

Chapter 4 Network layer

- [1. 网络层概述](#)
 - [1.1. 分组转发和路由选择](#)
 - [1.2. 网络层向其上层提供的两种服务](#)
 - [1.2.1. 面向连接的虚电路服务](#)
 - [1.2.2. 无连接的数据报服务](#)
- [2. 因特网协议IP](#)
 - [2.1. 异构网络互连](#)
 - [2.2. IPv4地址及其编址方法](#)
 - [2.2.1. 概述](#)
 - [2.2.2. 表示方法：点分十进制](#)
 - [2.2.3. 分类编址方法](#)
 - [2.2.4. 划分子网的编址方法](#)
 - [2.2.5. 无分类编址方法](#)
 - [2.3. IPv4地址的应用规划](#)
 - [2.3.1. 定长子网掩码](#)
 - [2.3.2. 变长子网掩码](#)
 - [2.4. IPv4地址与MAC地址](#)
 - [2.4.1. IPv4地址与MAC地址的封装位置](#)
 - [2.4.2. 数据包传送过程中IPv4地址与MAC地址的变化情况](#)
 - [2.4.3. IPv4地址与MAC地址的关系](#)
 - [2.5. 地址解析协议ARP](#)
 - [2.6. IP数据报的发送和转发流程](#)
 - [2.6.1. 主机发送IP数据报](#)
 - [2.6.2. 路由器转发IP数据报](#)
 - [2.7. IPv4数据报的首部格式](#)
 - [2.7.1. 版本](#)
 - [2.7.2. 首部长度](#)
 - [2.7.3. 可选字段](#)
 - [2.7.4. 填充](#)
 - [2.7.5. 区分服务](#)
 - [2.7.6. 总长度](#)
 - [2.7.7. IP数据包分片](#)
 - [2.7.7.1. 标识](#)
 - [2.7.7.2. 标志](#)
 - [2.7.7.3. 片偏移](#)
 - [2.7.8. 生存时间 \(Time To Live, TTL\)](#)
 - [2.7.9. 协议](#)
 - [2.7.10. 首部检验和](#)
 - [2.7.10.1. 首部检验和的计算方法](#)
 - [2.7.11. 源IP地址](#)
 - [2.7.12. 目的IP地址](#)
- [3. 静态路由配置](#)
 - [3.1. 人工配置静态路由](#)
 - [3.2. 默认路由0.0.0.0/0](#)
 - [3.3. 特定主机路由/32](#)
 - [3.4. 需注意的问题](#)
- [4. 因特网的路由选择协议](#)
 - [4.1. 路由选择分类](#)
 - [4.1.1. 静态路由选择](#)
 - [4.1.2. 动态路由选择](#)
 - [4.2. 因特网采用分层次的路由选择协议](#)
 - [4.2.1. 特点](#)
 - [4.2.2. 举例](#)
 - [4.3. 路由信息协议RIP \(封装在UDP\)](#)
 - [4.3.1. 基本概念](#)
 - [4.3.2. 基本工作过程](#)
 - [4.3.3. 距离向量算法](#)

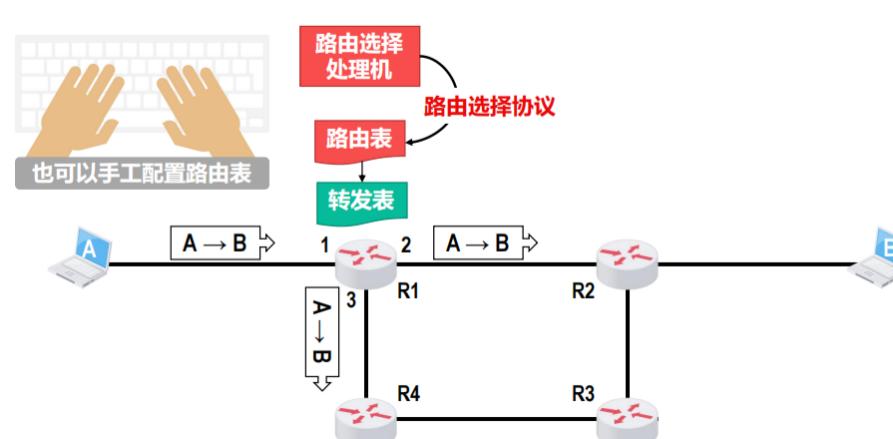
- 4.3.4. 存在的问题——“坏消息传播得慢”
- 4.3.5. 版本和相关报文的封装
- 4.3.6. 优缺点
- 4.4. 开放最短路径优先OSPF (封装在IP)
 - 4.4.1. 基本概念
 - 4.4.2. 五种分组类型
 - 4.4.3. 基本工作过程
 - 4.4.4. 多点接入网络中的OSPF路由器
 - 4.4.5. OSPF划分区域
- 4.5. 边界网关协议BGP (封装在TCP)
 - 4.5.1. 基本概念
 - 4.5.2. BGP-4的四种报文
- 4.6. 路由器的基本工作原理
- 5. 网际控制协议ICMP (封装在IP)
 - 5.1. 概述
 - 5.2. ICMP报文种类
 - 5.2.1. 差错报告报文
 - 5.2.2. 询问报文
 - 5.3. ICMP的典型应用
 - 5.3.1. 分组网间探测 (Packet InterNet Groper, PING)
 - 5.3.2. 跟踪路由 (traceroute)
- 6. 虚拟专用网VPN和网络地址转换NAT
 - 6.1. 虚拟专用网VPN
 - 6.2. 网络地址转换NAT
 - 6.2.1. 最基本的NET方法
 - 6.2.2. 网络地址与端口号转换方法 (NAPT)
- 7. IP多播技术
 - 7.1. 相关基本概念
 - 7.2. IP多播地址和多播组
 - 7.3. 在局域网上进行硬件多播
 - 7.4. 在因特网上进行IP多播需要的两种协议
 - 7.5. 网际组管理协议IGMP (封装在IP)
 - 7.5.1. 三种报文类型
 - 7.5.2. 基本工作原理
 - 7.5.2.1. 加入多播组
 - 7.5.2.2. 监视多播组的成员变化
 - 7.5.2.3. 退出多播组
 - 7.6. 多播路由选择协议 (封装在IP)
 - 7.6.1. 多播路由选择协议
 - 7.6.1.1. 基于源树多播路由选择
 - 7.6.1.2. 组共享树多播路由选择
 - 7.6.2. 因特网的多播路由选择协议
- 8. 移动IP技术概述
 - 8.1. 移动性对因特网应用的影响
 - 8.2. 移动IP技术的相关基本概念
 - 8.3. 移动IP技术的基本工作原理
 - 8.3.1. 代理发现与注册
 - 8.3.2. 固定主机向移动主机发送IP数据报
 - 8.3.3. 移动主机向固定主机发送IP数据报
 - 8.3.4. 同址转交地址方式
 - 8.3.5. 三角形路由问题
- 9. IPv6
 - 9.1. IPv6的诞生背景
 - 9.2. IPv6引进的主要变化
 - 9.3. IPv6数据包的基本首部
 - 9.4. IPv6数据包的扩展首部
 - 9.5. IPv6地址
 - 9.5.1. IPv6地址空间大小
 - 9.5.2. IPv6地址的表示方法

- [9.5.3. IPv6地址的分类](#)
- [9.6. 从IPv4向IPv6过渡](#)
 - [9.6.1. 使用双协议栈](#)
 - [9.6.2. 使用隧道技术](#)
- [9.7. 网际控制报文协议ICMPv6](#)
 - [9.7.1. 概述](#)
 - [9.7.2. ICMPv6报文的封装](#)
 - [9.7.3. ICMPv6报文的分类](#)
- [10. 软件定义网络SDN](#)
 - [10.1. 概述](#)
 - [10.2. 网络层的数据层面和控制层面](#)
 - [10.3. OpenFlow协议](#)
 - [10.3.1. 概述](#)
 - [10.3.2. 传统意义上的数据层面的任务](#)
 - [10.3.3. SDN中的广义转发](#)
 - [10.3.4. OpenFlow交换机和流表](#)
 - [10.4. SDN体系结构](#)
 - [10.4.1. SDN体系结构及其四个关键特征](#)
 - [10.4.2. SDN控制器](#)
 - [10.5. 总结](#)
- [11. 题目](#)
 - [11.1. IPv4分类编址方法](#)
 - [11.1.1. 【2017 36】](#)
 - [11.2. Ipv4划分子网编址方法](#)
 - [11.2.1. 【2012 39】](#)
 - [11.3. 无分类编址方法](#)
 - [11.3.1. 【2011 38】](#)
 - [11.3.2. 【2018 38】](#)
 - [11.3.3. 【2021 35】](#)
 - [11.4. IPv4地址的应用规划](#)
 - [11.4.1. 【2019 37】](#)
 - [11.5. 数据报传送过程中IPv4地址与MAC地址的变化情况](#)
 - [11.6. IP数据报的发送和转发过程](#)
 - [11.6.1. 【2019 47】](#)
 - [11.7. 片偏移](#)
 - [11.7.1. 【2020 36】](#)
 - [11.8. RIP](#)
 - [11.8.1. 【2010 35】](#)
 - [11.8.2. 【2021 37】](#)
 - [11.9. OSPF](#)
 - [11.9.1. 【2014 43改】](#)
 - [11.10. BGP](#)
 - [11.10.1. 【2013 46\(3\)】](#)
 - [11.10.2. 【2017 37】](#)
 - [11.11. ICMP](#)
 - [11.11.1. 【2010 36】](#)
 - [11.12. SDN题目](#)

1. 网络层概述

1.1. 分组转发和路由选择

- 网络层的主要任务就是将分组从源主机经过多个网络和多段链路传输到目的主机，可以将该任务划分为组转发和路由选择两种重要的功能。

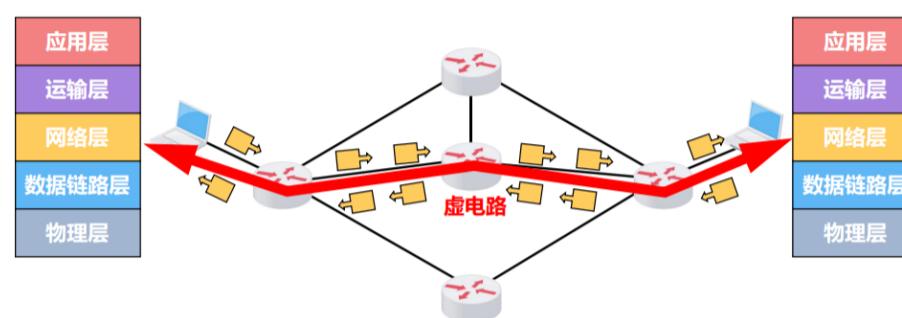


1.2. 网络层向其上层提供的两种服务

对比方面	虚电路服务	数据报服务
核心思想	可靠通信应当由网络自身来保证	可靠通信应当由用户主机来保证
连接	必须建立网络层连接	不需要建立网络层连接
目的地址	仅在连接建立阶段使用，之后每个分组使用短的虚电路号	每个分组都必须携带完整的目的地址
分组转发	属于同一条虚电路的分组均按同一路由进行转发	每个分组可走不同的路由
节点故障	所有通过出故障的节点的虚电路均不能工作	出故障的节点可能会丢失分组，一些路由可能会发生变化
分组顺序	总是按发送顺序到达目的主机	到达目的主机时不一定按发送顺序
服务质量	可以将通信资源提前分配给每一个虚电路，因此容易实现	很难实现

1.2.1. 面向连接的虚电路服务

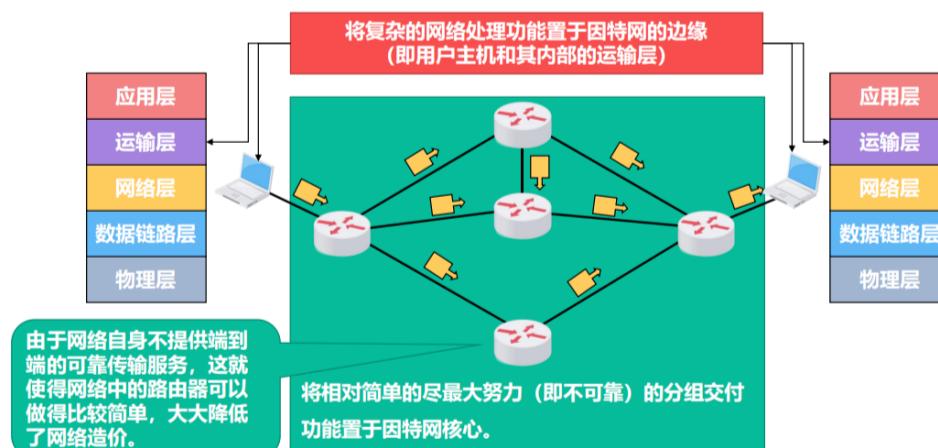
- 核心思想是“可靠通信应由网络自身来保证”。
- 必须首先建立网络层连接——虚电路（Virtual Circuit, VC），以保证通信双方所需的一切网络资源。
- 通信双方沿着已建立的虚电路发送分组。
- 通信结束后，需要释放之前所建立的虚电路。



- 这种通信方式如果再使用可靠传输的网络协议，就可使所发送的分组最终正确（无差错按序到达、不丢失、不重复）到达接收方。
- 很多广域分组交换网都使用面向连接的虚电路服务。例如，曾经的X.25和逐渐过时的帧中继（Frame Relay, FR）、异步传输模式（Asynchronous Transfer Mode, ATM）。
- 然而，因特网的先驱者并没有采用这种设计思想，而是采用了无连接的数据报服务。

1.2.2. 无连接的数据报服务

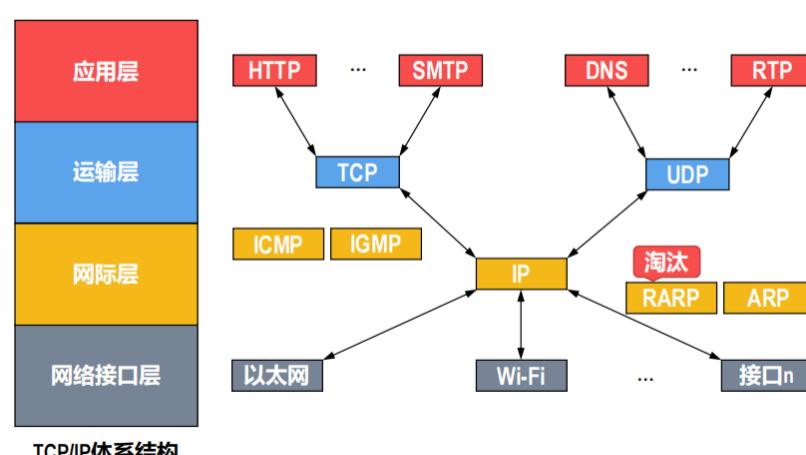
- 核心思想是“可靠通信应由用户主机来保证”。
- 不需要建立网络层连接。
- 每个分组可走不同的路径。因此，每个分组的首部都必须携带目的主机的完整地址。
- 通信结束后，没有需要释放的连接。



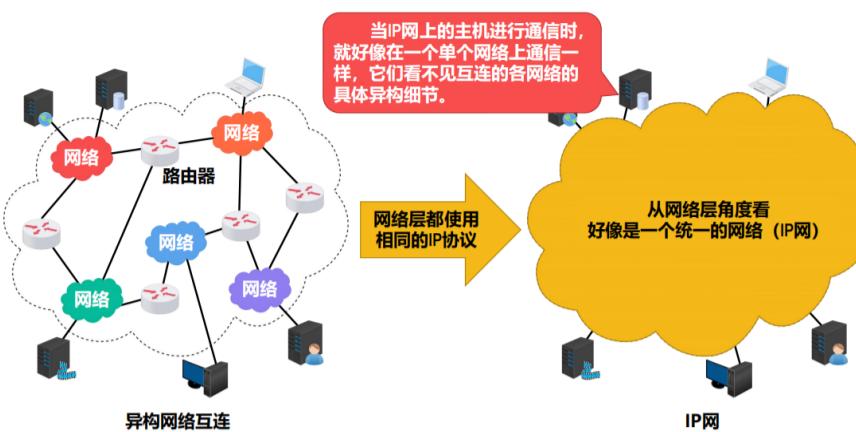
- 这种通信方式所传送的分组可能误码、丢失、重复和失序。

2. 网际协议IP

- 网际协议（Internet Protocol, IP）是TCP/IP体系结构网际层中的核心协议。

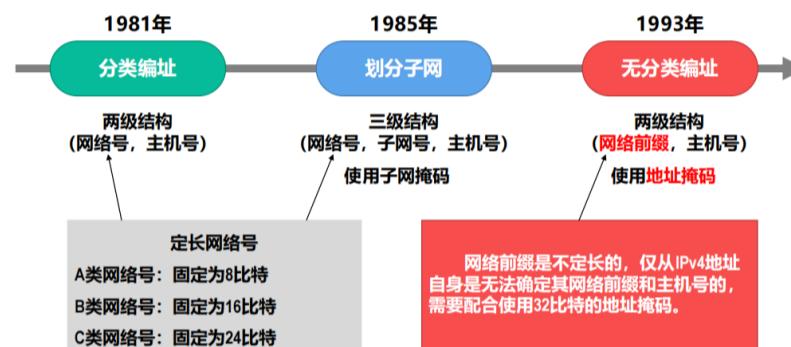


2.1. 异构网络互连



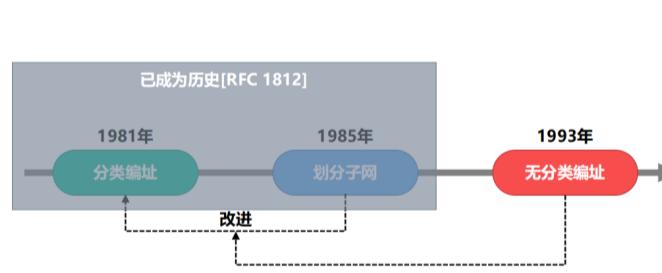
- 这些网络的拓扑、性能以及所使用的网络协议都不尽相同，这是由用户需求的多样性造成的，没有一种单一的网络能够适应所有用户的需求。
- 要将众多的异构型网络都互连起来，并且能够互相通信，则会面临许多需要解决的问题。
 - 不同的网络接入机制
 - 不同的差错恢复方法
 - 不同的路由选择技术
 - 不同的寻址方案
 - 不同的最大分组长度
 - 不同的服务（面向连接服务和无连接服务）

2.2. IPv4地址及其编址方法



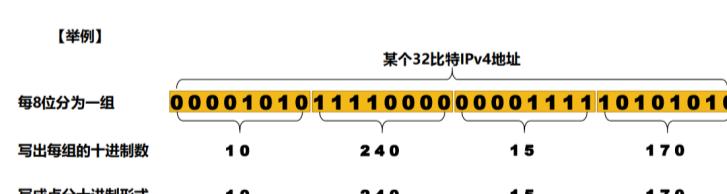
2.2.1. 概述

- IPv4地址是给因特网（Internet）上的每一个主机（或路由器）的每一个接口分配的一个在全世界范围内唯一的32比特的标识符。
- IPv4地址由因特网名字和数字分配机构（Internet Corporation for Assigned Names and Numbers, ICANN）进行分配。
 - 我国用户可向亚太网络信息中心（Asia Pacific Network Information Center, APNIC）申请IP地址，需要缴纳相应的费用，一般不接受个人申请。
 - 2011年2月3日，因特网号码分配管理局（Internet Assigned Numbers Authority, IANA）（由ICANN行使职能）宣布，IPv4地址已经分配完毕。
 - 我国在2014至2015年也逐步停止了向新用户和应用分配IPv4地址，同时全面开展商用部署IPv6。
- IPv4地址的编址方法经历了三个历史阶段

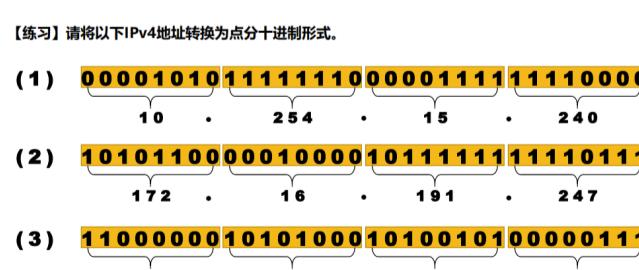


2.2.2. 表示方法：点分十进制

- 举例



- 练习



- 二进制转十进制

$$(b_7 b_6 b_5 b_4 b_3 b_2 b_1 b_0)_2 = (b_7 \times 2^7 + b_6 \times 2^6 + b_5 \times 2^5 + b_4 \times 2^4 + b_3 \times 2^3 + b_2 \times 2^2 + b_1 \times 2^1 + b_0 \times 2^0)_{10}$$

8位二进制数的每个位的权值: 128 64 32 16 8 4 2 1

【举例】

$$(10101010)_2 = (1 \times 2^7 + 0 \times 2^6 + 1 \times 2^5 + 0 \times 2^4 + 1 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 0 \times 2^0)_{10}$$

$$= (1 \times 128 + 0 \times 64 + 1 \times 32 + 0 \times 16 + 1 \times 8 + 0 \times 4 + 1 \times 2 + 0 \times 1)_{10}$$

$$= (170)_{10}$$

(11111100)₂ = (255 - 2 - 1)₁₀ = (252)₁₀ 比特0在整个8位无符号二进制整数中数量不多

(11110000)₂ = (255 - 8 - 4 - 2 - 1)₁₀ = (240)₁₀

(10000001)₂ = (128 + 1)₁₀ = (129)₁₀ 比特0在整个8位无符号二进制整数中数量较多

- 十进制转二进制

除2取余法 (逆序输出)		凑值法 (必须记住8位二进制数各位的权值 128 64 32 16 8 4 2 1)	
【举例】		【举例】	
(130) ₁₀ = (10000010) ₂		(171) ₁₀ = (10101011) ₂	
130 ÷ 2 = 65	余0	= (1 × 128 + 0 × 64 + 1 × 32 + 0 × 16 + 1 × 8 + 0 × 4 + 1 × 2 + 1 × 1) ₁₀	
65 ÷ 2 = 32	余1	↑ ↑ ↑ ↑ ↑ ↑ ↑ ↑	b ₇ b ₆ b ₅ b ₄ b ₃ b ₂ b ₁ b ₀
32 ÷ 2 = 16	余0	(b ₇ × 2 ⁷ + b ₆ × 2 ⁶ + b ₅ × 2 ⁵ + b ₄ × 2 ⁴ + b ₃ × 2 ³ + b ₂ × 2 ² + b ₁ × 2 ¹ + b ₀ × 2 ⁰) ₁₀	
16 ÷ 2 = 8	余0	↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓	128 64 32 16 8 4 2 1
8 ÷ 2 = 4	余0		
4 ÷ 2 = 2	余0		
2 ÷ 2 = 1	余0		
1 ÷ 2 = 0	余1		

2.2.3. 分类编制方法

- IPv4分类编址方法不够灵活、容易造成大量IPv4地址资源浪费。

32比特的IPv4地址	
网络号	主机号
● 标志主机 (或路由器) 的接口所连接到的网络	● 标志主机 (或路由器) 的接口
● 同一个网络中，不同主机 (或路由器) 的接口的IPv4地址的网络号必须相同，表示它们属于同一个网络。	● 同一个网络中，不同主机 (或路由器) 的接口的IPv4地址的主机号必须各不相同，以便区分各主机 (或路由器) 的接口。

■ A类、B类和C类地址都是单播地址，只有单播地址可以分配给网络中的主机 (或路由器) 的各接口。

■ 主机号为“全0”的地址是网络地址，不能分配给主机 (或路由器) 的各接口。

■ 主机号为“全1”的地址是广播地址，不能分配给主机 (或路由器) 的各接口。

			地址数量占比
A类地址	0 网络号	24位 主机号	$\frac{2^{(32-1)}}{2^{32}} = \frac{1}{2}$
B类地址	10 网络号	16位 主机号	$\frac{2^{(32-2)}}{2^{32}} = \frac{1}{4}$
C类地址	110 网络号	24位 主机号	$\frac{2^{(32-3)}}{2^{32}} = \frac{1}{8}$
D类地址	1110 多播地址		$\frac{2^{(32-4)}}{2^{32}} = \frac{1}{16}$
E类地址	1111 保留地址		$\frac{2^{(32-5)}}{2^{32}} = \frac{1}{32}$

网络类别	最小可指派网络号	最大可指派网络号	可指派网络数量	每个网络中最大可分配地址数量	不能指派的网络号	占总地址空间
A	1	126	$126 (2^{8-1} - 2)$	$16777214 (2^{24} - 2)$	0和127	50% ($2^{32-1}/2^{32}$)
B	128.0	191.255	$16384 (2^{16-2})$	$65534 (2^{16} - 2)$	无	25% ($2^{32-2}/2^{32}$)
C	192.0.0	223.255.255	$2097152 (2^{24-3})$	$254 (2^8 - 2)$	无	12.5% ($2^{32-3}/2^{32}$)

网络类别	作用	第一个地址	最后一个地址	地址数量	占总地址空间
D	多播地址	224.0.0.0	239.255.255.255	$268435456 (2^{28})$	6.25% ($2^{32-4}/2^{32}$)
E	保留	240.0.0.0	255.255.255.255	$268435456 (2^{28})$	6.25% ($2^{32-4}/2^{32}$)

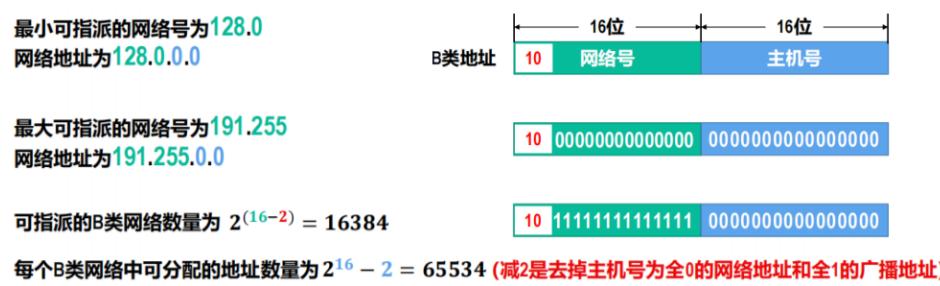
- A类细节

A类地址	0 网络号	24位 主机号
最小网络号为0，表示本网络，不能指派；	0 0000000	
最小可指派的网络号为1，网络地址为1.0.0.0	0 0000001	00000000000000000000000000000000
最大网络号为127，作为本地环回测试地址，不能指派	0 1111111	
最小的本地环回测试地址为127.0.0.1	0 1111111	00000000000000000000000000000001
最大的本地环回测试地址为127.255.255.254	0 1111111	11111111111111111111111111111110
最大可指派的网络号为126，网络地址为126.0.0.0	0 1111110	00000000000000000000000000000000

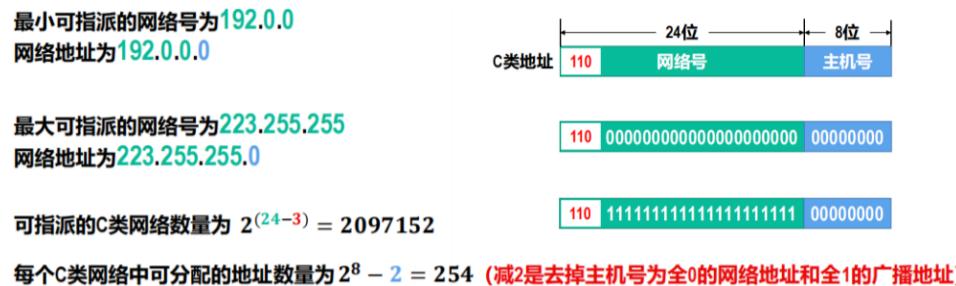
可指派的A类网络数量为 $2^{(8-1)} - 2 = 126$ (减2是去掉最小网络号0和最大网络号127)

每个A类网络中可分配的地址数量为 $2^{24} - 2 = 16777214$ (减2是去掉主机号为全0的网络地址和全1的广播地址)

- B类细节



- C类细节



- 一般不使用的IPv4地址

网络号	主机号	IP地址	作为源地址	作为目的地址	表示的意思
0	0	0.0.0.0	可以	不可以	在本网络上的本主机 (例如, DHCP协议)
0	host-id	0.host-id	可以	不可以	在本网络上的某台主机host-id
全1	全1	255.255.255.255	不可以	可以	只在本网络上进行广播 (各路由器均不转发)
net-id	全1	A类: net-id.255.255.255 B类: net-id.255.255 C类: net-id.255	不可以	可以	对网络net-id上的所有主机进行广播
127	非全0或全1的任何数	127.0.0.1~127.255.255.254	可以	可以	用于本地软件环回测试

