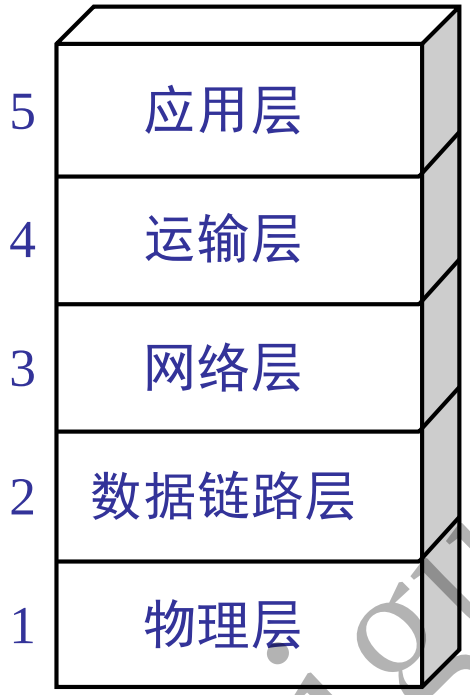
计算机网络

顾 军

计算机学院

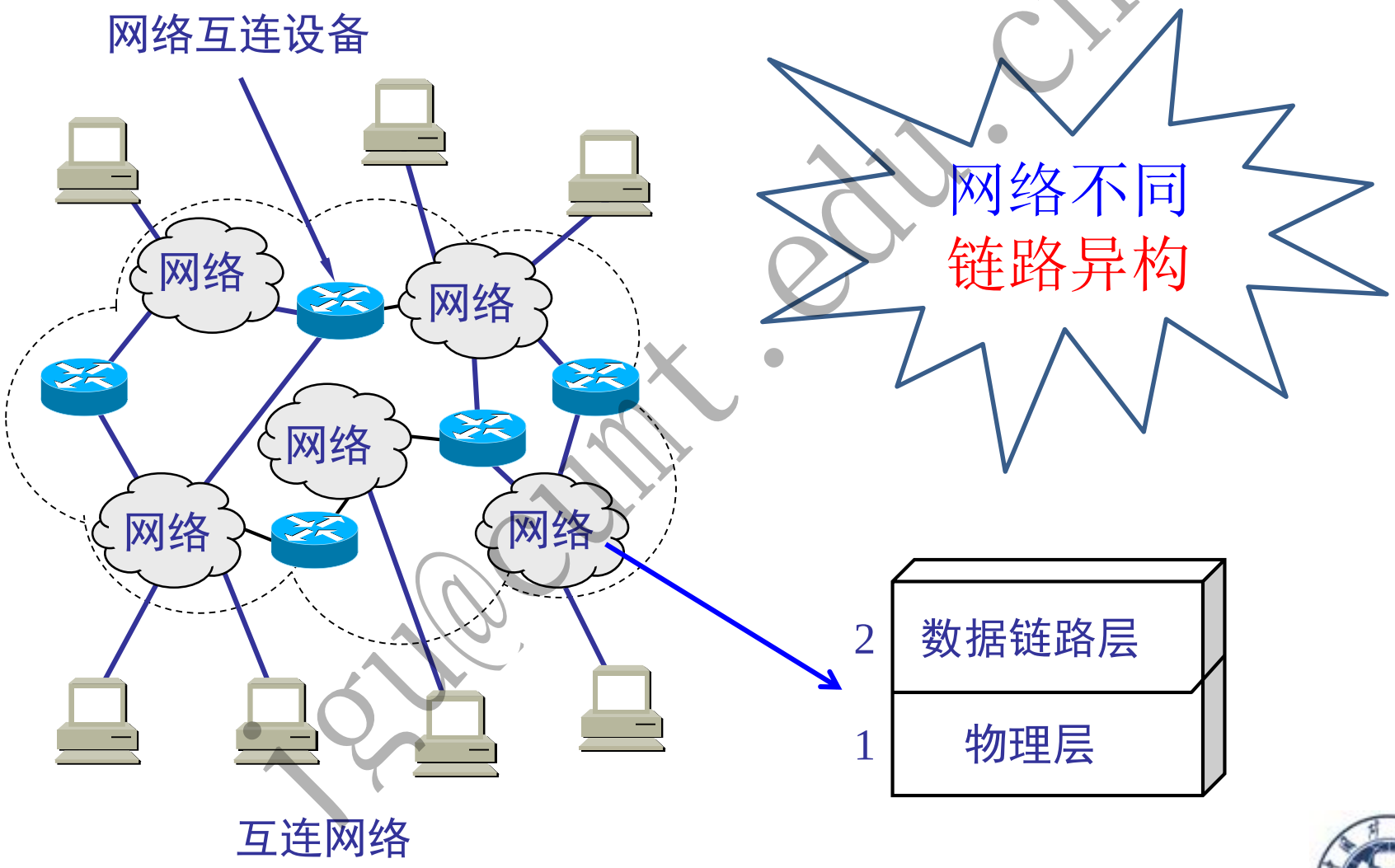
jgu@cumt.edu.cn

专题4：数据包怎么在互联网中寻路和转发？



- 应用层(application layer)
- 运输层(transport layer)
- 网络层(network layer)
- 数据链路层(data link layer)
- 物理层(physical layer)

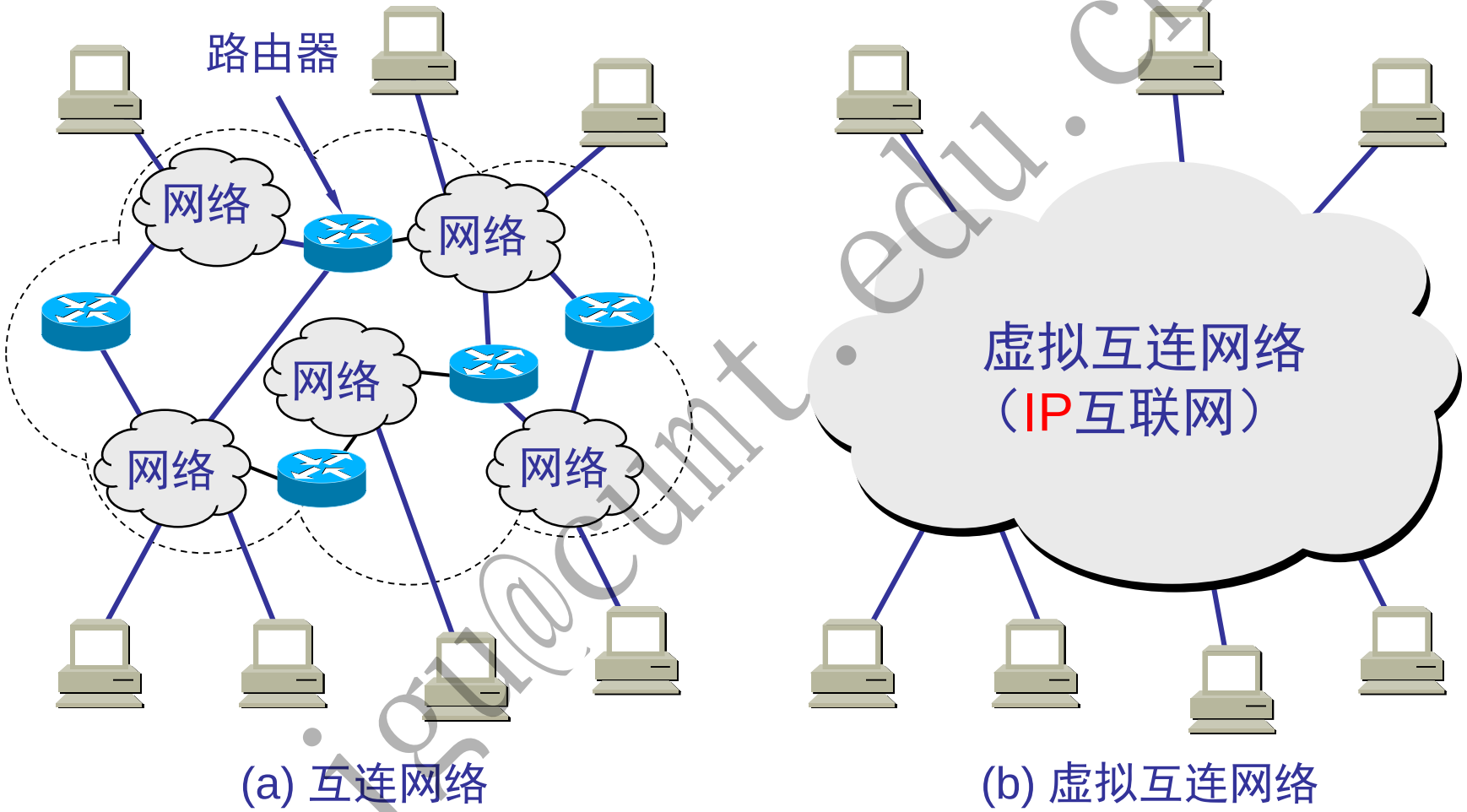
Q1：怎么互连不同的网络？



不同网络的互连并不容易

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

IP网络的引入

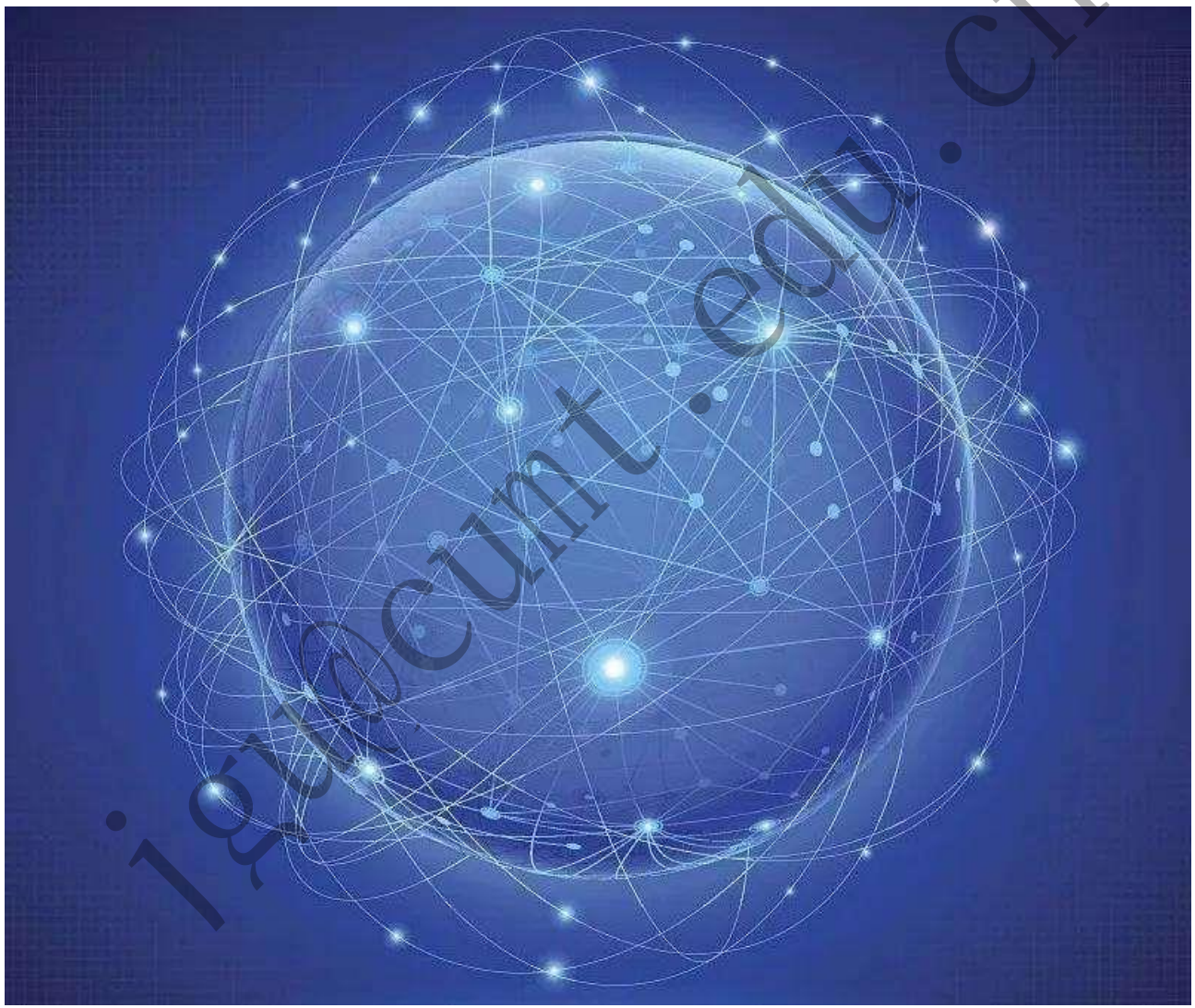


(a) 互连网络　　(b) 虚拟互连网络

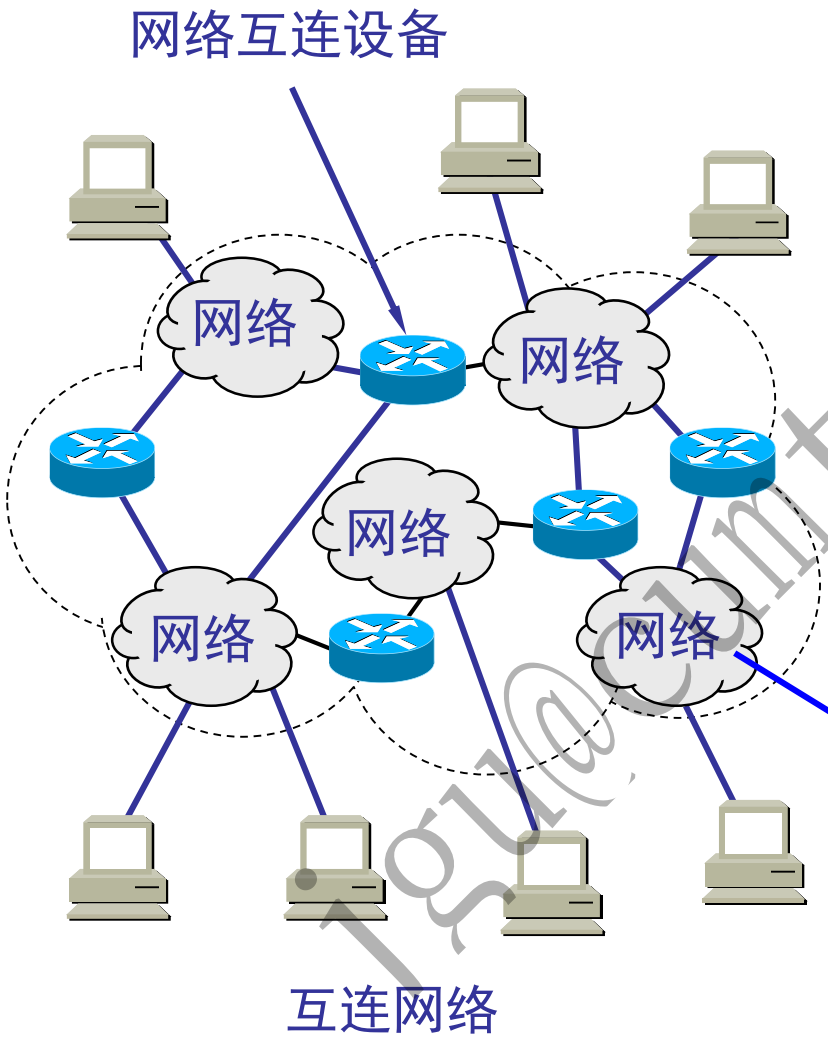
虚拟互连网络的意义

- 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。
- 使用虚拟互连网络的好处是：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。
- 使用 IP 协议的虚拟互连网络简称为IP网。
- 如果在这种覆盖全球的 IP 网的上层使用 TCP 协议，那么就是现在的互联网 (Internet)。

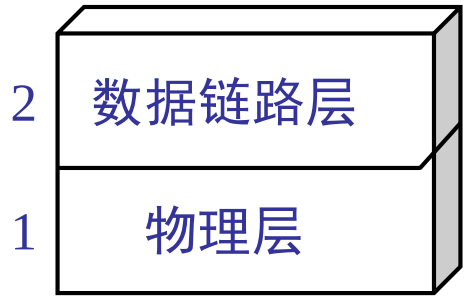
Q2：如何理解IP网络？



IP网络的要点1：IP地址标识设备



- 在互连网络范围使用硬件地址定位各种网络设备
- 基于硬件地址在互连网络范围内转发数据

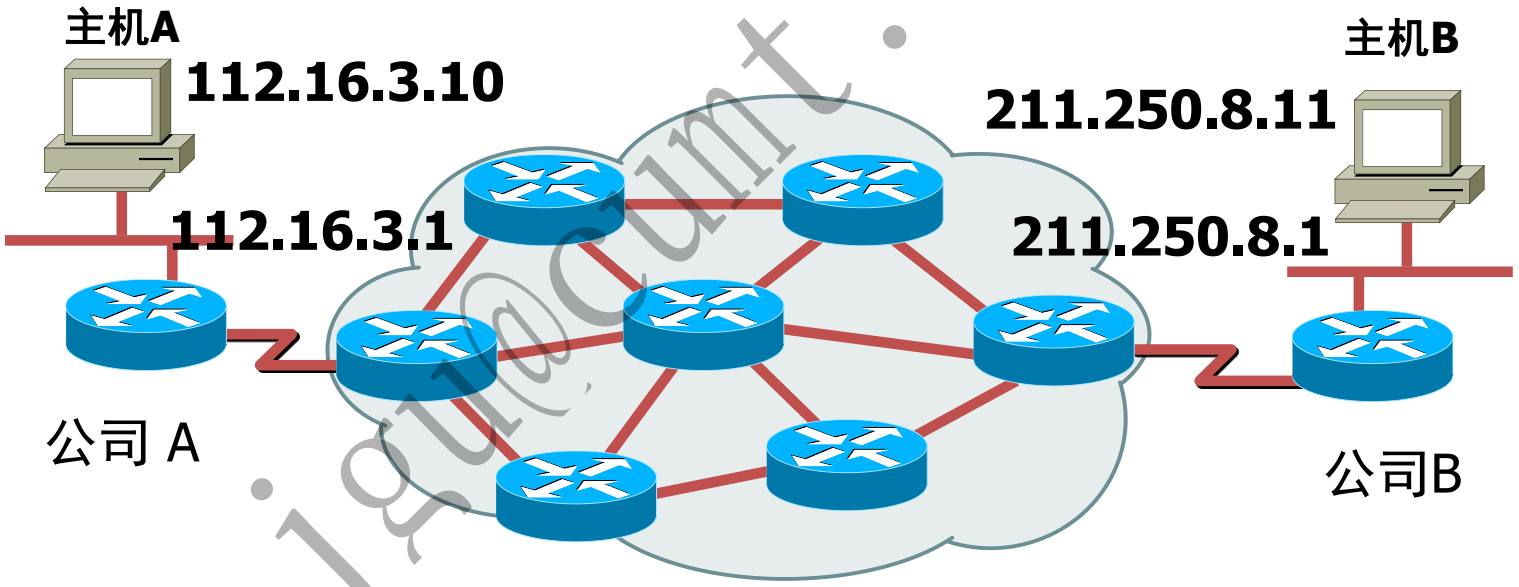


- 全世界存在着各式各样的网络，它们使用不同的硬件地址。要使这些异构网络能够互相通信就必须进行非常复杂的硬件地址转换工作，因此几乎是不可能的事。
- 即使用统一的硬件地址，也会因为硬件地址不包含地理位置信息而无法实现跨网络的互连网范围的有效寻址。

应用层	
表示层	
会话层	
传输层	
网络层	
数据链路层	网桥　交换机　网卡
物理层	中继器　集线器　调制解调器

全网惟一的IP地址使得互连网上的端系统间的通信成为可能

- IP 地址就是给每个连接在因特网上的主机（或路由器）分配一个在全世界范围是惟一的 32 位的标识符，也称为是IPv4地址。



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
则更加便于使用

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

IP 地址由互联网名字和数字分配机构进行分配

ICANN (Internet Corporation for Assigned Names and Numbers)

全球5大RIR(Regional Internet Registry)机构

1.RIPE(Registry IP Europeans)欧洲IP地址注册中心--服务于欧洲、中东地区和中亚地区；
2.LACNIC(Latin American and Caribbean Internet Address Registry)拉丁美洲和加勒比海Internet地址注册中心--服务于中美、南美以及加勒比海地区；
3.ARIN(American Registry for Internet Numbers)美国Internet编号注册中心--服务于北美地区和部分加勒比海地区；
4.AFRINIC(Africa Network Information Centre)非洲网络信息中心--服务于非洲地区；
5.APNIC(Asia Pacific Network Information Centre)亚太地址网络信息中心--服务于亚洲和太平洋地区的国家。

