第一章 网络概述

网络向用户提供的重要功能：
通信，资源共享。

因特网从工作方式上可以分为两大块：

边缘部分：

连接在计算中的主机组成，用户直接使用计算机进行通信（浏览网页，看视频）和资源共享。

计算机通信实际上指的是计算机进程间的通信。

进程间的通信可以分为两大类：

客户服务器方式：Client/Server （c/s）

Client 服务的请求方，Server 服务的提供方。

对等方式：Peer-to-peer(p2p)

平等的对等的连接通信。

核心部分：

大量网络和连接这些网络的路由器组成。它为边缘部分的主机提供连通性，使边缘部分的任何主机能与其他主机通信。

分组交换：发送端先将较长的报文划分为较短的、固定长度的数据段，在每个数据段前面添加首部构成分组，依次把分组发送到接收端，接收端把接收到的分组剥去首部还原成报文，最后把收到的数据恢复成原来的报文。

从通信资源的分配角度来看，“交换”就是按照某种方式动态地分配传输线路的资源。

路由器（route）对分组(根据目的主机所连接的网络号)进行存储转发，最后把分组交付目的主机。

路由器处理分组的过程：

把收到的分组放入缓存（暂时存储）。

查找转发表，找出到某个目的地址应该从哪个端口转发。

把分组送到适当的端口转发出去。

分组交换的优点：

高效　　动态分配传输带宽，对通信链路是逐段占用。

灵活　　以分组为传送单位和查找路由。

迅速　　不必先建立连接就能向其他主机发送分组。

可靠　　保证可靠性的网络协议；分布式的路由选择协议使网络有很好的生存性。

计算机网络性能指标

带宽（bandwidth）：数字信道所能传送的最高数据率（或速率）。单位是“比特每秒”或b/s(bit/s)。

吞吐量（throughput）：单位时间内通过某个网络（或信道、接口）的数据量。

传输时延（发送时延）：发送数据时，数据块从结点进入到传输媒体所需要的时间。即从发送数据帧的第一个比特算起，到该帧的最后一个比特发送完毕所需的时间。

$$\text{发送时延}=\frac{\text{数据块长度（比特）}}{\text{信道带宽（比特/秒）}}$$

传播时延：电磁波在信道中需要传播一定的距离而花费的时间。

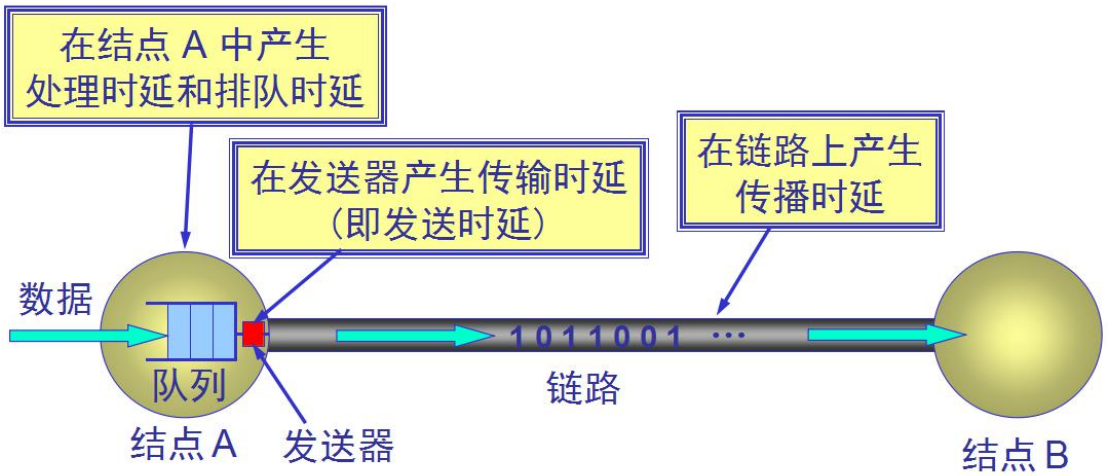
$$\text{传播时延}=\frac{\text{信道长度（米）}}{\text{信号在信道上的传播速率（米/秒）}}$$

处理时延：交换结点为存储转发而进行一些必要的处理所花费的时间。

排队时延：结点缓存队列中分组排队所经历的时延。排队时延的长短往往取决于网络中当时的通信量。

总时延 = 发送时延+传播时延+处理时延+处理时延

借鉴课件中的示意图来描述四种时延产生的地方：从节点 A 向节点 B 发送数据



理解带宽、传输时延（或发送时延）：

对于高速网络链路，我们提高的仅仅是数据的发送速率而不是比特在链路上的传播速率。提高链路带宽减小了数据的发送时延。

信道利用率：指出某信道有百分之几的时间是被利用的（有数据通过）。完全空闲的信道的利用率是零。网络利用率则是全网络的信道利用率的加权平均值。

时延与网络利用率的关系

根据排队论的理论，当某信道的利用率增大时，该信道引起的时延也就迅速增加。

若令 D0 表示网络空闲时的时延，D 表示网络当前的时延，则在适当的假定条件下，可以用下面的简单公式表示 D 和 D0 之间的关系：

$$D=\frac{D_0}{1-U}$$

U 是网络的利用率，数值在 0 到 1 之间。

计算机网络体系结构

计算机网络的体系结构（architecture）是计算机网络各层及其协议的集合。
五层协议结构：（tcp/ip 与 iso 体系整合）

应用层
传输层
网络层
数据链路层
物理层

电信号（或光信号）在物理媒体中传播从发送端物理层传送到接收端物理层。

实体、协议、服务和服务访问点

实体：任何可以发送或接收信息的硬件或软件进程。
协议：协议是“水平的”，它是控制两个对等实体进行通信的规则的集合。在协议的控制下，两个对等实体间的通信使得本层能够向上一层提供服务。要实现本层协议，还需要使用下层所提供的服务。
服务是“垂直的”，服务从下层向上层通过层间接口提供。
同一系统相邻两层的实体进行交互的地方，称为服务访问点 SAP (Service Access Point)。

