

Analysis of the Linear Probing Variant of Cuckoo Hashing

Stephan Beyer, TU Ilmenau

Diplomarbeit

Friday 10. February 2012, 11:00 Uhr, Raum Zusebau 4005

Cuckoo Hashing ist ein Hashingverfahren mit zwei Hashfunktionen, das Zugriffsoperationen im schlechtesten Fall in konstanter Zeit erlaubt. Leider wird dabei nur knapp die halbe Hash-Tabelle ausgenutzt. Eine bekannte Strategie zur Auflösung von Kollisionen ist lineares Sondieren, womit die gesamte Tabelle ausgenutzt werden kann. Allerdings werden Zugriffe mit zunehmender Tabellenauslastung extrem langsam. Wir analysieren den Platzverbrauch einer Cuckoo-Hashing-Variante, bei der zusätzlich ein Sondierungsschritt pro Hashfunktion erlaubt ist. Im Vergleich zu der Cuckoo-Hashing-Variante, die $k \geq 3$ Hashfunktionen nutzt, eignet sich die Cuckoo-Hashing-Variante mit zusätzlichem Sondierungsschritt eher für Cache-Architekturen. Experimente haben bereits gezeigt, dass bei unserer Cuckoo-Hashing-Variante eine Platzauslastung von ungefähr 96.3% der Hash-Tabelle möglich ist. Wir beweisen (mehr oder weniger formal) eine obere und untere Schranke der Platzauslastung. Wir zeigen, dass unsere Cuckoo-Hashing-Variante mit hoher Wahrscheinlichkeit funktioniert, wenn die Platzauslastung höchstens 82.9% ist. Wir zeigen, dass sie mit hoher Wahrscheinlichkeit nicht funktioniert, wenn die Auslastung mindestens 98.1% beträgt.