中国互联网信息中心
CNNIC

http://www.cnnic.net.cn/

IP地址先后有三种编排方法

- 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。

IP地址编排方法的不同

决定了路由转发方式的不同

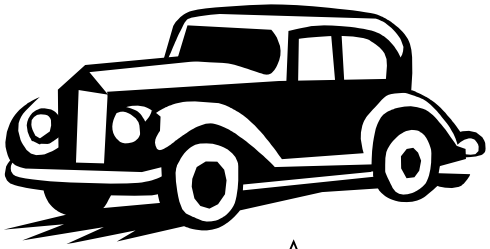
IP网络的要点2： 路由器互连网络

应用层	
表示层	
会话层	
传输层	
网络层	路由器
数据链路层	网桥　交换机　网卡
物理层	中继器　集线器　调制解调器

路由器在网际互连中的作用

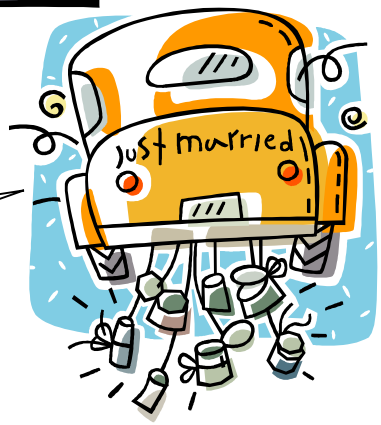
在一个十字路口…

快闪开，我要迟到了！



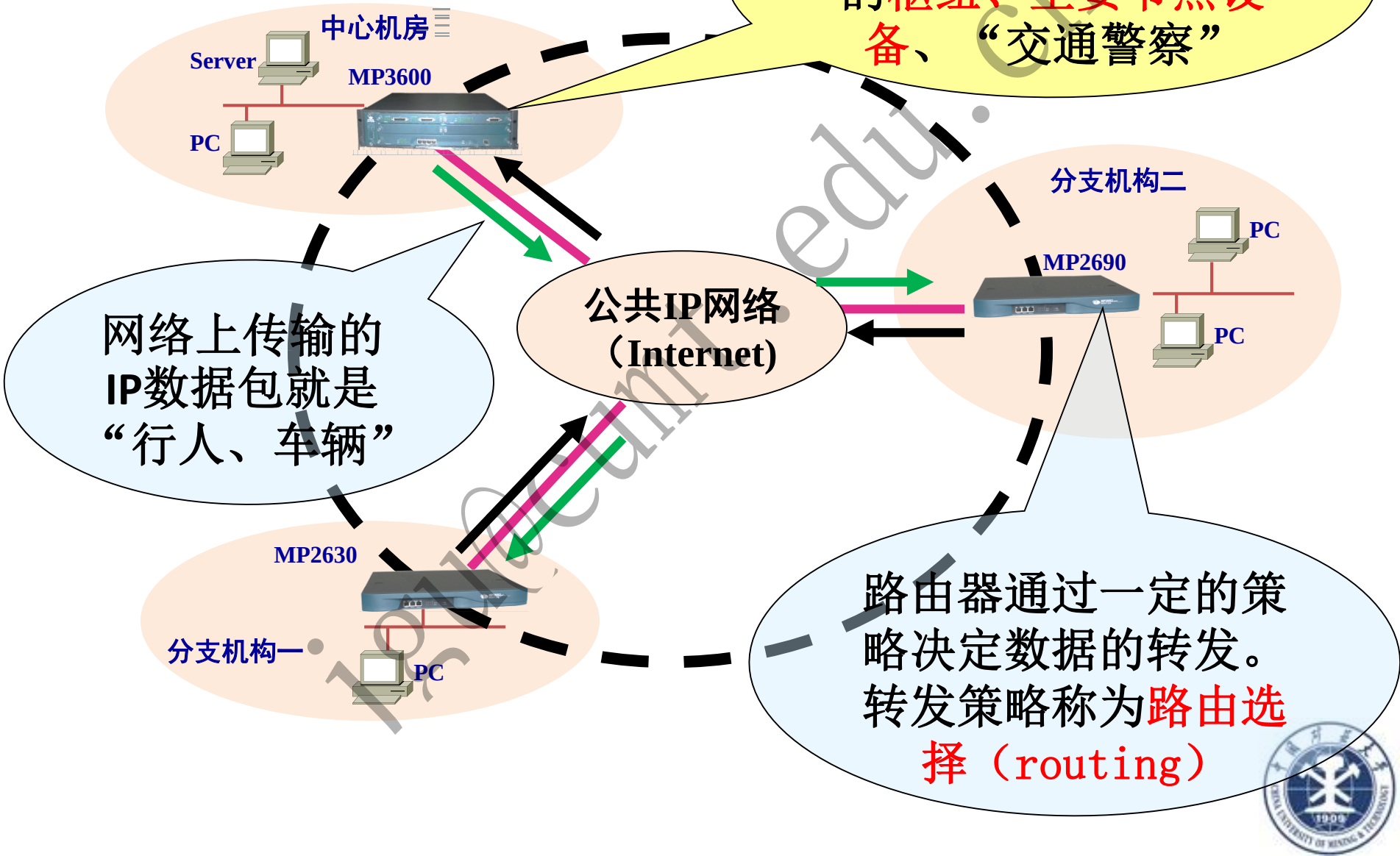
去九寨沟怎么走啊？

呵呵，开老爷车出去溜达溜达

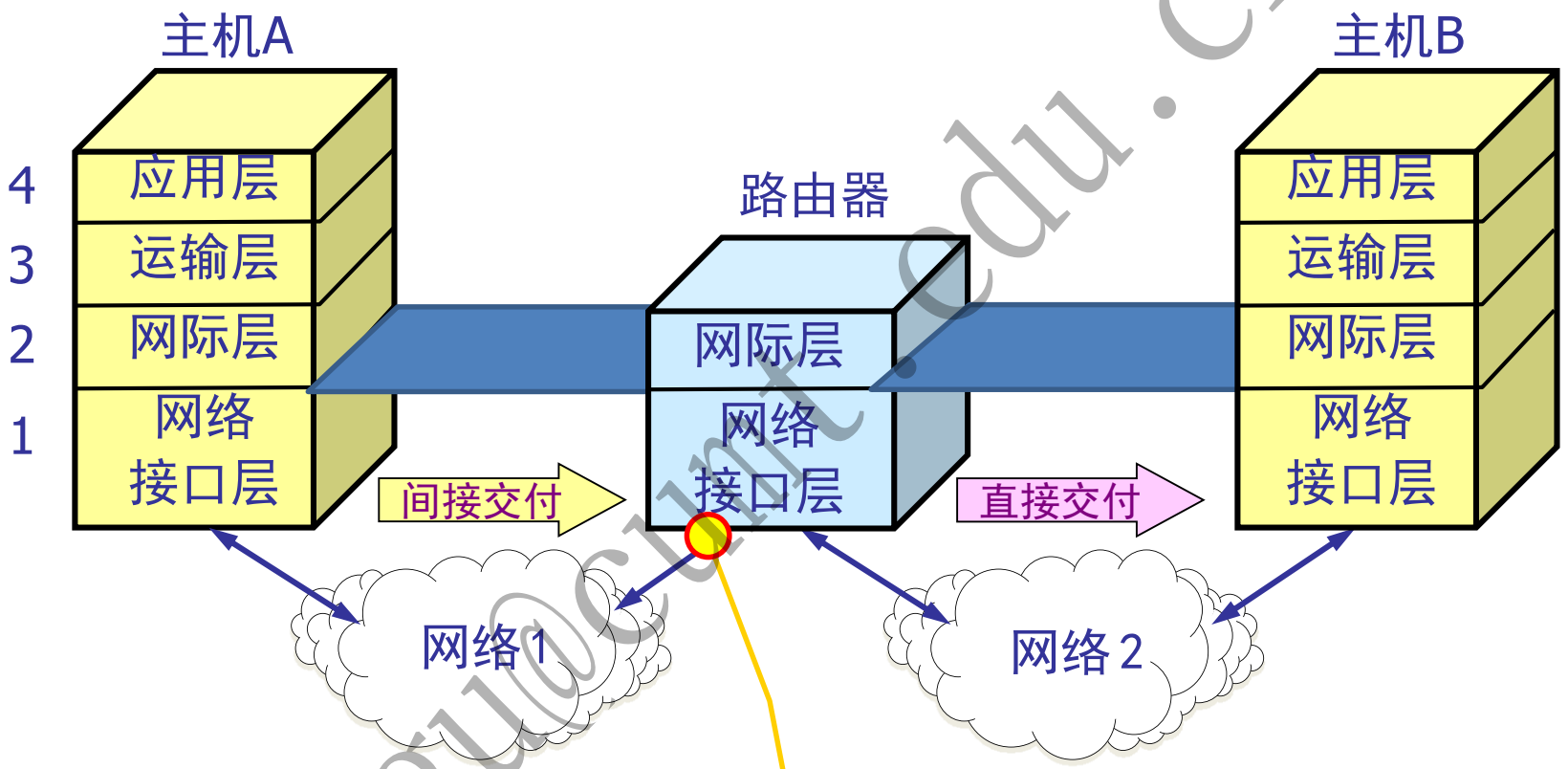


浪漫之旅开始了～

路由器基本概念

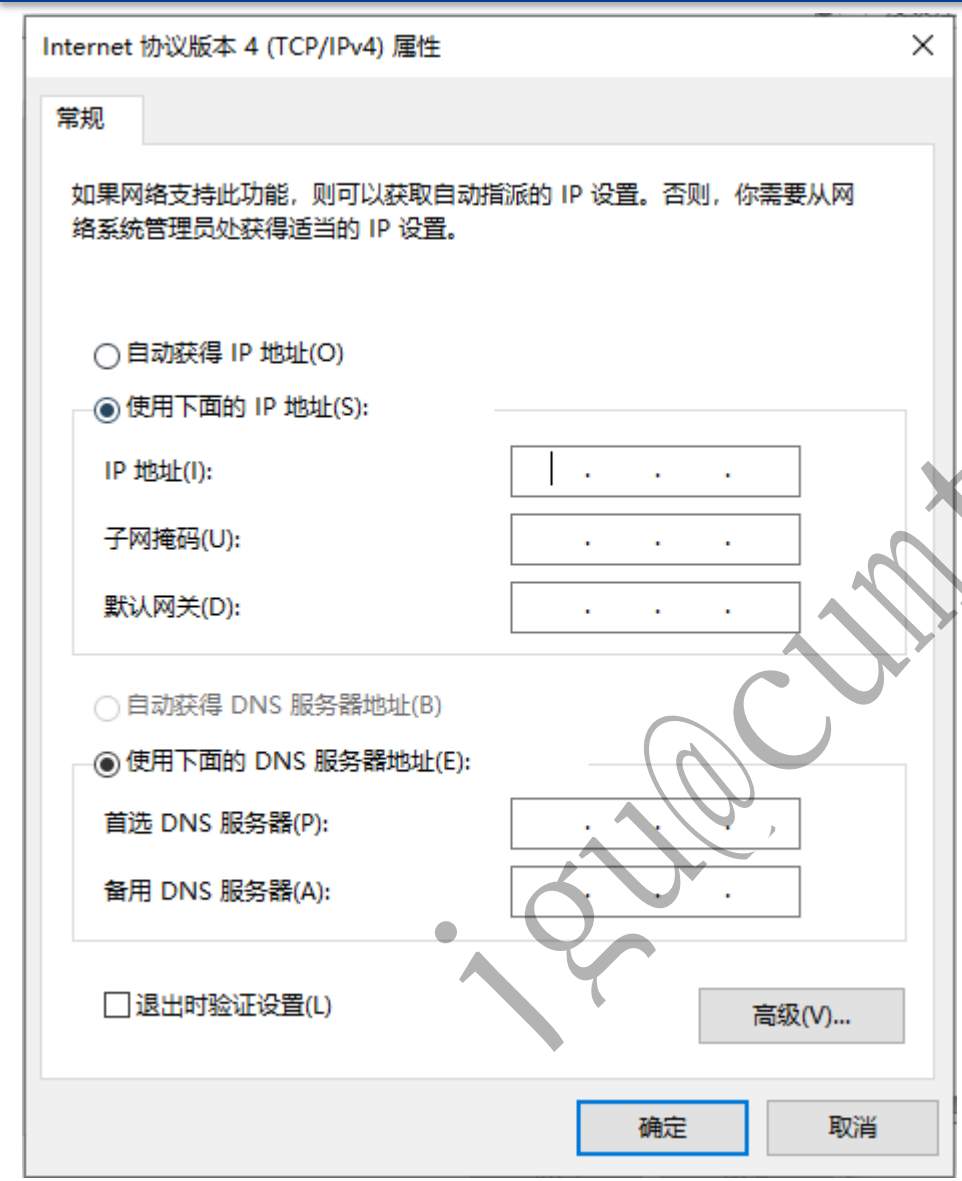

路由器是IP互联网络的枢纽、主要节点设备、“交通警察”
中心机房
Server
MP3600
PC
分支机构二
PC
MP2690
PC
公共IP网络（Internet）
网络上传输的IP数据包就是“行人、车辆”
MP2630
分支机构一
PC
路由器通过一定的策略决定数据的转发。转发策略称为路由选择（routing）

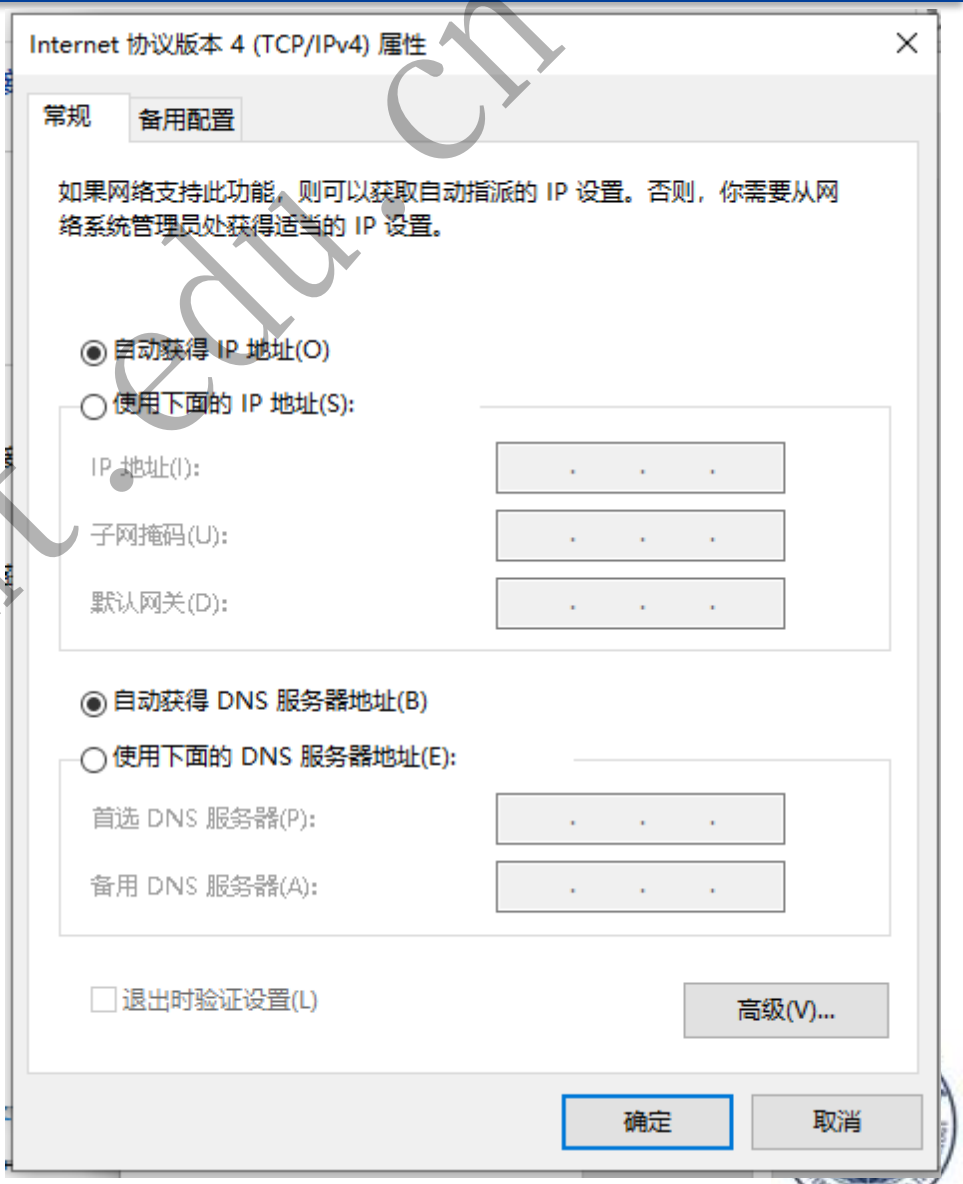
IP网络的要点3：间接交付+直接交付



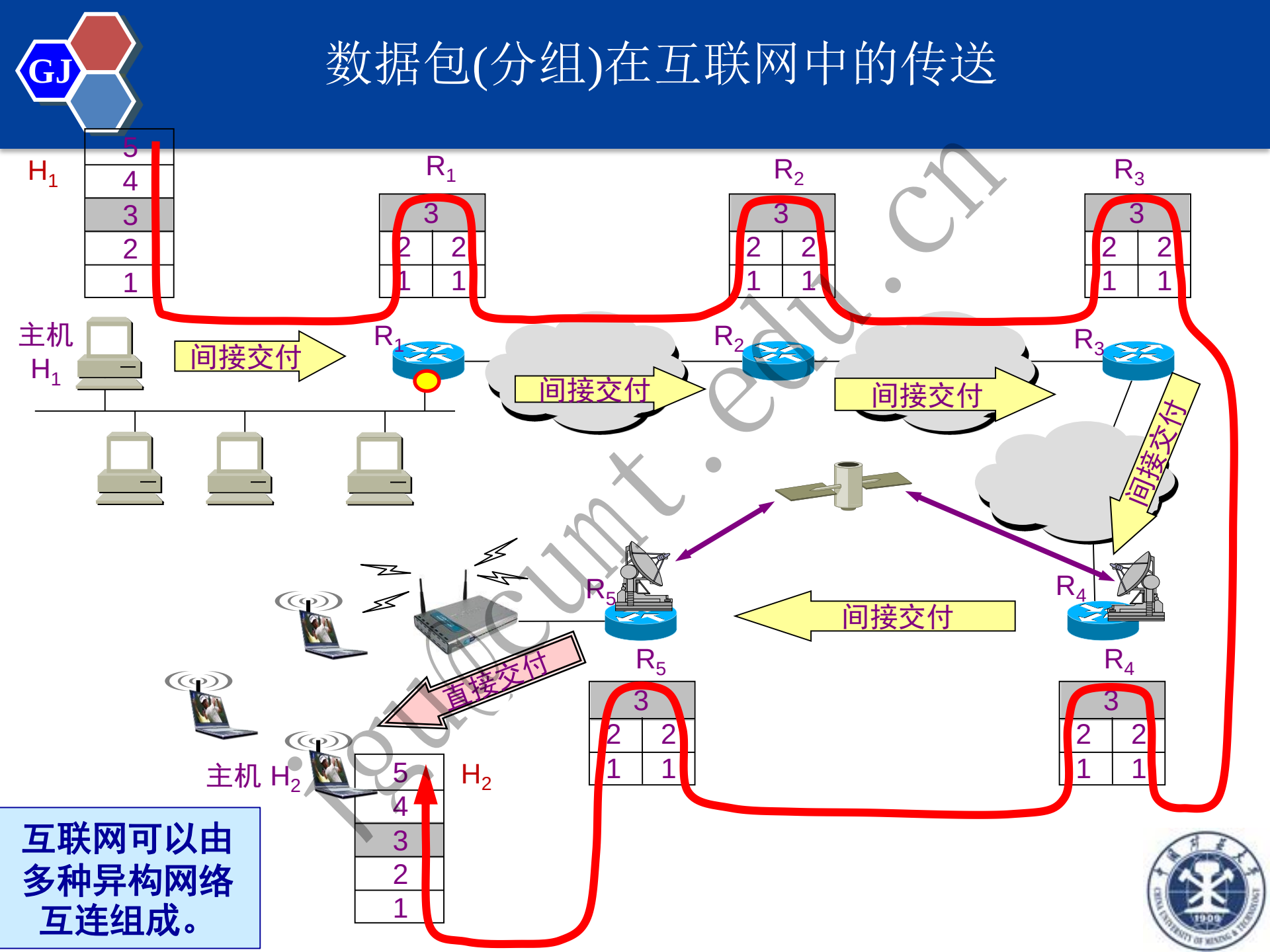
默认网关：一个网络通向其它网络的IP地址，一般是路由器某个接口的IP地址。当本地网络中的主机需要和外网的主机通信时，就把数据包发给默认网关，由这个网关来处理数据包。