- 练习

【练习】请填写以下两个表格的内容。

解析

(1) 根据地址左起第一个十进制数的值, 可以判断出地址类别:
小于127为A类;
128~191为B类;
192~223为C类。

(2) 根据地址类别, 可以找出地址中的网络号部分和主机号部分:
A类: 网络号为左起第一个字节
B类: 网络号为左起前两个字节
C类: 网络号为左起前三个字节

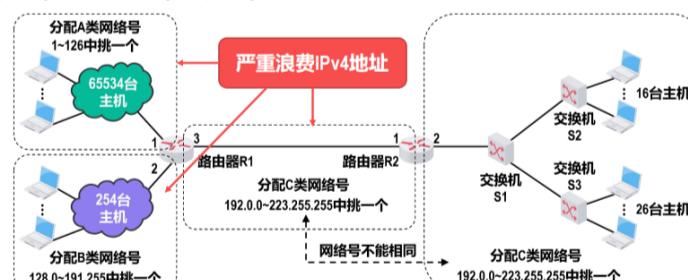
(3) 以下三种情况的地址不能分配给主机或路由器的接口:
A类网络号0和127;
主机号为全0, 这是网络地址;
主机号为全1, 这是广播地址。

IPv4地址	类别	是否可以分配给主机或路由器接口
0.1.2.3	A	不能分配, 网络号0是保留的网络号
1.2.3.4	A	可以分配, 网络号为1, 主机号为2.3.4
126.255.255.255	A	不能分配, 网络号为126, 主机号为255.255.255, 广播地址
127.0.0.1	A	不能分配, 网络号为127, 主机号为0.0.1, 本地环回测试地址
128.0.255.255	B	不能分配, 网络号为128.0, 主机号为255.255, 广播地址
166.16.18.255	B	可以分配, 网络号为166.16, 主机号为18.255
172.18.255.255	B	不能分配, 网络号为172.18, 主机号为255.255, 广播地址
191.255.255.252	B	可以分配, 网络号为191.255, 主机号为255.252
192.0.0.255	C	不能分配, 网络号为192.0.0, 主机号为255, 广播地址
196.2.3.8	C	可以分配, 网络号为196.2.3, 主机号为8
218.75.230.30	C	可以分配, 网络号为218.75.230, 主机号为30
223.255.255.252	C	可以分配, 网络号为223.255.255, 主机号为252

【练习】请给出下图各网络的IPv4地址分配方案, 要求尽量节约IP地址。

- 解析 1. 找出图中有哪些网络;
2. 根据各网络中主机和路由器的接口总数量给各网络分配相应类别的网络号;
3. 给各网络中的各主机和路由器的各接口分配IP地址。

同一个网络中, 不同主机和路由器的各接口的IPv4地址的主机号必须各不相同, 并且不能为“全0”(网络地址)和“全1”(广播地址)。



2.2.4. 划分子网的编址方法

- 在主机号中借用几位比特作为子网号。
- 子网掩码: 表明表明分类IPv4地址的主机号部分被借用了几个比特作为子网号。
 - 由32比特构成。
 - 用左起多个连续的比特1对应IPv4地址中的网络号和子网号。
 - 之后的多个连续的比特0对应IPv4地址中的主机号。
- 将划分子网的IPv4地址与相应的子网掩码进行逐比特的逻辑与运算, 就可得到该IPv4地址所在子网的网络地址。



- 默认子网掩码

- 指在未划分子网的情况下使用的子网掩码

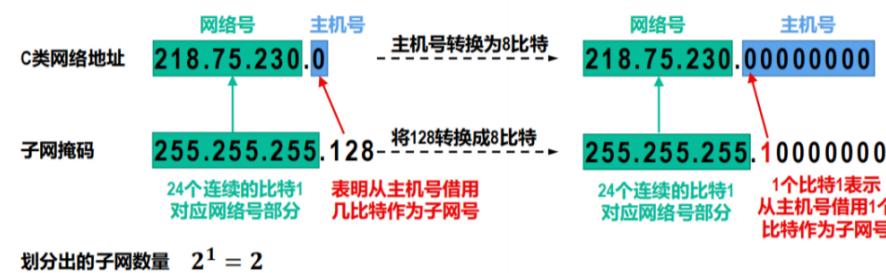
- ABC类主机号全置为0

点分十进制形式		
A类地址	8比特网络号	24比特主机号
A类地址的默认子网掩码	11111111	00000000 00000000 00000000
		255.0.0.0
B类地址	16比特网络号	16比特主机号
B类地址的默认子网掩码	11111111 11111111	00000000 00000000
		255.255.0.0
C类地址	24比特网络号	8比特主机号
C类地址的默认子网掩码	11111111 11111111 11111111	00000000
		255.255.255.0

- 子网划分细节——练习

【举例】已知某个网络的地址为218.75.230.0，使用子网掩码255.255.255.128对其进行子网划分，请给出划分细节。

解析



划分出的子网数量 $2^1 = 2$

每个子网可分配的地址数量 $2^{(8-1)} - 2 = 126$ (减2是去掉主机号为全0的网络地址和全1的广播地址)

网络号	子网号	主机号	子网0的网络地址	218.75.230.0
218.75.230.0	0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0	子网0的网络地址	218.75.230.0
218.75.230.0	0 0 0 0 0 0 0 1	0 0 0 0 0 0 0 1	可分配最小地址	218.75.230.1
⋮	⋮	⋮	⋮	⋮
218.75.230.0	1 1 1 1 1 1 1 0	0 0 0 0 0 0 0 0	可分配最大地址	218.75.230.126
218.75.230.0	1 1 1 1 1 1 1 1	0 0 0 0 0 0 0 0	子网0的广播地址	218.75.230.127
⋮	⋮	⋮	⋮	⋮
218.75.230.1	0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 1	子网1的网络地址	218.75.230.128
218.75.230.1	0 0 0 0 0 0 0 1	0 0 0 0 0 0 0 1	可分配最小地址	218.75.230.129
⋮	⋮	⋮	⋮	⋮
218.75.230.1	1 1 1 1 1 1 1 0	0 0 0 0 0 0 0 0	可分配最大地址	218.75.230.254
218.75.230.1	1 1 1 1 1 1 1 1	0 0 0 0 0 0 0 1	子网1的广播地址	218.75.230.255

【练习】已知某个网络的地址为145.13.0.0，使用子网掩码255.255.192.0对其进行子网划分，请给出划分细节。

解析

网络号	子网号	主机号	子网0的网络地址	145.13.0.0
145.13.0.0	0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0	子网0的网络地址	145.13.0.0
145.13.0.0	0 0 0 0 0 0 0 1	0 0 0 0 0 0 0 1	可分配最小地址	145.13.0.1
⋮	⋮	⋮	⋮	⋮
145.13.0.0	1 0 0 0 0 0 0 0	1 0 0 0 0 0 0 0	可分配最大地址	145.13.0.127
145.13.0.0	1 0 0 0 0 0 0 1	1 0 0 0 0 0 0 1	子网0的广播地址	145.13.0.128
⋮	⋮	⋮	⋮	⋮
145.13.0.1	0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 1	子网1的网络地址	145.13.0.128
145.13.0.1	0 0 0 0 0 0 0 1	0 0 0 0 0 0 0 1	可分配最小地址	145.13.0.129
⋮	⋮	⋮	⋮	⋮
145.13.0.1	1 0 0 0 0 0 0 0	1 0 0 0 0 0 0 0	可分配最大地址	145.13.0.254
145.13.0.1	1 0 0 0 0 0 0 1	1 0 0 0 0 0 0 1	子网1的广播地址	145.13.0.255

划分出的子网数量 $2^2 = 4$

每个子网可分配的地址数量 $2^{(16-2)} - 2 = 126$

2.2.5. 无分类编址方法

- 无分类编址方法使用的地址掩码与划分子网使用的子网掩码类似，由32比特构成。

- 用左起多个连续的比特1对应IPv4地址中的网络前缀。
- 之后的多个连续的比特0对应IPv4地址中的主机号。

- 用斜线记法标记网络前缀比特数。

- /30 应用在分配只有两个路由器接口的点对点链路。【2019 47(2)】

128.14.35.7 / 20 { 网络前缀: 20比特
主机号: 12比特 (32-20)

- 无分类域间路由选择 (Classless Inter-Domain Routing, CIDR)

- CIDR消除了传统A类、B类和C类地址以及划分子网的概念。
- CIDR可以更加有效地分配IPv4地址资源，并且可以在IPv6使用之前允许因特网的规模继续增长。

- CIDR地址块: 由将网络前缀都相同的、连续的多个无分类IPv4地址组成

- 地址块中的最小地址
- 地址块中的最大地址
- 地址块中的地址数
- 地址块中聚合某类网络 (A类、B类、C类) 的数量
- 地址掩码

- CIDR细节练习

【例1】给定的无分类编址的IPv4地址为128.14.35.7/20，请给出该地址所在CIDR地址块的全部细节。

解析

将左起第3、4个十进制数转换成二进制形式
 128.14.35.7/20 → 128.14.00100011.00000111
 最小地址 128.14.32.0 128.14.00100000.00000000
 最大地址 128.14.47.255 128.14.00101111.11111111
 地址数量 2^{32-20}

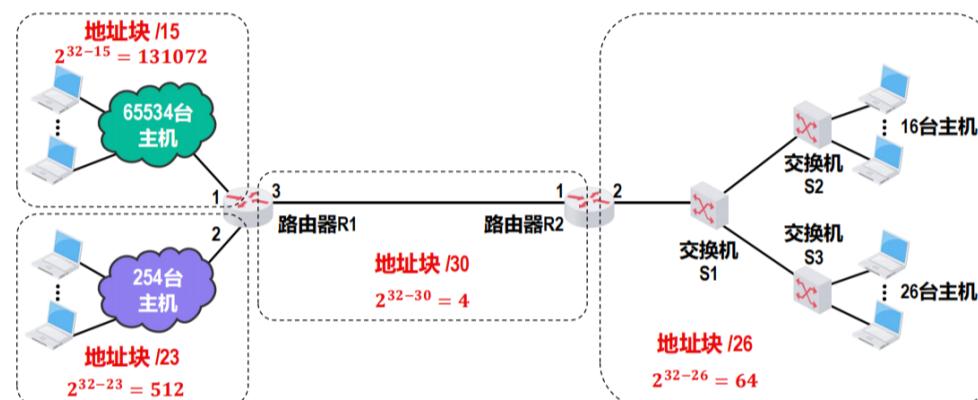
聚合C类网的数量 $2^{32-20} \div 2^8$
 地址掩码 255.255.240.0 11111111.11111111.11110000.00000000

【练习】给定的无分类编址的IPv4地址为206.0.64.8/18，请给出该地址所在CIDR地址块的全部细节。

解析

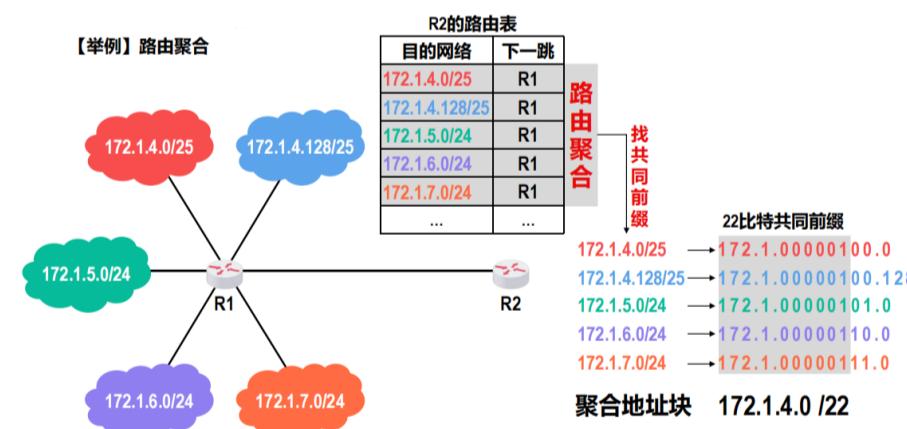
将左起第3、4个十进制数转换成二进制形式
 206.0.64.8/18 → 206.0.01 000000.00001000
 最小地址 206.0.64.0 206.0.01 000000.00000000
 最大地址 206.0.127.255 206.0.01 111111.11111111
 地址数量 2^{32-18}
 聚合C类网的数量 $2^{32-18} \div 2^8$
 地址掩码 255.255.192.0 11111111.11111111.11000000.00000000

- 使用无分类编址方法，可以根据客户的需要分配适当大小的CIDR地址块，因此可以更加有效地分配IPv4的地址空间。



- 使用无分类编址方法的另一个好处是路由聚合（也称为构造超网）。

- 网络前缀越长，地址块越小，路由越具体。
- 若路由器查表转发分组时发现有多条路由条目匹配，则选择网络前缀最长的那条路由条目，这称为最长前缀匹配，因为这样的路由更具体。



2.3. IPv4地址的应用规划

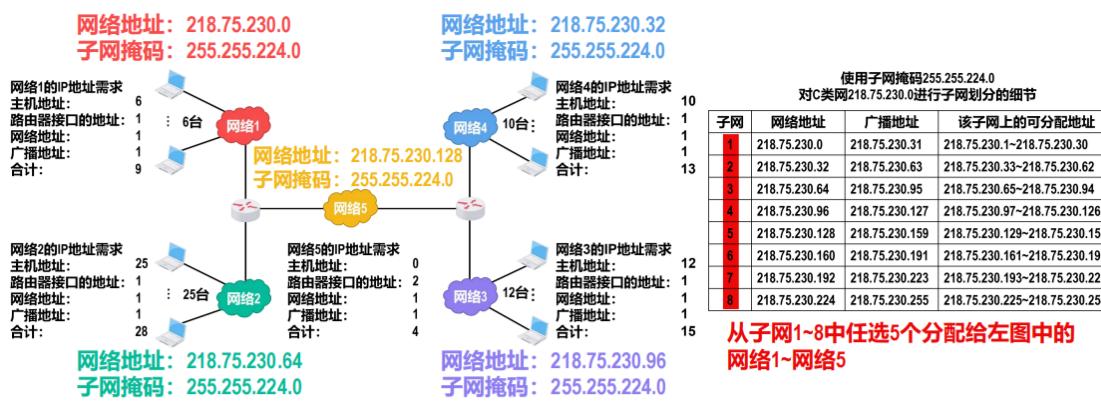
■ IPv4地址的应用规划是指将给定的IPv4地址块（或分类网络）划分成若干个更小的地址块（或子网），并将这些地址块（或子网）分配给互联网中的不同网络，进而可以给各网络中的主机和路由器的接口分配IPv4地址。

定长的子网掩码 (Fixed Length Subnet Mask, FLSM)
● 所划分出的每一个子网都使用同一个子网掩码。
● 每个子网所分配的IP地址数量相同，容易造成地址资源的浪费。

变长的子网掩码 (Variable Length Subnet Mask, VLSM)
● 所划分出的每一个子网可以使用不同的子网掩码。
● 每个子网所分配的IP地址数量可以不同，尽可能减少对地址资源的浪费。

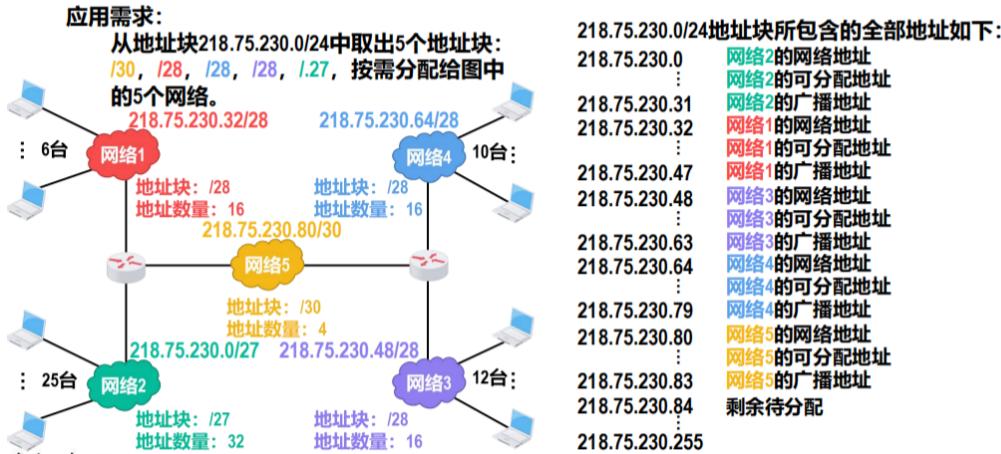
2.3.1. 定长子网掩码

【举例】假设申请到的C类网络为218.75.230.0，使用定长的子网掩码给下图所示的小型互联网中的各设备分配IPv4地址。



2.3.2. 变长子网掩码

【举例】假设申请到的地址块为218.75.230.0/24，使用变长的子网掩码给下图所示的小型互联网中的各设备分配IPv4地址。

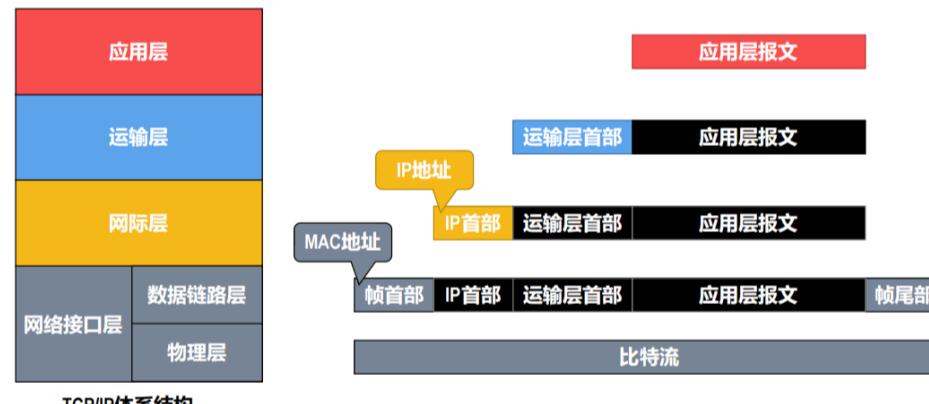


- 在地址块中选取子块的原则

- 每个子块的起点位置不能随便选取，只能选取主机号部分是块大小整数倍的地址作为起点。
- 建议先为大的子块选取。

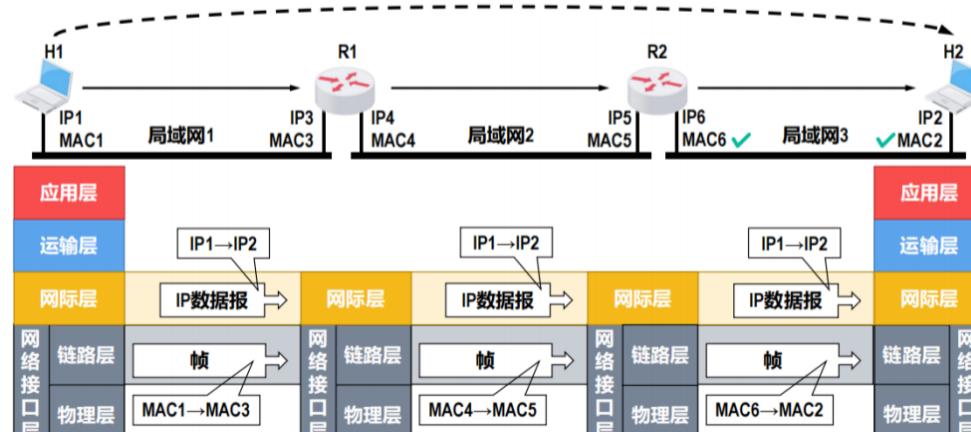
2.4. IPv4地址与MAC地址

2.4.1. IPv4地址与MAC地址的封装位置



2.4.2. 数据包传送过程中IPv4地址与MAC地址的变化情况

- 在数据包的传送过程中，数据包的源IP地址和目的IP地址保持不变；
- 在数据包的传送过程中，数据包的源MAC地址和目的MAC地址逐链路（或逐网络）改变。

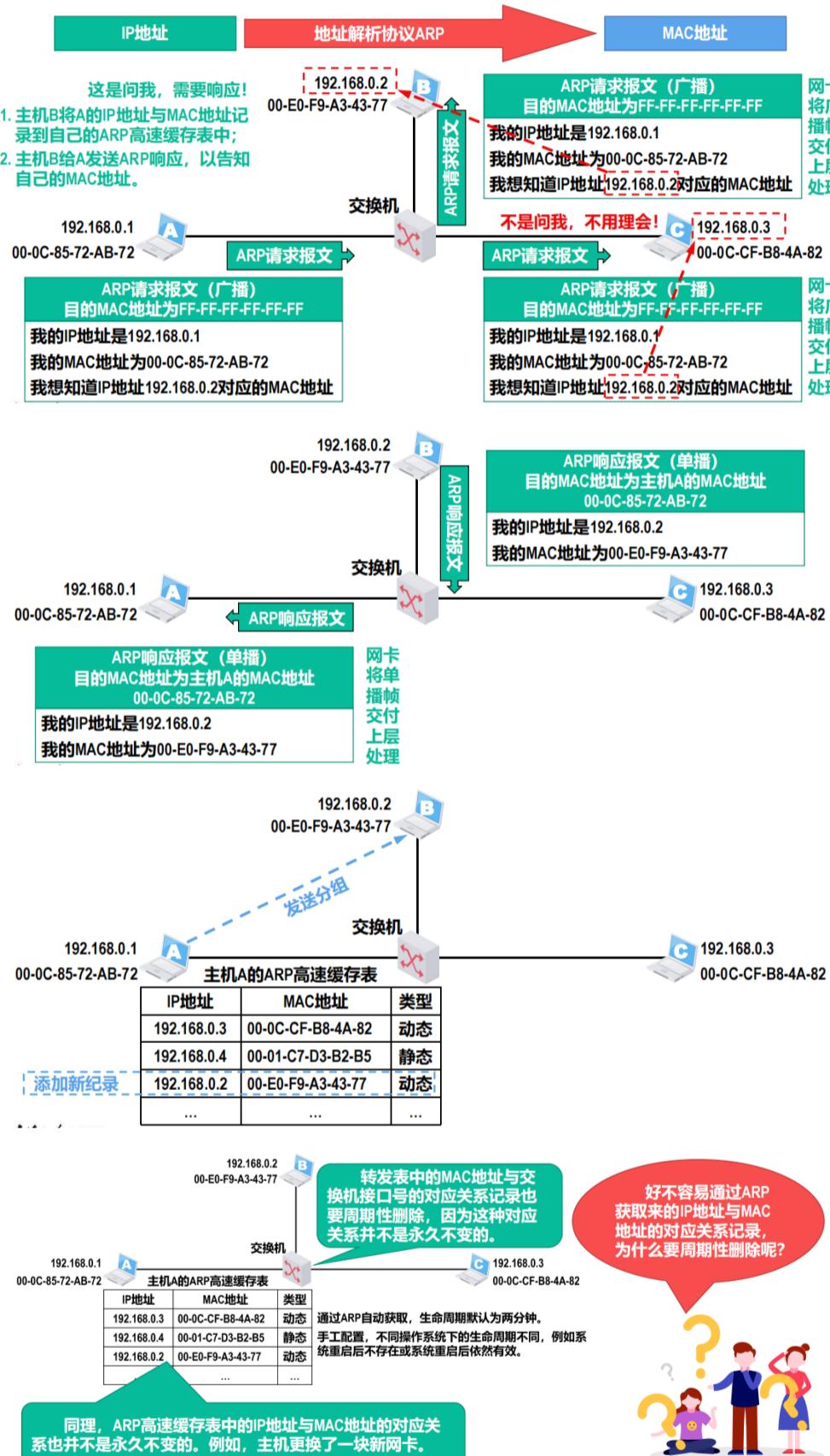


2.4.3. IPv4地址与MAC地址的关系

- 如果仅使用MAC地址进行通信，则会出现以下主要问题：
- 因特网中的每台路由器的路由表中就必须记录因特网上所有主机和路由器各接口的MAC地址。
 - 手工给各路由器配置路由表几乎是不可能完成的任务，即使使用路由协议让路由器通过相互交换路由信息来自动构建路由表，也会因为路由信息需要包含海量的MAC地址信息而严重占用通信资源。
 - 包含海量MAC地址的路由信息需要路由器具备极大的存储空间，并且会给分组的查表转发带来非常大的时延。
- 因特网的网际层使用IP地址进行寻址，就可使因特网中各路由器的路由表中的路由记录的数量大大减少，因为只需记录部分网络的网络地址，而不是记录每个网络中各通信设备的各接口的MAC地址。
- 路由器收到IP数据报后，根据其首部中的目的IP地址的网络号部分，基于自己的路由表进行查表转发。
查表转发的结果可以指明IP数据报的下一跳路由器的IP地址，但无法指明该IP地址所对应的MAC地址。因此，在数据链路层封装该IP数据报成为帧时，帧首部中的目的MAC地址字段就无法填写，该问题需要使用网际层中的地址解析协议ARP来解决。

2.5. 地址解析协议ARP

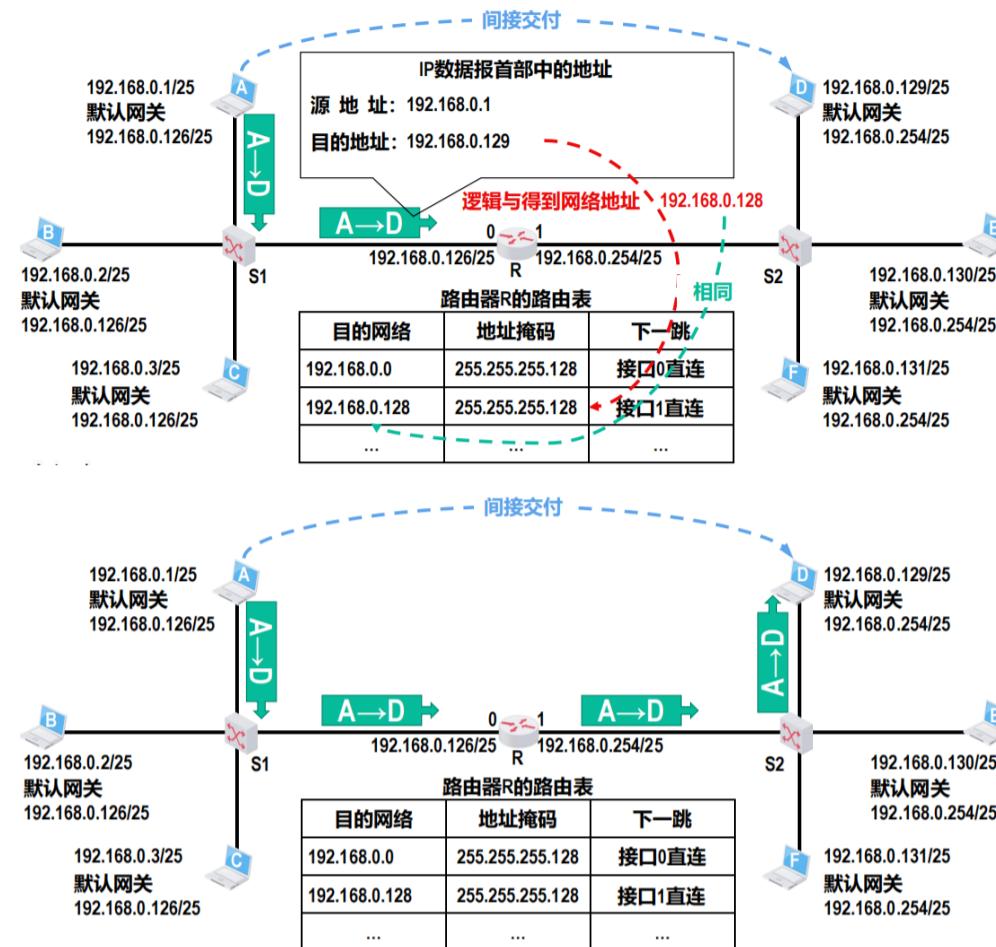
- Address Resolution Protocol



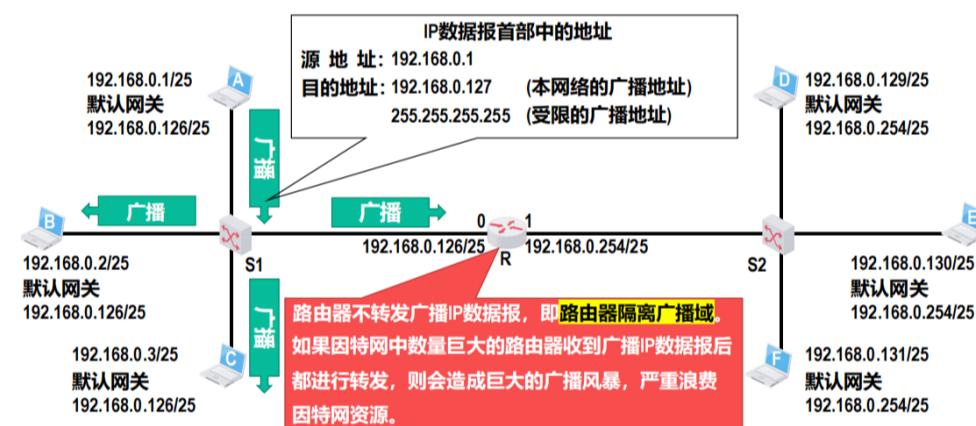
2.6. IP数据报的发送和转发流程

- 单播

- 默认网关：路由器的IP地址
- 将IP数据报中的目的地址与路由表中的地址掩码进行逻辑与运算找到下一跳



- 广播



2.6.1. 主机发送IP数据报

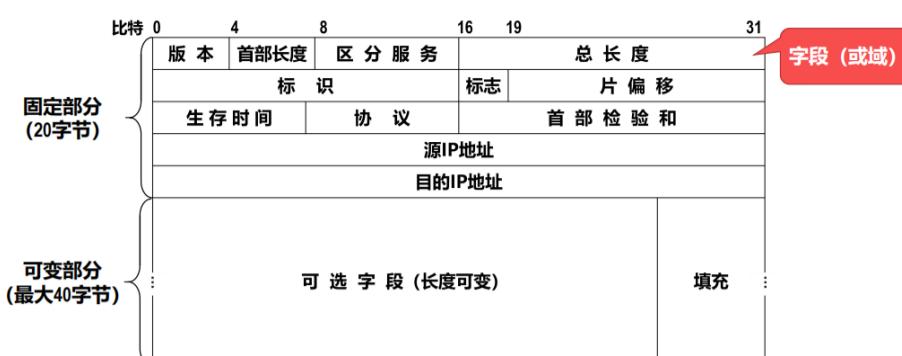
- 判断目的主机是否与自己在同一个网络
 - 若在同一个网络，则属于直接交付
 - 若不在同一个网络，则属于间接交付。发送给主机所在网络的默认网关（路由器），由默认网关帮忙转发。

