

计算概论笔试理论内容

计算机与信息社会

数字化进程

1. 内容数字化——20世纪80年代中期开始
2. 互联网和企业数字化——20世纪90年代中期开始
3. 信息、物理和生物世界的融合——现在

学术组织

ACM-Association for Computer Machinery

IEEE Computer Society-Institute of Electrical and Electronic Engineers, Computer Society

图灵奖

ACM于1966年设立的第一个奖项，是**计算机界的“诺贝尔奖”**，主要奖励在计算机科学理论和软件方面作出重要贡献的科学家。

华裔美国科学家姚期智(Andrew Chi-chih Yao)由于在计算理论方面做出了诸多“根本性的、意义重大的”贡献，获得2000年度图灵奖。这是首位（也是目前唯一一位）获得图灵奖的华裔科学家。

计算机

计算机是一种现代化的**信息处理工具**。它对信息进行处理并提供所需结果。其结果（输出）取决于所接收的信息（输入）及相应的处理算法。

计算机科学技术

是研究计算机的设计与制造和利用计算机进行**信息获取、表示、储存、处理、控制**等的**理论、原则、方法和技术**的学科。它包括科学与技术两方面。科学侧重研究现象与揭示规律；技术则侧重研制计算机及使用计算机进行信息处理的方法与技术手段。

计算机发展史

“史前”时代

普遍认为的第一台机械计算器：Pascal发明的一个用齿轮运作的加法器，可加八位长的数字

工业革命前的机械计算器：G.W.Leibnintz提出了“**二进制**”的概念，建造了能进行四则运算的计算机，结果可达到十六位

工业革命时代的计算器：Babbage在1822年制造出第一台**差分机**，可处理3个不同的5位数，计算精度达6位小数；在1834年提出了**分析机**的概念

机械计算器的不断发展：H.Hollerith等人创造“**高级分类统计机**”

采用电气元件的“计算机原型”：K.Zuse于1941年完成**全继电器式通用计算机Z3**，使用了大量的真空管

自动数字计算机：The Harvard Mark Series-1，IBM开发建造，1944年完成；后又有Mark-II，III，IV；全部使用继电器

Grace Murray Hopper发现世界上第一个BUG

现代计算机

现代计算机产生于图灵机 (Turing Machine)

第一台投入实用的电子计算机ENIAC (Electronic Numerical Integrator And Computer) 用于计算炮弹的弹道轨迹, 但没有存储器且用布线板进行控制, 控制效率极低

John Vincent Atanasoff在1937年设计出了第一台**电子数字计算机**, 制造了磁鼓来存储数据

John von Neumann(冯·诺伊曼) 确定了**现代存储程序式电子数字计算机**的基本结构和工作原理

- **存储器**
- **运算器**
- **控制器**
- **输入设备**
- **输出设备**

第一台现代意义上的计算机EDVAC (Electronic Discrete Variable Automatic Computer), 采用**二进制**表示数据和程序, 使用“**存储程序**”的方式对计算机进行控制

“101页报告”: 明确奠定了新机器由五个部分组成, 包括: **运算器、逻辑控制装置、存储器、输入和输出设备**, 并描述了这五部分的职能和相互关系。还有两个非常重大的改进, 即: 采用了二进制, 不但**数据**采用二进制, **指令**也采用二进制; 建立了**存储程序**, 指令和数据可一起放在存储器里, 并作同样处理, 简化了计算机的结构, 大大提高了计算机的速度。

现代计算机采用二进制的原因: 计算机容易区分电的on/off或者电压的高、低这类两种状态, 可以与二进制数字“1”和“0”相对应。

ISA电子计算机

但最早问世的内储程序式计算机既不是ISA, 也不是EDVAC, 而是EDSAC。Mauricee为此摘获了1967年的图灵奖。

John von Neumann的贡献

- **程序存储 (解决人工干预, 实现自动执行)**
 - 程序输入到计算机中, 存储在内存存储器中 (存储原理)
 - 在运行时, 控制器按地址顺序取出存放在内存存储器中的指令然后分析指令, 执行指令的功能, 遇到转移指令时, 则转移到转移地址, 再按地址顺序访问指令。
- **计算机组成结构 (两种说法):**
 - **五部分: 运算器、存储器、控制器、输入设备、输出设备**
 - **三部分: CPU (运算器、控制器)、存储器 (内部存储器和外部存储器)、I/O设备 (输入设备、输出设备)**

计算机发展的时代划分

- **第一代计算机(1946年-1957年): 电子管计算机**
 - 计算机的**逻辑元件**采用电子管, **主存储器**采用汞延迟线、磁鼓、磁芯; **外存储器**采用磁带; **软件**主要采用机器语言、汇编语言; 应用以**科学计算**为主。其**特点**是体积大、耗电大、可靠性差、价格昂贵、维修复杂, 但它奠定了以后计算机技术的基础。
 - 中国的第一台计算机DJS-103
- **第二代计算机(1957年-1964年): 晶体管计算机**

- **主存储器**采用磁芯，**外存储器**已开始使用更先进的磁盘；**软件**有了很大发展，出现了各种各样的高级语言及其编译程序，还出现了以批处理为主的**操作系统**，应用以**科学计算和各种事务处理**为主，并开始用于工业控制。
- 产生高级编程语言：Fortran, Cobol
- 晶体管之父：William Shockley
- **第三代**计算机(1964年-1970年)：**集成电路**计算机
 - **主存储器仍采用磁芯**。**软件**逐渐完善，分时操作系统、多种高级语言都有新的发展。
 - 集成电路之父：Jack Kilby
- **第四代**计算机(1970年-今)：**大规模集成电路**计算机
 - 大规模集成电路：单片硅片上集成1000~2000个以上晶体管的集成电路，其集成度比中、小规模集成电路提高了1~2个以上数量级。
- 第五代计算机 (??)：计算机发展的新方向

现代计算机：“通用电子数字计算机” (General -Purpose Electronic Digital Computer)

