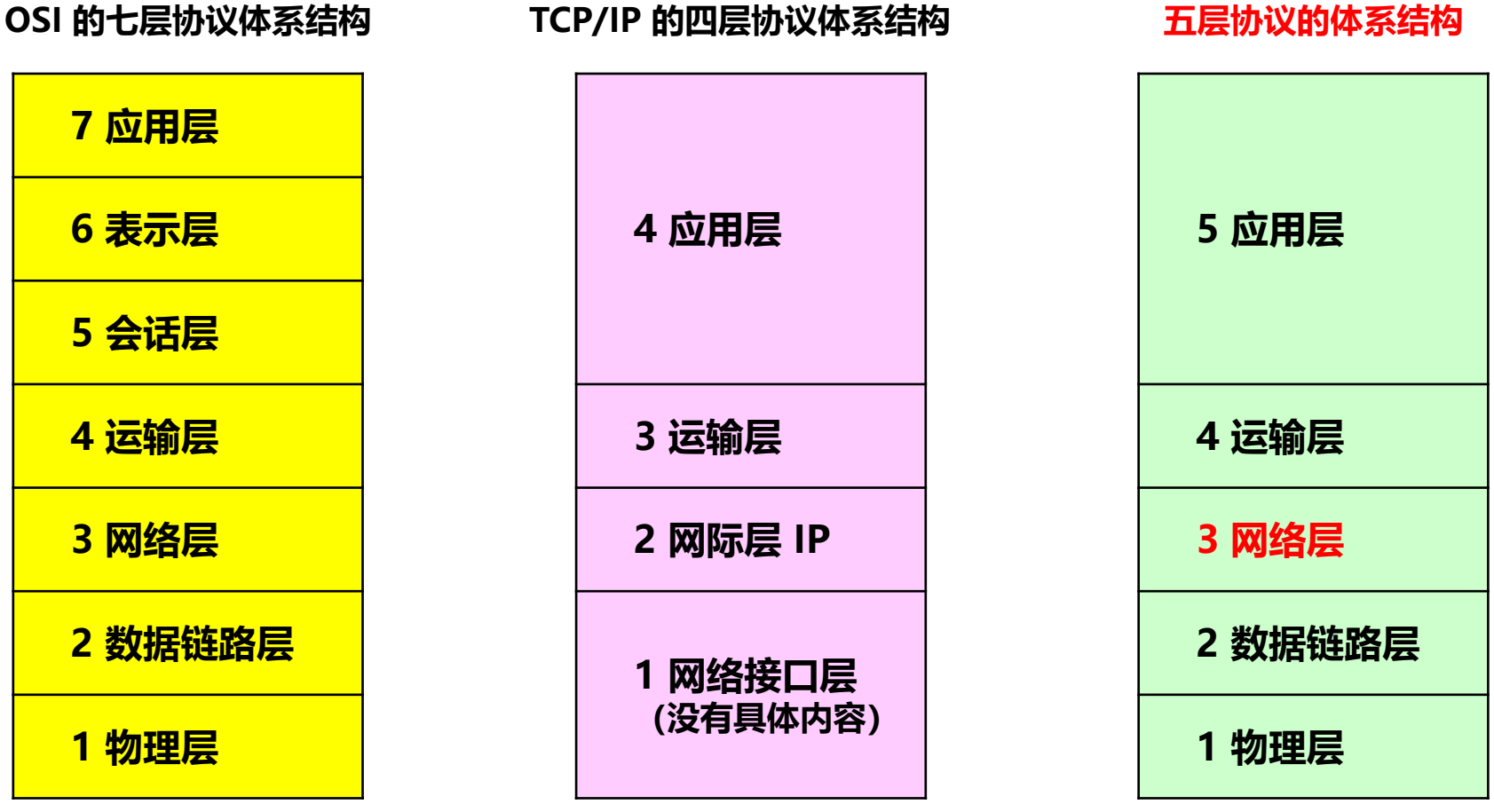
计算机网络原理

第4章：网络层

https://internet.hactcm.edu.cn

河南中医药大学信息技术学院互联网技术教学团队
河南中医药大学医疗健康信息工程技术研究所

2024.2


OSI 的七层协议体系结构
7 应用层
6 表示层
5 会话层
4 运输层
3 网络层
2 数据链路层
1 物理层
TCP/IP 的四层协议体系结构
4 应用层
3 运输层
2 网际层 IP
1 网络接口层
（没有具体内容）
五层协议的体系结构
5 应用层
4 运输层
3 网络层
2 数据链路层
1 物理层

网络层讨论

多个网络通过路由器互连成为一个互联网

本章教学计划

- ✓ 网络层的重要概念
- ✓ 网际协议 IP
- ✓ IP 层转发分组的过程
- ✓ 网际控制报文协议 ICMP
- ✓ IPv6
- ✓ 互联网的路由选择协议

网络层的基本内容

- ✓ IP 多播
- ✓ 虚拟专用网 VPN 和网络地址转换 NAT

网络层的扩展应用

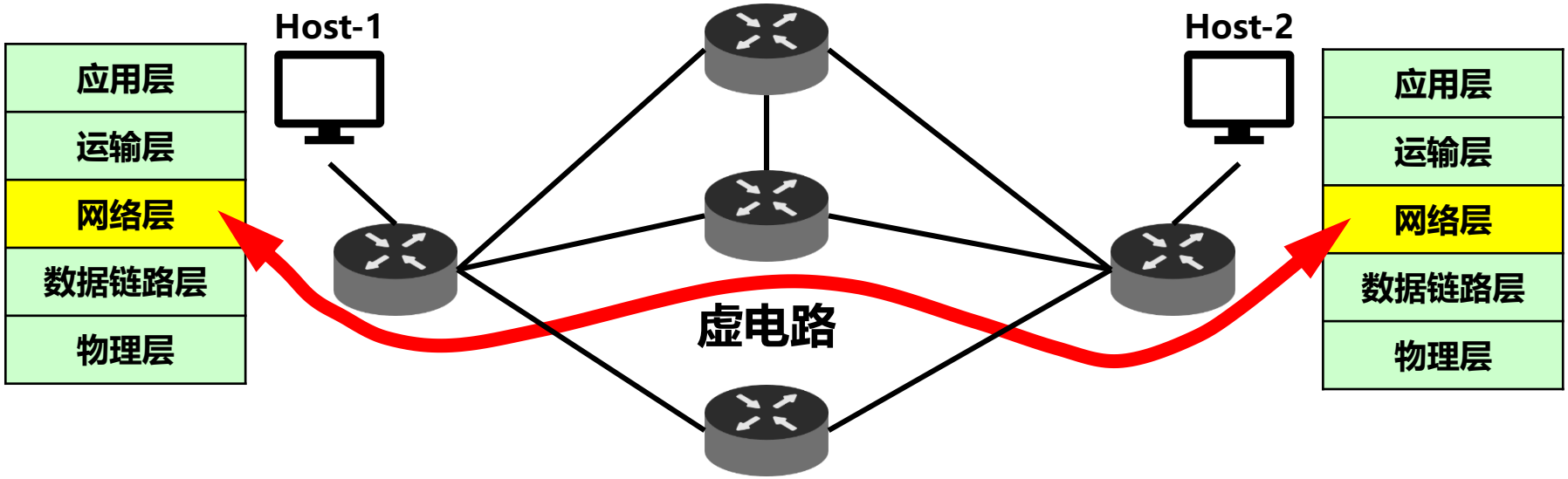
1. 网络层的重要概念

- 争论：
  - 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？
- 争论的实质是：
  - 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？
- 2 种观点：
  - 面向连接的可靠交付。
  - 无连接的、尽最大努力交付的数据报服务，不提供服务质量的承诺。

1. 网络层的重要概念

- 第 1 种观点：让网络负责可靠交付
  - 计算机网络模仿电信网络，使用面向连接的通信方式。
  - 通信之前先建立虚电路 VC (Virtual Circuit) (即连接)，以保证双方通信所需的一切网络资源。
  - 如果再使用可靠传输的网络协议，可使所发送的分组无差错按序到达终点，不丢失、不重复。

1. 网络层的重要概念



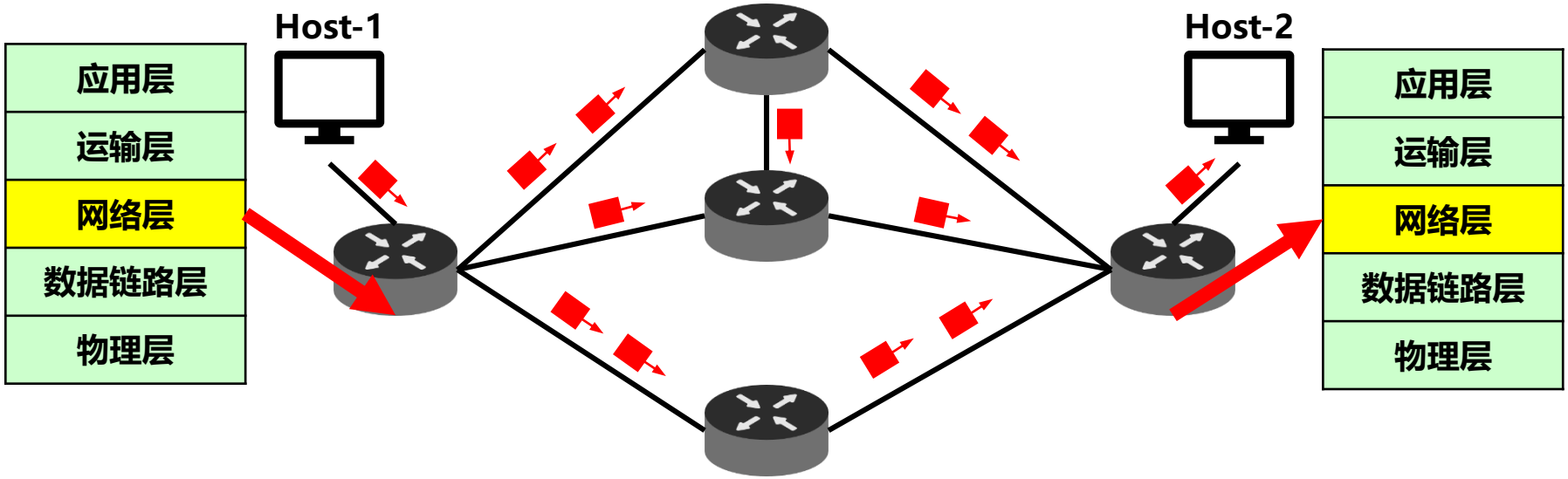
H1 发送给 H2 的所有分组都沿着同一条虚电路传送

虚电路只是一条逻辑上的连接，分组都沿着这条逻辑连接按照存储转发方式传送，并不是真正建立了一条物理连接。

1. 网络层的重要概念

- 第 2 种观点：网络提供数据报服务
  - 网络层要设计得尽量简单，向其上层只提供简单灵活的、无连接的、尽最大努力交付的数据报服务。
    - 网络在发送分组时不需要先建立连接。
    - 每一个分组（即 IP 数据报）独立发送，与其前后的分组无关（不进行编号）。
    - 网络层不提供服务质量的承诺。
      - 即所传送的分组可能出错、丢失、重复和失序（不按序到达终点），也不保证分组传送的时限。
  - 由主机中的运输层负责可靠的通信。

1. 网络层的重要概念



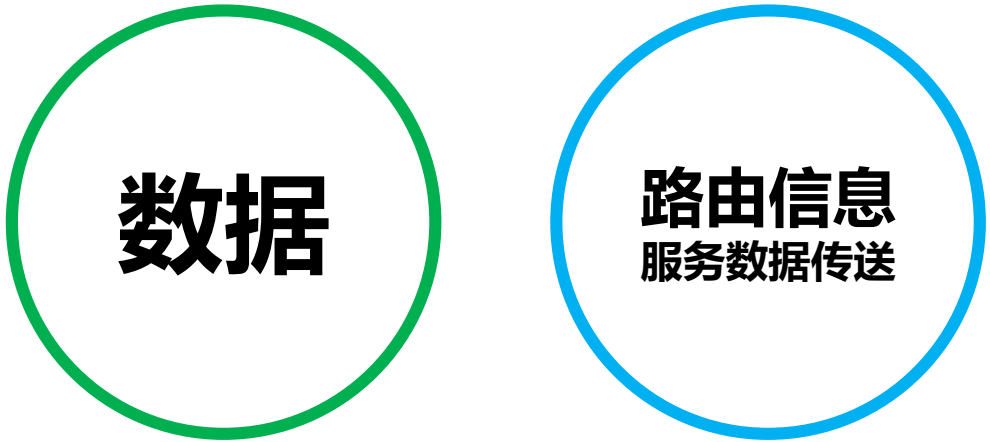
H1 发送给 H2 的分组可能沿着不同路径传送

虚电路只是一条逻辑上的连接，分组都沿着这条逻辑连接按照存储转发方式传送，并不是真正建立了一条物理连接。

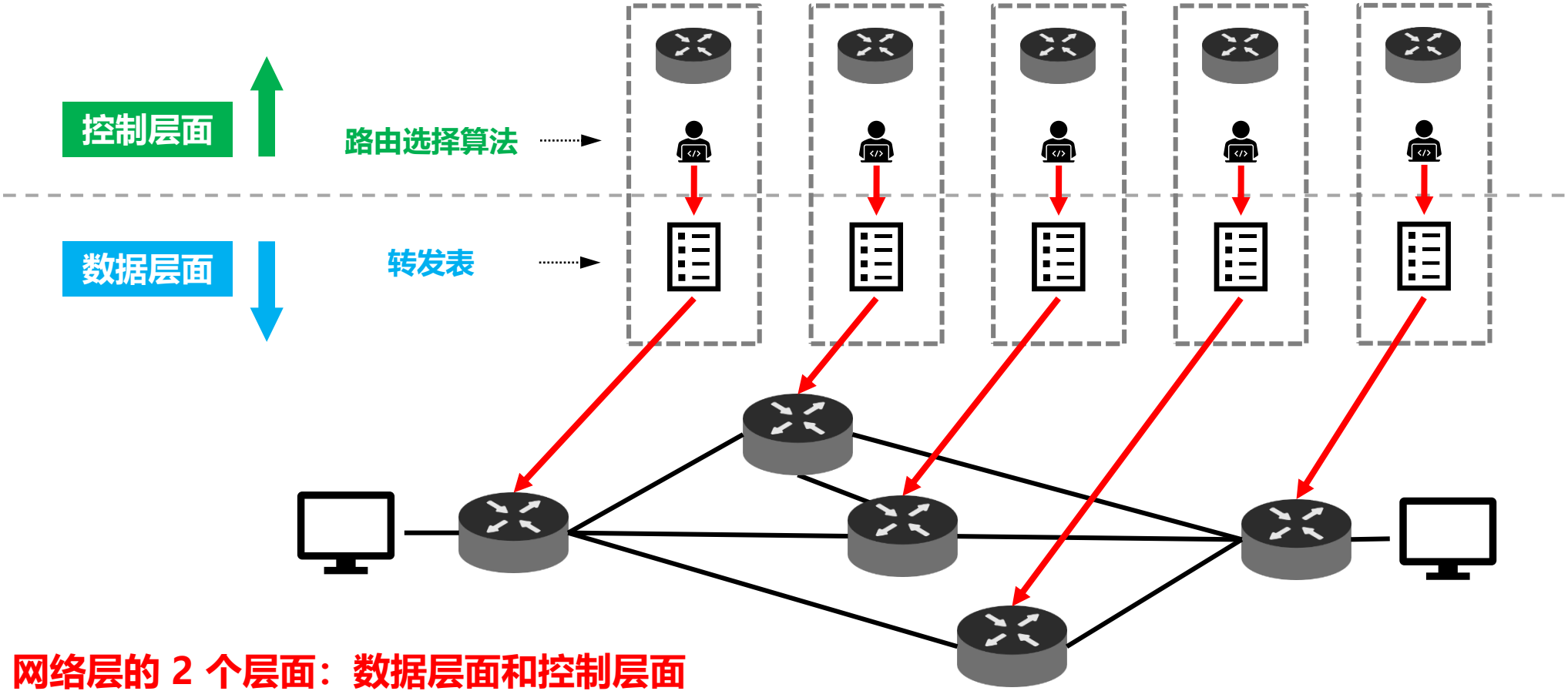
对比的方面	虚电路服务	数据报服务
思路	可靠通信应当由网络来保证	可靠通信应当由用户主机来保证
连接的建立	必须有	不需要
终点地址	仅在连接建立阶段使用，每个分组使用短的虚电路号	每个分组都有终点的完整地址
分组的转发	属于同一条虚电路的分组均按照同一路由进行转发	每个分组独立选择路由进行转发
当结点出故障时	所有通过出故障的结点的虚电路均不能工作	出故障的结点可能会丢失分组，一些路由可能会发生变化
分组的顺序	总是按发送顺序到达终点	到达终点时不一定按发送顺序
端到端的差错处理和流量控制	可以由网络负责，也可以由用户主机负责	由用户主机负责

1. 网络层的重要概念

- 不同网络中两个主机之间通信，要经过若干个路由器转发分组来完成。
- 在路由器之间传送的信息有以下 2 大类：



1. 网络层的重要概念



网络层的 2 个层面：数据层面和控制层面

1. 网络层的重要概念

网络层的 2 个层面：数据层面和控制层面

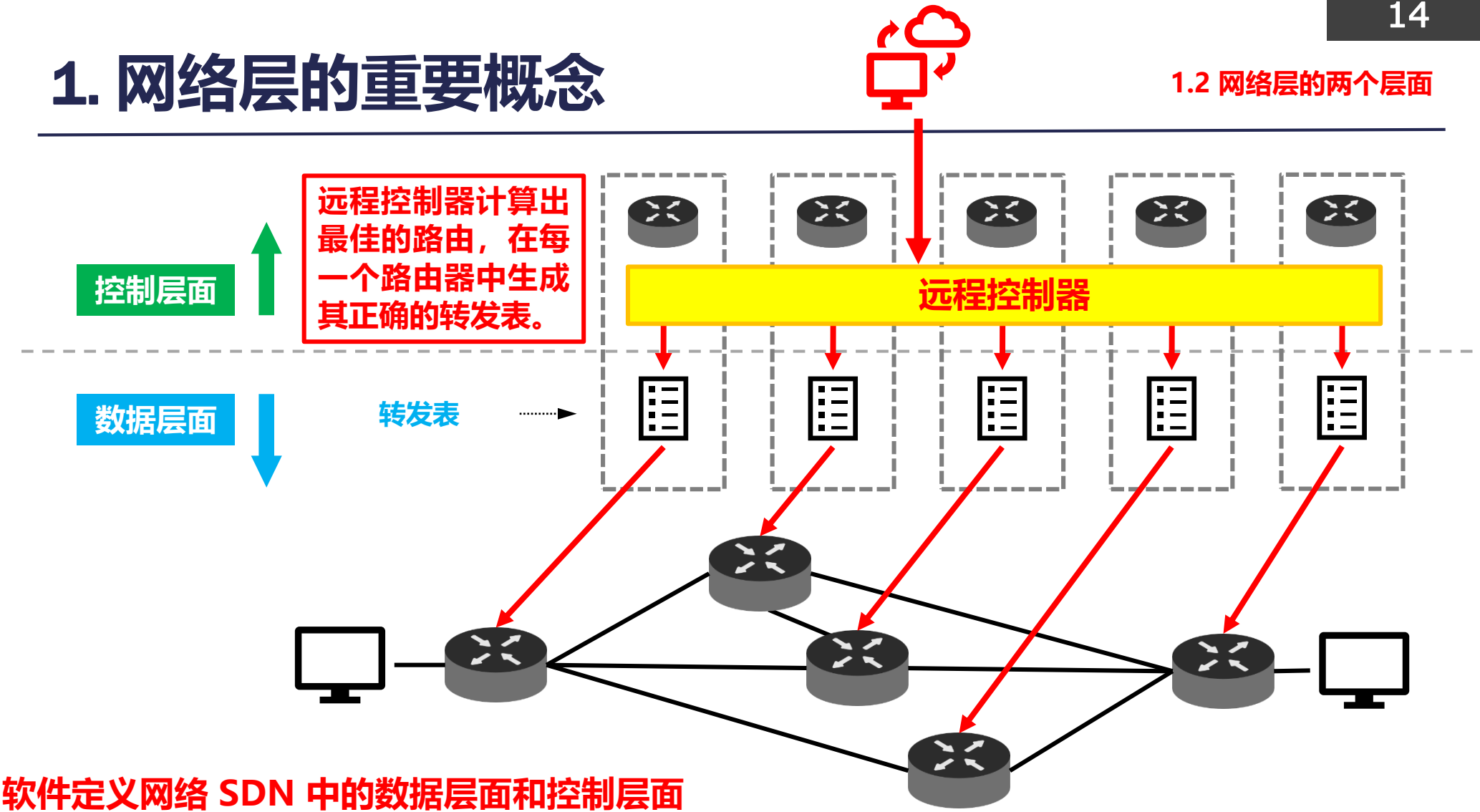
数据层面

- 路由器根据本路由器生成的转发表，把收到的分组从查找到的对应接口转发出去。
- 独立工作。
- 采用硬件进行转发，快。

控制层面

- 根据路由选择协议所用的路由算法计算路由，创建出本路由器的路由表。
- 许多路由器协同动作。
- 采用软件计算，慢。

1. 网络层的重要概念



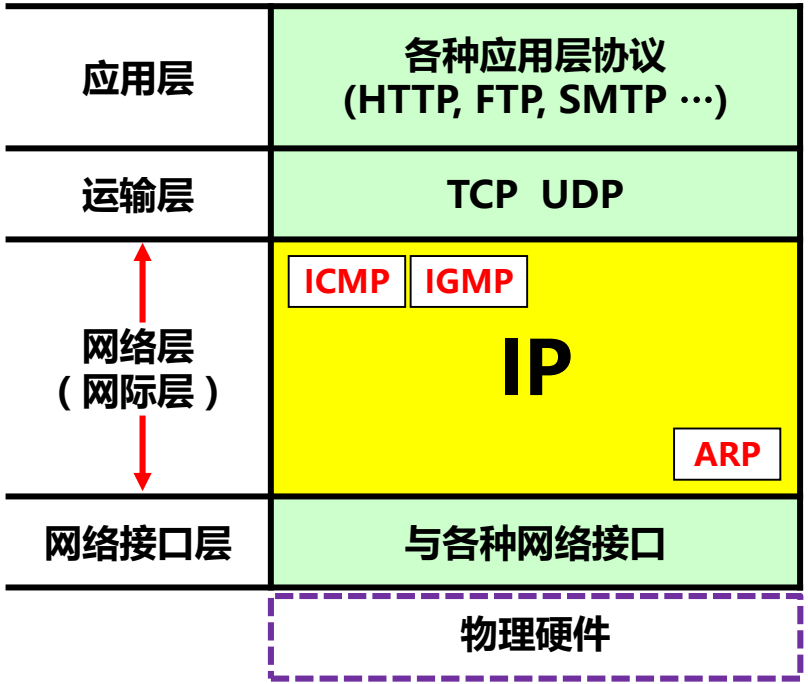
软件定义网络 SDN 中的数据层面和控制层面

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，称为 IPv4，目前仍然是最主要的互联网协议。
  - 世界各地正在积极部署 IPv6 。
  - 当然，也有IPv5、IPv7、IPv9、IPv10...

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 3 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

- 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

- 如果都是用相同的网络：
  - 不能满足不同用户需要。
    - 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。
  - 不适应技术发展
    - 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。
- **实现异构网络的互连互通方法：使用中间设备**
  - 可以满足不同需求
    - 实现市场上有不同性能、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。
  - 实用

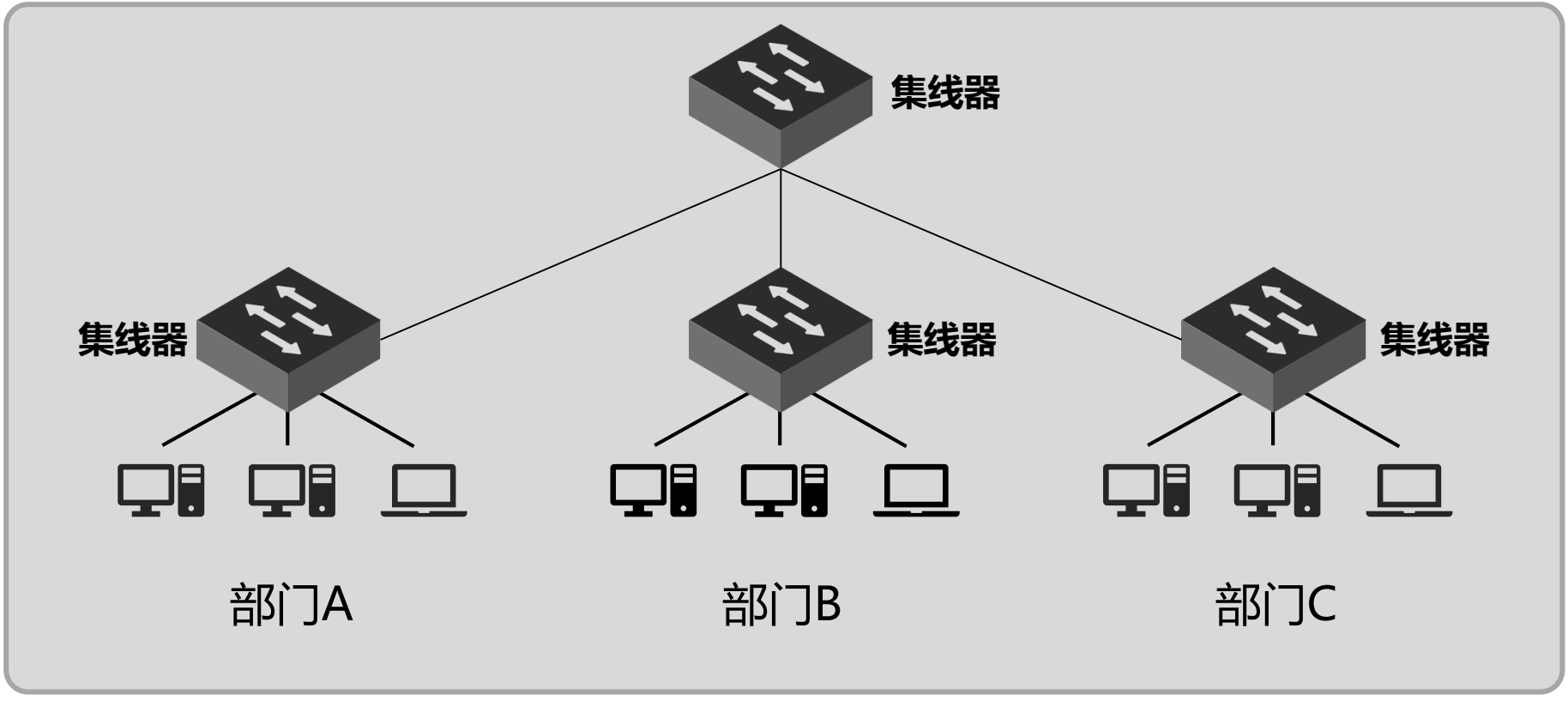
2. 网际协议 IP

- 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：

所在层次	中间设备
运输层及以上	网关 (gateway)
网络层	路由器 (router)
数据链路层	网桥或桥接器 (bridge) 交换机 (switch)
物理层	转发器 (repeater)

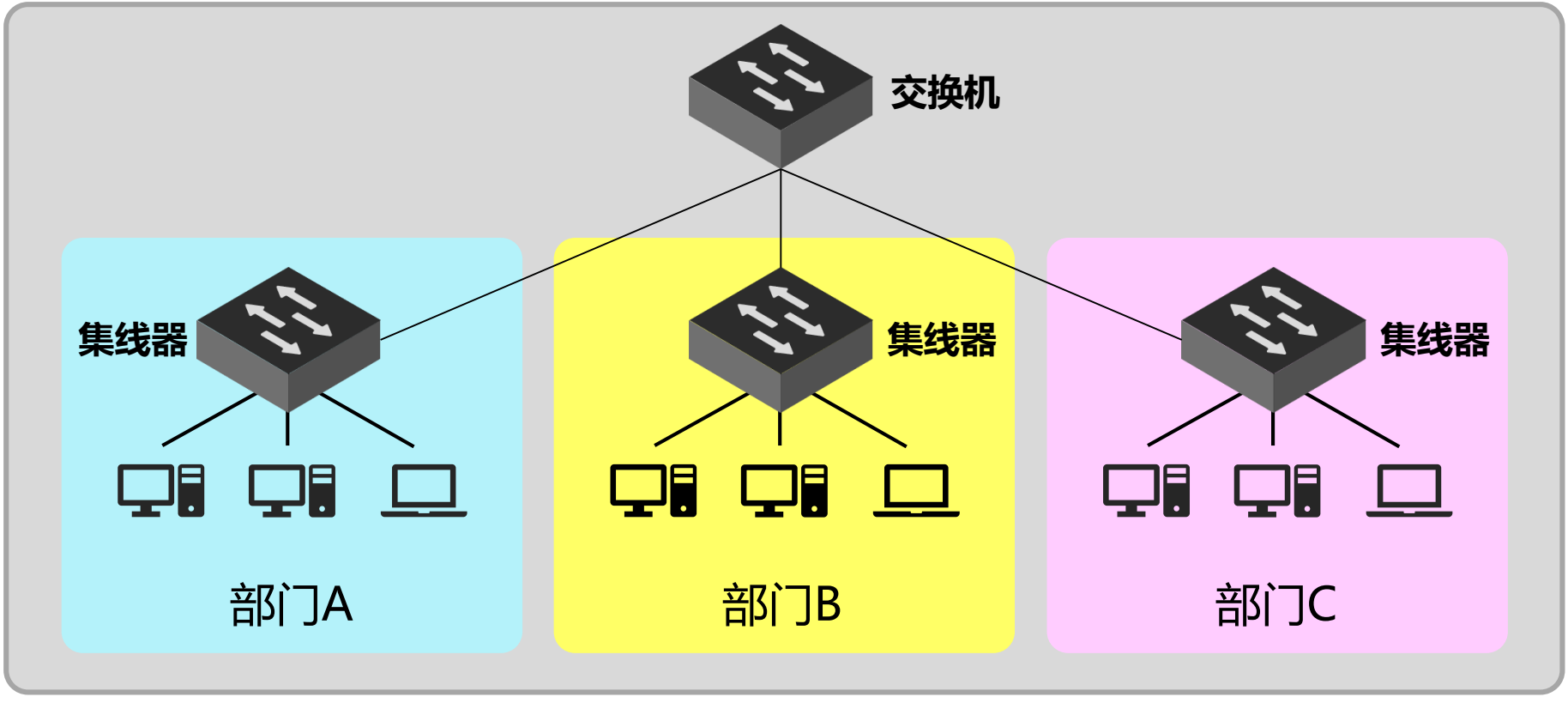
网络扩展

转发器、网桥或交换机仅把一个网络扩大了，仍然是一个网络



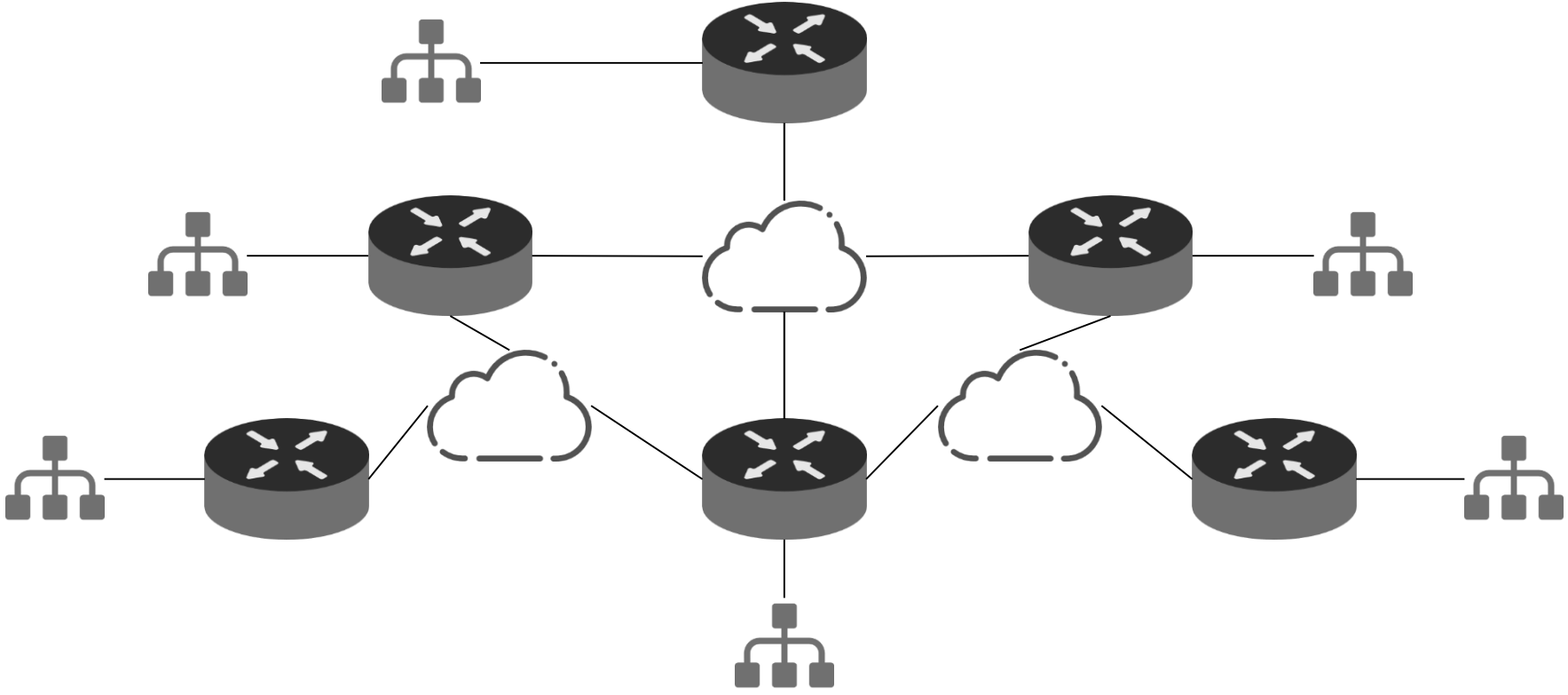
网络扩展

转发器、网桥或交换机仅把一个网络扩大了，仍然是一个网络



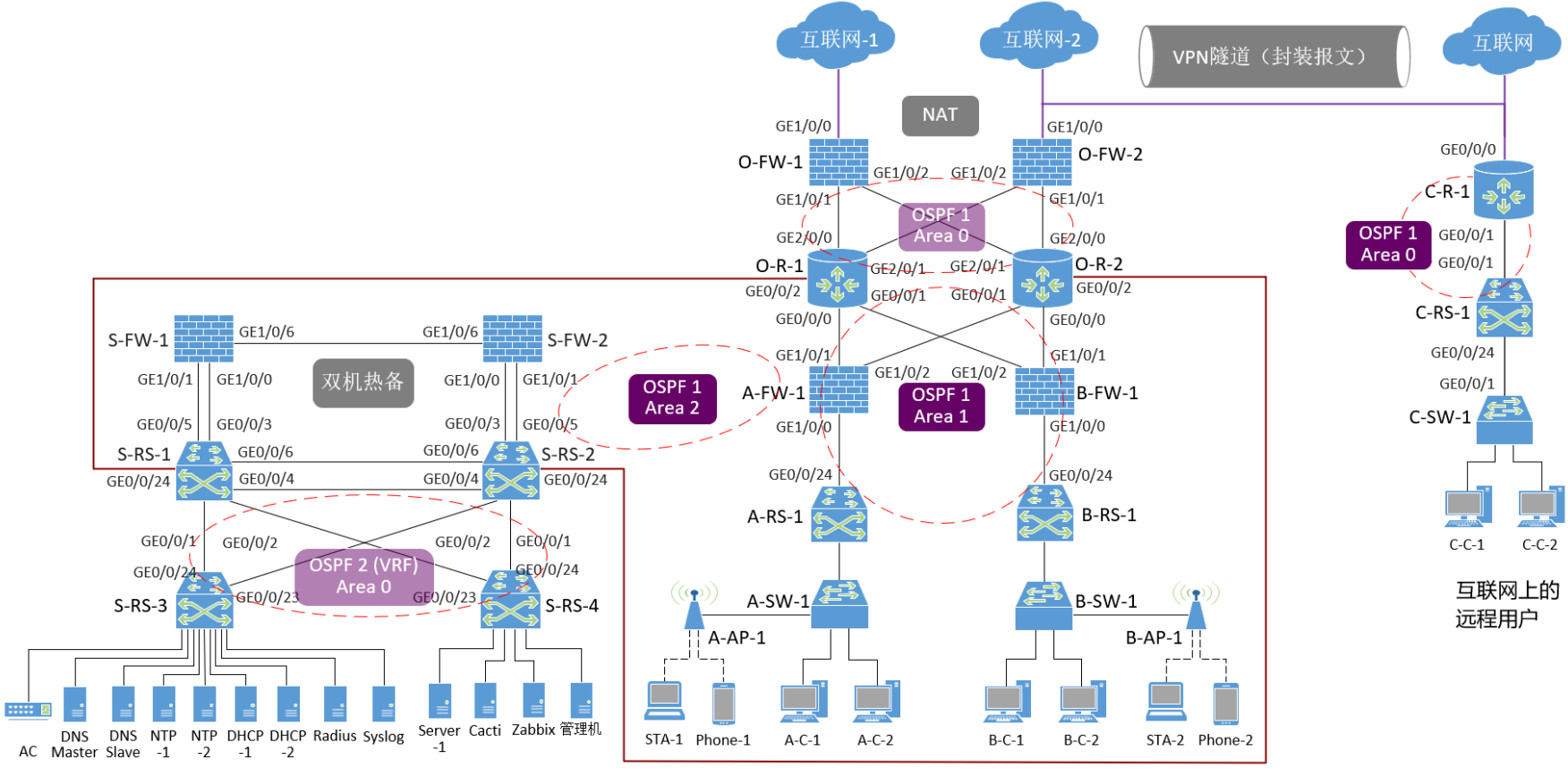
网络互联

网络互连使用路由器



网络互联

网络互连使用路由器



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

(a) 互连网络

(b) 虚拟互连网络

2. 网际协议 IP

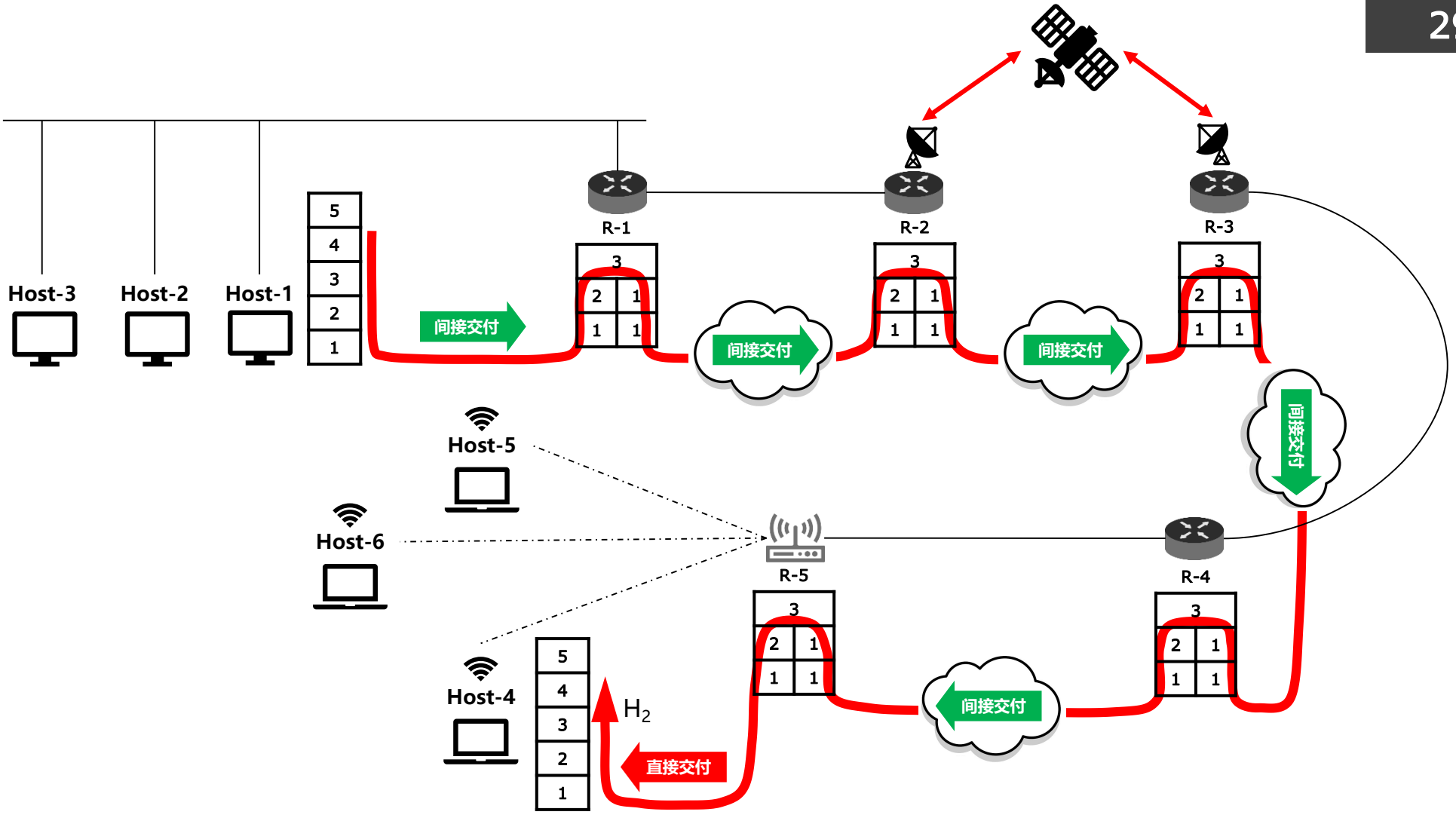
- TCP/IP 体系在网络互连上采用的做法是在网络层使用标准化协议，但相互连接的网络则可以是异构的。
- 参加互连的计算机网络都采用相同的网际协议（IP），因此可以把互连以后的计算机网络看成为一个虚拟互连网络（internet）。
- 当互联网上的主机进行通信时，就好像在一个网络上通信一样，看不见互连的各具体的网络异构细节。
- 如果在这种覆盖全球的 IP 网的上层使用 TCP 协议，那么就是现在的互联网 (Internet)。

2. 网际协议 IP

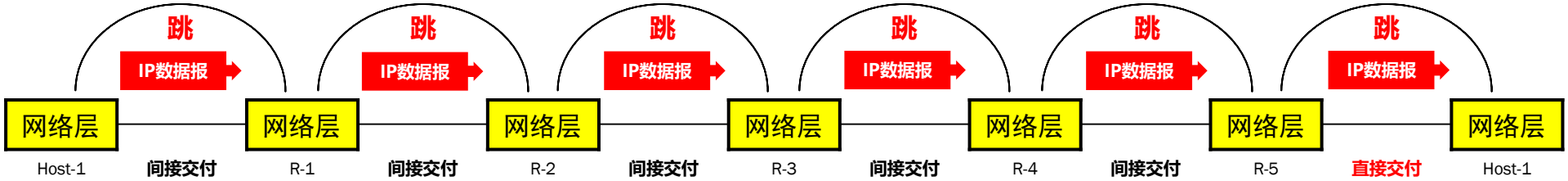
- 所谓虚拟互连网络，就是逻辑互连网络，它的意思是：
  - 互连起来的各种物理网络的异构性本来是客观存在的，但利用 IP 协议就可以使这些性能各异的网络从用户看起来好像是一个统一的网络。
  - 使用 IP 协议的虚拟互连网络可简称为IP网。
- 使用虚拟互连网络概念的好处是：
  - 当互联网上的主机进行通信时，就好像在一个网络上通信一样，而看不见互连的各具体的网络异构细节。

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。


Host-3
Host-2
Host-1
5
4
3
2
1
间接交付
R-1
R-2
R-3
3
2 1
1 1
间接交付
间接交付
间接交付
Host-5
Host-6
Host-4
R-5
R-4
H_2
直接交付
间接交付

2. 网际协议 IP



VisualRoute Overview

www.visualroute.com/overview.html

visualroute

Visualware®

VISUALROUTE DOWNLOAD PURCHASE COMPARE SUPPORT

VisualRoute
Network Tools and Diagnostics

You are here: Visualware > VisualRoute > VisualRoute Overview

VisualRoute Overview

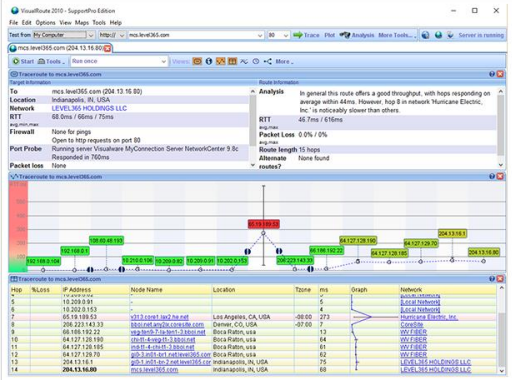
VisualRoute offers a wide variety of network tools that help users keep one step ahead of network issues such as bottle necks and packet loss/latency issues.

