Programming Project 1

Course 01435: Practical Cryptanalysis
June 2013

Andrey Bogdanov
anbog@dtu.dk

1 Remarks

- hand in via campusnet before Thursday, June 27, 11:59 pm (no print out necessary)
- work alone or in groups of up to 4 people
- free choice of programming language
- the code should be well-structured and contain a sufficient amount of comments such that it can be understand be an external programmer
- small additional written documentation of the code is necessary
- it is your responsiblity to be able to demonstrate your program at the colloquium (make sure that a computer is available your programme runs on etc)

Programming Project (P)

The first programming project is to build a tool for analysis of a simple substitution cipher, as described in the lecture. You should proceed in the following steps:

1. Implement a tool that counts the frequencies of letters, digrams, and trigrams in a given text file.
2. Use this tool to generate your own data files for the frequencies of letters, digrams, and trigrams in British English. Remember to "cleanse" the text first by leaving only capital letters! You can use sample texts e.g. from Project Gutenberg (www.gutenberg.org). Note that there is a difference in statistics between British English and American English, so make sure you pick texts from British authors!

3. Build a tool that assists the user in manually decrypting a substitution ciphertext. The minimum requirements are the following:

   - The tool displays the ciphertext (capital letters) and under each line the decryption so far (small letters).
   - The user can ask the program to assign a certain plaintext letter to a given ciphertext letter, and the program will update the ciphertext/plaintext view accordingly.
   - The user can take back an assignment.
   - The user can ask the program to display letter frequencies in the ciphertext and in the English language.
   - The user can ask the program for the most frequent digrams / trigrams in the ciphertext and in the English language.

If you want to do even more: (not mandatory!)
Extend your program to go all the way, i.e. to decrypt a substitution cipher automatically.

- Upon request, the program can offer the user a list of likely plaintext letters for each ciphertext letter. The plaintext letters should be chosen with a χ^2 test such that the error probability is under a certain threshold (e.g. less than 1%).
- Upon request, the program conducts the remaining analysis on its own, using the algorithm described in the lecture.
  Note: This is a difficult programming task; do not attempt it unless you have finished all of your mandatory tasks!