主机的IPv4协议属性配置



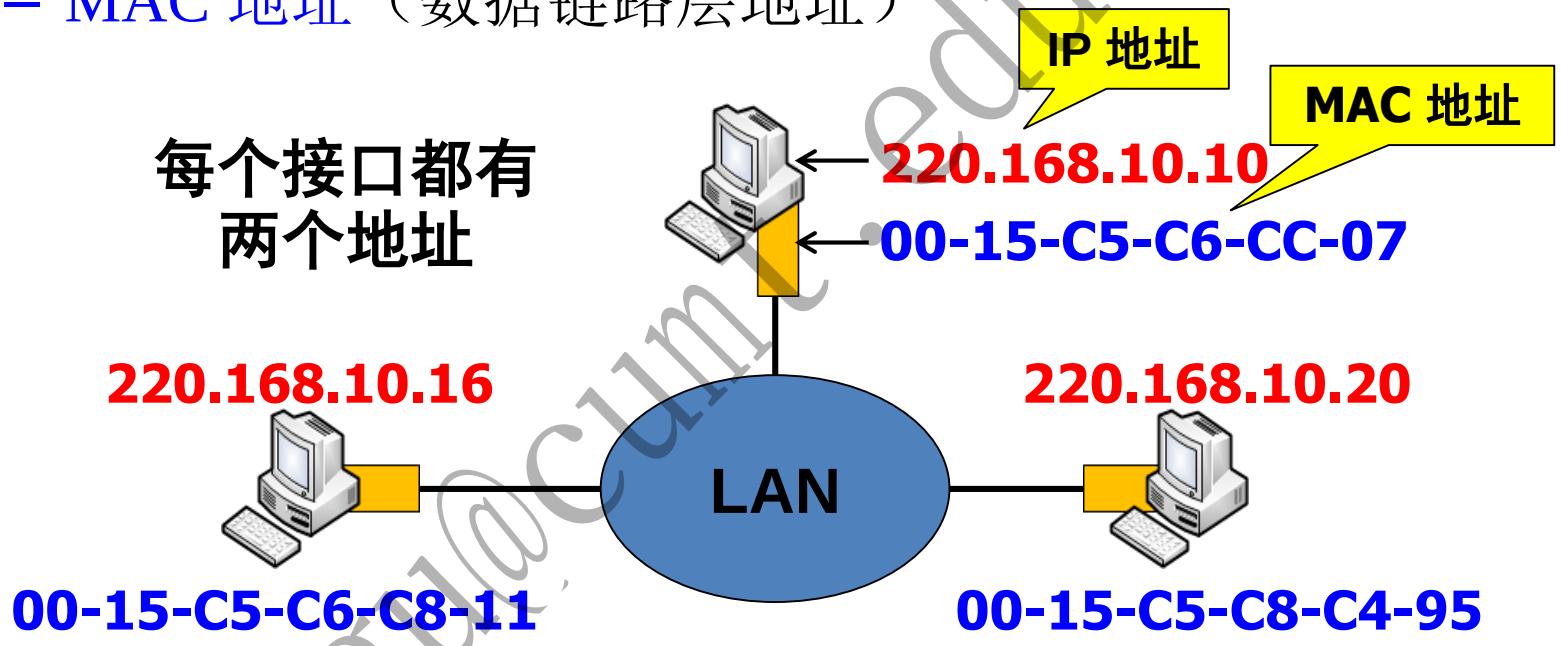


数据包(分组)在互联网中的传送



IP网络的要点4： IP地址+MAC地址

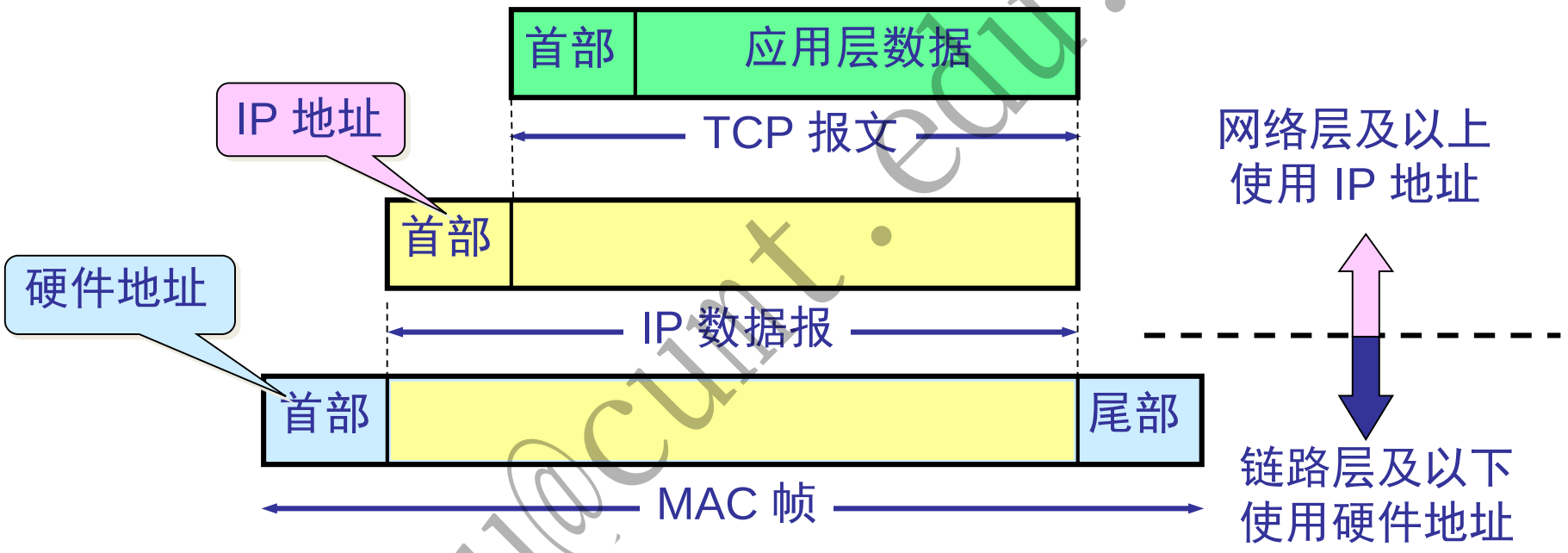
- 通信时使用了两个地址：
  - IP 地址（网络层地址）
  - MAC 地址（数据链路层地址）



