Karlsruher Institut für Technologie

Stefan Walzer
ITI Prof. Sanders
Wintersemester 2023/2024

Übungsblatt 14 – Cuckoo Hashing

Randomisierte Algorithmik

Aufgabe 1 – Cuckoo Hashing & Erdős-Renyi Graphen

Beachte: Für diese Aufgabe ist n eine Anzahl Kanten und m eine Ansatz Knoten.

Das „Sudden Emergence"-Resultat von Erdős und Renyi (Folie 19, Random Graphs Kapitel) gilt auch, wenn man $G(m, \lambda/m)$ durch $G^{\text{UE}}(m, \frac{\lambda m}{2})$ ersetzt. Beschreibe eine Variante von Cuckoo Hashing, der bei n Schlüsseln und m Tabellenplätzen der Graph $G^{\text{UE}}(m, n)$ zugrunde liegt.

(i) Wir wollen Schlüssel bis zu einer Beladung von $\frac{n}{m} = \alpha < \frac{1}{2}$ einfügen. Folgere aus dem „Sudden Emergence" Ergebnis, dass dies mit hoher Wahrscheinlichkeit möglich ist.

(ii) Es ergibt sich ein praktischer Nachteil bei der Implementierung von insert. Welcher?

Lösung 1

(i) Wir verwenden eine Tabelle der Größe m und zwei voll zufällige Hashfunktionen $h_1, h_2 : D \to [m]$. Jeder Schlüssel $x \in S$ entspricht einer Kante $\{h_1(x), h_2(x)\}$ im Graphen, dessen Knotenmenge die Tabellenpositionen sind. Für $n = |S|$ Schlüssel hat der Graph exakt die Verteilung von $G^{\text{UE}}(m, n)$ (auch Mehrfachkanten und Schleifen sind hier möglich). Falls die Beladung $\frac{n}{m} = \alpha < \frac{1}{2}$ gilt, so ist der durchschnittliche Knotengrad $\lambda = \frac{2n}{m} < 1$.

Aus dem „Sudden Emergence" Ergebnis folgt aus $\lambda < 1$, dass $G^{\text{UE}}(m, n)$ mit hoher Wahrscheinlichkeit nur aus Bäumen und Pseudobäumen besteht, also aus Komponenten mit höchstens einem Kreis. Für diese ist es stets möglich die Kanten kollisionsfrei zu richten und entsprechend die Schlüssel kollisionsfrei zu platzieren.

(ii) Beim Verdrängen eines Schlüssels ist zunächst nicht klar, ob dieser gemäß seiner ersten oder zweiten Hashfunktion platziert wurde. Man muss also im Zweifelsfall beide ausprobieren um die Ausweichposition zu finden. Dies bringt einen Branch oder doppelte Arbeit beim Hashen mit sich. In der Praxis ist daher die „bipartite" Variante wie in der Vorlesung gängig.