x86 汇编语言
从实模式到保护模式

李忠 王晓波 余洁◎著

- 你想知道汇编语言到底是什么吗？
- 你想探究x86处理器的内部原理吗？
- 你想从侧面理解操作系统的工作原理，并尝试打造一个自己的系统吗？
- 这本书将会带你学习和掌握汇编语言程序设计的基本方法，了解汇编语言和处理器、计算机系统以及操作系统之间的关系。

電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

x86 汇编语言
从实模式到保护模式

李 忠 王晓波 余 洁 著

電子工業出版社
Publishing House of Electronics Industry
北京 • BEIJING

内 容 简 介

本书采用开源的 NASM 汇编语言编译器和 VirtualBox 虚拟机软件，以个人计算机广泛采用的 Intel 处理器为基础，详细讲解了 Intel 处理器的指令系统和工作模式，以大量的代码演示了 16 / 32 / 64 位软件的开发方法，集中介绍处理器的 16 位实模式和 32 位保护模式，以及基本的指令系统。

这是一本有趣的书，它没有把篇幅花在计算一些枯燥的数学题上。相反，它教你如何直接控制硬件，在不借助于 BIOS、DOS、Windows、LINUX 或任何其他软件支持的情况下来显示字符、读取硬盘数据、控制其他硬件等。本书可作为大专院校相关专业学生和计算机编程爱好者的教程。

图书在版编目（CIP）数据

x86 汇编语言：从实模式到保护模式 / 李忠，王晓波，余洁著. —北京：电子工业出版社，2013.1
ISBN 978-7-121-18799-5

Ⅰ. ①x… Ⅱ. ①李… ②王… ③余… Ⅲ. ①汇编语言—程序设计 Ⅳ. ①TP313

中国版本图书馆 CIP 数据核字（2012）第 253290 号

责任编辑：董亚峰　　特约编辑：王　纲
印　　刷：涿州市京南印刷厂
装　　订：涿州市京南印刷厂
出版发行：电子工业出版社
　　北京市海淀区万寿路 173 信箱　邮编 100036
开　　本：787×1 092　1/16　印张：24.25　字数：620 千字
印　　次：2013 年 1 月第 1 次印刷
定　　价：56.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888。

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线：（010）88258888。

前 言

尽管汇编语言也是一种计算机语言，但却是与众不同的，与它的同类们格格不入。处理器的工作是执行指令并获得结果，而为了驾驭处理器，汇编语言为每一种指令提供了简单好记、易于书写的符号化表示形式。

一直以来，人们对于汇编语言的认识和评价可以分为两种，一种是觉得它非常简单，另一种是觉得它学习起来非常困难。

你认为我会赞同哪一种？说汇编语言难学，这没有道理。学习任何一门计算机语言，都需要一些数制和数制转换的知识，也需要大体上懂得计算机是怎么运作的。在这个前提下，汇编语言是最贴近硬件实体的，也是最自然和最朴素的。最朴素的东西反而最难掌握，这实在说不过去。因此，原因很可能出在我们的教科书上，那些一上来就搞一大堆寻址方式的书，往往以最快的速度打败了本来激情高昂的初学者。

但是，说汇编语言好学，也同样有些荒谬。据我的观察，很多人掌握了若干计算机指令，会编写一个从键盘输入数据，然后进行加减乘除或者归类排序的程序后，就认为自己掌握了汇编语言。还有，直到现在，我还经常在网上看到学生们使用 DOS 中断编写程序，他们讨论的也大多是实模式，而非 32 位或者 64 位保护模式。他们知道如何编译源程序，也知道在命令行输入文件名，程序就能运行了；又或者使用一个中断，就能显示字符。至于这期间发生了什么，程序是如何加载到内存中的，又是怎么重定位的，似乎从来不关汇编语言的事。这样做的结果，就是让人以为汇编语言不过如此，而且非常枯燥。

很难说我已经掌握了汇编语言的要义。但至少我知道，尽管汇编语言不适合用来编写大型程序，但它对于理解计算机原理很有帮助，特别是处理器的工作原理和运行机制。就算是为了这个目的，也应该让汇编语言回归它的本位，那就是访问和控制硬件（包括处理器），而不仅仅是编写程序，输入几个数字，找出正数有几个、负数有几个，大于 30 的有几个。

事实上，汇编语言对学习和理解高级语言，比如 C 语言，也有极大的帮助。老教授琢磨了好几天，终于想到一个好的比喻来帮助学生理解什么是指针，实际上，这对于懂得汇编语言的学生来说，根本就不算个事儿，并因此能够使老教授省下时间来喝茶。

