6.4

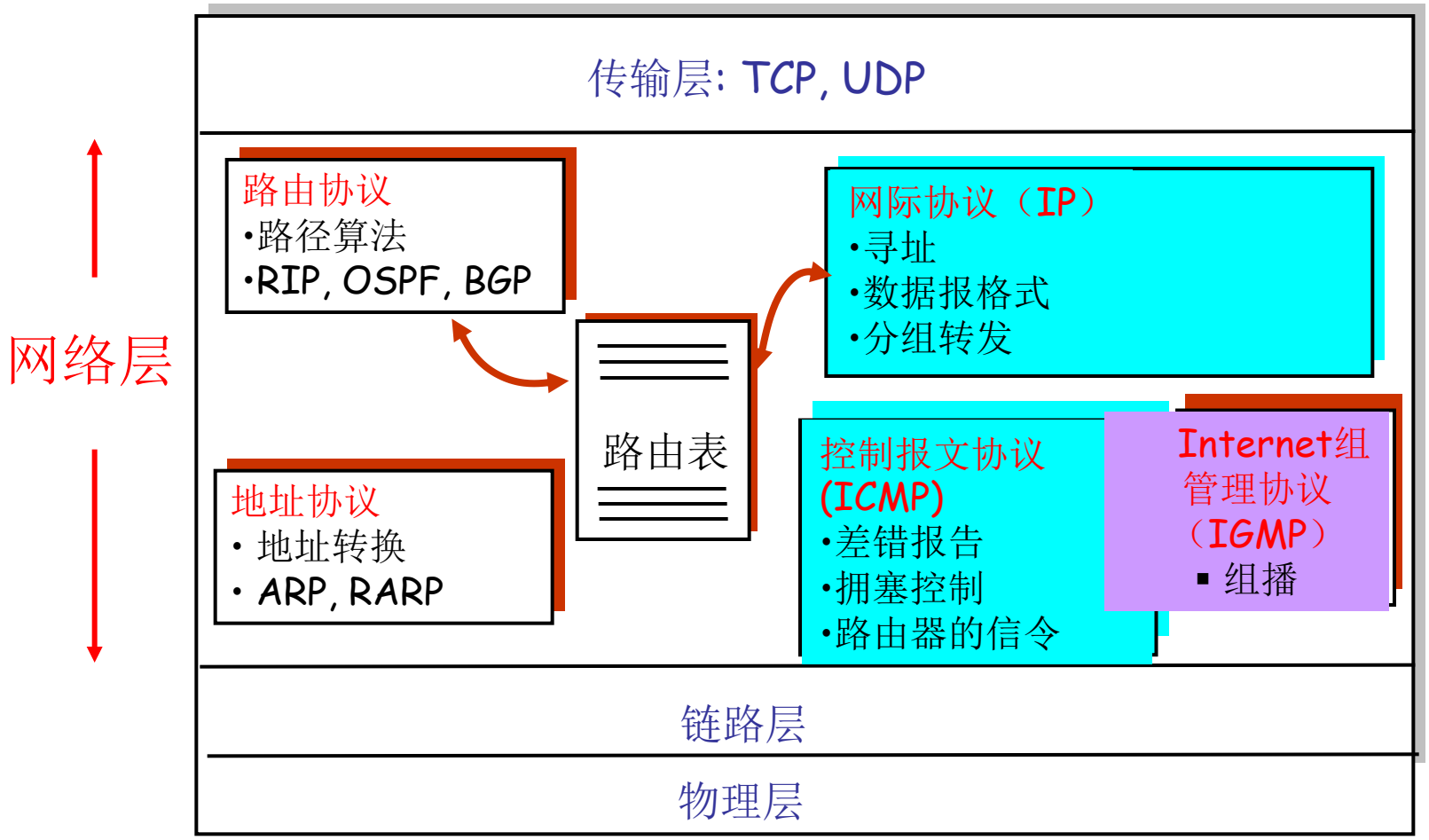
因特网网络层协议

设计、制作、讲授：谭献海

EMAIL：xhtan@home.swjtu.edu.cn

Internet 网络层

主机, 路由器 网络层功能:



6.4 Internet网络层协议

1. 报文传送----IP

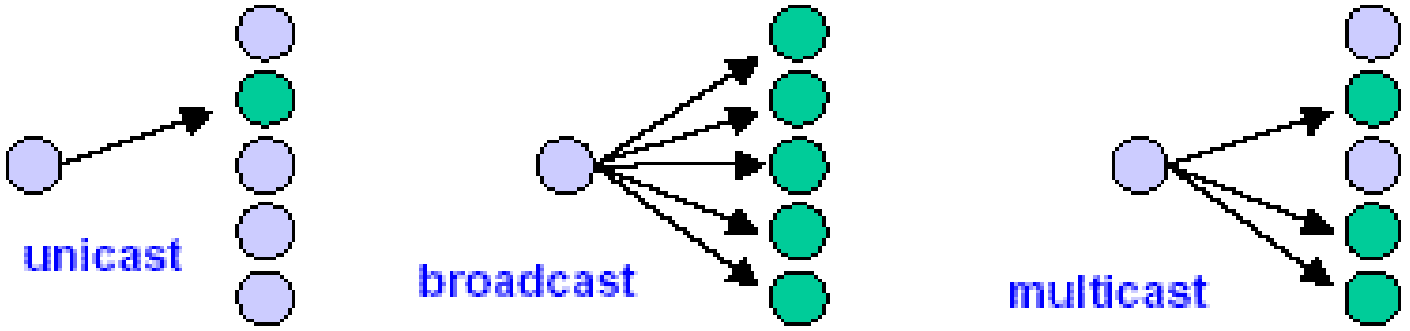
2. 网络控制----ICMP

3. 其他Internet网络层协议

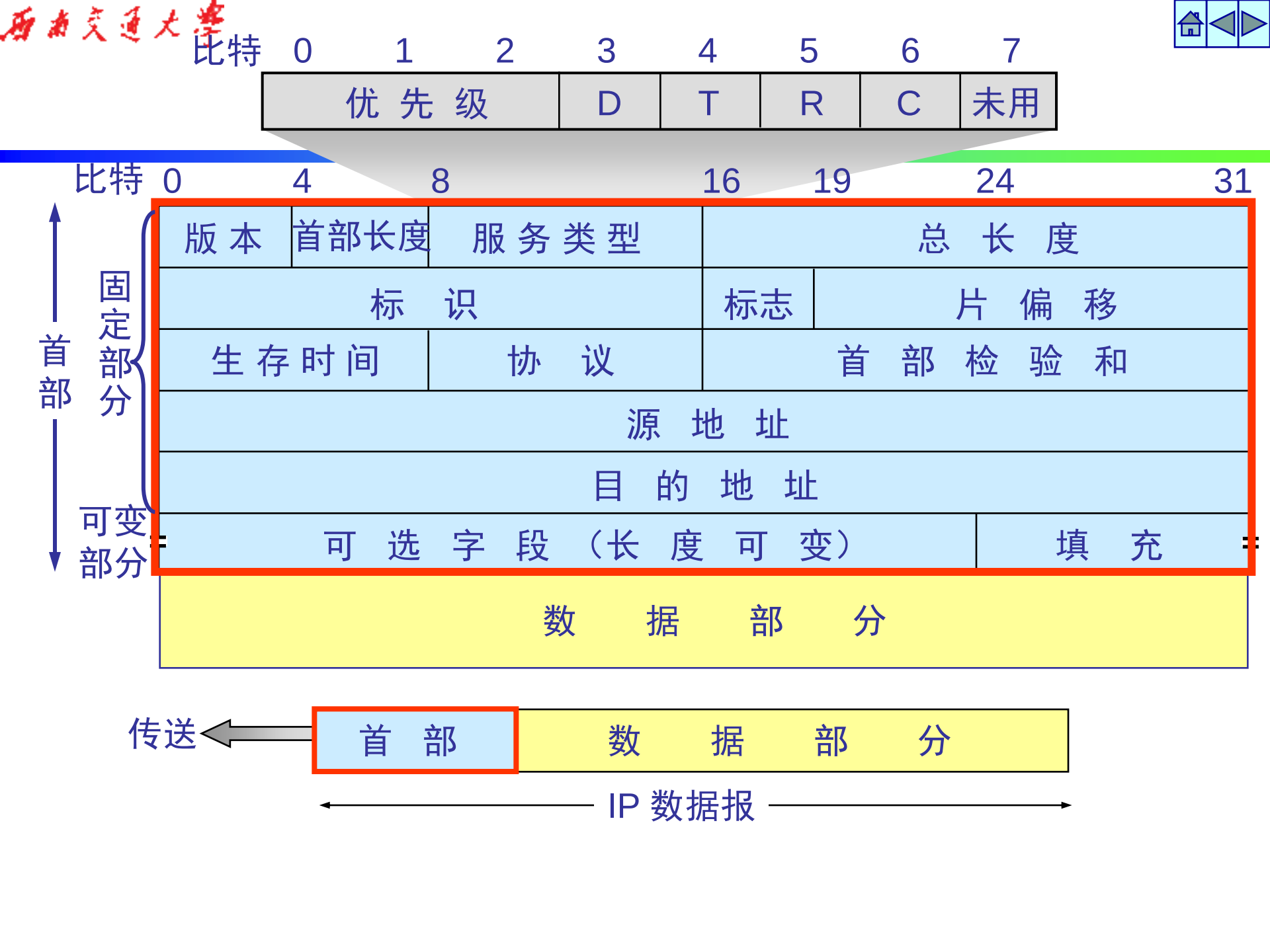
IP 服务

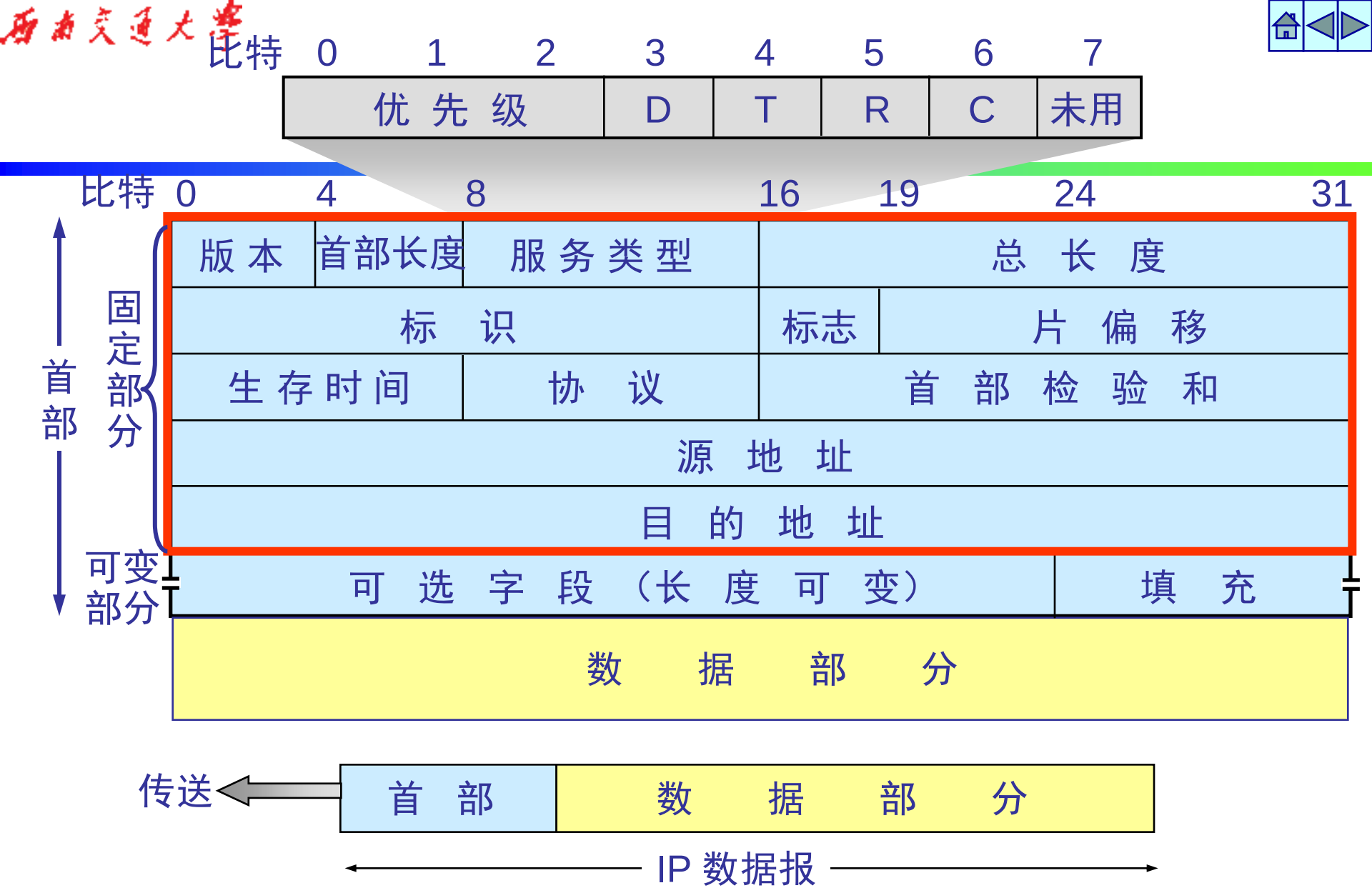
IP 支持以下的服务类型:

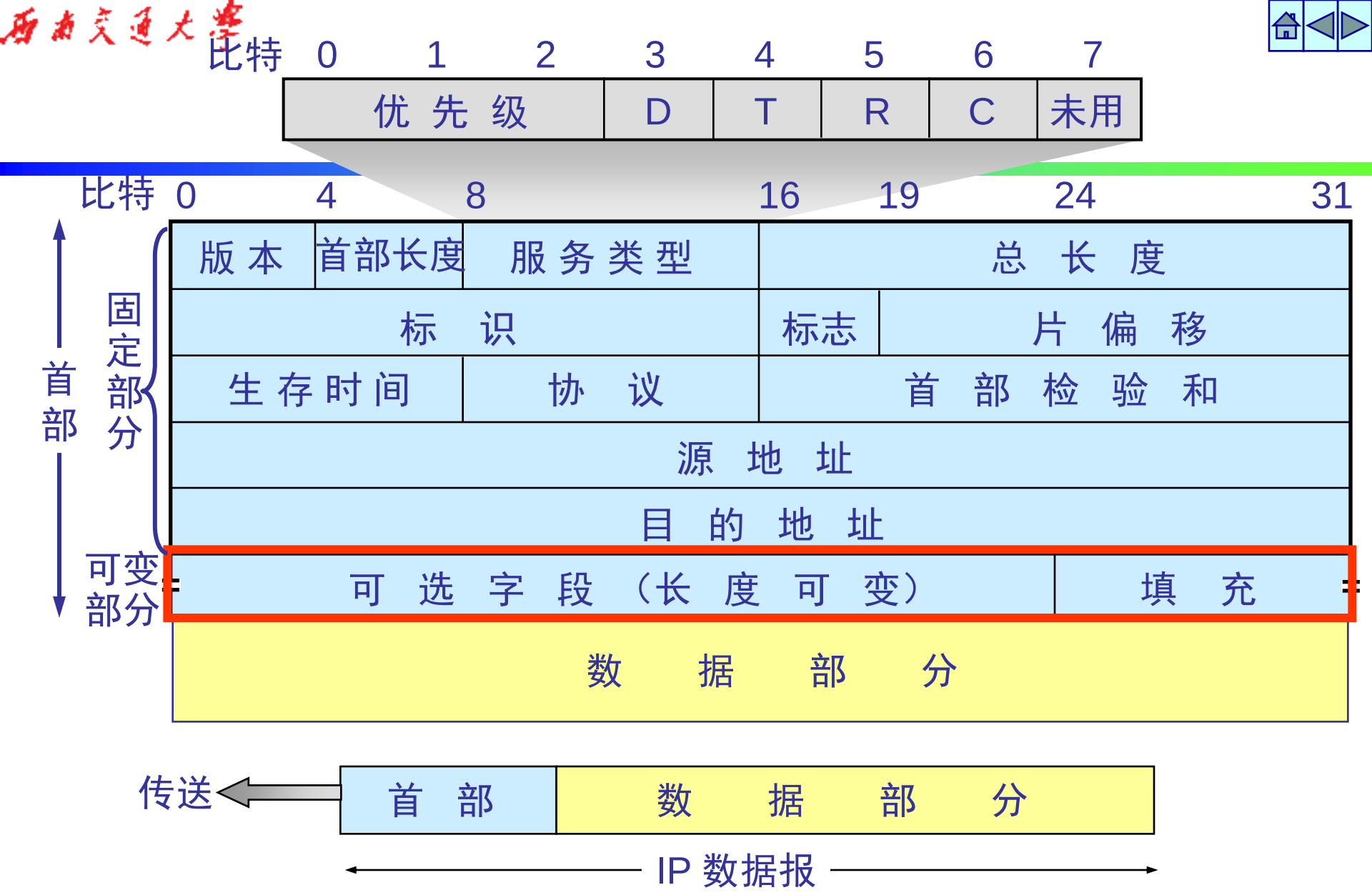
- 一到一 (单播)
- 一到所有 (广播)
- 一到多 (组播)



- IP 组播也支持多到多的服务
- IP 组播需要其他协议的支持(如IGMP)

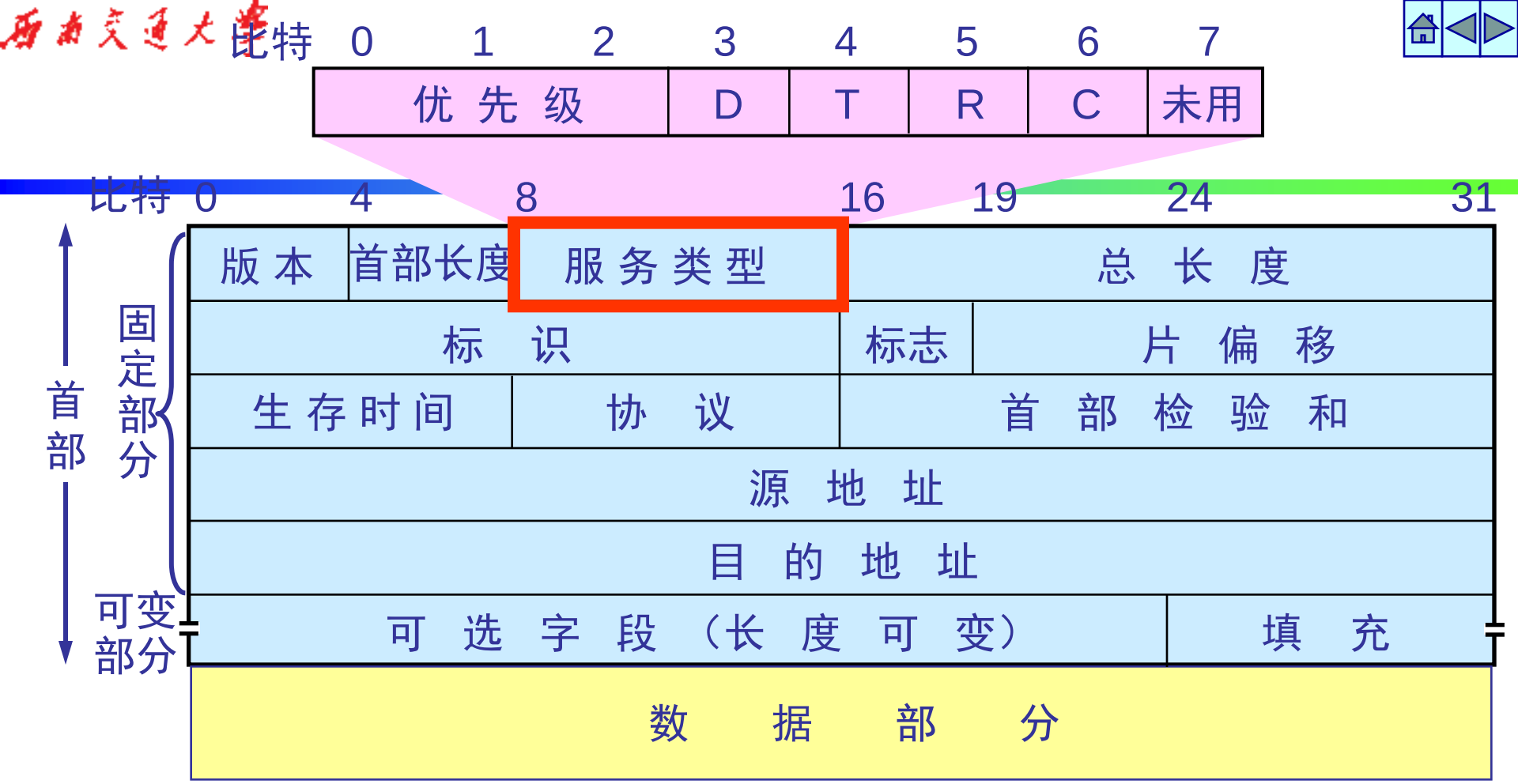

比特 0 1 2 3 4 5 6 7
优 先 级
D
T
R
C
未用
比特 0 4 8 16 19 24 31
首部
固定部分
可变部分
版 本
首部长度
服 务 类 型
总 长 度
标 识
标志
片 偏 移
生 存 时 间
协 议
首 部 检 验 和
源 地 址
目 的 地 址
可 选 字 段 （长 度 可 变）
填 充
数 据 部 分
传送
首 部
数 据 部 分
IP 数据报


比特 0 1 2 3 4 5 6 7
优 先 级
D
T
R
C
未用
比特 0 4 8 16 19 24 31
版 本
首部长度
服 务 类 型
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
固定部分
可变部分
传送
首 部
数 据 部 分
IP 数据报


比特 0 1 2 3 4 5 6 7
优 先 级
D
T
R
C
未用
比特 0 4 8 16 19 24 31
版 本
首部长度
服 务 类 型
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
固定部分
可变部分
传送
首 部
数 据 部 分
IP 数据报

版本——占 4 bit，指IP协议的版本
目前的 IP 协议版本号为 4 (即 IPv4)

首部长度——占 4 bit，以32比特(4 字节)为单位.
IP 首部长度的最小值是5。



服务类型——占 8 bit，用来表示服务质量
在区分服务（DiffServ）中使用

服务类型 (TOS)

0	1	2	3	4	5	6	7
优先级			D	T	R	C	0

优先级的说明如下：

111 - Network Control

110 - Internetwork Control

101 - CRITIC/ECP

100 - Flash Override

011 - Flash

010 - Immediate

001 - Priority

000 - Routine

服务类型 (TOS)

服务类型 (TOS): 包括 4个 TOS 比特, 每个比特表示一个需要的服务。

- 最小时延 (D)
- 最大吞吐量 (T)
- 最高可靠性 (R)
- 最小费用 (C)

--各比特：好----1；差---0

--每次只能设定一个比特。

--并不是所有的应用都支持服务类型。

--RFC1340描述标准应用如何使用TOS设置

--RFC1349进一步描述了TOS的特性

服务类型 (TOS)

下图列出了对不同应用建议的TOS值。

Telnet和Rlogin这两个交互应用要求最小的传输时延，因为人们主要用它们来传输少量的交互数据。

FTP文件传输则要求有最大的吞吐量。

网络管理（SNMP）和路由选择协议要求最高可靠性。

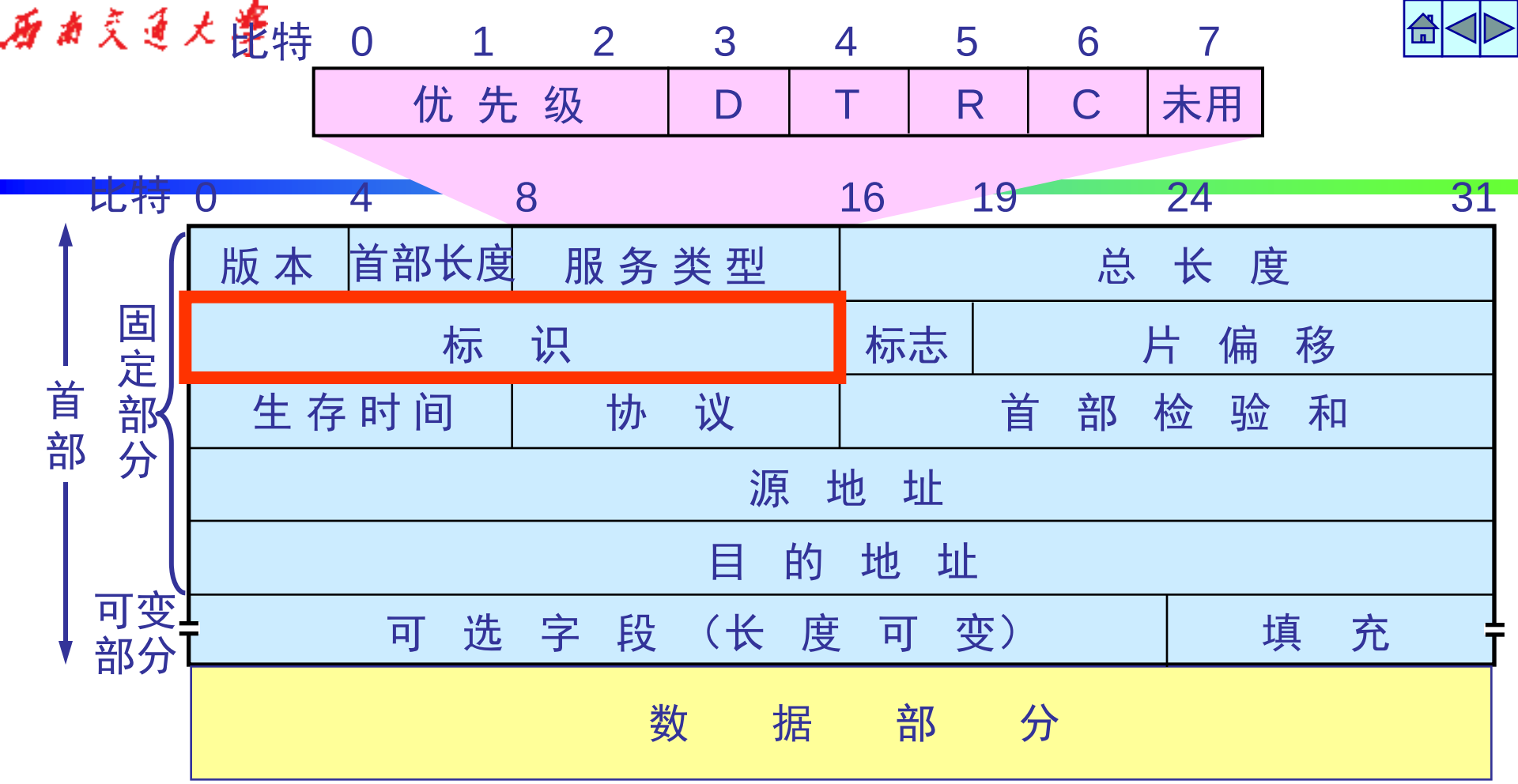
用户网络新闻（Usenet news, NNTP）是唯一要求最小费用的应用。

现在大多数的TCP/IP实现都不支持TOS特性。

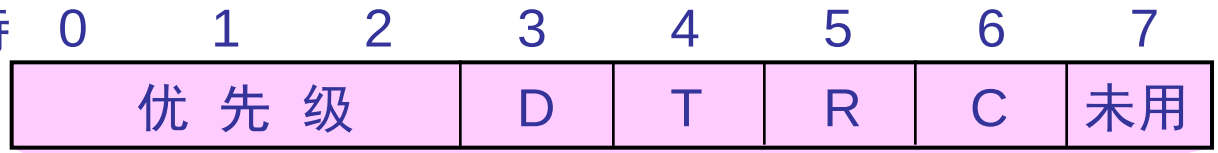
应用程序	最小时延	最大吞吐量	最高可靠性	最小费用	16进制值
Telnet/Rlogin	1	0	0	0	0x10
FTP					
控制	1	0	0	0	0x10
数据	0	1	0	0	0x08
任意块数据	0	1	0	0	0x08
TFTP	1	0	0	0	0x10
SMTP					
命令阶段	1	0	0	0	0x10
数据阶段	0	1	0	0	0x08
DNS					
UDP查询	1	0	0	0	0x10
TCP查询	0	0	0	0	0x00
区域传输	0	1	0	0	0x08
ICMP					
差错	0	0	0	0	0x00
查询	0	0	0	0	0x00
任何IGP	0	0	1	0	0x04
SNMP	0	0	1	0	0x04
BOOTP	0	0	0	0	0x00
NNTP	0	0	0	1	0x02

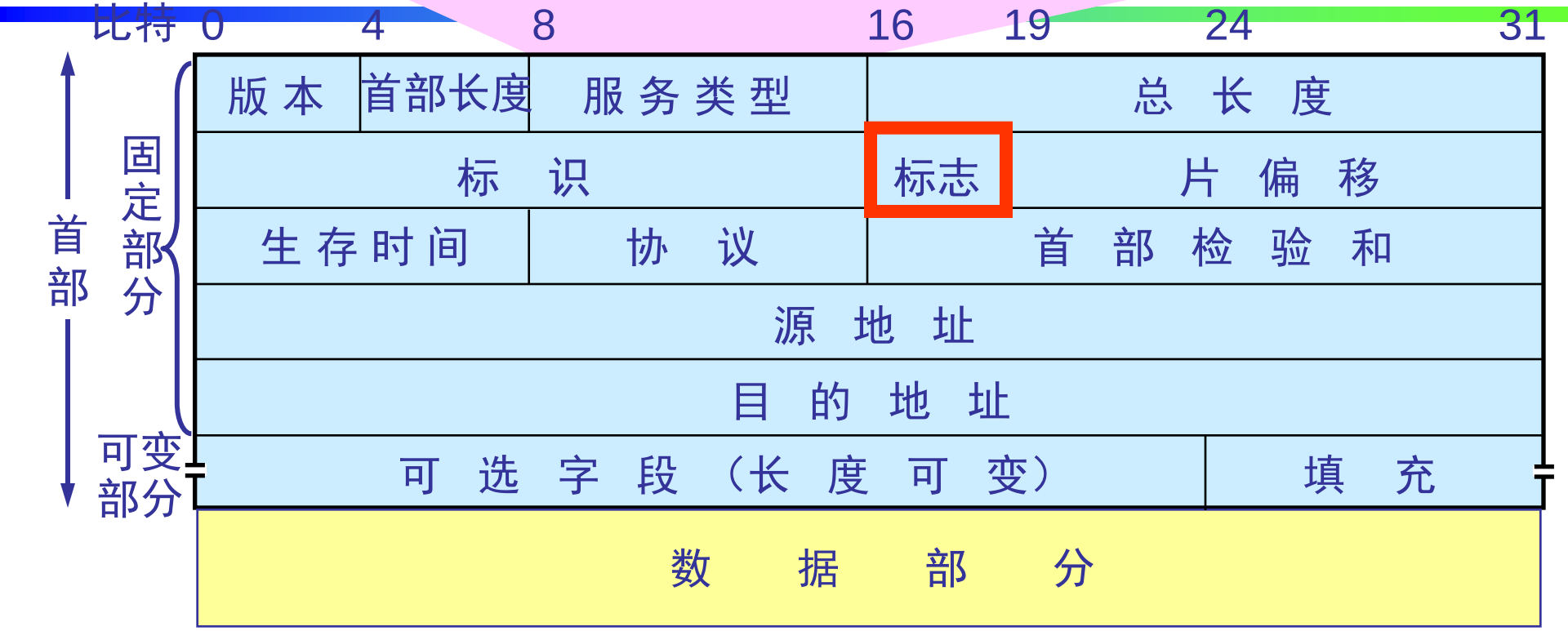
总长度——占 16 bit，指首部和数据之和的长度，单位为字节，因此数据报的最大长度为 65535 字节。

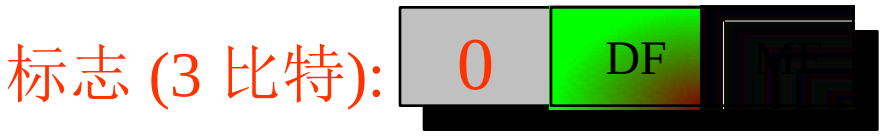
总长度必须不超过最大传送单元 MTU。



标识(identification),即分组序号,占 16 bit,
它是一个计数器，用来表示数据报的序号







DF：不能分片 （Don’t Fragment)

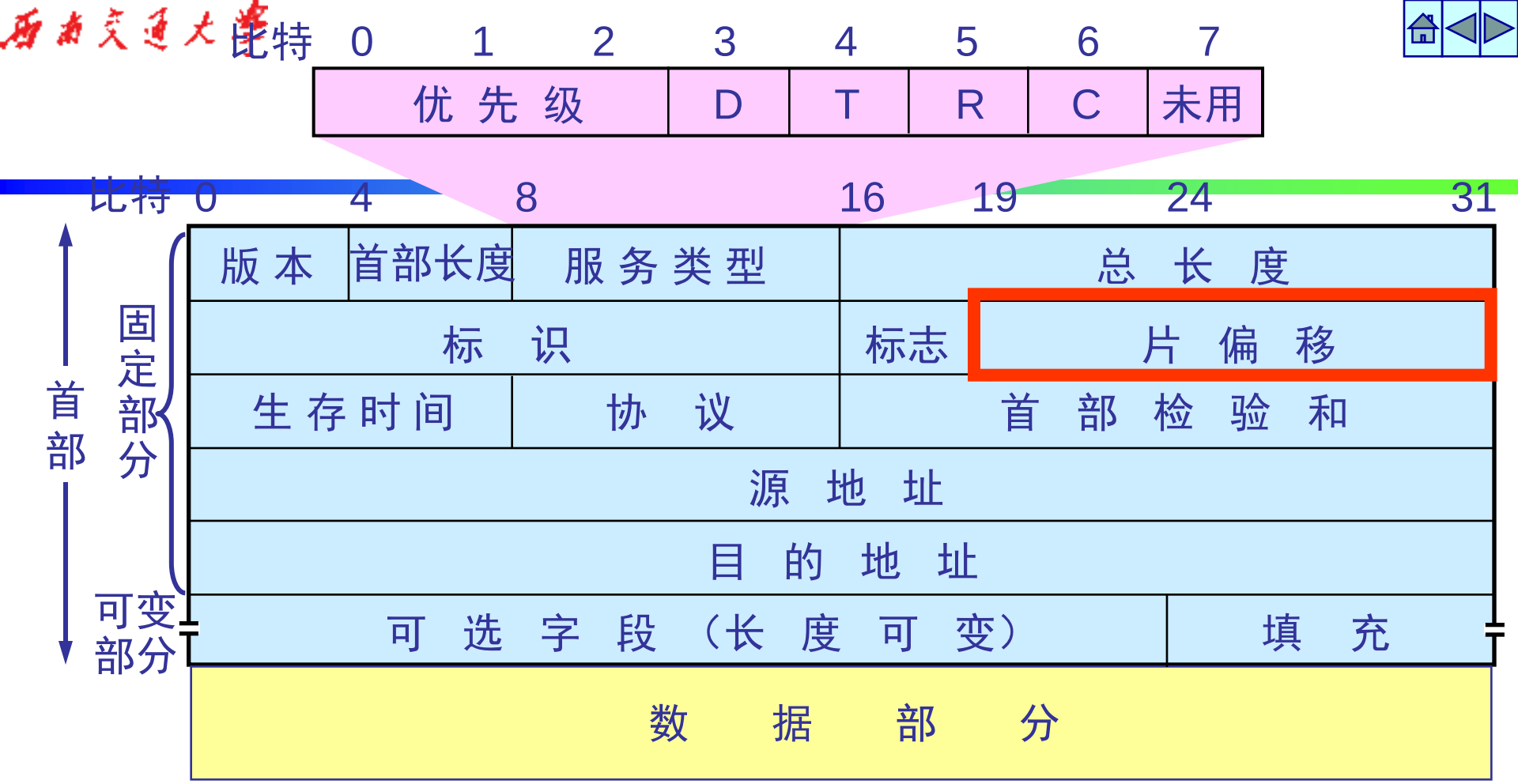
DF=1：不能分片

DF=0：可以分片

MF：还有分片 （More Fragment)