Click on the tools listed below for more information on how they can help.

- Continuous trace routing
- Reverse tracing
- Response time graphing
- Port Probing
- Network scanning
- Trace route history
- Side by side trace route comparison
- Route analysis (NetVu)
- Custom maps
- Remote access server

Extra features include:

- Whois Lookups
- IP Locations
- Traceroute tests from Visualware servers
- IPv6 Compatibility



VisualRoute

Home
Online Testing Portal
Download
Purchase
Version Comparison
Support

Visualware Products

MyConnection Server
MyConnection PC
eMailTrackerPro
Visual IP Trace

VisualRoute - Downloads

www.visualroute.com/download.html

Visualware®

VISUALROUTE DOWNLOAD PURCHASE COMPARE SUPPORT

VisualRoute
Network Tools and Diagnostics

You are here: Visualware > VisualRoute > Download Options

Download Options

***VisualRoute** has 4 versions but **ONLY ONE DOWNLOAD**. The license key unlocks the version purchased. The trial is a fully functional and lasts 15 days.

VisualRoute Full (all editions)

	2000/XP/2003/Vista/2008/7/8/Server 2012	4.4Mb	Download*
	Mac OS X (dmg) 10.3+, universal binary	4.4Mb	Download*

VisualRoute Lite (free for non-business use)

	2000/XP/2003/Vista/2008/7/8/Server 2012	4.4Mb	Download
	Mac OS X (dmg) 10.3+, universal binary	4.4Mb	Download

Learn more about VisualRoute Lite

Download FAQ

I'm upgrading, can I over install?

Requirements

System Requirements:

- Windows 2000/XP/2003/Vista/2008/7/8/Server 2012/10 or Mac OS X 10.4+
- 1Gb memory, 150Mb disk space, 1.5Ghz processor
- Internet Connection
- Sun Java Runtime Environment 1.7u25 or later. **NOTE**: currently only compatible with 32bit Java.

Remote Agent Platform (SupportPro Edition):

Windows XP/2003/Vista/2008/7/8/Server 2012/10

Java Requirements:

We recommend downloading the latest Java from here.

VisualRoute

Home
Online Testing Portal
Download
Purchase
Version Comparison
Support

Visualware Products

MyConnection Server
MyConnection PC
eMailTrackerPro
Visual IP Trace

VisualRoute Overvi... | VisualRoute - Down... | 4-1.计算机网络原理-... | 计算机网络第8版课件...

英 18:55

目标IP: 81.171.33.201 (www.centos.org@119.29.29.29)

本机网络 | www.centos.org | 119.29.29.29(DNSPOD) | + | 开始 | 清空

目标IP: 81.171.33.201 (www.centos.org@119.29.29.29)

同步请求 | TCP(80端口,IPv4) | 微软必应地图 | 导出

#	IP	时间(ms)	地址	AS	主机名
1	172.19.100.1	0 / 0 / 1	局域网	*	
2	192.168.179.1	0 / 0 / 0	局域网	*	
3	10.0.1.18	0 / 0 / *	局域网	*	
4	10.0.1.49	0 / 0 / 0	局域网	*	
5	42.228.55.161	0 / 0 / 16	中国 河南 郑州 chinaunicom.com 联通	AS4837	hn.kd.ny.adsl
6	61.168.12.81	0 / 1016 / *	中国 河南 郑州 chinaunicom.com 联通	AS4837	pc81.zz.ha.cn
7	61.168.30.145	0 / 0 / 0	中国 河南 郑州 chinaunicom.com 联通	AS4837	pc145.zz.ha.cn
8	219.158.108.205	16 / * / *	中国 北京 chinaunicom.com 联通	AS4837	
9	219.158.5.146	47 / * / *	中国 北京 chinaunicom.com 联通	AS4837	
10	219.158.3.30	47 / 47 / 47	中国 北京 chinaunicom.com 联通	AS4837	
11	219.158.12.170	156 / 171 / 219	德国 黑森州 法兰克福 chinaunicom.com 联通	AS4837	
12	219.158.42.234	250 / 265 / 281	德国 黑森州 法兰克福 chinaunicom.com 联通	AS4837	
13	84.116.137.18	266 / 281 / 281	德国 黑森州 法兰克福 libertyglobal.com	AS6830	de-fra02a-rc1-ae-30-0.aorta.net
14	84.116.130.150	265 / 281 / *	荷兰 北荷兰省 阿姆斯特丹 libertyglobal.com	AS6830	nl-ams02a-rc2-lag-11-0.aorta.net
15	84.116.130.97	250 / 281 / 282	荷兰 北荷兰省 阿姆斯特丹 libertyglobal.com	AS6830	nl-ams04a-ri3-ae-8-0.aorta.net
16	213.46.183.62	187 / 188 / 219	荷兰 北荷兰省 阿姆斯特丹 ziggo.nl	AS6830	
17	*	* / * / *			
18	81.171.33.201	203 / 203 / 282	荷兰 北荷兰省 阿姆斯特丹 eweka.nl	AS34343	

地址数据信息由 ipip.net 提供。

Visualware Newsle | 在线路由跟踪 地图 | best trace - 搜索 | Best Trace最专业的 | BestTrace 客户端 | App Store 上的"Be

https://www.ipip.net/product/client.html

IPIP 首页 产品价格 IP查询 技术支持 客户 关于 工具 | 登录 / 注册 | English

IPv4 数据库试用版 & 客户端工具

包括IPv4 数据库试用版及相关、BestTrace 软件下载、浏览器扩展

IP 数据库及相关

IP Free

试用版IP地址数据库

最后更新时间：20190703
试用版数据不含运营商，国外只具体到国家
试用版不能作为正式商业用途

IPv4 库试用版 下载

IP E&D

更精准的IP地址数据库

最后更新时间：2022-02-26
(每工作日更新)

更精准的IP数据库

IP API

http://freeapi.ipip.net/8.8.8.8

开放免费数据接口
限速为单 IP 每秒最多 5 次请求

文档

BestTrace 应用下载

Windows	Android	iOS	Mac	Linux
V3.8.0	V1.9.0	V1.39	V1.17	V1.2.0
从 Windows 设备上发起 traceroute 请求，附带链路可视化。	从你的 Android 设备上发起 traceroute 请求，附带链路可视化。	从你的 iPhone / iPad 上发起 traceroute 请求，附带链路可视化。	从你的 Mac 电脑上发起 traceroute 请求，附带链路可视化。	从你的 Linux(X86/ARM)/Mac/BSD 系统环境下发起 traceroute 请求，附带链路可视化，兼

2. 网际协议 IP

- 在 TCP/IP 体系中，IP 地址是一个最基本的概念。
- 没有 IP 地址，就无法和网上的其他设备进行通信。
- 本部分重点：
  - IP 地址及其表示方法
  - 分类的 IP 地址
  - 无分类编址 CIDR
  - IP 地址的特点

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：32 位二进制代码

10000000000010110000001100011111

分为每 8 位为一组

10000000 00001011 00000011 00011111

将每 8 位的二进制数
转换为十进制数

128 11 3 31

采用点分十进制记法

128.11.3.31

2. 网际协议 IP

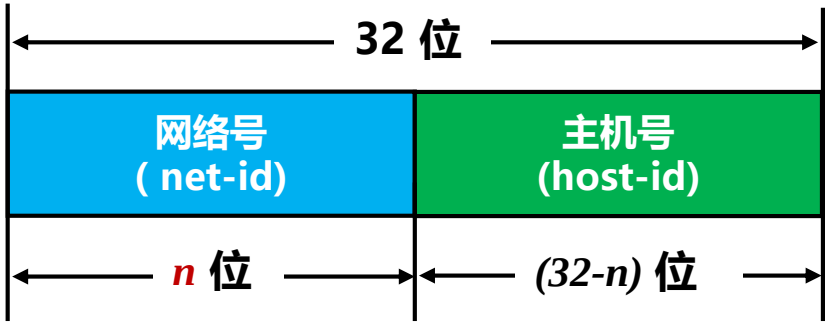
IP 地址采用 2 级结构

2 个字段：网络号和主机号

IP 地址 ::= { <网络号>, <主机号>}

IP地址在整个互联网范围内是唯一的。
IP 地址指明了连接到某个网络上的一个主机

网络号的位数 *n* 是多少?



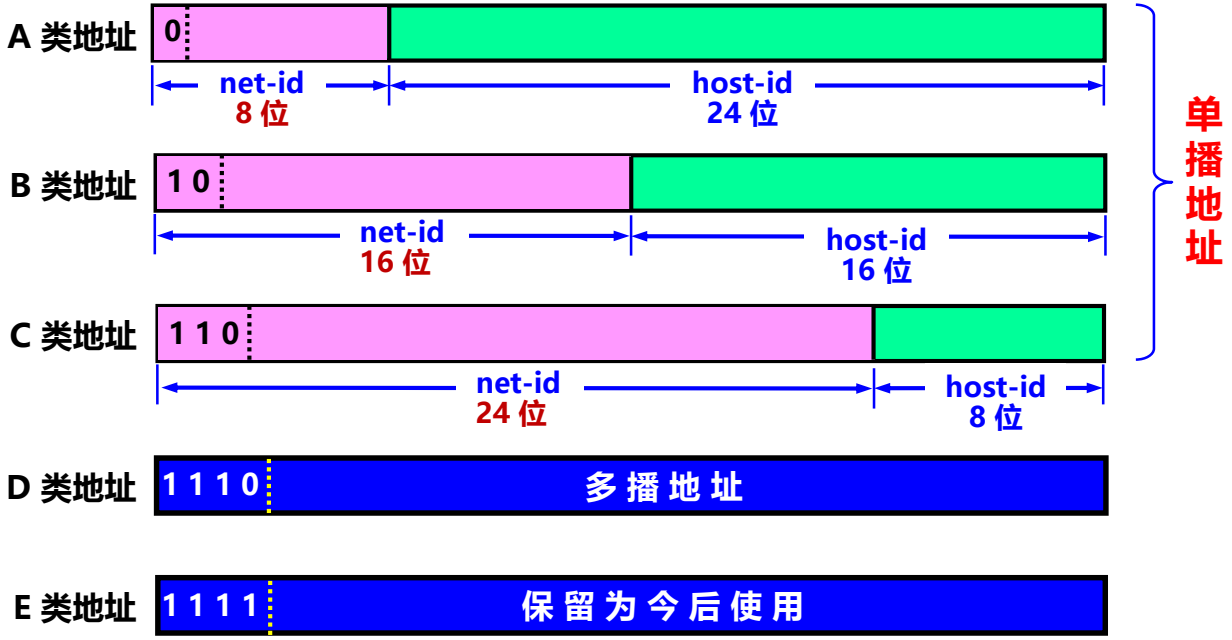
2. 网际协议 IP

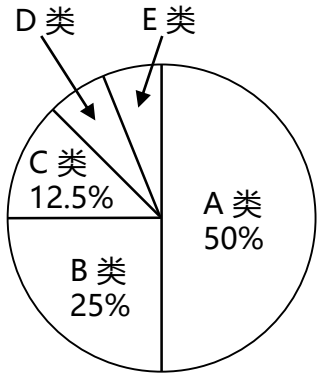
分类的 IP 地址

- 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





2. 网际协议 IP

各类 IP 地址的指派范围

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

注意：

- ✓ A 类网络地址中， 网络号 0 和 127 是保留地址，不指派。0 表示"本网络"，127 保留作为本地环回测试地址。
- ✓ B 类网络地址中，网络号 128.0 是被 IANA 保留的，不指派。采用无分类编址（CIDR）时可以指派。
- ✓ C 类网络地址中，网络号 192.0.0 是被 IANA 保留的，不指派。采用无分类编址（CIDR）时可以指派。
- ✓ 指派主机号时，要**扣除**全 0 和全 1 。全 0 和全 1 有特殊含义和用途。

2. 网际协议 IP

一般不使用的特殊的 IP 地址

网络号	主机号	源地址使用	目的地址使用	代表的意思
0	0	可以	不可	在本网络上的本主机（见 6.6 节 DHCP 协议）
0	X	可以	不可	在本网络上主机号为 X 的主机
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）
Y	全 1	不可	可以	对网络号为 Y 的网络上的所有主机进行广播
127	非全 0 或全 1 的任何数	可以	可以	用于本地软件环回测试

2. 网际协议 IP

分类的 IP 地址的优点和缺点

管理简单；

使用方便；

转发分组迅速；

划分子网，灵活地使用。

设计上不合理：

大地址块，浪费地址资源；

即使采用划分子网的方法，也无法解决 IP 地址枯竭的问题。

2. 网际协议 IP

无分类编址 CIDR

- CIDR (Classless Inter-Domain Routing)： 无分类域间路由选择。
- 消除了传统的 A 类、B 类和 C 类地址以及划分子网的概念，可以更加有效地分配 IPv4 的地址空间，但无法解决 IP 地址枯竭的问题。
- 要点：
  1. 网络前缀
  2. 地址块
  3. 地址掩码

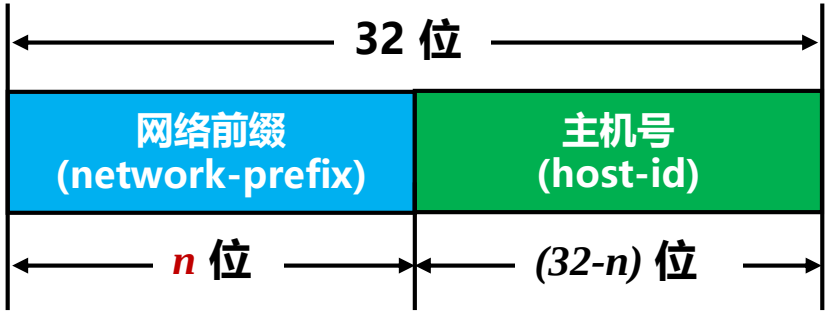
2. 网际协议 IP

无分类编址 CIDR 的网络前缀

2 个字段：网络前缀和主机号

IP 地址 ::= { <网络前缀>, <主机号>}

网络前缀的位数 *n* 是多少?



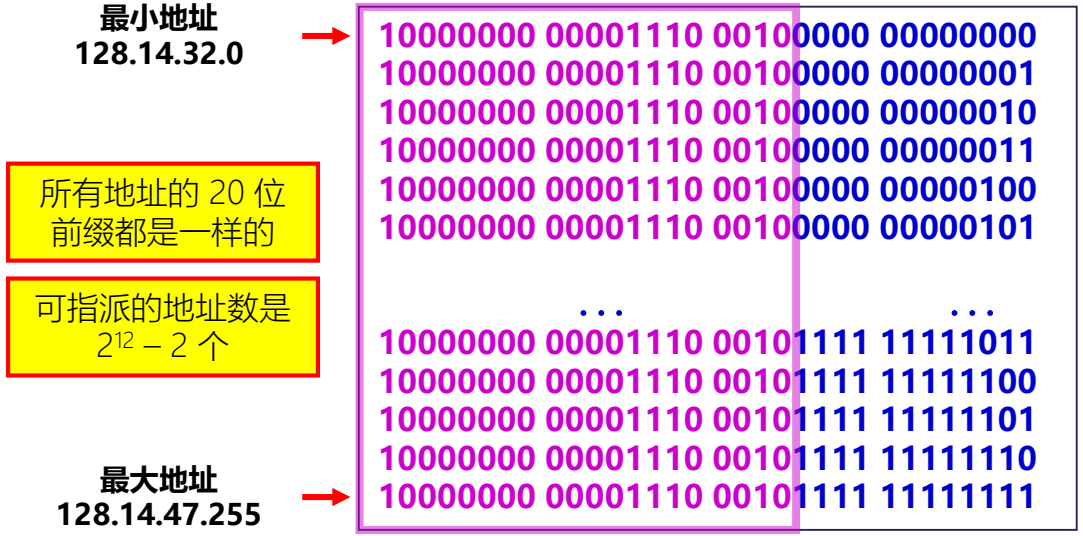
与分类的 IP 地址最大的区别：
前缀的位数 n 不固定，可以在 0 ~ 32 之间选取任意值。

CIDR 记法：斜线记法 (slash notation)
a.b.c.d / n：二进制 IP 地址的前 n 位是网络前缀。
例如：128.14.35.7/20：前 20 位是网络前缀。

2. 网际协议 IP

无分类编址 CIDR 的地址块

128.14.32.0/20 组成的地址块（共 2^{12} 个地址）



二进制代码表示：10000000 00001110 0010*

- ✓ CIDR 把**网络前缀都相同的所有连续的** IP 地址组成一个 CIDR 地址块。
- ✓ 一个 CIDR 地址块包含的 IP 地址数目，取决于网络前缀的位数。

2. 网际协议 IP

- 无分类编址 CIDR
  - 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。

2. 网际协议 IP

无分类编址 CIDR 的地址块

128.14.35.7/20	是 IP 地址，同时指明了网络前缀为 20 位。 该地址是 128.14.32.0/20 地址块中的一个地址。
128.14.32.0/20	是包含有多个 IP 地址的地址块，同时也是这个地址块中主机号为全 0 的 IP 地址。
128.14.35.7	是 IP 地址，但未指明网络前缀长度，不知道其网络地址。
128.14.32.0	不能指明一个网络地址，因为无法知道网络前缀是多少。

2. 网际协议 IP

无分类编址 CIDR 的地址掩码 (address mask)

- ✓ 又称为子网掩码 (subnet mask)。
- ✓ 位数：32 位。
- ✓ 目的：让机器从 IP 地址迅速算出网络地址。
- ✓ 形式：由一连串 1 和接着的一连串 0 组成， 1 的数量即网络前缀的长度。

/20 地址块：

- **地址掩码：11111111 11111111 11110000 00000000**
- **点分十进制记法：255.255.240.0**
- **CIDR 记法：255.255.240.0/20**

2. 网际协议 IP

无分类编址 CIDR 的默认地址掩码

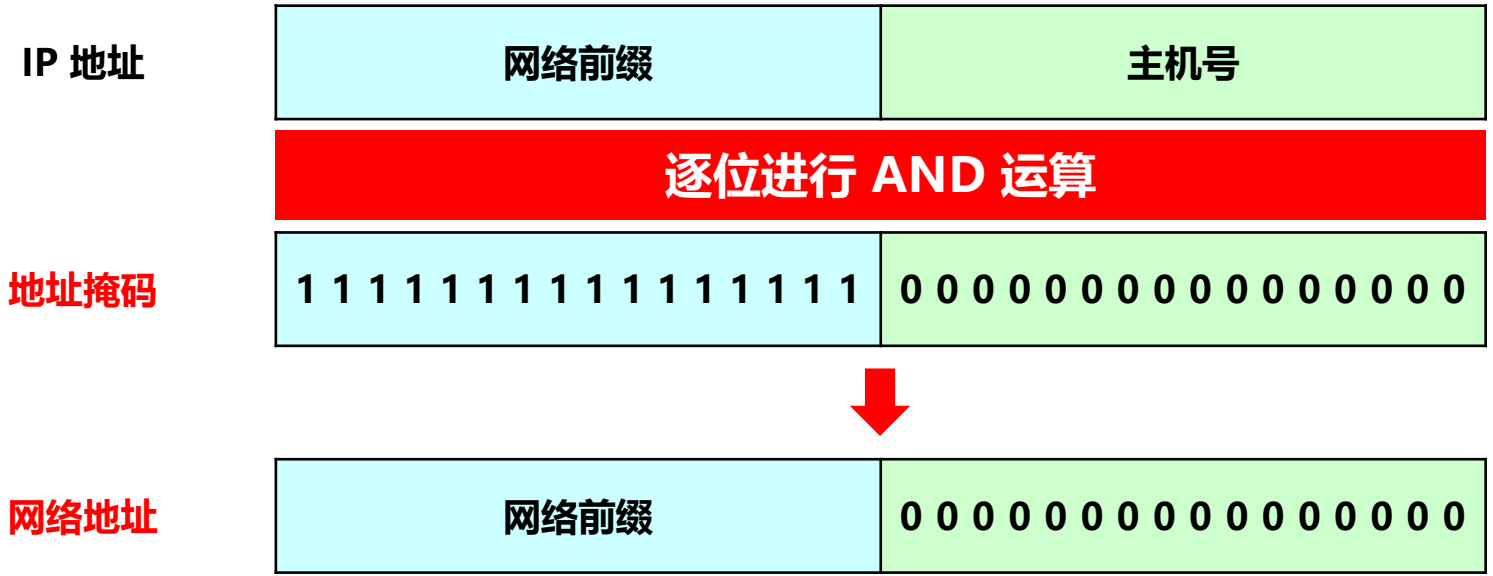
A类地址	默认地址掩码 255.0.0.0	网络号	主机号
		1 1 1 1 1 1 1 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

B类地址	默认地址掩码 255.255.0.0	网络号	主机号
		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	默认地址掩码 255.255.255.0	网络号	主机号
		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

2. 网际协议 IP

网络地址 = (二进制的 IP 地址) AND (地址掩码)



2. 网际协议 IP

【例】已知 IP 地址是 128.14.35.7/20。求网络地址。

(a) 点分十进制 IP 地址	128	14	35	7
(b) 二进制 IP 地址	10000000	00001110	00100011	00000111
(c) 地址掩码是 255.255.224.0	11111111	11111111	1111 0000	00000000
(d) IP 地址与地址掩码按位 AND	10000000	00001110	0010 0000	00000000
(e) 网络地址（点分十进制）	128	14	32	0

常用的CIDR地址块

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

2. 网际协议 IP

构建超网

- 每一个 CIDR 地址块中的地址数一定是 2 的整数次幂。
- 除最后几行外，CIDR 地址块都包含了多个 C 类地址（是一个 C 类地址的 2n 倍，n 是整数）。
- 因此在文献中有时称 CIDR 编址为“构造超网”。

三个特殊的 CIDR 地址块

网络前缀长度	点分十进制	说明
/32	255.255.255.255	就是一个 IP 地址。这个特殊地址用于主机路由
/31	255.255.255.254	只有两个 IP 地址，其主机号分别为 0 和 1。这个地址块用于点对点链路
/0	0.0.0.0	同时 IP 地址也是全 0，即 0.0.0.0/0。用于默认路由。

2. 网际协议 IP

路由聚合

聚合前

16 个 C 类地址，
地址掩码=255.255.255.0，
路由表中需要 16 个路由项目。

192.24.0
192.24.1
192.24.2
192.24.3
192.24.4
192.24.5
192.24.6
192.24.7
192.24.8
192.24.9
192.24.10
192.24.11
192.24.12
192.24.13
192.24.14
192.24.15

聚合后

聚合为 1 个 地址，
地址掩码=255.255.240.0，
路由表中只需 1 个路由项目。

192.24.0.0 /20

2. 网际协议 IP

- IP 地址的特点：

(1) 每个 IP 地址都由网络前缀和主机号两部分组成。

(2) IP 地址是标志一台主机（或路由器）和一条链路的接口。

(3) 转发器或交换机连接起来的若干个局域网仍为一个网络

(4) 在 IP 地址中，所有分配到网络前缀的网络都是平等的。

IP 地址是一种分等级的地址结构。

- 方便了 IP 地址的分配和管理。
- 实现路由聚合，减小了转发表所占的存储空间，以及查找转发表的时间。

2. 网际协议 IP

- IP 地址的特点：

(1) 每个 IP 地址都由网络前缀和主机号两部分组成。

(2) IP 地址是标志一台主机（或路由器）和一条链路的接口。

(3) 转发器或交换机连接起来的若干个局域网仍为一个网络

(4) 在 IP 地址中，所有分配到网络前缀的网络都是平等的。

- 当一台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，其网络号必须是不同的。
- 这种主机称为多归属主机 (multihomed host)。
- 一个路由器至少应当连接到两个网络，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 地址的特点：

(1) 每个 IP 地址都由网络前缀和主机号两部分组成。

(2) IP 地址是标志一台主机（或路由器）和一条链路的接口。

(3) 转发器或交换机连接起来的若干个局域网仍为一个网络

(4) 在 IP 地址中，所有分配到网络前缀的网络都是平等的。

- 按照互联网的观点，一个网络（或子网）是指具有相同网络前缀的主机的集合。
- 转发器或交换机连接起来的若干个局域网都具有同样的网络号，它们仍为一个网络。
- 具有不同网络号的局域网必须使用路由器进行互连。

2. 网际协议 IP

- IP 地址的特点：

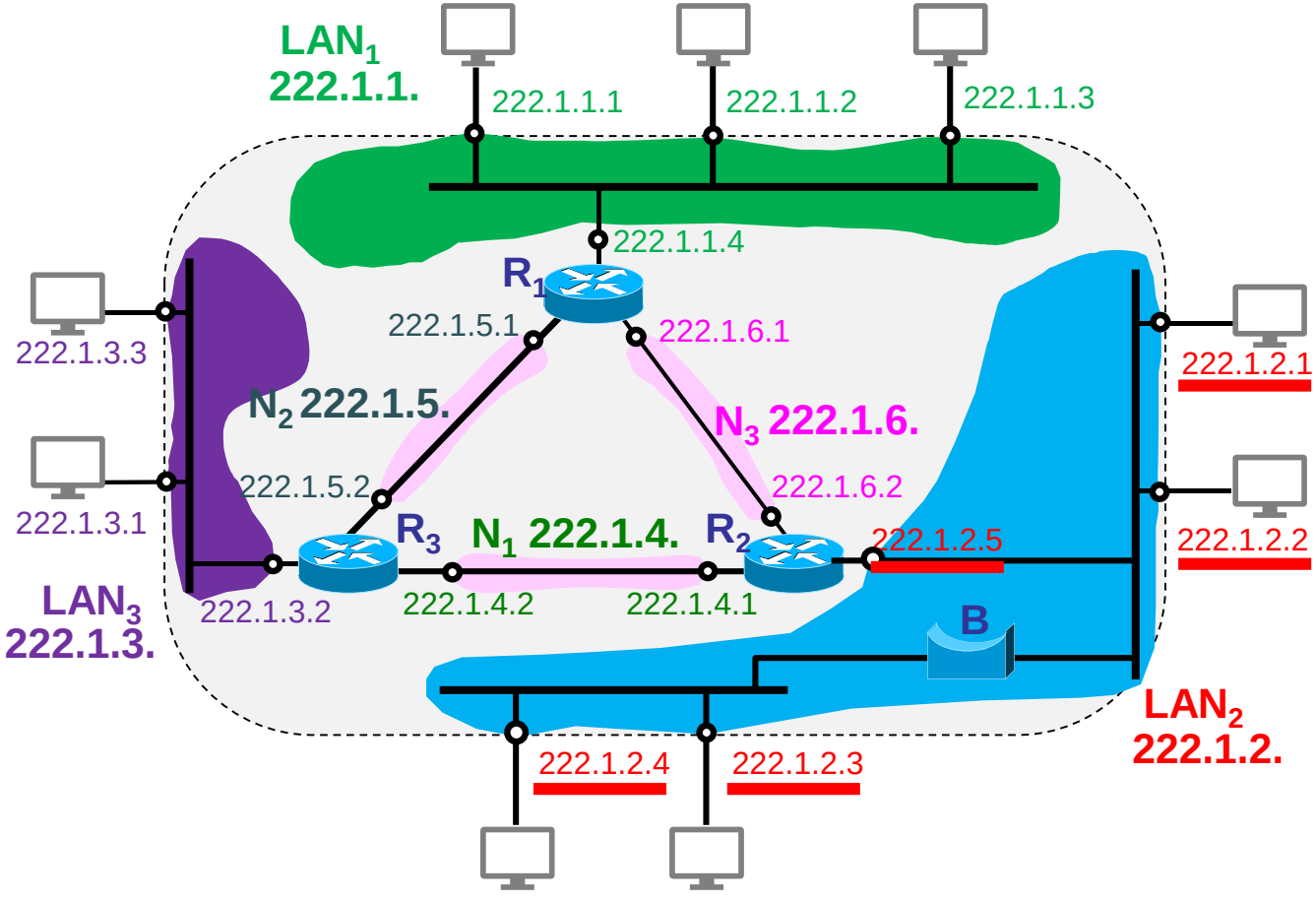
(1) 每个 IP 地址都由网络前缀和主机号两部分组成。

(2) IP 地址是标志一台主机（或路由器）和一条链路的接口。

(3) 转发器或交换机连接起来的若干个局域网仍为一个网络

(4) 在 IP 地址中，所有分配到网络前缀的网络都是平等的。

- 互联网同等对待每一个 IP 地址，不管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



① 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一样。

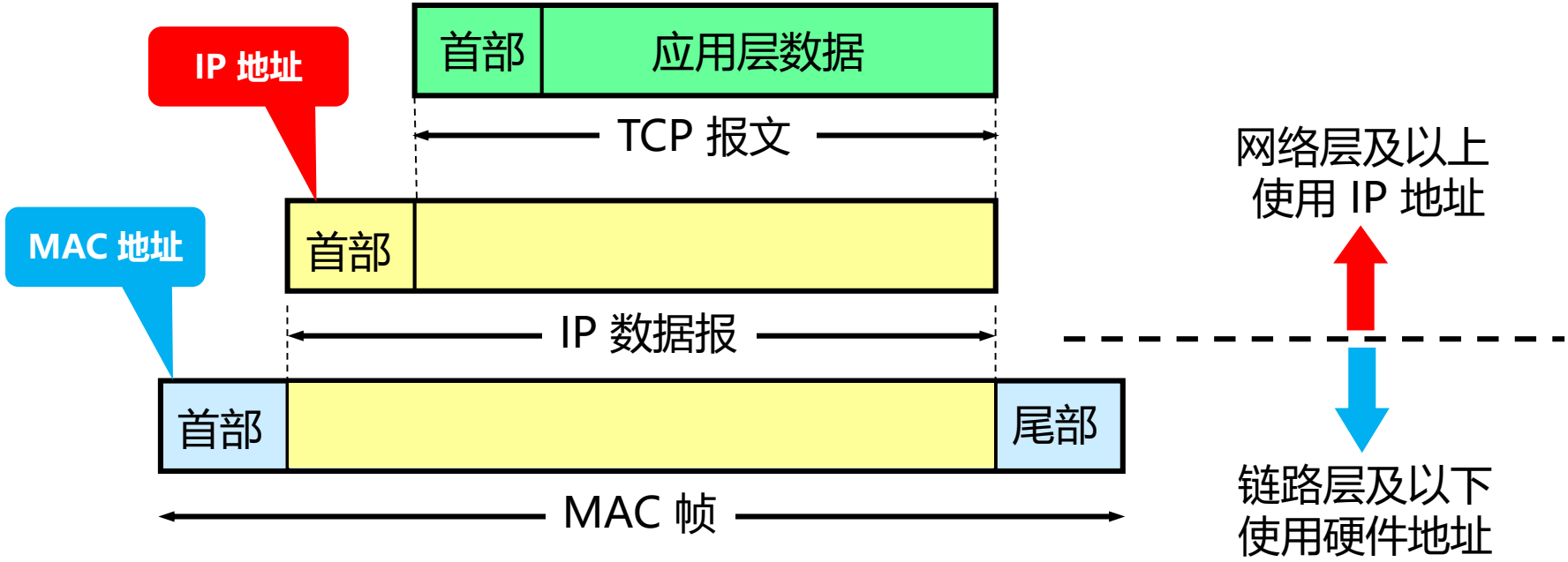
② 路由器的每一个接口都有一个不同网络号的 IP 地址。

③ 两个路由器直接相连的接口处，可指明也可不指明 IP 地址。

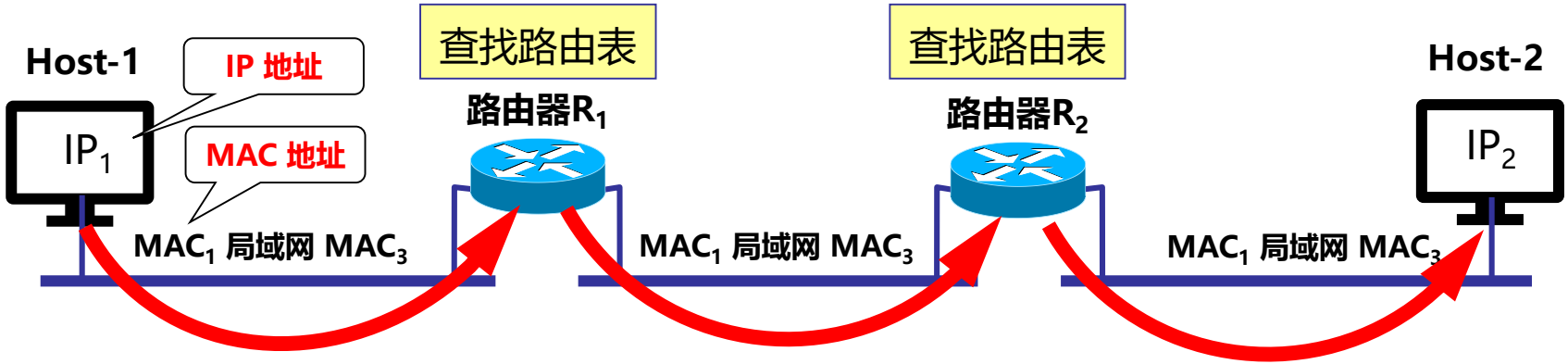
④ 如指明 IP 地址，则这一段连线就构成了一种只包含一段线路的特殊“网络”。这种网络仅需两个 IP 地址，可以使用 /31 地址块。主机号可以是 0 或 1。

2. 网际协议 IP

IP 地址	MAC 地址
• 虚拟地址、软件地址、逻辑地址。 • 网络层和以上各层使用。 • 放在 IP 数据报的首部。	• 固化在网卡上的 ROM 中。 • 硬件地址、物理地址。 • 数据链路层使用。 • 放在 MAC 帧的首部。


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

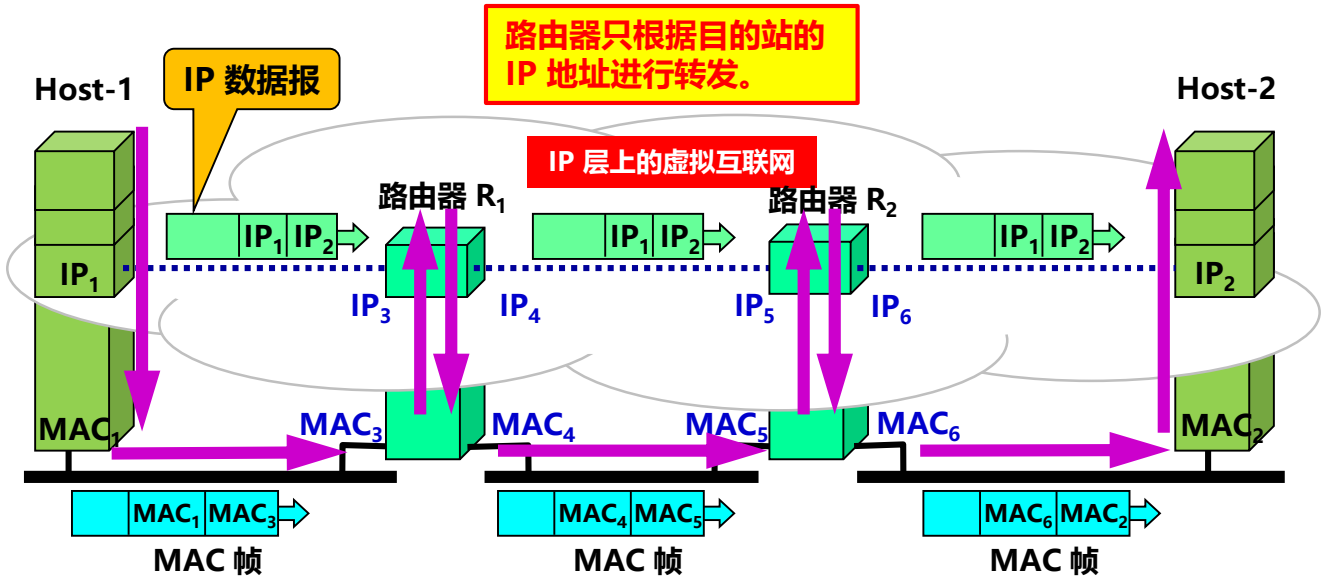
2. 网际协议 IP



通信的路径

H_1→经过 R_1 转发→经过 R_2 转发→H_2

从协议栈的层次上看 IP 地址和 MAC 地址



在局域网的链路层，只能看见 MAC 帧。

2. 网际协议 IP

不同层次、不同区间使用的源地址和目的地址

	在网络层 写入 IP 数据报首部的地址		在数据链路层 写入 MAC 帧首部的地址	
	源地址	目的地址	源地址	目的地址
从 Host-1 到 R_1	IP_1	IP_2	MAC_1	MAC_3
从 R_1 到 R_2	IP_1	IP_2	MAC_4	MAC_5
从 R_2 到 Host-2	IP_1	IP_2	MAC_6	MAC_2

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

主机或路由器怎样知道应当在 MAC 帧的首部填入什么样的 MAC 地址?

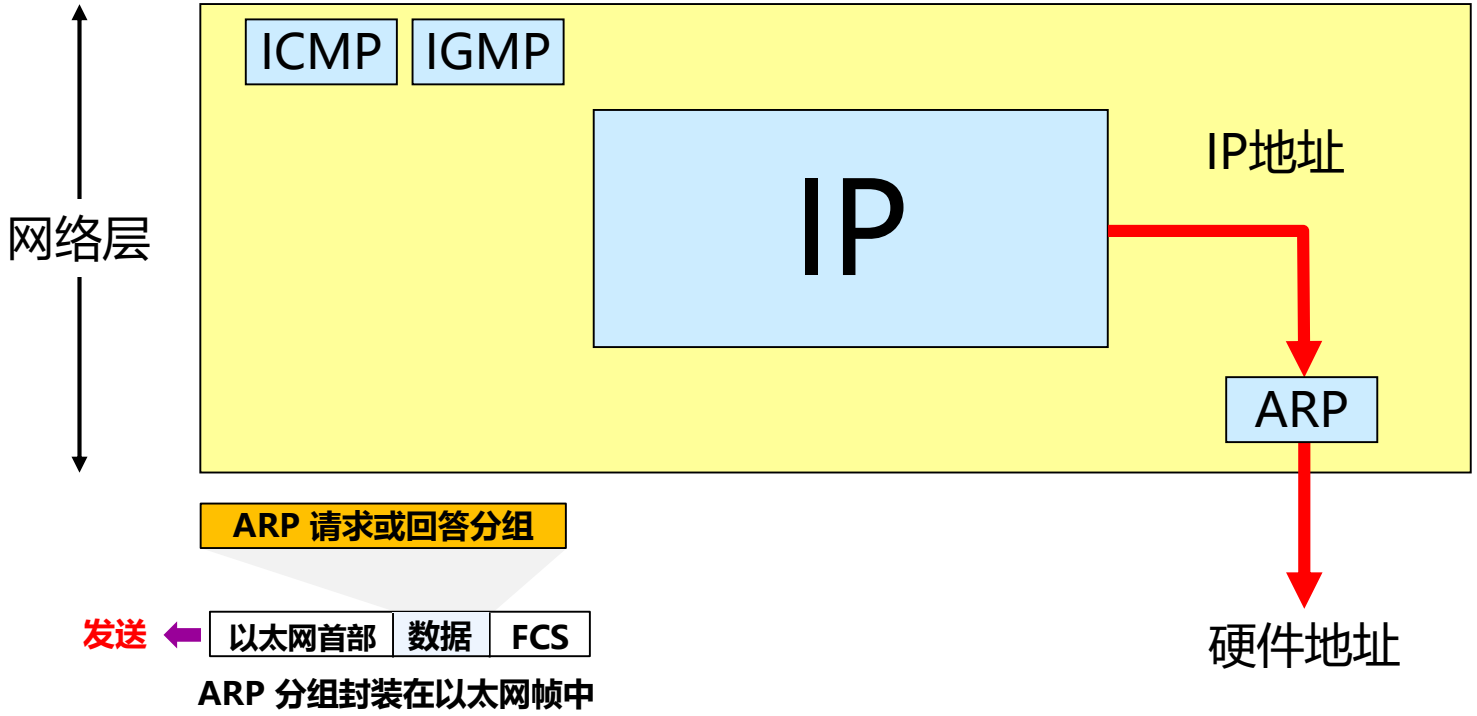
2. 网际协议 IP

主机或路由器怎样知道应当在 MAC 帧的首部填入什么样的 MAC 地址?

