Homework 3

CS6033 Design and Analysis of Algorithms I

Fall 2023

(Sec. B, Prof. Yi-Jen Chiang)

Due: Wed. 10/11 by 1pm
(submit online on NYU Brightspace; one submission per group)
Maximum Score: 105 points

Note: This assignment has 3 pages.

1. (20 points)

(a) Textbook Exercise 11.2-2 (page 281). Just write down your final result. **(5 points)**

(b) Consider inserting the keys 10, 22, 31, 4, 15, 28, 17, 88, 59 into a hash table of length $m = 11$ using open addressing with the auxiliary hash function $h'(k) = k$. Illustrate the result of inserting these keys using linear probing, using quadratic probing with $c_1 = 1$ and $c_2 = 3$, and using double hashing with $h_1(k) = k$ and $h_2(k) = 1 + (k \bmod (m-1))$.
For each method, just write down your final result. **(15 points — 5 points for each method)**

2. (13 points)

Consider the AVL-tree T as shown in Fig. 1 below, where the numbers are the keys stored.

(a) Copy the tree in your write-up, and for each node v label the height of the subtree rooted at v. (We define the height of a null node to be 0 and the height of a leaf to be 1). Verify that this is indeed a valid AVL-tree. **(3 points)**

(b) Suppose now we insert a key 19 to the tree T, and then insert another key 4 to T. For each insertion, show the resulting tree right after the insertion (i.e., before any re-balancing), and the tree after each structural change during re-balancing, as well as the final tree. **(3 + 3 = 6 points)**

(c) Ignoring the insertions in **part (b)**, suppose we delete the key 14 from the tree T as shown in Fig. 1. We use the policy that when an internal node with no null child is deleted, we replace the deleted key with its **successor** key in the tree whenever the successor key is available. Show the tree right after the deletion (i.e., before any re-balancing), and the tree after each structural change during re-balancing, as well as the final tree. **(4 points)**

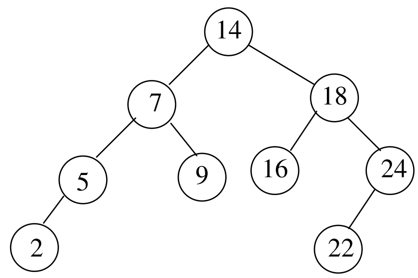

Figure 1: The AVL-tree for Question 2.

3. (20 points)
Show the results of inserting the keys

F, S, Q, K, C, L, H, T, V, W, M, R, N, P, A, B, X, Y, D

into an initially empty (2,4)-tree using the 2-pass insertion algorithm (as discussed in class and in the hand-out slides of (2,4)-trees), assuming that when a split operation occurs, the node in question is always split into a **left node** with **2 keys** and a **right node** with **1 key**. Draw only the configurations of the tree **right after some node has been split** (if inserting a key causes more than one split operation, draw the tree configuration right after **each** such split), and also draw the final configuration.
(**Note:** Here the keys are letters. We assume that there is a total order among the letter keys that is the alphabetical order of the letters, i.e., A $<$ B $<$ C $< \cdots <$ Z.) ($\mathbf{1 + 2 + 2 + 3 * 5 = 20}$ **points**)

4. (15 points)
Show the results of deleting keys 17, 16, 18, in that order, from the (2,4)-tree shown in Fig. 2 below (where the numbers shown are the keys stored), using the **2-pass deletion algorithm** (as discussed in class and in the hand-out slides of (2,4)-trees). Note that when deleting a key k from an internal node, we use the policy of replacing k with its **predecessor** key in the tree whenever the predecessor key is available. Also, for both the *transfer* and the *fusion* operations, if the immediate left and right siblings are both available for the operation, we use the policy of always using the **left** sibling. For each deletion, show the tree after each structural change and the final tree, and state the name of the re-structuring operation (transfer or fusion) if such operation(s) occur.
(**Note: 5 points for each deletion.**)

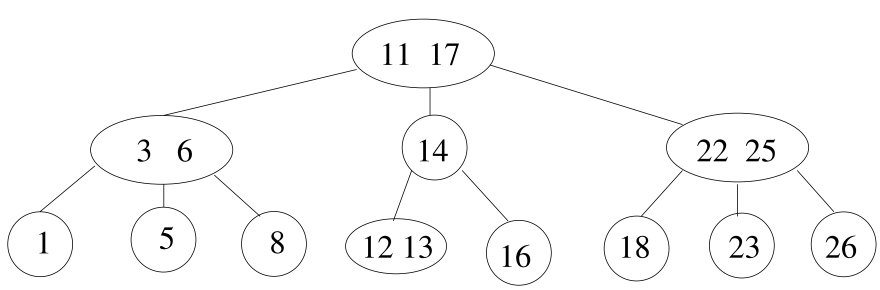

Figure 2: The (2,4)-tree for Question 4.

5. (25 points)
Textbook Problem 11-1 (page 308, **Longest-Probe Bound for Hashing**).
Do every part of the question.
Hints:
1. You may find the following **Boole's inequality** (also called the **union bound**, as discussed and proved in class) useful: Let A_i be an event for $i = 1, 2, \cdots, n$. Then $\Pr\{A_1 \cup A_2 \cup \cdots \cup A_n\} \le \sum_{i=1}^{n} \Pr\{A_i\}$.
2. For **part (d)**, observe that

$$E[X] = \sum_{x} x \Pr\{X = x\} = \sum_{x \le t} x \Pr\{X = x\} + \sum_{x > t} x \Pr\{X = x\},$$

where $\sum_{x \le t} x \Pr\{X = x\}$ is related to $\Pr\{X \le t\}$ and $\sum_{x > t} x \Pr\{X = x\}$ is related to $\Pr\{X > t\}$, for any reasonable value of $t > 0$. Use the result of **part (c)** to choose a suitable value for t, and use that result and this $E[X]$ formula to derive the $O()$ bound for $E[X]$. (For the second summation of this $E[X]$ formula, $\sum_{x > t} x \Pr\{X = x\}$, use a simple upper bound for the value of x.)
(**Notes: 5 points for each of (a), (b), (c), and 10 points for (d).**)

6. (12 points)
Given n distinct, unsorted integers where their value range is $>> n$, design and analyze an algorithm to report all pairs of integers (p, q) among them such that $p = 3q - 2$. Your algorithm should run in $O(n)$ **expected** time and $O(n)$ worst-case space.