绪论

计算机网络的功能和组成

计算机网络是相互连接的、自制的计算机组合。从物理组成来看，计算机网络包括阿硬件、软件、协议。

计算机网络的组成

核心部分：路由器，网络；边缘部分：主机（端系统）。
边缘部分有两种通信方式：客户-服务器方式（C/S），对等方式（P2P）。
核心部分要向边缘部分的主机提供连通器，由路由器实现分组交换和路由选择，任务是转发收到的分组。

通信服务可分为：面向连接服务、无连接服务。
面向连接服务必须建立连接、连接维护、释放连接（TCP），协议复杂，通信效率不高。
面向无连接服务的每个分组都携带完整的目的节点地址，传输过程不经过建立连接、链接维护、释放连接，传输时可能出现乱序、重复、丢失。可靠性不好，但协议简单、通信效率高。

可靠服务是具有纠错、检错、应答机制，能保证数据正确、可靠地传送到目的地。**不可靠服务**只尽量争取、可靠地传送，不能保证数据正确，一种尽力而为的服务。

有应答服务指接收方收到数据后向发送方给出相应答应（文件传输服务）。**无应答服务**则不自动给出应答。

只有TCP和PPP是**面向连接服务**，只有TCP是**可靠**的，只有TCP和CSMA/CA是**有应答的。**

计算机网络的特点

连通性：让两个用户之间实现连接，但不需要知道物理位置。
共享性：资源共享（信息、软件、硬件）。

计算机网络的功能

数据通信：最近本、最重要的功能，包括连接控制、传输控制、差错控制、流量控制、路由选择、多路复用等。
资源共享：数据资源、软件资源、硬件资源。
分布式处理：部分任务传送给网络中的其它计算机系统进行处理。
信息综合处理：分散各地计算机的数据进行集中处理或分级处理。

计算机网络分类

按交换技术分类：电路交换、报文交换、分组交换。
按传输介质分类：有线网络、无线网络。
按拓扑结构分类：星型网络、总线型网络、环形网络、网状型结构。 **按分布范围分类**：广域网、局域网、城域网、个人区域网。

计算机网络的性能指标

- 速率：数据率dataRate或比特率bitRate，单位有b/s, kb/s, Mb/s, Gb/s等
- 带宽：频带宽度单位为赫（千赫、兆赫等）

- 延迟时延：一段到另一端所需要的时间间隔。总时延=发送时延+传播时延+处理时延+排队时延(通常后二者时间很短，不考虑)
  - 发送时延：数据帧从节点进入到传输媒体所需时间，发送时延=数据帧长度/发射速度
  - 传播时延：电磁波在信道中需要传播一定距离而花费的时间，传播时延=信道长度/信号在信道上的传播速率
  - 处理时延：交换节点为存储转发而进行操作的耗时
  - 排队时延：节点缓存队列中分组排队所经历的耗时
- 时延带宽积：时延带宽积=传播时延×带宽，以比特b为单位计算两点距离
- 吞吐量：单位时间内通过某个网络或接口、信道的数据量

计算机体系结构【重点】

为了减少计算机网的复杂性，采用分层设计方法（按照信息流动过程将网络的整体功能分解为一个个功能层，不同计算机的同等功能层之间采用相同协议，相邻功能层之间通过接口进行信息交换。

- **实体**：任何可以发送或接收信息的硬件或软件进程
- **对等实体**：同一层的实体叫做对等实体
- **对等层**：不同计算机上同一层
- **相邻层**：相同计算机上的相邻层次
- n层实体实现的服务为n+1层所利用。在此情况下，n层被称为服务提供者，n+1层为服务用户

计算机网络的体系结构是计算机网络的各层机器协议的集合。体系结构就是该计算机网络机器部件锁应完成的功能的精确定义。
实现是遵循该体系结构的前提下用何种硬件或软件完成这些功能的为题。体系结构是抽象的，实现则是具体的，是真正在运行的计算机硬件和软件。

体系结构的三要素

- **协议**：计算机网络中数据交换必须遵守的事先约定好的规则。
  - 协议三要素：语法（数据与控制信息的结构或格式），语义（需要发出的控制信息），同步（事件实现顺序）
- **服务**：下层为紧邻的上层提供功能调用，为下一层提供服务
  - 分为面向连接服务与无连接服务，可靠服务与不可靠服务，有应答和无应答服务
- **接口**：同一节点内相邻两层交换信息的连接点

协议是“水平的”（控制对等实体之间通信等规则），服务是“垂直的”（下层向上层通过层间接口提供服务）。

网络模型

OSI/RM为七层，TCP/IP为四层。教学通常用五层模型（自上而下：应用层、运输侧、网络层、数据链路层、物理层）

ISO/OSU参考模型

由国际标准化组织ISO提出的网络体系结构模型，又称为系统互联参考模型OSI/RM，通常称为OSI参考模型。OSI/RM分为七层，自上而下为：应用层、表示层、会话层、传输层、网络层、数据链路层、物理层。（应表会传网数物。应表会：高层协议，传网数物：低层协议）

- **应用层**——数据：为用户程序提供网络服务（HTPP，FTP，TELNET，SMTP等）

- **表示层**：将不同数据格式转换成通用格式（ASCII，JPEG，MPEG等，加解密，解压缩）
- **会话层**session：会话的建立、管理、终止通信主机会话，为表示层提供服务（访问次序的安排）
- **传输层**——段：两台主机间建立端到端连接，实现可靠传输（TCP、UDP分别为面向连接和面向无连接协议）
- **网络层**——数据包：主机之间的连接、路径选择以及基于IP的寻址（三层交换机、路由器工作在此层）
- **数据链路层**——帧：提供数据在物理链路的传输、物理寻址、网络拓扑、错误检测（立案层交换机、网桥工作于此层设备）
- **物理层**——比特流：高低电平，数据传输速度，传输距离，物理连接器等（HUB，中继器等线路属于此层）

TCP/IP模型

从上至下：应用层、运输层、网际层、网络接口层

- **应用层**：提供系统与用户的接口
  - 文件传输、域名解析、电子邮件服务等
  - FTP、SMTP、POP3、HTTP
- **传输层**：负责主机中两个进程的通信。传输单元为报文段（TCP）或用户数据包（UDP）
  - 为端到端连接提供不可靠的传输服务 和 流量控制差错功能
  - TCP、UDP
- **网际层**（网络层、IP层）：报文封装成分组，选择适当路由令分组能够交付到目的主机
  - 为传输层提供服务，路由选择，分组转发
  - ICMP、ARP、RARP、IP、IGMP
- **网络接口层**：协作IP数据在已有网络介质上传输的协议。更像ARP协议

补充

- 广域网和局域网的差异主要在于所提供的服务不同
- RTT：往返传播时延
- 物理层不参与数据封装工作
- OSI数据链路层可以提供网络的接口，提供物理寻址、差错校验、流量控制
- 传输层提供拥塞控制功能
- TCP/IP的网络层提供无连接不可靠的数据报服务
- OSI中，流量控制的是234567层，提供建立、维护、拆除端到端连接的是传输层，网络中路由的功能是网络层，主机进程之间的数据传送是传输层，为网络层实体提供数据发送和接受是数据链路层

物理层

物理层通信

至少要让两台计算机存在信息通路

- 传输媒体：数据传输系统中在发送器和接收器之间的物理通路，分为*引导型传输媒体*和*非引导型传输媒体*
  - 光信号：光纤、红外线；电信号：双绞线、同轴电缆；电磁波：微波
  - 单向通信（单工）：只能有一个方向的通信而没有反方向的交互
  - 双向交替通信（半双工）：双方都能发送，但不能同时发送/接收
  - 双向同时通信（全双工）：双方可同时发送/接收
- 功能：传输媒体上传输数据比特流，尽可能拼比掉不同传输媒体和通信手段的差异，透明的传输比特流
  - 机械特性（接口外观标准）、电气特性（电缆的电压范围）、功能特性（不同电压的意义）、过程特性（不同功能的各种可能事件的顺序）

编码和调制

- 通信基础：需要有发送端和接收端，并由传输系统连接
  - `-输入信息-> 源点 -输入数据-> 发送器 -发送的信号-> 传输系统 -接收的信号-> 接收器 -输出数据-> 终点 -输出信息->`
  - 数据data：运送消息的实体
  - 信号signal：数据的电气或电磁表现
  - 模拟信号analogousSignal：消息的取值是连续的
  - 数字信号digitalSignal：消息的取值是离散的
- 调制和编码：
  - 系带信号：来自源的信号，往往包含较多低频甚至直流成分，因此需要对基带信号进行调制modulation。分为两类
    - 系带调制：`其它信号->基带信号（数字信号）`，**编码**
    - 带通调制：`其它信号->带通信号`，**调制**
  - 常见编码方式：
    - 不归零制：正电平1，负电平0
    - 归零制：正脉冲1，负脉冲0
    - 曼彻斯特编码：位于周期中心向上跳变代表0，向下代表1
    - 差分曼彻斯特编码：每一位的中心处始终有跳变。开始边界有跳变代表0，没有跳变代表1
  - 基本的二元调制方法：
    - 调幅AM（载波*振幅*随系带数字信号变化），调频FM（载波*频率*随系带数字信号变化），调相PM（最波初始相位随基带数字信变化）
  - 正交振幅调制QAM：多元制的振幅相位混合调制方法，具有4个bit共16个情形