- 通用：计算机是一种通用信息处理设备，只要有合适的软件，它能适用于各种用途
- 电子：是计算机硬件实现的物理基础，计算机的运行最终都通过电子电路中的电流、电位等实现
- 数字化：是计算机的信息表示方式。一切信息，无论原本是数值、文字、图形、声音等，在计算机里都统一到二进制的数字化表示上。数字化是计算机的一种基本特征，是通用性的重要基础现代计算机

计算机分类

- 微型计算机：PC
 - IBM PC世界上首台个人计算机
- 小型\中型计算机：服务器
- 大型计算机：一般的大型公司使用
- 巨型机：国家科技水平的代表
 - Flops (floating-point operations per second)

摩尔定律 (Moore)

Gordon Moore长期观察发现

指**集成电路IC上可容纳的晶体管数目约每隔18个月便会增加一倍，性能也将提升一倍。**

新摩尔定律：我国Internet联网主机数和上网用户人数的递增速度，大约每半年就翻一番。

麦卡夫定律 (Metcalfe)

一个网络的价值等于该网络内的节点数的平方，而且该网络的价值与联网的用户数的平方成正比。

计算

计算是指运用事先规定的规则，将一组**数值**变换为另一（所需的）**数值**的过程。

数

“数”是量度事物的概念

- 埃及数字
- 汉字数字

- 阿拉伯数字

IPv4和IPv6

IPv4:

- 采用32位(2^{32})二进制数表示IP地址
- 拥有42.95亿个 (2^{32})的地址资源
- IPv4的地址资源面临枯竭
- A、B、C类地址
 - 在分配IP地址时，主机标识不可以全部为0或1，因此C类地址每个网段只有254个主机地址

IPv6:

- 采用128位(2^{128})二进制数表示IP地址
- 拥有天文数字般 (2^{128})的地址资源，可以为全球的每一粒沙子分配一个IP地址
- 从IPv4向IPv6的迁移需要10年或更长的时间

程序设计语言

程序：按照一定顺序排列的一组指令

程序设计语言：

- 进行程序设计所使用的语言
- 规定了人们描述/编制程序所使用的规范和方法

演化历史：机器语言—汇编语言—高级语言

机器语言

直接使用机器指令进行程序设计的语言

计算机指令系统

- 任何一台计算机在硬件层上都有自己的指令系统
- 包含一组可被计算机直接理解和执行的机器指令

缺点：

- 难以阅读和理解
- 使得程序设计工作更加烦琐和容易出错
 - 要检查和发现程序中的错误非常困难
 - 程序的开发和维护效率非常低

汇编语言

基本特点

- 提供了一组具有直观含义的“汇编指令”
- 汇编指令与机器指令之间一一对应
- 操作和操作数采用具有直观含义的符号表示

但计算机不能直接执行，需要汇编器

缺点：

- 抽象层次太低
 - 基本描述单位仍然是计算机指令
 - 与人们描述计算过程所需要的抽象性之间的差距太大
- 与计算机的硬件结构和指令系统联系过于紧密
 - 程序的可移植性差
 - 针对一种指令系统编写的汇编语言程序，不能在另一种具有不同指令系统的计算机上运行

高级语言

基本特点：

- 在对计算过程的描述方面，完全脱离了简单的机器指令方式
 - 采用类似于数学公式的方式描述计算过程
 - 用更高级的流程控制结构取代简单的指令跳转
- 引入了变量的概念

变量：一种存放数据的容器

一个变量对应于存储器中的若干连续字节

每个变量都有一个字符串形式的名字

优点：

- 编程人员不需要关心变量的具体存放位置
- 变量的名字可以帮助人们直观的理解变量所代表的物理含义
- 在程序中，通过变量的名字就可以实现对变量中数据的读和写

表达式：用于描述计算过程的结构，类似于数学中的代数运算公式

语句：高级语言中的基本动作单位，不同类型的语句用于实现不同的功能

高级语言的执行方式：编译方式（一次性）、解释方式（即时）

C语言：

计算机信息表示和输入输出

信息：广义地说，信息就是消息。1948年，信息论创始人香农（C. E. Shannon）从研究通信理论出发，第一次从数学方法角度进行了定义，即“信息就是不确定性的消除”。

数据：指能够被计算机识别、存储和加工处理的信息载体。（数据是客观事物的符号表示。）

数据类型：

- 文本
- 数值
- 图像
- 音频
- 视频

数制

使用进位的方式进行计数的方法，称为进位计数制

- 数码：数制中表示基本数值大小的不同数字符号。例如，十进制的数码是0、1、2、3、4、5、6、7、8、9；二进制的数码是0、1。
- 基数：某数制可以使用的数码个数。例如，十进制的基数是10；二进制的基数是2。
- 数位：数码在一个数中所处的位置。
- 权：权是基数的幂，数制中某一位上的1所表示数值的大小（所处位置的价值）。

八进制：数的前面加0

十六进制：数的前面加0x；还有ABCDEF六个字母

进制转换

十进制到X进制：整数部分除X取余，小数部分乘X取整

X进制到十进制：把非十进制数各个数位上的数按权值展开求和

X进制到Y进制：以十进制为过渡

特殊情况：

2进制到8进制：3位一断

2进制到16进制：4位一断

数字化原理

- 二进制数表示：任何信息在计算机中都以二值的数字形式被存储、被处理，还通过各种通讯媒体被传输和接收。
- 数字化编码：在特定的上下文环境中，对表示某类信息的二值数据赋予意义，形成基本的共识性数据。
- 结构化数据：复杂数据，是各种基本共识性数据的组合，如Word文件、PowerPoint文件等，其中可以包括字符、数字、图像、声音等。

