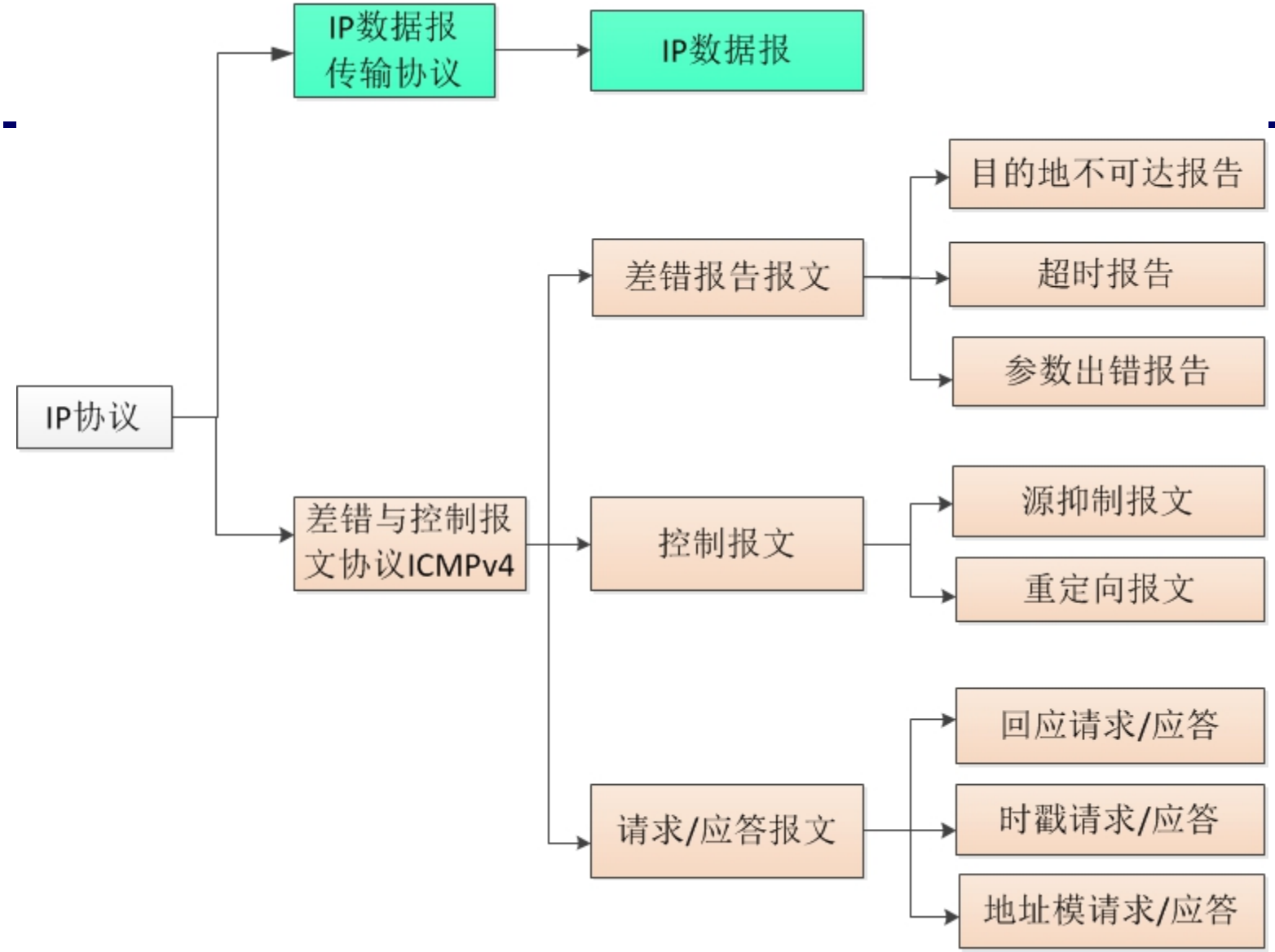
下一代Internet技术与协议

张冬梅
北京邮电大学 计算机学院
zhangdm@bupt.edu.cn


IP协议
IP数据报传输协议
IP数据报
差错与控制报文协议ICMPv4
差错报告报文
目的地不可达报告
超时报告
参数出错报告
控制报文
源抑制报文
重定向报文
请求/应答报文
回应请求/应答
时戳请求/应答
地址模请求/应答

IP协议

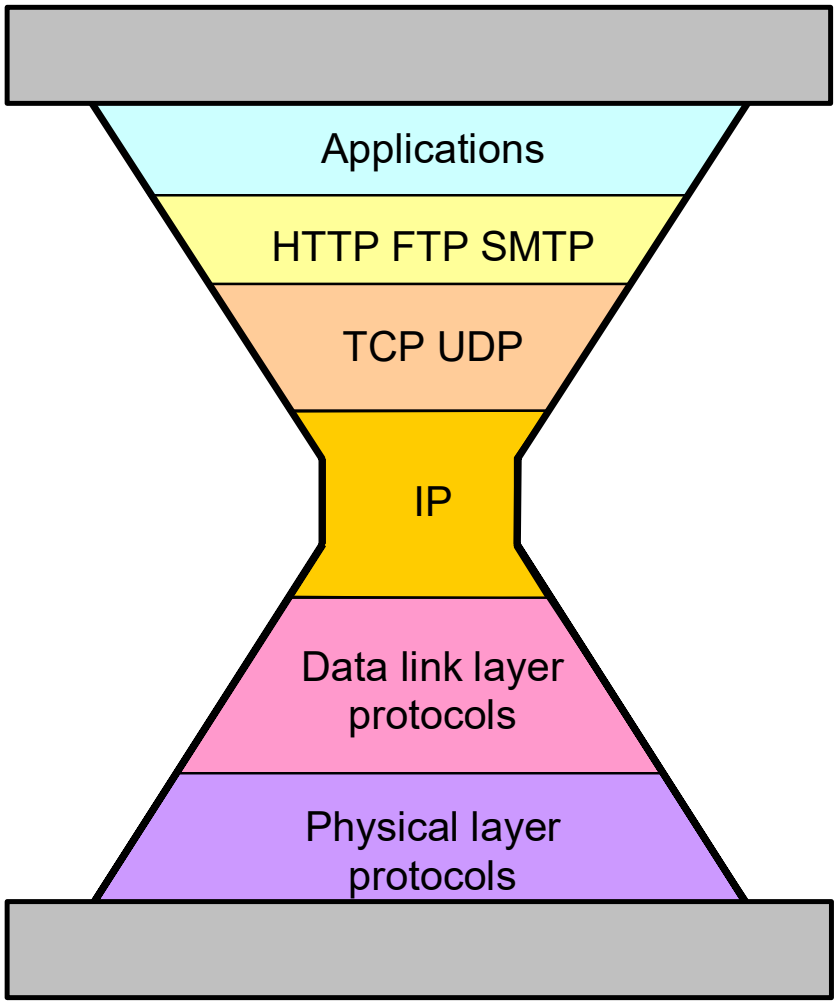
- IP协议概述
- IPv4
- IPv6

IP协议概述

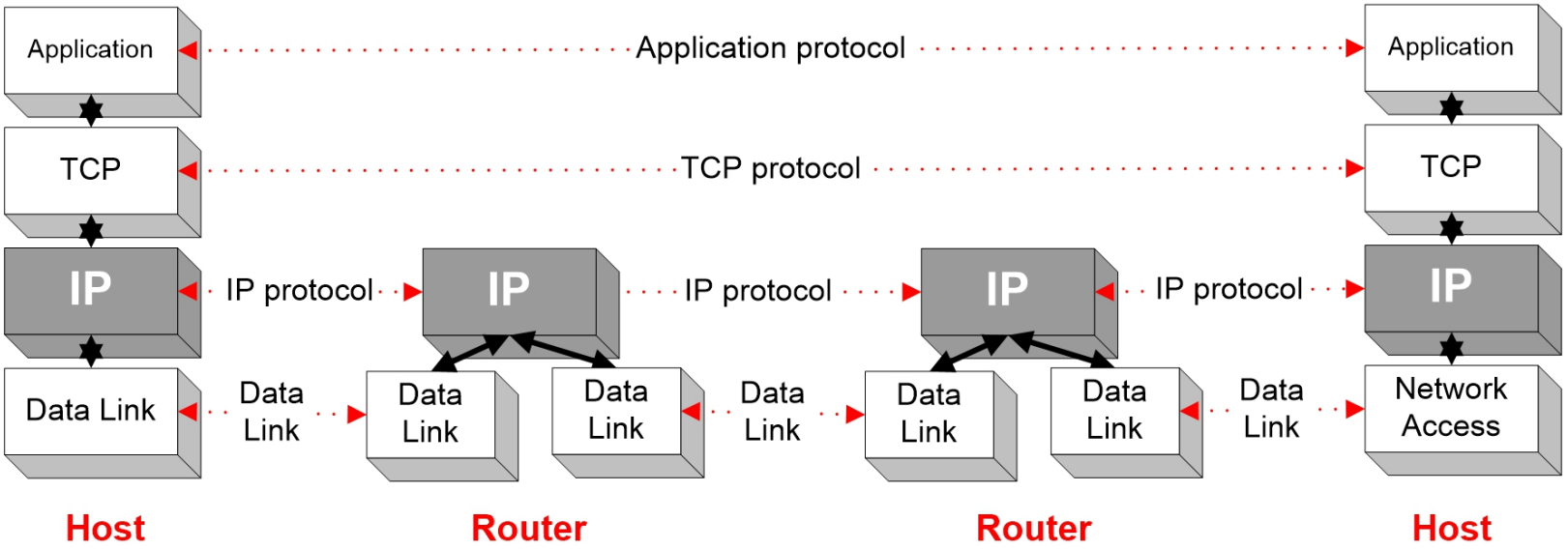
- 互联网通信协议的工作环境
  - 用户需求
  - 底层通道环境
- 互联网通信协议的功能
- IP协议提供的服务

