引言: 历史和组织*

历史：

1. 1960 – 1964：Baran 的自适应分组交换网络
2. 1969 ARPANET
3. 70年代： X.25 分组交换网 专用的网络体系结构：SNA，DNA；
4. 1979 TCP/IP
5. 80年代： ISO/OSI， LAN 空前发展， Internet 初具规模
6. 90年代: Internet 商业化, Web 技术, WWW
7. 今天的互联网：健壮性、适应性和互联程度都下降了（例如NAT）

IETF 制定了 RFC

网络体系结构

定义和组成

计算机网络：一批**独立自治的计算机系统**的**互连**集合体。

组成：

- 资源子网：服务器，客户机
- 通信子网：信道，网络互联设备

通信子网基本结构：

- 点到点通道(star, ring, tree)，路由选择，用于广域网、城域网
- 广播通道(bus, ring)，通道分配，用于局域网

计算机网络体系结构!

定义计算机网络及其部件所完成的**功能**，是**层次和层间关系**的集合。

- 仅定义功能，**不定义协议实现和接口。**

位于不同计算机上进行对话的**第 N 层通信各方**可分别看成是一种进程，称为**对等进程。**

协议：**同等层次通信双方**信息交换的规则。

- 由语法，语义，定时关系组成

服务：**同一实体上下层**信息交换时必须遵守的准则

接口：定义了**下层向上层**提供的原语操作和服务。

服务访问点 SAP：层间服务在接口的 SAP 上进行，每个 SAP 有唯一识别地址，每个接口可以有多个 SAP

接口数据单元 IDU: 通过 SAP 传送的层间数据单元,

- IDU = 服务数据单元 SDU + 接口控制信息 ICI

协议数据单元 PDU: 第 N 层实体通过网络传送给它的对等实体的信息单元。

- PDU = SDU + 协议控制信息 PCI

服务的分类:

- 面向连接的服务: 建立连接, 传输数据, 断开连接。**不代表可靠。**
- 无连接服务: 每个包独立进行路由选择

四种服务原语：request，respond，indicate，confirm.

设计原则

分层原则：**模块化、抽象、功能复用。**但带来**信息隐藏**的缺点。

端到端原则：只有对性能有提升时才将功能实现在底层，并且实现在底层的功能应该对其他应用的性能影响小。

参考模型!

ISO/OSI

- 物理层: 在物理线路上传输比特数据
- 数据链路层: 在有差错的物理线路上提供**无差错**的数据传输 (帧 Frame)
- 网络层: 控制**通信子网**提供源到目的的数据传送 (分组 Packet)
- 传输层: 为**用户**提供端到端数据传送服务
- 会话层: 为用户提供安全认证
- 表示层: 数据转换和表示
- 应用层

TCP/IP: 将 OSI 的前两层合为 Host-to-Network, 后三层合为应用层.

其它网络体系结构

X.25分组交换网

面向连接，支持永久/交换虚电路。三层协议:

- 物理层: X.21, X.3/X.28/X.29
- 数据链路层: LAP, LAPB
- 网络层: PLP

DTE：数据终端设备 DCE：数据电路端设备 PAD：包的封装和解封

数据通信基本原理

基本理论

信道有截止频率，高于截止频率的振幅衰减较多。

- 通过信道的谐波次数越多，信号越逼真。而数据传输速率越快，基频就越高，最高次谐波就越低。**有限的带宽限制了数据的传输速率。**

比特率和波特率：关系取决于每个信号表示几个比特

- 波特率：baud 每秒钟信号变化次数，也称调制速率
- 比特率：bit 每秒钟传送的二进制位数

信道的最大数据传输速率!:

- 奈奎斯特定理, 无噪声有限带宽:

  $$2H\log_2 V \text{ bps}$$

  H 为带宽，V 为电平级数
- 香农定理, 考虑随机噪声：$H\log_2(S/N+1)$ bps, 信噪比 $10\log_{10}(S/N)$ db,

已知信噪比可以计算 S/N，进而计算最大传输速率。

通信技术

连接方式

- 点掉点、点到多点
- 单工、半双工、全双工：单工传输中，监视信号可回送。
- 同步和异步传输：
    - 同步传输：接收方必须知道每一位信号的开始和持续时间。
    - 异步传输（以字符传输为例）：信息发送以字符为单位，需要辅助位。

数据表示 模拟数据和数字数据。

信号发送

- 模拟数据（声音） -- 电话系统 --> 模拟信号
- 数字数据（二进制脉冲） -- 调制解调器 --> 模拟信号
- 模拟数据 -- 编码解码器--> 数字信号
- 数字数据 -- 数字编码解码器 --> 数字信号

数字信号发送的优点是：**价格便宜，对噪声不敏感**；

缺点是：**易受衰减，频率越高，衰减越厉害**

数据编码

- 数字数据的数字传输（基带传输）：
    - 不归零制码 NRZ
    - 曼彻斯特码：LH = 0，HL = 1
    - 差分曼彻斯特码：每位中间都跳变，开始时有跳变表示 0，无跳变表示 1.
    - NRZ0/NRZ1：每位开始时，逢0/1跳变，否则不跳变。
- 数字数据的模拟传输（频带传输）
    - ASK，FSK，PSK：调幅，调频，调相
- 模拟数据的数字传输：
    - PCM (Pulse-code modulation)：将模拟信号振幅分成 2^n 级
    - 差分 PCM：根据前后两个采样值的差编码
    - delta 调制：根据每个采样值与前一个值之间的差来决定输出二进制1或0.

