有間大學

汇编语言与逆向技术课程实验报告

实验九: Reverse Engineering Exercises – Advanced



学	院	网络空间安全学院
专	业	信息安全、法学双学位
学	号	2212000
姓	名	宋奕纬
班	级	1061

一、实验目的

- 1、 进一步熟悉静态反汇编工具 IDA Freeware;
- 2、 熟悉将反汇编代码进行反编译的过程;
- 3、 掌握对于反编译伪代码的逆向分析;
- 4、 运用熟悉的编程语言, 实现简单的脚本编写

二、实验原理

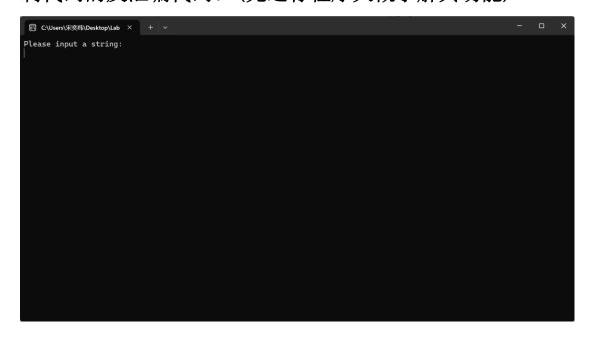
- 1、实验环境: 反汇编工具 IDA Freeware
- 2、实验原理:

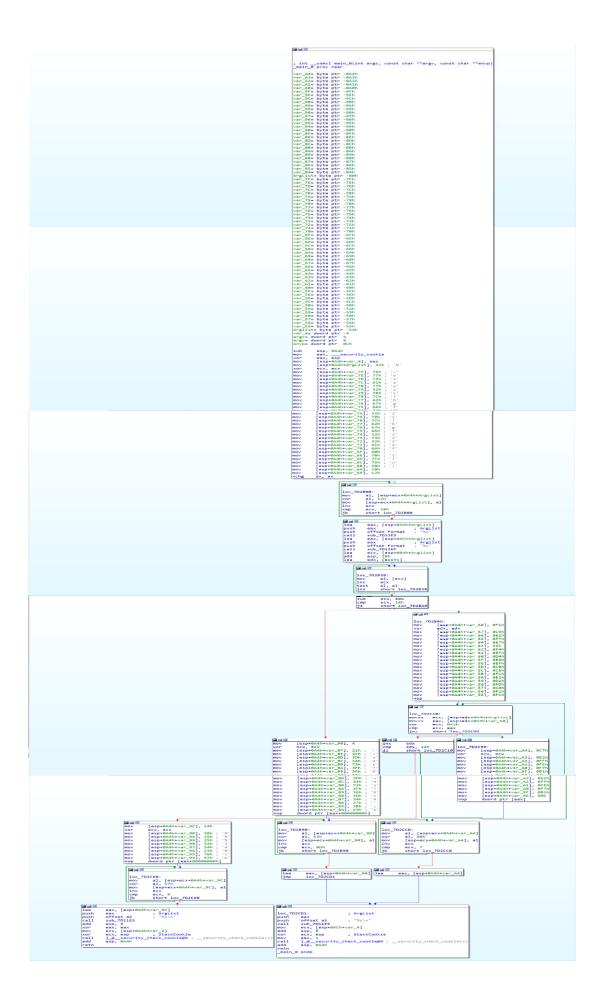
通过 IDA 得到二进制代码的反汇编代码,利用汇编所学知识对反汇编代码的数学计算、数据结构、条件判断、分支结构进行识别与分析。

三、实验过程

(一) Task3

1、 使用 IDA Freeware 打开 task3. exe 文件, 查看其二进制代码的反汇编代码。(先运行程序大概了解其功能)





