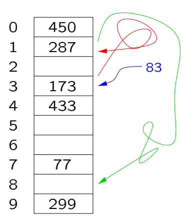
11—Hashing II

Double Hashing Analysis

CSE326 Spring 2002

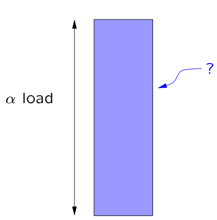
April 29, 2002

Double Hashing Analysis



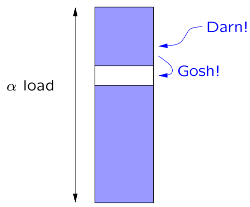
- Assume both hash functions look random
- Assume secondary probes are random
- Then *all* probes double hashing are *independent*: a probe doesn't depend on a previous probe
- *Much* easier to analyze

Analysis



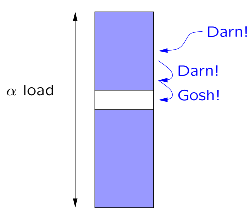
- What is probability of probing an *occupied* node?
- What is probability of probing an *unoccupied* node?

Analysis



- What is probability of an unsuccessful search. . .
  - ⋆ . . . first probing an occupied cell (but doesn't match key). . .
  - ⋆ . . . and then hitting an empty cell?

Analysis



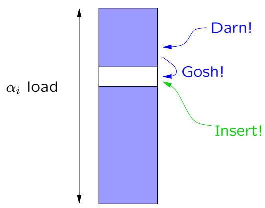
- What is probability of an unsuccessful search. . .
  - ⋆ . . . first probing an occupied cell (but doesn't match key). . .
  - ⋆ . . . then probing an occupied cell (but doesn't match key). . .
  - ⋆ . . . and then hitting an empty cell?

Analysis

Expected # probes for unsuccessful search is

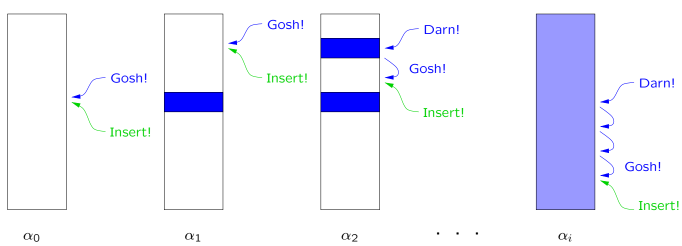
$$
\begin{aligned}
= \; & 1 \cdot \text{Pr[one probe to empty cell]} \\
& + 2 \cdot \text{Pr[1 probe to occupied and one probe to empty cell]} \\
& + 3 \cdot \text{Pr[2 probes to occupied and one probe to empty cell]} \\
& + 4 \cdot \text{Pr[3 probes to occupied and one probe to empty cell]} \\
& + \cdots \\
= \; &
\end{aligned}
$$

Analysis



To insert an item, we probe like an unsuccessful search for the item items, then insert it.

Analysis



After each insert, the load changes

Expected time for each insert is $U_{\alpha_i} = \frac{1}{1-\alpha_i} = \frac{1}{1-i/m} = \frac{m}{m-i}$

Analysis

- To search for an item, we do same number of probes as what took to insert it.
- Hence S_n is average over # probes to insert previous $i-1$ items.

The Scary Slide

$$
\begin{aligned}
S_n &= \frac{1}{n}\sum_{i=1}^{n} U_{i-1} \\
&= \frac{1}{n}\sum_{i=1}^{n} \frac{1}{1-\alpha_{i-1}} \\
&= \frac{1}{n}\sum_{i=1}^{n} \frac{m}{m-i+1} \\
&= \frac{m}{n}\sum_{i=1}^{n} \frac{1}{m-i+1} \\
&= \frac{m}{n}\left(\frac{1}{m} + \frac{1}{m-1} + \frac{1}{m-2} \cdots + \frac{1}{m-n+2} + \frac{1}{m-n+1}\right) \\
&= \frac{m}{n}\left(1 + \frac{1}{2} + \cdots + \frac{1}{m} - \left(1 + \frac{1}{2} + \cdots + \frac{1}{m-n}\right)\right) \\
&= \frac{m}{n}\left(H_m - H_{m-n}\right) = \frac{m}{n}(\ln m - \ln(m-n)) = \frac{1}{\alpha_n}\ln\frac{1}{1-\alpha_n}
\end{aligned}
$$

Summary

- $S_n \approx \frac{1}{\alpha_n}\ln\frac{1}{1-\alpha_n}$
- $U_n \approx \frac{1}{1-\alpha_n}$
- If table (of any size) is 90% full

  $S_n \approx \frac{1}{.9}\ln\frac{1}{1-.9} \approx 2.56$ *Not bad!*

  $U_n \approx \frac{1}{1-.9} \approx 10$ *Still constant*
- Compare with linear probing (analysis is harder):

$$
\begin{aligned}
S_n &\approx \frac{1}{2}\left(1 + \frac{1}{1-\alpha_n}\right) \\
U_n &\approx \frac{1}{2}\left(1 + \frac{1}{(1-\alpha_n)^2}\right)
\end{aligned}
$$

  At 90%, $S_n \approx 5.5$ and $U_n \approx 50.5$.