

南开大学

计算机学院和网络空间安全学院

《密码学》课程作业

SPN 线性密码分析 &SM4 差分密码分析

姓名:陆皓喆

学号: 2211044

专业:信息安全

指导教师: 苏明

2024年10月14日

目录

1	实验要求	2
2	github 仓库	2
3	实验原理 3.1 线性密码分析	2 2 2
4	Question1	3
5	Question2	7
6	附录 1:Random_Generate_Key_Pairs.py	7
7	附录 2:Linear_Attack.py	8
8	附录 3:Differential_Attack.py	13

1 实验要求

- 1. 请利用线性密码分析确定 SPN 分组加密算法的 (轮) 密钥;
- 2. 请计算国密分组加密算法 SM4 的 SBox 差分分布表。

2 github 仓库

本次实验的有关代码和文件,都已经上传至我的个人 github 中。 您可以通过访问**此链接**来查阅我的代码文件。

3 实验原理

3.1 线性密码分析

SPN 线性密码分析是一种基于 S 盒逼近的方法,已知明文和密文,分析出对应的密钥。此方法可以分析出最后一轮的 16 位密钥值,虽然课本上只提供了分析其中八位的伪代码,但是我们经过分析之后,得出了剩下八位的分析过程,进而完成了对整个密钥的破解。总体的流程为:

- 通过上一次实验中所使用的 SPN 生成代码, 随机生成 8000 对明文密文对;
- 构造出一个固定的表达式;
- 通过构建的表达式和前面所生成的明文密文对,来构建对应的线性逼近表;
- 选取线性逼近表中偏差最大的作为输出,即可得到对应的密钥值;
- 重复做两次上述操作,即可获得全部的16位密钥值。

对于本次实验,我们构建的第一轮表达式与课本上所提供的一致,为:

$$z_1 = x_5 \oplus x_7 \oplus x_8 \oplus u_6^4 \oplus u_8^4 \oplus u_{14}^4 \oplus u_{16}^4$$

通过查阅资料和分析,我们可以得到另外两组的密钥所对应的表达式为:

$$z_{2} = x_{1} \oplus x_{2} \oplus x_{4} \oplus u_{1}^{4} \oplus u_{5}^{4} \oplus u_{9}^{4} \oplus u_{13}^{4}$$

$$z_{3} = x_{9} \oplus x_{10} \oplus x_{12} \oplus u_{3}^{4} \oplus u_{7}^{4} \oplus u_{11}^{4} \oplus u_{15}^{4}$$

$$z^{final} = z_{2} + z_{3}$$

3.2 差分密码分析

差分密码分析(Differential Cryptanalysis)是一种密码分析方法,旨在通过观察密码算法在不同输入差分下产生的输出差分,来推断出密码算法的密钥信息。

差分密码分析的基本原理是利用输入差分对输出结果的影响来推断出密钥信息。具体来说,攻击者通过控制明文的差分,观察密文的差分变化,从而得到密钥的一些信息。这种方法的关键在于寻找输入差分和输出差分之间的关联,以便从中提取出有用的信息。

本次实验,我们需要求解出 SM4 的差分分布表,所以我们只需要对其分布情况进行求解,并不需要进行差分攻击。那就变得很简单了,我们根据课本上所提供的方法便可以很快完成对应的任务。

具体的算法为:

$$N_D(x', y') = |\{(x, x^*) \in \Delta(x') : \pi_S(x) \oplus \pi_S(x^*) = y'\}|$$

4 Question1

本问题中,我们主要使用上一次实验所编写的 SPN 脚本来生成本次实验所需要的明文密文对,然 后使用线性分析攻击算法来实现对 16 位全体密钥的破解。下面我对代码做一下详细的分析。

首先分析一下 Random_Generate_Key_Pairs.py 脚本。**该脚本主要完成了 8000 对明文密文对的随机生成,并且将生成结果存储到了 pairs.txt 中**。具体的代码由于过长,我将其放在了附录 1 中,此处主要做一下核心代码部分的分析。

```
def generate_random_binary_string(length=16, times=1):
binary_strings = [] # 创建一个空列表来存储生成的二进制字符串
for _ in range(times): # 循环指定的次数
binary_strings.append(''.join(random.choice('01') for _ in range(length)))
return binary_strings
```

