

# 计算机科学导论

June 27, 2019

# 写在前面

出于希望在电子设备阅读的目的，我把《计算机科学导论》一书的内容键入电脑，形成文档。此文档并不完整，原因是部分被我省略掉；有很多词用错，原因是键入错误。文档的章节编号是排版系统自动生成的，由于我省略掉了原书中的几个章节，文档章节编号与元书并不一致。文档在将来可能处于更新状态，也可能处于冻结状态。

葡萄红  
一八年八月十七日

# Contents

写在前面	i
<b>1 绪论</b>	<b>1</b>
1.1 图灵模型	1
1.1.1 数据处理器	1
1.1.2 可编程数据处理器	1
1.1.3 通用图灵模型	2
1.2 冯诺伊曼模型	2
1.2.1 4 个子系统	2
1.2.2 存储程序概念	2
1.2.3 指令的顺序执行	3
1.3 计算机组成部分	3
1.3.1 计算机硬件	3
1.3.2 数据	3
1.3.3 计算机软件	3
<b>2 计算机组成</b>	<b>5</b>
2.1 引言	5
2.2 中央处理器	5
2.2.1 算术逻辑单元	5
2.2.2 寄存器	6
2.2.3 控制单元	6
2.3 主存储器	6
2.3.1 地址空间	6
2.3.2 存储器的类型	6
2.3.3 存储器的层次结构	7
2.3.4 高速缓冲存储器	7
2.4 输入/输出子系统	8
2.4.1 非存储设备	8
2.4.2 存储设备	8
2.5 子系统的互连	9
2.5.1 CPU 和存储器的连接	9
2.5.2 I/O 设备的连接	9
2.5.3 输入/输出设备的寻址	10
2.6 程序执行	10
2.6.1 机器周期	10
2.6.2 输入/输出操作	10
2.7 不同的体系结构	10

2.8	简单计算机	10
2.8.1	可重用性	10
<b>3</b>	<b>计算机网络和因特网</b>	<b>11</b>
3.1	引言	11
3.1.1	网络	11
3.1.2	因特网	12
3.1.3	硬件和软件	12
3.1.4	协议分层	12
3.1.5	TCP/IP 协议族	14
3.2	应用层	14
3.2.1	提供服务	14
3.2.2	应用层模式	15
<b>4</b>	<b>操作系统</b>	<b>16</b>
4.1	引言	16
4.1.1	操作系统	16
4.1.2	自举过程	16
4.2	演化	17
4.2.1	批处理系统	17
4.2.2	分时系统	17
4.2.3	个人系统	17
4.2.4	并行式系统	17
4.2.5	分布式系统	17
4.2.6	实时系统	18
4.3	组成部分	18
4.3.1	用户界面	18
4.3.2	内存管理器	18
4.3.3	进程管理器	20
4.3.4	文件管理器	21
4.4	主流操作系统	21
4.4.1	UNIX	21
4.4.2	Linux	22
4.4.3	Windows	22
<b>5</b>	<b>算法</b>	<b>24</b>
5.1	概念	24
5.1.1	非正式定义	24
5.1.2	细化	24
5.1.3	泛化	25
5.2	三种结构	25
5.2.1	顺序	25
5.2.2	判断	25
5.2.3	循环	25
5.3	算法的表示	25
5.3.1	UML	25
5.3.2	伪代码	25
5.4	更正式的定义	25
5.4.1	定义良好	26
5.4.2	明确步骤	26

5.4.3	产生结果 . . . . .	26
5.4.4	在有限的时间内终止 . . . . .	26
5.5	基本算法 . . . . .	26
5.5.1	求和 . . . . .	26
5.5.2	乘积 . . . . .	26
5.5.3	最大和最小 . . . . .	27
5.5.4	排序 . . . . .	27
5.6	子算法 . . . . .	27
5.7	递归 . . . . .	27
5.7.1	迭代的定义 . . . . .	27
5.7.2	递归的定义 . . . . .	28

# Chapter 1

## 绪论

今天，计算机科学一词是一个非常广泛的概念。尽管如此，在本书里，我们将其定义为“和计算机有关的问题”。本章首先阐述什么是计算机，接着探索和计算机直接相关的一些问题。首先我们将计算机看成一个图灵模型，这是从数学上和哲学上对计算机的定义。然后，阐述当今的计算机是如何建立在冯·诺伊曼模型基础上。最后介绍计算机这一改变文明的装置的简明历史。

### 1.1 图灵模型

Alan Turing 在 1937 年首次提出了一个通用计算设备的设想。他设想所有的计算都能在一个中特殊的机器上执行，这就是现在所说的图灵机。尽管图灵对这样一种机器进行了数学上的描述，但他还是更有兴趣关注计算的哲学定义，而不是建造一台真是的机器。他将该模型建立在人们进行计算过程的行为上，并将这些行为抽象到用于计算的机器的模型中，这才真正改变了世界。

#### 1.1.1 数据处理器

在讨论图灵模型之前，让我们把计算机定义成一个数据处理器。依照这种定义，计算机就可以被看作是一个接受输入数据、处理数据并产生输出数据的黑盒。尽管这个模型能够体现现代计算机的功能，但是他的定义还是太宽泛。按照这种定义，也可以认为便携式计算器是计算机。

另一个问题是这个模型并没有说明它处理的类型以及是否可以处理一种以上的类型。换句话说，它并没有清楚地说明基于这个模型的机器能够完成操作的类型和数量。它是专用机器还是通用机器呢？

这种模型可以表示为一种设计用来完成特定人物的专用计算机，比如用来控制建筑物温度或汽车油料使用。尽管如此，计算机作为一个当今使用的术语，是一种通用的机器，它可以完成各种不同的工作。这表明我们需要该模型改变为图灵模型来反映当今计算机的实现。

#### 1.1.2 可编程数据处理器

图灵模型是一个适用与通用计算机的更好模型。该模型添加了一个额外的元素——程序到不同的计算机器中。程序是用来告诉计算机对数据进行处理指令集合。

在这个图灵模型中，输入数据是依赖来两个方面因素的结合作用：输入数据和程序。对于相同的数据输入，如果改变程序，则可以产生不同的输出。类此

的，对于同样的程序，如果改变输入数据，其输出结果也将不同。最后，如果输入数据和程序保持不变，输出结果也将不变。让我们看看下面三个示例。

### 1.1.3 通用图灵模型

通用图灵集是对现代计算机的首次描述，该机器只要提供了合适的程序就能做任何运算。可以证明，一台很强大的计算机和通用图灵机一样能进行同样的运算。我们所需要的仅仅是为这两者提供数据以及用于描述如何做运算的程序。实际上，通用图灵机能做任何可计算的运算。

## 1.2 冯诺伊曼模型

基于通用图灵机建造的计算机都是在存储器中存储数据。在 1944~1945 年期间，冯·诺依曼指出，鉴于程序和数据在逻辑上是相同的，因此程序也能存储在计算机的存储器中。

### 1.2.1 4 个子系统

基于冯·诺依曼模型建造的计算机分为 4 个子系统：存储器、算术逻辑单元、控制单元和输入/输出单元。

#### 1. 存储器

存储器是用来存储的区域，在计算机的处理过程中存储器用来存储数据和程序，我们将在这一章后边讨论存储数据和程序的原因。

#### 2. 算术逻辑单元

算术逻辑单元是用来进行计算和逻辑运算的地方。如果是一台数据处理计算机，它应该能够对数据进行算术运算。当然它也应该可以对数据进行一系列逻辑运算，正如我们将在第 4 章看到的那样。

#### 3. 控制单元

控制单元是对存储器、算术逻辑单元、输入/输出等子系统进行控制的单元。

#### 4. 输入/输出

输入子系统负责从计算机外部接收输入数据和程序；输入子系统负责将计算机的处理结果输出到计算机外部。输入/输出子系统的定义相当广泛，它们还包含辅助存储设备，例如，用来存储处理所需的程序和数据的磁盘和磁带等。当一个磁盘用于存储处理后的输出结果，我们一般就可以认为它是输出设备，如果从该磁盘上读取数据，则该磁盘就被认为是输入设备。

