Yehui Huang

CS 112

Professor Snyder

Lab 08

Partner: Mingyu Zhu

1.

*Which hash functions performed better?*

In my lab, the best-performed one is Sequential LC. This hash function is better than other ones which all has similar standard deviations. The worst one is Silly.

*Did you need the "industrial-strength" hash functions near the end of the list?*

I don’t think we need the “industrial-strength” hash functions because they are not performing MUCH better than other function except “Silly”.

*Which was the simplest (i.e., earliest in the list) hash function to do reasonably well?*

The simplest one is Naïve Add because all hash functions generate similar result except Silly. And Naïve Add is the earliest one among all these hash functions after Silly.

2.

*What do you observe?*

(A) has a standard deviation of 14.395

(B) has a standard deviation of 46.638

(C) has a standard deviation of 3.576

When I used prime numbers, it distributed more evenly in the buckets. However, when I used 2 and 12, the elements only go to some certain buckets.

And the standard deviation of bucket length is much smaller for prime numbers than 2 and 12. In all cases, 12 has the largest standard deviation of bucket length.

*Why did this happen?*

Because for prime numbers, there always has reminders when doing hash, so that all the buckets have probability of being filled. But for 2 and 12, when they are divided by some numbers, such as 2, there is no remainders.

3.

*What do you observe? Why did this happen? (Just answer what? for part C.)*

(A) All elements go to only one bucket. This happens because the algorithm of Add Lin Cong hash function could only be affected by character itself, but not the character ordering.

(B) When change the LC Multiplier, the standard deviation would not change like what happen in the previous exercise, which means increasing the LC Multiplier would not result in a better performance. This happens because Add Lin Cong has function could only be influenced by character itself. Since all of them have same character, it doesn't matter by the larger prime numbers.

(C) The “industrial strength” hash function performs better on the distribution of elements than Add Lin Cong does.