Report on Lab 2 - Attacking Classic Crypto Systems

Introduction:

In this report, the procedure followed in order to attack two classic crypto systems, namely the Caesar Cipher and the Simple Substitution Cipher, as per the requirements of Lab 2: Attacking Classic Crypto Systems, would be discussed. The main aim is to point out the vulnerabilities in these systems as a means of decrypting the respective ciphers.

Checkpoint 1: Attacking the Caesar Cipher

Methodology: Brute Force Attack
A Caesar cipher is a kind of substitution cipher in which every letter in the message is replaced by another letter a fixed number of positions down the alphabet. Because the English alphabet has only 26 possible keys, the best possible approach to crack a Caesar cipher would be a Brute Force Attack.

Weaknesses of Small Key Space: The key space for the Caesar cipher is very small, namely 25 possible non-zero shifts. It would be very cumbersome to check all 25 possible decryptions by hand. It could be done instantly by a computer.

Automation: The best approach would be to create a function named caesar_decrypt, which would require the cipher text as well as the key, then proceed to print the decrypted message.

Iteration: A simple loop is then used, which iterates over all possible keys from 1 to 25. The decrypted string is then printed for every key. Identification:The purpose here is to scan the 25 output messages to identify the string whose characters form a coherent English message. The plaintext message will make it apparent which is the correct key.

Checkpoint 2: Attacking the Simple Substitution Cipher

Methodology: Frequency Attack on Substitution Ciphers
The simple substitution cipher replaces every letter in a plaintext message according to a fixed substitution, where a given letter in the ciphertext always stands for a given letter in the plaintext. The number of possible keys in a simple substitution cipher is 26 factorial, making it unfeasible to use brute-force. Hence, the best way to attack it would be a frequency attack.

Principle of Frequency: The principle of frequency states that every language has a distinct frequency distribution of letters55. For English, it's a well-known fact that the most common letters are E, T, A, O, I, N, S, H, R etc.

Ciphertext Analysis: To attack a ciphertext, we could find a mapping based on the frequency of letters in the ciphertext. The most common letter in the ciphertext would correspond to English's common letter E, the next common one would be English's common letter T and so on.

Code Implementation: The calculate_frequency function has been coded as follows:
1. Remove spaces as well as punctuation to include only letters.
2. List the frequency of occurrence of A-Z.
3. Arrange it according to frequency from highest to lowest.
4. Display it along English frequency order.

Decryption
This frequency mapping table would be our first guess. Then on the basis of this, the final decryption, whether it's manual substitution decryption or automatic decryption, would require searches starting from a guess. Key substitutions would be confirmed on the basis of common phrases in English, such as "the", "and", "a", "is".

Which input was easier to break?
Cipher-2 could be broken relatively easily.

Explanation: The susceptibility of a substitution cipher to a frequency analysis attack is a direct function of the length of the ciphertext.

Statistical Reliability: The total number of letters in Cipher-2 is 1547, which is much larger than 406, the total number of letters in Cipher-1.

Convergence to English: The longer the example (Cipher-2), the stronger the probability of having a close match to the average English frequency distribution of characters. Thus, the starting guess, where the most common cipher symbol is matched to 'E,' the second most common to 'T,' and on, is much more plausible.

Close Rank Matching:

In the Cipher-2, the frequency of the top letter K, 10.53%, is close to the expected frequency of E, 12.22%.

The value of the second L is 8.00%, very close to T's value of 9.67%.The guess based on Cipher-2's frequency table would, therefore, involve fewer attempts at refinement than the guess based on Cipher-1's frequency table. Conclusion: The longer length of Cipher-2 allows for a more precise set of statistical information, so the frequency of letters in the Cipher-2 cipher is a truer indication of an English message, making it simpler to determine the keyword to unscramble the cipher.