目录

[第1章 引言 3](#_Toc529284356)

[1.1 使用计算机网络 3](#_Toc529284357)

[1.2 网络硬件 4](#_Toc529284358)

[1.3 网络软件 6](#_Toc529284359)

[1.4 参考模型 7](#_Toc529284360)

[1.5 网络实例 9](#_Toc529284361)

[第2章 物理层 10](#_Toc529284362)

[2.1 数据通信的理论基础 10](#_Toc529284363)

[2.2 引导性传输介质 10](#_Toc529284364)

[2.3 无线传输 11](#_Toc529284365)

[2.4 通信卫星 12](#_Toc529284366)

[2.5 数字调制与多路复用 13](#_Toc529284367)

[2.6 公共电话交换网络 15](#_Toc529284368)

[2.7 移动电话系统 18](#_Toc529284369)

[2.8 有线电视 20](#_Toc529284370)

[第3章 数据链路层 20](#_Toc529284371)

[3.1 数据链路层的设计问题 20](#_Toc529284372)

[3.2 差错检测和纠正 21](#_Toc529284373)

[3.3 基本数据链路层协议 22](#_Toc529284374)

[3.4 滑动窗口协议sliding window 23](#_Toc529284375)

[3.5 数据链路协议实例 24](#_Toc529284376)

[第4章 介质访问控制子层 24](#_Toc529284377)

[4.1 信道分配问题 24](#_Toc529284378)

[4.2 多路访问协议 25](#_Toc529284379)

[4.3 以太网（802.3） 26](#_Toc529284380)

[4.4 无线局域网 27](#_Toc529284381)

[4.5 宽带无线 29](#_Toc529284382)

[4.6 蓝牙Bluetooth 29](#_Toc529284383)

[4.7 RFID 30](#_Toc529284384)

[4.8 数据链路层交换 30](#_Toc529284385)

[第5章 网络层 31](#_Toc529284386)

[5.1 网络层的设计问题 31](#_Toc529284387)

[5.2 路由算法 32](#_Toc529284388)

[5.3 拥塞控制算法congestion 34](#_Toc529284389)

[5.4 服务质量 35](#_Toc529284390)

[5.5 网络互联internetwork 37](#_Toc529284391)

[5.6 Internet的网络层 37](#_Toc529284392)

[第6章 传输层 43](#_Toc529284393)

[6.1 传输服务 43](#_Toc529284394)

[6.2 传输协议的要素 44](#_Toc529284395)

[6.3 拥塞控制 45](#_Toc529284396)

[6.4 Internet传输协议：UDP 46](#_Toc529284397)

[6.5 Internet传输协议：TCP 47](#_Toc529284398)

[6.6 性能问题 49](#_Toc529284399)

[6.7 延迟容忍网络DTN，Delay-Tolerant Network 51](#_Toc529284400)

[第7章 应用层 51](#_Toc529284401)

[7.1 DNS——域名系统（Domain Name System） 51](#_Toc529284402)

[7.2 电子邮件 52](#_Toc529284403)

[7.3 万维网 53](#_Toc529284404)

[7.4 流式音视频 57](#_Toc529284405)

[7.5 内容分发 60](#_Toc529284406)

[第8章 网络安全 62](#_Toc529284407)

[8.1 密码学cryptology 62](#_Toc529284408)

[8.2 对称密钥算法symmetirc key algorithm 64](#_Toc529284409)

[8.3 公开密钥算法 65](#_Toc529284410)

[8.4 数字签名 66](#_Toc529284411)

[8.5 公钥的管理 66](#_Toc529284412)

[8.6 通信安全 67](#_Toc529284413)

[8.7 认证协议 69](#_Toc529284414)

[8.8 电子邮件安全性 69](#_Toc529284415)

[8.9 Web安全性 70](#_Toc529284416)

[8.10 社会问题 70](#_Toc529284417)

黑色页码 《计算机网络》(第5版) Andrew S. Tanenbaum, David J. Wetherall 清华大学出版社

# 第1章 引言

计算机网络（computer networks）：表示一组通过单一技术相互连接的自主计算机集合。如果两台计算机能够交换信息，则称这两台计算机是相互连接的（interconnected）。internet是计算机网络的一个实例。

分布式系统（distributed system）：由一组独立的计算机组成的关联系统，在操作系统之上有一层软件负责实现这个系统，称为中间件（middleware）。分布式系统和计算机网络存在区别，前者基于后者，区别在于有软件保持一组计算机之间的关联性。

## 1.1 使用计算机网络

1.1.1 商业应用

虚拟专用网络（VPN，Virtual Private Networks）

服务器（server）：性能强大的高级计算机，专用存储数据和网络信息

客户机（client）：也就是PC

客户机-服务器模型（client-server model）：最基本常见的网络应用模型

IP电话（IP telephony）：通过计算机网络打电话，如果采用internet技术则成为IP语音（VoIP，Voice over IP）

电子商务（e-commerce，electronic commerce）

1.1.2 家庭应用

连通性（connectivity）

对等（peer-to-peer）通信模式：没有固定的客户机和服务器分工的网络模型，对等系统没有任何中央数据库，BitTorrent是一个例子。

即时消息（instant messaging）

普适计算（ubiquitous computing）

射频识别（RFID，Radio Frequency Identification）：RFID标签是一种无源芯片（无电池），作用类似于条形码，但是比条形码储存更多信息，可以被RFID读写器在数米范围内定位和通信。是物联网（ITU）实现的一个设想中的基础。

1.1.3 移动用户

蜂窝网络：电话公司经营的手机无线网络（3G和4G）

无线热点（hotspot）：基于802.11标准

固定无线（fixed wireless）和移动无线（mobile wireless）：无线计算机也有非移动性的，比如无线电视等。

短消息（text messaging）和短信（texting）

智能手机（smart phone）

全球定位系统（Global Positioning System，GPS）和地理标记（geo-tagging）

移动商务（m-commerce，mobile-commerce）和近场通信（NFC，Near Field Communication）：手机可以充当RFID智能卡完成金融活动。

传感器网络（sensor network）：移动互联网络的前沿，传感器通过互联以及自动化分析处理可以实现更多功能与技术

可穿戴式计算机（wearable computer）

1.1.4 社会问题

网络公司就像电话公司或邮局，没有深刻的义务监管用户的言论。

短时玩家（small-time player）：只享用并支付少量服务的用户

网络中立（network neutrality）：通信不应该区分内容、来源，或者内容提供者的论点

僵尸网络（botnet）：受病毒攻击之后形成的

钓鱼（Phishing）：带有引诱受害者性质的非法活动

完全自动图灵测试（CAPTCHA，Completely Automated Public Turing test to tell Computers and Humans Apart）

## 1.2 网络硬件

计算机网络的两个重要维度：传输技术和网络尺度

数据包或包packet：宽泛而言，就是网络中传递的消息。很多时候，原本完整的信息受到一定传输条件的约束（如协议）而被分割成几部分。准确而言，这些才称为包。

线路测量速率：1Mbps=1000000比特/s；1Gbps=1000000000比特/s

传输技术：  
①广播式链路  
通信信道被网络上所有机器所共享，所有的机器都可以收到发送的信息，但是只有预期的接收方才能处理它，其他的机器则忽略不属于自己接收的信息。当然也可以通过设置数据包的地址编码，使其预期接收方为网络上所有机器（广播broadcasting）和一部分机器（组播multicasting）  
如无线网络  
②点到点链路point-to-point  
只有一个发送方和一个接收方，也称单播unicasting。

网络尺度：  
个域网，局域网，城域网，广域网，互联网

互联网络（internetwork）：两个或多个网络的连接。Internet只是一个互联网络的实例，同时也取了它的种类的名字。

星际互联网络（Interplanetary Internet）

1.2.1 个域网（PAN，Personal Area Network）

举例：  
•蓝牙（bluetooth）：一种无线网络，亦即一种个域网。使用主-从操作模式，即主设备负责决定一切和传输通信有关的细节。

1.2.2 局域网（LAN，Local Area Network）

局域网是一种私有网络。

举例：  
•企业网络enterprise network：一家公司内部的局域网

无线路由器wireless router：也称接入点Access Point，AP；或称基站base station。是中继无线计算机之间的数据包和计算机与Internet之间的数据包的设备。

Wifi：无线局域网的一个标准，也称为IEEE 802.11

有线局域网：使用铜线或光纤作为介质，因此性能全方面超过无线网，缺点是有线导致的不便。其拓扑结构是以点到点链路为基础的。  
如以太网（Ethernet）——IEEE 802.3，交换式以太网（switched Ethernet）——每台计算机按照以太网协议规定的方式运行，均通过点到点链路连接到一个交换机switch。

虚拟局域网（Virtual LAN，VLAN）：即在同一个物理局域网内建立的不同的“虚拟”形式的局域网，物理上仍为一体，功能上已经分离。

广播网络的静态设计和动态设计：（信道分配机制不同）  
•静态——将时间分成离散的时间间隔（称为时间槽time slot），每台机器只能在它分配的时间槽内广播，其他时间不行。而且如果没有广播需求，那么时间槽将被浪费。  
•动态——  
①集中式：有一个中心实体，决定信道的使用权和数据包发送的优先次序  
②分散式：无中心实体，每台机器自行决定（按照一定算法）是否可以传输

电线网络（power-line network）：一种网络概念，即通过电源线连接网络，或电源信号和数据信号同时通过电源线和插座传输。

1.2.3 城域网（MAN，Metropolitan Area Network）

覆盖一个城市范围的网络。

举例：  
有线电视网，WiMAX（IEEE 802.16）——高速无线Internet接入

自组织（ad hoc）系统：p18/32

头端（cable headend）：或称线缆前端，p18/32

1.2.4 广域网（WAN，Wide Area Network）

范围通常是一个国家、地区或者大陆。

广域网的组成：  
•主机host：或称用户机，专门运行用户应用程序的计算机或设备  
•通信子网communication subnet：或称子网，即连接所有主机的网络。包括以下：  
①传输线路transmission line——负责在机器之间移动比特  
②交换元素switching element——简称交换机switch、路由器router，负责连接两条或以上的传输线路，决定数据使用哪条线路

子网的传统含义：一组路由器和通信线路的集合。子网一般只在广域网的讨论范围内有意义  
更新含义——第5章

广域网和局域网的区别：  
•范围更大  
•网络的通信（子网）部分和应用（主机）部分分离，并由不同组织或团体运营  
•路由器通常连接不同类型的网络技术——意味着广域网和互联网络有交叉，或者称为复合网络  
•广域网的子网可以连接单个计算机也可以连接不同局域网

广域网的类型：  
①虚拟专用网络（VPN，Virtual Private Network）：没有专用的传输线路，使用底层Internet容量相互连接，即借用Internet连接创建的广域网。优点在于无需考虑使用专线的麻烦，缺点在于难于把握连接的质量和缺乏对其直接控制。  
②ISP网络（ISP network）：子网和传输线路由专门的公司（网络服务提供商network service provider）运营。这些公司多数会将其提供的网络与Internet连接，因此也被称为Internet服务提供商（ISP，Internet Service Provider）。  
③无线技术广域网：具体如卫星网络系统（通过无线电和天线传输，广播式）、蜂窝移动电话网络（通过无线电传输至基站，基站间有线传输，基站覆盖面积较无线LAN大，但数据传输速率非常小，1Mbps量级）

1.2.5 互联网络（internetwork）

定义：连接不同的网络（个域网、局域网、广域网等）的网络。不同的网络之间很可能使用不同的硬件和软件，甚至不相兼容，但是为了实现通信目的而连接在一起。

“网络”的定义：（标准、最初、传统定义——便于和其他相关事物定义区分）  
由一种单一技术相互连接在一起的计算机集合。

“互联网络”定义的判别：（由于存在定义争议而惯常使用的经验规则）  
•若不同组织出资构建了网络的不同部分，并各自维持其运营，则这个网络为互联网络  
•若网络的不同部分采用了不同的底层技术（广播、点到点、有线、无线等），则其为互联网络

网关gateway：将两个或多个网络连接起来并提供必要转换的硬件和软件的总称。路由器即为网关的一个实例。

## 1.3 网络软件

1.3.1 协议层次结构

（1）层次结构概览

绝大多数的网络结构都组织成一个层次栈（a stack of layer）或分级栈（a stack of level），每一层建立在下一层之上，每一层的特征和总层数等均视网络而异。每一层的目的均为向上一层提供特定的服务，并对上一层信息隐藏。

协议protocol：通信双方就如何进行通信的一种约定。机器与机器之间分别针对不同层次进行对应层次级别的会话，并遵循对应层次的协议。

对等体（peer）：不同机器上构成相应层次的实体，可能包括软件、硬件甚至人

通常意义上，称呼网络结构中最高几个层次为应用层，最低的几个层次为物理层，中间的为网络层。这只是一个宽泛概念，但是很常用。

（2）通信过程

数据和控制信息由顶层传递给其下一层，直至第1层。完成实际通信过程的是第1层之下的物理介质（physical medium）。当数据传递到目的地主机，又逐层向上进行反处理。因此，层与层之间的对等会话或对等进程是一种虚拟通信或概念。这种机制的目的在于方便设计不同的网络软件（并且符合软件的结构化设计），而仅需通过一种虚拟的同步来保证彼此能够通信。

相邻层之间通过接口来完成调用服务。合理的接口设计使得层与层之间处于相对独立的状态，因而可以改变某一层的协议或实现而不影响全局。另外，同一层次不同主机上的实现也是可以不同的，由不同的人编写，只需满足协议即可。

（3）网络体系结构（network architecture）

定义：即层和协议的集合。不包括每一层每个主机上的具体实现以及相邻层之间的接口，这些都是隐藏在本地的内容，不同主机之间这些内容可以完全不一样。

组成一个特定系统的一组协议称为协议栈（protocol stack）。

1.3.2 层次设计问题

（1）可靠性问题：  
①检错（error detection）和纠错（error correction）：检查接收到的信息是否正确，以及自动修复错误的信息  
②路由（routing）：找到通过网络的工作路径，包括避开故障点、选择最优路线等

（2）网络演进：  
•网络互联（internetworking）：随着网络的增长和演化、新设计的出现，旧网络需要和新技术连接或被改进重用，解决不同网络技术之间的差异造成的互联问题即为网络互联。网络规模变大时仍能工作良好的设计被称为可扩展的（scalable）  
•寻址（addressing）和命名（naming）：协议层中每一层在特定的消息中都需要有一种机制来标识发送方和接收方。这种机制在下层和高层分别称为寻址和命名

（3）资源分配  
•统计复用（statistical multiplexing）：根据主机的统计需求来动态共享带宽  
•流量控制（flow control）：从接收方到发送方的反馈机制，保证慢速接收方不会被快速发送方的信息淹没  
•拥塞（congestion）：主机发送总信息量过大，传输带宽不够  
•服务质量（quality of service）：有很多应用不单单只对网络带宽有要求，实时性（real-time）与延迟、传输时间、抖动、吞吐量都是对数据传输更高的需求。

（4）网络安全  
•保密性（confidentiality）  
•认证（authentication）  
•完整性（integrity）

1.3.3 面向连接与无连接服务

下层可以向上层提供两种不同类型的服务：

（1）面向连接的服务（connection-oriented service）  
•特征：按照电话系统建模，先建立发送与目的之间的连接，再进行数据传输，最后释放连接。有以下两种变异形式  
①报文序列：发送过程中，连续发送的信息的边界始终得到保持，比如传输文档  
②字节流：没有任何边界信息，比如传输视频  
•特点：信息到达顺序与发送的顺序相同  
•应用：文件传输

（2）无连接服务（connectionless service）  
•特征：按照邮政系统建模，没有确定的一对一连接，数据携带有目标地址，由路由器负责分配转发。有以下两种进一步分类  
①存储-转发交换（store-and-forward switching）：受到信息的全部内容之后再转发至下  
②直通式交换（cut-through switching）：没有完全接收完毕之前就向下一节点传输  
•特点：信息到达顺序不一定和发送顺序相同

服务的可靠性可以由反馈确认机制实现，但是会耗费额外的性能，视具体应用情况而定是否需要可靠性确认。  
•数据报服务（datagram service）：没有确认机制的无连接服务，是大多数网络中占主导地位的传输形式  
•有确认的数据报服务（acknowledged datagram service）  
•请求-应答服务（request-reply service）：发送方发送的是需要作出处理和回复的请求

注意：以上的种类划分不是绝对的，为了满足需要，实际情况下可能会出现各种复合的设计。

1.3.4 服务原语

原语（primitive）：说明一个服务的动作或功能的一组关键词。原语只是一种概括性描述，具体执行细节并不由其影响

例子：p29/43

1.3.5 服务与协议的关系

服务：某一层（服务提供者）向它上一层（服务用户）提供的一组原语（操作）。服务与两层之间的接口有关，但是并不涉及如何实现的细节

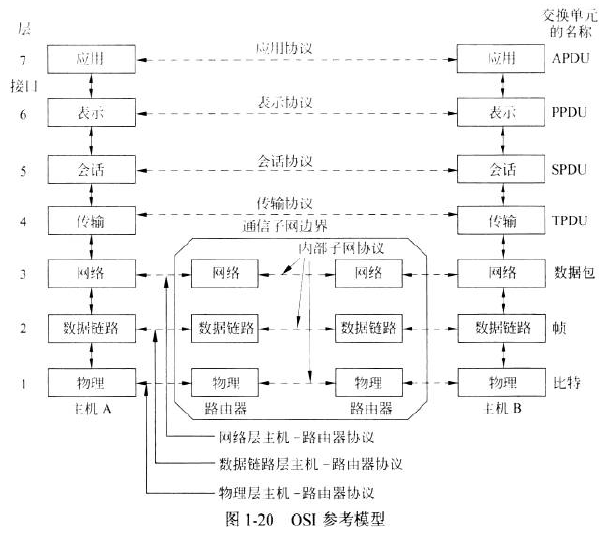
协议：一组规则，规定了同一层上对等实体之间所交换的数据包的格式和含义。协议决定了如何实现服务，但是不应该影响服务的接口

## 1.4 参考模型

两个重要的网络体系结构：

1.4.1 OSI参考模型

开放系统互连参考模型（OSI，Open Systems Interconnection）是一种供设计参考使用的、涉及如何连接开放系统的准网络体系结构模型。严格说来OSI模型不是网络体系结构模型，因为它并没有定义每层的服务和协议，只是指明了每一层应该做什么事。不过ISO已经为每一层制定了相应的标准，但是以独立的标准发布，而且这些协议应用得并不广泛。



OSI模型的基本原则：5条，p33/47

（1）物理层（physical layer）

关注在一条通信信道上传输原始比特。主要涉及机械、电子和时序接口，物理传输介质等。

（2）数据链路层（data link layer）

将一个原始的传输设施转变成一条没有漏检传输错误的线路。主要涉及流量调节，控制共享信道访问等。

数据帧（data frame）：发送方将数据拆分之后的产物

确认帧（acknowledgement frame）：接收方发给发送方确认正确收到数据的信息

（3）网络层（network layer）

控制子网的运行，处理多条链路之间的结合以及网络之间的互联。主要涉及如何将数据包从源端路由到接收方，处理拥塞，处理网络互连，提高服务质量等。

广播式网络中，网络层功能和意义不大。

（4）传输层（transport layer）

基本功能是接收来自上一层的数据，必要时对数据进行分割，然后传给网络层，并确保数据单元正确抵达。主要涉及决定向会话层甚至最终网络用户提供哪种类型的服务，如点到点、广播式等。

从传输层开始向上的层次都是端到端的层（协议一致），其下的层次需要经历源主机、中间路由器等中间传输步骤，每次传输都可能使用不同的协议。

（5）会话层（session layer）

允许不同机器上的用户建立会话。会话层提供各种服务，包括对话控制（dialog control）——记录该由谁来传递数据、令牌管理（token management）——禁止双方同时执行同一个关键操作、同步功能（synchronization）——在一个长传输中设置断点，支持断点续传

（6）表示层（presentation layer）

关注的是传递信息的语法语义，实际上是保证不同计算机上的数据结构和编码标准能够通畅交流。

（7）应用层（application layer）

应用层包含用户通常需要的各种协议，以及使用网络的应用程序。

广泛的应用层协议是HTTP，规定了如何访问和传输网络页面信息的准则。

1.4.2 TCP/IP参考模型（Reference Model）

ARPANET：所有广域计算机网络的鼻祖，Internet是其继任者。

（1）链路层（link layer）

不是真正意义上的一个层，是主机和传输线路之间的一个接口。描述链路必须完成什么功能才能满足无连接的互联网层的需求。

（2）互联网层（internet layer）

任务是允许主机将数据包注入到任何网络，并且让这些数据包独立地到达可能位于不同网络的接收方。

IP协议（Internet Protocol）：因特网协议，互联网层定义的官方的数据包格式和协议。具体见后文

ICMP协议（Internet Control Message Protocol）：因特网控制报文协议，一个辅助协议。

（3）传输层

任务是允许源主机和目标主机上的对等实体进行对话。

TCP协议（Transport Control Protocol）：传输控制协议，可靠的面向连接的协议，允许字节流准确无误传输到目的地。涉及分割传输信息和流量控制。

UDP协议（User Datagram Protocol）：用户数据报协议，不可靠的无连接协议，适用于不需要TCP协议的情况，或应用程序可以自行提供类似功能。

（4）应用层

应用层简单包含了所需的会话和表示功能，因此无需特别设置这两层。

应用层包含的高层协议种类：  
•虚拟终端协议（TELNET）  
•文件传输协议（FTP）  
•电子邮件协议（SMTP）  
•域名系统（DNS，Domain Name System）：将主机名字映射到其网络地址  
•HTTP：获取万维网页面  
•RTP：传送语音或电影等实时媒体

1.4.3 本书使用的模型

1.4.4 OSI参考模型与TCP/IP参考模型的比较

传输层及其以下层组成了传输服务提供者，之上各层为传输服务的用户。

OSI模型的核心：服务、接口、协议。服务定义的是某一层的语义和功能，不包括接口，也不包括实现细节。

TCP/IP模型只是已有协议的一个描述，亦即先有协议再有模型，模型本身不具有普适性，但是和自身的协议契合非常好。OSI模型则正好相反，具有普适性但也缺少特别性，比如模型设计之初根本没有考虑网络互联问题。

OSI网络层支持无连接和面向连接通信，传输层只支持面向连接通信。TCP/IP网络层支持无连接通信，传输层支持两种通信模式。

1.4.5 OSI模型和协议的评判（评价）p39/53

关于制定标准的时机的理论——两头大象

1.4.6 TCP/IP参考模型的评判（评价）p41/55

## 1.5 网络实例

1.5.1 因特网

①ARPANET

ARPA，Advanced Research Projects Agency：高级研究计划局

NPL，National Physical Laboratory：英国国家物理实验室

IMP，Interface Message Processors：接口报文处理器

ARPANET是因特网的雏形，最初由几所大学和机构互联形成，其他更多的局域网络逐渐加入便增加了规模。

②NSFNET

NSF，U.S. National Science Foundation：美国国家科学基金会

NSFNET最初由科学基金会自行创建，目的是提供ARPANET不能提供的社会性学术性需求，甚至继任ARPANET，后来被ANS接管，改称ANSNET

ANS，Advanced Networks and Services：高级网络与服务公司

网络接入点NAP，Network Access Point

Internet系统协会Internet Systems Consortium

③Internet的体系结构

三网融合triple play：电话，电视，互联网三网服务融合

内联网intranet：企业或机构内部将其内部的多个网络互连在一起形成的小型独立“互联网”

1.5.2 第三代移动电话网络

UMTS 3G移动电话网络体系结构p52/66

无线网络控制器RNC，Radio Network Controller：控制基站如何使用频谱

核心网络core network：移动电话系统的内部网络

移动交换中心MSC，Mobile Switching Center

网关移动交换中心GMSC，Gateway Mobile Switching Center

媒体网关MGW，Media Gateway

通用数据包无线业务GPRS，General Packet Radio Service

服务GPRS支撑节点SGSN，Serving GPRS Support Node

网关GPRS支撑节点GGSN，Gateway GPRS Support Node

归属用户服务器HSS，Home Subscriber Server：记录用户位置的服务器

1.5.3 无线局域网：802.11

略

1.5.4 RFID和传感器网络

无源RFID，passive RFID：无电池

有源RFID，active RFID：标签上有电池

超高频RFID，UHF RFID：货运托盘，驾照

高频RFID，HF RFID：护照，信用卡，书籍

低频RFID，LF：动物跟踪

多跳网络multihop network：各个网络节点自组织地中继各自的消息

## 1.6 网络标准化

1.6.1 电信领域有影响力的组织

国际电信联盟ITU，International Telecommunication Union

ITU-T：电信标准化部门

ITU-R：无线电通信部门

ITU-D：发展部门

1.6.2 国际标准领域有影响力的组织

国际标准化组织ISO，International Standards Organization

技术委员会TC，Technical Committee

联合技术委员会JTC，Joint Technical Committee

子委员会SC，subcommittee

工作组WG，working group

委员会草案CD，Committee Draft

国际标准草案DIS，Draft International Standard

国际标准IS，International Standard

国家标准和技术协会NIST，National Institute of Standards and Technology

国家标准局National Bureau of Standards

电气和电子工程师协会IEEE，Institute of Electrical and Electronics Engineers

1.6.3 Internet标准领域有影响力的组织

Internet活动委员会IAB，Internet Activities Board

Internet体系结构委员会IAB，Internet Architecture Board：更名之后的名称

请求注释RFC，Request For Comments：Internet技术报告

Internet研究任务组IRTF，Internet Research Task Force：IAB附属机构

Internet工程任务组IETF，Internet Engineering Task Force：IAB附属机构

Internet协会Internet Society

## 1.7 度量单位

一般在计算机领域为了避免歧义，KB、MB、GB等容量单位是2的幂指数，即KB是1024字节之意。而kbps、Mbps等码率单位，是10的幂指数，即kbps是1000位每秒