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。
- 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


网络层
ICMP
IGMP
IP
IP地址
ARP
硬件地址
ARP 请求或回答分组
发送
以太网首部
数据
FCS
ARP 分组封装在以太网帧中

2. 网际协议 IP

ARP 高速缓存 (ARP cache)

- 每一个主机都设有一个ARP高速缓存(ARP cache)
  - 存放 IP 地址到 MAC 地址的映射表。

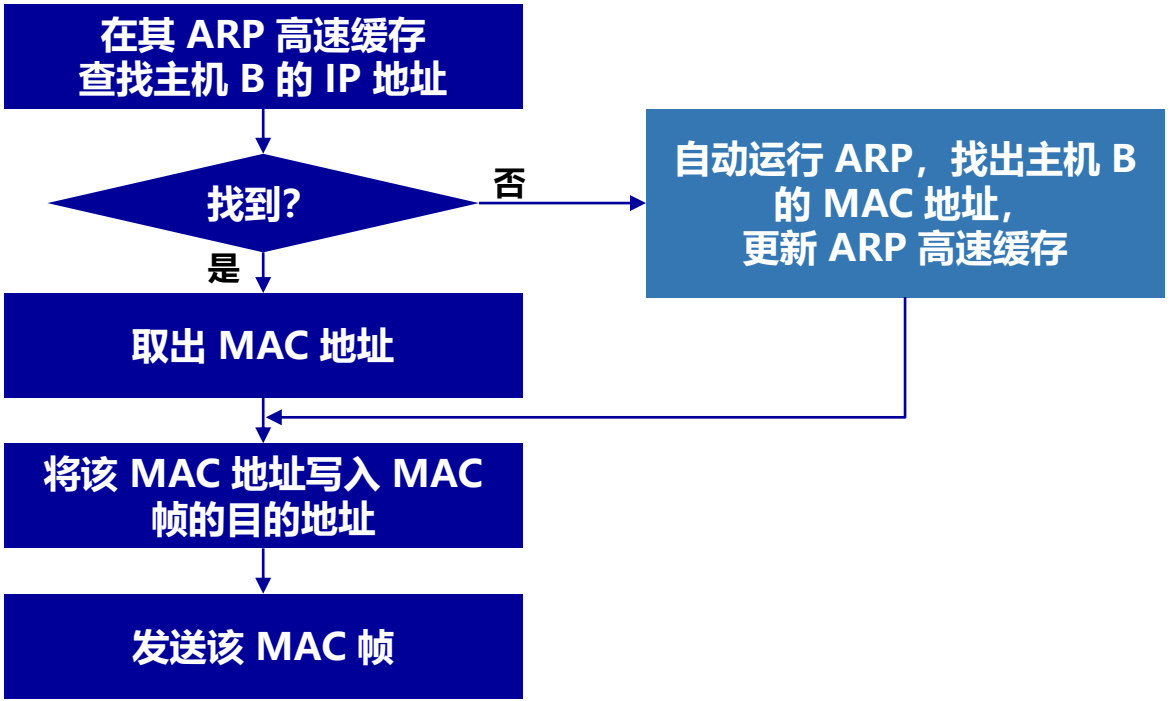
映射表 **< IP 地址；MAC 地址；生存时间 (Age)；类型等 >**

IP 地址	MAC 地址	生存时间 (Age)	类型	其他
10.4.9.2	0030.7131.abfc	00:08:55	Dynamic	
10.4.9.1	0000.0c07.ac24	00:02:55	Dynamic	
10.4.9.99	0007.ebea.44d0	00:06:12	Dynamic	

  - 映射表动态更新（新增或超时删除），超过生存时间的项目都从高速缓存中删除，以适应网络适配器变化。

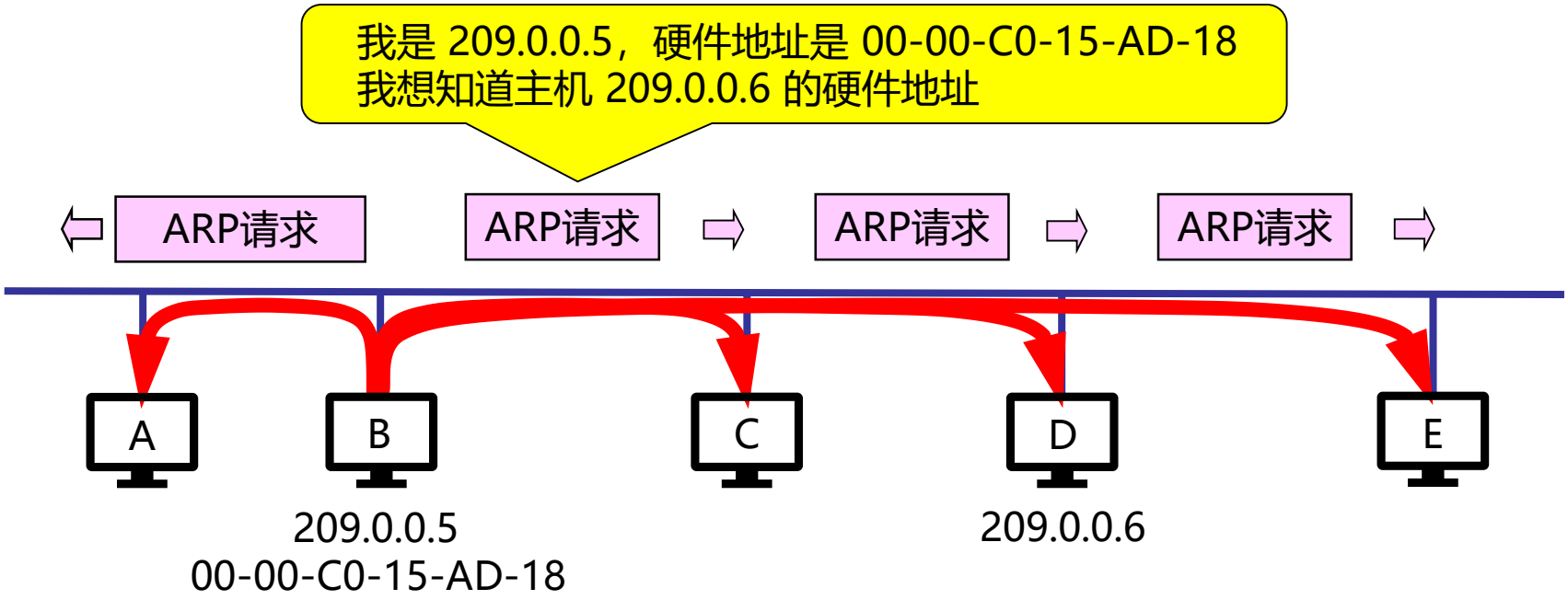
2. 网际协议 IP

当主机 A 欲向本局域网上的某个主机 B 发送 IP 数据报时：



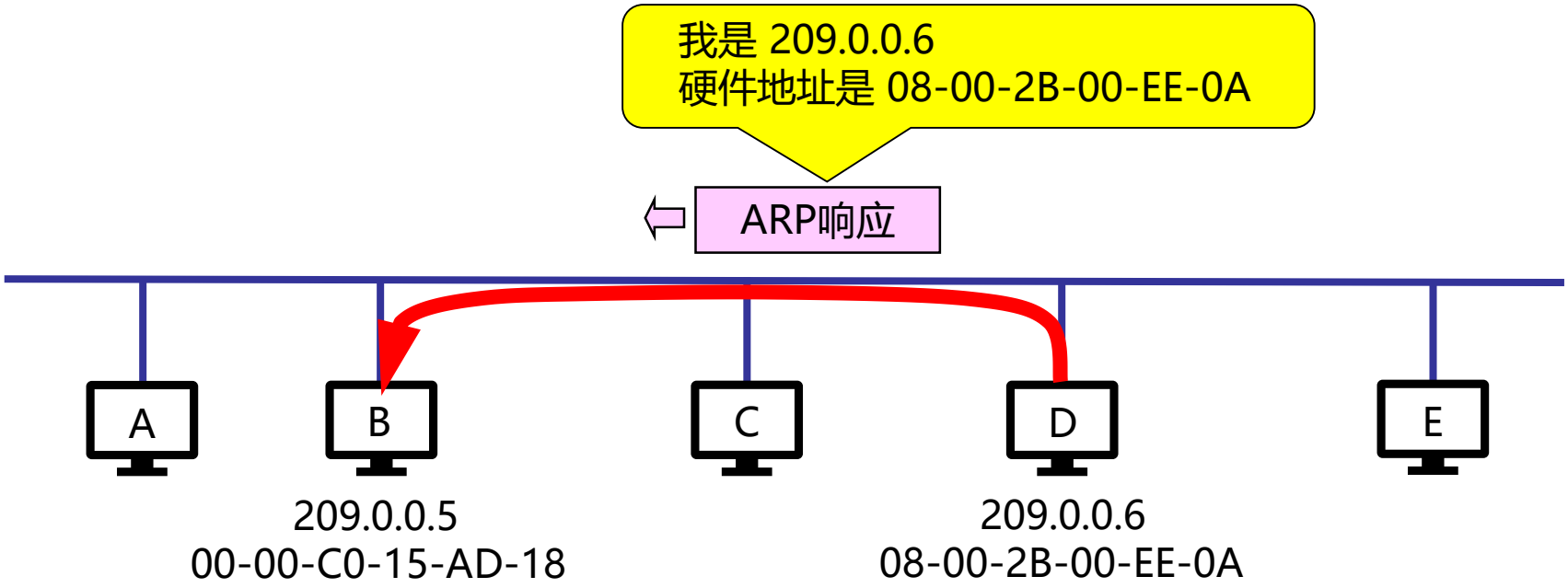
2. 网际协议 IP

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

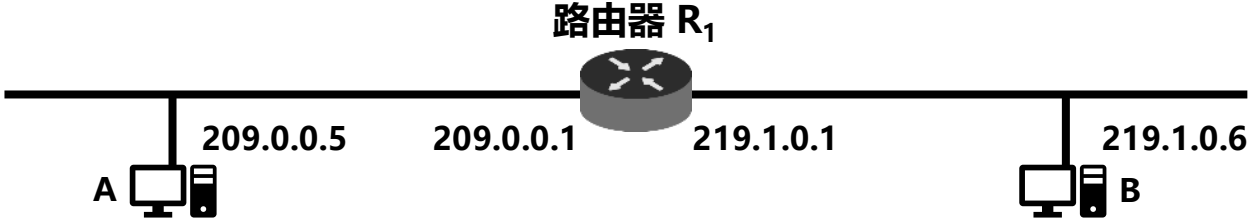


2. 网际协议 IP

- ARP 高速缓存 (ARP cache)
  - 存放最近获得的 IP 地址到 MAC 地址的绑定。
  - 可以减少 ARP 广播的通信量。
  - 为进一步减少 ARP 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 MAC 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 IP 地址及其对应的 MAC 地址映射写入主机 B 自己的 ARP 高速缓存中，不必再发送 ARP 请求。

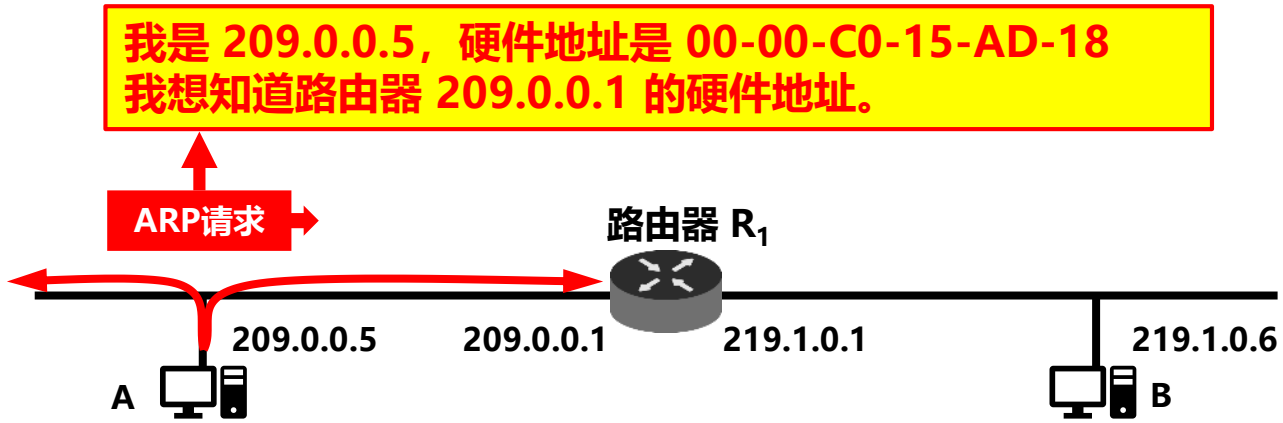
2. 网际协议 IP

- 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。

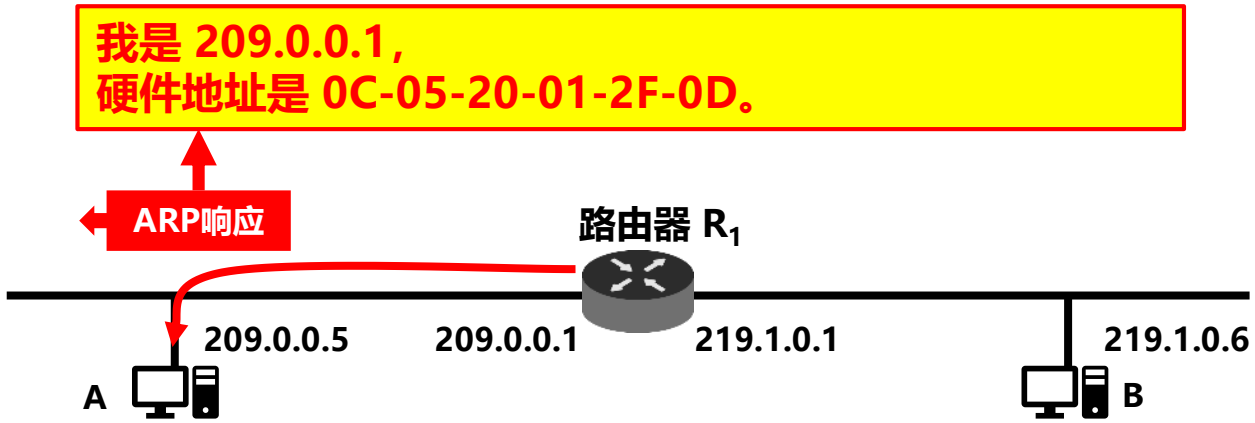


通信的路径：A → 经过 R_1 转发 → B。
因此主机 A 必须知道路由器 R_1 的 IP 地址，解析出其 MAC 地址。然后把 IP 数据报传送到路由器 R_1。

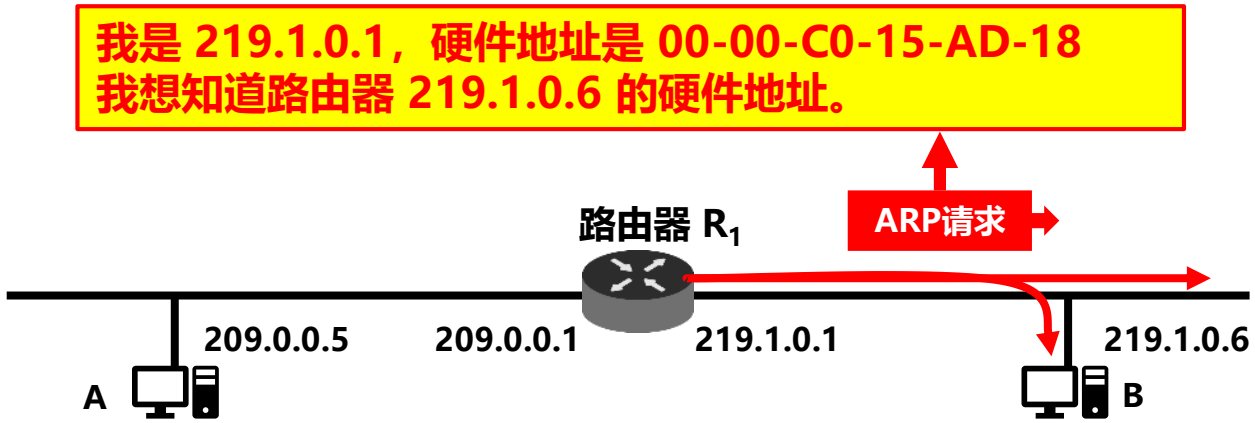
2. 网际协议 IP


我是 209.0.0.5，硬件地址是 00-00-C0-15-AD-18
我想知道路由器 209.0.0.1 的硬件地址。
ARP请求
路由器 R1
209.0.0.5
209.0.0.1
219.1.0.1
219.1.0.6
A
B

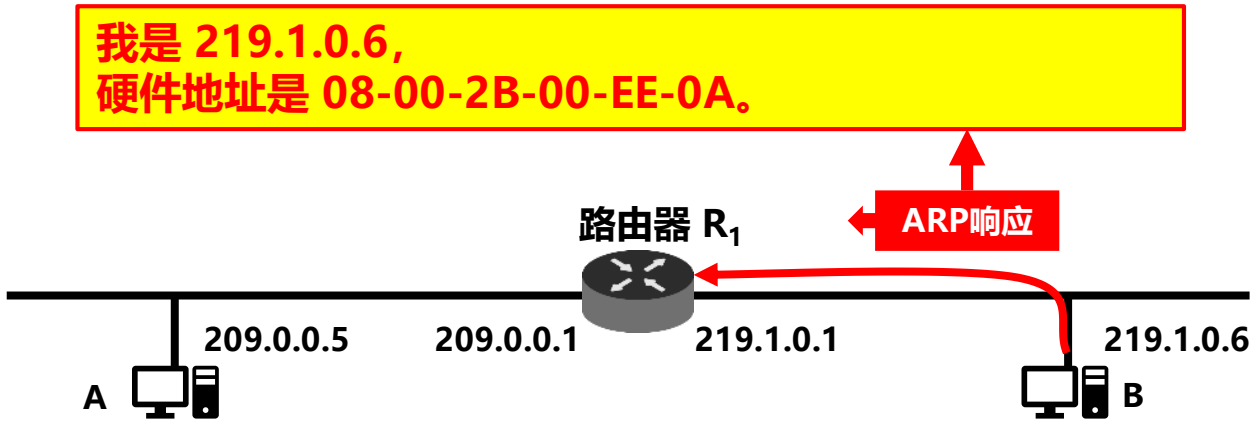
2. 网际协议 IP


我是 209.0.0.1，
硬件地址是 0C-05-20-01-2F-0D。
ARP响应
路由器 R_1
209.0.0.5
209.0.0.1
219.1.0.1
219.1.0.6
A
B

2. 网际协议 IP

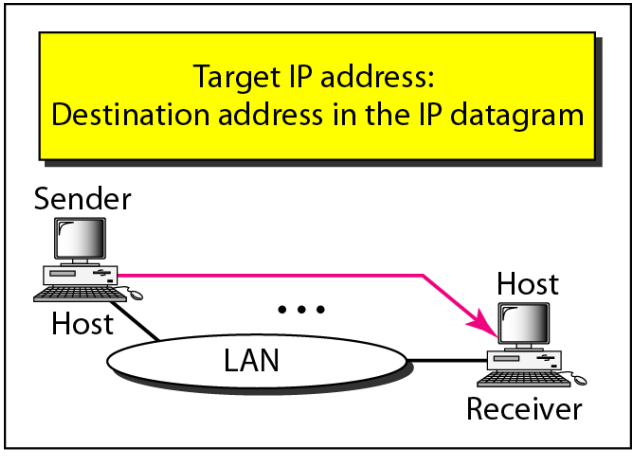

我是 219.1.0.1，硬件地址是 00-00-C0-15-AD-18
我想知道路由器 219.1.0.6 的硬件地址。
ARP请求
路由器 R1
209.0.0.5
209.0.0.1
219.1.0.1
219.1.0.6
A
B

2. 网际协议 IP

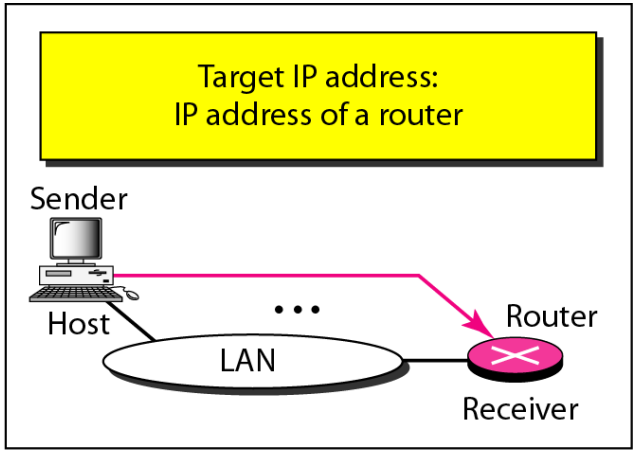

我是 219.1.0.6，
硬件地址是 08-00-2B-00-EE-0A。
ARP响应
路由器 R1
209.0.0.5
209.0.0.1
219.1.0.1
219.1.0.6
A
B

2. 网际协议 IP

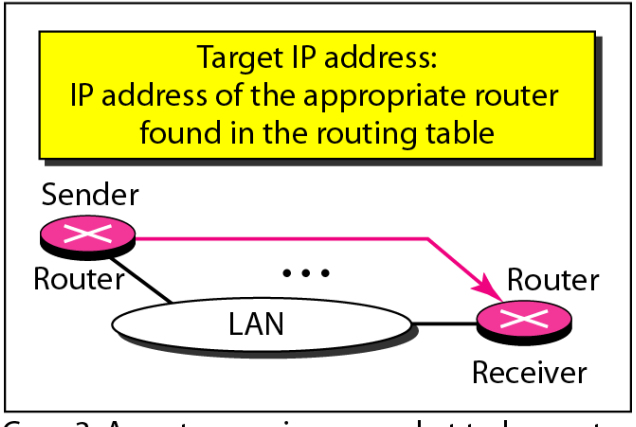
- 使用 ARP 的四种典型情况：
  - 发送方是主机，要把IP数据报发送到本网络上的另一个主机。这时用ARP找到目的主机的硬件地址。
  - 发送方是主机，要把 IP 数据报发送到另一个网络上的一个主机。这时用ARP找到本网络上的一个路由器的硬件地址。剩下的工作由路由器来完成。
  - 发送方是路由器，要把IP数据报转发到本网络上的一个主机。这时用ARP找到目的主机的硬件地址。
  - 发送方是路由器，要把IP数据报转发到另一个网络上的一个主机。这时用ARP找到本网络上另一个路由器的硬件地址。剩下的工作由路由器来完成。



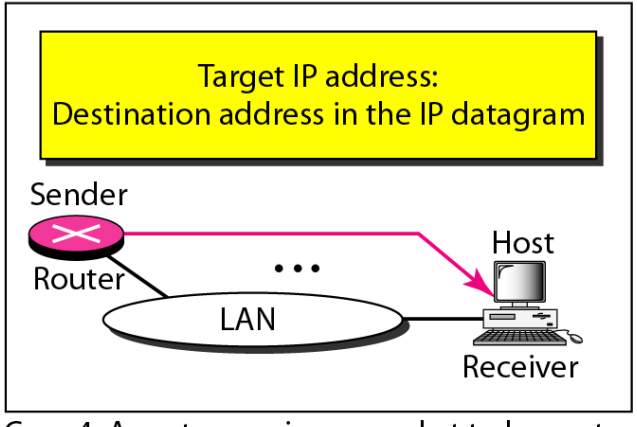
Case 1. A host has a packet to send to another host on the same network.



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.

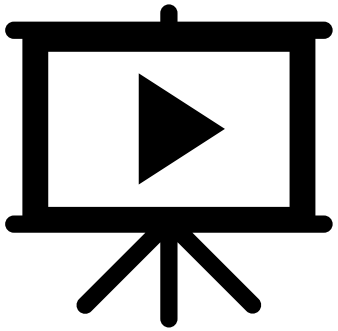


Case 4. A router receives a packet to be sent to a host on the same network.

2. 网际协议 IP

- 为什么不直接使用硬件地址进行通信?
  - 全世界存在着各式各样的网络，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，几乎是不可能的事。
  - IP 编址把这个复杂问题解决了。
  - 连接到互联网的主机只需各自拥有一个唯一的 IP 地址，它们之间的通信就像连接在同一个网络上那样简单方便，即使必须多次调用 ARP 来找到 MAC 地址，但这个过程都是由计算机软件自动进行的，对用户来说是看不见的。
  - 在虚拟的 IP 网络上用 IP 地址进行通信非常方便。

2. 网际协议 IP

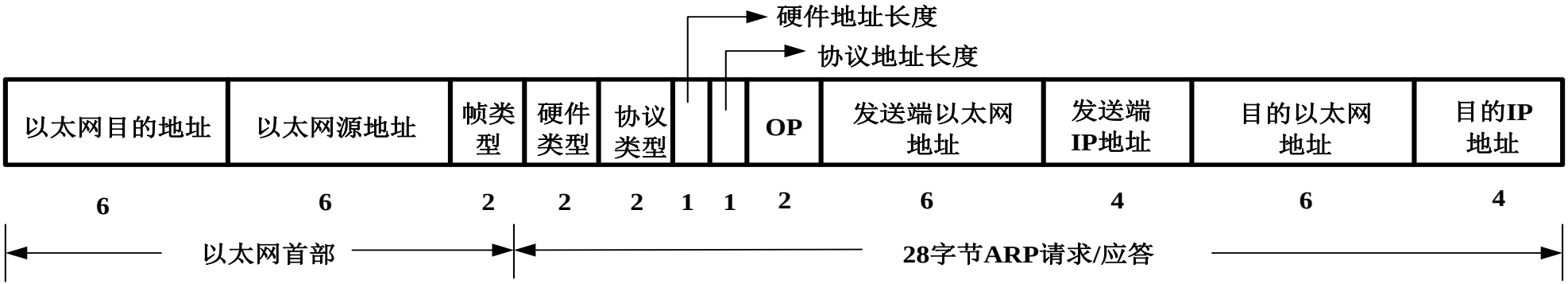


如何查看本地主机的 ARP 高速缓存?

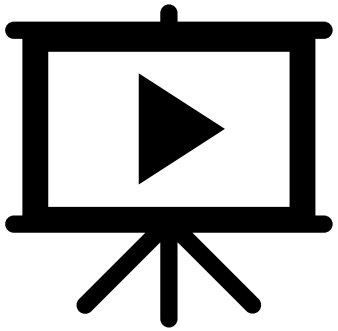
- ✓ 在 Linux 操作系统中，通过在 shell 环境下输入"arp"查看。
- ✓ 在 Windows 操作系统中，通过【运行】【cmd】，在命令窗体中输入"arp -a"查看。
- ✓ 可以通过 arp 命令进行更多操作。

2. 网际协议 IP

- ARP 的数据帧格式：



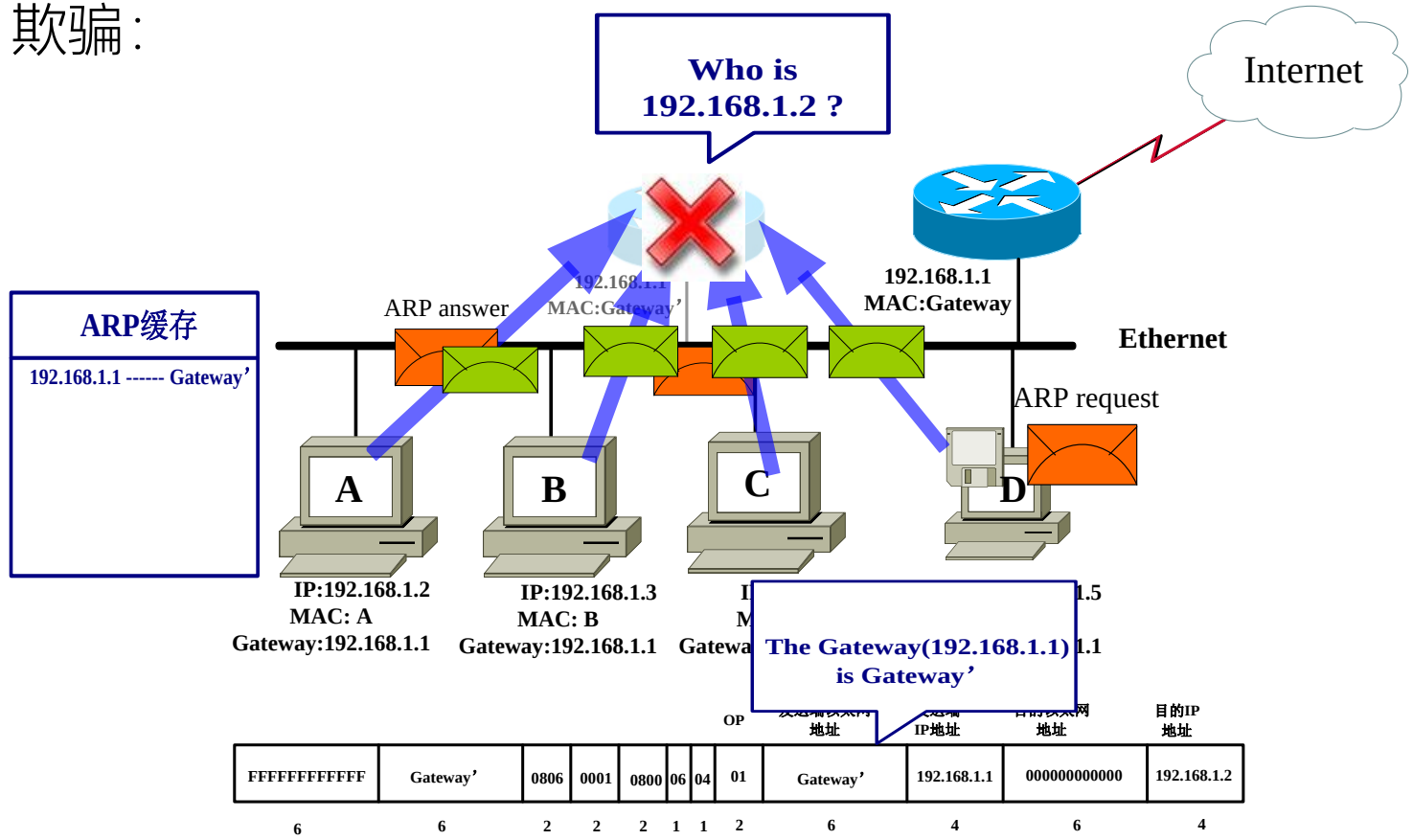
2. 网际协议 IP



通过 Wireshark 进行 ARP 数据报的分析

2. 网际协议 IP

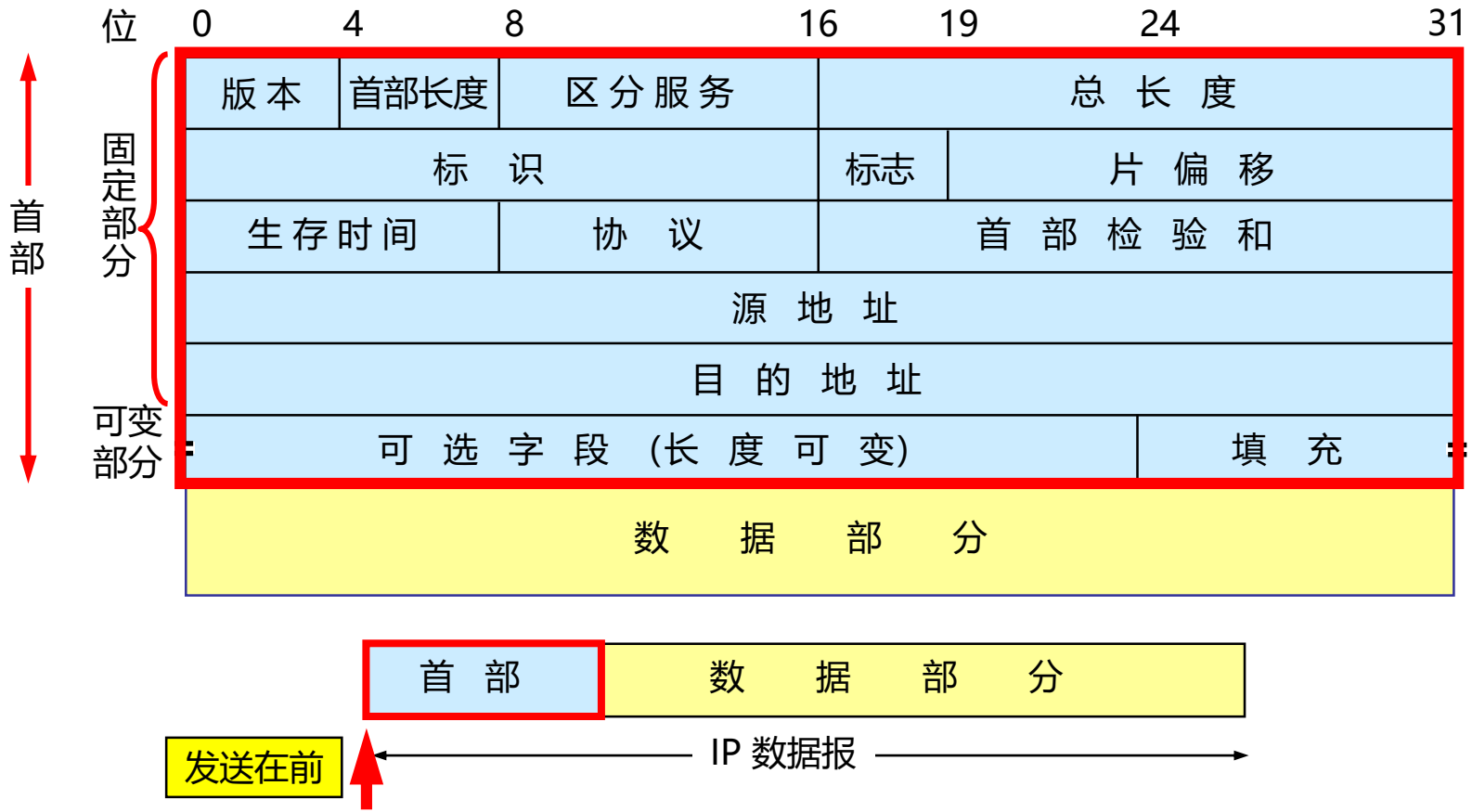
- ARP 欺骗：

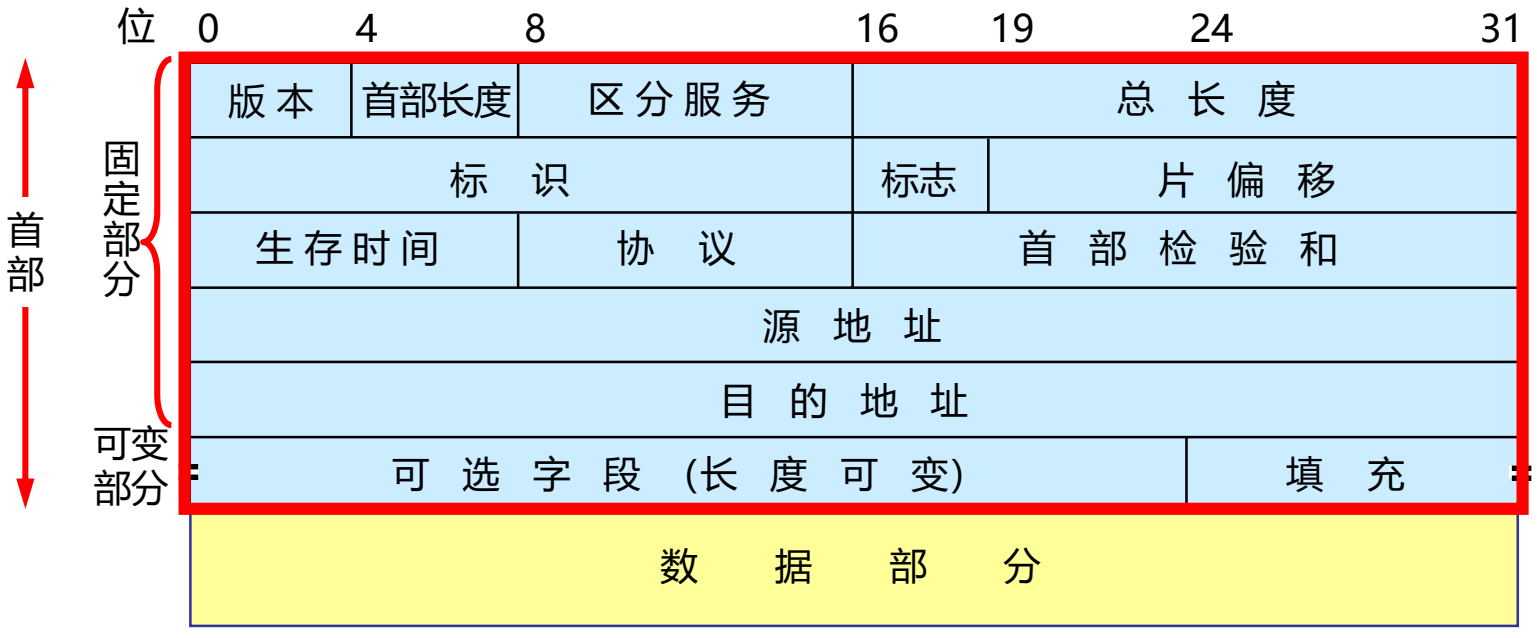


2. 网际协议 IP

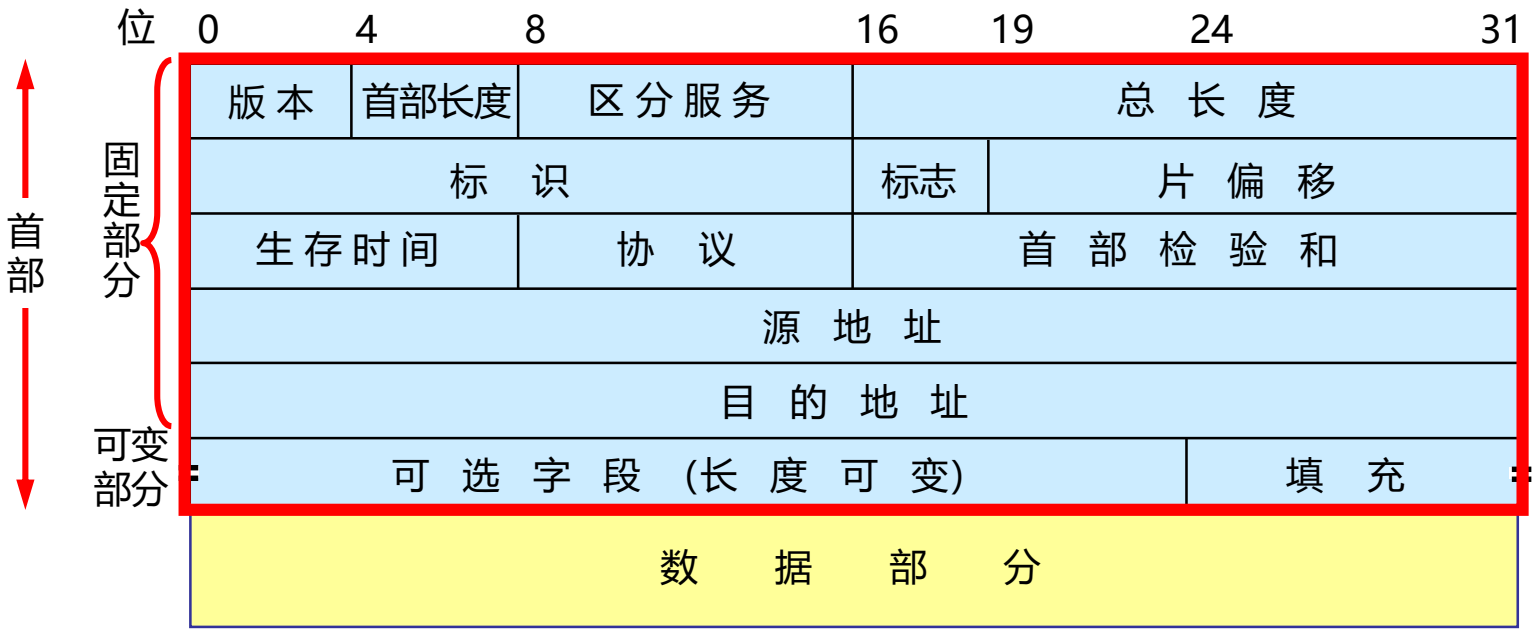
- 一个 IP 数据报由首部和数据两部分组成。
  - 首部的前一部分是固定长度，共20字节，是 IP 数据报必须具有的。
  - 在首部固定部分的后面是一些可选字段，其长度是可变的。
- IP数据报的格式也能够说明 IP 协议的功能。

2. 网际协议 IP

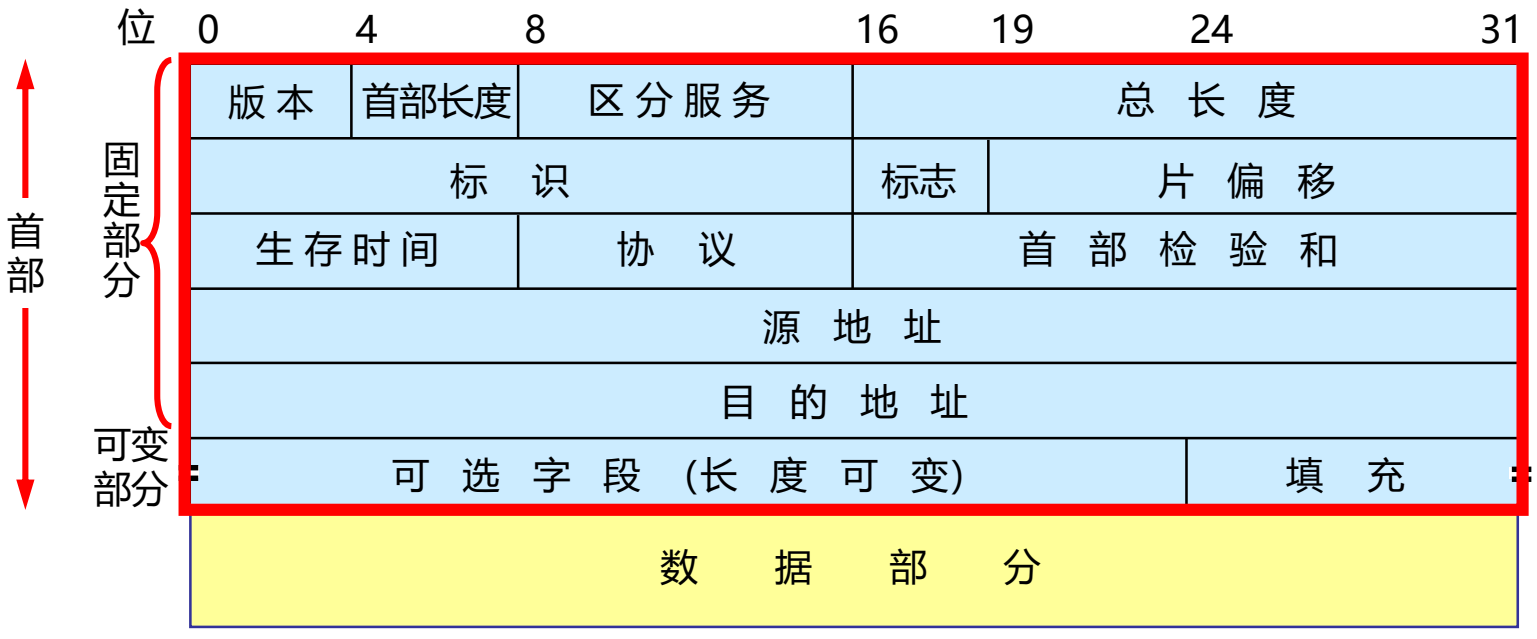

位
0
4
8
16
19
24
31
首部
固定部分
可变部分
版本
首部长度
区分服务
总长度
标识
标志
片偏移
生存时间
协议
首部检验和
源地址
目的地址
可选字段（长度可变）
填充
数据部分
首部
数据部分
发送在前
IP 数据报



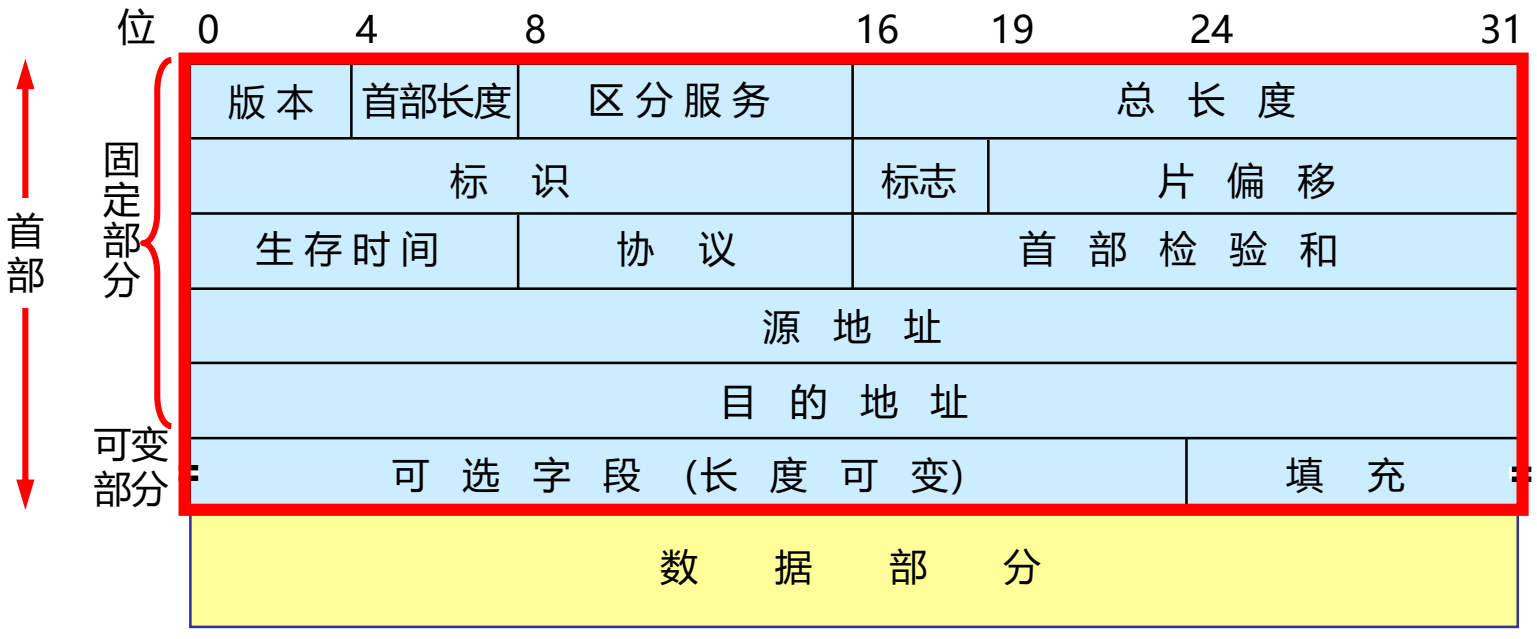
- 版本：4位，指IP协议的版本，目前的IP协议版本号为4(即 IPv4)。
- 首部长度：4位，可表示的最大数值是15个单位(一个单位为4字节)，因此IP的首部长度的最大值是60字节。
- 区分服务：8位，用来获得更好的服务。在旧标准中叫做服务类型，但实际上一直未被使用过。1998年这个字段改名为区分服务。只有在使用区分服务（DiffServ）时，这个字段才起作用。在一般的情况下都不使用这个字段。



