CSE 6512 Randomization in Computing

Fall 2023, Homework 1, Due on September 28, 2023

1. Let T be an ordered set and let S_1 be a random sample of T of size $n^2 2^{m/3}$. Sort S_1 and from the sorted list pick elements that are in positions $n^2, 2n^2, \ldots, (2^{m/3}-1)n^2$. Let this list be S_2. Keys in S_2 partition T in the obvious way. Let q be the maximum size of any of these parts. Show that

$$Pr[q > (1 + n^{-1/3})|T|/2^{m/3}] < 2^{-c_1 n}$$

for some constant $c_1 > 0$ and that

$$Pr[q < (1 - n^{-1/3})|T|/2^{m/3}] < 2^{-c_2 n}]$$

for some constant $c_2 > 0$. Assume that $n^4 2^{m/3} = o(|T|)$ and that $n^2 2^{m/3} = o(|T|^{2/3})$.

2. *[Problem 7.2 from MR95.]* Two rooted trees T_1 and T_2 are said to be *isomorphic* if there exists a one-to-one mapping f from the vertices of T_1 to those of T_2 satisfying the following condition: for each internal vertex v of T_1 with the children $v_1, \ldots, v_k$, the vertex $f(v)$ has as children exactly the vertices $f(v_1), f(v_2), \ldots, f(v_k)$. Observe that no ordering is assumed on the children of any internal vertex. Devise an efficient randomized algorithm for testing the isomorphism of rooted trees and analyze the performance.

3. *[Problem 7.10 from MR95].* Given a randomized algorithm for testing the existence of a perfect matching in a graph G, describe how you would actually construct such a matching. What is the run time of your algorithm if you use the testing algorithm described in class?

4. *[Problem 7.13 from MR95].* Consider the two-dimensional version of the pattern matching problem. The text is an $n \times n$ Boolean matrix X and the pattern is an $m \times m$ Boolean matrix Y. A pattern match occurs if Y appears as a (contiguous) sub-matrix of X. To apply the randomized algorithm described in class, we can convert Y into an m^2-bit vector using the row-major format. The possible occurrences of Y in X are the m^2-bit vectors $X(j)$ obtained by taking all $(n-m+1)^2$ sub-matrices of X in a row-major form. It is clear that the algorithm discussed in class can be used in this case. Analyze the error probability and run time in this case.

5. *[Problem 8.17 from MR95].* In defining a random leveling for a skip list, we sampled the elements from L_i with probability 1/2 to determine the next level L_{i+1}. Consider instead the skip list obtained by performing the sampling with probability p (at each level), where $0 < p < 1$. (a) Determine the expectation of the number of levels r, and prove a high probability bound on r; (b) Determine as precisely as you can the expected cost of each operation in this skip list; and (c) Discuss the relationship between the choice of p and the performance of the skip list in practice.

6. *[Exercise 8.13 from MR95].* Assume for simplicity that $n = s$. Show that for $m = 2^{\Omega(s)}$, there exist perfect hash families of size polynomial in m. (**Hint:** Use the probabilistic method.)

7. *[Problem 8.22 from MR95].* In this problem we consider a weakening of the notion of 2-universal families of hash functions. Let $g(x) = x \bmod n$. For each $a \in Z_p$, define the function $f_a(x) = ax \bmod p$, and $h_a(x) = g(f_a(x))$, and let $H = \{h_a | a \in Z_p, a \neq 0\}$. Show that H is nearly-2-universal in that, for all $x \neq y$, $\delta(x, y, H) \leq \frac{2|H|}{n}$.