下一代Internet技术与协议

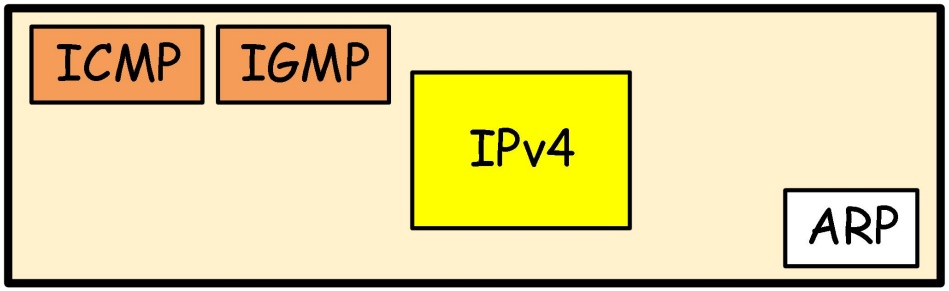
张冬梅
北京邮电大学 计算机学院
zhangdm@bupt.edu.cn

4.1 ICMP协议概述

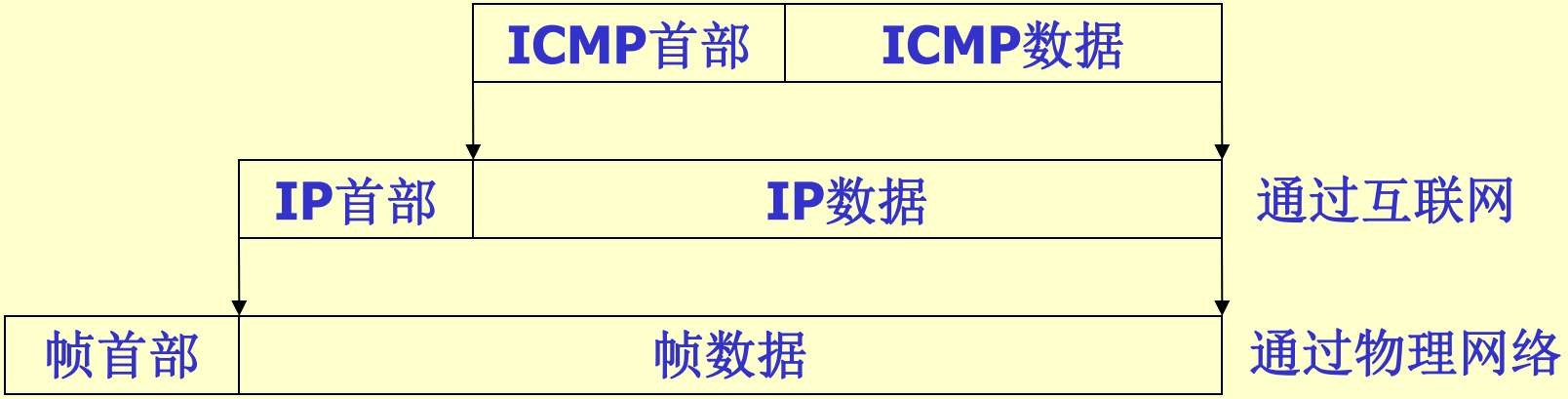
- 4.1.1 ICMP协议背景与功能
- 4.1.2 ICMPv4概述
- 4.1.3 ICMPv6格式与功能

4.1.1 协议背景与功能

- 协议背景
- 在协议栈位置
- IP与ICMP关系
  - IP与ICMP互相依赖
  - IP在发送一个差错报文时用到ICMP
  - ICMP用IP来封装传递报文
- ICMP功能
  - 使发送方了解为什么数据报无法投递(差错报告与诊断)
  - 管理查询（系统间调整）



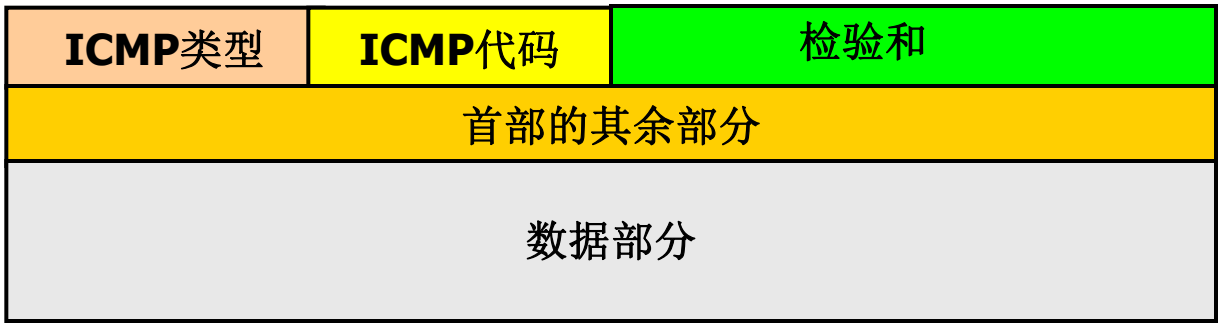
- ICMP报文的投递
  - 在一个IP数据报的数据部分通过互联网传送
  - 两级封装