2.6.2. 路由器转发IP数据报

- ① 检查收到的P数据报是否正确（生存时间是否结束；首部是否误码）
 - 若出错，则丢弃该P数据报并给发送该P数据报的源主机发送差错报告；
 - 若正确，则进行查表转发。
- ② 基于P数据报首部中的目的P地址在路由表中进行查找
 - 若找到匹配的路由条目，则按该路由条目的指示进行转发；
 - 若找不到匹配的路由条目，则丢弃该P数据报，并向发送该P数据报的源主机发送差错报告。
- ③ 查找步骤
 - 先根据路由表中目的地址的网络前缀，得出各目的地址的子网掩码。
 - 将子网掩码与所收到的IP数据包中的目的地址相与。
 - 若相与结果和路由表中对应的目的地址匹配，则选取网络前缀最长的那个目的地址进行转发。
 - 若相与结果不匹配任一目的地址，则选择默认路由0.0.0.0/0进行转发。

2.7. IPv4数据报的首部格式

- IPv4数据报的首部格式及其内容是实现IPv4协议各种功能的基础。
- 在TCP/IP标准中，各种数据格式常常以32比特（即4字节）为单位来描述。



- 固定部分是指每个IPv4数据报都必须要包含的部分。
- 某些IPv4数据报的首部，除了包含20字节的固定部分，还包含一些可选的字段来增加IPv4数据报的功能。
- IPv4数据报首部中的各字段或某些字段的组合，用来表达IPv4协议的相关功能。

2.7.1. 版本

- 长度为4个比特，用来表示IP协议的版本。
- 通信双方使用的IP协议的版本必须一致。目前广泛使用的IP协议的版本号为4（即IPv4）。

2.7.2. 首部长度

- 长度为4个比特，该字段的取值以4字节为单位，用来表示IPv4数据报的首部长度。
 - 最小取值为二进制的0101，即十进制的5，再乘以4字节单位，表示IPv4数据报首部只有20字节固定部分。
 - 最大取值为二进制的1111，即十进制的15，再乘以4字节单位，表示IPv4数据报首部包含20字节固定部分和最大40字节可变部分。

2.7.3. 可选字段

- 长度从1字节到40字节不等，用来支持排错、测量以及安全措施等功能。
 - 虽然可选字段增加了IPv4数据报的功能，但这同时也使得IPv4数据报的首部长度成为可变的，这就增加了因特网中每一个路由器处理IPv4数据报的开销。
 - 实际上，可选字段很少被使用。

2.7.4. 填充

- 用来确保IPv4数据报的首部长度是4字节的整数倍，使用全0进行填充。

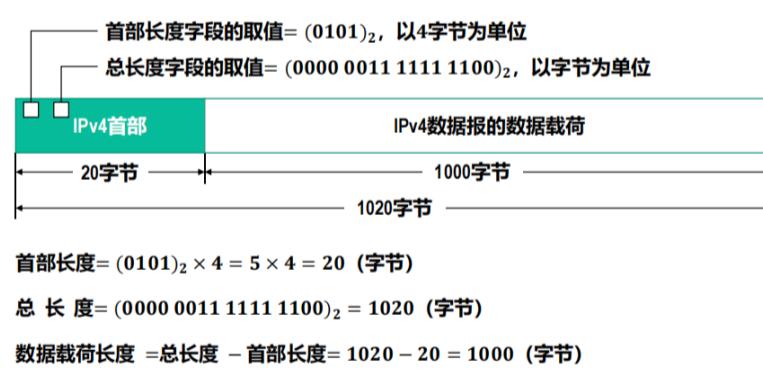
2.7.5. 区分服务

- 长度为8个比特，用来获得更好的服务。

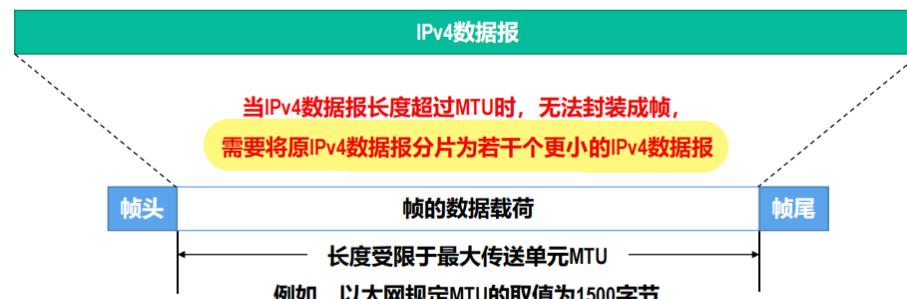
2.7.6. 总长度

- 长度为16个比特，该字段的取值以字节为单位，用来表示IPv4数据报的长度（首部长度+数据载荷长度）。
- 最大取值为二进制的16个比特1，即十进制的65535（很少传输这么长的IPv4数据报）。
- 计算举例

【举例】IPv4数据报首部中的首部长度字段和总长度字段。



2.7.7. IP数据包分片



2.7.7.1. 标识

- 长度为16个比特，属于同一个IPv4数据报的各分片数据报应该具有相同的标识。
- IP软件会维持一个计数器，每产生一个IPv4数据报，计数器值就加1，并将此值赋给标识字段。

2.7.7.2. 标志

- 最低位 (More Fragment, MF)
 - MF=1表示本分片后面还有分片
 - MF=0表示本分片后面没有分片
- 中间位 (Don't Fragment, DF)
 - DF=1表示不允许分片

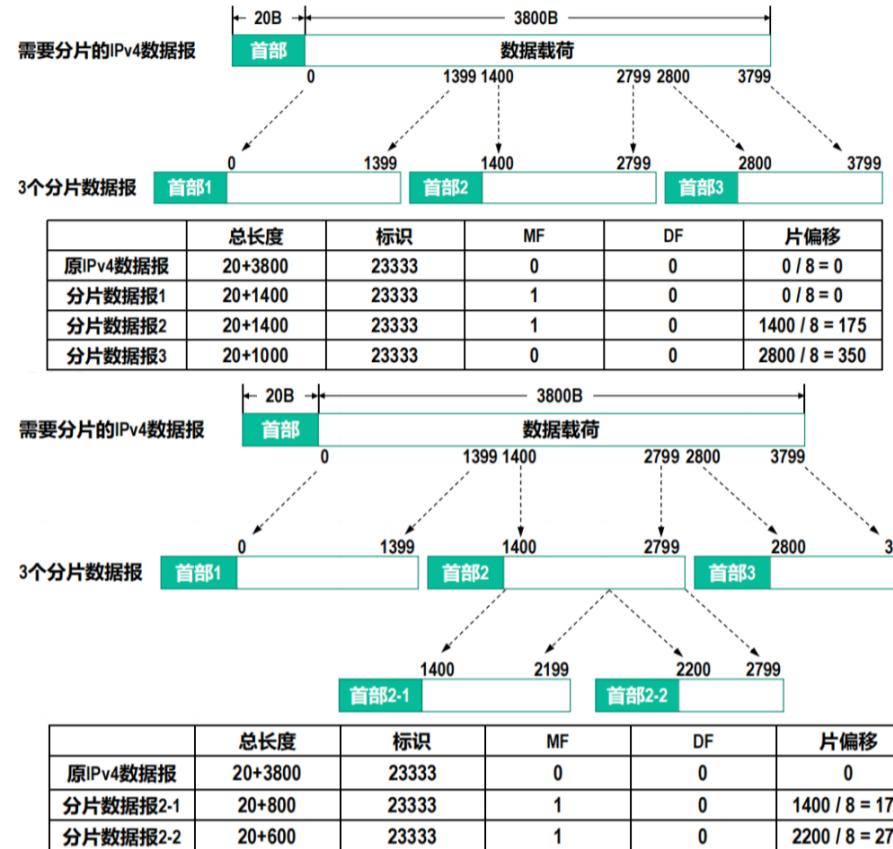
- DF=0表示允许分片
- 最高位为保留位，必须设置为0

2.7.7.3. 片偏移

- 长度为13个比特，该字段的取值以8字节为单位。
- 作用：指出分片IPv4数据报的数据载荷偏移其在原IPv4数据报的位置有多远。
- 计算
 - 分片第一个字节的序号/8
 - 片偏移必须是整数，并且以8字节为单位，若算出来小数，则需要调整分片长度为8的倍数。【2021 36】

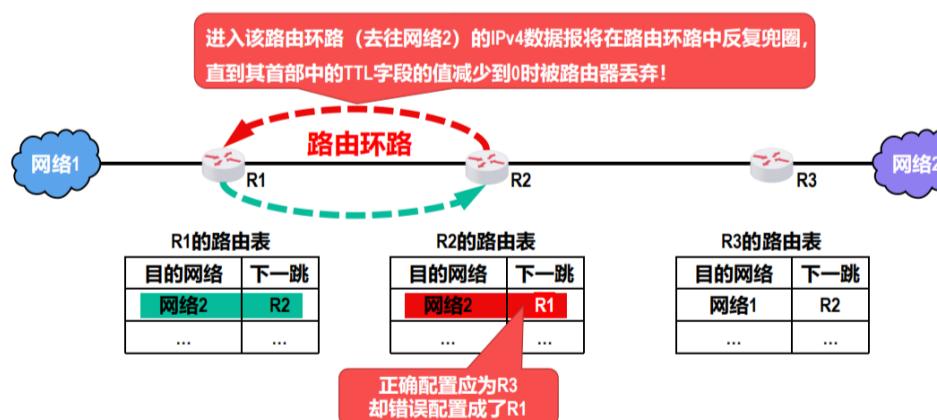
计算举例

【举例】某个IPv4数据报总长度为3820字节，采用20字节固定首部，根据数据链路层要求，需要将该IPv4数据报分片为长度不超过1420字节的数据报片。



2.7.8. 生存时间 (Time To Live, TTL)

- 长度为8个比特，最大取值为二进制的11111111，即十进制的255。
- 该字段的取值最初以秒为单位。
 - 因此，IPv4数据报的最大生存时间最初为255秒。
 - 路由器转发IPv4数据报时，将其首部中该字段的值减去该数据报在路由器上所耗费的时间，若结果不为0就转发，否则就丢弃。
- 生存时间字段后来改为以“跳数”为单位
 - 路由器收到待转发的IPv4数据报时，将其首部中的该字段的值减1，若结果不为0就转发，否则就丢弃。
- 作用：防止被错误路由的IPv4数据报无限制地在因特网中兜圈。

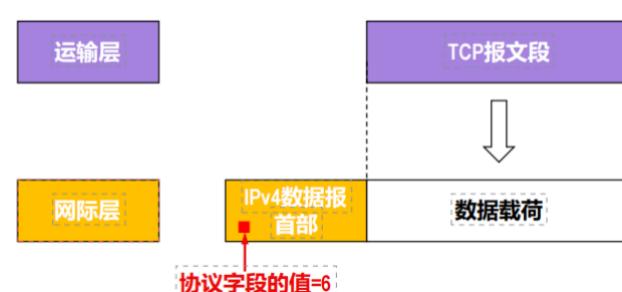


2.7.9. 协议

- 长度为8个比特
- 作用：指明IPv4数据报的数据载荷是何种协议数据单元PDU。

常用的一些协议和相应的协议字段值

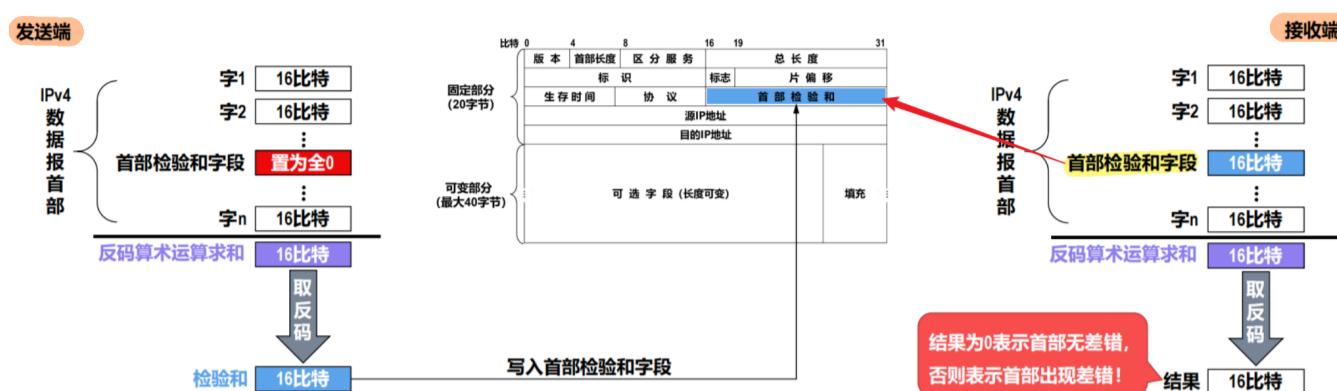
协议名称	ICMP	IGMP	TCP	UDP	IPv6	OSPF
协议字段值	1	2	6	17	41	89



2.7.10. 首部检验和

- 长度为16个比特
- 作用：用于检测IPv4数据报在传输过程中其首部是否出现了差错。
- IPv4数据报每经过一个路由器，其首部中的某些字段的值（例如生存时间TTL、标志以及片偏移等）都可能发生变化，因此路由器都要重新计算一下首部检验和。

2.7.10.1. 首部检验和的计算方法



- 由于网际层并不向其高层提供可靠传输的服务，并且计算首部检验和是一项耗时的操作，因此在IPv6中，路由器不再计算首部检验和，从而更快转发IP数据报。

2.7.11. 源IP地址

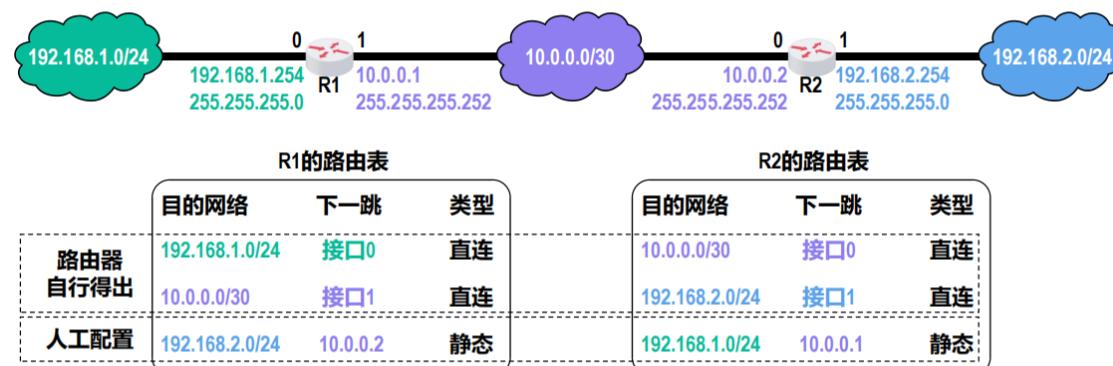
- 长度为32个比特
- 作用：用来填写发送IPv4数据报的源主机的IPv4地址。

2.7.12. 目的IP地址

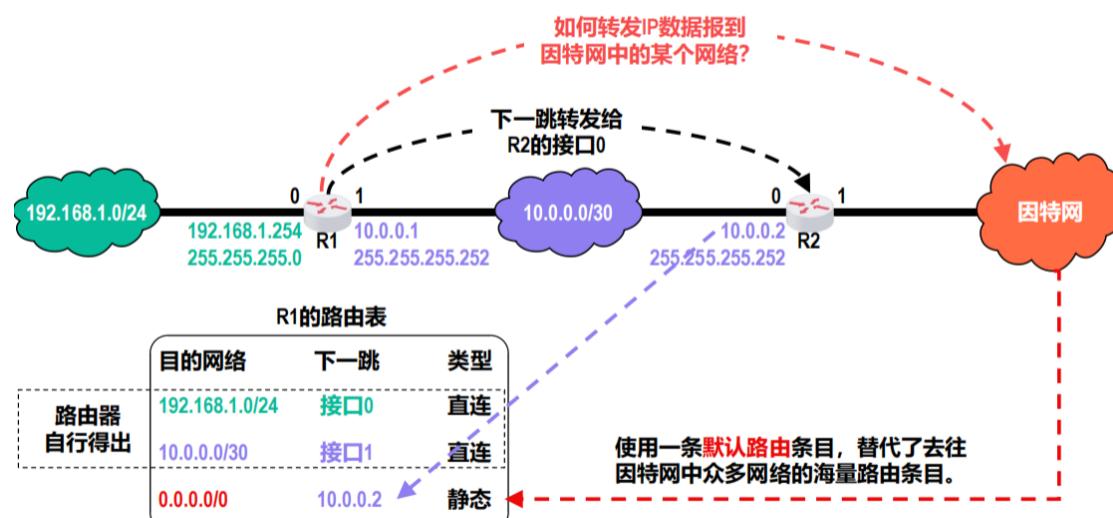
- 长度为32个比特
- 用来填写接收IPv4数据报的目的主机的IPv4地址。

3. 静态路由配置

3.1. 人工配置静态路由



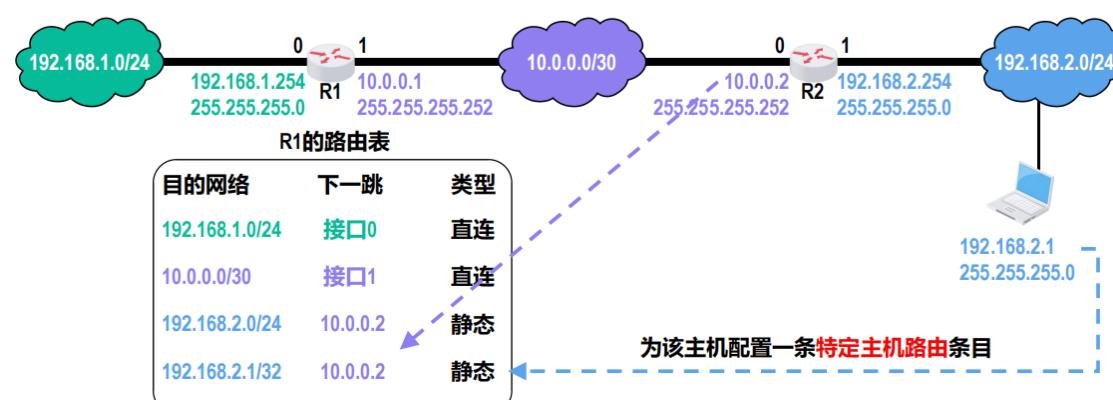
3.2. 默认路由0.0.0.0/0



- 默认路由条目中的目的网络0.0.0.0/0，其中0.0.0.0表示任意网络，而网络前缀“/0”（相应的地址掩码为0.0.0.0）是最短的网络前缀。
- 路由器在查找转发表转发IP数据报时，遵循“最长前缀匹配”的原则，因此默认路由条目的匹配优先级最低。
- 默认路由可以减少路由表所占用的存储空间和搜索路由表所耗费的时间。

3.3. 特定主机路由/32

- 特定主机路由条目的匹配优先级最高。



- 特定主机路由条目中的目的网络192.168.2.1/32，其中192.168.2.1是特定主机的IP地址，而网络前缀“/32”（相应地址掩码为255.255.255.255）是最长的网络前缀。
- 图中，在查表转发去往192.168.2.1这台特定主机的IP数据报时，192.168.2.0/24和192.168.2.1/32两个路由条目都可以匹配，遵循“最长前缀匹配”的原则，按照匹配优先级最高的特定主机路由条目进行转发。

3.4. 需注意的问题

- 进行静态路由配置需要认真考虑和谨慎操作，否则可能出现以下问题：
 - 路由条目配置错误，甚至导致出现路由环路。
 - 聚合路由条目时可能引入不存在的网络。

4. 因特网的路由选择协议

4.1. 路由选择分类

4.1.1. 静态路由选择

- 采用人工配置的方式给路由器添加网络路由、默认路由和特定主机路由等路由条目。
- 静态路由选择简单、开销小，但不能及时适应网络状态（流量、拓扑等）的变化。
- 静态路由选择一般只在小规模网络中采用。

4.1.2. 动态路由选择

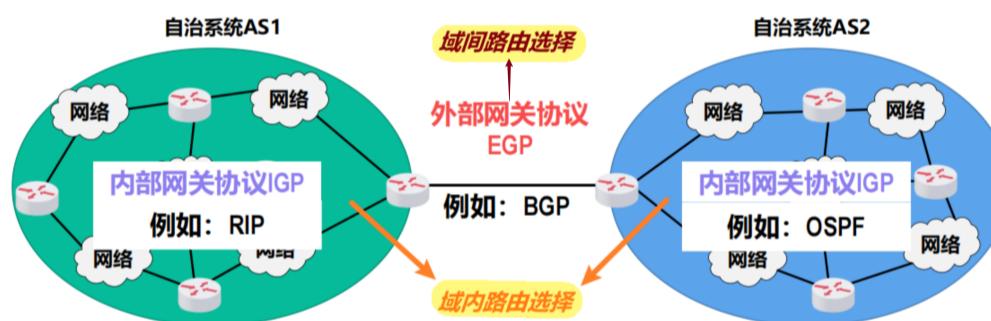
- 路由器通过路由选择协议自动获取路由信息。
- 动态路由选择比较复杂、开销比较大，但能较好地适应网络状态的变化。
- 动态路由选择适用于大规模网络。

4.2. 因特网采用分层次的路由选择协议

4.2.1. 特点

自适应	因特网采用 动态路由 选择，能较好地适应网络状态的变化。
分布式	因特网中的 各路由器 通过相互间的信息交互， 共同完成路由信息的获取和更新 。
分层次	将整个因特网划分为许多较小的 自治系统 (Autonomous System, AS) 。 在自治系统内部和外部采用不同类别的路由选择协议，分别进行路由选择。

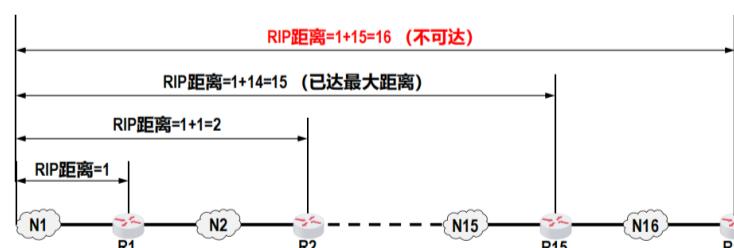
4.2.2. 举例



4.3. 路由信息协议RIP (封装在UDP)

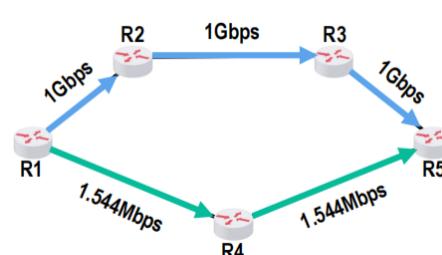
4.3.1. 基本概念

- 路由信息协议 (Routing Information Protocol, RIP) 是内部网关协议中最先得到广泛使用的协议之一，其相关标准文档为[RFC 1058]。
- 距离向量 (Distance-Vector, D-V)：RIP要求自治系统AS内的每一个路由器，都要维护从它自己到AS内其他每一个网络的距离记录。这组距离称为距离向量。
- 跳数 (Hop Count)：度量 (Metric) 来衡量到达目的网络的距离
 - 将路由器到直连网络的距离定义为1。
 - 将路由器到非直连网络的距离定义为所经过的路由器数加1。
 - 允许一条路径最多只能包含15个路由器，距离等于16时相当于不可达。因此RIP只适用于小型互联网。

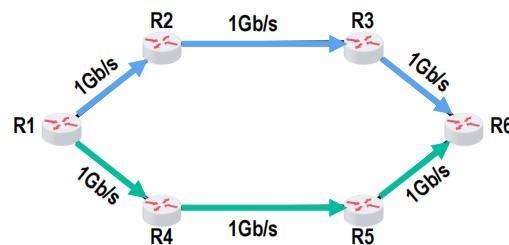


- 好的路由：距离短，即所通过路由器数量最少的路由。

- 如下图，RIP认为R1到R5的好路由是：R1→R4→R5



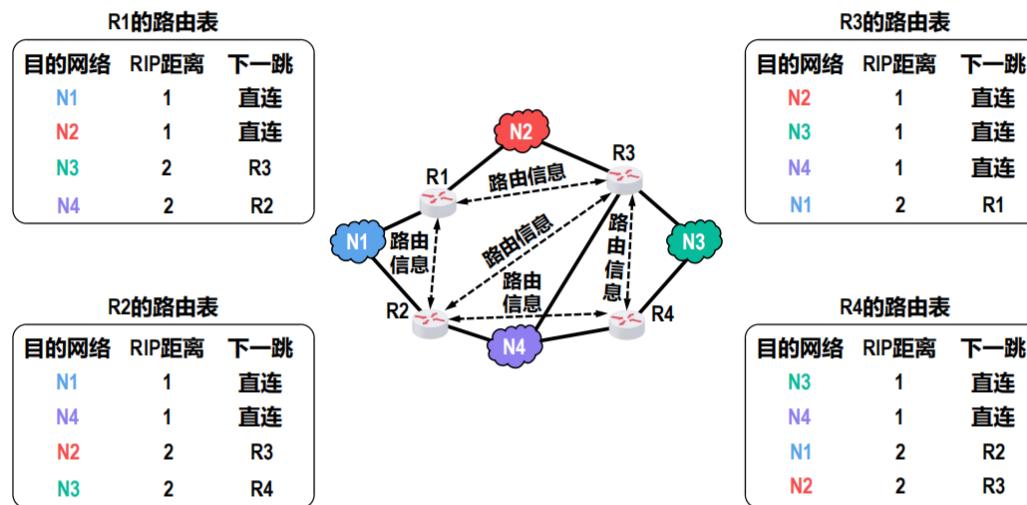
- 等价负载均衡：当到达同一目的网络有多条RIP距离相等的路由时，可以进行等价负载均衡，也就是将通信量均衡地分布到多条等价的路径上。



- 特点

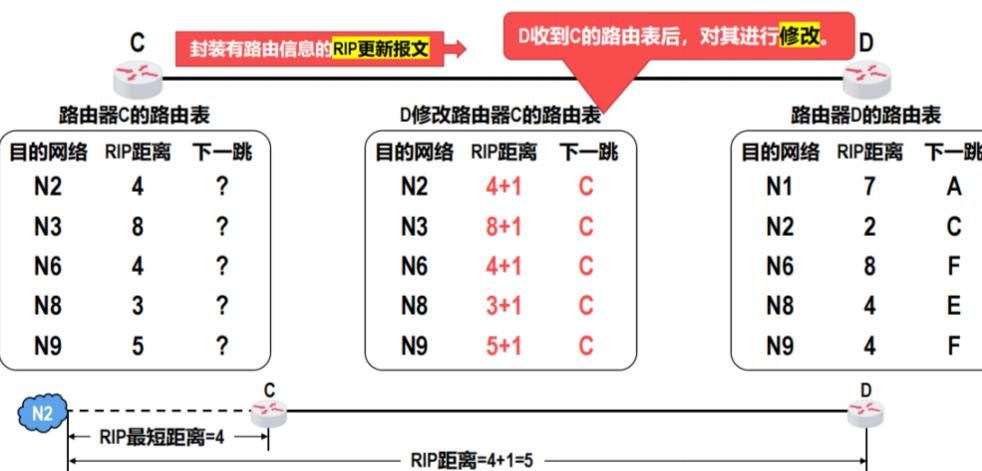
- 和谁交换信息
 - 仅和相邻路由器交换信息。
- 交换什么信息
 - 路由器自己的路由表。
 - 即本路由器到所在自治系统AS中各网络的最短RIP距离，以及到各网络应经过的下一跳路由器。
- 何时交换信息
 - 周期性交换（例如，每个约30秒）。
 - 为了加快RIP的收敛速度，当网络拓扑发生变化时，路由器要及时向相邻路由器通告拓扑变化后的路由信息，这称为触发更新。