该段代码实现了 01 二进制数的随机生成,规定了长度为 16, times 表示我们需要生成的对数,我们在程序中的最前面也申明了对数为 8000 对。

这段代码,我首先使用了上次实验中所编写的 C++ 文件生成了对应的 exe 可执行文件,然后使用 python 中的 subprocess 库对其进行了调用运算。

然后,我们使用在前面所生成的 8000 个随机明文数,通过 SPN_process 函数的运算,将输出的 密文结果存储到了 processing_strings_list 中,这样我们就拥有了两个长度为 8000 的列表,分别为我们生成的明文和所计算出的密文。

```
with open("./data/pairs.txt", "w") as file:
for original, encrypted in zip(binary_strings_list, processing_strings_list):
file.write(original + "\n")
file.write(encrypted + "\n")
```

然后我们将两个列表的值写人 pairs.txt 中即可,明文密文对所在的文件的格式为:明文 1、密文 1、明文 2、密文 2......

这样我们的明文密文对的生成过程就完成了,运行 py 程序,得到结果如图4.1所示。

图 4.1: 生成 8000 对明文密文对

打开 pairs.txt 文件查看,发现确实写入了 16000 行数据 (如图4.2所示)。

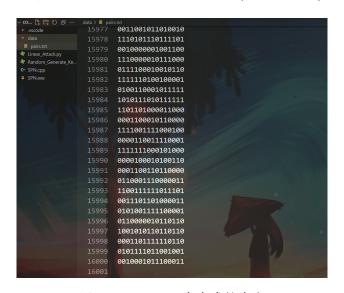


图 4.2: pairs.txt 中生成的内容

我们成功生成了 8000 对明文密文对, 并且将其存储在了 pairs.txt 中。

下面,我们就要进行相应的线性密码分析了。对应的 python 脚本由于长度问题,我将其放在了附录 2 中,此处仅仅对其进行关键部分代码的分析。

```
def solve_reverse_s_box(s_box):
    reverse_box=[0,0,0,0,0,0,0,0,0,0,0,0,0]
    for i in range(len(s_box)):
```

```
reverse_box[s_box[i]]=i
return reverse_box
```

solve_reverse_s_box 函数主要完成了对 S 盒求逆的操作。操作比较简单,此处不再赘述。

然后,我们完成了对代码所需要的列表的初始化操作,并打开 pairs.txt 文件,分别读取对应的明文和密文,存入 original 列表和 encrypted 列表中。

然后,开始我们的核心部分代码——线性攻击。我们主要分为两个步骤对其进行攻击,首先根据 课本上的伪代码,对其中的八位进行攻击;然后根据所构建的另一个表达式,完成剩下八位的攻击。最 后,将分析结果输出。

首先分析一下第一次攻击。

```
def Linear_Attack_First(original, encrypted): #攻击课本中的八位密钥
      for i in range(8000):
          original_temp=original[i]
          encrypted temp=encrypted[i]
          original_temp = original_temp.replace('\n', '').replace('\r', '')
          encrypted_temp = encrypted_temp.replace('\n', '').replace('\r', '')
          encrypted_test2=int(encrypted_temp[4:8],2)
          encrypted_test4=int(encrypted_temp[12:],2)
          x=[None]+[int(i) for i in original_temp]
          for i in range(16):
             for j in range(16):
11
                 u2 = [int(i) for i in format(reverse_pai_s[i ^ encrypted_test2],
                     '04b')]
                 u4 = [int(i) for i in format(reverse_pai_s[j ^ encrypted_test4],
13
                     '04b')]
                 u = [None] * 5 + u2 + [None] * 4 + u4
14
                 z = x[5] ^ x[7] ^ x[8] ^ u[6] ^ u[8] ^ u[14] ^ u[16]
                 if z&1==0:
                     count_list[i][j]=count_list[i][j]+1;
      for i in range(16):
          for j in range(16):
20
             count_list[i][j]=count_list[i][j]-4000
21
             if count list[i][j]<0:</pre>
                 count_list[i][j]=-count_list[i][j]
      temp1=0
25
      temp2=0
26
      count max=0
27
      for i in range(16):
          for j in range(16):
29
             if count_list[i][j]>count_max:
```

```
count_max=count_list[i][j]
temp1=i
temp2=j

return [temp1,temp2]
```

