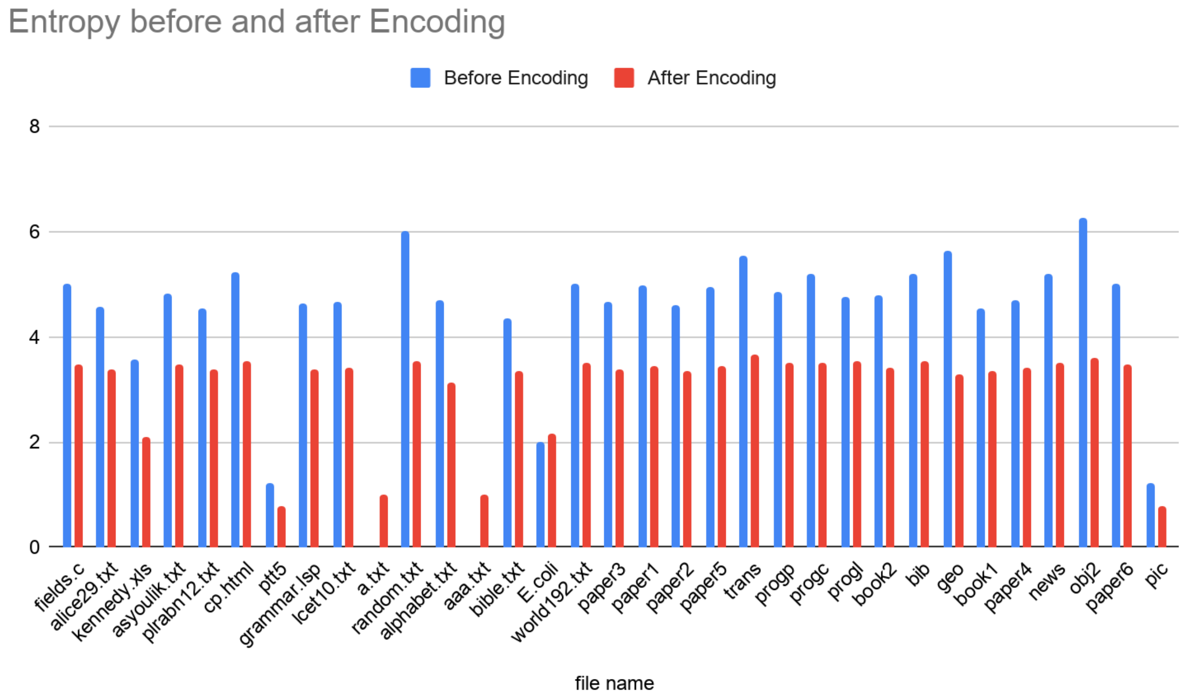
Rahul Vaidun
rvaidun@ucsc.edu
May 8 2021