第二章 物理层

预备知识：通信原理，模拟电子技术，计算机组成原理，数字电子技术，计算机接口技术。
物理层确定与传输媒体的接口的一些特性：机械特性、电气特性、功能特性和规程特性。

物理层考虑的是**怎样才能在连接各种计算机的传输媒体上传输数据比特流（制定媒体接口规范）**，而不是指具体的传输媒体。

物理层要形成适合数据传输需要的实体，为数据传送服务。保证数据能正确通过并且提供足够的带宽，减少信道上的拥塞。一次完整的数据传输，包括激活物理连接，传送数据，终止物理连接。

基带信号与调制

基带信号表示来自信源的信号。计算机输出的代表各种文字和图像的数据信号都属于基带信号。

基带信号往往包含有较多的低频成分，甚至有直流成分，而许多信道并不能传输这种低频分量或直流分量。因此必须对基带信号**调制**形成带通信号以便在信道中传输（即仅在一定频率范围内能通过信道）。

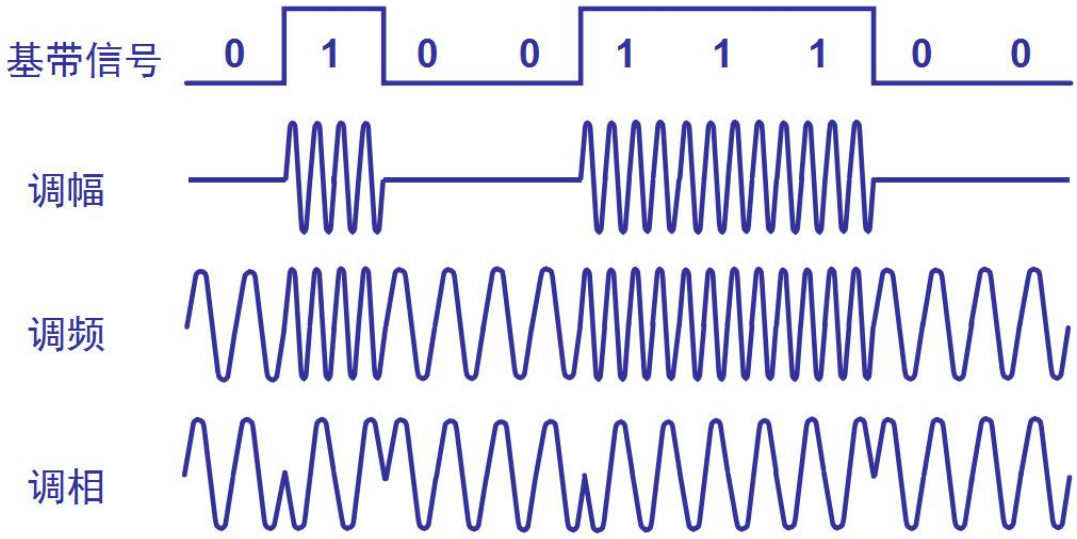
最基本的二元制调制方法有：

调幅(AM)：载波的振幅随基带数字信号而变化。

调频(FM)：载波的频率随基带数字信号而变化。

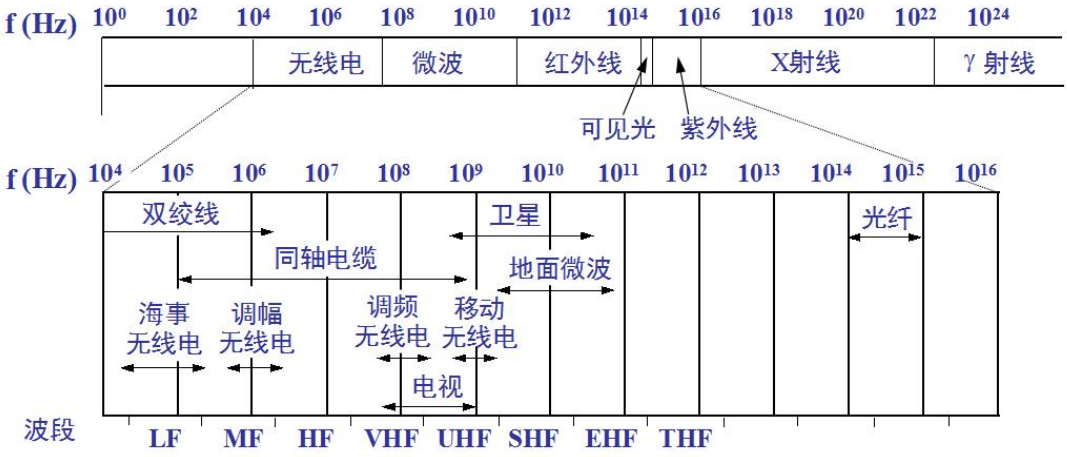
调相(PM)：载波的初始相位随基带数字信号而变化。

对基带信号（计算机输出的数据信号）的几种调制方法：



信道能通过的频率范围

信道带宽限定了允许通过该信道的信号下限频率和上限频率，即限定了一个频率通带。
电信领域使用的电磁波的频谱（在无线电-紫外线之间）：



码元(code)——在使用时间域（或简称为时域）的波形表示数字信号时，代表不同离散数值的基本波形。用编码的方法让每一个码元携带更多比特的信息量来提高信息传输速率。

