第 10 章 下一代因特网

第 10 章　下一代因特网（续）

10.1 下一代的网际协议 IPv6 (IPng)

10.1.1 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 位的 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。

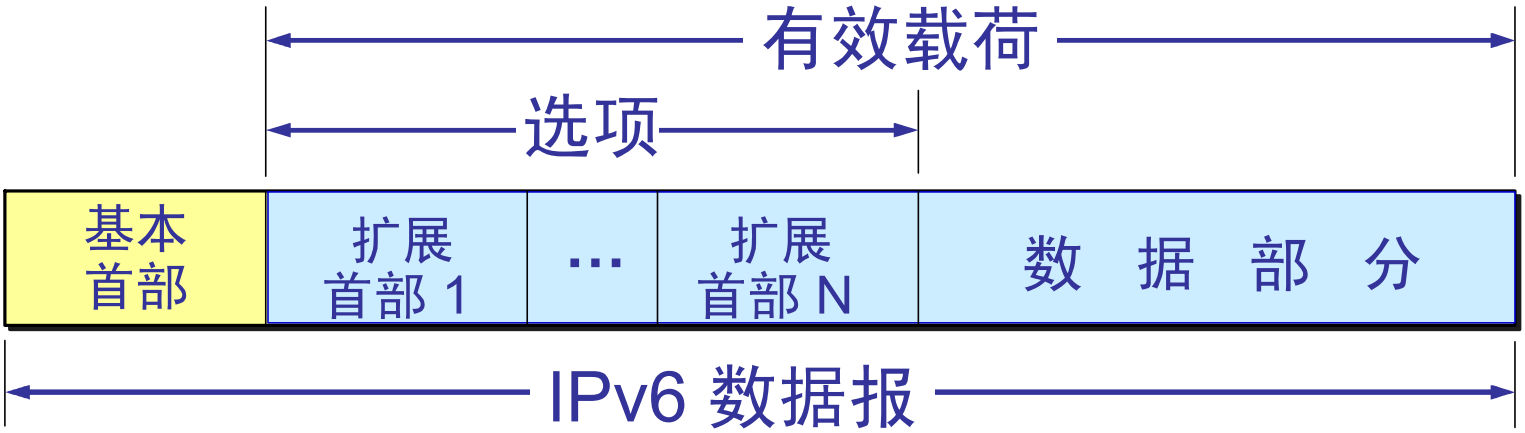
10.1.2 IPv6 的基本首部

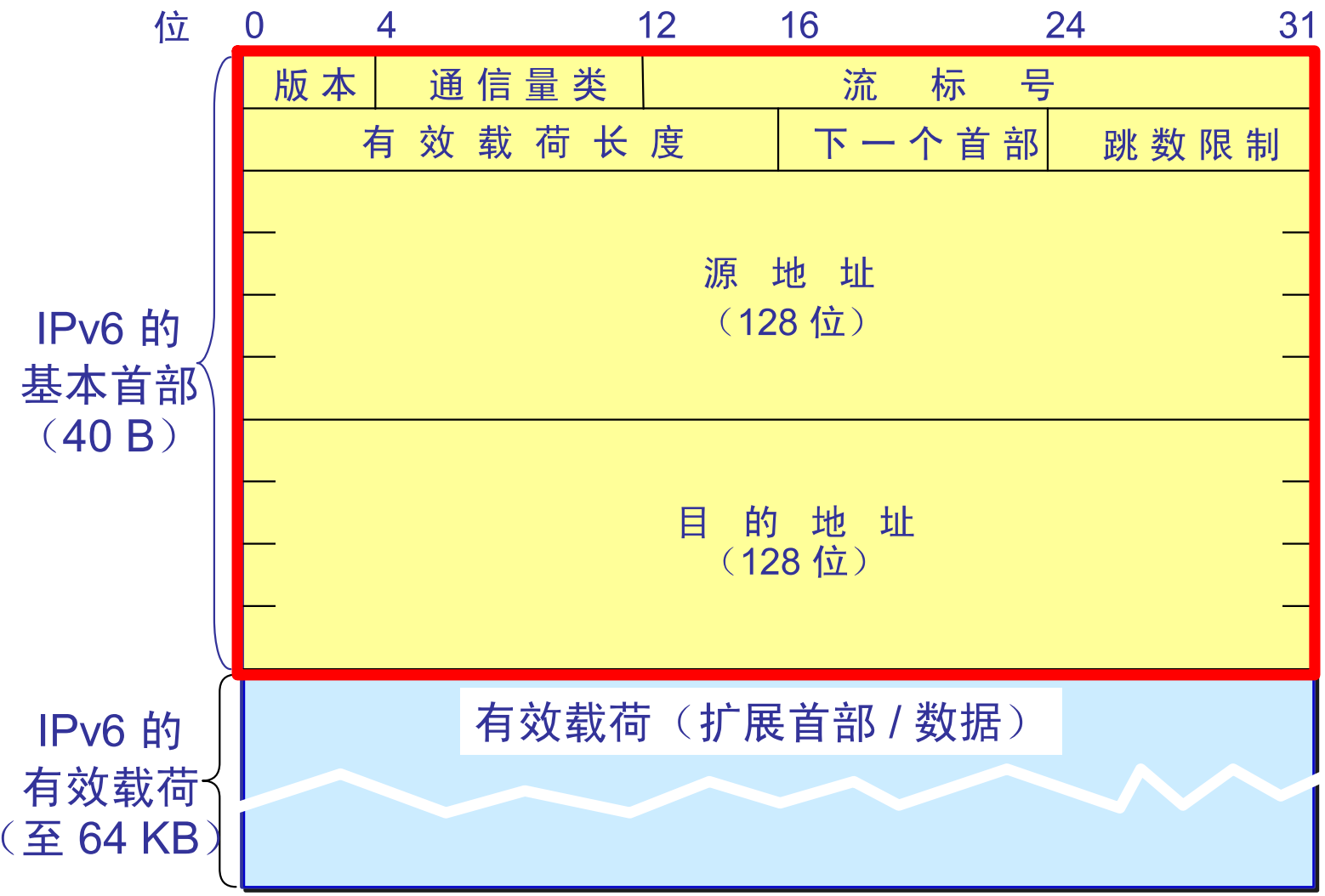
- IPv6 仍支持无连接的传送所引进的主要变化如下
- 更大的地址空间。IPv6 将地址从 IPv4 的 32 位 增大到了 128 位。
- 扩展的地址层次结构。
- 灵活的首部格式。
- 改进的选项。
- 允许协议继续扩充。
- 支持即插即用（即自动配置）
- 支持资源的预分配。

