网络层

概论

网络层的功能：**转发**，**路由**

网络层相关协议

- 路由选择
  - RIP，OSPF，BGP
- IP协议
- ICMP协议

数据平面

决定路由器的输入端口到达的分组如何转发到输出端口

控制平面

通过路由算法决定端到端路径

网络层的连接和无连接服务

- 数据报网络
  - 提供网络无连接服务
  - 网络层没有建立连接的过程
  - 转发使用目的IP地址，可能沿着不同路径向前转发
- VC网络
  - 提供网络面向连接的服务
  - 性能较好
  - VC的组成
    - 源到目的主机的路径
    - 路径上每段链路一个号码
    - 路由表中的向前转发表项
  - 每个VC路由都维护连接状态信息

传统方式

每一路由器中的单独路由算法元件，在控制平面进行交互

路由和转发相互作用：

- 路由算法决定端到端路径

- IP根据转发表决定IP数据报在此路由器上的局部转发

SDN模式

逻辑集中的控制平面

由一个远程控制器和本地代理交互

路由器组成

路由器结构概况

路由：运行路由选择算法/协议（RIP，OSPF，BGP）
转发：从输入到输出链路交换数据报，根据路由表进行分组的转发

输入端口功能

- 分布式交换
  - 基于目标的转发：仅依赖IP数据报的目标IP地址
  - 通用转发：基于头部字段的任意集合进行转发

采用**最长前缀匹配**目标的地址表项（硬件完成）

输入端口缓存

当交换机构的速率小于输入的汇聚速率时需要排队

存在**排队延迟**，当**输入缓存溢出**会造成丢失

交换结构

通过内存交换

第一代路由器采用

在CPU直接控制下的交换

通过总线交换

数据报共享总线，从输入端口转发到输出端口

总线竞争：交换速度受限于总线带宽

一次处理一个分组

通过互联网络交换

