CSci 4554 Assignment 3

Due Friday, February 15th in class

Problem 1 (12 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 die (equal probabilities of all sides)?

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

Problem 2 (10 points). Group assignment: 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 will use Wiki to coordinate this effort).

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

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. Run this program on 12 different online texts belonging to 3 or 4 different groups (say, technical documentation, literary texts of beginning of the 20th century or even earlier, if desired, and modern texts, such as online news or blogs). Each text should be at least a 1000 symbols. Use Wiki discussion to coordinate this and store links to the texts analyzed as well as the result of the frequencies 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 para-

graphs.

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.