计算机网络

计算机与信息学院
人 工 智 能 学 院

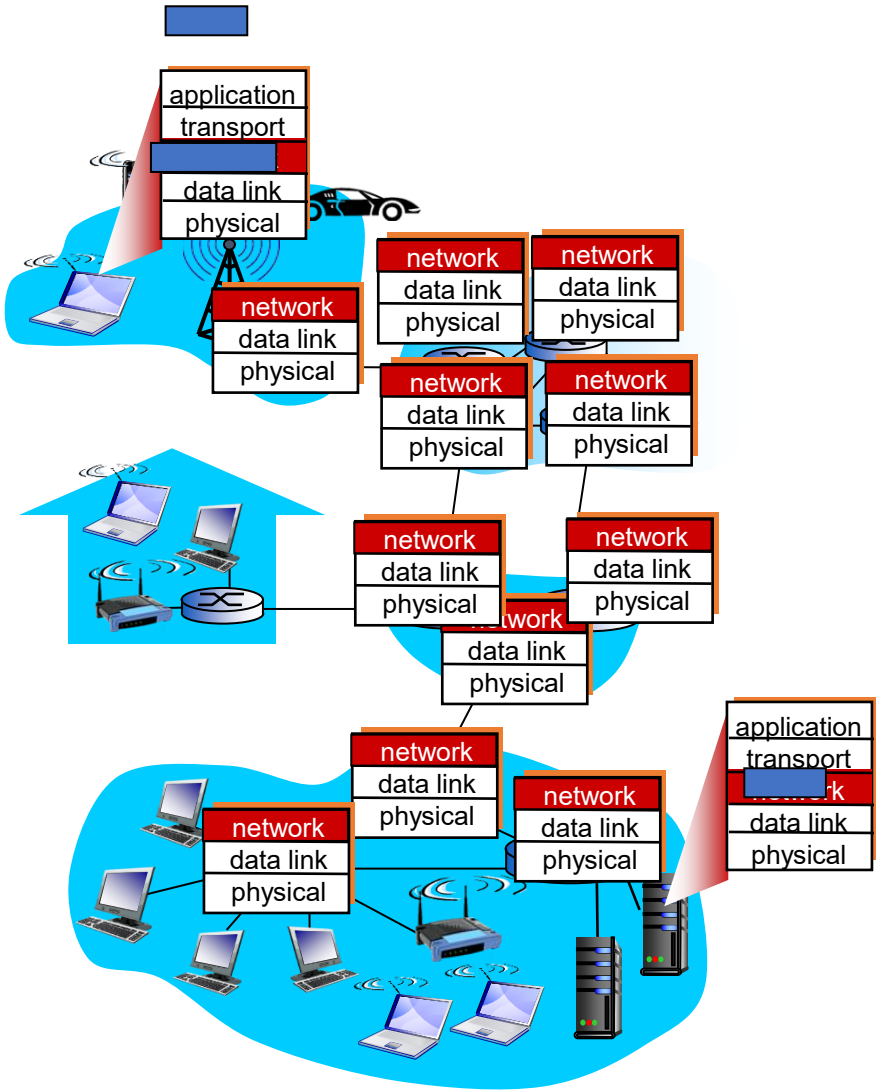
TCP/IP产生的背景

- 1972-1980： ARPAnet、AlOHAnet、 SNA（IBM）、Telenet、以及各种局域网涌现
- 异构网络如何互连：网络的网络

1983.1.1： TCP/IP作为ARPAnet新的标准主机协议，正式部署，替代了NCP协议

- 传输层：TCP/UDP，实现进程之间的通信
- 网络层：核心IP协议，实现不同网络中主机之间的通信

网络层互连设备：路由器



- 数据平面（data plane）
  - ✓局部：每个路由器功能
  - ✓决定从路由器输入端口到达的分组如何转发到输出端口
  - ✓转发表

✓控制平面（control plane）

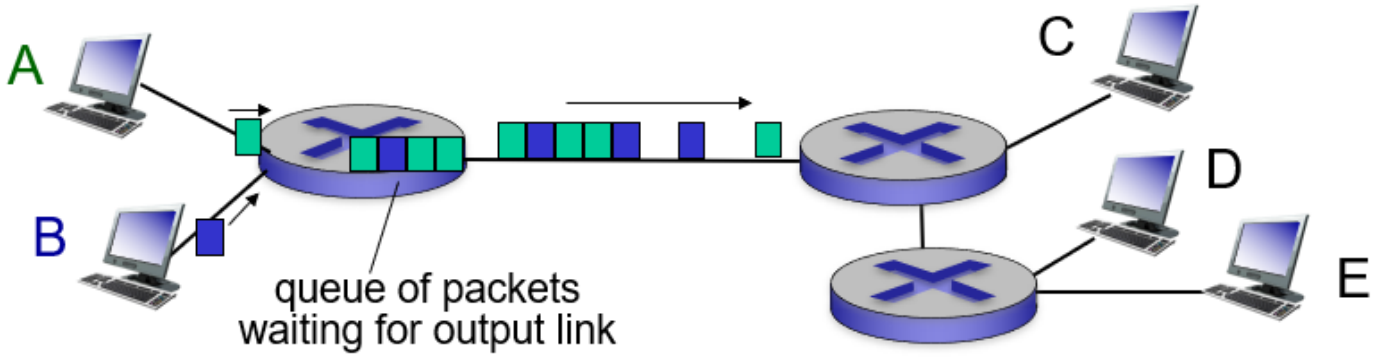
- ✓全局：多个路由器
- ✓决定数据报如何在路由器之间路由：从源到目标主机之间的端到端路径

本章内容

- 网络层提供的服务
- 数据平面（转发：IP、ARP、ICMP）
- 控制平面（路由：RIP、OSPF、BGP）

6.1 网络层提供的服务

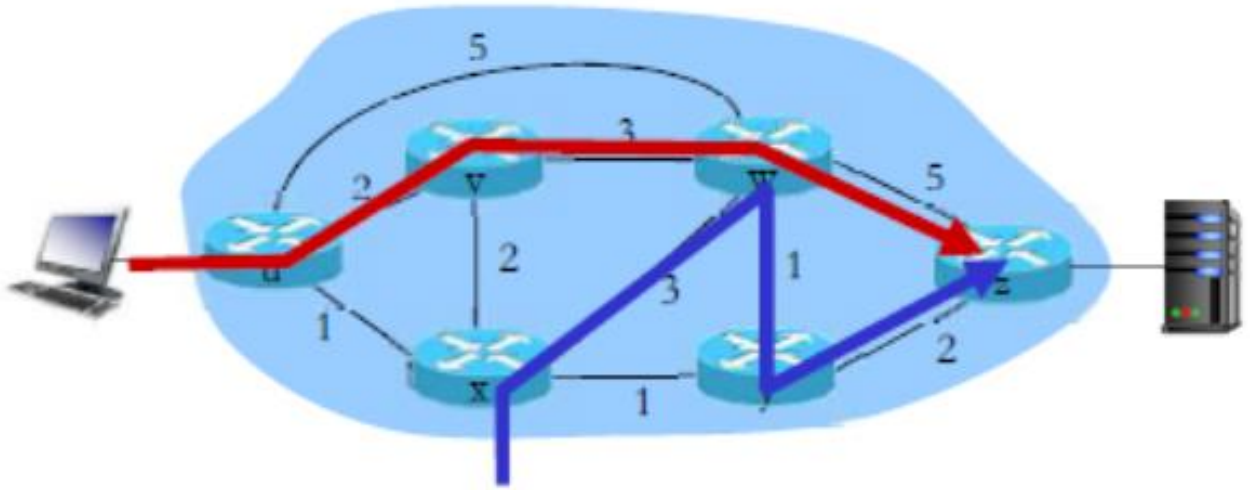
➢分组交换



- 虚电路
- 数据报

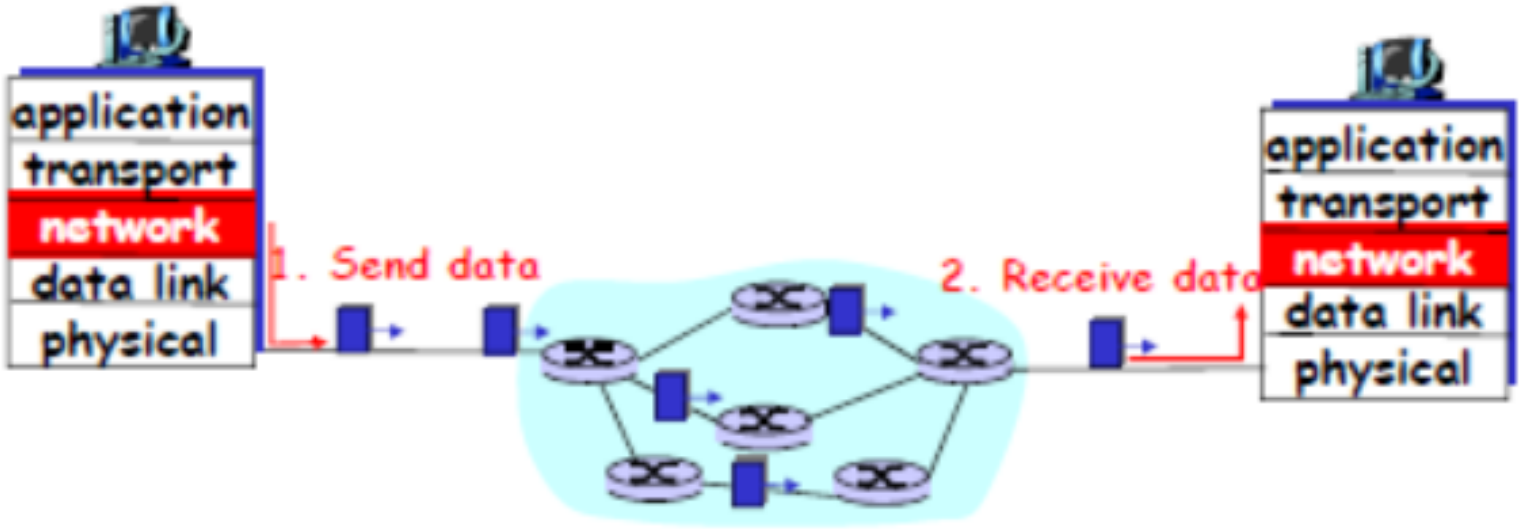
➢虚电路

- 在分组传输之前，两个主机之间，呼叫建立连接
- 后续分组都走相同路径
- 路由器维持该条路径信息



➢数据报

- 不建立连接
- 每个分组单独寻径



哪一种交换能够保证分组的按序到达？

1）虚电路

2）电路

3）数据报

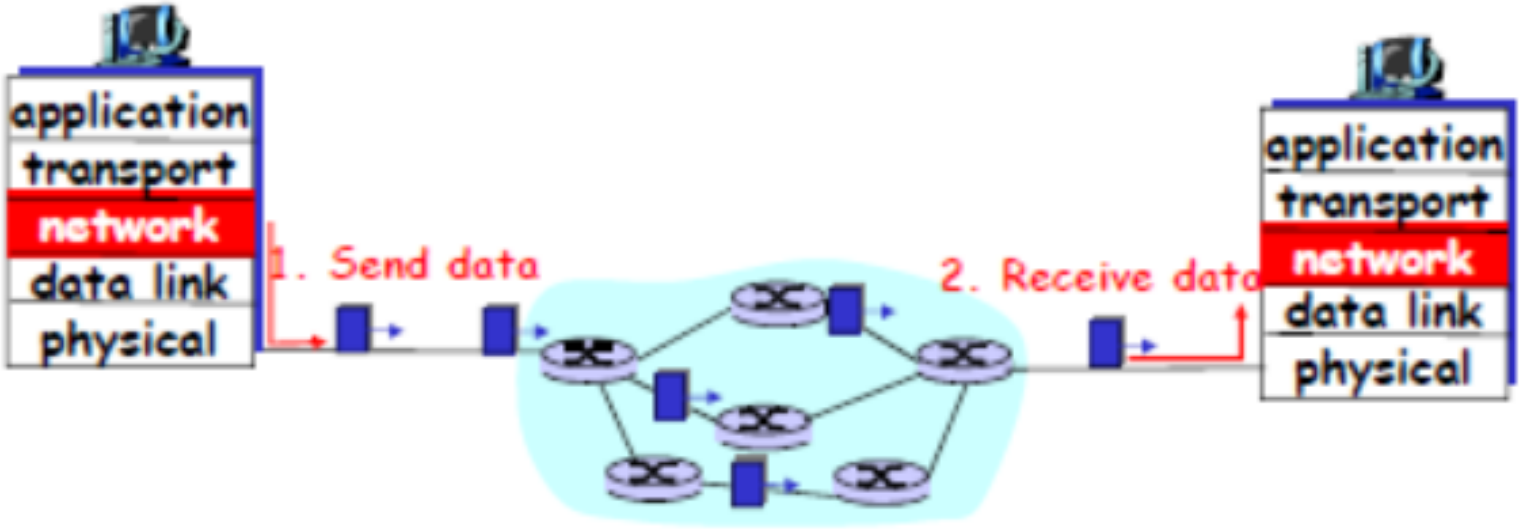
虚电路和数据报都属于分组交换

	数据报	虚电路
延时	分组传输延时	虚电路建立延时，分组传输延时
路由选择	每个分组单独选择路由	建立虚电路时选择路由，以后所有分组都使用该路由
状态信息	子网无需保存状态信息	每个结点要保存一张虚电路表
地址	每个分组携带完整的源/目的地址	每个分组分配一个较短的虚电路号
节点失败的影响	除了崩溃时正在该节点处理的分组都丢失外，无其他影响	所有经过失效节点的虚电路都要被终止
拥塞控制	难	容易