4.3.2. 基本工作过程



- 路由器刚开始工作时，只知道自己到直连网络的RIP距离为1。
- 每个路由器仅和相邻路由器周期性地交换并更新路由信息。
- 收敛：若干次交换和更新后，每个路由器都知道到达本自治系统AS内各网络的最短距离和下一跳路由器。

4.3.3. 距离向量算法



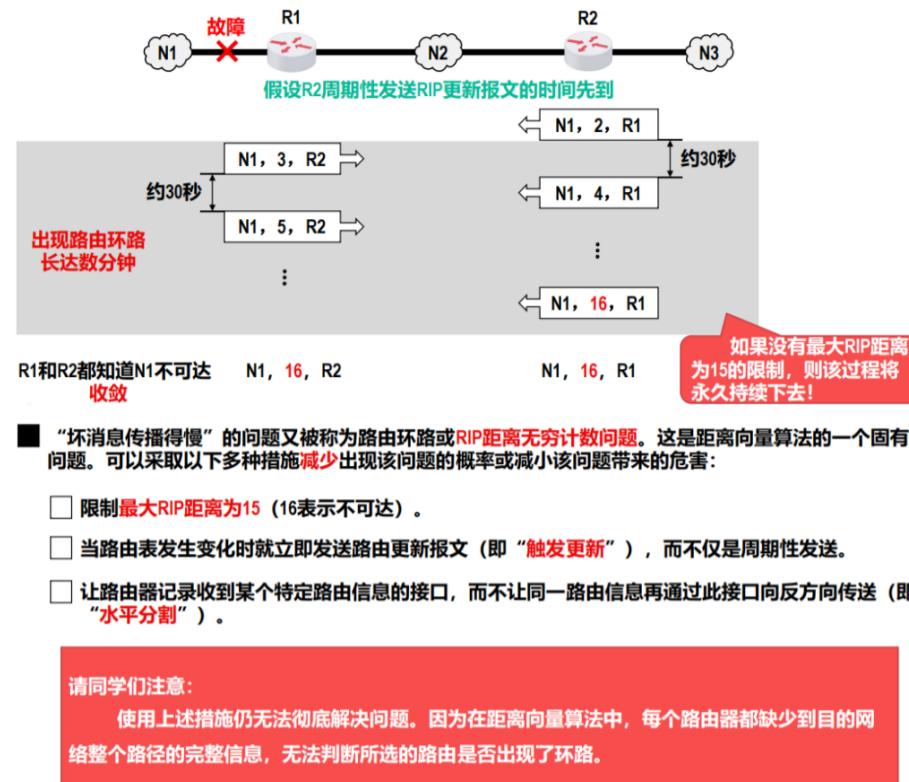
- D不需要关心C中下一跳的内容，收到后将C中的下一条均改为C。
- 从D修改后路由器C的目的网络往下看
 - N2：到达目的网络，相同的下一跳，表示信息为最新，要更新D中N2的RIP距离
 - N3：发现新的网络，向D的路由表中添加。
 - N6：到达目的网络，不同的下一条（F、C），新路由RIP更短（有优势），要更新D中N6的RIP距离
 - N8：到达目的网络，不同的下一条，RIP距离相等，可以等价负载均衡，向D的路由表中添加。
 - N9：到达目的网络，不同的下一条（F、C），新路由RIP更长（处于劣势），不更新D中N9的RIP距离
- 更新后D的路由表如下

目的网络	RIP距离	下一跳
N1	7	A
N2	5	C
N3	9	C
N6	5	C
N8	4	C
N9	4	F

- 除了上述RIP路由条目更新规则，在RIP的距离向量算法中还包含以下一些时间参数：
 - 路由器每隔大约30秒向其所有相邻路由器发送路由更新报文。

- 若180秒（默认）没有收到某条路由条目的更新报文，则把该路由条目标记为无效（即把RIP距离设置为16，表示不可达）。
- 若再过一段时间（如120秒），还没有收到该路由条目的更新报文，则将该路由条目从路由表中删除。

4.3.4. 存在的问题——“坏消息传播得慢”



4.3.5. 版本和相关报文的封装

- 现在较新的RIP版本是1998年11月公布的RIP2[RFC 2453]，已经成为因特网标准协议。与RIP1相比，RIP2可以支持变长子网掩码和CIDR。另外，RIP2还提供简单的鉴别过程并支持多播。
- RIP相关报文使用运输层的用户数据报协议UDP进行封装，使用的UDP端口号为520。
 - 从RIP报文封装的角度看，RIP属于TCP/IP体系结构的应用层。
 - 但RIP的核心功能是路由选择，这属于TCP/IP体系结构的网际层。

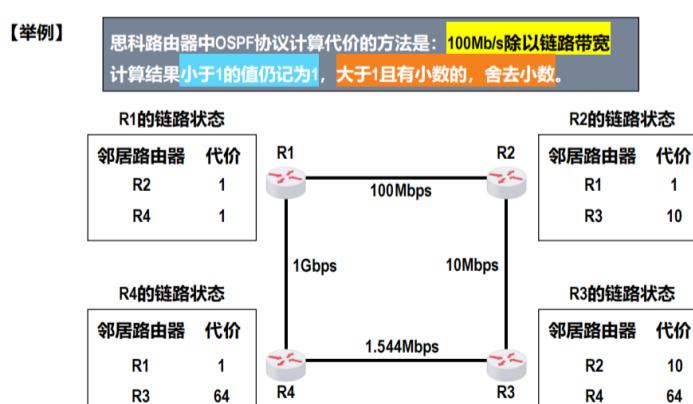
4.3.6. 优缺点

优点	缺点
<ul style="list-style-type: none"> ● 实现简单，路由器开销小。 ● 如果一个路由器发现了RIP距离更短的路由，那么这种更新信息就传播得很快，即“好消息传播得快”。 	<ul style="list-style-type: none"> ● RIP限制了最大RIP距离为15，这就限制了使用RIP的自治系统AS的规模。 ● 相邻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，开销也随之增大。 ● “坏消息传播得慢”，使更新过程的收敛时间过长。因此，对于规模较大的自治系统AS，应当使用OSPF协议。

4.4. 开放最短路径优先OSPF (封装在IP)

4.4.1. 基本概念

- 开放最短路径优先（Open Shortest Path First, OSPF）协议是为了克服路由信息协议RIP的缺点在1989年开发出来的。
 - “开放”表明OSPF协议不是受某一厂商控制，而是公开发表的。
 - “最短路径优先”是因为使用了Dijkstra提出的最短路径算法（Shortest Path First, SPF）。
- “开放最短路径优先”只是一个路由选择协议的名称，但这并不表示其他的路由选择协议不是“最短路径优先”。实际上，用于自治系统AS内部的各种路由选择协议（例如RIP），都要寻找一条“最短”的路径。
- OSPF是基于链路状态的，而不像RIP是基于距离向量的。
 - 链路状态（Link State, LS）是指本路由器都和哪些路由器相邻，以及相应链路的“代价（cost）”。
 - “代价”用来表示费用、距离、时延和带宽等，这些都由网络管理人员来决定。



• 优点

- OSPF基于链路状态并采用最短路径算法计算路由，从算法上保证了不会产生路由环路。
- OSPF不限制网络规模，更新效率高，收敛速度快。

• OSPF路由器邻居关系的建立和维护

- OSPF相邻路由器之间通过交互问候（Hello）分组来建立和维护邻居关系。

问候 (Hello) 分组封装在IP数据报中，发往组播地址224.0.0.5。IP数据报首部中的协议号字段的取值为89，表明IP数据报的数据载荷为OSPF分组。



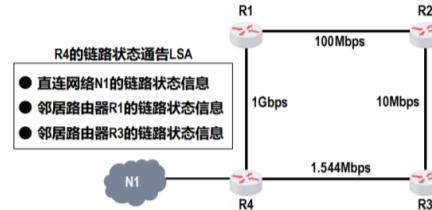
- 链路状态通告

■ 使用OSPF的每个路由器都会产生链路状态通告 (Link State Advertisement, LSA)。

■ LSA中包含以下两类链路状态信息：

直连网络的链路状态信息

邻居路由器的链路状态信息



- 链路状态更新分组
- 链路状态数据库

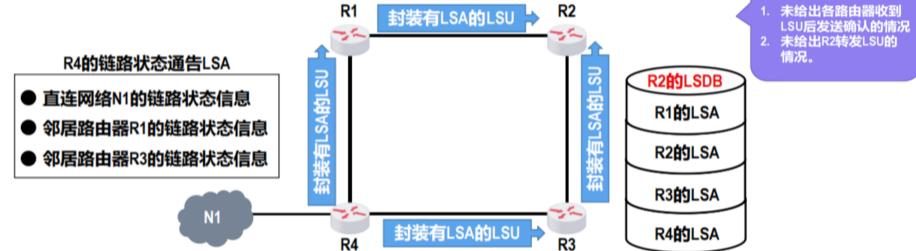
■ 链路状态通告LSA被封装在链路状态更新 (Link State Update, LSU) 分组中，采用可靠的洪泛法 (Flooding) 进行发送。

洪泛法的要点是路由器向自己所有的邻居路由器发送链路状态更新分组，收到该分组的各路由器又将该分组转发给自己所有的邻居路由器（但其上游路由器除外），以此类推。

可靠是指收到链路状态更新分组后要发送确认，收到重复的更新分组无需再次转发，但要发送一次确认。

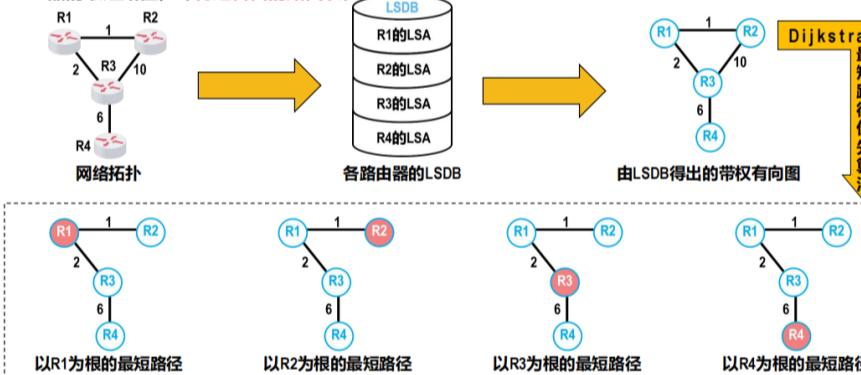
■ 使用OSPF的每一个路由器都有一个链路状态数据库 (Link State Database, LSDB)，用于存储链路状态通告LSA。

■ 通过各路由器洪泛发送封装有各自链路状态通告LSA的链路状态更新分组LSU，各路由器的链路状态数据库LSDB最终将达到一致。



- 基于链路状态数据库进行最短路径优先计算

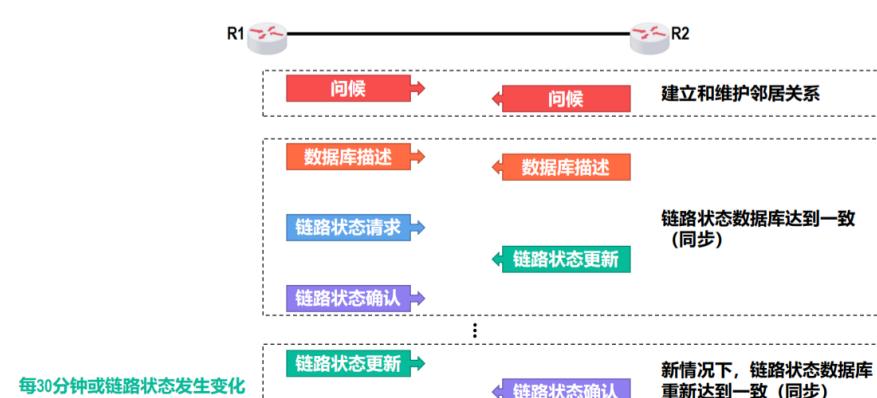
■ 使用OSPF的各路由器，基于链路状态数据库LSDB进行最短路径优先计算，构建出各自到达其他各路由器的最短路径，即构建各自的路由表。



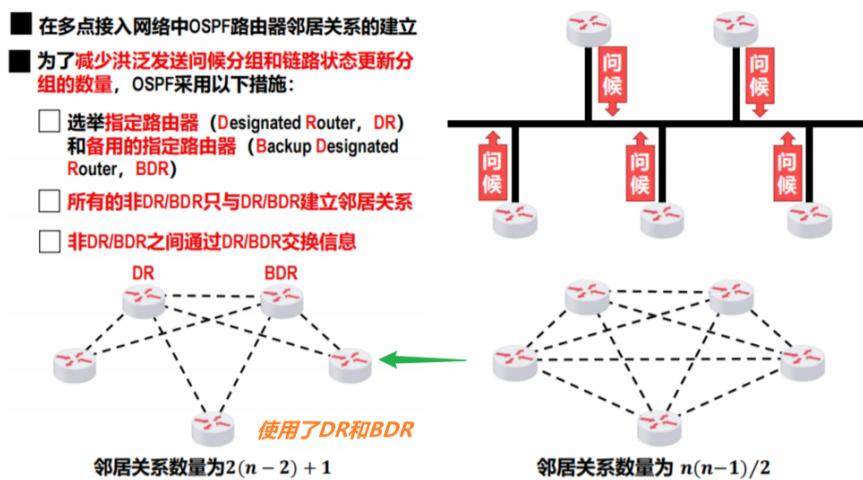
4.4.2. 五种分组类型

问候 (Hello)	用来发现和维护邻居路由器的可达性。
数据库描述 (Database Description)	用来向邻居路由器给出自己的链路状态数据库中的所有链路状态项目的摘要信息。
链路状态请求 (Link State Request)	用来向邻居路由器请求发送某些链路状态项目的详细信息。
链路状态更新 (Link State Update)	路由器使用链路状态更新分组将其链路状态信息进行洪泛发送，即用洪泛法对整个系统更新链路状态。
链路状态确认 (Link State Acknowledgement)	对链路状态更新分组的确认分组。

4.4.3. 基本工作过程

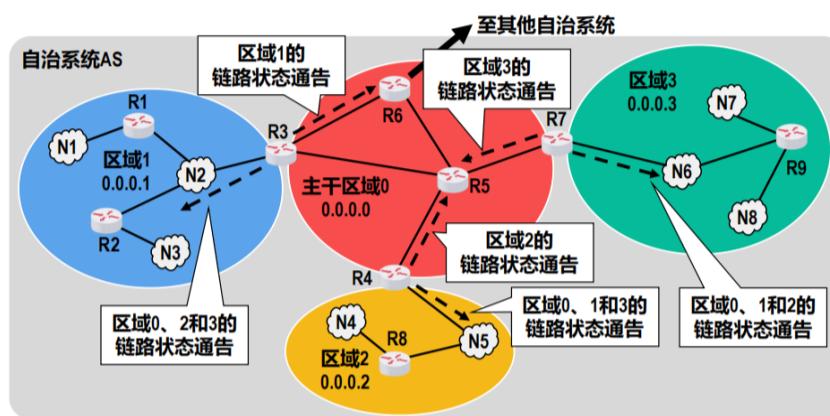


4.4.4. 多点接入网络中的OSPF路由器



4.4.5. OSPF划分区域

- 为了使OSPF协议能够用于规模很大的网络，OSPF把一个自治系统AS再划分为若干个更小的范围，称为区域（area）。
- 每个区域的规模不应太大，一般所包含的路由器不应超过200个。
- 划分区域的好处就是把利用洪泛法交换链路状态信息的范围局限于每一个区域，而不是整个自治系统AS，这样就减少了整个网络上的通信量。

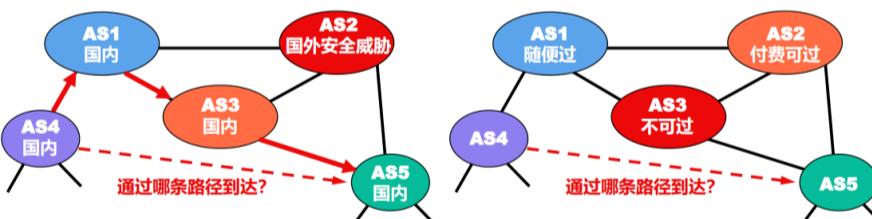


- 自治系统边界路由器（AS Border Router, ASBR）：R6
- 主干路由器（Backbone Router, BBR）：R3、R4、R5、R6和R7
- 区域内路由器（Internal Router, IR）：区域1内的R1和R2，区域2内的R8，区域3内的R9
- 区域边界路由器（Area Border Router, ABR）：R3、R4和R7

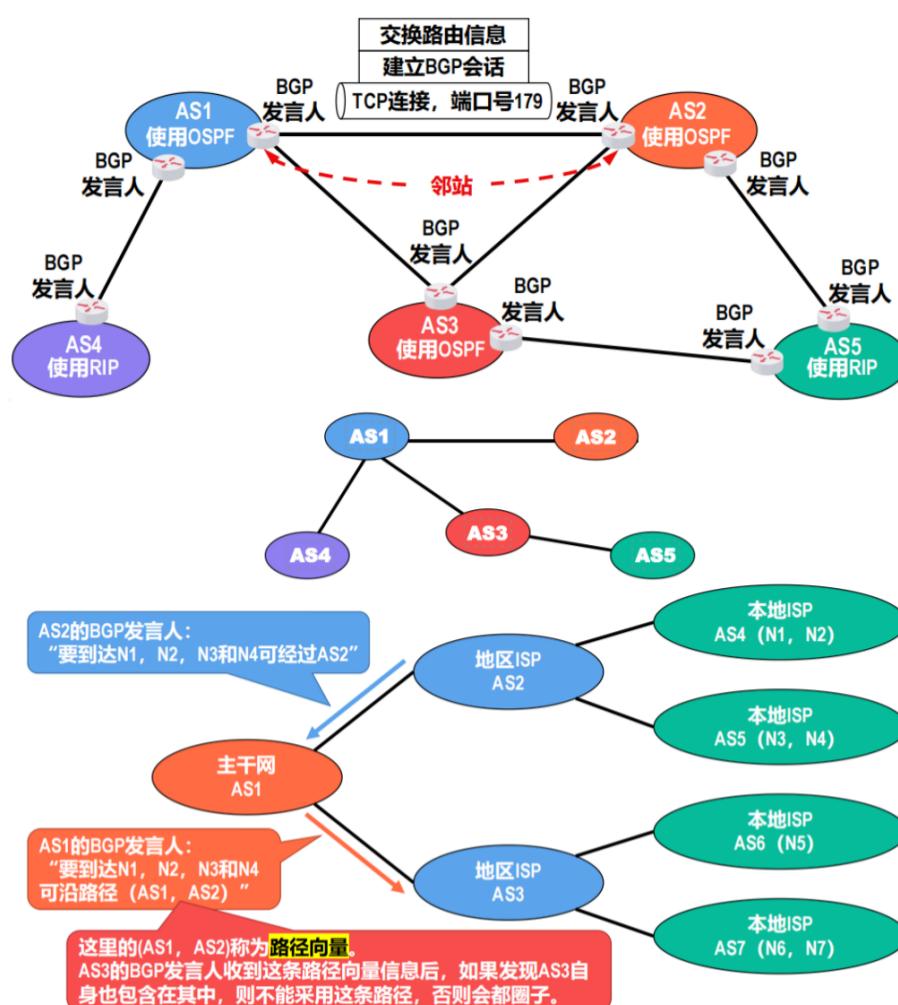
4.5. 边界网关协议BGP (封装在TCP)

4.5.1. 基本概念

- 边界网关协议（Border Gateway Protocol, BGP）属于外部网关协议EGP这个类别，用于自治系统AS之间的路由选择协议。
 - 由于在不同AS内度量路由的“代价”（距离、带宽、费用等）可能不同，因此对于AS之间的路由选择，使用统一的“代价”作为度量来寻找最佳路由是不行的。
 - AS之间的路由选择还必须考虑相关策略（政治、经济、安全等）。
 - 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（即不能兜圈子），而并非要寻找一条最佳路由。



- BGP发言人
 - 在配置BGP时，每个AS的管理员要选择至少一个路由器作为该AS的“BGP发言人”。
 - 一般来说，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。
 - 使用TCP连接交换路由信息的两个BGP发言人，彼此称为对方的邻站（neighbor）或对等站（peer）。
 - BGP发言人除了运行BGP协议外，还必须运行自己所在AS所使用的内部网关协议IGP，例如RIP或OSPF。
 - BGP发言人交换网络可达性的信息，也就是要到达某个网络所要经过的一系列自治系统。
 - 当BGP发言人相互交换了网络可达性的信息后，各BGP发言人就根据所采用的策略，从收到的路由信息中找出到达各自治系统的较好的路由，也就是构造出树形结构且不存在环路的自治系统连通图。
 - BGP适用于多级结构的因特网。



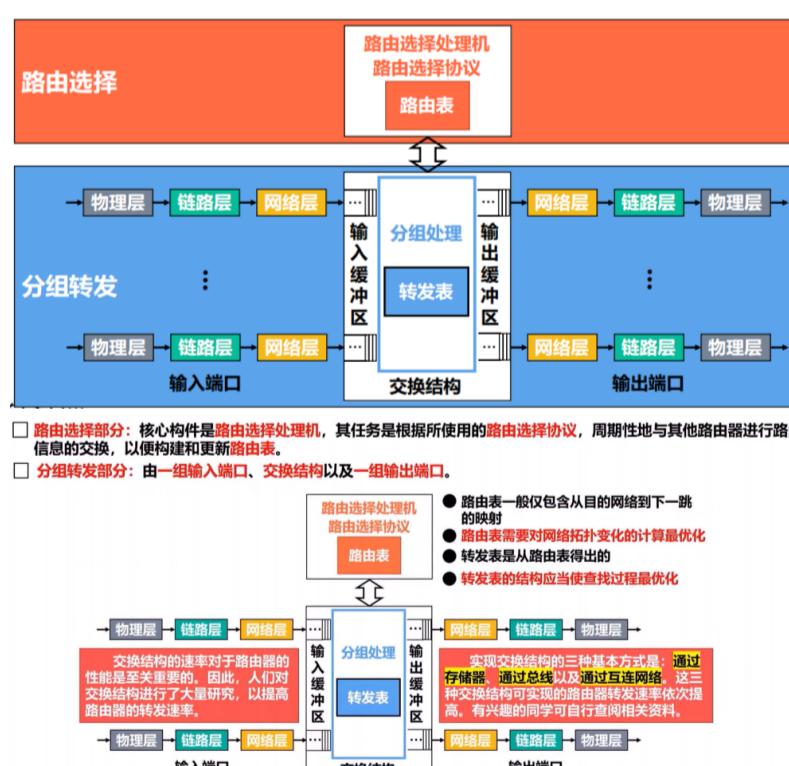
- BGP交换路由信息，要先建立TCP连接，在此基础上交换TCP报文。

4.5.2. BGP-4的四种报文

打开 OPEN	用来与相邻的另一个BGP发言人建立关系，使通信初始化。
保活 KEEPALIVE	用来周期性地证实邻站的连通性。
更新 UPDATE	用来通告某一条路由的信息，以及列出要撤销的多条路由。
通知 NOTIFICATION	用来发送检测到的差错。

4.6. 路由器的基本工作原理

- 路由器是一种具有多个输入端口和输出端口的专用计算机，其任务是转发分组。

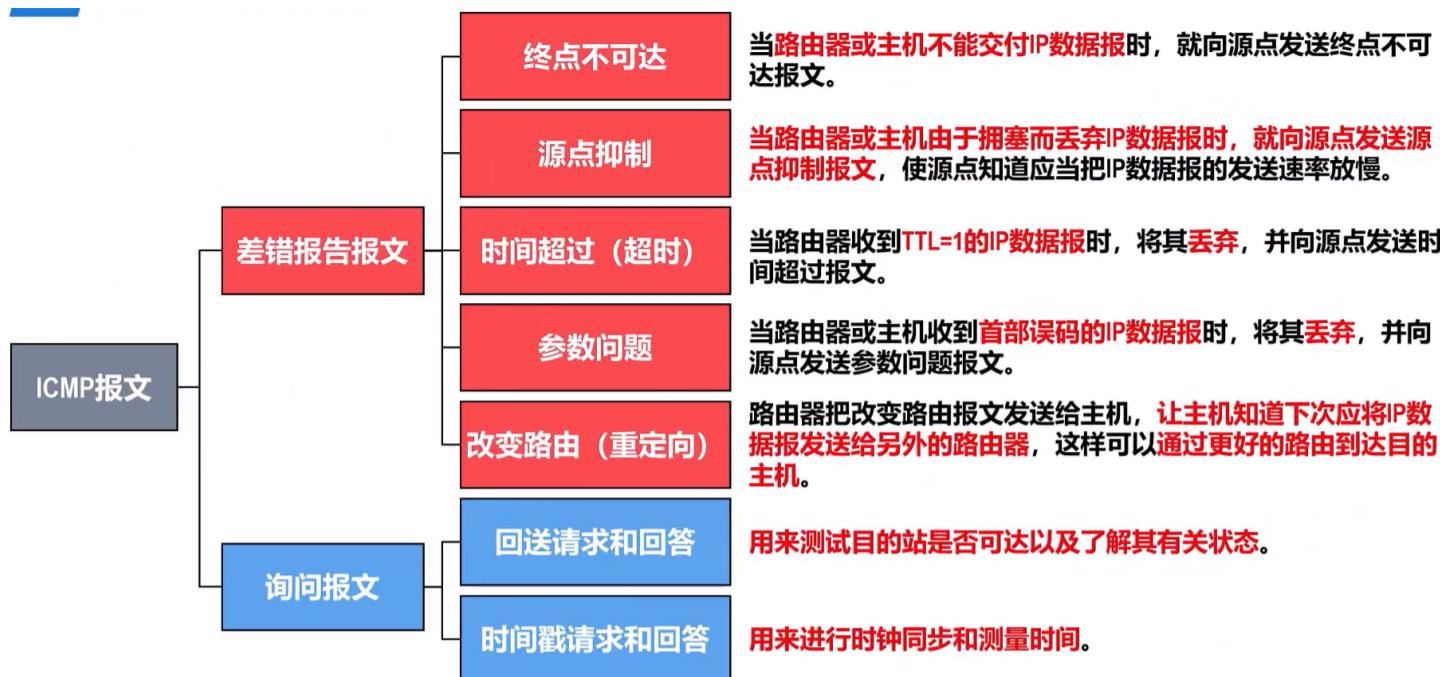


5. 网际控制协议ICMP (封装在IP)

5.1. 概述

- 为了更有效地转发IP数据报以及提高IP数据报交付成功的机会，TCP/IP体系结构的网际层使用了网际控制报文协议 (Internet Control Message Protocol, ICMP) [RFC 792]。
- 主机或路由器使用ICMP来发送差错报告报文和询问报文。
- ICMP报文被封装在IP数据报中发送。

5.2. ICMP报文种类



5.2.1. 差错报告报文

用来向主机或路由器报告差错情况。

- 终点不可达
- 源点抑制
- 时间超过 (超时)
- 参数问题
- 改变路由 (重定向)

以下情况不应发送 ICMP 差错报告报文：
 1 对 ICMP 差错报告报文不再发送 ICMP 差错报告报文。
 2 对第一个分片的 IP 数据报片的所有后续数据报片都不发送 ICMP 差错报告报文。
 3 对具有多播地址的 IP 数据报都不发送 ICMP 差错报告报文。
 4 对具有特殊地址（例如 127.0.0.0 或 0.0.0.0）的 IP 数据报不发送 ICMP 差错报告报文。
 5

5.2.2. 询问报文

用来向主机或路由器询问情况。

- 回送请求和回答
 - 由主机或路由器向一个特定的目的主机或路由器发出。
 - 收到此报文的主机或路由器必须给发送该报文的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
 - 用来请求某个主机或路由器回答当前的日期和时间。
 - 在 ICMP 时间戳回答报文中有一个 32 比特的字段，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。
 - 这种询问报文用来进行时钟同步和测量时间。

5.3. ICMP的典型应用

5.3.1. 分组网间探测 (Packet InterNet Groper, PING)

- 分组网间探测 PING 用来测试主机或路由器之间的连通性。
 - PING 是 TCP/IP 体系结构的应用层直接使用网际层 ICMP 的一个例子，它并不使用运输层的 TCP 或 UDP。
 - PING 应用所使用的 ICMP 报文类型为回送请求和回答。

```

1 C:\Users\ASUS>ping www.bilibili.com
2 正在 Ping a.w.bilicdn1.com [2409:8c3c:4:2::75] 具有 32 字节的数据:
3 来自 2409:8c3c:4:2::75 的回复: 时间=31ms
4 来自 2409:8c3c:4:2::75 的回复: 时间=26ms
5 来自 2409:8c3c:4:2::75 的回复: 时间=25ms
6 来自 2409:8c3c:4:2::75 的回复: 时间=29ms
7
8 2409:8c3c:4:2::75 的 Ping 统计信息:
9   数据包: 已发送 = 4, 已接收 = 4, 丢失 = 0 (0% 丢失),
10  往返行程的估计时间(以毫秒为单位):
11    最短 = 25ms, 最长 = 31ms, 平均 = 27ms
12
  