MF=1：后面还有分片

MF=0：这时候最后一个分片

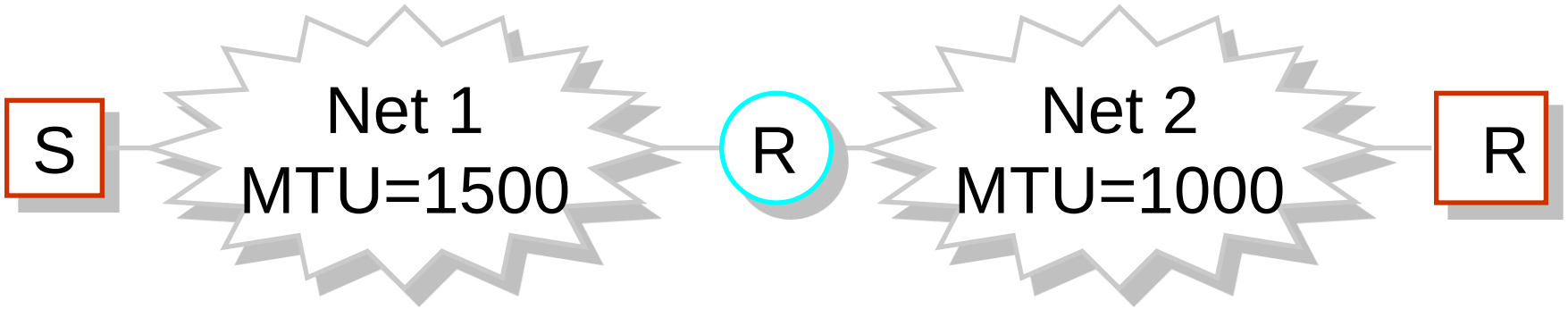


片偏移(13 bit)指出：分片后某片在原分组中的相对位置。

片偏移以 8 个字节为偏移单位。

最大传输单元(MTU)

- 每个子网都有一个 最大帧长度
  Ethernet: 1518 bytes　　FDDI: 4500 bytes
  Token Ring: 2 to 4 kB　　PPP 298
  ATM AAL5: 9180
- 传输单元 = IP 数据报 (数据 + 头部)
- 每个子网都有一个最大IP数据报长度（头部＋数据）=MTU

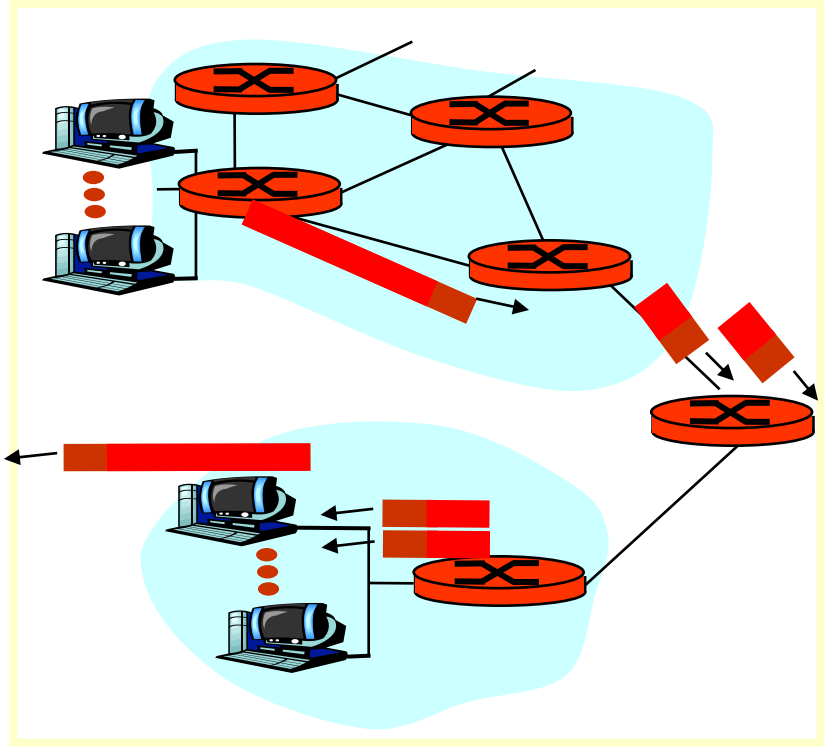


分片举例

MTU = 1500B

MTU = 280B

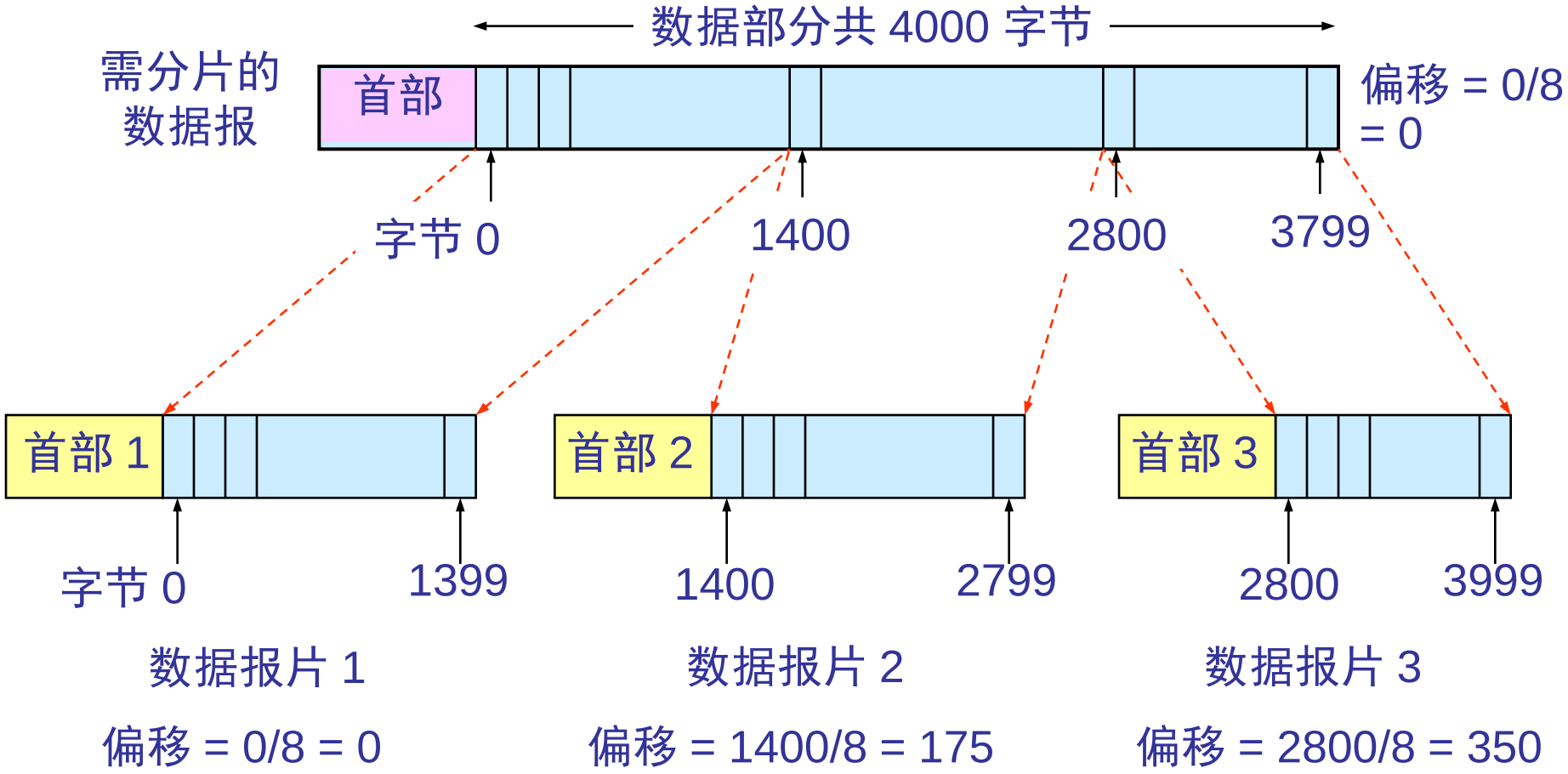
IHL = 5, ID = 111, More = 0
Offset = 0W, Len = 472B



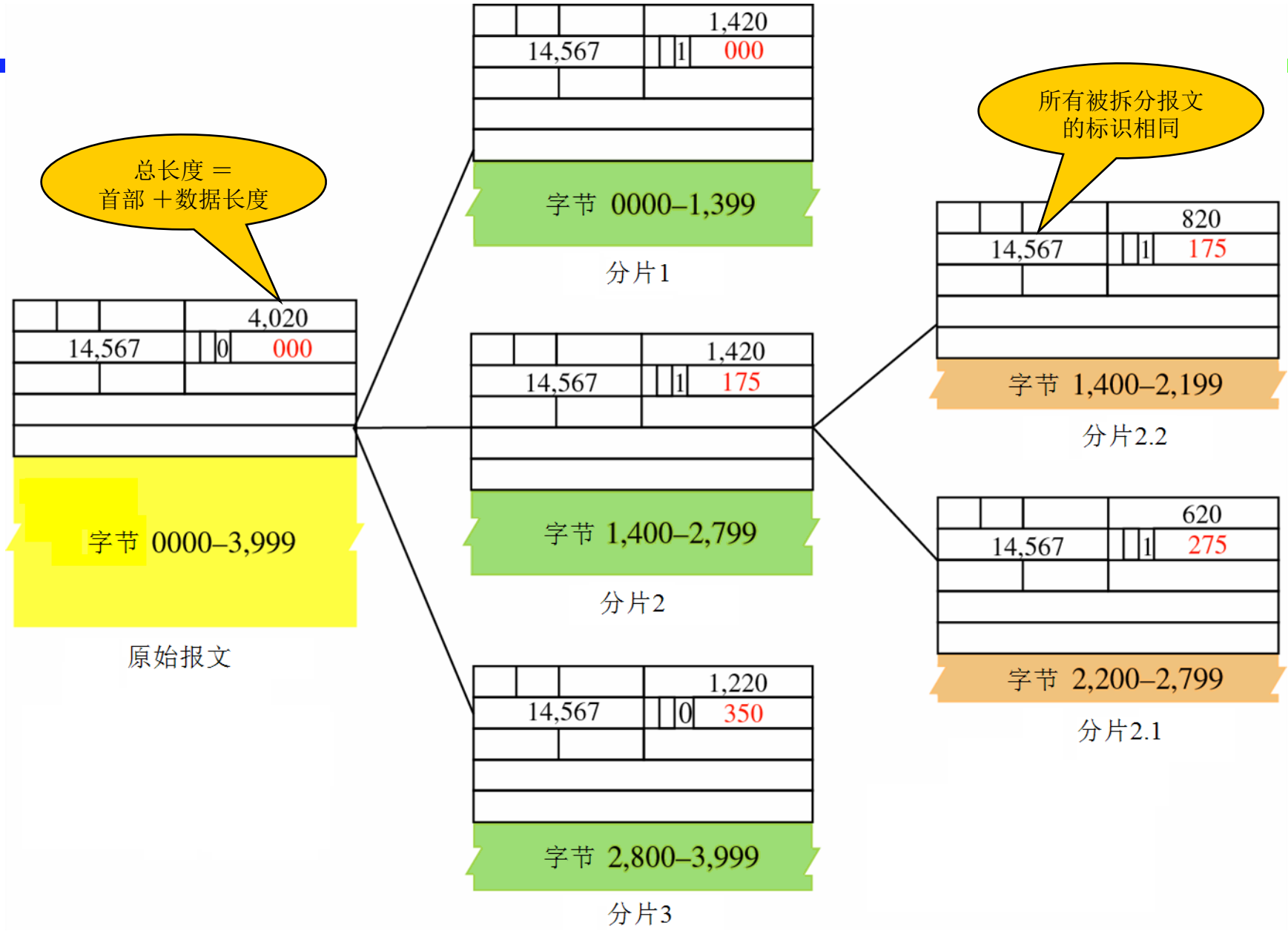
IHL=5, ID = 111, More = 1
Offset = 0W, Len = 276B

IHL=5, ID = 111, More = 0
Offset = 32W, Len = 216B

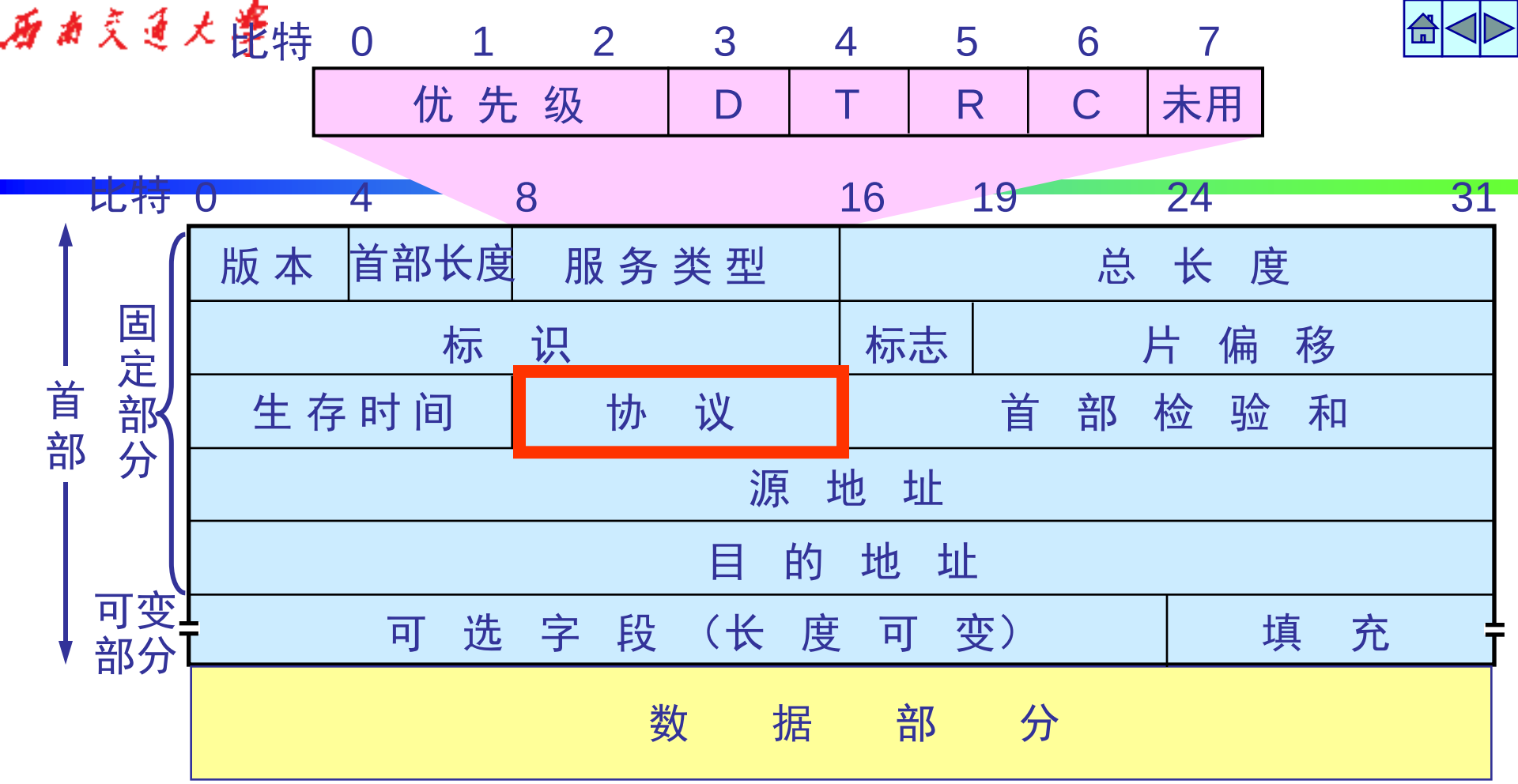
IP 数据报分片的举例


数据部分共 4000 字节
需分片的
数据报
首部
偏移 = 0/8
= 0
字节 0
1400
2800
3799
首部 1
首部 2
首部 3
字节 0
1399
1400
2799
2800
3999
数据报片 1
偏移 = 0/8 = 0
数据报片 2
偏移 = 1400/8 = 175
数据报片 3
偏移 = 2800/8 = 350

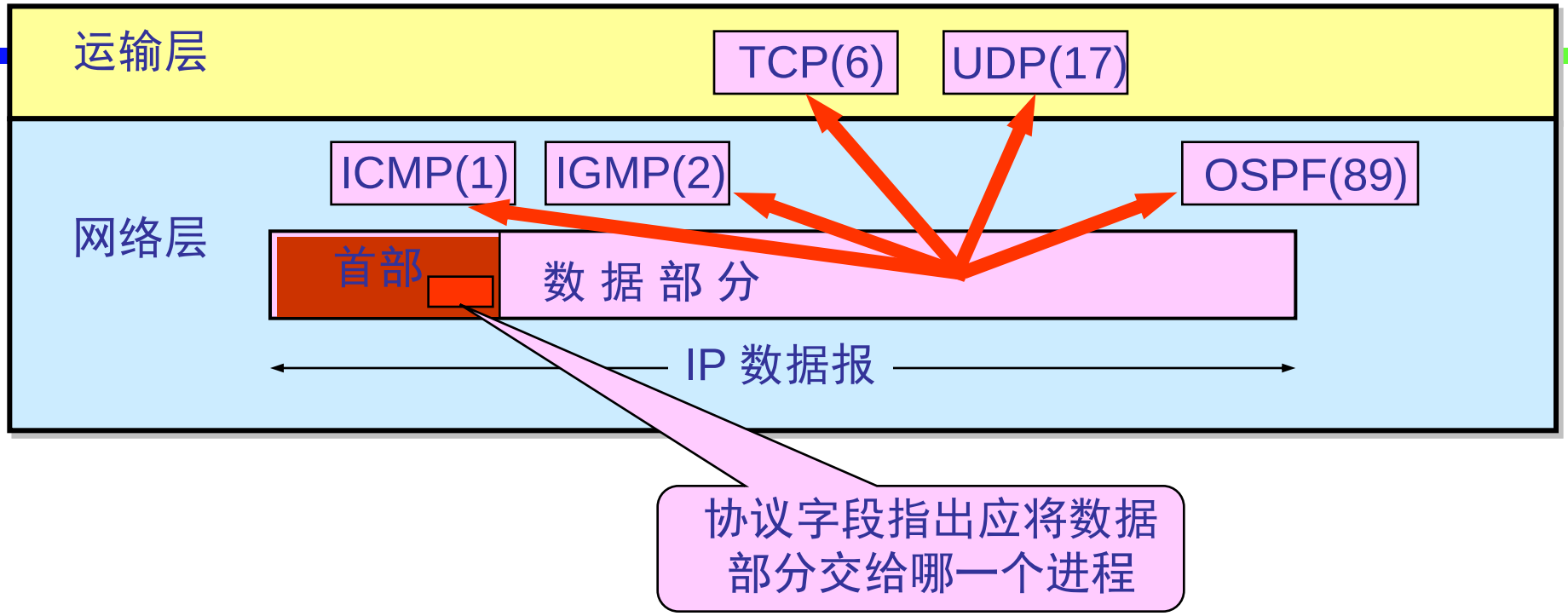
数据分片举例


总长度 =
首部 ＋数据长度
4,020
14,567
0
000
字节 0000–3,999
原始报文
1,420
14,567
1
000
字节 0000–1,399
分片1
1,420
14,567
1
175
字节 1,400–2,799
分片2
1,220
14,567
0
350
字节 2,800–3,999
分片3
所有被拆分报文
的标识相同
820
14,567
1
175
字节 1,400–2,199
分片2.2
620
14,567
1
275
字节 2,200–2,799
分片2.1

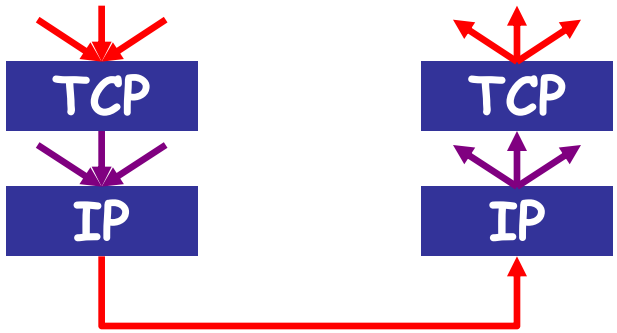
生存时间(8 bit)记为 TTL (Time To Live)
数据报在网络中的寿命，其单位为秒或hop。



协议(8 bit)字段指出此数据报携带的数据使用何种协议以便目的主机的 IP 层将数据部分上交给哪个处理进程

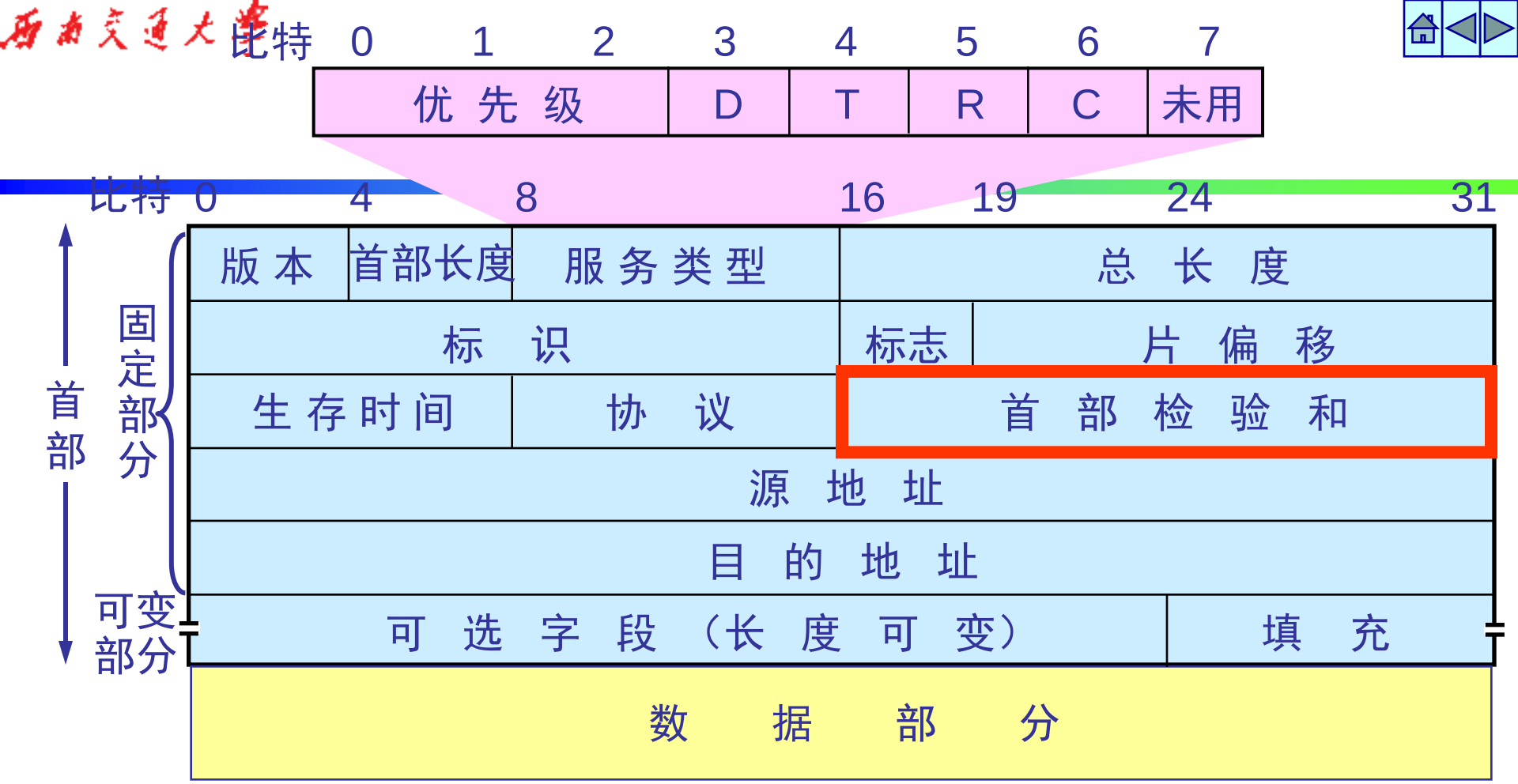


V/HL	TOS	Length
ID		Flags/Offset
TTL	Prot.	H. Checksum
Source IP address		
Destination IP address		
Options..		



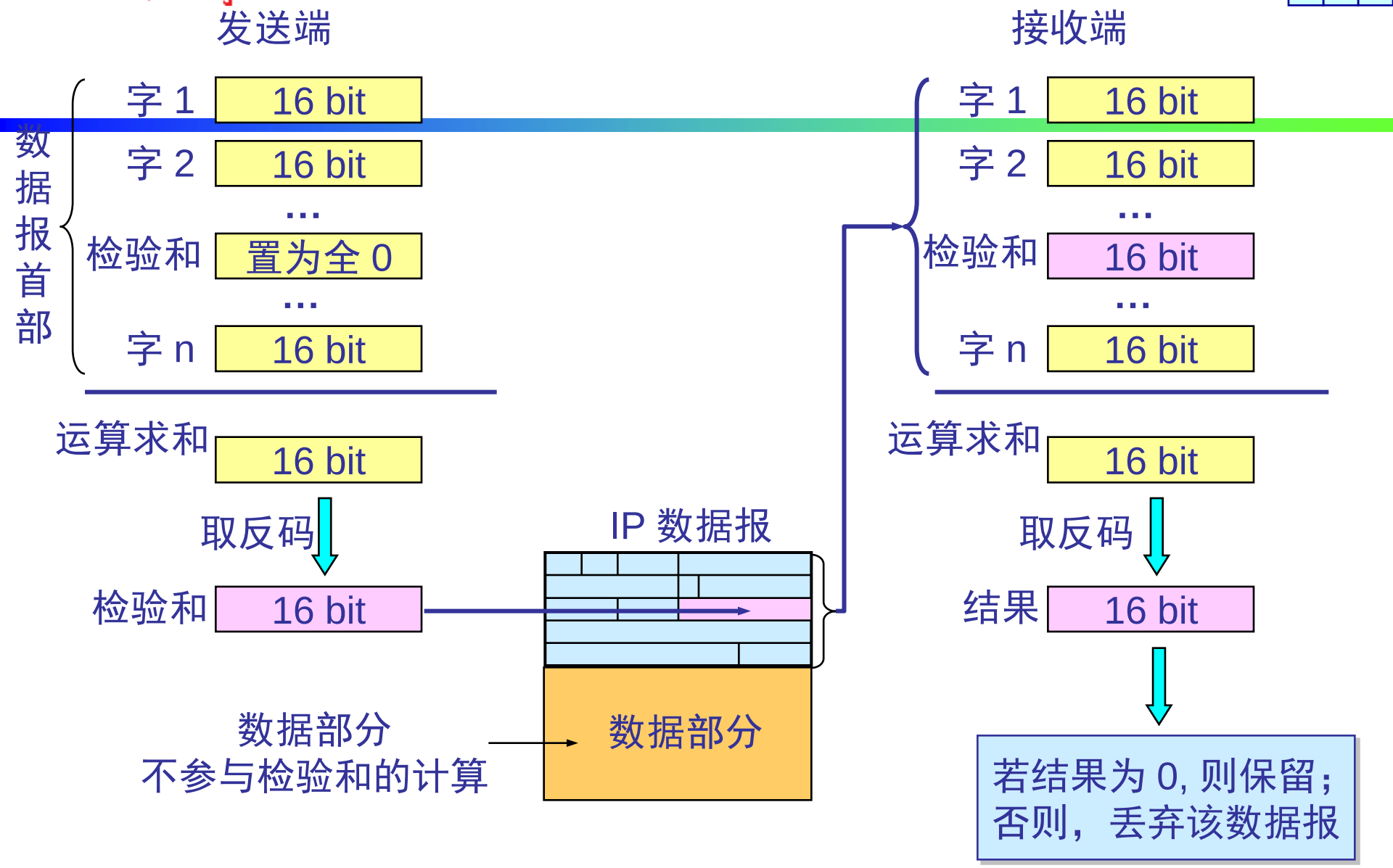
常用协议定义

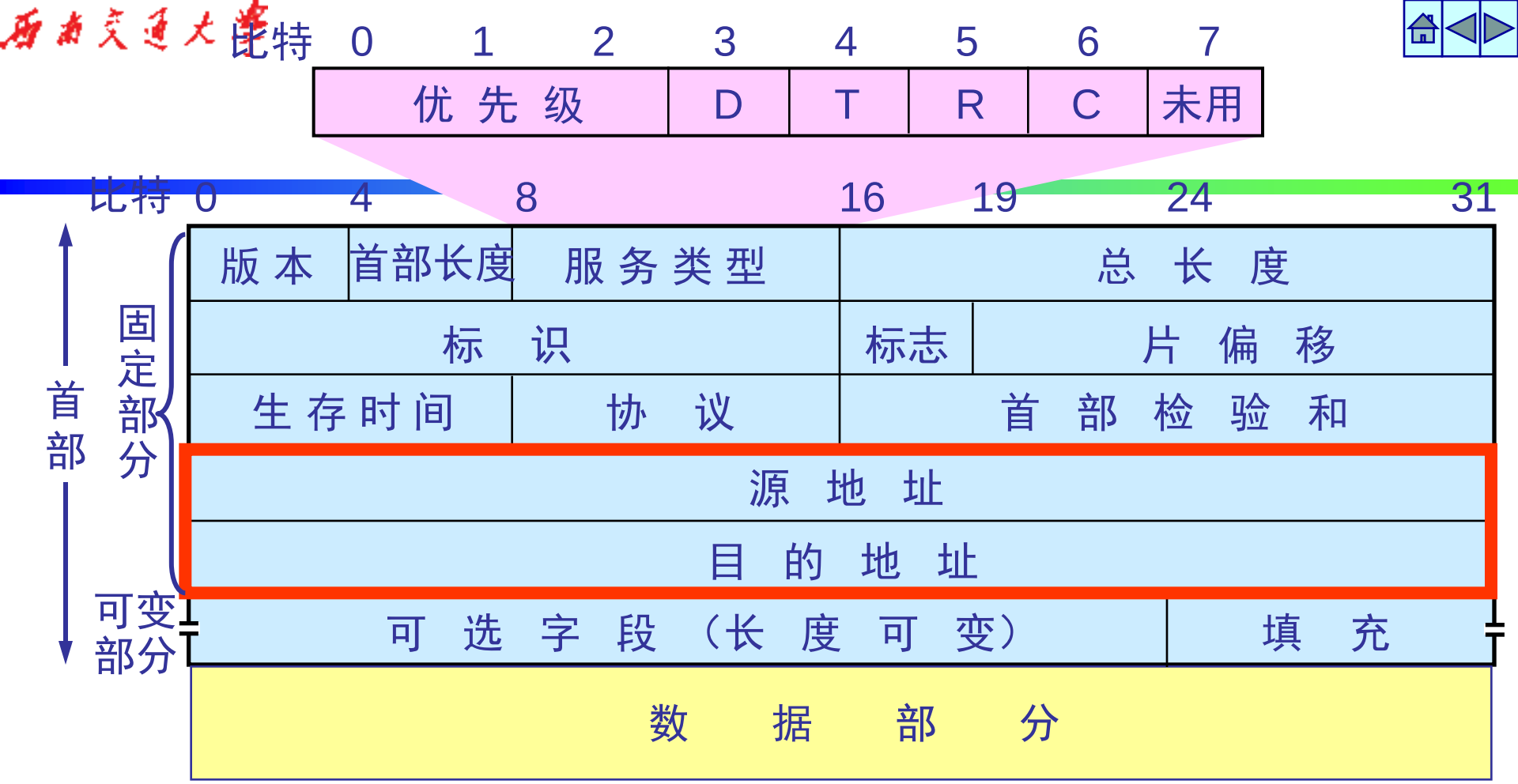
十进制编号	关键字	协议名称
0		保留
1	ICMP	因特网控制报文
2	IGMP	因特网组管理
3	GGP	网关-网关协议
4	IP	IP里的IP（IP in IP）
5	ST	数据流
6	TCP	传输控制
8	EGP	外部网关协议
17	UDP	用户数据报协议
41	IP v6	
46	**RSVP**	IDPR控制报文传输协议
80	ISO-IP	ISO因特网协议（CLNP）
88	IGRP	
89	OSPF	开放最短路径优先
255		保留



首部检验和(16 bit)字段只检验数据报的首部
不包括数据部分。

这里不采用 CRC 检验码而采用简单的计算方法。


发送端
接收端
数据报首部
字 1
16 bit
字 2
16 bit
...
检验和
置为全 0
...
字 n
16 bit
运算求和
16 bit
取反码
检验和
16 bit
IP 数据报
数据部分
不参与检验和的计算
数据部分
字 1
16 bit
字 2
16 bit
...
检验和
16 bit
...
字 n
16 bit
运算求和
16 bit
取反码
结果
16 bit
若结果为 0, 则保留；
否则， 丢弃该数据报



