Section 05: Open-Addressed Hash Tables and Heaps

1. Hash Tables – More Collision Resolution

Suppose you have a hash table of size 10, with initial hash function $h(x) = 2x$. Insert the following keys into the hash table:

$$2, 12, 3, 10, 5, 7$$

Do not resize the hash table while doing these insertions.

(a) Have your hash table use linear probing to resolve collisions.

(b) Have your hash table use quadratic probing to resolve collisions.

(c) Have your hash table use double hashing, with second hash function $f(x) = x + 1$ as the secondary hash function.

2. True or False

(a) An insertion can fail in a hash table using separate chaining when $\lambda \geq 1$.

(b) An insertion can fail in a hash table using linear probing when $\lambda = 3/4$.

(c) An insertion can fail in a hash table using quadratic probing when $\lambda = 3/4$.

3. Heaps – Basics

(a) Insert the following sequence of numbers into a *min heap*:

[10, 7, 15, 17, 12, 20, 6, 32]

(b) Now, insert the same values into a *max heap*.

(c) Now, insert the same values into a *min heap*, but use Floyd's `buildHeap` algorithm.

(d) Insert 1, 0, 1, 1, 0 into a *min heap*.

(e) Call `removeMin` three times on the min heap stored as the following array: $[1, 5, 10, 6, 7, 13, 12, 8, 15, 9]$

4. Food For Thought: Heaps

4.1. Running Times

Let's think about the best and worst case for inserting into heaps.

You have elements of priority $1, 2, \ldots, n$. You're going to insert the elements into a min heap one at a time (by calling `insert` not `buildHeap`) in an order that you can control.

(a) Give an insertion order where the total running time of all insertions is $\Theta(n)$. Briefly justify why the total time is $\Theta(n)$.

(b) Give an insertion order where the total running time of all insertions is $\Theta(n \log n)$.

4.2. Sorting and Reversing

(a) Suppose you have an array representation of a heap. Must the array be sorted?

(b) Suppose you have a sorted array (in increasing order). Must it be the array representation of a valid min-heap?

(c) You have an array representation of a min-heap. If you reverse the array, does it become an array representation of a max-heap?

(d) Describe the most efficient algorithm you can think of to convert the array representation of a min-heap into a max-heap. What is its running time?

5. Food For Thought: Heaps and Dictionaries

You just finished implementing your heap when your boss tells you they need you to add a new method.

```
/**
 * removes the element of priority k from the heap, and restores
 * the heap property
 * @param int k, the priority of the element to remove
 */
public void delete(int k)
```

(a) How efficient do you think you can make this method?

(b) Based on your answer to the previous part, should you use a heap to implement a dictionary?

6. Food For Thought: Recurrences

Suppose we have a min heap implemented as a tree, based on the following classes:

```
class HeapNode {
    HeapNode left;
    HeapNode right;
    int priority;

    // constructors and methods omitted.
}

class Heap {
    HeapNode root;
    int size;

    // constructors and methods omitted.
}
```

You just finished implementing your min heap and want to test it, so you write the following code to test whether the heap property is satisfied.

```
boolean verify(Heap h) {
    return verifyHelper(h.root);
}

boolean verifyHelper(HeapNode curr) {
    if (curr == null)
        return true;
    if (curr.left != null && curr.priority > curr.left.priority)
        return false;
    if (curr.right != null && curr.priority > curr.right.priority)
        return false;
    return verifyHelper(curr.left) && verifyHelper(curr.right);
}
```

In this problem, we will use a recurrence to analyze the worst-case running time of `verify`.

(a) Write a recurrence to describe the worst-case running time of the function above. **Hint:** our recurrences need an input integer, use the height of the subtree rooted at `curr`.

(b) Find an expression (using summations but no recursion) to describe the closed form. Leave the overall height of the tree h as a variable in your expression. You can use either unrolling or the tree method.

(c) Simplify to a closed form.

(d) If a complete tree has height h, how many nodes could it have? Use this to determine a formula for the height of a complete tree on n nodes.

(e) Use the formula from the last part to find the big-$\mathcal{O}$ of the `verify`.

7. Debugging

For this problem, we will consider a hypothetical hash table that uses linear probing and implements the `IDictionary` interface. Specifically, we will focus on analyzing and testing one potential implementation of the `remove` method.

(a) Come up with at least 4 different test cases to test this `remove(...)` method. For each test case, describe what the expected outcome is (assuming the method is implemented correctly).

Try and construct test cases that check if the `remove(...)` method is correctly using the key's hash code. (You may assume that you can construct custom key objects that let you customize the behavior of the `equals(...)` and `hashCode()` method.)

(b) Now, consider the following (buggy) implementation of the `remove(...)` method. List all the bugs you can find.

```java
public class LinearProbingDictionary<K, V> implements IDictionary<K, V> {
    // Field invariants:
    //
    // 1. Empty, unused slots are null
    // 2. Slots that are actually being used contain an instance of a Pair object

    private Pair<K, V>[] array;

    // ...snip...

    public V remove(K key) {
        int index = key.hashCode();

        while ((this.array[index] != null) && !this.array[index].key.equals(key)) {
            index = (index + 1) % this.array.length;
        }

        if (this.array[index] == null) {
            throw new NoSuchKeyException();
        }
        V returnValue = this.array[index].value;
        this.array[index] = null;
        return returnValue;
    }
}
```

(c) Briefly describe how you would fix these bug(s).