# 第2章 物理层

## 2.1 数据通信的理论基础

2.1.1 傅里叶分析

傅里叶级数Fourier series：任何一个行为合理周期为T的函数g(t)，都可以表示成用正弦和余弦函数组成的无穷级数。具体公式见p70/84

信号分析中将函数g(t)作为信号，组成级数的各个单独的三角函数称作谐波harmonics。

2.1.2 带宽有限的信号

导线传输衰减性质：从0到某个频率fc的范围内，振幅在传输过程中不会衰减，此频率以上会发生不同程度减弱，fc是为截止频率。因此实际情况下，传输数字信号或脉冲到接收端并无法保留原有的信号形状。有时人为约定的带宽限制也会注定导致这种情况——高频段的谐波衰减较大乃至消失。

带宽bandwidth：传输过程中振幅不会明显减弱的频率宽度。实际中，以从0到使得接受能量保留一半的频率为准。带宽取决于导线介质构成、厚度和长度等尺寸。有时候人为约定的用于传输数据的某个频率宽度也称为带宽。

滤波器一般可用来限制信号的带宽，某个信号只需一定频率宽度的若干谐波即可大致表示（用于数字信号传递已经足够）。滤波的意义在于只保留能够识别原信号的最小频率宽度，如此多个信号（信道）可以占用不同频率范围同时传输。另外，信号传输的谐波频率不一定非要从0开始，只要整个频率宽度足够即可。

基带baseband：从0到某个足够传输一个信号的谐波频率范围

通带passband：基带以上的某个更大的频率范围（用以传输另一个信号）。

数据传输速率（比特率）和带宽的关系：设比特率为b bps，发送一个字节（8bit）的时间为8/b秒。信号的第一个谐波的周期即信号本身的持续时间长度，可设信号有A个字节，即周期为8A/b，频率为b/(8A)。设带宽为B，则估计带宽可容纳谐波数为8BA/b。可见比特率越大，带宽可容纳谐波数越少，传输信号的质量也就越差。

模拟带宽和数字带宽：模拟带宽使用Hz来度量，是电气领域里的术语，数字带宽以bps来度量，是计算机网络领域术语。两者有密不可分的联系。

2.1.3 信道的最大数据速率

信号的离散等级：根据传输信号使用的不同种信号元素的数量划分，即信号采用的电压值等级数量。例如二进制信号，只有两级。

数字中继器：RP或Repeater，通过放大重生或重新发送信号以实现扩大信号传输距离的目的的设备。

信噪比SNR，Signal-to-Noise Ratio：信号功率S与噪声功率N的比值，写作S/N。但经常使用10log10S/N的表达形式来适用较大范围的信噪比书写，其单位为分贝dB，decibel（deci实际表示表达式中的10倍常数，bel才是单位的本体）。

尼奎斯特Nyquist定理：无噪声信道的最大数据速率 p73/87

香农容量Shannon capacity：实际信道（有噪声）能获得的最大比特率p74/88

## 2.2 引导性传输介质

引导性介质也称有线介质，同理非引导性意味着无线。

2.2.1 磁介质

即使用磁带或磁盘进行数据的输入输出，传输使用直接运送的方式（如快递等）。因为磁介质的容量极大，这种最原始的数据传输反倒有着极高的数字带宽和极低的传输成本。

缺点是传输过慢，即延迟特性太差，不适合在线需求。

2.2.2 双绞线（twisted pair）

双绞线使用两根相互绝缘的铜线以螺旋形式绞在一起。这样可以降低两根铜线之间的干扰，同时信号以两根线的电压差传递，可以降低噪声干扰影响（噪声对两根线影响一致）。

5类线（Category 5，Cat 5）：包含4对双绞线于一个保护套内的电线类型。

全双工（full-duplex）链路：双向可同时使用

半双工链路（half-duplex）：可双向使用，但一次只能使用一个方向

单工链路（simplex）：只允许一个方向上传输的链路

非屏蔽双绞线（UTP，Unshielded Twisted Pair）：6类（含6类）之前的双绞线的特点。7类开始在每对双绞线外加屏蔽层，整个线缆外再加屏蔽层（不同于绝缘层），是为屏蔽双绞线。

2.2.3 同轴电缆（coaxial cable）

比UTP具有更好的屏蔽特性和带宽。物理结构见p76/90

常用于有线电视和计算机城域网。

2.2.4 电力线

即使用用于电力传输的电线来传输数据信号。

2.2.5 光纤

光纤传输的带宽极限受制于电气信号和光学信号之间的转换速率。但可以通过多信道共享单条光纤的方式改进。

光纤到户（FttH，Fiber to the Home）：

光纤传输系统包括光源、传输介质和探测器。探测器用以转换光电信号。

多模光纤（multimode fiber）：光纤内部可容纳许多光束以不同的角度来回全反射向前传播

单模光纤（single-mode fiber）：光纤直径小到只有几个光波波长的时候，光纤如同波导，只能容纳一束按“直线”传播的光。

光的衰减定义为输入输出信号功率的比值，同样使用分贝度量。一个特殊值为3分贝表示衰减一半的能量。光在光纤中衰减程度取决于光的波长，p79/93。

光脉冲在光纤中会发生色散传播（chromatic dispersion）。通过将脉冲做成一种特殊的形状，可以消除色散效应，称作孤波soliton。

光缆的组成结构P80/94

光纤拼接方式：用连接器（损失10%-20%的光）、机械手段直接拼接（10%）、熔合手段

信号源的类型：LED（light emitting diodes）和半导体激光。二者比较p81/95  
在光源和光纤之间插入干涉仪（Fabry-Perot或Mach-Zehnder）对波长进行调节。

光纤和铜线的比较：绝大多数方面优于铜线p81/95

## 2.3 无线传输

2.3.1 电磁频谱

波段：用于数据传输的电磁波频率范围分类——低频（LF）、中频（MF）、高频（HF）、甚高频（VHF，Very）、超高频（UHF，Ultra）、特高频（SHF，Super）、极高频（EHF，Extremely）、巨高频（THF，Tremendously）。

宽带与窄带频段：大多数信息传输都使用相对窄的频段（Δf/f<<1），是为窄带频段。宽带频段同理即使用较宽的频段，这种技术也称为扩展频谱spread spectrum。分为三种方式，见下。

跳频扩频（frequency hopping spread spectrum）：以跳频的方式扩展频段。跳频即从一个频率跳到另一个频率，一般每秒几百次。用以避开干扰。

直接序列扩频（direct sequence spread spectrum）：CDMA

超宽带通信（UWB，Ultra-WideBand）：至少有500MHz带宽或中心频率至少占频率波段20%的信号。UWB使用特别宽的频段，抗干扰能力强，干扰其他信号能力弱，作为其他一般信号的衬底underlay。

2.3.2 无线电传输（RF，Radio Frequency）

无线电传输的发射和接收设备无需对齐，发射器向各个方向传播信号。电波本身也具备绕过障碍物的能力，可以在地面传输比较广的距离，但信号源之间容易彼此干扰。

路径损耗（path loss）：无线电波的能量在传播时至少以距信号源距离的平方的反比速度衰减。

无线电衰减规律和引导性介质衰减规律不同。后者信号在单位距离下降的能量比例相同，前者遵循平方反比规律。

甚低频、低频、中频无线电波沿着地面传播，一般用振幅调制AM，用于广播电台。高频、甚高频电波会被地球表面吸收，因此传播方式主要依靠地球电离层折射反弹，一般用频率调制FM，用于长距离通信。

2.3.3 微波传输

微波按照直线传播，微波传播的中继器之间的距离大致与塔高的平方根成正比。

多径衰落multipath fading：微波传输尽管集中，但仍有发散与折射，造成有延迟微波信号的存在，并且和原信号抵消。这个现象和天气与频率有关，可将一部分信道作为备用信道平时留用，其他信道多径衰落时起用备用信道。

4GHz左右的微波频段会被水分吸收，也就是微波炉的频率。在信号传输时只能避开该频段。

路权right of way：铺设线缆用于通信时，显然必须花费一些成本购买铺设线缆所占用的路径空间。

ISM频段：Industrial Scientific Medical，是一些被政府保留用于免费非许可性无线电应用的频段（不一定只有一个频段）。包括车库门遥控器（不包括电视遥控器）、遥控玩具、各种无线家用设备、蓝牙、Wifi等都使用ISM，但并没有具体划分频段分配。为了避免相互干扰，所有ISM发射器的功率都很低。

U-NII，Unlicensed National Information Infrastructure：非许可的国家信息基础设施

FCC，Federal Communication Commission：联邦通信委员会，美国负责频谱分配的机构

蓝牙和802.11b/g使用2.4GHz的ISM频段，会受到微波炉和雷达装置的干扰。

无线高速网络的频段是57-64GHz，也是被开放给民用或家用设备的频段。原因在于60GHz的频率会被氧气吸收，因而本质上无法传播太远。

2.3.4 红外传输

由于无法穿透固体，红外传输用于短程通信，也因此不需要政府许可经营。例如电视或类似设备的遥控器。

红外数据协会IrDA，Infrared Data Association

2.3.5 光通信

非引导性光通信依赖激光传递信号，但是要求传输和接收设备具备极高的对齐度，可是激光却会受到环境、气流、湿度等影响（太空中不是问题）。因此，实际中激光通信很难实现远距离通信，但是普通的可见光信号的编码和信息传递却可以在小范围内产生应用。

## 2.4 通信卫星

简单理解，通信卫星就是一个大型微波中继器。

工作原理：卫星包含几个转发器transponder，每个转发器侦听频谱中的某一部分，对入境信号进行放大，再在另一个频率上将放大信号重新广播。这种方式成为弯管模式bent pipe，为了避免出入信号相互干扰。弯管模式是直通式的，而存储-转发式可以滤去接收信号的噪声。

范艾伦辐射带Van Allen belts：受地球磁场影响的一些高带电粒子层，通信卫星不能在这里工作，甚至毁坏。

2.4.1 地球同步卫星

即GEO，Geostationary Earth Orbit卫星。

第一颗人造通信卫星叫Telstar。太空中同时最多能放置180颗同步卫星，小于2度的放置角度会使卫星之间干扰。

轨道控制station keeping：为了保持卫星在预定轨道上运行而进行的微调活动，需要使用火箭发动机。因此卫星运行大约10年后，便注定会坠落。

卫星的转发空间波束称为其足迹footprint，亦即其发送信号的覆盖范围。可以调整为很大的区域覆盖大面积地球表面，也可以调整为很小覆盖区域，称为点波束spot beam。

VSAT，Very Small Aperture Terminals，小孔径终端：一种低成本低功耗微型发射或接收站。如果是和卫星之间的单向传输，其功率足够，但不足够让VSAT之间相互通信。可以设置中继站hub来放大VSAT的信号，发送端通过卫星首先发送信号给中继站，然后中继站将信号放大后重新发送给卫星，进而发送给最终接收端。这种技术的特点是，用较长的延迟（一般卫星通讯来回的延迟为270ms，使用中继站一般为540ms）换取廉价的终端成本。使用地面中继站而不是卫星的原因是，地面中继站的能量供给充足，可以放大信号至很高的程度，而卫星显然受到各种条件限制，放大能力虽然具备但较弱。

卫星的通信特点是广播性质，因此需要广播功能时，其成本很低。而且传输成本和端到端的距离关系不大（实际上的传输距离都是端到卫星和卫星到端）。但其安全性不足，传输保密信息时必须加密。另外，卫星通讯错误率极低（和信号传输经过的路径有关，较少受到地形等因素阻碍），部署效率极高（卫星一直都在，只需布置收发端，这在军事活动或灾难时很重要）。

2.4.2 中地球轨道卫星

即MEO，Medium Earth Orbit卫星。大约6小时绕地一圈。普遍用于导航系统，而非通信。GPS，Global Positioning System使用的就是这种卫星。

2.4.3 低地球轨道卫星

即LEO，Low Earth Orbit卫星。距离近，运动速度快，延迟低。

铱星计划Iridium和全球星计划Globalstar是两个LEO通信系统的名称。前者由66颗LEO卫星组成，后者48颗。二者使用不同的信号中继方式，前者几乎完全依靠卫星中继信号，后者使用地面中继站中继远程信号（但仍然使用卫星以达到全球覆盖能力）。

2.4.4 卫星与光纤

ADSL，Asymmetric Digital Subscriber Line：非对称数字用户线，一种由电话公司提供的高带宽数据传输或联网服务用线。使用的是引导性介质，如双绞线等。

## 2.5 数字调制与多路复用

数字调制digital modulation：比特与代表它们的信号之间的转换过程

基带传输和通带传输是两种将数据比特直接转换成信号的方案。前者适合除光纤外的一般有线介质，因为这些介质的所有带宽一般都被唯一一个组织独占；后者适合无线和光纤传输，显然不同频率的带宽是固定分配的。

多路复用技术multiplexing：一个信道被多个信号共享，包括时分复用、频分复用、码分复用的方式。

2.5.1 基带传输baseband transmission

不归零NRZ，Non-Return-to-Zero编码方案：用正电压表示1，负电压表示0，即最直观的编码方式。

线路编码line codes：一些复杂的能够更加满足需求的信号编码方案

①带宽效率因素

符号symbol：用以携带多个比特信息的信号，比如用一种信号表示2个连续的比特信息，那么这种符号就会有4个不同级别。此时信号改变的速率称为符号率symbol rate，早期称作波特率baud rate，以区分比特率bit rate。

信号级别数不一定必须是2的幂。

②时钟恢复

时钟恢复简言之就是对接收器的采样周期进行频繁校对，保证其和入境信号时时同步。由于信号的速度极快，让接收器精准地采样很重要。但是不能依靠常规意义上的时钟，因为精度不可能达到信号的周期量级。所以才使用时钟信号来解决。但将时钟信号单独用线缆传输只适合计算机或短电缆环境下，网络链路中成本仍然太高。不过时钟信号也非唯一的解法，无论它是否占用单独的线路，只要信号的变化频繁，信号本身就可以起到时钟信号的效果。

曼彻斯特Manchester编码：使用两倍于数据信号比特率的时钟信号，将两种信号混合（异或）在一起。0混合结果为时钟信号从低到高跳变，1混合结果为从高至低转变。主要应用于经典以太网，缺点是需求带宽增大。

不归零逆转NRZI，Non-Return-to-Zero Inverted编码方案：1定义为信号有跳变（可升可降），0定义为信号无跳变。虽然比NRZ要强，但连续0仍然会导致信号不变。

通用串行总线USB，Universal Serial Bus：用来连接计算机外设的常用标准线，使用NRZI编码模式

T1 lines：T1线路，美国最早的数字电话线

4B/5B编码模式：将每4个比特按照一定规则映射成一个5比特模式，所有的5B信号只采用信号连续变化的模式（32种中选16种和原信号对应），增加25%带宽开销。除了16种对应信号外，5B信号还有一些连续变化的组合可用来表达线路控制信息。

扰频/倒频scrambling编码方式：发送数据之前扰频器scrambler用一个伪随机序列异或该数据，形成近乎连续变化的信号，接收端再用相同的伪随机序列异或解码数据。不增加额外带宽，能调节信号将信号的能量分散在整个频率成分上，降低了电磁辐射干扰。但仍然有几率产生不变的信号，或被恶意攻击产生该结果。

③平衡信号balanced signals

定义：很短时间内正电压和负电压一样多的信号，即均值为零，无直流电气分量。

平衡的信号即交流信号从物理性质的角度适合传输，因为线路中的变压器、接收器端连接到信道上的电容耦合capacitive coupling都强烈阻碍直流成分。平衡信号由于存在正负电压的混合，所以也有助于时钟恢复；另外平衡信号的平均值刚好可以作为解码符号的决策阈值，可用于校准接收器。

双极编码bipolar encoding或交替标记逆转AMI，Alternate Mark Inversion：后一种是电话网络中旧术语。使用两个电压级别表示1，用0V表示0。发射器根据平衡性自动选择用+1V或-1V表示1。但是如果存在连续的0，可能会导致时钟问题。

8B/10B线性编码模式：类似于4B/5B的映射方法。8B原始信号中分成5B和3B两组。5B映射成6B，3B成4B。映射的原则是尽量保证所有的输出模式都是平衡的，但实际上平衡的输出模式不足够映射。因此使用一个输入对应两个输出的模式，两个输出模式彼此取反。编码器根据当前累计的不均等性disparity，从两个输出模式中选择能够趋于整体平衡的输出模式。这种编码模式不仅保证平衡，也能保证时钟恢复，因为不会出现超过5个连续的1或0。

2.5.2 通带传输passband transmission

天线的大小与信号的波长成比例，低频信号需要大天线。

数字调制即调制或调节信号，通过调整信号的振幅、频率或相位来达到目的。对应的术语为幅移键控ASK，Amplitude Shift Keying；频移键控FSK，Frequency Shift Keying；相移键控PSK，Phase Shift Keying。其中PSK包括二进制相移键控BPSK，Binary（0和180两种相位）和正交相移键控QPSK，Quadrature（45,135,225,315四种相位）。

调制技术可以混合使用，但是频率和相位无法同时调制。

星座图constellation diagram：用来表示调制方式的图像p103/117。星座图之所以呈现方形样貌，是因为产生此类信号的硬件相较于同心圆形样貌的容易制造。

正交调幅QAM，Quadrature Amplitude Modulation：一种振幅和相位同时调制的方案。包含QAM-16和QAM-64。QAM-16包括16种不同的振幅相位组合，即16种信号，可用于每个符号传输4bit。QAM-64类似道理。

格雷码Gray code：一种对调制信号进行编码或分配比特表示的方案。在星座图中，格雷码的分配方案保证相邻的符号只有1bit的不同。因此如果信号解码出错，每个符号一般只会出错1bit。

2.5.3 频分复用FDM，Frequency Division Multiplexing

定义：将频谱分成几个频段，每个用户完全拥有其中的一个频段来发送自己的信号。具体过程可能包括将不同的用户原本同频段的信号频率上调至不同层次再混合。P104/118

保护带guard band：FDM中用于隔离不同信道的小空白频段区间。然而随着技术发展保护带不是必须的成分。而且保护带也无法完全保证能够隔离信道，仍然会存在边界处信号覆盖的非热噪声。

正交频分复用OFDM，Orthogonal：无保护带的FDM，相邻子载波彼此覆盖包裹，但是在每个子载波的中心频率上没有其他波干扰，可以用于采样。但是这种方案需要增加保护时间来及时重复符号信号的一部分，以便获得所需要的频率响应。不过这种开销小于保护带的开销。

OFDM已经广泛用于802.11、有线电视网络和电力线网络，以及第四代蜂窝系统。

信道退化degraded：在子载波一级更加易于应付

2.5.4 时分复用TDM，Time Division Multiplexing

定义：用户以循环的方式轮流发送数据，每个用户周期性（时间槽time slot）获得整个带宽。为了适应时钟，需要增加保护时间guard time间隔。

TDM广泛用于电话网络和蜂窝网络。

统计时分复用STDM，Statistical：即包交换的一个代名词。根据需求产生复用调度方案，而不是固定的。

2.5.5 码分复用CDM，Code Division Multiplexing

是扩频的一种形式，将窄带信号扩展到宽带上。更能容忍干扰同时允许多信号共享同频段。也称作码分多址CDMA，Code Division Multiple Access

码片chip：CDMA中，每个比特时间被再细分为m个更短的时间间隔，这些间隔称作码片

码片序列chip sequence：每个用户或发送站分配得到的一个专属m位码。用户只能通过发送码片序列或其反码来表达比特1或0，不同用户之间的码片序列彼此正交orthogonal（若将序列视为向量，则内积为零）。

Walsh码Walsh Code：用于产生正交码片序列

CDMA的基本工作原理p106/120，只是一个核心概念，实际应用中有大量复杂因素需要深入。

## 2.6 公共电话交换网络

公共交换电话网络PSTN，Public Switched Telephone Network

2.6.1 电话系统结构

电话系统包含3个主要成分：交换局（交换局有多个等级，每个等级之间的关系类似最低等级和客户之间的关系）、客户与交换局之间的线路（双绞线）、交换局之间的长距离线路

端局end office：也称为本地中心局local central office，最低级的交换局（1-10km覆盖）。客户机和端局之间的双线连接称为本地回路local loop，或“最后一英里”。

汇接局tandem office：同一个本地地区的交换中心，等级比端局高的交换局

长途局toll office：比汇接局更广的交换中心

长途连接中继线toll connection trunk：端局的出境线路名称

长途中继线intertoll trunk或interoffice trunk：长途局和更高层的交换局之间的连线

早期的电话系统使用分层路由来找到一条路径，收发方同时沿线追踪，直到同时到达同一个某层的交换局来建立连接。级别越高的交换局的出境线路所需带宽越大。

现在的电话系统，除了本地回路仍然使用模拟信号传输外，其他部分均使用数字传输。因为数字传输只要能够区分比特1或0即可，比模拟信号传输更加可靠。

2.6.2 电话政治化

AT&T公司：美国老牌电话公司

BOC，Bell Operating Company：贝尔运行公司，RBOC是区域性BOC

LATA，Local Access and Transport Areas：本地接入和传输区域，简单而言就是一个地区码覆盖的小块区域范围，即每个地区本地通话的范围。

LEC，Local Exchange Carrier：本地交换运营商，负责LATA范围内的本地通话的公司

IXC，IntereXchange Carrier：跨区运营商，负责不同LATA之间的长途通话业务的公司

POP，Point of Presence：IXC为了在某个LATA中展开业务而设置的代理点或称ISP存点

2.6.3 本地回路：调制解调器、ADSL和光纤

（1）电话调制解调器

调制解调器modem是调制器modulator和解调器demodulator的缩写。包括电话modem，DSL modem，有线电视modem和无线modem。指任何一种在数字比特和模拟信号之间作转换的设备。

宽带broadband：传统拨号上网速率被限制在56kbps，这是电话线的限制，以高于拨号速率接入网络就称为宽带。

网格编码调制TCM，Trellis Coded Modulation：为了减少由于符号和星座点数量增加导致的更高的错误率（噪声对信号接收影响更大），高速调制解调器标准采用一些额外符号进行差错检测。如每个符号传输4个数据比特和1个校验比特（V.32调制解调器）

电话modem的传输极限是56kbps或64kbps（取决于是否使用一个校验比特）。这个极限是由电话信道被滤波器限制为4kHz宽导致的。最终的结果是V.90和V.92调制解调标准。

（2）数字用户线DSL，Digital Subscriber Line

DSL：一种由电话公司提供的相对于电话调制解调器的新的数字服务。本地回路仍然使用电话线路，但是端局入境线路使用不同于原有电话线路的交换机（不存在将信号控制在语音信号范围内的滤波器）。

离散多音DMT，Discrete MultiTone：ADSL文档中对于OFDM的称呼

ADSL：非对称DSL，具体原理见p116/130。指上行和下行（上传和下载）信道数量不同或容量不等的DSL。

简单老式电话服务POTS，Plain Old Telephone Service：原有的电话服务信号，使用DSL的最底层信道。

传输信息时，可以通过改变传输编码方式来调节速度，比如使用不同点数的星座图。

ADSL设备配置图p117/131

网络接口设备NID，Network Interface Device：

分离器splitter：一个模拟滤波器，用于将POTS信号和数据信号分离。分离器可能和NID组成一体。

数字用户线路接入复用器DSLAM，Digital Subscriber Line Access Multiplexer：包含一个数字信号处理器，可从信号中恢复出比特，进而构造数据包发送给ISP

ADSL的国际标准称为G.dmt，另一种无住户分离器的设计方案称为G.lite

其他高带宽服务：ISDN

（3）光纤到户FttH，Fiber to the Home

无源光网络PON，Passive Optical Network：每个用户通过一根光纤连接到一个光分离器splitter和光合并器combiner上，经过合并的信号通过一根光纤连接到端局，反之下行信号被分离给各个用户。上行和下行的波长不同，但被所有用户共享。

千兆级PON，GPONs，Gigabit-capable PONs：ITU标准

以太PON，EPONs，Ethernet PONs：IEEE标准

光合并器并非同时将两个以上信号进行某种计算式的合并，那样会产生冲突。它只是将来自不同用户的信号精确串联在一起。实际上，上行过程每个用户的信号必须是独立发送的，因此需要用到“请求”和“时间槽”的概念。

2.6.4 中继线和多路复用

①数字化语音信号

群group：60-108kHz波段的12个电话呼叫组成

超群supergroup：5个群

编码解码器codec，coder-decoder：在端局将模拟信号数字化的设备

脉冲编码调制PCM，Pulse Code Modulation：Codec每秒采集8000个样值，每个信号的样值幅度被量化成一个8bit数字。

模拟信号量化数字信号过程采用对数尺度，对于相对较小的信号幅度用更多的比特表示。因此量化的错误率和整体信号的幅度成比例。

量化规则包括μ-规则μ-law和A-规则A-law。信号在量化之前其动态范围被压缩，数字化后也可以进一步压缩，在还原时被扩展恢复。也称为压缩扩展companding。

