汇编语言与逆向技术

第3章 汇编语言基础

王志
zwang@nankai.edu.cn

南开大学 网络空间安全学院

2024-2025学年

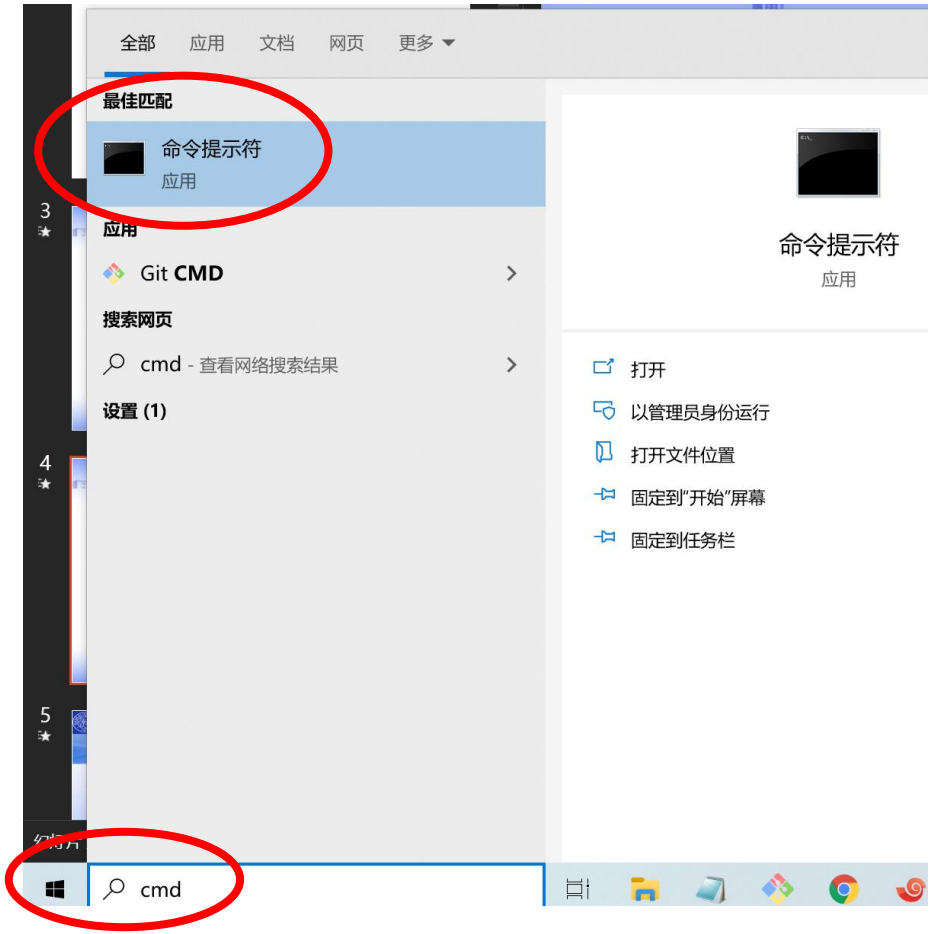
本章知识点

- 汇编语言的基本元素
  - 重点：标识符、指令、伪指令
  - 难点：伪指令
- 数据定义
  - 重点：BYTE、WORD、DWORD、DUP
- 符号常量
  - 难点：$符号常量