计数单位：

位：bit (Binary Digit) ;最小的信息单位，用0、1来表示一个二进制位

字节：Byte/B；数据存储中最常用的基本单位。由8位二进制数构成

- $1B = 8b$
- $1KB = 1024 B = 2^{10}B$ 近似值 10^3
- $1MB = 1024 KB = 2^{20}B$ 近似值 10^6
- $1GB = 1024 MB = 2^{30}B$ 近似值 10^9
- $1TB = 1024 GB = 2^{40}B$ 近似值 10^{12}
- 在描述距离、频率等数值时通常用10的幂次表示。
- 为了表示这种差别，通常用K表示1024，用k表示1000，其他前缀字母都为大写，表示的大小由其上下文决定。
- 由时钟频率计算得到的总线带宽或外设数据的传输频率中，度量单位通常也是10的幂次。
 - 比特/秒 (b/s)，有时也写为bps

- 千比特/秒 (kb/s) , $1\text{kb/s}=10^3\text{b/s}=1000\text{bps}$
- 兆比特/秒 (Mb/s) , $1\text{Mb/s}=10^6\text{b/s}=1000\text{kbps}$
- 吉比特/秒 (Gb/s) , $1\text{Gb/s}=10^9\text{b/s}=1000\text{Mbps}$
- 太比特/秒 (Tb/s) , $1\text{Tb/s}=10^{12}\text{b/s}=1000\text{Gbps}$

字长: 每个字所包含的二进制数码的位数

计算机信息编码

指令编码

数值表示

- 8位: $0 \sim 255$ ($2^8 - 1$) —— 字节
- 16位: $0 \sim 65535$ ($2^{16} - 1$)
- 32位: $0 \sim 40\text{亿}$ ($2^{32} - 1$) $= 4\,294\,967\,295$
- 64位: 天文数字 (2^{64}) $= 18446744073709551616 = 1.844674407371 \times 10^{19}$

整数的表示

- 有符号数
 - $-2^{15} \sim 2^{15}-1$ (32767)
- 无符号数
 - 常用16位二进制数
 - $0 \sim 65535$ ($2^{16} - 1$)

原码、反码、补码

原码: 符号用一位二进制数表示 (符号位) , 数的绝对值和符号位一起编码

- 将数的绝对值转换为2进制, 忽略符号
- 如果2 进制位数不足N -1, 则左边补0 至N -1 位
- 添加**符号位**: 如果是正数, 则左边补**0**; 若是负数, 则左边加**1**; 变为N 位
- 问题: 有两个0 (+0和-0) ; 且不方便加减

反码: 很少使用, 是求补码的中间过程

- 正数的反码和原码相同
- 负数的反码符号位和原码相同, 其余各位取反
- 问题: 0的表示不唯一 (0000 0000和1111 1111)

补码:

- 正数的补码和原码相同
- 负数的补码符号位和原码相同, 其余各位取反, **在最末位加一**
- 0的表示唯一 (0000 0000)
- 作为符号位的最高位可以和其他位一样参与运算

浮点数的表示

小数点的位置不固定。由阶码和尾数组成（**符号+阶码+尾数**）

- 阶码：指数部分，是一个整数
- 尾数：数的有效数值，整数或纯小数两种形式

并非任何一个十进数都可以转化为**完全相等**的对应的二进制数进行计算（**太大/太小会溢出**）

字符编码

数字化编码：就是用数值对现实生活中的某类对象进行一一标识，使得每个对象都拥有一个数值编号，然后我们可以用这个数值来唯一标识这个对象。

数字化编码的两个关键因素：

- 对象的数量
 - 有限
 - 无穷无尽：离散、采样
- 数值的范围（数值的位数）：与数量相关
 - 多了，浪费
 - 少了，不够，“溢出”

身份证号码的校验码：对前17位数按从右到左顺序，第1位乘2¹，第i位乘2ⁱ，将17个乘积求和，然后除以11得到余数。为保证十八位标准长度，10用X代替。

字符编码：计算机中非数值型数据的编码

- ASCII码（American national Standard Code for Information Interchange）
- Unicode
- 其他编码

中文：用两个字节来编码一个汉字；每个字节的最高位为0

在计算机中，采用汉字内码：把国标码两个字节的最高位变为1（避免和ASCII码发生冲突）

汉字的输入码（外码）：数字编码、拼音编码、字形编码

汉字的字形码（输出码）：矢量字符和点阵字符

Unicode

UTF-8以字节为单位对unicode字符进行编码，对不同范围的字符使用不同长度的编码

- 0x000000~ 0x00007f区间的字符用1字节编码，与标准ASCII码相同
- 0x000080~ 0x0007ff区间的符号用2字节编码，即110xxxxx 10xxxxxx
- 0x000800~ 0x00ffff区间的符号用3字节编码（汉字通常位于此区间），即1110xxxx 10xxxxxx 10xxxxxx
- 其他区间用4字节编码即11110xxx 10xxxxxx 10xxxxxx 10xxxxxx

声音编码

现实生活中的很多现象是连续的、无穷无尽的，而在计算机内部，表达数据的二进制位数是有限的，因此，不可能像字符那样对连续的现象进行穷举，需要有关现象的连续特性进行**离散化采样**，以减少相关现象的数量，从而能够用有限位数的二进制数来一一描述它们。

所谓**离散化采样**就是把连续现象划分为离散的区段，把**每一个区段内这些相近的现象看作一种现象**，并用一个特定的二进制数来表示它。

声波的离散化采样是在**时间**和**波形高度**这两个维度上独立进行的。

采样频率：**单位时间对声音波形的采样次数**，以Hz为单位

一般音乐CD的采样频率为44kHz。为了获得较好的声音效果，往往采用**双声道**，**每声道用2字节 (16 bit)** 来记录波形高度值。

例：一首4分钟长的歌曲需要的空间为 $44 \times 1000 \times 4 \times 60 \times (16/8) \times 2 = 42240 \times 10^3$

颜色编码

黑白图像：**1位模式**，**0表示黑，1表示白**

- 16级灰度，需要4 bit (2^{**4})
- 256级灰度，需要8 bit (2^{**8})

彩色图像：**RGB**，测出每种颜色的强度，并把8位模式赋给它

彩色系统：**彩色系统**

- **16色(4位)**
- **256色(8位)**
- **64K色(16位)**
- **16M色(24位)真彩色** (**一个像素点要3B**)

