计算机网络应用技术

第四章：网络层

信息技术学院互联网技术教学团队

阮晓龙 许成刚 高海波 李瑞昌

http://xslt.it.hactcm.edu.cn

2022.9

扫码访问课程学习平台

本章教学计划

- 网络层提供的两种服务
- 网际协议（IP）
- 划分子网和构建超网（IP地址管理）
- 网际控制报文协议（ICMP）
- 路由选择协议（RIP、OSPF、BGP）

基础内容

- IPv6
- 虚拟专用网（VPN）
- 网络地址转换（NAT）

扩展部分

本章教学计划

- 讨论多个网络通过路由器互连成为一个互联网的问题。
- 重点内容：
  - 虚拟互连网络
  - IP地址与物理地址的关系
  - IP地址分类与管理的方法
  - 路由选择协议
  - VPN的基本概念
  - NAT的基本原理

1.网络层提供的两种服务

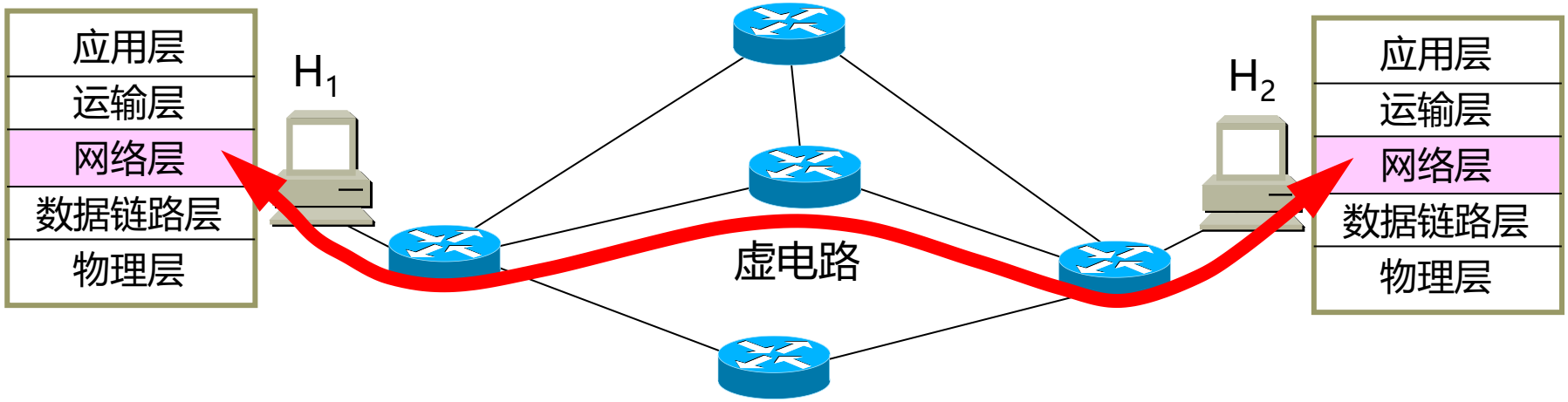
- 在计算机网络领域，网络层应该向运输层提供怎样的服务（“面向连接”还是“无连接”）曾引起了长期的争论。
- 争论焦点的实质就是：
  - 在计算机通信中，可靠交付应当由谁来负责？
  - 是网络还是端系统？
  - 通俗的讲：
    - 是网络设备负责可靠通信？
    - 是计算机负责可靠通信？

1.网络层提供的两种服务

- 电信网的成功经验：让网络负责可靠交付。
  - 电信网使用昂贵的程控交换机，用面向连接的通信方式，使电信网络能够向用户提供可靠传输的服务。
  - 虚电路就是当两个计算机进行通信时，应当先建立连接，以保证双方通信所需的一切网络资源，双方就沿着已建立的虚电路发送分组。
  - 如果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。

1.网络层提供的两种服务

1.1虚电路

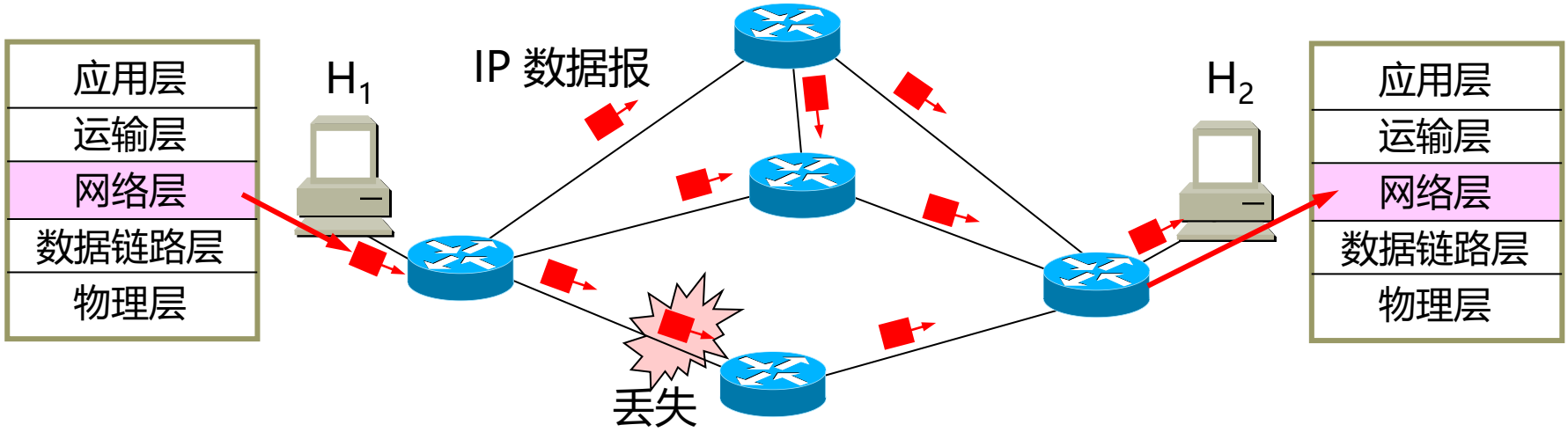


H_1发送给H_2的所有分组都沿着同一条虚电路传送

1.网络层提供的两种服务

- 因特网不提供端到端的可靠服务的优势：
  - 网络中的路由器可以做得比较简单，而且价格低廉。
  - 如果主机（端系统）中的进程之间的通信需要是可靠的，那么就由运输层负责（包括差错处理、流量控制等）。
  - 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。

1.网络层提供的两种服务



H_1发送给H_2的分组可能沿着不同路径传送

1.网络层提供的两种服务

1.3虚电路服务与数据报服务对比

对比的方面	虚电路服务	数据报服务
设计思路	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

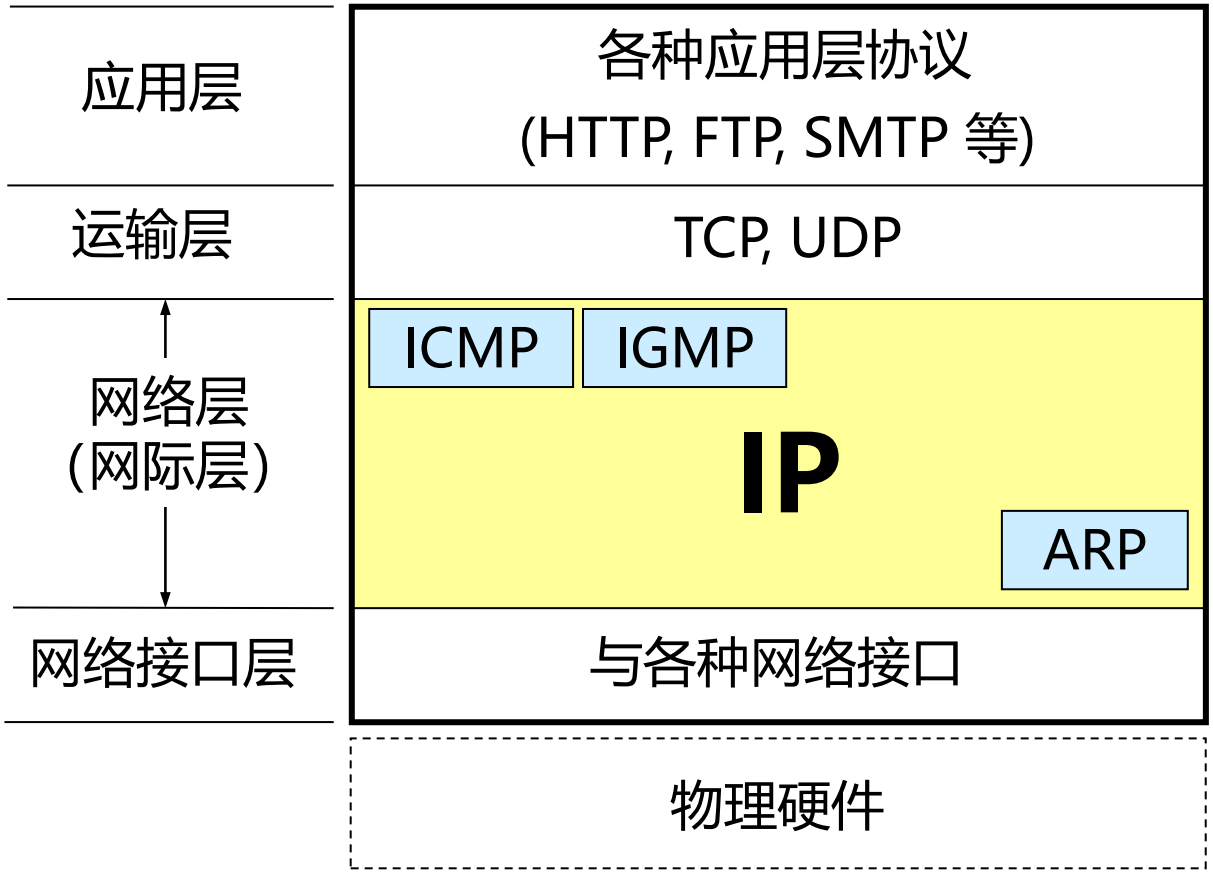
2.网际协议（IP）

- 网际协议（Internet Protocol，IP），或称互联网协议，是用于报文交换网络的一种面向数据的协议。
- IP是在TCP/IP协议中网络层的主要协议，任务是仅仅根据源主机和目的主机的地址传送数据。
  - IP定义了寻址方法和数据报的封装结构 。
  - IP第一个架构的主要版本，现在称为IPv4，目前仍然是最主要的互联网协议。
  - 尽管世界各地正在积极部署IPv6，但IPv4仍然是最重要的网际协议。

2.网际协议（IP）

- 网际协议IP是TCP/IP体系中两个最主要的协议之一。
- 与IP协议配套使用的三个协议：
  - 地址解析协议 ARP
    (Address Resolution Protocol)
  - 网际控制报文协议 ICMP
    (Internet Control Message Protocol)
  - 网际组管理协议 IGMP
    (Internet Group Management Protocol)

2.网际协议（IP）


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
ARP
网络接口层
与各种网络接口
物理硬件

2.网际协议（IP）

- 互连在一起的网络要进行通信，会遇到许多问题需要解决，例如：
  - 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

2.网际协议（IP）

- 因为用户的需求是多种多样的，所以没有一种单一的网络能够适应所有用户的需求。
- 网络技术是不断发展的，网络的制造厂家也要不停止的推出新产品，以获得更大的市场份额和持续利润。
- 市场上有不同性能、不同网络协议的网络，分布在不同的位置，由不同的组织和人员来管理。

2.网际协议（IP）

- 从一般概念上来讲，将网络互相连接起来要使用一些中间设备。
- 根据中间设备所在的层次，有五种不同的中间设备：
  - 转发器（repeater）：物理层中继系统。
  - 网桥（桥接器，bridge）：数据链路层中继系统。
  - 路由器（router）：网络层中继系统。
  - 桥路器（brouter）：数据链路层和网络层混合中继系统。
  - 网关（gateway）：网络层以上使用的中继系统。

2.网际协议（IP）

- 市场上主要的网络中间设备都有比较典型的产品名称。

中间设备	工作层次	主要产品
转发器（repeater）	物理层	集线器
网桥（桥接器，bridge）	数据链路层	交换机，二层交换机
路由器（router）	网络层	路由器
桥路器（brouter）	数据链路层和网络层	路由交换机，三层交换机
网关（gateway）	网络层以上	网关、七层交换机

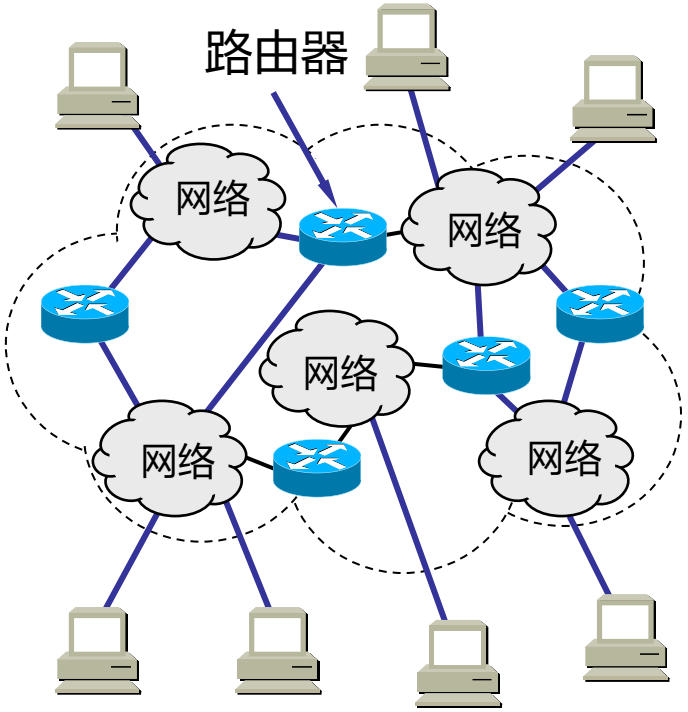
2.网际协议（IP）

- 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。
- 我们讨论网络互联是指用路由器进行网络互联和路由选择。
  - 路由器就是一台专用计算机，用来在互联网中进行路由选择。
  - 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。

2.网际协议（IP）

- TCP/IP体系在网络互连上采用的做法是在网络层使用标准化协议，但相互连接的网络则可以是异构的。
- 参加互连的计算机网络都采用相同的网际协议（IP），因此可以把互连以后的计算机网络看成为一个虚拟互连网络（internet）。

2.网际协议（IP）



(a) 互连网络

(b) 虚拟互连网络

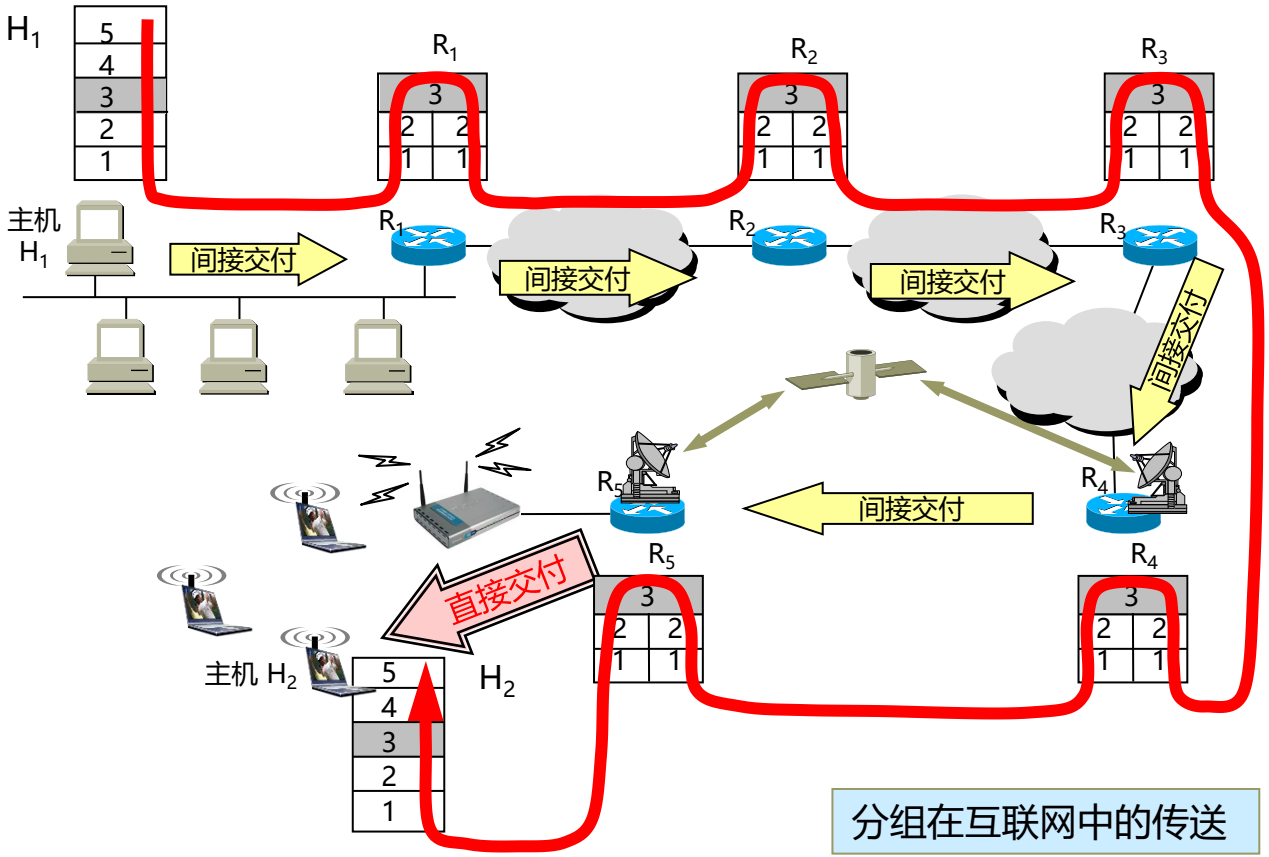
2.网际协议（IP）

- 所谓虚拟互连网络，就是逻辑互连网络，它的意思是：
  - 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。
- 使用虚拟互连网络概念的好处是：
  - 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。

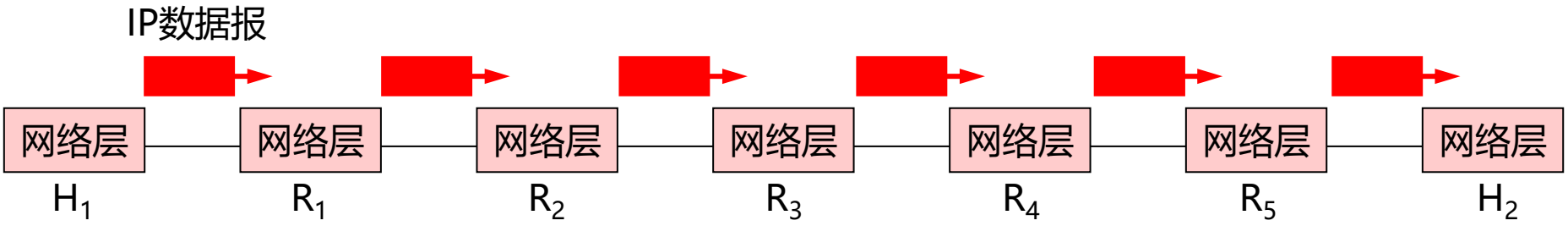
2.网际协议（IP）

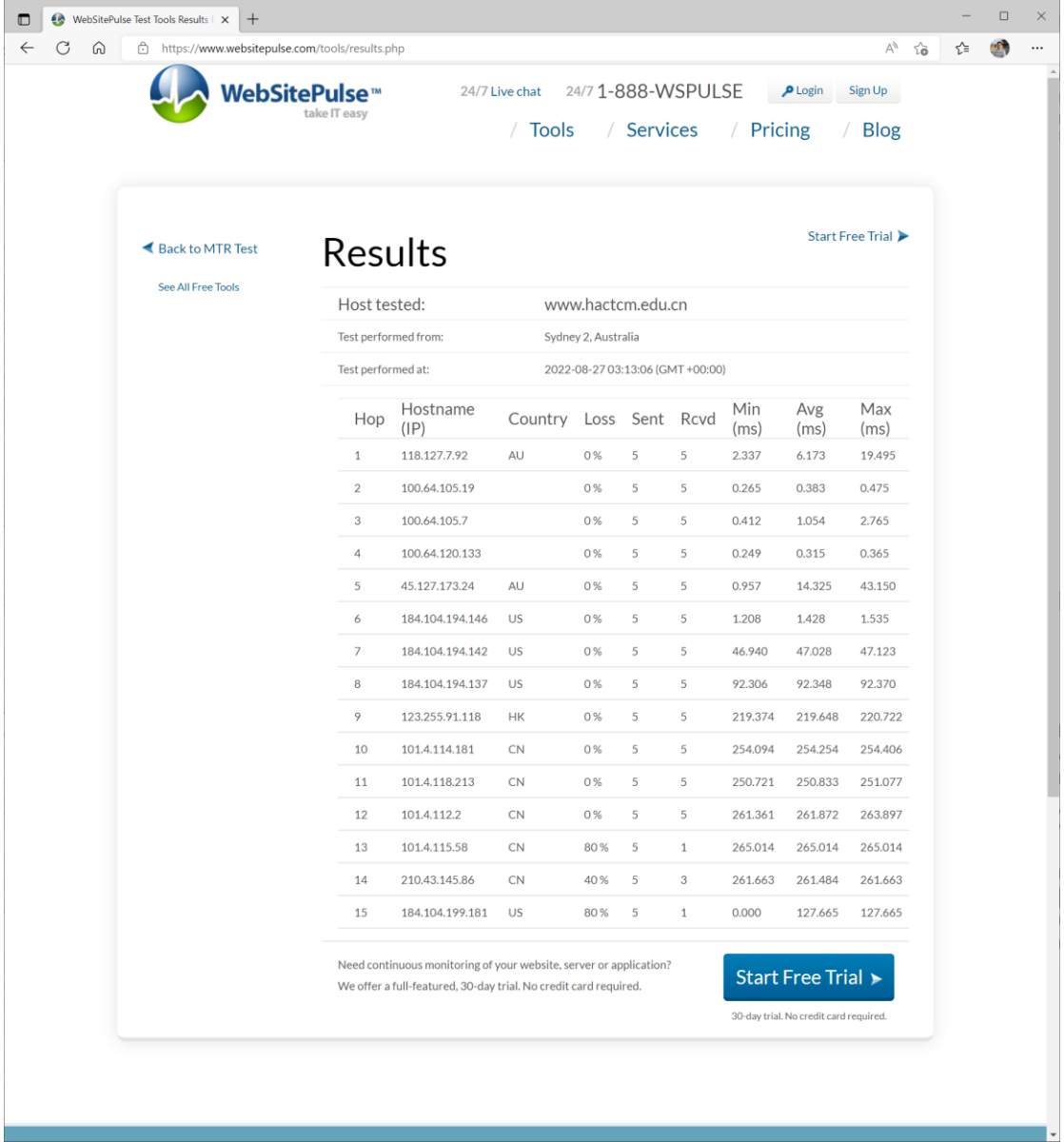
- 互联网可以由多种异构网络互连组成。
- 在网络上，两台主机通信分为两种形式：
  - 直接交付：
    - 在一个物理网络上，数据报被源主机直接传送到目标主机上。
  - 间接交付：
    - 当源主机和目标主机分别处于不同的物理网络上时，数据报由源主机通过网络上的路由器间接的传送到目标主机上。

2.网际协议（IP）


H1
5
4
3
2
1
R1
R2
R3
3
2 2
1 1
主机
H1
间接交付
间接交付
间接交付
间接交付
R5
R4
间接交付
直接交付
R5
R4
主机 H2
H2
5
4
3
2
1
分组在互联网中的传送

2.网际协议（IP）





WebSitePulse Test Tools Results

https://www.websitepulse.com/tools/results.php

WebSitePulse™ take IT easy

24/7 Live chat 24/7 1-888-WSPULSE Login Sign Up

Tools / Services / Pricing / Blog

Back to MTR Test

See All Free Tools

Results

Start Free Trial

Host tested:	www.hactcm.edu.cn
Test performed from:	Sydney 2, Australia
Test performed at:	2022-08-27 03:13:06 (GMT +00:00)

Hop	Hostname (IP)	Country	Loss	Sent	Rcvd	Min (ms)	Avg (ms)	Max (ms)
1	118.127.7.92	AU	0 %	5	5	2.337	6.173	19.495
2	100.64.105.19		0 %	5	5	0.265	0.383	0.475
3	100.64.105.7		0 %	5	5	0.412	1.054	2.765
4	100.64.120.133		0 %	5	5	0.249	0.315	0.365
5	45.127.173.24	AU	0 %	5	5	0.957	14.325	43.150
6	184.104.194.146	US	0 %	5	5	1.208	1.428	1.535
7	184.104.194.142	US	0 %	5	5	46.940	47.028	47.123
8	184.104.194.137	US	0 %	5	5	92.306	92.348	92.370
9	123.255.91.118	HK	0 %	5	5	219.374	219.648	220.722
10	101.4.114.181	CN	0 %	5	5	254.094	254.254	254.406
11	101.4.118.213	CN	0 %	5	5	250.721	250.833	251.077
12	101.4.112.2	CN	0 %	5	5	261.361	261.872	263.897
13	101.4.115.58	CN	80 %	5	1	265.014	265.014	265.014
14	210.43.145.86	CN	40 %	5	3	261.663	261.484	261.663
15	184.104.199.181	US	80 %	5	1	0.000	127.665	127.665

Need continuous monitoring of your website, server or application?
We offer a full-featured, 30-day trial. No credit card required.

Start Free Trial

30-day trial. No credit card required.

2.网际协议（IP）

- IP地址及其表示方法
  - 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）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。
    - ICANN：https://www.icann.org

2.网际协议（IP）

- IP地址及其表示方法
  - IP地址的编址方法共经过了三个历史阶段：

    分类的 IP 地址：

    这是最基本的编址方法，在1981年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。

2.网际协议（IP）

- 分类的IP地址
  - 分类的IP地址就是将IP地址划分为若干个固定类，每一类地址都有两个字段组成：网络号、主机号。
    - 网络号（net-id）：标志主机（或路由器）所连接到的网络。
      - 一个网络号在整个因特网范围内必须是唯一的。
    - 主机号（host-id）：标志该主机（或路由器）。
      - 一个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。
  - 一个IP地址在整个因特网范围内必须是唯一的。

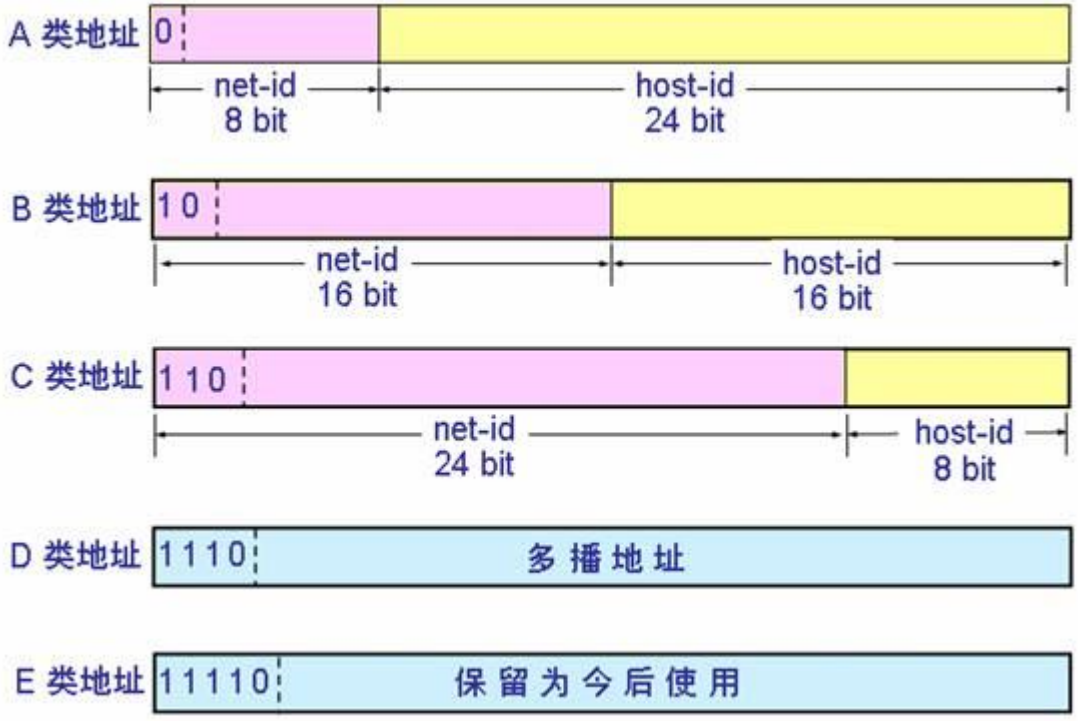
2.网际协议（IP）

- 分类的IP地址
  - 两级的IP地址可以“定义为”：

IP地址 ::= {<网络号>,<主机号>}

2.网际协议（IP）

- 分类的IP地址



2.网际协议（IP）

- 从IP地址的结构来看，IP地址并不仅仅指明一个主机，还指明了主机所连接到的网络。
  - 当初把IP地址划分为A、B、C三个类别，是因为在现实中，有的网络拥有很多主机，有的网络上的主机很少。
  - 把IP地址划分为A、B、C三个类别，能够更好满足用户的不同需求。
- 当某单位申请到一个IP地址时，实际上获得了具有同样网络号的一组地址。
  - 其中主机号是自行分配的，只要无重复即可。

2.网际协议（IP）

- 点分十进制记法：提高IP地址的可读性

计算机存放的IP地址是连续的二进制代码	11010011010001010010000000010010			
每隔8位进行分割,以便于阅读	11010011	01000101	00100000	00010010
将八位二进制数转为十进制数	211	69	32	18
点分十进制记录IP地址	211.69.32.18			

2.网际协议（IP）

- 常用的三种类别的IP地址

IP 地址的指派范围

网络类别	最大可指派的网络数	第一个可指派的网络号	最后一个可指派的网络号	每个网络中的最大主机数
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

2.网际协议（IP）

- 常用的三种类别的IP地址

一般不使用的特殊 IP 地址

网络号	主机号	源地址使用	目的地址使用	代表的意思
0	0	可以	不可	在本网络上的本主机（见 6.6 节 DHCP 协议）
0	host-id	可以	不可	在本网络上的某个主机 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

2.网际协议（IP）

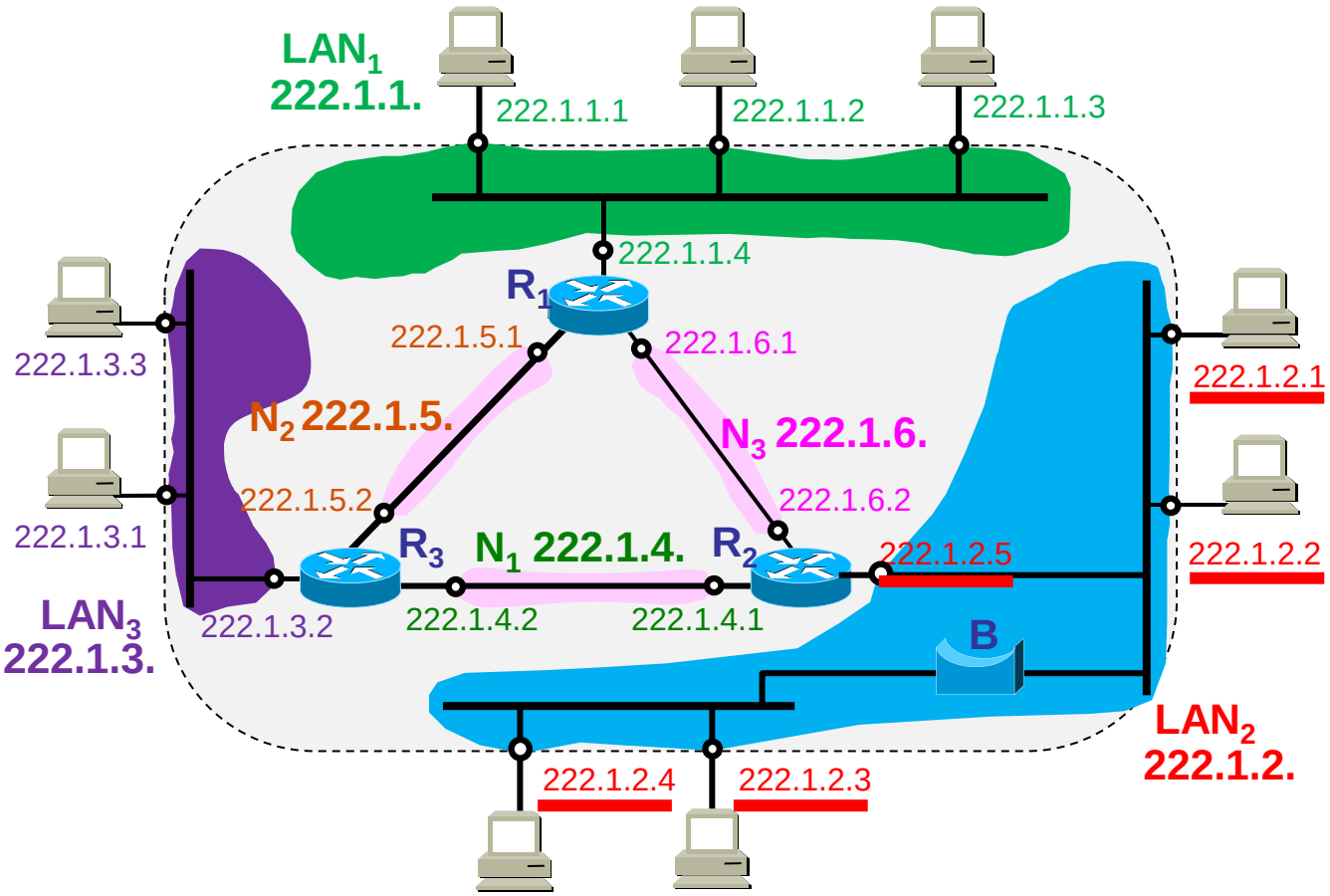
- IP地址具有的重要特点：
  - 每一个IP地址都由网络号和主机号两部分组成。
  - IP地址是一种分等级的地址结构。分等级的两个好处是：
    - 第一，IP 地址管理机构在分配IP地址时只分配网络号，而剩下的主机号则由得到该网络号的单位自行分配，方便了IP地址的管理。
    - 第二，路由器仅根据目的主机所连接的网络号来转发分组（而不考虑目的主机号），这样就可以使路由表中的项目数大幅度减少，从而减小了路由表所占的存储空间。

2.网际协议（IP）

- IP地址具有的重要特点：
  - 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。

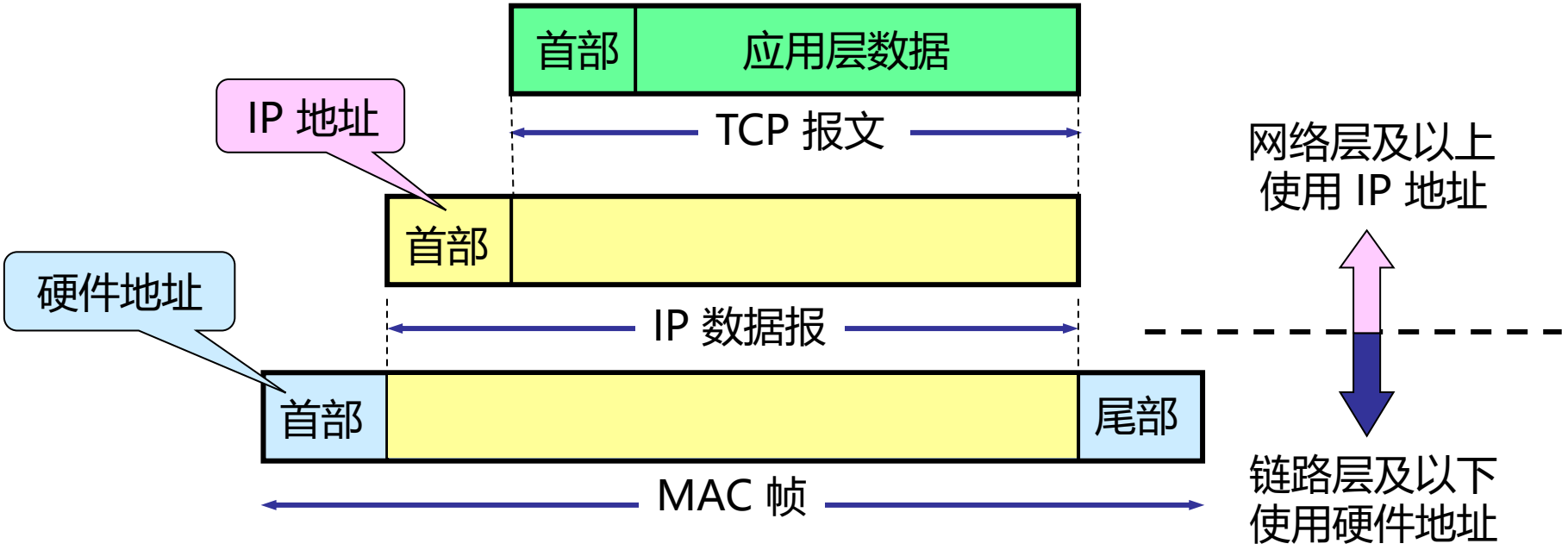
2.网际协议（IP）

- IP地址具有的重要特点：
  - 一个网络是指具有相同网络号net-id的主机的集合。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。
  - 在IP地址中，所有分配到网络号的网络都是平等的。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。