Hello World实验的问题

- 打开cmd命令行窗口
- 当前路径、相对路径、绝对路径
- 路径切换
- 查看目录中的文件列表

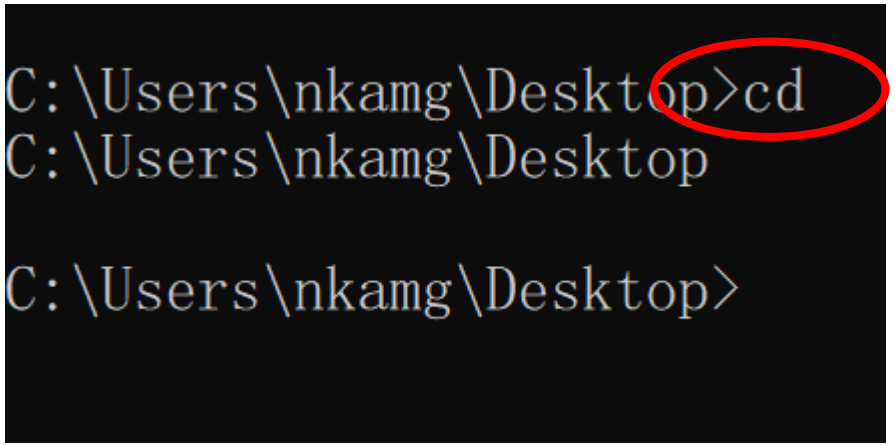
打开cmd命令行窗口



- 按Windows键
- 输入cmd
- 最佳匹配里面
  - “命令提示符”

当前路径

- 获得当前路径
  - cd命令

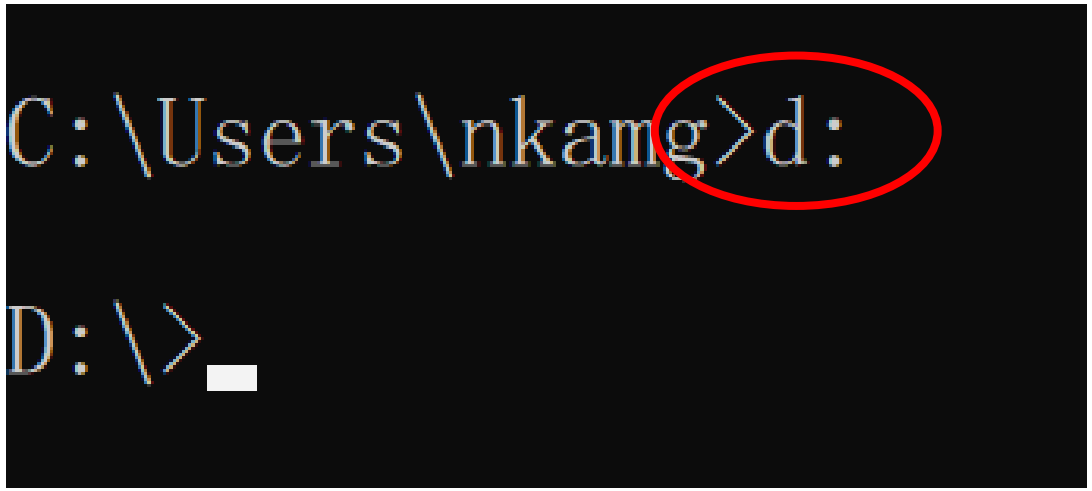


绝对路径

- 以盘符开始的路径

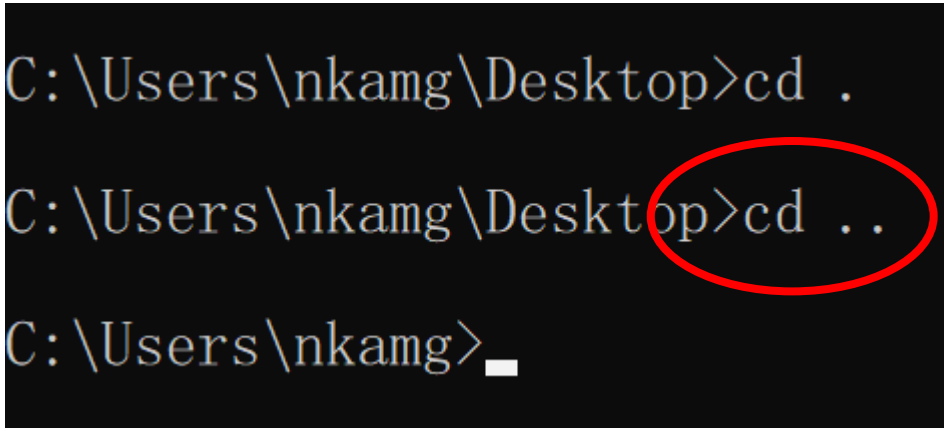
例如“C:\Users\nkamg\Desktop”