Internet的网络层采用了数据报方式

服务模型：尽力而为（best effort）

- ✓ 数据可能会丢失（不保证可靠交付）
- ✓ 数据会失序（不保证按序到达）
- ✓ 数据会延迟（不保证时延）
- ✓ 不提供拥塞控制



- 数据平面（data plane）：转发

  ✓IP地址、IP数据报格式

  ✓转发表和转发算法

  ✓辅助协议：ARP、ICMP、NAT

6.2 IP （ Internet Protocol ）

✓对主机或路由器的接口进行编址

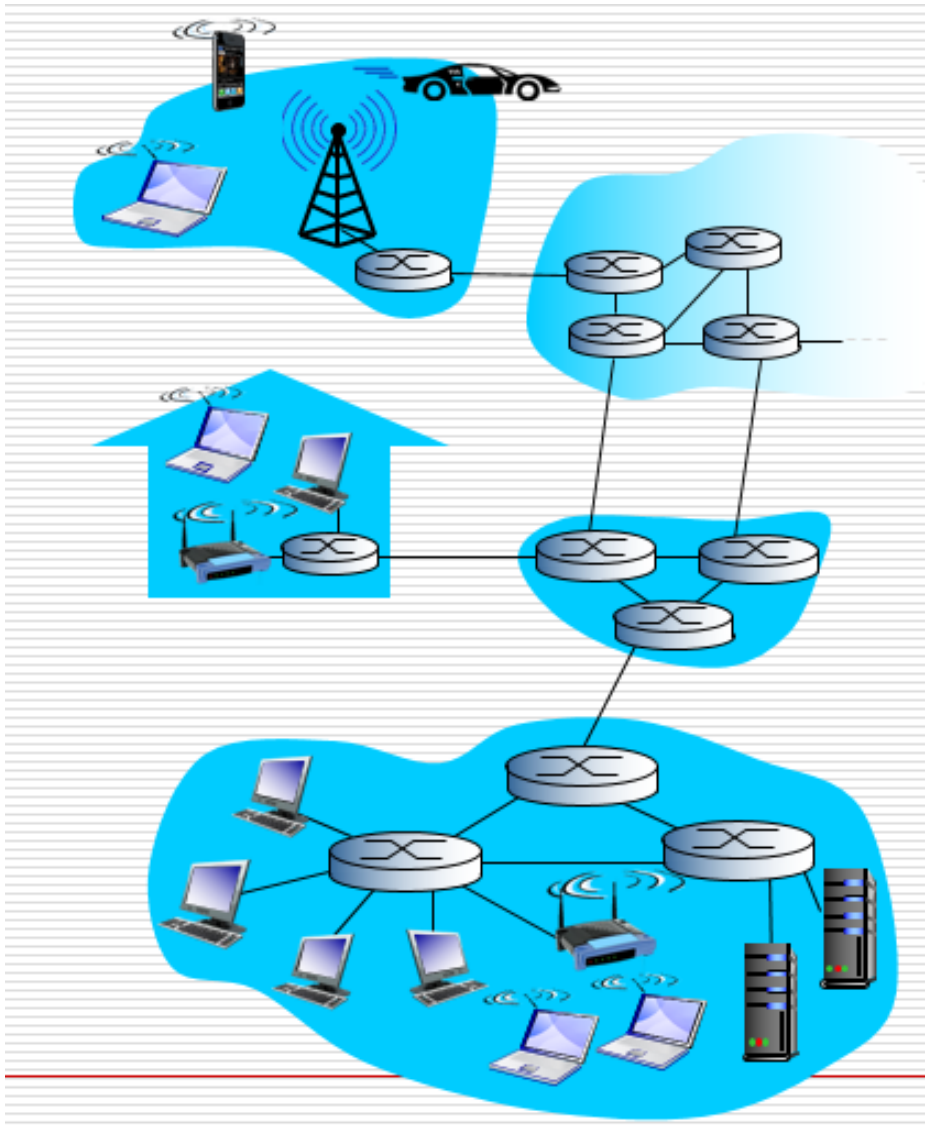
✓IPv4： 32bit

——点分十进制

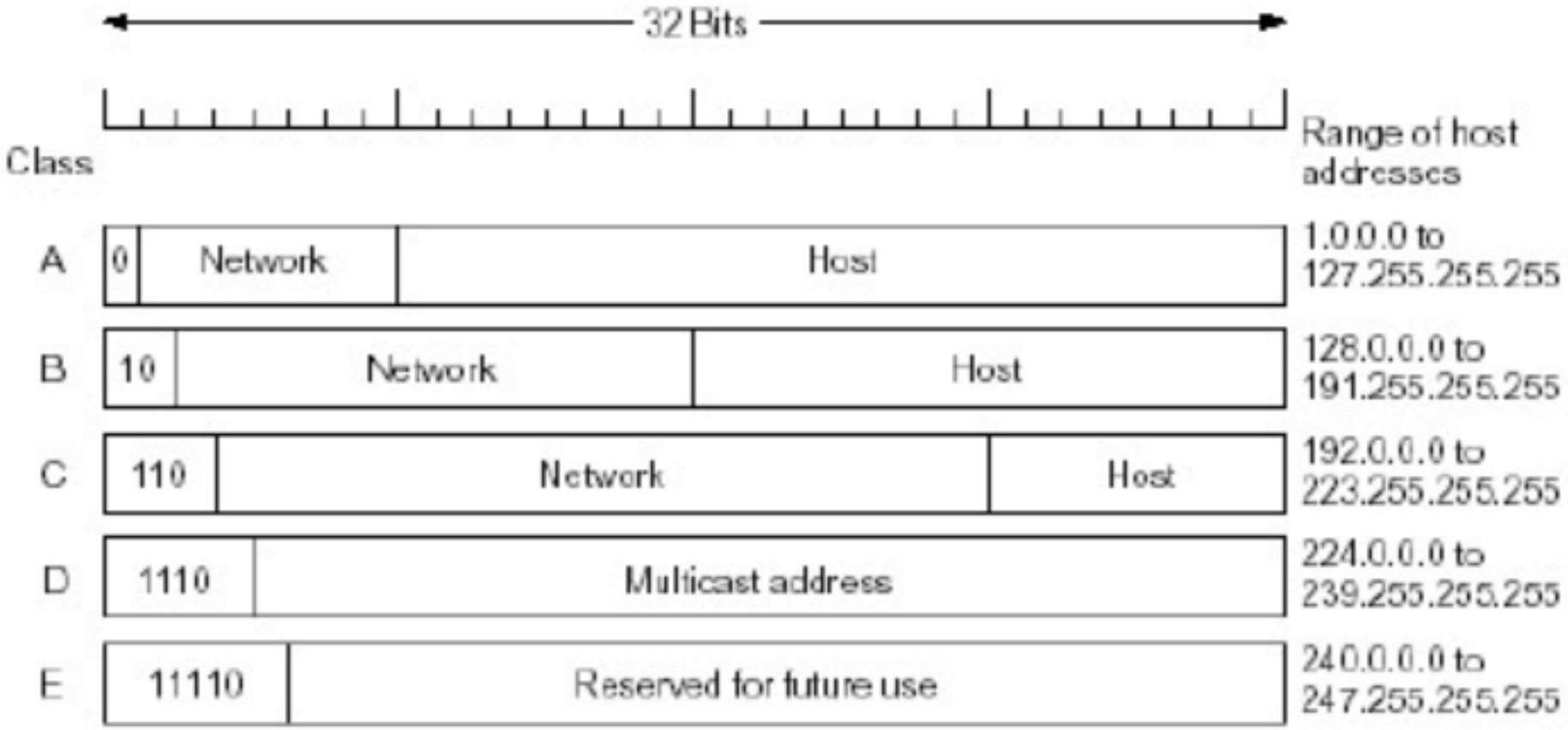
IP 地址 ::= { <网络号>, <主机号>}

网络号，由国际互联网信息中心(InterNIC)分配

主机号：系统管理员分配（动态分配）



A(8)、B(16)、C(24)三类网络



特殊IP地址

✓32bit 全1：255.255.255.255（广播地址）

✓32bit 全0：0.0.0.0 （网络内的某台主机，地址未分配）

✓ 主机号全0：为网络号，不作为主机IP地址分配

✓主机号全1：广播地址

特殊IP地址

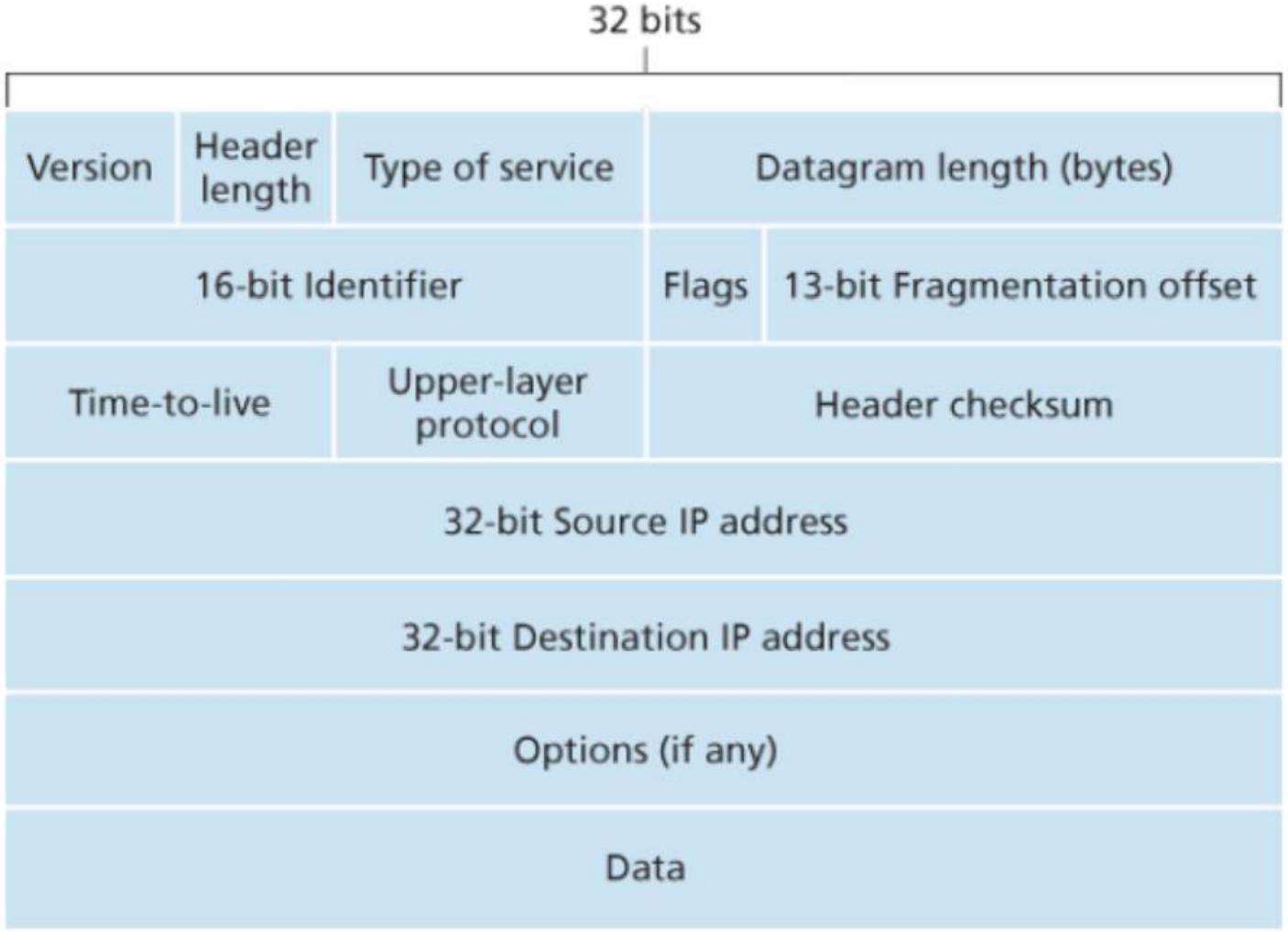
127.*.*.*：回环地址

- ✓网络层检测到目的地址为回环地址，则不发送到链路层，而是放入IP回环接口，传给本机的上层TCP
- ✓允许运行在同一台主机上的客户程序和服务器程序通过TCP/IP进行通信

- 某主机IP地址：180.172.32.12

- 1）所在网络的类别：
- 2）该网络的网络号
- 3）该网络的广播地址
- 4）可分配的IP地址范围

IP数据报格式


32 bits
Version
Header length
Type of service
Datagram length (bytes)
16-bit Identifier
Flags
13-bit Fragmentation offset
Time-to-live
Upper-layer protocol
Header checksum
32-bit Source IP address
32-bit Destination IP address
Options (if any)
Data

