INE5451 - Introdução à Criptoanálise (2015/1)

Trabalho 1, Grupo 3, Modelo 6
Gustavo Zambonin, Ranieri Schroeder Althoff

Nota: todos os algoritmos apresentados neste documento podem ser encontrados também neste repositório.

Questão 1

O processo utilizado para obter as frequências com relação à língua inglesa é listado abaixo:

- Lembrando: os cálculos realizados mostram que, desde que exista uma boa quantidade de textos (um livro já é o suficiente para construção do vocabulário diversificado), as frequências sempre respeitarão um intervalo intrínseco. Os valores descritos representam apenas uma aproximação das frequências da língua inglesa moderna. Comparações podem ser feitas através de dados presentes aqui, assim como resultados de outros arquivos analisados por este código.
- Obteve-se uma edição completa do livro *Les Misérables*, traduzida para inglês, disponível gratuitamente através do *Project Gutenberg*.
- O código utilizado para processar o arquivo é comentado logo abaixo.

```
1  #!/usr/bin/env python
2
3  import pprint
4  import string
5  import sys
6
7  allowed_letters = string.ascii_lowercase
8  digraph_index = {}
9  letter_index = {l : 0 for l in allowed_letters}
10
11 def process_text(text):
12     for word in text.lower().split():
13         if len(word) > 1:
14             digraphs = tuple(word[x:x + 2] for x in range(len(word) - 1))
15             for dig in digraphs:
16                 if dig[0] in allowed_letters and dig[1] in allowed_letters:
17                     letter_index[dig[0]] += 1
18                     digraph_index[dig] = digraph_index[dig] + 1 \
19                             if dig in digraph_index else 1
20
21         if word[-1] in allowed_letters:
22             letter_index[word[-1]] += 1
23
24 def process_data():
25     total = sum(digraph_index.values())
26
27     for let in letter_index:
28         letter_index[let] /= total
29
30     for dig in digraph_index:
31         digraph_index[dig] /= total
32
33     return sorted(letter_index.items(), key=lambda o: o[1], reverse=True), \
34         sorted(digraph_index.items(), key=lambda o: o[1], reverse=True)[:10]
35
36 if __name__ == '__main__':
37     for arg in sys.argv[1:]:
38         with open(arg) as raw:
39             for line in raw:
40                 process_text(line)
41
42     pprint.pprint(process_data())
```

- Linha 12: garante que todas as letras serão convertidas para suas versões minúsculas na lista de palavras.
- Linha 14: cria todos os digramas possíveis para uma palavra de tamanho válido.
- Linha 17: adiciona na contagem de letras a primeira letra de cada digrama, desconsiderando repetições.
- Linhas 18-19: trata de adicionar um digrama novo ou incrementar a contagem de um existente.

- Linhas 21-22: a última letra de cada palavra é finalmente tratada, visto que esta nunca será a primeira letra de um digrama, e portanto, nunca seria contada.
- Linhas 33-34: funções *inline* que ordenam as estruturas de dados por suas chaves numéricas.
- Linha 37: habilita a entrada de múltiplos arquivos de texto para análise conjunta.

a. A frequência relativa do alfabeto A-Z em inglês segue abaixo:

e	$\approx 12.900494344206656\%$	**d**	$\approx 4.045562489940431\%$	**b**	$\approx 1.5355173283924984\%$
t	$\approx 9.228622978258158\%$	**l**	$\approx 3.9625337394346226\%$	**y**	$\approx 1.3292832976368709\%$
a	$\approx 8.495812955201934\%$	**u**	$\approx 2.7658412228484253\%$	**v**	$\approx 1.0720053151572956\%$
o	$\approx 7.534992892953012\%$	**c**	$\approx 2.7387961613301724\%$	**k**	$\approx 0.5378550590525007 5\%$
i	$\approx 7.176120980751902\%$	**m**	$\approx 2.401041625930986\%$	**j**	$\approx 0.2396595862393739\%$
h	$\approx 7.16191923611903\%$	**w**	$\approx 2.2774658653863966\%$	**x**	$\approx 0.15218507219631834\%$
n	$\approx 6.63501392800086\%$	**f**	$\approx 2.264293232683442\%$	**q**	$\approx 0.10402263387614184\%$
s	$\approx 6.014994572463682\%$	**g**	$\approx 1.8497669473123507\%$	**z**	$\approx 0.07668942101751178\%$
r	$\approx 5.7699836041887324\%$	**p**	$\approx 1.7295255094206964\%$		