IP在协议栈中的位置

- 多个高层协议
- 多个低层协议
- 唯一的网络层协议



- 是路由器和主机共同支持的最高层协议



IP 提供的服务

- 不可靠（unreliable connectionless）的尽力投递服务
  - **不可靠:** IP does not make an attempt to recover lost packets
  - **无连接:** Each packet (“datagram”) is handled independently. IP is not aware that packets between hosts may be sent in a logical sequence
  - **尽力投递:** IP does not make guarantees on the service
- 影响
  - 高层协议需要处理丢包等问题
  - 分组乱序

IP数据报简介

- 通用的虚拟包
- 封装
  - IP数据报与帧
  - 底层封装
  - 互联网传输过程
- 分段与重组
  - 最大传送单元(Maxium Transmission Unit)
  - 原因
  - 方法

IPv4

- IP协议内容
- IPv4数据报
- IPv4路由技术的工作原理
- IPv4变革动机

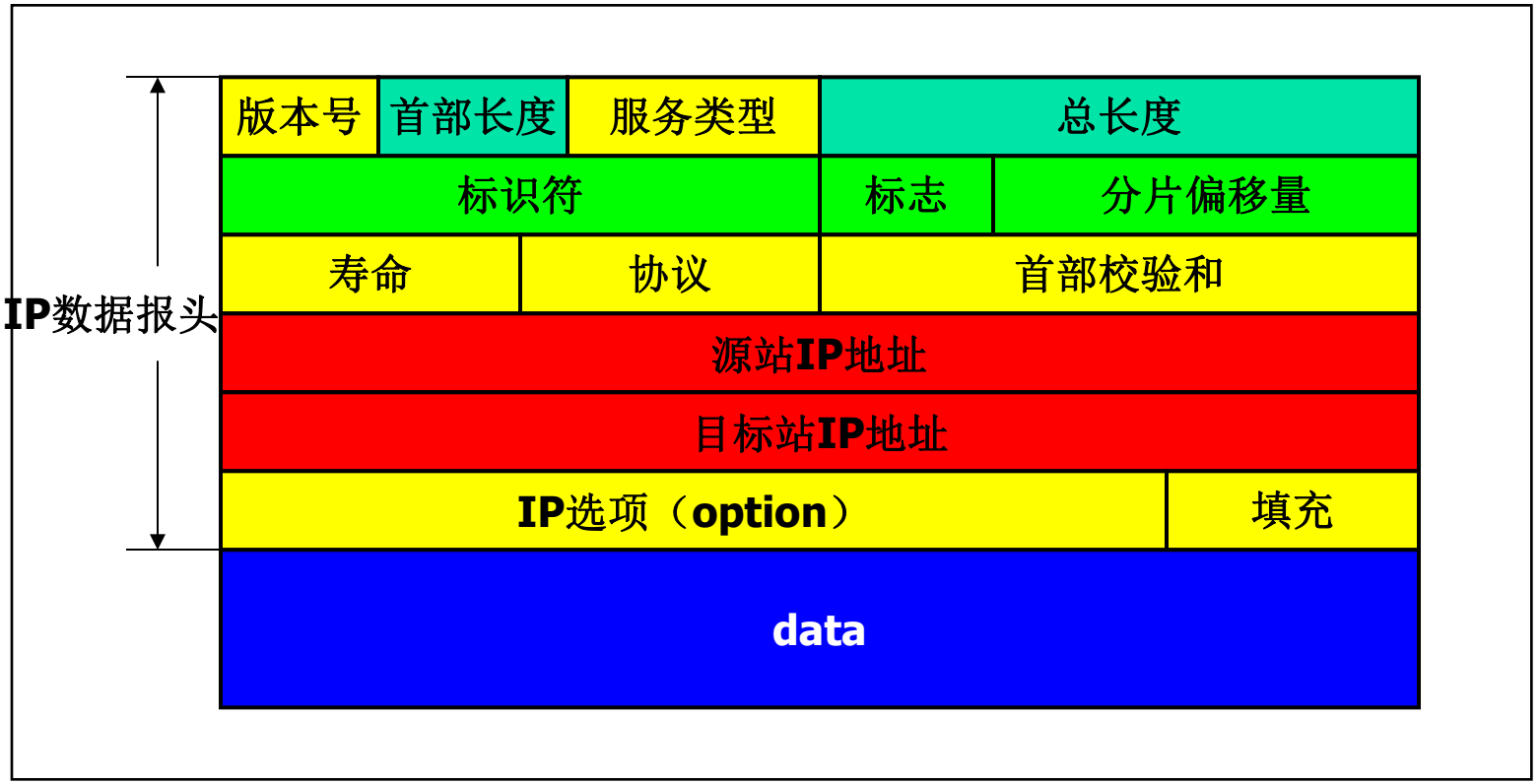
IP协议内容

- IP地址的编址方案(已经讲过)
- IP数据报格式
- IP路由和交换

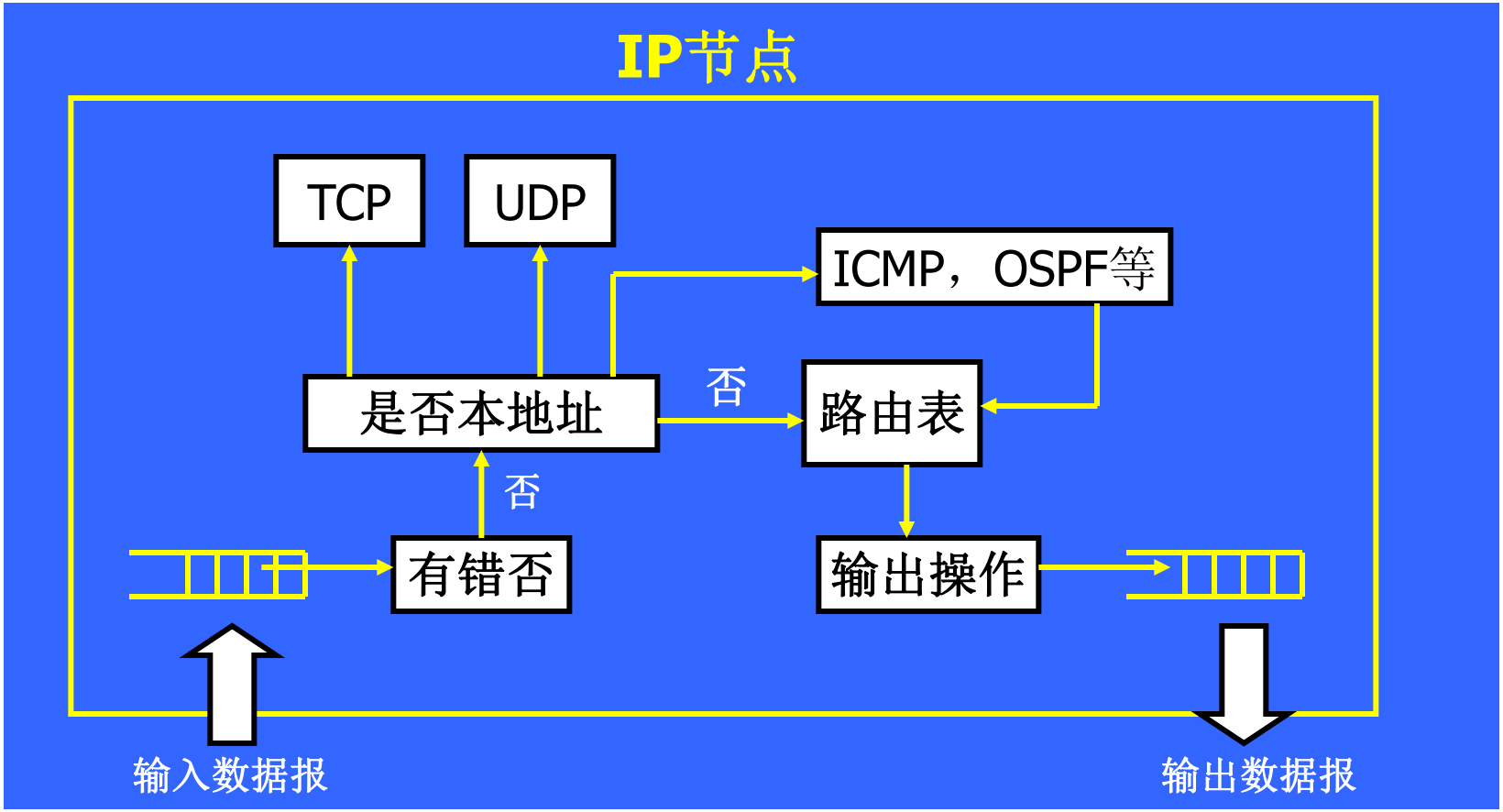
IPv4数据报

- 简介
  - 互联网的基本传输单元: IP数据报
  - IP数据报处理在软件中进行
- IP数据报格式

IPv4数据报格式

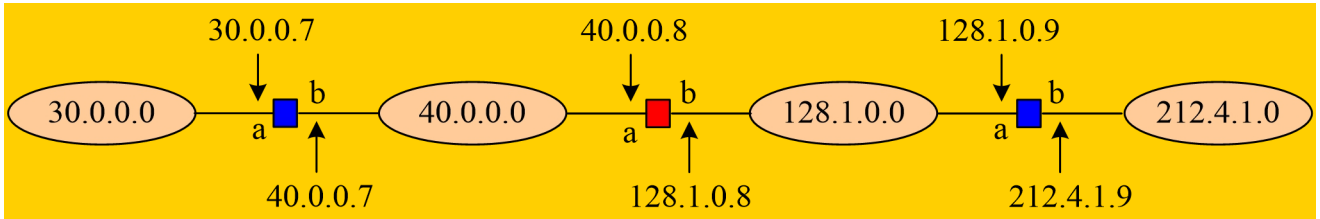

IP数据报头
版本号
首部长度
服务类型
总长度
标识符
标志
分片偏移量
寿命
协议
首部校验和
源站IP地址
目标站IP地址
IP选项（option）
填充
data