物理层的带宽计算【重点】

- **码元和比特率**：码元code是在使用时间域的波形表示数字信号；比特率是单位时间内数字通信系统传输的比特数，bit/s。若一个码元携带nbit信息量，则MBaud码元传输速率对应的信息传输速率为`M*n bit/s`
- **信道的极限容量**：传输信号时，码元传输速率越高/距离越远/媒体质量越差，输出端的波形失真越严重

    - 奈奎斯定理：理想条件下避免码间串扰的码元传输速率上限：`极限码元传输速率B=2W（带宽的2倍），极限数据传输速率C=2Wlog2(V) b/s`
    - 信噪比：信号和噪声的平均功率之比，即为S/N，用分贝dB做单位，`信噪比 = 10log10(S/N) dB`
    - 香农定理：带宽受限且有高斯白噪声干扰信道的极限且无差错的信息传输速率（香农公式）：`C = Wlog2(1+S/N) bit/s`
    - **常用的五个公式**：
        - 带宽：C = B×n = B×log2(V)
        - 极限码元传输速率：B = 2W
        - 奈奎斯特：C = 2Wlog2(V)
        - 香农：C = Wlog2(1+S/N)
        - 信噪比：dB = 10lg(S/N)
    - **注意！**香农定理不可能大于奈奎斯特定理，所以实际计算中取最小者

物理层设备

- **中继器**：
    - 碰撞域/冲突域：是指网络中一个站点发出的帧与其他站点的帧产生碰撞或冲突的那部分网络。任何信道都可能出现失真问题
    - 而中继器则是将已经衰减到不完整的信号经过整理后*重新产生*出完整的信号。中继器工作在物理层，每个结构仅仅简单地转发比特。信号/数据不经过任何筛选或过滤。不可以连接不同速率、不同规格的网段。不具有交换功能。
    - 使用中继器的局域网属于星形网，逻辑上是总网线，各站点会竞争对传输媒体的控制，同一时刻至多只允许一个站点发送数据
    - 中继器以太网是一个独立碰撞域
    - 放大器和中继器都起放大信号的作用。放大器会放大模拟信号，中继器再生的是数字信号
- **集线器**：
    - 实际是多端口中继器，由中继器发展而来

物理层下面的传输媒体

传输媒体/传输介质/传输媒介，就是数据传输系统在发送器和接收器间的物理通路。

- **导引型介质**：
    - **双绞线**：最常用。模拟和数字传输都可以用双绞线，通讯距离一般为几到几十公里
    - **同轴电缆**：具有较好抗干扰性，广泛应用于传输速率高的数据，带宽取决于电缆质量
    - **光纤**：传输带宽远大于其它传送媒体
        - 单模光纤：直径小到只有一个光波，此时光纤如同一个波导一直向前传播而不发生反射
        - 多模光纤：可存在多条不同角度入射的光在同一光线传输。折射容易导致衰减，所以传输距离不如单模
- **非导引型介质**：
    - 无线传输使用频段广。短波通信（高频通信）靠电离层反射，质量差速率低；微波通信为直线传播，有地面微波接力通信和卫星通信
- **宽带接入技术**：有线宽带接入、无线宽带接入
    - DSL的几种类型：ADSL，HDSL，SDSL，VDSL，DSL，RADSL
- **光纤同轴混合网HFC**：同时接入互联网和有线电视网等
- **FTTx技术**：实现宽带居民接入网的方案，代表多种宽带光纤接入方式

补充

- 利用模拟通信信道传输数字信号的方法称为频带传输
- 同步信息：时间信息。曼彻斯特和差分曼彻斯特包含同步信息
- 双相位编码中，代表每个比特的电信号中间都有跳变
- 曼彻斯特、差分曼彻斯特属于双相位编码
- 正交振幅调制的调制器中，实际传输比特率 = 每秒钟发送的波特数 × 每个波特携带的数据量 (log2(4×4)
- 不同的
    - PSK：改变载波信号的相位值来表示数字 1 0
    - ASK：改变载波信号的振幅
    - FSK：改变载波信号的频率
    - ATM

数据链路层

两台主机所经过的网络可以是多种不同类型的，中间伴随着封装、解封装的一系列形态。不同的链路层可能采用不同的数据链路层协议，传送的是帧。
包含有如下信道类型：

1. 点对点信道：采用一对一的点对点通信
2. 广播信道：采用一对多的广播通信，使用专用的共享信道协议来协调这些主机的数据发送

数据链路层的功能：封装成帧，透明传输，差错控制，流量控制

- **封装成帧**：在一段数据的前后添加首部SOH和尾部EOH构成一个帧，进行了帧定界。具体定界方法：
  - 字符计数法：在帧头部使用一个计数字段表明帧内字数，数值包含首部和尾部。（对帧的正确性影响最大）
  - 字符填充的首尾定界符：使用特殊字符来界定帧的开始DLE_STX与结束DLE_ETX
  - 比特填充的首尾标志：采用特定比特模式01111110来标志一帧的开始和结束
  - 违规编码法：例如曼彻斯特编码的比特1对应*高低*电平，比特零对应*低高*电平，而*高高*和*低低*电平在数据比特中是违规的
- **透明传输**：
  - 若在要传输的二进制代码恰好出现SOH和EOT一样的，会找到错误的帧边界。可采用字节填充或字符填充（转义字符）
  - 传输过程中可能会产生比特差错
- **差错控制**：传输错误的比特占总数的比例称为误码率BER。于是传送的帧广泛使用循环冗余检验CRC的检验技术（一般只检验，不纠错）
  - 在发送端把数据分组，设每组k个比特，每组M后再添加n位冗余码用于差错检测，并一起发出去
  - CRC会对每一个接收到的帧都会校验，但无法检测到帧丢失、重复
  - eg. 与串位101101对应的多项式为`x^5 + x^3 + x^2 + x^0`

额流量控制★★★

流量控制是让接收方来控制发送方发送数据的速度，便于接收方能够及时接收和处理数据，以免造成数据溢出和丢失。
发收双方AB分别维持一个发送窗口和接收窗口，**发送窗口**在没有收到确认的情况下可以连续把窗口内的数据全部发送出去，**接收窗口**只允许接收落入窗口内的数据。
停止-等待机制：每发送一帧都要等待接收方发来应答信号后才能发送下一帧。
滑动窗口机制：接收窗口向前时并且接收方发送了确认帧，控制发送窗口向前移动。

按窗口大小分类

按照窗口大小，流量控制可以分为三种：发送窗口1，接收窗口1（停止等待）；发送窗口N，接收窗口1（GBN）；发送窗口N，接收窗口M（SR）。
`N≥M>1`

停止等待协议

发收窗口数量都为1。发送端给接收端发送数据，等待接收端确认回复ACK，并停止发送新数据包，等待期间开启计时器。发送方发送数据后要对数据标号，第一次发送的数据标号为0，出错重传数据为1。

网络利用率（经常计算最大利用率）：
利用率 = 发送数据时间/总时间，
设发送数据时间t1，传播时间t2，确认帧的发送时间t3，传播时延t2，那么利用率h = t1/(t1+t2+t3+t2)。
有两种情况：① 等长确认帧/捎带：t1=t2；② 确认帧忽略不计：t3=0。

后退N帧滑动窗口协议GBN

发送窗口N，接收窗口1。发送N帧数据给接收端，接收端确认回复ACK，等待期间停止发送新的数据包并开启计时器。
连续发送，累计确认，哪里出错从哪儿传。当网络质量很好时，停止等待协议的性能很好。

网络利用率：
设发送窗口w，发送时间t1，传播时延t2，确认帧发送时间t3，传播时间t2，利用率h = w×t1/(t1+t2+t3+t2)≤1

选择重传滑动窗口协议SR

发送窗口N，接受窗口M（N≥M）。*连续发送，选择确认，哪里出错传哪里*。
对窗口编号确需确认的位数是n，满足2^n ≥ 发送窗口+接收窗口。

网络利用率：w×t1/(t1+t2+t3+t2)≤1

数据链路层流量控制

发送缓存和接收缓存；接收窗口大小为1时可保证帧的有序接收；窗口大小在传输过程中是固定的；只有窗口向前滑动并且接收方发送确认帧后，发送方才有可能向前滑动。

窗口大小的讨论：停止等待（发送1接收1），GBN（发送2^n-1 接收1），SR（发送2^(n-1) 接收2^(n-1)）

利用率的讨论：w×t1/(t1+t2+t3+t2)≤1，2×t2=RTT。
① 停止等待 w=1，GBN w=2^n-1，SR w=2^(n-1)；
② h ≤ 1
③ 捎带或等长确认，t1 = t3；
④ 确认帧时间很小或忽略不计时，t3 = 0

