计算机网络

11. IPv4 Datagram, Forwarding, Support Protocols, and IPv6

厦门大学软件学院

黄炜 助理教授

PART III Internetworking

Ch 20 IP Datagrams and Datagram Forwarding

IP数据报和数据报转发

20.3 Virtual Packet 虚拟包

- 网络互联协议定义独立于底层硬件的“分组”格式
  - 为解决异构问题
  - **The result is a universal, virtual packet.**
- 底层硬件不懂，路由器和主机（软件）懂
  - **The underlying hardware does not understand or recognize the internet packet format.**
  - **Each host or router in a internet contains protocol software that understands internet packets.**

20.4 The IP Datagram IP数据报

- TCP/IP protocol use the name IP datagram to refer to an internet packet.
- In IPv4, a datagram can contain as little as a single octet of data or at most 64K octets, including the header.

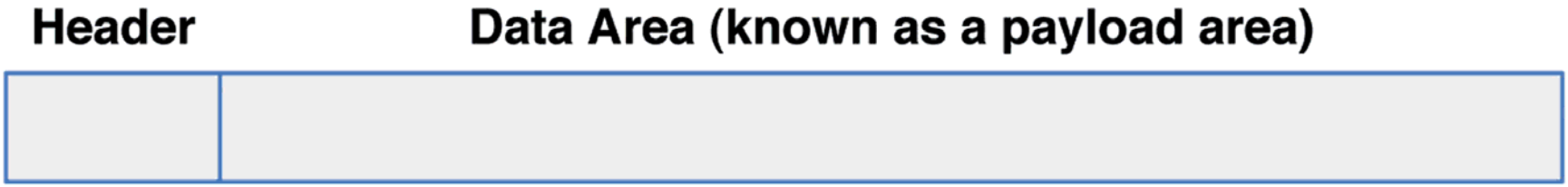


Figure 22.1 The general form of an IP datagram with a header followed by a payload.

The IP Datagram Header Format

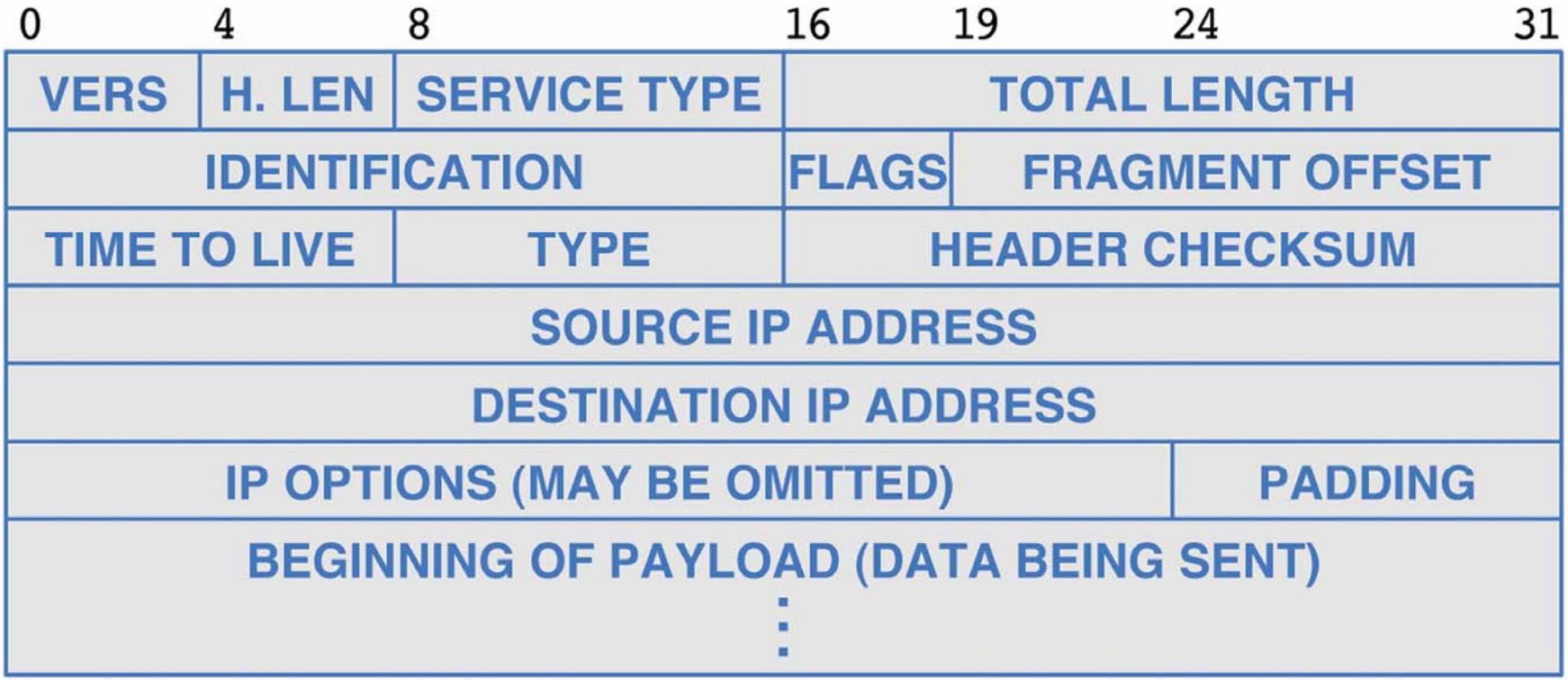


Figure 22.2 Fields in the IP version 4 datagram header.

IP报文格式

- IP报文头格式的组成（基本长度：20B）
  - 版本：4-bit，取值：4或6
  - 报头长度：4bits，单位为4Bytes
  - 服务类型：8bits，未实际使用
  - 报文总长度：16bits，单位为字节
  - 标识：16bits，IP软件在存储器中的计数器在产生一个数据报后自增1，并将值赋给标识字段。标识在分片时复制。
  - 分片标志：3bits，高到低位：无意义、不分片、还有分片

IP报文格式

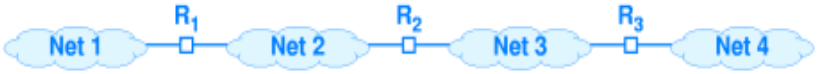
- IP报文头格式的组成（基本长度：20B）
  - 片偏移：13bits，分片在原始报文的位置，单位为8Bytes
  - TTL（生存时间）：8bits，单位为秒，路由器减去在其环节所消耗时间，直至零丢弃。
  - 协议类型：8bits，可能的取值有：ICMP、IGMP、TCP、UDP、OSPF等，用于将数据交给第四层的哪个软件。
  - 报头校验和：16bits，检验报头的完整性，不含数据部分。

IP报文格式

- IP报文头格式的组成（基本长度：20B）
  - 源IP地址：32bits
  - 目标IP地址：32bits
  - 选项内容：（可变长度，1～40bits），用来支持排错、测量以及安全等措施。
  - 填充部分：（根据选项部分改变），为了使得报文头部是4Bytes的整数倍

20.5 Forwarding An IP Datagram

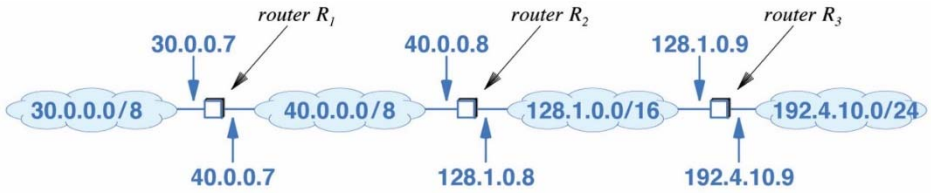
- 路由表的组成：目标网络号、子网掩码、下一跳
  - 下一跳：直接传送标志、子网IP地址
  - Internet路由表又大又复杂
    - 默认路由（default entry）



(a)

Destination	Next Hop
net 1	R_1
net 2	deliver direct
net 3	deliver direct
net 4	R_3

(b)



(a)

Destination	Mask	Next Hop
30.0.0.0	255.0.0.0	40.0.0.7
40.0.0.0	255.0.0.0	deliver direct
128.1.0.0	255.255.0.0	deliver direct
192.4.10.0	255.255.255.0	128.1.0.9

(b)

Figure 22.3 (a) An example internet with four networks, and (b) the forwarding table found in router R_2.

