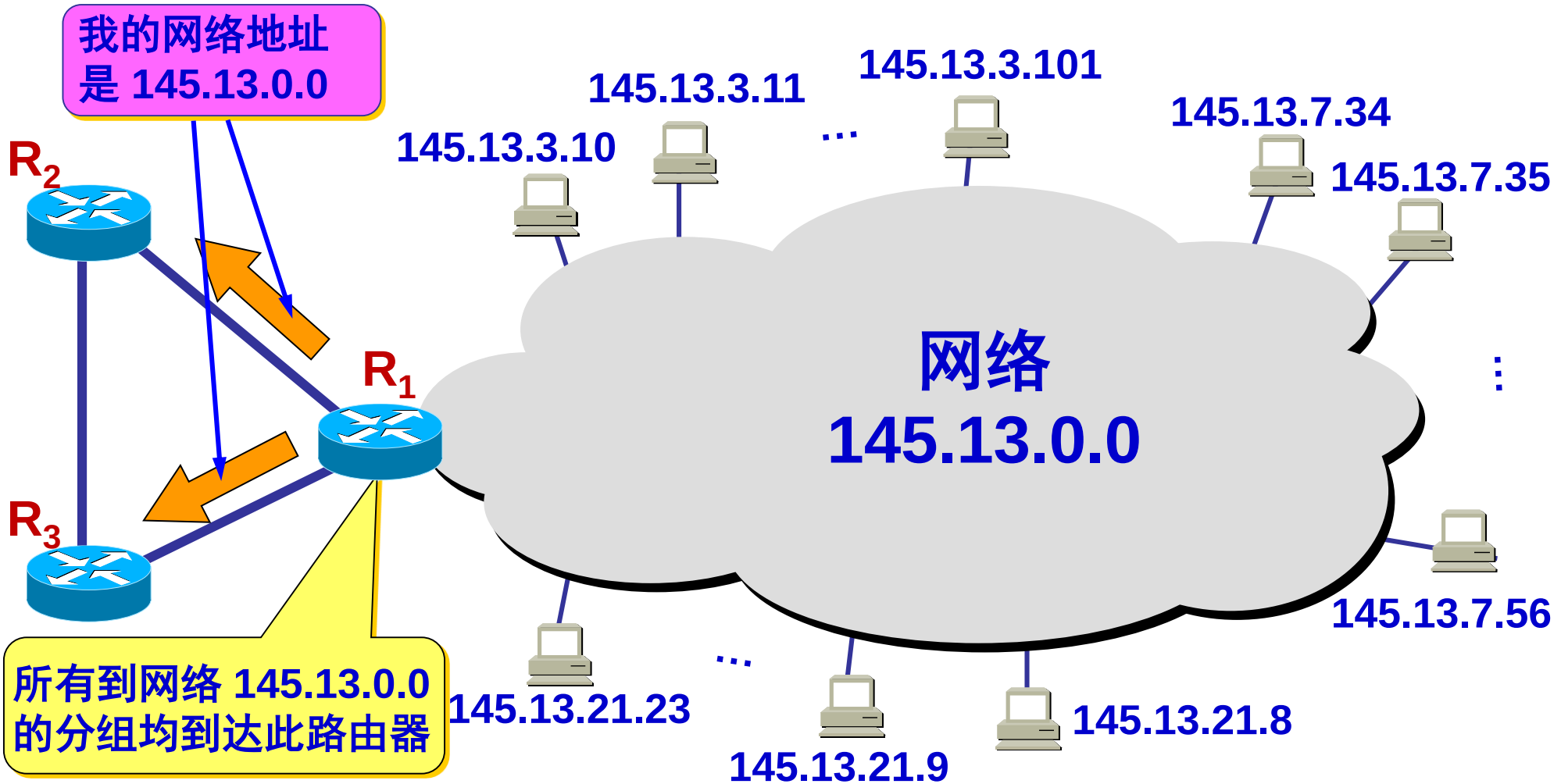


IP地址 ::= {<网络号>, <子网号>, <主机号>}　　(4-2)

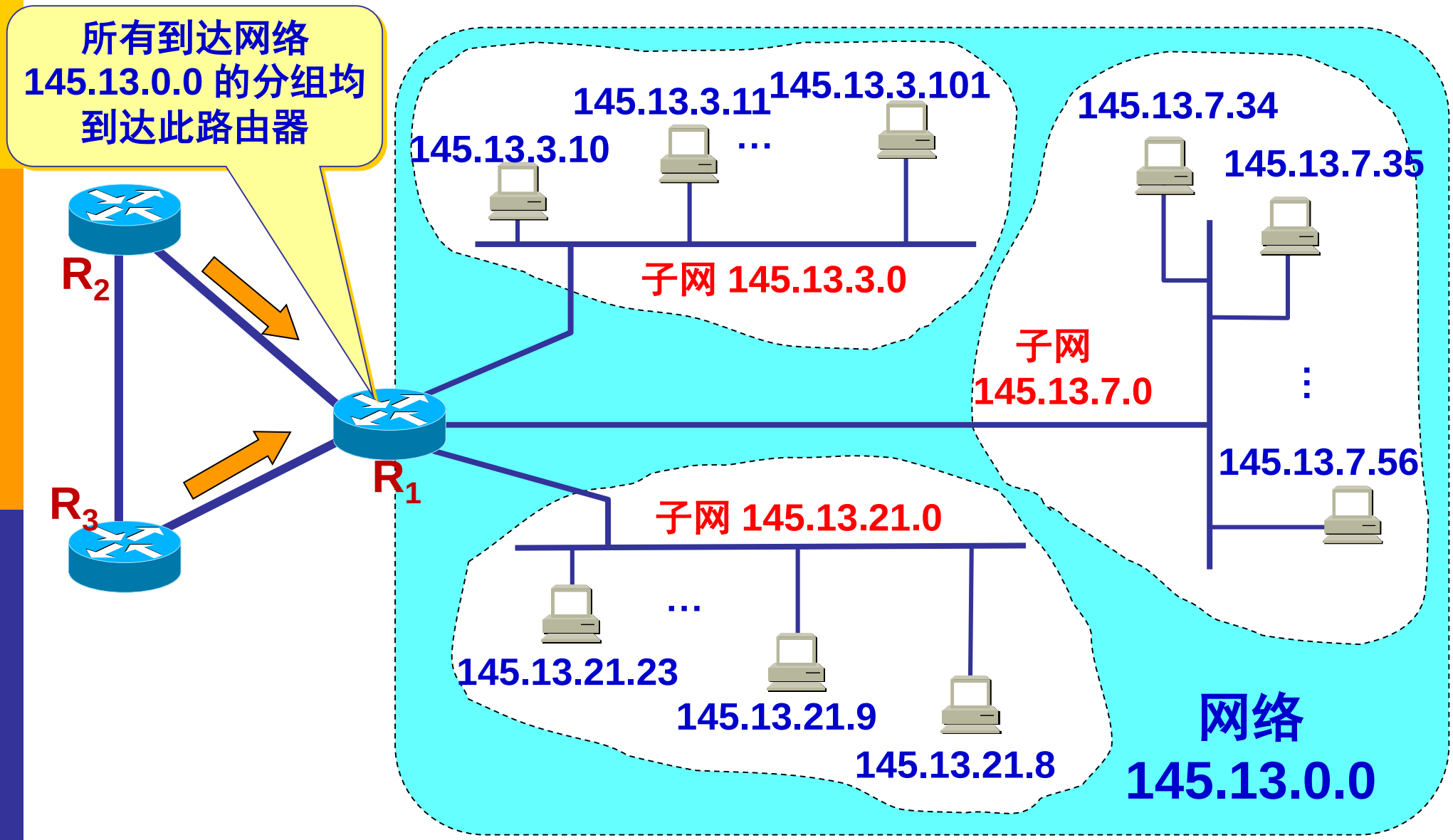
划分子网的基本思路（续）

- 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 数据报直接交付目的主机。
- 优点
  - 减少了 IP 地址的浪费
  - 使网络的组织更加灵活
  - 更便于维护和管理

