CSC738 – Cyber Security Assignment 1

Cryptanalysis

Done by Matthew Robin Arendse 3956525

In this assignment we were tasked to decrypt text from an image. The main task is to use cryptanalysis to find the decryption key using different cryptanalysis methods like frequency analysis, Index of Coincidence and making use of basic knowledge of the English languages. In this document I will discuss my thought process and the step-by-step process of what I done to fine the cypher key.

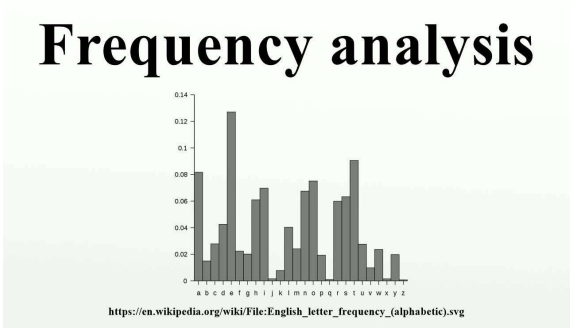
Step 1: Watch lecture videos and Learn methods to do cryptanalysis.

I started by watching lecturer videos on different ciphers provide. In the video I learnt how the different ciphers are implemented which assisted me in getting a starting point. One method that I had implemented on the side was the Ceasar/shift cipher as it was easy to implement. I did not uncover much useful information but found the only single letter in the cipher. While watching the presentations two methods stood out that were useful for finding the starting point was the single letter substitution as well as frequency analysis. Researching how to do cryptanalysis I discovered two methods that would assist in narrowing down what type of cipher was used to encrypt the text. IOC is used to determine how uniform the frequency of letters used in cipher. IOC stands for Index of Coincidence; the IOC for the English language is 0.07 if the IOC for a cipher is close of similar to 0.07 then the cipher used some sort of substitution cipher and not more complicated ciphers. Every langue has a different IOC number. Another method discovered just having an understanding of the English language as in the different probabilities that certain letters will appear next to one another in context of a sentence.

Step 2: Applying methods to our problem.

Understanding that in the English language the IOC is 0.07 and if a cipher text with an IOC being in a certain range it has to be a substitution cipher. After some light implementation of IOC algorithm on the cipher text it was found that that it was closer to 0.06 which was close enough to confirm it was a substitution cipher being used. After getting frequency of each letter used in the cipher text there was great outlier with having the use frequency of 88 times.

While looking at graphs of the frequency analysis, came across a graph of the English language, E is the most used letter in the English language. At this point I had already started getting a sense of possible substitutions to make, but nothing was concrete and confirmed at this point. I had implemented quick methods to confirm and reinforce what I have learnt about cryptanalysis.



Using the idea that A and I are the only characters that are single letter words in the English language it was used as a starting point to start figuring out the cipher key. I had some of the tools to start cracking the code.

Step 3: Implementing – Single Letter Substitution and Single letter analysis

To get the key I implemented simple python code to substitute the characters in the text. This was done to see the progress of the cipher and what letters still needs to be found.

Using the knowledge that A and I is that only characters that are single character in the English language. I found that in the cipher "M" was the only single character in the cipher text. It was most likely the letter "A" or "I", but I chose the letter "A" as it was more likely that the text was not written in the first person. The first substitution made was "M"- "A".

Using frequency analysis "S" had the highest frequency use in the cipher text, but knowing in the English language the highest frequency used letter is "E". With this in mind I made the substitution from "S" – "E". In the back of my mind was the 2nd highest used letter which was "O" with it being used 85 times in the cipher text, with frequency analysis of the English language the letter "T" is the second most used letter in the English language. Therefore made the substitution from "O"-"T".

After every substitution I look through the whole text to look for more clues for other letters.

One word stood out which was "ADD", not many words as a double letter after "A" which led me to choose the most logical swap from "D" – "L". Many words had L, A, E and T in them.