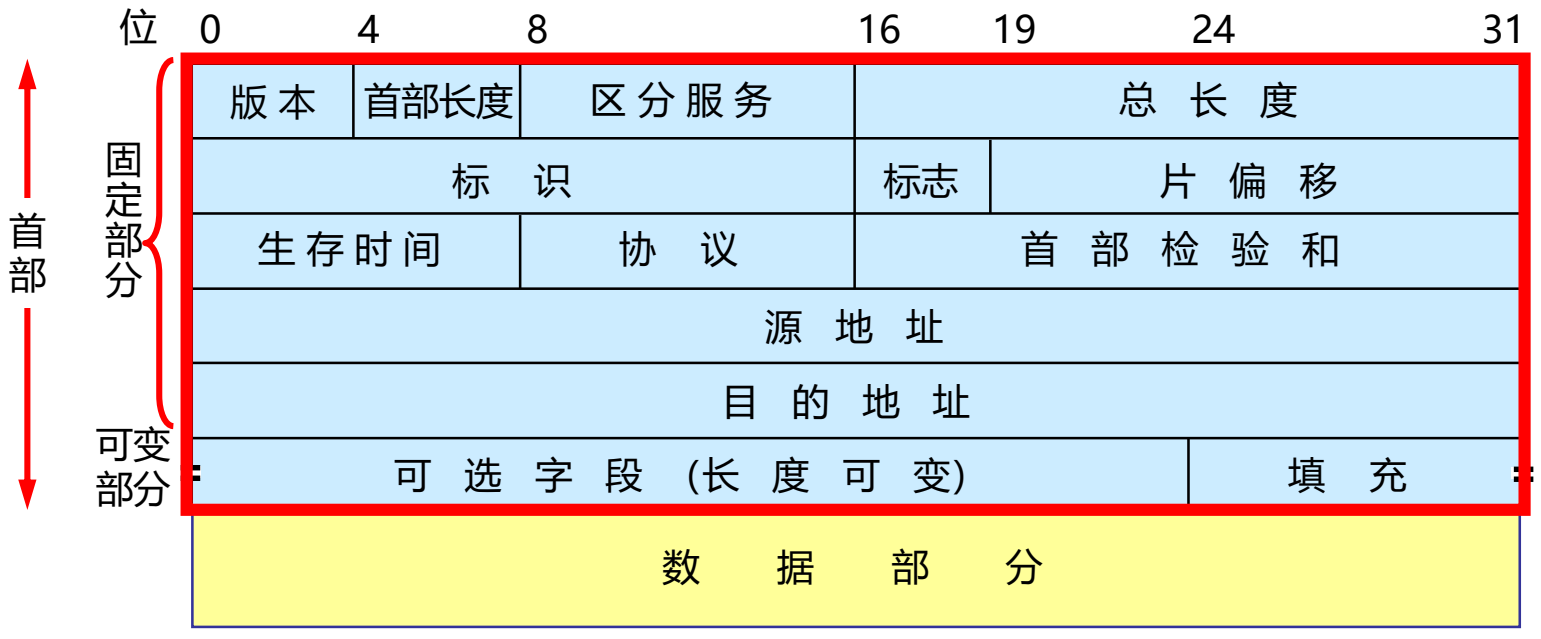
- 总长度：16位，指首部和数据之和的长度，单位为字节，因此数据报的最大长度为65535字节。总长度必须不超过最大传送单元MTU，如果超过了，就需要把过长的数据报进行分片处理。
- 标识(identification)：16位，它是一个计数器，用来产生数据报的标识。
- 标志(flag)：3位，目前只有前两位有意义。
  - 标志字段的最低位是MF(More Fragment)，MF=1表示后面“还有分片”，MF=0表示最后一个分片。
  - 标志字段中间的一位是DF(Don't Fragment)，只有当DF=0时才允许分片。



- 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。



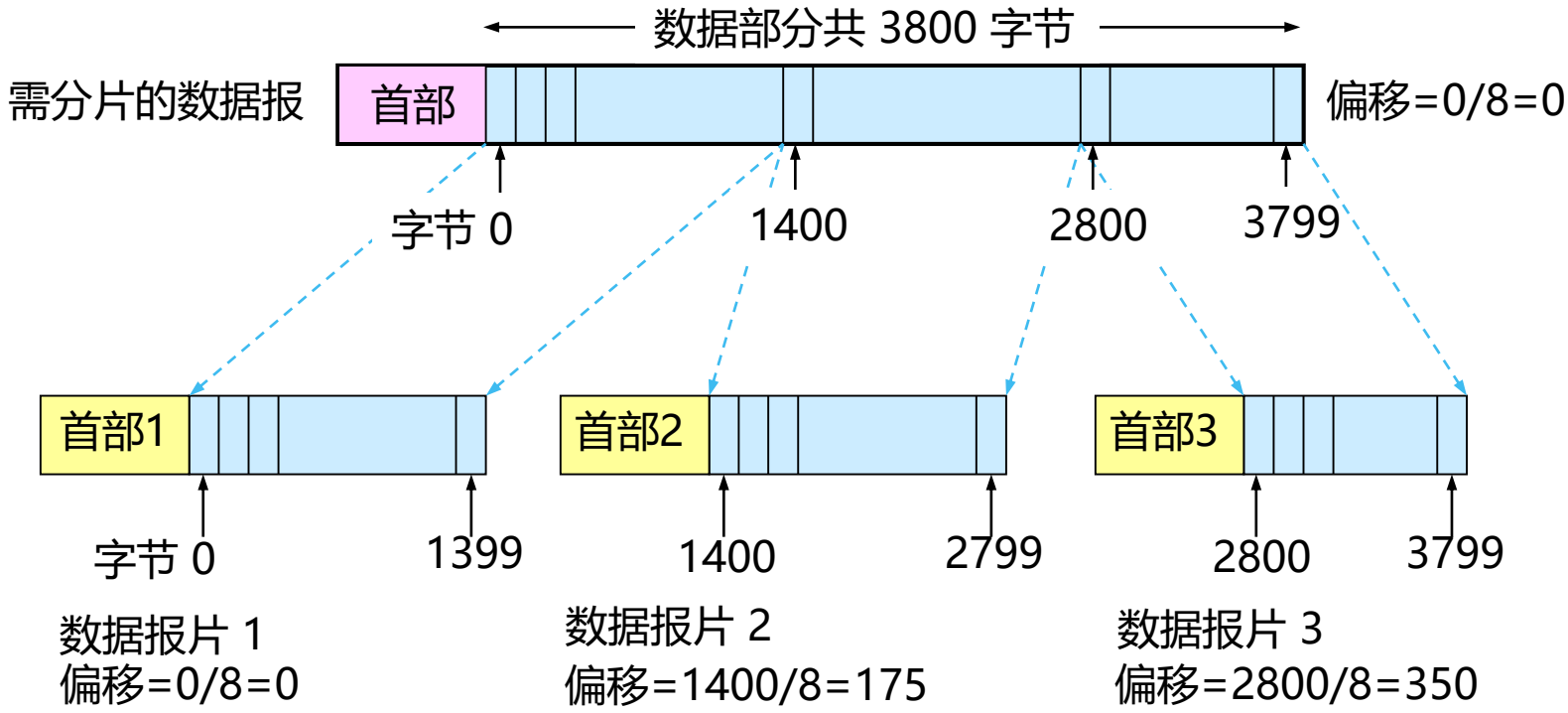
- 首部校验和：16位。这个校验只检测数据报的首部，不对数据报的数据部分进行校验。首部校验和没有使用CRC这样复杂的计算，而是采用了更加简单的方法。
  （算法参考教材的内容，并进行介绍）
- 源地址：32位，发送数据报的IP地址。
- 目的地址：32位，接收数据报的IP地址。



- 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：



2. 网际协议 IP

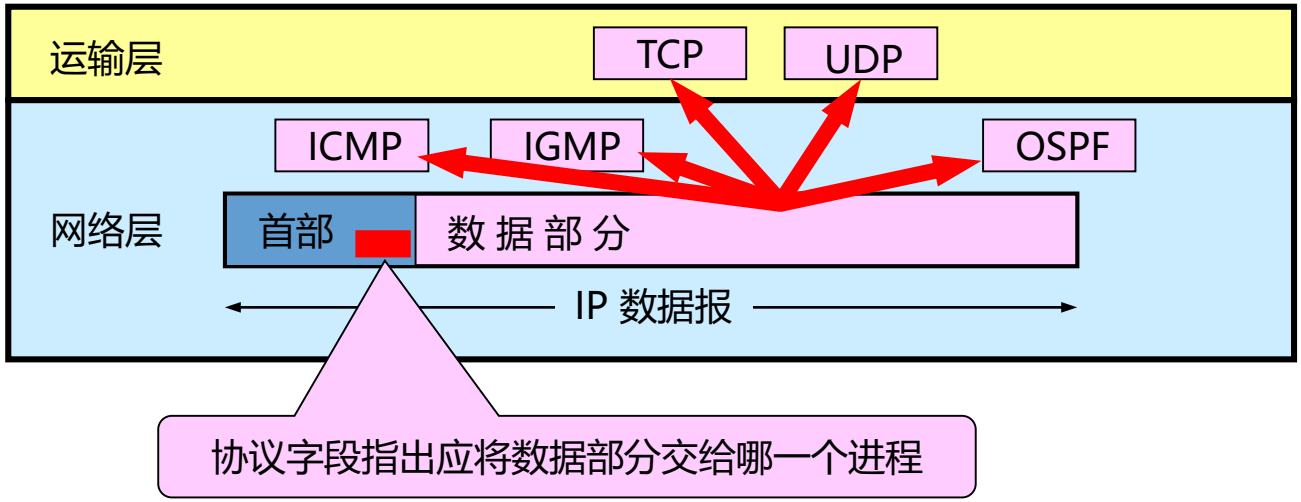
- IP 数据报分片：

IP 数据报首部中与分片有关的字段中的数值

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

2. 网际协议 IP

- 协议字段：

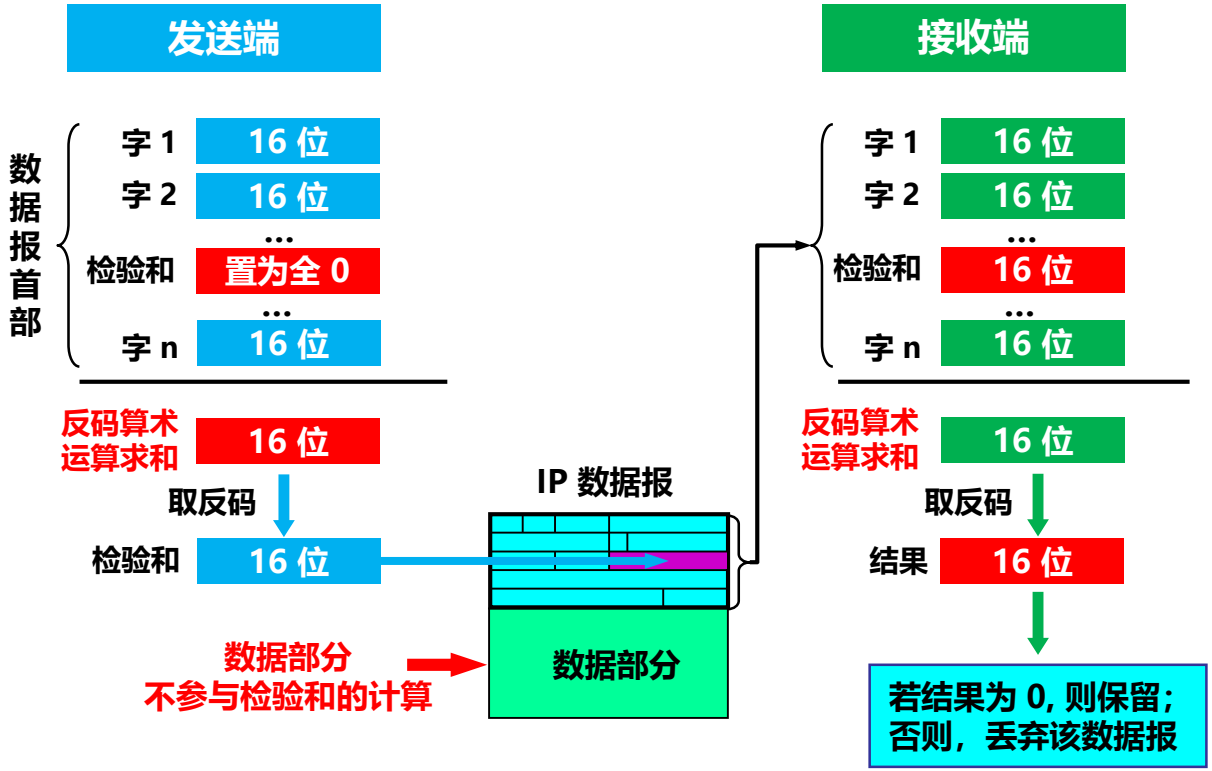


常用的一些协议和相应的协议字段值

协议名	ICMP	IGMP	IP	TCP	EGP	IGP	UDP	IPv6	ESP	AH	ICMP-IPv6	OSPF
协议字段值	1	2	4	6	8	9	17	41	50	51	58	89

2. 网际协议 IP

- 首部校验和：



3. IP 层转发分组的过程

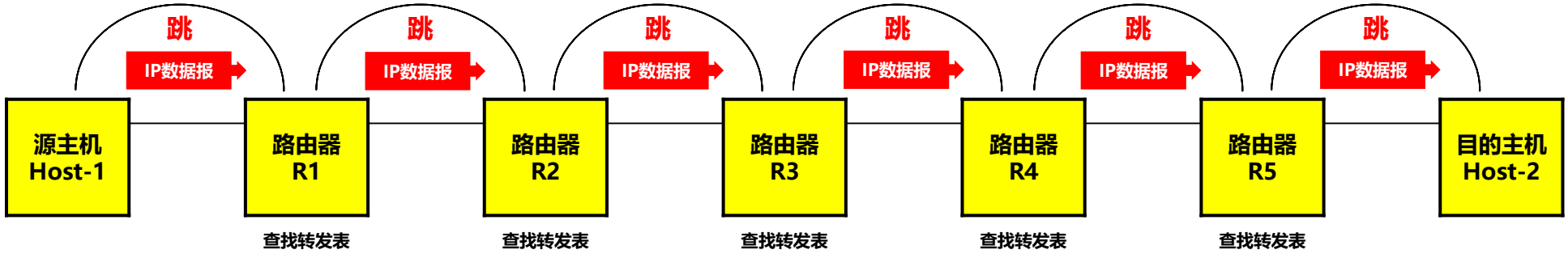
- 在互联网上转发分组时，是从一个路由器转发到下一个路由器。
- 在路由表中，对每一条路由最主要的是两个信息：

  （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。

3. IP 层转发分组的过程

- 分组在互联网中是逐跳转发的。
- 基于终点的转发：基于分组首部中的目的地址传送和转发。

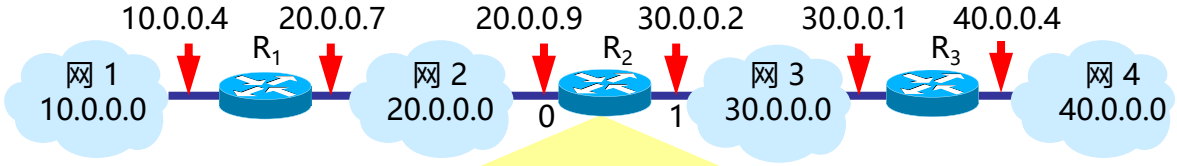


为了压缩转发表的大小，

转发表中最主要的路由是（目的网络地址，下一跳地址），~~不是（目的地址，下一跳地址）~~。

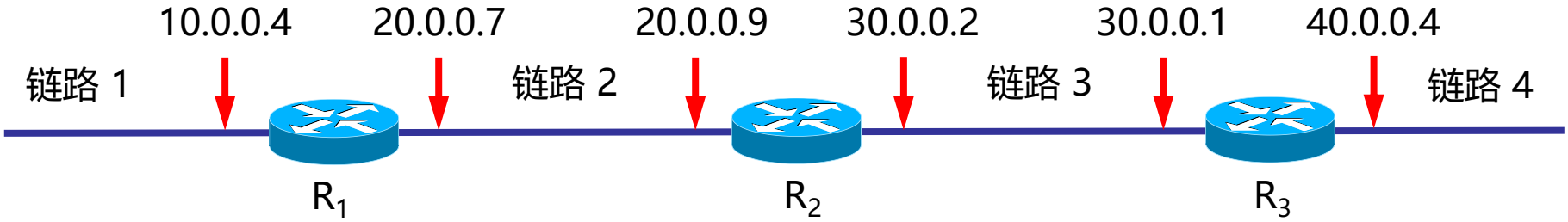
查找转发表的过程就是逐行寻找前缀匹配。

3. IP 层转发分组的过程

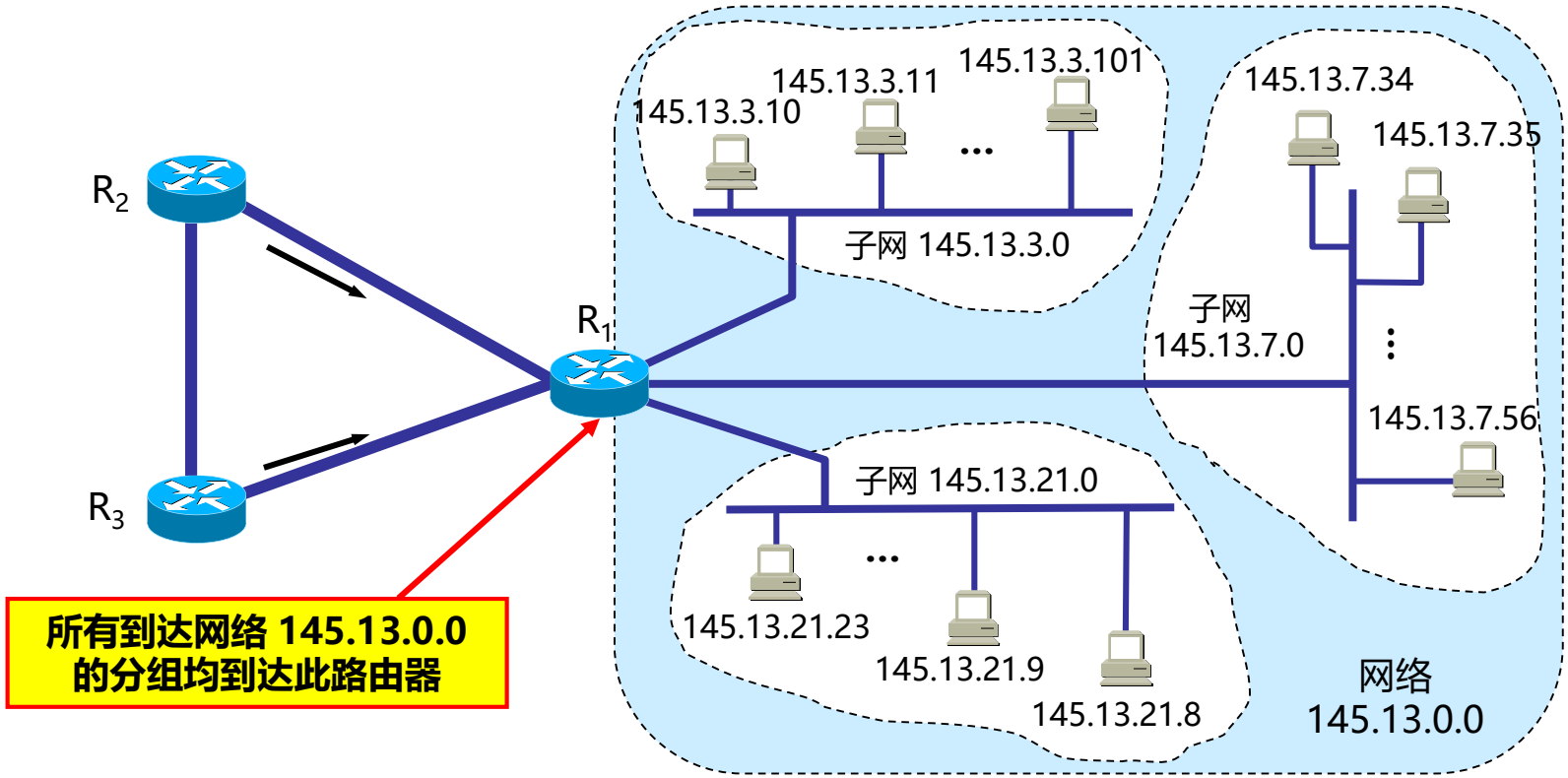


路由器 R_2 的路由表

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



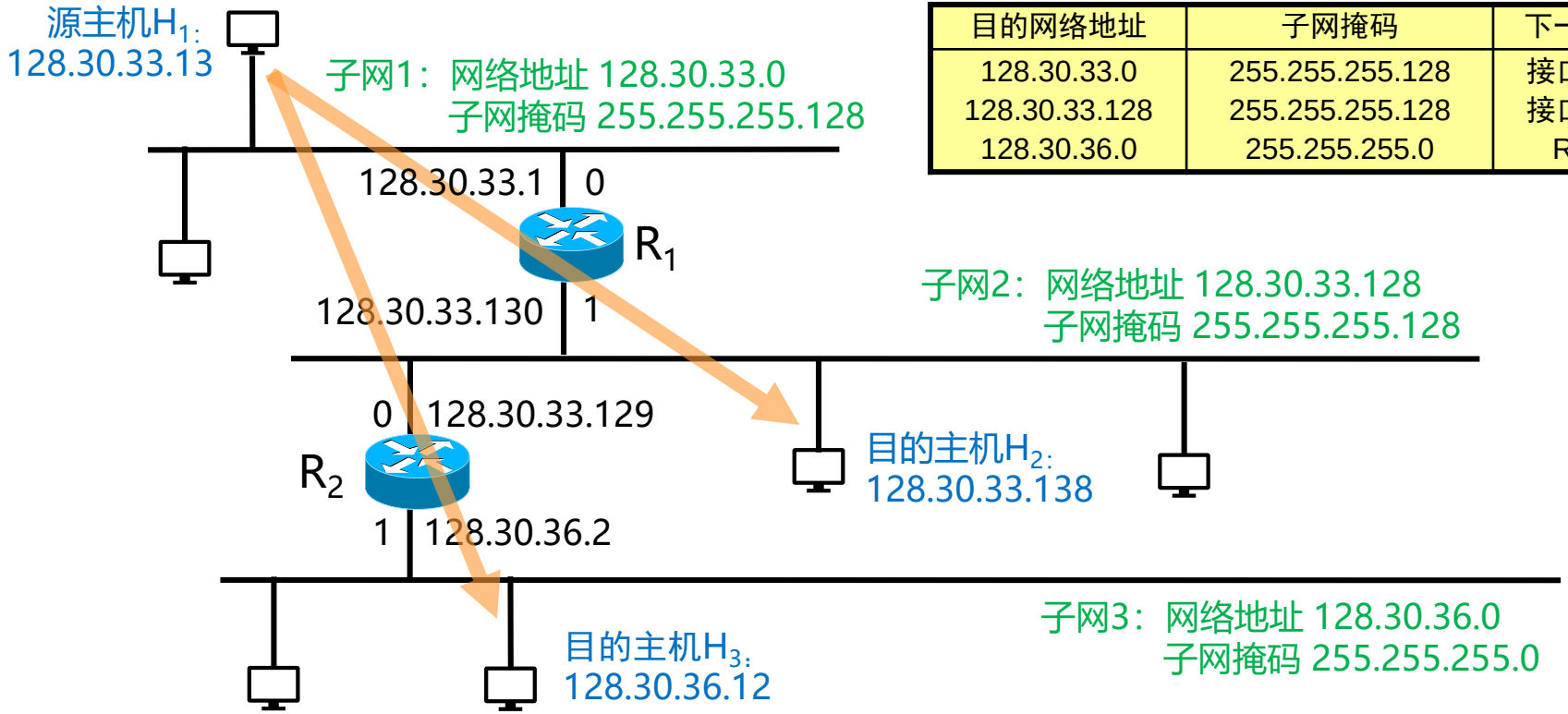
3. IP 层转发分组的过程



3. IP 层转发分组的过程

R_1的路由表（未给出默认路由器）

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



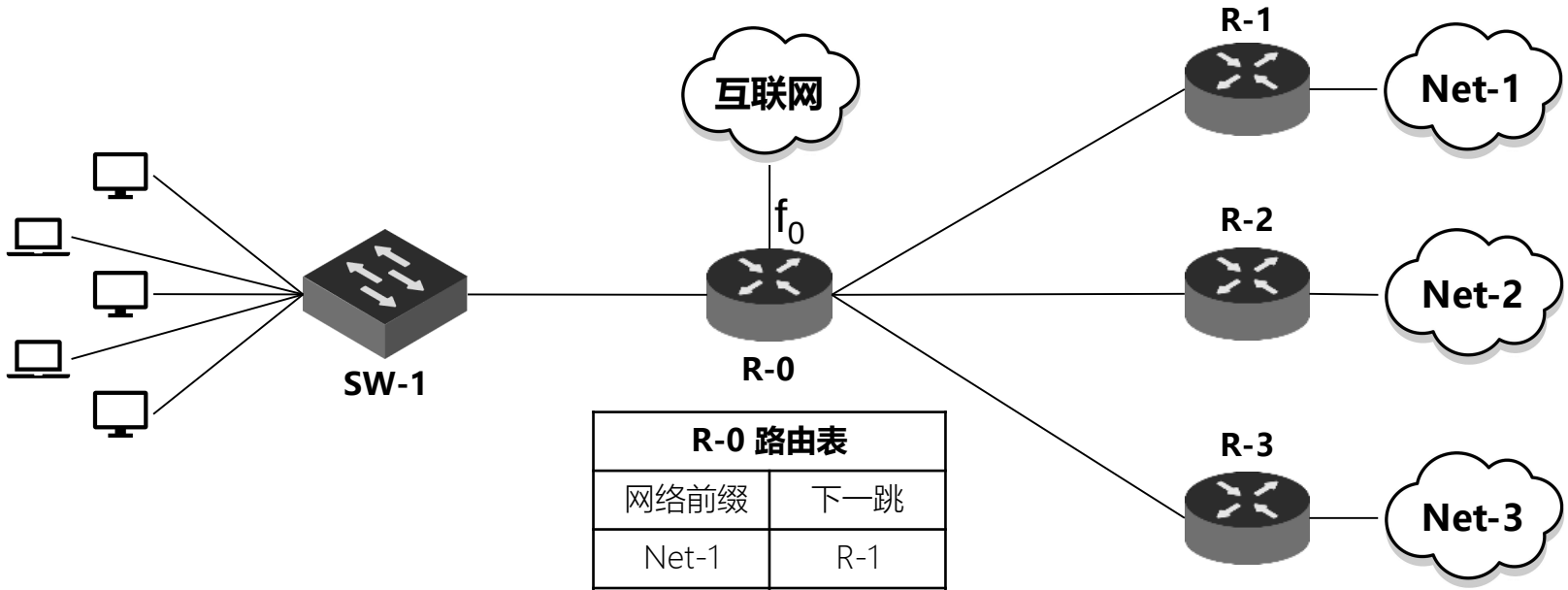
3. IP 层转发分组的过程

- 使用 CIDR 时，在查找转发表时可能会得到不止一个匹配结果。
- 最长前缀匹配 (longest-prefix matching) 原则：
  - 选择前缀最长的一个作为匹配的前缀。
  - 网络前缀越长，其地址块就越小，因而路由就越具体。
  - 把前缀最长的排在转发表的第 1 行。

3. IP 层转发分组的过程

- 转发表中的 2 种特殊的路由
  - **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。

3. IP 层转发分组的过程



R-0 路由表	
网络前缀	下一跳
Net-1	R-1
Net-2	R-2
Net-3	R-3
0.0.0.0/0	接口f0

只要目的网络不是 Net-1，Net-2，Net-3，就选择使用默认路由。

3. IP 层转发分组的过程

- 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。

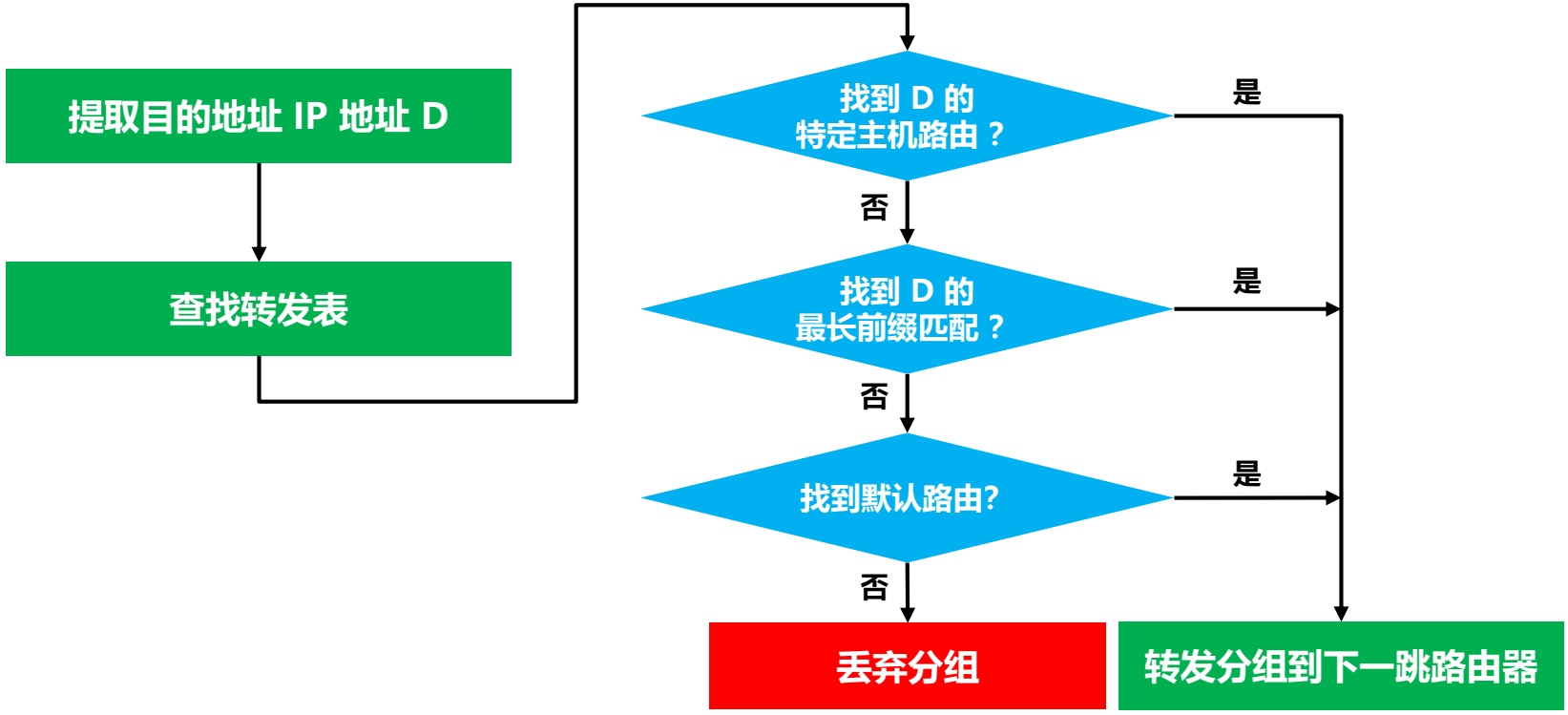
③ 若路由表中有目的地址为D的特定主机路由，则把数据报传送给路由表中所指明的下一跳路由器；否则，执行④。

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。

路由器分组转发算法

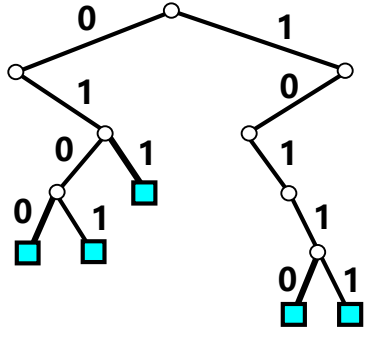

提取目的地址 IP 地址 D
查找转发表
找到 D 的
特定主机路由？
是
否
找到 D 的
最长前缀匹配？
是
否
找到默认路由?
是
否
丢弃分组
转发分组到下一跳路由器

3. IP 层转发分组的过程

- 二叉线索 (binary trie)：
  - 一种特殊结构的树，可以快速在转发表中找到匹配的叶节点。
  - 从二叉线索的根节点自顶向下的深度最多有 32 层，每一层对应于 IP 地址中的一位。
  - 为简化二叉线索的结构，可以用唯一前缀 (unique prefix) 来构造二叉线索。
  - 为了提高二叉线索的查找速度，广泛使用了各种压缩技术。

用 5 个唯一前缀构成的二叉线索

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



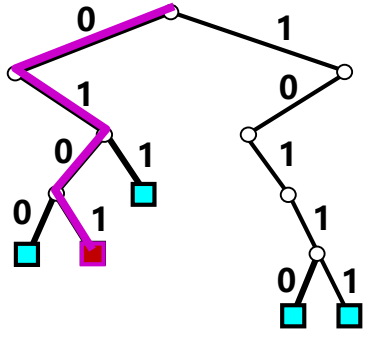
规则：

- ✓ 先检查 IP 地址左边的第一位，如为 0，则第一层的节点就在根节点的左下方；如为 1，则在右下方。
- ✓ 然后再检查地址的第二位，构造出第二层的节点。依此类推，直到唯一前缀的最后一位。每个叶节点代表一个唯一前缀。
- ✓ 为检查网络前缀是否匹配，必须使二叉线索中的每一个叶节点包含所对应的网络前缀和子网掩码。

在二叉线索中查找 IP 地址

32 位的 IP 地址

01010110 01111010 00000000 00000000



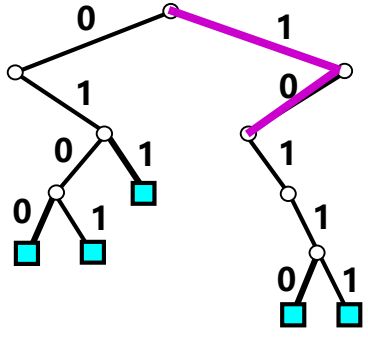
查找方法：

- ✓ 找到了一个叶节点。
- ✓ 将目的 IP 地址和该叶节点的子网掩码进行按位 AND 运算，看结果是否与叶节点的网络前缀相匹配。
- ✓ 若匹配，就按下一跳的接口转发该分组。否则，就丢弃该分组。

在二叉线索中查找 IP 地址

32 位的 IP 地址

10011011 01111010 00000000 00000000



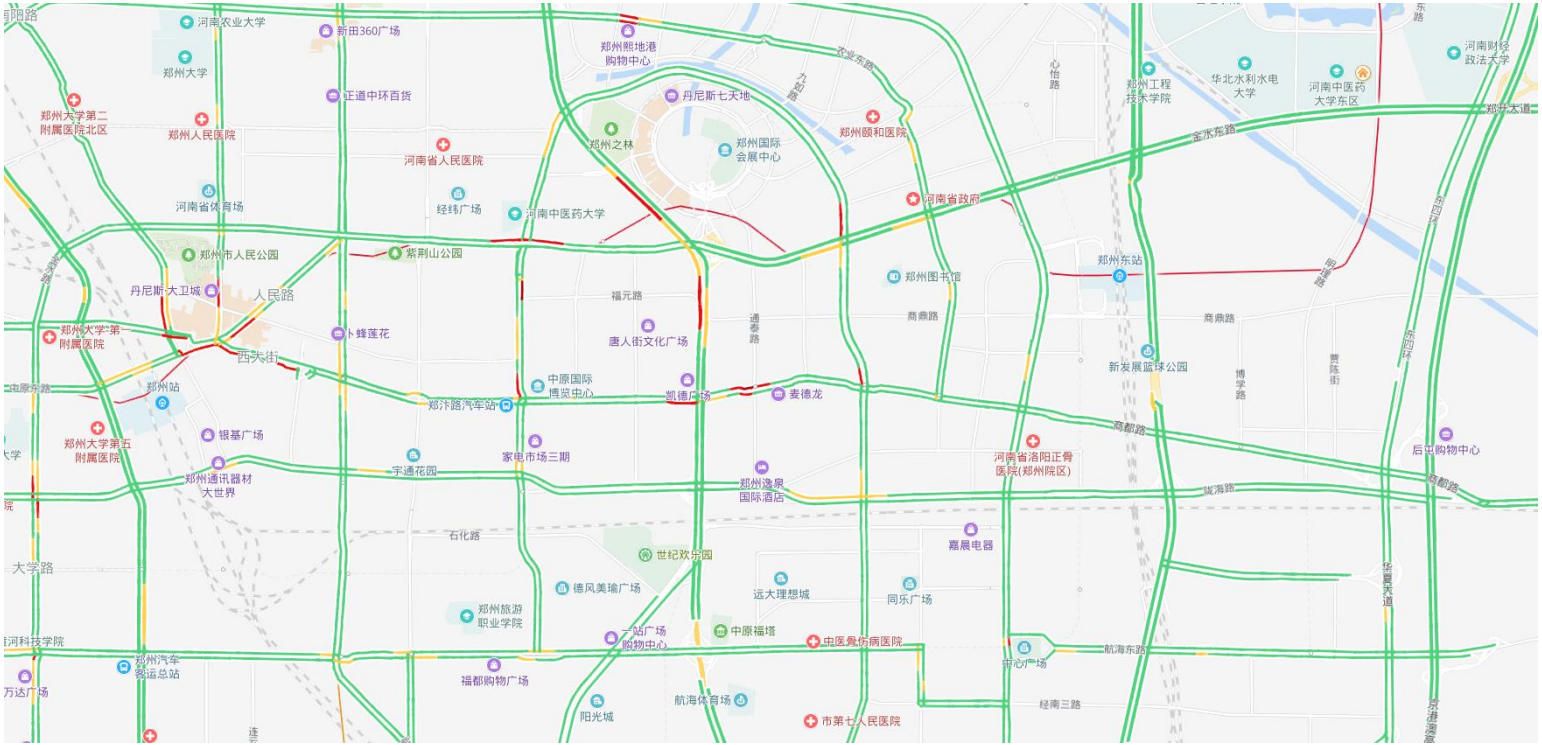
查找方法：

- ✓ 查到第三个字符 0 时，在二叉线索中找不到匹配的。说明这个地址不在这个二叉线索中。
- ✓ 检查是否存在默认路由。若有，把分组传送到指明的默认路由器，否则丢弃该分组。

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

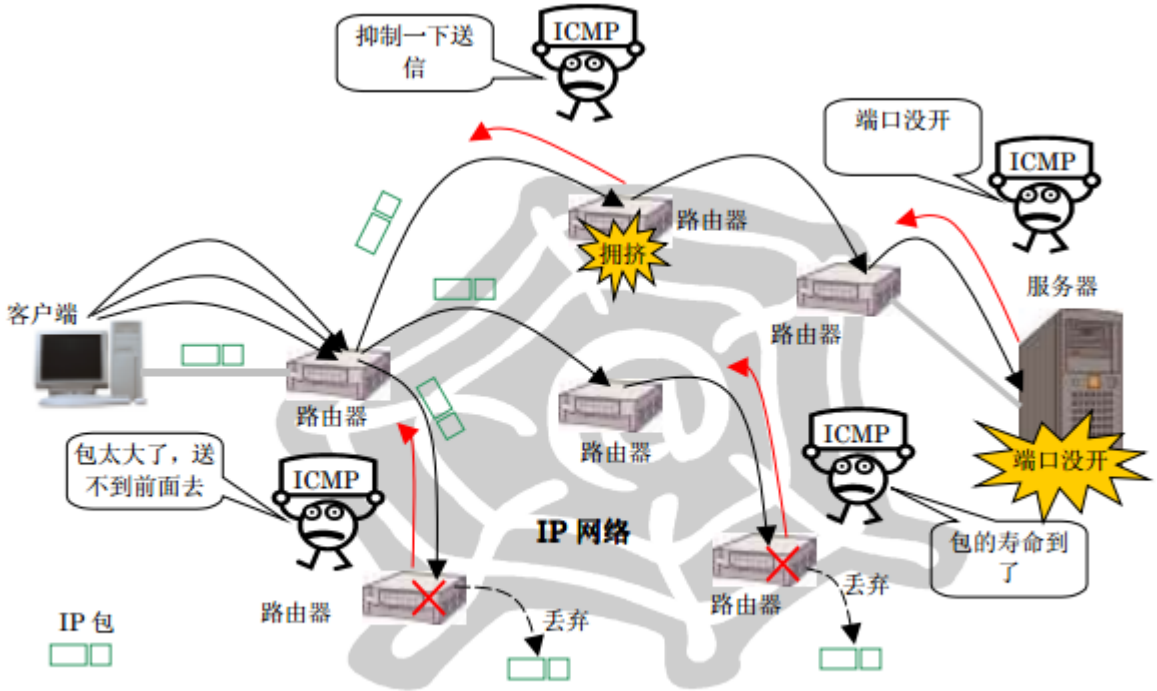
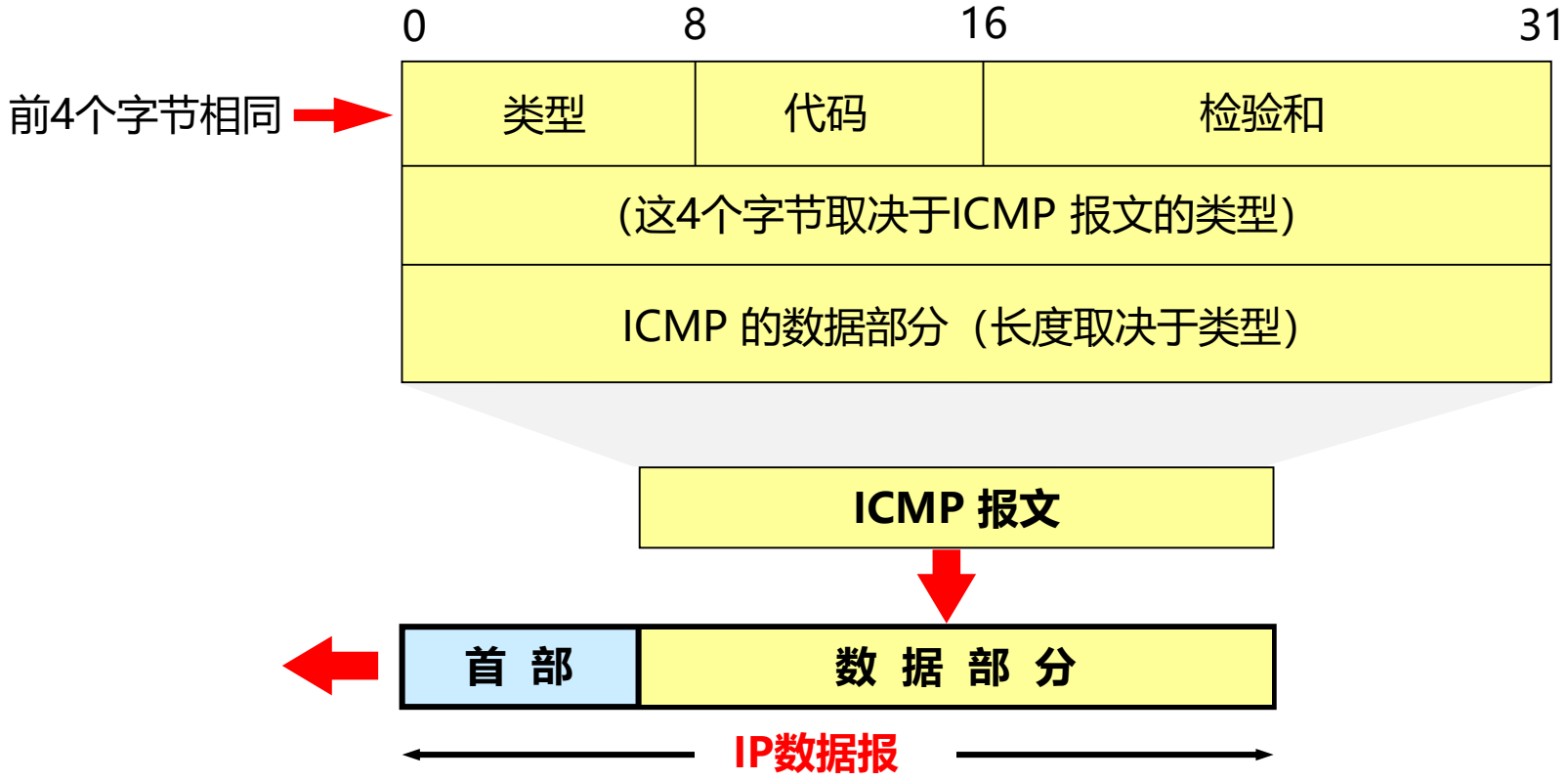


