**Table Join Operation Project Report**

It is common in applications of relational databases to compute the join of two tables, which is a generalization of the set intersection operation you will be computing. There are a variety of scenarios for which efficient join processing techniques may be pursued, but the scenario here is very specific:

1. 1)  there are two unordered random tables, one with m values and the other with n values, each without duplicate values,
2. 2)  all table values are integers in the range 0 . . . k – 1,
3. 3)  m ≤ n ≤ k, and
4. 4)  you are to compute the set intersection.

The four available processing techniques are:

1. Sort both tables and then use a merge-like intersection, i.e. use a modified version of the

method merge from the Mergesort algorithm to find the common values. That doesn’t necessarily mean that you have to use Merge sort to do the sorting but whichever sorting algorithm, with average running time O(nlogn), suits your needs.

1. Sort only the larger table, and then use binary search on the resulting table for each value in the smaller table. Here only the larger table should be sorted. The smaller table should be accessed sequentially (linearly).
2. Sort only the smaller table and then use binary search on the resulting table for each value in the larger table. Here only the smaller table should be sorted. The larger table should be accessed sequentially (linearly).
3. Search the larger table for each value in the smaller table. Here none of the tables should be sorted. Both tables should be accessed sequentially.

Note: the values m, n, and k should be given as input parameters to the program. Subsequently your program should generate the two random tables, ensuring that there are no duplicates. If the given value of k is small, there will be a lot of numbers in common between the two tables. The time for generating the two tables is not of interests and does not need to be measured.

Your task is to compare the time required by each of the four set intersection methods and provide general principles for choosing which method should be used in a particular situation. That means that you should suggest which processing techniques would be more appropriate

based on various properties of tables that you will identify as important e.g. size ratio, expected number of common values, etc. Use graph plots to explain and justify your results.