首先,我们分 8000 次来完成对 count 的计数累加。本次读取的 encrypted_test2 为 5-8 位密钥,encrypted_test4 为最后四位密钥。x 就是将我们的初始明文填充之后的结果。然后,通过 for 循环遍历 16*16 数组的每一个位置,通过 S 盒逆运算,将得到的值 u2 和 u4 存入列表 u 中,然后按照课本所提供的表达式进行计算即可,得到 z,通过判断 z 是否为 0,如果是 0 的话就将 count 的值加一。

然后,通过运算, count 减去 8000 的一半,再取绝对值。最后找出最大的那个,也就是偏差最大的,也就是我们所需要的部分密钥了!

另外的第二次攻击大体上与第一次攻击类似,我分析一下不同之处。

```
result1=Linear_Attack_First(original,encrypted)[0]
result2=Linear_Attack_First(original,encrypted)[1]
```

此处代码的含义是,将第一次分析的结果,直接在第二次分析中使用,分别为 result1 和 result2。

```
encrypted_test1=int(encrypted_temp[:4],2)
encrypted_test2=int(encrypted_temp[4:8],2)
encrypted_test3=int(encrypted_temp[8:12],2)
encrypted_test4=int(encrypted_temp[12:],2)
```

此处, 我们需要将四组密文数据均读入。

```
u = [None] + [int(i) for i in format(reverse_pai_s[i ^ encrypted_test1], '04b') + \
format(reverse_pai_s[result1 ^ encrypted_test2], '04b') + \
format(reverse_pai_s[j ^ encrypted_test3], '04b') + \
format(reverse_pai_s[result2 ^ encrypted_test4], '04b')]

z=x[1] ^ x[2] ^ x[4] ^ u[1] ^ u[5] ^ u[9] ^ u[13]

if z&1==0:
    count_list_first[i][j]=count_list_first[i][j]+1

z = x[9] ^ x[10] ^ x[12] ^ u[3] ^ u[7] ^ u[11] ^ u[15]

if z&1==0:
    count_list_second[i][j]=count_list_second[i][j]+1
```

此处,我们通过此处的数据以及前一轮分析的结果构造出列表 u,然后首先计算 z1,叠加 z1;再计算 z2,同样叠加 z2。最后,我们判断 z 是否为 0,去对应的计算 count 的值。

最后,我们需要取的值,为两个 count 之和偏差最大的横纵坐标的值。

两轮分析都结束后,我们只需要将分析结果转为二进制字符串进行输出即可。我们测试了不同的 10 组密钥,每组密钥测试了 5 轮,攻击成功率为 90% 以上。下面仅展示某一次测试时的结果。

```
0101111011001001
0010001011100011
Successfully generate the original number and the encrypted number!
PS F・、と学学、本科、大三上、密码学、Lab、Lab2、codes> & C:/Users/Lenovo/AppData/Local
1101011000111111
PS E:、学学学、本科、大三上、密码学、Lab、Lab2、codes>
```

图 4.3: 线性攻击结果

我们通过运行 py 脚本进行测试,得到了对应的密钥 1101011000111111,这和我们在 SPN 生成脚本中所使用的密钥的后 16 位 (第五轮)相同 (如图4.4所示)。这说明我们的攻击成功了!

```
int main() {
    string plaintext, key;//定义明文和密钥
    cin >> plaintext;//輸入明文
    key='00111010100101001101000111111';
```

图 4.4: SPN 生成中使用的 key

5 Question2

本部分我们需要求解 SM4 表的差分分布表,实际上这只是差分攻击的第一步,因此我们只需要编写几行代码就可以完成对其的求解。我将完整代码放到了附录 3 中,此处挑一些核心代码做一下说明。

```
for i in range(len(s_box)):

for j in range(len(s_box)):

s_box_diff[j^i][s_box[j]^s_box[i]]=s_box_diff[j^i][s_box[j]^s_box[i]]+1
```

该部分代码,我们完成了对 256*256 的差分分布的遍历,对于每一个差分值都进行计算,运算规则为: 遍历 256*256 的矩阵,对于横坐标为 j^i 和纵坐标为 s_box[j]^s_box[i] 的位置进行加一操作,全部循环一轮即可。