图 1 ICMP 是使 IP 通信平稳运行的辅助协议

4. 网际控制报文协议 ICMP



4. 网际控制报文协议 ICMP

- ICMP 报文种类有两种：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。

几种常用的 ICMP 报文类型

ICMP 报文种类	类型的值	ICMP报文的类型
差错报告报文	3	终点不可达
	11	时间超过
	12	参数问题
	5	改变路由 (Redirect)
询问报文	8 或 0	回送 (Echo) 请求或回答
	13 或 14	时间戳 (Timestamp) 请求或回答

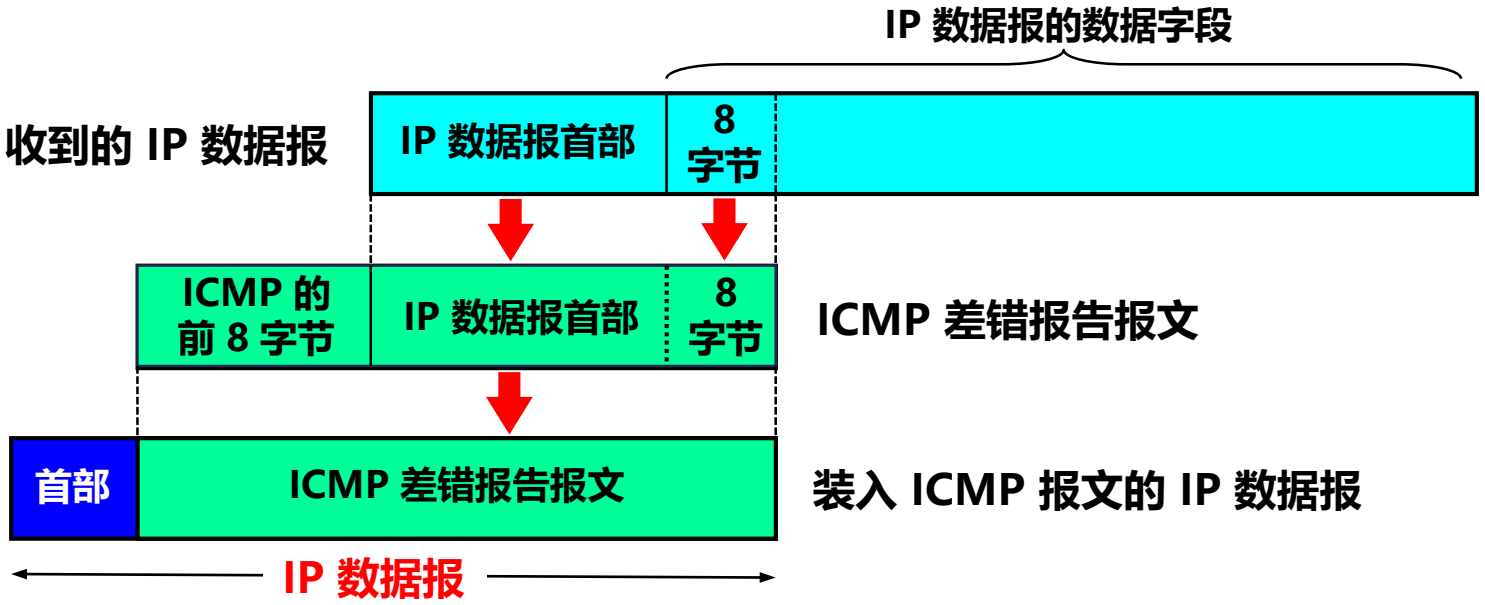
类型	代码	名称	查询	差错
0	0	回应应答(Echo Reply)	√	
3		目的地不可达		√
	0	网路不可达		√
	1	主机不可达		√
	2	协议不可达		√
	3	端口不可达		√
	4	需要分片和不需要分片标记置位		√
	5	源路由失败		√
	6	目的网络未知		√
	7	目的主机未知		√
	8	源主机被隔离		√
	9	目的网络的通告被禁止		√
	10	目的主机的通信被禁止		√
	11	对请求的服务类型 ToS，目的网路不可达		√
	12	对请求的服务类型 ToS，目的主机不可达		√
	13	由于过滤,通信被强制禁止		√
	14	主机越权		√
	15	优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

- 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

ICMP 差错报告报文的数据字段的内容



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。

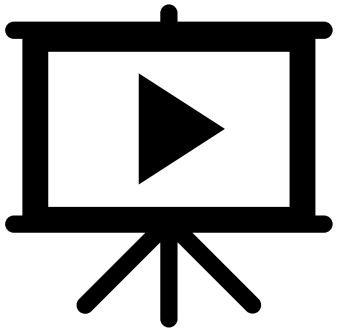
4. 网际控制报文协议 ICMP

- ICMP 询问报文有两种：
  - 回送请求和回答
    - 由主机或路由器向一个特定的目的主机发出的询问。
    - 收到此报文的主机必须给源主机或路由器发送 ICMP 回送回答报文。
    - 这种询问报文用来测试目的站是否可达，以及了解其有关状态。
  - 时间戳请求和回答
    - 请某台主机或路由器回答当前的日期和时间。
    - 时间戳回答报文中有一个 32 位的字段，其中写入的整数代表从1900 年 1 月 1 日起到当前时刻一共有多少秒。
    - 时间戳请求与回答可用于时钟同步和时间测量。

4. 网际控制报文协议 ICMP

- PING (Packet InterNet Groper)
  - 用来测试两个主机之间的连通性。
  - 使用了 ICMP 回送请求与回送回答报文。
  - 是应用层直接使用网络层 ICMP 的例子，没有通过运输层的 TCP 或 UDP。
- Traceroute
  - 这是UNIX操作系统中名字。在 Windows 操作系统中这个命令是 tracert。
  - 用来跟踪一个分组从源点到终点的路径。
  - 它利用 IP 数据报中的 TTL 字段、ICMP 时间超过差错报告报文和ICMP 终点不可达差错报告报文实现对从源点到终点的路径的跟踪。

4. 网际控制报文协议 ICMP



使用 PING 工具进行网络连通性测试

使用 Traceroute 工具进行网络性能分析

5. IPv6

- IP 是互联网的核心协议。
- IPv4 地址耗尽问题：
  - 到 2011 年 2 月，IANA IPv4 的 32 位地址已经耗尽。
  - 各地区互联网地址分配机构也相继宣布地址耗尽。
  - 我国在 2014 – 2015 年也逐步停止了向新用户和应用分配 IPv4 地址。
- 根本解决措施：
  - 采用具有更大地址空间的新版本的 IP，即 IPv6。

5. IPv6

- IPv6 仍支持无连接的传送，但将协议数据单元 PDU 称为分组。

教学中尽量按照 PDU 描述，偶然会延续使用数据报这一名词。

- IPv6 的主要变化：
  1. 更大的地址空间。
     - IPv6 将地址从 IPv4 的 32 位 增大到了 128 位。
  2. 扩展的地址层次结构。
  3. 灵活的首部格式。
     - IPv6 定义了许多可选的扩展首部。
  4. 改进的选项。
     - IPv6 允许数据报包含有选项的控制信息，其选项放在有效载荷中。

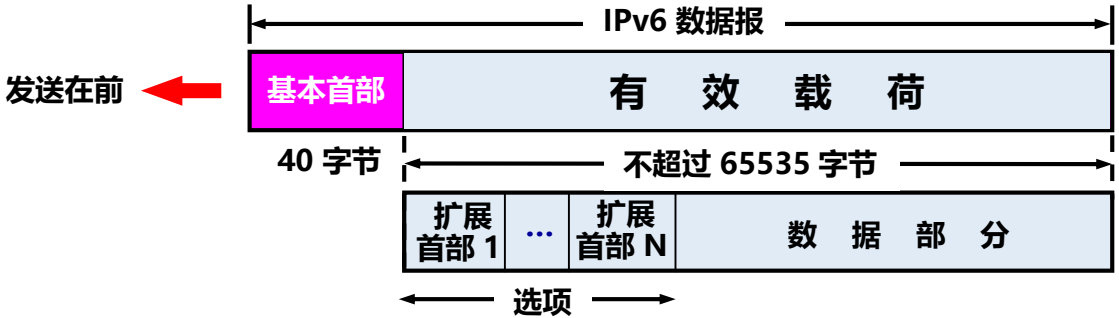
5. IPv6

- IPv6 的主要变化：
  5. 允许协议继续扩充。
  6. 支持即插即用（即自动配置）。
     - IPv6 不需要使用 DHCP。
  7. 支持资源的预分配。
     - IPv6 支持实时视像等要求，保证一定的带宽和时延的应用。
  8. IPv6 首部改为 8 字节对齐。
     - 首部长度必须是 8 字节的整数倍，IPv4 首部是 4 字节对齐。

What Is IPv6?
https://info.support.huawei.com/infofinder/encyclopedia/en/IPv6.html

5. 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。



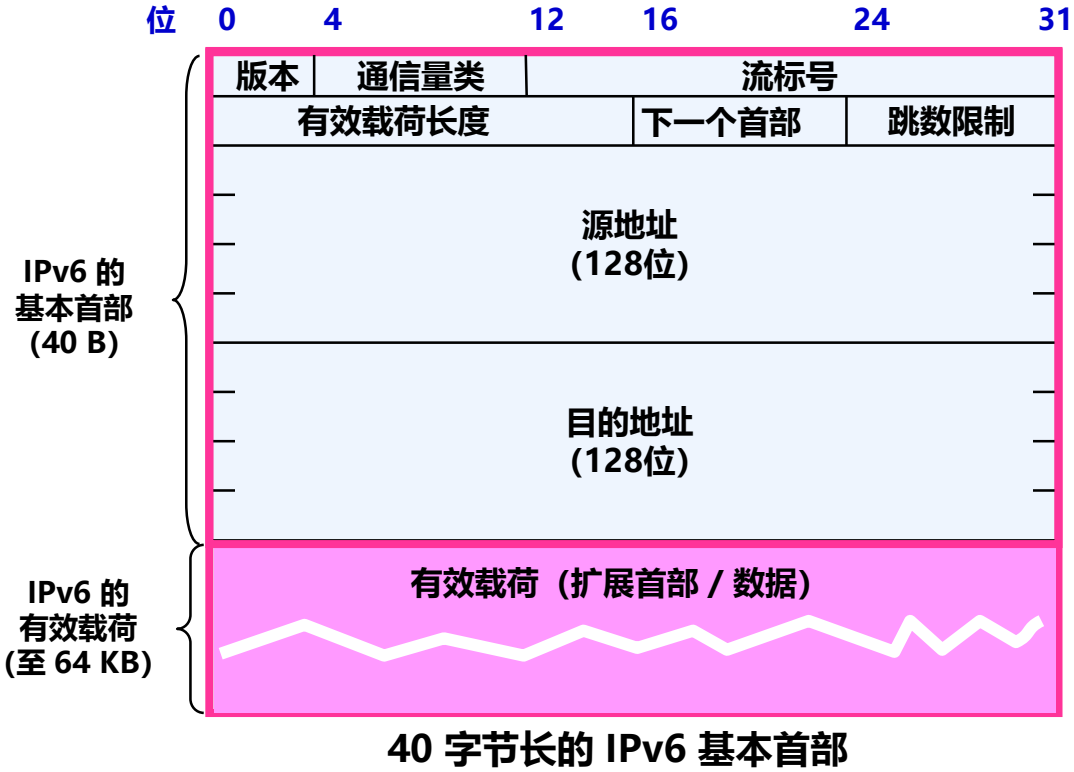
具有多个可选扩展首部的 IPv6 数据报的一般形式

5. 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。**

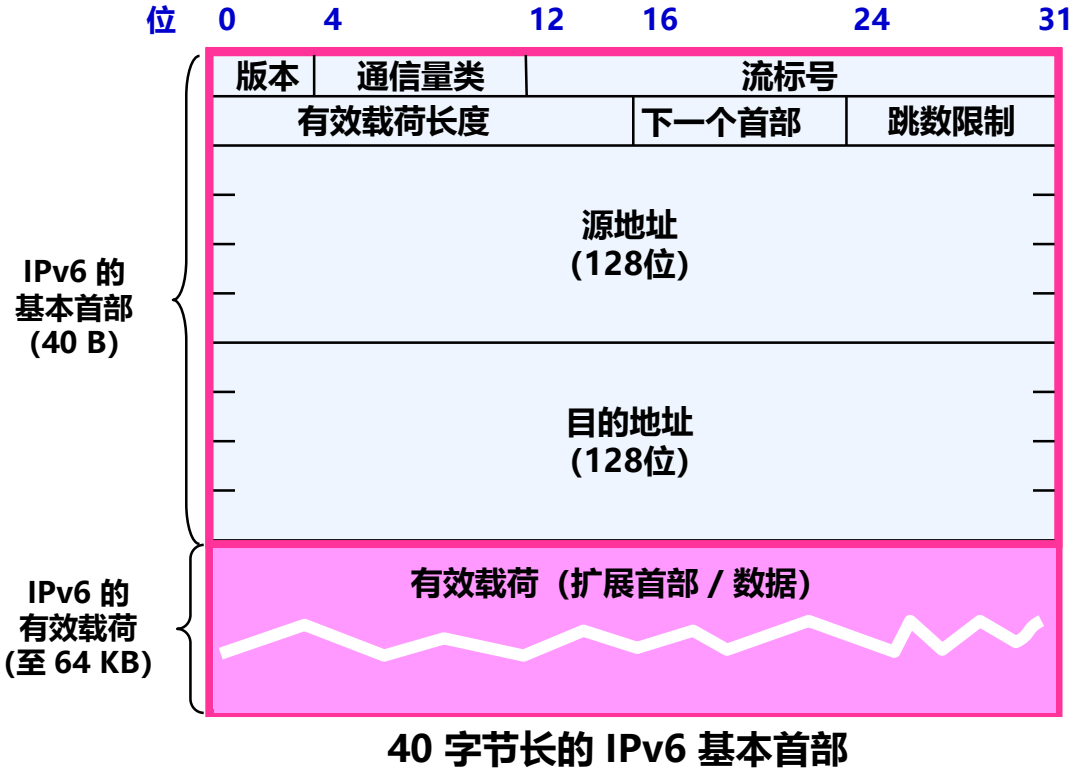
5. IPv6



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。

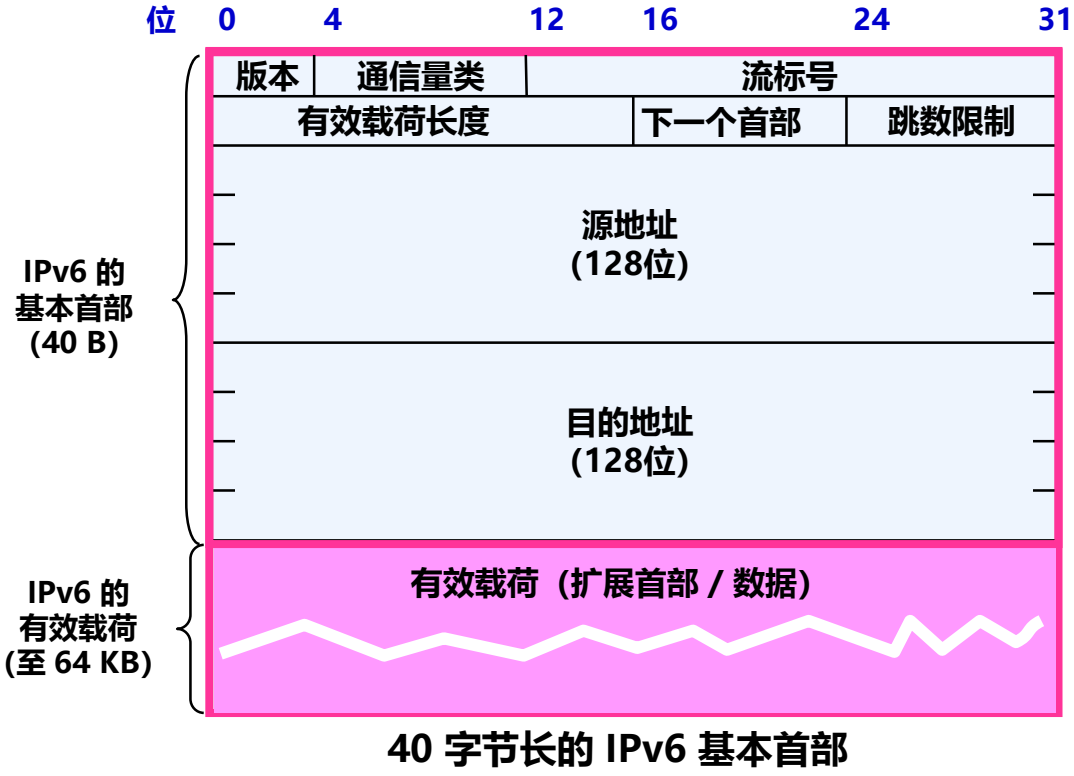
5. IPv6



40 字节长的 IPv6 基本首部

- 有效载荷长度(payload length)：16 位。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。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。

5. 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。

5. 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

5. 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。
    - 通常是按照路由算法得出的距离最近的一个。

5. IPv6

- IPv6 将实现 IPv6 的主机和路由器均称为**节点**。
- 一个节点可能有多个与链路相连的接口。
- IPv6 地址是分配给节点上**接口**的。
  - 一个具有多个接口的节点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节点的目的地址。

5. 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
  - 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

5. IPv6

冒号十六进制法的两个技术

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。

0:0:0:0:0:0:128.10.2.1	➡	::128.10.2.1
1080:0:0:0:8:800:200C:417A	➡	1080::8:800:200C:417A
FF01:0:0:0:0:0:0:101（多播地址）	➡	FF01::101
0:0:0:0:0:0:0:1（环回地址）	➡	::1
0:0:0:0:0:0:0:0（未指明地址）	➡	::

5. 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，但取消了子网掩码。
  - 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）

IPv6 地址分类

地址类型	二进制前缀	IPv6记法
未指明地址	00...0（128位），仅此一个	::/128
环回地址	00...1（128位），仅此一个	::1/128
多播地址	11111111（8位），功能和 IPv4 的一样	FF00::/8
本地链路单播地址	1111111010（10位），未连接到互联网，不能和互联网上的其他主机通信	FE80::/10
全球单播地址	除上述四种外，所有其他的二进制前缀	

5. IPv6

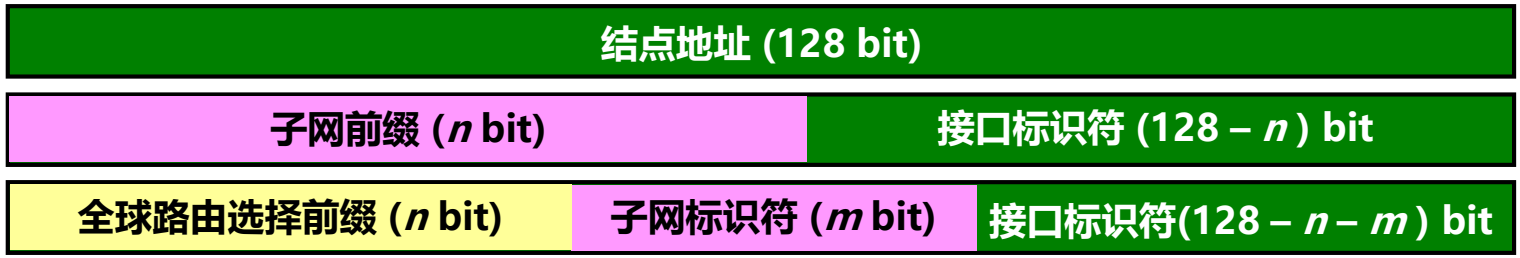
- IPv6 地址分类
  - 未指明地址
    - 这是 16 字节的全 0 地址，可缩写为两个冒号 “::” 。
    - 这个地址只能为还没有配置到一个标准的 IP 地址的主机当作源地址使用。
    - 这类地址仅此一个。
  - 环回地址
    - 即 0:0:0:0:0:0:0:1（记为 ::1）。
    - 作用和 IPv4 的环回地址一样。
    - 这类地址也是仅此一个。

5. 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。

5. 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 单播地址的划分非常灵活。



IPv6 单播地址的几种划分方法

5. IPv6

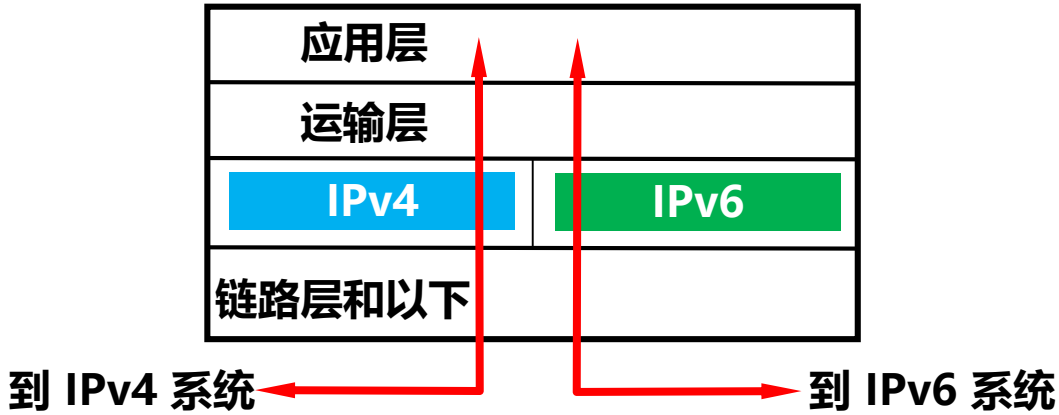
- 向 IPv6 过渡只能采用逐步演进的办法，还必须使新安装的 IPv6 系统能够**向后兼容**：
  - IPv6 系统必须能够接收和转发 IPv4 分组。
  - 能够为 IPv4 分组选择路由。
- 两种向 IPv6 过渡的策略：
  - 使用双协议栈
  - 使用隧道技术

5. 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。

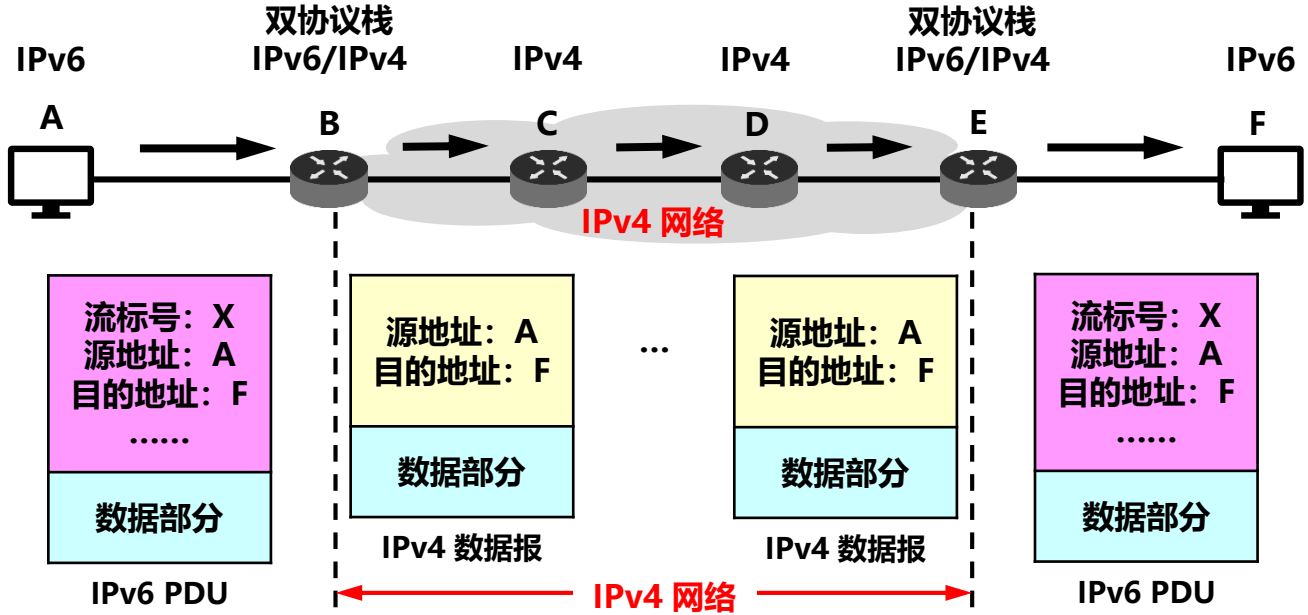
5. IPv6

- 双协议栈



IPv6/IPv4 双协议栈主机（或路由器）

5. IPv6

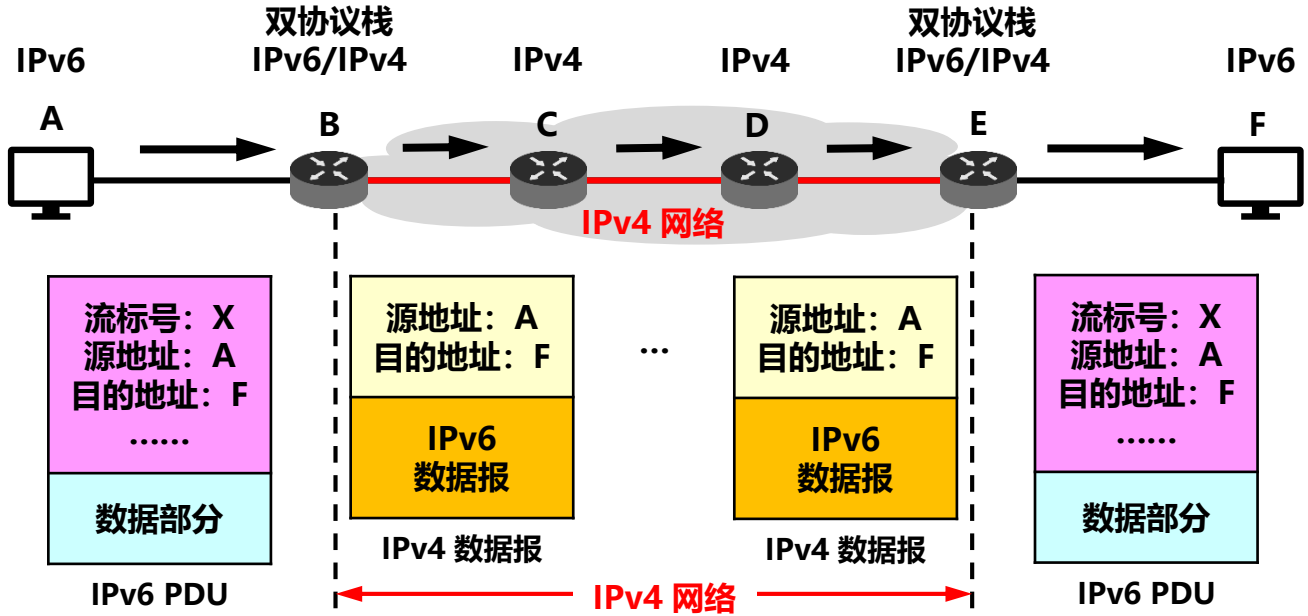


使用双协议栈进行从 IPv4 到 IPv6 的过渡

5. 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。

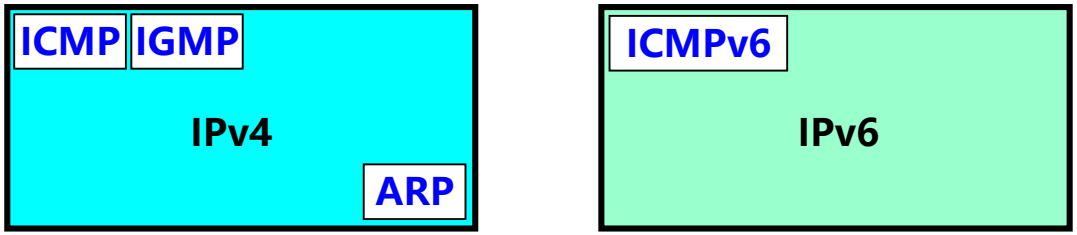
5. IPv6



使用隧道技术进行从 IPv4 到 IPv6 的过渡

5. IPv6

- ICMPv6
  - IPv6 也不保证数据报的可靠交付，互联网中的路由器可能会丢弃数据报。
  - 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。

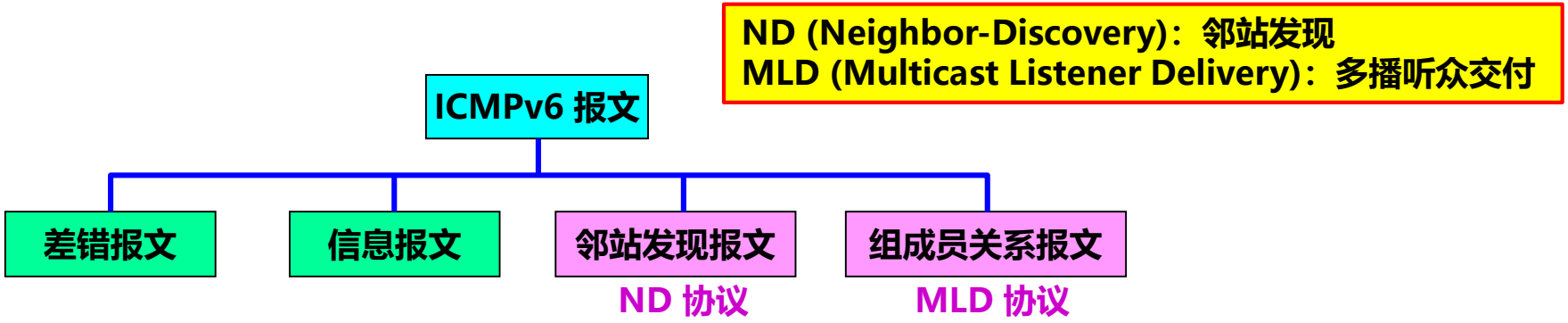


版本 4 中的网络层　　**版本 6 中的网络层**

新旧版本中的网络层的比较

5. IPv6

- 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

6. 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。

6. 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。

6. 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）

6. 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。

6. 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。

6. 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。

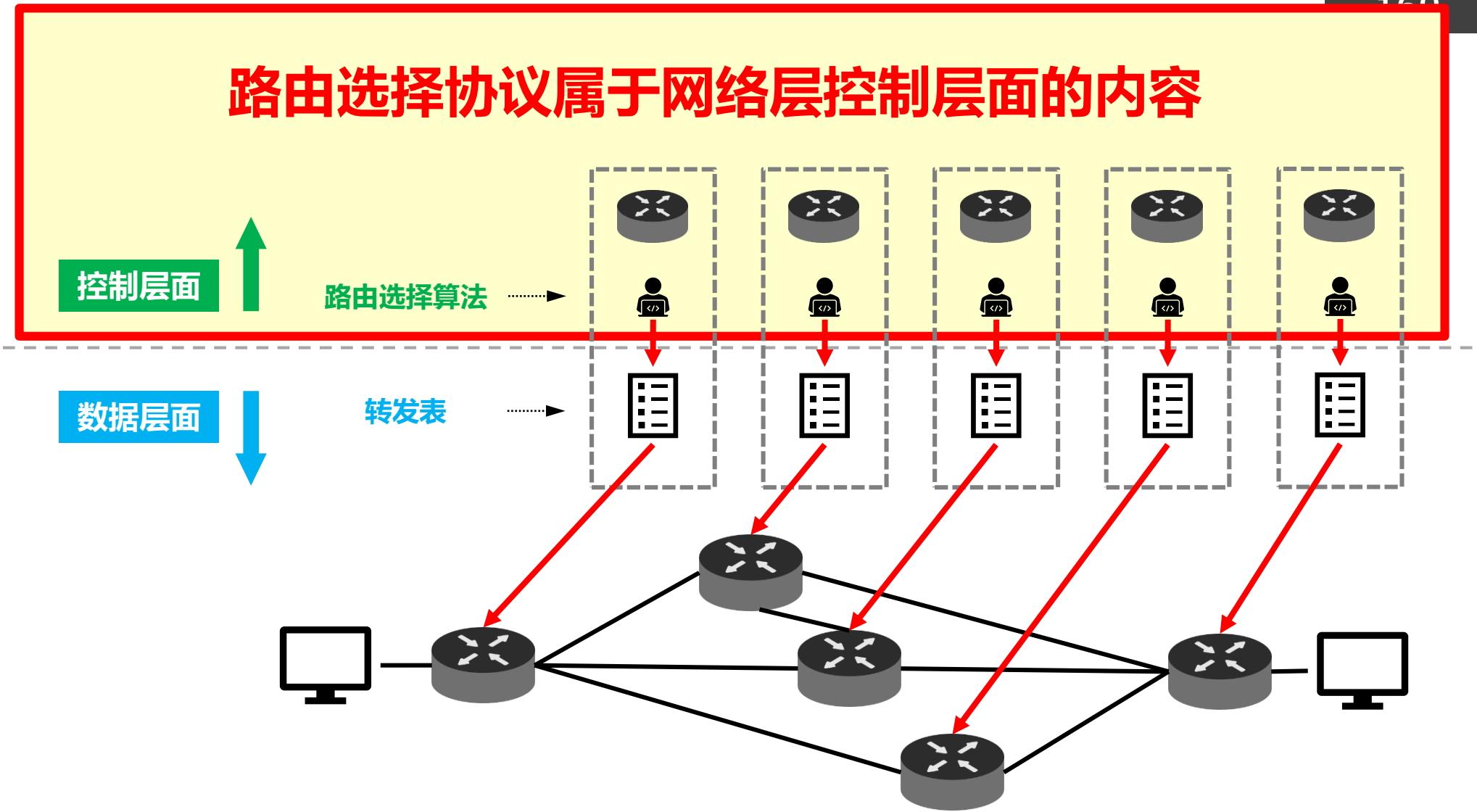
6. 路由选择协议

- 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。

6. 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。

路由选择协议属于网络层控制层面的内容



6. 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。

6. 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。

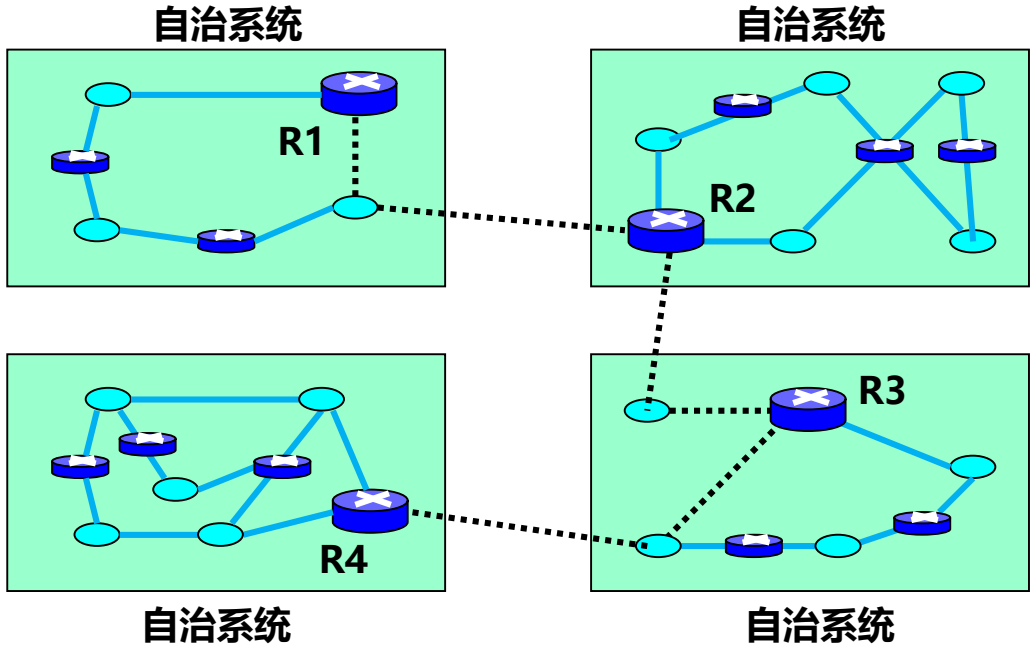
6. 路由选择协议

- 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。
  - 互联网采用自适应的（即动态的）、分布式路由选择协议。
    - 分为 2 个层次：
    - 自治系统之间的路由选择 或 域间路由选择 (interdomain routing)。
    - 自治系统内部的路由选择 或 域内路由选择 (intradomain routing)。

6. 路由选择协议

- 自治系统（Autonomous System，AS）：
  - 自治系统 AS 的定义：
    - 是在单一技术管理下的许多网络、IP地址以及路由器，而这些路由器使用一种自治系统内部的路由选择协议和共同的度量。
    - 每一个 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。
  - 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。

6. 路由选择协议

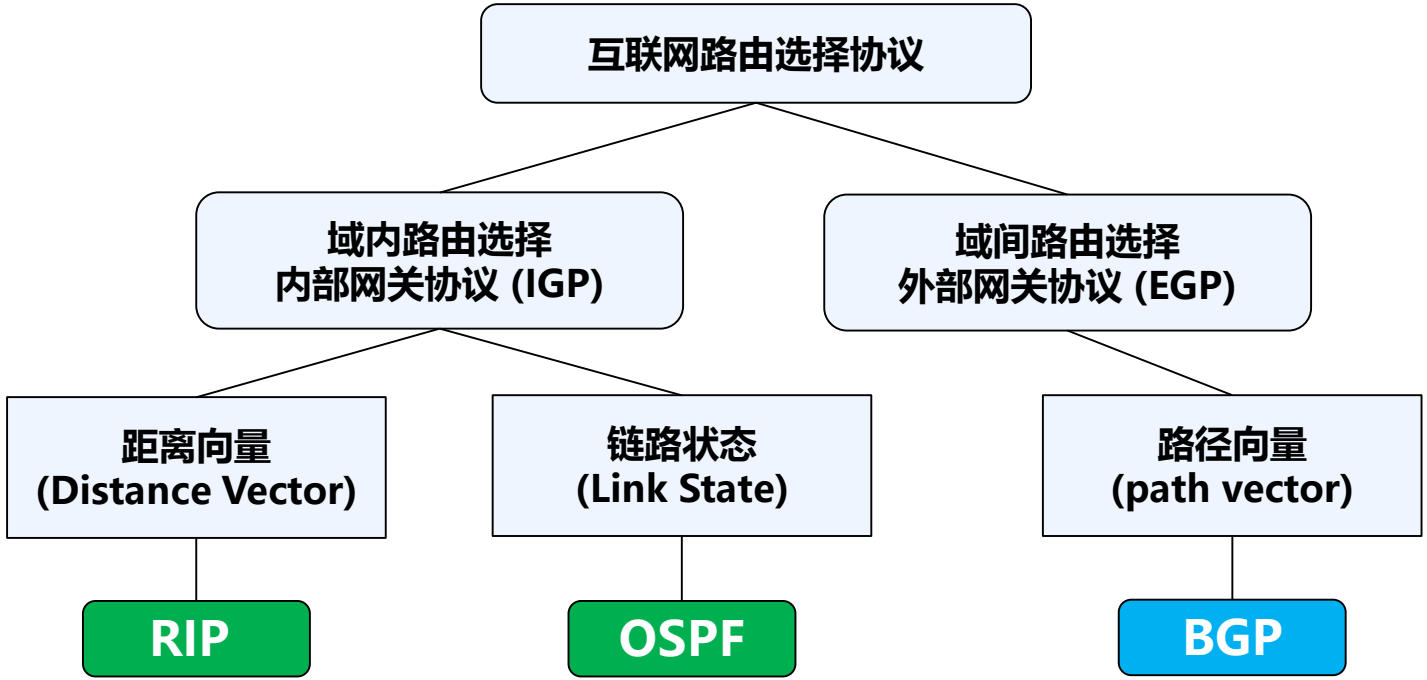


6. 路由选择协议

- 互联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。

6. 路由选择协议

- 因特网的两大类路由选择协议：

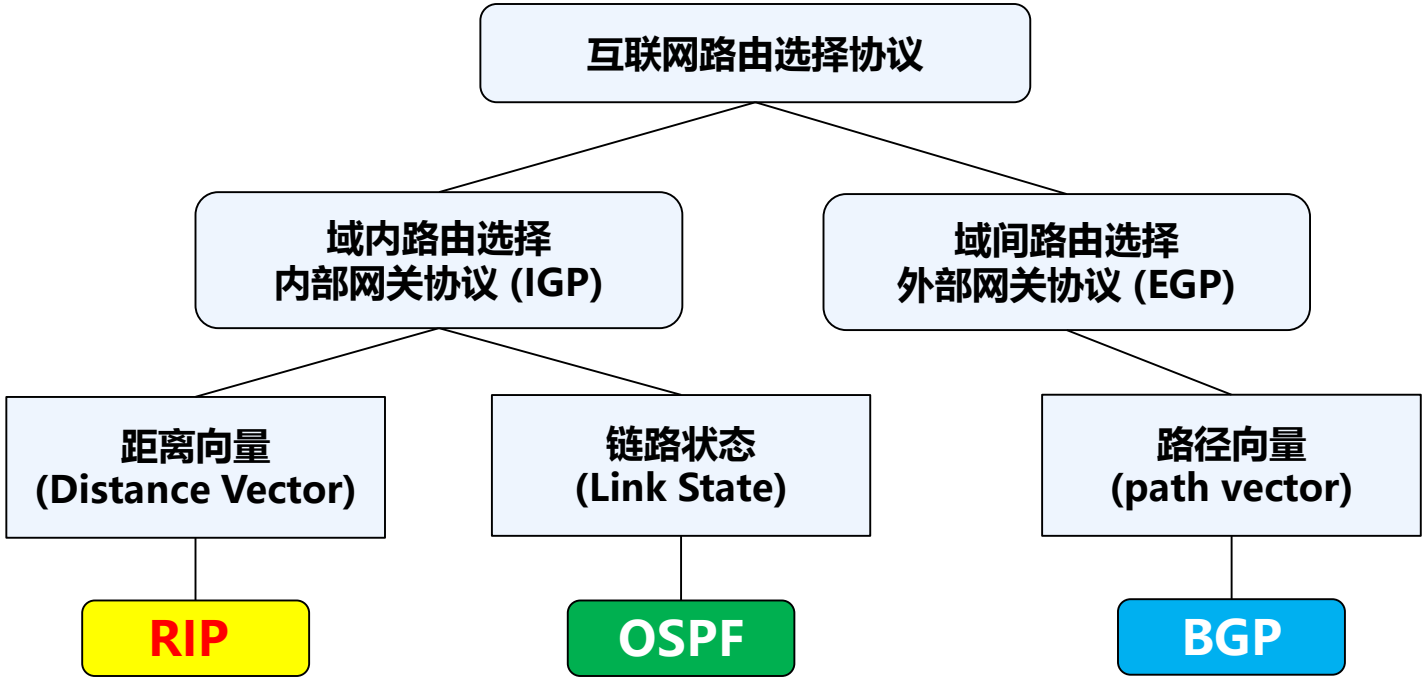


6. 路由选择协议

- 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)