同时并发转发多个分组，克服总线带宽限制

输出端口

当数据报从交换机构到达速率比传输速率快就需要**输出端口缓存**

由调度规则选择排队的数据报进行传输

存在**排队延迟**，当**输出缓存溢出**会造成丢失**

调度机制

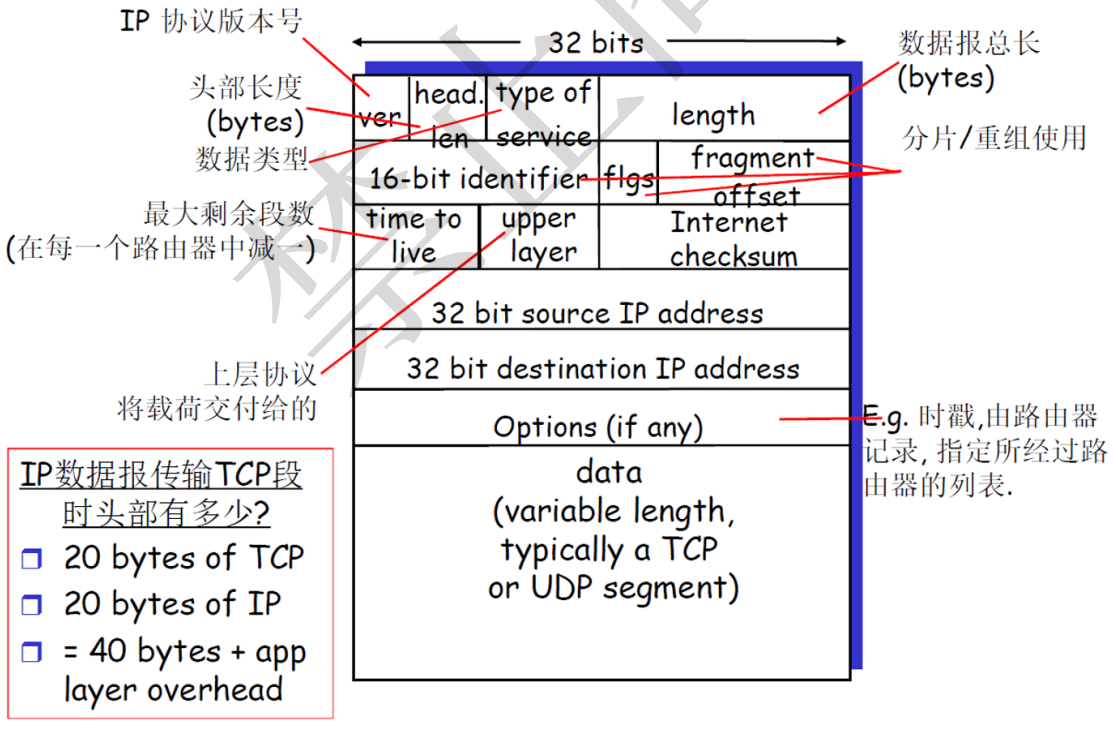
FIFO：使用先进先出策略
优先权调度：发送优先级高的分组

丢弃策略：

- 丢弃刚到达的分组
- 根据优先权移除分组
- 随机丢弃

IP协议

IP数据报格式

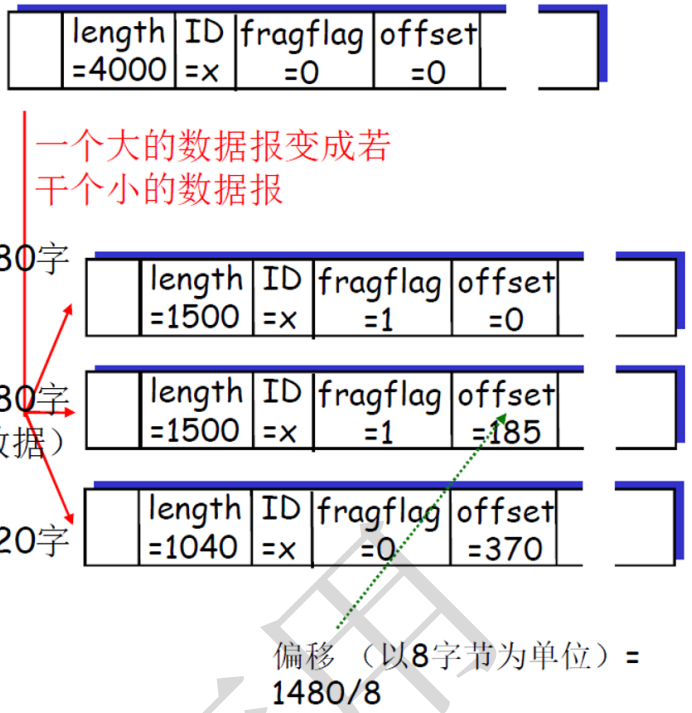


头部的40字节在一些题目中隐含

IP分片和重组

例子

- 4000 字节数据报
  - 20字节头部
  - 3980字节数据
- MTU = 1500 bytes
- 第一片：20字节头部+1480字节数据
  - 偏移量：0
- 第二片：20字节头部+1480字节数据（1480字节应用数据）
  - 偏移量：1480/8=185
- 第三片：20字节头部+1020字节数据（应用数据）
  - 偏移量：2960/8=370



子网

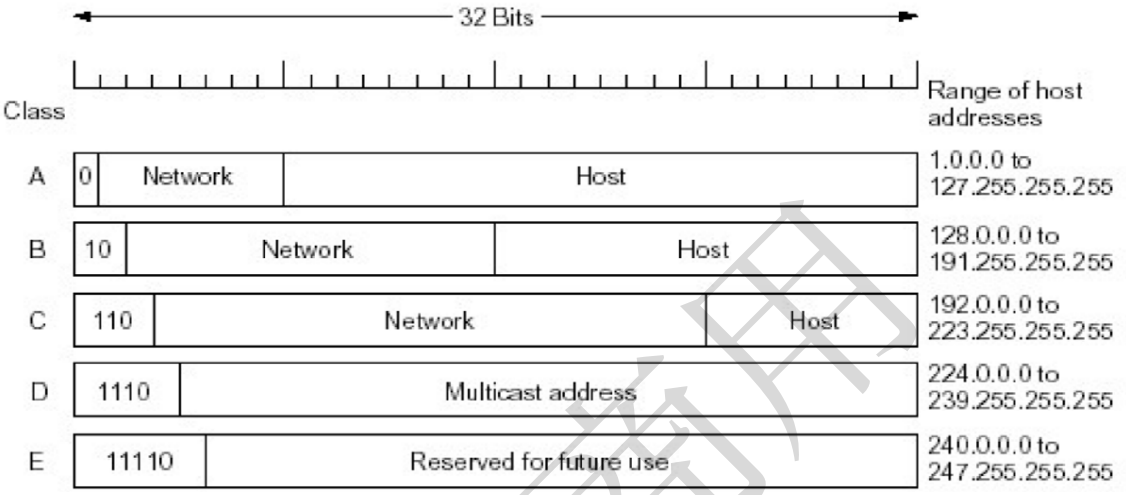
IP地址：

- 子网部分（高位）
- 主机部分（低位）

子网内的节点拥有相同的IP地址的高位部分，子网内部主机物理上相互到达**无需路由器介入**

IP地址的分类

- Class A：126 networks， 16 million hosts
- Class B：16382networks， 64 K hosts
- Class C：2 million networks， 254 host
- Class D：multicast
- Class E：reserved for future