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



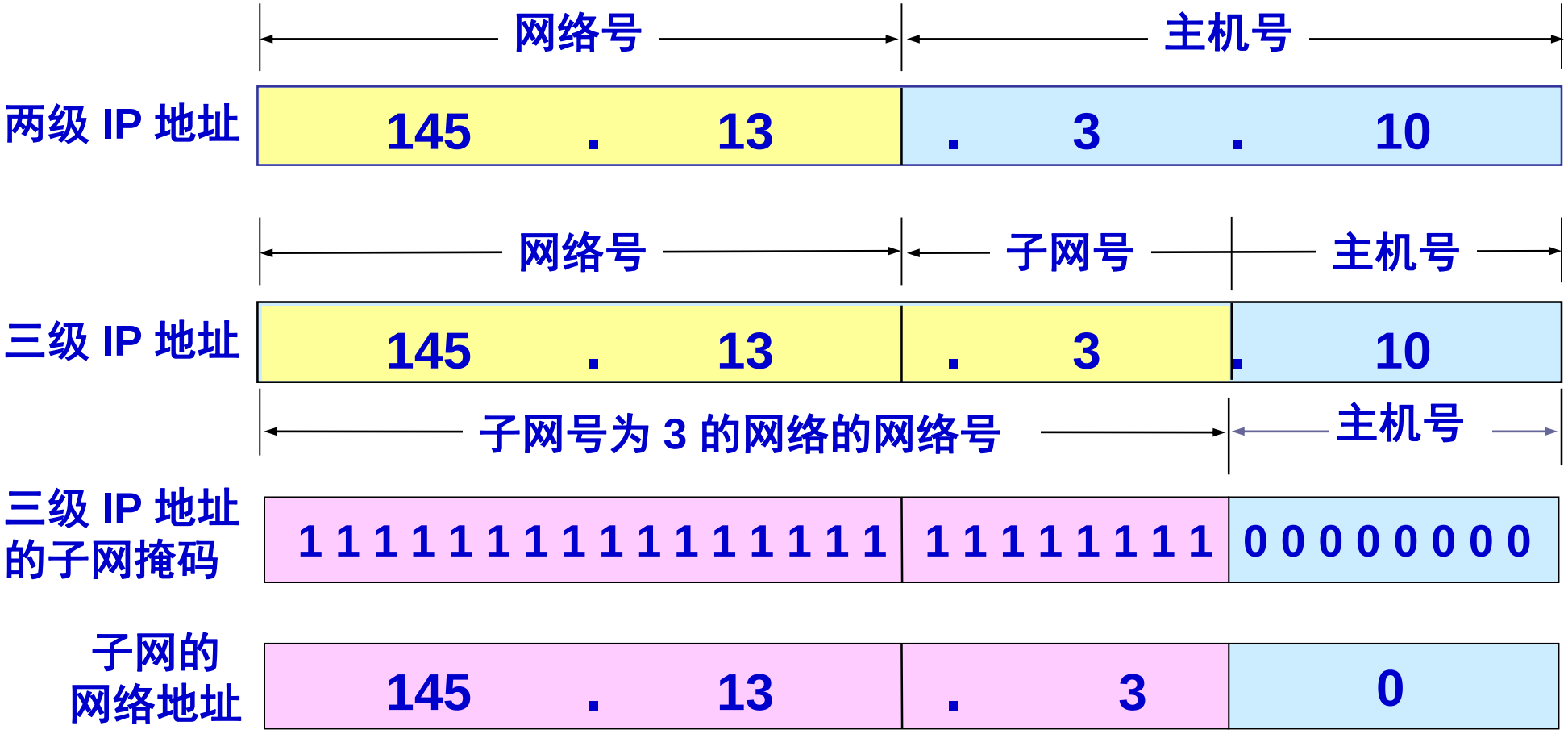
2. 子网掩码

- 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划分。
- 使用子网掩码 (subnet mask) 可以找出 IP 地址中的子网部分。

