

网络层 – 数据平面

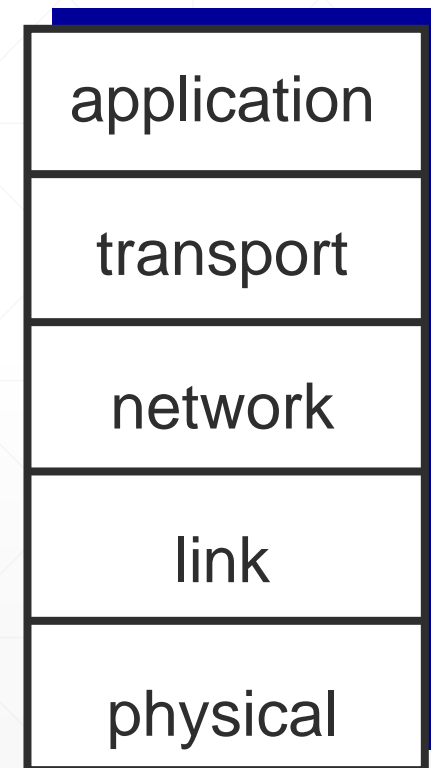
Network layer – data plane

本课内容

- 了解网络层服务及协议的原理，关注数据平面
 - 网络层服务模型
 - 转发和路由
 - 路由器基础
 - IP协议
 - 数据平面和控制平面
 - 路由器基础
 - IP协议
-

网络层的位置

- 网络层
- Network layer
- 以后每一次课件都会有这张图！

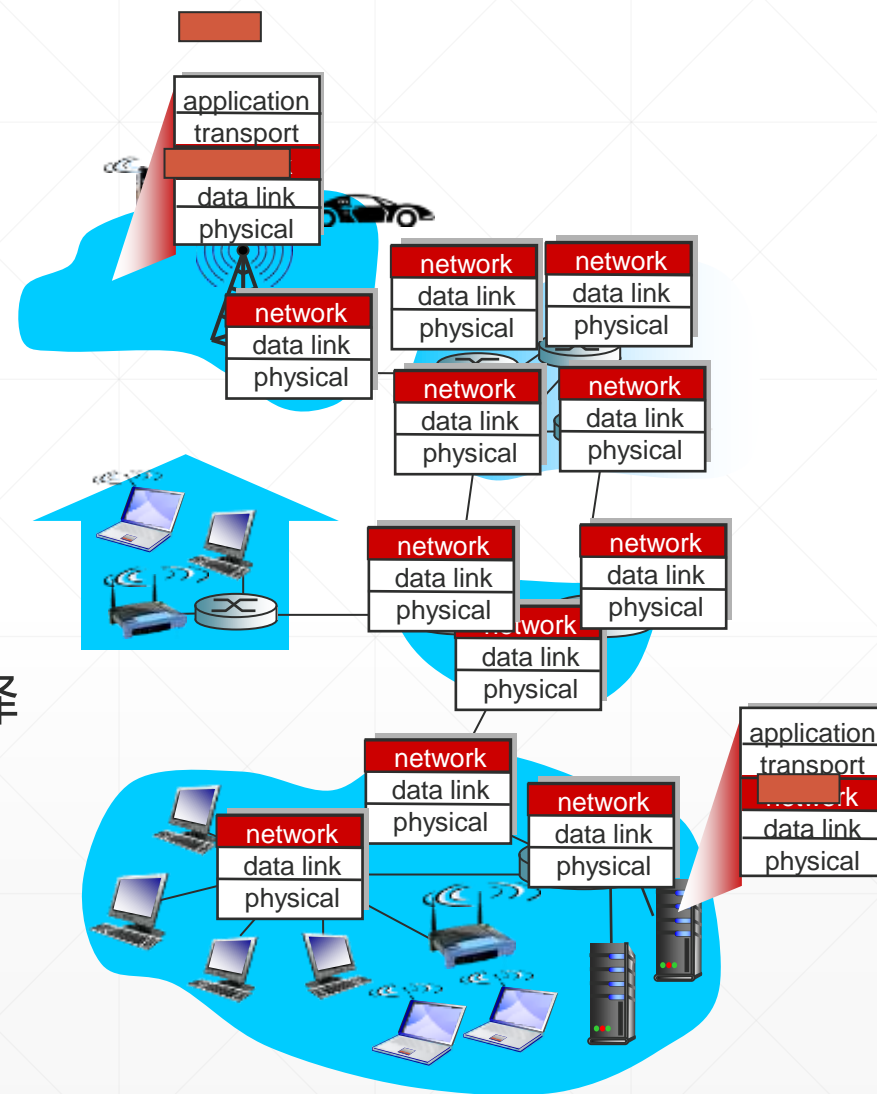


数据平面和控制平面

Data plane and control plane

网络层

- 从发送端到接收端传输数据段 (segment)
- 发送端：将传输层的数据段包装成数据报
- 接收端：将收到的数据报解包，然后传送到传输层
- 每个端/主机、路由都需要部署网络层
- 路由器检查每个IP数据报的首部，并根据路由表选择不同路径发送IP数据报



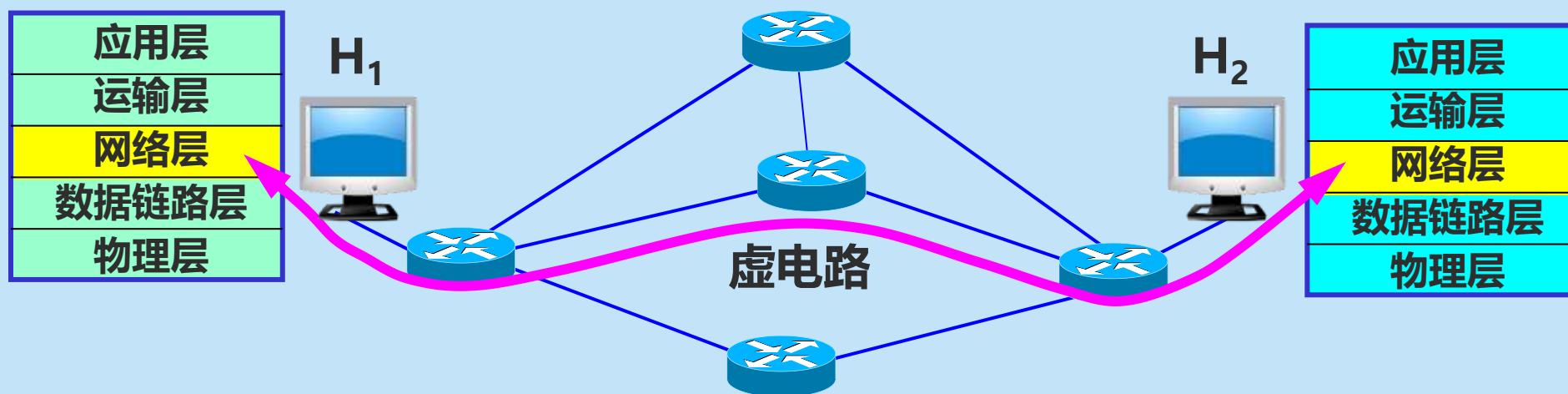
无连接的数据报服务

- 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？
 - 现在的互联网：“无连接” “尽最大努力的” 数据报服务
 - Connectionless
 - Best-effort service
-

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。
-

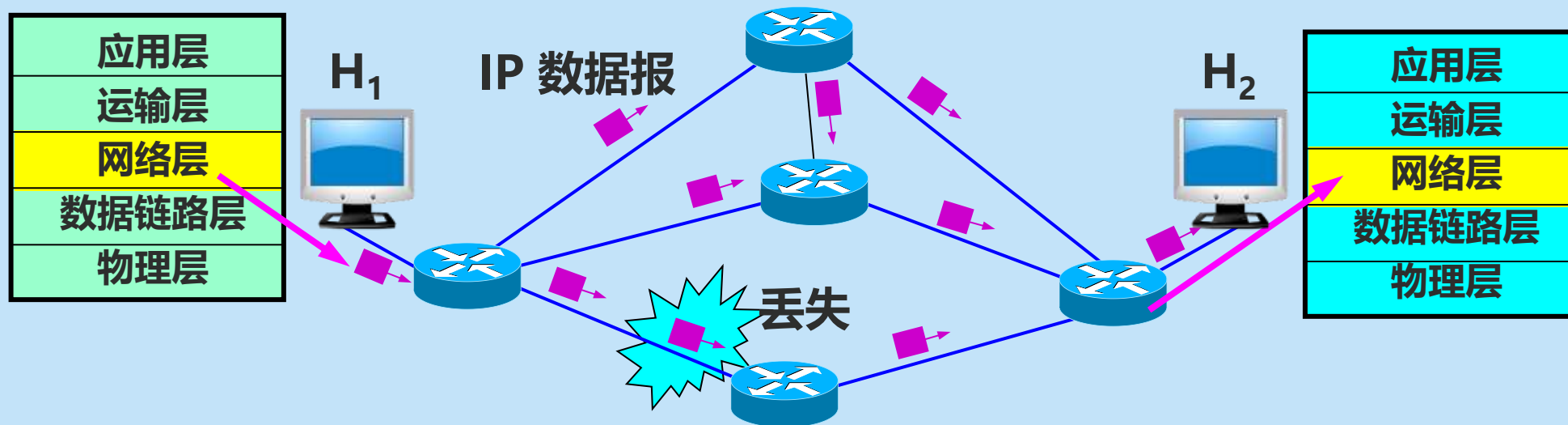
另一种观点：网络提供数据报服务

- 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。
-

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。
 - 互连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

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

两个网络层的关键作用 – 路由和转发

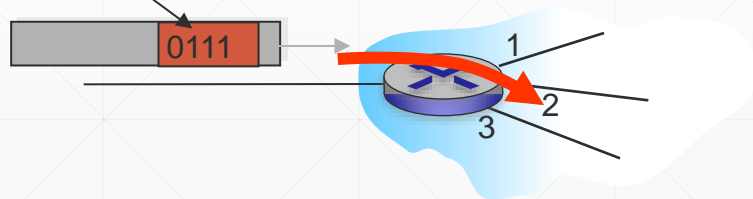
- **网络层关键作用：**
 - **转发：**将分组从路由器的输入端口转发到合适的输出端口
 - **路由：**决定分组从发送端到接收端的具体路径
 - 用到路由算法
 - **类比，旅行：**
 - **转发：**在机场，从入口进入，从哪个登机口出去？
 - **路由：**计划旅行过程，包括从出发到回家的整个过程
-

数据平面和控制平面

- **数据平面：**

- 本地、基于单个路由器的行为
- 主要行为：转发

values in arriving
packet header

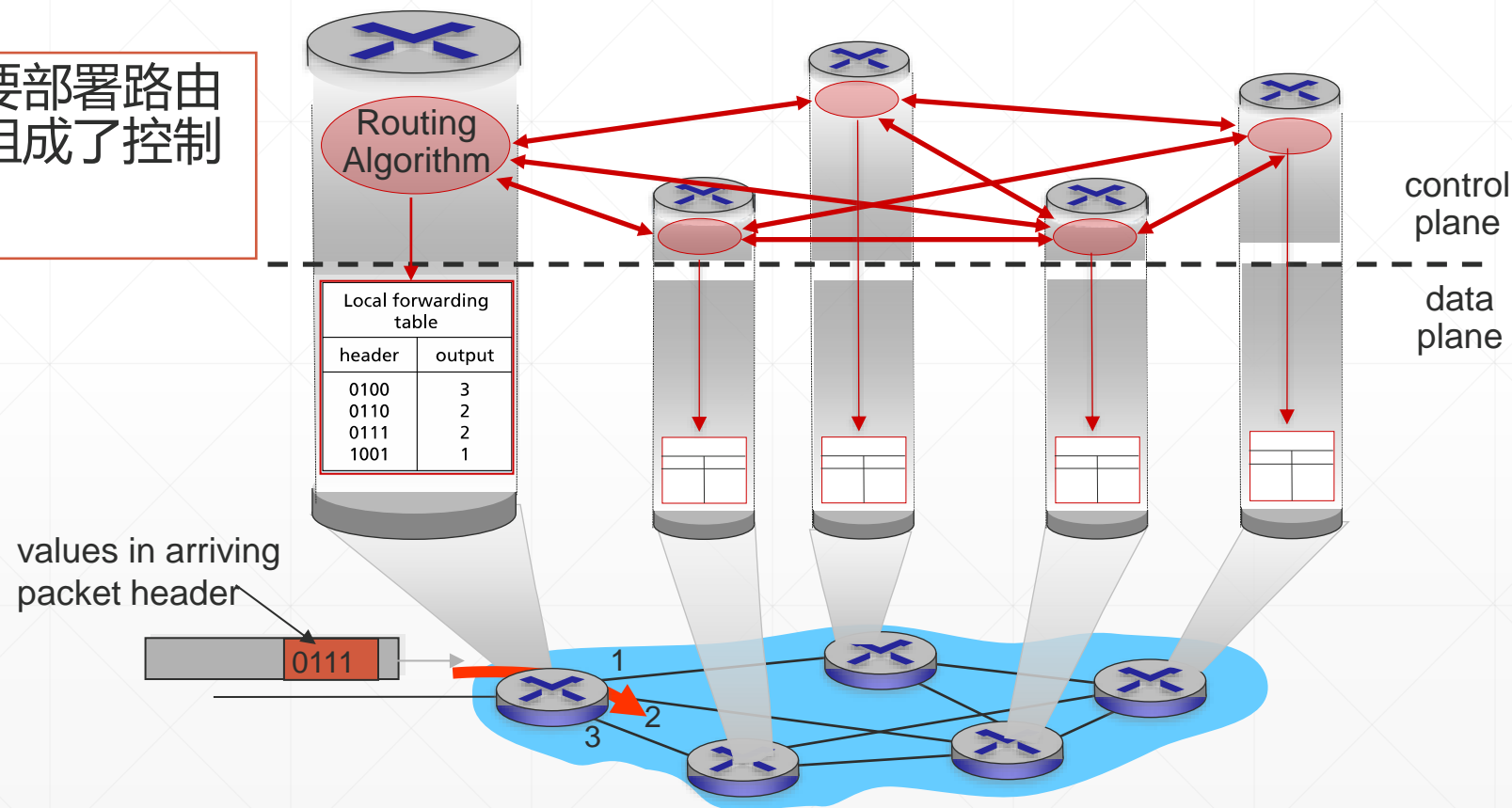


- **控制平面：**

- Network-wide的行为
- 主要行为：路由
- 两种控制平面的路由做法：
 - 传统的路由算法：在每个路由器上运行路由算法，决定分组的路由
 - SDN（软件定义网络Software-defined Networking）：在中心服务器上定义路由算法。

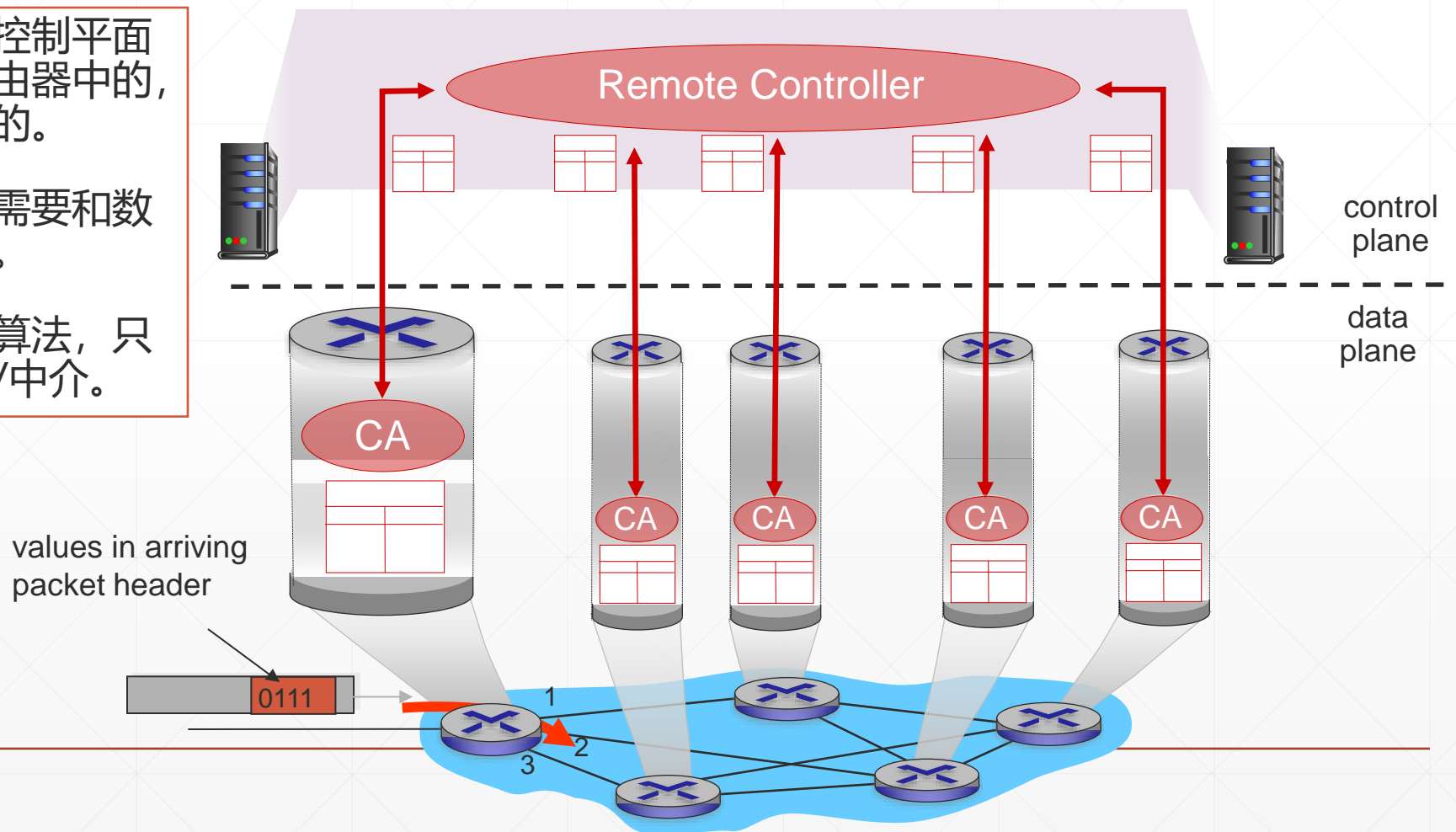
传统路由方法 per-router control plane

- 每一个路由器上都需要部署路由算法。这些路由算法组成了控制平面。



SDN的路由 logically centralized control plane

- 和传统路由不同，控制平面不是部署在每个路由器中的，而是逻辑上中心化的。
- 控制平面往往并不需要和数据平面部署在一起。
- 路由器中不再部署算法，只是控制平面的代理/中介。

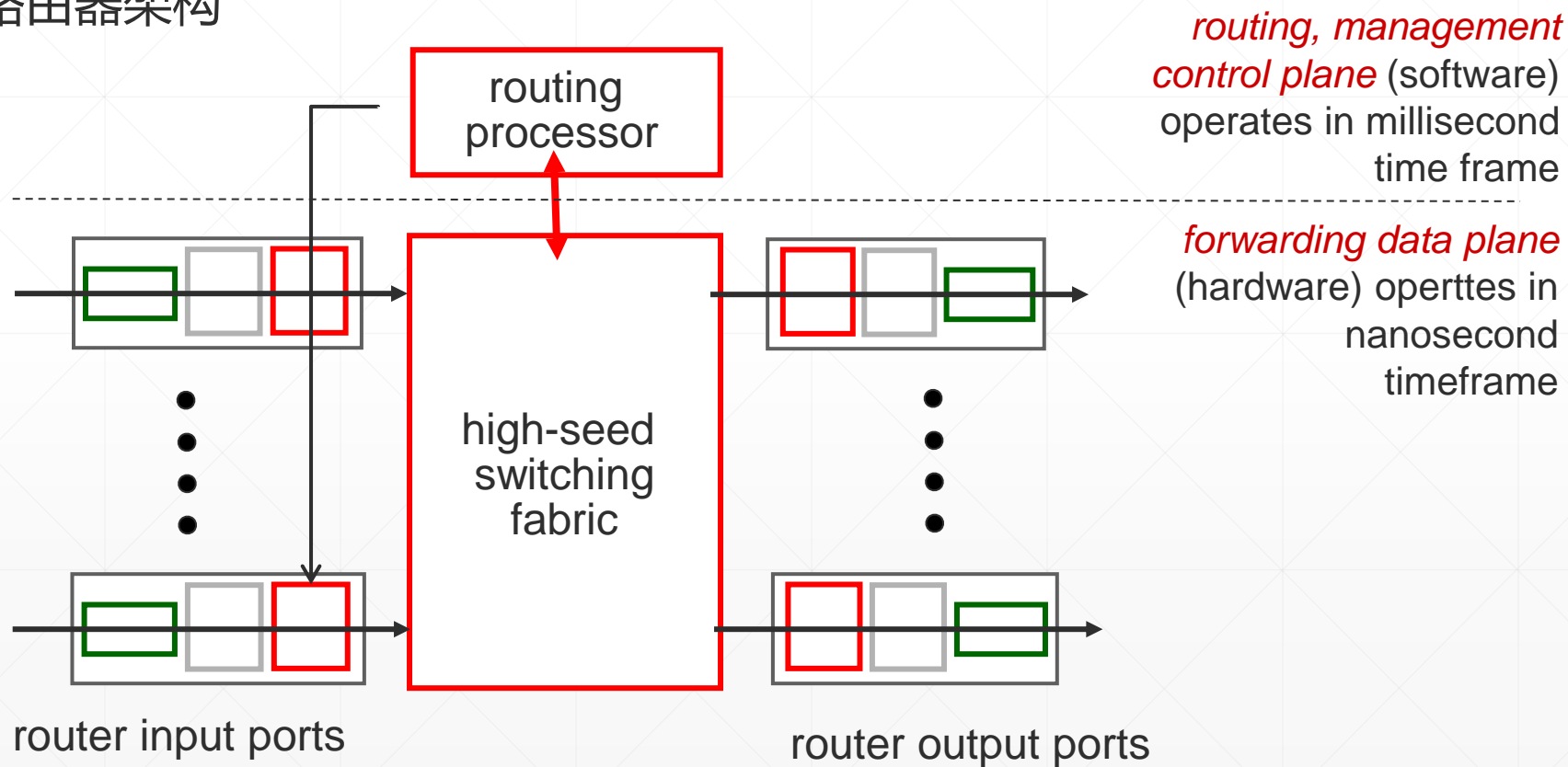


路由器基础

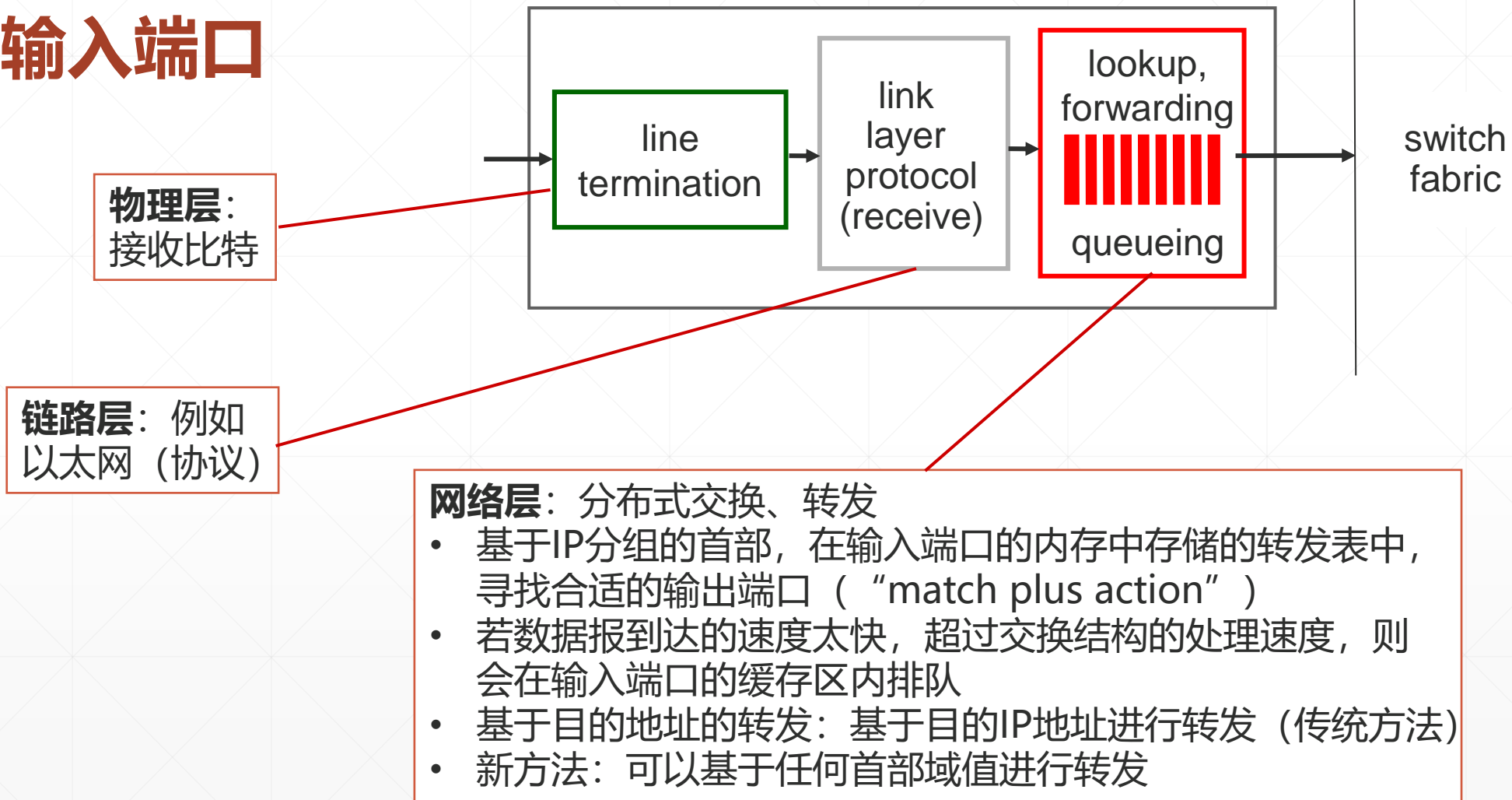
What' s inside a router?

