CPSC 319 HW5 Report

Mitchell Sawatzky

The size of the table was determined from the load factor by using the formula for load factor:

$$\alpha = \frac{elements\ in\ table}{table\ size}$$

Rearranging, we get

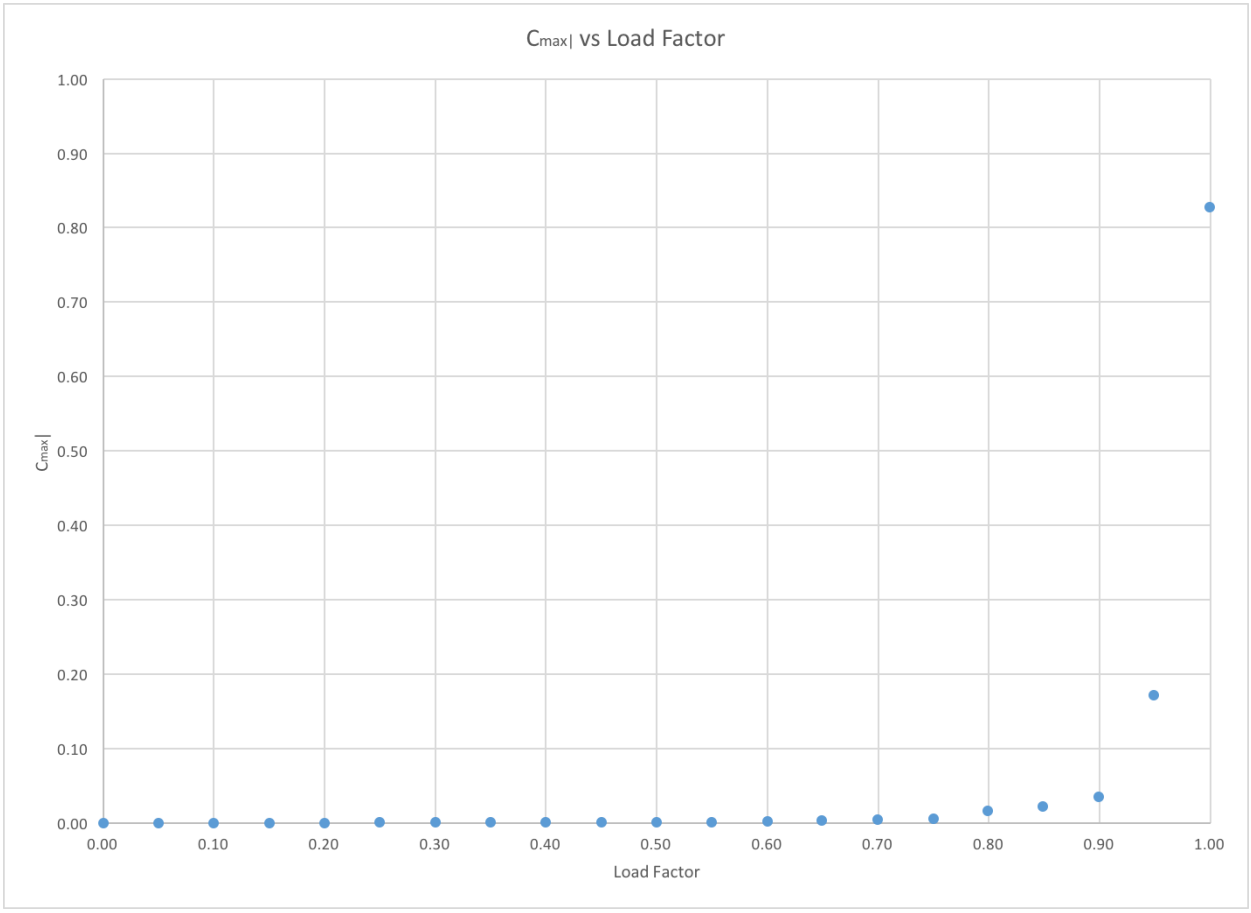
$$table\ size = \frac{elements\ in\ table}{\alpha}$$