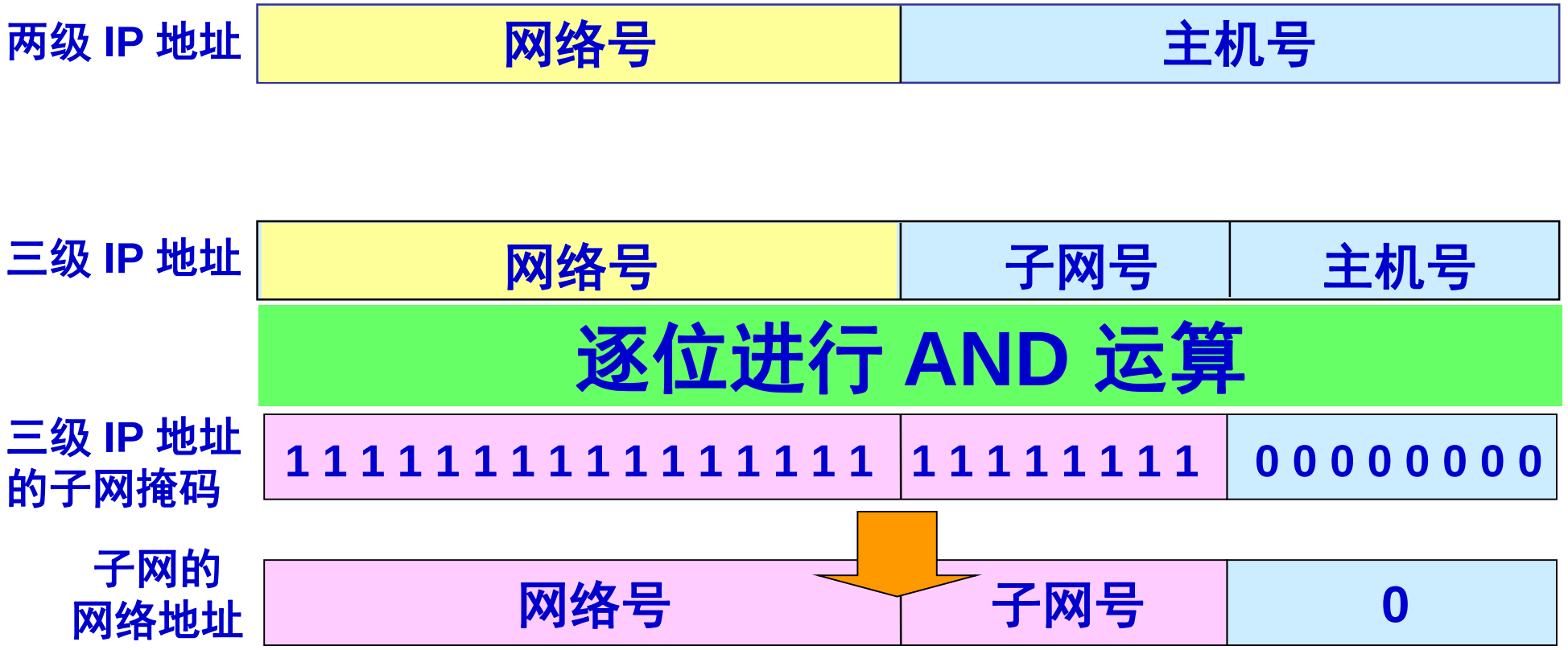
规则：

- 子网掩码长度 ＝ 32 位
- 某位 ＝ 1：IP地址中的对应位为网络号和子网号
- 某位 ＝ 0：IP地址中的对应位为主机号

IP 地址的各字段和子网掩码



(IP 地址) AND (子网掩码) =网络地址



默认子网掩码

A类地址	网络地址	网络号	主机号为全 0
	默认子网掩码 255.0.0.0	1 1 1 1 1 1 1 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
B类地址	网络地址	网络号	主机号为全 0
	默认子网掩码 255.255.0.0	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
C类地址	网络地址	网络号	主机号为全 0
	默认子网掩码 255.255.255.0	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	0 0 0 0 0 0 0 0

子网掩码是一个重要属性

- 子网掩码是一个网络或一个子网的重要属性。
- 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。

子网划分方法

- 有固定长度子网和变长子网两种子网划分方法。
- 在采用固定长度子网时，所划分的所有子网的子网掩码都是相同的。
- 若使用较少位数的子网号，则每一个子网上可连接的主机数就较多。
- 划分子网增加了灵活性，但却减少了能够连接在网络上的主机总数。

B 类地址的子网划分选择（使用固定长度子网）

子网号的位数	子网掩码	子网数	每个子网的主机数
2	255.255.192.0	2	16382
3	255.255.224.0	6	8190
4	255.255.240.0	14	4094
5	255.255.248.0	30	2046
6	255.255.252.0	62	1022
7	255.255.254.0	126	510
8	255.255.255.0	254	254
9	255.255.255.128	510	126
10	255.255.255.192	1022	62
11	255.255.255.224	2046	30
12	255.255.255.240	4094	14
13	255.255.255.248	8190	6
14	255.255.255.252	16382	2

表中的“子网号的位数”中没有 0, 1, 15 和 16 这四种情况，因为这没有意义。

【例4-2】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。

(a) 点分十进制表示的 IP 地址

141 . 14 . 72 . 24

(b) IP 地址的第 3 字节是二进制

141 . 14 .01001000. 24

(c) 子网掩码是 255.255.192.0

11111111 11111111 11000000 00000000

(d) IP 地址与子网掩码逐位相与

141 . 14 .01000000. 0

(e) 网络地址（点分十进制表示）

141 . 14 . 64 . 0

【例4-3】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。

不同的子网掩码得出**相同的网络地址**。
但不同的掩码的效果是不同的，可划分的子网数和每一个子网中的最大主机数是不一样的。

4.3.2 使用子网时分组的转发

- 在划分子网的情况下，从 IP 地址和数据报的首部并没有提供子网划分的信息。
- 因此分组转发的算法也必须做相应的改动。
- 注意，子网划分后，路由表必须包含以下三项内容：
  - 目的网络地址
  - 子网掩码
  - 下一条地址

在划分子网情况下路由器转发分组的算法

(1) 从收到的分组的首部提取目的 IP 地址 D。

(2) 先判断是否直接交付。用各网络的子网掩码和 *D* 逐位相“与”，看是否和相应的网络地址匹配。若匹配，则将分组直接交付。否则就是间接交付，执行 (3)。

(3) 若路由表中有目的地址为 *D* 的特定主机路由，则将分组传送给指明的下一跳路由器；否则，执行 (4)。

(4) 对路由表中的每一行，将子网掩码和 *D* 逐位相“与”。若结果与该行的目的网络地址匹配，则将分组传送给该行指明的下一跳路由器；否则，执行 (5)。

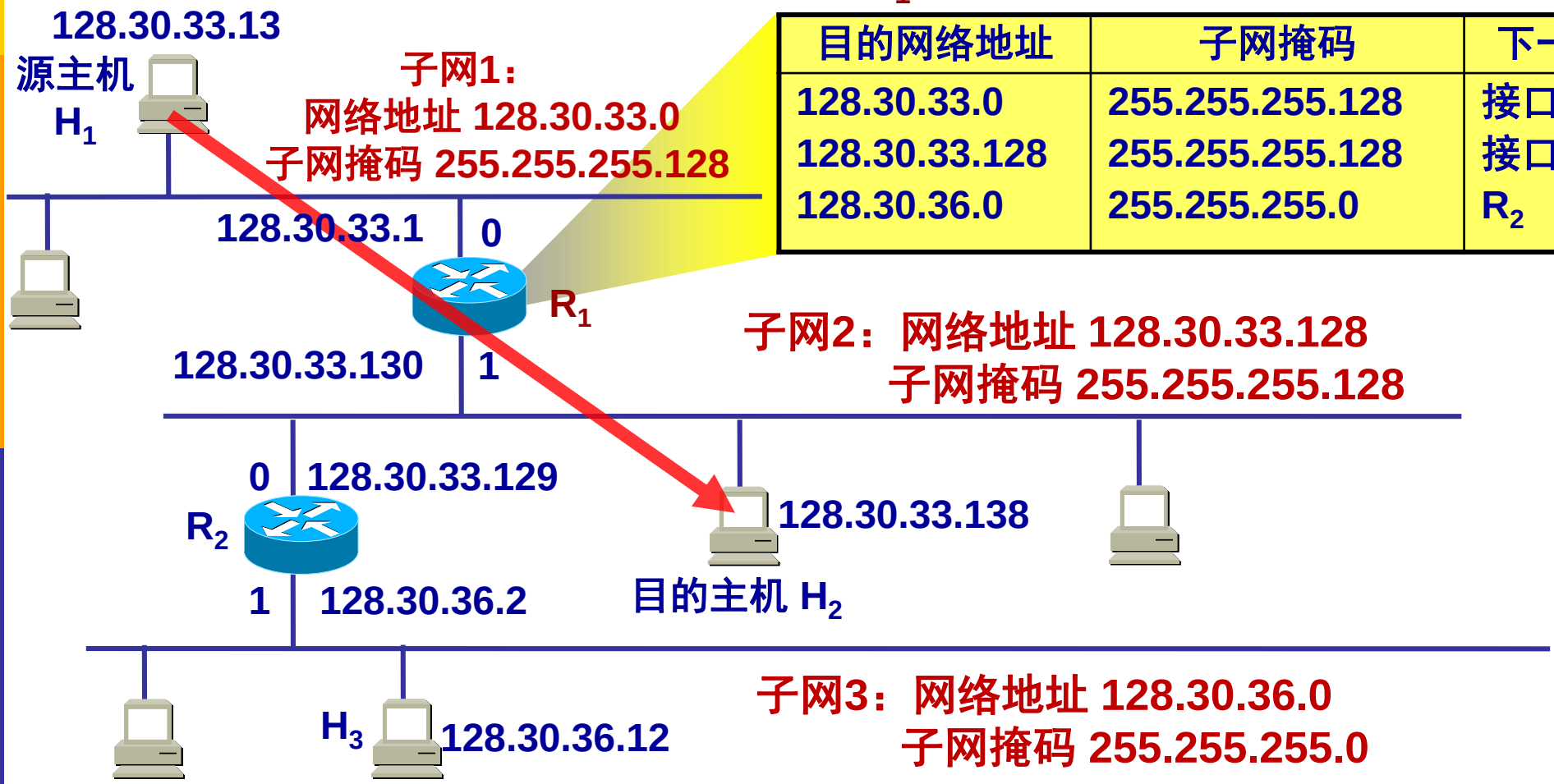
(5) 若路由表中有一个默认路由，则将分组传送给路由表中所指明的默认路由器；否则，执行 (6)。