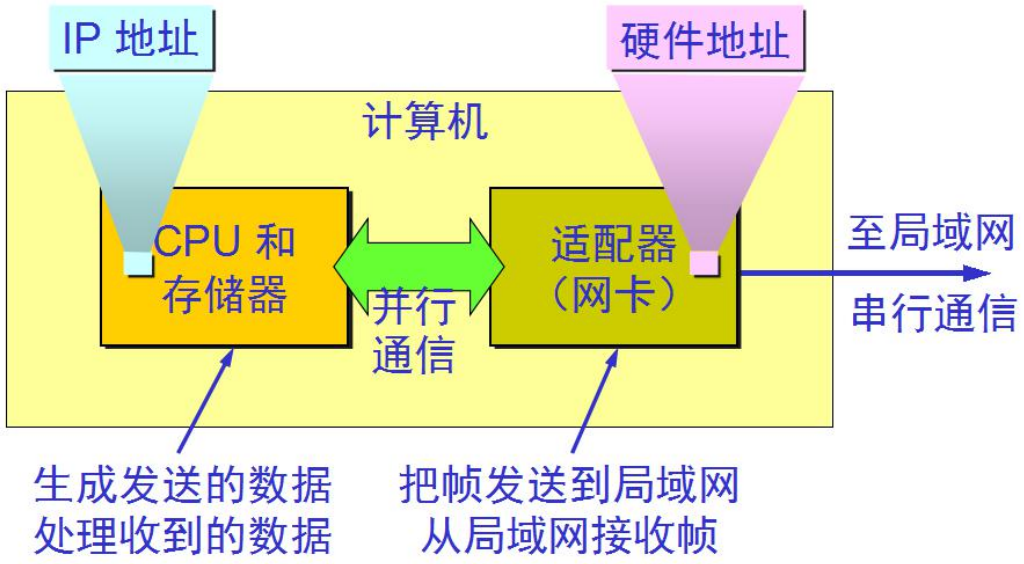
信道复用：一种将若干个彼此独立的信号，合并为一个可在同一信道上同时传输的复合信号的方法。比如，传输的语音信号的频谱一般在 300~3400Hz 内，为了使若干个这种信号能在同一信道上传输，可以把它们的频谱调制到不同的频段，合并在一起而不致相互影响，并能在接收端彼此分离开来。

关于物理层其他主题：信道、信道编码、信号传输参考通信原理相关书籍。

第三章 数据链路层

预备知识：计算机硬件，微机原理，计算机组成原理，通信原理，操作系统原理
关于本节中某些具体知识请参考其他网络相关书籍。
驱动程序：全称“设备驱动程序”，是计算机中央处理器（cpu）控制和使用设备的特殊程序，是硬件接口，操作系统通过这个接口来控制设备工作。
网卡驱动就是计算机 cpu 控制和使用网卡的程序。（IO 总线控制）

参考：网卡的组成工作原理 http://blog.csdn.net/evenness/article/details/7751992
网卡（网络接口卡）充当计算机和网络缆线之间的物理接口或连线将计算机中的数字信号转换成电或光信号,称为 nic（ network interface card ），所以网卡工作在物理层和数据链路层。网卡的核心是链路控制器，控制器包含特定芯片，它们提供成帧，链路接入，差错控制等服务。网卡的重要功能：进行串行/并行转换，对数据进行缓存，实现以太网协议。例如我电脑的网卡驱动：Realtek PCIe GBE Family Controller 是指 Realtek 公司 PCIe 接口千兆以太网系列控制器。
发送端，网卡控制器取得了由 tcp/ip 协议栈较高层（如运输层 tcp）生成并存储在主机内存中的数据报，在链路层帧中封装该数据报（添加首部），然后遵循链路接入协议将该帧传入通信链路中。接收端，控制器接收整个帧，提取出网络层数据报。



两种链路

链路层有两种类型的链路层信道(还记得信道吗？信号通道，个人看法...):
1、广播信道，通常用在局域网（local area network，LAN），无线 LAN，卫星网和混合光缆(HFC)接入网中。对于广播信道，许多主机连接到相同通信信道，需要媒体控制协议来协调传输和避免“碰撞”。
2、点对点通信链路，它要解决一些重要问题：成帧、可靠数据（透明）传输、差错检测和流量控制等。

链路层协议定义了在**链路**两端**结点**数据交互的**分组格式**，以及当发送和接受分组时这些结点采取的**动作（透明传输，差错检测和纠错技术等）**。

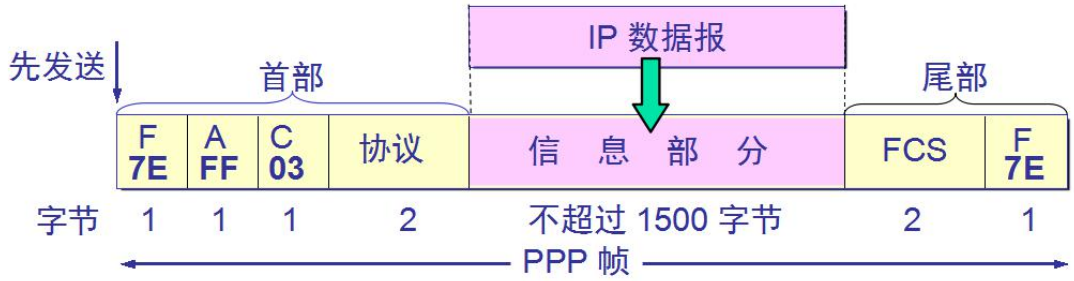
点对点协议 PPP (Point-to-Point Protocol)

使用最广泛的数据链路层(点对点通信链路)协议。用户使用拨号电话线接入因特网时，一般都是使用 PPP 协议。PPP 协议已经成为因特网正式标准[RFC 1661]。