带宽讨论：① 最大带宽（实际带宽）；② 理论带宽（题目中给定）。
实际带宽 ≤ 1 = 理论带宽；
h = w×L/(t1+t2+t3+t2)，发送方窗口大小w，帧长L。等长确认时t1=t3，帧时间很小时t3忽略不计

介质访问控制★★★

总线型介质访问：分时和共享，使用一对多的广播通信方式，因此必须用专用的共享信道协议来协调这些主机的数据发送。
介质访问控制分为：信道划分介质访问控制，随机访问介质访问控制（争用型介质访问控制），轮训访问介质访问控制。

信道划分介质访问控制

带宽访问时，可以在一条介质上同时携带多个传输信号来提高传输系统的利用率，也就是多路复用。具体分为：频分、时分、波分、码分 多路复用。

复用允许用户使用一个共享信道进行通信，降低成本，提高率用率。

- **频分复用FDMA**：用户分到一定频带后在通信过程中始终占用该频带。所有用户在同样时间占用不同的频率带宽资源
- **时分复用TDMA**：将时间切位各个时间片，让用户分时使用
- **波分复用WDMA**：光的分频复用，所以只适用于光纤。使用同一根光纤同事传世多个光载波信号
- **码分复用CDMA**：把每个比特时间分为m个短间隔，称为码片，每个站被指派一个唯一的m比特码片序列：发送比特1，则发送自己的m比特码片；发送比特0，则发送该码片序列的二进制反码
    - 让站的码片和各个分组进行向量内积（规格化内积）：内积结果正值，该站发送数据1；内积结果负值，该站发送数据0；内积结果0，该站不发送数据

随机访问介质控制

当若干计算机使用一条信道发送数据，需要去共享信道。随机接入就意味着所有用户都可以根据自己的意愿随机地发送信息，这样会容易产生冲突而导致冲突用户发送数据失败。

- **ALOHA协议**：网络中任何节点需要发送数据时，可以不进行任何检测就发送。若在一段时间内没有收到确认，则认为传输过程发生冲突。发生冲突后等待一段时间重发，直到发送成功。成功率18.4%
- **CSMA协议**：由ALOHA协议改进。发送数据前侦听其它设备是否在发送数据。侦听策略分为：
    - 坚持型：信道空闲时持续发送数据，忙时持续监听
    - 非坚持型：信道空闲时理解发送数据，忙时等待随机时间再监听
    - p-坚持型：信道空闲时以概率p发送数据、以概率1-p不发送数据，忙时等待随机时间再监听
- **CSMA/CD协议**：应用于有线网。当多个站点同时在总线发送数据，总线电压变化值增大，当检测到信号电压摆动超过一定门限值，就会认为总线上至少两个站在同时发送数据（碰撞）。
    - 工作过程分为：先听后发，边听边发，冲突停发，随机重发
    - 当某个站监听到总线空闲时，可能总线并非真的空闲。最先发送数据帧的站，最多在发送数据帧后至多2τ(两倍的端到端往返时延) 就可以知道发送的数据帧是否遭受到碰撞
- **CSMA/CA协议**：应用于无线网。发送数据前检查信道状态，等信道空闲时再等待一段时间后检测信道是否空闲。若空闲则直接发送数据，否则随机等待一段时间后重发。有三种信道空闲检测方式：
    - 能量检测：对信道能量进行检测，大于一定值时认为信道被占用
    - 载波检测：对接收信号与本地伪随机码PN码进行运算比较，超出一定值则表示信道被占用
    - 能量和载波混合检测：先向目标端发送请求传送报文RTS，等待收到目标端响应报文CTS，发送端才开始发送数据。若没有收到确认帧则会重传，经过若干次重传仍然失败后则会放弃重传

轮询访问介质访问控制

主要用在令牌环局域网。典型协议是令牌传递协议。
令牌环局域网把多个设备安排为一个物理或逻辑连接环，令牌在这个环上依次传递。若有设备需要发送数据，则在等待令牌传递到该设备后，由令牌承载数据发送到接收端。

局域网的数据链路层★

构建局域网后，局域网内的主机在同一网络。局域网覆盖地理范围小、只在相对独立的局部范围内，专门铺设的传输介质进行联网、传输效率高，通信延迟时间短、可靠性高，支持多种传输介质。
主要技术要素：网络拓扑结构，传输介质，介质访问控制方法。

局域网的标准：以太网

DIX EthernetV2是第一个局域网产品规约，IEEE802.3是第一个IEEE标准。二者只有很小的差别，可以将802.3局域网简称为以太网。
IEEE802将局域网分为 逻辑链路控制LLC子层、媒体介入控制MAC子层。
与接入到传输媒体有关的内容都在MAC子层，LLC子层与传输媒体无关。任何协议的局域网对LLC子层都是透明的。

以太网提供无连接的不可靠服务，尽最大努力交付。发送的数据采用曼彻斯特编码。
以太网资源的争用采用CSMA/CD协议。10Mbit/s的一台用以51.2μs为争用期，期内可发送512bit(64字节)，若前64字节均无冲突，则后续数据就不会发生冲突。

最短有效帧长 = 争用期×发送速度 = 2×(介质长度/传播速度) × 发送速度，
以太网规定最短有效帧为64字节，则小于64字节的帧都是由于冲突而一场终止的无效帧。

以太网的MAC层

以太网上的计算机都连接在一根总线上，易于实现广播通信。
对于一对一通信，则将接收站的硬件地址写入帧首部的目的地址字段，当数据帧中的目的地址与适配器的硬件地址一致，才能收到该数据帧。

局域网中**硬件地址/物理地址/MAC地址**，802标准所说的地址严格意义上是每个站的名字或标识符。
MAC地址由48位构成，前3字节是组织唯一标识符，后3字节是扩展唯一标识符。
MAC地址类型：单播帧(一对一)，广播帧(一对全体)，多播帧(一对多)。后二者只用于目的地址。

网络接口板/通信适配器/网络接口卡NIC/网卡。所配备重要功能：进行串并行转换，对数据进行缓存，在计算机的操作系统安装设备驱动程序，实现以太网协议。
适配器每从网络上收到一个MAC帧就先用硬件检查MAC帧中的MAC地址：若是发往本站的帧则收下，否则丢弃。

以太网的帧格式

以太网重用的MAC帧有：DIX EthernetV2，IEEE的802.3标准。最常用V2。 |目的地址6 |源地址6 |类型2 |数据46~1500 |FCS4 |

无效的MAC帧：数据字段的长度与字段的值不一致；帧的长度不是整数个字节；收到的FCS查出有误；数据字段长度不在46~1500字节之间（有效的MAC帧长度为64~1518字节）。
检查出无效的MAC帧丢弃，以太网不负责重传丢弃帧。

FCS的存在目的是校验，当传输媒体误码率为1×10^(-8)时，MAC子层可使未检测到的差错小于1×10(-14)。

在MAC帧前面插入由硬件生成的8个字节，前7字节是前同步码（迅速实现MAC帧的比特同步），后1字节是帧开始定界符。

广域网的数据链路层

广域网的目的是为了远距离传输而存在，大多数依赖于海底光缆。
范围涵盖很大的物理区域，覆盖从几十公里到几千公里，能链接多个城市或国家或洲以提供远距离通信。
使用协议：PPP、HDLC，由节点交换机组成，层次为下三层。

PPP协议

应满足的要求：简单，封装成帧，透明性，多种网络层协议，差错检测，检测连接状态，最大传送单元，网络层地址协商，数据压缩协商。

不需要的功能：纠错，流量控制，序号，多点线路，半双工或单工链路。

三个组成部分：一个将IP数据报凤凰岛串行链路的方法；链路控制协议LCP（创建链路完成链路的启动、测试、任选参数的协商和最终链路的断开）；网络控制协议NCP（调用链路层创建阶段选定的网络控制层协议）。

帧格式：

```
|F(7E) |A(FF) |C(03) |协议 |信息部分（IP数据报，不超过1500字节） |FCS |F(7E) |
|1 |1 |1 |2 |信息部分（IP数据报，不超过1500字节） |2 |1 |
```