(6) 报告转发分组出错。

【例4-4】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的过程。

R_1 的路由表（未给出默认路由器）

目的网络地址	子网掩码	下一跳
128.30.33.0	255.255.255.128	接口 0
128.30.33.128	255.255.255.128	接口 1
128.30.36.0	255.255.255.0	R_2

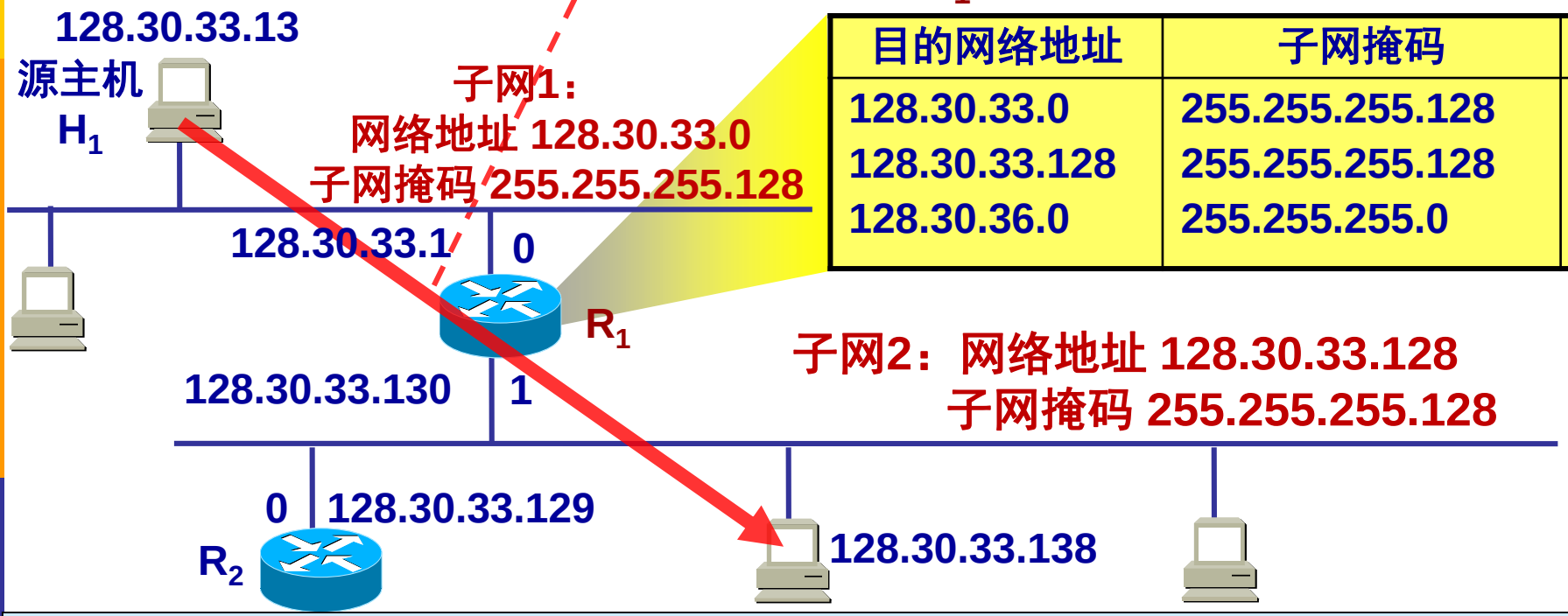


主机 H_1 要发送分组给 H_2

要发送的分组的目的 IP 地址：128.30.33.138

R_1 的路由表（未给出默认路由器）

目的网络地址	子网掩码	下一跳
128.30.33.0	255.255.255.128	接口 0
128.30.33.128	255.255.255.128	接口 1
128.30.36.0	255.255.255.0	R_2



请注意：H_1 并不知道 H_2 连接在哪一个网络上。

H_1 仅仅知道 H_2 的 IP 地址是

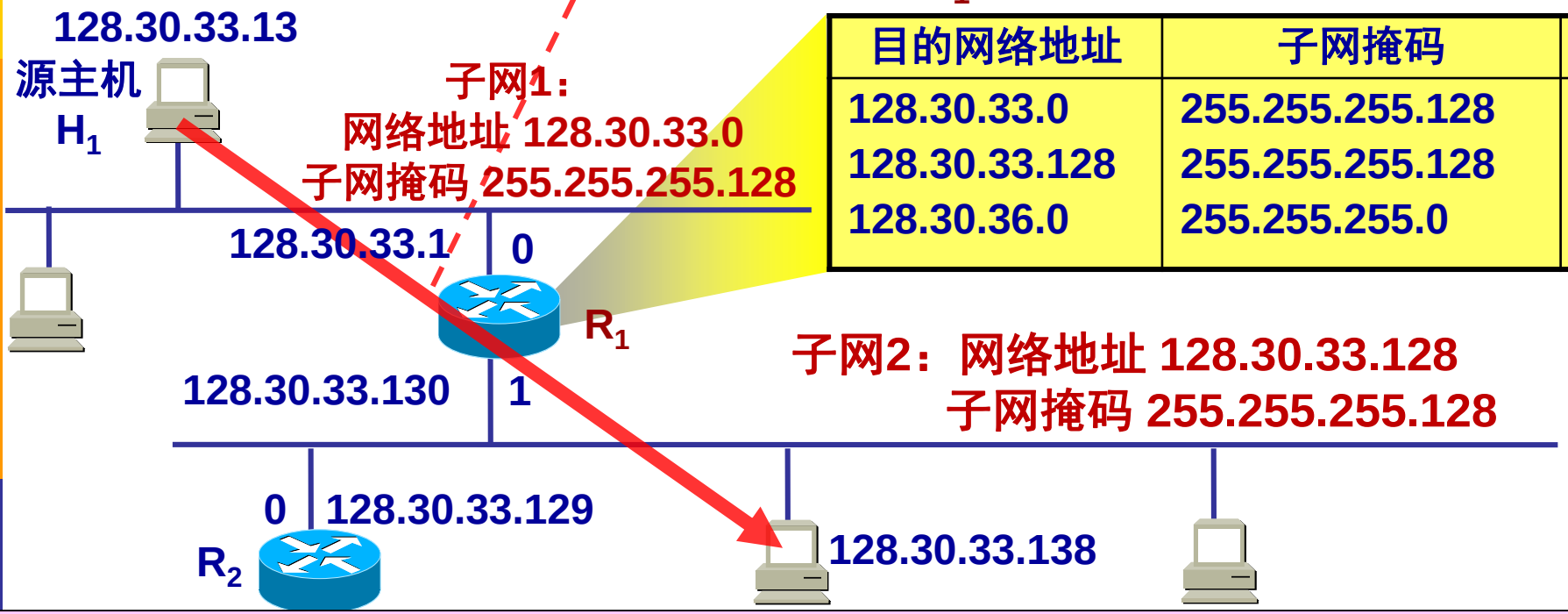
128.30.33.138

主机 H_1 要发送分组给 H_2

要发送的分组的目的 IP 地址：128.30.33.138

R_1 的路由表（未给出默认路由器）

目的网络地址	子网掩码	下一跳
128.30.33.0	255.255.255.128	接口 0
128.30.33.128	255.255.255.128	接口 1
128.30.36.0	255.255.255.0	R_2