IP子网掩码与数据转发

- 通过子网掩码进行计算的路由表匹配
  - 获得IP报文的目标IP地址 D
  - 用 D 顺序逐条匹配路由表各个条目 $T_1, T_2, T_3, \cdots$
  - 如果 $D \& T_i^{(m)}$ 等于 $T_i^{(d)}$，则下一跳为 $T_i^{(n)}$
    - D：目标端地址
    - $T_i^{(d)}$：路由表中第 i 条目标子网网络号
    - $T_i^{(m)}$：路由表中第 i 条目标子网掩码
    - $T_i^{(n)}$：路由表中第 i 条下一跳IP地址

IP子网掩码与数据转发

- 最长前缀匹配（ Longest Prefix Match ）
  - 设路由表有以下两个网络前缀：128.10.0.0/16; 128.10.2.0/24
  - What happens if a datagram arrives destined to 128.10.2.3?
    - 16-bit v.s. 24-bit：最长的那个
    - Internet forwarding will choose the entry 128.10.2.0/24
- 目的IP：真正的接收者；源和目的MAC：路由器
  - 传输过程中，路由器的IP地址不在IP报文里出现（而是数据链路层）

20.9 Best-Effort Delivery

- IP要适应不同硬件的需要，但底层硬件可能不工作
- IP提供一种尽力而为的传输 (Best-Effort Delivery)
  - IP datagrams may be lost (丢失), duplicated (重复), delayed (延迟), delivered out of order (乱序), or delivered with corrupted data (数据损坏).
  - Higher layers of protocol software are required to handle each of these errors.

PART III Internetworking

Ch 21 IP Encapsulation, Fragmentation, and Reassembly

IP封装、分段与重组

21.2 Datagram Transmission and Frames

- IP 软件选择下一站，并通过物理网络发送
  - 网络硬件不理解数据报文格式和因特网地址
    - 每个网络格式都有自己的硬件地址
  - **The sender and receiver must agree on the value used in the frame type field.**
- 发送者必须指定下一接收主机的物理地址
- **Encapsulation 封装**

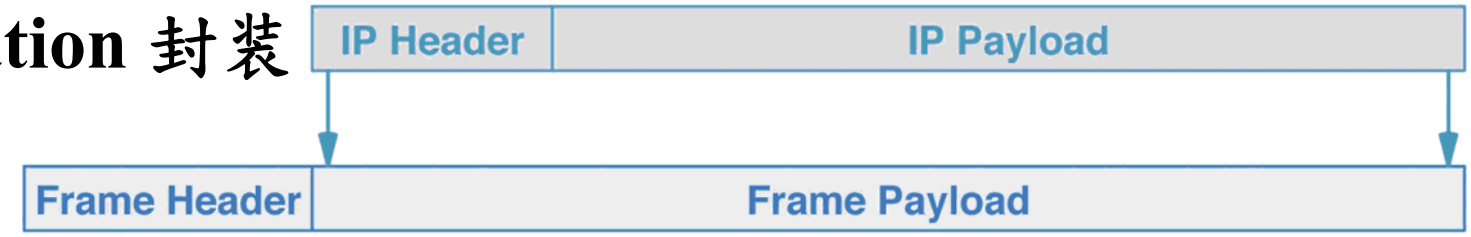


Figure 22.4 Illustration of an IP datagram encapsulated in a frame.

21.4 Transmission Across An Internet

- 帧到达下一跳，接收方软件提取IP数据报并丢弃帧头
- 若还需转发，则再封装
- 帧头部不积累
- 主机和路由器不存储额外的头部

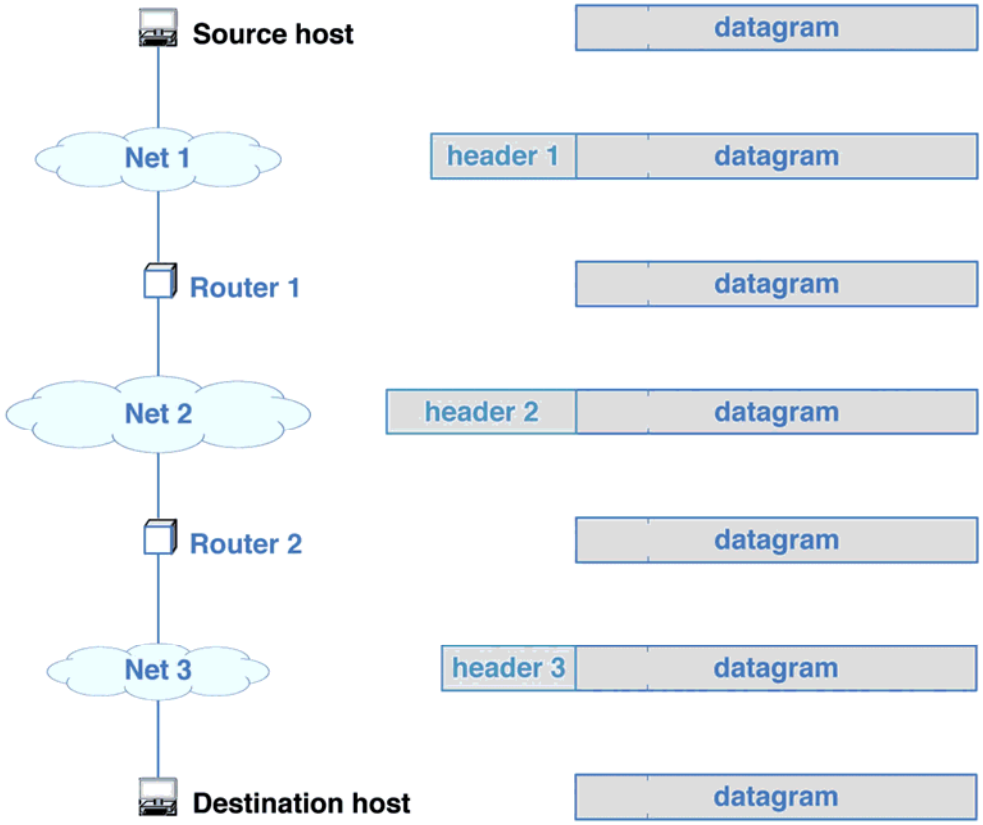


Figure 22.5 An IP datagram as it travels across the Internet.

最大传输单元（MTU）

- 什么是最大传输单元（人在屋檐下，不得不低头）
  - 数据链路层数据帧支持的最大传输字节数

Protocol	MTU
Token ring (16Mbps)	17914
Token ring (4Mbps)	4464
FDDI	4352
Ethernet	1500
X.25	576
PPP	296