  - 问题：为什么用IP进行封装？

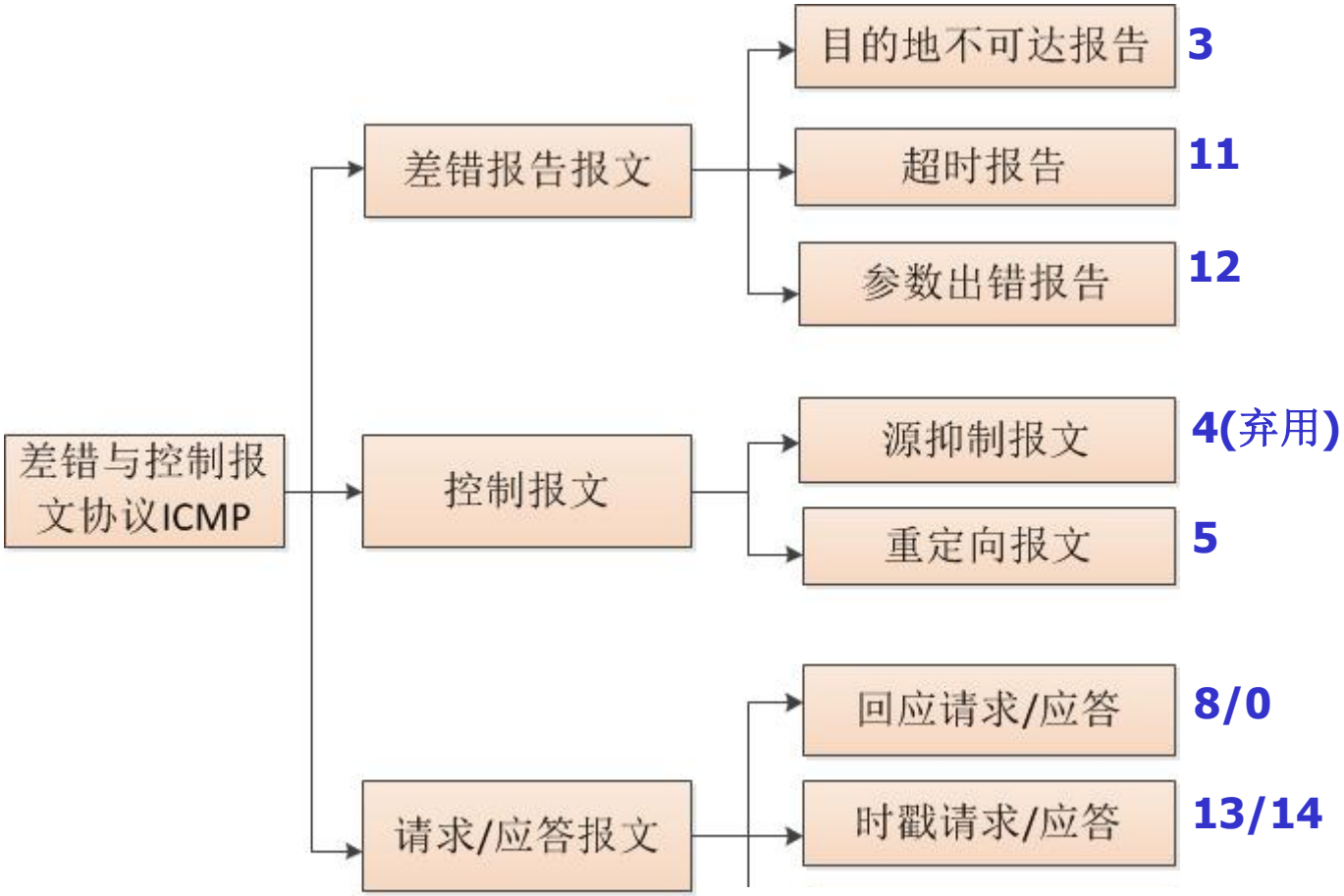
4.1.2 ICMPv4报文格式

- ICMP格式(8字节首部+可变长数据)



  - ICMP类型
  - ICMP代码：提供有关报文类型的进一步信息
  - 校验和: 包括整个报文，Checksum(见附录1)。
  - 数据：差错报文中，携带用于找出引起差错的原始分组的信息；查询报文中，携带的是基于查询类型的额外信息

ICMPv4报文类型


差错与控制报文协议ICMP
差错报告报文
控制报文
请求/应答报文
目的地不可达报告
3
超时报告
11
参数出错报告
12
源抑制报文
4(弃用)
重定向报文
5
回应请求/应答
8/0
时戳请求/应答
13/14

ICMP报文地址

- 信源IP地址的选择
  - 如果系统只对应一个接口的一个IP地址，则用此地址即可
  - 如果系统对应多个接口的多个IP地址，则按规则选取
    - case1：ICMP应答报文
      - 如果源报文的目的地址是单播地址，则信源地址为原报文的目的IP地址（问谁谁回答）；
      - 如果源报文的目的地址是组播或任播地址，则信源地址为收到原报文的接口的IP地址（谁收到谁回答）；

- case2：对于ICMP差错报文，则信源地址为报告出错信息的接口的IP地址；
- case3：对于其他ICMP报文，则信源地址为发送报文的链路的IP地址(主动发送的、以及不适用上述规则的)

- 信宿IP地址的选择
  - 提示：不同类型、不同用途的ICMP报文的信宿地址的选取规则不同

ICMPv4的几种使用

- 差错报告报文
- 控制报文
  - 拥塞控制与源抑制报文
  - 路由控制与重定向报文
- 请求/应答报文
  - 回应请求/应答
  - 时戳请求/应答

差错报告

- 功能

  ICMP对IP分组出现的差错进行报告，只报告错误，不纠错

- 把差错报文报告给最初的数据源
- 相关报文
  - 目的地不可达
  - 分组超时
  - 参数问题

差错报告

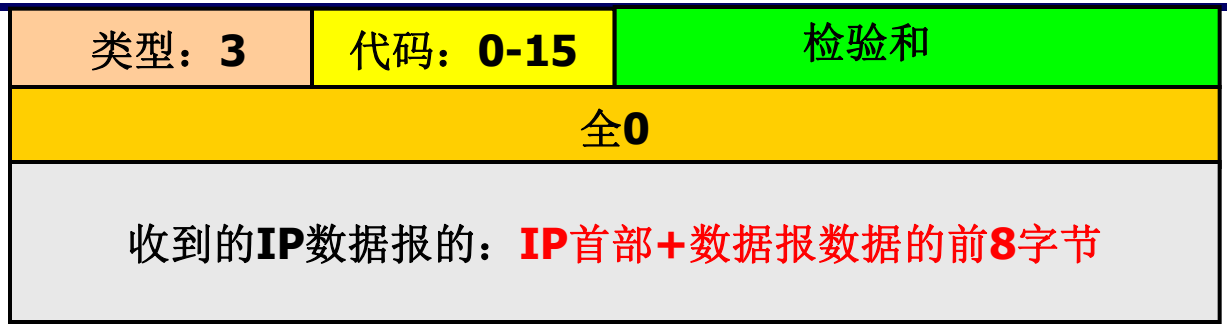
- 关于ICMP差错报文的要点
  - 对于携带ICMP差错报文的数据报，不再产生ICMP差错报文
  - 对于分片的数据报，如果不是第一个分片，则不产生ICMP差错报文
  - 对于具有多播地址的数据报，不产生ICMP差错报文
  - 对于具有特殊地址(127.0.0.0或0.0.0.0)的数据报，不产生ICMP差错报文

差错报告

- 差错报文的数据部分包含
  - 原始数据报的IP首部
  - 数据报数据的前8个字节

终点不可达

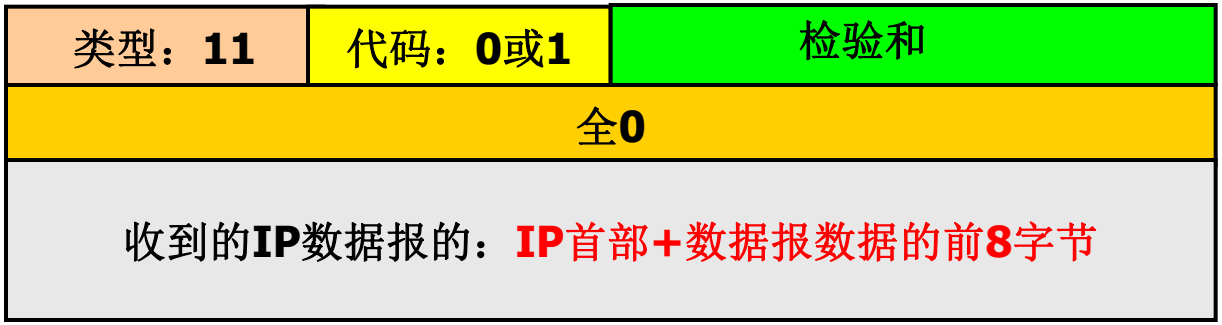
- 格式



代码	含义	代码	含义
0	网络不可达	8	源主机被隔离
1	主机不可达	9	管理上禁止与目的网络通信
2	协议不可达	10	管理上禁止与目的主机通信
3	端口不可达	11	对指明的服务类型，网络不可达
4	需要进行分片	12	对指明的服务类型，主机不可达
5	源路由不能完成	13	主机不可达，设置了过滤器
6	目的网络未知	14	主机不可达，违反了优先级策略
7	目的主机未知	15	主机不可达，优先级被截止

超时报文

- 两种情况
  - 报文超时(0)：TTL字段递减后为0，丢弃数据报，并发送超时报文
  - 重组超时(1)：终点在规定的时间内未收全全部报文分片，则丢弃已经收到的分片，并发送超时报文
- 格式



