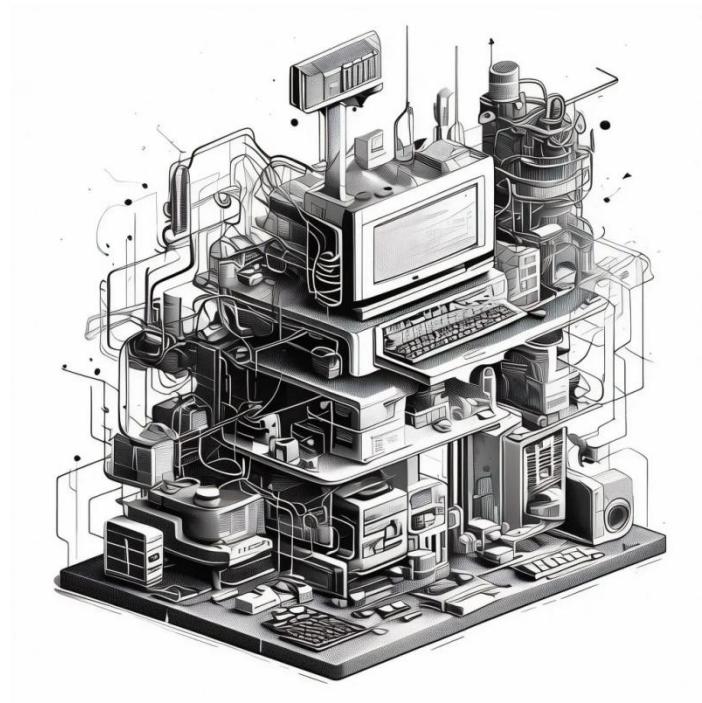


408 考研|计算机网络 知识点归纳总结



本文档适用于考前复习查漏补缺
和考场前快速回顾知识点使用

目录

第 1 章 计算机网络体系结构	1
1.1 计算机网络概述	1
·计算机网络的定义	1
·计算机网络的组成	1
·计算机网络的功能	1
·计算机网络的分类	1
·性能指标（速率、时延、利用率等）	2
*计算机中 KB 与 kb 的换算	2
*局域网与广域网的互联 P8T12 P8T16	2
1.2 计算机网络体系结构与参考模型	2
·PCI+SDU=PDU	2
·协议、接口、服务的概念	3
·网络体系结构	3
·ISO/OSI 参考模型	3
·TCP/IP 参考模型	4
·OSI 和 TCP/IP 差别	4
·五层参考模型	5
*服务访问点 P22 T19	5
*不同层的设备 P23 T25	5
第 2 章 物理层	6
2.1 通信基础	6
·基本概念（信源、信宿、信道）	6
·通信方式	6
·数据传输方式（串行/并行）	6
·同步/异步传输	6
·码元、波特率	6
·影响失真的因素	7
·奈氏准则	7
·香农定理	7
·奈奎斯特定理与香农定理的对比	7
·带宽	7
·基带信号/宽带信号	7
·数字数据编码为数字信号	8
·模拟数据编码为数字信号	9
·数据交换方式（电路交换、报文交换、分组交换）	9
·虚电路服务	10
*虚电路分类 P42 T29	10
2.2 传输介质	11
·导向性传输介质	11
·非导向性传输介质	11
2.3 物理层设备	11
·中继器	11
·集线器	11
第 3 章 数据链路层	12
3.1 数据链路层的功能	12

·为网络层提供服务	12
·链路管理	12
·组帧（帧定界、帧同步、透明传输）	12
·流量控制	12
·差错控制	12
*三个基本问题：封装成帧、透明传输、差错检测	12
3.2 组帧	13
·字符计数法	13
·字符填充的首尾定界符法	13
·零比特填充法	13
·违规编码法	13
3.3 差错控制	13
·差错	13
·检错编码（奇偶校验码、循环冗余码 CRC）	13
·纠错编码（海明码）	14
*海明距离与检错纠错 P71 T5	15
3.4 流量控制与可靠传输机制	15
·流量控制	15
·可靠传输	15
·停止-等待协议	15
·后退 N 帧协议 GBN	15
·选择重传协议 SR	15
·信道利用率和信道吞吐率	15
3.5 介质访问控制	16
·介质访问控制 MAC, Medium Access Control	16
·信道划分介质访问控制（多路复用技术）	16
·随机访问介质访问控制（ALOHA 协议、CSMA 协议、CSMA/CD 协议、CSMA/CA 协议）	17
·轮询访问介质访问控制（令牌传递协议）	18
3.6 局域网 LAN, Local Area Network	18
·局域网的基本概念和体系结构	18
·以太网的基本概念、传输介质与高速以太网	19
·网卡与 MAC 地址	19
·以太网的 MAC 帧	20
·无线局域网 IEEE 802.11	20
·虚拟局域网 VLAN, Virtual LAN	21
*放大器与中继器 P111 T4	22
*重复硬件地址 P112 T9	22
3.7 广域网	22
·广域网基本概念	22
·PPP (Point-to-Point Protocol) 协议	22
*PPP 协议认证 P120T6	23
3.8 数据链路层设备	23
·交换机	23
第 4 章 网络层	25
4.1 网络层的功能	25
·异构网络互连	25
·路由与转发	25
·软件定义网络 SDN 的基本概念	25

·拥塞控制	26
4.2 路由算法	26
·静态路由与动态路由	26
·距离-向量路由算法	26
·链路状态路由算法	26
·层次路由	27
*路由回路的根本原因 P142 T5	27
4.3 IPv4	27
·IPv4 分组	27
·IPv4 地址	28
·私有 IP 与网络地址转换 NAT	29
·子网划分与子网掩码, 无分类编址 CIDR 与链路聚合	30
·TCP/IP 协议栈	30
·地址解析协议 ARP, Address Resolution Protocol	30
·动态主机配置协议 DHCP, Dynamic Host Configuration Protocol	31
·网际控制报文协议 ICMP, Internet Control Message Protocol	31
4.4 IPv6	32
·IPv6 的主要特点	32
·IPv6 地址	33
4.5 路由协议	34
·自治系统 AS, Autonomous System	34
·域内路由与域间路由	34
·路由信息协议 RIP, Routing Information Protocol	34
·开放最短路径优先 OSPF 协议	35
·外部网关协议 BGP, Border Gateway Protocol	36
·三种路由协议的比较	36
4.6 IP 组播	37
·组播的概念	37
·组播地址	37
·网际组管理协议 IGMP, Internet Group Management Protocol	38
*组播路由避免路由环路 P194 T2	38
4.7 移动 IP	38
·移动 IP 相关概念	38
·移动 IP 通信过程	38
4.7 网络层设备	39
·冲突域和广播域	39
·路由器的组成和功能	39
·路由表与路由转发	39
第 5 章 传输层	41
5.1 传输层提供的服务	41
·传输层的功能	41
·传输层的寻址与端口	41
*各层服务访问点	41
·无连接服务 UDP 与面向连接服务 TCP	42
5.2 UDP 协议	42
·UDP 数据报特点	42
·UDP 数据报格式	42
·UDP 校验	42

5.3 TCP 协议	43
·TCP 特点	43
·TCP 报文段	44
·TCP 连接管理	45
·TCP 可靠传输	46
·TCP 流量控制	47
·TCP 拥塞控制	47
第 6 章 应用层	49
6.1 网络应用模型	49
·客户/服务器模型 C/S	49
·对等连接 P2P 模型	49
6.2 域名系统 DNS, Domain Name System	49
·DNS 概念	49
·层次域名空间	49
·域名服务器	50
·域名的解析过程	50
6.3 文件传输协议 FTP, File Transfer Protocol	51
·FTP 概念与特点	51
·控制连接和数据连接	51
6.4 电子邮件 E-mail	52
·电子邮件系统的组成结构	52
·电子邮件格式	53
·多用途网际邮件扩充 MIME, Multipurpose Internet Mail Extensions	53
·简单邮件传输协议 SMTP, Simple Mail Transfer Protocol	53
·邮局协议 POP, Post Office Protocol	54
·因特网报文存取协议 IMAP	54
*POP3 传输密码 P265T7	54
6.5 万维网 WWW, World Wide Web	54
·WWW 的概念与组成结构	54
·超文本传输协议 HTTP	55
*HTTP 1.0 P273 T6	56
*HTTP 请求报文中的 Connection 和 Cookie P273 T12	56

第1章 计算机网络体系结构

1.1 计算机网络概述

- 计算机网络的定义

广义观点：计算机网络是能实现远程信息处理的系统或进一步达到资源共享的系统。

资源共享观点：计算机网络是“以能够相互共享资源的方式互连起来的自治计算机系统的集合”

目的——资源共享

组成单元——分布在不同地理位置的多台独立的“自治计算机”

网络中的计算机必须遵循的统一规则——网络协议

用户透明性观点：计算机网络是一个能为用户自动管理资源的网络操作系统，它能够调用用户所需要的资源，完成相应的工作。

- 计算机网络的组成

按组成部分分：

硬件、软件、协议

按工作方式分：

边缘部分：用户直接使用 通信和资源共享

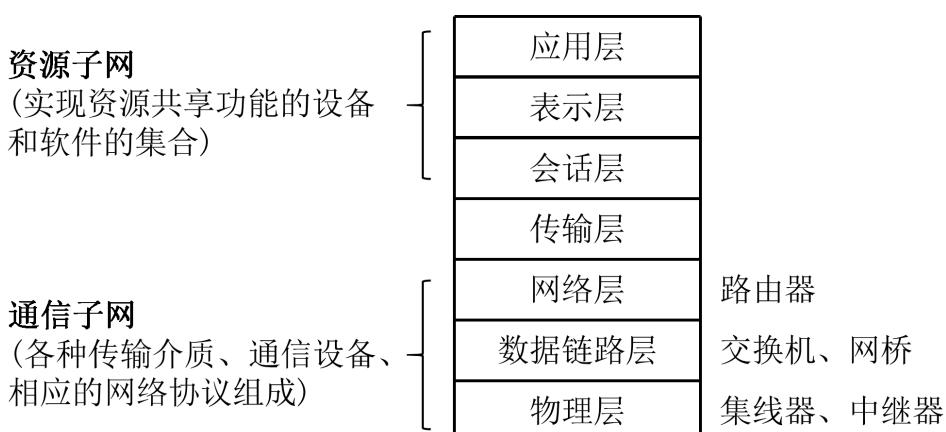
C/S 方式，B/S 方式，P2P 方式

核心部分：为边缘部分服务 提供连通性和交换服务

按功能组成分：

通信子网：实现数据通信

资源子网：实现资源共享/数据处理



- 计算机网络的功能

数据通信（最基本，最重要）

资源共享（软件，数据和硬件资源）

分布式处理

提高可靠性

负载均衡

- 计算机网络的分类

按分布范围：

广域网 WAN（交换技术）、城域网 MAN、局域网 LAN（广播技术）、个人区域网 PAN

按传输技术：

广播式网络（共享公用通信信道）

点对点网络（使用分组储存转发和路由选择机制）

按拓扑结构：

总线形、星形、环形、网状

按使用者：

公用网、专用网

按交换技术:

电路交换

过程: 建立连接、传输数据、断开连接

典型: 传统电话网络

优点: 直接传送、时延小

缺点: 线路利用率低、不能充分利用线路容量、不便于进行差错控制

报文交换

封装报文, 储存转发

分组交换

又称为包交换网络, 固定长度的数据块

• 性能指标 (速率、时延、利用率等)

速率 (数据率、数据传输率、比特率)

单位: b/s, kb/s, Mb/s, Gb/s, Tb/s

*计算机中 KB 与 kb 的换算

数据速率换算 10^3 $1\text{kb}/\text{s} = 10^3 \text{b}/\text{s}$

存储容量换算 2^{10} $1\text{KB} = 1024\text{B}$

带宽 见 2.1 带宽

原指频带宽度, 即最高频率和最低频率之差, 单位赫兹 Hz。该处指最高数据率

单位: b/s, kb/s, Mb/s, Gb/s, Tb/s

吞吐量

单位时间内通过某个网络的数据量

单位: b/s, kb/s, Mb/s, Gb/s, Tb/s

时延

数据从网络一端传送到另一端所需时间, 单位 s

发送时延

发送时延=数据长度 (b) /发送速率 (b/s)

传播时延

传播时延=信道长度 (m) /电磁波在信道上的传播速率 (b/s)

排队时延

处理时延

时延带宽积

时延带宽积=传播时延×带宽

往返时延 RTT

从发送方发送数据开始到收到接收方确认的时间

往返传播时延 RTT=传播时延×2+末端处理时间

利用率

信道利用率

信道利用率=有数据通过的时间/(有+无) 数据通过时间

网络利用率

信道利用率加权平均值

*局域网与广域网的互联 P8T12 P8T16

中继器和桥接器用于局域网的物理层和数据链路层的联网设备

局域网接入广域网通过路由器来实现

局域网工作在数据链路层

1.2 计算机网络体系结构与参考模型

- PCI+SDU=PDU

实体

第 n 层中的活动元素称为 n 层实体，该实体指任何可发送或接收信息的硬件或软件进程
同一层的实体叫做对等实体，n 层实体实现的服务为 n+1 层所利用

服务数据单元 SDU, Service Date Unit

为完成用户所需的功能而应传送的数据

协议控制信息 PCI, Protocol Control Information

控制协议操作的信息

协议数据单元 PDU, Protocol Data Unit

对等层次之间传送的数据单元

物理层的 PDU: 比特

链路层的 PDU: 帧

网络层的 PDU: 分组

传输层的 PDU: 报文

协议控制信息 PCI + 服务数据单元 SDU = 协议数据单元 PDU

• 协议、接口、服务的概念

协议 Network Protocol

为进行网络中的对等实体数据交换而建立的规则、标准或约定称之为网络协议【水平】

接口（典型：访问服务点 SAP, Service Access Point）

同一结点内相邻两层间交换信息的连接点，上层使用下层服务的入口

服务

下层为相邻上层提供的功能调用【垂直】

· 网络体系结构

网络体系结构是从功能上描述计算机网络结构

计算机网络体系结构简称网络体系结构是分层结构

每层遵循某个/些网络协议以完成本层功能

计算机网络体系结构是计算机网络的各层及其协议的集合。

第 n 层在向 n+1 层提供服务时，此服务包含第 n 层本身的功能和由下层服务提供的功能

仅仅在相邻层间有接口，且所提供的服务的具体实现细节对上一层完全屏蔽

体系结构是抽象的，而实现是指能运行的一些软件和硬件

• ISO/OSI 参考模型

七层上四层端到端 下三层点到点

OSI 七层模型 各层传输单位

7	应用层	数据	端到端
6	表示层		
5	会话层		
4	传输层	报文段或用户数据报	点到点
3	网络层	数据报	
2	数据链路层	帧	
1	物理层	比特流	

应用层

为用户的应用程序提供各种网络服务

协议：文件传输 FTP、电子邮件 SMTP、万维网 HTTP

表示层

用于处理两个通信系统中交换信息的表示方式（语法和语义）

功能：数据格式变换，数据加密解密，数据压缩和恢复
会话层

向表示层实体/用户进程提供建立链接并在连接上有序地传输数据
建立同步 SYN

功能：建立、管理、终止会话

使用校验点使会话在通信失效时从校验点/同步点继续恢复通信，实现数据同步

协议：ADSP、ASP

传输层

负责主机中两个进程的通信，即端到端的通信。

传输单位：报文段 TCP 或用户数据报 UDP

功能：差错控制，流量控制，复用分用

协议：TCP、UDP

网络层

把分组从源端传到目的端，为分组交换网上的不同主机提供通信服务。

传输单位：数据报

功能：路由选择，流量控制，差错控制，拥塞控制

协议：IP、IPX、ICMP、IGMP、ARP、RARP、OSPF

数据链路层

把网络层传下来的数据报组装成帧。

传输单位：帧

功能：成帧、差错控制（帧错+位错），流量控制，访问（接入）控制，控制对信道的访问

协议：SDLC、HDLC、PPP、STP

物理层

在物理媒体上实现比特流的透明传输。

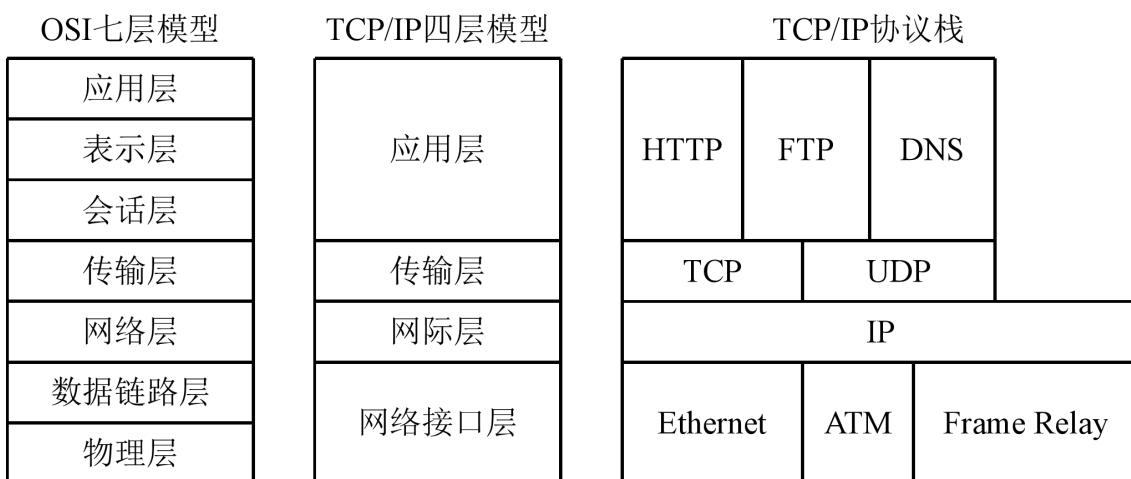
传输单位：比特

功能：定义接口特性、传输模式、传输速率，比特同步，比特编码

协议：RJ45、802.3

• TCP/IP 参考模型

四层 事实上的国际标准



• OSI 和 TCP/IP 差别

1.OSI 定义三点：服务、协议、接口

2.OSI 先出现，参考模型先于协议发明，不偏向特定协议

3.TCP/ IP 设计之初就考虑到异构网互联问题，将 IP 作为重要层次

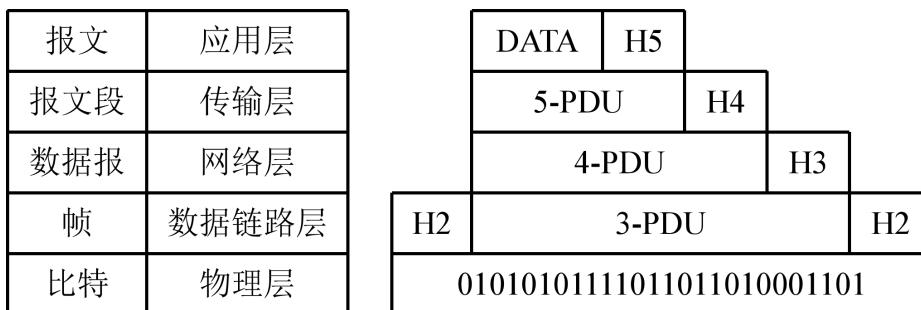
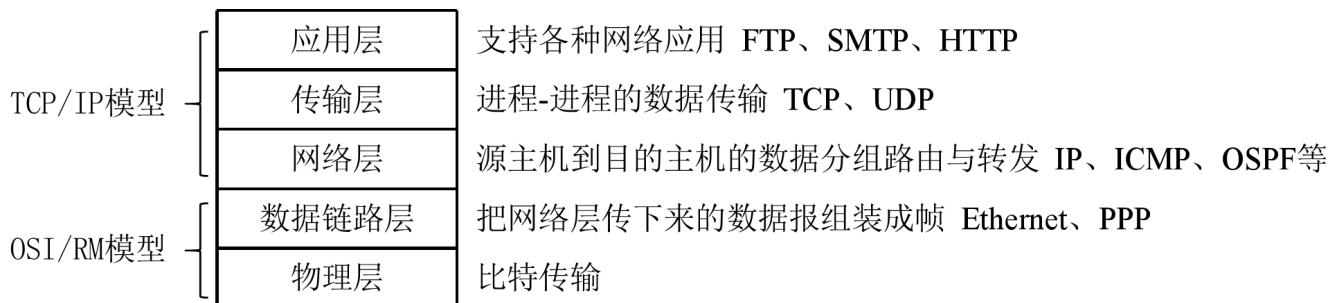
4.如下图所示

	ISO/OSI参考模型	TCP/IP参考模型
传输层	面向连接	无连接+面向连接
网络层	无连接+面向连接	无连接

面向连接：建立连接、数据传输、释放连接

无连接：进行数据传输

- 五层参考模型



*服务访问点 SAP P22 T19

物理层 “网卡接口”

数据链路层 “MAC 地址（网卡地址）”

网络层 “IP 地址（网络地址）”

传输层 “端口号”

应用层 “用户界面”

*不同层的设备 P23 T25

网络层 路由器

数据链路层 以太网交换机

物理层 集线器

第2章 物理层

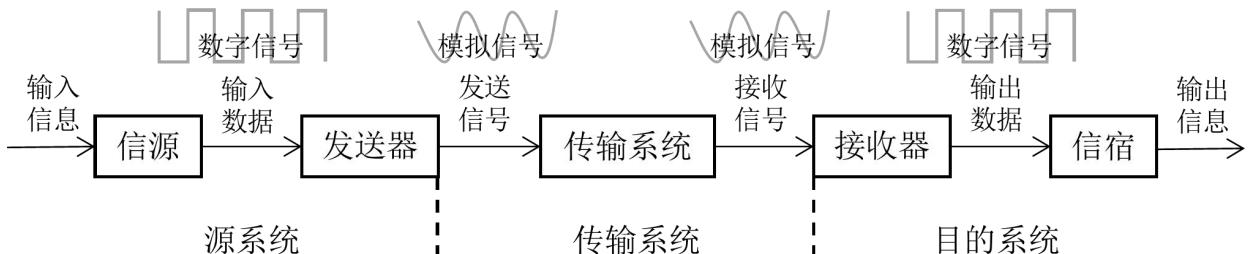
在物理媒体上实现比特流的透明传输。传输单位是比特

功能：定义接口特性，传输模式，传输速率，比特同步，比特编码

协议：RJ45、802.3

2.1 通信基础

- 基本概念（信源、信宿、信道）



数据 data: 传送信息的实体，通常是有意义的符号序列

信号 signal: 数据的电气/电磁的表现，是数据在传输过程中的存在形式

数字信号/离散信号：代表消息的参数的取值是离散的

模拟信号/连续信号：代表消息的参数的取值是连续的

信源: 产生和发送数据的源头

信宿: 接收数据的终点

信道: 信号的传输媒介。一般用来表示向某一个方向传递信息的介质

因此一条通信线路往往包含一条发送信道和一条接收信道

按传输信号分类：模拟信道、数字信道

按传输介质分类：有线信道、无线信道

- 通信方式

单工通信：

只有一个方向的通信而没有反方向的交互，仅需要一条信道

半双工通信/双向交替通信：

通信的双方都可以发送或接收信息，但任何一方都不能同时发送和接收，需要两条信道

全双工通信/双向同时通信：

通信双方可以同时发送和接受信息，需要两条信道

- 数据传输方式（串行/并行）

串行传输：将表示一个字符的 8 位二进制数按由低位到高位的顺序依次发送

速度慢，费用低，适合远距离

并行传输：将表示一个字符的 8 位二进制数同时通过 8 条信道发送

速度快，费用高，适合近距离

- 同步/异步传输

同步传输：在同步传输的模式下，数据的传送是以一个数据区块为单位，因此同步传输又称为区块传输。在传送数据时，需先送出 1个或多个同步字符，再送出整批的数据。

异步传输：异步传输将比特分成小组进行传送，小组可以是 8 位的 1 个字符或更长。发送方可以在任何时刻发送这些比特组，而接收方不知道它们会在什么时候到达。传送数据时，加一个字符起始位和一个字符终止位。

- 码元、波特率

码元：指用一个固定时长的信号波形（数字脉冲）代表一位 k 尽职数字，代表不同离散数值的基本波形，是数字通信中数字信号的计量单位，而该时长称为码元宽度。

波特率：又称码元传输速率，它表示单位时间内数字通信系统所传输的码元个数，也可以称为脉冲个数或者信号变化的次数单位是波特（Baud），1 波特表示每秒传输 1 个码元。