②时分多路复用

国际电信联盟ITU，旧称CCITT

T1载波T1 Carrier：格式上称为DS1，北美和日本地区使用。包含24个信道，每个信道包含8比特数据（或7比特数据1比特信令），另外1个比特作为控制，称为扩展超帧extended superframe。总共193比特。

信道内部用1比特作信令，称为带内in-band，或信道相关信令channel-associated signaling。表示信令和信道一一对应。

强占比特信令robbed-bit signaling：强占每个信道8bit中的一个最低有效位作信令

信道所有bit均用来发送数据，而专门使用一条独立信道发送信令，称为带外out-of-band处理，也叫公共信道信令common-channel signaling。

E1载波：北美及日本以外地区使用，有32个8bit数据样值作为一个125微秒的帧。2/32的信道用于信令。

4个T1输入流通过TDM复用形成一个T2载波，同理7个T2输出一个T3，6个T3输出一个T4。组合过程中还要添加一些额外的比特作为控制、同步、成帧等。不过这种输入输出比例是标准规定的，存在标准间的不同。

③SONET/SDH

同步光网络SONET，Synchronous Optical NETwork：是一种光纤TDM系统的标准，也就是光纤传输的协议标准。用来统一现存的不同数字系统和各种低速传输标准。

同步数字系列SDH，Synchronous Digital Hierarchy：一些类似于SONET的ITU建议标准

同步传输信号-1，STS-1，Synchronous Transport Signal-1：基本的SONET信道，每秒钟传输8000次，每次传输810字节的数据块。所有的SONET中继线都是STS-1的倍数。

段section开销在每一段的开始和结束时生成和校验

线路line开销在每条线路开始和结束时生成和校验

同步有效载荷信封SPE，Synchronous Payload Envelope：承载用户数据的部分

SONET和SDH的线路载波种类和相关信息表p123/137

若载波没有被复用，而是仅承载单个源的数据，则在原有线路名称后加字母c

④波分多路复用

定义：WDM，Wavelength Division Multiplexing。即光纤传输领域内的频分复用。但是和一般的电子FDM区别在于，光纤系统是无源的，因此可靠性更高

密集波分多路复用DWDM，Dense WDM：信道数目很大，波长间隔非常接近

2.6.5 交换

电话系统分为两个基本部分：局外部分（本地回路和中继线）与局内部分（交换机）

①电路交换技术circuit switching

电路交换的概念就是当一方发起呼叫时，系统的交换设备在呼叫双方之间建立起物理连线的过程。特点是专用线路一旦连接速度很快且不会拥塞，但寻找路径建立连接的过程可能很长。

②数据包交换技术

并不在发送接收的两端建立独立稳定的连接，而是将信息做成一定大小的数据包分别路由至接收端。每个包的路由路径可能不同，到达的顺序可能不同，都由网络环境决定。

包交换技术存在路由器处的存储-转发延迟、为了等待带宽的排队延迟queuing delay、拥塞问题等。因此对于单独用户而言，服务质量下降，但整体的效率提升，且容错率提升（可以绕过故障的交换机）。

包交换的服务重在流量大小，电路交换则重在连接时间和距离（距离对于个别产品不重要）。

IP电话技术使用数据包交换。

电路和数据包交换的比较细节p128/142

## 2.7 移动电话系统

移动电话mobile phone和蜂窝电话cell phone是一回事。

无绳电话cordless phone不是移动电话，只是家庭内部使用的无绳电话，如果离其基座过远则无法使用。

2.7.1 第一代移动电话（1G）：模拟语音

按钮通话系统push-to-talk system：只有一个信道用于发送和接收，发送消息时必须按按钮关闭接收功能，同时所有人都能接收到信息。出租车警车的无线电通常使用这种技术

改进型移动电话系统IMTS，Improved Telephone System：拥有两个频率，用于双向信道

①高级移动电话系统AMPS，Advanced Mobile Phone System

也称作TACS（英国）、MCS-L1（日本）。也就是蜂窝电话系统。

蜂窝cell：模拟电话系统中，将地理区域划分为小块，10-20千米。每个蜂窝使用一组频率，只是保证和其相邻蜂窝频率不同。必要时可以再进行细分成微蜂窝microcell以缓解过大的频率占用压力。

移动电话交换局MTSO，Mobile Telephone Switching Office或移动交换中心MSC，Mobile Switching Center：是移动电话网络的交换处，类似电话网络中的端局。手机和基站使用无线电通信，基站和交换局、PSTN则使用数据包交换网络进行通信。

切换handoff：也称移交handover，正在通话的手机转移了所在蜂窝所需的信道调整。

②信道

频分双工FDD，Frequency Division Duplex：全双工信道，每个信道由一对单工信道组成

信道种类：  
控制信道（基站到移动电话）用于管理用户电话  
寻呼信道（基站到移动电话）用于提醒移动用户有呼叫  
接入信道（双向）用于呼叫建立和信道分配  
数据信道（双向）承载语音或数据

一般每个蜂窝保留21条信道用于控制，实际可用的信道约45个。

③呼叫管理

移动电话的呼叫和收听过程p132/146

2.7.2 第二代移动电话（2G）：数字语音

数字化较模拟信号的优点在于数字信号的容量更小（数据更紧凑），数字化增加了加密功能提升安全性。

数字高级移动电话系统D-AMPS，Digital Advanced Mobile Phone System：数字版本的AMPS，使用TDM

全球移动通信系统GSM，Global System for Mobile communication：2G系统的另一个种类，使用FDM和TDM

2G系统中也有CDMA，但是不如3G普及

个人通信服务PCS，Personal Communications Services：2G系统的别称

①GSM

用户识别模块SIM，Subscriber Identity Module：手机卡，2G系统中的产物

空中接口air interface

基站控制器BSC，Base Station Controller：管理蜂窝基站的

访问位置寄存器VLR，Visitor Location Register：记录所有附近的移动电话

归属位置寄存器HLR，Home Location Register：每个移动电话的最后一个已知位置

多帧multiframe结构：即具有层次型结构的帧，如多个用户时间槽（数据帧）组成一个TDM帧，多个TDM帧和一些控制帧组成一个更高级的复杂帧

广播控制信道broadcast control channel：从基站输出一个连续流，包含该基站的标识和信道状态。所有移动站都要监视其以确认自身位置。

专用控制信道dedicated control channel：用于移动用户的位置更新、注册和呼叫的建立。VLR的信息在此发送。

公共控制信道common control channel：分3个子信道  
寻呼信道paging channel，基站用其通告有关入境呼叫的情况  
随机接入信道random access channel，允许用户在专用控制信道上请求一个时间槽  
接入授予信道access grant channel，用于宣布分配获得的专用控制信道的时间槽

移动辅助切换MAHO，Mobile Assisted HandOff：移动设备利用时间槽之外的空余时间参与辅助切换基站的操作。

2.7.3 第三代移动电话（3G）：数字语音和数据

宽带码分多址WCDMA，Wideband CDMA：现在也称为全球移动通信系统UMTS，Universal Mobile Telecommunications System。由爱立信公司提出

CDMA2000：另一个3G系统的样式，由Qualcomm提出

互相关性cross-correlation

自相关性auto-correlation

CDMA优点：  
a.可提升容量：普通的FDM和TDM技术仍然无法彻底利用信道的空闲期。一旦有空闲时间信道会被浪费。CDMA的信道由多个用户共享，个别用户的空闲会增加其他用户信号的质量。  
b.每个蜂窝可以使用相同的频率：区分用户的不再是频率或时间，而是密度更大差异更多的地址码。另外由于蜂窝可使用相同频率，基站也可以使用定向天线或扇形天线sectored antennas来提高信号质量  
c.软切换soft handoff技术：移动电话在移动中切换基站无需中断通话，因为频率一致。与前代的硬切换hard handoff对应。

增强数据率的GSM演进EDGE，Enhanced Data rates for GSM Evolution：2G和3G系统的过渡系统，仍然是GSM，只是使用每个符号运送更多比特。

长期演进LTE，Long Term Evolution：4G系统的别名。

WiMAX：802.16，4G性能的无线网络

## 2.8 有线电视

2.8.1 共用天线电视Community Antenna Television

这是有线电视的雏形。

2.8.2 线缆上的Internet

混合光纤电缆HFC，Hybrid Fiber Coax：远距离用光纤，用户端用同轴电缆的模式

光纤节点fiber node：光电转换器处

有线电视原有的同轴电缆带宽比双绞线要大，但是需要多用户共享。

2.8.3 频谱分配

有线电视网络接入Internet的数据流使用FDM技术，将低频段用作上行，高频段用作下行。中间部分仍用作电视节目。

线缆调制解调终端系统CMTS，Cable Modem Termination System：代替原有的有线电视头端的设备，用于接入internet

2.8.4 线缆调制解调器

线缆调制解调器cable modem：有线电视Internet用户的用具，不同于CMTS，那是头端非用户端设备。

线缆数据服务接口规范DOCSIS，Data Over Cable Service Interface Specification：即有线电视internet的标准。

迷你槽minislots：时间槽的另一个称呼

测距ranging：调制解调器测量其到头端的距离，以精确计算发送时间符合TDM。只有上行流才需要TDM，下行流只有头端才是发送者使用包交换。

解决时间槽竞争的方法：CDMA和等待随机时间之后再次尝试

2.8.5 ADSL与有线电视电缆（比较）

P145

# 第3章 数据链路层

## 3.1 数据链路层的设计问题

数据链路层功能：  
（1）向网络层提供一个定义良好的服务接口  
（2）处理传输错误  
（3）调节数据流，确保慢速的接收方不会被快速的发送方淹没

帧frame：数据链路层完整信息块的名称。包括帧头、有效载荷和帧尾。有效载荷用于存放网络层的数据包。数据链路层将数据包包装成帧来传输。

3.1.1 提供给网络层的服务

数据链路层提供的3种可能服务包括：  
（1）无确认的无连接服务：适合短距离可靠信道或实时传输  
（2）有确认的无连接服务：适合短距离不可靠信道  
（3）有确认的有链接服务：适合长距离不可靠链路

有连接的服务需要建立各种变量和计数器以便追踪记录帧的收发情况，确保每个帧只接收一次，且按照一定顺序。

3.1.2 成帧

拆分比特流成为离散的帧，有4种方法：  
（1）字节计数法：  
头部一个字段来标识帧中字符数，计数出错时难以更正，很少用。  
（2）字节填充的标志字节法：  
利用标志字节flag byte标记帧的起始和结束，标志字节并非也不包含帧头或帧尾，解码角度看没有实际意义。另外，每个帧都有两个标志字节，相邻不共享。如果标志字节必须作为数据内容出现，使用转义字节标记以区分，恢复数据包时再删除。这种技术称为字节填充byte stuffing。转义字节本身也可以被转义字节修饰，对于一串转义字节只有第一个有转义字节的实际意义。这种方法用于PPP协议中（PPP或P2P？Point-to-Point Protocol）。  
（3）比特填充的标志比特法  
道理类似方法2，只是使用比特填充而非字节填充来区分标志flag和内容。优点是减少了需要用于转义的额外帧长度。  
（4）物理层编码违禁法  
使用违反物理层编码方式的数据（无法解码）来作为标志。

以太网和802.11使用一个定义良好的比特模式标志帧，也称为前导码preamble。其他的数据链路协议也往往综合使用不同的方法。

3.1.3 差错控制

数据链路层实现一般需要有计时器和序号。计时器用以测量从发送之后经历的时间，一定时间未回复确认的，可以重传。序号用以编排帧，以保证收到的每个帧按照顺序且只保留一个。

3.1.4 流量控制

流量控制有两种：基于反馈的流量控制feedback-based flow control和基于速率的流量控制rate-based flow control。前一种使用较多。

网络接口卡NIC，Network Interface Cards：链路层实现硬件，简称网卡

## 3.2 差错检测和纠正

错误处理的两种基本策略：  
都是在数据中加入冗余信息以便接收方据此进行分析。  
纠错码error-correcting code，也称为前向纠错FEC，Forward Error Correction。可以推断出内容是什么。  
检错码error-detecting code，只能推断出是否错误。

错误模型：  
单个比特错误，偶尔出现的极端热噪声淹没信号。  
突发错误，偶尔出现的物理干扰导致的一串多个比特错误。  
擦除信道erasure channel，一个信号远离了所有预期值而被彻底舍弃。

3.2.1 纠错码

设一个帧由m个数据位（信息）和r个冗余位（校验）组成。n=m+r

块码block code：r个校验位是作为与之相关的m个数据位的函数计算获得的

系统码systematic code：直接发送m个数据位，然后发出r个校验位，而非在发送前编码

线性码line code：校验位是作为数据位的线性函数计算出的，一般包括异或和模2加法等。

码字codeword：即一串编码

码率code rate：码字中不包含冗余部分的比例，即m/n

①海明码

海明距离 Hamming distance：两个码字中不相同的位的个数

块码的检错和纠错特性和海明距离有关。为了检测d个错误，需要海明距离d+1的编码方案，为了纠正d个错误，需要距离为2d+1的编码方案。检错只是发现错误，无法确定原来正确的编码是啥，因为可能有至少两种可能，因此纠错的距离要求较高。

海明码的校验位是穿插在数据位中的，占据2的幂次方位，如1、2、4、8等。校验位是根据数据位计算得来的。数据位和校验位的对应关系由其位次决定，如第11位和1+2+8三个校验位有关。每个校验位由和所有和它有关的数据位反向通过某种计算得来。当接收方收到编码，首先重新计算校验位，然后再和收到的校验位比较，即可知哪里出错。

编码的规则、校验的协议等都是通过Galois领域或稀疏矩阵的高级数学知识制定的。不必自行研究获取，了解现成通用的规则即可。

②二进制卷积码

卷积码convolutional code：既不是块码也不是系统码，规则原理简述p161

约束长度constraint length：决定当前输出的以前输入位数称为代码的xxx。也就是内部状态寄存器的位数+1。内部寄存器用来存储最近几位输入位的值，而且是滚动移位式存储，因此需要寄存器位数+1次移位方可将上一次的所有输入位清空。

软判决解码soft-decision decoding：带有一位不确定性的工作方法，将纠错视为概率

硬判决解码hard-decision decoding：纠错前就决定了每个位是0或1的工作方法

③里德所罗门码Reed-Solomon code

线性块码、系统码

Massy, 1969

④低密度奇偶校验码LDPC, Low-Density Parity Check

Gallagher, 1962

3.2.2 检错码

①奇偶

使数据码中的所有0和1的个数全为奇数个或偶数个，以此可检验1位错误

②校验和checksum

Internet校验和，简单但是保护能力不高

Fletcher校验和，考虑了数据位置变化的影响，1982

③循环冗余校验CRC，Cyclic Redundancy Check

也称多项式编码polynominal code，将位串看成是多项式的系数（只有0和1），n位码代表一个n-1次的多项式。

利用校验码代表的多项式去除编码串代表的多项式，得到余数来判断校验结果。

## 3.3 基本数据链路层协议

协议公用的声明举例：p169/183

3.3.1 一个乌托邦式的单工协议

P171

3.3.2 无错信道上的单工停-等式协议

停-等式协议stop-and-wait：发送方发送一帧，等待对方确认（一个哑帧）到达后继续发送

3.3.3 有错信道上的单工停-等式协议

自动重复请求ARQ，Automatic Repeat reQuest或带有重传的肯定确认PAR，Positive Acknowledgement with Retransmission：发送方在前移到下一个数据之前必须等待一个肯定确认。P175/189

数据链路层协议正确性的基本原则包括：  
①接收端向网络层提交的数据包不能重复，且按照发送方的预定顺序。

## 3.4 滑动窗口协议sliding window

捎带确认piggybacking：暂时延缓确认以便将确认信息搭载在下一个出境数据帧上

滑动窗口是一类协议的名称，是双向传输数据的协议，且每个出境帧都具有一个序号，停-等式滑动窗口协议只用0和1两个序号。“窗口”的概念是指一组序号，发送方和接收方分别维护一组序号，对应于允许发送/接受的帧，称作发送/接收窗口sending/receiving window。每当发送、接收、收到确认等操作发生后，窗口内的序号根据情况进行调整。因为总是增加或删去最小或最大值，因此看上去像是窗口在滑动。

大多数情况下，无论接收端接收到的结果如何（只要不是没收到），都必须返回一个确认帧，只是确认帧的帧头包含不一样的序号。

3.4.1 1位滑动窗口协议

如果存在计时器过早超时或双方近乎同时发送数据导致彼此的数据流交叉而非交替时，会出现多次重复发送的情况，浪费带宽。

3.4.2 回退N协议go-back-n (p182/196)

长发送时间、高带宽和短帧是共同影响带宽使用效率的因素。

阻塞block：不能完全理解为某事物过多堆积造成拥堵（实际上这种情况称为拥塞congestion），可以广义的理解为因为系统其余部分没有足够快速的配合该部分的工作因而其暂时无法继续工作的状态，或称为挂起。

带宽-延迟乘积bandwidth-delay product：简称BD，即带宽（传输速率）和单向传输时间的乘积，结果是一个比特数，根据帧的大小可以进而得出单向信道同时可容纳的最大帧数。

链路利用率≤w/(1+2BD)。w是允许发送方在阻塞之前一次性连续发送的最大帧数，同时也是发送方的窗口大小（有几个帧）。

管道化pipelining：保持多个帧同时在传送的技术。应对管道化传输中，帧出错的问题，有两种基本协议类型，回退n和选择重传

回退n协议：当接收方遇到错误帧，简单丢球所有到达的后续帧，且不返回确认。使发送方超时重新从错误处再次发送一遍。

任何时候，可以发送的帧的最大个数w不能等于序号空间的大小，因为序号空间是循环利用的，这样做会导致另一方无法区分序号。这是对于顺序接收的协议而言的，非顺序接收的协议要求会更严苛。

累计确认cumulative acknowledgement：根据回退n协议的特点，当n号帧的确认到达，所有发送窗口内序号在n之前的帧都被自动确认。

多个超时计时器的实现方法：p184/198

3.4.3 选择重传协议selective repeat (p185/199)

选择重传：接收到坏帧之后，缓存后续的所有好帧，直到接收到连续的一组好帧，向网络层递交。

否定确认NAK，negative acknowledgement：当收到错误帧时，发送一个否定的确认以触发发送端重传操作，如此可以不完全依赖发送方超时来重传。一般和选择重传协议共用。

选择重传协议窗口尺寸最大为(MAX\_SEQ+1)/2，MAX\_SEQ为最大的序号值。这样保证了，每次接收窗口滑动时，滑动距离即便最大也不会出现与滑动前的序号重复。缓冲区的大小、计时器数量都和窗口大小一致。

通过设置辅助计时器来保证即便没有反向流量捎带确认时，确认帧也能以独立帧的形式尽快返回到发送端。辅助计时器的超时标准一定要明显小于发送超时计时器。

一般而言，当确认间隔的标准偏差与间隔本身相比非常大的时候，超时计时器应该设置得较松，利用NAK来加快重传速度。但若确认间隔较稳定，或信道始终有双向流量及时捎带确认信息，将超时计时器定得较紧，NAK的作用便不高。

NAK是一种重传请求，为了避免多次请求重传同一帧，应该设置变量进行管理。

## 3.5 数据链路协议实例

点到点协议PPP，Point-to-Point Protocol：一种数据链路协议的例子，被SONET和ADSL采用。

PPP由RFC1661定义，在RFC1662种进一步阐述，Simpson 1994a，1994b

串行线路Internet协议，SLIP，Serial Line Internet Protocol：一个早期的简化协议，PPP改进了该协议。

PPP包含的主要特性：  
①一种无歧义成帧区分帧的方法  
②一个链路控制协议LCP，Link Control Protocol  
③多种支持不同网络层的网络控制协议NCP，Network Control Protocol

PPP的帧格式类似高级数据链路控制协议HDLC，High-level Data Link Control，其为一个早期广泛使用的家庭协议实例。

PPP帧格式：p191/205

3.5.1 SONET上的数据包

PPP链路状态转换图和连接建立结束过程：p192/206

3.5.2 对称数字用户线

ADSL协议栈：p193/207

异步传输模式ATM，Asynchronous Transfer Mode：一种链路层协议，异步意味着信元不以连续的方式发送。传输基于信元，信元在头部带有虚电路virtual circuit标识符，道理上和帧类似。

ATM适应层5，AAL5，ATM Adaptation Layer 5：用于数据包数据的适配层，适配层的任务是将数据映射成一系列的信元cell，映射过程包括分段segmentation和重组reassembly。其帧的结构：p194/208

# 第4章 介质访问控制子层

广播信道也称多路访问信道multiaccess channel和随机访问信道random access channel，区别于点到点PPP，因为所有用户共用一个广播信道。注意广播信道不一定用来让所有站都接收数据，只是所有站都使用同一信道。可以仅支持两个站之间的通信，此时其他的站若接收到数据也不会去解码。广义上说，多路访问信道（特点是站发送时会产生冲突）不一定完全等同于广播信道（特点是所有信息会被所有站接收）。

介质访问控制子层MAC，Medium Access Control：是数据链路层的一个子层，且位于底部。用于确定多路访问信道使用者分配。在LAN，尤其是无线LAN中非常重要。注意局域网不代表就使用广播式信道，它可以使用不同的技术类型实现。

## 4.1 信道分配问题

4.1.1 静态信道分配

计算机系统中数据流量表现出极端突发性，通常峰值流量与平均流量之比能达到1000:1

排队理论分析静态信道效率：p200，泊松模型

划分了信道（FDM）比不划分信道的平均延迟差N倍

4.1.2 动态信道分配的假设

（1）流量独立independent traffic

（2）单信道single channel

只有一个信道可用

（3）冲突可观察observable collision

冲突即两帧同时传输

（4）时间连续或分槽continuous or slotted time

（5）载波侦听或不听carrier sense or no carrier sense

侦听即查看当前信道是否繁忙

多路访问协议不能保证可靠传送，要由链路层其他部分或更高层次来提供数据传输可靠性

## 4.2 多路访问协议

4.2.1 ALOHA

①纯ALOHA

竞争contention：系统中多个用户共享同一个信道导致冲突

基本思想：所有帧长度相同，所有用户随意发送帧，冲突后等待随机时间再次发送

分析过程：p204/218

②分槽ALOHA

基本思想：将时间分成离散槽，只有在每个空槽起始时才能发送帧

分析过程：p205/219

4.2.2 载波侦听多路访问协议carrier sense protocol

一般只要多个站之间能够相互通信（如局域网），就可以进行载波侦听，甚至做到发送前的冲突检测。而其他的一些网络（如无线网、有线电视网络）由于站与站之间无法直接通信，因此无法使用载波侦听。

①坚持和非坚持CSMA，Carrier Sense Multiple Access

1-坚持载波检测多路访问：当站有数据要发送，首先侦听信道，若信道空闲，则必然发送数据，反之等待至空闲。对于冲突的处理是，发送完毕后若反馈为冲突，则随机时间后再发送。

非坚持CSMA，nonpersistent：发送前侦听信道，若占用，则等待随机时间后再侦听，反之立即发送数据。

p-坚持CSMA，p-persistent：类似1-坚持，只是若监听到空闲，则发送数据的概率为p，而1-p概率下将发送推迟到下一个时间槽（适用有分槽的情况）。

②带冲突检测的CSMA，CSMA/CD，CSMA with Collision Detection

侦听和冲突检测不一样，前者为确定信道是否空闲，而本站未发送数据；后者是本站已经开始发送数据，检测信号是否被冲突影响。检测到影响后提前终止发送以节约带宽，否则需等到发送完毕接收方的反馈情况才能知道是否冲突。

只有当一个站传输了2t时间之后还没有监听到冲突，才可以确保已经占据了信道。t为距离当前站最远的站之间单向传播信号所需时间延迟。

4.2.3 无冲突协议

①位图协议

基本位图法basic bitmap method：每个竞争期包含对应N个站的N个槽，若有站要发送，则在其对应的槽内发送1。竞争期过后，要发送数据的站轮流发送1帧。之后重复过程。

预留协议reservation protocol：传输数据前，先广播自己发送意愿的协议类型

低负载情况下，位图协议信道利用率较低。低位站平均延迟较长。

②令牌传递token

令牌环协议token ring：令牌和帧以相同的顺序逐步发送给下一个站，整个网络的拓扑结构是个环。令牌环标准化为IEEE 802.5

令牌总线token bus：网络结构是个长总线

光纤分布式数据接口FDDI，Fiber Distributed Data Interface：一种更快的令牌环

弹性数据包环RPR，Resilient Packet Ring：IEEE 802.17，城域环组合标准

③二进制倒计数binary countdown

这是一种竞争仲裁型的协议，每个站具有一个2进制地址，且所有站的地址位数相同。当需要发送数据时，站广播自己的地址。高位站在仲裁后具有优先权发送数据。若能使发送方的地址恰好是帧内的第一个字段，则信道利用率为100%。