2、 使用 IDA 的反编译功能得到 task3. exe 的伪代码。

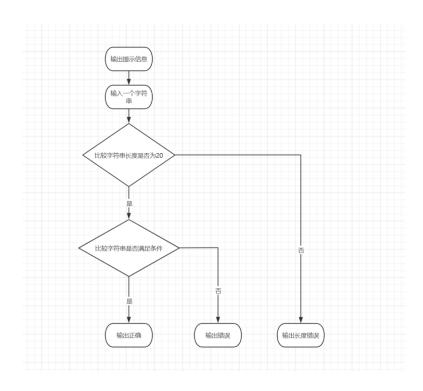
```
1 int __cdecl main_0(int argc, const char **argv, const char **envp)
                                                                                                                                                                                                    int __cosci man_@(int args, const char "args, c {
    unsigned int v3; // ecx
    unsigned int v4 // ecx
    char v5; // eax
    unsigned int v7; // ecx
    unsigned int v7; // ecx
    unsigned int v9; // ecx
    unsigned int 
be char viz: , (esp-m., 10 char viz: , (esp-m., 12 char viz: ), (esp-m., 12 char viz: ], (/esp-sh] [vv-, 12 char viz: ], (/esp-sh] [vv-, 12 char viz: ], (/esp-sh] [vb-p-sh] syrer (esp-sh) syrer (esp-sh) esp-sh] esp
                                                                                                                                                                                             | while (v3 < 0x18 );
| sub_TOLIES('Ne", (char)ArgList);
| sub_TOLIES('Ne", (char)ArgList);
| if (strlen(ArgList) == 20 )
| v16[0] = -15;
| v6 = 0;
| v16[2] = -31;
| v16[3] = -1;
| v16[3] = -1;
| v16[4] = -25;
| v16[5] = -169;
| v16[6] = -17;
| v16[7] = -17;
| v16[9] = -24;
| v16[1] = -64;
| v16[1] =
```

3、对代码和伪代码进行分析。

(1) 统观代码:

观察代码后,发现没有出现一开始运行程序时出现的"Please input a String:"字符串,同时也没有我们想看到的提示正确或者提示错误的提示信息。故猜测这个 task 需要先对提示信息进行分析,然后再去寻找可以输出"正确"提示信息的字符串。

对结构进行简单分析后作出以下流程图:



(2) 解密提示信息:

Ps:

- 1、 这里的伪代码直接将需要被处理的字符串给出,也可以跟据反汇编代码中的数字去分析(故写了多套脚本——在之后具体介绍)
- 2、对伪代码中函数和变量的理解: 伪代码中的变量、数组其实都是地址,指向这个地址下所存的数据; sub_7D11E5 这个函数的参数是输出的起始地址,从起始地址的数据开始输出
- ① "Please input a string: "

```
qmemcpy(ArgList, "B~wsaw2{|bgf2s2af`{|u(", 22);
v3 = 0;
ArgList[22] = 24;
ArgList[23] = 18;
do
    ArgList[v3++] ^= 0x12u;
while ( v3 < 0x18 );
sub_7D11E5("%s", (char)ArgList);</pre>
```

将这个字符串每一位的 ascII 码与 0x12 进行异或后作为新的 ascII 码,生成一个新的字符串(解出来是"Please input a string:")。

2 "Correct!"

```
v11 = 20;
v7 = 0;
qmemcpy(v12, "8%%24#vW", sizeof(v12));
do
  v12[v7++ - 1] ^= 0x57u;
while ( v7 < 9 );
sub_7D11E5("%s\n", (char)&v11);
```

将这个字符串每一位的 ascII 码与 0x54 进行异或后作为新的 ascII 码,生成一个新的字符串(解出来是"Correct!")。

此处有一个细节,这里是将 V12 从-1 开始进行异或和输出的(即 V12 地址前的数据,即 V11 这个被赋值为 20 的数字, 20 和 0x57(87)异或,解出了字符"C")。

3 "Wrong!"

```
v10[0] = -57;
v9 = 0;
v10[1] = -30;
v10[2] = -1;
v10[3] = -2;
v10[4] = -9;
v10[5] = -79;
v10[6] = -112;
do
    v10[v9++] ^= 0x90u;
while ( v9 < 7 );
</pre>
```

将 V10 每一位的 ascII 码与 0x90 进行异或后作为新的 ascII 码,生成一个新的 字符串(解出来是"Wrong!")。

4 "Wrong length"

```
v13 = 4;

v4 = 0;

qmemcpy(v14, "!<=4s?6=4';S", sizeof(v14));

do

v14[v4++ - 1] ^= 0x53u;

while ( v4 < 0xD );

v5 = &v13;
```