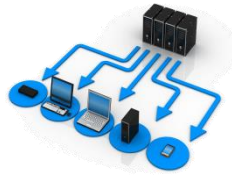


因此 H_1 首先检查主机 128.30.33.138 是否连接在本网络上

如果是，则直接交付；

否则，就送交路由器 R_1，并逐项查找路由表。

主机 H_1 首先将
本子网的子网掩码 255.255.255.128
与分组的 IP 地址 128.30.33.138 逐比特相“与”(AND 操作)



255.255.255.128 AND 128.30.33.138 的计算

255 就是二进制的全 1，因此 255 AND xyz = xyz，
这里只需计算最后的 128 AND 138 即可。

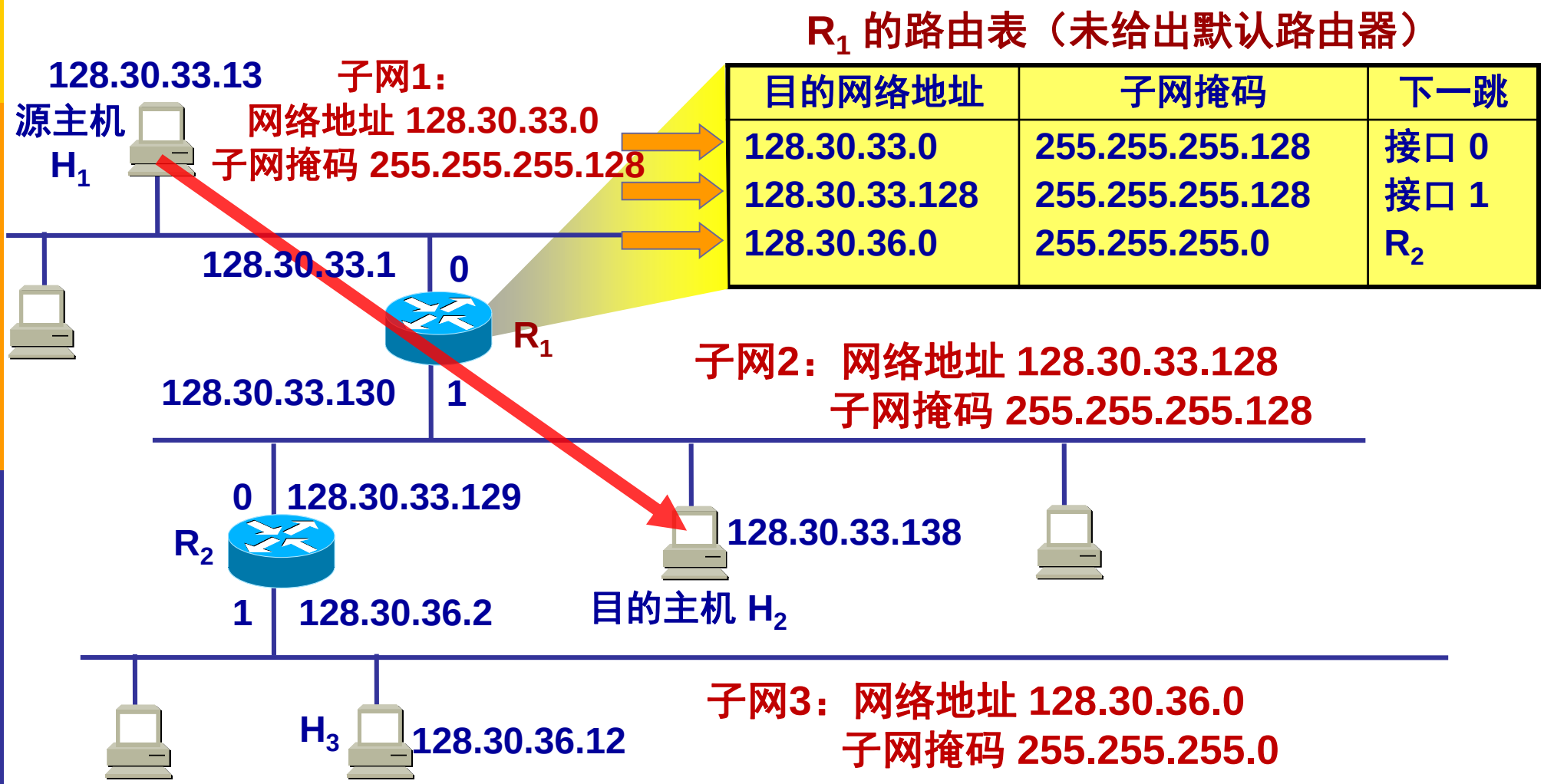
128 → 10000000
138 → 10001010

逐比特 AND 操作后　10000000 → 128

255.255.255.128
逐比特 AND 操作 128. 30. 33.138

128. 30. 33.128 ≠ H_1 的网络地址 128.30.33.0

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

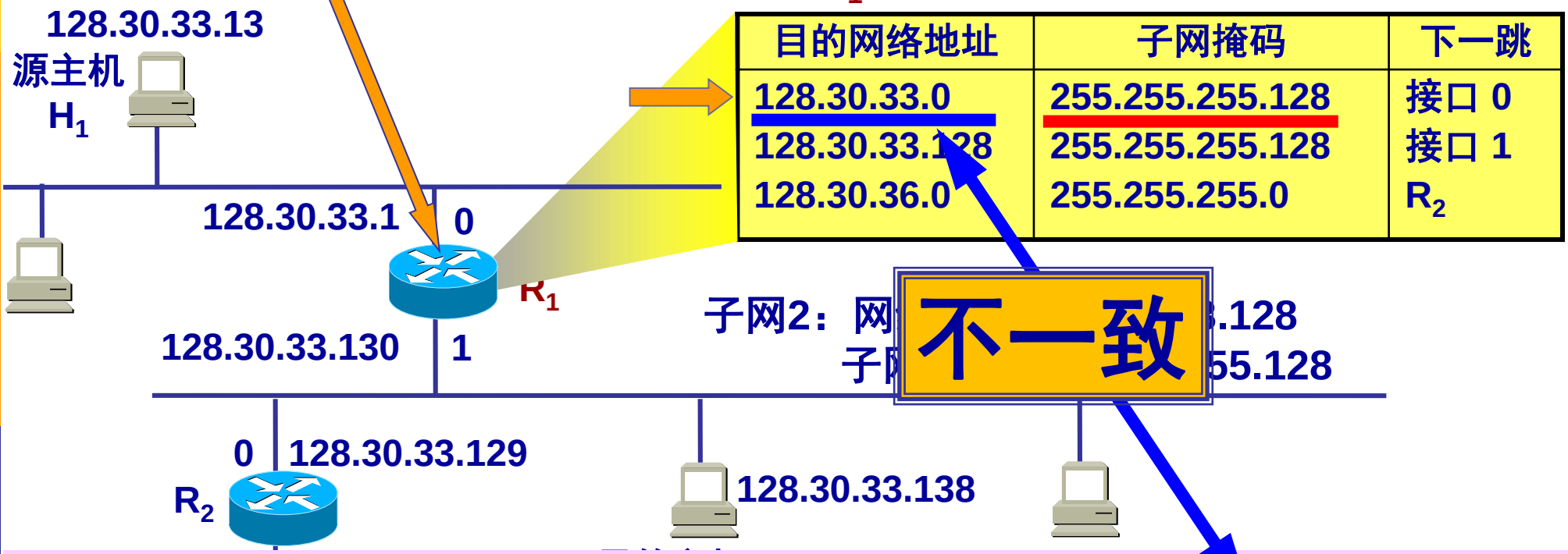


路由器 R_1 收到分组后就用路由表中第 1 个项目的子网掩码和 128.30.33.138 逐比特 AND 操作

R_1 收到的分组的目的 IP 地址：128.30.33.138

R_1 的路由表（未给出默认路由器）

目的网络地址	子网掩码	下一跳
128.30.33.0	255.255.255.128	接口 0
128.30.33.128	255.255.255.128	接口 1
128.30.36.0	255.255.255.0	R_2



255.255.255.128 AND 128.30.33.138 = 128.30.33.128

不匹配!