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

2.网际协议（IP）

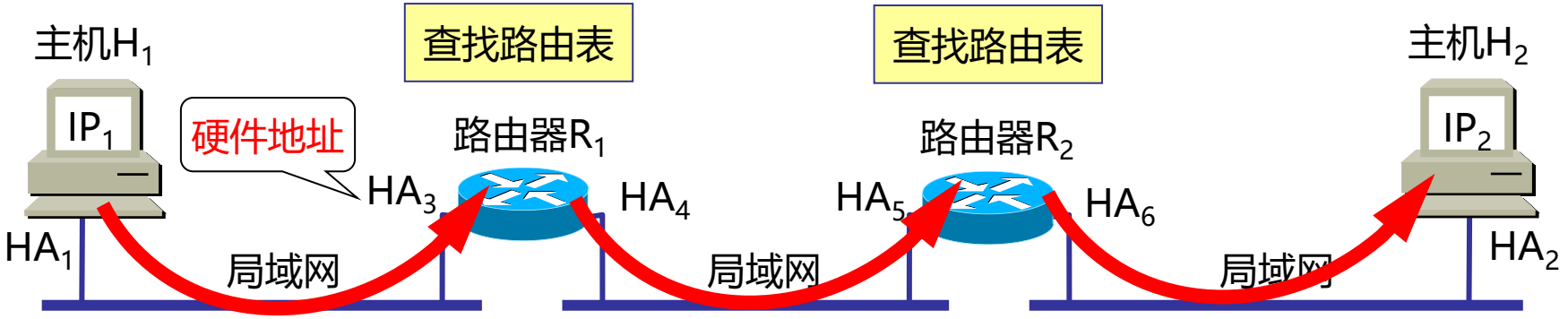
- IP地址和硬件地址的区别：
  - 物理地址是数据链路层和物理层使用的地址。
  - IP地址是网络层和以上各层使用的地址，IP地址是一种逻辑地址。
- 在发送数据时，数据从高层下到低层，然后才到通信链路上传输。使用IP地址的IP数据报一旦交给了数据链路层，就被封装成MAC帧。
  - MAC帧在传送时使用的源地址和目的地址都是硬件地址。
  - 连接在通信链路上的设备在接受MAC帧时，看不到IP地址。只有把数据帧提交给网络层，才能够识别到IP地址。


首部
应用层数据
TCP 报文
IP 地址
首部
硬件地址
IP 数据报
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

2.网际协议（IP）

- IP地址放在IP数据报的首部，硬件地址放在MAC帧的首部。
  - 在网络层和网络层以上使用IP地址，在数据链路层使用硬件地址。
  - 当IP数据报放入数据链路层的MAC帧中以后，整个的IP数据报就成为MAC帧的数据。
  - 在数据链路层看不到数据报的IP地址信息。

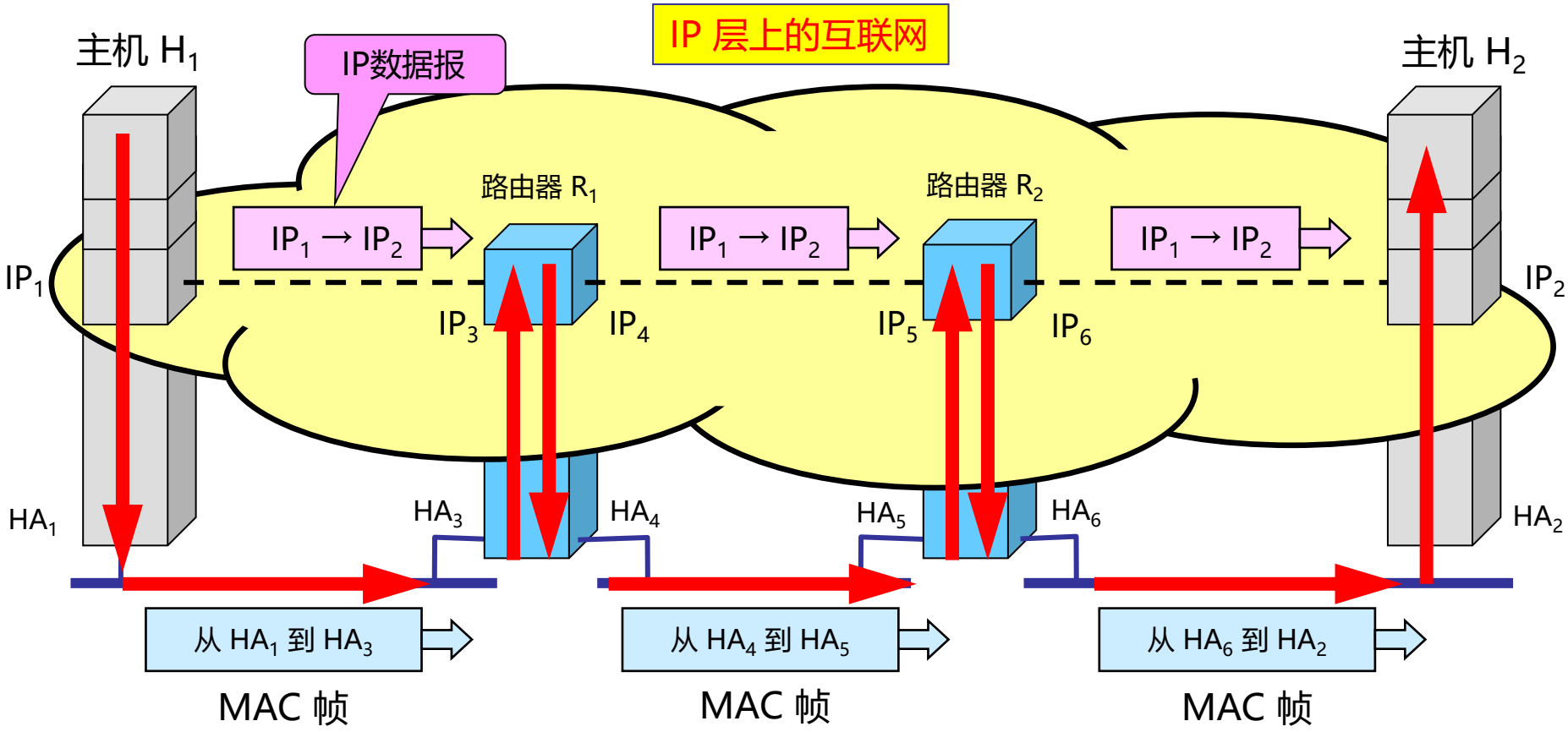
2.网际协议（IP）



通信的路径

H_1→经过 R_1 转发→经过 R_2 转发→H_2

2.网际协议（IP）


IP 层上的互联网
主机 H_1
IP数据报
主机 H_2
路由器 R_1
路由器 R_2
$IP_1 \rightarrow IP_2$
$IP_1 \rightarrow IP_2$
$IP_1 \rightarrow IP_2$
IP_1
IP_2
IP_3
IP_4
IP_5
IP_6
HA_1
HA_3
HA_4
HA_5
HA_6
HA_2
从 HA_1 到 HA_3
从 HA_4 到 HA_5
从 HA_6 到 HA_2
MAC 帧
MAC 帧
MAC 帧

2.网际协议（IP）

不同层次、不同区间的源地址和目的地址

网络层次中地址信息 / 不同的通信阶段	网络层 写入 IP 数据报首部的地址		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 H1 到 R1	IP1	IP2	HA1	HA3
从 R1 到 R2	IP1	IP2	HA4	HA5
从 R2 到 H2	IP1	IP2	HA6	HA2

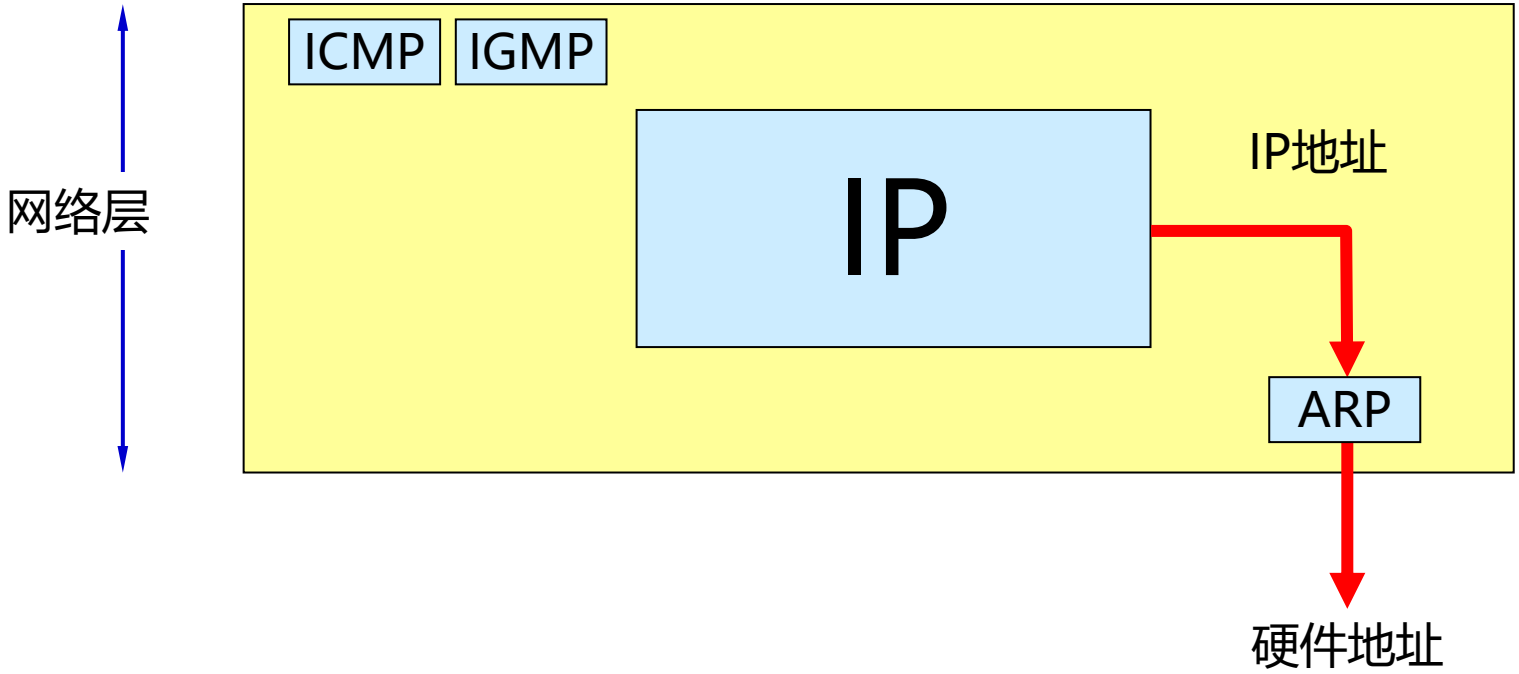
2.网际协议（IP）

- 重点和总结：
  - 在IP层抽象的互联网上只能看到IP数据报。
  - 虽然在IP数据报中有源站IP地址，但是路由器只根据目的站IP地址的网络号进行路由选择。
  - 在局域网的链路层，只能看见MAC帧。IP数据报被封装到MAC帧中作为数据部分。
  - 尽管互连在一起的网络的硬件地址体系各不相同，但IP层抽象的互联网却屏蔽了下层很复杂的细节。
  - 只要在网络层上讨论问题，就能够使用统一的、抽象的IP地址研究主机和主机或路由器之间的通信。

2.网际协议（IP）

- 地址解析协议（ARP）的作用是：在知道一个IP地址时，查找到该IP地址对应的硬件地址。
  - 由于IP协议用到了ARP，因此把ARP放到网络层来介绍。
  - 有些技术文档和书籍考虑到ARP最终解析的是硬件地址，把ARP放到数据链路层。
  - 放到哪一层讨论，不重要。
- 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。

2.网际协议（IP）

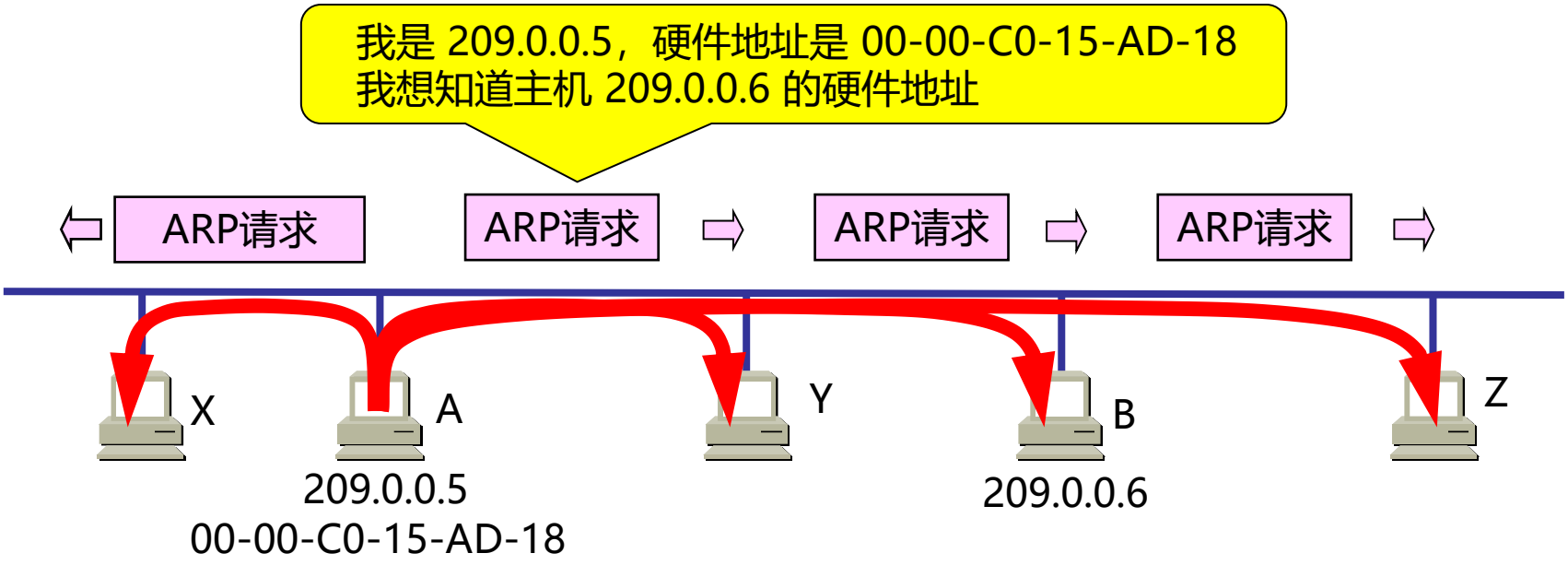


2.网际协议（IP）

- 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。

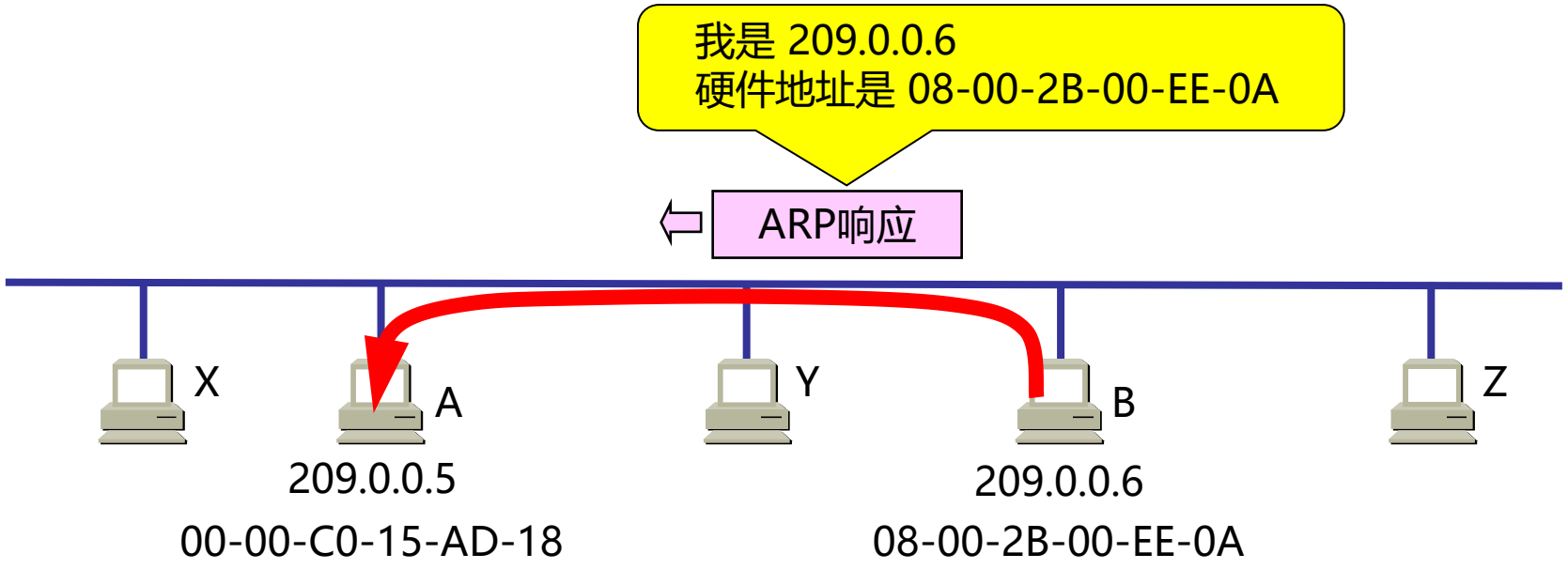
2.网际协议（IP）

主机A广播发送ARP请求分组



2.网际协议（IP）

主机B向A发送ARP响应分组



2.网际协议（IP）

- 为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。

2.网际协议（IP）

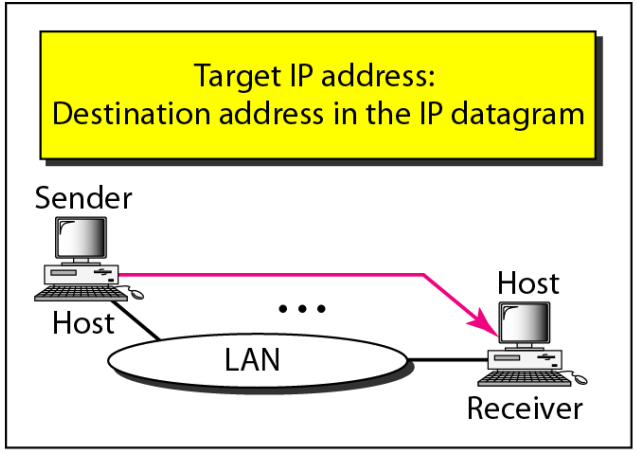
- ARP是解决同一个局域网上的主机或路由器的IP地址和硬件地址的映射问题。
- 如果所要找的主机和源主机不在同一个局域网上，那么就要通过ARP找到一个位于本局域网上的某个路由器的硬件地址，然后把分组发送给这个路由器，让这个路由器把分组转发给下一个网络。
- 剩下的工作就由下一个网络来做。

2.网际协议（IP）

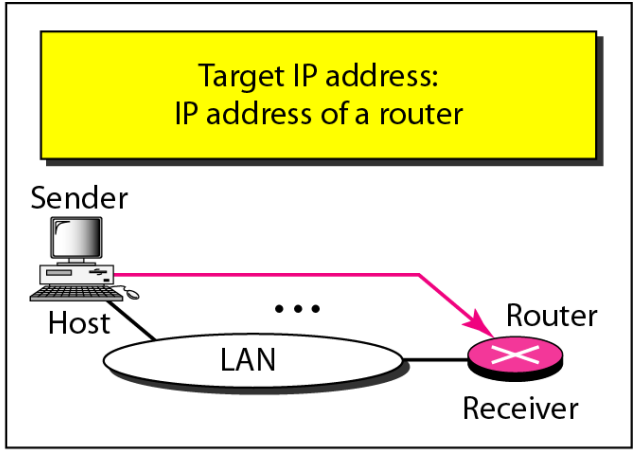
- 从IP地址到硬件地址的解析是自动进行的，主机的用户对这种地址解析过程是不知道的。
- 只要主机或路由器要和本网络上的另一个已知IP地址的主机或路由器进行通信，ARP协议就会自动地将该IP地址解析为链路层所需要的硬件地址。

2.网际协议（IP）

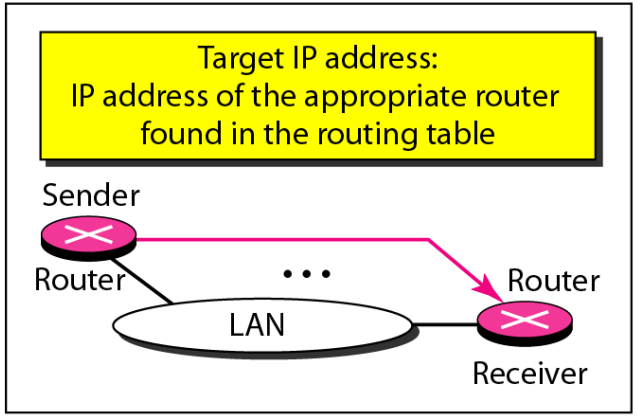
- 使用ARP的四种典型情况：
  - 发送方是主机，要把IP数据报发送到本网络上的另一个主机。这时用ARP找到目的主机的硬件地址。
  - 发送方是主机，要把 IP 数据报发送到另一个网络上的一个主机。这时用ARP找到本网络上的一个路由器的硬件地址。剩下的工作由路由器来完成。
  - 发送方是路由器，要把IP数据报转发到本网络上的一个主机。这时用ARP找到目的主机的硬件地址。
  - 发送方是路由器，要把IP数据报转发到另一个网络上的一个主机。这时用ARP找到本网络上另一个路由器的硬件地址。剩下的工作由路由器来完成。



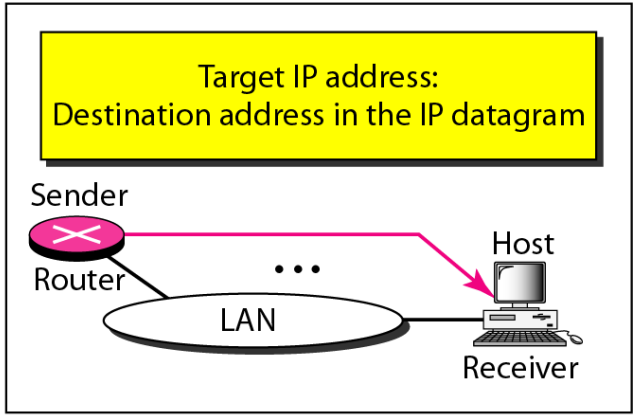
Case 1. A host has a packet to send to another host on the same network.



Case 2. A host wants to send a packet to another host on another network. It must first be delivered to a router.



Case 3. A router receives a packet to be sent to a host on another network. It must first be delivered to the appropriate router.



Case 4. A router receives a packet to be sent to a host on the same network.

2.网际协议（IP）

- 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。

2.网际协议（IP）

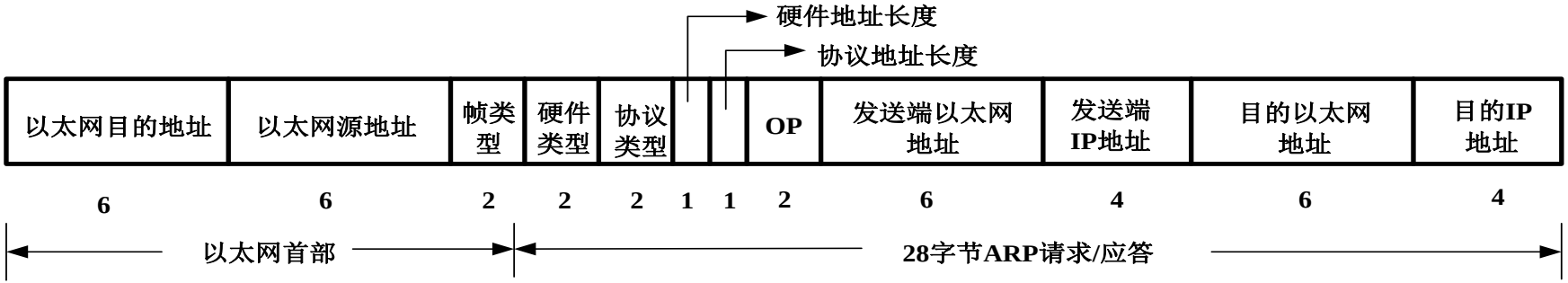
- 如何查看本地主机的ARP高速缓存?
  - 在Linux操作系统中，通过在shell环境下输入“arp”查看。
  - 在Windows操作系统中，通过【运行】【cmd】，在命令窗体中输入“arp -a”查看。
  - 可以通过arp命令进行更多操作。

- 现场演示arp命令的使用，并介绍ARP高速缓存。

```
ruanxiaolong@Teach-Ubuntu-Server-VMs:~$ arp
Address                  HWtype  HWaddress           Flags Mask            Iface
211.69.32.1              ether   e4:68:a3:a3:fa:7d   C                     eth0
211.69.32.122            ether   00:0c:29:16:25:97   C                     eth0
HACTCM-DNS-2             ether   bc:ae:c5:07:7a:82   C                     eth0
```

2.网际协议（IP）

- ARP的数据帧格式：



- 通过Wireshark进行ARP数据报的分析。

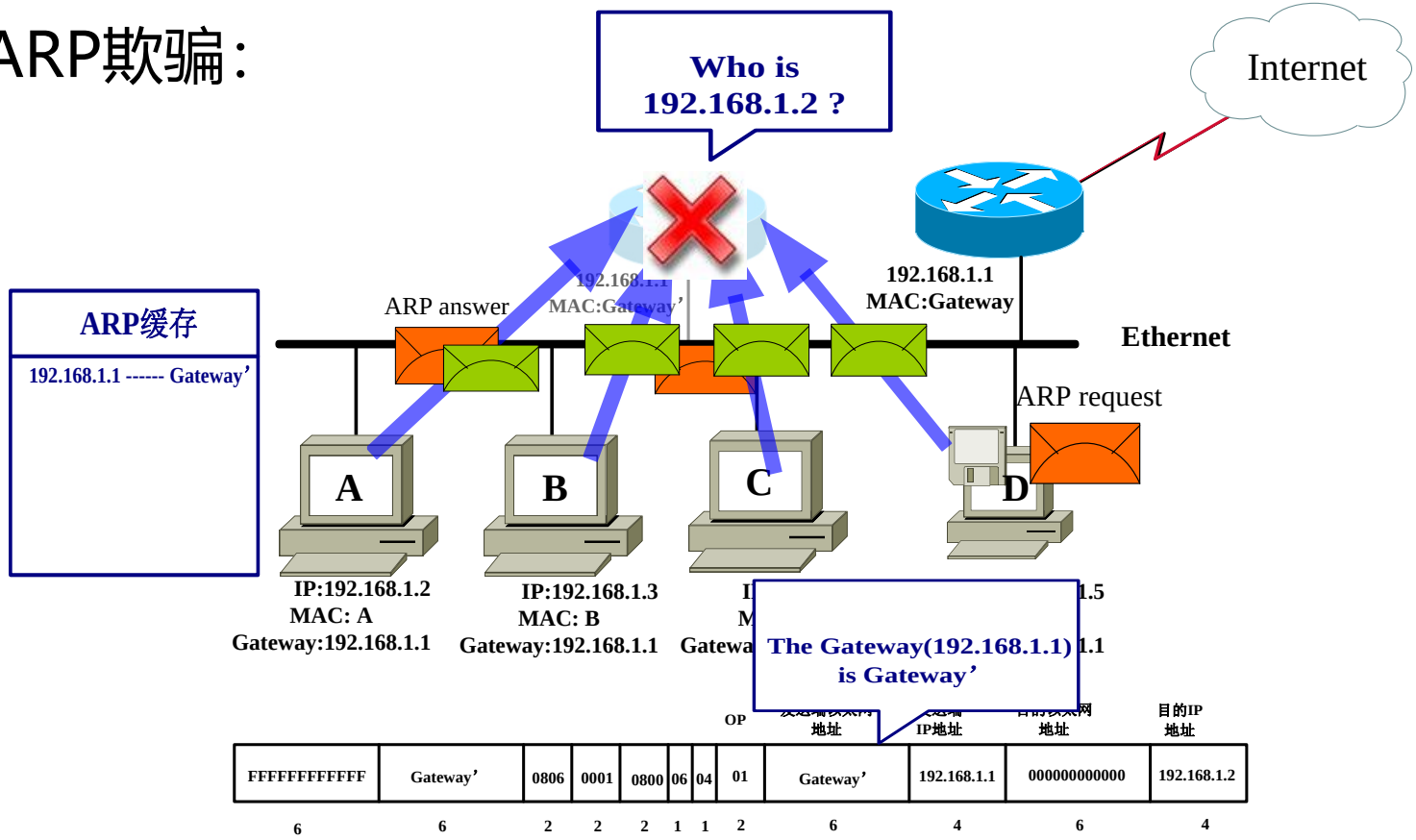
2.

(ARP)

```
3146 39.966249000    HuaweiTe_a3:fa:7d   Broadcast           ARP      60 Who has 211.69.32.179?  Tell 211.69.32.1
3150 40.123107000    HuaweiTe_a3:fa:7d   Broadcast           ARP      60 Who has 211.69.32.128?  Tell 211.69.32.1

Frame 3146: 60 bytes on wire (480 bits), 60 bytes captured (480 bits) on interface 0
    Interface id: 0
    Encapsulation type: Ethernet (1)
    Arrival Time: Apr  3, 2014 15:28:56.006516000 [illegible]
    [Time shift for this packet: 0.000000000 seconds]
    Epoch Time: 1396510136.006516000 seconds
    [Time delta from previous captured frame: 0.015001000 seconds]
    [Time delta from previous displayed frame: 0.041086000 seconds]
    [Time since reference or first frame: 39.966249000 seconds]
    Frame Number: 3146
    Frame Length: 60 bytes (480 bits)
    Capture Length: 60 bytes (480 bits)
    [Frame is marked: False]
    [Frame is ignored: False]
    [Protocols in frame: eth:arp]
    [Coloring Rule Name: ARP]
    [Coloring Rule String: arp]
Ethernet II, Src: HuaweiTe_a3:fa:7d (e4:68:a3:a3:fa:7d), Dst: Broadcast (ff:ff:ff:ff:ff:ff)
  Destination: Broadcast (ff:ff:ff:ff:ff:ff)
      Address: Broadcast (ff:ff:ff:ff:ff:ff)
      .... ..1. .... .... .... .... = LG bit: Locally administered address (this is NOT the factory default)
      .... ...1 .... .... .... .... = IG bit: Group address (multicast/broadcast)
  Source: HuaweiTe_a3:fa:7d (e4:68:a3:a3:fa:7d)
      Address: HuaweiTe_a3:fa:7d (e4:68:a3:a3:fa:7d)
      .... ..0. .... .... .... .... = LG bit: Globally unique address (factory default)
      .... ...0 .... .... .... .... = IG bit: Individual address (unicast)
    Type: ARP (0x0806)
    Padding: 000000000000000000000000000000000000
Address Resolution Protocol (request)
    Hardware type: Ethernet (1)
    Protocol type: IP (0x0800)
    Hardware size: 6
    Protocol size: 4
    Opcode: request (1)
    Sender MAC address: HuaweiTe_a3:fa:7d (e4:68:a3:a3:fa:7d)
    Sender IP address: 211.69.32.1 (211.69.32.1)
    Target MAC address: 00:00:00_00:00:00 (00:00:00:00:00:00)
    Target IP address: 211.69.32.179 (211.69.32.179)

0000  ff ff ff ff ff ff e4 68  a3 a3 fa 7d 08 06 00 01   .......h ...}....
0010  08 00 06 04 00 01 e4 68  a3 a3 fa 7d d3 45 20 01   .......h ...}.E .
0020  00 00 00 00 00 00 d3 45  20 b3 00 00 00 00 00 00   .......E  .......
0030  00 00 00 00 00 00 00 00  00 00 00 00              ........ ....
```

2.网际协议（IP）

- ARP欺骗：

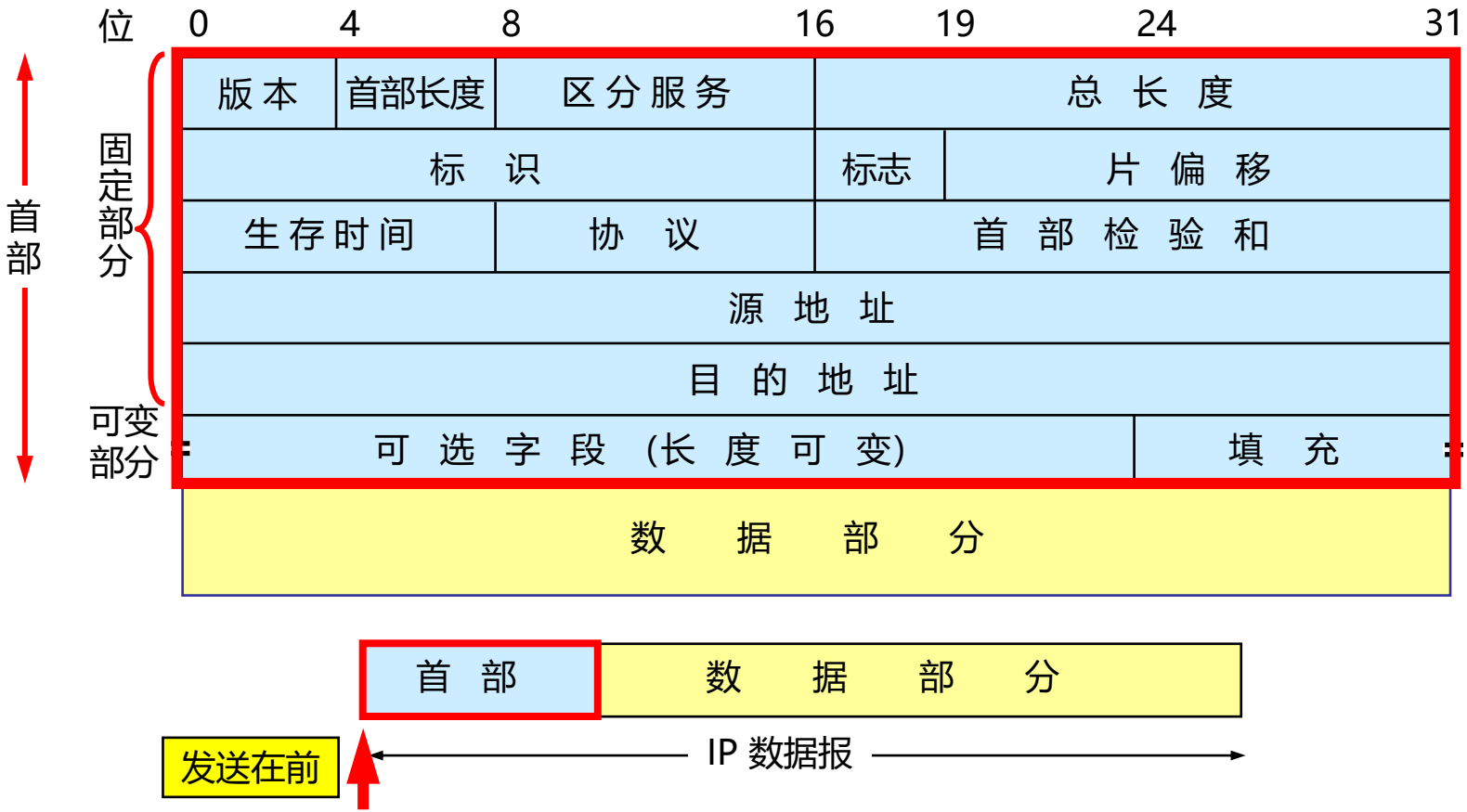


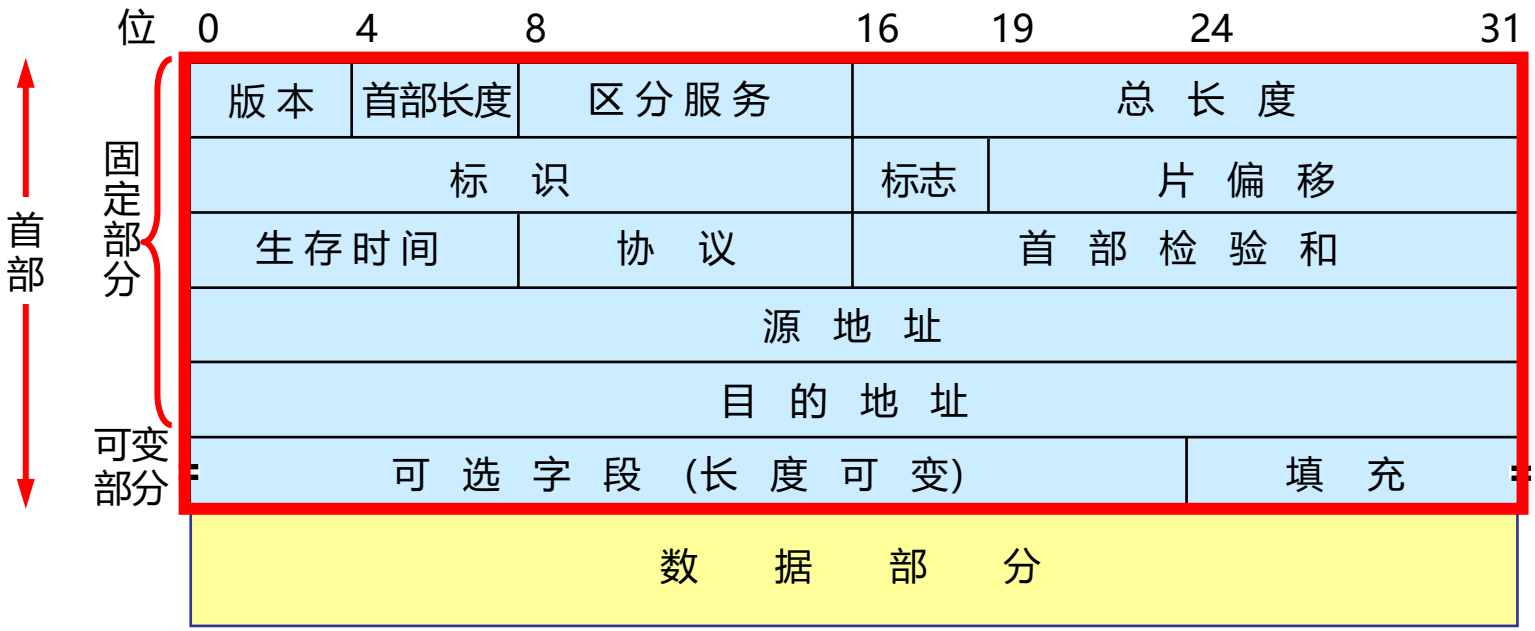
2.网际协议（IP）

- 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IP数据报必须具有的。
  - 在首部固定部分的后面是一些可选字段，其长度是可变的。
- IP数据报的格式也能够说明IP协议的功能。

