Project 1

Introduction

You are interning at the super secretive SLA (Three Letter Agency). The SLA has intercepted 50 encrypted documents from hostile enemy sources.

Your job. You have been tasked with the following:

- Decrypt the ciphertexts as best as possible. This means decrypting as many of the documents, and decrypting as much of each document as you can.
- Determine which sources sent which messages.
- Determine what ciphers were used by each source, to figure out the enemies' cryptographic capabilities.
- Report on any other intelligence gathered.

0 Part 0: Find a team

Due Friday, February 9, 11:59pm.

Cryptanalysis is a team effort. Therefore, form teams of **two or three** students to work on decrypting the documents. Try to form your teams in class, but also feel free to use the "Search for Teammates" functionality on Piazza.

If you are having trouble finding a team, please email the course staff.

0.1 Submission Instructions

Once you have found your team, go to the group sign-up sheet here: https://docs.google.com/spreadsheets/d/1TZt8oSrAvSV5NoCBBAlQOPJDAwLKhgfOAatQQzuj56g/edit?usp=sharing. Add your team members and preferred emails to the signup sheet (optionally, you can also specify a team name).

1 Part 1: Basic Analysis

Due Tuesday, February 20, 11:59pm.

Your first task will be to guess as much information as possible about the form of the cipher used to encrypt each plaintext, and to guess which plaintexts were encrypted using the same encryption method and key. This information will be crucial in Part 2, when you will actually decrypt the messages. This component of the project will also allow us to provide your group with valuable early feedback to make sure you are prepared for Part 2.

What you know. You have been told the following about the documents:

- They are from multiple sources. As such, multiple encryption methods may have been used. However, all information identifying the source has been lost.
- All of the documents were encrypted using paper and pencil ciphers. You expect the messages to be encrypted using ciphers similar to the ciphers you have seen in class, though there may be variations.
- There may be multiple documents from the same source and as such, multiple messages may by encrypted using the same method and key.
- All plaintexts are predominantly English language. All letters are lower case, and there may or may not be numerals, spaces, or punctuation.

Your Task. Using what you've learned so far, partition the intercepted documents into groups that you believe were encrypted with the same system using the same key. Conjecture about what type cipher was used for each grouping. Also, include a thorough description of your reasoning: describe what analyses you performed, what the results were, and explain how those results guided your answer. If you wrote any code to assist you, include the code in your submission.

1.1 Deliverables

Your submission consists of a writeup as well as any code you used.

The writeup consists of two sections: (1) your ciphertext groupings, and (2) your methodology.

Section 1: ciphertext groupings. list the ciphertext groups that you have found and believe are encrypted from the same source. Within each group, also write down

any information you can about the cipher used for that group (e.g. is it a substitution cipher? Mono-alphabetic/polyalphabetic? How many characters are substituted at a time? Key length for Vigenère cipher, etc.).

Please organize this information so that it is easy for us to read, such as:

```
Group 1:
     12.txt
     23.txt
     48.txt
Monoalphabetic substitution cipher, each
character mapped to 2-digit numbers

Group 2:
     ...
Group 3:
     ...
```

Also please order the ciphertexts within each group in increasing numerical order.

Section 2: Methodology and insights. In the second section, write down your methodology and any insights you gained in your preliminary analysis. How did you arrive at your groupings and decide on the encryption mechanism? Were there any patterns you observed that lead you to these conclusions? Also, please describe any code you used to assist in your analysis.

1.2 Submission Instructions

You will submit your writeup to the CS Dropbox: https://dropbox.cs.princeton.edu/COS433_S2018/Project_1. Please submit either a Microsoft Word document or a PDF (PDF preferred). The filename should be "writeup" with the appropriate extension.

2 Part 2: Decryption

Due Tuesday, March 6, 11:59pm.

Your second task is to actually decrypt the messages you have been given, and then use the answers to answer a few questions.

Your Task. Apply what you've learned so far to try to decrypt the documents you've been given. Also be sure to explain your reasoning, and include any code you used in the process of decrypting. Additionally, guess the sources of each document.

2.1 Deliverables

Writeup. Your writeup should expand on the writeup from Part 1 (in particular it should contain all of the information requested in Part 1 as well). As before, it should consist of the same two sections: the (1) ciphertext groupings, and (2) methodology. In each grouping, in addition to information about the cipher, guess the source of the ciphertext (this will always be a country). Determine the source by actually reading the decrypted plaintext you obtain. Also, very briefly describe what the group of messages appears to be about (reporting on enemy activities, giving directions to military units, etc).

