Abdoul-Nourou Yigo
Dr. Jiazhen Zhou
Cryptography and Cloud Security
February 2019

Cryptanalysis with Mono-alphabetic cipher

Letters from the cipher text	Letter frequency in the ciphertext (%)	Relative frequency in the English language (%)	Key for the First Iteration
A	6.745	8.167	N
B	2.326	1.492	G
C	0.075	2.782	Z
D	0.160	4.253	X
E	5.575	12.702	R
F	0.891	2.228	V
G	1.908	2.015	P
H	5.460	6.094	S
I	9.154	6.966	T
J	1.835	0.153	F
K	6.782	0.772	I
L	7.853	4.025	O
M	0.118	2.406	X
N	2.581	6.749	Y
O	5.856	7.507	H
P	12.934	1.929	E
Q	2.547	0.095	M
R	4.425	5.987	D

S	4.322	6.327	L
T	1.381	9.056	B
U	2.241	2.758	W
V	1.170	0.978	K
W	7.812	2.360	A
X	0.187	0.150	X
Y	2.503	1.974	C
Z	3.156	0.074	U

Table1: One Letter Frequencies from the CipherText

Most popular 2 Combinaison of Letter from The Ciphertext	String Frequencies in the Ciphertext (%)	Key	Key for the last Iteration
IO	3.523	TH	TH
OP	3.484	HE	HE
KA	2.550	IN	IN
PE	2.388	ER	ER
PR	2.136	ED	ED
EP	1.662	BE	RE
LZ	1.637	IT	OU
WA	1.605	BY	AU
WI	1.308	AN	AN
IL	1.226	OR	TO

Table2: Two Letters from the CipherText

Comparing the Ciphertext letters frequency with the relative letters frequency

The first of step of deciphering the ciphertext was possible with the obtention of the letter frequencies calculated from the ciphertext. However, even though those frequencies were an appropriate starting point the decrypted text was a little far from readability. The characters (one letter and two letters) needed to be adjusted accordingly. For instance, Some characters have given a confusing aspect because their appearance frequencies are fairly in the same range. We can see from table1 that the S and R percentages are fairly close.

Also, C, F, L, U, W characters have implicated some complexities because the percentages that match to theirs from the relatives frequency English language are fairly different. Therefore, some tricks have to be made to determinate the characters that could be used to substitute them in the plaintext.

On the other hand, there have been some satisfactory results with some frequencies because it was not too complex to map them to the right characters in reference to the relative frequency in the English language. For instance, we have the characters such as A, B, E, T, Z, N respectively. The percentages we got from the ciphertext gave sufficient information to determine them. Of course, those characters had given more understandable patterns in the first phase of the decrypted text.

The two letters frequencies have given a sharpen understanding of the pattern on the first phase of the decrypted text because the program has been run sequentially. For instance, the decryption was done first by replacing the characters in ciphertext with the supposedly right characters. Then, the program generates decrypted text file that

would be used as input file by the two letters decryption function to populate another decrypted text file. With that decrypted text file some logical thinking would be used to determine some common string patterns to give a complete decrypted text file that could be readable.

As shown in table2, the program has determined the top ten common patterns by sorting them accordingly. Therefore, some research have been made to try to find out what those common two characters string might be. Also, since we got the information about the one character appearances, some combination could also be used to determined those appearances. For instance, for the string that has the highest percentage IO we have assigned TH for the first iteration of decryption. Since we know that the likelihood P could be decrypted with the character E is considerable. I have made some adjustment. I have deduced that it could be appropriate to replace to P with E during the second iteration of the deciphering. Obviously, since we know that IO could probably be TH. We can see that OP is HE for the key.

For the other characters, I have also done some research to see what I could use for the first iteration. I have used the percentages I got to compare them to other resources that have taken the same approach to decrypt an encrypted text. For the first iteration, I have substituted the strings as the following:

OI -> TH, OP -> HE, KA -> IN, PE -> ER, PR -> ED, EP -> BE, LZ -> IT, WA -> BY , WI -> AN, IL -> OR.

After that, I have checked some patterns to decipher the text accordingly

Using words pattern to decipher

In this part, I have observed some recognizable patterns to decipher the text. For instance, I have checked in different lines for those patterns:

- In line 32, I observed ACTUADDA
    - In the ciphertext, we can see this string WYIZWSSN
    - By guessing we can see that could be matched ACTUALLY
        - The character S needed to be changed to L to decrypt the string
- In line 300 I found the string CAPTAYN
    - In the ciphertext we have the string YWGIWKA
    - Obviously, this for word could be CAPTAIN
    - Therefore, the character K was used to encrypt I
- In line 657 I got the pattern GUT
    - In the ciphertext we have TZI
    - The T was used to encrypt B
    - Therefore, we got the string BUT

With these adjustments some considerable patterns started forming in the decrypted text. For example, I started getting some recognizable phrase

- In line 54 I have observed this phrase: I CAN LET AHU PLAA
    - In the ciphertext we the following set of strings:
        - K YWA SPI NLZ GSWN
    - We could observe that the character Y was encrypted with N so the decrypted phase could be:
        - I CAN LET YOU PLAY

- In line 53 we have the phrase YOU XON'T VANT
    - In the ciphertext we have
        - NLZ RLA'I UWAI
    - Obviously, this could be
        - YOU DON'T WANT

As the deciphering was progressing more words are being decrypted. As a result, full sentences started to make sense.

- In line 1 we could observe this phrase: IT'D NOT EMACTLY WDAT I DAD IN AIND!
    - In the ciphertext we have the following strings:
        - KI'H ALI PMWYISN UOWI K OWR KA QKAR!
    - Using a little imagination we can see that S, X, H, and M were encrypted with H, M, O, Q respectively.
    - The plain is
        - IT'S NOT EXACTLY WHAT I HAD IN MIND!

The decryption was done on sequential manner. In other words, we went from some decrypted characters to a set of strings that have given sens to sentences.