- 从C盘进入到D盘
  - d:



相对路径

- 相对于当前的路径
- “.” 表示的是当前路径
- “..”表示的是上一级路径



查看目录中的文件列表

- dir命令

```
C:\Users\nkamg>d:

D:\>dir
 驱动器 D 中的卷没有标签。
 卷的序列号是 1234-5678

 D:\ 的目录

2021/09/26  01:03    <DIR>          HPSCANS
2021/05/27  14:01        14,321,039 信息安全新技术研究室-20210528.pptx
2021/04/06  11:34         4,477,682 ch6-Recognizing C Constructs in Assembly.pptx
2021/06/23  12:29            39,028 PPT活动背景.pptx
2021/04/13  12:02       219,287,064 RainClassroom_Full_4.3.0.2006.exe
2021/04/20  09:40            20,113 第9章-动态调试.docx
2021/04/20  10:34         5,188,096 ch7-Analyzing Malicious Windows Programs.ppt
2021/04/25  10:43         3,161,088 ch8-Debugging.ppt
2021/04/25  11:30         3,263,488 ch9-OllyDbg.ppt
2021/04/27  09:33         2,338,730 教育.pptx
```

主观题 10分

设置

“\masm32\bin\ml /c /Zd /coff hello_console.asm”中

\masm32\bin\ml是一个相对地址还是绝对地址？

如何判断当前文件夹中是否有hello_console.asm文件？

正常使用主观题需2.0以上版本雨课堂

作答

遇到的路径问题

- 相对路径错误
  - \masm32\bin\ml /c /Zd /coff hello_console.asm
- 当前路径没有需要的文件
  - hello_console.asm
- asm代码中include、includelib路径问题
  - include \masm32\include\windows.inc

汇编语言的基本元素

汇编语言的基本元素

- 整数常量、整数表达式
- 实数常量
- 字符常量、字符串常量
- 保留字、标识符
- 指令、伪指令、NOP指令

主观题 10分

设置

C++如何表示十六进制整数常量？例如十六进制的"1FF"在C++中如何表示。

作答

整数常量

- [{+|-}]数字[基数]
- 基数后缀（Radix）
  - d (**d**ecimal)、o(**o**ctonary)、h(**h**exadecimal)、b(**b**inary)
  - 如果整数常量后面没有基数后缀，默认是十进制整数
  - 10、10d、10o、10h、0A0h，10b
- 以字母开头的十六进制常量前面必须加0

单选题 10分

设置

FFh是有效的整数常量吗？

A 是

B 不是

提交

南開大學 Nankai University

整数表达式

- 包含整数值和算数运算符的数学表达式
- 表达式的结果不能超过32bits的表示范围
  - Carry Flag
- MOD：取余数运算

整数表达式

- 算术运算符的优先级
- （）优先级1
- *、/、MOD，乘、除、取余，优先级2
- +、-，加减，优先级3

填空题 10分

设置

表达式12 - 2 MOD 5 的计算结果是 [填空1]

正常使用填空题需3.0以上版本雨课堂

作答

南開大學 Nankai University

实数常量

- 十进制实数
- 编码（十六进制）实数

十进制实数常量

- -1.11E-5、2.、+3.0、2.E5
- 十进制实数常量由符号sign、整数、小数点、小数和指数组成
- [sign]integer.[integer][exponent]
- 至少要有一个数字和一个小数点

