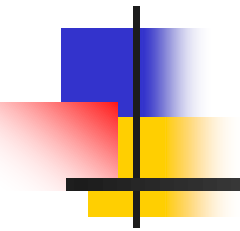


网络层：控制协议

刘志敏

liuzm@pku.edu.cn

问题回顾

- 地址分配：
  - IP地址，三种编址方式；
  - 如何分配IP地址？
  - IP地址数量不够如何解决？
- 隧道技术
- 分组传送
  - ARP：IP 地址到MAC的映射
  - 各段链路的帧长度不同，如何确定IP分组长度？
- 网络控制：超时控制、差错恢复、状态报告、拥塞检测与控制？
- 路由与转发：
  - RIP及距离矢量路由算法
  - 主机可以移动吗？

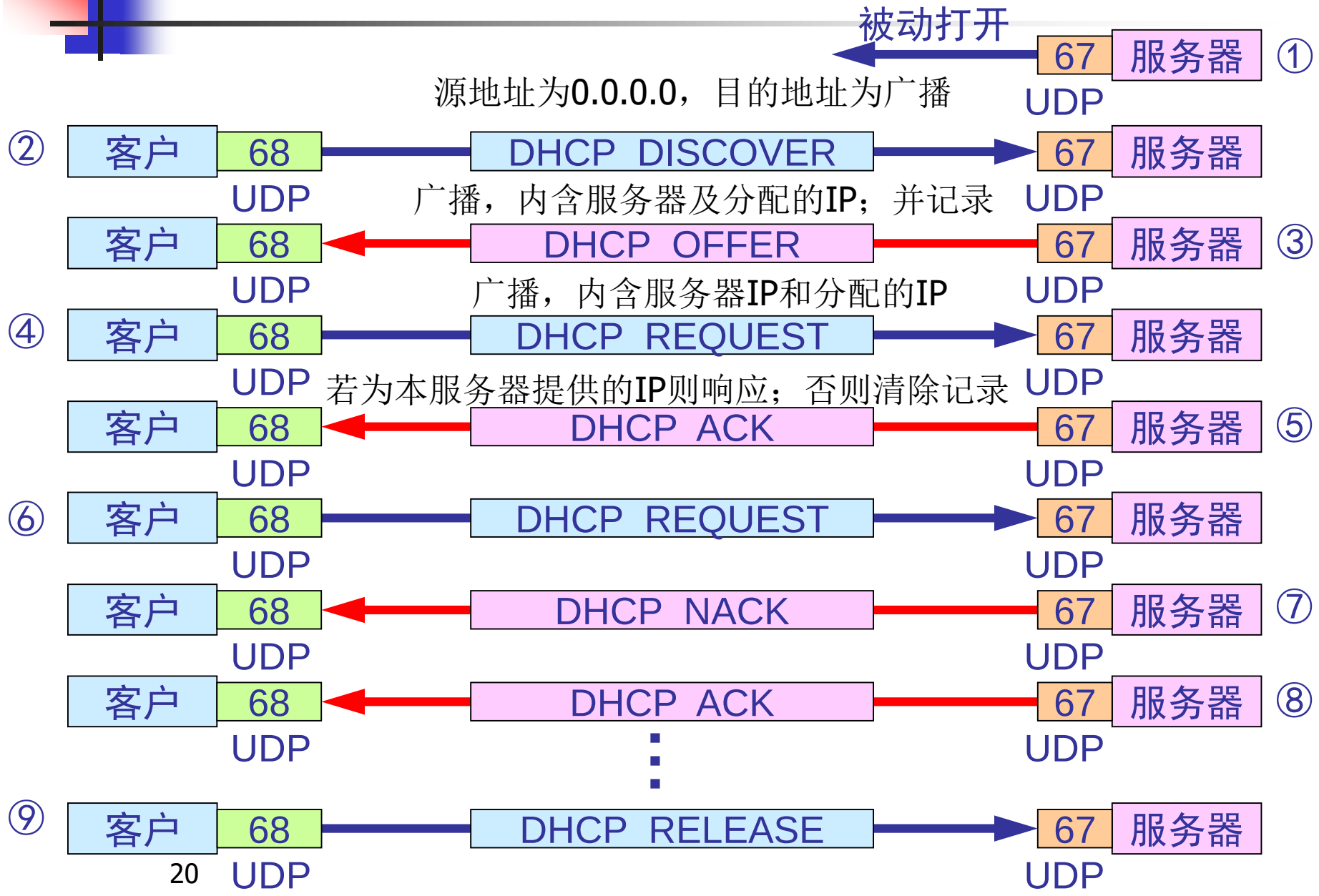
IP 地址

- IP地址是在互联网上标识主机（或路由器）的惟一方法
  - 分类IP地址
  - 子网划分
  - 构成超网
- 如何获得IP地址？
  - 静态地址：用于服务器及固定主机；
    - 如何保证唯一？网络管理员分配，人工配置
    - 如何方便配置及使用？IP，掩码，缺省路由器
    - 存在地址资源不足的问题
  - 动态地址：主机使用IP地址之前先申请
- 如何解决地址不足问题？
  - 动态分配
  - 专用地址，重复使用

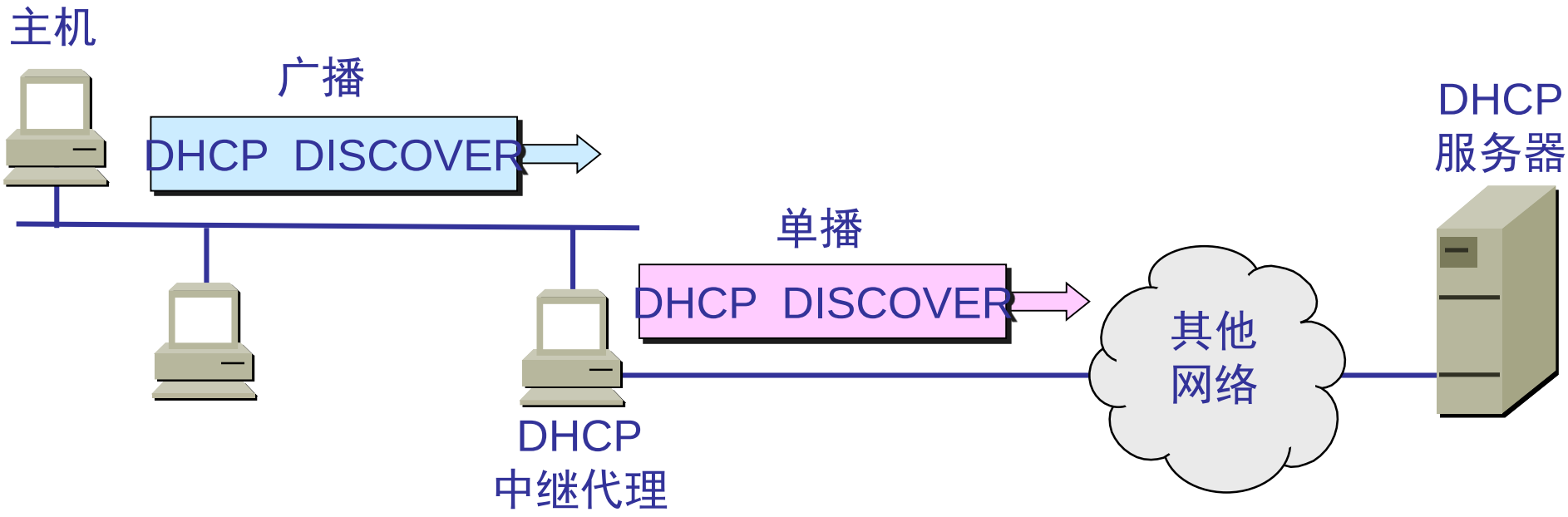
DHCP
(Dynamic Host Configuration Protocol)

- DHCP允许计算机加入新的网络并自动获取IP地址
- 主机启动时向DHCP服务器广播报文
  DHCP DISCOVER
- DHCP服务器以DHCPOFFER响应该报文
- DHCP服务器在其数据库中查找该计算机的配置信息。若找到，则返回信息；若找不到，则从IP地址缓存区取一个地址分配给该计算机