IP节点处理数据报过程



IP路由技术的工作原理(1/6)

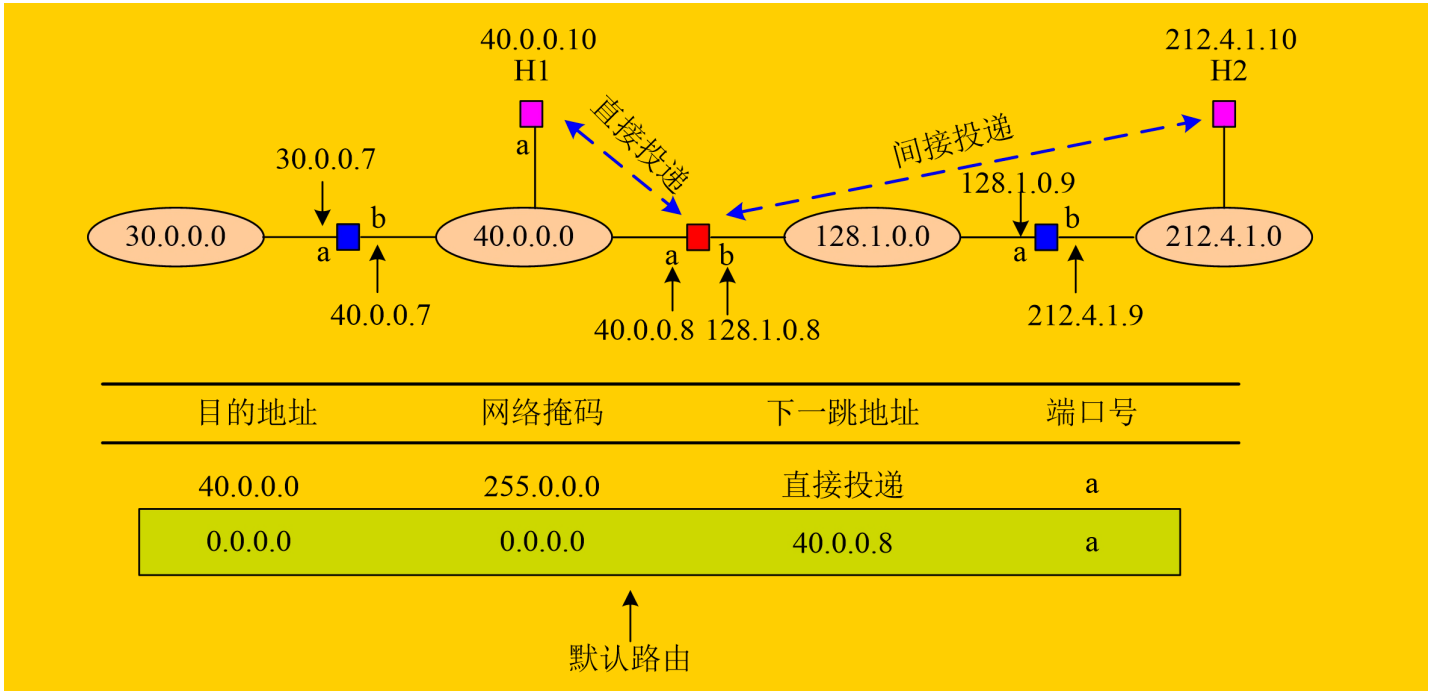
- 路由表的内容
  - 目的地址
  - 前缀长度
  - 下一跳地址
  - 端口号



目的地址	网络掩码	下一跳地址	端口号
30.0.0.0	255.0.0.0	40.0.0.7	a
40.0.0.0	255.0.0.0	直接投递	a
128.1.0.0	255.255.0.0	直接投递	b
212.4.1.0	255.255.255.0	128.1.0.9	b

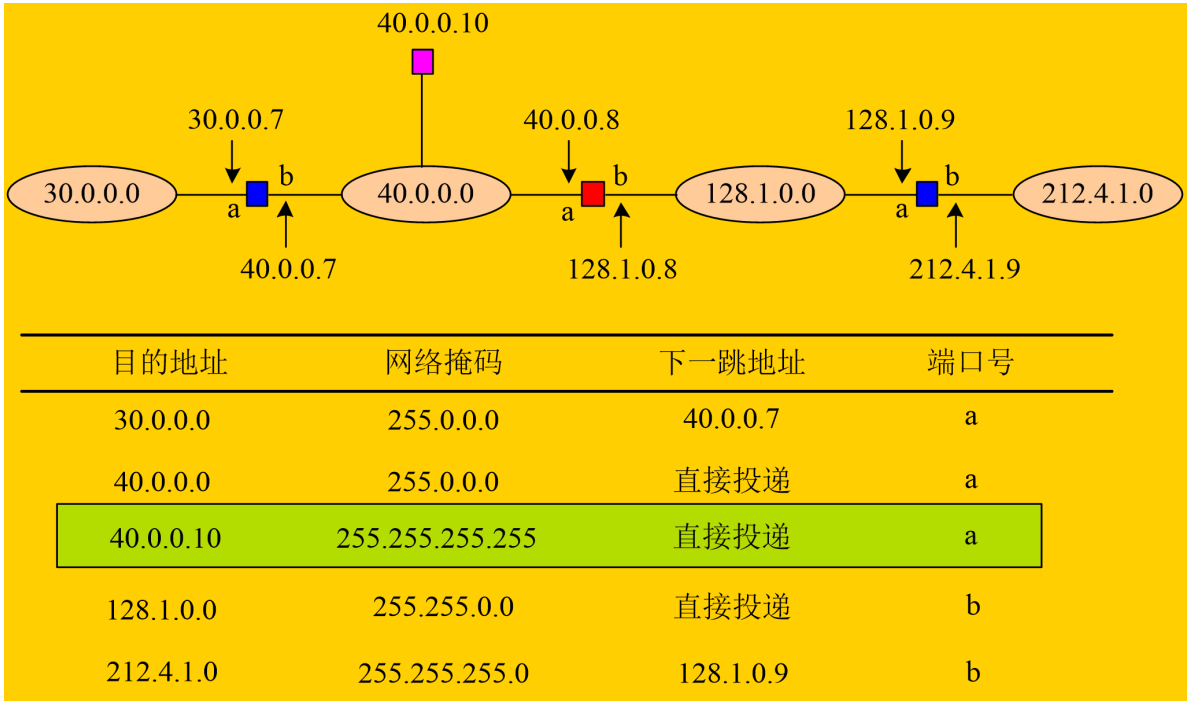
IP路由技术的工作原理(2/6)

- 基本概念
  - 直接投递：通信双方在同一个物理网络中
  - 间接投递：通信双方不在一个物理网络中
  - 默认路由：简化路由表的一种方法



IP路由技术的工作原理(3/6)

- 路由表表项的分类
  - 特定主机路由
  - 网络前缀路由
  - 缺省路由（默认路由）



目的地址	网络掩码	下一跳地址	端口号
30.0.0.0	255.0.0.0	40.0.0.7	a
40.0.0.0	255.0.0.0	直接投递	a
40.0.0.10	255.255.255.255	直接投递	a
128.1.0.0	255.255.0.0	直接投递	b
212.4.1.0	255.255.255.0	128.1.0.9	b

IP路由技术的工作原理(4/6)

- 路由匹配原则归纳
  - 首选：特定主机路由
  - 其次：最长网络前缀匹配
  - 最后：缺省路由
  - 路由错误，ICMP报错

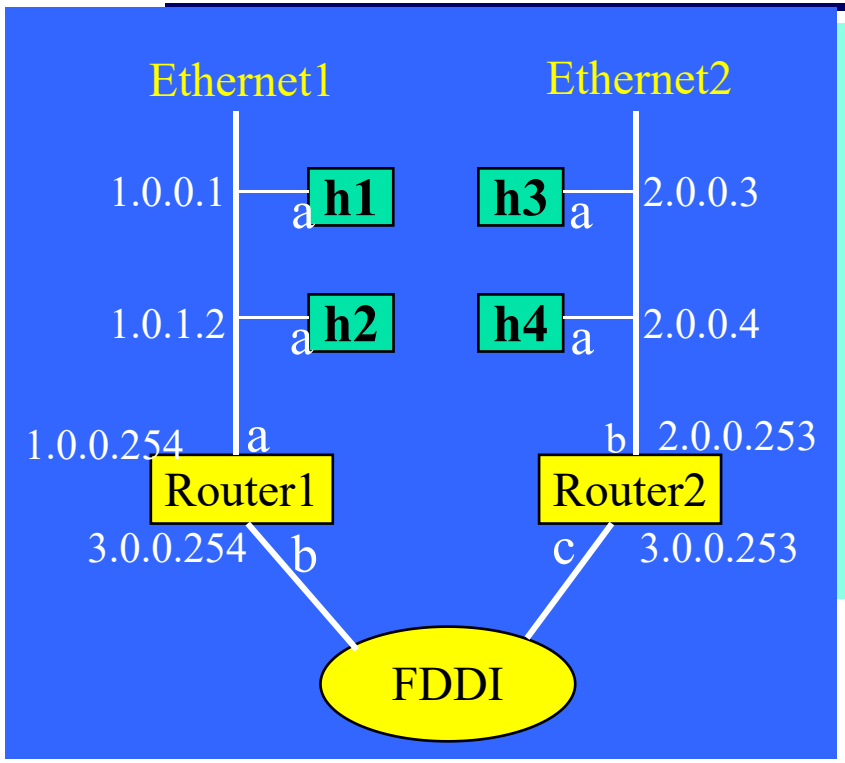
IP路由技术的工作原理(5/6)

