CS 641 : ASSIGNMENT : PART 2

A PREPRINT

Srinjay Kumar
170722

Abhyuday Pandey
170039

Kumar Shivam
170354

January 23, 2020

1 Initial Sequence

The keyword used to get to the cipher text are :
read

2 Problem

The second problem of our assignment is to decipher the ciphertext *"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."*.

3 Analytical Decryption

3.1 Why not Substitution Cipher ?

- In cipher text generated by substitution, if we have two same letters then they will correspond to the same letter from the English alphabet.
- In the first line of the ciphertext, we have a word "ccud" which means that if the ciphertext is generated by substitution, then it is a 4 letter english word with first letters being same. Such words are quite uncommon and so it cannot be a substitution cipher.We looked for such words in an online dictionary and found no such words.
- A general English text has repetition of words such as "is", "it", "the", "of" and "in". So, there should be repetition of at least few two letter words.
- We have done the frequency analysis of the cipher text in Table 1. In the analysis, we have all the 26 characters in the table. This is a very distant possibility because the characters such as q, x and z are quite rare in English alphabet. Also many characters seem to have not so distant frequencies unlike the normal behaviour of frequencies.
- Therefore, we can safely say that there exists no one to one mapping between letters of ciphertext and alphabets of English alphabet.
- So we can conclude that the ciphertext is not formed using substitution.

3.2 Why not Permutation Substitution Cipher ?

- A substitution followed by a permutation won't essentially change the mapping structure i.e. there will still be a one-to-one map between letters of ciphertext and plaintext.
- Owing to the frequency analysis from previous section, the possibility of a permutation (followed by substitution) can also be discarded.

Table 1: Frequency Analysis

Letter	Frequency	Percentage
G	16	8.65%
T	13	7.03%
U	13	7.03%
K	12	6.49%
J	10	5.41%
F	9	4.86%
V	9	4.86%
D	9	4.86%
H	7	3.78%
N	7	3.78%
Y	7	3.78%
O	7	3.78%
W	7	3.78%
C	7	3.78%
R	6	3.24%
Q	6	3.24%
I	6	3.24%
P	6	3.24%
Z	5	2.71%
E	5	2.71%
S	4	2.16%
A	4	2.16%
B	3	1.62%
M	3	1.62%
X	2	1.08%
L	2	1.08%

3.3 Type of Cipher

- In the previous ciphertext, we were given the password of the chamber. We were confident about a few things, first, "password" must be present in the ciphertext as it was there in the previous message, second, the string engulfed in quotes may be the password, third, "chamber" may be present in the message, and lastly, "the" may be present as a three letter word and preferably in front of a noun.
- **password:** "sfuuxytj" ⟶ "password" or "uzthsivi" ⟶ "password".
- **chamber:** "ejbwdkt" ⟶ "chamber" or "cncpggt" ⟶ "chamber" or "urtuwjm" ⟶ "chamber".
- **the:** There are many three letter words, near the above transformations only "vio" and "vng" are the words that may correspond to "the".
- We attempted to find hamming distance between all aabove words in attempt to realize if there was a cyclic shift encoded.
- "sfuuxytj" ⟶ "password" = "3 5 2 2 1 10 2 6"
- "uzthsivi" ⟶ "password" = "2 25 1 15 22 20 4 5"
- "ejbwdkt" ⟶ "chamber" ="2 2 1 10 2 6 2"
- "cncpggt" ⟶ "chamber" ="2 6 2 3 5 2 2"
- "urtuwjm" ⟶ "chamber" = "18 10 19 21 5 21"
- "vio" ⟶ "the" = "2 1 10"
- "vng" ⟶ "the" = "2 6 2"
- The recurrence of the pattern "3 5 2 2 1 10 2 6 2 3 5 ..." suggests that there is a recurrence in the hamming distance.
- On a little investigation, we found that such a cipher does exists and is named as **Vigenere Cipher**.

3.4 Vigenere Cipher

These ciphers do not have one to one mapping between the English alphabet and the alphabets of the ciphertext. These ciphers have a sequence of numbers as their key. For encypting a cipher in Vigenere cipher we move the alphabets of the plaintext forward in accordance with the corresponding number from the key.
For example if the plaintext is "cryptography" and the key is "3 1 2". Then the plaintext will be encoded as "fsasuqjscsia". We can do the decryption by just doing the inverse of the method for encryption.