PPP协议以`0x7E`开头

当PPP在异步传输时，就用一种特殊的字符填充法；当PPP在同步传输链路时，协议规定采用硬件来完成比特填充。

字符填充：
将信息字段中出现的`0x7E`替换为二字节序列`(0x7D),(0x5E)`，`0x7D`替换为二字节序列`(0x7D,0x5D)`。

透明传输问题：
在发送端发现5个连续的1会立刻填入一个0，在接收端则发现连续5个1则会将其后面一个0删除。

PPP协议的特点：不使用序号和确认机制；面向字节，所有PPP帧长度都是整数字节；只支持全双工链路；只支持全双工链路；面向连接的不可靠的协议；具有身份验证功能。

PPP协议的运行分为四个阶段

建立链路LCP，验证PAP、CHAP，网络控制协商NCP，终止PPP链路LCP

- *建立链路LCP*：PPP首先用LCP在链路两端建立连接，并且在两端动态协商一些参数，例如认证方式、是否支持压缩和MLP等
  - 若要完成建立、配置、测试、终止数据链路连接工作就需要通信两端互相交换LCP报文，基本上有三类：链路配置报文，链路维护报文，链路终止报文
  - LCP协议在建立两端链路链接还会经历的四种状态：初始化状态Initial 或 准启动状态Starting，请求发送Request-Sent，确认发送Ack-Sent，打开Open
  - LCP协商一些配置选项，发生LCP的配置请求帧：配置确认帧，配置否认帧，配置拒绝帧
- *验证PAP、CHAP*：验证协议有很多，包括 口令验证协议PAP、挑战握手身份验证协议CHAP、微软挑战握手身份验证协议2版本MS-CHAPv2，可扩展的身份验证协议EAP。其中PAP和CHAP用的最多
- *网络控制协商NCP*：负责建立并配置IP、IPX、AppleTalk等网络层协议，以及建立并中毒那多种第三层协议会话。NCP是PPP协议的另一个子层，主要作用是在通信亮度那协商网络层的参数（IP地址、DNS等）。PPP协议支持多协议栈，所以在不同协议栈中的NCP名称不一样（在IP协议栈中称为IPCP、在IPX协议栈中称为IPXCP）
- *终止PPP链路LCP*：LCP对链路直接终止

数据链路层设备★

碰撞域/冲突域指网络中一个站点发出的帧会与其他站点发出的帧产生碰撞或冲突的那部分。
广播域指网络中的任一设备发出的广播通信都能被网络中所有其它设备所接收。
在数据链路层使用**网桥**来扩展局域网。网桥工作在数据链路层，根据MAC帧目的地址对收到的帧进行转发。收到一个帧后会先检查此帧的目的MAC地址，再确定将该帧转发到对应接口。

网桥的优缺点： **优点**：过滤通信量，扩大了物理范围，提高了可靠性，可互联不同物理层、MAC子层、速率的局域网。
缺点：存储转发增加了时延，MAC子层没有流量控制功能，不同MAC子网的网段桥接在一起的时延更大，不适用于用户数高于几百个的大通信量局域网（容易广播风暴）。

透明网桥

目前使用最多的事透明网桥，“透明”指网络上的站点不知道所发送的帧将经过几个网桥，网桥对各站是不可见的。

透明网桥具有自学习算法来处理收到的帧和建立转发表。每收到一个帧后，将纪录其*源地址*和*进入网桥的接口*、*帧进入网桥的时间*，作为转发表中的一个项目。
收到帧后进行自学习，查找转发表与收到的源地址匹配。若无，则再转发表中增加项目[源地址，进入的接口，时间]。
转发帧时，匹配转发表：若无，则通过所有其它接口进行转发；若有，则按转发表给出的接口进行转发；若转发表给出的接口就是该帧进入网桥的接口，则应该丢弃。

广播时可能会存在回路问题，造成大规模网络资源浪费。于是引入一个生成树协议STP，不改变网络实际拓扑，但在逻辑上切断某些链路，使得一台主机到其它主机的路径都是无环路的树状结构。

原路由网桥

易安装但网络资源利用不充分。在发送帧时将详细路由信息放在帧首部。源站以广播形式向目的站发送一恶搞发现帧，每个发现帧都记录所经过的路由。
发现帧达到目的站时就会沿各自的路由返回源站，源站在得知这些路由后，从所有可能路由中选择一个最佳路由。凡从该源站向目的站发送帧的首部，都必须携带源站所确定的这一路由信息。

交换式网桥/以太网交换机/第二层交换机

通常有十多个接口（本质是一个多接口的网桥，可见交换机工作在数据链路层）。
每个接口以全双工的方式直接与主机相连，交换机能同时联通许多对接口让每一个相互通信的主机能进行无碰撞传输数据（如同独占通信媒体）。

工作方式

- **直通式交换**：检查前六个字节后转发，认为小于六个字节的数据报是碎片而不进行转发（转发延迟=发送6字节的发送延迟）。快速但缺乏安全性，无法支持不同速率的端口交换
- **存储转发式**：先把数据收下来，检查无误后再查找转发表发送（有误则丢弃）。可靠性高，但延迟较大

设备带宽的讨论

集线器和中继器处于同一冲突域，共享带宽。
网桥和交换机以全双工工作方式，独占带宽（对于N个接口的交换机而言，总带宽为N×单用户带宽）。但对于普通10Mb/s的共享式以太网，每个用户只有10Mb/s÷用户数。

虚拟局域网VLAN

在一个物理LAN内划分出多个虚拟LAN，每个VLAN是一个广播域，这就缩小了广播域范围，各个VLAN之间不能直接通信。

VLAN中的交换机端口可以分为：访问链接AccessLink；汇聚链接TrunkLink。
设置VLAN的顺序是：生成VLAN，设定访问链接（决定各端口属于哪个VLAN）。

设定访问链接的手法，分为两种：静态VLAN（事先固定），动态VLAN（根据所连计算机而动态改变设定）。

动态VLAN分为：基于MAC地址的VLAN，基于子网的VLAN，基于用户的VLAN

补充

- 数据链路层的功能：为网络层提供服务，帧定界、同步、透明传输，流量控制和差错控制
- 连续ARQ：GBN、SR。若窗口值以n比特编码，发送窗口最大值为2^n-1
- 有序接收：停止等待、GBN
- 发送窗口大小为a，则最少需要$n \geq \log_2(a+1)$位序列号来保证协议不出错
- 若GBN协议发送了0~7帧，收到了0、2、3号帧的确认，则发送方需要重复4
- A、B、C通过CDMA共享链路，A的码片序列是(1,1,1,1)，C从链路上收到的序列为(2,0,2,0,0,-2,0,-2,0,2,0,2)，计算C收到A发送的数据：
    - C收到的序列可排序为[[2,0,2,0],[0,-2,0,-2],[0,2,0,2]]，逐行乘以A的码片序列后，可得到(4,-4,4)，所以C收到A的数据为101
- CSMA/CD协议中的“争议期”：信号在最远两个断电之间往返传输的时间
- 以太网中发生介质访问冲突，则按照二进制指数回退算法决定下一次重发时间，因为该算法考虑到了网络负载对冲突的影响
- 采用二进制指数回退算法处理冲突时，首次重传的帧是再次发生冲突概率最低的
- 二进制回退算法中，设k为碰撞次数：$k \leq 10$，$k=k$；$10 \leq k \leq 16$，$k=10$；$k \geq 16$，报错
- 以太网的MAC协议提供无连接的不可靠服务

网络层

向上提供简单灵活的、无连接的、不可靠、尽最大努力交付的数据报服务。
网络在发送分组时不需要先建立连接。不提供服务质量的承诺。

网际协议IP 是TCP/IP体系的最主要协议，配套有 *地址解析协议ARP*、*网际控制报文协议ICMP*、*网际组管理协议IGMP*

IP地址

给连接在网络的主机或路由器分配一个唯一32的位地址，每8位为一组

基本分类IP

IP地址是一个分层架构，第一段是标志主机/路由器所连接网络的网络号net-id，第二段是标志主机/路由器的主机号host-id。
|网络号 |主机号 |，根据划分长度不同分为 A、B、C、D、E 类地址。
A：0~126；B：128~191；C：192~223；D：224~239；其余归为E留作备用。

一些特殊的IP地址：

网络号	主机号	源地址使用	目的地址使用	代表的意思
0	0	可以	不可	在本网络上的本主机
0	host-id	可以	不可	在本网络上的某台主机host-id
全1	全1	不可	可以	只在本网络上进行广播
net-id	全1	不可	可以	对net-id上的所有主机进行广播
127	非全0或全1的任何数	可以	可以	用于本地软件环回测试

> 广播地址：主机位全为1；网络地址：主机位全为0。
> 此二者无法分配给任何主机。

网络类别	最大可指派的网络数	第一个可指派的网络号	最后一个可指派的网络号	每个网络中最大主机数
A	126 (2^7-2)	1	126	16777214
B	16383 (2^14-2)	128.1	191.255	65534
C	2097151 (2^24-2)	192.0.1	223.255.255	254

划分子网

从两级IP划分为三级IP地址。从主机号host-id借用若干单位作为子网号subnet-id（主机号也相应减少若干位），不改变IP地址原来的网络号net-id。

发送到某个单位指定主机的IP数据报，由路由器接收后再按目的网络号和子网号找到目的子网，并交付目的主机。

子网划分是在分类IP的基础上完成的。规定子网号和主机号不能为全0或全1.子网掩码决定了子网号，即使网络地址写法相同，但子网掩码不同也表示了不同网络。

子网掩码subnetMask：

引入子网掩码来表示IP中的子网部分。
规则：子网掩码长度=32位；左边部分一连串1对应网络号和子网号；右边部分一连串0对应主机号。
默认子网掩码：A 255.0.0.0，B 255.255.0.0，C 255.255.255.0。
`（IP地址）AND（子网掩码）= 网络地址`