注意，全1和全0是保留IP地址，不能分配。

子网部分全为0：本网络
主机部分全为0：本主机
主机部分全为1：广播地址，这个网络的所有主机

内网专用IP地址

- Class A 10.0.0.0-10.255.255.255 MASK 255.0.0.0
- Class B 172.16.0.0-172.31.255.255 MASK 255.255.0.0
- Class C 192.168.0.0-192.168.255.255 MASK 255.255.255.0

永远不会当作公网地址来分配，不会与公用地址重复

IP编址

CIDR（无类域间路由）

- 子网部分可以在任意位置
- 地址格式a.b.c.d/x，x是指子网号的长度

DHCP（动态主机配置协议）：从服务器动态获得一个IP地址

- 工作流程
  - 主机广播"DHCP discover"
  - 服务器用"DHCP offer"提供报文响应
  - 主机请求IP地址，发送"DHCP request"
  - DHCP发送地址"DHCP ack"

DHCP返回的信息：

- IP地址
- 第一跳路由器地址（网关）
- DNS服务器的域名和IP地址
- 子网掩码

DHCP请求使用UDP

NAT(网络地址转换)

使用一个公网IP使得局域网中的所有设备都可以上网

外出数据包：替换源端口和目的端口为NAT IP地址和新的端口号，目标IP和端口不变

每个替换对都会记录在NAT转换表中

进入数据包：替换目标IP地址和端口号，使用NAT表中的mapping表项

ICMP（互联网控制报文协议）

用于主机和路由器交换网络层信息，返回错误报告和错误代码

IPV6

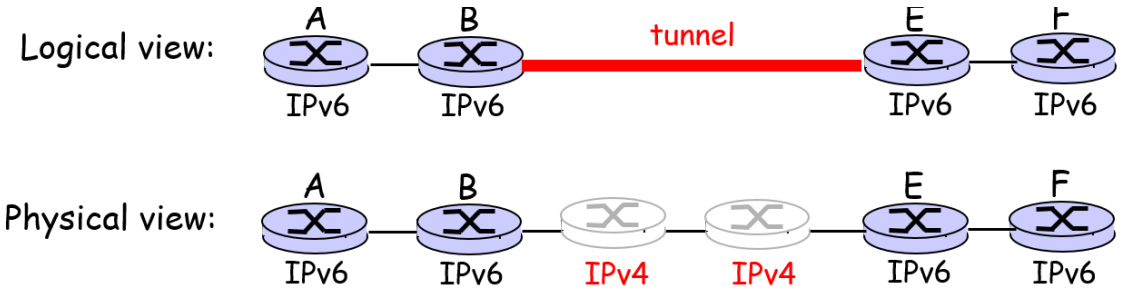
解决IPV4地址不够用，并且加速处理和转发

固定的40字节头部，并且传输过程中不允许分片

使用ICMP的新版本，Checksum被移除

IPV4的兼容问题

隧道：在IPV4路由器之间传输的IPV4数据报中携带IPV6数据报



SDN（软件定义网络）

控制平面和数据平面分离的优势：

- 水平集成控制平面的开放实现
- 集中式实现控制逻辑，网络管理容易
- 允许“可编程的”分组交换机

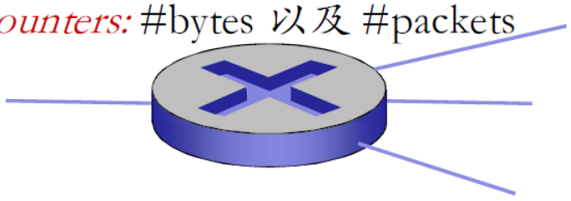
每一个路由器包含一个流表

特点

- 基于流的匹配+行动
- 控制平面和数据平面分离
- 控制平面功能在数据交换设备之外实现
- 可编程控制应用

❒ 通用转发：简单的分组处理规则

- 模式：将分组头部字段和流表进行匹配
- 行动：对于匹配上的分组，可以是丢弃、转发、修改、将匹配的分组发送给控制器
- 优先权Priority：几个模式匹配了，优先采用哪个，消除歧义
- 计数器Counters：#bytes 以及 #packets



路由器中的流表定义了路由器的匹配+行动规则
（流表由控制器计算并下发）