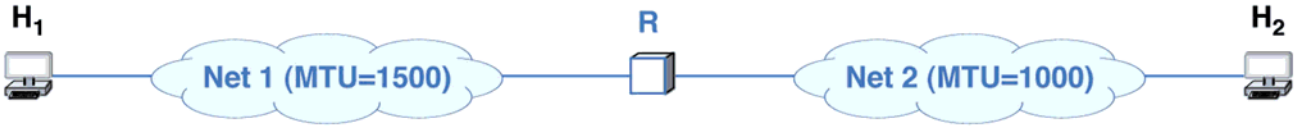
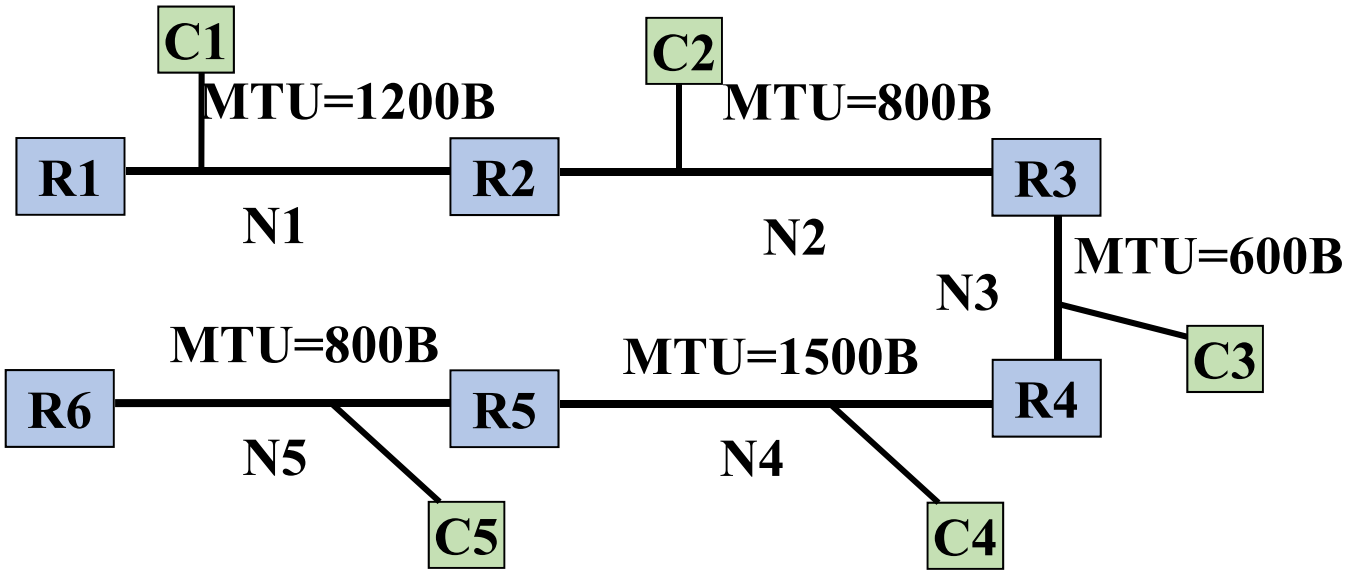


Figure 22.6 Illustration of a router that connects two networks with different MTUs.



21.5 MTU, Datagram Size, and Encapsulation

- The router (路由器) divides the datagram (数据报) into smaller pieces called fragments (分片).
- Each fragment uses the IP datagram format, and is sent independently (独立地).

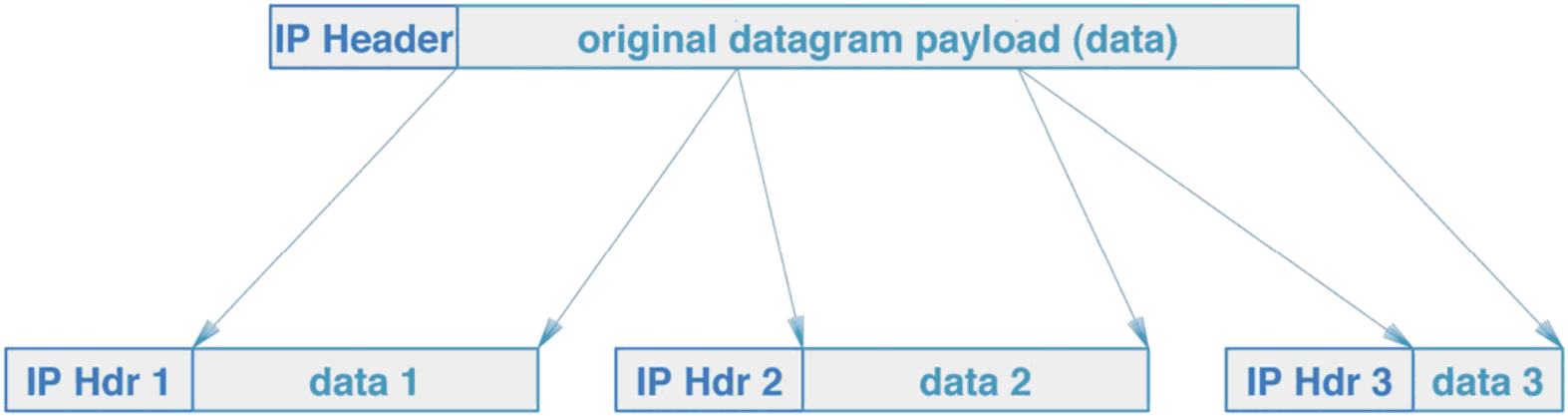


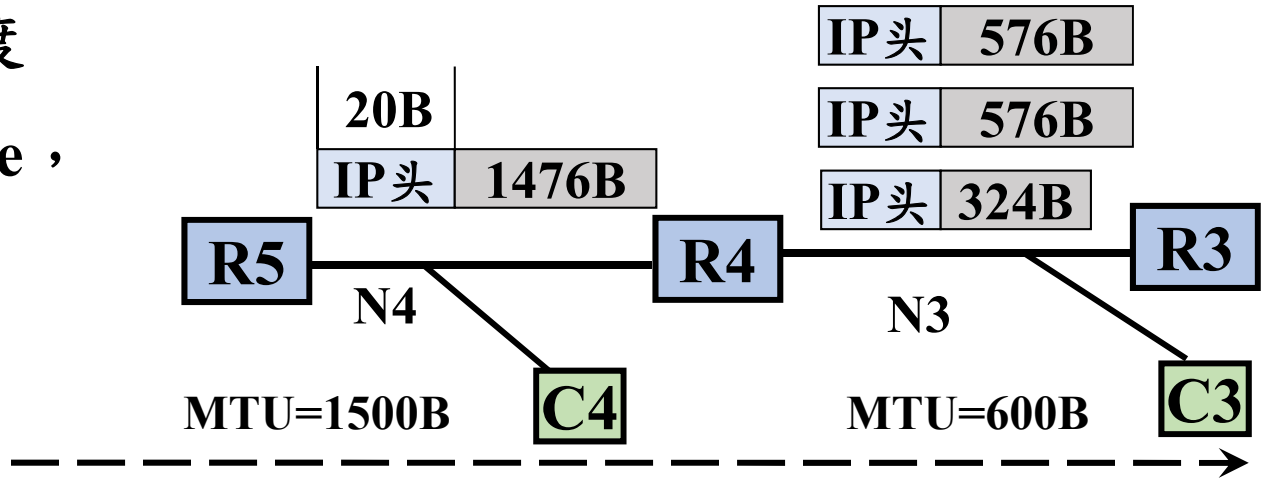
Figure 22.7 An IP datagram divided into three fragments, with the final fragment smaller than the others.

IP报文的分片策略

- 当前链路MTU小于IP报文长时，分成较小分片传输
- IP报文传输的分片原则
  - 各片尽可能大，但是必须能为帧所封装
  - 前面片大小必须为8的整数倍（FLAG已经占了3个bit了）
  - IP头部固有长度

    基本长度20Byte，

    不是8的倍数！



分片信息表示

- IP报文中的相关信息
  - IP报文中的ID：分片时复制，始终保持初始ID不变
  - 标志位：若第一次分片，则修改“是否分片”相应位
  - 片偏移量：表示当前分片在初始IP包中有效数据的偏移位置（8字节为单位）
    - 标志 MF(more fragment), MF=1 表示后面还有分片；MF=0 表示这是最后一个分片；DF(don’t fragment), DF=1 表示不允许分片；DF=0 表示允许分片。

