TTM4135 - Practical

Sander Lindberg - 493389 - Folder 107

Spring 2020, NTNU

1 Questions

1.1 Which file corresponds to which encryption algorithm?

The file 0.txt corresponds to the Hill cipher, 1.txt to the random substitution cipher, 2.txt to the transposition cipher and the last, 3.txt, to the Vingènere cipher.

1.2 Present the evidence for each decision, showing how the character frequency and/or autocorrelation is consistent with your decision

Figure 1 shows the character frequency for the random substitution cipher. Since the random substitution cipher maps one character to another, the character frequency for the ciphertext will look like the frequency for the English language, but with different characters. In this histogram, I can see we have two characters that are more frequent than the other - ”o” and ”q”. This is the same as the histogram for the English language, where ”e” and ”t” are the two most frequent characters. To further confirm my suspicion that this text was indeed encrypted using the random substitution, I took a look at the other histograms.

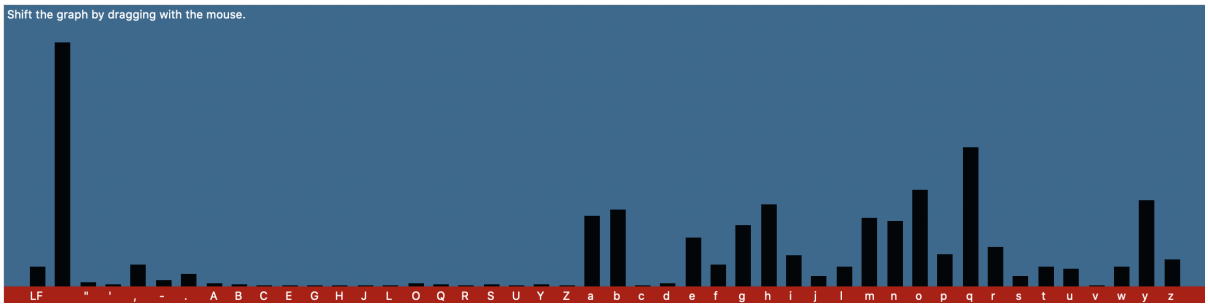

Figure 1: Character frequency for the random substitution cipher

Figure 2 shows the histogram for the character frequency for the transposition cipher. This also looks like the English frequency (high frequency of ”e” and ”t”), which could lead one to believe that the substitution cipher was used to encrypt the text. This is, however, very unlikely due do the randomness of the changes of the characters. Because the substitution cipher changes the characters in the plaintext, the highest frequency should be observed for other characters than ”e” and ”t”. Therefore, I deduced that ”2.txt” could not be the random substitution cipher.

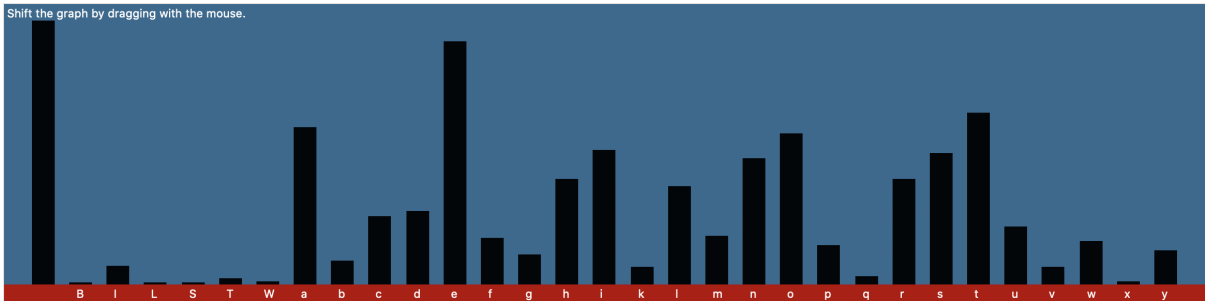

Figure 2: Character frequency for the transposition cipher

Figure 3 and figure 4 does not look like the English frequencies at all, therefore I could rule out ”0.txt” and ”3.txt” and know that ”1.txt” was indeed the substitution cipher. Since these two (figure 3 and figure 4) did not resemble the English language, and the transposition cipher does, I found that ”2.txt” was encrypted using the transposition cipher. I have also shown this above, substitution would not have high frequencies of ”e” and ”t”, but the transposition would. This is because the text is divided into blocks of columns, and then rearranged. Therefore, the characters stays the same in the ciphertext as in the plaintext.

