Lecture 9, Oct 20 2009

Analysis of Chaining

- Simplifying assumption:
  Assume each key equally likely hashed to any slot.
- n keys, m slots; $\alpha = \frac{n}{m} =$ "load factor"
- I_j – indicator random variable, $I_j=1$ if element j hits our chosen slot
- Expected length of a chain: $E[\sum I_j] = \sum E[I_j] = \sum_1^n \frac{1}{m} = \frac{n}{m} = \alpha$

  $\Rightarrow$ Access time $= O(1+\alpha)$
- Unsuccessful search:
  Expected length of a randomly chosen list + 1: $O(1+\alpha)$

Successful Search

- Expected time to find i-th element = time to insert i-th element
- Assume that the key being searched for is equally likely to be any one of the keys stored.
- Conditioned on "key was the i-th element inserted"
  - Conditional expectation = $\left(1+\frac{i-1}{m}\right)$
  - Expectation: $\sum_{i=1}^{n}\left(\Pr[\text{i-th element inserted}]E[\text{time conditioned on i}]\right)$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(1+\frac{i-1}{m}\right) = 1+\frac{1}{nm}\sum_{i=1}^{n}(i-1)$$

$$= 1+\frac{1}{nm}\frac{n(n-1)}{2} = 1+\frac{\alpha}{2}-\frac{1}{2nm} = O(1+\alpha)$$

- Intuition: need to search 1/2 of a list on the average.

Choosing Hash Functions

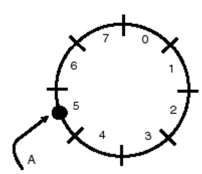
- Mostly black magic…
  division method: h(k)=k mod m
  - Do not use $m=2^p$ (will not use all the bits)
  - choose m=prime not too close to power of 2 or 10.
- Multiplication method: $h(k) = \lfloor m(kA - \lfloor kA \rfloor) \rfloor$
  - choose $m \neq 2^p$, $0 < A < 1$, not too close to 0 or 1.
  - If $m=2^p$, then all we do is scramble by multiplication, and choose p bits to the left of binary point.
  - Consider change from k-1 to k.

98

More multiplicative method

$h(k) = \lfloor m(kA - \lfloor kA \rfloor) \rfloor$

- Example: m=8:
  - each time k incremented:
    - go A around the circle,
    - Read off sector number.
  - Note what happens if A=.5 or $1/2^p$.



99