路由器架构

- 通用的路由器架构

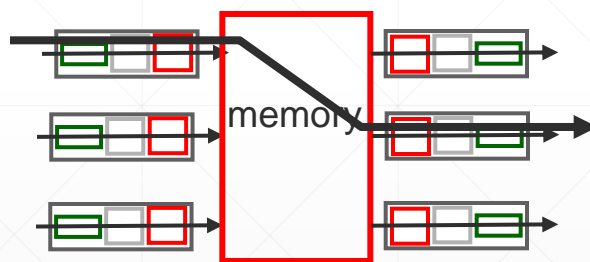


路由器输入端口



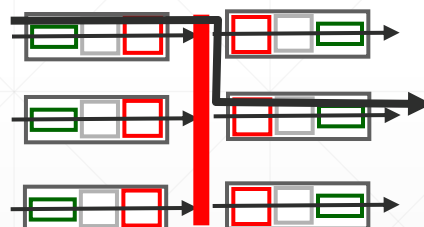
交换结构 Switching fabrics

- 负责将分组从输入端口的缓存发送到合适的输出端口缓存
- 交换速率：从所有输入端口到所有输出端口转发分组的速率
- 三种不同的交换结构



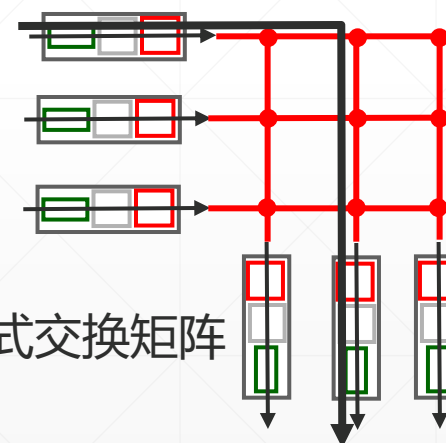
基于内存的

性能受限于内存带宽



基于总线的

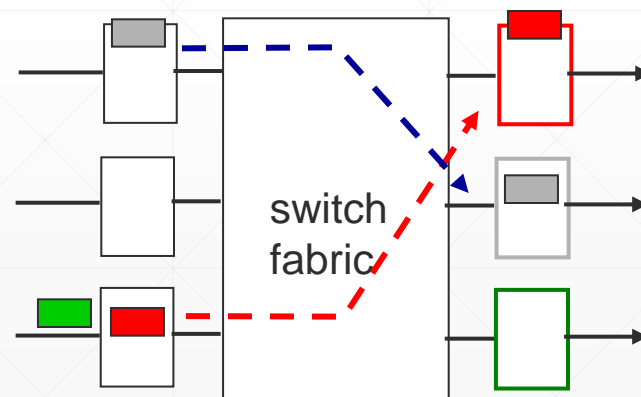
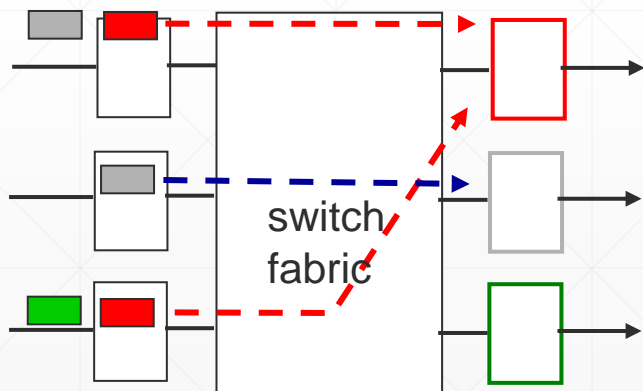
性能受限于总线带宽



纵横式交换矩阵

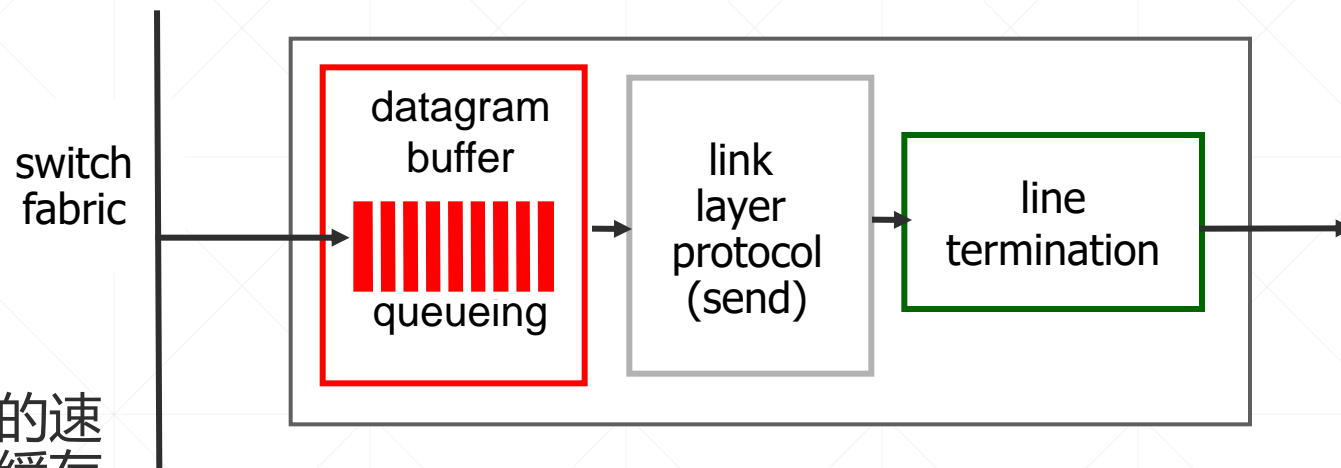
输入端口队列 Input port queueing

- 若交换结构转发分组的速度比所有输入端口分组到达的速度和要慢，则会在输入端口的缓存中发生排队
 - 若缓存溢出，则会产生排队延时或分组丢失
- HOL阻塞（Head-of-the-line）：排在队列前面的数据报可能会阻碍队列中其他的数据报的转发



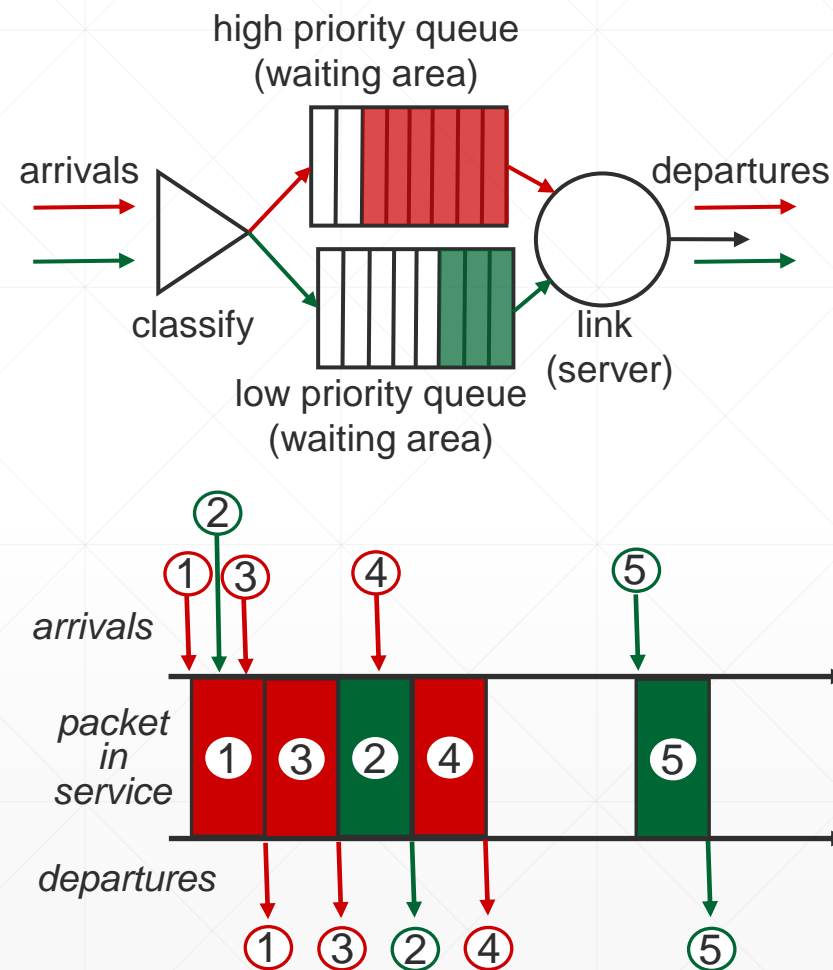
路由器输出端口

- 结构和输入端口类似
- 若发出分组的速度比交换结构的速度满，则分组会在输出端口的缓存中排队。
 - 若缓存大小不足，则分组可能因为发生拥塞而丢失
- 需要使用调度算法来选择传输哪些排队的分组
 - 优先级调度 – 选择哪个分组进行输出？



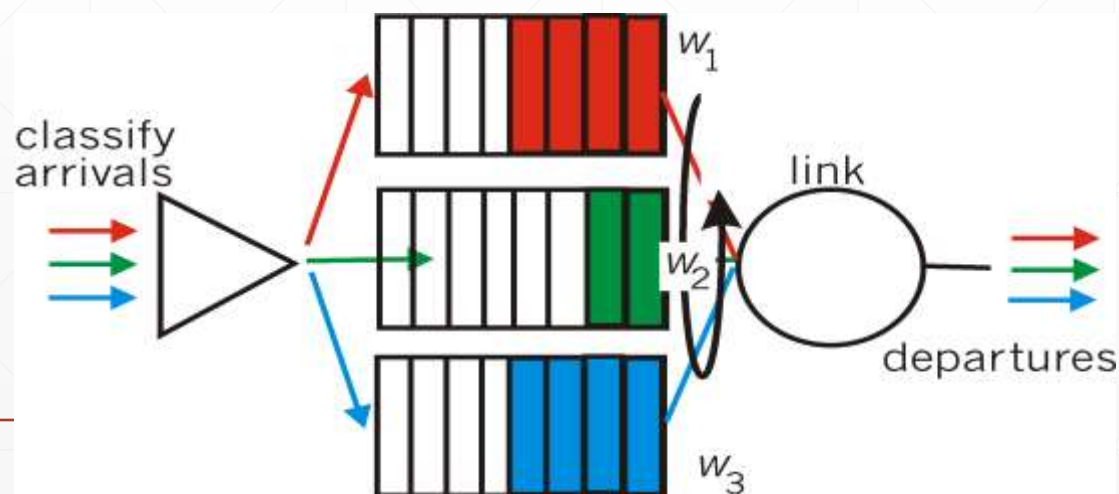
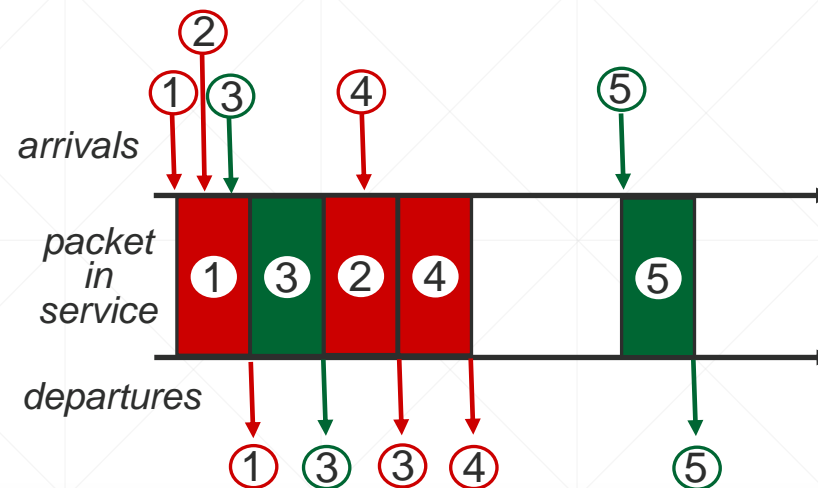
调度策略 – 基于优先级

- 基于优先级的调度：高优先级的排队分组最先发送
- 可能将分组分为多类，每一类可能具有多种优先级
 - 分类的标准可能是特殊标记、或其他首部信息。例如IP源地址、IP目的地址、端口号等
- 真实世界中的例子？



调度策略 – Round Robin

- Round Robin
 - 分组分为多类
 - 循环扫描每一类的队列，从该队列发出一个完整的分组
 - 真实世界的例子？
- Weighted Fair Queueing (WFQ)
 - 实际上是一般性的Round Robin策略
 - 一个循环中，每一类获得不同的发送数据的次数。
 - 真实世界的例子？



IP协议

IP: Internet Protocol

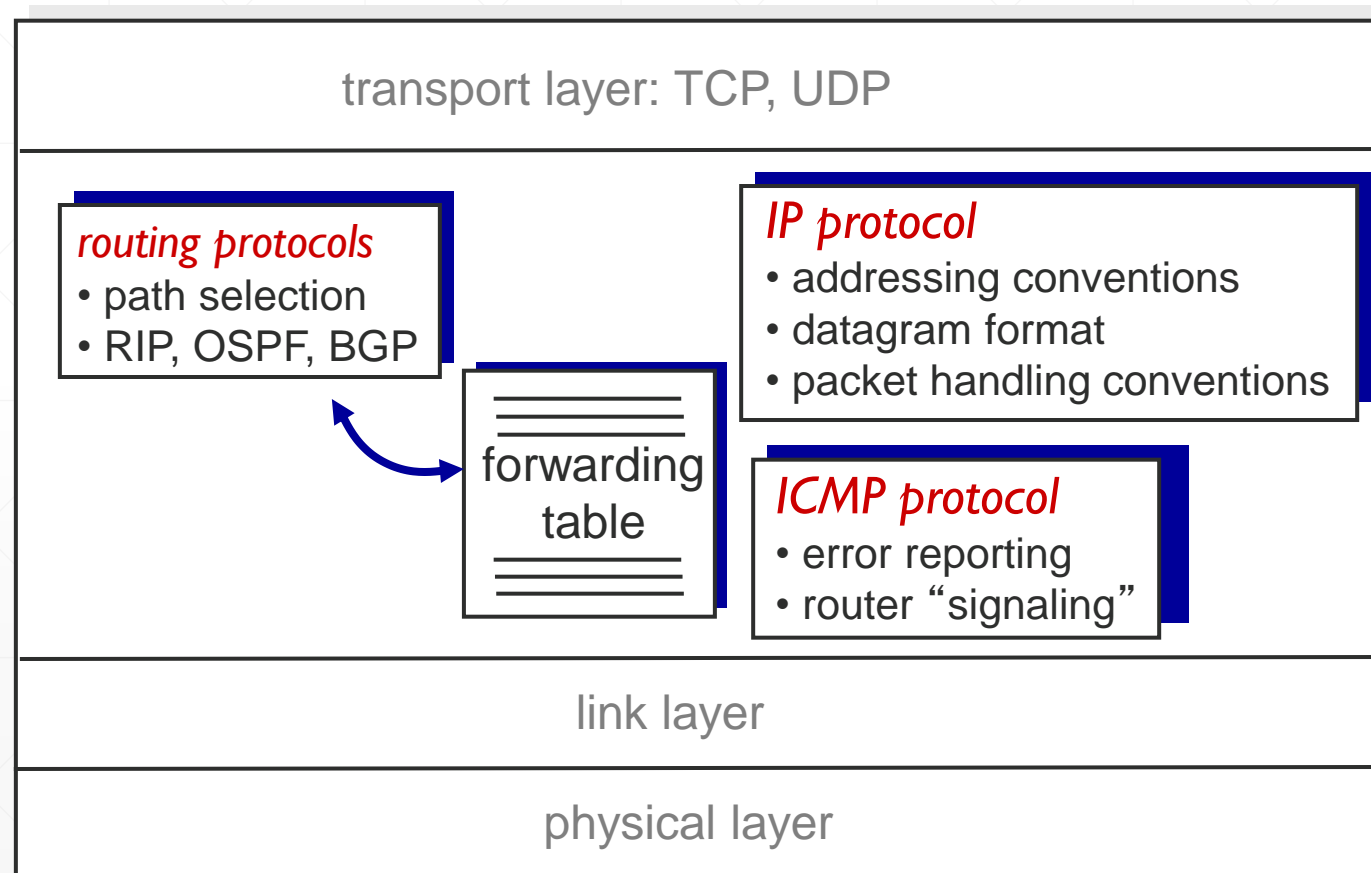
IP

- 网际协议IP是TCP/IP体系中两个最主要的协议之一
 - 网络层的主要协议
 - 本小节包括：
 - IPv4地址
 - 数据报格式
 - 分段和重组
 - NAT
 - IPv6
-

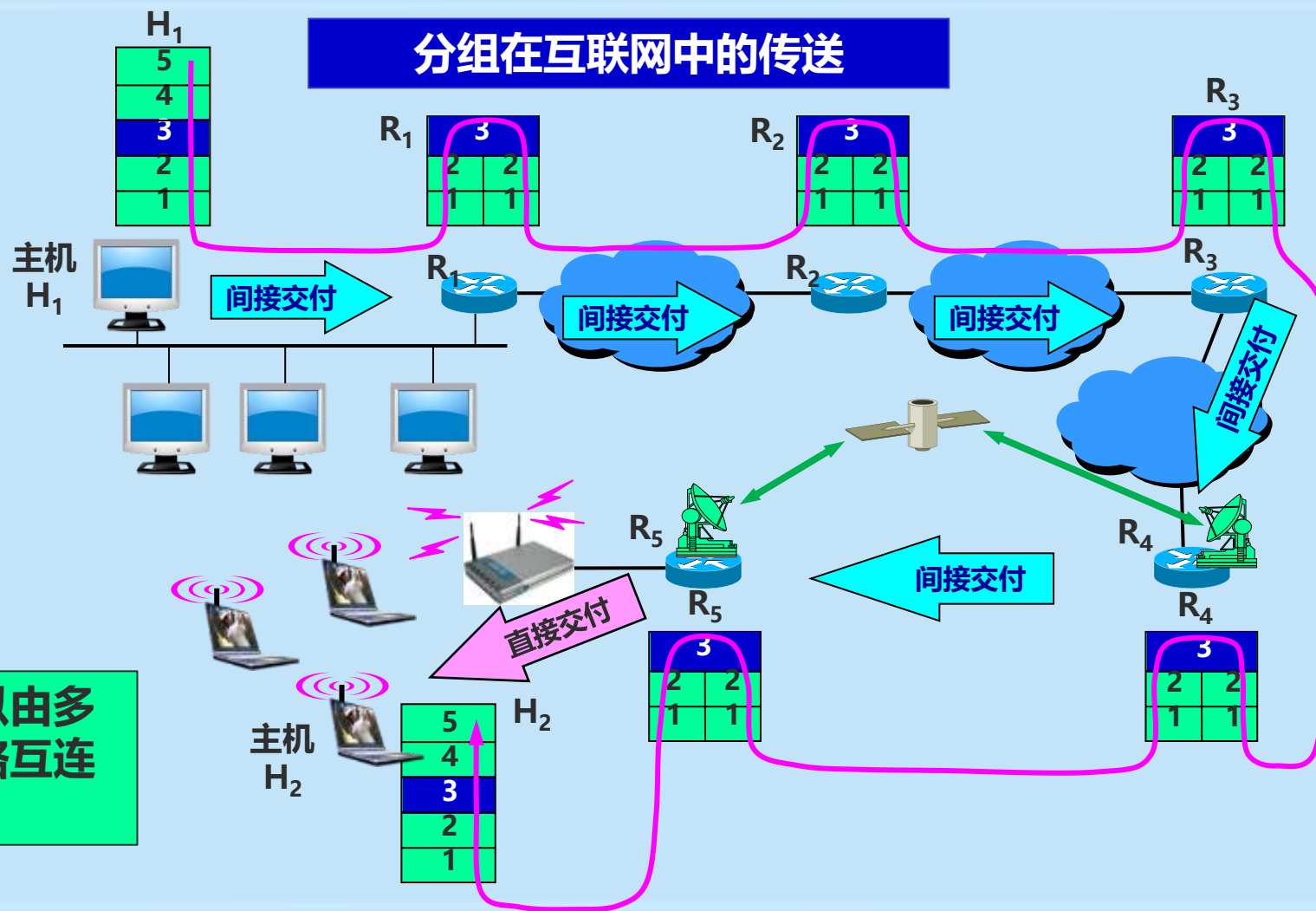
网络层、IP协议

- 网络层的作用
- IP协议
- ICMP协议
- 路由协议

network
layer



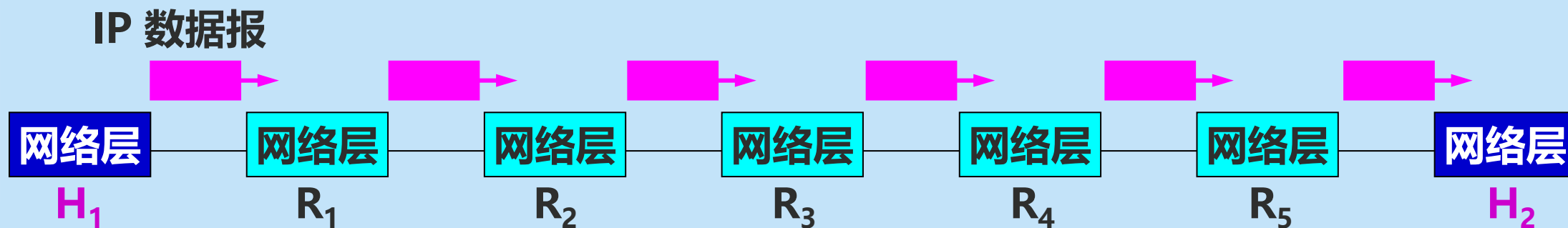
分组在互联网中的传送



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。



IP地址

- 我们把整个互联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。
 - 一般用形如X.X.X.X的方式表示，每个X对应8位的二进制数
 - IP 地址现在由互联网名字和数字分配机构ICANN (Internet Corporation for Assigned Names and Numbers)进行分配。
-

IP地址的编址方法

- 根据时间和划分方式的演进，有三种编址方法：
 - **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 (subnetting)**。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 (CIDR)**。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。
 - 现在广泛使用的是基于CIDR的编址方法。
-

分类IP地址

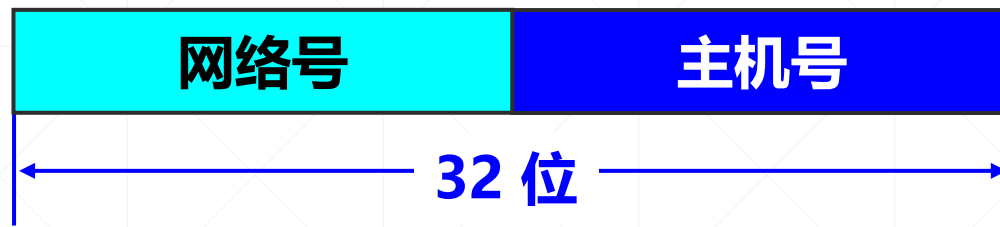


分类 IP 地址

- 将IP地址划分为若干个固定类。
 - 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：

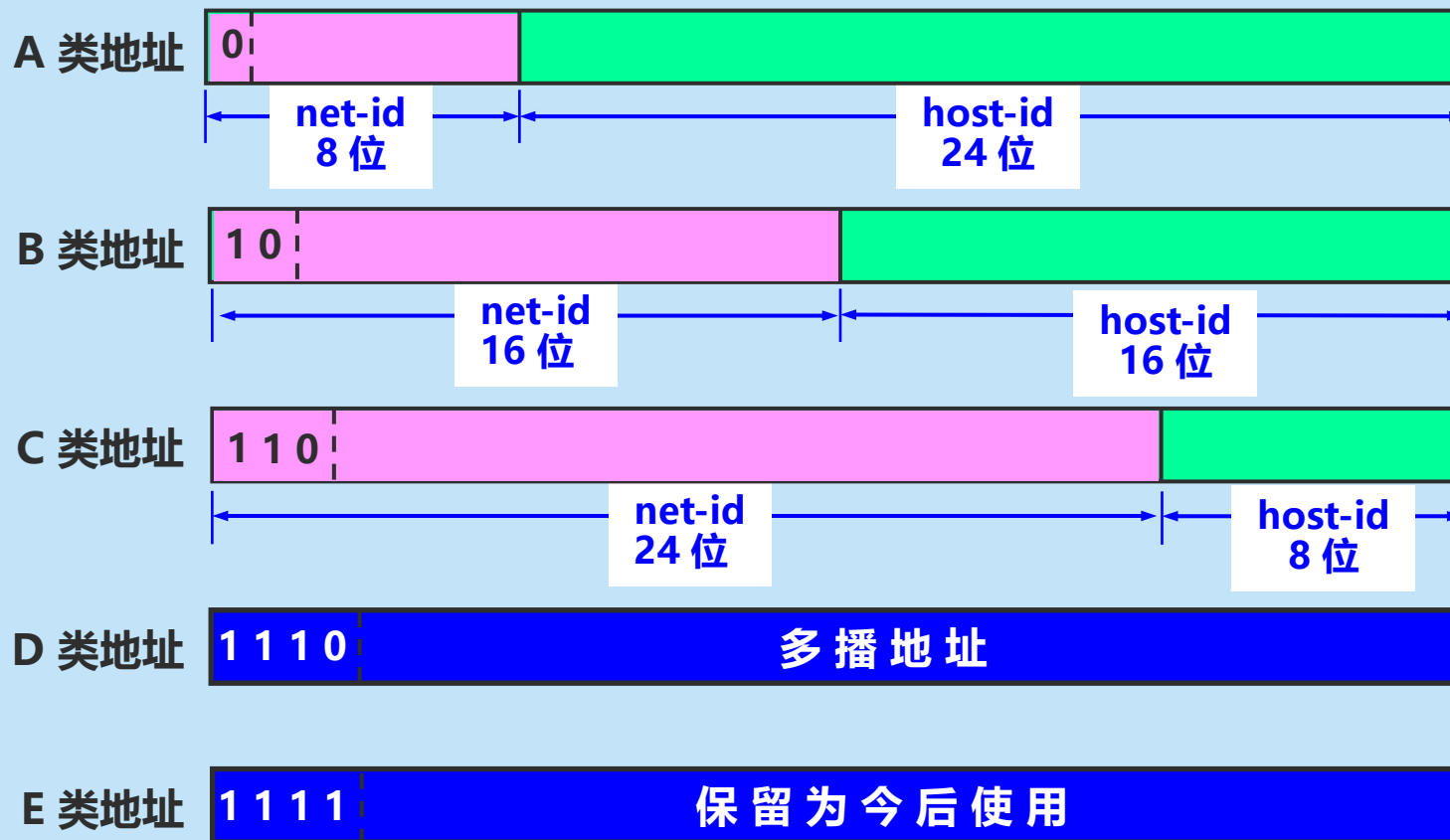


- 这种两级的 IP 地址可以记为：

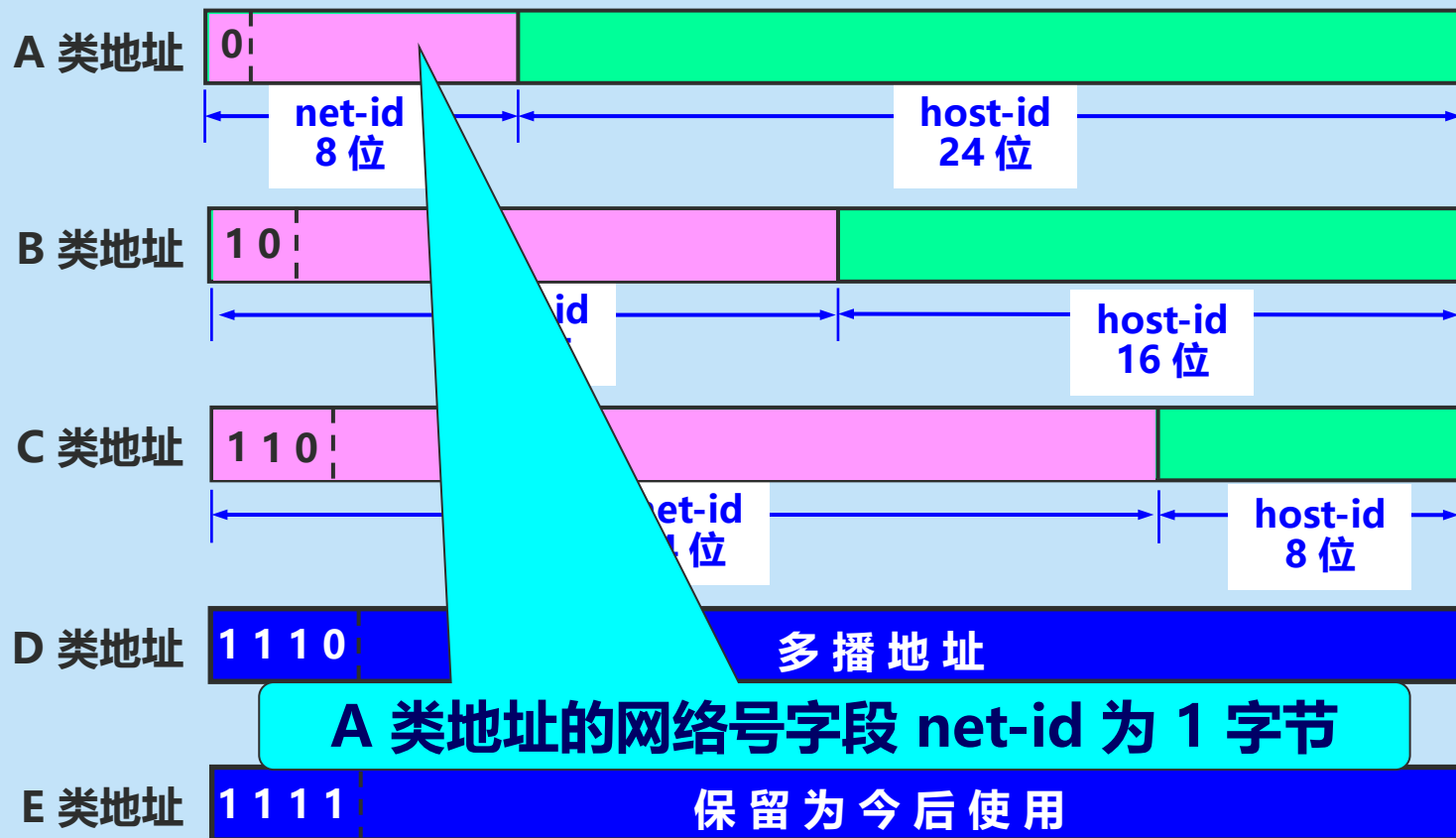
IP 地址 ::= { <网络号>, <主机号> }

::= 代表 “定义为”

