1. As per the hint, consider a random variable

Note that because for any two elements , we have two random variables and and we also have for any element . That makes for

.

Thus, is the same as .

Markov’s inequality is

We now try to get the expectation of the summation.

Note that

Let’s call . If we set the in Markov’s inequality to , Then

Because .

So

.

And

So

Which finally leads to

QED.

1. We first will determine how to get the number of collisions at a table entry given there are words stored there. A collision is defined as . If there is one word stored there, there are no collisions; if there are two stored, there is one collision; if there are words stored, there are collisions because there’s a collision between every pair of words. We now thus define a random variable that represents the total number of collisions in the hash table:

Simply:

Notice that the first term inside the parenthesis evaluates to the same thing as part a’s because the empty cells in our now larger hash table don’t matter to the summation – only the non-empty cells where the words reside matter. Taking the expectation of :

use part a’s work:

use the fact that , hence the sign:

Note that , so , so

The problem wanted us to show that there’s at least probability that there’s no collisions, which is logically equivalent to there’s less than probability that there is a collision. is the expected number of collisions. QED.

1. It is known from the previous parts that , and for all secondary level hash functions . So, it takes an expected 2 tries in getting a good first level hash function and an expected 2 tries in getting a good second level hash function for each cell that has collisions. Computing for all and computing for all is . There cannot be more than total entries in all the secondary hash tables added together. The time complexity in finding an overall good hash function is thus **.**