第4章 网络层

河北师范大学软件学院
Software College of Hebei Normal University

回顾

点对点信道　广播信道　帧　封装成帧

透明传输　差错检测　字节填充

零比特填充　以太网　CSMA/CD

争用期　最短帧长　MAC格式

集线器和网桥工作原理

碰撞域　广播域

虚拟局域网

指引

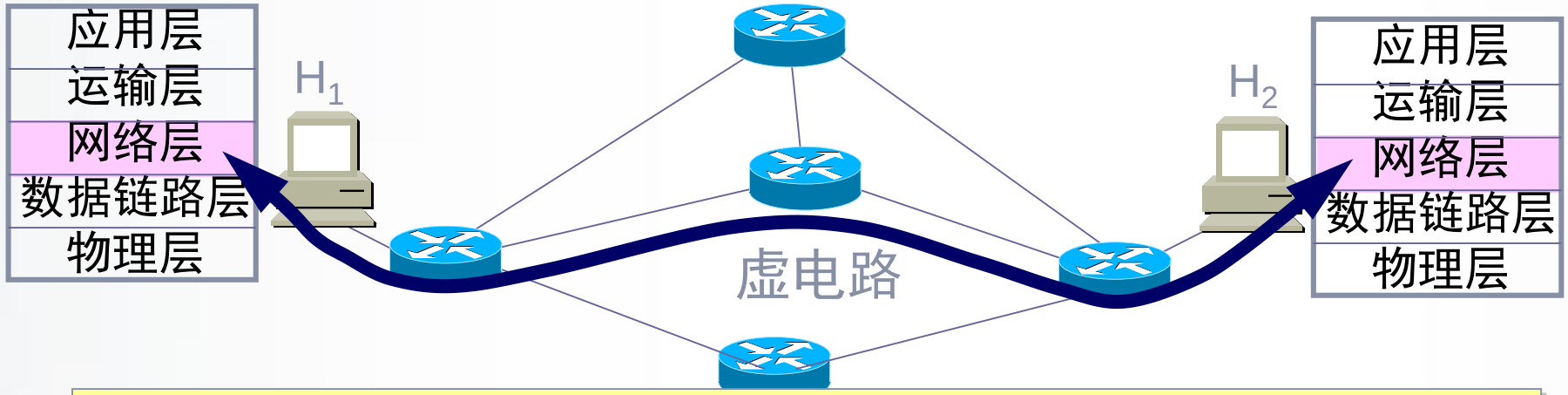
- 网络层提供的两种服务
- 网际协议 IP
- 划分子网和构造超网
- 网际控制报文协议 ICMP
- 因特网的路由选择协议
- IPv6
- IP 多播
- 虚拟专用网 VPN 和网络地址转换 NAT

网络层提供的两种服务

- 网络层关注的是如何将分组从源端沿着网络路径送达目的端。
- 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？
- 两种服务：网络层应该向运输层提供怎样的服务？
  - 虚电路服务
  - 数据报服务

电信网：虚电路

- 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。



H_1 发送给 H_2 的所有分组都沿着同一条虚电路传送

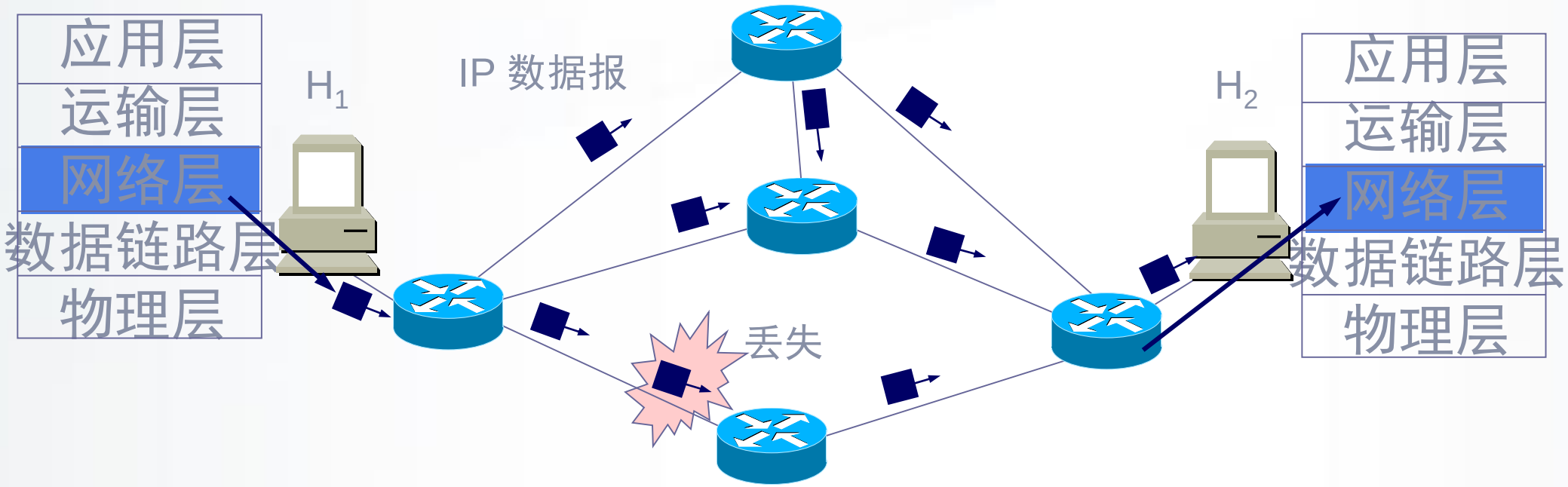
因特网：数据报服务

- 网络层向上只提供简单灵活的、无连接的、尽最大努力交付的数据报服务。
- 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）独立发送，与其前后的分组无关（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。

因特网：数据报服务

- 尽最大努力交付的好处：
  - 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（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。
  - 因特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。

因特网：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

虚电路与数据报服务比较

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

指引

- 网络层提供的两种服务
- 网际协议 IP
  - 虚拟互联网
  - IP地址
  - IP地址与硬件地址
  - IP数据报格式
  - IP转发分组的流程
- 划分子网和构造超网
- 网际控制报文协议 ICMP
- 因特网的路由选择协议
- IPv6
- IP 多播
- 虚拟专用网 VPN 和网络地址转换 NAT

网络互连的设备

- 中间设备又称为中间系统或中继(relay)系统。
  - 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)。
  - 网络层以上的中继系统：网关(gateway)。

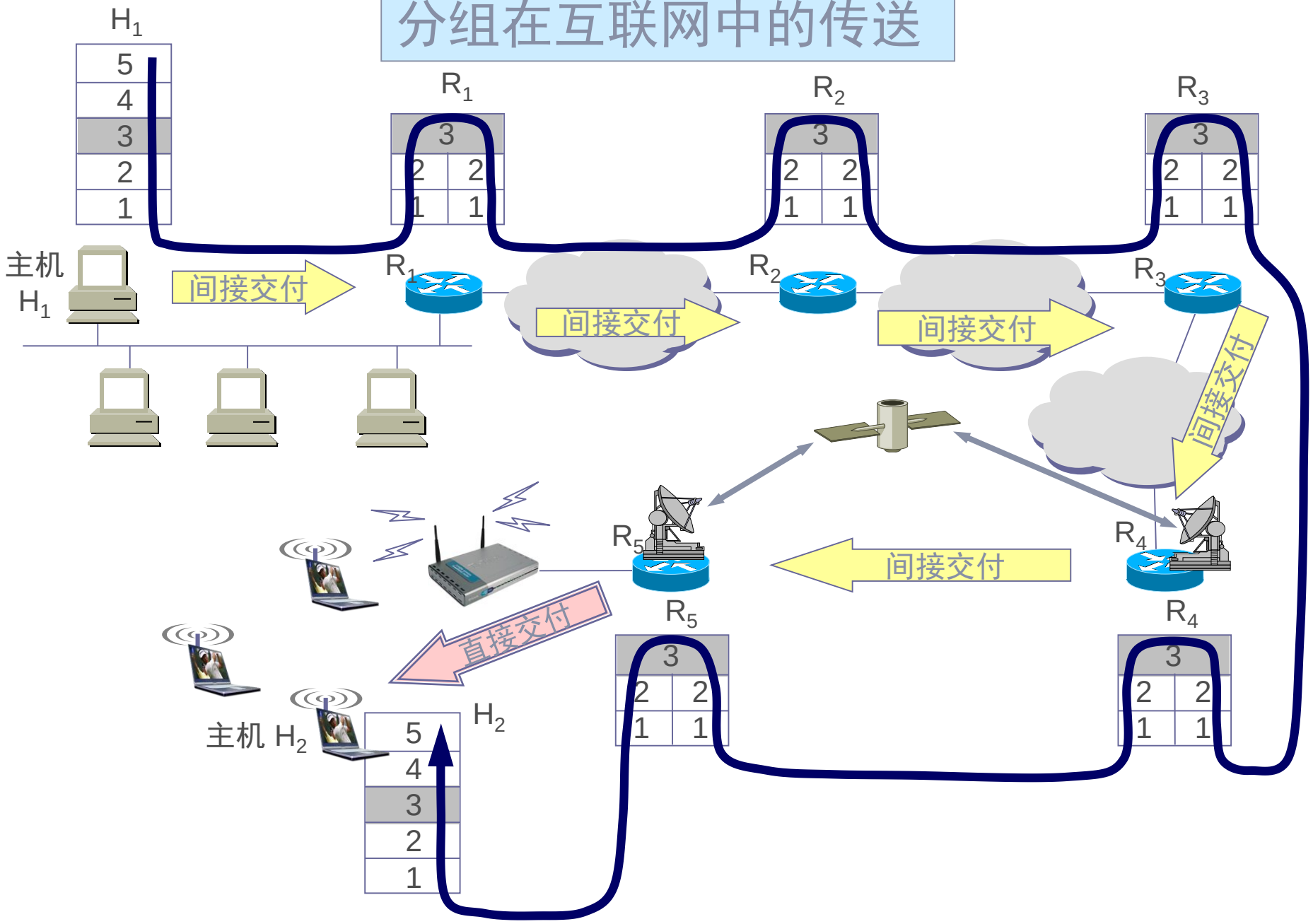
网络互连的设备：路由器

- 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。

网络互联的问题

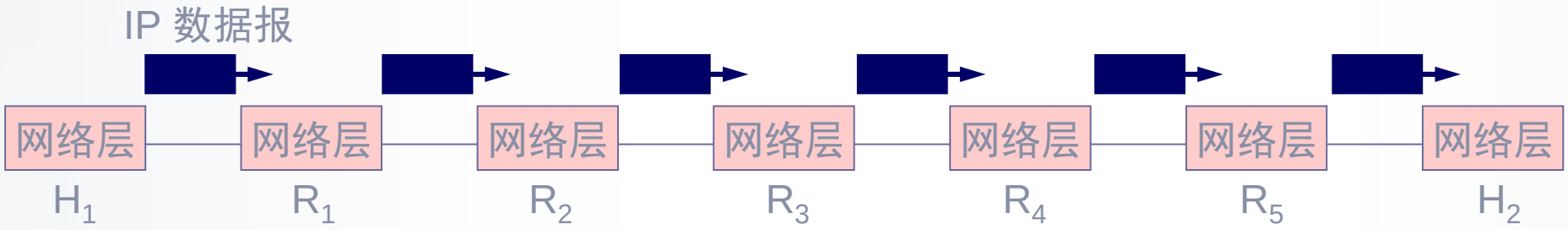
- 互连在一起的网络要进行通信，会遇到许多问题需要解决，如：
  - 不同的寻址方案
  - 不同的最大分组长度
  - 不同的网络接入机制
  - 不同的超时控制
  - 不同的差错恢复方法
  - 不同的状态报告方法
  - 不同的路由选择技术
  - 不同的用户接入控制
  - 不同的服务（面向连接服务和无连接服务）
  - 不同的管理与控制方式

分组在互联网中的传送

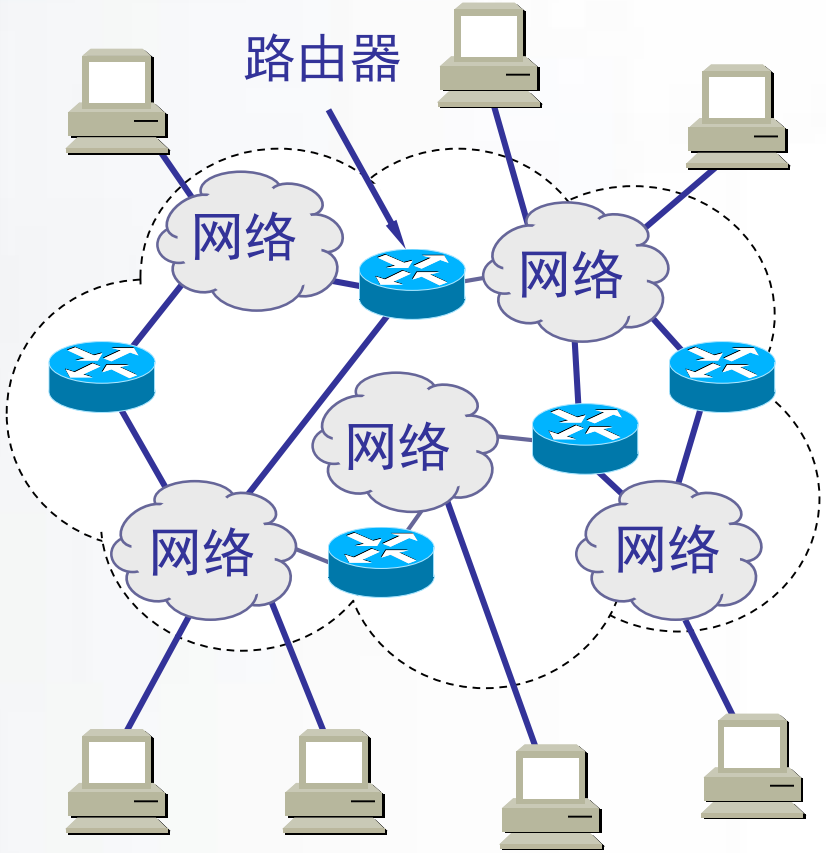

H1
5
4
3
2
1
R1
3
2 2
1 1
R2
3
2 2
1 1
R3
3
2 2
1 1
主机
H1
间接交付
R1
间接交付
R2
间接交付
R3
间接交付
R5
R4
间接交付
R5
R4
3
2 2
1 1
3
2 2
1 1
直接交付
主机 H2
H2
5
4
3
2
1

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。



互连网络与虚拟互连网络



(a) 互连网络

(b) 虚拟互连网络

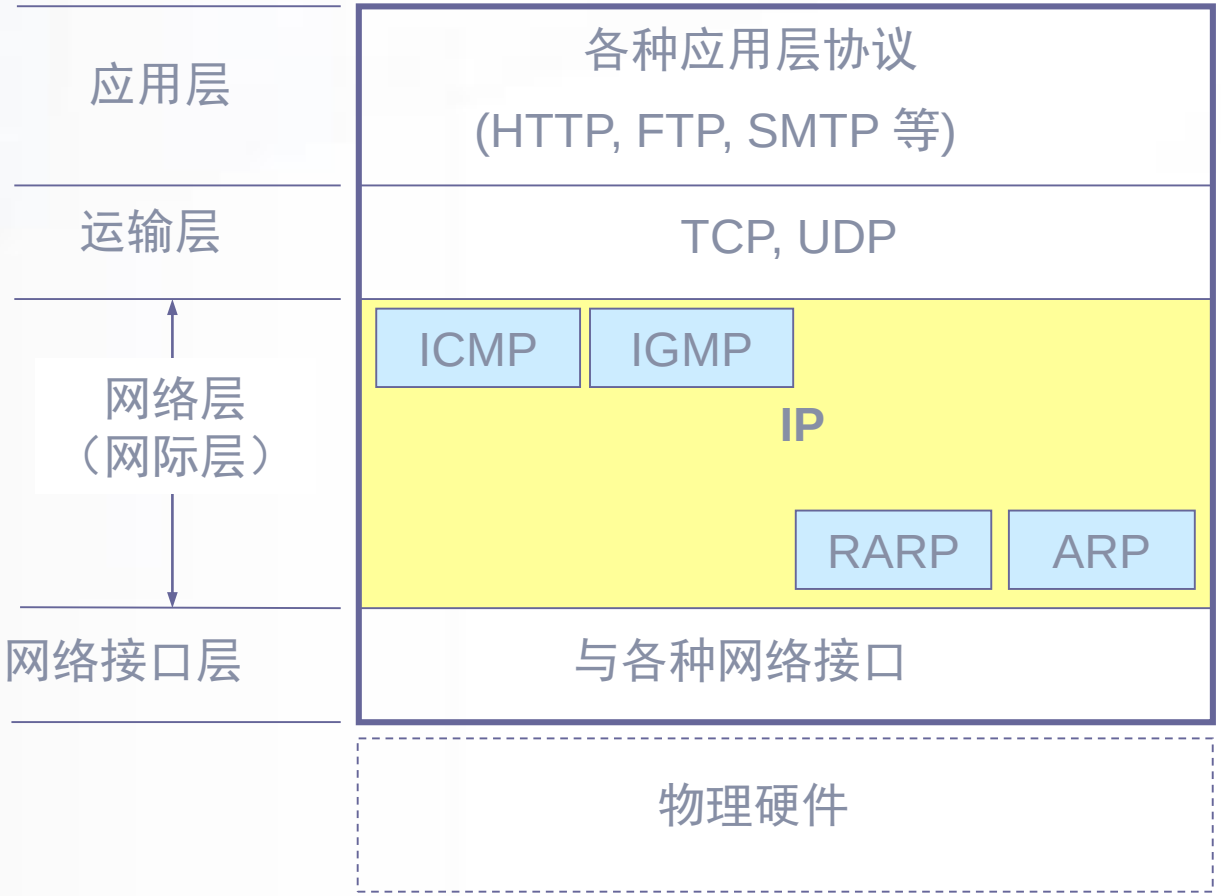
虚拟互连网络的意义

- 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从用户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。

IP协议简介

- 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)

IP协议简介


应用层
各种应用层协议
(HTTP, FTP, SMTP 等)
运输层
TCP, UDP
网络层
（网际层）
ICMP
IGMP
IP
RARP
ARP
网络接口层
与各种网络接口
物理硬件

指引

- 网络层提供的两种服务
- 网际协议 IP
  - 虚拟互联网
  - IP地址
  - IP地址与硬件地址
  - IP数据报格式
  - IP转发分组的流程
- 划分子网和构造超网
- 网际控制报文协议 ICMP
- 因特网的路由选择协议
- IPv6
- IP 多播
- 虚拟专用网 VPN 和网络地址转换 NAT

IP地址及其表示方法

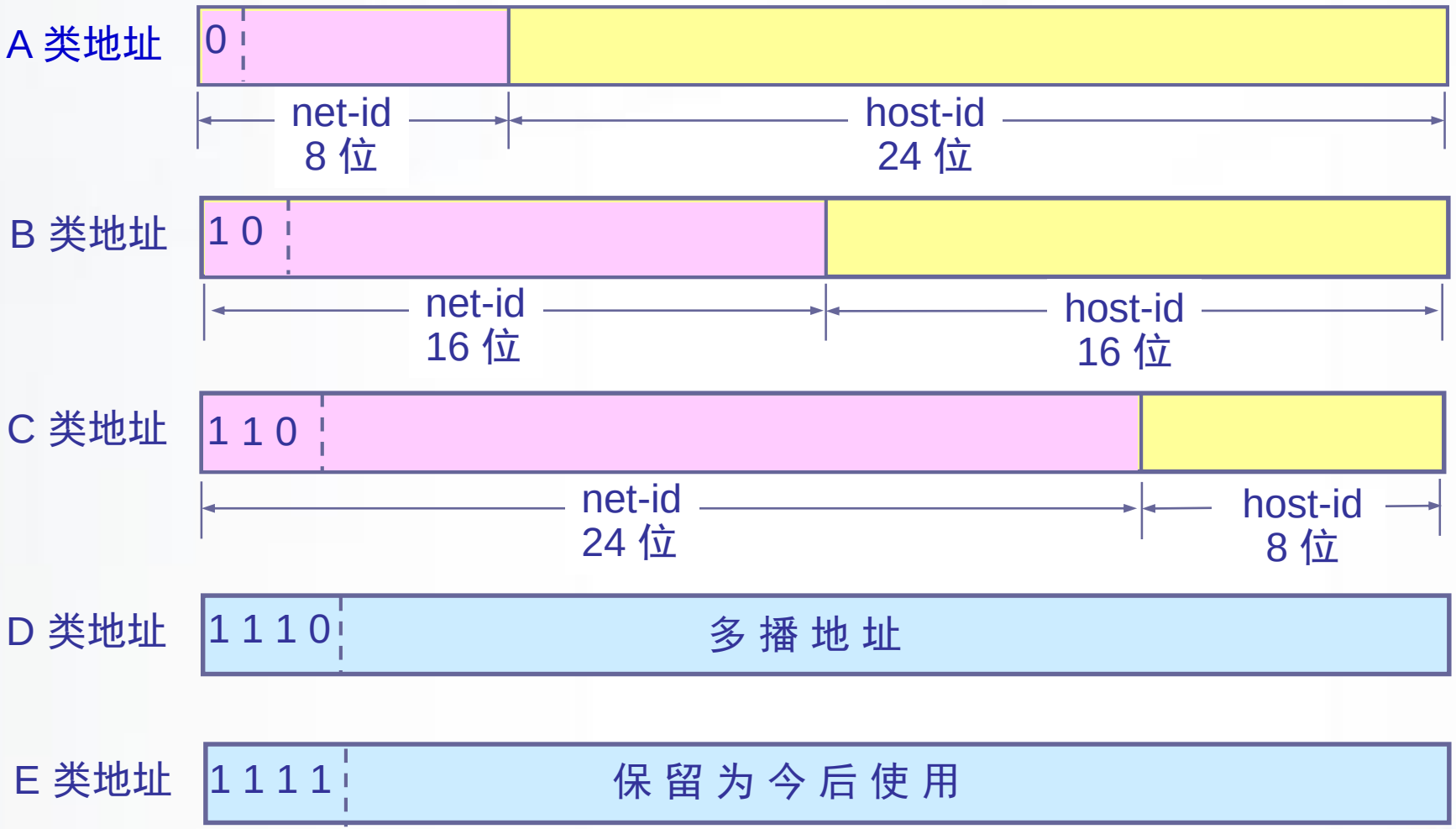
- 我们把整个因特网看成为一个单一的、抽象的网络。IP 地址就是给每个连接在因特网上的主机（或路由器）的每一个接口分配一个在全世界范围是唯一的 32 位的标识符。
- IP 地址现在由因特网名字与号码指派公司ICANN (Internet Corporation for Assigned Names and Numbers)进行分配。
- IP地址的编址方法
  - 分类的 IP 地址。这是最基本的编址方法，在 1981 年就通过了相应的标准协议。
  - 子网的划分。这是对最基本的编址方法的改进，其标准[RFC 950]在 1985 年通过。
  - 构成超网。这是比较新的无分类编址方法。1993 年提出后很快就得到推广应用。

分类 IP 地址

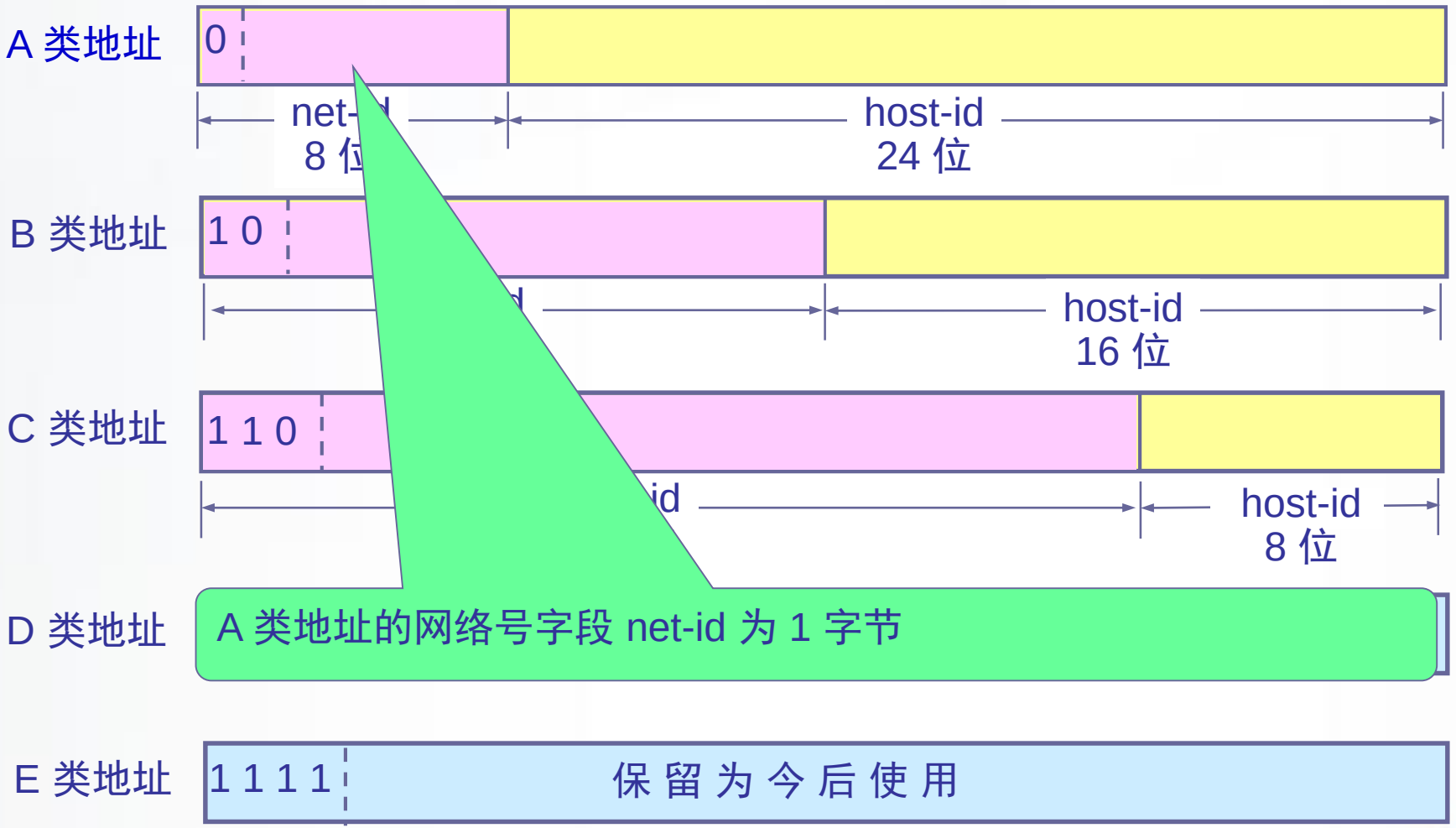
- 每一类地址都由两个固定长度的字段组成，其中一个字段是网络号 net-id，它标志主机（或路由器）所连接到的网络，而另一个字段则是主机号 host-id，它标志该主机（或路由器）。
- 两级的 IP 地址可以记为：

