

CS 4540, Fall 2014

Homework 3

due: Wednesday, September 24, 2014 (at the start of class).

This homework is worth 50 points. This should be done independently.

In this assignment you will implement a Bloom filter and then analyze the false positive rate that your implementation achieves versus the theoretical claim. You need to turn in your source code via TSquare. You can use any programming language, but you cannot use any libraries that implement Bloom filters. It is easy to code it up so just do it yourself.

Let  $N$  denote the size of the universe, so the items are coming from the set  $\{0, 1, \dots, N-1\}$  where  $N$  is a large number. We are adding  $m$  items into the subset  $S$  that our Bloom filter is maintaining. And our Bloom filter has a table of size  $n$ . Let  $c = n/m$ . Let  $k$  denote the number of hash functions used. Set  $k = c \ln 2$ .

For a particular choice of  $n$  and  $c$  calculate the false positive rate. Vary  $n$  and  $c$  and plot the results of the false positive rate compared to the theoretical claims.

Turn in a report with a figure (or a few figures) showing your results, an explanation of what exactly you did in your simulation, and a short conclusions section.