### 1.2.2 存储程序概念

冯·诺依曼模型中要求程序必须存储在内存中。这和早期只有数据才存储在存储器中的计算机结构完全不同。完成某一任务的程序是通过操作一系列的开关或改变其配线来实现。

现代计算机的存储单元用来存储程序及其响应数据。这意味着数据和程序应该具有相同的格式，这是因为他们都存储在存储器中。实际上它们都是以位模式存储在内存中的。

### 1.2.3 指令的顺序执行

冯·诺依曼模型中的一段程序是由一组有限的指令组成。按照这个模型，控制单元从内存中提取一条指令，解释指令，接着执行指令。换句话说，指令就一条接着一条地顺序执行。当然，一条指令可能会请求控制单元以便跳转到其前面或后面的指令去执行，但是这并不意味着指令没有按照顺序来执行。指令的顺序执行是基于冯·诺依曼模型的计算机初始条件。当今的计算机以最高的顺序来执行程序。

## 1.3 计算机组成部分

我们可以认为计算机由三大部分组成：计算机硬件、数据和计算机软件。

### 1.3.1 计算机硬件

### 1.3.2 数据

冯·诺依曼模型清楚地将一台计算机定义为一台数据处理机。它接受输入数据，处理并输出相应的结果。

#### 1. 存储数据

冯·诺依曼模型并没有定义数据如何存储在计算机中。如果一台计算机是一台电子设备，最好的数据存储方式应该是电子信号，例如以电子信号的出现和消失的定方式来存储数据，这意味着一台计算机可以以两种状态之一的形式存储数据。

显然，在日常使用的数据并不是以两种状态之一的形式存在，例如，我们数字系统中使用的数字可以是 0~9 十种状态中的任何一个。但是你不能将这类信息存储到计算机内部，除非将这类信息转换成另一种只使用两种状态的系统。同样，你也需要处理其他类型的数据，它们同样也不能直接存储到计算机中，除非将它们转变成合适的形式。

#### 2. 组织数据

尽管数据只能以一种形式存储在计算机内部，但在计算机外部却可以表现为不同的形式。另外，计算机开创了一门新兴的研究领域——数据组织。在将数据存储到计算机之前，能否有效地将数据组织成不同的实体和格式？如今，数据并不是按照杂乱无章的次序来组织信息的。数据被组织成许多小的单元，再由这些小的单元组成更大的单元，等等。

### 1.3.3 计算机软件

图灵或冯·诺依曼模型的主要特征是程序的概念。尽管早期的计算机并没有计算机存储器中存储程序，但它们还是使用了程序的概念。编程在早期的计算机中体现为系列开关的打开或关闭以及配线的改变。编程在数据实际开始处理之前是由操作员或工程师完成的一项工作。

#### 1. 程序必须是存储的

在冯·诺依曼模型中，这些程序被存储在计算机的存储器中，存储器中不仅要存储数据，还要存储程序。



## 2. 指令的序列

这个模型还要求程序必须是有序的指令集。每一条指令操作一个或者多个数据项。因此，一条指令可以改变它前面指令的作用。

也许我们会问为什么程序必须由不同的指令集组成，答案是重用性。如今，计算机完成成千上万的任务，如果每一项任务的程序都是相对独立而且和其他程序之间没有任何公用段，编程将会变成意见很困难的事情。图灵模型和冯·诺依曼模型通过仔细地定义计算机可以使用的不同指令集，从而使得编程变得相对简单。程序员通过组合这些不同的指令创建任意数量的程序。每个程序可以是不同指令的不同组合。

## 3. 算法

要求程序包含一系列指令是的编程变得可能，但也带来了另外一些使用计算机方面的问题。程序员不仅要了解每条指令所完成的任务，还要知道怎样将这些指令结合起来完成一些特定的任务。对于一些不同的问题，程序员首先应该以循序渐进的方式来解决，接着尽量找到合适的指令来解决问题。这种按照步骤解决问题的方法就是所谓的算法。算法在计算机科学中起到了重要的作用。

## 4. 语言

在计算机时代的早期，只有一种称为机器语言的计算机语言。程序员依靠写指令的方式（使用位模式）来解决问题。但是随着程序越来越大，采用这种模式来编写很长的程序变得单调乏味。计算机科学家们研究出利用符号来表示位模式，就像人们在日常中用符号（单词）来代替一些常用的指令一样。当然，人们在日常生活中所用的一些符号与计算机中所用的符号不同。这样计算机语言的概念诞生了。自然语言是丰富的语言，并有许多正确组合单词的规则；相对而言，九四 u 安吉语言只有比较有限的服从后和单词。第 9 章将介绍计算机语言。

## 5. 软件工程

在冯·诺依曼模型中没有定义软件工程，软件工程是指结构化程序的设计和编写。今天，它不仅仅是用来描述完成某一任务的应用程序，还包括程序设计中所要严格遵循的原理和规则。我们所讨论的这些原理和规则综合起来就是第 10 章中要说的软件工程。

## 6. 操作系统

在计算机发展的演变过程中，科学家们发现有一系列指令对所有程序来说是公用的。如果这些指令只编写一次就可以用于所有程序，那么效率将会大大提高。这样，就出现了操作系统。计算机操作系统最初是为了程序访问计算机部件提供方便的一种管理程序。今天，操作系统所完成的工作远不止这些，具体的内容将在第 7 章介绍。

## Chapter 2

# 计算机组成

本章我们将讨论计算机的组成。讲解计算机是如何由三个子系统组成的。我们还介绍了简单假想的计算机，它能运行简单程序，完成基本的算术或逻辑运算。

### 2.1 引言

计算机的组成部件可以分为三大类：中央处理单元、主存储器和输入/输出子系统。接下来的三个部分将讨论这些子系统以及如何将这些子系统组成一台计算机。

### 2.2 中央处理器

中央处理器用于数据的运算。在大多数体系结构中，它有三个组成部分：算术逻辑单元、控制单元、寄存器组、快速存储定位。

#### 2.2.1 算术逻辑单元

算术逻辑单元对数据进行逻辑、位移和算术运算

##### 1. 逻辑运算

在第 4 章中，我们讨论了几种逻辑运算，如：非、与、或和抑或。这些运算把输入数据作为二进制位模式，运算的结果也是二进制模式。

##### 2. 位移运算

在第 4 章中，我们讨论了数据的两种位移运算：逻辑移位和算术移位运算。逻辑移位运算用来对二进制位模式进行向左或右的移位，而算术运算被应用于整数。它们的主要用途是用 2 除或乘一个整数。

##### 3. 算术运算

在第 4 章我们讨论了整数和实数上的一些算术运算，我们提到有些运算能被更高效地硬件实现。

### 2.2.2 寄存器

寄存器是用来存放临时数据的高速独立的存储单元。CPU 的运算离不开大量寄存器的使用。

#### 1. 数据寄存器

在过去，计算机只有几个数据寄存器用来存储输入数据和运算结果。现在，由于越来越多的复杂运算改由硬件设备实现，所以计算机在 CPU 中使用几十个寄存器来提高运算速度，并且需要一些寄存器来保存这些运算的中间结果。

#### 2. 指令寄存器

现在，计算机存储的不仅是数据，还有存储在内存中相对应的程序。CPU 的主要职责是：从内存中逐条取出指令，并将取出的指令存储在指令寄存器中，解释并执行指令。

#### 3. 程序计数器

CPU 中另一个通用寄存器是程序计数器。程序计数器中保存这当前正在执行的指令。当前的指令执行完后，计数器将自动加 1，指向下一条指令的内存地址。

### 2.2.3 控制单元

CPU 的第三个部分是控制单元，控制单元控制各个子系统的操作。控制是通过从控制单元到其他子系统的信号进行的。

## 2.3 主存储器

主存储器是计算机内的第二个子系统。它是存储单元的集合，每一个存储单元都有唯一的标识，称为地址。数据以成为字的位组形式在内存中传入和传出。字可以是 8 位、16 位、32 位，甚至有时是 64 位，如果字是 8 位，一般称为 1 字节。术语字节在计算机科学中使用相当普遍，因此有时成为 16 位位 2 字节，32 位位 4 字节。