- 影响失真的因素

码元传输速率、信号传输距离、噪声干扰、传输媒体质量

- 奈氏准则

在理想低通（无噪声，带宽受限）条件下，为了避免码间串扰，极限码元传输速率为 $2WBaud$

$$\text{理想低通信道下的极限数据传输率} = 2W\log_2 V \text{ (单位为 b/s)}$$

其中 W : 理想低通信道的带宽, 单位是 Hz V : 离散电平数目

1. 在任何信道中, 码元传输的速率是有上限的。若传输速率超过此上限, 就会出现严重的码间串扰问题, 使接收端对码元的完全正确识别成为不可能。
2. 信道的频带越宽 (即能通过的信号高频分量越多), 就可以用更高的速率进行码元的有效传输。
3. 奈氏准则给出了码元传输速率的限制, 但并没有对信息传输速率给出限制。
4. 由于码元的传输速率受奈氏准则的制约, 所以要提高数据的传输速率, 就必须设法使每个码元能携带更多个比特的信息量, 这就需要采用多元制的调制方法。

- 香农定理

在带宽受限且有噪声的信道中, 为了不产生误差, 信息的数据传输速率有上限值

$$\text{信道的极限数据传输率} = W\log_2(1 + S/N) \text{ (b/s)}$$

其中 W : 信道的带宽 Hz S : 信道所传输信号的平均功率 N : 信道内部的高斯噪声功率

信噪比=信号的平均功率/噪声的平均功率, 常记为 S/N , 并用分贝 (dB) 作为度量单位

$$\text{信噪比} = 10\log_{10}(S/N) \text{ (dB)}$$

1. 信道的带宽或信道中的信噪比越大, 则信息的极限传输速率就越高
2. 对一定的传输带宽和一定的信噪比, 信息传输速率的上限就确定了
3. 只要信息的传输速率低于信道的极限传输速率, 就一定能找到某种方法来实现无差错的传输
4. 香农定理得出的为极限信息传输速率, 实际信道能达到的传输速率要比它低不少
5. 若信道带宽 W 或信噪比 S/N 没有上限 (不可能), 则信道的极限信息传输速率也就没有上限

- 奈奎斯特定理与香农定理的对比

1. 奈奎斯特定理只是给出了在无噪声情况下码元的最大传输速率。但是奈奎斯特定理并没有给出极限数据传输速率, 从概念上讲可以让每个码元具有无穷多离散电平, 那么就可以在有限码元速率的情况下, 让数据传输率无限大。
2. 香农定理证明了要使信息的极限传输速率提高, 就必须提高信道的带宽或信道中的信噪比。换句话说, 只要信道的带宽或信道中的信噪比固定了, 极限传输速率就固定了。
3. 只要信息的传输速率低于信道的极限传输速率, 就一定能找到某种方法来实现无差错的传输。
4. 实际信道上能够达到的信息传输速率要比香农的极限传输速率低。
5. 考试中, 如果两个公式都能用, 那么实际的数据传输率是二者中的较小者。

- 带宽

模拟信号系统中: 最高频率和最低频率间的差值就代表了系统的通频带宽

单位为赫兹 Hz, 目前仅在奈氏准则和香农定理中使用

数字设备中: 表示在单位时间内从网络中的某一点到另一点所能通过的“最高数据率”/单位时间内通过链路的数量, 常用来表示网络的通信线路所能传输数据的能力

单位是比特每秒 bps

- 基带信号/宽带信号

基带信号: 将数字信号 1 和 0 直接用两种不同的电压表示, 在数字信道上去传输 (基带传输) 像计算机输出的代表各种文字或图像文件的数据信号都属于基带信号。基带信号就是发出的直接表达了要传输的信息的信号, 比如我们说话的声波就是基带信号

宽带信号: 将基带信号进行调制后形成的频分复用模拟信号, 在模拟信道上去传输 (宽带传输)。把基带信号经过载波调制后, 把信号的频率范围搬移到较高的频段以便在信道中传输 (即仅在一段频率范围内能够通过信道)

距离较近时，采用基带传输方式（近距离衰减小，从而信号内容不易发生变化）

距离较远时，采用宽带传输方式（远距离衰减大，即使信号变化大也能最后过滤出来基带信号）

- 数字数据编码为数字信号

归零编码 RZ:

高 1 低 0 自同步

信号电平在一个码元之内都要恢复到零

非归零编码 NRZ:

高 1 低 0

容易实现，没有检错功能，无法判断一个码元的开始和结束，收发双方难以保持同步

反向非归零编码 NRZI:

保持 1 翻转 0

能传输时钟信号又尽量不损失系统带宽

曼彻斯特编码:

前高后低 1 前低后高 0

位中间的跳变既作时钟信号（可用于同步）又作数据信号

频带宽度是原始的基带宽度的两倍，数据传输速率只有调制速率的 1/2

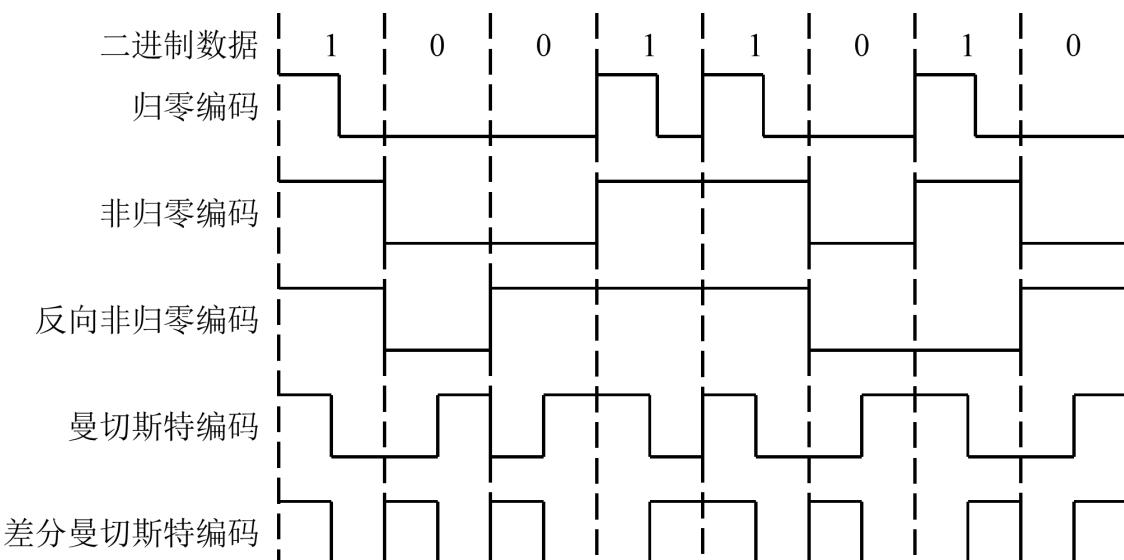
差分曼彻斯特编码:

同 1 异 0

每个码元中间有一次电平的跳转，可以实现自同步，且抗干扰性强于曼彻斯特编码

4B/5B 编码:

比特流中插入额外的比特以打破一连串的 0 或 1，用 5 个比特编码 4 个比特的数据，编码效率为 80%，只采用 16 种对应 16 种不同的 4 位码，其他的 16 种作为控制码（帧的开始和结束，线路的状态信息等）或保留



- 数字数据调制为模拟信号

幅移键控 ASK: 通过改变载波信号的振幅来表示数字信号 1 和 0，而载波的频率和相位都不改变。比较容易实现，但抗干扰能力差

频移键控 FSK: 通过改变载波信号的频率来表示数字信号 1 和 0，而载波的振幅和相位都不改变。容易实现，抗干扰能力强，目前应用较为广泛

相移键控 PSK: 通过改变载波信号的相位来表示数字信号 1 和 0，而载波的振幅和频率都不改变。它又分为绝对调相和相对调相

正交振幅调制 QAM: 在频率相同的前提下，将 ASK 与 PSK 结合起来，形成叠加信号。设波特率为 B，采用 m 个相位，每个相位有 n 种振幅

$$\text{数据传输率 } R = B \log_2(mn)$$

- 模拟数据编码为数字信号

典例：对音频信号进行编码的脉码调制 PCM

步骤：

抽样：对模拟信号周期性扫描，把时间上连续的信号变成时间上离散的信号

$$f_{\text{采样频率}} \geq 2f_{\text{信号最高频率}}$$

量化：把抽样取得的电平幅值按照一定的分级标度转化为对应的数字值，并取整数

把连续的电平幅值转换为离散的数字量

编码：把量化的结果转换为与之对应的二进制编码

- 数据交换方式（电路交换、报文交换、分组交换）

电路交换

在数据传输期间，源结点与目的结点之间有一条由中间结点构成的专用物理连接线路，在数据传输结束之前，这条线路一直保持

过程：连接建立、数据传输、连接释放

特点：独占资源，用户始终占用端到端的固定传输带宽

适用：远程批处理信息传输或系统间实时性要求高的大量数据传输的情况

报文交换

报文携带有目标地址、源地址等信息，无需在两个站点之间建立一条专用通路

单位：报文 方式：存储转发

分组交换

限制了每次传送的数据块大小的上限，把大的数据块划分为合理的小数据块

单位：分组 方式：存储转发

电路交换

优点

- 传输时延小
- 数据顺序传送，无失序问题
- 实时性强，双方一旦建立物理通路，便可以实时通信，适用于交互式会话类通信
- 全双工通信，没有冲突，通信双方有不同的信道，不会争用物理信道
- 适用于模拟信号和数字信号
- 控制简单，电路的交换设备及控制较简单

缺点

- 建立连接时间长
- 线路独占，即使通信线路空闲，也不能供其他用户使用，信道使用效率低
- 灵活性差，连接通路中任何一点出故障，必须重新拨号建立新连接，不适应突发性通信
- 无数据存储能力，难以平滑通信量
- 电路交换时，数据直达，不同类型、不同规格、不同速率的终端很难相互进行通信
- 无法发现与纠正传输差错，难以在通信过程中进行差错控制

报文交换

优点

- 无需建立连接，无建立连接时延，用户可随时发送报文
- 动态分配线路，动态选择一条报文通过的最佳路径
- 提高线路可靠性，某条传输路径发生故障，可重新选择另一条路径传输
- 提高线路利用率，通信双方在不同的时间一段一段地部分占有这条物理通道，多个报文可共

缺点

- 实时性差，不适合传送实时或交互式业务的数据。数据进入交换结点后要经历存储转发程，从而引起转发时延
- 只适用于数字信号
- 由于报文长度没有限制，而每个中间结点都要完整地接收传来的整个报文，当输出线路不空闲时，还可能要存储几个完整报文等待转发，要求网络中每个结点有较大的缓冲区。

享信道

- 提供多目标服务，一个报文可同时发往多个目的地址
- 在存储转发中容易实现代码转换和速率匹配，甚至收发双方可以不同时处于可用状态。这样就便于类型、规格和速度不同的计算机之间进行通信

为了降低成本，减少结点的缓冲存储器的容量，有时要把等待转发的报文存在磁盘上，进一步增加了传送时延

分组交换

优点

- 无建立时延，无需为通信双方预先建立一条专用通信线路。用户可随时发送分组
- 线路利用率高，通信双方在不同的时间部分占有这条物理通道，多个分组可共享信道
- 简化了存储管理。因为分组的长度固定，相应的缓冲区的大小也固定，在交换结点中存储器的管理通常被简化为对缓冲区的管理，相对比较容易
- 加速传输，后一个分组的存储可以和前一个分组的转发并行操作；传输一个分组比一份报文所需缓冲区小。减少等待发送时间
- 减少出错几率和重发数据量，提高可靠性减少传输时延
- 分组短小，适用计算机之间突发式数据通信

缺点

- 尽管分组交换比报文交换的传输时延少，但仍存在存储转发时延，而且其结点交换机必须具有更强的处理能力
- 每个分组都要加控制信息，一定程度上降低了通信效率，增加了处理的时间
- 当分组交换采用数据报服务时，可能出现失序、丢失或重复分组，分组到达结点时，要对分组按编号进行排序等工作，增加了麻烦。若采用虚电路服务，虽无失序问题，但有呼叫建立、数据传输和虚电路释放三个过程

• 虚电路服务

	数据报服务	虚电路服务
连接的建立	不需要	必须有
目的地址	每个分组都有完整的目的地址	仅在建立连接阶段使用，之后每个分组使用长度较短的虚电路号
路由选择	每个分组独立地进行路由选择和转发	属于同一条虚电路的分组按照同一路由转发
分组顺序	不保证分组的有序到达	保证分组的有序到达
可靠性	不保证可靠通信，可靠性由用户主机保证	可靠性由网络保证
对网络故障的适应性	出故障的结点丢失分组，其他分组路径选择发生变化，可正常传输	所有经过故障结点的虚电路均不能正常工作
差错处理和流量控制	由用户主机进行流量控制，不保证数据报的可靠性	可由分组交换网负责，也可由用户主机负责

*虚电路分类 P42 T29

永久性虚电路 PVC

交换性虚电路 SVC (临时的)

两台主机可以以存在多条虚电路为不同进程服务

2.2 传输介质

- 导向性传输介质

电磁波被导向沿着固体媒介（铜线/光纤）传播

包括：

双绞线：屏蔽双绞线、非屏蔽双绞线

同轴电缆

光纤：单模光纤、多模光纤

- 非导向性传输介质

无线电波：较强穿透能力，可传远距离，广泛用于通信领域（如手机通信）

信号向所有方向传播

微波：微波通信频率较高、频段范围宽，因此数据率很高

信号固定方向传播

包括：

地面微波接力通信

卫星通信

优点：通信容量大、距离远、覆盖广、广播通信和多址通信

缺点：传播时延长(250-270ms)、受气候影响大、误码率较高、成本高

红外线、激光：把信号分别转换为红外光信号和激光信号格式，再在空间中传播

信号固定方向传播

- 物理层接口特性

机械特性：指明接口所用接线器的形状和尺寸、引线数目和排列、固定和锁定装置等。

电气特性：指明在接口电缆的各条线上出现的电压的范围。

功能特性：指明某条线上出现的某一电平的电压表示何种意义。

过程特性：也叫时间特性、规程特性，指明对于不同功能的各种可能事件的出现顺序。

2.3 物理层设备

- 中继器

功能：对信号进行再生和还原，对衰减的信号进行放大，保持与原数据相同，以增加信号传输的距离，延长网络的长度

两端：

1. 两端的网络部分是网段，适用于完全相同的两类网络的互连，且两个网段速率要相同

2. 中继器只将任何电缆段上的数据发送到另一段电缆上，它仅作用于信号的电气部分，并不管数据中是否有错误数据或不适于网段的数据

3. 两端可连相同媒体，也可连不同媒体

4. 中继器两端的网段一定要是同一个协议（中继器不会存储转发）

5-4-3 规则：五段通信介质、四个中继器、三个挂接计算机

- 集线器

功能：对信号进行再生放大转发，对衰减的信号进行放大，接着转发到除输入端口外其他所有处于工作状态的端口上，以增加信号传输的距离，延长网络的长度。不具备信号的定向传送能力，是一个共享式设备。

星形拓扑

集线器不能分割冲突域——连在集线器上的工作主机平分带宽

第3章 数据链路层

结点：主机、路由器

链路：网络中两个结点之间的物理通道，传输介质有双绞线、光纤和微波。分为有线、无线链路

数据链路：网络中两个结点之间的逻辑通道，把实现控制数据传输协议的硬件和软件加到链路上就构成数据链路

帧：链路层的协议数据单元，封装网络层数据报

功能：为网络层提供服务、链路管理、组帧、流量控制、差错控制

3.1 数据链路层的功能

数据链路层在物理层提供服务的基础上向网络层提供服务，其最基本的服务是将源自网络层来的数据可靠地传输到相邻节点的目标机网络层。其主要作用是加强物理层传输原始比特流的功能，将物理层提供的可能出错的物理连接改造成为逻辑上无差错的数据链路，使之对网络层表现为一条无差错的链路

- 为网络层提供服务

无确认无连接服务

有确认无连接服务

有确认面向连接服务

- 链路管理

即连接的建立、维持、释放（用于面向连接的服务）

- 组帧（帧定界、帧同步、透明传输）

封装成帧：在一段数据的前后部分添加首部和尾部，这样就构成了一个帧。接收端在收到物理层上传交的比特流后，根据首部和尾部的标记，从收到的比特流中识别帧的开始和结束

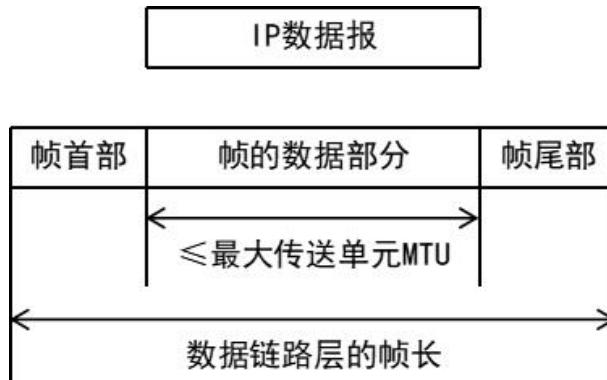
首部和尾部包含许多的控制信息，他们的一个重要作用：帧定界（确定帧的界限）

帧同步：接收方应当能从接收到的二进制比特流中区分出帧的起始和终止。

最大传送单元 MTU：帧的数据部分的长度上限

透明传输：当所传数据中的比特组合恰巧与某一个控制信息完全一样时，采取适当的措施，使收方不会将这样的数据误认为是某种控制信息。保证数据链路层的传输是透明的

组帧的四种方法：字符计数法、字符（节）填充法、零比特填充法、违规编码法



- 流量控制

限制发送方的数据流量，使其发送速率不超过接收方的接受能力

*对于数据链路层：控制的是相邻两结点之间数据链路上的流量

对于传输层：控制源端到目的端之间的流量

- 差错控制

位错：循环冗余校验 CRC

差错控制：自动重传请求 ARQ

帧错：定时器、编号机制

*三个基本问题：封装成帧、透明传输、差错检测

3.2 组帧

- 字符计数法

帧首部使用一个计数字段（第一个字节，八位）来标明帧内字符数。包括自身所占用的一个字节
 - 字符填充的首尾定界符法

使用特定字符来定界一帧的开始与结束
SOH 表示帧的开始 EOT 表示帧的结束 ESC 表示转义字符
 - 零比特填充法

允许数据帧包含任意个数的比特，允许每个字符编码包含任意个数的比特
用 01111110 来标志一帧的开始和结束
发送时遇五个连续“1”自动插入一个“0”
容易使用硬件来实现，性能优于字符填充法
 - 违规编码法

借用违规编码序列来定界帧的起始与终止
例如：曼切斯特编码利用“高-高”“低-低”
局域网 IEEE802 标准采用此方法
- 目前组帧常用方法是零比特填充法（用于 HDLC 协议）和违规编码法

3.3 差错控制

- 差错

来源：
由于线路本身电气特性所产生的随机噪声（热噪声），是信道固有的，随机存在的（全局性）
办法：提高信噪比来减少或避免干扰（对传感器下手）
外界特定的短暂原因所造成的冲击噪声，是产生差错的主要原因（局部性）
办法：通常利用编码技术来解决
- 分类：

位错：比特位出错，1 变成 0，0 变成 1
帧错：丢失，重复，失序
- 差错控制：

自动重传请求 ARQ
检测到差错时，通知发送端重发 → 检错编码
前向纠错 FEC
接收端发现差错并加以纠正 → 纠错编码
- 检错编码（奇偶校验码、循环冗余码 CRC）

奇偶校验码
n-1 位信息元，1 位校验元
奇校验码：码长 n 的码字中“1”的个数为奇数
偶校验码：码长 n 的码字中“1”的个数为偶数
只能检查出奇数个比特错误，检错能力为 50%
- 循环冗余码 CRC（多项式码）

对于 m bit 帧或报文，发送器生成 r bit 序列，即帧检测序列 FCS（冗余码）
使带检验码的帧能被双方预先确定的 r 阶多项式 $G(x)$ 整除
计算方法（对于 m 位的帧）：
 1. 加“0”。 $G(x)$ 的阶数为 r，则再帧的低尾端加 r 个“0”
 2. 模 2 除。利用模 2 除法，用 $G(x)$ 对应的数据串除 m+r 位的字符串，得余数即冗余码

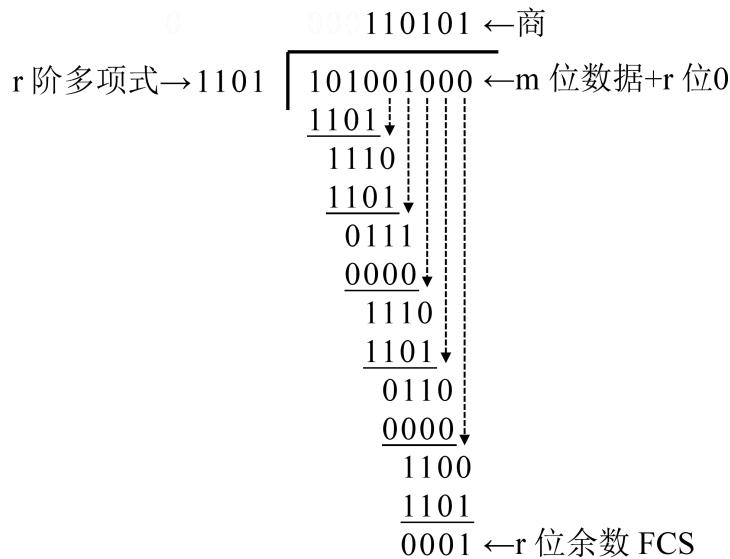
$$m \text{ 位数据 } r \text{ 位 } 0 \div r \text{ 阶多项式 } G(x) = \text{ 商} \dots \dots r \text{ 位 FCS}$$

检错：

1.接收到的 $m+r$ 位数据除以 $G(x)$ 对应的数据串

2.余数为 0 则认为正确接收，余数不为 0 则丢弃

注：循环冗余码有纠错功能，但数据链路层仅用其检错功能



• 纠错编码（海明码）

海明码（汉明码）：发现错误、找到位置、纠正错误

海明距离（码距）：两个合法编码（码字）的对应比特取值不同的比特数

一个有效编码集中，任意两个合法编码（码字）的海明距离的最小值

过程：（以数据码 1010 为例）

1. 确定校验码位数

数据/信息有 n 位，校验码有 k 位（若要检测两位错则再加 1 位校验位，即 $k+1$ 位）

校验码一共有 2^k 种取值

海明不等式： $2^k \geq n + k + 1$

设信息位为 $D_4D_3D_2D_1(1010)$ ，共 4 位 校验位为 $P_3P_2P_1$ ，共 3 位

对应的海明码为 $H_7H_6H_5H_4H_3H_2H_1$ ，共 7 位

2. 确定校验位的分布

规定校验位 P_i 在海明位号为 2^{i-1} 的位置上，其余为信息位

$P_1 \rightarrow 2^{1-1}=2^0=1$ ，位于 H_1

$P_2 \rightarrow 2^{2-1}=2^1=2$ ，位于 H_2

$P_3 \rightarrow 2^{3-1}=2^2=4$ ，位于 H_4

海明码对应关系： $H_7H_6H_5H_4H_3H_2H_1 \rightarrow D_4D_3D_2P_3D_1P_2P_1$

3. 分组以形成校验关系

被校验数据位的海明位号等于校验该数据位的各校验位海明位号之和

$D_1 \rightarrow H_3$, 由 P_2P_1 校验:	$3 =$	P_1	$+ P_2$	$+ P_3$
$D_2 \rightarrow H_5$, 由 P_3P_1 校验:	$5 =$	1	2	4
$D_3 \rightarrow H_6$, 由 P_3P_2 校验:	$6 =$		2	4
$D_4 \rightarrow H_7$, 由 $P_3P_2P_1$ 校验:	$7 =$	1	2	4

4. 校验位取值

校验位 P_i 的值为第 i 组（由该校验位校验的数据位）所有位求异或

$$\begin{aligned}P_1 &= D_1 \oplus D_2 \oplus D_4 = 0 \oplus 1 \oplus 1 = 0 \\P_2 &= D_1 \oplus D_3 \oplus D_4 = 0 \oplus 0 \oplus 1 = 1 \\P_3 &= D_2 \oplus D_3 \oplus D_4 = 1 \oplus 0 \oplus 1 = 0\end{aligned}$$

1010 对应的海明码为 1010010

5. 海明码的校验原理

每个校验组分别利用校验位和参与形成校验位的信息位进行奇偶校验检查

$$\begin{aligned}S_1 &= P_1 \oplus D_1 \oplus D_2 \oplus D_4 \\S_2 &= P_2 \oplus D_1 \oplus D_3 \oplus D_4 \\S_3 &= P_3 \oplus D_2 \oplus D_3 \oplus D_4\end{aligned}$$

