CS641A (Modern Cryptology)

Chapter-2 Report

Instructor: Prof. Manindra Agrawal

Group Name: CMEN

Names :	Pratyush	170500
	Pravar Deep Singh	160503
	Tushar Garg	160749

Date of submission of report : 23/01/2020

In Chapter 2, as we enter we see the following message:

The chamber is completely dark. You can see nothing!
After grouping around for a while, you remember that
you have (fortunately) brought a matchbox with you.
You quickly lightup a matchstick ...

The light slowly fill up the chamber. It is not very
bright but enough to find your way. Looking around,
you immediately notice another exit and a glass panel
next to it. As you get used to the dim light, you
notice that the chamber floor is very uneven. Several
boulders are lying on the floor and you were lucky
that you did not stumble on these when it was dark.
You look around these boulders searching for something
interesting and notice some funny patterns on one of
the distant boulder ...

Here we found out that we have two options, either to enter the command "read" or "go" which takes us to different sections. If we enter the command "read" then we obtained the following message:

Lg ccud qh urg tgay ejbwdkt, wmgtf su bgud nkudnk lrd
vjfbg. Yrhfm qvd vng sfuuxytj
"vkj_ecwo_ogp_ej_rnfkukf" wt iq urtuwjm. Ocz iqa jdag
vio uzthsivi pqx vkj pgyd encpggt. Uy hopg yjg fhkz
arz hkscv ckoq pgfn vu wwygt nkioe zttft djkth.

We realized that the message is encrypted and we need to decrypt it.

Now in the first screen of Chapter 2, if we enter “go” then we get the following message:

As you move closer to the boulder, you realize that it has something written on it! Wiping the dust from the boulder with your hand - and getting your hand very dirty in the process - you see a strange shape on the boulder -- it appears like a human face. Below the face, there is a message written:

The spirit of Cave Man is the keeper of the chamber. To navigate through the chamber, you must pay respect to him first. Bow, and then slowly look up. Count the number of lines in horizontal dimension -- they will stand in good stead.

After reading the above text we counted the number of horizontal lines in the image, as instructed in the text itself and got the number 9. At this point in time, we did not know what to do with it so we started to analyze the encrypted text.

Getting To The Encryption Method:-

We first did frequency analysis to find the frequency of various letters to get the histogram as shown in Fig.1. The graph didn't seem to be close to the English text so we ruled out the possibility of monoalphabetic substitution. We then tried to figure out if encryption might be a permutation or a combination of permutation and substitution. To check if that might be the case we calculated the "**Index of Coincidence**" of the whole ciphertext and found its value to be **0.04236 ~ 0.042**. Now the index of coincidence of Standard English text is close to **0.07** and applying substitution or permutation or combination of both does not change the Index of Coincidence of any text. Hence, we ruled out the above possibility too.