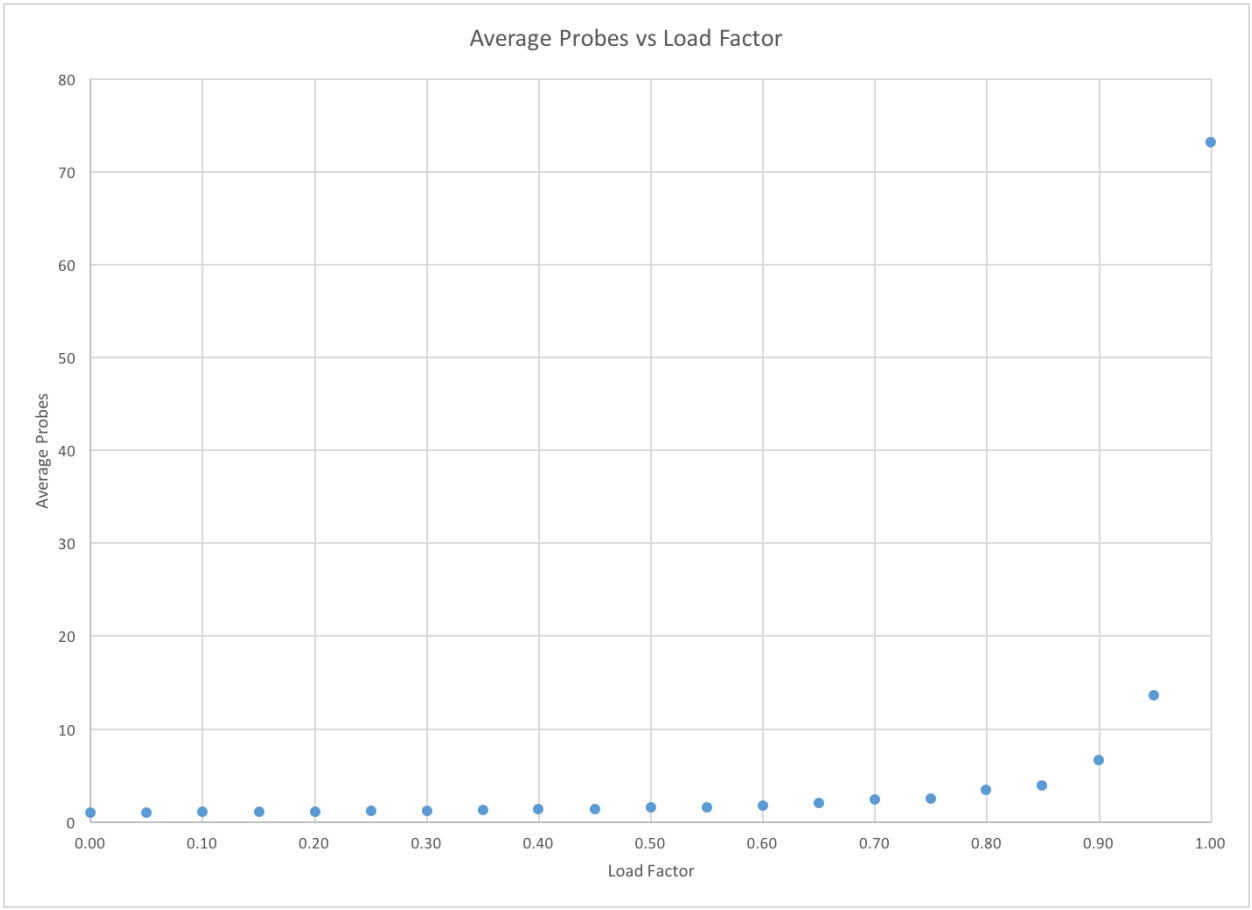
The *elements in table* can be found by counting every word in the input file.

```
While (fileScanner.hasNext()) {
        fileScanner.next();
        wordCount++;
}
```

After getting this minimum table size, all we need to do is find a prime number equal to or larger than this number. A number is **not prime** if it is even or if it is divisible by another number that isn't 1 or itself.

```
if (number % 2 == 0) {
    // number is not prime
}
for(int i = 3; i * i <= number; i += 2) {
    if (number % i == 0) {
        // number is not prime
    }
}
```


Cmax| vs Load Factor
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Cmax|
0.00
0.10
0.20
0.30
0.40
0.50
0.60
0.70
0.80
0.90
1.00
Load Factor