编码实数

- 编码实数是以十六进制数表示一个实数，遵循IEEE浮点数格式
- 《Intel汇编语言程序设计》第五版，第17章“浮点处理和指令编码”

字符常量

- 单引号或者双引号括起来的单个字符。
- 汇编器会将其转化为ASCII编码
- ‘A’、“B”

字符串常量

- 以单引号或者双引号括起来的一串字符
- ‘ABC’、“abc”
- 嵌套引号
- “print ‘Hello World’ on the terminal window”
- ‘print “Hello World” on the terminal window’

保留字

- 指令助记符：MOV、ADD
- 伪指令：INCLUDE、PROC
- 属性：BYTE、WORD
- 预定义符号：$、？
- 参考《Intel汇编语言程序设计》第五版 附录A

标识符

- 标识符是程序员选择用来标识变量、常量、过程、代码的标号
  - 包含1~247个字符
  - 大小写不敏感（MASM默认）
  - 第一个字符必须是字母、下划线、@、？或$
  - 第一个字符不能是数字（对比十六进制整数）
  - 标识符的名字不能与汇编器的保留字相同。

单选题 10分

设置

判断题：标识符可以用数字开头。

A 正确

B 错误

提交

指令

- 汇编语言中的指令是一条汇编语句
- 汇编器把汇编指令翻译成对应的机器指令
  - 标号
  - 指令助记符
  - 操作数
  - 注释

标号

- 标号是充当指令或数据位置标记的标识符
- 数据标号
  - 标识变量的地址
- 代码标号
  - 标识代码的地址

数据标号

- 标识变量的地址，方便变量的引用
- count DWORD 100
- array DWORD 100, 101, 102, 103
- 相对.data数据段在内存起始地址的偏移

OFFSET

- 获取数据标号的内存偏移地址

```
.data
    str_hello BYTE “Hello World！”，0
.code
    mov eax, OFFSET str_hello
```

代码标号

- 标识代码的地址，必须以冒号（:）结尾
- 通常作为跳转、循环指令的目标地址

```
target:
    mov eax，100h
    …
    jmp target
```

单选题 10分

设置

判断题：代码标号后面有冒号，数据标号后面没有冒号

A 正确

B 错误

提交

南開大學 Nankai University

指令助记符

- 指令助记符（instruction mnemonic）是一个简短的单词，用于表示一条指令。
  - mov、add、sub、mul、jmp、call

操作数

- 操作数是指令的操作对象
  - 寄存器
  - 内存
  - 常量
  - I/O端口

多选题 10分

设置

CPU指令可以直接访问的操作数类型有？

A 寄存器

B 内存

C I/O接口

D 硬盘

E 常量（立即数）

提交

操作数

- inc eax
  - eax寄存器的值加1
- mov count， ebx
  - mov指令有两个操作数：count、ebx
  - 第一个操作数是目的操作数
  - 第二个操作数是源操作数

注释

- 单行注释
  - mov count，ebx；save result to count
- 块注释: COMMENT伪指令和用户定义的符号

```
COMMENT !
        This is a comment
!
```

NOP指令

- NOP指令，空操作
  - 用于计时循环
- NOP指令占用1个字节的内存
  - 用于后继指令的对齐
  - IA-32处理器从偶数双字地址处加载代码和数据时更加快速

主观题 10分

设置

讨论题：什么是伪指令？ 我们为什么要学习伪指令？

作答

南開大學 Nankai University

伪指令

- 伪指令内嵌在汇编语言源代码中，由汇编器识别、执行相应动作的命令
- 用于定义变量、段、过程、汇编器选项等
- 参考《Intel汇编语言程序设计》第五版，附录A， MASM的伪指令

伪指令

- 定义变量

  my_var DWORD 100h；DWORD伪指令

  mov eax, my_var ； mov指令

伪指令

- 定义段（Segment）
  - .data、.code、.stack
- 定义过程（Procedure）
  - PROC、ENDP
- 允许或禁止汇编器的某些特性
  - OPTION、.386、.MODEL