然后我们对整个 s_box_diff 的矩阵进行输出,输出到 csv 文件中,就完成了本题的解答。本题的输出为一个 256*256 的矩阵,由于内容过于庞大,我将其放在了文件夹中,就不在此处展示了。

6 附录 1:Random_Generate_Key_Pairs.py

```
import random
import subprocess

counts=8000

# 生成了16位的二进制数
def generate_random_binary_string(length=16, times=1):
binary_strings = [] # 创建一个空列表来存储生成的二进制字符串
for _ in range(times): # 循环指定的次数
```

```
binary_strings.append(''.join(random.choice('01') for _ in range(length)))
      return binary_strings
11
   # 使用函数生成n个长度为16的随机二进制字符串
   binary_strings_list = generate_random_binary_string(16, times=counts)
   # 规定生成密钥的可执行文件路径
   exe_file_path = "./SPN.exe"
   def SPN_Process(times=1):
      processing strings = []
      for i in range(times):
21
         temp_data = subprocess.run([exe_file_path],
             input=binary_strings_list[i].encode(), stdout=subprocess.PIPE,
             stderr=subprocess.PIPE)
         processing_strings.append(temp_data.stdout.decode().strip())
      return processing_strings
   # 确保传递给SPN_Process的times参数与binary_strings_list中的元素数量相匹配
   processing_strings_list = SPN_Process(len(binary_strings_list))
   # 打印列表的内容
   for original, encrypted in zip(binary_strings_list, processing_strings_list):
      print(original)
      print(encrypted)
32
33
   #将8000对明文密文写入pairs.txt中
   with open("./data/pairs.txt", "w") as file:
35
      for original,encrypted in zip(binary_strings_list,processing_strings_list):
         file.write(original + "\n")
         file.write(encrypted + "\n")
  print("Successfully generate the original number and the encrypted number!")
```