IP地址虽然能够标识通信双方，但是并不能直接用来进行通信，还需要借助MAC地址来找到对方。

IP地址与硬件地址的关系

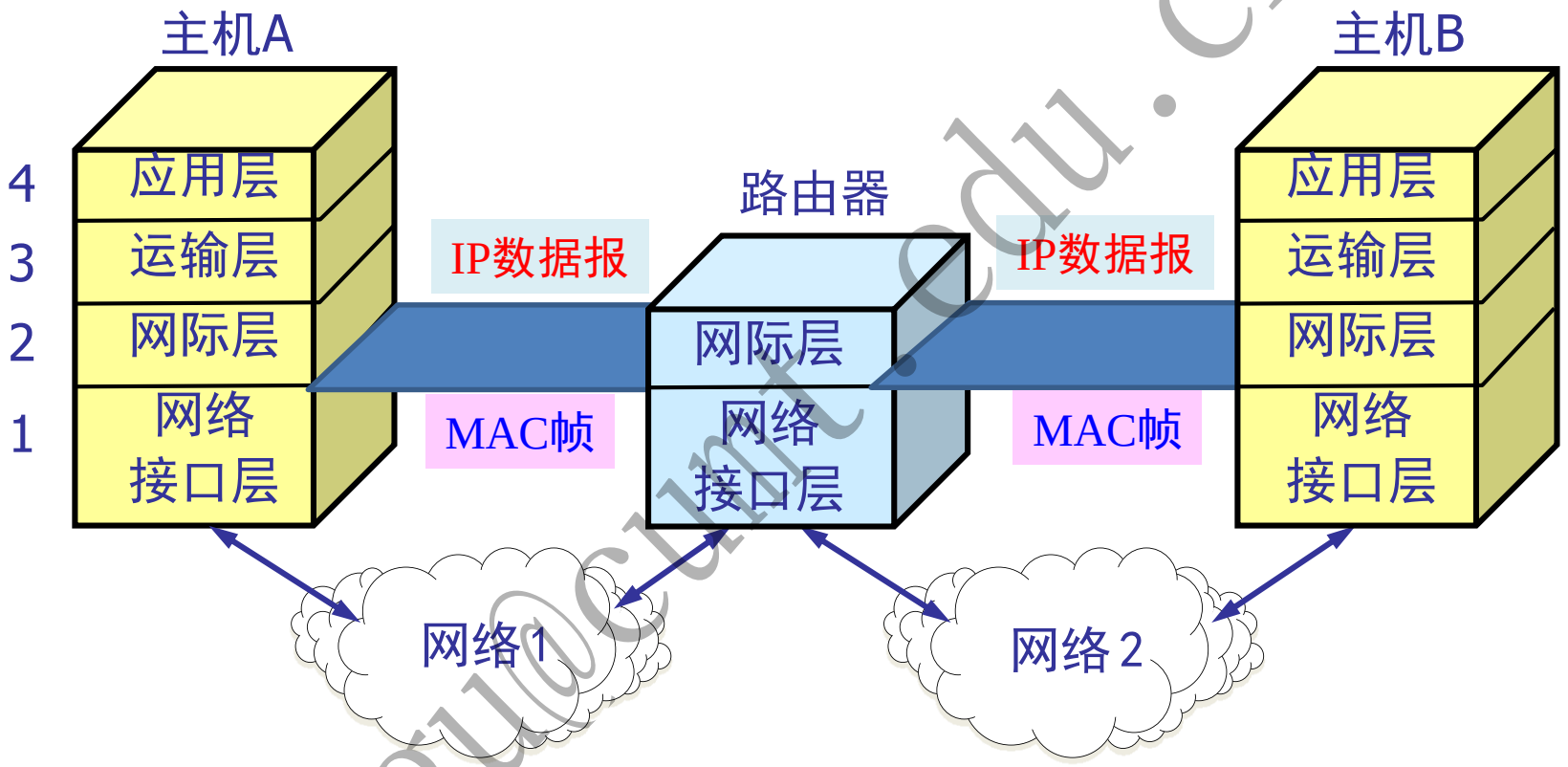
IP 地址是用软件实现的，是逻辑地址。



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。

硬件地址是用硬件实现的，是物理地址。

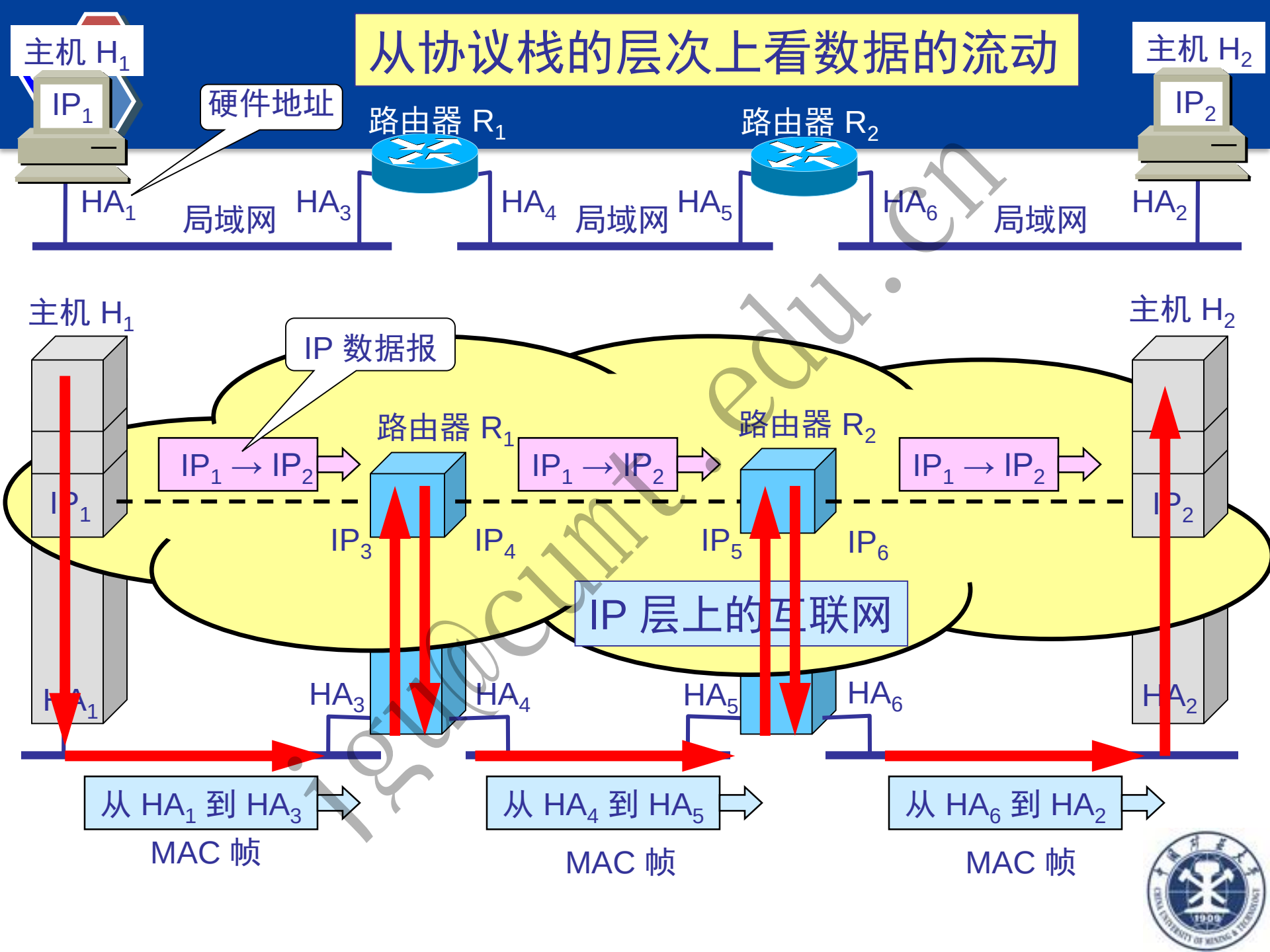
IP网络的要点5：IP数据报+MAC帧



通信的路径

H_1→经过 R_1 转发→再经过 R_2 转发→H_2

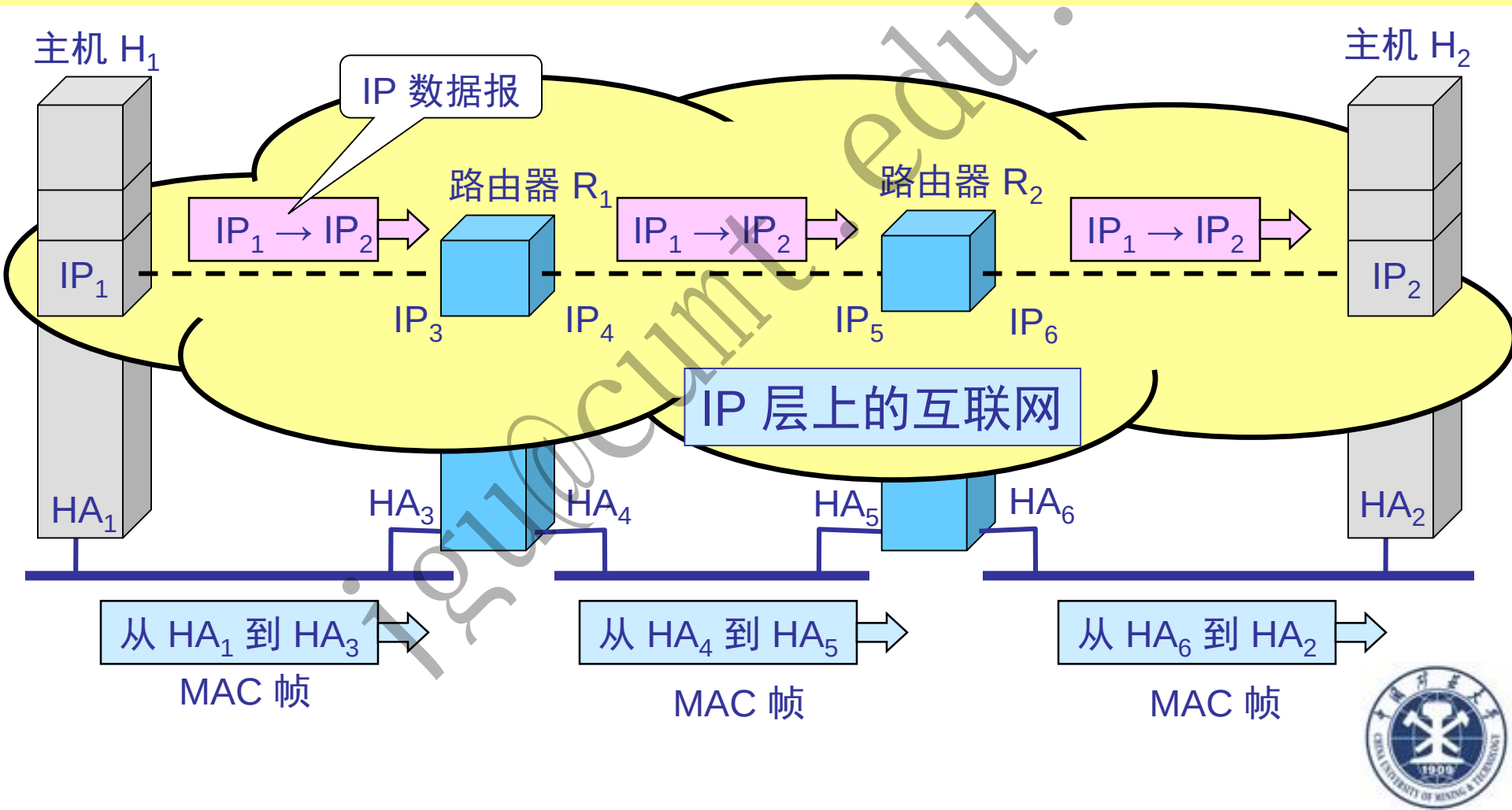
从协议栈的层次上看数据的流动



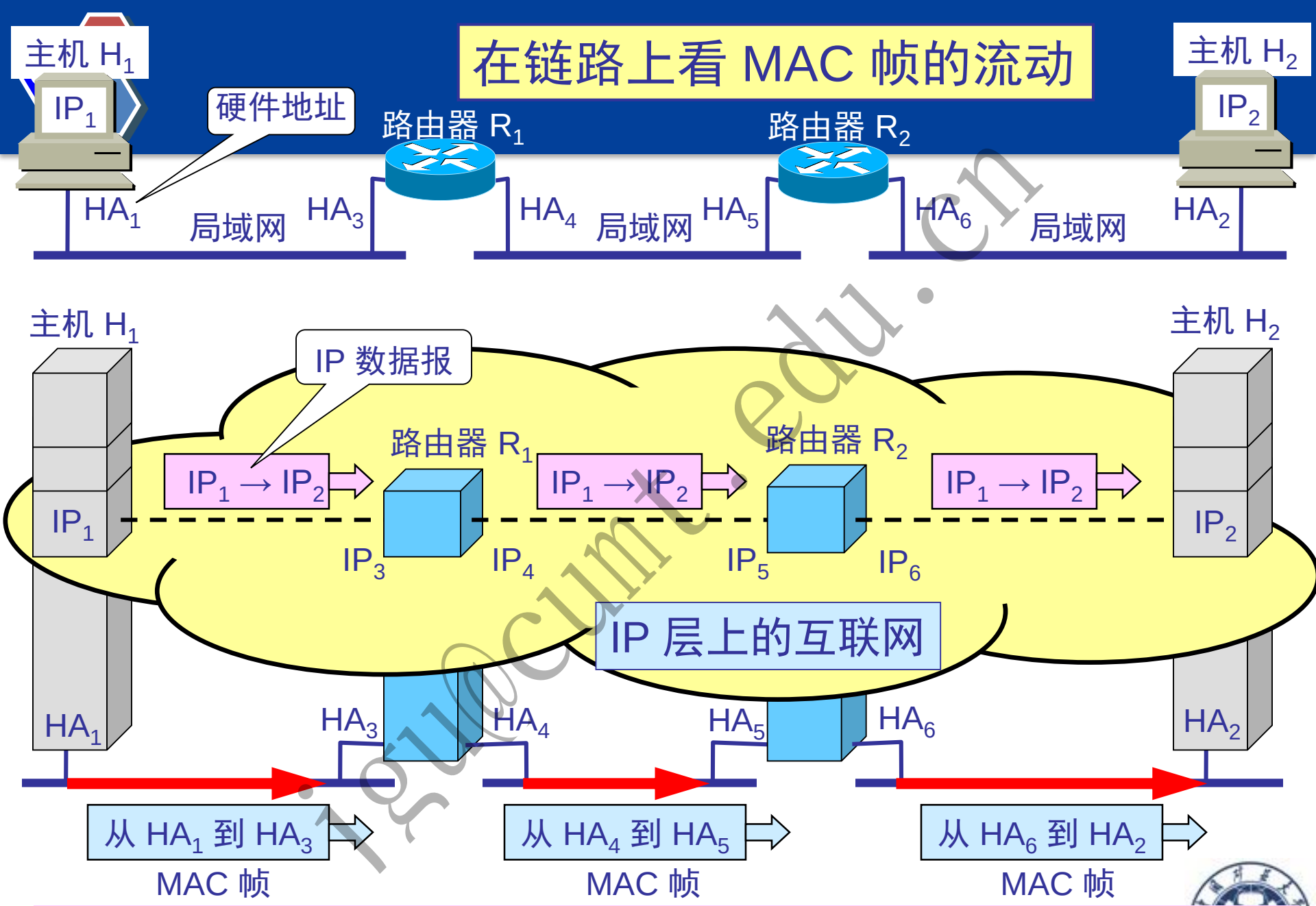
IP 数据报封装在MAC 帧中

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



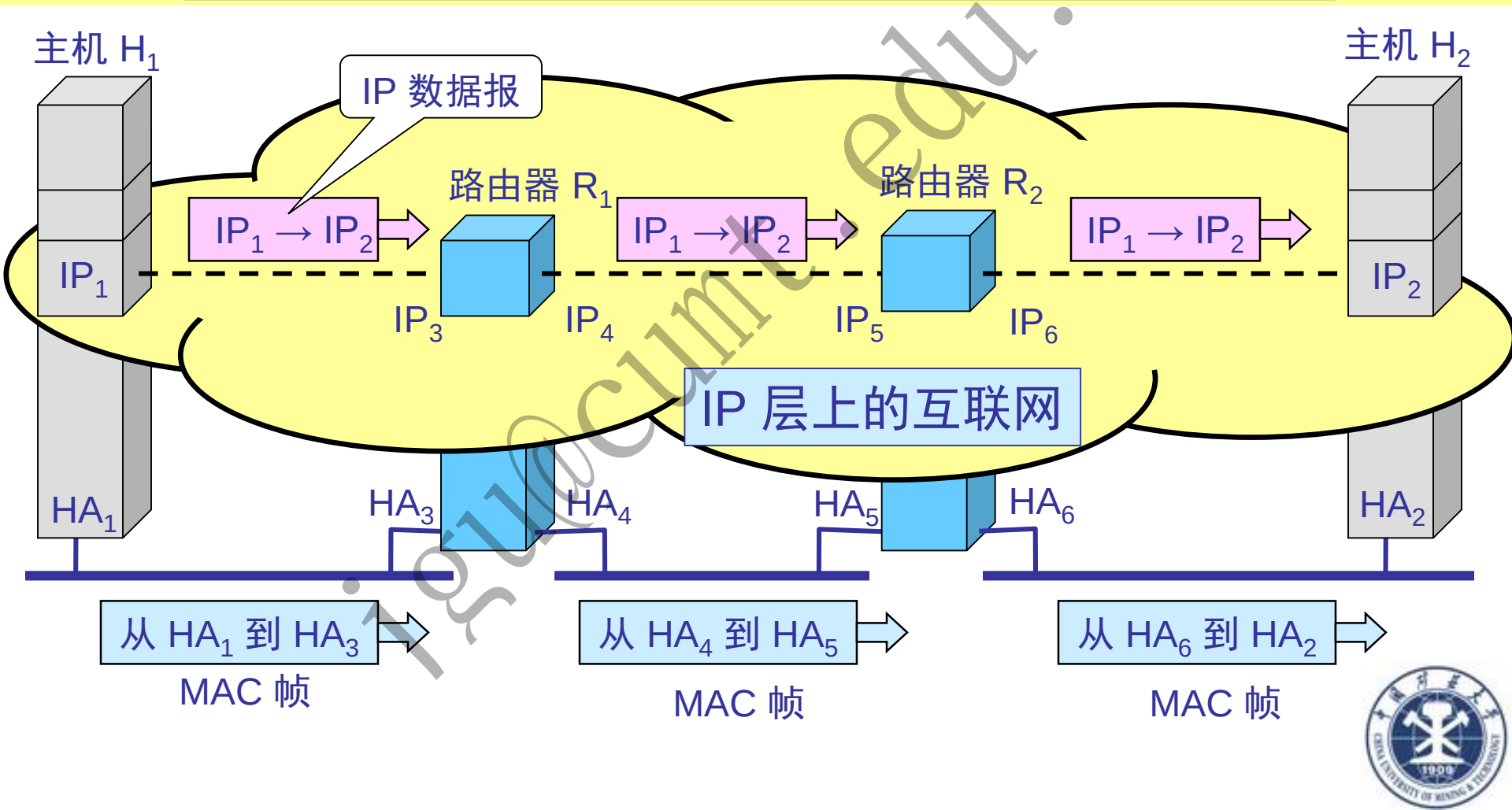
在链路上看 MAC 帧的流动



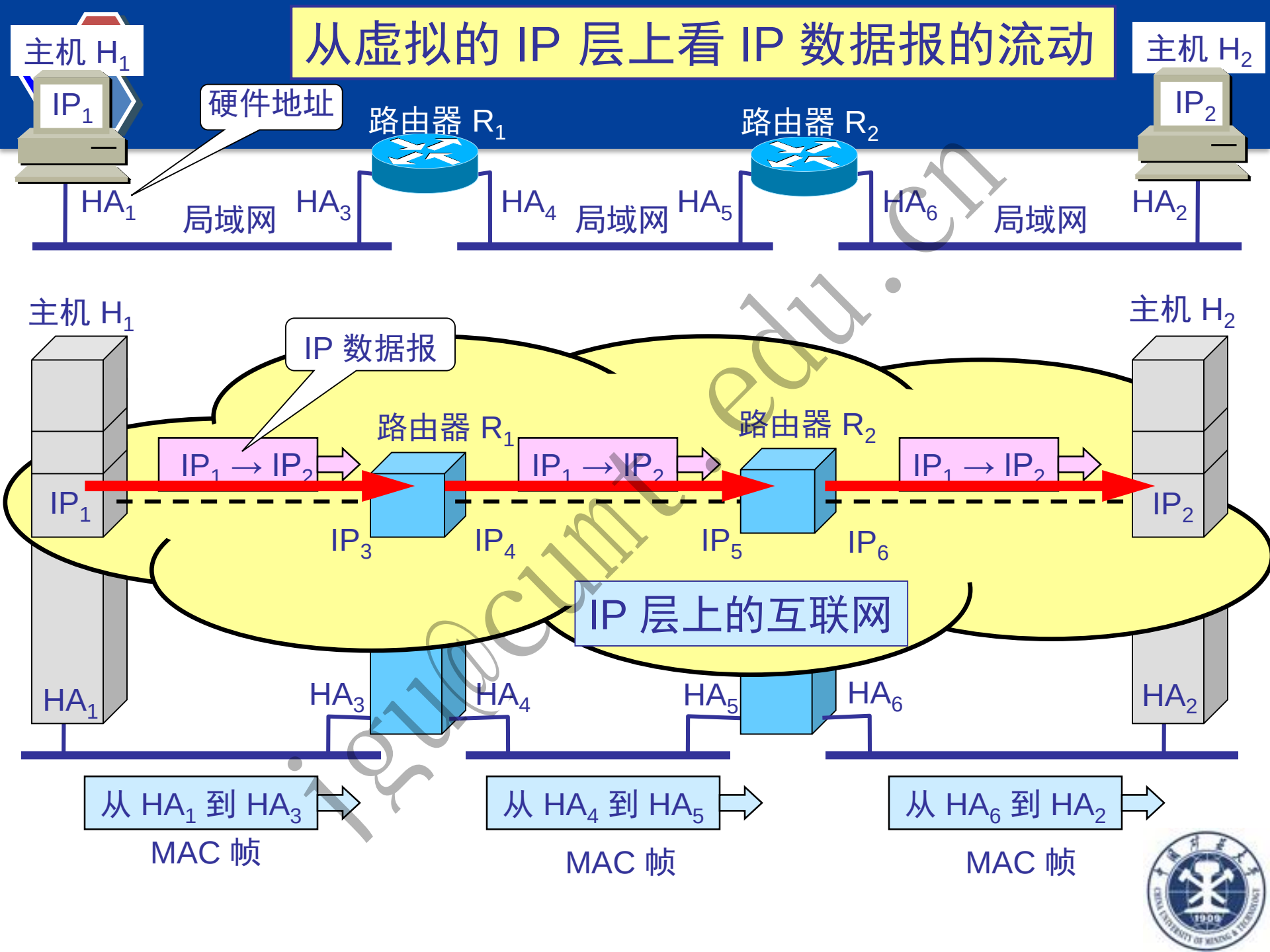
MAC帧在每个局域网链路上进行封装，因局域网异构而有差异。

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



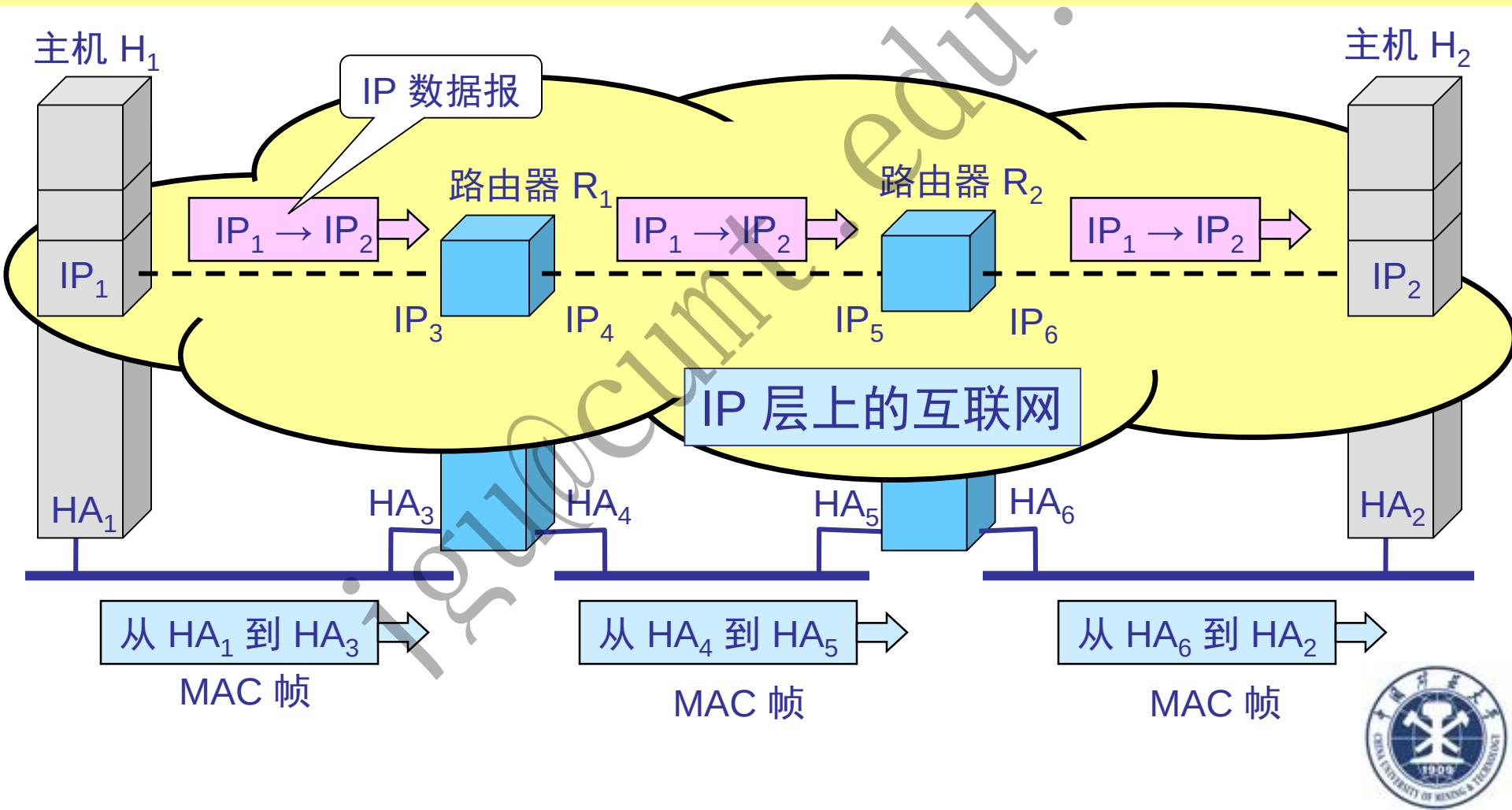
从虚拟的 IP 层上看 IP 数据报的流动



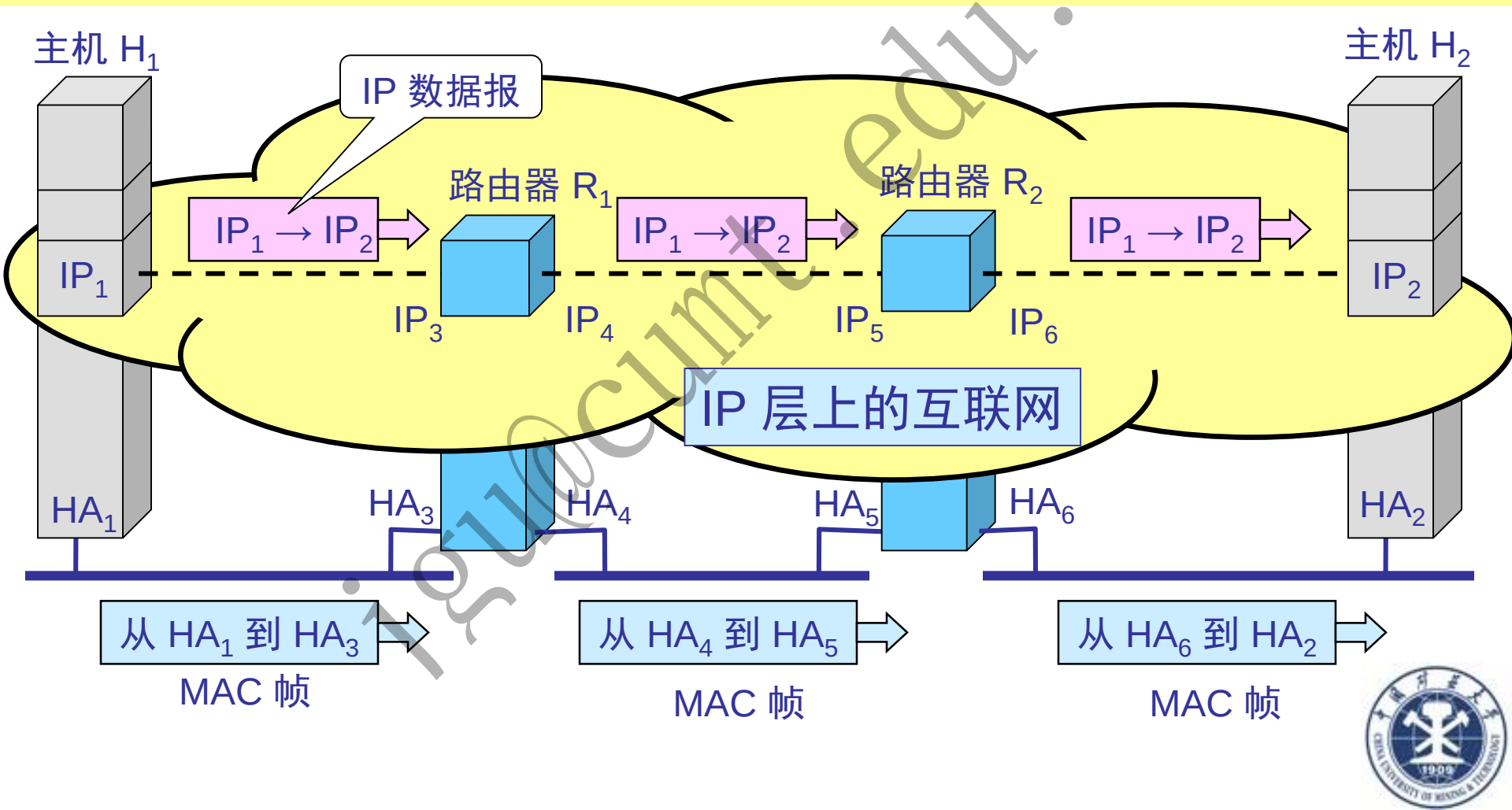
在 IP 层抽象的互联网上只能看到 IP 数据报

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

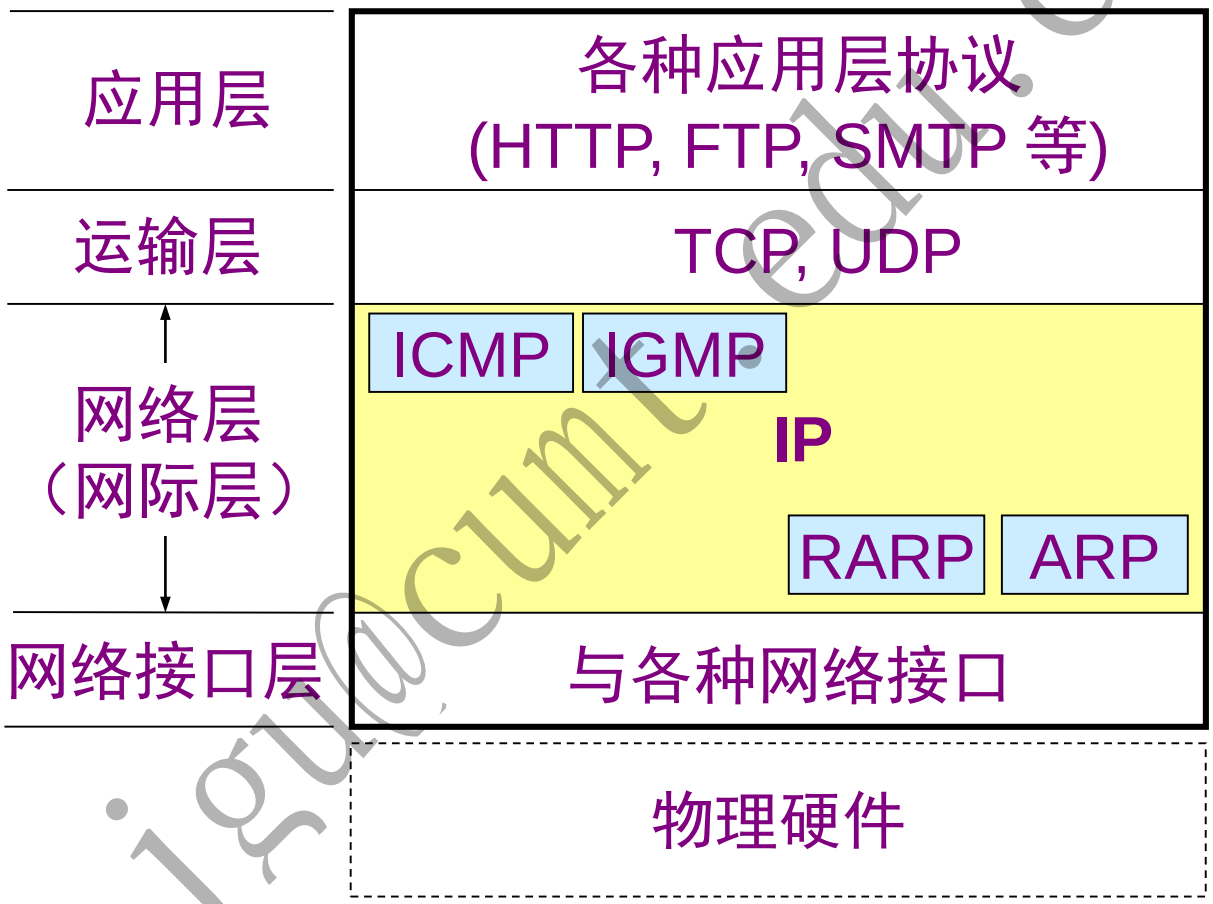
两个路由器的 IP 地址并不出现在 IP 数据报的首部中



路由器只根据目的站的 IP 地址的网络号进行路由选择

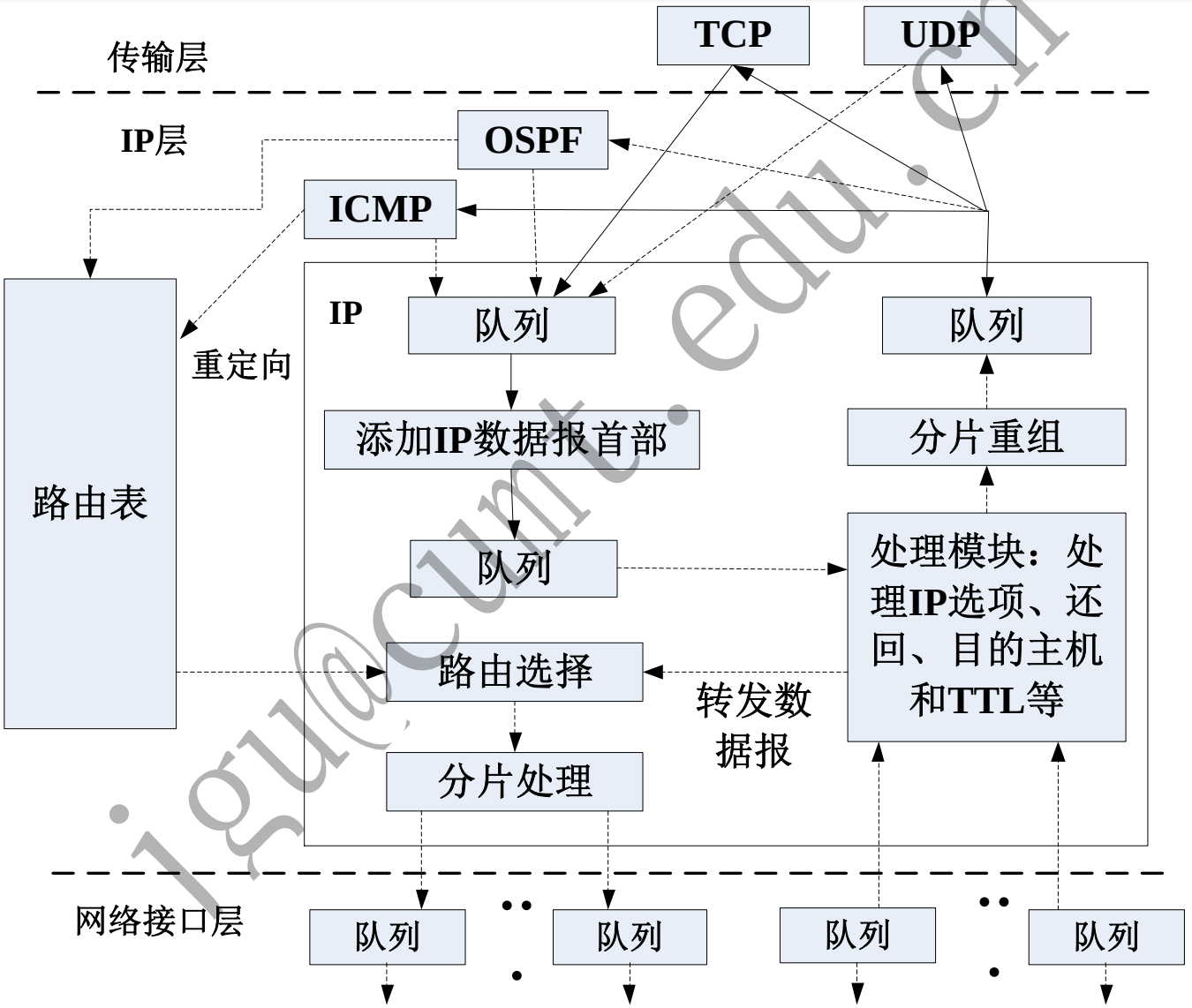


IP网络的要点6：网络层协议是配套的



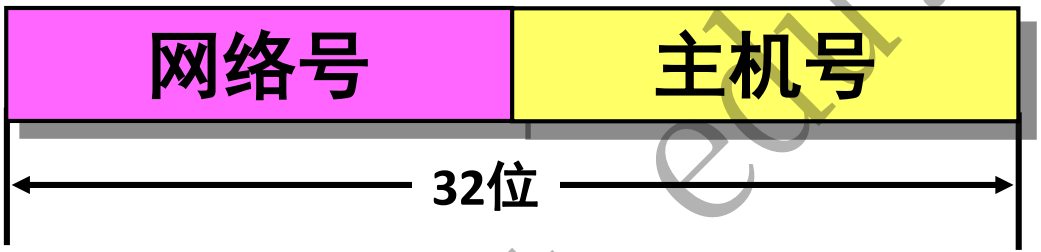
网络层的 IP 协议及配套协议

IP模块功能分解



Q3：分类 IP 地址如何表示？

- 这种两级的 IP 地址结构如下：

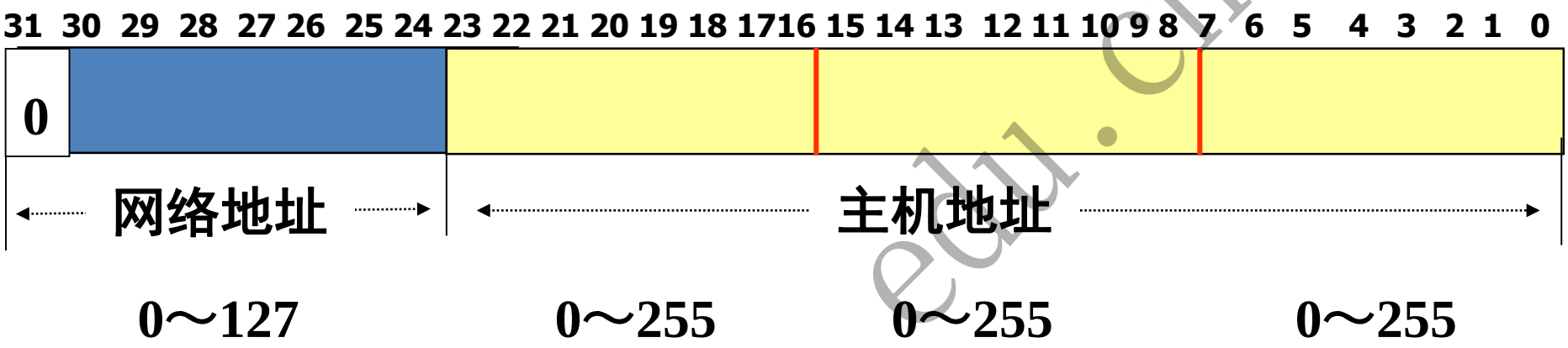


- 这种两级的 IP 地址可以记为：

IP 地址 ::= { <网络号>, <主机号>}

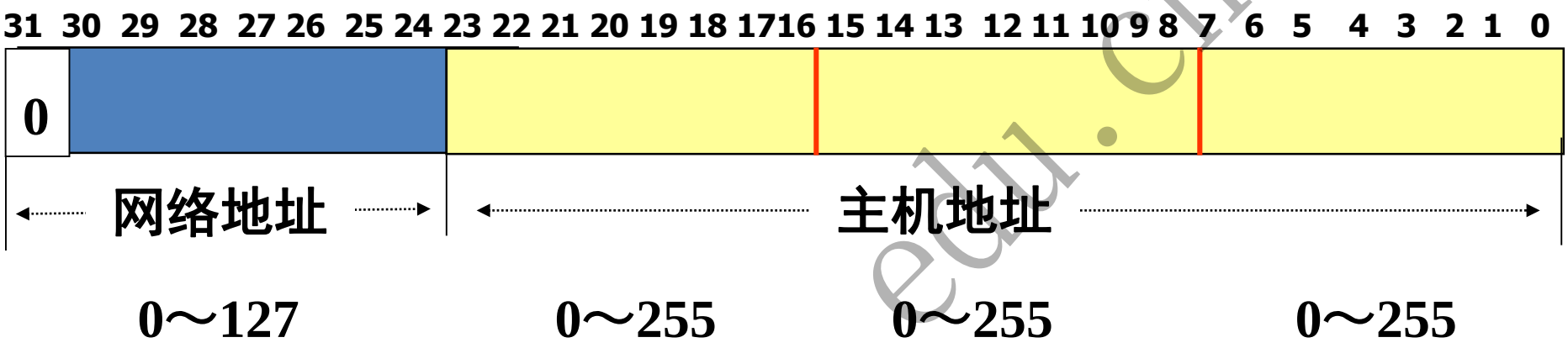
::= 代表“定义为”

A类地址