b. A frequência relativa de digramas na língua inglesa é apresentada abaixo, filtrando apenas os 10 resultados mais comuns. O conjunto total não será mostrado pois, além de ser demasiadamente extenso, não existe, neste arquivo, uma representação de todas as palavras presentes no vocabulário moderno da língua inglesa, utilizadas de modo contextualmente válido, impossibilitando a construção de todos os digramas válidos.

th	$\approx 4.276658660240072\%$
he	$\approx 4.0442594323565895\%$
in	$\approx 2.2615934381157863\%$
er	$\approx 2.111202678574571\%$
an	$\approx 1.9470851796970966\%$
re	$\approx 1.8199745678214054\%$
ed	$\approx 1.4616211932810376\%$
on	$\approx 1.4461346155683348\%$
ha	$\approx 1.4253685227263014\%$
at	$\approx 1.386953765023363\%$

Questão 2

- Seja $y = DES(c, K)$ a função que encripta o texto c com a chave K. c, por sua vez, será dividido em duas metades: L_i, a parte esquerda, e R_i, a parte direita, para cada rodada i. Portanto, existe uma função $DES(L_iR_i, K)$ que caracteriza uma rodada da rede de Feistel.
- Hipótese: $DES(L_iR_i, K)$ é complemento de $DES(\overline{L_iR_i}, \overline{K})$ e vice-versa.
- Em $DES(L_iR_i, K)$:
  - $L_{i+1} = R_i$
  - $R_{i+1} = L_i \oplus f(R_i, K_i)$
- Em $DES(\overline{L_iR_i}, \overline{K})$:
  - $\overline{L_{i+1}} = \overline{R_i}$
  - $\overline{R_{i+1}} = \overline{L_i} \oplus f(\overline{R_i}, \overline{K_i})$
  - A função $f(R_i, K_i)$ combina os bits de R_i e K_i utilizando o operador lógico $\oplus$. Como a operação de $\oplus$ é comutativa e associativa, $f(R_i, K_i) = f(\overline{R_i}, \overline{K_i})$. (É importante lembrar que a expansão de bits em R_i não afeta seu complemento pois apenas duplica bits e troca-os de lugar.)
  - Portanto, segue que $\overline{R_{i+1}} = \overline{L_i} \oplus f(R_i, K_i)$.
  - Porém, deseja-se obter $\overline{R_{i+1}} = \overline{L_i \oplus f(R_i, K_i)}$ a partir dos resultados acima. Isso pode ser mostrado através da tabela verdade entre os termos:

$\overline{L_i}$	$f(R_i, K_i)$	$\overline{L_i \oplus f(R_i, K_i)}$	$\overline{L_i} \oplus f(R_i, K_i)$
0	0	1	1
0	1	0	0
1	0	0	0
1	1	1	1

- Todas as operações que compõem as funções $DES(L_iR_i, K)$ e $DES(\overline{L_iR_i}, \overline{K})$ foram provadas complementos mútuos de acordo com as regras impostas, portanto a hipótese é verdadeira.
- Por fim, por indução fraca, prova-se que o complemento cifrado funciona em todas as rodadas do DES, de número arbitrário, usando o procedimento acima.

□

Questão 3

A máquina Enigma apresentada tem as seguintes características:

- *Steckerbrett* sem cabos conectados;
- *Ringstellung* com a configuração **AAA**;
- O texto cifrado é datado de maio de 1942.