不同的子网掩码得出相同的网络地址，但不同掩码的效果是不同的。

无分类编址CIDR

消除了传统ABC类地址以及子网划分，可更有效地分配IPv4的地址空间。使用各种长度的“网络前缀network-prefix”来代替分类地址中的网络号和子网号。从划分子网的三级回到了两级。

使用“斜线记法slashNotation”，在IP地址后面加一个斜线“/”写上网络前缀所占位数（此数值对应三级编址子网掩码中1的个数）。
eg. 128.14.32.0/20共有2^12个地址

路由聚合：

一个CIDR地址块可以表示很多地址，让路由表中的一个项目可以表示很多个原来传统分类地址的路由。路由聚合也称为构成超网supernetting。
路由聚合相当于取公共前缀。

私有IP

本地地址（专用地址、私有地址）——仅在某组织内部使用的IP地址，可由组织自行分配，不需要想互联网管理机构申请。
全球地址——全球唯一IP地址，必须向物联网管理结构申请。

三个专用的地址块：A类：10.0.0.0/8，24位块；B类：172.16.0.0/12，20位块；C类：192.168.0.0/16，16位块。
使用专用IP地址的网络称为专用互联网或本地互联网，仅在某个组织内部使用。

网络地址转换NAT：使用本地地址的主机在和外界通信时，都要在NAT路由器上将本地地址转换成全球IP地址，才可以使用互联网

IPv6

将地址从IPv4的32位扩大到128位。

由 40字节的基本首部 和 不超过65535字节的有效负荷 组成。有效载荷允许有零个或多个首部扩展，后面是数据部分。
`|基本首部 |有效载荷 |`，
每个16位值用16进制值表示，各值之间用冒号分隔。16进制记法中允许省略前面的0（eg. 0000写为0）

IPv6数据报地址：

地址类型	二进制前缀
未指明地址	::/128
环回地址	::1/128
多播地址	FF00::/8
本地链路单播地址	FE80::/10
全球单播地址	(除上述四种外的其它前缀)

IPv4到IPv6的过渡

逐步严谨，向后兼容（IPv6能够接受和转发IPv4分组，并能够成为IPv4分组选择路由）。有两种向IPv6过渡策略：
双协议栈（同时装有IPv4和IPv6协议栈），
隧道技术（把IPv6封装成IPv4数据报）。

ARP协议

网络层使用IP地址，数据链路层使用MAC。它们需要去做对应的映射。
ARP的作用，是从网络层使用的IP地址，解析出在数据链路层使用的硬件地址。

RARP是实现从MAC地址到IP地址的映射。每个主机都有ARP高速缓存(ARP Cache)，存放网络上各主机和路由器的IP地址到硬件地址的映射表。

广播和单播硬件地址：设A向B发送，建立请求和响应的过程（解析）

1. ARP广播：源IP是A，目的IP是B；源MAC是A，目的MAC是广播
2. ARP响应(单播)：源IP是B，目的IP是A；源MAC是B，目的MAC是A

如果A和B不在同一个网络，则需要通过ARP找到位于本网络的某个路由器硬件地址，然后把分组发给该路由器，并让其分组转发给下一个网络。剩下的工作由下一个网络完成。

ICMP协议

有效地转发IP数据报和高交付成功的机会，允许主机或路由器报告差错情况和提供有关异常情况的报告。报文格式如下：
`|类型0~7 |代码8~15 |检验和16~31 |`(前四个字节都一样，取决于ICMP报文类型)
`|ICMP数据部分(长度取决于类型)|`

ICMP差错报告报文

终点不可达，时间超过，参数问题，改变路由(重定向Redirect)，源抑制

不发送ICMP差错报文的情况：
对ICMP差错报告报文不再发送；对第一个分片的数据报片的所有有序数据报片都不发送；对具有多播地址的数据报都不发送；对具有特殊地址（127.0.0.0或0.0.0.0等）的数据报都不发送。

ICMP询问报文

请求和回答报文，事件请求和回答报文。

ICMP的应用

PING用来测试两台主机的连通性和联通质量。

Tracerout用于追踪一个分组从原点到终点的路径（windows命令为tracert）

IP首部格式

IP数据报由首部和数据两部分组成（首部采用固定时是20字节，所有IP数据报都要有。数据部分长度可变，1-40字节不等）。

当数据量很大时允许分片：标志flag占三位，取前两位有意义。最低位是MF，=1表示后面还有分片，=0表示最后一个分片；中间位是DF，=0时允许分片。
生存时间TTL，可通过路由器数的最大值（每经过一个路由器将会-1，为0时丢弃，并向源告知超时）。

IP组播

- IP多播：实现更好的一对多通信。对外只需发送一份，让多播路由器复制后发送给多个成员。尽最大努力交付
  - 软件多播：在网络层实现多播
  - 硬件多播：TCP/IP协议使用到的以太网地址块范围是`01-00-5E-00-00-00 ~ 01-00-5E-7F-FF-FF`。
  - IGMP协议：利用网络组管理协议把网络上的路由器协同工作，一遍以最小代价完成多播

路由选择

当两台非直连计算机需要经过几个网络通信时，路由器提供路由选择协议来开辟一个最佳网状连接路径。
路径必须需是正确且完整，计算上简单，能适应通信量和网络拓扑变化（要有自适应性，稳定性，公平性，最佳性）。

静态路由选择策略：简单，开销小，不能适应网络状态变化。
动态路由选择：自适应，实现复杂，开销大。

自治系统AS

在单一技术下管理的一组路由器，使用AS内部路由选择协议和共同的度量来确定分组在该AS内的路由，同时使用一种AS之间的路由选择协议用于确定分组在AS之间的路由。

虽然一个AS会使用多种内部路由选择协议和度量，但一种AS对另一种AS表现出的是一个单一和一致的路由选择策略。有两类主要协议：

内部网关协议IGP：自治系统内使用的路由选择协议，使用最多（例如RIP和OSPF）。
外部网关协议EGP：将一个自治系统的路由选择信息传递到另一个自治系统。BGP-4使用最多。

协议	层次	下层	范围	原理
RIP	应用层	UDP	AS内	距离向量

协议	层次	下层	范围	原理
OSPF	网络层	IP	AS内	链路状态
BGP	应用层	TCP	AS间	路径向量

内部网关协议IGP

RIP

分布式的、基于距离向量的路由协议。要求每个路由器都要维护从自己到其它每一个目的网络的距离记录。由此形成一个路由表：
|目的网络 |下一跳 |距离 |。
一个路由器下的网络到另一个直连网络的距离定义为1，从一个路由器到非直连网络的定义为所经过路由器数+1——也被称作跳数。

RIP认为一个好的路由即是经过的路由器数目最少——即为“距离”，允许一条路径最多只包含15个处理器。“距离”到达最大值16时相当于不可达。所以*RIP只适用于小型网络*。

刚开始工作时，只知道直连网络距离为1，且路由表为空。之后每个路由器也只和数目有限的相邻路由器交换，并更新路由信息。经过若干更新后所有路由器都会知道到达本自治系统中任何网络的最短距离，也就是“收敛”。

RIP仅和相邻路由器按固定时间间隔交换信息，交换的是本路由器的全部信息（路由表）。使用的是Bellman-Ford算法。RIP让互联网中的所有路由器都和自己相邻的路由器不断交换信息，最终每个路由器都知道相互的距离关系。

RIP好消息传得快，坏消息传的慢。当网络出现故障，要经过较长时间（数分钟）才能将此信息传递到所有路由器。

OSPF

基于Dijkstra算法找最短路径，采用分布式链路状态协议。
使用洪泛法向本治系统中所有路由器发送信息。发送的信息是与本路由器相邻的所有路由器的链路状态，链路状态说明本路由器都和哪些路由器相邻，以及该链路的“度量”。只有链路状态变化时才会发送信息。

适用于大型动态网络。

链路状态数据库：所有路由器建立起的全网拓扑结构图。

OSPF使用层次结构区域划分。在上层的称作主干区域，标识符规定为0.0.0.0，作用是连通下层区域。

OSPF的五种分组（确定可达性，达到数据库同步，新情况同步）：问候Hello分组；数据库描述databaseDescription分组；链路状态请求linkStateRequest分组；链路状态更新linkStateUpdate分组；链路状态确认linkStateAcknowledgment分组。

外部网关协议BGP-4

实现在AS之间交换路由信息的协议，简写BGP。力求寻找一条能够到达目的网络且比较好的路由（不能兜圈子），而且并非要寻找一条最佳路由。

BGP发言人：一个AS至少选择一个路由器作为该AS的BGP发言人。两个发言人通过一个共享网络连接在一起，而BGP发言人往往是BGP边界路由器。发言人交换路径向量。
交换过程要建立TCP连接，在此连接上交换BGP报文来建立BGP会话并交换路由信息。