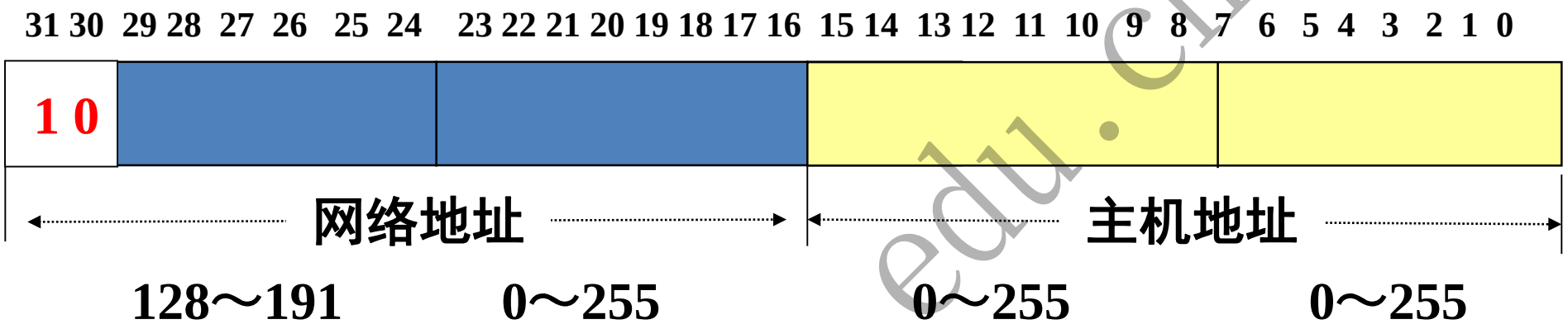
- 从高位起，前1位为“0”，第1字节用十进制表示的取值范围为“0～127”
- 前1字节标识网络地址，后3字节标识主机地址
- 具有A类地址特征的网络总数为（2^7 －2）= 126个
  - 网络地址全0表示的是“本网络”；
  - 网络地址全1，即127（01111111），表示回环地址

A类地址



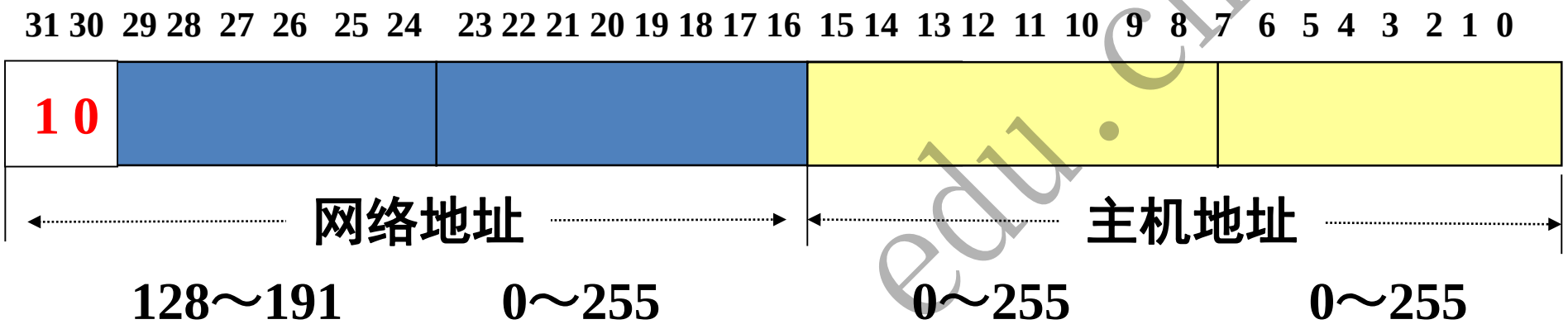
- ✓ 目的地址为环回地址的IP数据报只用于本主机的进程之间的通信之用，会发送到任何网络。
- ✓ 每个网络最多可容纳（$2^{24}-2$）台主机
  - 主机号全0表示的是网络地址，不能标识主机
  - 主机号全1标识的是网络上的所有主机，即广播地址
- ✓ A类地址空间为2^{31}，占整个IP地址空间的一半

B类地址



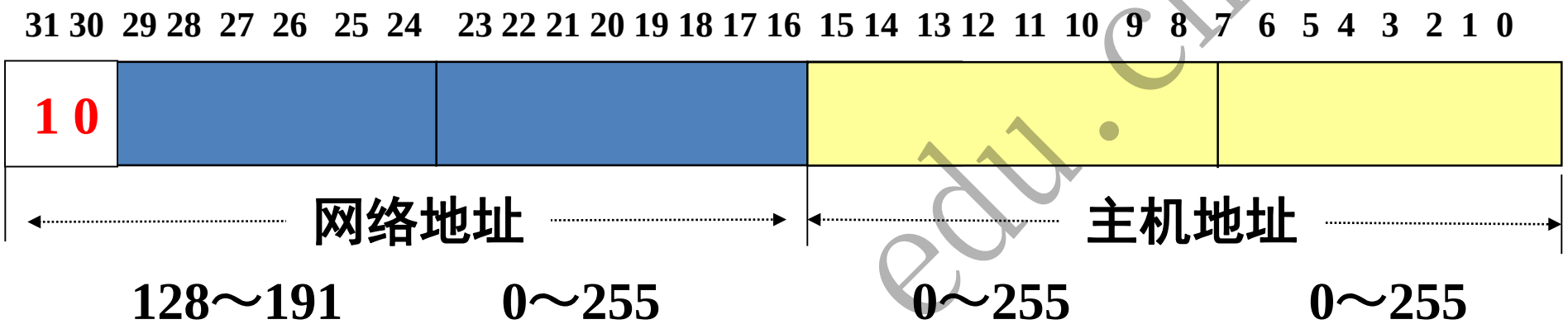
- 从高位起，前2位为“10”，第1字节用十进制表示的取值范围为“128～191”
- 前2字节标识网络地址，后2字节标识主机地址
- 网络地址不存在全0或全1，不存在网络总数减2的问题每个网络最多可容纳（$2^{16}-2$）台主机
- 具有B类地址特征的网络总数为 $2^{14}-1$个

B类地址



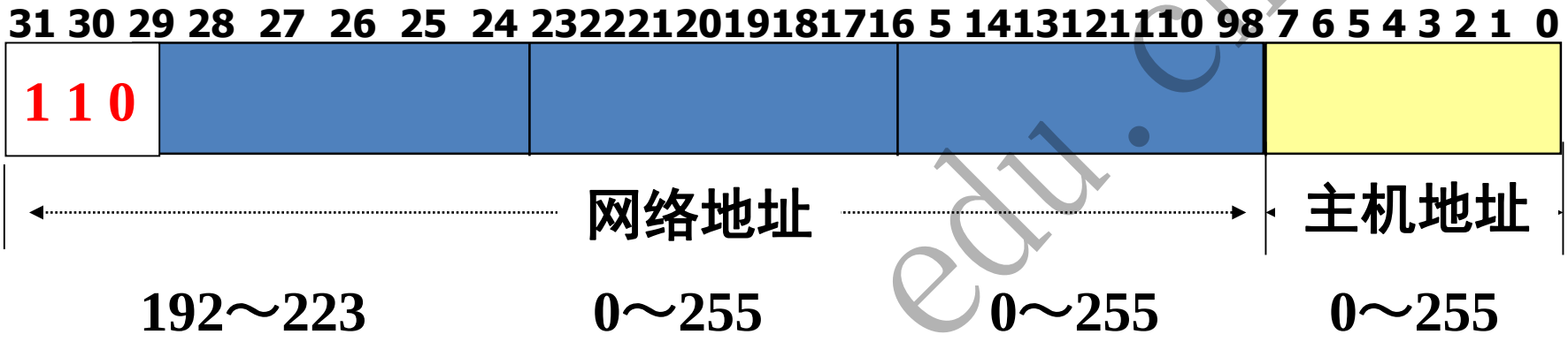
- 从高位起，前2位为“10”，第1字节用十进制表示的取值范围为“128～191”
- 前2字节标识网络地址，后2字节标识主机地址
- 网络地址不存在全0或全1，不存在网络总数减2的问题

B类地址



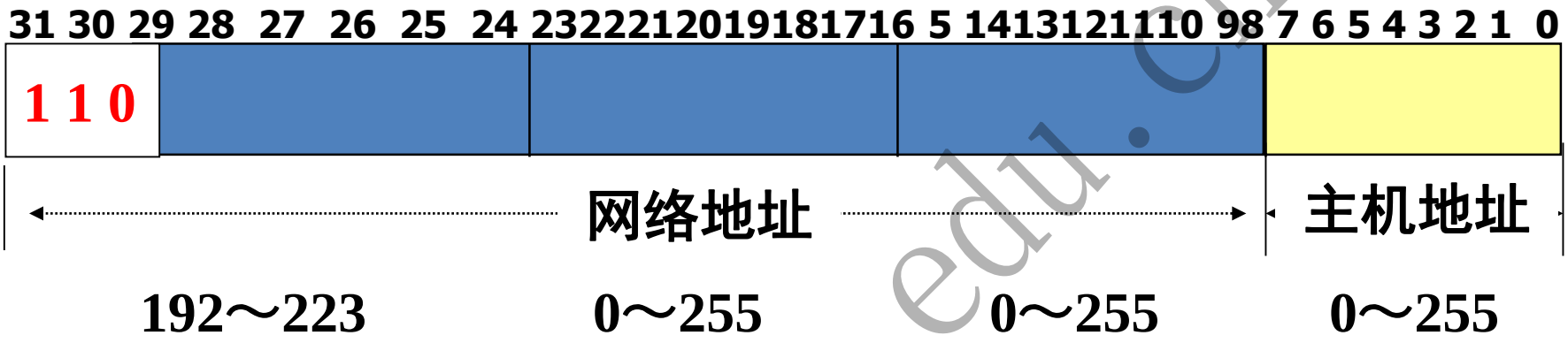
- 但是128.0.0.0（10000000.00000000.0.0）是不指派的，可以指派的B类最小网络地址是128.1.0.0（10000000.00000001.0.0）
- B类地址可指派的网络数为 2^{14} -1个
- 每个网络最多可容纳（ 2^{16} —2）台主机
  - 扣除全0和全1的主机号
- B类地址空间共约有 2^{30} 个，约占25%

C类地址



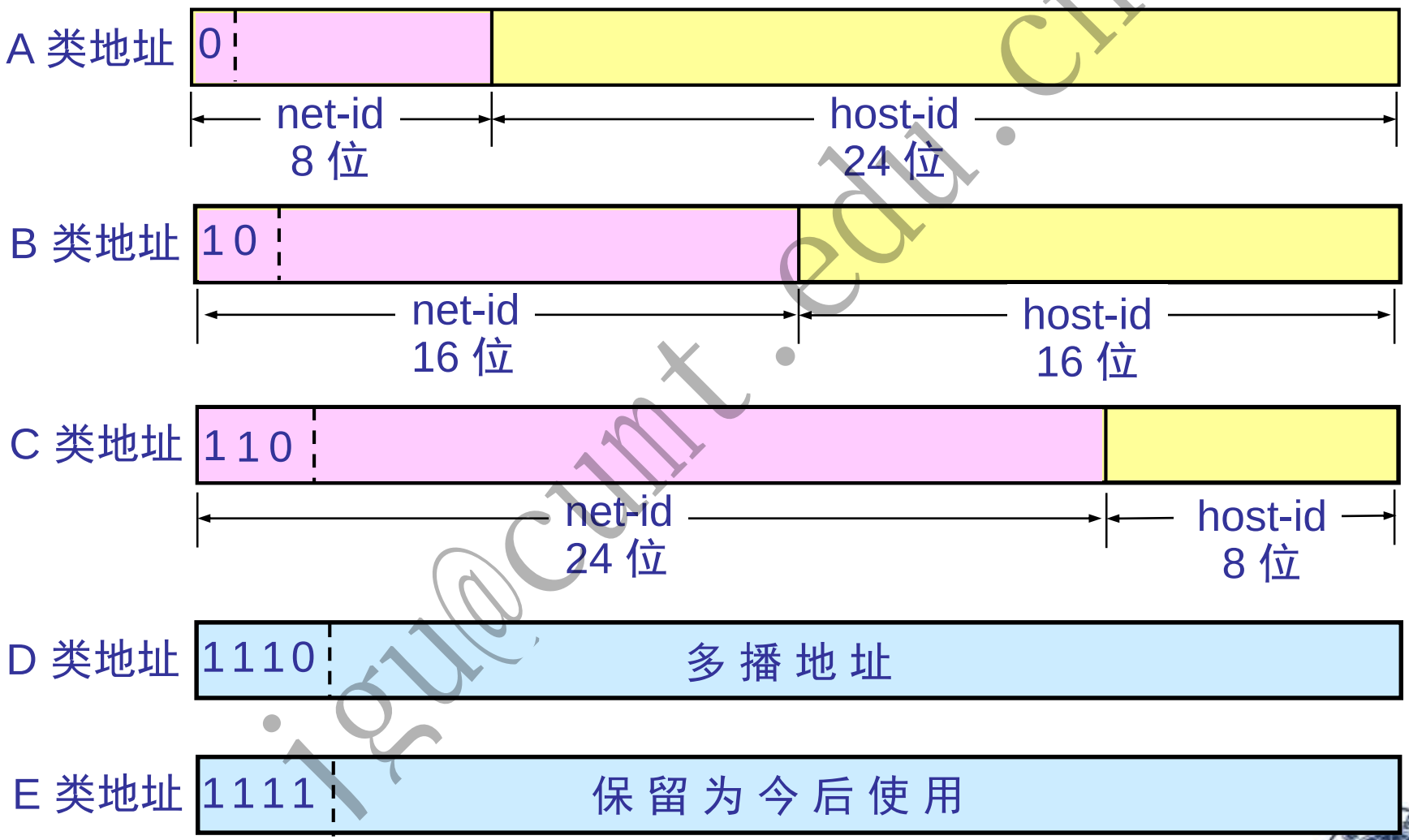
- 从高位起，前3位为“110”，第1字节用十进制表示的取值范围为“192～223”
- 前3字节标识网络地址，后1字节标识主机地址
- C类网络地址192.0.0.0（11000000.00000000.0.0）是不指派的，可以指派的C类最小网络地址是192.1.0.0（11000000.00000001.0.0）

C类地址



- 具有C类地址特征的网络总数为 2^{21} -1个
- 每个网络最多可容纳为（2^8 -2），即254台主机
- C类地址空间共约有 2^{29} 个，约占整个IP地址空间的12.5%