参数问题

- 格式

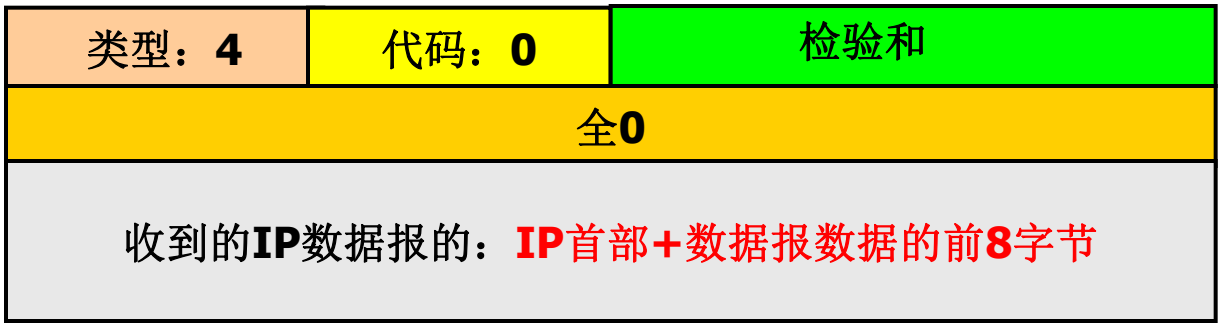
  - Code=0: 首部的某个字段有差错或二义性，指针字段值指向有问题的字节
  - Code=1：缺少所需要的选项部分。不使用指针。

ICMP控制报文—源抑制

- 功能：拥塞控制
- 条件：当路由器接收IP数据报的速度比其处理IP数据报的速度快，或者路由器传入IP数据报的速度大于其传出IP数据报的速度时，会产生拥塞现象
- 措施：路由器通过发送源站抑制报文(Source Quench)来抑制源主机发送IP数据报的速率，进而避免可能产生的拥塞和差错

源抑制报文

- 格式



- 作用
  - 通知源点，数据报已经被丢弃
  - 警告源点，在路径中的某处出现了拥塞，源点需要放慢(抑制)发送过程。
- 说明
  - 对每一个因拥塞而被丢弃的数据报，发送一个源抑制报文
  - 没有机制告诉源点拥塞得到缓解
  - 多对一通信不一定有效果

源抑制报文

- 利用源抑制报文进行拥塞控制的过程
  - 路由器发生拥塞时发出ICMP源抑制报文
    - 拥塞判别方法

    (1) 检查路由器缓存是否已满

    (2) 缓存区输出队列设置一个阀值，判断队列中数据报个数是否超过阀值

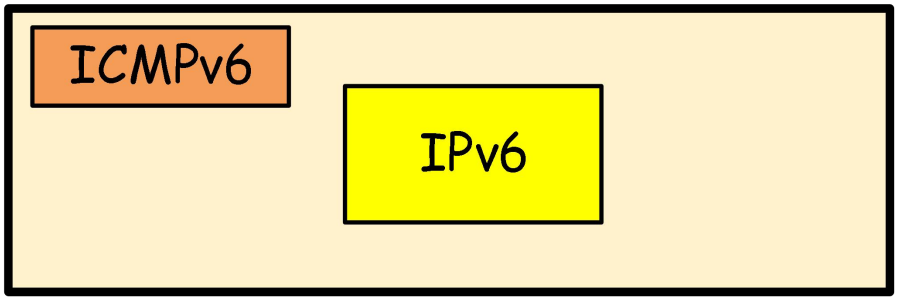
    (3) 检测某输入线路的传输率是否过高
  - 源主机收到抑制报文后按一定的速率降低发往目标主机的数据报传输率
  - 如果在一定的时间间隔内源主机没有再收到抑制报文，便认为拥塞已经解除，源主机可逐渐恢复到原来的发送速率

源抑制报文

- 相关说明
  - 对网络的拥塞控制效率低且不公平
  - 1995年，RFC1812正式禁止路由器产生和转发源抑制消息
  - 2012年，RFC6633正式宣布传输层协议不再对ICMP源抑制消息做出响应

ICMPv6概述

- ICMPv6与IPv6一起工作， IPv6网络中每一个节点均要实现ICMPv6

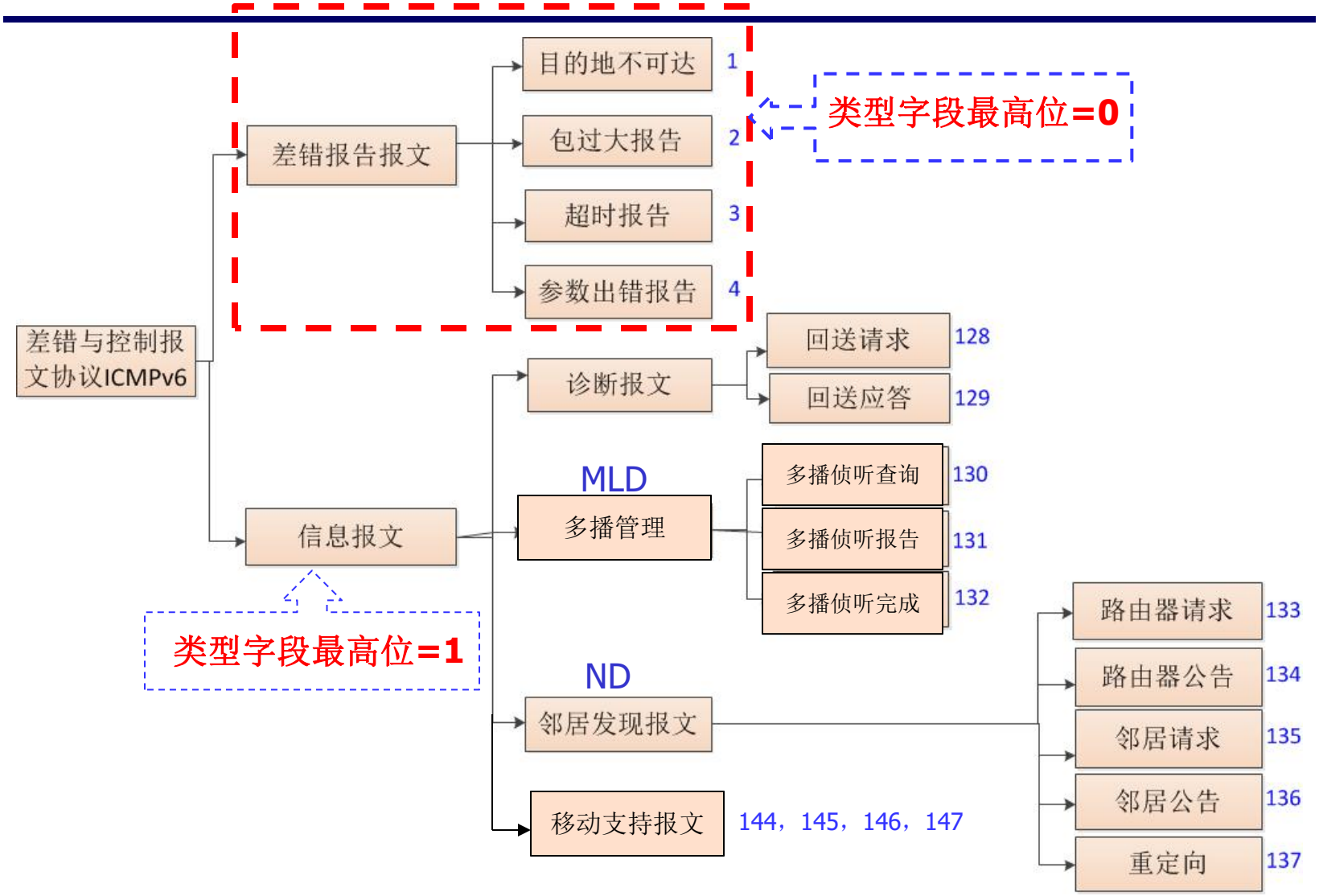


- 功能
  - 当IPv6分组不能被正确处理时，ICMPv6向源节点报告IPv6分组在传输过程中的出错信息和通告信息，使网络节点知道网络状态。
  - ICMPv6实现了ICMPv4、ARP(邻居发现ND)、IGMP(组播侦听者发现MLD)等协议的功能，并增加了对移动IPv6的支持,还包括安全邻居发现(SEND)协议。

