



GAZI UNIVERSITY

ENGINEERING FACULTY

COMPUTER ENGINEERING

CENG482 INTRODUCTION TO COMPUTER SECURITY

Gamze AKSU – 171180005

Assignment-5

APRIL 2022

USING LETTER FREQUENCY FOR CRYPTOANALYSIS

In cryptanalysis, letter frequency is based on counting letters or groups of letters in a text. Each language has its own letter frequency. Some letters or letter groups are used more often than others in a language. For example, in English, the letters e, t, a and o are used more often than other letters. The letters z, q, and x are the least used letters in English. Considered as a combination of two letters, it is called a bigram or digraph. Combinations of TH, ER, ON, and AN have the highest frequency. Figure 1 shows the frequency of use for each letter for the English language. Similarly, while the letters a, e and i are used more frequently in Turkish, the letters p, f and j have the least usage frequency. Figure 2 shows the letter frequency graph of Turkish. However, this only applies if the text is long enough. Letter frequency analysis cannot be performed on a very short text. [1] [2]

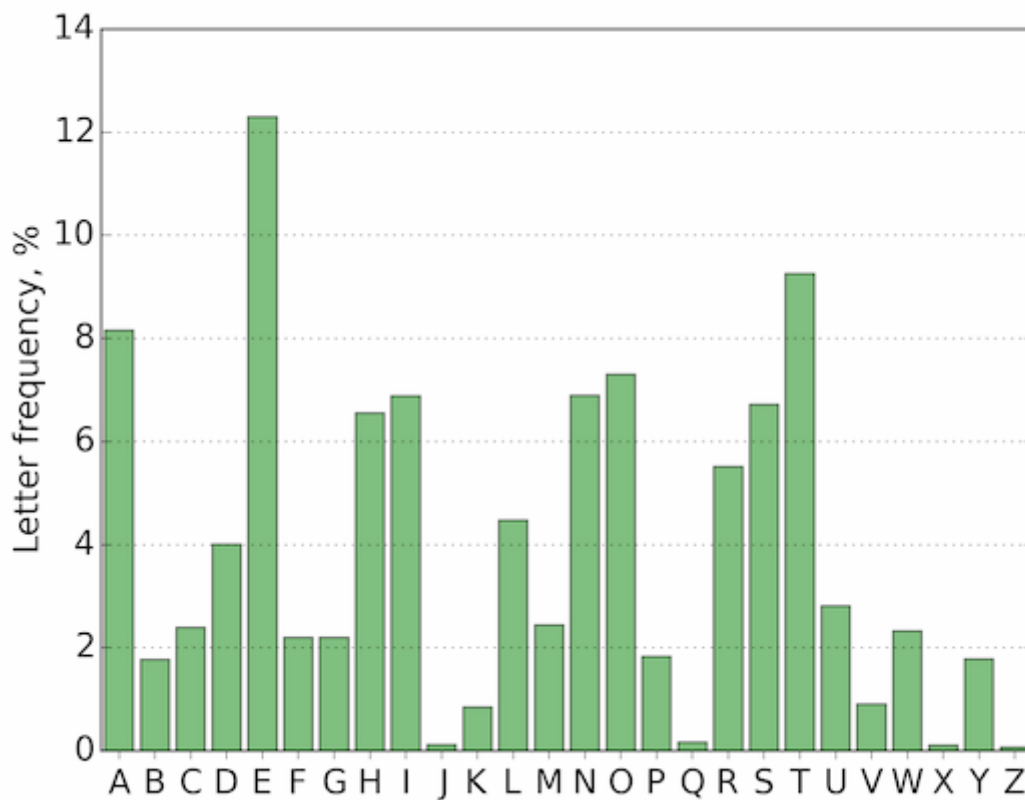


Figure 1 : Letter Frequency in English

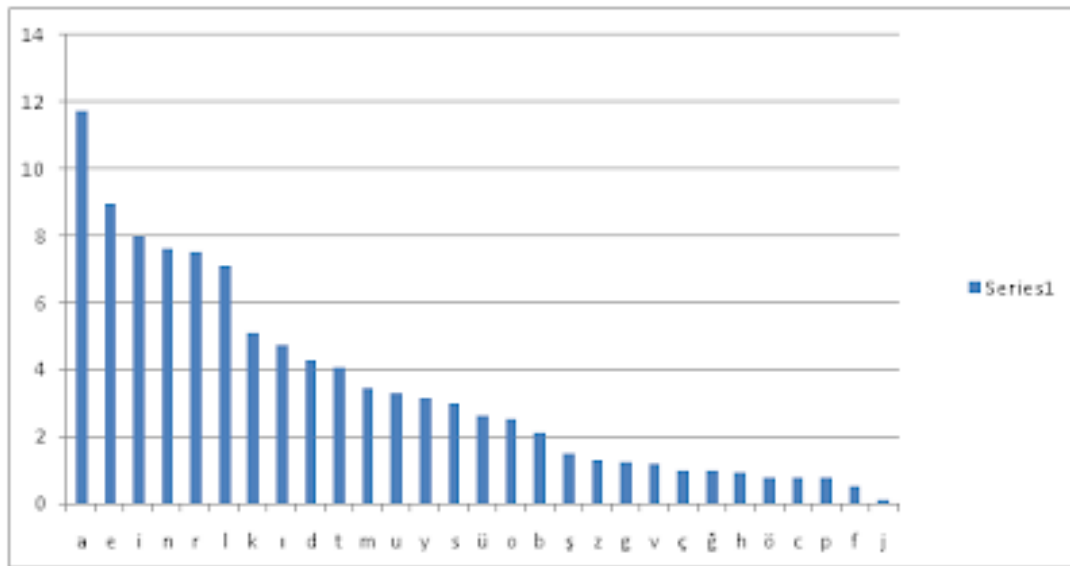


Figure 2: Letter Frequency in Turkish

Letter frequency analysis is useful for decrypting a ciphertext. With brute force, the time-consuming process is shortened. This method is used to decrypt texts encrypted with substitution encryption. Substitution is the replacement of letters with other letters with the help of a key in encryption. While the order of the letters remains the same, another letter is written for each letter instead of the letter. [3]

Example for Letter Frequency

Below is a summary of the Hansel and Gretel story. Letter frequency analysis of this text will be done.

Hansel and Gretel