6. 路由选择协议


互联网路由选择协议
域内路由选择
内部网关协议 (IGP)
域间路由选择
外部网关协议 (EGP)
距离向量
(Distance Vector)
链路状态
(Link State)
路径向量
(path vector)
RIP
OSPF
BGP

6. 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 最大优点是：简单。
  - RIP 要求网络中的每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。

广泛使用很重要，一种技术只有广泛使用，才能够形成生态体系。

6. 路由选择协议

- RIP 对“距离”的定义：
  - 从路由器到直接连接的网络的距离 = 1。
  - 路由器到非直接连接的网络的距离 = 所经过的路由器数 + 1。
  - RIP中的“距离”也称为“跳数” (hop count)，每经过一个路由器，跳数就加1。
- 好路由 = “距离短”的路由。
  - 最佳路由 = “距离最短”的路由。
  - 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，只选择距离最短”的路由。

6. 路由选择协议

- RIP 的三个特点：
  - 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。

6. 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。

6. 路由选择协议

- RIP 路由表的主要信息和更新规则：
  - RIP 路由表的主要信息是：目的网络、距离(最短)、下一跳地址

目的网络	距离（最短）	下一跳地址

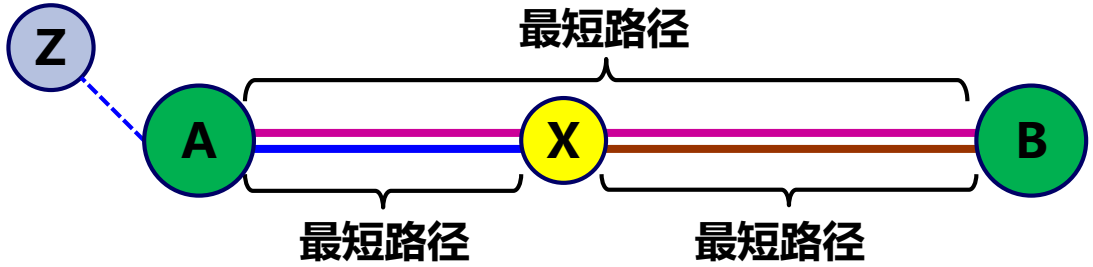
  - 路由表更新规则：
    - 使用距离向量算法找出到达每个目的网络的最短距离。

6. 路由选择协议

- 距离向量算法：
  - 对每个相邻路由器（假设其地址为 X）发送过来的 RIP 报文，路由器：
    1. 修改 RIP 报文中的所有项目（即路由）：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。
    2. 对修改后的 RIP 报文中的每一个项目，重复以下步骤：
       I. 若路由表中没有目的网络N，则把该项目添加到路由表中。否则
       II. 若路由表中网络 N 的下一跳路由器为 X，则用收到的项目替换原路由表中的项目。否则
       III. 若收到项目中的距离小于路由表中的距离，则用收到项目更新原路由表中的项目。否则
       IV. 什么也不做。
    3. 若 3 分钟还未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（表示不可达）。
    4. 返回。

6. 路由选择协议

- 距离向量算法：
  - 算法基础：
    - Bellman-Ford 算法（或 Ford-Fulkerson 算法）。
  - **算法要点**：
    - 设 X 是结点 A 到 B 的最短路径上的一个结点。
    - 若把路径 A→B 拆成两段路径 A→X 和 X→B，则每一段路径 A→X 和 X→B 也都分别是结点 A 到 X 和结点 X 到 B 的最短路径。



6. 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。

6. 路由选择协议

【例】已知路由器 R6 有表 4-8(a) 所示的路由表。现在收到相邻路由器 R4 发来的路由更新信息，如表 4-8(b) 所示。试更新路由器 R6 的路由表。

表 4-8(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
…	…	…

表 4-8(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

① 距离+1，修改下一跳地址

表 4-8(c) 修改后的表 4-8(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

②

计算更新

②

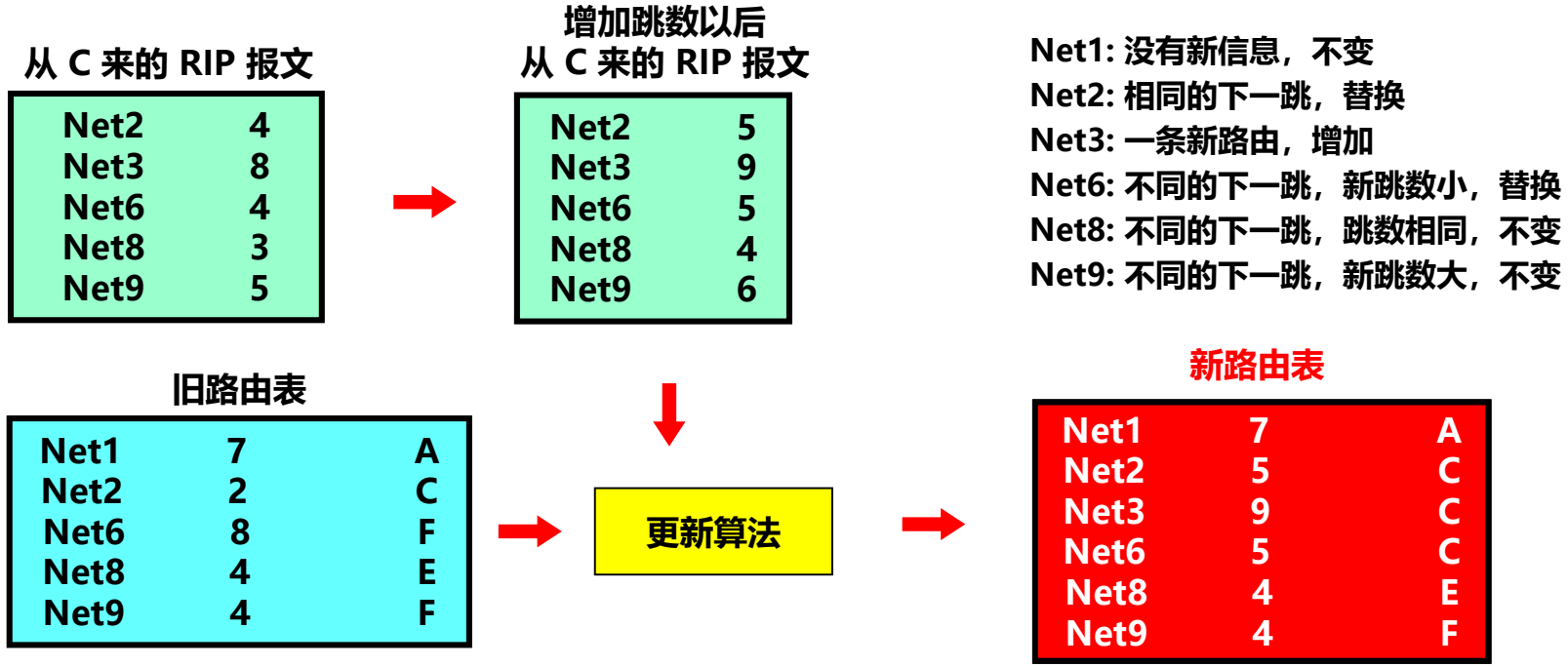
③

表 4-8(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
…	…	…

6. 路由选择协议

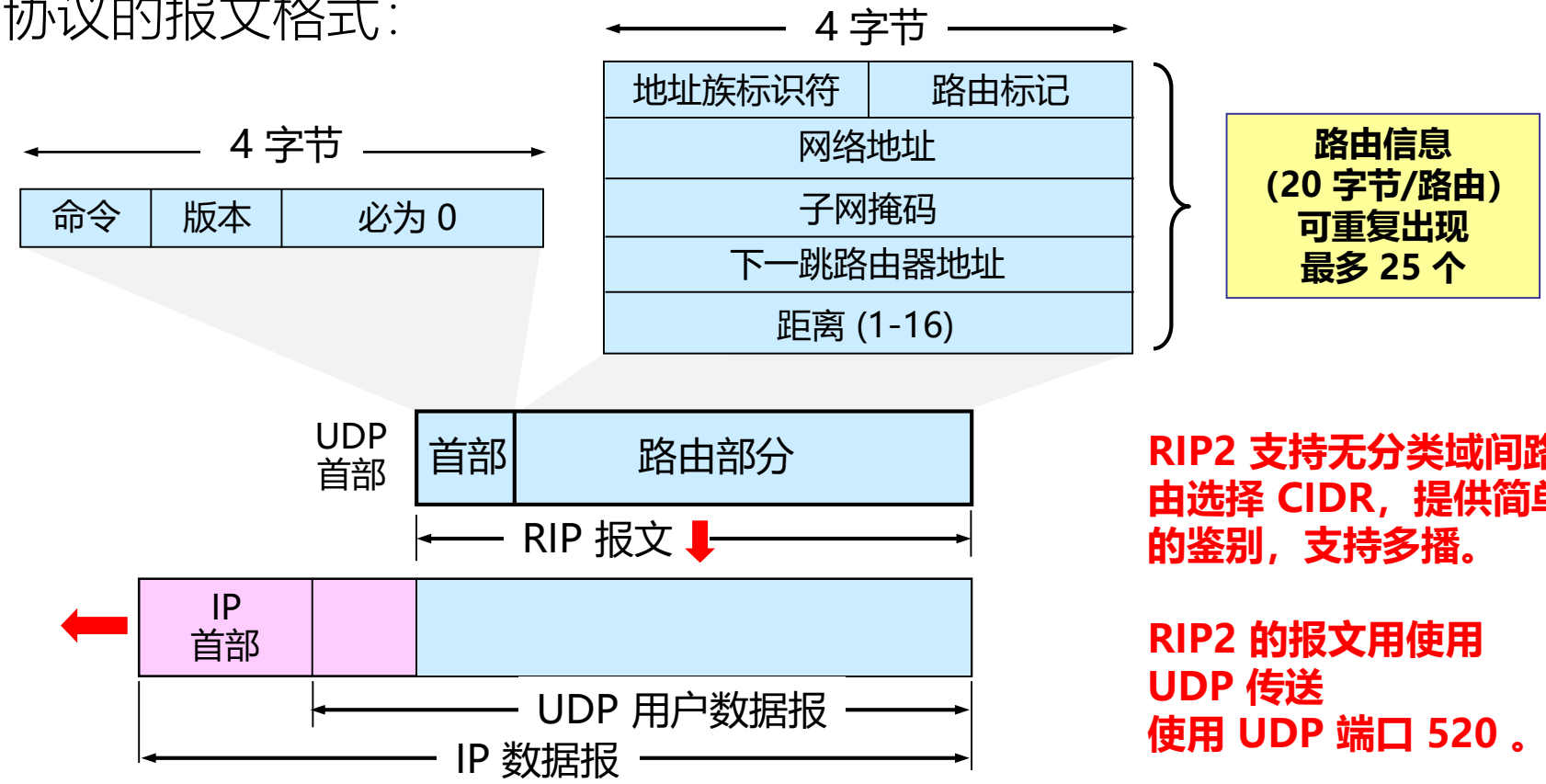
现场测试讨论举例：路由表更新。



Net1: 没有新信息，不变
Net2: 相同的下一跳，替换
Net3: 一条新路由，增加
Net6: 不同的下一跳，新跳数小，替换
Net8: 不同的下一跳，跳数相同，不变
Net9: 不同的下一跳，新跳数大，不变

6. 路由选择协议

- RIP2 协议的报文格式：



RIP2 支持无分类域间路由选择 CIDR，提供简单的鉴别，支持多播。

RIP2 的报文用使用 UDP 传送
使用 UDP 端口 520 。

6. 路由选择协议

- RIP2 报文：
  - 组成：首部和路由 2 个部分组成。
  - 路由部分：由若干个路由信息组成，每个路由信息共 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。
    - 后面为具体路由，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。

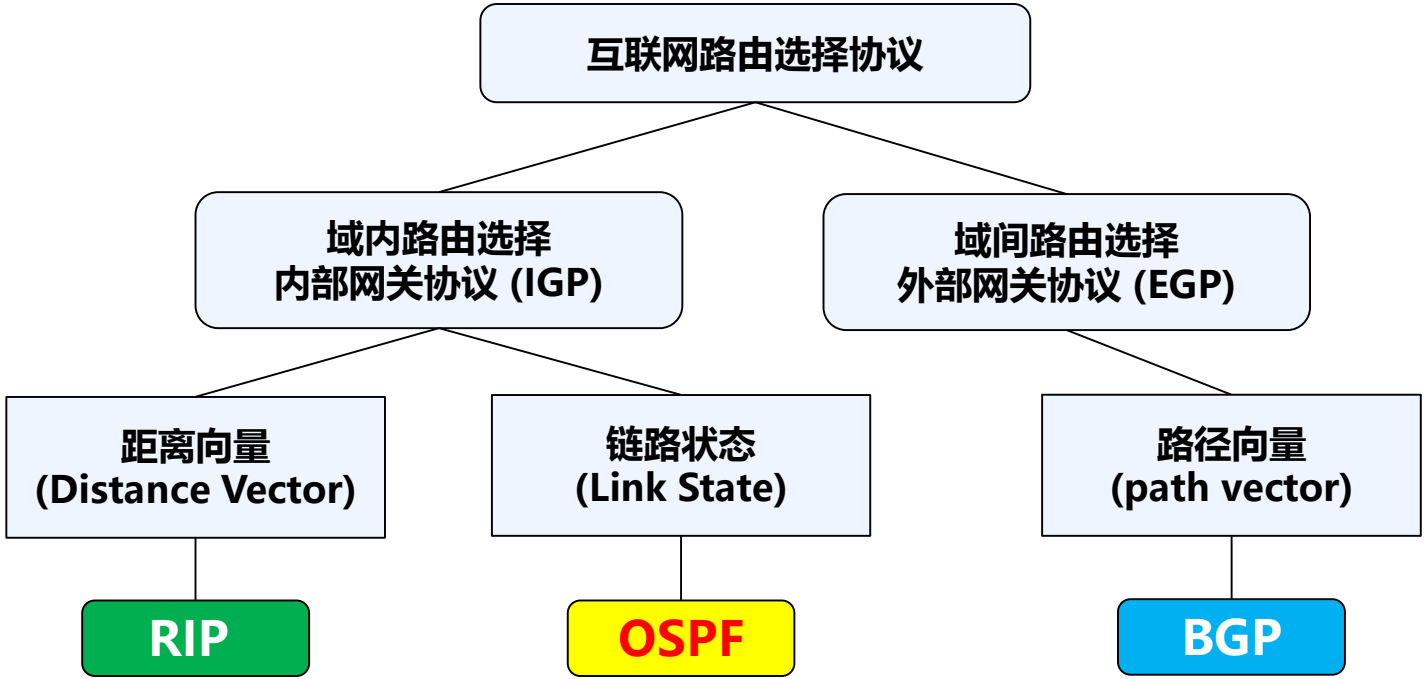
6. 路由选择协议

- 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。

6. 路由选择协议

- RIP 协议特点：
  - 特点：**好消息传播得快，坏消息传播得慢**
    - 当网络出现故障时，要经过比较长的时间才能将此信息（坏消息）传送到所有的路由器。慢收敛。
  - 优点：
    - 实现简单，开销较小。
  - 缺点：
    - 网络规模有限。最大距离为 15（16 表示不可达）。
    - 交换的路由信息为完整路由表，开销较大。
    - 坏消息传播得慢，收敛时间过长。 **（慢收敛）**

6. 路由选择协议


互联网路由选择协议
域内路由选择
内部网关协议 (IGP)
域间路由选择
外部网关协议 (EGP)
距离向量
(Distance Vector)
链路状态
(Link State)
路径向量
(path vector)
RIP
OSPF
BGP

6. 路由选择协议

- 开放最短路径优先（Open Shortest Path First，OSPF）是为了克服 RIP 的缺点，在1989年开发出来的。
  - 原理很简单，但实现很复杂。
  - 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：使用了 Dijkstra 提出的最短路径算法 SPF。
  - 采用分布式的链路状态协议 (link state protocol)。*
  - 现在使用 OSPFv2。
- OSPF只是协议名字，并不表示其他路由选择协议不是“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。

6. 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 的三个主要特点：
  - 采用洪泛法 (flooding)，向本自治系统中所有路由器发送信息。
  - 发送的信息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：说明本路由器都和哪些路由器相邻，以及该链路的度量 (metric)。
  - 当链路状态发生变化或每隔一段时间（如30分钟），路由器才用洪泛法向所有路由器发送此信息。

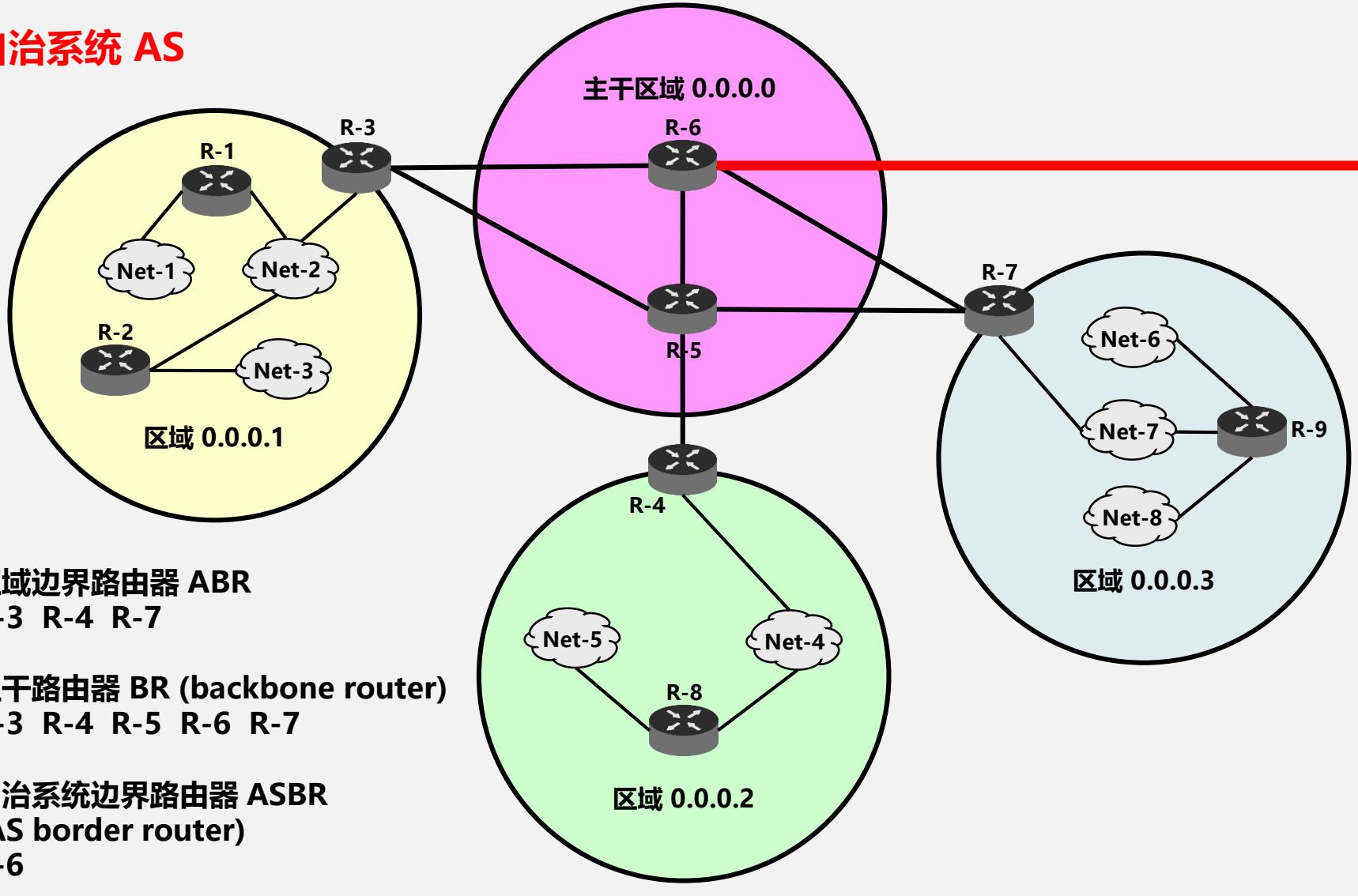
6. 路由选择协议

- 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）。
  - 每个路由器使用链路状态数据库中的数据构造自己的路由表
    - 例如，使用Dijkstra的最短路径路由算法，计算得出自己的路由表。
    - 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。
    - 重要优点：OSPF 更新过程收敛速度快。

6. 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。

自治系统 AS



区域边界路由器 ABR
R-3 R-4 R-7

主干路由器 BR (backbone router)
R-3 R-4 R-5 R-6 R-7

自治系统边界路由器 ASBR
(AS border router)
R-6

6. 路由选择协议

- OSPF 划分区域优点和缺点：
  - 优点：
    - 减少了整个网络上的通信量。
    - 减少了需要维护的状态数量。
  - 缺点：
    - 交换信息的种类增多了。
    - 使 OSPF 协议更加复杂了

分层次划分区域的好处：

使每一个区域内部交换路由信息的通信量大大减小，因而使 OSPF 协议能够用于规模很大的自治系统中。

6. 路由选择协议

- OSPF 划分区域优点和缺点：
  - 其他特点：
    - 对于不同类型的业务可计算出不同的路由。
    - 可实现多路径间的负载均衡（load balancing）。
    - 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。
    - 32 位的序号，序号越大状态就越新。
      - 全部序号空间在 600 年内不会产生重复号。

6. 路由选择协议

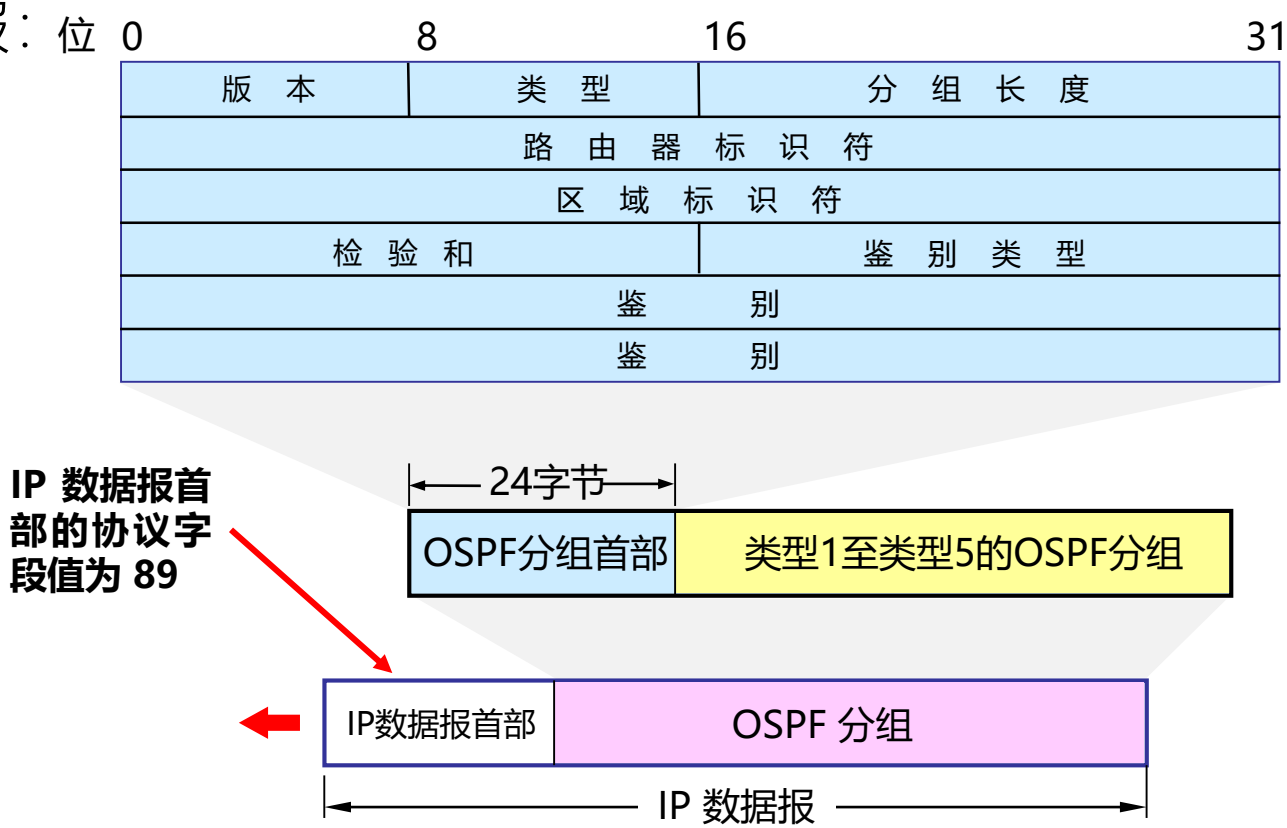
- OSPF 的五种分组类型：
  - 类型1：问候 (Hello) 分组。
  - 类型2：数据库描述 (Database Description) 分组。
  - 类型3：链路状态请求 (Link State Request) 分组。
  - 类型4：链路状态更新 (Link State Update) 分组。
    - 用洪泛法对全网更新链路状态。
  - 类型5：链路状态确认 (Link State Acknowledgment) 分组。

6. 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。

6. 路由选择协议

- OSPF 数据报：



6. 路由选择协议

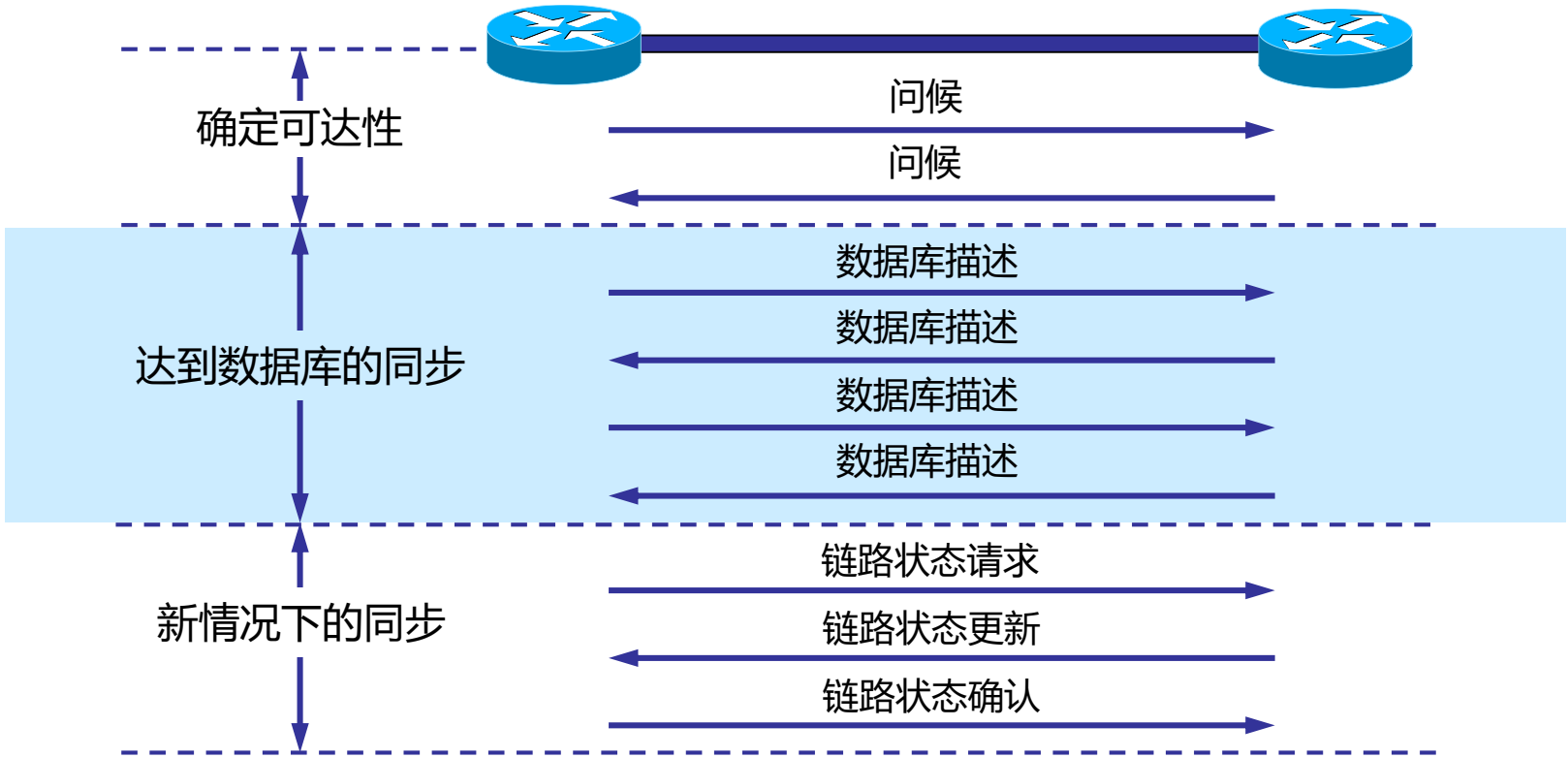
- OSPF 工作过程
  - 第1：确定邻站可达。
    - 相邻路由器每隔 10 秒钟要交换一次问候分组。
    - 若有 40 秒钟没有收到某个相邻路由器发来的问候分组，则可认为该相邻路由器是不可达的。
  - 第2：同步链路状态数据库。
    - 同步：指不同路由器的链路状态数据库的内容是一样的。
    - 两个同步的路由器叫做完全邻接的 (fully adjacent) 路由器。
    - 不是完全邻接的路由器：虽然在物理上是相邻的，但其链路状态数据库并没有达到一致。

6. 路由选择协议

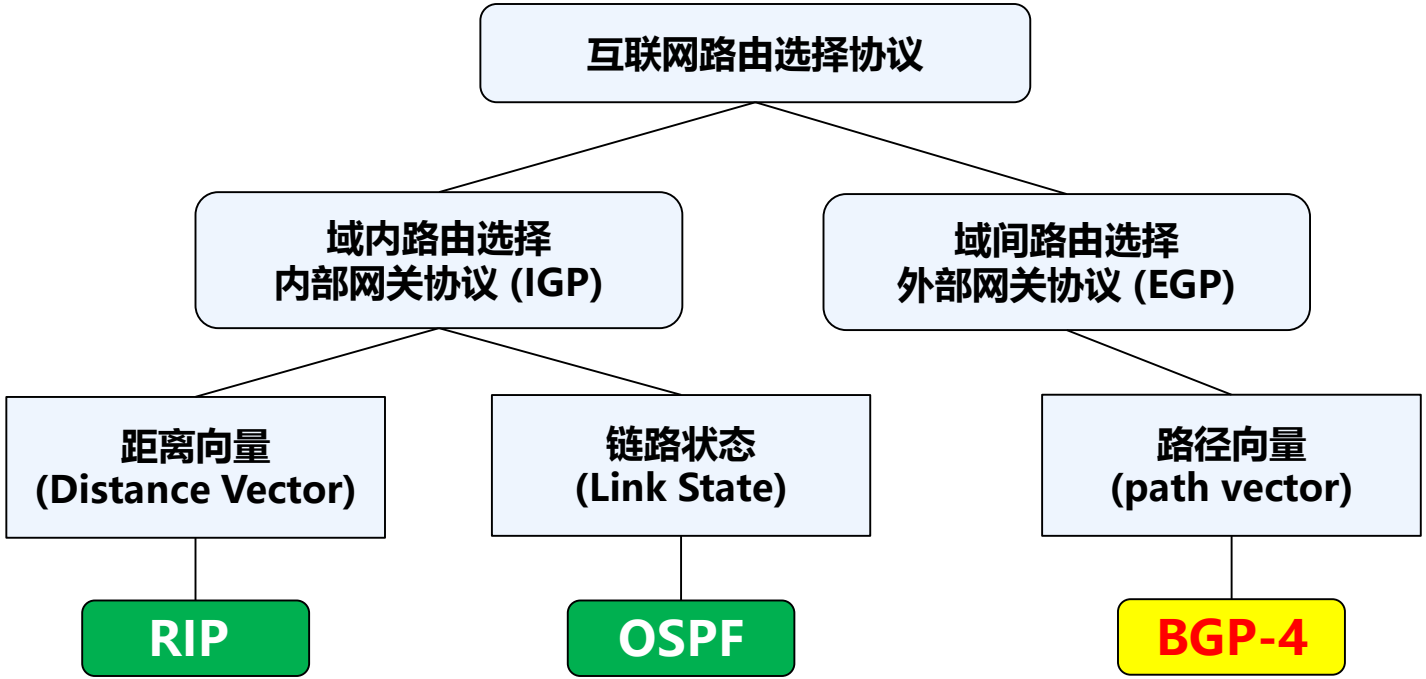
- OSPF 工作过程
  - 第3：更新链路状态。
    - 只要链路状态发生变化，路由器就使用链路状态更新分组，采用可靠的洪泛法向全网更新链路状态。
    - 为确保链路状态数据库与全网的状态保持一致，OSPF 还规定：每隔一段时间，如 30 分钟，要刷新一次数据库中的链路状态。
    - OSPF 链路状态只涉及相邻路由器，与整个互联网的规模并无直接关系，因此当互联网规模很大时，OSPF 协议要比距离向量协议 RIP 好得多。
    - OSPF 没有“坏消息传播得慢”的问题，收敛速度快。

6. 路由选择协议

- OSPF 工作过程：



6. 路由选择协议


互联网路由选择协议
域内路由选择
内部网关协议 (IGP)
域间路由选择
外部网关协议 (EGP)
距离向量
(Distance Vector)
链路状态
(Link State)
路径向量
(path vector)
RIP
OSPF
BGP-4

6. 路由选择协议

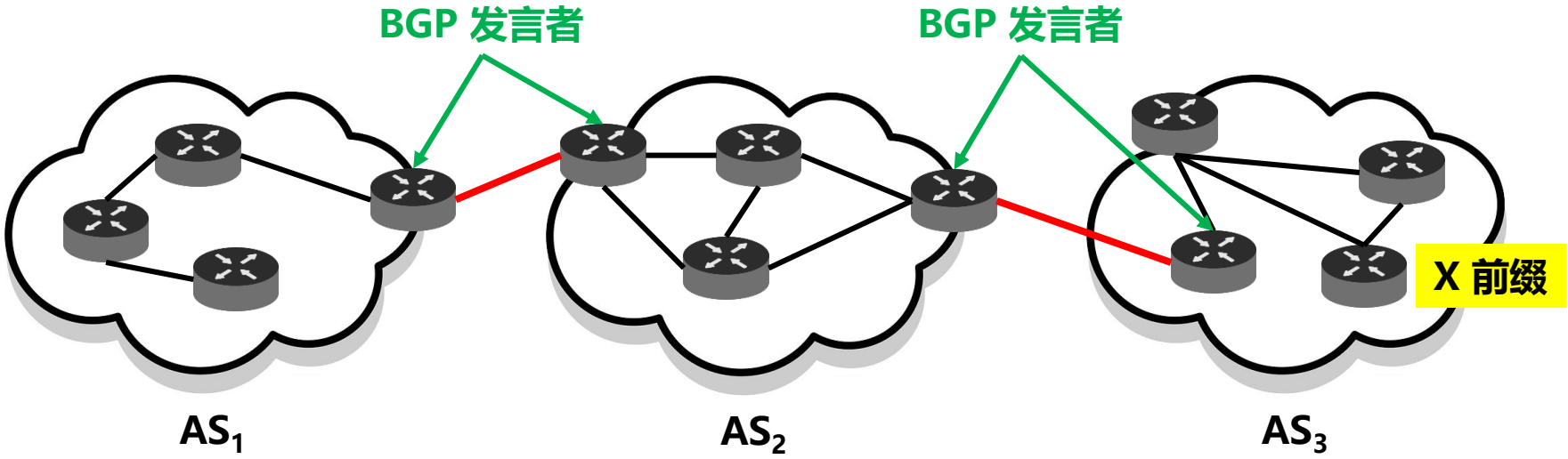
- 边界网关协议 BGP 是不同自治系统的路由器之间交换路由信息的协议。
- BGP 较新版本是2006年1月发表的 BGP-4。
  - BGP 第4个版本，即RFC 4271 ~ 4278。
  - 为了简单起见，通常将 BGP-4 简写为 BGP。

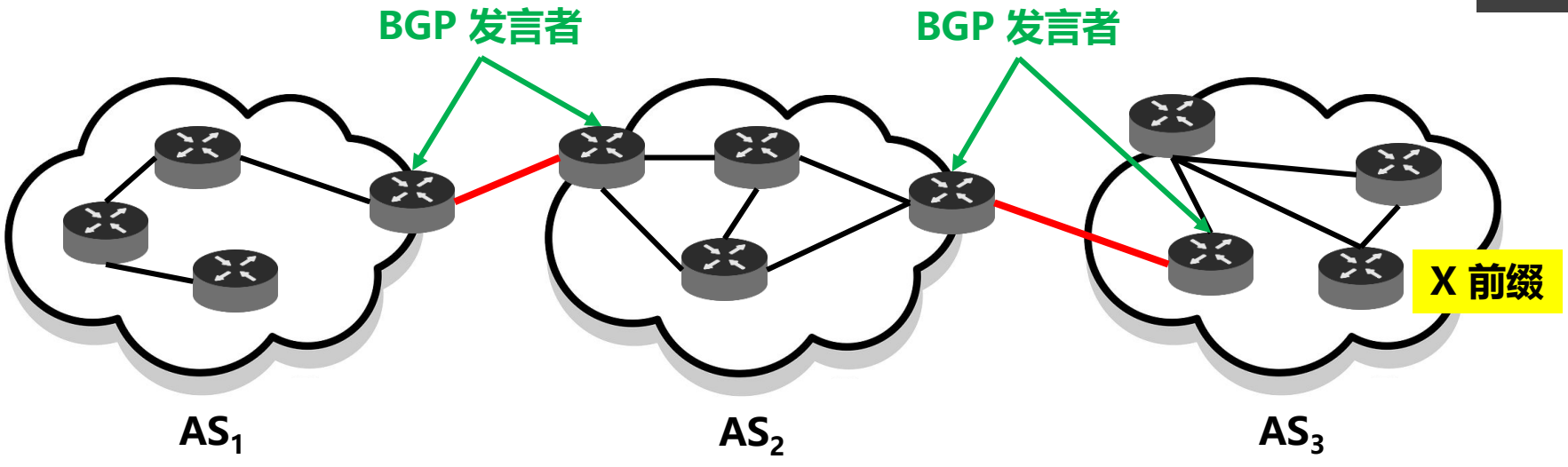
6. 路由选择协议

- BGP 的三个主要特点
  - 用于自治系统 AS 之间的路由选择。
  - 只能是力求选择出一条能够到达目的网络且**比较好的路由**（不能兜圈子），而并非要计算出一条最佳路由。
    - 因特网规模太大，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。
    - 自治系统之间的路由选择必须考虑有关策略。
      - 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路由是不太可能的。
      - 比较合理的做法是在 AS 之间交换“可达性”信息。
  - 采用了**路径向量 (path vector)** 路由选择协议。

6. 路由选择协议

- BGP 发言者（BGP speaker）
  - 每个自治系统选择至少一个路由器作为该自治系统的“ BGP 发言者”。
  - 两个 BGP 发言者通过一个共享网络连接在一起的。
    - BGP 发言者可以是 BGP 边界路由器，也可以不是。





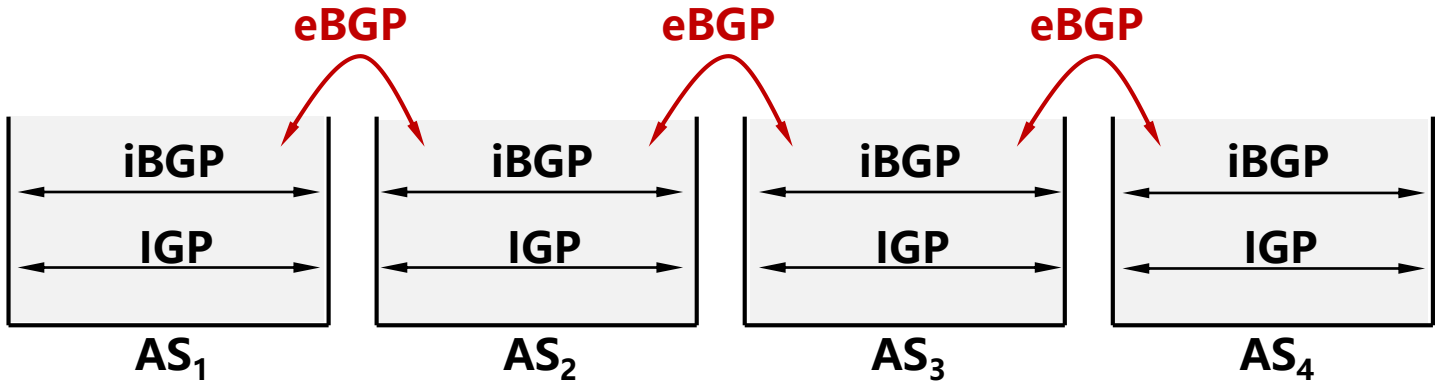
在 AS 之间， BGP 发言者在半永久性 TCP 连接（端口号为179）上建立 BGP 会话(session)。这种连接又称为 eBGP 连接。

在 AS 内部，任何相互通信的两个路由器之间必须有一个逻辑连接（也使用 TCP 连接）。AS 内部所有的路由器之间的通信是全连通的。这种连接常称为 iBGP 连接。

eBGP (external BGP) 连接：运行 eBGP 协议，在不同 AS 之间交换路由信息。

iBGP (internal BGP) 连接： 运行 iBGP 协议， 在 AS 内部的路由器之间交换 BGP 路由信息。

IGP、iBGP 和 eBGP 的关系



在 AS 内部运行：

- 内部网关协议 IGP，如 OSPF 或 RIP
- 协议 iBGP。

在 AS 之间运行：

- 协议 eBGP

eBGP 和 iBGP

- 同一个协议 BGP
  - 使用的报文类型、使用的属性、使用的状态机等都完全一样。
- 但它们在通报前缀时采用的规则不同：
  - 从 eBGP 连接的对等端得知的前缀信息，可以通报给一个 iBGP 连接的对等端。
  - 从 iBGP 连接的对等端得知的前缀信息，可以通报给一个 eBGP 连接的对等端。
  - 从 iBGP 连接的对等端得知的前缀信息，不能通报给一个 iBGP 连接的对等端。

6. 路由选择协议

- BGP 路由

BGP 路由 = [前缀, BGP属性] = [前缀, AS-PATH, NEXT-HOP]