4.2.4 有限竞争协议limited-contention protocol

考察协议性能的指标：  
•低负载下的延迟  
•高负载下的信道利用率

定义：低负载下采用竞争的做法提供较短延迟，高负载下采用无冲突/无竞争技术提高信道效率。

协议内容：将所有的站分组（不一定两两不相交），0号组的成员才可以竞争0号时间槽，以此类推。若产生冲突，1号组竞争1号槽（0号槽相当于浪费掉）。低负载下，每组或每槽中站点增多，高负载下，每组站点减少。

①自适应树遍历协议

有限竞争协议本质上仍然允许竞争，只是在负载增高时降低竞争。自适应树遍历是一种解决有限竞争协议分组或信道分配的算法。

4.2.5 无线局域网协议wireless LAN

接入点AP：无线局域网的“基站”，AP和站（笔记本电脑等无线设备）都是以无线电相互通讯，AP之间使用有线介质。

无线网不仅每个站之间很可能无法直接联系（有线网中所有的信号可以传播至所有的站），而且同一个站的发送和接收信号强度差别很大，很难进行侦听/冲突检测。但是，也因为这个与有线网的不同，无线网中不同的站有时可以同时传输数据且不干扰。

隐藏终端问题hidden station problem：由于竞争者太远而导致无法检测到潜在竞争者。竞争往往发生在接收端，因为发送端往往彼此无法互相直接通信。

暴露终端问题exposed station problem：检测到并不会影响传输的竞争者存在，该竞争者可能在向另一终端传输，但不会影响本站传输给其他站。

冲突避免多路访问MACA，Multiple Access with Collision Avoidance：发送方发送数据之前发送或广播（无线时）一个RTS（Request To Send）帧，预期接收方用一个CTS（Clear To Send）作为应答。完成应答之后发送方便开始传输数据，其他站或AP若接收到RTS，必须保持沉默至发送方收到CTS，若接收到CTS则必须沉默至数据发送完毕。若收发RTS和CTS过程中遇到冲突，则等待随机时间之后再行连接。

## 4.3 以太网（802.3）

简单来说，以太网是一种计算机局域网技术，是目前应用最普遍的局域网技术。

经典以太网classic Ethernet：使用本章协议解决多路访问问题

交换式以太网switched Ethernet：分为快速以太网，千兆以太网和万兆以太网

4.3.1 经典以太网物理层

DIX standard：DIX标准，1978年的10Mbps以太网标准

经典以太网结构p217/231

4.3.2 经典以太网MAC子层协议

以太网DIX和802.3的帧格式：p218/232

前导码preamble：以太网或IEEE802.3的帧的开头，用来同步时钟和预示帧的内容开始，使用曼彻斯特编码

定界符start of frame，SOF：前导码的最后一个字节

组织唯一标识符OUI，Organizationally Unique Identifier：IEEE分配的每个组织世界无二的以太网源地址，共6个字节。前三个字节即OUI，后三个字节为组织的设备编号（自行决定）。

以太网和802.3不完全一致，历史原因导致现在二者几乎完全兼容一致。

逻辑链路控制LLC，Logical Link Control：协议头，在数据内传达协议类型信息

①二进制指数后退的CSMA/CD

二进制指数后退binary exponential backoff：是一种确定冲突发生后，确定随机等待时间的方法。时间离散为槽，第i次冲突后，从0-2^i-1之间选择一个随机数n，然后等待n个时间槽后再次发送。一般最多16次冲突后报告失败，而不再试。

一般CSMA/CD和以太网（有线）都不支持确认机制，因为可靠性较高，没有冲突报错即认为发送成功。若仍有少量错误，只需检错后由高层完成。

4.3.3 以太网性能

数学推导p221/235

4.3.4 交换式以太网

交换机switch：交换式以太网使用的组件，每个端口连接一台或一组设备，每个端口之间传输数据彼此不影响（无需竞争同一信道），道理类似电路交换。但是每个端口同时只能接收一方的数据，反之亦然（使用缓存保存额外的数据）。

双绞线集线器：以太网发展过程中的中间形式，用以解决单根电缆的缺点

冲突域collision domain：会发生冲突的信道或频域

混杂模式promiscuous mode：每台机器都可以接收到所有帧，广播式

4.3.5 快速以太网（802.3u，fast Ethernet）

自动协商autonegotiation：一种允许两个站自动协调最佳速度和双工模式的机制。

4.3.6 千兆以太网（802.3ab，gigabit Ethernet）

载波扩充carrier extension：给短帧增加位数以凑够长度（可以延长冲突线路的长度），这些位数交给上层时自动删去。

帧突发frame bursting：一次发送多个帧（若有多个帧准备就绪），作为优于载波扩充的另一种帧延长技术（只是在特定情况下帧有效长度不足时才用）。

巨型帧jumbo frame：帧长度超过1500字节

4.3.7 万兆以太网10 gigabit Ethernet

4.3.8 以太网回顾

光纤分布式数据接口FDDI，Fiber Distributed Data Interface：

光纤信道Fibre Channel：以上两者基于光纤局域网的双环结构，不与以太网兼容，商业失败

电信级以太网carrier-grade Ethernet：

## 4.4 无线局域网

4.4.1 802.11体系结构和协议栈

分布式系统distribution system：若干个接入点AP通过有线网络连接在一起形成的扩展网络，AP是无线客户端接入网络的必经之路。

自组织网络ad hoc network：没有AP只有若干客户端的网络

4.4.2 802.11物理层

速率自适应rate adaption：根据信号强弱等状况自动调整传输速度

巴克序列Barker sequence：802.11b使用的扩展序列，低速率下

补码键控CCK，Complementary Code Keying：较高速率的802.11b采用的编码序列

多入多出MIMO，Multiple Input Multiple Output：802.11n采用的用以分离多根天线信息流的技术

4.4.3 802.11MAC子层协议

无线网络的冲突检测完全不可行，因为发送信号和接收信号的强度有很大不同，基本上都是半双工的。

带有冲突避免的CSMA，CSMA/CA，CSMA with Collision Avoidance：802.11采用的协议，虽类似CSMA/CD，但是区别在于不等待冲突发生，实际上也无法检测。p234

分布式协调功能DCF，Distributed Coordination Function：每个站独立行事，无中央控制

点协调功能PCF，Point Coordination Function：由AP全权负责自己范围内的活动，很少使用，因为绝大多数情况下，一个AP覆盖的区域很难做到像蜂窝一样彼此无交集。

隐藏终端和暴露终端问题：两种问题都是由于同一个事实原因造成的，那就是无线网覆盖有范围局限性，而有线网没有。可以部分理解为无线网络实际上是每两个用户独立相连，而非所有用户连在一起。

物理侦听：就是简单的检查介质中是否有物理信号

虚拟侦听：通过维护一个网络分配向量NAV，Network Allocation Vector来记录信道占用情况的信息，RTS或CTS帧可用以传输NAV向量，其他站接收到便相当于在某时段预定了一个时间槽。但这种方式只能解决隐藏终端问题，而该问题出现频率较低，用后退重发的方式解决也可。

可靠性：  
①降低传输速率以增加传输成功率，因为可以使用更复杂的调制解调方法  
②发送短帧——由AP可调整网络层消息尺寸，进行进一步分割为段fragment

节能：  
①信标帧beacon frames：由AP定期广播给所有客户端，用以通告AP存在和一些其他系统参数  
客户端可以用电源管理位通告AP其进入power-save mode，只在每个信标帧到来时查看和接收与其有关的信息，其他时间由AP缓存  
②自动省电交付APSD，Automatic Power Save Delivery：一个新机制，由客户端自主决定休眠时间长短，直到其自行活跃后才开始和AP交换并处理缓存

服务质量：802.11e  
通过定义不同优先级的帧之间的不同时间间隔来实现优先级划分，高优先级的帧以较短间隔发送  
DCF帧间隔DIFS，DCF InterFrame Spacing：普通数据帧  
短帧间间隔SIFS，Short InterFrame Spacing：用以VoIP等高优先级  
仲裁帧间间隔AIFS，Arbitration InterFrame Spacing：可用于一系列不同优先级  
扩展帧间间隔EIFS，Extended InterFrame Spacing：坏帧未知帧报告

传输机会TXOP，Transmission Opportunity：802.11a/g 用以协调不同速度发送站之间的吞吐量，不会以慢速方拖累快速方

4.4.4 802.11帧结构

详见p239

4.4.5 服务

①关联Association：用以AP和移动站之间建立第一次联系的服务

②重新关联Reassociation：用以移动站改变AP或二者之一遇到维护或关闭等情况

③认证Authentication：  
有线等效保密WEP，Wired Equivalent Privacy：有缺陷，发生在关联之前  
WiFi保护接入2（WPA2，WiFi Protected Access 2）：通过AP和认证服务器联系来认证移动站  
可配置的预共享密钥prehared key：通过AP和站之间交互认证，发生在关联之后

④分发distribution和集成integration：用以路由帧至Internet其他区域和协议转换

⑤数据传送data delivery：上层必须处理差错和纠正事宜

⑥隐私privacy  
高级加密标准AES，Advanced Encryption Standard：WPA2基于的标准

⑦QoS流量调度QoS traffic Scheduling

⑧发射功率控制transmit power control

⑨动态频率选择dynamic frequency selection

## 4.5 宽带无线

全球微波接入互操作性WiMAX，Worldwide Interoperability for Microwave Access：802.16或无线宽带的正式技术名称

4.5.1 802.16与802.11和3G的比较

802.16更加接近4G蜂窝网络的特点p242

4.5.2 802.16体系结构与协议栈

注意体系结构和协议栈的描述方式p243

4.5.3 802.16物理层

正交频分多址OFDMA，Orthogonal Frequency Division Multiple Access：OFDM的改进版，为不同站分配不同的子载波集，不是所有子载波都只能由一个站同时收发

时分双工TDD，Time Division Duplex：站交错着进行发送和接收，WiMAX使用TDD和FDD两种方式，首选TDD。其简单工作原理描述见p245

4.5.4 802.16的MAC子层协议

p245/259

4.5.5 802.16帧结构

p246/260

## 4.6 蓝牙Bluetooth

配对pairing：彼此发现并连接的行为

4.6.1 蓝牙体系结构

微网piconet：一个主节点和10米范围内最多7个活跃的从节点

散网scatternet：一组互联的微网

驻留节点parked node：处于低功耗状态的节点，或者休眠状态的节点，一个微网最多255个驻留节点。

蓝牙使用主从模式，主设备或节点负责通信和控制，从节点之间无法直接通信。

4.6.2 蓝牙应用

蓝牙的应用类型受到了开发方式由多个小组负责的影响，所以不同类型的应用有不同的协议栈，这些种类划分被称为轮廓profiles。这种分类方式可以肯定是有些多余的。

4.6.3 蓝牙协议栈

p249/263

主机-控制器接口：为了实现方便而定义的系统划分，接口线以下由蓝牙芯片实现，反之由设备实现。

逻辑链路控制适配协议L2CAP，Logical Link Control Adaption Protocol

服务发现service discovery：协议，用于网络中寻找可用服务

射频通信RFcomm，Radio Frequency Communication：模拟pc上的标准串行端口，用于连接pc外部设备。

4.6.4 蓝牙无线电层

p250/264

自适应跳频adaptive frequency hopping：蓝牙采用的跳频技术特点

4.6.5 蓝牙链路层

链路link：逻辑信道

安全简单配对secure simple paring

个人识别号码PIN，Personal Identification Number

同步有连接SCO，Synchronous Connection Oriented：一种链路，用于实时数据传输

异步无连接ACL，Asynchronous ConnectionLess：另一种链路，用于传输数据包

4.6.6 蓝牙帧结构

p251

蓝牙由于成本和使用环境限制，数据的传输采用大量冗余的形式，即多次发送相同数据来达到纠错的目的。

## 4.7 RFID

无线射频识别RFID，Radio Frequency Identification

电子产品码EPC，Electronic Product Code：RFID技术的一种应用形式，用于替代条形码，可读距离较远且无需可见（可穿透一般阻挡物）

4.7.1 EPC Gen 2体系结构

RFID体系包括标签和读写器，标签具备识别码和少量内存，读写器可以读取和写入信息，主要工作还包括发现附近的标签

1类标签：没有电池，只能依靠读写器传输功率运行的标签

4.7.2 EPC Gen 2物理层

RFID使用超高频UHF，Ultra High Frequency

后向散射backscatter：一种类似雷达反射形式的无线通讯方式，读写器永远在发送信号，而标签收到信号能量后才返回信号，二者不会在同一时间同时传输。

因为成本限制和低功耗的要求，信号的调制方法非常简单。

4.7.3 EPC Gen 2标签识别层

负责解决读写器和标签通讯过程中标签信道分配的问题，类似分槽ALOHA，因为标签彼此之间无法侦听

4.7.4 标签标识消息格式

即RFID消息帧结构，读写器发出的消息较复杂（相对于标签发出的），总体而言仍采用简短的消息格式。但是随着标签的工艺和功能复杂度增加，标签可以运行程序，消息格式可能会发生变化。

物联网Internet of things将会基于RFID和无线技术的进步。

## 4.8 数据链路层交换

网桥bridge：交换机的别称，是数据链路层设备，也是管理多个局域网之间连接的设备

虚拟局域网VLAN，Virtual LAN：

4.8.1 网桥的使用

网桥的深远意义在于通过连接多个LAN实现一体化，或者将一个大LAN分割化降低负载负担和增加网络总覆盖。

4.8.2 学习网桥

多点局域网：所有的站用一根线串起来的局域网结构，如经典以太网

点到点局域网：站与站之间不必用同一根线串联，站可能通过集线器或交换机（此时交换机就是网桥）与主干单联

后向学习法backward learning：一种可以让网桥通过尝试（泛洪）然后确认端口位置的算法

直通式交换cut-through switching：也称虫孔路由wormhole routing。不等帧完全被接收即开始转发，当然至少要接收到帧的目的地信息

中继模块relay

4.8.3 生成树网桥

生成树算法：网桥采用一种计算网络拓扑结构的算法，保证该结构无环，从而可以利用其它的算法而不致产生问题。算法标准化为IEEE 802.1D

4.8.4 中继器/集线器/网桥/交换机/路由器和网关

这些设备工作内容相似但不相同，实质上分属不同层，起决定的因素在于它们使用什么信息来决定如何交换。设备的工作层次见p263/277

交换机和网关既是严格的网络术语也是一种通用术语。通常交换机工作在较低的层负责转发，网关工作在较高层负责转发，而“转发”通常既包含继续发送的技术成分也包含格式转换的成分。

4.8.5 虚拟局域网VLAN，Virtual LAN

定义：通过软件的手段而不是物理实际的手段创建局域网或对原网络进行局域网划分。

VLAN感知VLAN-aware交换机：能够支持VLAN建立的交换机，通常VLAN使用颜色来作为标识符，目的是便于生成如地图一般的物理布局图

IEEE 802.1Q标准：修改了原有的以太网帧结构，增加了VLAN标识符字段。这种改动无需所有机器更换硬件，但显然需要发送和接收方附近必然有新型硬件。新旧硬件网卡可以混合搭配使用，旧硬件对于新帧中的内容不做任何工作而已。

802.1Q以太帧格式p269/283  
VLAN协议标识符VLAN protocol ID  
VLAN标识符VLAN identifier

VLAN技术的实现和VLAN标识符的存在使其更加接近面向连接的服务。

# 第5章 网络层

网络层是处理端到端数据传输的最底层，亦即如何将数据包从发送方路由到接收方。

数据链路层只是负责将帧从线路一边传送到另一边。

## 5.1 网络层的设计问题

5.1.1 存储转发数据包交换

5.1.2 提供给传输层的服务

（1）向上提供的服务应独立于路由器技术

（2）应该向传输层屏蔽路由器的数量、类型和拓扑关系

（3）传输层可用的网络地址应该有一个统一的编址方案

X.25和帧中继Frame Relay：上世纪70-80年代的数据网络

多协议标签交换MPLS，MultiProtocol Label Switching

5.1.3 无连接服务的实现

无连接服务的数据包必须带有完整的目标地址。

数据报datagram：无连接服务网络中数据包的别称

数据报网络datagram network：无连接服务网络别称

路由表：每个路由器对应的一个表格，写明了根据数据报最终地址（可以是路由器或其他设备）而决定的出境线路。P276/290

路由算法routing algorithm：管理路由表并作出路由选择的算法

IP协议Internet Protocol：互联网、无连接网络服务的协议，特点是每个数据包携带一个目标IP地址，分为IPv4的32位和IPv6的128位地址两种。

5.1.4 面向连接服务的实现p277/291

面向连接的服务仍然通过路由表路由实现，但数据包只带有连接标识符而非地址。标识符区分连接的虚电路，但实际上每条虚电路上所有的标识符不一定统一，存在由路由器更改标识符的情况，称为标签交换label switching。

虚电路VC，virtual circuit：面向连接服务中的所谓连接路径

虚电路网络virtual-circuit network：面向连接服务网络别称

5.1.5 虚电路与数据报网络的比较

P278/292

## 5.2 路由算法

会话路由session routing：虚电路连接时路由的别名

路由包括两个进程：转发forwarding和更新路由表

转发：数据包到达路由器时，所经历的处理、查找路由表、出境的过程。

路由算法的特性：正确性、简单性、鲁棒性（网络结构变化时无需停止所有工作）、稳定性（迅速收敛convergence）、公平性（每个客户都有使用权）、有效性（效率）

算法需要平衡的特性：公平性和有效性、有效性中的延迟最低和吞吐量最大（折衷：降低数据包经历的路由过程，减少延迟和带宽消耗，进而增加吞吐量）

非自适应算法nonadaptive algorithm：不会根据当前测量或者估计的流量和拓扑结构来调整路由决策。路由选择是网络启动时下载到路由器中，也称静态路由static routing，仍然适合可以直接确定路由路径的场合

自适应算法adaptive algorithm：也称动态路由dynamic routing。不同的自适应算法有几个要素的区别——获取网络信息的来源不同、改变路径的时机不同、用于路由优化的度量不同

跳数hop：一个路由度量，每一跳可以理解为经过一个路由器（转发）

5.2.1 优化原则

最优化原则optimality principle：p281/295

汇集树sink tree：所有的源到一个指定目标的最优路径集合，目标为树根

5.2.2 最短路径算法

Dijkstra算法

5.2.3 泛洪算法flooding

P284/298

不适合大多数情况，只适合广播、军事、科学验证等特殊场合。

5.2.4 距离矢量算法distance vector routing

P285/299 一种动态路由算法

该算法也称分布式Bellman-Ford路由算法，是RIP协议的一部分，曾用于Internet

算法基本原理：除了路由表之外，每个路由器维护一个距离矢量表以记录它到其他所有路由器的“距离”，每隔一段时间向其所有邻居发送该距离矢量表，通过此种方式更新传播路由状况。

①无穷计算问题count-to-infinity

对于距离矢量算法，路由表更新至更短更优的路径很容易，但是更新至更差的路径尤其是路由器宕机，路由表更新极慢。需要逐步计数至无穷大方能发现宕机和性能变差的区别。

带有染毒逆向的水平分裂法：一种尝试解决无穷计数的技术

5.2.5 链路状态路由link state routing

IS-IS和OSPF：链路状态路由算法的变种，应用最广

中间系统到中间系统链路状态协议Intermediate System-Intermediate System

开放最短路径优先Open Shortest Path First

过程步骤：

①发现邻居

路由器之间通过发送特殊的数据包以确认彼此地址和信息。对于LAN链路（多个路由器通过一个广播链路连接），可将LAN链路视为一个虚拟节点或虚拟路由器以简化拓扑结构。由其他的实际路由器代其执行路由协议。

②设置链路成本

即度量和设置每条链路的“距离”或其他成本

③构造链路状态包

链路状态包包括发送方标识符（源路由器标识符）、序号（序号越大的状态包越新，序号大的状态包更新序号小的）、年龄（从某个数值随时间减少，至0时状态包过时丢弃，相当于状态包有效时间）、邻居及延迟表等。每隔一段时间或特殊时机更新一次链路状态包

④分发链路状态包

泛洪法分发+其他优化机制

P289/303

维护一个状态包缓冲区，和一套标志位表，以增加算法强健性

⑤计算新路由

获得所有链路状态数据包之后执行最短路径算法以获得最优路径

5.2.6 层次路由

网络增大，每个路由器不可能为所有其他路由器维护表项。路由器按照等级和区域region进行划分成组，如同电话网络。

理论证明，N个路由器网络的最优层数为lnN。每个路由器所需的路由表现是elnN。

5.2.7 广播路由

多目标路由multidestination routing p293/307

逆向路径转发reverse path forwarding p294/308

生成树spanning tree：网络的一个子集，包括所有路由器，但是没有任何环路。汇集树即生成树中的一种。

5.2.8 组播路由multicasting

组播树：即组播生成树，一般的生成树用于全体广播，组播树是针对每个组的成员的广播生成树。组播树的创建一般都是通过生成树修剪得来

密集分布：接收者遍布网络的大部分区域，适合使用修剪式组播树的算法，如下两种

组播MOSPF，Multicast OSPF：使用链路状态协议获取网络拓扑结构，由每个路由器自行创建组播树（汇集树->组播树）

距离矢量组播路由协议，DVMRP，Distance Vector Multicast Routing Protocol：使用距离矢量路由算法的组播协议

稀疏分布：大部分网络都不属于组，适合使用基于核心树或共享树的技术core-based trees

共享树：一个组只有一个组播树，存在一个核心core或会聚点rendezvous point，所有的发送源都必须先发送数据至核心再传输到其他组内路由器（部分优化算法允许发送至核心的过程中，若存在和共享树同链路的情况，可同时向下转发）。共享树技术节省存储开销、计算开销等，但是不能保证最优传输效率。

协议独立组播PIM，Protocol Independent Multicast：使用共享树技术的协议

5.2.9 选播路由

选播Anycast：数据包被传递给最近的一个某组成员，对于某些应用这个技术很有用，如域名系统

选博路由方案由距离矢量和链路状态算法适当转化即可。

5.2.10 移动主机路由

移动主机mobile host：漫游、移动过程中的网络设备

家乡地址，家乡位置，home address，home location：移动路由中的一种思想，无论移动设备到何处，其固定有一个家乡地址用以建立发送和接收的连接。进一步只需单方面解决家乡代理home agent和移动设备的连接。

转交地址care of address：移动设备在外地获得的本地网络地址，一般不同于其旧地址，如新的IP地址

隧道tunneling：一种跨IP版本协议（v4和v6）传输的机制

三角路由triangle routing：家乡代理、发送方、移动接收方组成的连接，更复杂的发送方也可能是移动的

IPv6流动模型：一种移动路由算法，用于互联网和UMTS蜂窝网络中的IP部分

5.2.11 自组织网络路由

自组织网络Ad hoc：或称移动自组织网络MANET，Mobile Ad hoc NETworks。每个节点使用无线通信，并同时承担路由器和主机的任务。

Ad hoc按需距离矢量AODV，Ad hoc Ondemand Distance Vector：流行的自组织网络路由算法

动态源路由DSR，Dynamic Source Routing：另一个按需路由方案，Johnson，2001

贪婪边界无状态路由GPSR，Greedy Perimeter Stateless Routing：另一种路由方案，所有节点知道其地理位置，无需路由计算，只需简单的正向传递，Karp & Kung，2000

以下以AODV为例：

①路由发现

按需on demand：只有当需要建立连接才找寻路径或路由计算

ROUTE REQUEST路由请求：用以找寻到目标路由器的包

ROUTE REPLY路由应答：目的地路由器返回发送源的包

②路由维护

每个节点周期性广播HELLO消息，每个节点都观察邻居的动向，及时通报自己的其他活跃邻居那些失效的路由。

AODV为每条路由设置了序号以解决无穷计数问题。实际上，这种技术更像是面向连接的。

自组织网络的路由策略必须实践，因为这种网络类型变化繁多。

## 5.3 拥塞控制算法congestion

拥塞崩溃congestion collapse：由于过度拥塞导致的数据包延迟至允许寿命之外而不得不全部丢弃，因而造成网络实际吞吐量的大幅降低

实际吞吐量goodput：网络传递可用（非过时）数据包的速率

拥塞控制和流量控制的区别：前者是针对网络大局的协调，基于网络处理能力，后者是发送和接收方之间的协调，基于发送速度和接收方处理能力

5.3.1 拥塞控制的途径

由上至下时间尺度降低

慢（预防性）

网络供给provisioning：每隔几个月及时升级主要链路和路由器

流量感知的路由traffic-aware routing：根据流量模式量身定制路由，如每日时间变化等

准入控制admission control：控制能够建立的网络连接数（一般针对虚电路网路）

流量限制：反馈给发送源降低流量，必须及时反馈

负载脱落load shedding：选择性策略性的丢弃数据包，同时防止拥塞崩溃

快（反应性）

5.3.2 流量感知路由

早期流量感知路由算法通过将流量负载因子增加至链路权重，以利用已有的路由算法实现。但这种方法会导致路由计算不稳定，存在剧烈变化和波动的不收敛。

流量工程traffic engineering：Internet路由协议对流量的控制

5.3.3 准入控制

流量说明：流量很难用简单的模型描述，但是又必须将其描述以供网络分析是否可以准入

5.3.4 流量调节

拥塞避免congestion avoidance

监测拥塞情况的最佳参数是缓冲区内的排队数据包情况，延迟和数量等。

指数加权移动平均EWMA，Exponentially Weighted Moving Average：常用的路由器更新排队延迟的方法，p307/321

①抑制包choke packet

随机抽取排队数据包向其源主机发送抑制包

②显式拥塞通知ECN，Explicit Congestion Notification

路由器向每个出境数据包添加拥塞信号，由接收方向发送方发送通知（可附加于应答包）。

ECN是对早期拥塞信令协议的细化，尤其是对二进制反馈方案的细化（Ramakrishnan & Jain, 1988）。该方案曾用在DECNET体系结构。

③逐跳后压

抑制包回到发送方的过程中经历的每个路由器都产生抑制作用，每个路由器都为该路径的数据包提供更多缓冲区，减少出境。

5.3.5 负载脱落load shedding

葡萄酒策略：对于旧数据包更重要（没有旧的新的也没用），选择丢弃新包，如常规文件

牛奶策略：对于新数据包更重要的场合，选择丢弃旧数据包，如实时媒体

MPEG：一种视频压缩算法，只保留定期间隔的全帧，其余的部分使用和前一个全帧的差异帧

通过标记数据包的优先级，可以有效的支持负载脱落

①随机早期检测RED，Random Early Detection

路由器监视队列长度均值，超过阈值后随机选择丢弃数据包，被丢包的发送方可自行察觉以减缓速度。适用于ECN不可用的场合，否则其仍是更佳选择。

## 5.4 服务质量

过度配置overprovisioning：提供良好服务质量的最简单解决方案，即提供足够的容量

确保服务质量必须解决的问题：  
•应用程序需要网络什么样的质量？  
•如何规范进入网络的流量？  
•为了保障性能如何在路由器预留资源？  
•网络能否安全接受更多流量？

