6331 - Algorithms, Autumn 2016, CSE, OSU
Homework 6
Instructor: Anastasios Sidiropoulos

Problem 1. Let U be a finite set, and let H be the collection of *all* hash functions

$$f : U \to \{0, \ldots, m-1\}.$$

(a) Prove that H is universal.

(b) Suppose that you design an algorithm that samples uniformly at random a hash function $h \in H$, stores h in memory, and then uses h to hash n keys into a hash table of size m. Assume that $|U| \gg n \gg m$. How much space does your algorithm use?

Hint: Your findings should convince you that H is not a good choice of a universal collection of hash functions.

Problem 2. Let $r, m > 0$ be integers. Let $U = \{0, \ldots, m \cdot r - 1\}$ be the set of all possible keys. For any $i \in U$, let $f_i : U \to \{0, \ldots, m-1\}$ be the function such that for any $k \in U$, we have

$$f_i(k) = (k + i) \bmod m.$$

Let H be the collection of all such hash functions, i.e. $H = \{f_0, f_1, \ldots, f_{m \cdot r - 1}\}$. Is H universal?

Problem 3. Suppose we use a hash function h to hash n distinct keys into an array of length m. Assuming simple uniform hashing, what is the expected number of collisions? That is, you have to compute the expected cardinality of the set

$$S = \{\{k, l\} : k \neq l \text{ and } h(k) = h(l)\}.$$

Hint: For any pair of distinct keys k, l, define the indicator random variable $X_{kl} = I\{h(k) = h(l)\}$. That is, the random variable X_{kl} takes the value 1 when $h(k) = h(l)$, and the value 0 when $h(k) \neq h(l)$. Argue that $|S| = \sum_{k \neq l} X_{kl}$. Apply linearity of expectation to derive an estimate on the expected size of S, i.e. $E[|S|]$.