### 2.3.1 地址空间

在存储器中存取每个字都需要有相应的标识符。尽管程序员使用命名的方式来区分，但在硬件层次上，每个字都是通过地址来标识的。所以在内存中表示的独立的地址单元的总数称为地址空间。

### 2.3.2 存储器的类型

主要由两种类型的存储器：RAM 和 ROM。

#### 1. RAM

随机存储器（RAM）是计算机中存储器的主要组成部分。在随机存储设备中，可以使用存储单元地址来随机存储一个数据项，而不需要存取位于它前面的所有数据项。该术语有时因 ROM 也能随机存取而与 ROM 混淆，用户可读性 RAM，即用户可以在 RAM 中写信息，之后可以方便地通过覆盖来查除原有信息。RAM 的另一个特点是易失性。当系统断电后，

信息将丢失。换句话说，当计算机断电后，存储在 RAM 中的信息将被删除。RAM 技术又可分为两大类：SRAM 和 DRAM。

(a) SRAM

静态 RAM (SRAM) 技术是用传统的触发器门电路来保存出局。这些门保持状态 (0 或 1)，也就是说当通电的时候数据始终存在，不需要刷新。SRAM 速度快，但是价格昂贵。

(b) DRAM

动态 RAM (DRAM) 技术使用电容器。如果电容器充电，则这时的状态是 1；如果放电则状态是 0。因为电容器会随时间而漏掉一部分电，所以内存单元需要周期性地刷新。DRAM 比较慢，但是价格便宜。

## 2. ROM

只读存储器 (ROM) 的内容是由制造商写进去的。用户只能读但不能写，它的优点是：非易失性：当切断电源后，数据不会丢失。通常用来存储那些关机后也不能丢失的程序或数据。例如，用 ROM 来存储那些在开机时运行的程序。

## 3. PROM

称为可编程只读存储器 (EPROM) 的一种 ROM。这种存储器在计算机发货时是空白的。计算机用户借助一些特殊的设备可以将程序存储在上面。当程序被存储后，它就会像 ROM 一样不能重写。也就是说，计算机使用者可以用它来存储一些特定的程序。

## 4. EEPROM

称为电可擦除可编程只读存储器 (EEPROM) 的一种 EPROM。对它的编程和擦除用电子脉冲即可，无须从计算机上拆下来。

### 2.3.3 存储器的层次结构

计算机用户需要许多存储器，尤其是速度快而且价格低廉的存储器。但这种要求并不总能得到满足。存取速度快的存储器通常都不便宜。因此需要寻找一种折中的方法。解决的办法是采用存储器的层次结构。该层次结构如下：

1. 当对速度要求苛刻时可以使用少量高速存储器。CPU 中的寄存器就是这种存储器。
2. 用适量的中速存储器来存储经常需要访问的数据。例如下面将要讨论的告诉缓冲存储器就属于这一类。
3. 用大量的低速存储器存储那些不经常访问的数据。

### 2.3.4 高速缓冲存储器

高速缓冲存储器的存取速度要比主存快，但是比 CPU 及其内部的寄存器要慢。高速缓冲存储器通常容量较小，且常被置于 CPU 和主存之间。

高速缓冲存储器在任何使劲按都含有主存的一部分内容的副本。当 CPU 要存取主存中的一个字的时候，将按以下步骤进行：

1. CPU 首先检查高速缓冲存储器。

2. 如果要存取的字存在，CPU 就将它复制；如果不存在，CPU 将从主存中复制一份从需要读取的字开始的数据块。该数据块将覆盖高速缓冲存储器中的内容。
3. CPU 存取高速缓冲存储器并赋值该字。

这种方式将提高运算的速度；如果字在高速缓冲存储器中，就立即存取它。如果字不再高速缓冲存储器中，字和整个数据块就会被复制到高速缓冲存储器中。因为很有可能 CPU 在下次存取时需要存取从上次存取的第一个字后续的字，所以高速缓冲存储器可以大大提高处理器的速度。

读者可能会奇怪为什么高速存储器尽管存储容量小效率却很高，这是由于 80-20 规则。据观察，通常计算机花费 80% 的时间来读取 20% 的数据。换句话说，相同的数据往往被存储多次。高速缓冲存储器，凭借其高速，可以存储这 20% 的数据而使存取至少快 80%。

## 2.4 输入/输出子系统

计算机的第三个子系统是称为输入/输出 (I/O) 子系统的一系列设备。这个子系统可以使计算机与外部通信，并在断电的情况下存储程序和数据。输入/输出设备可以分为两大类：非存储设备和存储设备。

### 2.4.1 非存储设备

非存储设备使得 CPU/内存可以与外界通信，但它们不能存储信息。

#### 1. 键盘和监视器

两个最常见的非存储输入/输出设备使键盘和监视器。键盘提供输入功能；监视器显示输出并同时响应键盘的输入。程序、命令和数据的输入或输出都是通过字符串进行的。字符则是通过字符集（如 ASCII 码）进行编码。此类中其他的设备由鼠标、操纵杆等。

#### 2. 打印机

打印机是一种用于产生永久性记录的输出设备。它是非存储设备，因为要打印的材料不能够直接由打印机输入计算机中，而且也不能再次利用，除非有人通过打字或扫描的方式再次输入计算机中。

### 2.4.2 存储设备

尽管存储设备被分为输入/输出设备，但它可以存储大量的信息以备后用。它们要比主存便宜得多，而且存储的信息也不易丢失（即使断电信息也不会丢失）。有时称它们为辅助存储设备，通常分为磁介质和光介质两种。

#### 1. 磁介质存储设备

磁介质存储设备使用磁性来存储数据。如果一点有磁性则表示 1，如果没有磁性则表示 0。

- 磁盘

磁盘是由一张一张的磁片叠加而成的。这些磁片由薄磁膜封装起来。磁盘是通过盘上每一个磁片的读/写磁头写磁介质表面来进行读取和存储的。

- 磁带

磁带的大小不一。通常的一种是用厚磁膜封装的半英寸塑料磁带。

2. 光存储设备

光存储设备是一种新技术,它使用光技术来存储和读取数据。在发明了 CD 后人们利用光存储技术来保持音频信息。现在,相同的技术(稍作改进)被用于存储计算机上的信息。使用这种技术的设备有只读光盘(CD-ROM)、可刻录光盘(CD-R)、可重写光盘(CD-RW)、数字多功能光盘(DVD)。

- CD-ROM

只读光盘使用于 CD 相同的技术(该技术最初是由飞利浦和索尼公司为录制音乐而研发的)。两者间位移的区别在于增强程度不同;CD-ROM 更健壮,而且纠错能力较强。

## 2.5 子系统的互连

前面的几节中已经介绍了在单个计算机上的三个子系统的主要特点。本节将介绍它们三者之间在内部是如何连接的,内部连接扮演很重要的角色,因为 i 洗脑洗需要在这三个子系统中交换。

### 2.5.1 CPU 和存储器的连接

CPU 和存储器之间通常由称为总线的三组线路连接在一起,它们分别是:数据总线、地址总线和控制总线。

1. 数据总线

数据总线是由多跟线组成,每一根线上每次传送 1 位的数据。线的数量取决于计算机的字的大小。例如,计算机的字是 32 位(4 字节),那么需要有 32 根线的数据总线,以便于同一时刻能够同时传送 32 位的字。

2. 地址总线

地址总线允许访问存储器种的某个字,地址总线的线数取决于存储空间的大小。如果存储器容量为  $2^n$  个字,那么地址总线一次需要传送  $n$  位的地址数据。因此它需要  $n$  根线。