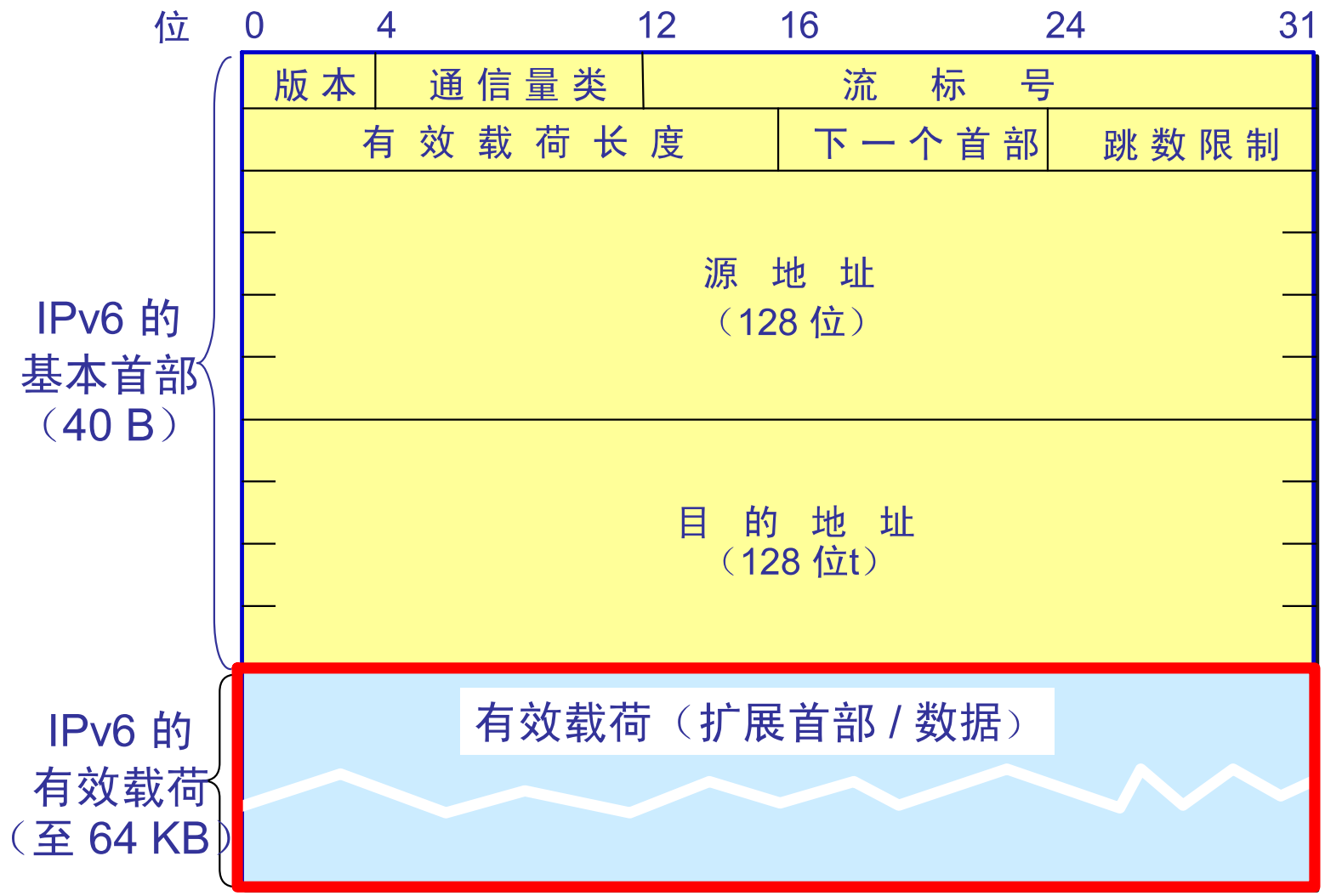
IPv6 数据报的首部

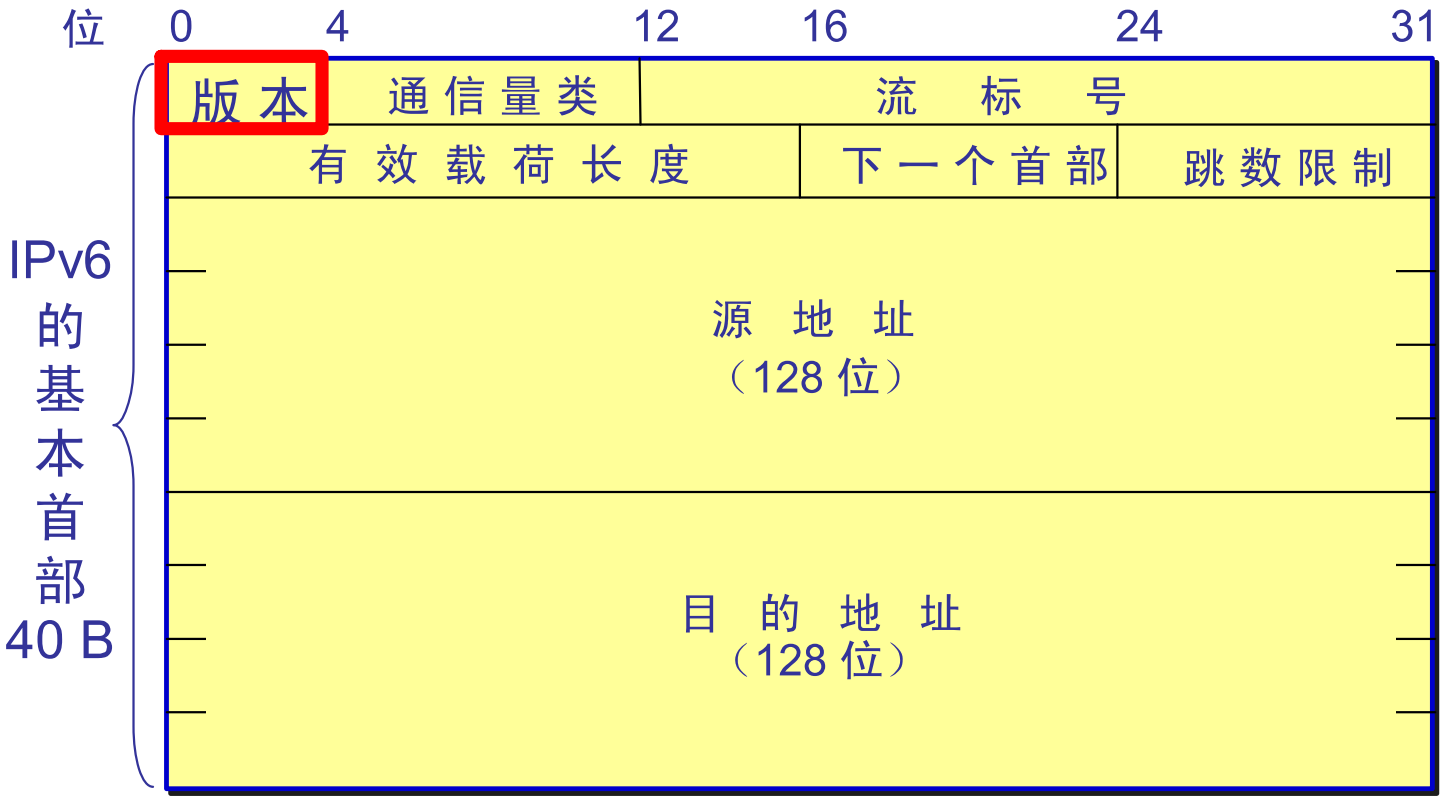
- IPv6 将首部长度变为固定的 40 字节，称为基本首部(base header)。
- 将不必要的功能取消了，首部的字段数减少到只有 8 个。
- 取消了首部的检验和字段，加快了路由器处理数据报的速度。
- 在基本首部的后面允许有零个或多个扩展首部。
- 所有的扩展首部和数据合起来叫做数据报的有效载荷(payload)或净负荷。

IPv6 数据报的一般形式

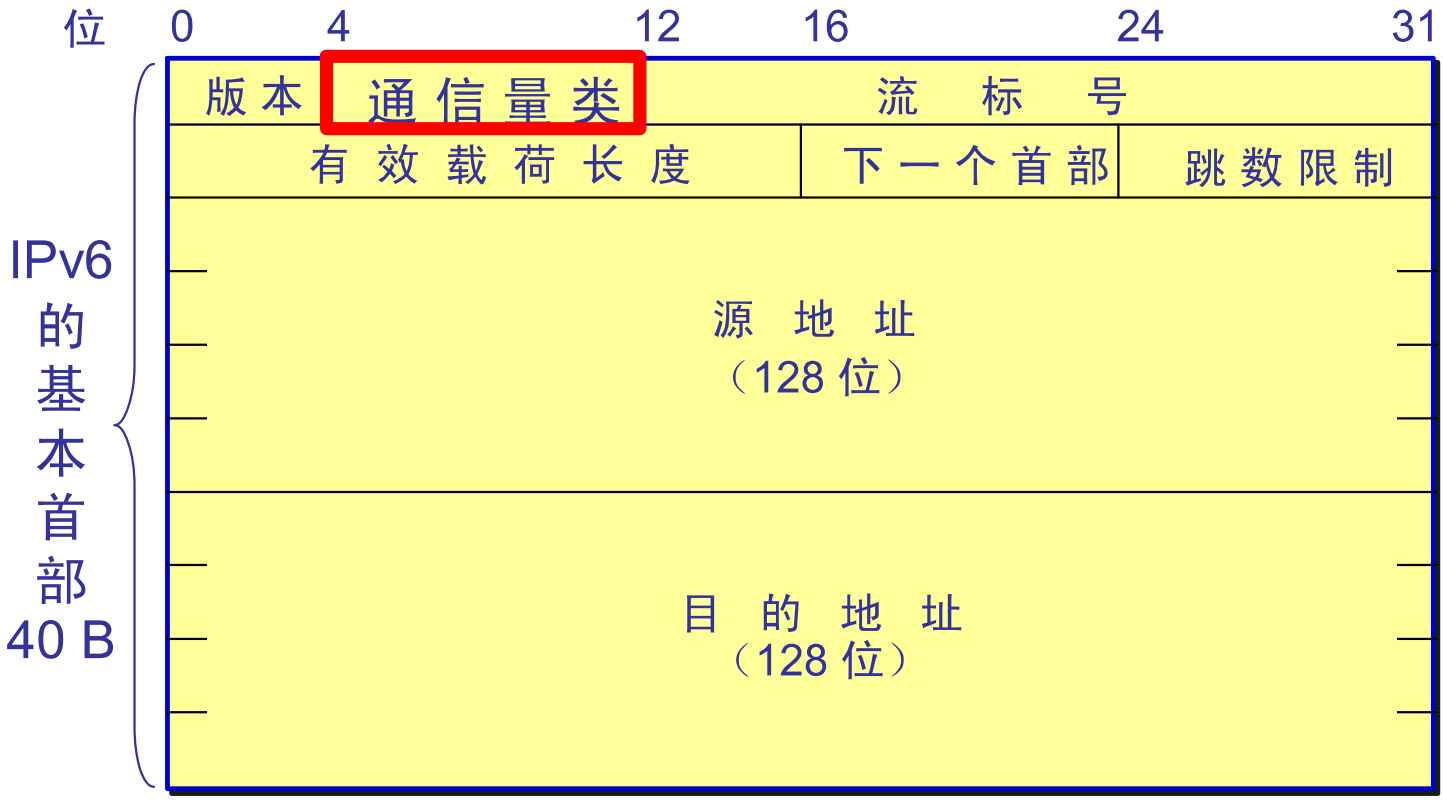



位
0
4
12
16
24
31
版本
通信量类
流 标 号
有效载荷长度
下一个首部
跳数限制
源 地 址
（128 位）
目 的 地 址
（128 位）
IPv6 的
基本首部
（40 B）
有效载荷（扩展首部 / 数据）
IPv6 的
有效载荷
（至 64 KB）

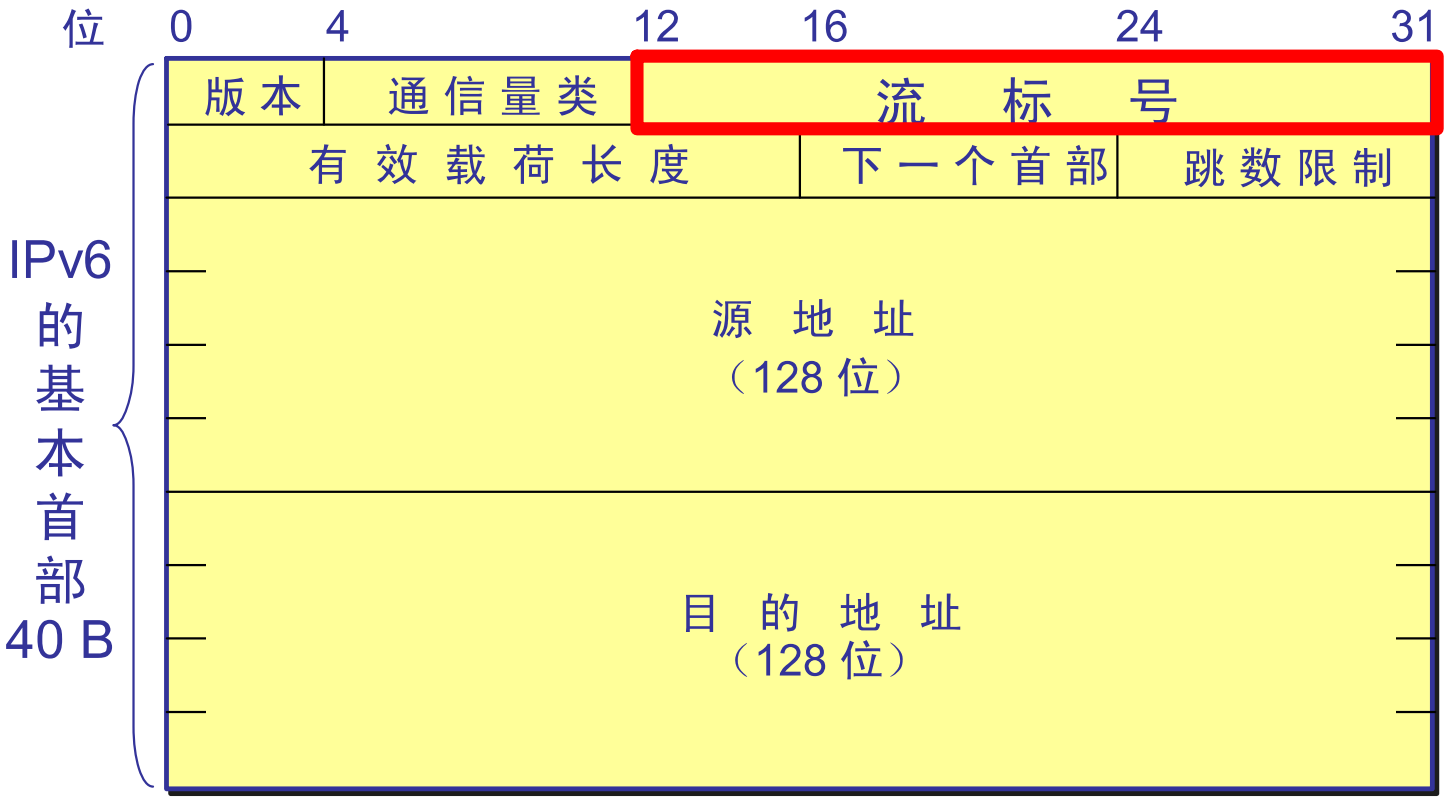

位
0
4
12
16
24
31
版本
通信量类
流 标 号
有效载荷长度
下一个首部
跳数限制
源 地 址
（128 位）
目 的 地 址
（128 位t）
IPv6 的
基本首部
（40 B）
IPv6 的
有效载荷
（至 64 KB）
有效载荷（扩展首部 / 数据）