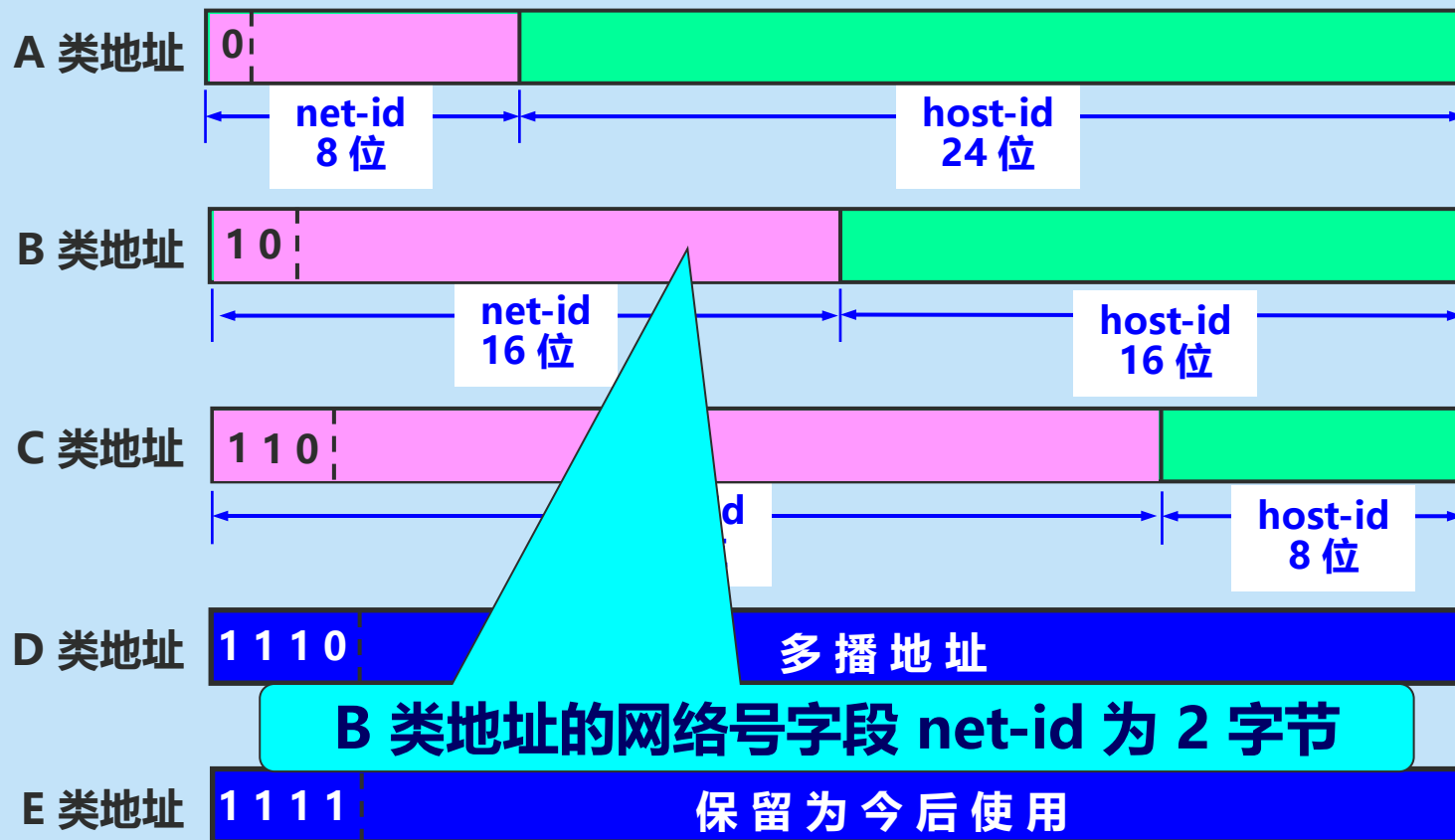
各类 IP 地址的网络号字段和主机号字段



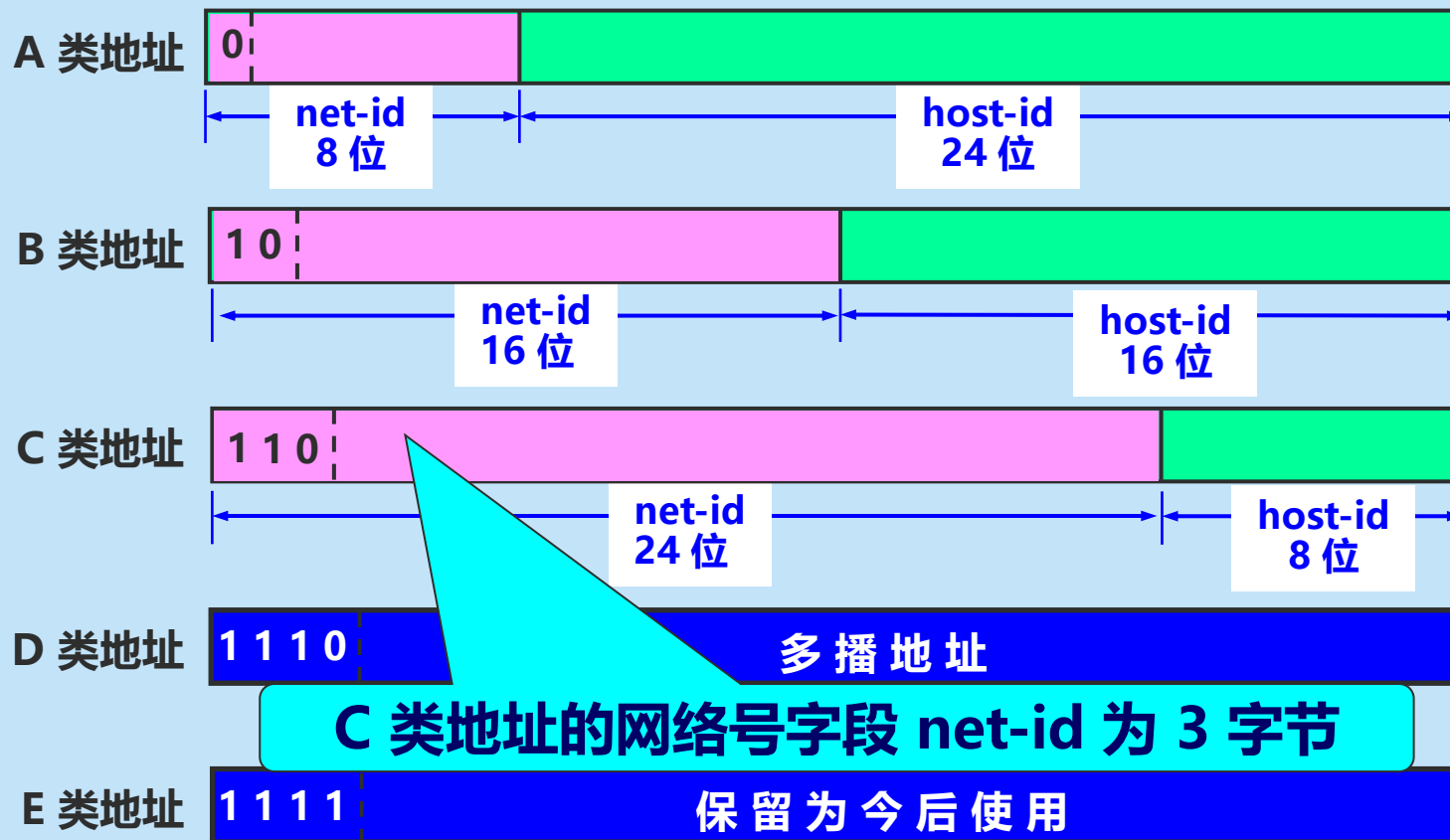
各类 IP 地址的网络号字段和主机号字段



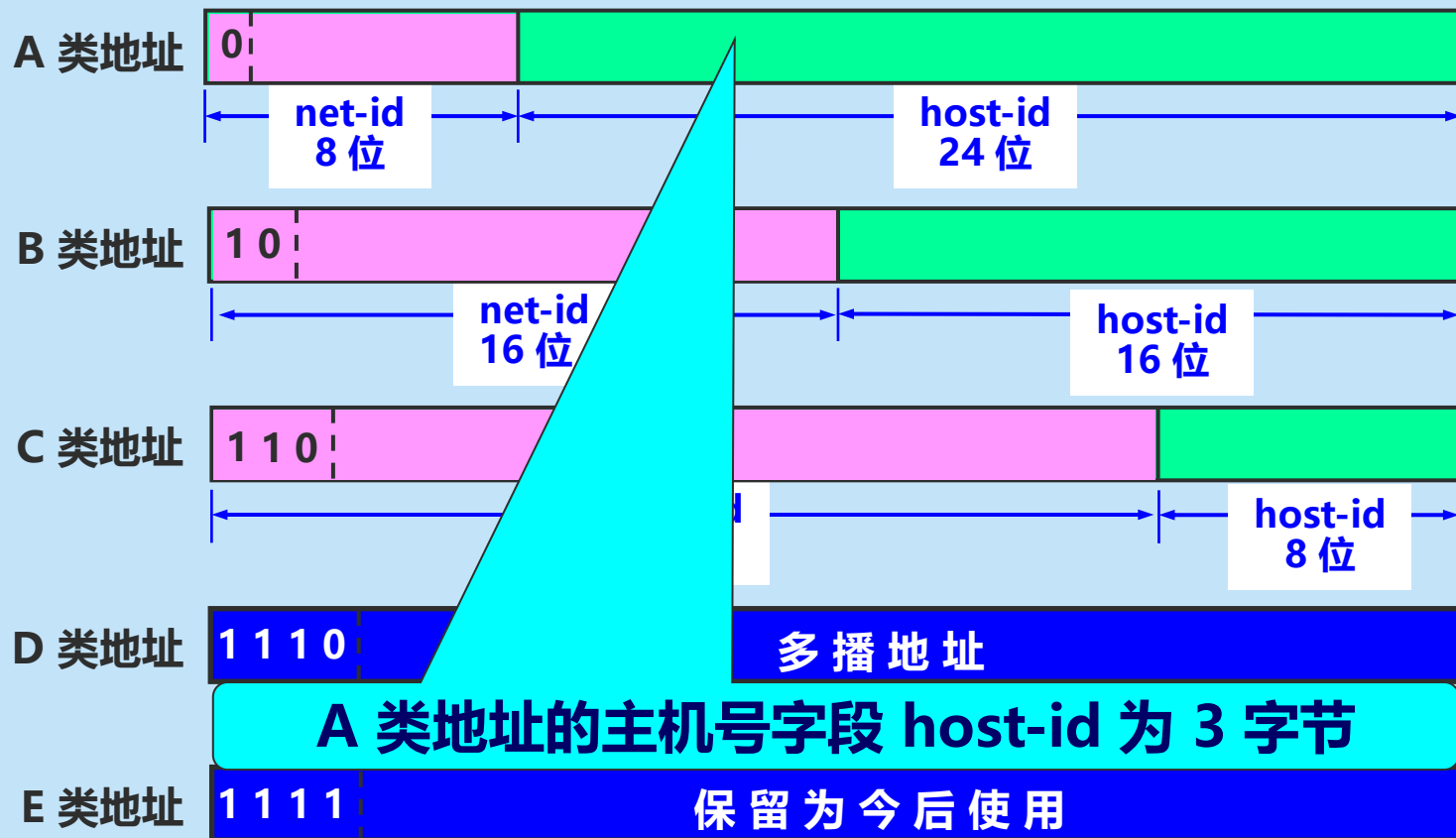
各类 IP 地址的网络号字段和主机号字段



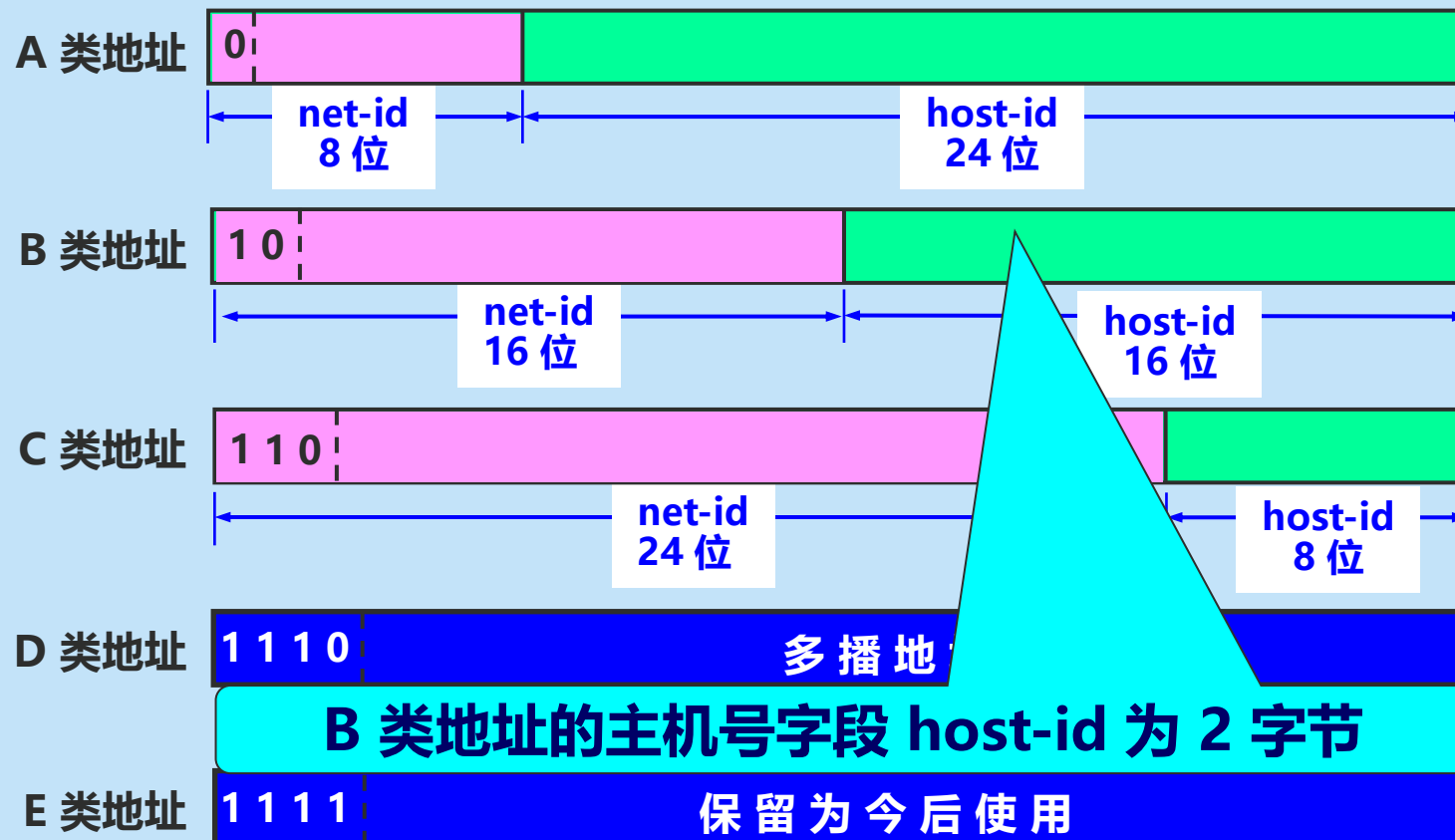
各类 IP 地址的网络号字段和主机号字段



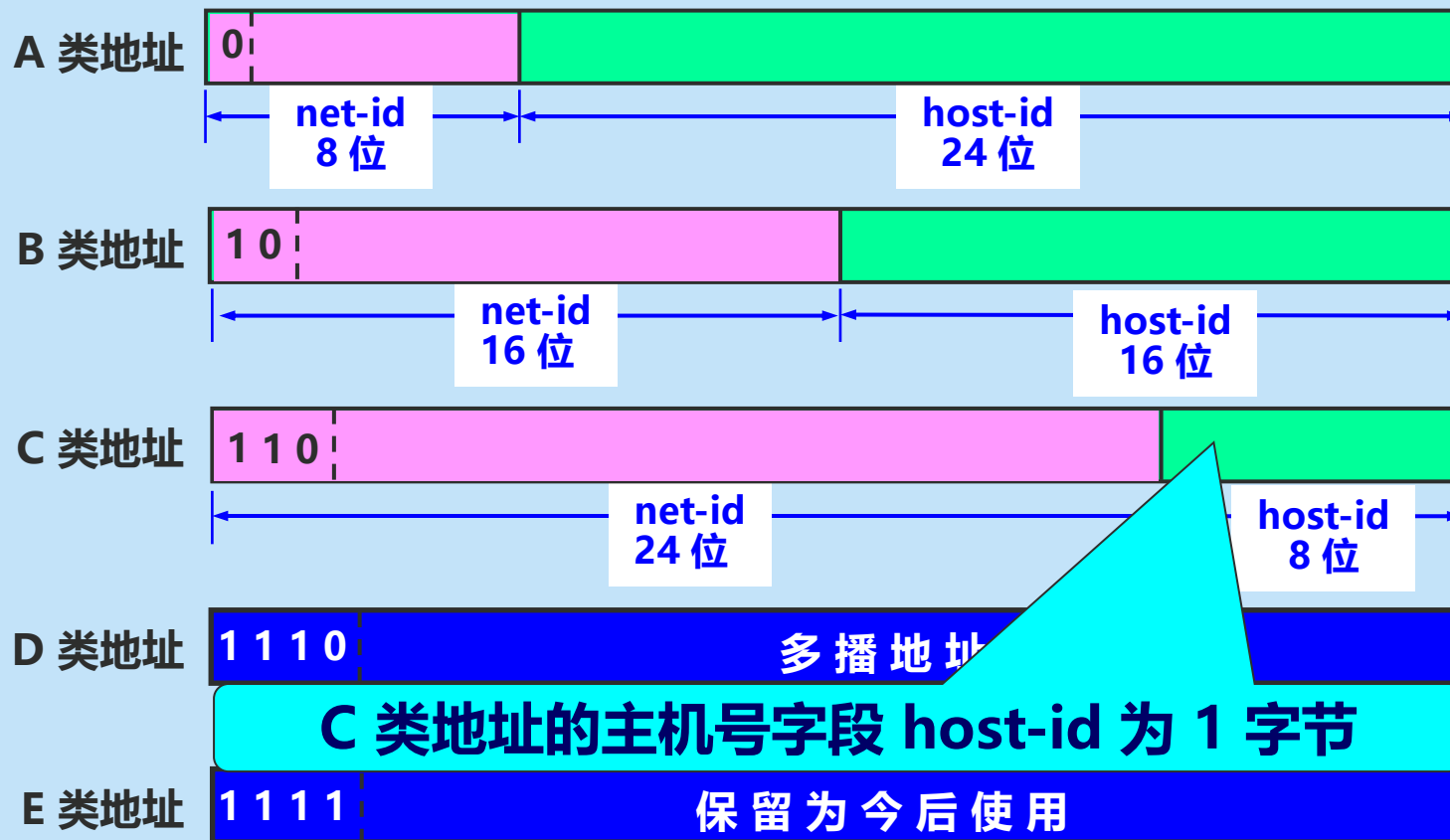
各类 IP 地址的网络号字段和主机号字段



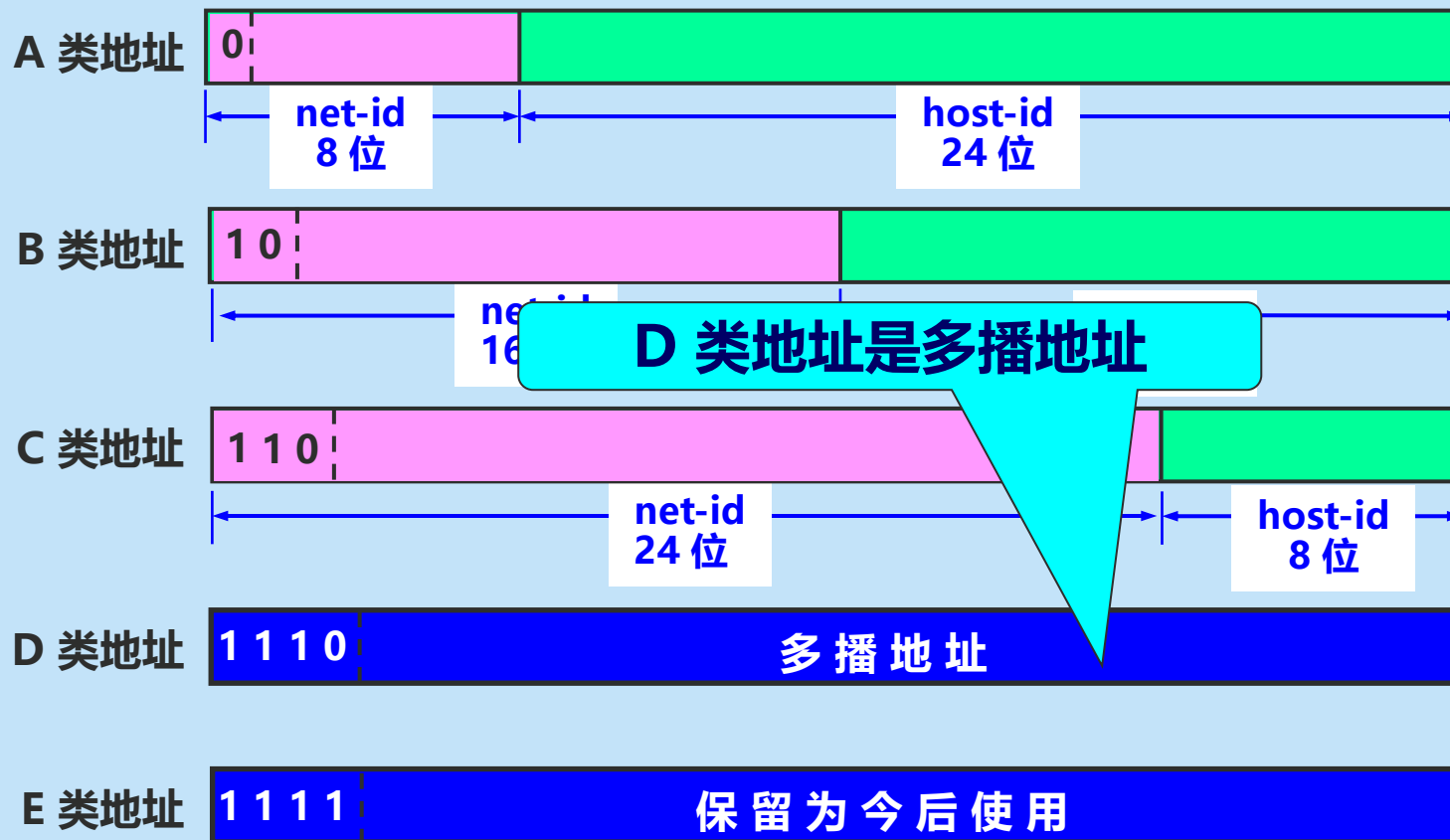
各类 IP 地址的网络号字段和主机号字段



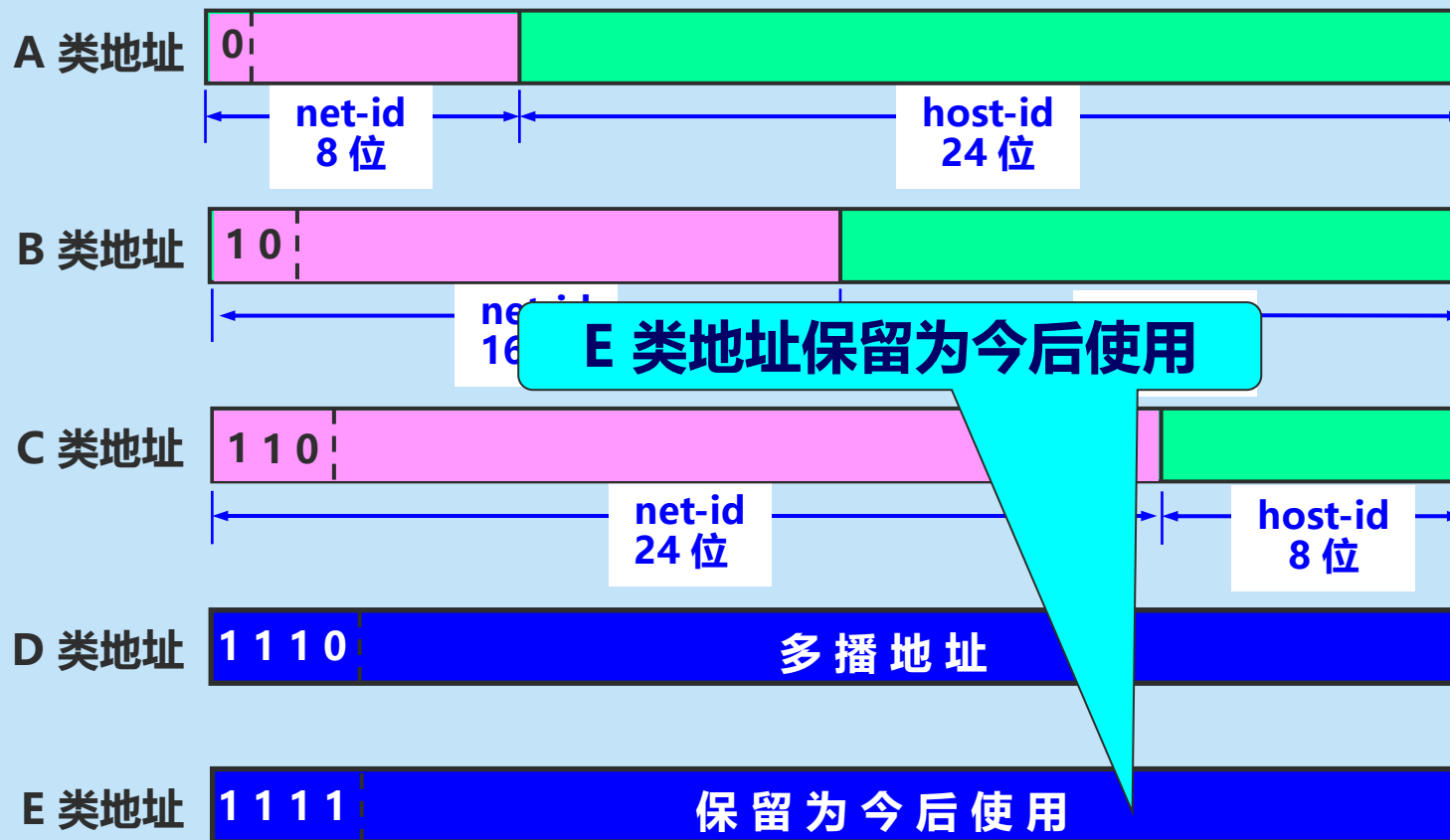
各类 IP 地址的网络号字段和主机号字段



各类 IP 地址的网络号字段和主机号字段



各类 IP 地址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 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
B	16383 ($2^{14} - 1$)	128.1	191.255	65534
C	2097151 ($2^{21} - 1$)	192.0.1	223.255.255	254

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（用于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	用于本地软件环回测试

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

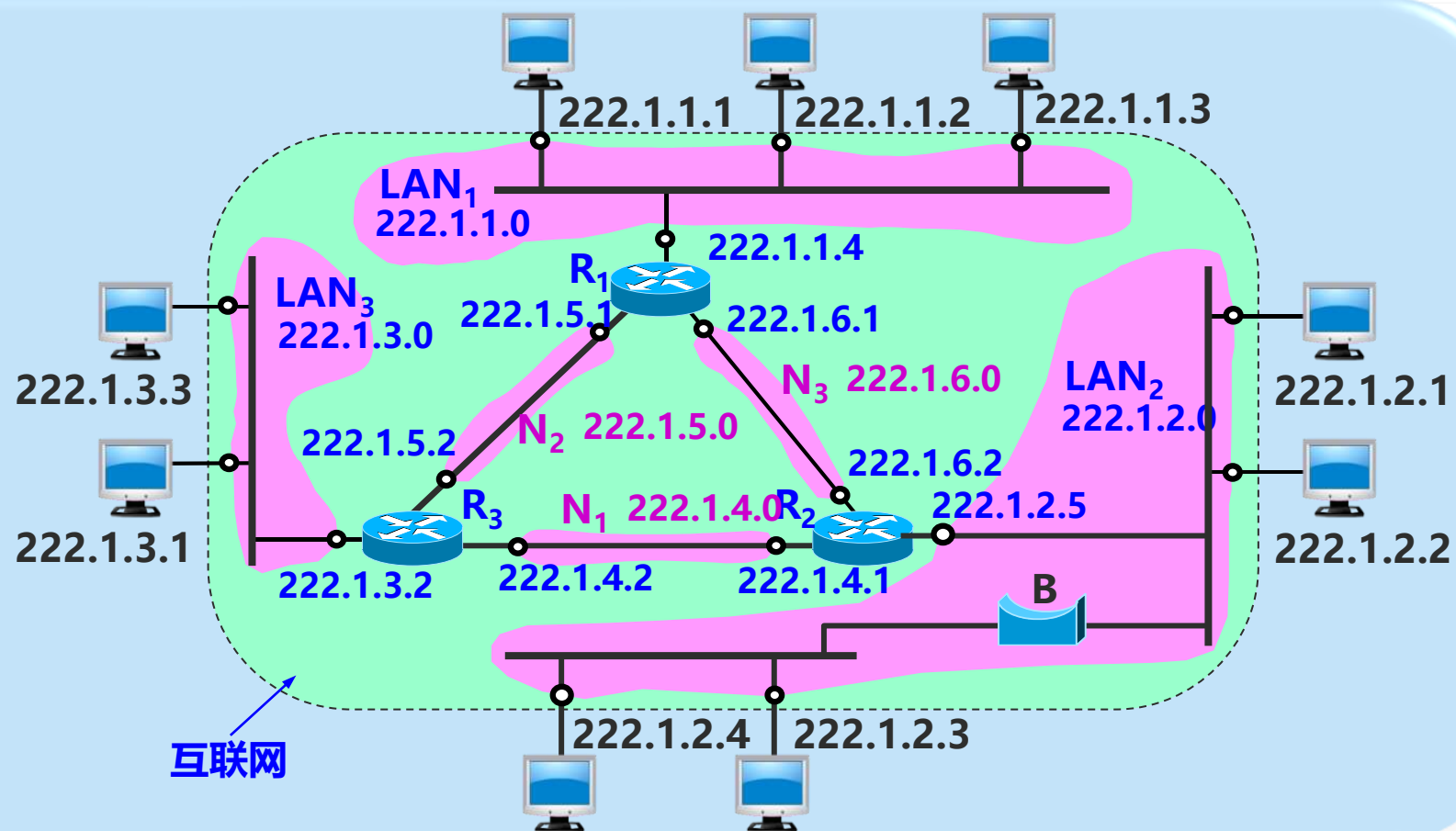
(2) 实际上 IP 地址是标志一个主机（或路由器）和一条链路的接口。

- 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

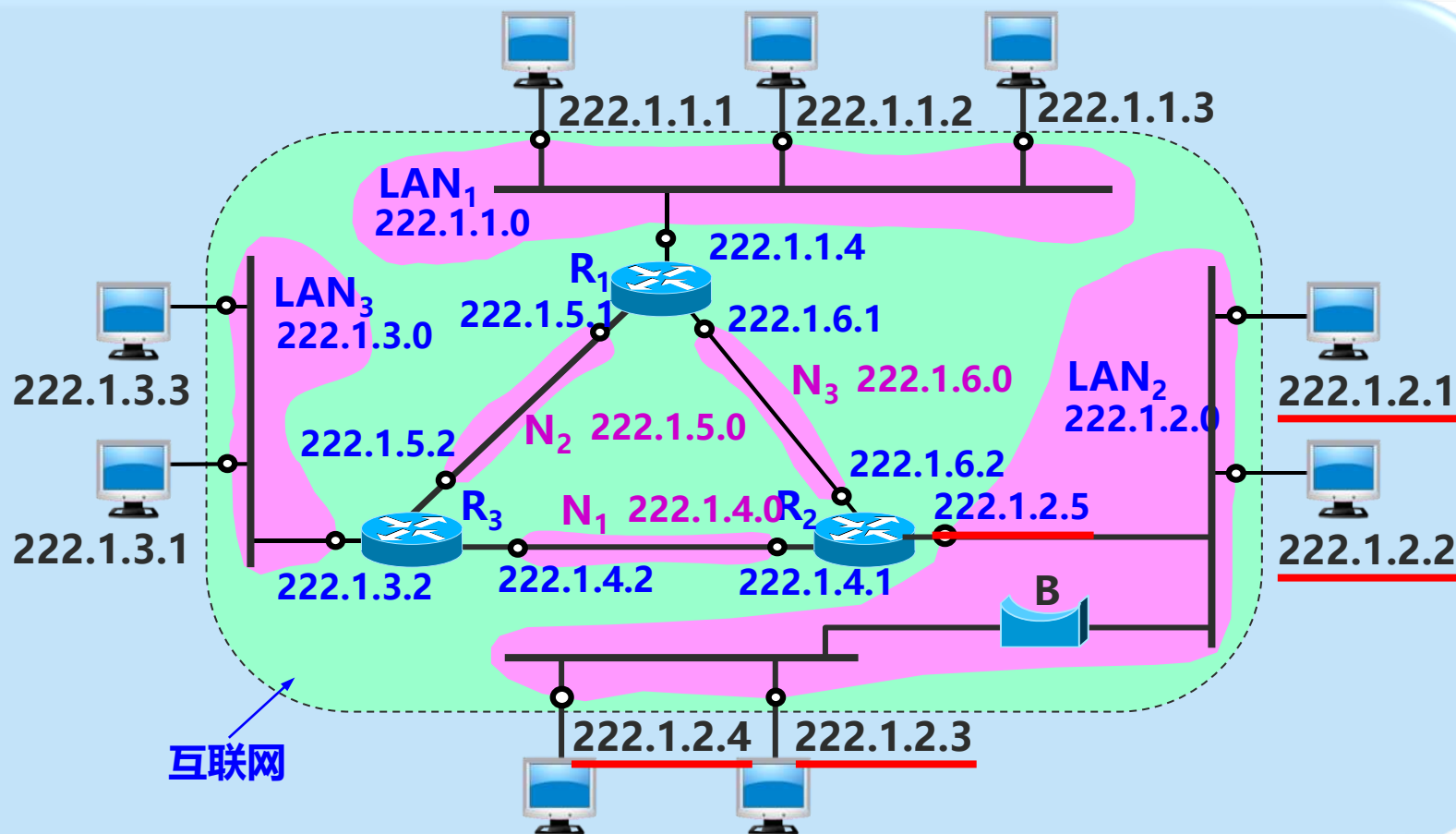
- (3) 用转发器或网桥连接起来的若干个局域网仍为一个网络，因此这些局域网都具有同样的网络号 net-id。
 - (4) 所有分配到网络号 net-id 的网络，无论是范围很小的局域网，还是可能覆盖很大地理范围的广域网，都是平等的。
-

