Albert-Ludwigs-Universität
Institut für Informatik
Prof. Dr. F. Kuhn
M. Fuchs, G. Schmid

Algorithmen und Datenstrukturen
Sommersemester 2024
Musterlösung Übungsblatt 5

Abgabe: Dienstag, 28. Mai, 2024, 10:00 Uhr

Aufgabe 1: Ungeeignete Hashfunktionen *(10 Punkte)*

Sei m die Größe einer Hashtabelle und $M \gg m$ der größte mögliche Schlüssel der Datenelemente, die wir in dieser Hashtabelle speichern möchten. Die folgenden "Hashfunktionen" sind aus unterschiedlichen Gründen schlecht gewählt. Erläutern sie für jede Funktion warum das so ist.

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

(b) $h : x \mapsto (2x + 1) \bmod m$ für ein gerades m. *(1,5 Punkte)*

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

(d) Bei jeder Berechnung des Hashwerts von x wählt man für $h(x)$ eine gleichverteilt zufällige Zahl aus $\{0, \ldots, m-1\}$ *(1,5 Punkte)*

(e) $h : x \mapsto \lfloor \frac{M}{x \cdot p \bmod M} \rfloor \bmod m$, wobei p prim und $\frac{M}{2} < p < M$ *(2 Punkte)*

(f) Für eine Menge "guter" Hashfunktionen $h_1, \ldots, h_\ell$ mit $\ell \in \Theta(\log m)$ berechnen wir erst $h_1(x)$, dann $h_2(h_1(x))$ etc. bis $h_\ell(h_{\ell-1}(\ldots h_1(x)) \ldots)$.

Die Funktion lautet also $h : x \mapsto h_\ell(h_{\ell-1}(\ldots h_1(x)) \ldots)$. *(2 Punkte)*

Musterlösung

(a) Die Funktion streut nicht gut. Je m aufeinanderfolgende Werte haben den gleichen Hashwert.

(b) Nur die Hälfte der Tabelle wird genutzt, die Zellen $0, 2, 4, \ldots, m-2$ bleiben frei.

(c) $h(m-1) = m$, aber die Hashtabelle hat nur die Positionen $0, \ldots, m-1$.

(d) Der Wert für x lässt sich nicht reproduzieren.

(e) Für Werte x mit $x \cdot p \equiv 0 \mod M$ ist der Hashwert nicht definiert (Division durch 0). Doch auch wenn man diese Werte aus der Schlüsselmenge heraus nimmt, ist h keine gute Hashfunktion, da sie die Tabelle ungleichmäßig füllt. Dazu betrachte zunächst die Funktion $h' : x \mapsto \lfloor \frac{M}{x} \rfloor \bmod m$. h' bildet alle $x > M/2$ (also die Hälfte der Schlüssel) auf Position 1 ab, alle Schlüssel mit $M/3 < x \leq M/2$ (also 1/6 der Schlüssel) auf Position 2 etc. Die Tabelle wird also ungleichmäßig genutzt. Da die Funktion $x \mapsto x \cdot p \bmod M$ eine Bijektion von $\{0, \ldots, M-1\}$ nach $\{0, \ldots, M-1\}$ ist, füllt h die Tabelle ebenso ungleichmäßig (aber im Vergleich zu h' werden keine langen Sequenzen aufeinanderfolgender Schlüssel auf die gleiche Position abgebildet, was eine weitere unerwünschte Eigenschaft wäre, siehe Teil (a)).

(f) Die Berechnung eines Hashwerts benötigt $\Omega(\log m)$. Zudem müssten die Funktionen $h_2, \ldots, h_\ell$ eingeschränkt auf $\{0, \ldots, m-1\}$ bijektiv sein, da sonst die Tabelle nicht vollständig genutzt wird.

Aufgabe 2: (Keine) Familien Universeller Hashfunktionen *(10 Punkte)*

(a) Sei $\mathcal{S} = \{0, \ldots, M-1\}$. Wir definieren $\mathcal{H}_1 := \{h : x \mapsto a \cdot x^2 \bmod m \mid a \in \mathcal{S}\}$. Zeigen Sie, dass $\mathcal{H}_1$ *nicht* c-universell ist, für konst. $c \geq 1$ (d.h., c ist fest und unabhängig von m). *(4 Punkte)*

(b) Sei nun m *prim* und $k = \lfloor \log_m M \rfloor$. Wir betrachten Schlüssel $x \in \mathcal{S}$ in ihrer Darstellung zur Basis m, d.h., $x = \sum_{i=0}^{k} x_i m^i$. Betrachten Sie die in der Vorlesung (Woche 5, Folie 15) vorgestellte Funktionsmenge

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

Zeigen Sie, dass $\mathcal{H}_2$ 1-universell ist. *(6 Punkte)*

Hinweis: Zwei Schlüssel $x \neq y$ müssen sich in mindestens einer Stelle $x_j \neq y_j$ ihrer Basis m-Darstellung unterscheiden.

Musterlösung

(a) Für ein $x \in \mathcal{S}$ sei $y = x + i \cdot m \in \mathcal{S}$ für ein $i \in \mathbb{Z} \setminus \{0\}$. D.h. y unterscheidet sich nur um ein Vielfaches von m von x. Solch ein y gibt es zu x für $M > 2m$. Sei $h \in \mathcal{H}_1$. Dann ist

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

Es folgt, dass $|\{h \in \mathcal{H}_1 \mid h(x) = h(y)\}| = |\mathcal{H}_1|$. Für $m > c$ gilt also

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

Für $m > c$ ist $\mathcal{H}_1$ daher nicht c-universell.

(b) Seien $x, y \in \mathcal{S}$ beliebig mit $x \neq y$. Sei $x_j \neq y_j$ die erste Stelle der Basis m Darstellung in welcher sich x und y unterscheiden. Sei $h \in \mathcal{H}_2$. Angenommen die Schlüssel $h(x) = h(y)$ kollidieren.

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

Dies bedeutet, dass für eine Funktion aus $\{h \in \mathcal{H}_2 \mid h(x) = h(y)\}$ der Parameter a_j bereits eindeutig durch die Wahl von $a_0, \ldots, a_{j-1}, a_{j+1}, \ldots, a_k$ bestimmt ist. Man hat also m^k Möglichkeiten eine Funktion aus $\{h \in \mathcal{H}_2 \mid h(x) = h(y)\}$ zu wählen. Es folgt

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