IP 地址 ::= { <网络号>, <主机号>}

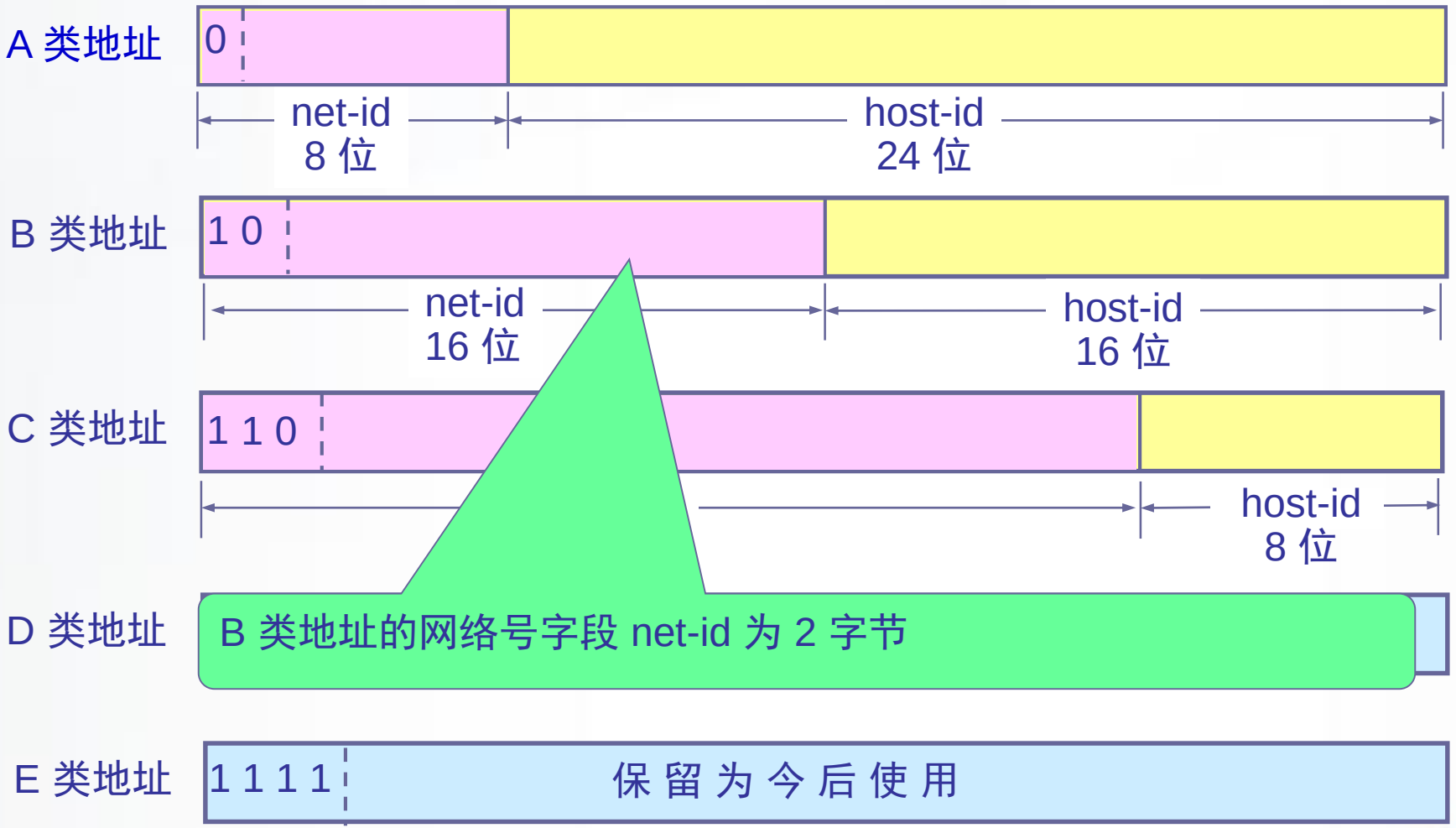
IP 地址中的网络号字段和主机号字段



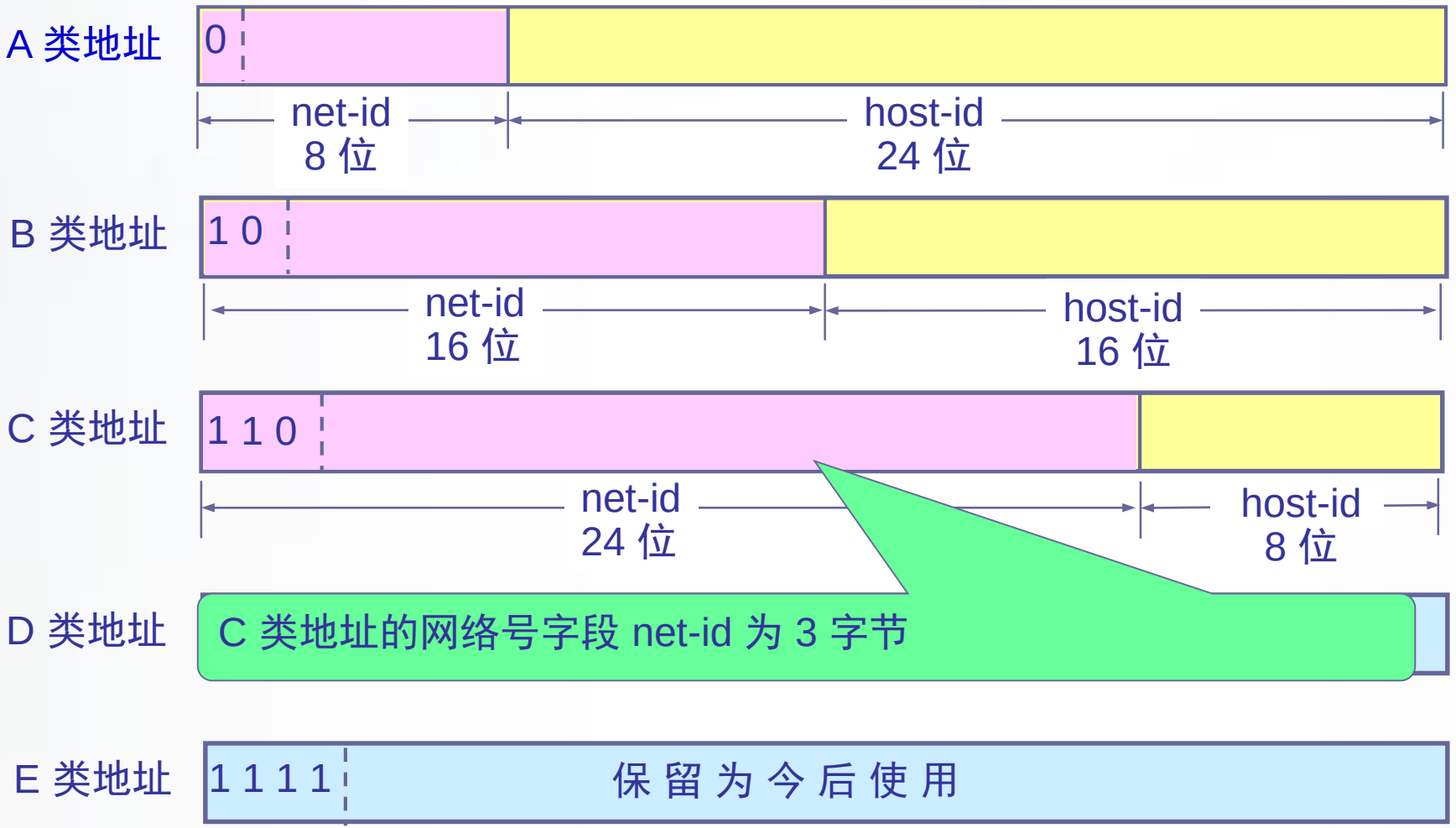
IP 地址中的网络号字段和主机号字段



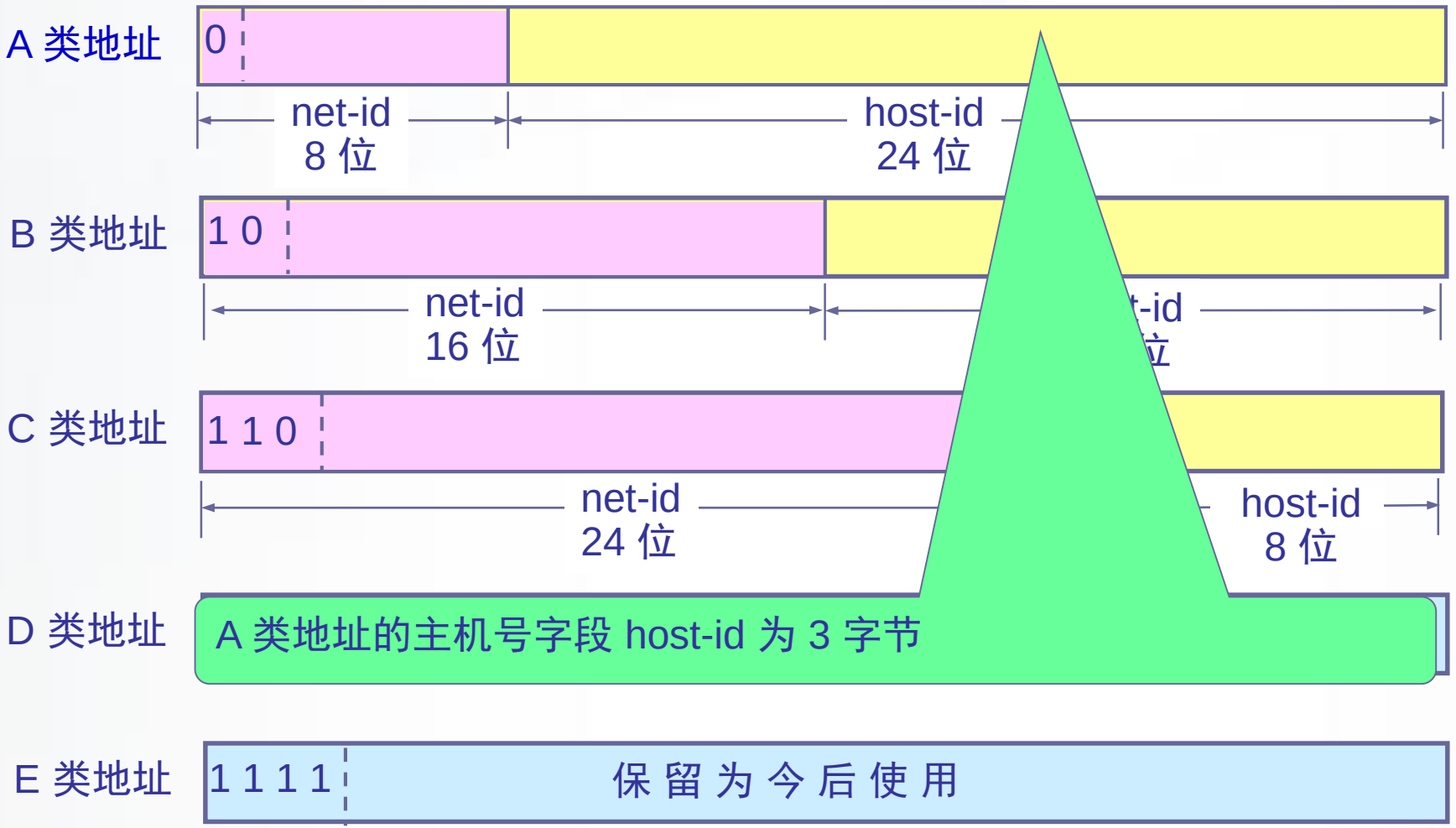
IP 地址中的网络号字段和主机号字段



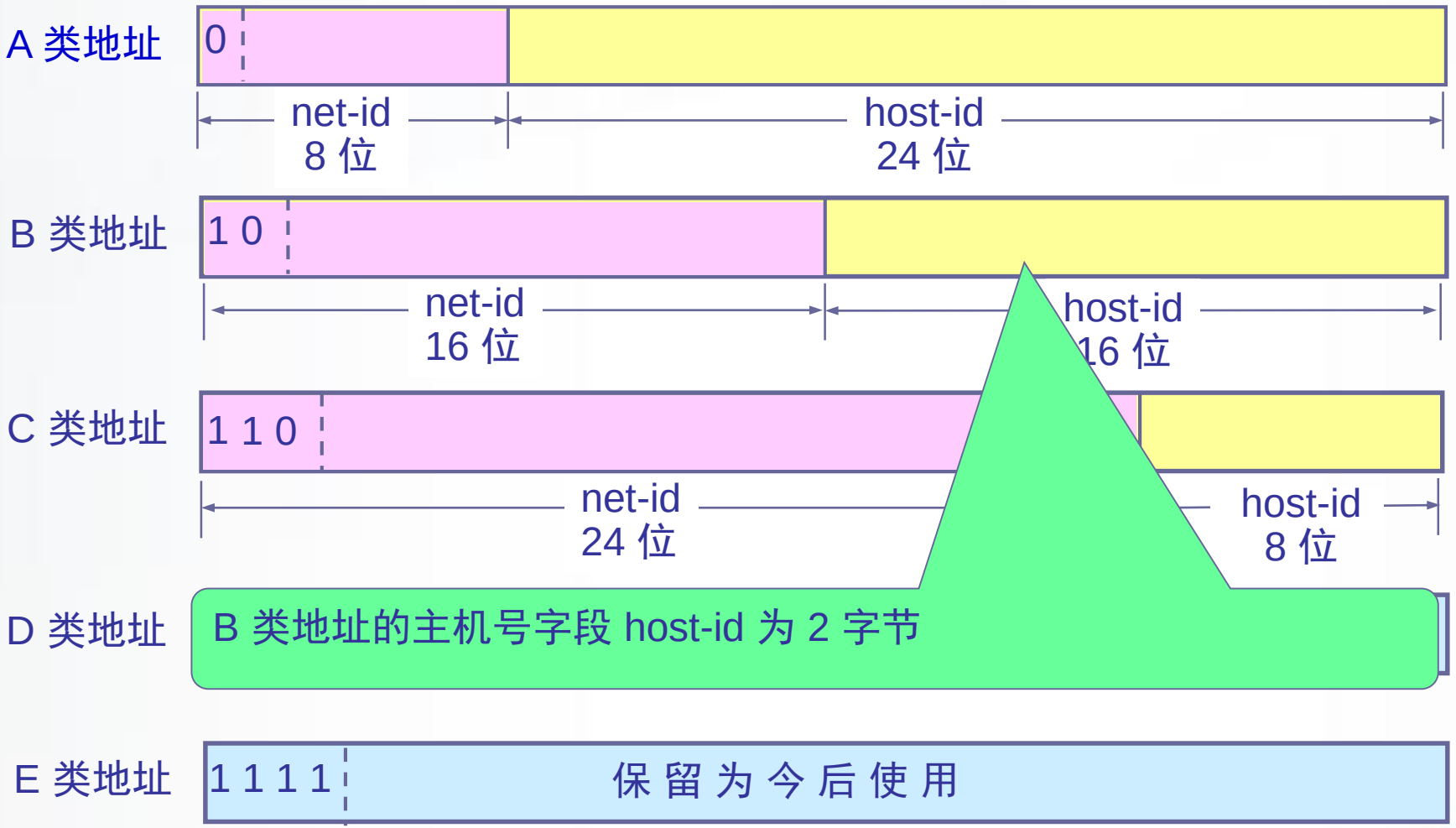
IP 地址中的网络号字段和主机号字段



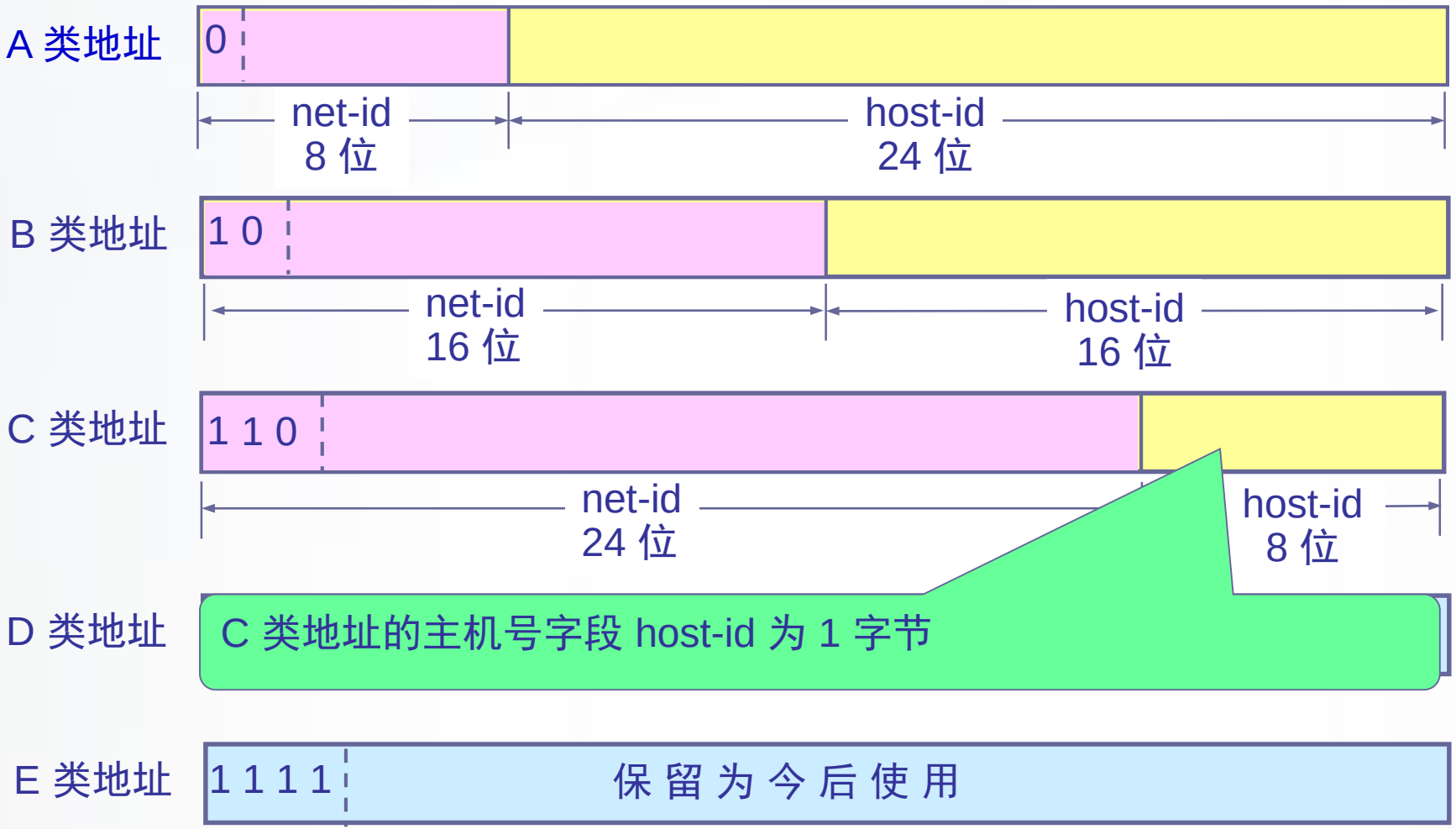
IP 地址中的网络号字段和主机号字段



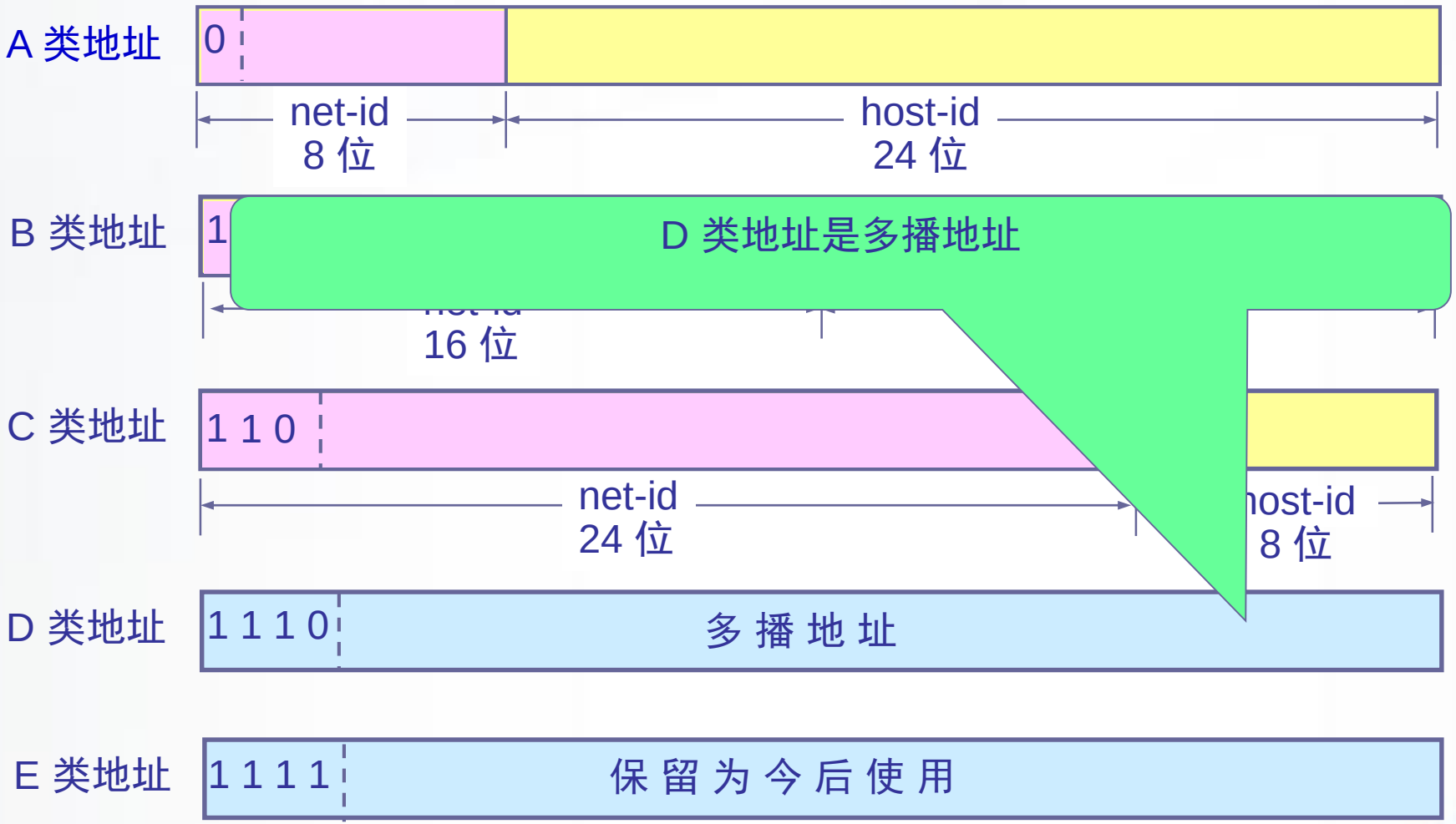
IP 地址中的网络号字段和主机号字段



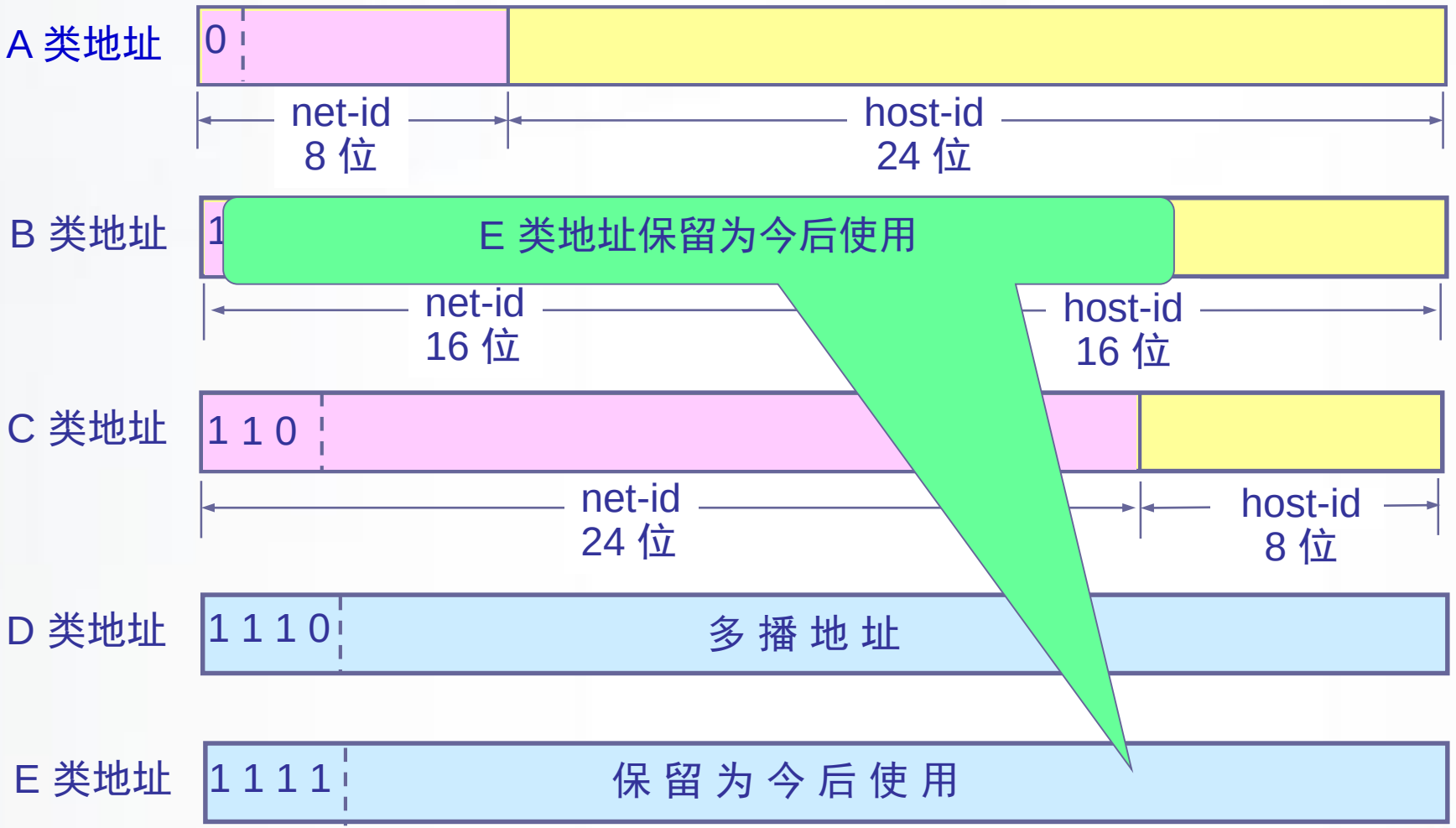
IP 地址中的网络号字段和主机号字段



IP 地址中的网络号字段和主机号字段



IP 地址中的网络号字段和主机号字段



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

每隔 8 位插入一个空格
能够提高可读性

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

常用的三种类别的 IP 地址

网络类别	最大网络数	第一个可用的网络号	最后一个可用的网络号	每个网络中最大的主机数
A	126 ($2^7 - 2$)	1	126	16,777,214
B	16,383($2^{14} - 1$)	128.1	191.255	65,534
C	2,097,151 ($2^{21} - 1$)	192.0.1	223.255.255	254

IP 地址的一些重要特点

- 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。

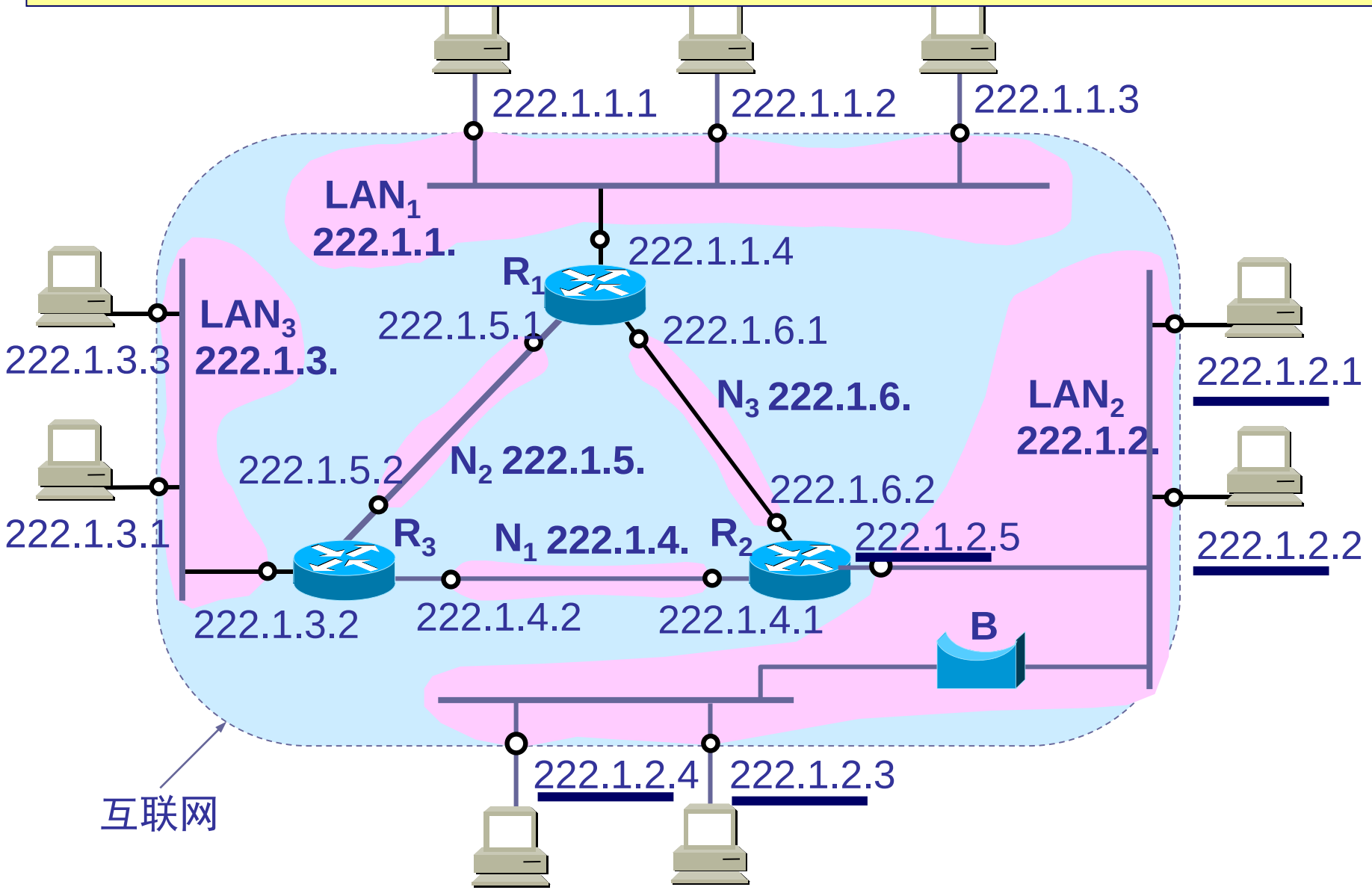
IP 地址的一些重要特点

- 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。

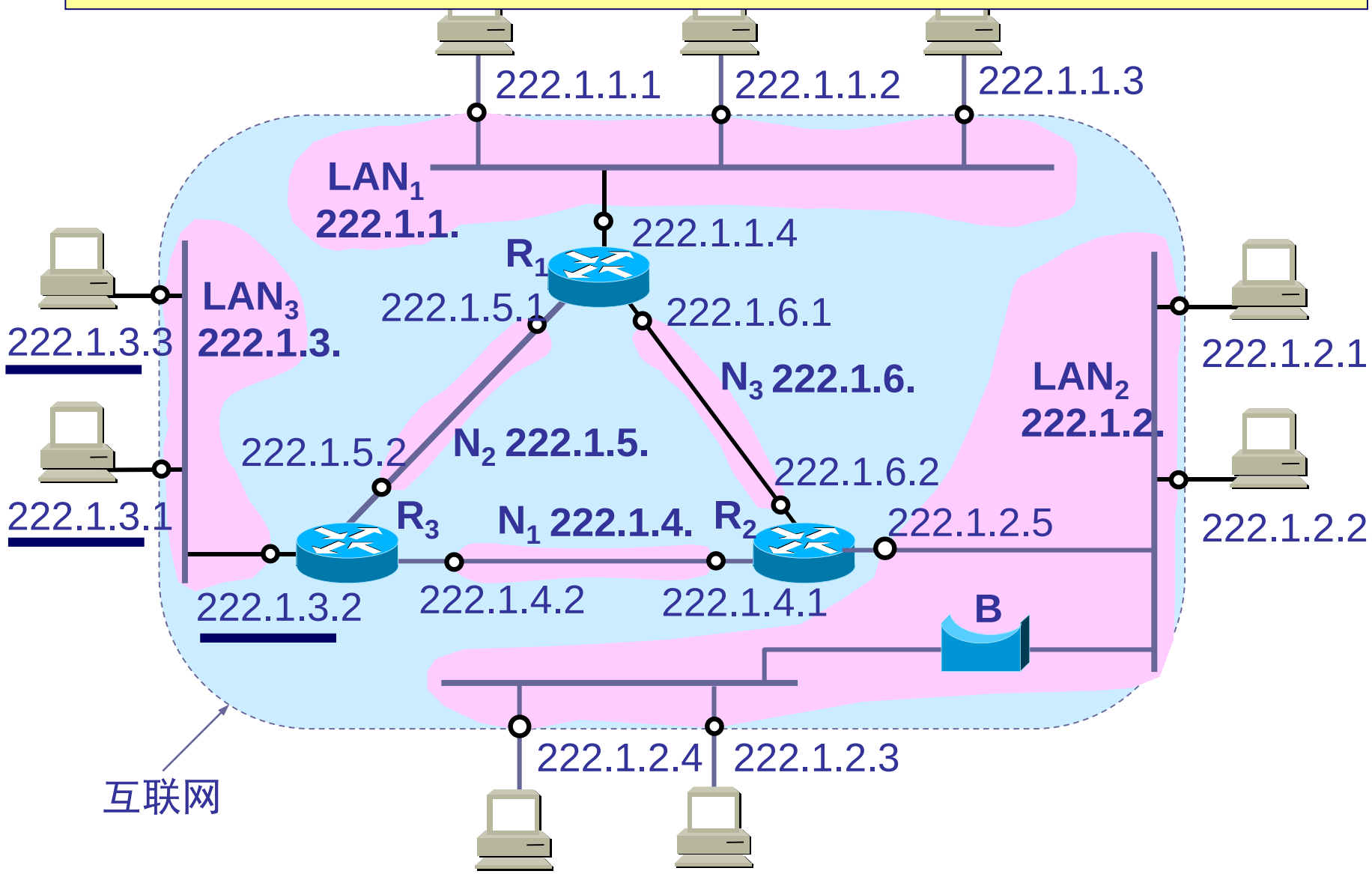
IP 地址的一些重要特点

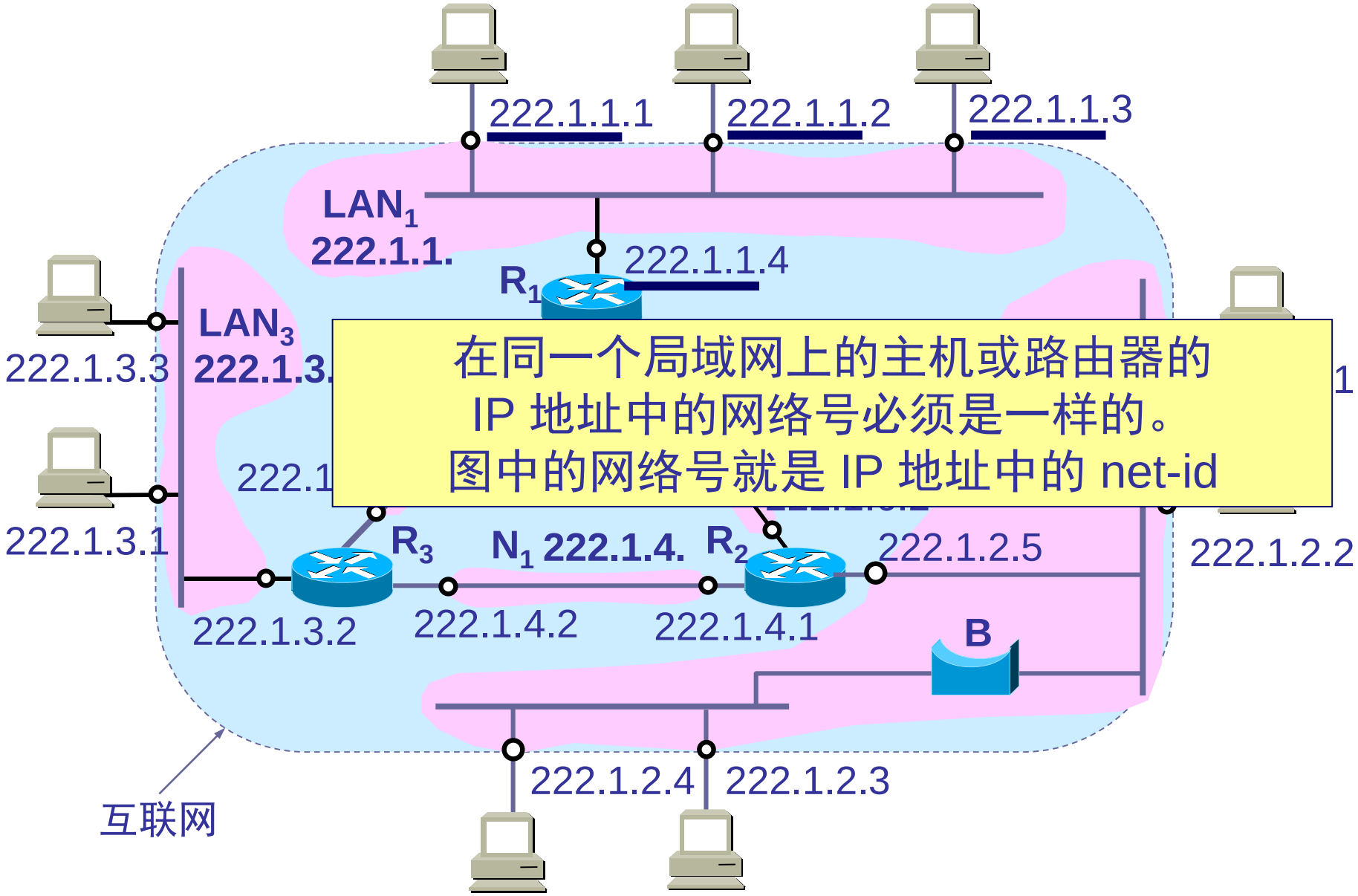
- 用转发器或网桥连接起来的若干个局域网仍为一个网络，因此这些局域网都具有同样的网络号 net-id。
- 所有分配到网络号 net-id 的网络，范围很小的局域网，还是可能覆盖很大地理范围的广域网，都是平等的。

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

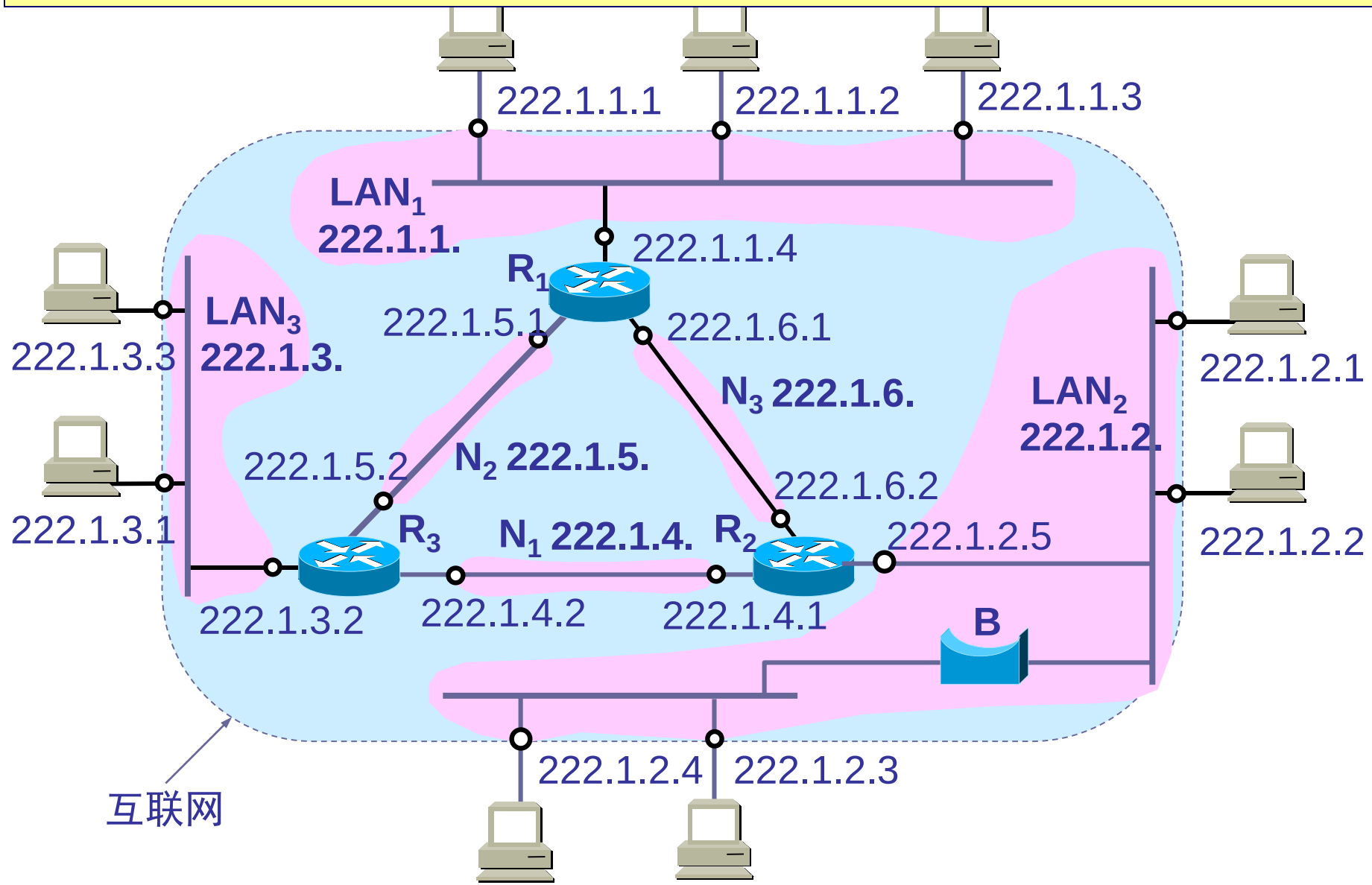


在同一个局域网上的主机或路由器的IP 地址中的网络号必须是一样的。图中的网络号就是 IP 地址中的 net-id

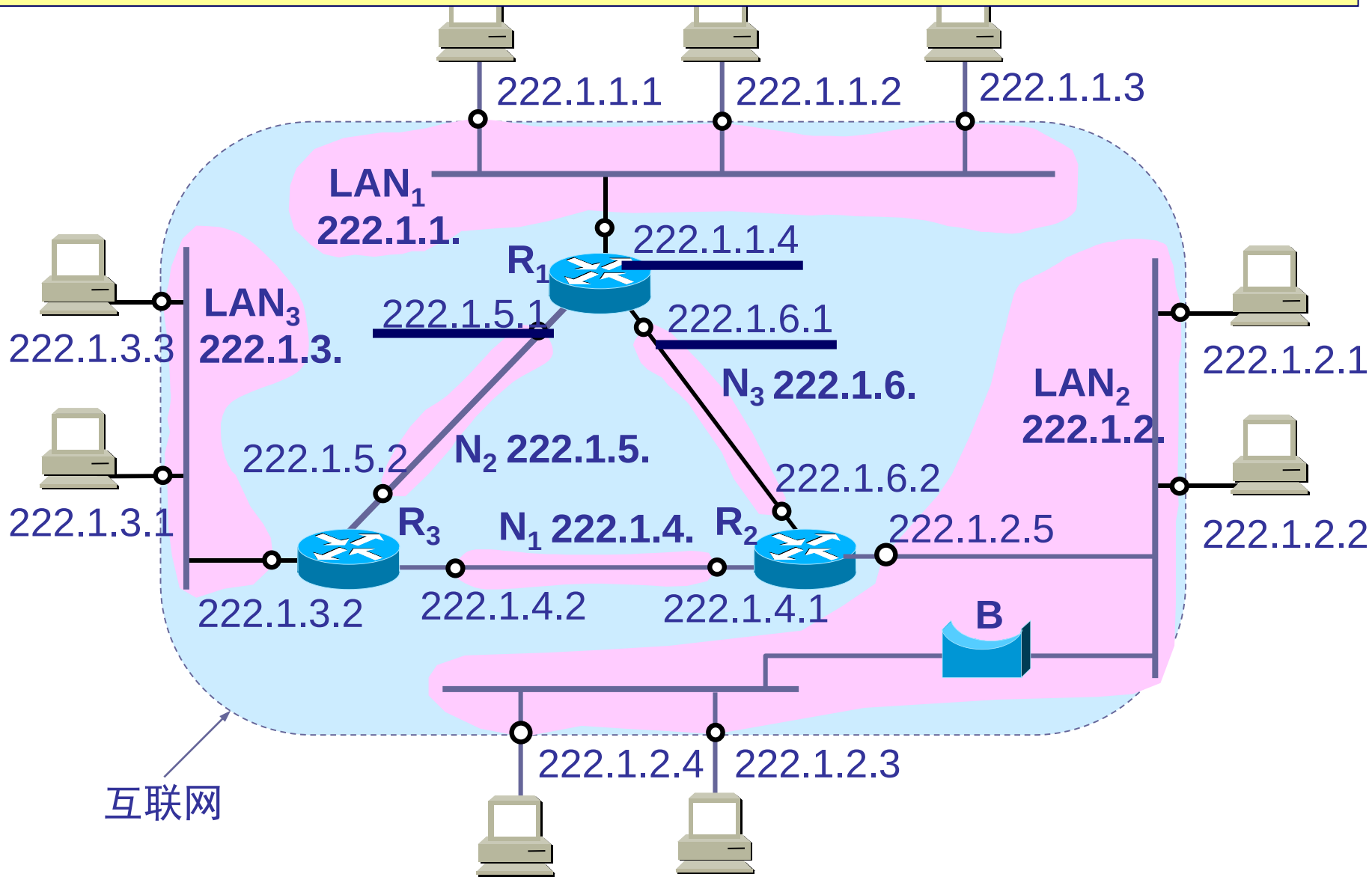



222.1.1.1
222.1.1.2
222.1.1.3
LAN1
222.1.1.
222.1.1.4
R1
LAN3
222.1.3.
222.1.3.3
222.1.3.1
222.1
在同一个局域网上的主机或路由器的
IP 地址中的网络号必须是一样的。
图中的网络号就是 IP 地址中的 net-id
R3
N1 222.1.4.
R2
222.1.2.5
222.1.2.2
222.1.3.2
222.1.4.2
222.1.4.1
B
222.1.2.4
222.1.2.3
互联网

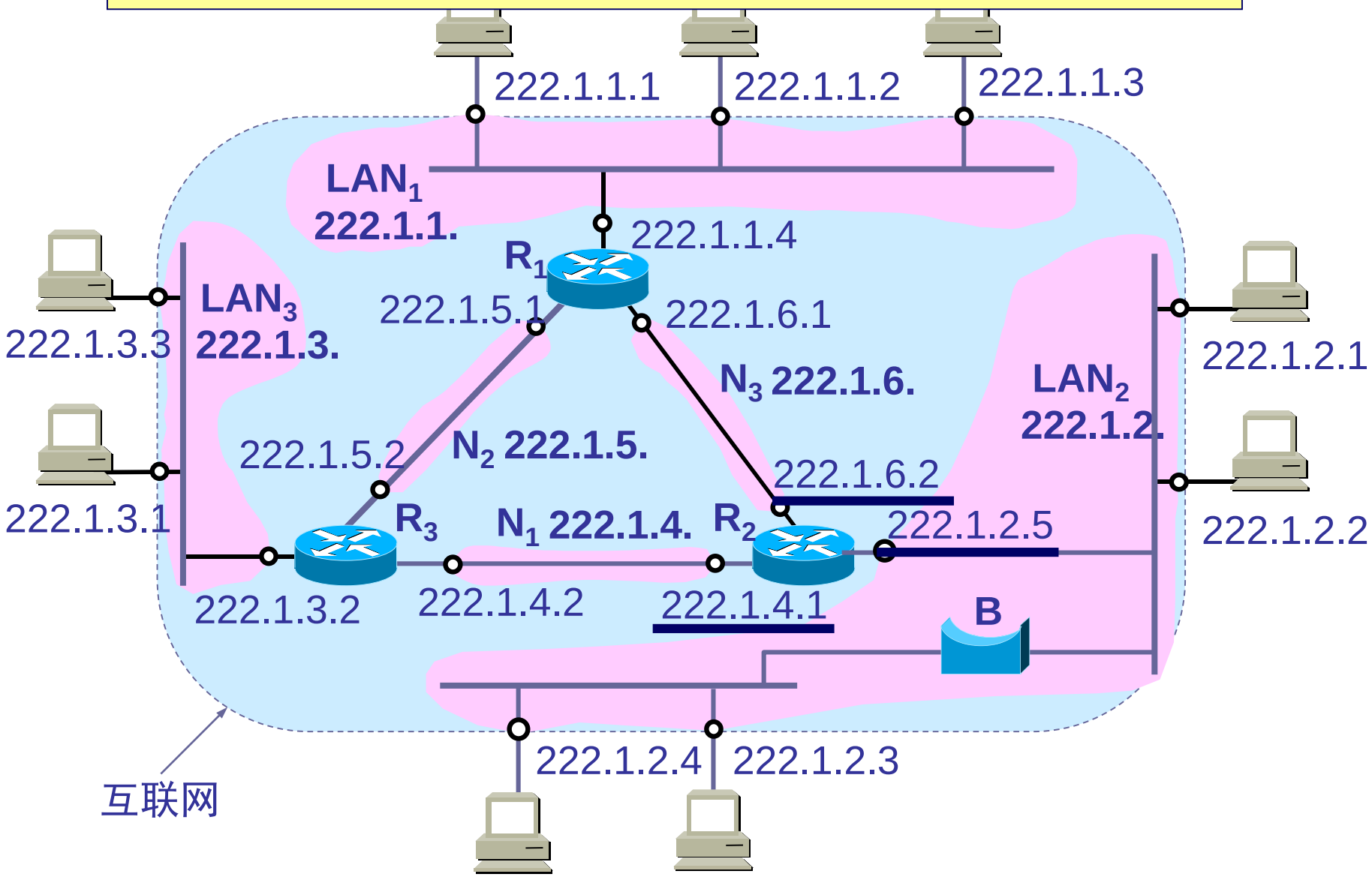
在同一个局域网上的主机或路由器的IP 地址中的网络号必须是一样的。图中的网络号就是 IP 地址中的 net-id



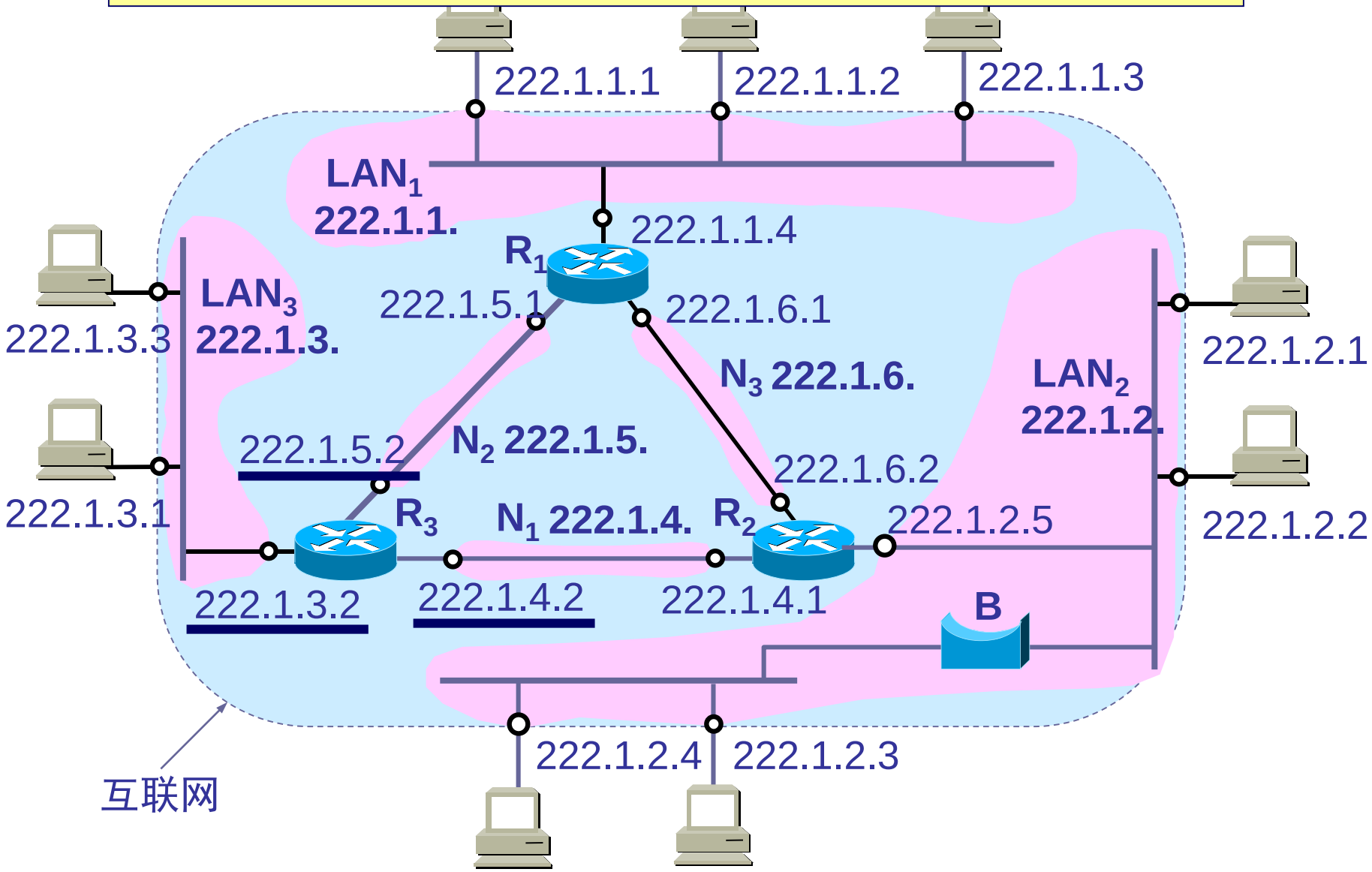
路由器总是具有两个或两个以上的 IP 地址。路由器的每一个接口都有一个不同网络号的 IP 地址。



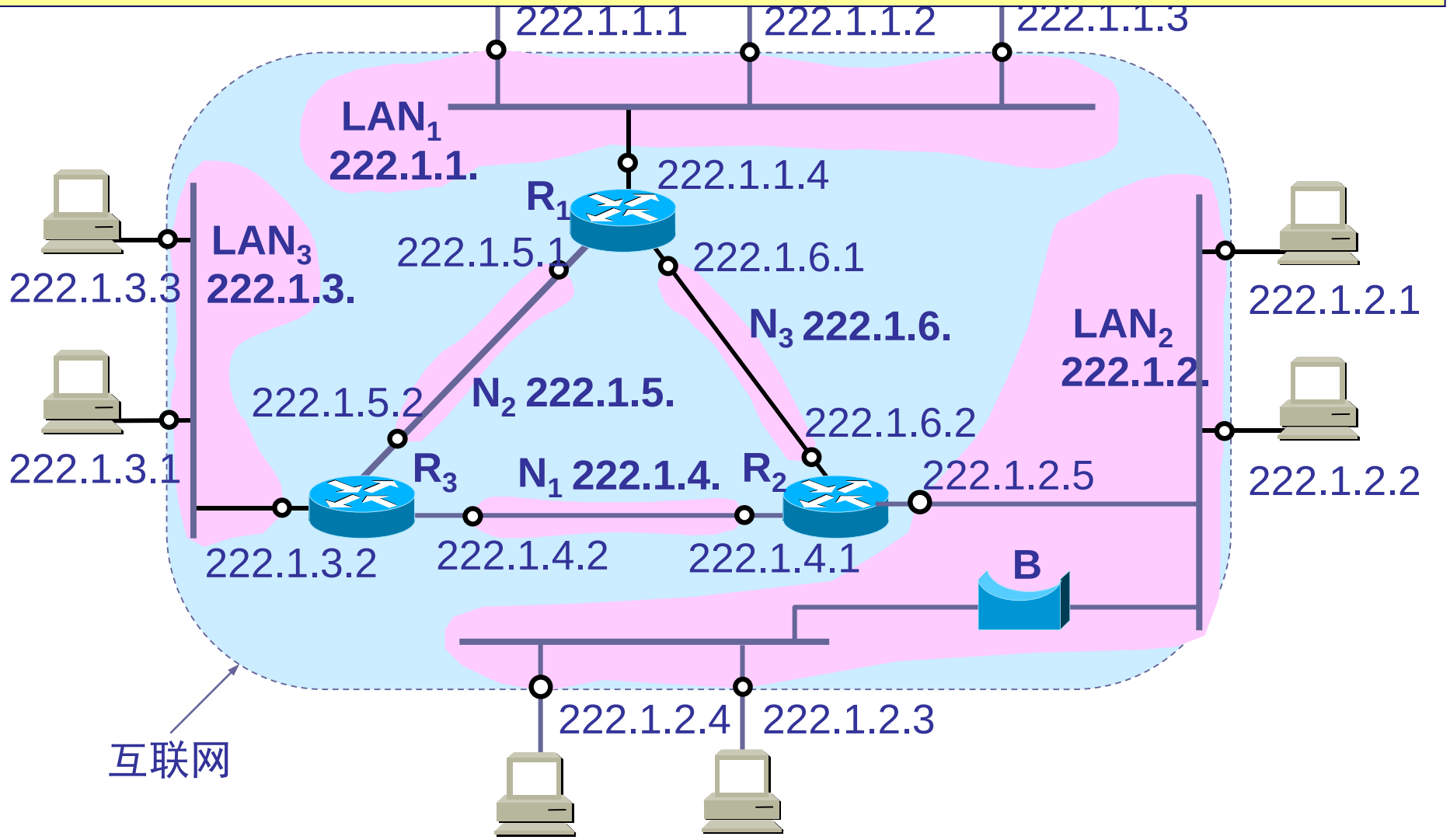
路由器总是具有两个或两个以上的 IP 地址。路由器的每一个接口都有一个不同网络号的 IP 地址。



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。



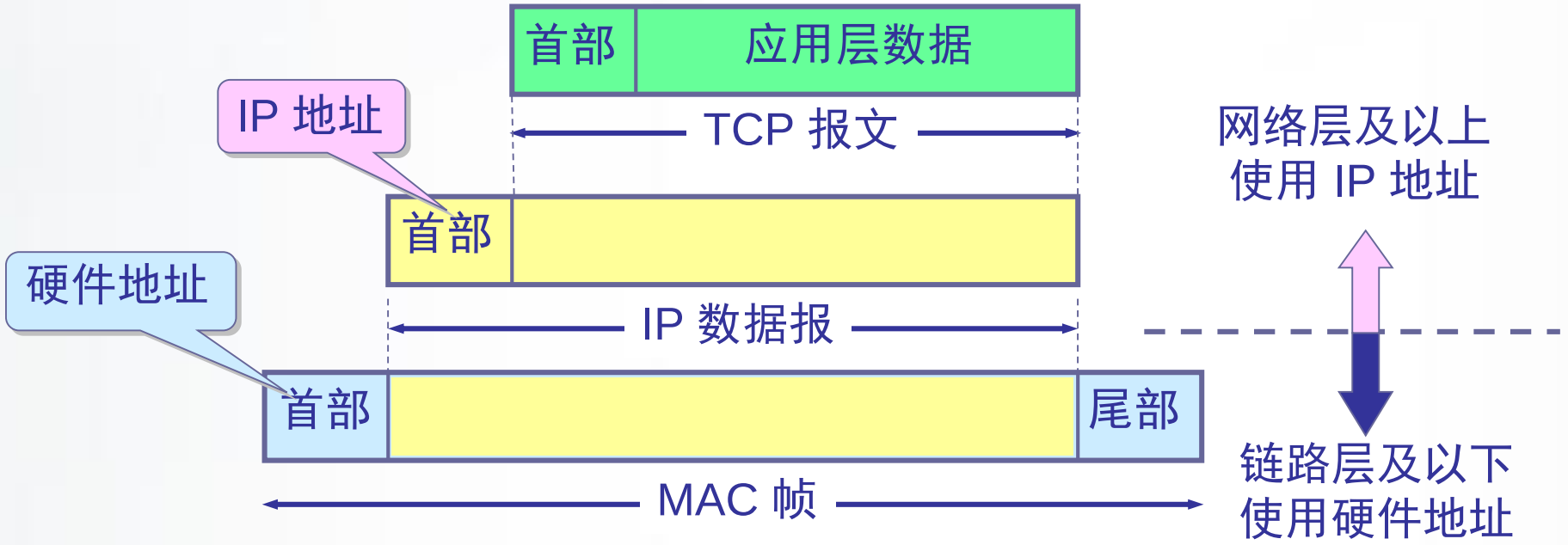
两个路由器直接相连的接口处，可指明也可不指明 IP 地址。如指明 IP 地址，则这一段连线就构成了一种只包含一段线路的特殊“网络”。现在常不指明 IP 地址。



指引

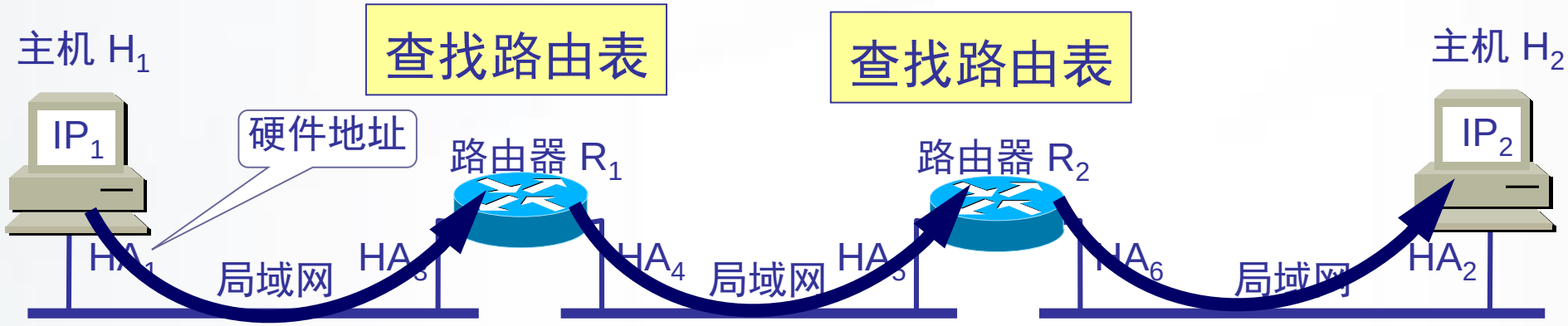
- 网络层提供的两种服务
- 网际协议 IP
  - 虚拟互联网
  - IP地址
  - IP地址与硬件地址
  - IP数据报格式
  - IP转发分组的流程
- 划分子网和构造超网
- 网际控制报文协议 ICMP
- 因特网的路由选择协议
- IPv6
- IP 多播
- 虚拟专用网 VPN 和网络地址转换 NAT

IP 地址与硬件地址


首部
应用层数据
TCP 报文
IP 地址
首部
IP 数据报
硬件地址
首部
尾部
MAC 帧
网络层及以上
使用 IP 地址
链路层及以下
使用硬件地址

为什么不直接使用硬件地址进行通信？

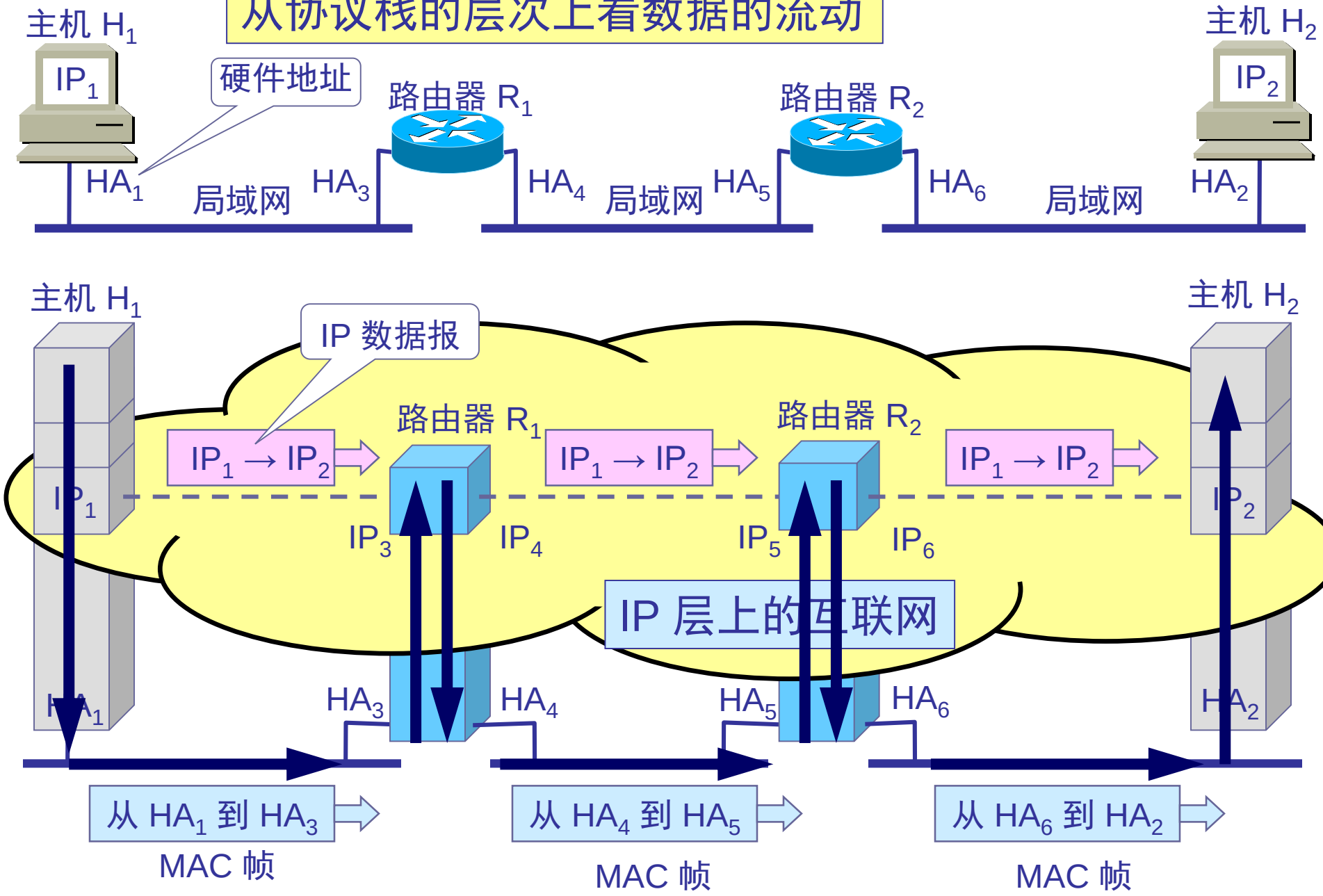
- 由于全世界存在着各式各样的网络，它们使用不同的硬件地址。要使这些异构网络能够互相通信就必须进行非常复杂的硬件地址转换工作，因此几乎是不可能的事。
- 连接到因特网的主机都拥有统一的 IP 地址，它们之间的通信就像连接在同一个网络上那样简单方便，因为调用 ARP 来寻找某个路由器或主机的硬件地址都是由计算机软件自动进行的，对用户来说是看不见这种调用过程的。