ICMPv6消息类型

- 错误类消息
  - 报告IPv6数据包转发或传输过程的错误
  - 目的节点或中间路由器报告
  - 四类错误消息，消息类型取值0-127
- 信息类消息
  - 提供诊断功能和其他主机功能
  - 消息类型取值128-255

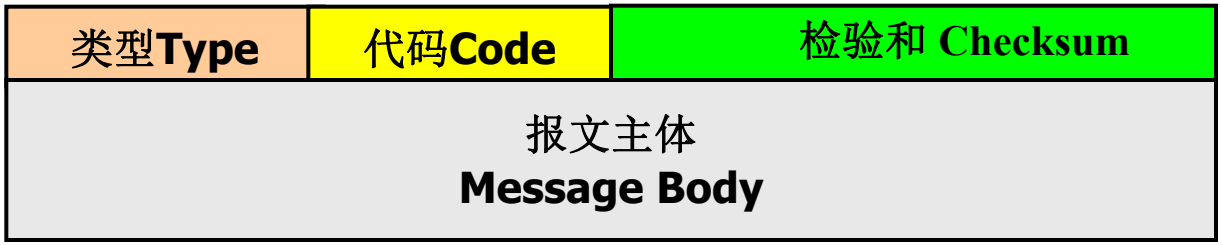
ICMPv6消息类型


差错与控制报文协议ICMPv6
差错报告报文
目的地不可达
1
包过大报告
2
超时报告
3
参数出错报告
4
类型字段最高位=0
信息报文
类型字段最高位=1
诊断报文
回送请求
128
回送应答
129
MLD
多播管理
多播侦听查询
130
多播侦听报告
131
多播侦听完成
132
ND
邻居发现报文
路由器请求
133
路由器公告
134
邻居请求
135
邻居公告
136
重定向
137
移动支持报文
144，145，146，147

报文名称	ICMPv4类型	ICMPv6类型
目的地不可达	3	1
协议包过大	类型3代码4	2
源抑制	4	无
重定向	5	137
回声请求/应答	8/0	128/129
超时	11	3
参数错误	12	4
时间戳/时间戳回复	13/14	无
路由器请求RS/公告RA	10/9	133/134
邻居请求NS/公告	ARP	135/136
家乡代理地址发现请求/公告	无	144/145
移动前缀请求/公告	无	146/147
组成员管理	IGMP	130,131,132

ICMPv6协议格式

- ICMP分组格式
  - ICMPv6首部+ICMPv6报文主体
  - IPv6首部中：下一个首部=58



  - 类型
    - 最高位为0：差错报文;最高位为1：查询报文
  - 代码：区分特定消息类型中的多个子类型，如果只有一个，则code=0
  - 校验和：伪头标校验

ICMPv6报文处理规则

- 当接收到ICMPv6差错报告报文时，如果无法识别具体的类型，必须将它交给上层协议模块进行处理；
- 当接收到ICMPv6信息报文时，如果无法识别具体的类型，将它丢弃；
- 所有的ICMPv6报文，都应该在IPv6所要求的最小MTU允许范围内，尽可能多地包括引发该ICMPv6差错报文的IPv6分组片段，以便给IPv6分组的源节点提供尽可能多的诊断信息

ICMPv6报文处理规则

- 不能产生ICMPv6差错报告报文的发送情况
  - 一个ICMPv6差错报告报文
  - 一个发往IPv6多播地址（或链路层多播地址）的分组
    - 例外情况：分组过大报文
  - IPv6分组的源地址无法唯一确定一个单独节点时，这种情况不能够引起ICMPv6差错报告报文的发送。

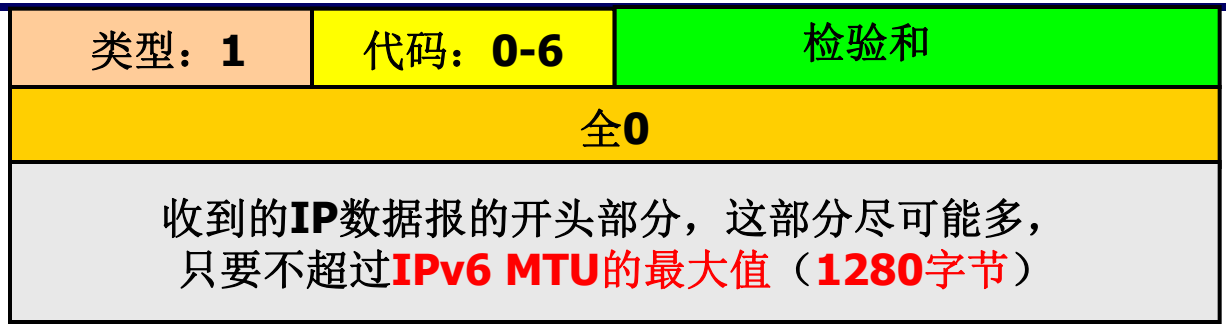
- IPv6节点必须限制其发送ICMPv6差错报文的速率。目前限制ICMPv6速率的方法：
  - 基于计时器的方法：T时间内只发送一个报文
  - 基于带宽的方法：ICMPv6差控报文占链路带宽的某个比例F

ICMPv6的几种使用

- 差错报告
  - 目的地不可达(Type=1)
  - 数据包过大(Type=2)
  - 超时(Type=3)
  - 参数问题(Type=4)
- 邻机发现：为了确定同一个链路上的邻居的链路层地址、发现路由器、随时跟踪哪些邻居可连接，以及检测更改的链路层地址。
- 组管理

目的地不可达

- 格式



- 代码

	含义		含义
0	没有路径到达目的地(R)	4	端口不可达(H)
1	与目的地的通信被禁止(R)	5	源地址未通过出入策略检查(R)
2	超出源地址的范围	6	拒绝路由到达目的地址
3	目的地址不可达(R)		

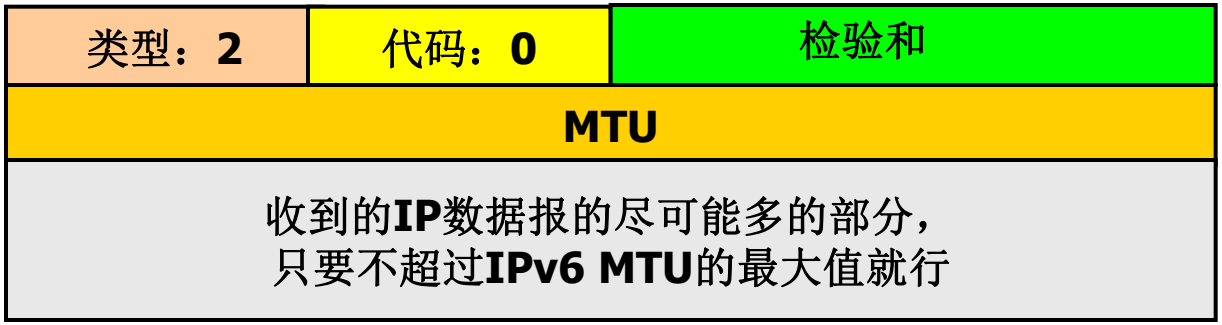
- 问题1：对于某个不包含扩展首部的IPv6数据包，原负载中有多少字节会在ICMPv6目的地址不可达消息中？

