This assignment consists of practical experiments with historical ciphers. It is expected that you will make use of the CrypTool package to perform the experiments. Although no other tools should be necessary, you are free to use any other software or to write your own. Note that different versions of CrypTool have slightly different tools.

For each experiment, describe clearly what outputs you found and explain why you think these results occurred using the theory that we have studied in the lectures. Use tables and figures as appropriate. This is an individual assignment. It is acceptable to discuss the general approach with your fellow students and the unit staff, but your answers should be your own.

Submission

There are 20 marks available in total for this assignment with part marks as shown. Your answers should be written in a report and submitted by the due date of **Friday 17 August, 2018 at 11:59pm**. Submission should be made online via Blackboard under *Assessment > Assignment 1 submission*. ***Your submission must be in PDF format.*** If you are submitting a scan, make sure that the scan is high quality.

Obtaining data

You need to obtain your individual folder of 4 files. To do this first go to Blackboard and find *Assessment > Assignment 1 > Folder Assignments* and look up your folder number based on your Student ID. Then go to the folder *Assessment > Assignment 1*, click on *assignment1files* under *Ciphertext files*. Then navigate to *???-Student* where ??? is the folder number that you previously looked up, left padded with zeros. You should find four files, *0.txt, 1.txt, 2.txt, 3.txt.*. These are the files that you will use for the questions below.

Tasks

This problem solving task concerns cryptanalysis of historical ciphers. You are given four ciphertext files. These are all written in English with a similar style of text. Each text is different and each is encrypted with a different cipher. The four ciphers used, in random order in your set, are:

- random simple substitution cipher

- Vigenère cipher
- transposition cipher
- 2×2 Hill cipher

All plaintexts are written in English and use upper and lower case letters.

1. [**4 marks**] For each ciphertext, using either a table or a histogram, present the following characteristics:

   - the single character frequencies;
   - the 20 most common digram frequencies;
   - the autocorrelation;
   - the index of coincidence.

   Include, for *each ciphertext* a description of the relevant properties of *each characteristic*.

   Summarise your results by providing a 4 by 4 grid, characteristics on one axis and ciphertexts on the other, with a brief comment in each square.

2. [**4 marks**] Explain clearly how each of the four characteristics measured in question 1 is expected to look for the ciphertext from each of the four cipher algorithms used.

   Organise your answer by providing a 4 by 4 grid, characteristics on one axis and ciphers algorithms on the other, with a brief comment in each square.

   Use the measured values to argue which ciphertext comes from which of the four ciphers.

3. [**4 marks**] Explain, with direct reference to the statistics that you found, the procedure you can use to cryptanalyse each cipher. You are not expected to perform the cryptanalysis in this part — you need to explain how it can be done in principle for each type of cipher and how the measured statistics can help.

4. [**3 marks**] Obtain the plaintext for the random simple substitution, the Vigenère, and the transposition cipher, explaining how this was done. (You can use the automatic tools in CrypTool or other where applicable. If you are using a tool, reference it.)

5. [**5 marks**] Now attempt to cryptanalyse the 2×2 Hill cipher as follows.

   (a) Determine the non-overlapping digram frequencies. Note that using the n-gram tool in CrypTool directly will not work because it counts overlapping bigrams. It can be achieved in CrypTool by:

   - defining the alphabet to include uppercase and lowercase letters with no space;
   - using the format text document tool to remove spaces;
   - using the format text document tool to split into blocks of size 2;
   - using the n-gram analysis tool to count the bigrams.

   (b) Next, try to find the bigram which maps to 'th'. This should be one of the most common of your non-overlapping bigrams. Because spaces in the plaintext are preserved in the ciphertext you can use the word structure of the original ciphertext to rule out most options. For example, 'th' will not occur on its own as a word and you are likely to see it at the start of some 3-letter words.

   (c) Next, try to find at least two more likely candidates for bigram plaintext/ciphertext pairs. Possible plaintext bigrams you might search for are:

   - 'ti' - this bigram cannot usually end a word or be a word on its own.
   - 'on' - this can be a word on its own or part of another word.
   - 'he' - this is likely to occur at the end of common 3-letter words and could be a word on its own.

   (d) Once you have identified likely candidates for two bigrams, perform the known-plaintext attack that we covered in the lecture, assuming that you know that your guesses are correct.

   You can get full marks for this question without finding the plaintext as long as you complete the method described above and show that at least five different valid candidate pairs do not result in a sensible plaintext.

Hints

1. The Hill cipher used here assumes that the character 'A' maps to the number 0. This is not the default in CrypTool1. You can change this by clicking the "Further Hill options" button.

2. The transposition cipher tool in both versions of CrypTool allow you to set various things to rows or columns. The encryption was done reading in by rows, permuting by columns, and reading out by rows. This is equivalent to the method used in class.

3. CrypTool can be installed without special permissions, so you should be able to install it on lab computers if needed.