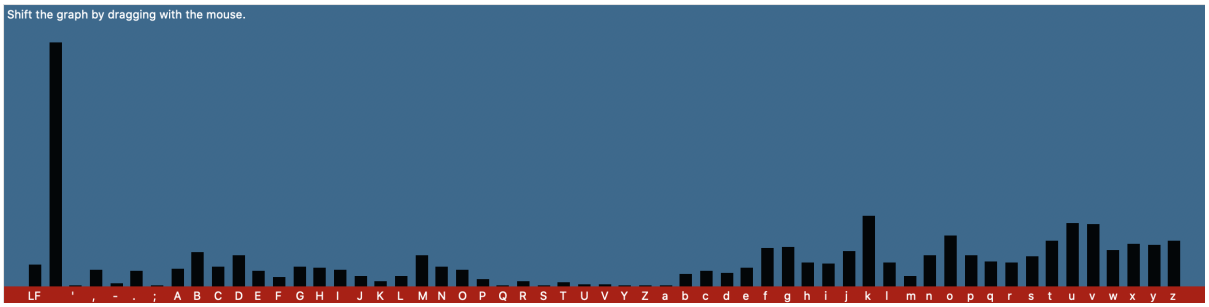

Figure 3: Character frequency for the Vigenère cipher

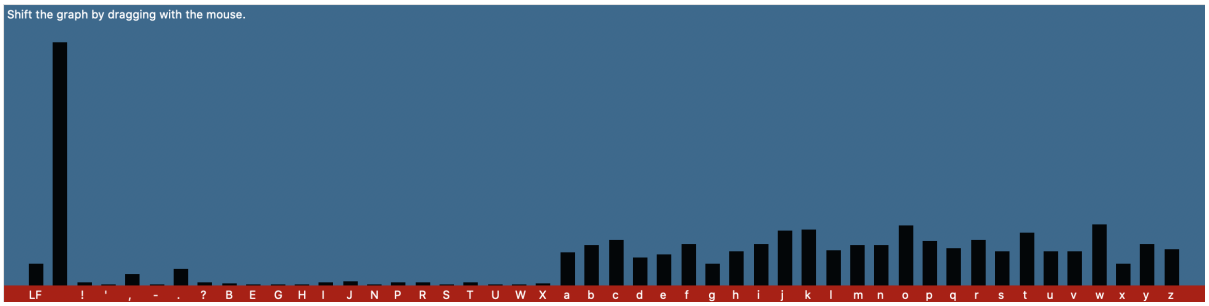

Figure 4: Character frequency for the Hill cipher

This left me with ”0.txt” and ”3.txt”, and the Hill and Vigenère cihpers. I took a look at the autocorrelation of the two texts, to find periodicity (see figure 5 and figure 6).

I quickly deduced that ”3.txt” was encrypted with the Vigenère cipher, based on the autocorrelation. As we can see in figure 6, a higher periodicity is observed than in figure 5, as the Vigenère would have, because of it’s repeating key. I also distinguished between the Hill and Vigenère by looking at the histograms for the character frequencies. If the Hill cipher had been encrypted with an alphabet consisting of both uppercase and lowercase letters, the ciphertext would have a mix of uppercase and lowercase. However, this was not the case in the text. Figure 4 shows the histogram for the Hill ciphertext, with very few uppercase letters, and figure 4 shows a little more uppercase letters. In addition, as the histogram could not be from the substitution or transposition ciphers and given the fact that the autocorrelation for the Hill cipher showed little periodicity, I concluded that ”3.txt” was encrypted using the Vigenère and ”0.txt” was encrypted using the Hill cipher.