PPP 协议的组成

（1）一个将 ip 数据报封装到串行链路的方法。
（2）一个用来建立、配置和测试数据链路协议的链路控制协议 LCP（link control protocol）。
（3）一套网络控制协议 NCP，每一个协议支持不同的网络层协议。

PPP 协议帧格式



PPP 是面向字节的，所有的 PPP 帧的长度都是整数字节。
首部：
标志字段 F = 0x7E（二进制 01111110）。
地址字段 A 只置为 0xFF。
控制字段 C 通常置为 0x03。
PPP 有一个 2 个字节的协议字段。
当协议字段为 0x0021 时，PPP 帧的信息字段就是 IP 数据报。
若为 0xC021，则信息字段是 PPP 链路控制数据。

若为 0x8021，则表示这是网络控制数据。

信息部分：实现透明传输，当信息部分出现首部的某些字段，接受方可能会错误的检测到 PPP 帧结束，因此 PPP 采用一种**字节填充**的技术来解决这个问题。

PPP 协议的工作状态

1、当用户拨号接入 ISP 时，路由器的调制解调器对拨号做出确认，并建立一条物理连接。PC 机向路由器发送一系列的 LCP 分组（封装成多个 PPP 帧）。

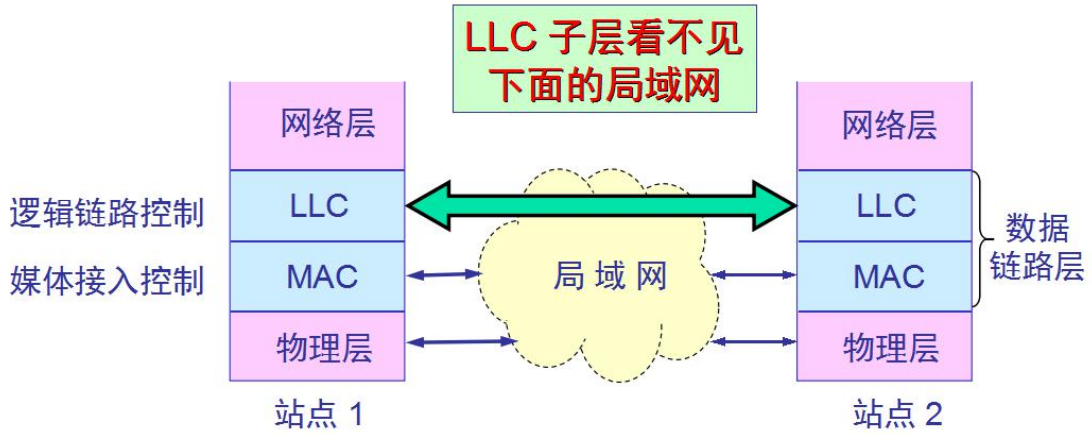
2、这些分组及其响应选择一些 PPP 参数，和进行网络层配置，NCP 给新接入的 PC 机分配一个临时的 IP 地址，使 PC 机成为因特网上的一个主机。

3、通信完毕时，NCP 释放网络层连接，收回原来分配出去的 IP 地址。接着，LCP 释放数据链路层连接。最后释放的是物理层的连接。

以太网--最流行的有线局域网技术

数据链路层分层

为了使数据链路层能更好地适应多种**局域网标准**，802 委员会就将局域网的数据链路层拆成两个子层：
逻辑链路控制 LLC (Logical Link Control)子层
媒体接入控制 MAC (Medium Access Control)子层。
与接入到传输媒体有关的内容都放在 MAC 子层，而 LLC 子层则与传输媒体无关，不管采用何种协议的局域网对 LLC 子层来说都是透明的。
由于 TCP/IP 体系经常使用的局域网是 DIX Ethernet V2 而不是 802.3 标准中的几种局域网，因此现在 802 委员会制定的逻辑链路控制子层 LLC（即 802.2 标准）的作用已经不大了。很多厂商生产的适配器上就仅装有 MAC 协议而没有 LLC 协议。



关于 LLC 层参考其他书籍。以下主要讨论 MAC 层

三个重要的问题：
CSMA/CD、以太网 MAC 子层、链路层交换机（数据链路层扩展局域网）

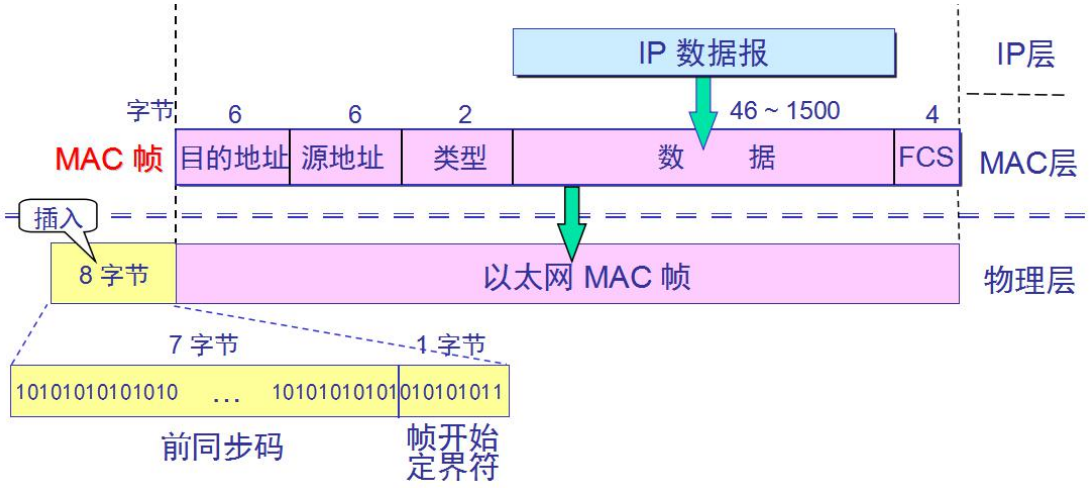
CSMA/CD

多点接入、载波监听、碰撞检测
以下用适配器代替网卡描述
传输时进行承载的信号的监听（**信道是否空闲**）、信号碰撞的检测（**检测是否有其他适配器发出的信号能量**）。
载波监听和碰撞检测都是形象的说法，以太网适配器通过传输前和传输中的电压水平来执行这两项任务。

