Cryptanalysis on Caeser Cipher

Kunal Jawale

10/12/2016

1 Abstract

Being the simplest keyed cipher, the Caesar cipher can be broken in milliseconds using automated tools. Since there are only 25 possible keys (each possible shift of the alphabet), we just try decrypting the ciphertext using each key and determine the fitness of each decryption. This form of solution is known as a cyrptanalysis solution, and is only possible for the very simplest of ciphers.

The crux of the approach depends on determining the fitness of a piece of decrypted text using chi-squared statistic. Chi-squared statistic is based on frequency distribution of English Alphabets.

2 Cryptanalysis

Since we have 25 possible keys to a message, we need to determine which key is exact message. The Chi-squared Statistic is a measure of how similar two categorical probability distributions(frequency distribution) are. If the two distributions are identical, the chi-squared statistic is 0, if the distributions are very different, some higher number will result. The formula for the chi-squared statistic is:

$$X^2(C,E) = \sum_{i=A}^{i=Z}(C_i - E_i)^2/E_i \tag{1}$$

where CA is the count (not the probability) of letter A, and EA is the expected count of letter A. Here expected count is taken as relative frequncy of a letter in English Alphabet distribution. (Please check graph attached seperately for frequency distribution of Englidh Alphabets).

3 Conclusion

Cryptanalysis of Caeser Cipher is performed using frequency analysis(chi-squared statistics).

4 Observations

Frequency analysis can be helpful in doing cryptanalysis of many cipher techniques. There are many algorithm/techniques avaialble such as bigram, quadgram, fitness measurement, index of coincidence, etc that can be helpful in implenting cyrptanalysis.

5 Suggestions

The implemented method can be very useful in performing cryptanalysis of Polyalphabetic Ciphers also.