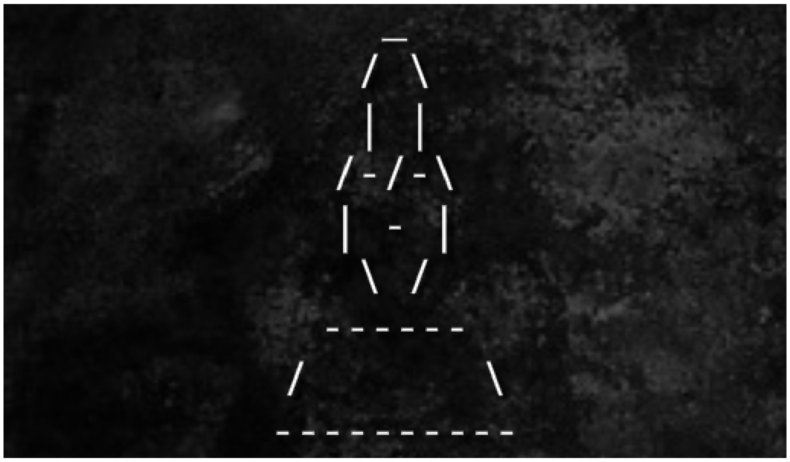
3.5 Decryption

- From our previous analysis, the following mappings were likely to be a match for the key.
- "sfuuxytj" ⟶ "password" = "3 5 2 2 1 10 2 6"
- "ejbwdkt" ⟶ "chamber" ="2 2 1 10 2 6 2"
- "cncpggt" ⟶ "chamber" ="2 6 2 3 5 2 2"
- "vio" ⟶ "the" = "2 1 10"
- "vng" ⟶ "the" = "2 6 2"
- The above words were shortlisted for the matching.
- We proceeded with "10 2 6 2 3 5 2 2 1" as the key for decryption keeping in mind the offset of above words from starting point.
- We have written a **C++** program which applies a key with all offsets to the ciphertext and outputs the plaintext. Currently, ciphertext and key is hardcoded in the code.

Since we have the key, we have the actual text, "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."

According to the text, the password for the next level is "the_cave_man_be_pleased".

4 Decryption by Image



- In the corresponding slide which contains this image, we had the text "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.".
- So after decryption, we verified our key with the image of the cave man.
- According to the text, we had to count the lines in the horizontal direction.
- Counting of the lines from downward to upward direction gave us the sequence of numbers "10 2 6 2 3 5 2 2 1" which is the actual key that we had got by analytical decryption.

- Since the key is "10 2 6 2 3 5 2 2 1", we get the plaintext "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.".
- According to the plaintext, the password for the next level is "the_cave_man_be_pleased".
- We realized that the image was meant to serve as a hint which we could only use as a cross-verifier.

5 Decryption by Code

Knowing very well that in general we may not have hints or be "lucky" enough to get the key by hit and trial. So, we coded a computer program which can decrypt any vignere cipher in general with key length < 16 approx on our PCs.

5.1 Chi-squared Analysis

- Chi-squared test calculates the distance of the ciphertext from an English Plaintext based on the frequency of English Alphabets. The frequencies of alphabets in any English plaintext are mentioned in Table 2.
- We calculate the percentage of the alphabets in the ciphertext and then calculate the distance between the actual percentage and the percentage in the ciphertext by using the following formula

  $$\chi^2 = \sum_{i=1}^{26} \frac{(y_i - x_i)^2}{y_i}$$

  where,
  x_i := the percentage of alphabets in the ciphertext
  y_i := the expected percentage of alphabets in any general Englishtext.
- For any instance of key length of key size k, we first make k strings by taking the i^{th} letter from each group of k letters where i lies between 1 and k. Then the string we get is simply a **Caeser** cipher.
- Now, we have 26 possibilities for Caeser cipher. We iterate through all the 26 possibilities and then calculate the χ^2 of the decrypted text. We then take the top 3 results and store it.
- After get a group of possible permutations, we examine the likelihood of each permutation by matching the words formed with a dictionary and then maximizing the number of words matched with the dictionary.
- The program takes key length as a command line input and we tried all possible key lengths from 1..9 and 9 was the correct key length.
- In this way, we get our answer for key length 9 and the key is "10 2 6 2 3 5 2 2 1".
- Since the key is "10 2 6 2 3 5 2 2 1", we get the plaintext "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.".
- According to the plaintext, the password for the next level is "the_cave_man_be_pleased".

6 References

- Free Dictionary
- Applied Cryptography 2ed
- Letter Frequency
- Chi-Squared Test

Table 2: Expected Frequency Analysis

Letter	Percentage
A	8.17%
B	1.49%
C	2.78%
D	4.25%
E	12.70%
F	2.29%
G	2.01%
H	6.09%
I	6.97%
J	0.15%
K	0.77%
L	4.02%
M	2.40%
N	6.74%
O	7.50%
P	1.92%
Q	0.09%
R	5.987%
S	6.37%
T	9.05%
U	2.75%
V	0.98%
W	2.36%
X	0.15%
Y	1.97%
Z	0.07%