BGP的四种报文：打开OPEN，更新UPDATE，保活KEEPALIVE，通知NOTIFICATION。

数据转发流程

IP层转发分组流程

按主机错在网络地址制作路由表，那么每个路由器中的路由表就只包含四个项目。在路由表中对每一条路由，最主要的是(目的网络地址，下一跳地址)。

特定主机路由：为特定目的主机指明一个路由，方便网络管理人员控制和测试网络。
默认路由：不会进行任何形式的检查（0.0.0.0/0），适用于只有一个路由器和互联网连接的小型网络。

查找路由表：根据目的网络地址就能确定下一跳路由器，IP数据报最终一定可以找到目的主机所在的目的网络上的路由器，只有到达最后一个路由器时才试图向目的主机进行直接交付。

使用子网分组转发

划分子网的情况下从IP地址不能唯一地得出网络，因为网络地址取决于那个网络所采用的子网掩码且数据报首部不包含。因此分组转发算法需要做改动。

最长前缀匹配

又被称为最长匹配or最佳匹配。

使用CIDR时，路由表的每个表项由“网路前缀”和“下一跳地址”组成。查找路由表时可能会得到不止一个匹配结果。
应当从匹配结果中选择具有最长网络前缀的路由：最长前缀匹配longest-profixMatching。网络前缀越长，地址块越小，路由越具体。

路由器

一种典型的具有多输入多输出端口的网络层设备，关键设备，主要作用：联通不同网络，路由选择，分组转发。

互联网设备总结：

设备名	中继器	集线器	网桥	交换机	路由器
冲突域	×	×	√	√	√
广播域	×	×	×	×	√
层次	物理层	物理层	数据链路层	数据链路层	网络层
工作原理	信号再生	信号再生	根据MAC地址转发	根据MAC地址转发	根据IP地址转发

SDN

软件定义网络software-DefinedNetworking是一种网络架构和管理方法，通过将网络控制平面controlPlane与数据转发平面dataPlane分离，实现网络的灵活性、可编程性和集中化管理。SDN将网络硬件设备的控制功能集中到一个中央控制器中。

其核心思想是将网络的控制逻辑从网络设备中抽离出来，使网络设备成为可编程的数据转发平面。网络管理员可以通过中央控制器对整个网络进行集中管理和编程。
SDN的优势包括简化网络管理、提高网络灵活性、降低网络成本和加速网络创新等，使网络能够更好地适应不断变化的需求和应用场景。

北向API面向用户，南向API面向资源。

移动网络

1. 归属代理：一个移动节点归属网上的路由器至少有一个接口在归属网上。当移动节点离开归属网时，通过IP通道吧数据报传给移动节点，并且负责维护移动节点当前的位置信息
2. 外区代理：移动节点当前所在网络上的路由器，向已登记的移动节点提供选路服务
3. 归属地址：识别端到端连接的静态地址。不论移动节点连接到网络何处，其归属地址保持不变
4. 转交地址：隧道重点地址。可能是外区代理转变地址，也可能是主流本地的转变地址。外区代表转交地址是外区代理的一个地址，移动节点利用其登记
5. 位置登记：移动节点必须向其所在位置进行登记
6. 代理发现：移动节点必须首先找到一个移动代理，便于随时随地与其他节点通信
7. 隧道技术：移动节点在外区网时，归属代理需将其原始数据包转发给已登记的外区代理

补充

- 在ISO/OSI参考模型中，网络层的主要功能是路由选择、拥塞控制、网络互联
- 常用的十进制转二进制：128 = 1000 000；192 = 1100 000；224 = 1110 0000；240 = 1111 0000；248 = 1111 1000；252 = 1111 1100；254 = 1111 1110；255 = 1111 1111
- .../30是点对点，只有2个有效位，通常作为PPP协议的两端
- 网络优先：当每个网络中主机数相等或等大时
- 主机优先：每个网络中主机数不等，优先分配大的网络
- 局域网的分配可以带入哈夫曼算法
- IPv6零压缩只能用一次
- 直接为ICMP提供服务协议的是IP
- IP数据报报头中，报头长度字段以32比特为计数单位，总长度字段以8比特为计数单位
- 互联网中，IP数据报传输的路径经由的主机和中途路由器，源主机和中途路由器通常不知道完整路径

传输层

面向用户的最底层，面向通讯的最高层。只有位于网络边缘的主机的协议栈才有运输层。
运输层提供的应用进程到应用进程的通信（端到端）。

传输层的功能

连通每个计算机的通信模式。实现**流量控制、可靠传输、拥塞控制**功能。
提供了TCP或UDP报文的复用和分用，根据端口号提供复用与分用的功能。

- TCP：连接的可靠的全双工信道。不支持多播、广播，用于大多数应用
- UDP：无连接的不可靠的信道。支持单播、多播、广播

两个对等运输实体在通信时传送的数据单位是运输协议数据单元TPDU。TCP传输TCP报文段，UDP传输UDP报文或用户数据段。

端口号

对TCP/IP体系的应用程序进行标志。把要传输报文交到目的主机的某一个合适的目的端口，最后交付目的进程由TCP完成。

端口是16位，允许65535个不同端口号。端口号仅具有标记本地计算机应用层各进程的作用。

常用端口：熟知端口0~1024；登记端口1024~49151，使用时需在IANA登记防止重复；客户端端口49152~65535，短暂端口号，给用户进程暂时使用。
常用熟知端口：
UDP：RPC111, DNS53, TFTP69, SNMP161, SNMP(trap)162;
TCP：SMTP25, FTP20、21, Telnet23, HTTP80, HTTPS443

UDP协议

UDP仅在IP数据报服务上增加了 复用/分用、差错检测 功能。
其中复用使用的是源端口，分用使用的是目的端口。

UDP是无连接的，发送数据之前不建立连接，减少开销和发送数据之前的时延。提供面向报文的不可靠交付，一次性交付一个完整报文。没有拥塞控制。支持一对一、一对多、多对一、多对多的交互通信。首部开销小，只有8字节：
|源端口 |目的端口 |长度 |校验和 |;
由于一次性交付完整报文，应用程序必须选择合适的大小。

基于端口的分用：UDP数据报到达时，根据首部中的目的端口分发给一个或多个端口。

UDP校验和：把首部和数据部分仪器检验。

TCP协议

面向连接的运输层协议，在无连接、不可靠的IP网络服务基础上提供可靠交付的服务。在IP数据报服务商增加了保证可靠性的一系列措施。
面向连接、面向字节流、点对点、可靠交付、全双工通信。

面向字节流：把应用程序交付的数据看成一连串无结构的字节流。每次发送的都是一个数据段，每个段都将会编号排序，对数据顺序进行控制，当其中一个数据段丢失将会重发。

TCP是基于socket的虚连接，而非物理连接；不关心应用进程的报文发送到TCP缓冲的长度；根据对方给出的窗口值和当前网络拥塞程度决定一个报文段应包含多少字节；可把过长的数据块划分变短后传送，也可积累足够多的字节后再构成报文段发送。

TCP的连接

TCP连接的端点是套接字socket，端口号拼接到IP地址构成套接字。每条TCP连接唯一地被通信两端的两个端点/套接字所确定。即：`TCP连接::= {socket1, socket2} = {(IP1:port1),(IP2:port2)}`

TCP首部格式

TCP报文段首部的前20个字节固定，后面4n个字节根据需要而增加。因此最小长度是20字节。

```
| 源端口16                    | 目的端口16                   |
| 序号32                                                    |
| 确认号32                                                  |
|数据偏移4 |保留6 |URG1|ACK1|PSH1|RST1|SYN1|FIN1 |窗口16 |
| 校验和16                    | 紧急指针16                   |
| 选项（长度可变）                                 |填充   |
```

- 源端口和目的端口：运输层与应用层的u无接口，复用/分用都要通过端口实现
- 序号：传送数据流中的序号
- 确认号：期望收到对方下一个报文段的数据的第一个字节的序号
- 数据偏移：报文段的数据起始处距TCP报文段的起始处的距离
- 保留：今后使用，目前为0
- URG=1时有高优先级；ACK=1时确认字段有效；PSH=1时会尽快交付接收应用进程，不再等到缓存满后在上交；RST=1时表示TCP连接出现严重差错，必须释放连接；SYN=1时表示这个是一个连接请求或连接接受报文；FIN=1时表示此报文段的发送端数据已发送完毕，并要求释放运输链接
- 窗口：让对方设置发送窗口的依据
- 检验和：检验和字段检验的范围包括首部和数据这两部分
- 紧急指针：报文段中紧急数据的字节数量
- 选项字段：常用的选项是MSS，它告诉对方TCP发送方的报文段的数据字段最大长度为MSS个字节
- 其它：
    - 窗口扩大选项3：其中一个字节表示移位值S，新窗口值=TCP首部中的窗口位数增大到(16+S)，相当于窗口值向左移动S后获得的实际窗口大小
    - 时间戳选项4：字段时间戳值字段，时间戳回送回答字段
    - 选择确认选项
- 填充：让整个首部长度为4字节的整数倍