7 附录 2:Linear_Attack.py

```
pai_s=[14,4,13,1,2,15,11,8,3,10,6,12,5,9,0,7]
l = 4
m = 4
Nr = 4
```

```
#define some functions
   def solve_reverse_s_box(s_box):
      reverse box=[0,0,0,0,0,0,0,0,0,0,0,0,0,0,0]
      for i in range(len(s_box)):
         reverse_box[s_box[i]]=i
      return reverse_box
19
  #init the s box
  reverse_pai_s=solve_reverse_s_box(pai_s)
   pairs_num=8000 #用于攻击的随机生成密钥对数量
   count_list = [[0 for _ in range(16)] for _ in range(16)] #初始化count数组
18
   original=[0 for _ in range(8000)]
19
   encrypted=[0 for _ in range(8000)]
20
   # 打开文件并读取内容
   with open('data/pairs.txt', 'r') as file:
      lines = file.readlines() # 读取所有行
   # 将两行一对分组
   for i in range(0, len(lines), 2):
      original[i//2]=lines[i]
      encrypted[i//2]=lines[i+1]
30
   def Linear_Attack_First(original,encrypted): #攻击课本中的八位密钥
31
      for i in range(8000):
32
          original_temp=original[i]
          encrypted_temp=encrypted[i]
          original_temp = original_temp.replace('\n', '').replace('\r', '')
          encrypted_temp = encrypted_temp.replace('\n', '').replace('\r', '')
          encrypted_test2=int(encrypted_temp[4:8],2)
          encrypted_test4=int(encrypted_temp[12:],2)
38
         x=[None]+[int(i) for i in original_temp]
39
         for i in range(16):
40
             for j in range(16):
                u2 = [int(i) for i in format(reverse_pai_s[i ^ encrypted_test2],
42
                u4 = [int(i) for i in format(reverse_pai_s[j ^ encrypted_test4],
43
                    '04b')]
                u = [None] * 5 + u2 + [None] * 4 + u4
                z = x[5] ^ x[7] ^ x[8] ^ u[6] ^ u[8] ^ u[14] ^ u[16]
```

```
if z&1==0:
46
                     count_list[i][j]=count_list[i][j]+1;
47
48
      for i in range(16):
49
          for j in range(16):
              count_list[i][j]=count_list[i][j]-4000
              if count_list[i][j]<0:</pre>
52
                 count_list[i][j]=-count_list[i][j]
53
      temp1=0
      temp2=0
       count_max=0
57
      for i in range(16):
58
          for j in range(16):
59
              if count_list[i][j]>count_max:
60
                  count_max=count_list[i][j]
61
                 temp1=i
                 temp2=j
      return [temp1,temp2]
65
66
   count_list_first = [[0 for _ in range(16)] for _ in range(16)]
   count_list_second = [[0 for _ in range(16)] for _ in range(16)]
   count_list_all=[[0 for _ in range(16)] for _ in range(16)]
   def Linear_Attack_Second(original, encrypted):
71
      result1=Linear Attack First(original, encrypted)[0]
      result2=Linear_Attack_First(original,encrypted)[1]
73
      for i in range(8000):
          original_temp=original[i]
          encrypted_temp=encrypted[i]
          original_temp = original_temp.replace('\n', '').replace('\r', '')
          encrypted_temp = encrypted_temp.replace('\n', '').replace('\r', '')
78
          encrypted_test1=int(encrypted_temp[:4],2)
79
          encrypted_test2=int(encrypted_temp[4:8],2)
80
          encrypted_test3=int(encrypted_temp[8:12],2)
          encrypted_test4=int(encrypted_temp[12:],2)
82
          x=[None]+[int(i) for i in original_temp]
83
84
          for i in range(16):
85
              for j in range(16):
86
                 u = [None] + [int(i) for i in format(reverse_pai_s[i ^
```

```
encrypted_test1], '04b') + \
                                             format(reverse_pai_s[result1 ^
88
                                                 encrypted test2], '04b') + \
                                             format(reverse_pai_s[j ^ encrypted_test3],
89
                                                 '04b') + \
                                             format(reverse_pai_s[result2 ^
                                                 encrypted_test4], '04b')]
                  z=x[1] ^x[2] ^x[4] ^u[1] ^u[5] ^u[9] ^u[13]
91
                  if z&1==0:
92
                     count_list_first[i][j]=count_list_first[i][j]+1
93
                  z = x[9] ^ x[10] ^ x[12] ^ u[3] ^ u[7] ^ u[11] ^ u[15]
                  if z&1==0:
95
                     count_list_second[i][j]=count_list_second[i][j]+1
96
97
       for i in range(16):
98
           for j in range(16):
              count_list_first[i][j]=count_list_first[i][j]-4000
              if count_list_first[i][j]<0:</pre>
                  count_list_first[i][j]=-count_list_first[i][j]
       for i in range(16):
           for j in range(16):
              count_list_second[i][j]=count_list_second[i][j]-4000
              if count_list_second[i][j]<0:</pre>
                  count_list_second[i][j]=-count_list_second[i][j]
108
109
       for i in range(16):
          for j in range(16):
              count_list_all[i][j]=count_list_first[i][j]+count_list_second[i][j]
       temp3=0
114
       temp4=0
       count_max=0
       for i in range(16):
117
          for j in range(16):
118
              if count_list_all[i][j]>count_max:
                  count_max=count_list_all[i][j]
120
                  temp3=i
                  temp4=j
       return [temp3,temp4]
```

```
result_list1=Linear_Attack_First(original,encrypted)
    result_list2=Linear_Attack_Second(original,encrypted)
128
129
    def Dec_To_Bin(a):
       match a:
131
           case 0:
               return "0000"
133
           case 1:
134
               return "0001"
           case 2:
               return "0010"
           case 3:
138
               return "0011"
139
           case 4:
140
               return "0100"
141
           case 5:
               return "0101"
           case 6:
144
               return "0110"
145
           case 7:
146
               return "0111"
147
           case 8:
               return "1000"
149
           case 9:
               return "1001"
           case 10:
               return "1010"
           case 11:
               return "1011"
           case 12:
               return "1100"
           case 13:
158
               return "1101"
159
           case 14:
               return "1110"
           case 15:
               return "1111"
164
    result1=Dec_To_Bin(result_list1[0])
    result2=Dec_To_Bin(result_list1[1])
166
   result3=Dec_To_Bin(result_list2[0])
```

```
result4=Dec_To_Bin(result_list2[1])

final_result=result3+result1+result4+result2

print(final_result)
```