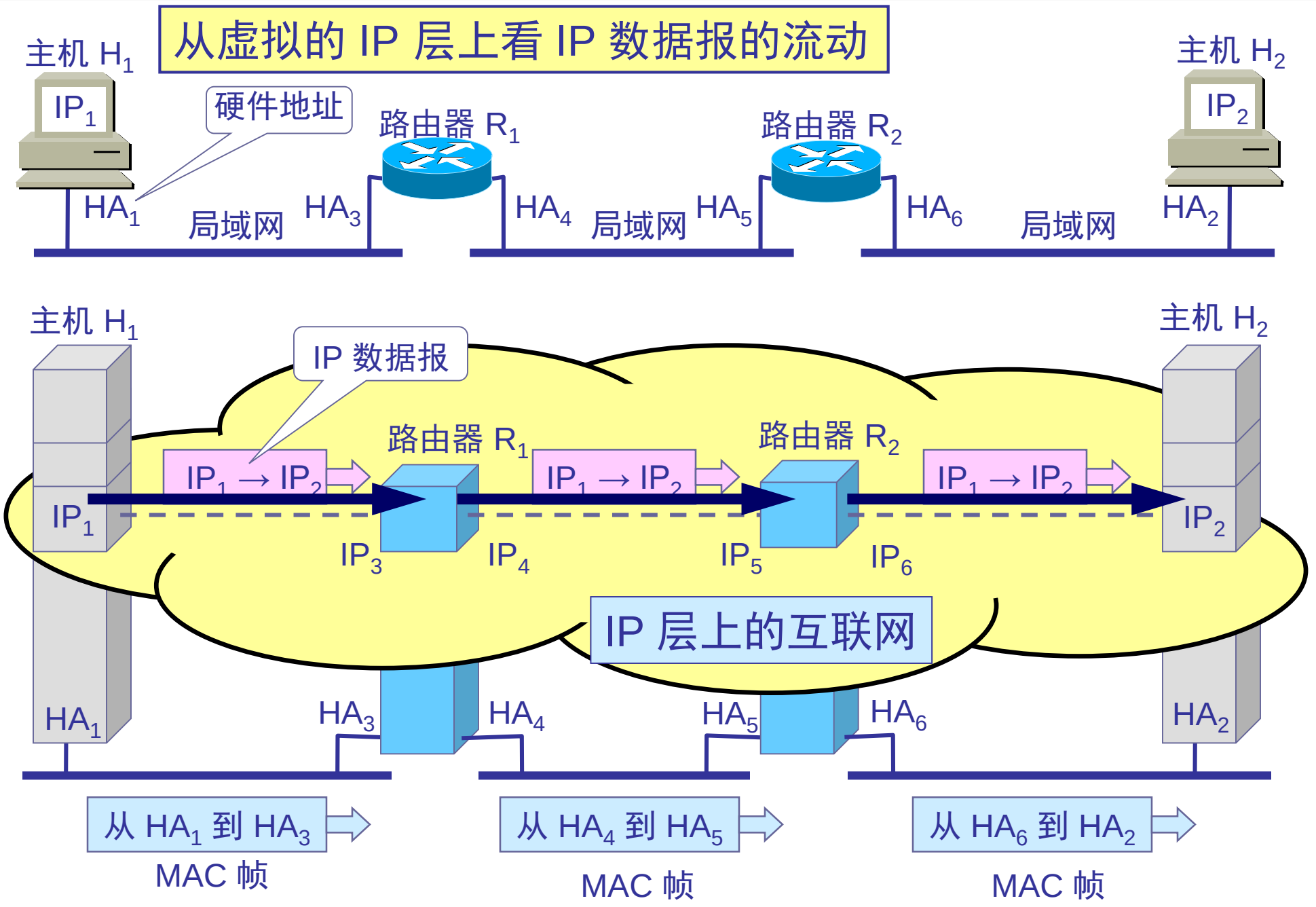


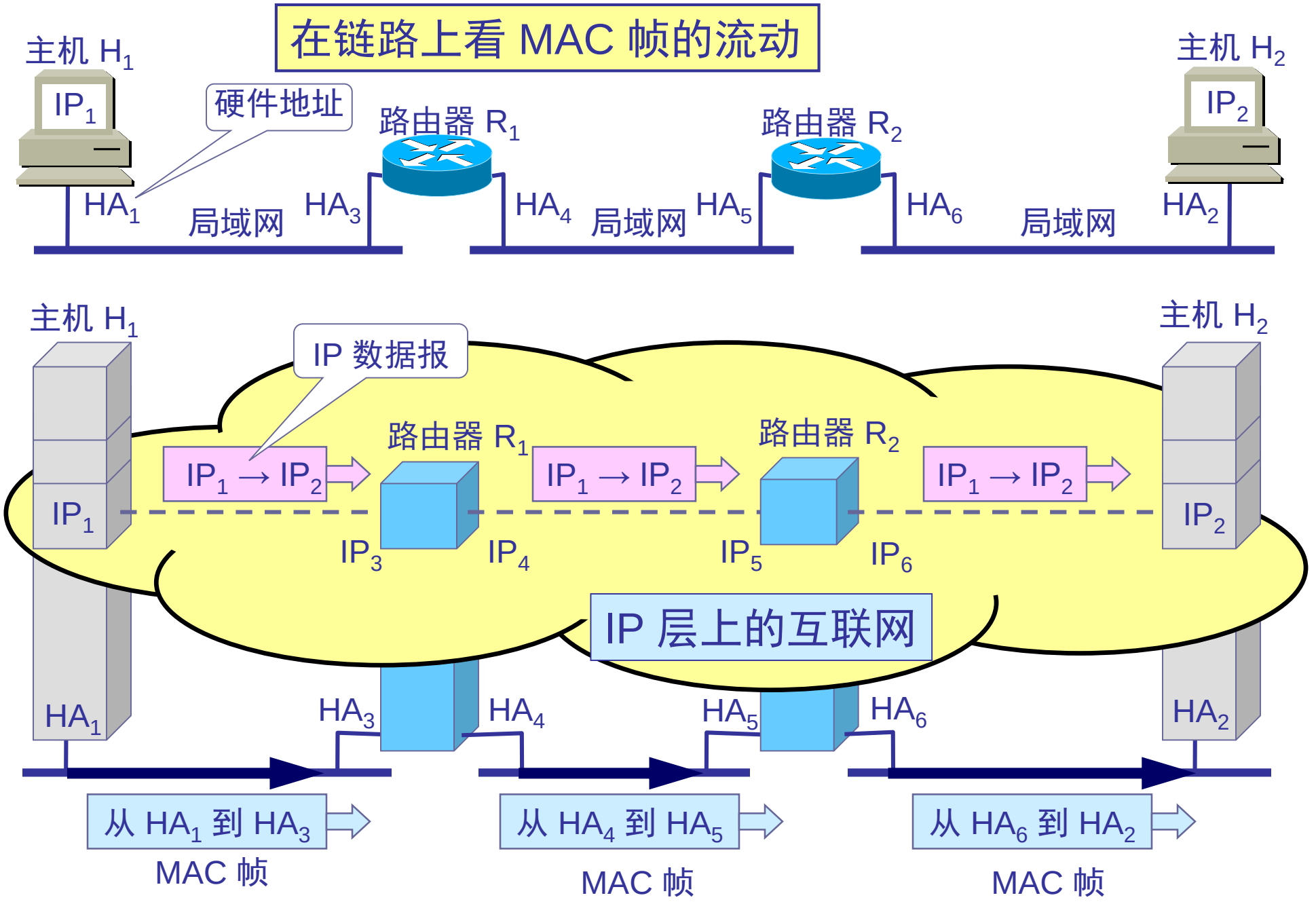
通信的路径

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



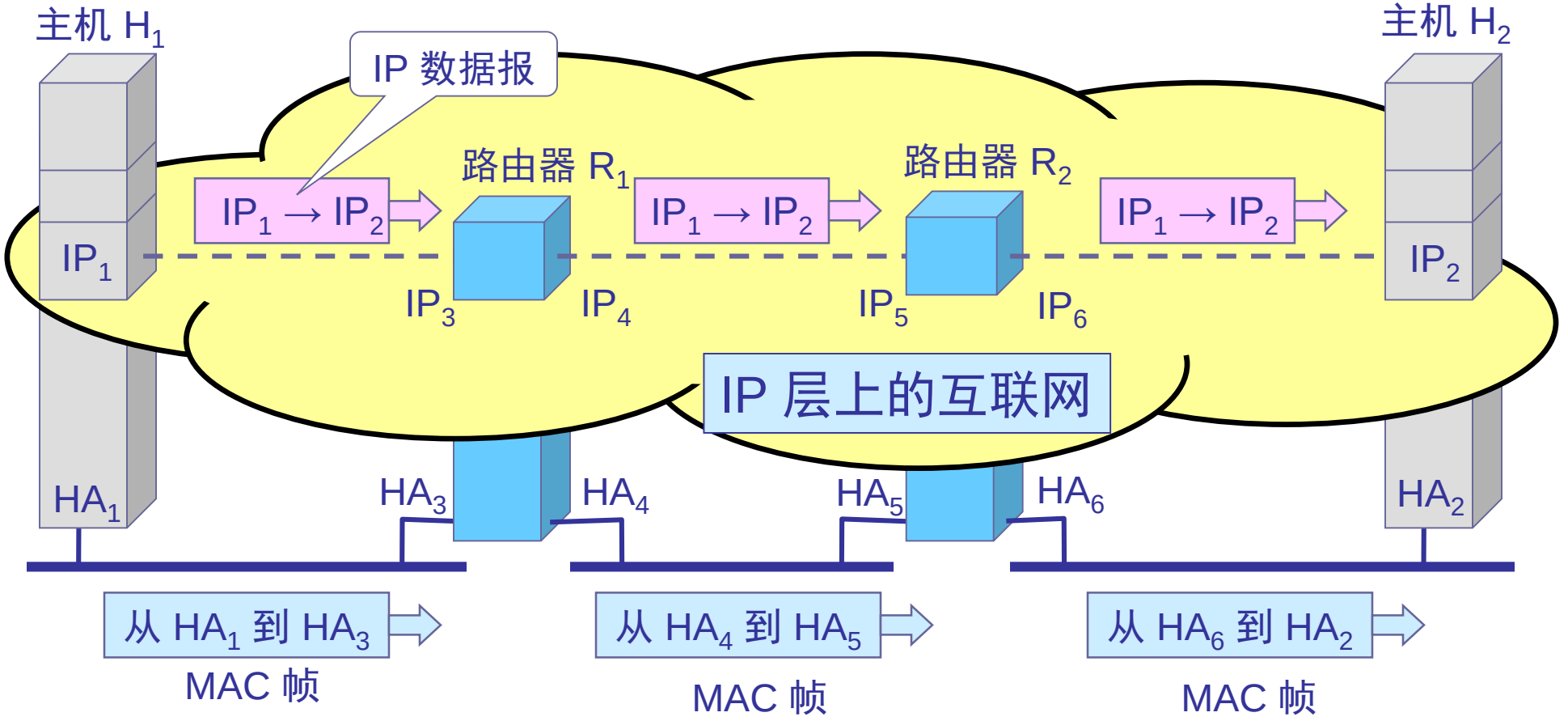

从虚拟的 IP 层上看 IP 数据报的流动
主机 H1
IP1
硬件地址
路由器 R1
路由器 R2
主机 H2
IP2
HA1
局域网
HA3
HA4
局域网
HA5
HA6
局域网
HA2
主机 H1
IP 数据报
主机 H2
IP1 → IP2
IP1 → IP2
IP1 → IP2
路由器 R1
路由器 R2
IP1
IP3
IP4
IP5
IP6
IP2
IP 层上的互联网
HA1
HA3
HA4
HA5
HA6
HA2
从 HA1 到 HA3
从 HA4 到 HA5
从 HA6 到 HA2
MAC 帧
MAC 帧
MAC 帧


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

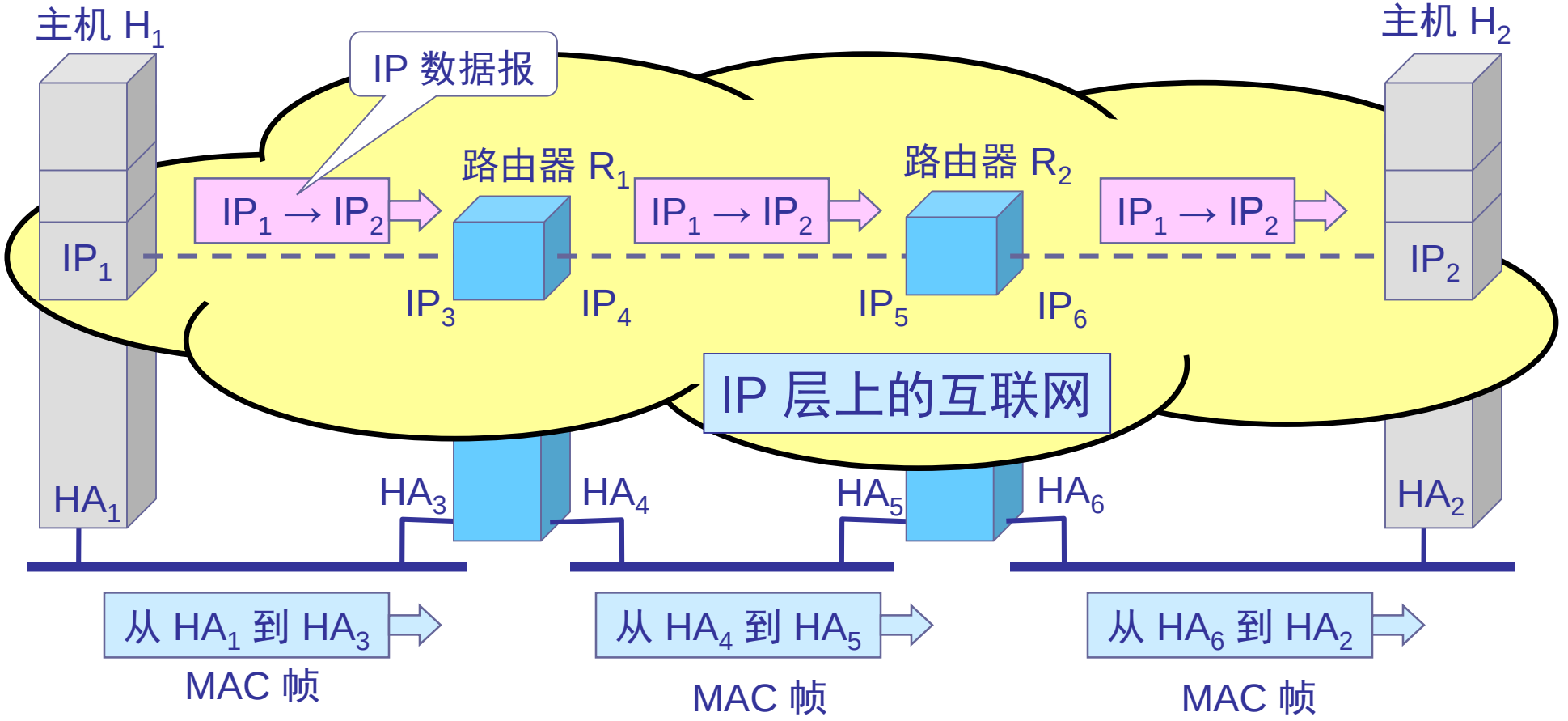
在 IP 层抽象的互联网上只能看到 IP 数据报

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

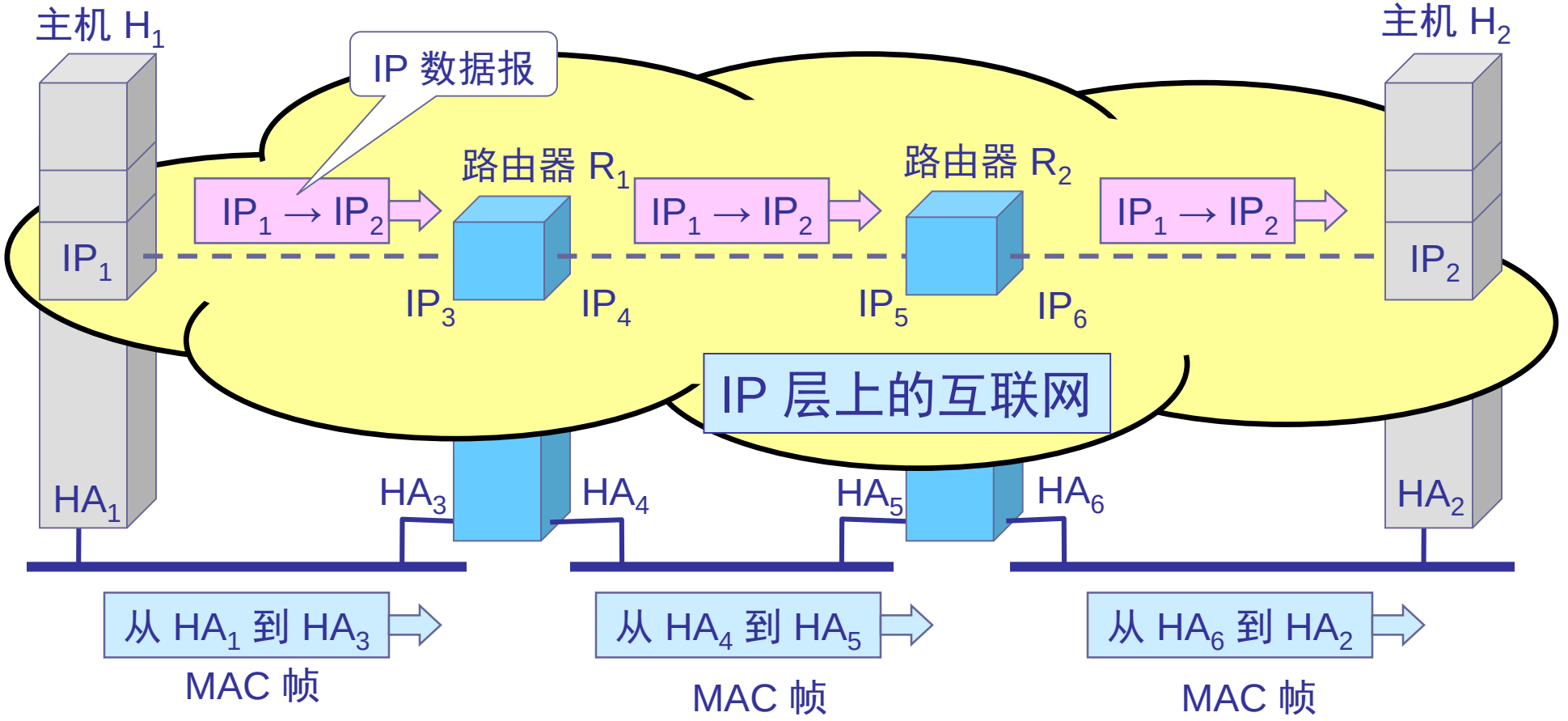
两个路由器的 IP 地址并不出现在 IP 数据报的首部中



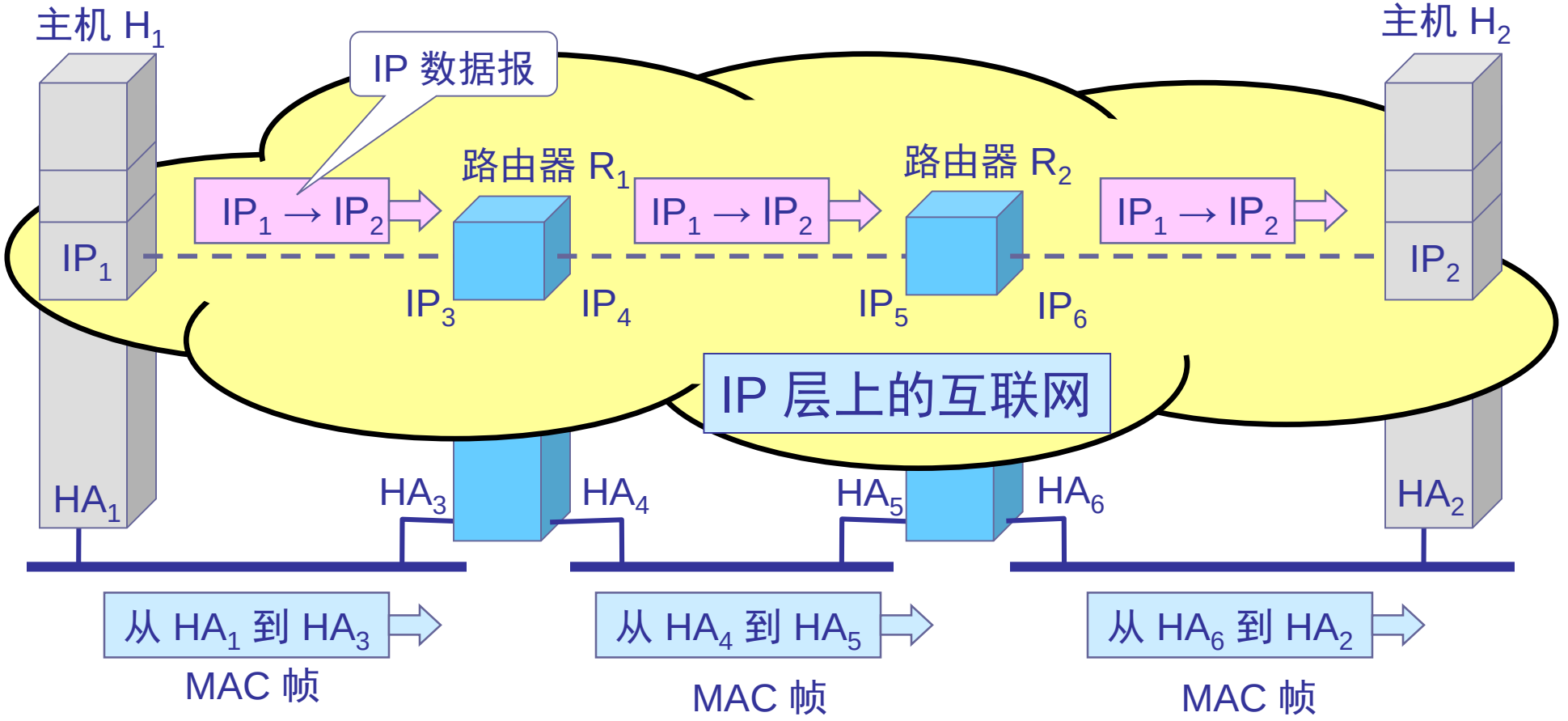
路由器只根据目的站的 IP 地址的网络号进行路由选择



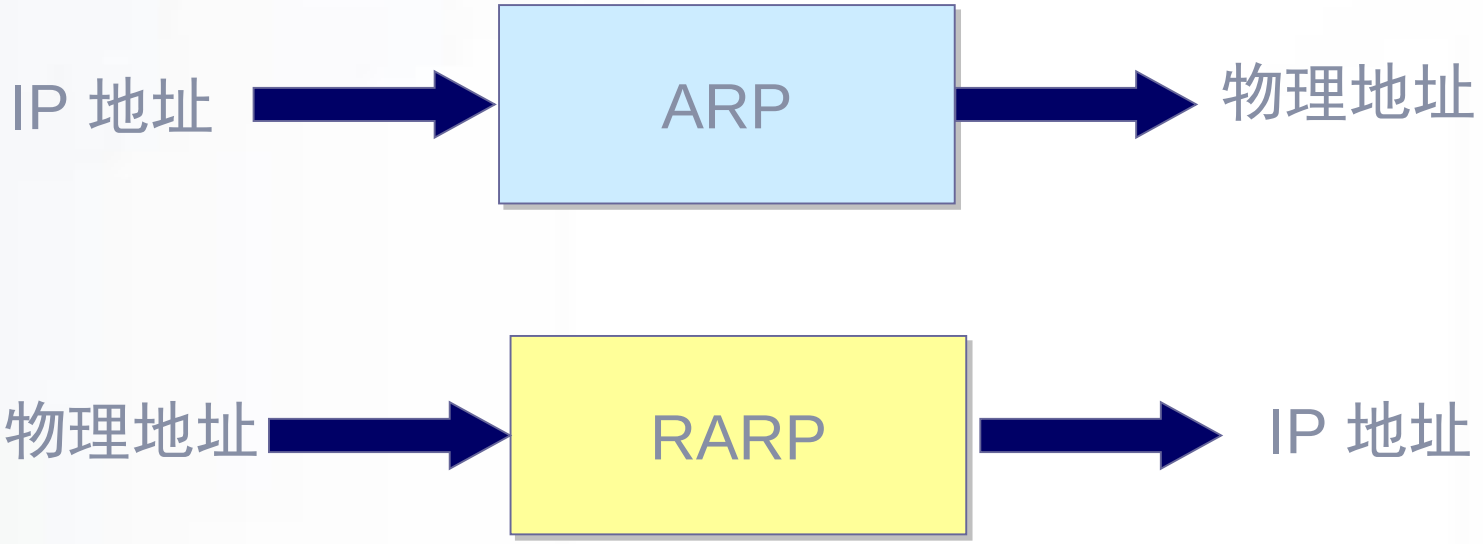
在具体的物理网络的链路层
只能看见 MAC 帧而看不见 IP 数据报



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



ARP & RARP

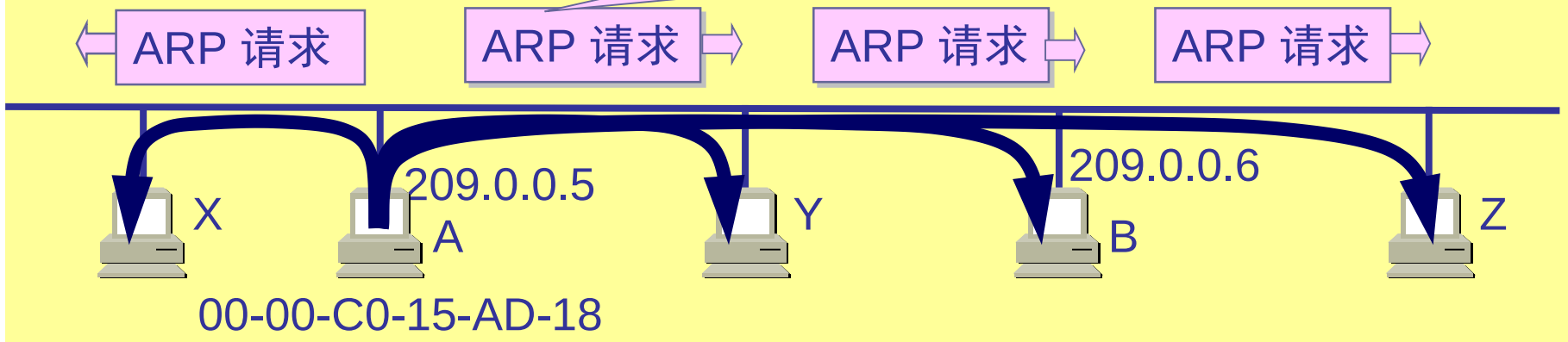


ARP简介

- 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。
- 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。

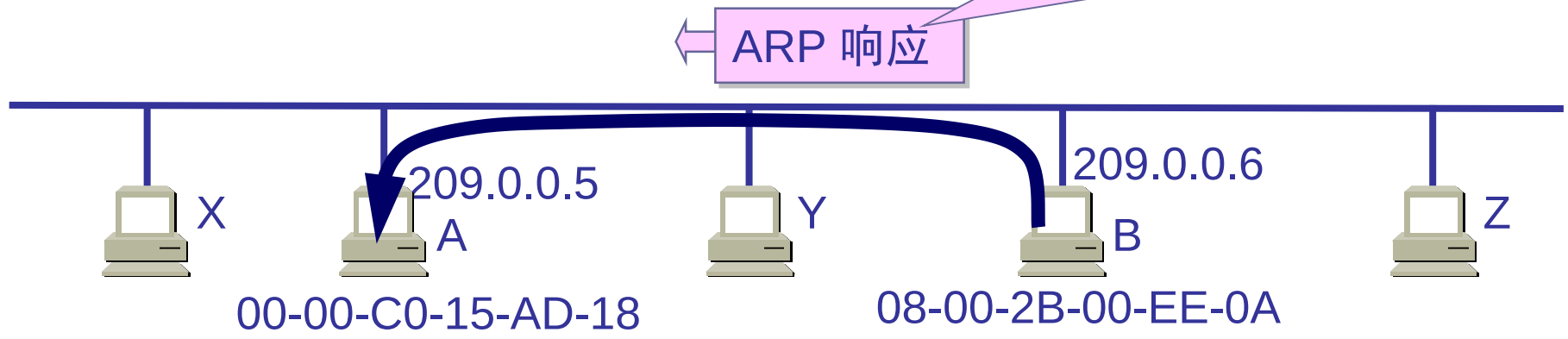
主机 A 广播发送
ARP 请求分组

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

- 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

- 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。

逆地址解析协议 RARP

- 逆地址解析协议 RARP 使只知道自己硬件地址的主机能够知道其 IP 地址。

动手实验

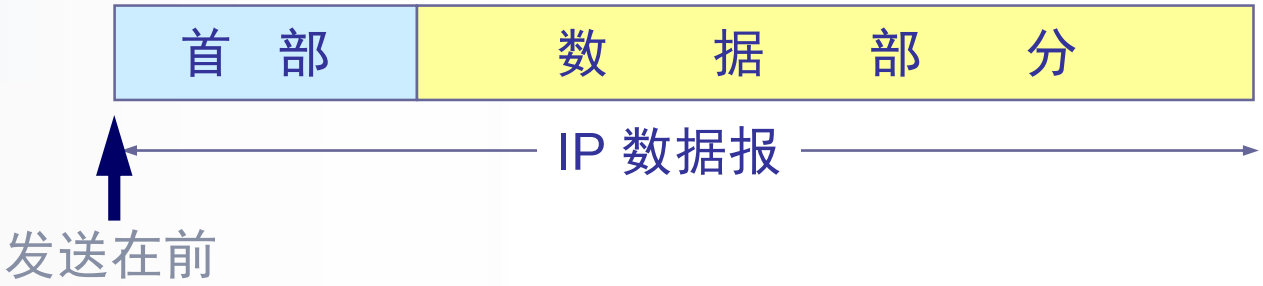
- 实验4-1：ARP命令

指引

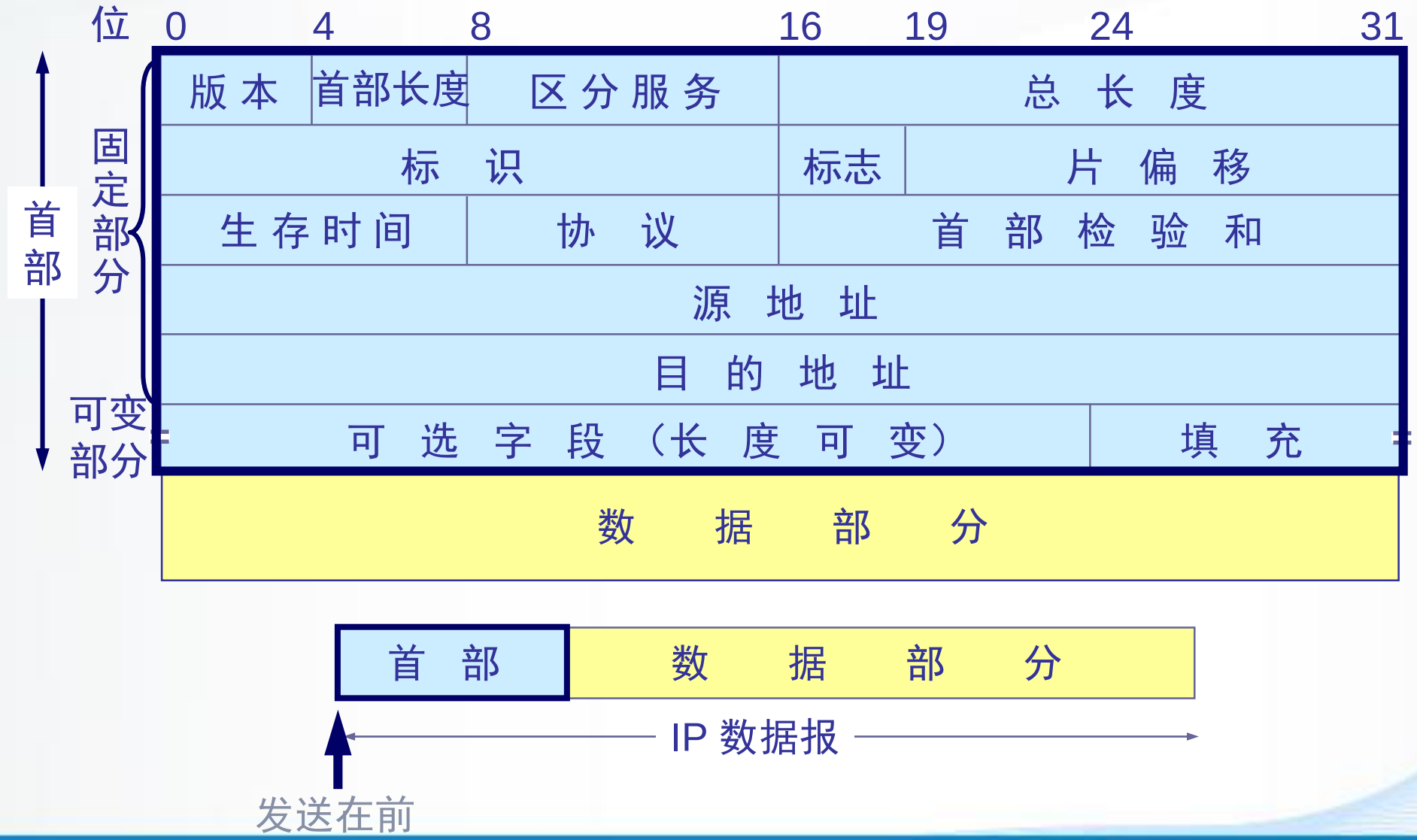
- 网络层提供的两种服务
- 网际协议 IP
  - 虚拟互联网
  - IP地址
  - IP地址与硬件地址
  - IP数据报格式
  - IP转发分组的流程
- 划分子网和构造超网
- 网际控制报文协议 ICMP
- 因特网的路由选择协议
- IPv6
- IP 多播
- 虚拟专用网 VPN 和网络地址转换 NAT

IP数据报

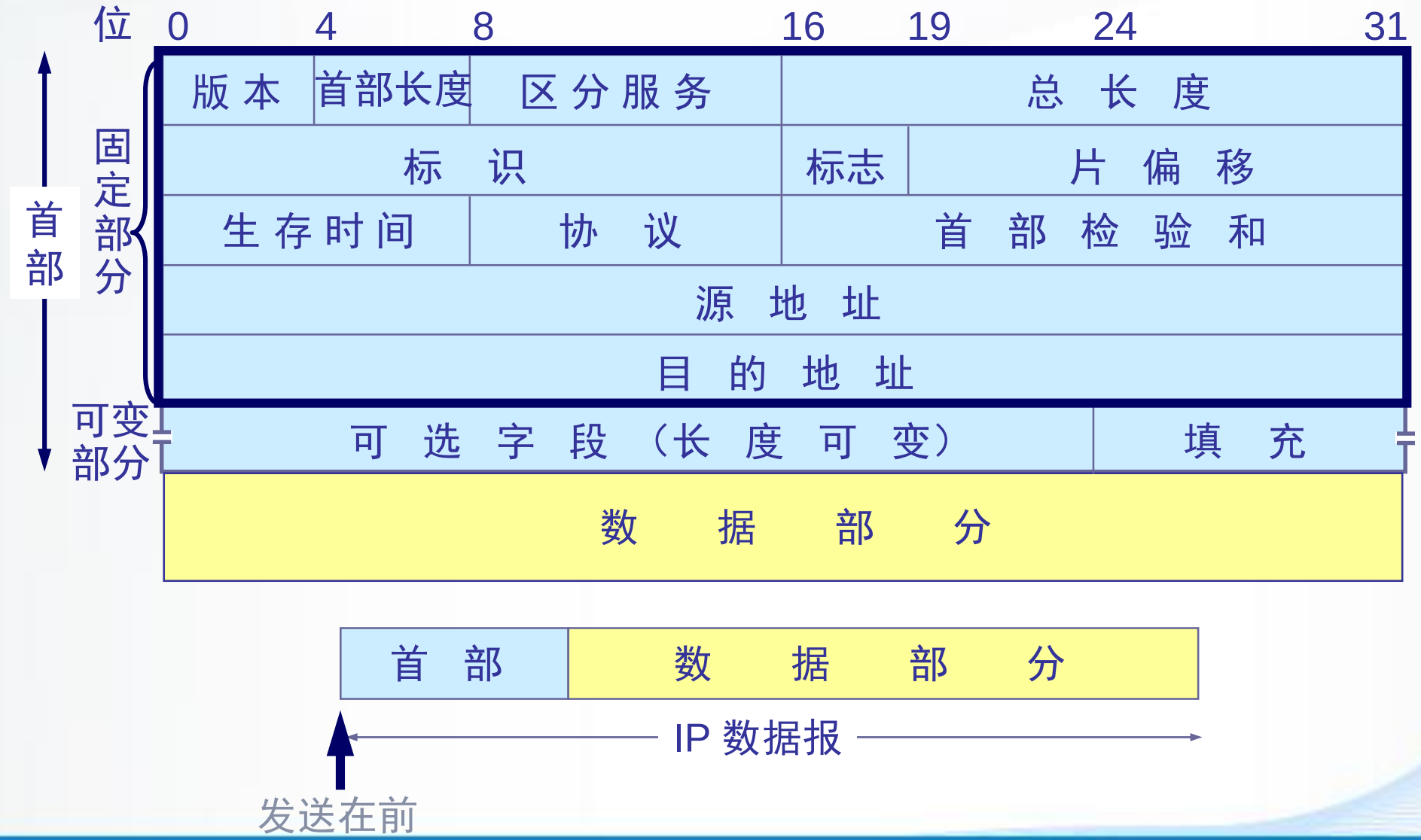
- 一个 IP 数据报由首部和数据两部分组成。
  - 首部的前一部分是固定长度，共 20 字节，是所有 IP 数据报必须具有的。
  - 在首部的固定部分的后面是一些可选字段，其长度是可变的。



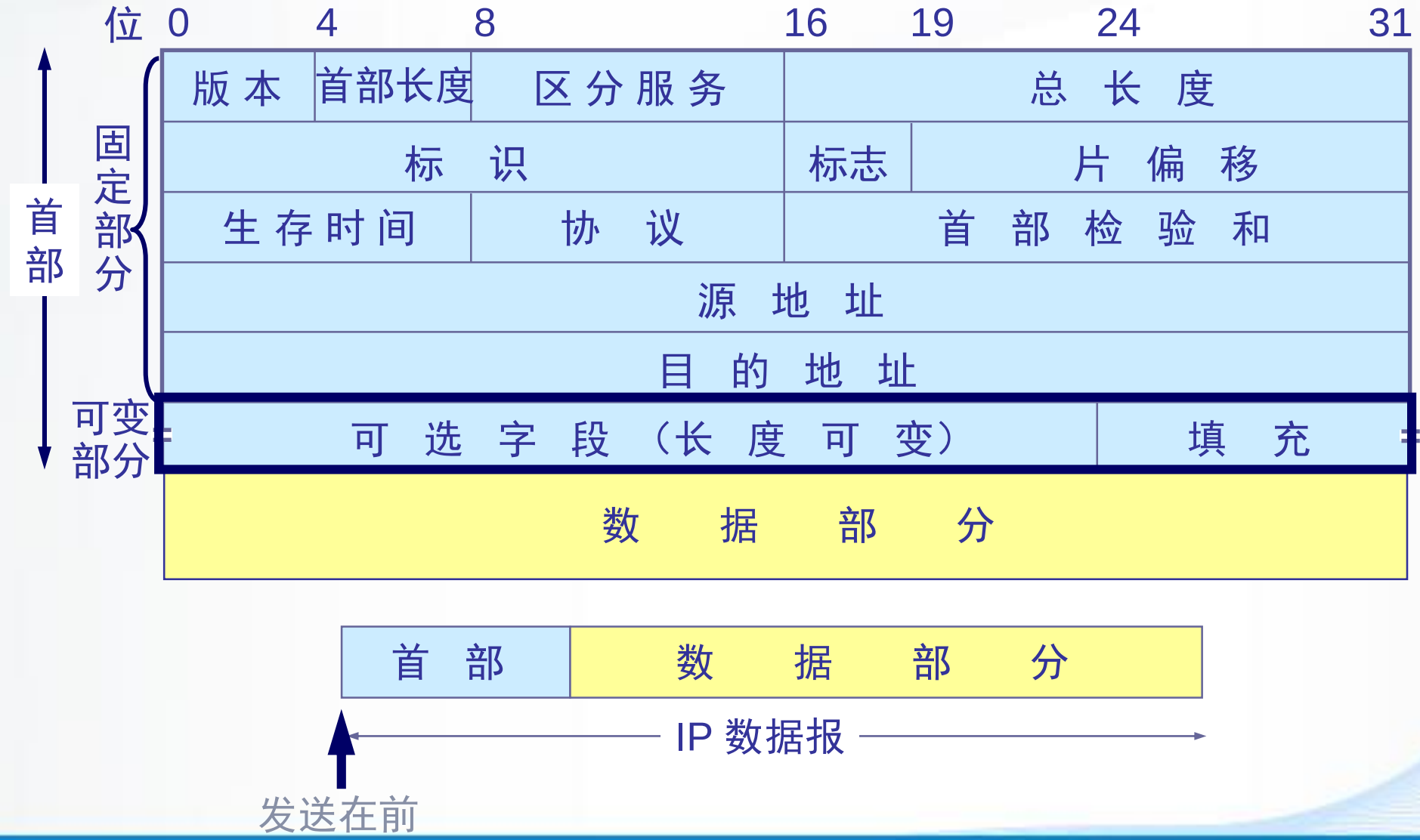
IP数据报


位
0
4
8
16
19
24
31
首部
固定部分
可变部分
版本
首部长度
区分服务
总长度
标识
标志
片偏移
生存时间
协议
首部检验和
源地址
目的地址
可选字段（长度可变）
填充
数据部分
首部
数据部分
IP 数据报
发送在前

IP数据报


位
0
4
8
16
19
24
31
首部
固定部分
版本
首部长度
区分服务
总长度
标识
标志
片偏移
生存时间
协议
首部检验和
源地址
目的地址
可变部分
可选字段（长度可变）
填充
数据部分
首部
数据部分
IP 数据报
发送在前

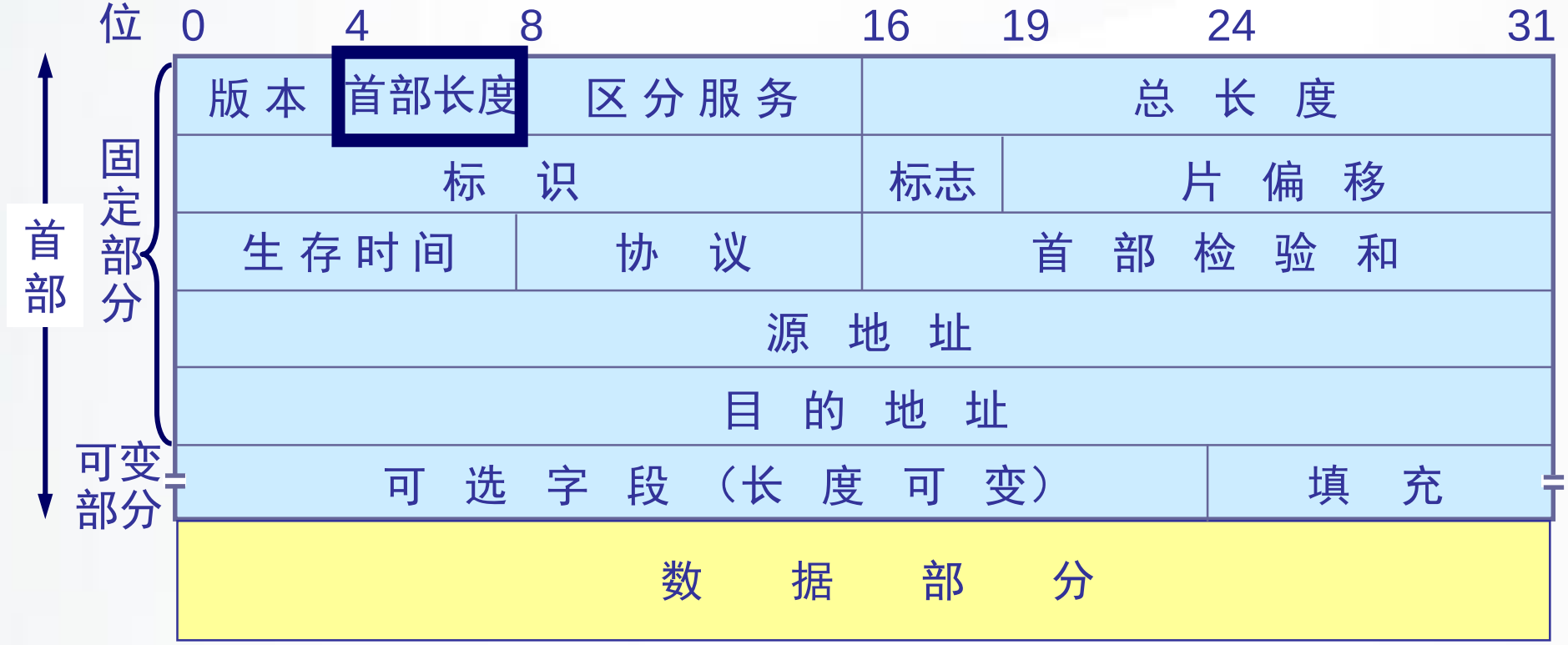
IP数据报


位 0
4
8
16
19
24
31
版本
首部长度
区分服务
总长度
标识
标志
片偏移
生存时间
协议
首部检验和
源地址
目的地址
可选字段（长度可变）
填充
数据部分
首部
固定部分
可变部分
首部
数据部分
IP 数据报
发送在前