源地址和目的地址都各占 4 字节

选项

- 安全性限制（用于军事领域，详细内容参见RFC 1108[Kent 1991]）
- 记录路由：每个处理数据报的路由器将它的IP地址加到头部。
- 时间标记：每个处理数据报的路由器将它的IP地址和时间加到头部。
- 松的源路由选择：给出一些不能漏掉的路由器地址列表(关键节点,其间可有其他路由节点)。
- 严格的源路由选择:预先设定数据报必须经过的传送路由——只能访问所列的路由节点

这些选项很少被使用，并非所有的主机和路由器都支持这些选项。

选项

复制标记	选项类	选项号
1位	2位	5位

复制标记:指明是否要在分片时将选项域复制到所有段中:　0 = 不复制；1 = 复制

选项

选项类	选项号	长度	DESCRIPTION
0	0	-	选项表结束。只占一个字节，没有长度字节
0	1	-	空操作。只占一个字节，没有长度字节
0	2	11	安全及限制处理，用于传送安全、Compartmentation、处理限制和用户组（TCC）标志
0	3	变长	松源地址路由（loose source routing），使用源地址提供的信息进行路由
0	9	变长	严格源地址路由(strict source souting)，使用源地址提供的信息进行路由
0	7	变长	记录路由，用于跟踪数据报采用的路由
0	8	4	数据流ID，用于传送流标记
2	4	变长	Internet时间标记

记录路由选项

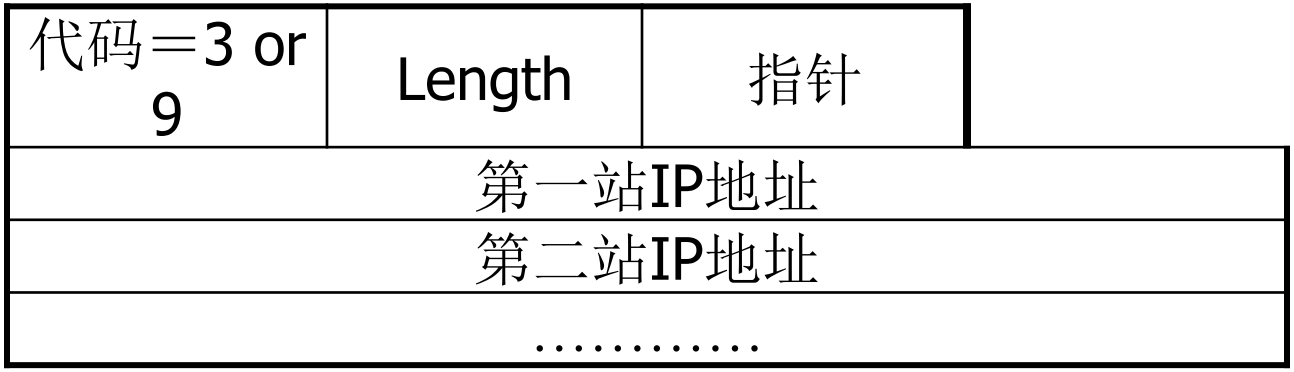
- IP报文每经过一个机器处理时，把自己的IP地址添加到该Option预留的空间中。
- 报文到达目的时，表中记录了该IP所走过的路由

代码＝7	Length	指针
第一站IP地址		
第二站IP地址		
……		

目的：用来监视和控制互连网中的路由器是如何转发数据报的

源站选路选项

- 在Option中给出一系列IP地址，指定报文传输的路径
- 严格源路由
  - 严格逐条按给定的IP地址转发（中间不允许经过其它IP地址）
- 松散源路由
  - 指定必须经由的关键IP地址（中间可经过其它IP地址）



目的：使网络管理者了解沿网络中某一条通路的通信状况是否正常

严格的源路由选择(SRS)

- 发送端指明IP数据包必须经过的确切地址。如果没有经过这一确切路径，数据包会被丢弃，并返回一个ICMP差错报文。
- 源站路由使用IP首部一个39个字节的源路由选项地址来工作，其中3个字节是附加信息，那么剩下的36个字节是地址信息。每一个地址是4个字节最多有9个地址的空间，因为最后一个地址必须是目的地址，所以它只留下8个地址的空间。随着互联网的发展，会出现IP地址的数目大于8的情况。在这些情况下，就只能使用宽松的源站选路，因为如果不能找到确切的路径，那么严格的源路由选路就会丢弃那个数据包。

时间戳选项

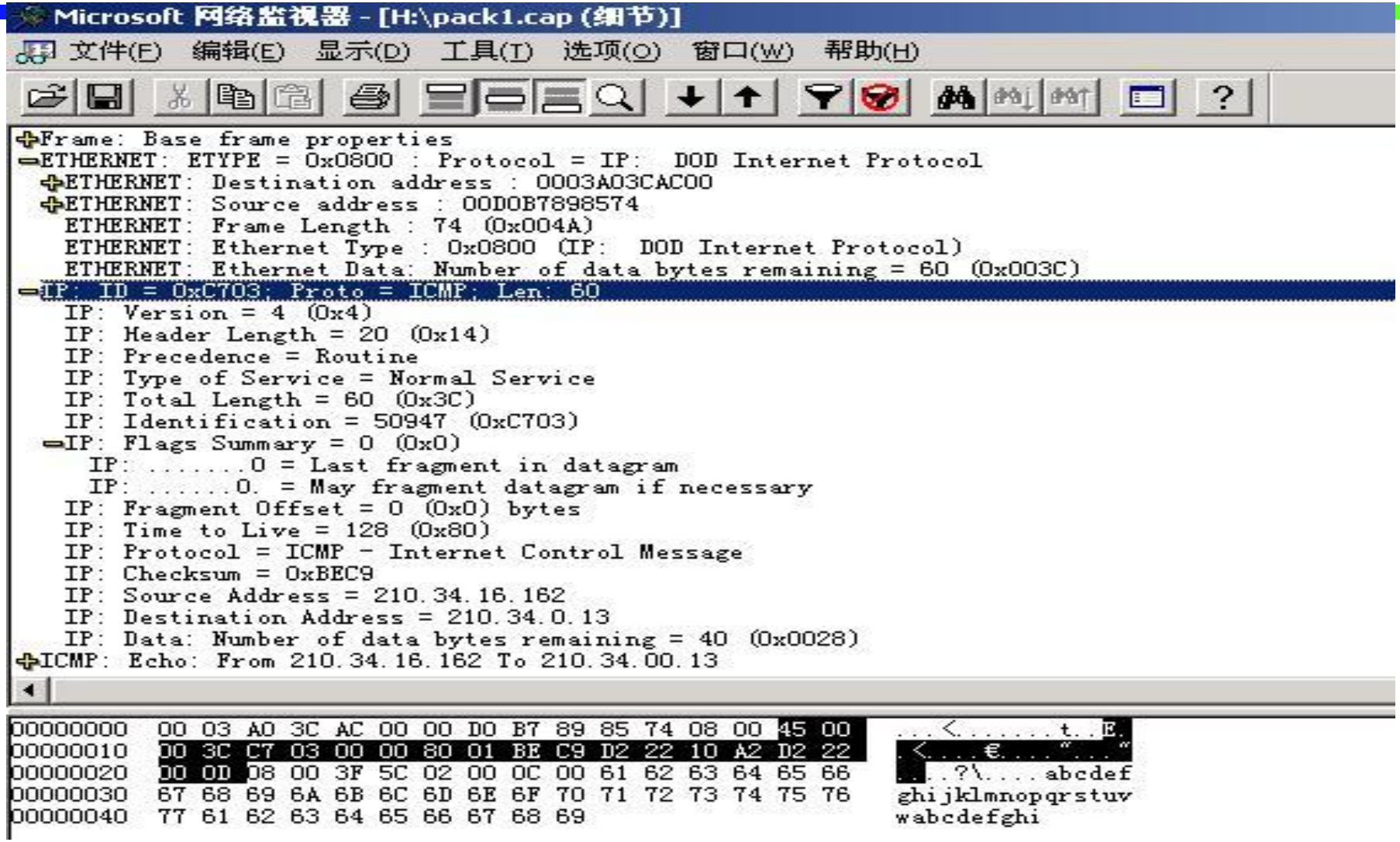
- IP报文每经过一个机器处理时，把自己的IP地址和到达时间添加该Option预留的空间中。
- 报文到达目的时，表中记录了该IP所走过的路径和时间。

代码＝7	Length	指针	溢出	标志
第1站IP地址				
第1个时间戳				
.....				

标志：
0: 仅记录时间，忽略IP地址
1: 记录IP和时间
3: 发方指定了IP，仅记录时间

目的：用来统计数据报经路由器产生的时延和时延的变化

IP包实例


Microsoft 网络监视器 - [H:\pack1.cap (细节)]
文件(F) 编辑(E) 显示(D) 工具(T) 选项(O) 窗口(W) 帮助(H)
Frame: Base frame properties
ETHERNET: ETYPE = 0x0800 : Protocol = IP: DOD Internet Protocol
ETHERNET: Destination address : 0003A03CAC00
ETHERNET: Source address : 00D0B7898574
ETHERNET: Frame Length : 74 (0x004A)
ETHERNET: Ethernet Type : 0x0800 (IP: DOD Internet Protocol)
ETHERNET: Ethernet Data: Number of data bytes remaining = 60 (0x003C)
IP: ID = 0xC703; Proto = ICMP; Len: 60
IP: Version = 4 (0x4)
IP: Header Length = 20 (0x14)
IP: Precedence = Routine
IP: Type of Service = Normal Service
IP: Total Length = 60 (0x3C)
IP: Identification = 50947 (0xC703)
IP: Flags Summary = 0 (0x0)
IP: .....0 = Last fragment in datagram
IP: .....0. = May fragment datagram if necessary
IP: Fragment Offset = 0 (0x0) bytes
IP: Time to Live = 128 (0x80)
IP: Protocol = ICMP - Internet Control Message
IP: Checksum = 0xBEC9
IP: Source Address = 210.34.16.162
IP: Destination Address = 210.34.0.13
IP: Data: Number of data bytes remaining = 40 (0x0028)
ICMP: Echo: From 210.34.16.162 To 210.34.00.13
00000000 00 03 A0 3C AC 00 00 D0 B7 89 85 74 08 00 45 00
00000010 00 3C C7 03 00 00 80 01 BE C9 D2 22 10 A2 D2 22
00000020 00 0D 08 00 3F 5C 02 00 0C 00 61 62 63 64 65 66
00000030 67 68 69 6A 6B 6C 6D 6E 6F 70 71 72 73 74 75 76
00000040 77 61 62 63 64 65 66 67 68 69
...<.....t..E.
.<...€....".....
....?\....abcdef
ghijklmnopqrstuv
wabcdefghi

ICMP

ICMP:网际控制报文协议

ICMP报文分为三大类（三大功能）：

a) 差错报告（用于网关→主机信息传输）

- ✓ 信宿（网络、主机、协议、端口）不可达报告
- ✓ 超时报告(TTL 值为0)
- ✓ 参数差错报告
- ✓ IP 校验和错误
- ✓ 重组失败

b) 控制报文（网关→主机）

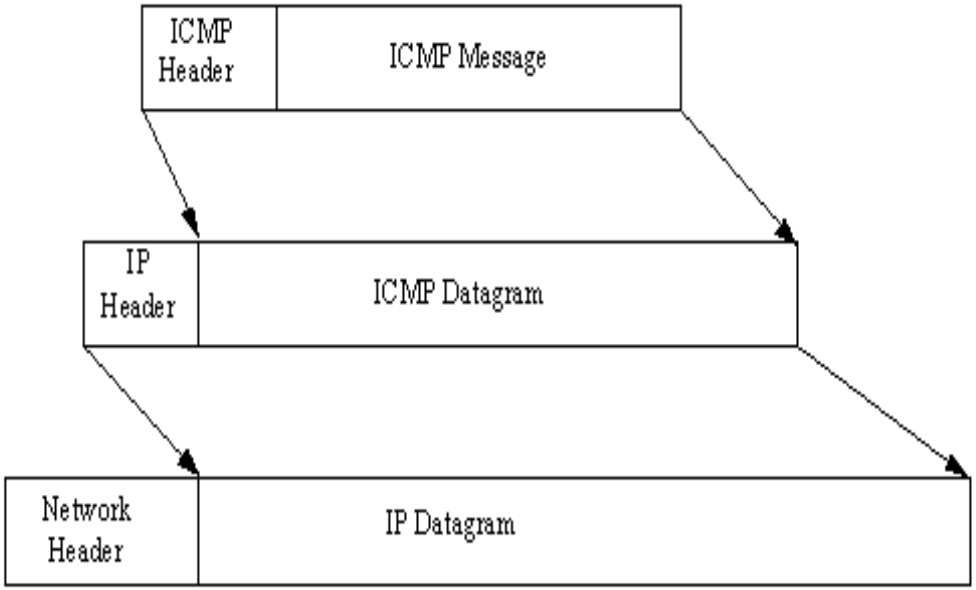
- ✓ 源抑制报文
- ✓ 重定向报文

c) 请求/应答报文（用于主机→主机信息传输）

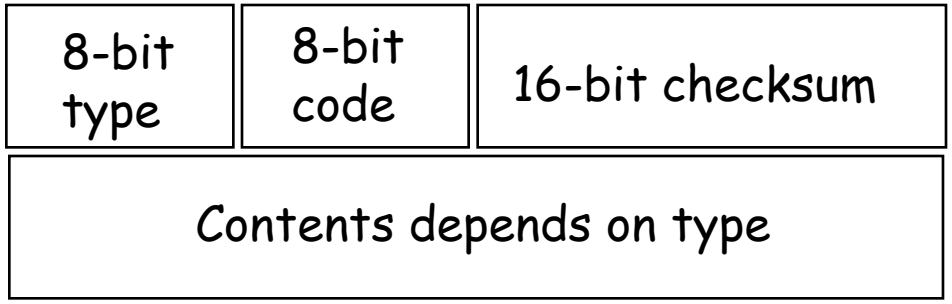
- ✓ ECHO请求/应答
- ✓ 时间戳请求/应答
- ✓ 地址掩码请求/应答

ICMP报文

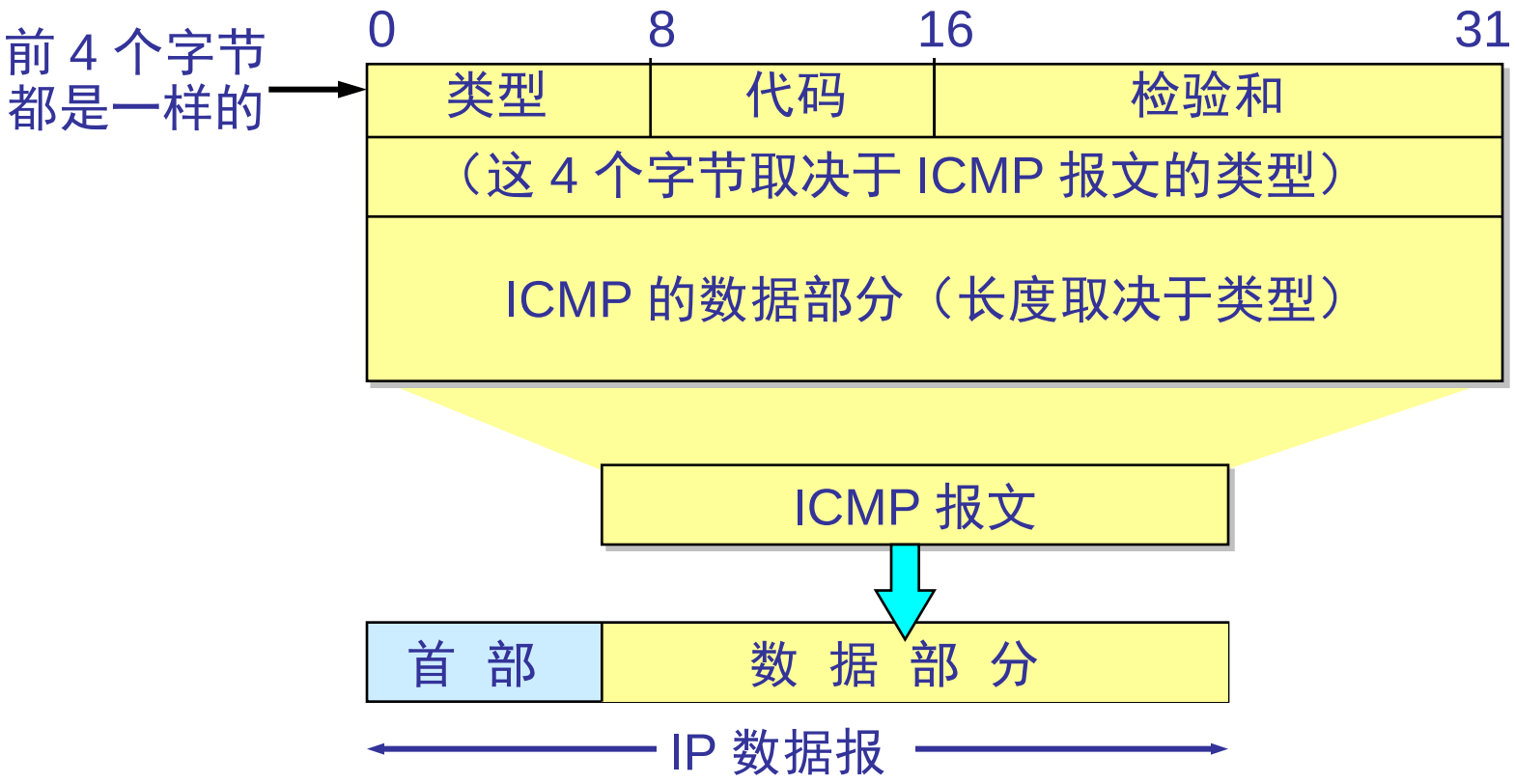
- 用于主机、路由器、网关传送网络层信息（如差错通知或询问网络状况）
- network-layer :"above" IP:
  - 由IP数据报携带ICMP报文（协议域=1）
- ICMP 报文: 类型、代码、出错的 IP分组的头8个字节



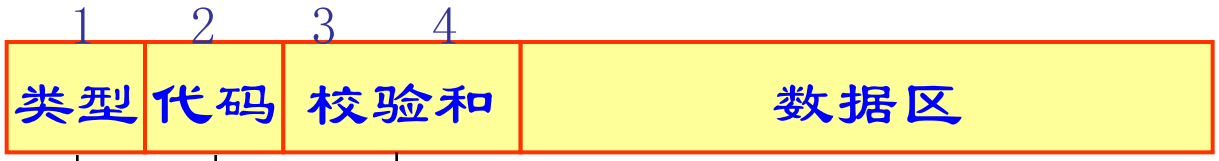
ICMP Message Format



ICMP 报文的格式


前 4 个字节
都是一样的
0
8
16
31
类型
代码
检验和
（这 4 个字节取决于 ICMP 报文的类型）
ICMP 的数据部分（长度取决于类型）
ICMP 报文
首 部
数 据 部 分
IP 数据报

ICMP 的报文



CHECKSUM：整个 ICMP 报文的校验和

算法与 IP 分组头的校验和算法相同

CODE：提供报文类型的详细信息

TYPE：指出 ICMP 报文的类型

- 3：目的不可达
- 11：分组超时
- 12：分组参数错
- 4：源抑制
- 5：路由重定向
- 8：回应请求
- 0：回应应答
- 13：时间戳请求
- 14：时间戳应答
- 17：地址掩码请求
- 18：地址掩码应答

ICMP报文的类型

类型码	ICMP报文类型
0	ECHO应答
3	信宿不可达
4	源抑制（Source Quench）
5	重定向
8	ECHO请求
9	路由广播
10	路由请求
11	数据报超时
12	数据报参数错
13	时戳请求
14	时戳应答
15	信息请求消息
16	信息响应消息
17	地址掩码请求
18	地址掩码应答

各类ICMP报文头部格式

Type	Code	Checksum
Unused		
Original IP header + 64 bits		

Destination unreachable, Source Quench, Time Exceeded

Type	Code	Checksum
Identifier		Sequence No.
Originating Timestamp		

Timestamp Request

Type	Code	Checksum
Ptr	Unused	
Original IP header + 64 bits		

Parameter Problem

Type	Code	Checksum
Identifier		Sequence No.
Originating Timestamp		
Receiving Timestamp		
Transmitting Timestamp		

Timestamp Reply

Type	Code	Checksum
Gatewat IP Address		
Original IP header + 64 bits		

Redirect

Type	Code	Checksum
Identifier		Sequence No.

Information Request and Reply, Address Mask Request

Type	Code	Checksum
Identifier		Sequence No.
Original IP header + 64 bits		

Echo Request and Echo Reply

Type	Code	Checksum
Identifier		Sequence No.
Address Mask		

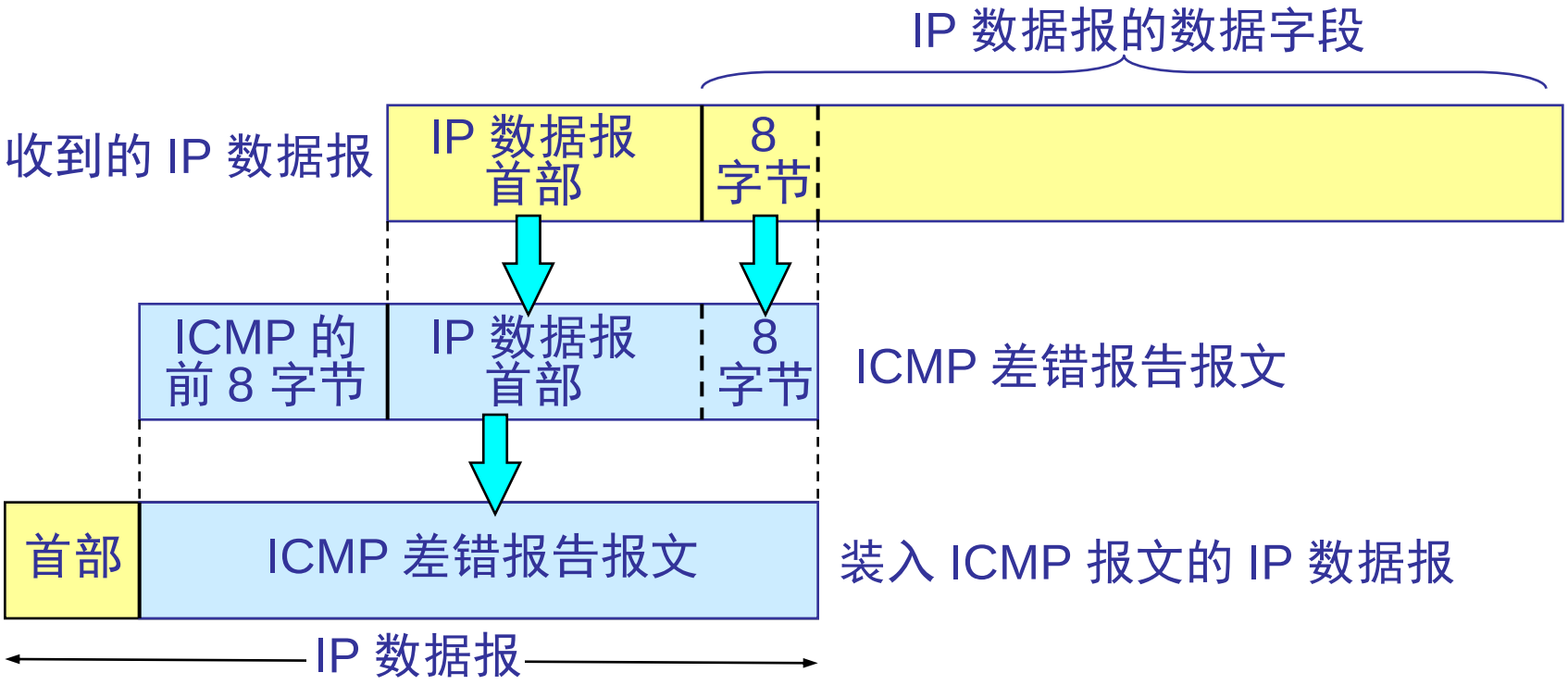
Address Mask Reply

ICMP 差错报告

- 差错报告是单方向的：路由器 -> 信源主机
- 目的地址不可达 TYPE = 3
- 分组超时 TYPE = 11
- 分组参数错

- ICMP 差
  - ICMP
    输的可
  - ICMP 提供差错报告的功能，但并没有严格规定对于某种差错应该采取何种差错处理措施
  - 路由器并不能发现所有的传输错误

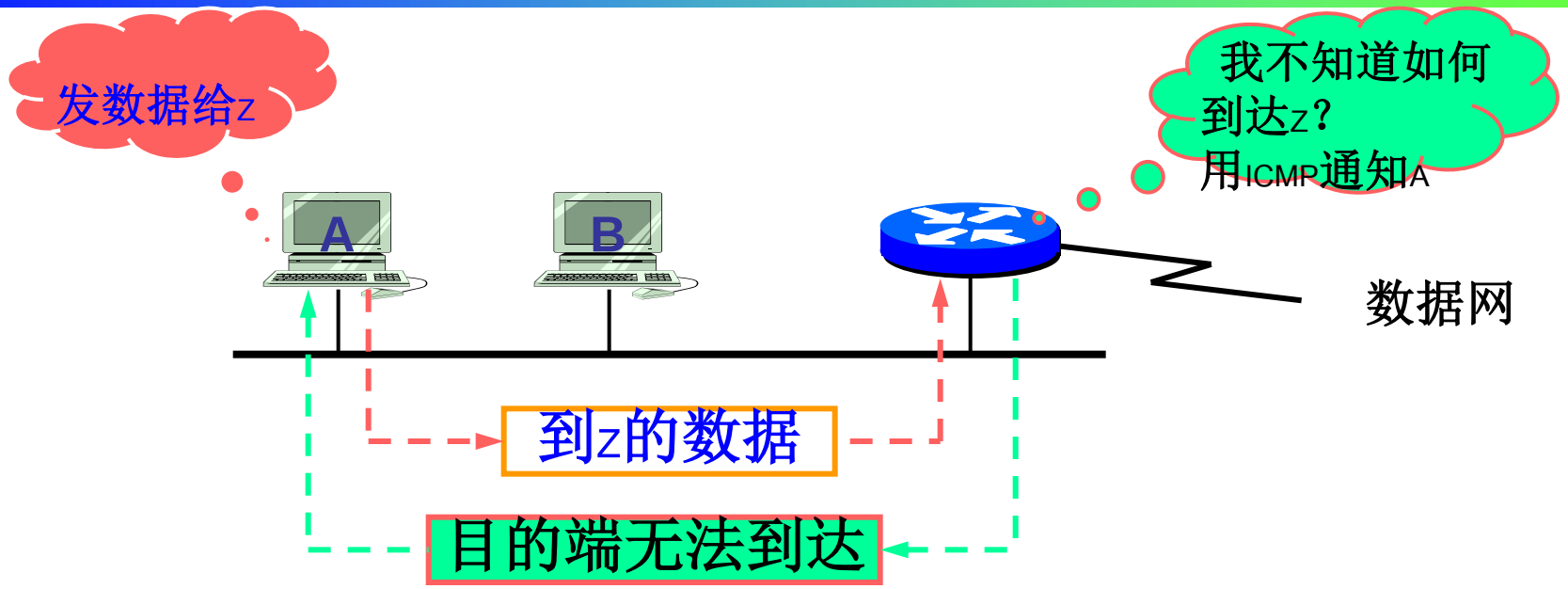
ICMP 差错报告报文的数据字段的内容



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。

ICMP 目的地不可达



路由器用ICMP通知目的地不可达的示意图

- RFC 792: 如果在目的主机中，IP模块因为指定的网络、主机、主机上的协议模块或者进程端口没有激活而不能传送数据报，那么目的主机可以发送一个目的地不可达消息到源端主机。

错误报文:端口不可达

- 有16种不同的“主机不可达”类型(Code = 3)的ICMP差错报文。

码值:	含义:
0	网络不可达
1	主机不可达
2	协议不可达
3	**端口不可达**
4	需要分片并且DF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

码值:	含义:
9	出于管理目的禁止了与目的网络的通信
10	出于管理目的禁止了与目的主机的通信
11	对所请求的服务类型，网络不可达
12	对所请求的服务类型，主机不可达
13	通过过滤来禁止通信
14	违反主机优先
15	有效地中止优先权

拥塞控制与 ICMP 源抑制报文

- 源抑制（Source Quench）

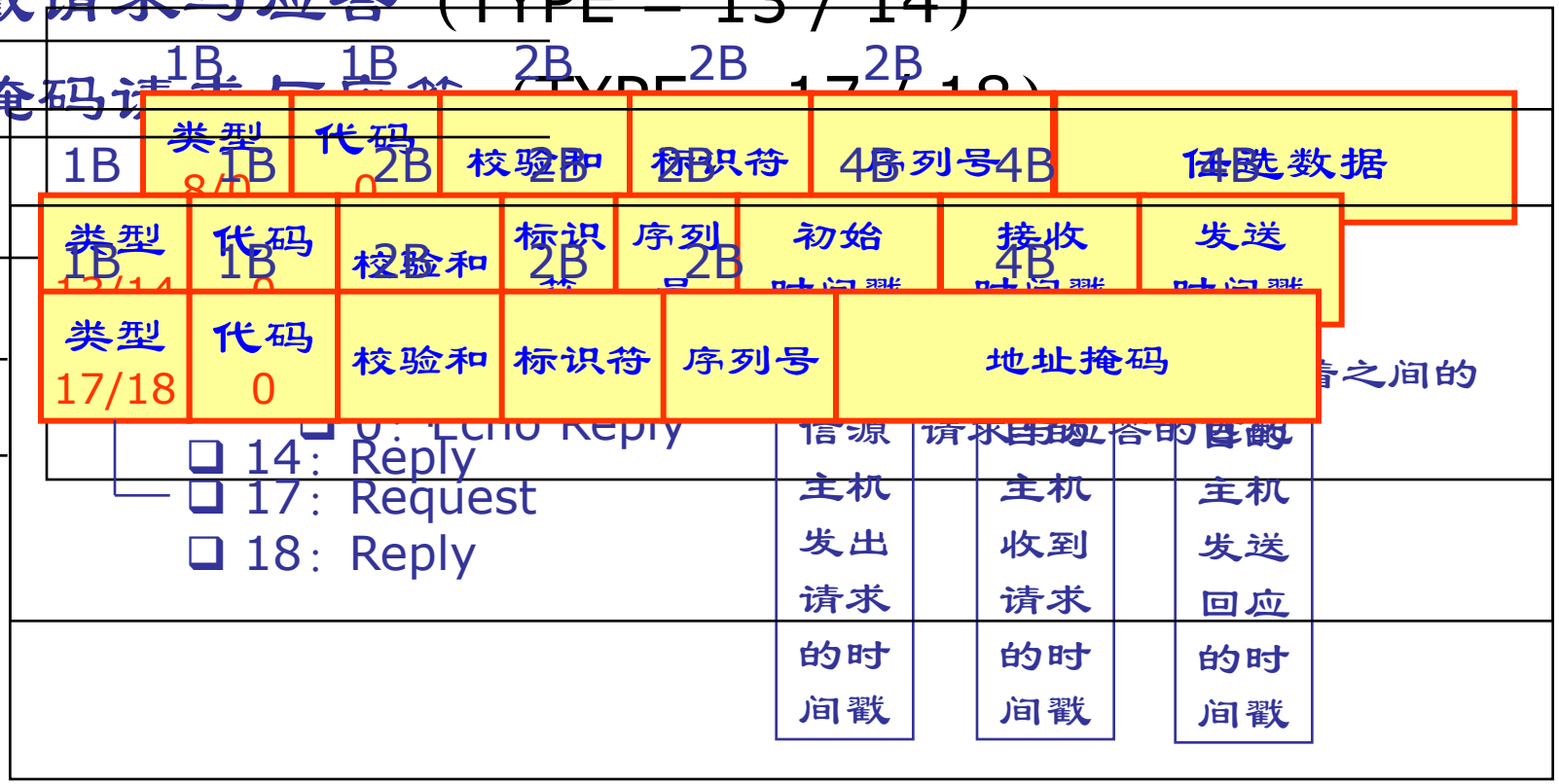
  通过限制信源主机发送 IP 分组的速率来降低拥塞的方法

- 源抑制的三个阶段：
  - 路由器发现拥塞，并向源端发送 ICMP 源抑制报文
  - 信源主机逐步降低发往目的主机的分组发送速率
  - 拥塞解除后，信源主机逐步恢复原有的分组发送速率
- 源抑制报文的格式

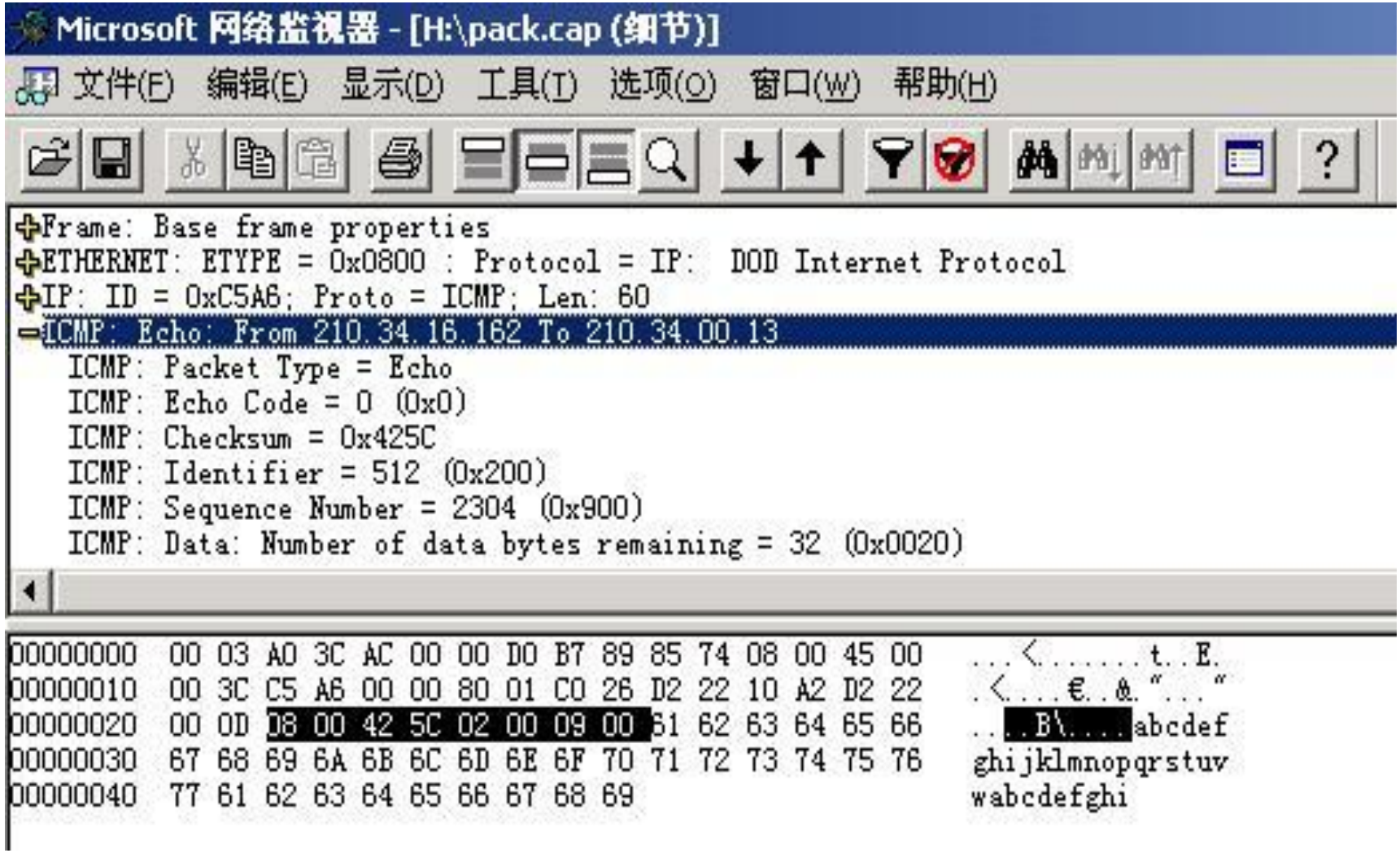
1B	1B	2B	4B	
类型 4	代码 0	校验和	未用 全 0	抑制分组的 IP 头 + 前 64 位的数据