图像编码

位图图像

- 图像被分成**像素矩阵**，每个像素是一个小点，每个像素用一个位模式表达 (**放大会失真**)
- GIF, JPEG, PNG, TIFF, BMP, PCX

图像质量

- 如果对其离散化后的网格点越多，即**分辨率越高**，则图像越精细，质量越好；
- 如果颜色编码所采用的二进制位数越多，即**所能表示的颜色数越多**，则图像质量越好。
- 例：一幅分辨率为640×480的真彩色(24位，3Byte)图像，如果不做压缩，其所需的存储空间约为：
 $640 \times 480 \times 3 \text{Byte} = 921600 \text{Byte} = 900 \text{KB} \approx 0.88 \text{MB}$

矢量图形

- **将图像分解为曲线和直线的集合**，**每条直线/曲线由数学公式表示** (**放大不会失真**)
- svg, flash, pdf, ps, Type 1 字体

影像编码

是在时间和空间上对活动场景的离散采样，影像中的一张图像是对某一时刻场景的空间离散采样，称为**影像的一帧**。

程序运行的基本原理

计算机信息处理过程：信息的输入，信息的存储和处理，信息的输出

二进制数的运算

算术运算

0x0=0

0x1=0

1x0=0

1x1=1

逻辑运算

表示推理的计算

布尔代数：

- 基本状态：是、非
- 基本计算：与、或、非

计算机中数的逻辑运算方法：

- 基本逻辑运算：
 - 与逻辑（与运算、逻辑乘）
 - &
 - 0&0=0; 0&1=0; 1&0=0; 1&1=1
 - 用途：可获取整型**某一位**的值；使整型数**某几位**的值置0

```
#include <stdio.h>
int main()
{
    int a=0xfe,b; //a的低8位1111 1110, 前24位为0, 略
    b=a&0xf9;
    //这里只给出低8位, 与1111 1001 &, 倒数第2,3位置0, 1111 1000
    printf("%x %x",a,b);
    return 0;
}
```

- 或逻辑（或运算、逻辑加）
 - |
 - 0|0=0; 0|1=1; 1|0=1; 1|1=1
 - 用途：可使整型**某一位**的值置1

```
#include <stdio.h>
int main()
{
    int a=0xf1,b; //a的低8位1111 0001, 前24位为0, 略
    b=a|0x8;
    // 这里只给出低8位, 与0000 1000 |, 倒数第4位置1, 1111 1001
    printf("%x %x",a,b);
    return 0;
}
```

- 非逻辑 (非运算、逻辑反)

- \sim
- $\sim 1=0$; $\sim 0=1$
- 用途: 可使整型数所有位取反

```
#include <stdio.h>
int main()
{
    int a=0,b; //a的二进制为32位0, 16进制表示为00000000
    b=~a; //b的二进制为32位1, 16进制表示为ffffffff
    printf("%x %x %d",a,b,b);
    return 0;
}
```

非按位取反用!

- 复合逻辑运算

- 异或

- 两数相同为0; 两数相异为1
- 位运算特点: 与1异或原数取反, 与0异或原数不变
- 逻辑异或: \wedge
- $0\wedge 0=0$; $0\wedge 1=1$; $1\wedge 0=1$; $1\wedge 1=0$
- 具有交换律和结合律
- 特殊用法: 整型数a和b的交换, $a=a\wedge b$; $b=a\wedge b$; $a=a\wedge b$; 对整数某些二进制位取反: 与1异或

```
#include <stdio.h>
int main()
{
    int a=0xfe,b; //a的低8位1111 1110, 前24位为0, 略
    b=a^0x9; //与0000 1001 ^, 倒数第1,4位取反, 1111 0111
    printf("%x %x",a,b);
    return 0;
}
```

- 同或

- 两数相同为1; 两数相异为0
- 加法的计算: 本位异或, 进位与
- 半加器: 考虑本位和+向高位的进位
- 全加器: 低位的进位+本位和+向高位的进位

- 与非

- 先与再非

- 或非

- 先或再非

- 与或非

CPU的内部结构和工作原理

CPU: Central Processing Unit

总线: 系统部件之间信息传送的公用信号线

- 数据总线
- 控制总线
- 地址总线

总线宽度: 一条总线上一次可传输的二进制位数

例: 32位地址总线宽度可以直接访问的主存地址空间为4GB

寄存器组

- 由一组寄存器组成的高速存储单元
- 用于暂时存放运算数据或其它信息
 - 整数类型的操作数或运算结果
 - 浮点数类型的操作数或运算结果
 - 指令
 - 指令地址
 - 各种内部标志信息

主存储器: 随机存取存储器 (Random Access Memory, RAM), 具有易失特性

高速缓存: 为了缓和CPU与主存储器之间的速度矛盾, 在CPU和主存储器之间设置一个缓冲性的高速存储部件(硬件), 这个部件称为高速缓存(简称“缓存”, 英文词是cache)

局部性原理:

- 程序执行时, CPU访问存储器, 无论是存取指令还是存取数据, 所访问的存储单元都趋于聚集在一个较小的连续区域中, 这种现象称为局部性, 有两种不同类型:
 - **时间局部性 (Temporal Locality):** 如果一个信息项正在被访问, 那么在近期它很可能还会被再次访问。程序循环、堆栈等是产生时间局部性的原因。
 - **空间局部性 (Spatial Locality):** 在最近的将来将用到的信息很可能与现在正在使用的信息在空间地址上是临近的。

外部存储器: 容量大得多但速度低很多, 具有非易失性

程序控制器

1. 分析/解释该指令的编码内容;
2. 确定为执行该指令应该完成的动作;
3. 确定指令相关的参数; 例如: 对于一个“加法指令”, 需要确定两个被加数的地址
4. 将所需的数据从主存储器读取到CPU的寄存器中;
5. 要求算术逻辑运算器进行相关的运算动作;
6. 指示算术逻辑运算器将运算结果放入寄存器或主存储器中。

