University of Freiburg
Dept. of Computer Science
Prof. Dr. F. Kuhn
P. Bamberger
P. Schneider

Algorithms and Datastructures
Summer Term 2020
Sample Solution Exercise Sheet 5

Due: Wednesday, 17th of June, 4 pm.

Exercise 1: Bad Hash Functions *(10 Points)*

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) $h : x \mapsto \lfloor \frac{x}{m} \rfloor \bmod m$ *(1,5 Points)*

(b) $h : x \mapsto (2x+1) \bmod m$ (m even). *(1,5 Points)*

(c) $h : x \mapsto (x \bmod m) + \lfloor \frac{m}{x+1} \rfloor$ *(1,5 Points)*

(d) For each calculation of the hash value of x one chooses for $h(x)$ a uniform random number from $\{0, \ldots, m-1\}$ *(1,5 Points)*

(e) $h : x \mapsto \lfloor \frac{M}{x \cdot p \bmod M} \rfloor \bmod m$, where p is prime and $\frac{M}{2} < p < M$ *(2 Points)*

(f) For a set of "good" hash functions $h_1, \ldots, h_\ell$ with $\ell \in \Theta(\log m)$, we first compute $h_1(x)$, then $h_2(h_1(x))$ etc. until $h_\ell(h_{\ell-1}(\ldots h_1(x))\ldots)$. That is, the function is $h : k \mapsto h_\ell(h_{\ell-1}(\ldots h_1(x))\ldots)$ *(2 Points)*

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 $0, 2, 4, \ldots, m-2$ stay empty.

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

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

(e) First, consider the function $h' : x \mapsto \lfloor \frac{M}{x} \rfloor \bmod m$. h' maps all $x > M/2$ (i.e., half of the keys) to position 1, all x with $M/3 < x \leq M/2$ (i.e. 1/6 of the keys) to position 2 etc. So the table is filled asymmetrically. As the function $x \mapsto x \cdot p \bmod M$ is a bijection from $\{0, \ldots, M-1\}$ to $\{0, \ldots, M-1\}$, h has the same property of an asymmetrical filled table (but compared to h' we do not have that a long sequence of subsequent keys are mapped to the same position which would be another undesirable property, cf. part (a)).

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

Exercise 2: (No) Families of Universal Hash Functions *(10 Points)*

(a) 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). *(4 Points)*

(b) Let m be a *prime* and let $k = \lfloor \log_m M \rfloor$. We consider the keys $x \in \mathcal{S}$ in base m presentation, i.e., $x = \sum_{i=0}^{k} x_i m^i$. Consider the set of functions from the lecture (week 5, slide 15)

$$\mathcal{H}_2 := \Big\{ h : x \mapsto \sum_{i=0}^{\mathbf{k}} a_i x_i \mod m \mid a_i \in \{0, \dots, m-1\} \Big\}.$$

Show that $\mathcal{H}_2$ is 1-universal.[1] *(6 Points)*

Hint: Two keys $x \neq y$ have to differ at some digit $x_j \neq y_j$ in their base m presentation.

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

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

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.

(b) Let $x, y \in \mathcal{S}$ with $x \neq y$. Let $x_j \neq y_j$ be the position where x and y differ in their base m representation. Let $h \in \mathcal{H}_2$ such that $h(x) = h(y)$. We have

$$\begin{aligned} & h(x) = h(y) \\ \iff \quad & \sum_{i=0}^{k} a_i x_i \equiv \sum_{i=0}^{k} a_i y_i \mod m \\ \iff \quad & a_j \underbrace{(x_j - y_j)}_{\neq 0} \equiv \sum_{i \neq j} a_i (y_i - x_i) \mod m \\ \iff \quad & a_j \equiv (x_j - y_j)^{-1} \sum_{i \neq j} a_i (y_i - x_i) \mod m && (x_j - y_j)^{-1} \textit{ exists because } m \textit{ is prime} \end{aligned}$$

This means that for any values $a_0, \dots, a_{j-1}, a_{j+1}, \dots, a_k$ there is a *unique* a_j such that the function h defined by $a_0, \dots, a_k$ is in $\{h \in \mathcal{H}_2 \mid h(x) = h(y)\}$. So we have m^k possibilities to choose a function from $\{h \in \mathcal{H}_2 \mid h(x) = h(y)\}$. It follows

$$\frac{|\{h \in \mathcal{H}_2 \mid h(x) = h(y)\}|}{|\mathcal{H}_2|} = \frac{m^k}{m^{k+1}} = \frac{1}{m} \ .$$

[1] This exercise and the according lecture slide was changed. Originally it stated $\mathcal{H}_2 := \big\{ h : x \mapsto \sum_{i=0}^{\mathbf{k-1}} a_i x_i \mod m \mid a_i \in \{0, \dots, m-1\} \big\}$ and $k = \lceil \log_m M \rceil - 1$. We are sorry for the inconvenience.