1280-40(IPv6首部)-8(ICMPv6首部)-40(原始包IPv6首部)=1192字节

- 问题2：哪些目的地不可达消息有Router发送，哪些由目的主机发送？

分组过大报文

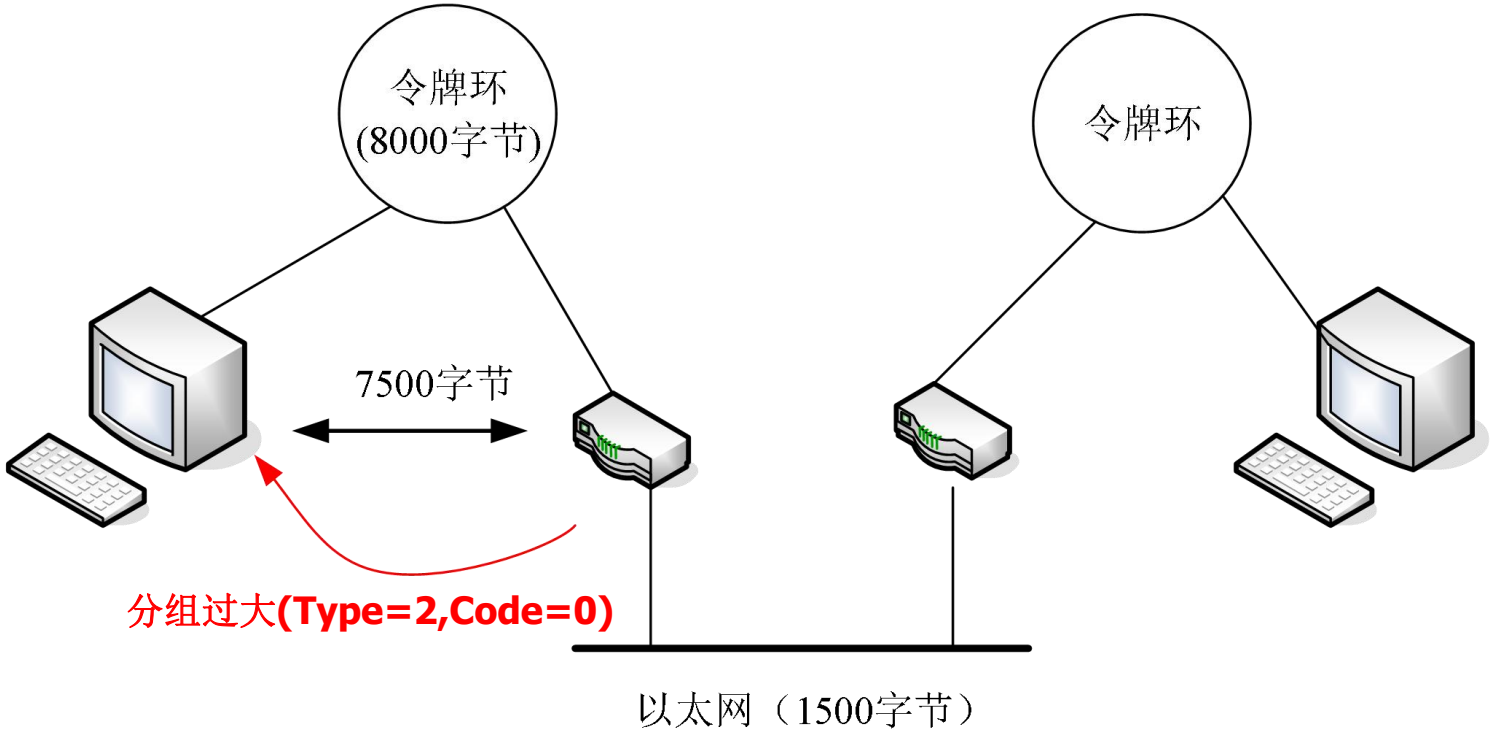
- 格式



- 主机收到该报文后必须通知上层进程
- IPv6路径MTU发现

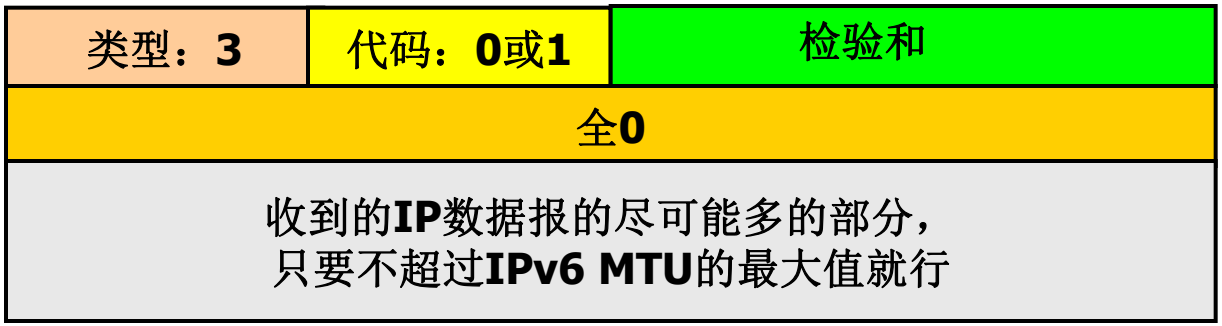
分组过大报文

- 工作过程举例



超时报文

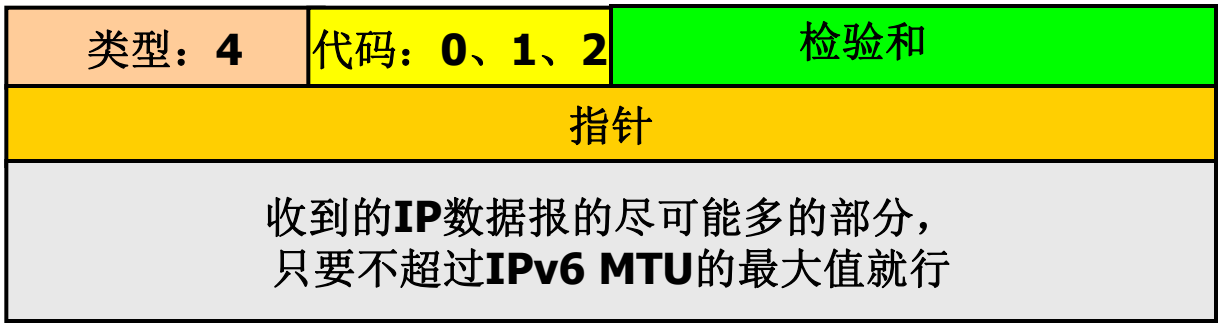
- 格式



  - Code=0：超出了跳数限制，路由器丢弃并报错
  - Code=1：分组重组超时，目的主机丢弃并报错（RFC2460指定为60s）
- 应用：路由跟踪（Traceroute）

参数问题报文

- IPv6基本首部或扩展首部出现问题，无法完成分组传输，目的节点或路由器会丢弃分组并发送ICMPv6参数问题报文
- 格式



  - Code=0：错误的首部字段
  - Code=1：无法识别的下一首部类型
  - Code=2：无法识别的IPv6可选项（问题：哪些扩展首部含可选项？是否都需要报错？）