将这个字符串每一位的 ascII 码与 0x54 进行异或后作为新的 ascII 码,生成一个新的字符串(解出来是"Wrong length")

这里与输出 "Correct!"信息处细节类似——这里是将 V14 从-1 开始进行异或和输出的(即 V14 地址前的数据,即 V13 这个被赋值为 4 的数字, 4 和 0x57(87) 异或,解出了字符 "W")。

(3) 解密需要输入的字符串:

```
sub_AEIIEF("%s", '(char)Arglist);
if (strlen(Arglist) == 20 )
{
vis(0) = -15;
vis = 0;
vis(1) = -55;
vis(2) = -31;
vis(3) = -165;
vis(3) = -189;
vis(3) = -189;
vis(6) = -12;
vis(7) = -12;
vis(7) = -17;
vis(8) = -44;
vis(9) = -24;
vis(1) = -84;
vis(9) = -24;
vis(11) = -64;
vis(13) = -80;
vis(14) = -80;
vis(15) = -80;
vis(15) = -80;
vis(15) = -80;
vis(15) = -80;
vis(16) = -80
```

先定义数组 V16

条件:

Arglist[i]^0x45==V6[i];

若要解出原始的 Arglist, 则利用 a^b=c、a^b^b=a=c^b,得

Arglist[i]=V6[i]^0x45

即将数组 V6 的每一个元素与 0x45 异或后推出字符串的每一个字符的 ascll 码, 进而推出字符 串。

4、 脚本实现

写了两个脚本:

第一个可以输入一个字符串,输入一个十六进制数,将字符串每个字符的 ascII 码和十六进制数异或后,生成新的字符串并输出;一个可以输入一个数组和一个数,将数组中的元素和这个数异或后得到对应的字符串。

这里可以只用第二个脚本,根据汇编代码挨着输入数据即可,但效率低。

第一个脚本程序虽然存在一些问题,只能处理字符串(伪代码给出的),但是效率较高,故将两者结合起来进行计算。

```
| ConsoleApplication17.cpp * | **
| Con
```

(1) 解密提示信息

解出 "Please input a string: "

解出 "Correct! "

解出 "Wrong length"

(2) 解密字符串

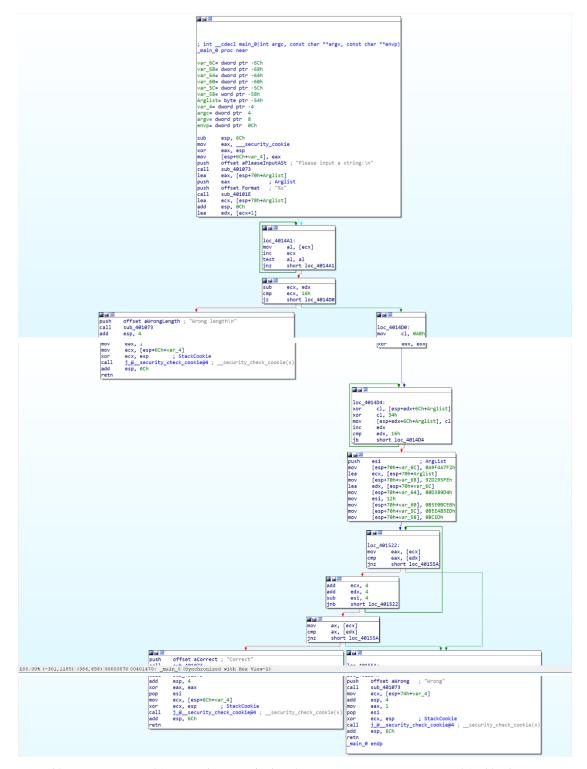
解出字符串为 "T1DZB6QJqMJekYGtXeTY"