以太网 MAC 子层

常用的以太网 MAC 帧格式有两种标准 ：
DIX Ethernet V2 标准
IEEE 的 802.3 标准
最常用的 MAC 帧是以太网 V2 的格式。以下简称 MAC 帧

以太网 MAC 帧格式



帧格式说明：
在帧的前面插入的 8 字节中的第一个字段共 7 个字节，是前同步码，用来迅速实现 MAC 帧的比特同步。
第二个字段是帧开始定界符，表示后面的信息就是 MAC 帧。 (为了达到比特同步，在传输媒体上实际传送的要比 MAC 帧还多 8 个字节。)帧间最小间隔为 9.6us，相当于 96 bit 的发送时间。一个站在检测到总线开始空闲后，还要等待 0.0096ms 才能再次发送数据。这样做是为了使刚刚收到数据帧的站的接收缓存来得及清理，做好接收下一帧的准备。
类型字段（2 字节）用来标志上一层使用的是什么协议，以便把收到的 MAC 帧的数据上交给上一层的这个协议。
数据字段的正式名称是 MAC 客户数据字段。
最小长度 64 字节 - 18 字节的首部和尾部 = 数据字段的最小长度

当数据字段的长度小于 46 字节时，应在数据字段的后面加入整数字节的填充字段，以保证以太网的 MAC 帧长不小于 64 字节。

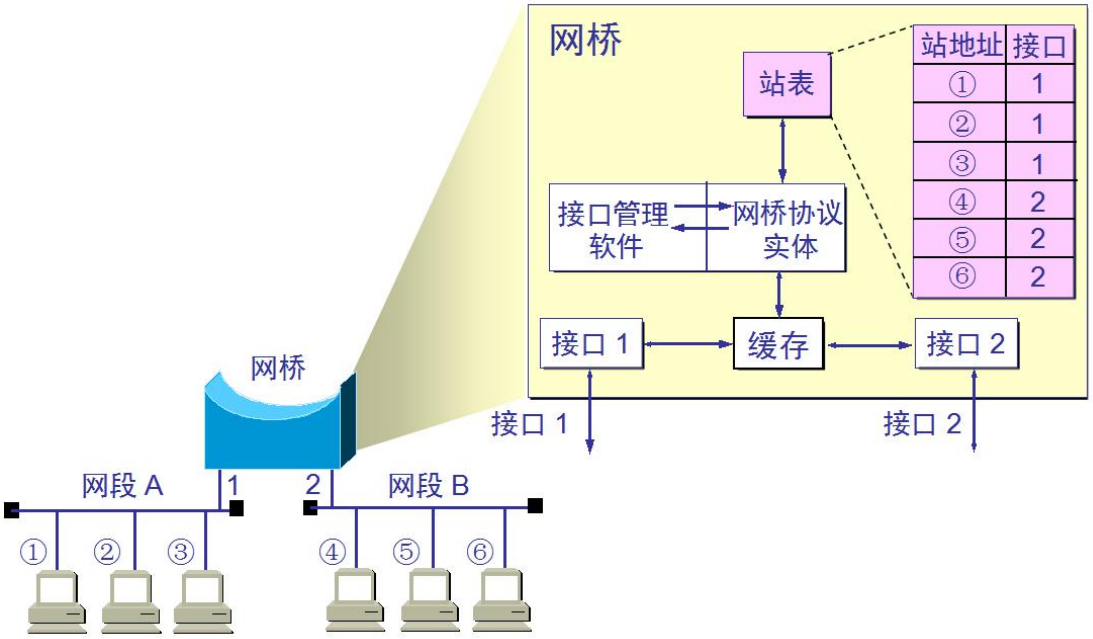
无效的 MAC 帧
数据字段的长度与长度字段的值不一致；
帧的长度不是整数个字节；
用收到的帧检验序列 FCS 查出有差错；
数据字段的长度不在 46～1500 字节之间。
有效的 MAC 帧长度为 64～1518 字节之间。
对于检查出的无效 MAC 帧就简单地丢弃。以太网不负责重传丢弃的帧。

链路层交换机--网桥

网桥工作在数据链路层，它根据 MAC 帧的目的地址对收到的帧进行转发。网桥具有过滤帧的功能。当网桥收到一个帧时，并不是向所有的接口转发此帧，而是先检查此帧的目的 MAC 地址，然后再确定将该帧转发到哪一个接口。

网桥的内部结构：

接口管理软件、协议实体、缓存、站表。



网桥自学习和转发帧的一般步骤

网桥收到一帧后先进行自学习。查找转发表中与收到帧的源地址有无相匹配的项目。如没有，就在转发表中增加一个项目（源地址、进入的接口和时间）。如有，则把原有的项目进行更新。

转发帧。查找转发表中与收到帧的目的地址有无相匹配的项目。

如没有，则通过所有其他接口（但进入网桥的接口除外）进行转发。

如有，则按转发表中给出的接口进行转发。

若转发表中给出的接口就是该帧进入网桥的接口，则应丢弃这个帧（因为这时不需要经过网桥进行转发）。

两个网桥之间还可使用一段点到点链路。

集线器与交换机比较

集线器工作在物理层在转发帧时，不对传输媒体进行检测。

网桥在转发帧之前必须执行 CSMA/CD 算法。若在发送过程中出现碰撞，就必须停止发送和进行退避。

第四章 网络层

网络层向上只提供简单灵活的、无连接的、尽最大努力交付的**数据报服务**。

网络服务--虚电路服务与数据报服务

网络在发送分组时不需要先建立连接。每一个分组（即 IP 数据报）独立发送，与其前后的分组无关（不进行编号）。网络层不提供服务质量的承诺。即所传送的分组可能出错、丢失、重复和失序（不按序到达终点），当然也不保证分组传送的时限。