Destas informações, é possível inferir os seguintes aspectos para início da criptoanálise:

- Não é possível aplicar um método que utilize um ataque de texto em claro conhecido (Bomba de Turing).
- Não é possível aplicar um método que dependa de mais de um texto cifrado (*females* de Zygalski).
- Não é possível aplicar um método que aproveite da repetição da chave secundária, já que isto não acontecia mais em 1942 (folhas perfuradas de Zygalski, método das características).
- Existem cinco modelos de rotor (*walzen* I, II, III, IV e V, presentes nas Enigma I e M3), visto que não se foi discutido em profundidade o funcionamento dos rotores de *notch* duplo utilizados pela Marinha alemã.
- Portanto, o número de testes para descobrir a ordem e escolha dos rotores (*walzenlage*) é $\binom{5}{3} = 60$.
- Ademais, deseja-se descobrir também o *grundstellung* (posição em que os rotores foram encaixados inicialmente). Existem 26^3 possibilidades, visto que todos os rotores podem ser encaixados com todas as letras apontando para a janela da carcaça da máquina.
- Logo, o número de possibilidades é igual a $60 \times 26^3 = \mathbf{1054560}$.
- O texto cifrado fornecido tem cerca de 400 caracteres. De acordo com o estudo de *Regan* sobre um ataque à Enigma usando índice de coincidência (IC), existe uma probabilidade muito baixa de que o texto com o maior IC não seja o correto para este número de caracteres. Deste modo, essa estratégia foi escolhida. É válido comentar que não se obteve sucesso em construir um método não-trivial caso existissem múltiplos textos de tamanho pequeno cifrados com diversas chaves.