路由器：转发表

网络号	下一跳地址	接口
……	……	……
……	……	……
……	……	……

✓网络号：路由器为每一个网络指定路由，而不是为每个主机指定路由

✓下一跳：相邻路由器的IP地址（相邻：两个路由器直接相连或连在同一个物理网络中）

✓接口：路由器的端口标识

转发：示例

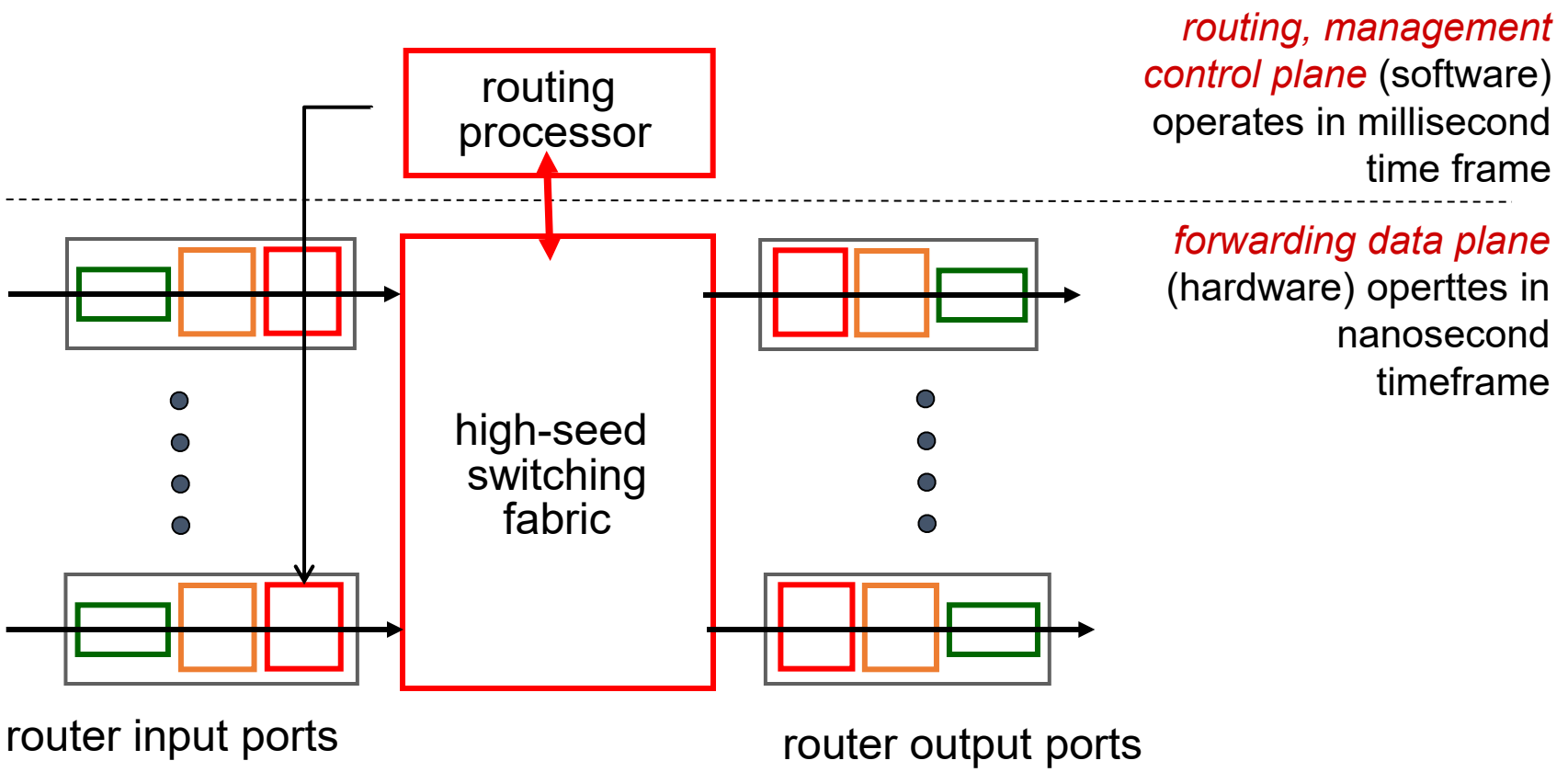
主机

- ✓ 目的主机与源主机在一个物理网络中，直接交到链路层，封装帧，发送
- ✓ 否则，发给默认路由器，由路由器转发

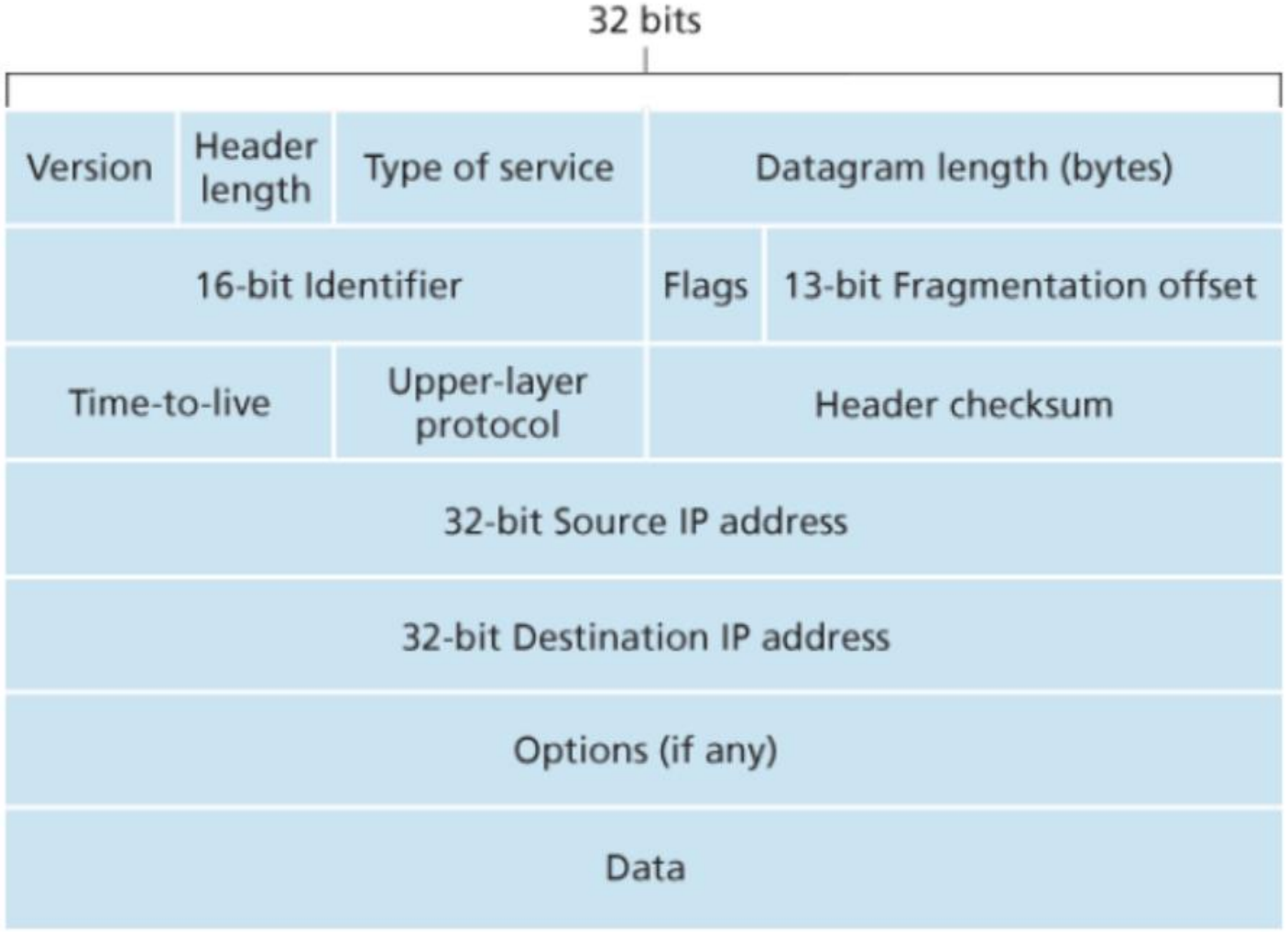
路由器

- ✓ 根据IP数据报中的目的地址，提取网络号，查找转发表，向对应接口转发

路由器：内部结构


routing processor
high-seed switching fabric
router input ports
router output ports
routing, management control plane (software) operates in millisecond time frame
forwarding data plane (hardware) operttes in nanosecond timeframe

分片与重组



- ✓ 不同网络，链路层的MTU（最大传输单元）不同
- ✓ 分片：一个IP数据报被分成若干小的数据报
- ✓ 重组：在目的主机进行

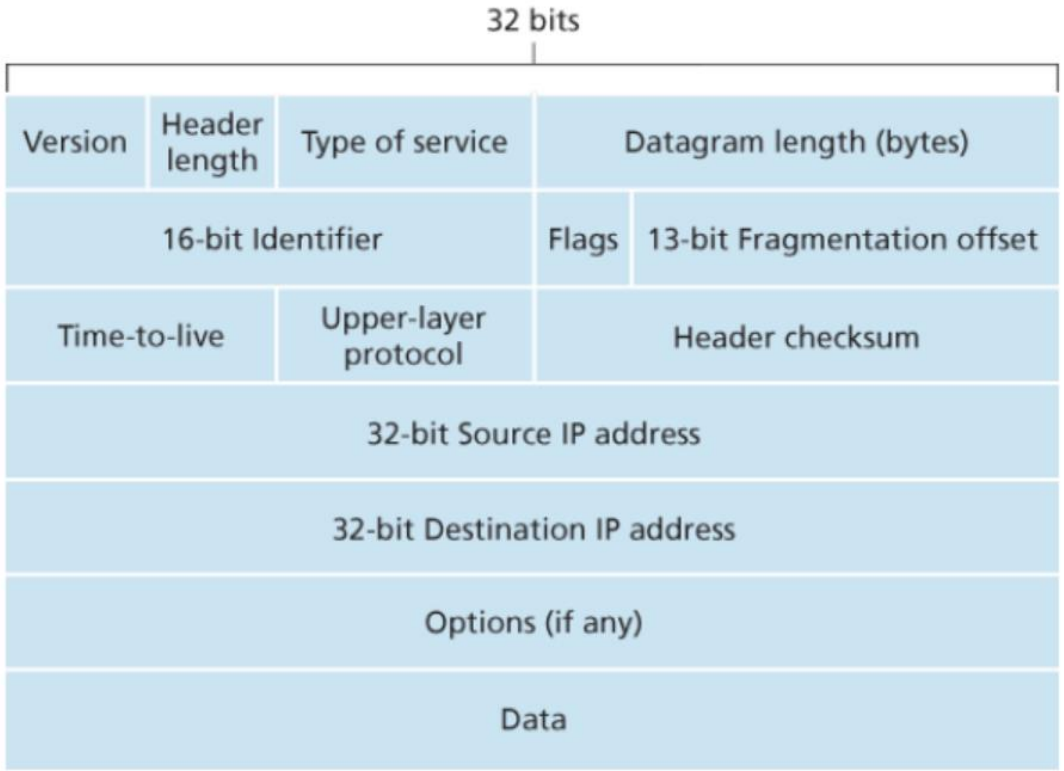
分片与重组

- 标志位

  3位，DF（Don't Fragment）：“0”能分片；“1”不能分片

  MF（More Fragment）：“1”后面还有分片；“0”最后一个分片

- 偏移量（13bit），单位：8bytes



- 2400字节数据报
  - 固定首部：20字节
  - 数据长度：2380字节
- MTU：1000字节
  1）20字节+976字节
  2）20字节+976字节
  3）20字节+428字节

Length: 2400	ID=X	DF:0 MF:0	Offset: 0

Length: 996	ID=X	DF:0 MF:1	Offset: 0
Length: 996	ID=X	DF:0 MF:1	Offset: 122
Length: 448	ID=X	DF:0 MF:0	Offset: 244

子网（subnet）

- 分类IP地址存在的问题：A类和B类可容纳的主机数太多，实际的网络没有容纳这么多主机，IP地址的空间利用率低

B类网：140.10.0.0

将一个网络划分若干地址空间不重复的网络——subnet

- 子网划分：从主机号取若干位作为子网号 subnet-id

B类网：140.10.0.0

划分2个子网

划分4个子网

划分6个子网

- 划分子网带来的问题：如何获得某个IP地址所在的网络号（+子网号）

子网掩码：32位

——对应网络号及子网号的位置1

——对应主机号的位置0

网络号（+子网号）=IP地址∩子网掩码

- 子网掩码

B类网：140.10.0.0

划分2个子网

划分4个子网

划分6个子网

A类、B类、C类

——255.0.0.0

——255.255.0.0

——255.255.255.0

固定长度子网：子网号长度固定，子网掩码相同，每个子网可容纳的主机数相同

变长子网：划分的子网号长度不一致，根据子网需要容纳的主机数划分

202.120.224.0，划分成5个子网，3个子网容纳主机数50台，2个子网容纳主机数30台

子网号能否为全0和全1？

路由器转发算法

网络号	子网掩码	下一跳地址	接口
特定主机	255.255.255.255	R1	E1
223.1.1.0	255.255.255.0	直连	E0
223.1.2.0	255.255.255.0	R1	E1
默认路由	0.0.0.0	R2	E2

1. 路由器从IP数据报的首部提取目的IP地址，与子网掩码 运算，获得网络号
2. 判断是否可直接交付
3. 判断是否为特定主机路由
4. 判断是否为间接路由
5. 默认路由/报告错误

- 子网划分，暂时缓解地址空间的分配不足
- 分类IP地址存在的问题
  - ✓ A类和B类越来越少
  - ✓ C类网络增多，路由表规模不断增长

无类别域间路由选择（CIDR：Classes InterDomain Routing）

- 取消分类IP地址的网络号长度规定，网络号可以为任意长度

  斜线记法:在 IP 地址面加上一个斜线“/”，写上网络号（网络前缀）所占的位数

128.14.32.0/20

子网掩码： 11111111 11111111 11110000 00000000（255.255.240.0）

CIDR 地址块，分配到一个CIDR地址块的组织，仍可以根据需要划分子网

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

✓ C类网络增多，路由表规模不断增长——CIDR支持路由聚合(地址聚会)，减小路由表的规模

目的网络	掩码	下一跳
192.60.128.0/24	255.255.255.0	R1
192.60.129.0/24	255.255.255.0	R1
192.60.130.0/24	255.255.255.0	R1
192.60.131.0/24	255.255.255.0	R1