分片信息表示

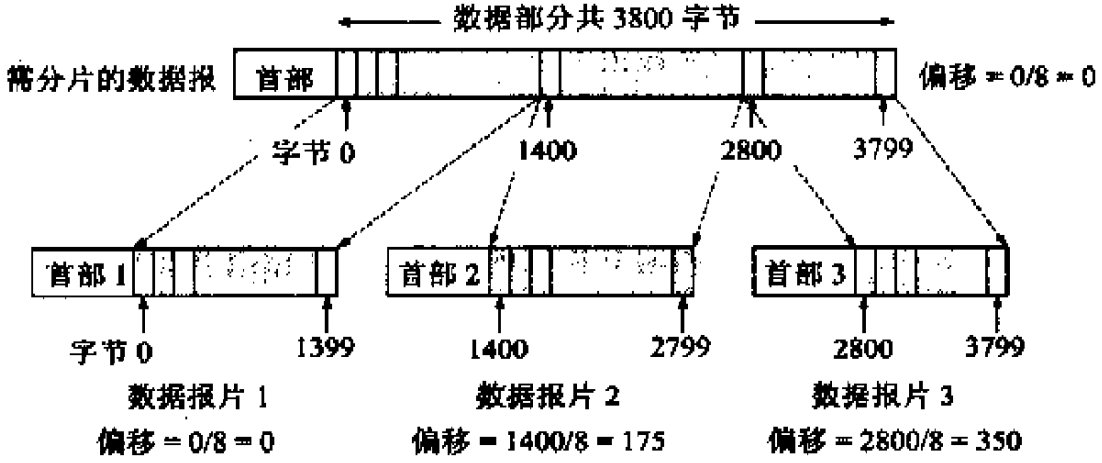


表 7-4　IP 数据报首部中与分片有关的字段中的数值

	总长度	标识	MF	DF	片偏移
原始数据报	4000	12345	0	0	0
数据报片 1	1420	12345	1	0	0
数据报片 2	1420	12345	1	0	175
数据报片 3	1020	12345	0	0	350

21.6 Reassembly重组

- 报文重组策略
  - 源端到目标端数据传输过程中可能有多次分片
  - 所有分片重组在目标端进行，中间路由设备不做分片重组
    - 减少中间节点的数据处理过程
  - 碎片还可以再分片(further fragment a fragment)

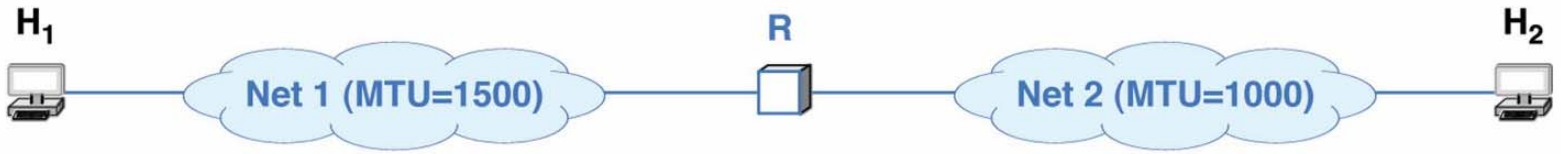


Figure 22.6 Illustration of a router that connects two networks with different MTUs.

IP报文丢失问题

- IP报文丢失判断
  - 目标端对IP报文分片作重组处理的时候进行丢失判断
  - 对应于源端发出的每一个报文，在收到第一个分片的时候，给出一个等待的有限时间T-out，如果T-out之后还没有收到全部分片，则为超时
  - 任何一个分片丢失或数据出错，则丢弃整个报文

PART III Internetworking

Ch 23 An Error Reporting Mechanism (ICMP)

差错报告机制 (ICMP)

23.2 Best-Effort Semantics and Error Detection

- **IP Datagrams can be lost, duplicated, delayed, or delivered out of order.**
- **IP attempts to avoid errors and to report problems when they occur.**

23.3 Internet Control Message Protocol

- The TCP/IP suite includes a protocol that IP uses to send error messages when conditions such as the one described above arise.
- ICMP: Internet Control Message Protocol.
- The protocol is required for a standard implementation of IP.
- IP uses ICMP when it sends an error message, and ICMP uses IP to transport messages.

ICMP 信息列表

Type	Name
0	Echo Reply
1	Unassigned
2	Unassigned
3	Destination Unreachable
4	Source Quench
5	Redirect
6	Alternate Host Address
7	Unassigned
8	Echo
9	Router Advertisement
10	Router Selection
11	Time Exceeded
12	Parameter Problem
13	Timestamp
14	Timestamp Reply
15	Information Request
16	Information Reply
17	Address Mask Request
18	Address Mask Reply
19	Reserved (for Security)
20-29	Reserved (for Robustness Experiment)
30	Traceroute
31	Datagram Conversion Error
32	Mobile Host Redirect
33	IPv6 Where-Are-You
34	IPv6 I-Am-Here
35	Mobile Registration Request
36	Mobile Registration Reply
37-255	Reserved

表 7-7　类型字段的值与 ICMP 报文的类型的关系

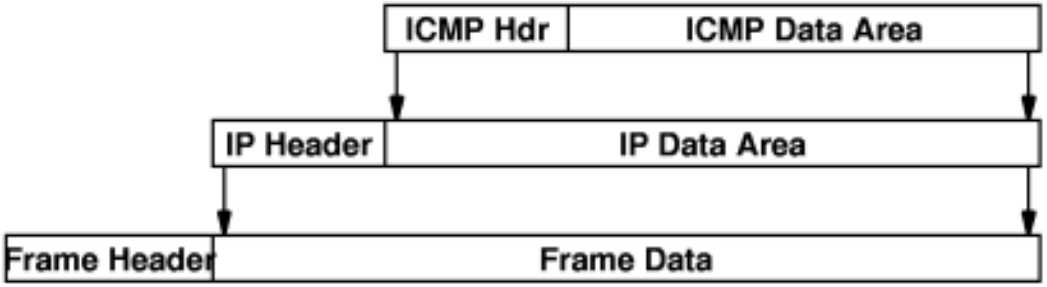
ICMP 报文种类	类型的值	ICMP 报文的类型
差错报告报文	3	目的站不可达
	4	源站抑制（Source quench）
	11	时间超过
	12	参数问题
	5	改变路由（Redirect，或重定向）
询问报文	8 或 0	回送（Echo）请求或回答
	13 或 14	时间戳（Timestamp）请求或回答
	17 或 18	地址掩码（Address mask）请求或回答
	10 或 9	路由器询问（Router solicitation）或通告

ICMP错误报文与信息报文

- **Examples of ICMP error messages:**
  - **Source Quench (源抑制)**
  - **Time Exceeded (超时)**
  - **Destination Unreachable (目的不可达)**
  - **Redirect (重定向)**
  - **Parameter Problem (参数问题)**
- **Examples of ICMP informational messages:**
  - **Echo Request/Reply (回应请求/应答)**
  - **Address Mask Request/Reply (地址屏蔽码请求/应答)**

23.4 ICMP Message Transport

- **ICMP uses IP to transport each error message.**
  - **The ICMP message is placed in the data area of the IP datagram.**



