Abhijay Gupta
Student ID: 260708548

Comp 251 Assignment 1

Task 1

Average collisions

Task 2

Task 3

Average collisions (y-axis: 0.0, 9.0, 18.0, 27.0, 36.0, 45.0, 54.0, 63.0, 72.0, 81.0, 90.0) vs Alpha value (x-axis: 0.0, 0.4, 0.8, 1.2, 1.6, 2.0, 2.4, 2.8, 3.2, 3.6, 4.0)

Chain — Open Addressing
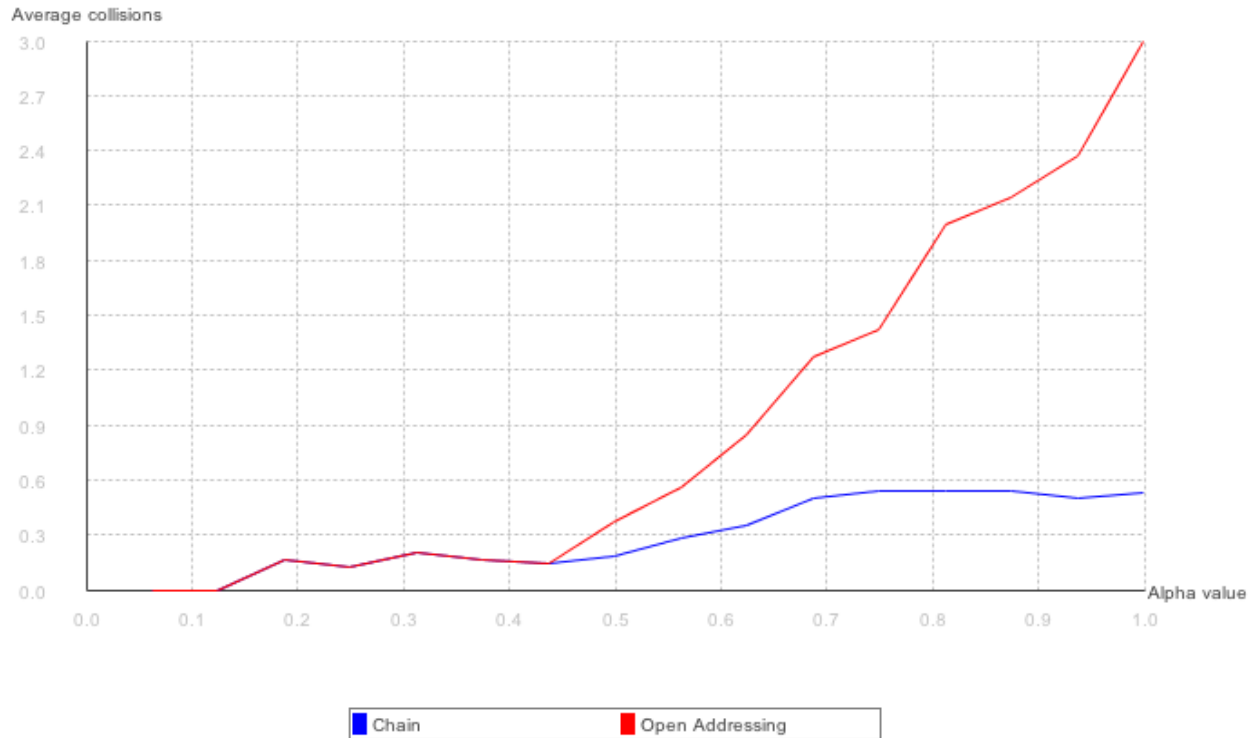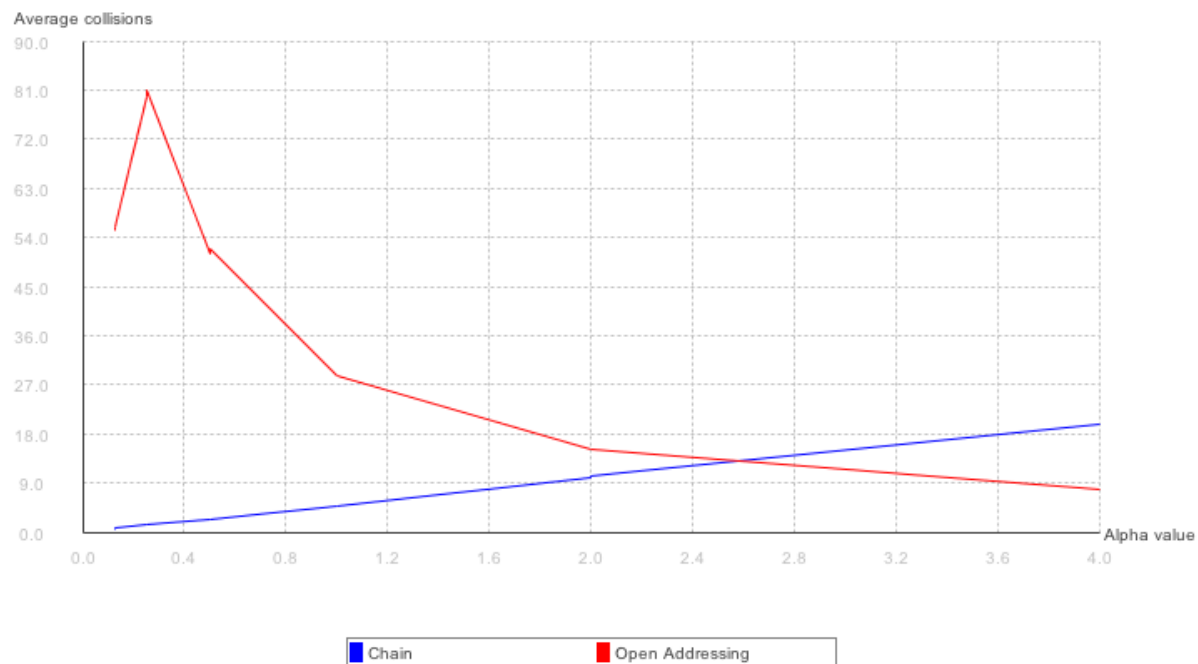
Using fixed n=32 keys for this task.