ICMP 的请求/应答报文

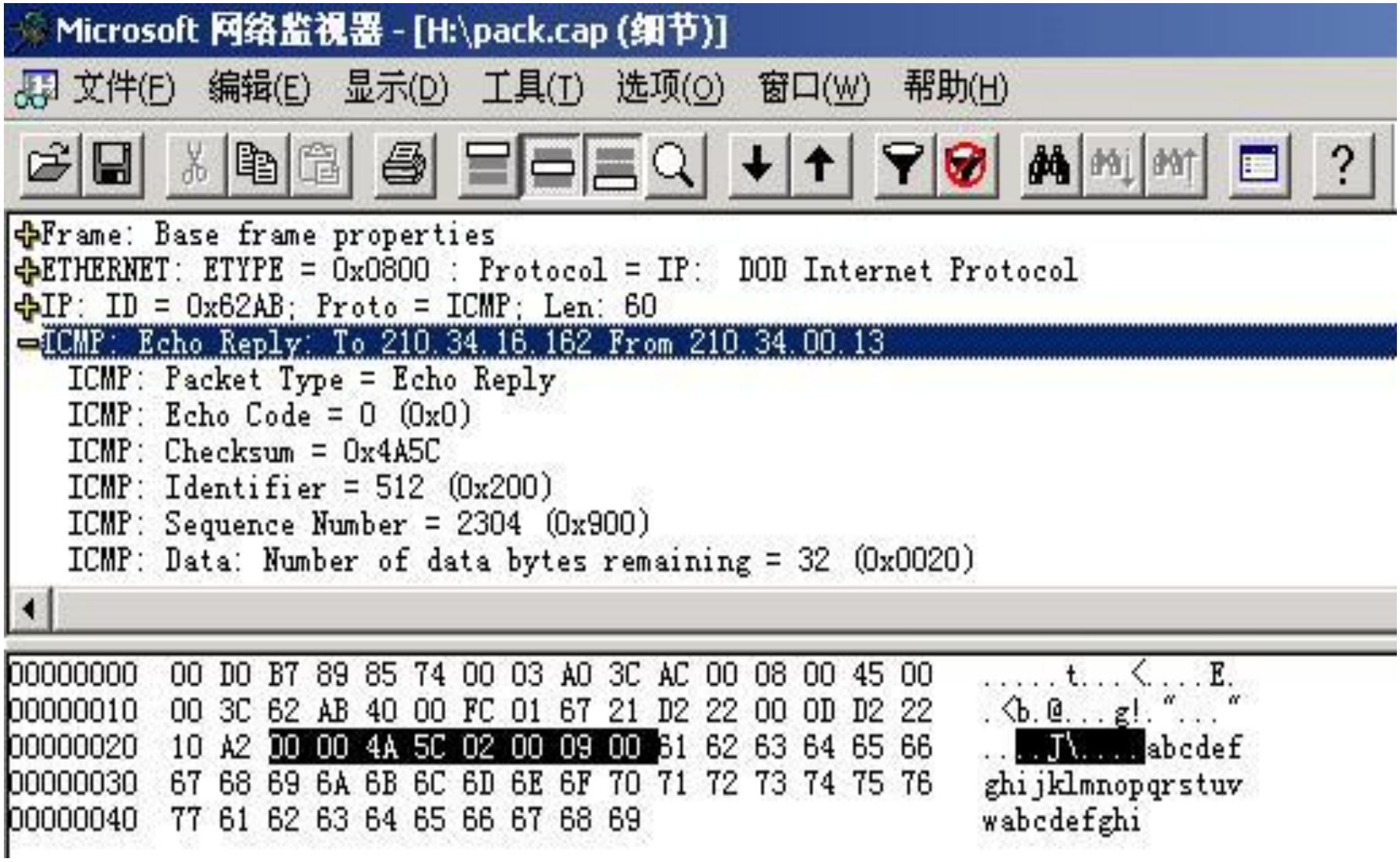
- 回应（Echo）请求/应答 [ping]（TYPE = 8 / 0）
- 时间戳请求与应答（TYPE = 13 / 14）
- 地址掩码请求与应答（TYPE = 17 / 18）



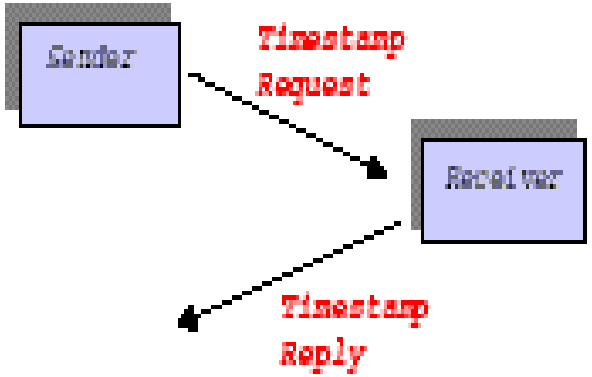
回送请求包实例


Microsoft 网络监视器 - [H:\pack.cap (细节)]
文件(F) 编辑(E) 显示(D) 工具(T) 选项(O) 窗口(W) 帮助(H)
Frame: Base frame properties
ETHERNET: ETYPE = 0x0800 : Protocol = IP: DOD Internet Protocol
IP: ID = 0xC5A6; Proto = ICMP; Len: 60
ICMP: Echo: From 210.34.16.162 To 210.34.00.13
ICMP: Packet Type = Echo
ICMP: Echo Code = 0 (0x0)
ICMP: Checksum = 0x425C
ICMP: Identifier = 512 (0x200)
ICMP: Sequence Number = 2304 (0x900)
ICMP: Data: Number of data bytes remaining = 32 (0x0020)
00000000 00 03 A0 3C AC 00 00 D0 B7 89 85 74 08 00 45 00
00000010 00 3C C5 A6 00 00 80 01 C0 26 D2 22 10 A2 D2 22
00000020 00 0D 08 00 42 5C 02 00 09 00 61 62 63 64 65 66
00000030 67 68 69 6A 6B 6C 6D 6E 6F 70 71 72 73 74 75 76
00000040 77 61 62 63 64 65 66 67 68 69
abcdef
ghijklmnopqrstuv
wabcdefghi

回送应答包实例


Microsoft 网络监视器 - [H:\pack.cap (细节)]
文件(F) 编辑(E) 显示(D) 工具(T) 选项(O) 窗口(W) 帮助(H)
Frame: Base frame properties
ETHERNET: ETYPE = 0x0800 : Protocol = IP: DOD Internet Protocol
IP: ID = 0x62AB; Proto = ICMP; Len: 60
ICMP: Echo Reply: To 210.34.16.162 From 210.34.00.13
ICMP: Packet Type = Echo Reply
ICMP: Echo Code = 0 (0x0)
ICMP: Checksum = 0x4A5C
ICMP: Identifier = 512 (0x200)
ICMP: Sequence Number = 2304 (0x900)
ICMP: Data: Number of data bytes remaining = 32 (0x0020)
00000000 00 D0 B7 89 85 74 00 03 A0 3C AC 00 08 00 45 00
00000010 00 3C 62 AB 40 00 FC 01 67 21 D2 22 00 0D D2 22
00000020 10 A2 00 00 4A 5C 02 00 09 00 61 62 63 64 65 66
00000030 67 68 69 6A 6B 6C 6D 6E 6F 70 71 72 73 74 75 76
00000040 77 61 62 63 64 65 66 67 68 69
abcdefghijklmnopqrstuvwabcdefghi

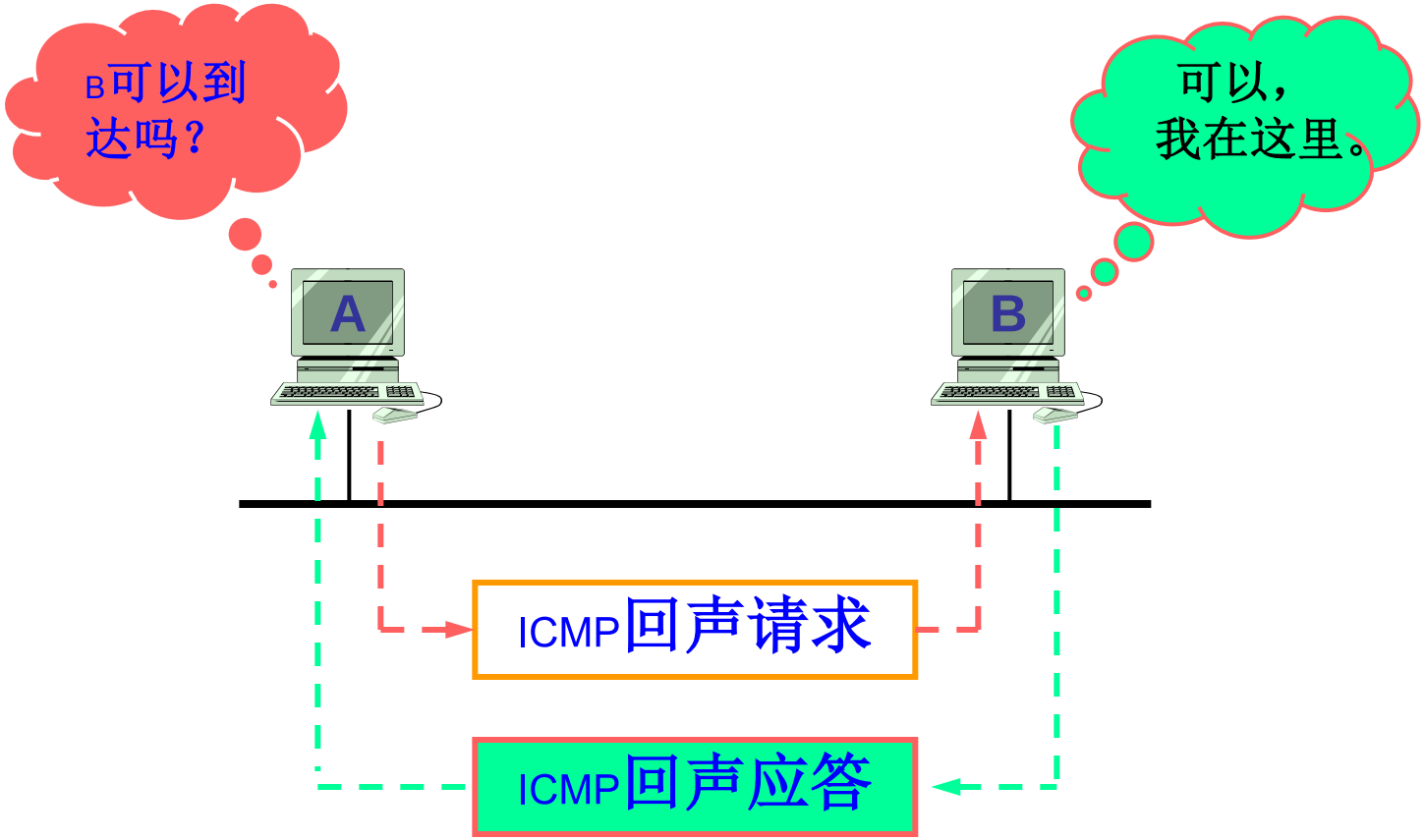
一个询问的例子:ICMP 时戳报文



- 一个系统（主机或者路由器）向另一个系统询问当前时间。
- 时间的单位是毫秒，从午夜开始的UTC（Universal Coordinated Time)
- 发送方发送一个 request, 接收方回应一个 reply

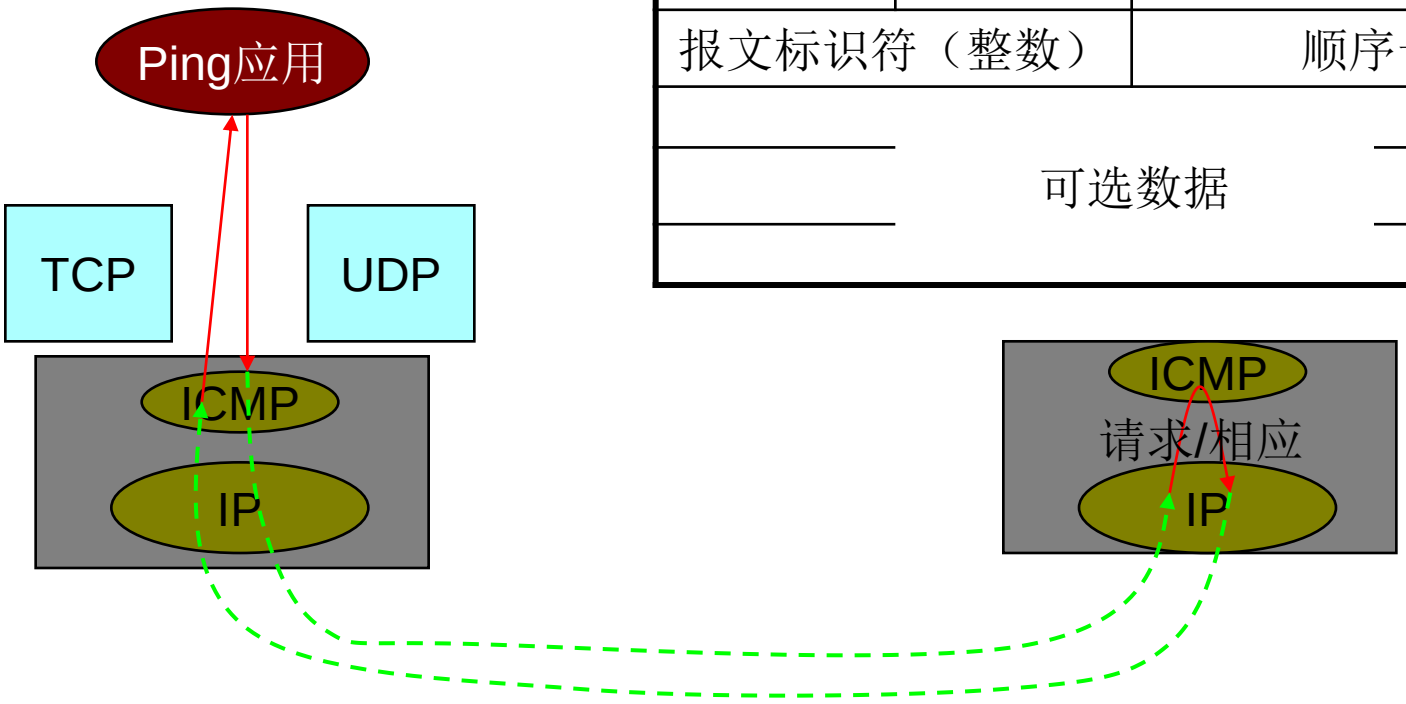
Type (= 17 or 18)	Code (=0)	Checksum
identifier		sequence number
32-bit sender timestamp		
32-bit receive timestamp		
32-bit transmit timestamp		

ICMP回声请求/应答



用PING命令产生的回声及其应答示意图

PING的实现原理



0 or 8	0	校验和
报文标识符（整数）		顺序号
可选数据		

PING

- PING (=Packet InterNet Groper) 是一个应用ICMP回送请求和回送应答消息的程序。
- PING 在使用的时候有很多选项，比如:

- f Flood ping. 尽快发送或是1/100秒的速度输出分组
- R 记录路由（ Record route ）. 包含ECHO_REQUEST分组中的 RECORD_ROUTE 选项并显示返回分组上的路由缓冲区的内容。
- s 分组大小（packet size）—指明要被发送的数据字段的大小(缺省值为 56)。

Ping

- Ping: 用来测试
  - 目的地的可达性,
  - 计算往返时间
  - 计算到目的地的跳数
  - 可以提供记录路由选项.
- Ping失败不确保目的不存在. 防火墙可以过滤pings.

利用PING来诊断网络问题

1. Ping 127.0.0.1——用于测试TCP/IP协议是否运行正常
2. Ping 本机地址——测试网络设置（网卡）是否正常
3. Ping 对外连接的路由器（网关）——测试内部网络与对外连网的路由器是否正常
4. Ping Internet上计算机的IP地址——随便找一台Internet上的计算机，如果有响应，代表IP设置全部正常。
5. Ping Internet上计算机的域名——例如ping www.sina.com.cn，如果有响应，代表DNS设置无误。

选路控制与 ICMP 重定向机制

主机获得路由信息的方式 —— ICMP 重定向机制

- ICMP 重定向机制
  - 在网络中，选路由路由器/主机上的路由表决定
    - 目的：使主机能维持一个动态的、小规模、最优路由表
    - 机制：在转发分组时，路由器同时检查被转发分组，一旦发现其使用非最优通路，则向信源发送重定向报文，指出去往目的端的最佳路径
  - 路由器利用路由协议，通过路由器之间定期的路由信息交换来获得转发分组的路由信息
  - 主机则通过 ICMP 的重定向报文来获得路由信息
- 重定向机制用于同一个网络中的主机和路由器之间

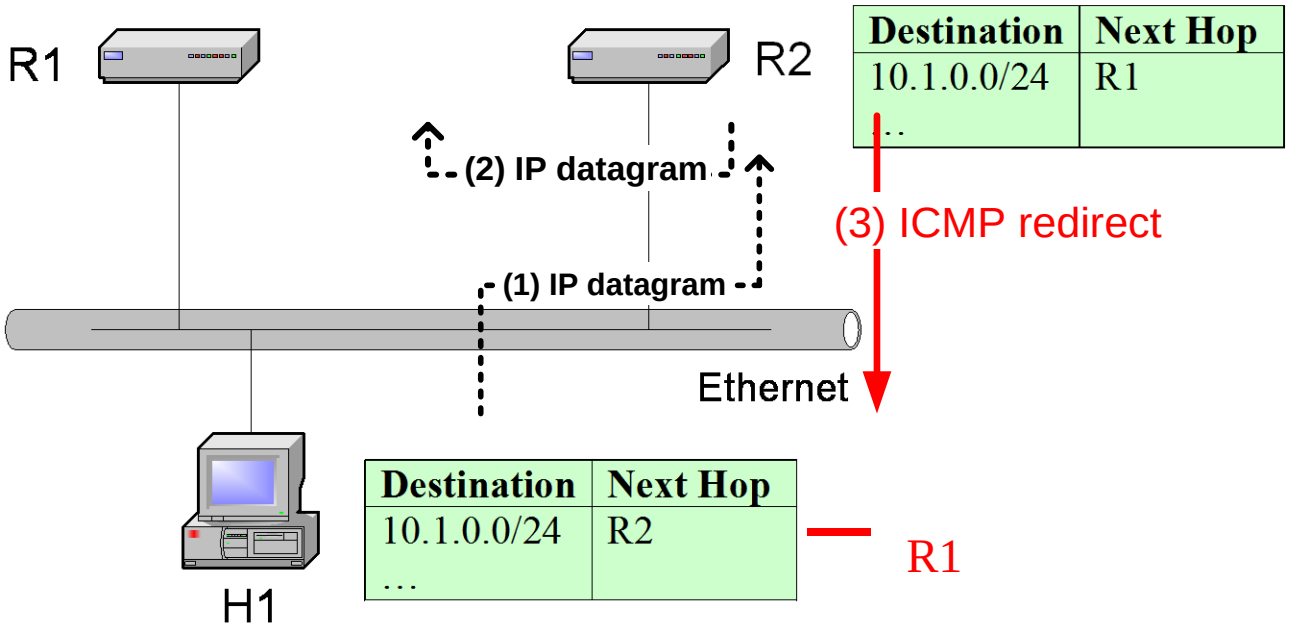
1B	1B	2B	4B	
类型 5	代码 0~3	校验和	路由器 IP 地址	重定向分组的 IP 头 + 前 64 位的数据

0：对网络重定向　　2：对服务类型和网络的重定向

1：对主机的重定向　　3：对服务类型和主机的重定向

Routing table manipulations with ICMP

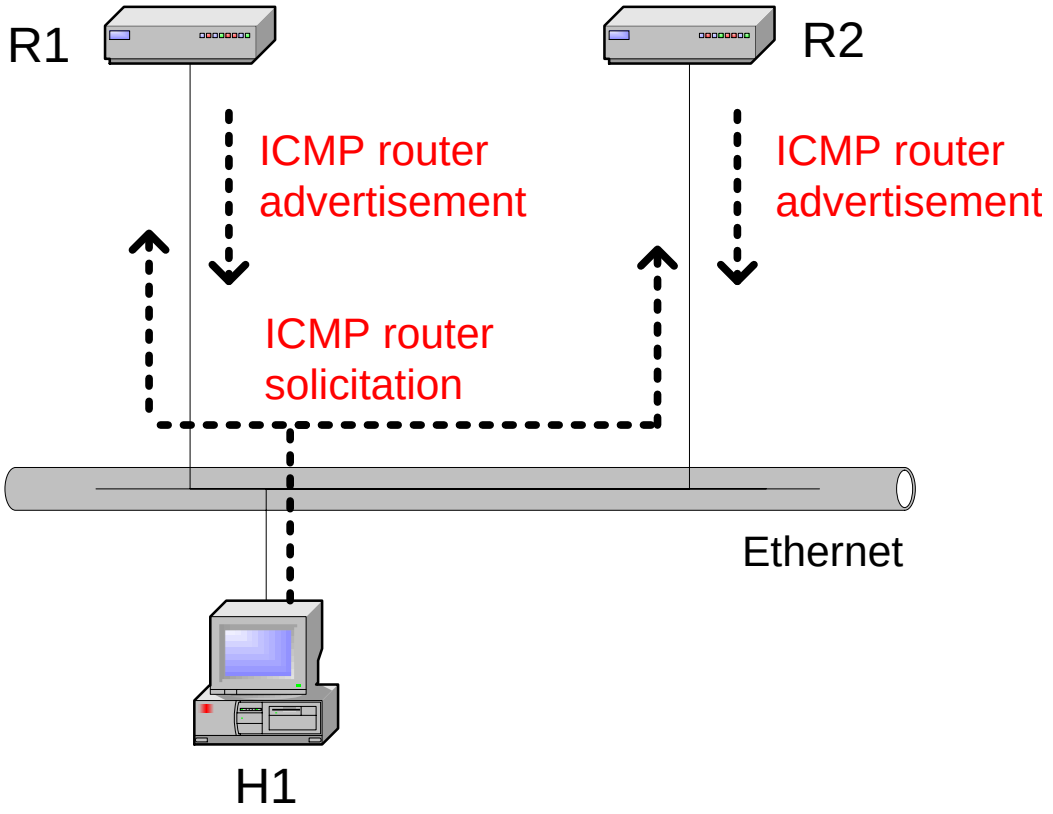
- When a router detects that an IP datagram should have gone to a different router, the router (here R2)
  - forwards the IP datagram to the correct router
  - sends an ICMP redirect message to the host
- Host uses ICMP message to update its routing table



ICMP Router Solicitation
ICMP Router Advertisement

- After bootstrapping a host broadcasts an **ICMP router solicitation**.
- In response, routers send an **ICMP router advertisement** message
- Also, routers periodically broadcast ICMP router advertisement

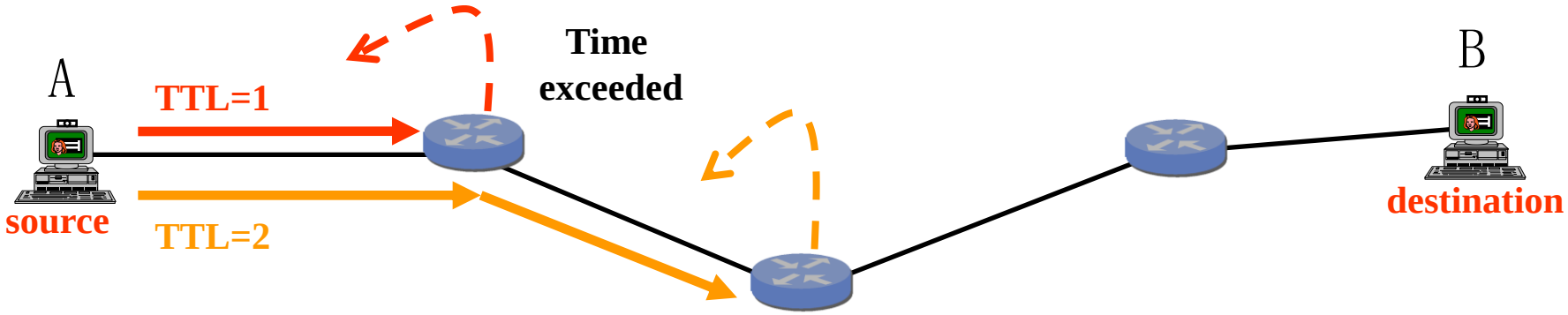
This is sometimes called the Router Discovery Protocol



基于ICMP的工具:跟踪路由

- 跟踪路由（Traceroute（unix)/tracert(win)）：利用 IP首部的TTL 和 ICMP，***TRACERT工具可找出至目的IP地址所经过的路有器。***

1. A发出Echo Request 1，目的地址为B，TTL=1;
2. R1路由器收到Echo Request 1后，因TTL=1便丢弃此封包，然后传送Timer Exceeded 1给A;
3. A收到Timer Exceeded 1之后，便可知道R1位路由过程中的第一个路由器。接着，A再发出Echo Request 2，目的地址仍为B，TTL=2;



Send packets with TTL=1, 2, 3, … and record source of "time exceeded" message

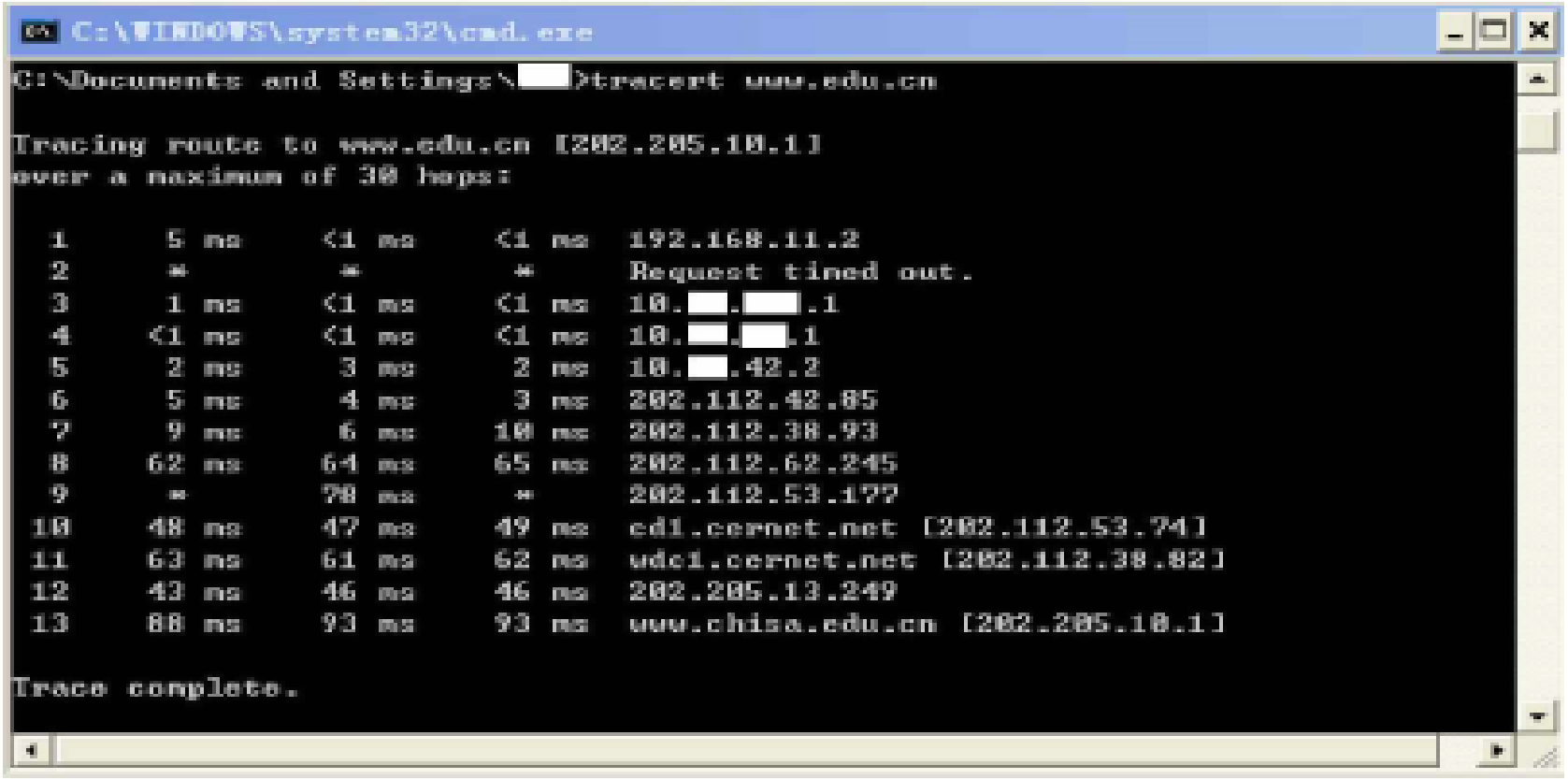
跟踪路由--TRACERT

4. Echo Request 2先送到R1，然后转送至R2，到达R2时， Echo Request 2的TTl=1，因此R2便丢弃此封包，然后传送Timer Exceeded 2给A;

5. A收到Timer Exceeded 2之后，便可知道R2位路由过程中的第二个路由器。接着，A再发出Echo Request 3，目的地址仍为B，TTL=3;

6. Echo Request 3经R1、R2，然后转送至B，B收到此封包后回应Echo Reply 1给A。A收到Echo Reply 1之后便大功告成。

跟踪路由--TRACERT



- 图 tracert路由追踪结果

- 第一跳1 5ms <lms <lms 192.168.11.2，其中192.168.11.2是本机网关。
- 第二跳2 * * * Request timed out.表明该网络设备没有响应，应该是二防火墙”。
- ……

PATHPING

- ***PATHPING可以认为是PING和TRACERT的结合，先找出至目的地IP地址所经过的路有器，然后依次对每部路***
- ***由器发出Echo Request包，以监测路由器是否正常。***

基于ICMP的工具:路径MTU 发现

- 发送一个IP数据报，并且设定不分片比特位。
  - 在链路中不能分片将导致ICMP错误消息。
  - 后续的ICMP版本在ICMP消息中指定MTU的大小。
- 减小MSS直到成功发送（没有收到ICMP错误消息）

其它因特网网络层协议

（自学）

- **IP V6**
- **IP组播**
- 虚拟专用网（ VPN ）
- IP 隧道传输（ Tunneling ）
- 移动 IP
- 在其它广域网技术上的IP
  - ATM
  - 帧中继
  - X.25

下一代的网际协议 IPv6 (IPng)
----解决 IP 地址耗尽的措施

- 从计算机本身发展以及从因特网规模和网络传输速率来看，现在 IPv4 已很不适用。
- 最主要的问题就是 32 bit 的 IP 地址不够用。
- 要解决 IP 地址耗尽的问题的措施：
  - 采用无类别编址 CIDR，使 IP 地址的分配更加合理。
  - 采用网络地址转换 NAT方法以节省全球 IP 地址。
  - 采用具有更大地址空间的新版本的 IP 协议 IPv6。

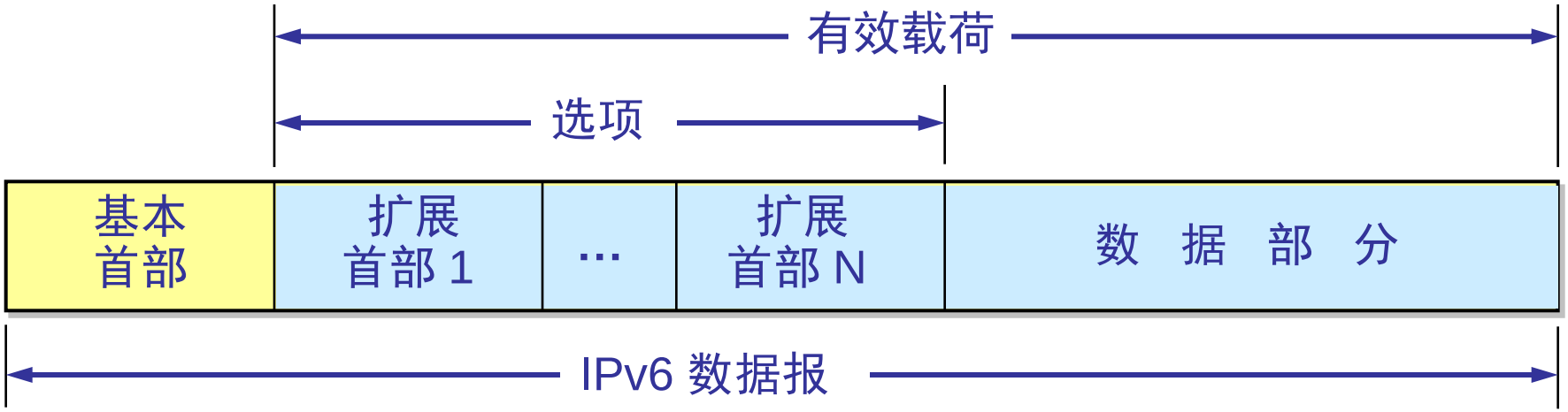
IPv6 的基本首部

- IPv6 所引进的主要变化如下
- 更大的地址空间。IPv6 将地址从 IPv4 的 32 bit 增大到了 128 bit，
- 扩展的地址层次结构。
- 灵活的首部格式。
- 改进的选项。
- 允许协议继续扩充。
- 支持即插即用（即自动配置）
- 支持资源的预分配。

