

External Memory Algorithms and Data Structures

Problem 2

Course Webpage: <http://www.ti.inf.ethz.ch/ew/courses/EMADS04/>

Topic: Sorting

Due Date: November 4, 2004 at the lecture

Exercise 2.1

Design an external-memory variant of the radix sort, and estimate the number of IOs it makes.

Exercise 2.2

Estimate the number of comparisons in the internal memory made by the external-memory multiway merge sort of N numbers. Do you get an upper bound of $O(N \log N)$?

Exercise 2.3

Given N elements in an unsorted array, we want to determine whether these elements are all distinct or not. In the comparison-based model, give a lower bound of $\text{sort}(N)$ for this problem on the number of IOs. Here $\text{sort}(N)$ represents the number of IOs which we need to sort N numbers.

Exercise 2.4

Consider the following scenario. We are given N distinct binary strings of length ℓ in the external memory. We can only access the strings by a stream. Namely, the strings come to us one by one. The external disk is a read-only tape, so we can just read strings from a stream; No manipulation is allowed. The stream can run many times and it is always the same. Assume that the size of the internal memory is small, namely at most $O(\ell)$ bits.¹

In such a scenario, our task is to output a binary string which is not stored in the external memory, if such a string exists at all; Otherwise output "No."²

Design an algorithm to solve this problem with as few runs of the stream as possible.

¹A distributed version wrote " $O(2^\ell)$ bits" but it was a mistake.

²If we could manipulate the strings on the external memory, we would solve the problem by sorting. But notice that the scenario forbids sorting.