Hansel and Gretel are brother and sister who lived in a small, humble house with their father and stepmom. Their house was at the edge of the forest, far away from all others. Since they were very poor the stepmom suggested to her husband that they should leave their kids somewhere deep inside the woods so they would survive. After they were left in the forest Hansel and Gretel found their way back because Hansel threw stones while they were walking so they just followed them back. The stepmom was persistent and suggested they left the kids in the forest again. This time they couldn't find their way back. They were in the forest, hungry and scared. They decided to find a shelter and they came across a house made of cake and bread. They were so hungry that they just started chewing the house. An older woman came out of the house and called them to come inside. She made them delicious meals and put them to bed. The kids didn't suspect she was a evil witch that wanted to eat Hansel. The next morning she locked

them up in a barn and made Gretel cook meals for Hansel so he would get fatter so she could eat him in the end. The witch checked his weight everyday but Hansel found a way to fool her. Instead of his finger he would have the witch a little bone to touch so that she would think he gained no weight. After a month of waiting the witch got tired and decided to eat him, no matter what his weight was. She ordered Gretel to get inside the oven and check if the temperature is alright. The girl knew what her intentions were so she pushed her inside the oven and she burned to death. Gretel freed Hansel and before they went home they gathered all of the gold and took it with them. They decided to go home to their father. He was thrilled about their arrival and they lived happily ever after. The stepmom died while the kids were held captive. [4]