算术逻辑运算器

ALU: Arithmetic Logical Unit

主要进行**算术运算**和**逻辑运算**

1. 一条加法指令（其中包含了两个被加数/操作数的地址）进入CPU；
2. 程序控制器分析该指令，判断两个操作数是在寄存器内，还是在主存内；
3. 如果在主存内，程序控制器 从主存内读入操作数；
4. 程序控制器将加法运算提交给 ALU；
5. ALU 进行加法运算；
6. ALU 根据程序控制器的指示，将运算结果存放到寄存器或主存中。

中断处理器

EOF

- 定义在<stdio.h>中
- 为了区分文件的有效数据与输入结束符，C语言规定在没有输入时，`getchar()` 函数返回一个特殊值，与任何都不同，称为EOF(End of File)
- 利用Ctrl+Z组合键，可使 `getchar()` 返回值为EOF。

Ctrl+C组合键结束程序

中断机制

- 一些事件会打断程序的正常执行：
 - 运算结果溢出，整数除法指令的除数为0
 - 无法获得有效数据，出现一些错误“segmentation fault”
 - 磁盘完成外部读写事件
- CPU除了能够正常不断地执行指令之外，还必须具有正常执行被打断时，处理一些特殊情况的处理机制，也被称为**异常控制**，或**中断机制**。

中断信号的处理

- 当发现中断信号后，**程序控制器**暂停正在运行的程序，保存该程序的运行现场（CPU内的各种状态信息）；
- **程序控制器**根据中断信号的编码，从特定位置启动**中断处理程序**（由操作系统提供）；
- **中断处理程序**运行完毕后，**程序控制器**恢复被暂停的程序。

中断的种类

- 内部异常：CPU内部异常引起的
 - 硬故障中断：硬件线路故障，比如主存校验线路出错
 - 程序性异常：执行CPU某个指令引起的的CPU内部异常，如除数为0、溢出、断点、单步跟踪、地址越界、访问超时等
- 外部中断：外设完成某些任务时发生的特殊事件：打印机缺纸等

中断信号的接收

中断处理器负责中断信号的接收，并将中断信号的编码、中断处理程序的起始地址 传给 程序控制器

中断信号的检测

程序控制器在每条指令执行完毕后，都会检测是否出现了新的中断信号

CPU的主要性能指标

- 工作主频：
 - 即CPU内核工作的**时钟频率**；1.7GHz~3.0GHz，
 - 在CPU内数字脉冲信号震荡的速度（时钟）
- 运算字长：
 - **CPU一次能够处理的二进制位数；32位/64位**
- 数据总线宽度、地址总线宽度
- 运算速度：
 - 每秒钟执行的指令数；例如：1000**MIPS**（Million Instructions Per Second）
- 指令集：指令系统的处理能力（复杂指令，精简指令）等。
 - 在实际选购CPU时，则主要从CPU的制造商、CPU型号、主频、高速缓存的容量等几个方面考虑。
 - 64位指令集就是运行64位数据的指令，处理器一次运行64bit数据。

指令系统

指令：组成程序的基本单位

每一条指令：

- 规定了CPU执行指令应该完成的工作（运算、或其它控制动作）
- 控制CPU的相关部件执行微操作，从而完成指令所规定的功能。

CPU的指令系统是CPU芯片的硬件与使用它的软件之间的一种严格的协议，反映了CPU能够完成的**全部功能**。

CPU的“指令系统”规定了：

- 它所能执行指令的**全部类别**，
- 指令的编码方式和每一类指令所涉及的参数等。

指令的分类

- 存储访问指令
- 算术运算指令
- 逻辑运算指令
- 条件判断和分支转移指令
- 输入输出指令
- 其他用于系统控制的指令

指令工作周期：程序控制器按照“**读取指令—执行指令**”的周期循环地工作

一个指令周期 一般需要占用**多个CPU时钟周期/时钟节拍**

CPU时钟周期/时钟节拍

- CPU完成一个原子动作的基本时间单位
- 如果 一个CPU的时钟频率是 y GHz, 那么这个CPU的时钟周期/时钟节拍 是 $1/y \times 10^{-9}$ 秒

主存储器及其与CPU之间的信息传输

存储空间的管理

为了更有效地进行管理, 通常以**8个比特** (一个字节) 为一个**存储管理单元** (简称: **存储单元**)

- 每个存储单元都有其特定且唯一的地址, 称为: **存储地址**。
- 存储地址为整数编码, 可表示为二进制整数。
- 由主存储器的所有存储单元的地址构成的集合, 称为**地址空间**。
- 表示地址空间所需的二进制位数, 称为**地址宽度**。
- 内存容量越大, 地址空间也就越大, 地址宽度也必须相应加大。

在一段程序中, 变量和存储单元相对应

- 变量对应于存储单元
- 变量内容对应于存储单元中的数据
- 指针型变量: 专门存放**存储单元地址**的变量

计算机的数学理论模型——图灵机

基本思想: 用机器来模拟人类用纸和笔进行数学运算的过程

计算机系统

计算机类型

按照性能和规模分类

- 巨型机
- 大型机
- 中型机
- 小型机
- 工作站
- 个人计算机
 - 台式计算机
 - 笔记本/便携式计算机
 - 个人数字助理 (PDA)

个人计算机系统

个人计算机硬件系统

硬件系统概述

总线技术

- 系统总线 (System Bus) – connects CPU to Memory
- 外部总线 (Local Bus or I/O Bus) – carries data to/from peripheral devices.
- 总线宽度 (Bus width) : measured in bits
- 总线频率 (Bus clock speed) : 指总线信号的时钟频率measured in MHz
- 总线带宽 (总线传输速率、总线吞吐率 (Throughput)) : 指单位时间传输的数据量measured in MB/s, 每秒位(b/s或bps)
- 总线带宽=传输的数据量/需要的时间
 - 例: 8088处理器的数据总线为8位, 典型的时钟频率是5MHz, 即每个时钟周期是 $1/5\text{MHz}=0.2\times 10^{-6}$ 秒, 8088需要四个时钟周期构成一个总线周期, 实现一次8位数据传送, 故8088处理器的总线带宽是总线带宽= (总线宽度/ (n ×时钟周期) =总线宽度×总线频率/n)
 $8/(4\times 0.2\times 10^{-6})\text{b/s}=10^7\text{b/s}=10\text{Mb/s}=1.25\text{MB/s}$

