University of Freiburg
Dept. of Computer Science
Prof. Dr. F. Kuhn

Algorithms and Data Structures
Winter Term 2021/2022
Sample Solution Exercise Sheet 5

Exercise 1: Bad Hash Functions

Let m be the size of a hash table and $M \gg m$ the largest possible key of the elements we want to store in the table. The following "hash functions" are poorly chosen. Explain for each function why it is not a suitable hash function.

(a) $h : x \mapsto \lfloor \frac{x}{m} \rfloor \mod m$

(b) $h : x \mapsto (2x + 1) \mod m$ (m even).

(c) $h : x \mapsto (x \mod m) + \lfloor \frac{m}{x+1} \rfloor$

(d) For each calculation of the hash value of x one chooses for $h(x)$ a uniform random number from $\{0, \dots, m-1\}$

(e) For a set of "good" hash functions $h_1, \dots, h_\ell$ with $\ell \in \Theta(\log m)$, we first compute $h_1(x)$, then $h_2(h_1(x))$ etc. until $h_\ell(h_{\ell-1}(\dots h_1(x))\dots)$. That is, the function is $h : k \mapsto h_\ell(h_{\ell-1}(\dots h_1(x))\dots)$

Sample Solution

(a) Values are not scattered. m subsequent values have the same hash value.

(b) Only half of the hash table is used. The cells $0, 2, 4, \dots, m-2$ stay empty.

(c) $h(m-1) = m$, but the table has only the positions $0, \dots, m-1$.

(d) The hash value of x can not be reproduced.

(e) The calculation of a single hash value needs $\Omega(\log m)$).

Exercise 2: (No) Families of Universal Hash Functions

Let $\mathcal{S} = \{0, \dots, M-1\}$ and $\mathcal{H}_1 := \{h : x \mapsto a \cdot x^2 \mod m \mid a \in \mathcal{S}\}$. Show that H_1 is not c-univeral for *constant* $c \geq 1$ (that is c is fixed and must not depend on m).

Sample Solution

(a) For an $x \in \mathcal{S}$ let $y = x + i \cdot m \in \mathcal{S}$ for some $i \in \mathbb{Z} \setminus \{0\}$. Such a y exists for any x if $M > 2m$. Let $h \in \mathcal{H}_1$. We obtain

$$\begin{aligned} h(y) &= a \cdot y^2 \mod m \\ &\equiv a \cdot (x + im)^2 \mod m \\ &\equiv a \cdot (x^2 + 2xim + (im)^2) \mod m \\ &\equiv ax^2 \mod m = h(x). \end{aligned}$$

It follows that $|\{h \in \mathcal{H}_1 \mid h(x) = h(y)\}| = |\mathcal{H}_1|$, so for $m > c$ we have

$$|\{h \in \mathcal{H}_1 \mid h(x) = h(y)\}| > \frac{c}{m}|\mathcal{H}_1| \ .$$

This means that for $m > c$, $\mathcal{H}_1$ is not c-universal.