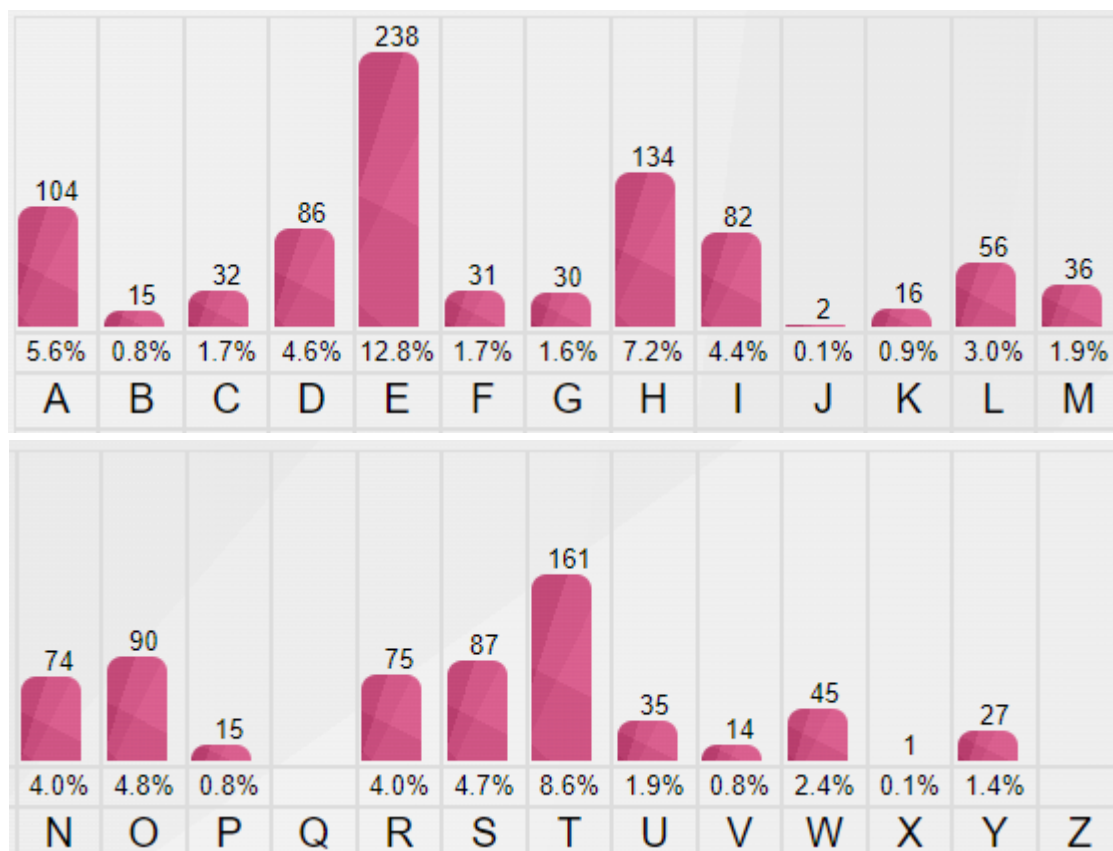


Figure 3: Letter Frequency for the Text

When the letter frequency for the text in the example is examined, it is seen that it is approximately similar to figure 1. The most common letters in this example are e with 238 repetitions, t with 161 repetitions, h with 134 repetitions, and a with 104 repetitions. The least repeated letters are j, x, q and z. While the letter j is repeated 2 times in the text and the letter x is repeated once in the text, the letters q and z are absent in the text.

REFERENCES

1. 101 Computing. (2019). Frequency Analysis. Retrieved from <https://www.101computing.net/frequency-analysis/>
2. Wikipedia contributors. (2021). Frequency analysis. In *Wikipedia, The Free Encyclopedia*. Retrieved from https://en.wikipedia.org/wiki/Frequency_analysis
3. Wikipedia contributors. (2022). Substitution cipher. In *Wikipedia, The Free Encyclopedia*. Retrieved from https://en.wikipedia.org/wiki/Substitution_cipher
4. Book Reports. (n.d.) Hansel and Gratel Summary. Retrieved from <https://www.bookreports.info/hansel-and-gretel-summary/>