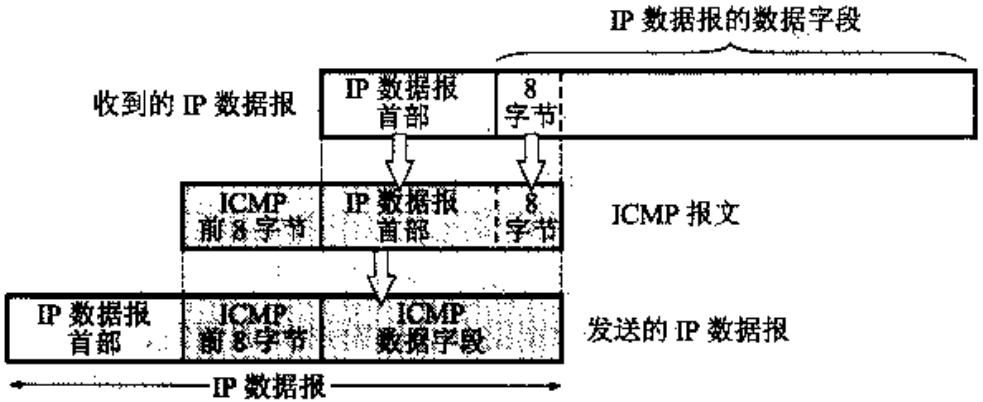


图 7-23 ICMP 差错报告报文的数据字段的内容

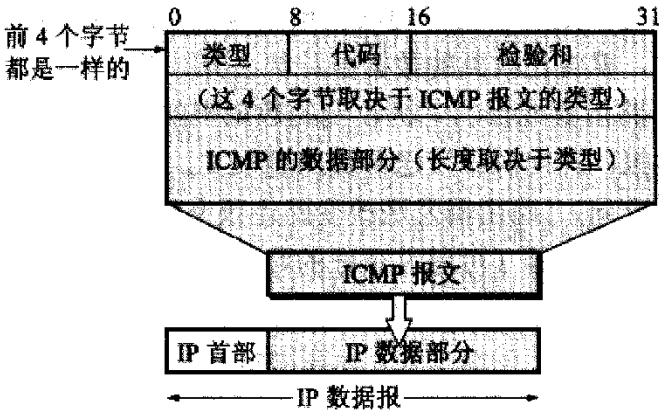


图 7-22 ICMP 报文的格式

ICMP: Echo Request

```
Ethernet Type 2
  Destination:          00:50:56:FC:52:95  VMware:FC:52:95  [0-5]
  Source:               00:0C:29:17:29:CA  VMware:17:29:CA  [6-11]
  Protocol Type:        0x0800  IP  [12-13]
IP Version 4 Header - Internet Protocol Datagram
  Version:              4  [14 Mask 0xF0]
...

  Fragment Offset:      0  (0 bytes)  [20-21 Mask 0x1FFF]
  Time To Live:         128  [22]
  Protocol:             1  ICMP - Internet Control Message Protocol  [23]
  Header Checksum:      0x0000  Checksum invalid. Should be:  0x1128  [24-25]
  Source IP Address:    192.168.7.4  [26-29]
  Dest. IP Address:     123.125.114.144  [30-33]
ICMP - Internet Control Messages Protocol
  ICMP Type:            8  Echo Request  [34]
  ICMP Code:            0  [35]
  ICMP Checksum:        0x4D5A  [36-37]
  Identifier:           0x0001  [38-39]
  Sequence Number:      1  [40-41]
  ICMP Data Area:       abcdefghijklmnopqrstuvwabcdefghi  [42-73]
```

ICMP: Echo Reply

```
Ethernet Type 2
  Destination:          00:0C:29:17:29:CA  VMware:17:29:CA  [0-5]
  Source:               00:50:56:FC:52:95  VMware:FC:52:95  [6-11]
  Protocol Type:        0x0800  IP  [12-13]
IP Version 4 Header - Internet Protocol Datagram
  Version:              4  [14 Mask 0xF0]
...

  Fragment Offset:      0  (0 bytes)  [20-21 Mask 0x1FFF]
  Time To Live:         128  [22]
  Protocol:             1  ICMP - Internet Control Message Protocol  [23]
  Header Checksum:      0xB747  [24-25]
  Source IP Address:    123.125.114.144  [26-29]
  Dest. IP Address:     192.168.7.4  [30-33]
ICMP - Internet Control Messages Protocol
  ICMP Type:            0  Echo Reply  [34]
  ICMP Code:            0  [35]
  ICMP Checksum:        0x555A  [36-37]
  Identifier:           0x0001  [38-39]
  Sequence Number:      1  [40-41]
  ICMP Data Area:       abcdefghijklmnopqrstuvwabcdefghi  [42-73]
```

ICMP: Time Exceeded

```
Ethernet Type 2
  Destination:          00:0C:29:17:29:CA  VMware:17:29:CA  [0-5]
  Source:               00:50:56:FC:52:95  VMware:FC:52:95  [6-11]
  Protocol Type:        0x0800  IP  [12-13]
IP Version 4 Header - Internet Protocol Datagram
  Version:              4  [14 Mask 0xF0]
  Header Length:        5  (20 bytes)  [14 Mask 0x0F]
...

  Fragment Offset:      0  (0 bytes)  [20-21 Mask 0x1FFF]
  Time To Live:         128  [22]
  Protocol:             1  ICMP - Internet Control Message Protocol  [23]
  Header Checksum:      0x8288  [24-25]
  Source IP Address:    192.168.7.2  [26-29]
  Dest. IP Address:     192.168.7.4  [30-33]
ICMP - Internet Control Messages Protocol
  ICMP Type:            11  Time Exceeded  [34]
  ICMP Code:            0  Time to Live count exceeded  [35]
  ICMP Checksum:        0xF4FF  [36-37]
  Unused (must be zero):0x00000000  [38-41]
```

ICMP: Time Exceeded (*cont.*)

```
Header of packet that caused error follows.
IP Version 4 Header - Internet Protocol Datagram
  Version:              4  [42 Mask 0xF0]
  Header Length:        5  (20 bytes)  [42 Mask 0x0F]
...

  Fragment Offset:      0  (0 bytes)  [48-49 Mask 0x1FFF]
  Time To Live:         1  [50]
  Protocol:             1  ICMP - Internet Control Message Protocol  [51]
  Header Checksum:      0x1C3F  [52-53]
  Source IP Address:    192.168.7.4  [54-57]
  Dest. IP Address:     210.34.0.12  [58-61]
ICMP - Internet Control Messages Protocol
  ICMP Type:            8  Echo Request  [62]
  ICMP Code:            0  [63]
  ICMP Checksum:        0xF7F7  [64-65]
  Identifier:           0x0001  [66-67]
  Sequence Number:      7  [68-69]
  ICMP Data Area:
.....................................................................  [70-133]
```

23.5 Using ICMP Message to test Reachability

- `ping` program uses the ICMP echo request and echo reply messages.
- `Ping` sends an IP datagram that contains an ICMP echo request message to the specified destination.
- Whenever an echo request arrives, the ICMP software must send an echo reply.

23.6 Using ICMP to Trace A Router

- Each router that handles a datagram decrements the TIME TO LIVE counter in the header.
- If a counter reaches zero, the router discards the datagram and sends an ICMP time exceeded error back to the source.
- `Traceroute` program simply sends a series of datagram and waits for a response to each.

23.7 The last Address printed by Traceroute

- Traceroute continues to increment the TIME TO LIVE until the value is large enough for the datagram to reach its final destinaation.
- Send an ICMP echo request message; the destination host will generate an ICMP echo reply.
- Send a datagram to a nonexistent application; the destination host will generate an ICMP destination unreachable message.

Ping 检查网络是否连通

```
C:\Windows\system32>ping www.xmu.edu.cn

正在 Ping www.xmu.edu.cn [210.34.0.12] 具有 32 字节的数据:
来自 210.34.0.12 的回复: 字节=32 时间=1ms TTL=128
来自 210.34.0.12 的回复: 字节=32 时间=1ms TTL=128
来自 210.34.0.12 的回复: 字节=32 时间=1ms TTL=128
来自 210.34.0.12 的回复: 字节=32 时间=1ms TTL=128

210.34.0.12 的 Ping 统计信息:
    数据包: 已发送 = 4，已接收 = 4，丢失 = 0 (0% 丢失)，
往返行程的估计时间(以毫秒为单位):
    最短 = 1ms，最长 = 1ms，平均 = 1ms
```

Route 操作网络路由表

```
C:\Windows\system32>route print -4
===========================================================
IPv4 路由表
===========================================================
活动路由:
网络目标        网络掩码          网关       接口   跃点数
          0.0.0.0          0.0.0.0      192.168.7.2    192.168.7.132     10
        127.0.0.0        255.0.0.0         在链路上         127.0.0.1    306
        127.0.0.1  255.255.255.255         在链路上         127.0.0.1    306
  127.255.255.255  255.255.255.255         在链路上         127.0.0.1    306
      192.168.7.0    255.255.255.0         在链路上     192.168.7.132    266
    192.168.7.132  255.255.255.255         在链路上     192.168.7.132    266
    192.168.7.255  255.255.255.255         在链路上     192.168.7.132    266
        224.0.0.0        240.0.0.0         在链路上         127.0.0.1    306
        224.0.0.0        240.0.0.0         在链路上     192.168.7.132    266
  255.255.255.255  255.255.255.255         在链路上         127.0.0.1    306
  255.255.255.255  255.255.255.255         在链路上     192.168.7.132    266
===========================================================================
永久路由:
  无
```

Tracert 搜索目标的跃点数

```
C:\Windows\system32>tracert www.xmu.edu.cn

通过最多 30 个跃点跟踪
到 www.xmu.edu.cn [210.34.0.12] 的路由:

  1    <1 毫秒    1 ms    12 ms  192.168.7.2
  2     *         *        *     请求超时。
  3     *         *        *     请求超时。
  4     *         *        *     请求超时。
  5     1 ms      1 ms     1 ms  210.34.0.12

跟踪完成。
```

Ping, Tracert, Route的高级用法，请进入控制台输入：XXXX /?