在这本书之前，我也写过《穿越计算机的迷雾》一书。它们是一个系列，没有基础的读者可以先看那本书，打一点计算机原理的基础再来学习汇编语言。

在计划写这本书的时候，我就给自己画了几条线。首先不能走老路，一上来就讲指令、寻址方式，而是采用任务驱动的方式来写，每一章都要做点事情，最好是比较有趣，有吸引力。在解决问题的过程中，不断地引入新指令，并进行讲解。一句话，我希望是润物细

无声式的；其次，汇编语言和硬件并举，完全抛弃 BIOS 中断和 DOS 中断，直接访问硬件，发挥汇编语言的长处，因为这才是我们学习汇编语言的目的。也只有这样，读者才能深刻体会到汇编语言的妙处。

王晓波和湖北经济学院的余洁共同参与了本书的创作。

本书主要讲述 INTEL x86 处理器的 16 位实模式、32 位保护模式，至于虚拟 8086 模式，则是为了兼容传统的 8086 程序，现在看来已经完全过时，不再进行讲述。本书的特色之一是提供了大量典型的源代码，这些代码以及相配套的工具程序可以到书中指定的网站，或者电子工业出版社华信教育资源网搜索下载。

很多读者在读书的时候会遇到这种情况：一开始读得很快，一口气读了好几章。随着内容的深入，学习越来越吃力，不得不频繁回到前面重新学习已经讲过的内容，这就是因为前面的知识没有完全掌握。为此，本书每一章都设有检测点，读者应当在通过检测点之后再继续往后阅读。

本书原来有 18 章，后来考虑到实模式的内容过多，而去掉了一章。这一章的标题是《聆听数字的声音》，讲述如何通过直接访问和控制 Sound Blaster 16 声卡来播放声音，感兴趣的朋友可以从下载的配书文件包中找到这部分内容。

在本书的写作和出版过程中，长春电视台台长王志强，副台长周武军和技术部主任刘贵先后对本书给予了关心和支持，在此表示衷心的感谢。

好友王南洋、桑国伟、刘维钊、蒋胜友、邱海龙、万利、李文心等负责了本书的一部分校对工作；好友周卫平帮我验证配书代码是否能在他的机器上正常工作，在这里向他们表示感谢，同时也谢谢所有关心和支持本书的朋友们。

感谢我的母亲、我的妻子和我的女儿，她们是我的精神支柱，是我努力创作这本书的动力来源。

在阅读本书的过程中，如果有任何问题，可以往电子邮件地址 leechung@126.com 给我写信；要了解其他更多的情况，请访问我的博客：http://blog.163.com/leechung@126。

二〇一二年十一月

目 录

第1部分 预备知识

第 2 部分 实模式

第 3 部分 32 位保护模式

第 1 部分

预 备 知 识

- ✧ 了解数制的基本知识和数制转换的方法。
- ✧ 了解 8086 处理器的结构和工作方式，初步认识所谓的针对处理器编程，是针对处理器的哪些部件和哪些方面进行的，理解分段的原理。
- ✧ 了解什么是汇编语言，以及如何书写、编译汇编语言源程序，掌握在虚拟机上运行程序的方法。

第1章　十六进制计数法

电子计算机，顾名思义，就是计算的机器。因此，学习汇编语言，就不可避免地要和数字打交道。在这个过程中，我们要用到三种数制：十进制（这是我们再熟悉不过的）、二进制和十六进制。本章的目标是：

1．熟悉后两种数制，了解这两种数制的计数特点。

2．能够在这三种数制之间熟练地进行转换，特别是看到一个二进制数时，能够口算出它对应的十六进制数，反之亦然。

3．对于0～15之间的任何一个十进制数，能够立即说出它对应的二进制数和十六进制数。

1.1　二进制计数法回顾

1.1.1　关于二进制计数法

在《穿越计算机的迷雾》那本书里我们已经知道，计算机也是一台机器，唯一不同的地方在于它能计算数学题，且具有逻辑判断能力。

与此同时，我们也已经在那本书里学到，机器在做数学题的时候，也面临着一个如何表示数字的问题，比如你采用什么办法来将加数和被加数送到机器里。

同样是在那本书里，我们揭晓了答案，那就是用高、低两种电平的组合来表示数字。如图 1-1 所示，参与计算的数字通过电线送往计算机器，高电平被认为是“1”，低电平被认为是“0”，这样就形成了一个序列“11111010”，这就是一个二进制数，在数值上等于我们所熟知的二百五，换句话说，等于十进制数250。