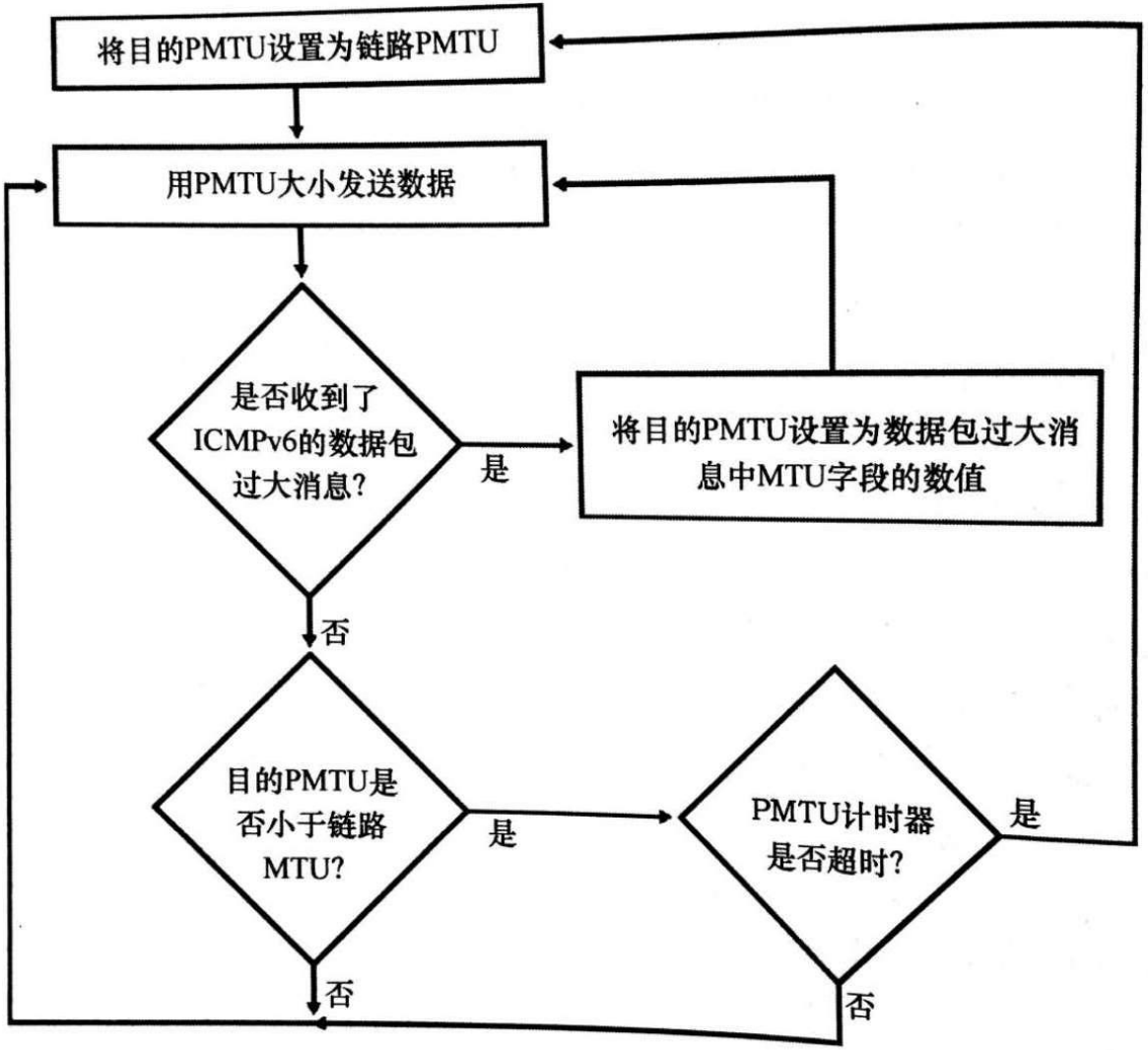
ICMP差错报告消息的应用--路径MTU发现

- 路径MTU发现(PMTUD-Path MTU Discovery)
  - Path MTU是路径上的最小接口MTU。
  - PMTUD的主要目的是发现某个时间路径上的MTU，当数据包被从源转发到目的地的过程中避免分段。
  - 实现方法：向目标节点发送“要求报告分片但又不被允许”的ICMP报文。
- 说明
  - 推荐IPv6节点支持PMTUD
  - 不支持该功能的节点必须使用1280字节的最小链路MTU作为所有目的的PMTU

ICMP差错报告消息的应用--路径MTU发现

- 实现过程
  - 步骤1：设置目的PMTU为发送流量的接口的链路MTU
  - 步骤2：按照PMTU发送数据包
  - 步骤3：分组过大，回送ICMP数据包过大报文（包含了转发失败的接口的链路MTU）
  - 步骤4：发送主机重置PMTU，重新发送。重复步骤2-步骤4,直到**发现PMTU**为止。
- 几个问题：
  - 发现PMTU的条件是什么？
  - PMTU修改的情况：网络拓扑变化造成PMTU随时间变化
    - 变小：响应，步骤3
    - 变大：主动发现，步骤1