路由聚合带来的问题：一个IP数据报可能会匹配到多个选项

目的网络	掩码	下一跳
192.60.128.0/22	255.255.252.0	R1
192.60.131.0/24	255.255.255.0	R2

最长网络前缀匹配

- 数据平面（data plane）：转发

  ✓IP地址、IP数据报格式

  ✓转发表和转发算法

  ✓辅助协议：ARP、ICMP、NAT

某路由表中的4条路由选项，下一跳相同

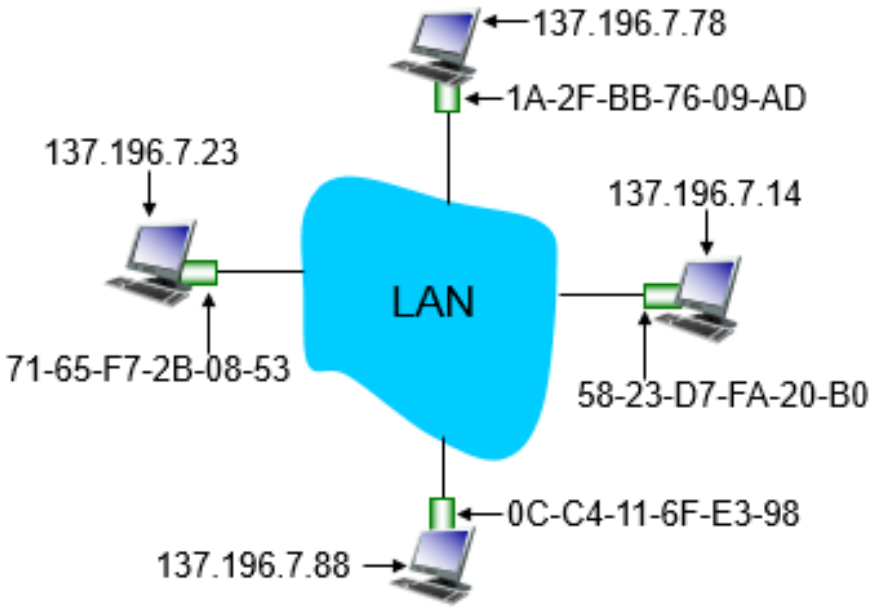
- 35.230.32.0/21
- 35.230.40.0/21
- 35.230.48.0/21
- 35.230.56.0/21

聚合后的路由_____ Mask _____。

6.3 ARP

✓Address Resolution Protocol

根据主机的IP地址，查找其对应的MAC地址



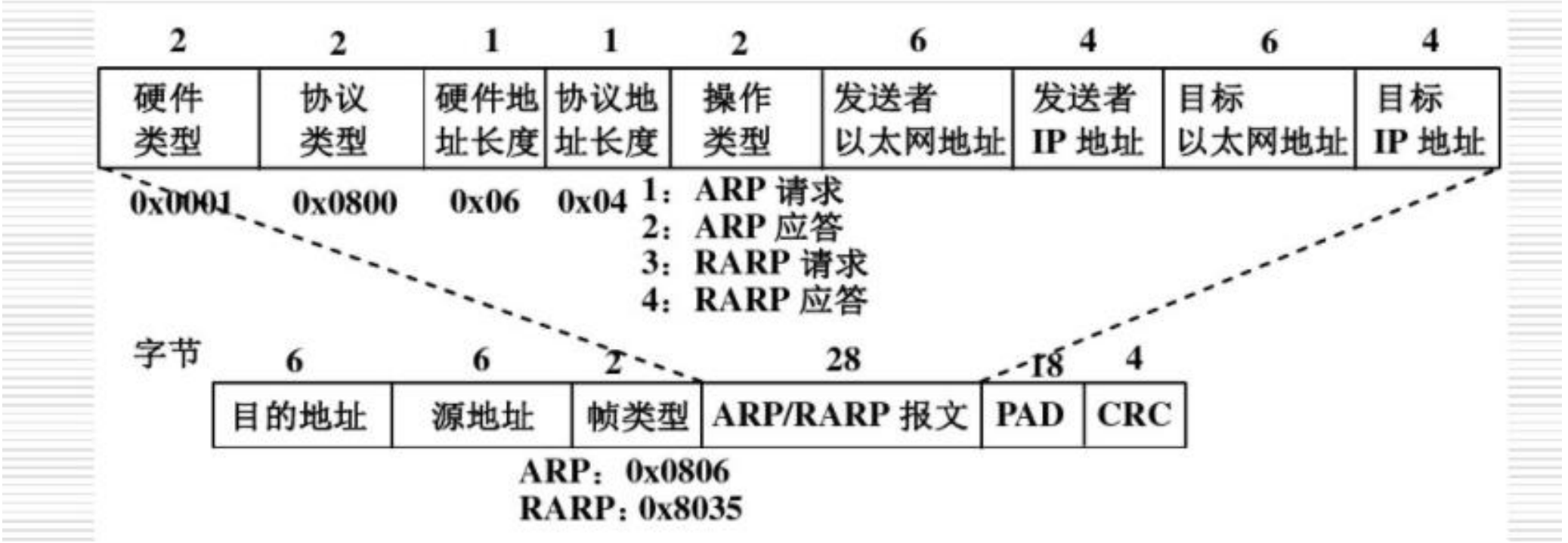
- 每台主机或路由器：在内存中维持ARP表

< IP address; MAC address; TTL>

A向B发送IP数据报，B的MAC地址不在ARP表中？

- A：发送ARP查询分组
  - ✓查询分组：A的IP地址、A的MAC地址、B的IP地址、B的MAC地址（全0）
  - ✓封装成帧（目的MAC地址：FF-FF-FF-FF-FF-FF）