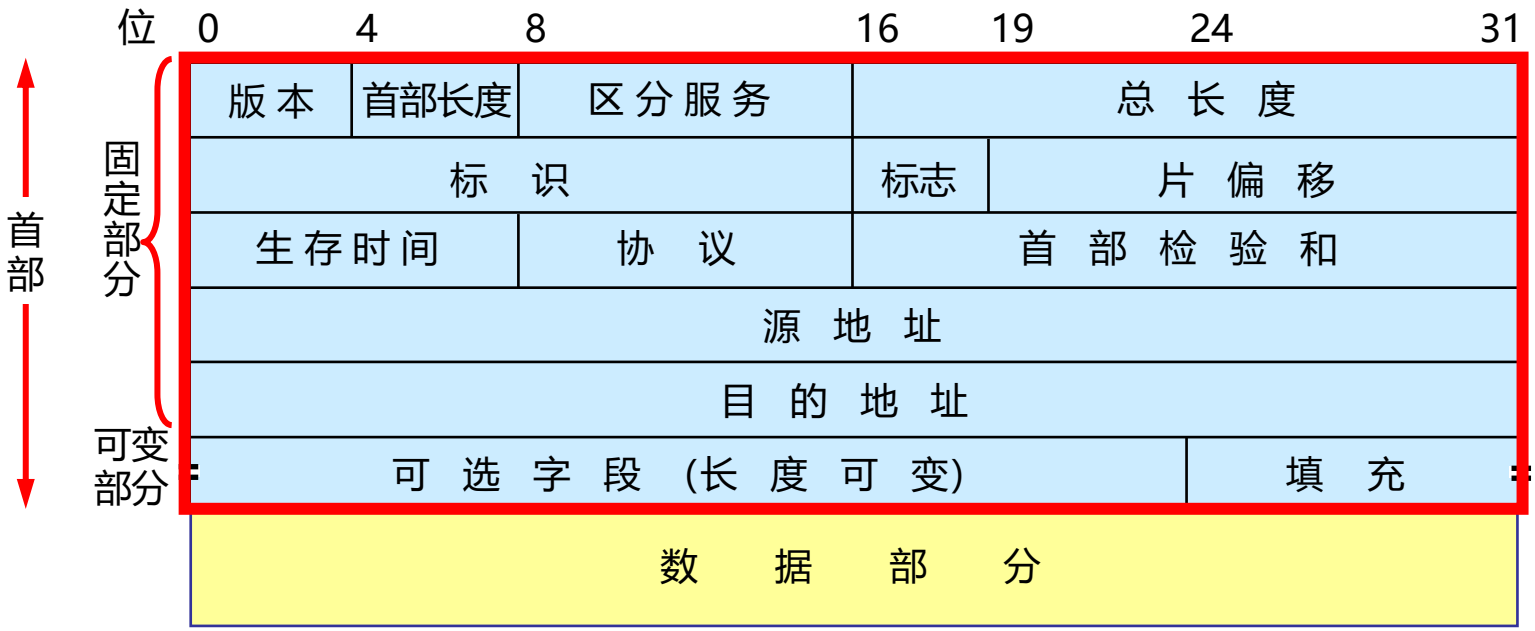
2.网际协议（IP）

2.5 IP数据报的格式

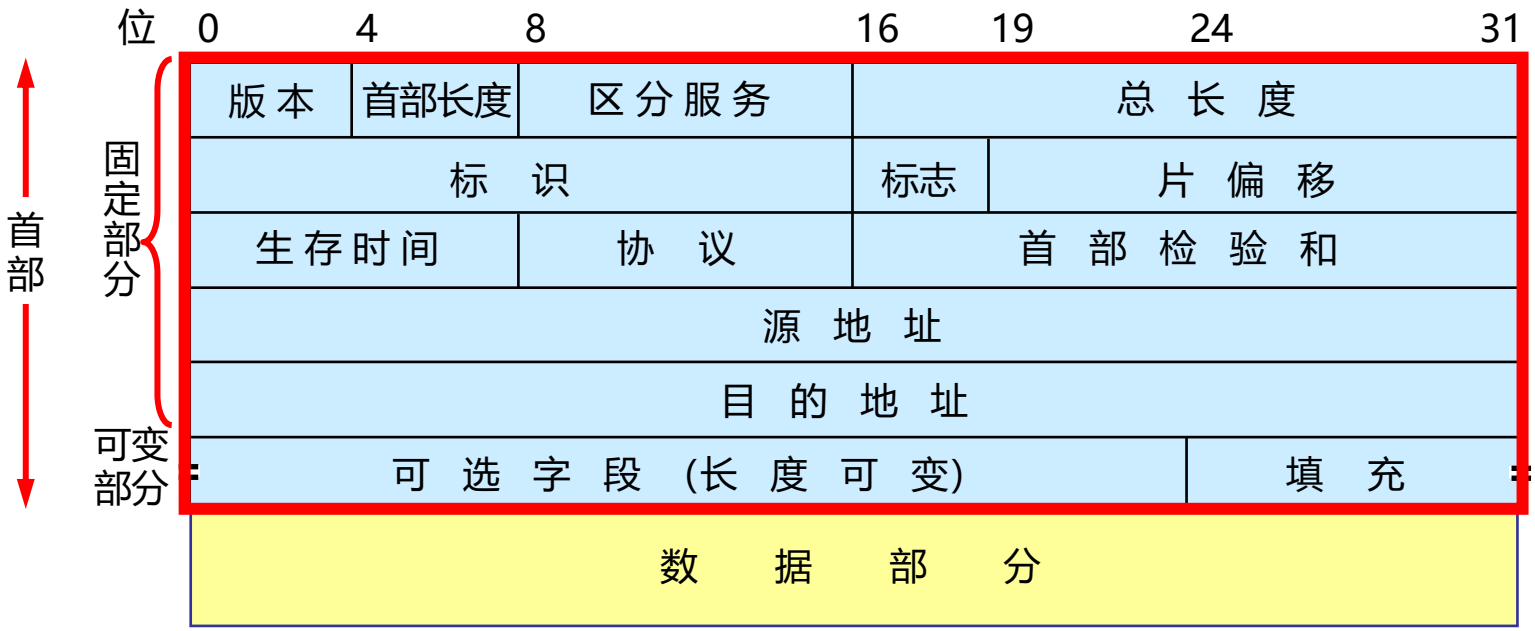




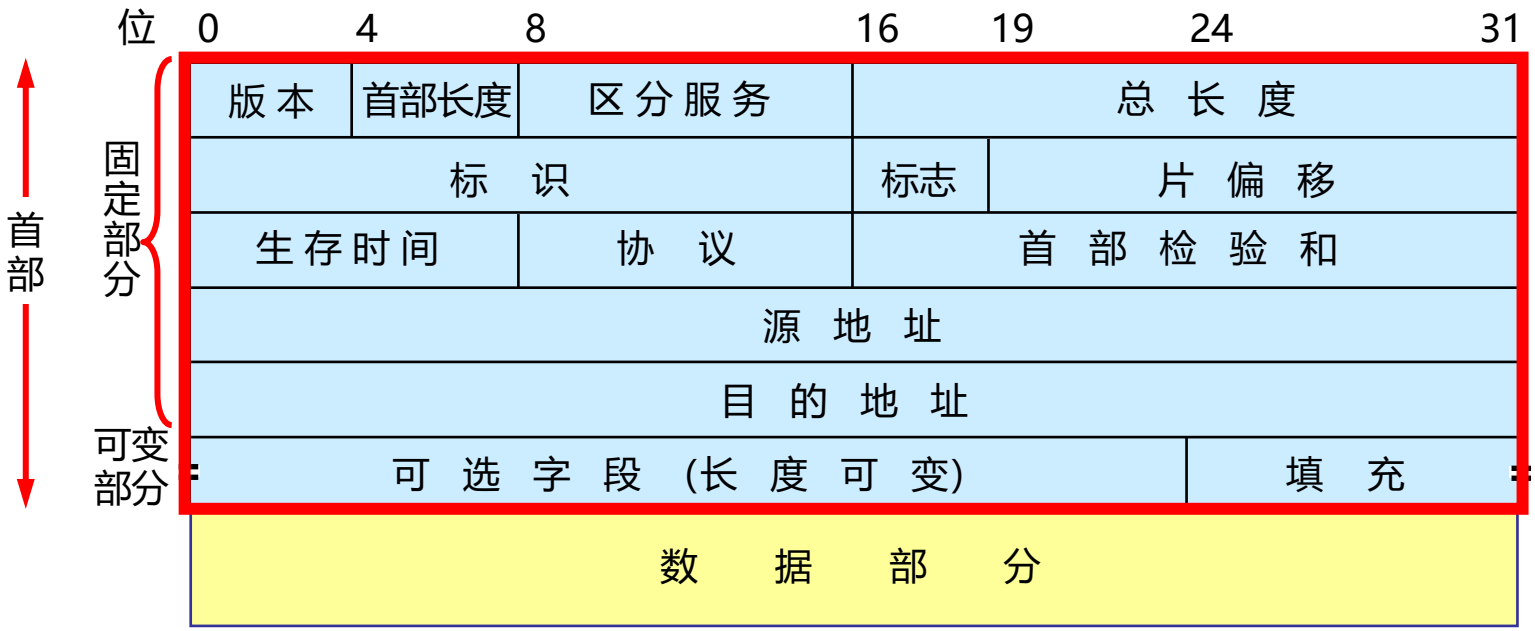
- 版本：4位，指IP协议的版本，目前的IP协议版本号为4(即 IPv4)。
- 首部长度：4位，可表示的最大数值是15个单位(一个单位为4字节)，因此IP的首部长度的最大值是60字节。
- 区分服务：8位，用来获得更好的服务。在旧标准中叫做服务类型，但实际上一直未被使用过。1998年这个字段改名为区分服务。只有在使用区分服务（DiffServ）时，这个字段才起作用。在一般的情况下都不使用这个字段。



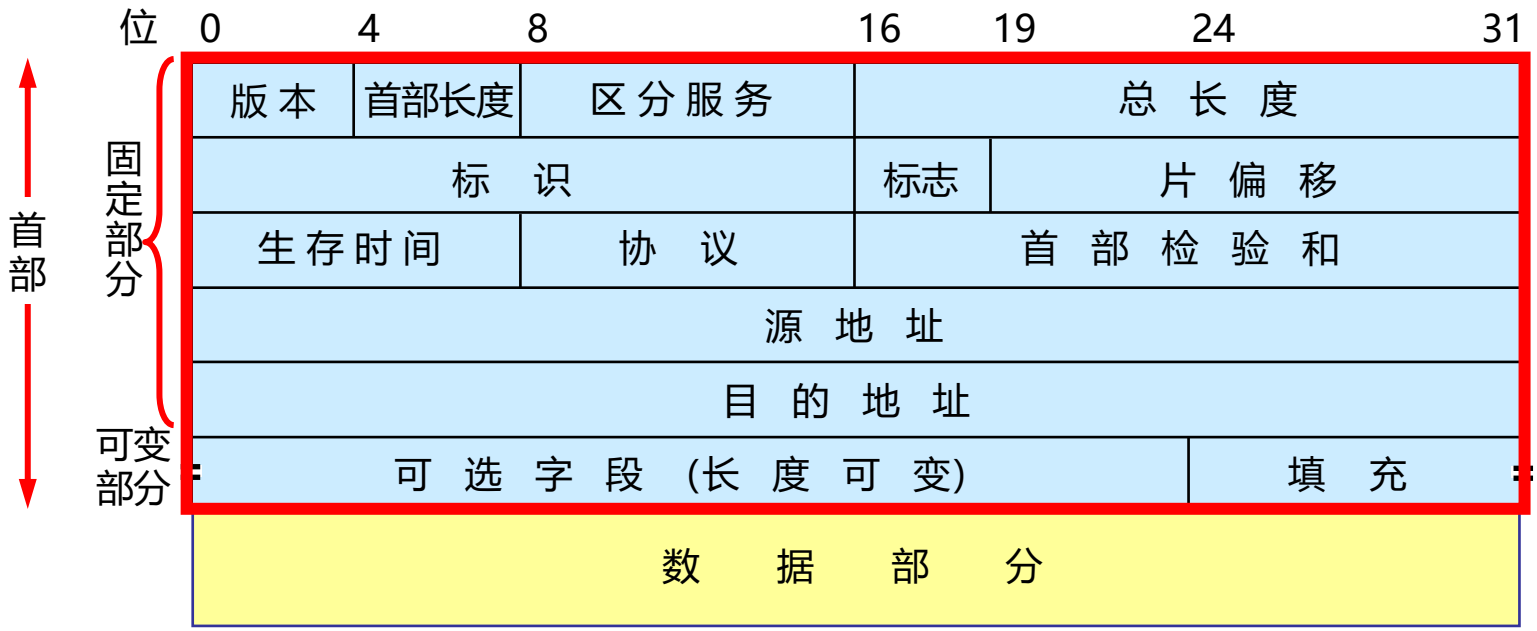
- 总长度：16位，指首部和数据之和的长度，单位为字节，因此数据报的最大长度为65535字节。总长度必须不超过最大传送单元MTU，如果超过了，就需要把过长的数据报进行分片处理。
- 标识(identification)：16位，它是一个计数器，用来产生数据报的标识。
- 标志(flag)：3位，目前只有前两位有意义。
  - 标志字段的最低位是MF(More Fragment)，MF=1表示后面“还有分片”，MF=0表示最后一个分片。
  - 标志字段中间的一位是DF(Don't Fragment)，只有当DF=0时才允许分片。



- 片偏移：13位。片偏移指出：较长的分组在分片后，某片在原分组中的相对位置。也就是说，相对于用户数据字段的起点，该片从何处开始。片偏移以8个字节为偏移单位，每个分片的长度一定是8个字节的整数倍。
- 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，TTL限制的为“跳数限制”，路由器转发数据报之前把TTL减1，如果TTL为0，路由器就丢弃这个数据报。因此TTL的单位是“跳数”。
- 协议：8位，协议字段指出此数据报携带的数据使用何种协议，以方便目的主机的 IP 层将数据部分上交给哪个处理程序进行处理。



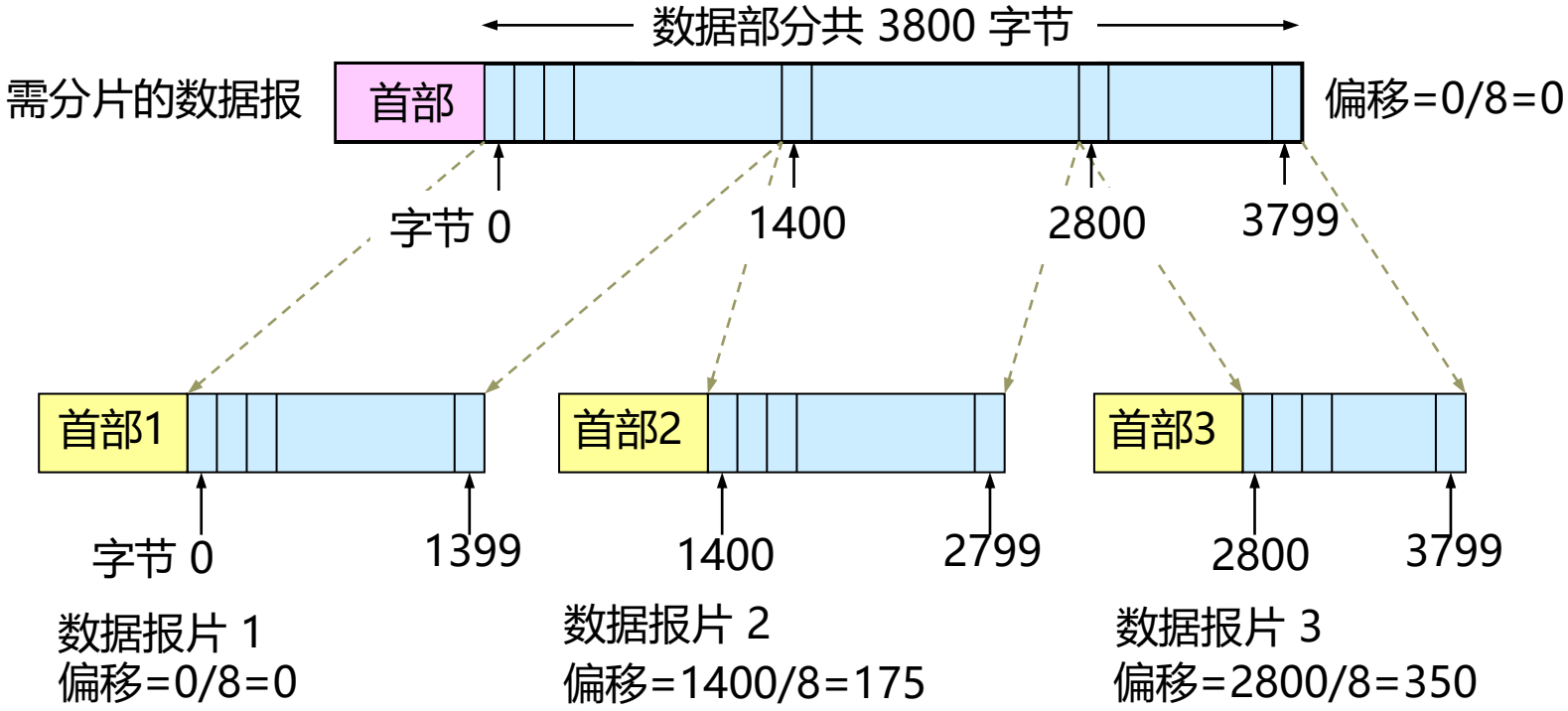
- 首部校验和：16位。这个校验只检测数据报的首部，不对数据报的数据部分进行校验。首部校验和没有使用CRC这样复杂的计算，而是采用了更加简单的方法。
  （算法参考教材的内容，并进行介绍）
- 源地址：32位，发送数据报的IP地址。
- 目的地址：32位，接收数据报的IP地址。



- IP首部的可变部分就是一个选项字段。
  - 选项字段用来支持排错、测量以及安全等措施，内容很丰富。
  - 此字段的长度为1-40个字节不等，取决于所选择的项目内容。
- 增加首部的可变部分是为了提高IP数据报的功能，但同时也使得IP数据报的首部变为可变的，增加了路由器的开销。
- IPv6将IP数据报的首部长度定为为固定的。

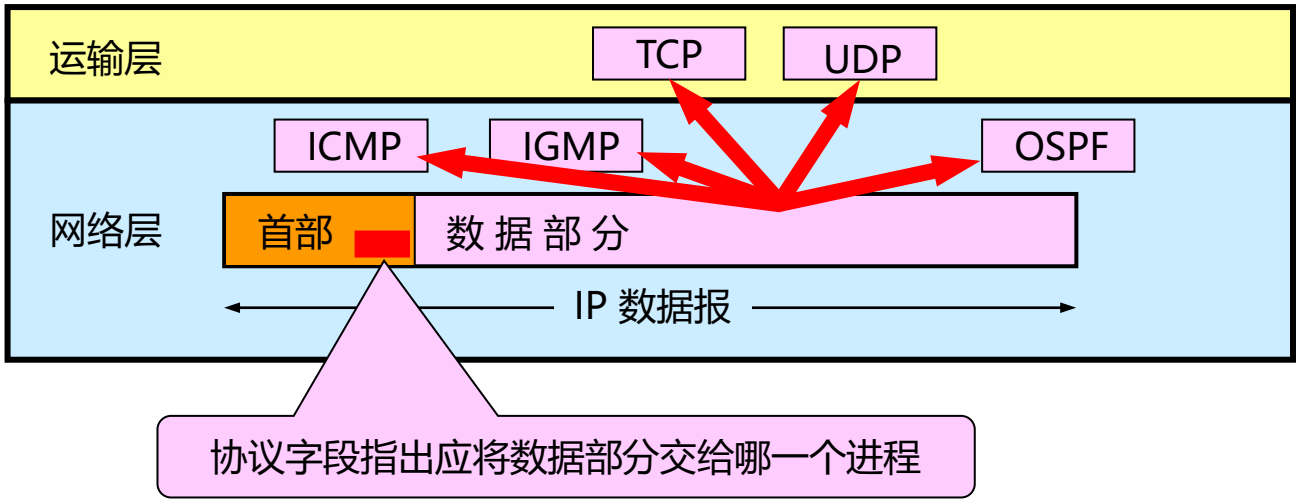
2.网际协议（IP）

- IP数据报分片的分析：



2.网际协议（IP）

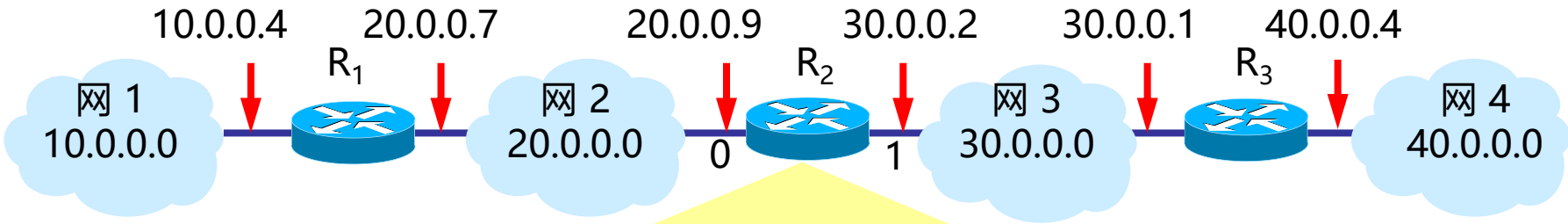
- 协议字段的作用：



2.网际协议（IP）

- 举个极端的例子：
  - 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。
  - 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。
  - 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。
  - 可使路由表大大简化。

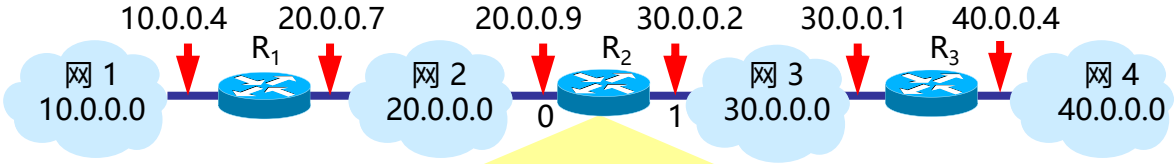
2.网际协议（IP）



路由器R_2的路由表

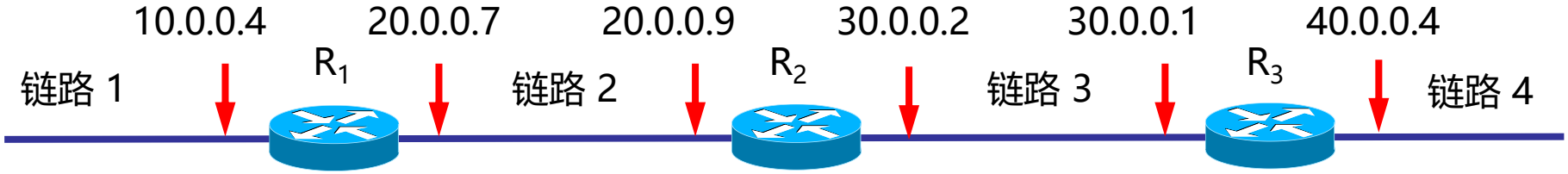
目的主机所在的网络	下一跳地址
20.0.0.0	直接交付，接口 0
30.0.0.0	直接交付，接口 1
10.0.0.0	20.0.0.7
40.0.0.0	30.0.0.1

2.网际协议（IP）



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



2.网际协议（IP）

- 在互联网上转发分组时，是从一个路由器转发到下一个路由器。
- 在路由表中，对每一条路由最主要的是两个信息：

  （目的网络地址，下一跳地址）
- 根据目的网络地址就能确定下一跳路由器，最终结果是：
  - 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。

2.网际协议（IP）

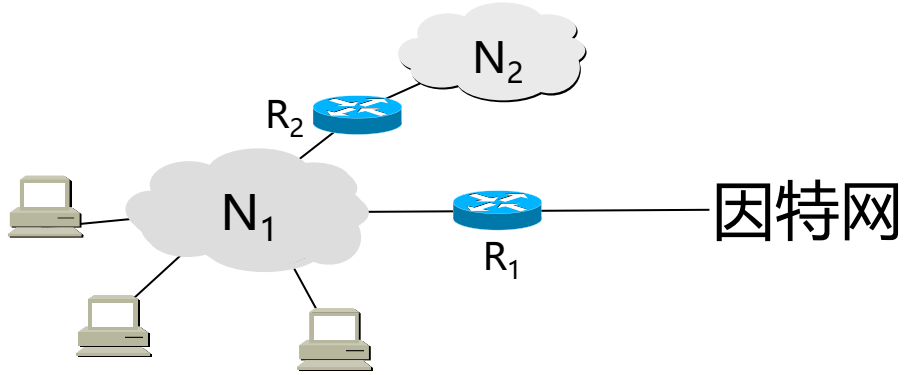
- 特定主机路由：
  - 是为特定的目的主机指明一个路由。
  - 采用特定主机路由可使网络管理人员能更方便地控制网络和测试网络，同时也可在需要考虑某种安全问题时采用这种特定主机路由。
- 默认路由：
  - 采用默认路由以减少路由表所占用的空间和搜索路由表所用的时间。这种转发方式在一个网络只有很少的对外连接时是很有用的。
  - 默认路由在主机发送IP数据报时往往更能显示出它的好处。
  - 如果一个主机连接在一个小网络上，而这个网络只用一个路由器和因特网连接，那么在这种情况下使用默认路由是非常合适的。

2.网际协议（IP）

- 默认路由：
  - 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。
  - 局域网主要使用默认路由接入互联网。

路由表

目的网络	下一跳
N_1	直接
N_2	R_2
默认	R_1



2.网际协议（IP）

- 需要注意的是：
  - IP数据报的首部中没有地方可以用来指明“下一跳路由器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。

2.网际协议（IP）

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。

3.划分子网与构建超网

- 举例讨论：
  - 某单位获得A类地址，却只有300台主机。
  - 路由器里面的路由表应该有多少条记录?
  - 某单位紧急扩展网络后，如何让新扩展网络接入互联网?

3.划分子网与构建超网

- 两级IP地址的不足：
  - 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。
- 解决思路：
  - 申请到地址后，可以再进行二次划分，把一个A类或B类地址划分为多个网络。
  - 让地址段更小、更灵活。

3.划分子网与构建超网

- 从1985年起，在IP地址中又增加了一个新的字段：“子网号字段”，两级的IP地址变成为三级的IP地址。
  - 从两级的IP地址变为三级的IP地址后，解决了分类IP地址管理的不足，让IP管理和使用变得更加灵活。
  - 将两级的IP地址变为三级的IP地址的做法，叫作划分子网(subnetting)，或叫子网划分、子网路由选择。
  - 划分子网是因特网的正式标准协议。

3.划分子网与构建超网

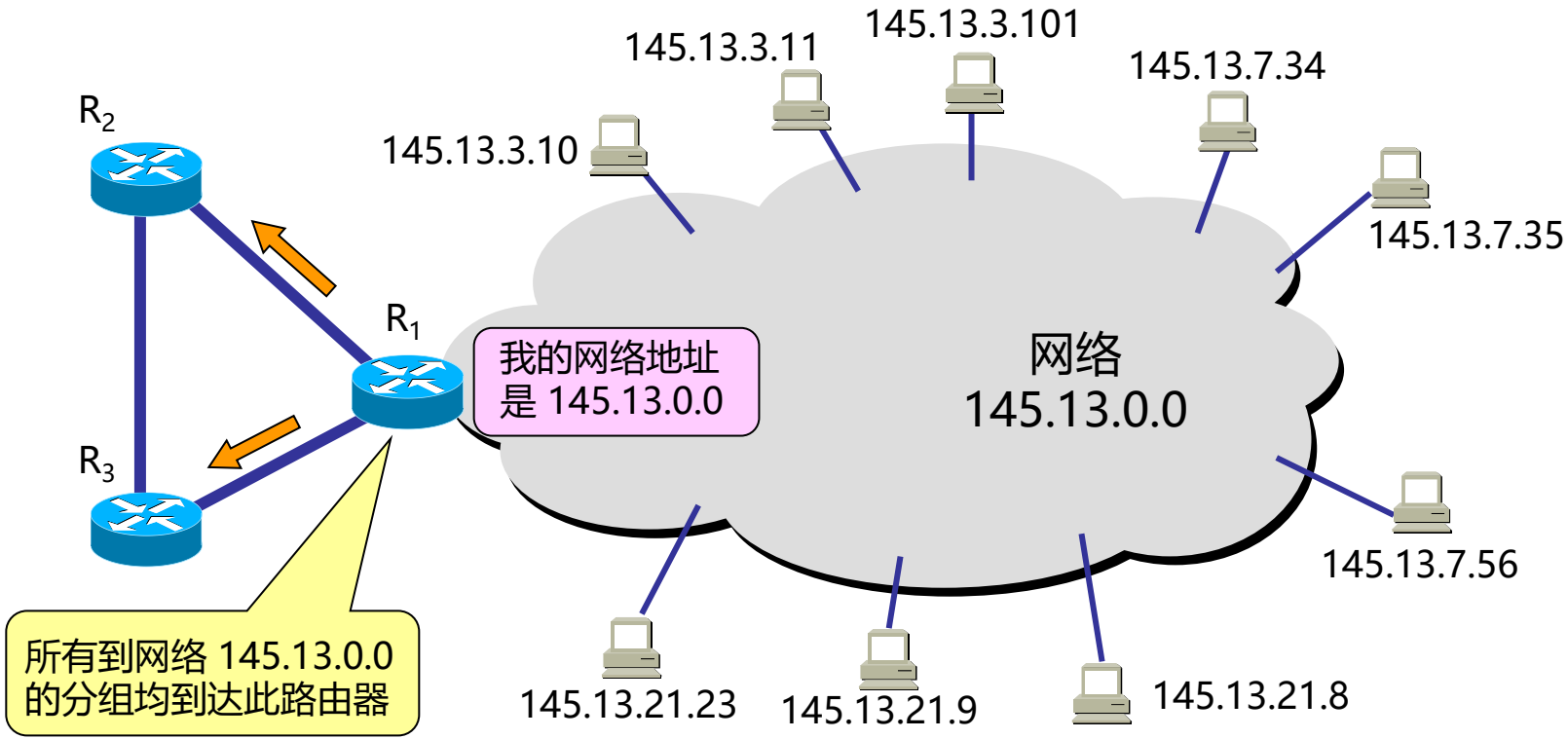
- 划分子网的思路：
  - 划分子网纯属一个单位内部的事情。
  - 单位对外仍然表现为没有划分子网的网络。
  - 从主机号借用若干个位作为子网号 subnet-id，而主机号 host-id 也就相应减少了若干个位。

IP地址 ::= {<网络号>, <子网号>, <主机号>}

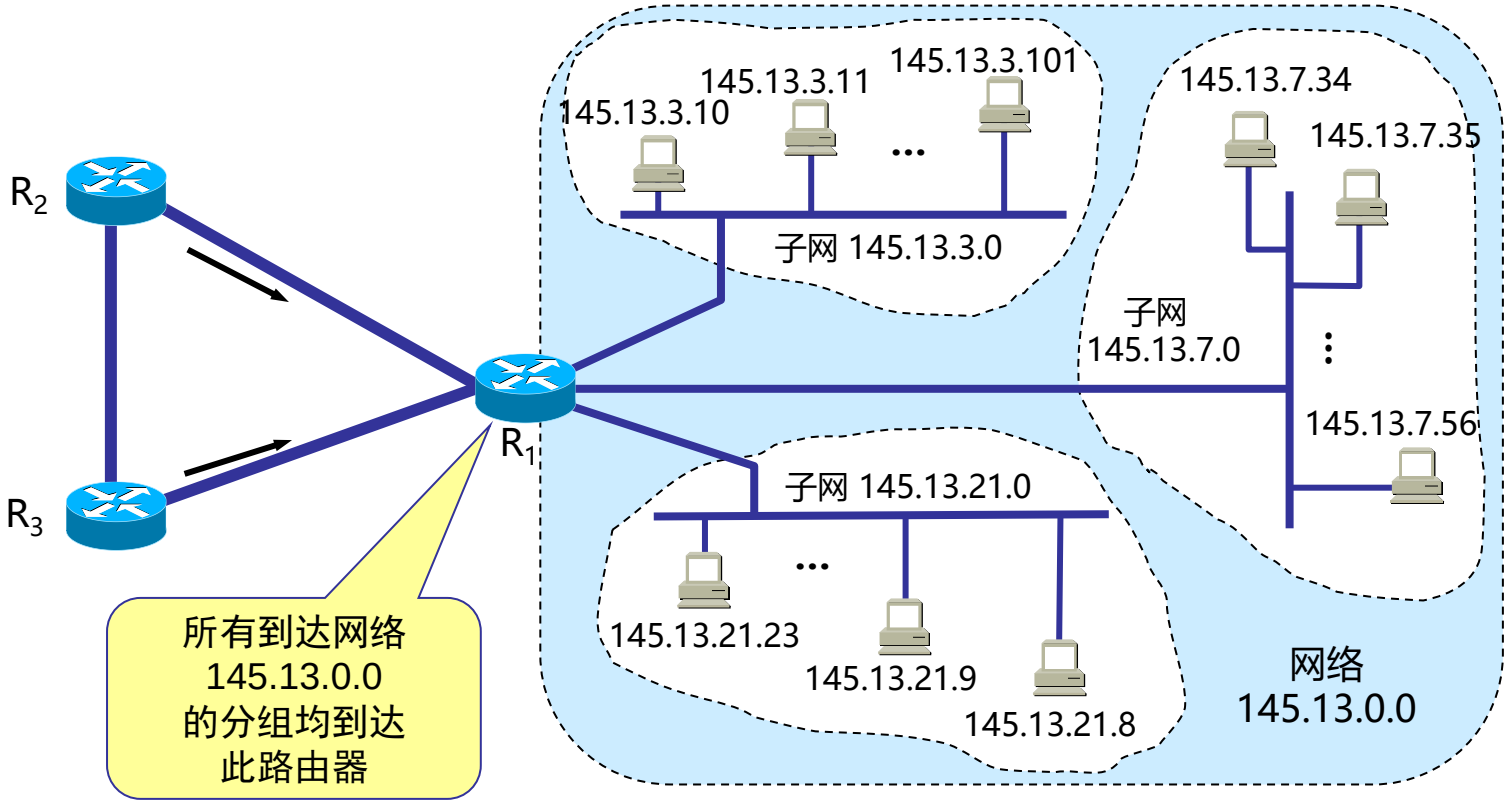
3.划分子网与构建超网

- 划分子网的思路：
  - 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。
  - 最后将IP数据报直接交付目的主机。