IP 地址中的网络号字段和主机号字段



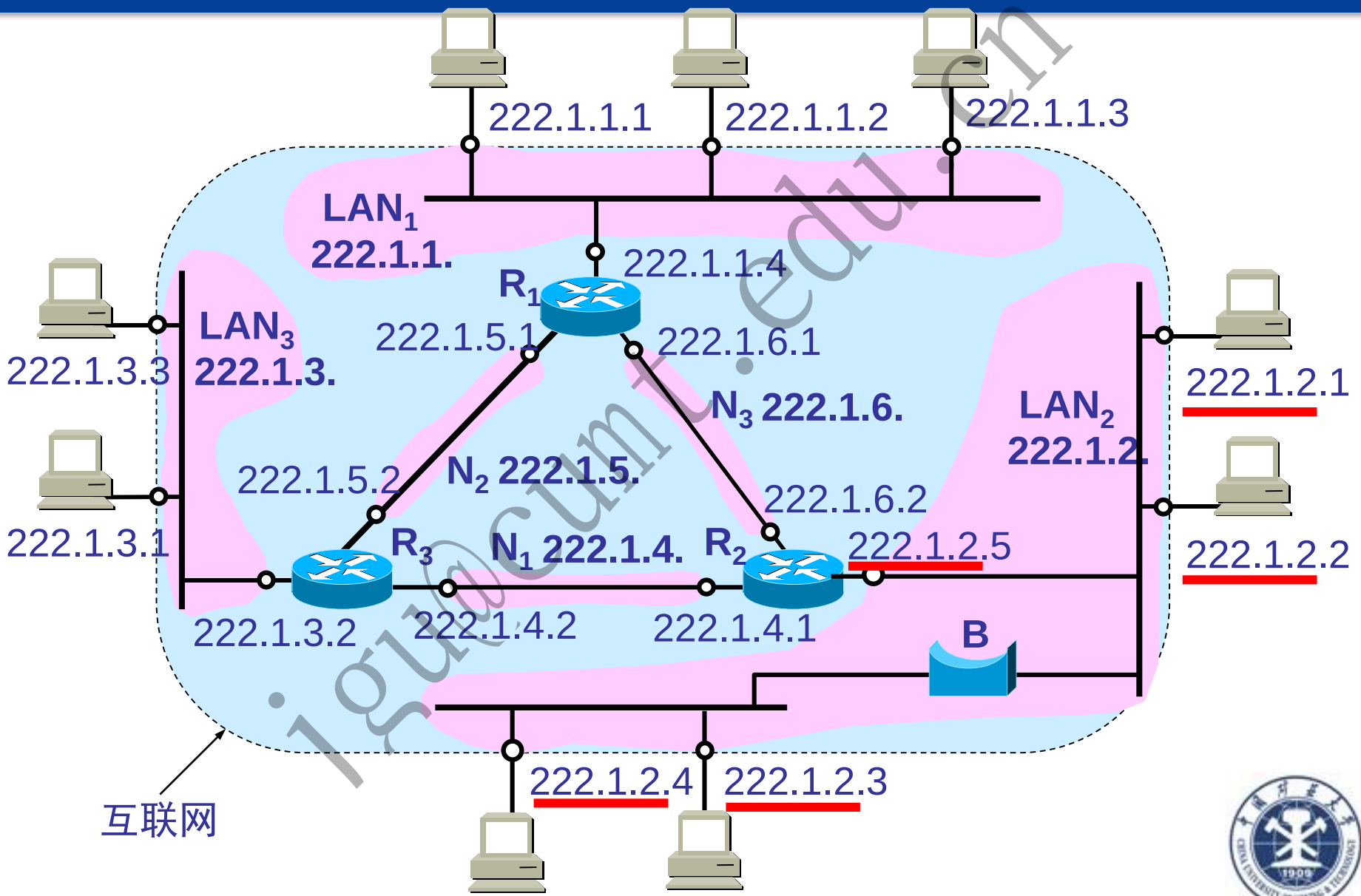
常用的三种类别的 IP 地址

IP 地址的使用范围

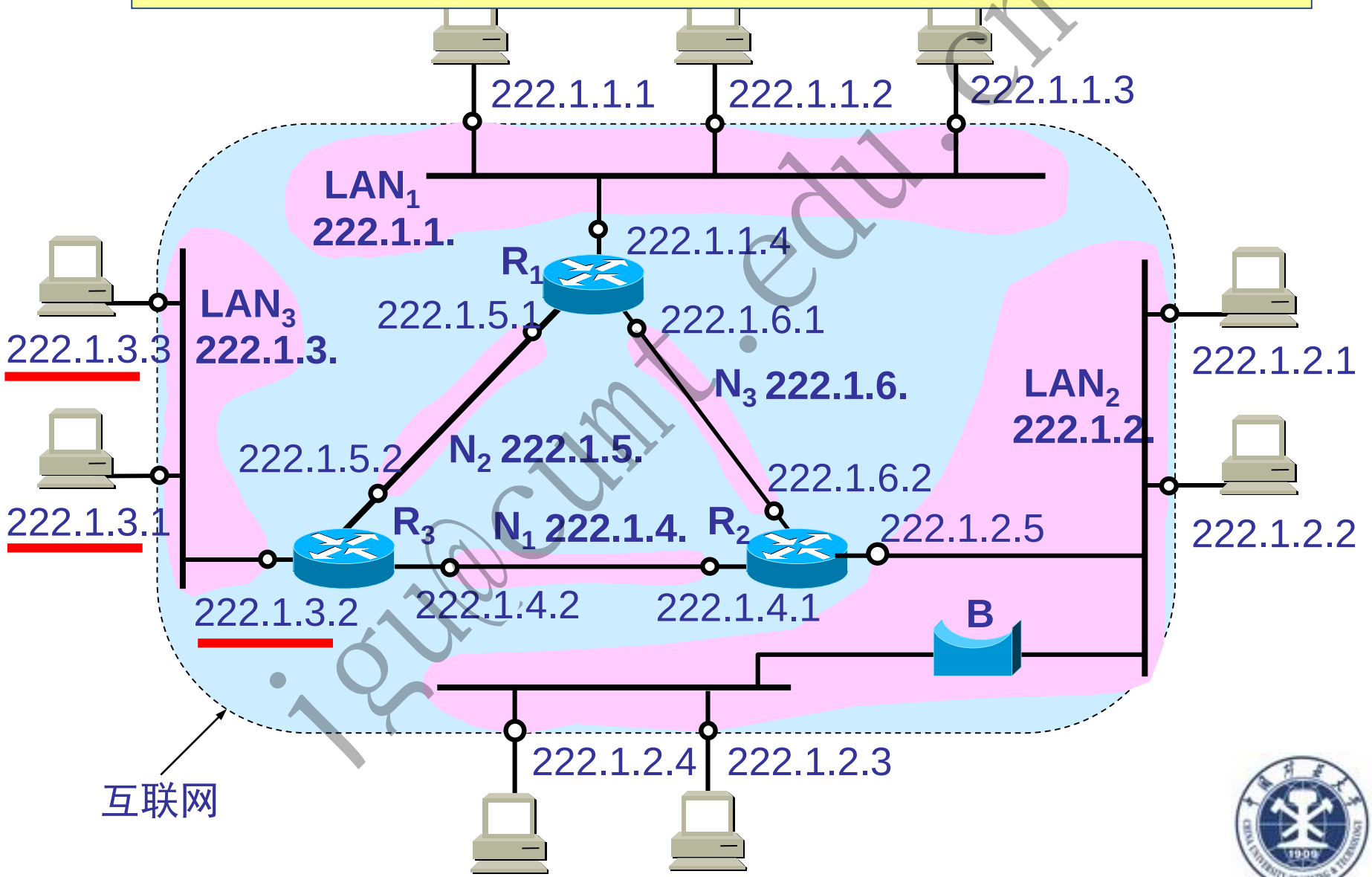
网络类别	最大网络数	第一个可用的网络号	最后一个可用的网络号	每个网络中最大的主机数
A	126 ($2^7 - 2$)	1	126	16,777,214
B	16,383 ($2^{14} - 1$)	128.1	191.255	65,534
C	2,097,151 ($2^{21} - 1$)	192.0.1	223.255.255	254

全0和全1的主机号不能分配给网络中的任何主机

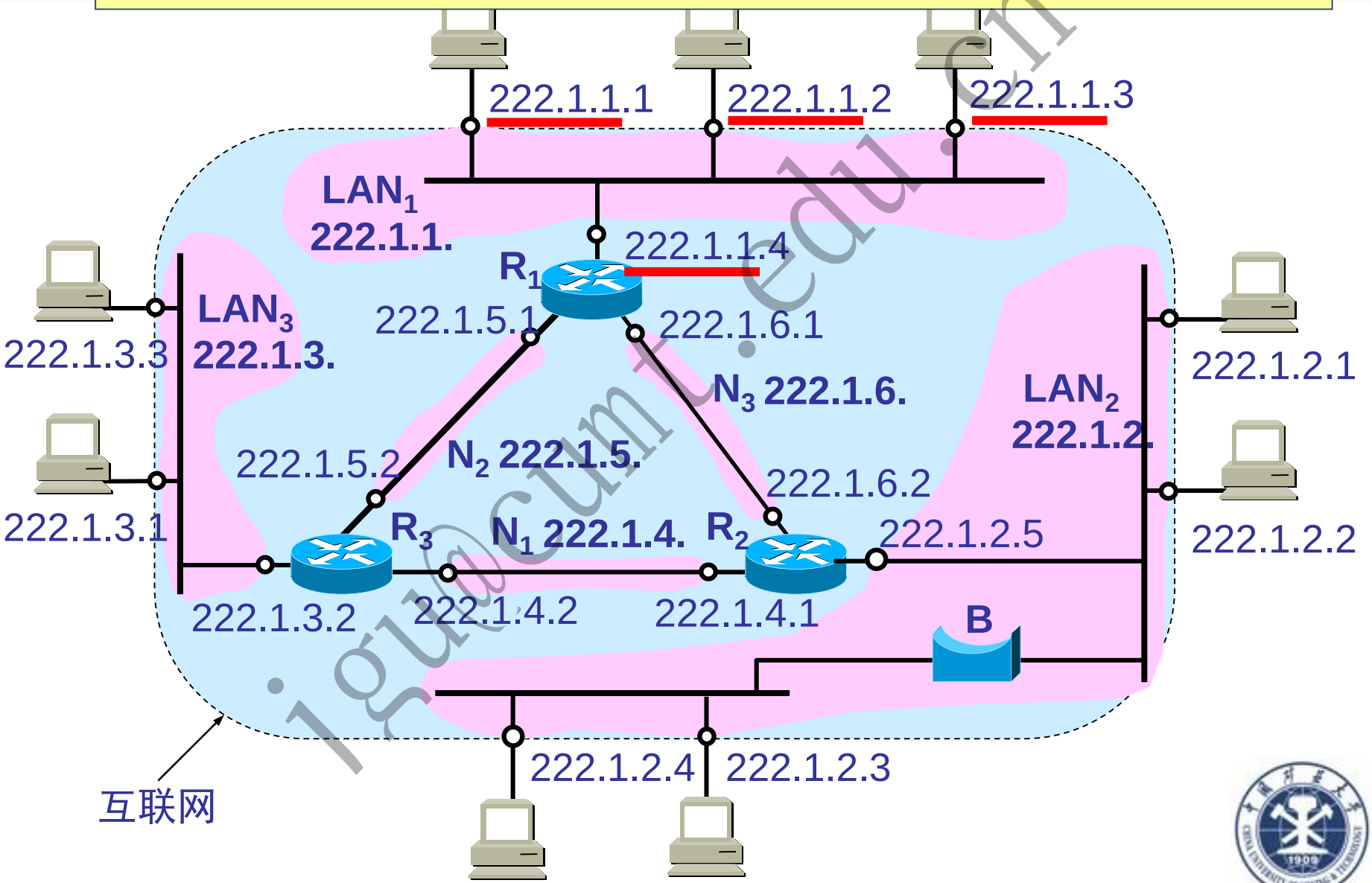
在同一个局域网上的主机或路由器的
IP 地址中的网络号必须是一样的。



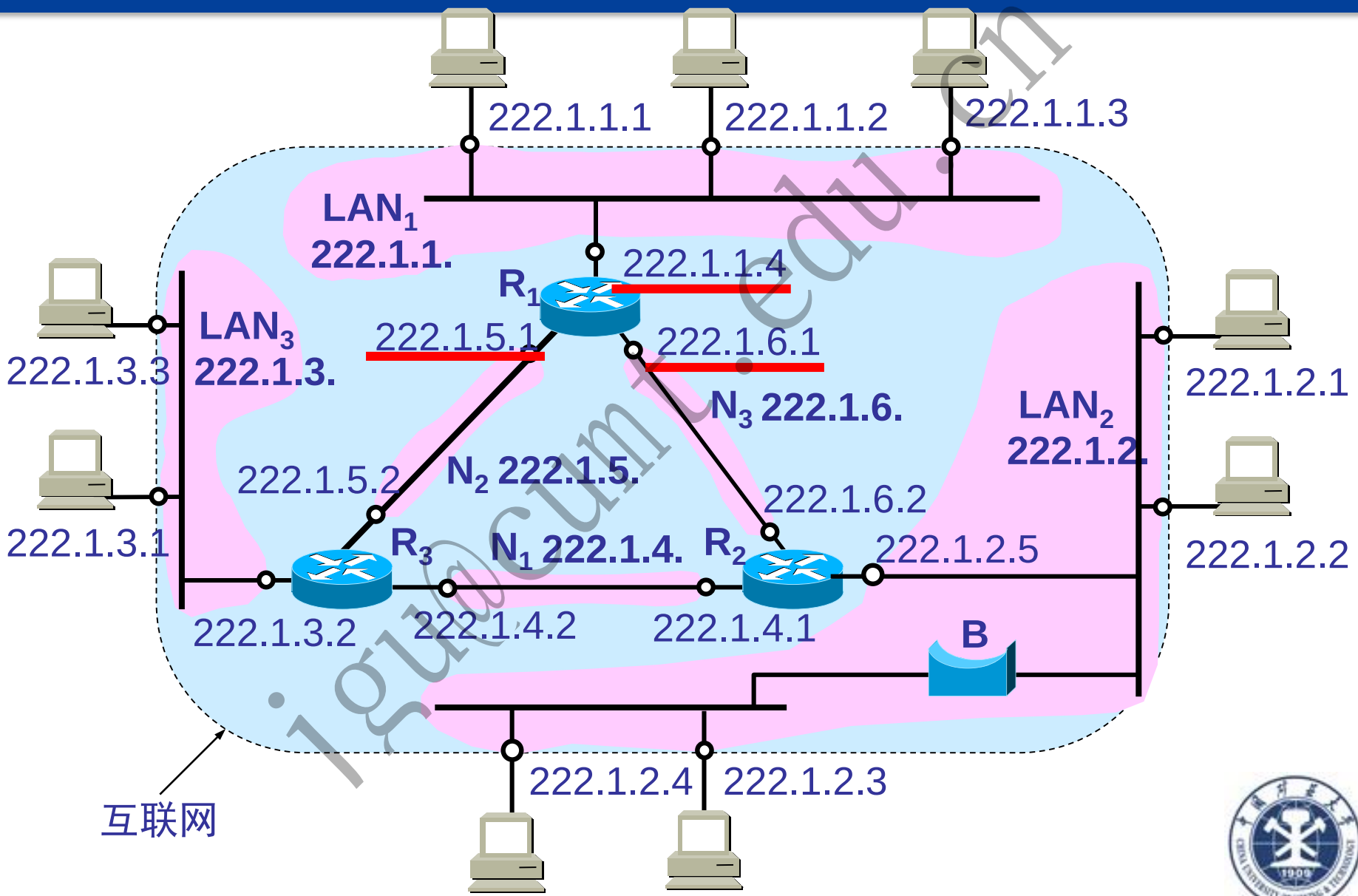
IP 地址管理机构在分配 IP 地址时只分配网络号，而剩下的主机号则由得到该网络号的单位自行分配。这样就方便了 IP 地址的管理。



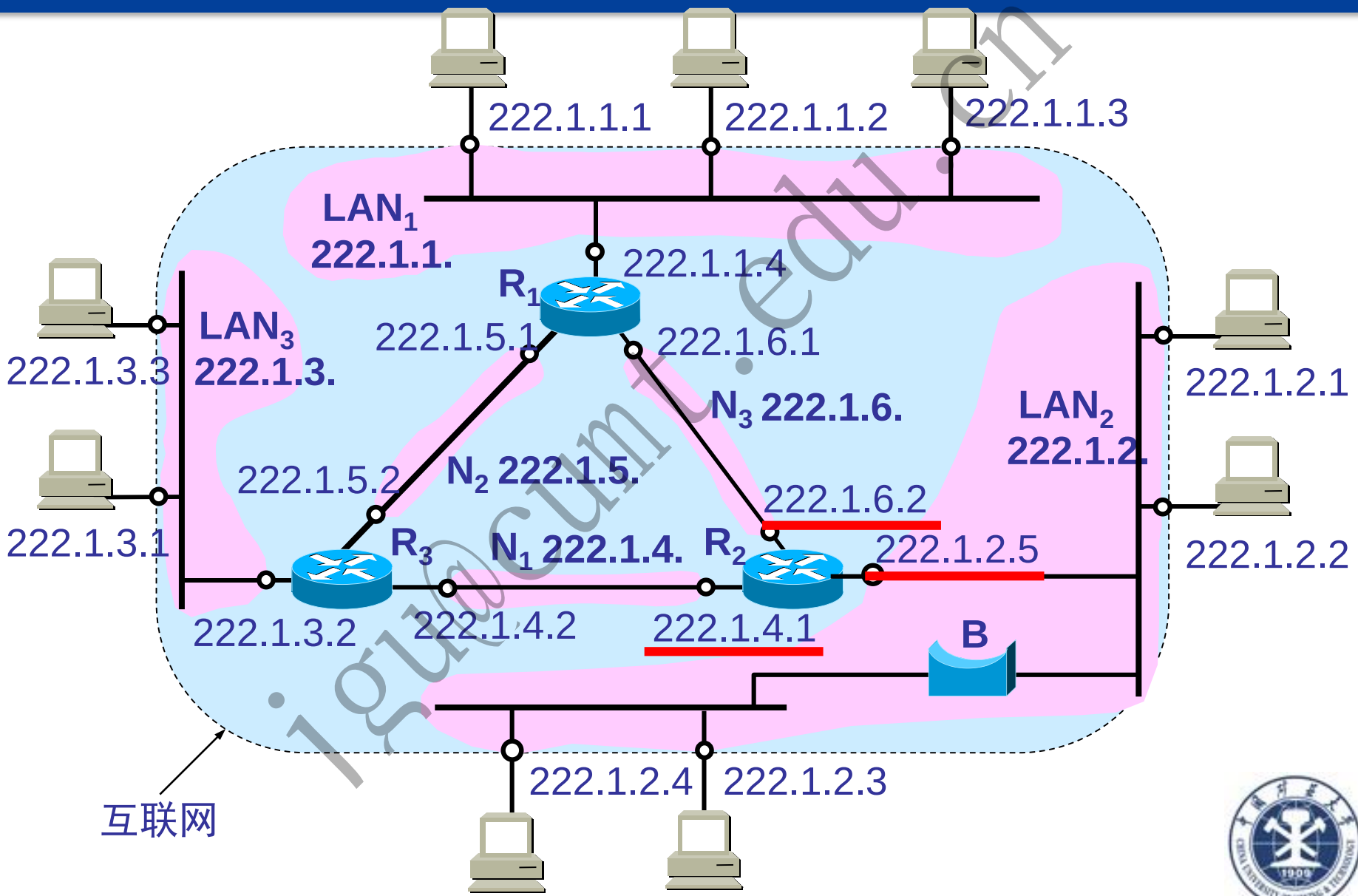
路由器仅根据目的主机所连接的网络号来转发分组（而不考虑目的主机号），这样就可以使路由表中的项目数大幅度减少，从而减小了路由表所占的存储空间。



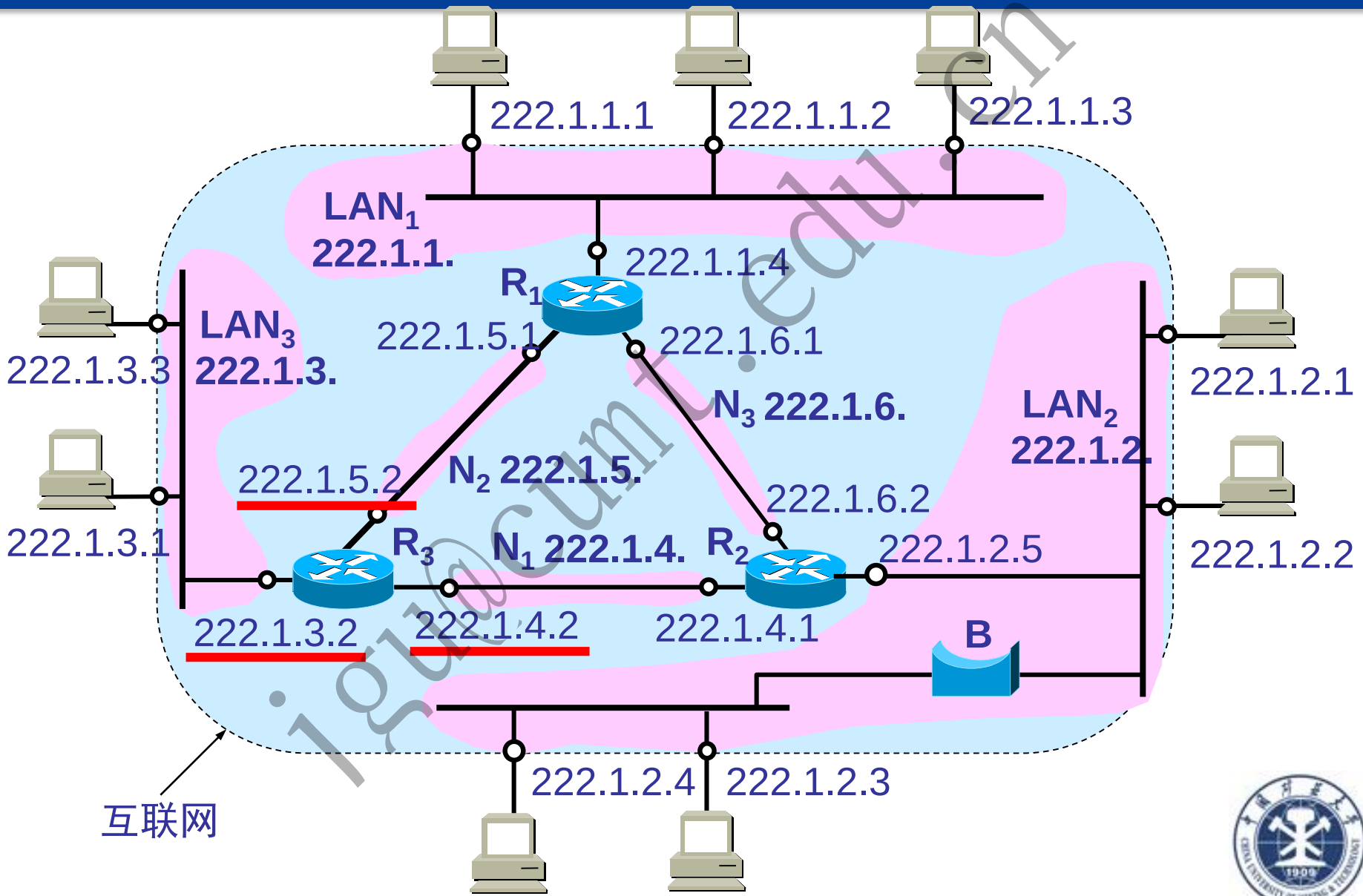
路由器总是具有两个或两个以上的 IP 地址。
路由器的每一个接口都有一个不同网络号的 IP 地址。