IP数据报首部的固定部分中的各字段

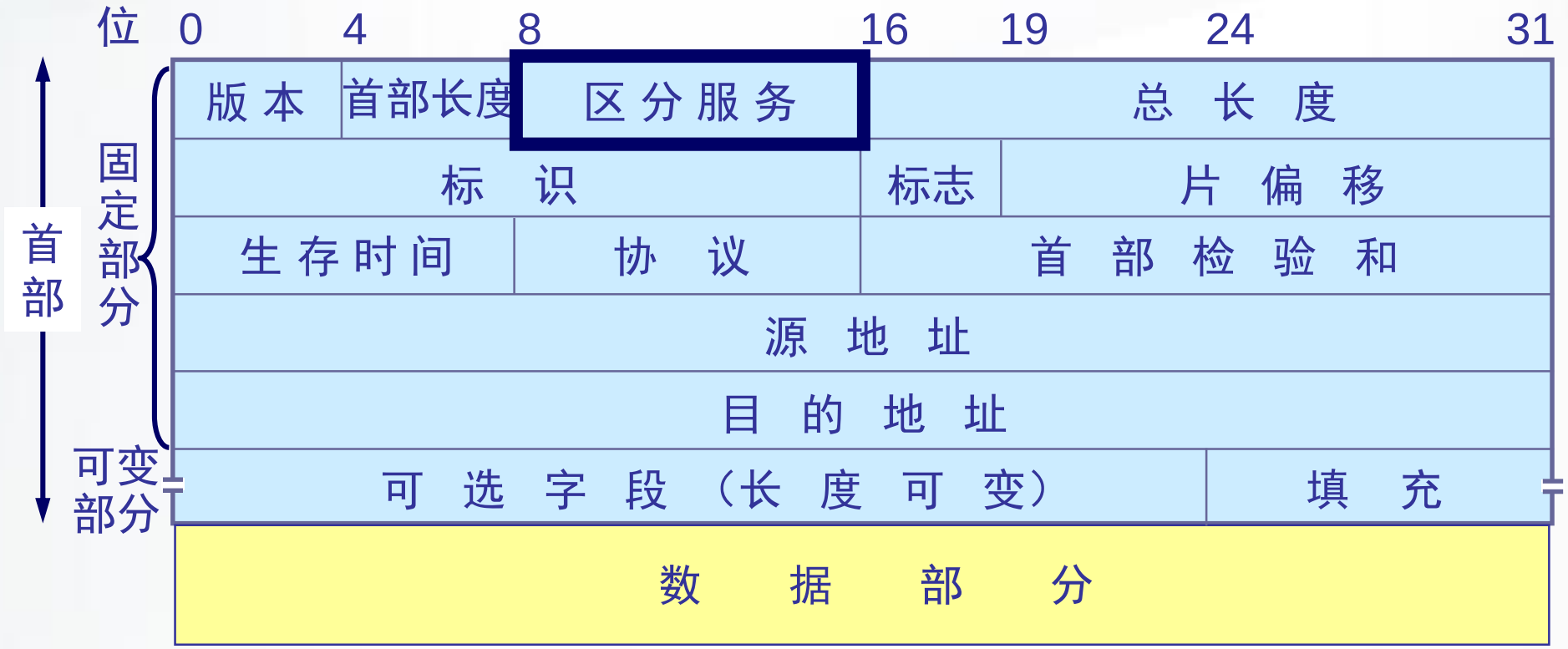
版本——占 4 位，指 IP 协议的版本
目前的 IP 协议版本号为 4 (即 IPv4)

IP数据报首部的固定部分中的各字段



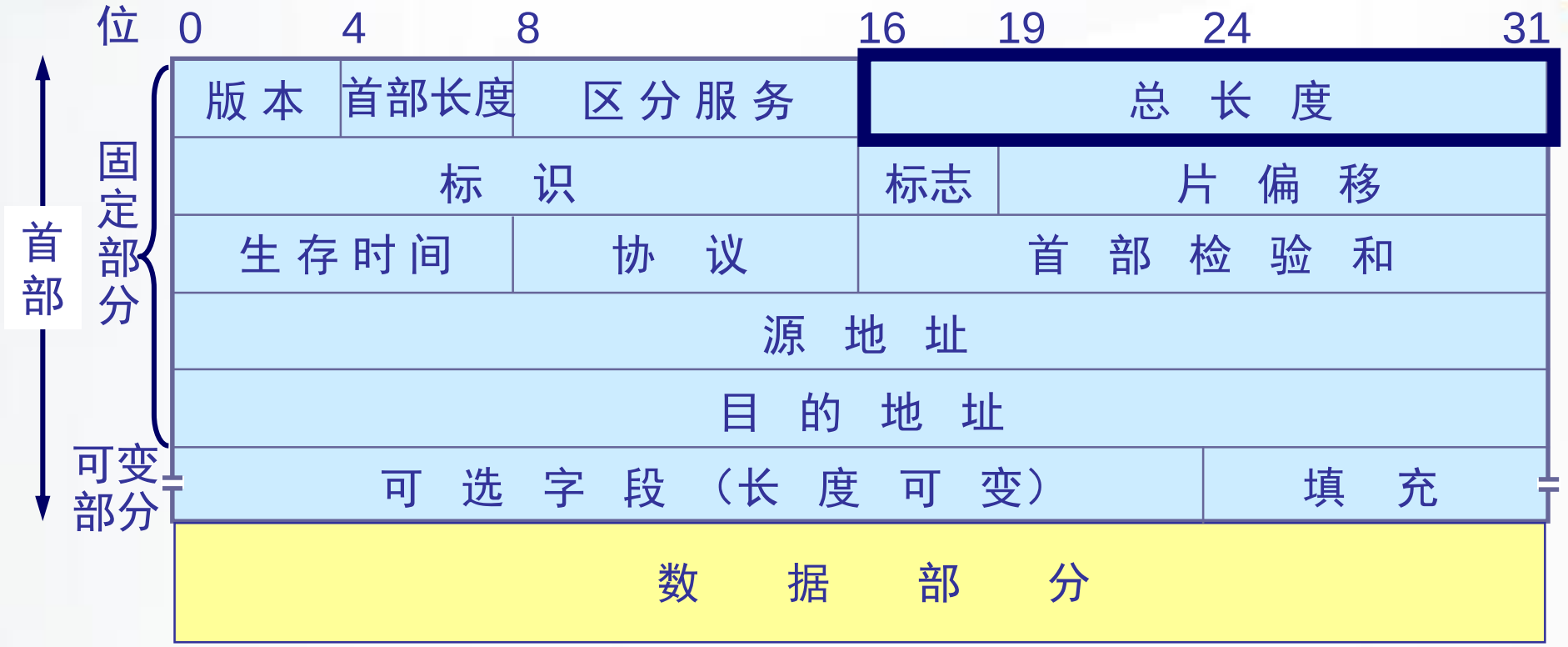
首部长度——占 4 位，可表示的最大数值是 15 个单位(一个单位为 4 字节)
因此 IP 的首部长度的最大值是 60 字节。

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

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。

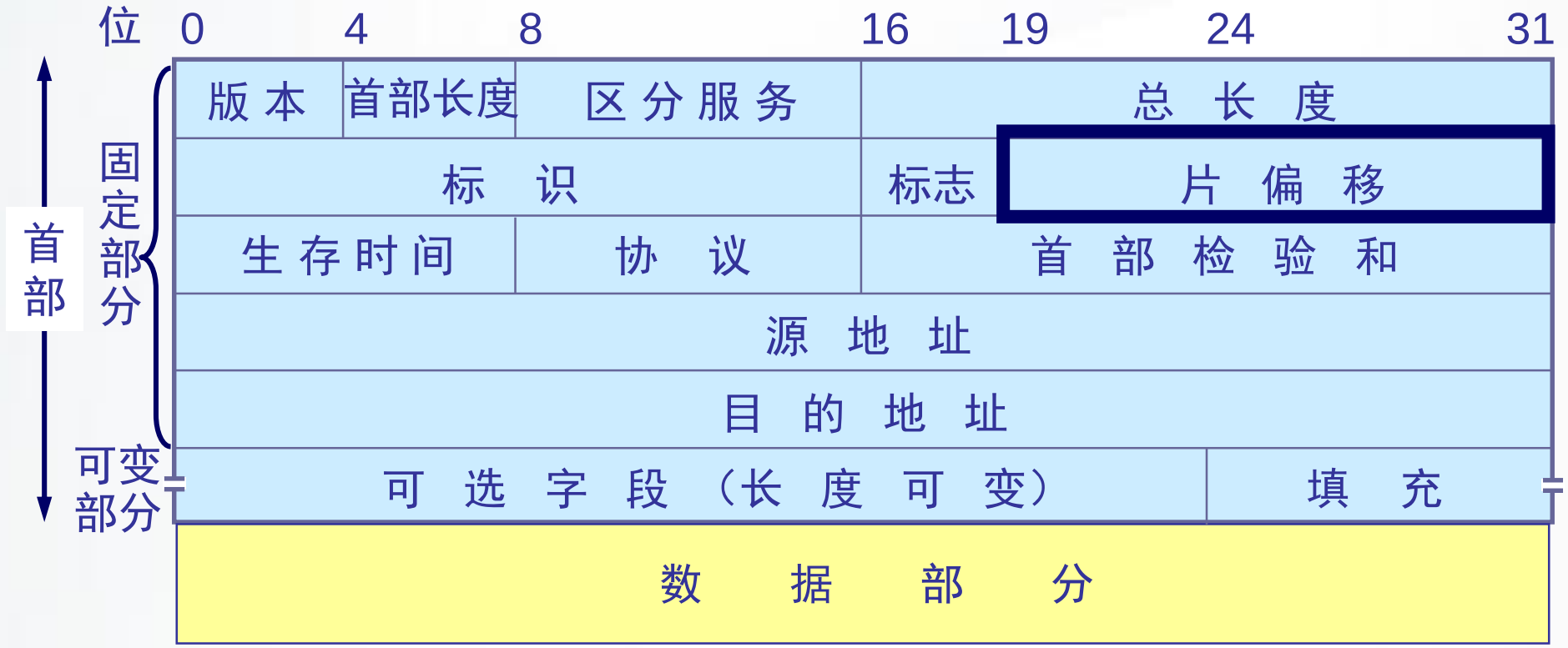
IP数据报首部的固定部分中的各字段

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

IP数据报首部的固定部分中的各字段

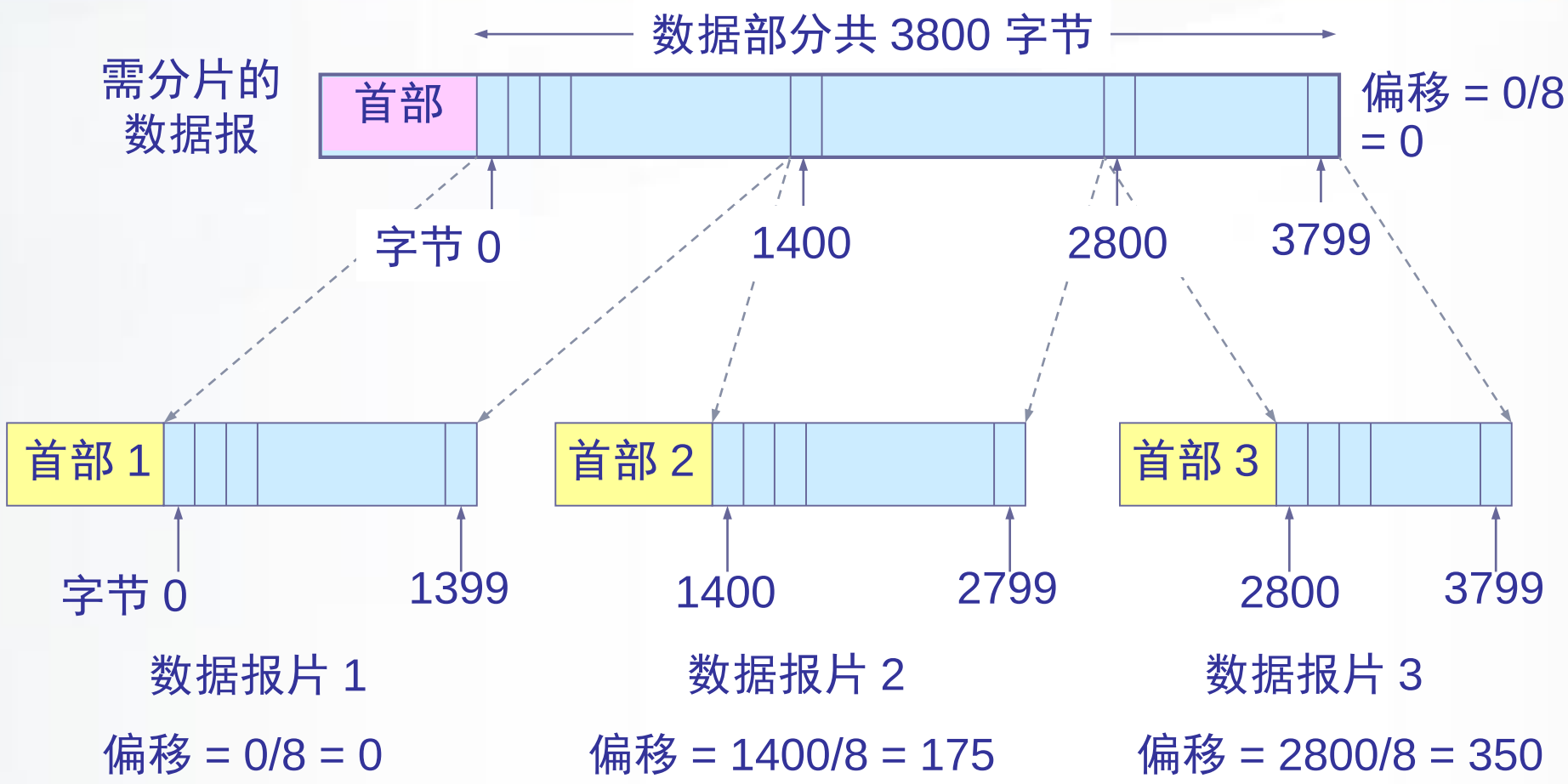
标志(flag)　占 3 位，目前只有前两位有意义。标志字段的最低位是 MF (More Fragment)。MF = 1 表示后面“还有分片”。MF = 0 表示最后一个分片。标志字段中间的一位是 DF (Don't Fragment)。只有当 DF = 0 时才允许分片。

IP数据报首部的固定部分中的各字段



片偏移(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。

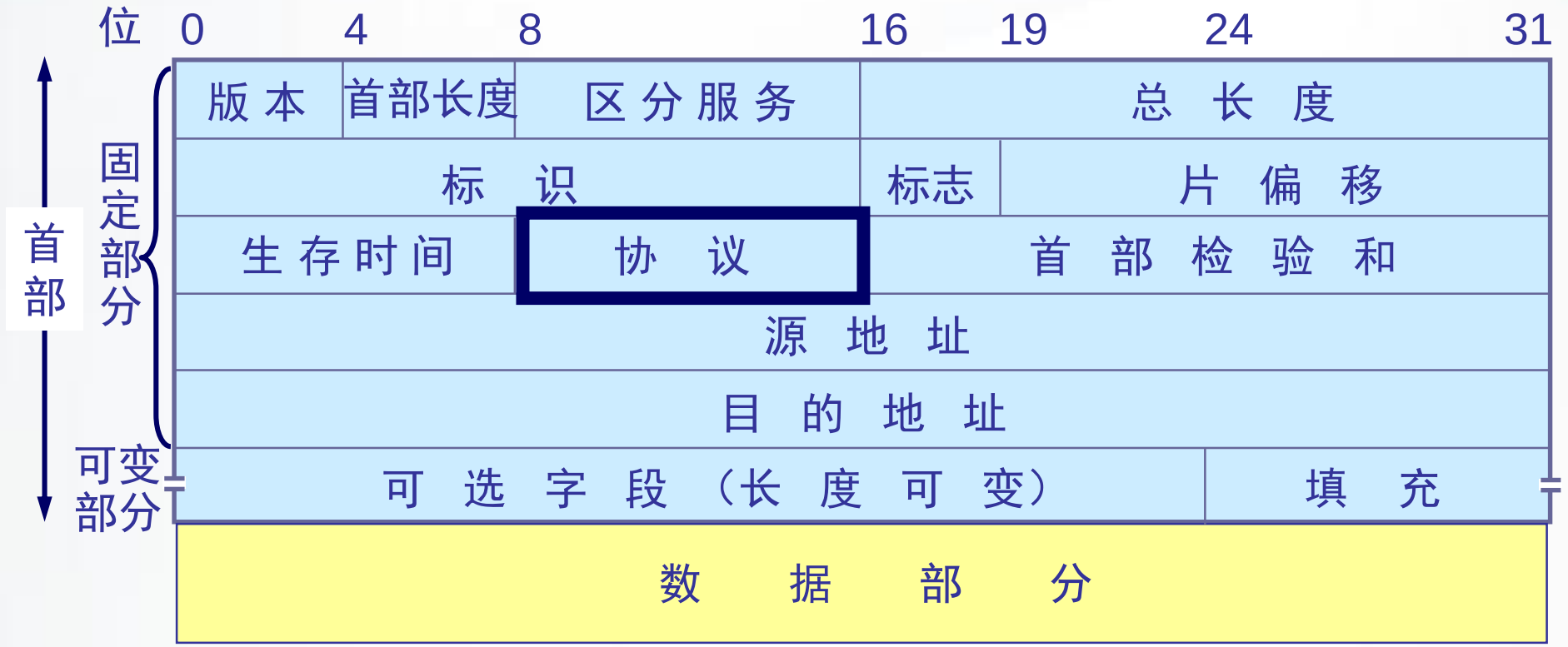
IP数据报分片举例


数据部分共 3800 字节
需分片的
数据报
首部
偏移 = 0/8
= 0
字节 0
1400
2800
3799
首部 1
首部 2
首部 3
字节 0
1399
1400
2799
2800
3799
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

IP数据报首部的固定部分中的各字段

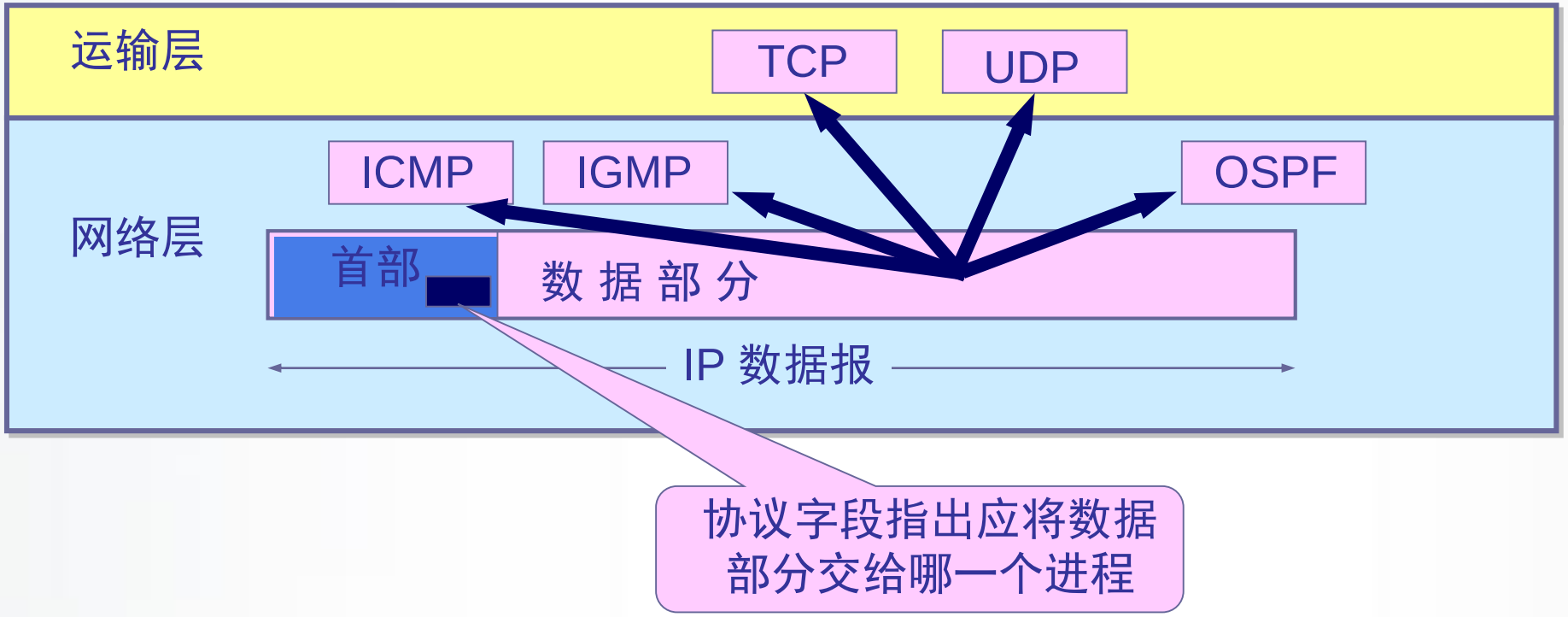
生存时间(8 位)记为 TTL (Time To Live)
数据报在网络中可通过的路由器数的最大值。

IP数据报首部的固定部分中的各字段



协议(8 位)字段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哪个处理过程

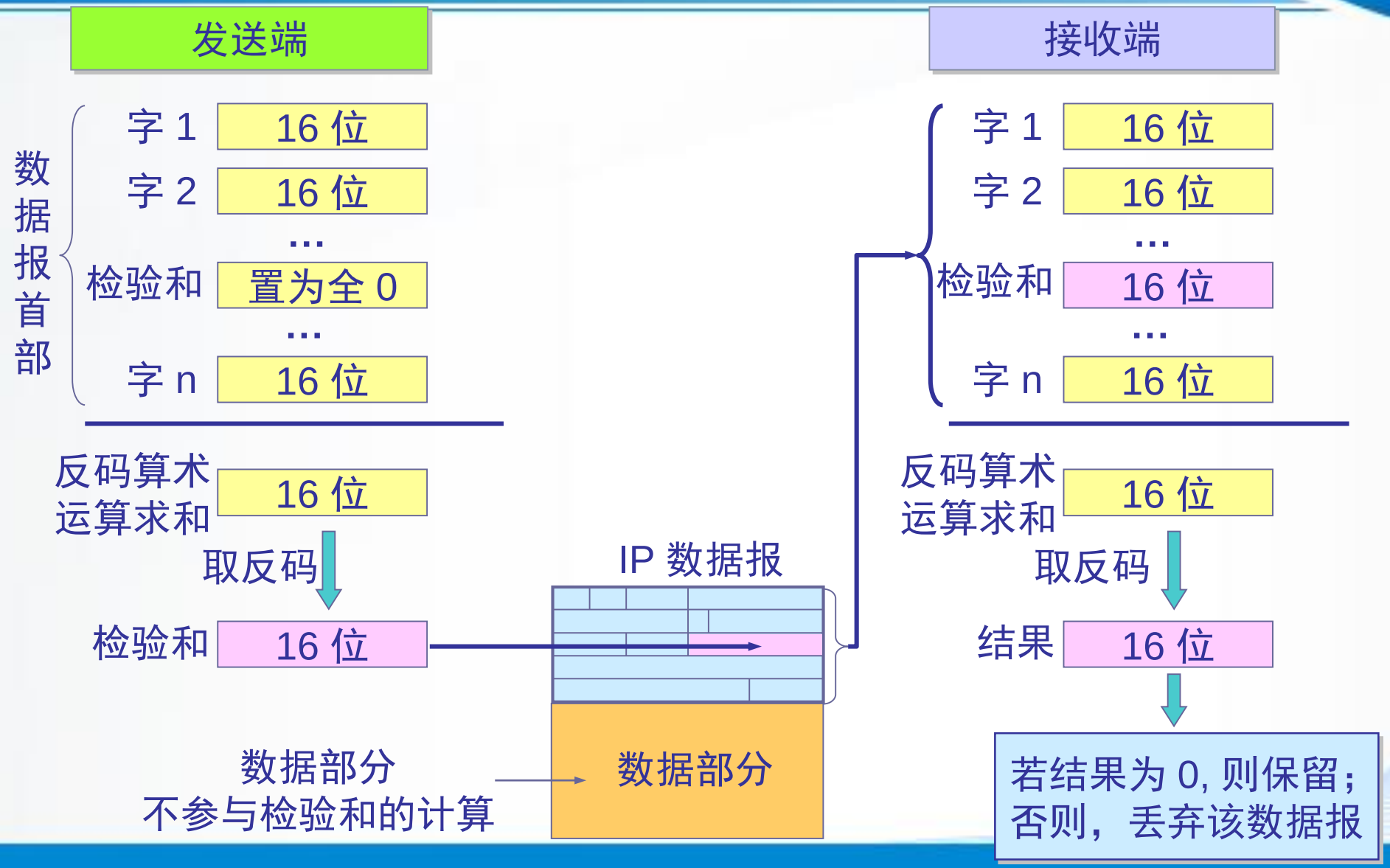
IP数据报首部的固定部分中的各字段



IP数据报首部的固定部分中的各字段

首部检验和(16 位)字段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。

校验过程


发送端
接收端
数据报首部
字 1
16 位
字 2
16 位
...
检验和
置为全 0
16 位
...
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
结果
16 位
若结果为 0, 则保留；
否则，丢弃该数据报

源地址和目的地址都各占 4 字节

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。

动手实验

实验4-2：抓取IP数据报

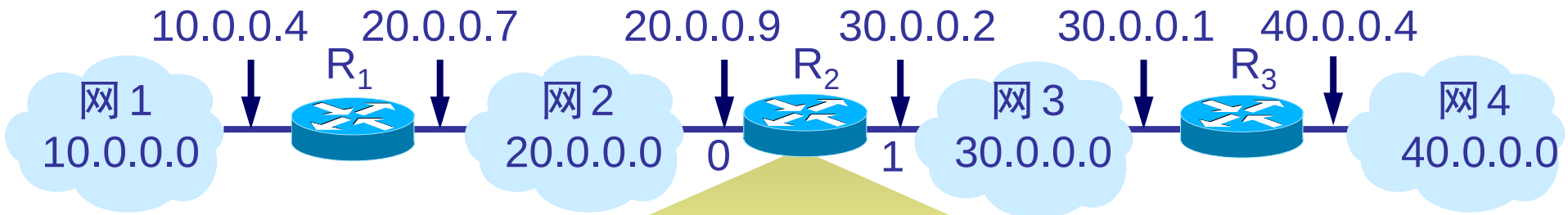
指引

- 网络层提供的两种服务
- 网际协议 IP
  - 虚拟互联网
  - IP地址
  - IP地址与硬件地址
  - IP数据报格式
  - IP转发分组的流程
- 划分子网和构造超网
- 网际控制报文协议 ICMP
- 因特网的路由选择协议
- IPv6
- IP 多播
- 虚拟专用网 VPN 和网络地址转换 NAT

路由表

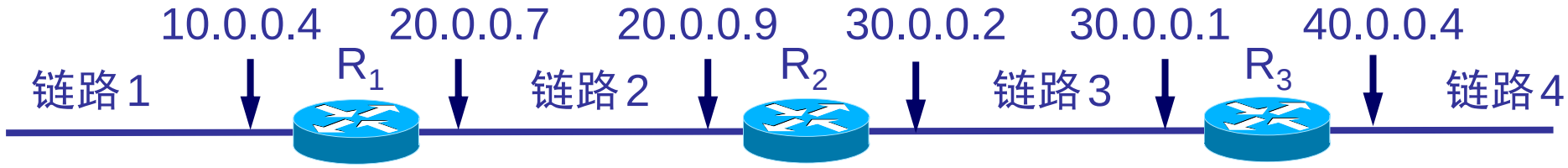
- 有四个 A 类网络通过三个路由器连接在一起。每一个网络上都可能有成千上万个主机。
- 可以想像，若按目的主机号来制作路由表，则所得出的路由表就会过于庞大。
- 但若按主机所在的网络地址来制作路由表，那么每一个路由器中的路由表就只包含 4 个项目。这样就可使路由表大大简化。

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



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。

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)

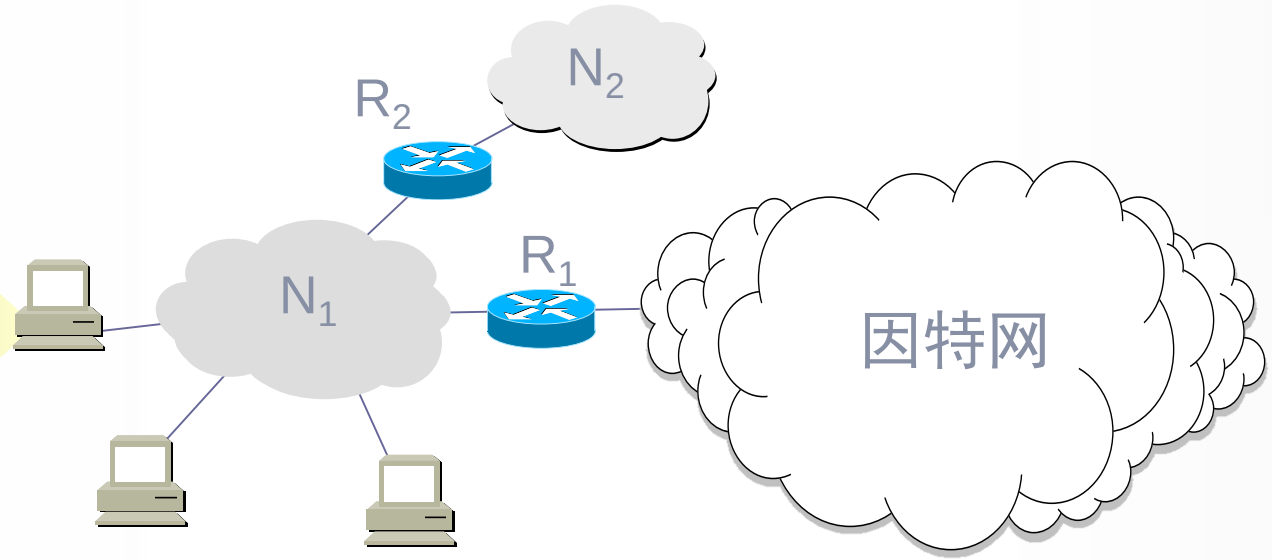
- 路由器还可采用默认路由以减少路由表所占用的空间和搜索路由表所用的时间。
- 这种转发方式在一个网络只有很少的对外连接时是很有用的。
- 默认路由在主机发送 IP 数据报时往往更能显示出它的好处。
- 如果一个主机连接在一个小网络上，而这个网络只用一个路由器和因特网连接，那么在这种情况下使用默认路由是非常合适的。

路由选择举例

- 只要目的网络不是 N1 和 N2，就一律选择默认路由，把数据报先间接交付路由器 R1，让 R1 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



注意

- 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。

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。

指引

- 网络层提供的两种服务
- 网际协议 IP
- 划分子网和构造超网
  - 划分子网
  - 无分址编址
  - 构造超网
- 网际控制报文协议 ICMP
- 因特网的路由选择协议
- IPv6
- IP 多播
- 虚拟专用网 VPN 和网络地址转换 NAT

为什么要划分子网

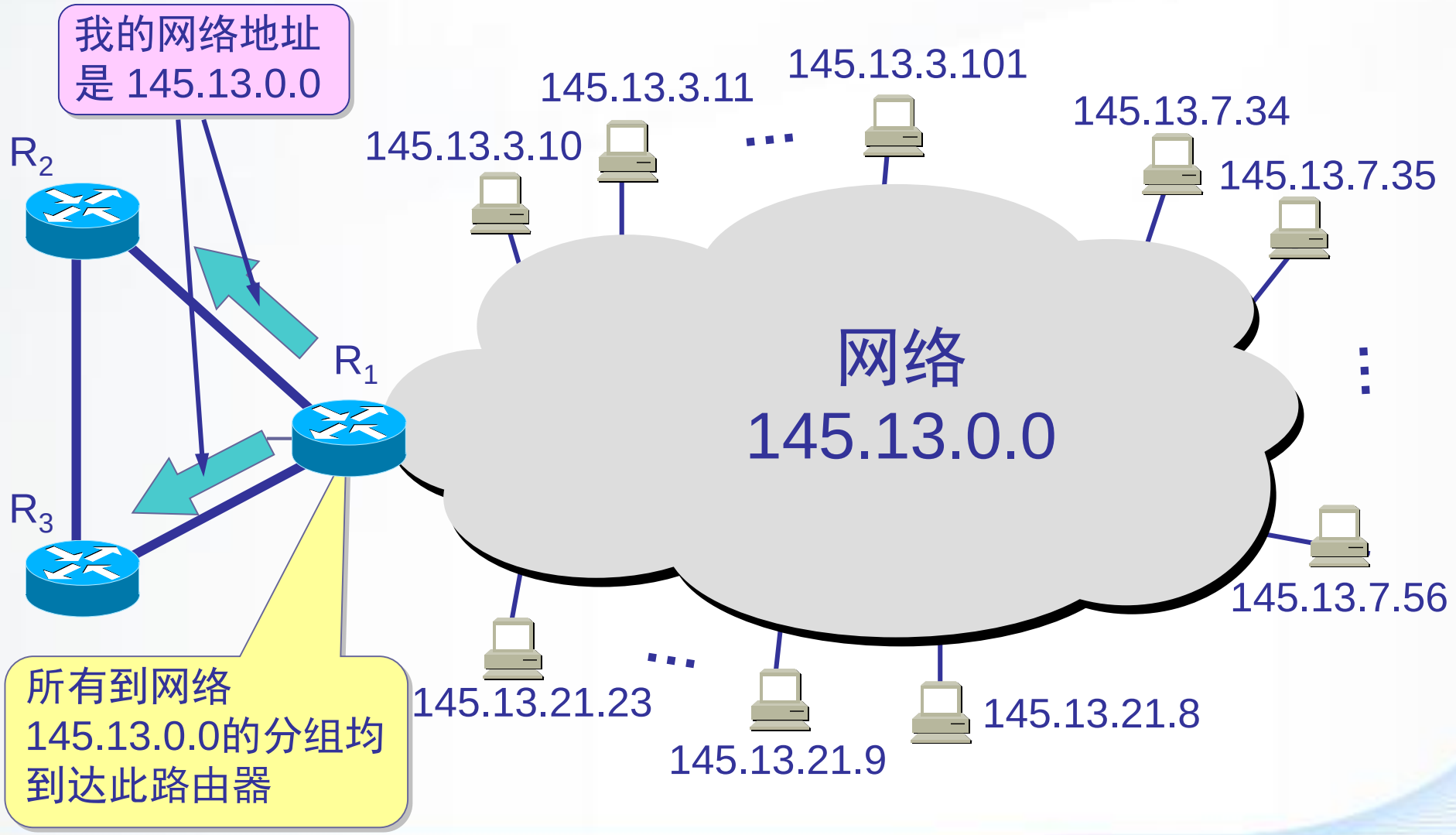
- 在 ARPANET 的早期，IP 地址的设计确实不够合理。
  - IP 地址空间的利用率有时很低。
  - 给每一个物理网络分配一个网络号会使路由表变得太大因而使网络性能变坏。
  - 两级的 IP 地址不够灵活。
- 解决的办法：从两级IP变成三级IP。
  - 从 1985 年起在 IP 地址中又增加了一个“子网号字段”，使两级的 IP 地址变成为三级的 IP 地址。
  - 这种做法叫作划分子网(subnetting) 。划分子网已成为因特网的正式标准协议。

划分子网的基本思路

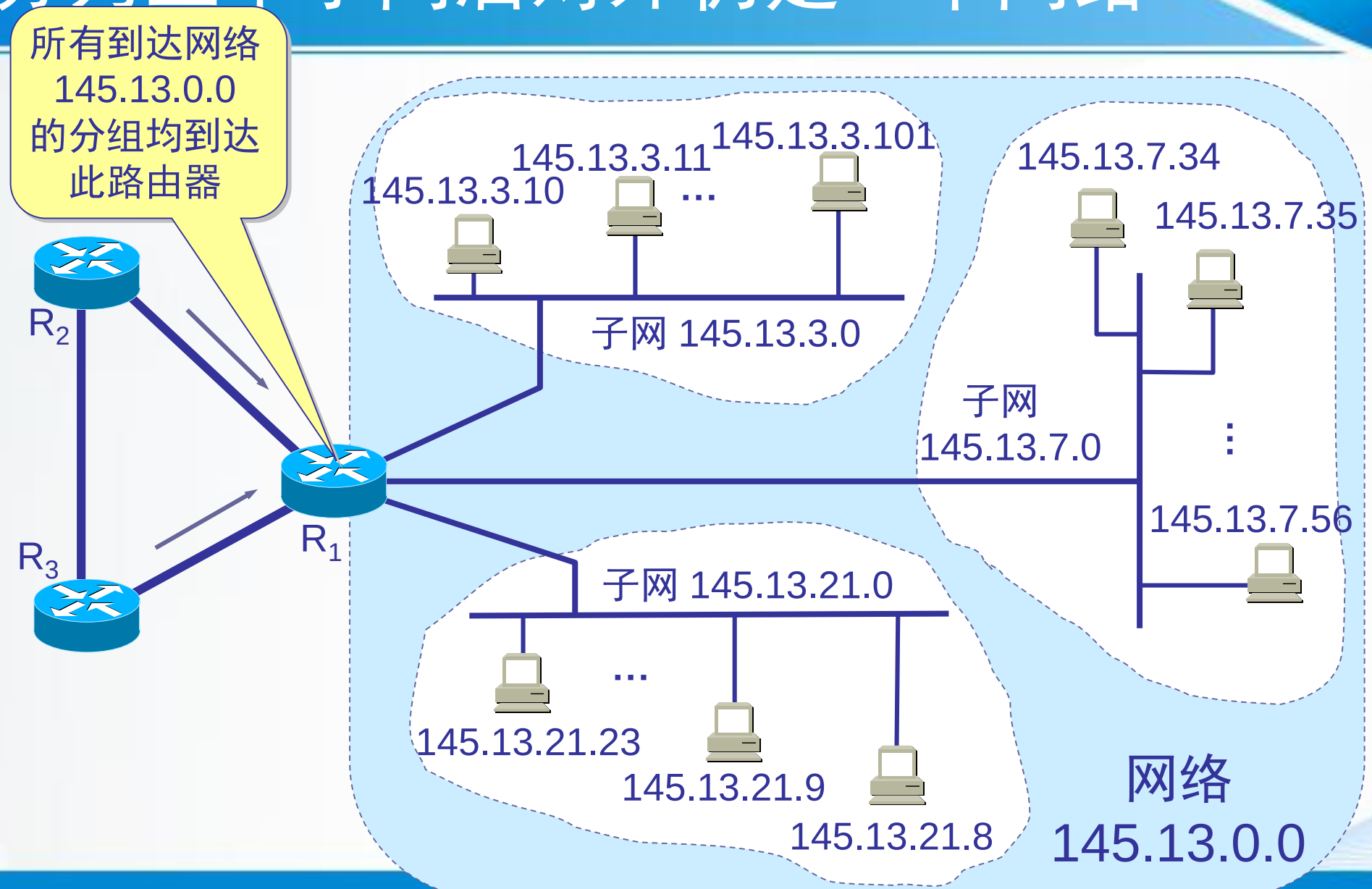
- 划分子网纯属一个单位内部的事情。单位对外仍然表现为没有划分子网的网络。
- 从主机号借用若干个位作为子网号 subnet-id，而主机号 host-id 也就相应减少了若干个位。凡是从其他网络发送给本单位某个主机的 IP 数据报，仍然是根据 IP 数据报的目的网络号 net-id，先找到连接在本单位网络上的路由器。
- 然后此路由器在收到 IP 数据报后，再按目的网络号 net-id 和子网号 subnet-id 找到目的子网。
- 最后就将 IP 数据报直接交付目的主机。

IP地址 ::= {<网络号>, <子网号>, <主机号>}

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



划分子网后变成了三级结构

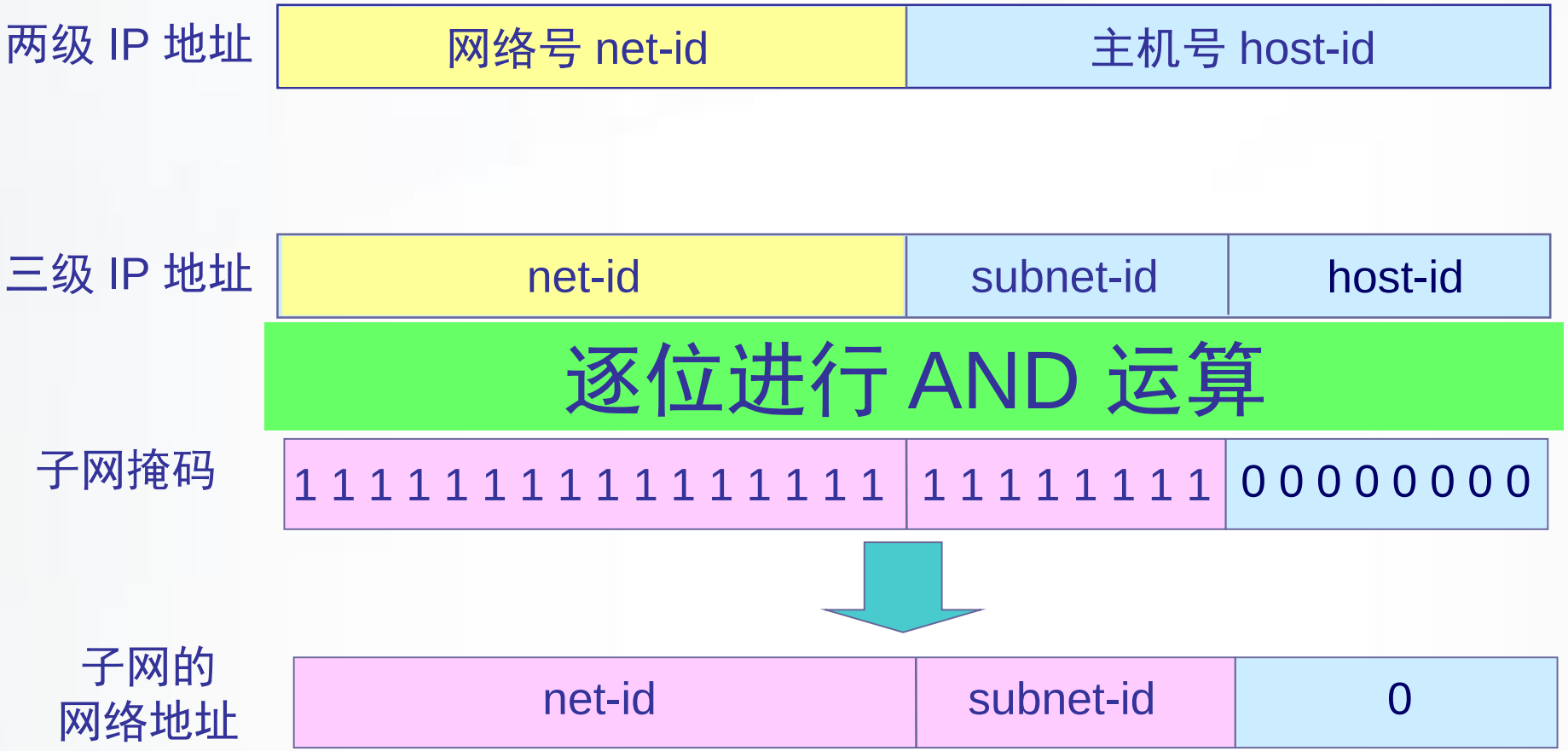
- 当没有划分子网时，IP 地址是两级结构。
- 划分子网后 IP 地址就变成了三级结构。
- 划分子网只是把 IP 地址的主机号 host-id 这部分进行再划分，而不改变 IP 地址原来的网络号 net-id。

子网掩码

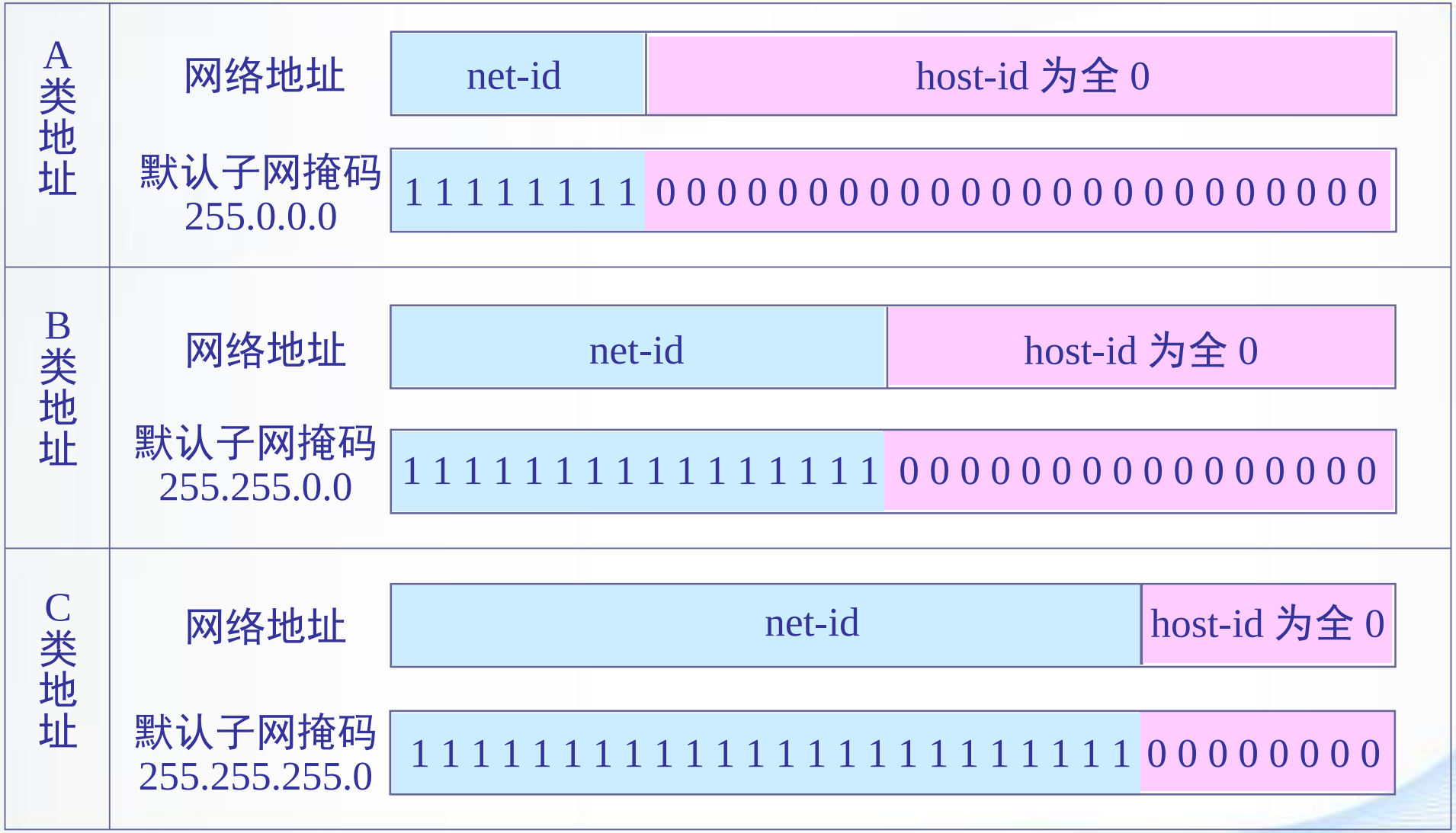
- 分成三级结构后，问题就来了：从一个 IP 数据报的首部并无法判断源主机或目的主机所连接的网络是否进行了子网划分。怎么办?
  - 使用子网掩码(subnet mask)可以找出 IP 地址中的子网部分。

