Problem Set #4
Quiz, 5 questions

1
point

1.

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

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

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

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

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

1
point

2.

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 $\{k : h(k) = 1\}$.

○ $1/n$

○ $1/m$

○ n/m

○ m/n

○ $n/(2m)$

○ $m/(2n)$

1

point

Problem Set #4

Quiz, 5 questions

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?

○ $1/n^2$

○ $1/n$

○ $1/m^2$

○ $1/(m-1)$

○ $1/m$

1

point

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^2/m

○ n/m^2

○ $n(n-1)/m$

○ n/m

○ $n(n-1)/2m$

1

point

5.

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.]

Problem Set #4
Quiz, 5 questions

○ Less than 1%

○ Less than .1%

○ Less than .01%

○ Less than .001%

Upgrade to submit