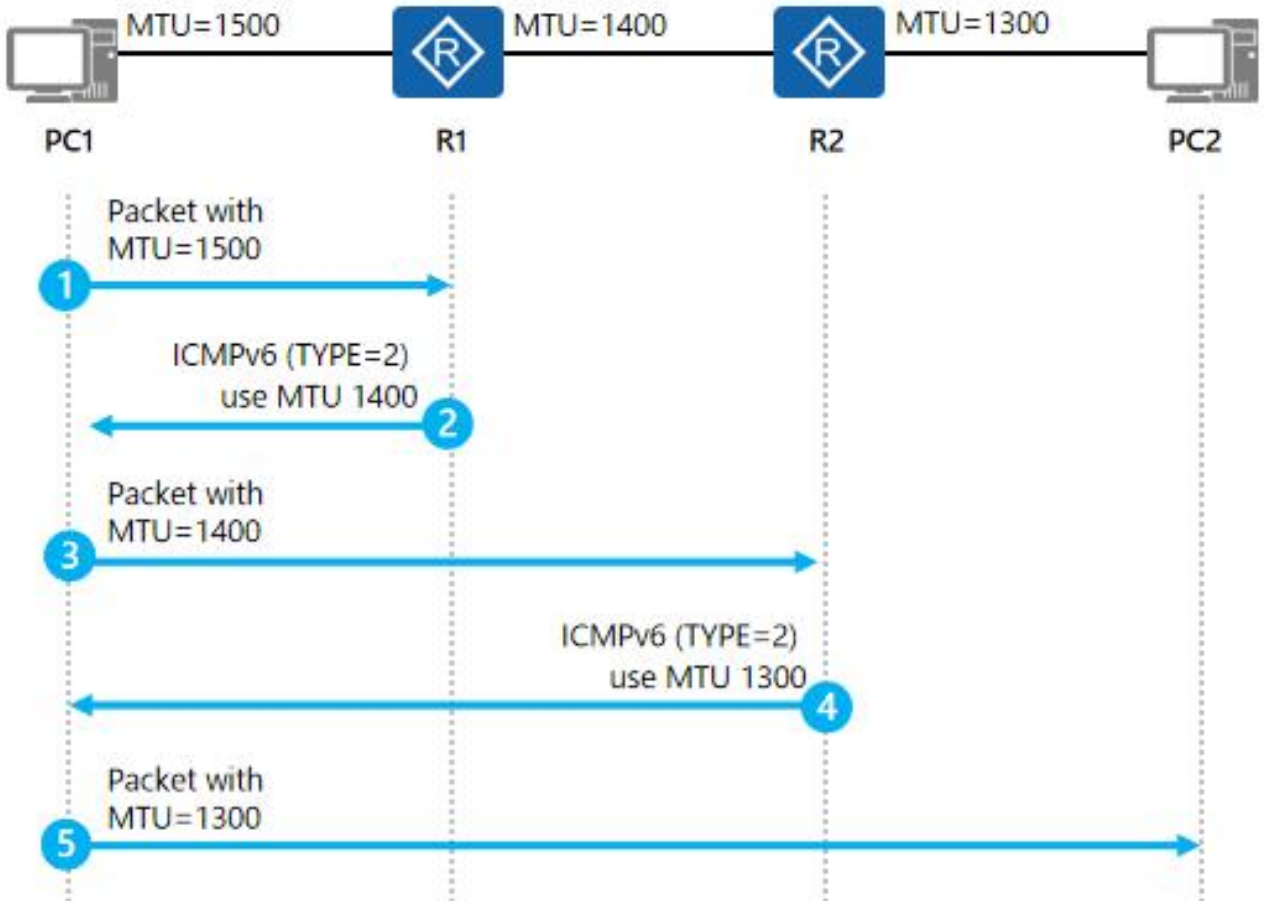
ICMP差错报告消息的应用--路径MTU发现



PMTU 发现过程

ICMP差错报告消息的应用--路径MTU发现

- 举例



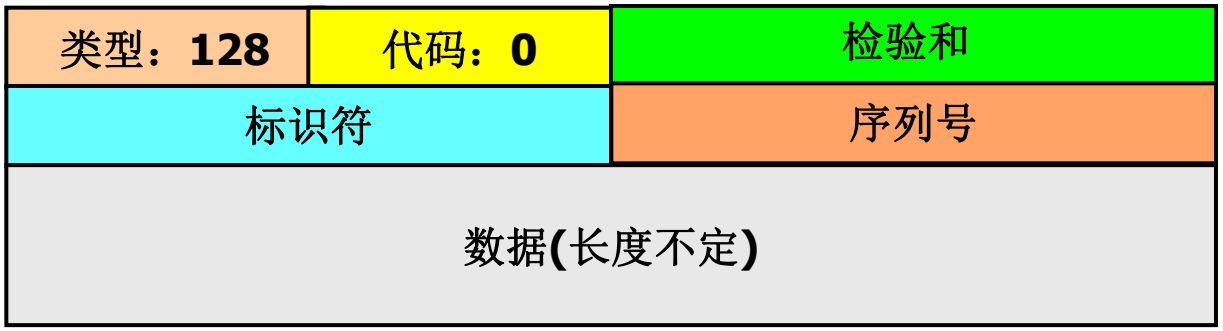
ICMP差错报告消息的应用--路径MTU发现

- 只有数据包超过路径上的最小MTU时，PMTUD机制才有意义
- 实际应用过程中存在的问题
  - 黑洞问题：防火墙或NAT阻塞ICMP消息
- 案例举例:

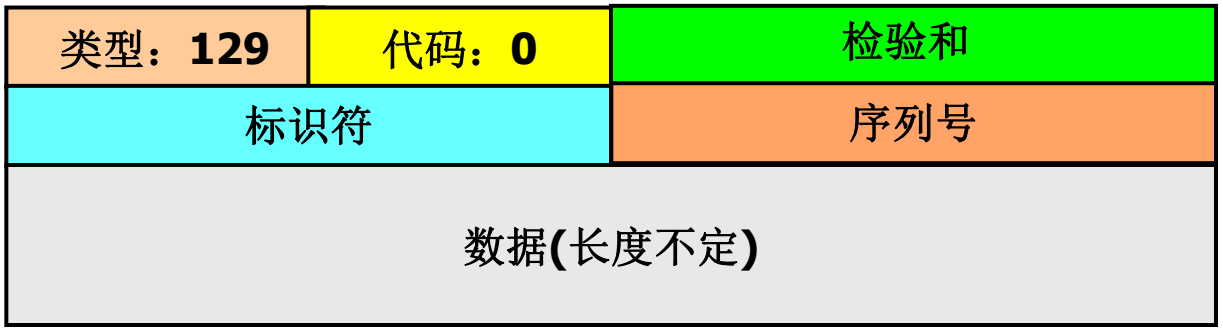
用 FTP 命令行工具成功地与 FTP 服务器建立连接并登录。但是，当您试图下载或者上载文件时，中间的 PMTU 黑洞路由器就会丢弃达到最大大小的 TCP 数据段，从而导致错误和文件传输失败。

查询报文（Echo报文）

- 回声请求报文
  - 单播或多播
  - 格式



- 回声应答报文
  - 单播
  - 标识符
  - 序列号
  - 数据



与回声请求报文相匹配，请求报文数据完整复制到应答报文。

ICMPv4与ICMPv6的比较

ICMPv4消息	ICMPv6消息
目的不可达-网络不可达(Type 3,Code 0)	目的不可达-没有目的地址的路由(Type 1,Code 0)
目的不可达-主机不可达(Type 3,Code 1)	目的不可达-地址不可达(Type 1,Code 3)
目的不可达-协议不可达(Type 3,Code 2)	参数问题-无法识别下一首部的类型(Type 4, Code 1)
目的不可达-端口不可达(Type 3,Code 3)	目的不可达-端口不可达(Type 1,Code 4)
目的不可达-需要进行分片并将FD置位(Type 3,Code 4)	数据包过大(Type 2,Code 0)
目的不可达-与目标主机的通信被管理策略禁止(Type 3,Code 10)	目的不可达-与目标主机的通信被管理策略禁止(Type 1,Code 1)
源抑制(Type 4,Code 0)	不支持这个消息
超时-传输中的TTL超时(Type 11,Code 0)	超时-超过传输中的Hop Limit(Type 3,Code 0)
超时-分片重组超时(Type 11,Code 1)	超时-分片重组超时(Type 3,Code 1)
参数问题(Type 12,Code 0)	参数问题(Type 4,Code 0 or 2)

ICMP差错消息的应用

- 路径MTU发现（PMTUD）
- 可达性(连通性)测试(PING)
- 网络路由跟踪（Traceroute）

可达性测试Ping

- 主机可达性测试：Ping
  - 方法：使用ICMP回送和应答消息来确定一台主机是否可达
  - 作用：Ping是因特网包探索器，Ping发送一个ICMP回声清求消息给目的地并报告是否收到所希望的ICMP回声应答使用ping命令，通过发送数据包，能够测试两台计算机之间的因特网连接是否正常、网卡配置是否正确、IP地址是否可使用等

Ping

- Ping是一个测试程序，用于确定本地主机是否能与另一台主机交换（发送与接收）数据报。如果Ping运行正确，就可以排除网络访问层、网卡、Modem的I/O线路、电缆和路由器等存在的故障。
- 按缺省设置，运行Ping命令时发送4个ICMP（网间控制报文协议）“回送请求”，每个32字节数据；若正常应得到4个回送应答。
- Ping能够以毫秒为单位显示发送“回送请求”到返回“回送应答”之间的时间量。如果应答时间短，表示数据报不必通过太多的路由器或网络连接，速度比较快

Ping

- 计算机进行TCP／IP通信的基本条件有：
  - 网卡安装正确；
  - 安装有TCP／IP协议；
  - TCP/IP协议的参数配置正确，TCP/IP涉及的基本参数有4个：IP地址、子网掩码、DNS和网关，任何一个设置错误都会导致故障发生；
  - 到有关节点(网关、服务器(如DNS等))网线连通。

Ping

- 按照以下步骤测试这些条件：
  - ping 127.0.0.1：网络地址127.0.0.1是一个保留地址，这个ＩＰ地址叫做回送地址（loopback address），用于测试本机的TCP/IP协议栈安装是否正确。无论网线是否连接，都能ping通本机的还回地址
  - ping <本机IP>：ping本机是试试网卡驱动和网卡是否连接网络。如果网线断掉，只能ping通你本机的还回地址。只有网线连接上才能ping通本机的ip地址。
  - ping 网关：测试到网关的局域网链路是否正常