23.8 Using ICMP for Path MTU Discovery

- In a router, IP software fragments any datagram that is larger than the MTU of the network over which the datagram is being transmitted.
- The smallest MTU along a path from a source to a destination is known as the path MTU.
- The error message consists of an ICMP message that reports fragmentation was required but not permitted.

PART III Internetworking

Ch 21 The Future IP (IPv6)

未来的IP：IPv6

IPv6的产生

- 2011年2月3日IPv4的42亿地址分配用尽。
- IP的瓶颈

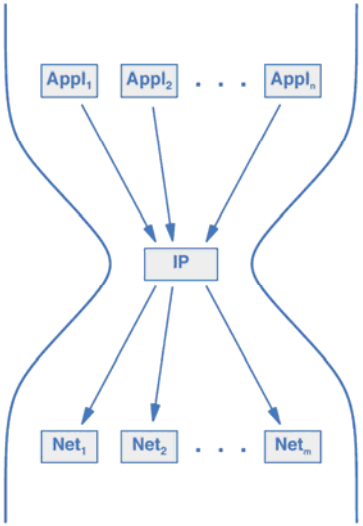


Figure 24.1 The hourglass model of Internet communication with IP at the center.

IPv6的特点

- 地址空间：128位
- 头部格式：新的头部
- 扩展头部：不同信息编码到不同头部中
  - 经济性、可扩展性
- 支持实时业务：允许底层网络建立搞质量通路
- 可扩充的协议：允许在数据报添加额外的信息

IPv6数据报格式

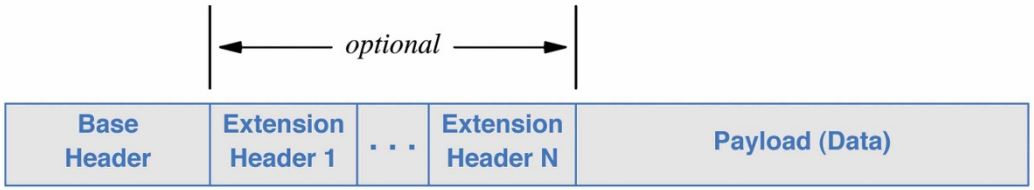


Figure 24.2 The general form of an IPv6 datagram.

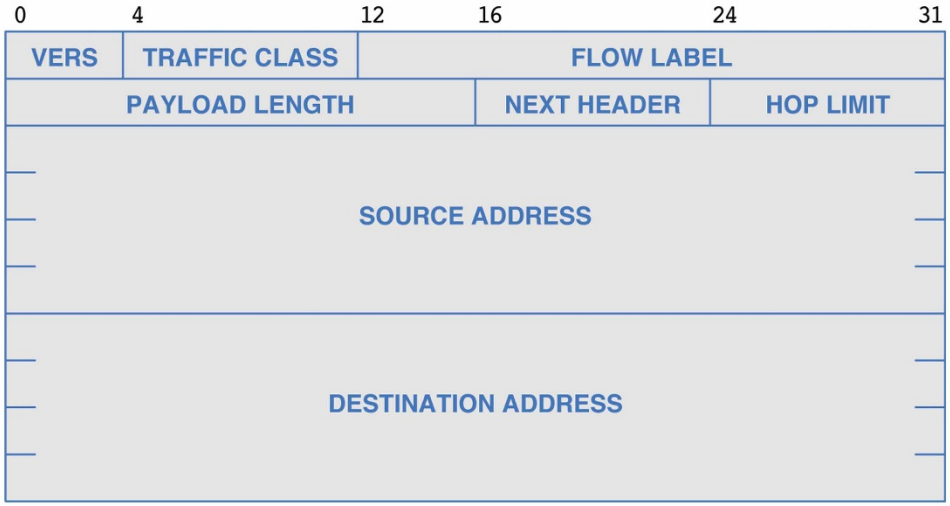
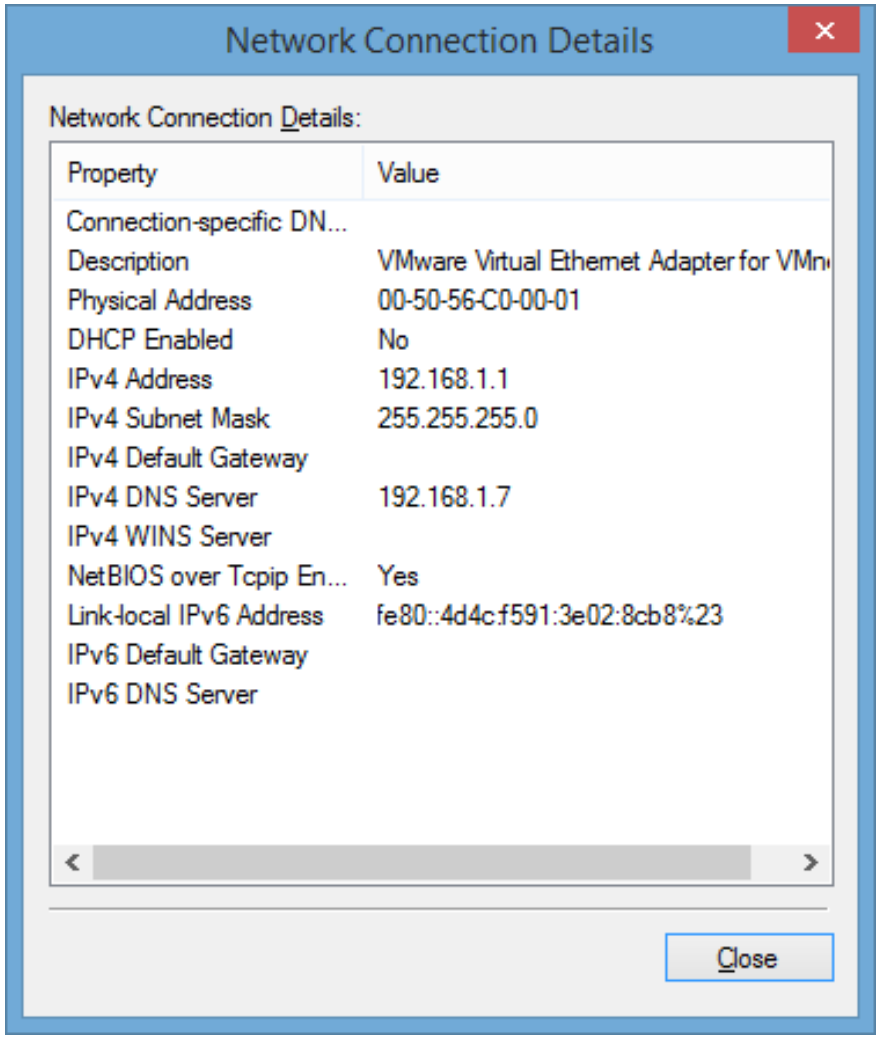


Figure 24.3 The format of the base header in an IPv6 datagram.