- 生成路由表的方法
  - 静态生成，手工配置
  - 利用ICMP消息
  - 利用动态路由协议
    - OSPF
    - RIP
    - BGP

IP路由技术的工作原理(6/6)

- 生成路由表的基本工作过程
  - 制定各物理网络的网络前缀
  - 配置路由器各端口的IP地址及前缀长度
  - 路由器学习邻局路由器的信息
  - 路由器根据路由协议，定期交换路由更新信息，更新自己的路由表

IP网络路由举例



Router1

目的地址	网络掩码	下一跳地址	端口号
1.0.0.0	255.0.0.0	直接投递	a
2.0.0.0	255.0.0.0	3.0.0.253	b
3.0.0.0	255.0.0.0	直接投递	b

Router2

目的地址	网络掩码	下一跳地址	端口号
1.0.0.0	255.0.0.0	3.0.0.254	c
2.0.0.0	255.0.0.0	直接投递	b
3.0.0.0	255.0.0.0	直接投递	c

源地址：**1.0.0.1**
目的地址：**1.0.1.2，2.0.0.3**
2.0.0.7,4.0.0.5

H1

目的地址	网络掩码	下一跳地址	端口号
0.0.0.0	0.0.0.0	1.0.0.254	a
1.0.0.0	255.255.255.240	直接投递	a

H3

目的地址	网络掩码	下一跳地址	端口号
0.0.0.0	0.0.0.0	2.0.0.253	a
2.0.0.0	255.0.0.0	直接投递	a

IPv4变革的动机

- 有限的地址空间
- 新的Internet应用
  - 传递音频和视频信息
  - 对更复杂的寻址和路由能力的需求
  - 移动服务

IPv6

- IPv6简介
- IPv6特征
- IPv6协议
- 常用的几种扩展头标简介

IPv6简介

- IPv6是IP协议的新版本
- 也叫做IPng(IP新一代)
- 1995年随RFC1883的出现而完成
- 1998年RFC2460取代了RFC1883
- IPv6的目的就是要解决IPv4遇到的问题
- IPv6将是未来唯一的第三层协议
- IPv6将是下一代Internet的基础协议

IPv6特征

- IPv6保留了IPv4的成功特征
- IPv6的新加特征
  - 地址尺寸
  - 头部格式
  - 扩展头标
  - 对音频和视频的支持
  - 可扩展的协议

IPv6协议

- 基本术语
  - 节点（node）：任何实现了IPv6的设备
  - 路由器：转发IPv6报文的节点
  - 主机：在网络上除了路由器的节点
  - 链路：节点利用来在链路层通信的通信设备或介质，如以太网、PPP链路或网络层隧道

相关术语

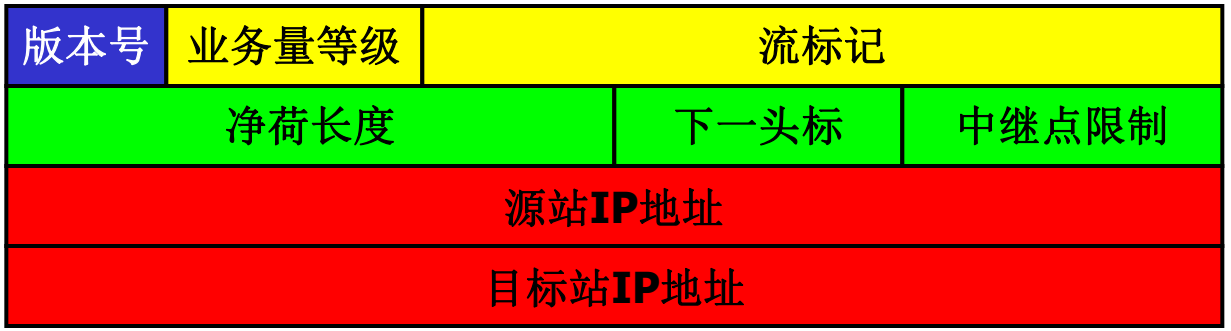
- 邻居：连接在同一链路上的节点
- 接口：结点与链路相连接的部件
- 链路MTU：在某一链路上的最大传输单元
- 路径MTU：出发点和目的节点之间的路径上所有链路的最小链路MTU.

IPv6格式

- IPv6报头特点
  - 报头大大简化
  - 固定的基本报头长度
  - 去掉报头校验和
  - IPv4报头中的一些字段被取消或是变成可选项
  - 用扩展报头代替了IPv4报头中的选项字段

IPv6格式

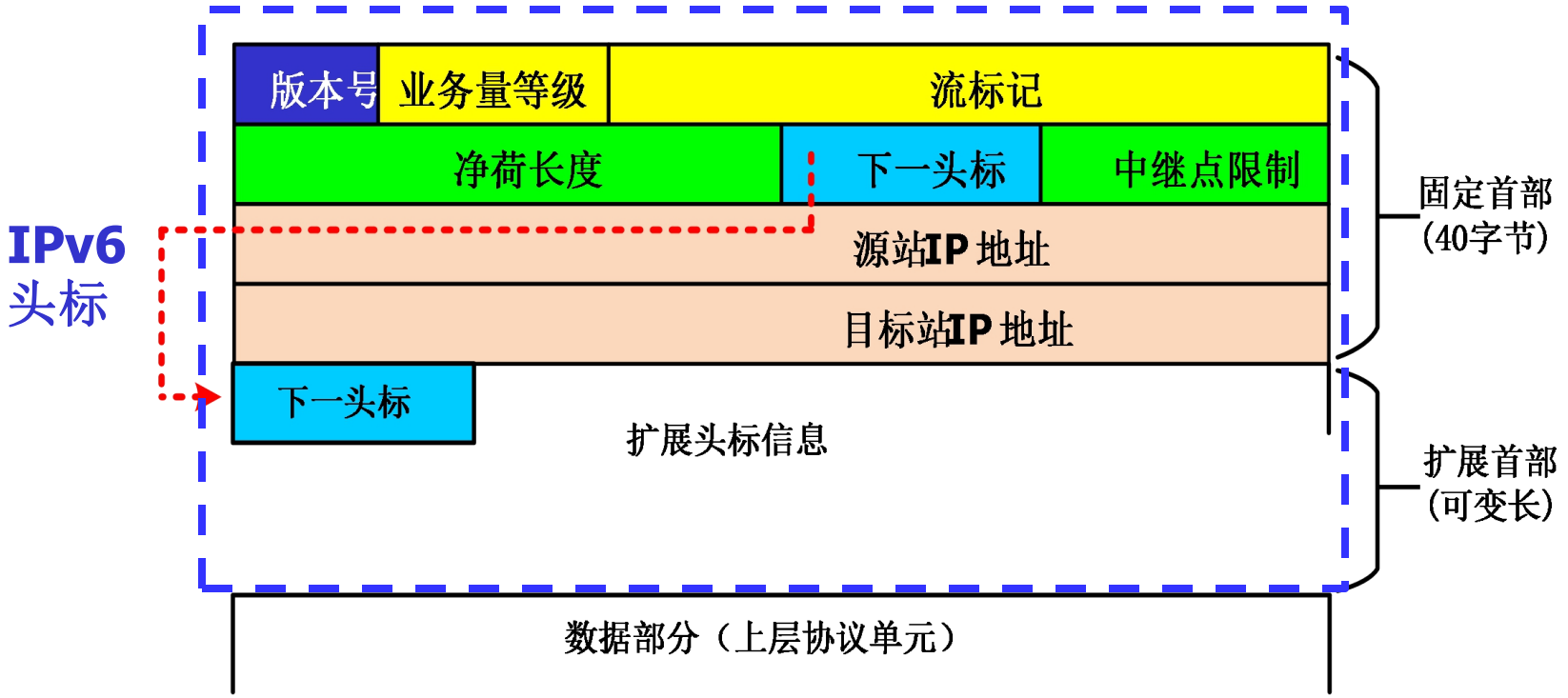
- IPv6数据报格式
  - 简化的头标（40字节基本/固定头标）
  - 参数的修订
  - 新增加的域



IPv6格式

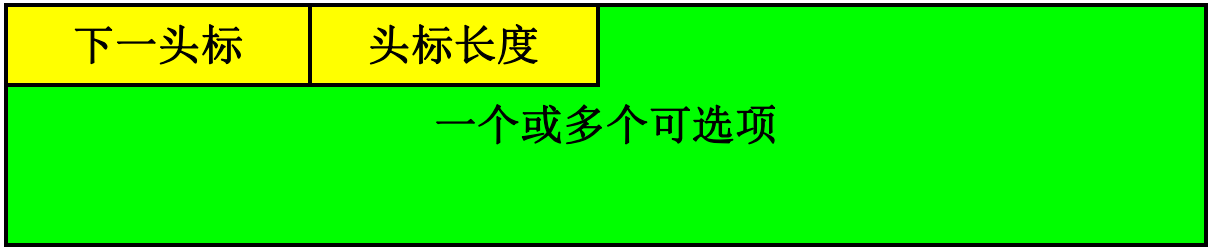
- 扩展报头
  - IPv6使用**扩展报头**来代替IPv4的**选项**字段。以此来减少IPv6信息包中途经过路由器时的处理时间。扩展报头可位于IPv6报头和上层协议之间，报头之间由下一个报头字段进行连接，这样组成一个菊花链式结构。
  - 一个IPv6信息包可以有0个，1个或多个扩展报头。

