Confusion and Diffusion – Cryptography

- **Confusion** and **Diffusion** are both cryptographic techniques that enhance the security of ciphers by making the relationship between the statistics of the ciphertext, the encryption key, and the plaintext as complicated and obscure as possible.

- Claude Shannon was one of the first to propose these techniques in order to make it very difficult if not technically impossible to break ciphers using statistical analysis.

- Block ciphers rely on confusion as well as diffusion, while stream ciphers only use confusion.

- Generally, **confusion** obscures the relationship between the **plaintext** and **ciphertext**, while **diffusion** dissipates the **plaintext statistics** through the **ciphertext**.

- A stream cipher is simply a weaker version of a one-time pad and hence stream ciphers employ only confusion.

- Block ciphers spread any local statistics throughout the block, thus implementing the principle of diffusion.

- Modern block ciphers employ both confusion and diffusion.

- The Vigenère and substitution ciphers do not have very poor properties of diffusion and confusion, hence their susceptibility to frequency analysis.

- The Vigenère cipher provides small-scale confusion: digrams are encoded by different keys, but does not provide any kind of diffusion: symbols are not moved around or spread.

- Simple transposition ciphers provide diffusion: they move symbols around, but it is difficult to achieve confusion.

- Modern ciphers use a combination of **substitution** and **permutation** to implement diffusion and confusion in what is known as **product ciphers**:

    - Substitution (confusion): replace one (symbol/bitstring) by another reversibly.
    - Permutation (diffusion): change the order of (symbols/bitstrings) reversibly.

Confusion

- **Confusion** obscures the relationship between the **ciphertext** and **key**. Its property is such that a **single bit change** in the key causes **many changes** in the ciphertext bits.
- For example, each binary digit (bit) of the ciphertext should depend on several parts of the key, obscuring the connections between the two.
- Specifically, the relationship between the statistics of the **ciphertext** and the value of the **encryption key** is designed to be as complex as possible.
- Even though the cryptanalyst is able to get some control over the statistics of the ciphertext, the key cannot be deduced.
- Confusion is achieved by using **substitution** algorithms and complex **scrambling** of the ciphertext.
- For example, a one-time pad relies entirely on confusion. A simple substitution cipher is another (very weak) example of a confusion-only cipher.
- In DES, the ***S-boxes*** (Substitution boxes) provide confusion rather than diffusion.
- In AES, the ***Add Round Key*** and ***Substitute Bytes*** implement provide confusion.
- Stream ciphers generally rely on confusion.

- **Diffusion**

- Diffusion distributes the plaintext statistics throughout the ciphertext in order to obscure the statistical characteristics of the plaintext.

- The affect of each individual plaintext digit is spread out over many of the ciphertext digits. The main principle is that bits from multiple positions in the original plaintext will contribute to a single bit in the ciphertext.

- For example, if a single bit of the plaintext is changed, then (statistically) half of the bits in the ciphertext should change.

- Similarly, if one bit of the ciphertext is changed, then approximately one half of the plaintext bits should change in such a way that the statistical properties of plaintext are be hidden.

- The statistical structure of the plaintext vanishes into the long-range statistics of the ciphertext, thus making the relationship between them so complex that the deduction of the original key is almost impossible.

- In block ciphers, diffusion is achieved by applying **permutation** on the data with a transformation function using a key.

- The permutation function ensures that the statistical structure of the plain text is dissipated throughout the long-range statistics of the cipher text.

- The classical double transposition cipher is an example of a diffusion-only cryptosystem.

- In **DES**, the ***p-boxes*** (permutation) boxes implement diffusion rather than confusion.

- In **AES**, the ***ShiftRows*** and ***MixColumns*** functions provide diffusion.