IP 地址的各字段和子网掩码

（IP 地址） AND （子网掩码）=网络地址



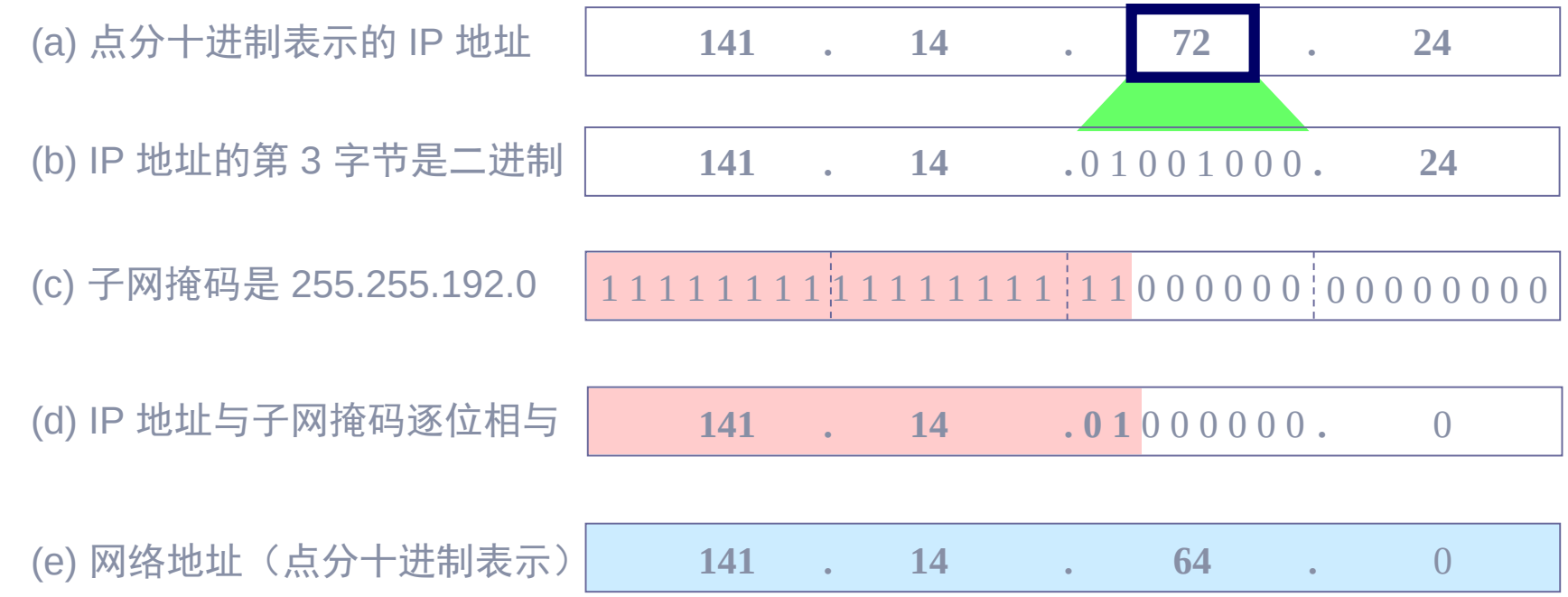
默认子网掩码


A类地址
网络地址
net-id
host-id 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
net-id
host-id 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
net-id
host-id 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

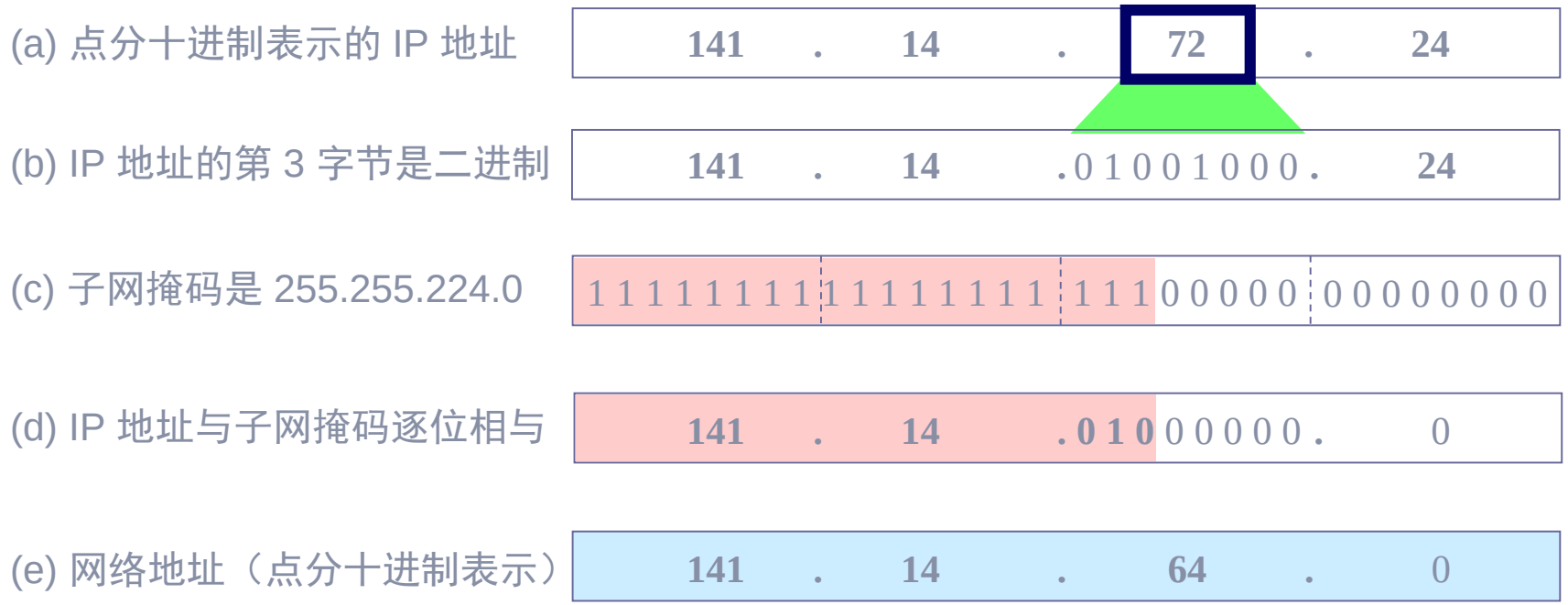
子网掩码是一个重要属性

- 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。

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。



在上例中，IP 地址是 141.14.72.24，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。

使用子网掩码的分组转发过程

- 在不划分子网的两级 IP 地址下，从 IP 地址得出网络地址是个很简单的事。
- 但在划分子网的情况下，从 IP 地址却不能唯一地得出网络地址来，这是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。

使用子网掩码的分组转发过程

(1) 从收到的分组的首部提取目的 IP 地址 *D*。

(2) 先判断是否可以直接交付。先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)。

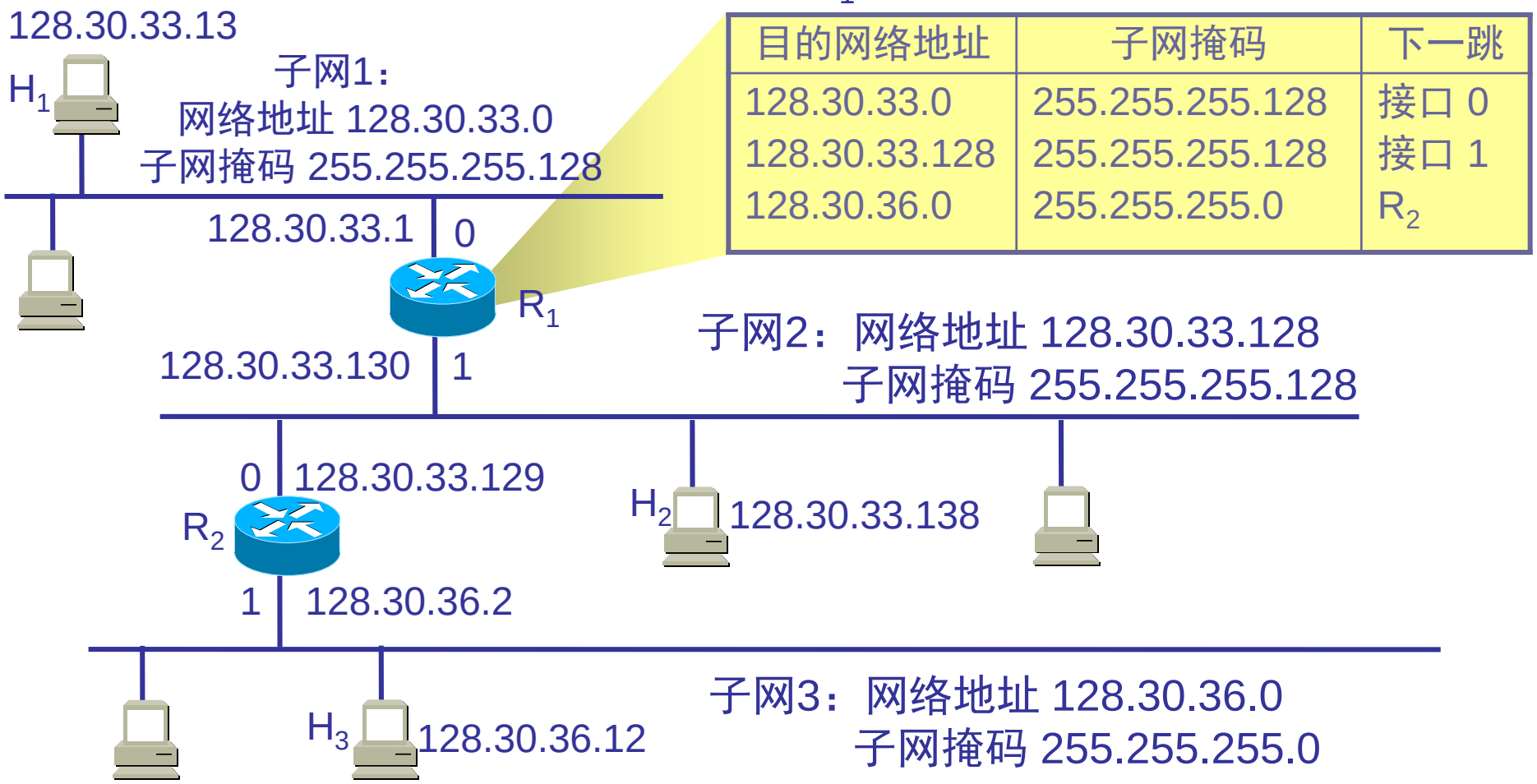
(3) 若路由表中有目的地址为 *D* 的特定主机路由，则将 分组传送给指明的下一跳路由器；否则，执行(4)。

(4) 对路由表中的每一行的子网掩码和 *D* 逐位相“与”，若其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。

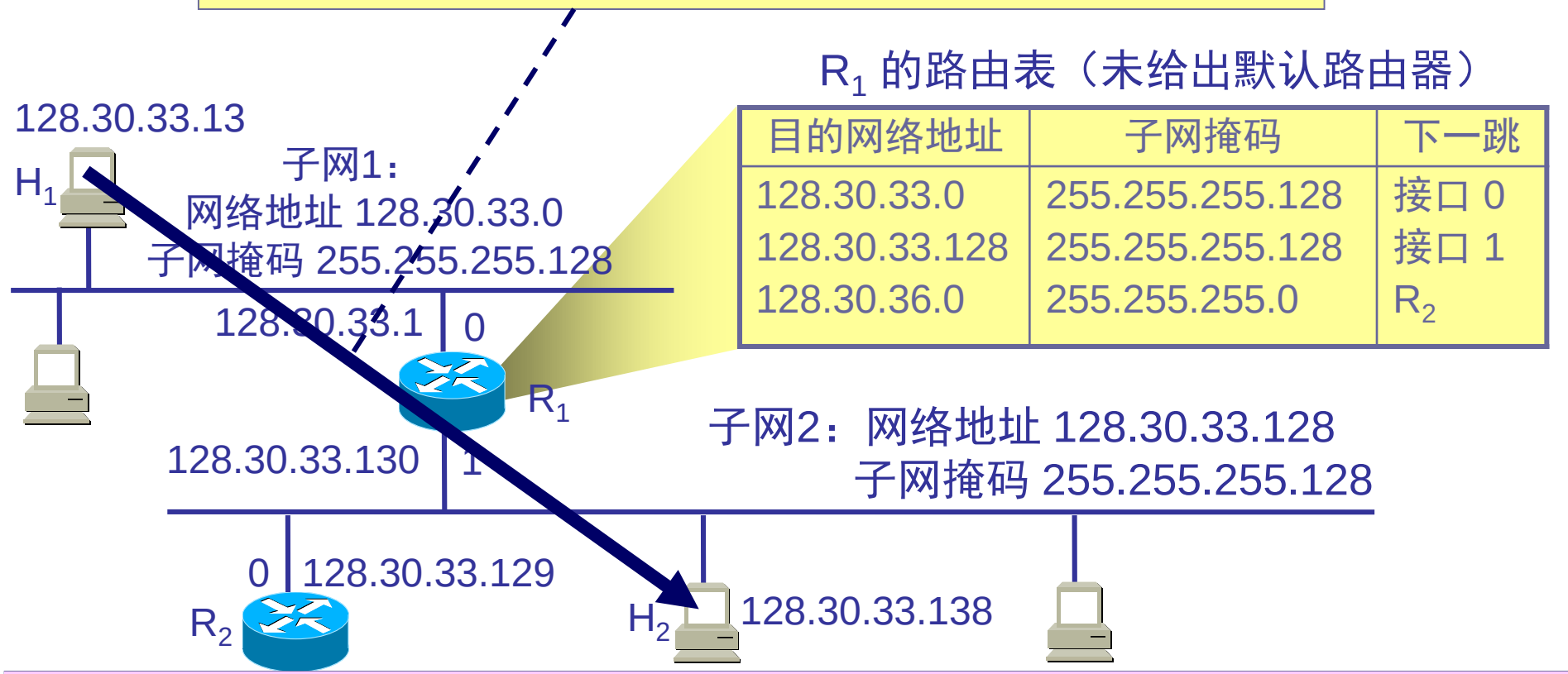
已知互联网和路由器 R_1 中的路由表。主机 H_1 向 H_2 发送分组。试讨论 R_1 收到 H_1 向 H_2 发送的分组后查找路由表的过程。



R_1 的路由表（未给出默认路由）

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

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）

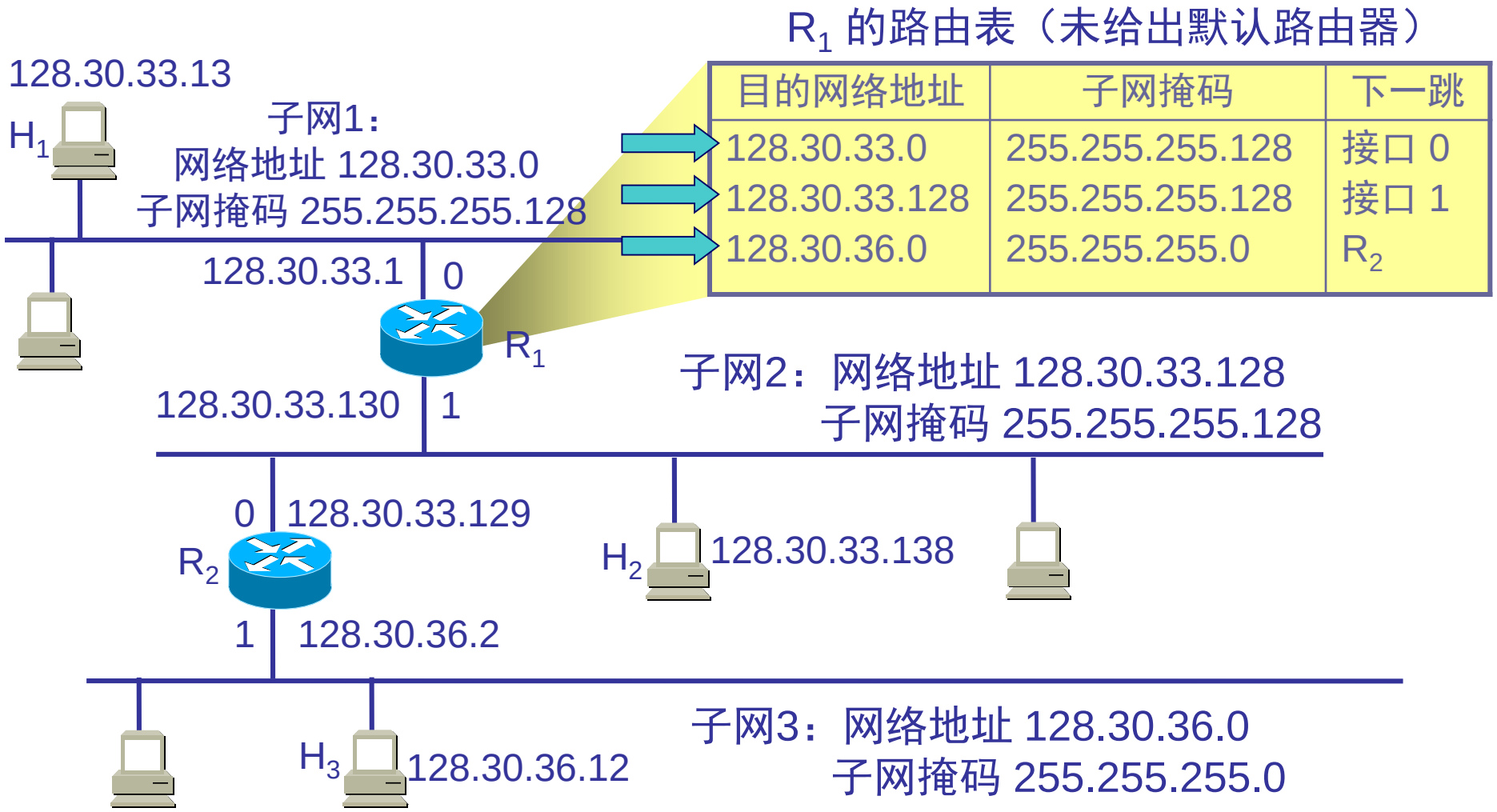
255.255.255.128 AND 128.30.33.138 的计算

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

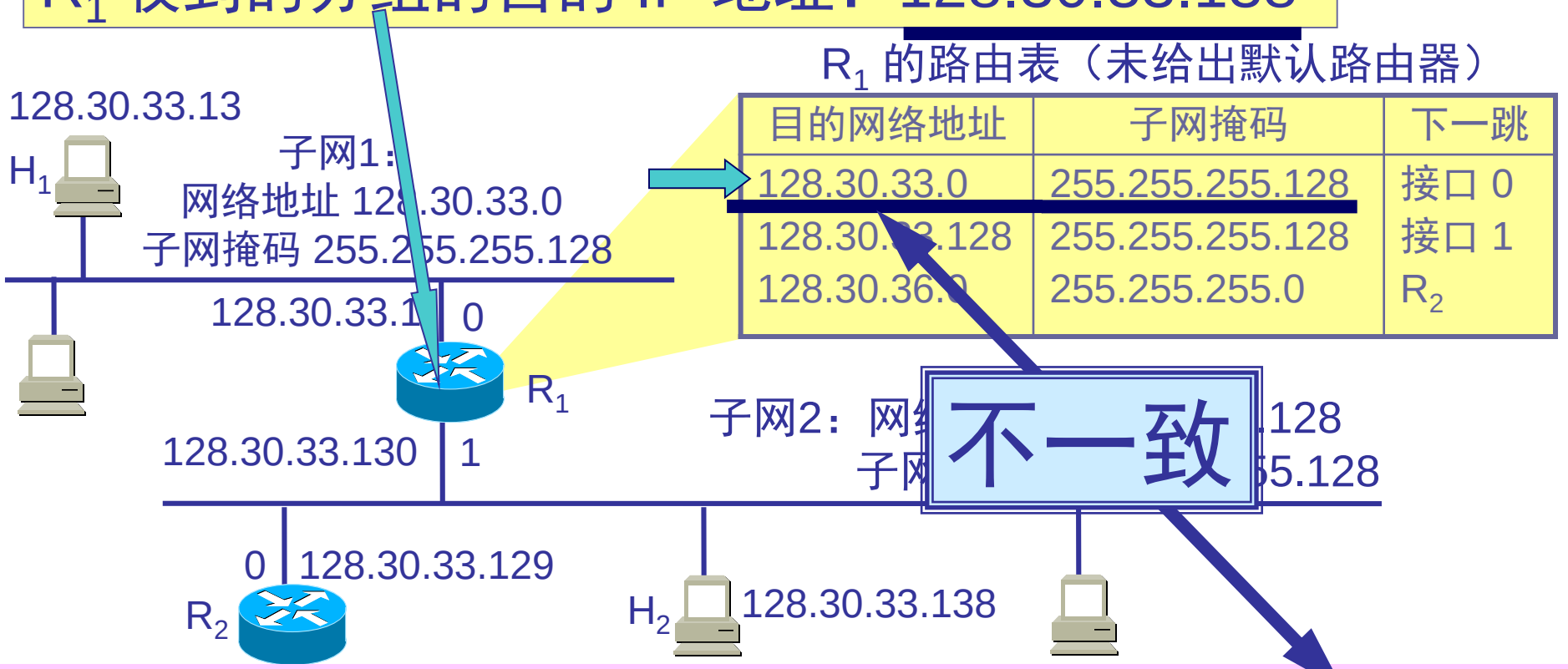
逐比特 AND 操作	255.255.255.128
	128. 30. 33.138
	128. 30. 33.128 ≠ H_1 的网络地址

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

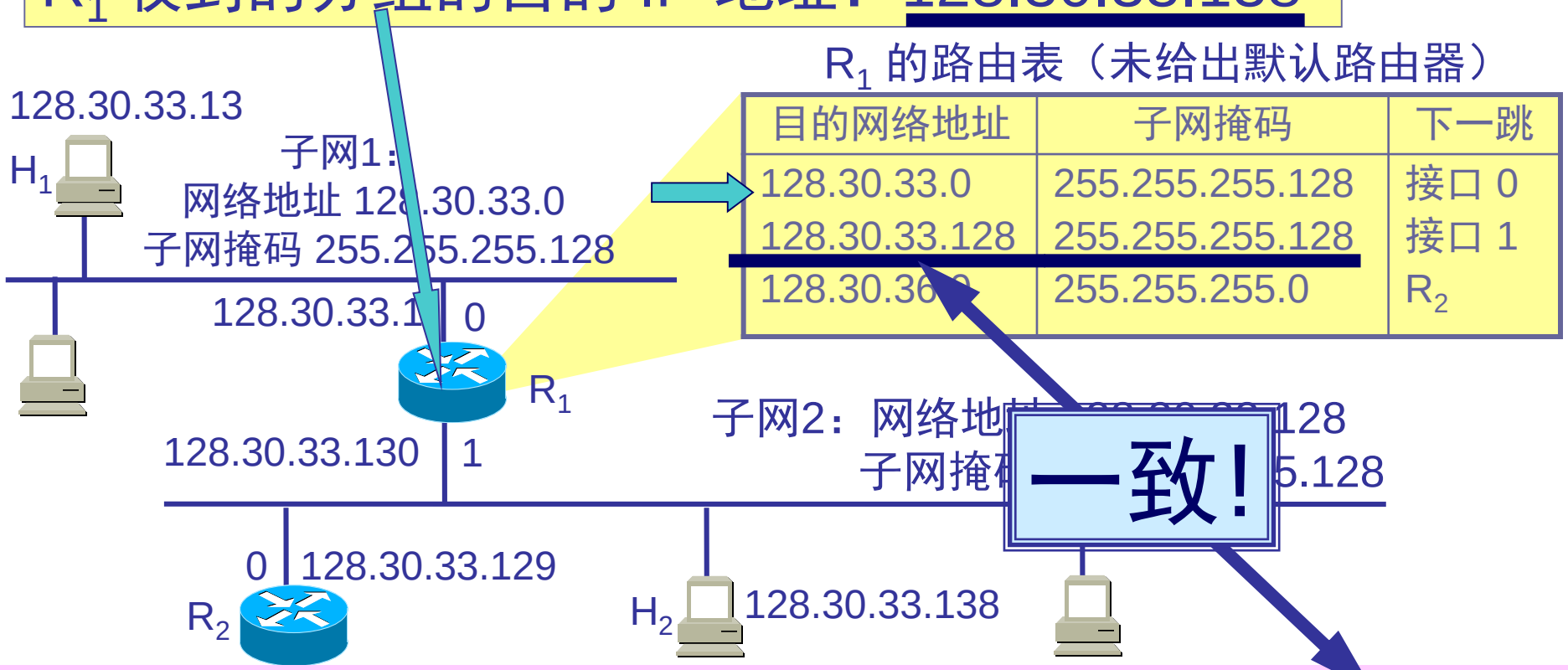
255.255.255.128 AND 128.30.33.138 = 128.30.33.128

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 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

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

指引

- 网络层提供的两种服务
- 网际协议 IP
- 划分子网和构造超网
  - 划分子网
  - 无分类编址
  - 构造超网
- 网际控制报文协议 ICMP
- 因特网的路由选择协议
- IPv6
- IP 多播
- 虚拟专用网 VPN 和网络地址转换 NAT

网络前缀

- 划分子网在一定程度上缓解了因特网在发展中遇到的困难。然而在 1992 年因特网仍然面临三个必须尽早解决的问题，这就是：
  - 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!
  - 因特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）。
  - 整个 IPv4 的地址空间最终将全部耗尽。
- 解决办法：
  - 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。

无分类的两级编址

- 无分类的两级编址的记法是：

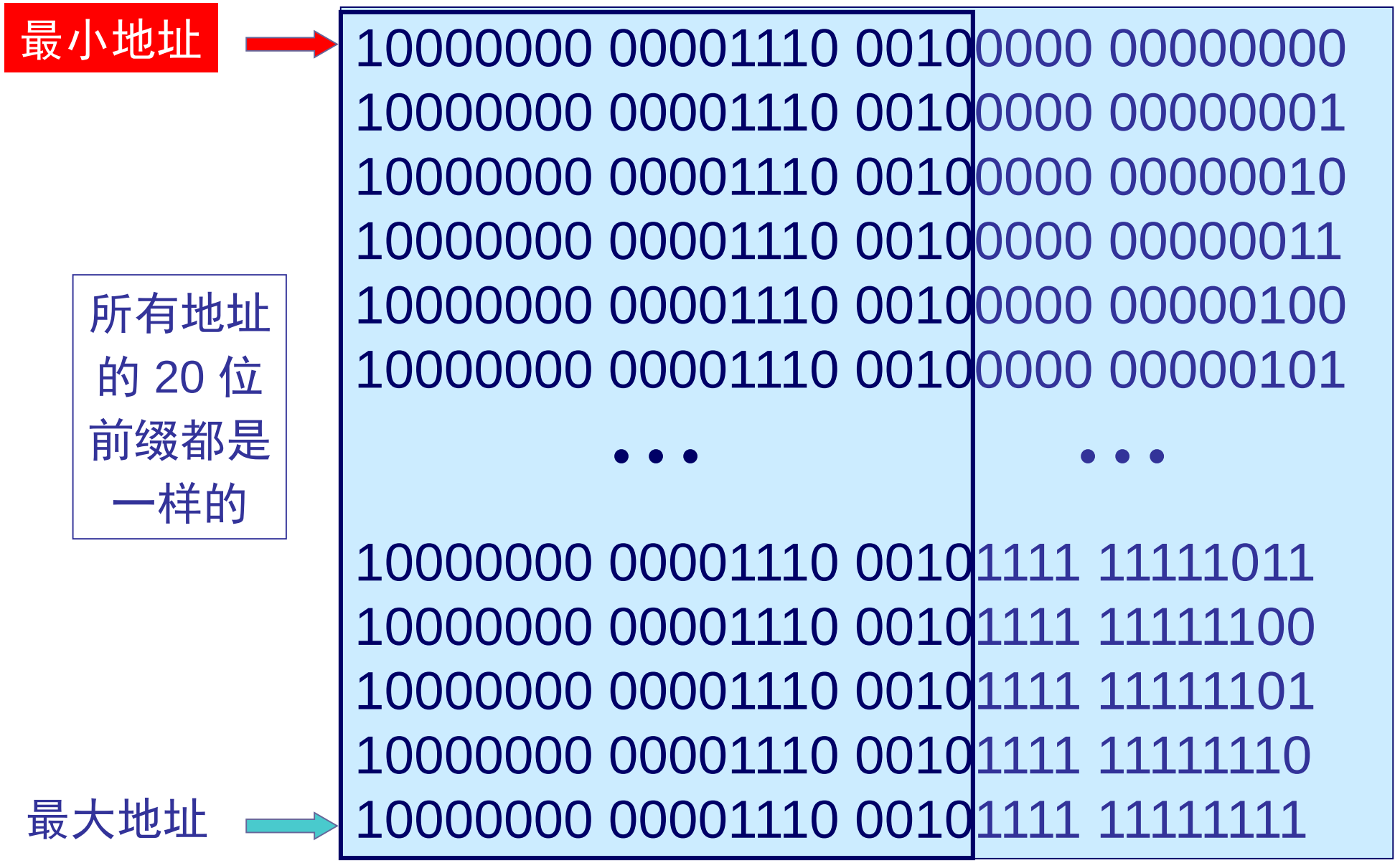
IP地址 ::= {<网络前缀>, <主机号>}

- CIDR 还使用“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）。
- CIDR 把网络前缀都相同的连续的 IP 地址组成“CIDR 地址块”。

CIDR地址块举例

- 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）



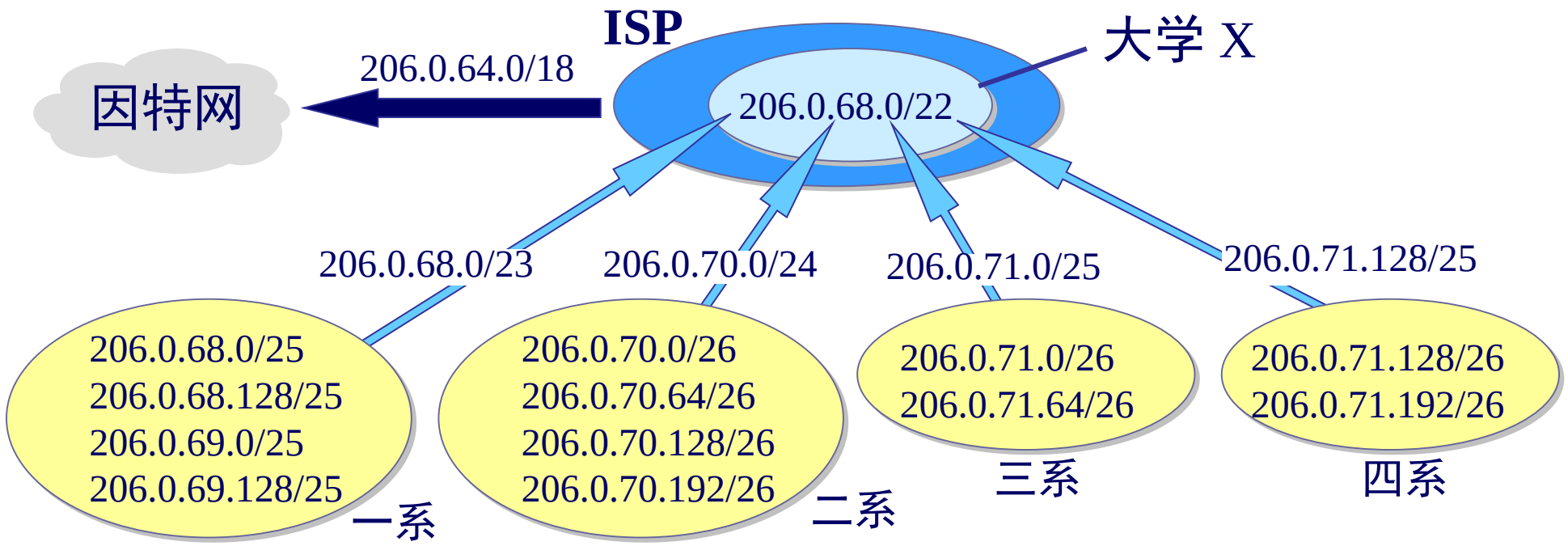
指引

- 网络层提供的两种服务
- 网际协议 IP
- 划分子网和构造超网
  - 划分子网
  - 无分址编址
  - 构造超网
- 网际控制报文协议 ICMP
- 因特网的路由选择协议
- IPv6
- IP 多播
- 虚拟专用网 VPN 和网络地址转换 NAT

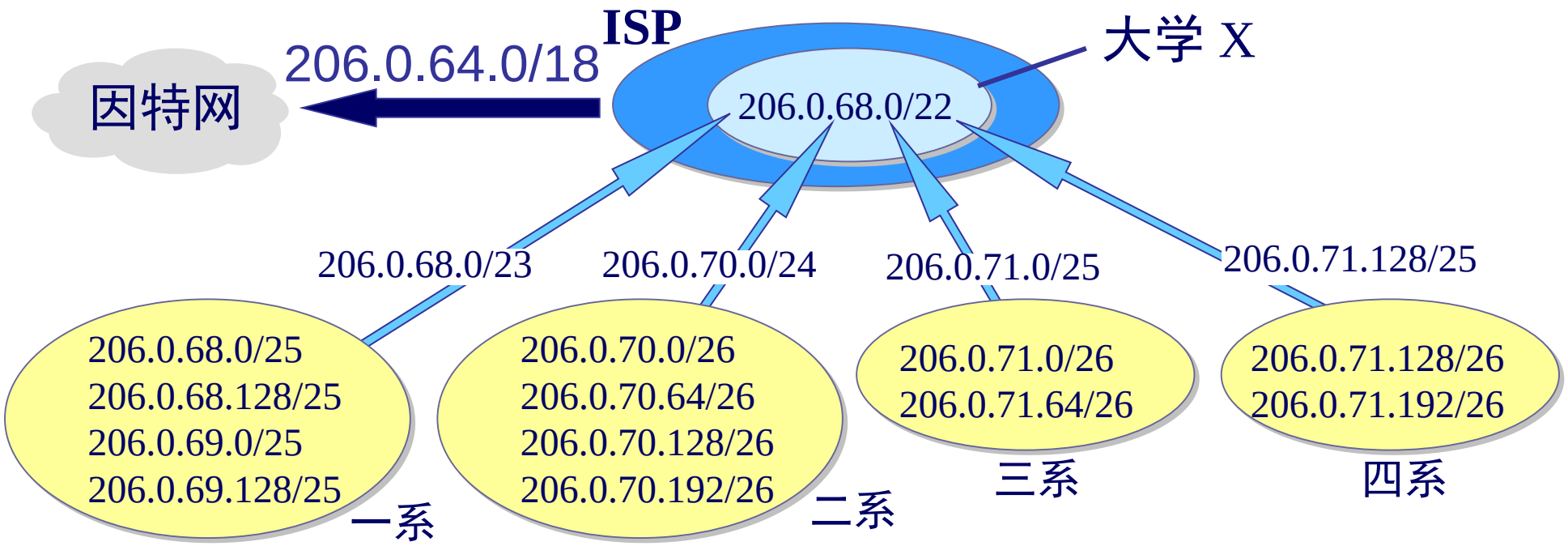
路由聚合(route aggregation)

- 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)。
- 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 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



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。

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　　（ISP）

　　206.0.71.128/25　（四系）

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. 10000010
		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　　（ISP）

　　206.0.71.128/25　　（四系）

再查找路由表中的第 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 11111111 10000000

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

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

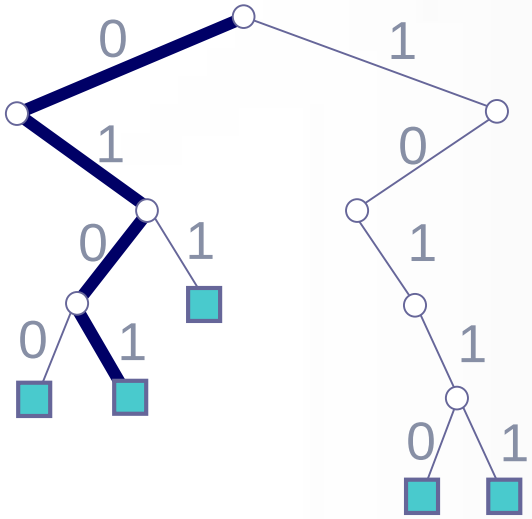
选择两个匹配的地址中更具体的一个，即选择最长前缀的地址。

二叉线索树与路由表查找

- 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。

二叉线索树与路由表查找举例

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



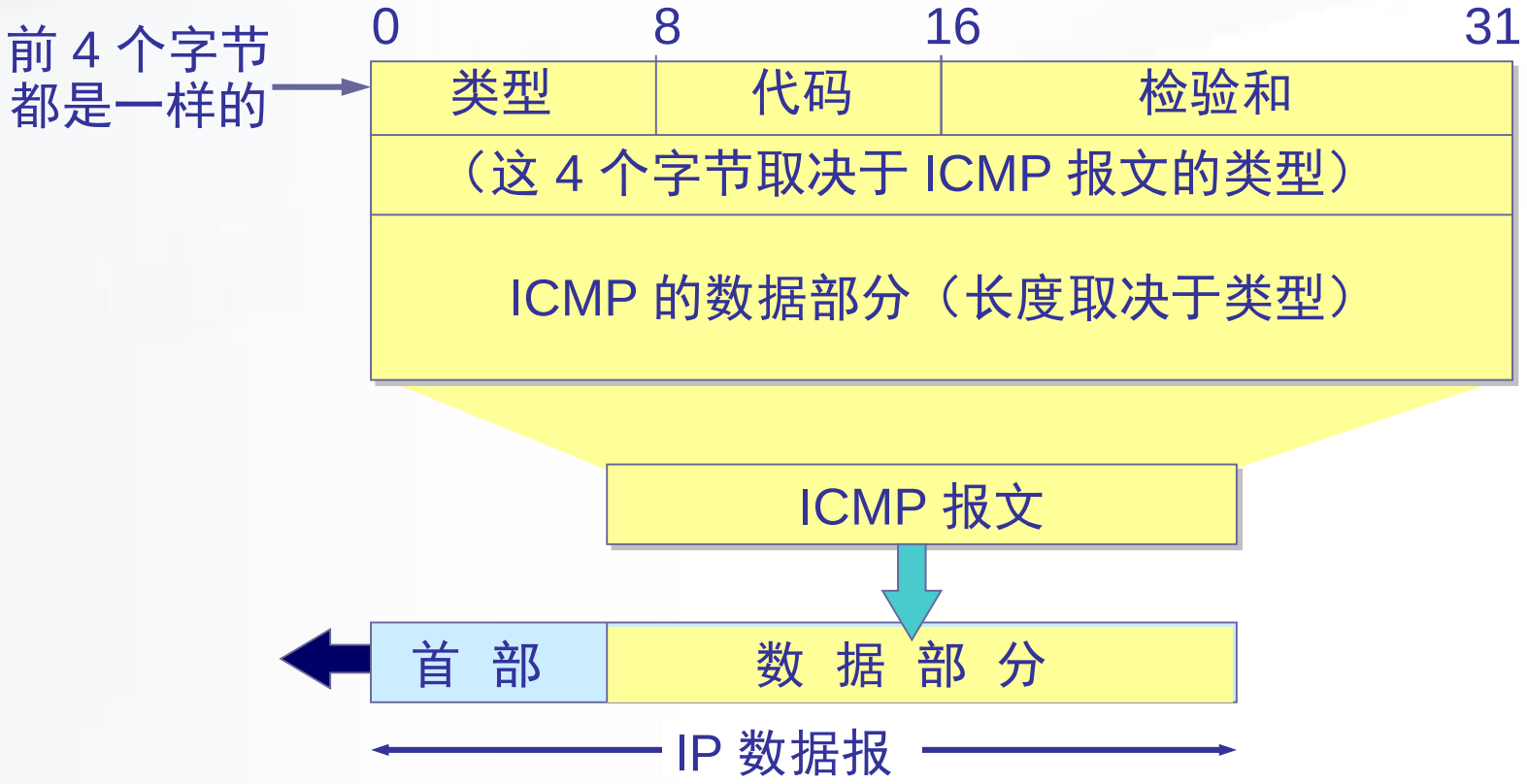
指引

- 网络层提供的两种服务
- 网际协议 IP
- 划分子网和构造超网
  - 划分子网
  - 无分址编址
  - 构造超网
- 网际控制报文协议 ICMP
- 因特网的路由选择协议
- IPv6
- IP 多播
- 虚拟专用网 VPN 和网络地址转换 NAT

ICMP简介

- 为了提高 IP 数据报交付成功的机会，在网际层使用了网际控制报文协议 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。
- ICMP 不是高层协议，而是 IP 层的协议。
- ICMP 报文作为 IP 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。

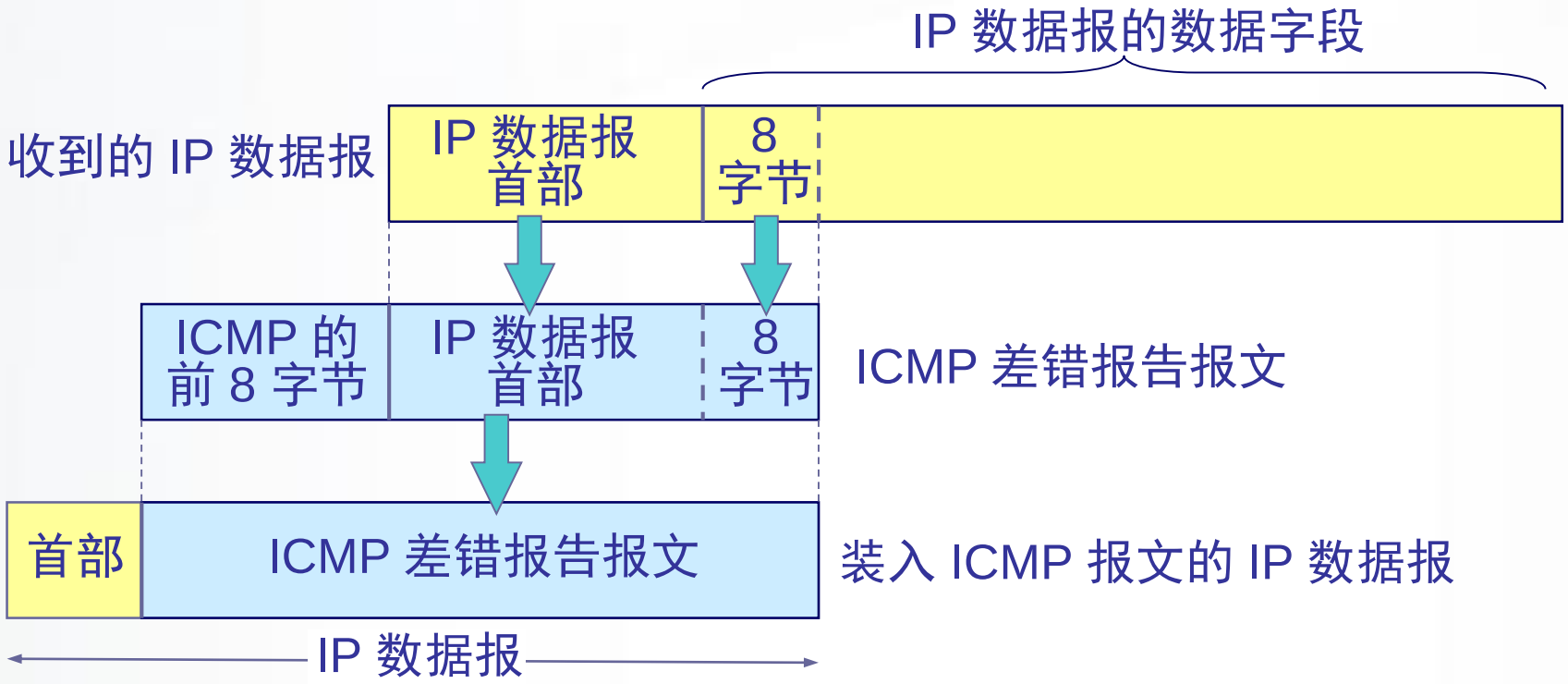
ICMP报文格式


前 4 个字节
都是一样的
0
8
16
31
类型
代码
检验和
（这 4 个字节取决于 ICMP 报文的类型）
ICMP 的数据部分（长度取决于类型）
ICMP 报文
首 部
数 据 部 分
IP 数据报

ICMP报文的类型

- ICMP 报文的种类有两种，即 ICMP 差错报告报文和 ICMP 询问报文。
- ICMP 报文的前 4 个字节是统一的格式，共有三个字段：即类型、代码和检验和。接着的 4 个字节的内容与 ICMP 的类型有关。
- 差错报告报文有五种：终点不可达 ，源点抑制(Source quench) ，时间超过 ，参数问题 ，改变路由（重定向）(Redirect)
- 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