IPv6格式


IPv6
头标
版本号
业务量等级
流标记
净荷长度
下一头标
中继点限制
源站IP 地址
目标站IP 地址
下一头标
扩展头标信息
固定首部
(40字节)
扩展首部
(可变长)
数据部分（上层协议单元）

IPv6扩展头标

- 基本格式



IPv6格式—下一头标

取值(十进制)	含义	取值(十进制)	含义
0	Hop-by-hop 逐跳选项首部	47	通用路由封装GRE
1	ICMPv4	50	ESP
2	IGMPv4	51	AH
4	IPv4封装	58	ICMPv6
5	IST(Internet Stream Protocol)	59	无下一头标
6	TCP	60	Destination option目的选项扩展头标
8	EGP	88	EIGRE(Enhanced Interior Gateway Routing Protocol)
9	IGP	89	OSPF
17	UDP	108	IP有效载荷压缩协议
41	IPv6封装	115	L2TP(二层隧道传输协议)
43	routing路由扩展首部	132	流控制传输协议SCTP
44	fragment分段扩展首部	135	Monility移动扩展头标，移动节点使用
46	RSVP		

路由器转发IPv6分组过程

- ~~检查首部校验和~~
- 检查版本字段
- 递减生存时间(跳数限制)字段的值
- √ 处理首部选项(下一首部)字段的值，依次处理
- 路由选择
  - 下一跳地址、默认路由、discard and ICMP
- √ 处理分组总长度(有效载荷)长度的问题
  - discard and ICMP
- ~~计算校验和~~
- 依据路由选择结果转发分组

扩展头标

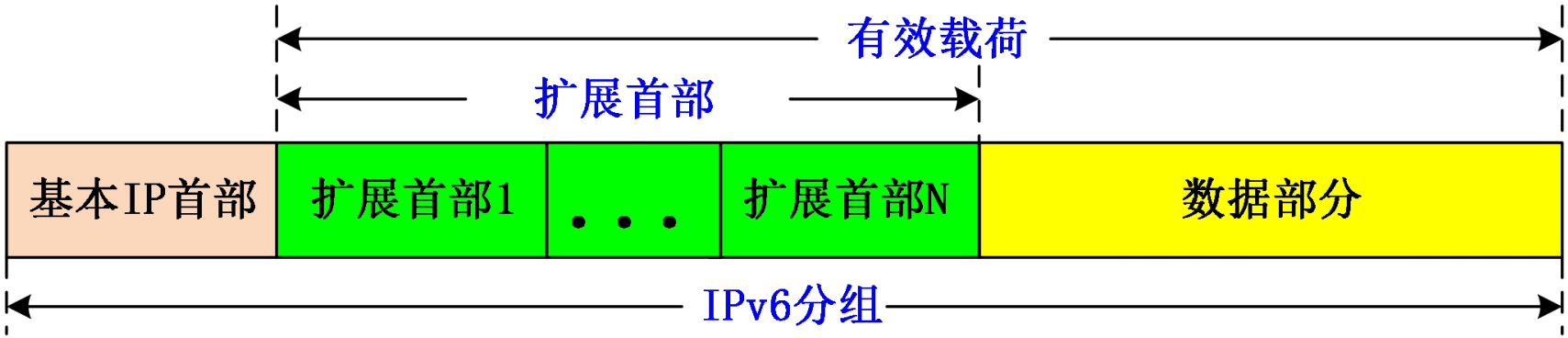
- 位置：在IPv6基本头标和有效载荷之间
- 特点（与IPv4的选项字段比较）
  - 灵活、高效：只在需要时才插入
  - 可扩展性好：可以根据需要定义新的扩展头标
- 处理位置：**路由器、目的节点**
- 问题
  - 路由器需要查看每个扩展头标吗？
  - 如何能够做到让路由器高效地选择出需要其处理的扩展头标

扩展头标

- 关于IPv6扩展头的几点说明：
  - 扩展包头必须严格按出现顺序处理，目的结点不能搜索某一特定的扩展头并对之优先处理
  - 如果要处理的下一个包头的类型不认识或0则返回ICMP（code 1）并丢弃包
  - 为了字边界对齐，每个扩展报头的长度是必须是8字节的整数倍

扩展头标

- 基本组合方式
  - IP基本头标+数据
  - IP基本头标+1个扩展头标+数据
  - IP基本头标+n个扩展头标+数据
  - IP基本头标+扩展头标



扩展头标

路由器处理

目的节点处理

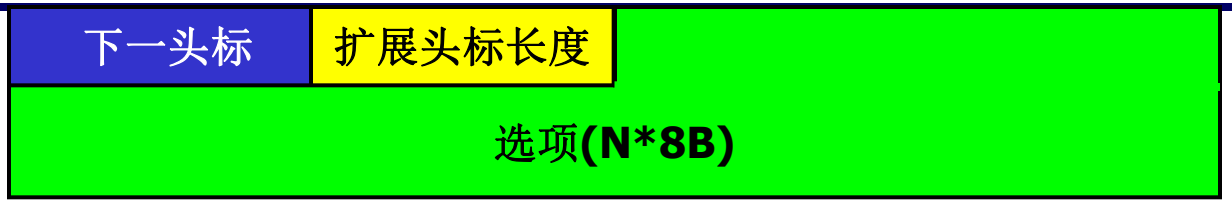
出现的顺序	首部名称
1	IPv6基本首部
2	Hop-by-hop 逐跳选项扩展首部
3	Destination目的选项扩展首部1(由首部中指定的网络节点依次进行处理)
4	routing扩展首部
5	fragment分片扩展首部
6	AH身份认证扩展首部
7	ESP封装安全净荷
8	Destination目的选项扩展首部2(仅由目的节点进行处理)
9	Mobility移动扩展首部
最后	无下一头标
最后	UDP,TCP，ICMP以及其他高层协议首部

逐跳选项扩展首部

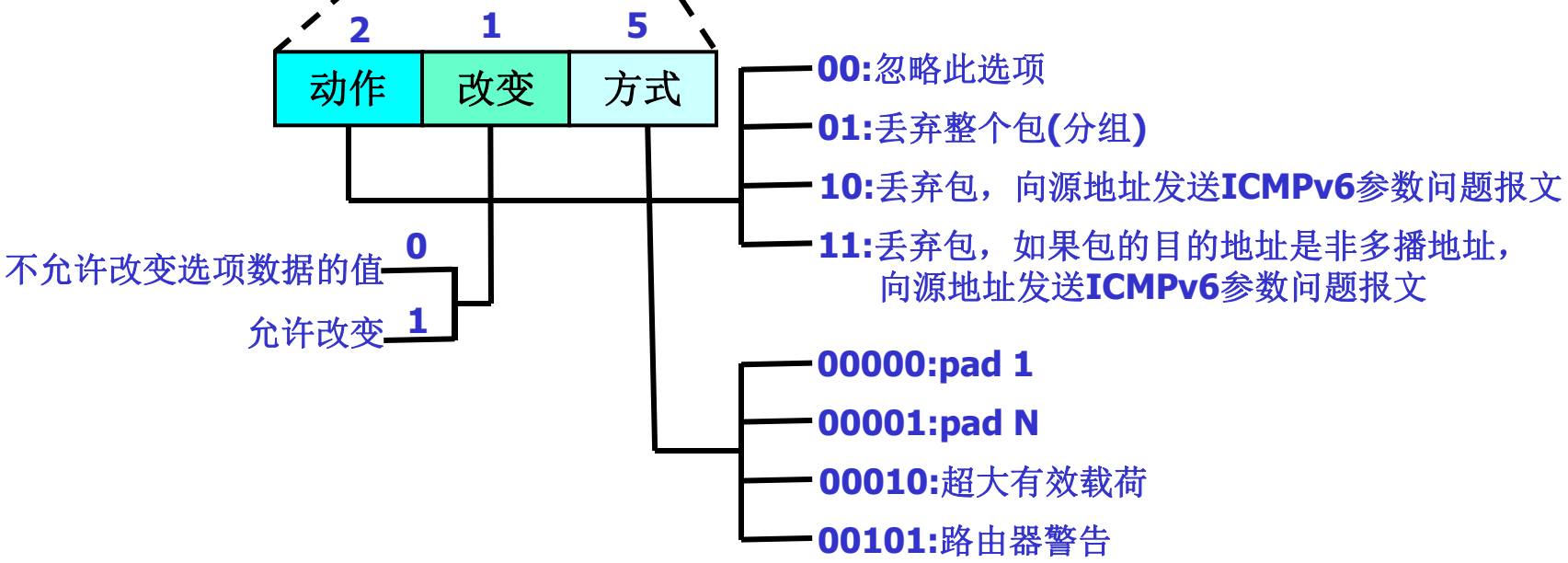
- Hop-by-hop Options Header
- 作用：描述了数据分组转发的特性
- 处理位置：从源节点到目的节点的路由上的每一个节点(即路由器)
- 说明：
  - 除了逐跳选项扩展首部，其余扩展头部与上层协议一样是根据目标地址判断是否需要解析处理。
  - 逐跳选项首部在沿途路由器上被无条件解析处理

逐跳选项扩展首部

- 格式



选项格式(TLV结构)



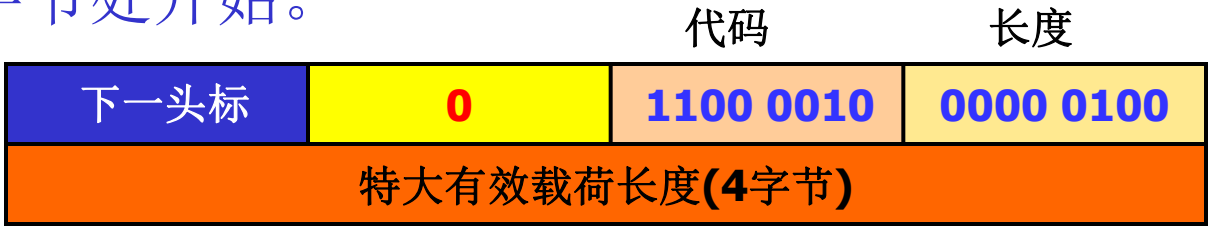
逐跳选项扩展首部

- 逐跳选项首部决定了数据包内容一定会被沿途的路由器处理
- 逐跳选项首部的选项类型决定了数据包如何被处理。
- Pad 1的选项
  - 用于边界对齐，插入一个填充字节
  - 1字节，格式：0000 0000
- Pad N的选项结构
  - 用于边界对齐，插入2个或多个填充字节
  - N字节
  - 格式

