

## Problem Set #4

**5/5 points (100%)**

Quiz, 5 questions

 **Congratulations! You passed!**[Next Item](#)1 / 1  
points

1.

Which of the following is *not* a property that you expect a well-designed hash function to have?



The hash function should be easy to compute (constant time or close to it).



The hash function should "spread out" every data set (across the buckets/slots of the hash table).

**Correct**

As discussed in lecture, unfortunately, there is no such hash function.



The hash function should be easy to store (constant space or close to it).



The hash function should "spread out" most (i.e., "non-pathological") data sets (across the buckets/slots of the hash table).

1 / 1  
points

2.

## Problem Set #4

Quiz, 5 questions

5/5 points (100%)

Suppose we use a hash function  $h$  to hash  $n$  distinct keys into an array  $T$  of length  $m$ . Assuming simple uniform hashing --- that is, with each key mapped independently and uniformly to a random bucket --- what is the expected number of keys that get mapped to the first bucket? More precisely, what is the expected cardinality of the set  $\{k : h(k) = 1\}$ .

- ☐  $1/m$
- ☐  $1/n$
- ☐  $m/(2n)$
- ☐  $m/n$
- ☒  $n/m$

**Correct**

Use linearity of expectation, with one indicator variable for each key. The probability that one key hashes to the first bucket is  $1/m$ , and by linearity of expectation the total expected number of keys that hash to the first bucket is just  $n/m$ .

- ☐  $n/(2m)$



1 / 1  
points

3.

Suppose we use a hash function  $h$  to hash  $n$  distinct keys into an array  $T$  of length  $m$ . Say that two distinct keys  $x, y$  *collide* under  $h$  if  $h(x) = h(y)$ . Assuming simple uniform hashing --- that is, with each key mapped independently and uniformly to a random bucket --- what is the probability that a given pair  $x, y$  of distinct keys collide?

- ☐  $1/n^2$
- ☒  $1/m$

**Correct**

For each of  $x$  and  $y$ , there are  $m$  choices for which bucket they map to. This means  $m^2$  choices for the pair of buckets  $(h(x), h(y))$ . Under uniform hashing, each of these choices is equally likely (i.e., occurs with probability  $1/m^2$ ). Since exactly  $m$  of these pairs yield a collision, the probability of a collision is  $m/m^2 = 1/m$ .

## Problem Set #4 $1/(m-1)$

5/5 points (100%)

Quiz, 5 questions

☐  $1/m^2$

☐  $1/n$



1 / 1  
points

4.

Suppose we use a hash function  $h$  to hash  $n$  distinct keys into an array  $T$  of length  $m$ . Assuming simple uniform hashing --- that is, with each key mapped independently and uniformly to a random bucket --- what is the expected number of pairs of distinct keys that collide? (As above, distinct keys  $x, y$  are said to collide if  $h(x) = h(y)$ .)

☒  $n(n-1)/2m$



**Correct**

There are  $n$  choose 2 (i.e.,  $n(n-1)/2$ ) pairs of distinct keys. By the previous problem, each pair has a  $1/m$  chance of colliding. The answer now follows from the linearity of expectation.

☐  $n/m^2$

☐  $n^2/m$

☐  $n(n-1)/m$

☐  $n/m$



1 / 1  
points

5.

## Problem Set #4

Quiz, 5 questions

To interpret our heuristic analysis of bloom filters in lecture, we considered the case where we were willing to use 8 bits of space per object in the bloom filter. Suppose we were willing to use twice as much space (16 bits per object). What can you say about the corresponding false positive rate, according to our heuristic analysis (assuming that the number  $k$  of hash tables is set optimally)? [Choose the strongest true statement.]

5/5 points (100%)

☐ Less than 1%☒ Less than .1%**Correct**☐ Less than .01%☐ Less than .001%