5、 测试运行。

将字符串"T1DZB6QJqMJekYGtXeTY"输入程序中,得到正确的结果。

(二) Task4

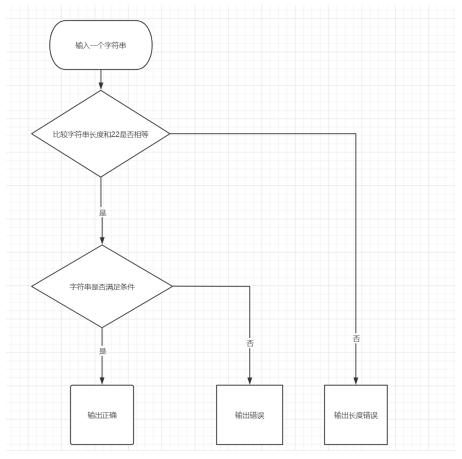
1、 使用 IDA Freeware 打开 task4. exe 文件, 查看其二进制代码的反汇编代码。



2、使用 IDA 的反编译功能得到 task4. exe 的伪代码。

3、 对代码和伪代码进行分析。

还是对代码进行总体分析,分析其逻辑与结构。作出一张流程图:

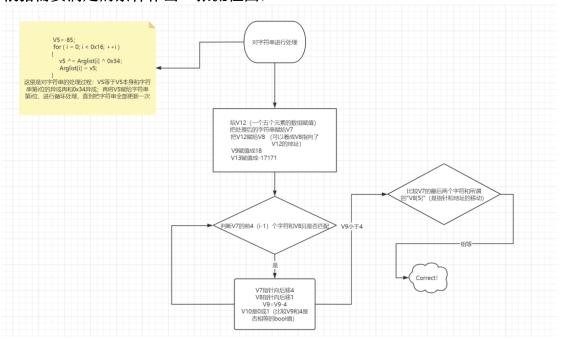


总体逻辑不复杂,主要是对字符串所满足条件的分析。

直接看字符串需要满足条件的伪代码进行分析:

```
v5 = -85;
for ( i = 0; i < 0x16; ++i )
  v5 ^= Arglist[i] ^ 0x34;
  Arglist[i] = v5;
v11 = v3;
v12[0] = -1443584014;
v7 = Arglist;
v12[1] = -1831692802;
v8 = v12;
v12[2] = -2133620268;
v9 = 18;
v12[3] = -1243562773;
v12[4] = -1092307475;
v13 = -17171;
while ( *(DWORD *) \lor 7 == * \lor 8 )
  v7 += 4;
  ++v8;
  v10 = v9 < 4;
  v9 -= 4;
  if ( v10 )
    if (*(_WORD *) \vee 7 == *(_WORD *) \vee 8)
       sub_EE1073("Correct", v11);
       return 0;
```

根据需要满足的条件作出一张流程图:



看上去是双层循环,其实是在比较处理后的字符串和数组之间的对应关系。 若想要输出 "Correct!":

- ①对字符串进行处理,得到新的字符串。
- ②新的字符串和从地址 V8 开始的数据进行比较。
- *(_DWORD *) v7 是从当前地址强制转化为双字类型(32 位),一个字符是 8 位,故一个双字可以存四个字符的数据。理解成每次比较 4 个字符——这样的话将 V8 存储的数组用完也只能比较 20 个字符。故用了下一条指令,即在前 20 个字符比较之后均符合的情况下,比较后两个字符是否满足条件:
- if (*(_WORD *) $v7 == *(_WORD *)v8$),从 V7 比较完前 20 个字符后的地址 开始,转换成一个单字,存储两个字符的信息,与 V8 地址之后的 V9 转化成的

单字类型进行比较,如果二者相等,则输出正确的提示语。 强制类型转化过程中,譬如处理后的字符串中有 4 位是 "abcd",则 a 的信息 先进行存储,存储到信息的第 4 位(简单理解为先入后出的栈操作)。

故若想求出原始字符串,就要先求出被处理后的字符串。 设处理后的是 B[22],处理前的是 A[22],可以通过如下方式进行计算。

```
int v5 = -85;

for (int i = 0; i < 22; ++i){
A[i] = v5 \land B[i] \land 0x34;
v5 = A[i];
```

被处理后的字符串则可以通过 V8 存储的数组及 V9 得出(这里直接观察汇编代码更容易得出)

