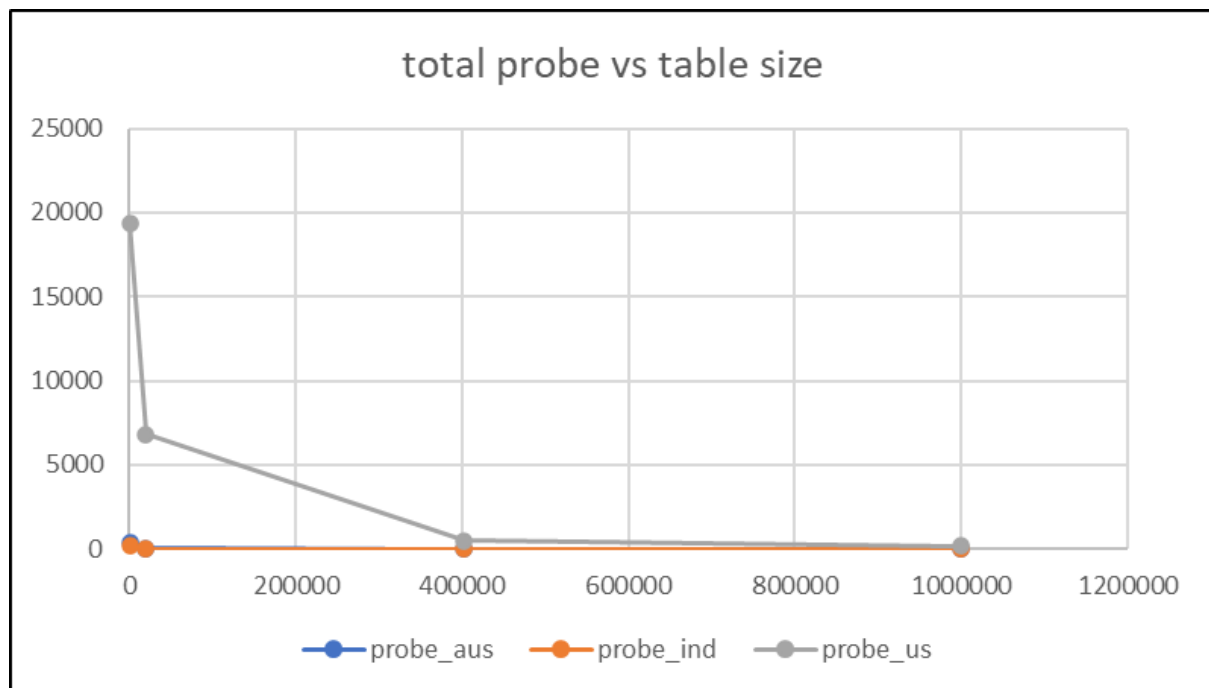


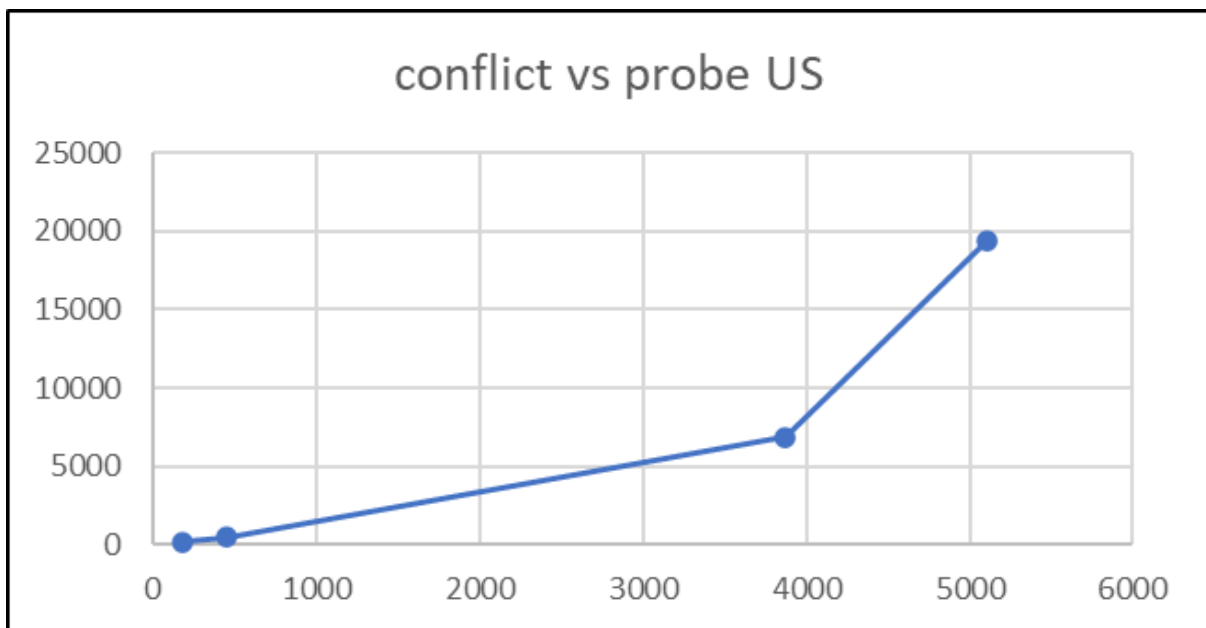
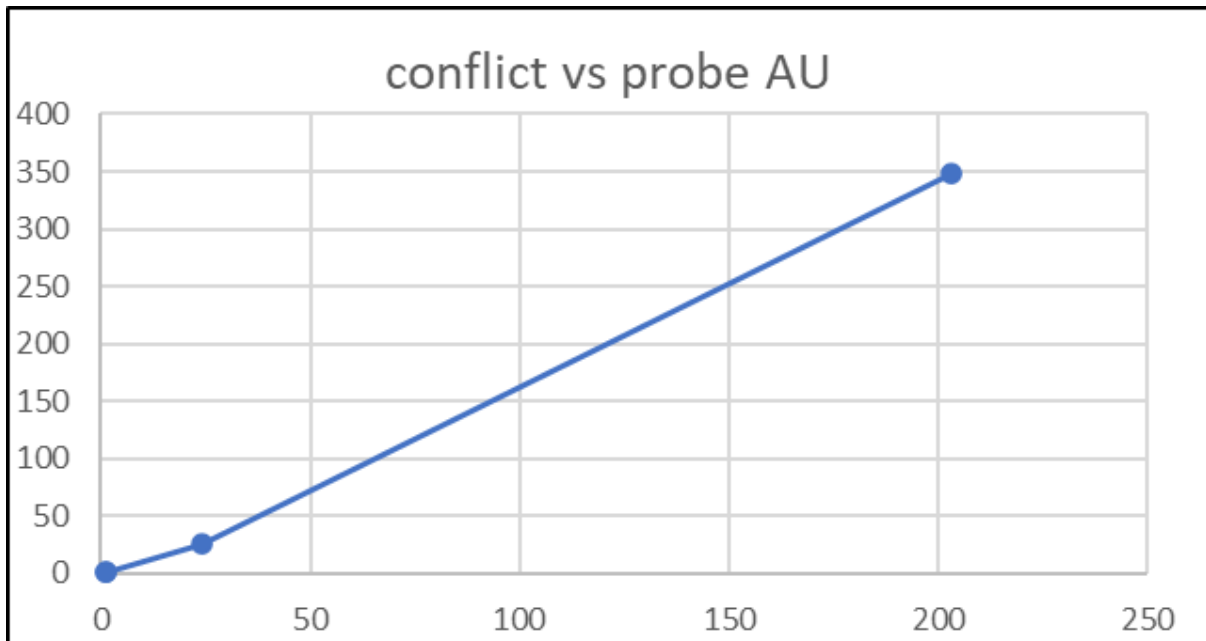
Group: i like oreos, FIT1008/2085 S2 2022

Analysis of hash table statistics

A good hash function would first make a large prime number which then will be used for our hash table size. For our hash table, it is expected as we have more data to work with, there is a higher chance that conflicts and probing happens, this is also related to the table size. As the input size increases, there is a higher chance that a key will be hashed into the same position which will cause conflict then triggers probing to resolve this issue. The higher the gap between the expected table size and the table size itself, the less that conflicts and probing could happen since the function `_linear_probing` has more room to work on.



Here, we can see that total probing done and number of conflicts relates to the table size, as the table size grows, the number of probing and the number of conflicts decreases showing that our statement before is correct.



As expected, here as the table size increases, the number of conflicts and probing decreases. This also means that conflict and probing are linearly related since when the number of conflicts goes up, so does the number of probing. The number of rehashing, which can be clearly seen in the excel sheet, happens when the gap between the table size and the number of input is small which can be seen by the US with 20021 as the table size and 19085 as the number of inputs. This shows that the hash table needs more space to work with the data.

For a more clear table that shows the conflict and the probe vs table size data for AU and IND, it is provided below. The complete data set of our analysis can be seen in our excel file in git or our submission.