```

- 某些主机或服务器为了防止恶意攻击，并会不理睬外界发来的 ICMP 回送请求报文。

5.3.2. 跟踪路由 (traceroute)

- 跟踪路由应用 traceroute，用于探测 IP 数据报从源主机到达目的主机要经过哪些路由器。
 - 在 UNIX 版本中，具体命令为“traceroute”，其在运输层使用 UDP 协议，在网络层使用 ICMP 报文类型只有差错报告报文。
 - 在 Windows 版本中，具体命令为“tracert”，其应用层直接使用网际层的 ICMP 协议，所使用的 ICMP 报文类型有回送请求和回答报文以及差错报告报文。

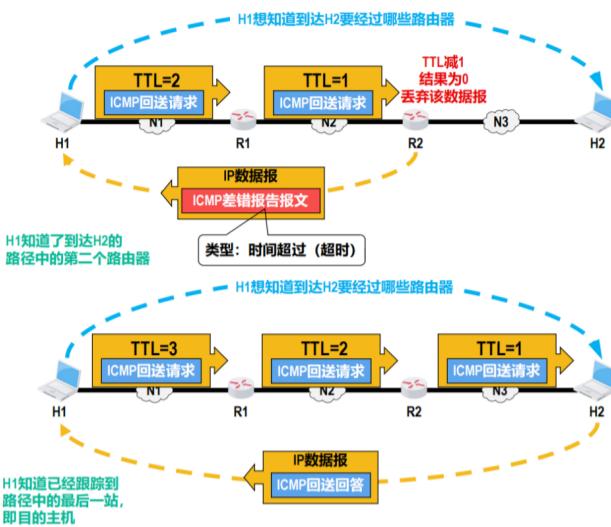
```

1 C:\Users\ASUS>tracert -d www.bilibili.com
2 通过最多 30 个跃点跟踪
3 到 a.w.bilicdn1.com [240e:bf:b800:4300:1::15] 的路由:
4
5     1       6 ms      *      33 ms  2001:da8:20c:a053::1
6     2       2 ms      1 ms      1 ms  2001:da8:20c:f0e8::1
7     3       3 ms      3 ms      3 ms  2001:250:215::1
8     4       *         *         * 请求超时。
  
```

```

10   5  4 ms   3 ms   6 ms  2001:da8:2:122::1
11   6  15 ms  5 ms   7 ms  2001:da8:2:4::1
12   7  4 ms   3 ms   4 ms  2001:da8:2:701:110:108:14:2
13   8  4 ms   3 ms   5 ms  240e::c:1:6200:302
14   9  *     23 ms  23 ms  240e::1:11:51:6303
15  10  23 ms  25 ms  24 ms  240e:f:b800:14e::3
16  11  36 ms  24 ms  24 ms  240e:f:b800:400::3
17  12  26 ms  26 ms  26 ms  240e:f:b800:c85::3
18  13  *     *     *     *     请求超时。
19  14  21 ms  23 ms  23 ms  240e:bf:b800:4300:1::15
20
21 跟踪完成。

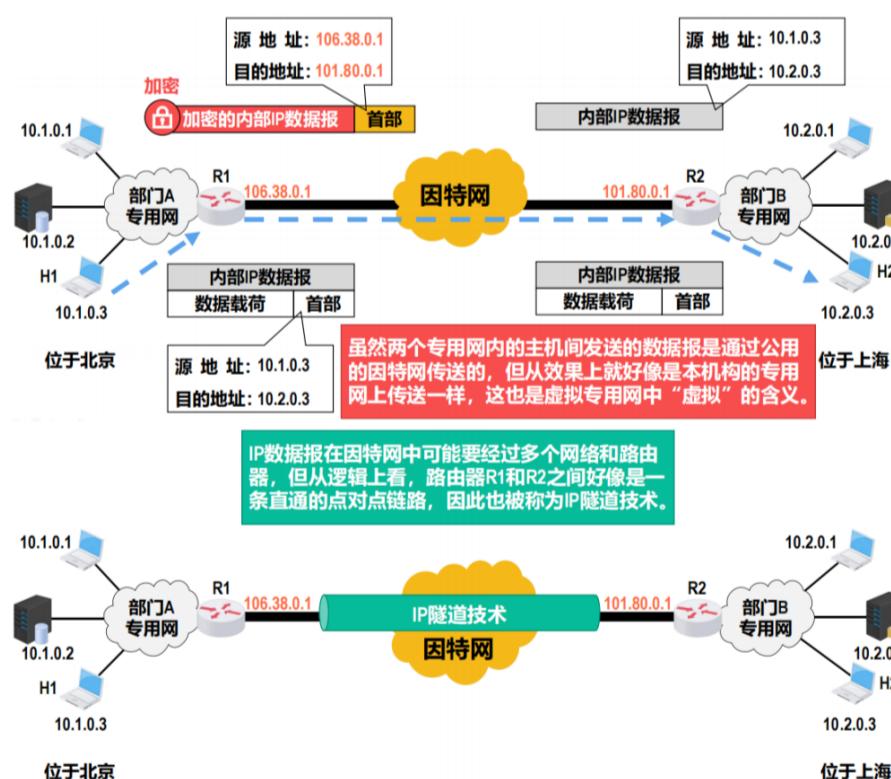
```



6. 虚拟专用网VPN和网络地址转换NAT

6.1. 虚拟专用网VPN

- 虚拟专用网（Virtual Private Network, VPN）利用公用的因特网作为本机构各专用网之间的通信载体，这样形成的网络又称为虚拟专用网。
- 给专用网内各主机配置的IP地址应该是该专用网所在机构可以自行分配的IP地址，这类IP地址仅在机构内部有效，称为专用地址（Private Address），不需要向因特网的管理机构申请。
- [RFC 1918]规定了以下三个CIDR地址块中的地址作为专用地址：
 - 10.0.0~10.255.255.255 (CIDR地址块10/8)
 - 172.16.0.0~172.31.255.255 (CIDR地址块172.16/12)
 - 192.168.0.0~192.168.255.255 (CIDR地址块192.168/16)
- 很显然，全世界可能有很多不同机构的专用网具有相同的专用IP地址，但这并不会引起麻烦，因为这些专用地址仅在机构内部使用。
- 在因特网中的所有路由器，对目的地址是专用地址的IP数据报一律不进行转发，这需要由因特网服务提供者ISP对其拥有的因特网路由器进行设置来实现。

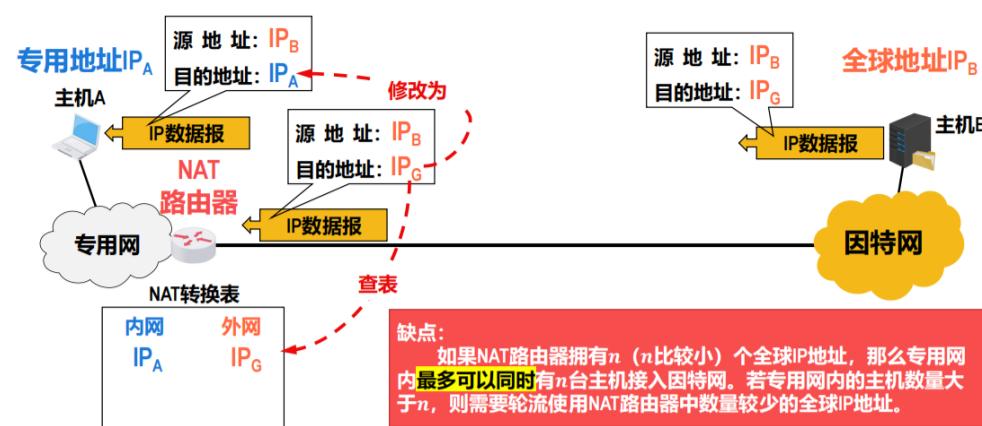


- 本例所示的是同一机构内不同部门的内部网络所构成的VPN，又称为内联网VPN。

6.2. 网络地址转换NAT

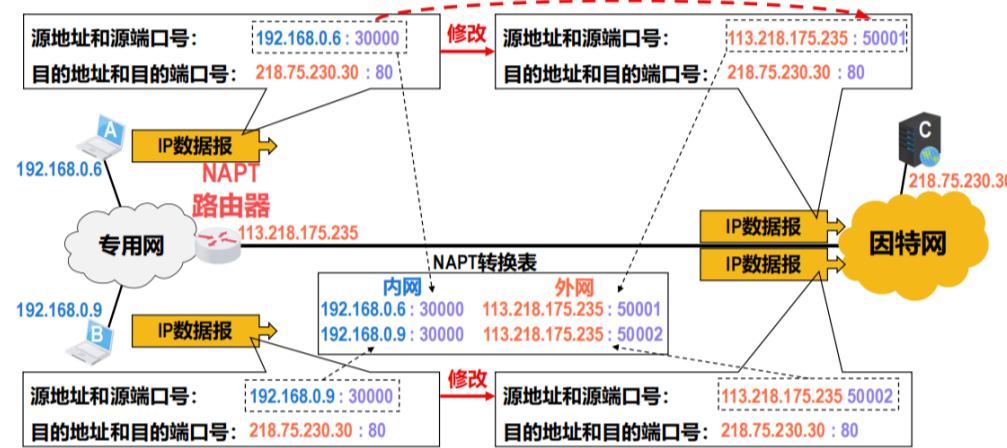
- 网络地址转换（Network Address Translation, NAT）技术于1994年被提出，用来缓解IPv4地址空间即将耗尽的问题。
 - NAT能使大量使用内部专用地址的专用网络用户共享少量外部全球地址来访问因特网上的主机和资源。
 - 这种方法需要在专用网络连接到因特网的路由器上安装NAT软件。
 - 装有NAT软件的路由器称为NAT路由器，它至少要有一个有效的外部全球地址 IP_G 。
 - 这样，所有使用内部专用地址的主机在和外部因特网通信时，都要在NAT路由器上将其内部专用地址转换成 IP_G 。

6.2.1. 最基本的NET方法

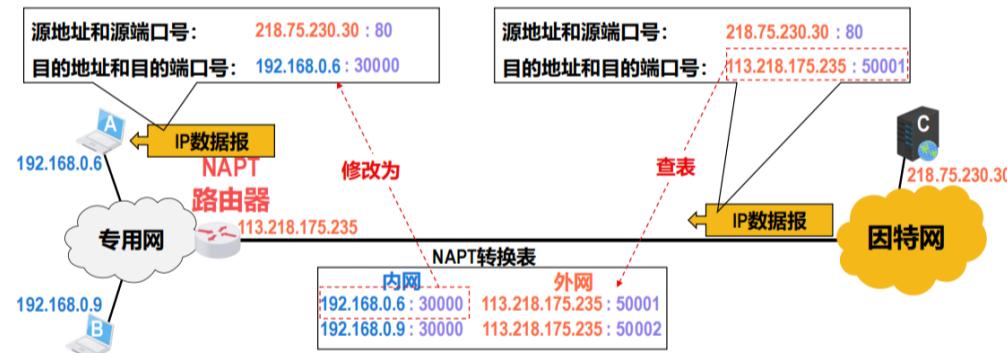


6.2.2. 网络地址与端口号转换方法 (NAPT)

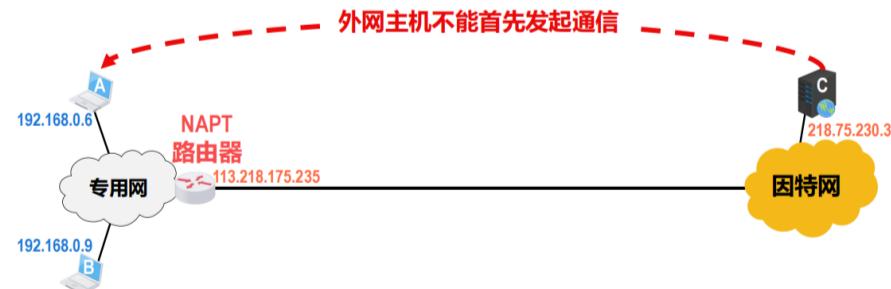
- 由于目前绝大多数基于TCP/IP协议栈的网络应用, 都使用运输层的传输控制协议TCP或用户数据报协议UDP, 为了更加有效地利用NAT路由器中的全球IP地址, 现在常将NAT转换和运输层端口号结合使用。
 - 这样就可以使内部专用网中使用专用地址的大量主机, 共用NAT路由器上的1个全球IP地址, 因而可以同时与因特网中的不同主机进行通信。
- 将NAT和运输层端口号结合使用, 称为**网络地址与端口号转换** (Network Address and Port Translation, NAPT)
 - 现在很多家用路由器将家中各种智能设备 (手机、平板、笔记本电脑、台式电脑、物联网设备等) 接入因特网, 这种路由器实际上就是一个NAPT路由器, 但往往并不运行路由选择协议。
- 主机向因特网发送
 - 与主机A选择的源端口号相同, 这纯属巧合 (端口号仅在本主机中才有意义)。特意这样举例, 就是为了能更好地说明NAPT路由器还会对源端口号重新动态分配。



- 因特网向主机发回



- 尽管NAT (和NAPT) 的出现在很大程度上缓解了IPv4地址资源紧张的局面, 但NAT (和NAPT) 对网络应用并不完全透明, 会对某些网络应用产生影响。
- NAT (和NAPT) 的一个重要特点就是通信必须由专用网内部发起, 因此拥有内部专用地址的主机不能直接充当因特网中的服务器。

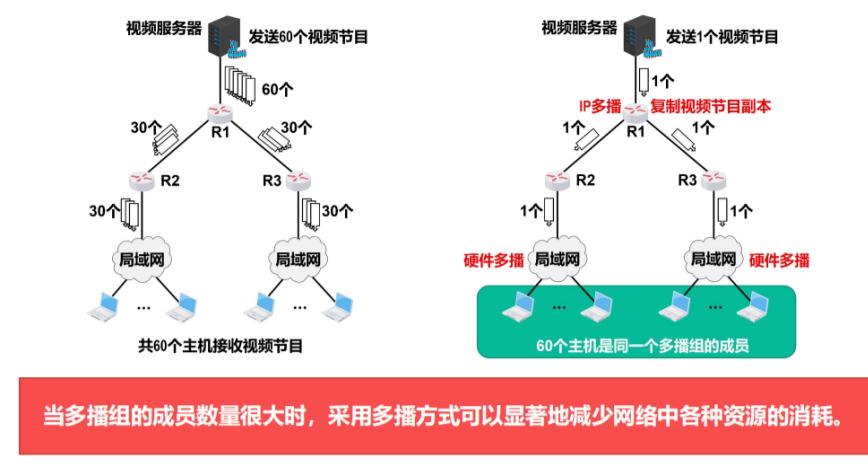


- 对于目前P2P这类需要外网主机主动与内网主机进行通信的网络应用, 在通过NAT时会遇到问题, 需要网络应用自身使用一些特殊的NAT穿透技术来解决。

7. IP多播技术

7.1. 相关基本概念

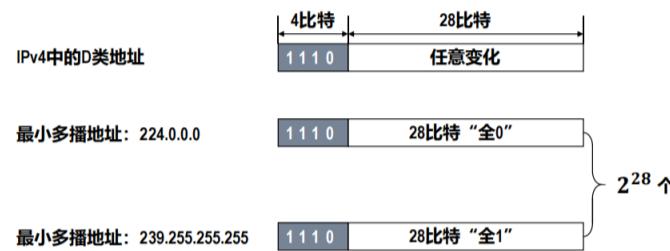
- 多播 (Multicast, 也称为组播) 是一种实现“一对多”通信的技术, 与传统单播“一对一”通信相比, 多播可以极大地节省网络资源。
 - 在因特网上进行的多播, 称为IP多播。



- 实现IP多播，则因特网中的路由器需要解决的问题
 - IP多播数据报的寻址问题
 - 多播路由选择问题

7.2. IP多播地址和多播组

- 在IPv4中，D类地址被作为多播地址。



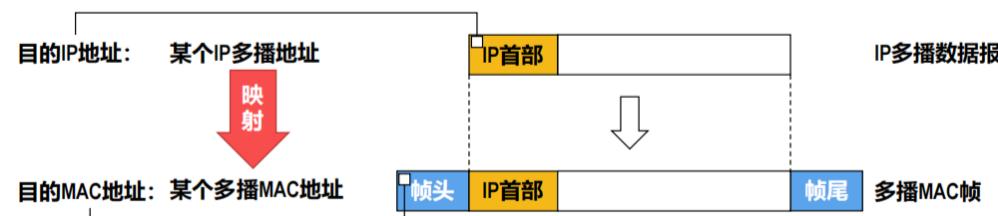
- 多播地址只能用作目的地址，而不能用作源地址。
- 用每一个D类地址来标识一个多播组，使用同一个IP多播地址接收IP多播数据报的所有主机就构成了一个多播组。
 - 每个多播组的成员是可以随时变动的，一台主机可以随时加入或离开多播组。
 - 多播组成员的数量和所在的地理位置也不受限制，一台主机可以属于几个多播组。
 - 非多播组成员也可以向多播组发送IP多播数据报
- 与IP数据报相同，IP多播数据报也是“尽最大努力交付”，不保证一定能够交付给多播组内的所有成员。
- IPv4多播地址又可分为预留的多播地址（永久多播地址）、全球范围可用的多播地址以及本地管理的多播地址[RFC 3330]。

224.0.0.0	基地址（保留）
224.0.0.1	仅在本子网上的所有参加多播的主机和路由器
224.0.0.2	仅在本子网上的所有参加多播的路由器
224.0.0.3	未指派
224.0.0.4	DVMRP路由器
224.0.0.5	OSPF路由器
.....	
永久多播地址	
224.0.1.0
238.255.255.255	全球范围内都可使用的多播地址
239.0.0.0
239.255.255.255	本地管理的多播地址，仅在特定的本地范围内有效

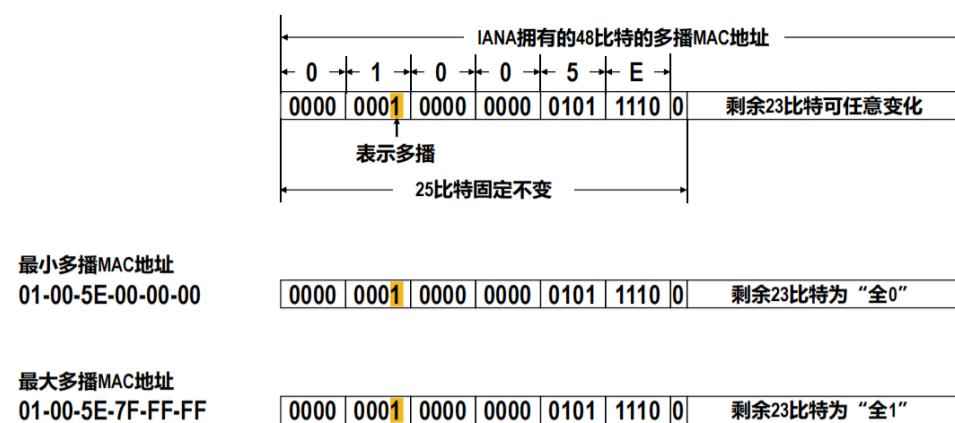
- IP多播可以分为以下两种
 - 只在本局域网上进行的硬件多播。
 - 在因特网上进行的多播。
- 目前大部分主机都是通过局域网接入因特网的。因此，在因特网上进行多播的最后阶段，还是要把IP多播数据报在局域网上用硬件多播交付给多播组的所有成员。

7.3. 在局域网上进行硬件多播

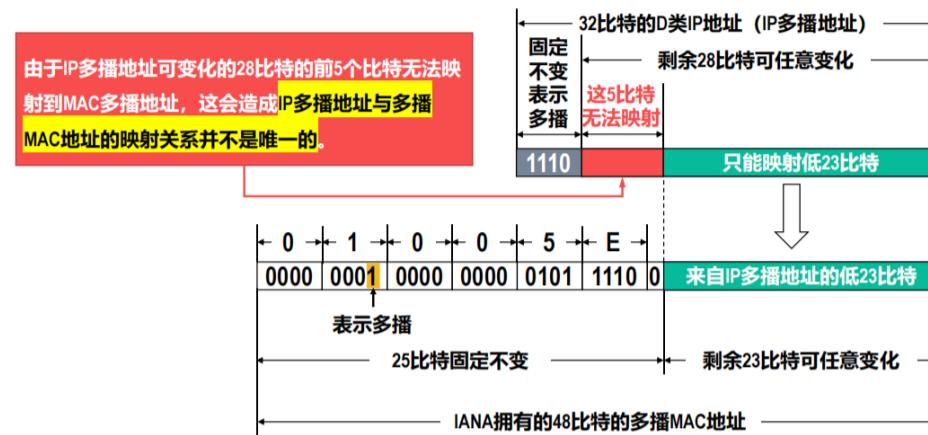
- 由于MAC地址（也称为硬件地址）有多播MAC地址这种类型，因此只要把IPv4多播地址映射成多播MAC地址，即可将IP多播数据报封装在局域网的MAC帧中，而MAC帧首部中的目的MAC地址字段的值，就设置为由IPv4多播地址映射成的多播MAC地址。这样，可以很方便地利用硬件多播来实现局域网内的IP多播。
- 当给某个多播组的成员主机配置其所属多播组的IP多播地址时，系统就会根据映射规则从该IP多播地址生成相应的局域网多播MAC地址。



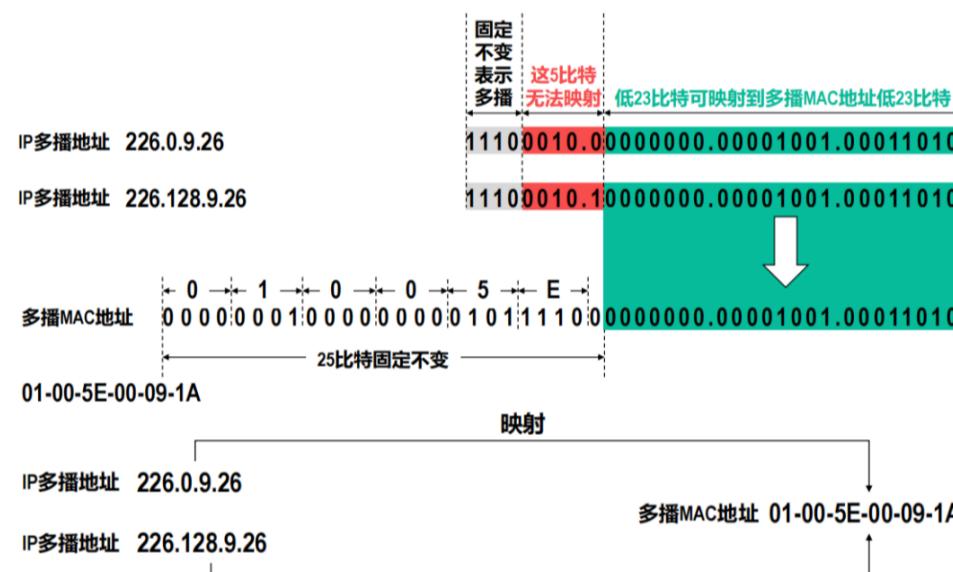
- 因特网号码指派管理局IANA，将自己从IEEE注册管理机构申请到的以太网MAC地址块中从01-00-5E-00-00-00到01-00-5E-7F-FF-FF的多播MAC地址，用于映射IPv4多播地址。
 - 这些多播MAC地址的左起前25个比特都是相同的，剩余23个比特可以任意变化，因此共有 2^{23} 。



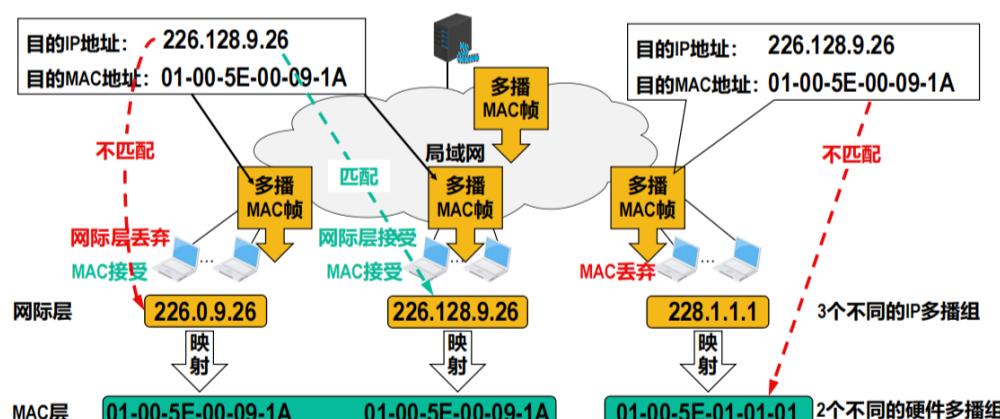
- IP多播地址与多播MAC地址的映射关系并不是唯一的



【举例】IP多播地址与多播MAC地址的映射关系并不是唯一的。

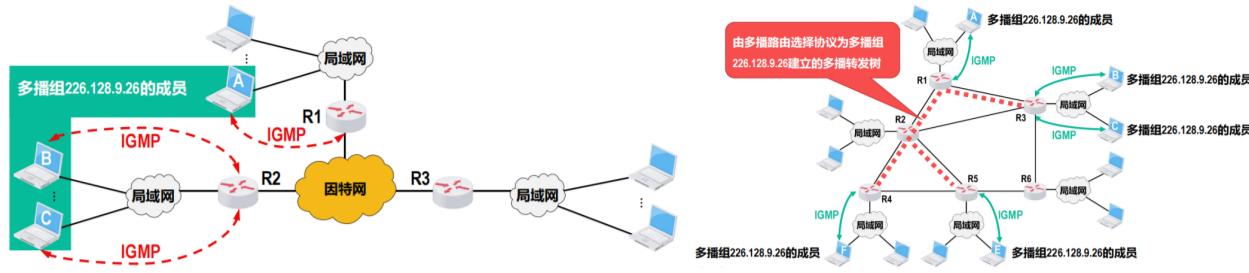


- 收到IP多播数据报的主机还要在网际层利用软件进行过滤，把不是主机要接收的IP多播数据报丢弃



7.4. 在因特网上进行IP多播需要的两种协议

网际组管理协议IGMP	多播路由选择协议
<ul style="list-style-type: none"> ■ 网际组管理协议 (Internet Group Management Protocol, IGMP) 是TCP/IP体系结构网际层中的协议，其作用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（实际上是主机中的某个进程）加入或退出了某个多播组。 ■ IGMP仅在本网络有效，使用IGMP并不能知道多播组所包含的成员数量，也不能知道多播组的成员都分布在哪些网络中。 ■ 仅使用IGMP并不能在因特网上进行IP多播。连接在局域网上的多播路由器还必须和因特网上的其他多播路由器协同工作，以便把IP多播数据报用最小的代价传送给所有的多播组成员，这就需要使用多播路由选择协议。 	<ul style="list-style-type: none"> ■ 多播路由选择协议的主要任务是：在多播路由器之间为每个多播组建立一个多播转发树。 □ 多播转发树连接多播源和所有拥有该多播组成员的路由器。 □ IP多播数据报只要沿着多播转发树进行洪泛，就能被传送到所有拥有该多播组成员的多播路由器。 □ 之后，在多播路由器所直连的局域网内，多播路由器通过硬件多播，将IP多播数据报发送给该多播组的所有成员。 ■ 针对不同的多播组需要维护不同的多播转发树，而且必须动态地适应多播组成员的变化，但此时网络拓扑并不一定发生变化，因此多播路由选择协议要比单播路由选择协议（例如RIP、OSPF等）复杂得多。 ■ 即使某个主机不是任何多播组的成员，它也可以向任何多播组发送多播数据报。 ■ 为了覆盖多播组的所有成员，多播转发树可能要经过一些没有多播组成员的路由器。



7.5. 网际组管理协议IGMP (封装在IP)