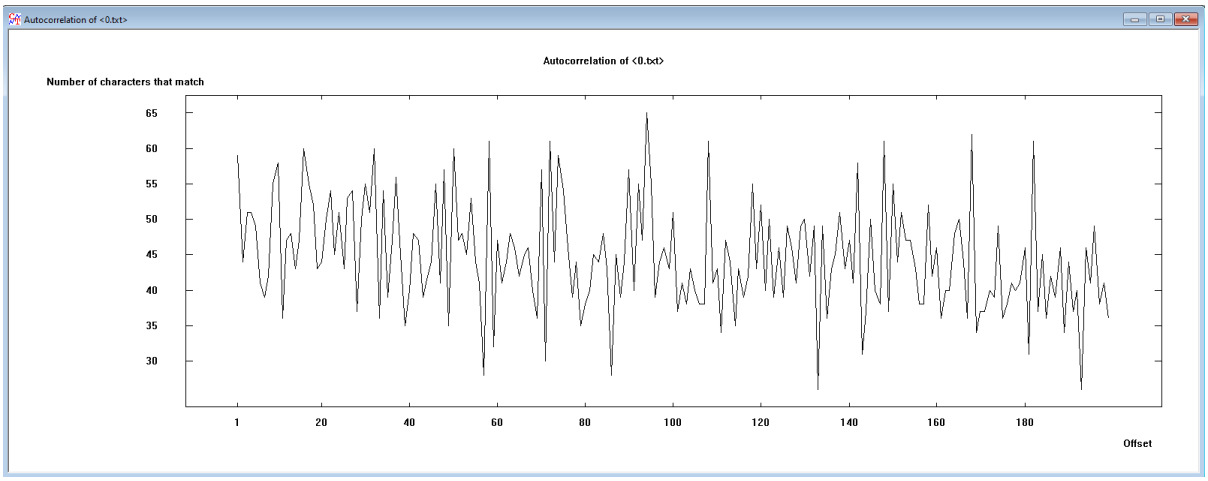

Figure 5: Autocorrelation for the Hill cipher

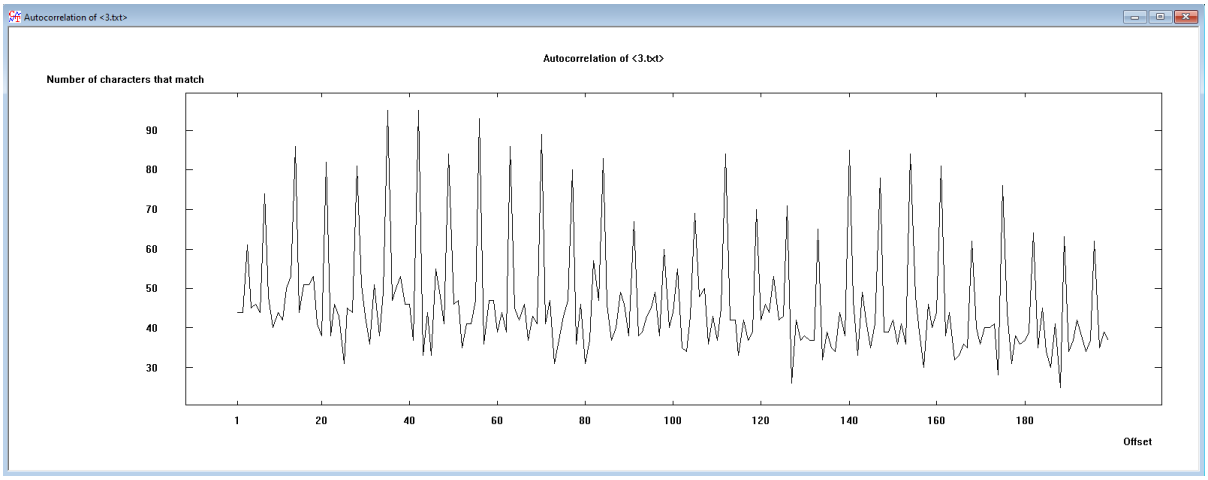

Figure 6: Autocorrelation for the Vigenère cipher

2 Breaking the ciphers

Now that I knew which file corresponded to which algorithm, it was time to decrypt them.