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

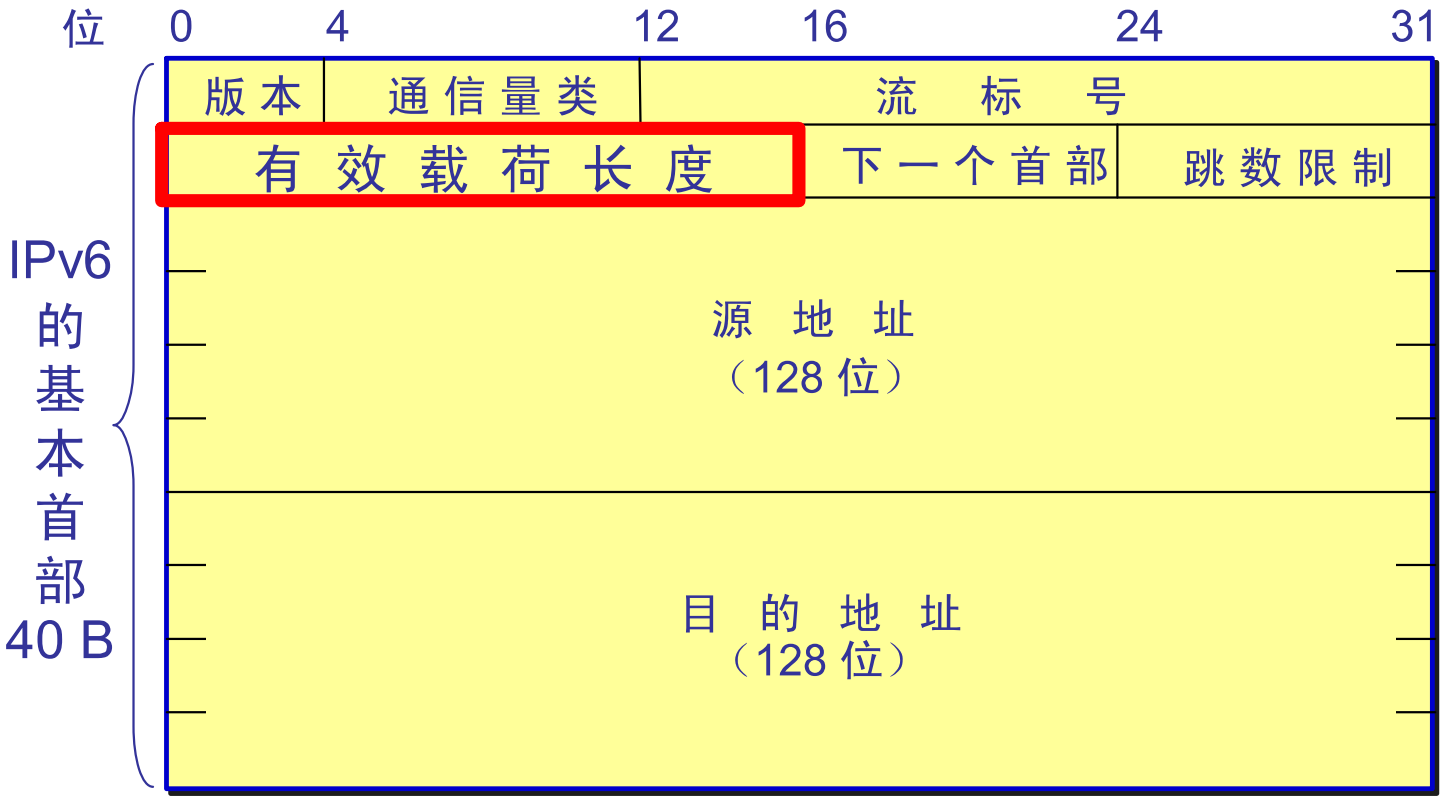


通信量类(traffic class)—— 8 位。这是为了区分不同的 IPv6 数据报的类别或优先级。目前正在进行不同的通信量类性能的实验。

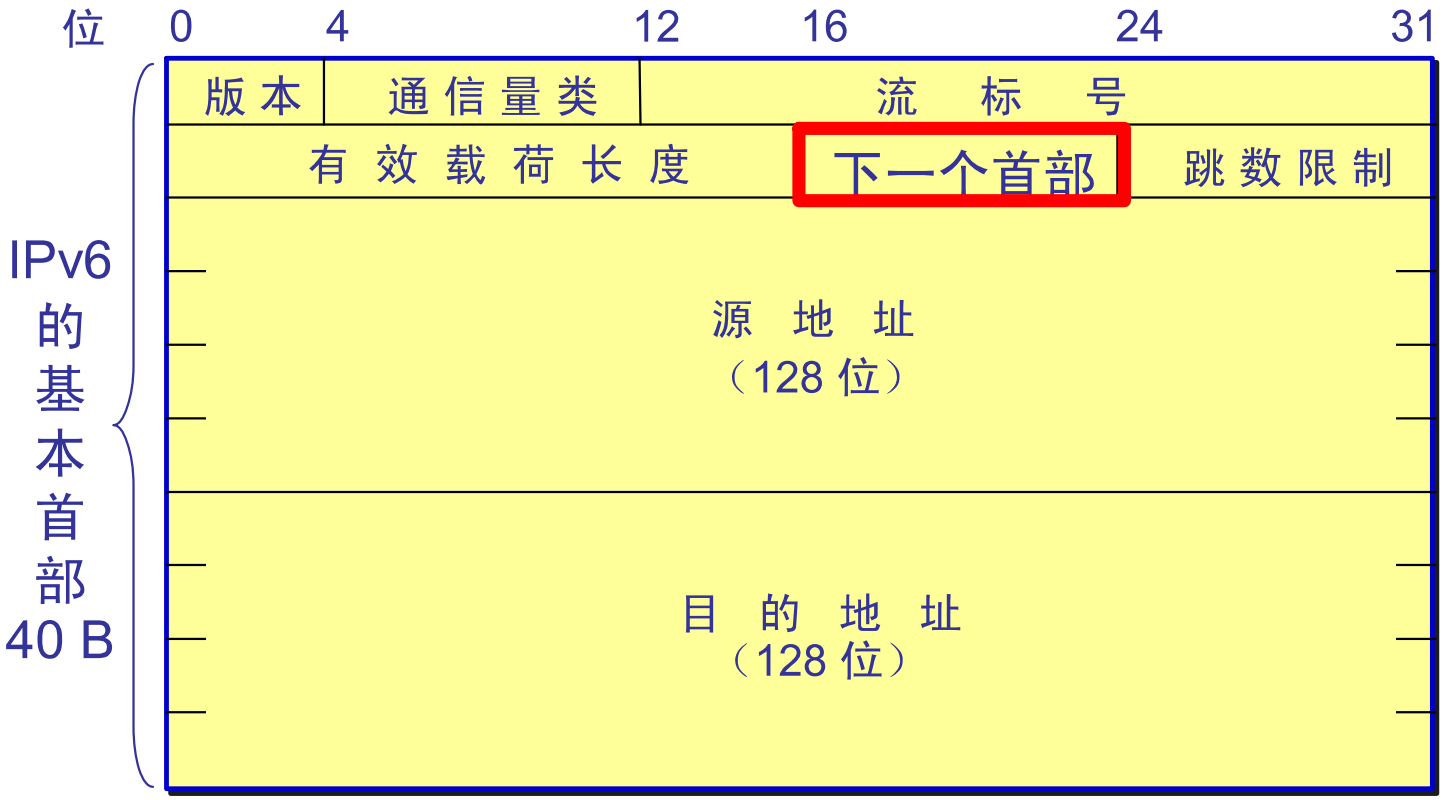


流标号(flow label)—— 20 位。“流”是互联网络上从特定源点到特定终点的一系列数据报，“流”所经过的路径上的路由器都保证指明的服务质量。

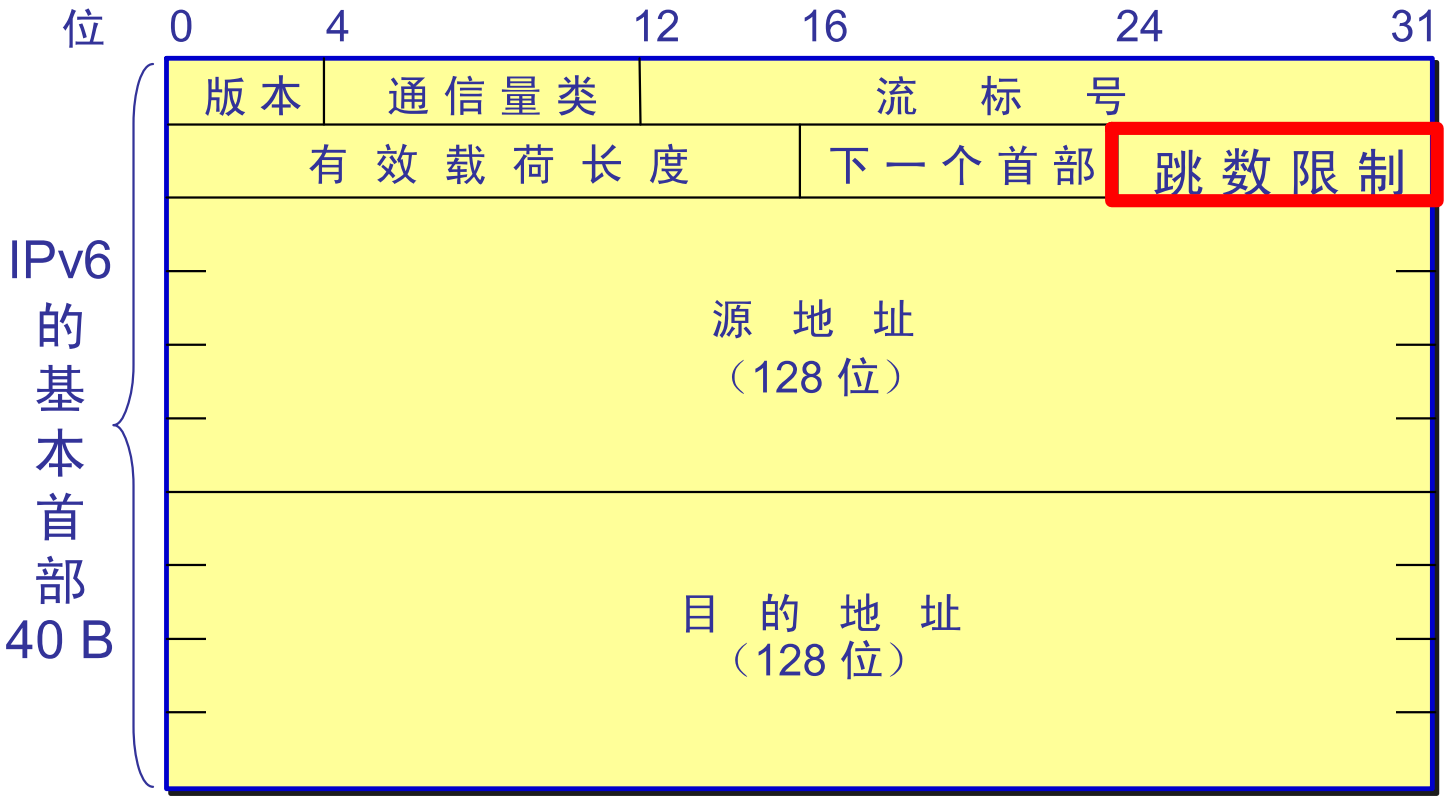
所有属于同一个流的数据报都具有同样的流标号。



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



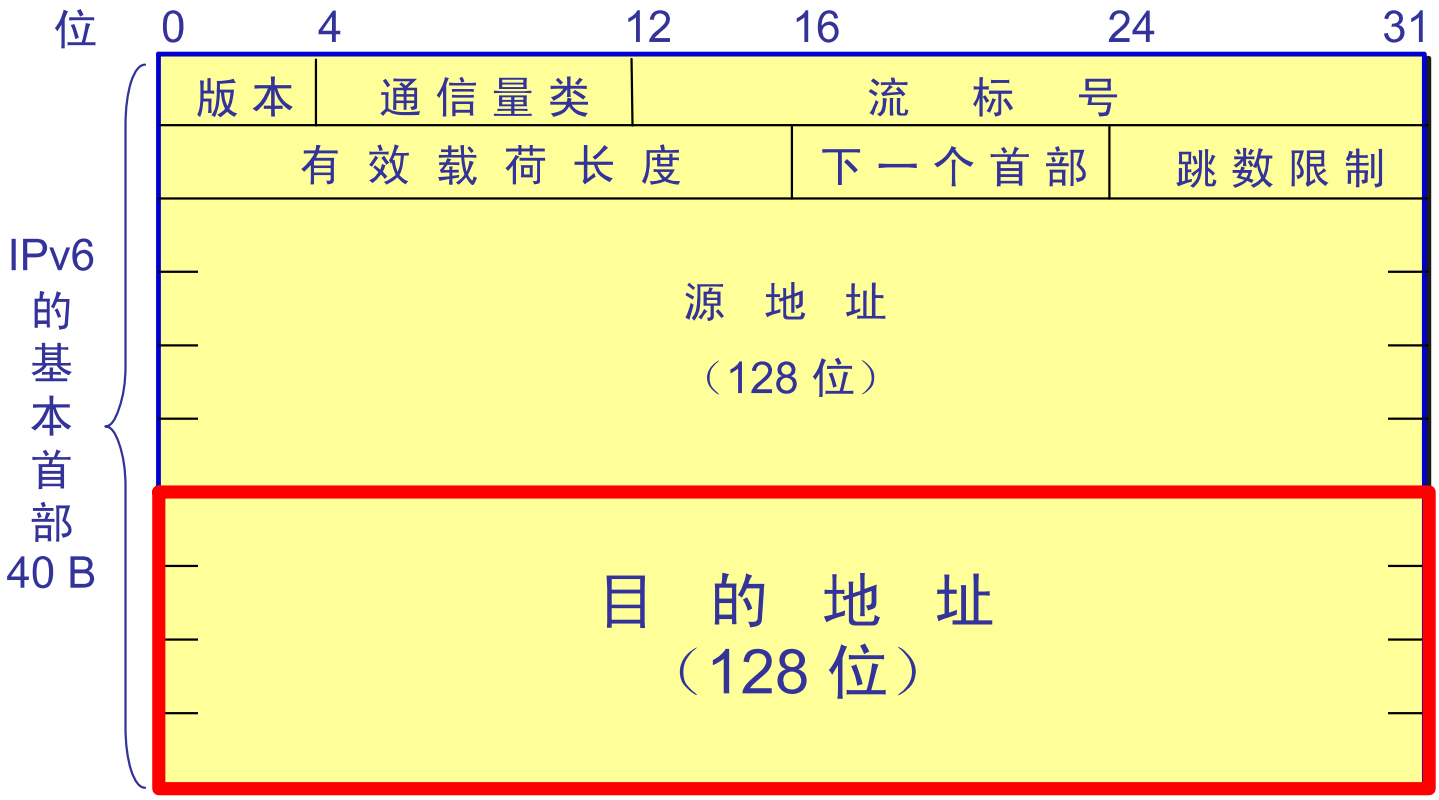
下一个首部(next header)—— 8 位。它相当于 IPv4 的协议字段或可选字段。



跳数限制(hop limit)—— 8 位。源站在数据报发出时即设定跳数限制。路由器在转发数据报时将跳数限制字段中的值减1。