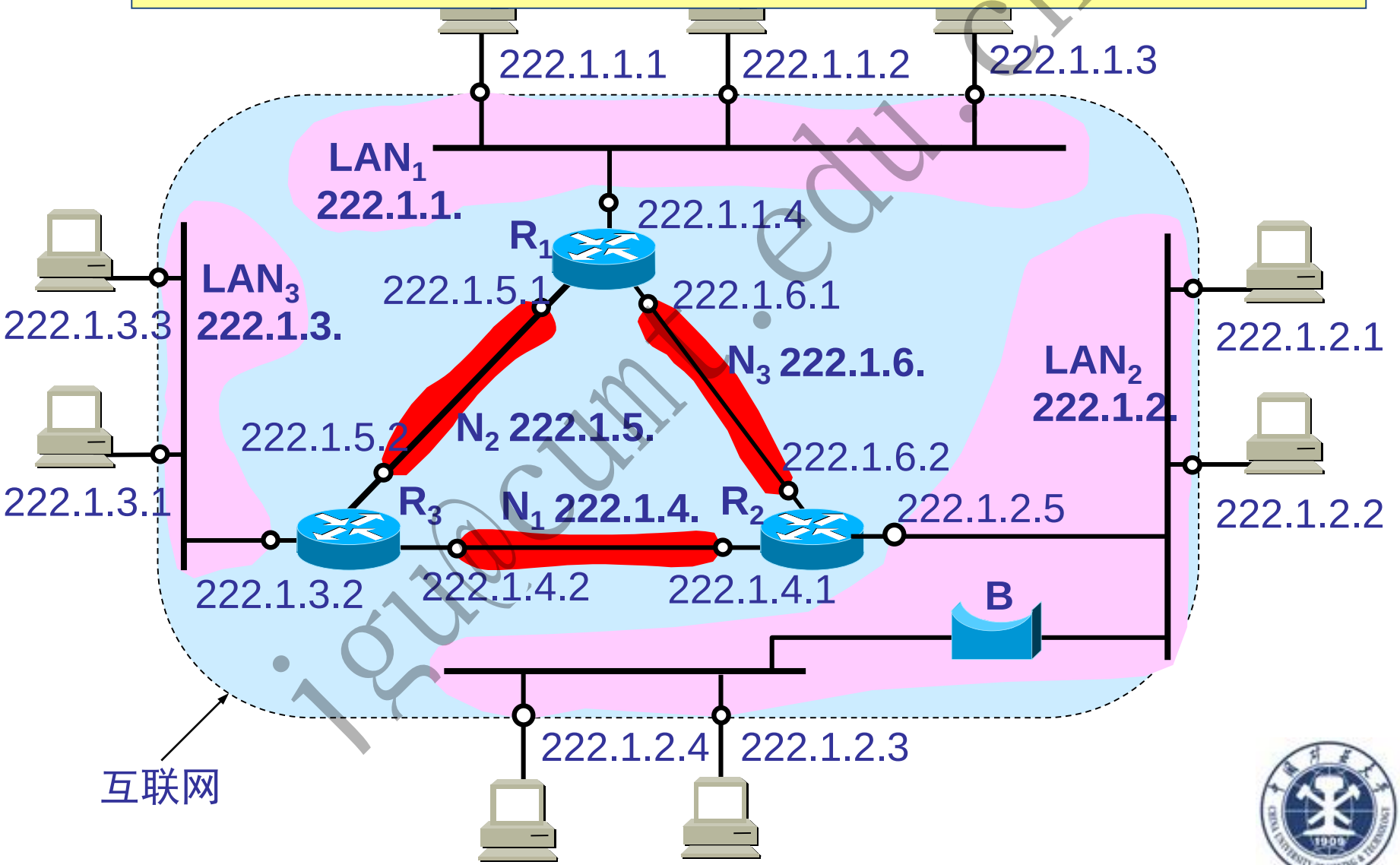
路由器总是具有两个或两个以上的 IP 地址。
路由器的每一个接口都有一个不同网络号的 IP 地址。



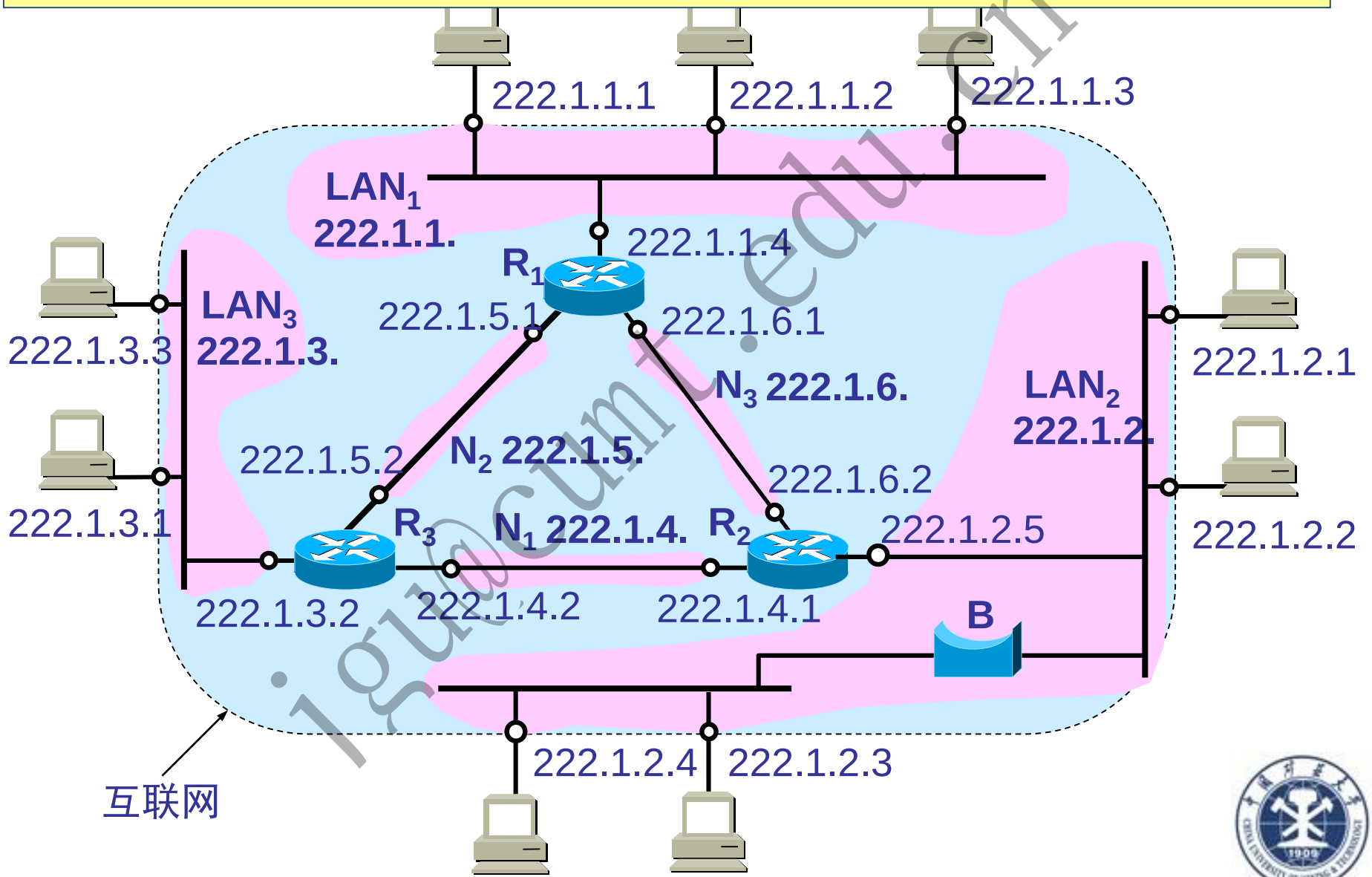
路由器总是具有两个或两个以上的 IP 地址。
路由器的每一个接口都有一个不同网络号的 IP 地址。



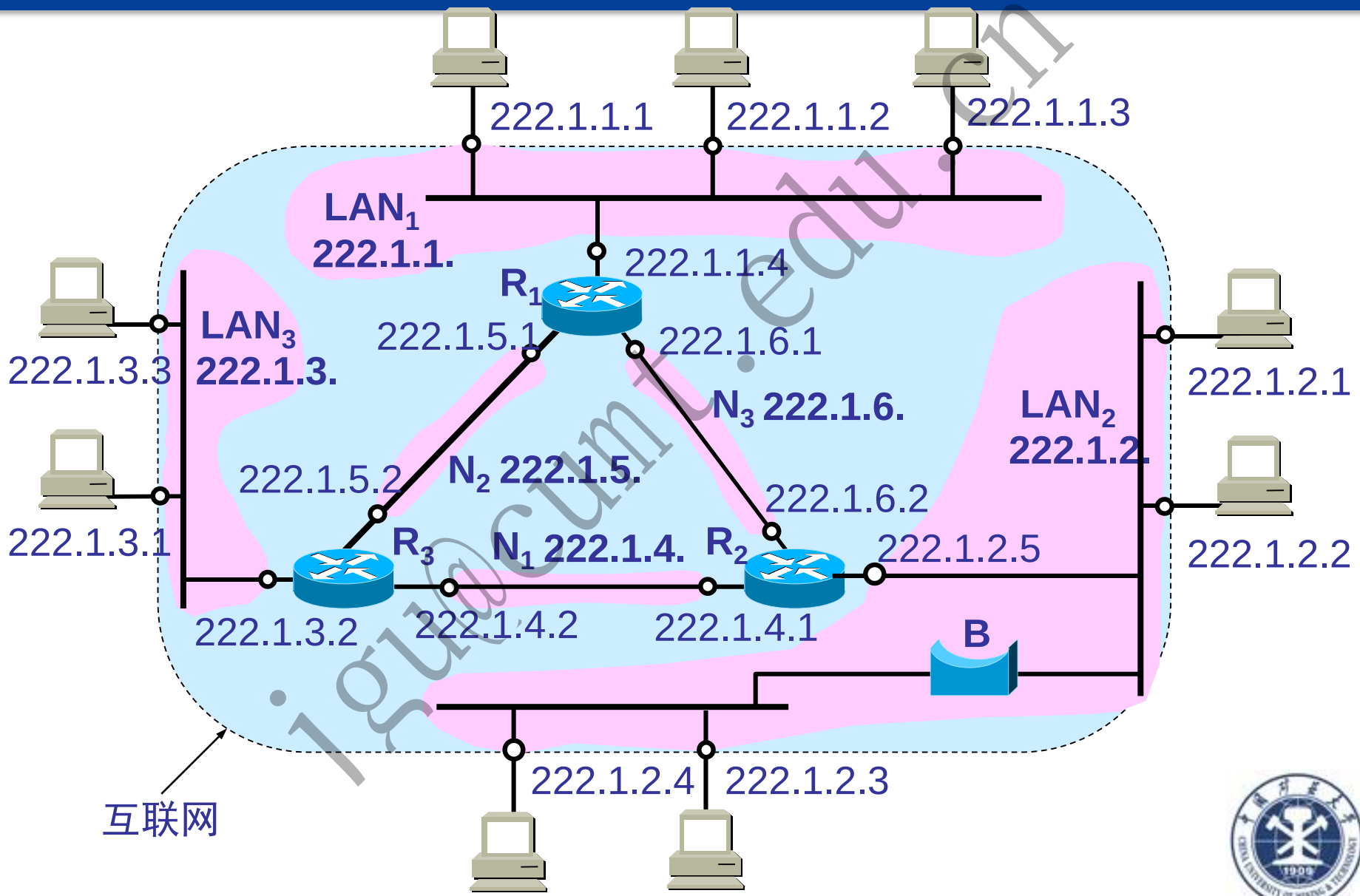
两个路由器直接相连的接口处，可指明也可不指明 IP 地址。如指明 IP 地址，则这一段连线就构成了一种只包含一段线路的特殊“网络”。现在常不指明 IP 地址。



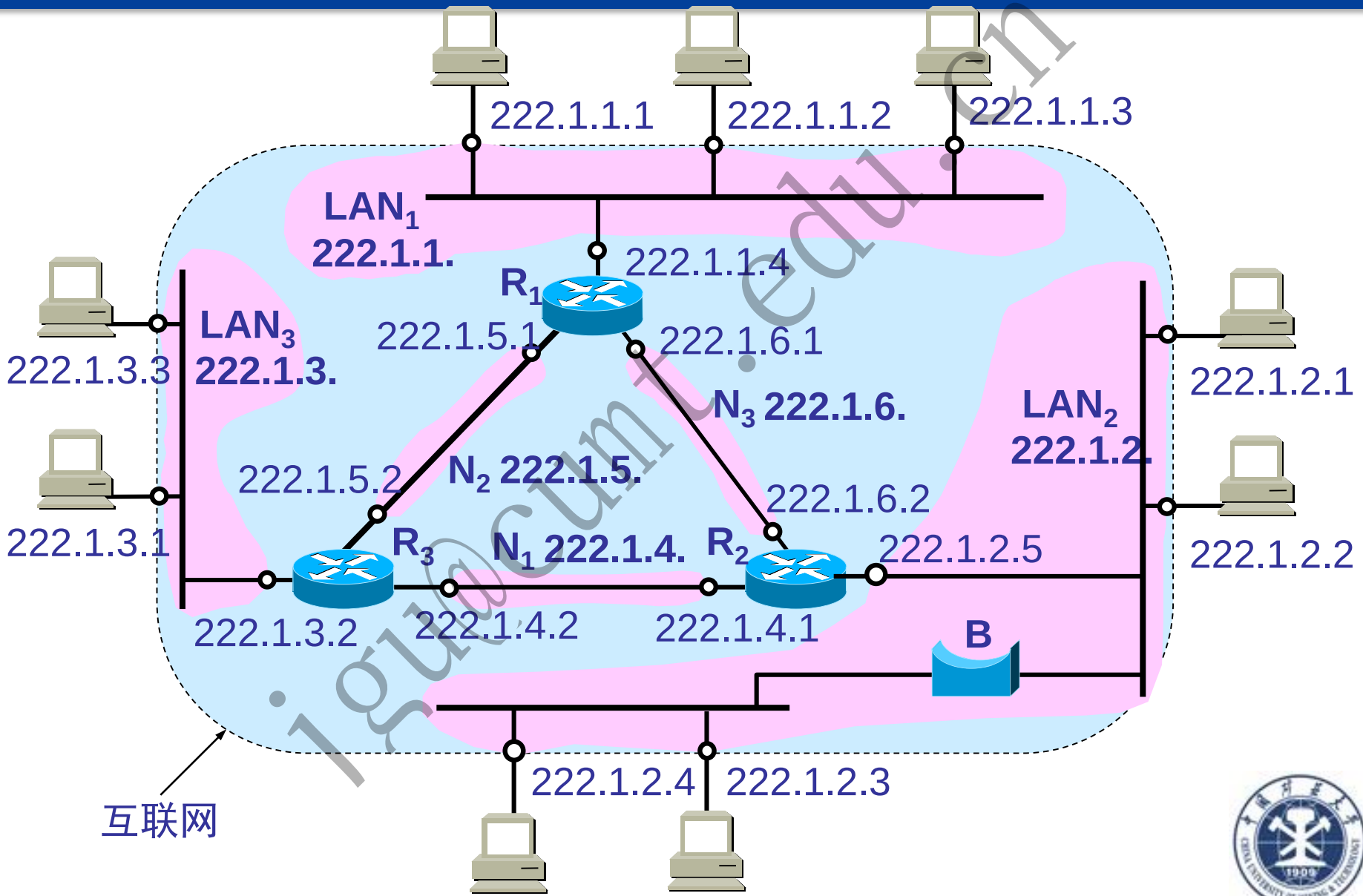
当一个主机同时连接到两个网络上时，该主机就必须同时具有两个相应的 IP 地址，其网络号 net-id 必须是不同的。这种主机称为多归属主机 (multihomed host)。



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，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。



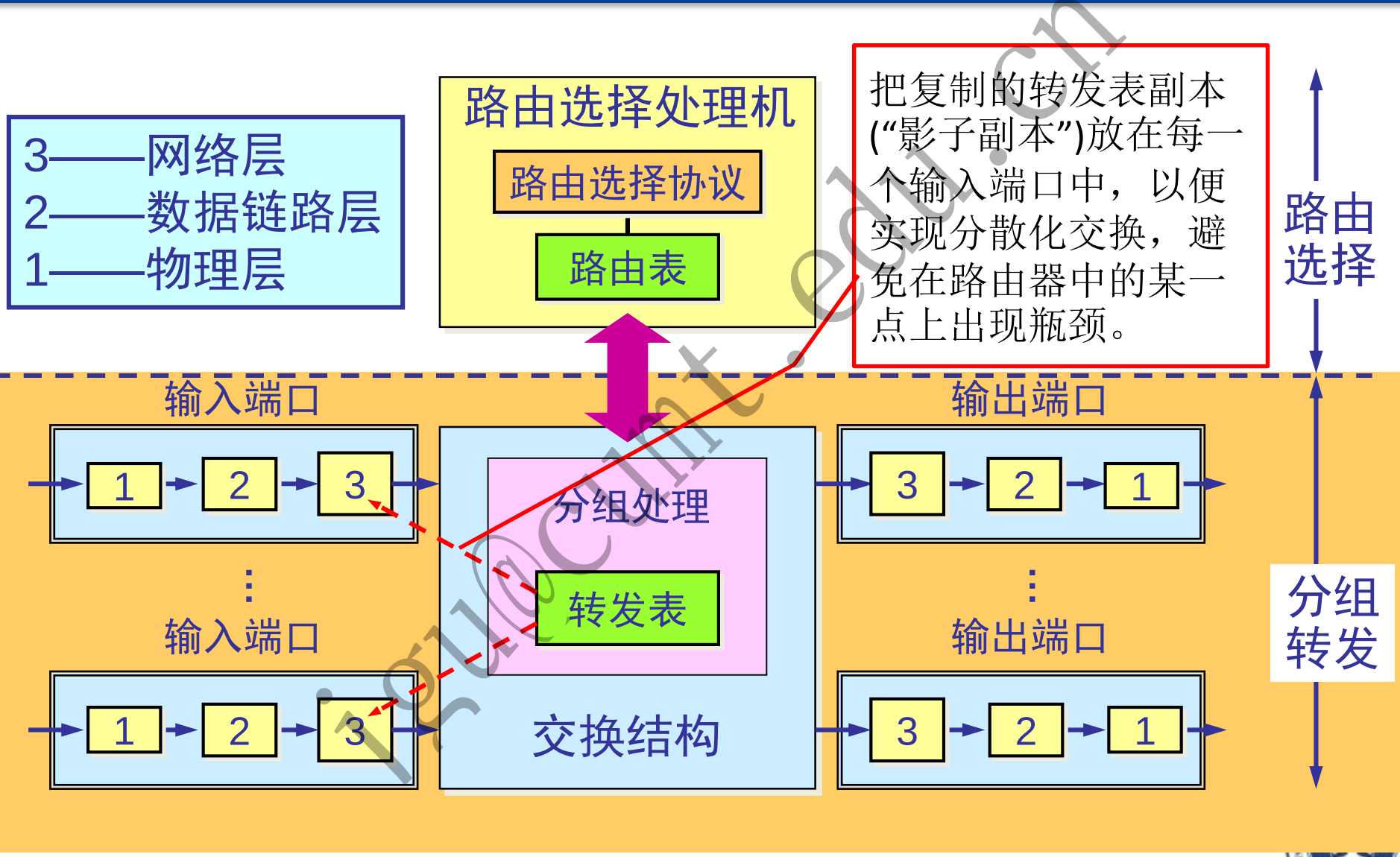
一般不使用的特殊的 IP 地址

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	用作本地软件环回测试之用

Q4：路由器是如何工作的？

- 路由器是一种具有多个输入端口和多个输出端口的专用计算机，其任务是转发分组。也就是说，将路由器某个输入端口收到的分组，按照分组要去的目的地（即目的网络），把该分组从路由器的某个合适的输出端口转发给下一跳路由器。
- 下一跳路由器也按照这种方法处理分组，直到该分组到达终点为止。