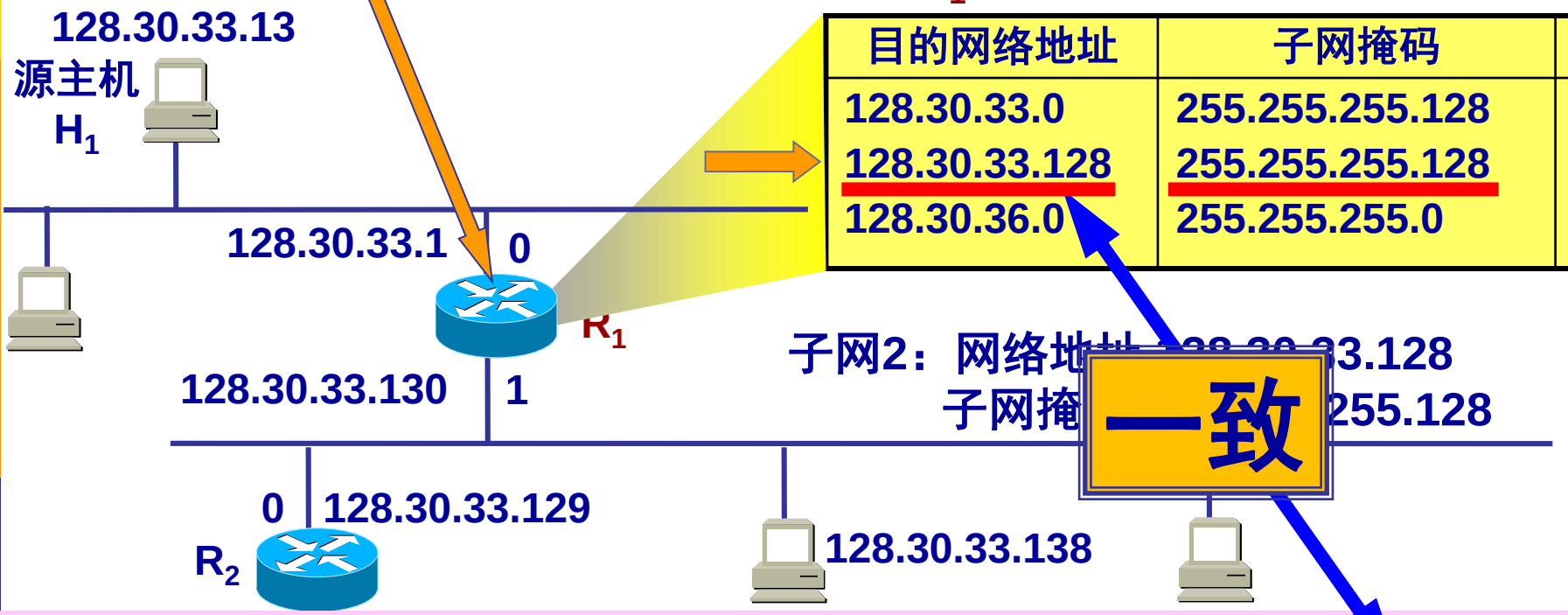
（因为128.30.33.128 与路由表中的 128.30.33.0 不一致）

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

R_1 收到的分组的目的 IP 地址：128.30.33.138

R_1 的路由表（未给出默认路由器）

目的网络地址	子网掩码	下一跳
128.30.33.0	255.255.255.128	接口 0
128.30.33.128	255.255.255.128	接口 1
128.30.36.0	255.255.255.0	R_2



255.255.255.128 AND 128.30.33.138 = 128.30.33.128

匹配!

这表明子网 2 就是收到的分组所要寻找的目的网络。

例题

- 某单位申请得到C类网络地址192.37.12.0,内部有2个LAN且没有扩充计划,给出合理的子网划分方案
  - 子网掩码？每个子网可容纳的主机数？地址范围？
- **解答：**
- 需要几位子网号？
  - 采用2位子网号，支持2^2-2=2个子网（全0和全1的子网不用）
- 子网掩码是多少？
  - 255.255.255.11000000，即为：255.255.255.192
- 每个子网有多少主机？地址范围？
  - 每个子网可有2^6-2=62台主机（全0表示本网络和全1表示广播地址）

一个练习

- 某网络的IP地址为181.189.0.0，子网掩码是255. 255. 192. 0(假设不支持CIDR)。请问：
  - （1）这个子网掩码最多可划分几个子网？（1分）
  - （2）写出每个子网的IP地址及其主机IP地址范围。（4分）

4.3.3 无分类编址 CIDR

1. 网络前缀

划分子网在一定程度上缓解了互联网在发展中遇到的困难。然而在 1992 年互联网仍然面临三个必须尽早解决的问题：

- (1) B类地址在 1992 年已分配了近一半，眼看就要在 1994 年 3 月全部分配完毕！
- (2) 互联网主干网上的路由表中的项目数急剧增长（从几千个增长到几万个）。
- (3) 整个 IPv4 的地址空间最终将全部耗尽。2011年IPv4地址已经耗尽了

IP 编址问题的演进

- 1987 年，RFC 1009 就指明了在一个划分子网的网络中可同时使用几个不同的子网掩码。
- 使用变长子网掩码 VLSM (Variable Length Subnet Mask)可进一步提高 IP 地址资源的利用率。
- 在 VLSM 的基础上又进一步研究出无分类编址方法，它的正式名字是无分类域间路由选择 CIDR (Classless Inter-Domain Routing)。

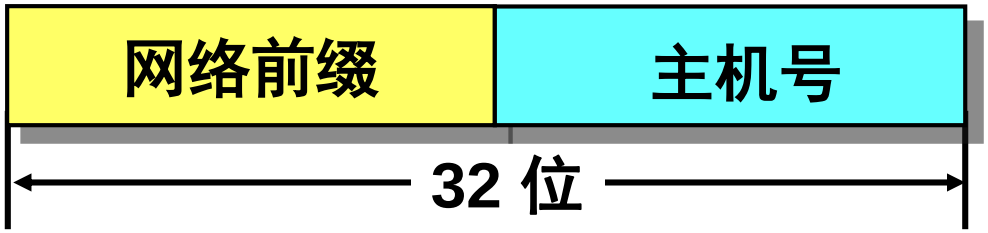
CIDR 最主要的特点

- 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。

IP地址 ::= {<网络前缀>, <主机号>}

无分类的两级编址

- 无分类的两级编址的记法是：

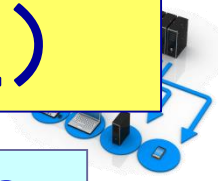


- CIDR 使用“斜线记法”(slash notation)，它又称为 CIDR 记法，即在 IP 地址面加上一个斜线“/”，然后写上网络前缀所占的位数（这个数值对应于三级编址中子网掩码中 1 的个数）。例如： 220.78.168.0/24
- CIDR 把网络前缀都相同的连续的 IP 地址组成“CIDR 地址块”。