DHCP 协议的工作过程



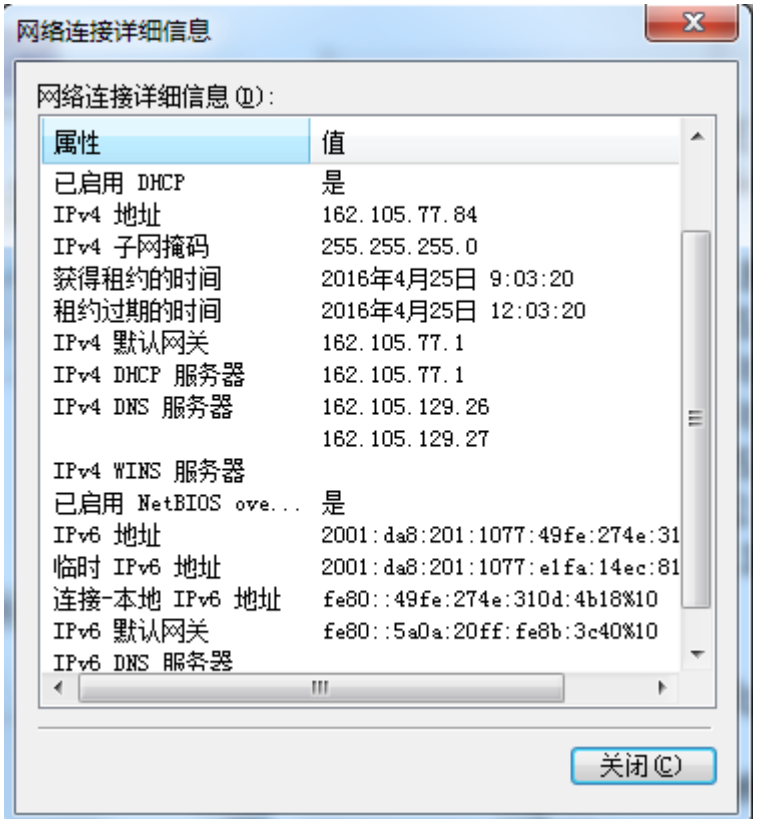
由DHCP服务器或DHCP代理以单播方式转发DHCP DISCOVER



注意：DHCP 报文只是 UDP 用户数据报中的数据

DHCP服务器

- 地址池开始和结束地址：设定地址池的范围
- 地址租期：分配给客户端的IP地址的有效使用时间
- 网关：路由器的IP地址
- DNS服务器：
  ISP提供的DNS服务器地址
- 例如，校园网中的DHCP



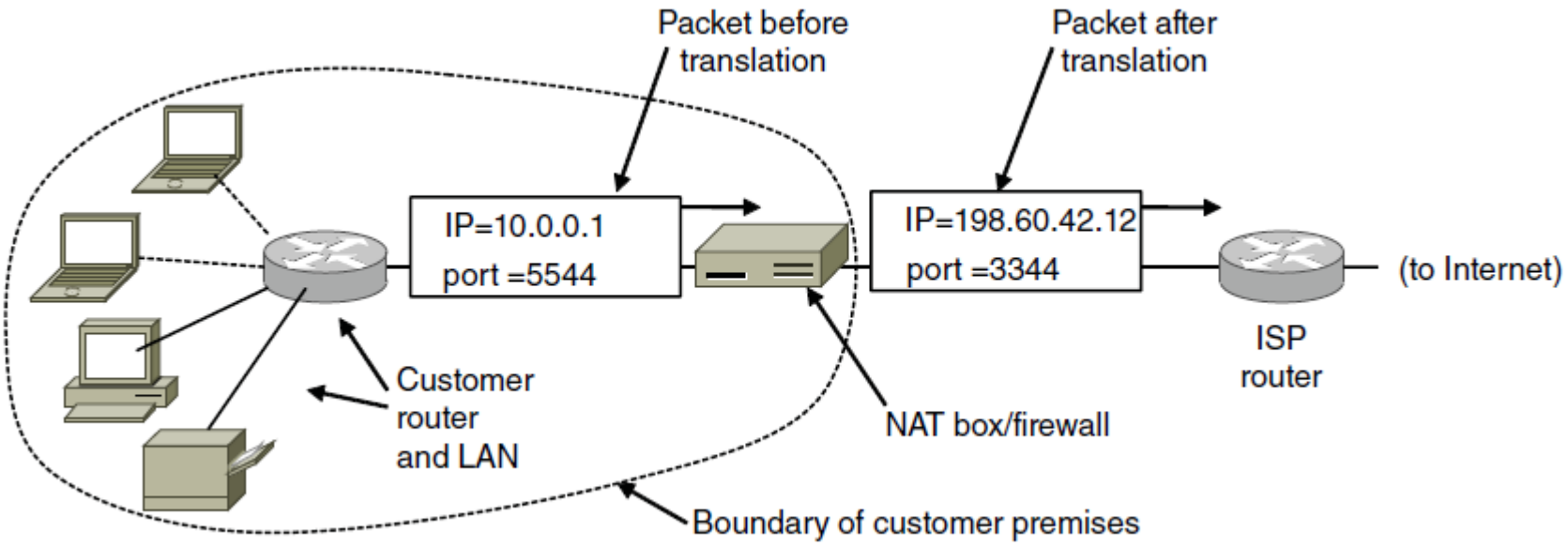
专用地址与虚拟专网VPN

- 全球地址：唯一的IP地址，须向ICANN（The Internet Corporation for Assigned Names and Numbers）申请
- 专用地址：无需申请，只能用于内部通信，路由器不转发目的地址是专用地址的分组
- RFC 1918 指明的专用地址(private address)
  - 10.0.0.0 到 10.255.255.255 （2^{24}个）
  - 172.16.0.0 到 172.31.255.255（2^{20}个）
  - 192.168.0.0 到 192.168.255.255（2^{16}个）
- 问题：使用专用地址如何与其他主机通信？
  - 在专用网内部，用专用路由器进行分组转发
  - 专用网与外网的通信，通过NAT路由器与互联网连接

网络地址转换NAT(Network Address Translation)

- NAT路由器，安装 NAT软件并至少有一个全球地址IP_G
- 专用地址经过NAT路由器后，将专用地址转换成 IP_G
- 网络地址的转换过程
  - 主机X用专用地址IP_X和主机Y通信，其数据报经过NAT路由器
  - NAT路由器将源地址IP_X转换成IP_G（利用IP_X及TCP端口得到NAT表索引值N），但目的地址IP_Y保持不变，N为新的TCP端口号，重新计算IP头及TCP头的检验和，然后发送到互联网
  - NAT路由器收到主机Y发回的数据报时，其源地址是IP_Y而目的地址是 IP_G
  - NAT路由器将目的地址IP_G转换为 IP_X（TCP端口号查表，得到原IP_G及TCP端口号），转发给内部主机X