差错报告报文的数据字段的内容



ICMP应用举例

- PING 用来测试两个主机之间的连通性。
- PING 使用了 ICMP 回送请求与回送回答报文。
- 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

```
C:\Documents and Settings\XXR>ping mail.sina.com.cn

Pinging mail.sina.com.cn [202.108.43.230] with 32 bytes of data:

Reply from 202.108.43.230: bytes=32 time=368ms TTL=242
Reply from 202.108.43.230: bytes=32 time=374ms TTL=242
Request timed out.
Reply from 202.108.43.230: bytes=32 time=374ms TTL=242

Ping statistics for 202.108.43.230:
    Packets: Sent = 4, Received = 3, Lost = 1 (25% loss),
Approximate round trip times in milli-seconds:
    Minimum = 368ms, Maximum = 374ms, Average = 372ms
```

指引

- 网络层提供的两种服务
- 网际协议 IP
- 划分子网和构造超网
  - 划分子网
  - 无分址编址
  - 构造超网
- 网际控制报文协议 ICMP
- 因特网的路由选择协议
- IPv6
- IP 多播
- 虚拟专用网 VPN 和网络地址转换 NAT

因特网的路由选择协议

- 根据路由算法的自适应性划分：
  - 静态路由选择策略
  - 动态路由选择策略

动手实验

实验4-3：配置网络地址

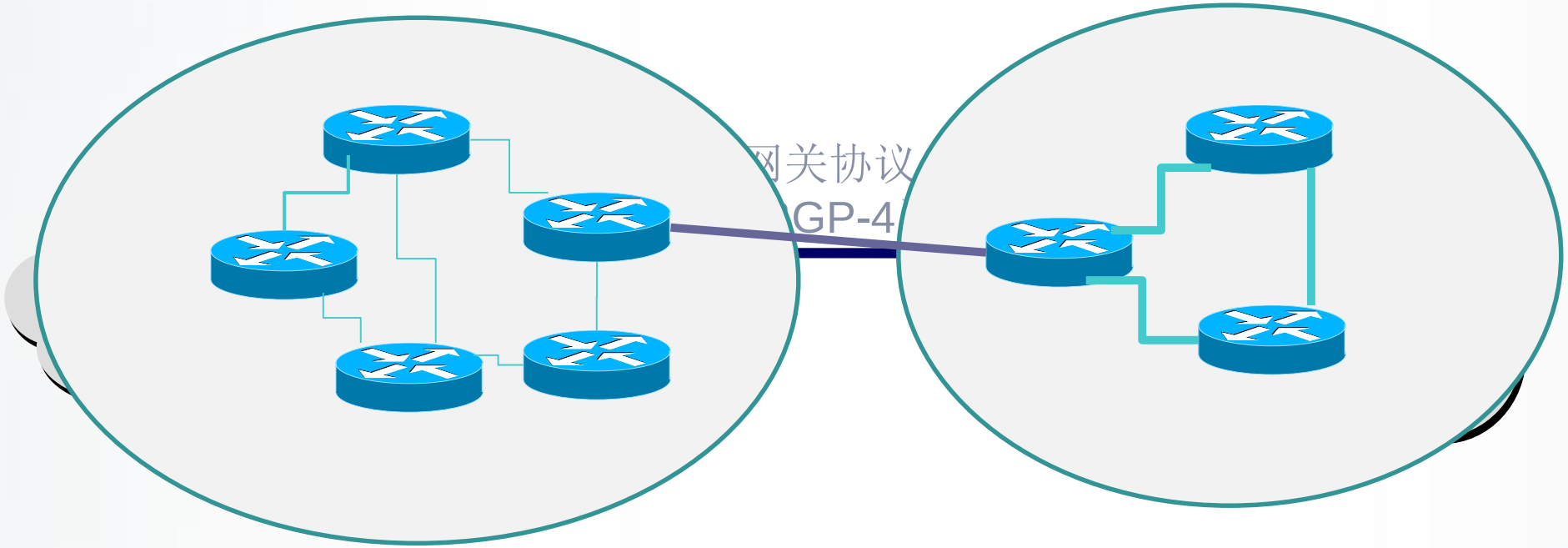
实验4-4：静态路由

实验4-5：配置默认路由

实验4-6：路由汇总

分层次的路由选择协议

自治系统AS、
内部网关协议IGP、外部网关协议 EGP



内部网关协议 RIP

- 工作原理
  - RIP 是一种分布式的基于距离向量的路由选择协议。

N1 1 直接
N2 1 直接
N3 2 R2

N1 2 R1
N2 1 直接
N3 1 直接

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。

RIP算法举例

表4-9(a) R_6路由表

目的网络	距离	下一跳路由器
Net2	3	R_4
Net3	4	R_5
…	…	…

表4-9(b) R_4发来的路由信息表

目的网络	距离	下一跳路由器
Net1	3	R_1
Net2	4	R_2
Net3	1	直接交付

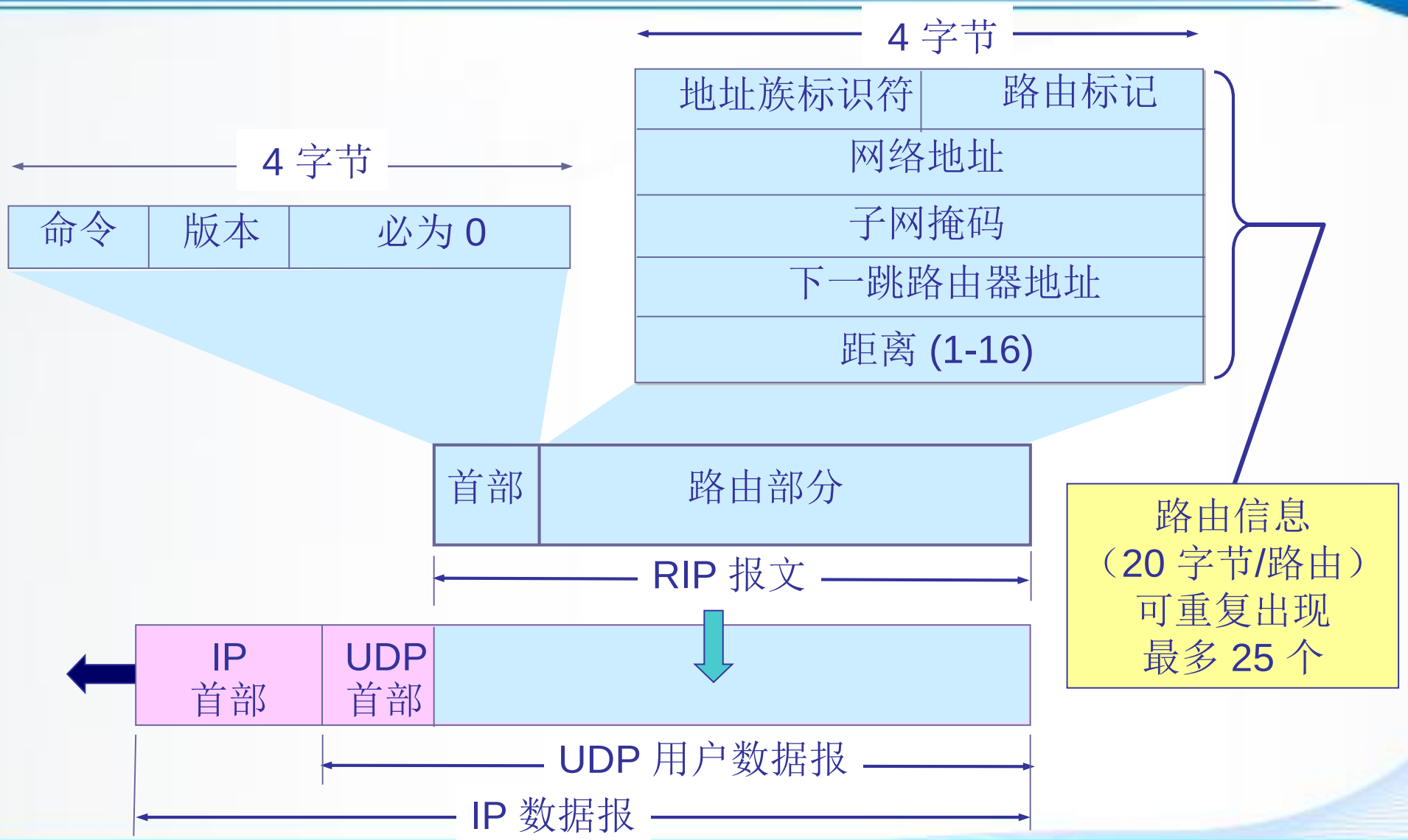
表4-9(c) 修改后的表4-9(b)

目的网络	距离	下一跳路由器

表4-9(d) R_6更新后的路由表

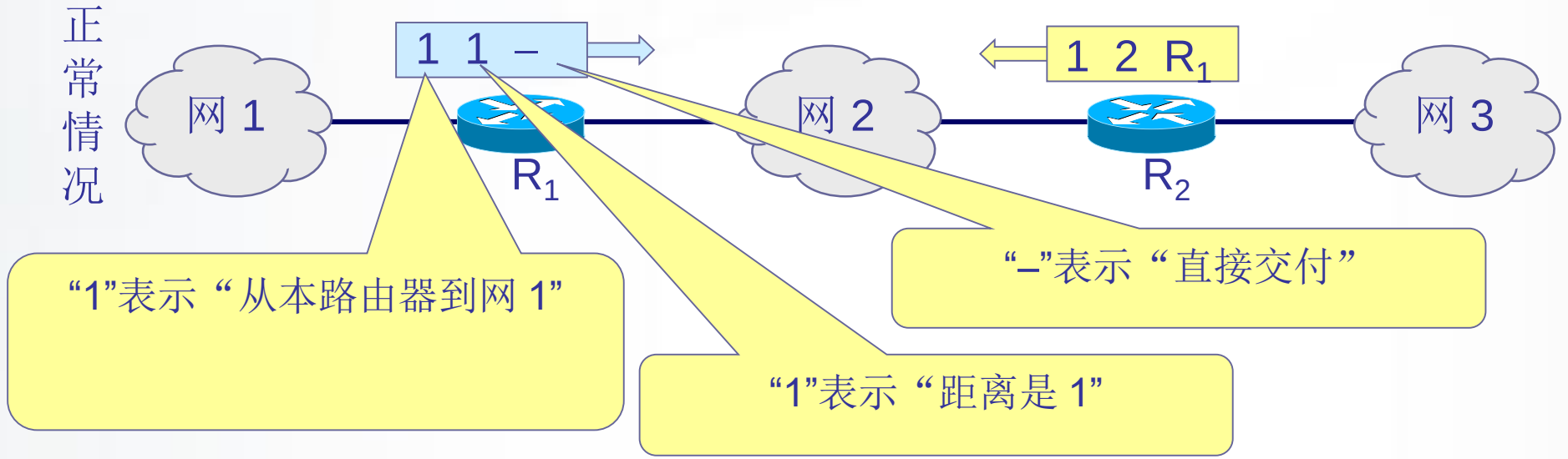
目的网络	距离	下一跳路由器

RIP2 协议的报文格式



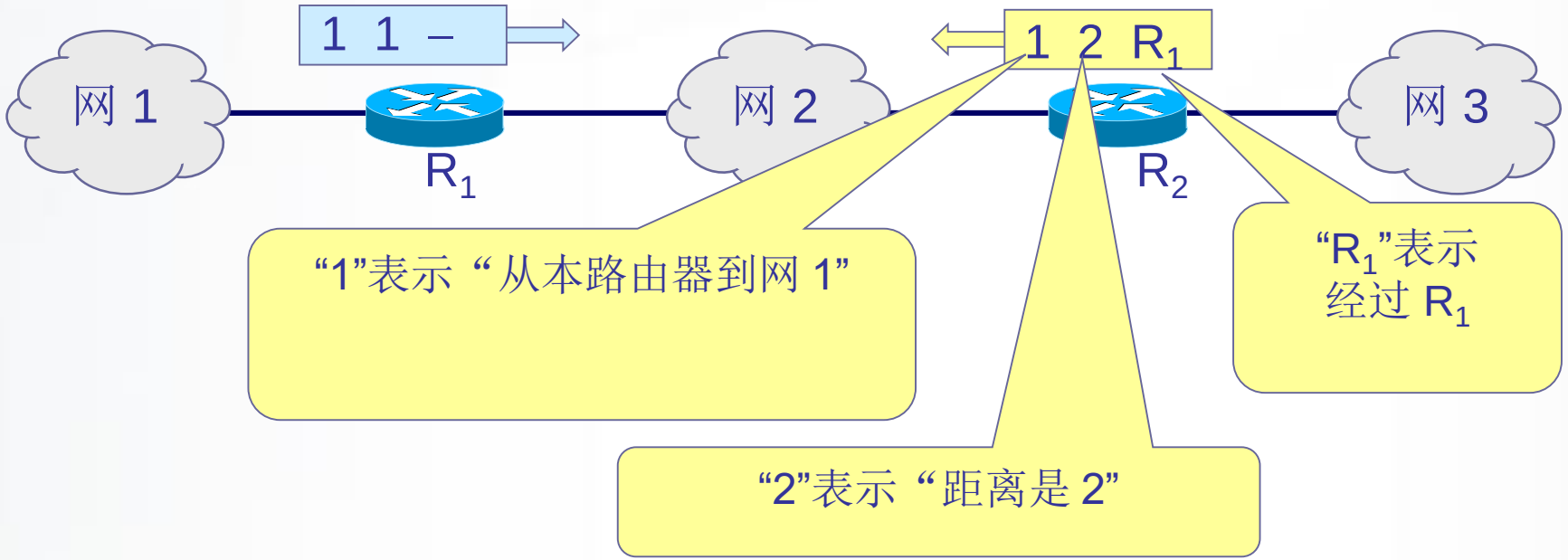
RIP 协议的优缺点

- RIP 存在的一个问题是当网络出现故障时，要经过比较长的时间才能将此信息传送到所有的路由器。
- RIP 协议最大的优点就是实现简单，开销较小。
- RIP 限制了网络的规模，它能使用的最大距离为 15（16 表示不可达）。
- 路由器之间交换的路由信息是路由器中的完整路由表，因而随着网络规模的扩大，开销也就增加。

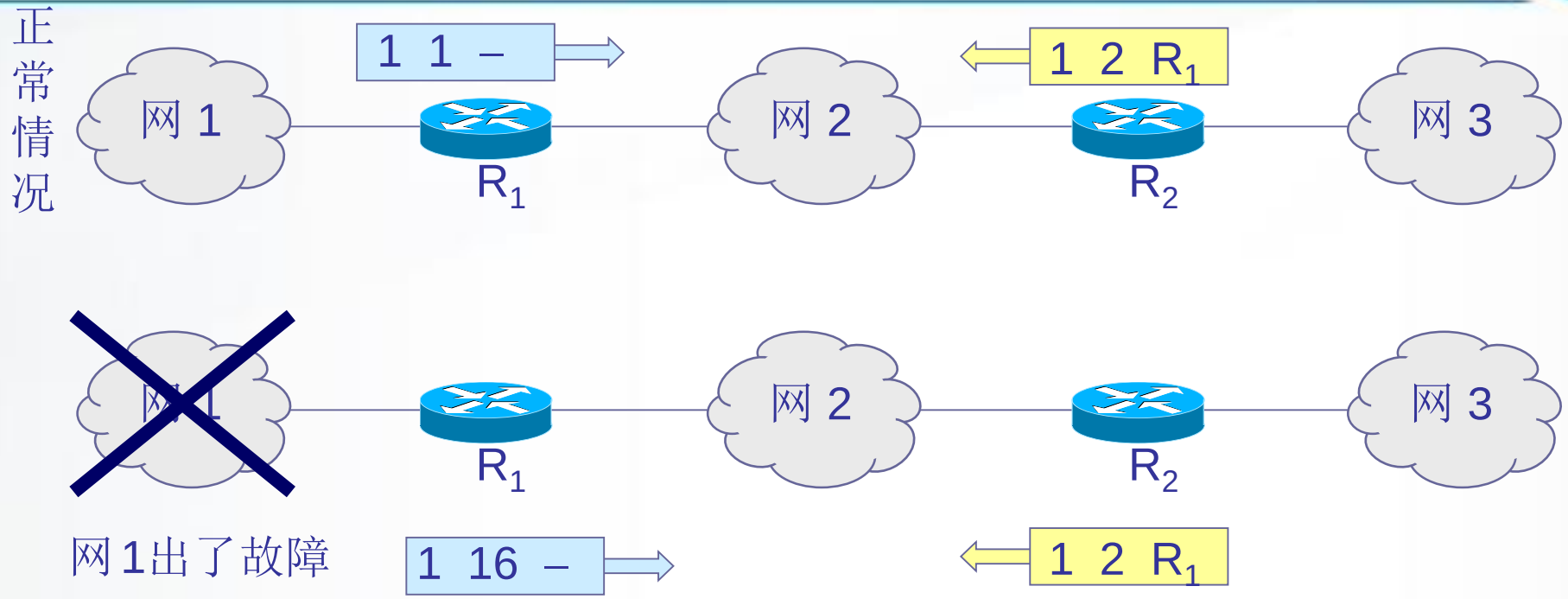


R_1 说：“我到网 1 的距离是 1，是直接交付。”

正常情况



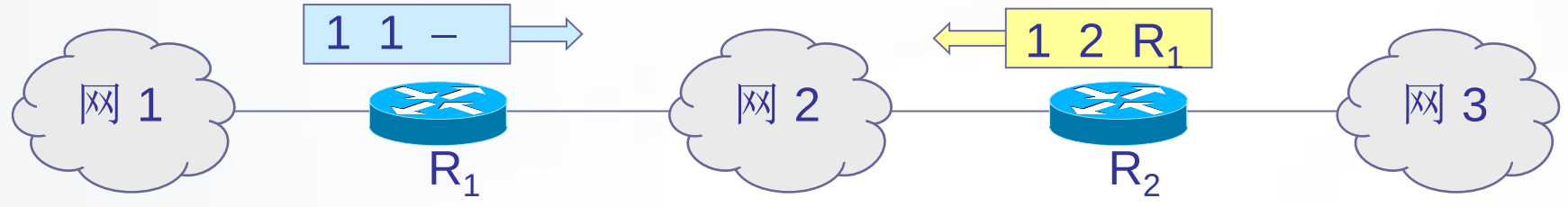
R_2 说：“我到网 1 的距离是 2，是经过 R_1。”

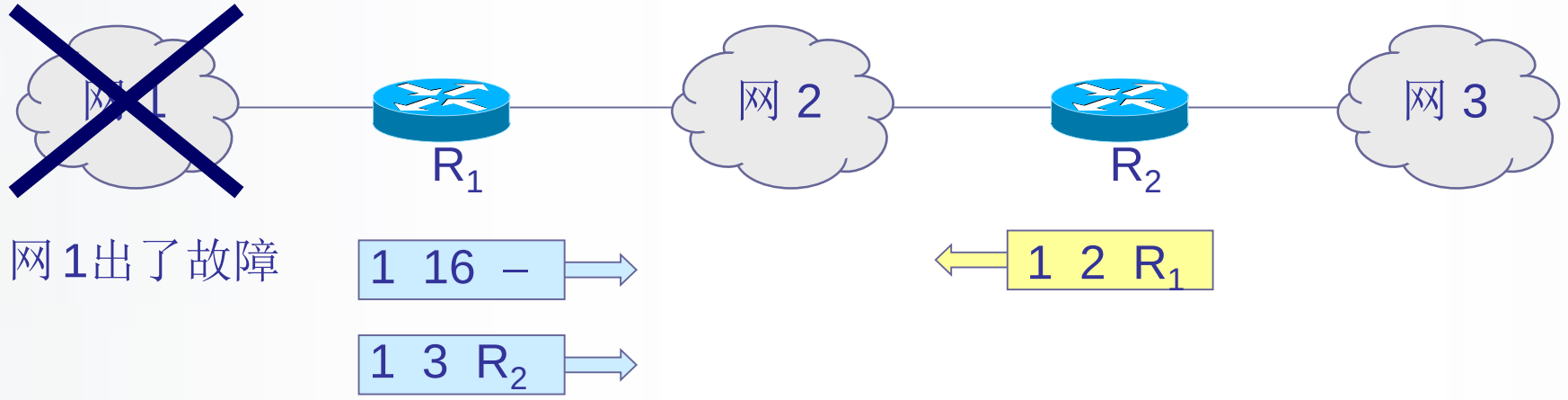


R_1 说：“我到网 1 的距离是 16 （表示无法到达），是直接交付。”

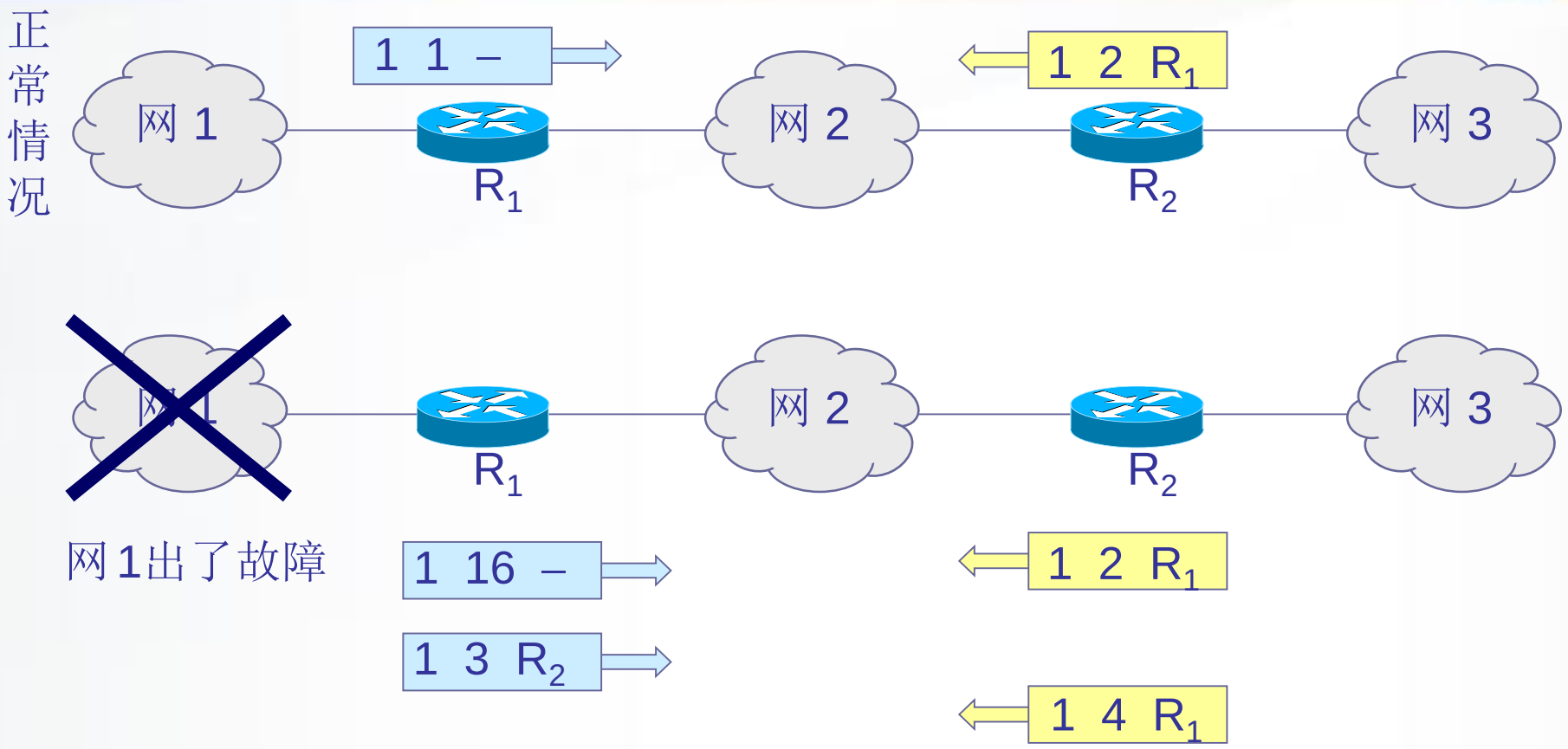
但 R_2 在收到 R_1 的更新报文之前，还发送原来的报文，因为这时 R_2 并不知道 R_1 出了故障。

正常情况





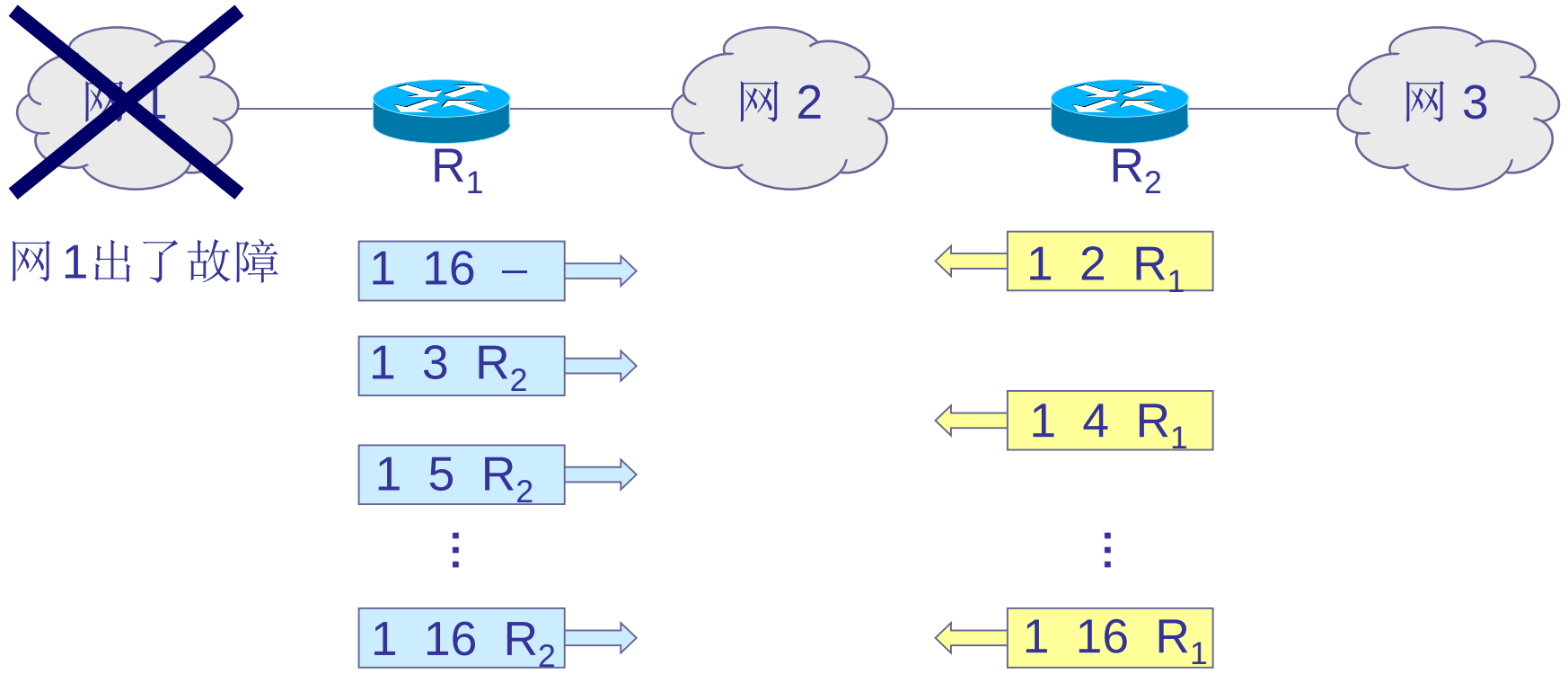
R_1 收到 R_2 的更新报文后，误认为可经过 R_2 到达网1，于是更新自己的路由表，说：“我到网 1 的距离是 3，下一跳经过 R_2”。然后将此更新信息发送给 R_2。



R_2 以后又更新自己的路由表为“1, 4, R_1”，表明 “我到网 1 距离是 4，下一跳经过 R_1”。

这就是好消息传播得快，而坏消息传播得慢。网络出故障的传播时间往往需要较长的时间(例如数分钟)。这是 RIP 的一个主要缺点。

情况



这样不断更新下去，直到 R_1 和 R_2 到网 1 的距离都增大到 16 时，R_1 和 R_2 才知道网 1 是不可达的。

动手实验

- 实验4-8：动态路由RIP

内部网关协议 OSPF

- 1. 向本自治系统中所有路由器发送信息，这里使用的方法是洪泛法。
- 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，但这只是路由器所知道的部分信息。
- 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。

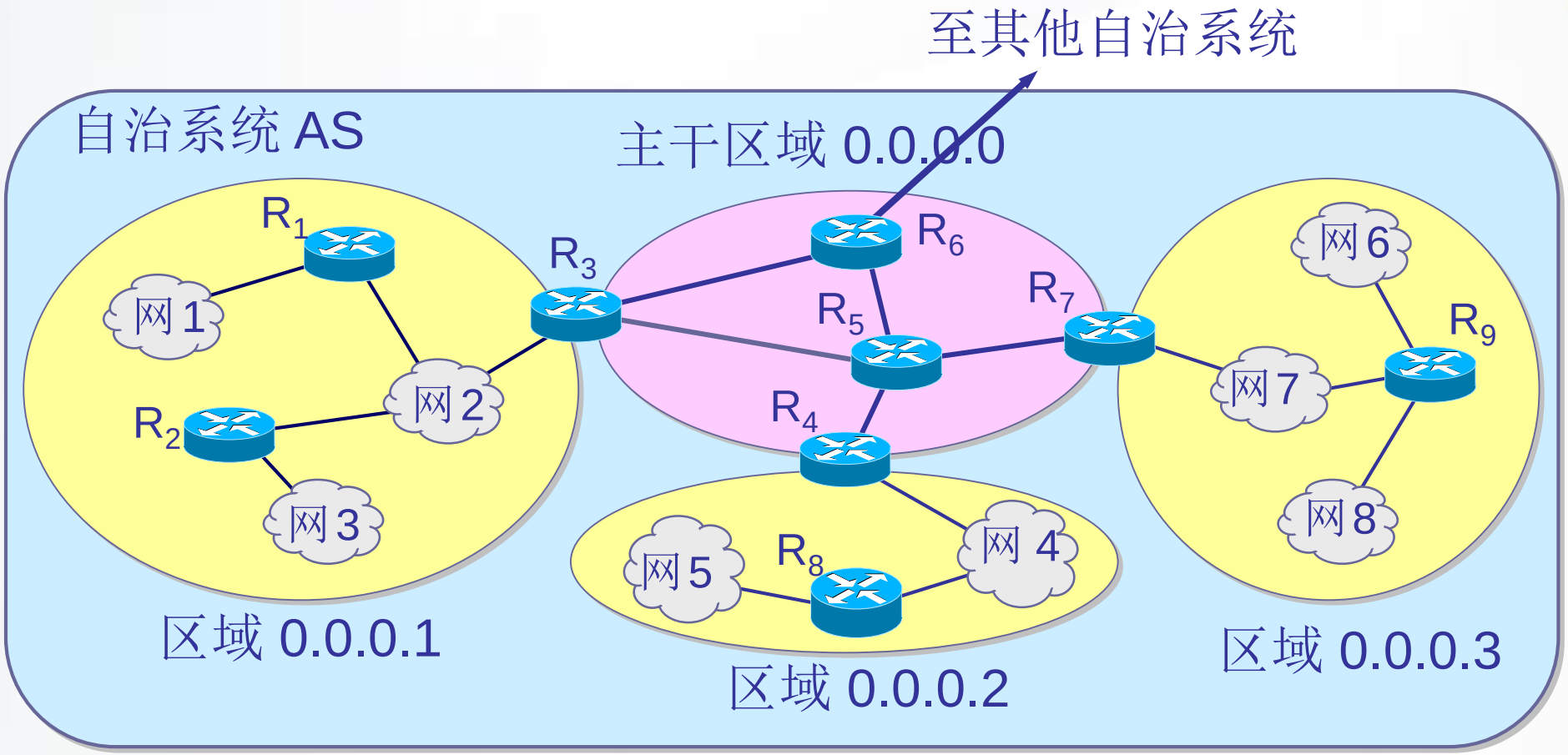
OSPF中每个路由器维护三部分内容

- 邻居状态
- 链路状态数据库

(link-state database)

- 路由表

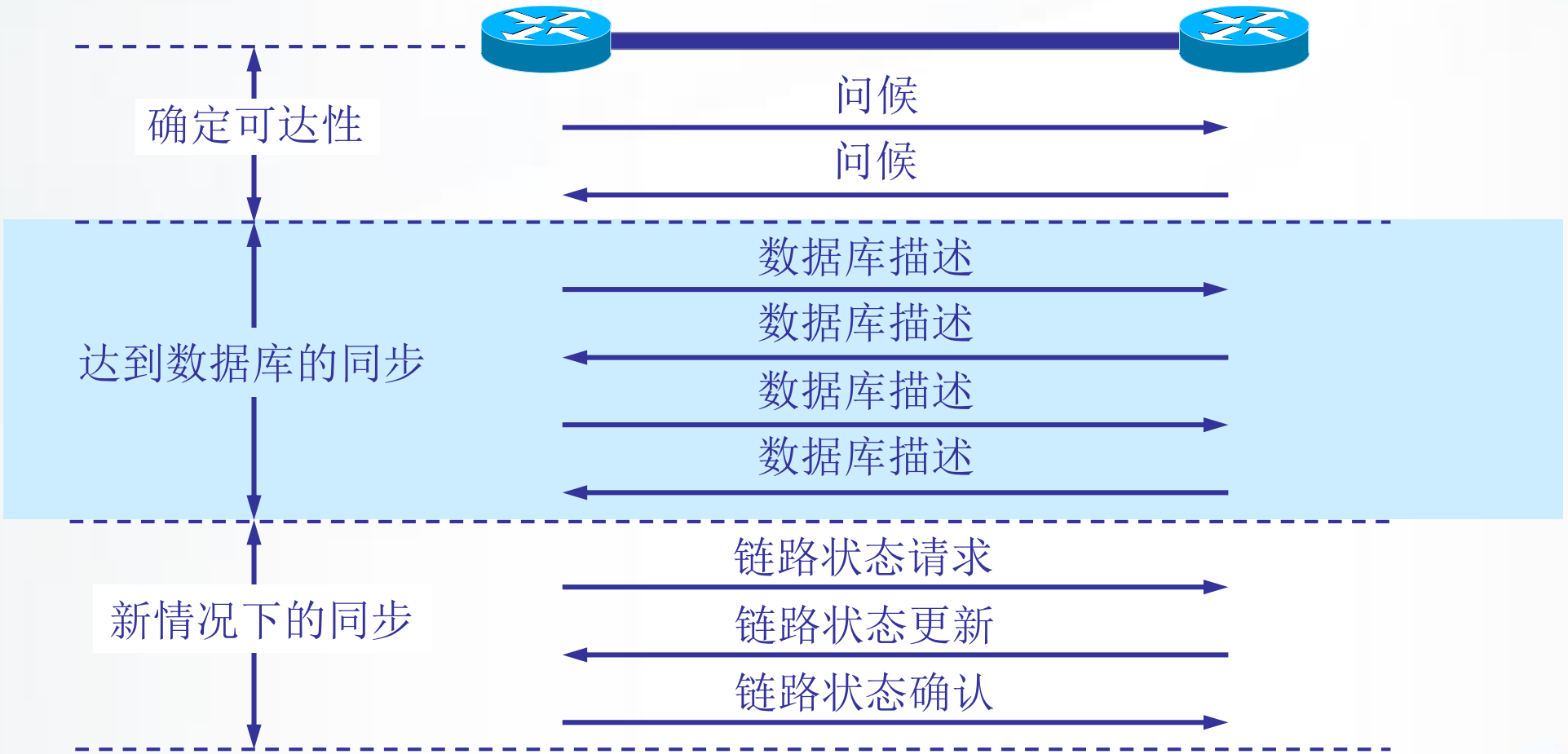
OSPF 划分为两种不同的区域



OSPF 的其他特点

- OSPF 不用 UDP 而是直接用 IP 数据报传送
- OSPF 对不同的链路可根据 IP 分组的不同服务类型而设置成不同的代价。因此，OSPF 对于不同类型的业务可计算出不同的路由。
- 如果到同一个目的网络有多条相同代价的路径，那么可以将通信量分配给这几条路径。这叫作多路径间的负载平衡。
- 所有在 OSPF 路由器之间交换的分组都具有鉴别的功能。
- 支持可变长度的子网划分和无分类编址 CIDR。
- 每一个链路状态都带上一个 32 位的序号，序号越大状态就越新。

OSPF的基本操作


确定可达性
问候
问候
达到数据库的同步
数据库描述
数据库描述
数据库描述
数据库描述
新情况下的同步
链路状态请求
链路状态更新
链路状态确认

指定的路由器(designated router)

- 多点接入的局域网采用了指定的路由器的方法，使广播的信息量大大减少。
- 指定的路由器代表该局域网上所有的链路向连接到该网络上的各路由器发送状态信息。

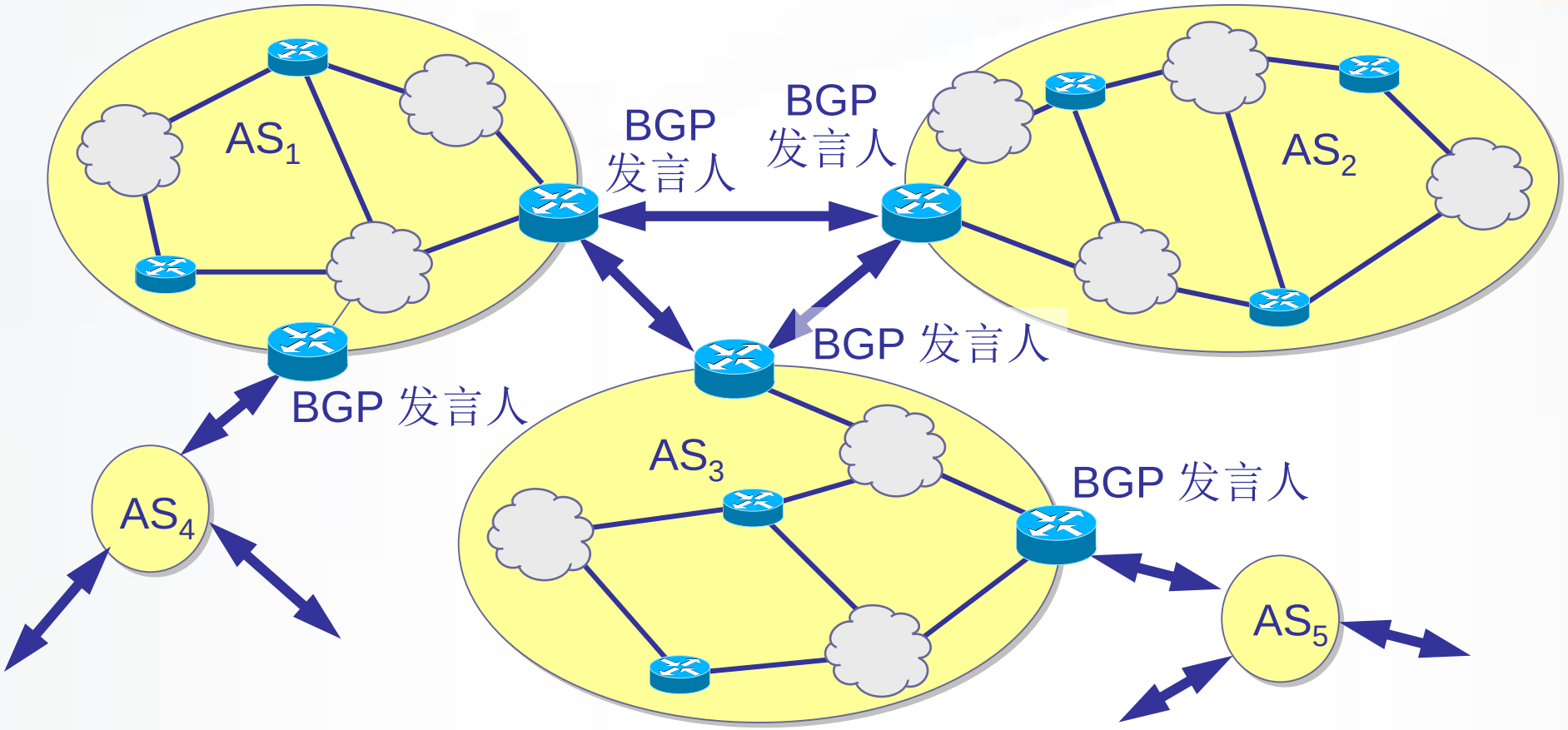
动手实验

- 实验4-9：OSPF

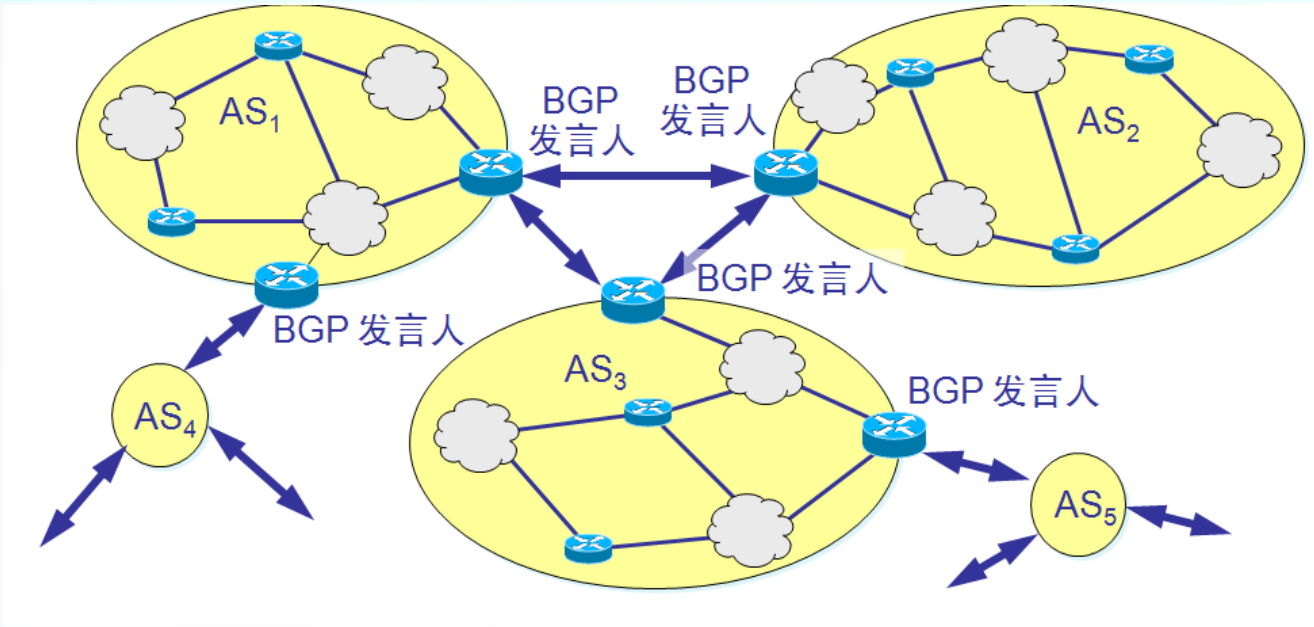
外部网关协议 BGP

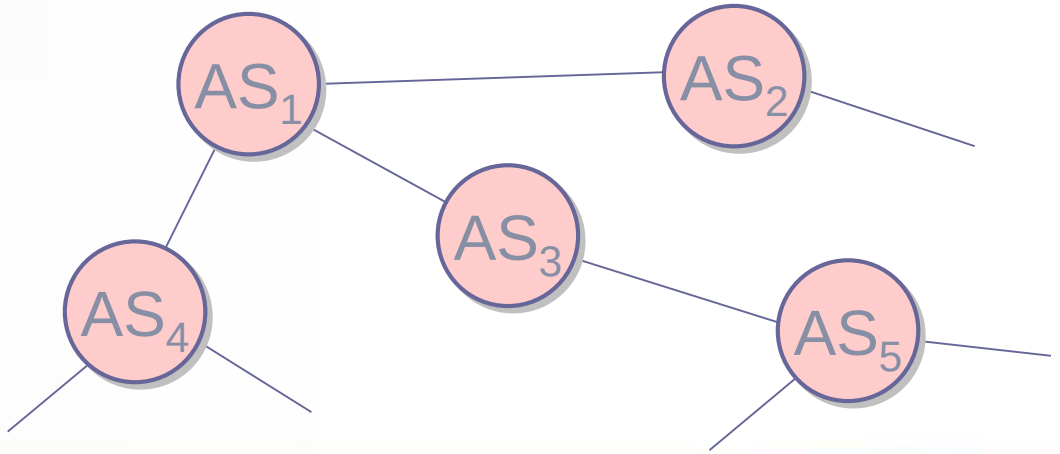
- BGP 是不同自治系统的路由器之间交换路由信息的协议，BGP 较新版本是BGP-4。
- 边界网关协议 BGP 只能是力求寻找一条能够到达目的网络且比较好的路由（不能兜圈子），而并非要寻找一条最佳路由。
- 每一个自治系统的管理员要选择至少一个路由器作为该自治系统的“ BGP 发言人”

