Solution for Problem Set 6

Mianzhi Pan, 181240045

October 24, 2020

1 Problem 1

First randomly choose a bucket from the m buckets. Suppose the length of the chosen bucket is k, we use $RANDOM(1, L)$ and suppose the return value is a. If $a \leq k$, return the a^{th} element in that bucket. Otherwise, we continue this process until $a \leq k$. By this way, each element is chosen with probability $\frac{1}{mK}$.

The probability that we succeed in chosing an element in a particular bucket is $\frac{k}{L}$, so the expected chosing times are $\frac{L}{k}$. Together with a times for retriving the element, total time is $O(a + \frac{L}{k}) = O(L \cdot (a/L + 1/k))$, so the excepted time is $O(L \cdot (1 + 1/\alpha))$(excepted value of k is α and $a/L \leq 1$).

2 Problem 2

(a) Suppose string x of length $l + 1$: $x_l x_{l-1} \cdots x_0$, x has key value $x_l \times (m + 1)^l + \cdots + x_0 \times (m + 1)^0$. We have

$$\begin{aligned} h(x) &= x \ mod \ m \\ &= ((x_l \times (m+1)^l) \ mod \ m + \cdots + (x_0 \times (m+1)^0) \ mod \ m) \ mod \ m \\ &= (x_l \ mod \ m + \cdots + x_0 \ mod \ m) \ mod \ m \end{aligned}$$

We can find the hash value of a string is determined by all its characters but is independent with the order of them. Hence, x and y hash to the same value.

(b) Linear probing, $h(k, i) = (h'(k) + i) \ mod \ 11$, let $h'(k) = k$

22
88
4
15
28
17
59
31
10

Quadratic probing: $h(k,i) = (k + i + 3i^2)\ mod\ 11$

22
88
17
4
28
59
15
31
10

Double hashing: $h(k,i) = h_1(k) + ih_2(k)\ mod\ 11$

22
59
17
4
15
28
88
31
10

3 Problem 3

Let $B_i = \{w|h(w) = i\}$, which respesents the i^{th} bucket in B. We have

$$\begin{aligned}
Pr[h(k) = h(l)] &= \frac{\sum_{i=1}^{|B|} C_{|B_i|}^2}{C_{|U|}^2} \\
&= \frac{\sum_{i=1}^{|B|} |B_i|(|B_i| - 1)}{|U|(|U| - 1)} \\
&= \frac{\sum_{i=1}^{|B|}(|B_i|^2 - |B_i|)}{|U|(|U| - 1)} \\
&= \frac{\sum_{i=1}^{|B|} |B_i|^2 - |U|}{|U|(|U| - 1)} \\
&= \frac{\sum_{i=1}^{|B|} |B_i|^2}{|U|(|U| - 1)} - \frac{1}{|U| - 1} \\
&= \frac{\frac{1}{|B|}(\sum_{i=1}^{|B|} |B_i|)^2}{|U|(|U| - 1)} - \frac{1}{|U| - 1} \\
&= \frac{|U|}{|B|(|U| - 1)} - \frac{1}{|U| - 1} \\
&\geq \frac{|U| - 1}{|U|}(\frac{|U|}{|B|(|U| - 1)} - \frac{1}{|U| - 1}) \\
&= \frac{1}{|B|} - \frac{1}{|U|}
\end{aligned}$$

Notice $Pr[h(k) = h(l)] \leq \epsilon$, hence $\epsilon \geq \frac{1}{|B|} - \frac{1}{|U|}$