IPv6 数据报的首部

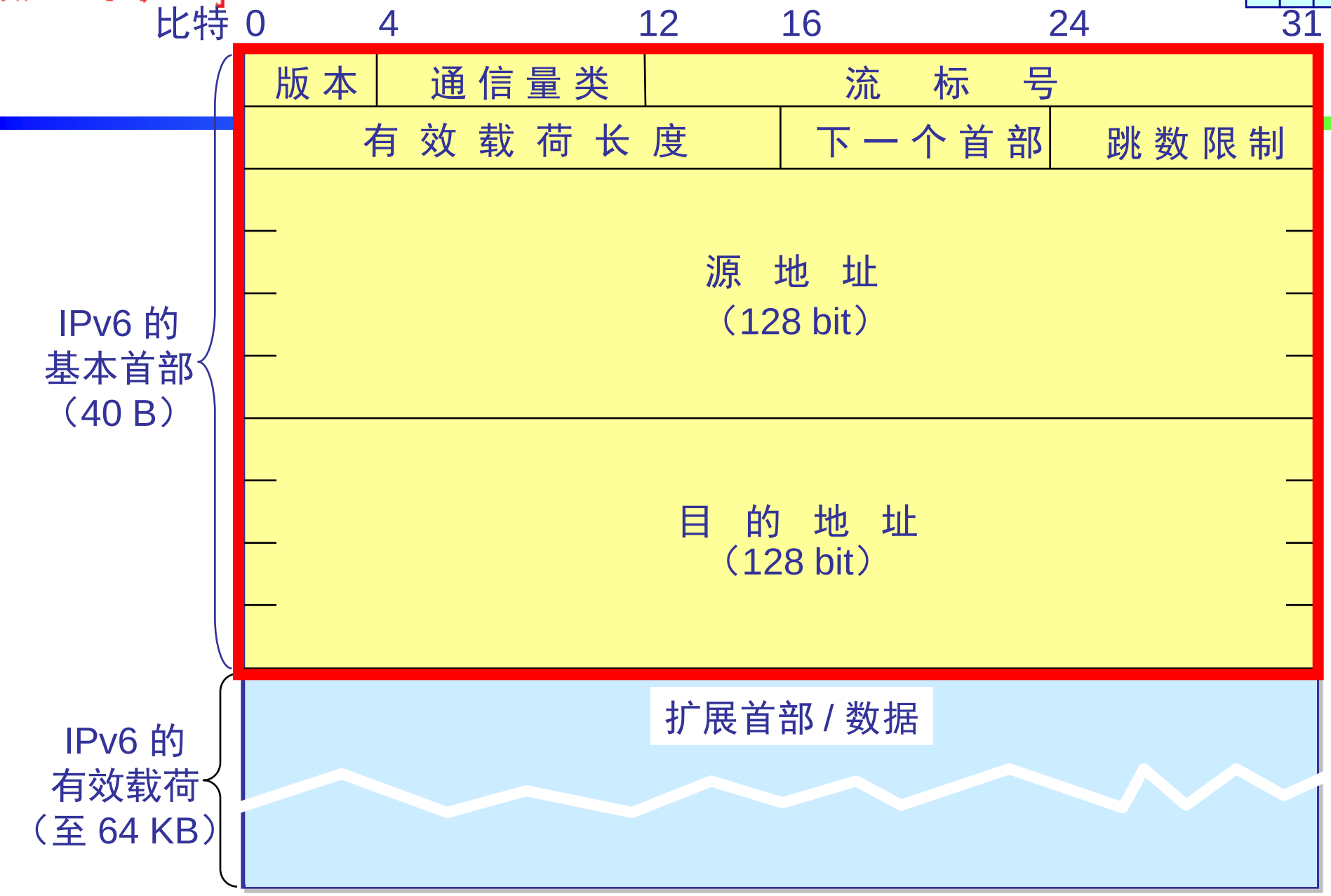
- IPv6 将首部长度变为固定的 40 字节，称为基本首部(base header)。
- 将不必要的功能取消了，首部的字段数减少到只有 8 个。
- 取消了首部的检验和字段，加快了路由器处理数据报的速度。
- 在基本首部的后面允许有零个或多个扩展首部。
- 所有的扩展首部和数据合起来叫做数据报的有效载荷(payload)或净负荷。

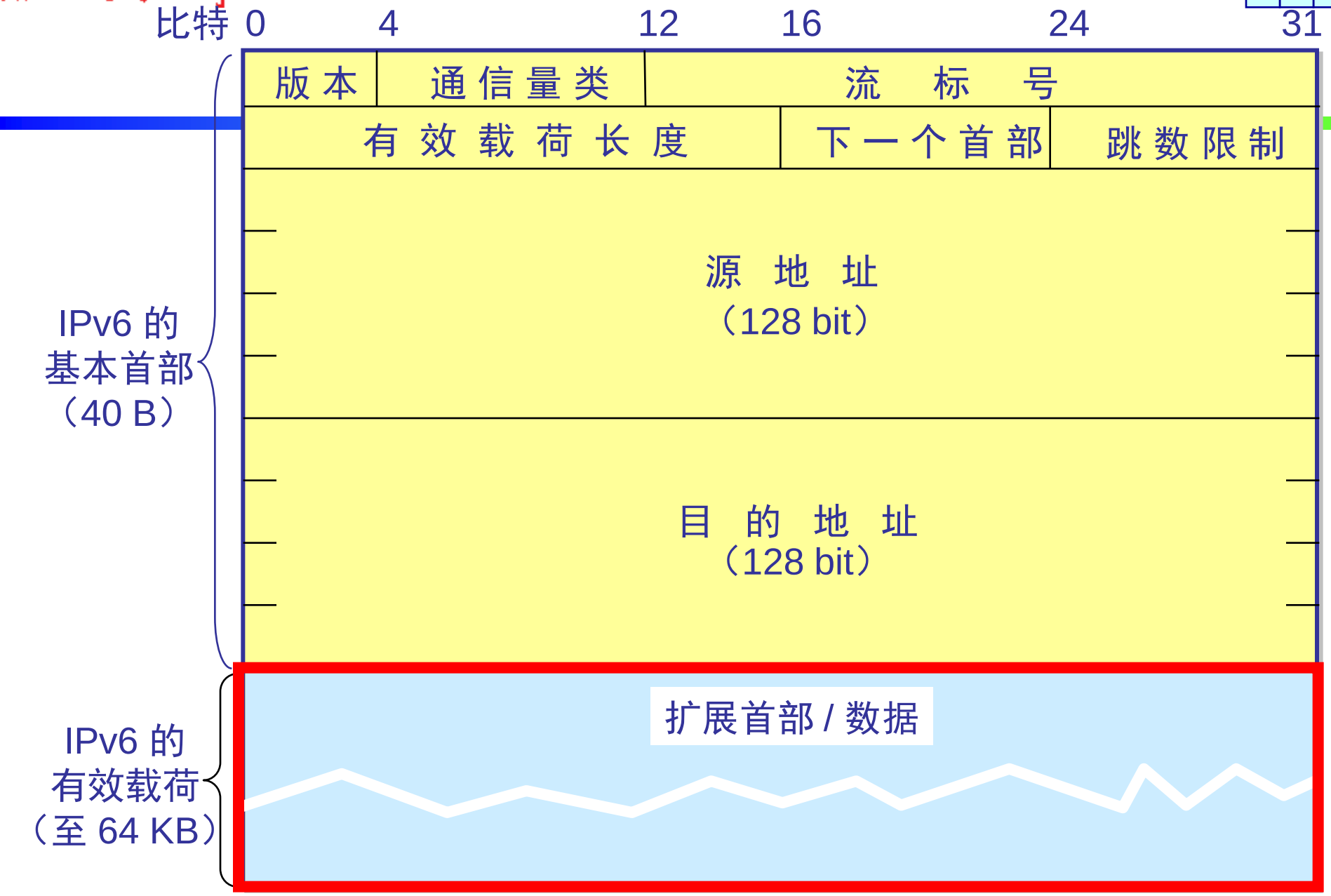
IPv6 数据报的一般形式

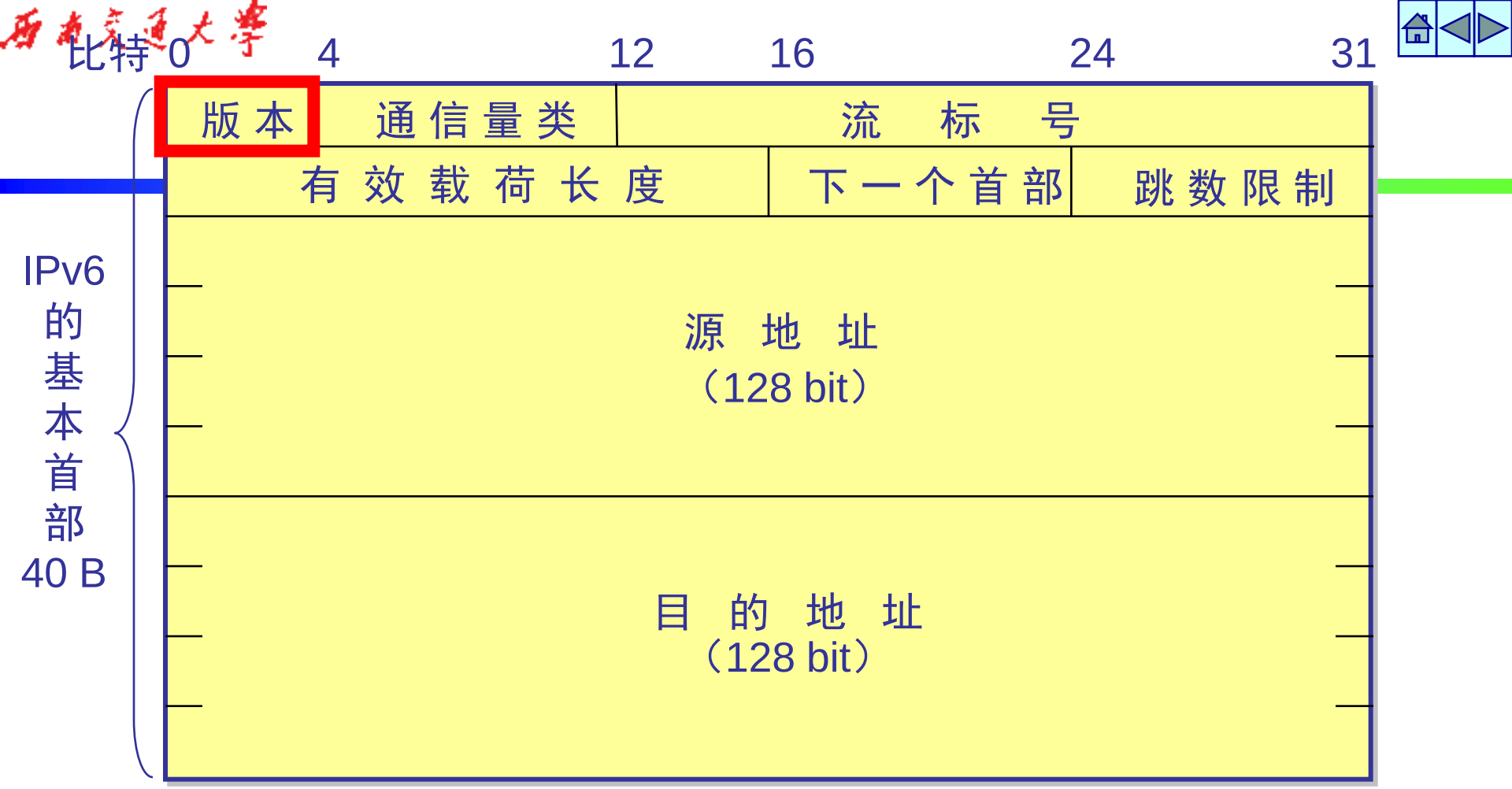


IPv6 数据报首部与 IPv4 数据报首部的对比

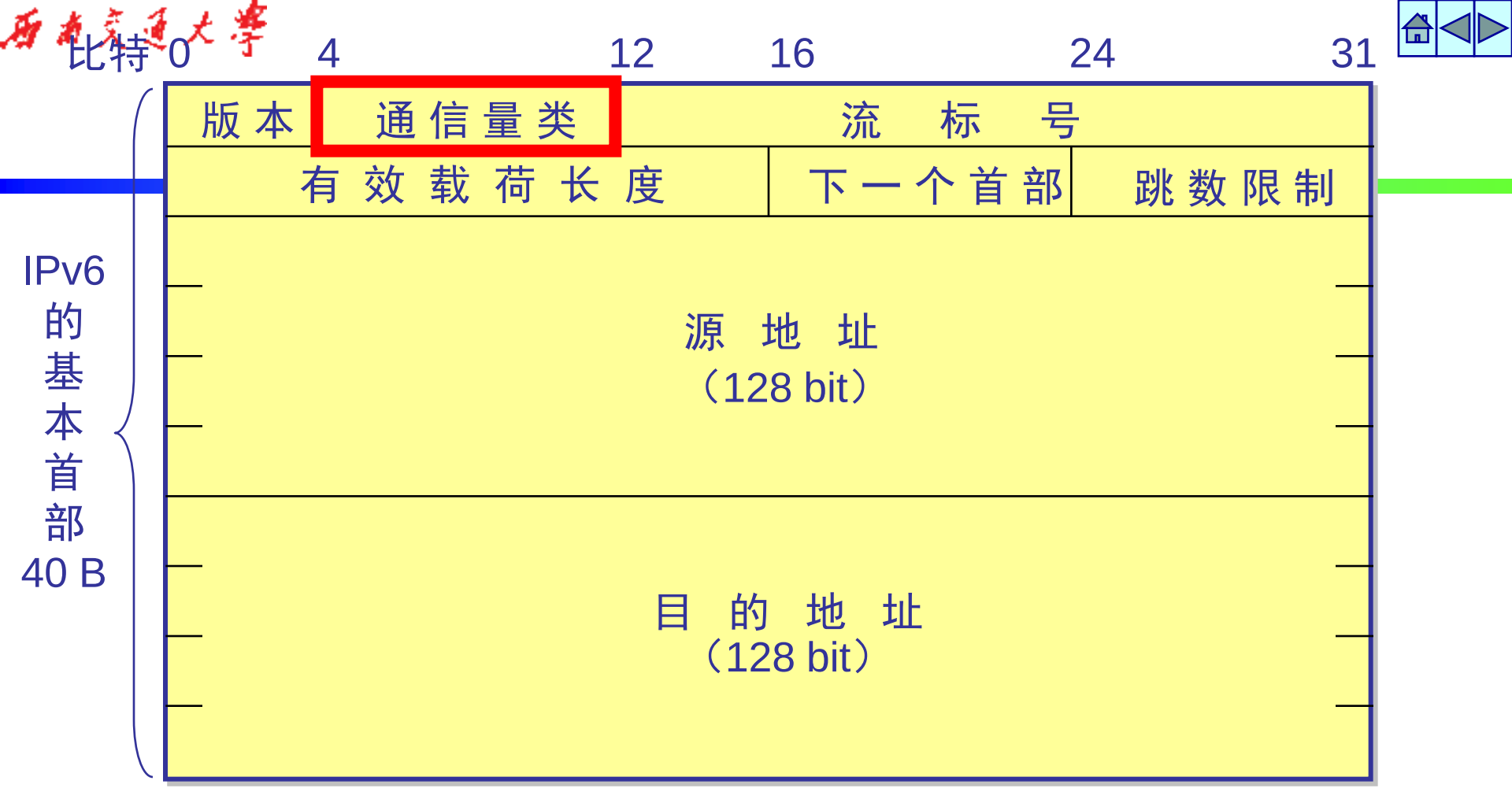
上面是 IPv4 数据报的首部


比特 0
4
12
16
24
31
版本
通信量类
流 标 号
有效载荷长度
下一个首部
跳数限制
源 地 址
（128 bit）
目 的 地 址
（128 bit）
IPv6 的
基本首部
（40 B）
扩展首部 / 数据
IPv6 的
有效载荷
（至 64 KB）

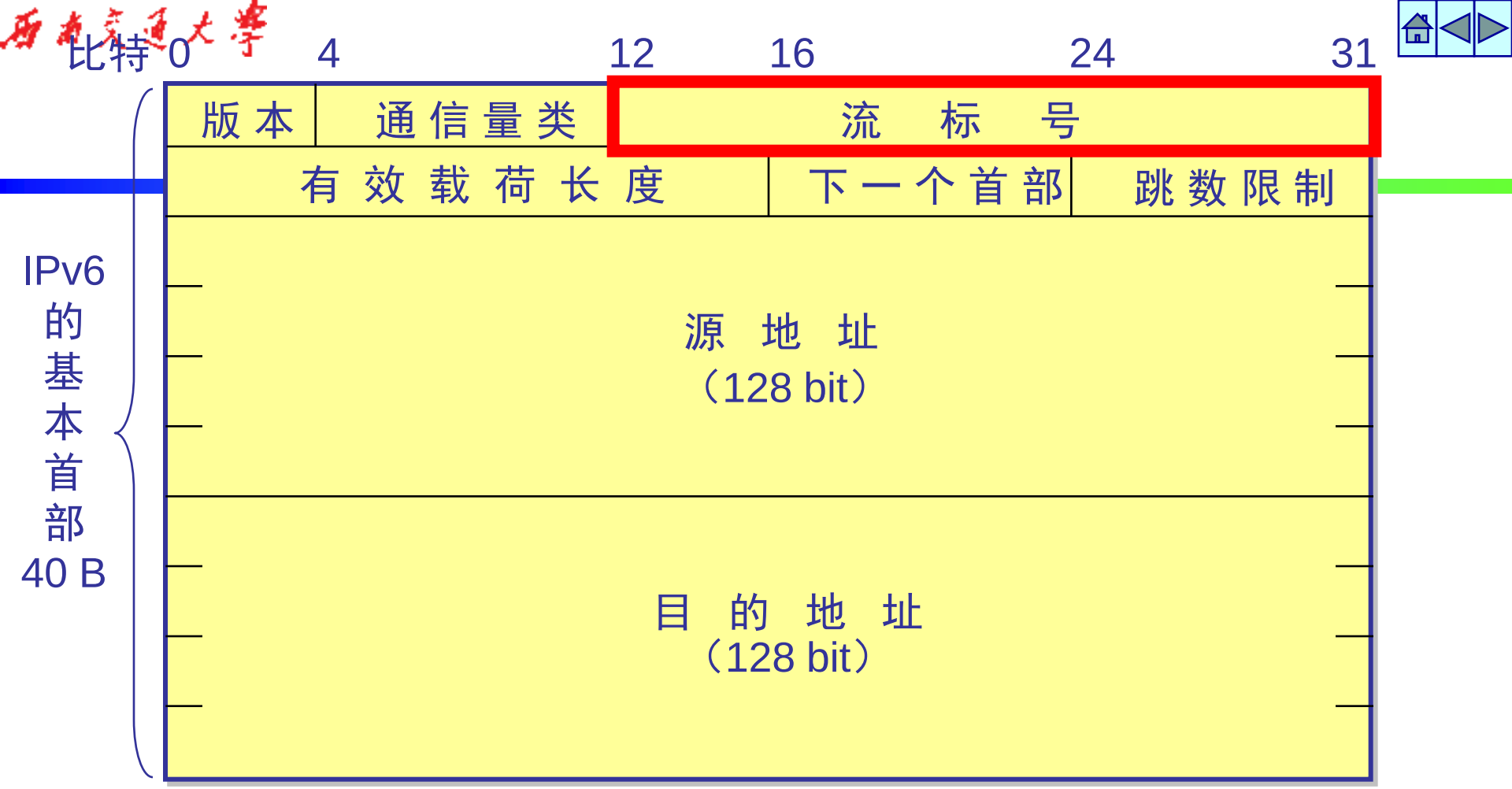

比特 0
4
12
16
24
31
版本
通信量类
流 标 号
有效载荷长度
下一个首部
跳数限制
源 地 址
（128 bit）
目 的 地 址
（128 bit）
IPv6 的
基本首部
（40 B）
扩展首部 / 数据
IPv6 的
有效载荷
（至 64 KB）



版本(version)—— 4 bit。它指明了协议的版本，对 IPv6 该字段总是 6。

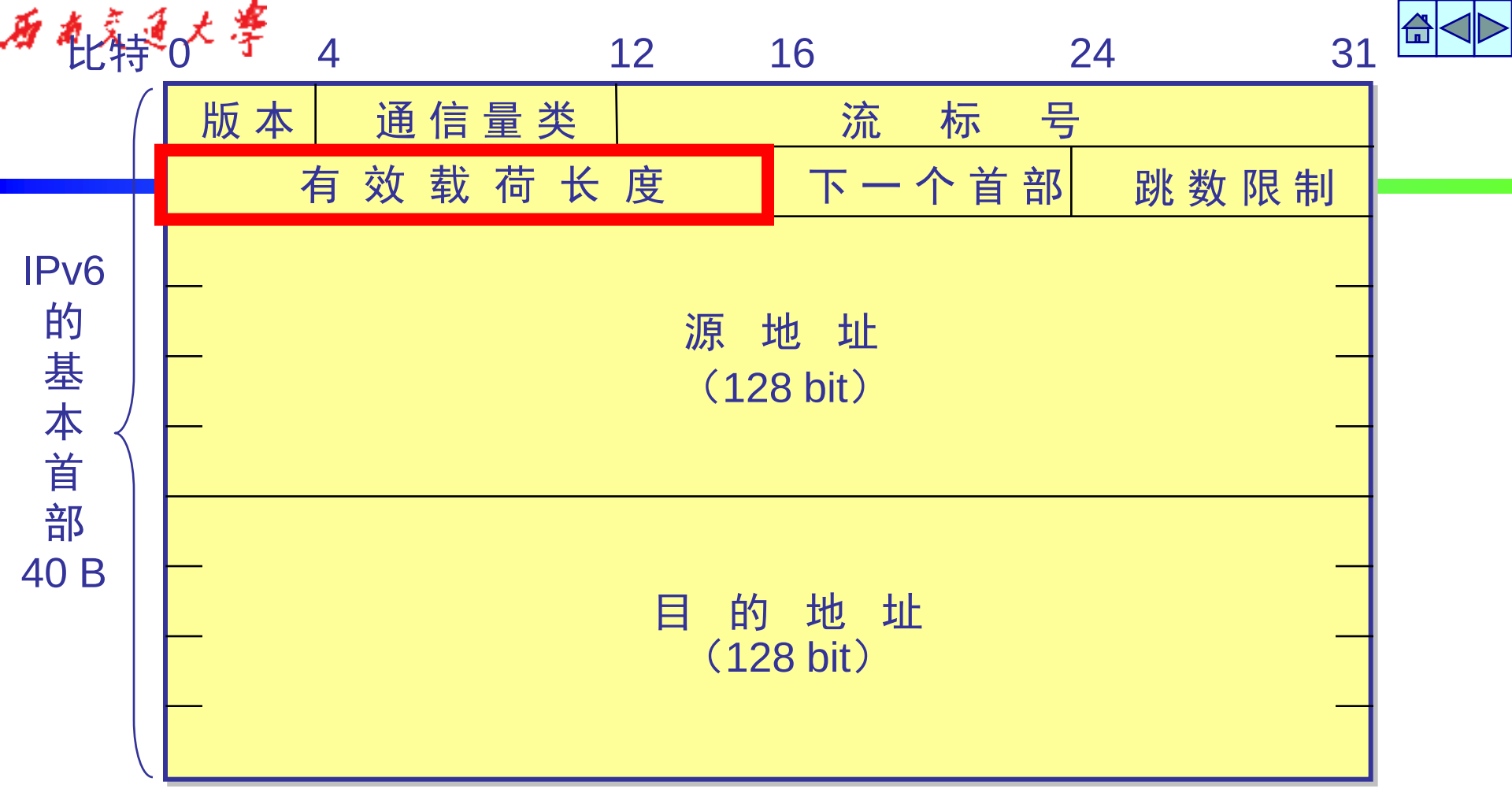


通信量类(traffic class)—— 8 bit。这是为了区分不同的 IPv6 数据报的类别或优先级。目前正在进行不同的通信量类性能的实验。

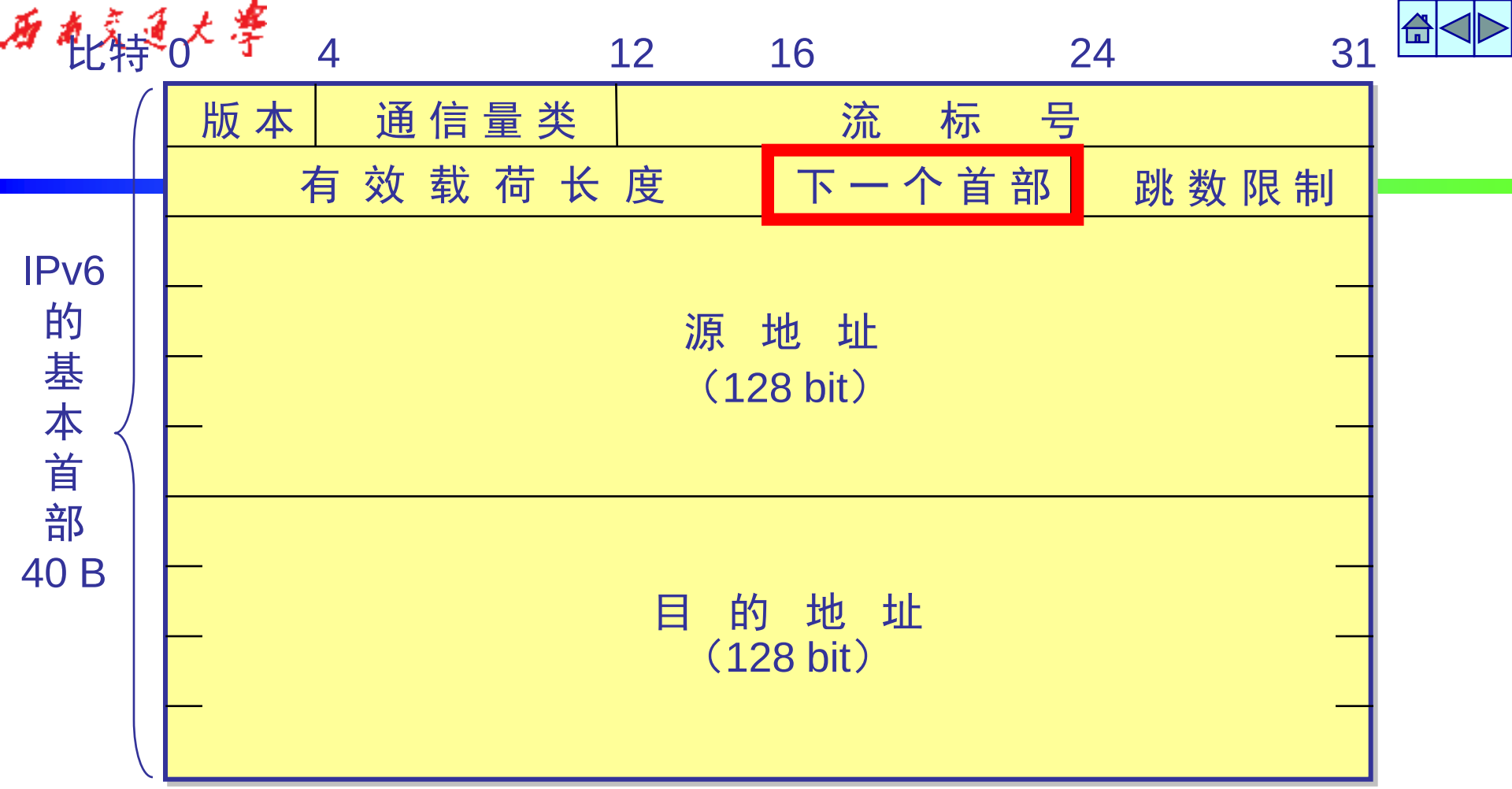


流标号(flow label)—— 20 bit。“流”是互联网络上从特定源点到特定终点的一系列数据报，“流”所经过的路径上的路由器都保证指明的服务质量。

所有属于同一个流的数据报都具有同样的流标号。



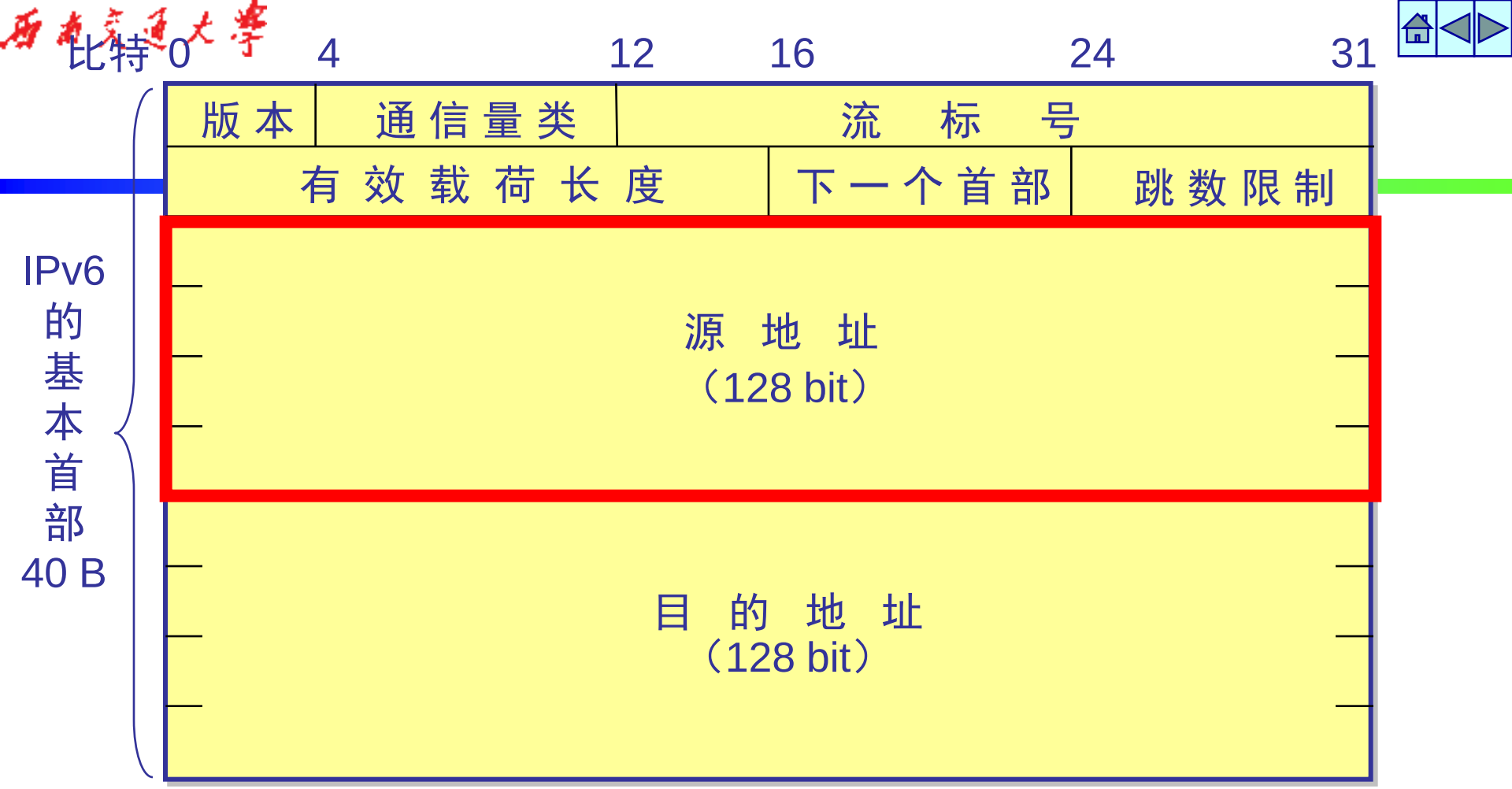
有效载荷长度(payload length)—— 16 bit。它指明 IPv6 数据报除基本首部以外的字节数（所有扩展首部都算在有效载荷之内），其最大值是 64 KB。



下一个首部(next header)—— 8 bit。它相当于 IPv4 的协议字段或可选字段。

跳数限制(hop limit)—— 8 bit。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。



源地址—— 128 bit。是数据报的发送站的 IP 地址。

目的地址—— 128 bit。是数据报的接收站的 IP 地址。

IPv6 的扩展首部

1. 扩展首部及下一个首部字段

- IPv6 将原来 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）。
- 这样就大大提高了路由器的处理效率。

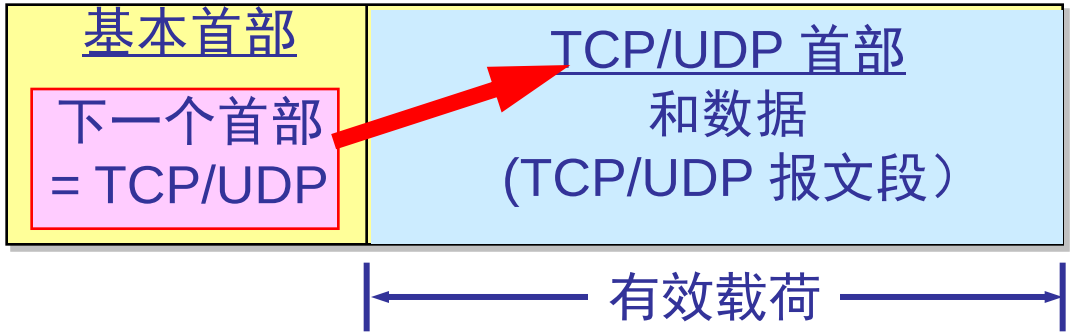
六种扩展首部

在[RFC 2460]中定义了六种扩展首部：

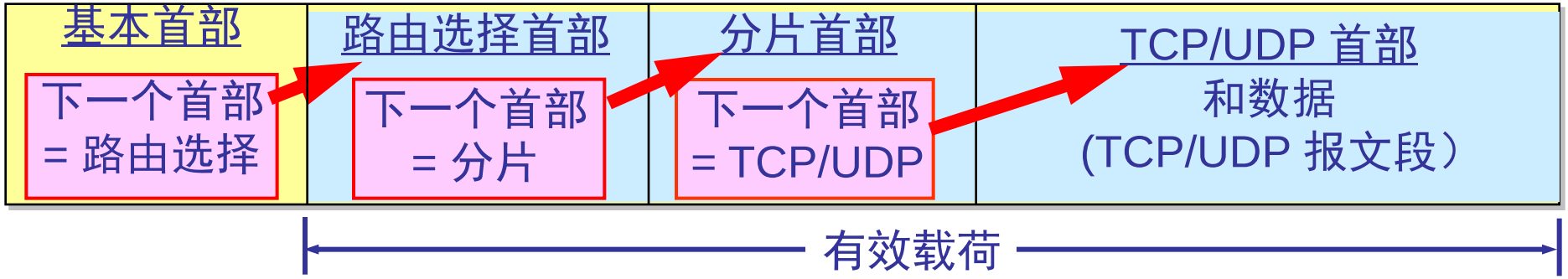
- 逐跳选项
- 路由选择
- 分片
- 鉴别
- 封装安全有效载荷
- 目的站选项

