CSci 4554 Assignment 3

Due Monday, February 15th in class

Problem 1 (10 points). An unfair 6-sided die has the following probabilities of each side:

Value	Probability
6	30%
5	22%
4	18%
3	15%
2	10%
1	5%

Question 1. What is the entropy of one throw of the unfair die? Show your work.

Question 2. What is the entropy of one throw of a fair 6-sided die (equal probabilities of all sides)?

Question 3. Compare the two entropies. Which of the dice has more uncertainty?

Problem 2 (15 points). Group assignment (groups of 2-3): write a program to compute entropy of a language based on frequency of letters. Then use the link on the resources page to compute entropy of one of the languages given there. We would like different groups to do different languages so please use sign up on the wiki page before you choose a language. Make sure that all languages are different. One of the groups should choose English.

Individual part of the assignment: based on the computed entropy for each language explain which ones provide better cryptographic defense against frequency analysis methods?

Please submit your program and instructions on how to run it. A sample data file would help. You may use any programming language runnable in the dungeon; clear compilation/running instructions are required for less common languages.

Problem 3 (15 points). Group assignment: write a program to compute frequencies of letters in a text and the corresponding entropy. For the purposes of this assignment you may ignore all other symbols in a text. Each group will run their program on 4 English texts belonging to at least 2 different groups (technical documentation, modern literary texts, older literary texts, online news, blogs, etc.). The texts cannot be a translation. Each text must be at least a 1000 symbols.

Use Wiki discussion to coordinate this and post links to the texts analyzed as well as the result of the frequency analysis.

Individual part of the assignment: did you observe differences between the groups of text? Please explain your findings. This should be a couple of paragraphs.

Please submit your program and instructions on how to run it. A sample data file would help. You may use any programming language runnable in the dungeon; clear compilation/running instructions are required for less common languages.

Problems 2 and 3 will be graded based on your contribution to writing the programs and recording the results, as well as on your individual conclusions.

Problem 4 (5 points). Show the step-by-step work of the extended Euclidean algorithm for $a = 130$ and $b = 25$. You have to show the final represenation in the form

$$\lambda a + \mu b = gcd(a, b)$$

and show all the steps for computing λ and μ.