主机内的硬件设备

- 电源适配器
- CPU (Central Processing Unit)
- 主存储器(Main Memory)/内部存储器
- 外部存储器(External Storage)
 - 硬盘/硬盘驱动器
 - 磁盘、固态硬盘 (SSD)
 - 光盘/光盘驱动器; 软盘/软盘驱动器
- 外围设备适配器/控制器
 - 显示器适配器(显卡)
 - 以太网适配器(网卡)
 - 声音适配器(声卡)
 - 鼠标适配器、键盘适配器
 -
- 主板(Mother Board, Main Board, System Board)

主板上的主要部件:

- 总线(Bus)
 - 前端总线(Front-Side Bus, FSB) 【CPU↔芯片组】
 - 内存总线(Memory Bus) 【内存↔芯片组】
 - 高速图形总线 【显卡↔芯片组】
 - PCI-E, AGP
 - 外部存储器总线 【硬盘适配器、光驱适配器↔芯片组】

- SATA, IDE, SCSI
- 低速设备总线(LPC)
 - 【BIOS→芯片组】
 - 【键盘适配器、鼠标适配器、软驱适配器、...↔芯片组】
- 通用型总线【其它外设适配器↔芯片组】
 - USB (Universal Serial Bus)
 - PCI-E, PCI, ISA
- 内部总线(Internal Bus)【芯片组内部互联】
- 接口：插槽(Slot), 插座(Socket), 端口(Connector/Port)
 - CPU插槽/座, 内存插槽, PCI(-E)插槽, ISA插槽
 - 鼠标端口, 键盘端口, USB端口
- 芯片组(Chipset)
 - 北桥芯片(North Bridge): ↔CPU、内存、显卡
 - 南桥芯片(South Bridge): ↔BIOS、其它外围设备适配器
- BIOS (Basic Input/Output System)

总线：

系统部件之间信息传送的公用信号线；

- 具有**汇集与分配信息、选择发送信息的部件与接收信息的部件、总线控制权的建立与转移**等功能；
(在芯片组的控制下)
- 内部含有三种类型的信号线：**数据总线、控制总线、地址总线**；
- 常见的总线标准
 - ISA (Industrial Standard Architecture Bus) : 8/16位
 - PCI (Peripheral Component Interconnection Bus) : 32/64位
 - AGP (Accelerated Graphics Port) : 显卡专用总线
 - PCI-E (PCI Express) : PCI的升级版；替代PCI和AGP
 - IDE (Integrated Drive Electronics) : 用于连接硬盘、光驱
 - SATA (Serial Advanced Technology Attachment) : IDE的升级版
 - USB (Universal Serial Bus) : 由IBM、Intel及Microsoft等多家公司共同开发的新型外设连接技术

BIOS和CMOS

- BIOS (Basic Input/Output System)
 - 基本输入输出系统
 - 主板上的一块只读的芯片，其中固化了一组程序
 - 用于：操作系统的引导、对CMOS中参数的设置
- CMOS (Complementary Metal Oxide Semiconductor)
 - 互补金属氧化物半导体
 - 主板上的一块可读写的芯片，用来保存当前系统的硬件配置和用户对系统某些参数的设定
 - 需要一块电池为其供电（CMOS电池）

硬盘：

- 在硬盘上存储信息是以**扇区**为信息单位成批进行读写的。一个扇区的位置（地址）由它所在的**盘面编号、磁道编号和扇区在磁道中的位置（编号）**三者联合确定。
- 磁盘读写动作过程分为三个阶段：磁头定位（磁道）、扇区定位和实际读写

磁带：

- 磁带从固定的磁头下面滑过，电信号流经磁头时产生的磁场使得磁带的某些部分被磁化，将信息记录在磁带上，信息读出的过程正相反。
- 但数据磁带上记录的是**数字信息**，也就是二进制形式的信息，不像录音磁带，那里记录的是**模拟形式**的信息。
- 适合**顺序读写**，缺点是**信息查找速度慢**

光盘：

- 利用**激光原理存储**
- 分类：只读光盘、可刻录光盘、可重复刻录光盘

半导体存储器：

- 从存取功能上可以分为两大类
 - 只读存储器(ROM，即Read Only Memory)
 - PROM 指的是“可编程只读存储器”这样的产品只允许写入一次，所以也被称为“一次可编程只读存储器”
 - EPROM 指的是“可擦写可编程只读存储器”。它的特点是具有可擦除功能，擦除后即可进行再编程，但是缺点是擦除需要使用紫外线照射一定的时间。
 - **EEPROM** 指的是“电可擦除可编程只读存储器”，它的最大优点是可直接用电信号擦除，也可用电信号写入。（闪存）
 - 随机存取存储器(RAM，即Random Access Memory)

基本输入输出设备

- 基本输入设备：键盘、鼠标
- 基本输出设备：显示器

显示器：

- CRT (Cathode Ray Tube，阴极射线管)
- LCD (Liquid Crystal Display，液晶显示器)
- LED (Light Emitting Diode，发光二极管)

显示器的主要技术指标

- 尺寸
- 点距
- 分辨率：屏幕上可以容纳的像素的个数
- 刷新频率
- 辐射性和绿色环保能力

通信设备

- 网络适配器：用于将计算机连接到局域网
- 无线网卡：用于将计算机连接到无线网络
- 调制解调器(Modem)
 - 基本功能：模拟信号和数字信号之间的变化
 - 用于通过普通公用电话网络与计算机网络建立连接，包括：
 - ISDN (Integrated Service Digital Network) 接入卡
 - ADSL (Asymmetrical Digital Subscriber Loop) 接入卡
- 电缆调制解调器
 - 让计算机利用有线电视网进行数据传输