IPv6 的扩展首部

无扩展首部



有扩展首部

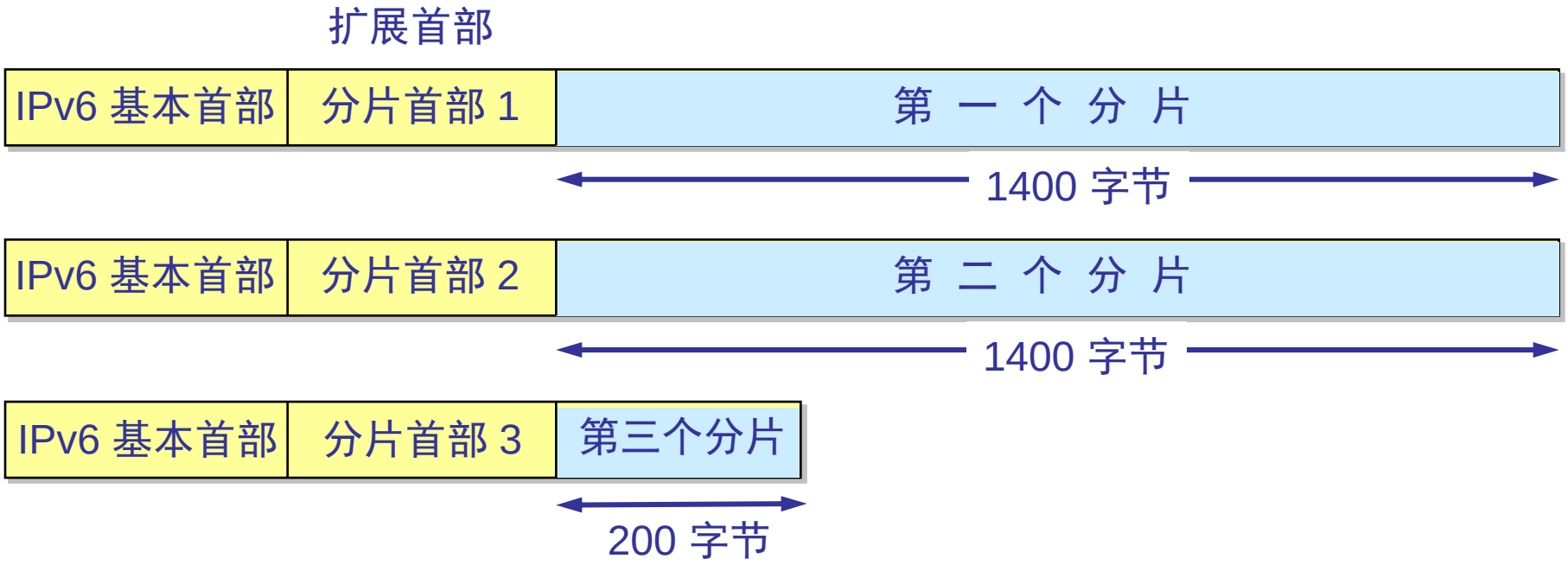


2. 扩展首部举例

- IPv6 将分片限制为由源站来完成。源站可以采用保证的最小 MTU（1280字节），或者在发送数据前完成路径最大传送单元发现(Path MTU Discovery)，以确定沿着该路径到目的站的最小MTU。
- 分片扩展首部的格式如下：

扩展首部举例

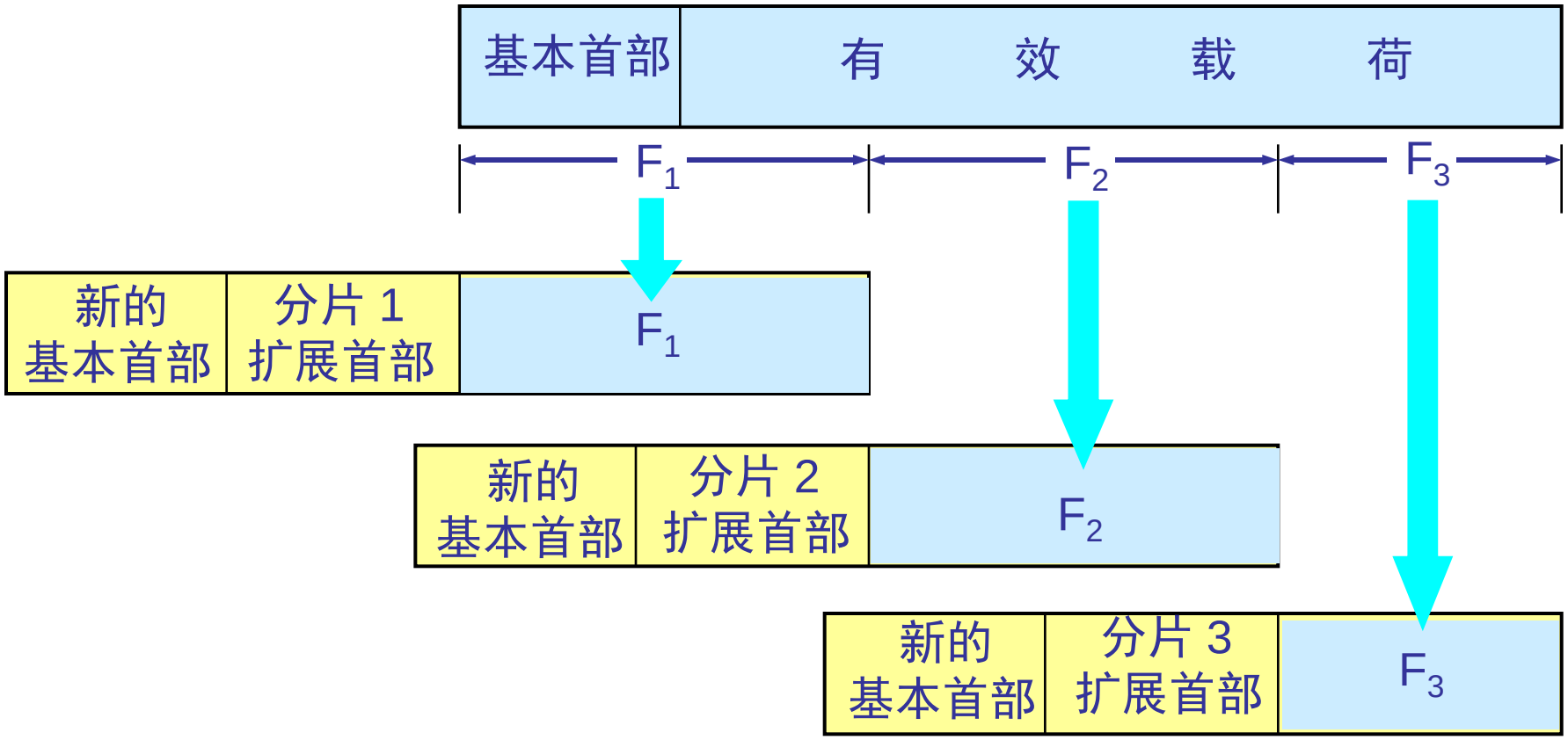
- IPv6 数据报的有效载荷长度为 3000 字节。下层的以太网的最大传送单元 MTU 是 1500 字节。
- 分成三个数据报片，两个 1400 字节长，最后一个是 200 字节长。



用隧道技术来传送长数据报

- 当路径途中的路由器需要对数据报进行分片时，就创建一个全新的数据报，然后将这个新的数据报分片，并在各个数据报片中插入扩展首部和新的基本首部。
- 路由器将每个数据报片发送给最终的目的站，而在目的站将收到的各个数据报片收集起来，组装成原来的数据报，再从中抽取出数据部分。

用隧道技术将一个 IPv6 数据报分成 3 个数据报片



IPv6 的地址空间

1. 128 bit 的地址空间

IPv6 数据报的目的地址可以是以下三种基本类型地址之一：

(1) 单播(unicast)　单播就是传统的点对点通信。

(2) 多播(multicast)　多播是一点对多点的通信。

(3) 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给其中的一个，通常是距离最近的一个。

结点与接口

- IPv6 将实现 IPv6 的主机和路由器均称为结点。
- IPv6 地址是分配给结点上面的接口。
  - 一个接口可以有多个单播地址。
  - 一个结点接口的单播地址可用来惟一地标志该结点。

冒号十六进制记法 (colon hexadecimal notation)

- 每个 16 bit 的值用十六进制值表示，各值之间用冒号分隔。

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

- 零压缩(zero compression)，即一连串连续的零可以为一对冒号所取代。
- FF05:0:0:0:0:0:0:B3　　可以写成：
- FF05::B3

点分十进制记法的后缀

- 0:0:0:0:0:0:128.10.2.1

  再使用零压缩即可得出： ::128.10.2.1
- CIDR 的斜线表示法仍然可用。
- 60 bit的前缀 12AB00000000CD3 可记为：

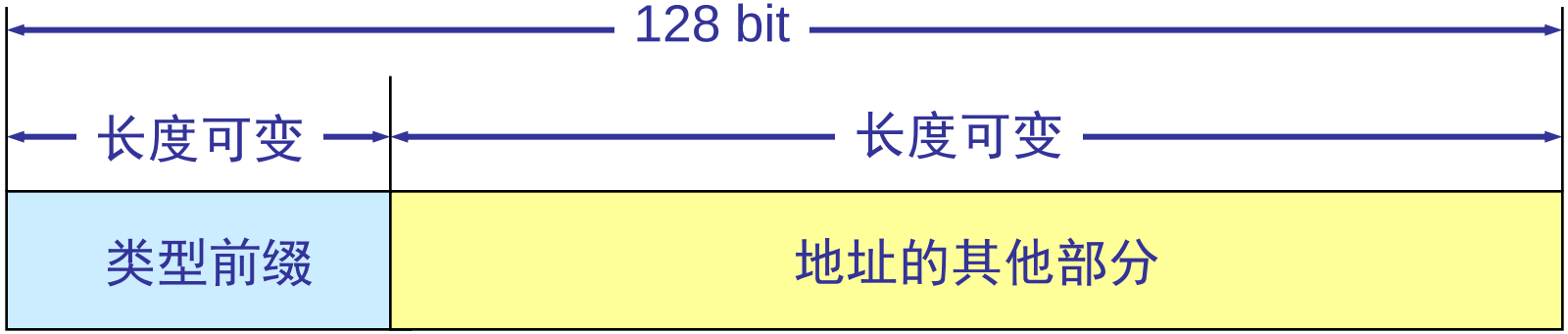
  12AB:0000:0000:CD30:0000:0000:0000:0000/60

或12AB::CD30:0:0:0:0/60

或12AB:0:0:CD30::/60

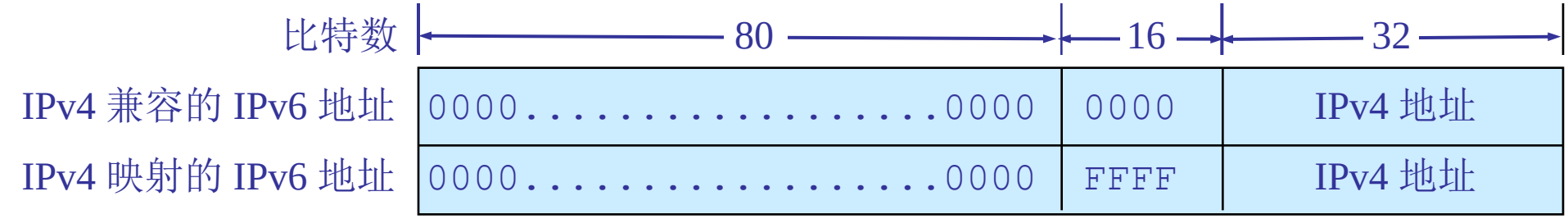
2. 地址空间的分配

- IPv6 将 128 bit 地址空间分为两大部分。
  - 第一部分是可变长度的类型前缀，它定义了地址的目的。
  - 第二部分是地址的其余部分，其长度也是可变的。



前缀为 0000 0000 的地址

- 前缀为 0000 0000 是保留一小部分地址与 IPv4 兼容的，这是因为必须要考虑到在比较长的时期 IPv 4和 IPv6 将会同时存在，而有的结点不支持 IPv6。
- 因此数据报在这两类结点之间转发时，就必须进行地址的转换。



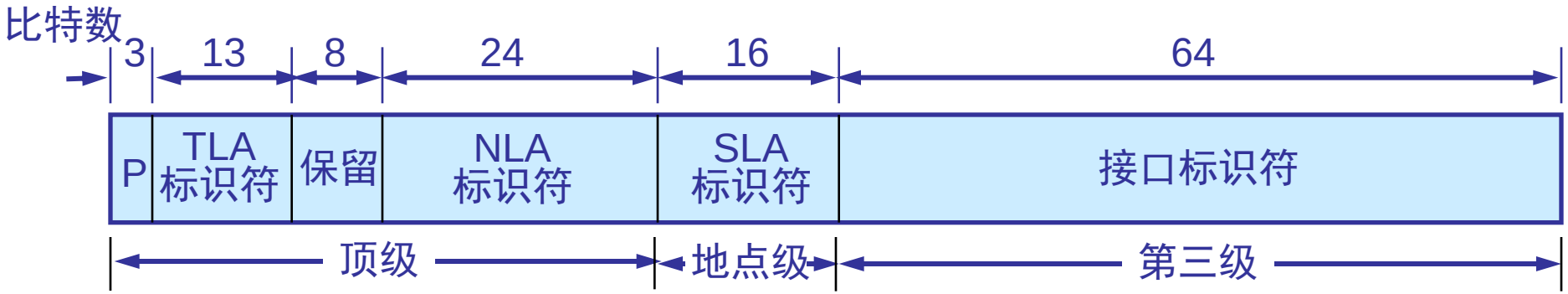
IPv6 单播地址的等级结构

IPv6 扩展了地址的分级概念，使用以下三个等级：

(1) 第一级（顶级），指明全球都知道的公共拓扑。

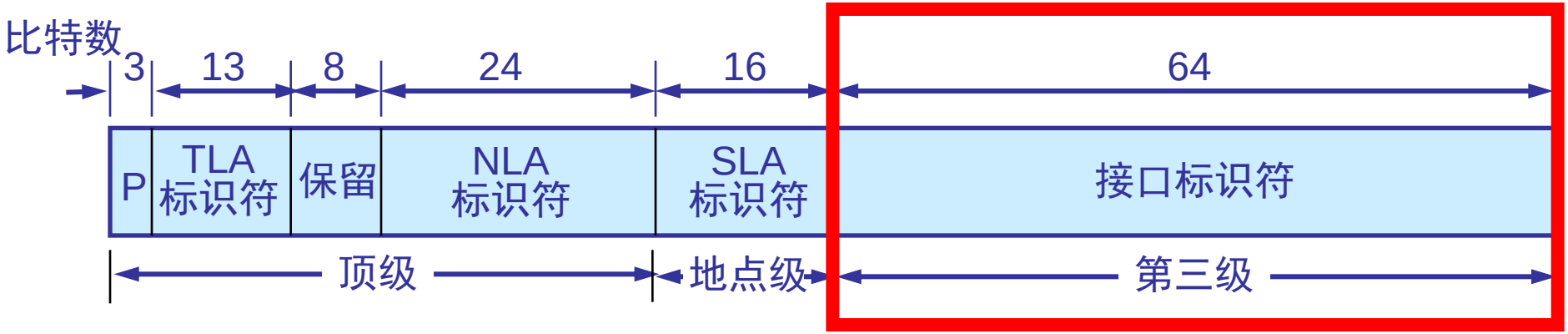
(2) 第二级（地点级），指明单个的地点。

(3) 第三级，指明单个的网络接口。



第三级地址

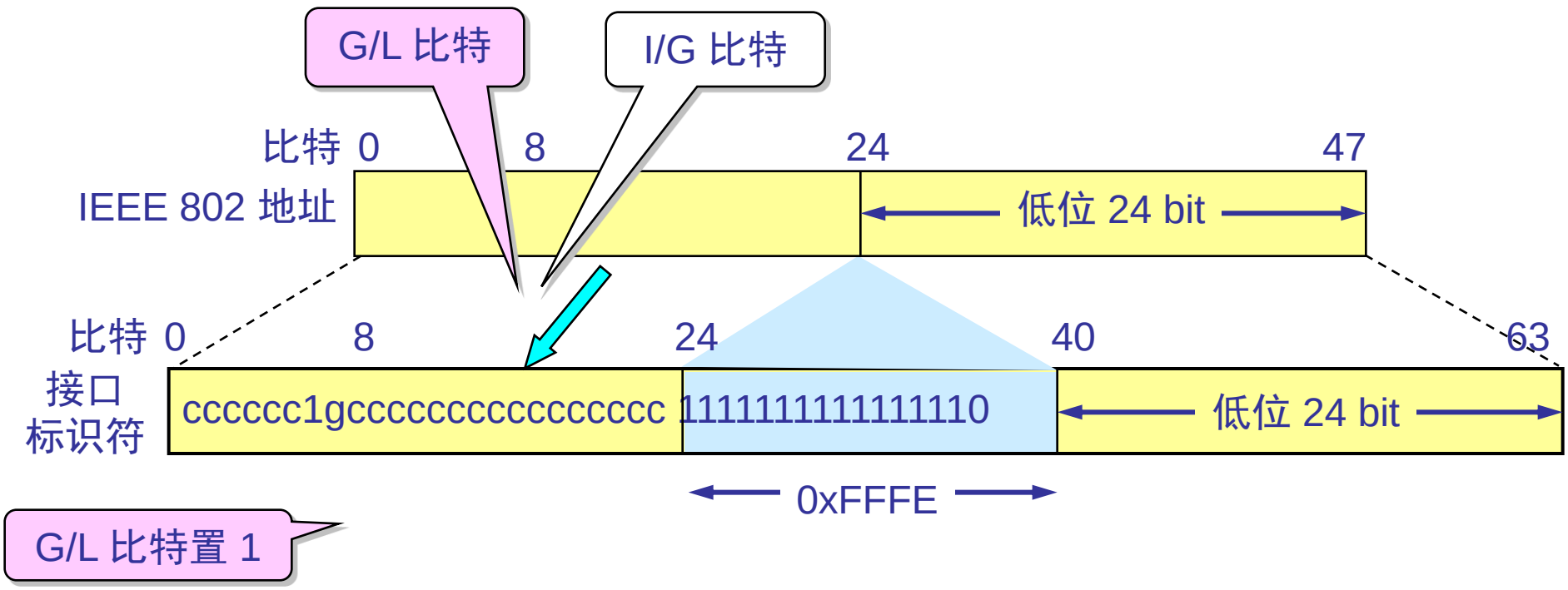
- IPv6 地址的最低的第三级对应于计算机和网络的单个接口。
- IPv6 地址的后缀有 64 bi t之多，它足够大，因而可以将各种接口的硬件地址直接进行编码。
- IPv6 使用邻站发现协议使结点能够确定哪些计算机是和它相邻接的。



EUI-64

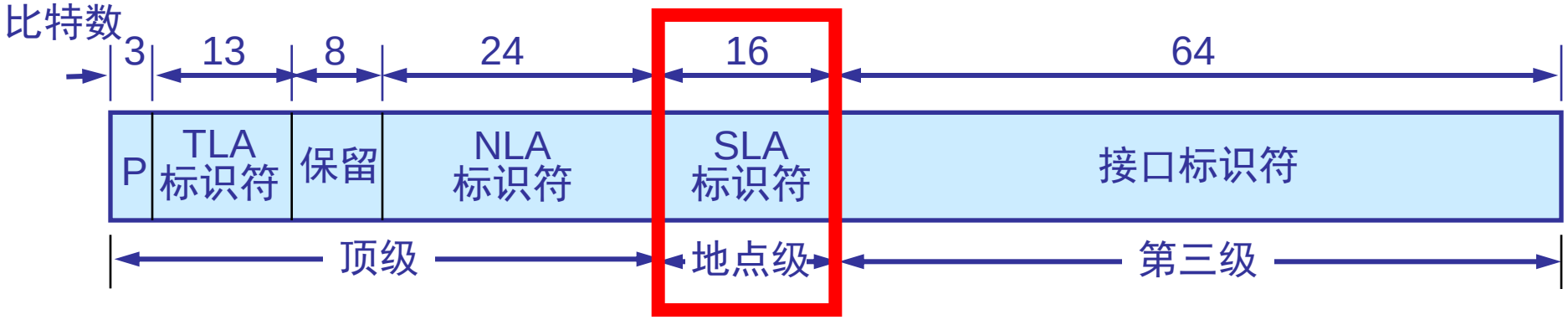
- IEEE定 义了一个标准的 64 bit 全球惟一地址格式 EUI-64。
- EUI-64 的前三个字节（24 bit）仍为公司标识符，但后面的扩展标识符是五个字节（40 bit）。
- 较为复杂的是当需要将 48 bit 的以太网硬件地址转换为 IPv6 地址。

将以太网地址转换为 IPv6 地址



第二级地址

- IPv6 地址中间的第二级对应于在一个地点的一组计算机和网络，它们通常是相距较近的且都归一个单位来管理。
- SLA 级表示 Site Level Aggregation，即地点级聚合，它和 IPv4 中的子网字段相似。



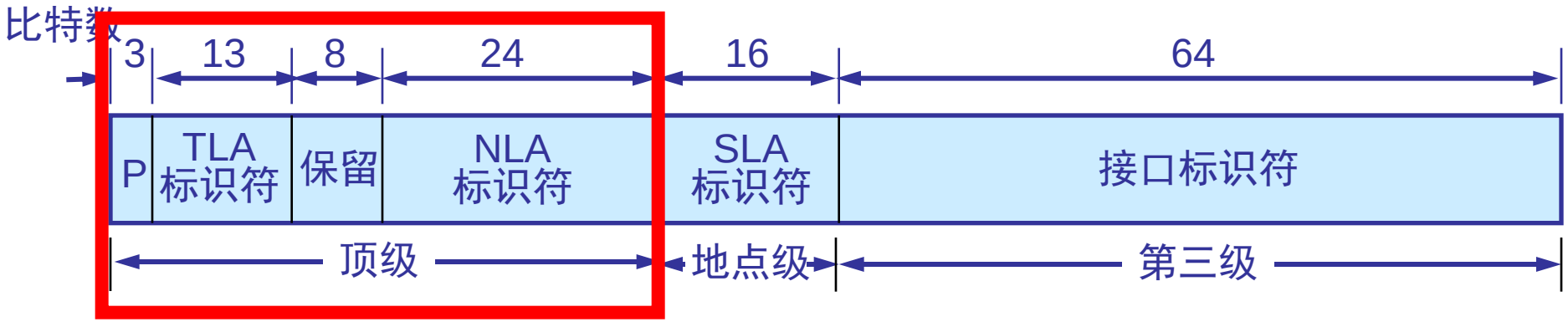
第一级地址（有四个字段）

(1) P字段—— 3 bit，即格式前缀。

(2) 顶级聚合标识符 TLA ID——13 bit，指派给ISP或拥有这些地址的汇接点(exchange)。

(3) 保留字段—— 8 bit。

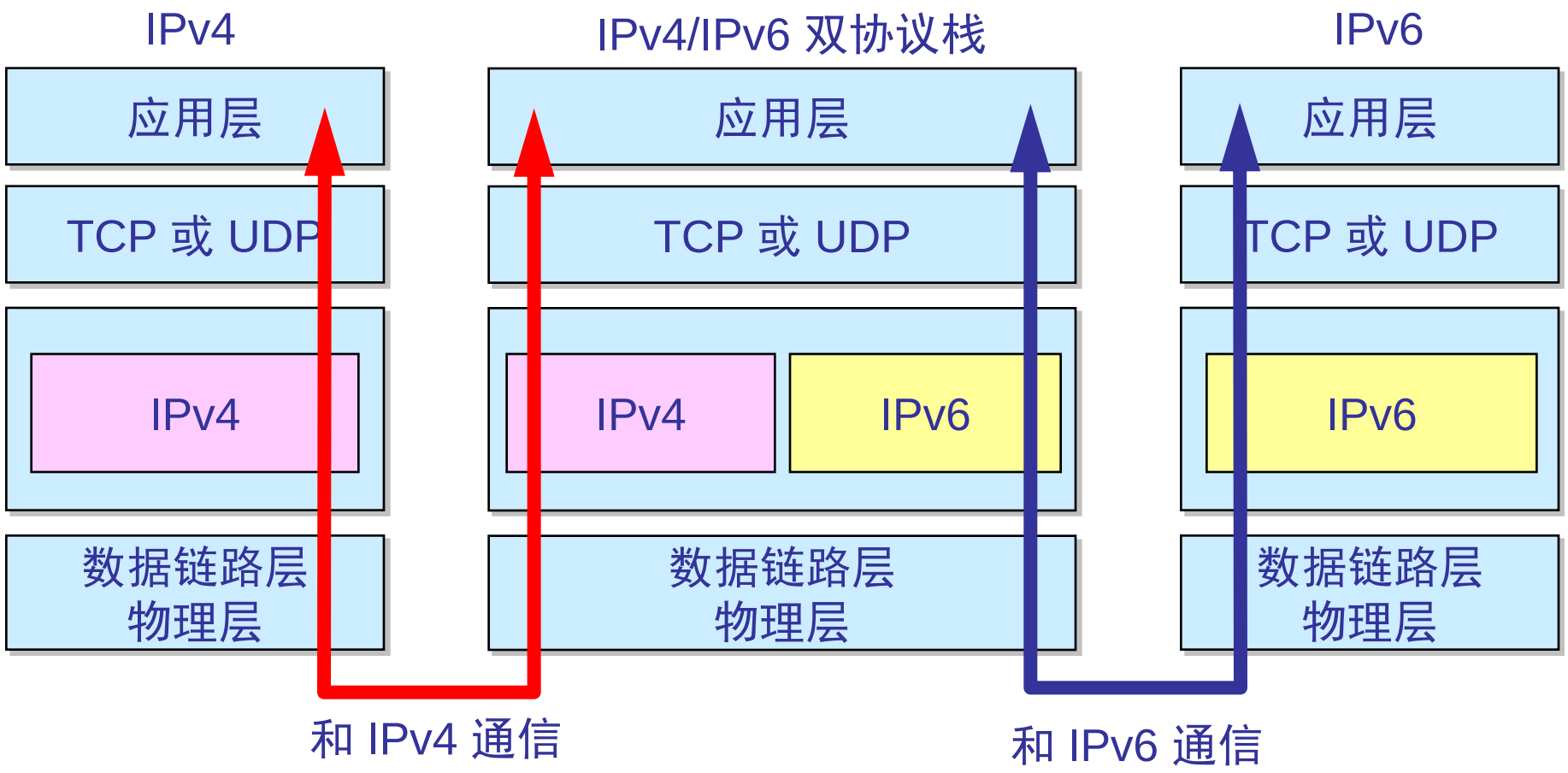
(4) 下一级聚合标识符 NLA ID—— 16 bit。指派给一个特定的用户。



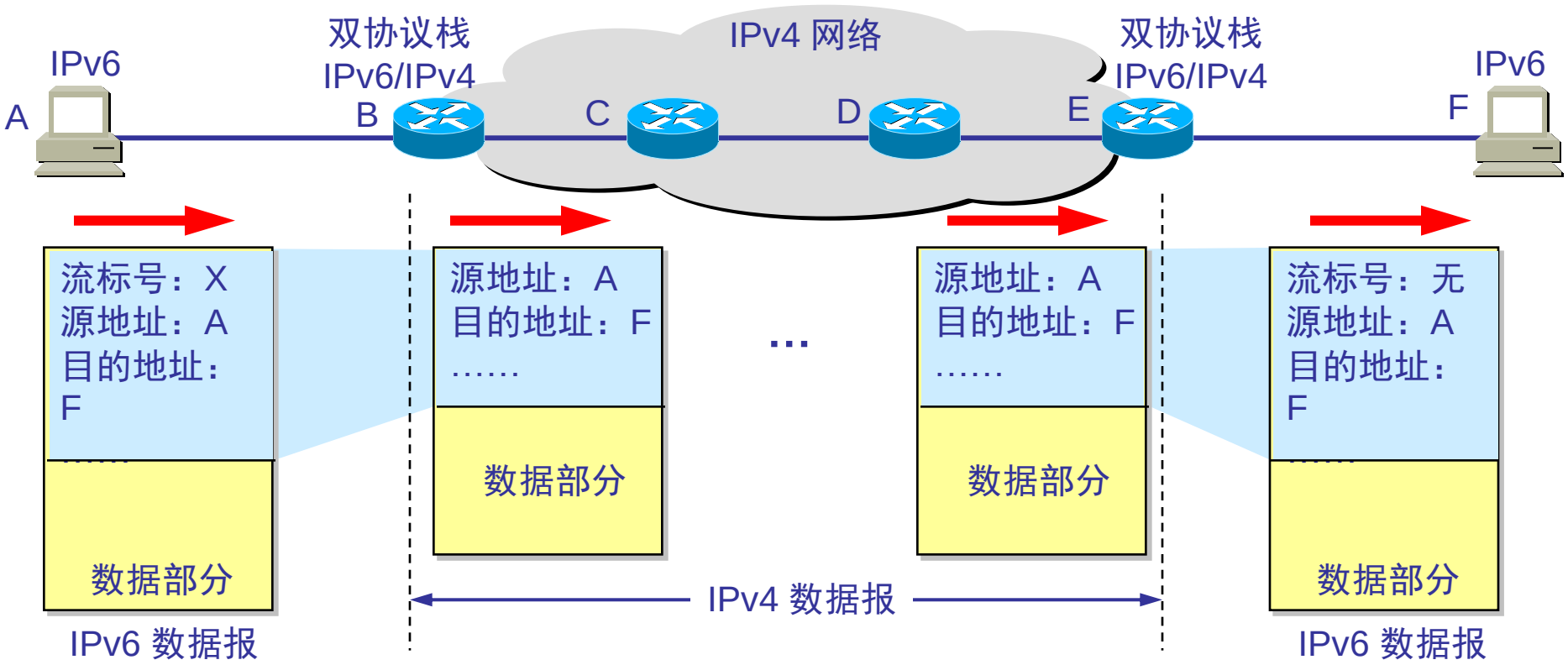
从 IPv4 向 IPv6 过渡

- 向 IPv6 过渡只能采用逐步演进的办法，同时，还必须使新安装的 IPv6 系统能够向后兼容。
- IPv6 系统必须能够接收和转发 IPv4 分组，并且能够为 IPv4 分组选择路由。
- 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。

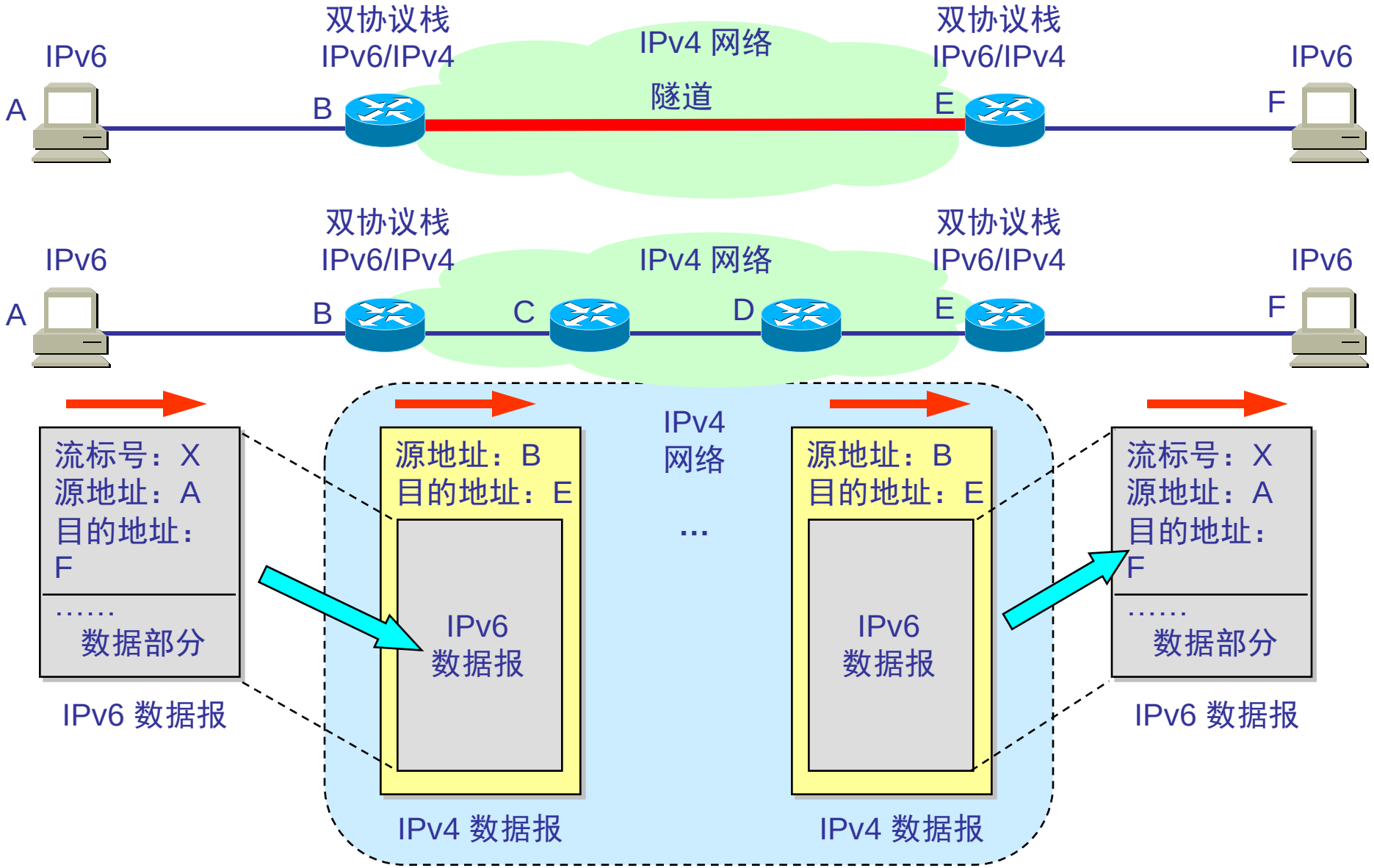
双协议栈


IPv4
IPv4/IPv6 双协议栈
IPv6
应用层
TCP 或 UDP
IPv4
IPv4
IPv6
IPv6
数据链路层
物理层
数据链路层
物理层
数据链路层
物理层
和 IPv4 通信
和 IPv6 通信