3.划分子网与构建超网


R2
R1
R3
145.13.3.11
145.13.3.101
145.13.7.34
145.13.3.10
145.13.7.35
我的网络地址
是 145.13.0.0
网络
145.13.0.0
145.13.7.56
所有到网络 145.13.0.0
的分组均到达此路由器
145.13.21.23
145.13.21.9
145.13.21.8

3.划分子网与构建超网


R2
R3
R1
145.13.3.10
145.13.3.11
145.13.3.101
子网 145.13.3.0
145.13.7.34
145.13.7.35
子网
145.13.7.0
145.13.7.56
子网 145.13.21.0
145.13.21.23
145.13.21.9
145.13.21.8
网络
145.13.0.0
所有到达网络
145.13.0.0
的分组均到达
此路由器

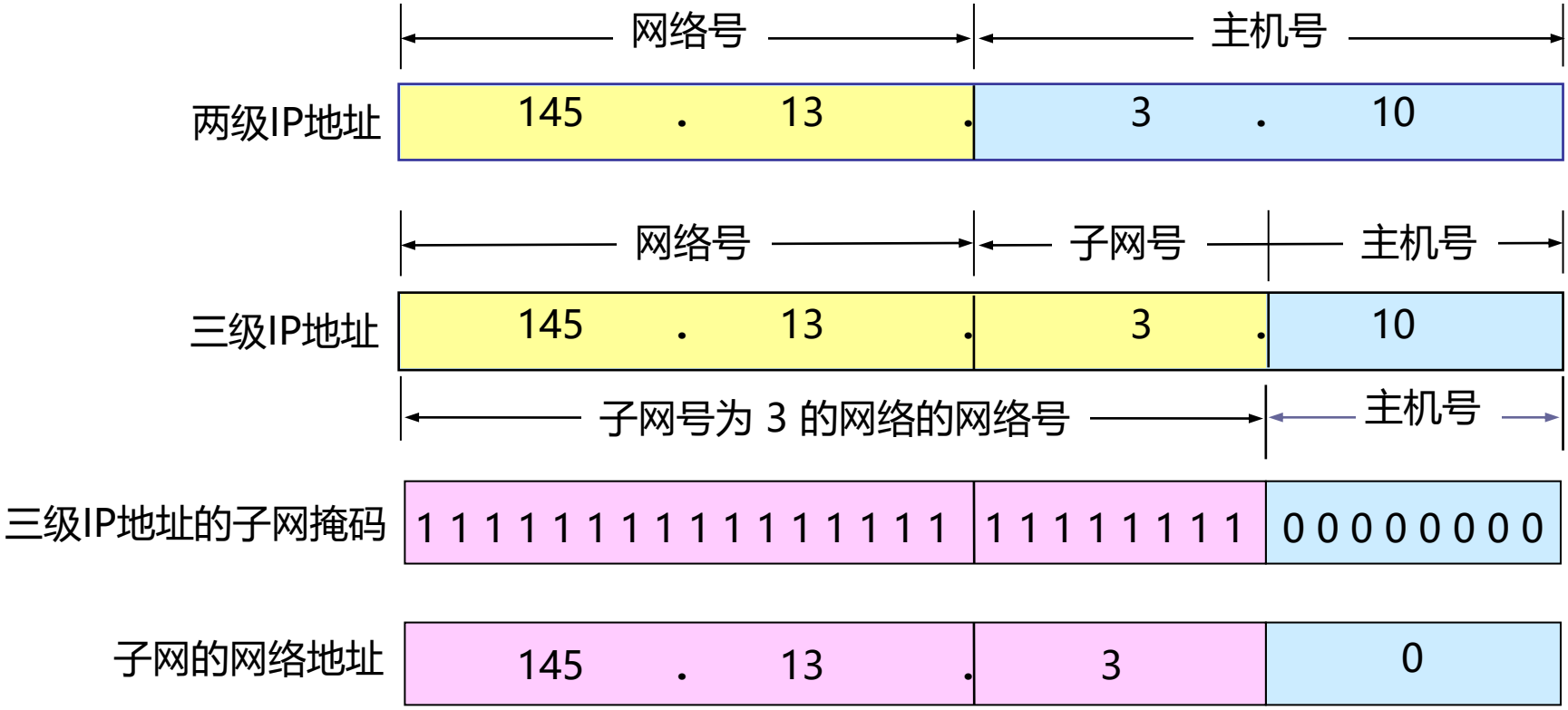
3.划分子网与构建超网

- 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。
- 划分子网是单位内部为了管理而自行开展的工作，对于上层接入网络而讲没有任何影响。

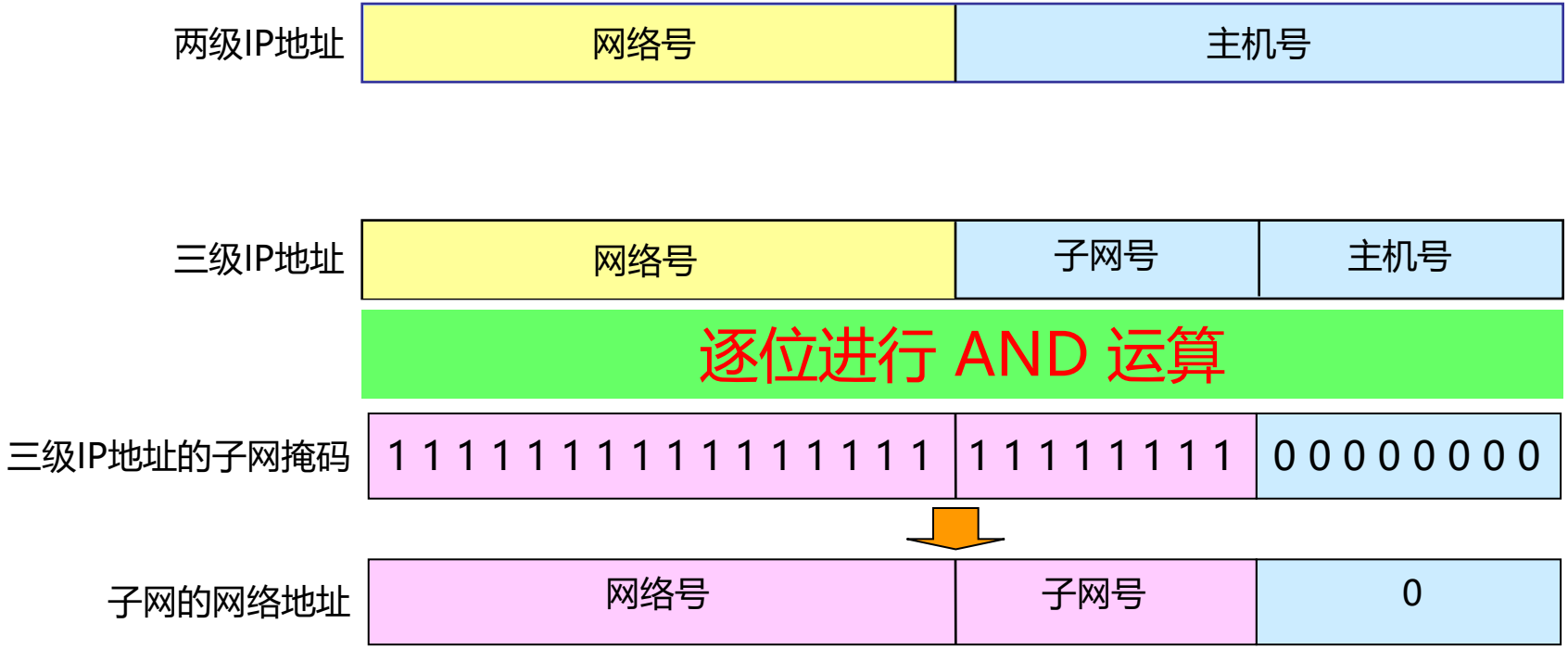
3.划分子网与构建超网

- 从IP数据报的首部无法看出源主机或目的主机所连接的网络是否进行了子网划分。
- 为了看到一个IP地址的网络号和主机号，于是使用了子网掩码 (subnet mask) 来作为辅助手段，以计算某一个IP地址的网络号。

3.划分子网与构建超网


网络号
主机号
两级IP地址
145 . 13 . 3 . 10
网络号
子网号
主机号
三级IP地址
145 . 13 . 3 . 10
子网号为 3 的网络的网络号
主机号
三级IP地址的子网掩码
11111111111111111111 11111111 00000000
子网的网络地址
145 . 13 . 3 0

3.划分子网与构建超网


两级IP地址
网络号
主机号
三级IP地址
网络号
子网号
主机号
逐位进行 AND 运算
三级IP地址的子网掩码
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
子网的网络地址
网络号
子网号
0

3.划分子网与构建超网

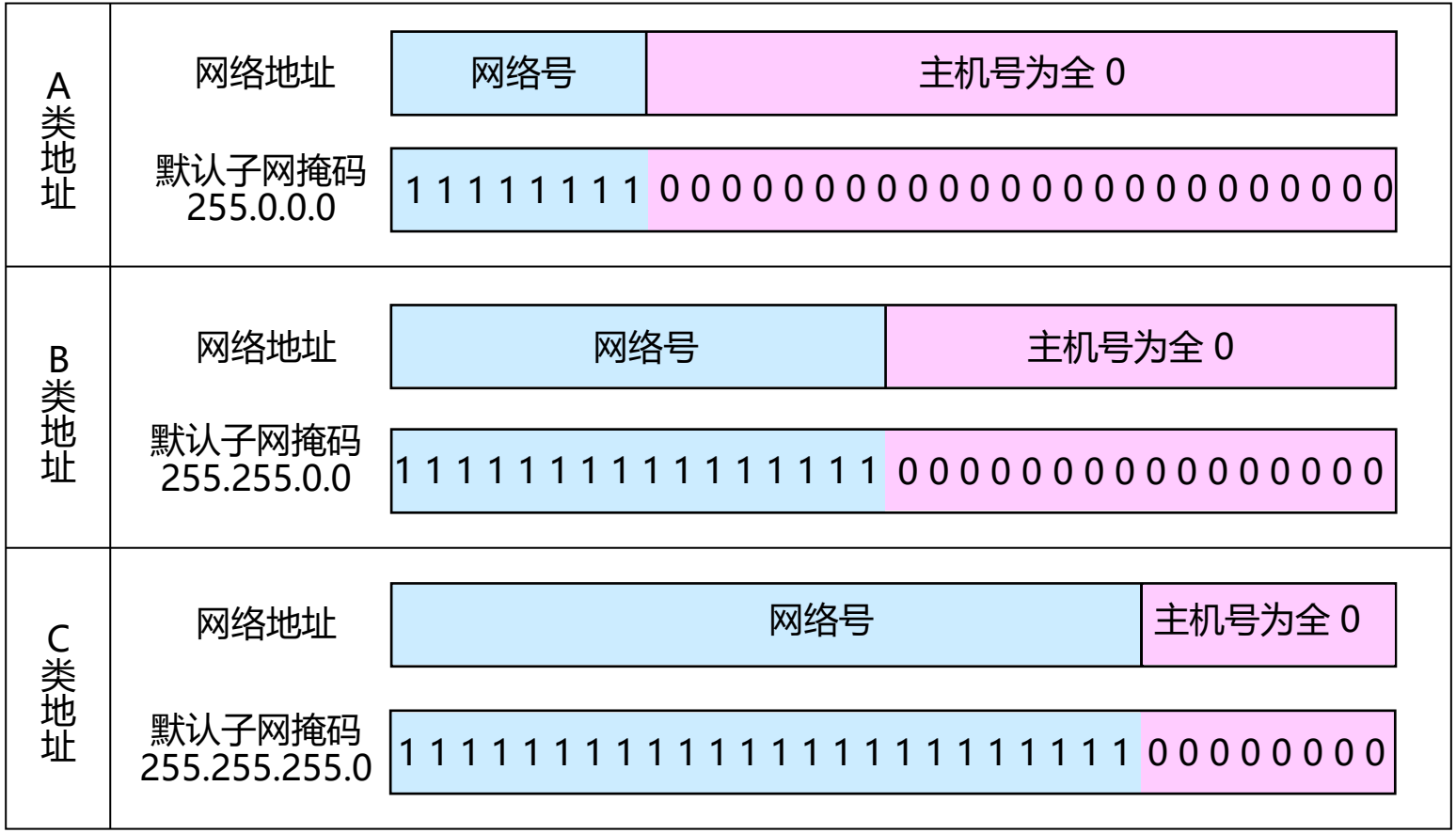
- 使用子网掩码之后，不管网络是否划分了子网，都可以通过子网掩码和IP地址逐位的“与”运算，方便的得出网络地址。
- 路由器在进行分组转发时，就不需要考虑是否划分子网，通过一个方法进行处理即可。

3.划分子网与构建超网

- 讨论：
  - IP地址：211.69.32.18是C类地址，在使用时没有进行子网划分。
  - 但在具体应用中，依然需要使用子网掩码，这是为什么？

- 不划分子网也要使用子网掩码的原因，就在于让路由器以一个方法进行分组处理，以便于以一种方法查找路由表。
- 子网掩码是一个网络或一个子网的重要属性。

3.划分子网与构建超网

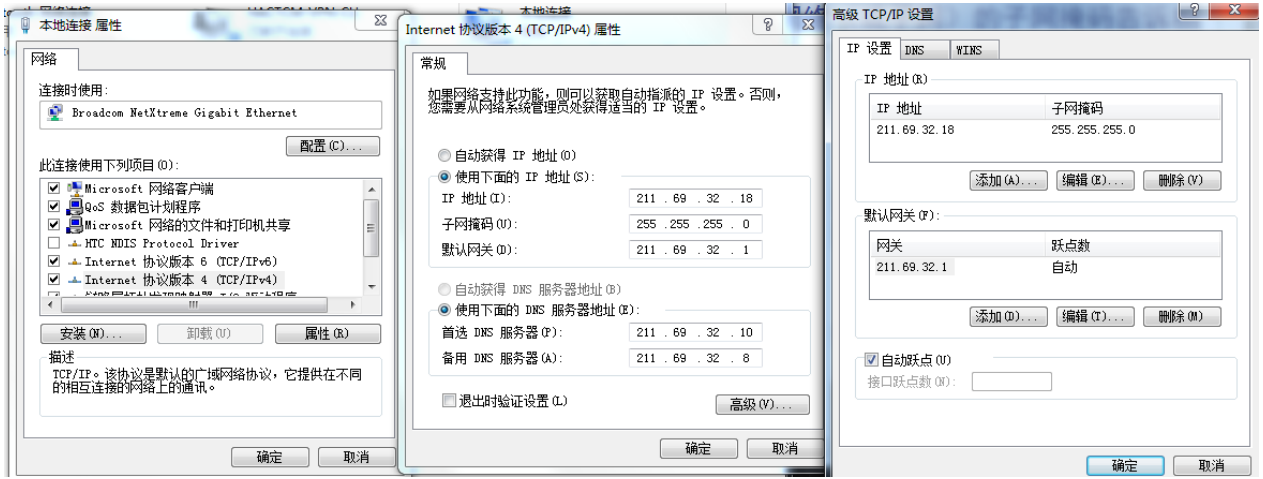

A类地址
网络地址
网络号
主机号为全 0
默认子网掩码
255.0.0.0
1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
B类地址
网络地址
网络号
主机号为全 0
默认子网掩码
255.255.0.0
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
C类地址
网络地址
网络号
主机号为全 0
默认子网掩码
255.255.255.0
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0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

3.划分子网与构建超网

- 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。

3.划分子网与构建超网

- 现场演示：
  - 让一个主机同时接入到两个网络上。
    - 为一个网络接口卡配置多个IP地址



3.划分子网与构建超网

- 计算：
  - 已知IP地址211.69.32.18，子网掩码是255.255.255.0，请计算其网络地址，并列出该网络内的所有IP地址。
  - 已知IP地址211.69.32.18，子网掩码是255.255.240.0，请计算其网络地址，并列出该网络内的所有IP地址。
  - 已知IP地址211.69.32.18，子网掩码是255.255.255.252，请计算机其网络地址，并列出该网络内的所有IP地址。

3.划分子网与构建超网

- 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。（讨论：两级IP地址下的分组转发）
- 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，这是因为网络地址取决于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。
- 在划分子网的情况下，分组转发算法也必须做相应的改动。

3.划分子网与构建超网

- 使用子网划分后，路由表必须包含以下三项内容：目的网络地址、子网掩码和下一跳地址。

192.168.183.0/24	OSPF	10	82	D	10.0.1.30	Vlanif1129
192.168.254.0/24	OSPF	10	42	D	10.0.1.30	Vlanif1129
192.168.255.0/24	OSPF	10	42	D	10.0.1.30	Vlanif1129
202.4.128.0/19	Static	60	0	RD	222.21.219.73	Vlanif1222
202.38.64.0/18	Static	60	0	RD	222.21.219.73	Vlanif1222
202.38.140.0/23	Static	60	0	RD	222.21.219.73	Vlanif1222
202.38.184.0/21	Static	60	0	RD	222.21.219.73	Vlanif1222
202.38.192.0/18	Static	60	0	RD	222.21.219.73	Vlanif1222
202.112.0.0/13	Static	60	0	RD	222.21.219.73	Vlanif1222
202.120.0.0/15	Static	60	0	RD	222.21.219.73	Vlanif1222
202.127.216.0/21	Static	60	0	RD	222.21.219.73	Vlanif1222
202.127.224.0/19	Static	60	0	RD	222.21.219.73	Vlanif1222
202.179.240.0/20	Static	60	0	RD	222.21.219.73	Vlanif1222

3.划分子网与构建超网

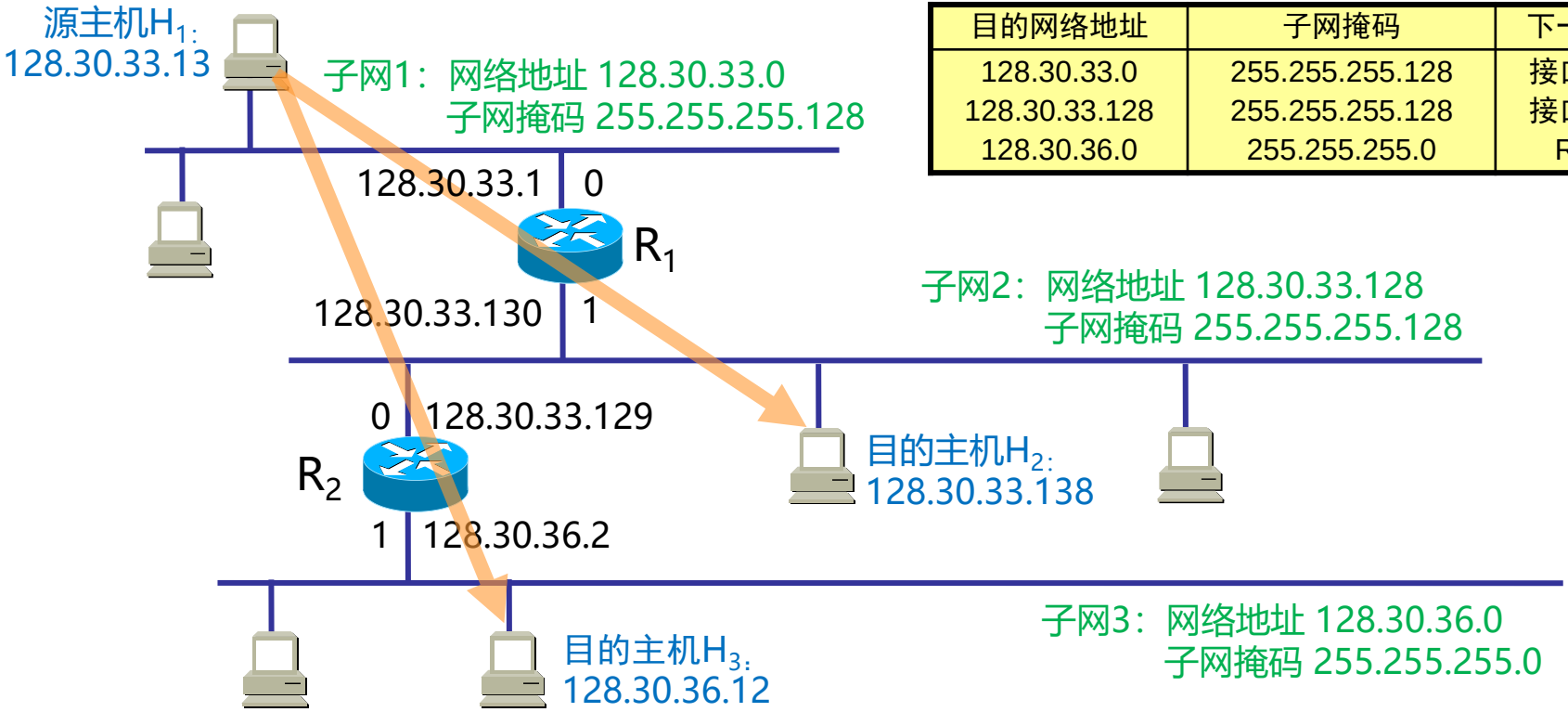
- 划分子网的情况下，路由器转发分组的算法：
  - ① 从收到的分组的首部提取目的IP地址D。
  - ② 先用各网络的子网掩码和D逐位相“与”，看是否和相应的网络地址匹配。若匹配则直接交付。否则就是间接交付，执行③ 。
  - ③ 若路由表中有目的地址为D的特定主机路由，则将分组传送给指明的下一跳路由器；否则，执行④ 。
  - ④ 对路由表中的每一行的子网掩码和D逐位相“与”，若其结果与该行的目的网络地址匹配，则将分组传送给该行指明的下一跳路由器；否则，执行⑤ 。
  - ⑤ 若路由表中有一个默认路由，则将分组传送给路由表中所指明的默认路由器；否则，执行⑥。
  - ⑥ 报告转发分组出错。

3.划分子网与构建超网

3.2使用子网掩码的分组转发过程

R_1的路由表（未给出默认路由器）

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



3.划分子网与构建超网

- 讨论：
  - H1向H2发送分组数据，R1在接收到H1发送的数据报后，查找路由表的过程。
  - H1向H3发送分组数据，R1、R2在接收到H1发送的数据报后，查找路由表的过程。
  - R2的路由表的具体内容是什么？

3.划分子网与构建超网

- 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。

3.划分子网与构建超网

- 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)。

3.划分子网与构建超网

- CIDR消除了传统的A类、B类和C类地址以及划分子网的概念，因而可以更加有效地分配IPv4的地址空间。
- CIDR使用各种长度的“网络前缀”(network-prefix)来代替分类地址中的网络号和子网号。
- IP地址从三级编址（使用子网掩码）又回到了两级编址。

IP地址 ::= {<网络前缀>, <主机号>}

3.划分子网与构建超网

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

IP地址 ::= {<网络前缀>, <主机号>}

3.划分子网与构建超网

- CIDR地址表示方法：
  - 211.69.32.0/20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 211.69.32.0。
    - 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”。
    - 211.69.32.0/20地址块的最小地址：211.69.32.0
    - 211.69.32.0/20地址块的最大地址：211.69.47.255

3.划分子网与构建超网

- CIDR地址表示方法：
  - 10.0.0.0/10表示的地址块共有2^{22}个地址。
  - 10.0.0.0/10可简写为10/10，也就是将点分十进制中低位连续的0省略。
  - 10.0.0.0/10相当于指出IP地址10.0.0.0的掩码是255.192.0.0，即

    11111111 11000000 00000000 00000000
  - 10.0.0.0/10可以表示为在网络前缀的后面加一个星号 * 的方法，即：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。

3.划分子网与构建超网

- CIDR地址表示方法讨论：

Network：211.69.32.0
Netmask：255.255.255.0 ＝ CIDR：211.69.32.0/24

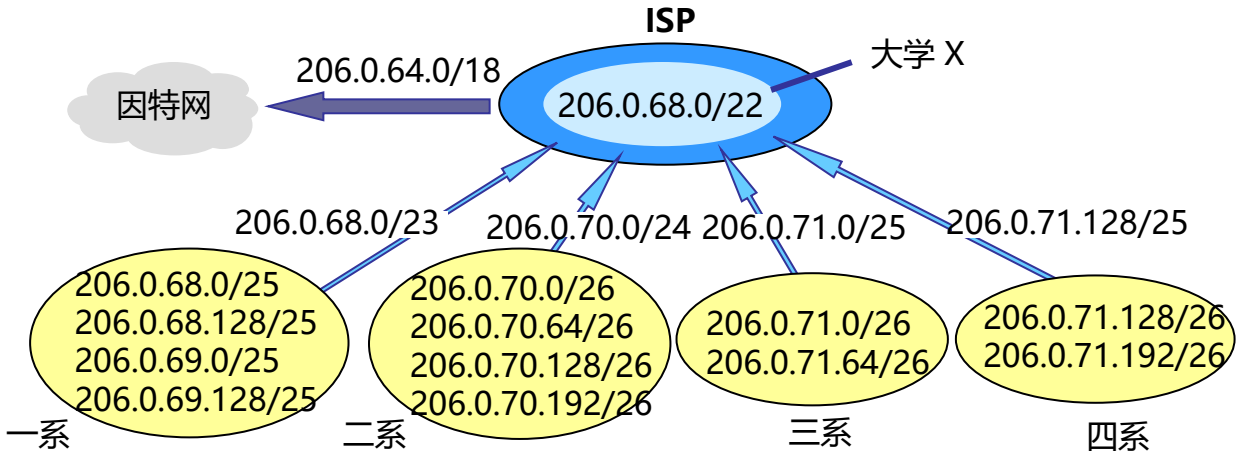
Network：211.69.32.0
Netmask：255.255.240.0 ＝ CIDR：211.69.32.0/?

3.划分子网与构建超网

- 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。
  - 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。

3.划分子网与构建超网

- 路由聚合（route aggregation）：



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

3.划分子网与构建超网

- 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。

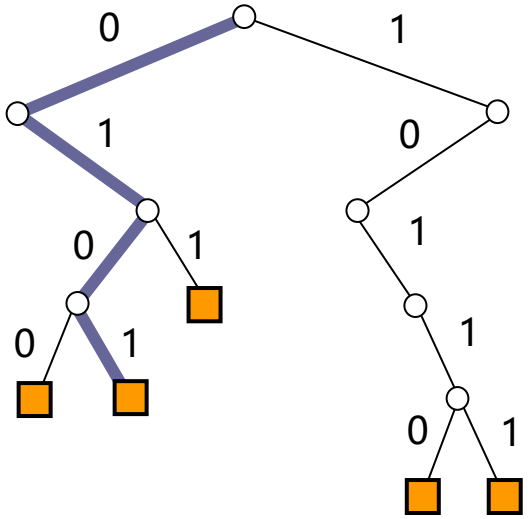
3.划分子网与构建超网

- 使用二叉线索查找路由表：
  - 当路由表的项目数很大时，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。
    - 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。

3.划分子网与构建超网

- 使用二叉线索查找路由表：

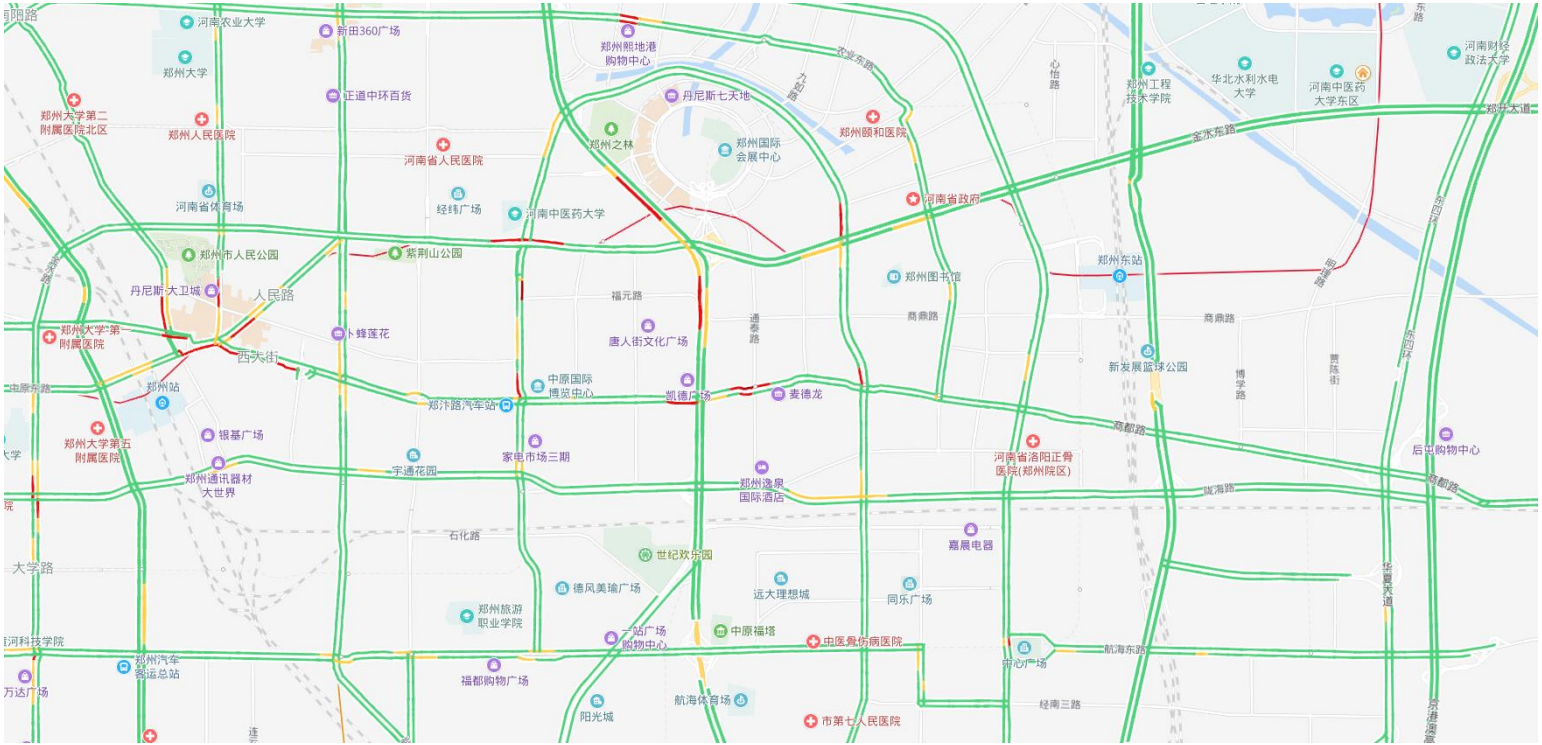
32 位的 IP 地址	唯一前缀
01000110 00000000 00000000 00000000	0100
01010110 00000000 00000000 00000000	0101
01100001 00000000 00000000 00000000	011
10110000 00000010 00000000 00000000	10110
10111011 00001010 00000000 00000000	10111



4.网际控制报文协议ICMP

- 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是IPv4协议簇中的一个子协议，用于在IP主机、路由器之间传递控制消息。
  - 控制消息是指网络通不通、主机是否可达、路由是否可用等网络本身的消息。
  - 控制消息虽然并不传送用户数据，但是对于用户数据的传递起着重要的作用。

4.网际控制报文协议ICMP



ICMP就如同地图中的路况信息

4.网际控制报文协议ICMP

- 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协议与ARP协议不同，ICMP依靠IP协议来完成任务，所以ICMP报文中要封装IP头部，组成IP数据报发送出去。
- ICMP一般并不用来在端系统之间传送数据，不被用户网络程序直接使用。
  - 端系统中，Ping和Traceroute等诊断网络的工具才会直接使用ICMP协议。

4.网际控制报文协议ICMP

- ICMP报告无法传送的数据报的错误，并帮助对这些错误进行疑难解答。
  - 例如，如果IPv4不能够将数据报传送到目标主机，则路由器或目标主机上的ICMP就会向主机发送ICMP的“无法到达目标”的消息。
- 在下列情况中，通常自动发送ICMP消息：
  - IP数据报无法访问目标。
  - IP路由器（网关）无法按当前的传输速率转发数据报。
  - IP路由器将发送主机重定向为使用到达目标的更佳路由。