单选题 10分

设置

判断题：伪指令是在程序运行时执行的

A 正确

B 错误

提交

定义数据

内部数据类型

- MASM内部以数据位的个数定义了多种数据类型
  - BYTE，db，8位
  - WORD，dw， 16位
  - DWORD，dd， 32位
  - QWORD ，dq，64位

内部数据类型

- MASM汇编器默认情况下，大小写不敏感
- DWORD
  - Dword
  - dword
  - dWord

数据定义语句

- 为变量在内存中保留存储空间
- 为变量指定一个名字（数据标号）
- [变量名] 数据定义伪指令 初始值

数据定义伪指令

- BYTE，db，8 bits
- WORD，dw， 16 bits
- DWORD，dd， 32 bits
- QWORD ，dq，64 bits

初始值

- 数据定义语句中要指定初始值
- 多个初始值用逗号隔开
  - my_var DWORD 0, 1, 2, 3
- 0: 可以指定初始值为0
- ?: 表示在程序运行的时候初始化该变量

数据声明的位置

- .data段声明初始化的变量

```
.data
    dw_var1 DWORD 0
```

- .data?段声明未初始化的变量

```
.data?
    dw_var2 DWORD ?
```

定义字符串

```
str_hello BYTE “Hello World!”, 0Dh, 0Ah,
          BYTE “I love assembly language”,
          BYTE 0Dh, 0Ah, 0
```

- 0Dh和0Ah是CR/LF（回车、换行）的ASCII编码
- 字符串的结尾是0

DUP伪指令

- 为字符串或者数组分配内存空间
- BYTE 20 DUP（0）；20个字节的内存空间
- BYTE 4 DUP（“Hello”）：20个字节，连续的4个“Hello”，每个“Hello”5字节

定义WORD和SWORD数据

- 在数据定义语句中如果使用WORD（用于定义字）和SWORD（用于定义有符号字）伪指令就可以为一个或多个16位整数分配存储空间
- word1 WORD 65535 ； 最大无符号字
- 字数组：可以通过显示指令初始化每个元素，或使用DUP操作符创建字数组。

定义DWORD和SDWORD数据

- 在数据定义语句中使用DWORD（定义双字）和SDWORD（定义有符号双字）伪指令，可以为一个或多个32位的整数分配存储空间
- val1 DOWRD 12345678h ； 无符号数
- 双字数组：所谓双字数组，就是指可以通过显式地初始化数组每个元素，或使用DUP操作符来创建双字数组。

定义QWORD数据和TBYTE数据

- 使用QWORD（定义8字节）伪指令可以定义64位的数据。
- val1 QWORD 1234567812345678h
- 使用TBYTE（定义10字节）伪指令可以定义80位的数据。
- val1 TBYTE 100000000123456789Ah

定义实数

- REAL4定义4字节的单精度实数，REAL8定义8字节的双精度实数，REAL10定义10字节的扩展精读实数.
- 对于每个伪指令，都要求一个或多个与其数据尺寸相匹配的实数常量初始值

填空题 10分

设置

声明一个包含单词“TEST”重复50次的字符串变量 [填空1]

正常使用填空题需3.0以上版本雨课堂

作答

单选题 10分

设置

以下3条BYTE伪指令的效果是一致的吗?

(1) str_hello BYTE “Hello World”,0

(2) str_hello BYTE 48h, 65h, 6ch, 6ch, 6fh, 20h, 57h, 6fh, 72h, 6ch, 64, 0

(3) str_hello BYTE ‘H’,’e’,’l’,’l’,’o’,’ ’,’W’,’o’,’r’,’l’,’d’,0

A 是

B 不是

提交

符号常量

符号常量

- 符号常量（或符号定义），将标识符与整数表达式或文本联系起来
- 符号常量不占用存储空间
- 变量占用存储空间

等号伪指令

- 等号伪指令，将符号名和整数表达式联系起来

COUNT = 500

mov eax， COUNT

- 易于阅读与维护
- 减少程序修改时的查找与替换次数

=

Article • 08/04/2021 • 2 minutes to read • 6 contributors

Assigns the numeric value of *expression* to *name*.

