Design and Analysis of Algorithms I

Data Structures

Performance Guarantees (Open Addressing)

Open Addressing

Recall : one object per slot, hash function produces a probe sequence for each possible key x.

Fact : difficult to analyze rigorousely.

Heuristic assumption : (for a quick & dirty idealized analysis only) all n! probe sequences equally.

Heuristic Analysis

Observation : under heuristic assumption, expected Insertion time is ~ $\frac{1}{1-\alpha}$, where α = load

Proof : A random probe finds an empty slot with probability $1-\alpha$

So : Insertion time ~ the number N of coin flips to get "heads", where Pr["heads"] = $1-\alpha$

Let N denote the number of coin flips need to get “heads”, with a coin whose probability of “heads” is $1 - \alpha$. What is E[N]?

- ○ $1/(1 - \alpha)$
- ○ $1/\alpha$
- ○ $1 - \alpha$
- ○ α

Heuristic Analysis

Observation : under heuristic assumption, expected Insertion time is ~ $\frac{1}{1-\alpha}$, where α = load

Proof : A random probe finds an empty slot with probability $1-\alpha$

So : Insertion time ~ the number N of coin flips to get "heads", where Pr["heads"] = $1-\alpha$

Note : $E[N] = 1 + \alpha \cdot E[N]$

1st coin flip Probability of tails Expected # of further coin flips needed

Solution : $E[N] = \frac{1}{1-\alpha}$

Tim Roughgarden

Linear Probing

Note : heuristic assumption completely false.

Assume instead : initial probes uniform at random independent for different keys. ("less false")

Theorem : [Knuth 1962] under above assumption, expected Insertion time is

$$= \frac{1}{(1-\alpha)^2}, \ where\ \alpha = \ load$$

The Allure of Algorithms

"I first formulated the following derivation in 1962… Ever since that day, the analysis of algorithms has in fact been one of the major themes in my life."

-D. E. Knuth, *The Art of Computer Programming, Volume 3.* (3rd ed., P. 536)