路由跟踪

- Traceroute命令
  - Windows下是tracert命令
  - 作用：定位本计算机和目标计算机之间的所有路由器，即跟踪数据包访问网络中某个节点时所走的路径，进行路由跟踪，以用来分析网络和排查网络故障。
  - 实现方法：通过发送小的UDP数据包到目的设备直到其返回，来测量其需要多长时间。把一个TTL=1的数据报发送给目的主机，第1个路由器把TTL减小到0，丢弃该数据报并把ICMP超时消息返回给源主机，这样路径上第1个路由器就被标识了。随后不断增大TTL值重复该过程。
  - 如何判断探测数据包已经到达目的主机？

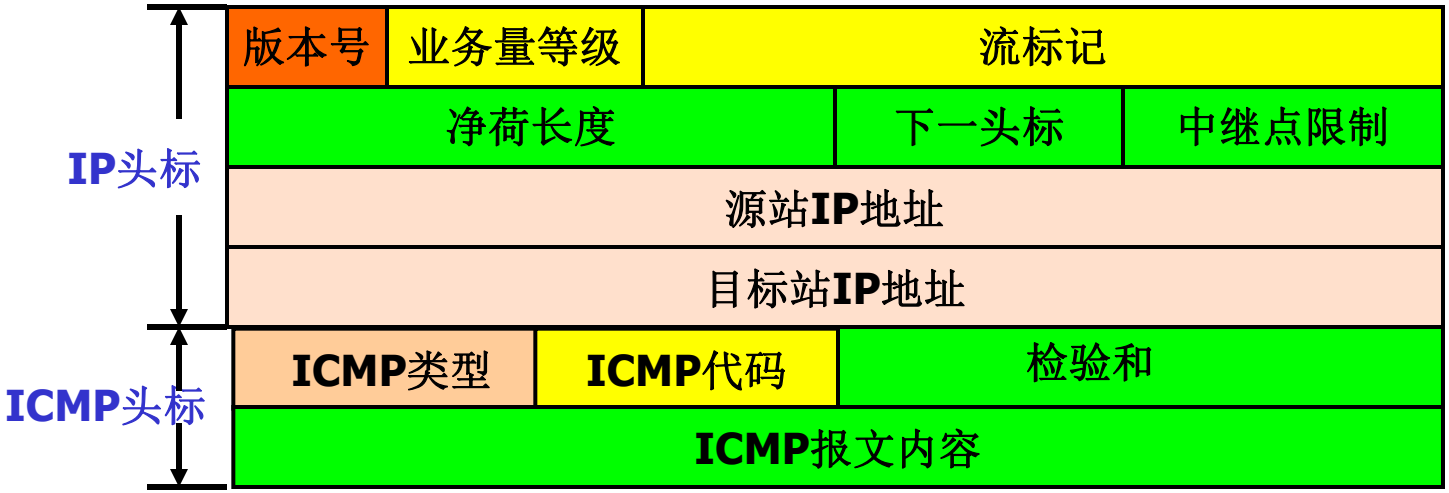
路由跟踪

- 探测数据包的设计
  - UDP模式：UDP探测数据包(目标端口大于30000) + 中间路由器返回 ICMP超时 + 目标主机返回ICMP 端口不可达数据包
  - ICMP模式：ICMP Echo Request + 中间路由器返回ICMP超时 + 目标主机返回ICMP Echo reply 数据包
  - TCP模式：TCP[SYN]探测数据包(目标端口为Web服务端口) + 中间路由器返回ICMP超时 + 目标主机返回TCP[SYN ACK] 数据包

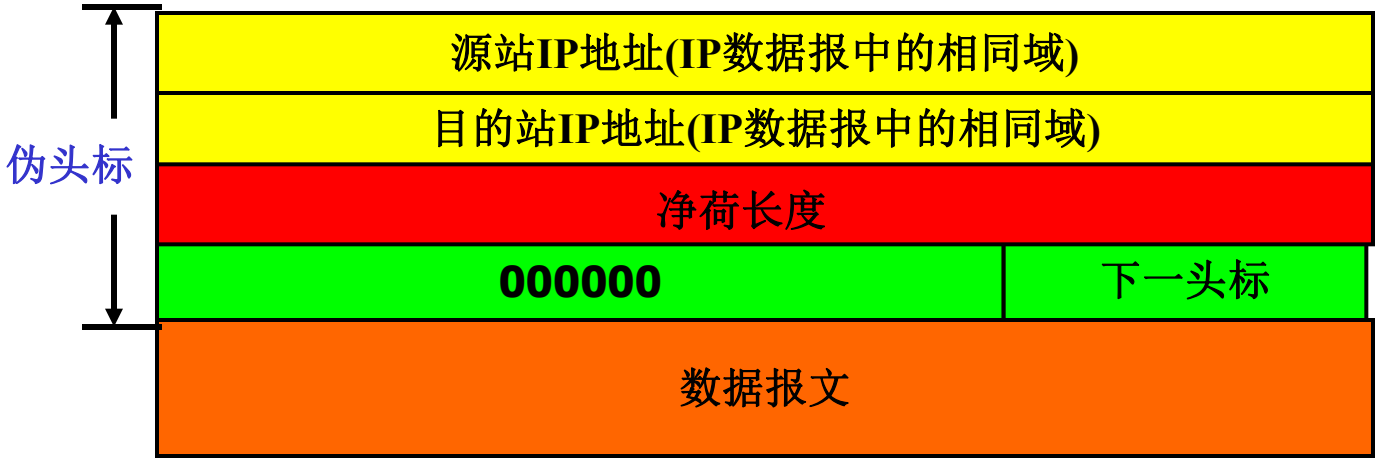
谢　谢！

附录A：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



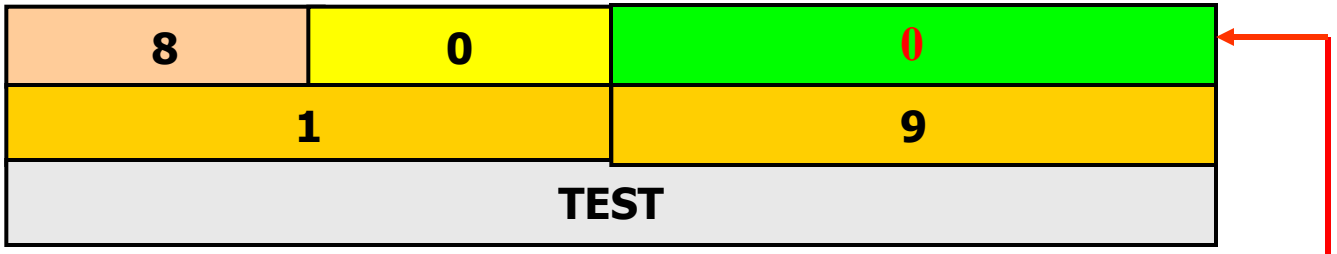
- IP基本头标中的关键数据
  - 信源地址
  - 信宿地址
  - 下一头标
  - 净荷长度



附录B: 16比特校验

- 校验和的计算
  - 校验和字段设置为0
  - 计算校验域内的所有16位字之和
  - 取反，得到校验和
- 校验和的检测
  - 计算校验域内的所有16位字之和
  - 把得到的和求反
  - 结果如果为16个0，则接受，否则拒绝。

附录B: 16特加法校验



8 和 0	00001000	00000000
0	00000000	00000000
1	00000000	00000001
9	00000000	00001001
T 和 E	01010100	01000101
S 和 T	01010011	01010100
和	10101111	10100011
校验和	01010000	01011100