N个字节的填充

逐跳选项扩展首部

- 特大有效载荷（jumbo payload）
  - 作用：IP数据报的载荷长度超过65535字节时使用
  - 选项从(4n+2)字节处开始。
- 结构



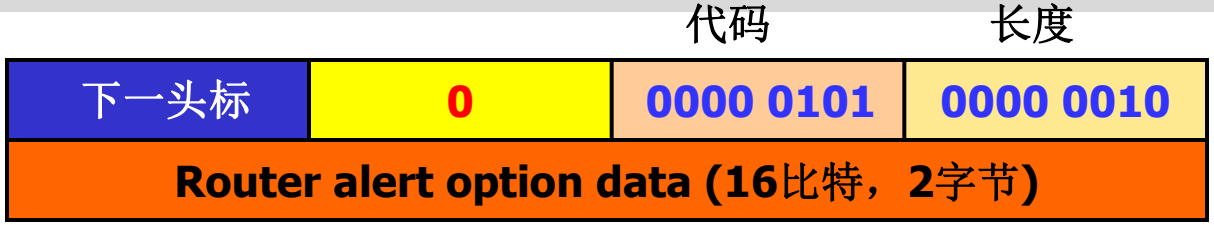
- 能表示的最大IP分组长度JPL：$65535 \le JPL < 2^{32}$
  - 该长度不包含IPv6基本首部
  - 包含逐跳选项扩展首部在内的字节数
  - 使用该选项，则IPv6基本头中的有效载荷字段设置为0
- 只有沿途每个路由器都能处理时才可使用该选项
- 如果使用了分片扩展首部，则hop-by-hop选项扩展首部中不能包含特大有效载荷选项

逐跳选项扩展首部

- 路由器警告(警示)选项（router alert option）
  - 用于告知路由器该IPv6分组中的内容需要进行特殊的处理，用于RSVP、MLD(Multicast Listener Discovery Protocol)等

This memo describes a new IP Option type that alerts transit routers to more closely examine the contents of an IP packet. This is useful for, but not limited to, new protocols that are addressed to adestination but require relatively complex processing in routers along the path.

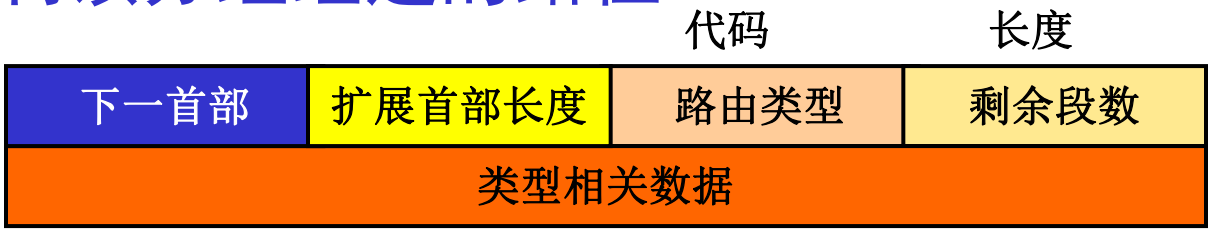
- 格式



  - Router处理行为：被应用层进程在用户态处理，将剩余跳数递减后重新注入协议栈，继续转发

路由扩展首部

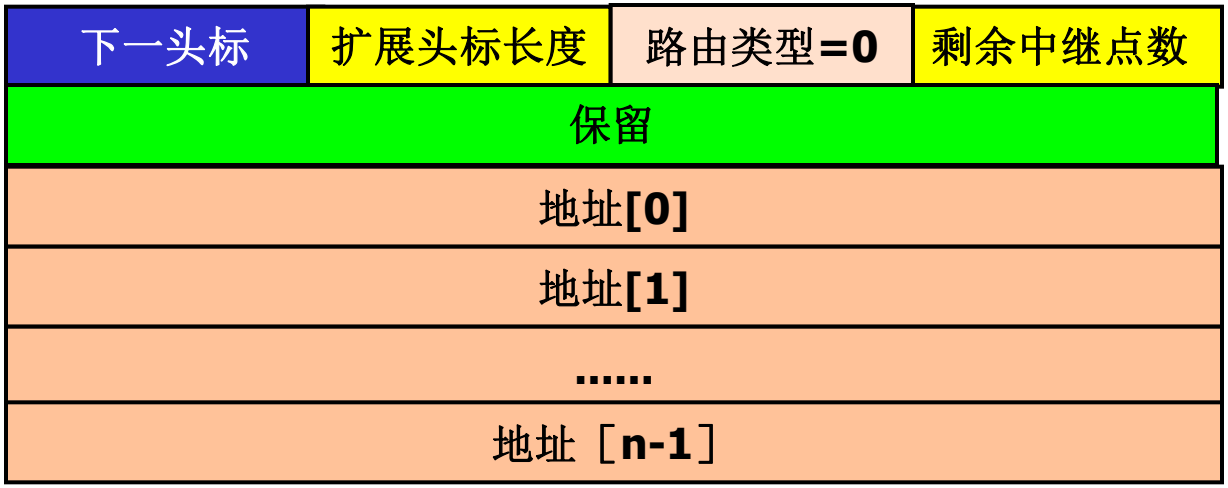
- Routing Header, RH
- 作用：控制路径，用来指出IPv6分组在从源节点到目的节点的过程中需要经过的一个或多个路由器
- 应用场景：信源将分组发往信宿时，在某些情况下希望控制该分组经过的路径。
- 格式



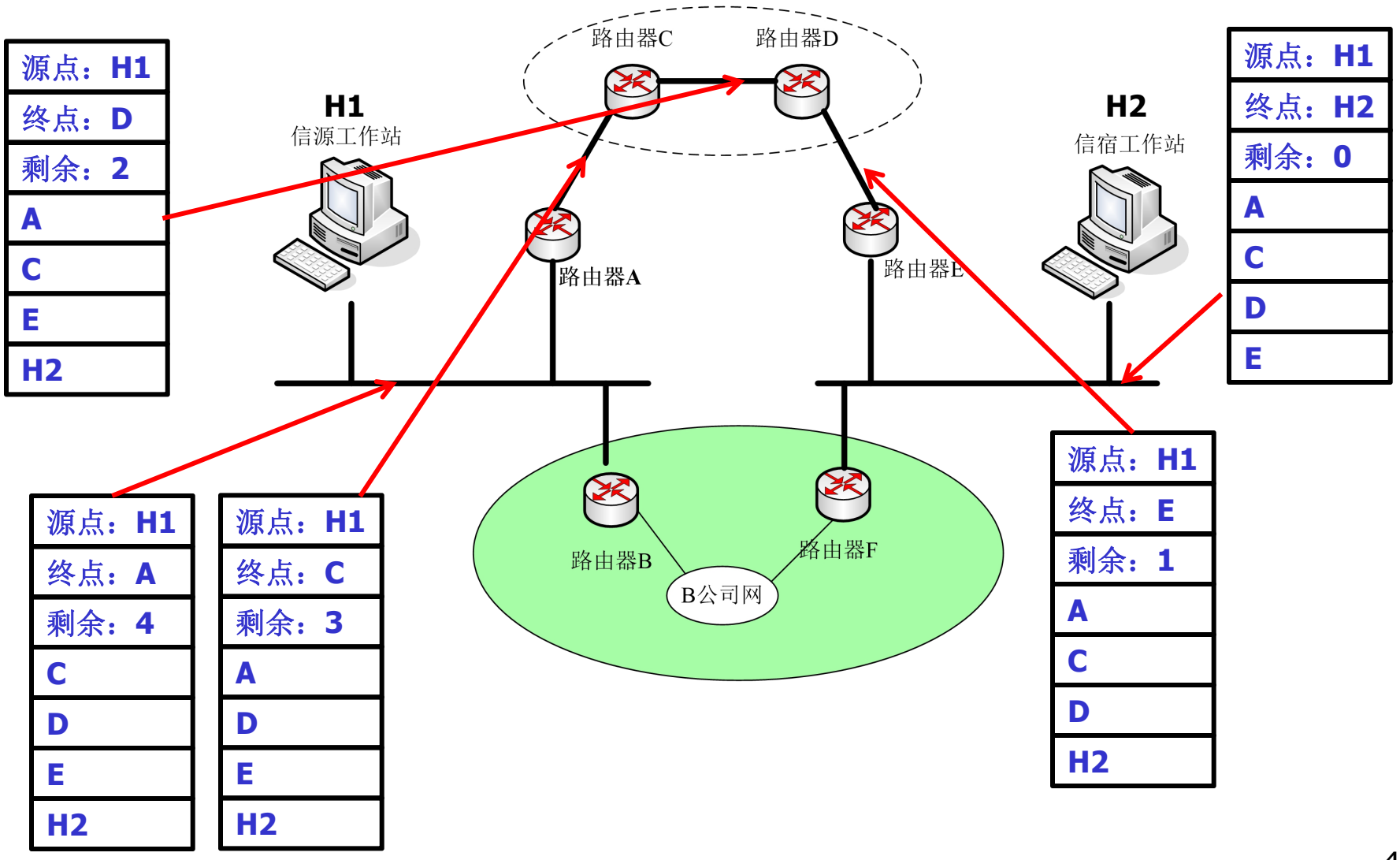
- 路由类型
  - 类型0：支持源站选路
  - 类型2：支持IPv6移动性

路由扩展首部

- 路由类型0的路由扩展首部



路由扩展首部应用举例


源点：H1
终点：D
剩余：2
A
C
E
H2
H1
信源工作站
路由器C
路由器D
路由器A
路由器E
H2
信宿工作站
源点：H1
终点：H2
剩余：0
A
C
D
E
源点：H1
终点：A
剩余：4
C
D
E
H2
源点：H1
终点：C
剩余：3
A
D
E
H2
路由器B
路由器F
B公司网
源点：H1
终点：E
剩余：1
A
C
D
H2