BGP 发言人和自治系统 AS 的关系

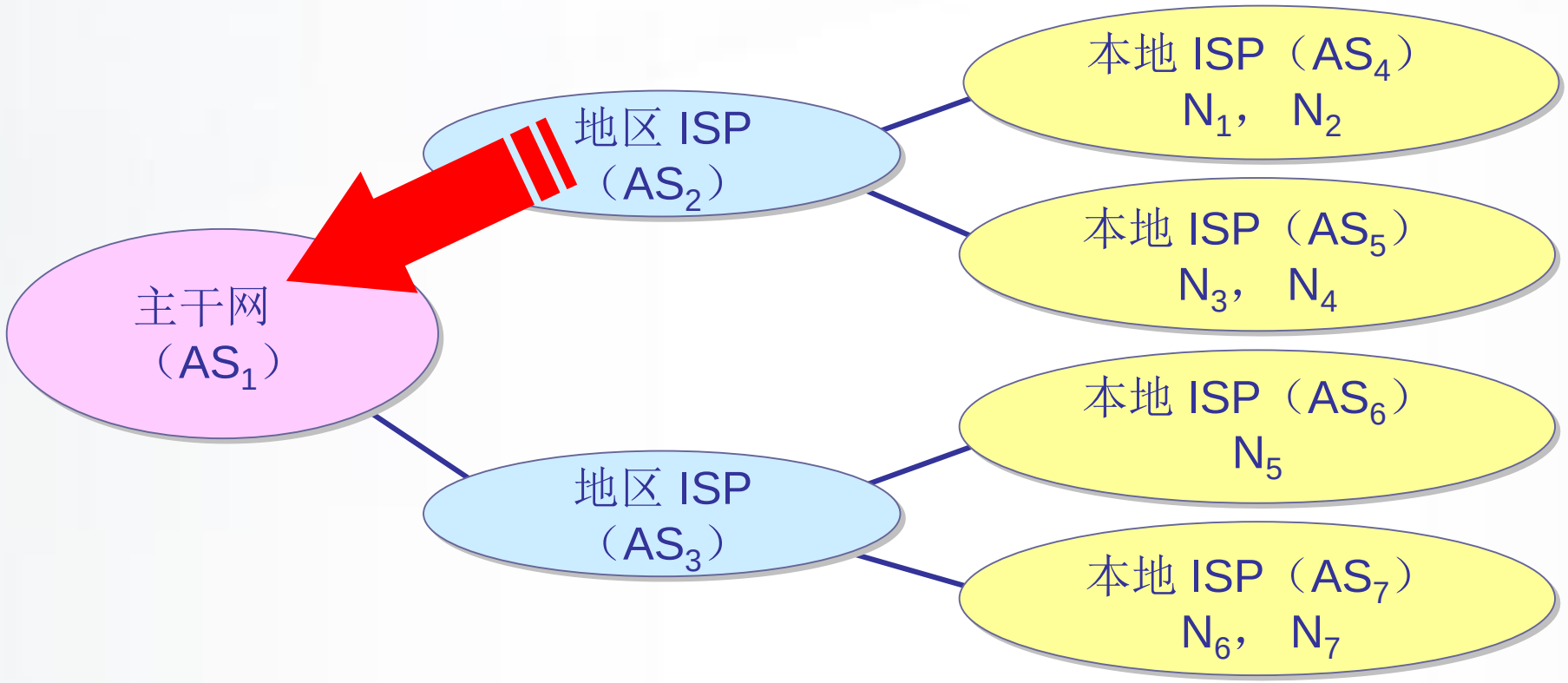


AS 的连通图举例


AS1
AS2
AS3
AS4
AS5
BGP
发言人
BGP
发言人
BGP 发言人
BGP 发言人
BGP 发言人


AS1
AS2
AS3
AS4
AS5

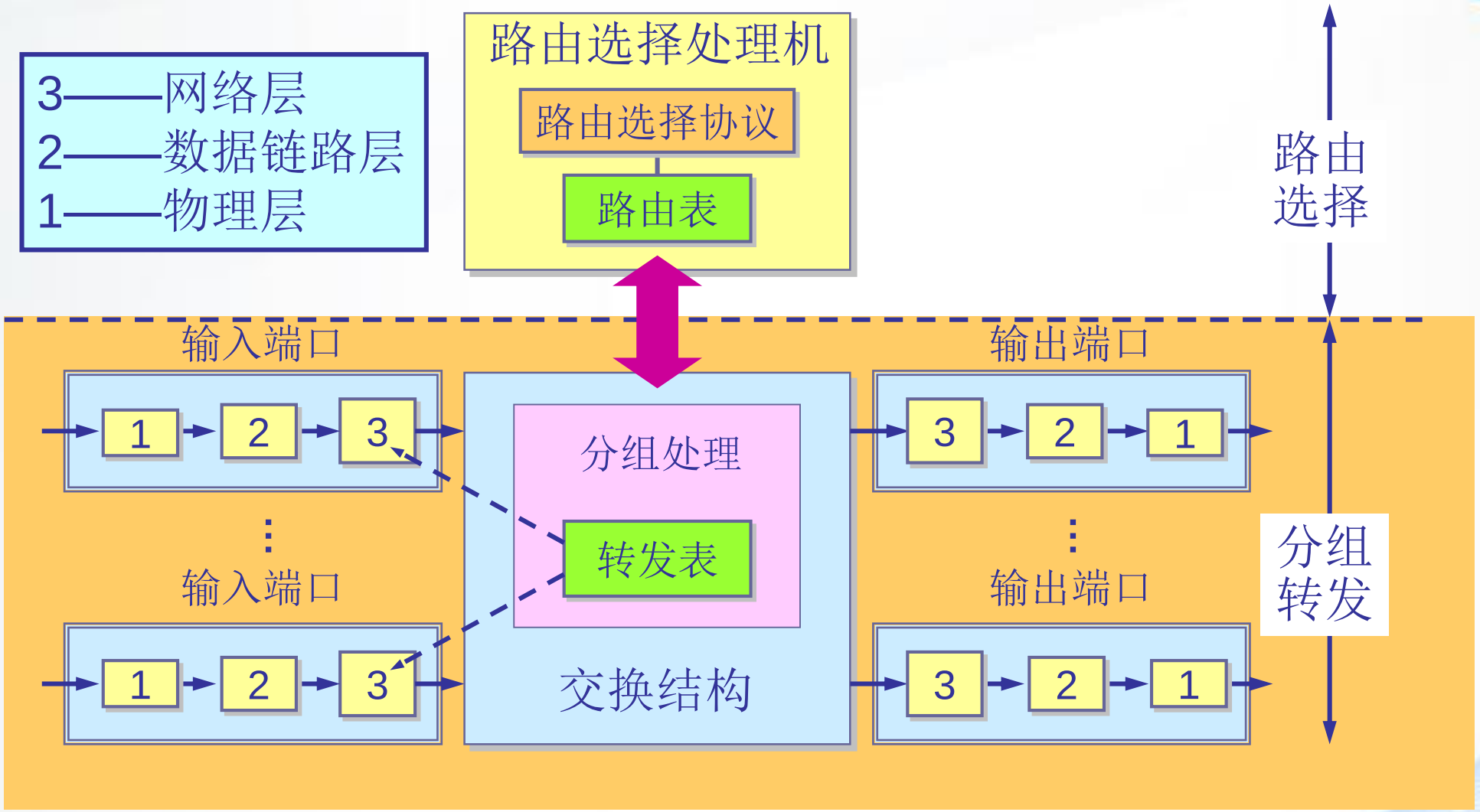
BGP 发言人交换路径向量



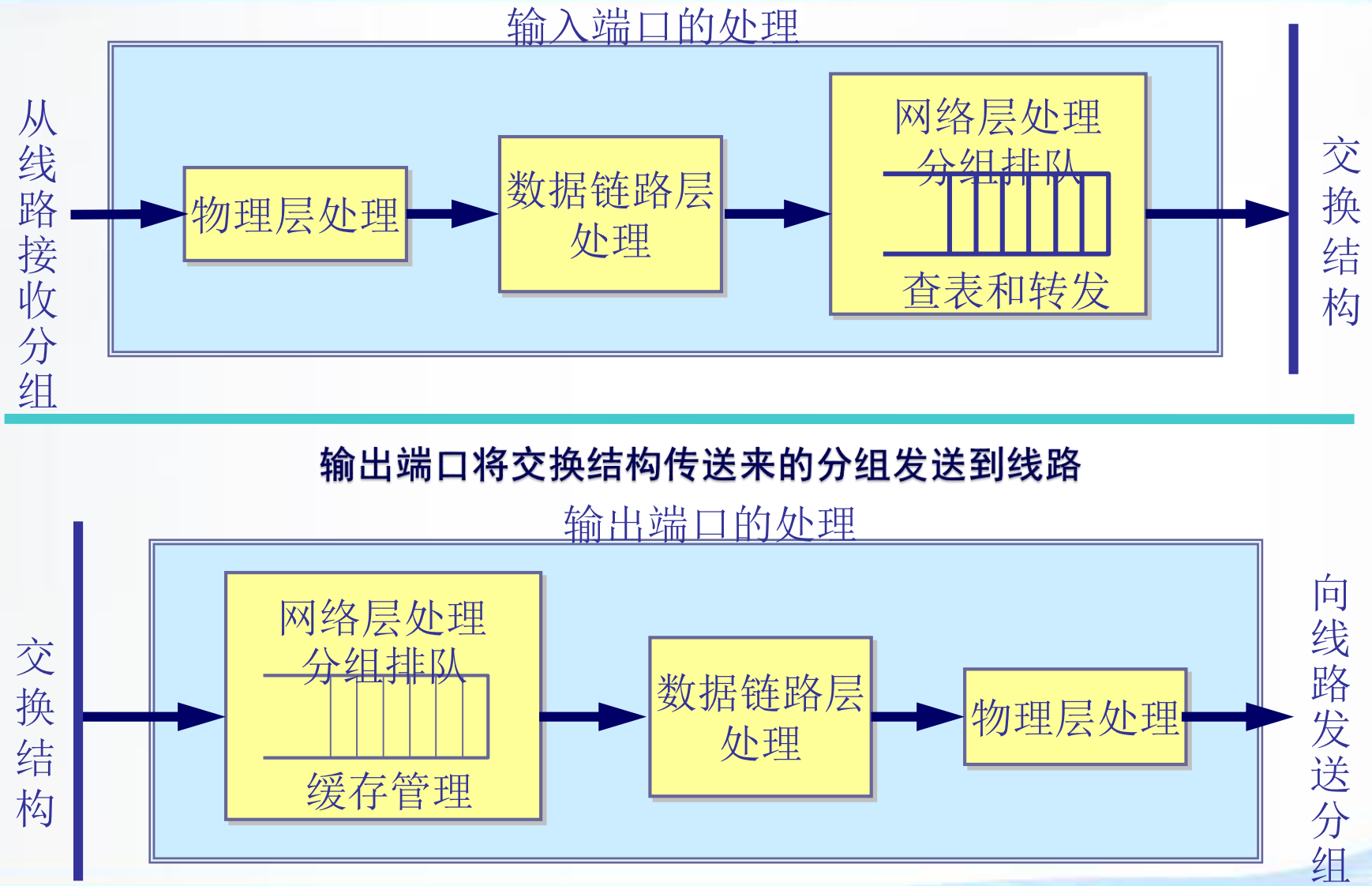
BGP 协议的特点

- BGP协议交换路由信息数量不是很多
- BGP发言人数目不多，路由选择相对简单
- BGP协议支持CIDR
- BGP建立时，交换整个路由表但之后只交换变化部分

典型的路由器的结构


3——网络层
2——数据链路层
1——物理层
路由选择处理机
路由选择协议
路由表
路由选择
输入端口
输出端口
1
2
3
分组处理
转发表
交换结构
分组转发

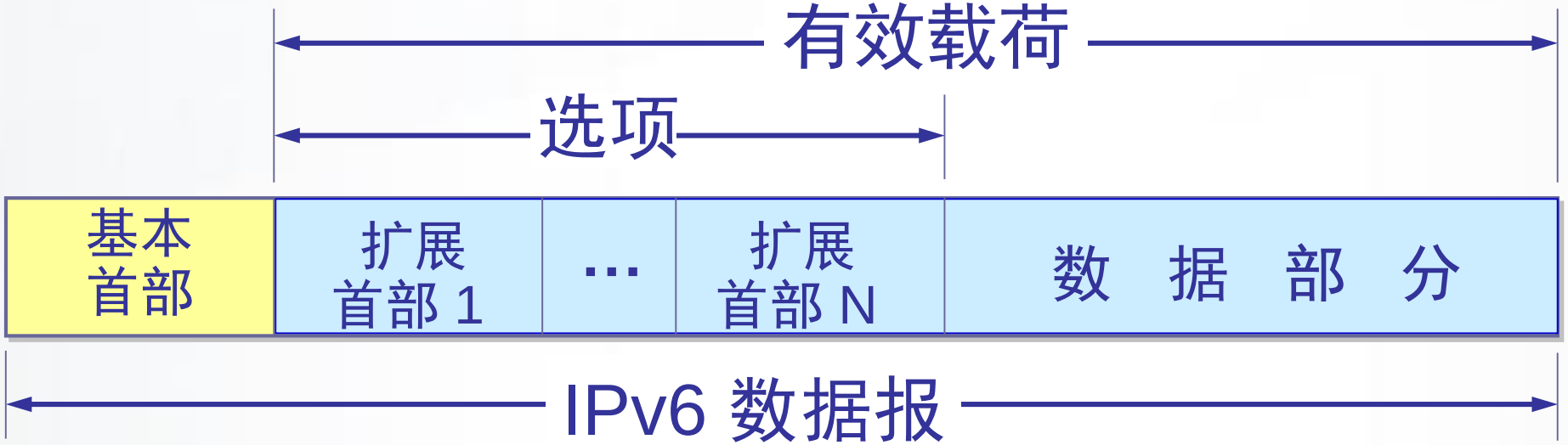
输入端口对线路上收到的分组的处理

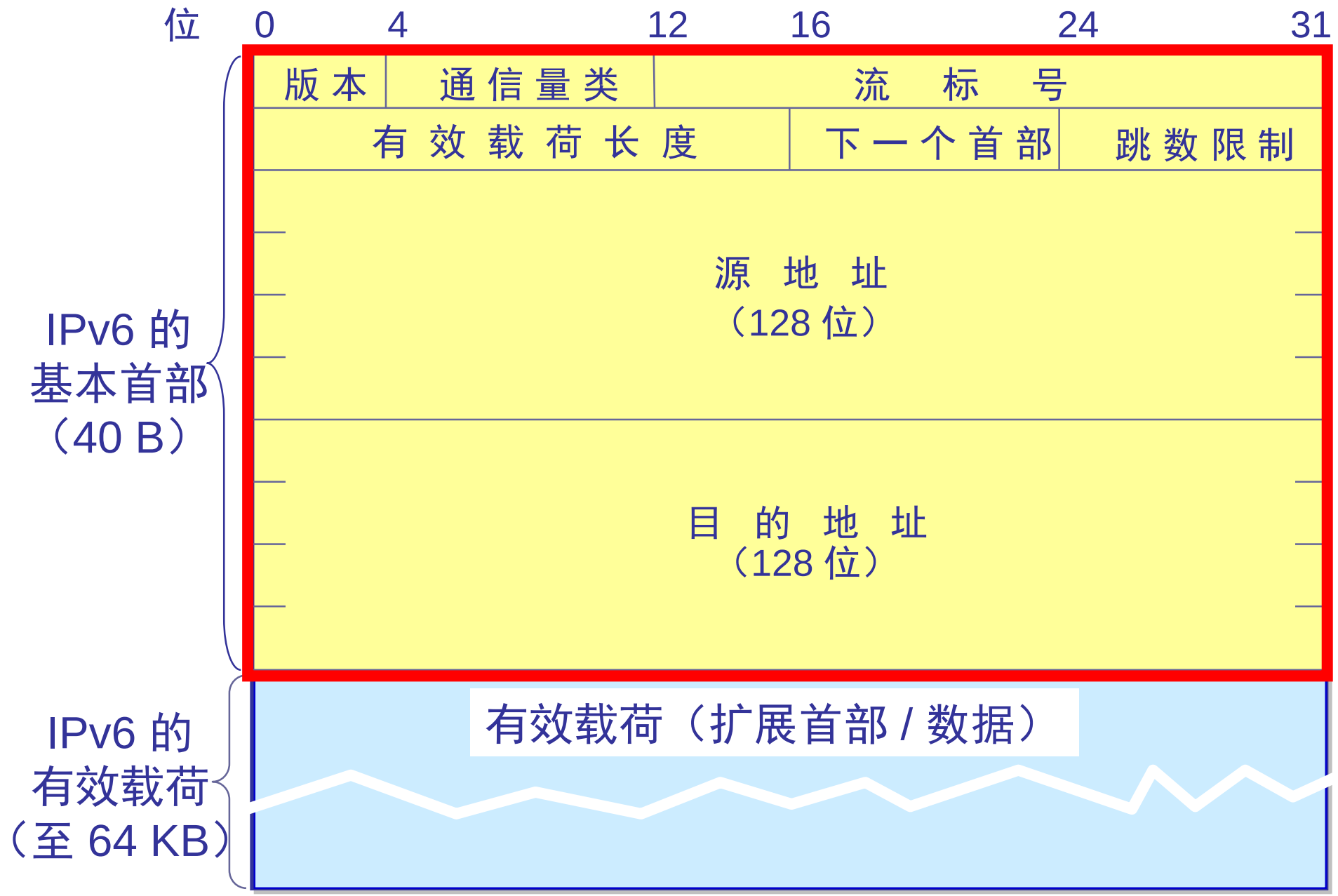


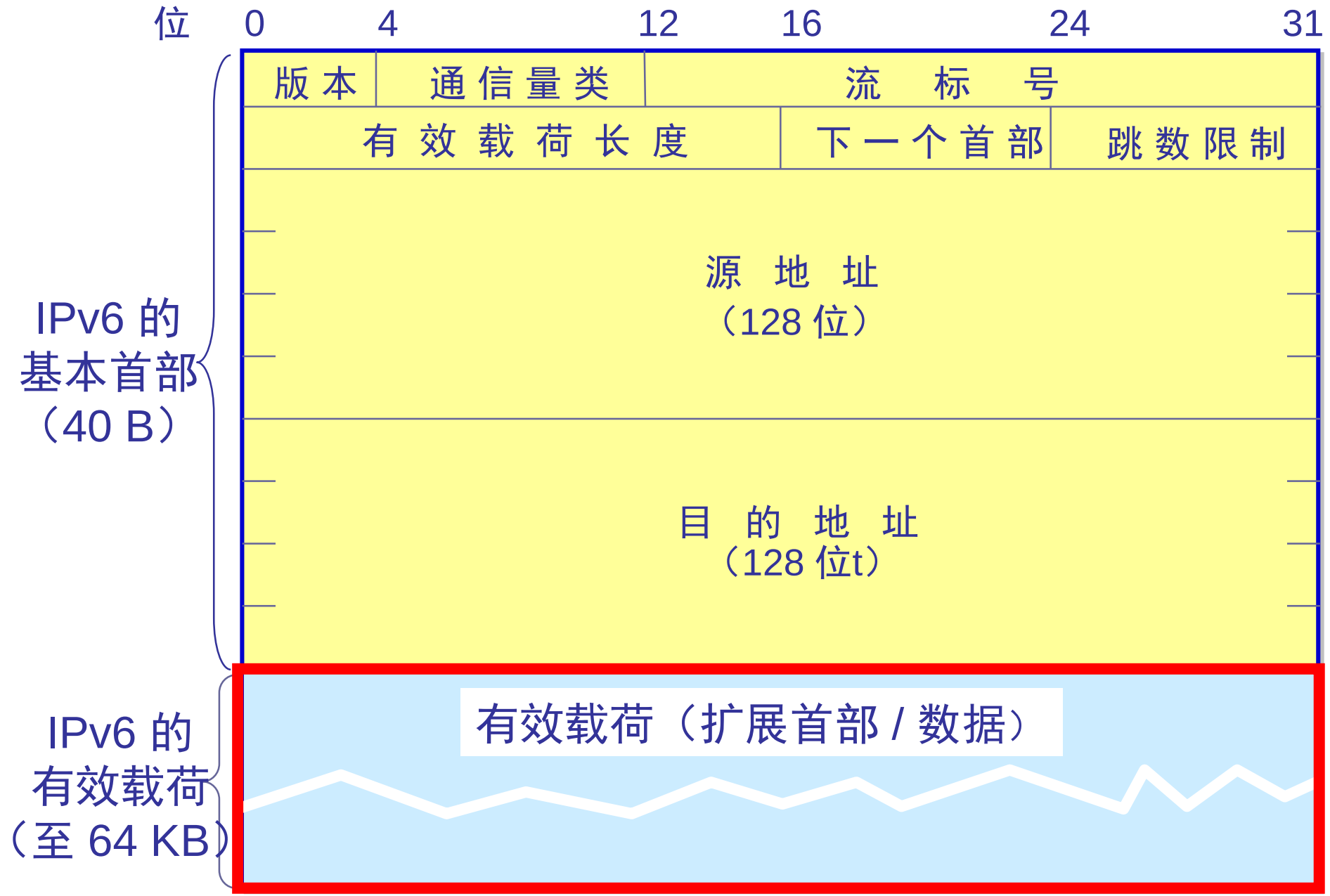
指引

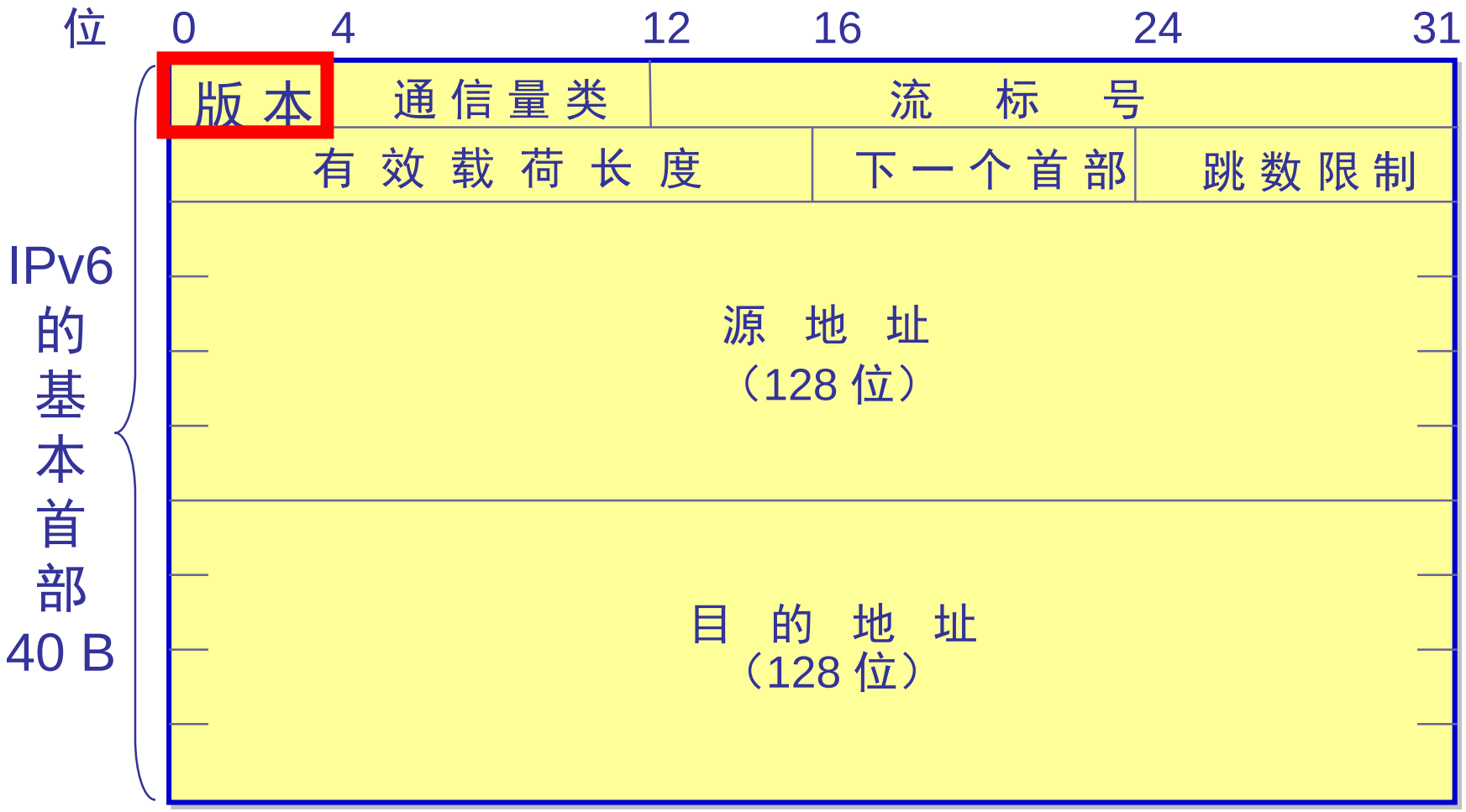
- 网络层提供的两种服务
- 网际协议 IP
- 划分子网和构造超网
  - 划分子网
  - 无分址编址
  - 构造超网
- 网际控制报文协议 ICMP
- 因特网的路由选择协议
- IPv6
- IP 多播
- 虚拟专用网 VPN 和网络地址转换 NAT

IPv6 数据报的一般形式

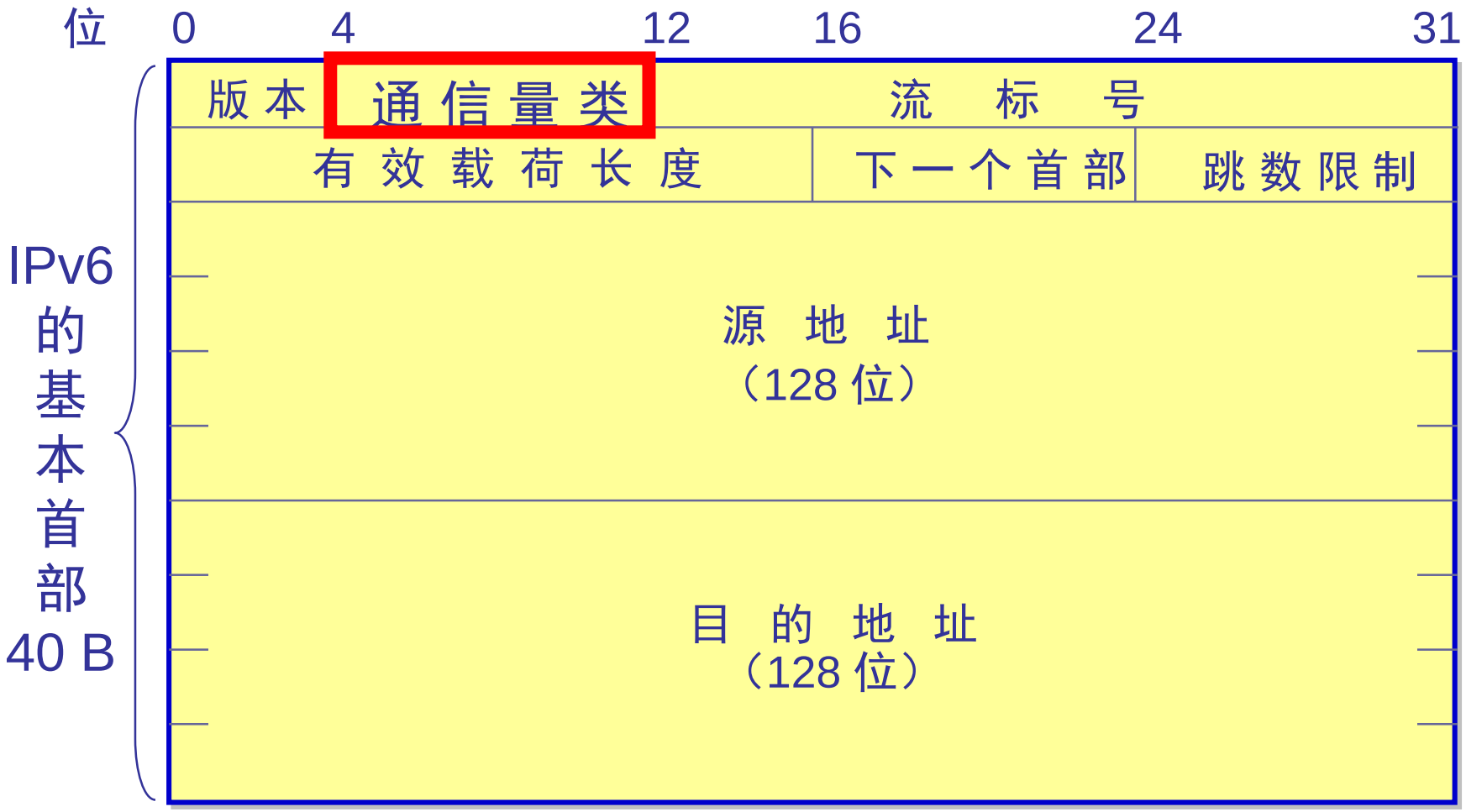



位
0
4
12
16
24
31
版本
通信量类
流 标 号
有效载荷长度
下一个首部
跳数限制
源 地 址
（128 位）
目 的 地 址
（128 位）
IPv6 的
基本首部
（40 B）
有效载荷（扩展首部 / 数据）
IPv6 的
有效载荷
（至 64 KB）

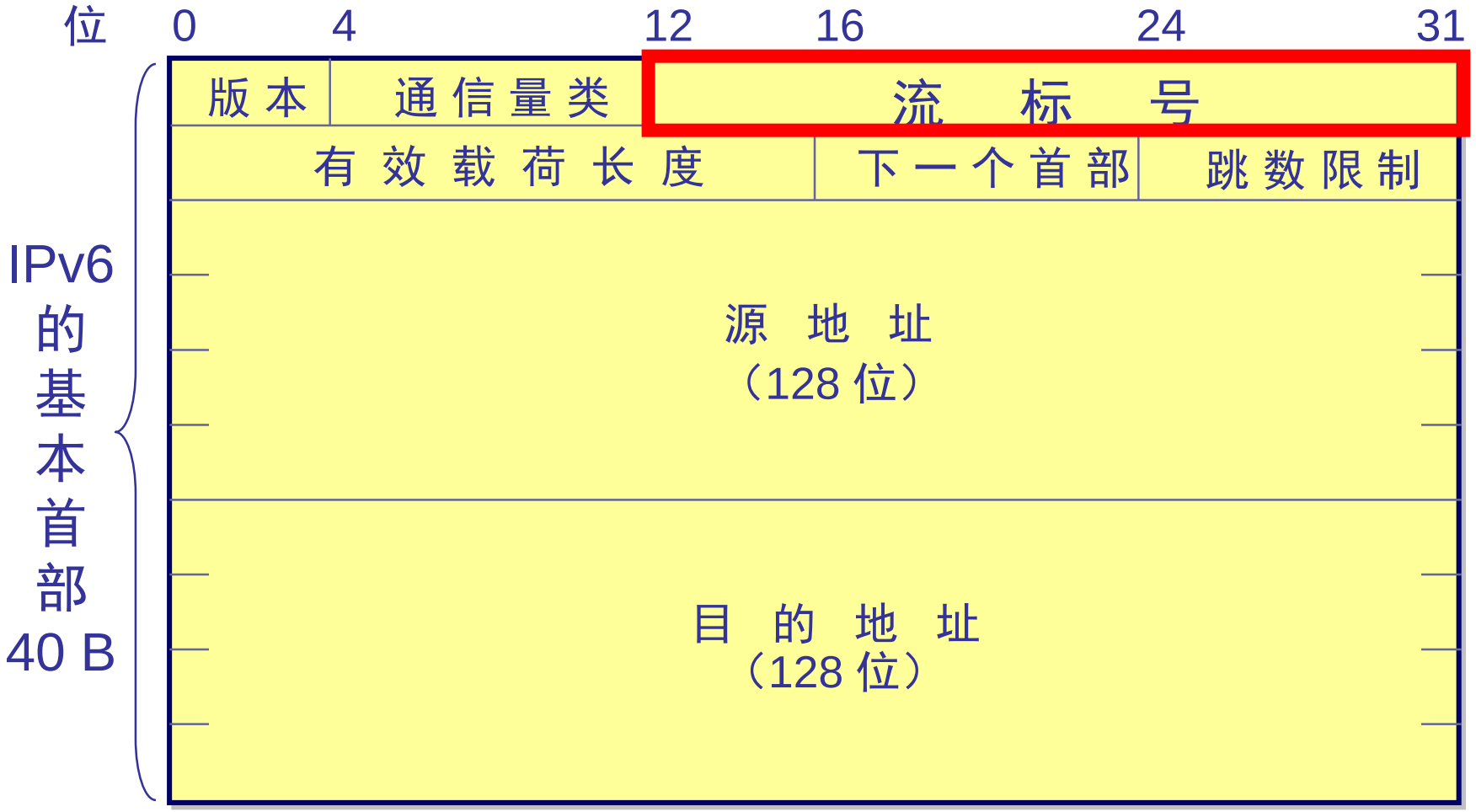

位
0
4
12
16
24
31
版本
通信量类
流 标 号
有效载荷长度
下一个首部
跳数限制
源 地 址
（128 位）
目 的 地 址
（128 位t）
IPv6 的
基本首部
（40 B）
有效载荷（扩展首部 / 数据）
IPv6 的
有效载荷
（至 64 KB）



版本(version)—— 4 位。它指明了协议的版本，对 IPv6 该字段总是 6。

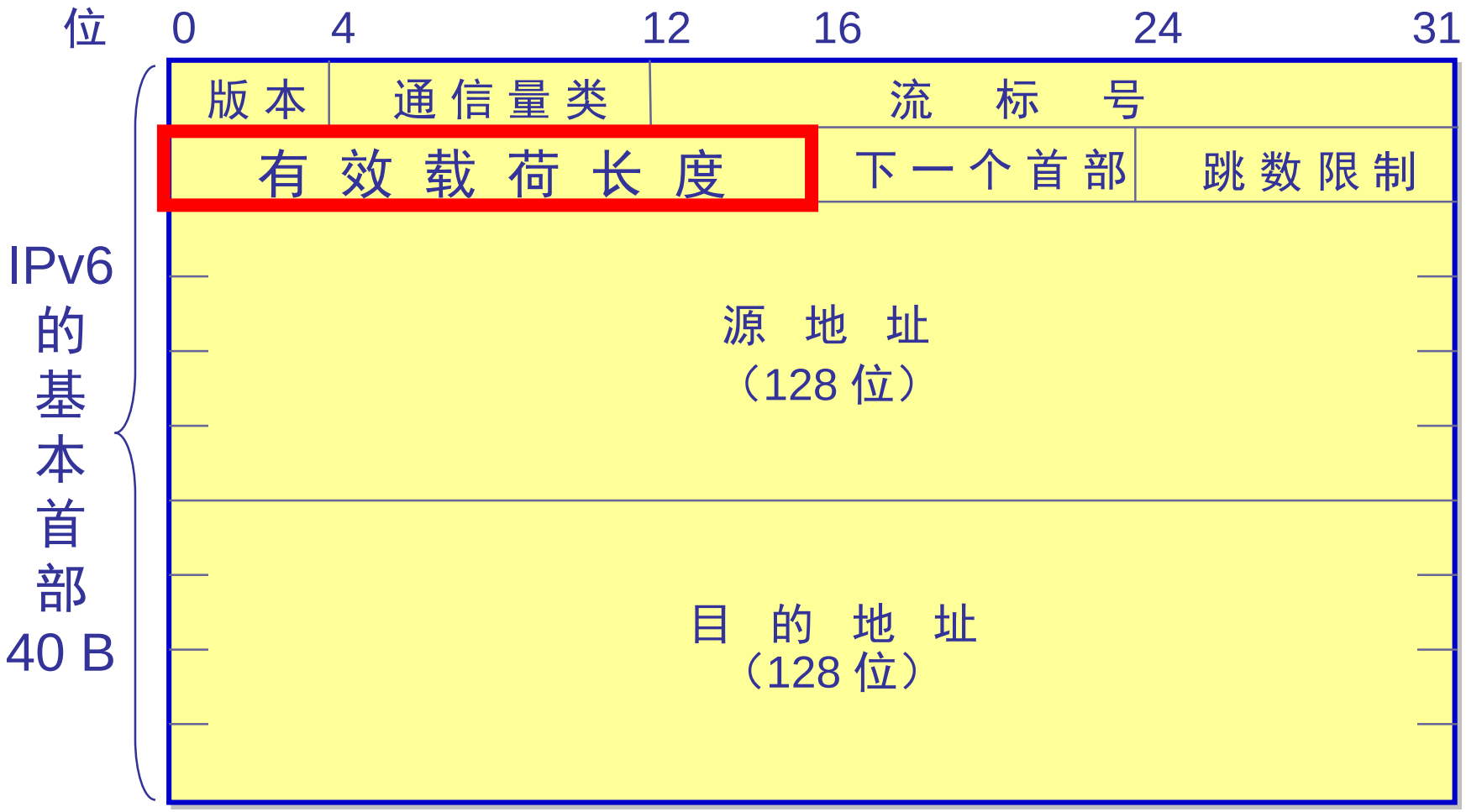


通信量类(traffic class)—— 8 位。这是为了区分不同的 IPv6 数据报的类别或优先级。目前正在进行不同的通信量类性能的实验。

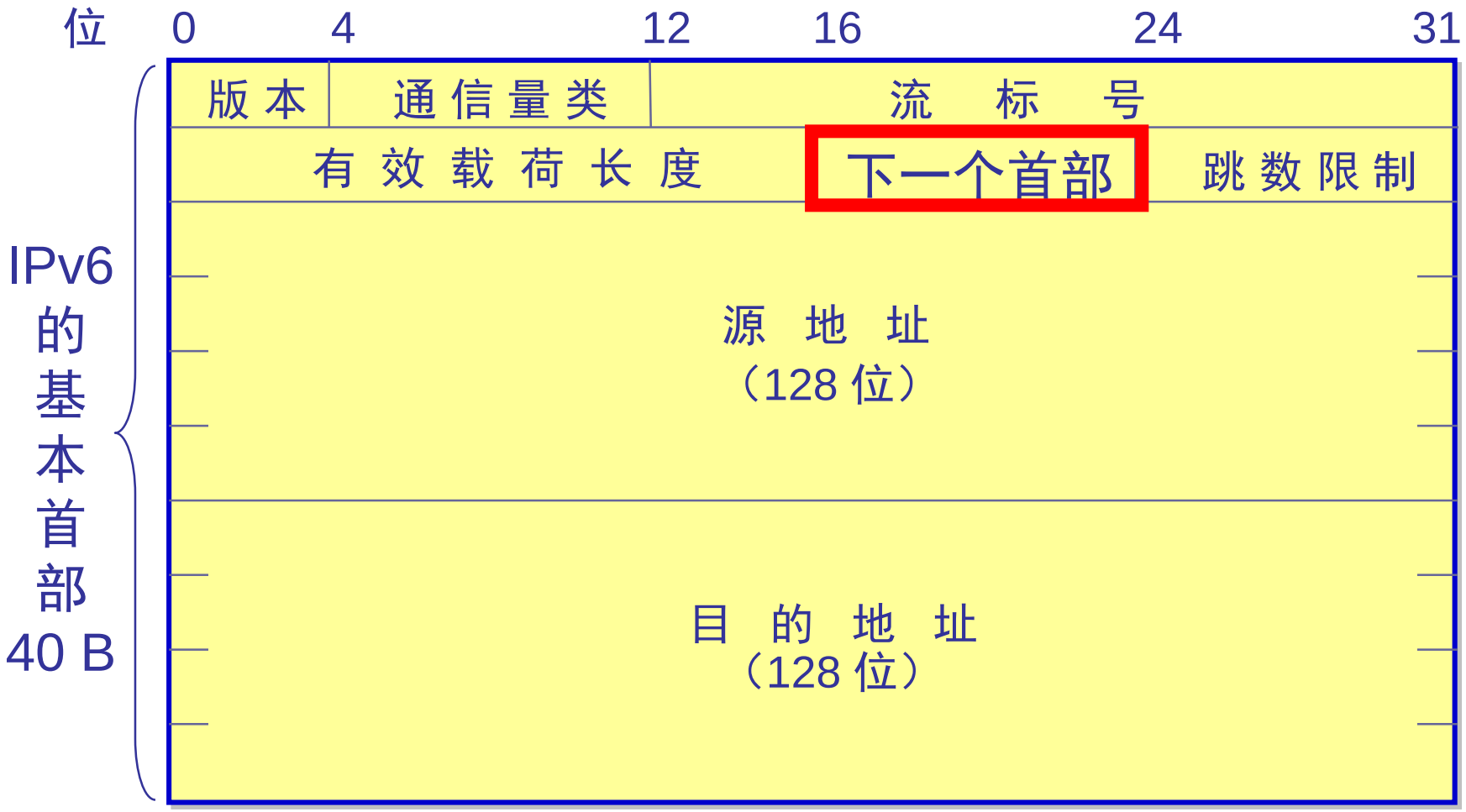


流标号(flow label)—— 20 位。“流”是互联网络上从特定源点到特定终点的一系列数据报，“流”所经过的路径上的路由器都保证指明的服务质量。

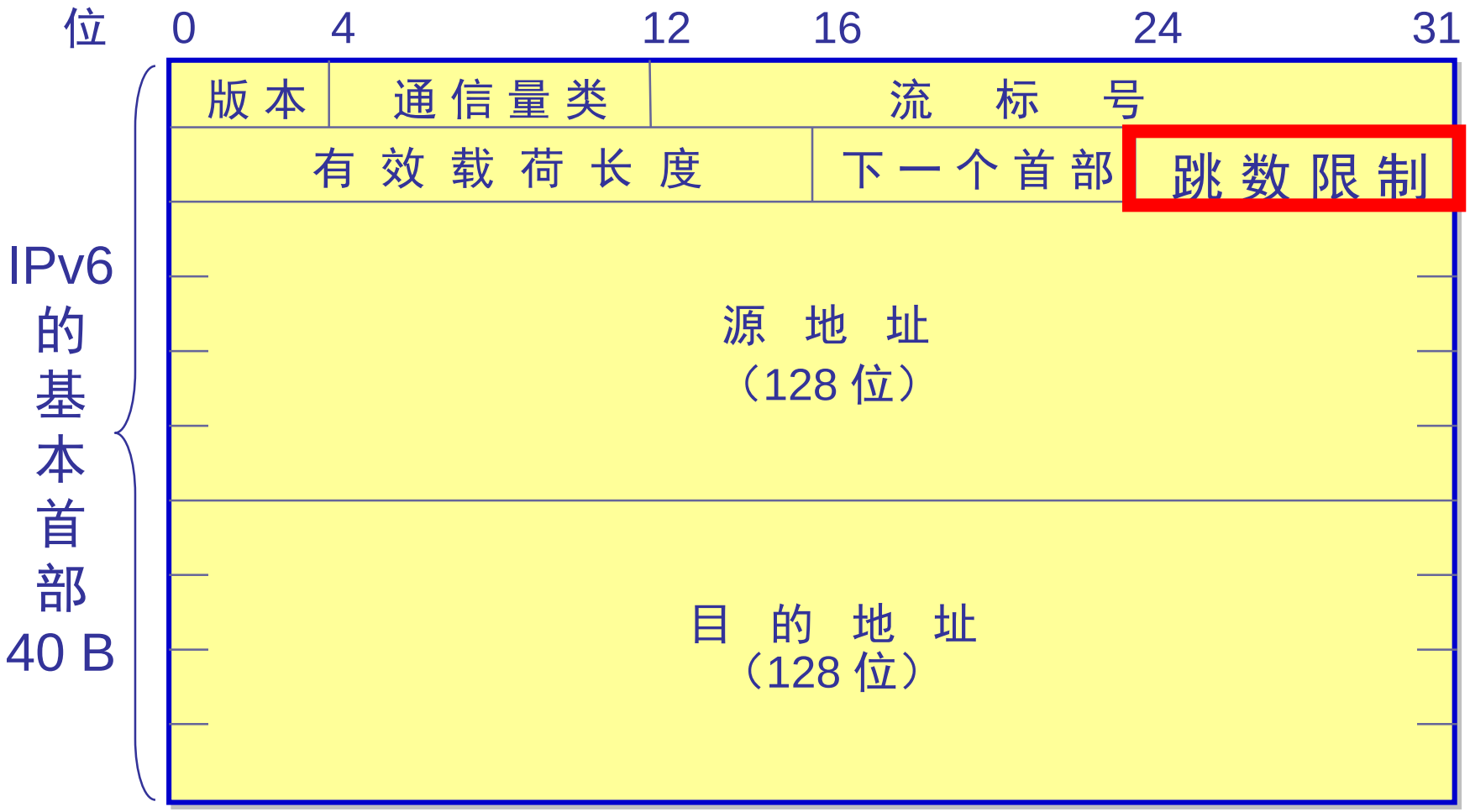
所有属于同一个流的数据报都具有同样的流标号。



有效载荷长度(payload length)—— 16 位。它指明 IPv6 数据报除基本首部以外的字节数（所有扩展首部都算在有效载荷之内），其最大值是 64 KB。