去哪里

BGP 前缀：

✓ 指明到哪一个子网（用 CIDR 记法表示）

如何去

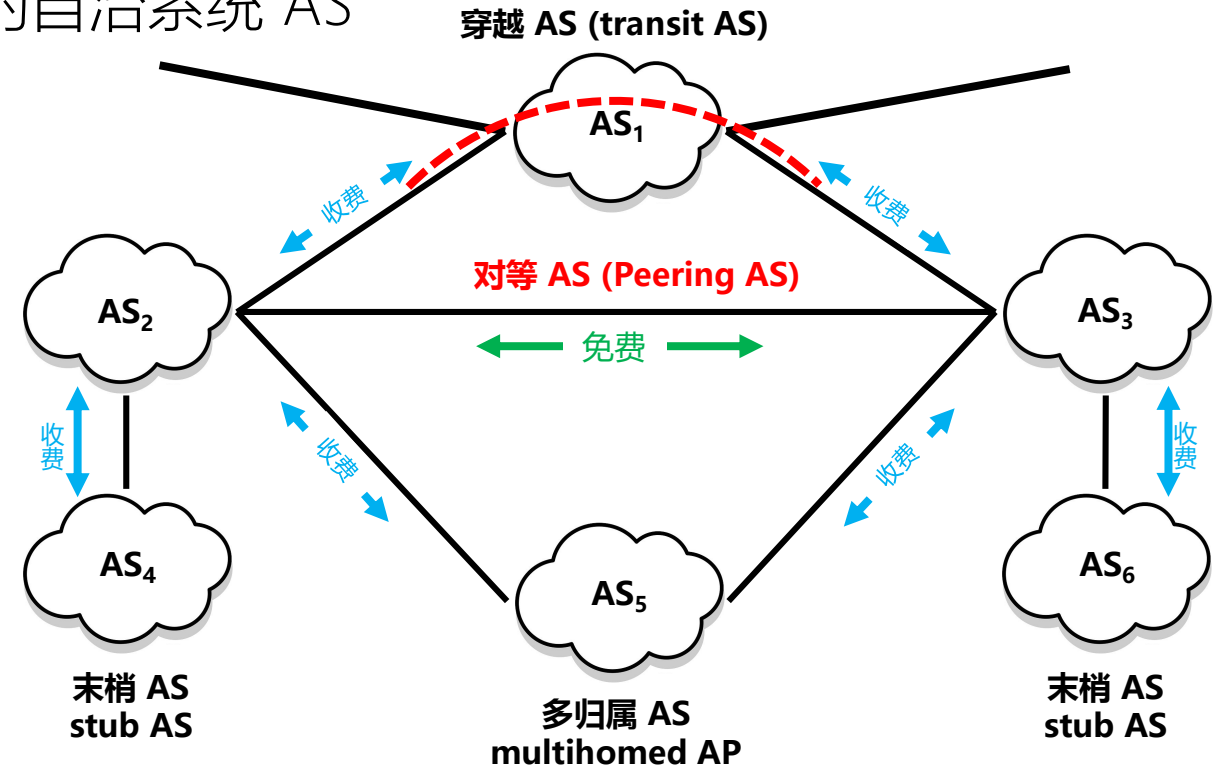
BGP 重要的两个属性是：

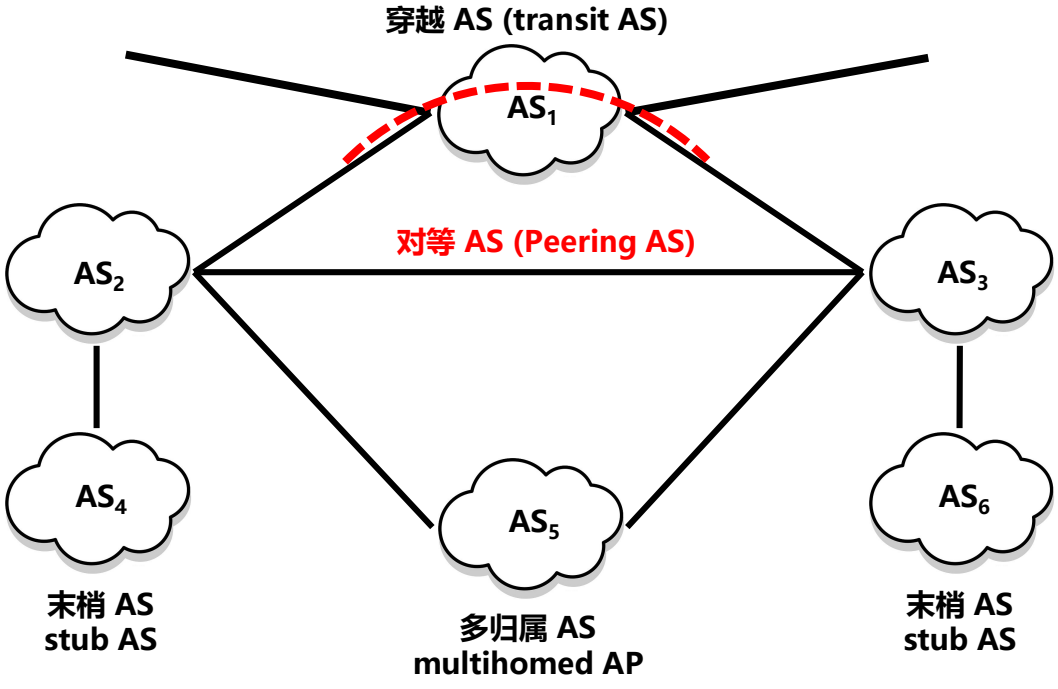
✓ 自治系统路径　　AS-PATH

✓ 下一跳　　NEXT-HOP

6. 路由选择协议

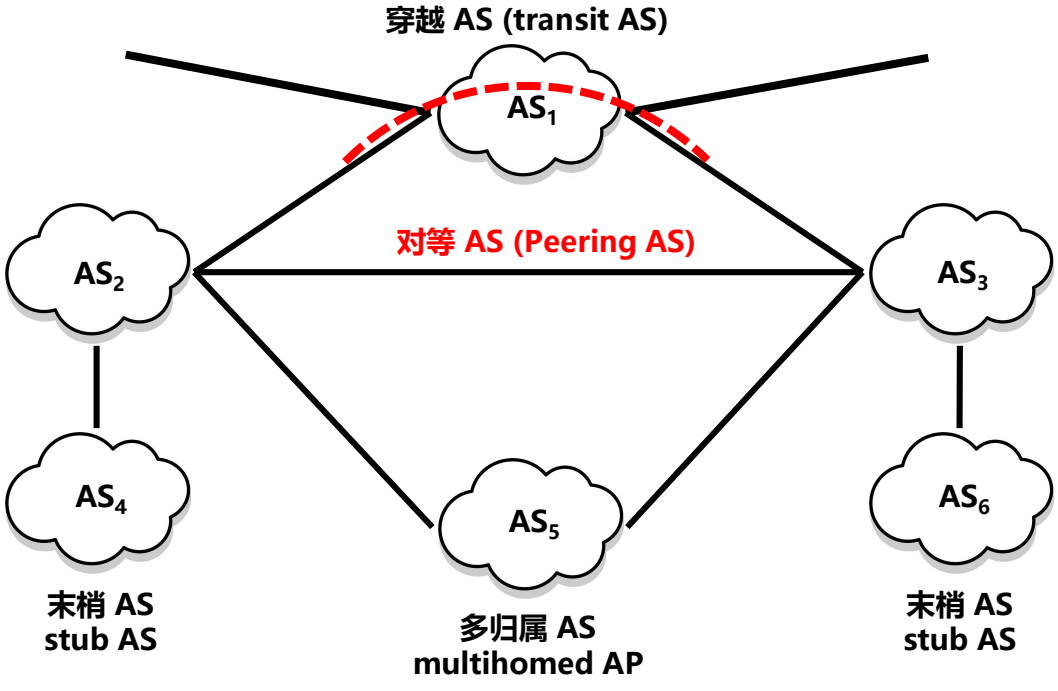
- 三种不同的自治系统 AS





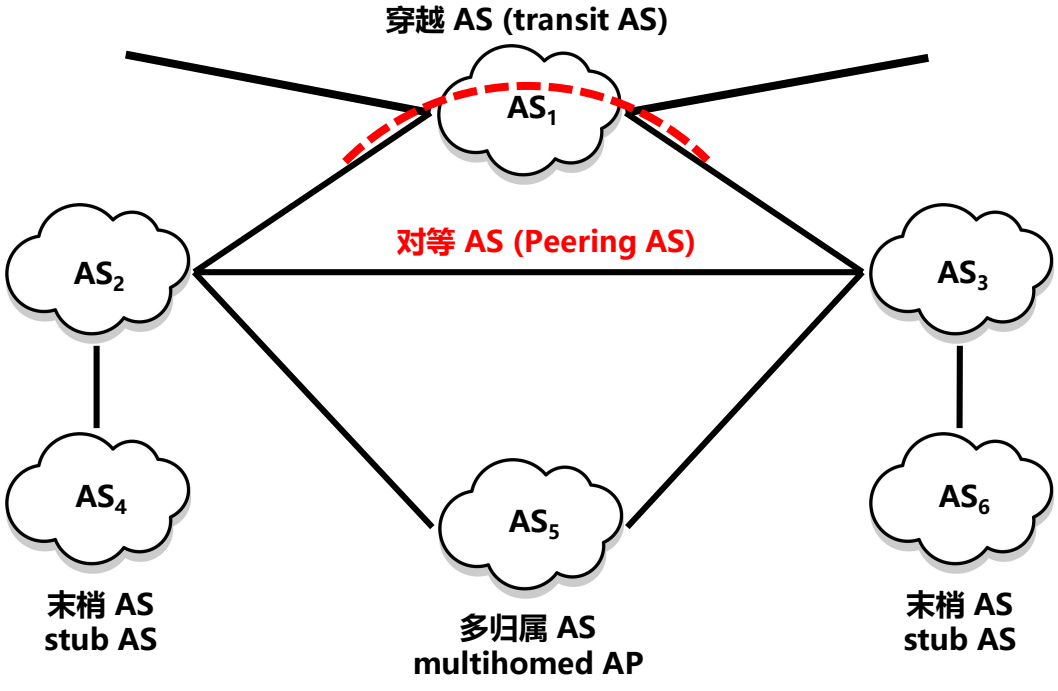
末梢 AS，stub AS：

- 比较小的 AS，把分组发送给直接连接的 AS，或者接收直接连接的 AS 发来的分组。
- 不会把来自其他 AS 的分组再转发到另一个 AS。
- 末梢 AS 必须向所连接的 AS 付费才能够发送或接收分组。



多归属 AS，multihomed AS：

- 末梢 AS 可以同时连接到两个或者多个 AS，称为多归属 AS。
- 多归属 AS 可以增加连接的可靠性，一条连接故障，其他连接可用。
- 多归属 AS 不能成为穿越 AS。



对等 AS，peering AS：

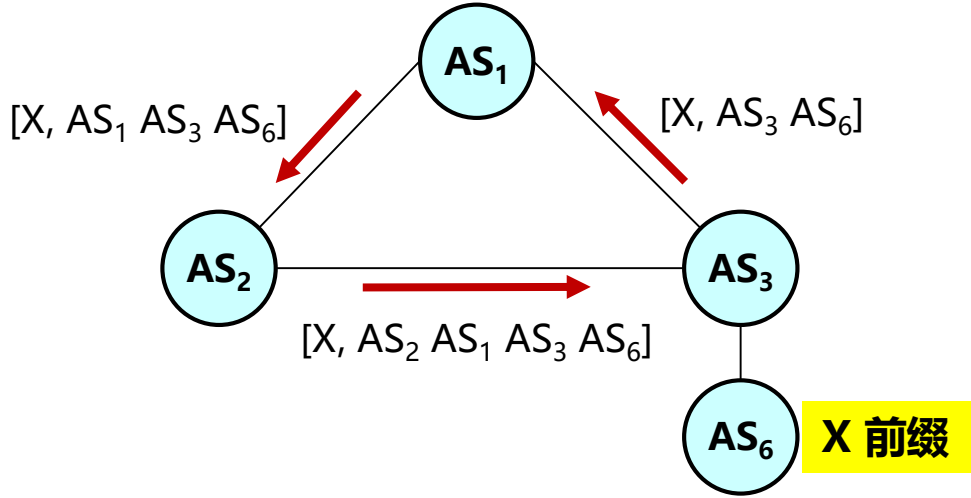
- 经过事先协商的两个 AS。
- 对等 AS 之间发送和接收 AS 不收费。

穿越 AS，transit AS：

- 拥有高速通信干线的主干 AS。
- 用于为其他 AS 有偿转发分组。
- 很多 AS 连接到穿越 AS 上。

BGP 路由如何避免兜圈子?

- 在属性 AS-PATH 中，不允许出现相同的 AS 号。
  - AS3 检查收到的 BGP 路由的 AS-PATH 中已经有了自己，
  - 立即删除掉这条路由，从而避免兜圈子路由的出现。



6. 路由选择协议

- BGP 的路由选择
  - 从一个 AS 到另一个 AS 中的前缀 X 只有一条路由，不存在选择的问题。
  - 从一个 AS 到另一个 AS 中的前缀 X 不止一条路由，选择较好的 BGP 路由。

1. **本地偏好 (local preference) 值最高的路由 （默认值=100）**
2. **AS 跳数最小的路由**
3. **使用热土豆路由选择算法（分组在 AS 内的转发次数最少）**
4. **路由器 BGP ID 数值最小的路由。具有多个接口的路由器有多个 IP 地址，BGP ID 就使用该路由器的 IP 地址中数值最大的一个。**

6. 路由选择协议

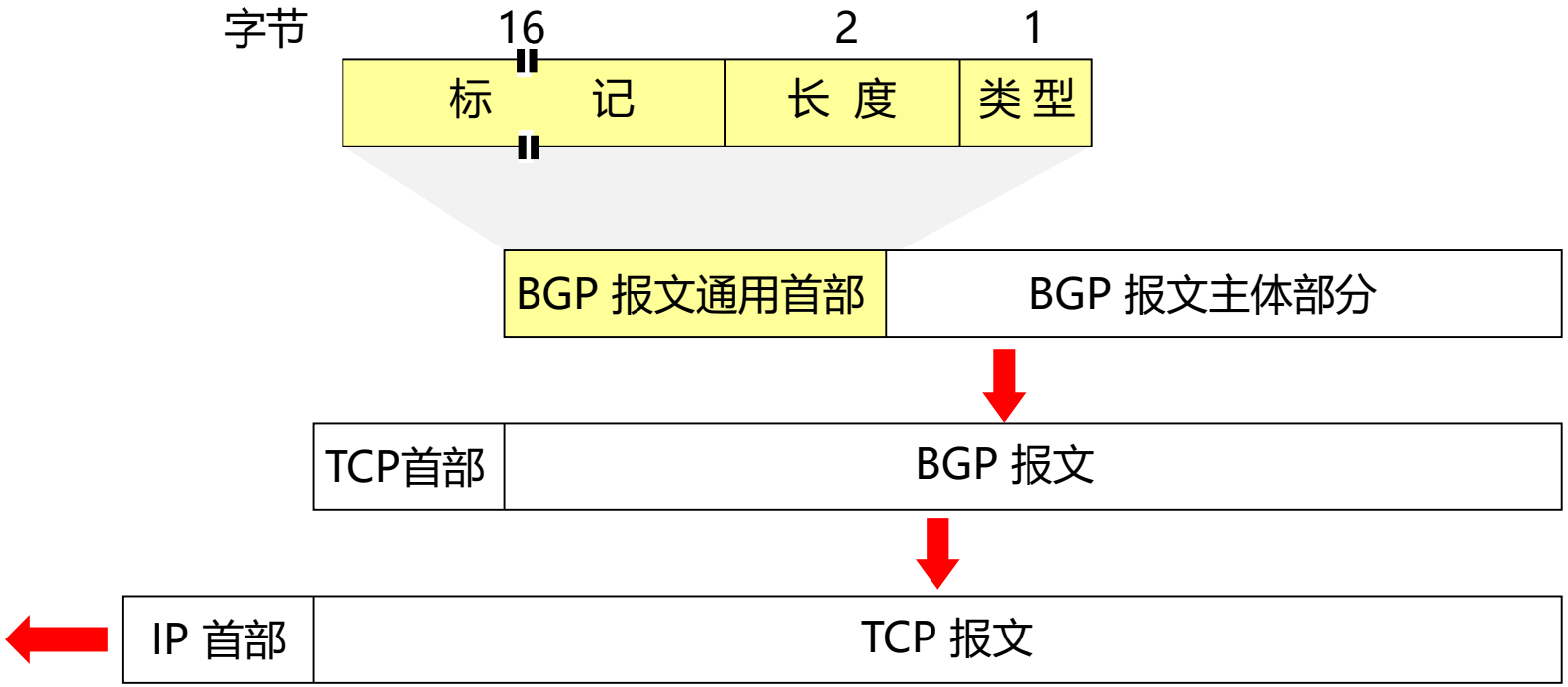
- BGP-4 的四种报文：

OPEN (打开)	• 用来与相邻的另一个 BGP 发言者建立关系，使通信初始化。
UPDATE (更新)	• 用来通告某一路由的信息，以及列出要撤销的多条路由。
KEEPALIVE (保活)	• 用来周期性地证实邻站的连通性。
NOTIFICATION (通知)	• 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。

6. 路由选择协议

- BGP 报文结构：

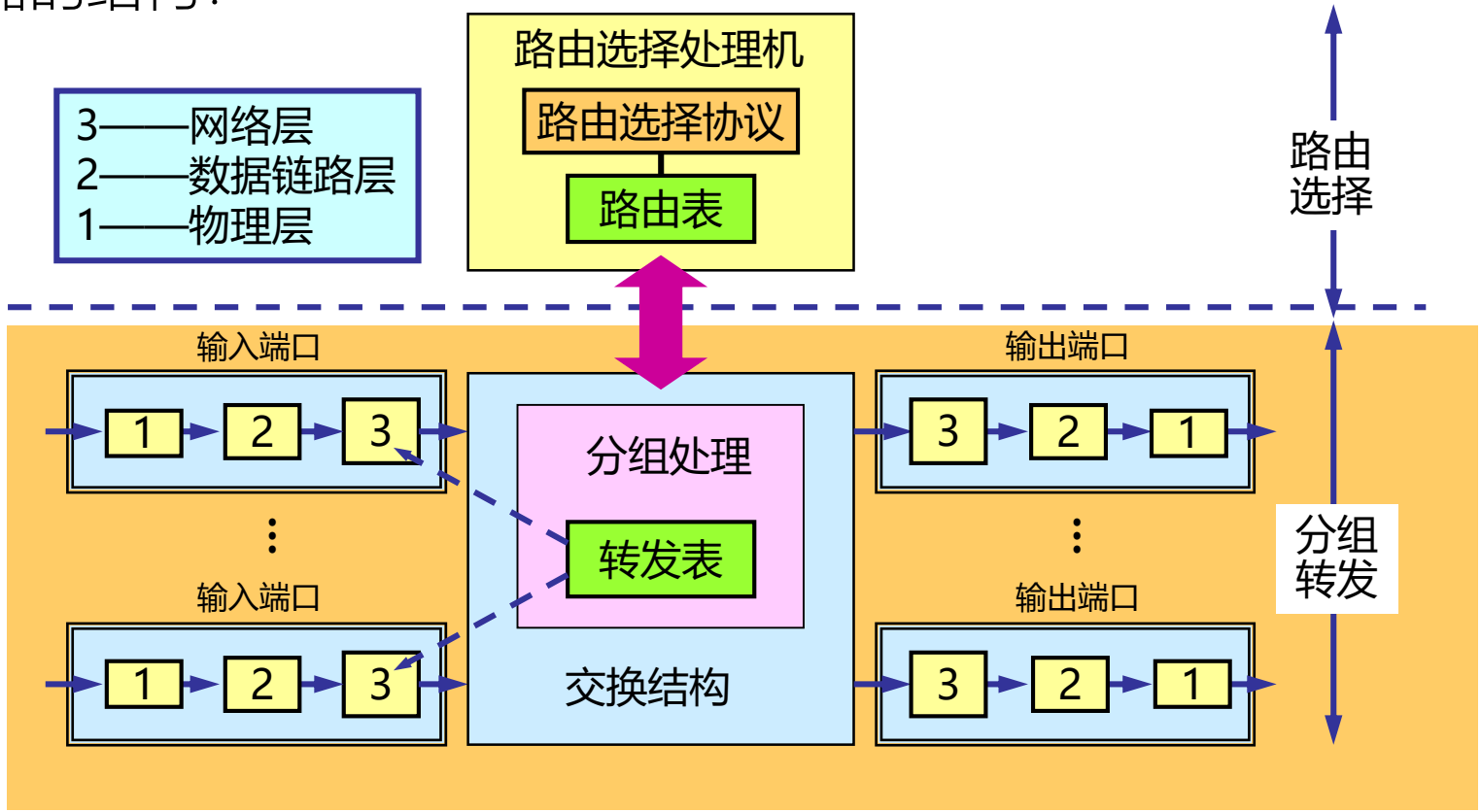


6. 路由选择协议

- 路由器的结构：
  - 路由器工作在网络层，用于互连网络。
  - 路由器是互联网中的关键设备。
  - 路由器的主要工作：转发分组。
    - 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。
    - 路由器某个输入端口收到分组，按照分组要去的目的网络，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。

6. 路由选择协议

- 路由器的结构：



6. 路由选择协议

- 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。

6. 路由选择协议

- 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

“转发”和“路由选择”的区别

转发

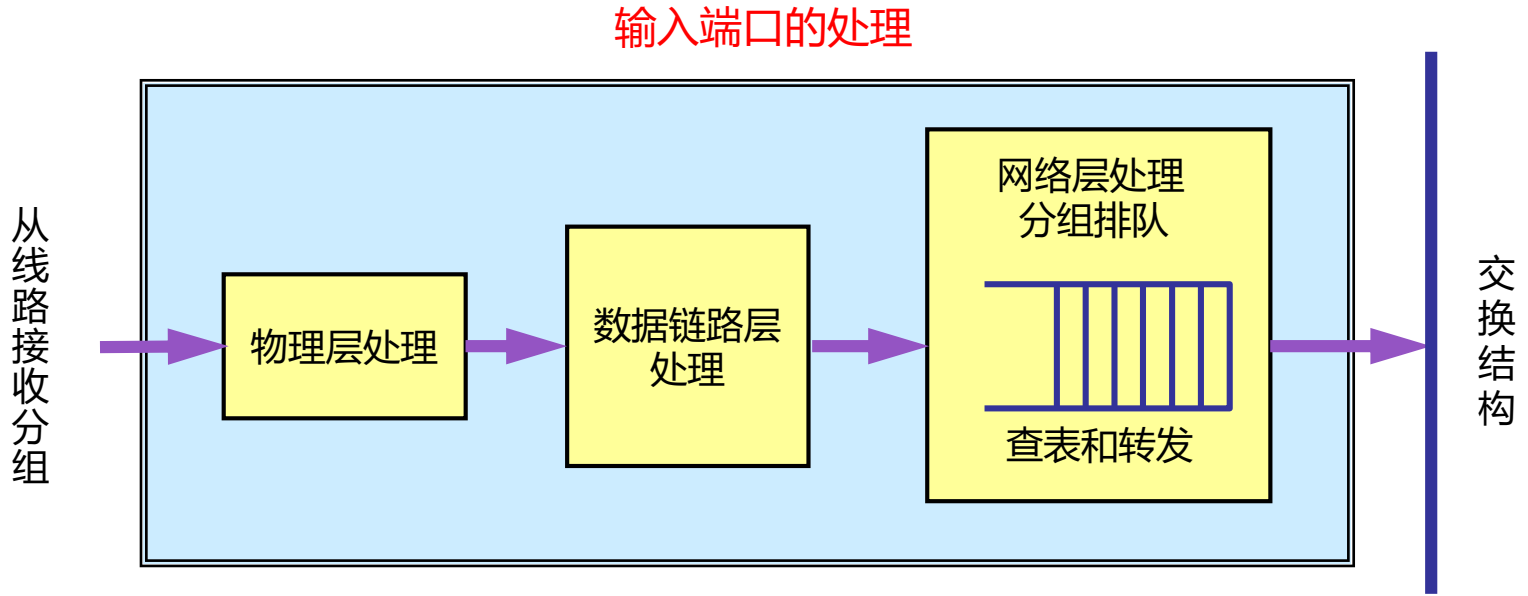
- 根据转发表将用户的 IP 数据报从合适的端口转发出去。
- 仅涉及到一个路由器。
- 转发表是从路由表得出的。
- 转发表必须包含完成转发功能所必需的信息，每一行必须包含从要到达的目的网络到输出端口和某些 MAC 地址信息（如下一跳的以太网地址）的映射。

路由选择

- 按照路由选择算法，根据网络拓扑的变化情况，动态地改变所选择的路由，并由此构造出整个路由表。
- 涉及到很多路由器。
- 路由表一般仅包含从目的网络到下一跳（用 IP 地址表示）的映射。

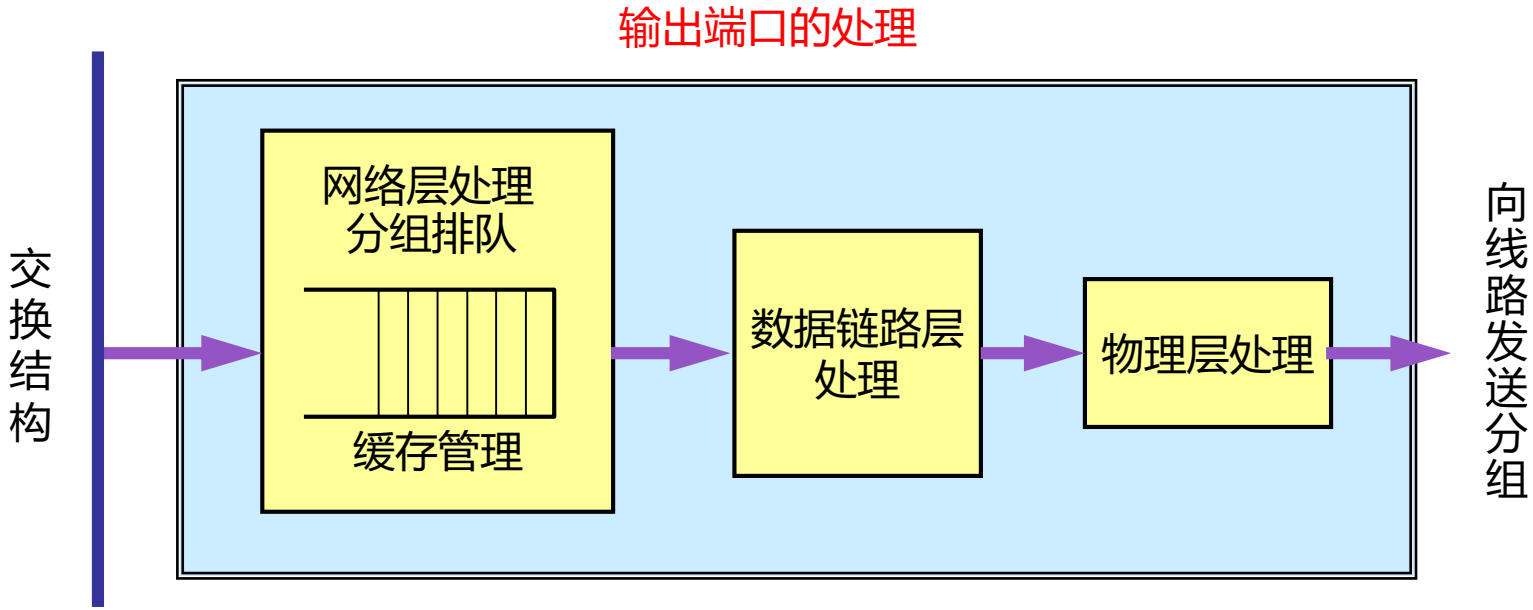
6. 路由选择协议

- 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。



6. 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。



6. 路由选择协议

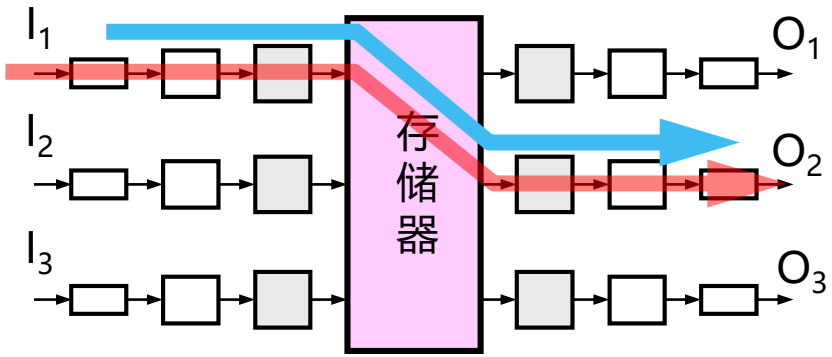
- 分组丢弃：
  - 若路由器处理分组的速率赶不上分组进入队列的速率，则队列的存储空间最终必定减少到零，这就使后面再进入队列的分组由于没有存储空间而只能被丢弃。
  - 路由器中的输入或输出队列产生溢出是造成分组丢失的重要原因。

6. 路由选择协议

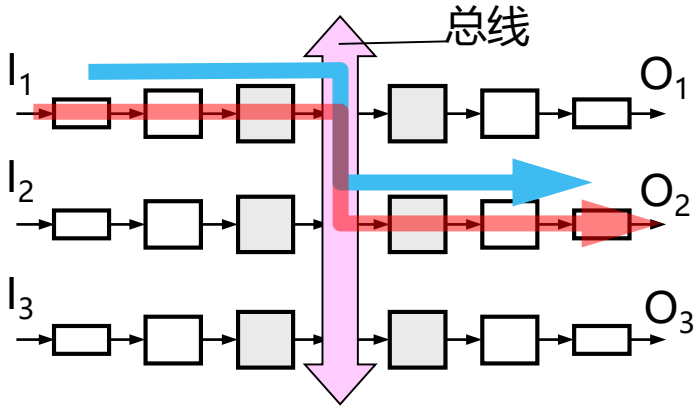
- 交换结构：
  - 交换结构是路由器的关键构件。
  - 正是这个交换结构把分组从一个输入端口转移到某个合适的输出端口。
  - 实现交换有多种方法，常用交换方法有三种：
    - 通过存储器
    - 通过总线
    - 通过纵横交换结构

6. 路由选择协议

- 交换结构：



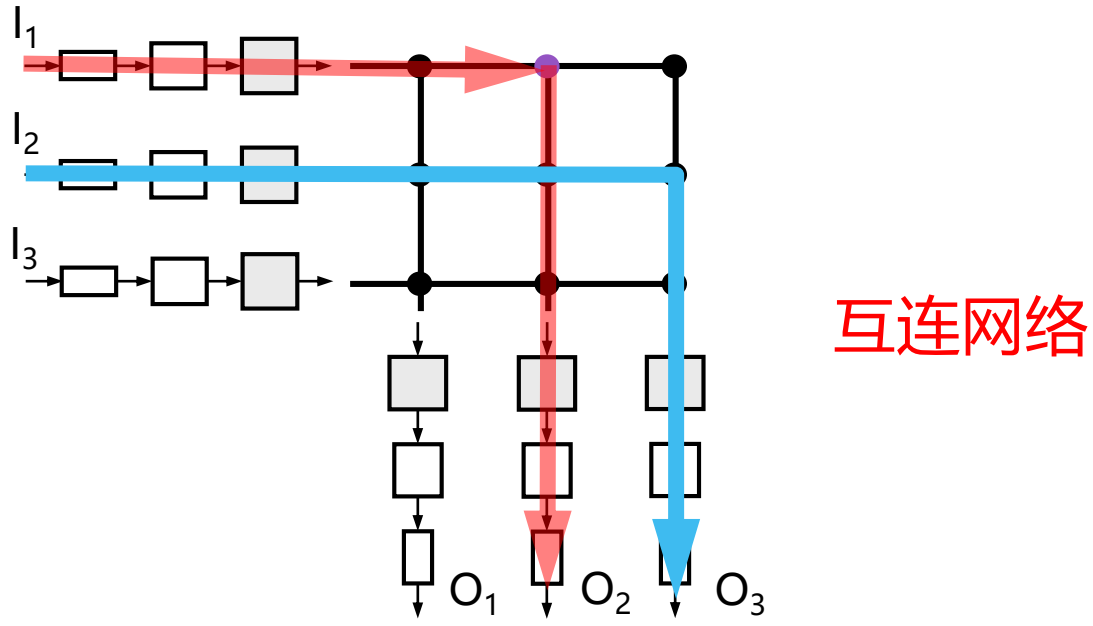
(a) 通过存储器



(b) 通过总线

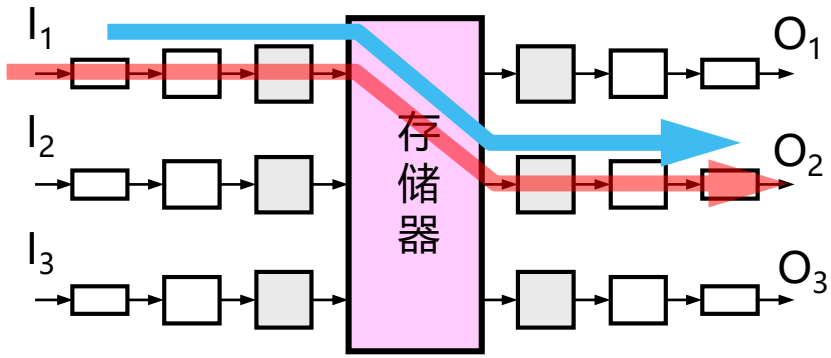
6. 路由选择协议

- 交换结构：



(c) 通过互连网络

6. 路由选择协议

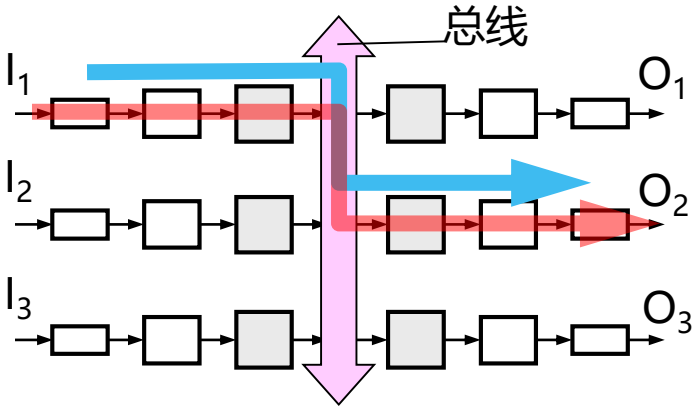


(a) 通过存储器

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 $M/2$。

6. 路由选择协议

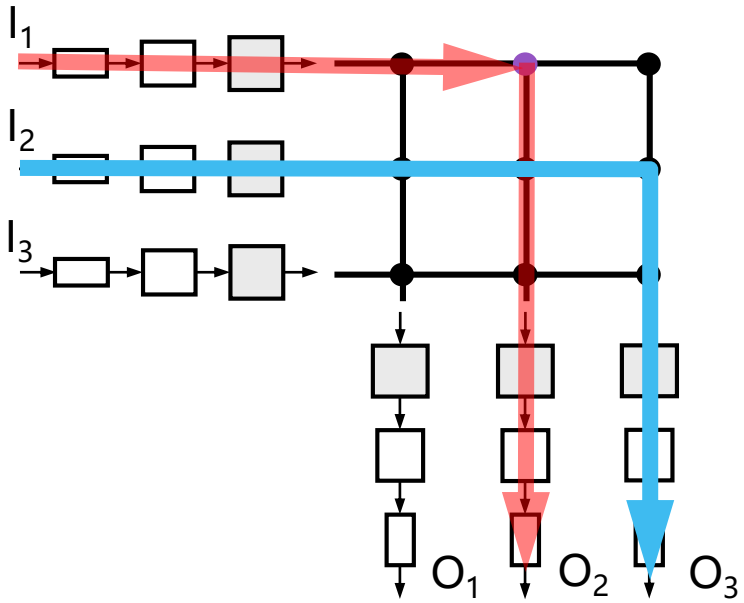


(b) 通过总线

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。
- 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。

6. 路由选择协议



(c) 通过互连网络

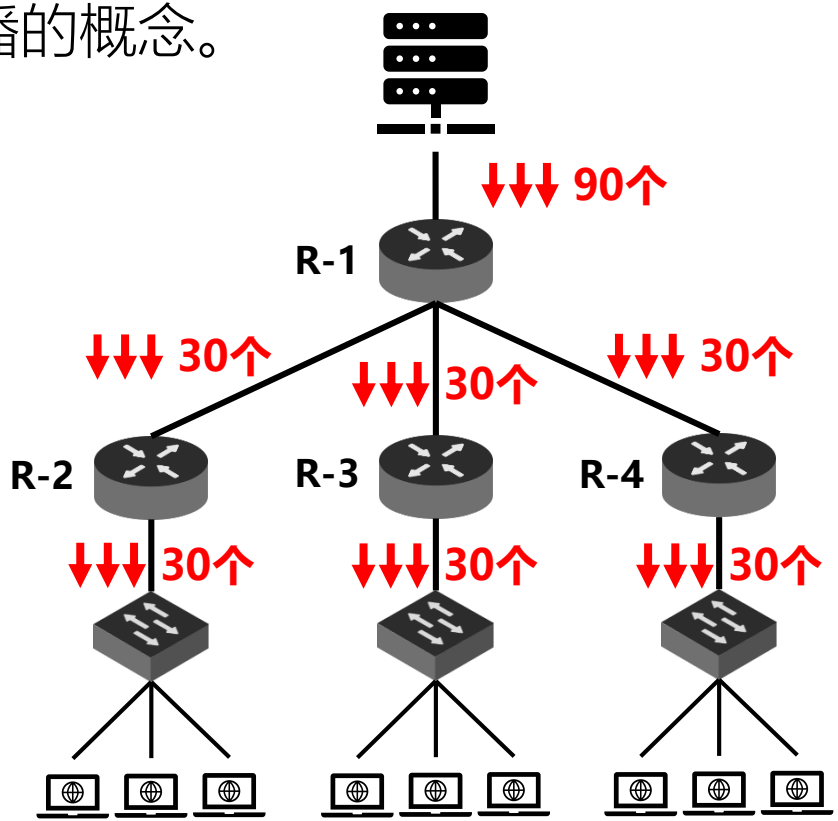
通过纵横交换结构 (crossbar switch fabric)

- 常被称为互连网络 (interconnection network)。
- 它有 2N 条总线，控制交叉节点可以使 N 个输入端口和 N 个输出端口相连接。
- 当输入端口收到一个分组时，就将它发送到水平总线上。
- 若通向输出端口的垂直总线空闲，则将垂直总线与水平总线接通，把该分组转发到这个输出端口。若输出端口已被占用，分组在输入端口排队等待。
- 特点：是一种无阻塞的交换结构，分组可以转发到任何一个输出端口，只要这个输出端口没有被别的分组占用。

7. IP 多播

- 1988 年，Steve Deering 首次提出 IP 多播的概念。
  - 多播 (multicast)：以前曾译为组播。
  - 目的：更好地支持一对多通信。
  - 一对多通信：一个源点发送到许多个终点。

采用单播方式
向 90 台主机传送
同样的视频节目
需要发送 90 个单播



7. IP 多播

- 1988 年，Steve Deering 首次提出 IP 多播的概念。
  - 多播 (multicast)：以前曾译为组播。
  - 目的：更好地支持一对多通信。
  - 一对多通信：一个源点发送到许多个终点。

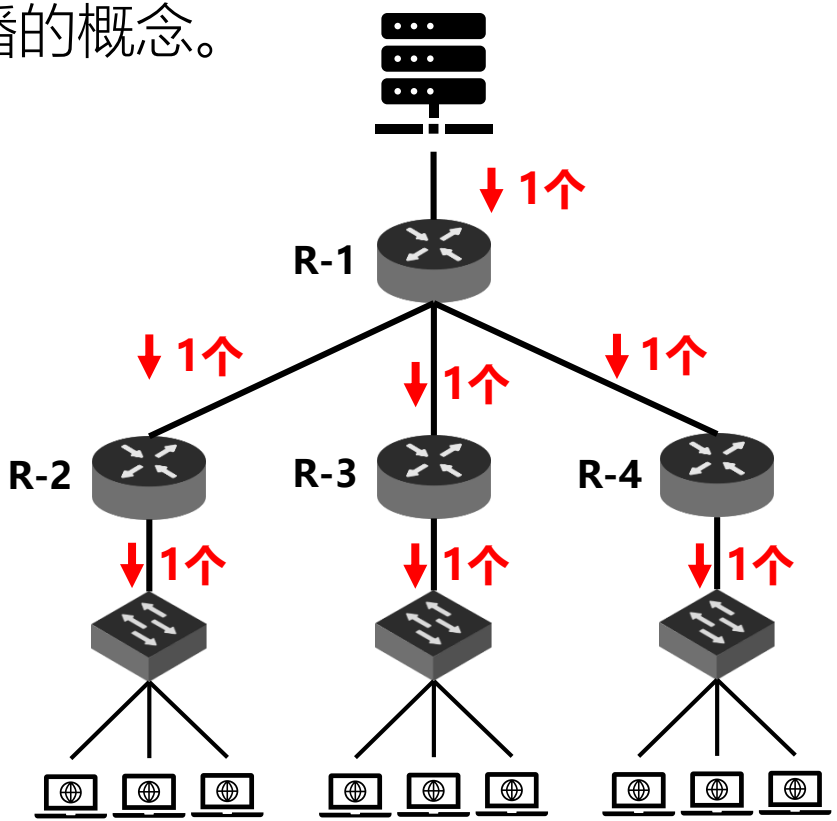
采用多播方式

只需发送一次到多播组。

路由器复制分组。

局域网具有硬件多播功能，不需要复制分组。

当多播组的主机数很大时（如成千上万个），采用多播方式可明显地减轻网络中资源消耗。



7. IP 多播

- IP 多播
  - 在互联网上进行多播就叫做 IP 多播。
  - 互联网范围的多播要靠路由器来实现。
  - 能够运行多播协议的路由器称为多播路由器 (multicast router)。
  - 多播路由器也可以转发普通的单播 IP 数据报。
  - 从 1992 年起，在互联网上开始试验虚拟的多播主干网 MBONE (Multicast Backbone On the InterNEt)。

7. IP 多播

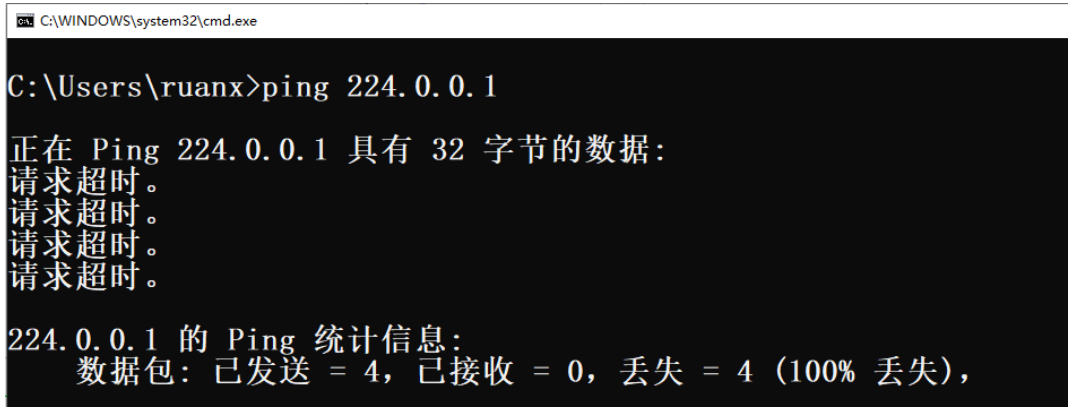
- 多播使用的 IP 地址
  - 在 IP 多播数据报的目的地址需要写入多播组的标识符。
  - 多播组的标识符就是 IP 地址中的 D 类地址（多播地址）。

