91.673 Advanced Database Systems Homework3

Chang Liu
chang_liu@student.uml.edu

October 8, 2015

1 Problem 1

Exercise 4.3.1: For the situation of our running example (8 billion bits, 1 billion members of the set S), calculate the false-positive rate if we use three hash functions? What if we use four hash functions?

2 Problem 2

Exercise 4.4.1: Suppose our stream consists of the integers 3, 1, 4, 1, 5, 9, 2, 6, 5. Our hash functions will all be o f the form h(x)=ax+b mod 32 for some a and b. You should treat the result as a 5-bit binary integer. Determine the tail length for each stream element and the resulting estimate of the number of distinct elements if the hash function is:

(a) h(x)=2x+1 mod32.
(b) h(x)=3x+7 mod32.
(c) h(x) = 4x mod 32.

3 Problem 3

Exercise 4.6.1: Suppose the window is as shown in Fig. 4.2. Estimate the number of 1s the the last k positions, for k = (a) 5 (b) 15. In each case, how far off the correct value is your estimate?