NAT



NAT的位置及作用：将一个专用地址转换为一个全球地址，实现多个专用地址共享一个全球地址，缓解了IP地址资源紧缺的压力

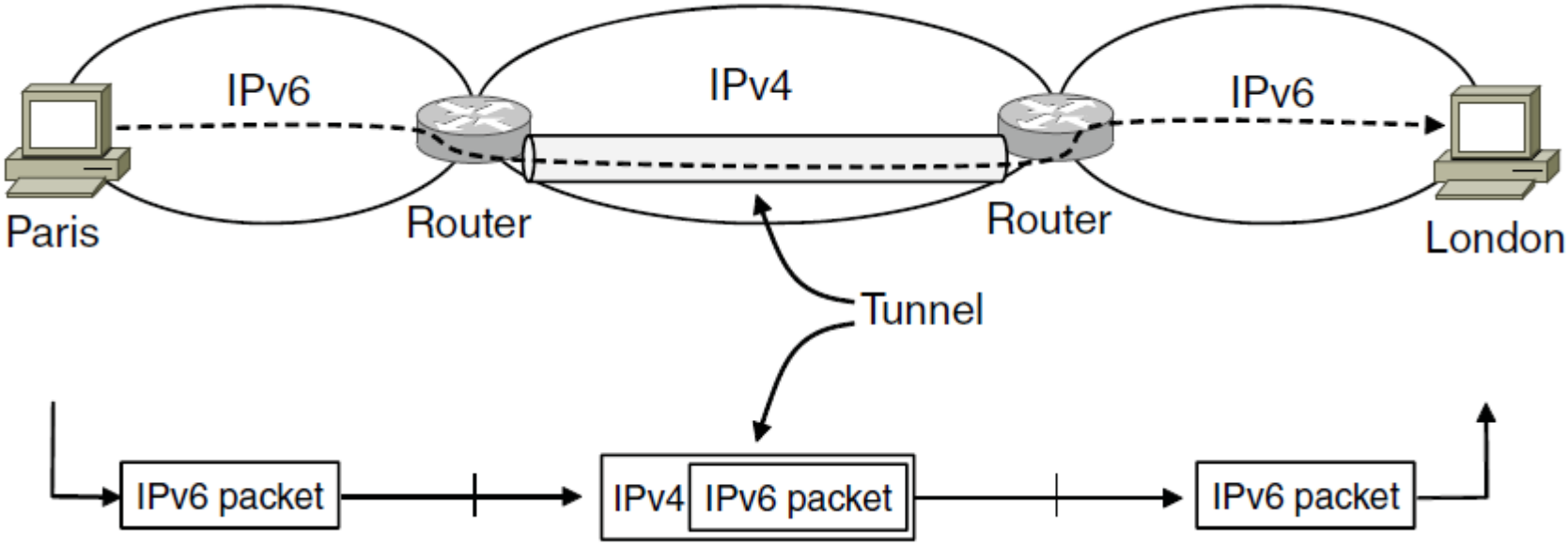
问题回顾

- 地址分配：
  - IP地址，三种编址方式；
  - 如何分配IP地址？
  - IP地址数量不够如何解决？
- 隧道技术
- 分组传送
  - ARP：IP 地址到MAC的映射
  - 各段链路的帧长度不同，如何确定IP分组长度？
- 网络控制：超时控制、差错恢复、状态报告、拥塞检测与控制？
- 路由与转发：
  - RIP及距离矢量路由算法
  - 主机可以移动吗？

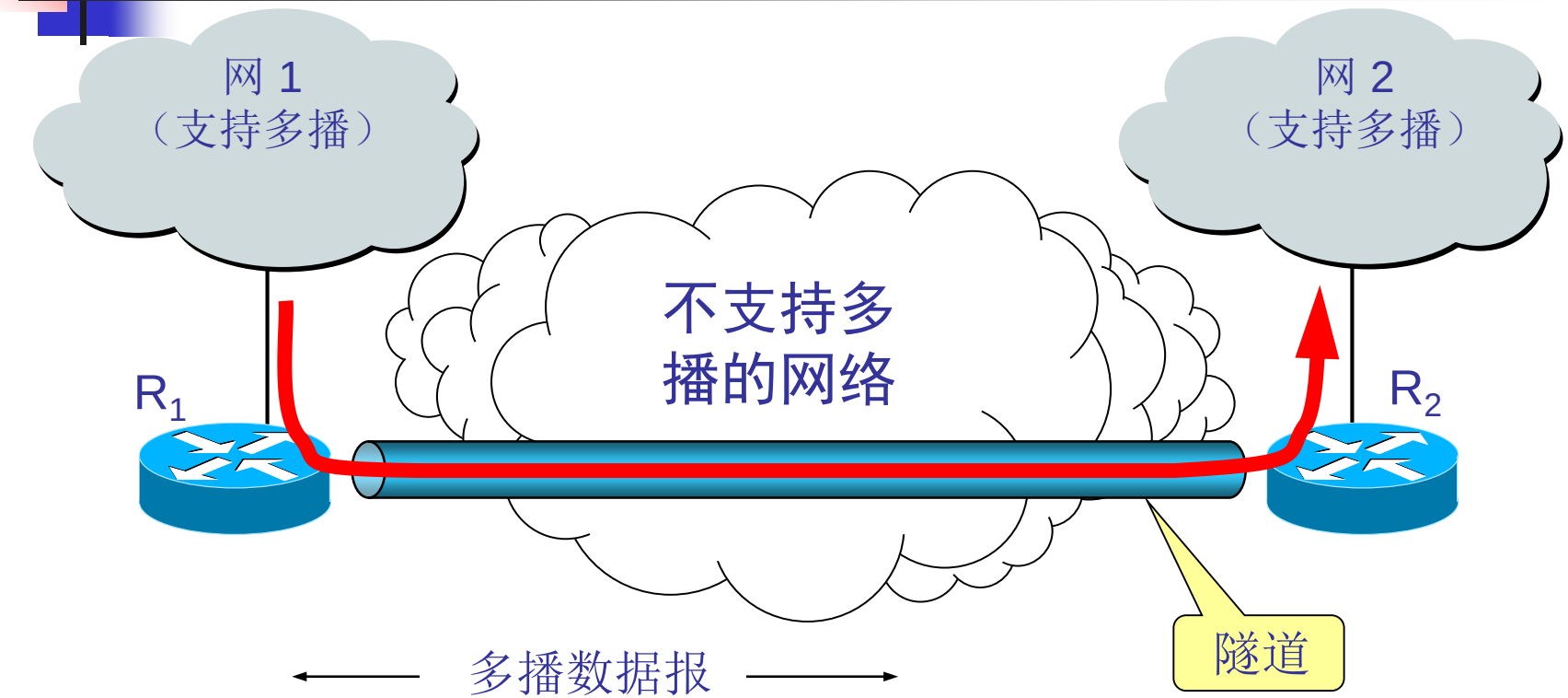
隧道技术(tunneling)

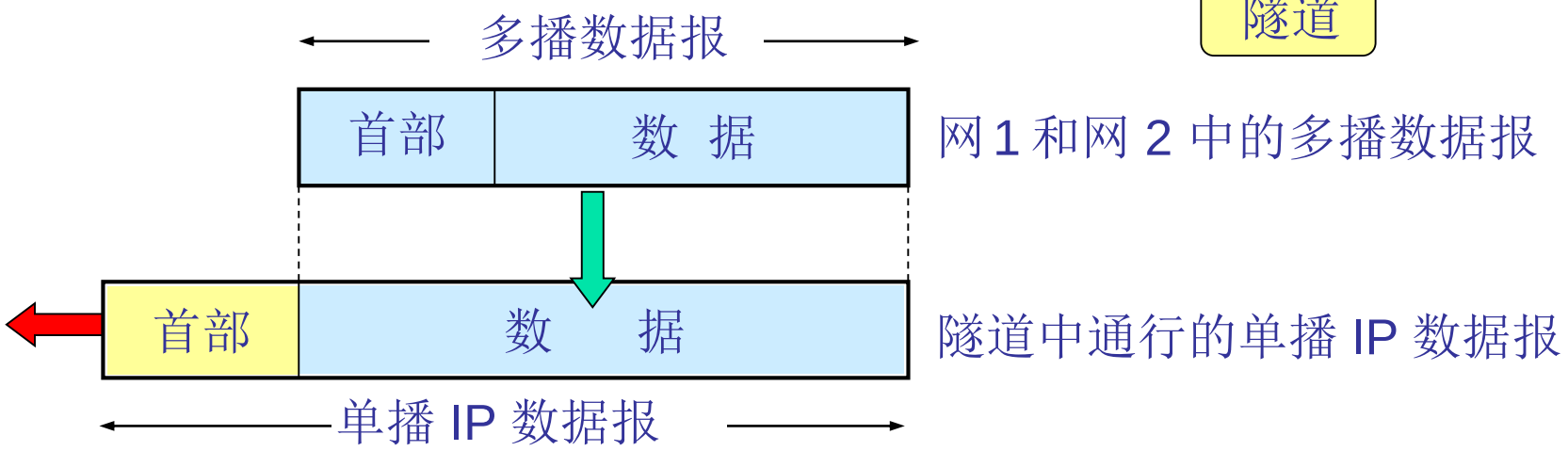
- 隧道是路由器把一种网络层协议封装到另一个协议中以跨过网络传送到另一个路由器的处理过程。发送路由器将被传送的分组进行封装，经过网络传送，接收路由器解开收到的分组，取出原始的分组；而在传输过程中的中间路由器并不在意封装的分组是什么。
- 隧道技术的应用
  - 虚拟专用网
  - 移动IP
  - 组播网络的互联
  - IPV4、IPV6互联