IP地址

fe80::4d4c:f591:3e02:8cb8%23

IPv6地址

- 冒分十六进制数表示法（兼容CIDR表示法）
  - 按16位一组，以冒号分隔每个组
    - 2001:0db8:85a3:0000:0000:8a2e:0370:7334
  - 前导0压缩：0db8写成db8
  - 零压缩：两个冒号代替连续出现两个以上的零，最多1次
    - 2001:db8:85a3::8a2e:370:7334
  - IPv4扩展到IPv6
    - ::ffff:0:0/96前缀

IP地址哪里来

- 向ISP购买一个（或段）IP的使用权
- 大量设备如何使用有限的地址上网
  - DHCP服务：“时分多路复用”，轮流使用IP地址
  - NAT、NAPT服务：“频分多路复用”，共用一个IP地址

动态主机配置协议（DHCP）

- 早期：反向地址解析协议（RARP）
- 作用：从服务器获得IP地址。
- 已知条件
  - 本地机器：没有IP（本机IP：0.0.0.0；MAC已知）
  - 目的机器：有IP但不知道（目的IP、MAC，全1广播）
- DHCP获得的IP地址有租期，可附加其他配置
- DHCP提供一个好心的服务（防君子不防小人）

监听结果

- 用Omnipeek软件解析DHCP包
  - 拔出网线，开软件，勾选DHCP，插入网线，再解析

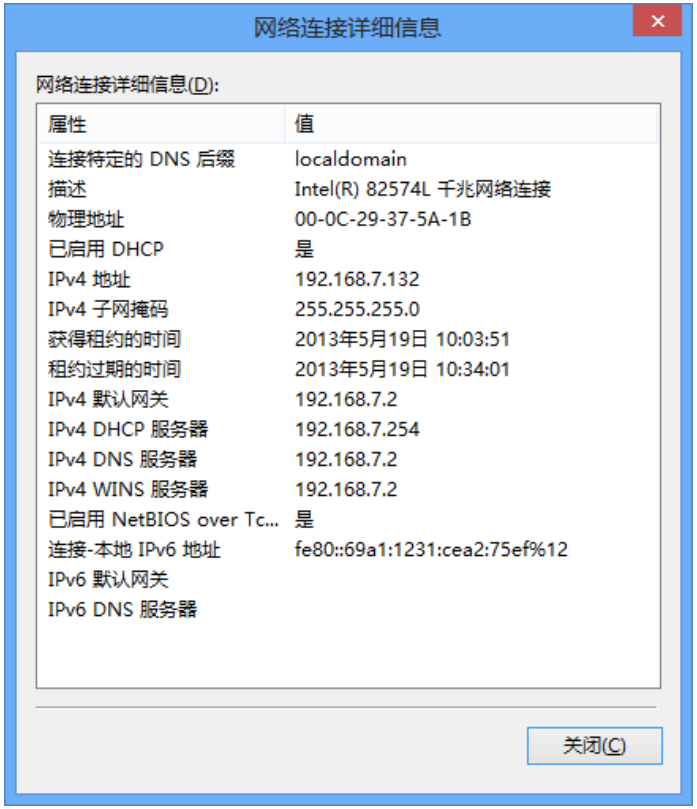
ID	Src. Logical	Src. Physical	Src. Port	Dest. Log.	Dest. Phy.	Dest. Prt.	Summary	Expert
1	0.0.0.0	00:0C:29:37:5A:1B	UDP 68	255.255.255.255	FF:FF:FF:FF:FF:FF	UDP 67	C DISCOVER 192.168.7.132 WIN-KG9CLM76UIA	
2	192.168.7.254	00:50:56:E2:AF:04	UDP 67	192.168.7.132	00:0C:29:37:5A:1B	UDP 68	R OFFER 192.168.7.132	
3	0.0.0.0	00:0C:29:37:5A:1B	UDP 68	255.255.255.255	FF:FF:FF:FF:FF:FF	UDP 67	C REQUEST 192.168.7.132 WIN-KG9CLM76UIA	
4	192.168.7.254	00:50:56:E2:AF:04	UDP 67	192.168.7.132	00:0C:29:37:5A:1B	UDP 68	R ACK	DHCP Low Lease Time (30 minutes, threshold=30 minutes)

监听结果节选

```
Packet #1
Ethernet Type 2
  Destination:          FF:FF:FF:FF:FF:FF  Ethernet Broadcast  [0-5]
  Source:               00:0C:29:37:5A:1B  VMware:37:5A:1B  [6-11]
  Protocol Type:        0x0800  IP  [12-13]
IP Version 4 Header - Internet Protocol Datagram
  Version:              4  [14 Mask 0xF0]
  Protocol:             17  UDP  [23]
  Source IP Address:    0.0.0.0  [26-29]
  Dest. IP Address:     255.255.255.255  IP Broadcast  [30-33]
UDP - User Datagram Protocol
  Source Port:          68  bootpc  [34-35]
  Destination Port:     67  bootps  [36-37]
BootP - Bootstrap Protocol
  IP Address Known By Client:   0.0.0.0  IP Address Not Known By Client  [54-57]
  Client IP Addr Given By Srvr: 0.0.0.0  [58-61]
  Server IP Address:            0.0.0.0  [62-65]
  Gateway IP Address:           0.0.0.0  [66-69]
  Client Hardware Addr:         00:0C:29:37:5A:1B  VMware:37:5A:1B  [70-75]
DHCP - Dynamic Host Configuration Protocol
Requested IP Address
  Address:              192.168.7.132  [296-299]
Host Name Address
  String:               WIN-KG9CLM76UIA  [302-316]
```

DHCP的配置

- Windows提供DHCP Client服务



网络地址转换（NAT）

- NAT应用场景
  - 多台主机上网，但是只有一个公网IP地址
- NAT动机：IP地址紧张，端口号并不紧张

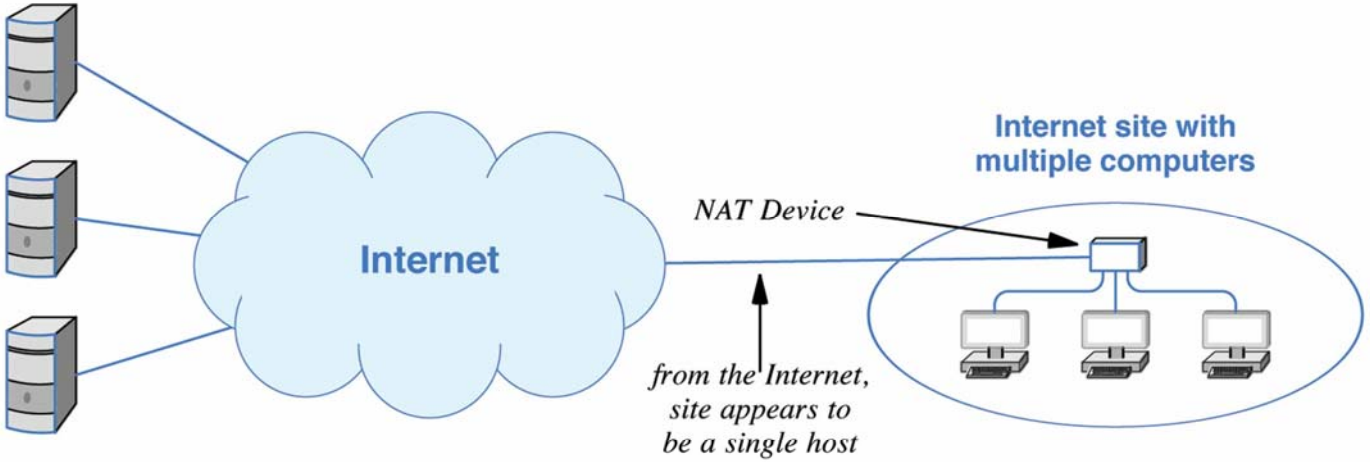


Figure 23.9 The conceptual architecture used with NAT.