5.4.1 应用需求

流flow：从一个进程到另一个进程的所有数据包

服务质量QoS，Quality of Service：是一种网络服务需求，其决定参数包括带宽、延迟、抖动、丢失。实际中需求包括应用需求和网络需求，应用需求高于网络需求，因为应用程序可以在上层完成一些工作使得网络传输可以不必尽善尽美。当然，没有适当的网络需求，应用程序难为无米之炊。

各种常见应用程序对网络服务质量需求p312/326

抖动jitter：延迟的变化（即标准方差）或者数据包到达时间的变化

ATM网络对于不同类别的QoS的分类p313/327（可广泛适用）

5.4.2 流量整形traffic shaping

定义：调节进入网络的数据流的平均速率和突发性所采用的技术

服务等级约定SLA，service level agreement：客户和服务提供者之间的约定

流量监管traffic policing：对于一个流进行监测，看其是否按照约定

①漏桶和令牌桶

漏桶算法leaky bucket algorithm：p314/328

令牌桶算法token bucket algorithm：p314/328

漏桶算法使得流量绝对不会超过某个值，令牌桶使得流量可以出现峰值但一定时间尺度上总流量不会超过某个值（亦即均值不超过），桶都作为一个整形器存在

桶算法的实现需要离散，仍然依靠机器时钟周期性实现代表其连续性。

5.4.3 包调度packet scheduling

数据包调度算法：在同一个流的数据包之间以及在竞争流之间分配路由器资源的算法，适用于虚电路网络

流占用的潜在资源：带宽、缓冲区、CPU周期

先入先出FIFO，First-In First-Out：也称先来先服务FCFS，First-Come First-Serve，一种最基本的队列实现的包调度算法思想。这种算法丢弃后来的超额数据包，称作尾丢包tail drop

公平队列fair queueing：p317/331 所有的流轮流发送数据包或更公平地发送字节（以字节数量计量发送数据包的多少）

加权公平队列：WFQ，Weighted Fair Queueing。部分流获得相对较多带宽的设定

赤字循环deficit round robin：一种近似算法，用于WFQ中将数据包编排至一个有序队列

5.4.4 准入控制

QoS路由QoS routing：由于保证服务质量而产生的路由行为，不一定选取最佳路由路线，而是保证服务质量第一

流规范flow specification：一组能够精确描述流的参数

使用令牌桶和WFQ的流规格说明法和资源预留技术p321/335

5.4.5 综合服务integrated service

定义：流式多媒体体系结构

①RSVP——资源预留协议Resource reSerVation Protocol

由RFC2205-2210文档标准化

5.4.6 区分服务differentiated service

定义：也称为基于类别的服务质量class based，之前介绍的都是基于流的服务质量算法。由RFC2474、2475标准化。区分服务即按照数据包的服务类别或等级进行对应质量的服务，不再提前对每个/整个流进行预留和协商。后者实际上在大型网络中难以实现。

单跳行为PHP，per hop behaviors：定义服务类别的方式，实际上只是数据包在每个路由器上的待遇

①加速转发expedited forwarding

定义：一个与网络无关、用于网络之间的转发数据的服务类别。RFC 3246

②确保转发assured forwarding

定义：一种管理服务类别的方案，RFC 2597。规定4种优先级和3种丢弃数据包概率，共12种服务类别。

一种可能的实现p326/340

## 5.5 网络互联internetwork

5.5.1 网络如何不同

网络层之上的一些不同网络之间的差异举例p327/341

5.5.2 何以连接网络

物理层设备：中继器、集线器

链路层设备：网桥、交换机（针对同网络或单一网络间的传输，使用MAC地址）

网络层设备：路由器（针对不同网络之间的互联，使用IP地址）

多协议路由器multiprotocol router：可以处理多个网络协议的路由器，可翻译协议

5.5.3 隧道tunneling

定义：网络互联中，若发送和接收方使用相同的网络协议，途中经历的其他类型的网络可以简单地将整个数据包进行封装，形似过隧道。

覆盖网络overlay：一种网络技术或协议被另一种技术分割或隔离开，形似覆盖。

隧道技术无法使得数据包到达“隧道”代表的网络中的目的地，因为两种网络技术不同，转换需要额外处理。但这种限制正是VPN的优势。

5.5.4 互联网路由

互联网路由使用2级路由算法，一个域内intradomain或内部网关协议interior gateway protocol，一个域间interdomain或外部网关协议exterior gateway protocol。

网关gateway：路由器的旧称

域内协议每个网络可能不同，但是域间协议必定相同才能互联。Internet的域间路由协议也称为边界网关协议BGP，Border Gateway Protocol、

自治系统AS，Autonomous System：每个独立运营的网络的称呼，不一定只是一个单一网络。

路由政策routing policy：由社会、政治、法律等非技术因素组成的条件

5.5.5 数据包分段

路径最大传输单元MTU，Path Maximum Transmission Unit：数据包尺寸的限度

数据包分段的策略：第一种是每次遇到较小数据包限制的网络时都进行分段，并在最后离开该网络时重组，但下次遇到小数据包网络时重复整个过程；第二种是只有在遇到更小的数据包限制网络时进行分段，直到达到最终的目标主机才重组。IP采用第二种策略

路径MTU发现path MTU discovery：现代互联网使用的代替数据包分段的技术，数据包分段势必增加了性能损耗和路由算法复杂性，因此该技保证数据包在发送的途中不必分段。P334/348

## 5.6 Internet的网络层

驱动Internet协议设计的原则：RFC1958, p335/349

①保证工作——实现了通信功能完全验证之后方可确立设计或标准，即能够工作优于一切

②保持简单——只要一项特性并非绝对必要，就没必要

③明确选择——

④模块开发——协议栈思想

⑤期望异构性——设计简单灵活

⑥避免静态选项和参数——协商解决参数

⑦寻找好的而不是完美的设计——完美只针对绝对的需求，不完美的能针对大多数需求即足够

⑧严格发送，宽容接收——

⑨考虑可扩展性——分散资源利用

⑩考虑性能和成本

一级网络Tier 1 networks：Internet骨干网backbone中最大的核心，和每个其他骨干网相连。一般也由大型ISP公司运营，覆盖全世界主要地区。

Internet服务提供商ISP，Internet Service Provider：连接至骨干网上的执行机构，负责家庭企业和其他区域网络的接入服务。

Internet协议，IP：提供一种尽力而为best-effort的服务的协议，其作用正是通过网络层将所有互联网组成网络连接在一起。

5.6.1 IPv4协议

IP数据报包括两部分：头和正文

Big-endian和little-endian：是字节序的选项。前者表示高序字节优先，后者相反。IP协议采用前者，部分计算机采用后者。

头的格式和结构：p337/351  
版本Version——  
IHL——指明头的长度  
区分服务differentiated services——最低2位用来发送显示拥塞通知  
总长度Total length——  
标识Identification——用于表明分段数据包属于哪个整体数据报，同报的标识相同  
DF，不分段标志位Don't Fragment——指示路由器不许对包分段  
MF，更多分段标志位More Fragments——指示当前包是否为所在报的最后一段  
分段偏移量Fragment offset——  
生存期Time to live——时钟TTL  
协议Protocol——指明数据包下一步需要进行的传输协议  
头校验Header checksum——  
源地址和目标地址Source address & destination address——  
选项options——额外的IP协议选项p339/353

5.6.2 IP地址

IPv4采用32位地址，一个地址指向一个网络接口，而非一个主机。因此路由器和主机都可以有多个地址，只是一般主机都只有一个网络接口。

①前缀prefix

定义：IP地址中高位的可变长网络值即为前缀，同一网络上的所有主机其网络地址的网络部分相同。IP地址由前缀即网络部分和主机部分两部分组成，但对于每个网络前缀长度可能不同。

子网掩码subnet mask：用于和IP地址进行AND操作以提取出其网络部分或前缀的序列，常见的为255.255.255.0，对应24位网络值。这是一种对于给定IP地址判断其前缀长度的方法。

前缀书写方法：如128.208.0.0/24表示24位前缀。

注意：前缀长度不一定是8或16或24中的其一，因此第二种前缀表示法的前缀IP地址只是该网络块的最小IP地址。为0的部分不一定可取至255，非零部分也并非在该网络中不存在其他取值的地址。实际上，IP地址的点分十进制只是一种不绝对直观的写法。

IP地址的层次性使得路由器不必储存所有的IP地址，而只需储存网络地址。当数据包到达具体网络后，可自行分发。

②子网

Internet域名和地址分配机构ICANN，Internet Corporation for Assigned Names and Number

子网subnet：（新定义）一个原本同网络值的大型网络进行划分之后形成的功能上独立的网络部分。这个行为也称子网划分subnetting。

注意：子网划分不是随意的，如果限定了子网的大小，那么其网络部分的位数也被限定，亦即前缀或网络部分的取值也被限定。

子网彼此之间通过不同的子网掩码进行区分，主网络路由器接收到一个数据包后，只需使用不同的子网掩码检验数据包IP地址即可找到匹配的对应子网的前缀。

③CIDR——无类域间路由Classless Inter-Domain Routing

默认自由区default-free zone：准确翻译应该为无默认区域，即Internet的核心路由器所在的骨干网络。边缘和地方组织路由器往往存在路由的默认选项——发给ISP，但核心路由器必须处理IP地址寻址。

路由聚合route aggregation：即子网划分的对立面，将原本不同的网络聚合成同属一个前缀的名义网络。主要用于减少主干路由器存储的表项，当然能进行路由聚合的网络其地理位置宜接近。

超网supernet：子网的对立体

CIDR：路由聚合和子网划分的统称，目的就是为了减少全局路由表的大小

最长匹配前缀longest matching prefix：IP地址的分发不一定严格按照地理位置，当存在多个前缀表项均匹配时，路由器选择最具体的地址转发，亦即前缀最长的表项。P344/358

实际使用的复杂的路由前缀匹配算法Ruiz-Sanchez 2001

VLSI

④分类和特殊寻址

分类寻址classful addressing：1993年以前，IP地址分为5类，p345/359

分类寻址虽然操作和算法较最长匹配前缀简单，但是由于互联网规模扩大和IP地址起初设计局限的冲突，这种方式已经不适实际应用。

特殊的IP地址：p346/360  
0.0.0.0——意味着当前网络的当前主机；网络号全0可用于机器不知当前网络号的情况，只填入主机号即可向本地主机发送数据，但须知网络掩码  
255.255.255.255——用来标识本地网络中的所有主机；每个网络地址中主机字段全1的地址都是用来向全部网络主机广播的，但实际上多数情况因为安全原因而禁用  
127.x.y.z——保留给回环测试

⑤NAT——网络地址转换Network Address Translation

NAT是短期内解决IP地址短缺问题的技术，长期解决仍需普及IPv6，由RFC 3022描述，Dutcher 2001。

NAT技术保留了3个IP地址范围以供网络内部私用p347/361。这三个范围内的IP地址绝对不会出现在Internet之上，因此每个网络内部如何使用这些地址无所谓。每个出境数据包的内部IP地址会被转换为公有的IP地址，而后在出境NAT盒子上留下一个索引项用以区分内部IP地址。在数据包中，索引项占据TCP或UDP的源端口字段。P348/362

NAT盒子NAT box：一般至于网络出境路由器和ISP路由器之间用以执行地址转换，同时一般也和防火墙整合成一个设备，或与路由器和ADSL调制解调器整合。

NAT使用带来的问题：  
①违反了IP结构模型——不是每台机器都有唯一的IP地址  
②打破了Internet的端-端连接模型——亦即数据发送和接收的两端不对等，必须由一方发起后，建立地址映射，另一方才能发送并能够被反向接收。除非使用特殊的配置技术，或NAT穿越技术NAT traversal。  
③改变了Internet的特性为面向连接  
④违反了协议分层规则  
⑤限制于用户使用TCP或UDP传输协议  
⑥标准文件传输协议FTP等协议不可用，除非另行增添补丁

NAT的一个优势是，能够默认阻止未经请求的入境数据包，因为没有建立映射的连接NAT无法确定数据包该传输给哪个主机。

5.6.3 IPv6协议

IPv6由RFC 2460-2466定义

CLNP：一种由OSI设计的网络层协议，弃用

简单Internet协议+：SIPP，Simple Internet Protocol Plus；未被选中为IPv6标准前的非正式Internet协议称呼

IPv6不与v4兼容，但是兼容辅助性协议：TCP, UDP, ICMP, IGMP, OSPF, BGP, DNS

①主要的IPv6头结构：p352/366  
版本——  
区分服务——  
流标签flow label——用以结合数据报和虚电路网络的设计  
有效载荷长度payload length——  
Next header——指明此头后还有哪种扩展头，若此头为最后的，则指定上层传输协议  
跳数限制hop limit——类似于IPv4的TTL，实际上二者实现相同，IPv6的名称更确切  
源地址和目标地址

IPv6地址的书写方式：16字节地址分成8组，每组使用4个16进制数字表示，用冒号隔开  
例——8000:0000:0000:0000:0123:4567:89AB:CDEF  
其余书写规则——p353/367

②扩展头extension header

IPv6扩展头p354/368  
逐跳头hop-by-hop header——存放沿途所有路由器必须要检查的信息，只用于巨型数据报jumbogram（超过64kb）的发送  
目标选项扩展头destination options header——用于只需被目标主机翻译的字段  
路由扩展头——类似松散源路由  
分段扩展头fragment header——IPv6中只有源主机才能进行分段，降低路由器压力  
认证扩展头authentication header——  
加密安全有效载荷扩展头encrypted security payload header——

③争论（IPv6的特点）

增长地址长度

跳数限制hop limit没有增加

最大数据包长度64kb，巨型数据包使用扩展头完成

没有网络层校验和

5.6.4 Internet控制协议

IPv4的辅助控制协议：ICMP，ARP和DHCP

IPv6的辅助控制协议：ICMP和DHCP的类似版本，NDP邻居发现协议（ARP的等价协议）

①ICMP——Internet控制消息协议Internet Control Message Protocol

含义：该协议针对路由器处理数据包过程中发生各种意外情况，或纯粹的测试目的的情况，路由器向源端报告问题的规程和消息格式。所有的消息实际上都是一个既定义的IP数据包。

主要的ICMP消息类型及其含义：p358/372  
目的地不可达DESTINATION UNREACHABLE  
超时TIME EXCEEDED  
参数问题PARAMETER PROBLEM  
源抑制SOURCE QUENCH——用于拥塞控制，但现今Internet的拥塞主要由传输层完成  
回显ECHO 回显应答ECHO REPLY——用于ping工具  
请求时间戳TIMESTAMP REQUEST 应答时间戳TIMESTAMP REPLY  
路由器通告ROUTER ADVERTISEMENT 路由器恳求ROUTER SOLICITATION

②ARP——地址解析协议Address Resolution Protocol

含义：该协议用于IP地址和其他本地网络地址之间的转换、映射与解析，RFC 826定义ARP。如在以太网内部，数据以以太网帧形式发送，需要解决以太网地址和IP地址之间的解析才能实现IP技术。大体上，ARP协议包括发送方请求所有以太网络上的主机自检IP地址，和目标主机以自身以太网地址应答。

免费ARP，gratuitous ARP：执行ARP协议过程中的一个术语，指每次更新ARP地址解析映射的动作不一定由主动发送请求等动作发起，传递其他路由器或主机的ARP信息时可顺带更新它们的相关信息。

默认网关default gateway：一个把控一个本地网络外部流量的路由器，其一般具有本地网络的最低IP地址。

ARP代理，proxy ARP：不在同一本地网络的发送和接收方使用一个中间路由器代替进行ARP解析。从一方看来，执行过程就像中间路由器即另一方。这只是ARP实现的一种方式。

③DHCP——动态主机配置协议Dynamic Host Configuration Protocol

含义：即动态为本地网络内的流动主机分配IP地址的协议，由RFC2131和RFC2132描述。

DHCP不仅仅可以配置IP地址，还包括其他各种网络信息配置。其前身是RARP和BOOTP。

5.6.5 标签交换和MPLS

多协议标签交换MPLS，MultiProtocol Label Switching：一种非传统的面向连接的网络的技术。具体实现方法和虚电路网络类似（如标签的作用），但是可用于各种IP或非IP网络。其目的是更快速的转发数据包，以及适合服务质量的需求。由RFC3031和其他文档描述。

MPLS技术使用的帧格式：p362/376；MPLS介于网络层和链路层之间，有自己的信息头，也称作2.5层协议。

标签交换路由器LSR，Label Switched Router：能够处理MPLS增强型数据包的路由器

路由器的转发和交换：  
转发——读表找寻与目标地址最佳匹配项  
交换——使用标签作为索引来查询（更简单迅速）

标签边缘路由器LER，Label Edge Router：MPLS网络边缘用于将其他数据包转换为MPLS数据包的设备

转发等价类FEC，Forwarding Equivalence Class：MPLS技术中，将多个流合并成为一组，并配发同一个标签。不必关心接收方是否可以区分，因为不同于虚电路网络，此网络的数据包一般还有其他的独特地址信息。

标签栈stack of label：每个数据包的MPLS标签可以不止一个，由多层的标签指导数据包一个阶段一个阶段的路由。

MPLS在设置标签转发表的阶段不会涉及用户，因为它与虚电路的不同之处也在于，它创建的路径并非是针对某个用户的虚电路。

MPLS的细节复杂，其他信息：Davie & Farrel 2008, Davie & Rekhter 2000

5.6.6 OSPF内部网关路由协议

域内路由算法intradomain routing：一个组织或网络内部的路由算法。域内路由协议也称内部网关协议interior gateway protocol。

域间路由interdomain routing：不同网络之间的路由协议，也称外部网关协议exterior gateway protocol。

路由信息协议RIP，Routing Information Protocol：早期的域内路由协议，距离矢量设计

开放最短路径优先OSPF，Open Shortest Path First：基于一个分层网络（骨干、区域、边界等）的拓扑图的链路状态（路由器之间彼此交换网络拓扑信息，自行计算最优路径）协议，开放open指的是算法必须公开发表，而非某私有方案。原理见p366/380

中间系统到中间系统IS-IS，Intermediate-System to Intermediate-System：一个ISO标准协议，同OSPF非常类似。RFC2328描述。OSPF更多用于公司网络，IS-IS更多应用在ISP网络

等价成本多路径ECMP，Equal Cost MultiPath：OSPF不仅仅选取所谓的最短路径，可能会将数据分流至所有同样短的路径上实现负载均衡。

区域area：对于大型AS，OSPF将其划分成编号的区域。区域不能相互重叠，但是不必完全面面俱到，即该AS中不是所有的路由器都必定属于某个区域。区域彼此间是独立而且不透明的。

内部路由器internal router：全部属于一个区域的路由器

骨干区域backbone area：也称0号区域，其中的路由器称骨干路由器backbone router。每个AS中的所有区域都和骨干区域相连。

区域边界路由器border router：每个连接到两个或更多区域的路由器，且必须是骨干区域的一部分。

存根区域stub area：区域只有一个边界路由器，所有出境路由必定经此

AS边界路由器AS boundary router：负责通往其他AS的外部路由，其他AS不一定直接与该AS相邻。

邻接adjacent：邻接和邻居的概念不同，同一个LAN上的路由器均为邻居，这是一种地理或物理上的事实。邻接的前提是邻居，另一个必要条件是交换链路信息的事实。也就是说，OSPF中有邻居路由器但是彼此间不交换路由信息。

指定路由器designated router：OSPF中采用类似LAN核心路由器的策略，它与LAN中其他所有路由器邻接，其他路由器之间只是邻居而不邻接。同时存在一个备份指定路由器以转接和必要时替换。

OSPF正常工作流程中的5类消息：p368/382  
HELLO  
LINK STATE UPDATE  
LINK STATE ACK  
DATABASE DESCRIPTION  
LINK STATE REQUEST

5.6.7 BGP——外部网关路由协议

外部网关路由协议exterior gateway protocol：AS之间所需使用的另一个不同于内部网关协议的协议。原因主要是外部路由需要考虑大量地缘政治等因素的需求，而非仅仅是最低成本等。

中转服务transit service：一种服务类型，流量只是经由服务提供者中转，其非流量的源也非目的地。

对等传输peering：不同于中转服务的一种对等服务，双方互相为彼此流量的源或目的地。

Internet交换点IXP，Internet eXchange Point：用以连接至其他AS或Internet站点，亦即不同ISP经营的网络之间的连接点，用以彼此交换流量，分布于世界各地城市。

多穴寻址multihoming：一个AS连接到多个ISP，以望提高Internet连接可靠性

边界网关协议BGP，Border Gateway Protocol：是Internet中采用的外部网关路由协议，是一种距离矢量协议，但是更确切的说是一种路径矢量协议。由RFC4271定义规范。

路径矢量协议path vector protocol：BGP不仅仅维护到达每个目的地的成本，还必须跟踪到达每个目的地所使用的路径，即后者相当于覆盖在成本要求之上的具体路由要求。

由于是类似距离矢量协议，很难完全杜绝无穷计数和满收敛的问题，实际情况还有中转循环。

BGP路由通告传播：p371/385

内部BGP，iBGP，internal BGP：常规BGP的变种，一个域内协议用于在ISP内部穿越式传输BGP路由通告。常规的BGP也称外部BGP，eBGP，external BGP。

BGP常见路由策略：  
对等网络优先  
短路径更好作为默认规则  
优先选择具有ISP内最小成本的路由——也称为提前退出early exit或热土豆路由hot-potato routing

5.6.8 Internet组播

IP使用D类地址用于组播（ABC类现在已不可用或不区分），共有28位用于标识组，即可同时并存250万个组。

224.0.0.0/24范围内的地址保留用作本地网络组播。其中具有特殊含义的例子：p373/387

Internet组管理协议IGMP，Internet Group Management Protocol：RFC3376，指明互联网中如何确定主机或系统所在组的规程协议，包括组播路由器查询主机信息和主机应答等细节。

协议独立组播协议PIM，Protocol Independent Protocol：在同一AS内的组播路由，密集模式下采用逆向路径转发树，稀疏模式采用核心树。

特定源组播PIM，Source-Specific Multicast PIM：PIM的变体，专用于只有一个发送方的情况

不同AS内的组播路由需要借助BGP或隧道扩展。

5.6.9 移动IP

家乡代理home agent

转交地址：移动主机在外地网点获得的新IP地址，若无空余IP地址，也可使用外地代理

IPv4的移动IP解决方案由RFC3344给出。大致细节上，移动主机的原始IP地址由家乡代理接管，移动主机在转交地址上和家乡代理通信，即一种三角路由。

入口过滤ingress filtering：一种安全措施，路由器会丢弃那些看似不正确地址的数据包。

移动IPv6由RFC3775定义。该协议中，除了最初的数据包遭受三角路由，其余的数据包可直接在移动主机和其他发送方之间直接传输。

IPv6的网络移动由RFC3963定义。网络移动指的是一个网络，如LAN的移动，比如飞机上的内置无线网络。

# 第6章 传输层

网络层使用数据报或虚电路技术为端到端通信提供数据包交付服务。传输层在此基础上，将数据传递服务建立在两个计算机的进程之间。传输层为应用层使用网络提供了抽象的模式。

## 6.1 传输服务

6.1.1 提供给上层的服务

传输实体transport entity：传输层内完成传输任务的硬件和软件。

传输层很多特点特性上与网络层类似（如有无连接的服务类型），可以说是直接构架在其上的层。之所以这样设计是因为传输层在主机上执行，用户有一定的操作性，而网络层实现在路由器上，不属于用户的权限范围。网络层的一些问题若无法被妥善处理，传输层仍然有机会弥补。另外，网络的技术有很多种，传输层将网络层的技术细节和区别隐藏，应用层调用传输服务时不必在意网络技术的细节。