3. 控制总线

控制总线负责在中央处理器和内存之间传送信息。例如,必须有一个代码从 CPU 发往内存,用于制定进行的是读操作还是写操作。控制总线的线数取决于计算机所需要的控制命令的总数。如果计算机有  $2^m$  条控制命令,那么控制总线就需要有  $m$  根,因为  $m$  位可以定义  $2^m$  个不同的操作。

### 2.5.2 I/O 设备的连接

输入/输出设备不能够直接与连接 CPU 和内存的总线相连。因为输入/输出设备的本质与 CPU 和内存的本质不同,输入/输出设备都是些机电、磁性或光学设备,而 CPU 和内存是电子设备。与 CPU 和内存相比,输入/输出设备的操作速度要慢很多。因此必须要有中介来处理这种差异,输入/输出设备是通过一种被称为输入/输出控制器或接口的器件连接到总线上的。每一个输入/输出设备都有一个特定的控制器。

有几种控制器至今还在使用,如今最常用的有 SCSI、火线、USB 和 HDMI。

### 2.5.3 输入/输出设备的寻址

通常 CPU 使用相同的总线在主存和输入/输出设备之间读写数据。

## 2.6 程序执行

当今，通用计算机使用称为程序的一系列指令来处理数据。计算机通过执行程序，将输入数据转换成输出数据。程序和数据都放在内存中。

本章最后将给出假象简单计算机执行程序的例子。

### 2.6.1 机器周期

CPU 利用重复的机器周期来执行程序种的指令，一步一条，从开始到结束。一个简化的周期包括 3 步：取指令、译码和执行。

### 2.6.2 输入/输出操作

## 2.7 不同的体系结构

在最近十几年中，计算机的体系结构和组织经历了许多变化。本书将讨论一些与我们前面讨论的简单计算机体系结构所不同的常见的体系结构和组织。

## 2.8 简单计算机

### 2.8.1 可重用性

与不可变成的计算器相比，计算机的一个优点是我们能反复使用相同的程序。沃恩能运行程序多次，每次键入不同的输入，得到不同的输出。

## Chapter 3

# 计算机网络和因特网

个人计算机的发展带动了商业、工业、科学和教育的巨大改变。网络也发生了类似的变革。技术的进步使得通信线路能传送更多、更快的信号。而不断发展的服务使得我们能够使用这些扩展的能力。计算机网络的研究导致了新技术的产生——在全球各个地方交换文本、音频和视频等数据，在任何时候快速、准确地下载或上载信息。

### 3.1 引言

虽然本章的目标是讨论因特网，一个将世界上几十亿台计算机互相连接的系统，我们对互联因特网的认识不应该是一个单独的网络，而是一个网络结合体，一个互连网络。因此，我们的旅程将从定义一个网络开始。然后我们将要展示如何通过网络连接来建造小型的互连网络。最终我们会展示因特网的结构并且在本章的以下部分开启研究因特网的大门。

#### 3.1.1 网络

网络是一系列可用于通信的设备相互连接构成的。在这个定义里面，一个设备可以是一台主机，比如一台大型计算机、台式机、便携式计算机、工作站、手机或安全系统。在这种定义中，设备也可以是一个连接设备，比如用来将一个网络与另一个网络相连接的路由器，一个将不同设备连接在一起的交换机，或者一个用于改变数据形式的调制器，等等。在一个网络中，这些设备都通过有线或无线传输媒介互相连接。当我们在家通过即插即用路由器连接两台计算机时，虽然规模很小，但已经建造了一个网络。

##### 1. 局域网

局域网通常是与单个办公室、建筑或校园内的几个主机相连的私有网络。基于机构的需求，一个局域网既可以简单到某人家庭办公室中的两台个人计算机和一台打印机，也可以扩大至一个公司范围，并包括音频和视频设备。在一个局域网中的每一台主机都作为这台主机在局域网中唯一定义的一个标识和一个地址。一台主机向另一台主机发送的数据包中包括源主机和目标主机的地址。

##### 2. 广域网

广域网也是通信设备互相连接构成的。但是广域网与局域网之间有一些差别。局域网的大小通常是受限的，跨越一个办公室、一座大楼或一个校园；



而广域网的地理跨度更大，可以跨越一个城镇、一个州、一个国家，甚至横跨世界。局域网将主机互联，广域网则将交换机、路由器或调制解调器之类的连接设备互连。通常，局域网为机构私有，广域网则由通信公司创建并运营，并且租给使用它的机构。我们可以看到广域网的两种截然不同的案例：点对点广域网和交换广域网。

点对点广域网是通过传输媒介连接两个通信设备的网络。

交换广域网是一个有至少两个端的网络。就像我们很快就会看到的那样，交换广域网用于今天全球通信的骨干网。我们也可以这么说，交换广域网是一个点对点广域网通过开关连接产生的结合体。

### 3. 互联网络

现在很难看见独立存在的局域网或广域网，它们现在都是互相连接的。当两个或多个网络相互连接时，它们构成一个互联网络，或者说网际网。

#### 3.1.2 因特网

正如我们之前讨论过的，一个网际网是两个或多个可以互相通信的网络。最值得注意的网际网是因特网，它由成千上万个互连的网络组成。图展示了因特网的一个概念图像。

这幅图将因特网展示为几个骨干网、供应商网络和客户网络。在顶层，骨干网位通信公司所拥有，这些骨干网通过一些复杂的交换系统相互连接。我们把这些交换系统称为网络对等交汇点（peering point）。在第二层，有一些规模较小的网络，这些网络成为供应商网络，他妈呢付费使用骨干网上的一些服务。这些供应商网络与骨干网相连，有时也连接其他应用商网络。在因特网的边缘有一些真正使用基于因特网的服务的网络，这些网络是客户网络，它们向供应商网络付费来得到服务。

骨干网和供应商网络也被称为因特网服务供应商（ISP），骨干网通常被称为国际因特网服务提供商，供应商网络则被称为国内或地域性因特网服务提供商。

#### 3.1.3 硬件和软件

我们已经展示了因特网通过连接设备将大大小小的网络相互交织在一起构成的基本结构。然而，如果仅仅将这些部分连接在一起，很明显什么都不会发生。为了产生沟通，即需要硬件也需要软件设备。这就像当进行一个复杂的计算时，我们同时需要计算机和程序。

#### 3.1.4 协议分层

当讨论因特网时，有一个词我们总是会听见，这个词就是协议。协议定义了发送器、接收器以及所有中间设备必须遵守以保证有效地通信的规则。简单的通信可能只需要一条简单的协议，当通信变得复杂时，可能需要将任务分配到不同的协议层中，在这种情况下，我们在每一个协议层都需要一个协议，或者协议分层。

##### 1. 情景

为了更好地理解协议分层的必要，首先我们开发一个简单的场景。假定 Ann 和 Maria 是有很多共同想法的邻居，她们每次都会为了一个何时退休的计划互相见面和进行沟通。突然，Ann 所在的公司为她提供一个升职

的机会，但是同时要求她搬到离 Maria 很远的一个城市中的分部去住。由于她们想出了一个具有创新型的计划——在退休后开始做新的生意，这两个朋友仍然向继续它们的通信并且就这个新计划交换想法。她们决定通过到邮局使用普通邮件通信来继续她们的对话，但是如果邮件被拦截了，她们不想别人知晓它们的想法。她们在邮件的加密/解密方式上达成了一致，寄信人对邮件进行加密，这样对于入侵者而言邮件是无法阅读的；同时收件人对邮件进行解密以得到原始邮件。第 16 章将讨论加密/解密方法，但是现在我们可以推测 Maria 和 Ann 使用了其中一种使没有密钥的人很难那对邮件进行解密的技术。这样，可以把 Maria 和 Ann 之间的通信分为三个协议层，如图所示。假设 Ann 和 Maria 都各自拥有三台机器（或机器人）来完成每一个协议层的任务。