用双协议栈进行从 IPv4 到 IPv6 的过渡



使用隧道技术从 IPv4 到 IPv6 过渡



ICMPv6

- ICMPv6 的报文格式和 IPv4 使用的 ICMP 的相似，即前4个字节的字段名称都是一样的。
- 但 ICMPv6 将第 5 个字节起的后面部分作为报文主体。
- ICMPv6 的报文划分为两大类
  - 差错报文(error message)
  - 提供信息的报文(informational message
  - 取消了使用得很少的 ICMP 报文

IP组播

- 单个数据流可以发送到多个客户端的组播能力已成为大多数多媒体应用的传输手段。
- 组播技术利用一个IP地址使IP数据报文发送到用户组。IP组播采用了特殊定义的目的IP地址和目的MAC地址。
- IGMP为客户端提供加入和离开组播组的方式。CGMP使路由器为交换机配置组播转发表，并告诉交换机当前的组播成员。
- 指派路由器根据对网络中的组播成员的分布和使用的不同采用密集模式DM或稀疏模式SM组播路由协议来构造组播的分布树，而这个分布树将在源子网和组播组之间确定一条唯一路径以提高数据传输效率。

IP组播与组播协议

- 在Internet上，多媒体业务诸如：流媒体，视频会议和视频点播等，正在成为信息传送的重要组成部分。点对点传输的单播方式不能适应这一类业务传输特性--单点发送多点接收，因为服务器必须为每一个接收者提供一个相同内容的IP报文拷贝，同时网络上也重复地传输相同内容的报文，占用了大量资源。如图1.1所示。虽然IP广播允许一个主机把一个IP报文发送给同一个网络的所有主机，但是由于不是所有的主机都需要这些报文，因而浪费了网络资源。在这种情况下组播（multicast）应运而生，它的出现解决了一个主机向特定的多个接收者发送消息的方法。1989年，IETF通过RFC1112，定义了Internet上的组播方式。

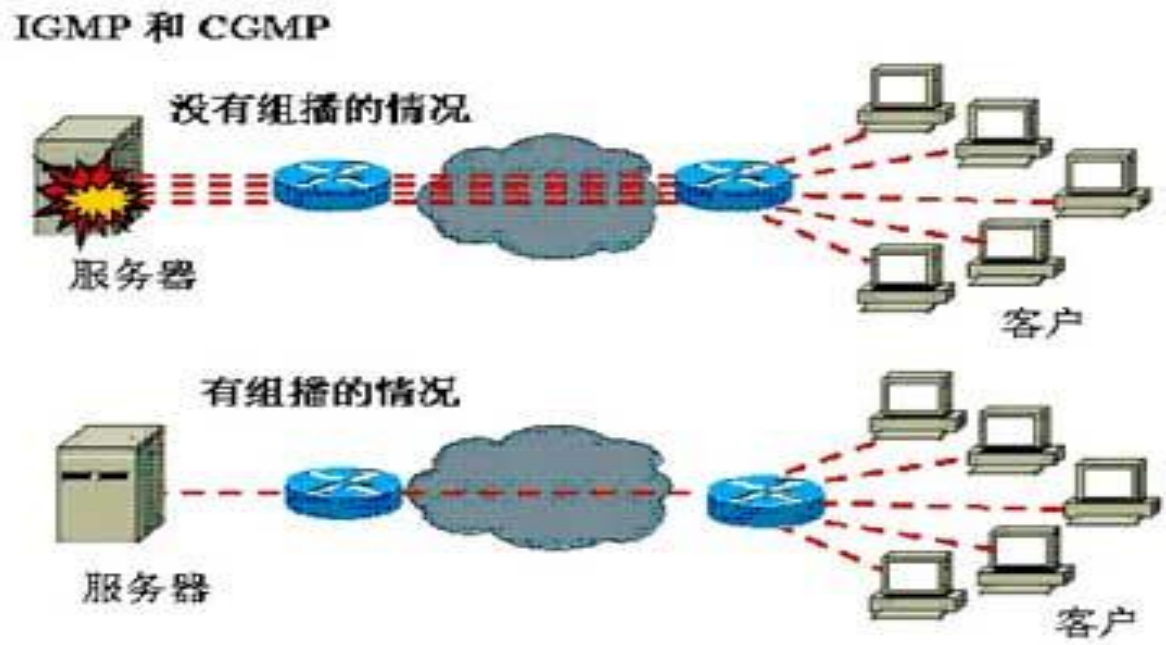


图1.1

IP组播

- IP组播是指一个IP报文向一个“主机组”的传送，这个包含零个或多个主机的主机组由一个单独的IP地址标识。主机组地址也称为“组播地址”，或者D类地址。除了目的地址部分，组播报文与普通报文没有区别，网络尽力传送组播报文但是并不保证一定送达。
- 主机组的成员可以动态变化，主机有权选择加入或者退出某个主机组。主机可以加入多个主机组，也可以向自己没有加入的主机组发送数据。主机组有两种：永久组和临时组。永久组的IP地址是周知的，由Internet管理机构分配，是保留地址。临时组的地址则使用除永久组地址外的非保留D类地址。
- IP组播分组在互联网上的转发由支持组播的路由器来处理。主机发出的IP组播分组在本子网内被所有主机组成员接收，同时与该子网直接相连的组播路由器会把组播报文转发到所有包含该主机组成员的网络上。组播报文传递的范围由报文的生存期值(TTL, Time-to-Live)决定，如果TTL值等于或者小于设置的路由器端口TTL门限值（TTL Threshold），路由器将不再转发该报文。

组播地址

- IP组播地址，或称为主机组地址，由D类IP地址标记。D类IP地址的最高四位为“1110”，起范围从224.0.0.0到239.255.255.255。如前所述，部分D类地址被保留，用作永久组的地址，这段地址从224.0.0.0-224.0.0.255。比较重要的地址有：
- 224.0.0.1 — 网段中所有支持组播的主机
  224.0.0.2 — 网段中所有支持组播的路由器
  224.0.0.4 — 网段中所有的DVMRP路由器
  224.0.0.5 — 所有的OSPF路由器
  224.0.0.6 — 所有的OSPF指派路由器
  224.0.0.9 — 所有RIPv2路由器
  224.0.0.13 —所有PIM路由器
- 临时主机组的组播地址由网络管理员选择，他需要保证这个地址在一定的范围内没有其他的主机组在使用这个组播地址。

组播地址

- 第2层的组播地址（组播MAC地址）可以从IP组播地址中衍生。计算方法是把IP地址的最后23位拷贝到MAC地址的最后23位，然后把这23位前面的那一位置为0。MAC地址的前24位必须为0x01-00-5E。例如：组播IP地址224.0.1.128，16进制表示为0xE0-00-01-10，最低的23位为0x00-01-10，计算得出的MAC地址为：0x01-00-5E-00-01-10。

Internet组管理协议（IGMP）

- IGMP协议由主机成员关系协议发展而来，目前有两个版本：IGMPv1（RFC1112），IGMPv2 (RFC2326)。主机使用IGMP消息通告本地的组播路由器它想接收组播流量的主机组地址。如果主机支持IGMPv2，它还可以通告组播路由器它退出某主机组。组播路由器通过IGMP协议为其每个端口都维护一张主机组成员表，并定期的探询表中的主机组的成员，以确定该主机组是否存活。
IGMP消息被置于IP报文中传送。

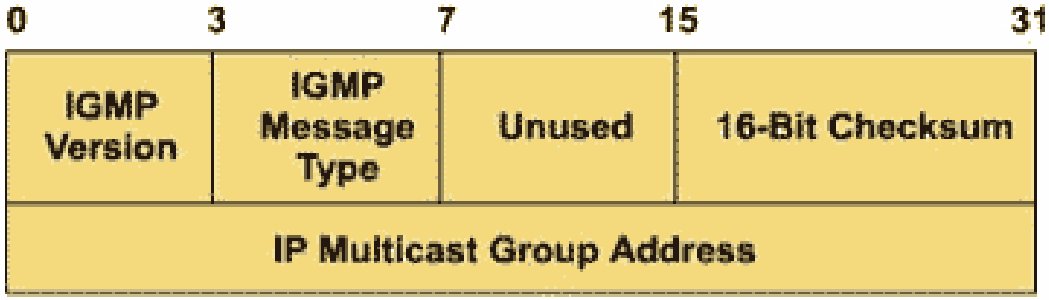


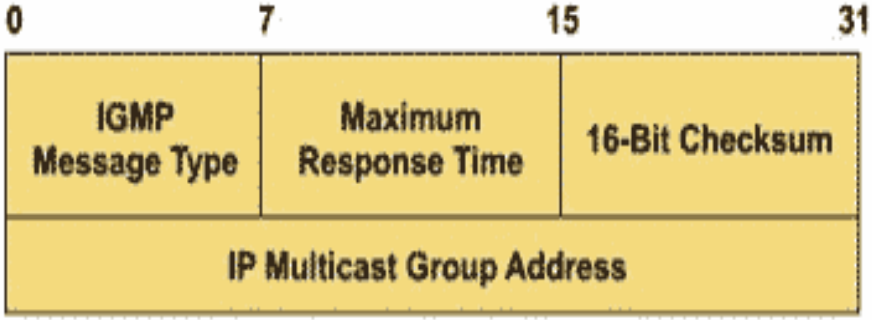
图1.2　IGMPv1的报文格式

Internet组管理协议

- IGMPv1中定义了两种消息类型：主机成员询问和主机成员报告。当某主机想要接收某个组播流量时，它向本地的组播路由器发送"主机成员报告"消息，告知欲接收的组播地址。组播路由器收到"主机成员报告"消息后把该主机加入指定的主机组，并在设定的周期内向组播地址224.0.0.1(代表所有支持组播的主机) 发送"主机成员询问"消息。主机如果还想继续接收组播流量，必须发送"主机成员报告"消息。

Internet组管理协议

- IGMPv2的报文如下图所示。与IGMPv1不同的是它将版本字段和消息类型字段融合，把未使用字段作了“最大响应时间”字段。IGMPv2报文的消息类型字段定义了四种消息类型：



- 0x11 - 成员询问
  0x12 - IGMPv1 成员报告
  0x16 - IGMPv2 成员报告
  0x17 - 退出主机组

IGMPv2向前兼容IGMPv1协议， IGMPv1的设备可以接收处理IGMPv2的消息报文。 IGMPv2中允许路由器对指定的主机组地址做"成员询问"，非该组的主机不必响应。如果某主机想退出，它可以主动向路由器发送"推出主机组"消息，而不必像IGMPv1中那样只能被动退出。

CGMP协议

- 在交换网络中，2层交换机可能既不了解哪个端口有哪些组播组，也不能在其源MAC地址表中找到组播MAC地址的表项。从而，交换机只能简单地把组播报文向所有端口转发，组播的优势将大大削弱。因此，Cisco提出CGMP协议，让组播路由器来配置交换机的组播转发表，从而彻底解决交换网络中的组播问题。
- CGMP (Cisco Group management protocol)全称Cisco组管理协议，采用CGMP的路由器将主机加入或者退出组播组的IGMP消息通知交换机，交换机则根据该消息将该主机所在端口从组播转发表中加入或者删除。通过CGMP协议的使用，2层交换机可以掌握接收组播的主机的情况，从而提高整个网络的性能和利用率。

分布树（Distribution Tree）

- 在传送组播分组时，指派路由器需要构造一个连接所有组播组成员的树。根据这个树，路由器得出转发分组的一条唯一路径。这个树就称为分布树。由于成员可以动态的加入和退出，分布树也必须动态更新。
- 根据构造方法的不同，分布树分为源分布树（Source Distribution Tree）和共享分布树（Shared Distribution Tree）。源分布树以组播源为根节点构造到所有组播组成员的生成树，通常也称为最短路径树（SPT）。共享分布树，也称为RP树或基于核心的树（CBT, Core_based Tree）。它的构造方法是以网络中的某一个指定的路由器为根节点，该路由器称为集合点或中心点，由此节点生成包含所有组成员的树。使用共享分布树时，组播源需要首先把组播分组发送给集合点路由器，再由这个路由器转发给其他的组成员。

组播路由协议

- 组播路由协议的主要任务就是构造组播的分布树，使组播分组能够传送到相应的组播组成员。根据对网络中的组播成员的分布和使用的不同，组播路由协议分为两类：密集模式路由协议（DM）和稀疏模式路由协议（SM）。
- DM路由协议通常用于组播成员较为集中、数量较多—网络的大部分用户、并且有足够带宽的网路环境，比如公司或园区的局域网。因此，DM路由协议用定期广播组播报文的方法维护组播分布树。DM协议只使用源分布树（SPT），组播流量被广播到网络中所有的组播路由器。
- DM路由协议有：

组播路由协议

- DVMRP：距离向量组播路由协议。这是一种基于距离向量算法的组播路由协议。目前已基本上被PIM和MOSPF所取代。
- MOSPF：组播OSPF协议。
- PIM-DM：协议无关组播协议—密集模式。它不需要单独的组播协议，利用路由器上单播路由协议的路由表作反向路径转发检查，由此获得组播分布树。相比另两种协议，PIM-DM的开销要小很多，它用于组播源和目的非常靠近、接收者数量大于发送者数量并且组播流量比较大的环境中效果很好。

组播路由协议

- 在网路中稀疏分布、网络也没有充足带宽的情况，如广域网环境，可以使用SM路由协议。因此，SM路由协议采用选择性的建立和维护分布树的方式，由空树开始，仅当成员显式的请求加入分布树才做出修改。SM路由协议有：
- CBT：基于中心的分布树协议（RFC 2201）。协议由以一个中心的路由器为根构造一个共享分布树，所有的组播流量都经由这个中心路由器转发。
- PIM-SM：协议无关组播协议—稀疏模式。工作原理与PIM—DM类似，但专门针对稀疏环境优化。适用于组播组中接收者较少、间歇性组播流量的情况。不同于PIM-DM的广播方式，PIM-SM定义了一个集合点(RP)，所有的接收者在RP注册，组播分组由RP转发给接收者。

VPN

Virtual Private Networks

虚拟专用网

虚拟专用网

- 定义
  - 虚拟专业网是一种在公用因特网内构建的专用网络
- 目标
  - 用共享公共的网络基础设施来连接专用网络
- 举例
  - 连接两个商业场所
  - 可以使人们在家工作的时候能够接入公司的网络

如何实现?

- IP 封装和隧道传输
- 在隧道的一端，专用网的IP包被放置在一个新的IP包的数据域中（可以被加密），然后新的IP包被发送到隧道的另一端

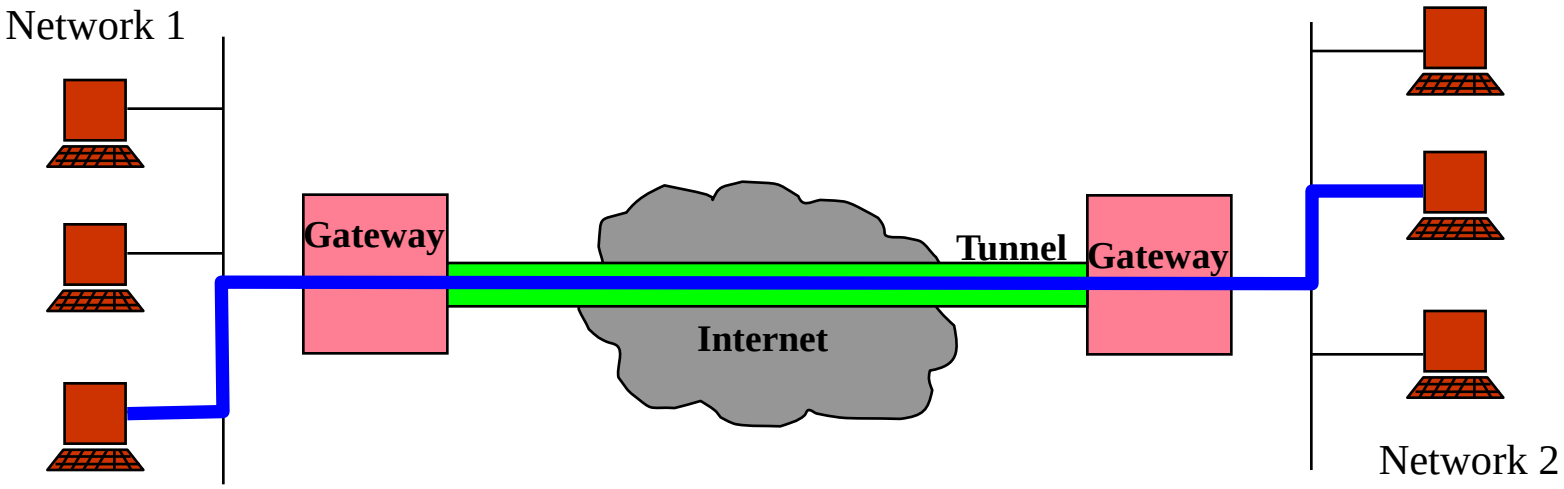
Motivations(动机)

- 经济性
  - 使用共享的基础设施可以节省网络成本
  - 可以减少对租用线路连接的需求
- 通信保密
  - 如果需要，通信可以加密
  - 确保第三方不能使用虚拟网络
- 虚拟的设备位置
  - 在同一网络中的主机并不需要在同一地点
  - 可以在分离的物理网络间构建逻辑网络
- 支持专用网络的特性
  - 组播，IPX或者Apple talk协议，等等

举例

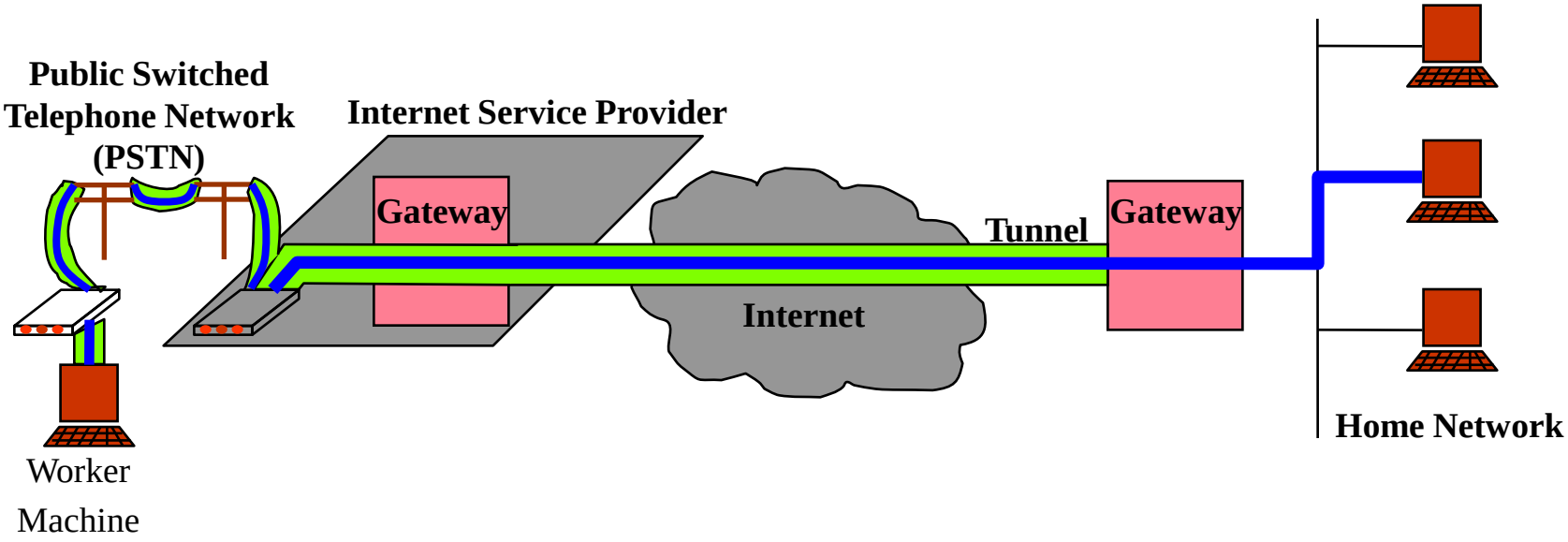
- 逻辑网络构建
- 虚拟拨号

举例：逻辑网络构建



- 远程网络1和网络2构建一个逻辑网络
- 在底层保证通信的安全性

举例：虚拟拨号



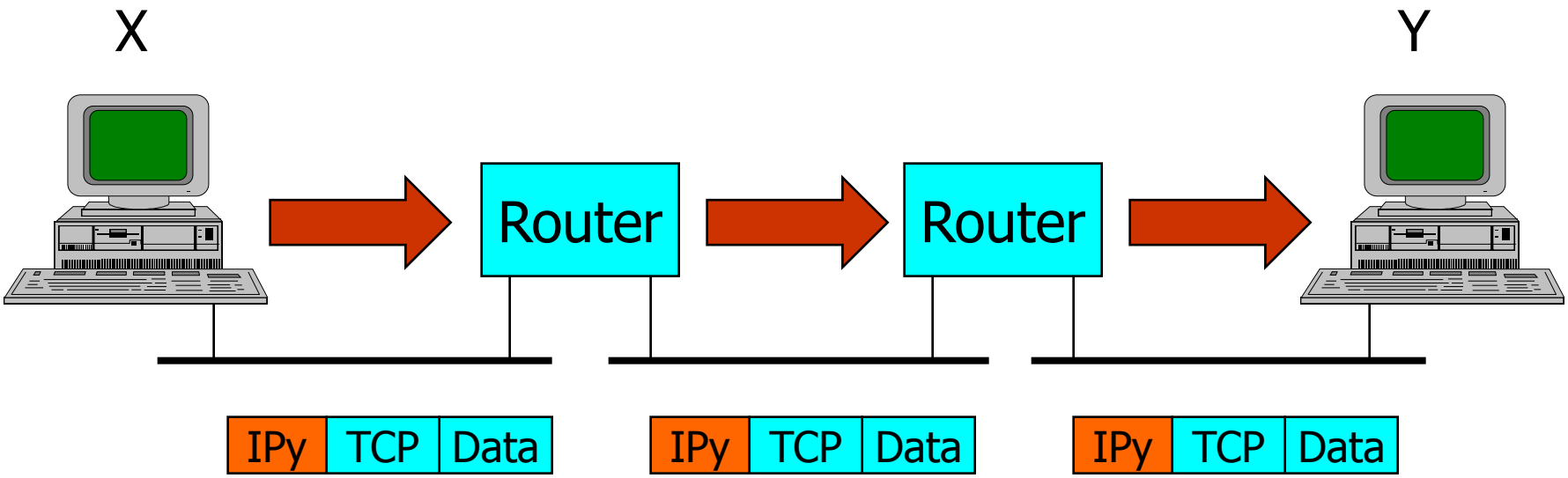
- 工作人员拨号ISP，以获取基本的IP服务
- 工作人员创建到Home网络的隧道

IP 隧道

- 一个通常的网络层包有如下的形式:
  - (IP header (TCP header (data)))
- 隧道传输是将以上的包封装在另外一个IP头部:
  - (IP header (IP header (TCP header (data))))
- 利用隧道技术可以在一对路由器间创建虚拟链路
- 对于网络中的其他部分，将经过隧道传输技术处理后的包被看做一个普通的IP包
- 例子：移动IP使用隧道传输技术来将包传输到正确的目的地。

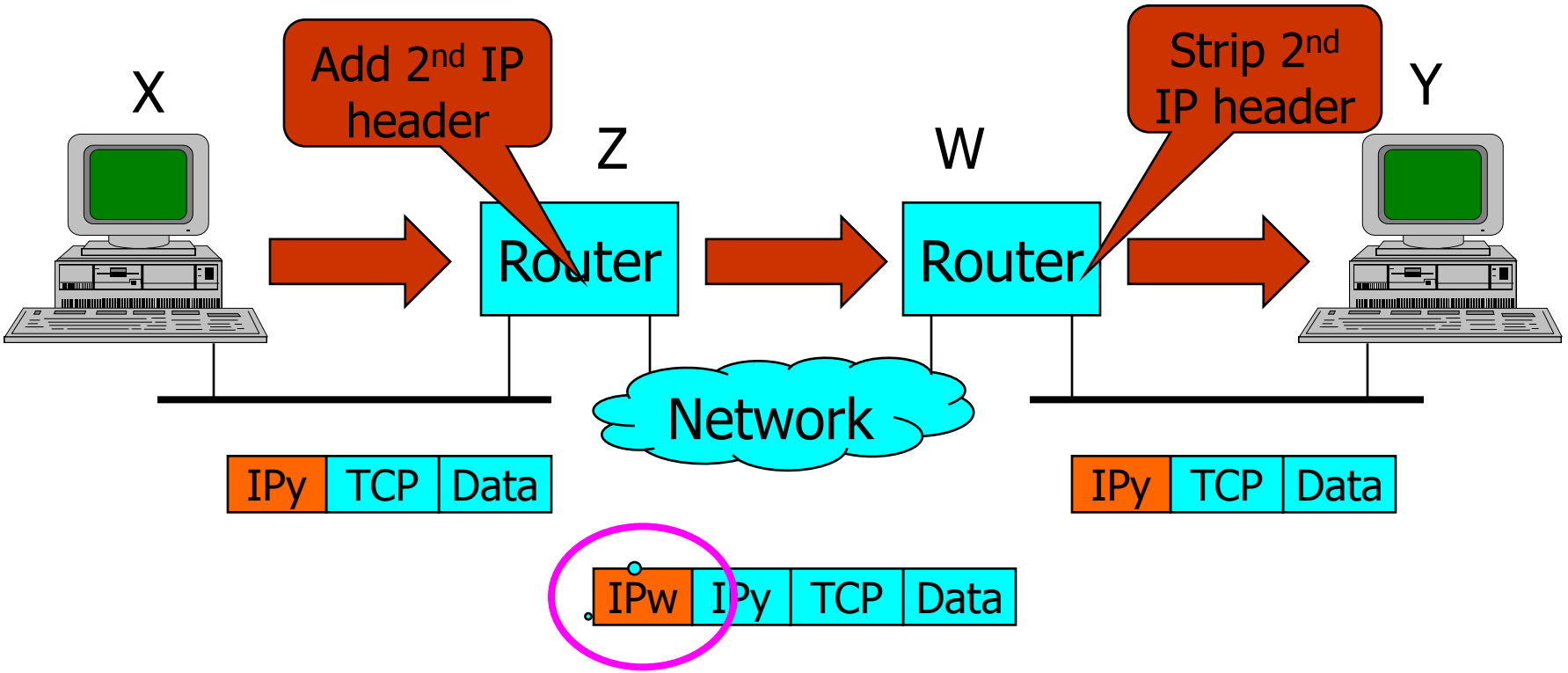
正常IP包传送

- 每个路由器以如下的方式传递IP包



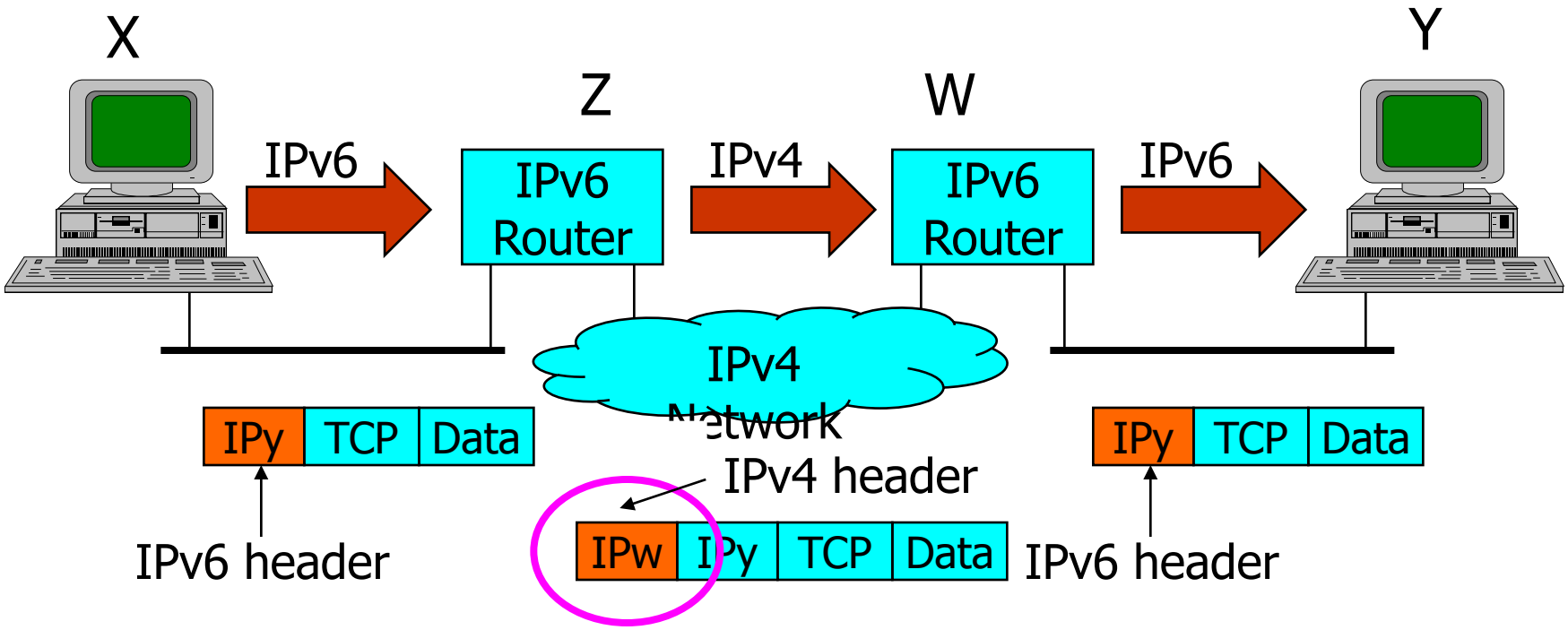
隧道传输（Tunneling）

- 每个路由器以如下的方式传递IP包



为什么使用隧道传输？

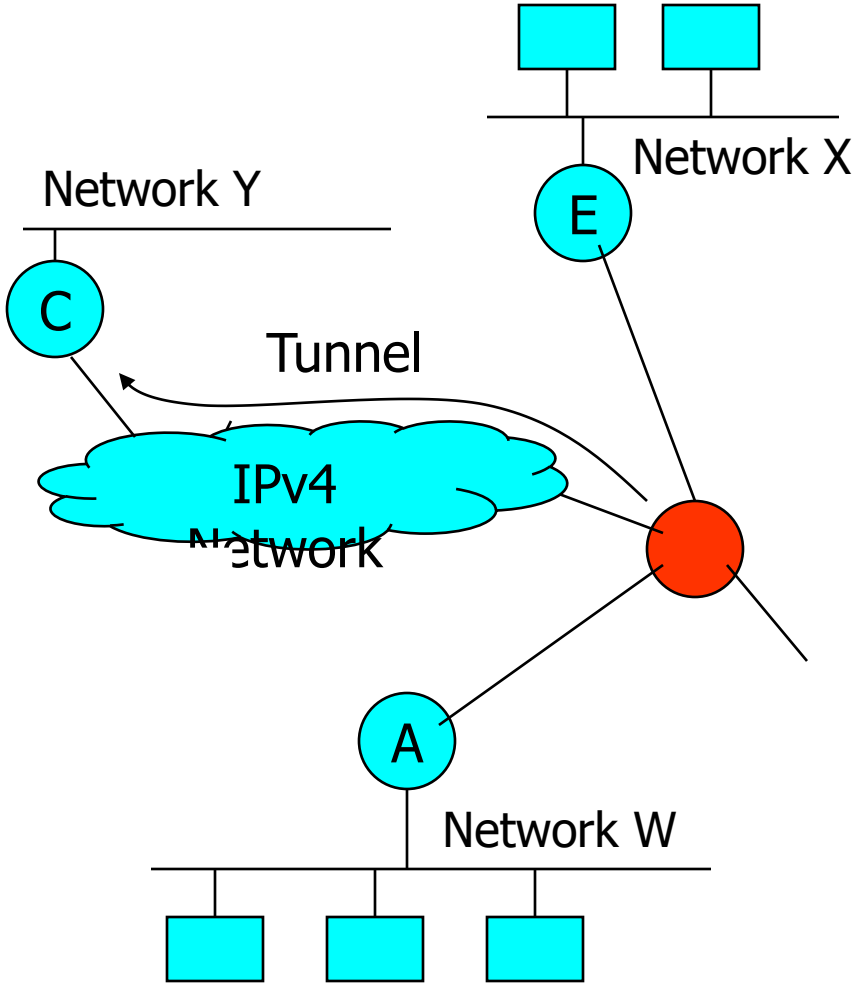
- 虚拟专用网络（安全性）
- 在一个IPV4网络中连接IPv6（或者其他异构的路由器）
  - 例子:

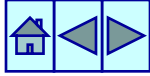


IP 路由表

- 支持隧道传输

Destination	Next hop (port)
Network X	E
Network Y	Encapsulate in C
Network W	A
……	……





Mobile IP

移动IP

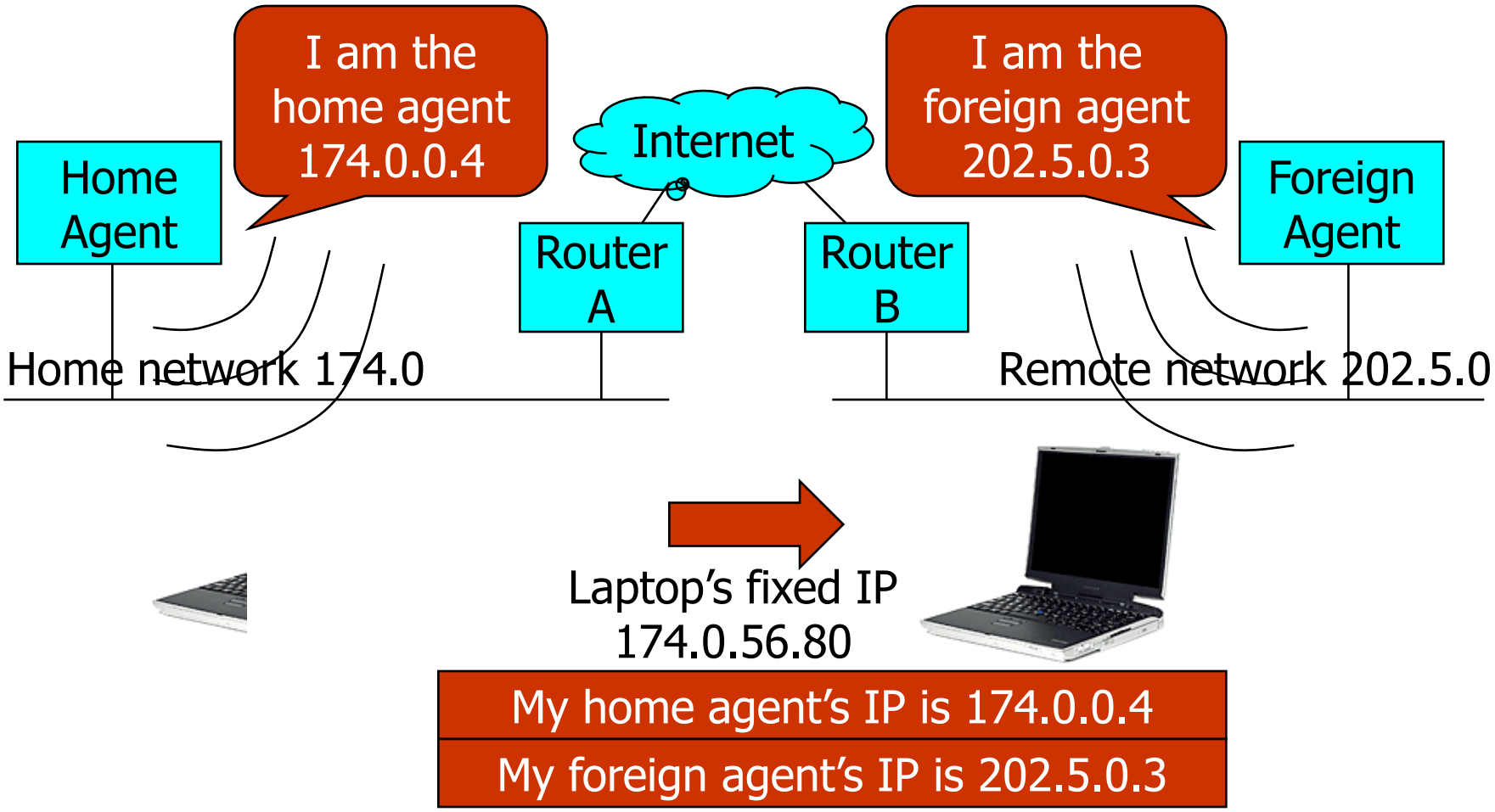
移动IP

- IP地址编码了网络位置
- 改变位置→不同的网络→不同的IP地址
- 如何支持主机移动性
  - 简单的解决方法：当一个漫游的主机（比如你的笔记本电脑）加入到一个新的网络中时使用DHCP来动态分配IP地址
  - 问题?