Chaining
As we can see from the graph, the average number of collisions for a Chaining table increases in a linear fashion with a low slope as the load factor (or alpha value) increases.

A low load factor means we have ample number of slots(left side of graph). This leads to less number of collisions being made. As load factor increases and we have fewer slots, the number of collisions increases. The average collisions increase in a linear fashion here.

Open Addressing
Formulae:

$$\text{Chaining}: h(k) = \left((A \cdot k) \bmod 2^w \right) >> (w - r)$$
$$\text{Open Addressing}: g(k,i) = (h(k) + i) \bmod 2^r$$
$$\text{Load factor } \alpha = \frac{n}{m}; \; r = \lceil w/2 \rceil; \; m = 2^r; m \propto w \Longrightarrow \alpha \propto \frac{1}{w}$$

High value of α corresponds with low value of w (same as low number of slots m due to direct proportionality between m and w).

The values of w that I chose were w = {6, 7, 8 ..... 15, 16}. These correspond with a minimum number of slots m = 8 (for w = 6) and maximum number of slots m = 256 (for w = 16). So, we have a range of values for m, some of which are not sufficient to store 32 keys while others are sufficient. The number of keys n, is fixed to be 32 for all the different values of w (hence different values of m).

Let's start with the right end of the graph: α = 4 (n/m = 32/8). There are 32 keys to be inserted but there are only 8 slots available. After the insertion of the first 8 keys, for the 24 keys left, they will all encounter 8 collisions since all the slots are already filled. This repetition of '8' for 24 out of the 32 keys brings the average number of collisions to 8, for w = 6 (m = 8).

So, this pattern continues (towards the left of the graph) until we have more keys to insert than the number of slots (The average number of collisions increases gradually as we obtain more slots).

When we have more slots than number of keys (lower α value), the keys go about being inserted according to their hash values. From the formula for Open Addressing, our probe number i increases by 1 each time. This ends up selecting consecutive slots upon repeated probing and it takes a while to find an empty slot. This might be what is leading to higher number of collisions on the left side of the graph.