传输服务提供者transport service provider：1-4层

传输服务用户transport service user：4层之上

6.1.2 传输服务原语

服务原语：服务原语在形式上是一组操作或动作关键词，它们在整体上组成了一个完整的服务或事务。服务原语只是描述一个事务过程中各参与方所需执行的动作的关键词。

段segment：传输层实体之间发送的消息名称，对应网络层的包，链路层的帧

传输协议数据单元TPDU，Transport Protocol Data Unit：段的旧称呼

6.1.3 Berkeley套接字

套接字socket：广义上是一种传输模型，狭义上是该模型中的一个数据结构，用以处理并发的多个连接。套接字模型或API广泛用于Internet程序设计（传输服务）中，尤其是UNIX系统。Windows的套接字风格API是winsock

TCP的套接字原语p386/400

可靠字节流reliable byte stream：套接字API与TCP协议结合的面向连接服务。套接字传输模型或API可以和各种协议搭配，面向连接或无连接，消息流或字节流。

数据报拥塞控制协议DCCP，Datagram Congestion Controlled Protocol：带有拥塞控制机制的UDP版本。

流控制传输协议SCTP，Stream Control Transmission Protocol

结构化流传输SST，Structured Stream Transport

6.1.4 套接字编程实例：Internet文件服务器

可用来了解套接字模型的过程p388/402

## 6.2 传输协议的要素

传输层和数据链路层的协议需要考虑的课题类似，如重复/错误控制、顺序性、流量控制等。使用的算法也有类似之处，如使用序号编排数据段（帧）、滑动窗口等，但是两者的层面不同，程度不同，只能说道理相像。

6.2.1 寻址

端口port：传输层是面向主机的进程的，网络层上的端点是主机和路由器及其地址，传输层需要为每个主机上需要传输服务的进程分配一个传输地址，即端口

传输服务访问点TSAP，Transport Service Access Point：传输层的端点end-point或地址，端口的书面称呼

网络服务访问点NSAP，Network Service Access Point：网络层的端点，即网络层地址，IP便是其中一例

端口映射器portmapper：端口映射器也是一个绑定在某个端口的进程，只是这个端口号一定众所周知。外部连接到来时若连接的服务不是知名的公共服务而是某个自定义服务，且不知道其临时端口号，可首先与端口映射器连接，通过服务名称和端口号的映射，端口映射器可以将外部连接转至它所想要连接的端口。其作用机理类似电话系统的查号员。

初始连接协议initial connection protocol：对于一组不太常用的服务端口，没必要对每一个每时每刻都进行监听。可使用一个特殊的进程服务器process server来代理一组不太常用的服务器，同时监听它们所有的端口。当有连接请求到来时，进程服务器创建并转接给所求的服务器。UNIX中的进程服务器或守护进程称为inetd（Internet daemon）

服务器进程：服务器可以指代硬件，即一种特殊功能的计算机。也可以指代运行其功能的软件，即一个进程。因此道理上说，一个计算机其上拥有多个服务器即拥有多个服务进程。之所以称为服务进程，是对应于客户进程而言的供求关系描述，因此这种关系不是固定的。

6.2.2 连接建立

为了让传输层的连接能够建立，网络层的数据包必须要有生存期限制等机制。在网络层看来，这种机制是以优化网络层性能为目的（避免数据包滞留网络内）。传输层而言，这种机制可避免重复的请求多次前后出现而难以区分真实情况。

最大数据包生存期：在数据包有生存期的限制下，一个网络中的数据包在各种情况下的最大生存期。一般是一个保守常数或经验估计数值，Internet中习惯取120s。最大数据包生存期的某个不太大的倍数，可以估计成为某个数据包的痕迹（包括其确认ack）在网络中的最大存留时间T。

日时钟time-of-day clock：顾名思义，就是一个标记日常时间的实际时钟。一般算法中的时钟，准确的讲是计时器（秒表）或倒计时器，可以标记时间段。另外一个区别是，日时钟在主机关机（或停机崩溃等）后也依然运行。

TCP建立连接使用当前日时钟作为当前连接的初始序号，而后发送的段的序号自行递增即可，不过不能发送得太快超过时钟（导致若重新建立相同连接可能出现相同序号），也不能太慢导致序号回绕。这样可以保证机器崩溃后序号不会重复，也无需等待T时间以确保原数据段无残留。

三次握手three-way handshake：建立连接时采用的协议，要求一方检查连接请求是否的确是当前的。之所以用在建立连接时，因为连接过程中采用的滑动窗口协议可以记住序号，但连接建立初始没有记忆。P398/412，Tomlinson，1975

防止序号回绕PAWS，Protection Against Wrapped Sequence numbers：由RFC1323定义，是TCP中的一个面向越来越快速的连接的扩展机制。使用时间戳来扩展32位序号，以免序号回绕。

6.2.3 连接释放

连接释放或终止连接有两种方式：非对称释放和对称释放；前者任何一方终止会话都会导致连接完全结束，后者双方的终止行为彼此独立，即一方释放连接后还可以单向接受另一方的发送。

两军对垒问题two-army problem：对称断开连接的问题比喻，如果连接两端不能独立决定是否断开连接，则问题是无解的。p400/414

半开连接half-open：即对称释放中，只有一方释放了连接的情况。对于这种情况可以采用检测机制处理，若一段时间内没有检测到活动，则自动断开连接。

连接释放的问题需要传输层用户（应用）来参与解决，亦即根据应用层的需求和事务逻辑可以针对性制定出有效的断开连接部分协议，而试图制定出普适的协议很难。

6.2.4 差错控制和流量控制

链路层：  
针对的是单条链路上的差错，不包括路由器内部的损坏。  
无线链路传输速度快，距离近，使用小窗口的停等式协议即可  
有线链路错误率低，可以忽略链路层差错控制交给传输层

传输层：  
端到端的整体校验。链路层的检错机制的存在对于提高性能也重要，否则每次错误必须整条路径重传。  
连接路径较长，使用大窗口的控制协议

6.2.5 多路复用

逆向多路复用inverse multiplexing: 一般的主机只有一个网络地址，而可能存在多条传输连接，若一个传输连接使用多条网络路径和地址则为逆向多路复用，SCTP就是一例

6.2.6 崩溃恢复

从第N层崩溃中的恢复工作只能由第N+1层完成，当且仅当高层保留了问题发生前的状态信息。

## 6.3 拥塞控制

Internet的拥塞控制严重依赖于传输层。一般意义上的网络，拥塞控制是传输层和网络层的共同责任。

6.3.1 理想的带宽分配

带宽分配方法的标准：

①效率和功率p410/424

功率=负载/延迟 Kleinrock，1979

功率存在最大值，对应的负载即网络的最大理想负载程度，或最佳可用带宽。这个带宽低于网络实际带宽，余下的带宽留给突发流量。

②最大-最小公平性max-min fairness

含义：如果分配给一个流A的带宽在不减少分配给另一个流B的带宽的前提下无法得到进一步增长，且B的带宽不比A的更大，就不给这个流更多带宽。

分配方法：所有的流速率从0开始增加，遇到瓶颈就停止，直到所有的流遇到各自瓶颈。

这只是公平性的一种实现方式（带宽），另一种是基于网络路径长度的，路径较长的流同等流量或带宽下占用网络资源更大。

公平性一般基于连接，而不是基于单个主机设备，但这种方式也存在着问题——鼓励主机打开更多连接以竞争更多带宽。

③收敛

即拥塞算法必须快速收敛到正确结论，以便实时更新。

6.3.2 调整发送速率

显式拥塞协议XCP，eXplicit Congestion Protocol：一个显式且精确的拥塞控制协议，路由器告诉源端其最大发送速率。

显式拥塞通知ECN，Explicit Congestion Notification：路由器发送通知警告放慢速度，但没有明确量化。TCP中使用此法

其他的一些拥塞控制机制的反馈信号：p414/428

①控制法则control law

加法递增乘法递减AIMD，Additive Increase Multiplicative Decrease：一种带宽调整策略，可以收敛到公平和效率的最佳点，TCP采用。P414/428

TCP是Interent拥塞控制的主要形式，设计新的传输协议时，需要做到TCP友好（TCP-friendly）的拥塞控制，才能和Internet兼容。

6.3.3 无线问题

无线网络条件下的几个有关拥塞控制的问题：  
（1）无线网络可靠性差，丢包现象时常发生，但并非由于拥塞  
（2）无线网络可变性强，总体容量和各种条件可能随时间剧烈变化

一般情况下，无需为无线网络使用特殊的传输协议。  
（1）因为链路层的可靠性差而导致的超时重传，其时间尺度小于端到端的传输。因此传输层检测到丢包时，足够链路层尝试重传数次。对于链路往返时间较长的网络，如卫星，不适用。  
（2）现有的有线协议足够应付一般程度的网络条件变化，和有线条件下的用户连接流动同等处理。

## 6.4 Internet传输协议：UDP

6.4.1 UDP概述

用户数据报协议User Datagram Protocol：为应用程序提供了无需建立连接就可以发送封装的IP数据报，RFC 768。

UDP段的头格式p418/432

UDP协议适合客户机-服务器应用和多媒体应用（已有专门的基于UDP的更加适用的协议）开发，步骤简单。如DNS系统。

6.4.2 远程过程调用RPC，Remote Procedure Call

定义：允许本地程序调用远程主机上的过程，调用过程称为客户，被调用过程称为服务器。这是一个网络通信的模型或模式，将其仿造为一个普通的“本地”过程调用，不同于套接字模型。

客户存根client stub：代表客户地址空间中的服务器过程（虚拟）

服务器存根client stub：同理

列集marshalling：客户存根将过程调用的参数封装到一个消息中

散集unmarshaled：服务器存根逆向通过消息获得参数

复制-恢复调用call-by-copy-restore：一种过程（函数）调用机制，即将参数复制修改后再更新其原地址数据（粘贴回去）

引用调用call-by-reference：

幂等：形容一个操作其重复并不导致结果的改变，亦即重复安全。比如请求或应答过程无论重复多少次结果都一样。非幂等操作如递增计数器，若无法确认操作已经按需执行，盲目重复会导致结果改变。

RPC模式因为本身的实现机制原因，使用起来需要很多限制，尤其是和C语言等弱类型语言。

6.4.3 实时传输协议

RTP是UDP的派生协议，适合实时多媒体应用领域。但其设计层次高于传输层（即RTP头包含在UDP段内），但是又低于常规应用层（为不同类型的应用提供服务），可理解为应用层上实现的通用传输协议。RFC 3550，p421/435

①RTP——实时传输协议Real-time Transport Protocol

RTP的基本功能是将几个实时数据流复用到一个UDP段包流中。RTP没有确认和请求重传的机制，但是为段包编号以发现丢包和排序，丢包交由应用层处理。

RTP头结构p423/437  
贡献源contributing source  
同步源synchronization source identifier

②RTCP——实时传输控制协议Realtime Transport Control Protocol

该协议也由RFC3550描述，是一个控制协议，不传输媒体样值。处理反馈、同步和用户接口等。Perkins，2003

编码进程可以根据当前网络状况动态改变多媒体文件的编码算法。

③带有缓冲和抖动控制的播放

播放点playback point：多媒体数据传输到达客户端后，可能存在延迟不均等的现象，称为抖动jitter。为了避免，需要先缓冲一定量的数据以确保可以连续顺畅播放。播放点即定义缓冲量或缓冲播放时间长度的点。

大的抖动相对于小的抖动（即便相同的平均延迟），需要更加靠后（晚）的播放点。如果没有抖动，则播放点=传输延迟。

会话峰talkspurts：应用程序可以自适应播放点，当延迟改变时。不过这会造成媒体流不连续，可是对于会话式音频，可以跳过静默时间，不影响应用需求。

## 6.5 Internet传输协议：TCP

6.5.1 TCP概述Transmission Control Protocol

TCP由RFC4614描述，这是一个RFC指南，包括一系列发展至今的RFC。

6.5.2 TCP服务模型

套接字socket：套接字是用于面向连接的传输模型，套接字拥有一个套接字编号或地址。该编号由主机IP地址和一个主机端口号组成。简而言之，套接字=主机IP+端口号，代表连接的一端。

知名端口well-known port：端口号1024以下的端口，p427/441

TCP是全双工，点到点的服务（不支持组播或广播，每个连接只有两个端点）。

紧急数据urgent data：TCP中粗糙的信号机制

6.5.3 TCP协议

特征：连接上的每个字节都有其独特的32位序号

基本协议：具有动态窗口大小的滑动窗口协议

6.5.4 TCP段的头

头结构：p429/443

连接标识符：一个5元组5 tuple。包括协议、源IP地址和源端口号，目标IP和目标端口号；实际上就是协议+2个套接字。

累计确认cumulative acknowledgement：一种接收确认方式，确认的编号是接收方下一个期待的编号，而不是已经正确接收到的最后一个编号。这种确认表示编号之前的所有段或包都已经被接收。

类型-长度-值Type-Length-Value：协议头的选项options字段的常用结构

最大段长MSS，Maximum Segment Size

选择确认SACK，Selective ACKnowledgement

6.5.5 TCP连接建立

SYN泛洪，SYN flood：一种恶意攻击web服务器的方式，通过不断请求建立连接而不完成连接建立行为，导致服务器资源消耗殆尽。

SYN Cookie：一种应对SYN泛洪的策略，但是无法处理TCP选项。所以应该针对性使用，p433/447

6.5.6 TCP连接释放

两个方向独立的释放连接和确认对方的释放连接请求，同时使用超时机制在未被确认的情况下自行断开连接。

6.5.7 TCP连接管理模型

TCP连接管理模型的有限状态机描述p434/448+p435/449

6.5.8 TCP滑动窗口

窗口探测window probe：接收端向发送端表明窗口大小为0时，一般不可以继续发送，但是紧急数据和窗口探测消息除外，后者用于强制接收端重新宣告下一期望字节和窗口大小。

延迟确认delayed acknowledgement：一种TCP的优化方法，将确认和窗口更新消息延迟，以期获得数据免费搭载的机会，减少接收端回复消息的带宽消耗。

Nagle算法：Nagle 1984.一种用来避免发送端发送多个小数据包的机制，或缓冲机制p437/451。不过缺点就是可能会造成短暂死锁，从而延迟网络传输，对部分实时应用难以支持。禁用Nagle算法称为TCP\_NODELAY选项。

低能窗口综合征silly window syndrome：接收端应用层处理数据过慢，导致接收端缓冲区淤积，并以缓慢的速度清空，如一次一个字节。若接收端每次针对一个小空档就向发送方更新一次窗口，则会造成性能浪费。

Clark算法：用来解决低能窗口综合征的算法，可以和Nagle算法互补共同使用。P437/451

6.5.9 TCP计时器管理

重传计时器RTO，Retransmission TimeOut：顾名思义，TCP中的RTO最小值为1s

Jacobson算法：1988.一个动态更新重传超时时限的算法。该算法维护一个变量SRTT，Smoothed Round-Trip Time，也称为平滑的往返时间，和另一个变量RTTVAR，Round-Trip Time VARiation，往返时间变化。每次计时动作（超时或非超时操作）都会有一个时间用来更新SRTT和RTTVAR（使用EWMA方式）。最后RTO间隔由二者计算得出。P439/453

Karn算法：Karn & Patridge 1987。重传后，接收到的确认到底是第一次发送还是第二次的问题，会对RTO的更新有显著影响。这是一个修正算法，用来解决（实际上是换条路）这个问题。P440/454

TCP中采用的计时机制，详细可参考RFC2988

持续计时器persistence timer：用来避免接收端反馈的窗口更新消息丢失而造成的发送端死锁，计时器超时后，发送端再次探询接收端窗口情况。

保活计时器keepalive timer：不常用，用来维系连接

连接终止阶段计时器：超时值为2倍最大数据包生存期，用来确保连接上所有数据包都消失

6.5.10 TCP拥塞控制

拥塞窗口congestion window：窗口大小是任何时候发送端可以往网络发送的字节数，和流量控制窗口独立，后者是接收端缓冲区大小决定。两个窗口或者两个端的所有窗口相互作用时以最小的为主。

确认时钟ack clock：确认包返回的速率和整条路径上最慢的链路的速率相同，发送端可以参考这个速率进行发送，从而不会堵塞沿途路由器缓冲区

慢速启动slow start：拥塞窗口启动的算法，从一个较小的窗口大小，逐渐增加至一个合理大小。采用指数增长方式。

慢启动阈值slow start threshold：慢速启动算法中的一个控制量，当拥塞窗口大小抵达阈值，则采用线性递增的方式缓慢增加窗口大小。

重复确认duplicate acknowledgement：携带相同确认号的一系列确认，说明数据包丢失。也有可能是发送的数据包没有按照顺序被接收，不过这种情况较少。TCP中以3次重复确认视为丢包，以第三次重复掩盖低概率的非顺序接收情况。TCP是面向连接的，非顺序接收的情况确实少见。

快速重传fast retransmission：由重复确认激发的重传，而非超时

TCP Tahoe的拥塞窗口控制示意图或拥塞算法示意图p445

快速恢复fast recovery：一种快速重传之后调节拥塞的改进式算法p446，拥塞窗口大小呈锯齿sawtooth

TCP Reno：相当于Tahoe+快速恢复，是现代TCP的核心

CUBIC TCP——Linux的TCP

Compound TCP——Windows

选择确认SACK，Selective ACKnowledgement的描述：p447

拥塞窗口减少CWR，Congestion Window Reduced：TCP段标志位，用于发送方告知接收端它已经收到拥塞信号

6.5.11 TCP未来

TCP不能保证为每一种应用程序解决一切所需问题，因此需要应用程序本身作为。

随着网络速度加快，丢包率下降，以丢包为信号的拥塞控制不再那么有效。

## 6.6 性能问题

6.6.1 计算机网络中的性能问题

广播风暴broadcast storm：由于广播数据发生错误，导致所有接收端都反馈错误报告，而造成的突发大量数据包

同步过载：在某些特定情况下，很多机器同时进行相同的发送动作造成的网络冲击。

6.6.2 网络性能测量

测量过程需要注意：

①确保样值空间足够大

100万次取平均

②确保样值具有代表性

针对网络流量的时间周期均匀采样

③缓存可以破坏测量结果

必须先对网络缓存机制了解才能准确测量性能数据

④确保测试期间不会发生不可知的事情

意外事件导致的流量

⑤小心使用粗粒度时钟

⑥小心推断结果

不要肆意推断结果的趋势

6.6.3 针对快速网络的主机设计

线速率wire speed：即数据链路上速率，这里特指路由器和网卡的处理速度现在已经和数据链路层的速度相当。

①主机速度比网络速度更重要

②减少包计数来降低开销

降低用在处理段和包的头部的开销

③最小化数据预取

数据预取即对数据复制至内存以待处理，一般而言每一层协议的执行都需要一次数据预取。TCP/IP协议的一个特点就是两个层的协议对同一份数据的内存复制进行处理。

④最小化上下文切换

上下文切换context switching：一种程序运行环境的切换，此处特指内核模式和用户模式的切换。上下文切换需要消耗CPU的性能，进而消耗网络性能。

复制操作会导致CPU中断进入内核模式，传输协议一般实现在内核中以减少上下文切换。

⑤避免拥塞比从中恢复更好

⑥避免超时（计时器的设计实现与优化）

计时器和超时机制的慎重使用，实际上大部分有线链路的可靠性极高，超时等现象发生的概率较低。虽然必须处理低概率的问题，但可以优化其性能减少对正常情况的资源占用。

6.6.4 快速处理段

头预测header prediction：首先快速检查头，然后再详细处理（调用TCP进程）的通用模式，可提高速度

计时轮timing wheel：一种高效实现计时器功能的数组结构p457

6.6.5 头压缩header compression

在无线网络中，最重要的性能因素是带宽的使用而非软件的性能，尽管后者也有意义。使用各种压缩头的技术可以有效利用无线带宽。

鲁棒头压缩ROHC，Robust Header Compression：RFC5795定义的现代头压缩技术框架

6.6.6 长肥网络的协议long fat networks

定义：速度快，长延迟的远距离高速网络

设计原则：为了处理速度而不是带宽优化进行设计

长肥网络的特点和特殊问题：  
数据包和段序号不够用而回绕  
流量控制窗口相对网络的带宽延迟积不能太小，否则大部分资源无法利用  
高延迟导致一些重传协议运行不佳，而且高延迟是长肥网络的主要瓶颈而非带宽  
高延迟导致一些反馈机制运行不佳，可以使用基于速率或面向连接预留带宽的协议  
高速高带宽导致计算机计算能力不再占据优势反倒也成为瓶颈，有必要使用简洁的协议

网络协处理器network coprocessor：用硬件来构建快速网络接口的设计

## 6.7 延迟容忍网络DTN，Delay-Tolerant Network

定义：也称为中断容忍网络disruption-tolerant network。基于消息交换通信方法构建的网络，消息交换体系针对终端始终处于移动或其他极端条件而无法保持网络连接的情况，对消息进行暂存和延迟转发。类似邮件。

6.7.1 DTN体系结构

由RFC4838描述，虽看似路由器存储转发，但两者延迟转发的时间尺度不同，且DTN节点可以移动。

数据束bundle：DTN体系中的一条消息的术语

接触contact：一条DTN工作链路，非工作链路不算

6.7.2 数据束协议Bundle protocol

延迟容忍网络协议栈p464/478，数据束协议的层定位类似RTP

数据束协议的消息格式p465/479  
主块primary block  
有效载荷块payload block  
可选块optional block

# 第7章 应用层

## 7.1 DNS——域名系统（Domain Name System）

定义：RFC1034,1035,2181。将主机名映射成IP地址。

工作方式：应用程序调用名为解析器resolver的库程序，以名字作为参数。解析器向本地DNS服务器发送请求查询的报文，获得响应回复后，IP地址即返回调用方。

7.1.1 DNS名字空间

ICANN：Internet名字与数字地址分配机构，Internet Corporation for Assigned Names and Numbers。

顶级域名top-level domains：DNS域名由后至前逐渐降低级别，域名的最后一个部分即该域名的顶级域名。顶级域名分为两类，通用的p473/487和国家或地区的。每一级的域名申请需要向上一级的管理部门提交和统筹。

一个域无论是什么级别都可能包含一台或数台主机，即域名和主机并不一一对应。如主机有两个以上网络接口即IP地址，那么一个域名也不与一个IP地址一一对应。域名可以理解为一种高级命名语言，没有实质上的技术层面的内涵。

域名抢注cybersquatting

域名不区分大小写，每一级的域名最多63字符，总长度最多255个字符。

7.1.2 域名资源记录resource record

实际上，DNS是一个域名与资源记录的映射。资源记录是一个五元组数据结构，不仅仅包含IP地址信息。资源记录组成了DNS数据库。  
Domain\_name Time\_to\_live Class Type Value  
域名 生存期 类别 类型 值

主要的DNS资源记录类型Type：p475/489

DNS数据库内容举例p477/491

7.1.3 域名服务器

域名服务器是解析域名查询name resolution的机器，为了效率实现，域名服务器和DNS数据库是分布式的，每个域拥有自己的一台或数台本地域名服务器和本地域的所有DNS资源记录。具体如何分配，尤其是低级域由区域管理员决定。

权威记录authoritative record：本地域名服务器存有和管理本地域的资源记录，这些资源记录就是权威的可信的

缓存记录cached record：暂存在某个域名服务器中的其他域的资源记录，可能是最近查询备份的

解析过程实例p478/492

根域名服务器：共有13个根域名服务器分布于世界各处，存有顶级域名的服务器资源记录

递归查询recursive query：本地域名服务器代替本地主机的查询过程，只返回最终结果

迭代查询iterative query：本地域名服务器逐步解析域名的查询过程，每次返回部分结果

## 7.2 电子邮件

传统的纸质邮件snail mail

微笑图标smiley：像:-)的图标，其他种类的情绪图标如emoticon

7.2.1 体系结构和服务

用户代理user agent：电子邮件子系统之一，即收发邮件的客户端程序，或称电子邮件阅读器

邮件传输代理message transfer agent：电子邮件系统之一，也称邮件服务器，负责转发邮件至目的地

邮箱mailbox：存储用户的邮件，由邮件服务器维护

7.2.2 用户代理

自动响应和休假代理vacation agent

罐装回复canned reply

签名块signature block

邮箱地址格式（SMTP）：用户（主机名）@DNS地址

邮箱地址格式（X.400）：p485/499，已经被淘汰的ISO标准

邮件列表mailing list：群发邮件至列表中所有人，邮件列表可以在用户代理或传输代理上维护。传输代理上的邮件列表在格式上可能和一般的个人邮箱地址相似，无法区分，但它却是等效代表了多个邮箱地址。这种情况下，邮件列表中的每个邮箱地址不一定和邮件列表本身具有相同的域名，可以理解为邮件首先被路由到邮件列表所在的域，然后分别路由到所有成员。

7.2.3 邮件格式

邮件的传输格式RFC821-822和RFC5321-5322修订

①RFC5322——Internet邮件格式

邮件由一个信封envelope（包含了传输代理所需的路由信息）、数个头字段（header用户代理的控制信息）、一个空行和邮件体body组成

邮件传输的主要头字段：p487/501，若将信封和消息头区分开，这实际是信封段  
Cc：Carbon copy抄送，名字源于复写纸  
Bcc：Blind carbon copy密件抄送