当跳数限制的值为零时，就要将此数据报丢弃。

源地址—— 128 位。是数据报的发送站的 IP 地址。



目的地址—— 128 位。是数据报的接收站的 IP 地址。

10.1.3 IPv6 的扩展首部

1. 扩展首部及下一个首部字段

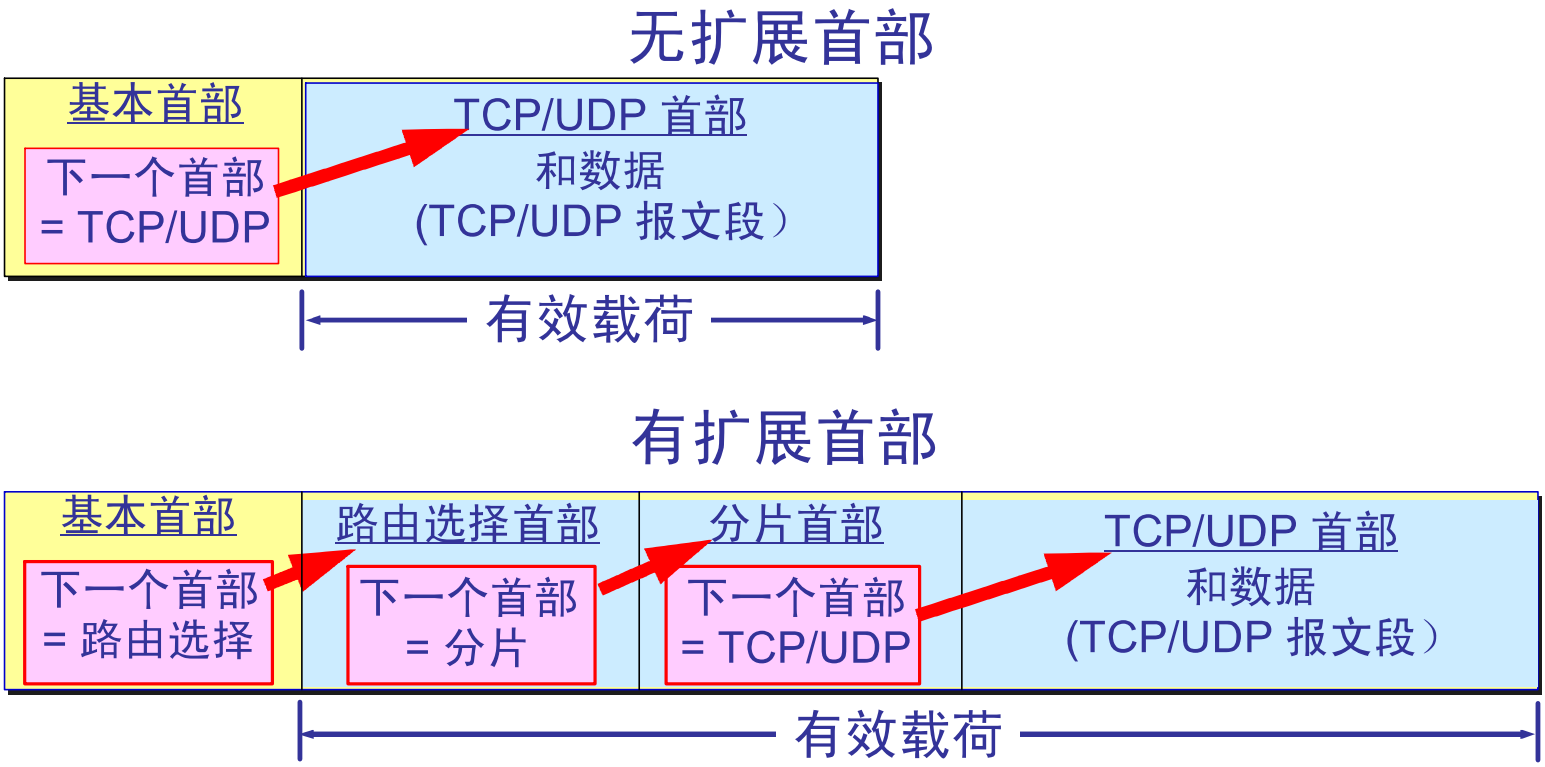
- IPv6 把原来 IPv4 首部中选项的功能都放在扩展首部中，并将扩展首部留给路径两端的源站和目的站的主机来处理。
- 数据报途中经过的路由器都不处理这些扩展首部（只有一个首部例外，即逐跳选项扩展首部）。
- 这样就大大提高了路由器的处理效率。

六种扩展首部

在 RFC 2460 中定义了六种扩展首部：

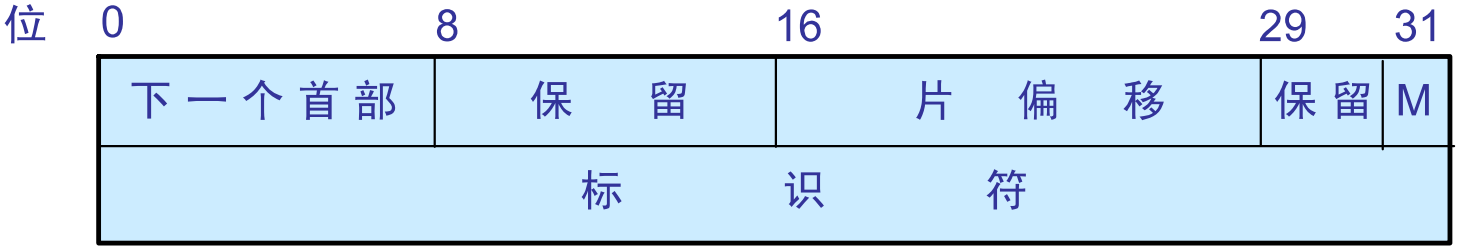
- 逐跳选项
- 路由选择
- 分片
- 鉴别
- 封装安全有效载荷
- 目的站选项

IPv6 的扩展首部


无扩展首部
基本首部
下一个首部
= TCP/UDP
TCP/UDP 首部
和数据
(TCP/UDP 报文段）
有效载荷
有扩展首部
基本首部
下一个首部
= 路由选择
路由选择首部
下一个首部
= 分片
分片首部
下一个首部
= TCP/UDP
TCP/UDP 首部
和数据
(TCP/UDP 报文段）
有效载荷

