Karlsruher Institut für Technologie

Stefan Walzer
ITI Prof. Sanders
Wintersemester 2023/2024

Übungsblatt 15 – Peeling

Randomisierte Algorithmik

Aufgabe 1 – Deterministisches Schälen in Linearzeit

(a) Der Algorithmus constructByPeeling ist nichtdeterministisch (in der Wahl von i in der While-Schleife). Zeige, dass dennoch für zwei mögliche Ausführungen von constructByPeeling stets die gleiche Menge von Schlüsseln unplatziert bleibt. (Insbesondere beeinflusst der Nichtdeterminismus nicht Erfolg/Misserfolg.)

(b) Verfeinere den Pseudocode so, dass ein deterministischer Algorithmus herauskommt, der erkennbar Laufzeit $O(n)$ hat. Nutze geeignete Hlifsdatenstrukturen deiner Wahl.

Aufgabe 2 – Peeling mit 2 Hashfunktionen

Angenommen wir verwenden den Schälalgorithmus constructByPeeling in einem Setting mit nur zwei Hashfunktionen h_1 und h_2. Wir gehen vereinfachend davon aus, dass $h_1(x) \neq h_2(x)$ für alle $x \in D$ gilt und unter dieser Einschränkung die Paare $(h_1(x), h_2(x))$ uniform zufällig und für verschiedene $x \in D$ unabhängig sind.[1] Zeige:

(a) Es werden alle Schlüssel platziert genau dann wenn folgender Graph kreisfrei ist:

$$G = ([m], \{\{h_0(x), h_1(x)\} \mid x \in S\})$$

Beachte: G ist als Multigraph zu verstehen.

(b) Sei $\alpha > 0$ eine Konstante und $n = \lfloor \alpha m \rfloor$. Zeige, dass es (für $m \to \infty$) mit Wahrscheinlichkeit $\Omega(1)$ einen Kreis gibt.
Hinweis: Es genügt nach Kreisen der Länge 2 (also Doppelkanten) zu fanden.
Siehe `https://en.wikipedia.org/wiki/Birthdayproblem#Arbitrarynumberofdays`.

Es gibt damit keinen Schälbarkeitsschwellwert, wie es ihn für $k \geq 3$ gibt, denn für kein $\alpha = \Theta(1)$ hat constructByPeeling mit hoher Wahrscheinlichkeit Erfolg.

[1] Um das sicherzustellen können wir zum Beispiel $h_2(x) := (h_1(x) + h_{\text{diff}}(x)) \bmod m$ definieren für ein $h_{\text{diff}} : D \to [m-1]$.

Aufgabe 3 – Der Schälalgorithmus bleibt nicht erst spät stecken[2]

Für eine Schlüsselmenge $S \subseteq D$ der Größe $n = |S|$ und Hashfunktionen $h_1, h_2, h_3 \sim \mathcal{U}([m]^D)$ betrachten wir den Cuckoo Graphen wie in der Vorlesung:

$$G = G_{S,h_1,h_2,h_3} = (S, [m], \{(x, h_i(x)) \mid x \in S, i \in [3]\})$$

Wir nehmen lediglich $\alpha = \frac{n}{m} \leq 1$ an. Sei $S' \subseteq S$ die Menge der Schlüssel, die vom Schälalgorithmus nicht entfernt werden können (es gilt $|S'| \in \{0, \ldots, n\}$). Wir wollen zeigen, dass $\Pr[|S'| \in \{1, \ldots, \delta m\}] = O(1/m)$ ist für eine Konstante $\delta > 0$ (die später gewählt wird).

Intution: Entweder gilt $S' = \emptyset$ oder $|S'|$ ist $\Omega(m)$; alles dazwischen ist unwahrscheinlich.

(a) Beobachte: $|S'| = 1$ ist nicht möglich.

(b) Zeige: $|N(S')| \leq \frac{3}{2}|S'|$. Hierbei ist $N(S')$ die Menge aller Nachbarn von S' in G.

(c) Zeige, dass es eine Konstante C gibt, sodass für $s \in \{2, \ldots, n\}$ folgendes gilt:

$$p_s := \Pr[\exists X \subseteq S, |X| = s : \exists Y \subseteq [m], |Y| = \lfloor \tfrac{3}{2}s \rfloor : N(X) \subseteq Y] \leq \left(C \cdot \frac{s}{m}\right)^{s/2}.$$

Hinweis: Einfach brutal Union-Bound über alle Möglichkeiten von X und Y verwenden. Nützlich ist außerdem die Abschätzung $\binom{n}{k} \leq (\frac{ne}{k})^k$ für Binomialkoeffizienten. Ignoriere die Gaussklammern, das machen alle so.

(d) Zeige (i) $p_2 + p_3 + p_4 + p_5 = O(1/m)$, (ii) $\sum_{s=6}^{\sqrt{m}} p_s = O(1/m)$, (iii) $\sum_{s=\sqrt{m}}^{m/(2C)} p_s = O(1/m)$.

(e) Wähle $\delta = 1/2C$ und zeige $\Pr[|S'| \in \{1, \ldots, \delta m\}] \leq \sum_{s=2}^{\delta m} p_s = O(1/m)$.

[2]Diese Aufgabe ist etwas aufwändig. Sie ist vor allem auf dem Blatt weil sonst der Beweis der Vorlesung unvollständig wäre.