邮件头主要字段：p487/501，若不加区分信封和消息头，这些字段可以以任意顺序混合在一起

用户可以自定义自己的邮件头，必须以”X-“开头命名即可。

②MIME——多用途Internet邮件扩展Multipurpose Internet Mail Extensions

描述：RFC2045-2047、4288、4289以及2049

MIME扩充了原有的邮件协议，添加了一些邮件头：p488/502  
以及对于在原有的邮件传输系统上如何发送二进制类型文件（编码）进行扩展

MIME类型Content-Type：RFC1521, 7种基本类型p490/504  
运动图像专家组MPEG，Moving Picture Experts Group：视频格式  
MP3音频格式：MPEG3  
MP4：MPEG4

除了基本的MIME类型，一些商业机构根据自己的产品而增加了一些MIME类型，包括用户自定义的MIME类型。这些类型和子类型一般由vnd.带头标识，采用如下的类似格式：vnd.android.cursor.dir/vnd.xxx.xx

多部分multipart带MIME扩展的邮件举例：p492/506

7.2.4 邮件传送

简单邮件传输协议SMTP，simple mail transfer protocol：邮件传输的协议，RFC5321

SMTP包含两部分，用户邮件提交mail submission和传输代理之间传送，最终邮件交付使用不同的协议

①SMTP及其扩展ESMTP，Extended SMTP

扩展功能p495

②邮件提交

开放邮件中继open mail relay

③邮件传送

DNS查询MX记录确定接收方邮件服务器地址

7.2.5 最后传递

SMTP是一种基于推的协议即push-based，即旨在推送消息。而用户接收邮件的需求往往是基于拉的。

①IMAP——Internet邮件访问协议Internet Message Access Protocol

描述：RFC3501，版本4

用户需使用IMAP协议的客户端（用户代理）访问邮件服务器，而邮件服务器也必须运行IMAP服务器，且使用端口143，也有一些其他的客户端专用协议，如Microsoft Exchange

IMAP命令：p498/512

邮局协议版本3，POP3，Post Office Protocol version 3：IMAP的早期版本，RFC1939定义

②Webmail

使用Web程序代替用户代理，邮件提交采用Web协议。

## 7.3 万维网

Web是万维网world wide web的俗称，确切定义是一个应用体系结构框架，用于将Internet上的主机中的内容供他人访问。网站和网页称为Web内容。

世界上第一个图形浏览器叫Mosaic

W3C，World Wide Web Consortium：万维网联盟

7.3.1 体系结构概述

Web页面Web page或page：可能是文本、图形、格式化命令等，确切定义是一个存储在Internet主机上的资源，可用于呈现页面表象和后台程序逻辑等

超文本hypertext：确切定义是让一个页面指向另一个页面

超链接hyperlink：可导向另一个关联页面相关的文字、图标或图片

浏览器browser：能够将页面内容或Web程序适当展示出来的另一种程序

Web体系结构p501/515

①客户端

统一资源定位符URL，Uniform Resource Locator：针对每个页面的标识符，包括  
协议或方案scheme  
页面所在机器的DNS域名  
唯一指向特定页面的路径（可能和实际文件目录相同也可能无关，解释权归服务器）

点击链接或URL获取页面的过程p502/516

各种协议的URL方案举例p503/517  
https, ftp, file, mailto, rtsp, sip, about

统一资源标识符URI，Uniform Resource Identifiers：URI包括URL和URN，URL指出了页面资源的获取地址（DNS），而URN仅仅指出了资源名或本地路径，没有给出获取资源的主机的域名。可用于不关心资源获取位置的场合，如有多个资源获取点。

统一资源名URN，Uniform Resource Names

书写URI的规则，见RFC3986

②MIME类型

MIME通过其特征性的“类型/子类型”表示，作为页面中包含的文件内容的类型描述方式。道理上和在电子邮件中一样。

插件plugin：第三方代码模块，浏览器扩展，通过扩展插件可以处理更多文件类型

辅助应用程序helper application：另一种扩展浏览器的方式，即浏览器遇到特定文件类型时，直接自动打开第三方程序本身，如Word，PowerPoint等

对于exe类型，应该小心安全漏洞。

③服务器端

服务器循环的基本步骤p506/520

服务器循环的较详细步骤p507/521

多线程模式multithreaded：服务器由一个前端模块front-end module和k个处理模块processing module组成。处理模块采用多线程模式运行。

缓存模式：服务器将最近发送过的文件存入缓存，每次查询文件时优先从缓存中查找，降低读取磁盘的性能消耗

④Cookie

定义：一小段文本信息，伴随着用户请求页面在Web服务器和浏览器之间传递，既可以将一些信息存储在用户本地主机，也可以在下一次访问对应网站时为对方服务器提供这些信息。Cookie是和域名相关联的，每次用户访问一个网站，发送请求的同时会将本地的所有该域曾经留下的Cookie一并发送。RFC2109

非持续Cookie，nonpersistent cookie：没定义过期时间Expires的cookie，浏览器退出则删除

持续Cookie，persistent cookie：定义了过期时间的

间谍软件spyware

第三方Cookie，third-party cookie：指除了用户本地和被访问域的另外一个域的Cookie。第三方域可能会在被访问的网页上（第二方）留下一个URL，从而跨站点监视用户的踪迹。第二方和第三方可能存在合作关系。

7.3.2 静态Web页面static page

定义：每次显示的是一个不变的内容，除非手动更改

①HTML——超文本标记语言HyperText Markup Language

定义：来源于早期印刷出版行业的术语，编辑将有关如何印刷的信息标记于文件之上，交给工人印刷。

指示符directive：标签<sth></sth>中的内容

HTML使用ISO 8859-1 Latin-1字符集编写，对于普通ASCII键盘，必须使用转义序列（以&开头）表示特别的一些字符，如空格、<、>、&本身和其他非英文字符。

HTML版本功能的演进表1代-5代p515/529

②输入和表单

举例p517/531

表单信息从浏览器返回服务器的可能形式举例（响应）：p518/532

这里的例子是针对Web浏览器作为前端平台的情况，而前端页面由HTML定义。表单信息按照一定的格式形成一个连续字符串被发送回服务器，进一步交由对应进程处理。

②CSS——层叠样式表Cascading Style Sheets

样式表style sheet：用来以文本内容的结构划分定义网页内容的格式

CSS文档的例子：p519/533

HTML有两种方式引用CSS，第一种是使用<style>标签直接将CSS内嵌，不推荐。第二种是引用外部连接的CSS文档，p519/533

7.3.3 动态Web页面dynamic page和Web应用

定义：页面每次显示的内容是程序（服务器上或浏览器内）按需产生的，或页面本身包含程序。

云计算cloud computing：一种Web应用的工作模式，即Web应用相对于传统应用将计算从用户主机转移到Internet服务器上

①服务器端动态Web页面生成

以下列出4种常用技术和API——

注：这些是服务器端脚本技术，由服务器运行计算

（1）公共网关接口CGI，Common Gateway Interface

RFC3875，后端程序和脚本使用CGI与Web服务器通信，交换信息返回页面。后端程序一般使用Python，Ruby和Perl等脚本语言，但也可以使用任何其他编程语言。后端程序和服务器分属不同进程。

（2）超文本预处理器PHP，Hypertext Preprocessor

在HTML页面中嵌入少量脚本，让服务器（进程）执行这些脚本以便生成最终页面，页面文件扩展名为.php

PHP应用举例p521/535

（3）Java服务器页面JSP，JavaServer Pages

与PHP非常相似，只是页面的动态部分用Java编写而不是PHP。页面文件扩展名为.jsp

（4）活动服务器页面ASP.NET，Active Server Page

微软的PHP和JSP对等体。使用.NET网络应用框架来生成动态内容。页面文件扩展名.aspx

②客户端动态Web页面生成

动态HTML，dynamic HTML：产生交互式Web页面的技术

注：这里讲的是客户端脚本技术，由客户机运行计算以及和用户交互（交互程度较复杂于表单输入等，如响应鼠标移动事件）

（1）JavaScript

是最流行的客户端脚本语言

（2）VBScript

基于Visual Basic的适用于Windows平台的脚本

（3）Applet小程序

使用Java编写，已被编译成一种虚拟机JVM的机器指令。ActiveX控件是其中一个例子，特点是运行速度快，跨平台性强。

③AJAX——异步JavaScript和XML，Asynchronous JAvascript and Xml

AJAX是一组技术，包括：  
HTML和CSS  
DOM——浏览时改变部分页面  
XML——程序与服务器交换应用数据  
异步方式——程序发送和检索XML  
JavaScript——组合起所有部分

文档对象模型DOM，Document Object Model：通过程序访问HTML页面的一种表示，这种模型实际上是一个树模型，p526/540。结构化的HTML文档对象便于对部分页面进行更改而无需重写整个页面。

可扩展标记语言XML，eXtensible Markup Language：一个说明结构化内容的语言。  
XML是一个自由的文本语言，没有通用的单一明确的语法和编译规则等。但是常用来表示一种结构化的内容（不在乎其显示方式），用于程序之间的数据交换和通信等，XML文件可以通过一些辅助技术转换成HTML，后者是必须关注显示内容的方式的。

可扩展样式表语言转换XSLT，eXtensible Stylesheet Language：用于将XML内容样式化成HTML的语言技术。

扩展超文本标记语言XHTML，eXtendend HTML：严格遵循XML语言规范的HTML版本。HTML使用初期依靠手写，语法规则不明确，相反XML虽然没有确切的使用目的，但是语法规则明确严格。

异步工作方式的含义：Web应用的用户接口不应该因为其正在等待请求的响应而被阻塞，应该采用异步方式保持和用户的交互同时等待数据到达。这是一种用户界面的设计思想。

Web服务，Web service：由HTTP协议承载的通信服务

简单对象访问协议SOAP，Simple Object Access Protocol：运行在HTTP之上的XML RPC。实现了Web服务的方法，以一种语言无关且系统平台无关的方式执行程序之间的RPC，实际上是依赖XML承载消息内容p527/541

7.3.4 HTTP——超文本传输协议HyperText Transfer Protocol

含义：是一个请求-响应协议，运行在TCP之上，RFC2616。起初是一种应用层协议，针对Web浏览器和服务器，但随着应用广泛，开始面向更多应用需求。特征上出现传输层的特点。

①连接

持续连接persistent connection：早期HTTP的TCP连接，只处理一次请求响应就关闭。现在已经更改，也称为连接重用connection reuse

并行连接parallel connection：持续连接之前使用的技术，同时并行打开多个TCP连接

②方法

HTTP协议支持请求一个Web页面，也支持请求方法method，即对服务器调用一个方法（实际上请求一个页面用的就是GET方法）。

内置的请求方法如p531/545，且区分大小写。最常用的（浏览性）Web服务方法是POST和GET。

HTTP请求：由一行或多行ASCII文本组成，第一行的第一个词是请求方法名称

HTTP响应：由一个状态行及可能的附加信息组成，状态行包括一个3位数字的状态码。状态码含义和使用方法见p532/546

③消息头

消息头分为请求头request header和响应头 response header，是跟随在请求行之后的额外信息行，类似调用过程的参数或属性。

消息头举例p532/546，有些头是强制性的必须添加，如Host

④缓存caching

定义：积攒已经获取的网页供日后使用的处理方式

HTTP缓存策略和查验过期性过程示意图p534/548

代理缓存proxy caching

⑤HTTP实验

HTTP消息举例p535/549

7.3.5 移动Web

移动Web或mobile web指的是手机或小型设备上使用Web服务的技术。移动平台有诸多相较于PC平台的性能劣势。

无线应用协议WAP，Wireless Application Protocol：早期的针对性能有限的无线设备而设计的协议栈。随着无线设备的硬件性能提升，尽管和PC仍存在差距，但使用单独的协议逐渐变得不再必要。不过仍然有必要使用一些类似头压缩技术来降低移动Web的带宽和能耗。

移动友好Mobile-friendly：使用相同的协议情况下，仍然可以让服务器通过请求信息分辨出用户浏览器或机器的种类

标准化移动Web内容的最佳实践best practice：一组由W3C提供的由普通Web内容转换到移动Web内容的推荐

基本XHTML，XHTML Basic：精简版的HTML，用于移动Web领域，也可以起到轻量化Web的作用。基本的XHTML定义的元素见p537/551

内容转换或转码content transformation or transcoding：由一台机器工作在移动客户端和服务器之间，负责将普通Web内容转换或转码。

7.3.6 Web搜索

Web爬虫，Web crawling：Web搜索引擎普遍使用的搜索方法，即从少数页面开始通过遍历所有网页和链接来获取查询结果。为此，多数搜索引擎背后维护了一个存储所有外部可见Web页面索引的数据库。Web crawler称为爬虫程序。

深层Web，deep Web：一般爬虫遍历碰触不到的页面，多为需要登录验证或者程序生成的动态页面。

## 7.4 流式音视频

实时音频/视频，real-time audio/video

IP语音，voice over IP：等同于Internet电话，Internet telephony

多媒体multimedia：准确的定义是两个以上的连续媒体continuous media同时播放，最常见的即为音频和视频同时播放。不能播放的静态媒体不属于，但是现在已经不再追究纯粹的音频是否属于。

流媒体streaming media：描述基于网络的视频音频应用技术

7.4.1 数字音频

描述声音大小的分贝单位是一种相对性的对数性的比率，即分贝只是一个比率的对数表示，分贝单位本身不具有物理意义。一般常见的声音描述，是用1kHz（能听度低限）作为比率的分母，即相当于0dB。

模数转换器ADC，Analog Digital Converter

数模转换器DAC，Digital-to-Analog Converter

量化噪声quantization noise：由于数字采样值的位数有限而引入的误差

压缩光盘CD，Compact disc

①音频压缩

数字音频原始信号会占用很大的带宽，CD级立体声音频会占用1.411Mbps的带宽。所以传输和存储音频文件需要对其进行编码压缩。

音频压缩和解压实际上是一种编码encoding和解码decoding过程。目的就是为了减少音频文件的带宽需求等，尽管音频文件比视频文件小很多。

压缩算法和解压算法并不对称，为了让解压算法简单快速，实际应用中可以让非实时压缩算法不计成本。因为压缩计算往往只进行一次，所有的用户都需要进行解压。但对于实时音频应用而言，压缩算法或编码算法不能太慢，因此往往使用不同的算法，进而压缩程度降低。

压缩解压算法另一不对称性表示为这两个过程不必是可逆的。经编码解码后的信号如果和原始信号有误差，称为有损耗lossy。反之称为无损lossless。有损系统的意义在于，丢失少量不重要的信息可能换来巨大的性能提升。

声码器vocoder：负责为特殊语音进行压缩音频的机器，也称语音编码器。用于电话系统。

MPEG音频层3：即MP3，MPEG audio layer 3是一种音频压缩算法。MPEG目前共有三个版本，MPEG-1、MPEG-2和MPEG-4，是一种提供了音频和视频编码算法的标准。MP3是MPEG-1标准中的音频压缩部分，也是其第三个部分。

高级音频编码AAC，Advanced Audio Coding：是MPEG-4标准中使用的默认音频编码。因此MP3是一种音频编码算法，而MP4才是MPEG-4的缩写，包括了音频视频两种技术。MPEG-2标准中已经可以使用MP3和AAC两种音频。

波形编码waveform coding：完全基于数学和物理实际的转换，使用傅里叶变换编码。

感知编码perceptual coding：另一种音频编码方式，基于心理声学psychoacoustics。将声音信号压缩成人耳觉察不出区别的形式，但实际上已经很大程度改变了信号。MP3和AAC都是此类。

屏蔽mask：感知编码方式的核心特性，利用了某些声音听觉上掩盖了其他的声音的性质，来压缩音频。同一频段中较大的声音掩盖另一个较小的声音，称为频率屏蔽frequency masking。根据人耳特性，较大的声音停下时，短时间内较小的声音也无法听到，称为暂时屏蔽temporal masking。

可听度阈值：对于人耳而言，不同的频率的声音能够被人耳感知到时所需要的最低功率是不一样的。可以根据这个阈值对声音信号进行裁剪屏蔽。但是注意有其他声音存在时，会影响到可听度阈值曲线在该声音频率周围的分布。P544/558

采样率；对原始信号波形进行采样的频率，采样值根据不同情况进行离散选择，如8位对应256个不同的采样值。采样后得到的是初步的数字信号。

输出比特率：压缩编码之后的信号文件的比特率

音频编码过程简单介绍p544/558。编码压缩的关键作用点在于将听不到的次要的信号剔除，给主要的信号分配更多的字节，同时使用Huffman编码再次对字节进行加权编码进一步压缩。

7.4.2 数字视频

视频的数字表示法就是一个帧序列，每个帧包含一组像素pixel矩阵或称为图像元素。每个像素由一组数字进行单一的色彩描述。帧序列的理想速率是50个全帧每秒，此时人眼觉察不出任何离散性。

隔行扫描interlacing：将每一全帧分为两个域field，一个由奇数扫描线构成，另一个为偶数扫描线。实际上是两个相互穿插的0.5个帧交错顺序播放，可将全帧的帧速率提高一倍。一般用于电视机。

计算机显示图像可以使用显卡上的缓冲区，显卡可以使用缓冲区中的图像以比全帧速率更高的速率重新绘制屏幕，以保证屏幕不会闪烁。隔行扫描的视频在计算机上播放时，会产生梳理效果combing。

高清晰电视HDTV：1280\*720p

①视频压缩

类似音频的原始文件，无压缩视频文件的规模很大，标清480p，24位色彩，30FPS的视频需要200Mbps。

一种视频压缩的思想是将视频处理成一组连续图像的静态编码，然后增加一些额外的适应视频特性的处理。

②JPEG标准

联合图像专家组JPEG，Joint Photographic Experts Group：是一个由一组图像专家在标准制定组织的主持下共同制定的压缩连续色调的静止图片的标准

JPEG使用的颜色表征是亮度Y和色度Cb,Cr。而不是RGB。人眼对于亮度比对色度更加敏感，两种颜色表征可以转换。

JPEG图片压缩过程简述p546/560  
1）块准备  
2）离散余弦变换DCT，Discrete Cosine Tra nsformation  
3）量化quantization  
4）差值替代  
5）行程编码run-length encoding

JPEG算法较复杂，能够达到20:1的压缩率，且解码和编码花一样长的时间。

③MPEG标准

MPEG-1标准目标是产生录像机品质的输出，压缩率40:1，编码速率1Mbps。

MPEG-2目标是压缩广播品质的视频，是DVD视频编码和DVB数字广播电视的基础，编码速率4-8Mbps。

MPEG-4由两种视频格式，第二种是H.264或高级视频编码AVC，Advanced Video Coding。目标是编码率相对前期同质量视频的一半。适合输出蓝光光盘和HDTV。AVC的压缩率超过50:1。

MPEG视频压缩的过程简述p550/564。MPEG的视频帧包括三种  
I-帧，Intracoded Frames——静态图片编码  
P-帧，Predictive Frames——宏块macroblock之间的差异编码  
B-帧，Bidirectional Frames——双向前后帧差异编码

7.4.3 流式存储媒体

定义：不是下载之后播放媒体，也不是在线直播媒体，而是在线观看网站存储的媒体。

视频点播VOD，Video on Demand：流式存储媒体的服务名称

简单的下载之后播放媒体的模型示意图p551/565

使用Web和媒体双服务器的流式媒体示意图p552/566  
RTSP实时流协议，Real Time Streaming Protocol：媒体播放器对服务器远程控制p557/571

媒体播放器：  
（1）管理用户界面  
（2）处理传输错误——前向纠错FEC、交错编码interleaving  
（3）解压缩内容  
（4）消除抖动——缓冲、启动延迟

低水位标记low-water mark：缓冲区的低标记，用以通知媒体服务器开始发送或媒体播放器开始播放。媒体服务器使用TCP响应播放器时，由于建立连接成本较大，需要更高的低水位标记，但是更加安全可靠。

高水位标记high-water mark：缓冲区高标记，用以通知媒体服务器暂停发送

7.4.4 流式直播媒体

IP电视IPTV，IP TeleVision

网络广播Internet radio

播客podcast广播电台的系统模型p560/574

Internet广泛使用的流直播方式是基于TCP的，也就是说每个用户都要与媒体服务器进行一对一连接。这里面有很多特殊的原因。UDP方式直播或组播一般仅限于单独的ISP网络内部。

7.4.5 实时会议

IP语音voice over IP和IP电话Internet telephone

最大的挑战在于降低延迟。有以下几个技术要点来达到：  
使用UDP  
使用小数据包，牺牲带宽效率降低延迟  
使用较小的缓冲区，在延迟和抖动丢失之间权衡  
使用服务质量机制，如区分服务  
保证有足够的带宽

①H.323

实际名称：无服务质量保障的局域网可视电话系统与设备Visual Telephone Systems and Equipment for Local Area Networks Which Provide a Non-Guaranteed Quality of Service。后更为基于数据包的多媒体通信系统

H.323实际上是一个IP电话系统的描述，即引用了其他的多种协议的整体。结构模型见p563/577

网关gateway和网守gatekeeper

H.323协议栈p564/578  
音频视频编码解码——G.711（标准电话音频编码）和MPEG（H.264）  
允许终端协商使用何种编码解码算法——H.245  
建立和释放连接与提供标准电话功能——Q.931  
终端与网守通信——H.225

H.323中不包括服务质量的定义，是由底层网络具体解决的。

②SIP——会话发起协议Session Initiation Protocol

H.323是一个完整协议集，SIP只是一个应用层模块，只处理会话建立、管理和终止，仍需其他协议来实施数据传输。SIP也是一个文本协议，类似HTTP发送文本消息。

SIP的常用消息类型/方法 p566

SIP与H.323对比 p567

## 7.5 内容分发

7.5.1 内容和Internet流量

齐普夫定律Zipf’s law：是幂定律power law的一种。即相对频率和频率排名成反比，用以描述常见字使用（起初），也可以用来描述网站使用频率分布

指数衰减exponential decay：用以描述原子衰减的规律，不同于幂定律，其衰减速度很快使得忽略“长尾效应”long tail是可以接受的。

7.5.2 服务器农场和Web代理

①服务器农场

使用多台服务器在同一地点替代单独的服务器进行工作

负载均衡load balancing：服务器农场的前端front end将所有请求均衡映射到农场内所有服务器

中间件盒middlebox：插入到一条网络路径中间的路由器或交换机，但是不起到网络层的作用，而是根据传输层或更高层信息行动，也就是农场前端的一种形式。

②Web代理Web proxy

可以被多个用户共享的高速缓存。用户向代理发送请求而非直接向目标服务器，只有代理没有请求页面时才会向目标服务器请求，返回用户结果后缓存新页面备用

用户的浏览器也具有缓存能力，代理缓存相当于二级缓存。

上行流代理upstream proxy：对应下行流代理downstream，是一种代理等级之间的相对描述，用户浏览器缓存相对于代理就是下行流，可以有更多级代理。

优点：  
代理可以保护和隐匿用户的身份  
代理可以过滤用户请求以及过滤网站等

问题：  
无法处理特别大的用户群，100人左右为宜，特别多的人群导致存储空间匮乏，不流行的网站得不到缓存，由于长尾效应，此时代理缓存的优势将不会再增长。

7.5.3 内容分发网络CDN，Content Distribution Network

内容分发网络采用上下级树形结构，CDN源服务器作为主服务器将其所有内容复制给若干个CDN节点，这些节点也成为镜像点。

使用CDN的具体方式有多种：采用将CDN节点作为Web代理（最不流行），让用户自行选择CDN镜像点来下载内容（实质上是不同的web站点），以及DNS重定向

DNS redirection：CDN技术最有效的实现技术，特点是自动分配CDN节点以及外观上是一个单独的网站。每一次对网站内容的请求，在网站自身域的DNS服务器最终不会将全域名解析为一个固定的IP，而是根据请求方的IP来返回不同的地址。

CDN可以由网站公司自己实现，也可以由第三方公司作为一项服务实现，后者的技术示例和过程见p577

优点：  
增加公司服务大量客户和处理大量信息的能力  
可以增加使用CDN服务的公司网站应对突发流量的能力（闪群flash crowd）

CDN技术的更多细节【Dilley，2002】

7.5.4 对等网络P2P，Peer-to-Peer

对等节点peer：即一个节点既可以充当客户也可以充当服务器，所有节点之间彼此都是平等一致的。

对等网络能够自我调整，理论上它的网络能力依赖的是用户群的网络能力，而用户群增加并不会消耗而是增加网络的能力。

BitTorrent和DHT是两种对等网络技术，前者较早，一般只适用于小范围的共享文件（至少发明之初如此）；后者是P2P系统作为整体运行在大范围的技术核心。

①BitTorrent

定义：BitTorrent是一个协议，由Brahm Cohen,2001开发。

种子文件torrent：描述下载内容的文件，用以为用户提供某些信息（跟踪器名称，片或块名称清单），以及根据其描述验证下载数据的完整性。

跟踪器tracker：一个服务器，维护着所有上传和下载某一内容的所有对等用户列表。这些用户称为用户集群swarm。所有加入的用户都要定期向跟踪器汇报自身的活跃情况。

块chunk：组成下载内容的一部分，不同种子文件定义的块可以不同大小（64KB-512KB），同一种子文件定义的块大小一致。

种子seeder：拥有组成内容的所有块的对等用户。

对等节点之间通过交换各自拥有的块清单，选择其中自己需要的但是最“罕见”的块首先下载，以此扩展来提高所有块的可用性。

抑制算法choking algorithm：用以根据对等节点之间上传行为来控制节点之间连接的算法，该算法支持鼓励有良好上传行为的节点之间合作，反之抑制或阻塞choke没有良好上传行为的节点。

