MysteryTwister C3

THE CRYPTO CHALLENGE CONTEST

Homophonic Encrytion – Part 3

Authors: Nils Kopal, Bernhard Esslinger

May 2019

Introduction (1/2)

A *monoalphabetic substitution cipher* is a cryptographic technique which uses only one (fixed) ciphertext alphabet for encryption.

A modified version of the monoalphabetic substitution is the *homophonic encryption*. This encryption method was already widely used in the 17th century. In this case (in comparison to the „simple“ monoalphabetic substitution), one plaintext symbol (character) does not need to be mapped on the same ciphertext symbol every time.

Introduction (2/2)

Instead, for each plaintext symbol you have the choice to substitute it by one of several (different) ciphertext symbols. This means that the alphabet for the ciphertext is longer than the alphabet for the plaintext.

The aim of the homophonic substitution is to smoothen the different frequencies of the plaintext symbols so that the ciphertext symbols occur roughly equally frequent.

Example (1/2)

In English texts, the frequency of the letter „A“ is roughly 8.17 % and the frequency of the letter „B“ is about 1.49 %. We consider an alphabet for the ciphertext with 100 symbols so that the symbols of the ciphertext are almost uniformly distributed.

More precisely, we substitute the 26 letters of the English alphabet by the numbers „00“ up to „99“. Now we can substitute 8 possible numbers for the letter „A“, and one number for the letter „B“. If we continue this procedure, we achieve that analyzing the frequency of the ciphertexts symbols brings almost no advantage any more.

Example (2/2)

The key of a homophonic substitution shows how the plaintext symbols can be substituted.

If we know part of the key (substitution table)

- G:[26|43]
- O:[02|34|44|52|57|66|98]
- D:[28|45|63|95]

we can decrypt the example ciphertext „43 57 98 45“:
„G O O D“

Challenge (1/2)

This is a ciphertext-only challenge: Given a ciphertext and the information, which cryptographic technique is used, you as a cryptanalyst have to find the plaintext and extract a specific word from the plaintext.

The alphabet of the plaintext consists of the English alphabet (only uppercase letters, in total 26 symbols).

We consider a uniform distribution of the homophones. Every plaintext symbol should have three possible substitutions.

Challenge (2/2)

The alphabet of the ciphertext consists of 78 (26*3) different symbols. Every ciphertext symbol is represented by two capital letters.

The unknown plaintext consists of 306 symbols. Then the given ciphertext consists of 917 characters: 306 ciphertext symbols (each consists of two characters) and 305 blanks.

You can find the ciphertext in the additional text file of this challenge.

The solution consists of the last word of the plaintext (uppercase).

Overview Series Homophonic Encryption

- Part 1: homophonic encryption with blank (uniform distribution of homophones)
- Part 2: homophonic encryption with blank (distribution based on language statistics)
- Part 3: homophonic encryption without blank (uniform distribution of homophones)
- Part 4: homophonic encryption without blank (distribution based on language statistics)