地址范围：224.0.0.0 ~ 239.255.255.255

  - 每一个 D 类地址标志一个多播组。
    - 多播地址只能用于目的地址，不能用于源地址。

7. IP 多播

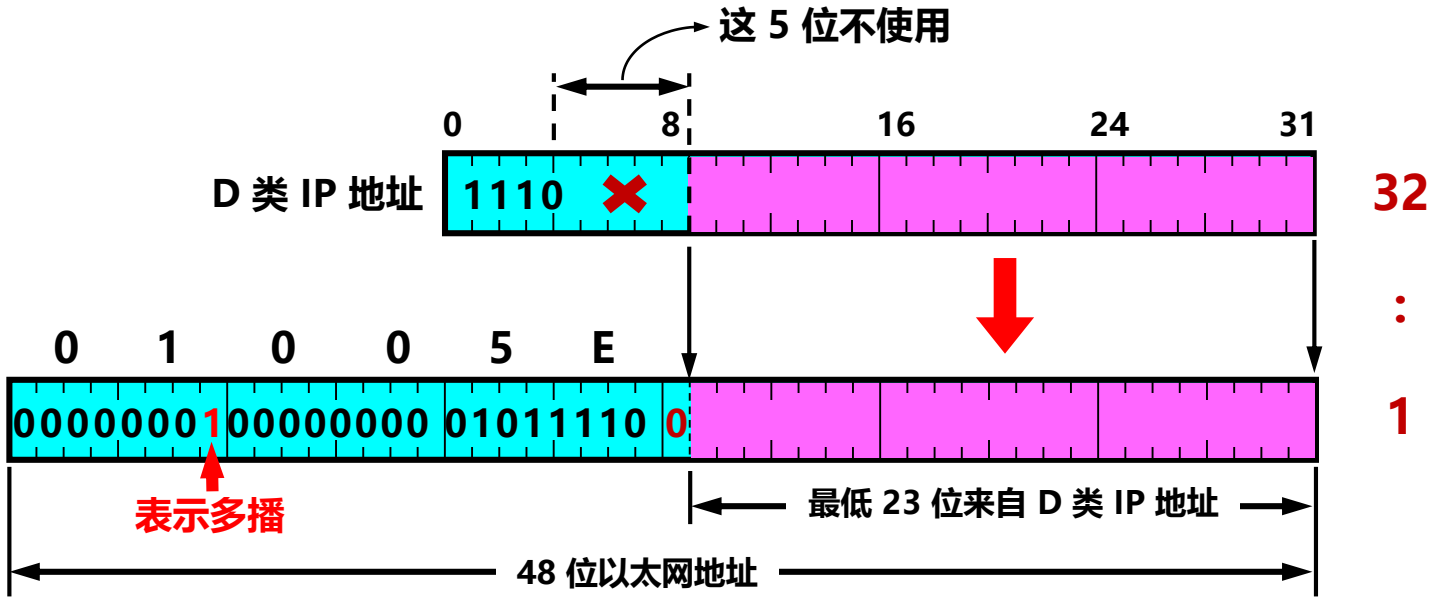
- 多播数据报
  - 多播数据报和一般的 IP 数据报的区别：
    - 目的地址：使用 D 类 IP 地址。
    - 协议字段 = 2，表明使用网际组管理协议 IGMP。
  - 尽最大努力交付，不保证一定能够交付多播组内的所有成员。
  - 对多播数据报不产生 ICMP 差错报文。
    - 在 PING 命令后面键入多播地址，
    - 将永远不会收到响应。



7. IP 多播

- IANA 拥有的以太网地址块的高 24 位为 00-00-5E。
- TCP/IP 协议使用的以太网地址块的范围是：
- 从　00-00-5E-00-00-00
- 到　00-00-5E-FF-FF-FF
- IANA 只拿出 01-00-5E-00-00-00 到 01-00-5E-7F-FF-FF （223 个地址）作为以太网多播地址。
  - 或者说，在 48 位的多播地址中，前 25 位都固定不变，只有后 23 位可用作多播。

7. IP 多播

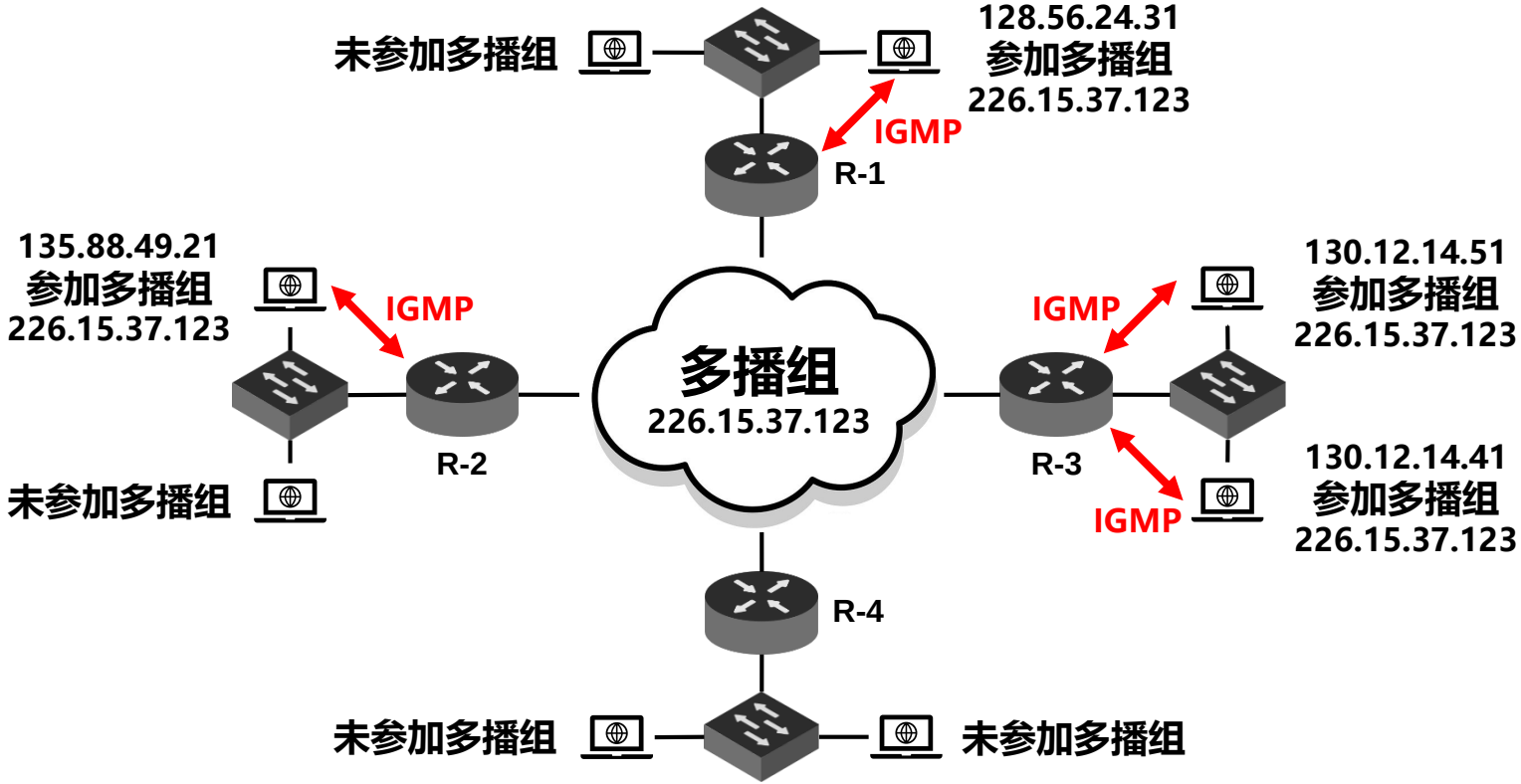


收到多播数据报的主机，还要在 IP 层对 IP 地址进行过滤，
把不是本主机要接收的数据报丢弃。

7. IP 多播

- IP 多播需要两种协议
  - 网际组管理协议 IGMP (Internet Group Management Protocol)
    - IGMP 使多播路由器知道多播组成员信息（有无成员）
      - IGMP 不知道 IP 多播组包含的成员数，也不知道这些成员都分布在哪些网络上。
      - IGMP 协议是让连接在本地局域网上的多播路由器知道本局域网上是否有主机参加或退出了某个多播组。
  - 多播路由选择协议
    - 使多播路由器协同工作，把多播数据报用最小代价传送给多播组的所有成员。

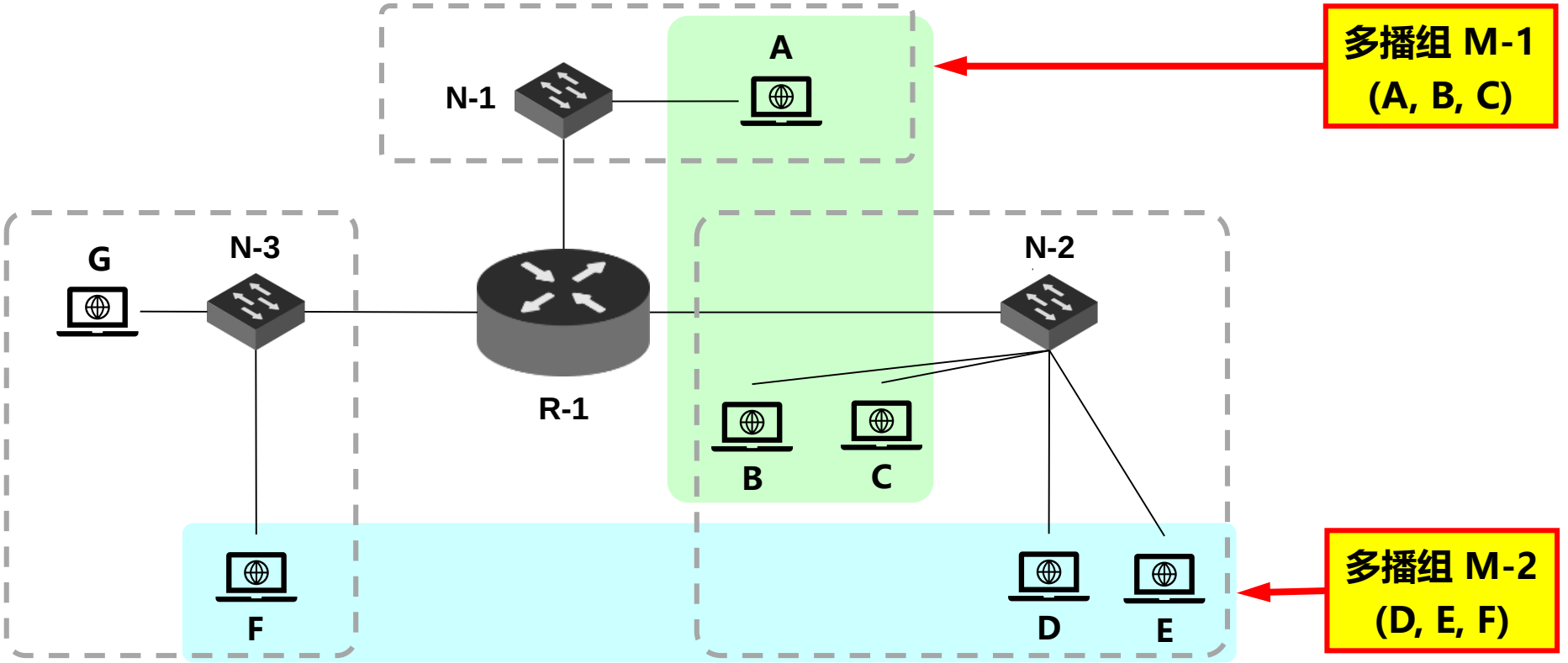
7. IP 多播


未参加多播组
128.56.24.31
参加多播组
226.15.37.123
IGMP
R-1
135.88.49.21
参加多播组
226.15.37.123
IGMP
130.12.14.51
参加多播组
226.15.37.123
IGMP
多播组
226.15.37.123
R-2
R-3
未参加多播组
IGMP
130.12.14.41
参加多播组
226.15.37.123
R-4
未参加多播组
未参加多播组

7. IP 多播

- IP 多播需要两种协议
  - 网际组管理协议 IGMP (Internet Group Management Protocol)
  - **多播路由选择协议**
    - 多播路由选择协议更为复杂。
    - 多播转发必须动态地适应多播组成员的变化（这时网络拓扑并未发生变化），因为每一台主机可以随时加入或离开一个多播组。
    - 多播路由器在转发多播数据报时，不能仅仅根据多播数据报中的目的地址，还要考虑这个多播数据报从什么地方来和要到什么地方去。
    - 多播数据报可以由没有加入多播组的主机发出，也可以通过没有组成员的接入网络。

7. IP 多播



路由器 R 不应当向网络 N-3 转发多播组 M-1 的分组，因为网络 N-3 上没有多播组 M-1 的成员。

7. IP 多播

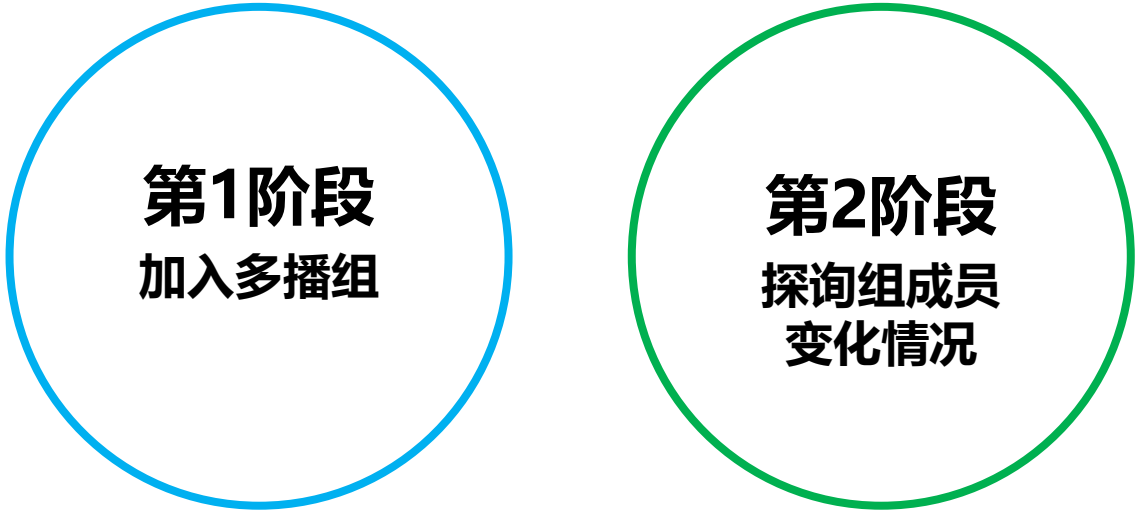
- 网际组管理协议 IGMP
  - 1989 年公布的 RFC 1112（IGMPv1）已成为了互联网的标准协议。
  - 1997 年公布的 RFC 2236（IGMPv2，建议标准）对 IGMPv1 进行了更新。
  - 2002 年 10 月公布了 RFC 3376（IGMPv3，建议标准）。

7. IP 多播

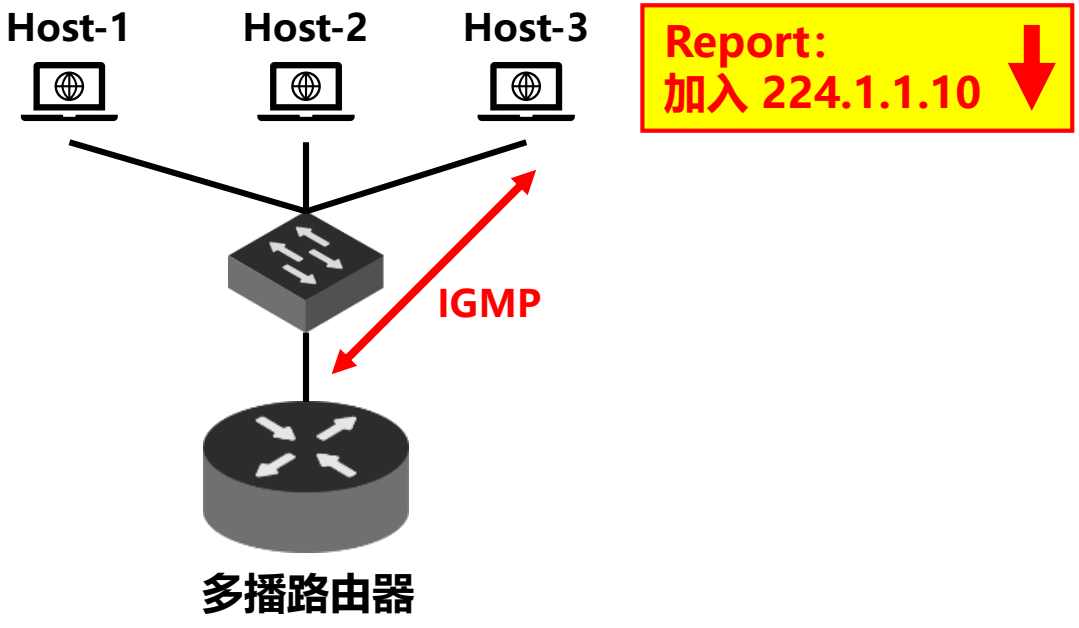
- 网际组管理协议 IGMP
  - IGMP 使用 IP 数据报传递其报文。
  - 在 IGMP 报文加上 IP 首部构成 IP 数据报。
  - 但 IGMP 也向 IP 提供服务。
  - 因此：
    - 不把 IGMP 看成是一个单独的协议，而是整个网际协议 IP 的一个组成部分。

7. IP 多播

- IGMP 工作可分为两个阶段

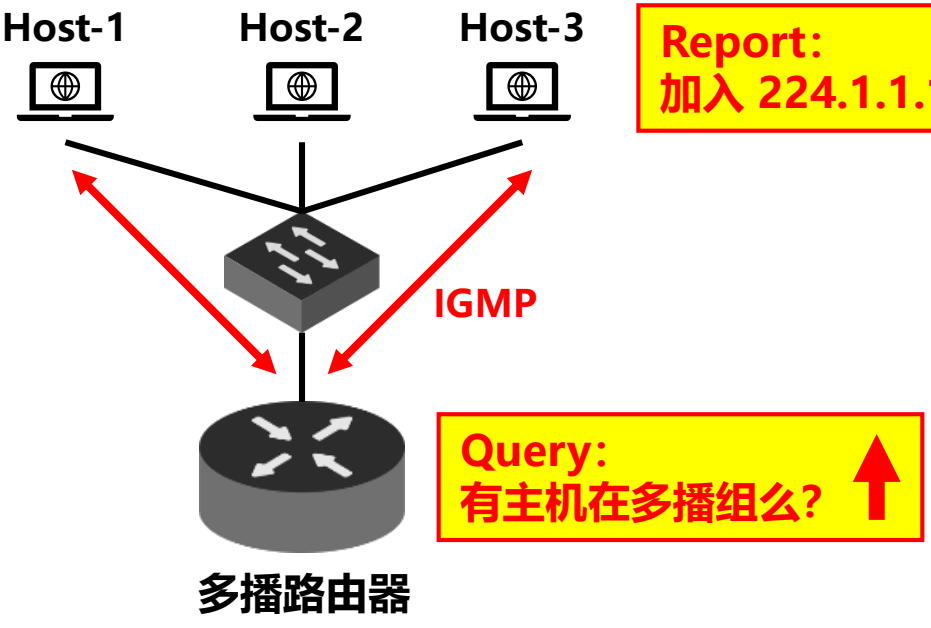


第1阶段：加入多播组



1. 当某个主机加入多播组时，该主机向多播组的多播地址发送 IGMP 报文，声明自己要成为该组的成员。
2. 本地的多播路由器收到 IGMP 报文后，将组成员关系转发给互联网上的其他多播路由器。

第2阶段：探询组成员变化情况



1. 本地多播路由器周期性地探询本地局域网上的主机，以便知道这些主机是否还继续是组的成员。
2. 只要对某个组有一个主机响应，那么多播路由器就认为这个组是活跃的。
3. 但一个组在经过几次的探询后仍然没有一个主机响应，则不再将该组的成员关系转发给其他的多播路由器。

7. IP 多播

- IGMP 采用的一些具体措施，以避免增加大量开销
  - 所有通信都使用 IP 多播。
    - 只要有可能，都用硬件多播来传送。
  - 对所有的组只发送一个请求信息的询问报文。
    - 默认询问速率是每 125 秒发送一次。
  - 当同一个网络上连接有多个多播路由器时，能迅速和有效地选择其中的一个来探询主机的成员关系。

7. IP 多播

- IGMP 采用的一些具体措施，以避免增加大量开销
  - 分散响应。
    - 在 IGMP 的询问报文中有一个数值 N，它指明一个最长响应时间（默认值为 10 秒）。
    - 当收到询问时，主机在 0 到 N 之间随机选择发送响应所需经过的时延。若一台主机同时参加了几个多播组，则主机对每一个多播组选择不同的随机数。
    - 对应于最小时延的响应最先发送。
  - 采用抑制机制。
    - 同一个组内的每一个主机都要监听响应，只要有本组的其他主机先发送了响应，自己就不再发送响应了。

7. IP 多播

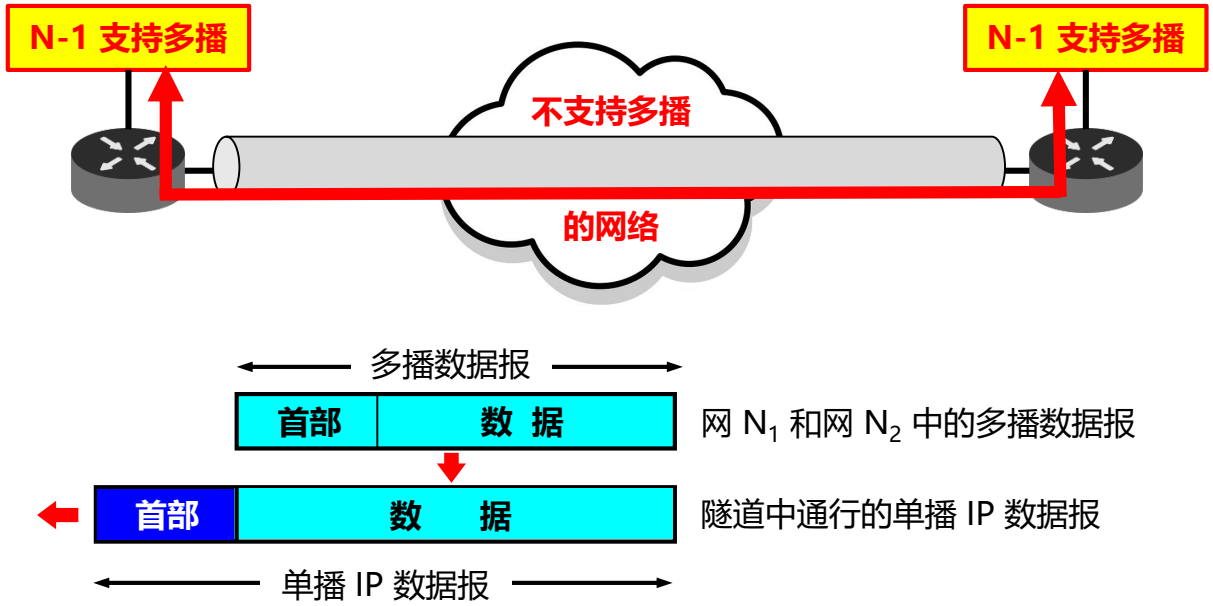
- 多播路由选择
  - IP 多播协议已成为建议标准，但多播路由选择协议尚未标准化。
  - 在多播过程中，一个多播组中的成员是动态变化的。
  - 多播路由选择实际上就是要找出以源主机为根节点的多播转发树。
  - 不同的多播组对应于不同的多播转发树。
  - 同一个多播组，对不同的源点也会有不同的多播转发树。

7. IP 多播

- 多播路由选择在转发多播数据报时使用三种方法：
  - 洪泛与剪除
    - 适合于较小的多播组，所有组成员接入的局域网也是相邻接的。
    - 开始时，路由器转发多播数据报使用洪泛的方法（这就是广播）。
    - 为避免兜圈子，采用反向路径广播 RPB (Reverse Path Broadcasting) 的策略。
      - RPB 的要点：形成以源为根节点的多播转发树。如果存在几条同样长度的最短路径，选择 IP 地址最小的。 最后就得出了以源为根节点的、用来转发多播数据报的多播转发树。
      - 剪枝：如果在多播转发树上的某个路由器发现它的下游树枝（即叶节点方向）已没有该多播组的成员，就把它和下游的树枝一起剪除。
      - 嫁接：当某个树枝有新增加的组成员时，可以再接入到多播转发树上。

7. IP 多播

- 多播路由选择在转发多播数据报时使用三种方法：
  - 隧道技术 (tunneling)
    - 隧道技术适用于多播组的位置在地理上很分散的情况。



7. IP 多播

- 多播路由选择在转发多播数据报时使用三种方法：
  - **基于核心的发现技术**
    - 对于多播组的大小在较大范围内变化时都适合。
    - 对每一个多播组 G 指定一个核心 (core) 路由器，并给出它的 IP 单播地址。
    - 核心路由器按照前面讲过的 2 种方法创建出对应于多播组 G 的转发树（核心路由器为根节点）。
      - 为一个多播组构建一棵转发树，而不是为每个（源，组）组合构建一棵转发树。
      - 构建转发树开销较小，扩展性较好。

7. IP 多播

- 多播路由选择在转发多播数据报时使用三种方法：
  - **基于核心的发现技术**
    - 如果有一个路由器 R1 向核心路由器发送数据报，那么它在途中经过的每一个路由器都要检查其内容。
    - 当数据报到达参加了多播组 G 的路由器 R2 时，R2 就处理这个数据报。
      - 如果 R1 发出的是一个多播数据报，其目的地址是 G 的组地址，R2 就向 G 的成员转发这个多播数据报。
      - 如果 R1 发出的数据报是一个请求加入多播组 G 的数据报，R2 就把这个信息加到它的路由中，并用隧道技术向 R1 转发每一个多播数据报的副本。

7. IP 多播

- 多播路由选择在转发多播数据报时使用三种方法：
  - 基于核心的发现技术
    - 距离向量多播路由选择协议 DVMRP (Distance Vector Multicast Routing Protocol)
      - 互联网上使用的第一个多播路由选择协议。
    - 基于核心的转发树 CBT (Core Based Tree)
    - 开放最短通路优先的多播扩展 MOSPF (Multicast Extensions to OSPF)
    - 协议无关多播-稀疏方式 PIM-SM (Protocol Independent Multicast-Sparse Mode) 。
      - 唯一成为互联网标准的一个协议。
    - 协议无关多播-密集方式 PIM-DM (Protocol Independent Multicast-Dense Mode)

8. 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。

8. 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。

8. 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

8. 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)。

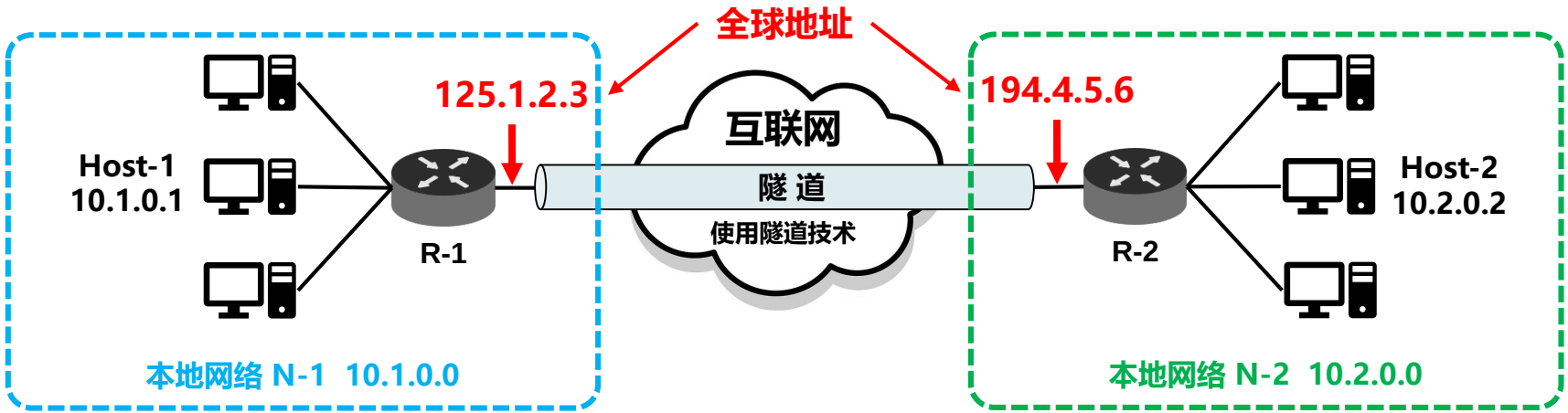
8. VPN 和 NAT

- 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。

8. VPN 和 NAT

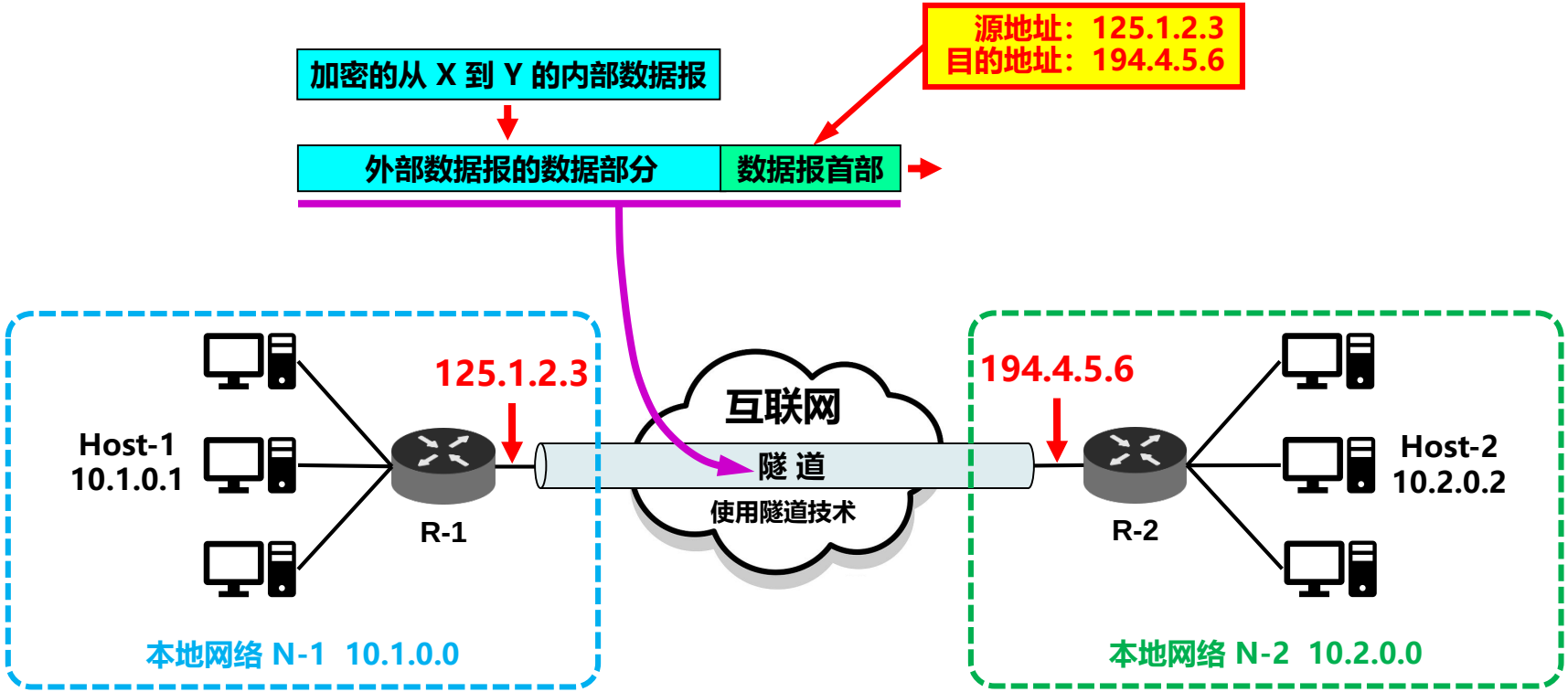
- 虚拟专用网 VPN 的构建方法
  - 如果专用网不同网点之间的通信必须经过公用的互联网，但又有保密的要求，那么所有通过互联网传送的数据都必须加密。
  - 一个机构要构建自己的 VPN 就必须为它的每一个场所购买专门的硬件和软件，并进行配置，使每一个场所的 VPN 系统都知道其他场所的地址。

- 虚拟专用网 VPN 的实现：基于隧道技术



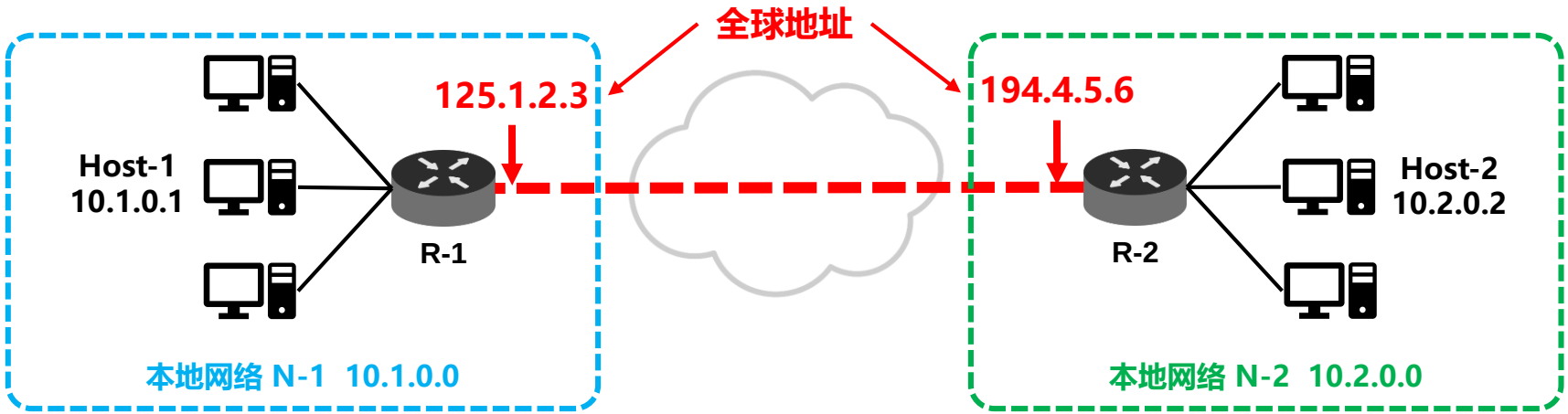
8. VPN 和 NAT

- 虚拟专用网 VPN 的实现：基于隧道技术



8. 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)**。



8. 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。

8. VPN 和 NAT

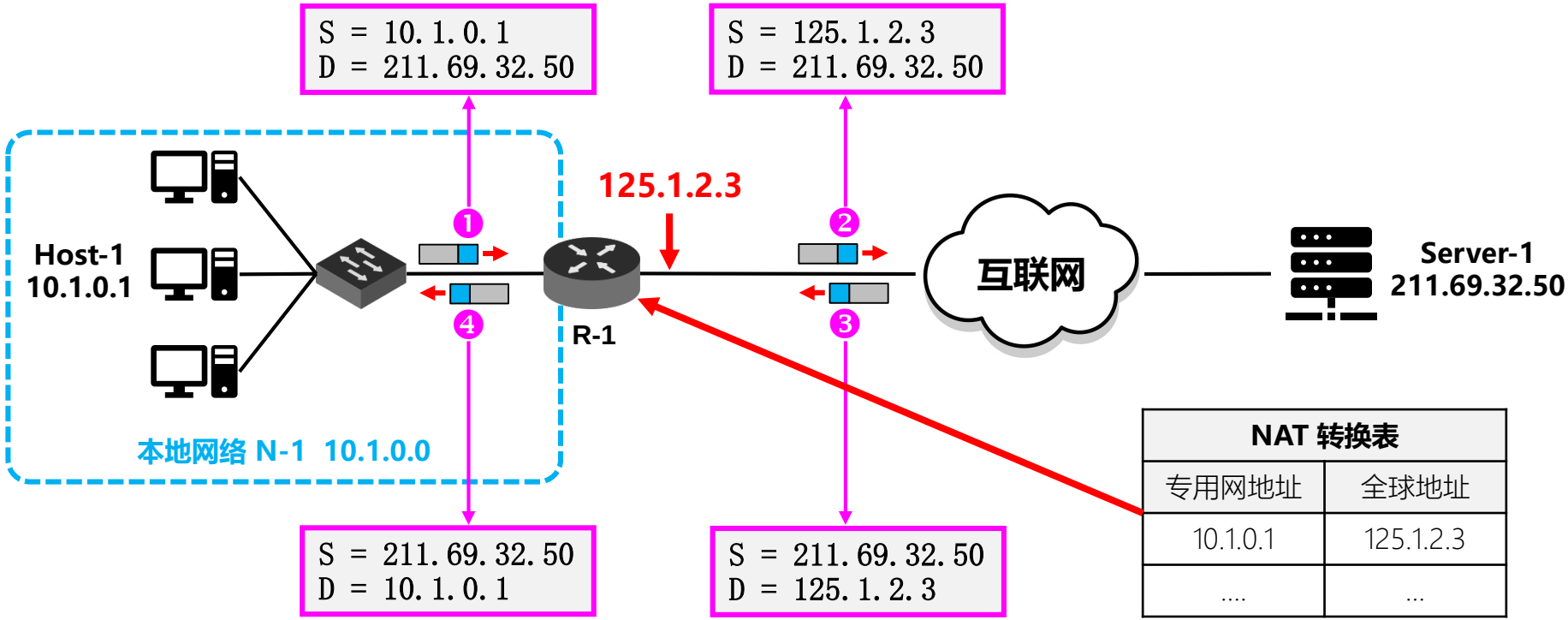
- 远程接入 VPN
  - 对于个人“翻墙”行为的规定早已明确。
  - 1996年1月23日的国务院常务会议通过《计算机信息网络国际联网管理暂行规定》，该规定在1997年进行修正。
  - 该《规定》第六条提出，计算机信息网络直接进行国际联网，必须使用邮电部国家公用电信网提供的国际出入口信道。该规定明确要求，**任何单位和个人不得自行建立或者使用其他信道进行国际联网**。如果违反第六条规定，由公安机关责令停止联网，给予警告，可以并处15000元以下的罚款;有违法所得的，没收违法所得。

8. 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。

8. VPN 和 NAT

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



8. 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。

8. 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，在 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（发往互联网）	源 IP 地址	10.1.0.1	172.38.1.5
入（进入专用网）	目的 IP 地址	172.38.1.5	10.1.0.1

8. 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

8. VPN 和 NAT

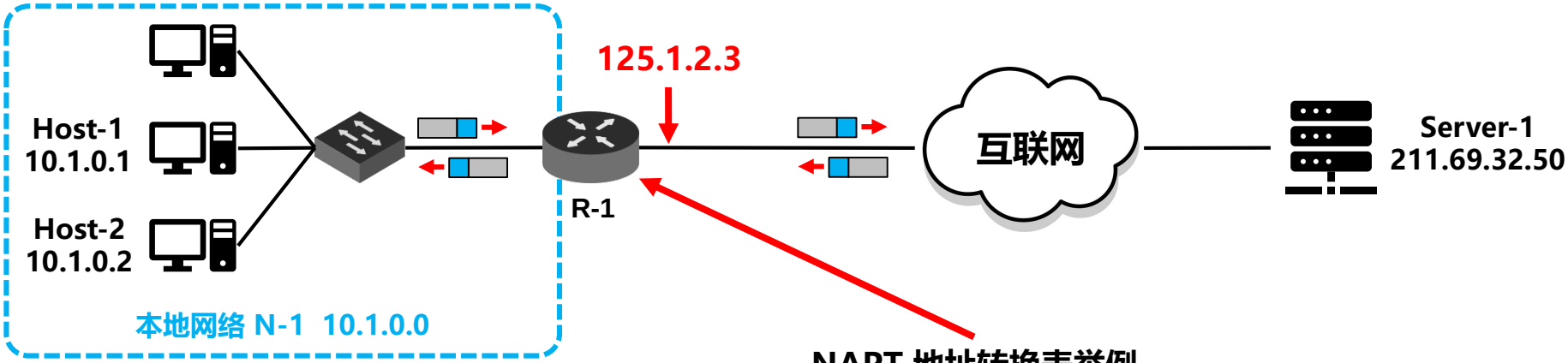
- 网络地址与端口号转换 NAPT
  - NAT 并不能节省 IP 地址。
    - 为了更加有效地利用 NAT 路由器上的全球 IP 地址，现在常用的 NAT 转换表把运输层的端口号也利用上。
    - 这样就可以使多个拥有本地地址的主机，共用一个 NAT 路由器上的全球 IP 地址，因而可以同时和互联网上的不同主机进行通信。
  - 使用端口号的 NAT 叫做**网络地址与端口号转换NAPT** (Network Address and Port Translation)。
    - NAPT 可以使多台拥有本地地址的主机，共用一个 全球 IP 地址，同时和互联网上的不同主机进行通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

8. 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。

8. VPN 和 NAT

- 网络地址与端口号转换 NAPT

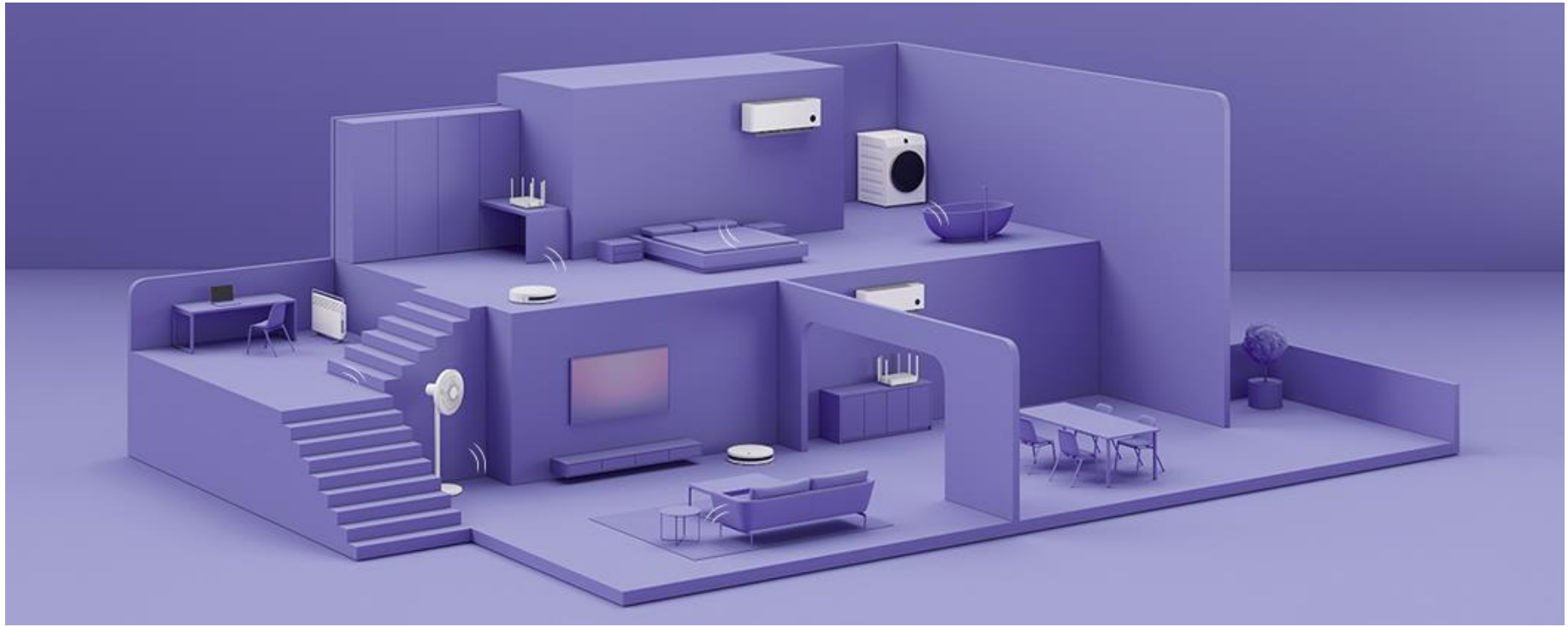


NAPT 地址转换表举例

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	10.1.0.1:30000	125.1.2.3:40001
出	源IP地址:TCP源端口	10.1.0.2:30000	125.1.2.3:40002
入	目的IP地址:TCP目的端口	125.1.2.3:40001	10.1.0.1:30000
入	目的IP地址:TCP目的端口	125.1.2.3:40002	10.1.0.2:30000

8. VPN 和 NAT

- 网络地址与端口号转换 NAPT



智能运维课程体系


棒棒堂智能运维
UID: 510401430
抖音
@棒棒堂智能运维
抖音号：ID2003079563
只要思想不滑坡，办法总比困难多。
保存图片
到相册
打开抖音
搜索页扫一扫