- 网际组管理协议IGMP目前的最新版本是2002年10月公布的IGMPv3[RFC 3376]。
- IGMP报文被封装在IP数据报中传送



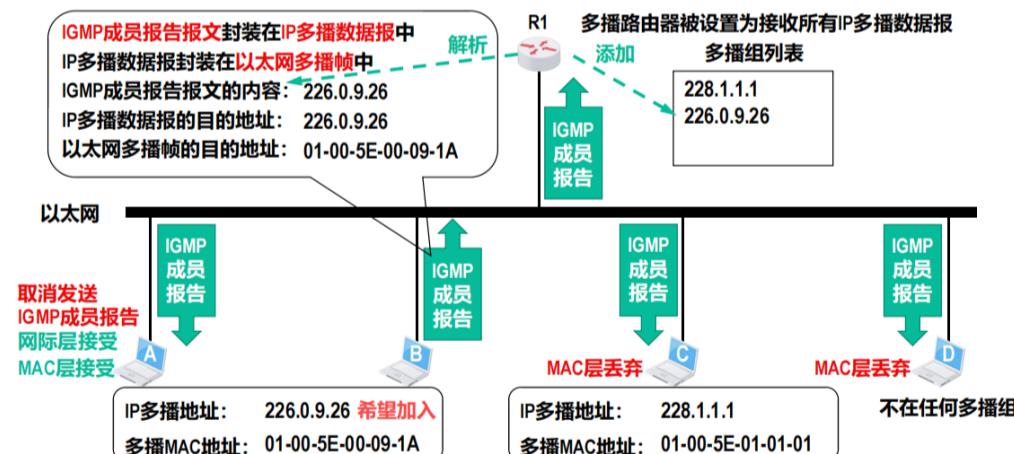
7.5.1. 三种报文类型

- 成员报告报文
- 成员查询报文
- 离开组报文

7.5.2. 基本工作原理

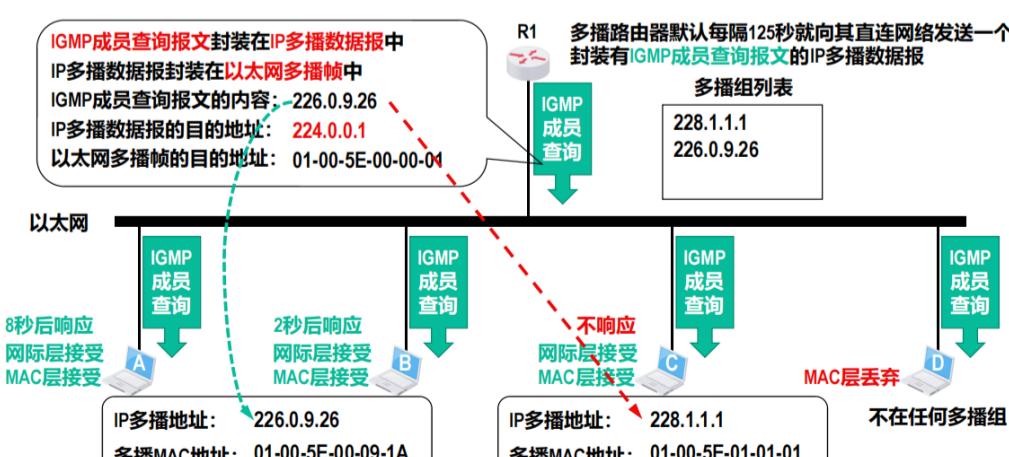


7.5.2.1. 加入多播组



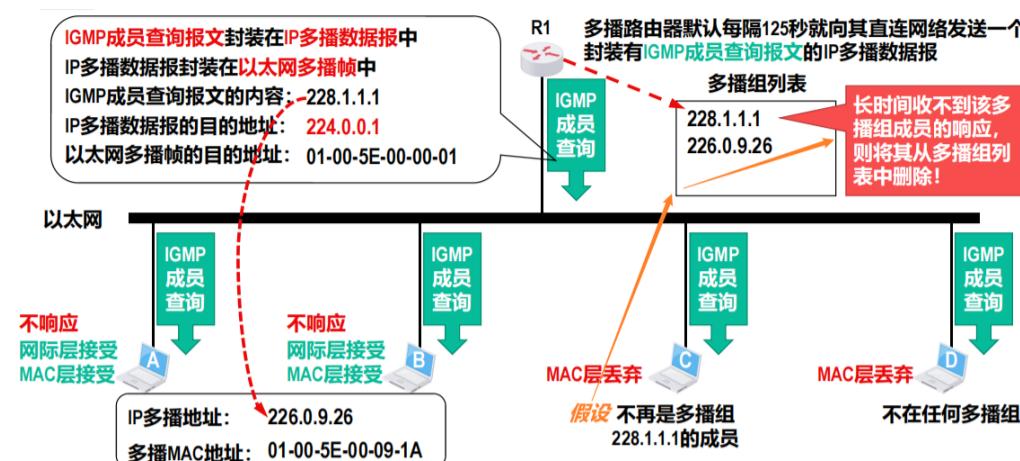
7.5.2.2. 监视多播组的成员变化

- 多播路由器默认每隔125秒就向其直连网络发送一个封装有IGMP成员查询报文的IP多播数据报。
- 成员查询报文中，224.0.0.1为特殊的IP多播地址，在本网络中所有参加多播的主机和路由器的网际层都会接受该多播数据报。



- 收到IGMP成员查询报文的被查询多播组的任何成员，都会发送IGMP成员报告文作为应答。
- 为了减少不必要的重复应答，每个多播组只需要一个成员应答就行了。

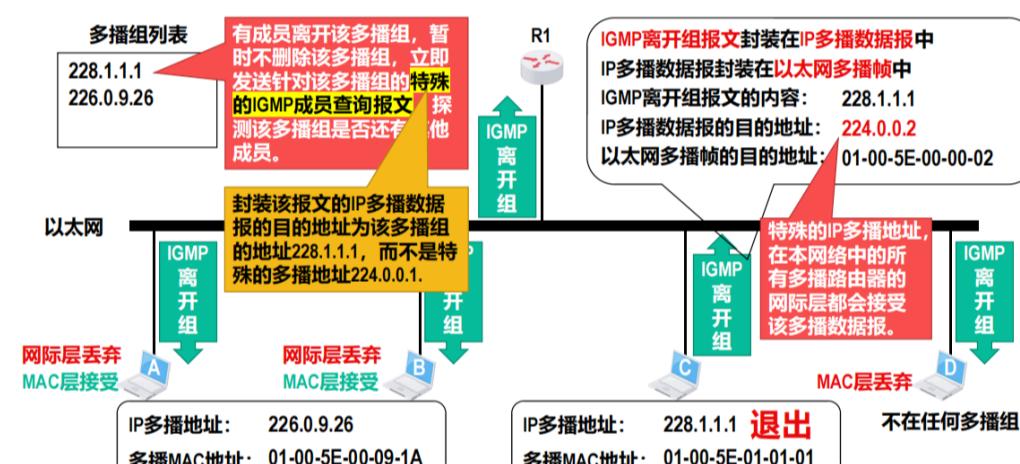
- 应答采用延迟响应的策略，主机收到查询报文后再1-10秒内等待一段随机时间后进行响应（不是立即响应）。
 - 在这段时间内如果收到了同组成员发送的成员报文报文，则取消响应。
 - 本例中B先发送成员报告报文，则A收到B的成员报告文后取消相应。
 - 路由器R1收到B的成员报告报文后对其进行解析并更新多播组列表。



- 同一网络中的多播路由器可能不止一个，但没有必要每个多播路由器都周期性地发送IGMP成员查询报文。
 - 只要在这些多播路由器中选择一个作为查询路由器，由查询路由器发送IGMP成员查询报文，而其他的多播路由器仅被动接收响应并更新自己的多播组列表即可。
 - 选择查询路由器的方法：
 - 每个多播路由器若监听到源IP地址比自己的IP地址小的IGMP成员查询报文则退出选举。
 - 最后，网络中只有IP地址最小的多播路由器成为查询路由器。

7.5.2.3. 退出多播组

- IGMPv2在IGMPv1的基础上增加了一个可选项：当主机要退出某个组时，可主动发送一个离开组报文而不必等待多播路由器的查询。这样可使多播路由器能够更快地发现某个组有成员离开。



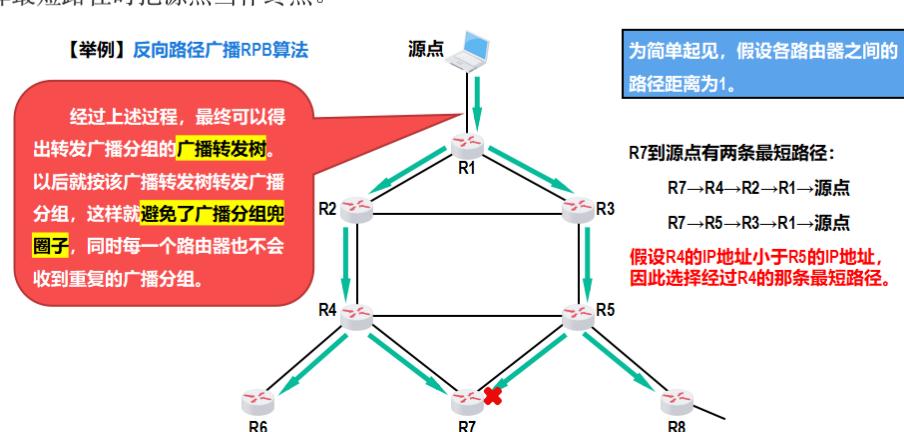
7.6. 多播路由选择协议

7.6.1. 多播路由选择协议

- 多播路由选择协议的主要任务是：在多播路由器之间为每个多播组建立一个多播转发树。
 - 多播转发树连接多播源和所拥有该多播组成员的路由器。
 - 建多播转发树的方法
 - 基于源树（Source-Base Tree）多播路由选择
 - 组共享树（Group-Shared Tree）多播路由选择

7.6.1.1 基于源树多播路由选择

- 反向路径多播算法 (Reverse Path Multicasting, RPM)
 - 利用反向路径广播 (Reverse Path Broadcasting, RPB) 算法建立一个广播转发树。
 - 每一台路由器在收到一个广播分组时，先检查该广播分组是否是从源点经最短路径传送来的。
 - 若是，本路由器就从自己除刚才接收该广播分组的接口的所有其他接口转发该广播分组。
 - 否则，丢弃该广播分组。
 - 如果本路由器有好几个邻居路由器都处在到源点的最短路径上，也就是存在好几条同样长度的最短路径，那么只能选取一条最短路径。选取的规则是这几条最短路径中的邻居路由器的IP地址最小的那条最短路径。
 - RPB中“反向路径”的意思是：在计算最短路径时把源点当作终点。

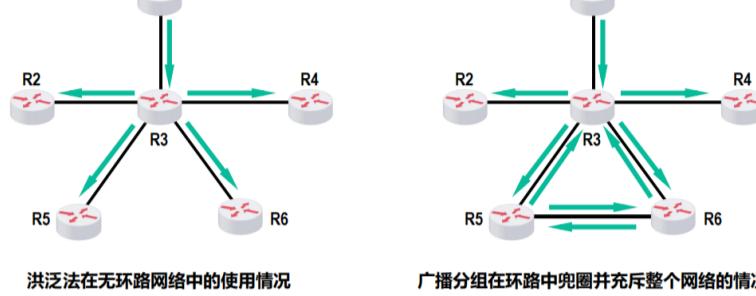




- 若被剪枝的路由器通过IGMP又发现了新的多播组成员，则会向上游路由器发送一个嫁接报文，并重新加入到多播转发树中。
- 尽管R2没有多播组成员，但也要保留R2以确保多播转发树的连通性。

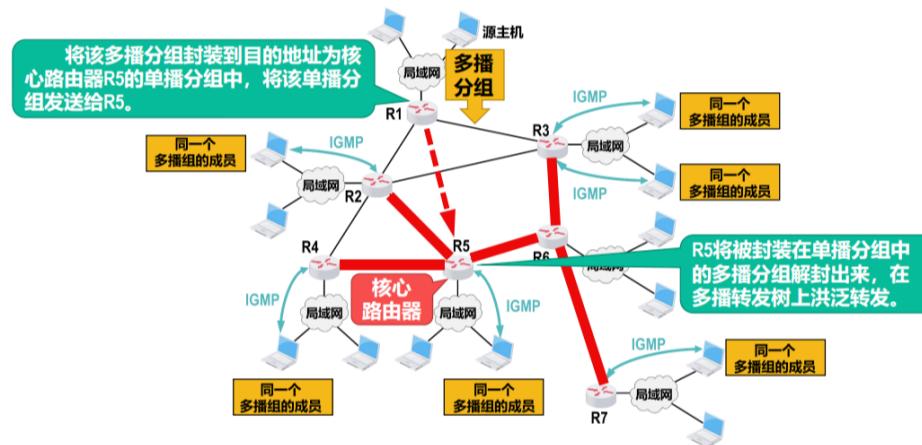
- 要建立广播转发树，可以使用洪泛（Flooding）法。

- 利用反向路径广播RPB算法生成的广播转发树，不会存在环路，因此可以避免广播分组在环路中兜圈。



7.6.1.2. 组共享树多播路由选择

- 组共享树多播路由选择采用基于核心的分布式生成树算法来建立共享树。
 - 该方法在每个多播组中指定一个核心（core）路由器，以该路由器为根，建立一棵连接多播组的所有成员路由器的生成树，作为多播转发树。
- 每个多播组中除了核心路由器，其他所有成员路由器都会向自己多播组中的核心路由器单播加入报文。
 - 加入报文通过单播朝着核心路由器转发，直到它到达已经属于该多播生成树的某个节点或者直接到达该核心路由器。
 - 加入报文所经过的路径，就确定了一条从单播该报文的边缘节点到核心路由器之间的分支，而这个新分支就被嫁接到现有的多播转发树上。



7.6.2. 因特网的多播路由选择协议

■ 目前还没有在整个因特网范围使用的多播路由选择协议。下面是一些建议使用的多播路由选择协议：

- 距离向量多播路由选择协议 (Distance Vector Multicast Routing Protocol, DVMRP) [RFC 1075]。
- 开放最短路径优先的多播扩展 (Multicast Extensions to OSPF, MOSPF) [RFC 1585]。
- 协议无关多播-稀疏方式 (Protocol Independent Multicast-Sparse Mode, PIM-SM) [RFC 2362]。
- 协议无关多播-密集方式 (Protocol Independent Multicast-Dense Mode, PIM-DM) [RFC 3973]。
- 基于核心的转发树 (Core Based Tree, CBT) [RFC 2189, RFC 2201]。

■ 尽管因特网工程任务组IETF努力推动着因特网上的全球多播主干网 (Multicast Backbone On the Internet, MBONE) 的建设，但至今在因特网上的IP多播还没有得到大规模的应用。

- 主要原因是：改变一个已成功运行且广泛部署的网络层协议是一件及其困难的事情。
- 目前IP多播主要应用在一些局部的园区网络、专用网络或者虚拟专用网中。

另外，P2P技术的广泛应用推动了应用层多播技术的发展，许多视频流公司和内容分发公司，通过构建自己的应用层多播覆盖网络来分发它们的内容。但上述多播路由选择协议的算法思想在应用层多播中依然适用。

8. 移动IP技术概述

8.1. 移动性对因特网应用的影响

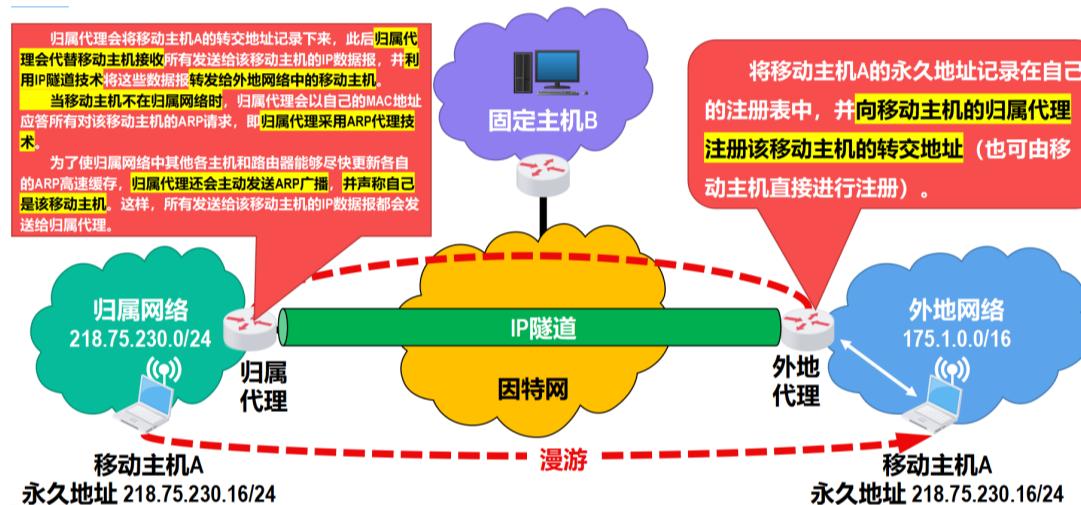


8.2. 移动IP技术的相关基本概念

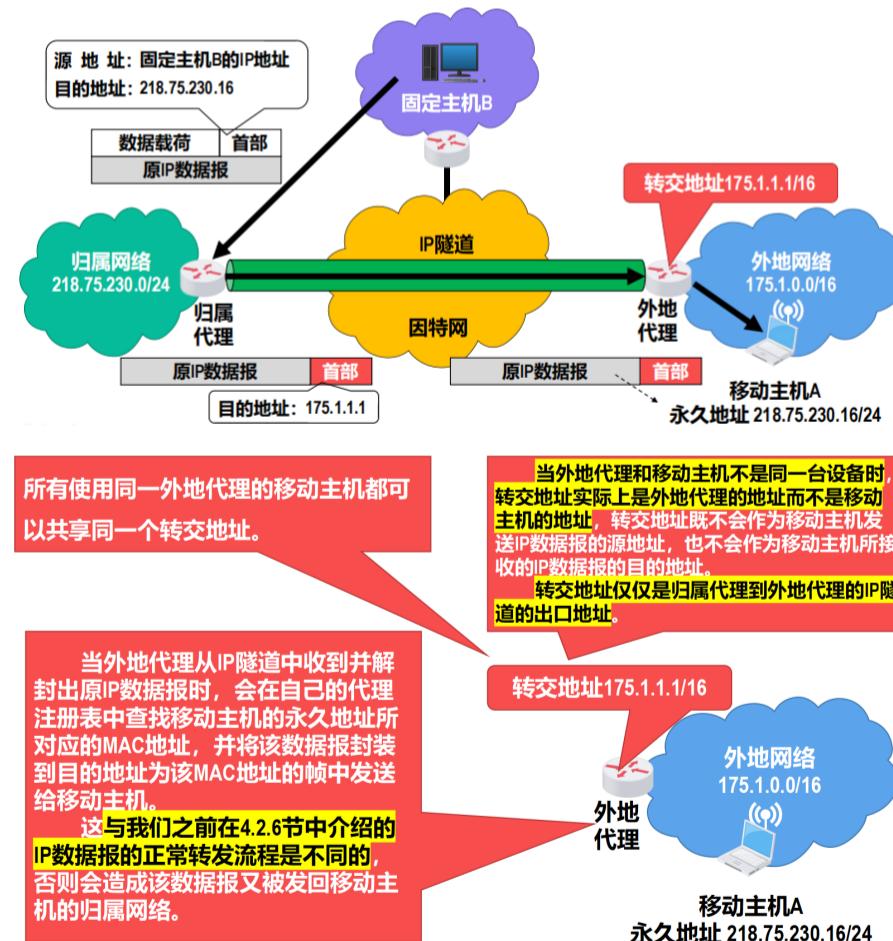
- 移动IP (Mobile IP) 是因特网工程任务组IETF开发的一种技术[RFC 3344]，该技术使得移动主机在各网络之间漫游时，仍然能够保持其原来的IP地址不变。
- 移动IP技术还为因特网中的非移动主机提供了相应机制，使得它们能够将IP数据报正确发送到移动主机。
- 基本概念
 - 归属网络 (Home Network)
 - 每个移动主机都有一个默认连接的网络或初始申请接入的网络。
 - 永久地址 (Permanent Address) 或归属地址 (Home Address) 。
 - 移动主机在归属网络中的IP地址 (在其整个移动通信过程中是始终不变的)
 - 归属代理 (Home Agent)
 - 在归属网络中，代表移动主机执行移动管理功能的实体。
 - 归属代理通常就是连接在归属网络上的路由器，然而它作为代理的特定功能则是在网络层完成的。
 - 外地网络 (Foreign Network) 或被访网络 (Visited Network)
 - 移动主机当前漫游所在的网络。
 - 外地代理 (Foreign Agent)
 - 在外地网络中，帮助移动主机执行移动管理功能的实体。
 - 外地代理通常就是连接在外地网络上的路由器。
 - 转交地址 (Care-of Address)
 - 外地代理会为移动主机提供一个临时使用的属于外地网络的转交地址。

8.3. 移动IP技术的基本工作原理

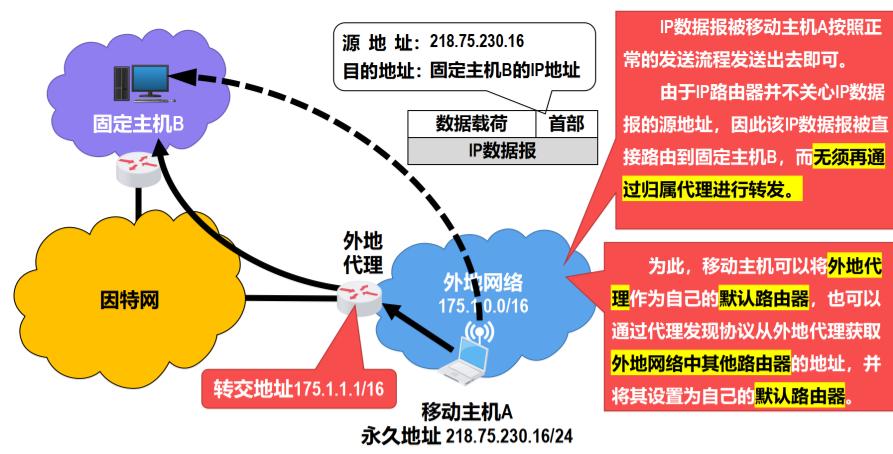
8.3.1. 代理发现与注册



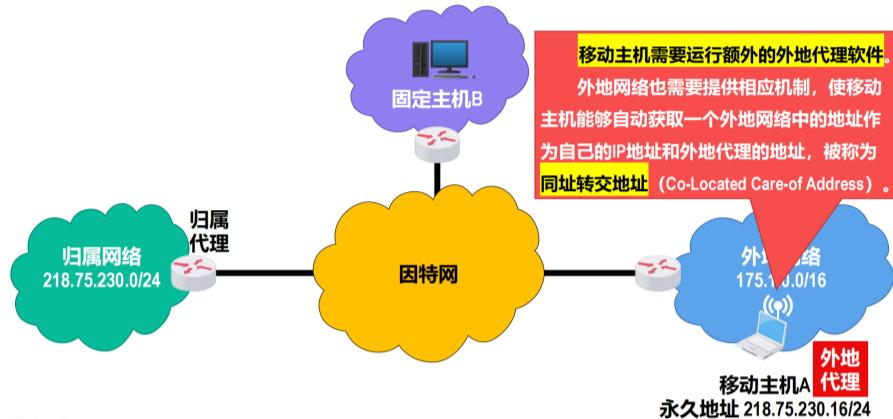
8.3.2. 固定主机向移动主机发送IP数据报



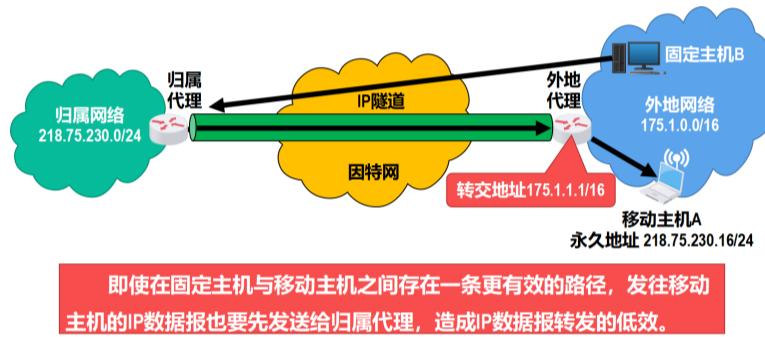
8.3.3. 移动主机向固定主机发送IP数据报



8.3.4. 同址转交地址方式



8.3.5. 三角形路由问题



- 解决三角形路由问题的一种方法

- 给固定主机配置一个通信代理，固定主机发送给移动主机的IP数据报，都要通过该通信代理转发。
- 通信代理先从归属代理获取移动主机的转交地址，之后所有发送给移动主机的IP数据报，都利用转交地址直接通过IP隧道发送给移动主机的外地代理，而无须再通过移动主机的归属代理进行转发。
- 这种方法以增加复杂性为代价，并要求固定主机也要配置通信代理，也就是对固定主机不再透明。

9. IPv6

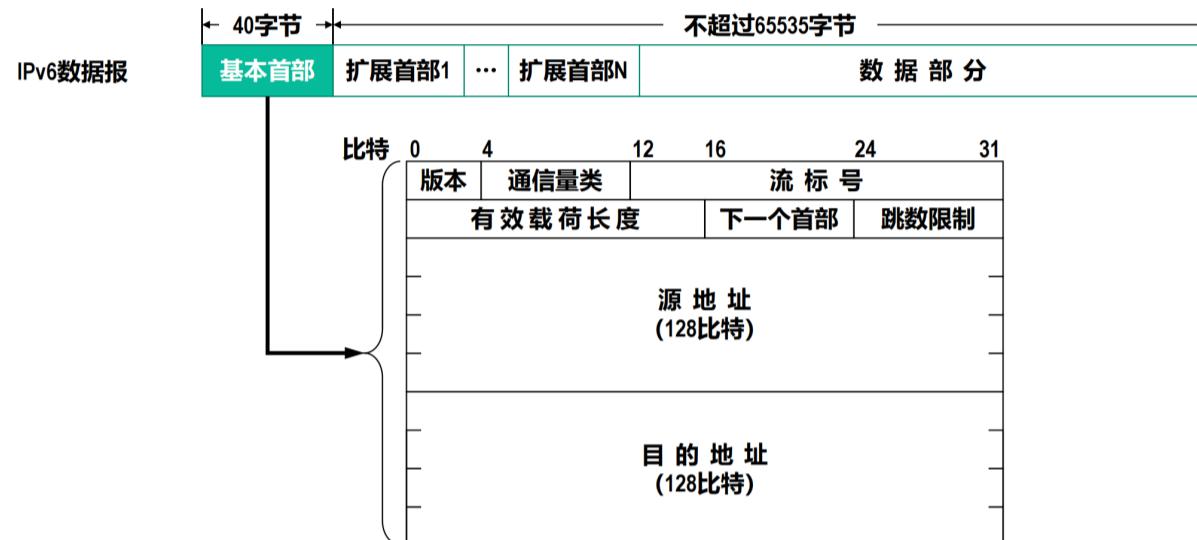
9.1. IPv6的诞生背景

- IPv4地址存在缺陷
 - IPv4地址的长度仅为32比特
 - 早期的编址方法不够合理，造成IPv4地址资源的浪费。
- 2011年2月3日，因特网号码分配管理局IANA宣布IPv4地址已经分配完毕
- 如果没有网络地址转换NAT技术的广泛应用，IPv4早已停止发展。
- 解决IPv4地址耗尽的根本措施就是采用具有更大地址空间（IP地址的长度为128比特）的新版本IP，即IPv6。
- 到目前为止，IPv6还只是草案标准阶段
- 尽早开始过渡到IPv6的好处
 - 有更多时间来平滑过渡
 - 有更多时间来培养IPv6的专业人才
 - 及早提供IPv6服务比较便宜

9.2. IPv6引进的主要变化

更大的地址空间	IPv6将IPv4的 32比特 地址空间增大到了 128比特 ，在采用合理编址方法的情况下，在可预见的未来是不会用完的。
扩展的地址层次结构	可划分为更多的层次，这样可以更好地反映出因特网的拓扑结构，使得对寻址和路由层次的设计更具有灵活性。
灵活的首部格式	与IPv4首部并不兼容。IPv6定义了许多 可选的的扩展首部 ，不仅可提供比IPv4更多的功能，而且还可以 提高路由器的处理效率 ，因为路由器对逐跳扩展首部外的其他扩展首部都不进行处理。
改进的选项	IPv6允许分组 包含有选项的控制信息 ，因而 可以包含一些新的选项 。然而IPv4规定的选项却是固定不变的。
允许协议继续扩充	这一点很重要，因为技术总是在不断地发展，而新的应用也会层出不穷。然而IPv4的功能却是固定不变的。
支持即插即用 (即自动配置)	IPv6支持主机或路由器自动配置IPv6地址及其他网络配置参数。因此 IPv6不需要使用DHCP 。
支持资源的预分配	IPv6能为实时音视频等要求保证一定带宽和时延的应用，提供 更好的服务质量保证 。

9.3. IPv6数据包的基本首部



9.4. IPv6数据包的扩展首部

- IPv4数据报如果在其首部中使用了选项字段，则在数据报的整个传送路径中的全部路由器，都要对选项字段进行检查，这就**降低了路由器处理数据报的速度**。
- 实际上，在路径中的路由器对很多选项是不需要检查的。因此，为了提高路由器对数据包的处理效率，IPv6把原来IPv4首部中的选项字段都放在了扩展首部中，由路径两端的源点和终点的主机来处理，而数据报传送路径中的所有路由器都不处理这些扩展首部（除逐跳选项扩展首部）。
- 在[RFC 2460]中定义了以下六种扩展首部：