让我们假设 Maria 将第一封邮件寄给 Ann。Maria 假设第三协议层机器是正在听她说话的 Ann 并对其说话，第三层协议层的机器听她说话并且创作出明文，并被传到第二层的机器，第二层的机器对文本进行加密，将它创作成密文，并传送到第一协议层的机器。这个第一协议层的机器，也许是一个机器人，把密文装进信封，加上发信人地址，然后将信寄出。

在 Ann 这边，第一协议层的机器从 Ann 的邮件中取出邮件并通过发信人地址找出来自 Maria 的这一封。它从信封中取出密文并传递给第二协议层的机器，第二协议层的机器对密文进行解密，创作出明文并传递给第三协议层的机器，第三协议层的机器接受明文并且将它读出来，就像 Maria 在说话一样。

协议分层使我们的大任务简化成几个更小、更简单的任务。例如，在上述中，我们可以只用一台机器来完成三台机器全部的工作，但是，如果 Maria 和 Ann 觉得机器完成的现有加密或解密无法保证她们保密，他们呢需要替换整个机器。她现在的情况下，她们只需要替换第二协议层的机器就足够了，另外两层可以保持不变；这称为模块化。在这里模块化指的是独立的协议层。一个协议层（模块）可以定义为一个具有输入和输出而不需要考虑输入时如何变成输出的黑匣子。当两台机器提供相同输入得到相同输出时，它们就可以相互替换。

协议分层的一个优势就是可以将服务和其实施分开来。每层使用更底层的服务，并向较高一层提供服务；并且我们不需要考虑该层时如何实施的。例如，Maria 也许可以自己完成第一协议层的工作而决定不买第一协议层的计算机（机器人）。只要 Maria 可以完成第一协议层提供的工作，这个通信系统就可以正常的双向运行。

协议层的另一个优势，虽然可在简单的例子中无法看出，但是在我们讨论因特网中的协议分层时会表现出来，因为通信系统往往不仅仅具有两个端系统，还有一些只需要几个协议层而不是所有协议层的中间系统。如果我们不使用协议分层，整个系统会变得更复杂，因为那样我们得把每一个中间系统都变得和端系统一样复杂。

协议分层有什么劣势吗？也许有人会认为单一协议层可以是整个工作变简单，而且没有每个协议层都使用低一级协议层的服务并向高一级协议层提供服务的必要。但是，就像之前提到的那样，一旦密码被破解，她们每一人得将整个机器换成新的而不是仅仅更换第二协议层。

## 2. 协议分层的原则

我们讨论一下分层的原则。第一原则规定，如果我们想达到双向通信，我们需要保证每一个协议层都可以进行两个对立且方向相反的工作。

### 3. 逻辑连接

在遵循以上两条原则之后，我们可以如图所示理解每个协议层之间的逻辑连接。这说明了层与层之间通信的存在。Maria 和 Ann 可以认为她们能够发送该协议层创作的对象是基于每层的逻辑（假想的）连接。逻辑连接的概念会帮助我们更好地理解在数据通信和建立数据关系的网络中遇到的分层工作。

### 3.1.5 TCP/IP 协议族

通过第二个场景我们了解了协议分层和协议层之间的逻辑通信，这样就可以介绍传输控制协议/网际协议（TCP/IP）。如今因特网中使用的协议集（一组通过不同分层进行组织的协议）被称为 TCP/IP 协议族。TCP/IP 协议族是一个分层协议，它由提供特定功能的交互式模块组成。分层这个术语说明每一个高层协议都基于一个或多个低层协议提供的服务。TCP/IP 协议族被定义成图所示布局的软件层。

#### 1. 分层架构

为了展示 TCP/IP 协议族中的分层是如何在两台主机通信中作用的，我们假设要使用的一套痛惜是在一个由 3 个 LAN（连接）构成的小网络，且链路层开关与每个 LAN 相连。同时我们假定这些链接都与同一个路由器相连。

#### 2. 地址和数据包名称

在因特网中，另外两个和协议分层有关的概念很值得一提，它们是地址和数据包名称。就像我们在之前讨论过的那样，在这个模型中，我们由层组之间的逻辑通信。任何涉及两步校验的通信需要两个地址：源地址和目标地址。虽然看上去我们需要 5 组地址，每个协议层一组，但是正常情况下我们只有 4 组，因为物理层不需要地址。这是由于物理层数据交换的单位是位，这使它无法得到地址。

在应用层，我们通常使用名称来第一提供服务的站点，比如 `someorg.com`，或者邮箱地址，比如 `somebody@coldmail.com`。在传输层，地址被称为端口号，这些端口号的作用是在源和目的之间定义应用层程序。端口号的作用是通过各程序的本地地址来辨别多个同时运行的本地程序。在网络层，这些地址在整个因特网范围下是全球化的，网络层的地址独一无二得顶一个了该设备与因特网的连接。链路层地址，有时称为 MAC 地址，是在本地定义的地址，每一个链路层地址在计算机网络或广域网中定义一个特定的主机或者路由器。

## 3.2 应用层

在简单讨论网络、互联网和因特网之后，我们准备好了来进行一些关于 TCP/IP 协议族中每一层的讨论。我们从第 5 层像第 1 层讨论。

TCP/IP 协议的第 5 层叫做应用层。应用层像用户提供服务。通信由逻辑连接提供，也就是说，假设两个应用层通过之间假想的直接连接发送和接受消息。

### 3.2.1 提供服务

应用层与其他层不同的地方在于，它是协议族中的最高层。在这层中的协议不向其他协议提供服务，它们只接收在传输层的协议提供的服务。这意味着

该层的协议可以轻易去除。只要新的协议可以使用传输层中任意一个协议提供的服务，这个新的协议就可以添加到应用层上。

### 3.2.2 应用层模式

在网络的生命周期中，应用程序发展出了两种模式来解答这个问题：客户机-服务器模式和端到端模式。这里简单介绍这两种模式。

## Chapter 4

# 操作系统

### 4.1 引言

计算机系统是由两个主题部分组成：硬件和软件。硬件是计算机的物理设备。软件则是使得硬件能够正常工作的程序的集合。计算机软件分为两大类：操作系统和应用程序。应用程序使用计算机硬件来解决用户的问题。另一方面，操作系统则控制计算机用户对硬件的访问。

#### 4.1.1 操作系统

操作系统是一个非常复杂的系统，因此很难给予它一个普遍认同的简单定义。在这里列举一些常见的定义：

1. 操作系统是介于计算机硬件和用户之间的接口。
2. 操作系统时一种用来使得其他程序更加方便有效运行的程序。
3. 操作系统作为通用管理程序管理着计算机系统中每个部件的活动，并确保计算机系统硬件和软件资源能够更加有效地使用。当资源使用出现冲突时，操作系统应能及时处理，排除冲突。

操作系统是计算机硬件和用户的一个接口，它使得其他程序更加方便有效地运行，并能方便地对计算机硬件和软件资源进行访问。

操作系统的两个主要设计目标：有效地使用硬件；容易地使用资源。

#### 4.1.2 自举过程

基于上面的定义，操作系统位其他程序提供支持。例如，它负责把其他程序装入内存以便运行。但是，操作系统本身也是程序，它需要被装入内存和运行，这个困难如何解决呢？如果使用 ROM 技术把操作系统存储在内存中，这个问题就能解决。CPU 的程序计数器可以被设置到这个 ROM 的开始处。当计算机被加电时，CPU 从 ROM 中读取指令，执行它们。但这种解决方案是非常低效的，因为内存的很大一部分需要由 ROM 构成，而不能被其他程序使用。如今的技术是仅需要分配小部分的内存给部分操作系统。

如今使用的解决方案采用两个阶段过程。很小一部分内存用 ROM 构成，其中存有成为自举程序的小程序。当计算机被加电时，CPU 计数器被设置位自举程序的第一条指令，并执行程序中的指令，操作系统就被执行。