O algoritmo que simula a Enigma é apresentado abaixo:

```
#!/usr/bin/env python

import string
import sys

alphabet = string.ascii_uppercase
desloc = alphabet[1:] + alphabet[:1]
shift_up = {alphabet[i]:desloc[i] for i in range(len(alphabet))}

class Steckerbrett:
    def __init__(self, *args):
        self.map = {}

        for arg in args:
            if arg[0] in self.map or arg[1] in self.map:
                raise KeyError('Same letter used twice in plugboard')

            self.map[arg[0]] = arg[1]
            self.map[arg[1]] = arg[0]

    def swap(self, letter):
        if letter in self.map:
            return self.map[letter]
        return letter

class Umkehrwalze:
    def __init__(self, wiring):
```

```
28          assert isinstance(wiring, str)
29          self.wiring = wiring
30
31      def reflect(self, letter):
32          return self.wiring[alphabet.index(letter)]
33
34  class Walzen:
35      def __init__(self, index, grundstellung, ringstellung):
36          if index == 'I':
37              self.wiring = 'EKMFLGDQVZNTOWYHXUSPAIBRCJ'
38              self.notch = 'Q'
39          elif index == 'II':
40              self.wiring = 'AJDKSIRUXBLHWTMCQGZNPYFVOE'
41              self.notch = 'E'
42          elif index == 'III':
43              self.wiring = 'BDFHJLCPRTXVZNYEIWGAKMUSQO'
44              self.notch = 'V'
45          elif index == 'IV':
46              self.wiring = 'ESOVPZJAYQUIRHXLNFTGKDCMWB'
47              self.notch = 'J'
48          elif index == 'V':
49              self.wiring = 'VZBRGITYUPSDNHLXAWMJQOFECK'
50              self.notch = 'Z'
51
52          if isinstance(self.notch, str) and len(self.notch) == 1:
53              self.notch = alphabet.index(self.notch)
54
55          if isinstance(ringstellung, str) and len(ringstellung) == 1:
56              self.ringstellung = alphabet.index(ringstellung)
57
58          if isinstance(grundstellung, str) and len(grundstellung) == 1:
59              self.grundstellung = alphabet.index(grundstellung)
60
61      def wire(self, forward):
62          permut = {}
63          for i in range(len(self.wiring)):
64              permut[alphabet[i]] = self.wiring[i]
65          if not forward:
66              return {v: k for k, v in permut.items()}
67          return permut
68
69      def move(self, char, steps, permutation):
70          if steps < 0:
71              permutation = {v: k for k, v in permutation.items()}
72              steps *= -1
73          for i in range(steps):
74              char = permutation[char]
75          return char
76
77  class Enigma:
78      def __init__(self, rotors, plugboard=None):
79          assert plugboard is None or isinstance(plugboard, Steckerbrett)
80
81          assert isinstance(rotors, tuple)
82          for index in range(len(rotors)):
83              assert isinstance(rotors[index], Walzen)
84
85          self.plugboard = plugboard
86          self.rotors = rotors
87          self.reflector = Umkehrwalze(wiring='YRUHQSLDPXNGOKMIEBFZCWVJAT')
88
89      def cipher(self, message):
90          assert isinstance(message, str)
91
92          message = message.upper()
93          ciphered = ''
94
95          p1 = self.rotors[0].grundstellung
96          p2 = self.rotors[1].grundstellung
97          p3 = self.rotors[2].grundstellung
98
99          n1 = self.rotors[0].notch
100         n2 = self.rotors[1].notch
101
102         r1 = self.rotors[0].ringstellung
103         r2 = self.rotors[1].ringstellung
```

```
104         r3 = self.rotors[2].ringstellung
105
106         m1 = (n1 - p1) % 26
107         m2 = m1 + 26*((n2 - p2 - 1) % 26) + 1
108
109         i1 = p1 - r1 + 1
110
111         for j in range(len(message)):
112             letter = message[j]
113
114             k1 = int((j - m1 + 26) / 26)
115             k2 = int((j - m2 + 650) / 650)
116
117             i2 = p2 - r2 + k1 + k2
118             i3 = p3 - r3 + k2
119
120             letter = self.rotors[0].move(letter, i1 + j, shift_up)
121             letter = self.rotors[0].move(letter, 1, self.rotors[0].wire(True))
122             letter = self.rotors[0].move(letter, -i1 - j, shift_up)
123             letter = self.rotors[1].move(letter, i2, shift_up)
124             letter = self.rotors[1].move(letter, 1, self.rotors[1].wire(True))
125             letter = self.rotors[1].move(letter, -i2, shift_up)
126             letter = self.rotors[2].move(letter, i3, shift_up)
127             letter = self.rotors[2].move(letter, 1, self.rotors[2].wire(True))
128             letter = self.rotors[2].move(letter, -i3, shift_up)
129             letter = self.reflector.reflect(letter)
130             letter = self.rotors[2].move(letter, i3, shift_up)
131             letter = self.rotors[2].move(letter, 1, self.rotors[2].wire(False))
132             letter = self.rotors[2].move(letter, -i3, shift_up)
133             letter = self.rotors[1].move(letter, i2, shift_up)
134             letter = self.rotors[1].move(letter, 1, self.rotors[1].wire(False))
135             letter = self.rotors[1].move(letter, -i2, shift_up)
136             letter = self.rotors[0].move(letter, i1 + j, shift_up)
137             letter = self.rotors[0].move(letter, 1, self.rotors[0].wire(False))
138             letter = self.rotors[0].move(letter, -i1 - j, shift_up)
139
140             ciphered += letter
141
142         return ciphered
143
144 if __name__ == '__main__':
145     # INSERT FROM RIGHT TO LEFT LIKE NORMAL ENIGMA
146     rotors = (Walzen(index='I',   grundstellung='C', ringstellung='A'),
147               Walzen(index='II',  grundstellung='S', ringstellung='A'),
148               Walzen(index='III', grundstellung='F', ringstellung='A'))
149
150     print(Enigma(rotors=rotors).cipher(sys.argv[1]))
```

- Linhas 6-8: Definição do deslocamento em relação ao alfabeto na sua ordem normal.
- Linhas 10-24: Classe que define a *plugboard*, não utilizada neste programa, que contém apenas o método de troca de letras.
- Linhas 26-32: Classe que define um refletor e sua permutação simples de acordo com a definição.
- Linhas 34-75: Classe que define um rotor, levando em conta seu *ringstellung* e *grundstellung*, além do seu nome, para que sejam atribuídas suas características intrínsecas. O método *wire* constrói as permutações de cada rotor. O método *move* guia as letras ao passo dos rotores.
- Linhas 77-142: A classe de nível mais alto, que define a Enigma, sua composição e seu propósito, o método *cipher*, que implementa matematicamente a lógica de movimentação de rotores.
- Linhas 144-150: Uma instância da máquina com os rotores certos para a mensagem cifrada (mais abaixo) é criada, que consegue cifrar qualquer texto enviado por linha de comando com estas configurações.