多路复用技术

TDM, FDM, WDM: 时分复用, 频分复用, 波分复用

- T1 载波为 TDM, 分成 24 个信道，128 级 PCM, 1.544 Mbps

交换技术!

动态地接通或断开通信线路.

- 电路交换: 直接利用可切换的物理通信线路，连接通信双方
    - 建立物理通路时间长, 数据传输延迟短
    - 一般为 TDM, 时间被分为帧 (frame) ，帧被分为时槽 (slot)。非永久会话需要动态绑定时槽.
- 报文交换:信息以报文为单位进行**存储转发**
    - 线路利用率高, 延迟长, 要求缓冲大.
- 分组交换: 信息以分组为单位进行**存储转发。**统计复用, 按需分配信道资源.

    - 数据报分组交换: 每个分组均带有网络地址 (分组头)，可走不同的路径
        - 每个分组都要路由选择, 可扩展性更好
    - 虚电路分组交换: 来自同一流的分组通过一个预先建立的路径（虚电路）传输
        - 建立连接时做一次路由选择, 路由器需要维护虚电路的状态信息

对各类通信子网定义下列参数：

N = 两个给定站点之间所经过的段数；

L = 报文长度（L为分组大小P的整数倍），单位：位；

B = 所有线路上的数据传输速率，单位：位/秒；

P = 分组大小（P <= L），单位：位；

H = 每个分组的分组头，单位：位；

S1 = 线路交换的呼叫建立时间，单位：秒；

S2 = 虚电路的呼叫建立时间，单位：秒；

D = 各段内的传播延迟，单位：秒。

计算延迟：

1. 电路交换：$S_1 + ND + L/B$
2. 报文交换：$ND + NL/B$

   每级都必须收到完整报文后才传递给下一级
3. 数据报分组交换：$ND + \frac{P+H}{B}(\frac{L}{P} + N - 1)$

   计算时需要注意是否考虑头的overhead
4. 虚电路分组交换：$S_2 + ND + \frac{P}{B}(\frac{L}{P} + N - 1)$

   不考虑头的overhead

物理层

物理层提供**机械、电气、功能、规程**的特性。目的是**启动、维护和关闭数据链路实体间**进行比特传输的物理连接。

四个特性!

- 机械特性：物理连接的边界点，即插接装置。规格、引脚数量、排列
- 电气特性：传输时电压高低、阻抗匹配、传输速率和距离限制
- 功能特性：线路功能。数据/控制/定时/地
- 规程特性：线路的规程和时序关系

传输介质：

- 双绞线：模拟传输/数据传输。带宽依赖于线的类型和传输距离。
- 同轴电缆: 基带（数据传输），宽带（模拟传输）
- 光纤：多模（短距离），单模（长距离），都支持波分复用

SONET/SDH 光纤传输，采用TDM技术的同步系统。

数据链路层

- **基本功能**：数据传输，成帧，差错控制
- 对于误码率低的链路，链路层协议可以不实现可靠传输功能。

提供给网络层的服务：

- 无确认无连接：适用于错误率低的场合，或实时通信
- 有确认无连接：WiFi
- 有确认有连接

成帧

- 字符计数：在帧头中用一个域来表示整个帧的字符个数
- 字符填充的字符定界法：起始字符 DLE STX，结束字符 DLE ETX

  发送方在数据中的 DLE 字前再填充一个 DLE

  缺点：局限于8位字符传送
- 位填充的标记定界法：帧的开始和结束用 "01111110" 标记

  发送方若在数据中遇到连续5个1，就插入一个0
- 物理层编码违例法：如曼彻斯特编码中 HH/LL 不表示数据，可作定界符

纠错检错

纠错码：海明码

检错码：CRC 循环冗余码

```
若生成多项式 G(x) 有 r 位
在待编码数据后添 r-1 个 0
用生成多项式串除(模2除法)填零后的数据，得到余数
发送数据 = 待编码数据后添余数
```

模 2 加减等同于异或，只要最高位为 1 就商 1，保证余数是 r -1 位.

链路层数据传输协议!

1. 无约束单工协议
2. 单工停等协议：增加响应帧

   数据传输速度为 B，数据长度为 L，传输延迟为 D，则效率 $\frac{L/B}{L/B+D}$
   - 数据传输速度越快，距离越远，效率越低。
3. 有噪声信道的单工协议：增加 1 位序号，标记需要重传的帧。

滑动窗口协议!

单工 --> 全双工

设发送窗口大小为 n，网络吞吐率 $\min(nL/RTT, B)$, 要最大限度发挥网络带宽，需 $n \geq \frac{RTT}{L/B}$

- 发送方窗口：大小不固定，表示已发送但尚未确认的帧的序号表。上界是下一个要发送的序号，大小 = 上界 - 下界. 发送帧时，序号取上界，上界+1； **收到 ack 帧时，下界 = ack+1。**
- 接收方窗口：大小固定，表示允许接受的信号帧。 等于下界的帧被正确接收，上下界都加 1.