TCP可靠传输

目前唯一的*可靠的*网络协议，面相连接的传输服务。

可靠传输的机制：
编号和确认机制；超时机制；自动重传机制。其中会有两种不同的协议：停止等待协议，连续ARQ协议。

> 停止等待机制：每发送一个分组就停止发送并等待对方确认，确认后再发下一个分组。
>
> 连续ARQ协议：发送方一次发出多个分组。使用滑动窗口协议控制发送方和接收方所传播分组的数量和编号，发送方每收到一个确认就把发送窗口向前滑动，接收方一般采用累计确认的方式。重传时使用回退N `Go-Back-N` 方法进行重传。

TCP拥塞控制

拥塞是指某短时间内网络中某资源的需求超出了该资源所能提供的可用部分，使网络性能变坏，最坏结果是系统崩溃。
拥塞的原因有：缓存容量太小，链路容量不足，处理机处理速率太慢，拥塞本身进一步加剧拥塞。

拥塞控制和**流量控制**是不同的。流量控制是一直发送端发送的速率，是点对点通信量的控制；拥塞控制是一个全局性的过程，涉及与降低网络传输性能的所有因素。

有两种控制类型：

- **开环控制**：设计网络时事先考虑周全，力求工作时不拥塞，力争避免发生拥塞
- **闭环控制**：根据网络当前状态采取相应控制措施，发生拥塞后采取措施进行控制已达到消除拥塞的目的

TCP拥塞控制的基本概念：
一个传输轮次所经历的时间是往返时间RTT。传输轮次更加强调把拥塞窗口cwnd所允许发送的报文段都连续发送出去，并受到了对已发送的最后一个字节的确认。

控制拥塞窗口的原则：
只要网络没有出现拥塞，拥塞窗口就可以再增大一些以便发送更多分组，提高网络利用率。出现拥塞或可能出现拥塞时，则反之来缓解拥塞。

拥塞判断方式：重传定时器超时，收到三个重复的ACK。

拥塞控制算法：

- **慢开始slowStart**：用于确定网络的负载能力或拥塞程度，由小到大逐渐增加拥塞窗口数值，两个变量：拥塞窗口（1MSS逐渐增大），慢开始门限
  - 拥塞窗口cwnd控制方法：没收到一个对新报文段的确认后，可以把拥塞窗口最多增加一个SMSS数值。`拥塞窗口cwnd每次增加量 = min(N, SMSS)`
  - 每经过一个传输轮次，拥塞窗口加倍，指数增长，cwnd+1
- **拥塞避免算法**（已经超时时使用）：让拥塞窗口cwnd缓慢增大，避免拥塞
  - 每经过一个传输轮次，`拥塞窗口cwnd = cwnd+1`，线性增长
  - 拥塞避免并非指完全避免了拥塞，仅仅是把拥塞窗口控制为按现行规律增长来尽量不容易出现拥塞
  - 当拥塞出现时：`ssThresh = max(cwnd/2, 2), cwnd = 1, 执行慢开始算法`。当拥塞窗口cwnd增长到慢开始门限值ssThresh时，就改为拥塞避免算法
- **快重传算法fastRetransmission**：发送方一连收到三个重复确认就知道接收方没有收到报文段，因此立刻重传，避免出现超时使发送方误认为出现网络拥塞
- **快速恢复算法fastRecovery**：发送端收到连续三个重复确认，认为网络很可能没有发生拥塞，因此执行快速回复算法而不是慢开始算法：

    - 慢开始门限ssThresh = 当前拥塞窗口cwnd / 2，新拥塞窗口cwnd = 慢开始门限ssThresh，开始执行拥塞避免算法

发送窗口的上限值：发送窗口的上限值应当取位接收方窗口rwnd和拥塞窗口cwnd这两个变量中较小者，按 发送窗口上限值 = min(rwnd, cwnd) 来确定

TCP连接管理

注：下面的+1表示加一个有效载荷。

连接建立的三个阶段（三次握手）：客户端C向服务器S发出请求连接的报文段，S收到后若同意则发回确认，C收到报文段后向S给出确认。此时C的TCP向上层应用进程通知连接已建立。
整体过程描述如下：

```
C                                    S
| ----------- SYN=1, seq=x ----------> |
| <-- SYN=1, ACK=1, seq=y, ack=x+1 --- |
| ----- ACK=1, seq=x+1, ack=y+1 -----> |
```

连接释放：C的应用进程向TCP发出连接释放报文段，主动关闭TCP连接，S发出确认。此时C到S的连接已释放，TCP连接处于半关闭状态。若S发送数据，C仍要接收；若S没有需要向C发送数据，其应用进程通知TCP释放连接。C收到释放报文段后发出确认。
整体过程描述如下：

```
C                                    S
| ----------- FIN=1, seq=u ----------> |
| <----- ACK=1, seq=v, ack=u+1 ------- |
| <-- FIN=1, ACK=1, seq=w, ack=u+1 --- |
| ----- ACK=1, seq=u+1, ack=w+1 -----> |
```

TCP连接必须经过时间2MSL后才真正释放掉。MSL是报文最大生存时间。所以最后C发出确认报文段时需要等待2MSL。

数据交换格式

- **电路交换**
    - 两部电话机用一对电线相互连接，“交换”就是按照某种方式分配专用的物理线路
    - 必须是面向连接的，三个阶段包括：建立连接，通信，释放连接
    - *优点*：用户专用，随时通信，实时性强，无时序问题，适用于模拟和数字信号，控制简单
    - *缺点*：建立连接时间较长，因为专用而利用率低，不同标准的中断难以通信，仅适合语音、不适合数据通信
- **报文交换**
    - 交换单位是携带有目标地址、源地址的报文，在交换节点采用存储转发的方式

- *优点*：通讯双方不需要建立专用线路，不存在连接建立时延，随时发送报文。采用存储转发提高了传输时的可靠性，易实现代码转换和速率匹配（便于不同标准的计算机之间通信）；提供多目标服务；允许建立数据传输的优先级。通信双方在不同时段分段占用物理通路，提高了线路利用率
- *缺点*：存储转发的时延大；只适用于数字信号；报文长度没有限制，每个中间结点都要完整报文

- **分组交换**
  - 在报文交换的基础上限制每次传送数据块的大小，并加上必要的控制信息(首部)构成分组（IP分组)
  - 每个分组都携带地址等控制信息，交换网中的节点交换机根据收到分组的首部来转发给下一节点，并采用存储转发的方式让所有分组到达目的地
  - *两种实现方式*:
    - 数据报：无连接方式，类似报文交换，每个分组都携带目的信息，在发送时每个分组可能走不同路径使得具有不同时延(多数是通顺序的)，但最终都会到达目的地点
    - 虚电路：面向连接，类似线路交换，建立虚连接，传输路径相对确定地从一个节点到下一节点不断存储转发，直到到达目的地。不存在数据报方式的排序问题，但需要建立呼叫请求建立路线

补充

- UDP协议没有全双工的说法
- UDP校验出错时直接将数据丢弃
- 发送窗口始终 = min(接收窗口rwnd, 拥塞窗口cwnd)

应用层

应用层是为了解决某一类问题而设立的，其具体内容是规定应用进程在通信时所遵守的协议。许多协议都是基于客户服务器模式，描述了进程之间服务和被服务的关系（客户是服务请求方，服务器是服务提供方）。

具体应用包含有：域名解析DNS，文件传输FTP，电子邮件Email，万维网WWW，动态主机配置DHCP，建立在FTP的电子邮件SMTP，远程主机TELNET

DNS

域名到IP地址的解析由域名解析服务器完成的（域名 -> IP映射），域名服务器程序在专设的节点上运行，运行该程序的机器称为域名服务器。
域名分为不同的层级，从根到顶级域名、二级域名、三级域名、四级域名...，呈现树形结构。

域名服务器：一个服务器负责管辖的范围叫区zone。
域名服务器也是分层的，从根域名服务器到顶级域名服务器、权限域名服务器。

> 根域名服务器是最高层次的、最重要的域名服务器，所有跟域名服务器都知道所有的顶级域名服务器域名和IP地址。任何本地域名服务器解析任一域名时若无法自己解析，就首先求助于根域名服务器。
> 顶级域名服务器负责挂你该顶级域名服务器的所有二级域名。收到DNS查询请求时就给出相应回答（可能是最后结果，也可能是下一步应当找的域名服务器的IP地址）。
> 权限域名服务器负责一个区的域名服务器，当权限域名服务器还不能给出最后查询回答时，就会告诉DNS客户下一步应当找拿一个权限域名服务器。
> 本地域名服务器也被叫做默认域名服务器，对域名系统很重要，一个主机发出DNS查询请求时，该查询请求报文就发送给本地域名服务器。