O algoritmo responsável por analisar o IC de cada possível texto é mostrado abaixo:

```
1 #!/usr/bin/env python
2
3 import string
4 import sys
5 from enigma import Enigma, Walzen
6
7 rotors=('I', 'II', 'III', 'IV', 'V')
```

```
8
9  def index_of_coincidence(text):
10     def freq(i):
11         c = text.count(i)
12         return c * (c - 1)
13
14     l = len(text)
15     return sum(map(freq, string.ascii_uppercase)) / (l * (l - 1))
16
17 def main(text):
18     for i in rotors:
19         for j in rotors:
20             if i == j:
21                 continue
22             for k in rotors:
23                 if i == k or j == k:
24                     continue
25                 for l in string.ascii_uppercase:
26                     for m in string.ascii_uppercase:
27                         for n in string.ascii_uppercase:
28                             deciphered = Enigma(rotors=
29                             (Walzen(index='%s' % i, ringstellung='A', grundstellung='%s' % l),
30                              Walzen(index='%s' % j, ringstellung='A', grundstellung='%s' % m),
31                              Walzen(index='%s' % k, ringstellung='A', grundstellung='%s' % n))
                                    ).cipher(text)
32                             yield ('(%s%s%s␣%s␣%s␣%s):␣%s' % (
33                                 l, m, n, i, j, k,
34                                 deciphered
35                             ), index_of_coincidence(deciphered))
36
37 if __name__ == '__main__':
38     ics = main(sys.argv[1])
39
40     filtered_ics = filter(lambda v: 0.05 < v[1] <= 0.07, ics)
41
42     sorted_ics = sorted(filtered_ics, key=lambda o: o[1], reverse=True)
43
44     with open('result.out', 'w') as fp:
45         for i in sorted_ics:
46             fp.write(str(i))
```

- Linhas 9-15: calcula $IC = \sum_{i=A}^{Z} \frac{f_i \times (f_i - 1)}{n \times (n-1)}$ tal como descrito por *Smart* em seu livro, p. 73.
- Linhas 37-40: Armazena o texto em um gerador, que será tratado para os resultados desejados.
- Linha 45: Filtra o conjunto de ICs. Por precaução, a amplitude do intervalo continua sendo alta.
- Linha 46: Ordena os resultados pelo IC, para que o resultado seja facilmente distinguível.
- Linhas 48-50: Armazena o resultado em disco para acesso posterior e pesquisa, se necessário.

O resultado final foi modificado apenas para adição de espaços. O texto original foi adicionado abaixo para comparação de conteúdo, visto que a pontuação e ênfase contextual podem ser ambíguas. Após o texto decifrado, o IC apresenta um valor muito próximo ao IC da língua inglesa (0.068). Por conta da quantidade de caracteres no texto cifrado, nenhuma outra possibilidade de texto ultrapassou a marca de 0.043 em relação ao IC.

```
('(FSC III II I): AS ECONOMICS IS KNOWN AS THE MISERABLE SCIENCE SOFTWARE ENGINEERING SHOULD BE KNOWN AS THE DOOMED
DISCIPLINE DOOMED BECAUSE IT CANNOT EVEN APPROACH ITS GOAL SINCE ITS GOAL IS SELFCONTRADICTORY SOFTWARE ENGINEERING
OF COURSE PRESENTS ITSELF AS ANOTHER WORTHY CAUSE BUT THAT IS EYEWASH IF YOU CAREFULLY READ ITS LITERATURE AND ANAL-
YSE WHAT ITS DEVOTEES ACTUALLY DO YOU WILL DISCOVER THAT SOFTWARE ENGINEERING HAS ACCEPTED AS ITS HOW TO PROGRAM IF
YOU CANNOT EDSGER DIJKSTRA', 0.06592372083555963)
```

As economics is known as "The Miserable Science", software engineering should be known as "The Doomed Discipline", doomed because it cannot even approach its goal since its goal is self-contradictory. Software engineering, of course, presents itself as another worthy cause, but that is eyewash: if you carefully read its literature and analyse what its devotees actually do, you will discover that software engineering has accepted as its charter "How to program if you cannot.".

Edsger W. Dijkstra

Finalmente, o *grundstellung* é igual a **F, S, C** e o *walzenlage* é igual a **III, II, I**.