*海明距离与检错纠错 P71 T5

检测 d 位错误——海明距离 $d+1$

纠正 d 位错误——海明距离 $2d+1$

3.4 流量控制与可靠传输机制

• 流量控制

基本思路：有接收端控制发送方发送数据的速率

解决方法：停止-等待协议、滑动窗口协议

数据链路层的流量控制是点对点的，而传输层的流量控制是端到端的

数据链路层流量控制手段：接收方收不下就不回复确认

传输层流量控制手段：接收端给发送端一个窗口公告

• 可靠传输

解决方法：确认、超时重传

捎带确认：将确认捎带在一个回复帧中

超时重传：发送方发送某个数据帧后开启计时器，一定时间内未收到确认则重新发送该帧

自动重传请求 ARQ：通过接收方请求发送方重传出错的数据帧来恢复出错的帧

• 停止-等待协议

发送窗口大小=1，接收窗口大小=1

发送方每发送一帧，都要等待接收方的确认信号，再发下一帧

接收方每接收一帧，都要反馈一个确认信号 ACK

发完一个帧后，必须保留它的副本

数据帧和确认帧必须编号

• 后退 N 帧协议 GBN

发送窗口大小>1，接收窗口大小=1

那么发送窗口的尺寸 W_T 应满足： $1 \leq W_T \leq 2^n - 1$

接收窗口为 1，保证了按序接收数据帧

对 n 号帧的确认采用累积确认的方式，标明接收方已经收到 n 号帧和它之前的全部帧

• 选择重传协议 SR

发送窗口大小>1，接收窗口大小>1

那么发送窗口的尺寸 W_T 应满足： $W_{T_{max}} = W_{T_{min}} \leq 2^{n-1}$

发送窗口大小等于接收窗口大小

• 信道利用率和信道吞吐率

信道利用率：发送方在一个发送周期内，有效地发送数据所需要的时间占整个发送周期的比率

$$\text{信道利用率} = (L/C)/T$$

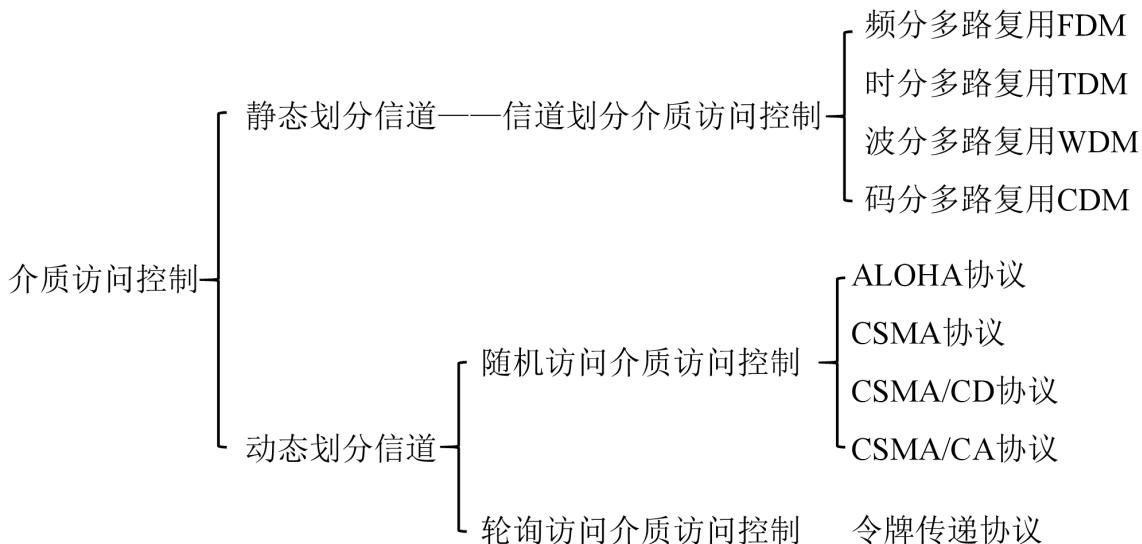
T: 发送周期, 从开始发送数据, 到收到第一个确认帧为止

L: T 内发送 L 比特数据

C: 发送方数据传输率

信道吞吐率=信道利用率×发送方的发送速率

3.5 介质访问控制



- 介质访问控制 MAC, Medium Access Control

为使用介质的每个结点隔离来自同一信道商其他结点所传送的信号, 以协调活动节点的传输

介质访问控制子层: 属于数据链路层的子层, 包括用来决定广播信道中信道分配的协议

- 信道划分介质访问控制 (多路复用技术)

在一条介质上同时携带多个传输信号

利用分时分频分码等方法把一条广播信道, 逻辑上分成几条用于两点通信互不干扰的子信道

1. 频分多路复用 FDM

将多路基带信号调制到不同频率载波上, 再叠加形成一个复合信号

实际应用中为防止子信道间干扰, 相邻信道需加入“保护频带”

2. 时分多路复用 TDM

将一条物理信道按时间分呈若干时间片, 轮流分配给多个信号使用

统计时分多路复用 STDM

采用 STDM 帧, 不固定分配时隙, 按需动态的分配时隙

3. 波分多路复用 WDM

在光纤中传输多种不同波长的光信号

4. 码分多路复用 CDM

1 个比特分为多个码片/chip, 每一个站点被指定一个唯一的 m 位的芯片序列

以 $S = 00011011 = (-1 -1 -1 +1 +1 -1 +1 +1)$ 、 $T = 00101110 = (-1 -1 +1 -1 +1 +1 +1 -1)$ 为例

1. 多个站点同时发送数据的时候, 要求各个站点芯片序列相互正交, 规格化内积为 0

$$S \cdot T \equiv \frac{1}{m} \sum_{i=1}^m S_i T_i = 0$$

2. 发送 1 时发送芯片序列 (通常把 0 写成 -1), 发送 0 时发送芯片序列反码

S 发送数据 1 则发送向量 $(-1 -1 -1 +1 +1 -1 +1 +1)$

T 发送数据 0 则发送向量 $(+1 +1 -1 +1 -1 -1 +1)$

3.两个向量到了公共信道上，线性相加

$$S + T = (0 \ 0 \ -2 \ 2 \ 0 \ -2 \ 0 \ 2)$$

4.数据分离：合并的数据和源站规格化内积

$$S \cdot (S + T) = +1$$

$$T \cdot (S + T) = -1$$

注：如果计算结果为 0 说明未发送数据

• 随机访问介质访问控制（ALOHA 协议、CSMA 协议、CSMA/CD 协议、CSMA/CA 协议）

1) ALOHA 协议

纯 ALOHA 协议

当网络中的任何一个结点需要发送数据时，可以不进行任何检测就发送数据

如果在一段时间内没有收到确认，该结点就认为传输过程中发生了冲突

发生冲突的结点需要等待一段随机时间后再发送数据，直至发送成功

发送成功率不高，最大值只有 18.4%

时隙 ALOHA 协议

把所有各站在时间上同步起来，并将时间划分为一段段等长的时隙(Slot)

规定只能在每个时隙开始时才能发送一个帧

避免了用户发送数据的随意性，减少了数据产生冲突的可能性，提高了信道的利用率

2) CSMA 协议

载波监听多路访问（CSMA，Carrier Sense Multiple Access）协议

每个结点发送数据之前都使用载波监听技术来判定通信信道是否空闲

常用的 CSMA 有以下 3 种策略

	1-坚持CSMA	非坚持CSMA	p坚持CSMA
信道空闲时	立即发送数据	立即发送数据	以 p 概率发送数据 以 (1-p) 概率不发送数据
信道忙时	持续监听信道	等待随机时间再监听	等待随机时间再监听

3) CSMA/CD 协议

载波监听多路访问/碰撞检测（Carrier Sense Multiple Access/Collision Detection）协议

适用于总线型网络或半双工网络

先听后发：每个站在发送数据之前要先检测一下总线上是否有其他计算机在发送数据

若有，则暂时不发送数据，以免发生冲突；若没有，则发送数据

边听边发：在发送的过程需要监听是否有碰撞，因为存在信号的传播延迟

冲突停发：只有经过争用期这段时间还没有检测到碰撞，才能肯定不会发生碰撞

以太网的端到端往返时延 2τ 称为争用期，或碰撞窗口

随机重发：发生碰撞的站在停止发送数据后，要推迟（退避）一个随机时间才能再发送数据

基本退避时间取为争用期 2τ 。从整数集合 $[0, 1, \dots, 2^{k-1}]$ 中随机地取出一个数，记为 r。重传所需的时延就是 r 倍的基本退避时间

$$k = \min(\text{重传次数}, 10)$$

当重传达 16 次仍不能成功时即丢弃该帧，并向高层报告

4) CSMA/CA 协议

载波监听多路访问/冲突避免（Carrier Sense Multiple Access with Collision Avoidance）协议

用于无线网络，802.11 协议

采用能量检测 ED、载波检测 CS 和能量载波混合检测三种方式检测信道空闲

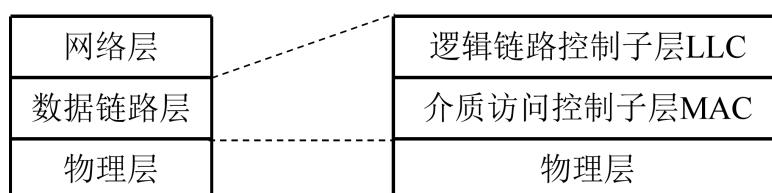
发送数据前，先检测信道是否空闲

空闲则发出 RTS (request to send)，信道忙则等待

- RTS 包括发射端的地址、接收端的地址、下一份数据将持续发送的时间等信息
- 接收端收到 RTS 后，将响应 CTS (clear to send)
- 发送端收到 CTS 后，开始发送数据帧
 - 同时预约信道：发送方告知其他站点自己要传多久数据
 - 接收端收到数据帧后，将用 CRC 来检验数据是否正确，正确则响应 ACK 帧
 - 发送方收到 ACK 就可以进行下一个数据帧的发送，若没有则一直重传至规定重发次数为止
 - 采用二进制指数退避算法来确定随机的推迟时间
- 轮询访问介质访问控制（令牌传递协议）
 - 令牌传递协议
 - 令牌 (Token)：一个特殊格式的 MAC 控制帧，不含任何信息
 - 控制信道的使用，确保同一时刻只有一个结点独占信道
 - 令牌环网无碰撞
 - 每个结点都可以在令牌持有时间获得发送数据的权利，并不是无限制地持有令牌
 - 问题：1.令牌开销 2.等待延迟 3.单点故障
 - 应用于令牌环网（物理星型拓扑，逻辑环形拓扑）
 - 采用令牌传送方式的网络常用于负载较重、通信量较大的网络中

3.6 局域网 LAN, Local Area Network

- 局域网的基本概念和体系结构
 - DIX Ethernet V2 是世界上第一个局域网产品（以太网）的规定，802.3 局域网可简称为“以太网”
 - 主要特点：
 - 覆盖的地理范围较小，只在一个相对独立的局部范围内联
 - 各站为平等关系，共享传输信道
 - 通信延迟时间短，误码率低，可靠性较高
 - 多采用分布式控制和广播式通信，能进行广播和组播
 - 主要技术要素：网络拓扑结构、传输介质与介质访问控制方法。
 - 主要拓扑结构：星形网、环形网、总线型网和树形网（星形网和总线型网的结合）
 - 主要传输介质：双绞线、铜缆和光纤等，其中双绞线为主流传输介质
 - 主要介质访问控制方法：括 CSMA/CD、令牌总线（总线型网）和令牌环（环形网）
 - 三种特殊的局域网拓扑：
 - 以太网，逻辑拓扑是总线形结构，物理拓扑是星形或拓展星形结构
 - 令牌环 (Token Ring, IEEE 802.5)，逻辑拓扑是环形结构，物理拓扑是星形结构
 - FDDI (光纤分布数字接口, IEEE 802.8)，逻辑拓扑是环形结构，物理拓扑是双环结构
 - ATM 网 (Asynchronous Transfer Mode) 单元交换技术，使用 53 字节的单元进行交换
 - 无线局域网 WLAN (Wireless Local Area Networks) 采用 IEEE 802.11 标准
 - IEEE 的 802 标准定义的局域网参考模型只对应于 OSI 参考模型的数据链路层和物理层
 - 并且将数据链路层拆分为两个子层：逻辑链路控制 LLC 子层和媒体接入控制 MAC 子层
 - MAC 子层：与接入到传输媒体有关的内容
 - 主要功能包括：组帧和拆卸帧、比特传输、差错检测、透明传输
 - LLC 子层与传输媒体无关
 - 提供无确认无连接、面向连接、带确认无连接、高速传送 4 种连接服务类型



OSI 模型与 IEEE 802 协议层的比较

- 以太网的基本概念、传输介质与高速以太网

以太网采用总线拓扑结构，信息以广播方式发送，使用了 CSMA/CD 技术对总线进行访问控制

以太网规定 51.2s 为争用期的长度。最短有效帧长为 64B

采用无连接的工作方式，提供的是不可靠服务，对于差错的纠正则由高层完成

发送的数据都使用曼彻斯特编码的信号

以太网的传输介质：

参数	10BASE5	10BASE2	10BASE-T	10BASE-FL
传输媒体	基带同轴电缆（粗缆）	基带同轴电缆（细缆）	非屏蔽双绞线	光纤对（850nm）
编码	曼彻斯特编码	曼彻斯特编码	曼彻斯特编码	曼彻斯特编码
拓扑结构	总线形	总线形	星形	点对点
最大段长	500m	185m	100m	2000m
最多结点数目	100	30	2	2

高速以太网分类：

速率达到或超过 100Mb/s 的以太网称为高速以太网

- 1) 100BASE-T 以太网

在双绞线上传送 100Mb/s 基带信号的星形拓扑结构以太网

支持全双工方式，又支持半双工方式（CSMA/CD 协议）

MAC 帧格式仍然是 802.3 标准规定的，保持最短帧长不变

一个网段的最大电缆长度减小到 100m

帧间时间间隔从原来的 9.6us 改为现在的 0.96us

- 2) 吉比特以太网

又称千兆以太网，速率 1Gb/s

用全双工和半双工（CSMA/CD 协议）两种方式工作

使用 802.3 协议规定的帧格式

与 10BASE-T 和 100BASE-T 技术向后兼容

- 3) 10 吉比特以太网

与 10Mb/s、100Mb/s 和 1Gb/s 以太网的帧格式完全相同

保留了 802.3 标准规定的以太网最小和最大帧长，便于升级

只使用光纤作为传输媒体

只工作在全双工方式

- 网卡与 MAC 地址

网络接口板又称为通信适配器（adapter）、网络接口卡（NIC，Network Interface Card）或“网卡”

适配器的功能：进行串行（对局域网）/并行（对计算机）转换，帧的发送与接收、帧的封装与拆

封、介质访问控制、数据的编码与解码及数据缓存功能等

每块网卡在出厂时都有一个唯一的代码，称为介质访问控制 MAC 地址

这个地址用于控制主机在网络上的数据通信

数据链路层设备（网桥、交换机等）都使用各个网卡的 MAC 地址

网卡控制着主机对介质的访问，也工作在物理层，关注比特，而不关注地址信息和高层协议信息

适配器从网络上每收到一个 MAC 帧就首先用硬件检查 MAC 帧中的 MAC 地址

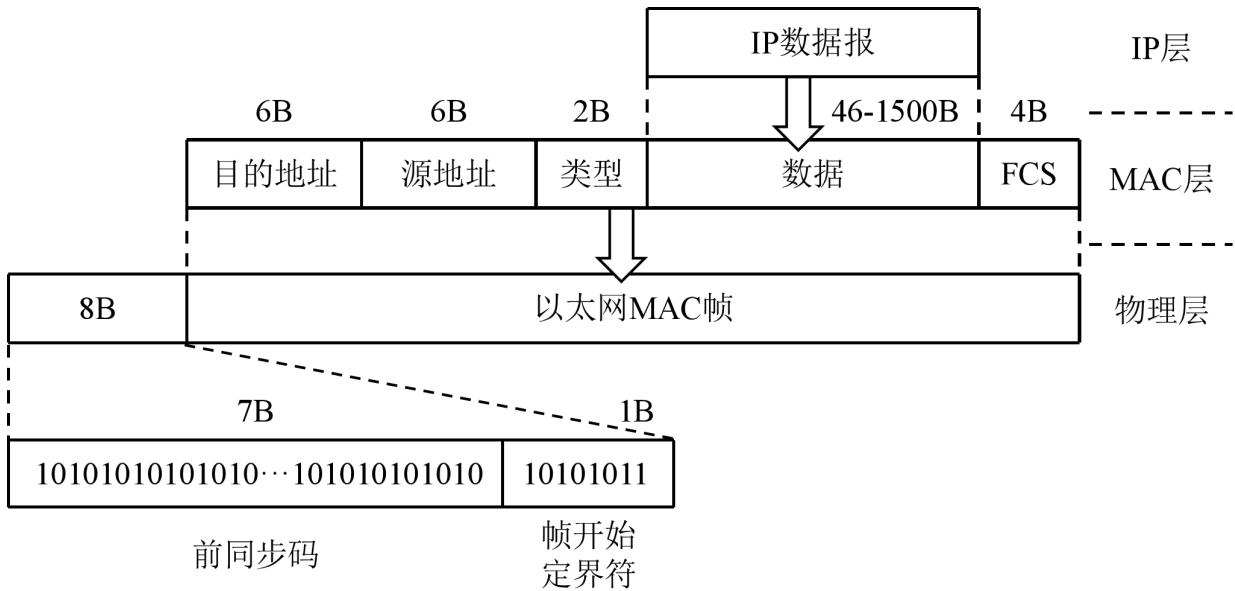
如果是发往本站的帧则收下，然后再进行其他的处理。否则就将此帧丢弃，不再进行其他的处理

每块网卡中的 MAC 地址也称物理地址，记录在 ROM 上

MAC 地址长 6 字节，一般用由连字符或冒号分隔的 12 个十六进制数表示，如 02-60-8c-e4-b1-21
高 24 位为厂商代码，低 24 位为厂商自行分配的网卡序列号

- 以太网的 MAC 帧

MAC 帧的格式有两种：DIX Ethernet V2 标准和 IEEE802.3 标准



以太网 V2 标准的 MAC 帧格式

前导码：8B 数据，使接收端与发送端时钟同步

第一个字段共 7 字节，是前同步码，用来快速实现 MAC 帧的比特同步

第二个字段共 1 字节，帧开始定界符，表示后面的信息就是 MAC 帧

目的地址、源地址：均使用 48b (6B) 的 MAC 地址

类型：占 2B，指出数据域中携带的数据应交给哪个协议实体处理

数据：占 46~1500B，包含高层的协议信息，数据较少时必须加以填充

校验码 FCS：4B，采用循环冗余码，校验目的地址、源地址、类型、数据字段，不校验前导码

802.3 帧格式用长度域替代了 DIX 以太帧中的类型域，指出数据域的长度

长度/类型两种机制可以并存，从 1501 到 65535 的值可用于类型段标识符

- 无线局域网 IEEE 802.11

协议标准：IEEE 802.11，包括 IEEE 802.11a 和 IEEE 802.11b 等

无线局域网的分类：

有固定基础设施的无线局域网的组成

IEEE 802.11 标准规定其最小构件为基本服务集 BSS

一个基本服务集包括一个基站和若干个移动站，所有站在本 BSS 内可直接通信

但在和本 BSS 以外的站通信时都必须通过本 BSS 的基站

因此，BSS 中的基站称为接入点 AP，Access Point

一个基本服务集可以是孤立的，也可通过接入点连接到一个主干分配系统 DS (Distribution System)，然后再接入到另一个基本服务集，构成扩展的服务集 ESS (Extended Service Set)。ESS 还可通过门桥 (Portal) 设备为无线用户提供到非 IEEE 802.11 无线局域网（如到有线连接的因特网）的接入。门桥的作用就相当于一个网桥。

无固定基础设施的无线局域网的组成

无固定基础设施的无线局域网又称为自组网络 (ad hoc network)

自主网络没有上述基本服务集中的接入点

由一些处于平等状态的移动站之间相互通信组成的临时网络

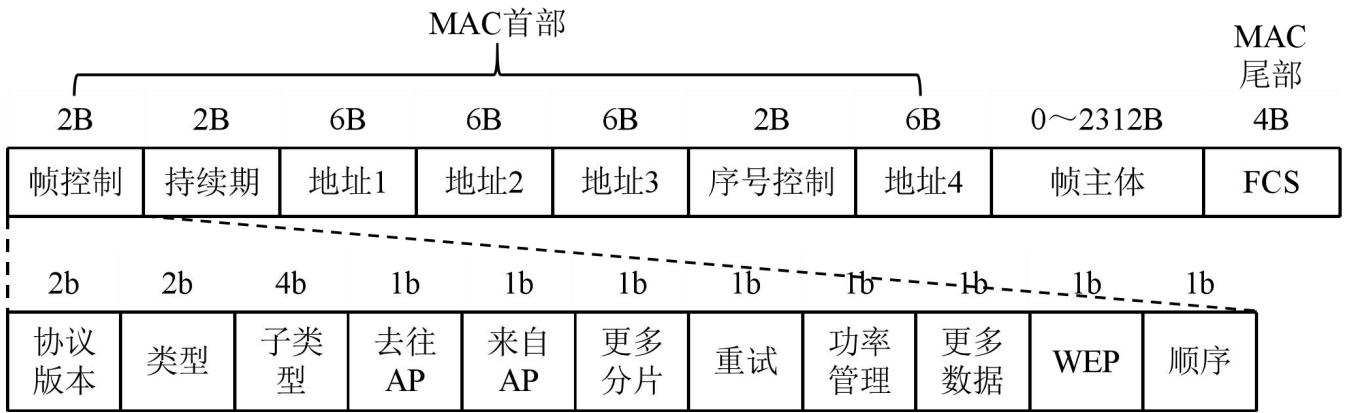
这些移动站都具有路由器的功能

802.11 帧共有三种类型：数据帧、控制帧和管理帧

MAC 首部：共 30B，帧的复杂性都在 MAC 首部

帧主体：即帧的数据部分，不超过 2312B，它比以太网的最大长度长很多

帧检验序列：FCS 是尾部，共 4B

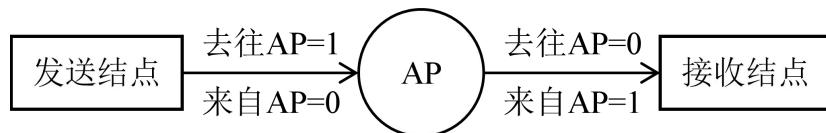


802.11 局域网的数据帧

802.11 帧的 MAC 首部中最重要的是 4 个地址字段（都是 MAC 地址）

前三个地址的内容取决于帧控制字段中的“去往 AP”和“来自 AP”这两个字段的数值
地址 4 用于自组网络

地址 1 是直接接收数据帧的结点地址，地址 2 是实际发送数据帧的结点地址



去往 AP	来自 AP	地址1	地址2	地址3	地址4
1	0	接收地址=AP地址	发送地址=源地址	目的地址	——
0	1	接收地址=目的地址	发送地址=AP地址	源地址	——

802.11 帧的地址字段最常用的两种情况

- 虚拟局域网 VLAN, Virtual LAN

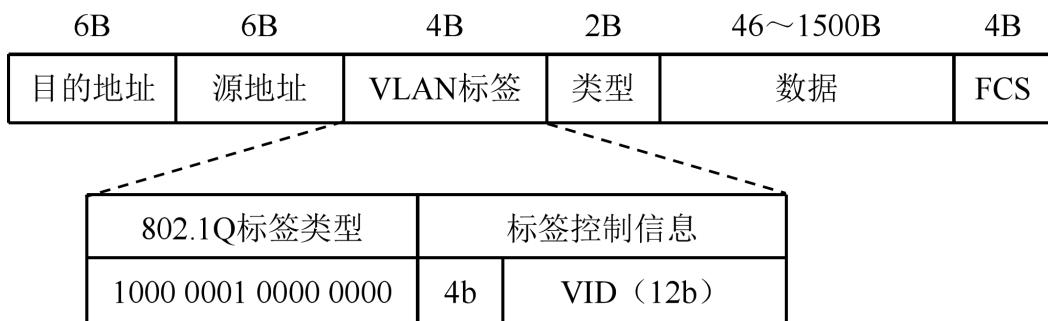
虚拟局域网把一个较大的局域网分割成一些较小的与地理位置无关的逻辑上的 VLAN

802.3ac 标准定义了支持 VLAN 的以太网帧格式的扩展

在以太网帧中的源地址字段和类型字段之间插入一个 4 字节的标识符

该标识符指明发送该帧的计算机所属的虚拟局域网，称为 VLAN 标签

对应的帧称为 802.1Q 帧



VLAN 标签的前两个字节置为 0x8100，表示这是一个 802.1Q 帧

VLAN 标签的后两个字节中，前 4 位没有用，后 12 位是该 VLAN 的标识符 VID

标识符 VID 唯一标识了该 802.1Q 帧属于哪个 VLAN，12 位的 VID 可识别 4096 个不同的 VLAN
插入 VID 后，802.1Q 帧的 FCS 必须重新计算