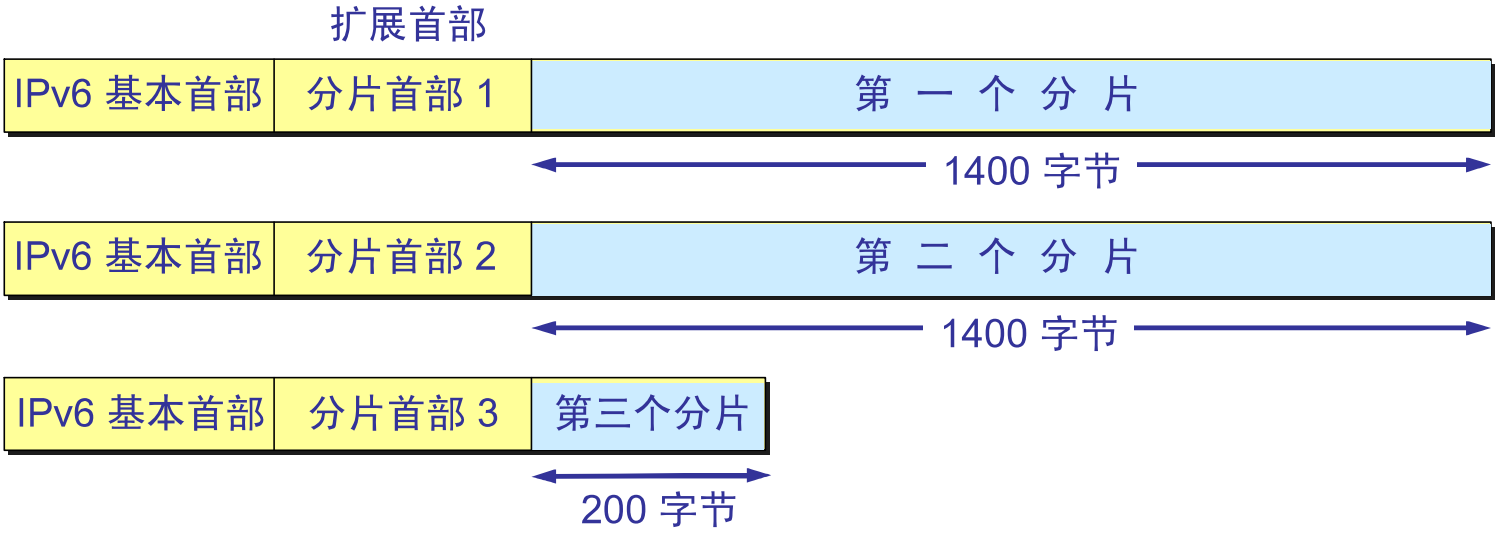
2. 扩展首部举例

- IPv6 把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。

10.1.4 IPv6 的地址空间

1. 地址的类型与地址空间

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唯一地标志该结点。

冒号十六进制记法 (colon hexadecimal notation)

- 每个 16 位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

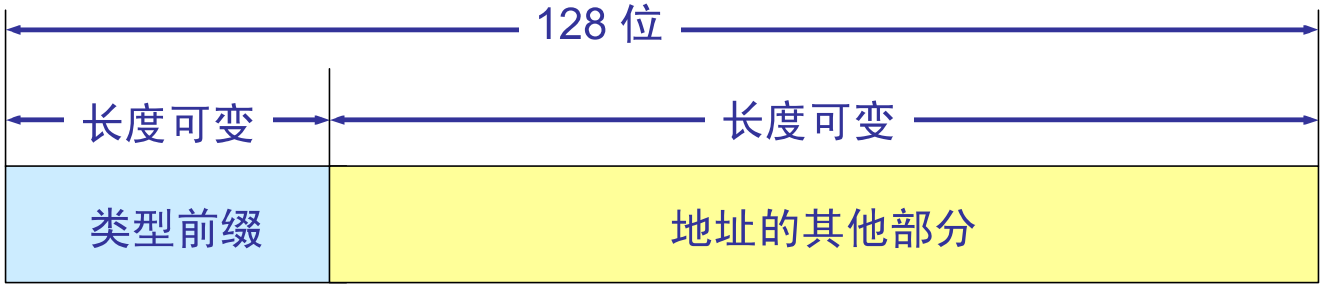
- 0:0:0:0:0:0:128.10.2.1
  再使用零压缩即可得出： ::128.10.2.1
- CIDR 的斜线表示法仍然可用。
- 60 位的前缀 12AB00000000CD3 可记为：
  12AB:0000:0000:CD30:0000:0000:0000:0000/60

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 位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。



3. 特殊地址

未指明地址　这是 16 字节的全 0 地址，可缩写为两个冒号“::”。这个地址只能为还没有配置到一个标准的 IP 地址的主机当作源地址使用。

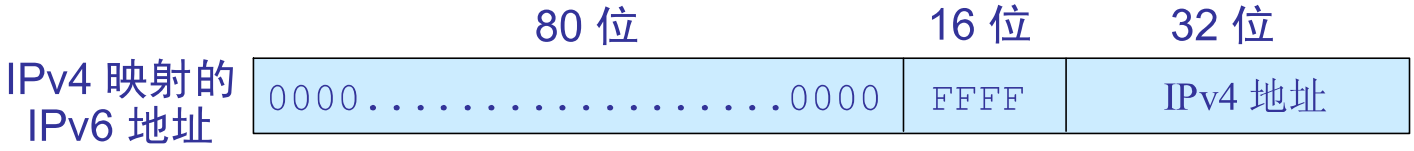
环回地址　即 0:0:0:0:0:0:0:1（记为 ::1）。

基于 IPv4 的地址　前缀为 0000 0000 保留一小部分地址作为与 IPv4 兼容的。

本地链路单播地址

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。



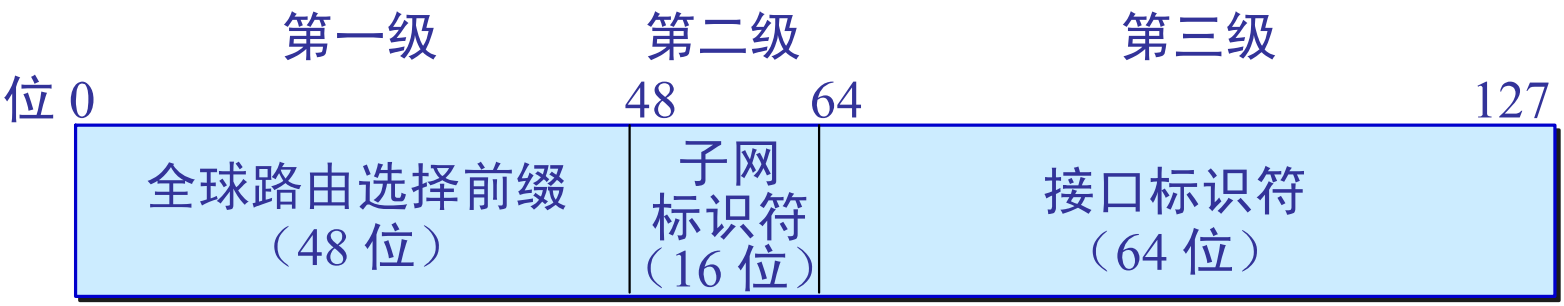
4. 全球单播地址的等级结构

IPv6 扩展了地址的分级概念，使用以下三个等级：

(1) 全球路由选择前缀，占 48 位。

(2) 子网标识符，占16 位。

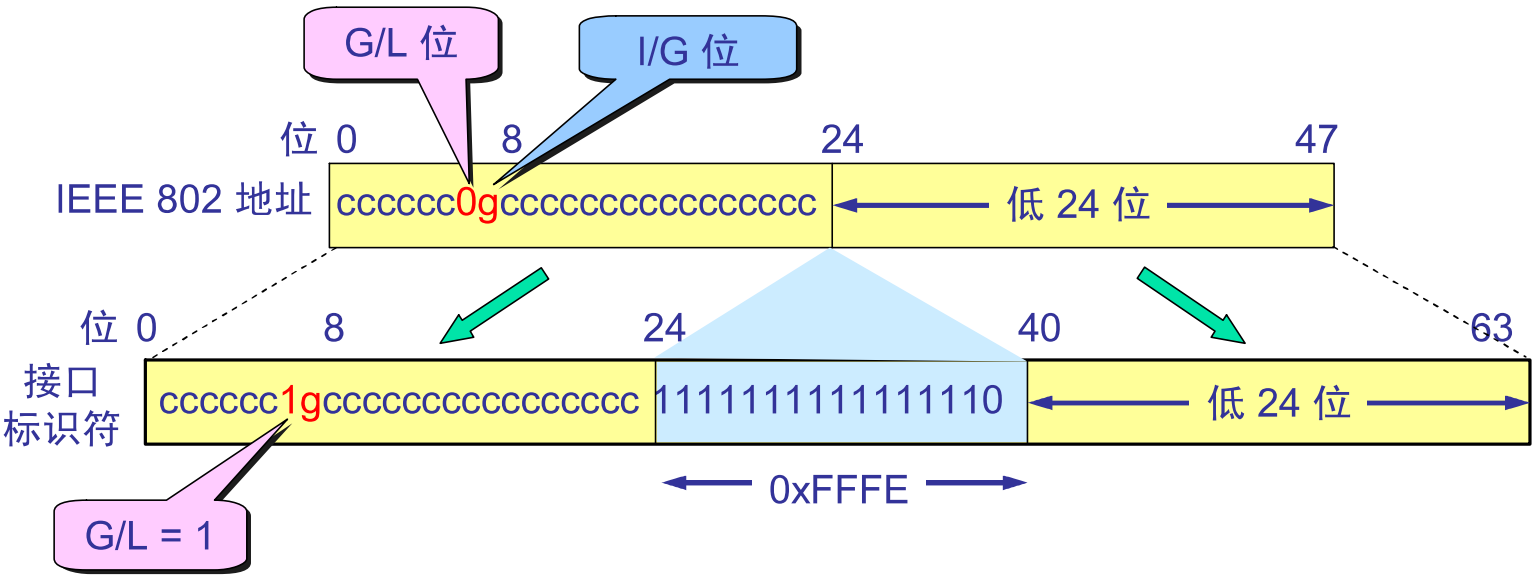
(3) 接口标识符，占 64 位。



EUI-64

- IEEE定 义了一个标准的 64 位全球唯一地址格式 EUI-64。
- EUI-64 前三个字节（24 位）仍为公司标识符，但后面的扩展标识符是五个字节（40 位）。
- 较为复杂的是当需要将 48 位的以太网硬件地址转换为 IPv6 地址。