## 4.2 演化

操作系统已经经历了很长的一段发展历程，我们将在下面加以总结。

### 4.2.1 批处理系统

批处理操作系统设计于 20 世纪 60 年代，目的是控制大型计算机。当时计算机十分庞大。用穿孔卡片进行输入数据，用行式打印机输出结果，用磁带设备作为辅助存储介质。

### 4.2.2 分时系统

每个运行的程序叫一个作业。想要运行程序的程序员通过穿孔卡片将程序和数据输入计算机，并向控制器发出作业请求。穿孔卡片由操作员处理。如果程序运行成功，打印结果将传给程序员，如果不成功，则报错。

这个时代的操作系统非常简单：它们只保证计算机所有资源被从一个作业转换到另一个作业。

为了有效的使用计算机资源，多道程序的概念被引进。它可以将多个作业同时装进内存，并且当且仅当资源可用时分配给需要它的作业。例如，当一个程序正使用输入/输出设备时，CPU 则处于空闲状态，并可以供其他程序使用。我们将在本章后面详细介绍多道操作系统。

多道程序带类了分时的概念：资源可以被不同的作业分享。每个作业可以分到一段时间来使用资源。因为计算机运行速度很快，所以分时系统对于用户是隐藏的，每个用户都感觉整个系统在为自己服务。

最终利用分时技术的多道程序极大地改进了计算机的使用效率。但是，它们需要一个更复杂的操作系统，它必须可以调度：给不同的程序分配资源并决定哪一个程序什么时候使用哪一个资源。在这个时代中用户和计算机的关系也改变了。用户可以直接与操作系统进行交互而不必通过操作员。一个新的术语也随之产生：进程。一个作业是一个要运行的程序，一个进程则是在内存中等待分配资源的程序。

### 4.2.3 个人系统

个人计算机产生后，需要有一类适合这类计算机的操作系统。于是，单用户操作系统就应运动而生了。

### 4.2.4 并行式系统

人们对更快和更有效的需求导致了并行系统的设计：在统一计算机中安装多个 CPU，每个 CPU 可以处理一个程序或一个程序的一部分。意味着很多任务可以并行地处理而不是在串行处理。当然这种操作系统要比单 CPU 的操作系统复杂得多。

### 4.2.5 分布式系统

网络化和网络互联的发展，扩大了操作系统的内涵。一个以往必须在一台计算机上运行的作业现在可以由远隔千里的多台计算机共同完成。程序可以在一台计算机上运行一部分而在另一台计算机上运行另一部分，只要它们通过网络连接即可。资源可以是分布式的，一个程序需要的文件可能分布在世界的不同地方。分布式系统结合了以往系统的特点和新的功能，例如安全控制。

### 4.2.6 实时系统

实时系统是指在特定时间限制内完成任务。它们被用在实时应用程序中，这些应用程序监控、响应或控制外部过程或环境。在交通控制、病人监控或军事控制系统中可以找到实时系统的例子。

## 4.3 组成部分

现在的操作系统十分复杂，它必须可以管理系统中的不同资源。它像是一个有多个上层部门经理的管理机构，每个部门经理负责自己的部门管理，并且互相协调。现代操作系统至少具有以下 4 中功能：存储管理、进程管理、设备管理、文件管理。就像很多组织由一个部门不归任何经理管理一样，操作系统也有这样一个部分，被称为用户界面或命令解释程序，它负责操作系统与外界通信。

### 4.3.1 用户界面

每个操作系统都有用户界面，即指用来接收用户的输入并向操作系统解释这些请求的程序。一些操作系统的用户界面，被称为命令解释程序。在其他操作系统中，则被称为窗口，以指明它是一个由菜单驱动的并有着 GUI 的部件。

### 4.3.2 内存管理器

现在计算机操作系统的一个重要职责是内存管理。计算机中存储器的容量近年来得到激增，同样所处理的程序和数据也越来越大。内存分配必须进行管理一面“内存溢出”的错误。操作系统按照存储管理可以分为两大类：单道程序和多道程序。

#### 1. 单道程序

单道程序属于过去，但它还是值得学习，因为它有助于理解多道程序。在单道程序中，大多数内存用来转载单一的程序，仅仅一小部分用来装载操作系统。在这种配置下，整个程序装入内存运行，运行结束后，程序区域由其他程序取代。

这里内存管理器的工作是简单明了的，即将程序载入内存、运行它、再装入新程序。但是，在技术方面仍然有很多问题：

程序必须能够载入内存。如果内存容量比程序小，程序就无法运行。

当一个程序正在运行时，其他程序不能运行。一个程序在执行过程中经常需要从输入设备得到数据，并且把数据发送至输出设备。但输入/输出设备的速度远远小于 CPU，所以当输入/输出设备运行时，CPU 处于空闲状态。而此时由于其他程序不再内存中，CPU 不能为其服务。这种情况下 CPU 和内存的使用效率很低。

#### 2. 多道程序

在多道程序下，同一时刻可以装载多个程序并且能够同时被执行。CPU 轮流为其服务。

从 20 世纪 60 年代开始，多道程序已经经过了一系列的改进。

有两种技术属于非交换范畴，这意味着程序在运行期间始终驻留在内存中。另外两种技术属于交换范畴。也就是说，在运行过程中，程序可以在内存和硬盘之间多次交换数据。

多道程序的第一种技术称为分区调度。在这种模式中，内存被分为不定长的几个分区。那个部分或分区保存一个程序。CPU 在各个程序之间交替服务。它由一个程序开始，执行一些指令，直到有输入/输出操作或者分配给程序的时限到达为止。CPU 保存最近使用的指令所分配的内存地址后转入下一个程序。对下一个程序采用同样的步骤反复执行下去。当所有程序服务完毕后，再转回第一个程序。当然，CPU 可以进行优先级管理，用于控制分配给每个 CPU 时间。

在这种技术下，每个程序完全载入内存，并占用连续的地址。分区调度改进了 CPU 的使用效率，但仍然有以下一些问题：

分区的大小必须由内存管理器预先决定。如果分区小了，由的程序就不能载入内存。如果分区大了，就会出现空闲区。

即使分区在刚开始时比较合适，但随着新程序的交换载入内存后有可能出现空闲区。

当空闲区过多时，内存管理器能够紧缩分区并删除空闲区和创建新区，但这将会增加系统额外开销。

分页调度提高了分区调度的效率。在分页调度下，内存被分成大小相等的若干部分，称为帧。程序被分为大小相等的部分，称为页。页和帧的大小通常时一样的，并且与系统用于从存储设备中提取信息的块的大小相同。

页被载入内存中的帧。如果一个程序有 3 页，它就在内存中占用 3 个帧。在这种技术下，程序在内存中不必时连续的：来嗯个连续的页可以占哟哦嗯内存中不连续的两个帧。分页调度对分区调度的优势在于，一个需要 6 个帧的程序可以代替两个各占有不连续的 3 个帧的程序。而不必等到有 6 个连续的帧出现后再载入内存。

分页调度在一定程度上提高了效率，但整个程序仍需要在运行前全部载入内存。这意味着只有 4 个不连续帧时，一个需要 6 个空闲帧的程序不能载入。

分页调度不需要程序转载在连续的内存中，但仍需要程序载入内存中运行。请求分页调度该表了后一种限制。在请求分页调度中，程序被分为页，但是页可以依次载入内存、运行，然后被另一个页替代。换句话说，内存可以同时载入多个程序的页。此外，来自同一个程序的连续页可以不必载入同一个帧，一个页可以载入任何空闲帧。

类此于分页调度的技术时分段调度。在分页调度中，不像程序员以模块来考虑程序，程序实际是分为大小相同的页。你将在后面的章节中看到，程序通称由主程序和子程序组成，在请求分段调度中，程序将按照程序员的角度划分成段，它们载入内存中、执行，然后被来自统一程序或其他程序的模块代替。