8 附录 3:Differential_Attack.py

```
import csv
   s_box=[
      0xd6, 0x90, 0xe9, 0xfe, 0xcc, 0xe1, 0x3d, 0xb7, 0x16, 0xb6, 0x14, 0xc2, 0x28,
          0xfb, 0x2c, 0x05,
      0x2b, 0x67, 0x9a, 0x76, 0x2a, 0xbe, 0x04, 0xc3, 0xaa, 0x44, 0x13, 0x26, 0x49,
          0x86, 0x06, 0x99,
      0x9c, 0x42, 0x50, 0xf4, 0x91, 0xef, 0x98, 0x7a, 0x33, 0x54, 0x0b, 0x43, 0xed,
          0xcf, 0xac, 0x62,
      0xe4, 0xb3, 0x1c, 0xa9, 0xc9, 0x08, 0xe8, 0x95, 0x80, 0xdf, 0x94, 0xfa, 0x75,
          0x8f, 0x3f, 0xa6,
      0x47, 0x07, 0xa7, 0xfc, 0xf3, 0x73, 0x17, 0xba, 0x83, 0x59, 0x3c, 0x19, 0xe6,
          0x85, 0x4f, 0xa8,
      0x68, 0x6b, 0x81, 0xb2, 0x71, 0x64, 0xda, 0x8b, 0xf8, 0xeb, 0x0f, 0x4b, 0x70,
          0x56, 0x9d, 0x35,
      0x1e, 0x24, 0x0e, 0x5e, 0x63, 0x58, 0xd1, 0xa2, 0x25, 0x22, 0x7c, 0x3b, 0x01,
          0x21, 0x78, 0x87,
      0xd4, 0x00, 0x46, 0x57, 0x9f, 0xd3, 0x27, 0x52, 0x4c, 0x36, 0x02, 0xe7, 0xa0,
          0xc4, 0xc8, 0x9e,
      Oxea, Oxbf, Ox8a, Oxd2, Ox40, Oxc7, Ox38, Oxb5, Oxa3, Oxf7, Oxf2, Oxce, Oxf9,
          0x61, 0x15, 0xa1,
      0xe0, 0xae, 0x5d, 0xa4, 0x9b, 0x34, 0x1a, 0x55, 0xad, 0x93, 0x32, 0x30, 0xf5,
          0x8c, 0xb1, 0xe3,
      0x1d, 0xf6, 0xe2, 0x2e, 0x82, 0x66, 0xca, 0x60, 0xc0, 0x29, 0x23, 0xab, 0x0d,
          0x53, 0x4e, 0x6f,
      0xd5, 0xdb, 0x37, 0x45, 0xde, 0xfd, 0x8e, 0x2f, 0x03, 0xff, 0x6a, 0x72, 0x6d,
          0x6c, 0x5b, 0x51,
      0x8d, 0x1b, 0xaf, 0x92, 0xbb, 0xdd, 0xbc, 0x7f, 0x11, 0xd9, 0x5c, 0x41, 0x1f,
          0x10, 0x5a, 0xd8,
      0x0a, 0xc1, 0x31, 0x88, 0xa5, 0xcd, 0x7b, 0xbd, 0x2d, 0x74, 0xd0, 0x12, 0xb8,
          0xe5, 0xb4, 0xb0,
      0x89, 0x69, 0x97, 0x4a, 0x0c, 0x96, 0x77, 0x7e, 0x65, 0xb9, 0xf1, 0x09, 0xc5,
17
          0x6e, 0xc6, 0x84,
      0x18, 0xf0, 0x7d, 0xec, 0x3a, 0xdc, 0x4d, 0x20, 0x79, 0xee, 0x5f, 0x3e, 0xd7,
```

```
0xcb, 0x39, 0x48
  ]
19
   s_box_diff = [[0 for _ in range(len(s_box))] for _ in range(len(s_box))]
21
   for i in range(len(s_box)):
      for j in range(len(s_box)):
          s_box_diff[j^i][s_box[j]^s_box[i]] = s_box_diff[j^i][s_box[j]^s_box[i]] + 1
26
   with open('result/s_box_diff.csv', 'w', newline='') as csvfile:
      writer = csv.writer(csvfile)
30
      for i in range(len(s_box)):
31
          row = [s_box_diff[i][j] for j in range(len(s_box))]
32
          writer.writerow(row)
33
```