4.网际控制报文协议ICMP

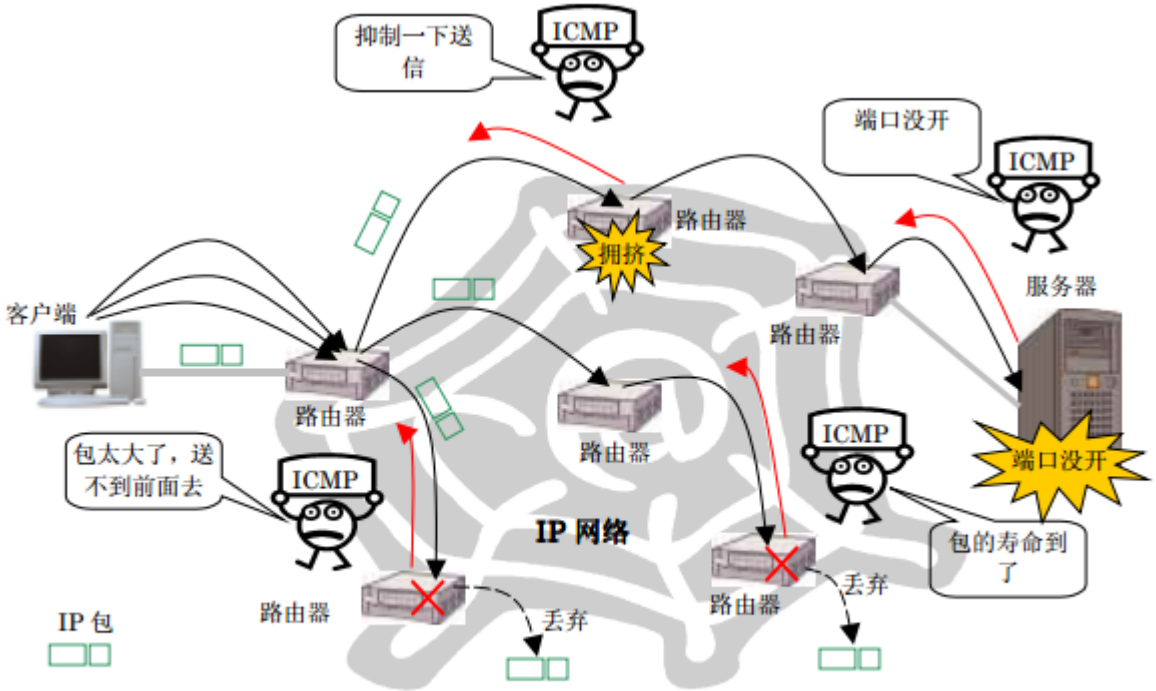
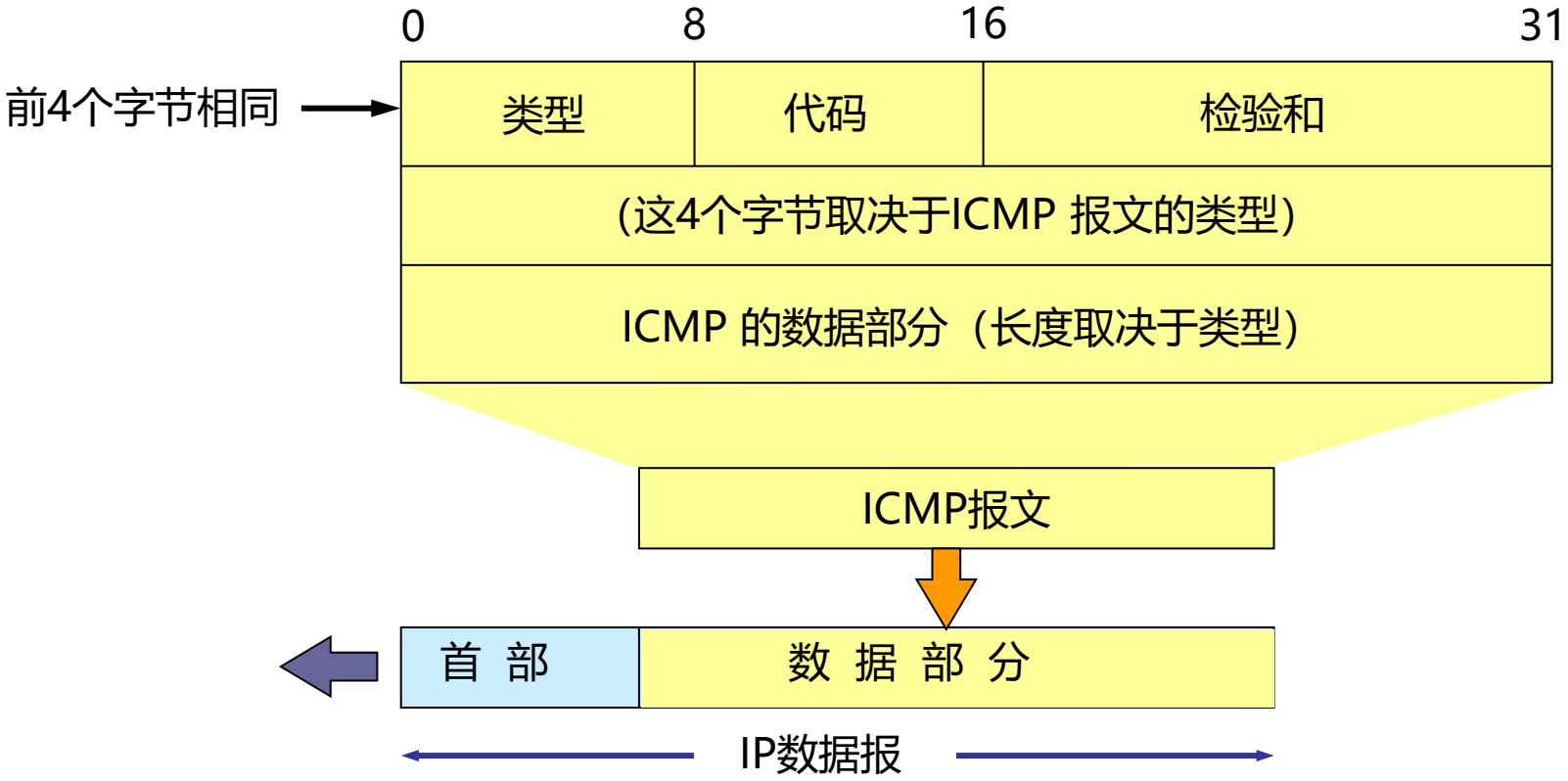
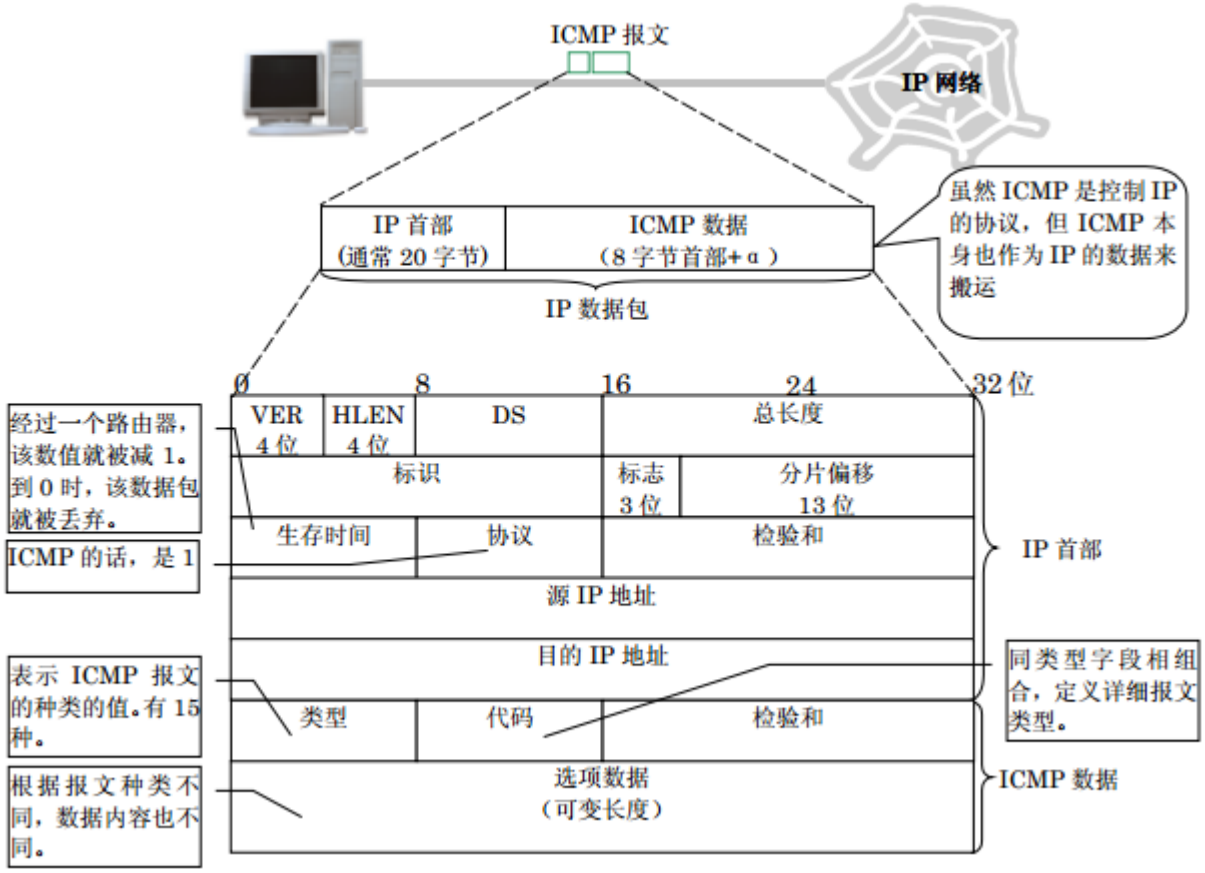


图 1 ICMP 是使 IP 通信平稳运行的辅助协议

4.网际控制报文协议ICMP



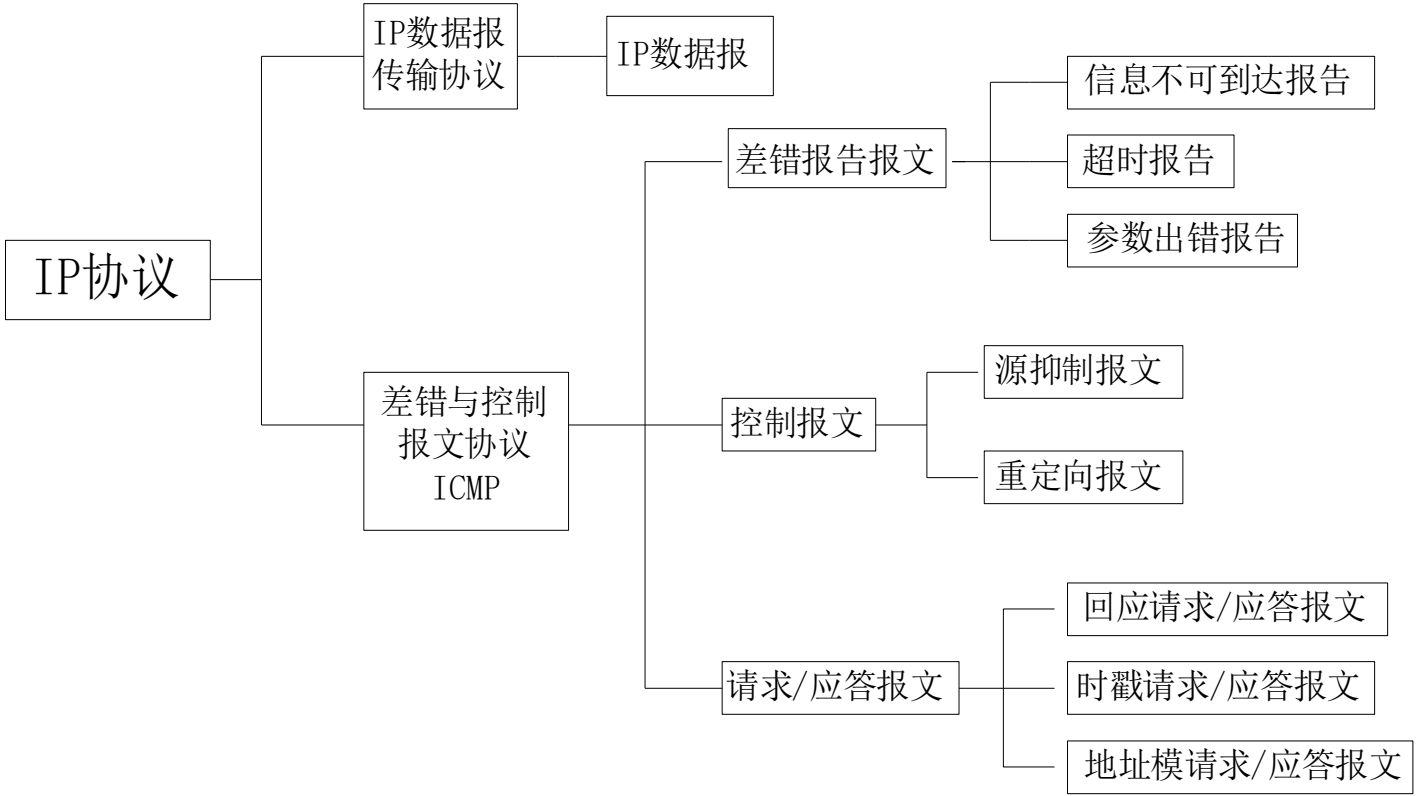
4.网际控制报文协议ICMP



4.网际控制报文协议ICMP

- 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。

4.网际控制报文协议ICMP



4.网际控制报文协议ICMP

- ICMP差错报告报文共有5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)

4.网际控制报文协议ICMP

- 不应发送ICMP差错报告报文的几种情况：
  - 对ICMP差错报告报文不再发送ICMP差错报告报文。
  - 对第一个分片的数据报片的所有后续数据报片都不发送ICMP差错报告报文。
  - 对具有多播地址的数据报都不发送ICMP差错报告报文。
  - 对具有特殊地址（如127.0.0.0或0.0.0.0）的数据报不发送 ICMP 差错报告报文。

4.网际控制报文协议ICMP

- ICMP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

- 停止使用的几种ICMP报有：
  - 信息请求与回答报文
  - 掩码地址请求和回答报文
  - 路由器询问和通告报文

类型	代码	名称	查询	差错
0	0	回应应答(Echo Reply)	√	
3		目的地不可达		√
	0	网路不可达		√
	1	主机不可达		√
	2	协议不可达		√
	3	端口不可达		√
	4	需要分片和不需要分片标记置位		√
	5	源路由失败		√
	6	目的网络未知		√
	7	目的主机未知		√
	8	源主机被隔离		√
	9	目的网络的通告被禁止		√
	10	目的主机的通信被禁止		√
	11	对请求的服务类型 ToS，目的网路不可达		√
	12	对请求的服务类型 ToS，目的主机不可达		√
	13	由于过滤,通信被强制禁止		√
	14	主机越权		√
	15	优先权中止生效		√
4	0	源抑制（Source Quench）		√
5		重定向		√
	0	为网络（子网）重定向数据报		√
	1	为主机重定向数据报		√
	2	为网络和服务类型重定向数据报		√
	3	为主机和服务类型重定向数据报		√

6	0	选择主机地址		
8	0	请求回应	√	
9	0	路由器通告	√	
10	0	路由器选择请求	√	
11		超时		
	0	传输中超出 TTL=0		√
	1	分片重组 TTL=0		√
12		参数问题		
	0	指定错误的指针(坏的 IP 头部)		√
	1	缺少需要的选项		√
	2	错误长度		
13	0	时间戳请求	√	
14	0	时间戳回复	√	
15	0	信息请求（已作废不用）	√	
16	0	信息回复（已作废不用）	√	
17	0	地址掩码请求	√	
18	0	地址掩码回复	√	
30		跟踪路由		
31		数据报会话错误		
32		移动主机重定向		
33		IPv6 你在哪里		
34		IPv6 我在这里		
35		移动注册请求		
36		移动注册回复		

4.网际控制报文协议ICMP

- Ping程序是ICMP协议的最常见应用程序。
  - 由Mike Muuss 编写。
  - 用来测试目的主机是否可到达。
  - 使用ICMP回应请求和回应应答报文实现。

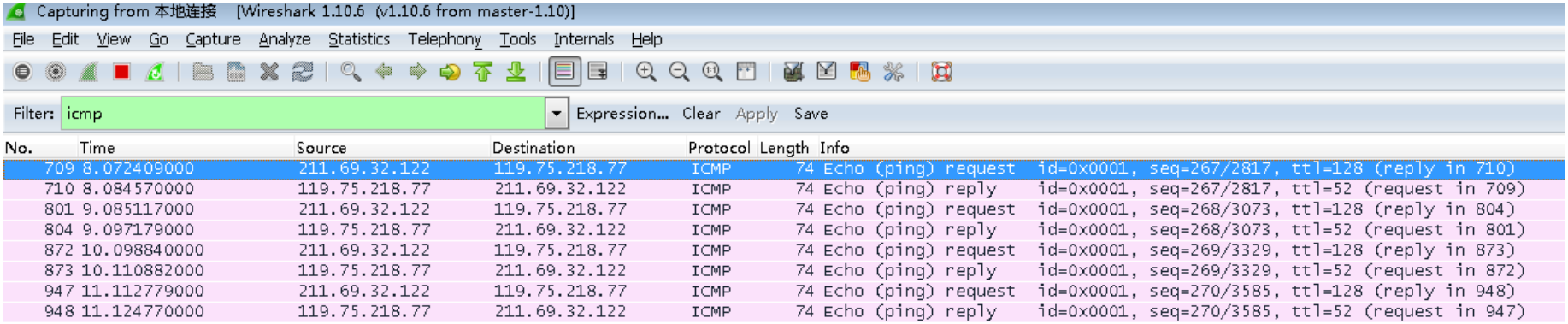
4.网际控制报文协议ICMP

- 当调用ping程序时，其发送一个包含ICMP回应请求的报文给目的地，然后等待一段很短的时间。
  - 如果没有收到应答，则重新传送请求。
  - 如果重传的请求仍没有收到应答（或收到一个ICMP目的不可达报文），ping报告该远程机器为不可达。
- 远端主机上的ICMP软件应答该回应请求报文。
- 按照协议只要收到回应请求，ICMP软件必须发送回应应答。

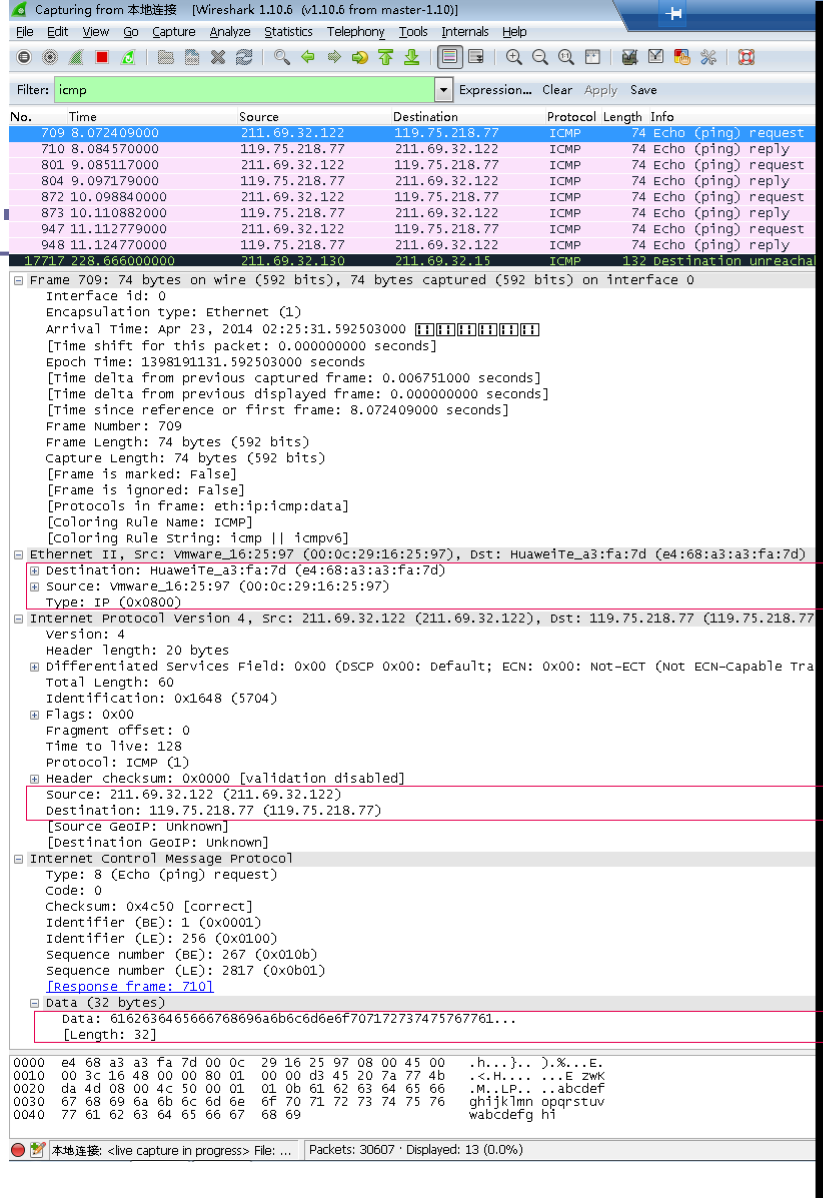
4.网际控制报文协议ICMP

执行Ping操作：ping www.baidu.com

4.网际控制报文协议ICMP



执行Ping操作：ping www.baidu.com
通过Wireshark捕获的ICMP数据报文


Capturing from 本地连接 [Wireshark 1.10.6 (v1.10.6 from master-1.10)]
File Edit View Go Capture Analyze Statistics Telephony Tools Internals Help
Filter: icmp
Expression... Clear Apply Save
No. Time Source Destination Protocol Length Info
709 8.072409000 211.69.32.122 119.75.218.77 ICMP 74 Echo (ping) request
710 8.084570000 119.75.218.77 211.69.32.122 ICMP 74 Echo (ping) reply
801 9.085117000 211.69.32.122 119.75.218.77 ICMP 74 Echo (ping) request
804 9.097179000 119.75.218.77 211.69.32.122 ICMP 74 Echo (ping) reply
872 10.098840000 211.69.32.122 119.75.218.77 ICMP 74 Echo (ping) request
873 10.110882000 119.75.218.77 211.69.32.122 ICMP 74 Echo (ping) reply
947 11.112779000 211.69.32.122 119.75.218.77 ICMP 74 Echo (ping) request
948 11.124770000 119.75.218.77 211.69.32.122 ICMP 74 Echo (ping) reply
17717 228.666000000 211.69.32.130 211.69.32.15 132 Destination unreachal
Frame 709: 74 bytes on wire (592 bits), 74 bytes captured (592 bits) on interface 0
Interface id: 0
Encapsulation type: Ethernet (1)
Arrival Time: Apr 23, 2014 02:25:31.592503000
[Time shift for this packet: 0.000000000 seconds]
Epoch Time: 1398191131.592503000 seconds
[Time delta from previous captured frame: 0.006751000 seconds]
[Time delta from previous displayed frame: 0.000000000 seconds]
[Time since reference or first frame: 8.072409000 seconds]
Frame Number: 709
Frame Length: 74 bytes (592 bits)
Capture Length: 74 bytes (592 bits)
[Frame is marked: False]
[Frame is ignored: False]
[Protocols in frame: eth:ip:icmp:data]
[Coloring Rule Name: ICMP]
[Coloring Rule String: icmp || icmpv6]
Ethernet II, Src: Vmware_16:25:97 (00:0c:29:16:25:97), Dst: HuaweiTe_a3:fa:7d (e4:68:a3:a3:fa:7d)
Destination: HuaweiTe_a3:fa:7d (e4:68:a3:a3:fa:7d)
Source: Vmware_16:25:97 (00:0c:29:16:25:97)
Type: IP (0x0800)
Internet Protocol Version 4, Src: 211.69.32.122 (211.69.32.122), Dst: 119.75.218.77 (119.75.218.77
Version: 4
Header length: 20 bytes
Differentiated Services Field: 0x00 (DSCP 0x00: Default; ECN: 0x00: Not-ECT (Not ECN-Capable Tra
Total Length: 60
Identification: 0x1648 (5704)
Flags: 0x00
Fragment offset: 0
Time to live: 128
Protocol: ICMP (1)
Header checksum: 0x0000 [validation disabled]
Source: 211.69.32.122 (211.69.32.122)
Destination: 119.75.218.77 (119.75.218.77)
[Source GeoIP: Unknown]
[Destination GeoIP: Unknown]
Internet Control Message Protocol
Type: 8 (Echo (ping) request)
Code: 0
Checksum: 0x4c50 [correct]
Identifier (BE): 1 (0x0001)
Identifier (LE): 256 (0x0100)
Sequence number (BE): 267 (0x010b)
Sequence number (LE): 2817 (0x0b01)
[Response frame: 710]
Data (32 bytes)
Data: 6162636465666768696a6b6c6d6e6f707172737475767761...
[Length: 32]
0000 e4 68 a3 a3 fa 7d 00 0c 29 16 25 97 08 00 45 00 .h...}..).%...E.
0010 00 3c 16 48 00 00 80 01 00 00 d3 45 20 7a 77 4b .<.H.... ...E zwK
0020 da 4d 08 00 4c 50 00 01 01 0b 61 62 63 64 65 66 .M..LP.. ..abcdef
0030 67 68 69 6a 6b 6c 6d 6e 6f 70 71 72 73 74 75 76 ghijklmn opqrstuv
0040 77 61 62 63 64 65 66 67 68 69 wabcdefg hi
本地连接: <live capture in progress> File: ...
Packets: 30607 · Displayed: 13 (0.0%)
Capturing from 本地连接 [Wireshark 1.10.6 (v1.10.6 from master-1.10)]
File Edit View Go Capture Analyze Statistics Telephony Tools Internals Help
Filter: icmp
Expression... Clear Apply Save
No. Time Source Destination Protocol Length Info
709 8.072409000 211.69.32.122 119.75.218.77 ICMP 74 Echo (ping) reques
710 8.084570000 119.75.218.77 211.69.32.122 ICMP 74 Echo (ping) reply
801 9.085117000 211.69.32.122 119.75.218.77 ICMP 74 Echo (ping) reques
804 9.097179000 119.75.218.77 211.69.32.122 ICMP 74 Echo (ping) reply
872 10.098840000 211.69.32.122 119.75.218.77 ICMP 74 Echo (ping) reques
873 10.110882000 119.75.218.77 211.69.32.122 ICMP 74 Echo (ping) reply
947 11.112779000 211.69.32.122 119.75.218.77 ICMP 74 Echo (ping) reques
948 11.124770000 119.75.218.77 211.69.32.122 ICMP 74 Echo (ping) reply
17717 228.666000000 211.69.32.130 211.69.32.15 ICMP 132 Destination unreac
Frame 710: 74 bytes on wire (592 bits), 74 bytes captured (592 bits) on interface 0
Interface id: 0
Encapsulation type: Ethernet (1)
Arrival Time: Apr 23, 2014 02:25:31.604664000
[Time shift for this packet: 0.000000000 seconds]
Epoch Time: 1398191131.604664000 seconds
[Time delta from previous captured frame: 0.012161000 seconds]
[Time delta from previous displayed frame: 0.012161000 seconds]
[Time since reference or first frame: 8.084570000 seconds]
Frame Number: 710
Frame Length: 74 bytes (592 bits)
Capture Length: 74 bytes (592 bits)
[Frame is marked: False]
[Frame is ignored: False]
[Protocols in frame: eth:ip:icmp:data]
[Coloring Rule Name: ICMP]
[Coloring Rule String: icmp || icmpv6]
Ethernet II, Src: HuaweiTe_a3:fa:7d (e4:68:a3:a3:fa:7d), Dst: Vmware_16:25:97 (00:0c:29:16:25:97
Destination: Vmware_16:25:97 (00:0c:29:16:25:97)
Source: HuaweiTe_a3:fa:7d (e4:68:a3:a3:fa:7d)
Type: IP (0x0800)
Internet Protocol Version 4, Src: 119.75.218.77 (119.75.218.77), Dst: 211.69.32.122 (211.69.32.1
Version: 4
Header length: 20 bytes
Differentiated Services Field: 0x00 (DSCP 0x00: Default; ECN: 0x00: Not-ECT (Not ECN-Capable T
Total Length: 60
Identification: 0x1648 (5704)
Flags: 0x00
Fragment offset: 0
Time to live: 52
Protocol: ICMP (1)
Header checksum: 0x2b21 [validation disabled]
Source: 119.75.218.77 (119.75.218.77)
Destination: 211.69.32.122 (211.69.32.122)
[Source GeoIP: Unknown]
[Destination GeoIP: Unknown]
Internet Control Message Protocol
Type: 0 (Echo (ping) reply)
Code: 0
Checksum: 0x5450 [correct]
Identifier (BE): 1 (0x0001)
Identifier (LE): 256 (0x0100)
Sequence number (BE): 267 (0x010b)
Sequence number (LE): 2817 (0x0b01)
[Request frame: 709]
[Response time: 12.161 ms]
Data (32 bytes)
Data: 6162636465666768696a6b6c6d6e6f707172737475767761...
0000 00 0c 29 16 25 97 e4 68 a3 a3 fa 7d 08 00 45 00 ..).%..h ...}..E.
0010 00 3c 16 48 00 00 34 01 2b 21 77 4b da 4d d3 45 .<.H..4. +!wK.M.E
0020 20 7a 00 00 54 50 00 01 01 0b 61 62 63 64 65 66 z..TP.. ..abcdef
0030 67 68 69 6a 6b 6c 6d 6e 6f 70 71 72 73 74 75 76 ghijklmn opqrstuv
0040 77 61 62 63 64 65 66 67 68 69 wabcdefg hi
本地连接: <live capture in progress> File: ...
Packets: 32657 · Displayed: 13 (0.0%)

4.网际控制报文协议ICMP

- windows的 tracert 和 linux/UNIX/router 的 traceroute 都用于探测数据包从源到目的经过路由的IP，但两者探测的方法却有差别。
- 默认情况下，tracert 是向目的地址发出ICMP请求回显数据包，而 traceroute 是向目的地址的某个端口（大于30000）发送UDP数据报。
  - 两者用于探测的数据类型不同。
  - 但他们也有一个共同点：都是通过设置发送包的TTL的值从1开始、逐次增1的方法来探测。

4.网际控制报文协议ICMP

4.4 tracert

Filter: icmp Expression... Clear Apply Save

No.	Time	Source	Destination	Protocol	Length	Info
776	7.100759000	211.69.32.122	119.75.217.56	ICMP	106	Echo (ping) request id=0x0001, seq=271/3841, ttl=1
777	7.103338000	211.69.32.1	211.69.32.122	ICMP	70	Time-to-live exceeded (Time to live exceeded in transit)
778	7.105919000	211.69.32.122	119.75.217.56	ICMP	106	Echo (ping) request id=0x0001, seq=272/4097, ttl=1
779	7.108839000	211.69.32.1	211.69.32.122	ICMP	70	Time-to-live exceeded (Time to live exceeded in transit)
780	7.110844000	211.69.32.122	119.75.217.56	ICMP	106	Echo (ping) request id=0x0001, seq=273/4353, ttl=1
781	7.113511000	211.69.32.1	211.69.32.122	ICMP	70	Time-to-live exceeded (Time to live exceeded in transit)
1311	12.971001000	211.69.32.122	119.75.217.56	ICMP	106	Echo (ping) request id=0x0001, seq=274/4609, ttl=2
1312	12.971850000	222.21.219.73	211.69.32.122	ICMP	70	Time-to-live exceeded (Time to live exceeded in transit)
1313	12.974177000	211.69.32.122	119.75.217.56	ICMP	106	Echo (ping) request id=0x0001, seq=275/4865, ttl=2
1314	12.975011000	222.21.219.73	211.69.32.122	ICMP	70	Time-to-live exceeded (Time to live exceeded in transit)
1315	12.977696000	211.69.32.122	119.75.217.56	ICMP	106	Echo (ping) request id=0x0001, seq=276/5121, ttl=2
1316	12.978554000	222.21.219.73	211.69.32.122	ICMP	70	Time-to-live exceeded (Time to live exceeded in transit)
1852	18.524059000	211.69.32.122	119.75.217.56	ICMP	106	Echo (ping) request id=0x0001, seq=277/5377, ttl=3
1853	18.524992000	210.43.145.209	211.69.32.122	ICMP	70	Time-to-live exceeded (Time to live exceeded in transit)
1854	18.527417000	211.69.32.122	119.75.217.56	ICMP	106	Echo (ping) request id=0x0001, seq=278/5633, ttl=3
1855	18.528302000	210.43.145.209	211.69.32.122	ICMP	70	Time-to-live exceeded (Time to live exceeded in transit)
1856	18.530291000	211.69.32.122	119.75.217.56	ICMP	106	Echo (ping) request id=0x0001, seq=279/5889, ttl=3
1857	18.531190000	210.43.145.209	211.69.32.122	ICMP	70	Time-to-live exceeded (Time to live exceeded in transit)
2488	24.077668000	211.69.32.122	119.75.217.56	ICMP	106	Echo (ping) request id=0x0001, seq=280/6145, ttl=4
2489	24.079077000	210.43.145.214	211.69.32.122	ICMP	70	Time-to-live exceeded (Time to live exceeded in transit)
2490	24.081326000	211.69.32.122	119.75.217.56	ICMP	106	Echo (ping) request id=0x0001, seq=281/6401, ttl=4
2491	24.082795000	210.43.145.214	211.69.32.122	ICMP	70	Time-to-live exceeded (Time to live exceeded in transit)
2492	24.085184000	211.69.32.122	119.75.217.56	ICMP	106	Echo (ping) request id=0x0001, seq=282/6657, ttl=4
2493	24.086592000	210.43.145.214	211.69.32.122	ICMP	70	Time-to-live exceeded (Time to live exceeded in transit)
3040	29.631898000	211.69.32.122	119.75.217.56	ICMP	106	Echo (ping) request id=0x0001, seq=283/6913, ttl=5
3041	29.633837000	101.4.116.25	211.69.32.122	ICMP	70	Time-to-live exceeded (Time to live exceeded in transit)
3042	29.636127000	211.69.32.122	119.75.217.56	ICMP	106	Echo (ping) request id=0x0001, seq=284/7169, ttl=5
3043	29.637911000	101.4.116.25	211.69.32.122	ICMP	70	Time-to-live exceeded (Time to live exceeded in transit)
3046	29.640450000	211.69.32.122	119.75.217.56	ICMP	106	Echo (ping) request id=0x0001, seq=285/7425, ttl=5
3047	29.642757000	101.4.116.25	211.69.32.122	ICMP	70	Time-to-live exceeded (Time to live exceeded in transit)
3657	35.184932000	211.69.32.122	119.75.217.56	ICMP	106	Echo (ping) request id=0x0001, seq=286/7681, ttl=6
3658	35.196870000	101.4.117.145	211.69.32.122	ICMP	110	Time-to-live exceeded (Time to live exceeded in transit)
3659	35.199353000	211.69.32.122	119.75.217.56	ICMP	106	Echo (ping) request id=0x0001, seq=287/7937, ttl=6
3660	35.204513000	101.4.117.145	211.69.32.122	ICMP	110	Time-to-live exceeded (Time to live exceeded in transit)
3661	35.206887000	211.69.32.122	119.75.217.56	ICMP	106	Echo (ping) request id=0x0001, seq=288/8193, ttl=6
3662	35.210351000	101.4.117.145	211.69.32.122	ICMP	110	Time-to-live exceeded (Time to live exceeded in transit)
4212	40.754192000	211.69.32.122	119.75.217.56	ICMP	106	Echo (ping) request id=0x0001, seq=289/8449, ttl=7
4213	40.765805000	101.4.112.1	211.69.32.122	ICMP	182	Time-to-live exceeded (Time to live exceeded in transit)
4214	40.768497000	211.69.32.122	119.75.217.56	ICMP	106	Echo (ping) request id=0x0001, seq=290/8705, ttl=7
4215	40.780166000	101.4.112.1	211.69.32.122	ICMP	182	Time-to-live exceeded (Time to live exceeded in transit)
4216	40.782562000	211.69.32.122	119.75.217.56	ICMP	106	Echo (ping) request id=0x0001, seq=291/8961, ttl=7

5.路由选择协议

- 网络层的主要功能是将分组从源节点路由到目的节点，而且在大多数计算机网络中，采用的是数据报分组交换方式，数据报分组需要经过多跳（Hop）才能到达目的地。
- 路由功能是一种数据报分组交换路径选择行为，是网络层的一种基本功能。
- 路由功能和日常旅行时选择最佳路线的道理是相通的，路由选择就是综合考虑多种因素，例如线路长度、信道带宽、线路稳定性等。

5.路由选择协议

- 路由（Routing）是把信息从源节点通过网络传送到目的节点的行为。
  - 简单的讲：路由就是指网络层设备从一个接口上收到数据包，根据数据包的目的地址进行定向，并转发到另一个接口的过程。
- 路由与桥接对比的主要区别在于：
  - 桥接发生在数据链路层，连接的是同一网络或同一子网的不同网段。
  - 路由发生在网络层，连接的是不同网络或不同子网。

5.路由选择协议

- 路由功能的实现是依靠路由器或路由交换机中的路由表进行的。
- 从路由算法的自适应性考虑：
  - 静态路由选择策略——即非自适应路由选择，其特点是简单和开销较小，但不能及时适应网络状态的变化。
    - 静态路由（Static Routing）
  - 动态路由选择策略——即自适应路由选择，其特点是能较好地适应网络状态的变化，但实现起来较为复杂，开销也比较大。
    - 动态路由（Dynamic Routing）

5.路由选择协议

- 静态路由：
  - 静态路由是手动配置的路由，主要应用在小型局域网中。
  - 静态路由的配置和管理都比较简单。
  - 静态路由的特点是：
    - 手动配置：静态路由需要管理员手动进行逐条配置。
    - 路由路径固定不变：静态路由不会随着网络的拓扑结构或链路的状态变化而变化，是静态固定的。
    - 不可通告性：静态路由信息是私有的，不会通告给其他路由器。
    - 单向性：静态路由仅为数据提供沿着下一跳的方向进行路由，不提供反向路由。

5.路由选择协议

- 静态路由：
  - 静态路由明确的指明了如何到达目的网络。
  - 在所有相同目的地址的路由记录中，静态路由的优先级是除“直连路由”外最高的。
  - 如果配置了到达某一网络或者某一结点的静态路由，则优先采用静态路由，只有当静态路由不可用时，才会考虑其他路由。
  - 静态路由一般适合比较简单的小型网络环境，因为在这样环境中，网络管理员易于清楚的了解网络的拓扑结构，能够设置正确的路由信息。

5.路由选择协议

- 动态路由：
  - 对于较为大型的广域网来说，由于拓扑结构复杂，且网络结构可能经常变动，通常会采用更加灵活、更具自动特性的动态路由。

5.路由选择协议

- 动态路由的特点：
  - 自动生成：
    - 路由器在启动了动态路由协议后，将会通告所直接连接的网络，则路由器间就会自动生成路由器直接连接的网络间的路由表项。
  - 自动调整：
    - 当网络结构发生改变，动态路由可以随时根据网络拓扑结构的变化调整路由表项，并删除无效的路由表项。
  - 自动通告：
    - 动态路由可以在相邻路由器上相互通告，以便及时反映拓扑结构的变化，生成新的动态路由表项。
  - 自动生成双向路由：
    - 路由器在生成某条路由的动态路由时会自动生成回程路由表项，也就是会同时双向路由表项。

5.路由选择协议

- 动态路由的特点：
  - 仅可生成网络间的路由表项：
    - 动态路由不能够生成到达具体节点或主机的动态路由表项。
  - 不同动态路由不兼容：
    - 动态路由根据所采用的算法的不同分为不同类型，如RIP、OSPF、EIGRP、IS-IS、BGP等。
    - 不同的动态路由协议主要适用的网络环境不一样，也是不兼容的，但是支持互相重发布。

5.路由选择协议

- 理想的路由算法：
  - 路由选择协议的核心是路由算法，即需要何种算法来获得路由表中的各项目。
  - 路由算法应具有的特点是：
    - 算法必须是正确的和完整的。
    - 算法在计算上应简单。
    - 算法应能适应通信量和网络拓扑的变化，这就是说，要有自适应性。
    - 算法应具有稳定性。
    - 算法应是公平的。
    - 算法应是最佳的。

