MysteryTwister C3
THE CRYPTO CHALLENGE CONTEST

A MODIFIED HOMOPHONIC CIPHER WITH REDUCED ALPHABETS – PART 2

Author: Bernhard Esslinger

October 2010 (Updated: 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)

By assumption let the plaintext alphabet be $A = \{0, 1\}$ and the ciphertext alphabet $B = \{y, z\}$. In common texts the „0“ occurs with a probability of 60 %, and the „1“ with a probability of 40 %.

If one encrypts the plaintext with a „simple“ monoalphabetic substitution (e.g. by replacing „0“ by „y“ and „1“ by „z“), the ciphertext shows the same frequency distribution of single symbols as the plaintext. An attacker would just have to calculate the frequencies of „y“ and „z“ and compare them to the frequencies of „0“ und „1“. That way he very probably gets the right substitution.

Example (2/2)

A slight increase of security is gained by using a homophonic substitution.

So, for instance, a homophonic substitution cipher could replace „0“ by „v“, „w“ or „x“, and „1“ by „y“ or „z“. This implies that the ciphertext is built of the alphabet $B = \{v, w, x, y, z\}$.

If „v“, „w“, „x“, „y“ and „z“ have been chosen adequately before encryption and an attacker now calculates the frequencies of the ciphertext symbols, he finds out that „v“, „w“, „x“, „y“ and „z“ occur equally frequent. In this case the frequency of single symbols doesn't give an indication for an attack.

Challenge (1/3)

We again assume that our words in plaintext form are made up of the signs „0“ and „1“, i.e. the plaintext alphabet is $A = \{0, 1\}$. Like in the example, the ciphertext alphabet is $B = \{v, w, x, y, z\}$.

In common texts the „0“ occurs with a probability of 60 %, and the „1“ with a probability of 40 %. Furthermore, the bigrams „00“, „01“, „10“ respectively „11“ occur with a probability of 32 %, 29 %, 29 % respectively 10 %.

Challenge (2/3)

1. Calculate the frequency distribution of the bigrams of the following ciphertext.
2. Afterwards, determine the most probable plaintext.

vwzzzxyyyxyywzzxzwyyyxzwywvxvvwwvywyzyyyvwzzxyyxvx
ywvvxxvxvxzxvvvvwyyyzzzwyyyxzzzwzzywywzxzxvvxvvzxzzy
zyzwvywywwvxyywxzzzxvzvzzwvxzxvvwzzxywvwyyyxvvvwyxv
xvxxzzzzxvxvxvzzzwzyyyxyywzzzzzzwzzvvzxvxxxyxyxywwzz
xxzzwwvvxxxvwwvvvvwvxxyyywvvwwyzxzwyywyzyzvvxxvvxxxvw
wyywzwzwwyxxyyyxxzwwwwvwwyyxvxzxxxyywwvxxxvxzzvywzzzzx

Challenge (3/3)

Hint:
If you separate the looked for plaintext (sequence of 312 „zeros“ and „ones“) into 8-bit-blocks and interpret each byte as one ASCII character, you will get the secret codeword (or rather secret sentence).

Please enter the secret sentence as the solution.