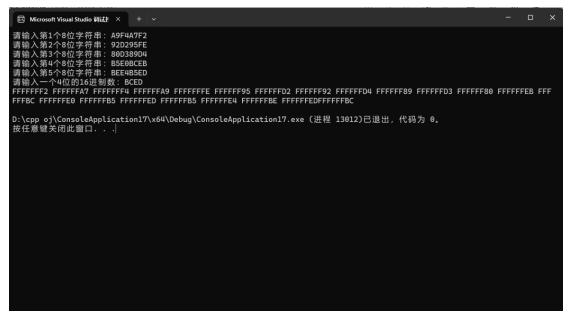
```
[esp+70h+var_6C], 0A9F4A7F2h|
mov
        ecx, [esp+70h+Arglist]
lea
        [esp+70h+var 68], 92D295FEh
mov
lea
      edx, [esp+70h+var_6C]
mov
      [esp+70h+var 64], 80D389D4h
       esi, 12h
mov
moν
      [esp+70h+var_60], 0B5E0BCEBh
       [esp+70h+var 5C], 0BEE4B5EDh
mov
       [esp+70h+var 58], 0BCEDh
mov
```

V8 中存 8 位 16 进制数(一个双字大小,每两位对应一个字符)
V9 是一个 4 位 16 进制数(每两位对应一个字符)
具体实现就是对每个元素从后两位开始处理。
比如 A9F4A7F2, 就是从 F2 开始, 写成 FFFFFFF2
FFFFFFA7 FFFFFF4 FFFFFFA9, 每个都对应一个有符号的十进制数,再对应处理后的字符串的ascll 码。每一个元素都如上处理。

故整个流程就是利用 V8 和 V9 求出转换后的字符串的 ascII 码,然后逆推出原来字符串的 ascII 码,最后根据 ascII 码推出对应的字符串。

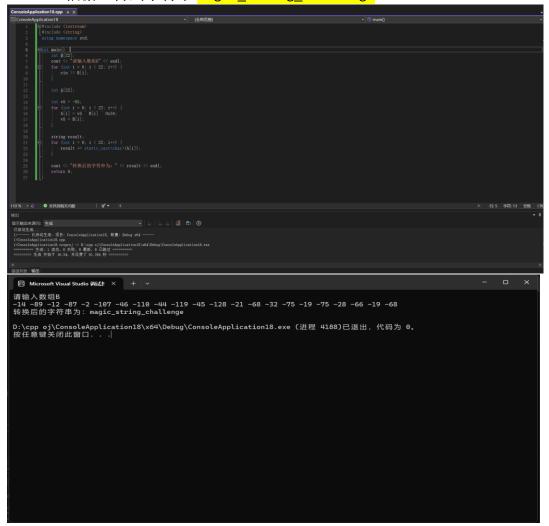
4、脚本计算。

(1) 对 V8 和 V9 进行处理,整合出一个数组,存放 22 个转化后字符串中字符的 8 位 16 进制数。



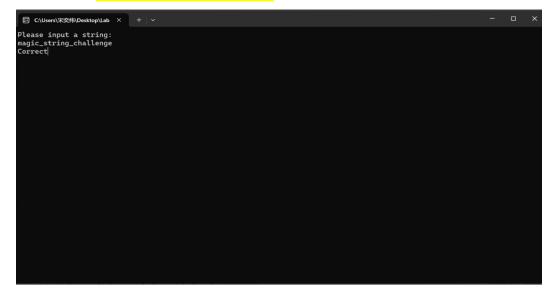
(2) 对得到的数组进行处理,全部转化成对应的 10 进制数(作为数组 B)

(3) 对得到的数组 B 进行处理,得到存放正确字符串 AscII 码的数组 A, 根据 A 得到字符串 magic_string_challenge



5、测试运行。

将字符串"magic_string_challenge"输入,得到正确答案。



四、实验结论及心得体会

- 1、 熟悉静态反汇编工具 IDA Freeware。
- 2、 熟悉反汇编代码的逆向分析过程。
- 3、掌握反汇编语言中的数学计算、数据结构、条件判断、分支结构的识别和逆向分析。
- 4、 对地址、变量、指针都有了更深入的理解
- 5、 对汇编语言有了更深入的了解。
- 6、提升了解决问题的能力,学会多种编程语言综合使用。