请求分页和分段调度结合了两者的有点以提高系统效率。一个段也许太大而不能载入内存中的空闲区。内存可以分成很多帧，一个模块可以份额成很多页，一次装入内存运行。

### 3. 虚拟内存

请求分页调度和请求分段调度移位这当程序运行时，一部分程序驻留在内存中个一部分则放在硬盘上。这就意味着...。虚拟内存。



### 4.3.3 进程管理器

操作系统的第二个功能是进程管理，在介绍该概念之前，我们先定义一些术语。

#### 1. 程序、作业和进程

现代操作系统关于指令集有三个术语：程序、作业和进程。尽管这些术语比较模糊，并且不同操作系统对于它们的定义不一致，我们还是可以作出非正式的定义。

- 程序是由程序员编写的一组稳定的指令，存在磁盘上，它可能会也可能不会成为作业。
- 从一个程序被选中执行，到其运行结束并再次成为一个程序的这段过程中，该程序称为作业。在整个过程中，作业可能会或不会被执行，或者驻留在磁盘上等待调入内存，或者在内存中等待 CPU 执行，或者驻留在硬盘或内存中等待一个输入/输出事件，或者在内存中等待知道被 CPU 运行。在所有这些情况下程序才称为作业。当一个作业执行完毕，它又变成程序代码并再次留在硬盘中，操作系统不再支配该程序。
- 进程是一个运行中的程序。该程序开始运行但还未结束。换句话说，进程是一个驻留在内存中运行的作业，它是从众多等待作业中选取出来并转入内存中的作业。一个进程可以处于运行或者的等待 CPU 调用。只要作业装入内存就称为一个进程。

#### 2. 状态图

当明白程序怎样变成作业和作业怎样变成进程时，程序、作业、进程的关系也就很明显了。状态图显示了每个实体的不同状态。

#### 3. 调度器

将一个作业或者进程从一个状态改变为另一个状态，进程管理器使用了两个调度器：作业调度器和进程调度器。

- 作业调度器  
作业调度器将一个作业从保持状态转入就绪状态，或是从运行状态转入终止状态。换句话说，作业调度其负责从作业中创建一个进程或终止一个进程。
- 进程调度器  
进程调度器将一个进程从一个状态转入另一个状态。当一个进程等待某事件发生时，它使这一进程从运行状态进入等待状态。当事件发生时，进程将从等待状态进入就绪状态。
- 其他调度器  
一些操作系统使用其他的调度器使进程之间的转换更为有效。

#### 4. 列队

#### 5. 进程同步

#### 6. 设备管理器

设备管理器负责访问输入/输出设备。在计算机系统中输入/输出设备存在这数量和速度的限制。由于这些设备与 CPU 和内存比起来速度要慢很多，

多以当一个进程访问输入/输出设备时，在该段时间内这些设备对其他进程而言是不可用的。设备管理器负责让输入/输出设备使用起来更有效。

对设备管理器的细节的讨论需要掌握有关操作系统原理的高级知识，这些都不在本书讨论之列。但是我们可以在这里简要地列出设备管理器的功能。

#### 4.3.4 文件管理器

现今的操作系统使用文件管理器来控制对文件的访问。对文件管理器细节的讨论同样需要掌握有关操作系统原理和文件访问的高度概念，这些超出了本出的讨论范围。我们将在地 13 章讨论一些有关文件访问的问题，但是对于了解文件管理器实际的操作还是不够。下面简述一下文件管理器的功能：

- 文件管理器控制文件的访问。只有那些获得允许的应用程序才能够访问，访问方式也不同。例如，一个进程也许可以读取文件，但却不允许写操作。另一个进程也许被允许执行文件和进程，但却不允许读取文件的内容。
- 文件管理器管理文件的创建、删除和修改。
- 文件管理器可以给文件命名。
- 文件管理器管理文件的存储：怎样存储，存在哪里等。
- 文件管理器负责归档和备份。

### 4.4 主流操作系统

在这一节，我们将介绍一些常用的操作系统，以促进将来的学习。我们选择三种计算机用户熟悉的操作系统：UNIX、Linux 和 Windows。

#### 4.4.1 UNIX

UNIX 是由贝尔实验室的计算机科学研究小组的 Thomson 和 Ritchie 在 1969 年首先开发出来的。从那时起，UNIX 经历了许多版本。它是一个在程序员和计算机科学家中较为流行的操作系统。它可以不经较大的改动而方便地从一个平台移植到另一个平台。原因是它主要是由 C 语言编写的。第二，UNIX 拥有一套功能强大的工具，它们能够组合起来去解决许多问题，而这一工作在其他操作系统则需要通过编程来完成。第三，它具有设备无关性，因为操作系统本身就包含了设备驱动程序，这意味着它可以方便地配置来运行任何设备。

UNIX 是多用户、多道程序、可移植的操作系统，它被设计来方便编程、文本处理、通信和其他许多希望操作系统来完成的任务。它包含几百个简单、单一目的的函数，这些函数能组合起来完成任何可以想象的处理任务。它的灵活性通过它可以在三种不同的计算机环境中而得到证明，这三种环境为：单机个人环境、分时系统和客户、服务器系统。

UNIX 由 4 个主要部分构成：内核、命令解释器、一组标准工具和应用程序。这些组成部分显示在图中。

内核是 UNIX 系统的核心。它包含操作系统最基本的部分：内存管理、进程管理、设备管理和文件管理。系统所有其他部分均调用内核来执行这些服务。

命令解释器是 UNIX 中用户最可见的部分。它接收和解释用户输入的命令。在许多方面，这使它成为 UNIX 结构的最重要的组成部分。它肯定也是用户最知道的部分。为了在系统做任何事情，我们必须向命令解释器输入命令。如果

命令需要一个工具，命令解释器将请求内核执行该工具。如果命令需要一个应用程序，命令解释器需要内核运行它。有些操作系统有几种不同的命令解释器。

UNIX 中有几百个工具。工具是 UNIX 标准程序，它为用户提供支撑过程。常用的三个工具是：文本编辑器、搜索程序和排序程序。

许多系统工具实际上复杂的应用程序。例如，UNIX 的电子邮件系统被看成一个工具，就像三种常见的文本编辑器：vi、emacs 和 pico。所有这 4 个工具本身都是很大的系统。其他工具是简短函数。例如，list 工具显示磁盘目录中的文件。

UNIX 的应用是指一些程序，它们不是操作系统发布中的标准部分。它们是由系统管理员、专职程序员或用户编写的，提供了对系统的扩展能力。事实上，许多标准工具自多年前都是作为应用出现的，后来被证明非常有用，现在就成了系统的一部分。

#### 4.4.2 Linux

在 1991 年，芬兰 Helsinki 大学的学生 Linux Torvalds 开发了一个新的操作系统，这就是如今所知的 Linux。初始内核如今成长为全面的操作系统。1997 年发布的 Linux2.0 内核成为商业操作系统，它具有传统 UNIX 的所有特性。

##### 1. 组成 Linux 有下列组成部分。

- 内核  
内核负责处理所有属于内核的职责。
- 系统库  
系统库含有一组被应用程序使用的函数，用于与内核交互。
- 系统工具  
系统工具是使用系统库提供的服务，执行管理任务的各个程序。

##### 2. 网络功能

Linux 支出第 6 章中讨论的标准因特网协议。它支持三层：套接字接口、协议驱动和网络设备驱动。

##### 3. 安全

Linux 的安全机制提供了传统上为 UNIX 定义的安全特性。如身份验证和访问控制。

#### 4.4.3 Windows

20 世纪 80 年代后期，在 Dave Cutler 的领导下，微软开始开发代替 MS-DOS 的新的单用户操作系统。Windows 就是结果。后来又有几个 Windows 的版本，我们统称这些版本为 Windows。