互联网中的 IP 地址



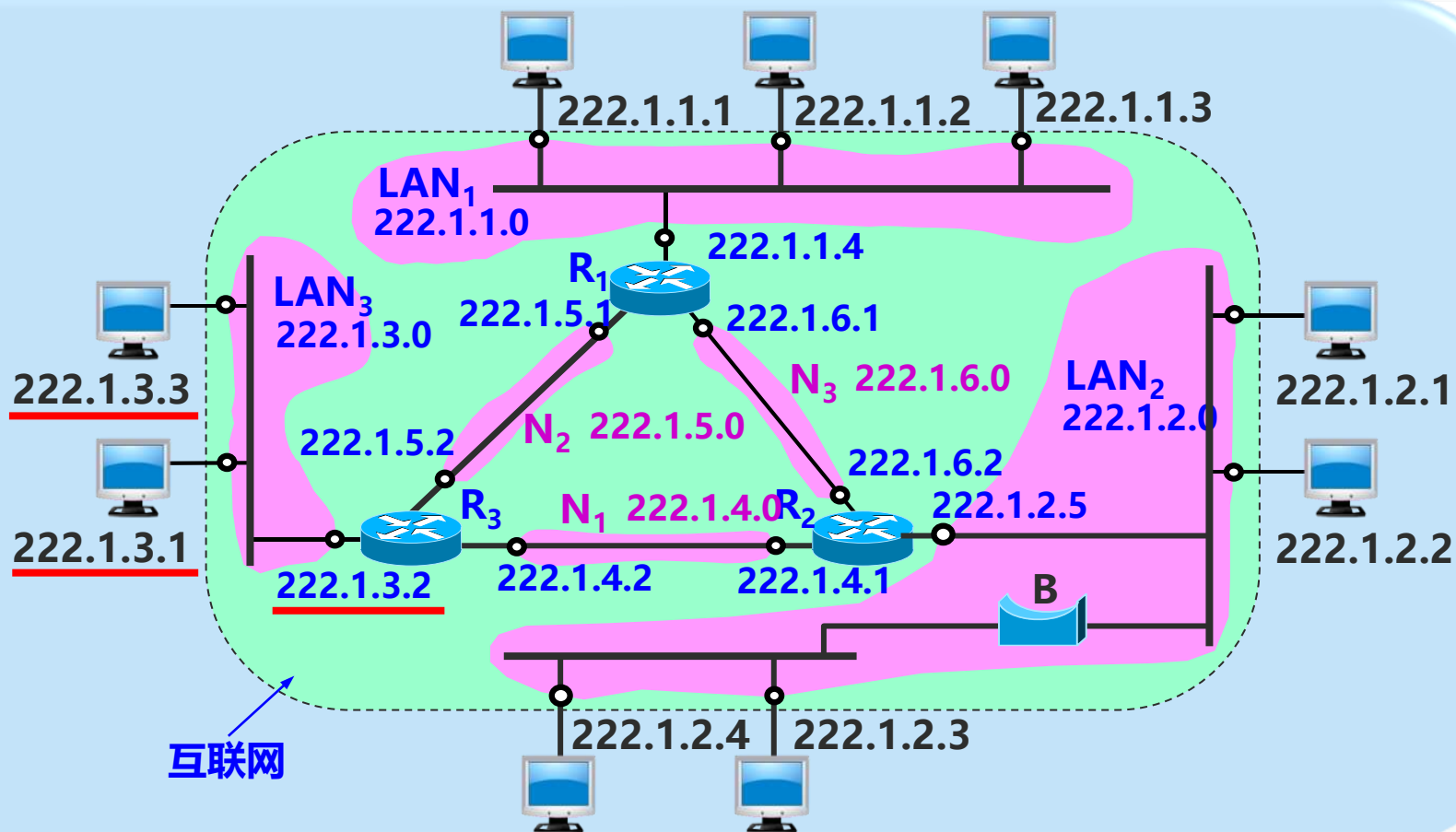
互联网中的 IP 地址

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。



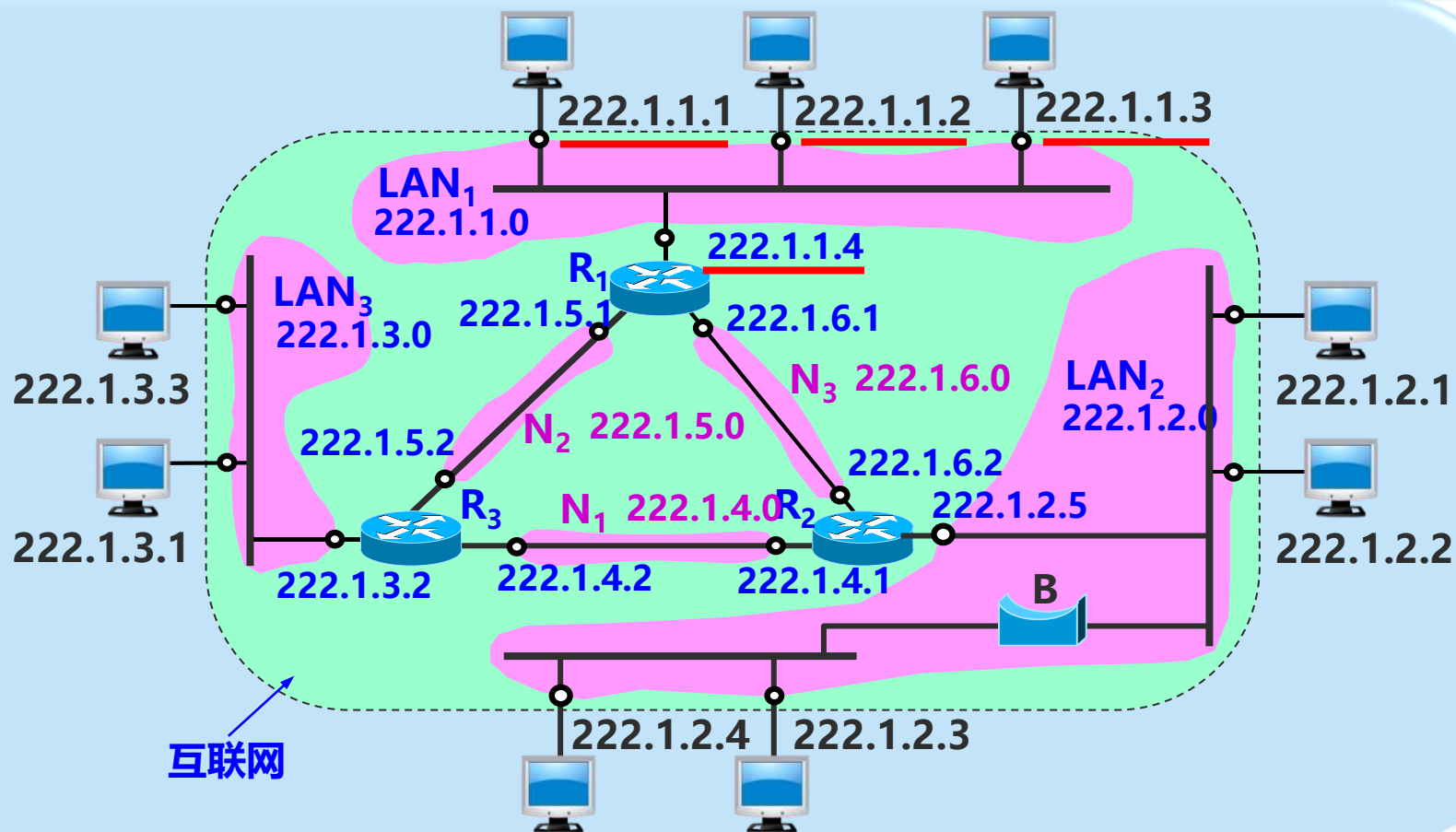
互联网中的 IP 地址

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。



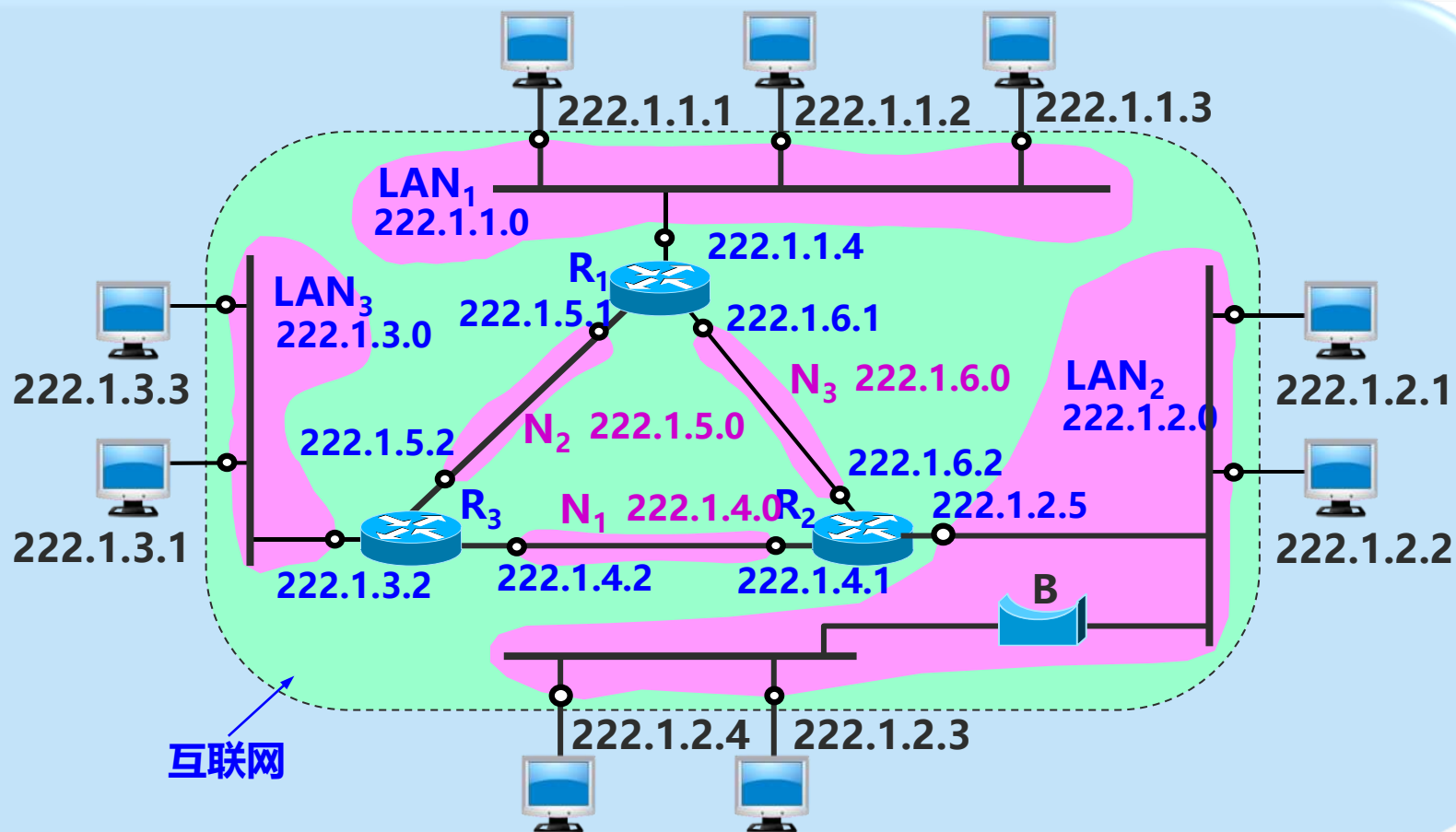
互联网中的 IP 地址

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。



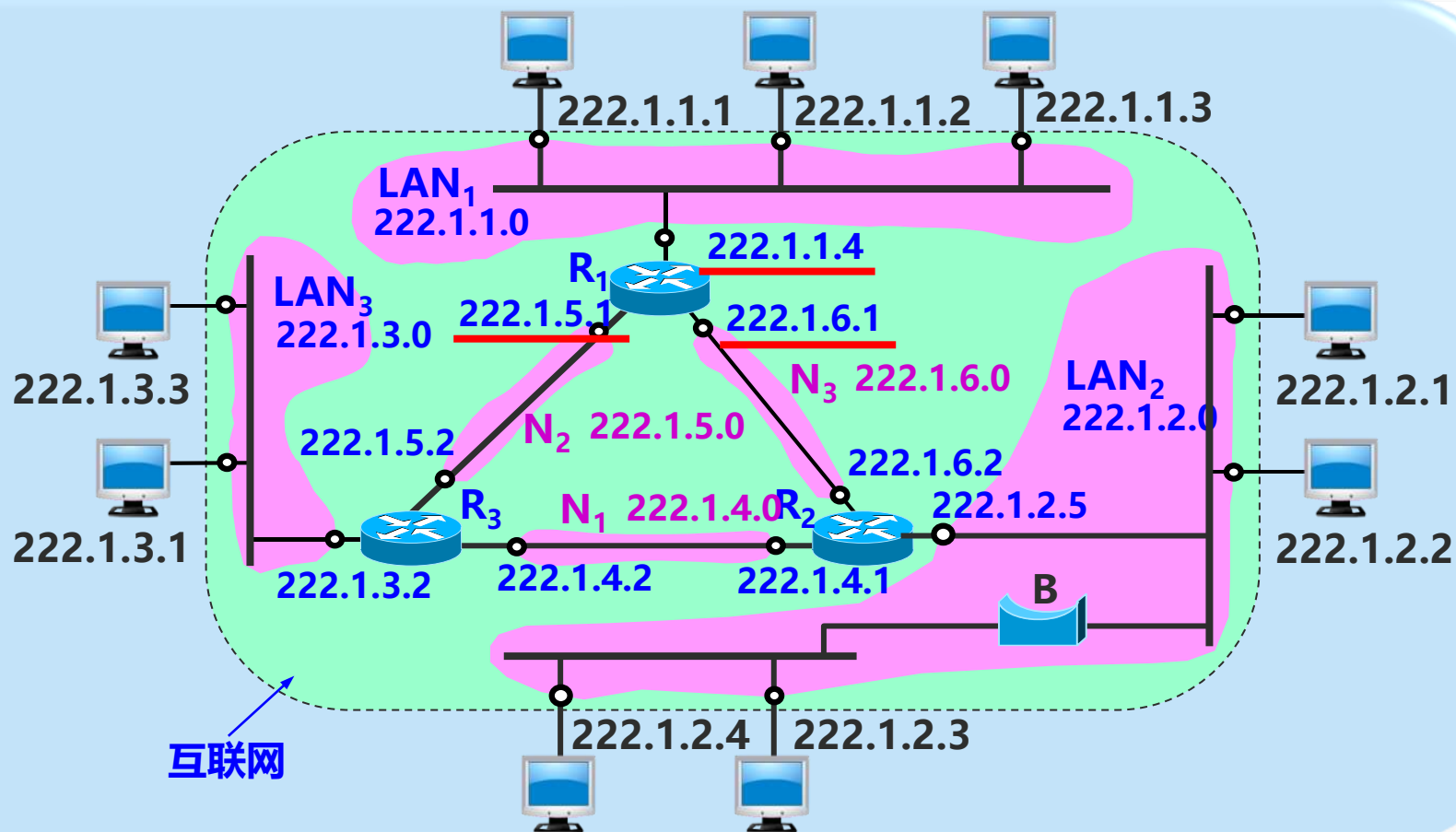
互联网中的 IP 地址

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。



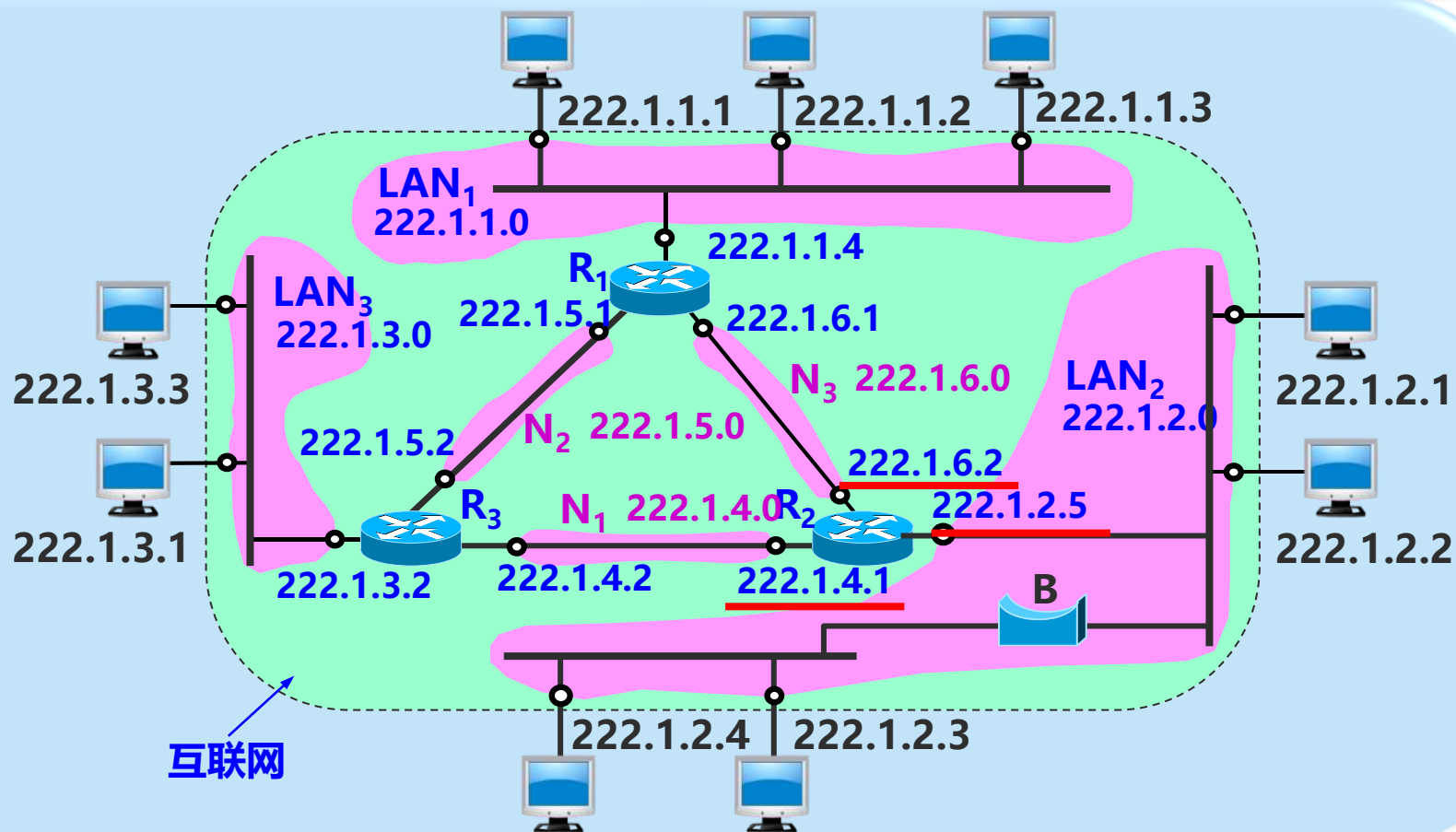
互联网中的 IP 地址

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。



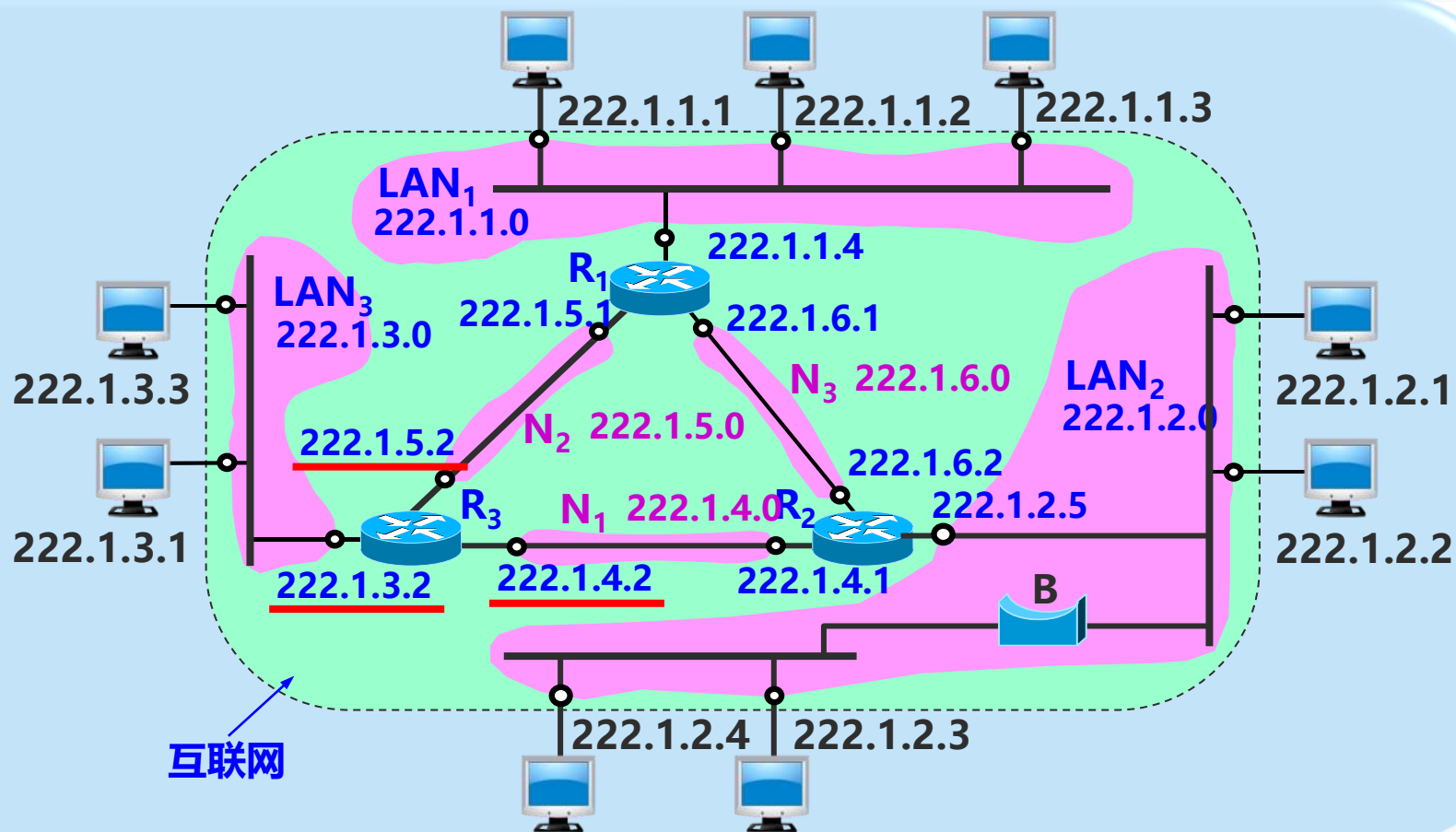
互联网中的 IP 地址

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。



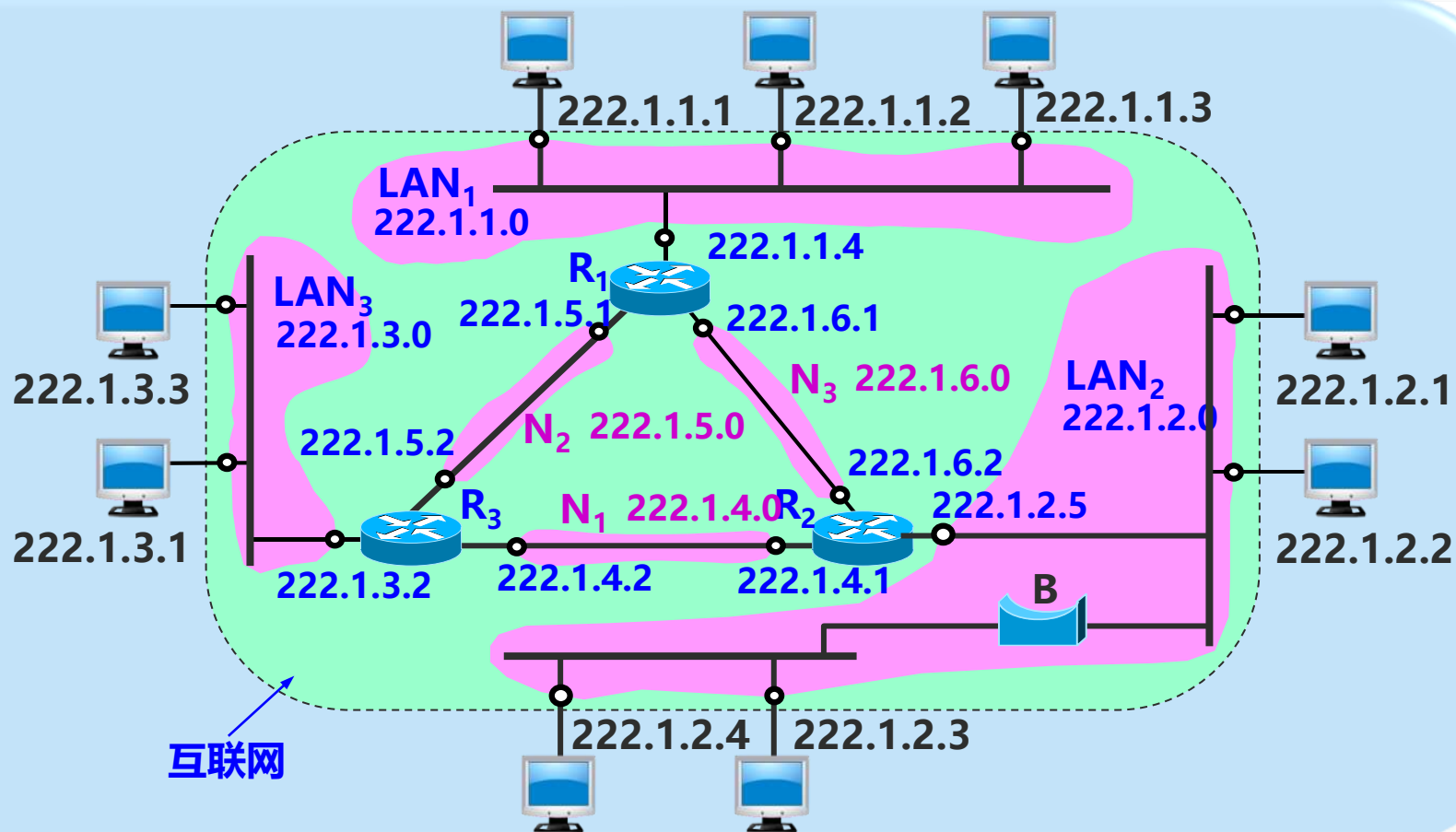
互联网中的 IP 地址

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。



互联网中的 IP 地址

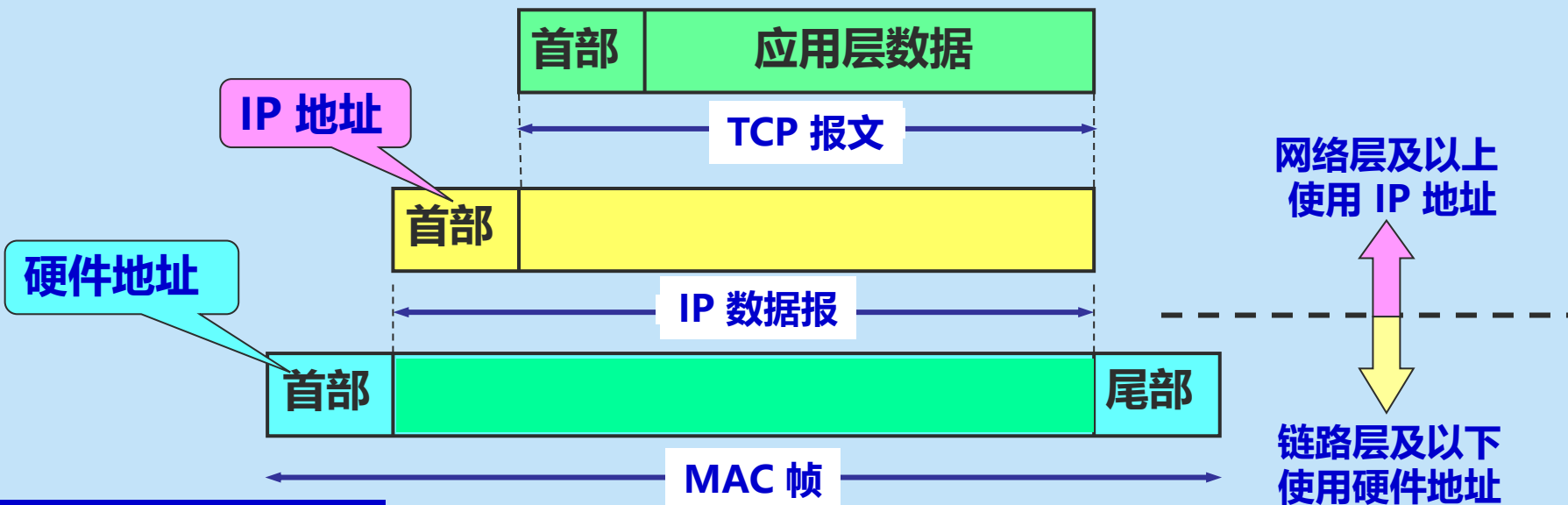
两个路由器直接相连的接口处，可指明也可不指明IP地址。如指明IP地址，则这一段连线就构成了一种只包含一段线路的特殊“网络”。现在常不指明IP地址。