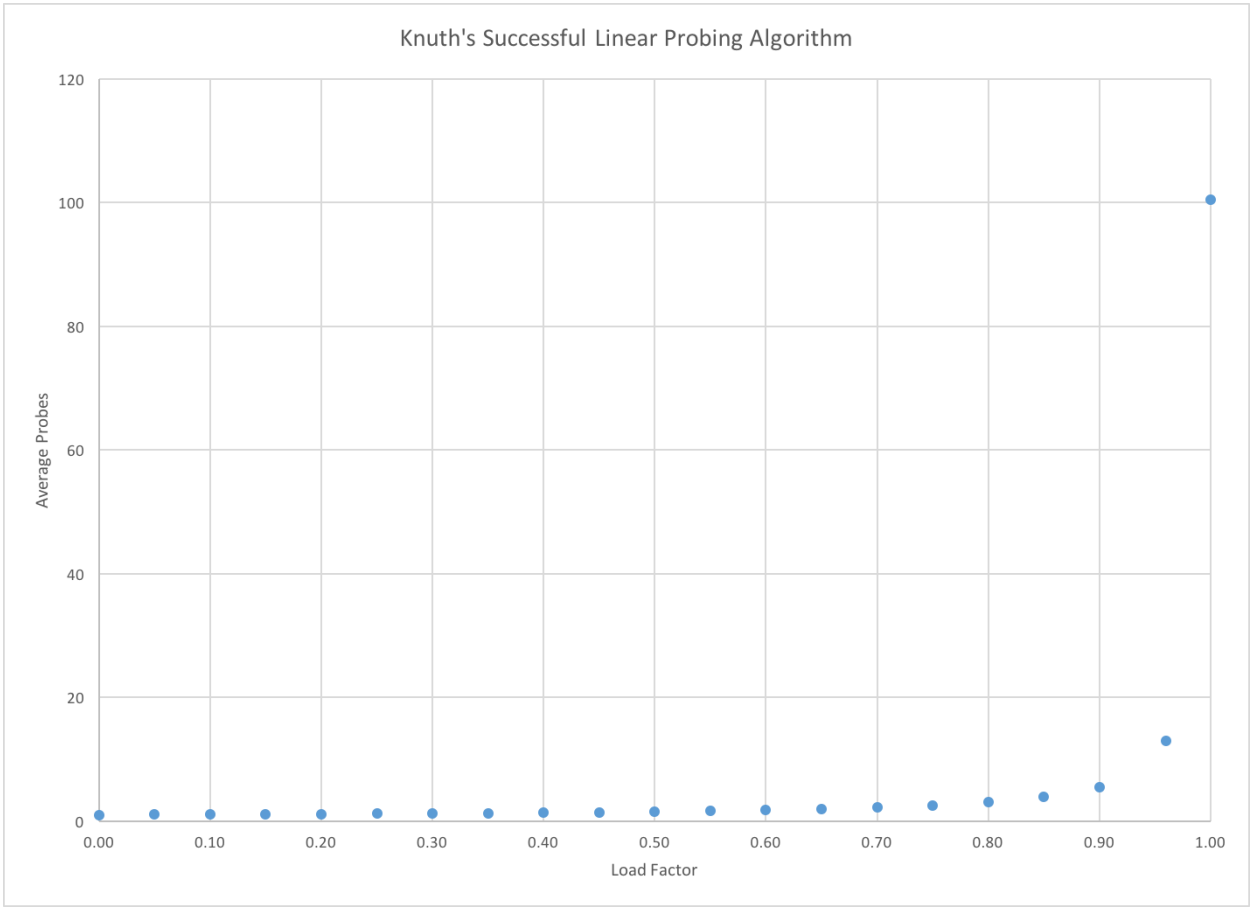

Average Probes vs Load Factor
80
70
60
50
40
30
20
10
0
Average Probes
0.00
0.10
0.20
0.30
0.40
0.50
0.60
0.70
0.80
0.90
1.00
Load Factor

Data used for plots

Average Probes	Cmax|	Load Factor
1.00017630465	0.00000001763	0.00010000000
1.02450634697	0.00001322163	0.04999537243
1.05650564175	0.00003525689	0.09998854151
1.08762341326	0.00007933465	0.14999537215
1.12552891396	0.00015864342	0.19996122050
1.17031029619	0.00024241356	0.24999449060
1.20839210155	0.00026433348	0.29985990325
1.28005994358	0.00052448092	0.34998303150
1.35269746121	0.00063409307	0.39961954416
1.42727433004	0.00091201079	0.44981958048
1.53314527504	0.00132211009	0.49993389450
1.61142454161	0.00135744413	0.54995879187
1.80518335684	0.00222092962	0.59986251388
2.08947461213	0.00343504895	0.64945325471
2.37367771509	0.00443978541	0.69951285688
2.53816995769	0.00568369572	0.74971911969
3.42727433004	0.01655279284	0.79904205114
3.97408321580	0.02177003067	0.84865714072
6.70548307475	0.03481087939	0.89953215447
13.60525387870	0.17142140049	0.94905044759
73.13628349788	0.82768038058	0.99938331425

We can see that there is a striking similarity between these and Knuth's formula,

$$Average\ Probes = \frac{1}{2}\left(1 + \frac{1}{1-\alpha}\right)$$



Based on this data, the load factor should be kept under **0.6** if the average number of probes is to be kept under 2.

$C_{max}|$ represents the maximum size of cluster as a percent of the total hash table size. As the load factor approaches 1.0, $C_{max}|$ approaches 1.0. In the plot, $C_{max}|$ only around 83% due to the fact that the table size was larger than the word count. The table size is larger because it needed to be increased in order to be prime.