CSEC 507 - Applied Cryptology
20221
Homework 1 Solutions

ÖZTÜRK, Muhammed Tayyip
e238080@metu.edu.tr

December 1, 2022

---

1. **Caesar's Cipher**
   Source code for observing all possibilities of encoding is placed in **caesarCipher.py** file in **src** folder.

   (a) i. BXADQRDBTQHSX
   ii. AWZCPQCASPGRW
   iii. ZVYBOPBZROFQV
   iv. DZCFSTFDVSJUZ

   (b) According to the encoding section above, k denotes the number of right shifts needed to decode the cipher texts:
   i. Plaintext: APPLIEDCRYPTOLOGY, k=13
   ii. Plaintext: CAESARCIPHER, k=24

2. **XOR Operation**

   (a) i. EEEEEEEE
   ii. 33333333
   iii. A2C07A
   iv. 11111111
   v. EEEEEEEE

3. **Frequency Analysis**

   (a) Frequency analysis has been done by running a Python script. Source file and outputs are attached. Moreover, ngrams of both text, where n $\in \mathbb{Z}^+$ and $2 \leq n \leq 4$, are also in the folder named *engciphertext_ngrams, turciphertext_ngrams* folders respectively under the *output* folder. Those outputs are obtained by running the shell command **python3 src/frequencyAnalysis.py** inside the HW1 directory.

   While analyzing cipher texts:
   `https://www3.nd.edu/~busiforc/handouts/cryptography/Letter%20Frequencies.html` is utilized to compare the analysis we have with for English Cipher Text,
   (Serengil and Akin, Attacking Turkish texts encrypted by homophonic cipher) for Turkish Cipher Text.

   Table 1 and Table 2 are the frequency analysis results for English Cipher Text and Turkish Cipher Text respectively:

Letter	Count	Ratio
n	11.86%	315
c	7.38%	250
y	7.38%	196
o	7.27%	193
z	7.23%	192
h	6.82%	181
e	6.67%	177
b	6.29%	167
i	5.39%	143
d	4.18%	111
x	4.14%	110
t	4.03%	107
v	3.54%	94
r	3.13%	83
s	2.79%	74
p	2.18%	58
k	2.0%	53
l	1.85%	49
a	0.98%	26
j	0.83%	22
u	0.72%	19
f	0.68%	18
q	0.49%	13
g	0.15%	4
m	0.0%	0
w	0.0%	0
Total	˜100.0%	2655

Table 1: Letters and corresponding counts and frequency of existence for English Cipher Text

Letter	Count	Ratio
e	12.57%	391
m	8.45%	263
ç	8.33%	259
g	7.91%	245
b	7.2%	224
ö	5.88%	183
c	5.18%	161
l	4.79%	149
y	4.28%	133
f	4.18%	130
v	3.89%	121
t	3.54%	110
j	2.7%	84
d	2.6%	81
s	2.44%	76
k	2.09%	65
ş	1.83%	57
a	1.54%	48
z	1.45%	45
ğ	1.38%	43
r	1.32%	41
ı	1.25%	39
h	1.19%	37
n	0.96%	30
o	0.9%	28
u	0.87%	27
p	0.74%	23
ü	0.42%	13
i	0.13%	4
Total	˜100.0%	3111

Table 2: Letters and corresponding counts and frequency of existence for Turkish Cipher Text

(b) Firstly, I started by examining the numbers and symbols. I recognized that plaintext is something about A5/1 algorithm. I looked for resembling articles and tried to place corresponding letters in the article, starting from conjunctions and word pairs inside parenthesis. When I stuck to continue, started looking at the most frequent trigrams in Turkish Language from On the Cryptographic Patterns and Frequencies in Turkish Language (Dalkilic, Mehmet & Dalkılıç, Gökhan. (2002)). With the help of the paper, started with the most used trigram ”LAR”. Letter by letter, by replacing cipher keys with a letter which would not disrupt the meaning of the possible plaintext, a meaningful plaintext is obtained. It is also attached to the submission with a name **turplaintext.txt** in the **output** folder.

Alphabet	A	B	C	Ç	D	E	F	G	Ğ	H	I	İ	J	K	L	M	N	O	Ö	P	R	S	Ş	T	U	Ü	V	Y	Z
Key	E	S	A	I	T	M	N	R	U	O	L	B	İ	C	Ç	F	Ö	D	Ü	Ğ	G	J	Z	Y	V	Ş	P	K	H

Table 3: Corresponding Ciphertext Alphabet for Turkish Alphabet

(c) As well as it has been done in the decryption of Turkish cipher text, replaceLetter method has been used to replace each letter and observe if anything meaningful appears. However, this is not done randomly. ngrams for n $\in$ 2,3,4 are produced from the ciphertext. Starting from ngrams where n=4 at first, frequency analysis has been observed. There was a trigram pattern also exists in various tetragrams which is ”zxv”. Besides this, letters which are the same letter and consecutive to each other is queried and their frequency has been obtained by getDoubles() function. These ratios are also used to derive tetragrams to corresponding plaintexts by observing the English tetragram ratios from `https://www3.nd.edu/~busiforc/handouts/cryptography/Letter%20Frequencies.html`. Word pairs in parenthesis are also ease the way predict some unknown cipher keys, as their word structure resembles each other.

Alphabet	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Key	E	A	I	R	N	L	K	D	O	M	U	T	S	B	Y	V	G	Z	H	C	P	F	J	Q	X	W

Table 4: Corresponding Ciphertext Alphabet for English Alphabet

(d) The frequency analysis made in the first part does not completely matches with the plaintext corresponding. This result may be affected by miscellaneous factors:

i. Since hose both plaintext has a genre, a specific theme, some keywords are repeated.

ii. Even though they would not have any specific genre, their length are not adequate to have a homogeneous distribution for letters similar to the research results. Thus, some letters are more probable to be encountered.

iii. Even usage of a conjunction instead of its counterpart, some letters can be skewed, may affect the ratio of each letter in the text significantly.

iv. The ratios we have utilized for the original English and Turkish alphabets may not be suitable for referencing, since they may not be adequate as well depending on the homogeneity and length of the texts ratios were derived.

v. Counting may be erroneous in the first part. Ratios may not be calculated properly.

References

1. https://www3.nd.edu/~busiforc/handouts/cryptography/Letter%20Frequencies.html

2. Serengil, S. I., & Akin, M. (n.d.). Attacking Turkish texts encrypted by homophonic cipher. Retrieved November 30, 2022, from https://www.researchgate.net/profile/Sefik-Serengil/publication/262407127_Attacking_Turkish_texts_encrypted_by_homophonic_cipher/links/5667429408aea62726ee72fe/Attacking-Turkish-texts-encrypted-by-homophonic-cipher.pdf

3. Dalkilic, Mehmet & Dalkılıç, Gökhan. (2002). On the Cryptographic Patterns and Frequencies in Turkish Language. 144-153. 10.1007/3-540-36077-8_14. From https://www.researchgate.net/publication/221581391_On_the_Cryptographic_Patterns_and_Frequencies_in_Turkish_Language