各主机不知道自己的 VID 值（但交换机必须知道），主机与交换机之间交互的都是标准以太网帧
一个 VLAN 的范围可以跨越不同的交换机，前提是所用的交换机能够识别和处理 VLAN
同一个 VLAN 下的主机属于同一个广播域

*放大器与中继器 P111 T4

放大器是用来加强宽带信号（用于传输模拟信号）的设备

中继器是用来加强基带信号（用于传输数字信号）的设备（大多数以太网采用基带传输）

*重复硬件地址 P112 T9

在使用静态地址的系统中，如果有重复的硬件地址，那么这两个设备都不能正常通信

*同轴电缆分类 P112 T10

同轴电缆分 50Ω 基带电缆和 75Ω 宽带电缆两类

基带电缆又分细同轴电缆和粗同轴电缆

3.7 广域网

• 广域网基本概念

广域网通常是指覆盖范围很广（远远超出一个城市的范围）的长距离网络

广域网由结点交换机以及连接这些交换机的链路组成

结点交换机完成分组存储转发的功能

结点交换机在单个网络中转发分组

路由器是在多个网络构成的互联网中转发分组

广域网与局域网的对比：

	广域网	局域网
覆盖范围	很广，通常跨区域	较小，通常在一个区域内
连接方式	点对点通信方式	广播通信方式
OSI模型	物理层、数据链路层、网络层	物理层、数据链路层
联系	1) 广域网和局域网都是互联网的重要组成构件，从互联网的角度上看，二者平等，没有包含关系 2) 连接在广域网或局域网上的主机在该网内进行通信时，只需要使用其网络的物理地址即可	
重点	强调资源共享	强调数据传输

• PPP (Point-to-Point Protocol) 协议

应用在直接连接两个结点的链路上，使用串行线路通信的面向字节的协议

通过拨号或专线方式建立点对点连接发送数据

各种主机、网桥和路由器之间简单连接的一种共同的解决方案

PPP 协议有三个组成部分：

一个将 IP 数据报封装到串行链路的方法

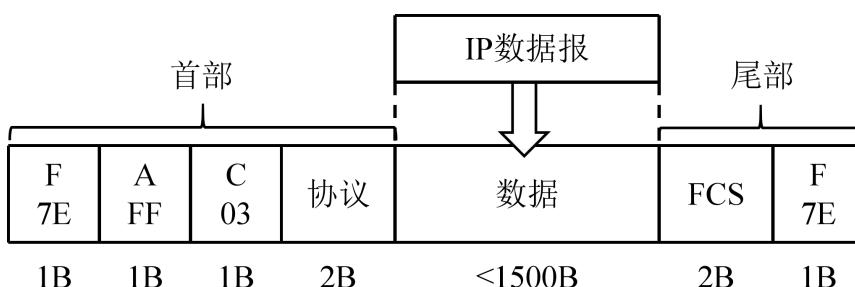
一个链路控制协议 LCP, Link Control Protocol

用于建立、配置和测试数据链路连接，并在不需要时将它们释放

一套网络控制协议 NCP, Network Control Protocol

其中每个协议支持不同的网络层协议，用来建立和配置不同的网络层协议

PPP 帧的格式：



标志字段 F：首部和尾部各占 1 个字节，规定为 0x7E（0111110），前后都加标志字段
地址字段 A：占 1 个字节，规定为 0xFF
控制字段 C：占 1 个字节，规定为 0x03
协议字段：占 2 个字节。

字段为 0x0021 时，PPP 帧的信息字段就是 IP 数据报

字段为 0xC021 时，则信息字段是 PPP 链路控制数据

字段为 0x8021 时，则表示这是网络控制数据

信息部分：占 0~1500B。转义字符 0x7D（01111101）

帧检验序列 FCS：占 2 个字节，循环冗余码。检验地址、控制、协议和信息字段

PPP 协议的特点：

PPP 协议是一个面向字节的协议

PPP 协议是点对点协议，支持一对一通信。

PPP 协议提供差错检测但不提供纠错功能，只保证无差错接收（通过硬件进行 CRC 校验）

PPP 协议是不可靠的传输协议，因此也不使用序号和确认机制

PPP 协议只支持全双工链路

PPP 的两端可以运行不同的网络层协议，但仍然可使用同一个 PPP 进行通信

当 PPP 协议用在同步传输链路时，协议规定采用硬件来完成比特填充（和 HDLC 做法一样）

当 PPP 协议用在异步传输（默认）时，就使用一种特殊的字符填充法

*PPP 协议认证 P120T6

PPP 支持两种认证：一种是 PAP，一种是 CHAP。相对来说，CHAP 的安全性更高。

PAP 在传输密码时是明文

PAP 认证通过 2 次握手实现

PAP 认证由被叫方提出连接请求，主叫方响应

CHAP 在传输过程中不传输密码，取代密码的是哈希 hash 值

CHAP 认证则通过 3 次握手实现

CHAP 认证则由主叫方发出请求，被叫方回复一个数据报，其中有主叫方发送的随机哈希值
主叫方在确认无误后发送一个连接成功的数据报连接

PPP 可用于拨号连接，因此支持动态分配 IP 地址

3.8 数据链路层设备

• 交换机

局域网交换机，又称以太网交换机，以太网交换机实质上就是一个多端口的网桥

以太网交换机的特点：

以太网交换机的每个端口都直接与单台主机相连，并且一般都工作在全双工方式

以太网交换机能同时连通多对端口，无碰撞地传输数据

以太网交换机是一种即插即用设备，其转发表通过自学习算法自动地逐渐建立

以太网交换机由于使用专用的交换结构芯片，交换速率较高

以太网交换机独占传输媒体的带宽

以太网交换机对工作站是透明的

以太网交换机工作在数据链路层，通常都工作在全双工方式

以太网交换机可以实现虚拟局域网 VLAN，VLAN 可以隔离冲突域和广播域

以太网交换机的工作原理：

检测从以太端口来的数据帧的源和目的地的 MAC（介质访问层）地址

然后与系统内部的动态查找表进行比较

若数据帧的源 MAC 地址不在查找表中，则将该地址加入查找表

并将数据帧发送给相应的目的端口

若数据帧的目的 MAC 地址不在查找表中，则广播这个帧

以太网交换机的交换模式：

直通式交换机

只检查帧的目的地址，这使得帧在接收后几乎能马上被传出去
无法支持具有不同速率的端口的交换。

存储转发式交换机

先将接收到的帧缓存到高速缓存器中，并检查数据是否正确
确认无误后通过查找表转换成输出端口将该帧发送出去
如果发现帧有错，那么就将其丢弃
优点是可靠性高，并能支持不同速率端口间的转换，缺点是延迟较大

第4章 网络层

4.1 网络层的功能

• 异构网络互连

所谓虚拟互联网络也就是逻辑互联网络

互相联接起来的各种物理网络的异构性本来是客观存在的

我们利用 IP 协议就可以使这些性能各异的网络从用户层面看起来好像是一个统一的网络

使用 IP 协议的虚拟互联网络可简称为 IP 网

互联网上的主机进行通信时，像在一个网络上通信一样，看不见互联的各具体的网络异构细节

• 路由与转发

路由选择：

指按照复杂的分布式算法，根据从各相邻路由器所得到的关于整个网络拓扑的变化情况

动态地改变所选择的路由，根据特定的路由选择协议构造出路由表

同时经常或定期地和相邻路由器交换路由信息而不断地更新和维护路由表

分组转发：

路由器根据转发表将用户的 IP 数据报从合适的端口转发出去

处理通过路由器的数据流，关键操作是转发表查询、转发及相关的队列管理和任务调度等
路由表是根据路由选择算法得出的，而转发表是从路由表得出的。

转发表的结构应当使查找过程最优化，路由表则需要对网络拓扑变化的计算最优化

• 软件定义网络 SDN 的基本概念

传统互联网中的路由器既有转发表又有路由选择软件

网络层可以抽象为数据层面（也称转发层面）和控制层面

数据层面实现转发功能，控制层面实现路由选择功能

软件定义网络 SDN 采用集中式的控制层面和分布式的数据层面，两个层面相互分离

1) 控制层面利用控制-数据接口对数据层面上的路由器进行集中式控制，方便软件来控制网络

在网络的控制层面有一个逻辑上的远程控制器（可以由多个服务器组成）

用于掌握各主机和整个网络的状态，为每个分组计算出最佳路由

通过 Openflow 协议（也可以通过其他途径）将转发表（在 SDN 中称为流表）下发给路由器

2) 数据层面的路由器不再需要相互交换路由信息，只要实现查找转发表、接收-转发分组功能

SDN 的接口：

SDN 提供的编程接口称为北向接口，提供了一系列丰富的 API

上层应用的开发者可以在此基础上设计自己的应用，无需了解底层的硬件细节

SDN 通过为开发者们提供强大的编程接口，使得网络具有很好的编程性

SDN 控制器和转发设备建立双向会话的接口称为南向接口

通过不同的南向接口协议（如 Openflow），SDN 控制器就可兼容不同的硬件设备

同时可以在设备中实现上层应用的逻辑

SDN 控制器集群内部控制器之间的通信接口称为东西向接口

用于增强整个控制层面的可靠性和可拓展性

SDN 的优点：

1. 全局集中式控制和分布式高速转发，利于控制层面的全局优化和高性能的网络转发

2. 灵活可编程与性能的平衡，控制和转发功能分离后使得网络可以由专有的自动化工具以编程方式配置

3. 降低成本，控制和数据层面分离，在使用开放的接口协议后，就实现了网络设备的制造与功能软件的开发相分离，从而有效降低了成本

SDN 的缺点：

1. 安全风险，集中管理容易受到攻击，如果崩溃，整个网络会受到影响

2. 瓶颈问题，分布式的控制层面集中化后，随网络规模扩大，控制器可能成为网络性能瓶颈

- 拥塞控制

在通信子网中，因出现过量的分组而引起网络性能下降的现象称为拥塞

判断网络是否进入拥塞状态的方法是，观察网络的吞吐量与网络负载的关系

网络的负载继续增大，而网络的吞吐量下降到零，那么网络就可能已进入死锁状态

拥塞控制的作用是确保子网能够承载所达到的流量，这是一个全局性的过程

涉及各方面的行为：主机、路由器及路由器内部的转发处理过程等

4.2 路由算法

- 静态路由与动态路由

静态路由算法（非自适应路由算法）

由网络管理员手工配置的路由信息当网络的拓扑结构或链路的状态发生变化时，网络管理员需要手工去修改路由表中相关的静态路由信息。它不能及时适应网络状态的变化，对于简单的小型网络，可以采用静态路由

动态路由算法（自适应路由算法）

指路由器上的路由表项是通过相互连接的路由器之间彼此交换信息，然后按照一定的算法优化出来的，而这些路由信息会在一定时间间隙里不断更新，以适应不断变化的网络，随时获得最优的寻路效果。常用的动态路由算法有距离-向量路由算法和链路状态路由算法

- 距离-向量路由算法

所有结点都定期地将它们的整个路由选择表传送给所有与之直接相邻的结点

路由选择表包含：每条路径的目的地（另一结点），路径的代价（距离）

所有结点都必须参与距离向量交换，以保证路由的有效性和一致性

所有的结点都监听从其他结点传来的路由选择更新信息

在下列情况下更新它们的路由选择表：

- 1) 被通告一条本结点的路由表中不存在的新的路由，此时本地系统加入这条新的路由
- 2) 发来的路由信息中有一条到达某个目的地的路由，该路由与当前使用的路由相比，有较短的距离（较小的代价）。此种情况下，就用经过发送路由信息的结点的新路由替换路由表中到达那个目的地的现有路由

实质：迭代计算一条路由中的站段数或延迟时间，从而得到到达一个目标的最短（最小代价）通路。

它要求每个结点在每次更新时都将它的全部路由表发送给所有相邻的结点

最常见的距离-向量路由算法是 RIP 算法，它采用“跳数”作为距离的度量

- 链路状态路由算法

要求每个参与该算法的结点都具有完全的网络拓扑信息

执行两项任务：

- 1) 主动测试所有邻接结点的状态。
两个共享一条链接的结点是相邻结点，它们连接到同一条链路或同一广播型物理网络
- 2) 定期地将链路状态传播给所有其他结点（或称路由结点）
典型的链路状态算法是 OSPF 算法

三个特征：

- 1) 向本自治系统中所有路由器发送信息，这里使用的方法是洪泛法，即路由器通过所有端口向所有相邻的路由器发送信息。而每个相邻路由器又将此信息发往其所有相邻路由器（但不再发送给刚刚发来信息的那个路由器）
- 2) 发送的信息是与路由器相邻的所有路由器的链路状态，这只是路由器所知道的部分信息。
所谓“链路状态”，是指说明本路由器与哪些路由器相邻及该链路的“度量”。对于 OSPF 算法，链路状态的“度量”主要用来表示费用、距离、时延、带宽等
- 3) 只有当链路状态发生变化时，路由器才向所有路由器发送此信息

每当链路状态报文到达时，路由结点便使用这些状态信息去更新自己的网络拓扑和状态“视野图”，一旦链路状态发生变化，结点就对更新的网络图利用 Dijkstra 最短路径算法重新计算路由，从单一的源出发计算到达所有目的结点的最短路径

主要优点：

- 1) 每个路由结点都使用同样的原始状态数据独立地计算路径，而不依赖中间结点的计算；链路状态报文不加改变地传播，因此采用该算法易于查找故障
- 2) 当一个结点从所有其他结点接收到报文时，它可以在本地立即计算正确的通路，保证一步汇聚
- 3) 由于链路状态报文仅运载来自单个结点关于直接链路的信息，其大小与网络中的路由结点数目无关，因此链路状态算法比距离-向量算法有更好的规模可伸展性
- 4) 链路状态路由算法可以用于大型的或路由信息变化聚敛的互联网环境

比较：

在距离-向量路由算法中

每个结点仅与它的直接邻居交谈

它为它的邻居提供从自己到网络中所有其他结点的最低费用估计

在链路状态路由算法中

每个结点通过广播的方式与所有其他结点交谈

但它仅告诉它们与它直接相连的链路的费用

相较之下，距离-向量路由算法有可能遇到路由环路等问题

• 层次路由

因特网将整个互联网划分为许多较小的自治系统（一个自治系统中包含很多局域网），每个自治系统有权自主地决定本系统内应采用何种路由选择协议。如果两个自治系统需要通信，那么就需要一种在两个自治系统之间的协议来屏蔽这些差异

因特网把路由选择协议划分为两大类：

- 1) 一个自治系统内部所使用的路由选择协议称为内部网关协议 IGP，也称域内路由选择
具体的协议有：RIP 和 OSPF 等
- 2) 自治系统之间所使用的路由选择协议称为外部网关协议 EGP，也称域间路由选择
在不同自治系统的路由器之间交换路由信息，为分组在不同自治系统间选择最优的路径
具体的协议有：BGP

使用层次路由时，OSPF 将一个自治系统再划分为若干区域 Area

每个路由器都知道在本区域内如何把分组路由到目的地的细节，但不知道其他区域的内部结构

*路由回路的根本原因 P142 T5

在距离-向量路由协议中，“好消息传得快，而坏消息传得慢”

慢收敛是导致发生路由回路的根本原因

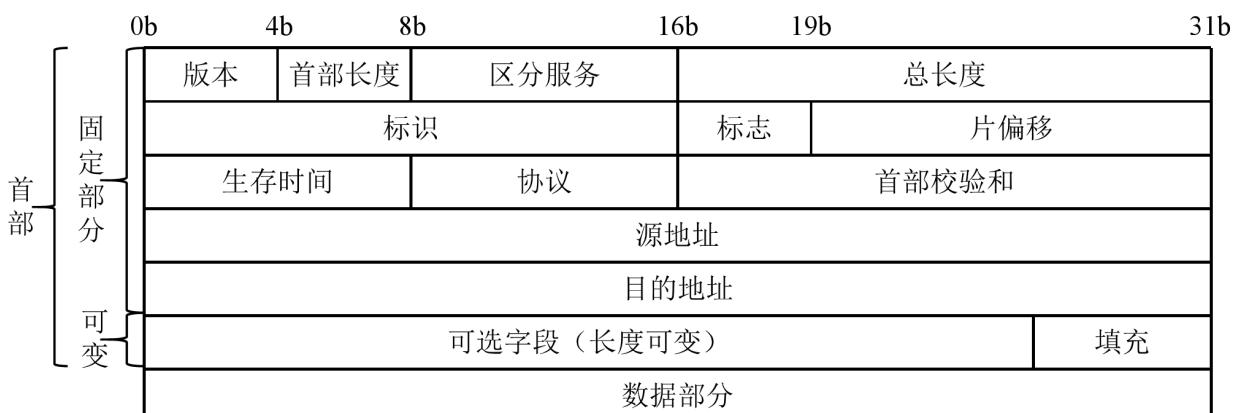
4.3 IPv4

• IPv4 分组

IP 协议定义数据传送的基本单元——IP 分组及其确切的数据格式

一个 IP 分组由首部和数据部分组成

首部前一部分的长度固定，共 20B，是所有 IP 分组必须具有的



- 1) 版本，占 4 位。指 IP 协议的版本，目前广泛使用的版本号为 4
 - 2) 首部长度，占 4 位。可表示的最大十进制数是 15，以 32 位 (4B) 为单位
最大值为 60B ($15 \times 4B$)，最常用的首部长度是 20B，此时不使用任何选项（即可选字段）
 - 3) 区分服务，占 8 位。此字段是用来获得更好的服务，实际上一直没有被使用过
 - 4) 总长度，占 16 位。指首部和数据之和的长度，单位为字节 (1B)，最大长度为 $2^{16}-1=65535B$
IP 数据报的总长度一定不能超过数据链路层的 MTU 值
以太网帧的最大传送单元 MTU 为 1500B，须把过长的数据报进行分片处理
 - 5) 标识 identification，占 16 位。它是一个计数器，每产生一个数据报就加 1，并值赋给标识字段
但它不是“序号”，IP 是无连接服务
分片时，此时每个数据报片都复制一次标识号，以便能正确重装成原来的数据报
 - 6) 标志 flag，占 3 位。最低位记为 MF (More Fragment)
MF=1 表示后面还有分片，MF=0 表示这是最后一个分片
标志字段中间的一位记为 DF (Don't Fragment)，只有当 DF=0 时才允许分片
 - 7) 片偏移，占 13 位。代表较长的分组在分片后，某片在原分组中的相对位置
片偏移以 64b (8B) 为偏移单位，除最后一个分片外，每个分片的长度一定是 8B 的整数倍
 - 8) 生存时间 TTL, Time To Live，占 8 位。数据报在网络中可通过的路由器数的最大值
标识分组在网络中的寿命，以确保分组不会永远在网络中循环
路由器在转发分组前，先把 TTL 减 1。若 TTL 被减为 0，则该分组必须丢弃
 - 9) 协议，占 8 位。指出此分组携带的数据使用何种协议，即数据部分应上交给哪个协议进行处理
如：TCP (6)、UDP (17) 等
 - 10) 首部校验和，占 16 位。首部校验和只校验分组的首部，而不校验数据部分
 - 11) 源地址字段，占 4B。标识发送方的 IP 地址
 - 12) 目的地址字段，占 4B。标识接收方的 IP 地址
- 在首部固定部分的后面是一些可选字段，其长度可变，用来提供错误检测及安全等机制

常见协议名与对应字段值：

协议名	ICMP	IGMP	IP	TCP	EGP	IGP	UDP	IPv6	ESP	OSPF
协议字段值	1	2	4	6	8	9	17	41	50	89

注：三个基本单位首部长度 4B、总长度 1B、片偏移 8B

• IPv4 地址

连接到因特网上的每台主机（或路由器）都分配一个 32 比特的全球唯一标识符，即 IP 地址
IP 地址由 ICANN (Internet Corporation for Assigned Names and Numbers) 进行分配

IP 地址分类：

A类 (1~126)	0	网络号	主机号
B类 (128~191)	1 0	网络号	主机号
C类 (192~223)	1 1 0	网络号	主机号
D类 (224~239)	1 1 1 0		多播地址
E类 (240~255)	1 1 1 1		保留今后使用

由网络号和主机号组成的 IP 地址 (IP 地址 ::= {<网络号>, <主机号>}) 在因特网范围内是唯一的
网络号标志主机（或路由器）所连接到的网络。一个网络号在整个因特网范围内必须是唯一的
主机号标志该主机（或路由器）。一个主机号在它的网络号所指明的网络范围内必须是唯一的

特殊 IP 地址：

- 1) 主机号全为 0 表示本网络本身，如 202.98.174.0
- 2) 主机号全为 1 表示本网络的广播地址，又称直接广播地址，如 202.98.174.255
- 3) 127.x.x.x 保留为环回自检 Loopback Test 地址，此地址表示任意主机本身
 目的地址为环回地址的 IP 数据报永远不会出现在任何网络上
- 4) 32 位全为 0，即 0.0.0.0 表示本网络上的本主机
- 5) 32 位全为 1，即 255.255.255.255 表示整个 TCP/IP 网络的广播地址，又称受限广播地址
 由于路由器对广播域的隔离，255.255.255.255 等效为本网络的广播地址

三类地址的适用范围：

网络类别	最大可用网络数	第一个可用网络号	最后一个可用网络号	每个网络中最大主机数
A	2^7-2	1	126	$2^{24}-2$
B	2^{14}	128	191.255	$2^{16}-2$
C	2^{21}	192.0.0	223.255.255	2^8-2

IP 地址特点：

- 1) 每个 IP 地址都由网络号和主机号两部分组成，因此 IP 地址是一种分等级的地址结构
 - ①IP 地址管理机构在分配 IP 地址时只分配网络号，而主机号则由得到该网络的单位自行分配，方便了 IP 地址的管理
 - ②路由器仅根据目的主机所连接的网络号来转发分组而不考虑目标主机号，减小了路由表所占的存储空间
- 2) IP 地址是标志一台主机（或路由器）和一条链路的接口
当一台主机同时连接到两个网络时，该主机就必须同时具有两个相应的 IP 地址
每个 IP 地址的网络号必与所在网络的网络号相同，且这两个 IP 地址的主机号是不同的
IP 网络上的一个路由器必然至少应具有两个 IP 地址（每个端口至少分配一个 IP 地址）
- 3) 用转发器或桥接器（网桥等）连接的若干 LAN 仍然是同一个网络（同一个广播域）
因此该 LAN 中所有主机的 IP 地址的网络号必须相同，但主机号必须不同
- 4) 在 IP 地址中，所有分配到网络号的网络（无论是 LAN 还是 WAN）都是平等的
- 5) 在同一个局域网上的主机或路由器的 IP 地址中的网络号必须是一样的
路由器总具有两个或两个以上的 IP 地址，每个端口都有一个不同网络号的 IP 地址

• 私有 IP 与网络地址转换 NAT

IP 地址分为公网 IP 地址和私有 IP 地址

公网 IP 地址是在因特网上使用的 IP 地址

私有 IP 地址则是在局域网中使用的 IP 地址，在互联网上不使用

网络地址转换 NAT (Network Address Translation) 通过将专用网络地址转换为公用地址

对外隐藏内部管理的 IP 地址，使得整个专用网只需要一个全球 IP 地址就可以与因特网连通
隐藏了内部网络结构，从而降低了内部网络受到攻击的风险

专用网本地 IP 地址是可重用的，所以 NAT 大大节省了 IP 地址的消耗

在因特网中的所有路由器，对目的地址是私有地址的数据报一律不进行转发

采用私有 IP 地址的互联网络称为专用互联网或本地互联网。私有 IP 地址也称可重用地址

私有 IP 地址网段：

A 类：1 个 A 类网段，即 **10.0.0.0~10.255.255.255**

B 类：16 个 B 类网段，即 **172.16.0.0~172.31.255.255**

C 类：256 个 C 类网段，即 **192.168.0.0~192.168.255.255**

装有 NAT 软件的路由器叫做 NAT 路由器，它至少有一个有效的外部全球地址 IP

NAT 路由器使用 NAT 转换表进行本地 IP 地址和全球 IP 地址的转换

NAT 转换表的表项需要管理员添加，存放着{本地 P 地址：端口}到{全球 P 地址：端口}的映射
通过这种映射方式，可让多个私有 IP 地址映射到一个全球 IP 地址

- 子网划分与子网掩码，无分类编址 CIDR 与链路聚合

子网划分：

把 IP 从两级 IP 地址提升到三级 IP 地址，从主机号借用若干个位作为子网号

IP 地址={<网络号>, <子网号>, <主机号>}

划分子网仅提高 IP 地址的利用率，并不增加网络数量

数据报的接收：

从其他网络发送给本单位某个主机的 IP 数据报，根据 IP 数据报的目的网络号，先找到连接在本单位网络上的路由器。然后此路由器在收到 IP 数据报后，再按目的网络号和子网号找到目的子网。最后将 IP 数据报直接交付目的主机