隧道技术，实现IPV6子网的互联



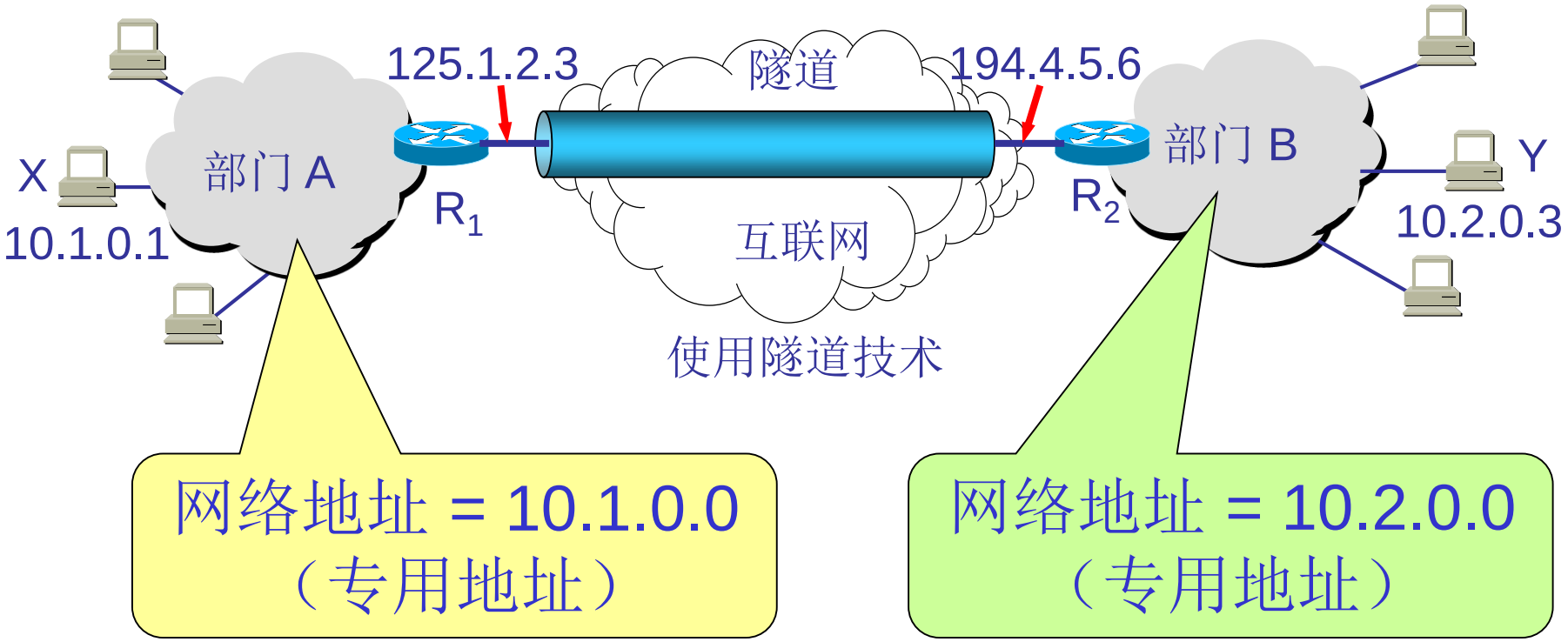
隧道技术(tunneling)



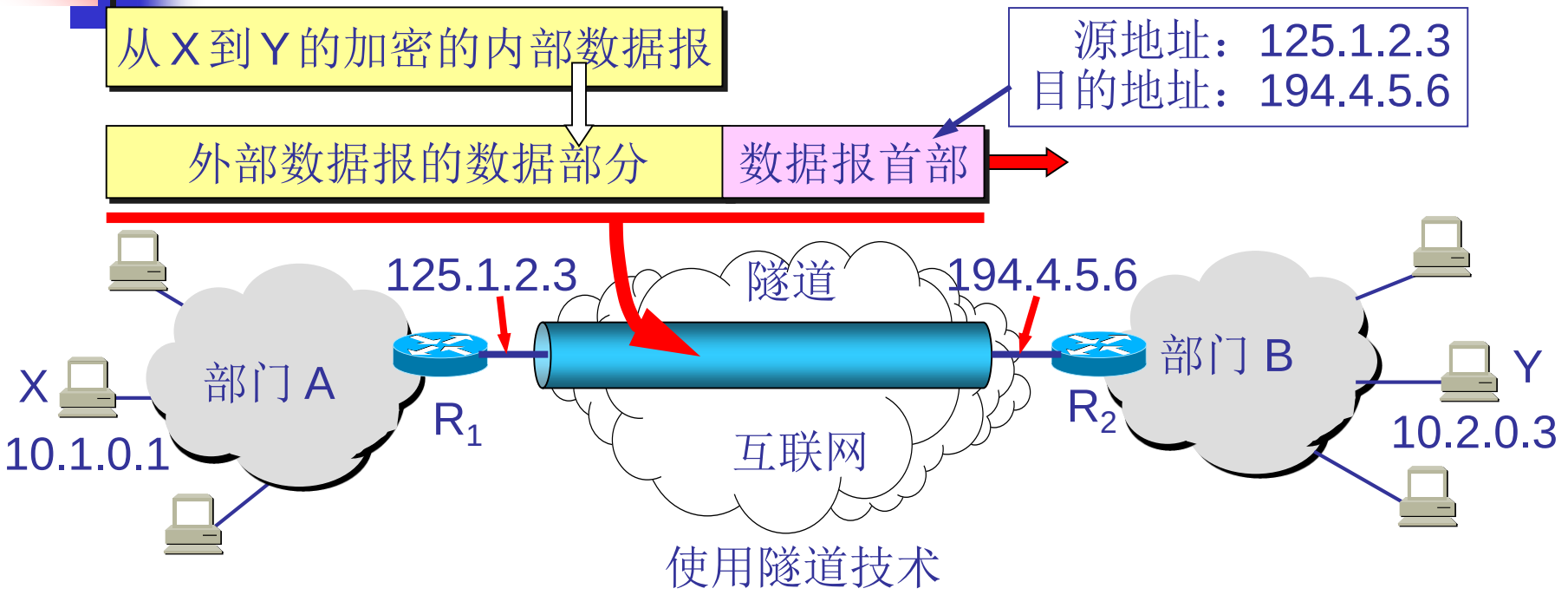


用隧道技术实现虚拟专用网

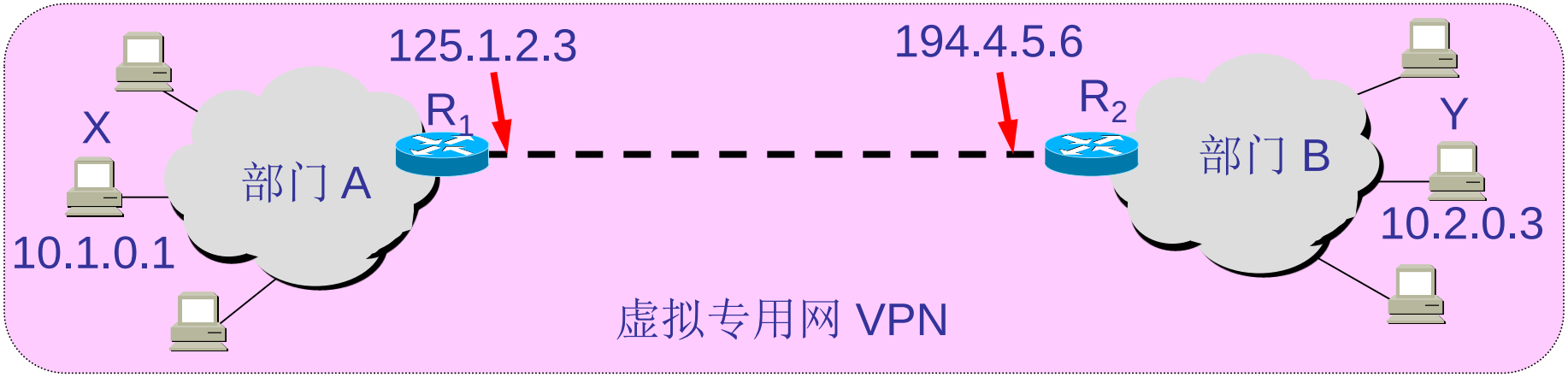
专用地址　　全球地址　　专用地址



用隧道技术实现虚拟专用网



使用隧道技术



虚拟专用网 VPN

问题回顾

- 地址分配：
  - IP地址，三种编址方式；
  - 如何分配IP地址？
  - IP地址数量不够如何解决？
- 隧道技术
- 分组传送
  - ARP：IP 地址到MAC的映射
  - 各段链路的帧长度不同，如何确定IP分组长度？
- 网络控制：超时控制、差错恢复、状态报告、拥塞检测与控制？
- 路由与转发：
  - RIP及距离矢量路由算法
  - 主机可以移动吗？

