COS 433: Cryptography (Spring 2018) Princeton University
Project 1 Due: March 10, 11:59pm

Project 1

1 Introduction

You are interning at the super secretive TLA (Three Letter Agency). The TLA 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.

2 Preliminaries: Find a team

Cryptanalysis is a team effort. Therefore, form teams of **up to four** 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. Working alone is ok, but you are strongly encouraged to find a team to work with.

2.1 Submission Instructions

Once you have found your team, go to the group sign-up sheet here: https://docs.google.com/spreadsheets/d/1_0ZDOfac5aDMq23x4nvRTgvvwp3O8pj6Gv2ry1VFhfY/edit?usp=sharing. Add your team members and preferred emails to the signup sheet (optionally, you can also specify a team name).

3 Project Details

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.

What you must figure out. Your goal is to figure out the following:

- The actual plaintexts for each message, or at least as much of each plaintext as you can
- Which ciphertexts were encrypted by the same method under the same key. Presumably these ciphertexts all came from the same source
- Any intelligence gathered as the result of decryption.

Your submission will consist of two parts: a writeup of your findings, and the actual plaintexts.

3.1 Writeup

Your writeup should be either a Microsoft Word document or a PDF (PDF preferred). The filename should be “writeup” with the appropriate extension. Your writeup should consist of the following sections:

Section 1: Basic Analysis The first step is to make an informed guess about the types of ciphers used for each ciphertext, and which ciphertexts may have been encrypted with the same cipher. For this section, perform a basic analysis of the various ciphertexts, without necessarily decrypting. Some things to think about:

- 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?
- 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. How might you determine the number of ciphertext characters corresponding to each plaintext character?

In this section, your writeup should discuss your thought process and steps you took to try to make your guesses. How did you arrive at your guesses for the ciphers used? Of course you won't be able to confirm your guess until you actually try to decrypt, so this section will be graded on the basis of how well thought out your approach is, rather than the actual guesses.

Section 2: Actual Cryptanalysis Next, you should proceed to cryptanalyze the ciphertexts group by group. For the writeup, please explain how you went about this process, and explain any code you may have used. Note that some portions of documents will not be decryptable. For these cases, try to explain why decryption is impossible.

Section 3: Summary of results In this section, you will summarize your results, namely the ciphertext groupings and a basic decryption of the cipher used (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.

4 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.

Your score for each document will be

$$1 - ed_L/\ell$$

where ℓ is the length of the document, and ed_L is the Levenshtein edit distance: the number of insertions, deletions, or replacements needed to transform the document you provided into the actual plaintext document. This means if you do not know a particular character, it is equivalent to put a single-character placeholder character like "*", or to simply omit the character. However, if you're placeholder is several characters, such as "*** unknown character ***", this will un-necessarily increase the edit distance. In fact, you might as well make a guess for the character, since you might guess correctly.

5 Grading and Evaluation

The project will be out of 100 points.

- **Writeup**: 75 points. Sections 1 & 2: 30 points each. Section 3: 15 points.
- **Decrypted documents:** 25 points, or 0.5 points per document

Remember that full decryption of every ciphertext will not be possible, so it will not be possible to actually get all 25 points for the decrypted documents. However, you should be able to get very close.

6 Hints

Through its various intelligence gathering activities, the TLA has learned several potentially useful facts:

- The following message was recovered from an unknown source:

  > From: Admiral Commanding U Boats
  > To: Entire Fleet
  >
  > Highly classified documents will be sent over the next several days. Due to heightened security concerns, documents will additionally be encrypted using the one time pad tapes sent previously.

- Below are a number of plaintext/ciphertext pairs that have been observed previously, but the sources and the ciphers used have been lost:

Plaintext	Ciphertext
“`attack`”	“457685368536457623926839”
“`careful`”	“66305866189114926341 7”"
“`enemy`”	“17583754929890125455”
“`discussion`	“`lprx rrp.a`”
“`radio`”	“661058928391182”