4. 一比特滑动窗口协议：若双方同时开始发送，会有一半重复帧。
5. 退后 n 帧重传：发送窗口 > 1，接受窗口 = 1。
   - 发送窗口 < 序号个数，设多个计数器
   - 超时时，将窗口上界置成窗口下界，重传原窗口内的所有帧
6. 选择重传：发送、接收窗口均 > 1

    - 发送窗口 + 接收窗口 <= 序号个数

- 发送方和接收方的缓冲区大小应等于各自窗口大小
- 接收方收到不等于接收窗口下界的包时发 NAK，如果**这个包在接受窗口内则会被缓存**；checksum 错误时也发 NAK；收到等于接收窗口下界的包时，回送 Ack，并启动 Ack timer，如果一段时间内没有新的帧传来，就重传 Ack.
- 发送方 Ack timeout 时，重传超时的这一帧；收到 ACK 时，增加发送窗口下界；收到 NAK 时，重传失效的帧而不改变发送窗口。

协议工程

协议说明：定义一个协议实体给用户的服务及其内部操作

协议验证：验证协议说明是否完整正确

协议实现：用硬软件实现协议说明中规定的功能

协议测试：用测试的方法来检查协议实现是否满足要求

对协议 3 的形式化验证：看不懂

- 有限状态机
- Petri网模型

常用数据链路层协议

HDLC (High-level data link control)

- 带**位**填充的标记定界, 用滑动窗口技术, CRC 检错
- 分为信息帧、监控帧、无序号帧

X.25的链路层协议LAPB

- 可看作 HDLC 的子集

PPP 协议

- **字符**填充，通常不使用滑动窗口，以帧为单位发送
- 包括链路控制协议 LCP 和 网络控制协议 NCP
- 与 HDLC 的主要区别的是面向字符

MAC 子层

数据链路层中管理局域网信道分配的子层。

信道分配

静态：FDM，TDM，WDM

适用于用户少，数目固定，通行量都很大的情况

无法灵活地适应站点数及其通信量的变化

ALOHA 协议

纯 ALOHA 协议：发送数据后监听是否冲突，若冲突则随机等待一段时间重发。利用率最高 18.4 %.

- 分析冲突危险区

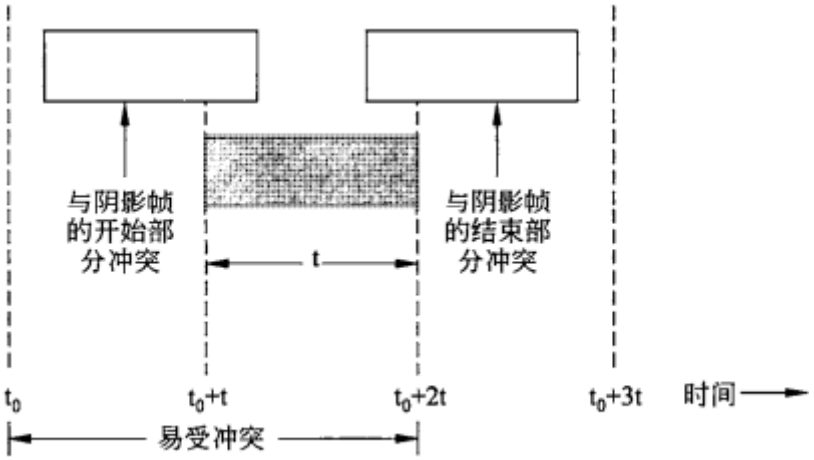


图 4-2　阴影帧的易受冲突周期

分槽 ALOHA： 把信道分成时间槽，槽长为帧发送时间，站点只能在时槽开始时发送帧。冲突危险区缩短一半，利用率最高 36.8 %.

载波监听多路访问协议!

1. 1-坚持型 CSMA:
    - 发送数据前监听信道，若信道空闲，则发送
    - 若信道忙，则监听至信道空闲发送
    - 若冲突，则等待随机时间重新发送
2. 非坚持型 CSMA：在 1 - 坚持型基础上，第二条改为等待随机时间重新发送。
3. p-坚持型 CSMA: 适用于分槽信道
    - 发送数据前监听信道，若站点发现信道空闲，则以概率 p 发送数据，以概率 1- p 等待下一时槽；
    - 若信道忙，则等待下一个时槽；
    - 若产生冲突，等待一随机时间，然后重新开始发送。
4. CSMA/CD： 边发送边监听，监听到冲突之后立即停止发送，并告知所有站点冲突发生。
    - 最坏需要两倍电缆传输时间确定冲突，必须满足： 发送有效帧时间 >= 冲突窗口
    - 半双工以太网使用

无冲突协议*

- 基本位图协议：N 个竞争槽 + 传输周期
    - 效率：轻负载下 $d/(N+d)$ 重负载下 $d/(1+d)$
- 二进制下数法 效率 $d/(d+\log_2 N)$

序号大的站得到的服务好。

有限竞争协议*

适应树搜索协议：站点组织成二叉树，每次一半站点参与竞争

无线局域网协议*

无线局域网与有线网的不同：

- 隐藏站点问题：不能发现潜在竞争者
- 暴露站点问题：非竞争者距离站点太近，导致非竞争者无法发送数据

MACA：发送站点刺激接收站点发送应答短帧， 从而使得接收站点周围的站点监听到该帧，并在 一定时间内避免发送数据

MACAW：改进 MACA，提高性能，增加确认帧，载波监听等

LAN 参考模型

物理层 --> MAC LLC --> 网络层

逻辑链路控制子层 LLC：提供**确认机制**和**流量控制**，基于 HDLC，为网络层提供统一接口

介质访问控制子层 MAC：数据帧封装（成帧，检错纠错），介质访问管理

IEEE 802.3 和以太网!

