8. Aufgabenblatt zur Vorlesung

Höhere Algorithmik

WiSe 2016/17

Wolfgang Mulzer, Boris Klemz

Abgabe bis zum 16. Dezember 2016, 12 Uhr, in den Briefkasten neben Raum 111

Aufgabe 1 Hashing mit Verkettung I *10 Punkte*

Angenommen, wir haben eine Hashtabelle mit n Plätzen, die eine Menge $S \subseteq K$ der Größe n speichert. Konflikte werden mit Hilfe von Verkettung gelöst. Wir nehmen an, dass sich die Hashfunktion wie eine zufällige Funktion verhält.

Für einen Platz i in der Hashtabelle, bezeichne mit Q_i die *Anzahl* der Schlüssel, die auf i gehasht werden. Unser Ziel ist es, $\mathbf{E}[\max_{i=0}^{n-1} Q_i]$ zu bestimmen, also den Erwartungswert für die maximale Anzahl an Elementen, die auf denselben Platz abgebildet werden.

Hinweis: Bei jedem Aufgabenteil, mit Ausnahme von (d), genügt eine Zeile als Antwort.

(a) Begründen Sie in einem Satz: Es gilt für $i = 0, \ldots, n-1$ und $r = 0, \ldots, n$,

$$\Pr[Q_i = r] = \left(\frac{1}{n}\right)^r \left(1 - \frac{1}{n}\right)^{n-r} \binom{n}{r}.$$

(b) Zeigen Sie,

$$\Pr\left[\max_{i=0}^{n-1} Q_i = r\right] \leq n \Pr[Q_0 = r].$$

(c) Aus der Stirling-Formel folgt die Abschätzung $\binom{n}{r} \leq (\frac{ne}{r})^r$. Benutzen Sie dies, um zu zeigen: $\Pr[Q_0 = r] \leq e^r / r^r$.

(d) Definiere $r_0 := c \log n / \log\log n$, für eine Konstante $c > 1$. Zeigen Sie, dass man c so wählen kann, dass $\Pr[Q_0 = r] < 1/n^3$ ist, für alle $r \geq r_0$.

Folgern Sie daraus:

$$\Pr\left[\max_{i=0}^{n-1} Q_i \geq r_0\right] \leq 1/n.$$

(e) Zeigen Sie

$$\mathbf{E}\left[\max_{i=0}^{n-1} Q_i\right] \leq r_0 \cdot \Pr\left[\max_{i=0}^{n-1} Q_i < r_0\right] + n \cdot \Pr\left[\max_{i=0}^{n-1} Q_i \geq r_0\right].$$

Folgern Sie daraus: $\mathbf{E}[\max_{i=0}^{n-1} Q_i] = O(\log n / \log\log n)$.

In der Vorlesung haben wir gezeigt, dass für alle i gilt: $\mathbf{E}[Q_i] = 1$. Ist das ein Widerspruch?

Aufgabe 2 Hashing mit Verkettung II *10 Punkte*

Auf einem Tisch stehen N Kisten. In diese Kisten werden nacheinander n Bälle geworfen, wobei jede Kiste unabhängig mit gleicher Wahrscheinlichkeit getroffen wird.

(a) Sei Y_i die Zufallsvariable, die den Wert 1 annimmt, falls Kiste i leer ist, und 0 sonst. Berechnen Sie die Wahrscheinlichkeit, daß Kiste i leer ist. Geben Sie auch den Erwartungswert $\mathbf{E}[Y_i]$ an.

(b) Sei X die Zufallsvariable, welche die Anzahl von leeren Kisten angibt. Berechnen Sie den Erwartungswert von X mit Hilfe der Erwartungswerte $\mathbf{E}[Y_i]$.

(c) Geben Sie eine möglichst gute Schranke $f(N)$ an, so dass gilt: wenn $n \geq f(N)$, dann ist die die Wahrscheinlichkeit, dass eine Kiste mindestens zwei Bälle enthält, größer als $1/2$. (*Hinweis*: Verwenden Sie die ungemein nützliche Abschätzung $1 + x \leq e^x$, welche für alle $x \in \mathbb{R}$ gilt.)

(d) Professor Pinocchio hat eine Idee, um Hashtabellen zu vereinfachen. Wenn wir die Zahl N der Plätze in der Hashtabelle im Verhältnis zur Anzahl der zu speichernden Einträge n groß genug wählen, sollte die Wahrscheinlichkeit, dass Kollisionen auftreten, verschwindend gering werden (unter der Annahme, dass sich die Hashfunktion wie eine zufällige Funktion verhält). Dann könnte man auf die Kollisionsbehandlung verzichten. In Anbetracht von (c), was halten Sie von dem Vorschlag? (Wenn Sie Teil (c) nicht gelöst haben, dann bearbeiten Sie diesen Teil unter der Annahme, dass $f(N) = N^{1/3}$ ist. Achtung: Das ist nicht die Lösung für (c).)

Aufgabe 3 Universelles Hashing und Authentifizierung *10 Punkte*

Sei H eine Menge von Hashfunktionen, wobei jedes $h \in H$ eine Funktion von der Schlüsselmenge K nach $\{0, \ldots, m-1\}$ ist. Wir sagen H ist *l-universell*, wenn für jede Folge von l paarweise verschiedenen Schlüsseln $(k_1, k_2, \ldots, k_l)$ gilt: wählen wir $h \in H$ zufällig gleichverteilt, so ist die Folge $(h(k_1), h(k_2), \ldots, h(k_l))$ zufällig gleichverteilt in $\{0, 1, \ldots, m-1\}^l$, der Menge aller l-Tupel über $\{0, 1, \ldots, m-1\}$

(a) Zeigen Sie: Wenn H 2-universell ist, so ist H universell (im Sinne der Vorlesung).

(b) Sei p eine Primzahl und sei K die Menge der n-Tupel über $\mathbb{Z}_p$. Sei $m := p$. Für jede Folge $a = (a_1, a_2, \ldots, a_n) \in \mathbb{Z}_p^n$ und $b \in \mathbb{Z}_p$, definiere eine Hashfunktion $h_{a,b} : K \to \mathbb{Z}_p$ als

$$h_{a,b}(x) = \Big(\sum_{j=1}^{n} a_j x_j + b\Big) \bmod p,$$

und sei $H := \{h_{a,b}\}$. Zeigen Sie: H ist 2-universell.

(c) Sei H wie in (b). Nehmen Sie an, dass sich Alice und Bob heimlich auf eine Hashfunktion $h \in H$ einigen. Später sendet Alice eine Nachricht $m \in K$ an Bob über das Internet. Alice unterschreibt die Nachricht, indem sie die Unterschrift

$t = h(m)$ mitsendet, und Bob überprüft, ob das empfangene Paar (m, t) die Bedingung $t = h(m)$ erfüllt. Nehmen Sie an, ein Gegner fängt die Nachricht unterwegs ab und ersetzt sie durch ein anderes Paar (m', t'). Argumentieren Sie, dass der Gegner Bob nur mit Wahrscheinlichkeit höchstens $1/p$ überlisten kann, egal wie viel Rechenkraft ihm zur Verfügung steht. Welche Annahmen müssen Sie machen?