Letters found in the cipher key:

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

Step 4: Implementation – Double letter words analysis and Di-gram analysis.

Tackling the two letter words, I started by looking at the words after the last substitution. The letter "T" had a big presence in two letter words appearing at the end and beginning of two letter words. Looking at words that started with the letter "T" the next common letter in the cipher was "G". knowing that most double letter words in alphabet has a vowel in it. The most logical substation to make "G" - "O", as "TI", "TA", "TE", "TY" aren't words in the English language. Looking at the words that ended in "T", since "A" was already found there was only one other word that had "T" on the

end which was “PT”, I assumed that the other letter was most likely the letter “I”. This was not a confident decision but I make the substitution non the less. The substitution being “P” – “I”

After making the substitution more two letter words, stood out that ended and started with “O”.

Starting with the words that ended in “O”, after being “T”, “S” is the next best fit to be used next to “O”. Therefore, I make the substitution from “L” – “S”.

Letters found in the cipher key:

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

Before looking for the next step, bigger word appeared, like “ITS” which gave me confirmation that substitution “P” – “I” was a good substitution.

Step 5: Implementation –Triple digit word analysis and Trigram analysis.

With “T” and “E” being discovered, a three-letter word that appeared frequently throughout the cipher text was “OIS” which mapped to “TIE”, knowing that “I” is found the most common three letter word in the English language that starts with “T” and ends with “E” is most definitely “THE”, therefore the mostly logical next substitution would be “I” – “H”.

Having the letters “O” and “E” the first word was “OFE”, The letter “F” appeared in double letter words next to letters like “O” and “I” and other instances. This confirmed that “F” could be substituted to “N”.

Before moving on I looked at the text again and found that the letters “V”, “W”, “Y”, “X”, were not substituted in this cipher text. As full words like made sense in context like “SOLVE”, “WAY”, “WE”, “WANT”, “NEXT” etc. The first few words were found being “ONE WAY TO SOLVE AN EN…”, finding this many new words started appearing and not a lot of analyse had to be done. This was because the humans can pick up on patterns and put stuff in to context very easily.

Letters found in the cipher key:

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

Step 6: Manually filling the left-over words.

Looking at the text in context, the text appeared to be speaking about encryption. Using this picking making words was simple and could be read even though one or two letters were not in correct. Using this idea, the next two words in the sentence is “ENCRYPTED” and “MESSAGE”. This gave us the letters “C”, “P”, “G”, “R”, “D” and “M” … so doing the substitution, “N”-“C” , “H”-“P”, “R”-“G”, “K”-“R”, “U”-“D” and “E”- “M” .

Letters found in the cipher key:

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

Step 7: getting the rest of the letters the text.

Looking at the cipher text again the sentence read as “ONE WAY TO SOLVE AN ENCRYPTED MESSSAGE, IC WE BNOW ITS LANGQAGE, IS TO CIND A DICCERENT PLAINTEXT …” as one can see that the letter “C”, “Q”, “B” can be mapped to “F”,”U” and “K” make the words “IF” , “KNOW” , “FIND”, “LANGUAGE”, “DIFFERENT”

Letters found in the cipher key:

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

Step 8: Conclusion

In this step the “B” and “Q” was found in the words “SYMAOLS” and “FREJUENTLY”, Therefore mapping “A”-“B” and “J”-“Q”. The last two letters that weren’t used in the cipher was T and Z, and the only two letters left is J and Z, one has no other indication as to was the letters maps to. I had to guess and just mapped “T” – “J” and “Z”

Letters found in the cipher key:

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

That is how I found the key to solve the cipher text. The process was a lot more manual than thought.

Link to GitHub repository: https://github.com/SlipperySatyr6/CSC738-Assignment1_ImagetoText_decyphering_app.git

Google Drive for the apk file and video:
https://drive.google.com/drive/folders/1NgbHdTxEPbkPGeSnotfnAv1JhXxJsFzo?usp=sharing