##### 1. 设计目标

微软发布的设计目标是：可扩展性、可移植性、可靠性、兼容性和性能。

- 可扩展性  
Windows 被设计成具有多层的模块化结构。意图是允许高层随时间而变化，而不影响底层。

- 可移植性

像 UNIX 一样, Windows 是用 C 或 C++ 编写的, 这个语言是独立于它所运行的计算机的机器语言的。

- 可靠性

Windows 被设计成能处理包括防止恶意软件的错误条件。

- 兼容性

Windows 被设计对运行成能运行为其他操作系统编写的程序, 或 Windows 早期版本。

- 性能

Windows 被设计成对运行在操作系统顶部的应用程序, 具有快速响应时间。

## 2. 体系结构

Windows 使用层次结构。

## Chapter 5

# 算法

### 5.1 概念

本节将给出算法的非正式定义，然后通过一个例子来详细讲述算法的概念。

#### 5.1.1 非正式定义

算法一种非正式定义如下：算法是一种逐步解决问题或完成任务的方法。

按照这种定义，算法完全独立于计算机体系。更特别的是，还应该记住算法接受一组输入数据，同时产生一组输出数据。

下面用一个例子来对这种简单的定义进行分析。我们要生成从一组正整数中找到最大整数的一个算法。这个算法应该能从一组任意整数中找出其最大值。这个算法必须具有通用性并与整数的个数无关。

很明显，要完成从许多整数中找到最大值的这个任务不可能只要一步完成。算法必须一个个地测试每一个整数。

要解决这个问题，可以用一种直接的方法。先用一组少量的整数，然后将这种解决方法扩大到任意多的整数。其实对 5 个整数所采取的解决方法的原理和约束条件与对 1 000 个或 1 000 000 个整数采取的是一样的。可以假设，即使是 5 个整数的例子，算法也必须一个接一个地处理那些整数。看到第一个整数，并不知道剩下的整数的值。等处理完第一个整数，算法才开始处理第二个整数，依次进行。

我们称这个算法为求最大值算法。每个算法都有自己不同于其他算法的名字。这个算法接收一组 5 个整数作为输入，然后输出其中的最大值。

#### 5.1.2 细化

细化。为了使算法能在所有的程序中应用，还需要进行细化。现在有两个问题，首先，第一步中的动作与其他步骤中不一样。其次第二步到第五步中的程序描述语言不通。我们只要很简单地改进一下算法就可以解决以上两个问题。把第二步到第五步的程序段都写成“如果当前整数大于 Largest，那么当前整数就称为 Largest”。第一步不通于其他步是因为那时 Largest 还没有初始化。如果开始就把 Largest 初始化为负无穷，那么第一步就可写成和其他步一样，所以。增加一个新的步骤，可称为第 0 步，也就是表明它要在处理任何其他整数之前完成。

### 5.1.3 泛化

可以把这个算法泛化吗？假设要从  $n$  个正整数中找到最大值， $n$  的值可能是 1 000 或 1 000 000，或者更多。当然，可以按照图所示那样重复每一步。

## 5.2 三种结构

计算机专家为结构化程序或算法定义了三种结构。这种想法认为程序必定是由顺序、判断（选择）和循环这三种结构组成。已经证实其他结构都是不必要的。仅仅使用这三种结构就可以是程序或算法容易理解、调试或修改。

### 5.2.1 顺序

第一种结构称为顺序结构。算法都是指令序列。它可以是一简单指令或其他两种结构之一。

### 5.2.2 判断

有些问题只用简单的指令序列是不能够解决的。有时候需要检测一个条件是否满足。假如测试的结果为真，则可以继续顺序往下执行指令；假如结果为假，程序将从另外一个顺序结构的指令继续执行。这就是所谓的判断（选择）结构。

### 5.2.3 循环

有些问题中，相同指令序列需要重复。可以用重复或循环结构来解决这个问题。

## 5.3 算法的表示

到目前为止，我们已经使用图来表示算法的基本概念。在最近几十年中，还出现了其他几种用来表示算法的工具。这里讲介绍 UML 和伪代码这两种工具。

### 5.3.1 UML

统一建模语言（UML）是算法的图形表示法。它使用“大图”的形式掩盖了算法的所有细节，它只显示算法从开始到结束的整个流程。

### 5.3.2 伪代码

伪代码是算法的一种类似英语的表示法。现在还没有伪代码的标准。有些人使用得过细，有些人则使用得过粗。有些人用一种很像英语的代码，有些则用和 Pascal 编程语言相似的语法。

## 5.4 更正式的定义

既然我们讨论了算法的概念并且给出了它的表示，下面给出算法更为正式的定义。

### 5.4.1 定义良好

算法必须是一组定义良好且有序的指令集合。

### 5.4.2 明确步骤

算法的每一步必须有清晰、明白的定义。如某一步是将两个整数相加，那么必须定义相加的两个整数和加法符号，相同的符号不能再某处用作假发符号，而再其他故方用作乘法符号。

### 5.4.3 产生结果

算法必须产生结果，否则该算法就没有意义。结果可以是返回给调用算法的数据或其他效果（如打印）。

### 5.4.4 在有限的时间内终止

算法必须能够终止（停机）。如果不能，说明不是算法。

## 5.5 基本算法

有一些算法在计算机科学中应用的非常普遍，我们称之为“基本”算法。这里讲讨论一些最常用的算法。讨论只是概括性的，具体的实现则取决于采用何种语言。

### 5.5.1 求和

计算机科学中经常用到的一种算法是求和。你可以容易地实现两个或三个整数的相加，但是怎样才能实现一系列整数相加呢？答案很简单：在循环中使用加法操作。

求和算法可以分为三个逻辑部分：

1. 将和（sum）初始化。
2. 循环，在每次迭代总将一个新数加到和（sum）上。
3. 推出循环后返回结果。

### 5.5.2 乘积

另一个常用算法是求出一系列整数的乘积。方法也很简单：在循环中使用乘法操作。乘法算法有三个逻辑部分：

1. 将乘积（product）初始化。
2. 循环，在每次迭代中将一个新数与乘积相乘。
3. 推出循环后返回结果。

### 5.5.3 最大和最小

在本章开头讨论了求一组整数中最大值的算法。它的思想是通过一个判断结构求到两个数中的较大值。如果是把这个结构放在循环中，就可以求出一组数中的最大值。

求一组整数中的最小值和上面的方法相似，只有两个小小的不同。

### 5.5.4 排序

计算机科学总的一个最普遍应用的是排序，即根据数据的值对它们进行排列。人们的周围充满了数据，如果这些数据都是无序的，可能会花很多时间去查找一条简单信息。

本节将介绍三种排序算法：选择排序、冒泡排序、插入排序。这三种方法是当今计算机科学中使用的快速排序的基础。

#### 1. 选择排序

在选择排序中，数字列表可分为两个字列表，它们通过假想的一堵墙分开。求未排序子列表中最小的元素并把它和未排序子列表中第一个元素进行交换，进过每次选择和交换，两个子列表中假想的这堵墙向前移动一个元素，这样每次排序列表中将增加一个元素而未排序列表中将减少一个元素，每次把一个元素从未排序列表一道已排序列表就完成了一轮排序。一个还有  $n$  个元素的数字列表需要  $n - 1$  轮排序来完成数据的重新排序。

#### 2. 冒泡排序

## 5.6 子算法

根据在 8.2 节描述的三种编程结构，可以为每一个可解的问题创建算法。结构化编程的原则要求将算法分成几个单元，称为子算法。每个子算法依次又分为更小的子算法。

使用子算法至少有两个优点：

- 程序更容易理解。
- 子算法可在主算法中不同地方调用，而无需重写。

程序员使用的另一个编程工具就是结构图。结构图是一种高级设计工具，它显示了算法和子算法之间的关系。它一般在设计阶段使用，而不是在编程阶段。

## 5.7 递归

通常，有两种途径用于编写解决问题的算法。一种使用迭代，另一种使用递归。递归是算法的自我调用过程。

### 5.7.1 迭代的定义

学习一个简单的例子，考虑一个阶乘的计算。阶乘的因子是从 1 到该数的整数。迭代的定义如图所示。如果算法的定义不涉及算法本身，则算法是迭代的。



### 5.7.2 递归的定义

每一个算法出现在它本身定义中，该算法就是递归定义的。