- 以太网标准：10Mbps 的 CSMA/CD
- 802.3: 1-坚持型CSMA/CD技术，曼彻斯特编码

物理层类型用以下域表示：

- <data rate in Mb/s> <medium type> <maximum segment length (*100m)>
- 后缀-T最大长度100m，-F最大长度2000m

两个收发器之间最多使用 4 个**中继器**，最长 2500 米

二进制指数后退：将冲突发生后的时间划分为长度为 51.2 微秒的时槽。第 i 次冲突后，在 0 至 2^{i-1} 间随机地选择一个等待的时槽数，再开始重传；16 次冲突后，发送失败。

为什么要有最大最小帧长：

以太网的最大帧长为 1500 字节，其存在的原因是：

- 如果帧长过长，会造成接收端的缓存溢出，发生错误；也将导致某一主机占据信道过久，不公平。

以太网的最小帧长需根据实际往返时间而定，一般**至少为 46 字节，不足时需要填0。**

- 如果帧长过短，会导致在 CSMA/CD 中无法检测出碰撞

$$\frac{\text{帧长}}{\text{数据传输速率}} \leq RTT = \frac{2 \times \text{物理距离}}{\text{物理传输速率}}$$

IEEE 802.5 令牌环

- 差分曼彻斯特编码
- 环由点到点链路组成，当某站点要发送数据时，抓住令牌环，数据被接收后释放令牌环
- 各个站点是公平的，获得信道的时间有上限。重负载下，效率接近100%.
- 环上存在一个监控站，负责环的维护，通过站的竞争产生。
- 为解决环断裂导致整个环无法工作的问题，使用线路中心进行布线，线路中心设有旁路中继器。

网桥!

网桥（bridge）是工作在数据链路层的一种网络互连设备，它在互连的 LAN 之间实现帧的存储和转发.

- 连接距离超过 2500 m (IEEE 802.3上限) 的 LAN
- 中继器不能隔离冲突域，但网桥/交换机可以隔离冲突域
- 连接不同类型的 LAN，以有助于安全保密

透明网桥

工作在混杂（promiscuous）方式，接收所有的帧。

采用**逆向学习**(backward learning)算法收集MAC地址。学习**源MAC地址和端口的对应关系。**注意每收到一个包就有一次学习的机会。

转发策略：

- 目的LAN与源LAN相同，则丢弃帧
- 目的LAN与源LAN不同，则转发帧
- 目的LAN未知，则洪泛帧（向除了源端口外的所有端口发送）

生成树网桥

多个网桥并行可能产生回路

生成树构造：

1. 每个桥广播自己的桥编号，号最小的桥称为生成树的根
2. **每个网桥计算自己到根的最短路径**，构造出生成树
3. 算法持续运行，应对网络拓扑的动态变化

源路由网桥

帧的发送者知道目的主机是否在自己的 LAN 内 。

如果不在，在发出的帧头内构造一个准确的路由序列， 包含**要经过的网桥、LAN**的编号。

- 对带宽进行最优的使用。但网桥的插入对于网络是不透明的，需要人工干预。
- 可看作是面向连接的

高速以太网*

FDDI 光纤分布式数据接口:

- **多模光纤**作为传输介质, MAC 协议与令牌环类似
- 通常作为连接LAN的主干网络，连接双环的 A 类站和单环的 B 类站
- 为提高信道利用率，站点**发完数据后立即产生新令牌，环上可能同时存在多个帧。**

快速 (百兆) 以太网：比特时间100ns -> 10ns。

千兆以太网：在一个冲突域内，只允许一个repeater。

万兆以太网：10GE只工作在全双工方式，不使用CSMA/CD协议，传输距离大大提高。使用单模或多模光纤。

网络层

网络层为一个网络连接的两个**传送实体**间**交换网络服务数据单元**提供功能和规程的方法，它使传送实体独立于**路由选择和交换**的方式。

- 通信子网的最高层，端到端传输的最底层。

内部结构(交换技术):

- 数据报子网
- 虚电路子网

路由算法

找出并使用汇集树。

洪泛算法

选择性洪泛算法：将接收的每个分组仅发送到与正确方向接近的线路上

基于流量的路由算法!

- 前提：每对结点间平均数据流相对稳定和可预测。

  需要预知网络拓扑，通信量矩阵，线路带宽矩阵，路由算法。

距离向量路由算法

DV，用于 RIP.

- 让坏消息快速传播：水平分裂，从邻居结点学到的到X的距离不向邻居结点报告
- 缺点：**不考虑链路带宽**，路由收敛慢，存在无穷计算问题，路由**报文开销大**（不是增量更新），**不适用于大规模网络。**

链路状态路由算法

LS，用于 OSPF.

- 发现邻居结点，两个或多个路由器连在一个 LAN 时，引入人工节点：

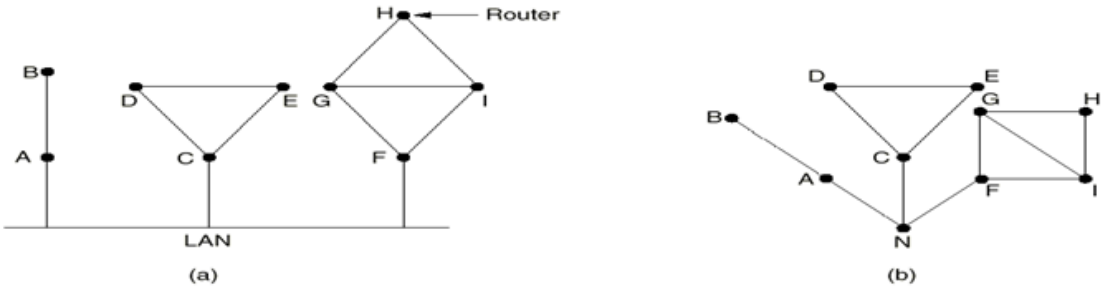


Fig. 5-13. (a) Nine routers and a LAN. (b) A graph model of (a).

