Homophonic Substitution Cipher

Introduction

The Homophonic Substitution cipher is a substitution cipher in which single plaintext letters can be replaced by any of several different ciphertext letters. They are generally much more difficult to break than standard substitution ciphers.

The number of characters each letter is replaced by is part of the key, e.g. the letter 'E' might be replaced by any of 5 different symbols, while the letter 'Q' may only be substituted by 1 symbol.

The easiest way to break standard substitution ciphers is to look at the letter frequencies, the letter 'E' is usually the most common letter in english, so the most common ciphertext letter will probably be 'E' (or perhaps 'T'). If we allow the letter 'E' to be replaced by any of 3 different characters, then we can no longer just take the most common letter, since the letter count of 'E' is spread over several characters. As we allow more and more possible alternatives for each letter, the resulting cipher can become very secure.

An Example

Our cipher alphabet is as follows:

```
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
D X S F Z E H C V I T P G A Q L K J R U O W M Y B N
9       7       3         5 0       4 6
        2
        1
```

To encipher the message `DEFEND THE EAST WALL OF THE CASTLE`, we find 'D' in the top row, then replace it with the letter below it, 'F'. The second letter, 'E' provides us with several choices, we could use any of '2', '7', '2' or '1'. We choose one of these at random, say '7'. After continuing with this, we get the ciphertext:

```
plaintext:  DEFEND THE EAST WALL OF THE CASTLE
ciphertext: F7EZ5F UC2 1DR6 M9PP 0E 6CZ SD4UP1
```

The number of ciphertext letters assigned to each plaintext letter was chosen to flatten the frequency distribution as much as possible. Since 'E' is normally the most common letter, it is allowed more possibilities so that the frequency peak from the letter 'E' will not be present in the ciphertext.

Cryptanalysis

Breaking homophonic substitution ciphers can be very difficult if the number of homophones is high. The usual method is some sort of hill climbing, similar to that used in breaking substitution ciphers. In addition to finding which letters map to which others, we also need to determine how many letters each plaintext letter can become. This is handled in this attempt by having 2 layers of nested hill climbing: an outer layer to determine the number of symbols each letter maps to, then an inner layer to determine the exact mapping.

Contents

- Introduction
- An Example
- Cryptanalysis

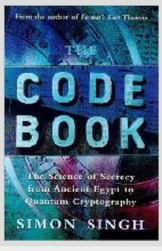
Further reading

We recommend these books if you're interested in finding out more.

Elementary Cryptanalysis: A Mathematical Approach

ASIN/ISBN: 978-0883856475

Buy from Amazon.com

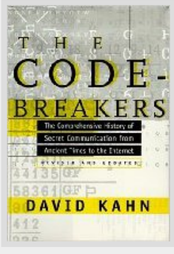


The Code Book: The Science of Secrecy from Ancient Egypt to Quantum Cryptography

ASIN/ISBN: 978-1857028799

"Simon Singh's 'The Code Book' is an excellent introduction to ciphers and codes"

Buy from Amazon.com



The Codebreakers - The Story of Secret Writing

ASIN/ISBN: 0-684-83130-9

Buy from Amazon.com

y ngp'i zpgo avce ge lgm avce vj oscc vj y jagmcn cyzs; vpn y cyzs csjj iavp avce ge lgm avce vj oscc vj lgm nsjsuds
– q.u.u. igczysp. (ias esccgojayk ge ias uyph)

Questions/Feedback

Notice a problem? We'd like to fix it!
Leave a comment on the page and we'll take a look.