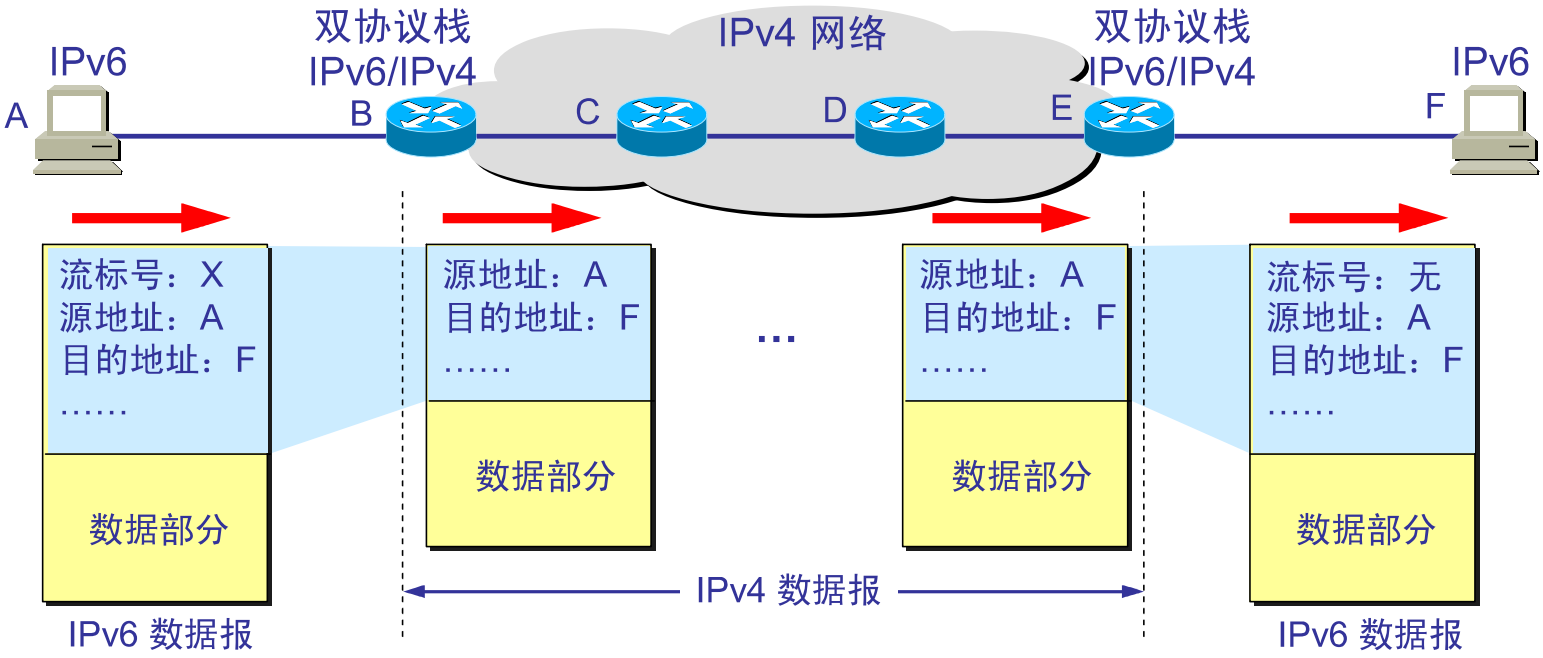
把以太网地址转换为 IPv6 地址



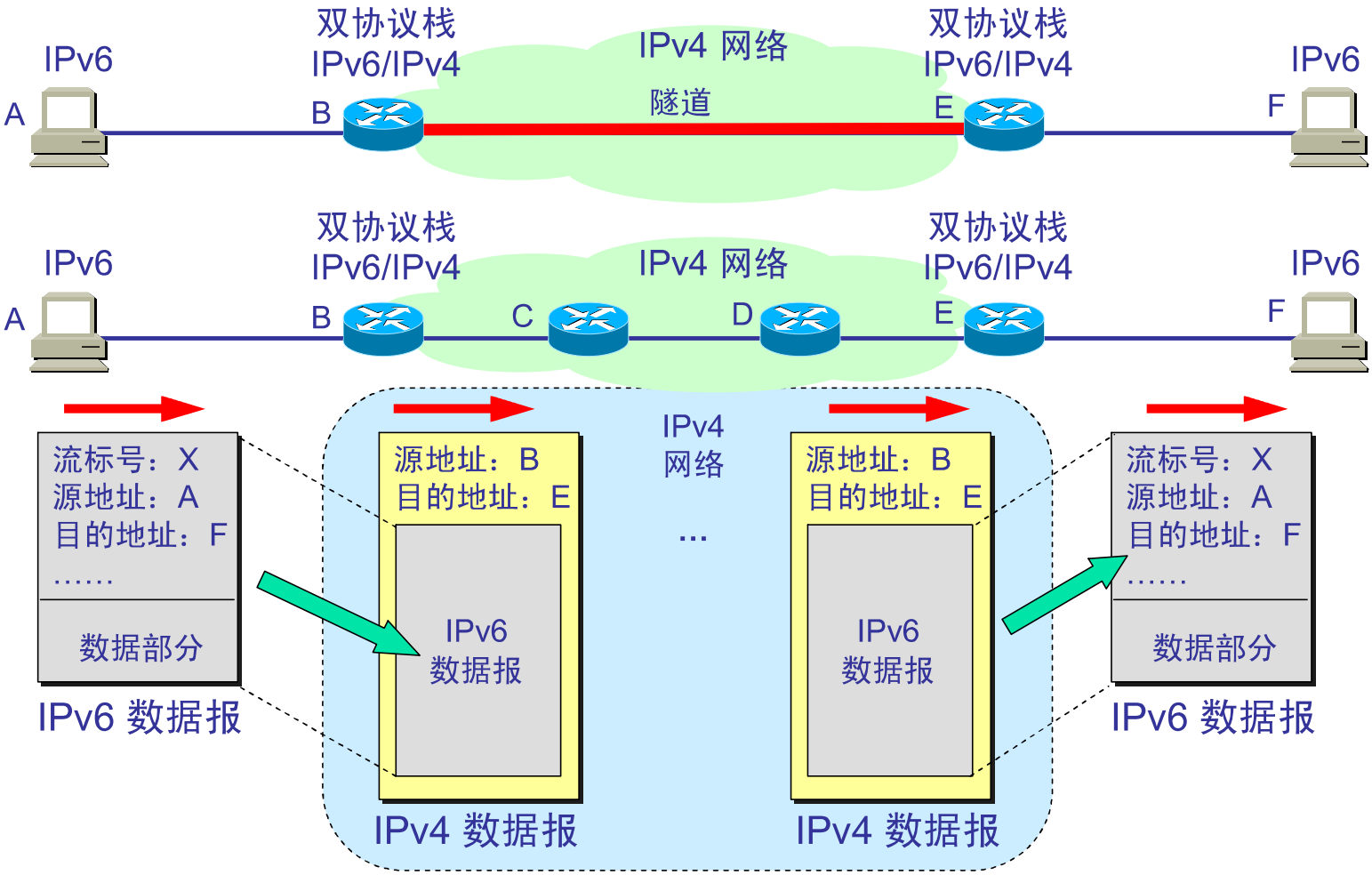
10.1.5 从 IPv4 向 IPv6 过渡

- 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。

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



10.1.6 ICMPv6

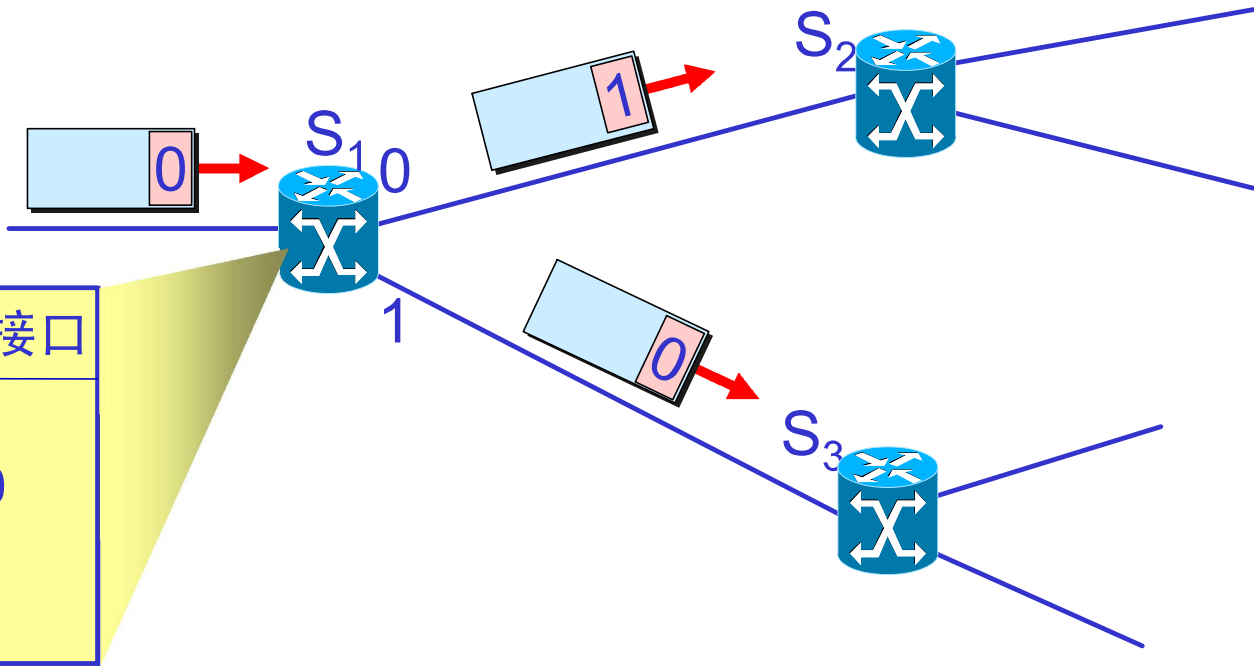
- ICMPv6 的报文格式和 IPv4 使用的 ICMP 的相似，即前 4 个字节的字段名称都是一样的。
- 但 ICMPv6 将第 5 个字节起的后面部分作为报文主体。
- ICMPv6 的报文划分为四大类
  - 差错报告报文
  - 提供信息的报文
  - 多播听众发现报文
  - 邻站发现报文

10.2 多协议标记交换 MPLS (MultiProtocol Label Switching)

10.2.1 MPLS 的产生背景

- 在 20 世纪 80 年代，出现了一种思路：用面向连接的方式取代 IP 的无连接分组交换方式，这样就可以利用更快捷的查找算法，而不必使用最长前缀匹配的方法来查找路由表。
- 这种基本概念就叫做交换(switching)。
- 人们经常把这种交换概念与异步传递方式 ATM (Asynchronous Transfer Mode)联系起来，
- 在传统的路由器上也可以实现这种交换