子网掩码：

子网掩码 Subnet Mask 是一个与 IP 地址相对应的、长 32bit 的二进制串

它由一串 1 和跟随的一串 0 组成。其中，1 对应网络号及子网号，0 对应主机号

将 IP 地址和其对应的子网掩码逐位“与”（逻辑 AND 运算），就可得出相应子网的网络地址

无分类编址 CIDR, Classless Inter-Domain Routing:

无分类域间路由选择 CIDR 是在变长子网掩码的基础上提出的一种消除传统 A、B、C 类网络划分，并可在软件支持下实现超网构造的一种 IP 地址的划分方法

IP::= {<网络前缀>, <主机号>}

CIDR 还使用“斜线记法”（或称 CIDR 记法），即 IP 地址/网络前缀所占比特数

路由聚合：

将网络前缀都相同的连续 IP 地址组成“CIDR 地址块”

一个 CIDR 地址块可以表示很多地址，这种地址的聚合称为路由聚合，或称构成超网

路由聚合使得路由表中的一个项目可以表示多个原来传统分类地址的路由

有利于减少路由器之间的信息的交换，从而提高网络性能

最长前缀匹配（最佳匹配）：

使用 CIDR 时，路由表中的每个项目由“网络前缀”和“下一跳地址”组成

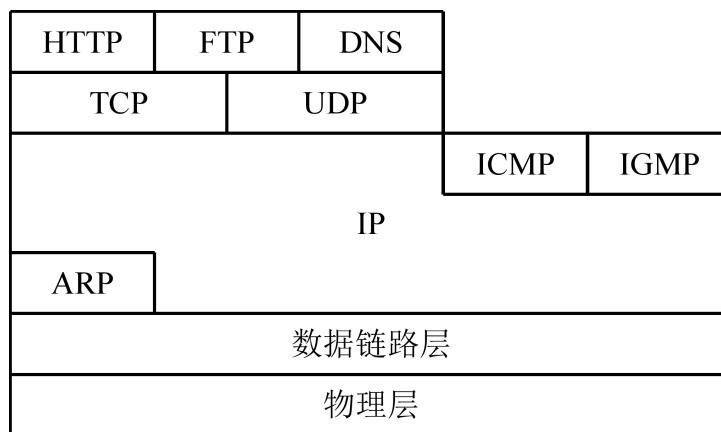
在查找路由表时从匹配结果中选择具有最长网络前缀的路由

因为网络前缀越长，其地址块就越小，因而路由就越具体

通常将无分类编址的路由表存放在一种层次式数据结构中，然后自上而下地按层次进行查找

这里最常用的数据结构就是二叉线索

- TCP/IP 协议栈



- 地址解析协议 ARP, Address Resolution Protocol

逆地址解析协议 RARP, Reverse Address Resolution Protocol

地址解析协议 ARP 已知一个机器（主机或路由器）的 IP 地址，需要找出其相应的硬件地址

逆地址解析协议 RARP 使知道自己硬件地址的主机，找出其 IP 地址

主机 ARP 高速缓存中存放一个从 IP 地址到硬件地址的映射表，称 ARP 表

使用 ARP 来动态维护此 ARP 表，新增或超时删除

主机 A 在 ARP 高速缓存中找不到主机 B，按以下步骤找出主机 B 的硬件地址：

1. ARP 进程在本局域网上广播发送一个 ARP 请求分组（包含自身 A 硬件地址）目的 MAC 地址为 FF-FF-FF-FF-FF-FF
2. 在本局域网上的所有主机上运行的 ARP 进程都收到此 ARP 请求分组
3. 如果主机 B 的 IP 地址与 ARP 请求分组中要查询的 IP 地址一致，就收下这个 ARP 请求分组，并向主机 A 单播发送 ARP 响应分组（包含自身 B 硬件地址）

从 IP 地址到硬件地址的解析是自动进行的，主机的用户并不知道这种地址解析过程

- 动态主机配置协议 DHCP, Dynamic Host Configuration Protocol

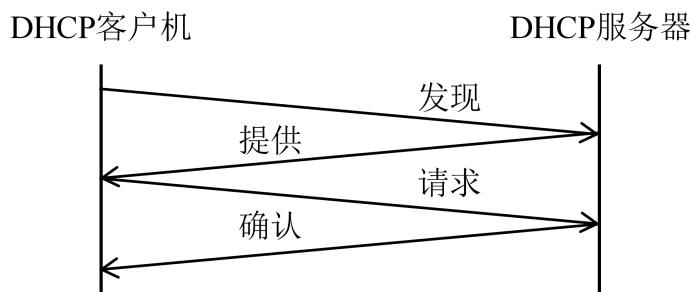
用于给主机动态地分配 IP 地址，它提供了即插即用的联网机制

这种机制允许一台计算机加入新的网络和获取 IP 地址而不用手工参与

DHCP 是应用层协议，它是基于 UDP 的

DHCP 使用客户/服务器模式，工作原理如下：

- 1) DHCP 客户机广播“DHCP 发现”消息，试图找到网络中的 DHCP 服务器
源地址为 0.0.0.0，目的地址为 255.255.255.255
- 2) DHCP 服务器收到“DHCP 发现”后广播“DHCP 提供”消息，包括提供的 IP 地址
源地址为 DHCP 服务器地址，目的地址为 255.255.255.255
- 3) DHCP 客户机收到“DHCP 提供”消息，如果接受该 IP 地址，就广播“DHCP 请求”
源地址为 0.0.0.0，目的地址为 255.255.255.255
- 4) DHCP 服务器广播“DHCP 确认”消息，将 IP 地址分配给 DHCP 客户机
源地址为 DHCP 服务器地址，目的地址为 255.255.255.255



DHCP 允许网络上配置多台 DHCP 服务器，当 DHCP 客户机发出“DHCP 发现”消息后有可能收到多个应答消息。DHCP 客户机只会挑选最先到达的一个

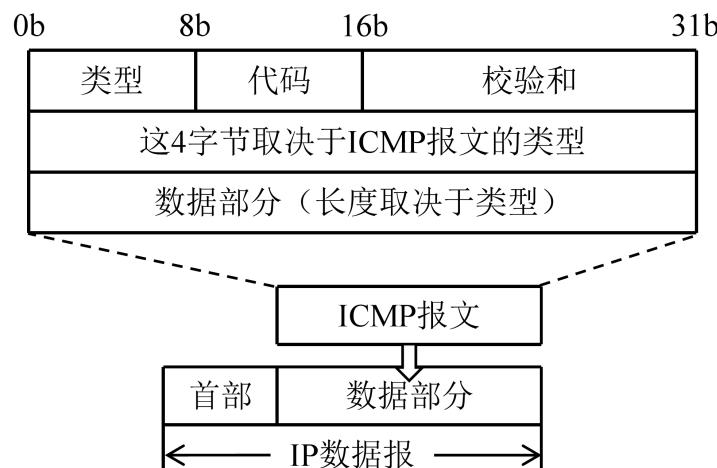
DHCP 服务器分配给 DHCP 客户的 IP 地址是临时的，DHCP 客户只能在一段有限的时间（租用期）内使用这个分配到的 IP 地址。租用期的数值应由 DHCP 服务器自己决定，DHCP 客户也可在自己发送的报文中提出对租用期的要求

- 网际控制报文协议 ICMP, Internet Control Message Protocol

ICMP 允许主机或路由器报告差错情况和提供有关异常情况的报告

ICMP 报文作为 IP 层数据报的数据，加上数据报的头部，组成 IP 数据报发送

ICMP 是网络层协议，ICMP 报文格式：



ICMP 差错报告报文：

- 1) 终点不可达。当路由器或主机不能交付数据报时就向源点发送终点不可达报文
- 2) 源点抑制。已不再使用
- 3) 时间超过。路由器收到 TTL 为 0 的数据报时，丢弃该数据报，向源点发送时间超过报文
 当终点在预先规定的时间内不能收到一个数据报的全部数据报片时
 把已收到的数据报片都丢弃，并向源点发送时间超过报文
- 4) 参数问题。当路由器或目的主机收到的数据报的首部中有的字段的值不正确时
 就丢弃该数据报，并向源点发送参数问题报文
- 5) 改变路由（重定向）。路由器把改变路由报文发送给主机
 让主机知道下次应将数据分组发送给哪个路径

ICMP 询问报文：

- 1) 回送请求和回答。由主机或路由器向一个特定的目的主机发出的询问
 收到此报文的主机必须给源主机或路由器发送 ICMP 回送回答报文
- 2) 时间戳请求和回答。请某台主机或路由器回答当前的日期和时间
- 3) 地址掩码请求和回答报文
- 4) 路由器询问和通告报文

不应发送 ICMP 差错报告报文的情况：

- 1) 对 ICMP 差错报告报文不再发送 ICMP 差错报告报文
- 2) 对第一个分片的数据报片的所有后续数据报片都不发送 ICMP 差错报告报文
- 3) 对具有组播地址的数据报都不发送 ICMP 差错报告报文
- 4) 对具有特殊地址（如 127.0.0.0 或 0.0.0.0）的数据报不发送 ICMP 差错报告报文

ICMP 的应用举例：

分组网间探测 PING, Packet Inter Net Groper

用来测试两台主机之间的连通性，使用了 ICMP 回送请求与回送回答报文

PING 命令工作在应用层，直接使用网络层 ICMP

Traceroute (Windows 中是 Tracert)

定位源计算机和目标计算机间的所有路由器，工作在网络层，使用 ICMP 时间超过报文

4.4 IPv6

• IPv6 的主要特点

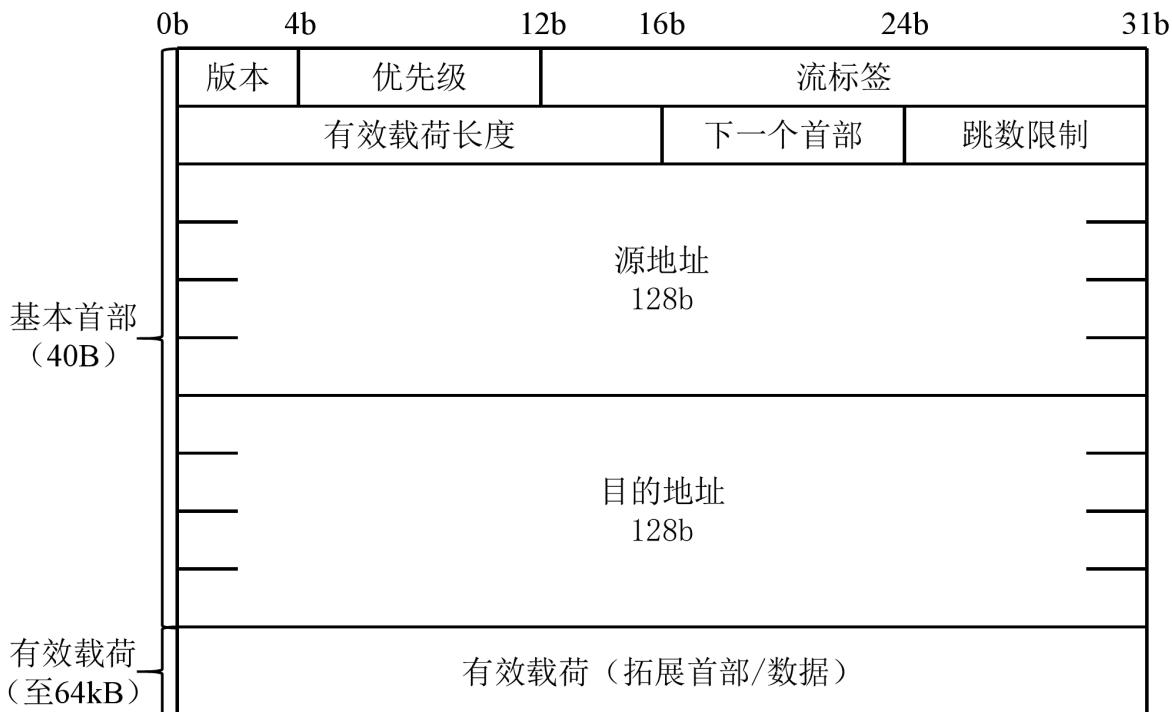
解决“IP 地址耗尽”问题的措施：

- ①采用无类别编址 CIDR，使 IP 地址的分配更加合理
- ②采用网络地址转换 NAT 方法以节省全球 IP 地址
- ③采用具有更大地址空间的新版本的 IPv6（从根本上解决）

IPv6 的主要特点如下：

- 1) IPv6 将地址从 32 位 (4B) 扩大到 128 位 (16B)，更大的地址空间
- 2) IPv6 将 IPv4 的校验和字段彻底移除，以减少每跳的处理时间
- 3) IPv6 将 IPv4 的可选字段移出首部，变成了扩展首部，成为灵活的首部格式
 路由器通常不对扩展首部进行检查，大大提高了路由器的处理效率
- 4) IPv6 支持即插即用（即自动配置），不需要 DHCP 协议
- 5) IPv6 首部长度必须是 8B 的整数倍，IPv4 首部是 4B 的整数倍
- 6) IPv6 只能在主机处分片，IPv4 可以在路由器和主机处分片
- 7) ICMPv6：附加报文类型“分组过大”
- 8) IPv6 支持资源的预分配，支持实时视像等要求，保证一定的带宽和时延的应用
- 9) IPv6 取消了协议字段，改成下一个首部字段
- 10) IPv6 取消了总长度字段，改用有效载荷长度字段
- 11) IPv6 取消了服务类型字段
- 12) 增大了安全性。身份验证和保密功能是 IPv6 的关键特征

13) IPv6 定义了三种不同的地址类型，包括单播、多播、任意播



• IPv6 地址

IPv6 基本类型地址：

- 1) 单播。单播就是传统的点对点通信。
 - 2) 多播。多播是一点对多点的通信，分组被交付到一组计算机的每台计算机。
 - 3) 任播。IPv6 新增加的。目的站是一组计算机，只交付其中一台计算机，通常距离最近的
- IPv6 地址表示方法：

IPv6 不使用点分十进制进行表示，而是用冒分十六进制表示法基本格式为 X:X:X:X:X:X:X:X
其中每个 X 代表 16 个 bit，以十六进制显示，前导的 0 可以省略表示

例如：ABCD:EF01:2345:6789:ABCD:EF01:2345:6789

每个 X 中前面连续的 0 可以省略不写，若整个 X 都为 0，则用一个 0 表示整个 X

例如：2001:0DB8:0000:0023:0008:0800:200C:417A=2001:DB8:0:23:8:800:200C:417A

同时为了方便可以对多个零进行压缩，也就是0位压缩表示法，可以把连续的一段 0 压缩为“::”
为保证地址解析的唯一性，地址中 “:” 只能出现一次

例如：FF01:0:0:0:0:0:1101=FF01::1101

IPv6 地址分级：

第一级（顶级）指明全球都知道的公共拓扑

第二级（场点级）指明单个场点

第三级指明单个网络接口

IPv6 地址采用多级体系主要是为了使路由器能够更快地查找路由

IPv6 向 IPv4 过渡的策略：

双栈协议：在一台设备上同时启用 IPv4 协议栈和 IPv6 协议栈

隧道技术：通过使用互联网络的基础设施在网络之间传递数据的方式。使用隧道传递的数据（或
负载）可以是不同协议的数据帧或包。隧道协议将其它协议的数据帧或包重新封装
然后通过隧道发送

4.5 路由协议

- 自治系统 AS, Autonomous System

单一技术管理下的一组路由器，这些路由器使用一种 AS 内部的路由选择协议和共同的度量来确定分组在该 AS 内的路由，同时还使用一种 AS 之间的路由选择协议来确定分组在 AS 之间的路由

- 域内路由与域间路由

内部网关协议 IGP, Interior Gateway Protocol

即在一个自治系统内部使用的路由选择协议，而这与在互联网中的其他自治系统选用什么路由选择协议无关。目前这类路由选择协议使用得最多，如路由信息协议 RIP (Routing Information Protocol) 和开放最短路径优先协议 OSPF (Open Shortest Path First)

外部网关协议 EGP, External Gateway Protocol

若源主机和目的主机处在不同的自治系统中（这两个自治系统可能使用不同的内部网关协议），当数据报传到一个自治系统的边界时，就需要使用一种协议将路由选择信息传递到另一个自治系统中。这样的协议就是外部网关协议。目前使用最多的外部网关协议 BGP (Border Gateway Protocol) 的版本 4 (BGP-4)

自治系统之间的路由选择叫做域间路由选择 Interdomain Routing

自治系统内部的路由选择叫做域内路由选择 Intradomain Routing



- 路由信息协议 RIP, Routing Information Protocol

RIP 是内部网关协议 IGP 中最先得到广泛应用的协议

RIP 是一种分布式的基于距离向量的路由选择协议，其最大优点就是简单

RIP 是应用层协议，它使用 UDP 传送数据（端口 520）

RIP 选择的路径不一定是时间最短的，但一定是具有最少路由器的路径

RIP 协议首部开销是 20B

RIP 规定：

- 1) 每个路由器都要维护从它自身到其他每个目的网络的距离记录（即“距离向量”）
- 2) 距离也称跳数 Hop Count，从一个路由器到直接连接网络的距离（跳数）为 1
每经过一个路由器，距离（跳数）加 1
- 3) RIP 认为好的路由就是它通过的路由器的数目少，即优先选择跳数少的路径
- 4) RIP 允许一条路径最多只能包含 15 个路由器，距离等于 16 时，它表示网络不可达
防止距离向量路由出现环路时，减少网络拥塞的可能性，只适用于小型互联网
- 5) RIP 默认在任意两个使用 RIP 的路由器之间每 30 秒广播一次 RIP 路由更新信息
以便自动建立并维护路由表（动态维护）
- 6) 在 RIP 中不支持子网掩码的 RIP 广播，所以 RIP 中每个网络的子网掩码必须相同
但在新的 RIP2 中，支持变长子网掩码和 CIDR

RIP 路由表的更新：

①仅和相邻路由器交换信息

②路由器交换的信息是当前路由器所知道的全部信息，即自己的路由表

③按固定的时间间隔交换路由信息，如每隔 30 秒

每个路由表项目都有三个关键数据：<目的网络 N, 距离 d, 下一跳路由器地址 X>

对于每个相邻路由器发送过来的 RIP 报文，执行如下步骤：

1) 对地址为 X 的相邻路由器发来的 RIP 报文，先修改此报文中的所有项目：

把“下一跳”字段中的地址都改为 X，并把所有“距离”字段的值加 1

2) 对修改后的 RIP 报文中的每个项目，执行如下步骤：

原路由表中没有目的网络 N 时

 把该项目添加到路由表中

原路由表中有目的网络 N, 且下一跳路由器的地址是 X 时

 用收到的项目替换原路由表中的项目

原路由表中有目的网络 N, 且下一跳路由器的地址不是 X 时

 如收到项目中的距离 d 小于路由表中的距离, 就用收到的项目替换原项目

 否则什么也不做

3) 如果 180 秒 (RIP 默认超时时间为 180 秒) 还没有收到相邻路由器的更新路由表:

 把此相邻路由器记为不可达路由器, 即把距离设置为 16 (表示不可达)

4) 返回

RIP 的优点: 实现简单、开销小、收敛过程较快

RIP 的缺点:

1) RIP 限制了网络的规模, 它能使用的最大距离为 15 (16 表示不可达)

2) 路由器之间交换的是路由器中的完整路由表, 因此网络规模越大, 开销也越大

3) 网络出现故障时, 会出现慢收敛现象 (可能发生环路)

• 开放最短路径优先 OSPF 协议

OSPF 协议是分布式路由选择协议

每一个路由器都要不断地和其他一些路由器交换路由信息

OSPF 是网络层协议, 直接用 IP 数据报传送 (其 IP 数据报首部的协议字段为 89)

OSPF 最主要的特征就是使用分布式的链路状态协议 Link State Protocol

OSPF 协议的三个要点:

1) 使用洪泛法向本自治系统中所有路由器发送信息

2) 发送的信息就是与本路由器相邻的所有路由器的链路状态,

 “链路状态”就是说明本路由器都和哪些路由器相邻, 以及该链路的“度量 metric”

3) 只有当链路状态发生变化时, 路由器才用洪泛法向所有路由器发送此信息

OSPF 特点:

1) OSPF 对不同的链路可根据 IP 分组的不同服务类型 TOS 而设置成不同的代价

2) 多路径间的负载平衡, 出现等价路由 ECMP

 如果到同一个目的网络有多条相同代价的路径, 则可以将通信量分配给这几条路径

3) 所有在 OSPF 路由器之间交换的分组都具有鉴别功能

4) 支持可变长度的子网划分和无分类编址 CIDR

5) 每个链路状态都带上一个 32 位的序号, 序号越大, 状态就越新

6) OSPF 规定每隔一段时间, 如 30 分钟, 要刷新一次数据库中的链路状态

7) 路由器的链路状态只涉及到与相邻路由器的连通状态, 与整个互联网的规模无直接关系
链路状态数据库 Link-State Database:

通过各路由器之间频繁地交换链路状态信息, 所有路由器都能建立一个链路状态数据库

该数据库就是全网的拓扑结构图, 它在全网范围内是一致的 (链路状态数据库的同步)

每个路由器根据全网拓扑结构图, 使用 Dijkstra 最短路径算法, 构造自己的路由表

链路状态发生变化时, 每个路由器重新计算到各目的网络的最优路径, 构造新的路由表

OSPF 的区域划分:

为用于规模很大的网络, OSPF 将一个自治系统再划分为若干更小的范围, 称为区域 area

好处是利用洪泛法交换链路状态信息的范围局限于区域, 减少了整个网络上的通信量

一个区域内部的路由器只知道本区域的完整网络拓扑, 不知道其他区域的网络拓扑情况

区域有层次之分, 上层的域称为主干区域, 负责连通其他下层的区域, 并连接其他自治域

OSPF 五种分组类型:

1) 问候 Hello 分组, 用来发现和维持邻站的可达性

2) 数据库描述分组, 向邻站给出自己的链路状态数据库中的所有链路状态项目的摘要信息

3) 链路状态请求分组, 向对方请求发送某些链路状态项目的详细信息

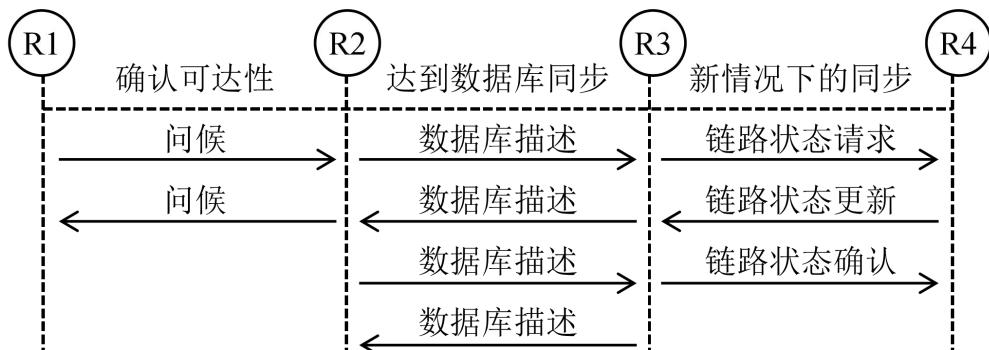
4) 链路状态更新分组，用洪泛法对全网更新链路状态

5) 链路状态确认分组，对链路更新分组的确认

OSPF 规定，每两个相邻路由器每隔 10s 要交换一次问候 Hello 分组，确定哪些邻站是可达的
若 40s 内没有收到某个相邻路由器发来的问候分组，则可认为该相邻路由器是不可达的
其他的四种分组都是用来进行链路状态数据库的同步

同步就是指不同路由器的链路状态数据库的内容是一样的

两个同步的路由器叫做“完全邻接的 Fully Adjacent”路由器



- 外部网关协议 BGP, Border Gateway Protocol

BGP 采用的是路径向量 (Path vector) 路由选择协议

BGP 是应用层协议，它基于 TCP (端口号为 179)

只求寻找一条能够到达目的网络且比较好的路由，而非寻找一条最佳路由

BGP-4 的原理：

每一个自治系统的管理员要选择至少一个路由器作为该自治系统的“BGP 发言人”