Questão 4

O ataque diferencial da SPN será discutido logo abaixo. O código responsável por cifrar uma mensagem através da SPN será omitido neste documento, mas pode ser encontrado aqui. Este ataque está direcionado para uma certa subchave e *input-xor* definido, para que um paralelo com *Stinson* (p. 94) seja realizado.

```
1  import random
2  from q4_spn import SPN
3
4  bits = ["{0:0>4}".format(bin(i)[2:]) for i in range(16)]
5  key = '00111010100101001101011000111111'
6
7  def gen_quadruples(chosen_xor='1011'):
8          tuples = []
9          for i in bits:
10                 u = ''
11                 for j in range(4):
12                         u += SPN().xor(chosen_xor[j], i[j])
13                 tuples += [(i, u, SPN().sbox[i], SPN().sbox[u])]
14
15         quadruples = []
16         for each in tuples:
17                 for i in range(16):
18                         for j in range(4):
19                                 bits1 = bits[i] + each[0] + '1001' + bits[j]
20                                 bits2 = bits[i] + each[1] + '1001' + bits[j]
21                                 quadruples.append((bits1, bits2, SPN().encode(bits1, key),
22                                                                  SPN().encode(bits2, key)))
23
24         return random.sample(quadruples, 200)
25
26 def multi_xor(arg1, arg2):
27         xor = ''
28         for i in range(len(arg1)):
29                 xor += SPN().xor(arg1[i], arg2[i])
30         return xor
31
32 def diff_attack(quadruples=gen_quadruples()):
33         inverse_sbox = {v: k for k, v in SPN().sbox.items()}
34
35         tentatives = []
36         for i in range(16):
37                 for j in range(16):
38                         tentatives.append([bits[i], bits[j]])
39
40         count = [0] * len(tentatives)
41
42         for each in quadruples:
43                 if each[2][0:4] == each[3][0:4] and each[2][8:12] == each[3][8:12]:
44                         for key in tentatives:
45                                 v2 = multi_xor(key[0], each[2][4:8])
46                                 v4 = multi_xor(key[1], each[2][12:16])
47                                 v2_star = multi_xor(key[0], each[3][4:8])
48                                 v4_star = multi_xor(key[1], each[3][12:16])
49
50                                 u2 = inverse_sbox[v2]
51                                 u4 = inverse_sbox[v4]
52                                 u2_star = inverse_sbox[v2_star]
53                                 u4_star = inverse_sbox[v4_star]
54
55                                 u2_line = multi_xor(u2, u2_star)
56                                 u4_line = multi_xor(u4, u4_star)
57
58                                 if u2_line == '0110' and u4_line == '0110':
59                                         count[tentatives.index(key)] += 1
60
61         return tentatives[count.index(max(count))]
62
63 assert diff_attack() == ['0110', '1111']
```

- Linhas 7-24: Criam todas as quádruplas de texto em claro, e cifrado, disponíveis sendo um *input-xor* fixo. Após o procedimento, 200 instâncias aleatórias, sem repetição, são escolhidas e repassadas ao próximo método. Relembrando: a quádrupla consiste de um x e um x^* amarrados por um *input-xor*, e seus resultados y e y^* após serem codificados pela SPN.
- Linhas 26-30: Declaração de uma operação $\oplus$ *multibit* para praticidade nas iterações do ataque diferencial.

- Linhas 32-61: O método principal consiste da construção de todas as possibilidades de subchave para que a contagem seja realizada, e da lógica de ataque probabilístico. Após o filtro de pares inúteis, os bits v' são manipulados de modo que passam pela permutação inversa da caixa-S e, se forem iguais aos resultados desejados, contribuem para um contador de subchave correta.
- Linha 63: Como este caso é fixo, porém probabilístico, o algoritmo não mostrará sucesso em todas as situações.

Entretanto, com esta implementação, não foi possível concluir que $T = c \times e^{-1}$, onde $e \approx 38$ e $50 < T < 100$. A quantidade observada foi um pouco maior – cerca de 200 quádruplas foram necessárias para obtenção da subchave em 80% das vezes. Com $T = 100$, inferiu-se que em menos de 50% das vezes isso acontece.