CSE 4502/5717 Big Data Analytics

Homework 1, due on September 24, 2019 at 2 PM

1. Input is a sorted array $a[1:n]$ of arbitrary real numbers. The array could only be of one of the following two types: 1) **Type I:** All the elements in the array are distinct; or 2) **Type II:** The array has $\sqrt{n}$ copies of one element, the other elements being distinct. Present a Monte Carlo algorithm that determines the type of the array in $O(\sqrt{n}\log n)$ time. Show that the output of your algorithm will be correct with high probability.(**Fact:** $(1-x)^{1/x} \leq 1/e$ for any $1 > x > 0$.)

2. Show that the maximum of n given elements can be found in $O(1)$ time using $n^{1+\epsilon}$ common CRCW PRAM processors, where ϵ is any constant > 0.

3. Present an $O(\sqrt{n})$ time algorithm for the selection problem. You can use up to $\sqrt{n}$ CREW PRAM processors.

4. What happens to the I/O complexity of the sorting algorithm we discussed in class if we choose k to be $\frac{cM}{B}$ for some integer $c > 1$?

5. Present an efficient implementation of Djikstra's algorithm for the single source shortest paths problem on an out-of-core computing model with a single disk. What is the I/O complexity of your implementation?