- 测量到每个邻居结点的延迟或开销
- 将所有学习到的内容封装成一个分组：分组以发送方标识开头，后面是序号、年龄和一个邻居节点列表。
  - 分组定时或发生重大事件时创建。
- 将分组发给所有路由器
  - 用 32 位序号标记分组，防止循环使用序号造成混淆
  - 增加 age 域，防止序号出错

DV 和 LS 的对比

路由信息：

- DV：将本路由结点对全网拓扑的认识告诉给邻居
- LS：将本节点对邻居的认识洪泛给全网，增量更新

收敛速度：

- DV：不确定
- LS：$O(n \log n)$ 存在路由震荡问题

健壮性：结点会广播错误链路开销

- DV：每个节点只计算自己的路由表
- LS：每个节点的路由表被别的节点使用，错误会传播到全网

分层路由

- 解决规模增长带来的空间、时间压力
- 分层后计算的路由不保证最优性

移动主机路由

- 移动用户（mobile users）：位置发生变化，包括通过固定方式或移动方式与网络连接的两类用户。
- 家乡位置（home location）：所有用户都有一个永久的家乡位置，用一个地址来标识。
- 家乡代理（home agent）：每个区域有一个家乡代理，负责**记录家乡在该区域，但是目前正在访问其它区域的用户。**
- 外部代理（foreign agent）：每个区域（一个LAN或一个wireless cell）有一个或多个外部代理，它们**记录正在访问该区域的移动用户。**

移动用户进入新区域时，必须先向外部代理注册。

```
给用户的包 --> 家乡局域网 -- 家乡代理隧道传输 -> 外部代理 --> 用户

家乡代理告知发送方后续分组直接发给外部代理
```

拥塞控制！

网络上有太多的分组时，性能会下降，这种情况称为拥塞。

拥塞控制是全局性的，流控制是局部问题。

闭环控制：基于反馈机制

开环控制：通过设计解决拥塞，不考虑网络当前状态

流量整形：强迫分组以可预测的速率发送

- 漏桶算法：
  - 将用户发出的不平滑的数据分组流转变成网络中平滑的数据分组流
  - 漏斗已满在进水，水会溢出到漏斗外，**不灵活**
- 令牌桶算法
  - 漏桶存放令牌，每 T 秒产生一个令牌，令牌累积到超过漏桶上界时就不再增加。分组传输之前必须获得一个令牌，传输之后删除该令牌
  - 允许空闲主机积累发送权，
  - 令牌桶高速维持时间 = 桶大小 / (流量速率-积累速率)

都可用于定长/变长分组协议。

1. 流说明：描述发送数据流的模式和希望得到的服务质量的数据结构
2. 准入控制：根据流说明和网络资源分配情况，进行准入控制
3. 子网根据协议在虚电路上为连接预留资源

抑制分组：

- 路由器监控输出线路及其它资源的利用情况，超过某个阈值，则此资源进入警戒状态。在警戒状态下，向源主机发送抑制分组，分组中指出发生拥塞的目的地址。同时将原分组打上标记

逐条抑制分组

公平队列算法、加权公平队列

上述算法都不能消除拥塞时，只能将负载丢弃。

网络互联

互联设备

- 中继器：物理层，在电缆段间拷贝比特，放大弱信号，延长传输距离
- 网桥：链路层，在不同的 LAN 间存储转发帧
- 多协议路由器：网络层，在网络之间存储转发分组，必要时做协议转换
- 传输网关：在传输层转发字节流
- 应用网关：在应用层实现互联

隧道技术、防火墙：网安已学了

分片和重组

各种网络都限制了分组的最大长度。

重组策略：

- 重组过程对其他网络透明：

  大分组在入口网关被分片后，在同一出口网关重组

  - 出口网关需要知道何时所有片段都到齐

    所有片段必须从同一出口网关离开

    大分组经过一系列小分组网络时，需要反复分片重组，开销大
- 重组过程对其他网络不透明：

  由目的主机重组分片。

  - 对主机要求高，每个片段都需要分组头，网络开销增大。

标记分片：

- 树形标记法
- 偏移量法

网络层协议

IP!

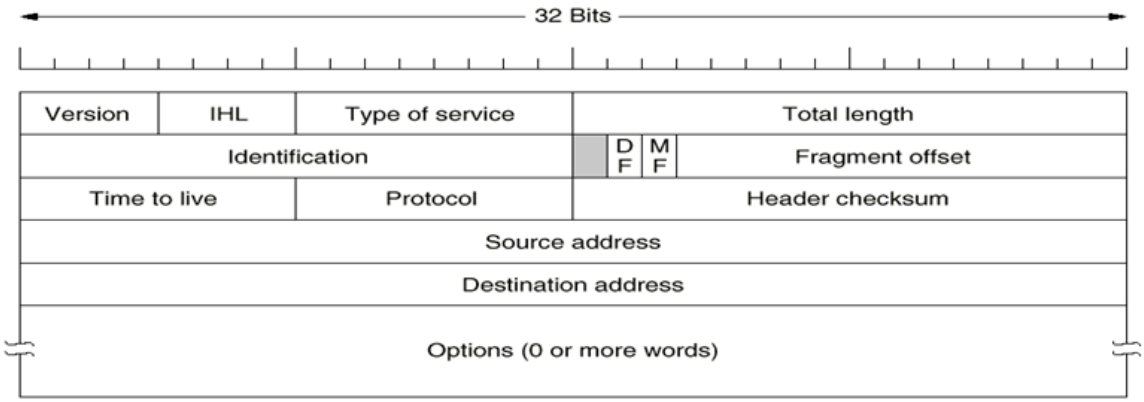


Fig. 5-45. The IP (Internet Protocol) header.