2.1 Vigenère

I started with the Vigenère cipher. As instructed, I used the automatic tool in CrypTool. This analysis gave me the key length 7 and the key ”VKTCGBD”. Using this key to decrypt, I got the plaintext (100-200 charaters):

> *he could see the strange, bottle-shaped kilns with their orange, fanlike tongues of fire. A dog barked as they went by, and far away in the darkness some wandering sea-gull screamed.*

(A quick Google search told me this was from the book ”The Picture of Dorian Gray and Three Stories” by Oscar Wilde)

2.2 Random substitution

Next, I tried to decrypt the random substitution cipher. The automatic analyisis gave me the key ”YFZRQILABKDETGHSCMNOPUJVWX”, which seemed *almost* correct. I had words such as ”peggar” and ”charitaple”, that told me the key was only partly correct. I tweaked it a little bit and got the key ”YSZRQILABCDEFGHJKMNOPTUVWX”[1]. Now ”peggar” became ”beggar” and ”charitaple” became ”charitable”. This told me I had gotten the right key! The plaintext I got are as follows (100-200 chars):

> *one's self. Of course, they are charitable. They feed the hungry and clothe the beggar. But their own souls starve, and are naked. Courage has gone out of our race.*

(A quick Google search told me this was also from the book ”The Picture of Dorian Gray and Three Stories” by Oscar Wilde)

2.3 Transposition

It was given in the problem description that the transposition cipher had block length 6, 7 or 8. Therefore, I just fired up JCrypTool's interactive tool to swap columns (as recommended). After playing with it for a while, I got block size 7 and the key 2137645[2]. The plaintext I got was as follows (100-200 chars):

> *kicked out by a schoolmaster was a mystification without end Let me add that in their company now and I was careful almost never to be out of it*

[1] This is the key I got using CrypTool 1. With JCrypTool, I had to add lowercase also, so I got the key ”YSZRQILABCDEFGHJKMNOPTUVWXyszrqilabcdefghjkmnoptuvwx”. I suspect that is is because CrypTool 1 does not differentiate between lovercase and uppercase, but JCrypTool does.

[2] or 2|1|3|7|6|4|5

(A quick Google search told me this was from the book ”The Turn of The Screw and Other Short Novels” by Henry James)

2.4 Hill

For the Hill cipher, I did as instructed in the problem's description. I split the text into digrams, and defined the alphabet as [A-Za-z]. I then used the n-gram tool in CrypTool and found that the most common digrams was ”ib” and ”rw”. I thought these could map to ”th” and ”he”, respectively, and knowing that $K = C \cdot P^{-1}$, I did the following calculations:

$$ibrw = \begin{bmatrix} 8 & 17 \\ 1 & 22 \end{bmatrix} \pmod{26} = C$$

$$thhe = \begin{bmatrix} 19 & 7 \\ 7 & 4 \end{bmatrix} \pmod{26} = P$$

$$\begin{aligned} K &= \begin{bmatrix} 8 & 17 \\ 1 & 22 \end{bmatrix} \cdot \begin{bmatrix} 19 & 7 \\ 7 & 4 \end{bmatrix}^{-1} \pmod{26} \\ &= \begin{bmatrix} 8 & 17 \\ 1 & 22 \end{bmatrix} \cdot \frac{1}{27} \cdot \begin{bmatrix} 4 & -7 \\ -7 & 19 \end{bmatrix} \pmod{26} \\ &= \begin{bmatrix} 8 & 17 \\ 1 & 22 \end{bmatrix} \cdot 27^{-1} \cdot \begin{bmatrix} 4 & -7 \\ -7 & 19 \end{bmatrix} \pmod{26} \\ &= \begin{bmatrix} 8 & 17 \\ 1 & 22 \end{bmatrix} \cdot 1 \cdot \begin{bmatrix} 4 & -7 \\ -7 & 19 \end{bmatrix} \pmod{26} \\ &= \begin{bmatrix} -87 & 267 \\ -150 & 411 \end{bmatrix} \pmod{26} = \begin{bmatrix} 17 & 7 \\ 6 & 21 \end{bmatrix} \pmod{26} \end{aligned}$$

It turned out that $\begin{bmatrix} 17 & 7 \\ 6 & 21 \end{bmatrix}$[3] actually was the key used for encryption. Putting these values into CrypTool's decryption option, I got the plaintext (100 - 200 chars):

> *dim light the hideous face on the canvas grinning at him. There was something in its expression that filled him with disgust and loathing.*

[3] or ”rghv”

(A quick Google search told me this was also from the book ”The Picture of Dorian Gray and Three Stories” by Oscar Wilde)