②DHT分布式哈希表Distributed Hash Tables——p582

结构化的P2P网络structured P2P networks：DHT需要节点之间使用正规结构连接，是为结构化，而传统的P2P协议是非结构化的，即节点之间可使用任何连接

以下以Chord【Stoica etc., 2001】方案为例介绍一种DHT技术实现：

DHT核心是一个哈希表，每个槽对应一个或0个对等节点，槽位号hash(节点IP)一般作为节点标识符node identifier。关键字key=hash(torrent)是将种子文件内容同样散列到哈希表中（注意这里的key定义不是“散列表的关键字”定义）的槽位号，key之后的第一个非空槽位对应的节点（用successor(key)表示）存有该torrent文件对应用户集群信息的列表。获取了该列表之后，即可进行传统意义上的P2P下载。

指取表finger table：每个节点维护的一张“小”表格，里面记录了一部分的槽位号k和successor(k)的IP地址（即k后第一个非空节点的IP）的信息。表格的规模，记录哪些key，都有准确定义。所有的节点通过指取表接力式查找某一特定k对应的IP地址（这个对应关系如上所述，较复杂），以此提高搜索效率。

为了总结这个复杂的过程，可以将DHT理解为一个过程，它将本应在跟踪器上的用户集群信息分布式存储在节点上，并通过哈希表等方式索引查找存有某一特定集群信息（对应某一特定种子文件）的节点IP。即DHT将跟踪器分散或分布化到各个节点本身了。

# 第8章 网络安全

网络安全可分为4个领域：  
①保密secrecy  
也称机密confidentiality  
②认证  
③不可否认  
防止对方篡改和抵赖  
④完整性控制  
确认信息不是伪造的

链路加密link encryption：数据链路层的加密手段，对所有信息无区别对待，但是在路由器出入口处会进行加密解密过程，易受到路由器内部攻击

在网络系统中的每一层都存在网络安全相关的内容和技术，这并不是一个层独立的技术话题。

网络安全和操作系统安全/应用系统安全是独立的话题，虽然有时区分不明显。对于后者的学习，如木马病毒、代码注入、口令安全、缓冲区溢出攻击等，参见Modern Operating Systems【Tanenbaum，2007，第9章】

## 8.1 密码学cryptology

密码学是除了物理层安全之外的其他层安全的技术基础。其他相关参考书籍列表见p593

密码cipher：逐个字符或者逐位进行变换，不涉及信息的内容或语言结构

编码code：用一个词或符号来代替另一个词，通常目的不是加密而是提高数据性能，因此编码方式一般和被编码的内容有关。虽然编码有时也可以起到密码的作用，而编码也可以逐位进行，以至于二者没有明显界限。

8.1.1 密码学概论

明文plaintext：待加密消息

密钥key：加密方法或函数的参数

密文ciphertext：加密后的信息，它等于以key为参数的加密方法对明文的处理结果，实际上加密方法本身多为两个参数的函数

密码分析学cryptanalysis：破解密码的技术，假定分析者一定知道加解密方法本身

密码编码学cryptography：设计密码的技术

密码学cryptology：上面两者的集合

主动攻击者或入侵者：可以修改信息或密文

被动攻击者或入侵者：只能监听密文

Kerckhoff原理Kerckhoff’s principle：让密码分析者知道加解密算法，把所有的加密信息放在密钥中。这是弗兰德斯（地名）军事密码学家Auguste Kerckhoff的思想。

安全模糊：企图使算法保密的做法，这是很难奏效而且以此为基础的密码体系非常脆弱和更换代价高昂。公开算法反过来可以通过社会力量验证算法的能力。

工作因子work factor：用于破解密码，即确认密钥所需要的时间指数

密码分析的问题种类，也是对应攻击方式的名称  
①唯密文ciphertext only（攻击）：只有密文供分析  
②已知明文known plaintext（攻击）：有匹配的明文和密文分析  
③选择明文chosen plaintext（攻击）：能够加密一些分析者自己选择的明文

8.1.2 置换密码subsitution cipher

定义：每个字母或每一组字母被另一个字母或另一组字母取代。最早的置换密码是凯撒密码Caesar cipher，它将所有的字母用其后第n个字母来替代，亦即将字母表整体移动n个字母。这种密码只需要枚举26种可能情况即可破解。

单字母置换monoalphabetic substitution：采用符号对符号置换策略的通用系统名称，一般而言替换字母的规律是26！种可能中的一种。然而仍然可以通过词频统计和语言特性对其进行唯密文攻击。

8.1.3 替代密码transposition cipher

定义：不改变字符本身，但重新对字母排序的加密法。密钥用一个不包含任何重复字母的单词或短语。具体加解密过程见p597

替代密码由于不隐藏字符本身，而且将字符改变位置的策略有一定规律，可以结合环境猜测一些明文词组，进而推测出密钥长度和顺序等，即已知明文攻击。不过破解难度也和密钥实际长度和明文长度等相关。

8.1.4 一次性密钥one-time pad

定义：随机产生一串序列作为密钥，与明文的位串（如ASCII）进行异或得到密文，再次异或即可解密。

优点：技术上或理论上无法破解

缺点：实用中需要对难于记忆的密钥进行保护（尤其是接收方和发送方必须对密钥同时掌握）。而且加解密操作需逐位进行，对于丢失字符或同步问题非常敏感。

理解：这种策略的名称有一定的误导性，本意是指密钥由随机数产生，具有“一次性”的感觉，实际上该策略的特点不只在于密钥（其他算法的密钥也不见得必须重复使用）也在于算法。密文由随机性产生，不包含词频等信息用于分析；另外试探性破解出的各种“明文”很难验证是否错误（会产生很多看似有效的明文）。

①量子密码学quantum cryptography：

BB84协议（Bennet，Brassard，1984）

主角pincipal：密码学案例中的用户

光子photon：量子密码学基于光子的偏振特性和信息的光纤传输，尤其是偏振光会等概率随机的偏振到与自己成45角的两个正交基的任何一个方向上；另外依靠现在的技术无法复制光子，所以入侵者对于光信息的操作受到非传统信息可比的限制。

量子位qubit：一次发送一个光子的数据位

保密增强privacy amplification：量子密码中，被动入侵者可以获取有效密钥的一部分内容，但是若发送接收方对已达成的密钥进行适当变换，其结果依赖于所有的输入位，将新结果作为正式密钥，则入侵者将完全失去密钥的内容。

8.1.5 两个基本的密码学原则

①冗余度

消息必须包含一定的冗余度，即明文消息必须包含一定的冗余信息来证明其有效性。

这种原则有助于阻止主动攻击者发送垃圾信息，但是反过来也有助于被动攻击者窃取信息，或验证窃取信息的有效性。

②新鲜度

需要采取某种方法来对抗重放攻击

可以使用时间戳

## 8.2 对称密钥算法symmetirc key algorithm

定义：使用同样的密钥完成加解密操作

块密码block cipher：明文信息块转化为等位长的密文块的算法

P盒P-box：用于进行一次替换操作的硬件元件，P代表permutation排列

S盒S-box：用于进行一次置换操作的硬件，S盒的内部实现一般包含一个P盒

乘积密码product cipher：将两种盒串并联形成的复杂的加密方式，仍然是块密码的一种。p604

8.2.1 DES——数据加密标准Data Encryption Standard

DES由IBM开发，基于公司的专利加密算法Lucifer，但是密钥长度从128减少至56位

原理：p605

白化whitening：一种特殊的加强DES强度的技术，实际上就是增加一些更加复杂的操作来增加密钥长度。

三重DES：使用加密-解密-加密EDE的方式进行加密，反之DED进行解密。对于加密过程，一般而言，两次加密使用相同密钥，中间解密使用另一个密钥，反过程类似。之所以这样设计是为了和单重DES兼容，只需将两个密钥设为同一个即可实现。

8.2.2 AES——高级加密标准Advanced Encryption Standard

美国国家标准及技术委员会NIST，National Institute of Standards and Technology

①Rijndael

即AES的美国政府标准，FIPS197联邦信息处理标准

两个变种：数据块128位，密钥128位；数据块128，密钥256位

算法过程简介p608

8.2.3 密码模式

①电码本模式ECB，Electronic Code Book

定义：无论多么复杂的算法，明文和密文一一对应。实际上也就是单纯地应用某些块密码算法对明文加密一次即结束。

缺点：一一对应的特点可以利用来破坏文件的内容，甚至篡改。

②密码块链模式cipher block chaining

定义：在仍然使用块密码对明文逐段加密的情况下，某一段的明文在被加密前先于上一段密文XOR，这样相同的明文一般将导致不同的密文。另一方面，这个方式也屏蔽了对密文进行一般意义上的修改（如整块拷贝覆盖）所造成的破坏。

初始矢量IV，Initialization vector：对第一段明文进行XOR的一个数据

③密码反馈模式cipher feedback mode

定义：用于替代块链模式在逐字节加密的条件下实现相似的功能，明文不是直接加密的（因为长度不够），而是和一段密文的一部分进行异或来输出自己的密文。也需要初始矢量来启动加密解密过程。见p612

缺点：某一位密文的传输错误影响的不仅仅是一位明文，而将会导致一个局部区域的错误。

④流密码模式stream cipher mode

定义：使用密钥循环加密一段初始矢量生成一个密钥流keystream，然后使用一次性密钥方法加密

优点：不会产生密码反馈模式的缺点，也不会产生电码本模式的缺点

密钥流重用攻击keystream reuse attack：先后使用两个同样的密钥流，造成入侵者可以获得两个明文的XOR结果，进而便于进行语言统计学分析。

⑤计数器模式counter mode

定义：类似于流密码模式，但是密钥流不是反复加密形成的，而是通过对初始矢量增加一个常量再加密形成的。

优点：如果用于磁盘存储大量加密数据，则计数器模式可以便于随机访问各个位置的信息，但其他加密方式（除了ECB）必须从头解密。而计数器模式也不会产生一一对应的明文密文关系。

8.2.4 其他密码模式

对称密钥算法简表p614

8.2.5 密码分析

4个密码分析领域的进展

①区分密码分析differential cryptanalysis

使用一对仅在少量位有差别的明文来加密进行分析

②线性密码分析linear cryptanalysis

③功率分析

④时间分析

后两种分析方法不基于算法本身，而是基于计算机执行代码时的表现。

## 8.3 公开密钥算法

公开密钥密码系统public key cryptography：加密算法和加密密钥均公开的系统，解密密钥显然与加密不同，即非对称密钥系统。发送方只需要知道接收方的公开加密密钥即可传输密文，解密方无需从发送方获取密钥即可解密。避免现实中密钥传输导致的安全问题。

公开密钥：即加密密钥

私有密钥：即解密密钥

公开密钥算法的设计要求：  
从加密密钥推断出解密密钥非常困难  
加密过程不能被选择明文破解  
显然的——加密解密后必须得到原明文

8.3.1 RSA

方法名称为发现人首字母Rivest，Shamir，Adleman，1978 p616

方法基于数论知识，以及分解大数的难度上

缺点：密钥必须很长才能保证安全，算法执行太慢，所以一般用于分发对称密钥算法的密钥

8.3.2 其他公开密钥算法

背包算法Merkle，Hellman 1978：不安全

公开密钥算法一般基于数学难题，这个难题是由公钥直接破解密文的基础。主要分两类：分解大数难度和以大素数为模来计算离散对数的难度。

## 8.4 数字签名

数字签名的必备条件：  
可以验证身份  
不可否认nonrepudiation  
不可伪造

8.4.1 对称密钥签名

原理：设立一个中心权威机构负责认证消息，所有用户的密钥被中心获知，中心使用不公开的自身密钥进行“盖章” p619

8.4.2 公开密钥签名

原理：只要公开密钥算法具备可用解密过程来“加密”明文的特性，就可以利用来作为签名。不需要中心机构支持。

业界常用的公开密钥签名算法是RSA

数字签名标准DSS，Digital Signature Standard：NIST推出的El Gamal算法为基础的公开密钥签名标准，原理不同，速度较慢，接受性不高

8.4.3 消息摘要message digest

定义：实际上是一个单向散列技术，用来将一串明文信息缩减至一个固定长度的位串，达到缩减有效信息同时仍然能够认证原消息的作用。这项技术相比于前面介绍的签名，单纯实现了认证功能而剥离了一般并不需要的保密功能。然而消息摘要使用时的协议并非不包含加密算法，加密算法执行在消息摘要上起到保护作用，只是不执行在明文或原文上。

特性：  
①散列函数很容易计算  
②根据结果反求明文原信息不可能  
③根据原文，再找到一个可以得出相同散列结果的信息不可能  
④原文信息哪怕只有1位变化，结果也会完全不同

安全散列算法1（SHA-1，Secure Hash Algorithm 1）：一种消息摘要函数，已有SHA-2版本，算法细节见p622

MD5：一种消息摘要函数，现已不够安全，即不满足特性第3条。细节见p623

8.4.4 生日攻击

原理：一种利用“生日匹配”概率问题作为指导思想的破解消息摘要的方法。p624

## 8.5 公钥的管理

公开密钥算法具有很多优点，然而公钥的管理仍然是个问题，虽然不怕泄露公钥但是仍怕公钥被篡改，亦即公钥的传输和确认同步仍然是个问题。

8.5.1 证书

证书在这里仍然是传统意义的证书，只是为了方便，它的消息摘要（散列值）也是证书。

密钥分发中心KDC，key distribution center：分发密钥的机构，这种体系不提倡

认证中心CA，Certification Authority：用于为公钥的证书（消息摘要）进行加密（公开密钥签名）的机构。该机构用其私钥为证书签名，其公钥用于公众查看认证的结果（散列值），公众通过对比被认证方提供的证书明文产生的散列值和认证结果的散列值，即可知道其有效性。

应用：  
①用于确认公钥的正确性和所有权，如前所述  
②用于确认证书中的公钥的所有者满足一定标准条件  
具备证书声明条件的人，一定具有证书中公钥所匹配的私钥，通过发送加密信息并要求回执即可确认  
③用于面向对象的分布式系统确定用户类的权限

8.5.2 X.509

X.509的IETF版本由RFC5280描述。是一个专门针对证书格式的标准。

X.509标准证书的基本字段p628

抽象语法标记1（ASN.1 Abstract Syntax Notation 1）：OSI的一个编码标准，被X.509采用作为证书的编码

8.5.3 公钥的基础设施

公开密钥基础设施PKI，Public Key Infrastructure：另一种证明公钥身份的方法，用于缓解CA认证的压力和安全问题。大体思想是采用树结构将CA职能逐级下放，上一级CA只证明相邻下一级CA的合法性；而根CA同时也不只采用单一的机构，而是根据地区和政治因素分担的。

区域管理机构RA，Regional Authority：地方CA的别称

信任链chain of trust或证书路径certification path：即逐级证明合法性的证书链

信任锚trust anchor：根CA的别称

①目录

即目录服务器，专门用于管理所有的X.509证书，便于查询

另一种设想是通过DNS查询来完成，即返回查询结果包括IP地址以及证书

②撤销

即对于证书因为各种原因需要撤回的相关问题的处理。

证书撤销列表CRL，Certificate Revocation List：一个动态更新的列表，存有尚未过期但是已被撤销的证书信息。

## 8.6 通信安全

8.6.1 IPSec

因特网的安全性被实现在网络层之上，便于大部分用户的使用，尤其是对于安全性一无所知的用户。

IP安全IP Security：是一种设计方案，RFC2401, 2402, 2406。其所有的服务都建立在对称密钥系统上，因其性能较快。同时支持多算法、多粒度。

空算法null algorithm：IPSec不支持不加密的选项，但是可以通过选择空算法来实现

安全关联SA，security association：IPSec环境下，一个点到点之间的“连接”，是单向的。若两个方向的安全通信需要两个SA

安全关联及密钥管理协议ISAKMP，Internet Security Association and Key Management Protocol：处理建立密钥的框架，其核心是Internet密钥交换（IKE，Internet Key Exchange）。

IPSec两种模式：传输模式transport mode，IPSec头插入到IP头后；隧道模式tunnel mode，IP数据包被封装在一个新的IP包中，新包的头负责携带安全信息

流量分析traffic analysis：研究数据包的流模式

IPSec含有两个新的头，这些头可以独立加入到数据包中以携带安全信息。

认证头AH，Authentication Header：p634/648  
下一个头Next header  
有效载荷长度Payload length  
安全参数索引Security parameters index  
序号Sequence number  
认证数据Authentication data：散列消息认证码HMAC，Hashed Message Authentication Code

封装的安全有效载荷ESP，Encapsulating Security Payload：另一个IPSec头p635/649  
ESP将会逐渐取代AH，因为它兼具保密和认证功能。

8.6.2 防火墙Firewall

防火墙本质上就像一个包过滤器packet filter，实质上是一个局域网唯一通向外界网络的路由器。

非军事化区域DMZ，DeMilitarized Zone：一个位于内部网络防火墙之外，但是仍属于同一个局域网的网络外围缓冲区。这个区域可以设置局域网的Web服务器和邮件服务器，因此不会阻断来自外界的访问。具体如何设置DMZ和Firewall显然取决于局域网的本地策略。

状态防火墙Stateful firewalls：能够跟踪连接状态的防火墙，可以实现一些更加细化的过滤规则。

应用层网关application-level gateway：为了让防火墙能够更加细致的工作，为其增加了能够窥视上层信息的功能。虽然违反了协议分层，但是这是目前为了安全的唯一解法。

拒绝服务攻击DoS，Denial of Service：通过短时间内发送大量数据包将服务器导致瘫痪或无法接收新请求。

分布式拒绝服务DDoS，Distributed Denial of Service：将攻击方分布至很多地点

8.6.3 虚拟专用网络VPN，Virtual Private Networks

一种建立在公共网络Internet连接之上的层叠网络或一种概念上功能上和专用网络一致的网络。一般由防火墙，VPN局域网，ESP安全隧道模式组成，实质上就是利用互联网进行远程数据传输，但是利用安全措施将数据和Internet上其他信息“隔离”，以此模拟专用网络的结构。

8.6.4 无线安全性

①802.11安全性

无线安全和有线安全不同，无线接入点向四面八方传输信息。

有线等效保密WEP，Wired Equivalent Privacy：802.11的安全协议，可破解

WiFi保护接入2WPA2，WiFi Protected Access 2：无线数据链路层安全协议，防止无线节点之间干扰或检查与其不相关的传输。802.11i

802.1X：一个协议，用于接入点让客户端和认证服务器对话并检查结果，WPA2的一部分

可扩展的身份验证协议EAP，Extensible Authentication Protocol：RFC3748，规定客户端如何和认证服务器互动的协议框架。

WPA2用于大型企业时可使用认证服务器，家用时只需一个共享密码。后者不同的客户端之间可以相互破解出加密密钥（因为密码相同），前者每个客户端的流量加密密钥完全不可推导。

四次包four-Packet握手：p641/655客户端与接入点建立会话密钥的方式。客户端的主密钥（由密码推导）并不直接用于加密数据，而是用来和其他信息计算出临时会话密钥来加强保护。

消息完整性检查MIC，Message Integrity Check：为了避免和MAC术语冲突，该术语通常用来替代网络协议。和HMAC术语概念类似。

临时密钥完整性协议TKIP，Temporary Key Integrity Protocol：802.11i使用密钥提供安全服务的协议之一。已破解

具有密码块链的计数器模式消息认证码协议Counter Mode with Cipher block chaining Message authentication code Protocol，CCMP：另一种802.11i安全协议，是规定如何使用会话密钥来加密信息、使用何种密码模式等的协议。

②蓝牙安全性

万能密钥passkey：蓝牙设备中的预设密钥或硬置密钥，用于配对设备

加密使用E0流密码，完整性控制使用SAFER+

## 8.7 认证协议

8.7.1 基于共享密钥的认证

共享密钥：双方在其他安全的通讯方式中预先建立的一个共同密钥

质询-回应challenge-response：一方给另一方发送一个随机数，反过来另一方返回随机数的一个特殊替代

反射攻击reflection attack：打开多个并行会话，通过将一个会话的内容在另一个会话中“反射”回去，达到攻破协议的目的

认证协议设计的4条通用规则：p645/659

基于HMAC的共享密钥认证协议（有效）p647/661

8.7.2 建立共享密钥：Diffie-Hellman密钥交换key exchange

密钥交换：用于在双方无法预先在安全条件下建立共享密钥的情况

中间人攻击man-in-the-middle attack：也称水桶队列攻击bucket brigade attack（所有人站成一排逐个传递水桶），攻击者截获双方的消息，并且和双方同时建立会话密钥。

8.7.3 使用密钥分发中心的认证

重放攻击replay attack：攻击者发送有效的但是重复的消息给被攻击方达到目的，如银行转账。

Needham-Schroeder认证协议：一种多路径的质询-回应协议

Otway-Rees协议：可防止重放攻击的一种KDC认证协议p651/665

8.7.4 使用Kerberos的身份认证

Kerberos第5版是广泛用于业界的认证协议。

认证服务器AS：用户登录过程中验证用户的身份，虽然与每个用户共享一个秘密口令，但是并不是完整意义上的KDC。可以理解为TGS为用户分发另一个用户的票据和会话密钥，而AS为用户分发TGS的票据以及和TGS联络的会话密钥。用户和AS之间不需要创建双向会话密钥。

票据发放服务器TGS，Ticket-Granting Server：发放身份票据的证明

票据：在这个系统中的含义和作用与签名类似，都是对一个包含若干信息的数据结构的加密，唯有对应方才能解密，以确认彼此身份。

对于不同的域realm之间的AS和TGS，采用彼此认证的方式联络，即不同域的TGS相互成为彼此的认证对象，传递票据。

8.7.5 使用公开密钥密码学的认证

协议简述p653/667

## 8.8 电子邮件安全性

8.8.1 PGP——良好的隐私性Pretty Good Privacy

国际数据加密算法IDEA，International Data Encryption Algorithm：一种块密码算法，类似于DES和AES

PGP流程p656/670

PGP邮件格式结构p657/671

私钥环private key ring：用户在本地维护的自身的多个公钥私钥对

公钥环public key ring：用户在本地保存的其他用户的公钥

8.8.2 S/MIME

安全的MIME，Secure/MIME：IETF开发，RFC2632-2643

增强保密邮件PEM，Privacy Enhanced Mail

## 8.9 Web安全性

8.9.1 威胁

骇客cracker：一般意义上的Web攻击者

黑客hacker：也可以做骇客的意思，也有超级程序员的意思（非贬义）

8.9.2 安全命名

DNS欺骗DNS spoofing：攻击DNS服务器以篡改其条目，达到刻意引导其他用户至篡改站点的攻击。

染毒缓存poisoned chache：存放了被篡改条目的缓存

DNS安全DNSsec，DNS Security：RFC2535，致力于让DNS更加安全的项目名称。3个基本服务p662/676

资源记录集RRSet，Resource Record Sets：DNSsec中管理多个DNS记录的集合，通过加密和摘要保证安全性

DNSsec新的记录类型：KEY，SIG，CERT等

8.9.3 SSL——安全套接层

安全套接字层SSL，Secure Sockets Layer：一种安全软件应用，功能包括p664/678

安全的HTTP，HTTPS，Secure HTTP：在SSL之上使用的HTTP

SSL连接建立子协议：p665/679

SSL传输数据子协议：p666/680

预设主密钥premaster key：用于和临时值一起推导会话密钥

传输层安全TLS，Transport Layer Security：RFC5246，基于SSL第三个版本的、由IETF标准化的安全应用。二者不完全兼容，因此以SSL/TLS的形式存在。

8.9.4 移动代码安全性

移动代码mobile code：Web页面包含的脚本程序代码

①Java Applet安全性

沙箱sandbox：解释器对不信任的小程序applet在隔离沙箱中运行，监视其调用其他资源的行为

特点：从用户端来控制的机制理论上更为安全，不能查询出代码发行源头，但出错率较高

②ActiveX

代码签名code signing：即针对代码本身进行散列签名的安全机制

认证码authenticode：微软用于验证ActiveX控件的系统

特点：可以查询代码源头，但是只能依赖源头是否可信，用户端无法现场控制

③JavaScript

不安全

④浏览器扩展Extension

⑤病毒

略

## 8.10 社会问题

8.10.1 隐私

加密芯片clipper chip

密钥托管key escrow

①匿名邮件转发器anonymous remailer

将发信人真实身份转换为假身份的服务器，但是可能保有真实身份信息

加密朋克邮件转发器cypherpunk remailer

洋葱路由onion routing：一个节点只知道链中的下一个节点，而不是知道最终的目的地，以此保护web请求方的身份。

8.10.2 言论自由

永恒服务eternity service：上传材料一经上传无法被任何人包括上传者更改的服务

①信息隐藏学steganography

将信息压缩、加密并且隐藏在表观的图像或音频等信息中的技术

数字水印watermarking：在具有版权的图像上隐藏版权声明信息

8.10.3 版权copyright

知识产权IP，Intellectual Property

数字千禧年版权法案DMCA，Digital Millennium Copyright Act

DMCA删除通告DMCA takedown notices

公平使用原则fair use doctrine

可信计算组TCG，Trusted Computing Group

可信平台模块TPM，Trusted Platform Module：在操作系统更低层次上限制用户行为的模块