为了实现交换，可以利用面向连接的概念，使每个分组携带一个叫做标记(label)的小整数。当分组到达交换机时，交换机读取分组的标记，并用标记值来检索分组转发表。



标号	转发接口
0	1
1	0
2	1
3	1

S_1 的转发表

MPLS 的特点

(1) 支持面向连接的服务质量。

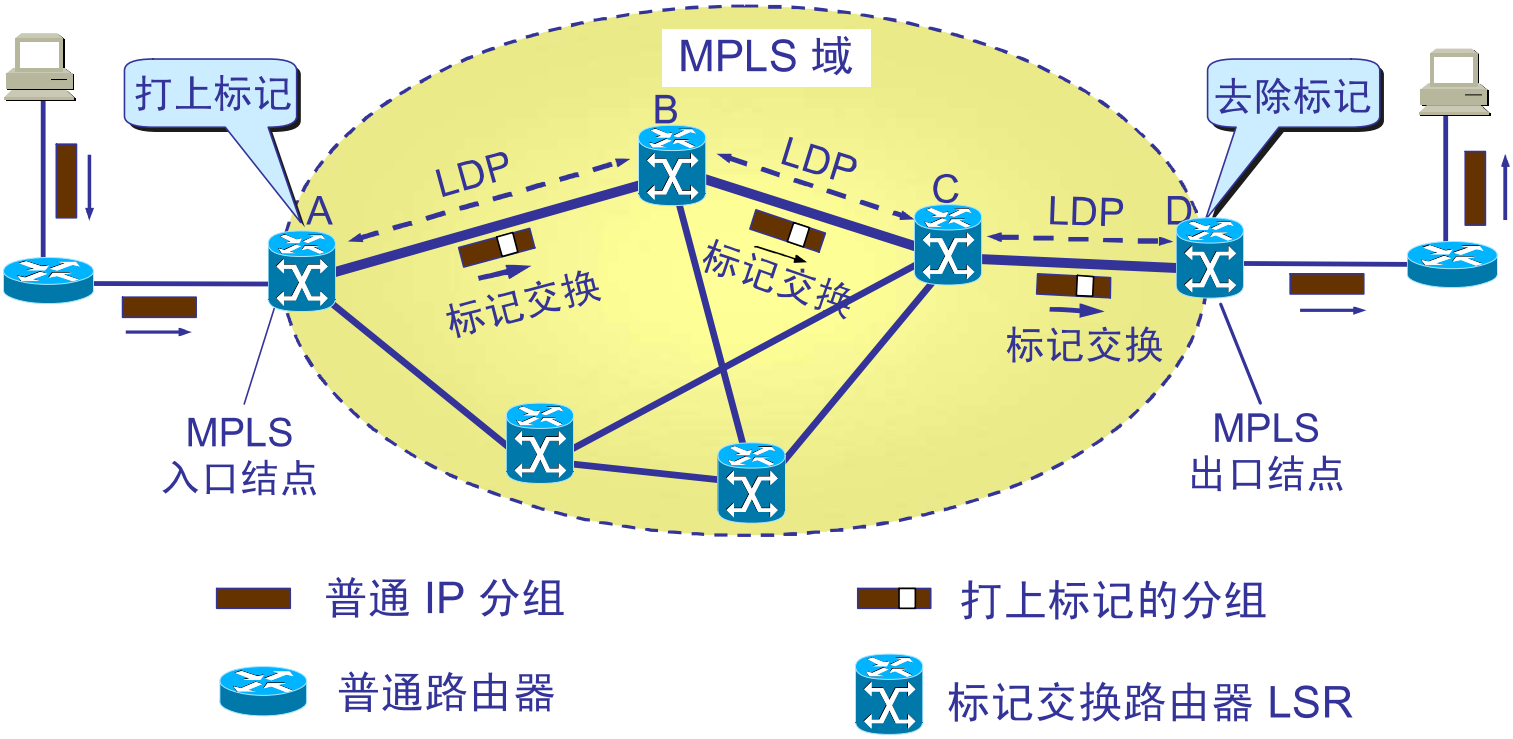
(2) 支持流量工程，平衡网络负载。

(3) 有效地支持虚拟专用网 VPN。

10.2.2 MPLS 的工作原理
1. 基本工作过程

- MPLS 对打上固定长度“标记”的分组用硬件进行转发，使分组转发过程中省去了每到达一个结点都要查找路由表的过程，因而分组转发的速率大大加快。
- 采用硬件技术对打上标记的分组进行转发称为标记交换。“交换”也表示在转发分组时不再上升到第三层用软件分析 IP 首部和查找转发表，而是根据第二层的标记用硬件进行转发。

MPLS 协议的基本原理



普通 IP 分组

打上标记的分组

普通路由器

标记交换路由器 LSR

MPLS 的基本工作过程

(1) MPLS 域中的各 LSR 使用专门的标记分配协议 LDP 交换报文，并找出标记交换路径LSP。各 LSR 根据这些路径构造出分组转发表。

(2) 分组进入到 MPLS 域时， MPLS 入口结点把分组打上标记，并按照转发表将分组转发给下一个 LSR。

(3) 以后的所有 LSR 都按照标记进行转发。每经过一个 LSR，要换一个新的标记。

(4) 当分组离开 MPLS 域时，MPLS 出口结点把分组的标记去除。再以后就按照一般分组的转发方法进行转发。

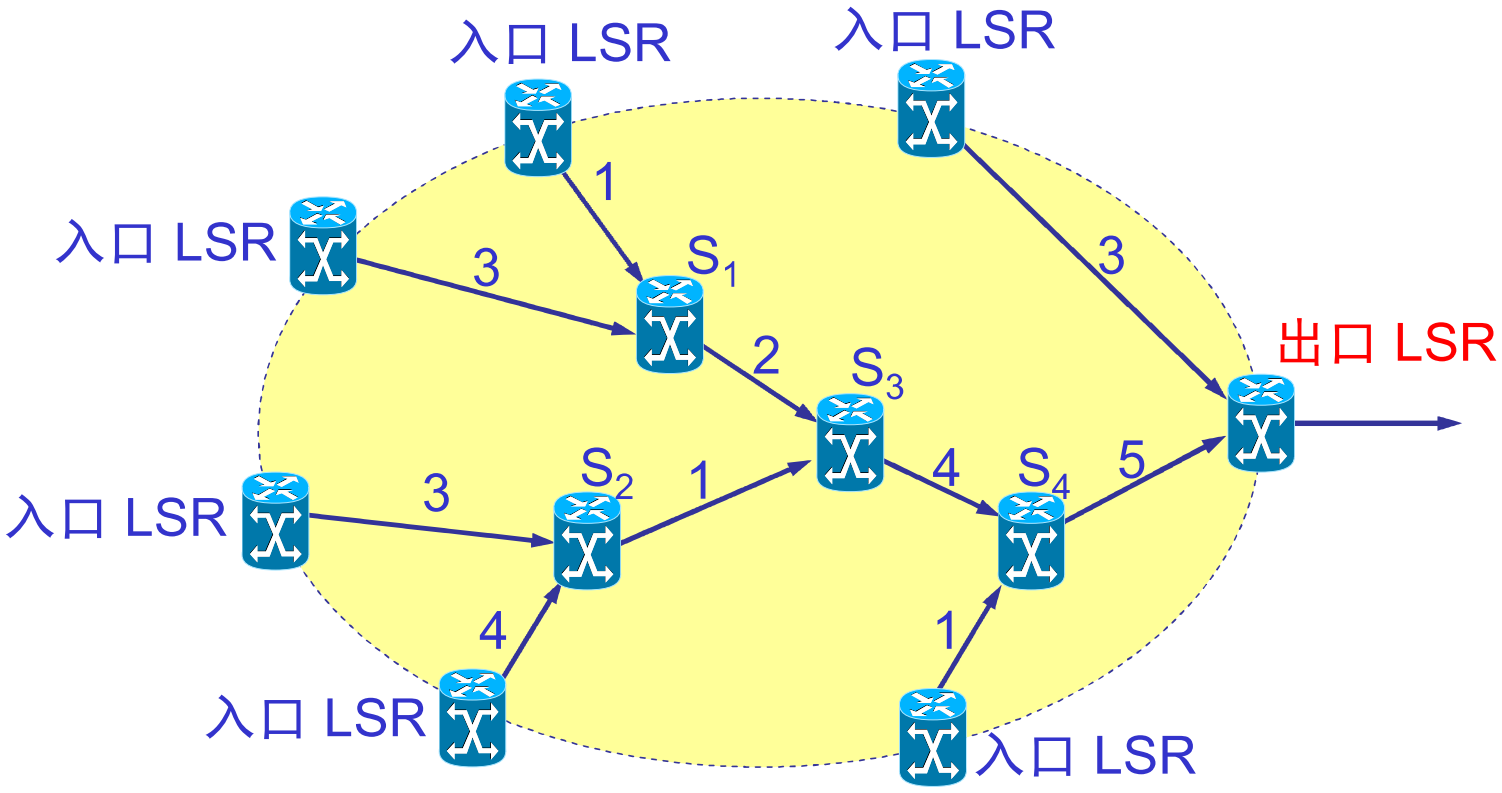
2. 转发等价类 FEC (Forwarding Equivalence Class)

- “转发等价类”就是路由器按照同样方式对待的分组的集合。
- 划分 FEC 的方法不受什么限制，这都由网络管理员来控制，因此非常灵活。
- 入口结点并不是给每一个分组指派一个不同的标记，而是将属于同样 FEC 的分组都指派同样的标记。FEC 和标记是一一对应的关系。 .

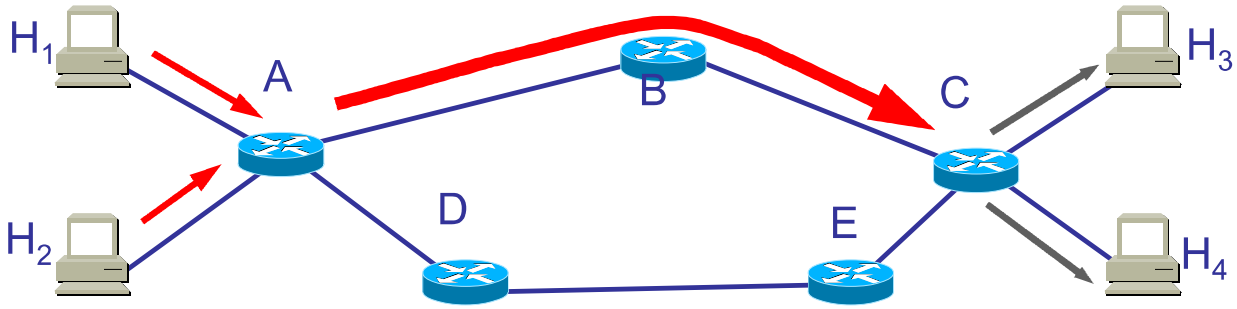
虚电路合并
(VC merging)

- FEC 可以有不同的粒度。
- 细粒度的例子：为特定源主机和目的主机之间的特定应用指派的 FEC。
- 粗粒度的例子：与特定出口 LSR 相关联的 FEC 是。许多应用流聚合到出口 LSR 离开 MPLS 域，它的根在出口 LSR。这种应用流的聚合也称为虚电路合并。这样做可以大大减少转发表中的项目数。