每个域名服务器都维护一个高速缓存来存放最近用过的名字和从何处获得名字映射信息的纪录，用以提高可靠性和速度。

DNS解析类型包括：
迭代解析（本地DNS有缓存时，最少0次 最多N次；无缓存时，最少1次 最多N次）；
递归解析（本地DNS有缓存时，最少0次 最多1次；无缓存时，最少1次 最多1次）。

FTP

互联网上最广泛使用的文件传送协议，提供交互式访问，允许客户知名文件类型并允许文件具有存取权限。FTP屏蔽了计算机系统的细节，适用于异构网络中任意计算机间传送文件。
屏蔽了计算机存储数据的格式不同、屏蔽文件目录结构和文件命名的规定不同、冰壁操作系统使用命令不同、屏蔽访问控制方法不同。

FTP使用客户服务器模式，一个FTP服务器进程可同时为多个客户服务。

FTP的两个连接：
*控制连接*在整个会话期间一直保持打开，FTP客户发出传送请求通过控制链接发送给服务器端的控制进程。
*数据连接*用于实际的传送文件，服务器端的控制进程接收到FTP客户发来文件传输请求后就创建“数据传送进程”和“数据连接”，连接客户端和服务器端的数据传送进程。

FTP的两个端口：
在客户进程向服务器进程发出建立连接请求时寻找*熟知端口21*，并向服务器进程告诉自己的另一个端口号，用于建立连接。
接着服务器进程用自己传送数据的*熟知端口20*与客户进程提供的端口号建立数据传送连接。
21是控制端口，20是数据端口（主动模式下）。在被动模式时由服务器端和客户端协商而定。

WWW

能够让web客户端浏览访问web服务器上页面的应用。万维网是超分布式超媒体系统，是超文本系统的扩展。
WWW的特性是：统一资源定位符URL，超文本传送协议HTTP，万维网的文档，万维网的信息检索系统。

URL：对从互联网上得到资源的位置和访问方式的一种简介表示。URL给资源位置提供一种抽象识别方法，并用这种方法给资源定位。其一般格式为（其中主机位存放资源的主机）：
`<协议>://<主机>:<端口>/<路径>`

HTTP：万维网客户程序和服务器程序间进行交互使用的协议。协议本身是无连接的，但借助了TCP协议的80端口进行可靠传输。面向事务的客户服务器协议，HTTP1.0协议是无状态的。
工作流程：
浏览器分析超链指向页面的URL，向DNS请求解析域名IP地址；DNS解析出IP地址；浏览器与服务器建立TCP连接，发出取文件命令；服务器给出相应，把文件发送给浏览器；TCP释放连接；浏览器显示文件中的所有内容。
HTTP连接的两种方式：
非持续性连接（每个请求/响应都是经一个单独TCP连接发送）：`(n+1) * RRT`；
持续性连接（所有请求/响应都用相同TCP连接发送）：`2n * RRT`。

Email

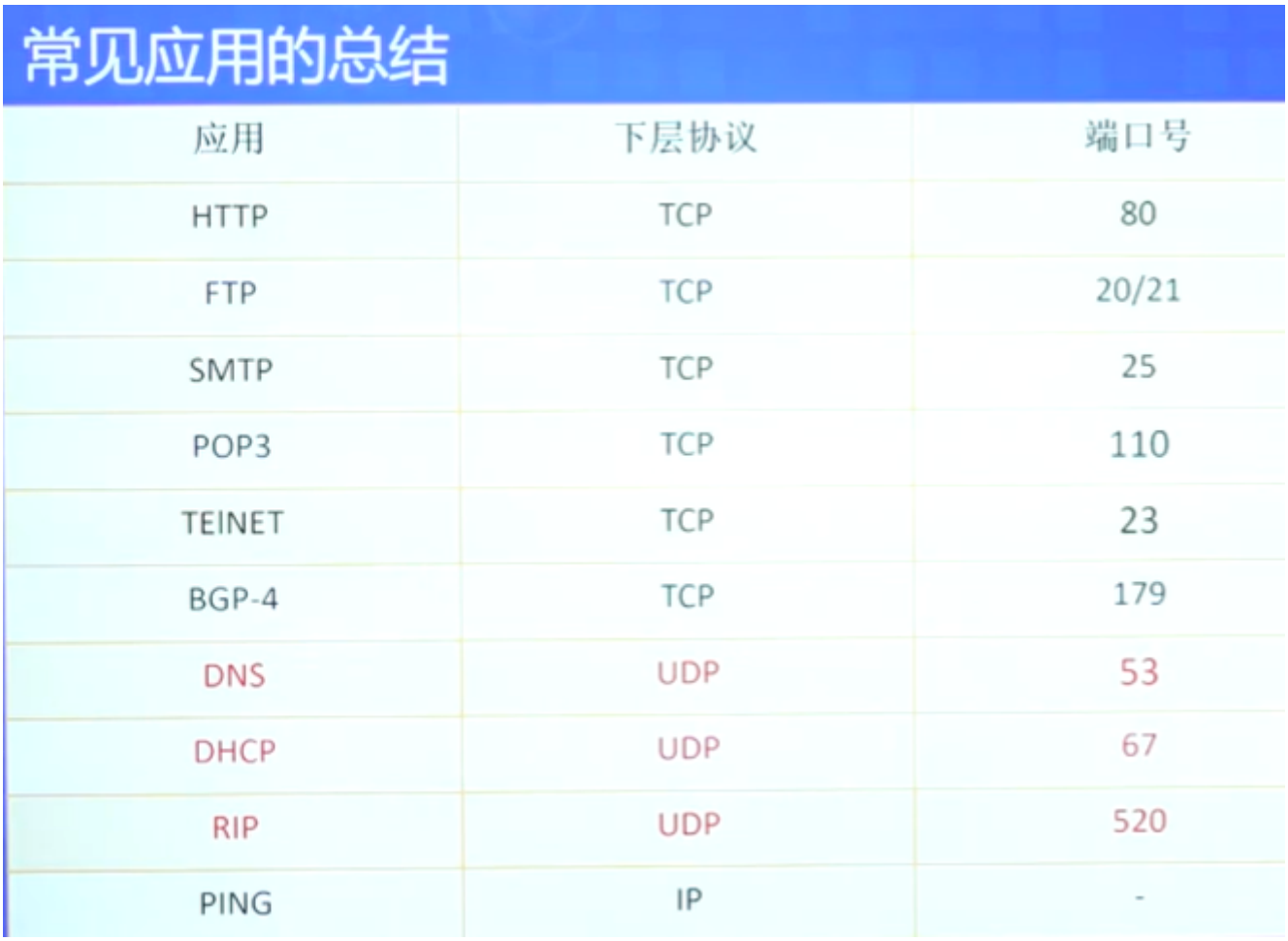
电子设备交换的邮件及其方法。包括用户代理，邮件服务器，以及邮件发送和读取协议。
电子邮件基于TCP连接，通过SMTP发送邮件。

相关Email协议：

1. **MIME**：用于电子邮件的标准，它扩展了原始的文本邮件格式，支持多种字符集和各种附件，使其传输和呈现多种类型的数据
2. **SMTP**：*发送*电子邮件的标准协议。使用客户端-服务器模型，邮件客户端通过与邮件服务器建立连接，将邮件发送到目标服务器，然后由目标服务器将邮件传递给接收方的邮件服务器。但SMTP不能传送可执行文件或其它二进制对象，限于传送7位的ASCII码，SMTP会拒绝超过一定长度的邮件
3. **POP3**：*接收*电子邮件的协议。将邮件从邮件服务器下载到本地设备，并在下载后将邮件从服务器上删除。常用于单个设备上的简单邮件访问，不支持多个设备同步邮件状态
4. **IMAP**：*接收*电子邮件的协议。与POP3不同，IMAP允许用户在邮件服务器上管理和组织邮件，不仅仅是将邮件下载到本地。支持在多个设备同步邮件状态

DHCP

互联网广泛使用的动态主机配置协议，提供了即插即用的联网机制，允许一台计算机加入新的网络和获取IP地址，而非手工配置。
需要IP地址的主机在启动时想DHCP服务器广播发送发现报文DHCPDISCOVER，主机称为DHCP客户，本地网络上所有主机都能收到报文，但只有DHCP服务器才能回答此报文。



常见应用的总结

应用	下层协议	端口号
HTTP	TCP	80
FTP	TCP	20/21
SMTP	TCP	25
POP3	TCP	110
TEINET	TCP	23
BGP-4	TCP	179
DNS	UDP	53
DHCP	UDP	67
RIP	UDP	520
PING	IP	-

补充

- DNS查询通常用UDP协议
- 采用迭代解析比递归解析更能减轻DNS服务器压力，也因此互联网上的计算机普遍采用此策略
- 在递归查询中，若本地域名服务器不知道被查询域名的IP地址，则会以DNS客户身份向其它根域名服务器发出请求报文
- 从FTP服务器下载文件时，FTP服务器对数据进行封装的五个转换步骤是：数据，数据段，数据包，数据帧，比特
- FTP中的数据连接会在整个会话期间保持打开状态
- WWW中的超文本标记语言HTML属于OSI协议体系中的表示层