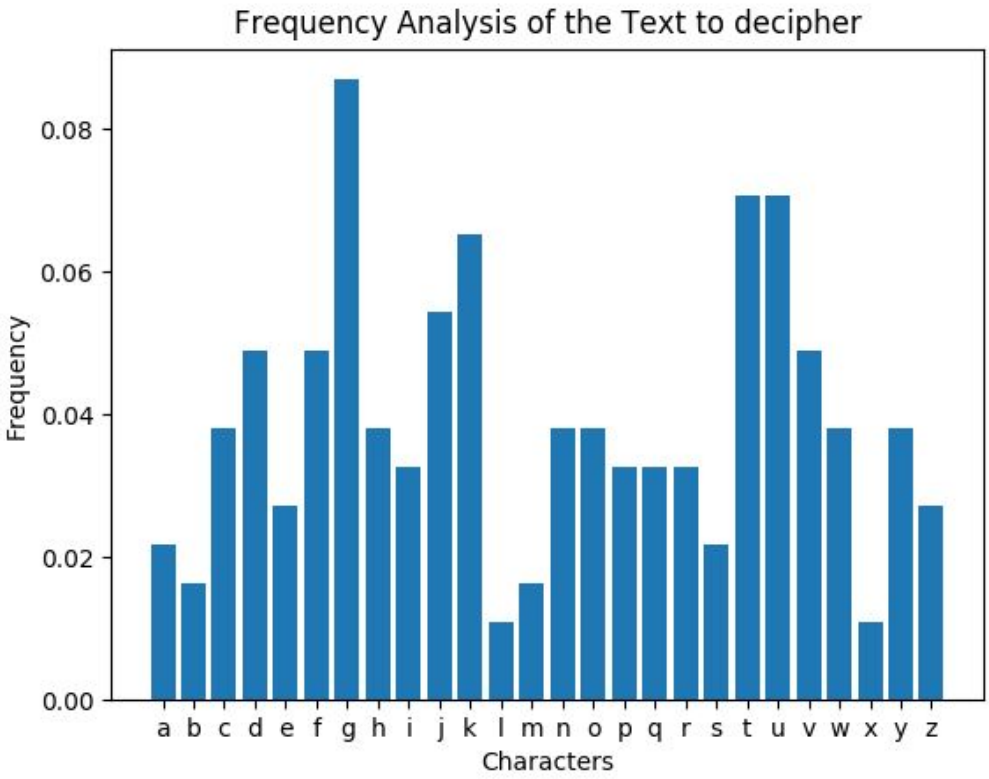

Fig. 1: Frequency Analysis of text to decrypt

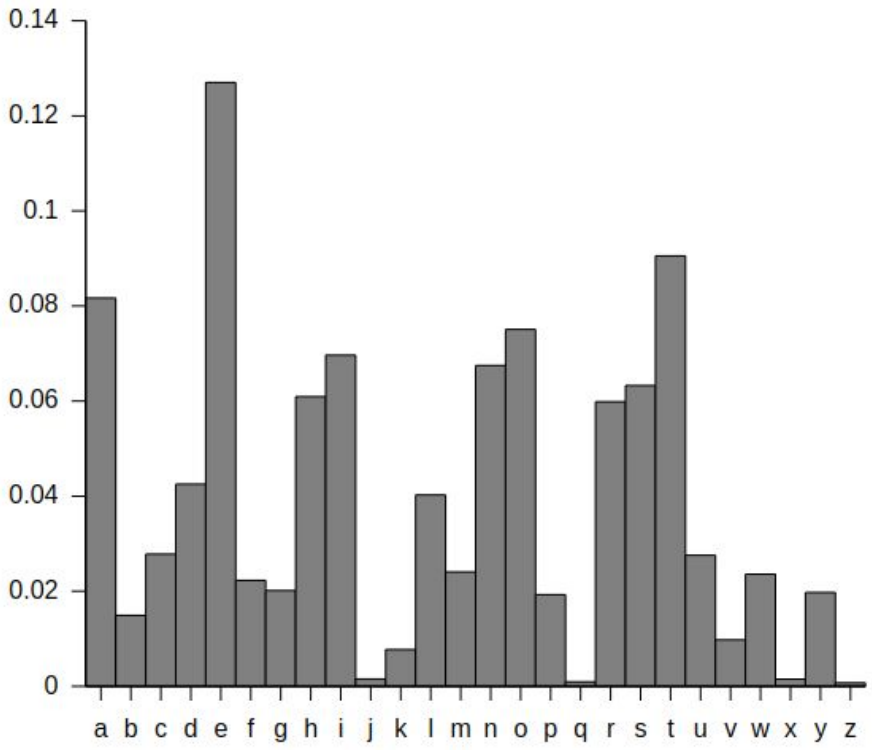

Fig. 2: Frequency Analysis of the Standard English Text

Next, we tried the possibility of "**polyalphabetic substitution**". We started with the simplest case in which k different monoalphabetic substitutions might be used on blocks of length k i.e. character number i, $i+k$, $i+2k, \ldots$ for $i = 1, 2, \ldots, k$ is part of the same substitution rule.

Decrypting the text:-

To find the value of k, let us create k strings where the i^{th} string contains all the character from the text with the indices $i,\ i+k,\ i+2k,\ ...,\ i+n_ik$ where n_i is such that $i+n_ik \leq length\ of\ the\ text < i+(n_i+1)k$.