IP 地址与硬件地址

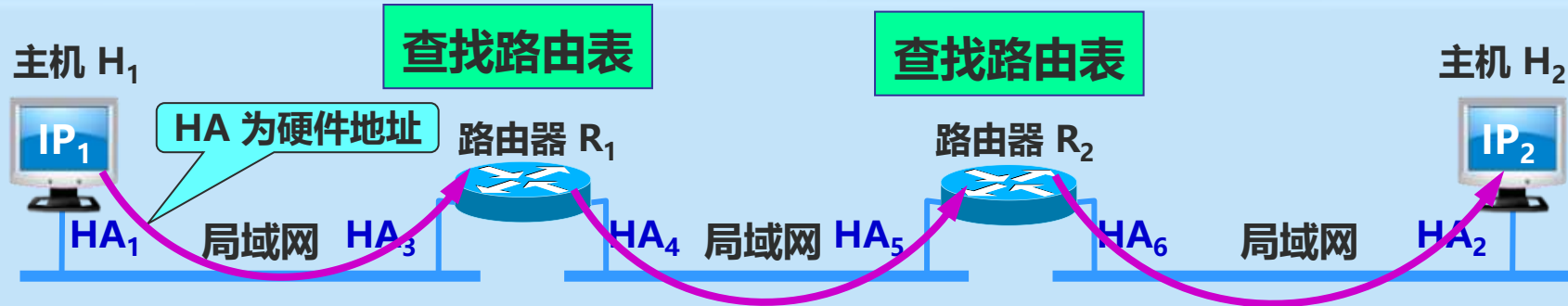
- IP 地址与硬件地址是不同的地址。
 - 从层次的角度看，
 1. **硬件地址（或物理地址）**是数据链路层和物理层使用的地址。
 2. **IP 地址**是网络层和以上各层使用的地址，是一种逻辑地址（称 IP 地址是逻辑地址是因为 IP 地址是用软件实现的）。
-

IP 地址与硬件地址



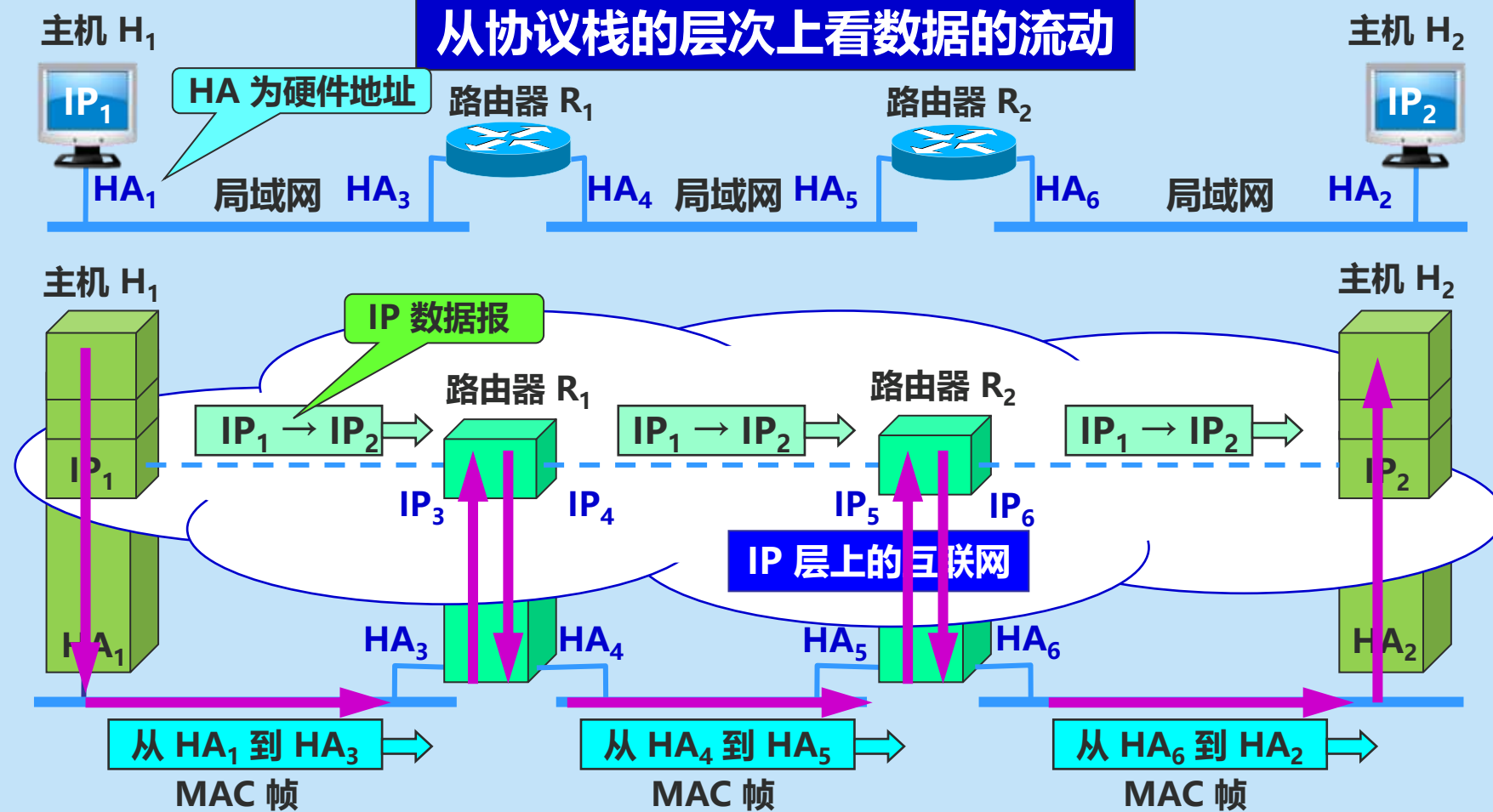
IP 地址放在 IP 数据报的首部，而硬件地址则放在 MAC 帧的首部。

IP 地址与硬件地址的区别

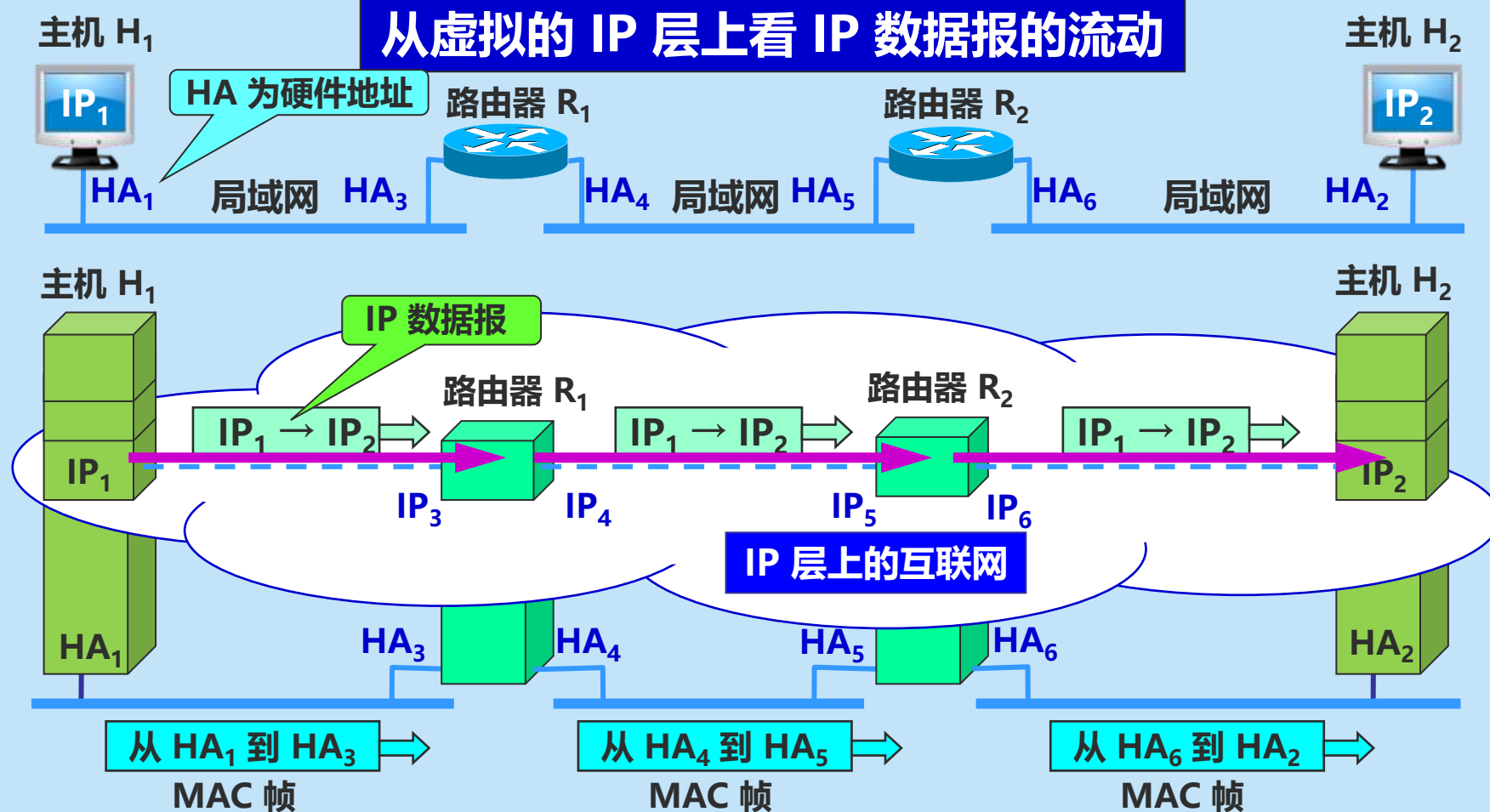


通信的路径：
 $H_1 \rightarrow$ 经过 R_1 转发 \rightarrow 再经过 R_2 转发 $\rightarrow H_2$

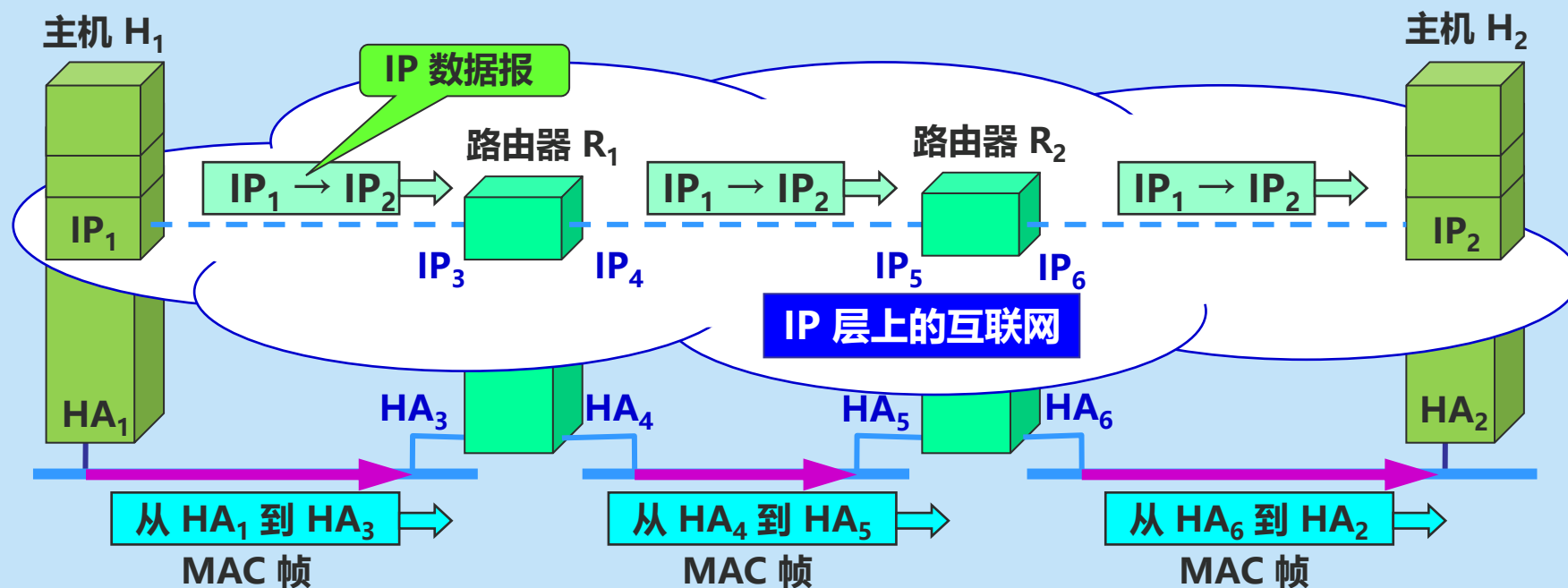
从协议栈的层次上看数据的流动



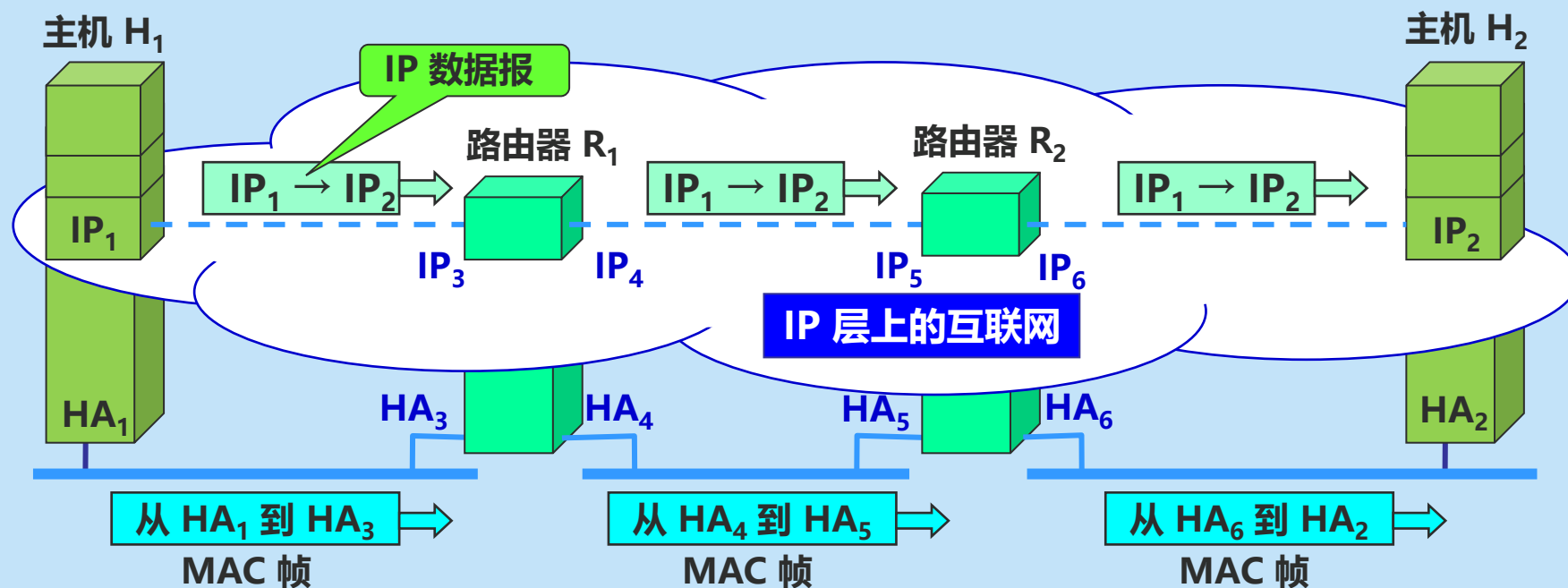
从虚拟的 IP 层上看 IP 数据报的流动



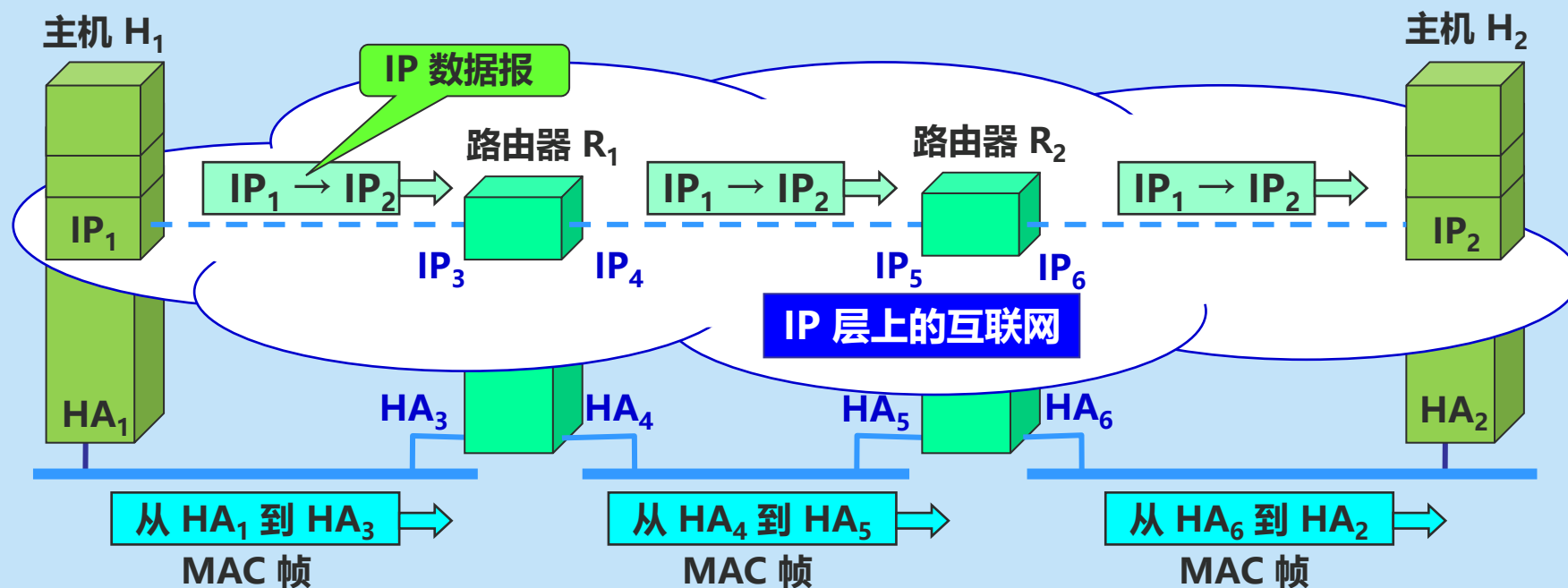
在链路上看 MAC 帧的流动



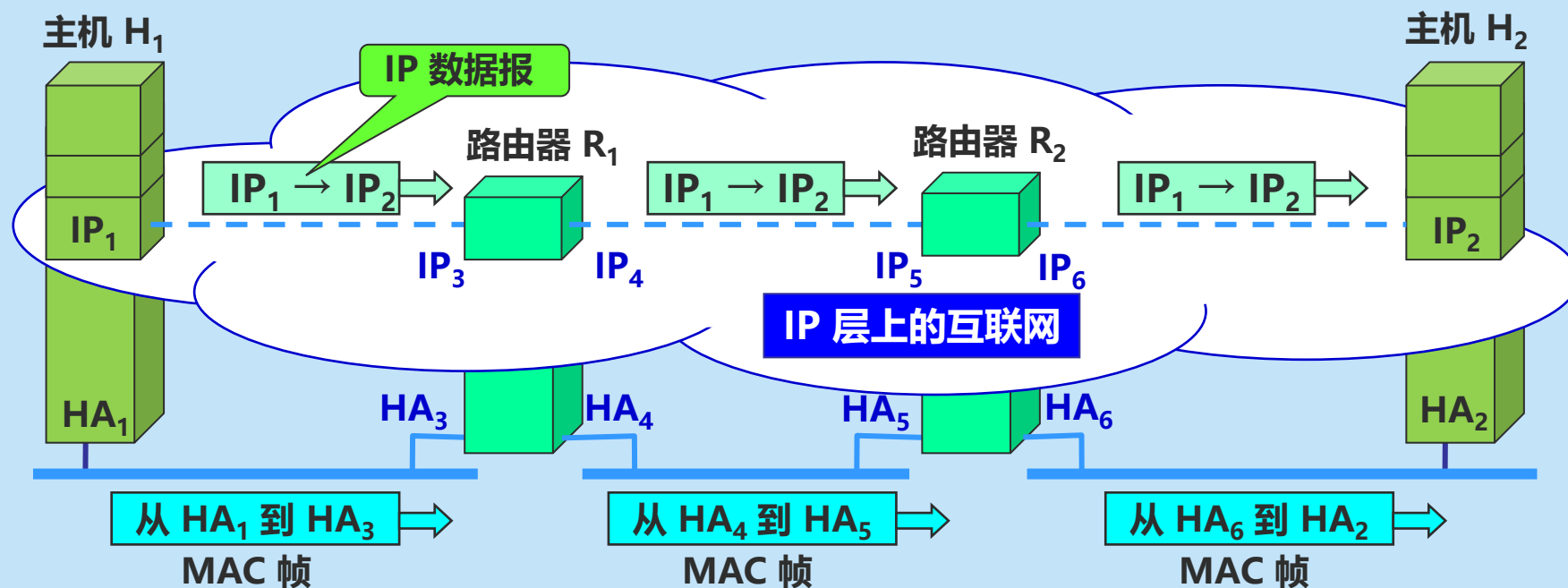
在 IP 层抽象的互联网上只能看到 IP 数据报。
图中的 $IP_1 \rightarrow IP_2$ 表示从源地址 IP_1 到目的地址 IP_2 。
两个路由器的 IP 地址并不出现在 IP 数据报的首部中。



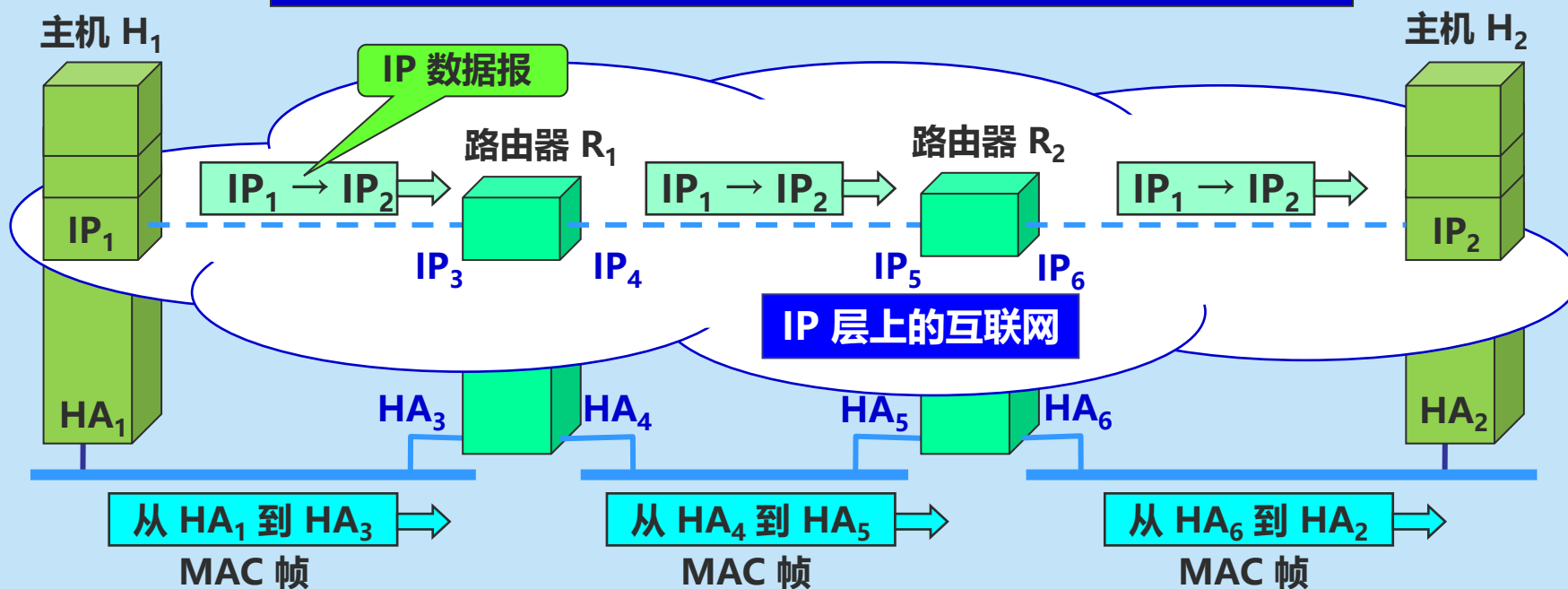
路由器只根据目的站的 IP 地址的网络号进行路由选择。



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。



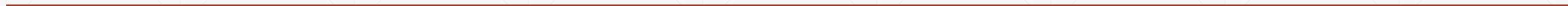
主机 H1 与 H2 通信中使用的IP地址 与 硬件地址 HA

	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

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。
 - **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。**
-

子网划分



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：

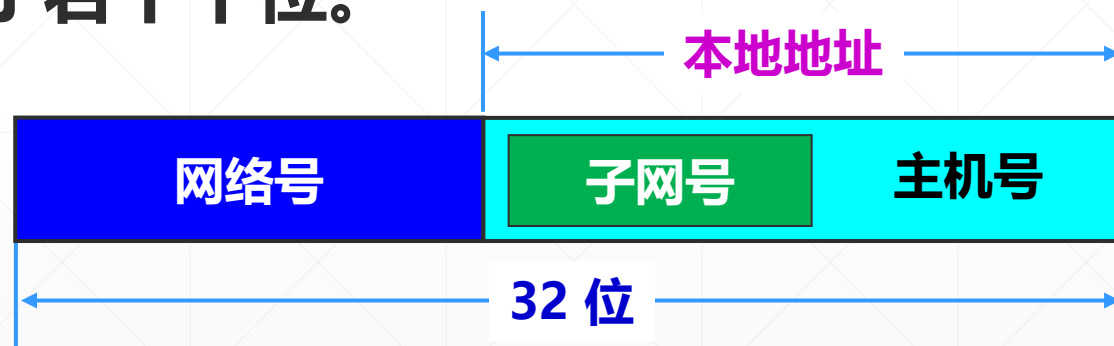
- (1) IP 地址空间的利用率有时很低。
 - (2) 给每一个物理网络分配一个网络号会使路由表变得太大因而使网络性能变坏。
 - (3) 两级的 IP 地址不够灵活。
-

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)。
 - 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。