一个 BGP 发言人与其他自治系统中的 BGP 发言人要交换路由信息

先建立 TCP 连接 (端口号为 179)，在此连接上交换 BGP 报文来建立 BGP 会话

利用 BGP 会话交换路由信息，如：增加了新的路由、撤销过时的路由、报告出差错的情况

所有发言人相互交换网络可达性信息后，各 BGP 发言人找出到达各个自治系统的较好路由

使用 TCP 连接交换路由信息的两个 BGP 发言人，彼此成为对方的邻站 neighbor 或对等站 peer

BGP-4 的 4 种报文：

1) 打开 Open 报文。用来与相邻的另一个 BGP 发言人建立关系，使通信初始化

2) 更新 Update 报文。用来发送某一路由的信息，以及列出要撤销的多条路由

3) 保活 Keepalive 报文。用来确认打开报文并周期性地证实邻站关系

4) 通知 Notification 报文。用来发送检测到的差错

一开始向邻站进行商谈时发送打开报文，如邻站接受邻站关系，用保活报文响应

BGP 的特点：

1) BGP 交换路由信息的结点数量级是自治系统的数量级，比自治系统中的网络数少很多

2) 每个自治系统中 BGP 发言人的数目很少，使得自治系统之间的路由选择不致过分复杂

3) BGP 支持 CIDR，BGP 路由表包括：

目的网络前缀、下一跳路由器、到达该目的网络所经过的各个自治系统序列

4) 在 BGP 刚运行时，BGP 的邻站交换整个 BGP 路由表，

以后只需在发生变化时更新有变化的部分，节省网络带宽和减少路由器的处理开销

- 三种路由协议的比较

协议	RIP	OSPF	BGP	
类型	内部	内部	外部	
路由算法	距离-向量	链路状态	路径-向量	
传递协议	UDP	IP	TCP	
路径选择	跳数最少	代价最低	较好，非最佳	
交换节点	本结点相邻的路由器	网络中所有的路由器	和本结点相邻的路由器	
交换内容	当前路由器知道的全部信息，即路由表	与本路由器相邻路由器的链路状态	首次	整个路由表
			非首次	有变化的部分

4.6 IP 组播

- 组播的概念

IP 组播（多址广播）是一种允许主机（多播源）发送单一数据包到多台主机的 TCP/IP 网络技术
多播作为一点对多点的通信，是节省网络带宽的有效方法之一，应用 UDP 协议

主机使用因特网组管理 IGMP 协议加入组播组

它们使用该协议通知本地网络上的路由器关于要接收发送给某个组播组的分组的愿望
通过扩展路由器的路由选择和转发功能，可以支持硬件组播的网络上面实现因特网组播
主机组播时仅发送一份数据，只有数据在传送路径出现分岔时才将分组复制后继续转发
组播需要路由器的支持才能实现，能够运行组播协议的路由器称为组播路由器

- 组播地址

组播包括软件组播和硬件组播

硬件组播只在本局域网上进行

软件组播是使用 D 类地址作为目的地址，每个 D 类 IP 地址标志一个组播组
D 类地址的前四位是 1110，地址范围是 224.0.0.0~239.255.255.255

首部中的协议字段值是 2，表明使用 IGMP

因特网上组播的最后阶段，仍要把组播数据报在局域网上用硬件组播交付给组播组的所有成员
组播特点：

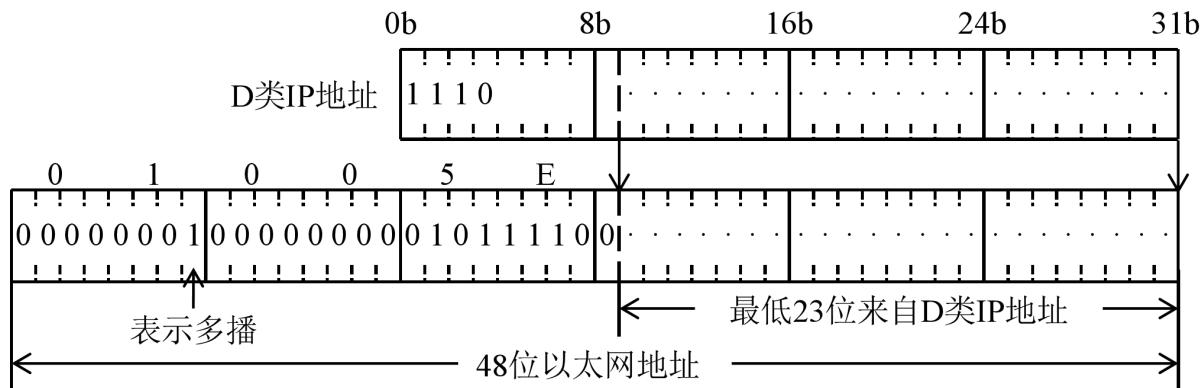
- 1) 组播数据报也是“尽最大努力交付”，不提供可靠交付
- 2) 组播地址只能用于目的地址，而不能用于源地址
- 3) 对组播数据报不产生 ICMP 差错报文
- 4) 并非所有的 D 类地址都可作为组播地址

IANA 拥有的以太网组播地址的范围是从 01-00-5E-00-00-00 到 01-00-5E-7F-FF-FF

在每个地址中，只有 23 位可用作组播，和 D 类 P 地址中的 23 位有一一对应关系

D 类 IP 地址可供分配的有 28 位，可见在这 28 位中，前 5 位不能用来构成以太网的硬件地址
组播 P 地址与以太网硬件地址的映射关系不是唯一的

收到组播数据报的主机，还要在 IP 层利用软件进行过滤，把不是本主机要接收的数据报丢弃



- 网际组管理协议 IGMP, Internet Group Management Protocol

IGMP 的使用范围是本地，并非在因特网范围内对所有多播组成员进行管理的协议

IGMP 不知道 IP 多播组包含的成员数，也不知道这些成员都分布在哪些网络上

IGMP 协议是让多播路由器知道本局域网上是否有主机（进程）参加或退出了某个多播组

IGMP 的工作流程：

第一阶段：当某个主机加入新的多播组时，该主机应向多播组的多播地址发送 IGMP 报文，声明自己要成为该组的成员。本地的多播路由器收到 IGMP 报文后，将组成员关系转发给因特网上的其他多播路由器

第二阶段：因为组成员关系是动态的，因此本地多播路由器要周期性地探询本地局域网上的主机，以便知道这些主机是否还继续是组成员。只要某个组有一个主机响应，那么多播路由器就认为这个组是活跃的

组播路由：

组播路由选择实际上就是要找出以源主机为根结点的组播转发树

每个分组在每条链路上只传送一次，在组播转发树上的路由器不会收到重复的组播数据报

不同的多播组对应于不同的多播转发树

同一个多播组，对不同的源点也会有不同的多播转发树

三种组播路由算法：

第一种是基于链路状态的路由选择

第二种是基于距离-向量的路由选择

第三种可以建立在任何路由器协议之上，因此称为协议无关的组播 PIM

- * 组播路由避免路由环路 P194 T2

树有不存在环路的特性，通过组播转发树既能将组播分组传送到组内的每台主机，又能避免环路
水平分割用于避免距离-向量路由算法中的无穷计数问题

4.7 移动 IP

- 移动 IP 相关概念

移动 IP 技术是移动结点（计算机/服务器等）以固定的网络 IP 地址实现跨越不同网段的漫游功能并保证了基于网络 IP 的网络权限在漫游过程中不发生任何改变

移动结点：具有永久 IP 地址的移动设备（移动站）

本地代理（归属代理）：通常就是连接在归属网络（原始连接到的网络）上的路由器

外地代理（外部代理）：通常就是连接在被访网络（移动到另一地点所接入的网络）上的路由器

永久地址（归属地址/主地址）：移动站点在归属网络中的原始地址

转交地址（辅地址）：可以是外部代理的地址或动态配置的一个地址

- 移动 IP 通信过程

移动节点在归属网络

按传统的 TCP/IP 方式进行通信

移动节点刚进入外部网络

向外地代理进行登记，以获得一个临时的转交地址

外地代理要向移动站的归属代理登记移动站的转交地址
 归属代理知道移动站的转交地址后，会构建一条通向转交地址的隧道
 后续到达该归属代理且要发往移动节点的数据报将被封装并以隧道方式发给外地代理
 外地代理把收到的封装的数据报进行拆封，恢复成原始的 IP 分组，然后发送给移动站
 移动站在被访网络对外发送数据报时，仍然使用自己的永久地址作为数据报的源地址
 移动站移动到另一外地网络
 在新外地代理登记后，新外部代理给归属代理发送新的转交地址（覆盖旧的）
 移动站回到归属网络
 移动站向归属代理注销转交地址

4.7 网络层设备

- 冲突域和广播域

冲突域：连接到同一物理介质上的所有结点的集合，这些结点之间存在介质争用的现象
 广播域：是指接收同样广播消息的结点集合

领域	物理层		数据链路层		网络层
	中继器	集线器	网桥	交换机	路由器
冲突域	不能隔离	不能隔离	能隔离	能隔离	能隔离
广播域	不能隔离	不能隔离	不能隔离	不能隔离	能隔离

在虚拟局域网 VLAN 中，交换机可以隔离广播域，这是一个特殊情况

- 路由器的组成和功能

路由器是一种具有多个输入端口和多个输出端口的专用计算机

其任务是连接不同的网络（连接异构网络）并完成路由转发

在多个逻辑网络（即多个广播域）互连时必须使用路由器

路由器的结构：

路由选择部分也叫做控制部分

其核心构件是路由选择处理

路由选择处理器的任务：

是根据所选定的路由选择协议构造出路由表

经常或定期地和相邻路由器交换路由信息来不断地更新和维护路由表

分组转发部分由三部分组成：交换结构、一组输入端口和一组输出端口（硬件接口）

交换结构 Switching Fabric 又称为交换组织

它的作用：

就是根据转发表 Forwarding Table 对分组进行处理

将某个输入端口进入的分组从一个合适的输出端口转发出去

交换结构本身就是一种网络，这种网络完全包含在路由器之中“在路由器中的网络”

路由器的功能：

分组转发

处理通过路由器的数据流，操作有转发表查询、转发及相关的队列管理和任务调度等

路由计算

通过和其他路由器进行基于路由协议的交互，完成路由表的计算

- 路由表与路由转发

路由表是根据路由选择算法得出的，主要用途是路由选择

标准的路由表有 4 个项目：

目的网络 IP 地址、子网掩码、下一跳 IP 地址、接口

转发表是从路由表得出的，其表项和路由表项有直接的对应关系

转发表中含有：

一个分组将要发往的目的地址，分组的下一跳（MAC 地址）

为了减少转发表的重复项目，可以使用一个默认路由代替所有具有相同“下一跳”

路由表总是用软件来实现的；转发表可以用软件来实现，也可以用特殊的硬件来实现

第 5 章 传输层

5.1 传输层提供的服务

- 传输层的功能

- 1) 传输层提供应用进程之间的逻辑通信（即端到端的通信）

- 2) 复用和分用

复用是指发送方不同的应用进程都可使用同一个传输层协议传送数据

分用是指接收方的传输层在剥去报文的首部后能够把这些数据正确交付到目的应用进程

网络层的复用是指发送方不同协议的数据都可以封装成 IP 数据报发送出去

网络层的分用是指接收方的网络层在剥去首部后把数据交付给相应的协议

- 3) 传输层还要对收到的报文进行差错检测（首部和数据部分）

网络层只检查 IP 数据报的首部，不检验数据部分是否出错

- 4) 提供两种不同的传输协议，即面向连接的 TCP 和无连接的 UDP

- 传输层的寻址与端口

端口的作用：

硬件端口是不同硬件设备进行交互的接口

软件端口是应用层的各种协议进程与传输实体进行层间交互的一种地址

传输层使用的是软件端口

端口用一个 16 位端口号进行标识，端口标识的是主机中的应用进程

端口号只具有本地意义，即端口号只是为了标识本计算机应用层中的各进程

让应用层的各种应用进程将其数据通过端口向下交付给传输层

让传输层知道应当将其报文段中的数据向上通过端口交付给应用层相应的进程

端口是传输层服务访问点 TSAP

*各层服务访问点

数据链路层的 SAP: MAC 地址

网络层的 SAP: IP 地址

传输层的 SAP: 端口

端口号：

端口号长度为 16 位，能够表示 65536 个不同的端口号

- 1) 服务器端使用的端口号

1. 熟知端口号，数值为 0~1023

IANA (互联网地址指派机构) 把这些端口号指派给 TCP/IP 最重要的一些应用程序

2. 登记端口号，数值为 1024~49151

供没有熟知端口号的应用程序使用的，使用这类端口号必须在 IANA 登记

- 2) 客户端使用的端口号，数值为 49152~65535

这类端口号仅在客户进程运行时才动态地选择，又称短暂端口号（也称临时端口）

常见熟知端口号：

应用程序	FTP数据	FTP控制	TELNET	SMTP	DNS	DHCP	TFTP	HTTP	POP3	SNMP	RIP
传输层协议	TCP	TCP	TCP	TCP	UDP	UDP	UDP	TCP	TCP	UDP	UDP
熟知端口号	20	21	23	25	53	67	69	80	110	161	520

套接字 socket:

通过 IP 地址来标识区别不同主机，通过端口号标识区分一台主机中的不同应用进程

端口号拼接到 IP 地址构成套接字 Socket，采用发送方和接收方的套接字组合来识别端点

套接字 Socket= (主机 IP 地址, 端口号)

唯一地标识了网络中的一个主机和其上的一个应用（进程）

- 无连接服务 UDP 与面向连接服务 TCP

无连接的用户数据报协议 UDP

一个无连接的、非可靠的传输层协议，在传送数据之前不需要先建立连接

在 IP 之上仅提供两个服务：多路复用和对数据的错误检查

远程主机的传输层收到 UDP 报文后，不需要给出任何确认

小文件传输 TFTP、域名服务 DNS、简单网络管理 SNMP、路由信息协议 RIP、实时传输 RTP

面向连接的传输控制协议 TCP

TCP 提供面向连接的服务，在传送数据之前必须先建立连接

TCP 只能提供一对一的服务，不提供一对多、多对一或多对多的服务

议数据单元的头部增大很多，还要占用许多的处理机资源

有更多开销，如确认、流量控制、计时器以及连接管理等

文件传输协议 FTP、超文本传输协议 HTTP、远程登录 TELNET、SMTP、POP3 等

5.2 UDP 协议

- UDP 数据报特点

UDP 仅在 IP 的数据报服务之上增加了两个最基本的服务：复用和分用以及差错检测

1) UDP 是无连接的，不会引入建立连接的时延，因此 UDP 具有较高的系统效率

2) UDP 使用尽最大努力交付，即不保证可靠交付，同时也不使用拥塞控制

3) UDP 支持一对一、一对多、多对一和多对多的交互通信

4) UDP 的首部只有 8 个字节，相比于 TCP 的 20 字节，具有较小的首部开销

5) UDP 是面向报文的。即一次发送（交付）一个报文，应用程序必须选择大小合适的报文

- UDP 数据报格式

UDP 数据报包含两部分：UDP 首部和用户数据

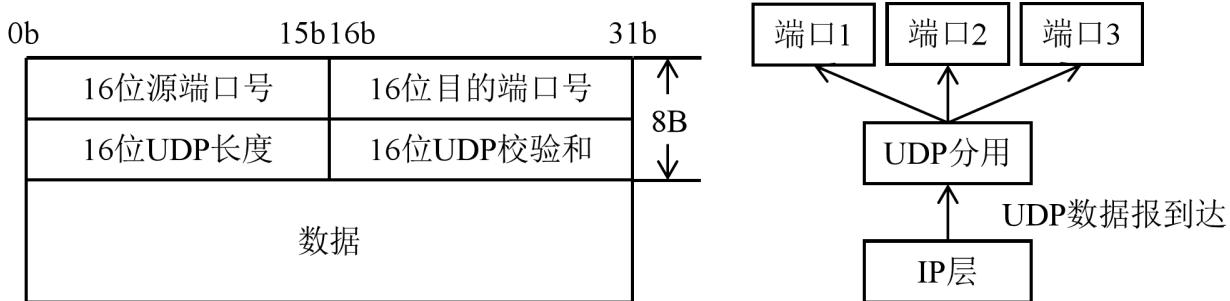
UDP 首部有 8B，由 4 个字段组成，每个字段的长度都是 2B

1) 源端口，源端口号。在需要对方回信时选用，不需要时可用全 0

2) 目的端口，目的端口号。这在终点交付报文时必须使用到

3) 长度。UDP 数据报的长度（包括首部和数据），其最小值是 8（仅有首部）

4) 校验和。检测 UDP 数据报在传输中是否有错。有错就丢弃。该字段是可选的，当源主机不想计算校验和时，则直接令该字段为全 0



传输层根据首部中的目的端口，把 UDP 数据报通过相应的端口上交给应用进程

如收到的 UDP 报文中的目的端口号不正确，则丢弃该报文

并由 ICMP 发送“端口不可达”差错报文给发送方

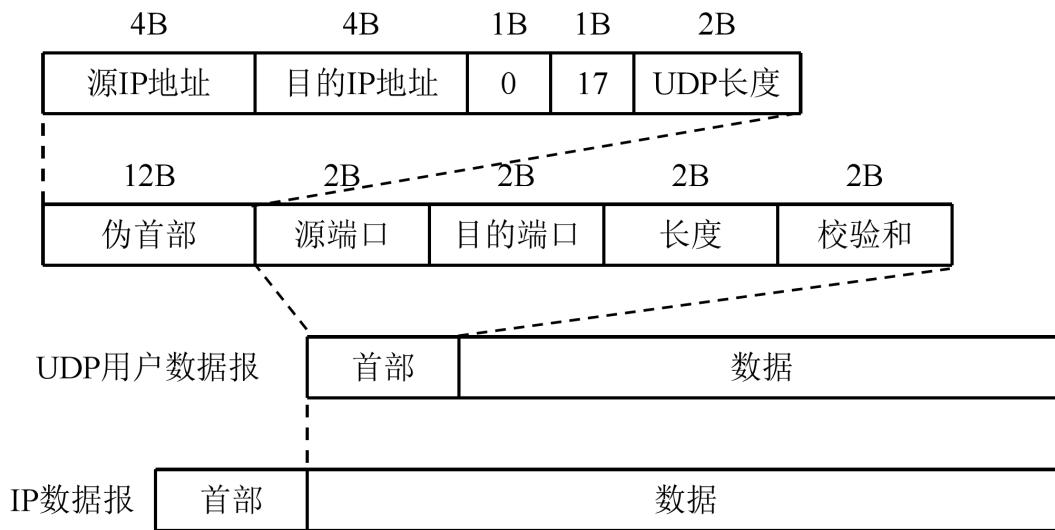
- UDP 校验

在计算校验和时，要在 UDP 数据报之前增加 12B 的伪首部，伪首部并不是 UDP 的真正首部只是在计算校验和时，临时添加在 UDP 数据报的前面，得到一个临时的 UDP 数据报

校验和就是按照这个临时的 UDP 数据报来计算的

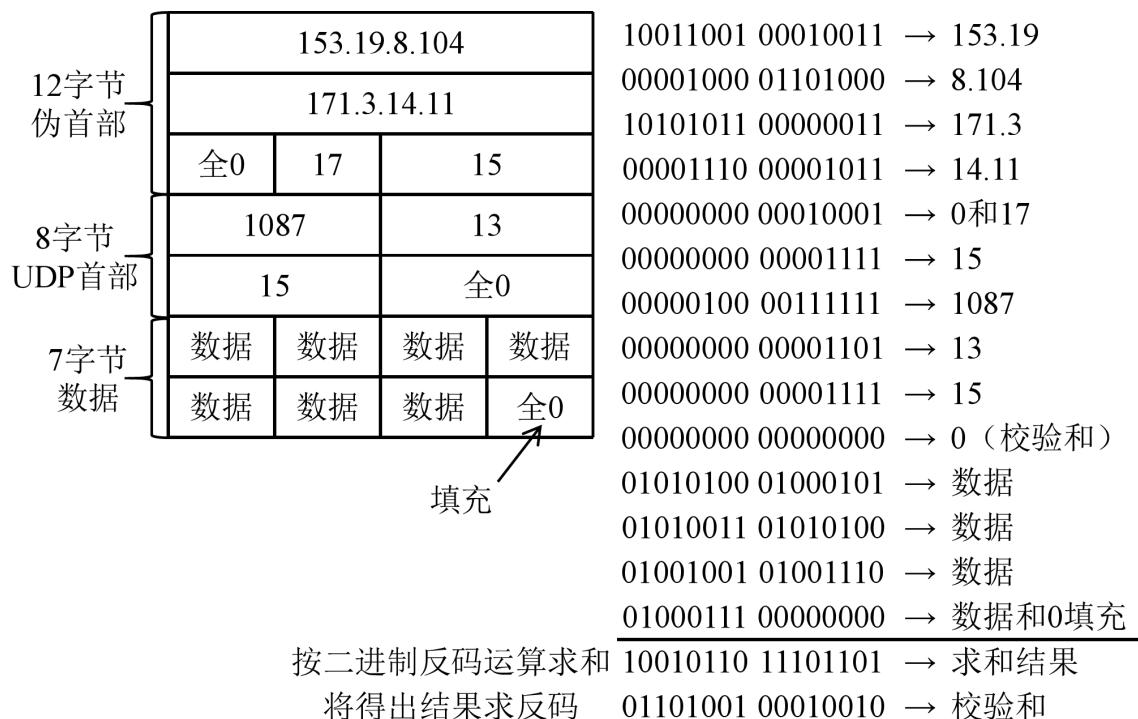
伪首部既不向下传送又不向上递交，而只是为了计算校验和

包含伪首部的校验方式校验了 UDP 数据报首部，IP 数据报的源 IP 地址和目的 IP 地址，数据部分 UDP 校验和的计算方法不适用 CRC 或者奇偶校验等方法，而是使用二进制反码运算求和再取反



校验过程:

1. 发送方首先把全零放入校验和字段并添加伪首部
2. 然后把 UDP 数据报视为许多 16 位的字串接起来
3. 若 UDP 数据报的数据部分非偶数个字节，则在数据末尾填入一个全 0 字节（此字节不发送）
4. 然后按二进制反码计算出这些 16 位字的和，将此和的二进制反码写入校验和字段，并发送
5. 接收方把收到的 UDP 数据报加上伪首部（非偶补零）后按二进制反码求这些 16 位字的和
6. 当无差错时其结果应为全 1，否则就表明有差错出现，接收方就应该丢弃这个 UDP 数据报



5.3 TCP 协议

- TCP 特点

- 1) TCP 是面向连接的传输层协议，TCP 连接是一条逻辑连接
- 2) 每一条 TCP 连接只能有两个端点 socket，每一条 TCP 连接是点对点的
- 3) TCP 提供可靠交付的服务，保证传送的数据无差错、不丢失、不重复且有序
- 4) TCP 提供全双工通信，TCP 连接的两端都设有发送缓存和接收缓存

发送端缓存：

- ①发送应用程序传送给发送方 TCP 准备发送的数据
- ②TCP 已发送但尚未收到确认的数据

接收端缓存：

- ① 按序到达但尚未被接收应用程序读取的数据
- ② 不按序到达的数据

5) TCP 是面向字节流的，把应用程序交下来的数据仅视为一连串的无结构的字节流

6) TCP 根据对方给出的窗口值和当前网络拥塞的程度来决定一个报文段应包含多少个字节

• TCP 报文段

TCP 传送的数据单元称为报文段

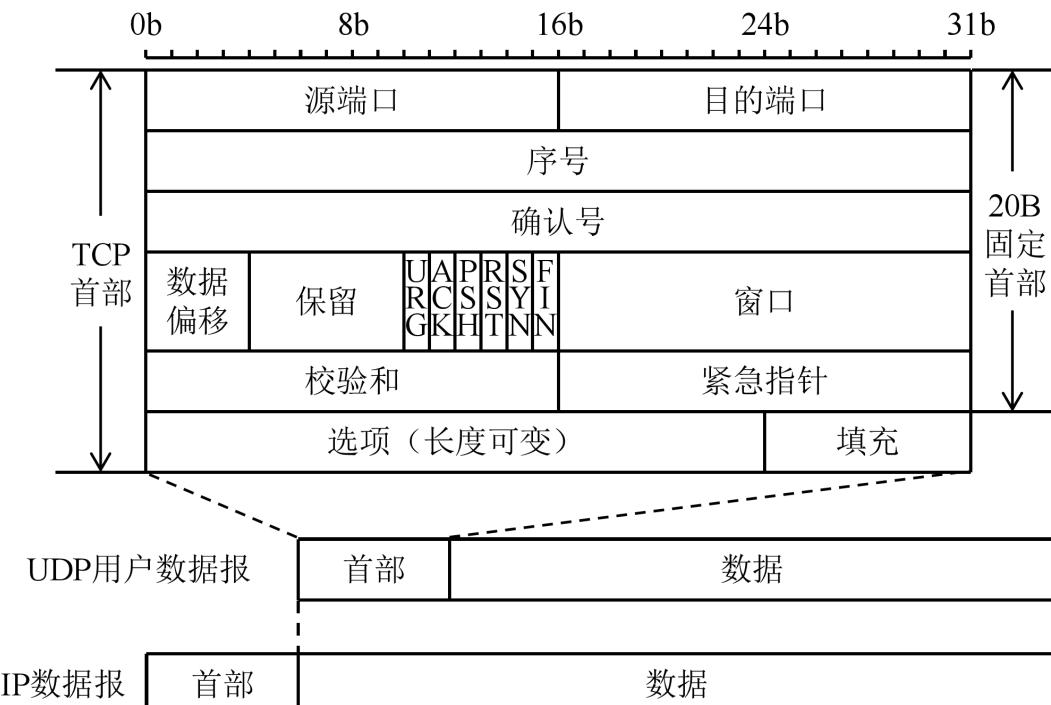
TCP 报文段既可以用来运载数据，又可以用来建立连接、释放连接和应答

一个 TCP 报文段分为首部和数据两部分

整个 TCP 报文段作为 IP 数据报的数据部分封装在 IP 数据报中

首部的前 20B 是固定的。TCP 首部最短为 20B

固定首部后面有 4N 字节是根据需要而增加的选项，长度为 4B 的整数倍



1) 源端口和目的端口字段，各占 2B

端口是传输层与应用层的服务接口，传输层的复用和分用功能都要通过端口才能实现

2) 序号字段，占 4B。序号的值指的是本报文段所发送的数据的第一个字节的序号

3) 确认号字段，占 4B，是期望收到对方的下一个报文段的数据的第一个字节的序号
若确认号为 N，则表明 1~N-1 的所有数据都已正确收到

4) 数据偏移，占 4b，表示首部长度，单位是 32b（以 4B 为计算单位）
当此字段的值为 15 时，达到 TCP 首部的最大长度字节 60B

5) 保留字段，占 6 位，保留为今后使用，但目前应置为 0，该字段可以忽略不计

6) 紧急位 URG，URG=1 时，表明紧急指针字段有效，此报文段中有紧急数据，应尽快传送
需要和紧急指针配套使用，数据从第一个字节到紧急指针所指字节就是紧急数据