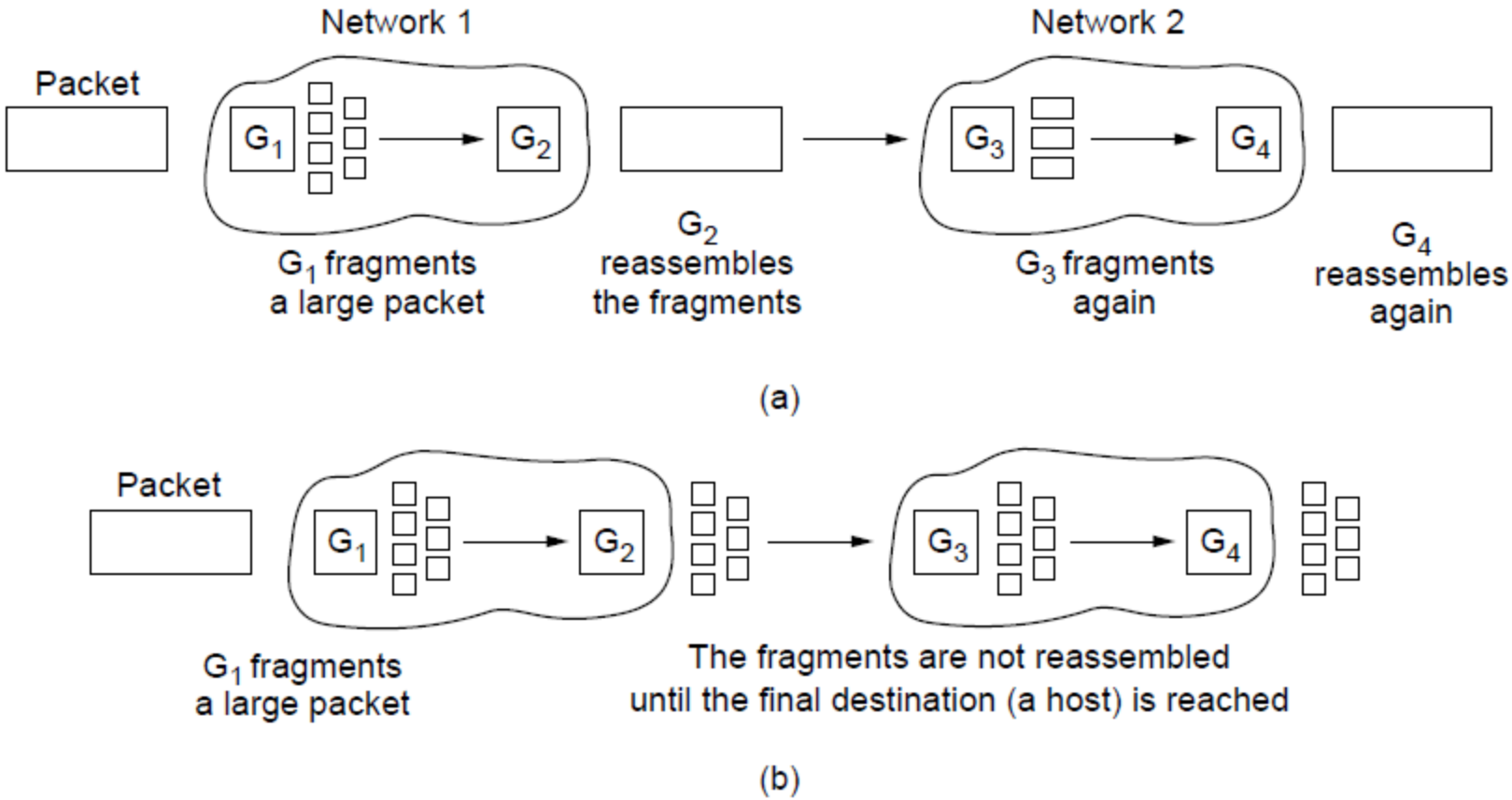
分组分段（1）

与分组长度相关的问题

MTU(Maximum Transmission Unit)：最大传输单元

1. 硬件：与链路层传输技术有关，例如LAN为1500B；WLAN为2312B
2. 操作系统
3. 协议
4. 与标准的兼容性
5. 期望降低传输分组出错的概率以及重传次数
6. 避免分组占用信道时间太长

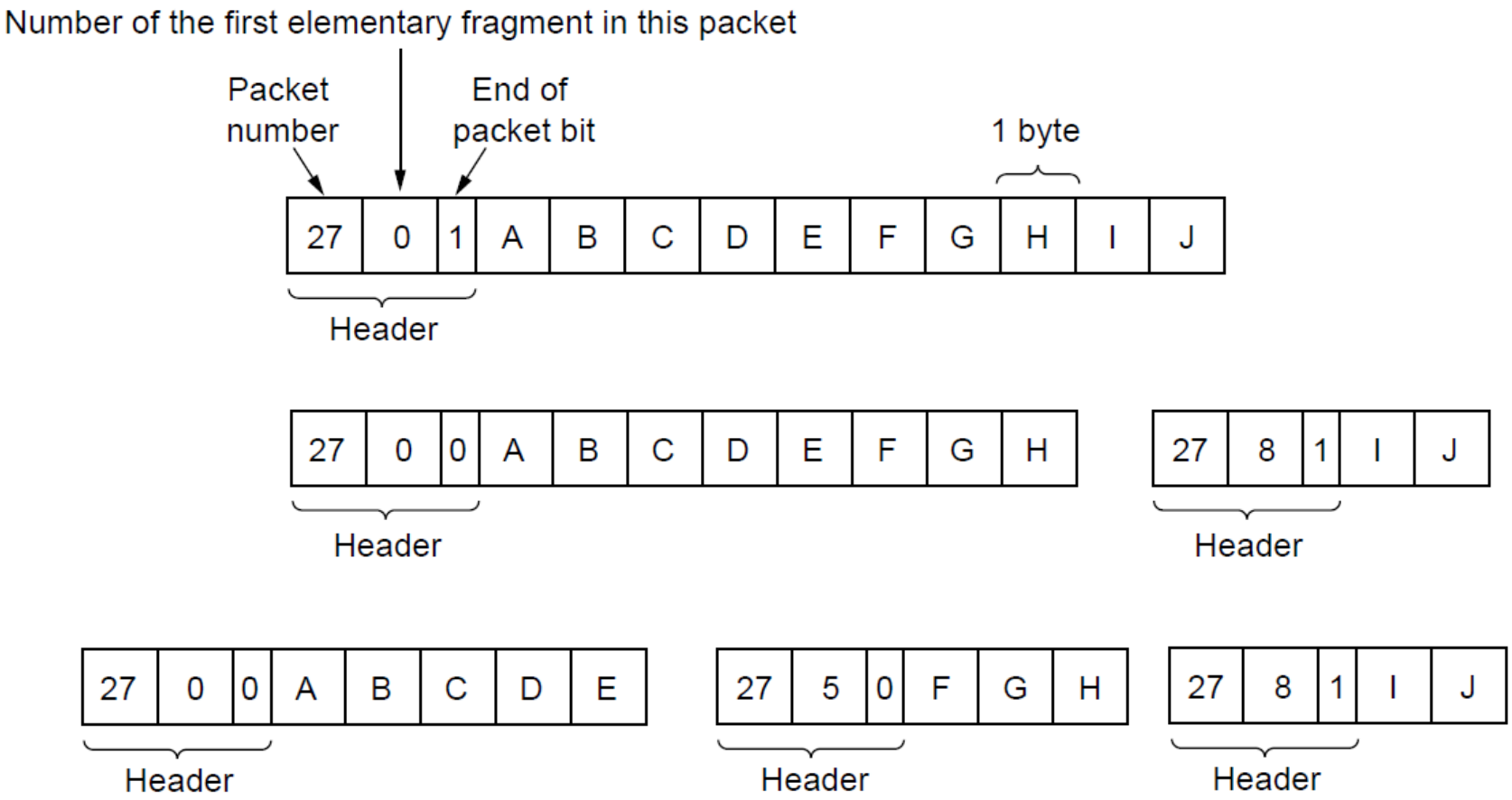
分组分段(2)



