Name: _____

CSE 4502/5717 Big Data Analytics

Exam I; October 6, 2022

Note: You are supposed to give proofs to the time and processor bounds of your algorithms. Read the questions carefully before attempting to solve them.

1. (17 points) Input are arrays $A_1, A_2, \ldots, A_{\sqrt{n}}$ each of size $\sqrt{n}$. It is known that all the elements in these arrays are distinct except for the elements $k_1, k_2, \ldots, k_{\sqrt{n}}$. These $\sqrt{n}$ elements are themselves distinct. Also, k_i has n_i copies in A_i, for $1 \leq i \leq \sqrt{n}$. (The element k_i does not appear in any array other than A_i, for $1 \leq i \leq \sqrt{n}$). It is also given that $\sum_{i=1}^{\sqrt{n}} n_i = n^{0.8}$. The problem is to identify one of these repeated elements. Present a Las Vegas algorithm to solve this problem in $\widetilde{O}(n^{0.9} \log n)$ time.

2. (15 points) What will be the worst case run time of quicksort if we use the BFPRT algorithm to find the median of the input elements and use it as the pivot?

3. (17 points) Input is a sequence $X = k_1, k_2, \ldots, k_n$. It is known that each k_i is an integer in the range$[1, n^{2/3}]$ (for $1 \leq i \leq n$). The problem is to find the maximum element in X. Show how to solve this problem in $O(1)$ time using n common CRCW PRAM processors.

4. (17 points) Input is a sequence $X = k_1, k_2, \ldots, k_n$ of real numbers and an integer m, $1 \leq m \leq n$. Let $A_i = \sum_{j=0}^{m-1} k_{i+j}$, for $1 \leq i \leq (n-m+1)$. The problem is to find the minimum of $A_1, A_2, \ldots, A_{n-m+1}$. Present an algorithm to solve this problem that uses $\frac{n}{\log n}$ CREW PRAM processors and runs in $O(\log n)$ time. (**Hint:** Use prefix computations).

5. (17 points) Input is a sequence $X_1, X_2, \ldots, X_q$ such that each X_i is a sorted sequence (for $1 \leq i \leq q$), and $\sum_{i=1}^{q} |X_i| = n$. X_i's are in a disk. The problem is to merge these q sorted sequences to produce one sorted sequence and write it in the disk. Show how to do this in $O\left(\frac{n}{B}\frac{\log q}{\log(M/B)}\right)$ I/O operations.

6. (17 points) Input is a (not necessarily sorted) sequence $X = k_1, k_2, \ldots, k_n$ residing in a disk. Assume that these n elements are distinct. The problem is to partition X into four equal sized parts X_1, X_2, X_3, and X_4, such that any element in X_i is less than any element in X_{i+1}, for $i = 1, 2, 3$. X_1, X_2, X_3, and X_4 should be written in the disk. Show how this partition can be done in $O\left(\frac{n}{B}\right)$ I/O operations.