S.No.	Value of *k*	Index of Coincidence
1	1	0.04236
2	2	0.04356
3	3	0.04580
4	4	0.04266
5	5	0.03874
6	6	0.04835
7	7	0.04739
8	8	0.04788
9	9	0.05564
10	10	0.03340
11	11	0.03917
12	12	0.04881
13	13	0.04903
14	14	0.04919
15	15	0.03807
16	16	0.04299
17	17	0.04433
18	18	0.06173
19	19	0.03509
20	20	0.03139

Table 1: Average Index of Coincidence for different block lengths

Now all the characters with the indices: $i,\ i+k,\ i+2k,\ ...,\ i+n_ik$, where $1 \leq i \leq k$, are part of the same monoalphabetic substitution rule, so their Index of

Coincidence (IOC) should be close to that of English text (if our value of k is correct). So we calculated IOC for each of the k strings and then took their average. We tried all the values of k starting from 1 till 20 (after which length of the k strings will become too small to solve). The values thus found are shown in table 1. Now we see that we have maxima at $k = 9$ and $k = 18$. We also spotted the hint from caveman's face that horizontal lines were 9, therefore we concluded that $k = 9$.

Now that we figured out that there are 9 different monoalphabetic substitutions used we still needed to figure out what they are. Firstly we need to figure out what kind of monoalphabetic substitution they are. If a Permutation Cipher is used then we have too few letters to figure out in legible time. So we started with the case that simple Caesar Cipher might be used in each of the 9 monoalphabetic substitutions.

Now to figure out the shift values, we first divided our ciphertext into 9 strings as defined above. Now for each of 9 strings, we tried all possible 26 shifts and for each shift, we performed χ^2 test to find how close our string becomes with respect to the English text. We have tabulated the top three shifts for each of the 9 strings in table 2.

String Number	First Possibility		Second Possibility		Third Possibility	
	Shift Amount	χ^2 Value	Shift Amount	χ^2 Value	Shift Amount	χ^2 Value
1	4	31.030	0	32.187	16	36.314
2	24	7.484	25	62.381	20	64.101
3	20	9.309	4	53.358	8	74.024
4	24	37.912	22	42.651	11	68.919
5	4	32.517	23	35.709	12	40.724
6	21	23.939	15	40.315	8	92.799
7	24	18.107	25	84.718	13	51.727
8	24	4.637	1	37.870	8	51.727
9	9	27.964	25	42.604	14	51.521

Table 2: Average Index of Coincidence for different key lengths

We noticed that the top 3 χ^2 values in the case of strings 1 are very close. So, we have to try all three possibilities. Top 2 χ^2 values in case of string numbers 4,5,6,9 are pretty close so we need to try the top 2 possibilities. But in the case of string numbers 2, 3, 7, 8 first possibility is quite dominant so we need not try second and third possibility.

So we tried all 48 combinations and found out one of them was meaningful which is shown below:

> *Be wary of the next chamber, there is very little joy there. Speak out the password "the_cave_man_be_pleased" to go through. May you have the strength for the next chamber. To find the exit you first will need to utter magic words there.*

This is achieved by following shift amounts in order 16,24,20,24,23,21,24,24,25
Then we simply typed the password “the_cave_man_be_pleased”, as instructed to go to chapter 3.

Explanation of codes:-

1. The file input.txt contains the message to be deciphered.
2. The file preprocess.py removes all the characters except alphabets from the input.txt and it also changes the upper case letters into lower case letters and saves the final text into the file level_2.py.
3. The file Frequency_Analysis.py creates the frequency histogram for the text to be deciphered.
4. The file Index_of_Coincidence.py calculates the index of coincidence of the text for different key lengths through which we confirmed the value of key size.
5. The file Shift.py take the 9 string and calculate the χ^2 value for a different amount of shifts and print the possible deciphered text to output.txt file.
6. The file Decrypted.py changes the text from the file output.txt to the way it was before applying the preprocess.py file.

References:-

1. *https://en.wikipedia.org/wiki/Frequency_analysis*
2. *https://en.wikipedia.org/wiki/Index_of_coincidence*
3. *http://practicalcryptography.com/cryptanalysis/text-characterisation/chi-squared-statistic/*