Universität Siegen
Lehrstuhl Theoretische Informatik
Markus Lohrey

Exercise 6

Task 1
We generalize the definition on slide 121 in the following way: Let $\mathcal{H} \subseteq \{h \mid h : A \to B\}$ be a family of hash functions. We call $\mathcal{H}$ a *family of k-wise independent hash functions*, if for all $a_1, \dots, a_k \in A$ (pairwise different) and $b_1, \dots, b_k \in B$ we have

$$\text{Prob}[\bigwedge_{i=1}^{k} h(a_i) = b_i] = 1/|B|^k$$

for a randomly chosen $h \in \mathcal{H}$ (uniform distribution). Show that

$$\mathcal{H} = \{h_x : \mathbb{F}_p \to \mathbb{F}_p \mid h_x(a) = \sum_{i=0}^{k-1} x_i a^i, x = (x_0, \dots, x_{k-1}) \in \mathbb{F}_p^k\}$$

is such a k-wise independent family if $k \leq p$.

Task 2 (AMS algorithm)
Consider the stream $S = (101, 011, 010, 111, 011, 101, 000, 001)$ and the corresponding set A. Approximate the cardinality of A by using the hash functions $h_{x,y}(u) = xu + y$ over $\mathbb{F}_{2^3}$ with

1. $x = 101$ and $y = 001$,
2. $x = 100$ and $y = 101$.

Hint: You can use that + over the field $\mathbb{F}_{2^3}$ works like a bitwise XOR and $x \cdot u$ is given by the following table:

u	000	001	010	011	100	101	110	111
$100 \cdot u$	000	100	011	111	110	010	101	001
$101 \cdot u$	000	101	001	100	010	111	011	110

Task 3 (Average height of binary search trees)

(a) Write down all binary search trees (BSTs) with 4 nodes.

(b) Compute the average height of a BST with 4 nodes (uniform distribution).

(c) Compute the expected value $E[H_4]$ (slide 153).