电信网采用面向连接的网络通信方式,通信时会会建立虚电路保证双方通信所需的一切网络资源。

虚电路服务与数据报服务的对比

对比的方面	虚电路服务	数据报服务
思路	可靠通信应当由网络来保证	可靠通信应当由用户主机来保证
连接的建立	必须有	不需要
终点地址	仅在连接建立阶段使用，每个分组使用短的虚电路号	每个分组都有终点的完整地址
分组的转发	属于同一条虚电路的分组均按照同一路由进行转发	每个分组独立选择路由进行转发
当结点出故障时	所有通过出故障的结点的虚电路均不能工作	出故障的结点可能会丢失分组，一些路由可能会发生变化
分组的顺序	总是按发送顺序到达终点	到达终点时不一定按发送顺序
端到端的差错处理和流量控制	可以由网络负责，也可以由用户主机负责	由用户主机负责

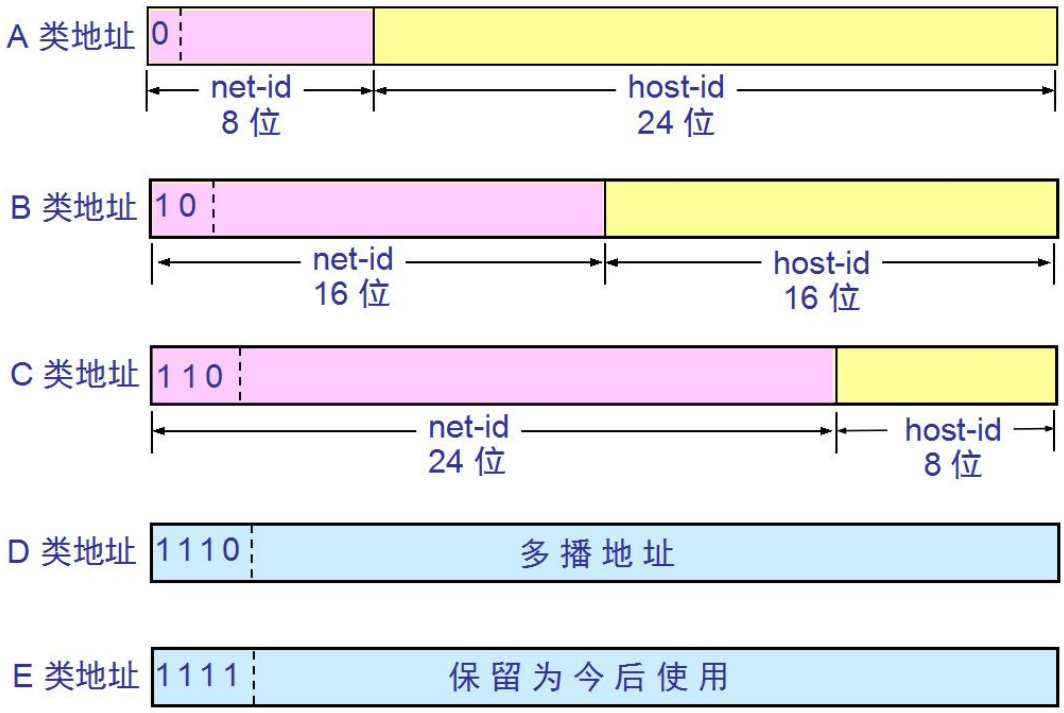
IP 分类编址

IP 地址就是给每个连接在因特网上的主机（或路由器）分配一个在全世界范围是唯一的 32 位的标识符。IP 地址现在由因特网名字与号码指派公司 ICANN (Internet Corporation for Assigned Names and Numbers)进行分配。

IP 地址中网络号与主机号字段

IP 地址 ::= {<网络号>, <主机号>}

IP 地址分类



IP 地址的使用范围

网络类别	最大网络数	第一个可用的网络号	最后一个可用的网络号	每个网络中最大的主机数
A	126 ($2^7 - 2$)	1	126	16,777,214
B	16,383($2^{14} - 1$)	128.1	191.255	65,534
C	2,097,151 ($2^{21} - 1$)	192.0.1	223.255.255	254

IP 数据报格式

首部的固定部分共 20 字节。

IP 数据报分组转发

(1) 从数据报的首部提取目的主机的 IP 地址 D，得出目的网络地址为 N。
(2) 若网络 N 与此路由器直接相连，则把数据报直接交付目的主机 D；否则是间接交付，执行(3)。
(3) 若路由表中有目的地址为 D 的特定主机路由，则把数据报传送给路由表中所指明的下一跳路由器；否则，执行(4)。
(4) 若路由表中有到达网络 N 的路由，则把数据报传送给路由表指明的下一跳路由器；否则，执行(5)。
(5) 若路由表中有一个默认路由，则把数据报传送给路由表中所指明的默认路由器；否则，执行(6)。
(6) 报告转发分组出错。

tips：当路由器收到待转发的数据报，不是将下一跳路由器的 IP 地址填入 IP 数据报，而是送交下层的网络接口软件。网络接口软件使用 ARP 负责将下一跳路由器的 IP 地址转换成硬件地址，并将此硬件地址放在链路层的 MAC 帧的首部，然后根据这个硬件地址找到下一跳路由器。