5.路由选择协议

- 最佳路由：
  - 不存在一种绝对的最佳路由算法。
  - 所谓“最佳”只能是相对于某一种特定要求下得出的较合理选择。
  - 实际的路由选择算法，应尽可能接近于理想的算法。
  - 路由选择是个非常复杂的问题：
    - 路由选择是网络中的所有结点共同协调工作的结果。
    - 路由选择的环境往往是不断变化的，而这种变化有时无法事先知道。

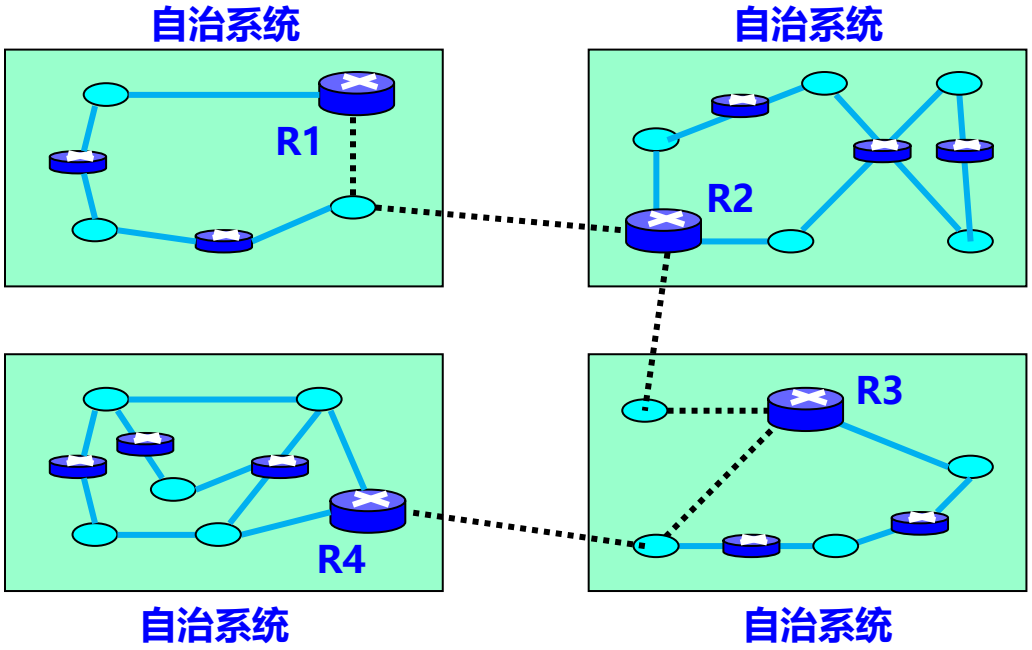
5.路由选择协议

- 分层次的路由选择协议：
  - 因特网采用分层次的路由选择协议。
  - 因特网的规模非常大。
    - 如果让所有的路由器知道所有的网络应怎样到达，则这种路由表将非常大，处理起来也太花时间。
    - 所有这些路由器之间交换路由信息所需的带宽就会使因特网的通信链路饱和。
  - 许多单位不愿意外界了解自己单位网络的布局细节和本部门所采用的路由选择协议（这属于本部门内部的事情），但同时还希望连接到因特网上。

5.路由选择协议

- 自治系统（Autonomous System，AS）：
  - 自治系统AS的定义：
    - 在单一的技术管理下的一组路由器，而这些路由器使用一种AS内部的路由选择协议和共同的度量以确定分组在该AS内的路由，同时还使用一种AS之间的路由选择协议用以确定分组在AS之间的路由。
  - 现在对自治系统AS的定义是强调下面的事实：
    - 尽管一个AS使用了多种内部路由选择协议和度量，但重要的是一个AS对其他AS表现出的是一个单一的和一致的路由选择策略。

5.路由选择协议

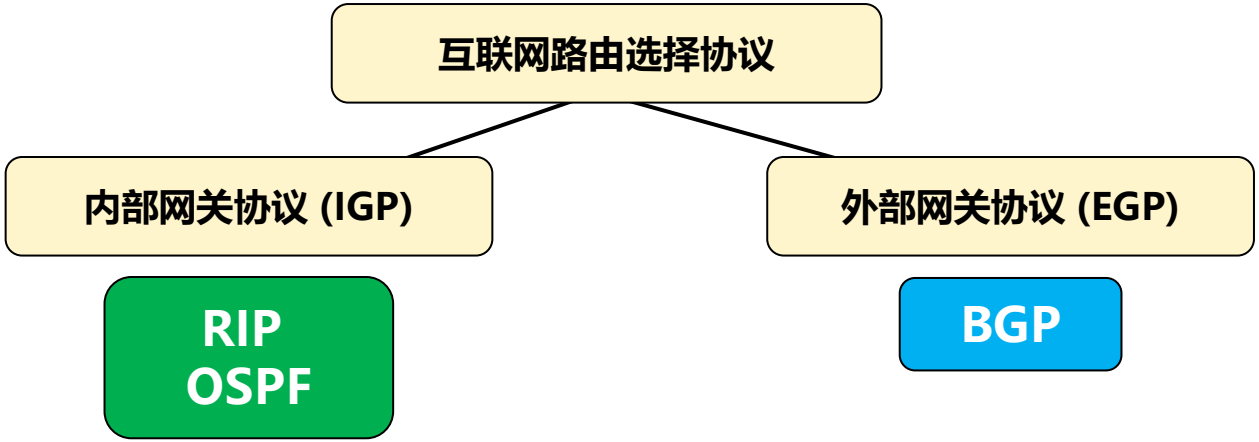


5.路由选择协议

- 因特网的两大类路由选择协议：
  - 内部网关协议 IGP (Interior Gateway Protocol)： IRP
    - 在一个自治系统内部使用的路由选择协议。
    - 目前这类路由选择协议使用得最多，如RIP和OSPF协议。
  - 外部网关协议 EGP (External Gateway Protocol)： ERP
    - 若源站和目的站处在不同的自治系统中，当数据报传到一个自治系统的边界时，就需要使用一种协议将路由选择信息传递到另一个自治系统中，这样的协议就是外部网关协议EGP。
    - 在外部网关协议中目前使用最多的是BGP-4。

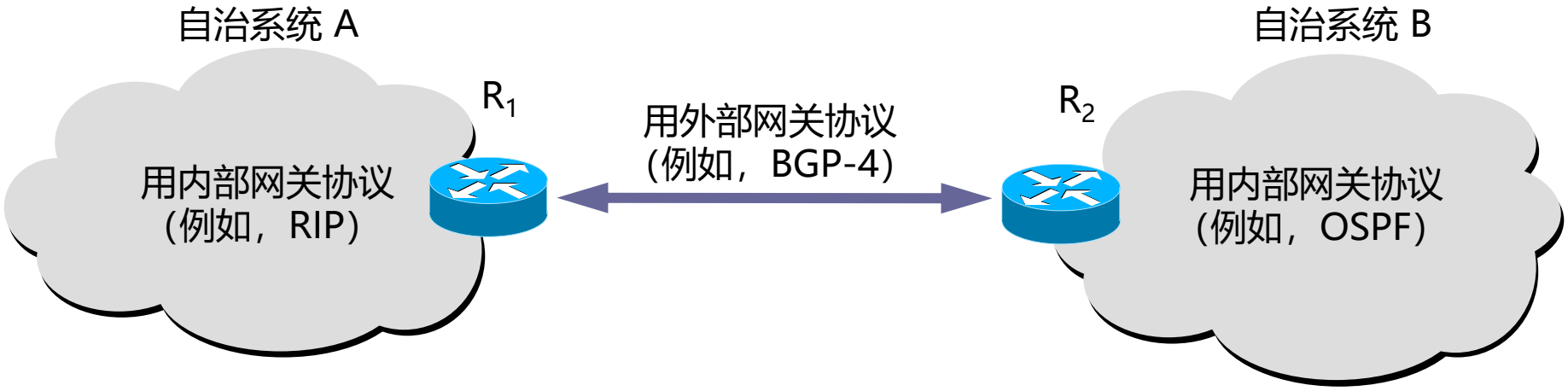
5.路由选择协议

- 因特网的两大类路由选择协议：

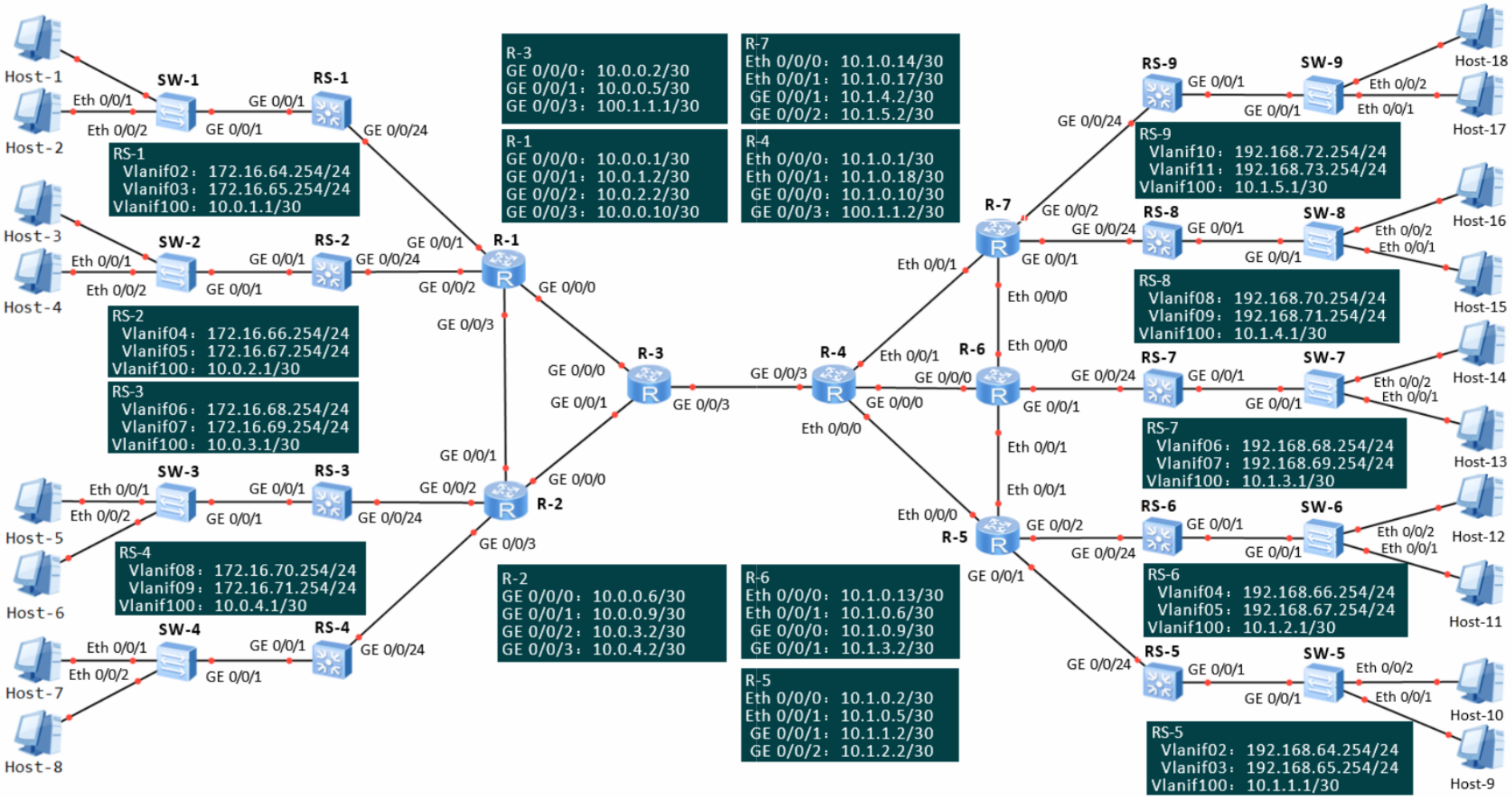


5.路由选择协议

- 因特网的两大类路由选择协议：



自治系统之间的路由选择叫做域间路由选择(interdomain routing)
自治系统内部的路由选择叫做域内路由选择(intradomain routing)


Host-1
Host-2
Host-3
Host-4
Host-5
Host-6
Host-7
Host-8
SW-1
SW-2
SW-3
SW-4
RS-1
RS-2
RS-3
RS-4
Eth 0/0/1
Eth 0/0/2
GE 0/0/1
GE 0/0/24
RS-1
Vlanif02：172.16.64.254/24
Vlanif03：172.16.65.254/24
Vlanif100：10.0.1.1/30
RS-2
Vlanif04：172.16.66.254/24
Vlanif05：172.16.67.254/24
Vlanif100：10.0.2.1/30
RS-3
Vlanif06：172.16.68.254/24
Vlanif07：172.16.69.254/24
Vlanif100：10.0.3.1/30
RS-4
Vlanif08：172.16.70.254/24
Vlanif09：172.16.71.254/24
Vlanif100：10.0.4.1/30
R-1
R-2
R-3
R-4
R-5
R-6
R-7
GE 0/0/0
GE 0/0/1
GE 0/0/2
GE 0/0/3
R-3
GE 0/0/0：10.0.0.2/30
GE 0/0/1：10.0.0.5/30
GE 0/0/3：100.1.1.1/30
R-1
GE 0/0/0：10.0.0.1/30
GE 0/0/1：10.0.1.2/30
GE 0/0/2：10.0.2.2/30
GE 0/0/3：10.0.0.10/30
R-2
GE 0/0/0：10.0.0.6/30
GE 0/0/1：10.0.0.9/30
GE 0/0/2：10.0.3.2/30
GE 0/0/3：10.0.4.2/30
R-7
Eth 0/0/0：10.1.0.14/30
Eth 0/0/1：10.1.0.17/30
GE 0/0/1：10.1.4.2/30
GE 0/0/2：10.1.5.2/30
R-4
Eth 0/0/0：10.1.0.1/30
Eth 0/0/1：10.1.0.18/30
GE 0/0/0：10.1.0.10/30
GE 0/0/3：100.1.1.2/30
R-6
Eth 0/0/0：10.1.0.13/30
Eth 0/0/1：10.1.0.6/30
GE 0/0/0：10.1.0.9/30
GE 0/0/1：10.1.3.2/30
R-5
Eth 0/0/0：10.1.0.2/30
Eth 0/0/1：10.1.0.5/30
GE 0/0/1：10.1.1.2/30
GE 0/0/2：10.1.2.2/30
RS-9
RS-8
RS-7
RS-6
RS-5
SW-9
SW-8
SW-7
SW-6
SW-5
RS-9
Vlanif10：192.168.72.254/24
Vlanif11：192.168.73.254/24
Vlanif100：10.1.5.1/30
RS-8
Vlanif08：192.168.70.254/24
Vlanif09：192.168.71.254/24
Vlanif100：10.1.4.1/30
RS-7
Vlanif06：192.168.68.254/24
Vlanif07：192.168.69.254/24
Vlanif100：10.1.3.1/30
RS-6
Vlanif04：192.168.66.254/24
Vlanif05：192.168.67.254/24
Vlanif100：10.1.2.1/30
RS-5
Vlanif02：192.168.64.254/24
Vlanif03：192.168.65.254/24
Vlanif100：10.1.1.1/30
Host-18
Host-17
Host-16
Host-15
Host-14
Host-13
Host-12
Host-11
Host-10
Host-9

5.路由选择协议

- 路由信息协议 (Routing Information Protocol，RIP) 是内部网关协议IGP中最先得到广泛使用的协议。
  - RIP是一种分布式的基于距离向量的路由选择协议，是因特网的标准协议。
  - RIP要求网络中的每一个路由器都要维护从它自己到其他每一个目的网络的距离记录。

5.路由选择协议

- 距离：
  - 从一路由器到直接连接的网络的距离定义为1。从一个路由器到非直接连接的网络的距离定义为所经过的路由器数加1。
  - RIP中的“距离”也称为“跳数”(hop count)，因为每经过一个路由器，跳数就加1。
  - RIP中的“距离”实际上指的是“最短距离”，RIP认为一个好的路由就是它通过的路由器的数目少，即“距离短”。
  - RIP允许一条路径最多只能包含15个路由器，“距离”的最大值为16时即相当于不可达。
  - RIP不能在两个网络之间同时使用多条路由。
  - RIP选择一个具有最少路由器的路由（即最短路由），哪怕还存在另一条高速(低时延)但路由器较多的路由。

5.路由选择协议

- RIP协议的特点：
  - 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，例如每隔30秒，然后路由器根据收到的路由信息更新路由表。
  - 当网络拓扑发生变化时，路由器也及时向相邻路由器通告拓扑变化后的路由信息。

5.路由选择协议

- 路由表的建立过程：
  - 首先：路由器在刚刚开始工作时，路由表是空的。
  - 然后：路由器得出到直接连接的网络的距离，距离定义为1。
  - 接着：每一个路由器也只和数目非常有限的相邻路由器交换并更新路由信息。
  - 以后：经过若干次更新后，所有的路由器最终都会知道到达本自治系统中任何一个网络的最短距离和下一跳路由器的地址。

  - RIP的收敛(convergence)过程较快，即在自治系统中所有的结点都得到正确的路由选择信息的过程较短。

5.路由选择协议

- 距离向量算法：
  - 收到相邻路由器（其地址为 X）的一个RIP报文：
    - 先修改此RIP报文中的所有项目：把“下一跳”字段中的地址都改为X，并把所有的“距离”字段的值加1。
    - 对修改后的RIP报文中的每一个项目，重复以下步骤：
    - 若项目中的目的网络不在路由表中，则把该项目加到路由表中。
      - 否则若下一跳字段给出的路由器地址是同样的，则把收到的项目替换原路由表中的项目。
      - 否则若收到项目中的距离小于路由表中的距离，则进行更新。
      - 否则什么也不做。
    - 若3分钟还没有收到相邻路由器的更新路由表，则把此相邻路由器记为不可达路由器，即将距离置为16（距离为16表示不可达）。

5.路由选择协议

- 路由器之间交换信息：
  - RIP协议让互联网中的所有路由器都和自己的相邻路由器不断交换路由信息，并不断更新其路由表，使得从每一个路由器到每一个目的网络的路由都是最短的（即跳数最少）。
  - 虽然所有的路由器最终都拥有了整个自治系统的全局路由信息，但由于每一个路由器的位置不同，它们的路由表也是不同的。

5.路由选择协议

已知路由器 R_6 有表a 所示的路由表。现在收到相邻路由器 R_4 发来的路由更新信息，如表b所示。试更新路由器 R_6 的路由表。

表a 路由器 R_6 的路由表

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

1

表b R4 发来的路由更新信息

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

2

表d 路由器 R_6 更新后的路由表

目的网络	距离	下一跳路由器
Net1	4	R_4
Net2	5	R_4
Net3	2	R_4
...	...	...

4

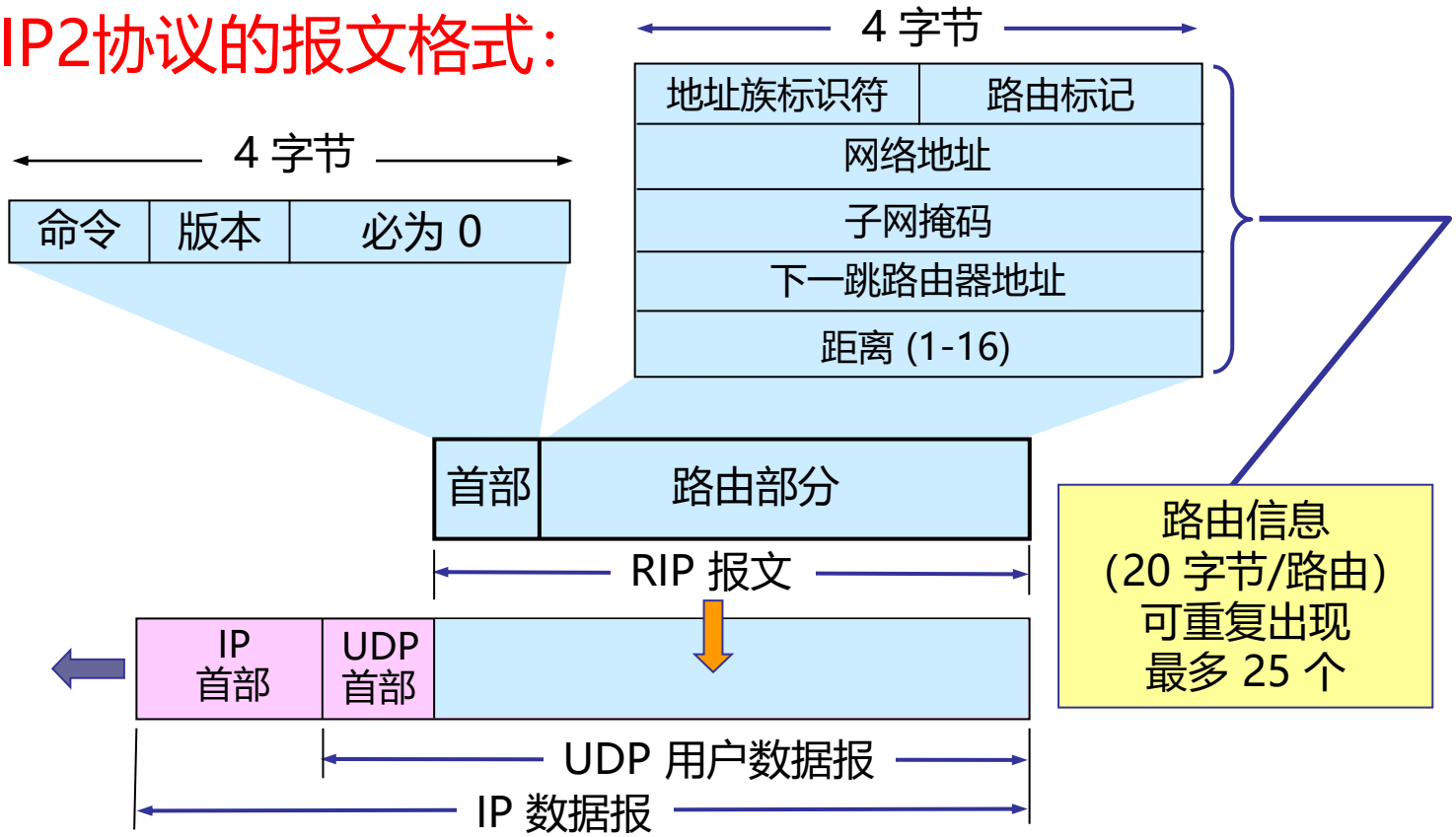
表c 修改后的表b

目的网络	距离	下一跳路由器
Net1	4	R_4
Net2	5	R_4
Net3	2	R_4

3

5.路由选择协议

- RIP2协议的报文格式：



5.路由选择协议

- RIP2 报文：
  - RIP2 报文由首部和路由部分组成。
  - RIP2 报文中的路由部分由若干个路由信息组成。每个路由信息需要用 20 个字节。地址族标识符（又称为地址类别）字段用来标志所使用的地址协议。
  - 路由标记填入自治系统的号码，这是考虑使 RIP 有可能收到本自治系统以外的路由选择信息。
  - 再后面指出某个网络地址、该网络的子网掩码、下一跳路由器地址以及到此网络的距离。

5.路由选择协议

- RIP2 报文：
  - 一个 RIP 报文最多可包括 25 个路由，因而 RIP 报文的最大长度是 4+20 x25=504 字节。如超过，必须再用一个 RIP 报文来传送。
  - RIP2 具有简单的鉴别功能。
    - 若使用鉴别功能，则将原来写入第一个路由信息（20 个字节）的位置用作鉴别。
    - 在鉴别数据之后才写入路由信息，但这时最多只能再放入 24 个路由信息。

5.路由选择协议

- 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，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。

好消息传播得快，坏消息传播得慢

5.路由选择协议

- 开放最短路径优先（Open Shortest Path First，OSPF）协议是为了克服RIP的缺点，在1989年开发出来的。
  - 开放：表明OSPF协议不是受某一家厂商控制，而是公开发表的。
  - 最短路径优先：因为使用了Dijkstra提出的最短路径算法SPF。
  - OSPF只是一个协议的名字，它并不表示其他的路由选择协议不是“最短路径优先”。
    - 实际上，所有的在自治系统内部使用的路由选择协议都是要寻找最短路径的。

5.路由选择协议

- OSPF协议最主要的特征就是使用分布式的链路状态协议（Link State Protocol），而不是像RIP那样的距离向量协议。
- OSPF的三个要点：
  - 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。
  - 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。“链路状态”就是说明本路由器都和哪些路由器相邻，以及该链路的“度量”(metric)。
  - 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。

5.路由选择协议

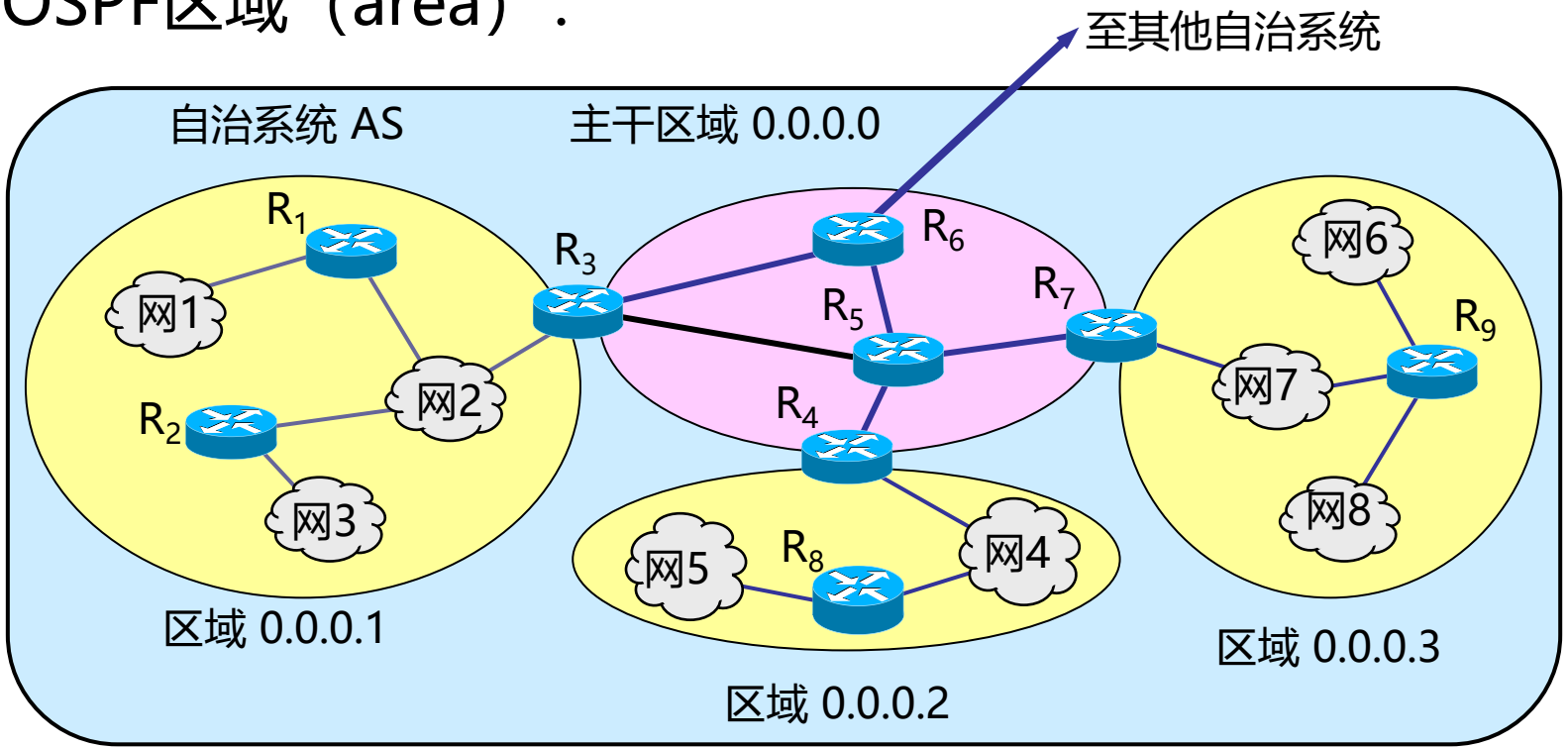
- 链路状态数据库（Link-state Database）：
  - 由于各路由器之间频繁地交换链路状态信息，因此所有的路由器最终都能建立一个链路状态数据库。
  - 这个数据库实际上就是全网的拓扑结构图，它在全网范围内是一致的（这称为链路状态数据库的同步）。
  - OSPF的链路状态数据库能较快地进行更新，使各个路由器能及时更新其路由表。
  - OSPF的更新过程收敛得快是其重要优点。

5.路由选择协议

- OSPF区域（area）：
  - 为了使OSPF能够用于规模很大的网络，OSPF将一个自治系统再划分为若干个更小的范围，叫作区域。
  - 每一个区域都有一个32位的区域标识符（用点分十进制表示）。
  - 区域也不能太大，在一个区域内的路由器最好不超过200个。

5.路由选择协议

- OSPF区域（area）：

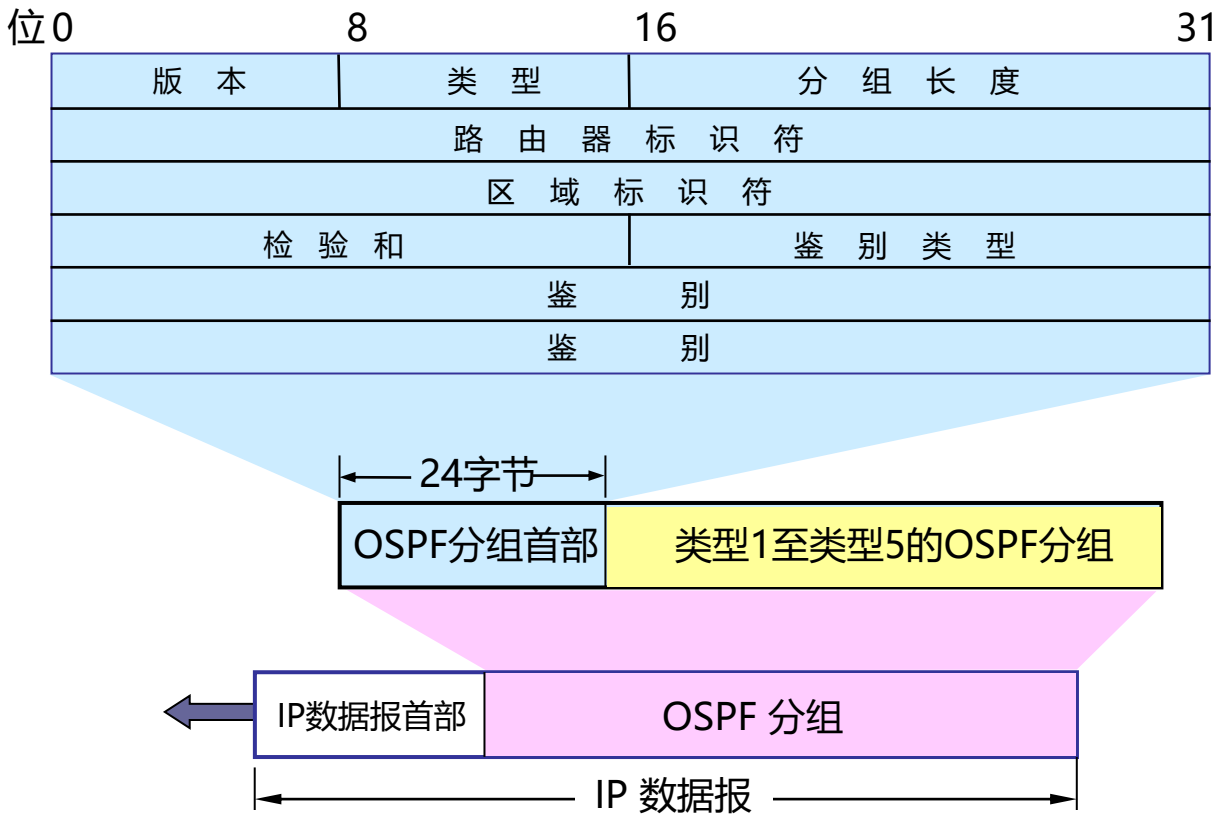


5.路由选择协议

- OSPF数据报：
  - OSPF不用UDP而是直接用IP数据报传送。
  - OSPF构成的数据报很短：
    - 可减少路由信息的通信量。
    - 不必将长的数据报分片传送。分片传送的数据报只要丢失一个，就无法组装成原来的数据报，而整个数据报就必须重传。

5.路由选择协议

- OSPF数据报：



5.路由选择协议

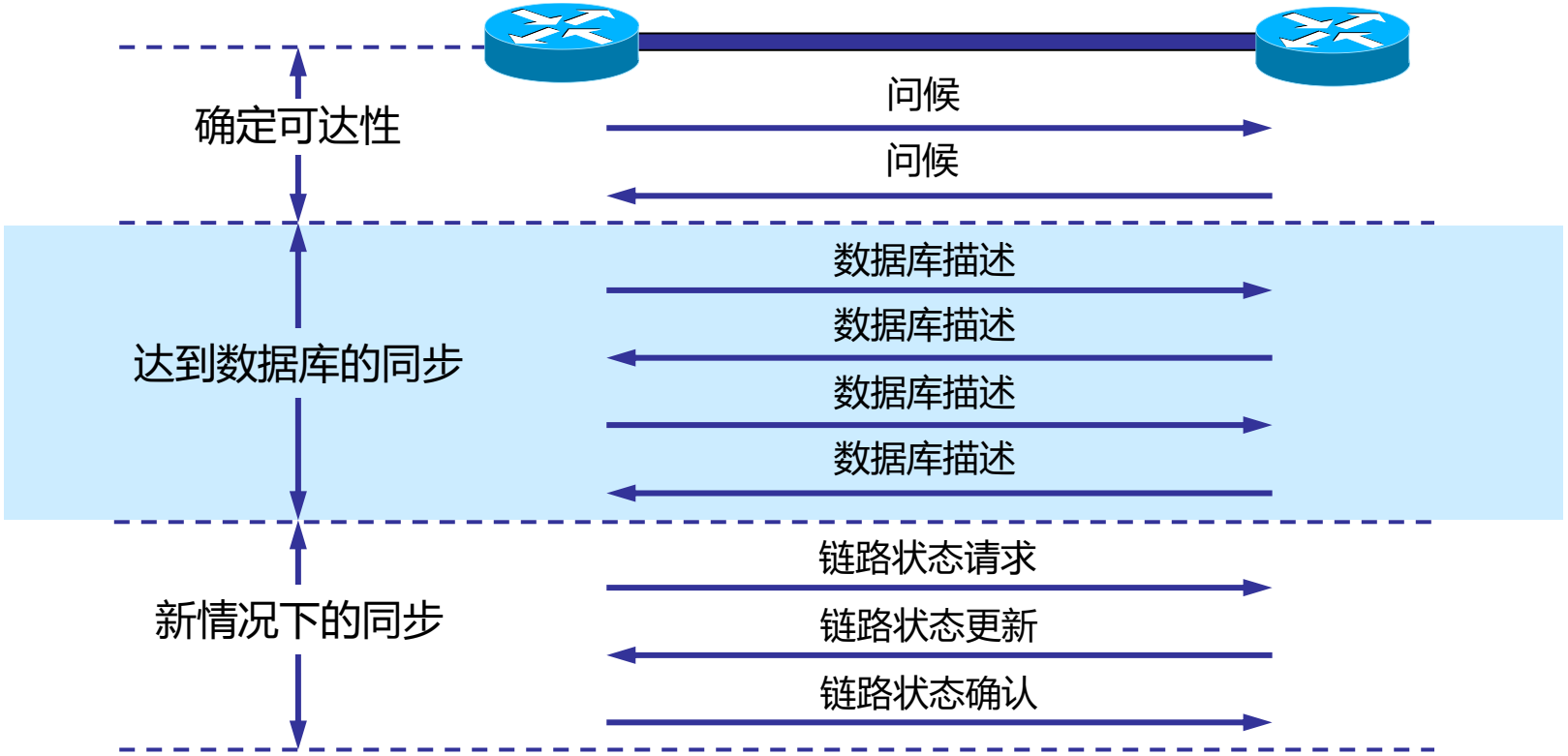
- OSPF数据报：
  - OSPF对不同的链路可根据IP分组的不同服务类型TOS而设置成不同的代价。因此，OSPF对于不同类型的业务可计算出不同的路由。
  - 如果到同一个目的网络有多条相同代价的路径，那么可以将通信量分配给这几条路径。这叫作多路径间的负载平衡。
  - 所有在OSPF路由器之间交换的分组都具有鉴别的功能。
  - 支持可变长度的子网划分和无分类编址CIDR。
  - 每一个链路状态都带上一个32位的序号，序号越大状态就越新。