7) 确认位 ACK。ACK=1 时，有效。ACK=0 时，无效
TCP 规定，在连接建立后，所有传送的报文段都必须把 ACK 置 1

8) 推送位 PSH (Push)。收到 PSH=1 的报文段，不等整个缓存都填满，尽快交付应用进程

9) 复位位 RST (Reset)。RST=1，表明 TCP 连接出现严重差错，必须释放连接，重新建立连接
10) 同步位 SYN。SYN=1 表示这是一个连接请求或连接接收报文

当 SYN=1，ACK=0 时，表明这是一个连接请求报文

当 SYN=1，ACK=1 时，在响应报文中使用，表明同意建立连接
即 SYN=1 表示这是一个连接请求或连接接收报文

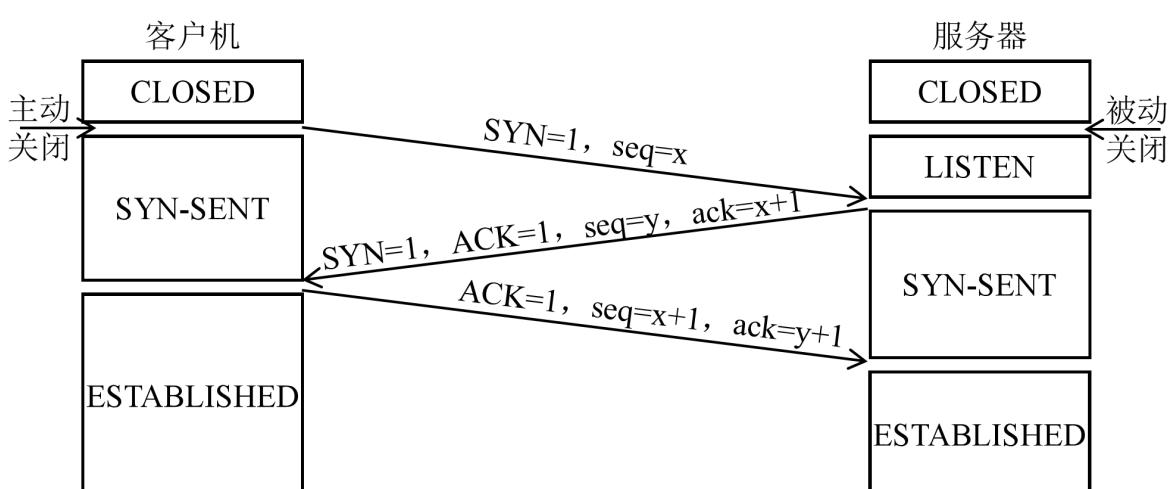
- 11) 终止位 FIN (Finish)。用来释放一个连接
FIN=1 表明此报文段的发送方的数据已发送完毕，并要求释放传输连接
- 12) 窗口字段，占 2 字节。它规定现在允许对方发送的数据量，单位为字节
- 13) 校验和，占 2 字节。校验和字段校验的范围包括首部和数据两部分
在计算校验和时，要在 TCP 报文段的前面加上 12 字节的伪首部
只需将 UDP 伪首部的第 4 个字段，即协议字段的 17 改成 6，其他的和 UDP 一样
- 14) 紧急指针字段，占 16 位，指出报文段中有多少字节紧急数据
紧急数据放在本报文段数据的最前面
- 15) 选项字段，长度可变
TCP 最初只规定了一种选项，即最大报文段长度 MSS，Maximum Segment Size
MSS 是 TCP 报文段中的数据字段的最大长度（注意仅仅是数据字段）
- 16) 填充字段。使整个首部长度是 4 字节的整数倍。

• TCP 连接管理

TCP 是面向连接的协议，每一个 TCP 连接都有三个阶段：建立连接、数据传送和释放连接
TCP 连接的建立都是采用客户服务器方式

主动发起连接请求的应用进程叫做客户机 client，被动等待连接的应用进程叫做服务器 server
TCP 连接的建立：

经历 3 个步骤，通常称为“三次握手”



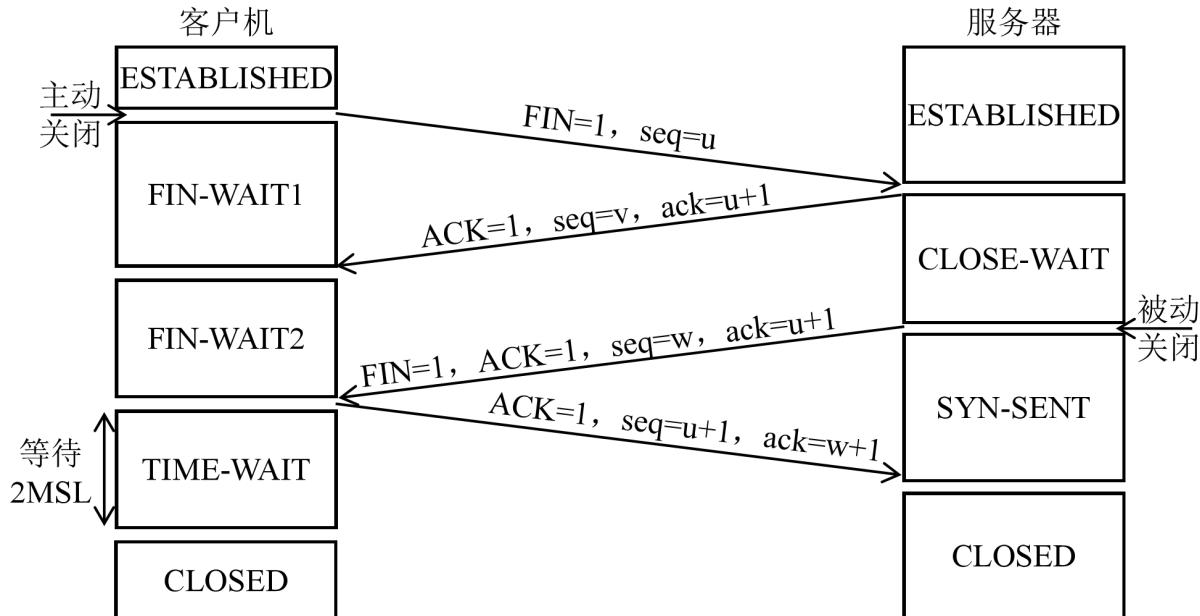
连接建立前，服务器进程处于收听 LISTEN 状态，等待客户的连接请求

- 1) 客户机的 TCP 向服务器发出连接请求报文段
首部中的同步位 SYN=1，并选择序号 seq=x
SYN 报文段不能携带数据，但要消耗掉一个序号
TCP 客户进程进入同步已发送 SYN-SENT 状态
- 2) 服务器的 TCP 收到连接请求报文段，如同意则发回确认，并为 TCP 连接分配缓存和变量
确认报文段中应使 SYN=1，使 ACK=1，确认号 ack=x+1，自己选择的序号 seq=y
确认报文段不能携带数据，也要消耗掉一个序号
TCP 服务器进程进入同步收到 SYN-RCVD 状态
- 3) 当客户机收到确认报文段后，还要向服务器给出确认
这个报文段的 ACK=1，seq=x+1，确认号字段 ack=y+1
该报文段可以携带数据，若不携带数据则不消耗序号
这时，TCP 客户进程进入已建立连接 ESTABLISHED 状态
 - ①发送 TCP 连接请求是： SYN=1 ACK=0 seq=x
 - ②同意 TCP 连接请求是： SYN=1 ACK=1 seq=y ack=x+1
 - ③发送 TCP 连接确认是： SYN=0 ACK=1 seq=x+1 ack=y+1

注：服务器端的资源是在完成第二次握手时分配的，而客户端的资源是在完成第三次握手时分配的，这就使得服务器易于受到 SYN 洪泛攻击

TCP 连接的释放：

经历 4 个步骤，通常称为“四次握手”



参与 TCP 连接的两个进程中的任何一个都能终止该链接

- 1) 客户机打算关闭连接，向 TCP 发送释放连接报文段，停止发送数据，主动关闭 TCP 连接
该报文段的终止位 FIN 置为 1，seq=u (u=已经传送过的最后一字节的序号+1)
FIN 报文段即使不携带数据，也消耗掉一个序号
TCP 客户进程进入终止等待 1FIN-WAIT-1 状态
 - 2) 服务器收到释放连接报文段后即发出确认
确认号是 ack=u+1，seq=v (v=已经发送过的最后一个字节的序号+1)
然后服务器进入关闭等待 CLOSE-WAIT 状态
从客户机到服务器这个方向的连接就释放了，TCP 连接处于半关闭状态
 - 3) 若服务器已经没有要向客户机发送的数据，就通知 TCP 释放连接
发出 FIN=1 的连接释放报文段，seq=w，ack=u+1
服务器进入最后确认 LAST-ACK 状态
 - 4) 客户机收到释放连接报文段后，必须发出确认
在确认报文段中，ACK 字段被置为 1，确认号 ack=w+1，序号 seq=u+1
经过时间等待计时器设置的报文最大生存时间 2MSL (Maximum Segment Lifetime) 后
客户机才进入到连接关闭 CLOSED 状态
- ①客户机发送释放连接: FIN=1 ACK=0 seq=u
②服务器发确认报文段: FIN=0 ACK=1 seq=v ack=u+1
③服务器发出释放通知: FIN=1 ACK=0 seq=w ack=u+1
④客户机发确认报文段: FIN=0 ACK=1 seq=u+1 ack=w+1

• TCP 可靠传输

TCP 提供面向字节的序号、确认、超时和自动重传等机制来达到可靠传输

序号：

TCP 首部的序号字段用来保证数据能有序提交给应用层。

TCP 把数据看成有序的字节流，对发送数据中的每一个字节进行编号

在 TCP 首部中的序号字段指出本报文段所发送数据的第一个字节的序号

不同 TCP 数据段中的数据是按照同一个顺序进行编号

在实际的传输中，TCP 协议使用面向字节的窗口机制对字节的序号和数据的发送进行管理
确认：

确认机制是接收方向发送方发送的每一个字节进行确认

TCP 首部的确认号是期望收到对方的下一个报文段的数据的第一个字节的序号
接收方希望收到的下一个报文段的序号是当前收到数据段的最大编号的下一个字节
TCP 默认使用累计确认，即 TCP 只确认数据流中至第一个丢失字节为止的字节
发送方在规定时间内没有收到接收方发送的确认数据，将进行自动重传未收到确认的数据段超时与自动重传：

导致 TCP 对报文段进行重传的事件：超时和冗余 ACK

1) 超时

TCP 每发送一个报文段，就对这个报文段设置一次计时器
只要计时器设置的重传时间到期但还没有收到确认，就要重传这一报文段
为了计算超时计时器的重传时间，TCP 采用一种自适应算法
记录一个报文段发出的时间，以及收到相应确认的时间
这两个时间之差称为报文段的往返时间 RTT (Round-Trip Time)
保留加权平均往返时间 RTTs，且会随新测量 RTT 样本值的变化而变化
超时重传时间 RTO (Retransmission Time-Out) 略大于 RTTs

2) 冗余 ACK (冗余确认)

发送方通常可在超时事件发生之前通过注意冗余 ACK 来检测丢包情况
冗余 ACK 就是再次确认某个报文段的 ACK，而发送方之前已经收到过该报文段的确认
比期望序号大的失序报文段到达时，发送一个冗余 ACK，指明下一个期待字节的序号
发送方收到 3 个相同冗余 ACK 时，可认为跟在此被确认报文段之后的报文段已经丢失
这时发送方可以立即对 M3 报文执行重传，这种技术通常称为快速重传

• TCP 流量控制

接收方来控制发送方发送数据的速度，以便及时接收和处理数据，以免造成数据溢出和丢失
TCP 提供一种基于滑动窗口协议的流量控制机制
在通信过程中，接收方根据自己接收缓存的大小，动态调整接收窗口 rwnd 的大小

• TCP 拥塞控制

拥塞控制是指防止过多的数据注入网络，保证网络中的路由器或链路不致过载
端点并不了解拥塞发生的细节，对通信连接的端点来说，拥塞往往表现为通信时延的增加
在通信过程中，发送方全局考虑不要使网络发生拥塞，动态调整拥塞窗口 cwnd 的大小
流量控制与拥塞控制的区别：

流量控制：往往是指点对点的通信量的控制，是个端到端的问题（接收端控制发送端）
它所要做的是抑制发送端发送数据的速率，以便使接收端来得及接收
拥塞控制：是让网络能够承受现有的网络负荷，是一个全局性的过程

涉及所有的主机、所有的路由器，以及与降低网络传输性能有关的所有因素

接收窗口与拥塞窗口：

接收窗口 rwnd：接收方根据目前接收缓存大小所许诺的最新窗口值，反映接收方的容量
由接收方根据其放在 TCP 报文的首部的窗口字段通知发送方

拥塞窗口 cwnd：发送方根据自己估算的网络拥塞程度而设置的窗口值，反映网络的当前容量
只要网络未出现拥塞，拥塞窗口就再增大一些，以便把更多的分组发送出去
但只要网络出现拥塞，拥塞窗口就减小一些，以减少注入网络的分组数

发送窗口的上限值应取接收窗口 rwnd 和拥塞窗口 cwnd 中较小的一个，即

$$\text{发送窗口的上限值} = \min[rwnd, cwnd]$$

因特网建议标准定义了进行拥塞控制的 4 种算法：慢开始、拥塞避免、快重传和快恢复
慢开始：

TCP 刚刚连接好，开始发送 TCP 报文段时，令拥塞窗口 cwnd=1，即一个最大报文段长度 MSS
在每收到一个对新的报文段的确认后，将 cwnd 加 1

用这样的方法逐步增大发送方的拥塞窗口 cwnd，可使分组注入到网络的速率更加合理
每经过一个传输轮次（即往返时延 RTT），拥塞窗口 cwnd 就会加倍（指数形式增长）
拥塞窗口 cwnd 增大到一个规定的慢开始门限 ssthresh（阈值）后改用拥塞避免算法

慢开始阶段，若 $2 \times \text{cwnd} > \text{ssthresh}$ ，则下一个 RTT 的 cwnd 应等于 ssthresh
即 cwnd 不能跃过 ssthresh 值

拥塞避免算法：

每经过一个往返时延 RTT，拥塞窗口 cwnd 加 1，按线性规律缓慢增长（加法增大）

当出现一次网络拥塞时，令慢开始门限 ssthresh 等于当前 cwnd 的一半（乘法减小）

根据 cwnd 的大小执行不同的算法，慢开始门限 ssthresh 的用法如下：

当 $\text{cwnd} < \text{ssthresh}$ 时，使用慢开始算法

当 $\text{cwnd} > \text{ssthresh}$ 时，停止使用慢开始算法而改用拥塞避免算法

当 $\text{cwnd} = \text{ssthresh}$ 时，既可使用慢开始算法，又可使用拥塞避免算法（通常做法）

网络拥塞的处理：

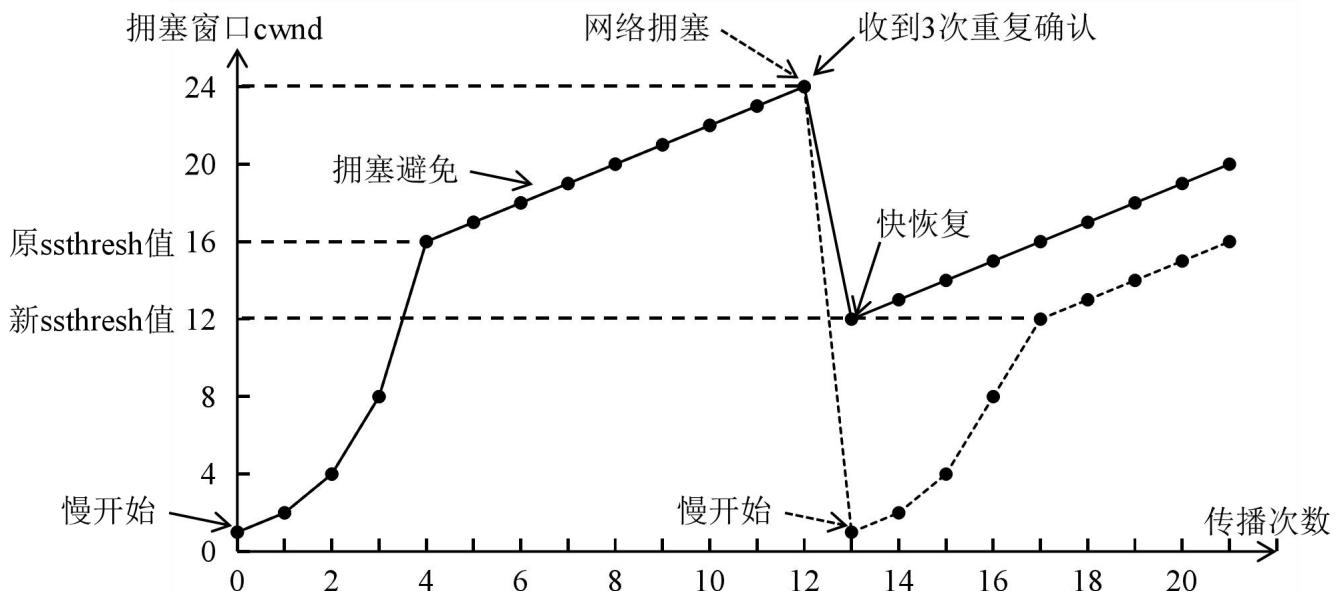
当网络出现拥塞时，只要发送方检测到超时或者事件的发生

把慢开始门限 ssthresh 设置为出现拥塞时发送方 cwnd 值的一半（但不能小于 2）

把拥塞窗口 cwnd 重新设置为 1，执行慢开始算法

拥塞避免并非完全能避免拥塞

拥塞避免是指在拥塞避免阶段把拥塞窗口控制为按线性规律增长，使网络不容易出现拥塞



快重传：

使用冗余 ACK 来检测网络拥塞

首先要求接收方每收到一个失序的报文段后就立即发出重复确认

发送方只要连续收到 3 个重复确认，就应当立即重传对方尚未收到的报文段

不必等待那个报文段所设置的重传计时器超时

快恢复算法：

当发送方收到连续 3 个冗余 ACK（重复确认）时，就执行“乘法减小”算法

把慢开始门限 ssthresh 设置为出现拥塞时发送方 cwnd 的一半

把 cwnd 的值设置为慢开始门限 ssthresh 改变后的数值

然后开始执行拥塞避免算法（加法增大）使拥塞窗口缓慢地线性增大，被称为快恢复

第6章 应用层

6.1 网络应用模型

• 客户/服务器模型 C/S

客户 client 和服务器 server 都是通信中所涉及的两个应用进程

C/S 描述的是进程之间服务和被服务的关系，客户是服务的请求方，服务器是服务的提供方

C/S 工作流程：

- 1) 服务器处于接收请求的状态
- 2) 客户机发出服务请求，并等待接收结果
- 3) 服务器收到请求后，分析请求，进行必要的处理，得到结果并发送给客户机

客户程序必须知道服务器程序的地址，客户机上一般不需要特殊的硬件和复杂的操作系统

服务器上运行的软件则是专门用来提供某种服务的程序，可同时处理多个远程或本地客户的要求

服务器程序不需要知道客户程序的地址，被动地等待并接收来自各地客户的请求

C/S 特点：

- 1) 网络中各计算机的地位不平等
 服务器可以通过对用户权限的限制来达到管理客户机的目的
 整个网络的管理工作由少数服务器担当，因此网络的管理非常集中和方便
- 2) 客户机相互之间不直接通信
- 3) 可扩展性不佳，受服务器硬件和网络带宽的限制，服务器支持的客户机数有限

• 对等连接 P2P 模型

对等连接 P2P (Peer to Peer) 模型在两个主机在通信时并不区分其是服务请求方还是服务提供方
只要两个主机都运行了对等连接软件 (P2P 软件)，它们就可以进行平等的、对等连接通信
任意一对计算机——称为对等方 Peer，双方都可以下载对方已经存储在硬盘中的共享文档

P2P 优点：

- 1) 减轻了服务器的计算压力，消除了对某个服务器的完全依赖，
- 2) 多个客户机之间可以直接共享文档。
- 3) 可扩展性好，传统服务器有响应和带宽的限制，因此只能接受一定数量的请求
- 4) 网络健壮性强，单个结点的失效不会影响其他部分的结点。

P2P 缺点：在获取服务的同时，还要给其他结点提供服务，会占用较多的内存，影响整机速度

6.2 域名系统 DNS, Domain Name System

• DNS 概念

域名系统 DNS 是描述名字—地址映射的分布式计算机系统

DNS 的本质是一种层次结构的、基于域的命名方案和实现这种命名方案的分布式数据库

其作用是提供主机名和 IP 地址间的映射关系，以及提供电子邮件的路由信息

域名在不同的时间可以解析出不同的 IP 地址，可以把多个域名指向同一台主机 IP 地址

采用客户/服务器模型，其协议运行在 UDP 之上，使用 53 号端口

从概念上可将 DNS 分为 3 部分：层次域名空间、域名服务器和解析器

• 层次域名空间

因特网采用层次树状结构的命名方法，域 Domain 是名字空间中一个可被管理的划分

每一个连接在因特网上的主机或路由器，都有唯一的层次结构的名字，即域名 Domain Name

域可以划分为子域，子域还可以划分为子域的子域，这样就形成了顶级域、二级域、三级域等
每个域名都由标号序列组成，而各标号之间用点 “.” 隔开



关于域名中的标号：

- 1) 标号中的英文不区分大小写。
- 2) 标号中除连字符（-）外不能使用其他的标点符号
- 3) 每个标号不超过 63 个字符，多标号组成的完整域名最长不超过 255 个字符
- 4) 级别最低的域名写在最左边，级别最高的顶级域名写在最右边

顶级域名 TLD (Top Level Domain) 分为三大类：

- 1) 国家（地区）顶级域名 nTLD，国家和某些地区的域名
如 “.cn” 表示中国，“.us” 表示美国，“.uk” 表示英国
- 2) 通用顶级域名 gTLD
如 “.com” 公司，“.net” 网络服务机构，“.org” 非营利性组织，“.gov” 政府部门
- 3) 基础结构域名。这种顶级域名只有一个，即 arpa
用于反向域名解析，因此又称反向域名

• 域名服务器

域名服务器负责域名到 IP 地址的解析，采用 C/S 模型，使用 53 号端口

一个服务器所管辖的（或有权限的）范围叫做区 zone

各单位根据具体情况来划分自己管辖范围的区，在一个区中的所有结点必须是能够连通的

每一个区设置相应的权限域名服务器，用来保存该区的所有主机的域名到 IP 地址的映射

DNS 服务器的管辖范围不是以“域”为单位，而是以“区”为单位

DNS 使用大量的域名服务器，它们以层次方式组织

没有一台域名服务器具有因特网上所有主机的映射

域名服务器的类型：

- 1) 根域名服务器是最高层次的域名服务器，也是最重要的域名服务器
所有的根域名服务器知道所有的顶级域名服务器的域名和 IP 地址
本地域名服务器，若对因特网上任何一个域名无法解析，可先求助于根域名服务器
根域名服务器用来管辖顶级域，如.com
不直接把待查询的域名直接转换成 IP 地址
而是告诉本地域名服务器下一步应当找哪个顶级域名服务器进行查询
- 2) 顶级域名服务器 (TLD 服务器)
这些域名服务器负责管理在该顶级域名服务器注册的所有二级域名
当收到 DNS 查询请求时，就给出相应的回答
可能是最后的结果，也可能是下一步应当找的域名服务器的 IP 地址
- 3) 授权域名服务器 (权限域名服务器)
每台主机都必须在授权域名服务器处登记
域名服务器总能将其管辖的主机名转换为该主机的 IP 地址
当一个权限域名服务器还不能给出最后的查询时
就会告诉发出查询请求的 DNS 客户，下一步应当找哪一个权限域名服务器
- 4) 本地域名服务器
当一个主机发出 DNS 查询请求时，这个查询请求报文就发送给本地域名服务器

