

Exercise Sheet 4

Due date: Thursday, 8. December 2005

Problem 4-1 (15 points) Bloom filters can be used to estimate set differences. Suppose you have a set X and I have a set Y , both with n elements. For example, the set might represent our 100 favorite songs. We both create Bloom filters of our sets, using the same number of bits m and the same k hash functions. Determine the expected number of bits where our Bloom filters differ as a function of m, n, k , and $|X \cap Y|$. Explain how this could be used as a tool to find people with the same taste in music more easily than comparing lists of songs directly.

Problem 4-2 (20 points) Consider a uniform rooted tree of height h - every leaf is at distance h from the root. The root, as well as any internal node, has three children. Each leaf has a Boolean value associated with it. Each internal node returns the value returned by the majority of its children. The evaluation problem consists of determining the value of the root; at each step, an algorithm can choose one leaf whose value it wishes to read.

- (a) Show that for any deterministic algorithm, there is an instance (a set of Boolean values for the leaves) that forces it to read all $n = 3^h$ leaves.
- (b) Consider the recursive randomized algorithm that evaluates two sub-trees of the root chosen at random. If the values returned disagree, it proceeds to evaluate the third sub-tree. Show that the expected number of leaves read by this algorithm (on any instance) is at most $n^{0.9}$.

Problem 4-3 (20 points) Use Yao's Minimax Principle to prove a lower bound on the expected running time of any Las Vegas algorithm for sorting n numbers.