(a) 透明分段：路由器分段并重组；路由器的开销很大！

(b) 非透明分段：路由器对传输超过MTU的分组分段，目的主机负责重组；增加了传输的分组数，需要更多的传输时间！

分组分段(3)



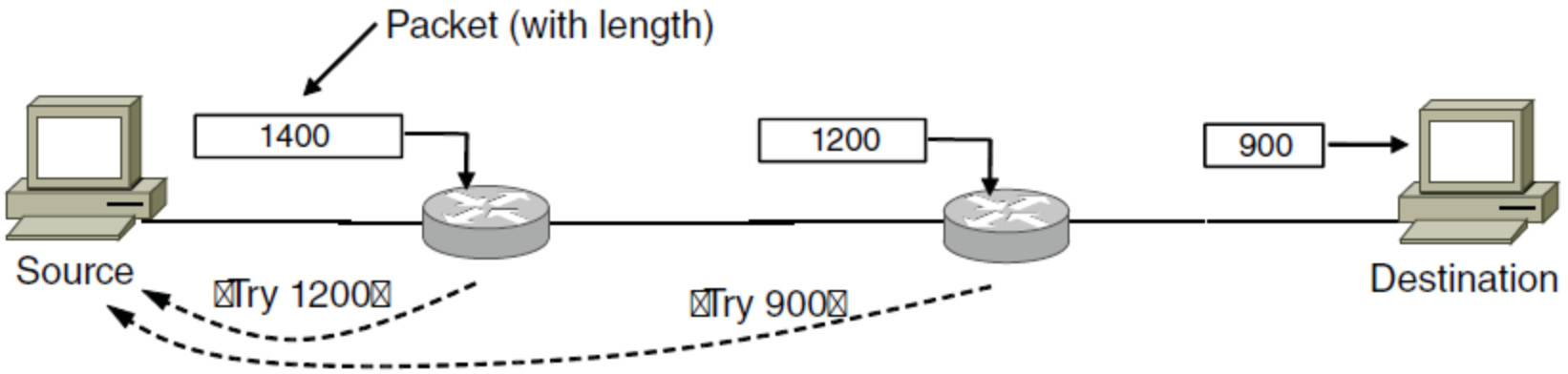
(1)原分组含10个字节的数据，头中含有起始位置及结束标志

(2)经过MTU=8的网络分段后，加上头

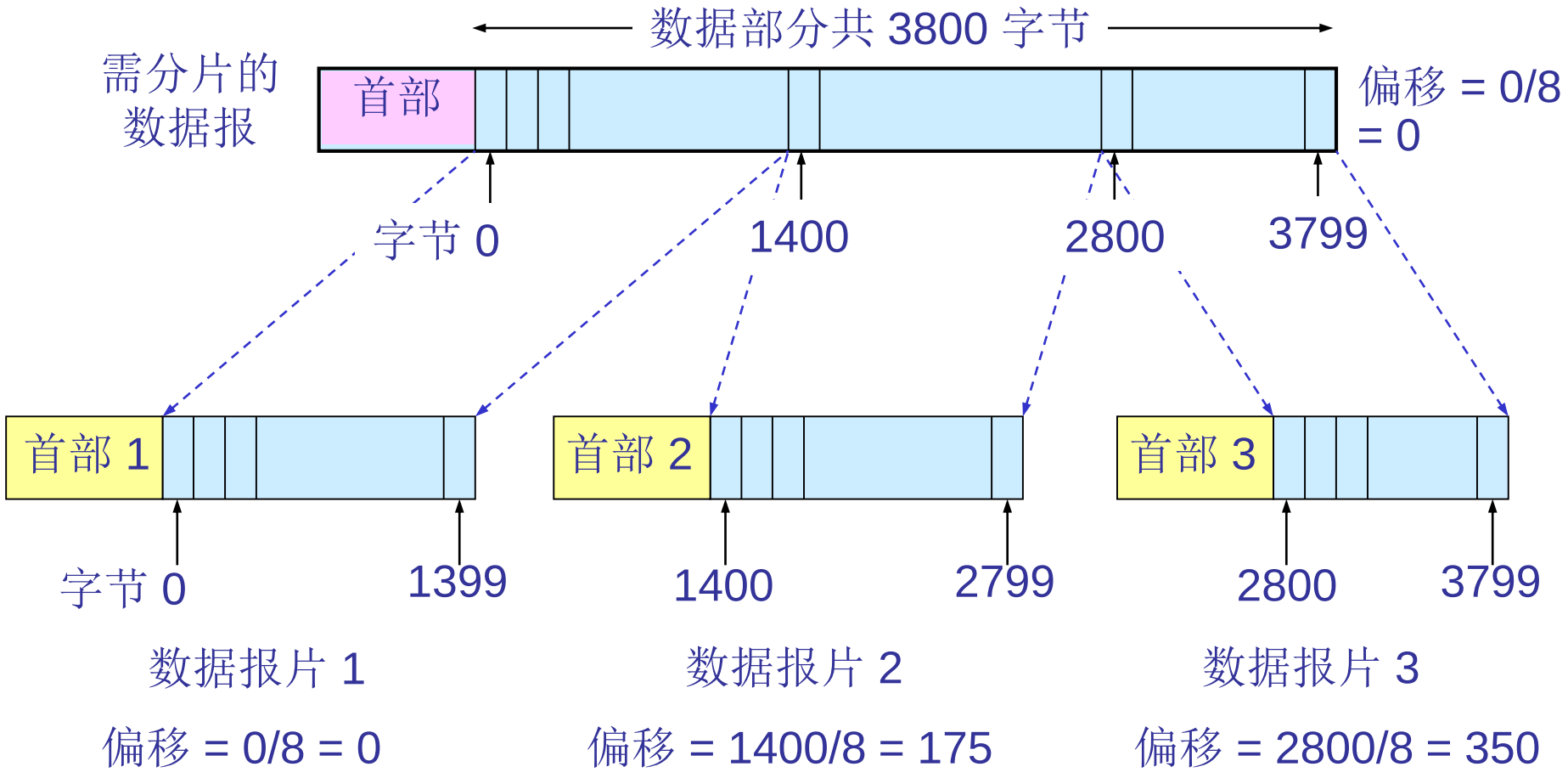
(3)再经过MTU=5的网络后的分段

分组分段(4)

