Universal Hashing

universal family $\qquad m =$ table size

Keys $k_1 \neq k_2$, then $h(k_1) = h(k_2)$ w. prob $\leq \frac{1}{m}$

$\mathbb{Z}_p \{0, 1, \dots, p-1\} \qquad \mathbb{Z}_p^* = \mathbb{Z}_p \setminus \{0\}$

where p larger than any key

$a \in \mathbb{Z}_p^*$, $b \in \mathbb{Z}_p$

$h_{ab}(k) = ((ak + b) \bmod p) \bmod m$

Th. 11.4 $\quad H_{pm}$ is universal

Proof $\quad k_1 \neq k_2$

$r_1 = (ak_1 + b) \bmod p$, $r_2 = (ak_2 + b) \bmod p$

Show $r_1 \neq r_2$

$r_1 - r_2 \equiv a(k_1 - k_2) \pmod p$

Assume $r_1 = r_2 \Rightarrow \underset{p}{\underset{\wedge}{a}}(\underset{p}{\underset{\wedge}{k_1 - k_2}}) = 0 \bmod p = \overset{\in \mathbb{Z}}{x} \cdot p \quad ↯$

Different a, b give different r_1, r_2

$k_1 - k_2$, p are co-prime

$a = (r_1 - r_2) \cdot ((k_1 - k_2)^{-1} \bmod p) \pmod p$

$b = r_1 - ak_1 \pmod p$

same Prob. of collision if I choose a, b at random

or r_1, r_2 at random

$r_1 \neq r_2$: $p(p-1)$

Fix r_1

Get $r_1 \equiv r_2 \pmod m$?

$$\leq \left\lceil \frac{p}{m} \right\rceil - 1 \leq \frac{p+m-1}{m} - 1$$

$$= \frac{p-1}{m}$$

r_2 equal (mod p) to r_1 w. prob.

$$\le \frac{\frac{p-1}{m}}{p-1} \le \frac{1}{m}$$

Perfect hashing

- Static data structure

Th 11.9

n keys, table size $m = n^2$, universal hash function h
prob of any collision is $\le \frac{1}{2}$

proof $\quad Pr\{X \ge t\} \le \frac{E[X]}{t}$

any pair collides with prob. $\frac{1}{m} = \frac{1}{n^2}$

$X = \#$ collisions

$$E[X] = \binom{n}{2} \cdot \frac{1}{n^2} = \frac{n \cdot (n-1)}{2} \cdot \frac{1}{n^2} \le \frac{1}{2}$$

markovs Ineq : $Pr\{X \ge 1\} \le \frac{E[X]}{1} \le \frac{1}{2}$

Th 11.10

n keys, $m = n$, univ. h,

$E\left[\sum_{j=0}^{m-1} n_j^2\right] < 2n$, n_j = # keys hashing to slot j.

$$E\left[\sum_{j=0}^{m-1} n_j^2\right] = E\left[\sum_{j=0}^{m-1}\left(n_j + 2\binom{n_j}{2}\right)\right]$$

$$= E\left[\sum_{j=0}^{n-1} n_j\right] + 2 \cdot E\left[\sum_{j=0}^{m-1}\binom{n_j}{2}\right]$$

$$= E[n] + 2\binom{n}{2} \cdot \frac{1}{m}$$

$$= n + \cancel{2}\,\frac{\cancel{n}(n-1)}{\cancel{2}} \cdot \frac{1}{\cancel{n}}$$

$$= 2n - 1$$

Collary 11.12

Prob. that $\sum_{j=0}^{m-1} n_j^2 \geq 4n$ is $\frac{1}{2}$

$$\Pr\left\{\sum_{j=0}^{m-1} n_j^2 \geq 4n\right\} \leq \frac{E\left[\sum_{j=0}^{m-1} n_j^2\right]}{4n}$$

$$< \frac{2n}{4n} = \frac{1}{2}$$

Construction

repeat

choose h from H_{pm}^{n} - hash all n keys

compute $s = \sum_{j=0}^{m-1} n_j^2$

until $s < 4n$

waiting times 2

for $j = 0$ to $m-1$

repeat

choose h_j from $H_{pm}^{n_j^2}$

expected waiting time 2

hash into table size $m_j = n_j^2$

until no collision

Expected construction time

$$2n + \left(\sum_{j=0}^{m-1} 2 \cdot n_j\right)$$

$$= 2n + 2n = 4n$$

Space: $n + 4n \in O(n)$

look up time: $O(1)$