- B收到ARP查询分组，发送：ARP响应分组（单播）



硬件类型：值1（以太网）

协议类型：值0800（IPv4）

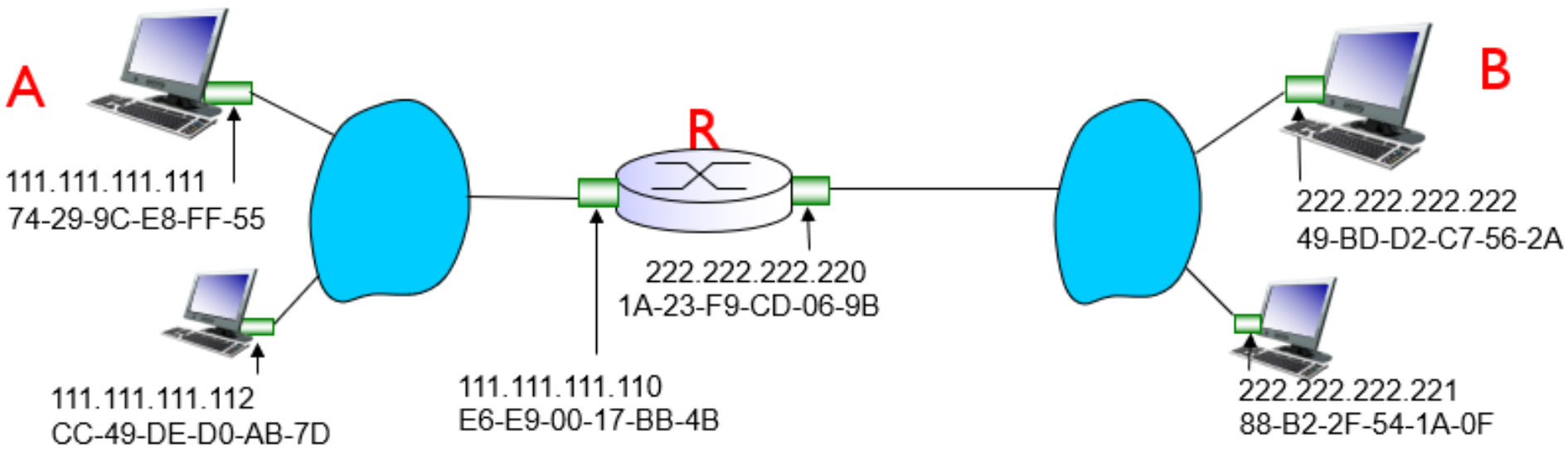
硬件地址长度：值6

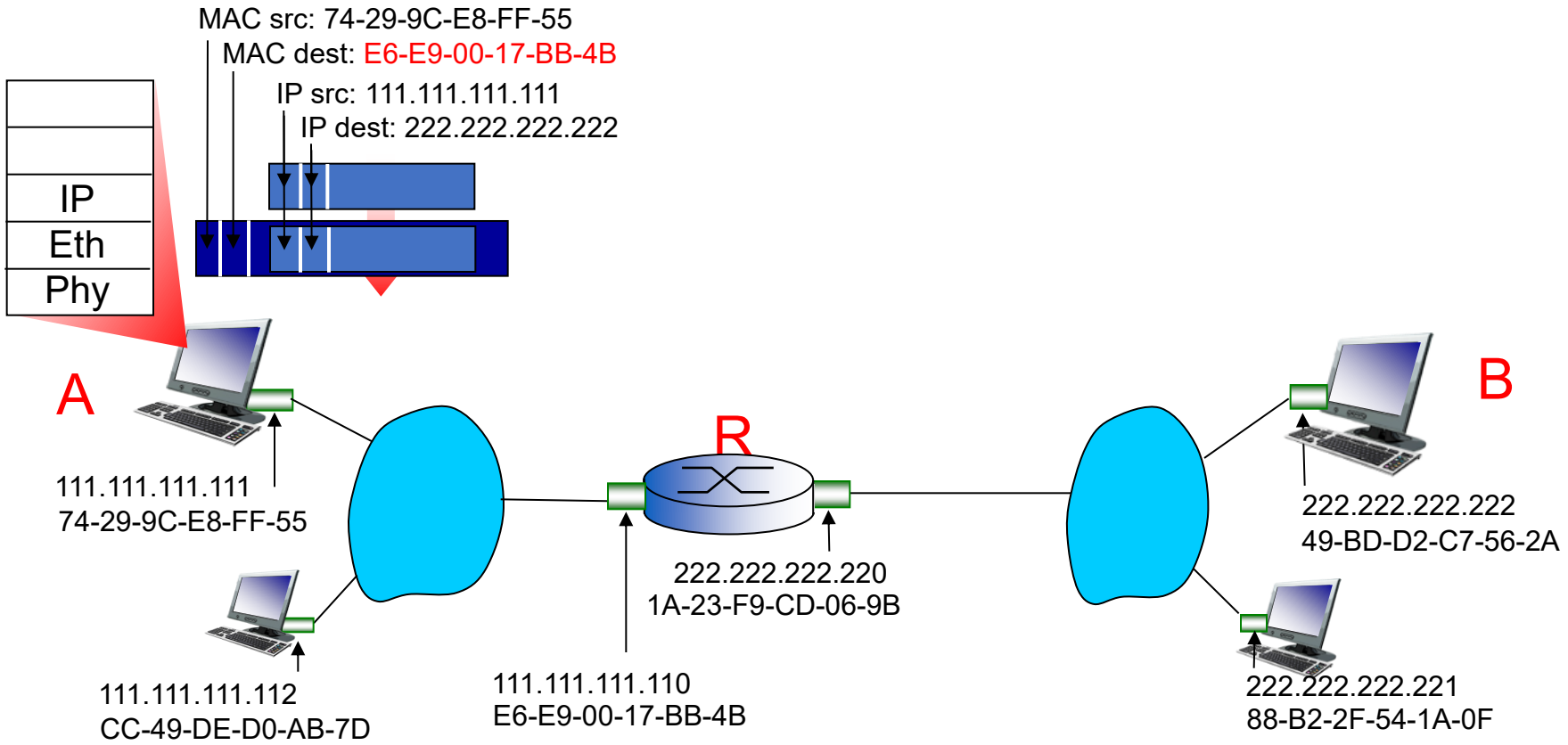
协议地址长度：值4

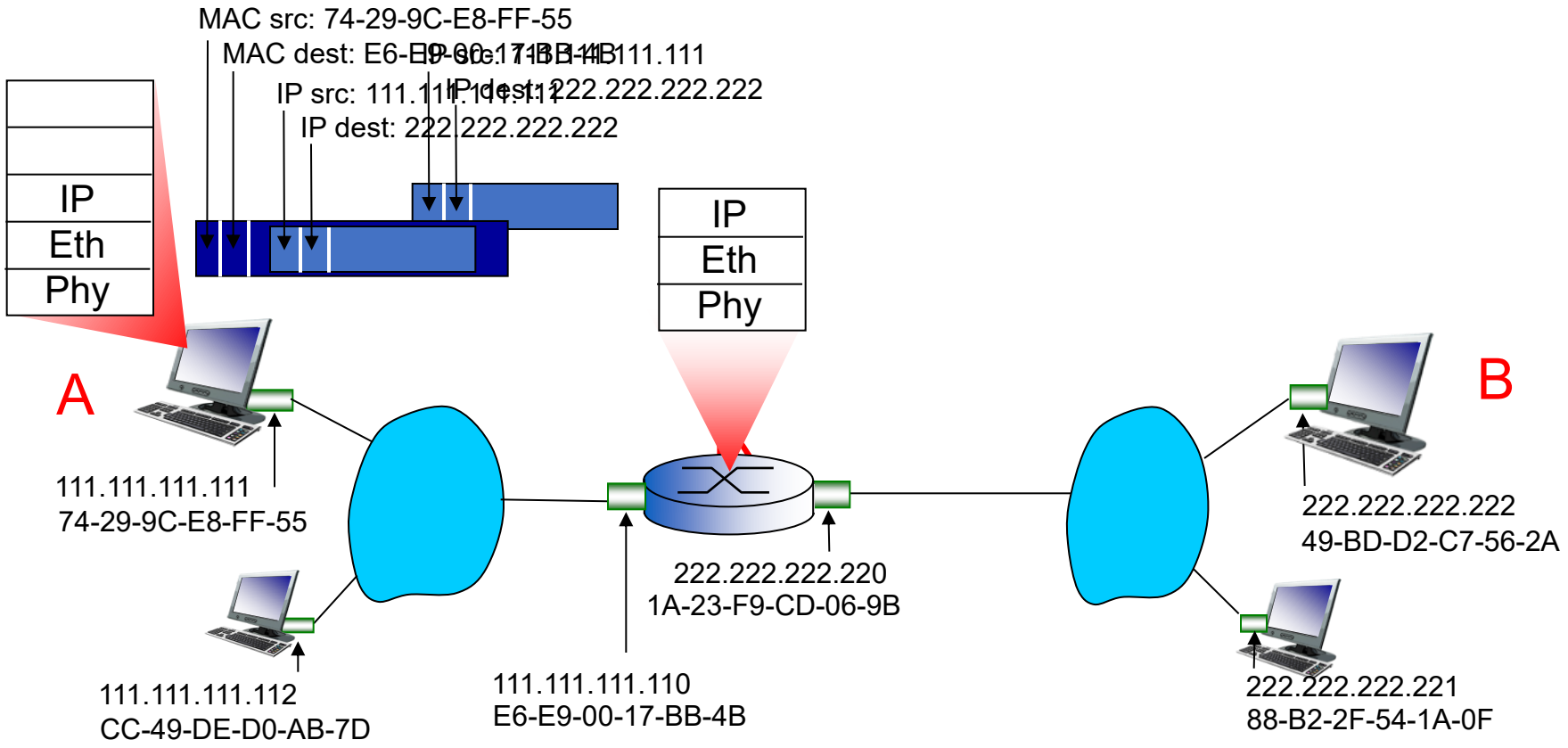
操作：请求1，响应2

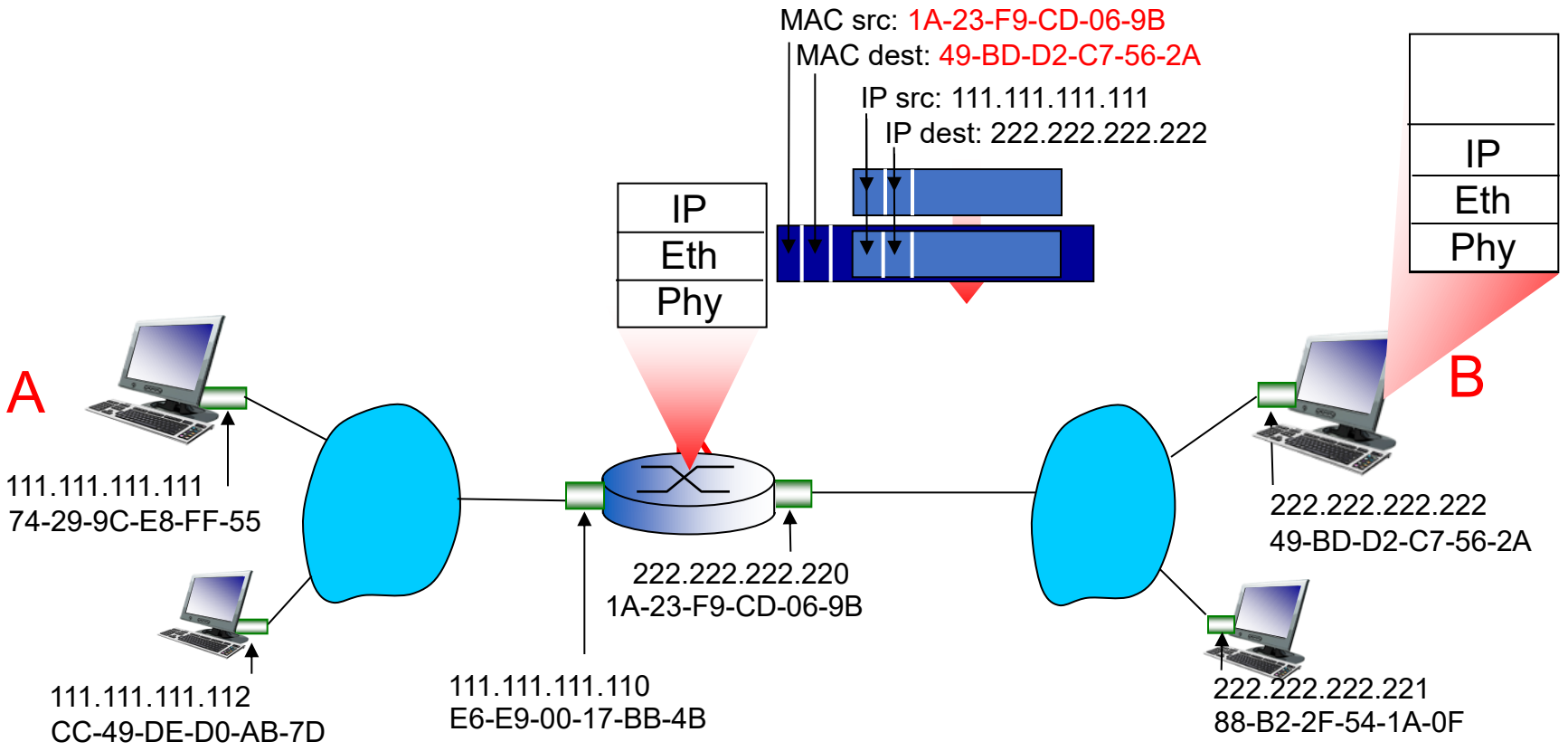
- 发送方硬件地址（以太网：6个字节）
- 发送方协议地址（IP：4个字节）
- 目标硬件地址（以太网：6个字节）
- 目标协议地址（IP：4个字节）

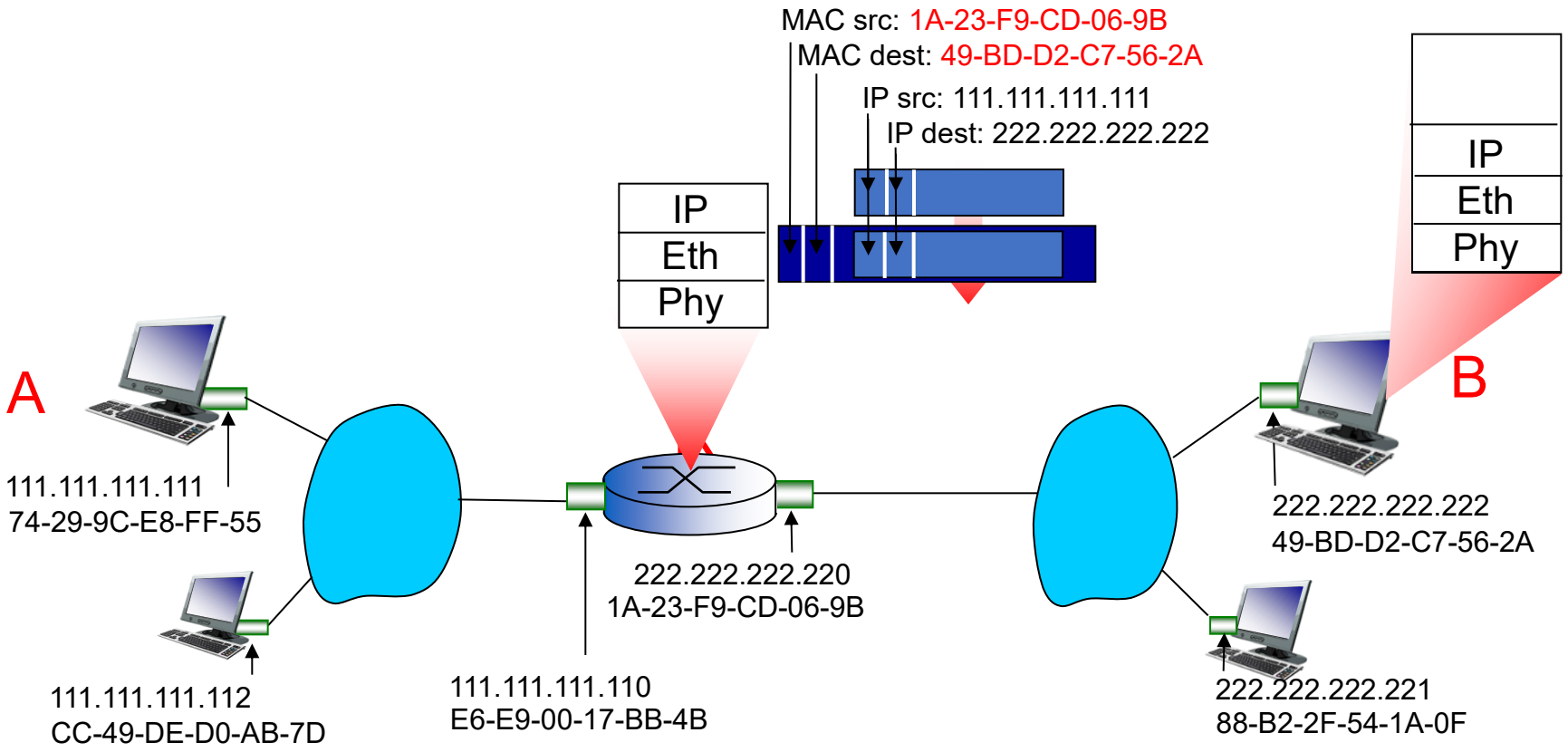
不在一个LAN?

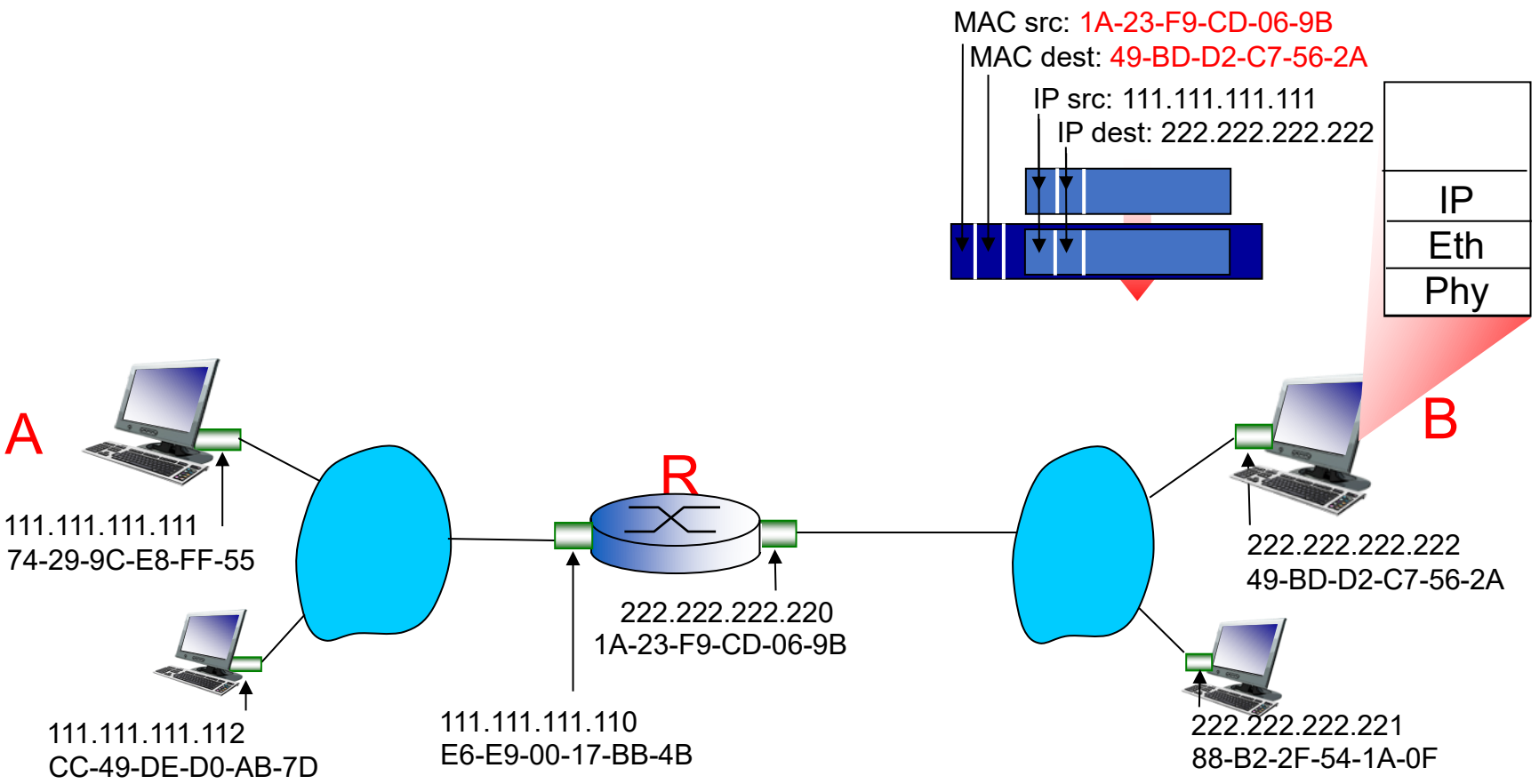

A
111.111.111.111
74-29-9C-E8-FF-55
111.111.111.112
CC-49-DE-D0-AB-7D
R
222.222.222.220
1A-23-F9-CD-06-9B
111.111.111.110
E6-E9-00-17-BB-4B
B
222.222.222.222
49-BD-D2-C7-56-2A
222.222.222.221
88-B2-2F-54-1A-0F


