Name: _____

CSE 4502/5717 Big Data Analytics

Exam I; October 3, 2019

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. (16 points) Input is an array $A[1:n]$ of real numbers. There are 5 elements in this array that have $\frac{n}{10}$ copies each. The other elements occur exactly once each. The problem is to output any one of these 5 elements. Present an $\widetilde{O}(\log n)$ time Las Vegas algorithm to solve this problem.

2. (17 points) Input is an array A of size n. The array can only be of one of the following three types:

   - **Type I**: A has all zeros.
   - **Type II**: A has all ones.
   - **Type III**: A has $\frac{n}{4}$ ones and $\frac{3}{4}n$ zeros.

   How will you identify the type of A in $O(\log n)$ time using a Monte Carlo algorithm? Show that the probability of a correct answer will be $\geq 1 - n^{-\alpha}$ for any constant $\alpha \geq 1$.

3. (16 points) Input is a sequence X of n real numbers. The problem is to check if X is in sorted order. Present a common CRCW PRAM algorithm to solve this problem in $O(1)$ time. You can use up to n processors.

4. (17 points) Input are two **sorted** sequences X and Y of real numbers with $|X| = |Y| = n$. The problem is to merge them. Show how to do this in $O(\log n)$ time using n CREW PRAM processors.

5. (17 points) Input are two sequences X and Y of real numbers in a disk. Each sequence is of length n. The problem is to check if they have a common element. Show how to solve this problem in $O\left(\frac{n}{B}\frac{\log(n/M)}{\log(M/B)}\right)$ I/O operations. Here M is the main memory size and B is the block size.

6. (17 points) Input is a sequence X of n distinct real numbers residing in a disk. The problem is to output all the elements of X whose ranks in X are in the interval $\left[\frac{n}{2} \pm \frac{M}{2}\right]$. Present an algorithm to solve this problem that makes $O\left(\frac{n}{B}\right)$ I/O operations. Here M is the main memory size and B is the block size. (The rank of an element k in a sequence X of elements is defined as $|\{q \in X : q < k\}| + 1$.)