In the methodology section, in addition to explaining your grouping, explain what steps you took to actually decrypt the documents.

Decrypted documents. Please submit an archive (zip, tar, or rar) containing the decrypted messages. The archive name should be "messages" with the appropriate extension.

The file names within the archive should be the same as the original source file: if the ciphertext was contained in file "37.txt", the plaintext you submit should be in the file "37.txt". You're plaintext files should contain only lower-case English letters, numerals, spaces, and punctuation.

If you have only decrypted some files, there is no need to generate empty placeholder files (but you can if you wish). Also, if you have only been able to partially decrypt any particular file, go ahead and submit what you have. Any unknown characters can just be replaced with an arbitrary character, or a "*". Note you will be given credit for any portion you get correct; points will not be taken off for incorrect guesses, so you might as well guess a random character in any position where you don't know the answer.

2.2 Submission Instructions

Your submission will consist of (1) a writeup, (2) an archive containing all of the decrypyted documents, and (3) any code you used in this part of the project. You will submit via CS Dropbox just as in part 1: https://dropbox.cs.princeton.edu/COS433_S2018/Project_1.

As before, please submit either a Microsoft Word document or a PDF. The filename

should be "writeup" with the appropriate extension. In your writeup, clearly explain what the code does and how you used it. Your writeup should replace the writeup from Part 1. Also, submit any code you used; you should make sure you have also submitted any code used for Parts 1 or 2.

3 Grading and Evaluation

The project will be out of 100 points.

- **Part 1**: 15 points
- **Part 2**: 85 points
  - Write-up: 60 points
  - Accuracy of decrypted messages: 25 points

Evaluation. Part 1 will be graded based on the thoroughness of your analysis. As this part is mostly meant to provide early feedback, the grade will mostly be on effort and the quality of your write-up.

For Part 2, there are two components to your grade. The first component (60 points) will be based on your final write-up. Here, we are looking to see that you performed any relevant analysis we learned in class and made reasonable conclusions based on the analysis.

Your decryptions in Part 2 will be scored according to the formula $s = \sum_i 0.5 \times f_i$ where f_i is the fraction of the ith message correctly decrypted. Thus if you decrypt every message perfectly, $s = 25$. Note that our scoring function means long and short messages are assigned the same number of points.

Also note that we do not expect 100% accuracy, though for this project it should be possible in principle to decrypt most of every message. Indeed, with code breaking, it is often not possible to, given a limited number of ciphertexts, decrypt every message in its entirety, and some messages might be completely undecipherable.

4 Competition for Bonus Points

Every Monday morning, we will compare your decrypted messages to the original plaintexts. The group with the highest accuracy (using the scoring function from above) at the beginning of each week will receive **2 bonus points** added to the final project score. The second-place team will receive **1 bonus point**.

There will be three opportunities for bonus pints:

- February 19th,
- February 26th, and
- March 5th

Note that this means the first bonus points will be given **before** Part 1 is due, and the last bonus points will be given out the day before Part 2 is due. So please submit your decrypted messages early and often to take advantage of this competition.

Note that when assigning final grades, any bonus points earned in the course will be ignored when deciding on the curve. Only after the curve is decided will the bonus points be added back to your score. This ensures that bonus pursuing bonus points is purely optional; if you do not earn any bonus points, it will not negatively impact your grade in the course.

5 Hints

If you get stuck, here are some hints that may be of use:

- A quick visual inspection of each document will reveal some information about the form of the cipher used.
- The alphabet used for the ciphertext may be different than the plaintext alphabet. Moreover, each character in the plaintext may correspond to multiple ciphertext characters.
- Remember that frequency analysis can be done character-by-character, but sometimes must also be performed on groupings of characters.
- For each type of cipher, think about how you would identify if that cipher is being used. What could you do to distinguish between a substitution and permutation cipher? How would you identify a Vigenère cipher? A one-time-pad?
- For each type of cipher, think about how you would identify if multiple messages were encrypted with the same key. How would you identify multiple uses of the same substitution cipher? Same Vigenère cipher? Same one-time-pad?
- Given a group of documents encrypted under the same system and same key, it will often be useful to look at all the documents in the group simultaneously to help decrypting. What should you do if you suspect a substitution cipher? What about the one-time pad?