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

- 每一个扩展首部都由若干个字段组成，它们的长度也各不相同。
- 所有扩展首部中的第一个字段都是8比特的**下一个首部**字段。该字段的值指出在该扩展首部后面是何种扩展首部。
- 当使用多个扩展首部时，应按以上的先后顺序出现。

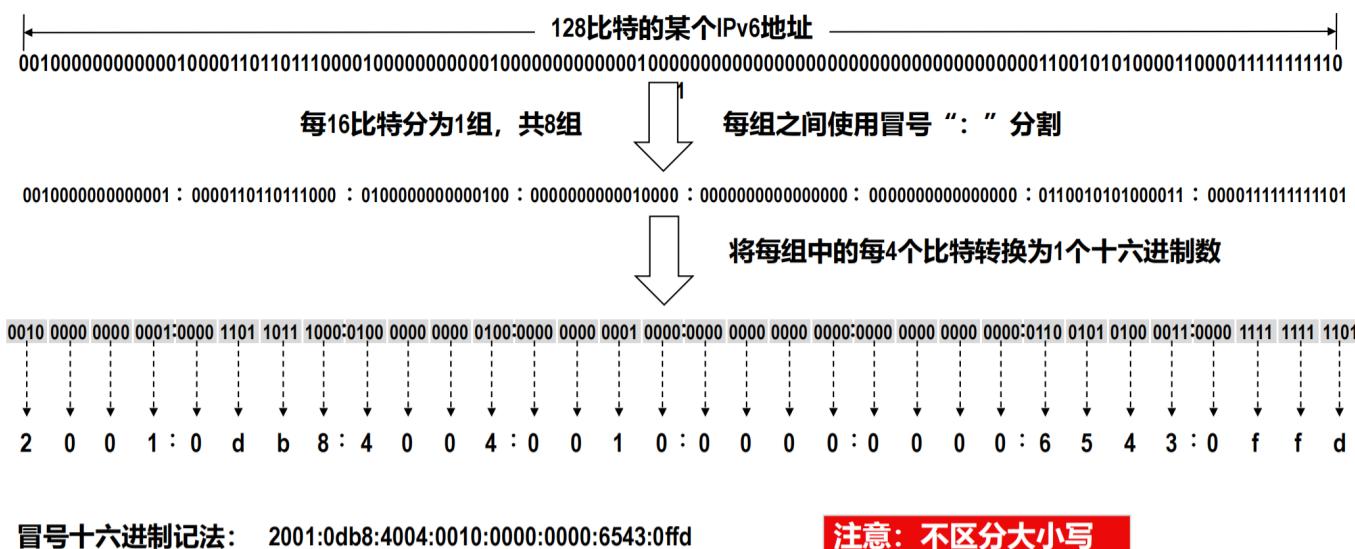
9.5. IPv6地址

9.5.1. IPv6地址空间大小

- 在IPv6中，每个地址占**128个比特**。

IPv6地址空间大小为 2^{128} (大于 3.4×10^{38})

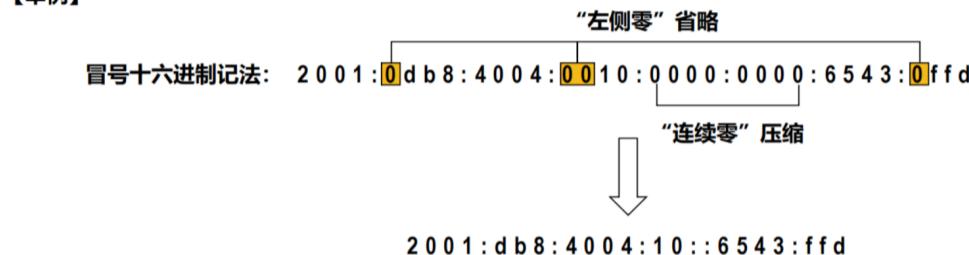
9.5.2. IPv6地址的表示方法



■ 在IPv6地址的冒号十六进制记法的基础上，再使用“左侧零”省略和“连续零”压缩，可使IPv6地址的表示更加简洁。

- 在IPv6地址的冒号十六进制记法的基础上，再使用“左侧零”省略和“连续零”压缩更加简洁。
 - “左侧零”省略是指两个冒号间的十六进制数中最前面的一串0可以省略不写。
 - “连续零”压缩是指一连串连续的0可以用一对冒号取代。

【舉例】



■ 在一个IPv6地址中只能使用一次“连续零”压缩，否则会导致歧义。

【举例】

2001:0000:0000:abcd:0000:0000:0000:1234	对每个地址进行多次“连续零”压缩 得到同一个有歧义的地址	2001::abcd::1234
2001:0000:0000:0000:abcd:0000:0000:1234		2001::abcd::1234
2001:0000:abcd:0000:0000:0000:0000:1234		2001::abcd::1234
2001:0000:0000:0000:0000:abcd:0000:1234		2001::abcd::1234
2001:0000:0000:abcd:0000:0000:0000:1234	只使用一次“连续零”压缩，并使用“左侧零”省略	2001:0:0:abcd::1234
2001:0000:0000:0000:abcd:0000:0000:1234		2001::abcd:0:0:1234
2001:0000:abcd:0000:0000:0000:0000:1234		2001:0:abcd::1234
2001:0000:0000:0000:0000:abcd:0000:1234		2001::abcd:0:1234

■ 冒号十六进制记法还可结合点分十进制的后缀。这在IPv4向IPv6过渡阶段非常有用。

【举例】



■ CIDR的斜线表示法在IPv6中仍然可用

【举例】

2001:0db8:0000:cd30:0000:0000:0000/60
↓
“左侧零”省略，“连续零”压缩
2001:db8:0:cd30::/60

9.5.3 IPv6地址的分类

■ IPv6数据报的目的地址有三种基本类型：

单播
(unicast)

传统的点对点通信

多播
(multicast)

语，而将广播看作多播的一个特例。

任播
(anycast)

这是IPv6新增的一种类型。任播的终点是一组计算机，但数据报只交付其中的一个，通常是按照路由算法得出的距离最近的一个。

未指明地址	128个比特为“全0”的地址，可缩写为两个冒号“::”。 该地址不能用作目的地址，只能用于还没有配置到一个标准IPv6地址的主机用作源地址。 未指明地址仅有一个。
环回地址	最低比特为1，其余127个比特为“全0”，即0:0:0:0:0:0:1，可缩写为::1。 该地址的作用与IPv4的环回地址相同。 IPv6的环回地址只有一个。
多播地址	最高8比特为“全1”的地址，可记为FF00::/8。 IPv6多播地址的功能与IPv4多播地址相同。 这类地址占IPv6地址空间的1/256。
本地链路单播地址	最高10比特为11111110的地址，可记为FE80::/10。即使用户网络没有连接到因特网，但仍然可以使用TCP/IP协议。连接在这种网络上的主机都可以使用本地链路单播地址进行通信，但不能和因特网上的其他主机通信。这类地址占IPv6地址空间的1/1024。
全球单播地址	全球单播地址是使用得最多的一类地址。 IPv6全球单播地址采用三级结构，这是为了使路由器可以更快地查找路由。

全局路由选择前缀 48比特 子网标识符 16比特 接口标识符 64比特

分配给公司和机构，用于因特网中路由器的路由选择，相当于IPv4分类地址中的网络号。

用于各公司和机构构建自己的子网。

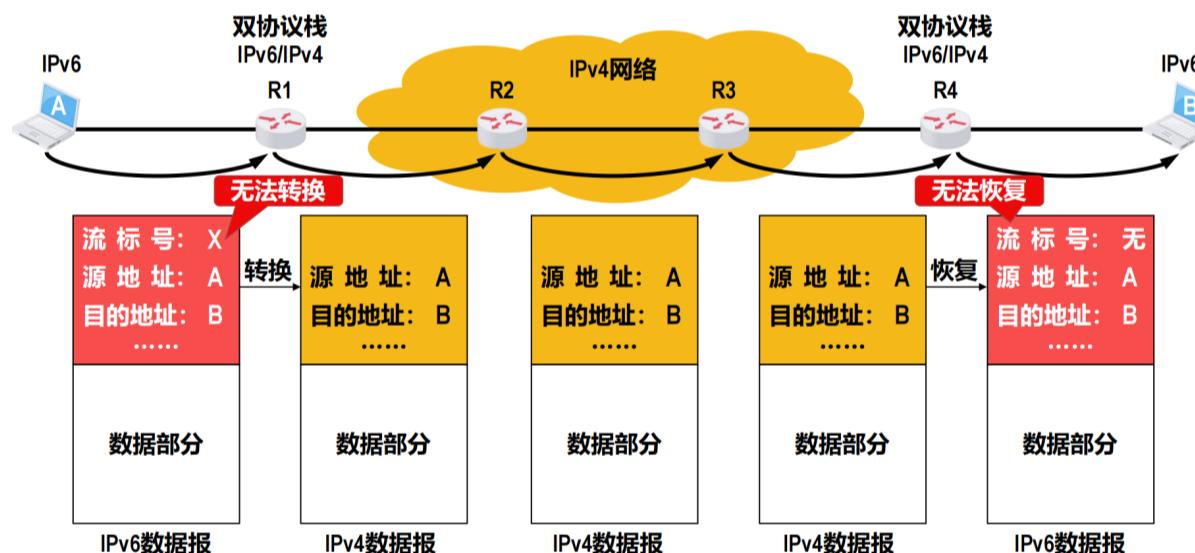
用于指明主机或路由器的单个网络接口，相当于IPv4分类地址中的主机号。有64个比特，足以将各种接口的硬件地址直接进行编码，这样就不需要使用ARP。

9.6. 从IPv4向IPv6过渡

- 因特网上使用IPv4的路由器的数量太大，要让所有路由器都改用IPv6并不能一蹴而就。因此，从IPv4转变到IPv6只能采用逐步演进的办法。
- 另外，新部署的IPv6系统必须能够向后兼容，也就是IPv6系统必须能够接收和转发IPv4数据报，并且能够为IPv4数据报选择路由。

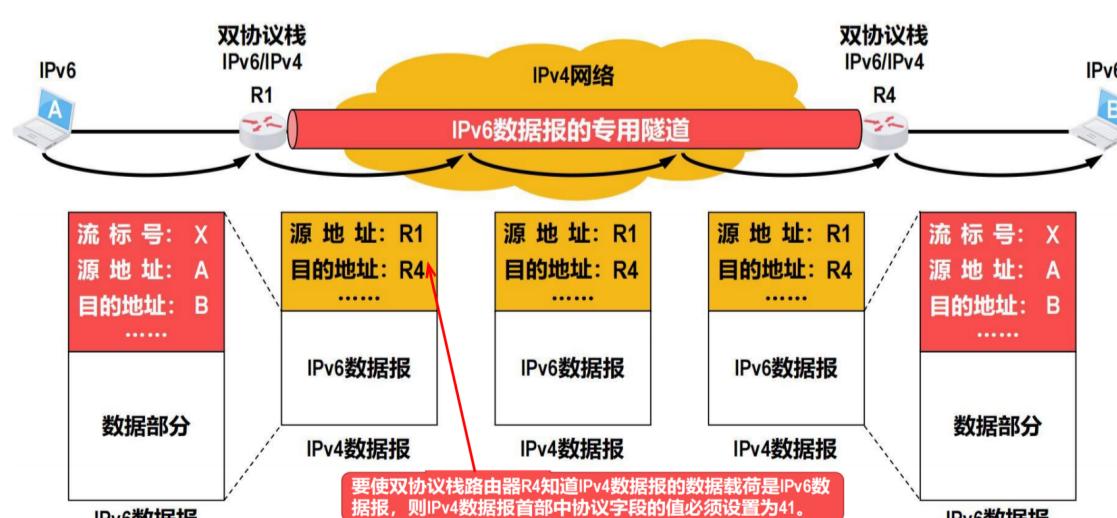
9.6.1. 使用双协议栈

- 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IPv4和IPv6两套协议栈。
- 双协议栈主机或路由器既可以和IPv6系统通信，又可以和IPv4系统通信。
- 双协议栈主机或路由器记为IPv6/IPv4，表明它具有一个IPv6地址和一个IPv4地址。
 - 双协议栈主机在与IPv6主机通信时采用IPv6地址，而与IPv4主机通信时采用IPv4地址
 - 双协议栈主机通过域名系统DNS查询目的主机采用的IP地址：
 - 若DNS返回的是IPv4地址，则双协议栈的源主机就使用IPv4地址
 - 若DNS返回的是IPv6地址，则双协议栈的源主机就使用IPv6地址



9.6.2. 使用隧道技术

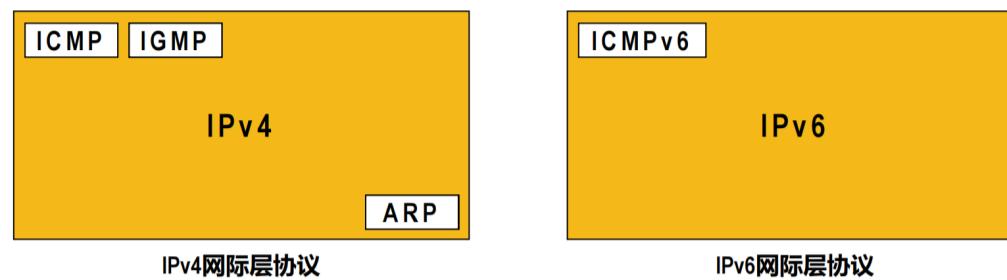
- 隧道技术（Tunneling）的核心思想是：
 - 当IPv6数据报要进入IPv4网络时，将IPv6数据报重新封装成IPv4数据报，即整个IPv6数据报成为IPv4数据报的数据载荷。
 - 封装有IPv6数据报的IPv4数据报在IPv4网络中传输。
 - 当IPv4数据报要离开IPv4网络时，再将其数据载荷（即原来的IPv6数据报）取出并转发到IPv6网络。



9.7. 网际控制报文协议ICMPv6

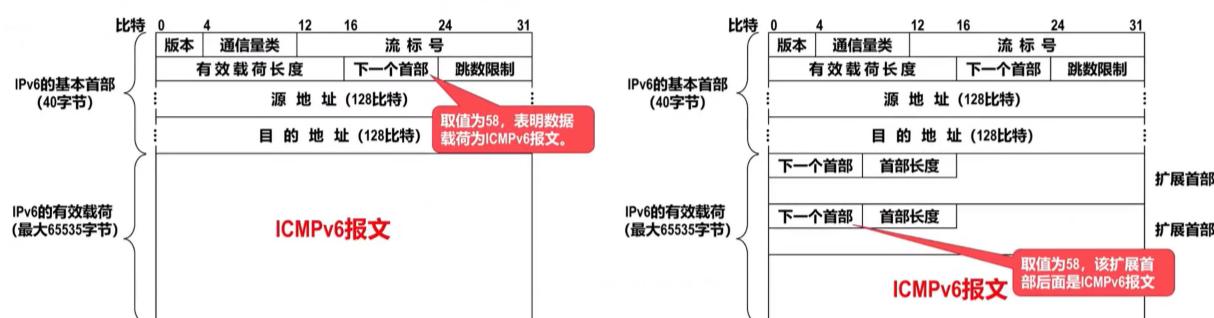
9.7.1. 概述

- 由于IPv6与IPv4一样，都不确保数据报的可靠交付，因此IPv6也需要使用网际控制报文协议ICMP来向发送IPv6数据报的源主机反馈一些差错信息，相应的ICMP版本为ICMPv6。
- ICMPv6比ICMPv4要复杂得多，它合并了原来的地址解析协议ARP和网际组管理协议IGMP的功能。因此与IPv6配套使用的网际层协议就只有ICMPv6这一个协议。



9.7.2. ICMPv6报文的封装

- ICMPv6报文需要封装成IPv6数据报进行发送。



9.7.3. ICMPv6报文的分类

- ICMPv6报文可被用来报告差错、获取信息、探测邻站或管理多播通信。
- 在对ICMPv6报文进行分类时，不同的RFC文档使用了不同的策略：
 - 在[RFC 2463]中定义了六种类型的ICMPv6报文
 - 在[RFC 2461]中定义了五种类型的ICMPv6报文
 - 在[RFC 2710]中定义了三种类型的ICMPv6报文

常用的几种ICMPv6报文		
ICMP报文种类	类型的值	ICMP报文的类型
差错报告报文	1	目的站不可达
	2	分组太长
	3	时间超过
	4	参数问题
回送请求与回答报文	128	回送请求
	129	回送回答
多播听众发现报文	130	多播听众查询
	131	多播听众报告
	132	多播听众完成
邻站发现报文	133	路由器询问
	134	路由器通告
	135	邻站询问
	136	邻站通告
	137	改变路由

替代原来的IGMP协议
替代原来的ARP协议

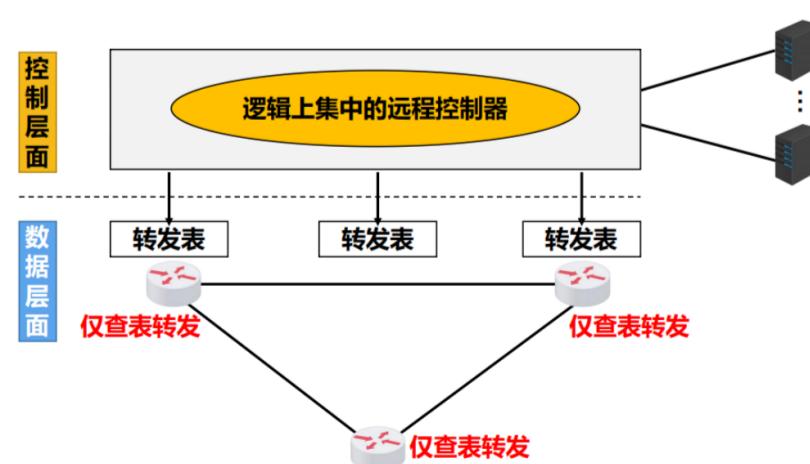
10. 软件定义网络SDN

10.1. 概述

- 软件定义网络（Software Defined Network, SDN）的概念最早由斯坦福大学的Nick McKeown教授于2009年提出。
- SDN最初只是学术界讨论的一种新型网络体系结构。
- SDN成功案例：谷歌于2010~2012年间建立的数据中心网络B4。
- SDN是当前网络领域最热门和最具发展前途的技术之一，成为近年来的研究热点。

10.2. 网络层的数据层面和控制层面

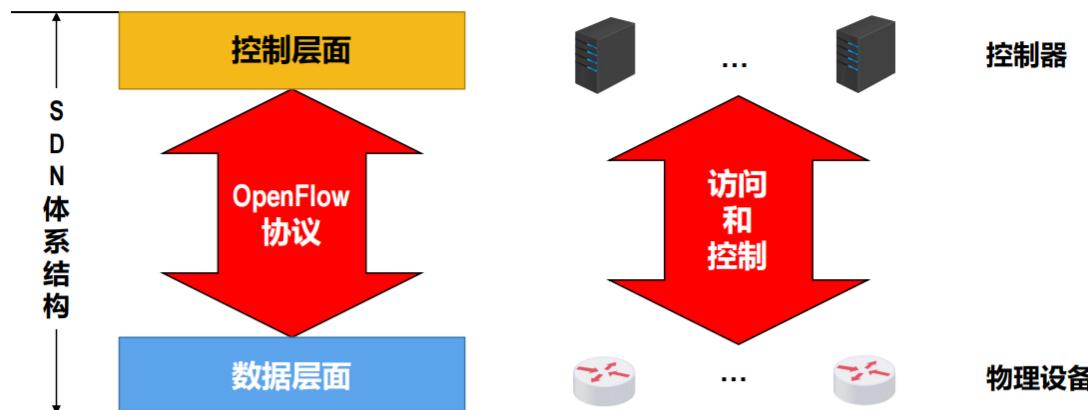
- 在SDN体系结构中，路由器中的路由软件都不存在了。因此，路由器之间不再交换路由信息。
- 在控制层面中，有一个在逻辑上集中的远程控制器。
- 逻辑上集中的远程控制器在物理上可由不同地点的多个服务器组成。
- 远程控制器掌握各主机和整个网络的状态。
- 远程控制器能够为每一个分组计算出最佳的路由。
- 远程控制器为每一个路由器生成其正确的转发表。
- SDN这种新型网络体系结构的核心思想：把网络的控制层面和数据层面分离，而让控制层面利用软件来控制数据层面中的许多设备。



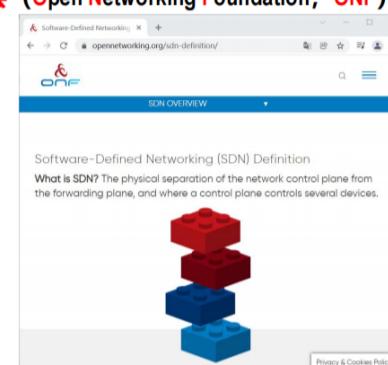
10.3. OpenFlow协议

10.3.1. 概述

- OpenFlow协议是一个得到高度认可的标准，在讨论SDN时往往与OpenFlow一起讨论。
- OpenFlow协议可被看成是SDN体系结构中控制层面与数据层面之间的通信接口。
- OpenFlow协议使得控制层面的控制器可以对数据层面中的物理设备进行直接访问和控制。

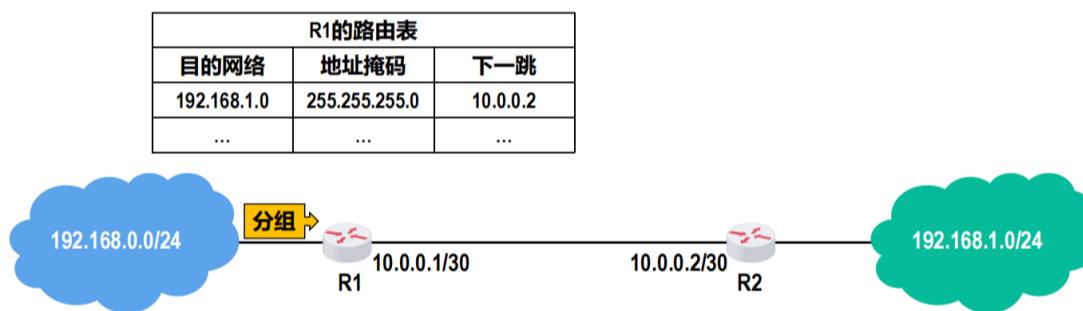


- OpenFlow协议的技术规范由非营利性的产业联盟开放网络基金会 (Open Networking Foundation, ONF) 负责制定。
 - ONF的任务是致力于SDN的发展和标准化。
 - SDN并未规定必须使用OpenFlow，只不过大部分SDN产品采用了OpenFlow作为其控制层面与数据层面的通信接口。
 - OpenFlow从2009年底发表的1.0版开始，每年都被更新，历经12次更新，到2015年3月发布了1.5.1版，目前较为成熟的是1.3版本。



10.3.2. 传统意义上的数据层面的任务

- 传统意义上的数据层面的任务：根据转发表转发分组
 - 转发分组分为以下两个步骤：
 - ① 进行“匹配”：查找转发表中的网络前缀，进行最长前缀匹配。
 - ② 执行“动作”：把分组从匹配结果指明的接口转发出去。



10.3.3. SDN中的广义转发

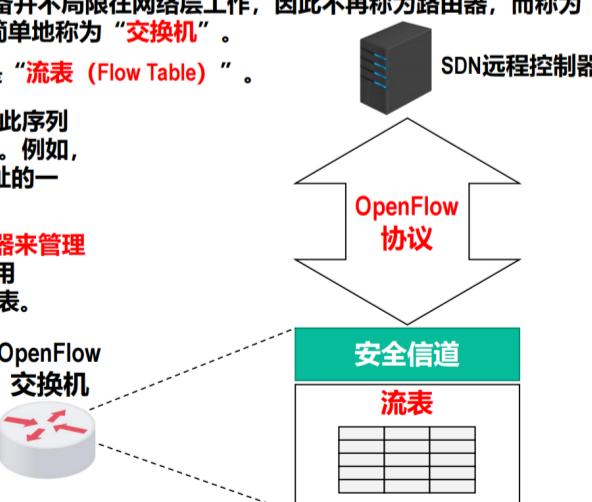
- SDN的广义转发分为以下两个步骤：
 - ① 进行“匹配”：能够对网络体系结构中各层（数据链路层、网络层、运输层）首部中的字段进行匹配。
 - ② 执行“动作”：不仅转发分组，还可以负载均衡、重写IP首部（类似NAT路由器中的地址转换）、人为地阻挡或丢弃一些分组（类似防火墙一样）。

10.3.4. OpenFlow交换机和流表

- 在SDN的广义转发中，完成“匹配+动作”的设备并不局限在网络层工作，因此不再称为路由器，而称为“OpenFlow交换机”或“分组交换机”，或更简单地称为“交换机”。
- 相应的，在SDN中取代传统路由器中转发表的是“流表（Flow Table）”。

- 一个流就是穿过网络的一种分组序列，而在此序列中的每个分组都共享分组首部某些字段的值。例如，某个流可以是具有相同源IP地址和目的IP地址的一连串分组。
- OpenFlow交换机中的流表是由SDN远程控制器来管理的。SDN远程控制器通过一个安全信道，使用OpenFlow协议来管理OpenFlow交换机中的流表。

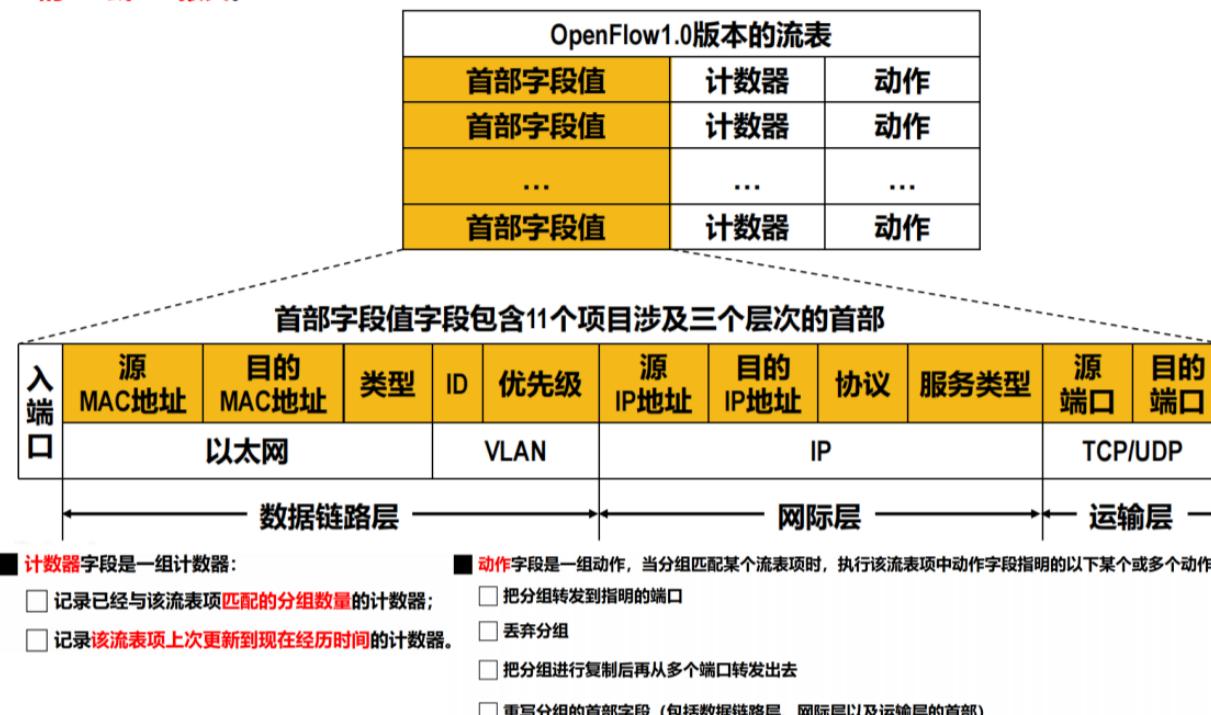
1. 网络设备可以由不同厂商来生产，可以使用在不同类型的网络中。
2. 从SDN远程控制器看到的，是统一的逻辑交换功能。