• 域名的解析过程

域名解析是指把域名映射成为 IP 地址（正向解析）或把 IP 地址映射成域名（反向解析）的过程

客户端需要域名解析时，通过本机的 DNS 客户端构造一个 DNS 请求报文

以 UDP 数据报方式发往本地域名服务器

域名解析有两种方式：

递归查询

主机向本地域名服务器的查询一般都采用递归查询

如主机所询问的本地域名服务器不知道被查询域名的 IP 地址

本地域名服务器就以 DNS 客户的身份，向其他根域名服务器继续发出查询请求报文

递归与迭代相结合的查询

本地域名服务器向根域名服务器的查询通常是采用迭代查询

当根域名服务器收到本地域名服务器的迭代查询请求报文时

给出所要查询的 IP 地址或告诉本地域名服务器下一步应向哪一个域名服务器进行查询

注意事项：

- 1) 本地域名服务器，若对因特网上任何一个域名无法解析，就首先求助于根域名服务器
- 2) 最终结果都可以在对应负责一个区域的权限域名服务器中查到
- 3) DNS 服务器接收到 DNS 查询结果时，将该 DNS 信息缓存在高速缓存，提高查询效率
- 4) DNS 把数据复制到几个域名服务器来保存其中的一个是主域名服务器，其他的是辅助域名服务器。当主域名服务器出故障时，辅助域名服务器可以保证 DNS 的查询工作不会中断

6.3 文件传输协议 FTP, File Transfer Protocol

• FTP 概念与特点

文件传输协议 FTP (File Transfer Protocol) 是因特网上使用得最广泛的文件传输协议
FTP 提供交互式的访问，允许客户指明文件的类型与格式，并允许文件具有存取权限
它屏蔽了各计算机系统的细节，适合于在异构网络中的任意计算机之间传送文件

FTP 采用客户/服务器的工作方式，一个 FTP 服务器进程可同时为多个客户进程提供服务

FTP 使用 TCP 可靠的传输服务

FTP 提供以下功能：

- ① 提供不同种类主机系统（硬、软件体系等都可以不同）之间的文件传输能力
- ② 以用户权限管理的方式提供用户对远程 FTP 服务器上的文件管理能力
- ③ 以匿名 FTP 的方式提供公用文件共享的能力

FTP 的服务器进程组成：

一个主进程，负责接收新的请求；若干从属进程，负责处理单个请求

FTP 工作流程：

- 1) 服务器端打开熟知端口（端口号为 21），使客户进程能够连接上
- 2) 等待客户进程发出连接请求
- 3) 启动从属进程来处理客户进程发来的请求
 从属进程对客户进程的请求处理完毕后即终止
 从属进程在运行期间根据需要还可能创建其他一些子进程
- 4) 回到等待状态，继续接收其他客户进程发来的请求

FTP 服务器必须在整个会话期间保留用户的状态信息

特别是服务器必须把指定的用户账户与控制连接联系起来

服务器必须追踪用户在远程目录树上的当前位置

• 控制连接和数据连接

FTP 在工作时使用两个并行的 TCP 连接，一个是数据连接，一个是控制连接

FTP 使用了一个分离的控制连接，所以也称 FTP 的控制信息是带外 (Out-of-band) 传送的
控制连接（服务器端口号 21）

服务器监听 21 号端口，等待客户连接，建立在这个端口上的连接称为控制连接

控制连接以 7 位 ASCII 格式来传输控制信息（如连接请求、传送请求等）

FTP 客户发出的传送请求，通过控制连接发送给服务器端的控制进程

控制连接并不用来传送文件，在传输文件时还可以使用控制连接

控制连接在整个会话期间一直保持打开状态

数据连接（服务器端口号 20）

服务器端的控制进程在接收到文件传输请求后，就创建“数据传送进程”和“数据连接”

数据连接用来连接客户端和服务器端的数据传送进程

数据传送进程完成文件的传送，在传送完毕后关闭“数据传送连接”并结束运行

数据连接有两种传输模式：主动模式 PORT 和被动模式 PASV

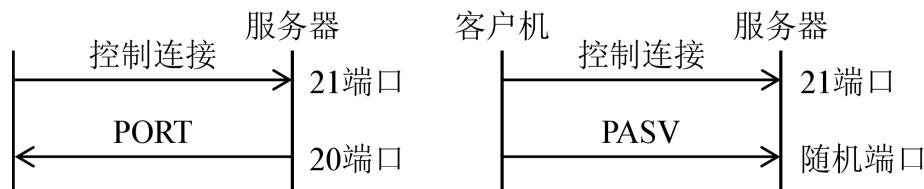
PORT 模式：客户端连接到服务器的 21 端口，登录成功后要读取数据时，客户端随机开放一个端口，并发送命令告知服务器，服务器收到 PORT 命令和端口号后，通过 20 端口和客户端开放的端口连接，发送数据

PASV 模式：客户端要读取数据时，发送 PASV 命令到服务器，服务器在本地随机开放一个端口，并告知客户端，客户端再连接到服务器开放的端口进行数据传输

用 PORT 模式还是 PASV 模式，选择权在客户端

主动模式传送数据是“服务器”连接到“客户端”的端口

被动模式传送数据是“客户端”连接到“服务器”的端口



使用 FTP 时，若要修改服务器上的文件，需要先将此文件传送到本地主机

然后再将修改后的文件副本传送到原服务器，来回传送耗费很多时间

网络文件系统 NFS：

允许进程打开一个远程文件，并能在该文件的某个特定位置开始读写数据

这样，NFS 可使用户复制一个大文件中的一个很小的片段，而不需要复制整个大文件

6.4 电子邮件 E-mail

- 电子邮件系统的组成结构

电子邮件是一种异步通信方式，通信时不需要双方同时在场

电子邮件系统组成构件：用户代理 User Agent、邮件服务器、电子邮件使用的协议

用户代理 UA：用户与电子邮件系统的接口

向用户提供一个很友好的接口来发送和接收邮件，具有撰写、显示和邮件处理的功能

通常就是一个运行在 PC 上的程序（电子邮件客户端软件）

邮件服务器：

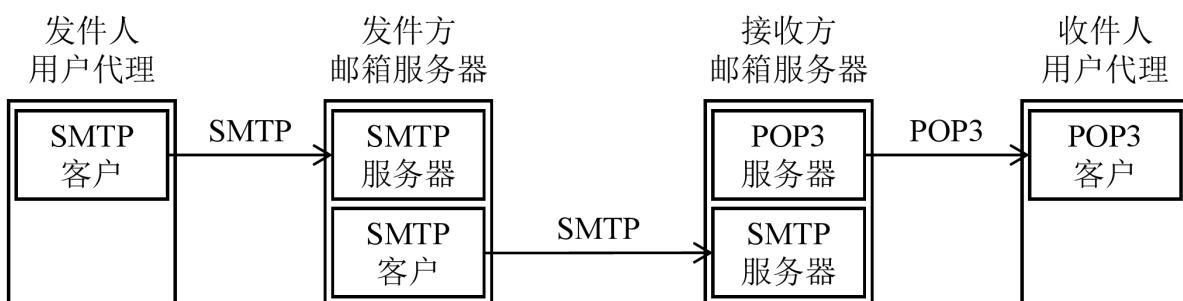
它的功能是发送和接收邮件，同时还要向发信人报告邮件传送的情况

邮件服务器采用客户/服务器方式工作，但它必须能够同时充当客户和服务器

邮件发送协议和读取协议：

邮件发送协议，用户代理向邮件服务器发送邮件或邮件服务器之间发送邮件，如 SMTP

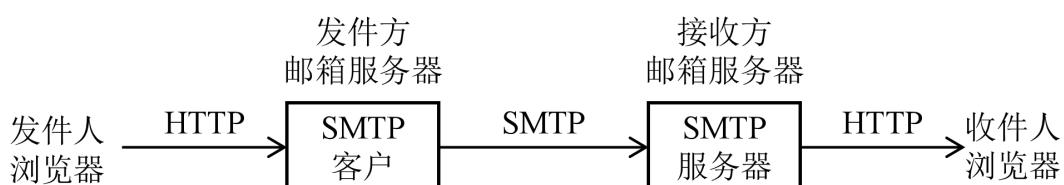
邮件读取协议，用户代理从邮件服务器读取邮件，如 POP3



基于万维网的电子邮件，如 Hotmail、Gmail 等

用户浏览器与邮件服务器之间的邮件发送或接收使用的是 HTTP

在不同邮件服务器之间传送邮件时才使用 SMTP



• 电子邮件格式

一个电子邮件分为信封和内容两大部分，邮件内容又分为首部和主体两部分

RFC 822 规定了邮件的首部格式，而邮件的主体部分则让用户自由撰写

邮件系统自动地将信封所需的信息提取出来并写在信封上，用户不需要亲自填写信封上的信息

邮件内容的首部包含一些首部行，每个首部行由一个关键字后跟冒号再后跟值组成

有些关键字是必需的，有些则是可选的。最重要的关键字是 To 和 Subject

To 是必需的关键字，后面填入一个或多个收件人的电子邮件地址

Subject 是可选关键字，是邮件的主题，反映了邮件的主要内容

From 是必填的关键字，但它通常由邮件系统自动填入

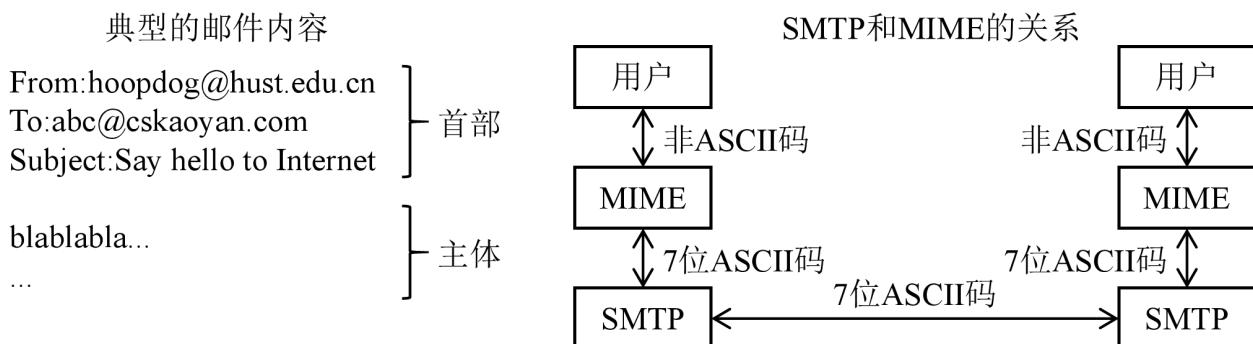
电子邮件地址的规定格式：收件人邮箱名@邮箱所在主机的域名，如 abc@abc.com

收信人邮箱名即用户名，abc 在 abc.com 这个邮件服务器上必须是唯一的

这也保证了 abc@abc.com 这个邮件地址在整个因特网上是唯一的

首部与主体之间用一个空行进行分割

典型的邮件内容如下：



• 多用途网际邮件扩充 MIME, Multipurpose Internet Mail Extensions

SMTP 只能传送一定长度的 ASCII 码邮件

MIME 增加了邮件主体的结构，并定义了传送非 ASCII 码的编码规则

MIME 邮件可在现有的电子邮件程序和协议下传送

MIME 主要内容：

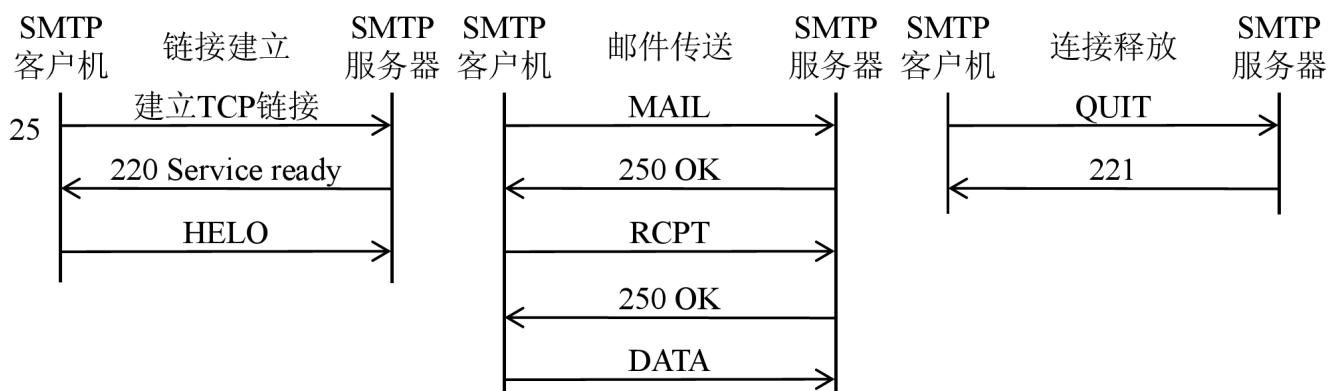
- ①5 个新的邮件首部字段，包括 MIME 版本、内容描述、内容标识、传送编码和内容类型
- ②定义了许多邮件内容的格式，对多媒体电子邮件的表示方法进行了标准化
- ③定义了传送编码，可对任何内容格式进行转换，而不会被邮件系统改变

• 简单邮件传输协议 SMTP, Simple Mail Transfer Protocol

SMTP 是一种提供可靠且有效的电子邮件传输的协议，控制两个相互通信的 SMTP 进程交换信息

SMTP 使用客户/服务器方式，使用 TCP 连接进行数据传输，端口号为 25

SMTP 通信三个阶段：



1) 建立连接

- ①SMTP 客户（发送方邮箱服务器）每隔一定时间对邮件缓存扫描一次
- ②如发现邮件，使用熟知端口号 25 与接收方邮件服务器的 SMTP 服务器建立 TCP 连接

- ③连接建立后，接收方 SMTP 服务器发出 220 Service ready（服务就绪）
- ④然后 SMTP 客户向 SMTP 服务器发送 HELO 命令，附上发送方的主机名
SMTP 不使用中间邮件服务器，TCP 连接在发送方和接收方邮件服务器之间直接建立

2) 邮件传送

- ①连接建立后，就可开始传送邮件
- ②邮件的传送从 MAIL 命令开始，MAIL 命令后面有发件人的地址
如：MAIL FROM: <hoopdog@hust.edu.cn>
- ③若 SMTP 服务器已准备好接收邮件，则回答 250 OK
- ④SMTP 客户端发送一个或多个 RCPT (recipient) 命令，格式为 RCPT TO:<收件人地址>
- ⑤每发送一个 RCPT 命令，都应有相应的信息从 SMTP 服务器返回
如：250 OK 或 550 No such user here（无此用户）

RCPT 命令的作用：先确定接收方系统是否做好接收邮件的准备，然后才发送邮件

- ⑥获得 OK 的回答后，客户端就使用 DATA 命令，表示要开始传输邮件的内容

通常，SMTP 服务器回复的信息是 354 Start mail input; end with <CRLF>.<CRLF>

此时 SMTP 客户端就可开始传送邮件内容，并用<CRLF>.<CRLF>表示邮件内容的结束其中，<CRLF>表示回车换行

3) 释放连接

- ①邮件发送完毕后，SMTP 客户应发送 QUIT 命令
- ②SMTP 服务器返回的信息是 221（服务关闭），表示 SMTP 同意释放 TCP 连接
- ③邮件传送的全部过程即结束

SMTP 缺点：

- 1) SMTP 不能传送可执行文件或其他的二进制对象
- 2) SMTP 限于传送 7 位的 ASCII 码。许多其他非英语国家的文字，如中文就无法传送
- 3) SMTP 服务器会拒绝超过一定长度的邮件

• 邮局协议 POP，Post Office Protocol

邮局协议 POP 是一个非常简单但功能有限的邮件读取协议，现在使用的是它的第 3 个版本 POP3
当用户读取邮件时，用户代理向邮件服务器发出请求，“拉”取用户邮箱中的邮件

POP 也使用客户/服务器的工作方式，在传输层使用 TCP，端口号为 110

接收方的用户代理上必须运行 POP 客户程序，接收方的邮件服务器上则运行 POP 服务器程序
POP 有两种工作方式：

下载并保留方式

用户从邮件服务器上读取邮件后，邮件仍保存在服务器上，用户可再次读取该邮件

下载并删除方式

邮件一旦被读取，就被从邮件服务器上删除，用户不能再次从服务器上读取

• 因特网报文存取协议 IMAP

IMAP 提供了创建文件夹、在不同文件夹之间移动邮件及在远程文件夹中查询邮件等联机命令

IMAP 服务器维护会话用户的状态信息，且允许用户代理只获取报文的某些部分

*POP3 传输密码 P265T7

POP3 协议在传输层是使用明文来传输密码的，并不对密码进行加密

6.5 万维网 WWW，World Wide Web

• WWW 的概念与组成结构

万维网（World Wide Web，WWW）是一个分布式、联机式的信息存储空间

WWW 的构成：

- 1) 统一资源定位符 URL。负责标识万维网上的各种文档，并使每个文档在整个万维网的范围内具有唯一的标识符 URL
- 2) 超文本传输协议 HTTP。一个应用层协议，它使用 TCP 连接进行可靠的传输，HTTP 是万

维网客户程序和服务器程序之间交互所必须严格遵守的协议

3) 超文本标记语言 HTML。一种文档结构的标记语言，它使用一些约定的标记对页面上的各种信息（包括文字、声音、图像、视频等）、格式进行描述

URL 是对可以从因特网上得到的资源的位置和访问方法的一种简洁表示

URL 相当于一个文件名在网络范围的扩展

URL 的一般形式是：

<协议>://<主机>:<端口>/<路径>。

<协议>指用什么协议来获取万维网文档，常见的协议有 http、ftp 等

<主机>是存放资源的主机在因特网中的域名或 IP 地址

<端口>和<路径>有时可省略。在 URL 中不区分大小写

万维网以客户/服务器方式工作，使用端口 80

浏览器是在用户主机上的客户程序，万维网文档所在的主机运行服务器程序

客户程序向服务器程序发出请求，服务器程序向客户程序送回客户所要的万维网文档

工作流程如下：

1) Web 用户使用浏览器（指定 URL）与 Web 服务器建立连接，并发送浏览请求

2) Web 服务器把 URL 转换为文件路径，并返回信息给 Web 浏览器

3) 通信完成，关闭连接

• 超文本传输协议 HTTP

HTTP 是面向事务的（Transaction Oriented）应用层协议，采用客户/服务器协议

它是万维网上能够可靠地交换文件（包括文本、声音、图像等各种多媒体文件）的重要基础

HTTP1.0 协议是无状态的，简化服务器设计，使服务器更容易支持大量并发的 HTTP 请求

使用 cookie 与数据库相结合的方式来跟踪用户的活动

cookie 是储存在用户主机中的文本文件，由服务器产生，作为识别用户的手段

HTTP 协议使用面向连接的 TCP 向上提供的服务

HTTP 有两类报文：

请求报文，从 Web 客户端向 Web 服务器发送服务请求

响应报文，从 Web 服务器对 Web 客户端请求的回答

浏览器浏览网页工作过程（以清华大学网站为例）：

1) 浏览器分析超链接指向页面的 URL

2) 浏览器向 DNS 请求解析清华大学官网的 IP 地址

3) 域名系统 DNS 解析出清华大学服务器的 IP 地址

4) 浏览器与服务器建立 TCP 连接（端口号 80）

5) 浏览器发出 HTTP 请求：GET/chn/yxsz/index.htm

6) 服务器通过 HTTP 响应，把文件 index.htm 发给浏览器

7) TCP 连接释放

8) 浏览器显示“清华大学院系设置”文件 index.htm 中的所有文本

HTTP 协议两种工作方式：

非持久连接：每个连接处理一个请求-响应事务，即当请求-响应过程完成后立即断开连接，下次再次进行传输时，需要重新建立连接

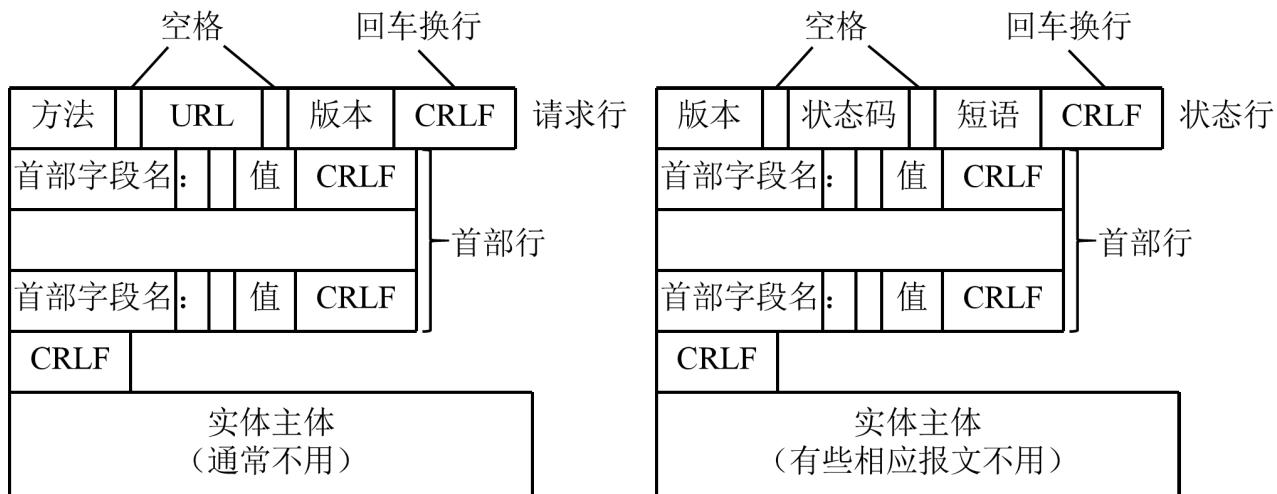
持久连接：HTTP/1.1 协议使用持续连接。万维网服务器在发送响应后仍然在一段时间内保持这条连接，使同一个客户（浏览器）和该服务器可以继续在这条连接上传送后续的 HTTP 请求报文和响应报文

持久连接又分为两种方式：

非流水线方式：客户在收到前一个响应后才能发出下一个请求。虽然这比非持续连接的两倍 RTT 的开销节省了建立 TCP 连接所需的一个 RTT 时间，但服务器在发送完一个对象后，其 TCP 连接就处于空闲状态，浪费了服务器资源

流水线方式：客户在收到 HTTP 的响应报文之前就能够接着发送新的请求报文。连续的请求报文到达服务器后，服务器就可连续发回响应报文。使用流水线方式时，客户访问所有的对

象只需花费一个 RTT 时间，使 TCP 连接中的空闲时间减少，提高了下载文档效率
HTTP 的报文结构：



HTTP 是面向文本的 Text-Oriented

报文中的每个字段都是一些 ASCII 码串，且每个字段的长度都是不确定的

HTTP 请求报文和响应报文都由开始行、首部行、实体主体三个部分

这两种报文格式的区别就是开始行不同

开始行：用于区分是请求报文还是响应报文。

在请求报文中的开始行称为请求行，而在响应报文中的开始行称为状态行

开始行的三个字段之间都以空格分隔，最后的“CR”和“LF”代表“回车”和“换行”

请求报文的请求行有三个内容：方法、请求资源的 URL 及 HTTP 的版本

方法是对所请求对象进行的操作，这些方法实际上也就是一些命令

首部行：用来说明浏览器、服务器或报文主体的一些信息。首部可以有几行，也可以不使用

每个首部行中都有首部字段名和它的值，每一行在结束的地方都要有“回车”和“换行”

整个首部行结束时，还有一空行将首部行和后面的实体主体分开

实体主体：在请求报文中一般不用这个字段，而在响应报文中也可能没有这个字段

*HTTP 1.0 P273 T6

只支持非持久连接，每请求一个对象需要建立一次 TCP 连接

*HTTP 请求报文中的 Connection 和 Cookie P273 T12

Connection：连接方式，Close 表明为非持续连接方式，keep-alive 表示持续连接方式

Cookie 值由服务器产生，HTTP 请求报文中有 Cookie 报头表示曾经访问过服务器