子网掩码--划分子网

当没有划分子网时，IP 地址是两级结构。
划分子网后 IP 地址就变成了三级结构。

IP 地址 ::= {<网络号>, <子网号>, <主机号>}

划分子网只是把 IP 地址的主机号 host-id 这部分进行再划分，而不改变 IP 地址原来的网络号 net-id。

使用子网掩码的分组转发过程

(1) 从收到的分组的首部提取目的 IP 地址 D。
(2) 先用各网络的子网掩码和 D 逐位相“与”，看是否和相应的网络地址匹配。若匹配则将分组直接交付。否则就是间接交付，执行(3)。
(3) 若路由表中有目的地址为 D 的特定主机路由，则将分组传送给指明的下一跳路由器；否则，执行(4)。
(4) 对路由表中的每一行的子网掩码和 D 逐位相“与”，若其结果与该行的目的网络地址匹配，则将分组传送给该行指明的下一跳路由器；否则，执行(5)。
(5) 若路由表中有一个默认路由，则将分组传送给路由表中所指明的默认路由器；否则，执行(6)。
(6) 报告转发分组出错。

无分类编址--构成超网

为了解决 ipv4 地址空间耗尽问题，研究出一种无分类编址方法：无分类域间路由选择 CIDR (Classless Inter-Domain Routing)。CIDR 消除了 ip 地址分类以及划分子网的概念。

CIDR 记法：IP 地址 ::= {<网络前缀>, <主机号>}

例如：128.14.32.0/20 表示的地址块共有 2 的 12 次方 个地址（因为斜线后面的 20 是网络前缀的位数，也是掩码的 1 的位数，所以这个地址的主机号是 12 位）。

路由器

这里只做简单描述，具体内容参考其他书籍。
路由器是工作在网络层，可以连接不同类型的网络，能够选择数据传送路径并对数据进行转发的网络设备。
一个通用的路由器体系结构由 4 个部分组成：
输入端口、交换结构、选路处理器、输出端口
输入端口：数据链路处理（协议、拆封），查找、转发、排队。
交换结构：分组从输入端口转发到输出端口。交换可以有很多种方式：内存交换、经一根总线交换、经一个互联网络交换。

输出端口：分组排队（缓存）管理、数据链路处理。
选路处理器：在路由器中运行、交换和计算，以配置转发表信息。

选路算法

选路算法的目的很简单：给定一组路由以及连接路由器的链路，选路算法要找到一条从源路由器到目的路由器的相对最佳的路径。

关于网络层其他协议：OSPF、BGP、ICMP、IGMP 等参考其他书籍。

第五章 运输层

从通信和信息处理的角度看，运输层向它上面的应用层提供通信服务，向下公用网络层提供的服务，它属于面向通信部分的最高层，同时也是用户功能中的最低层。
当网络的边缘部分中的两个主机使用网络的核心部分的功能进行端到端的通信时，只有位于网络边缘部分的主机的协议栈才有运输层，而网络核心部分中的路由器在转发分组时都只用到下三层的功能。
运输层向高层用户屏蔽了下面网络核心的细节（如网络拓扑、所采用的路由选择协议等），它使应用进程看见的就是好像在两个运输层实体之间有一条端到端的逻辑通信信道。
运输层提供进程间通信，网络层提供主机间通信。

硬件端口与软件端口

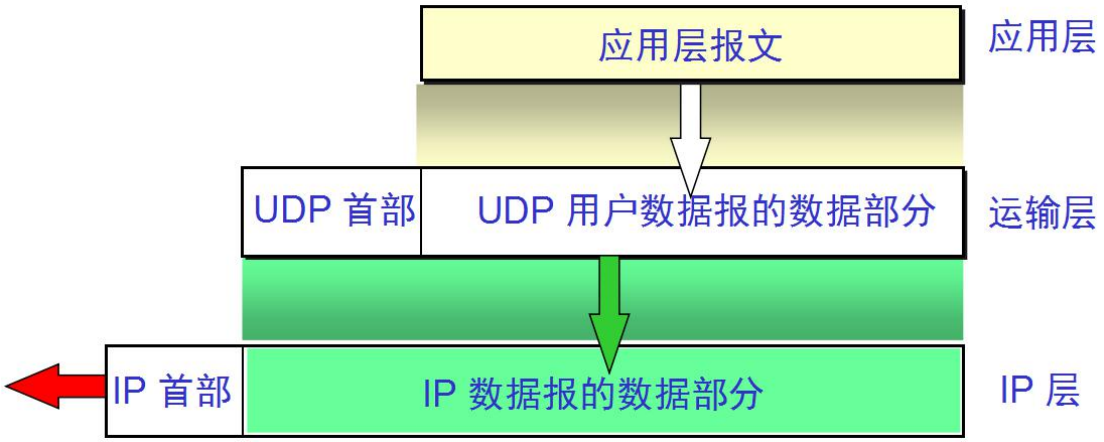
在协议栈层间的抽象的协议端口是软件端口。
路由器或交换机上的端口是硬件端口。
硬件端口是不同硬件设备进行交互的接口，而软件端口是应用层的各种协议进程与运输实体进行层间交互的一种地址。