外围设备

- 打印机
- 绘图仪
- 投影仪
- 手写板/图形输入板/数字化板
- 扫描仪/数码照相机/数码摄像机
- 声卡/麦克/音箱

计算机软件系统与信息安全

计算机软件

- 应用含义：指示计算机完成任务的，以电子格式存储的指令序列和相关数据
- 学科含义：软件= 程序+ 文档

计算机软件系统

- 系统软件：
 - 操作系统
 - 其它系统软件：数据库、设备驱动程序、编译器、...
- 应用软件：各类领域、行业的应用软件，如聊天软件、办公软件、娱乐软件

操作系统：是连接**硬件**和**应用软件**的中间层软件

- 一些著名的操作系统
 - 工作站、大型机操作系统：UNIX
 - 个人计算机操作系统：DOS、Windows、Macintosh OS、Linux
- 操作系统的引导：在计算机的电源接通，硬件开始工作后，首先必须把操作系统的**常驻内核**从磁盘装入主存储器，并且使它进入正常工作状态，这样的过程称为操作系统的引导（System Booting）。
- 操作系统的功能
 - CPU管理和任务管理：
 - 对CPU的工作时间进行划片
 - 将CPU的各个时间片断分配给各项任务

- 管理CPU在各个任务之间的切换
- 存储资源管理：
 - 对主存空间的分配和收回等管理工作
- 输入/输出设备管理
 - 管理计算机与各种输入/输出设备之间的数据交换
- 信息资源管理
 - 通过“**文件系统**”来进行，因此又称为文件管理
- 用户界面

文件：是在逻辑上具有完整意义的并赋有名称的信息集合体。

- 操作系统以**文件**为单位，在外存上管理并存取信息
- 文件的重要属性
 - 文件名
 - 主文件名、拓展名
 - 类型
 - 位置
 - 大小
 - 文件属性
 - 其他：文件创建者、创建时间、修改时间等
- 分类
 - 文本文件
 - 二进制文件：应用程序、图形/图像文件、声音文件
 - 可执行文件
 - 数据文件

文件系统：文件系统涉及计算机的**文件的结构**和**文件的组织**，以及负责**管理文件**的软件系统，一般把这两者的总和称为计算机的文件系统。文件系统是由**文件和文件夹**组成的**树状结构**。

- 文件的组织结构
 - 用一种称为**目录 (Directory)** 或**文件夹 (File Folder)** 的分层树状结构把文件组织起来，使各种文件分散到目录结构的不同位置，以利于对文件的分类管理和使用。
 - 一个文件的**路径名**是由根目录到该文件的通路上**所有目录名和该文件的符号名**组成的。
- 各种文件系统
 - FAT32：文件大小不能大于4G
 - NTFS：单文件最大64GB
 - NTFS5.0：单文件最大2TB

实体鉴别：实体鉴别就是确认实体是它所声明的

- 实现途径：
 - 三种途径之一或他们的组合
 - 所知 (Knowledge) :密码、口令
 - 所有 (Possesses):身份证、护照、信用卡、钥匙

- 个人特征：指纹、笔迹、声纹、手型、血型、视网膜、虹膜、DNA以及个人动作方面的一些特征

访问控制

- 概念：是针对越权使用资源的防御措施
- 基本目标：防止对任何资源（如计算资源、通信资源或信息资源）进行未授权的访问
 - 未授权的访问包括：
 - 非法用户进入系统。
 - 合法用户对系统资源的非法使用。

信息安全

缓冲区溢出Buffer Overflow

- 缓冲区：程序运行期间，在内存中分配的一个连续的区域，用于保存包括字符数组在内的各种数据类型。
- 溢出：所填充的数据超出了原有缓冲区的边界，并非法占据了另一段内存区域。
- 缓冲区溢出：由于填充数据越界而导致程序原有流程的改变，黑客借此精心构造填充数据，让程序转而执行特殊的代码，最终获得系统的控制权。

在C语言中，指针和数组越界不保护是Buffer overflow的根源，而且，在C语言标准库中就有许多能提供溢出的函数，如 `strcat()`，`strcpy()`，`sprintf()`，`vsprintf()`，`gets()`

计算机病毒

- 定义：凡能够引起计算机故障，破坏计算机数据的程序统称为计算机病毒(Computer Virus)
- 我国的条例：计算机病毒，是指编制或者在计算机程序中插入的破坏计算机功能或者毁坏数据，影响计算机使用，并能**自我复制的一组计算机指令或者程序代码**。
- 基本特征
 - 寄生性
 - 传染性
 - 隐蔽性
 - 潜伏性
 - 可触发性
 - 破坏性
 - 产生的必然性
 - 非授权性
- 检测技术
 - 计算机感染病毒后，会引起一系列变化，检测正是以此为依据的。
 - 比较法:进行原始的或正常特征与被检测对象的特征比较
 - 校验和法：计算出正常文件的程序代码的校验和
 - 搜索法：用病毒的特定字符串作为被检测对象进行扫描
 - 行为监测法：检测病毒发作时的共同行为

防火墙：墙是位于两个或多个网络之间，执行访问控制策略的一个或一组系统，是一类防范措施的总称。

- 防火墙的访问控制能力
 - 服务控制，确定哪些服务可以被访问
 - 方向控制，对于特定的服务，可以确定允许哪个方向能够通过防火墙
 - 用户控制，根据用户来控制对服务的访问
 - 行为控制，控制一个特定的服务的行为

互联网及应用

互联网简介

Internet（网际网）：相互连接的计算机网络

- 是一种 将处于不同地理位置且具有独立功能的多个计算机系统通过**通信设备和线路**连接起来，在功能完善的**网络软件**的支持下，实现彼此之间的数据通信和资源共享的系统
- 是**计算机科学**和**通信科学**密切结合的产物