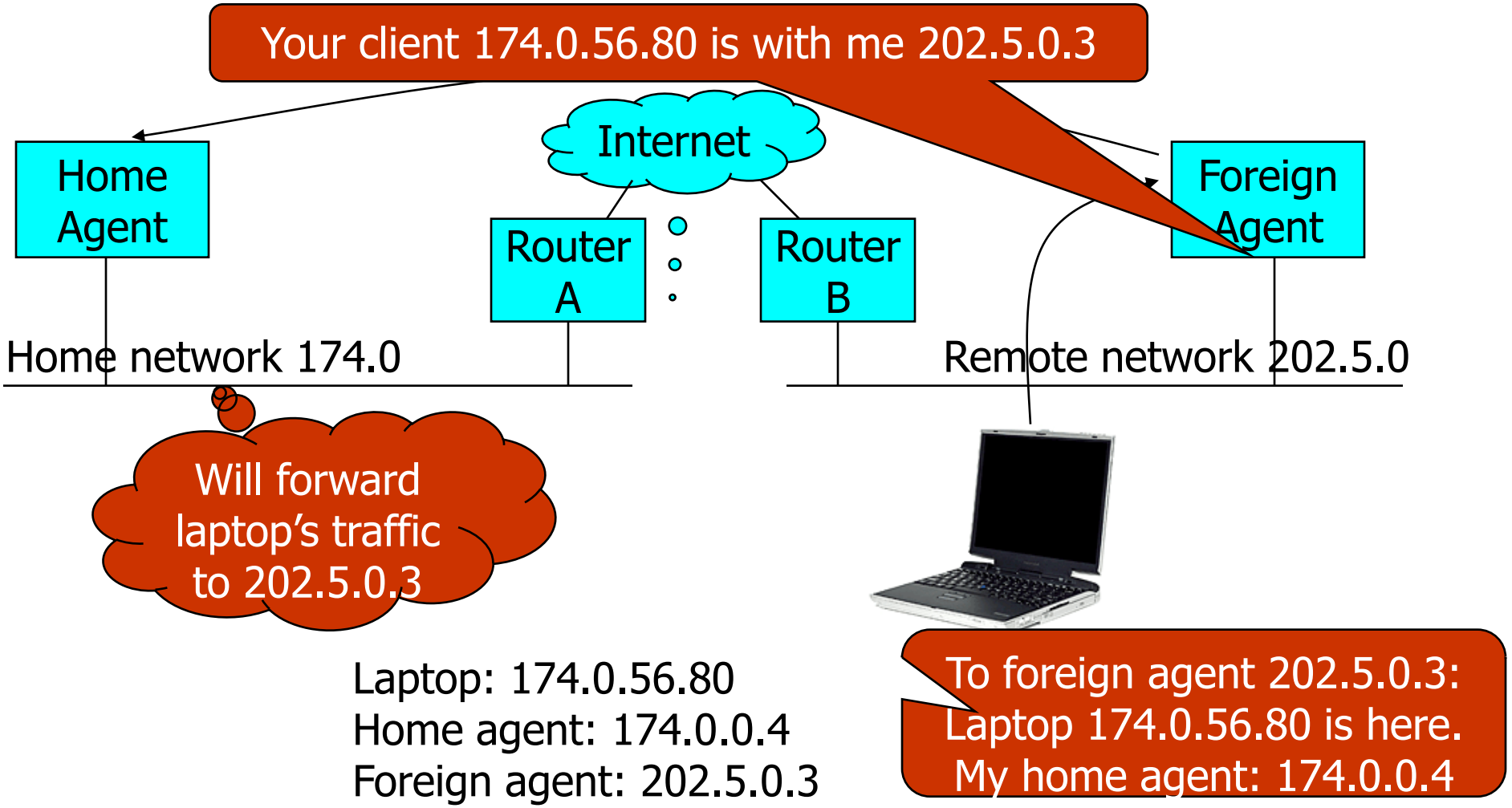
移动IP的目标：通过固定的IP实现可移动性

- 考虑一个漫游的用户。这个用户的笔记本电脑可能需要从网络A分离，连接到网络B，而不需要改变其IP地址！
- 透明性:正在运行的应用程序不必关心笔记本电脑改变了网络

移动IP:本地代理和远程代理



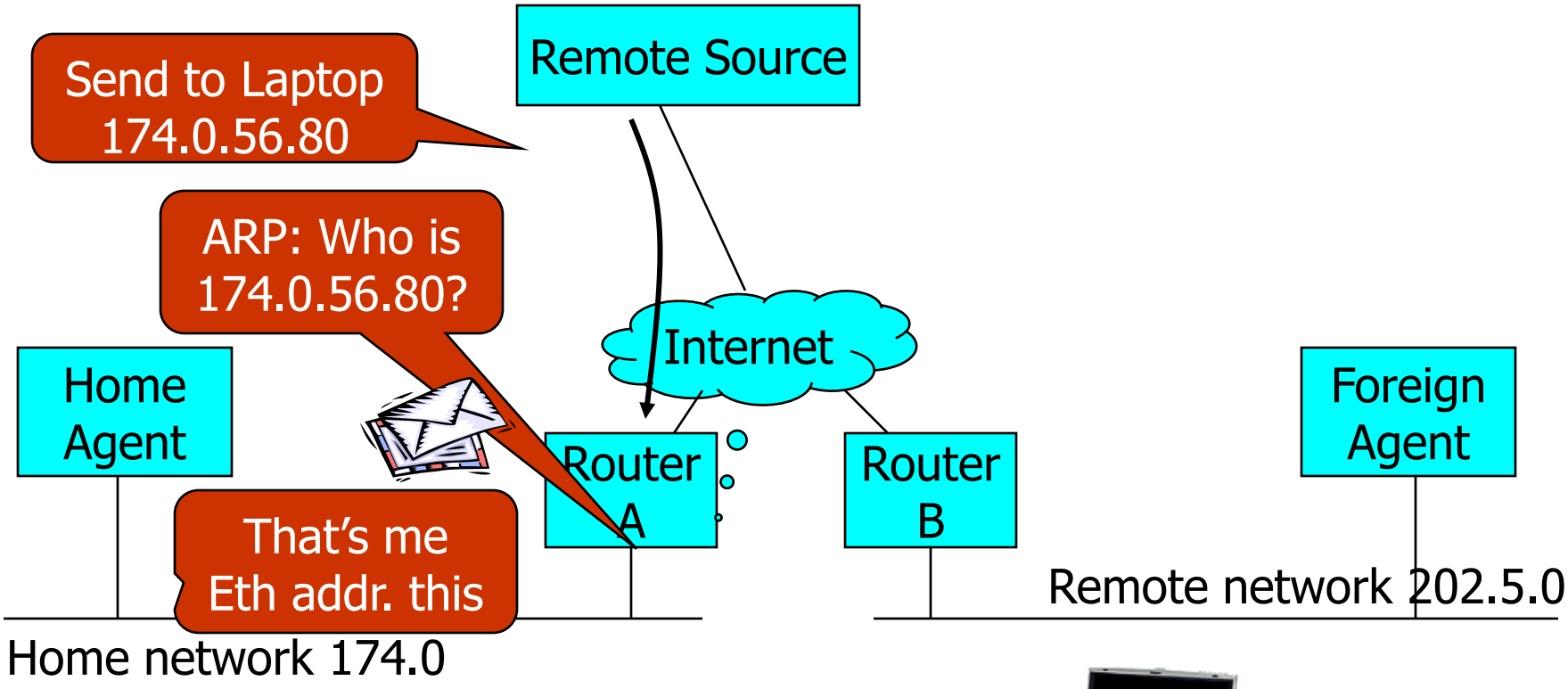
移动IP:Home and Foreign Agent



路由到移动主机

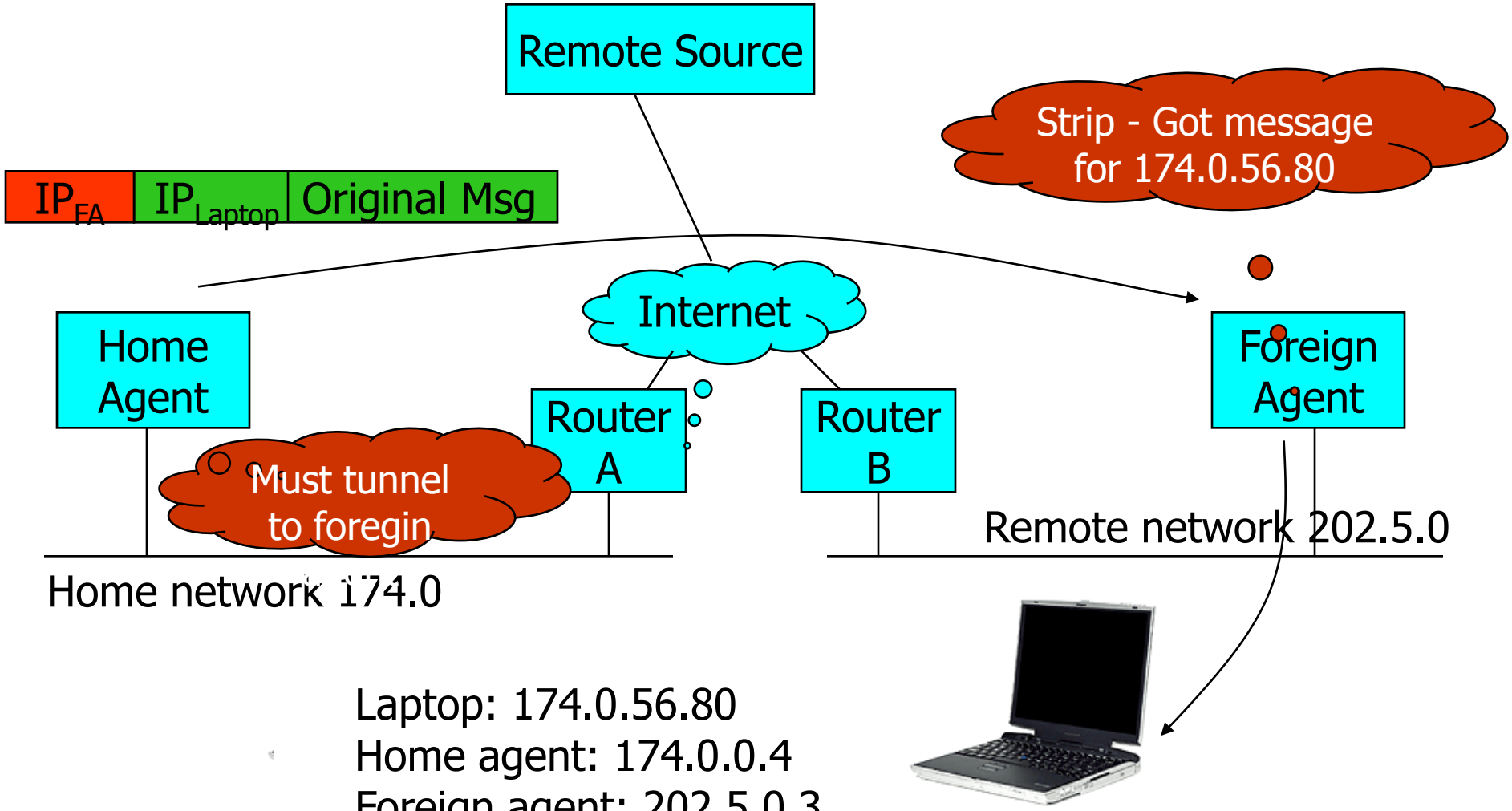
- 本地代理扮演了移动主机的角色：
  - 到达的分组中含有目的地址的移动IP
  - 发出ARP 请求
  - 本地代理答复它的硬件地址
  - 分组被送到本地代理
- 本地代理把远程代理的地址封装到分组IP头中
  - 分组被远程代理接收
- 远程代理去掉IP头部，找出与移动主机相关的分组。

路由到移动主机



Laptop: 174.0.56.80
Home agent: 174.0.0.4
Foreign agent: 202.5.0.3

路由到移动主机


Remote Source
Strip - Got message for 174.0.56.80
IP FA
IP Laptop
Original Msg
Internet
Home Agent
Foreign Agent
Router A
Router B
Must tunnel to foregin
Remote network 202.5.0
Home network 174.0
Laptop: 174.0.56.80
Home agent: 174.0.0.4
Foreign agent: 202.5.0.3

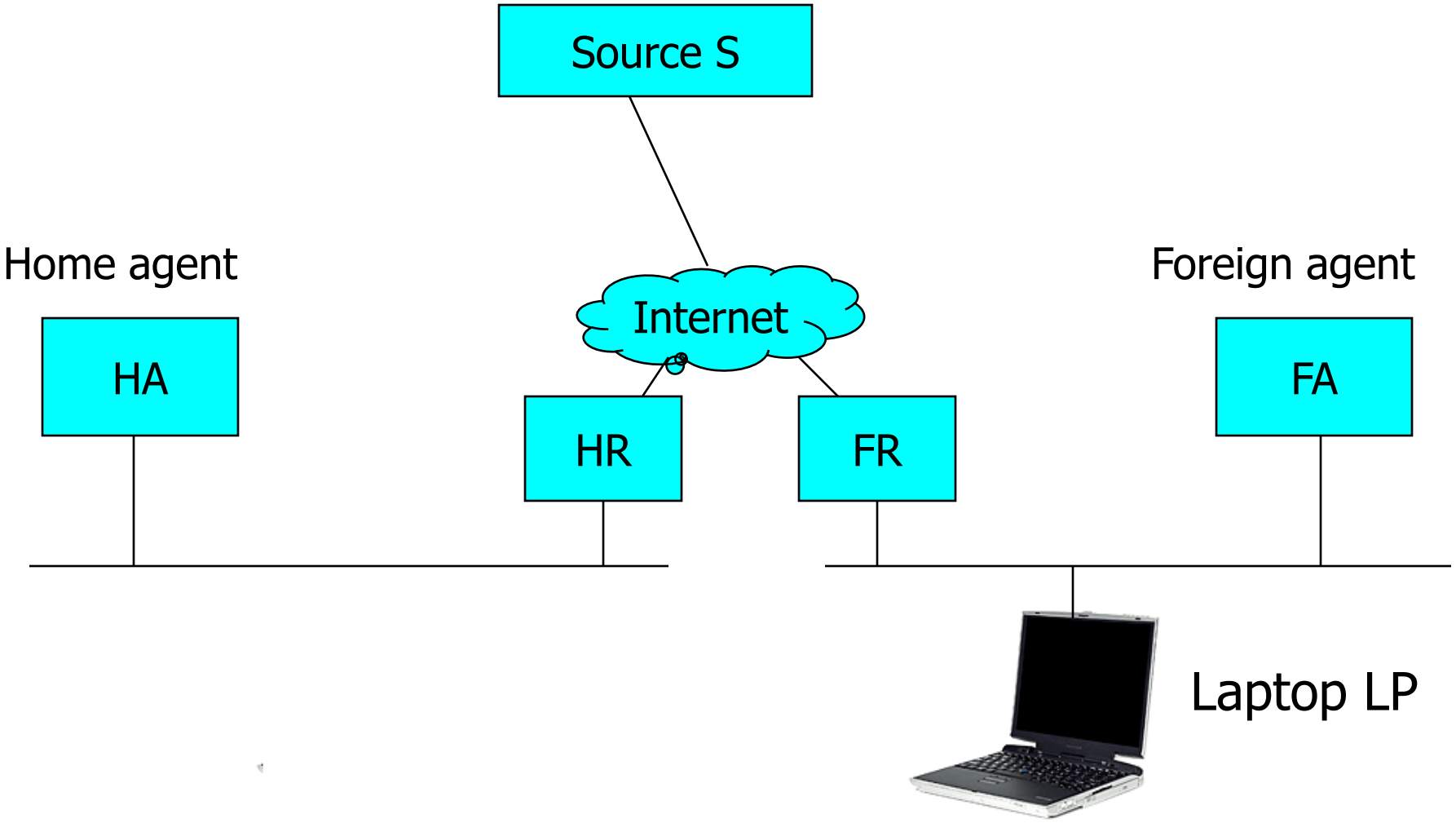
练 习

分组头部中含有什么？

假设每个主机X含有IP地址IP_X和以太网地址 Eth_X 。同时考虑源端S、笔记本LP、本地代理HA、远程代理FA、本地路由器HR及远程路由器FR。

Step #	Source		Destination	
	IP	Eth	IP	Eth

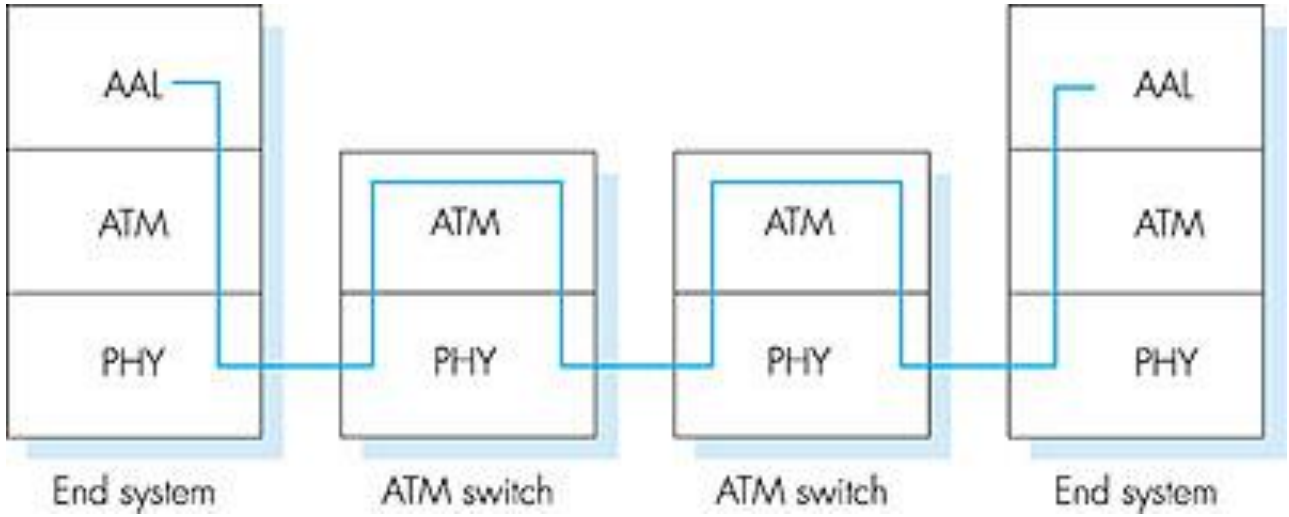
练习


Source S
Home agent
Internet
Foreign agent
HA
HR
FR
FA
Laptop LP

其它广域网技术上的IP

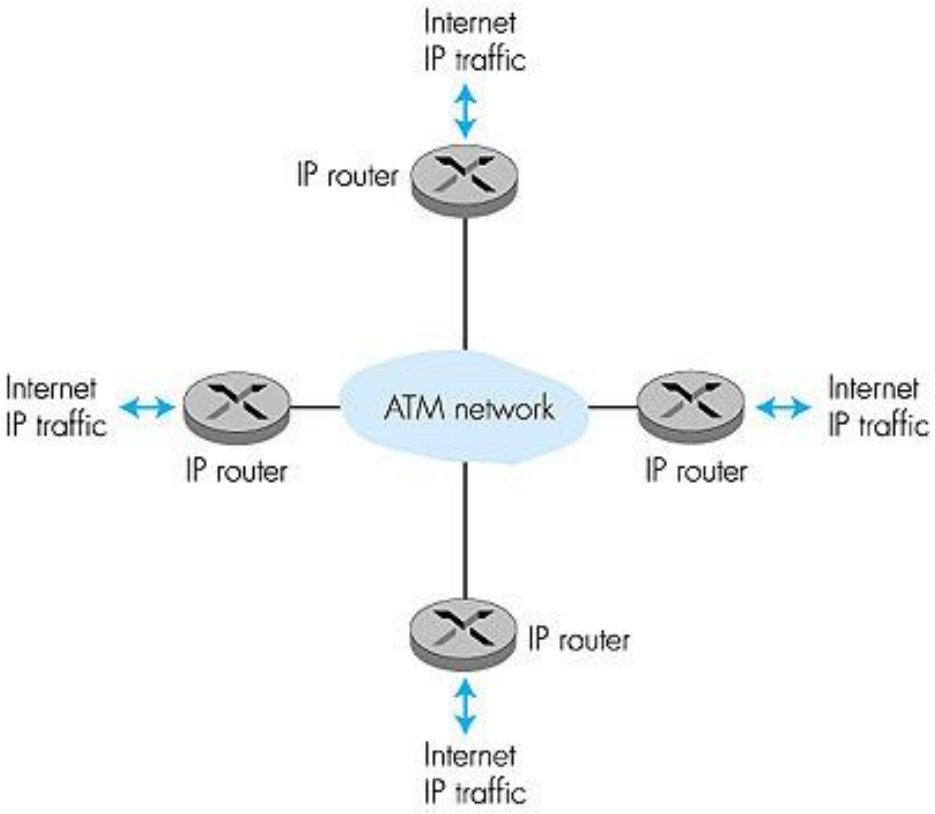
- ATM
- 帧中继
- X.25

ATM 体系结构



- 适配层 (AAL): 仅在ATM网络的边缘
  - 数据分段/重组
  - 与Internet传输层相似
- ATM层: “网络层”
  - 虚电路、路由、信元交换
- 物理层

ATM：网络层或链路层？



表象：端到端的传输：从一台主机到另一台主机。

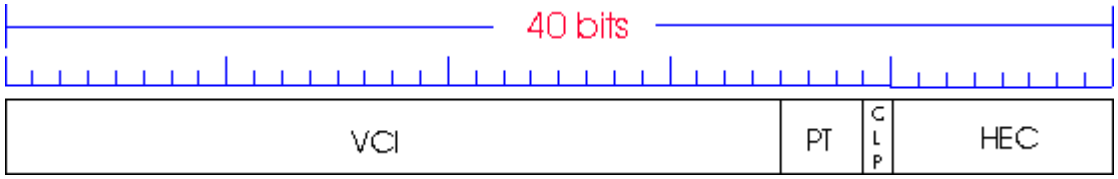
- ATM 是一种网络技术

事实：用于连接IP主干路由器 。

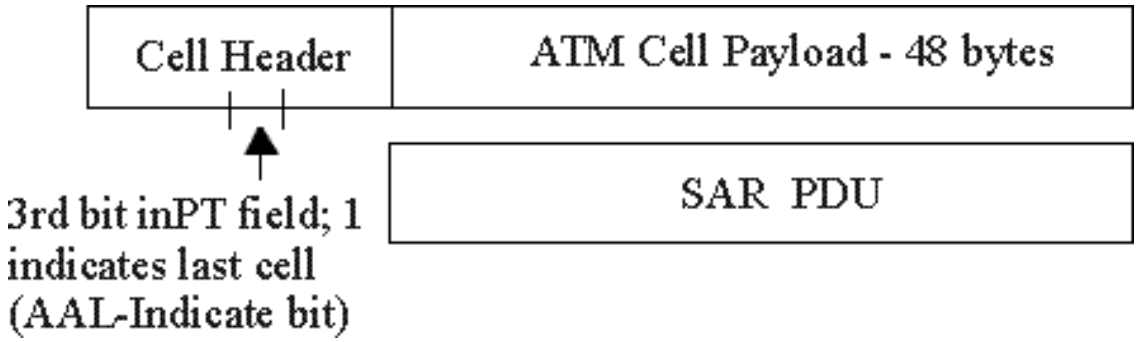
- “IP over ATM"
- 作为交换的链路层，ATM用于连接IP路由器。

ATM 层: ATM 信元

Cell header

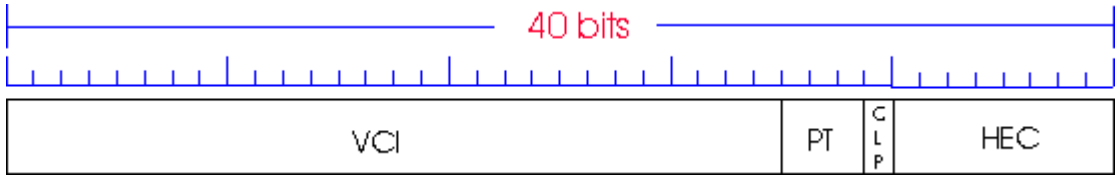


Cell format



- 5字节的ATM信元头部
- 48字节的有效荷载
  - 为什么?: small payload 小的有效荷载->对于数字化的话音，有较小的信元产生的延时
  - 在32和64之间进行折衷

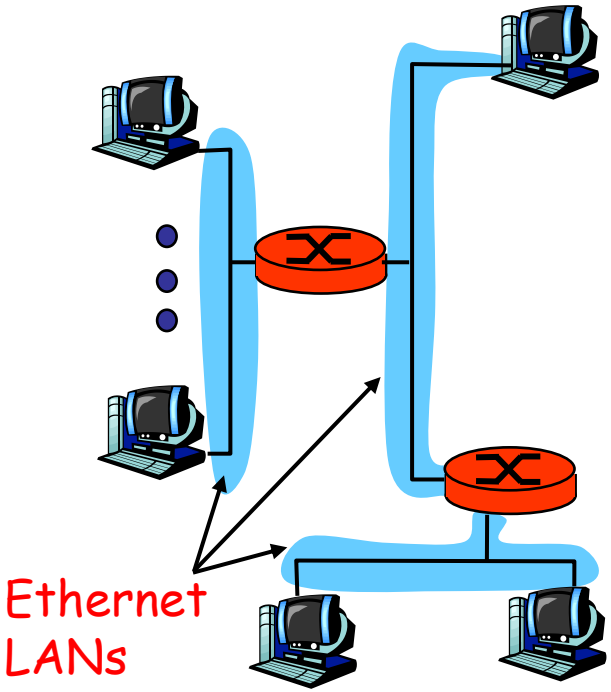
ATM 信元头部



- **VCI:** 虚通道 ID
  - 经过网络中的链路时,虚通道ID会改变
- **PT:** 负荷类型 (如 RM 信元、数据信元)
- **CLP:** 信元丢失优先级位
  - CLP = 1 表示信元优先级低，将在拥塞时丢弃该信元。
- **HEC:** 头部差错校验和
  - 循环冗余校验

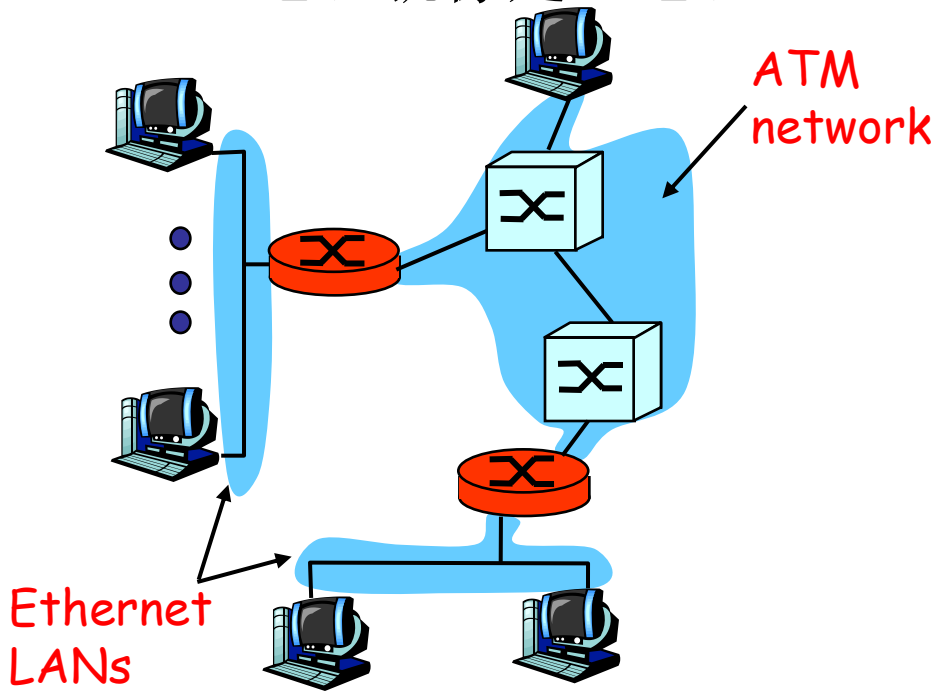
典型的IP

- 3个LANS
- MAC (802.3) 和IP地址

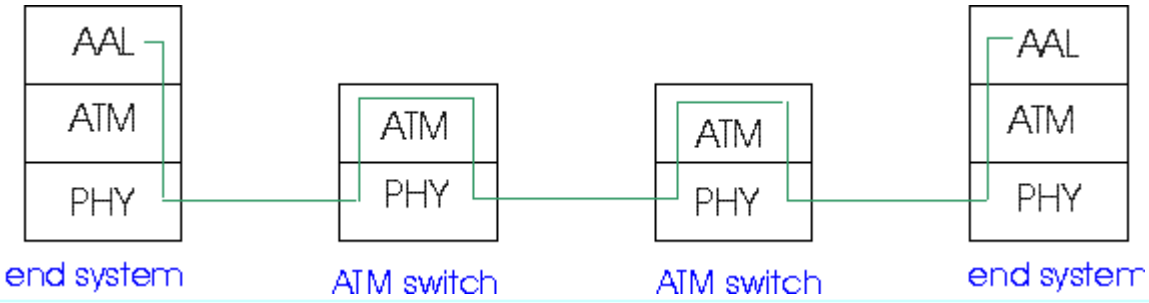


IP over ATM

- 用ATM替代一个LAN
- IP 地址 -> 对于MAC(802.3) 地址，ATM 地址就像是IP地址



IP-over-ATM网中的数据报



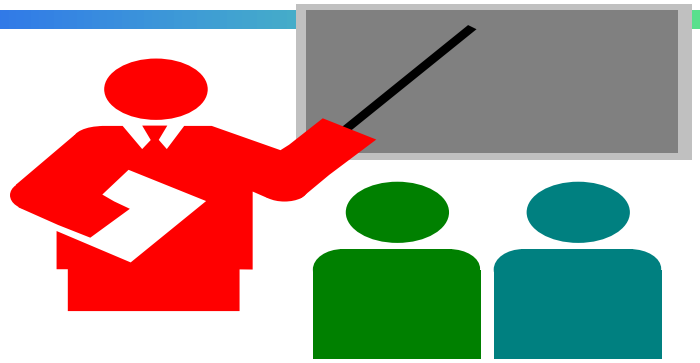
- 源主机:
  - IP层查找IP地址与ATM地址的映射表 (利用ARP)
  - 把数据报送到 AAL5
  - AAL5 把数据和分段封装到信元中，把它送到ATM层 。
- ATM 网: 沿着虚电路把信元传送到目的地（利用现有的VC或新建一个VC）。
- 目的主机:
  - AAL5 把信元重装为原来的数据报
  - 如果通过循环冗余校验, 数据报将被送到IP。

帧中继 -VC 速率控制

- 确保信息速率 (CIR)
  - 为每一个虚电路（ VC ）保证数据率
  - 在VC建立时协商
  - 用户根据CIR付费

- DE 位: 丢弃条件指示（ Discard Eligibility ）位
  - 边缘的帧中继交换测定了每个VC的通信速率；标记DE位。
  - DE = 0: 高优先级, 以允许数率传送的帧; 不惜代价地进行传送。
  - DE = 1:低优先级,当拥塞发生时应先丢弃

小结



- IP
- ICMP
- VPN
- IP 隧道传输
- 移动IP
- 其它广域网技术中的IP

Thank You !