- 避免分段：需要发现路径上的MTU；随着跳数的增加；尝试的次数也增加了



IP数据报分片：采用非透明分段



IP 首部

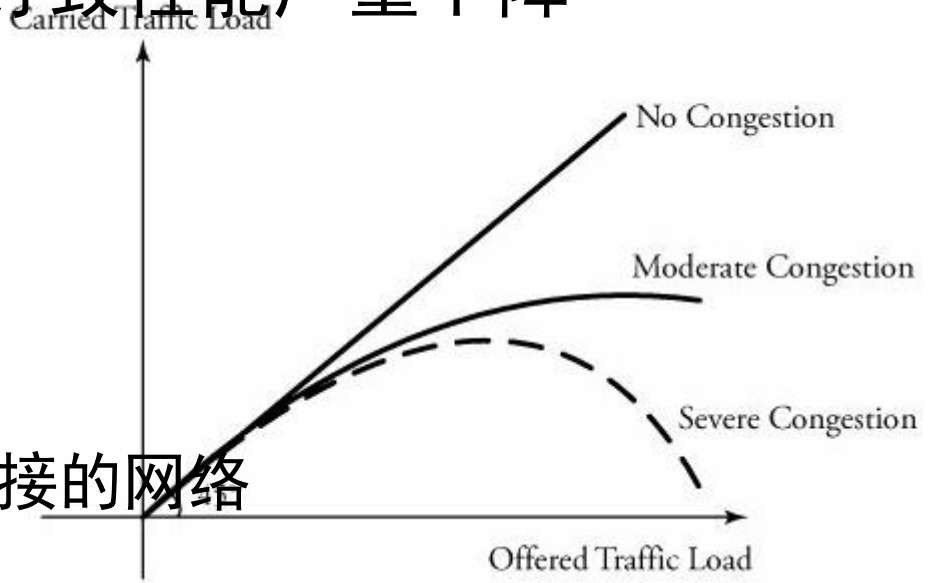
- 标志(flag)占 3 位，前2位有意义。最低位是 **MF** (More Fragment)。MF = 1 “还有分片”。MF = 0 最后一个分片。中间一位是 **DF** (Don‘t Fragment)，当 DF = 0 时才允许分片
- 片偏移(13 位)：给出在分片后某片在原分组中的相对位置；片偏移以 8 个字节为偏移单位。

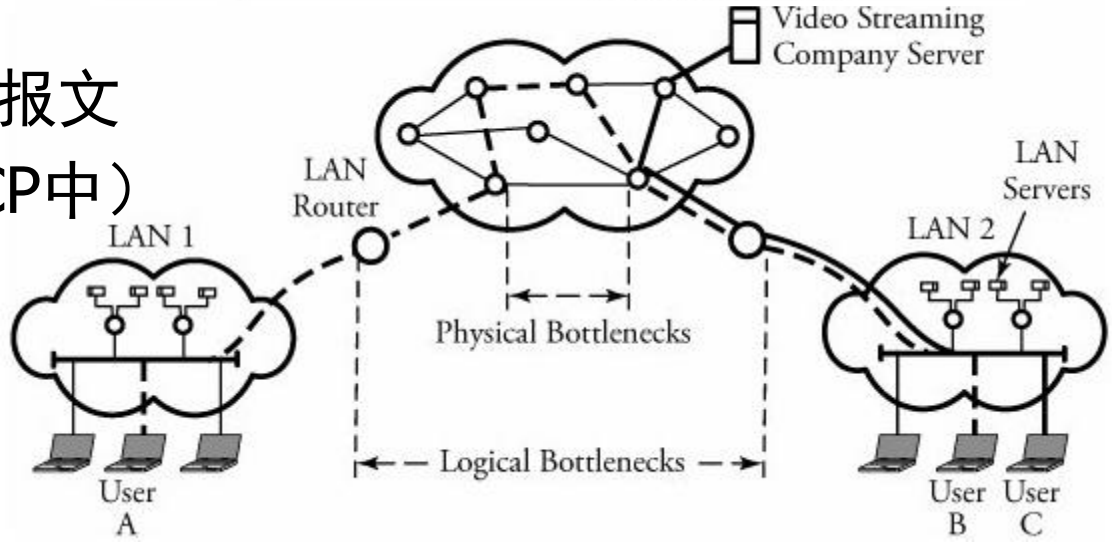
问题回顾

- 地址分配：
  - IP地址，三种编址方式；
  - 如何分配IP地址？
  - IP地址数量不够如何解决？
- 隧道技术
- 分组传送
  - ARP：IP 地址到MAC的映射
  - 各段链路的帧长度不同，如何确定IP分组长度？
- 网络控制：超时控制、差错恢复、状态报告、拥塞检测与控制？
- 路由与转发：
  - RIP及距离矢量路由算法
  - 主机可以移动吗？

网络拥塞控制

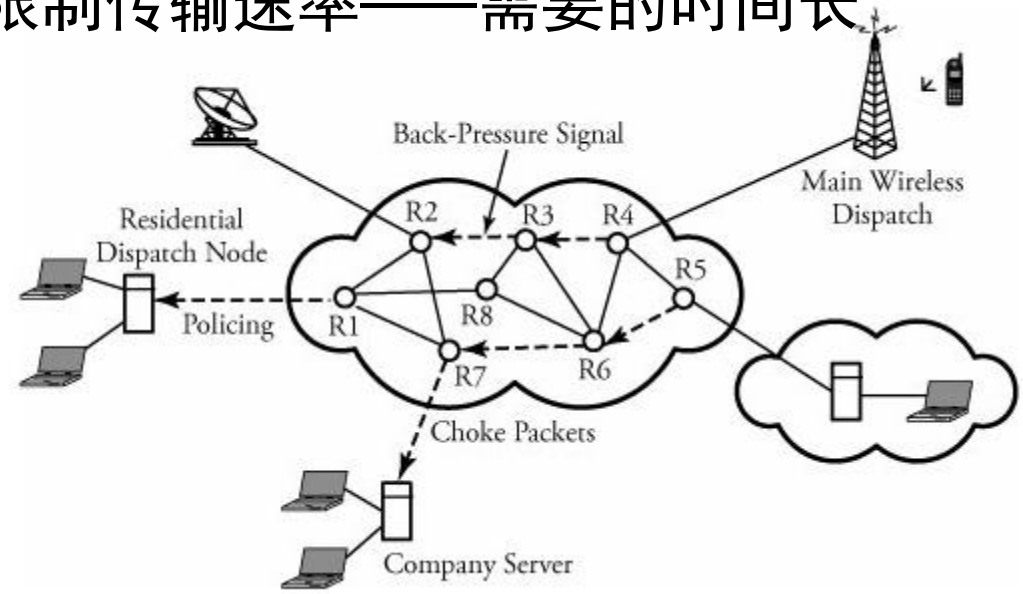
- 拥塞：因网络过载而导致性能严重下降
- 拥塞的现象：
  - 信道的利用率
  - 缓存队列长度
  - 分组丢失率
- 准入控制：适于面向连接的网络
- 流量调节
  - 向源主机发送抑制报文
  - 显式拥塞通知（TCP中）
  - 逐跳后压
  - 随机早期检测RED





流量调节方法

- 向源主机发送抑制报文
  - 路由器产生抑制报文，沿数据流的反向传送到源节点，源节点限制传输速率——需要的时间长



- 逐跳后压
  - 在拥塞的沿路传输抑制报文，路由器减缓发送速率；快速缓解拥塞，但消耗路由器的缓存空间

流量调节方法

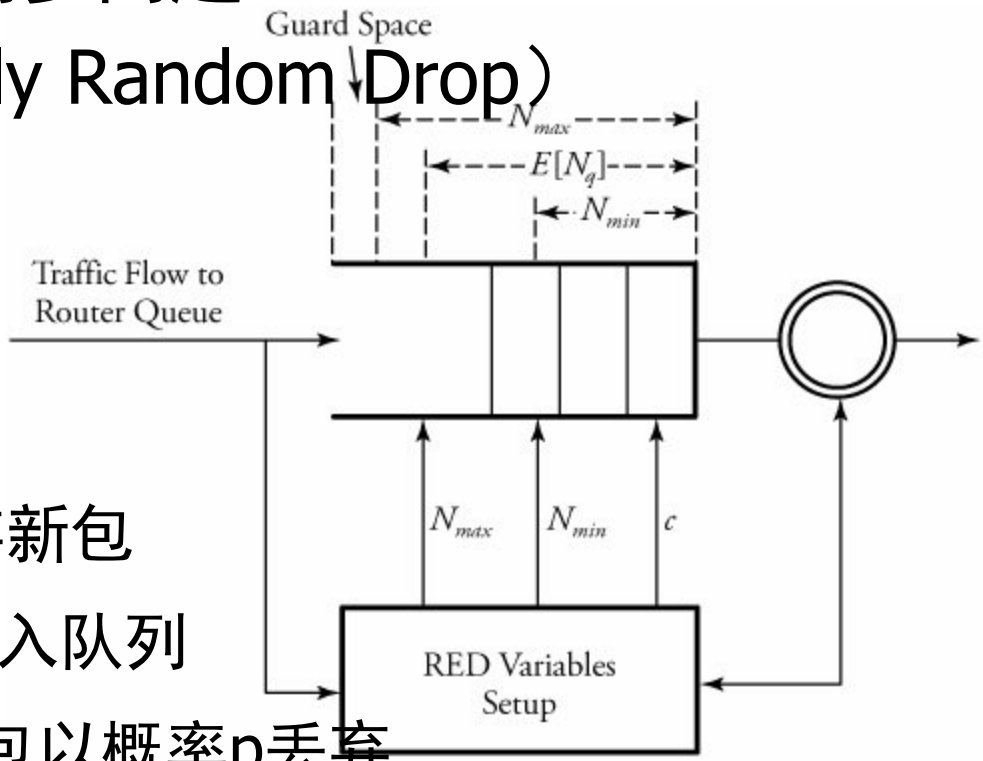
- 对缓存的尾部丢包：连续丢多个包，导致业务性能下降，以及TCP连接同步问题
- 随机早期检测RED（Early Random Drop）
  - 计算平均队列长度