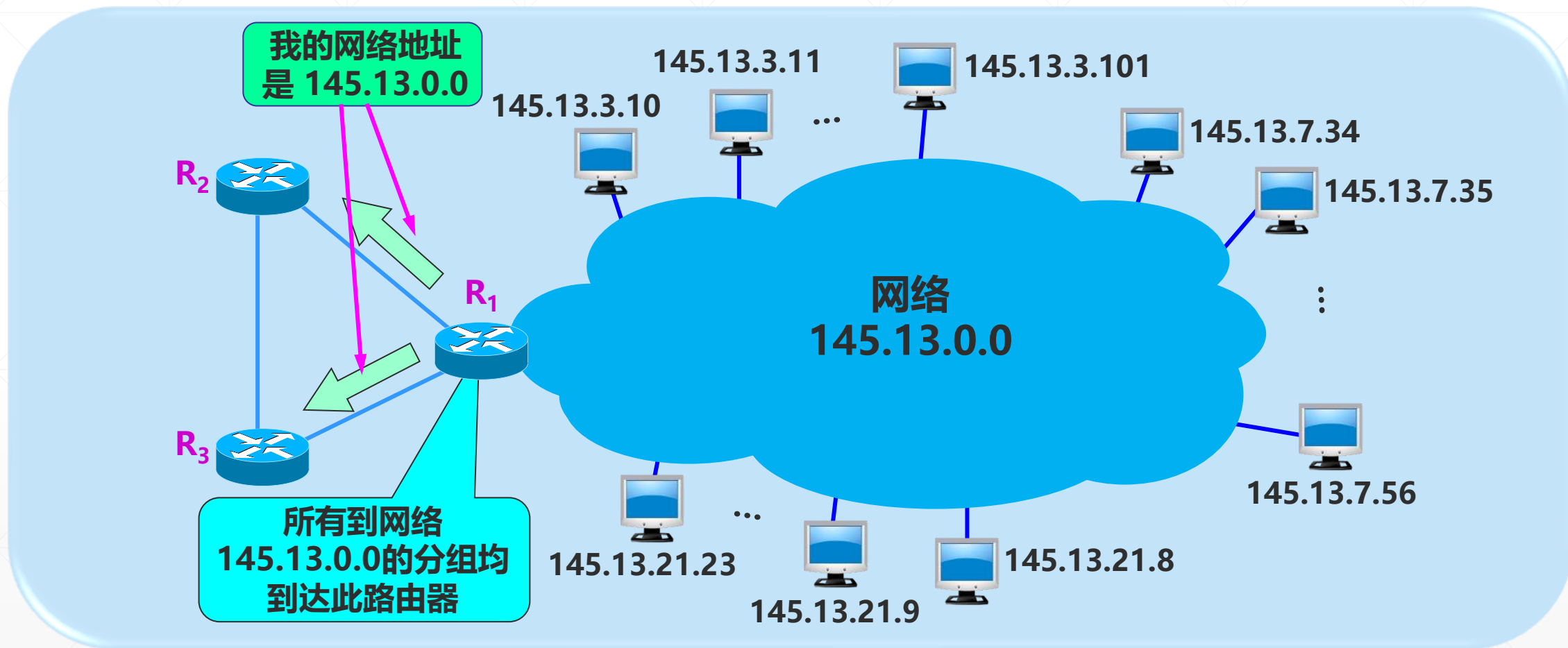


IP地址 ::= {<网络号>, <子网号>, <主机号>}

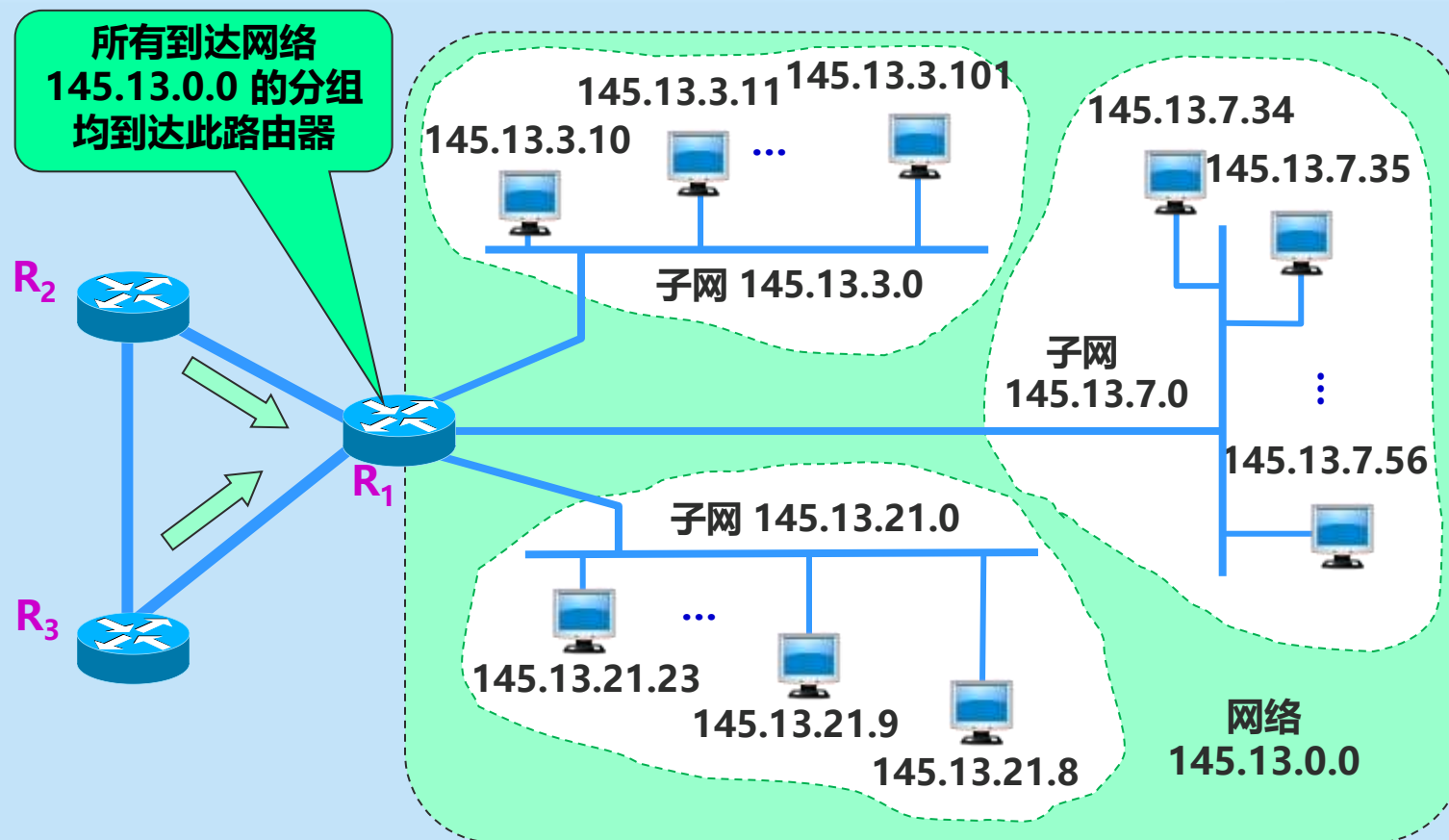
划分子网的基本思路（续）

- 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。
-

一个未划分子网的 B 类网络145.13.0.0

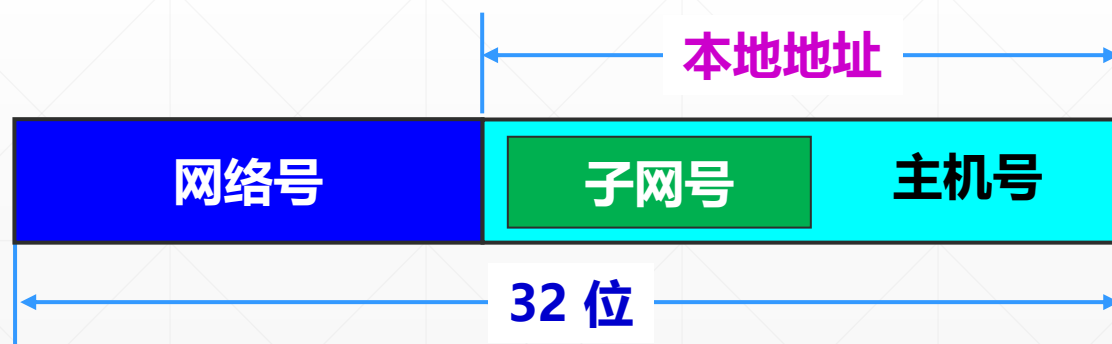


划分为三个子网后对外仍是一个网络



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。



划分子网后变成了三级结构

- **优点**

1. 减少了 IP 地址的浪费
2. 使网络的组织更加灵活
3. 更便于维护和管理

- **划分子网纯属一个单位内部的事情，对外部网络透明，对外仍然表现为没有划分子网的一个网络。**
-

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：

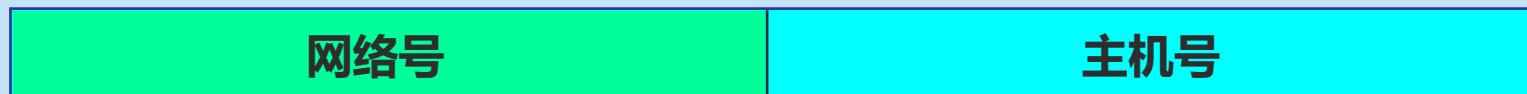
- 子网掩码长度 = 32 位
 - 子网掩码左边部分的一连串 1，对应于网络号和子网号
 - 子网掩码右边部分的一连串 0，对应于主机号
-

IP 地址的各字段和子网掩码

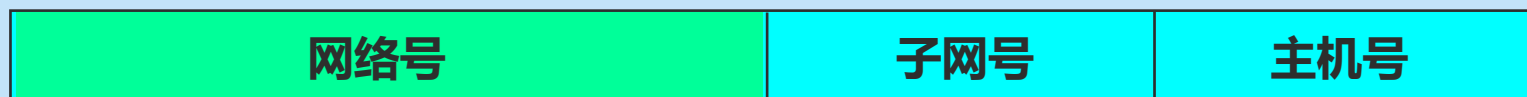


(IP 地址) AND (子网掩码) = 网络地址

两级 IP 地址

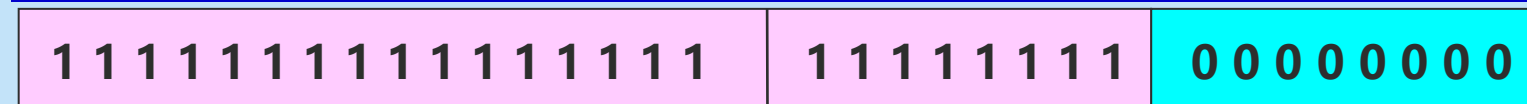


三级 IP 地址

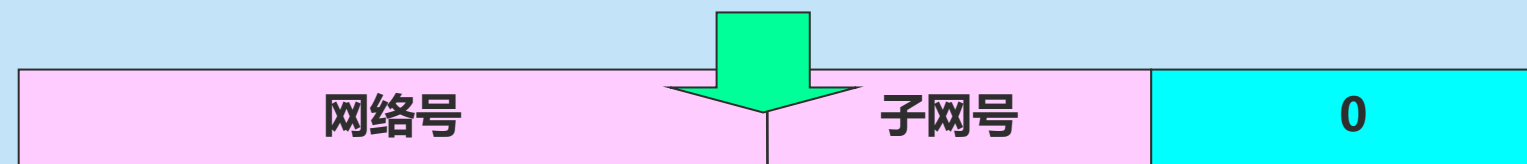


逐位进行 AND 运算

三级 IP 地址
的子网掩码



子网的
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
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 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	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。
 - 虽然根据已成为互联网标准协议的 RFC 950 文档，子网号不能为**全 1**或**全 0**，但随着无分类域间路由选择 CIDR 的广泛使用，现在全 1 和全 0 的子网号也可以使用了，但一定要谨慎使用，确认你的路由器所用的路由选择软件是否支持全 0 或全 1 的子网号这种较新的用法。
 - **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。

(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 255.255.224.0

11111111	11111111	11100000	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
-----	---	----	---	----	---	---

**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。**

超网 CIDR

- 无分类编址 CIDR – Classless Inter-Domain Routing



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：

- 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！
 - 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）。
 - 整个 IPv4 的地址空间最终将全部耗尽。
-

IP 编址问题的演进

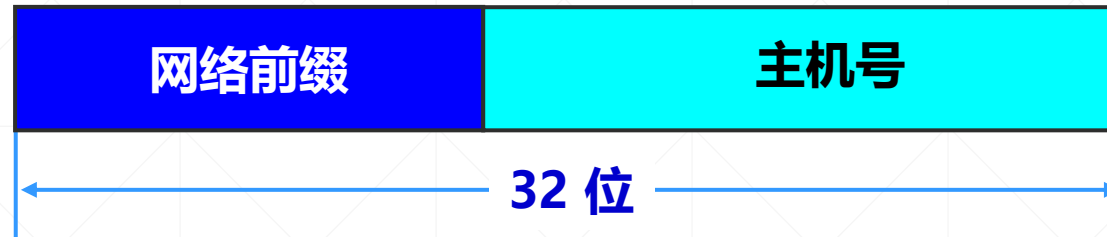
- 1987 年, RFC 1009 就指明了在一个划分子网的网络中可同时使用几个不同的子网掩码。
 - 使用**变长子网掩码 VLSM** (Variable Length Subnet Mask)可进一步提高 IP 地址资源的利用率。
 - 在 VLSM 的基础上又进一步研究出无分类编址方法, 它的正式名字是**无分类域间路由选择 CIDR** (Classless Inter-Domain Routing)。
-

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。**
-

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>} (4-3)

- 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 地址块

- CIDR 把网络前缀都相同的连续的 IP 地址组成 “**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）。
 1. 这个地址块的起始地址是 128.14.32.0。
 2. 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**”。
 3. 128.14.32.0/20 地址块的最小地址：128.14.32.0
 4. 128.14.32.0/20 地址块的最大地址：128.14.47.255
 5. **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	00100000	00000000
10000000	00001110	00100000	00000001
10000000	00001110	00100000	00000010
10000000	00001110	00100000	00000011
10000000	00001110	00100000	00000100
10000000	00001110	00100000	00000101
...		...	
10000000	00001110	00101111	11111011
10000000	00001110	00101111	11111100
10000000	00001110	00101111	11111101
10000000	00001110	00101111	11111110
10000000	00001110	00101111	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。

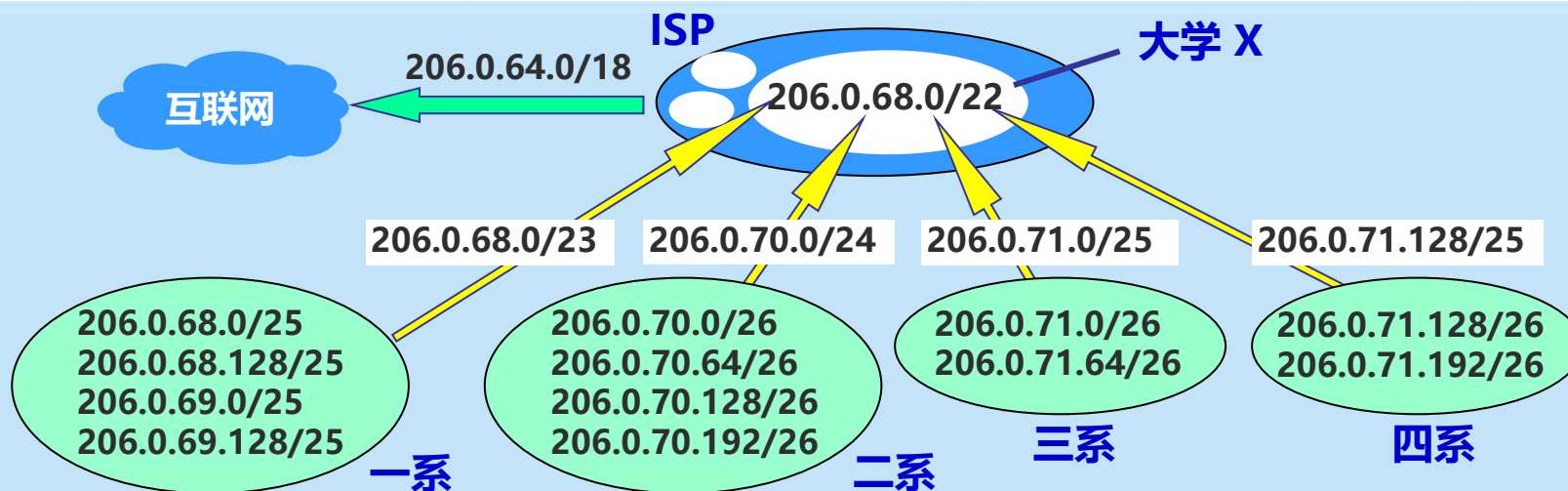
常用的 CIDR 地址块

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

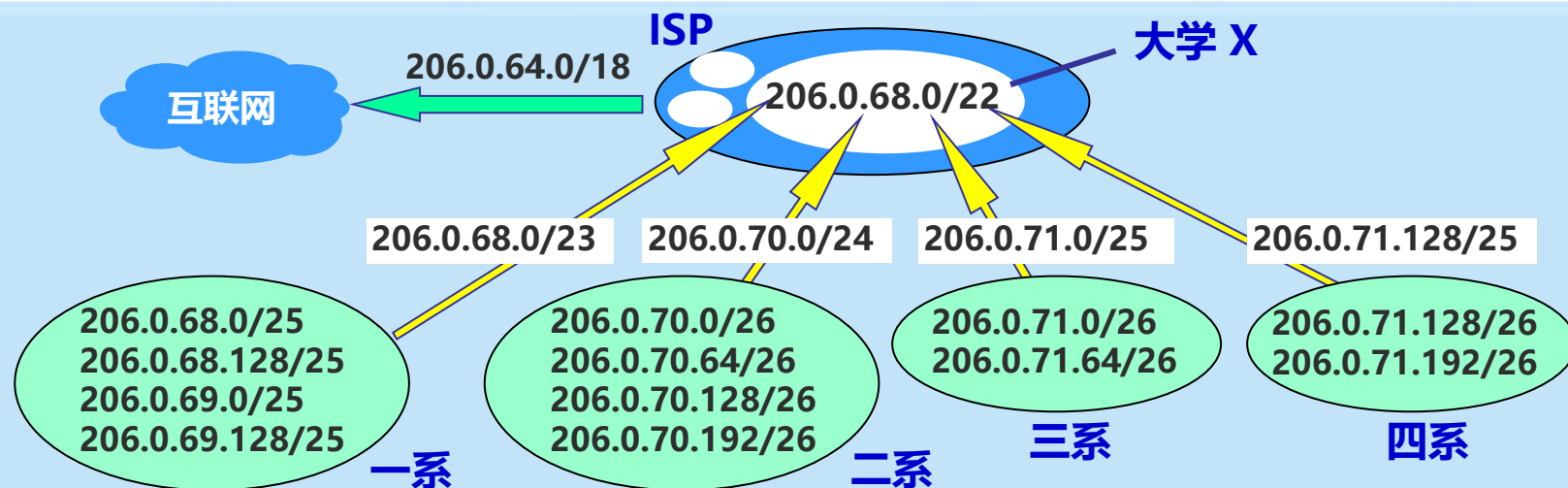
- 前缀长度不超过 23 位的 CIDR 地址块都包含了多个 C 类地址。
 - 这些 C 类地址合起来就构成了超网。
 - CIDR 地址块中的地址数一定是 2 的整数次幂。
 - 网络前缀越短，其地址块所包含的地址数就越多。而在三级结构的 IP 地址中，划分子网是使网络前缀变长。
 - CIDR 的一个好处是：可以更加有效地分配 IPv4 的地址空间，可根据客户的需要分配适当大小的 CIDR 地址块。
-

CIDR 地址块划分举例



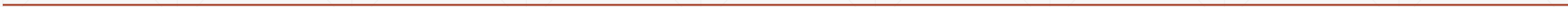
单位	地址块	二进制表示	地址数
ISP	206.0.64.0/18	11001110.00000000.01*	16384
大学	206.0.68.0/22	11001110.00000000.010001*	1024
一系	206.0.68.0/23	11001110.00000000.0100010*	512
二系	206.0.70.0/24	11001110.00000000.01000110.*	256
三系	206.0.71.0/25	11001110.00000000.01000111.0*	128
四系	206.0.71.128/25	11001110.00000000.01000111.1*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

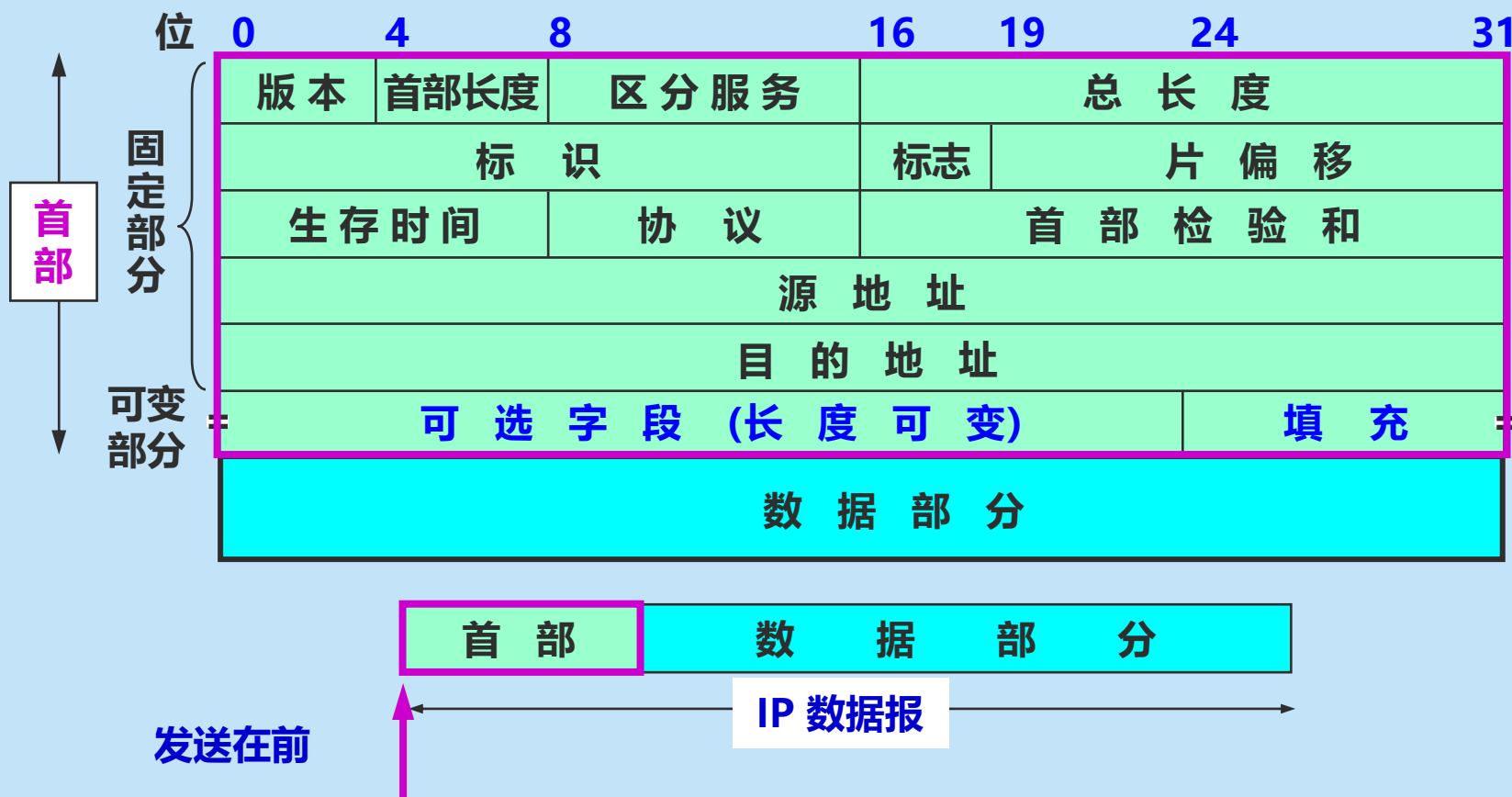
IP数据报格式 + 分片与重组



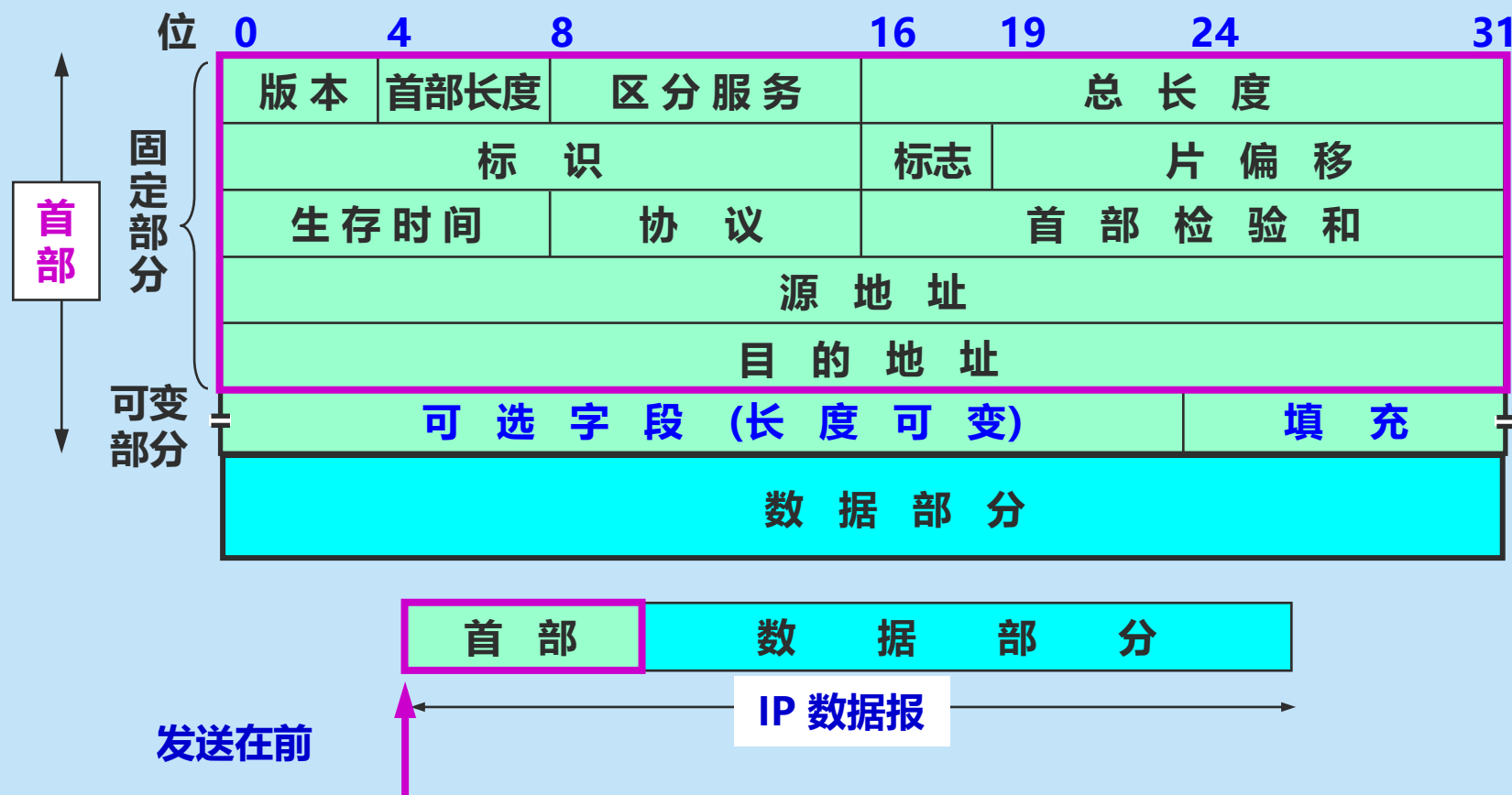
IP 数据报的格式

- 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

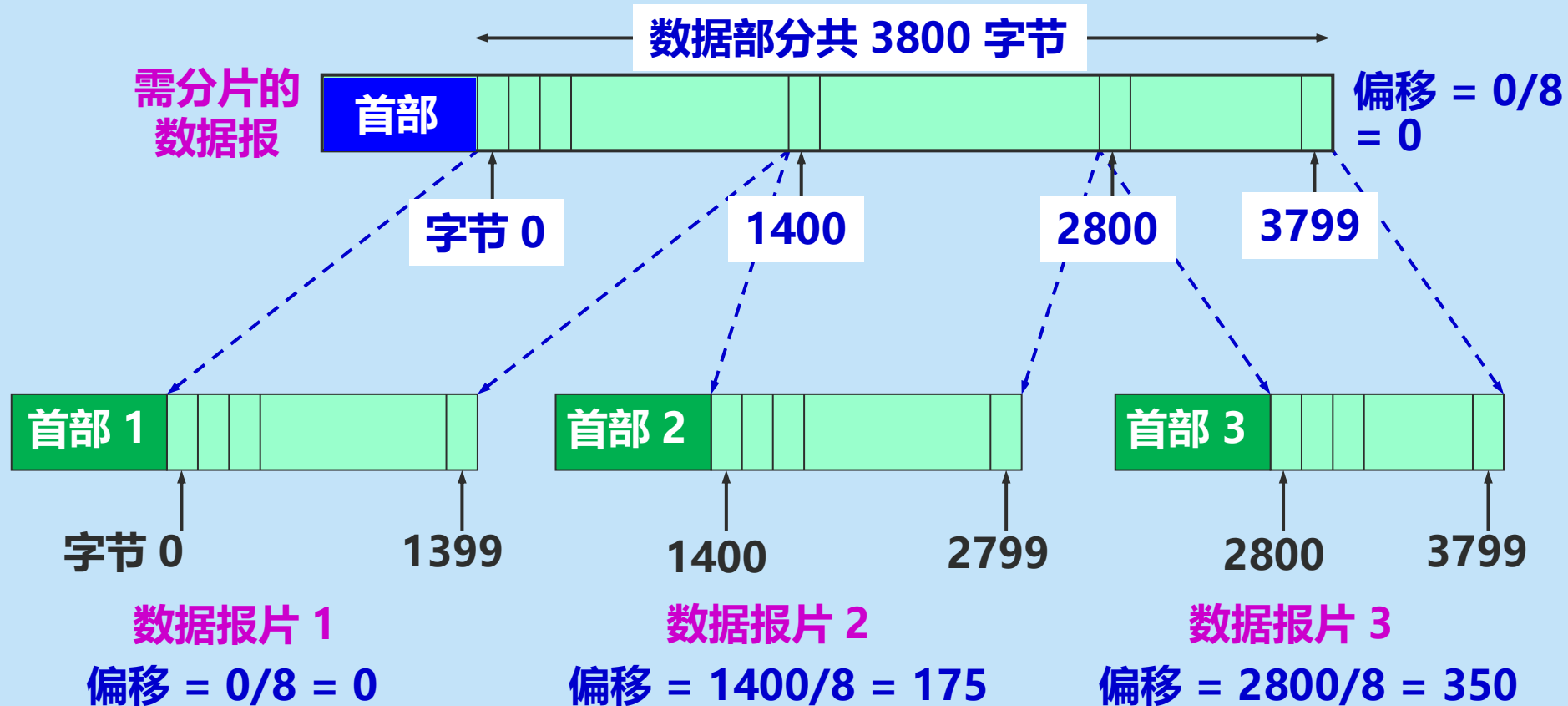


片偏移——占13位，指出：较长的分组在分片后某片在原分组中的相对位置。
片偏移以8个字节为偏移单位。

【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



【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

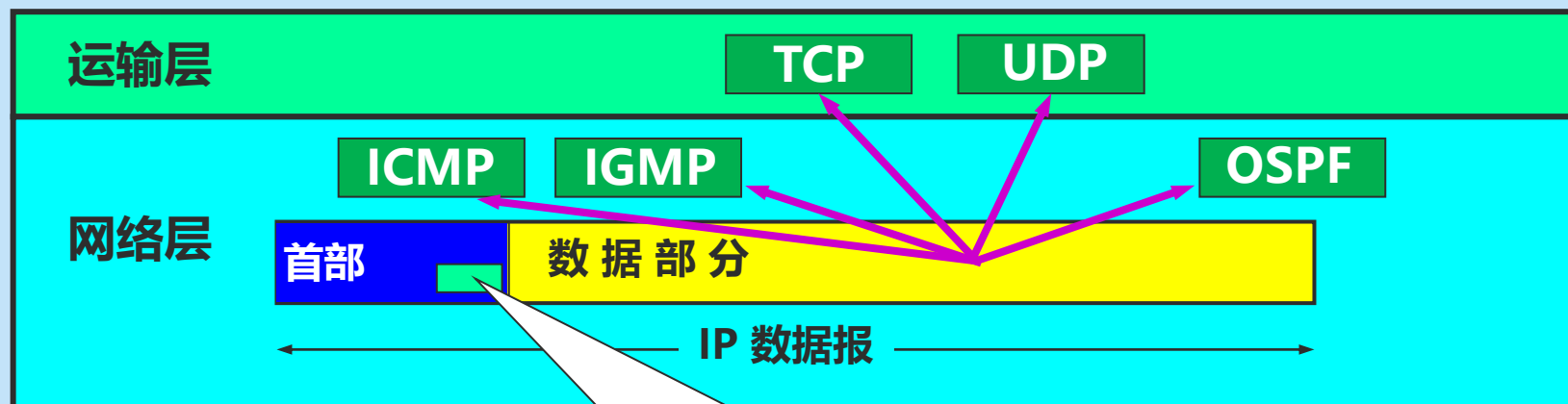


生存时间——占8位，记为 TTL (Time To Live)，指示数据报在网络中可通过的路由器数的最大值。

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



协议字段指出应将数据部分交给哪一个进程

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。

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

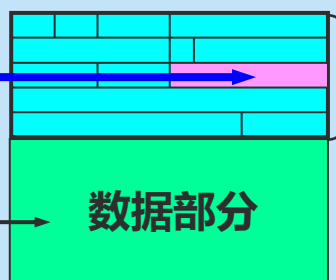
取反码

检验和 16 位

数据部分

不参与检验和的计算

IP 数据报



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

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

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。
-

NAT – Network Address Translation

- 网络地址转换 – 为什么需要网络地址转换?
 - 有很多原因:
 - 本地互联网地址 (或称之为专用互联网)
 - ISP并不需要知道一个范围内的所有地址, 很多情况下, 知道一个地址指的是多个设备即可
 - 透明性:
 - 可以更改本地设备的IP地址而不需要通知外部世界
 - 可以更改ISP而不需要更改本地设备的IP地址
 - 安全性: 本地网络内的设备并不能直接被外界网络访问或可见
-