路由扩展首部

STEP1：信源发出分组时，基本头标的目的地址是预定路径上第一个中继点地址，沿该路径的各HOP地址依次列于寻路头标地址表内，剩余中继点数为地址表中地址的总数

STEP2：中间节点需要改变的域

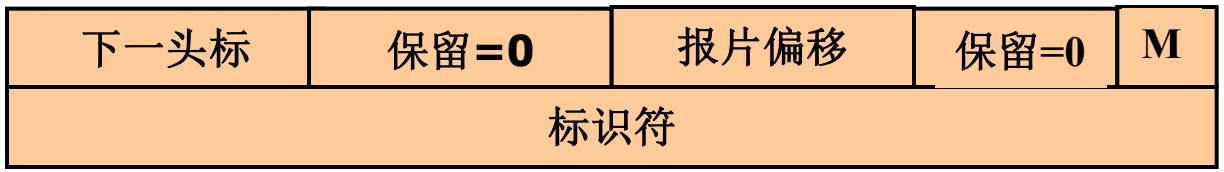
- 基本头标
  - 中继点限制（减1）
  - 目的站IP地址（地址表中的一个地址）
- 扩展头标
  - 剩余中继点数（SI=SL-1）
  - 地址表中的某个地址（与基本头标中目的地址对调）

路由扩展首部

- 相关讨论
  - 效率高：只有目的地址指示的路由器处理路由扩展首部，其他中间路由器不处理
  - 路由类型字段不可识别时：
    - 剩余字段数≠0，忽略，继续处理下一个首部
    - 剩余字段数=0，丢弃并发送ICMPv6报参数错误
  - 路由类型=0时，目的地址不能为多播地址

分片扩展首部

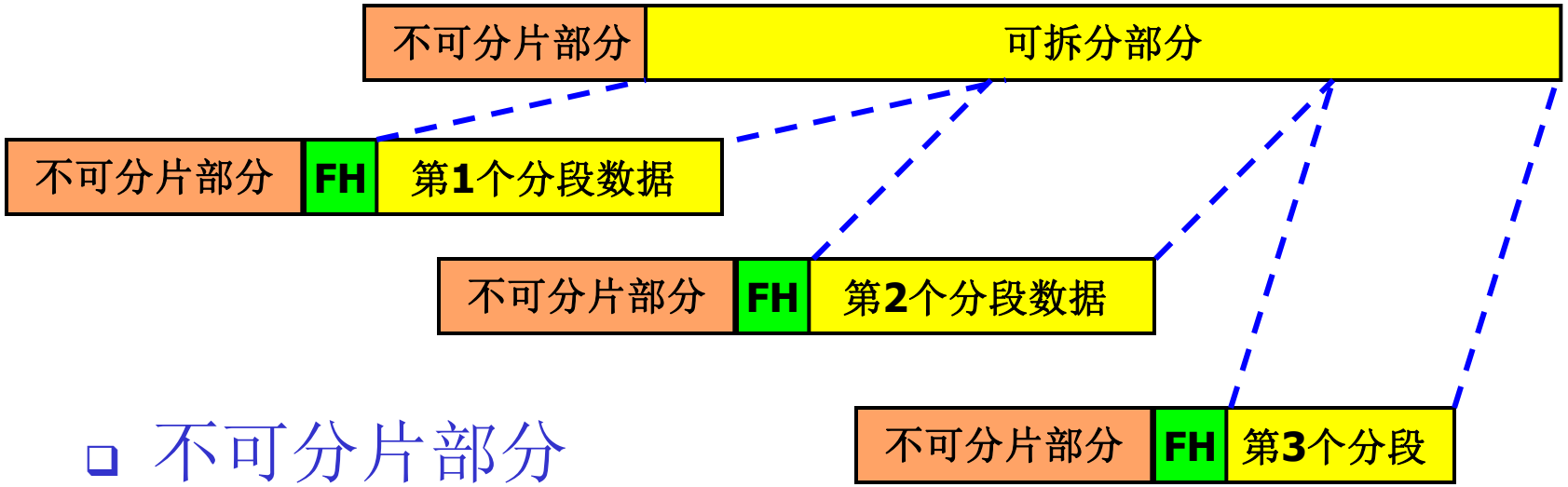
- Fragment Header: FH
- 作用：数据报分片（头标类型：44）用于将大于路径MTU的信息包从源节点发送到目的节点
- 相关技术：路径MTU发现技术
- 格式（8字节）



  - M(ore)：最后的报片置为“0”，其余报片置为“1”；
  - 报片偏移(13bit)：以64比特为单位（即8字节的整数倍），最大值=$2^{13}-1=8191*8=65528$字节
  - 标识符(32bit)：唯一标识最近(在分组的生存期内)从源地址发现目的地址的分组

分片扩展首部

- 原始IPv6数据报：未被分片的分组



- 不可分片部分
  - IPv6分组基本首部
  - 需要路由器处理的扩展首部
    - 逐跳选项扩展首部
    - 目的选项扩展首部（放在路由选项之前的）
    - 路由扩展首部

分片扩展首部

- 可拆分部分
  - 有效载荷
  - 只需要目的节点处理的扩展首部
- 重组
  - 具有相同源地址、目的地址和分片标识符
  - 不可分片部分=第1个分片分组的不可分片部分
  - 下一头标：第1个分片分组的分片扩展首部的下一首部字段

- 原始IPv6数据报
- 路径MUT=1500B

IPv6基本头(40字节) + IP数据(3960字节)
净荷长度=3960
下一头标=17(UDP协议)

第1片：IPv6基本头(40字节)+分片扩展首部(8字节)+IP数据(1448字节)
净荷长度=1456　　下一头标=17
下一头标=44　　偏移量=0
M=1
标识符=1234567

第2片：IPv6基本头(40字节)+分片扩展首部(8字节)+IP数据(1448字节)
净荷长度=1456　　下一头标=17
下一头标=44　　偏移量=181
M=1
标识符=1234567

第3片：IPv6基本头(40字节)+分片扩展首部(8字节)+IP数据(1064字节)
净荷长度=1072　　下一头标=17
下一头标=44　　偏移量=362
M=0
标识符=1234567

分片扩展首部

- 路径MTU
  - 作用：为了传送大于路径MTU的信息包，节点可使用IPv6分段报头，在源节点将信息包分段，而在目的节点将信息包重装配。
  - 特点：IPv6 的分段处理不同于IPv4，IPv6仅在源节点通过扩展报头中的分段报头进行分段处理，简化了中间节点对分组的处理。
  - 思考问题：主机如何发现路径的MTU?
    - **路径MTU发现机制**

分片扩展首部

- 异常情况处理
  - 60s内没有收到全部分片，则终止重组并报错
  - M=1且数据部分长度非8字节整数倍，丢弃并报错
  - 重组后有效载荷长度>65535字节，丢弃并报错
  - MTU<1280字节的数据包为非法
- Fragment attack
  - Ping of Death---发送大于65536字节的ICMP包使操作系统崩溃
  - Teardown attack---偏移字段设置成不正确的值(DoS攻击)
  - Tiny fragment attack---发送设计过的分片来绕过防火墙等包过滤系统或者入侵检测系统

目的选项扩展首部

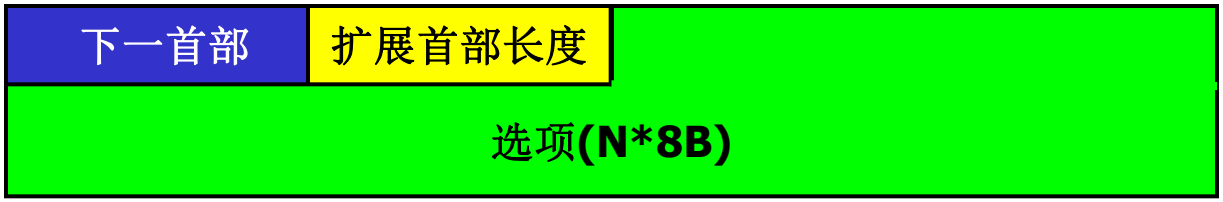
- Destination Option Header, DOH
- 问题的引出：如何对IPv6增加新的功能？
- 常用方法
  - 定义一个新扩展头标，该头标仅由目的地址标识的主机来处理；
  - 不分配新的头标类型，仅定义一个通用的、自由度高的由目的地主机处理的扩展头标
- 目的选项扩展首部作用
  - 携带只需要目的站点检验的可选信息(便于用户增加新IP层功能)，为中间节点或目的节点指定分组的转发参数

目的选项扩展首部

- 使用方式
  - IF(存在路由首部AND目的选项在路由首部前)，THEN 目的选项指定中间节点(Router)均需要转发或处理的选项
  - ELSE(即不存在路由首部OR目的选项在路由首部后)目的选项指定目的节点处理的选项
- 为什么要定义两个位置？
  - 在某些情况下(例如使用路由头部)，当数据报被转发到最终目的地时，IPv6头部中的目的IP地址字段将会改变。