典型的路由器的结构



“转发”和“路由选择”的区别

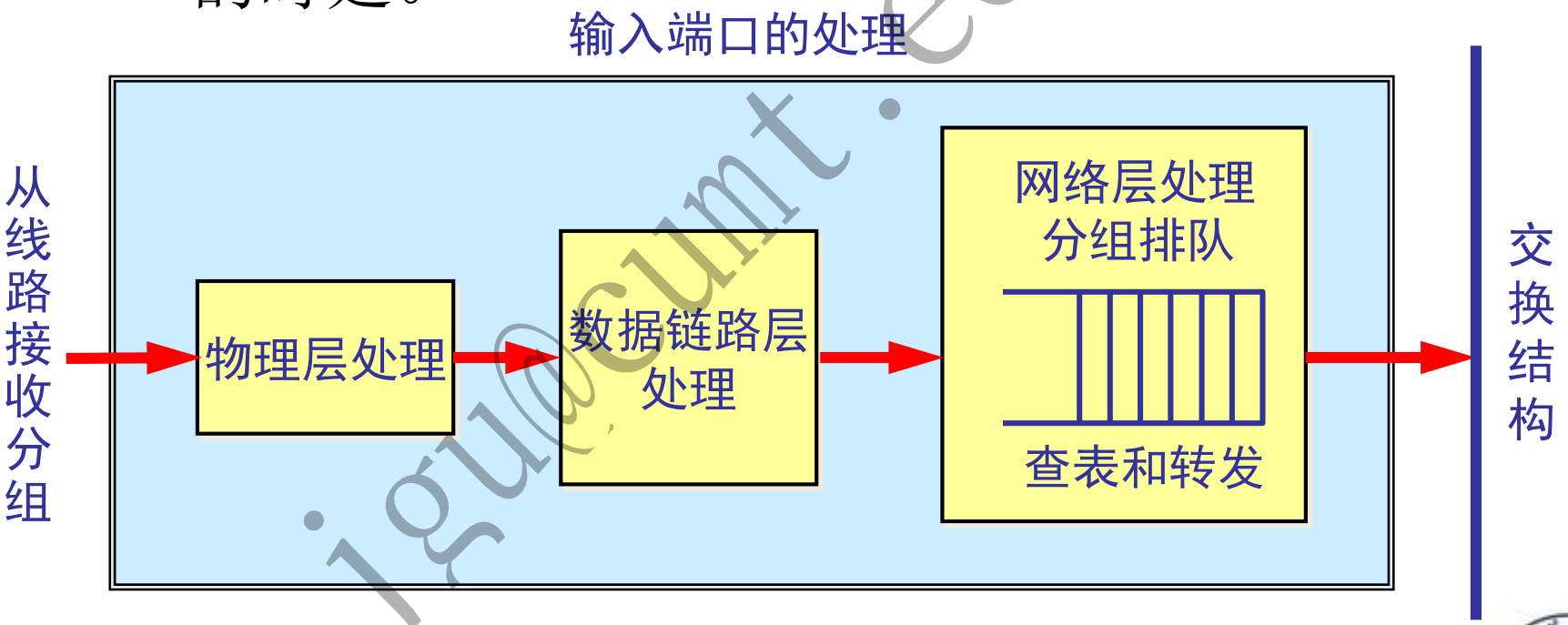
- “转发”(forwarding)就是路由器根据转发表将用户的 IP 数据报从合适的端口转发出去。
- “路由选择”(routing)则是按照分布式算法，根据从各相邻路由器得到的关于网络拓扑的变化情况，动态地改变所选择的路由。
- 路由表是根据路由选择算法得出的。而转发表是从路由表得出的。
- 在讨论路由选择的原理时，往往不去区分转发表和路由表的区别，

线速

- 路由器必须以很高的速率转发分组。
- 最理想的情况是输入端口的处理速率能够跟上线路把分组传送到路由器的速率。这种速率称为线速(line speed 或 wire speed)。
  - 设线路是OC-48链路，即2.5Gb/s。若分组长度为256字节，那么线速就应当达到每秒能够处理100万以上的分组。
  - 现在常用Mpps(百万分组每秒)为单位来说明一个路由器对收到的分组的处理速率有多高。

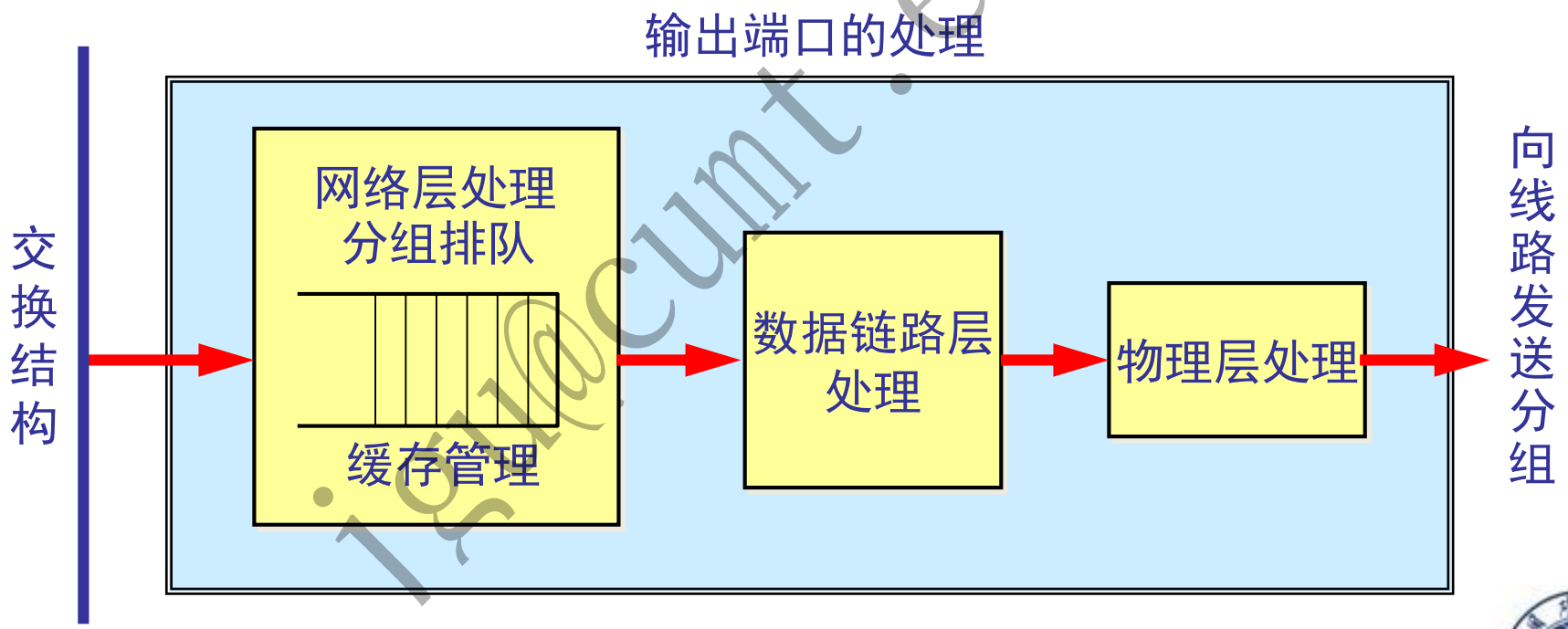
输入端口对线路上收到的分组的处理

- 数据链路层剥去帧首部和尾部后，将分组送到网络层的队列中排队等待处理。这会产生一定的时延。



输出端口将交换结构传送来的分组发送到线路

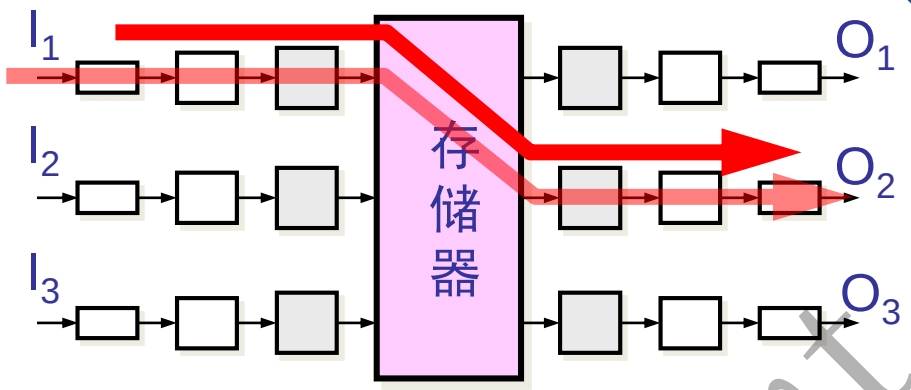
- 当交换结构传送过来的分组先进行缓存。数据链路层处理模块将分组加上链路层的首部和尾部，交给物理层后发送到外部线路。



分组丢弃

- 若路由器处理分组的速率赶不上分组进入队列的速率，则队列的存储空间最终必定减少到零，这就使后面再进入队列的分组由于没有存储空间而只能被丢弃。
- 路由器中的输入或输出队列产生溢出是造成分组丢失的重要原因。

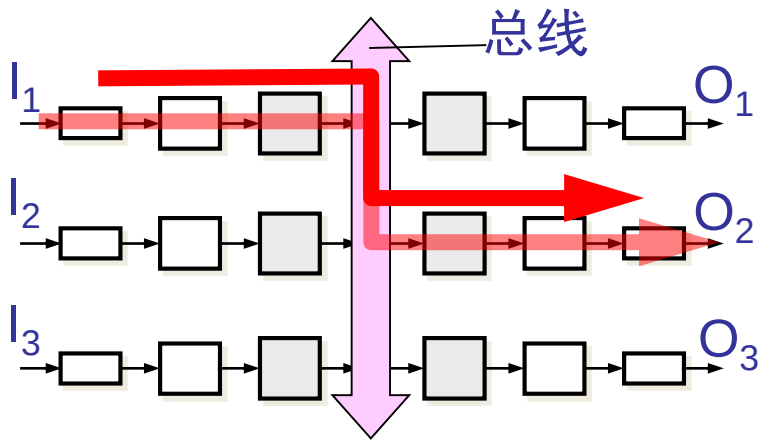
路由器的交换结构



(a) 通过存储器

- 当路由器的某个输入端口收到一个分组时，就用中断方式通知路由选择处理机。然后分组就从输入端口复制到存储器中。路由器处理机从分组首部提取目的地址，查找路由表，再将分组复制到合适的输出端口的缓存中。
- 若存储器的带宽(读或写)为每秒M个分组，那么路由器的交换速率(即分组从输入端口传送到输出端口的速率)一定小于M/2。这是因为存储器对分组的读和写需要花费的时间是同一个数量级。

路由器的交换结构

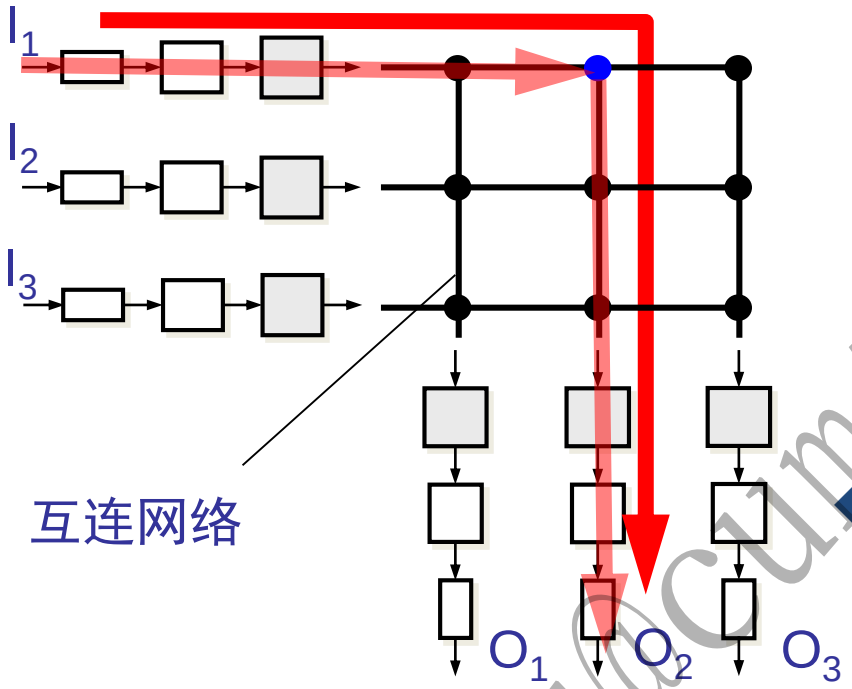


(b) 通过总线

- 数据报从输入端口通过共享的总线直接传送到合适的输出端口，而不需要路由选择处理机的干预，但是同一时间只能有一个分组在共享式总线上传送。

- 当分组到达输入端口时，若发现总线忙，则被阻塞而不能通过交换结构，并在输入端口排队等待。
- 因为每一个要转发的分组都要通过这一条总线，因此路由器的转发带宽就受总线速率的限制。
- 现代技术可以将总线的带宽提高到每秒吉比特的速率，因此许多的路由器产品都采用这种总线的交换方式。

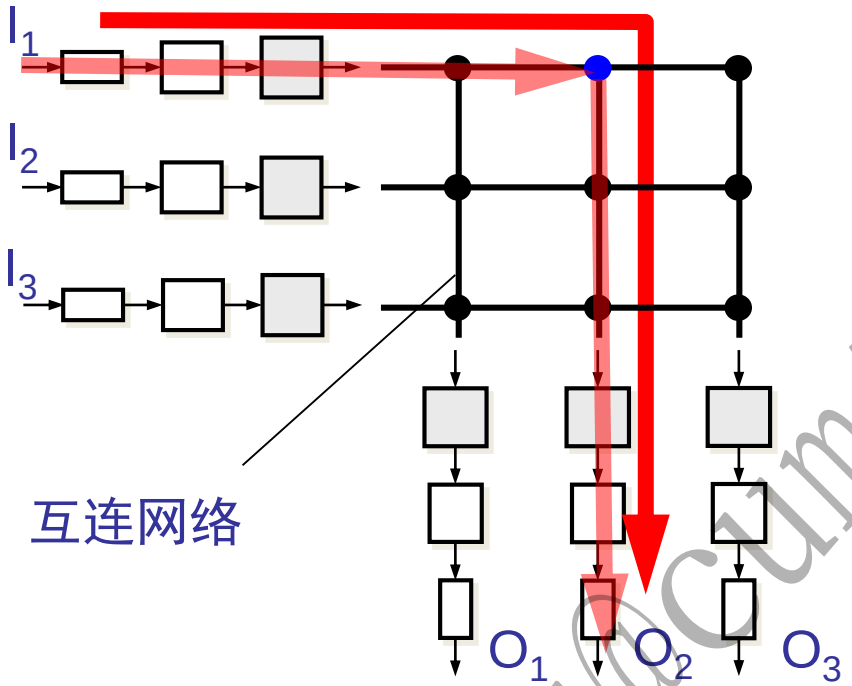
路由器的交换结构



(c) 通过纵横交换结构

- 纵横交换结构(crossbar switch fabric)的互连网络(interconnection network)有2N条总线，可以使N个输入端口和N个输出端口相连接，这取决于相应的交叉点是使水平总线和垂直总线接通还是断开。
- 当输入端口收到一个分组时，就将它发送到与该输入端口相连的水平总线上。若通向所要转发的输出端口的垂直总线是空闲的，则在这个结点将垂直总线与水平总线接通，然后将该分组转发到这个输出端口。

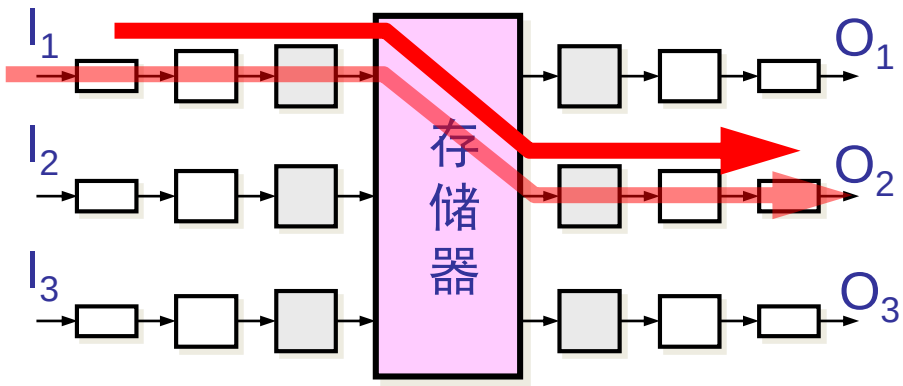
路由器的交换结构



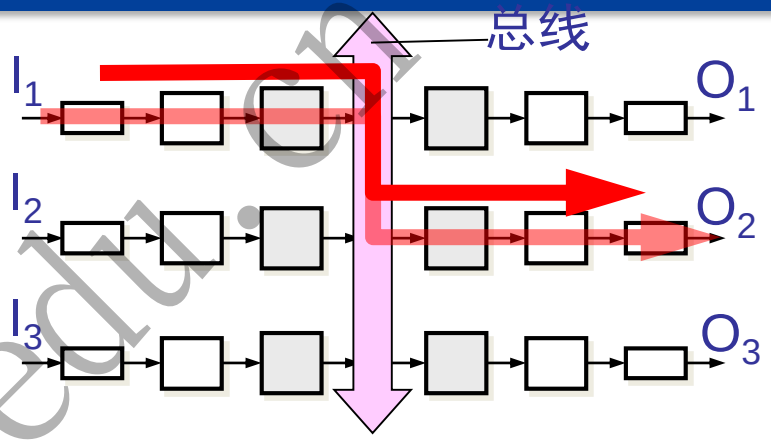
(c) 通过纵横交换结构

- 但若该垂直总线已被占用（有另一个分组正在转发到同一个输出端口），则后到达的分组就被阻塞，必须在输入端口排队。
- 矩阵交换的最大优点是允许多个相互不冲突的交换同时进行，并支持点对多点（Multicast）的交换。

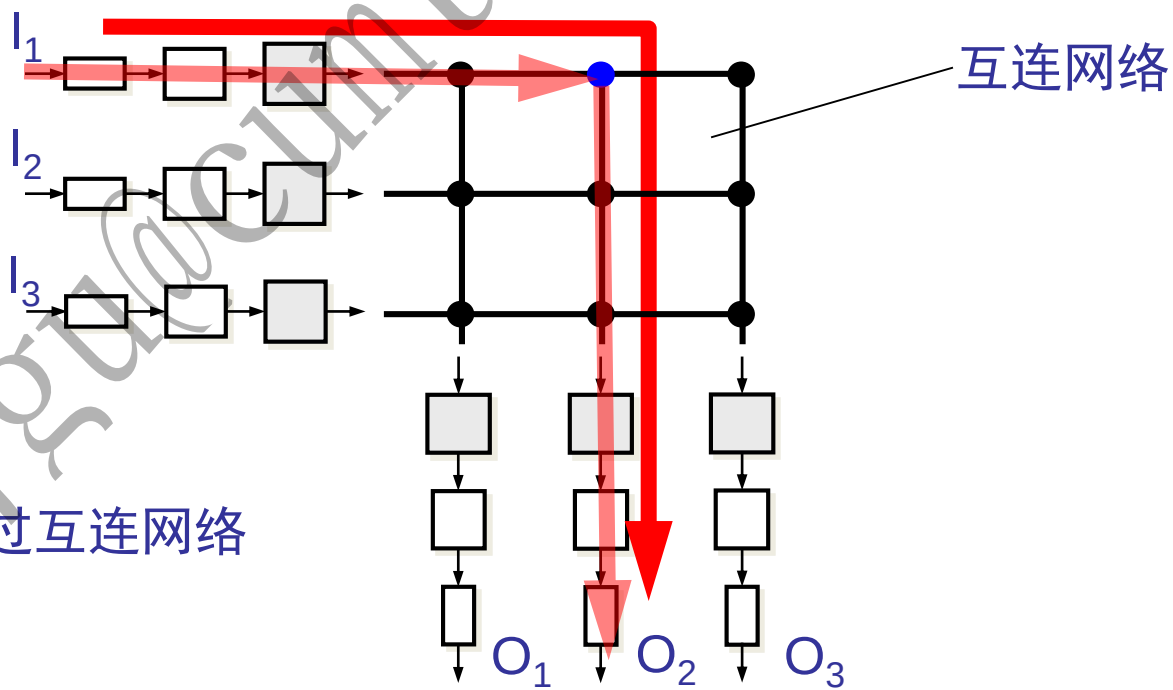
路由器的交换结构



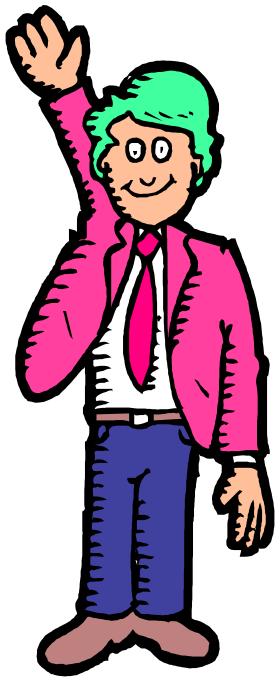
(a) 通过存储器



(b) 通过总线



(c) 通过互连网络



THANK YOU!