- 由于 **IP 地址的紧缺**，一个机构能够申请到的IP地址数往往远小于本机构所拥有的主机数。
 - 考虑到**互联网并不很安全**，一个机构内也并不需要把所有的主机接入到外部的互联网。
 - 假定在一个机构内部的计算机通信也是采用 TCP/IP 协议，那么从原则上讲，对于这些仅在**机构内部使用**的计算机就可以由本机构**自行分配其 IP 地址**。
-

本地地址与全球地址

- **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，而不需要向互联网的管理机构申请。
 - **全球地址**——全球唯一的 IP 地址，必须向互联网的管理机构申请。
 - **问题**：在内部使用的本地地址就有可能和互联网中某个 IP 地址重合，这样就会出现地址的二义性问题。
-

本地地址与全球地址

- **问题**：在内部使用的本地地址就有可能和互联网中某个 IP 地址重合，这样就会出现地址的**二义性**问题。
 - **解决**：RFC 1918 指明了一些**专用地址** (private address)。专用地址只能用作本地地址而不能用作全球地址。在互联网中的所有路由器，对目的地址是专用地址的数据报一律不进行转发。
-

RFC 1918 指明的专用 IP 地址

三个专用 IP 地址块:

(1) 10.0.0.0 到 10.255.255.255

A类, 或记为10.0.0.0/8, 它又称为 24 位块

(2) 172.16.0.0 到 172.31.255.255

B类, 或记为172.16.0.0/12, 它又称为 20 位块

(3) 192.168.0.0 到 192.168.255.255

C类, 或记为192.168.0.0/16, 它又称为 16 位块

专用网

- 采用这样的专用 IP 地址的互连网络称为**专用互联网**或**本地互联网**，或更简单些，就叫做**专用网**。
 - 因为这些专用地址仅在本机构内部使用。专用IP地址也叫做**可重用地址** (reusable address)。
-

NAT – Network Address Translation

- 基于以上的原因，1994年提出了NAT的概念
 - 本地网络内的主机如何与互联网上的其他主机进行通信（在不需要加密的情况下）
 - 在申请一些全球IP地址 – 当然这个是不现实的
 - 采用网络地址转换NAT – 这个实际上是目前最常见的方法
 - NAT实际上是本地网络和互联网中间的媒介
 - 需要在专用网连接到互联网的路由器上安装 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IP地址。
 - 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全球 IP 地址，才能和互联网连接。
-

NAT – 两类实现方式

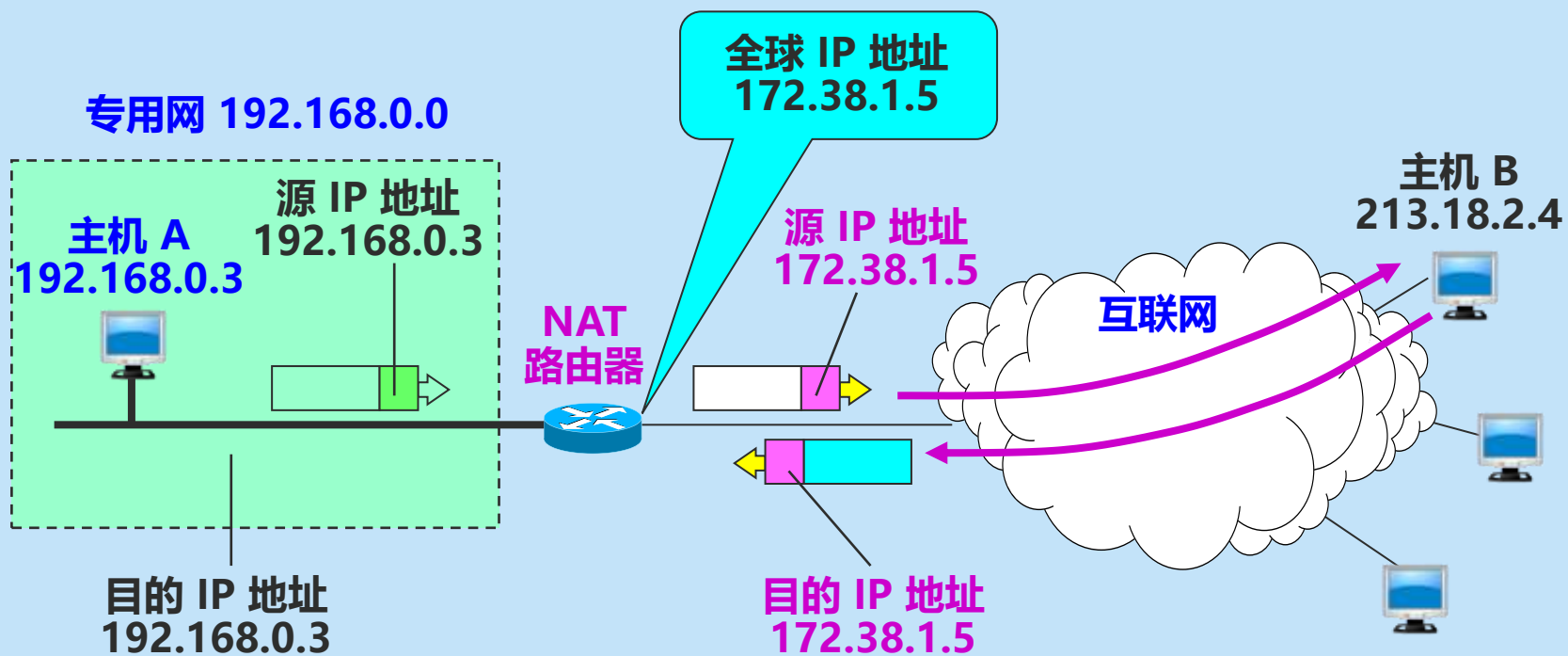
- NAT
 - 纯粹的地址转换
 - 现实中不会使用
 - NAT (Network Address and Port Translation)
 - 利用地址加端口进行转换
 - 现实中一般都使用的是这种，我们往往称之为NAT，而不会加上“端口”这个限定。
 - 课件中为了定义方便，我们还是分开使用NAT和NAPT这两种称呼
-

NAT

- 我们先来看NAT，也就是纯粹的地址转换



网络地址转换的过程



NAT 路由器的工作原理

网络地址转换的过程

- 内部主机 A 用本地地址 IP_A 和互联网上主机 B 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_A 转换成全球地址 IP_G ，并把转换结果记录到 NAT 地址转换表中，目的地址 IP_B 保持不变，然后发送到互联网。
- NAT 路由器收到主机 B 发回的数据报时，知道数据报中的源地址是 IP_B 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_A ，转发给最终的内部主机 A。

网络地址转换的过程

- 可以看出，在内部主机与外部主机通信时，在NAT路由器上发生了**两次地址转换**：
 1. **离开专用网时**：替换源地址，将内部地址替换为全球地址；
 2. **进入专用网时**：替换目的地址，将全球地址替换为内部地址；

NAT地址转换表举例

方向	字段	旧的IP地址	新的IP地址
出	源IP地址	192.168.0.3	172.38.1.5
入	目的IP地址	172.38.1.5	192.168.0.3
出	源IP地址	192.168.0.7	172.38.1.6
入	目的IP地址	172.38.1.6	192.168.0.7

网络地址转换 NAT

- 当 NAT 路由器具具有 n 个全球 IP 地址时，专用网内**最多可以同时有 n 台主机接入到互联网**。这样就可以使专用网内较多数量的主机，轮流使用 NAT 路由器有限数量的全球 IP 地址。
- 通过 NAT 路由器的通信必须由专用网内的主机发起。**专用网内部的主机不能充当服务器用**，因为互联网上的客户无法请求专用网内的服务器提供服务。

NAPT

- 我们注意到，这种NAT的方式中：
 - 需要NAT路由器具具有n个全球IP地址，在本地网内才可以有n台主机连接到互联网
 - 如果本地网内多于n台主机，就需要轮流使用全球IP地址
 - 这种方法，在IP地址稀缺的今天，显然是不够的
 - 因此就提出了NAPT
 - 传输层 – 端口号的概念
 - 下一层我们要讲的才是传输层，现在先把端口号借用过来
-

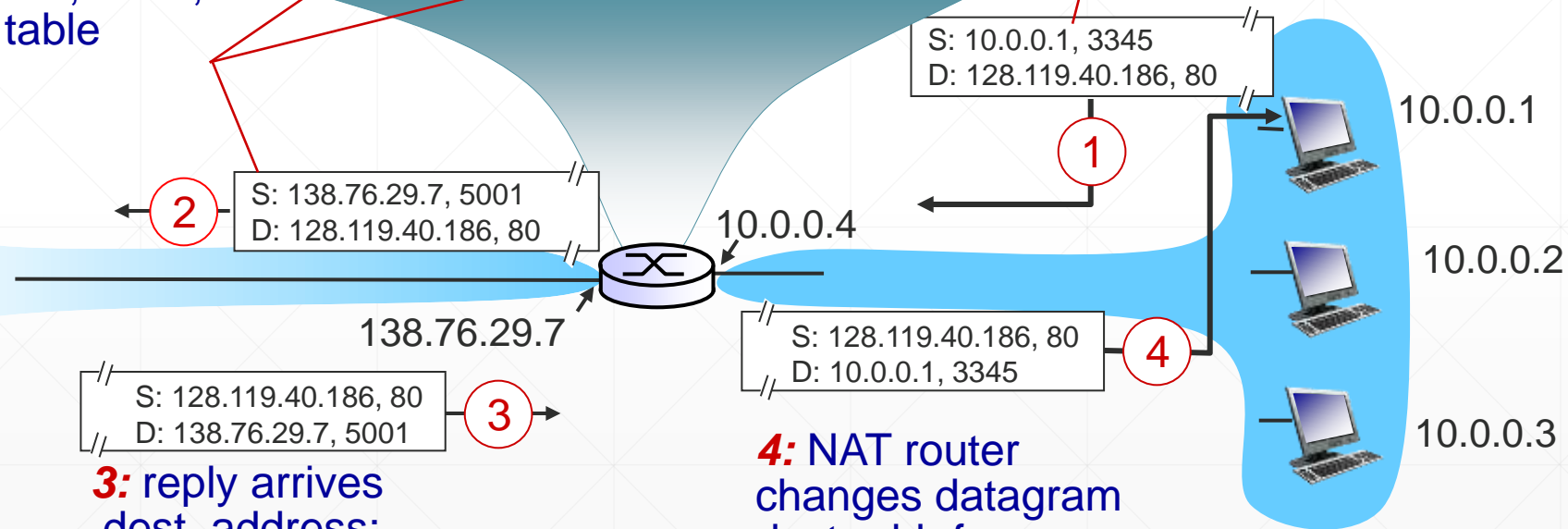
NAPT

- NAT路由器只需要具有一个全球IP地址就足够了。不通过IP地址来区分不同的本地、源IP地址，而通过NAT路由器上的端口号来区别。
 - 所有经过NAT路由器发出的数据报
 - (source IP addr, port #) → (NAT IP addr, new port #)
 - 所有从外界发回来的数据报也会使用(NAT IP addr, new port #)作为目的地址
 - 在NAT地址转换表中记录每一个映射对
 - 所有经过NAT路由器收到的数据报
 - (NAT IP addr, new port #) → (source IP addr, port #)
 - 再发送给正确的主机
-

2: NAT router changes datagram source addr from 10.0.0.1, 3345 to 138.76.29.7, 5001, updates table

NAT translation table	
WAN side addr	LAN side addr
138.76.29.7, 5001	10.0.0.1, 3345
.....

1: host 10.0.0.1 sends datagram to 128.119.40.186, 80



3: reply arrives
dest. address:
138.76.29.7, 5001

4: NAT router
changes datagram
dest addr from
138.76.29.7, 5001 to 10.0.0.1, 3345

另外一个例子

- 上图的例子是只有一个映射对的，下面的例子是多个映射对的

方向	IP数据报字段	转换前的IP地址：端口号	转换后的IP地址：端口号
发出	源IP：源端口	192.168.0.3:30000	172.38.1.5:40001
发出	源IP：源端口	192.168.0.4:30000	172.38.1.5:40002
进入	目的IP：目的端口	172.38.1.5:40001	192.168.0.3:30000
进入	目的IP：目的端口	172.38.1.5:40002	192.168.0.4:30000

- NAT路由器的全球地址是多少？
- 如何将本地网地址的 “IP地址：端口号” 映射成 “全球IP地址：端口号” ？

NAT (&NAPT) 总结

- NAPT采用端口号区分本地网内的不同主机
 - 端口号在传输层分组中，占用16bit。因此，即使只使用一个全球IP地址，也能同时连接超过六万台本地主机
 - NAT是“有争议性”的、“临时”的解决手段
 - 路由器应至多只处理第三层；端口号是第四层的
 - 地址短缺应由IPv6彻底解决，而不应该让NAT作临时解决
 - 违背了端对端原则（end-to-end argument）
 - 网络不应该修改分组的内容！应该由端系统来做这件事情
 - 例如，p2p应用，由应用程序本身来做这件事情，而不是由网络来做
 - NAT后的主机不能同时作为服务器向外提供服务
-

IPv6



IPv6

- IP 是互联网的核心协议。
 - 互联网经过几十年的飞速发展，到 2011 年 2 月，IPv4 的 32 位地址已经耗尽。
 - ISP 已经不能再申请到新的 IP 地址块了。
 - 我国在 2014 – 2015 年也逐步停止了向新用户和应用分配 IPv4 地址。
 - 解决 IP 地址耗尽的根本措施就是采用具有更大地址空间的新版本的 IP，即 IPv6。
-

IPv6 的基本首部

- IPv6 仍支持**无连接的传送**，但将协议数据单元 PDU 称为**分组**。为方便起见，本书仍采用数据报这一名词。
 - 所引进的**主要变化**如下：
 1. **更大的地址空间**。IPv6 将地址从 IPv4 的 32 位 增大到了 128 位。
 2. **扩展的地址层次结构**。
 3. **灵活的首部格式**。IPv6 定义了许多可选的扩展首部。
 4. **改进的选项**。IPv6 允许数据报包含有选项的控制信息，其选项放在有效载荷中。
-

IPv6 的基本首部

- 所引进的主要变化如下（续）：

5. 允许协议继续扩充。

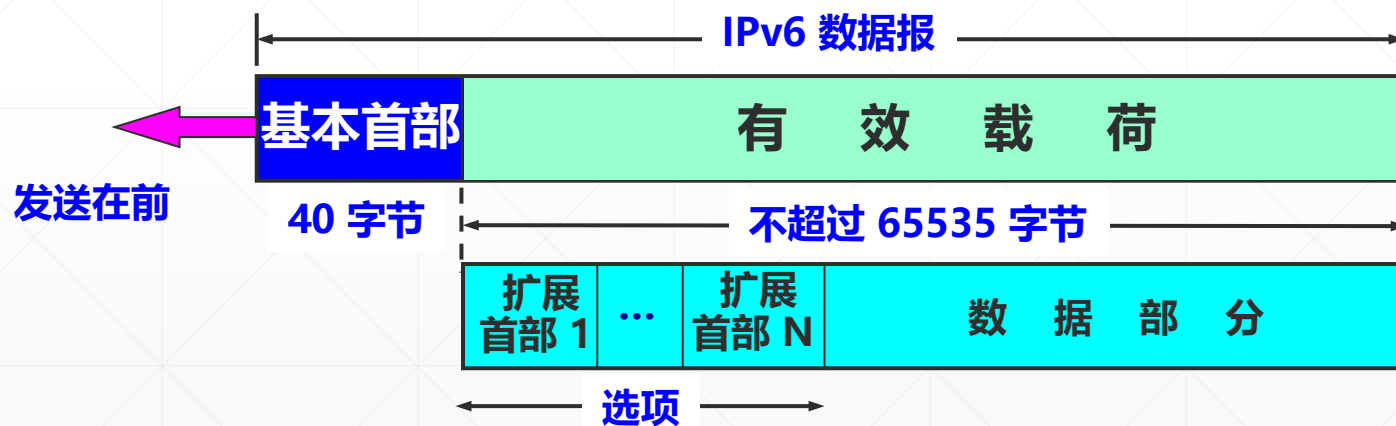
6. 支持即插即用（即自动配置）。因此 IPv6 不需要使用 DHCP。

7. 支持资源的预分配。IPv6 支持实时视像等要求，保证一定的带宽和时延的应用。

8. IPv6 首部改为 8 字节对齐。首部长度必须是 8 字节的整数倍。原来的 IPv4 首部是 4 字节对齐。

IPv6 数据报的一般形式

- IPv6 数据报由两大部分组成：
 1. 基本首部 (base header)
 2. 有效载荷 (payload)。有效载荷也称为净负荷。有效载荷允许有零个或多个扩展首部 (extension header)，再后面是数据部分。



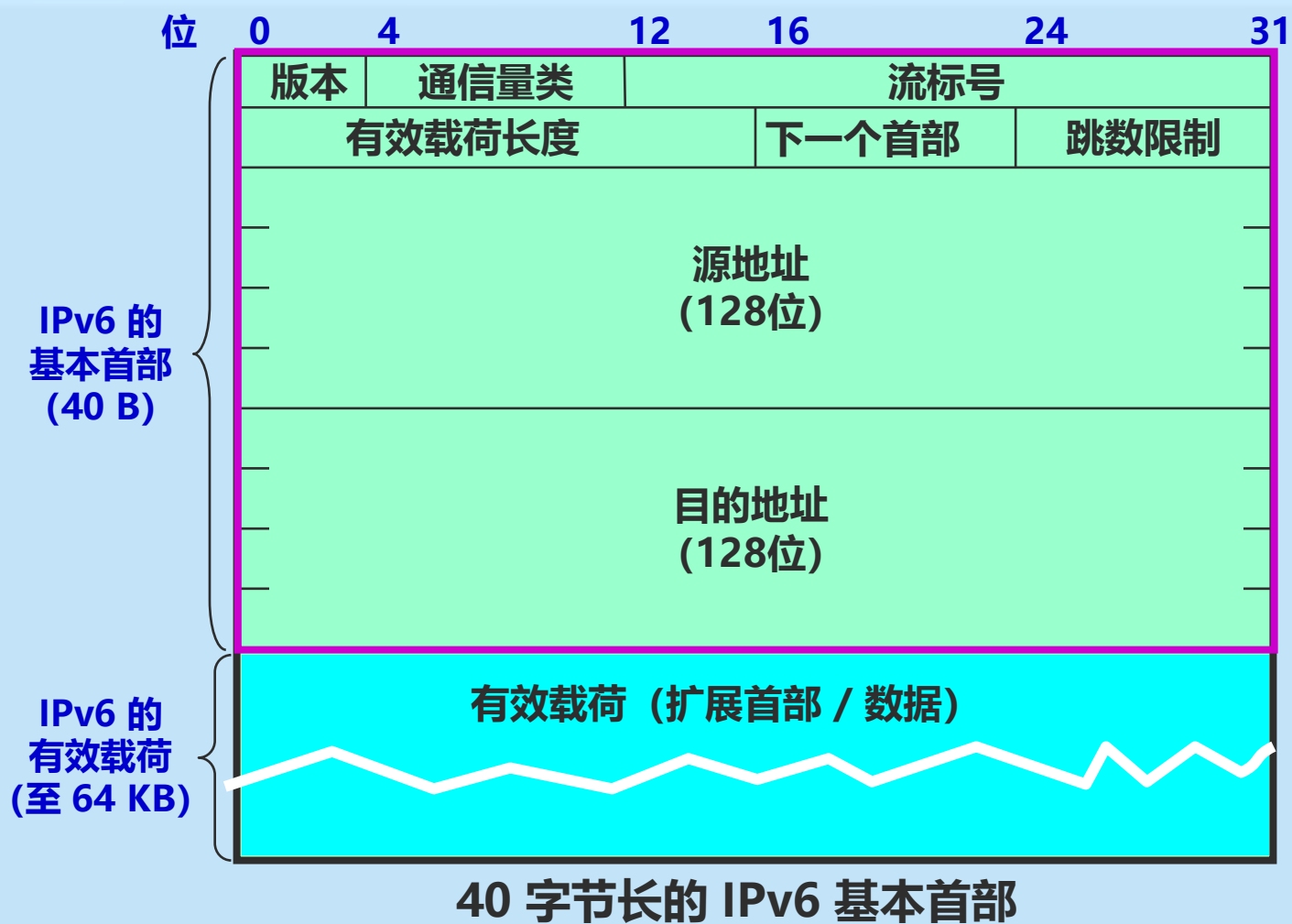
具有多个可选扩展首部的 IPv6 数据报的一般形式

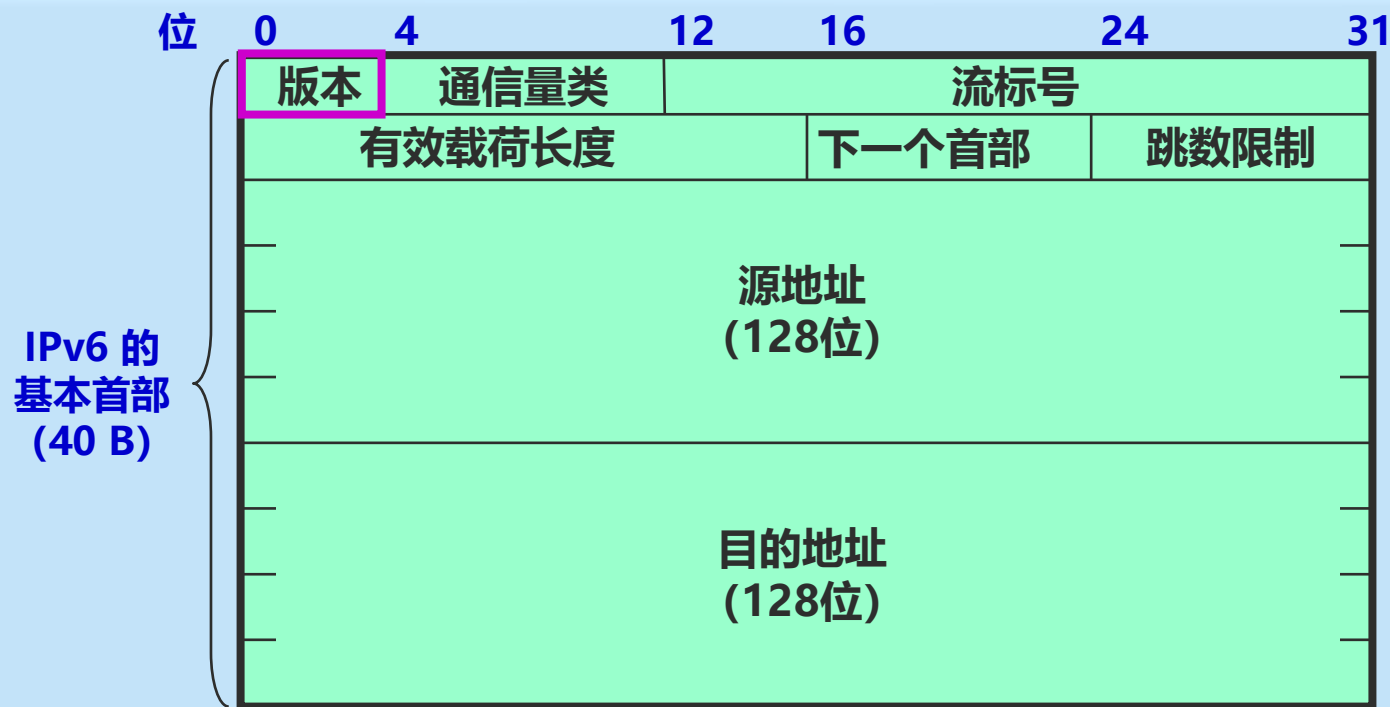
IPv6 数据报的基本首部

- IPv6 将首部长度变为**固定的 40 字节**，称为**基本首部**。
- 把首部中不必要的功能取消了，使得 IPv6 首部的字段数减少到只有 8 个。
- IPv6 对首部中的某些字段进行了如下的**更改**：

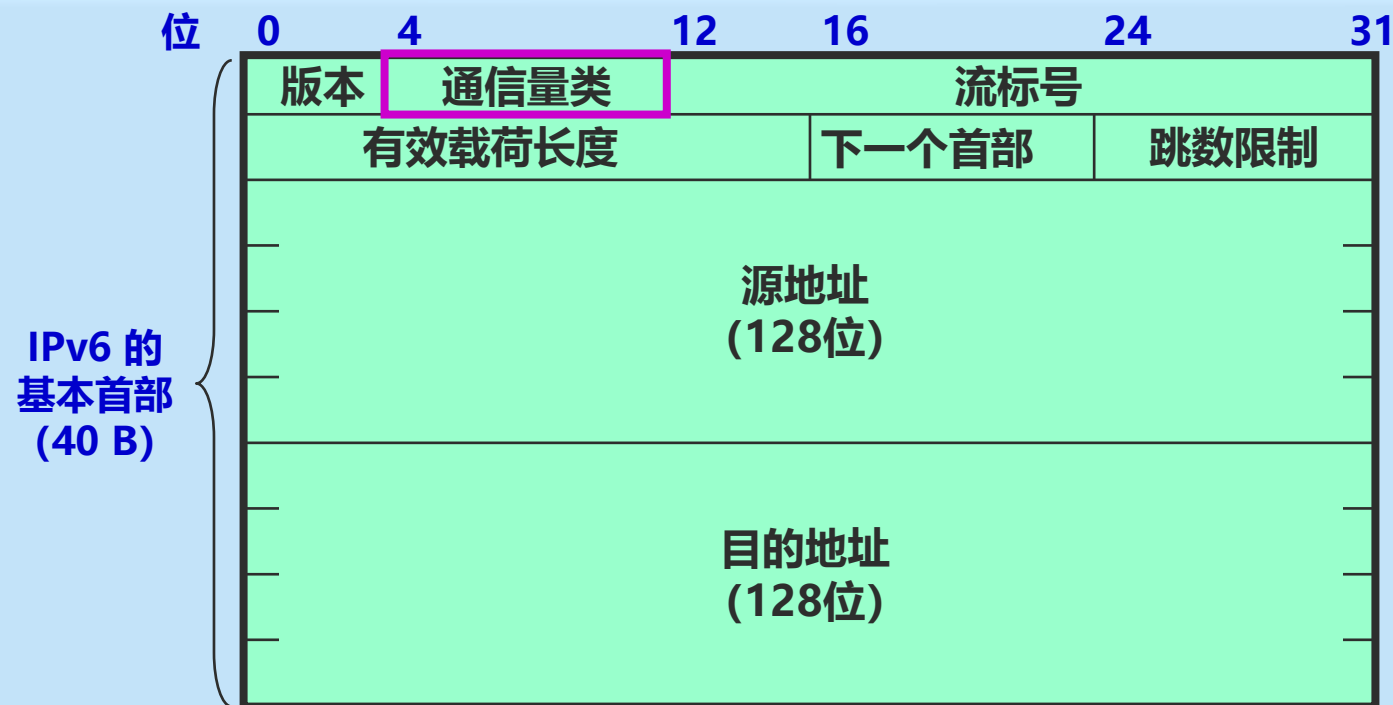
- 取消了首部长度字段，因为首部长度是固定的 40 字节；
- 取消了服务类型字段；
- 取消了总长度字段，改用有效载荷长度字段；