- 定长段 20 byte，变长段 0-40 byte
- 头长度以 32 bit 为单位
- DF 要求所有机器必须能够接收小于等于 576 字节的片段
- 分组后除最后一片段都要置 MF

- 除最后一个片段外的所有片段的长度必须是8字节的倍数, (Fragment offset 以 8 byte 为单位)
- 头校验和只对头做计算: 每 16 位求反，循环相加，截断为 16 位再取反。

IPv4 地址

- 全 0 表示 host，全 1 表示 broadcast，127.x.x.x 用于 loopback
- 一个 IP 地址并不真正指向一台主机，而是指向一个网络接口
- ABCDE 的标志是 0,10,110,1110,11110，ABC 的主机号都按字节对齐，D 没有 网络号 + 主机号，表示多播，E 是保留

CIDR

基于分类的IP地址空间的组织浪费了大量的地址，罪魁祸首是B类地址。

CIDR 将剩余的 C 类地址分成大小可变的地址空间：

- 路由表中增加一个32位的掩码（mask）域
- 最长前缀匹配原则：路由查找时，若多个路由表项匹配成功，选择掩码最长（1比特数多）的路由表项

ICMP!

Internet Control Message Protocol，用于报告错误和测试

- 封装在 IP 头内

ARP!

Address Resolution Protocol，解决网络层地址 (IP) 到数据链路层地址 (MAC) 的映射问题。

- 主机启动时，广播其 (IP, MAC)
- 主机维护一个 ARP 表，根据 IP 查找 MAC：

```
if 目的主机在同一子网内:
    ARP.find(目的IP)
else:
    ARP.find(缺省网关IP)
if not find:
    广播分组，等待应答并更新分组
```

- ARP攻击存在于**局域网**

RARP

由 MAC 到 IP 的映射，主要用于无盘工作站启动，

- 缺点：由于路由器不转发广播帧 ，RARP 服务器必须与无盘工作站在同一子网内
- 替代协议 BOOTP

网关协议

IGP：内部网关协议，包括 RIP，OSPF

EGP：外部网关协议，包括 BGP

RIP

距离向量算法，基于 **UDP**. 实验已写了。

OSPF

链路状态算法, 不经过 TCP/UDP

- 支持**多种距离尺度，基于服务类型的路由**等
- 分层路由：

  自治系统可以划分成区域（areas），每个AS有一个主干（backbone）区域，称为区域0.

  - 所有其他区域都和主干区域相连，其它区域间不能相连。

  四类路由器：

  1. 区域内部路由器
  2. 区域边界路由器
  3. 主干区域内部路由器
  4. 自治系统边界路由器 ASBR

BGP

边界网关协议，通过 **TCP** 连接传送路由信息，采用**路径向量算法**：

- 与距离向量协议相似，每个BGP网关向邻居广播**所有通往目的地的路径**

IPv6

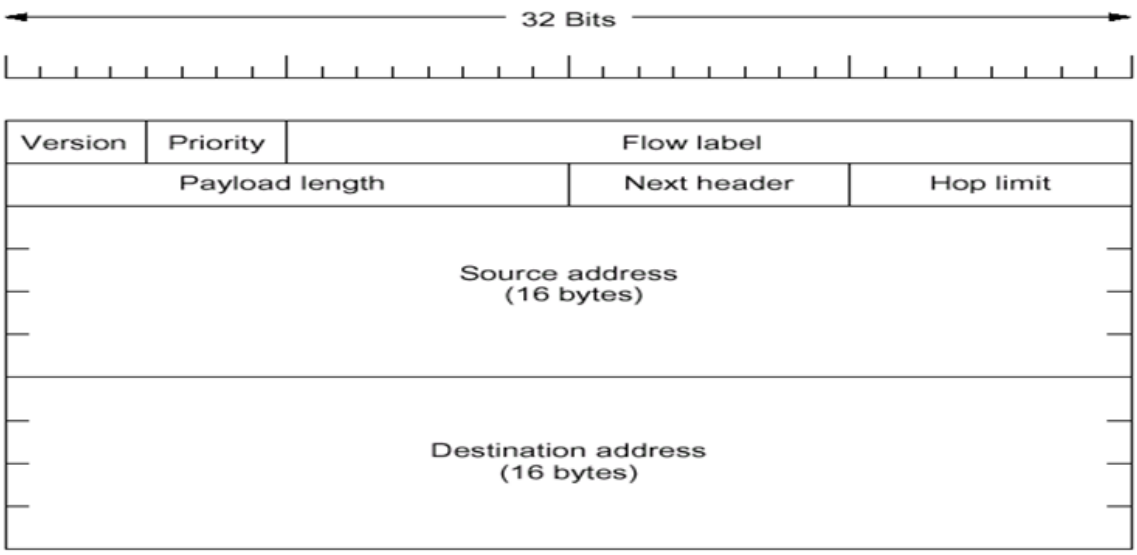


Fig. 5-56. The IPv6 fixed header (required).