目的选项扩展首部

- 格式

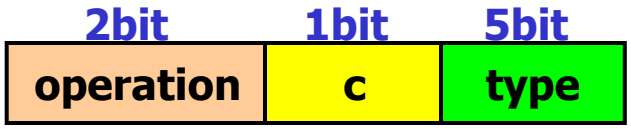


  - 下一首部（8bit）
  - 扩展首部长度（8bit）
  - 选项（变长）

目的选项扩展首部

- 选项

  - 选项类型T



    - 动作(Operation)：指明处理节点不能识别选项时的操作
      - 00：忽略，继续处理下一选项
      - 01：丢弃IP分组，不回送ICMP报文
      - 10：丢弃IP分组，回送ICMP差错报文
      - 11：丢弃IP分组，如果目的地址不是组播地址，就回送ICMP报文（目的信息作为一个单独的扩展首部是使用)
    - 改变(C)：选项在传输的路径上是否改变（1：改变；0：不改变）

目的选项扩展首部

- 类型0目的选项扩展首部
  - 填充1选项
  - 填充N选项
- 类型2目的选项扩展首部
  - MIPv6章节介绍

选项扩展首部

选项名	首部	动作	改变C	类型(8bit)	长度(Byte)	RFC
Pad 1	H,D	00	0	0	N	RFC2460
Pad N	H,D	00	0	1	可变	RFC2460
超大有效载荷	H	11	0	194	4	RFC2675
路由器警告	H	00	0	5	2	RFC2711
家乡地址	D	11	0	201	16	RFC6275

IPv6扩展首部与IPv4选项的比较

IPv4中的情况	IPv6中的情况
无操作和选项结束选项	Pad1, Pad N(Hop by Hop)
记录路由	无
时间戳	无
源路由(严格、松散)	路由扩展首部(Routing)
基本头中的分片字段	分片扩展首部(Fragment)
无	认证首部AH
无	封装安全净荷ESP

关于定义新的扩展首部和选项

- RFC8200（2017）
  - **不推荐**定义新类型的IPv6扩展首部，除非有充分的证据说明现存的IPv6扩展首部(通过增加新选项等的方法)无法实现需要的功能
  - **禁止**定义新的具有hop-by-hop行为的扩展首部
  - **不推荐**定义hop-by-hop新选项
  - **推荐**利用目的选项扩展首部实现更多可选功能

IPv6与IPv4的比较

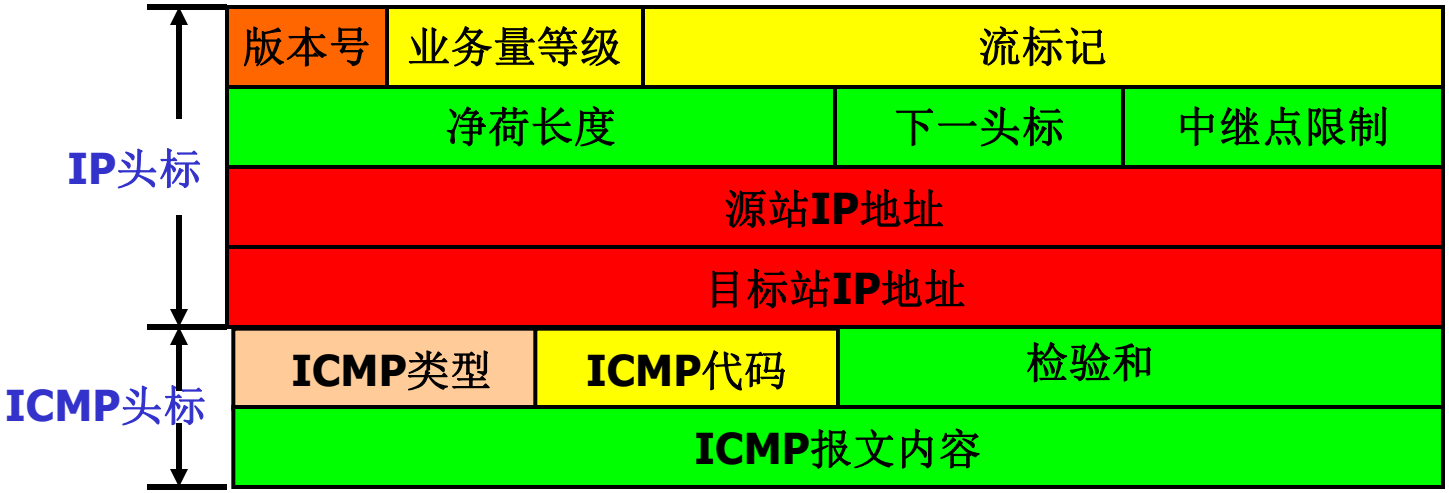
- IPv6中，要求所封装的UDP首部中必须有校验和字段，而IPv4封装的UDP首部中校验和是可选的
- IPv6要求每条链路最小MTU为1280字节，IPv4的链路最小MTU为68字节.
- IPv6的分组最大长度是2^{32}，IPv4为2^{16}.

IPv6协议与相邻协议关系

- 上层协议和计算
  - IPv6中，UDP的校验和是必需的，IPv4是可选的
  - TCP、UDP和ICMPv6的校验和采用伪头标校验

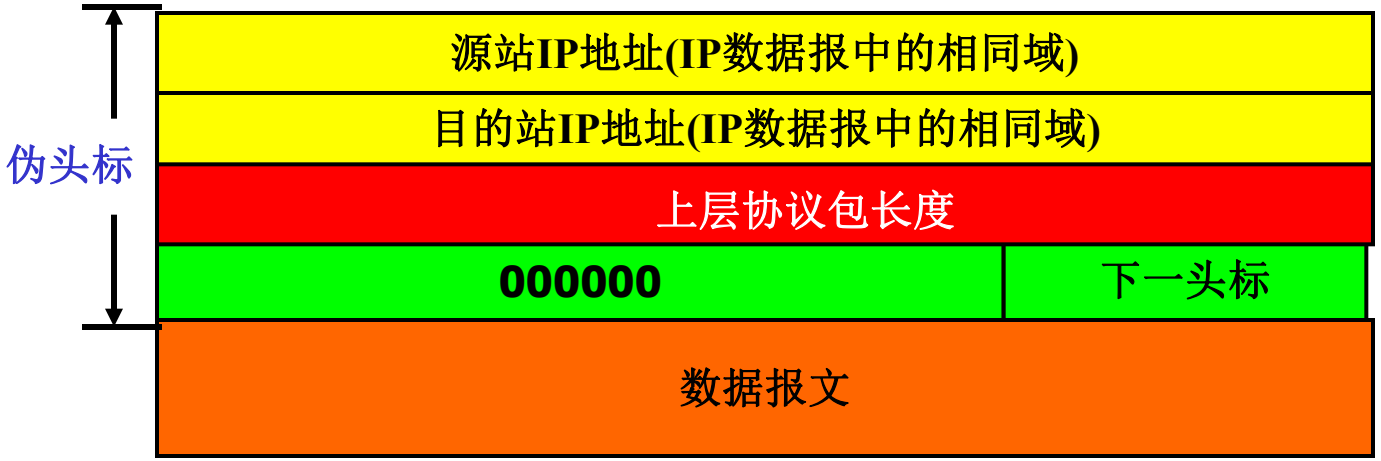
伪头标校验

- 关于数据的校验问题
  - IP基本头标无校验
  - IP包中的数据部分的报文格式中如果有校验和域，则该校验和的计算需要使用伪头标

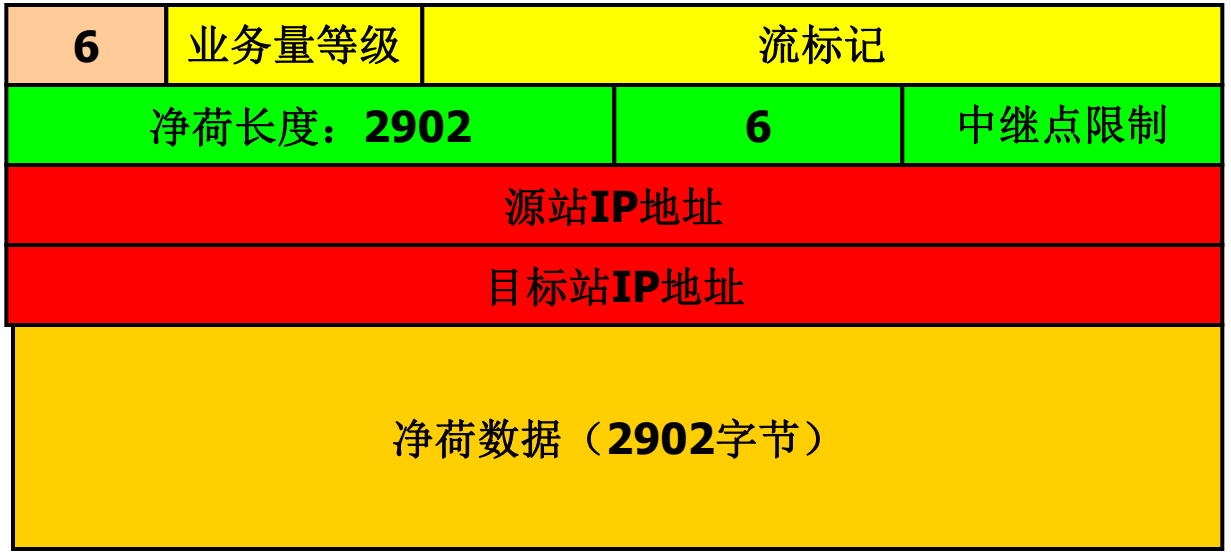


伪头标校验

- IP基本头标中的关键数据
  - 信源地址
  - 信宿地址
  - 下一头标
  - 净荷长度



IPv6基本首部



谢　谢！