CS6033 Homework Assignment 3*

Due October 1st at 5:00 p.m.
Turn in this assignment as a PDF file on NYU classes
No late assignments accepted

1. Suppose you are given the following keys: $115, 2545, 995, 505$ and the following hash function $h(x) = x \mod 10$.

   Hash the keys using the hash function. How many keys collide?

   Choose a random hash function, h_R from $H_{10007,10}$. Include your random choice of h_R with your answer to this problem. Before you rehash the numbers with the new hash function, determine the probability that the keys 115 and 2545 collide when hashed with h_R? Hash each of the keys $115, 2545, 995, 5025$ with h_R. How many keys collided?

   If you had 1000 keys (all keys were positive integers less than 10000) inserted into a hash table of size 2000 using a hash function, h_R, randomly choosen from $H_{10007,2000}$ (the family of universal hash functions we defined in class). What is the expected number of collisions you would have if you inserted a new key, x, into the hash table?

2. We proved that in perfect hashing that the probability that we use more that $4n$ space for the secondary hash tables was at most $1/2$.

   What is the probability that you use at most $8n$ space for the secondary hash tables?

   What is the probability (in terms of k) that you use at most kn space (for some constant k)?

3. How many times must you attempt to construct one of a perfect hash table's sub-tables so you have one with probability greater than 99.9999%? This is a failure rate of 1 out of million.

4. Consider the following family of hash functions:

   $$H_m = \{h_a(x) = (ax \texttt{ mod } m) \texttt{ where } a \in \{1, ..., m-1\}\}$$

   .

   Is H_m universal? If so, prove H_m is universal. If not, provide a counter example to prove it is not universal. There is no restriction on the size of the keys.

5. News alert!

   Great hoards of space rats[1] have caused food shortages in the the eastern galaxy. As part of the relief effort, the federation is sending cheese to all the colonies in the eastern galaxy.

   The huge battleship INS Engorged Larder has been requisitioned to distribute the food. The cheese has been put into *missile* like containers, and the battleship will *shoot* it down to a hungry colony. One cheese "missile" per colony.

*Many of these questions came from outside sources.
[1]The effects of genetic engineering at the beginning of the 20th century.

Unfortunately the warehouse was attacked by the space rats, who before they were stopped, used a laser to cut the food containers into two pieces; those pieces now lay in a jumbled pile. Using duck tape[2], the containers can be put back together. Unfortunately you don't know which two pieces belong together.

You are given the weight of every broken piece of container in an array w of length n. Your job is to find which *two* pieces belong together. Two pieces go together if they weigh w.[3].

Find an $O(n)$ *average case* running time algorithm to determine which pieces belong together and save the Chancellors vacation - oops, I mean save the people from starvation! Justify your run time.

6. Unfortunately, the ungrateful colonies didn't respond well to their new cheese diets[4]. Piracy is on an all time high. New space recruits have been gathered from all over the galaxy to combat the pirates. The recruits are ranked from 1 to k.

   In $O(n+k)$ *worst case* time, develop an algorithm to sort the n fresh faced recruits based on their ranking where those ranked 1 occur before those ranked 2, etc. Justify your run time.

7. Political unrest has developed again. Some colonies say they are providing more recruits than others. Rumors are flying that *The Spiral Colonies* provide the most recruits. Others say *The Galactic Core* provide the most. To settle the rumors before the big academy party, you want to set the story straight.

   Given a list of recruits and which colony they came from, find the colonies that provided the most recruits. You should find the list of colonies sorted by which colony provided the most recruits in decreasing order. (i.e If $c = 10$, you find the colony providing the most recruits listed first, then colony providing the second most recruits listed second, etc. The last item in the list returned is the colony which provided the 10th most recruits.)

   You must find this list quickly or there will be no food at the party since everyone will be arguing instead of preparing. Your algorithm must run in $O(n + c \log c)$ *average case* running time where n is the number of recruits and c is the number of colonies. Justify your running time.

8. Peace has come again to the galaxy! The space pirates have been defeated![5]

   Celebrations are happening all over the galaxy. The best party will be at the Chancellor's house. The Chancellor has invited all star captains to the party. This is a very large list.

   To speed things up, the guest list will not be a list of names, instead it will be a hash table containing only zeros and ones. The guest list is created by the following steps. Initially, the hash table T is initialized to 0 for all locations. Then for each person on the guest list their name is hashed[6], let $i = h(\text{name})$, and that location in the hash table is set to one, $T[i] = 1$.

   When a guest arrives, their name will be hashed, let $i = h(\text{name})$; if the array T has a one in the ith position, they will be allowed in. If the array's ith position has a zero, they will be put in prison.

[2]Space vessels and missile are routinely patched up with duct tape to save money.

[3]This is really large number, much larger than the number of colonies needing food.

[4]Who knew that piracy could occur from the peace loving Tortuga.

[5]History will note that the victory came just as the federation started to provide crackers as well as cheese to the colonies. I am sure that is just a coincidence.

[6]For this problem assume the hash function has the simple uniform hashing assumption.

- What is the probability that a person on the guest list was put in prison?
- What is the probability that a person who wasn't on the guest list was allowed into the party (i.e. wasn't put in prison)?
- If instead of a single hash function, what if two hash functions h_1, h_2 were used. Could you improve the probability that only people on the guest list were invited into the party? The table size T will remain the same.

9. (3 bonus points) Think of a *good*[7] exam/homework question for the material covered in Lecture 3.

[7] Only well thought out questions will receive the bonus points.