与 IPv4 的主要变化：

- 地址由 32 位变为 128 位
- 头简化，定长 40 字节
  - 由于 IPv6 分组头定长，取消 IHL 域 Protocol 域取消，用 Next header 域表示
  - 取消与分片有关的域，IPv6的分片方法：所有主机和路由器必须支持1280字节的分组，路由器不做分片，由主机做分片
  - 取消Checksum域
  - Flow label，用来允许源和目的建立一条具有特殊属性和需求的伪连接
  - Payload length 指示IP分组中40字节分组**头后面部分**的长度

IPv6 地址：

- 每个数字开头的 0 可省略，**连续的 16 bit “0” 可以被一对冒号替代，但是一对冒号只能出现一次**
- 打头的“0”可以省略，0123可以写成123

IPv4 向 IPv6 过渡：

- 双栈：实现 IPv4/v6 两套协议栈，主机根据 DNS 返回的结果或对方发来报文的版本号决定采用哪个协议，路由器根据收到 IP 分组的版本号决定采用哪个协议。
- 隧道：IPv6的报文作为IPv4报文的净负荷在IPv4网络中传输
- 翻译

传输层

提供**源主机应用程序到目的主机应用程序**间的**有效、可靠或尽力而为**的连接服务。

简单连接管理

拆除连接方式：不对称/对称

Berkeley Sockets：对称释放的一个例子，全双工

传输层编址

TSAP = (IP, Local Port) = socket

- TSAP 预先约定，或从名称服务器/目录服务器获取。
- 进程服务器：监听多个端口，必要时唤起对应服务进程。

TCP

面向连接的、可靠的、基于字节流 (消息边界传输中不能得到保留)

- 点到点，不支持组播和广播
- 每条连接用 (src socket, dst socket) 标识, 256 以下的 port 号被标准服务保留。
- TPDU 被称作段 (segment)，段长受 IP 包长度 (65535) 限制，也受链路层 MTU (ethernet 1500) 限制。

TCP 头

内容	长度	注
源端口和目的端口	各 16 bit	
序号和确认号	各 32 位	**以字节为单位编号**
TCP 头长	4 位	单位 32 bit
unused	6 位	
标志位	6 位	
窗口大小	16 位	
checksum	16 位	对**TCP头，数据和伪头**计算

连接管理!

建立连接，三次握手：

1. A 发出序号为 X 的 TPDU
2. B 发出序号为 Y 的 TPDU，确认序号 X
3. A 发出序号 X + 1 的 TPDU，确认序号 Y

- 如果仅仅只有两次握手，若客户端已经连接超时失败，放弃连接；而超时的信号又到了服务端，服务器认为建立了连接，资源就被浪费了。
- 若两个主机同时试图建立彼此间的连接，则**只能建立一条连接。**

释放连接，三次握手 + 定时器

- 两军问题，不存在安全的通过 N 次握手实现对称式连接释放的方法
- 释放连接时，客户端发出 FIN 位置 1 的 TCP 并启动定时器，收到确认后关闭连接；无确认且超时，也关闭连接。
- 服务器收到关闭连接的包，就发送 ACK, 并释放连接
- 唯一的失败情况：客户端的请求没有到达服务器，超时后客户端关闭连接，但服务器没有。

窗口管理

基于确认和可变窗口大小

- 窗口大小为0时，正常情 况下，发送方不能再发 TCP段，但有两个例外 -
    - 紧急数据可以发送
    - 为防止死锁，发送方可以发送1字节的TCP段，以便让接收方重新声明确认号和窗口大小

改善性能

1. 发送方缓存应用程序的数据，等到形成一个比较大的段再发出
2. 在没有可能进行"捎带"的情况下，接收方延迟发送确认段
3. Nagle 算法：当应用程序向传输实体发出一个字节时，传输实体只发出第一个字节并缓存所有其后的字节直至收到对第一个字节的确认，然后将已缓存的所有字节组段发出并对再收到的字节缓存，直至收到下一个确认.
4. Clark 算法解决傻窗口症状

   傻窗口症状：当应用程序一次从传输层实体读出一个字节时，传输层实体会产生一个一字节的窗口更新段，使得发送方只能发送一个字节.

   解决办法：限制收方只有在具备一半的空缓存或最大段长的空缓存时，才产生一个窗口更新段

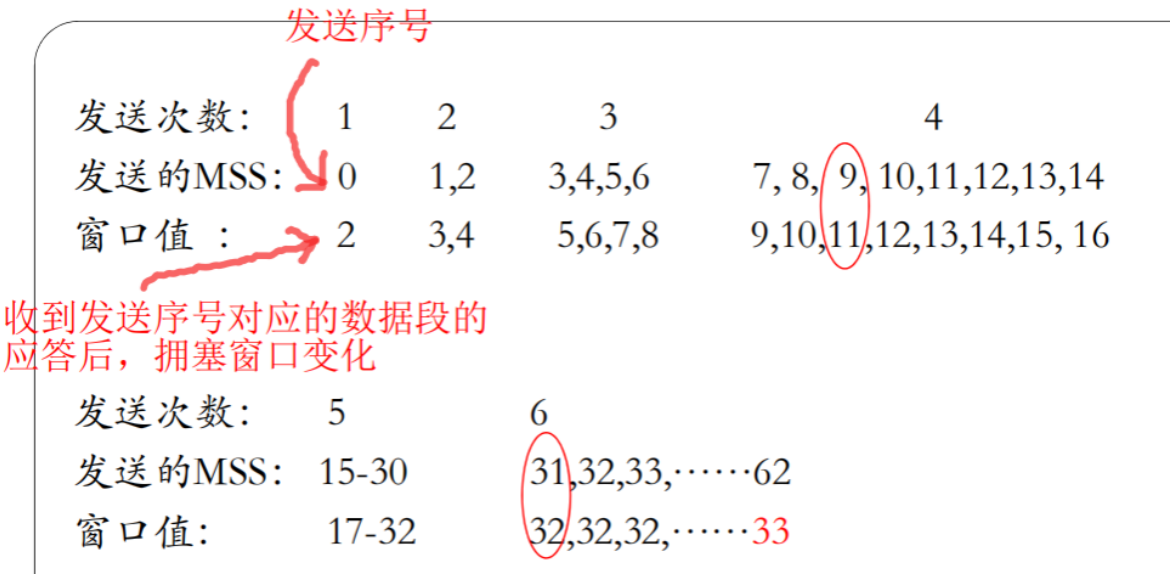
拥塞控制!