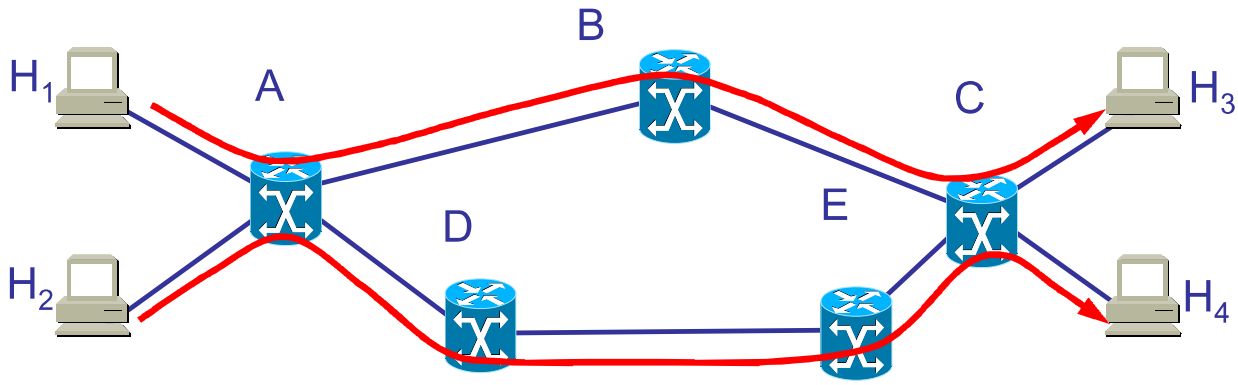
应用流聚合到出口 LSR



FEC 用于负载平衡



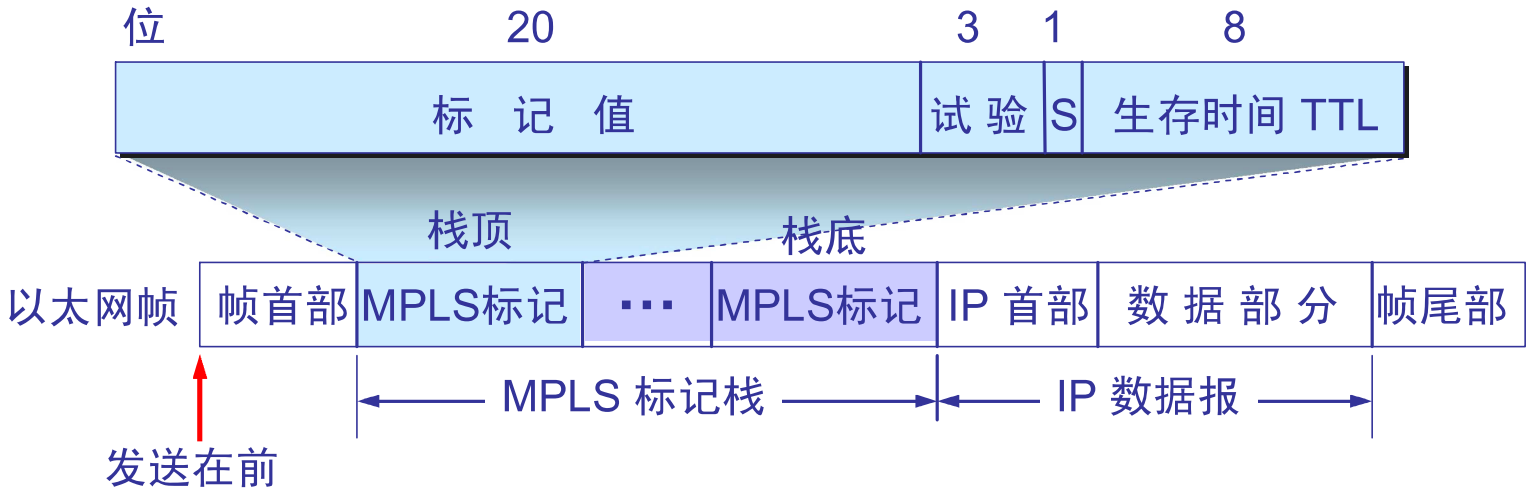
(a) 传统路由选择协议使最短路径 A→B→C 过载



(b) 利用 FEC 使通信量分散

10.2.3 MPLS 首部的位置与格式

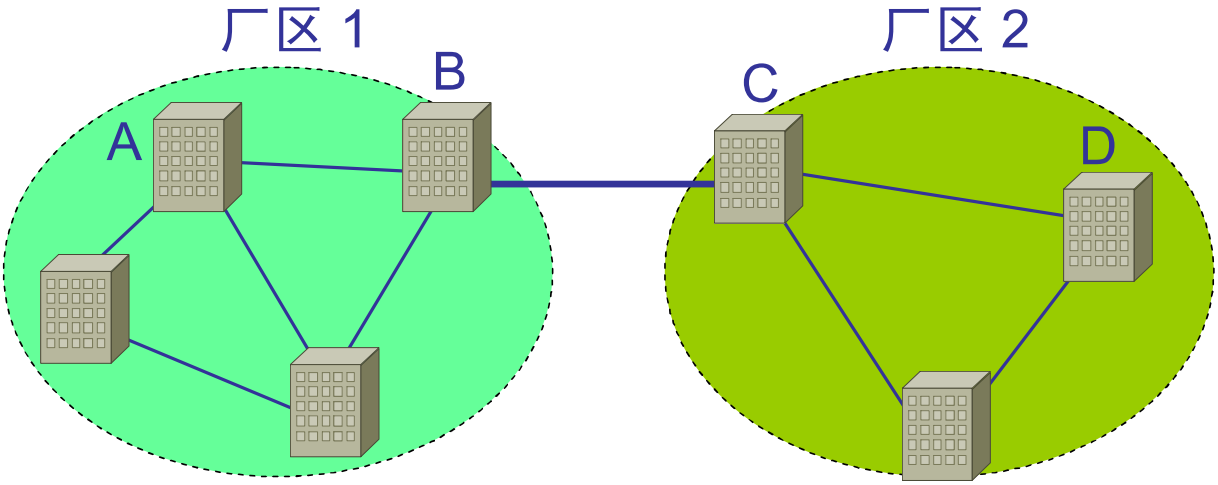
- MPLS 的一个重要功能就可以构成标记栈。
- MPLS 标记的格式以及标记栈：

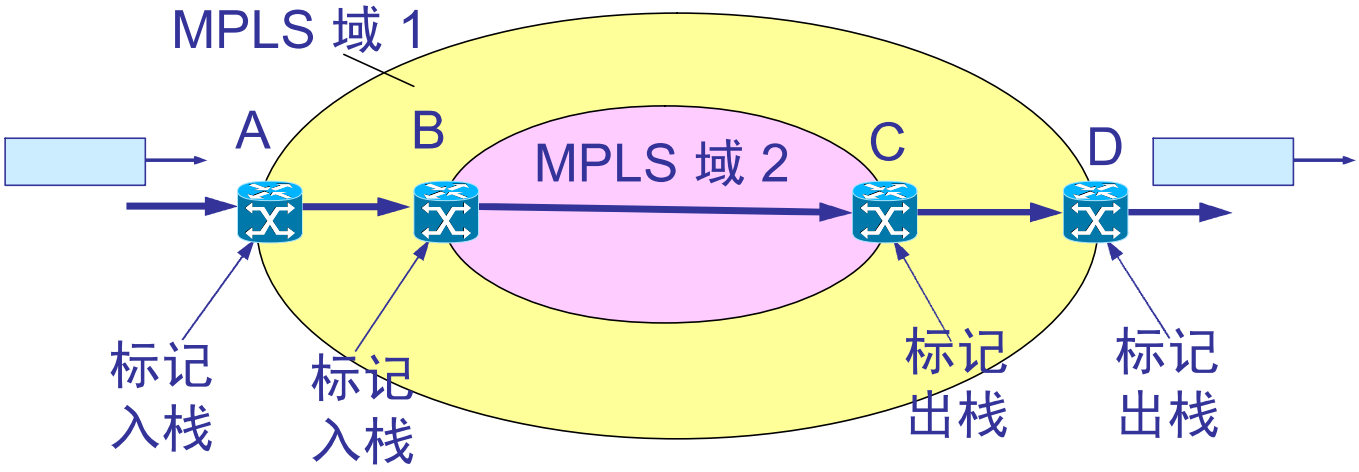


MPLS 标记

- MPLS 标记一旦产生就压入到标记栈中，而整个标记栈放在数据链路层首部和IP首部之间。
- 栈是一种后进先出的数据结构。MPLS 协议规定，标记栈的栈顶（最后进入栈的标记）最靠近数据链路层首部，而栈底最靠近 IP 首部。
- 在最简单的情况下，标记栈中只有一个标记。

MPLS 标记栈的使用





10.3 P2P 文件共享

- 自从因特网能够提供音频/视频服务后，宽带上网用户数也急剧增长。很多用户使用宽带接入的目的就是为了更快地下载音频/视频文件。
- P2P 工作方式受到广大网民的欢迎。这种工作方式解决了集中式媒体服务器可能出现的瓶颈问题。
- 在 P2P 工作方式下，所有的音频/视频文件都是在普通的因特网用户之间传输。这是相当于有很多分散在各地的媒体服务器向其他用户提供所要下载的音频/视频文件。

Napster

- 最早出现的 P2P 技术，可提供免费下载 MP3 音乐。
- Napster 能够搜索音乐文件，能够提供检索功能。所有的音乐文件地址集中存放在一个 Napster 目录服务器中。使用者可很方便地下载需要的 MP3 文件。
- 用户要及时向 Napster 的目录服务器报告自己存有的音乐文件。当用户想下载某个 MP3 文件时，就向目录服务器发出询问。目录服务器检索出结果后向用户返回存放此文件的 PC 机的 IP 地址。Napster 的文件传输是分散的，但文件的定位则是集中的。
- 这种集中式目录服务器的缺点就是可靠性差。Napster 被判决属于“间接侵害版权”，因此在 2000 年 7 月底 Napster 网站就被迫关闭了。

Gnutella

- Gnutella 是第二代 P2P 文件共享程序，它全分布方法定位内容的P2P 文件共享应用程序。 。
- Gnutella 与 Napster 最大的区别就是不使用集中式的目录服务器，而是使用洪泛法在大量 Gnutella 用户之间进行查询。
- 为了不使查询的通信量过大，Gnutella 设计了一种有限范围的洪泛查询。这样可以减少倾注到因特网的查询流量，但由于查询的范围受限，因而这也影响到查询定位的准确性。

第三代 P2P 共享文件程序

- eMule 使用分散定位和分散传输技术，把每一个文件划分为许多小文件块，并使用多源文件传输协议 MFTP 进行传送。因此用户可以同时从很多地方下载一个文件中的不同文件块。由于每一个文件块都很小，并且是并行下载，所以下载可以比较快地完成。
- eMule 用户在下载文件的同时，也在上传文件，因此，因特网上成千上万的 eMule 用户在同时下载和上传一个个小的文件块。

eMule 的其他特点

- eMule 使用了一些服务器。这些服务器并不是保存音频/视频文件，而是保存用户的有关信息，因而可以告诉用户从哪些地方可以下载到所需的文件。
- eMule 使用了专门定义的文件夹，让用户存放可以和其他用户共享的文件。
- eMule 的下载文件规则是鼓励用户向其他用户上传文件。用户上传文件越多，其下载文件的优先级就越高（因而下载就越快）。