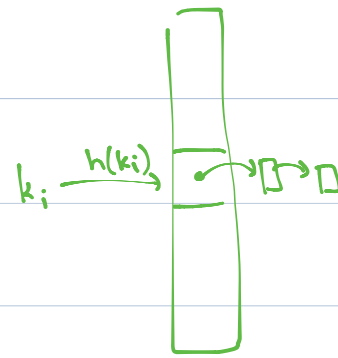
universe U table of size m hashing w/ chaining

Claim: with random hash function, expected time to perform n operations is $\frac{n^2}{m}$



Proof: Ops: $k_1, k_2, \ldots, k_n$

T_i : time to perform op i $= 1 + L_{h(k_i)}$

$$E(T_i) = E\left(1 + L_{h(k_i)}\right) = 1 + \underbrace{E\left(L_{h(k_i)}\right)}$$

$$L_{h(k_i)} = \sum_{j \neq i} 1_{h(k_j) = h(k_i)}$$

$$E\left(L_{h(k_i)}\right) = E\left[\sum_{j \neq i} 1_{h(k_j) = h(k_i)}\right]$$

$$= \sum_{j \neq i} \underbrace{E\left(1_{h(k_j) = h(k_i)}\right)}_{= Pr(h(k_i) = h(k_j))} \leq \frac{n}{m}$$

$$E(T) = E\left(\sum_{i=1}^{n} T_i\right) = \sum_{i=1}^{n} E(T_i) = n\left(1 + \frac{n}{m}\right)$$

load factor

If $n = cm$ $E(T) = O(n)$

Crucial fact: linearity of expectation

$$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)$$

$\mathcal{H}$ is $\boxed{\text{universal}}$ family of hash functions where $h : U \to [m]$ $\forall h \in \mathcal{H}$

$\forall x \neq y \in U$

$$\Pr(h(x) = h(y)) \leq \frac{1}{m}$$

where probability taken over random choice of hash fn.

$\mathcal{H}$ is $\boxed{\text{k-wise independent}}$ if $\forall$ $x_1, x_2, \ldots, x_k \in U$

$\forall$ $y_1, y_2, \ldots, y_k \in [m]$

$$\Pr(h(x_1) = y_1, h(x_2) = y_2, \ldots, h(x_k) = y_k) = \frac{1}{m^k}$$

$|U| = 2^n$ $\qquad$ $m = 2^k$

$$A = \begin{array}{cc} & n \text{ columns} \\ k \text{ rows} & \begin{pmatrix} \quad & \quad & \quad \\ & & \\ & & \end{pmatrix} \end{array}$$

$h(x) = Ax$ $\qquad$ All calculations mod 2

hash fn selected by choosing all entries in A at random

This is a universal class

Pf: Suppose $x \neq y$

$$\Pr(Ax = Ay) = \Pr(A(x-y) = 0)$$

$z = x - y \neq 0 \Rightarrow \exists i$ s.t. $z_i = 1$

$$\begin{array}{c} i^{th}\text{ column} \\ \begin{pmatrix} & a_1 & \\ & a_2 & \\ & & \\ & a_k & \end{pmatrix} \end{array} \begin{pmatrix} \\ \\ z_i \\ \\ \\ \end{pmatrix} = \begin{pmatrix} c_1 + a_1 z_i \\ c_2 + a_2 z_i \\ c_3 + a_3 z_i \\ c_k + a_k z_i \end{pmatrix}$$

$z_i = 1$

$$\Pr(Az = 0 \mid c_1, \dots, c_k) = \frac{1}{2^k} \Rightarrow \sum_{c_1,\dots,c_k} \Pr(Az=0 \mid c_1,\dots,c_k)\Pr(c_1,\dots,c_k) = \frac{1}{2^k}$$

law of total probability

Perfect Hashing

Given S, how can I construct a collision-free hash table?

n items, table of size m

How good is universal hash fn?

$$E(\#\text{collisions}) = \sum_{x \neq y} \Pr(h(x)=h(y)) = \binom{n}{2}\frac{1}{m} < \frac{n}{2} \quad \text{if } n=m$$

If I want $E(\#\text{ collisions}) < 1$

need to take $m = \Omega(n^2)$

$m=n^2 \Rightarrow \quad E(\#\text{collisions}) < \frac{1}{2} \quad \equiv E\left(\sum_j \binom{L_j}{2}\right) \leq \frac{n}{2}$

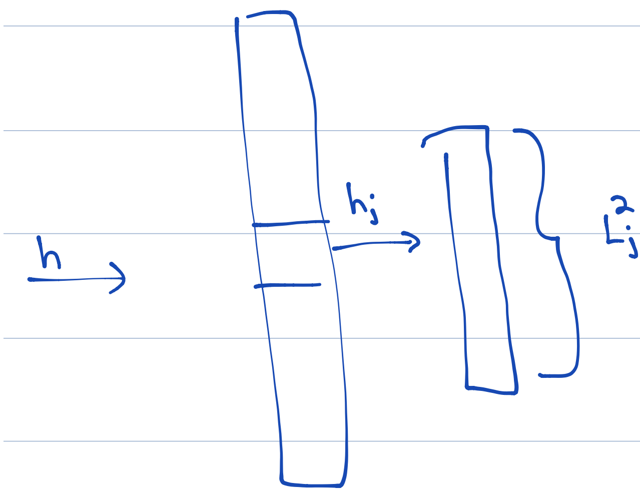
$\Rightarrow \Pr(\exists \text{ collision}) < \frac{1}{2}$

also $\Rightarrow \quad \Pr(\exists\, j \text{ s.t. } L_j \geq \sqrt{2n}) \leq \frac{1}{2}$

$$\sum_j \frac{E(L_j^2)}{2} - \sum_j \frac{E(L_j)}{2} \leq \frac{n}{2} \quad \Rightarrow \sum_j E(L_j^2) \leq 2n$$

Failure modes:

$\sum_j E(L_j^2) > 4n$ prob $\le \frac{1}{2}$



Collision in subtable
prob $\le \frac{1}{2}$

How long to successfully construct? $E(T) = O(n)$

How much space? $O(n)$

Lookup time once constructed 2