CSE 13S Spring 2021
Assignment 5: Hamming Codes
Writeup

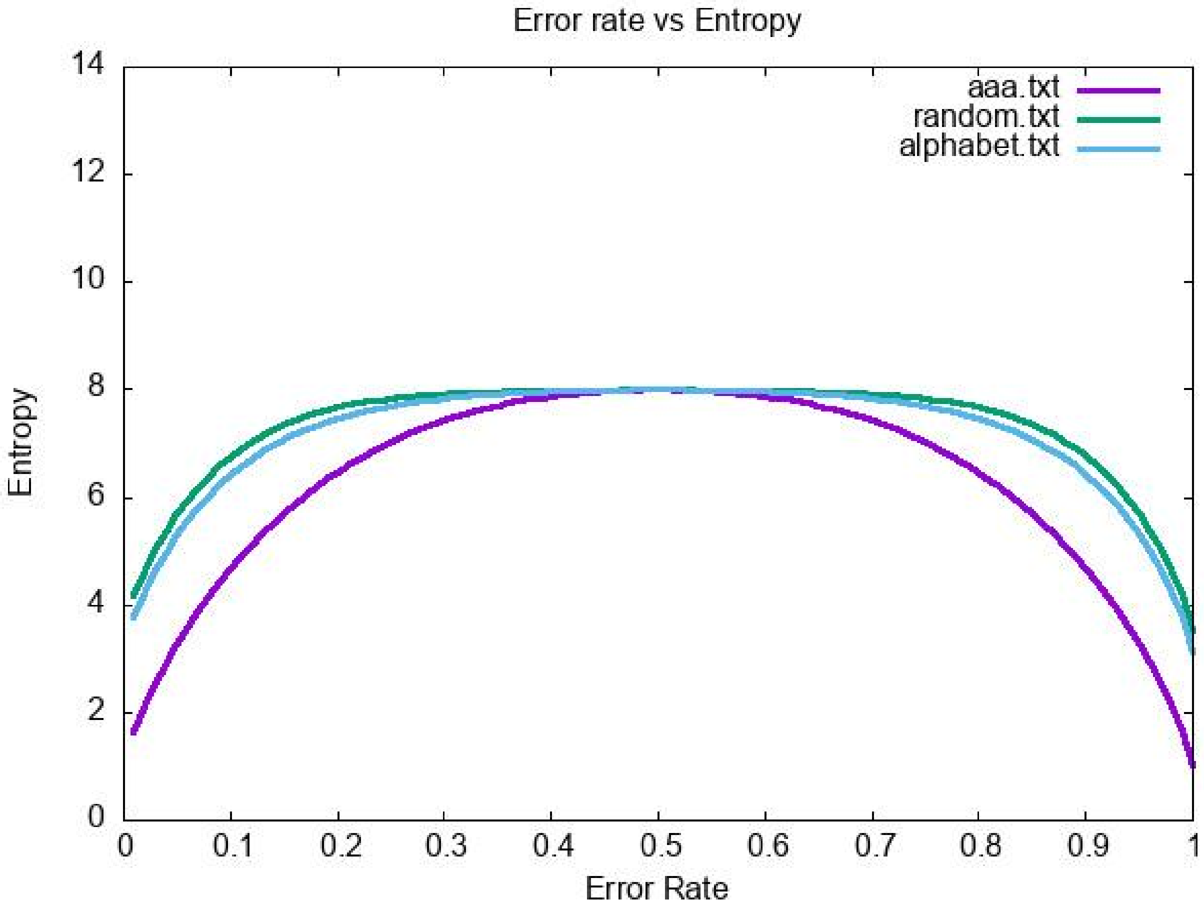
This assignment I wrote multiple C programs to encode and decode hamming codes. These codes are used to correct and detect natural noise that happens when data is transmitted over the internet. I conducted several experiments with my encoded data and decoded data to learn more about how the encoding and decoding processes worked and how errors affected the entropy or randomness of the data. The programs to introduce errors and entropy were given by Professor Long.

The first experiment I did was looking at the difference in entropy between files before they were encoded and the file after the encoding. Intuitively it may seem like the entropy should be the same but this is not the case. I graphed the results in a bar graph which is shown below.



The bar graph shows that the encoded data always has less entropy than the data before encoding except for a couple of files. a.txt, aaa.txt have more entropy but this is because these files just contained the character a. The entropy has reduced after the encoding because there are only 16 different combinations of an encoded byte while there are a lot more combinations of the byte when it is not encoded. This explains why the entropy is lower after encoding. The opposite is going to hold true when decoding a file. The decoder takes a maximum of 16 different bytes of data as input but there are 256 possible bytes that the decoder can output, which makes the entropy increase.

Next I wanted to see how noise affected the entropy of the encoded data compared to the original data. I conducted a similar experiment to the one above and noticed that the entropy increased as the error rate got closer to the middle and then decreased when the error rate was too high. The plot is shown below



This graph is pretty interesting as it shows when the error rate goes closer to the middle all the entropies also get closer to the middle but at the edges the entropies start to deviate. This graph can be explained because as the error rate gets closer to 0.5 the entropy will steadily increase but after 0.5 the entropy starts decreasing because there is a limited amount of errors that can happen in 1 byte. So when the error rate gets closer to 1 more bits start getting flipped and when every bit is flipped it the randomness goes away since it is the opposite of the original.