MAC src: 74-29-9C-E8-FF-55
MAC dest: E6-E9-00-17-BB-4B
IP src: 111.111.111.111
IP dest: 222.222.222.222
IP
Eth
Phy
A
B
R
111.111.111.111
74-29-9C-E8-FF-55
222.222.222.222
49-BD-D2-C7-56-2A
222.222.222.220
1A-23-F9-CD-06-9B
111.111.111.112
CC-49-DE-D0-AB-7D
111.111.111.110
E6-E9-00-17-BB-4B
222.222.222.221
88-B2-2F-54-1A-0F


MAC src: 74-29-9C-E8-FF-55
MAC dest: E6-E9-00-17-BB-4B
IP src: 111.111.111.111
IP dest: 222.222.222.222
IP
Eth
Phy
IP
Eth
Phy
A
B
111.111.111.111
74-29-9C-E8-FF-55
222.222.222.220
1A-23-F9-CD-06-9B
222.222.222.222
49-BD-D2-C7-56-2A
111.111.111.112
CC-49-DE-D0-AB-7D
111.111.111.110
E6-E9-00-17-BB-4B
222.222.222.221
88-B2-2F-54-1A-0F


MAC src: 1A-23-F9-CD-06-9B
MAC dest: 49-BD-D2-C7-56-2A
IP src: 111.111.111.111
IP dest: 222.222.222.222
IP
Eth
Phy
IP
Eth
Phy
A
B
111.111.111.111
74-29-9C-E8-FF-55
222.222.222.222
49-BD-D2-C7-56-2A
222.222.222.220
1A-23-F9-CD-06-9B
111.111.111.112
CC-49-DE-D0-AB-7D
111.111.111.110
E6-E9-00-17-BB-4B
222.222.222.221
88-B2-2F-54-1A-0F


MAC src: 1A-23-F9-CD-06-9B
MAC dest: 49-BD-D2-C7-56-2A
IP src: 111.111.111.111
IP dest: 222.222.222.222
IP
Eth
Phy
IP
Eth
Phy
A
B
111.111.111.111
74-29-9C-E8-FF-55
222.222.222.222
49-BD-D2-C7-56-2A
222.222.222.220
1A-23-F9-CD-06-9B
111.111.111.112
CC-49-DE-D0-AB-7D
111.111.111.110
E6-E9-00-17-BB-4B
222.222.222.221
88-B2-2F-54-1A-0F


MAC src: 1A-23-F9-CD-06-9B
MAC dest: 49-BD-D2-C7-56-2A
IP src: 111.111.111.111
IP dest: 222.222.222.222
IP
Eth
Phy
A
B
R
111.111.111.111
74-29-9C-E8-FF-55
111.111.111.112
CC-49-DE-D0-AB-7D
222.222.222.220
1A-23-F9-CD-06-9B
111.111.111.110
E6-E9-00-17-BB-4B
222.222.222.222
49-BD-D2-C7-56-2A
222.222.222.221
88-B2-2F-54-1A-0F

6.4 ICMP

✓Internet Control Message

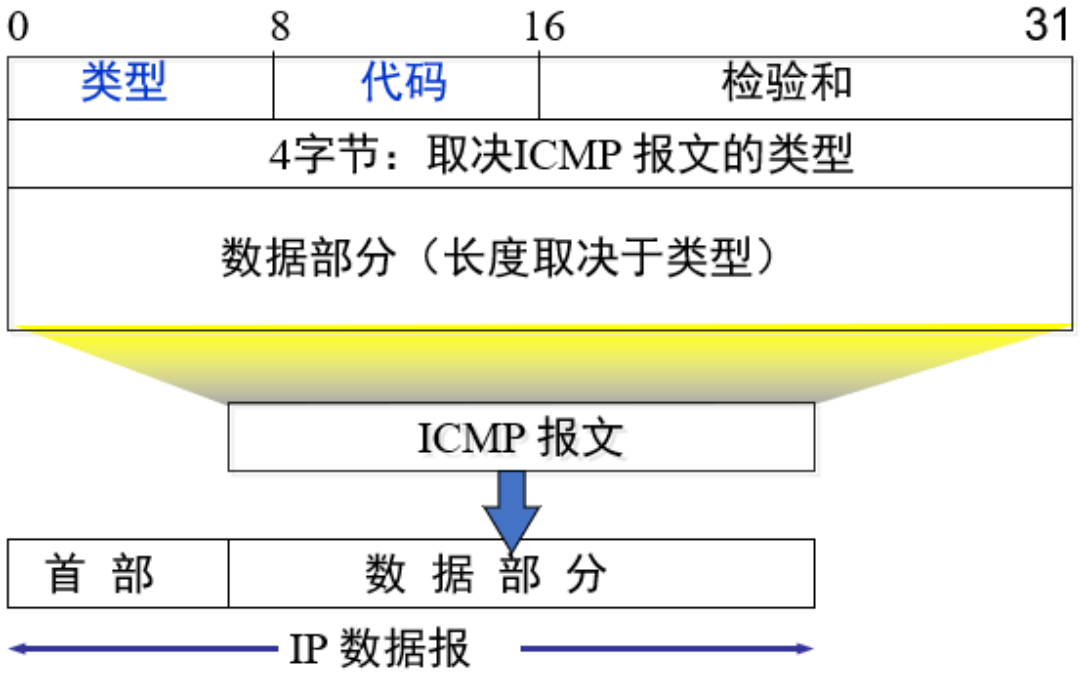
Internet 控制报文协议

✓差错报文：IP分组无法递交到目的主机（或对应进程上），路由器向源主机返回差错报文

✓查询报文：测试主机或路由器在网络层是否可达

ICMP

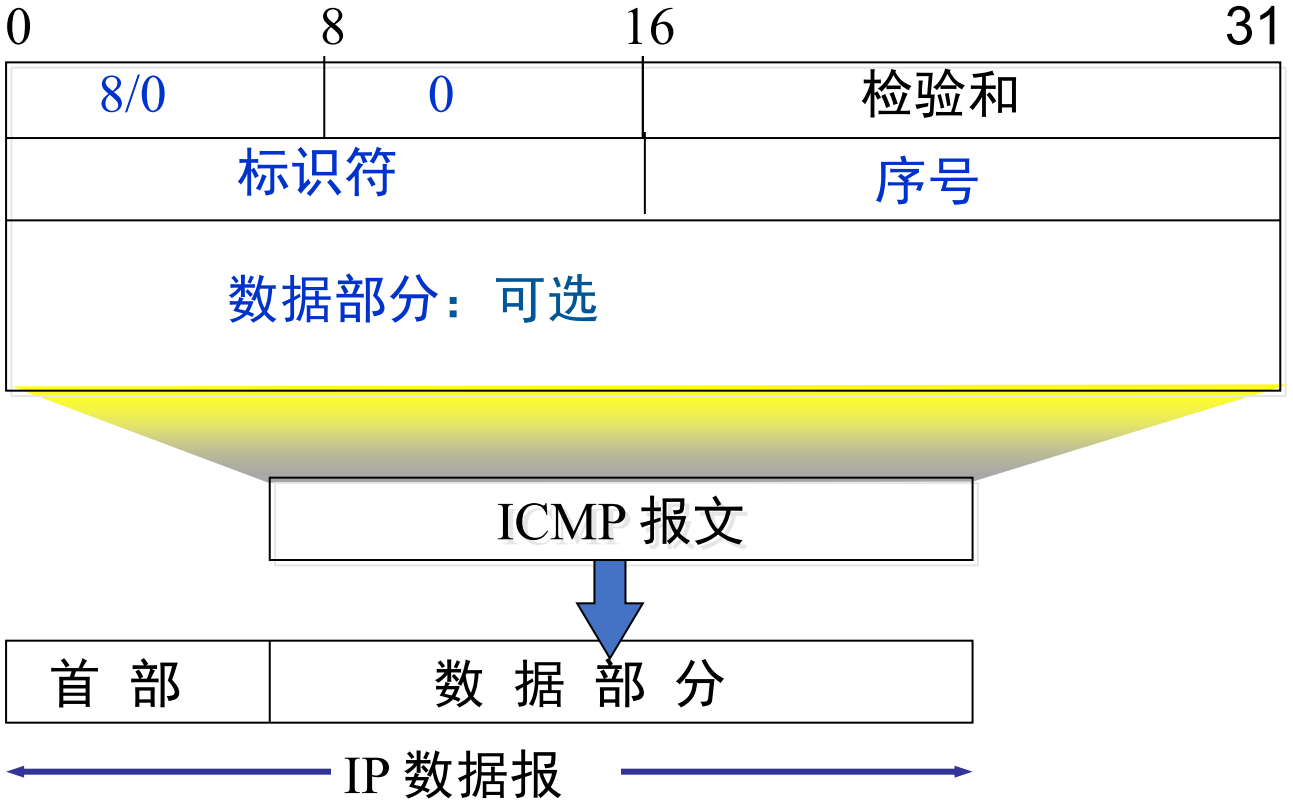
✓查询报文：测试主机或路由器在网络层是否可达



ICMP报文类型	类型对应的值	ICMP报文的类型
差错报告报文	3	终点不可达：路由器或者主机发现信息不可达时
	11	时间超时（TTL）：当路由器收到TTL为1，然后-1后是0
	12	参数问题
	5	改变路由：默认路由器发现有跟好的路由线路
询问报文	8或者0	回送请求或回答
	13或者14	时间戳请求或回答

