CS553 Cryptography

BitBees

Question 2

Approach 1

We take the ciphertext and try all the possible shifts from 0 to 25. Now we calculate the Chi-Square Statistic of the entire plaintext we just obtained from shifting. We have the probabilities of frequency all the letters occurring in the English language.

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

where,
O_i is the occurance of letter i in alphabet,
E_i is the expected occurance of letter i in alphabet,

Using this formula we get the statistic for all the elements and one with the lowest Chi-Square statistic is most likely the shift which is correct.

Please refer to `q2.py` for the implementation of this approach.

a) Decrypted message 1: `ITHINKTHATISHALLNEVERSEEABILLBOARDLOVELYASATREE`,
Shift 3
b) Decrypted message 2: `LOVEISNOTLOVEWHICHALTERSWHENITALTERATIONFINDS`,
Shift 9
c) Decrypted message 3: `INBAITINGAMOUSETRAPWITHCHEESEALWAYSLEAVEROOMFORTHEMOUSE`,
Shift 19

Approach 2

We have to find the readable text from the decrypted messages coming from different shifts from 0 to 25. After applying all the shifts, we got the readable text as follows:

a) Decrypted message 1: ITHINKTHATISHALLNEVERSEEABILLBOARDLOVELYASATREE,
Shift 3
b) Decrypted message 2: LOVEISNOTLOVEWHICHALTERSWHENITALTERATIONFINDS,
Shift 9
c) Decrypted message 3: INBAITINGAMOUSETRAPWITHCHEESEALWAYSLEAVEROOMFORTHEMOUSE,
Shift 19

The above ones were extracted by manual reading. We can have multiple methods to automate the process. Few to name are: counting the number of E's, NLP, etc. NLP gives us almost 100 percent accurate results. With the help of libraries, each string is tokenized into the words belonging to the English dictionary and in case the tokenized word doesn't belong to the English dictionary then we jump to the next shift. We check for all the keys if all the tokenized words in a string belong to the English dictionary and then only consider a string as readable text. We get the following results from NLP:

a) Readable message 1: I THINK THAT IS HALL NEVER SEE A BILLBOARD LOVELY AS A TREE
b) readable message 2: LOVE IS NOT LOVE WHICH ALTERS WHEN IT ALTERATION FINDS
c) readable message 3: IN BAITING A MOUSETRAP WITH CHEESE ALWAYS LEAVE ROOM FOR THE MOUSE

To see the output, here is the link to the Google colab notebook:
https://colab.research.google.com/drive/1heI1oAU0BgCcCUao7qs0kU3cbpzxFqpr?usp=sharing