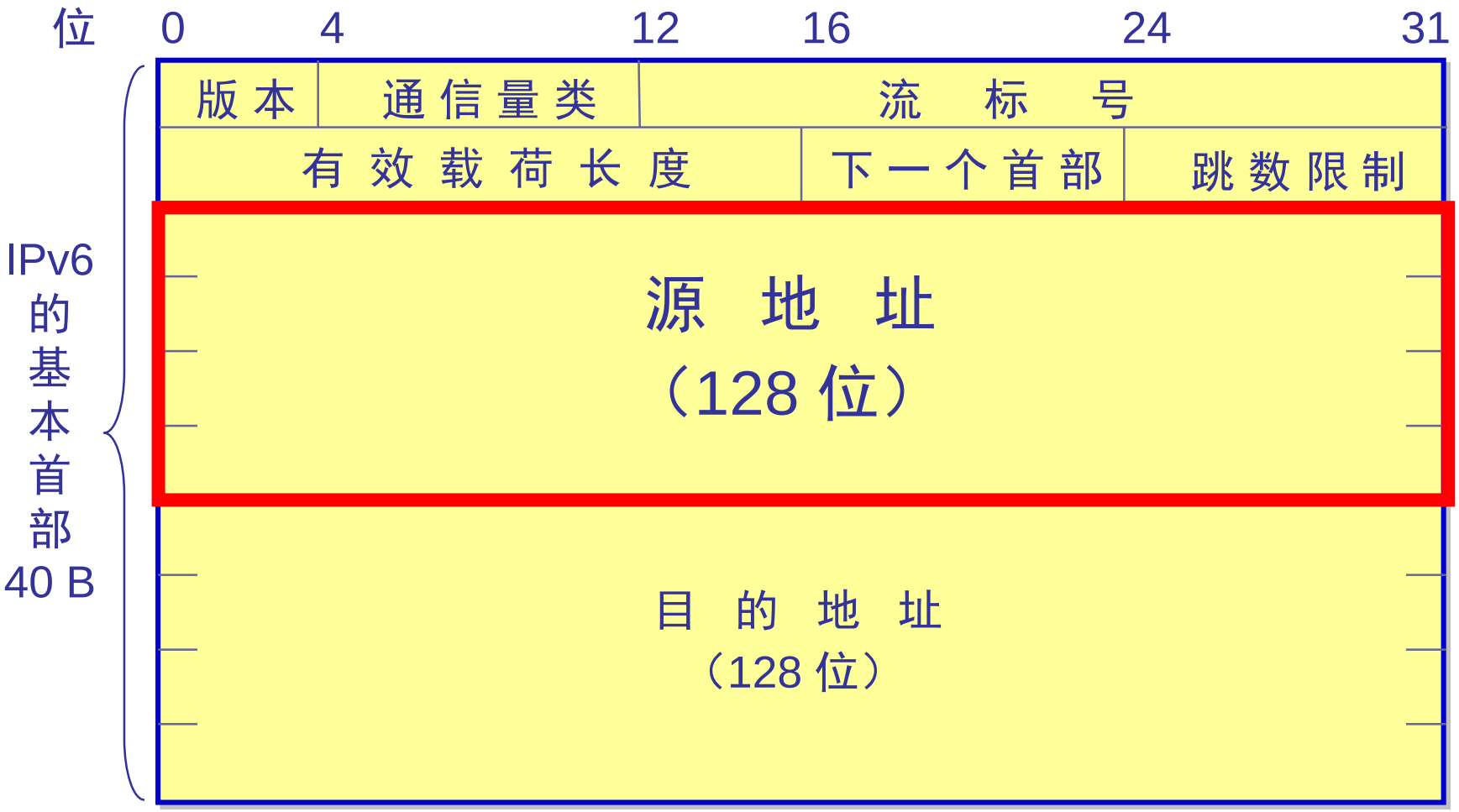


下一个首部(next header)—— 8 位。它相当于 IPv4 的协议字段或可选字段。

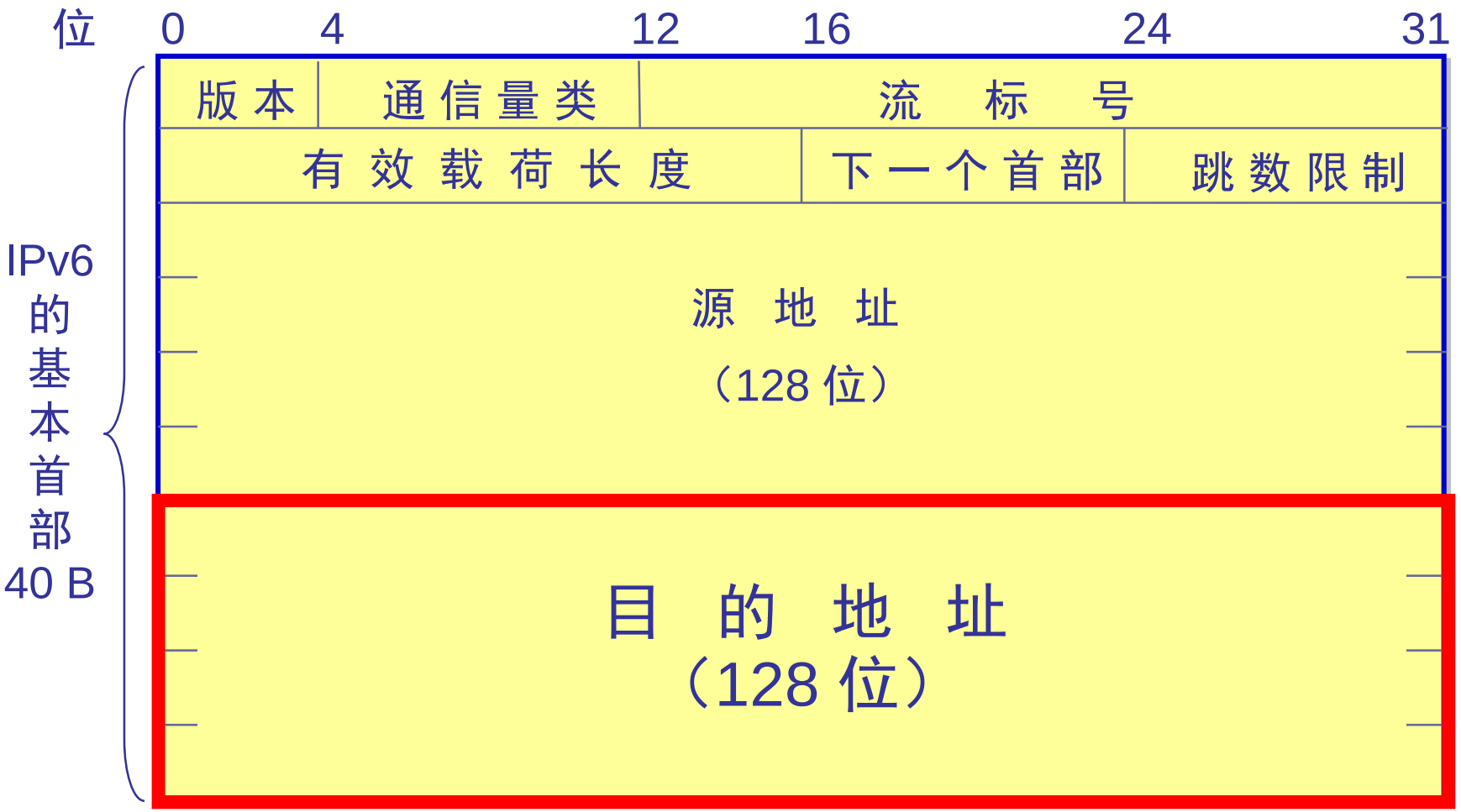


跳数限制(hop limit)—— 8 位。源站在数据报发出时即设定跳数限制。路由器在转发数据报时将跳数限制字段中的值减1。

当跳数限制的值为零时，就要将此数据报丢弃。



源地址—— 128 位。是数据报的发送站的 IP 地址。



目的地址—— 128 位。是数据报的接收站的 IP 地址。

IPv6 的地址空间

- IPv6 数据报的目的地址可以是以下三种基本类型地址之一：
  - (1) 单播(unicast) 单播就是传统的点对点通信。
  - (2) 多播(multicast) 多播是一点对多点的通信。
  - (3) 任播(anycast) 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。

结点与接口

- IPv6 将实现 IPv6 的主机和路由器均称为结点。
- IPv6 地址是分配给结点上面的接口。
  - 一个接口可以有多个单播地址。
  - 一个结点接口的单播地址可用来唯一地标志该结点。

冒号十六进制记法

- 每个 16 位的值用十六进制值表示，各值之间用冒号分隔。
  - 68E6:8C64:FFFF:FFFF:0:1180:960A:FFFF
- 零压缩,例:

  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

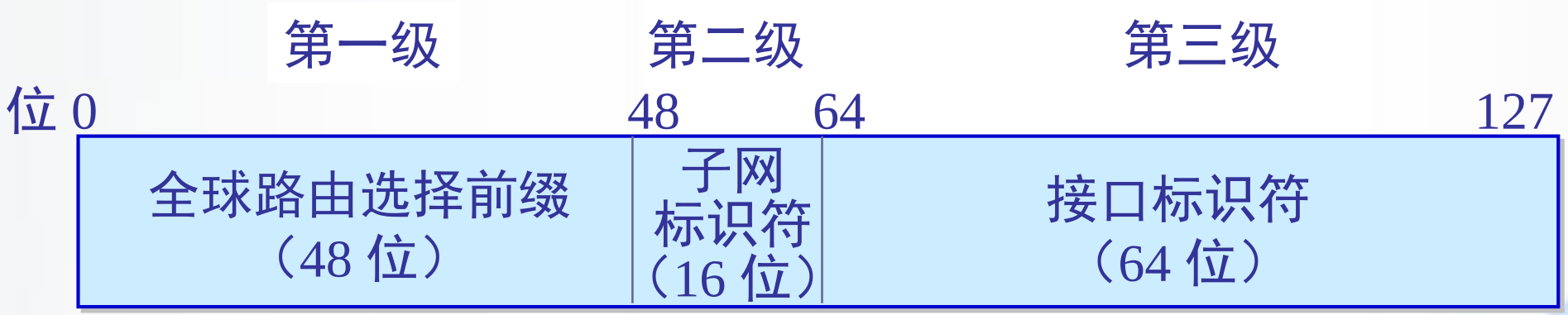
  或12AB:0:0:CD30::/60

特殊地址

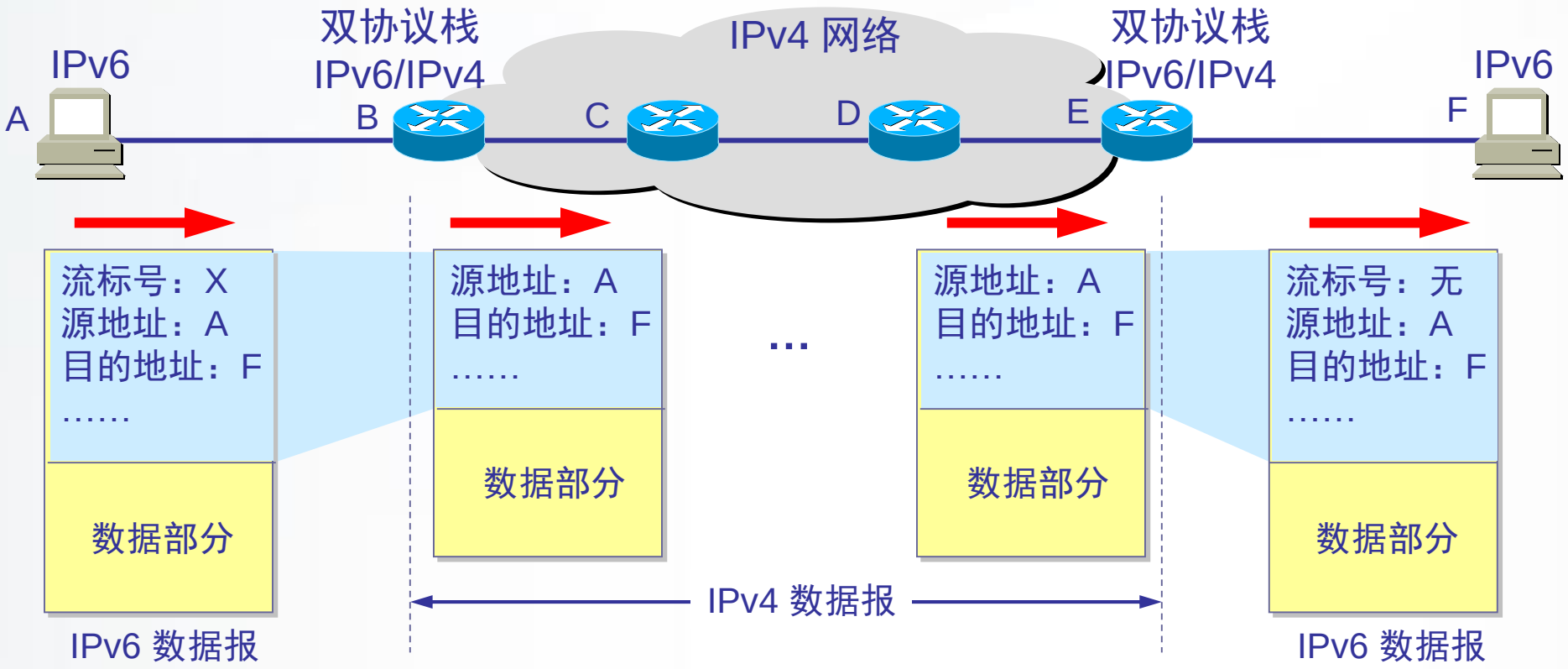
- **未指明地址**　这是 16 字节的全 0 地址，可缩写为两个冒号“::”。这个地址只能为还没有配置到一个标准的 IP 地址的主机当作源地址使用。
- **环回地址**　即 0:0:0:0:0:0:0:1（记为 ::1）。
- **基于 IPv4 的地址**　前缀为 0000 0000 保留一小部分地址作为与 IPv4 兼容的。
- **本地链路单播地址**

全球单播地址的等级结构

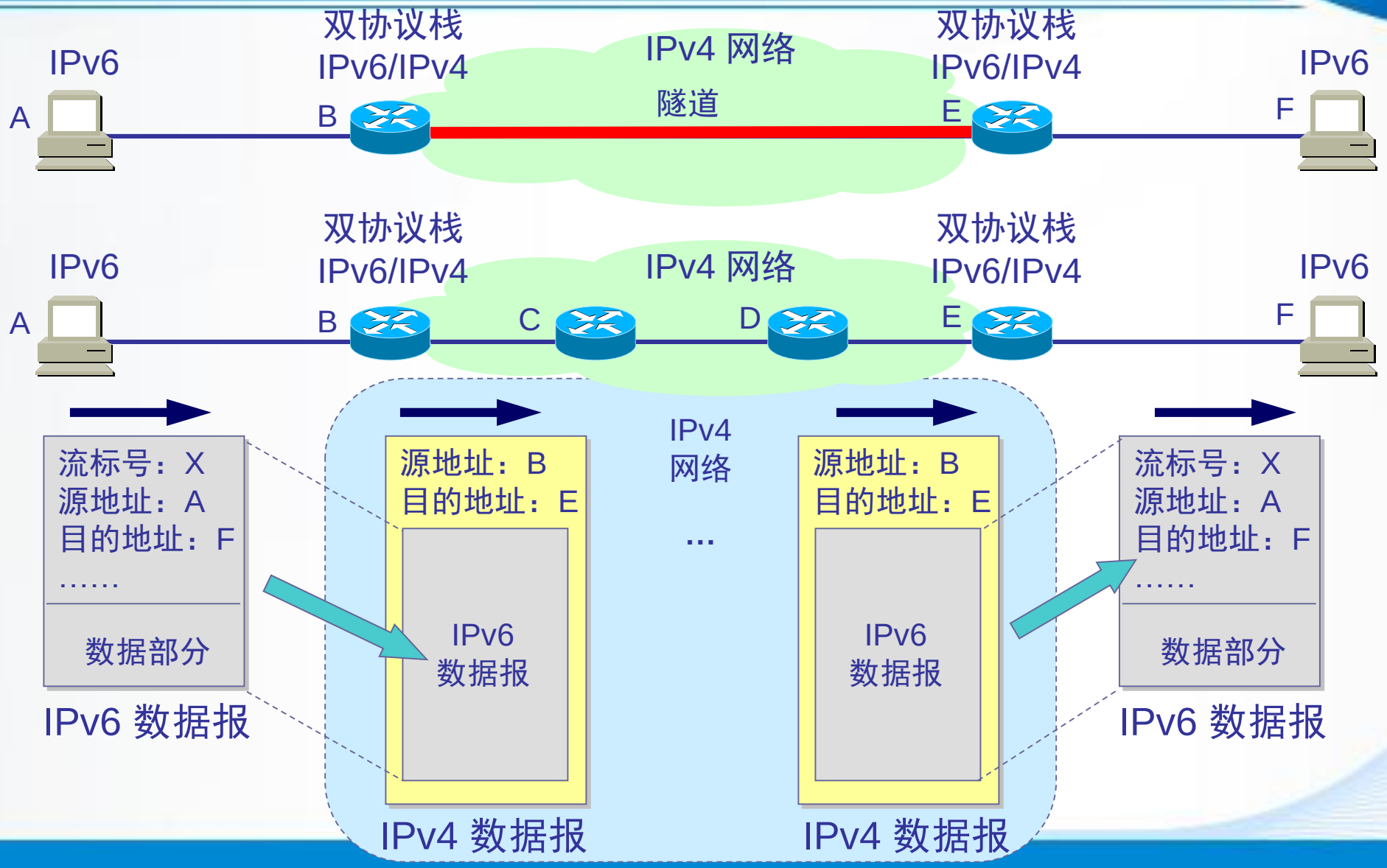
- IPv6 扩展了地址的分级概念，使用以下三个等级：
  - (1) 全球路由选择前缀，占 48 位。
  - (2) 子网标识符，占16 位。
  - (3) 接口标识符，占 64 位。



用双协议栈进行从 IPv4 到 IPv6 的过渡



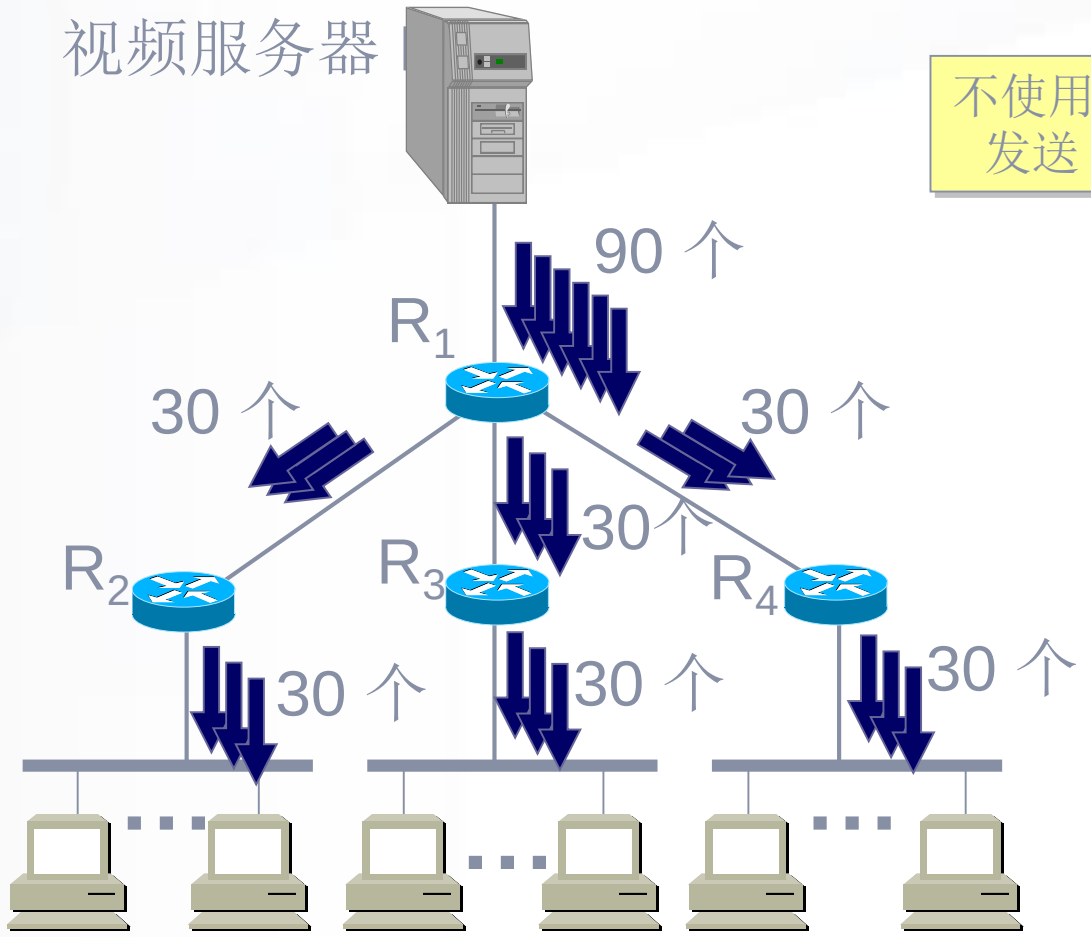
使用隧道技术从 IPv4 到 IPv6 过渡



指引

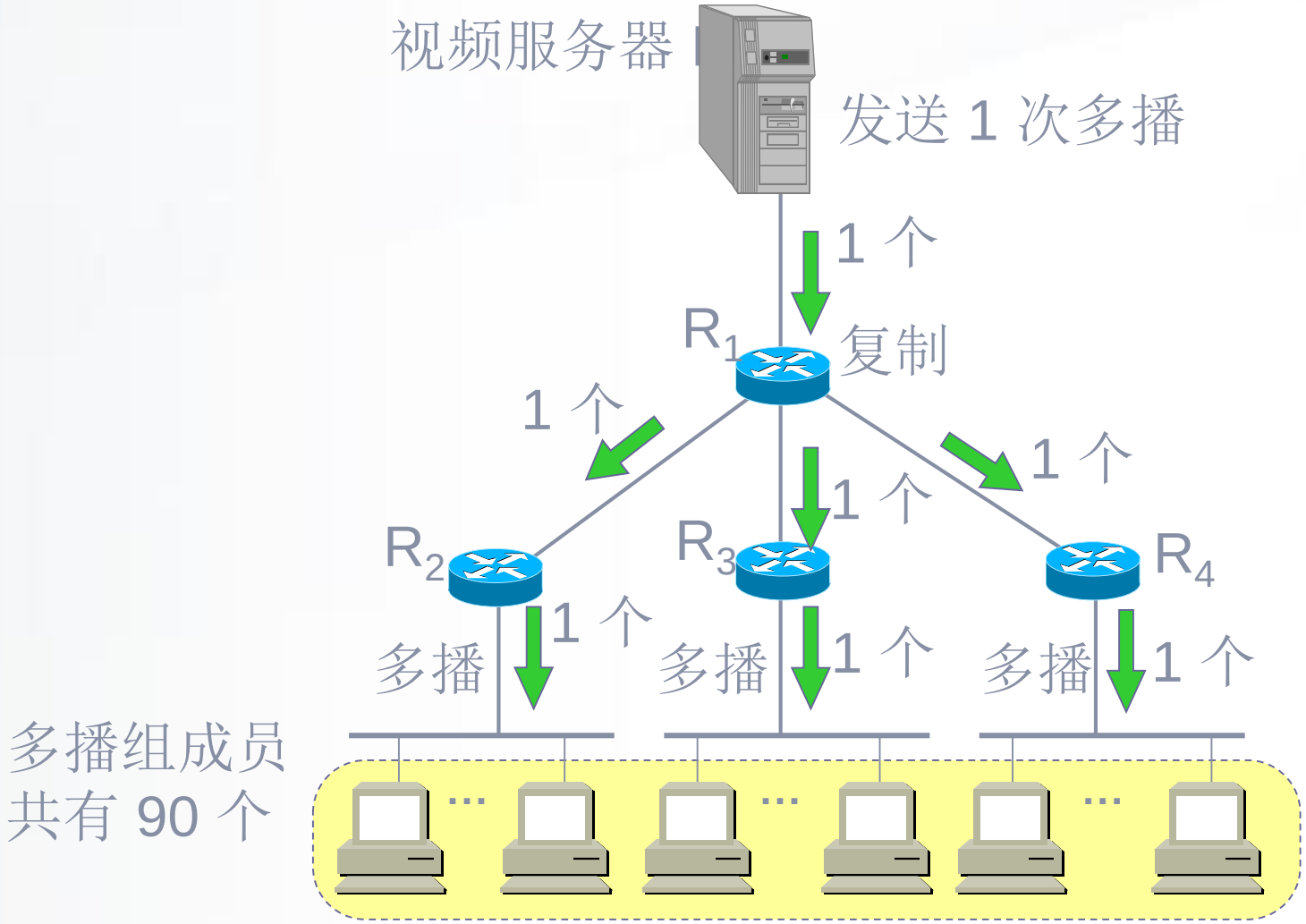
- 网络层提供的两种服务
- 网际协议 IP
- 划分子网和构造超网
  - 划分子网
  - 无分址编址
  - 构造超网
- 网际控制报文协议 ICMP
- 因特网的路由选择协议
- IPv6
- IP 多播
- 虚拟专用网 VPN 和网络地址转换 NAT

IP 多播的基本概念



共有 90 个主机接收视频节目

多播可明显地减少网络中资源的消耗



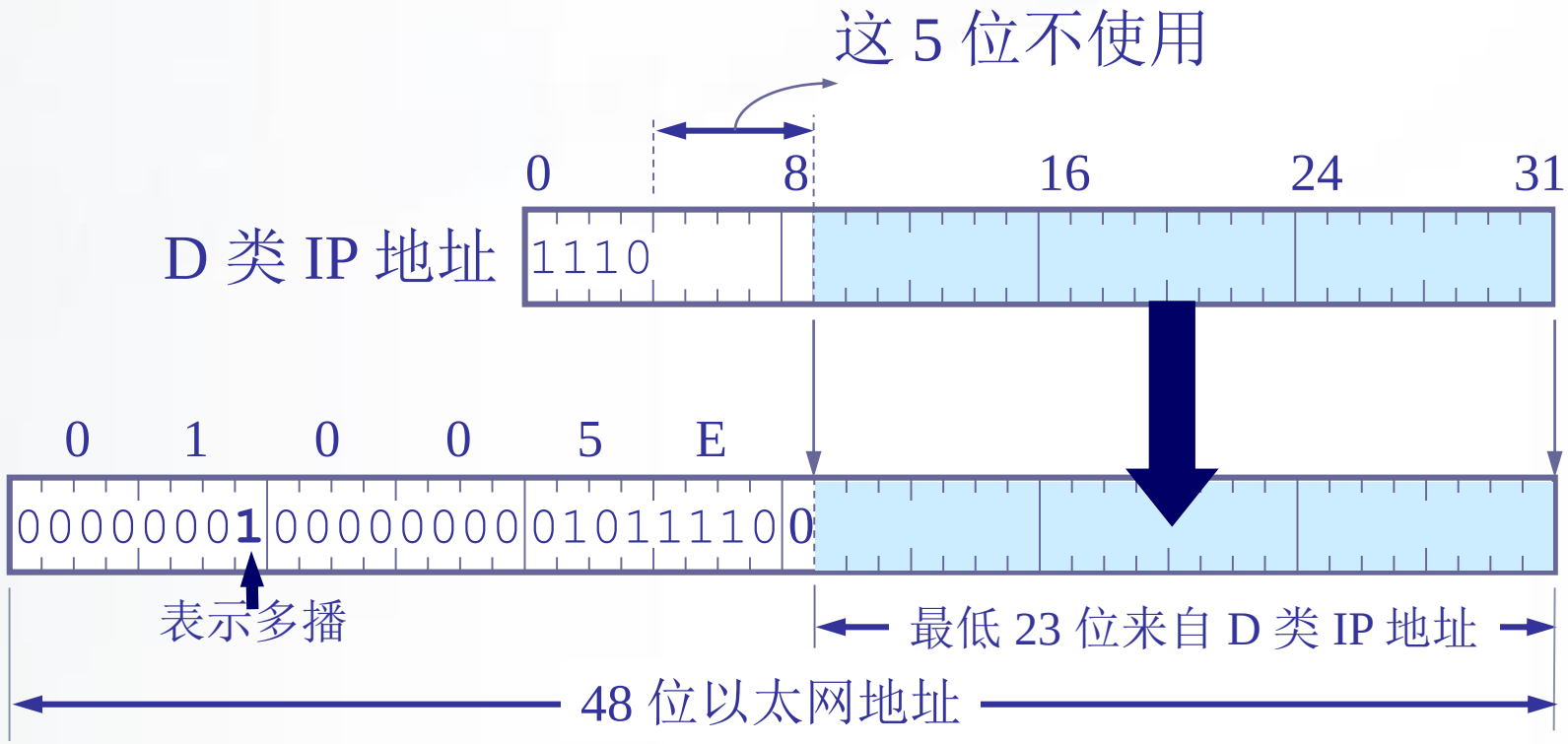
IP 多播的一些特点

- 多播使用组地址— IP 使用 D 类地址支持多播。多播地址只能用于目的地址，而不能用于源地址。
- 永久组地址—由因特网号码指派管理局 IANA 负责指派。
- 动态的组成员
- 使用硬件进行多播

在局域网上进行硬件多播

- 因特网号码指派管理局 IANA 拥有的以太网地址块的高 24 位为 00-00-5E。
- 因此 TCP/IP 协议使用的以太网多播地址块的范围是：从 00-00-5E-00-00-00 到 00-00-5E-FF-FF-FF
- D 类 IP 地址可供分配的有 28 位，在这 28 位中的前 5 位不能用来构成以太网硬件地址。

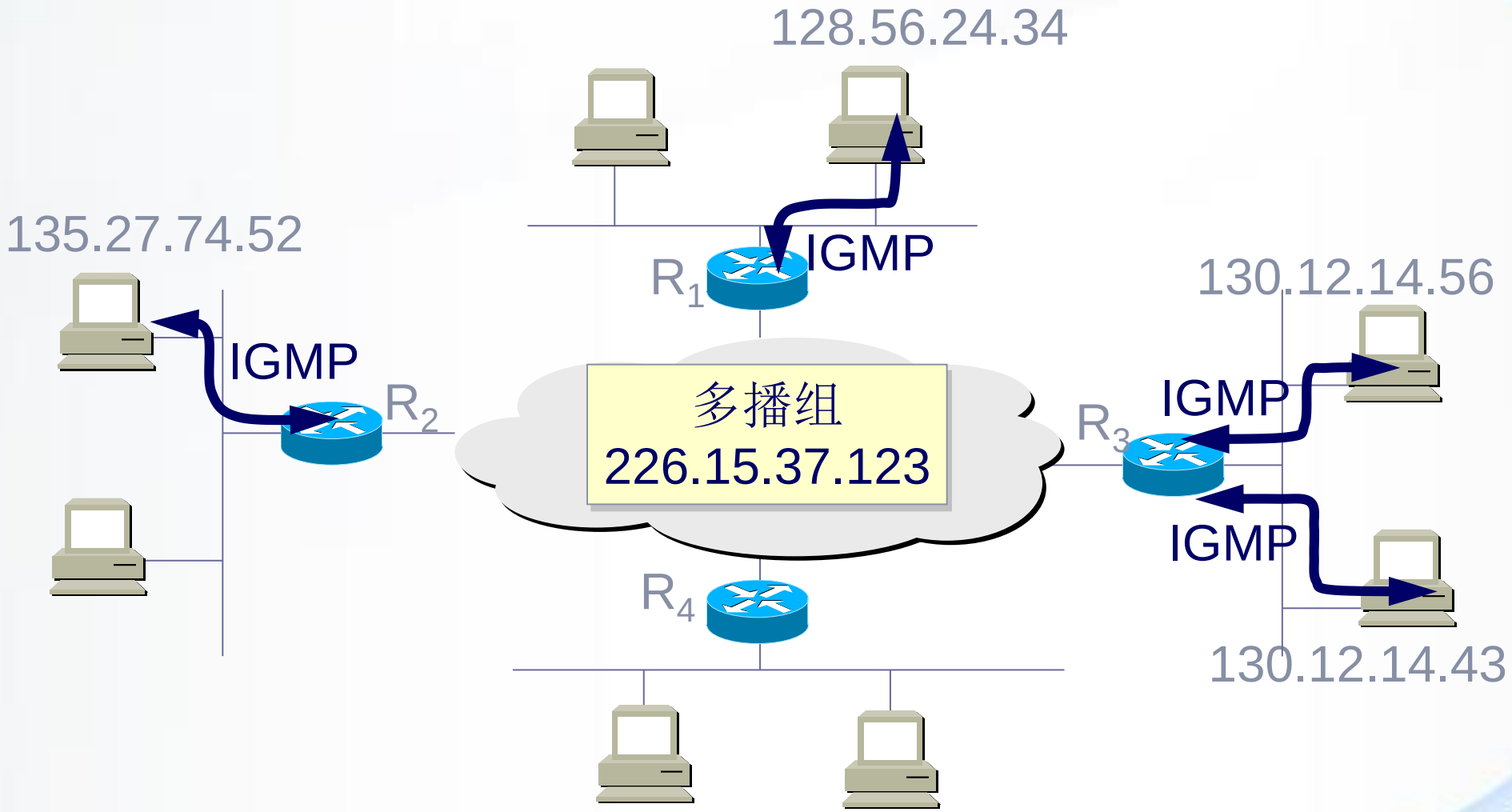
D 类 IP 地址与以太网多播地址的映射关系



IP多播需要两种协议

- IGMP网际组管理协议
- 多播路由选择协议

IGMP 使多播路由器知道多播组成员信息



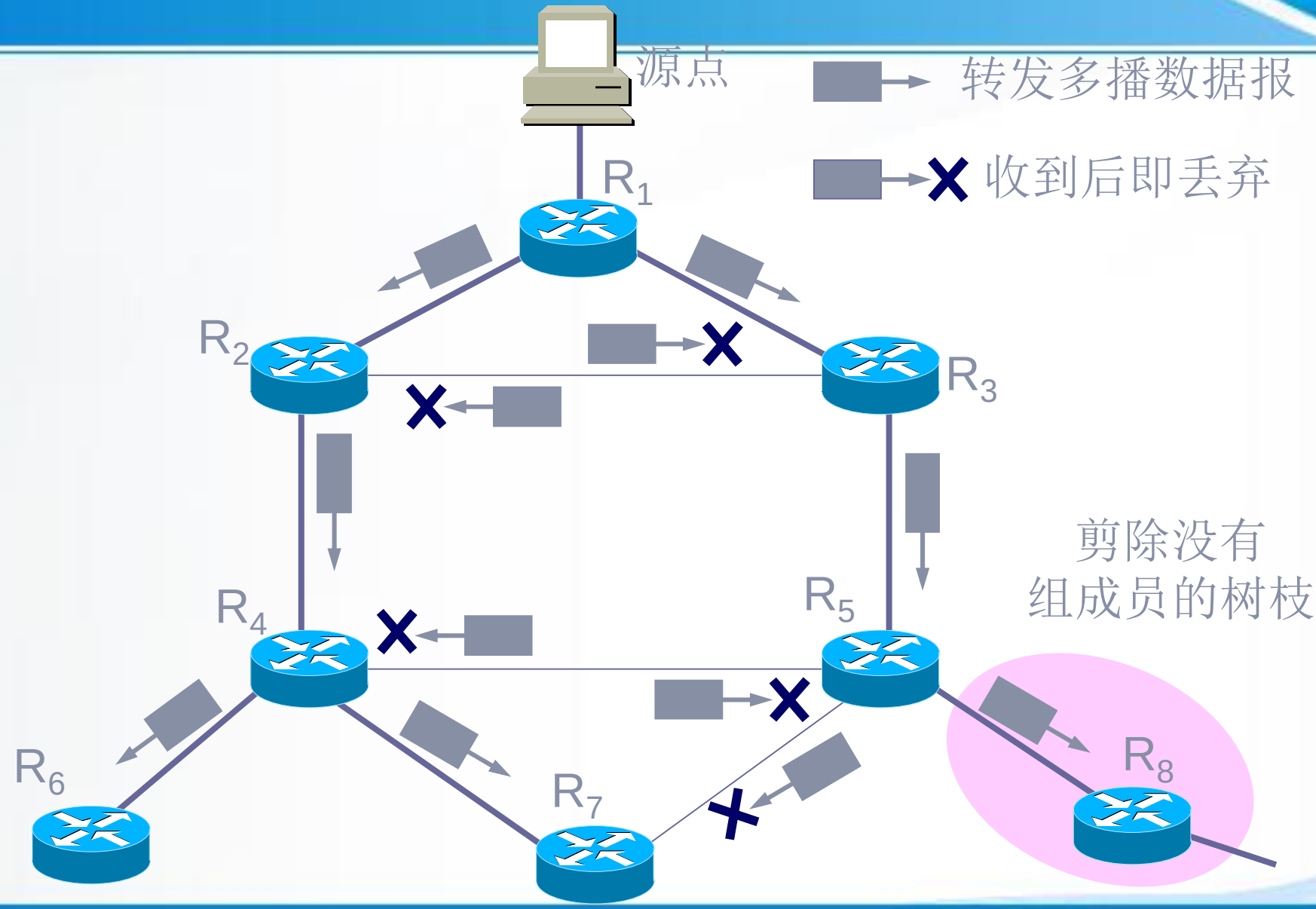
IGMP 可分为两个阶段

- 第一阶段：当某个主机加入新的多播组时，该主机应向多播组的多播地址发送IGMP 报文，声明自己要成为该组的成员。本地的多播路由器收到 IGMP 报文后，将组成员关系转发给因特网上的其他多播路由器。
- 第二阶段：因为组成员关系是动态的，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。

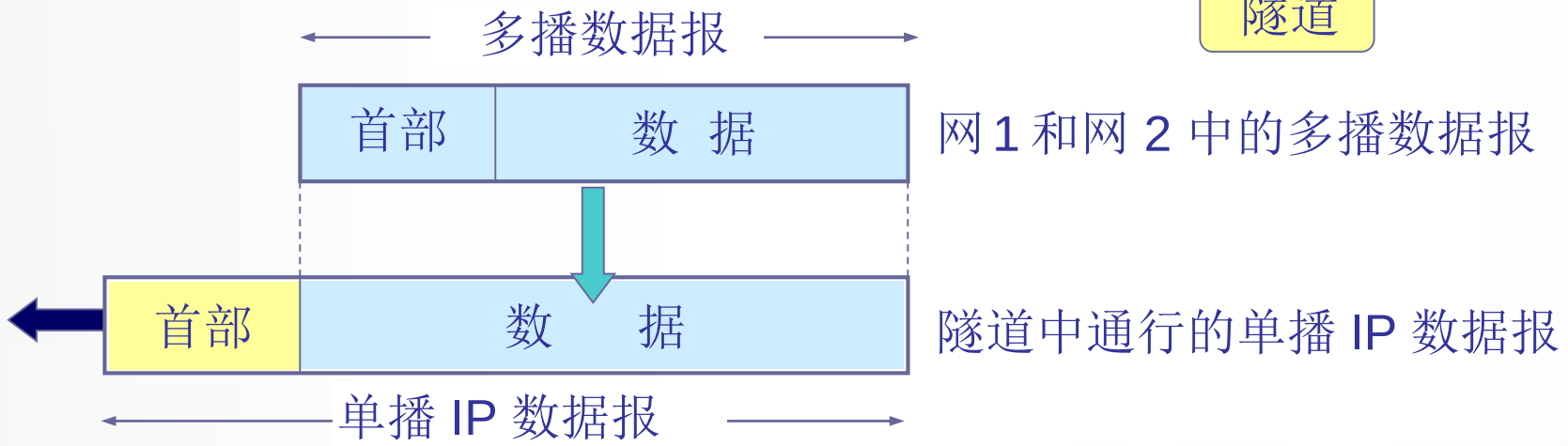
多播路由选择

- 洪泛和剪除
- 隧道技术(tunneling)
- 基于核心的发现技术

洪泛和剪除（RPB）



隧道技术(tunneling)



基于核心的发现技术

- 这种方法对于多播组的大小在较大范围内变化时都适合。
- 这种方法是对每一个多播组 G 指定一个核心(core)路由器，给出它的 IP 单播地址。
- 核心路由器按照前面讲过的方法创建出对应于多播组 G 的转发树。

指引

- 网络层提供的两种服务
- 网际协议 IP
- 划分子网和构造超网
  - 划分子网
  - 无分址编址
  - 构造超网
- 网际控制报文协议 ICMP
- 因特网的路由选择协议
- IPv6
- IP 多播
- 虚拟专用网 VPN 和网络地址转换 NAT

虚拟专用网 VPN

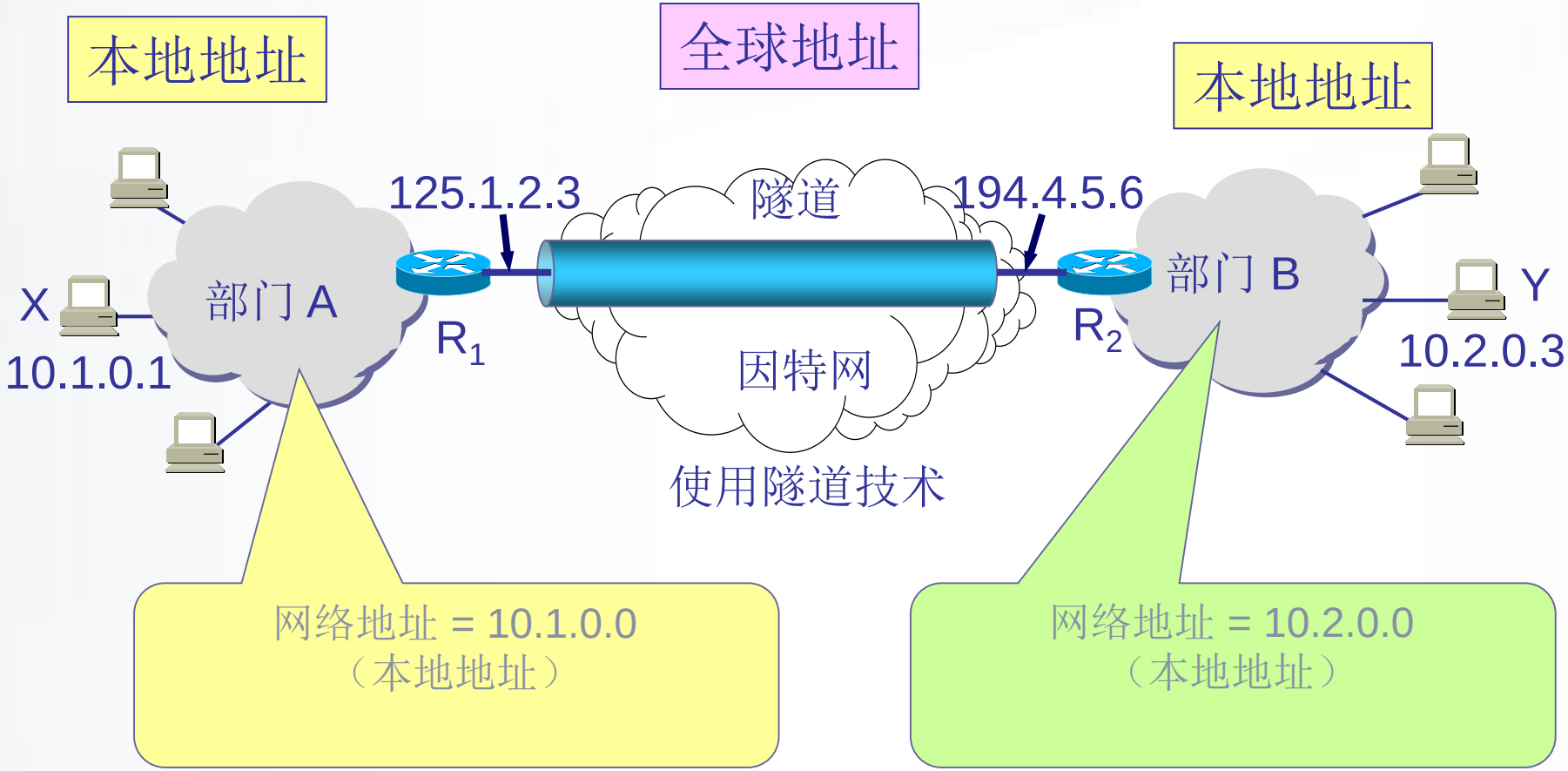
- 本地地址——仅在机构内部使用的 IP 地址，可以由本机构自行分配，而不需要向因特网的管理机构申请。
- 全球地址——全球唯一的IP地址，必须向因特网的管理机构申请。

RFC 1918 指明的专用地址

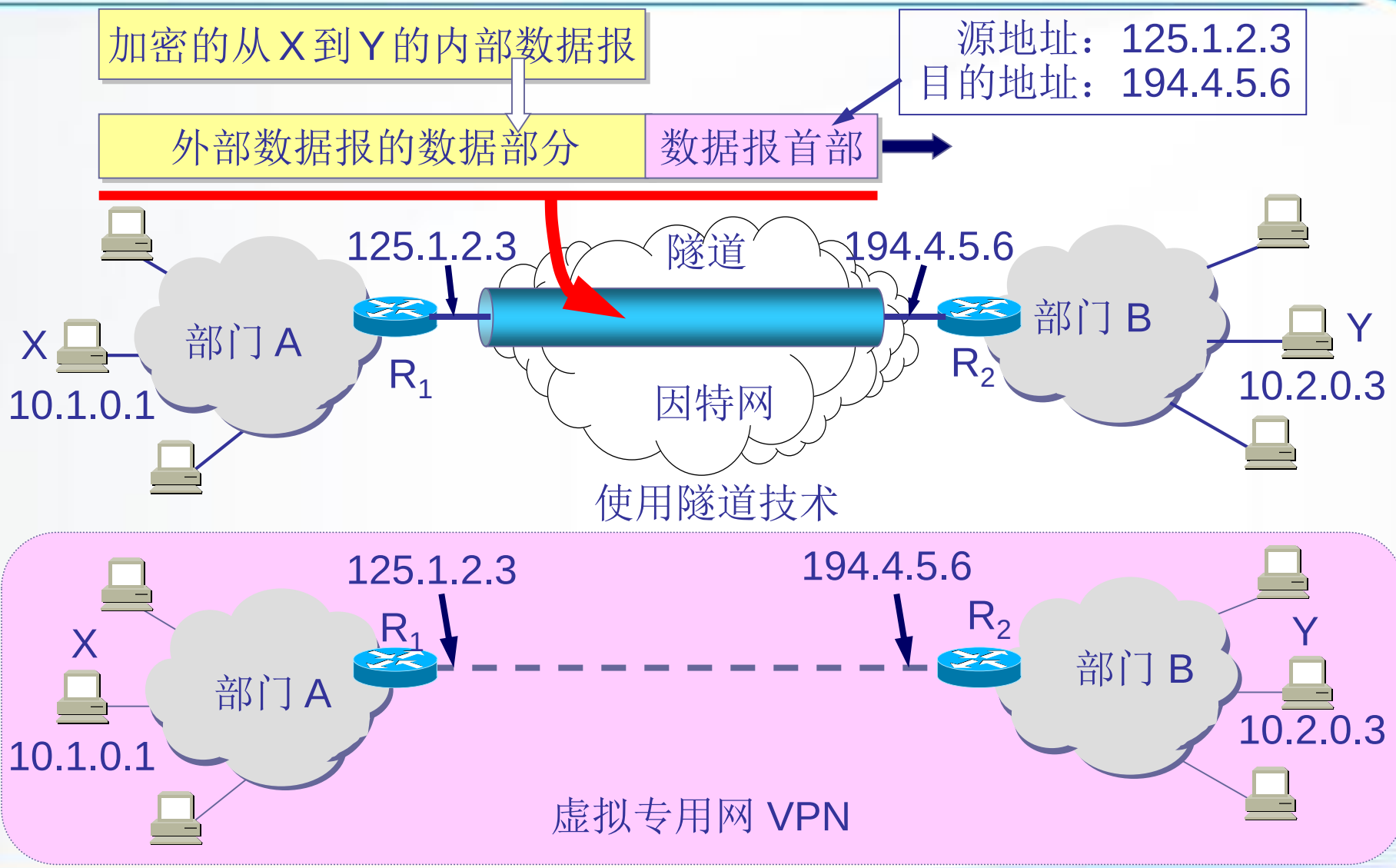
(private address)

10.0.0.0 到 10.255.255.255
172.16.0.0 到 172.31.255.255
192.168.0.0 到 192.168.255.255

用隧道技术实现虚拟专用网



用隧道技术实现虚拟专用网



远程接入VPN(remote access VPN)

- 有的公司可能没有分布在不同场所的部门，但有很多流动员工在外地工作。公司需要和他们保持联系，远程接入 VPN 可满足这种需求。
- 在外地工作的员工拨号接入因特网，而驻留在员工 PC 机中的 VPN 软件可在员工的 PC 机和公司的主机之间建立 VPN 隧道，因而外地员工与公司通信的内容是保密的，员工们感到好像就是使用公司内部的本地网络。

动手实验

- 实验4-10：配置远程接入VPN

网络地址转换 NAT

- 网络地址转换 NAT 方法于1994年提出。
- 需要在专用网连接到因特网的路由器上安装 NAT 软件。装有 NAT 软件的路由器叫做 NAT路由器，它至少有一个有效的外部全球地址 IP_G。
- 所有使用本地地址的主机在和外界通信时都要在 NAT 路由器上将其本地地址转换成 IP_G 才能和因特网连接。

网络地址转换的过程

- 内部主机 X 用本地地址 IP_X和因特网上主机 Y 通信所发送的数据报必须经过 NAT 路由器。
- NAT 路由器将数据报的源地址 IP_X转换成全球地址 IP_G，但目的地址 IP_Y保持不变，然后发送到因特网。
- NAT 路由器收到主机 Y 发回的数据报时，知道数据报中的源地址是 IP_Y而目的地址是 IP_G。
- 根据 NAT 转换表，NAT 路由器将目的地址 IP_G转换为 IP_X，转发给最终的内部主机 X。

动手实验

- 实验4-11：配置PAT

本章小结

- 网络层的两种服务
- IP
- 划分子网
- 路由选择协议
- VPN
- 多播
- NAT

Thank You!
Any Questions?