私有地址

- 目的：虚拟的寻址机制
  - The goal of NAT is to provide an illusion (错觉).
- Blocks of private addresses (私有地址) used by NAT
  - 10.0.0.0/8：Class A private address block
  - 169.254.0.0/16：Class B private address block
    - 一般开启DHCP客户端又无法获取到IP时使用
  - 172.16.0.0/12：16 contiguous Class B blocks
  - 192.168.0.0/16：256 contiguous Class C blocks
- 防止IP冲突，私有地址不被路由

NAT的地址转换

- The most basic form of NAT replaces the IP source address in datagrams passing from the site to the Internet, and replaces the IP destination address in datagrams passing from the Internet to the site

Direction	Field	Old Value	New Value
out	IP Source	192.168.0.1	128.210.24.6
	IP Destination	198.133.219.25	-- no change --
in	IP Source	198.133.219.25	-- no change --
	IP Destination	128.210.24.6	192.168.0.1

NAT的地址转换

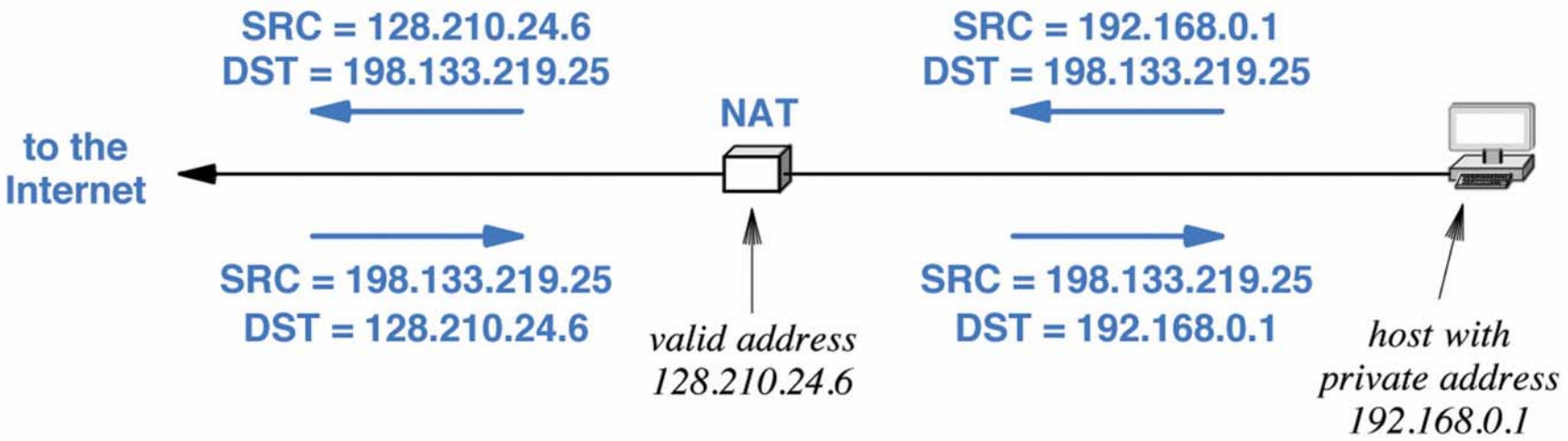


Figure 23.11 Illustration of basic NAT translation that changes the source address of an outgoing datagram and the destination address of an incoming datagram.

传输层的NAT（NAPT）

- 传输层的特别之处：端口号
- 端口号也参与转换
  - 因为终究是主机上的应用在上网
- NAT有时候也用于负载均衡

Dir.	Fields	Old Value	New Value
out	IP SRC:TCP SRC	192.168.0.1 :30000	128.10.24.6 :40001
out	IP SRC:TCP SRC	192.168.0.2 :30000	128.10.24.6 :40002
in	IP DEST:TCP DEST	128.10.19.20 :40001	192.168.0.1 :30000
in	IP DEST:TCP DEST	128.10.19.20 :40002	192.168.0.2 :30000

FTP Login (VMWare)

```
  Timestamp:            21:00:57.444125300 04/11/2014
Ethernet Type 2
  Destination:          00:50:56:FC:52:95  VMware:FC:52:95  [0-5]
  Source:               00:0C:29:17:29:CA  VMware:17:29:CA  [6-11]
  Protocol Type:        0x0800  IP  [12-13]
IP Version 4 Header - Internet Protocol Datagram
...
  Fragment Offset:      0  (0 bytes)  [20-21 Mask 0x1FFF]
  Time To Live:         128  [22]
  Protocol:             6  TCP - Transmission Control Protocol  [23]
  Header Checksum:      0x0000  Checksum invalid. Should be:  0xB059  [24-25]
  Source IP Address:    192.168.7.4  [26-29]
  Dest. IP Address:     59.77.7.25  [30-33]
TCP - Transport Control Protocol
  Source Port:          4425  netrockey6  [34-35]
  Destination Port:     21  ftp  [36-37]
  Sequence Number:      1304971726  [38-41]
  Ack Number:           1171416600  [42-45]
  TCP Offset:           5  (20 bytes)  [46 Mask 0xF0]
...
FTP Control - File Transfer Protocol
  Line  1:              USER student<CR><LF>  [54-65]
```

FTP Login (NAT)

```
  Timestamp:            21:00:57.764403200 04/11/2014
Ethernet Type 2
  Destination:          3C:E5:A6:D0:**:**  HangzhouH3:D0:**:**  [0-5]
  Source:               F8:B1:56:B5:**:**  [6-11]
  Protocol Type:        0x0800  IP  [12-13]
IP Version 4 Header - Internet Protocol Datagram
...
  Fragment Offset:      0  (0 bytes)  [20-21 Mask 0x1FFF]
  Time To Live:         128  [22]
  Protocol:             6  TCP - Transmission Control Protocol  [23]
  Header Checksum:      0x0000  Checksum invalid. Should be:  0x0B26  [24-25]
  Source IP Address:    59.77.5.***  [26-29]
  Dest. IP Address:     59.77.7.25  [30-33]
TCP - Transport Control Protocol
  Source Port:          10405  [34-35]
  Destination Port:     21  ftp  [36-37]
  Sequence Number:      2633766987  [38-41]
  Ack Number:           300260607  [42-45]
  TCP Offset:           5  (20 bytes)  [46 Mask 0xF0]
...
FTP Control - File Transfer Protocol
  Line  1:              USER student<CR><LF>  [54-65]
```

FTP Response (NAT)

```
  Timestamp:            21:00:57.764979200 04/11/2014
Ethernet Type 2
  Destination:          F8:B1:56:B5:**:**  [0-5]
  Source:               3C:E5:A6:D0:**:**  HangzhouH3:D0:**:**  [6-11]
  Protocol Type:        0x0800  IP  [12-13]
IP Version 4 Header - Internet Protocol Datagram
...
  Fragment Offset:      0  (0 bytes)  [20-21 Mask 0x1FFF]
  Time To Live:         63  [22]
  Protocol:             6  TCP - Transmission Control Protocol  [23]
  Header Checksum:      0xB59D  [24-25]
  Source IP Address:    59.77.7.25  [26-29]
  Dest. IP Address:     59.77.5.***  [30-33]
TCP - Transport Control Protocol
  Source Port:          21  ftp  [34-35]
  Destination Port:     10405  [36-37]
  Sequence Number:      300260607  [38-41]
  Ack Number:           2633767001  [42-45]
  TCP Offset:           5  (20 bytes)  [46 Mask 0xF0]
...
FTP Control - File Transfer Protocol
  Line  1:              331 User name okay, need password.<CR><LF>  [54-87]
```

FTP Response (VMWare)

```
  Timestamp:            21:00:57.444794300 04/11/2014
Ethernet Type 2
  Destination:          00:0C:29:17:29:CA  VMware:17:29:CA  [0-5]
  Source:               00:50:56:FC:52:95  VMware:FC:52:95  [6-11]
  Protocol Type:        0x0800  IP  [12-13]
IP Version 4 Header - Internet Protocol Datagram
...
  Fragment Offset:      0  (0 bytes)  [20-21 Mask 0x1FFF]
  Time To Live:         128  [22]
  Protocol:             6  TCP - Transmission Control Protocol  [23]
  Header Checksum:      0xF389  [24-25]
  Source IP Address:    59.77.7.25  [26-29]
  Dest. IP Address:     192.168.7.4  [30-33]
TCP - Transport Control Protocol
  Source Port:          21  ftp  [34-35]
  Destination Port:     4425  netrockey6  [36-37]
  Sequence Number:      1171416600  [38-41]
  Ack Number:           1304971740  [42-45]
  TCP Offset:           5  (20 bytes)  [46 Mask 0xF0]
...
FTP Control - File Transfer Protocol
  Line  1:              331 User name okay, need password.<CR><LF>  [54-87]
```

11. THANK YOU.

厦门大学软件学院

黄炜 助理教授