

Question 1- Hash Table Readme and analysis

Objective:

To study the run times and average number of probes for different collision resolution methods for hashing and compare them.

Observations:

Table for Separate chaining: note alpha is for 2000 elements

Alpha	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Run time (in seconds)	0.014566	0.11968	0.15968	0.012514	0.013150	0.013736	0.015189	0.016136	0.016979	0.019527
Average number of probes = (total probes for given alpha/no. of queries)	0.005	0.0175	0.0433	0.207500	0.115000	0.2	0.263571	0.292500	0.272778	0.417500

Table for Linear Probing: note alpha is for 2000 elements

Alpha	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Run time (in seconds)	0.003955	0.003853	0.004036	0.004284	0.004284	0.004294	0.004702	0.004111	0.004259	0.004734
Average number of probes = (total probes for given alpha/no. of queries)	0.075	0.2	0.316667	0.507500	0.674	0.945833	1.336429	2.418125	3.605556	33.425999

Table for Quadratic Probing: note alpha is for 2000 elements and rehashing was done for alpha greater than 0.5

Alpha	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Run time (in seconds)	0.007024	0.009763	0.004126	0.004014	0.001210	0.001385	0.001602	0.000525	0.001818	0.000570
Average number of probes = (total probes for given alpha/no. of queries)	0.1	0.24	0.291667	0.51	0.276	0.36	0.37	0.4775	0.535556	0.616

The highlighted areas are for when the table was rehashed and garbage array was removed

Note: to be able to successfully rehash to a table size of 4000 from 2000 it was required to remove the garbage array with 10^6 elements. The maximum allowable rehash size with the garbage array was ~2500 elements. Malloc failed to allocate a table size greater than 2500 and hence we had to remove the garbage array. Notice the run times for the highlighted part of the quadratic probing table. The run times are significantly smaller. Thus we obtain a valuable comparison of run times with and without the added stress load of the garbage arrays.

Another important observation is that the run time for linear and quadratic probing is always lesser than the run time for separate chaining for the corresponding value of alpha. However even for alpha = 0.1 the number of probes for separate chaining is very low as compared to linear probing and quadratic probing.

Number of probes:

Separate-chaining < linear-probing < quadratic-probing

However, when the table is rehashed, the number of probes for quadratic probing significantly reduces.

1. Another interesting observation is the run time for alpha = 0.1 and alpha = 0.2 in the table for Quadratic probing. The run time seems to be quite large as compared for larger values of alpha. This may be attributed to the distribution of even numbers. As the numbers were randomly generated it is possible that it took a longer time to input them into the table. But this is in contrast to the average number of probes which seems to be as expected in comparison to the average number of probes for larger values of alpha.
2. Hence, I ran the program again for alpha = 0.1 and 0.2 without changing the test case but this time I found the total number of probes which was in line with the other observations. So instead, I found the individual number of probes for each element inserted which revealed that there was a large spike for certain elements and no. of probes were zero for most of the other elements. This made the average same as the expected average value but increased the run time significantly for the elements which required a greater number of probes.

Take a look at the table for Linear probing for alpha = 1. The difference between average number of probes for alpha = 0.9 and alpha = 1 is almost 10 times. This means that for alpha = 1, on average each element needed to be moved approximately 33 spaces before finding a suitable space for it to be inserted. In one test case the value of average number of probes was found to be 44 for alpha = 1 but it was omitted as the average for alpha = 1 was around 33.

Conclusions:

On average linear probing appears to be faster than both separate chaining and quadratic probing (in last 5 cases garbage array had to be removed). However, number of probes for quadratic probing (after rehashing) were way lower than the number of probes for linear probing. The number of probes for separate chaining were lower than linear probing and even quadratic probing after re-hashing.

Number of probes:

- Separate-chaining < Linear-probing < Quadratic-Probing (without rehashing)
- Separate-chaining < Quadratic-Probing (with rehashing) < Linear-probing

Run time:

- Linear-Probing < Quadratic-Probing (without rehashing) < Separate-chaining
- Results for quadratic probing after rehashing unavailable as garbage array had to be removed.