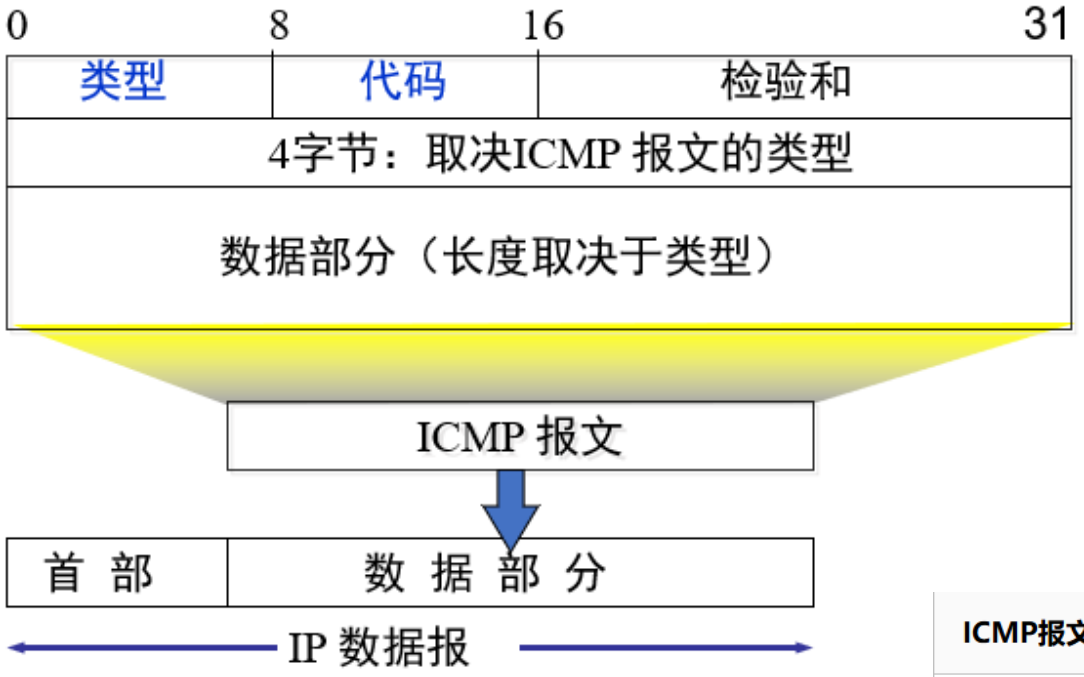
✓回送请求，type：8，code：0

✓回送应答，type：0，code：0



- Ping: ICMP回送请求与应答报文

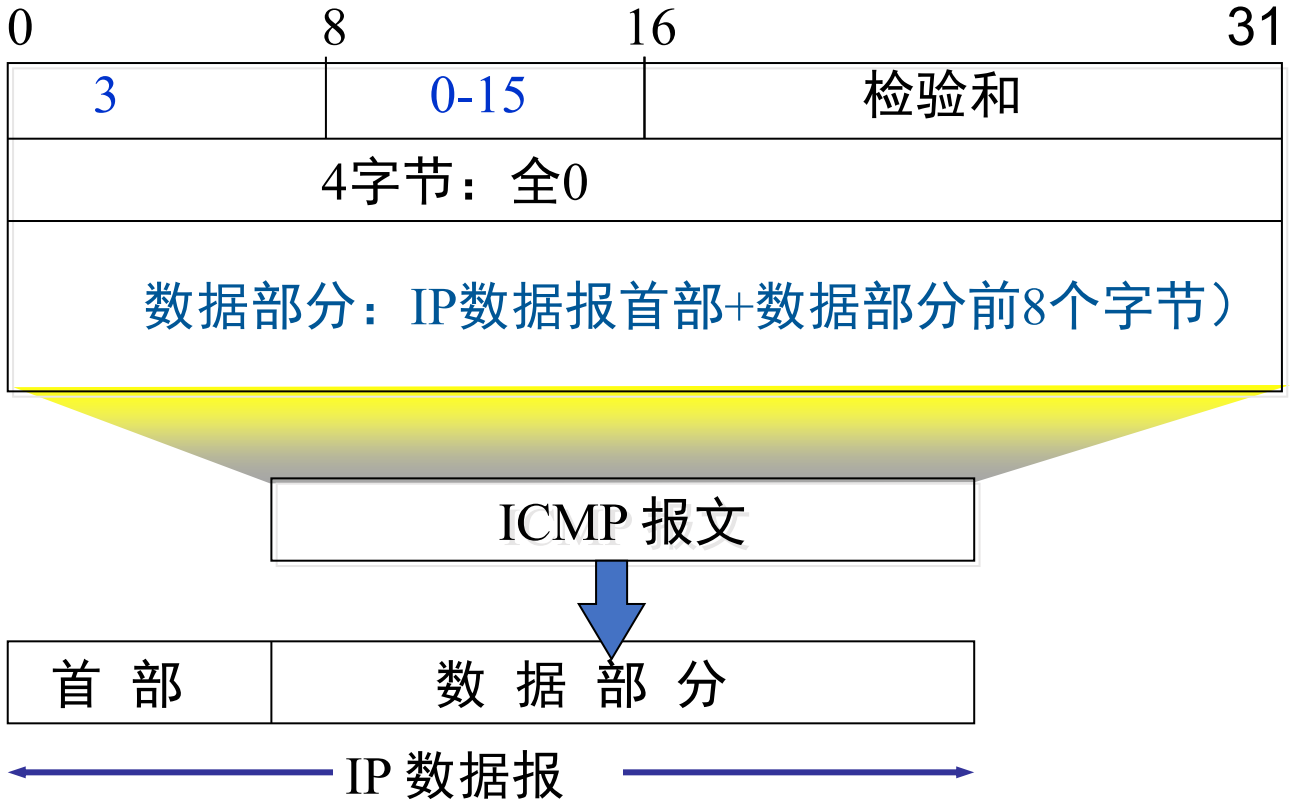
```
C:\>ping www.ustc.edu.cn

正在 Ping www.ustc.edu.cn [202.38.64.246] 具有 32 字节的数据
来自 202.38.64.246 的回复: 字节=32 时间=68ms TTL=44
来自 202.38.64.246 的回复: 字节=32 时间=61ms TTL=44
来自 202.38.64.246 的回复: 字节=32 时间=71ms TTL=44
来自 202.38.64.246 的回复: 字节=32 时间=61ms TTL=44

202.38.64.246 的 Ping 统计信息:
    数据包: 已发送 = 4, 已接收 = 4, 丢失 = 0 (0% 丢失)
往返行程的估计时间(以毫秒为单位):
    最短 = 61ms, 最长 = 71ms, 平均 = 65ms
```



ICMP报文类型	类型对应的值	ICMP报文的类型
差错报告报文	3	终点不可达：路由器或者主机发现信息不可达时
	11	时间超时（TTL）：当路由器收到TTL为1，然后-1后是0
	12	参数问题
	5	改变路由：默认路由器发现有跟好的路由线路
询问报文	8或者0	回送请求或回答
	13或者14	时间戳请求或回答

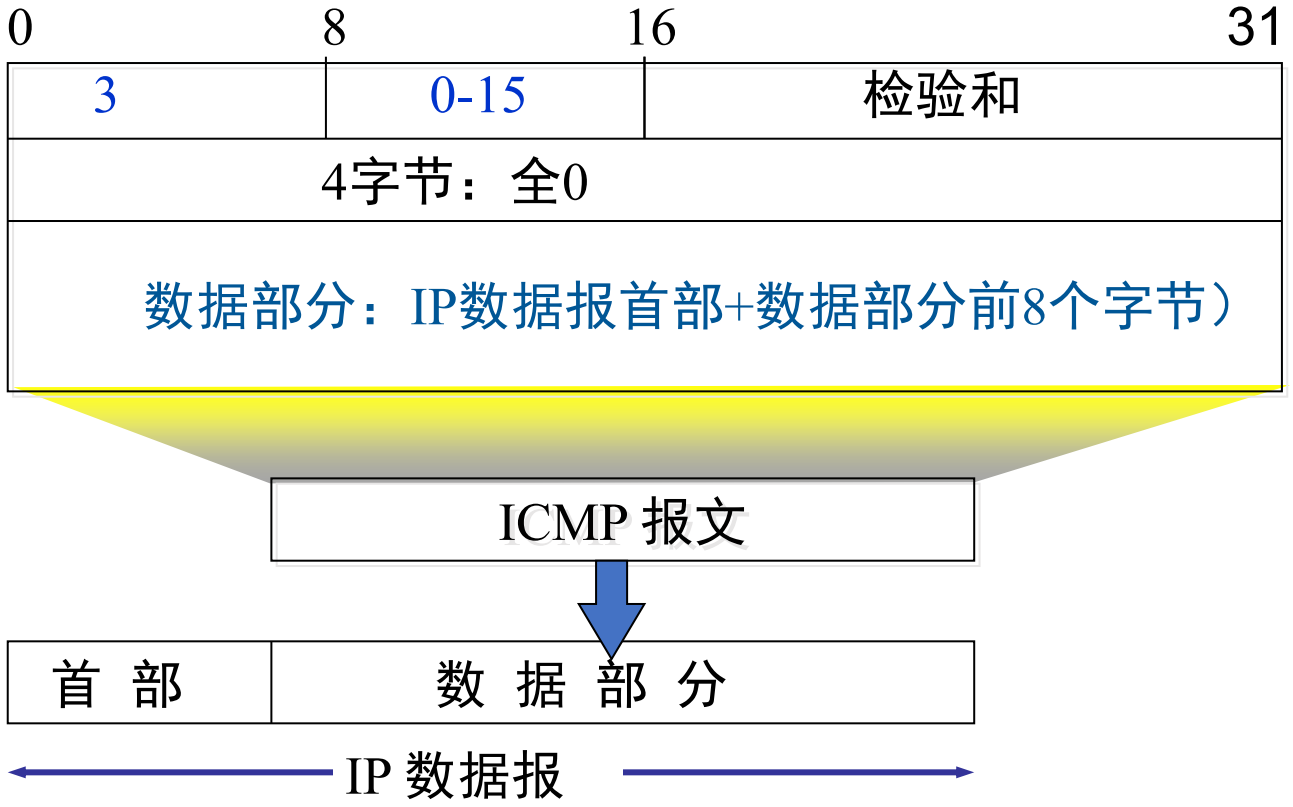
✓不可达，type：3

- 网络不可达：code 0
- 主机不可达：code 1
- 协议不可达：code 2
- 端口不可达：code 3
- …..



- Tracert/Traceroute: 探测两个主机/间的路由

  - ✓ ICMP：TTL超时差错报文
  - ✓ ICMP：端口不可达差错报文



- Tracert/Traceroute: 探测两个主机/间的路由

  - ✓ ICMP：TTL超时差错报文
  - ✓ ICMP：端口不可达差错报文

```
C:\>tracert www.ustc.edu.cn

通过最多 30 个跃点跟踪
到 www.ustc.edu.cn [202.38.64.246] 的路由:

  1     7 ms     1 ms     4 ms  192.168.0.1
  2     *        *        *     请求超时。
  3    47 ms    38 ms    39 ms  172.21.1.1
  4    41 ms    41 ms    36 ms  10.138.211.66
  5     *        *        *     请求超时。
  6     *       54 ms     *     120.193.85.169
  7     *       37 ms     *     221.183.48.45
  8     *        *        *     请求超时。
  9    77 ms     *       36 ms  221.183.90.150
 10    62 ms    65 ms    45 ms  221.183.151.254
 11    99 ms    65 ms    78 ms  101.4.115.186
 12    98 ms    71 ms    86 ms  101.4.115.14
 13    95 ms    83 ms    74 ms  210.45.224.60
 14     *        *        *     请求超时。
 15    95 ms    73 ms    58 ms  202.38.64.246

跟踪完成。
```

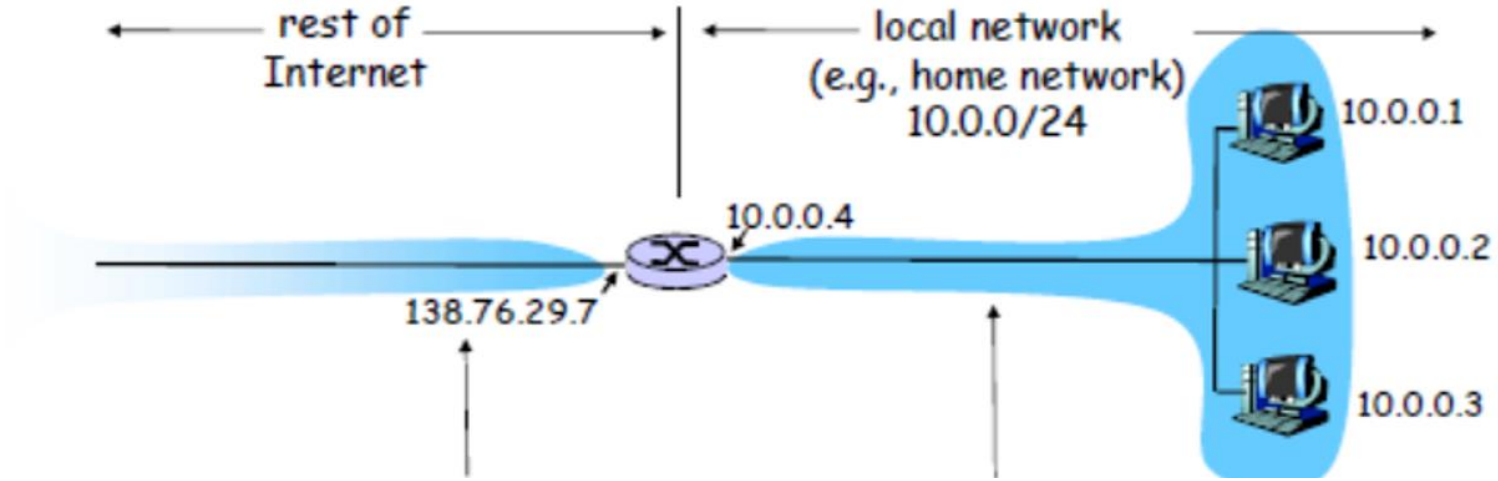
NAT

- 内部网络（private：专有）地址

仅在机构内部使用的 IP 地址，可以由本机构自行分配，不需要向互联网管理机构申请（仅在机构内网中有意义，区分不同的设备）

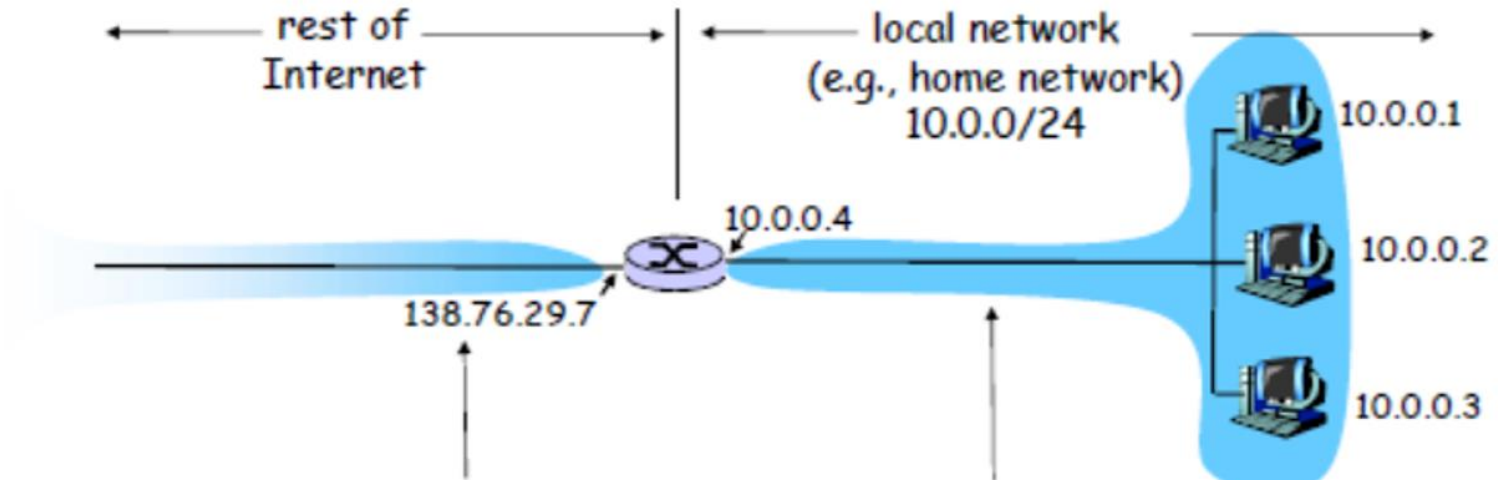
- Class A 10.0.0.0-10.255.255.255　　MASK 255.0.0.0
- Class B 172.16.0.0-172.31.255.255　　MASK 255.255.0.0
- Class C 192.168.0.0-192.168.255.255　　MASK 255.255.255.0