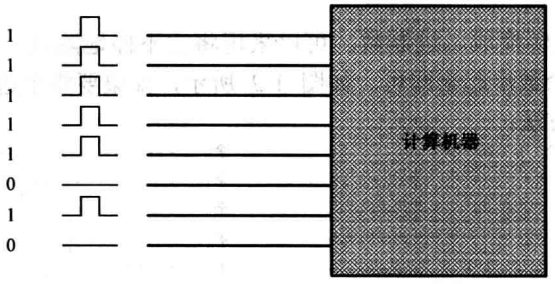

图 1-1　在计算机里，二进制数字对应着高低电平的组合

从数学的角度来看，二进制计数法是现代主流计算机的基础。一方面，它简化了硬件设计，因为它只有两个符号“0”和“1”，要得到它们，可以用最少的电路元件来接通或者关断电路就行了；另一方面，二进制数与我们熟悉的十进制数之间有着一对一的关系，任何一个十进制数都对应着一个二进制数，不管它有多大。比如，十进制数 5，它所对应的二进制数是 101，而十进制数 5785478965147 则对应着一长串“0”和“1”的组合，即 10101000011000010010110101100100111100011011。

组成二进制数的每一个数位，称为一个比特（bit），而一个二进制数也可以看成是一个比特串。很明显，它的数值越大，这个比特串就越长，这是二进制计数法不好的一面。

1.1.2 二进制到十进制的转换

每一种计数法都有自己的符号（数符）。比如，十进制有 0、1、2、3、4、5、6、7、8、9 这十个符号；二进制呢，则只有 0、1 这两个符号。这些数字符号的个数称为基数。也就是说，十进制有 10 个基数，而二进制只有两个。

二进制和十进制都是进位计数法。进位计数法的一个特点是，符号的值和它在这个数中所处的位置有关。比如十进制数 356，数字 6 处在个位上，所以是“6 个”；5 处在十位上，所以是“50”；3 处在百位上，所以是“300”。即：

$$\text{百位 } 3、\text{十位 } 5、\text{个位 } 6=3\times10^2+5\times10^1+6\times10^0=356$$

这就是说，由于所处的位置不同，每个数位都有一个不同的放大倍数，这称为“权”。每个数位的权是这样计算的（这里仅讨论整数）：从右往左开始，以基数为底，指数从 0 开始递增的幂。正如上面的公式所清楚表明的那样，“6”在最右边，所以它的权是以 10 为底，指数为 0 的幂 10^0；而 3 呢，它的权则是以 10 为底，指数为 2 的幂 10^2。

上面的算式是把**十进制**数“翻译”成**十进制**数。从十进制数又算回到十进制数，这看起来有些可笑，注意这个公式是可以推广的，可以用它来将二进制数转换成十进制数。

比如一个二进制数 10110001，它的基数是 2，所以要这样来计算与它等值的十进制数：

$$10110001B=1\times2^7+0\times2^6+1\times2^5+1\times2^4+0\times2^3+0\times2^2+0\times2^1+1\times2^0=177D$$

在上面的公式里，10110001B 里的“B”表示这是一个二进制数，“D”则表示 177 是个十进制数。“B”和“D”分别是英语单词 Binary 和 Decimal 的头一个字母，这两个单词分别表示二进位和十进位的意思。

◆ **检测点 1.1**

将下列二进制数转换成十进制数：

1101、1111、1001110、11111111、10000000、1101101100011011

1.1.3 十进制到二进制的转换

为了将一个十进制数转换成二进制数，可以采用将它不停地除以二进制的基数 2，直到商为 0，然后将每一步得到的余数串起来即可。如图 1-2 所示，如果要将十进制数 26 转换成二进制数 11010，那么可采用如下方法：

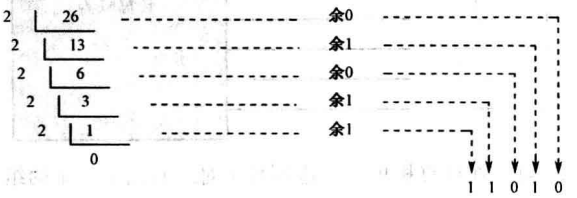

图 1-2 将十进制数 26 转换成二进制数

第 1 步，将 26 除以 2，商为 13，余数为 0；

第 2 步，用 13 除以 2，商为 6，余数为 1；

第 3 步，用 6 除以 2，商为 3，余数为 0；

第 4 步，用 3 除以 2，商为 1，余数为 1；

第 5 步，用 1 除以 2，商为 0，余数为 1，结束。

然后，从下往上，将每一步得到的余数串起来，从左往右书写，就是我们所要转换的二进制数。