    $E[N_q]=(1-\alpha)E[N_q]+\alpha N_i$

    N_i为队列长度，

    α为权重因子，$0<\alpha<1$
  - 若$E[N_q]>N_{max}$则直接丢弃新包
  - 若$E[N_q]<N_{min}$则新包直接入队列
  - 若$N_{min}<E[N_q]<N_{max}$则新包以概率p丢弃，

    $p=\delta/(1-c\delta)$　$\delta=(E[N_q]-N_{min})/(N_{max}-N_{min})$



互联网报文控制协议 ICMP

- 互联网报文控制协议 ICMP (Internet Control Message Protocol)
- ICMP允许主机或路由器报告差错情况和提供异常报告
- ICMP报文作为IP层数据报的数据，加上数据报的首部，组成 IP分组发送

ICMP 报文

- 两大类： ICMP 差错报告报文和ICMP 询问报文
- ICMP 差错报告报文
  - 目的地不可达，路由器无法转发分组时产生
  - 超时，路由器转发分组时TTL-1，当减为0时产生
  - 参数问题，无效的IP头字段
  - 源站抑制，路由器检测到拥塞时产生
  - 重定向，告知路由器有关错误的路由信息
- ICMP 询问报文
  - 回显请求和回显应答，检查一台机器是否工作
  - 时间戳请求和应答，与回显相同，但含有时间戳
  - 路由器询问和通告报文，用于发现附件的路由器

Traceroute 的应用举例

- 利用ICMP超时报文，其TTL依次为1，2，3等，测试主机的路由及时间开销

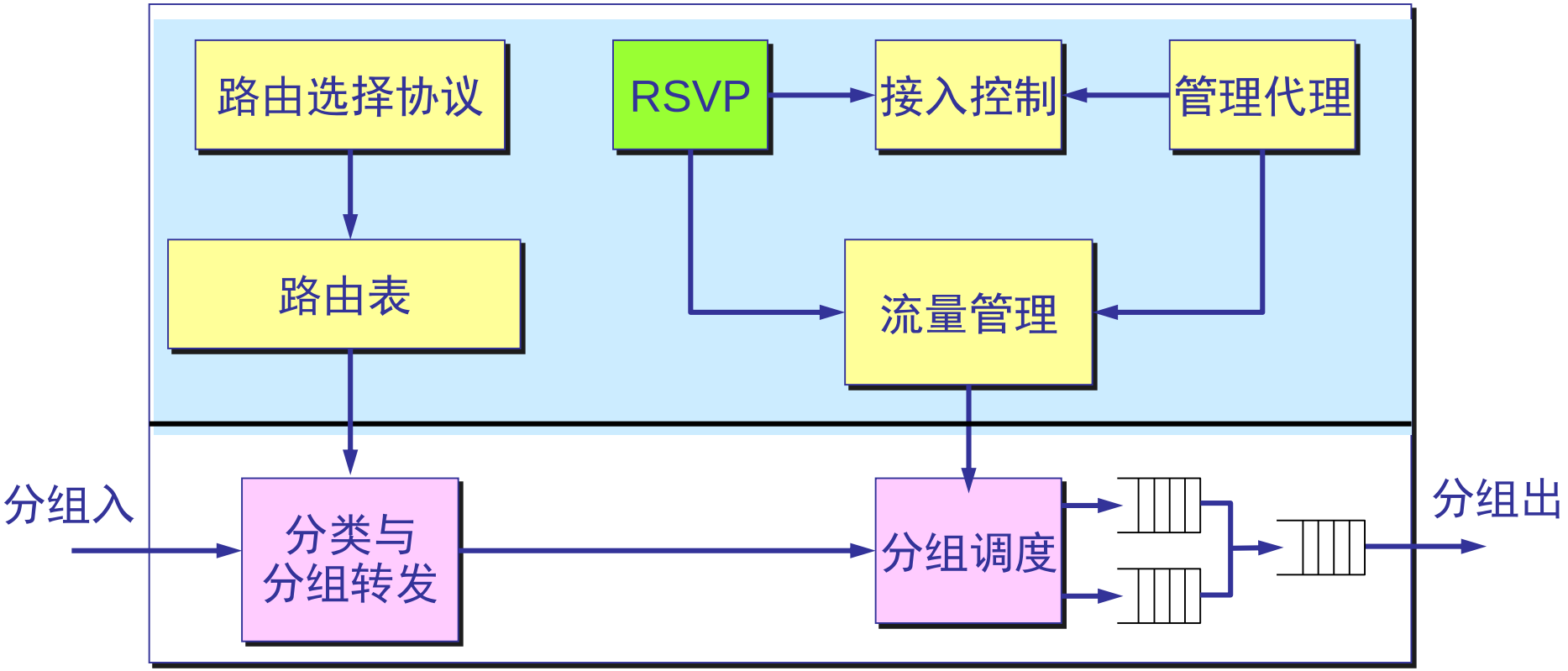
```
C:\Documents and Settings\XXR>tracert mail.sina.com.cn

Tracing route to mail.sina.com.cn [202.108.43.230]
over a maximum of 30 hops:

  1    24 ms    24 ms    23 ms  222.95.172.1
  2    23 ms    24 ms    22 ms  221.231.204.129
  3    23 ms    22 ms    23 ms  221.231.206.9
  4    24 ms    23 ms    24 ms  202.97.27.37
  5    22 ms    23 ms    24 ms  202.97.41.226
  6    28 ms    28 ms    28 ms  202.97.35.25
  7    50 ms    50 ms    51 ms  202.97.36.86
  8   308 ms   311 ms   310 ms  219.158.32.1
  9   307 ms   305 ms   305 ms  219.158.13.17
 10   164 ms   164 ms   165 ms  202.96.12.154
 11   322 ms   320 ms  2988 ms  61.135.148.50
 12   321 ms   322 ms   320 ms  freemail43-230.sina.com [202.108.43.230]

Trace complete.
```

PING 的应用举例

```
C:\Documents and Settings\XXR>ping mail.sina.com.cn

Pinging mail.sina.com.cn [202.108.43.230] with 32 bytes of data:

Reply from 202.108.43.230: bytes=32 time=368ms TTL=242
Reply from 202.108.43.230: bytes=32 time=374ms TTL=242
Request timed out.
Reply from 202.108.43.230: bytes=32 time=374ms TTL=242

Ping statistics for 202.108.43.230:
    Packets: Sent = 4, Received = 3, Lost = 1 (25% loss),
Approximate round trip times in milli-seconds:
    Minimum = 368ms, Maximum = 374ms, Average = 372ms
```

网络控制算法

路由选择：根据业务服务质量的需求选择路由
资源分配：在路由器上预留资源
接入控制：接受或拒绝接入业务流
流量管理：监控及管理流量
分组调度：为发送的分组设置时间节点



小结

- 地址分配：
  - IP地址，三种编址方式；
  - DHCP
  - 专用地址、NAT
- 隧道技术
- 分组传送：ARP、分段与重组
- 网络控制：超时、差错、状态报告、拥塞检测与控制
- 路由与转发：
  - 单播路由算法
  - 移动主机路由
  - 组播业务及其路由
  - 自组织网络路由：动态路由