- 局域网内部主机对外不可见
- 路由器不对外转发源地址为内部IP地址的分组，确保不会出现重复的IP分组



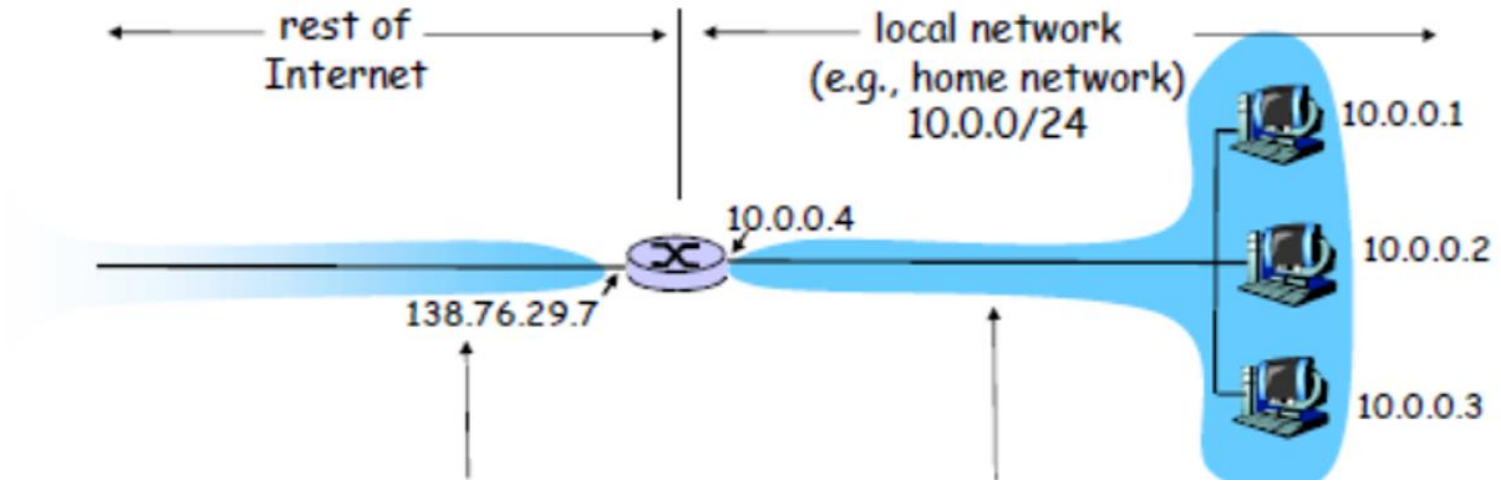
NAT: Network Address Translation

NAT 转换表	
WAN side addr	LAN side addr
138.76.29.7	10.0.0.1
……	……



本地网络只有1个或几个外部IP地址，不需要从ISP申请一个地址块

NAT 转换表	
WAN side addr	LAN side addr
138.76.29.7	10.0.0.1
……	……



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

NAPT（Network Address and Port Translation），提高公网地址的利用率

- 路由器转发工作在网络层，使用了传输层的端口号

NAT/NAPT：

- 解决了内网访问外网的问题
- 外网到内网的访问？——NAT穿越

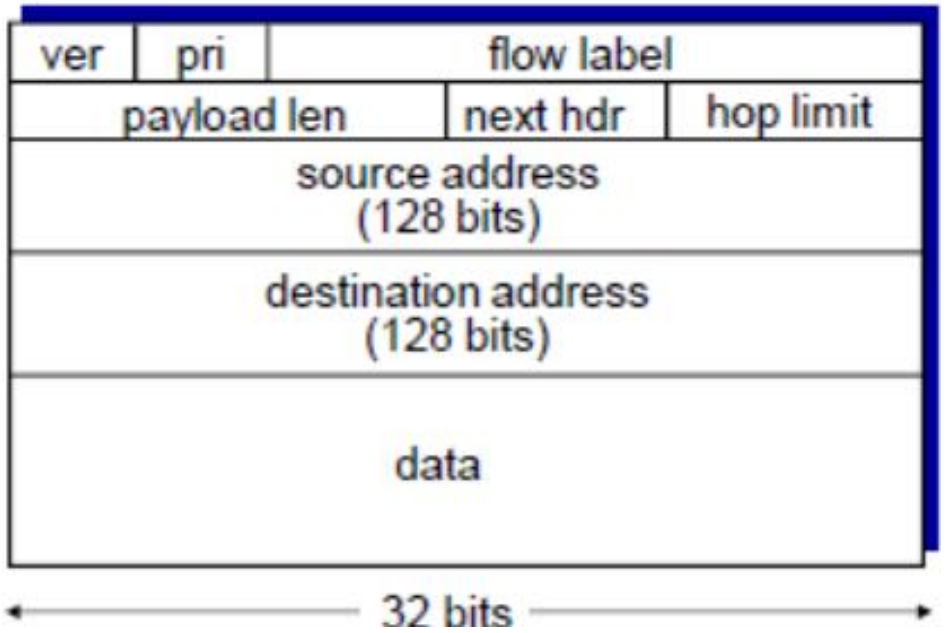
IPv6

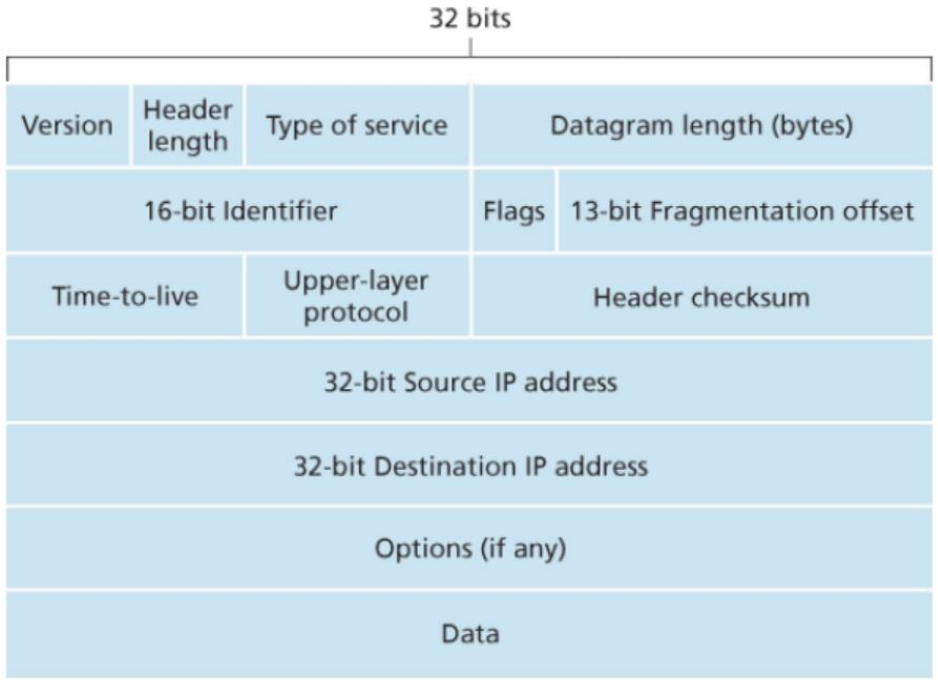
20世纪90年代：Internet工程任务组开始研究开发一种替代IPv4的协议

- 32-bit的地址空间即将用尽（物联网、车联网，海量设备联网）
- IPv6：128-bit 的地址空间

——简化头部字段，提升转发速度

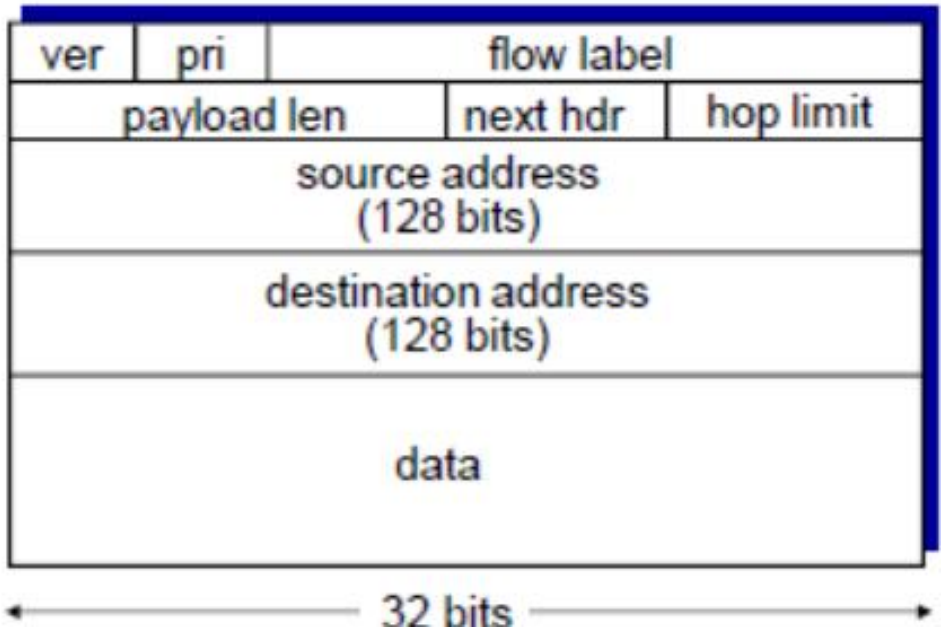
——默认支持Ipsec协议

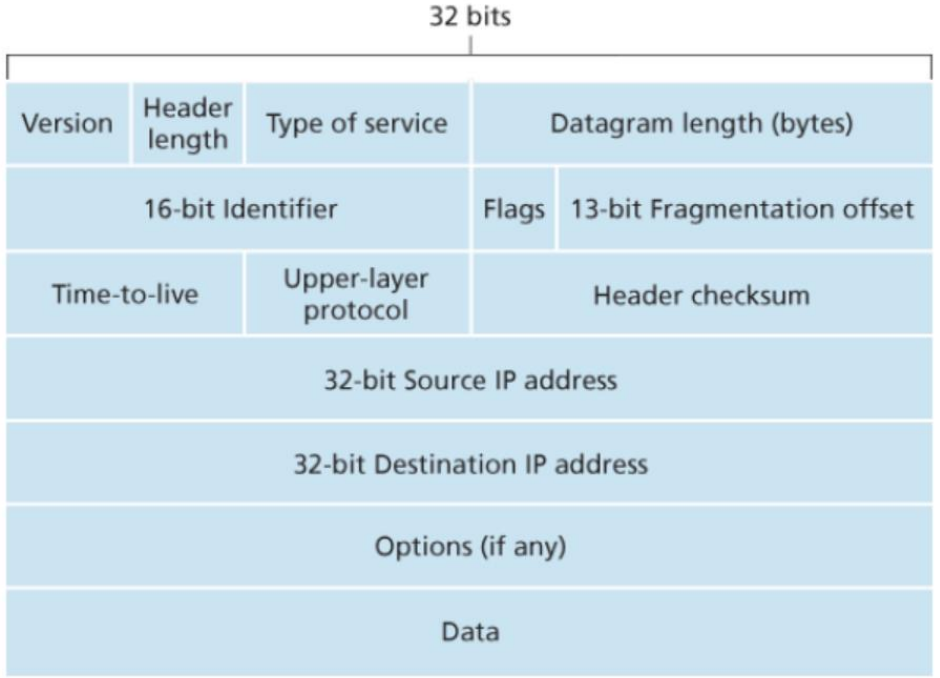




固定：40字节

- 源/目的ip地址： 16字节
- Pri（Traffic Class）：区分数据报优先级
- Flow label：**标记同一 “流” 的数据包**
- Payload Length：数据长度
- **Next Header**： 协议类型/IP首部扩展
- **Hop Limit**： TTL

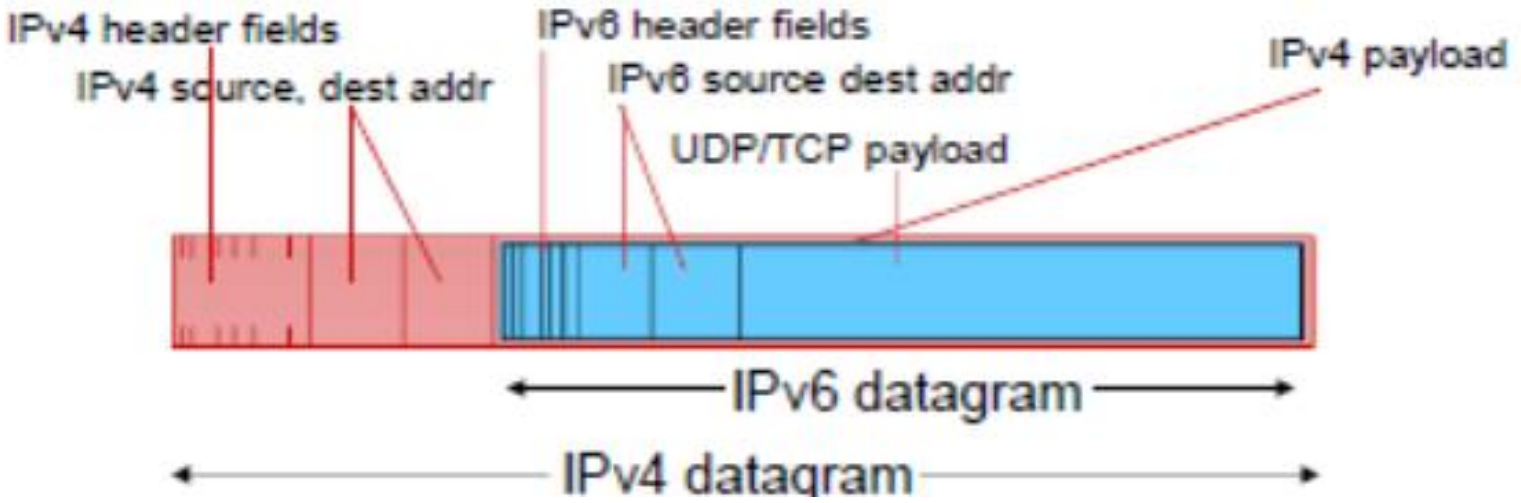




- 取消分片，不允许分片
- 取消checksum，加快处理速度

IPv4到IPv6的升级

- 数十亿设备同时升级？（资金和人力成本）
- 多种过渡解决方案，比如隧道技术：，在IPv4路由器之间传输的IPv4数据报中携带IPv6数据报



相比应用层的变化，网络层改变要慢很多，IPv6的广泛部署需要很长的时间

作业

P249：5.1

P250：5.8，5.11，5.12，5.13