CIDR 地址块

- 128.14.32.0/20 表示的地址块共有 2^{12} 个地址（因为斜线后面的 20 是网络前缀的位数，所以这个地址的主机号是 12 位）。
  - 这个地址块的起始地址是 128.14.32.0。
  - 在不需要指出地址块的起始地址时，也可将这样的地址块简称为“/20 地址块”。
  - 128.14.32.0/20 地址块的最小地址：128.14.32.0
  - 128.14.32.0/20 地址块的最大地址：128.14.47.255
  - 全 0 和全 1 的主机号地址一般不使用。

128.14.32.0/20 表示的地址（2^{12} 个地址）



最小地址 →

所有地址的 20 位前缀都是一样的

10000000 00001110 0010	0000 00000000
10000000 00001110 0010	0000 00000001
10000000 00001110 0010	0000 00000010
10000000 00001110 0010	0000 00000011
10000000 00001110 0010	0000 00000100
10000000 00001110 0010	0000 00000101
. . .	. . .
10000000 00001110 0010	1111 11111011
10000000 00001110 0010	1111 11111100
10000000 00001110 0010	1111 11111101
10000000 00001110 0010	1111 11111110
10000000 00001110 0010	1111 11111111

最大地址 →

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。
- 路由聚合有利于**减少**路由器之间的路由选择信息的交换，从而提高了整个互联网的性能。
- **路由聚合也称为构成超网 (supernetting)。**
- 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**）。
- 对于 **/20** 地址块，它的掩码是 20 个连续的 1。 斜线记法中的数字就是掩码中1的个数。

CIDR 记法的其他形式

- 10.0.0.0/10 可简写为 10/10，也就是把点分十进制中低位连续的 0 省略。
- 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

255 192 0 0

掩码中有 10 个连续的 1

- 网络前缀的后面加一个星号 * 的表示方法，如 00001010 00*，在星号 * 之前是网络前缀，而星号 * 表示 IP 地址中的主机号，可以是任意值。

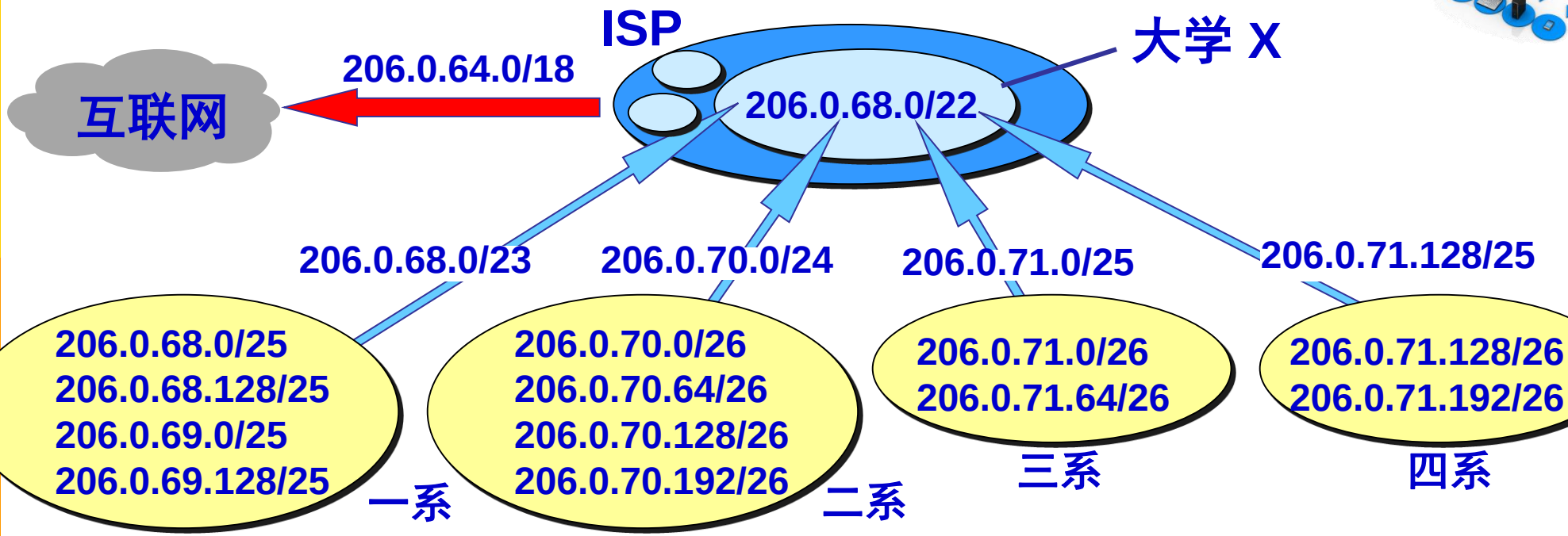
常用的 CIDR 地址块

CIDR 前缀长度	点分十进制	包含的地址数	相当于包含分类的网络数
/13	255.248.0.0	512 K	8 个 B类或 2048 个 C 类
/14	255.252.0.0	256 K	4 个 B 类或1024 个 C 类
/15	255.254.0.0	128 K	2 个 B 类或512 个 C 类
/16	255.255.0.0	64 K	1 个 B 类或256 个 C 类
/17	255.255.128.0	32 K	128 个 C 类
/18	255.255.192.0	16 K	64 个 C 类
/19	255.255.224.0	8 K	32 个 C 类
/20	255.255.240.0	4 K	16 个 C 类
/21	255.255.248.0	2 K	8 个 C 类
/22	255.255.252.0	1 K	4 个 C 类
/23	255.255.254.0	512	2 个 C 类
/24	255.255.255.0	256	1 个 C 类
/25	255.255.255.128	128	1/4 个 C 类
/26	255.255.255.192	64	1/4 个 C 类
/27	255.255.255.224	32	1/8 个 C 类