Syntax

```
name = expression
```

计算数组和字符串的大小

- MASM用$运算符存储当前语句的地址偏移值。
- $可以用来计算数组或字符串的大小

计算字符串大小

```
str_hello BYTE “Hello World!”, 0Dh, 0Ah,
          BYTE “I love assembly language”,
          BYTE 0Dh, 0Ah, 0
```

- str_size = ($ - str_hello)

主观题 10分

设置

使用$运算符计算数组中有几个dw数据？

dw_array DWORD 0, 1, 2, 3, 4
array_size = ?

作答

计算数组大小

```
dw_array DWORD 0, 1, 2, 3, 4

array_size = ($ - dw_array)/4
```

EQU伪指令

- EQU伪指令将符号名与整数表达式或任意文本联系起来
  - name EQU expression
  - name EQU symbol
  - name EQU <text>

EQU 伪指令

PI EQU 3.1415926

press_key EQU <“Press any key to continue…”, 0>

.data

prompt BYTE pressKey ; 变量

EQU伪指令

- EQU伪指令不能在程序中重定义
- “=”伪指令可以在程序中重定义

=

Article • 08/03/2021 • 6 contributors

In this article

Syntax
Remarks
See also

Assigns the numeric value of *expression* to *name*.

Syntax

name = *expression*

Remarks

The symbol can be redefined later.

EQU

Article • 08/03/2021 • 6 contributors

In this article

Syntax
Remarks
See also

The first directive assigns numeric value of *expression* to *name*.

Syntax

name EQU *expression*

name EQU <*text*>

Remarks

The *name* cannot be redefined later.

TEXTEQU伪指令

- TEXTEQU伪指令与EQU非常相似，同样可以用来创建文本宏（text macro）。
- 它有三种不同的使用格式：第一种格式会将文本赋值给符号；第二种格式则将已定义的文本宏内容赋值给符号；第三种格式将整数表达式常量赋值给符号。
- name TEXTEQU <text>
- name TEXTEQU textmacro
- name TEXTEQU %constExpr

实验2：dex2hex

实验目的

- 熟悉汇编语言的数据传送、寻址和算术运算；
- 熟悉汇编语言过程的定义和使用；
- 熟悉十进制和十六进制的数制转换

实验环境

- MASM32编译环境
- Windows命令行窗口

实验内容

- 编写汇编程序dec2hex.asm，编译成dec2hex.exe。
- dec2hex.exe能够将Windows命令行输入的十进制无符号整数，转换成对应的十六进制整数，输出在Windows命令行中

```
D:\>dec2hex.exe
Please input a decimal number( 0~ 4294967295): 100

The hexdecimal number is : 00000064
```

StdIn函数

- 获得用户输入的十进制整数。
- 定义在\masm32\include\masm32.inc
- 库文件是\masm32\lib\masm32.lib。
- StdIn函数的定义“StdIn PROTO :DWORD,:DWORD”
  - 内存存储空间的起始地址
  - 内存存储空间的大小。

StdIn函数

- 用户输入的十进制数对应的ASCII编码字符串存储在内存中
- 输入：100
- 内存存储：49h, 48h, 48h, 00

汇编语言与逆向技术

第3章 汇编语言基础

王志
zwang@nankai.edu.cn

南开大学 网络空间安全学院

2024-2025学年

本章知识点

- 汇编语言的基本元素
  - 重点：标识符、指令、伪指令
  - 难点：伪指令
- 数据定义
  - 重点：BYTE、WORD、DWORD、DUP
- 符号常量
  - 难点：$符号常量