处理拥塞：按可变滑动窗口和拥塞窗口最小值发送。

- 慢启动算法: 快速探测网络承载力
- 拥塞窗口大于阈值时，使用拥塞避免算法
- 快速重传算法：不必等到计时器超时才判定丢包，连续收到3个重复确认后，就判定 timeout。

```
congwin = MSS
cnt = 0
for every ack:
    if congwin < thredhold:
        congwin += 1
    else:
        cnt += 1/congwin
        if cnt == 1:
            congwin += 1

for every loss:
    threshold = congwin // 2
    congwin = MSS
```



拥塞控制分析：

- MIAD：公平性不收敛也不稳定。有效性不收敛，在 x1=x2=bIaD/(1-bI)时稳定。
- AIAD：公平性不收敛但稳定。有效性不收敛，在 aI+aD=0 时稳定。
- MIMD：公平性不收敛但稳定。有效性收敛。条件是平凡的。
- AIMD：公平性收敛。有效性收敛。

UDP

无连接、端到端。

- 尽力而为的服务，报文可能会丢失、乱序

应用：**RIP，DNS，SNMP**，流媒体，可靠传输可在应用层实现

应用层

用户代理（user agent）指**用户和网络应用程序**间的接口。API 定义用户程序和传输层之间的接口。

客户/服务器模型

网络应用的基础。

- 服务器要支持并发，一般分成两部分，一部分用于接受请求并创建新的进程/线程，另一部分用于处理实际通信过程。

DNS

基于 UDP. 大小写无关. 最长 255 字符，各部分最长 63 字符。

- 将域名空间划分为许多无重叠区域，每个区域覆盖了域名空间的一部分，区域的边界划分是人工设置的.
- 存储于分布式数据库中

DNS 解析域名，得到**资源记录五元组 (IP)**。

域名查询：

- Recursive query：问一个人就是把任务完全交给它
- Iterated query："我不知道，但可以问它"

解释以下 URL 各部分的意义 `http://info.tsinghua.edu.cn:80/index.jsp`

> `http`：协议
>
> `info.tsinghua.edu.cn`：主机的 DNS 域名
>
> `80`：主机的 HTTP 端口号
>
> `index.jsp`：路径名

能访问域名，不能访问 IP：可能这个域名对应多个 IP，其中这个 IP 对应的主机失效了；或对应服务器禁止了通过 IP 访问。

能访问 IP，不能访问域名：DNS 失效，无法解析域名。

SNMP

简单网络管理协议，基于 UDP.

管理协议用于管理工作站(**客户端**)查询和修改被管理节点(**服务器**)的状态，被管理节点可以使用管理协议向管理站点产生"陷阱（trap）"报告.

抽象语法表示法 ASN.1:

- 基本数据类型，构造数据类型
- 对象命名树，是基本数据类型的一种
- 传输语法：标识符，数据域长度，数据域
  - 标识符：[tag(2) type(1) number(5)]

    number <= 30 时，用低五位表示

    > 30 时，低五位为 11111，用后面字节表示。最后一个字节最高位为 1，其他字节最高位为 0
  - 数据域长度：

    < 128 时，用一个字节表示，且其最高位为 0

    >= 128 时，第一个字节最高位为 1，第七位表示后面表示长度的字节数。
  - 数据域：

    INTEGER/OCTED STRING：二进制编码

    BIT STRING: 传位串前，先传一个字节表示位串最后一个字节不用的位数。

    NULL：长度域为 0，不传数据

    对象树标识符：前两个数a，b可用一个字节编码为 40a + b

SNMP 扩展了 ASN.1 的数据结构，成为管理信息结构 SMI. 相关的**对象**被集合成组。

SNMP 的**管理信息库 MIB** 包含十个组。**网络管理工作站(客户端)**通过使用 SNMP 协议，向**被管理节点(服务器)**中的 SNMP 代理发出请求，查询这些对象的值。

电子邮件

SMTP 基于 TCP，是一组用于从源地址到目的地址传送邮件的规则，并且控制信件的中转方式。

组成：

- 信封：接收方的信息，如名字、地址、邮件的优先级和安全级别
- 信件内容：由信头和信体组成，信头包含了用户代理所需的控制信息，信体是真正的内容

```
发送方用户代理 -- SMTP -> 发送方邮件server -- SMTP -> 接收方邮件server -- POP3/IMAP -> 接收方用户代理
```

SMTP 只能将邮件发给服务器，客户接受服务器数据需要 POP3/IMAP.

WWW

Web 对象，即网页，用 URL 标识。URL 组成：协议、域名(IP)、路径

Http: TCP, 80 port

- 非持久化连接：服务器每次响应报文后都关闭连接，每个 object 都要两个 RTT
- 持久化连接：较少的 RTT (Round-trip delay)

Web 缓存：不访问源服务器而满足用户的请求，浏览器实现。

主机访问 www.sina.com 所需用到的协议：

应用层协议为 DNS 和 HTTP，即除了 HTTP 之外还需要 DNS 协议； 传输层协议为 UDP 和 TCP。 DNS 用于将服务器主机名解析成 IP 地址；

FTP

文件传输协议，建立两个双向并行 TCP 连接，控制和数据。

FTP 服务器维护状态：当前目录、身份认证。