可靠传输与不可靠传输

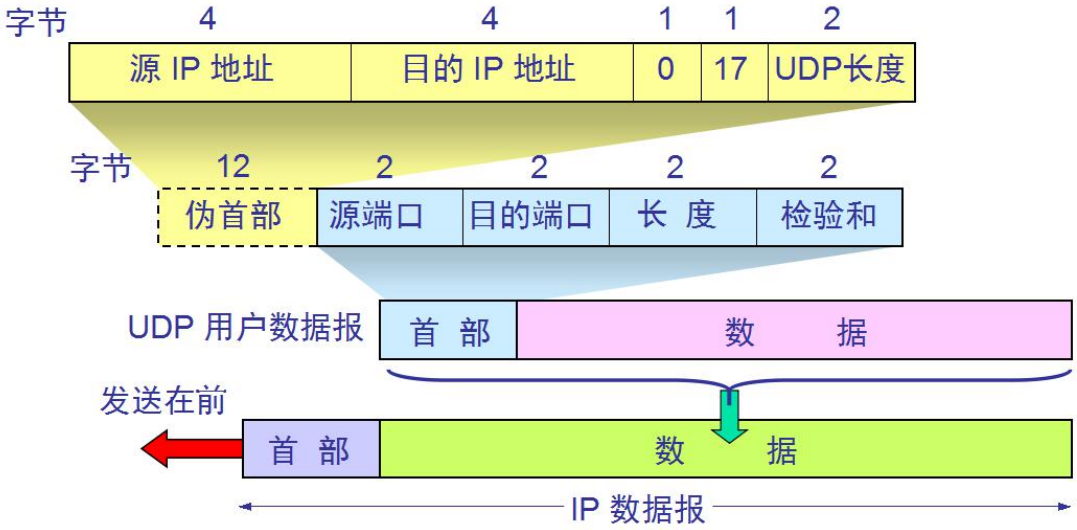
UDP 在传送数据之前不需要先建立连接。对方的运输层在收到 UDP 报文后，不需要给出任何确认。虽然 UDP 不提供可靠交付，但在某些情况下 UDP 是一种最有效的工作方式。
TCP 则提供面向连接的服务。TCP 不提供广播或多播服务。由于 TCP 要提供可靠的、面向连接的运输服务，因此不可避免地增加了许多的开销。这不仅使协议数据单元的首部增大很多，还要占用许多的处理机资源。
运输层的 UDP 用户数据报与网际层的 IP 数据报有很大区别。IP 数据报要经过互连网中许多路由器的存储转发，但 UDP 用户数据报是在运输层的端到端抽象的逻辑信道中传送的。
TCP 报文段是在运输层抽象的端到端逻辑信道中传送，这种信道是可靠的全双工信道。但这样的信道却不知道究竟经过了哪些路由器，而这些路由器也根本不知道上面的运输层是否建立了 TCP 连接。

面向报文的 UDP

发送方 UDP 对应用程序交下来的报文，在添加首部后就向下交付 IP 层。UDP 对应用层交下来的报文，既不合并，也不拆分，而是保留这些报文的边界。
应用层交给 UDP 多长的报文，UDP 就照样发送，即一次发送一个报文。
接收方 UDP 对 IP 层交上来的 UDP 用户数据报，在去除首部后就原封不动地交付上层的应用进程，一次交付一个完整的报文。
应用程序必须选择合适大小的报文。



UDP 的首部格式

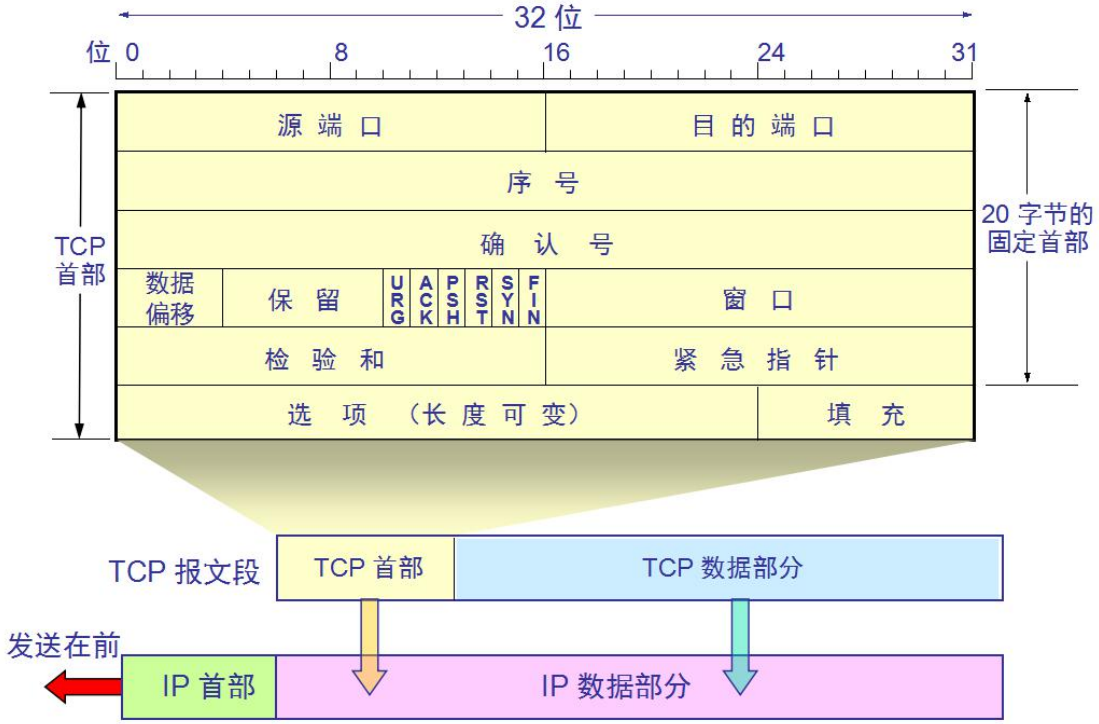


传输控制协议 TCP

可靠传输的实现（确认和重传机制）：自动重传请求 ARQ (Automatic Repeat reQuest)，ARQ 表明重传的请求是自动进行的。接收方不需要请求发送方重传某个出错的分组 。

TCP 可靠传输具体实现

tcp 报文格式



TCP 流量控制