互联网络分类（按地域）

- 局域网（LAN，Local Area Network）：一个建筑物范围内、利用以太网技术形成的计算机网络（通常是公司或学校内构建）
- 广域网（WAN，Wide Area Network）：跨越国家和省市地域的网络（通信服务商提供）
- 城域网（MAN，Metropolitan Area Network）：城市内建立的通信网络，比LAN范围要广
- 内联网（Intranet）：企业/组织的内部网（可能是WAN、LAN）

互联网的通信协议

- **HTTP**
 - Hyper Text Transfer Protocol
 - 超文本传输协议
 - 浏览网页时，客户端和服务器的通讯协议
- **TCP**
 - Transmission Control Protocol
 - 传输控制协议
 - 可靠的传输协议
- **IP**
 - Internet Protocol
 - 网际协议
 - 非可靠的传输协议（best-efforts,尽力投递）

MAC地址

- 用于识别**数据链路**中互连的节点。
- MAC地址长48比特，任何一个网卡（NIC）的MAC地址都是**唯一**的。
- 第3 - 24位标识**厂商识别码**，每个NIC厂商都有唯一的识别数字。
- 第25 ~ 48位是每个厂商内部为**识别每个网卡**而用。

子网与子网掩码

- 与运算的一个应用就是把一个位模式的指定位置0。这种情况下，第二个输入称为**掩码**。掩码中的0位对第一个输入中相应的位复位（置0）。
- 网络标识相同的计算机必须同属于同一个链路，例如，架构B类IP网络时，理论上一个链路内允许6万5千多台计算机，但实际不会存在这种情况。
- 现在，一个IP地址的网络标识和主机标识已经不再受限于该地址的类别，而是由一个叫“子网掩码”的识别码，通过子网网络地址细分出比A类、B类、C类更小粒度的网络。

特殊意义的IP地址

- 保留地址，只用于内部通信，包含在这个范围内的IP地址都属于私有IP，而在此之外的IP地址称为全局IP。
 - 10.0.0.0 - 10.255.255.255
 - 172.16.0.0 - 172.31.255.255
 - 192.168.0.0 - 192.168.255.255
- 回送地址:网络号是127的地址（127.xx.yy.zz）保留做回路(loopback)测试，大多数系统使用127.0.0.1

IPv6地址

- IPv6地址结构= **前缀+ 接口标识**
 - 前缀：相当于v4地址中的网络ID
 - 接口标识：相当于v4地址中的主机ID
- IPv6地址的基本表达方式是X:X:X:X:X:X:X:X,
 - 4位一组，中间用“:”隔开
 - 其中X是一个4位十六进制整数(16位二进制位)。
 - 每个地址包括8个整数X，共计128位($4 \times 4 \times 8 = 128$)
- IPv6地址示例
 - CFDE:710A:3222:7416:2435:1121:4900:3020
 - 地址中的每个整数都必须表示出来
- 压缩格式
 - 每段的前导0可以省略，但至少留1个0
 - “::”符号在一个地址中只能出现一次。
 - 地址前缀长度用“/xx”来表示，如：1::1/64
- 以下是同一个地址不同表示法的例子：
 - 0001:0123:0000:0000:0000:ABCD:0000:0001/96
 - 1:123:0:0:0:ABCD:0:1/96
 - 1:123::ABCD:0:1/96
 - 2001:0410:0000:0001:0000:0000:0000:45ff
 - 2001:410:0:1:0:0:0:45ff
 - 2001:410:0:1::45ff
 - 0:0:0:0:0:0:0:1
 - 00:1

DHCP

- 为了实现自动设置IP地址，统一管理IP地址分配，就产生了DHCP (Dynamic Host Configuration Protocol) 协议，有了DHCP，计算机可以即插即用。

DNS

- IP地址不便记忆，产生了一个可以**有效管理主机名和IP地址之间对应关系的系统**，就是域名系统 (Domain Name System,DNS)。
- 域名
 - **用文字的方式标识特定的IP地址**
 - www.pku.edu.cn, www.sohu.com, ...
 - 域名空间：顶级域名、二级域名、三级域名、四级域名
- **DNS (Domain Name System)**
 - **一个将域名映射成相应的IP地址的服务系统**
 - www.pku.edu.cn → 162.105.130.161
 - www.sohu.com → 211.159.191.76
 - mail.pku.edu.cn → 162.105.129.99
 - 由全世界数量众多的DNS服务器组成
 - 北大校内的两个DNS服务器：
 - 162.105.129.122, 162.105.129.88

互联网的应用

- 初期主要是文字类的信息服务
 - 电子邮件、文件传输、远程终端
- 逐步涌现出丰富多样的多媒体类的信息服务
 - WWW、即时通信、视频点播、P2P下载
 - 电子政务、电子商务、网络游戏、...

基本模式：客户端/服务器模式 (client/server)

万维网

- 信息存储的文件称为**网页**，Web页面存储在Web服务器上。
- 阅读网页的计算机被称为**Web客户端**。
- Web客户端查看网页的程序称为**Web浏览器**。
- 流行的浏览器有Chrome、Firefox、Internet Explorer等。
- 超链接 (超文本连接, Hyper Text Link)
- 统一资源定位符 (Uniform Resource Locator, URL)
 - 格式：访问资源的协议: 资源地址
 - HTTP URL的格式：http://主机名称: 端口/路径
 - <http://hi.baidu.com>
 - <https://news.sohu.com/>
 - <https://portal.pku.edu.cn/portal2017/#/index>

电子邮件

- 协议：**SMTP** (Simple Mail Transfer Protocol) 、POP3 (Post Office Protocol 3)

文件传输

- 文件传输协议 (File Transfer Protocol, FTP)
- **FTP服务器** (File Transfer Protocol Server) 是在互联网上提供文件存储和访问服务的计算机，它们依照FTP协议提供服务。用来在两台计算机之间传输文件，是网络中经常采用的资源共享方式之一。