- 把 TTL 字段改称为跳数限制字段；
- 取消了协议字段，改用下一个首部字段；
- 取消了检验和字段；
- 取消了选项字段，而用扩展首部来实现选项功能。

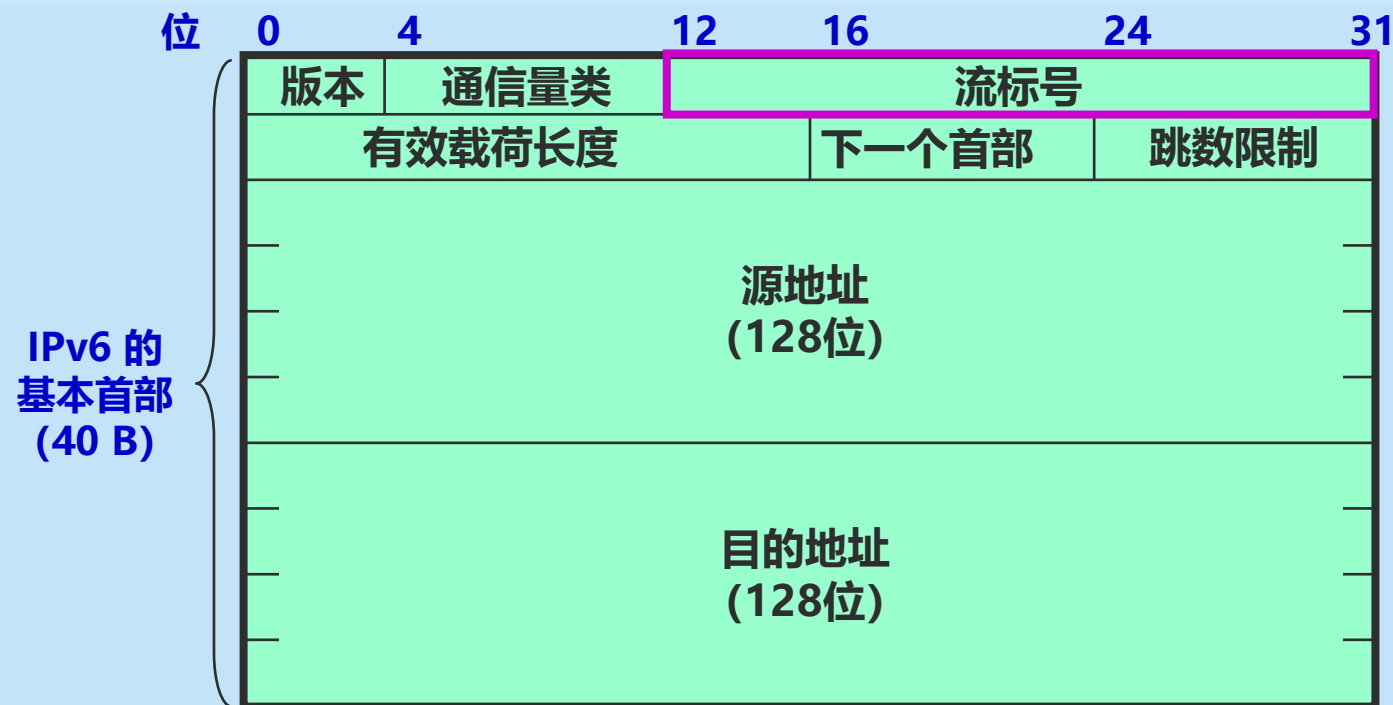




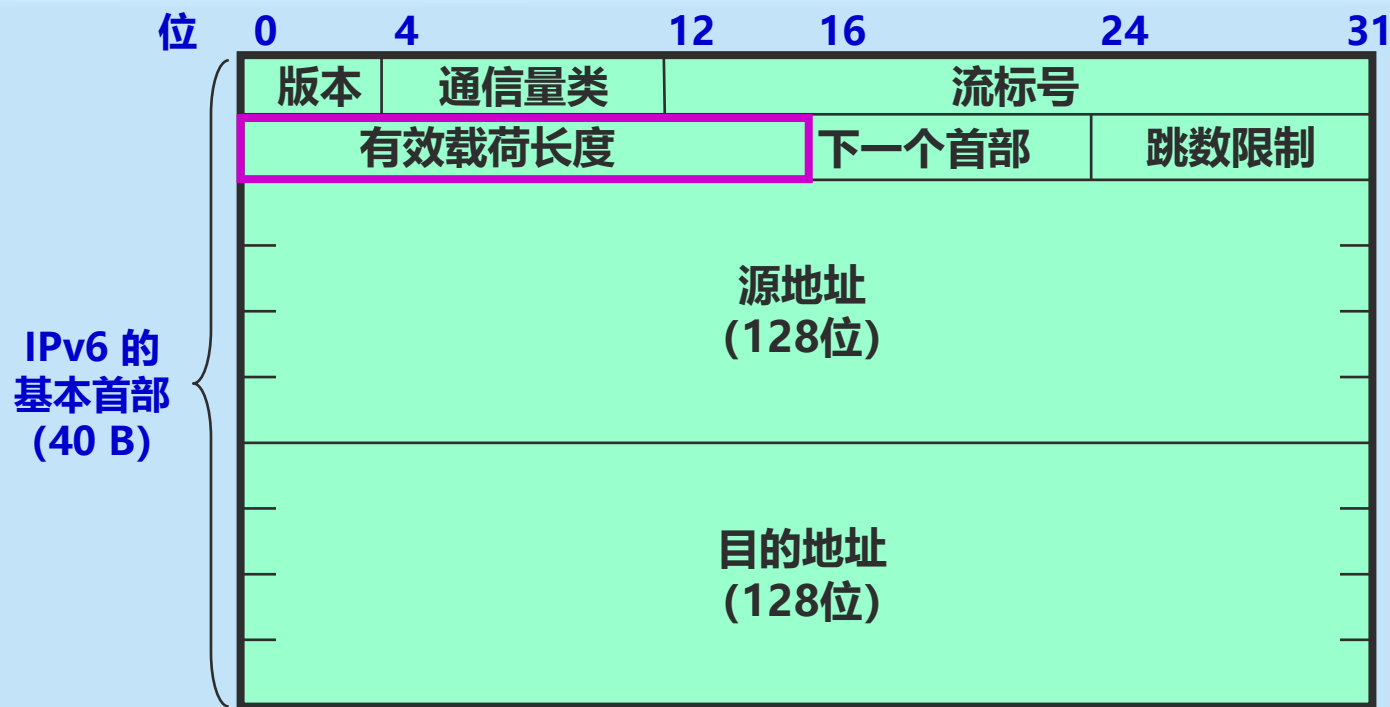
版本(version)—— 4 位。它指明了协议的版本，对 IPv6 该字段总是 6。



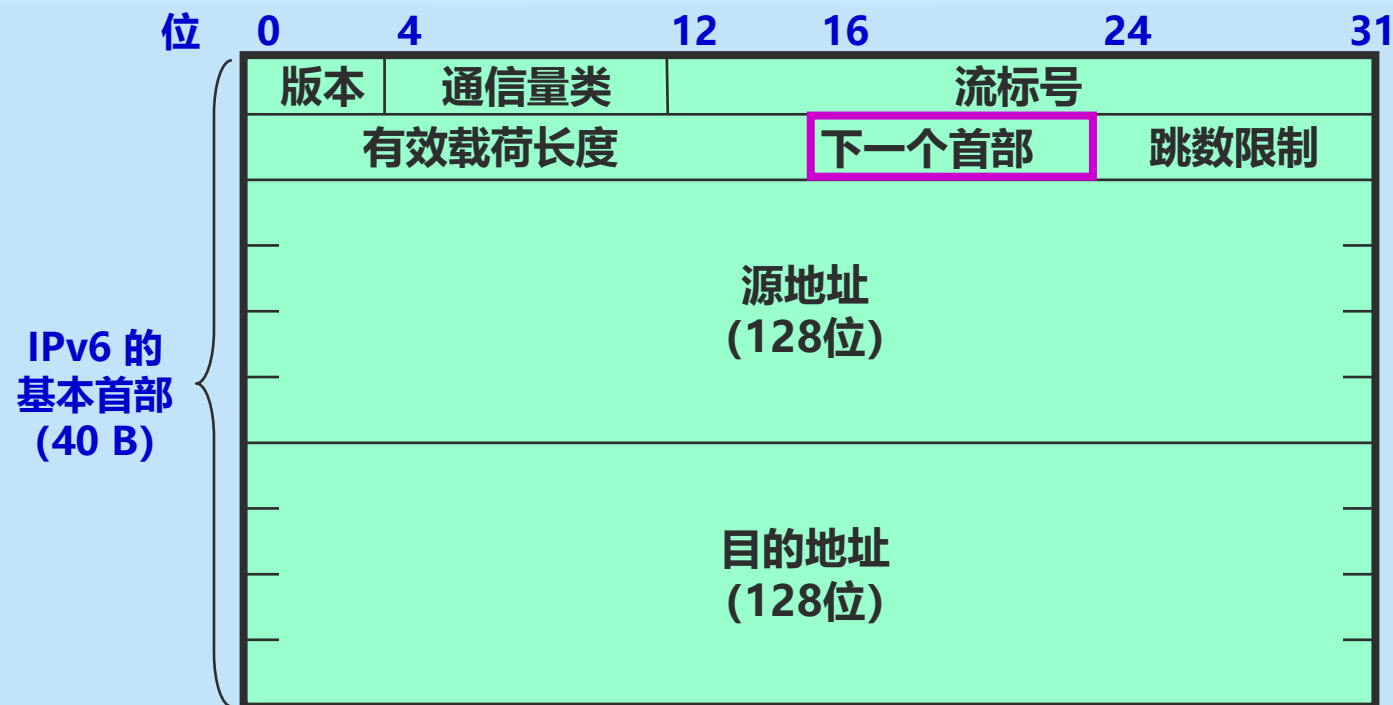
通信量类(traffic class)—— 8 位。这是为了区分不同的 IPv6 数据报的类别或优先级。目前正在进行不同的通信量类性能的实验。



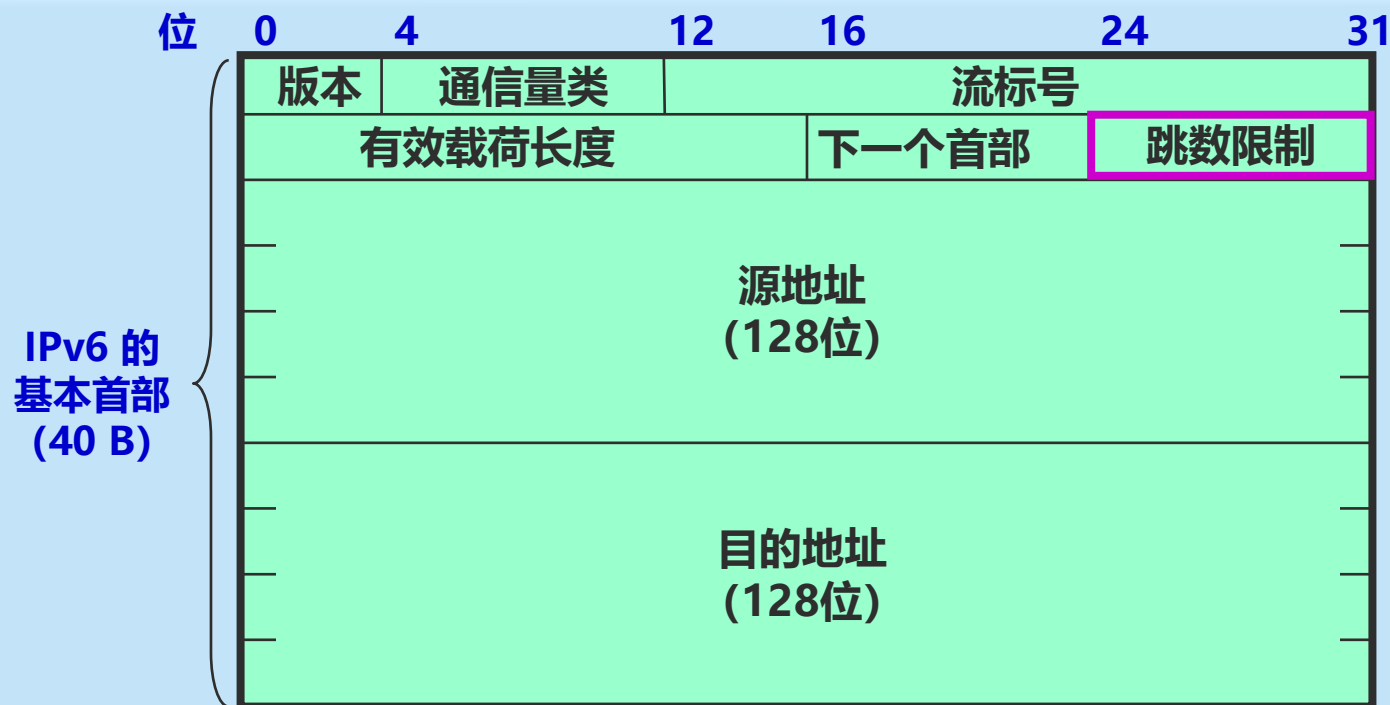
流标号(flow label)—— 20 位。 “流” 是互联网络上从特定源点到特定终点的一系列数据报，“流” 所经过的路径上的路由器都保证指明的服务质量。所有属于同一个流的数据报都具有同样的流标号。



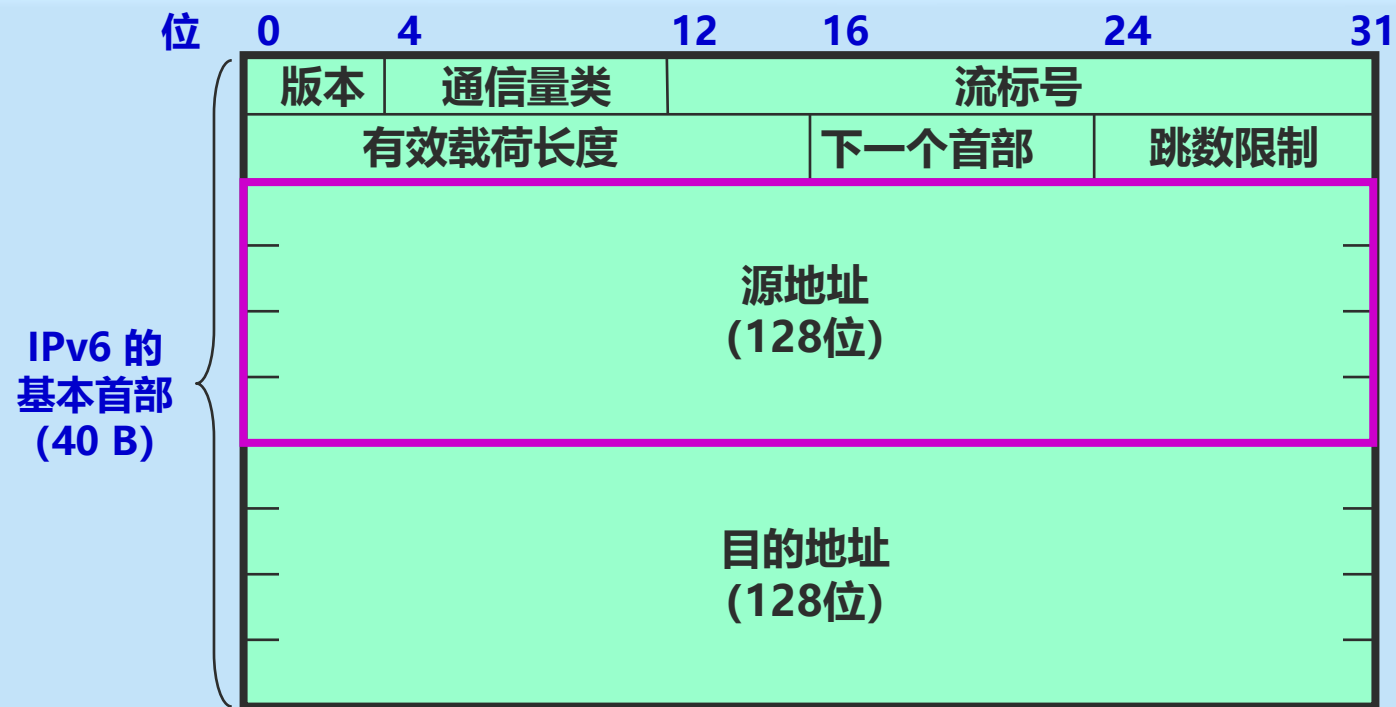
有效载荷长度(payload length)—— 16 位。它指明 IPv6 数据报除基本首部以外的字节数（所有扩展首部都算在有效载荷之内），其最大值是 64 KB。



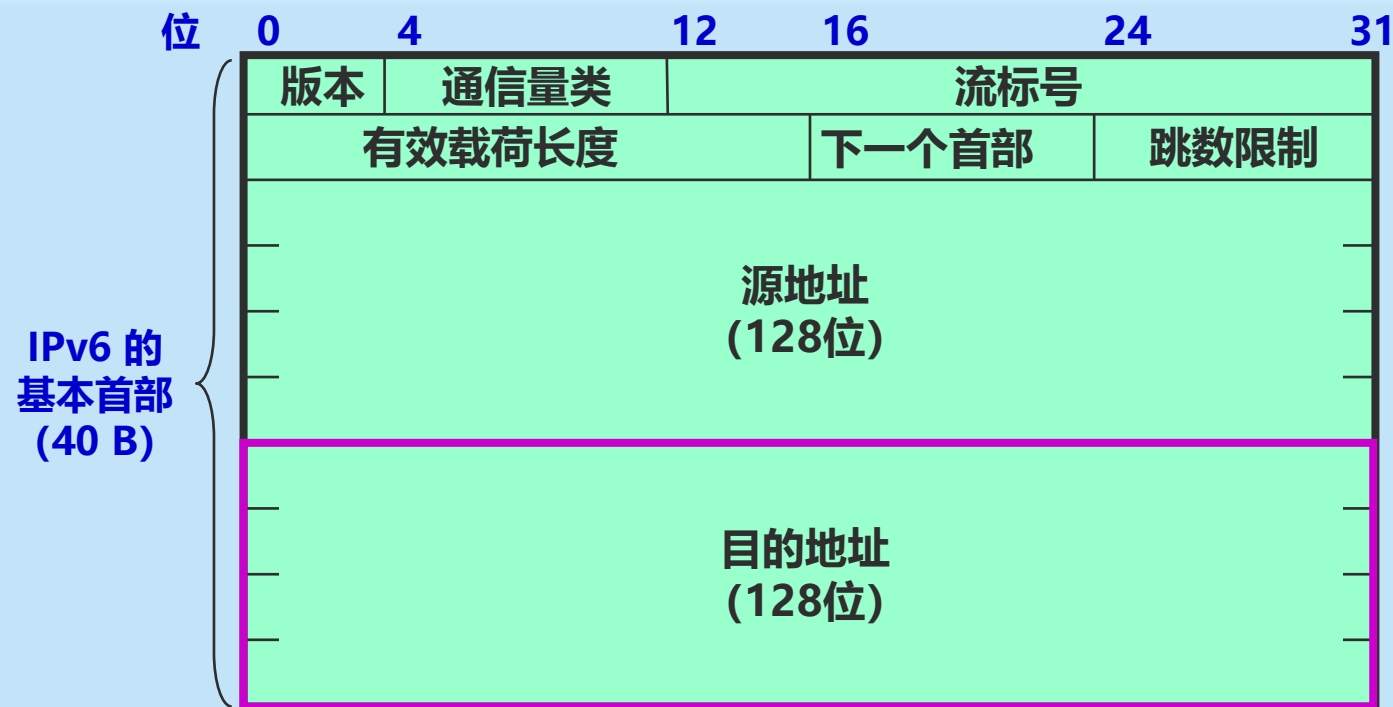
下一个首部(next header)—— 8 位。它相当于 IPv4 的协议字段或可选字段。



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 的扩展首部

- IPv6 把原来 IPv4 首部中选项的功能都放在**扩展首部**中，并将扩展首部留给路径两端的源站和目的站的主机来处理。
 - **数据报途中经过的路由器都不处理这些扩展首部（只有一个首部例外，即逐跳选项扩展首部）。**
 - **这样就大大提高了路由器的处理效率。**
-

六种扩展首部

在 RFC 2460 中定义了六种扩展首部：

1. 逐跳选项
 2. 路由选择
 3. 分片
 4. 鉴别
 5. 封装安全有效载荷
 6. 目的站选项
-

IPv6 的地址

- 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): 传统的点对点通信。
 2. **多播** (multicast): 一点对多点的通信。
 3. **任播** (anycast): 这是 IPv6 增加的一种类型。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**。
- 一个结点就可能有多多个与链路相连的接口。
- IPv6 地址是分配给结点上面的接口的。
 1. 一个接口可以有多个单播地址。
 2. 其中的任何一个地址都可以当作到达该结点的目的地址。即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

- 在 IPv6 中，每个地址占 128 位，地址空间大于 3.4×10^{38} 。
- 为了使地址再稍简洁些，IPv6 使用**冒号十六进制记法**(colon hexadecimal notation, 简称为 colon hex)。
- 每个 16 位的值用十六进制值表示，各值之间用冒号分隔。例如：
68E6:8C64:FFFF:FFFF:0:1180:960A:FFFF
- 在十六进制记法中，允许把数字前面的 0 省略。例如把 0000 中的前三个 0 省略，写成 1 个 0。

零压缩

- 冒号十六进制记法可以允许**零压缩** (zero compression), 即一连串连续的零可以为一对冒号所取代。

FF05:0:0:0:0:0:0:B3 可压缩为:

FF05::B3

- **注意:** 在任一地址中只能使用一次零压缩。
-

点分十进制记法的后缀

- 冒号十六进制记法可结合使用点分十进制记法的后缀，这种结合在 IPv4 向 IPv6 的转换阶段特别有用。
- 例如：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。
- 例如：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 (零压缩)
或 12AB:0:0:CD30::/60 (零压缩)

IPv6 地址分类

地址类型	二进制前缀
未指明地址	00...0 (128位) , 可记为 ::/128。
环回地址	00...1 (128位) , 可记为 ::1/128。
多播地址	11111111 (8位) , 可记为 FF00::/8。
本地链路单播地址	1111111010 (10位) , 可记为 FE80::/10。
全球单播地址	(除上述四种外, 所有其他的二进制前缀)

IPv6 地址分类

- 未指明地址

1. 这是 16 字节的全 0 地址，可缩写为两个冒号 “::”。
2. 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。
3. 这类地址仅此一个。

- 环回地址

1. 即 0:0:0:0:0:0:0:1（记为 ::1）。
 2. 作用和 IPv4 的环回地址一样。
 3. 这类地址也是仅此一个。
-

IPv6 地址分类

- **多播地址**

1. 功能和 IPv4 的一样。
2. 这类地址占 IPv6 地址总数的 $1/256$ 。

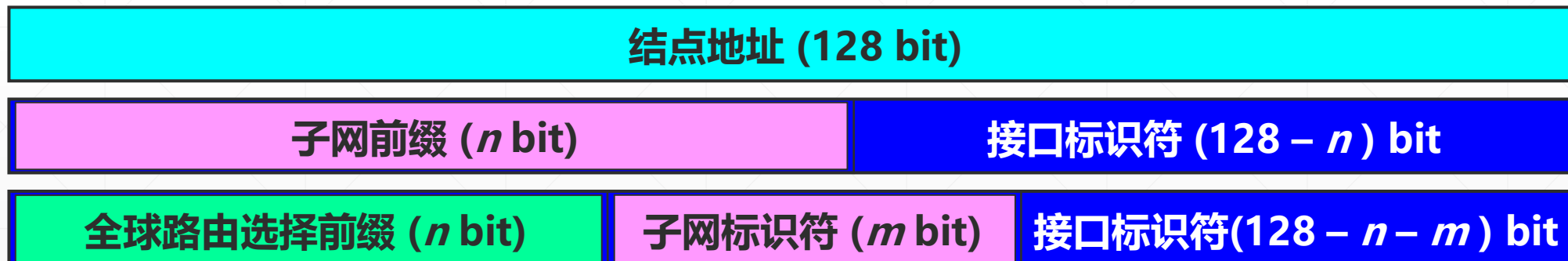
- **本地链路单播地址** (Link-Local Unicast Address)

1. 有些单位的网络使用 TCP/IP 协议，但**并没有连接到互联网上**。
连接在这样的网络上的主机都可以使用这种本地地址进行通信，但不能和互联网上的其他主机通信。
 2. 这类地址占 IPv6 地址总数的 $1/1024$ 。
-

IPv6 地址分类

- 全球单播地址

1. IPv6 的这一类单播地址是使用得最多的一类。
2. 曾提出过多种方案来进一步划分这 128 位的单播地址。
3. 根据 2006 年发布的草案标准 RFC 4291 的建议， IPv6 单播地址的划分方法非常灵活。



IPv6 单播地址的几种划分方法

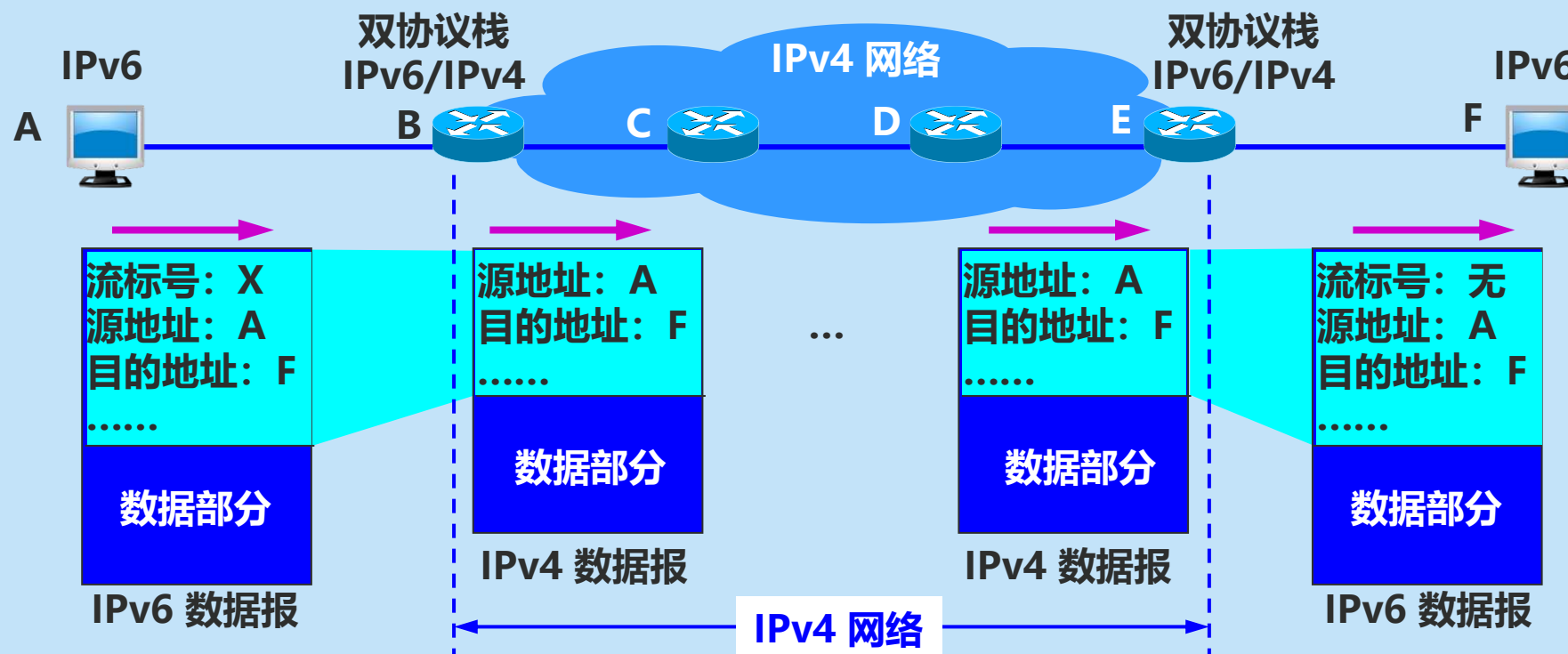
从 IPv4 向 IPv6 过渡

- 向 IPv6 过渡**只能采用逐步演进的办法**，同时，还必须使新安装的 IPv6 系统能够**向后兼容**：**IPv6 系统必须能够接收和转发 IPv4 分组，并且能够为 IPv4 分组选择路由。**
 - 两种向 IPv6 过渡的策略：
 1. **使用双协议栈**
 2. **使用隧道技术**
-

双协议栈

- 双协议栈 (dual stack) 是指在完全过渡到 IPv6 之前, 使一部分主机 (或路由器) 装有两个协议栈, 一个 IPv4 和一个 IPv6。
 - 双协议栈的主机 (或路由器) 记为 IPv6/IPv4, 表明它同时具有两种 IP 地址: 一个 IPv6 地址和一个 IPv4 地址。
 - 双协议栈主机在和 IPv6 主机通信时是采用 IPv6 地址, 而和 IPv4 主机通信时就采用 IPv4 地址。
 - 根据 DNS 返回的地址类型可以确定使用 IPv4 地址还是 IPv6 地址。
-

双协议栈

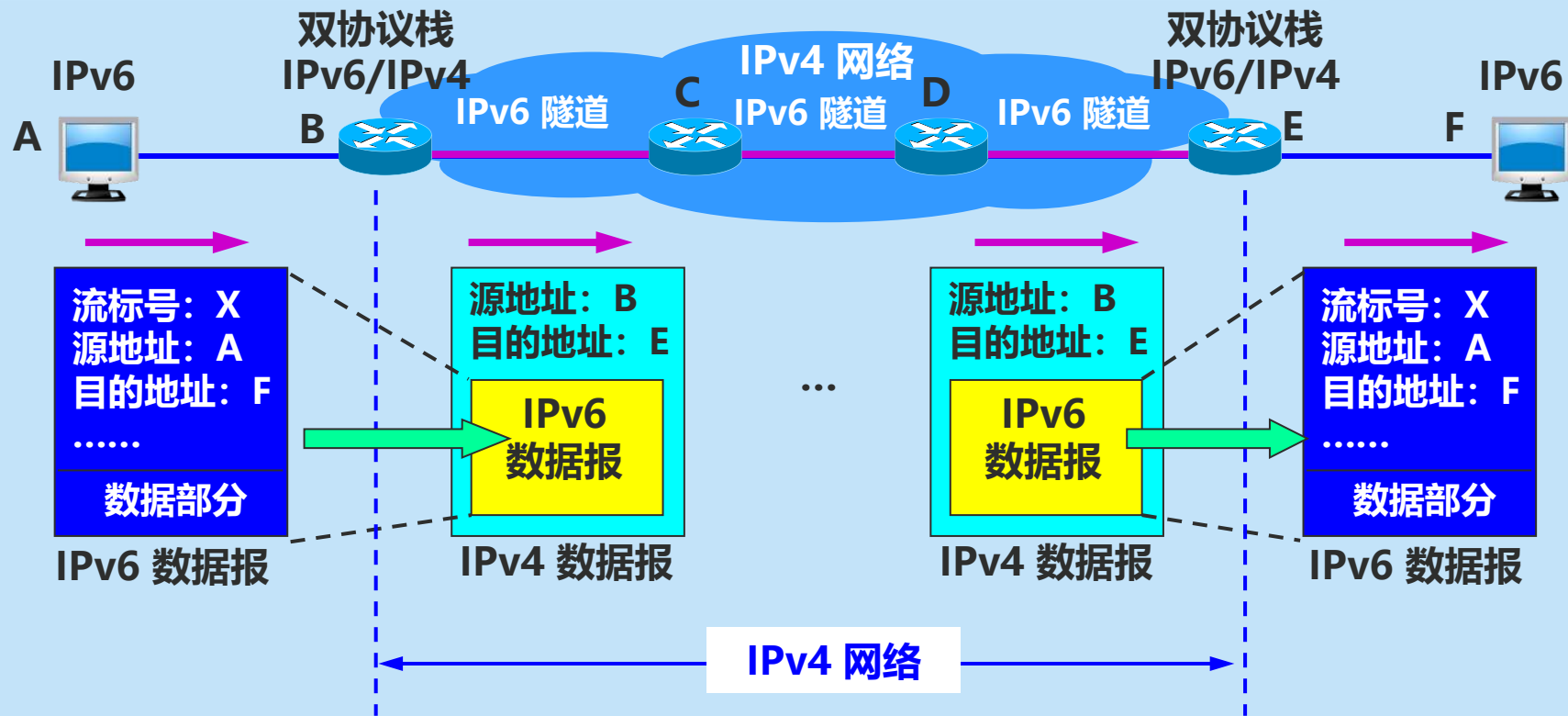


使用双协议栈进行从 IPv4 到 IPv6 的过渡

隧道技术

- 在 IPv6 数据报要进入 IPv4 网络时，把 IPv6 数据报封装成为 IPv4 数据报，整个的 IPv6 数据报变成了 IPv4 数据报的数据部分。
 - 当 IPv4 数据报离开 IPv4 网络中的隧道时，再把数据部分（即原来的 IPv6 数据报）交给主机的 IPv6 协议栈。
-

隧道技术



使用隧道技术进行从 IPv4 到 IPv6 的过渡

IPv6

- IPv6真的能像现在的IPv4一样投入非常大规模的使用，仍需要相当长的一段时间
 - 20年？
 - 想想20年在过去的互联网里面发生了什么进化？
 - 不可能像过去的IPv4和TCP/IP那样采用 “clean-slate” 方法
 - 不仅是技术因素，也有政治、经济因素
 - 从IPv4到IPv6的转化，仍是活跃的研究问题之一
-