构成超网

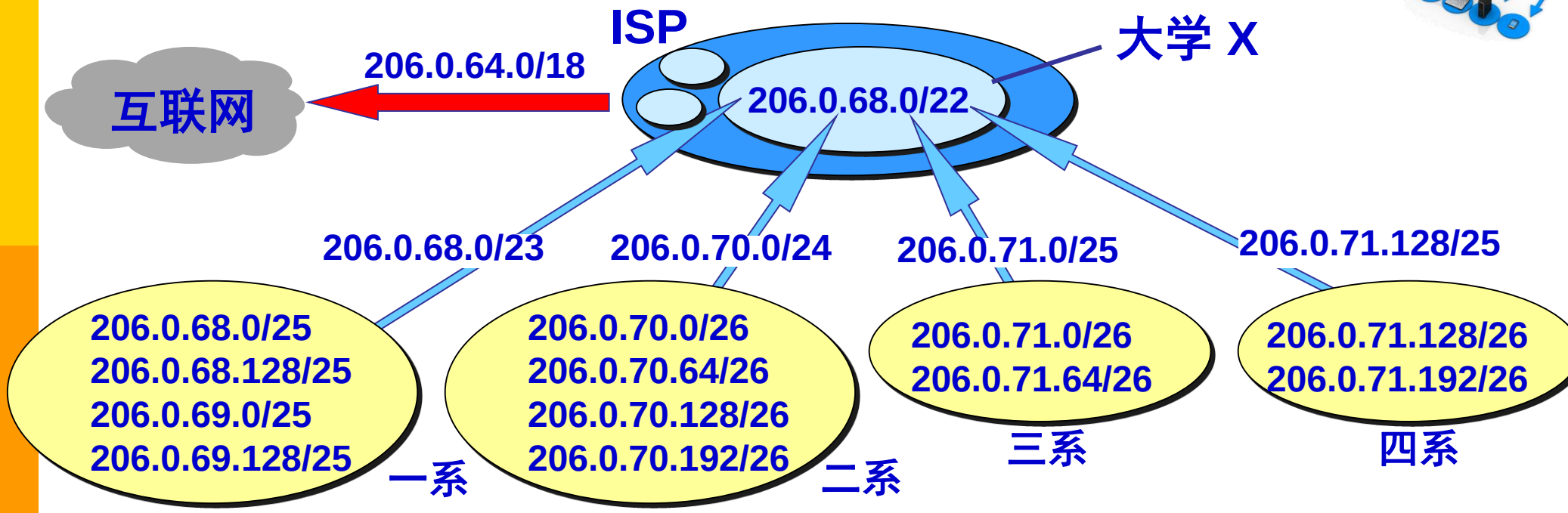
- 前缀长度不超过 23 位的 CIDR 地址块都包含了多个 C 类地址。
- 这些 C 类地址合起来就构成了超网。
- CIDR 地址块中的地址数一定是 2 的整数次幂。
- 网络前缀越短，其地址块所包含的地址数就越多。
- 而在三级结构的IP地址中，划分子网是使网络前缀变长。
- CIDR 的一个好处是：可以更加有效地分配 IPv4 的地址空间，可根据客户的需要分配适当大小的 CIDR 地址块。

CIDR 地址块划分举例



单位	地址块	二进制表示	地址数
ISP	206.0.64.0/18	11001110.00000000.01*	16384
大学	206.0.68.0/22	11001110.00000000.010001*	1024
一系	206.0.68.0/23	11001110.00000000.0100010*	512
二系	206.0.70.0/24	11001110.00000000.01000110.*	256
三系	206.0.71.0/25	11001110.00000000.01000111.0*	128
四系	206.0.71.128/25	11001110.00000000.01000111.1*	128

CIDR 地址块划分举例



这个 ISP 共有 64 个 C 类网络。如果不采用 CIDR 技术，则在与该 ISP 的路由器交换路由信息的每一个路由器的路由表中，就需要有 64 个项目。但采用地址聚合后，只需用路由聚合后的 1 个项目 206.0.64.0/18 就能找到该 ISP。

- 有四个/24的地址块，分别是212.56.132.0/24，212.56.133.0/24，212.56.134.0/24，212.56.135.0/24，试计算其最大可能的聚合地址是多少？

2. 最长前缀匹配

- 使用 CIDR 时，路由表中的每个项目由“网络前缀”和“下一跳地址”组成。在查找路由表时可能会得到不止一个匹配结果。
- 应当从匹配结果中选择具有最长网络前缀的路由：最长前缀匹配 (longest-prefix matching)。
- 网络前缀越长，其地址块就越小，因而路由就越具体 (more specific) 。
- 最长前缀匹配又称为最长匹配或最佳匹配。

最长前缀匹配举例

收到的分组的目的地址 *D* = 206.0.71.130

路由表中的项目：206.0.68.0/22　　1

206.0.71.128/25　　2

查找路由表中的第 1 个项目：

第 1 个项目 206.0.68.0/22 的掩码 *M* 有 22 个连续的 1。

M = 11111111 11111111 11111100 00000000

因此只需把 *D* 的第 3 个字节转换成二进制。

	M =	11111111	11111111	11111100 00000000
AND	*D* =	206.	0.	01000111. 130
		206.	0.	01000100. 0

与 206.0.68.0/22 匹配!

最长前缀匹配举例

收到的分组的目的地址 *D* = 206.0.71.130

路由表中的项目：206.0.68.0/22　　1

206.0.71.128/25　　2

查找路由表中的第 2 个项目：

第 2 个项目 206.0.71.128/25 的掩码 *M* 有 25 个连续的 1。

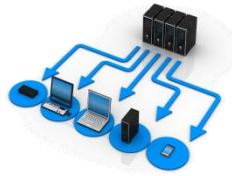
M = 11111111 11111111 11111100 00000000

因此只需把 *D* 的第 4 个字节转换成二进制。

	M =	11111111.	11111111.	11111111.	10000000
AND	*D* =	206.	0.	71.	10000010
		206.	0.	71.	10000000

与 206.0.71.128/25 匹配!

最长前缀匹配举例



D AND (11111111 11111111 11111100 00000000)
= 206.0.68.0/22　　匹配

D AND (11111111 11111111 11111111 10000000)
= 206.0.71.128/25　　匹配

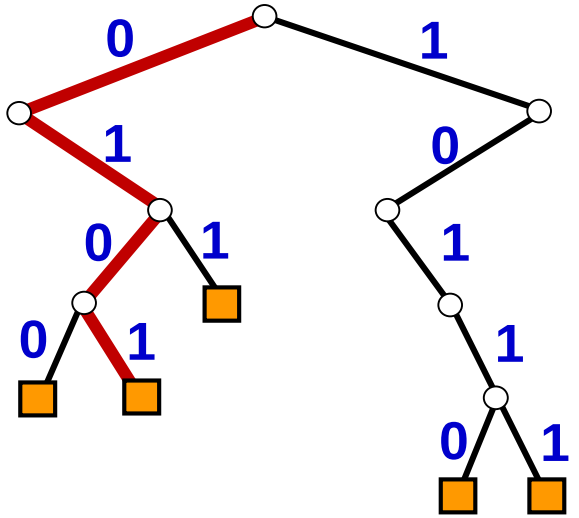
选择两个匹配的地址中更具体的一个，即选择**最长前缀的地址**。

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。
- 为了进行更加有效的查找，通常是将无分类编址的路由表存放在一种层次的数据结构中，然后自上而下地按层次进行查找。这里最常用的就是二叉线索 (binary trie)。
- IP 地址中从左到右的比特值决定了从根结点逐层向下层延伸的路径，而二叉线索中的各个路径就代表路由表中存放的各个地址。
- 为了提高二叉线索的查找速度，广泛使用了各种压缩技术。

用 5 个前缀构成的二叉线索

32 位的 IP 地址	唯一前缀
01000110 00000000 00000000 00000000	0100
01010110 00000000 00000000 00000000	0101
01100001 00000000 00000000 00000000	011
10110000 00000010 00000000 00000000	10110
10111011 00001010 00000000 00000000	10111



从二叉线索的根节点自顶向下的深度最多有 32 层，每一层对应于IP地址中的一位。一个IP地址存入二叉线索的规则很简单。先检查IP地址左边的第一位，如为 0，则第一层的节点就在根节点的左下方；如为 1，则在右下方。然后再检查地址的第二位，构造出第二层的节点。依此类推，直到唯一前缀的最后一位。