5.路由选择协议

- OSPF的五种分组类型：
  - 类型1：问候(Hello)分组。
  - 类型2：数据库描述(Database Description)分组。
  - 类型3：链路状态请求(Link State Request)分组。
  - 类型4：链路状态更新(Link State Update)分组，用洪泛法对全网更新链路状态。
  - 类型5：链路状态确认(Link State Acknowledgment)分组。

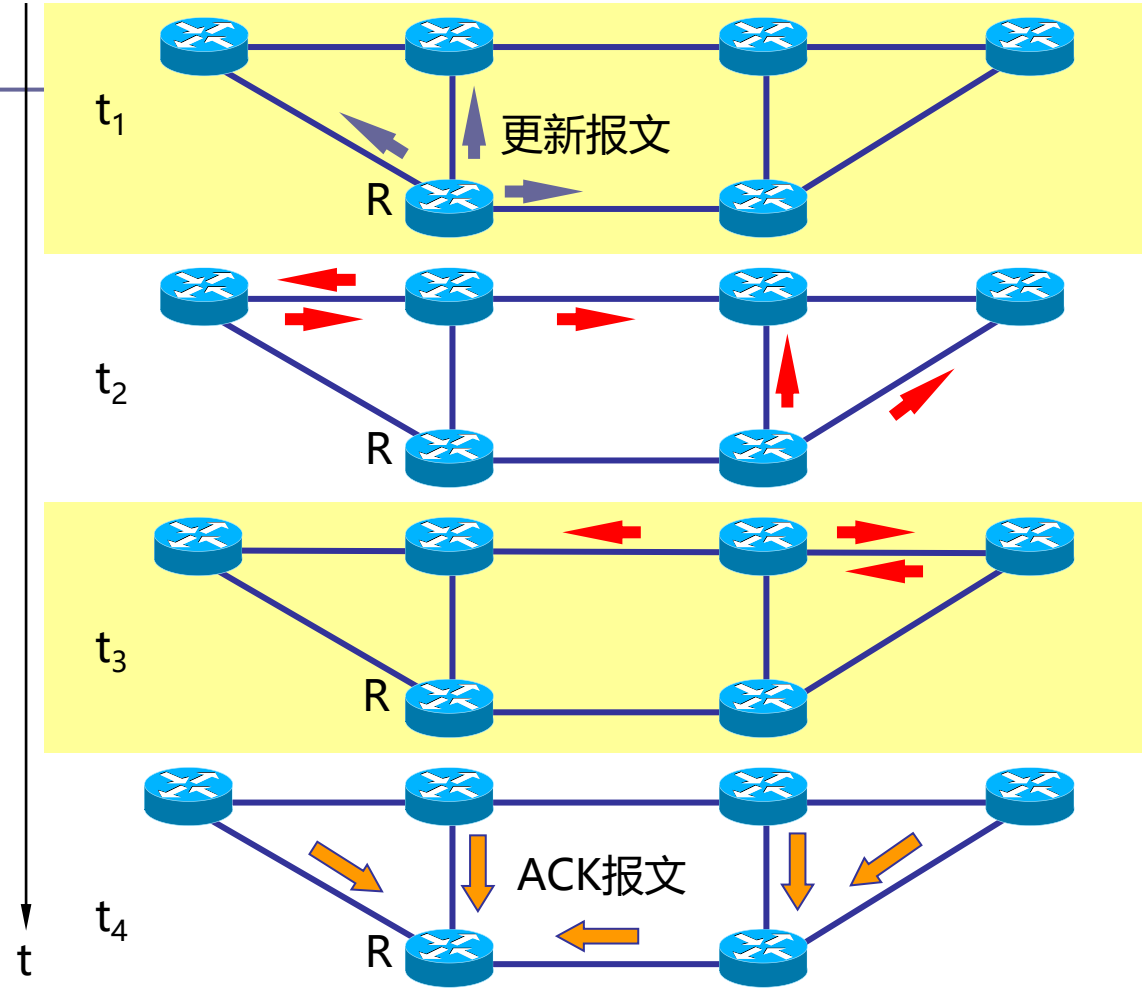
5.路由选择协议

- OSPF的五种分组类型：



5.路由选择协议

- OSPF使用的是可靠的洪泛法：



5.路由选择协议

- OSPF的其他特点：
  - OSPF规定每隔一段时间，如30分钟，要刷新一次数据库中的链路状态。
  - 由于一个路由器的链路状态只涉及到与相邻路由器的连通状态，因而与整个互联网的规模并无直接关系。因此当互联网规模很大时，OSPF协议要比距离向量协议RIP好得多。
  - OSPF没有“坏消息传播得慢”的问题，据统计，其响应网络变化的时间小于100ms。
  - 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。

5.路由选择协议

- 边界网关协议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2006年1月发表的BGP-4。
  - BGP第4个版本，即RFC 4271 ~ 4278。
  - 为了简单起见，通常将BGP-4简写为BGP。

5.路由选择协议

- 为什么在不同的AS之间不能够使用RIP或OSPF?
  - 因特网的规模太大，使得自治系统之间路由选择非常困难。对于自治系统之间的路由选择，要寻找最佳路由是很不现实的。
  - 当一条路径通过几个不同 AS 时，要想对这样的路径计算出有意义的代价是不太可能的。
  - 比较合理的做法是在 AS 之间交换“可达性”信息。
- 自治系统之间的路由选择必须考虑有关策略。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。

5.路由选择协议

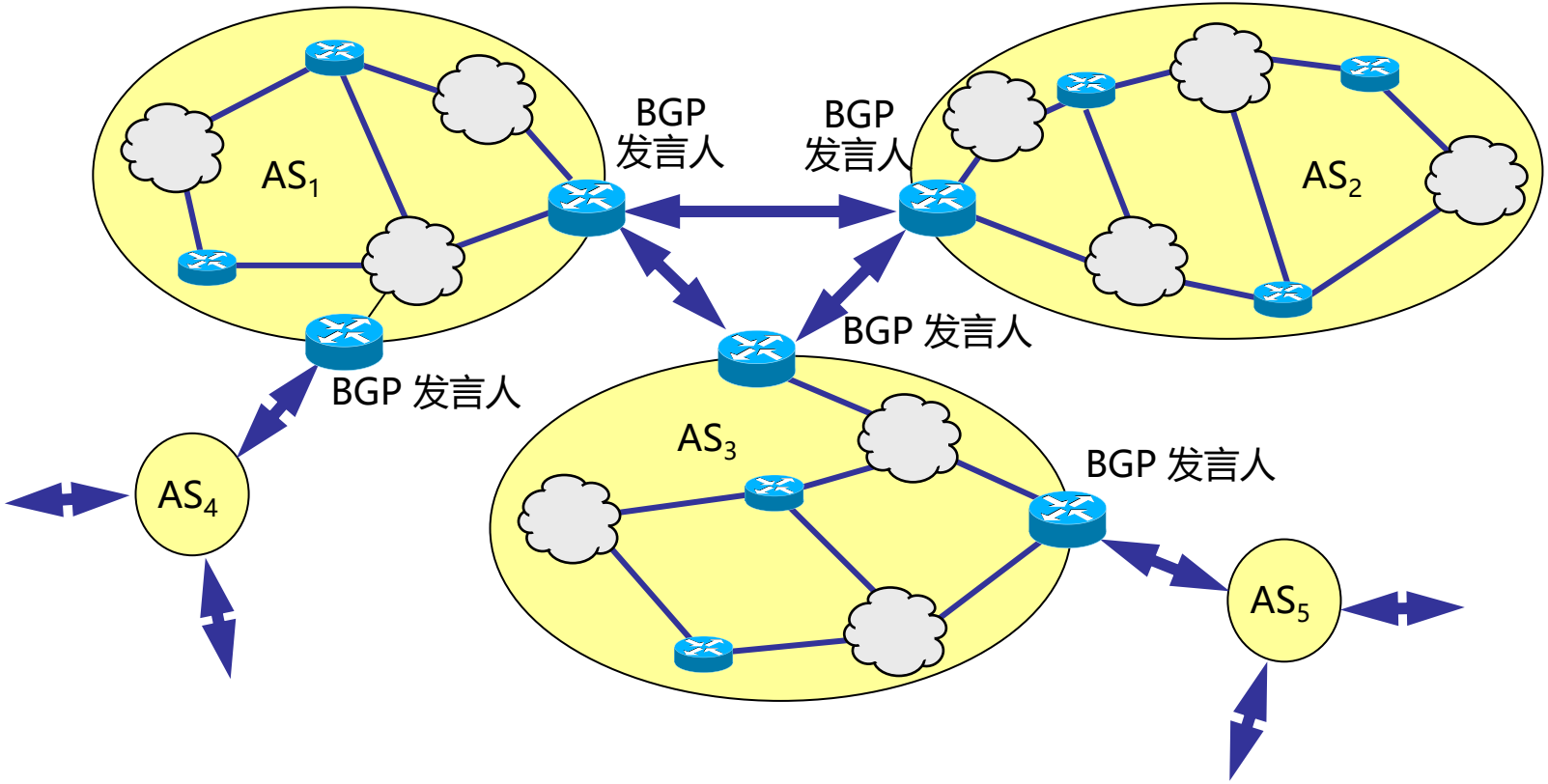
- BGP发言人（BGP speaker）
  - 每一个自治系统的管理员要选择至少一个路由器作为该自治系统的“ BGP 发言人” 。
  - 一般说来，两个 BGP 发言人都是通过一个共享网络连接在一起的，而 BGP 发言人往往就是 BGP 边界路由器，但也可以不是 BGP 边界路由器。

5.路由选择协议

- BGP发言人（BGP speaker）
  - 一个BGP发言人与其他自治系统中的BGP发言人要交换路由信息，就要先建立TCP连接，然后在此连接上交换BGP报文以建立BGP会话(session)，利用BGP会话交换路由信息。
  - 使用TCP连接能提供可靠的服务，也简化了路由选择协议。
  - 使用TCP连接交换路由信息的两个BGP发言人，彼此成为对方的邻站或对等站。

5.路由选择协议

- BGP发言人自治系统AS的关系



5.路由选择协议

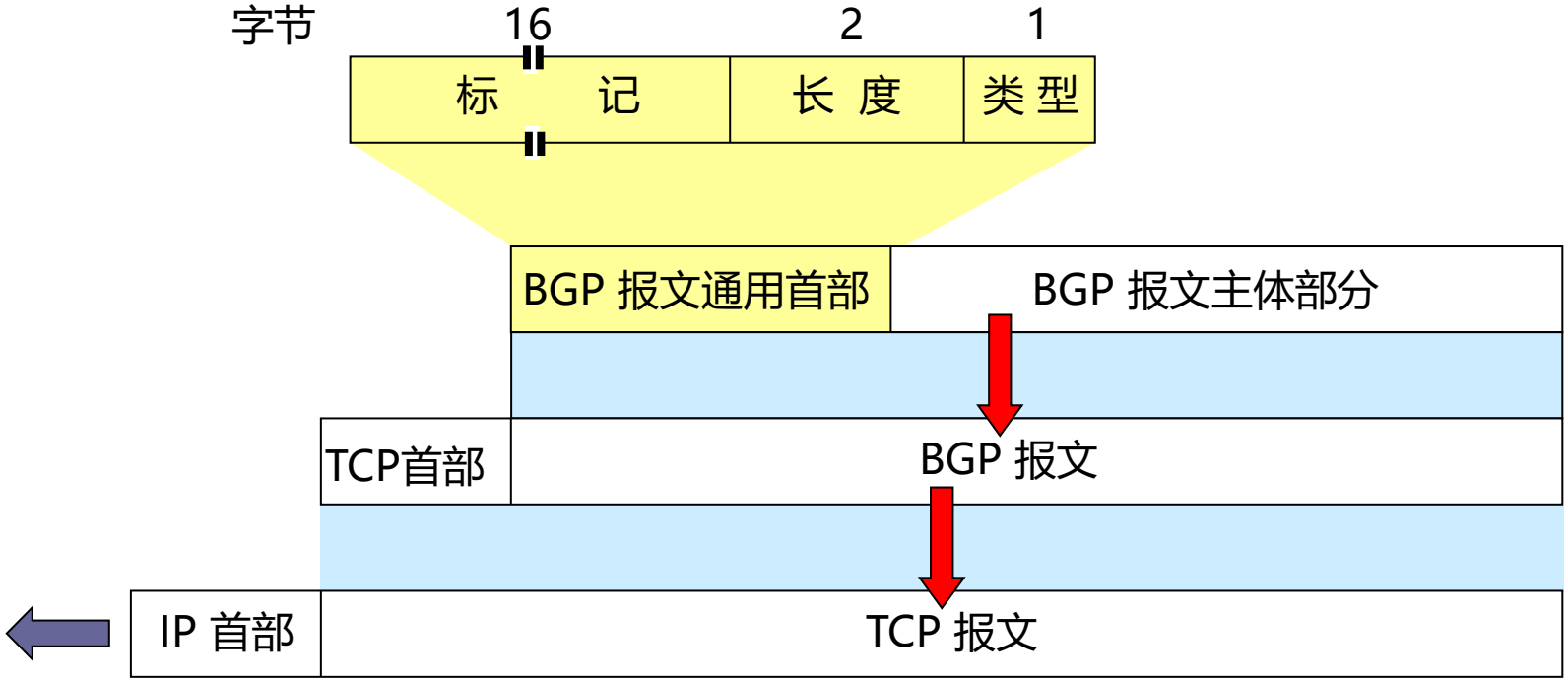
- BGP协议的特点：
  - BGP协议交换路由信息的结点数量级是自治系统数的量级，这要比这些自治系统中的网络数少很多。
  - 每一个自治系统中BGP发言人（或边界路由器）的数目是很少的。这样就使得自治系统之间的路由选择不致过分复杂。
  - BGP支持CIDR，因此BGP的路由表也就应当包括目的网络前缀、下一跳路由器，以及到达该目的网络所要经过的各个自治系统序列。
  - 在BGP刚刚运行时，BGP的邻站是交换整个的BGP路由表。但以后只需要在发生变化时更新有变化的部分。这样做对节省网络带宽和减少路由器的处理开销方面都有好处。

5.路由选择协议

- BGP-4共使用四种报文：
  - 打开(OPEN)报文，用来与相邻的另一个BGP发言人建立关系。
  - 更新(UPDATE)报文，用来发送某一路由的信息，以及列出要撤消的多条路由。
  - 保活(KEEPALIVE)报文，用来确认打开报文和周期性地证实邻站关系。
  - 通知(NOTIFICATION)报文，用来发送检测到的差错。
    - 在RFC2918中增加ROUTE-REFRESH报文，用来请求对等端重新通告。

5.路由选择协议

- BGP报文结构：

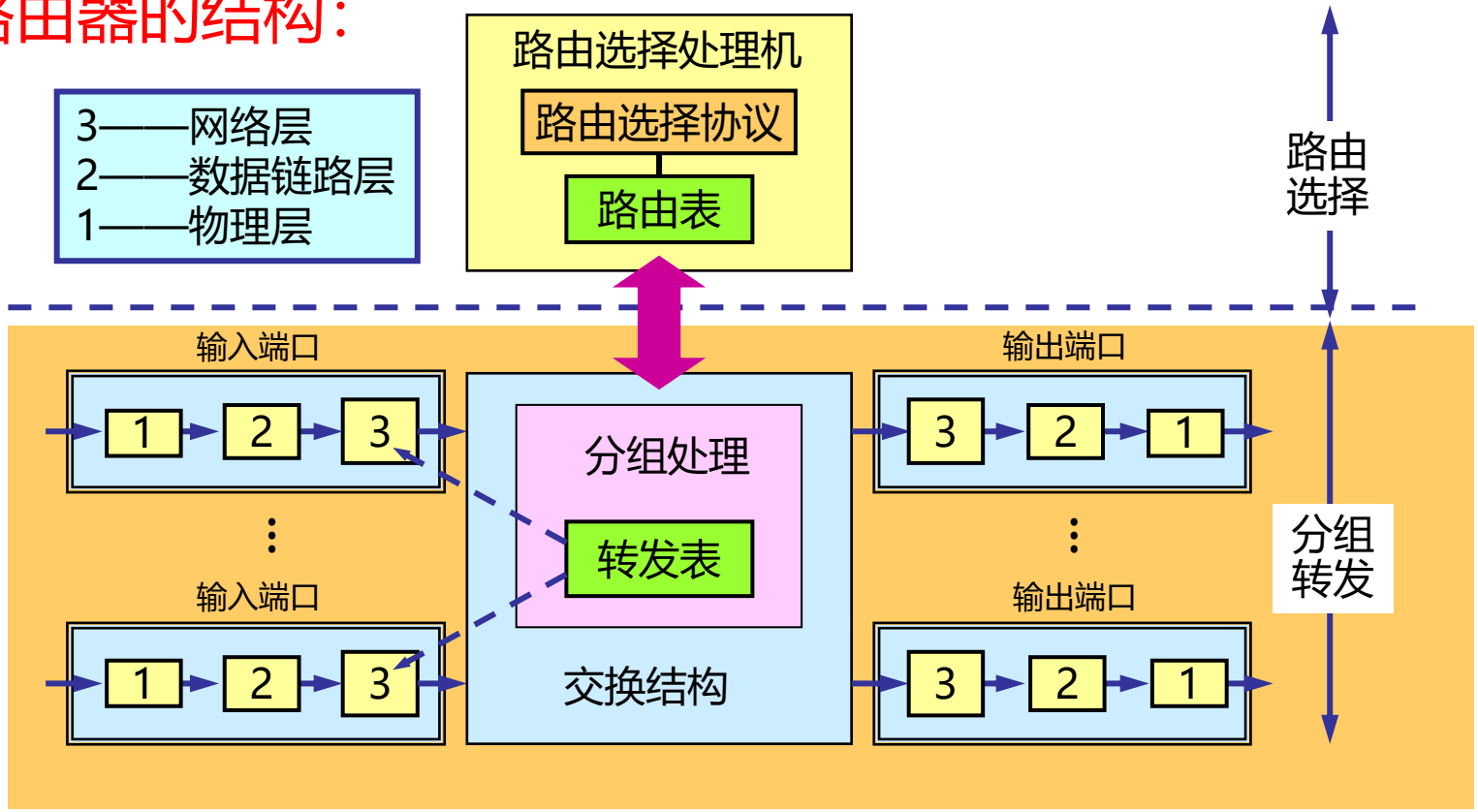


5.路由选择协议

- 路由器的结构：
  - 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。

5.路由选择协议

- 路由器的结构：



5.路由选择协议

- 典型的路由器的结构：
  - 整个的路由器结构可划分为两大部分：
    - 路由选择部分
    - 分组转发部分
  - 路由选择部分
    - 也叫做控制部分，其核心构件是路由选择处理机。
    - 路由选择处理机的任务是根据所选定的路由选择协议构造出路由表，同时经常或定期地和相邻路由器交换路由信息而不断地更新和维护路由表。

5.路由选择协议

- 典型的路由器的结构：
  - 整个的路由器结构可划分为两大部分：
    - 路由选择部分
    - 分组转发部分
  - 分组转发部分：由三部分组成
    - 交换结构 (switching fabric)：又称为交换组织，其作用是根据转发表 (forwarding table) 对分组进行处理。
    - 一组输入端口
    - 一组输出端口

端口就是硬件接口

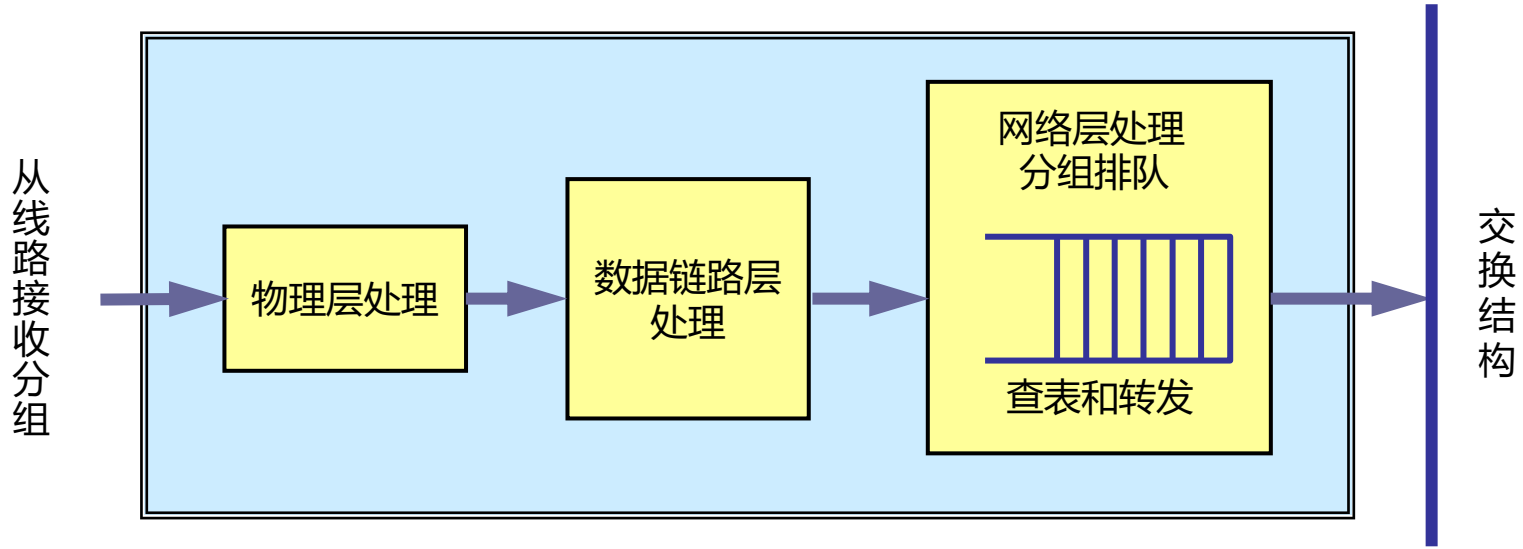
5.路由选择协议

- “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。
  - 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。

5.路由选择协议

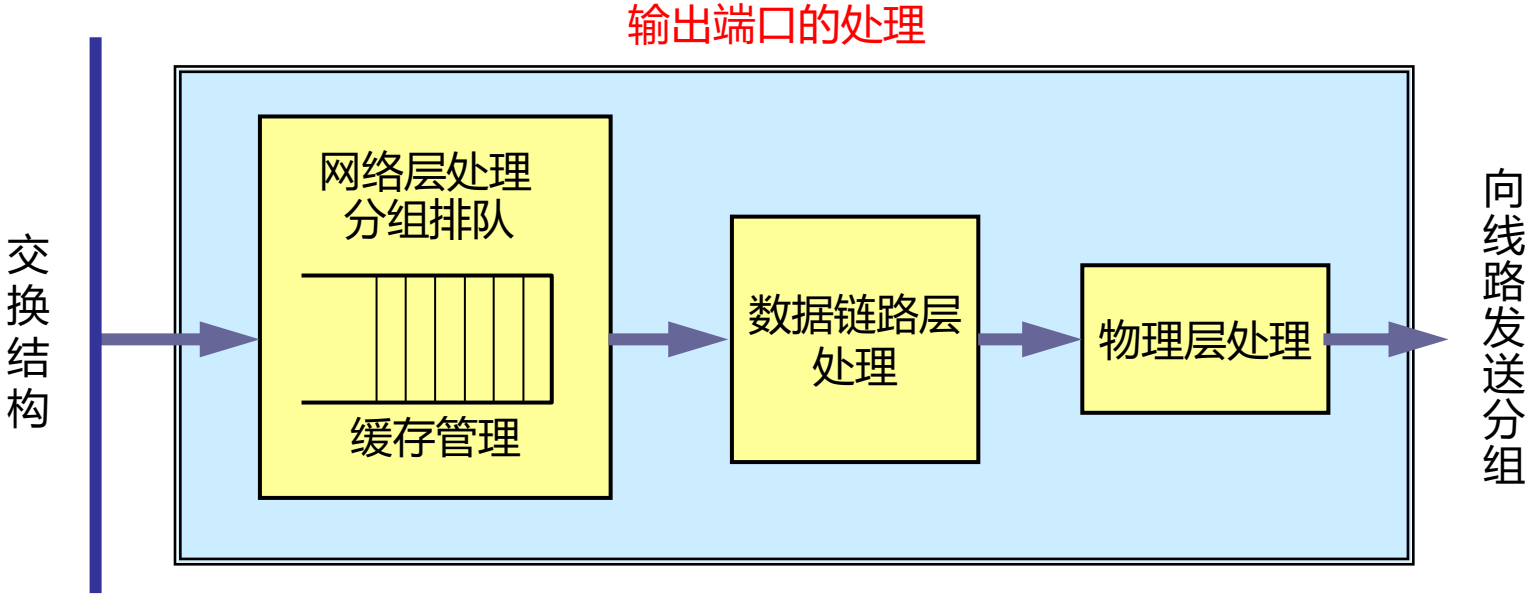
- 输入端口对线路上收到的分组的处理：
  - 数据链路层剥去帧首部和尾部后，将分组送到网络层的队列中排队等待处理。这会产生一定的时延。

输入端口的处理



5.路由选择协议

- 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。



5.路由选择协议

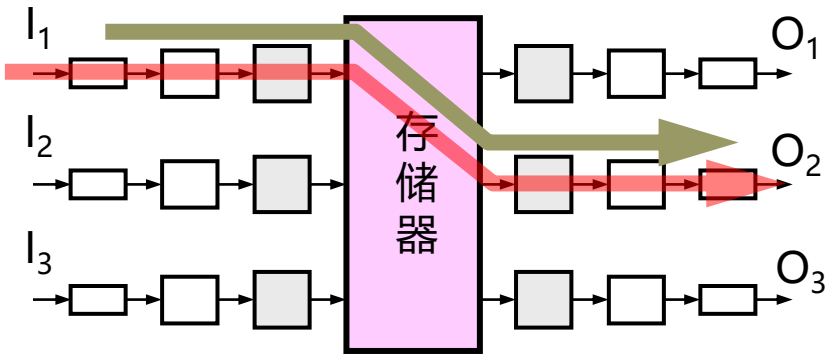
- 分组丢弃：
  - 若路由器处理分组的速率赶不上分组进入队列的速率，则队列的存储空间最终必定减少到零，这就使后面再进入队列的分组由于没有存储空间而只能被丢弃。
  - 路由器中的输入或输出队列产生溢出是造成分组丢失的重要原因。

5.路由选择协议

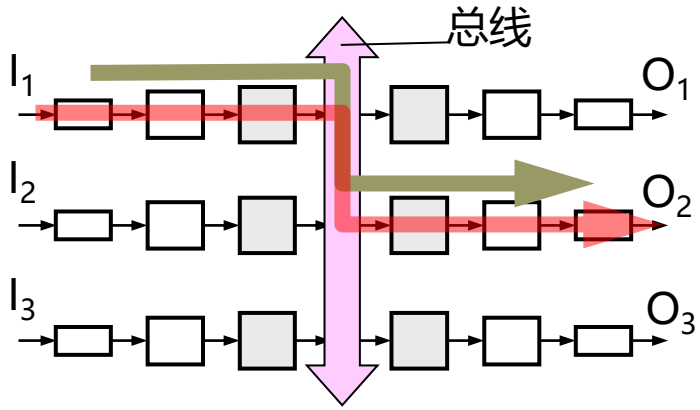
- 交换结构：
  - 交换结构是路由器的关键构件。
  - 正是这个交换结构把分组从一个输入端口转移到某个合适的输出端口。
  - 实现交换有多种方法，常用交换方法有三种：
    - 通过存储器
    - 通过总线
    - 通过纵横交换结构

5.路由选择协议

- 交换结构：



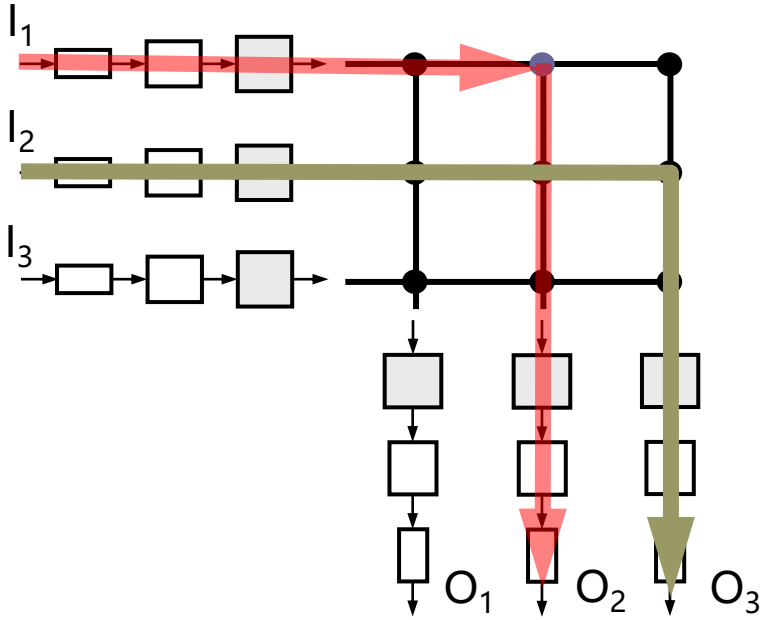
(a) 通过存储器



(b) 通过总线

5.路由选择协议

- 交换结构：



互连网络

(c) 通过互连网络

5.路由选择协议

5.路由选择协议

AR2220

AR2240

5.路由选择协议


HUAWEI


CISCO


Juniper
NETWORKS


锐捷网络
Ruijie


Extreme
networks


H3C
ITOIP解决方案专家

6.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：
  - 采用具有更大地址空间的新版本的 IP，即 IPv6。

6.IPv6

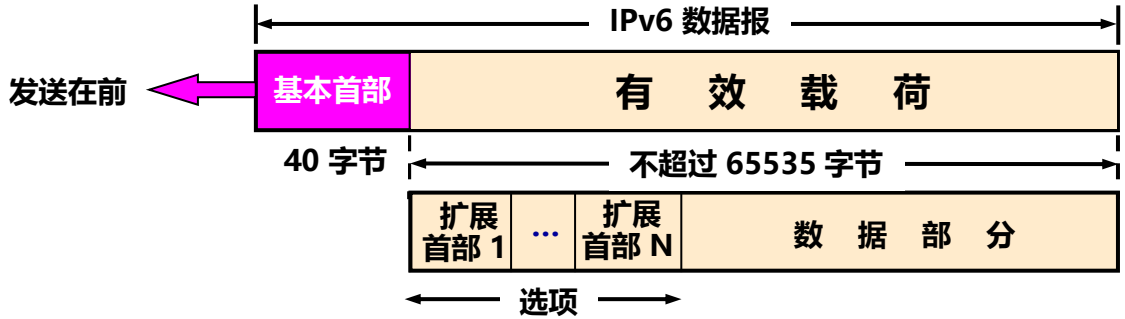
- IPv6 仍支持无连接的传送，但将协议数据单元 PDU 称为分组。教学中仍然采用数据报这一名词。
- IPv6 的主要变化如下：
  - 更大的地址空间。
    - IPv6 将地址从 IPv4 的 32 位 增大到了 128 位。
  - 扩展的地址层次结构。
  - 灵活的首部格式。
    - IPv6 定义了许多可选的扩展首部。
  - 改进的选项。
    - IPv6 允许数据报包含有选项的控制信息，其选项放在有效载荷中。

6.IPv6

- IPv6 仍支持无连接的传送，但将协议数据单元 PDU 称为分组。教学中仍然采用数据报这一名词。
- IPv6 的主要变化如下：
  - 允许协议继续扩充。
  - 支持即插即用（即自动配置）。
    - IPv6 不需要使用 DHCP。
  - 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。
  - IPv6 首部改为 8 字节对齐。
    - 首部长度必须是 8 字节的整数倍，IPv4 首部是 4 字节对齐。

6.IPv6

- IPv6 数据报由两大部分组成：
  - 基本首部 (base header)
  - 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。



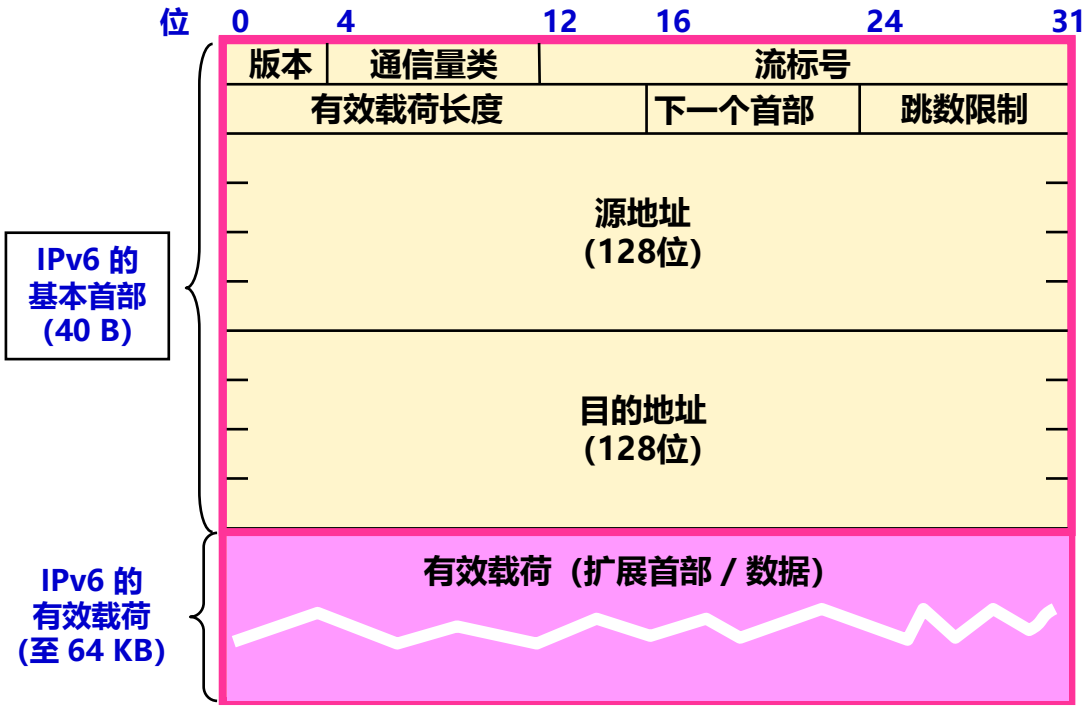
具有多个可选扩展首部的 IPv6 数据报的一般形式

6.IPv6

- IPv6 数据报由两大部分组成：
  - 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。**

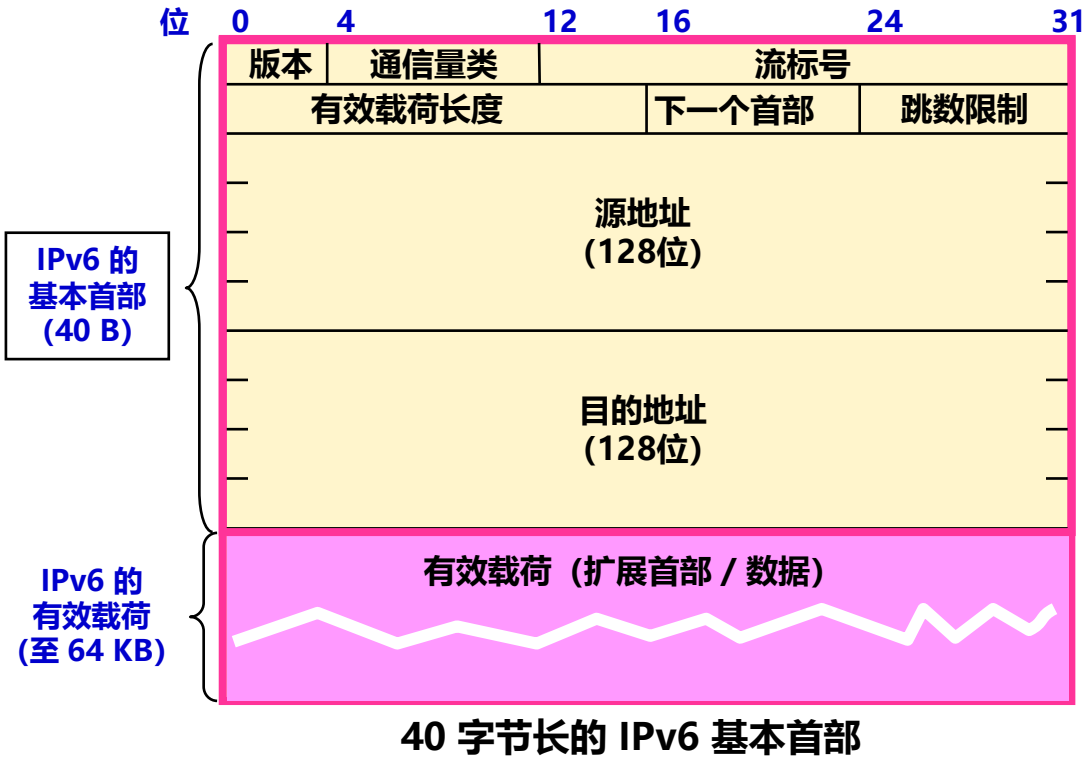
6.IPv6



40 字节长的 IPv6 基本首部

- 版本(version)：4 位。指明了协议的版本，对 IPv6 该字段总是 6。
- 通信量类(traffic class)：8 位。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。

