Optional Homework (9)

Statistical Analysis

Frequency Distribution

The **simple substitution cipher** replaces each letter of the original text with another letter according to a substitution key. This type of cipher alters the correspondence between the original letters and the ciphered letters, but **does not change the frequency of the letters** in the ciphered text compared to the original text. The encrypted letters retain the same frequency as the original letters, only with a different label.

This means that a frequency analysis on the cipher text can reveal information about the original language and potentially allow decryption of the message through frequency-based cryptanalysis techniques.

Effect on Frequency Distribution

- **Original Text**: has a frequency distribution characteristic of the language in which it is written. For example, in Italian, letters such as ‘A’, ‘E’, ‘I’ are more frequent.
- **Crypted Text**: Even if the letters are replaced, the frequency distribution remains the same in terms of quantity, but associated with different letters.

Entropy

Entropy measures the degree of uncertainty or randomness in a data set. In information, the entropy H of a discrete source can be calculated as:

$$H = -\sum_i p_i \log_2 p_i$$

where p_i is the probability of occurrence of the symbol i.

- **Entropy of Original Text**: Reflects the predictability of the text based on the frequency of letters in the language. Languages with non-uniform frequency distributions have lower entropy.
- **Entropy of Cipher Text**: In the case of a simple substitution cipher, the entropy remains the same as in the original text, since the probabilities p_i do not change, only the symbols are relabelled.

Permutation Phase

The **permutation** reverses the order of the letters in the cipher text.

Permuting the order of letters does not alter the frequency of each letter in the text. Each letter appears the same number of times both before and after the permutation.

- **Frequency Distribution Unchanged**: The frequency of individual letters remains the same, but permutation can break up common letter sequences, making n-gram based analysis (bigrams, trigrams, etc.) more difficult.

Since the probabilities p_i of each symbol do not change, the entropy of the message remains unchanged even after permutation.

- **Less obvious Patterns**: Permutation breaks up the structure of the text, making position-based patterns, such as repetitions of words or common sequences, less obvious.
- **Difficulties in Sequential Statistical Analysis**: Tools that analyse the sequence of letters are less effective.

Encryption/Decryption

Decryption by Frequency Analysis

- **Letter Frequency**: By analysing the frequency of letters in the cipher text and comparing it with the known frequencies of the language, assumptions can be made about substitutions.
- **Letter Patterns**: Despite permutation, certain letter sequences may suggest certain common words or letters.

Deciphering with Substitution Key

If you know the substitution key, you can decipher the message by reversing both substitution and permutation.

Statistical Discussion

Changes in Frequency Distribution after Permutation

- **Invariance of Individual Frequencies**: The frequencies of individual letters remain unchanged.
- **Alteration of Sequence Frequencies**: The frequency of bigrams and trigrams changes, making the analysis based on these sequences less effective.
- **Significance in Cryptography**: Altering sequence structures increases security against attacks that exploit positional patterns.

Entropy Considerations

- **Invariant Entropy**: Since letter probabilities do not change, entropy remains the same.
- **Security Implications**: An unchanged entropy implies that the overall level of uncertainty does not increase, but permutation can complicate the analysis without increasing entropy.

Comparison with RSA

- **RSA and Frequency Distribution**: The RSA algorithm, being based on mathematical operations on integers, tends to **disrupt the frequency structure** of letters completely. The numerical transformations make the frequencies of the letters in the ciphertext uniform or seemingly random.
- **Key Management**: RSA uses a public key for encryption and a private key for decryption, further complicating the possibility of decryption without the correct key.
- **Decryption Difficulty**: Unlike substitution ciphers, RSA is not vulnerable to frequency analysis.

Entropy and Security

- **Importance of Entropy**: A higher entropy in the encrypted message increases security, making it more difficult for an attacker to predict or analyse the content.
- **Entropy Increase**: Techniques such as using ciphers that produce uniform frequency distributions or adding operations such as XOR with a random key can increase entropy.
- **AES and Entropy**: The Advanced Encryption Standard (AES) significantly changes the frequency distribution and entropy of the original text through complex transformations, including multiple substitutions, permutations and XOR operations with a key. This makes it extremely difficult to analyse the cipher text using statistical methods.