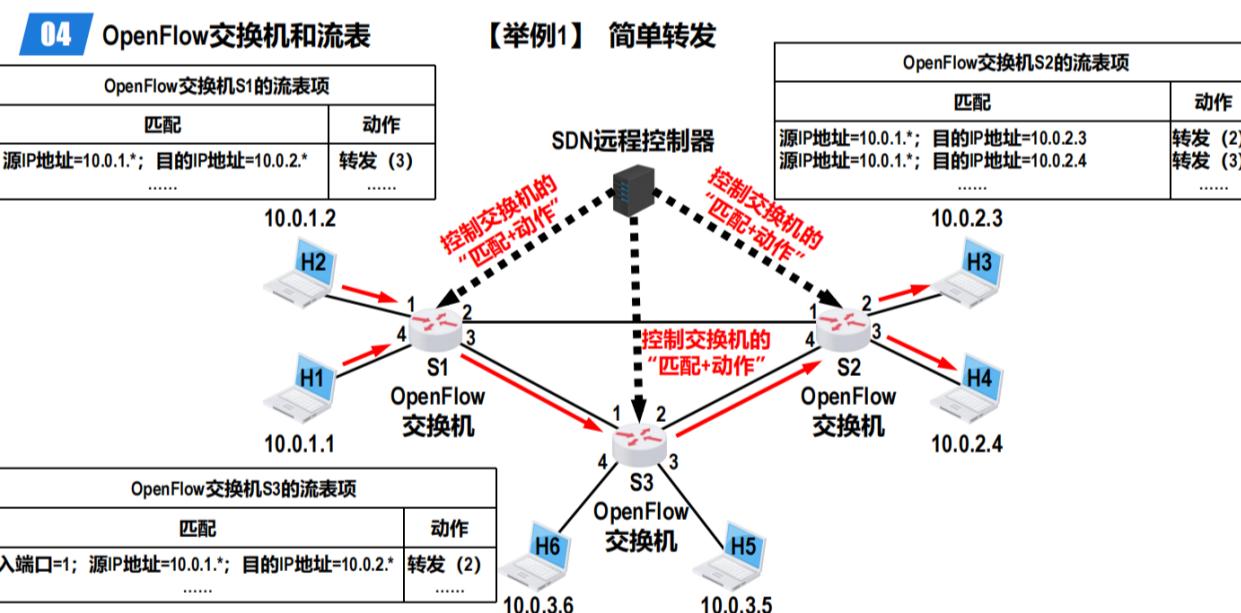
- 每个OpenFlow交换机必须有一个或多个流表。
- 每一个流表可以包含多个流表项。
- 每个流表项包含三个字段：首部字段值（或称匹配字段）、计数器、动作。

OpenFlow1.0版本的流表		
首部字段值	计数器	动作
首部字段值	计数器	动作
...
首部字段值	计数器	动作

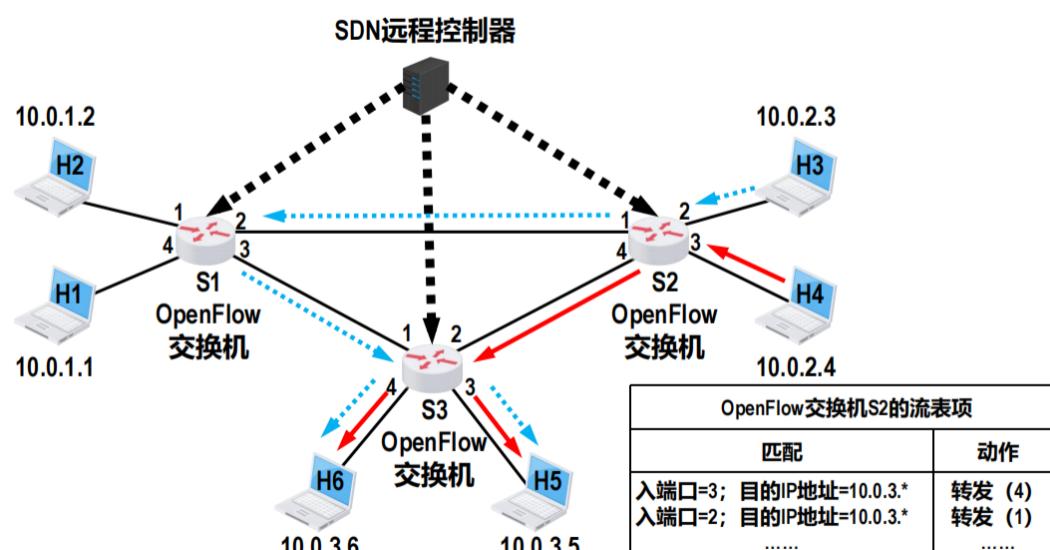
- 首部字段值字段包含有一组字段，用来使入分组（Incoming Packet）的对应首部与之匹配，因此又称为匹配字段。匹配不上的分组就被丢弃，或被发送到SDN远程控制器做更多的处理。
- 在OpenFlow交换机中，既可以处理数据链路层的帧，也可以处理网际层的IP数据报，还可以处理运输层的TCP或UDP报文。



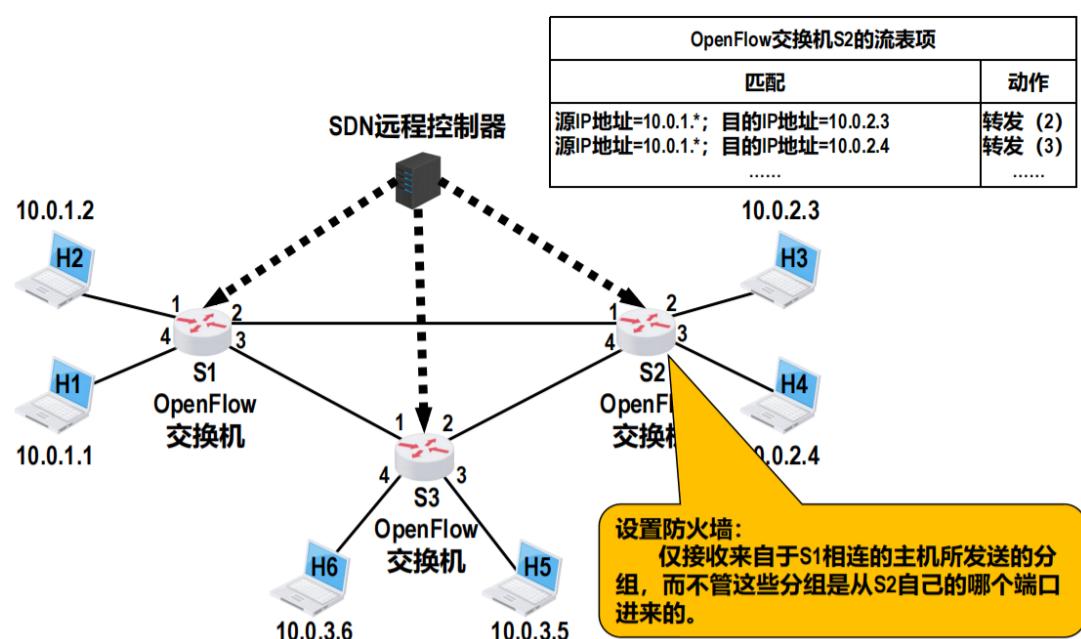
- 简单转发



- 负载均衡



- 防火墙

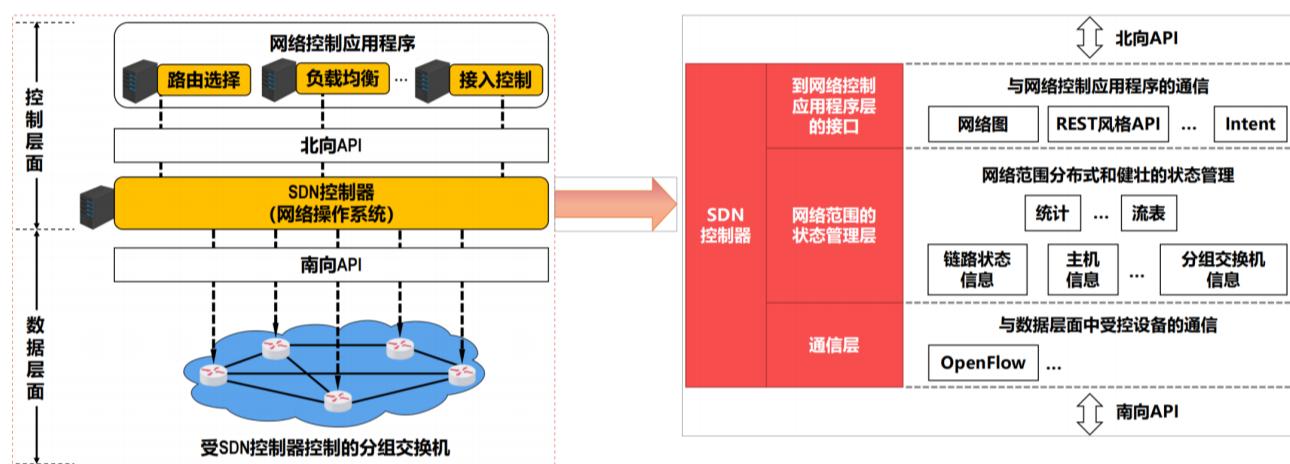


10.4. SDN体系结构

10.4.1. SDN体系结构及其四个关键特征

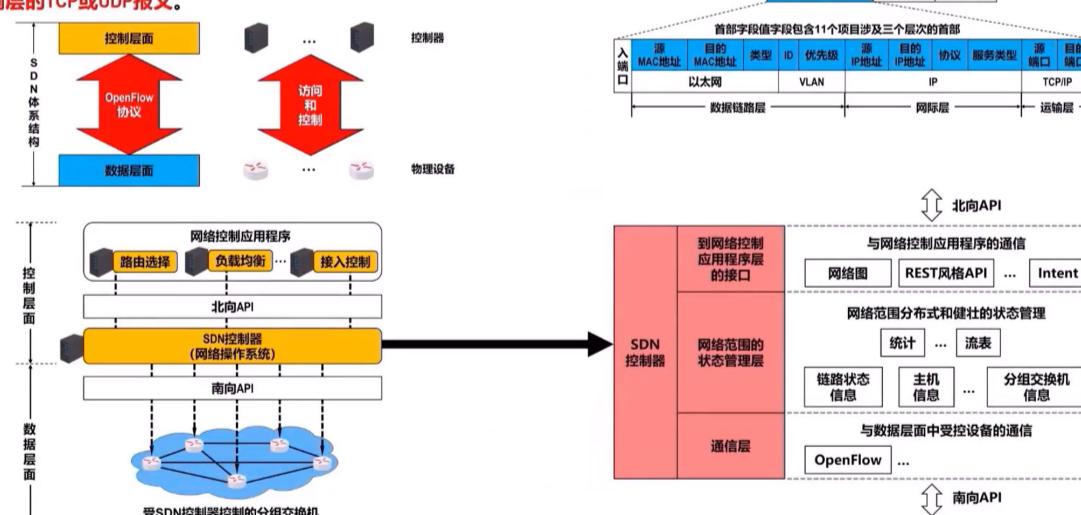
- 基于流的转发
- 数据层面与控制层面分离
- 位于数据层面分组交换机之外的网络控制功能
- 可编程的网络

10.4.2. SDN控制器



10.5. 总结

- SDN这种新型网络体系结构的核心思想：把网络的控制层面和数据层面分离，而让控制层面利用软件来控制数据层面中的许多设备。
- OpenFlow协议可被看成是SDN体系结构中控制层面与数据层面之间的通信接口。
- 在SDN中取代传统路由器中转发表的是“流表（Flow Table）”。在OpenFlow交换机中，既可以处理数据链路层的帧，也可以处理网际层的IP数据报，还可以处理运输层的TCP或UDP报文。



11. 题目

11.1. IPv4分类编址方法

11.1.1. 【2017 36】

【2017年题36】下列IP地址中，只能作为IP分组的源IP地址但不能作为目的IP地址的是（A）。

A. 0.0.0.0 B. 127.0.0.1 C. 20.10.10.3 D. 255.255.255.255

解析

普通的A类地址
既可以作为源地址，
也可以作为目的地址

一般不使用的特殊IPv4地址

	网络号	主机号	IP地址	作为源地址	作为目的地址	表示的意思
选项A	0	0	0.0.0.0	可以	不可以	在本网络上的本主机（例如，DHCP协议）
	0	host-id	0.host-id	可以	不可以	在本网络上的某台主机host-id
选项D	全1	全1	255.255.255.255	不可以	可以	只在本网络上进行广播（各路由器均不转发）
	net-id	全1	A类：net-id.255.255.255 B类：net-id.255.255 C类：net-id.255	不可以	可以	对网络net-id上的所有主机进行广播
选项B	127	非全0或全1的任何数	127.0.0.1-127.255.255.254	可以	可以	用于本地软件环回测试

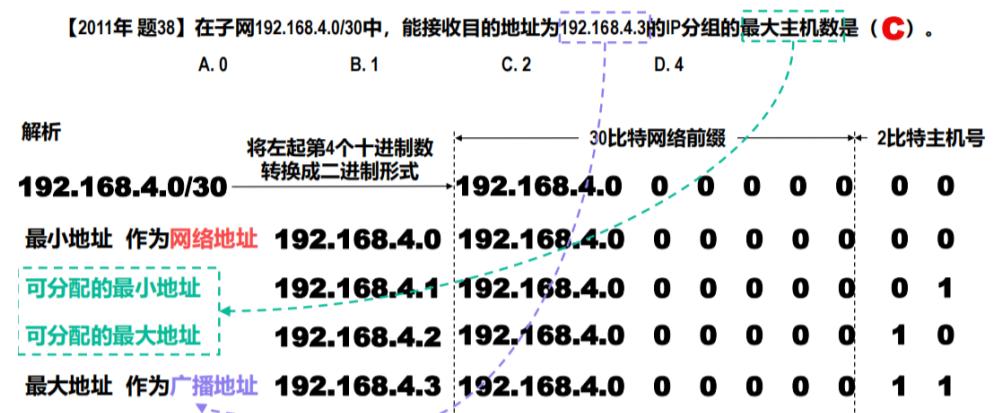
11.2. IPv4划分子网编址方法

11.2.1. 【2012 39】



11.3. 无分类编址方法

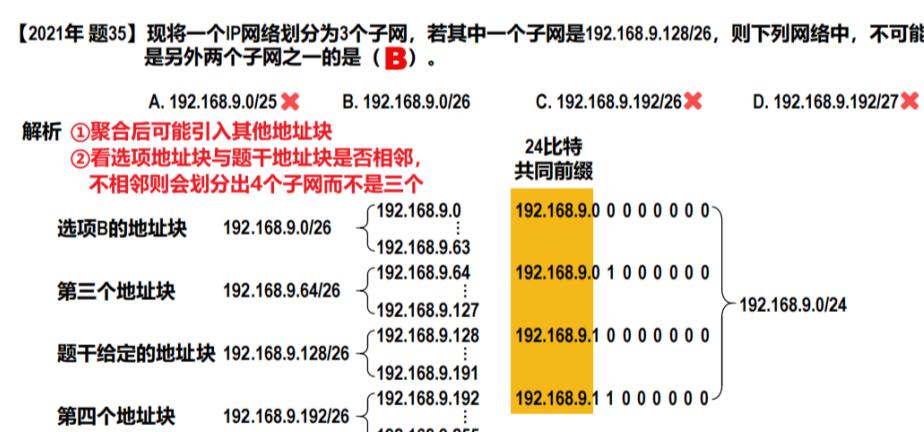
11.3.1. 【2011 38】



11.3.2. 【2018 38】



11.3.3. 【2021 35】



11.4. IPv4地址的应用规划

11.4.1. 【2019 37】

【2019年题37】若将101.200.16.0/20划分为5个子网，则可能的最小子网的可分配IP地址数是（B）。

- A. 126 B. 254 C. 510 D. 1022

解析

需要使用变长子网划分的方法

101.200.16.0/20 { 网络前缀：20比特
主机号：12比特

$$\text{地址数量 } 2^{12} = 4096$$

根据题意，需要将这4096个地址划分成5个地址块，其中4个地址块都尽量大，则剩余1个地址块就最小。

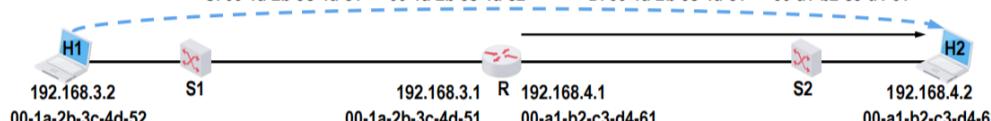
101.200.16.0/20 $2^{12} = 4096$ 个地址	101.200.16.0	/21地址块	$2^{32-21} = 2048$ 个地址
		/22地址块	$2^{32-22} = 1024$ 个地址
		/23地址块	$2^{32-23} = 512$ 个地址
		/24地址块	$2^{32-24} = 256$ 个地址
	101.200.31.255	/24地址块	$2^{32-24} = 256$ 个地址

去掉全0的网络地址
和全1的广播地址

11.5. 数据报传送过程中IPv4地址与MAC地址的变化情况

【2018年题37】路由器R通过以太网交换机S1和S2连接两个网络，R的接口、主机H1和H2的IP地址与MAC地址如下图所示。若H1向H2发送一个IP分组P，则H1发出的封装P的以太网帧的目的MAC地址、H2收到的封装P的以太网帧的源MAC地址分别是（D）。

- A. 00-a1-b2-c3-d4-62 B. 00-a1-b2-c3-d4-62 C. 00-a1-b2-c3-d4-51 D. 00-a1-b2-c3-d4-61



解析

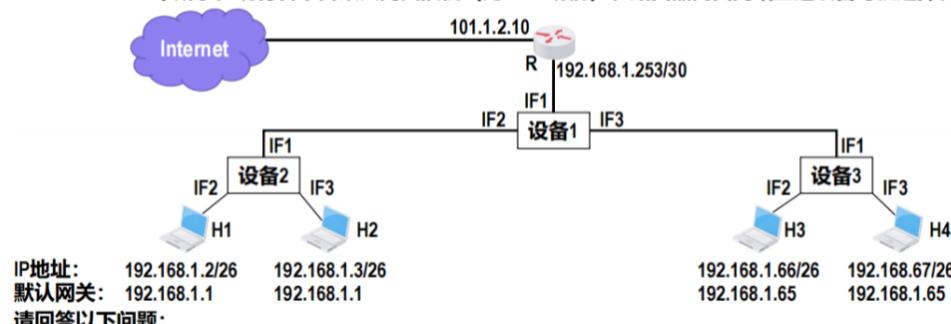
在数据包的传送过程中，源IP地址和目的IP地址保持不变，而源MAC地址和目的MAC地址逐链路（或逐网络）改变。

数据包传输区间	在网络层写入IP数据报首部的IP地址		在数据链路层写入帧首部的MAC地址	
	源IP地址	目的IP地址	源MAC地址	目的MAC地址
H1→R	192.168.3.2	192.168.4.2	00-a1-b2-c3-d4-52	00-a1-b2-c3-d4-51
R→H2	192.168.3.2	192.168.4.2	00-a1-b2-c3-d4-61	00-a1-b2-c3-d4-62

11.6. IP数据报的发送和转发过程

11.6.1. 【2019 47】

【2019年题47】某网络拓扑如下图所示，其中R为路由器，主机H1-H4的IP地址配置以及R的各接口IP地址配置如图中所示。现有若干台以太网交换机（无VLAN功能）和路由器两类网络互连设备可供选择。



请回答以下问题：

- (1) 设备1、设备2和设备3分别应选择什么类型网络设备？
- (2) 设备1、设备2和设备3中，哪几个设备的接口需要配置IP地址？并为对应的接口配置正确的IP地址。
- (3) 若主机H3发送一个目的地址为192.168.1.127的IP数据报，网络中哪几个主机会收到该数据报？

- (1) 设备1为路由器，设备2为交换机。
(2) 设备1的IF2的ip地址为H1、H2的默认网关192.168.1.1
设备1的IF3的ip地址为H3、H4的默认网关192.168.1.65
设备1的IF1的ip地址根据R的接口ip的CIDR形式推出，
/30用来分配给只有两个路由器接口的点对点链路，故IF1: 192.168.1.253
(3) 容易发现该目的地址为广播地址，故H4会接收改数据报。

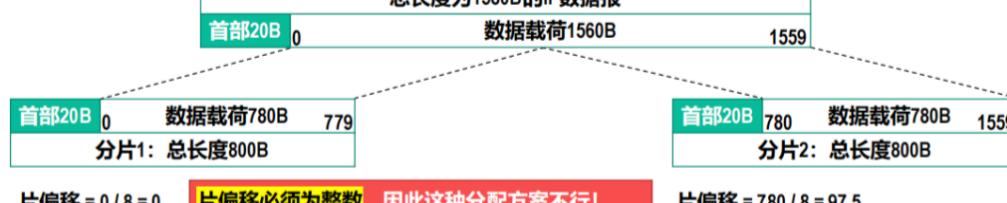
11.7. 片偏移

11.7.1. 【2020 36】

【2021年题36】若路由器向MTU=800B的链路转发一个总长度为1580B的IP数据报（首部长度为20B）时，进行了分片，且每个分片尽可能大，则第2个分片的总长度字段和MF标志位的值分别是（B）。

- A. 796, 0 B. 796, 1 C. 800, 0 D. 800, 1

解析

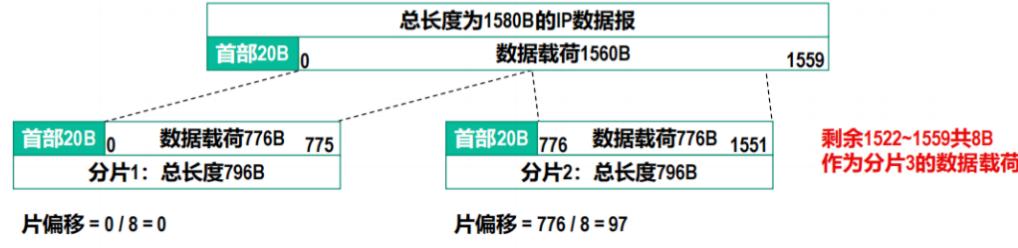


$$\text{片偏移} = 0 / 8 = 0$$

片偏移必须为整数，因此这种分配方案不行！

$$\text{片偏移} = 780 / 8 = 97.5$$

分片的数据载荷的最大长度取小于780且能整除8的最大整数 $[780 \div 8] \times 8 = 776$



$$\text{片偏移} = 0 / 8 = 0$$

$$\text{片偏移} = 776 / 8 = 97$$

MF = 1

11.8. RIP

11.8.1. 【2010 35】

【2010年题35】某自治系统内采用RIP协议，若该自治系统内的路由器R1收到其邻居路由器R2的距离矢量，距离矢量中包含信息<net1,16>，则能得出的结论是 (D)。

- A. R2可以经过R1到达net1，跳数为17
- B. R2可以到达net1，跳数为16
- C. R1可以经过R2到达net1，跳数为17
- D. R1不能经过R2到达net1

解析

在RIP协议中，**距离16表明目的网络不可达**。

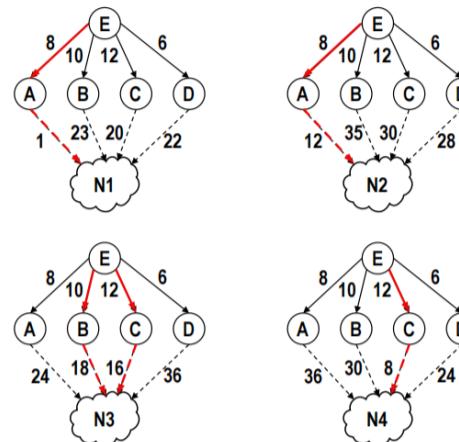
因此，R2无法到达net1，R1也无法通过R2到达net1。

11.8.2. 【2021 37】

【2021年题37】某网络中的所有路由器均采用距离向量路由算法计算路由。若路由器E与邻居路由器A、B、C和D之间的直接链路距离分别是8, 10, 12和6，且E收到邻居路由器的距离向量如下表所示，则路由器E更新后的到达目的网络Net1~Net4的距离分别是 (D)。

目的 网络	A的 距离向量	B的 距离向量	C的 距离向量	D的 距离向量
Net1	1	23	20	22
Net2	12	35	30	28
Net3	24	18	16	36
Net4	36	30	8	24

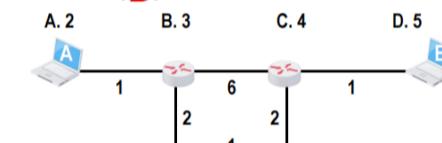
- A. 9, 10, 12, 6
- B. 9, 10, 28, 20
- C. 9, 20, 12, 20
- D. 9, 20, 28, 20



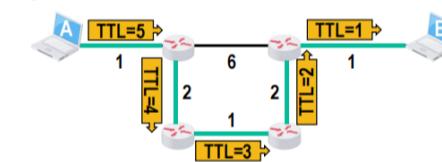
11.9. OSPF

11.9.1. 【2014 43改】

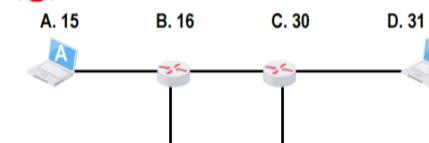
【改自2014年题43】网络拓扑如下图所示，假设各路由器使用OSPF协议进行路由选择且已收敛，各链路的度量已标注在其旁边，主机A给B发送一个IP数据报，为了让IP数据报能够到达主机B，其首部中的TTL字段的取值至少应设置为 (D)。



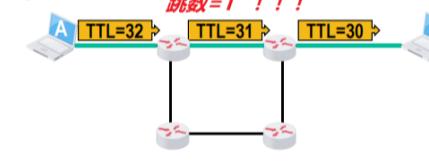
解析



【举一反三】网络拓扑如下图所示，假设各路由器使用RIP协议进行路由选择且已收敛，主机A给B发送一个IP数据报，其首部中的TTL字段的值设置为32，则当主机B正确接收到该IP数据报时，其首部中的TTL字段的值为 (C)。



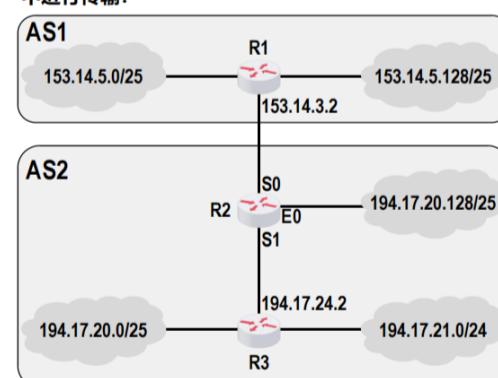
解析



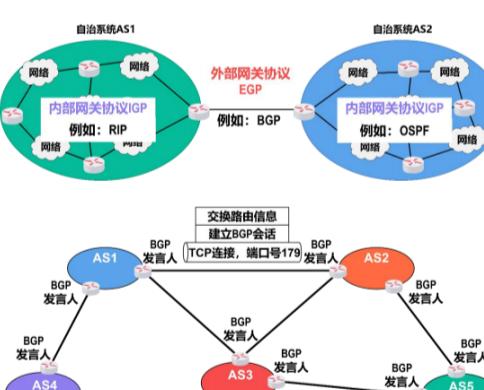
11.10. BGP

11.10.1. 【2013 46(3)】

【2013年题47（3）】R1与R2之间利用哪个路由协议交换路由信息？该路由协议的报文被封装到哪个协议的分组中进行传输？



解析

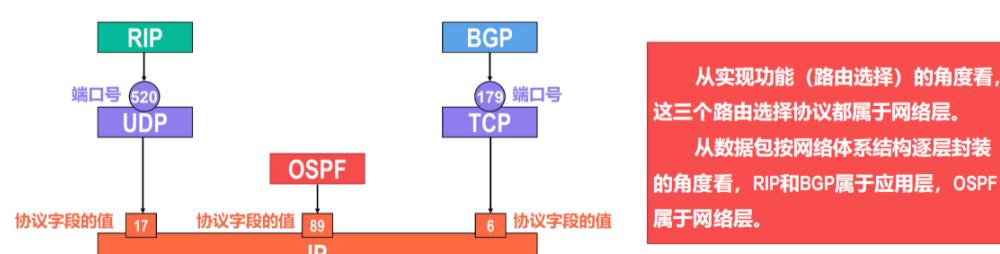


11.10.2. 【2017 37】

【2017年题37】直接封装RIP、OSPF、BGP报文的协议分别是 (D)。

- A. TCP、UDP、IP
- B. TCP、IP、UDP
- C. UDP、TCP、IP
- D. UDP、IP、TCP

解析



11.11. ICMP

11.11.1. 【2010 36】

【2010年 题36】若路由器R因为拥塞丢弃IP分组，则此时R可向发出该IP分组的源主机发送的ICMP报文类型是
(C)。

- A. 路由重定向
- B. 目的不可达
- C. 源点抑制
- D. 超时

11.12. SDN题目

【习题1】下列有关SDN的描述中，**正确**的是（ ）。

- A. SDN是近年来出现的一种新型物理网络
- B. SDN等同于OpenFlow
- C. SDN将网络的控制层面和数据层面分开
- D. OpenFlow交换机就是IP路由器

【习题2】下列有关SDN的描述中，**错误**的是（ ）。

- A. SDN是近年来出现的一种新型网络体系结构
- B. OpenFlow可被看作是SDN的控制层面与数据层面的通信接口
- C. SDN远程控制器位于OpenFlow交换机中
- D. OpenFlow交换机基于“流表”转发分组

【习题3】下列各种首部中的字段，**不能**在OpenFlow1.0中匹配的是是（ ）。

- A. 目的MAC地址
- B. VLAN ID
- C. 源IP地址
- D. 窗口