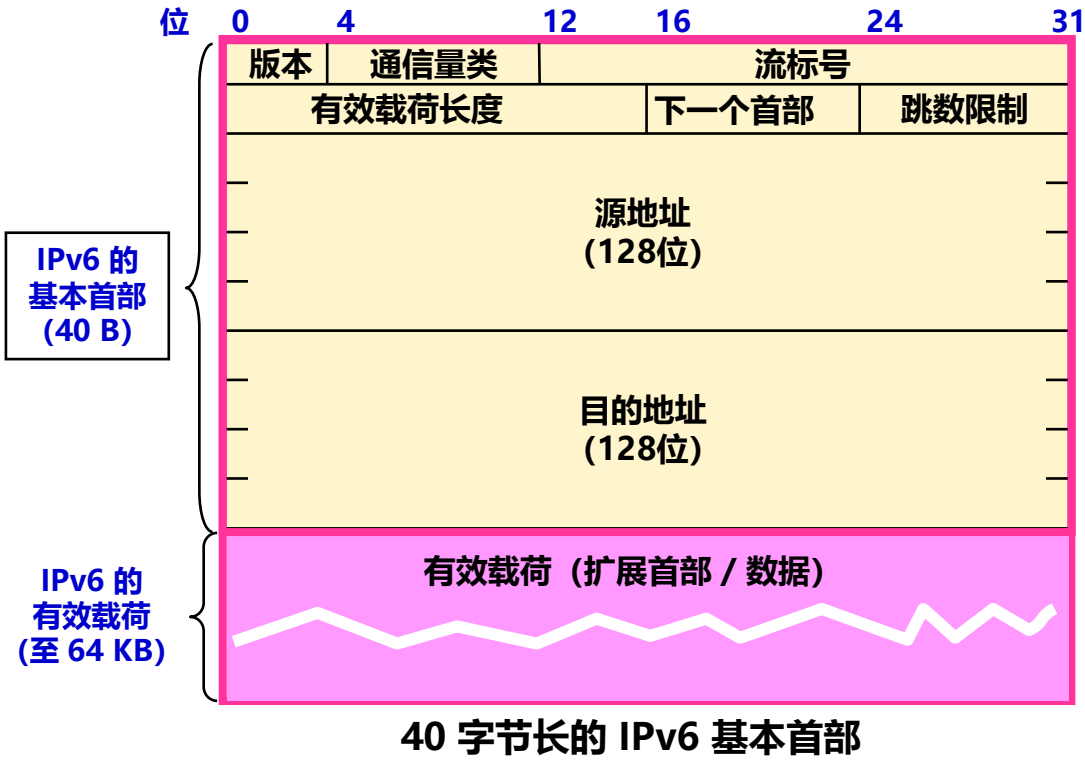
6.IPv6



40 字节长的 IPv6 基本首部

- 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），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。

6.IPv6



40 字节长的 IPv6 基本首部

- 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。

6.IPv6

- 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），极大提高了路由器的处理效率。
- 在 RFC 2460 中定义了六种扩展首部：
  - 逐跳选项
  - 路由选择
  - 分片
  - 鉴别
  - 封装安全有效载荷
  - 目的站选项

6.IPv6

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

6.IPv6

- IPv6 将实现 IPv6 的主机和路由器均称为结点。
- 一个结点可能有多个与链路相连的接口。
- IPv6 地址是分配给结点上的接口。
  - 一个接口可以有多个单播地址。
  - 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。

6.IPv6

- 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, 简写为 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:0000:1180:960A: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。例如：

    68E6:8C64:FFFF:FFFF:0:1180:960A:FFFF

6.IPv6

- 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。
    - 例如：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**。

6.IPv6

- 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
    - 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
    - 或记为：**12AB::CD30:0:0:0:0/60** （零压缩）
    - 或记为：**12AB:0:0:CD30::/60** （零压缩）

6.IPv6

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）

6.IPv6

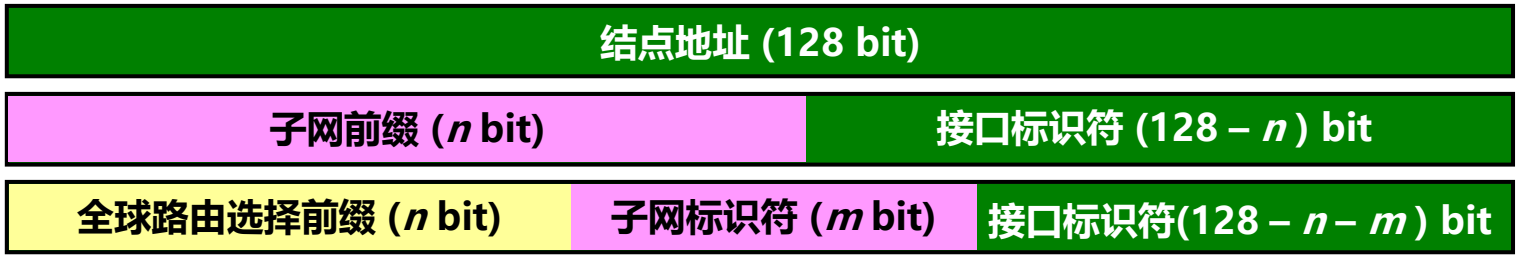
- IPv6 地址分类
  - 未指明地址
    - 这是 16 字节的全 0 地址，可缩写为两个冒号“::”。
    - 这个地址只能为还没有配置到一个标准的 IP 地址的主机当作源地址使用。
    - 这类地址仅此一个。
  - 环回地址
    - 即 0:0:0:0:0:0:0:1（记为 ::1）。
    - 作用和 IPv4 的环回地址一样。
    - 这类地址也是仅此一个。

6.IPv6

- IPv6 地址分类
  - 多播地址
    - 功能和 IPv4 的一样。
    - 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)
    - 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。
    - 这类地址占 IPv6 地址总数的 1/1024。

6.IPv6

- IPv6 地址分类
  - 全球单播地址
    - IPv6 的这一类单播地址是使用得最多的一类。
    - 曾提出过多种方案来进一步划分这 128 位的单播地址。
    - 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

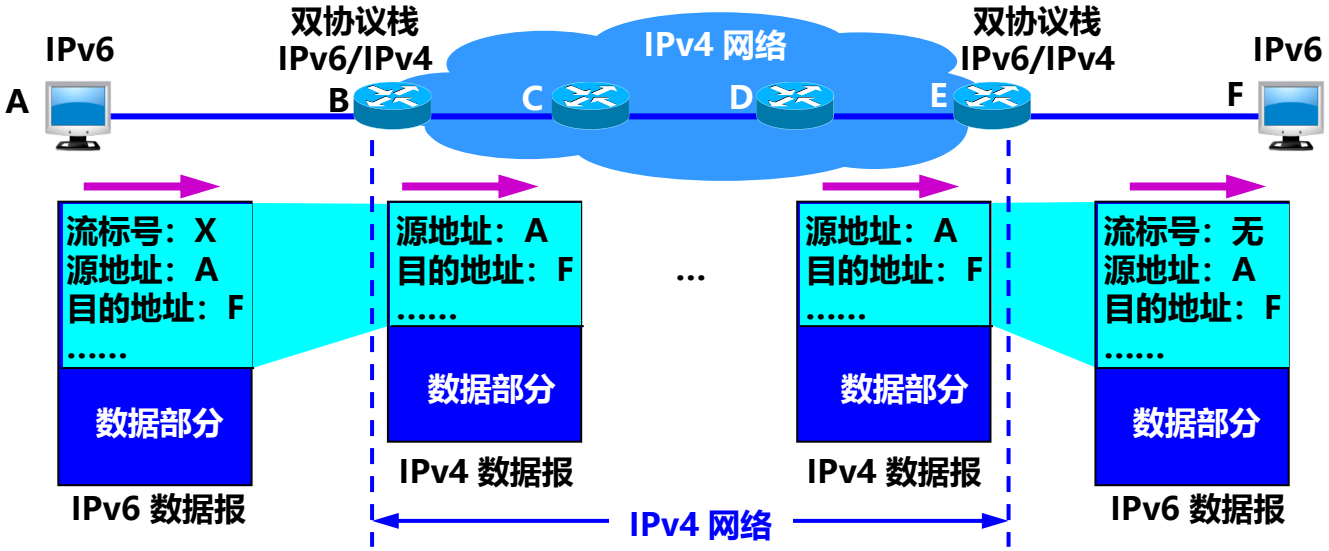
6.IPv6

- 向 IPv6 过渡只能采用逐步演进的办法，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：
  - 使用双协议栈
  - 使用隧道技术

6.IPv6

- 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。

6.IPv6

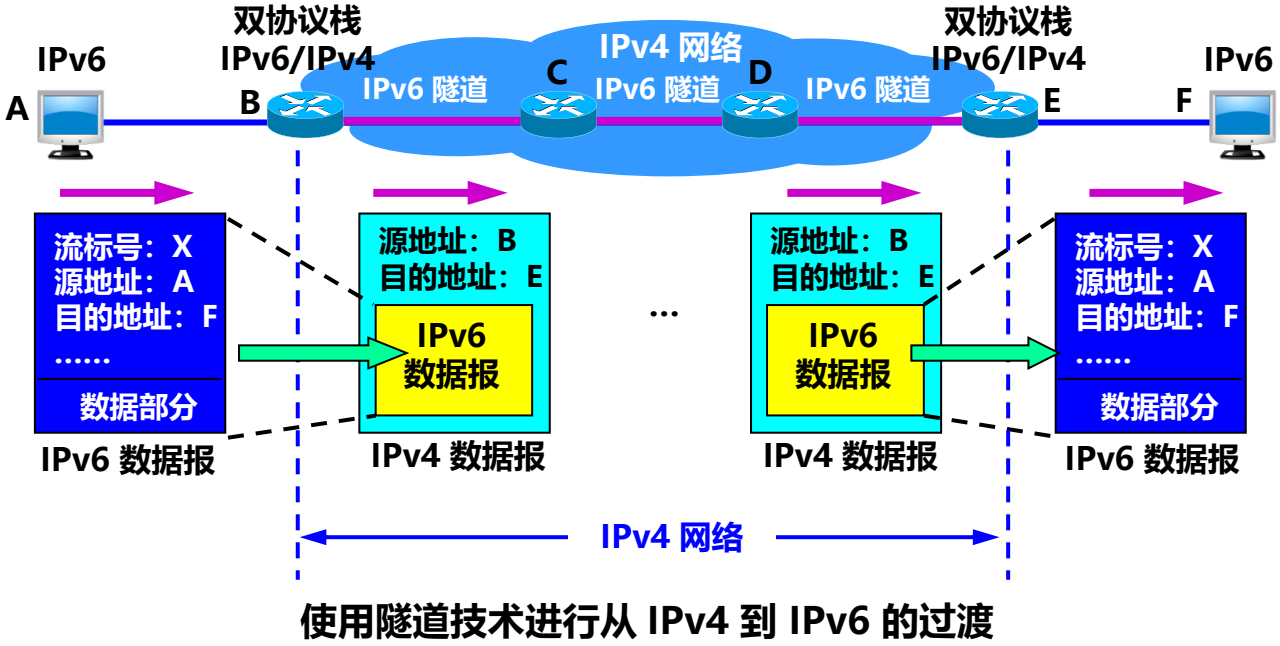


使用双协议栈进行从 IPv4 到 IPv6 的过渡

6.IPv6

- 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。

6.IPv6



使用隧道技术进行从 IPv4 到 IPv6 的过渡

6.I

过渡

```
C:\Users\Lenovo>ipconfig /all

Windows IP 配置

   主机名  . . . . . . . . . . . . . : LAPTOP-PEBPCHLE
   主 DNS 后缀 . . . . . . . . . . . :
   节点类型  . . . . . . . . . . . . : 混合
   IP 路由已启用 . . . . . . . . . . : 否
   WINS 代理已启用 . . . . . . . . . : 否

以太网适配器 以太网:

   连接特定的 DNS 后缀 . . . . . . . :
   描述. . . . . . . . . . . . . . . : Intel(R) Ethernet Connection (6) I219-V
   物理地址. . . . . . . . . . . . . : F8-75-A4-A1-DD-E2
   DHCP 已启用 . . . . . . . . . . . : 是
   自动配置已启用. . . . . . . . . . : 是
   IPv6 地址 . . . . . . . . . . . . : 2408:8220:40:3cb0:5101:3249:9b66:530(首选)
   临时 IPv6 地址. . . . . . . . . . : 2408:8220:40:3cb0:15b2:2bf9:b2e:f9c5(首选)
   本地链接 IPv6 地址. . . . . . . . : fe80::5101:3249:9b66:530%3(首选)
   IPv4 地址 . . . . . . . . . . . . : 192.168.1.83(首选)
   子网掩码  . . . . . . . . . . . . : 255.255.255.0
   获得租约的时间  . . . . . . . . . : 2020年10月30日 19:02:58
   租约过期的时间  . . . . . . . . . : 2020年10月31日 19:02:59
   默认网关. . . . . . . . . . . . . : fe80::1%3
                                       192.168.1.1
   DHCP 服务器 . . . . . . . . . . . : 192.168.1.1
   DHCPv6 IAID . . . . . . . . . . . : 116946340
   DHCPv6 客户端 DUID  . . . . . . . : 00-01-00-01-25-DE-C6-DC-F8-75-A4-A1-DD-E2
   DNS 服务器  . . . . . . . . . . . : fe80::1%3
                                       192.168.1.1
   TCPIP 上的 NetBIOS  . . . . . . . : 已启用

C:\Users\Lenovo>
```

6.I

过渡

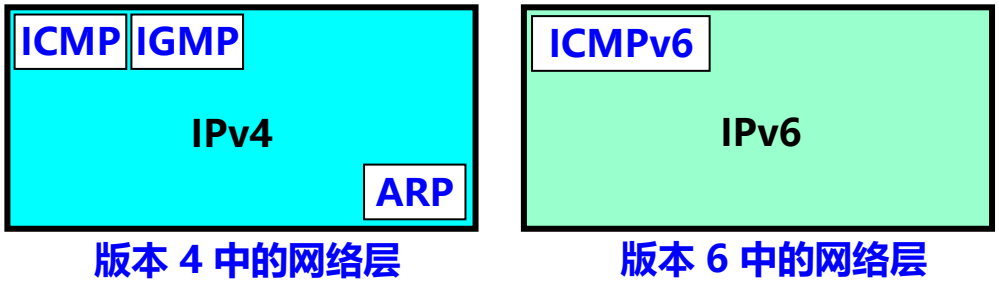
```
C:\Users\Lenovo>nslookup
默认服务器:  UnKnown
Address:  fe80::1

> www.baidu.com
服务器:  UnKnown
Address:  fe80::1

非权威应答:
名称:    www.a.shifen.com
Addresses:  61.135.169.121
          61.135.185.32
Aliases:  www.baidu.com

> www.youku.com
服务器:  UnKnown
Address:  fe80::1

非权威应答:
名称:    ipv6-aserver-heyi.m.taobao.com.gds.alibabadns.com
Addresses:  2408:4001:f10::8
          2408:4001:f00::1af
          2408:4001:f00::3a
          2408:4001:f10::fd
          106.11.35.97
Aliases:  www.youku.com
          ipv6-aserver-heyi.m.taobao.com

> www.aliyun.com
服务器:  UnKnown
Address:  fe80::1

非权威应答:
名称:    aliyun-adns.aliyun.com.gds.alibabadns.com
Addresses:  2401:b180:1:50::f
          2401:b180:1:60::6
          106.11.172.51
Aliases:  www.aliyun.com
          www-jp-de-intl-adns.aliyun.com
```

6.I

过渡

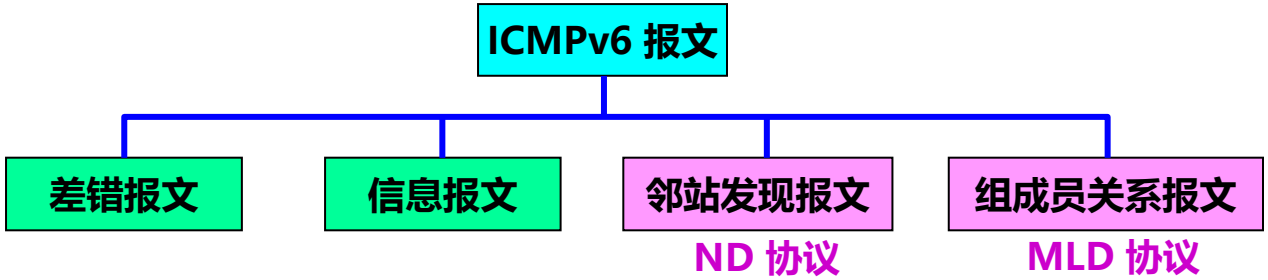
```
C:\Users\Lenovo>ping www.baidu.com

正在 Ping www.a.shifen.com [61.135.185.32] 具有 32 字节的数据:
来自 61.135.185.32 的回复: 字节=32 时间=21ms TTL=55
来自 61.135.185.32 的回复: 字节=32 时间=20ms TTL=55
来自 61.135.185.32 的回复: 字节=32 时间=20ms TTL=55
来自 61.135.185.32 的回复: 字节=32 时间=20ms TTL=55

61.135.185.32 的 Ping 统计信息:
    数据包: 已发送 = 4, 已接收 = 4, 丢失 = 0 (0% 丢失),
往返行程的估计时间(以毫秒为单位):
    最短 = 20ms, 最长 = 21ms, 平均 = 20ms

C:\Users\Lenovo>ping youku.com

正在 Ping youku.com [2408:4001:f10::fd] 具有 32 字节的数据:
来自 2408:4001:f10::fd 的回复: 时间=28ms
来自 2408:4001:f10::fd 的回复: 时间=27ms
来自 2408:4001:f10::fd 的回复: 时间=28ms
来自 2408:4001:f10::fd 的回复: 时间=27ms

2408:4001:f10::fd 的 Ping 统计信息:
    数据包: 已发送 = 4, 已接收 = 4, 丢失 = 0 (0% 丢失),
往返行程的估计时间(以毫秒为单位):
    最短 = 27ms, 最长 = 28ms, 平均 = 27ms

C:\Users\Lenovo>ping aliyun.com

正在 Ping aliyun.com [2401:b180:1:50::f] 具有 32 字节的数据:
来自 2401:b180:1:50::f 的回复: 时间=24ms
来自 2401:b180:1:50::f 的回复: 时间=24ms
来自 2401:b180:1:50::f 的回复: 时间=23ms
来自 2401:b180:1:50::f 的回复: 时间=24ms

2401:b180:1:50::f 的 Ping 统计信息:
    数据包: 已发送 = 4, 已接收 = 4, 丢失 = 0 (0% 丢失),
往返行程的估计时间(以毫秒为单位):
    最短 = 23ms, 最长 = 24ms, 平均 = 23ms
```

- ICMPv6
  - IPv6 也不保证数据报的可靠交付，因为互联网中的路由器可能会丢弃数据报。因此 IPv6 也需要使用 ICMP 来反馈一些差错信息，新的版本称为 ICMPv6。
  - 地址解析协议 ARP 和网际组管理协议 IGMP 协议的功能都已被合并到 ICMPv6 中。



新旧版本中的网络层的比较

- ICMPv6
  - ICMPv6 是面向报文的协议，它利用报文来报告差错，获取信息，探测邻站或管理多播通信。
  - ICMPv6 还增加了几个定义报文的功能及含义的其他协议。



ICMPv6 报文的分类

7.VPN和NAT

- 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。

7.VPN和NAT

- 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。
  - 问题与解决方案：
    - 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。
    - 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。

7.VPN和NAT

- 本地地址与全球地址
  - RFC 1918 指明了一些**专用地址 (private address)**。

(1) 10.0.0.0 到 10.255.255.255

A类，或记为10.0.0.0/8，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，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，又称为 16 位块

7.VPN和NAT

- 专用网
  - 采用专用 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)。

7.VPN和NAT

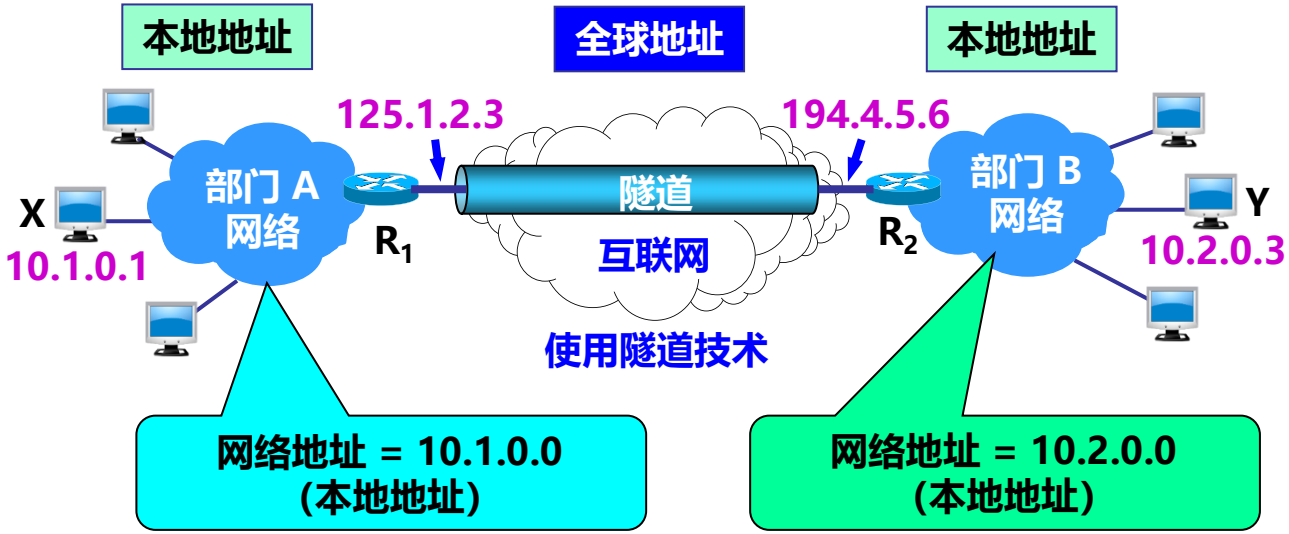
- 虚拟专用网 VPN
  - 利用公用的互联网作为本机构各专用网之间的通信载体，这样的专用网又称为**虚拟专用网VPN** (Virtual Private Network)。
  - “专用网”是因为这种网络是为本机构的主机用于机构内部的通信，而不是用于和网络外非本机构的主机通信。
  - “虚拟”表示“好像是”，但实际上并不是，因为现在并没有真正使用通信专线，而VPN只是在效果上和真正的专用网一样。

7.VPN和NAT

- 虚拟专用网 VPN 的构建方法
  - 如果专用网不同网点之间的通信必须经过公用的互联网，但又有保密的要求，那么所有通过互联网传送的数据都必须加密。
  - 一个机构要构建自己的 VPN 就必须为它的每一个场所购买专门的硬件和软件，并进行配置，使每一个场所的 VPN 系统都知道其他场所的地址。

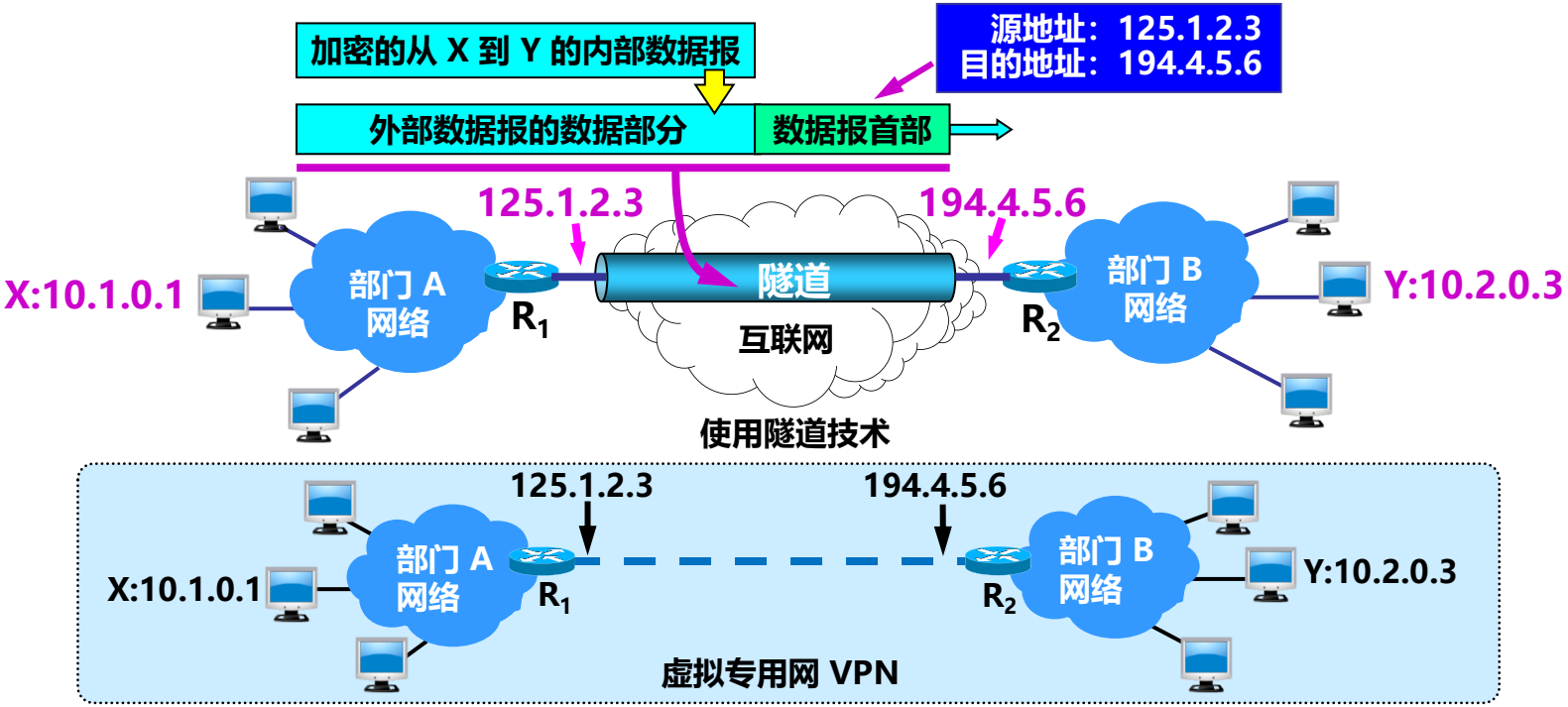
7.VPN和NAT

- 虚拟专用网 VPN 的实现：基于隧道技术



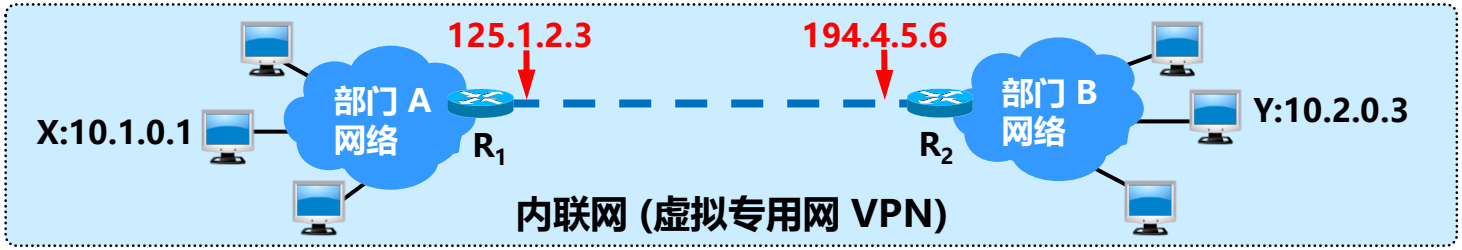
7.VPN和NAT

- 虚拟专用网 VPN 的实现：基于隧道技术



7.VPN和NAT

- 内联网 intranet 和外联网 extranet
  - 均基于 TCP/IP 协议。
  - 由部门 A 和 B 的内部网络所构成的虚拟专用网 VPN 又称为**内联网 (intranet)**，表示部门 A 和 B 都是在同一个机构的内部。
  - 某机构和其他的外部机构共同建立的虚拟专用网VPN 又称为**外联网 (extranet)**。



7.VPN和NAT

- 远程接入 VPN
  - 远程接入 VPN (remote access VPN) 可以满足外部流动员工访问公司网络的需求。
  - 在外地工作的员工拨号接入互联网，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。
  - 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。

7.VPN和NAT

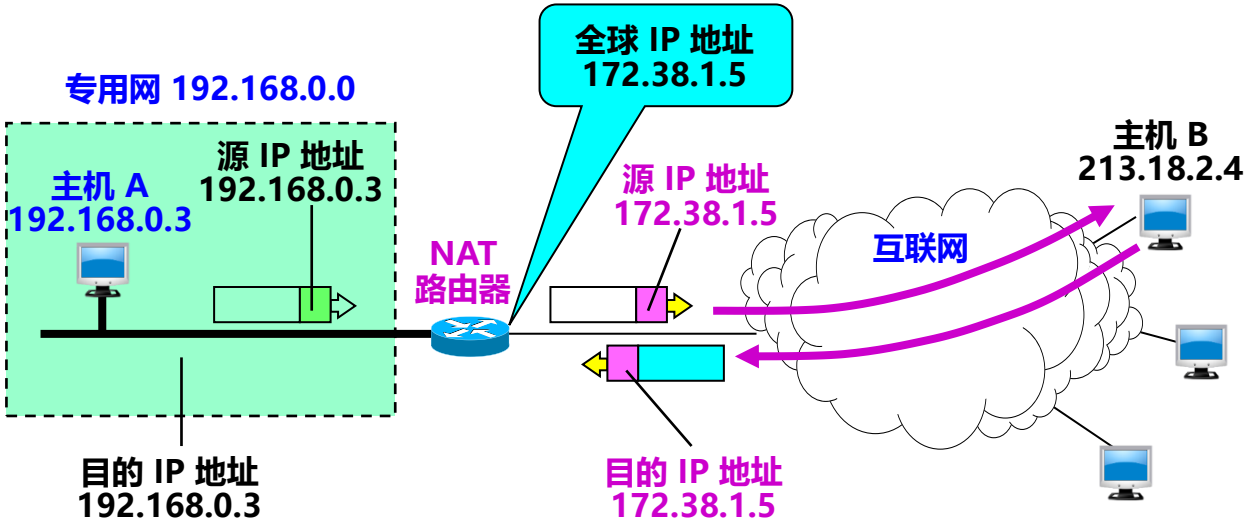
- 远程接入 VPN
  - 对于个人“翻墙”行为的规定早已明确。
  - 1996年1月23日的国务院常务会议通过《计算机信息网络国际联网管理暂行规定》，该规定在1997年进行修正。
  - 该《规定》第六条提出，计算机信息网络直接进行国际联网，必须使用邮电部国家公用电信网提供的国际出入口信道。该规定明确要求，**任何单位和个人不得自行建立或者使用其他信道进行国际联网**。如果违反第六条规定，由公安机关责令停止联网，给予警告，可以并处15000元以下的罚款;有违法所得的，没收违法所得。

7.VPN和NAT

- 网络地址转换 NAT (Network Address Translation)
  - 1994年提出。
  - 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。

7.VPN和NAT

- 网络地址转换 NAT (Network Address Translation)



NAT 路由器的工作原理

7.VPN和NAT

- 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。

7.VPN和NAT

- 网络地址转换的过程
  - 在内部主机与外部主机通信时，NAT路由器上发生了两次地址转换：
    - 离开专用网时：替换源地址，将内部地址替换为全球地址
    - 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

7.VPN和NAT

- 网络地址转换
  - 当 NAT 路由器具有 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。

NAT显然有巨大的缺陷
无法应用于较大规模的园区网

7.VPN和NAT

- 网络地址与端口号转换 NAPT
  - 为了更加有效地利用 NAT 路由器上的全球IP地址，现在常用的 NAT 转换表把运输层的端口号也利用上。这样就可以使多个拥有本地地址的主机，共用一个 NAT 路由器上的全球 IP 地址，因而可以同时和互联网上的不同主机进行通信。
  - 使用端口号的 NAT 叫做**网络地址与端口号转换NAPT** (Network Address and Port Translation)，而不使用端口号的 NAT 就叫做传统的 NAT (traditional NAT)。

7.VPN和NAT

- 网络地址与端口号转换 NAPT
  - NAPT把专用网内不同的源 IP 地址，都转换为同样的全球 IP 地址。但对源主机所采用的 TCP 端口号（不管相同或不同），则转换为不同的新的端口号。
  - 当 NAPT 路由器收到从互联网发来的应答时，就可以从 IP 数据报的数据部分找出运输层的端口号，然后根据不同的目的端口号，从 NAPT 转换表中找到正确的目的主机。

7.VPN和NAT

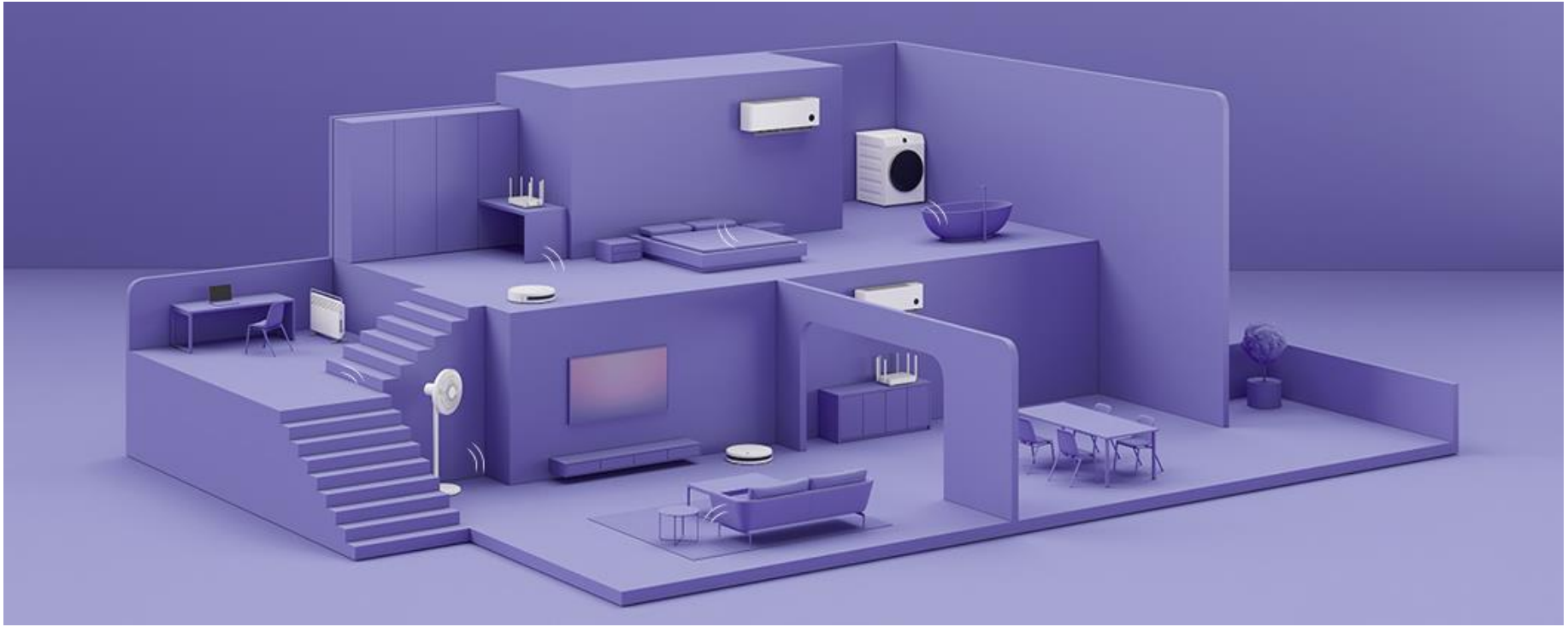
网络地址与端口号转换 NAPT

NAPT 地址转换表举例

方向	字段	旧的IP地址和端口号	新的IP地址和端口号
出	源IP地址:TCP源端口	192.168.0.3:30000	172.38.1.5:40001
出	源IP地址:TCP源端口	192.168.0.4:30000	172.38.1.5:40002
入	目的IP地址:TCP目的端口	172.38.1.5:40001	192.168.0.3:30000
入	目的IP地址:TCP目的端口	172.38.1.5:40002	192.168.0.4:30000

7.VPN和NAT

- 网络地址与端口号转换 NAPT



Thanks