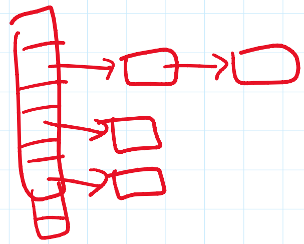
Store array $A[0..m-1]$

where $A[i]$ = list of all $x \in S$ with $h(x) = i$

("bucket")



space $O(m+n)$

query(y): do linear search in $A[h(y)]$

Ex: $h(x) = x \bmod m$

if input is rand, unif. distrib,

each bucket has $\sim \frac{n}{m}$ elems "on average"

choose $m \approx n \Rightarrow$ $O(1)$ "average" query time

$O(n)$ space

but don't want to assume input is random!

key idea [Carter, Wegman '77]

- randomize choice of hash fn h.

Ex Assume U is prime.

Pick rand $a \in [U] - \{0\}$.

$b \in [U]$

Define $h_{a,b} : [U] \to [m]$

$h_{a,b}(x) = ((a \cdot x + b) \bmod U) \bmod m$

Define $h_{a,b} : [U] \to [m]$

$$\boxed{h_{a,b}(x) = ((ax+b) \bmod U) \bmod m}$$

$O(1)$ time

Prop For any fixed $x, y \in [U]$, $x \neq y$,

$$\Pr_{a,b}[h_{a,b}(x) = h_{a,b}(y)] \leq O\left(\frac{1}{m}\right)$$

called universality

More strongly, for any fixed i, j, $x, y \in [U]$, $x \neq y$,

$$\Pr_{a,b}[h_{a,b}(x) = i \;\&\; h_{a,b}(y) = j] \leq O\left(\frac{1}{m^2}\right)$$

called 2-universality

(similar to pairwise indep)

Pf: $x, y \in [U]$, $x \neq y$

Fix $s, t \in [U]$.

rand. $a \in [U] - \{0\}$
$b \in [U]$

$$\Pr_{a,b}\left[\begin{array}{l}(ax+b) \bmod U = s \;\& \\ (ay+b) \bmod U = t\end{array}\right]$$

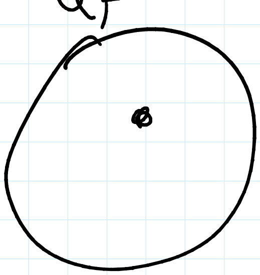
$3^{-1} \equiv 2 \pmod 5$

$$= \Pr_{a,b}\left[\begin{array}{ll} xa + 1 \cdot b \equiv s & \bmod U \\ \& \; ya + 1 \cdot b \equiv t & \end{array}\right]$$

2x2 linear sys of eqns!

$\left(a \equiv (s-t)(x-y)^{-1} \;\&\; b \equiv s - xa \right)$

$a\phi$



$$= \frac{1}{(U-1)U}$$

$$\Pr_{a,b}[\, h_{a,b}(x) = i \ \& \ h_{a,b}(y) = j \,]$$

$$= \sum_{\substack{s,t\,: \\ s \bmod m = i \\ \wedge\ t \bmod m = j}} \Pr\left[\begin{array}{l} (ax+b) \bmod U = s \ \& \\ (ay+b) \bmod U = t \end{array} \right]$$

$$\le O\left(\frac{U}{m} \cdot \frac{U}{m} \cdot \frac{1}{(U-1)U} \right) = O\left(\frac{1}{m^2}\right). \ \square$$

Expected query time:

$$= O\Big(E_{a,b}\big[\, \#x \in S - \{y\} \text{ with } \underbrace{h_{a,b}(x) = h_{a,b}(y)}_{\text{collision}} \,\big] + 1 \Big)$$

$$= O\left(n \cdot \frac{1}{m} + 1 \right)$$

Set $m \approx n \Rightarrow$ $\boxed{O(1)}$ expected query time

$\boxed{O(n)}$ space

$\boxed{O(1)}$ expected insert/delete time

Q: can we get worst case query time?

YES, in static case (Fredman, Komlós, Szemerédi '84)

Rmk: other alternative hash families

e.g. "tabulation hashing"

say $U = 2^w$, $m = 2^q$

say $U = 2^w$, $m = 2^q$

write $x = x_{\ell-1} \cdots x_0$ in base $2^{w/\ell}$

define $h(x) = T_{\ell-1}[x_{\ell-1}] \oplus T_{\ell-2}[x_{\ell-2}] \oplus \cdots \oplus T_0[x_0]$

& where T_i rand. table $T_i[x_i] \in [2^q]$

is 3-universal ...

Part I. perfect hashing

$$E\left[\#\text{colliding pairs } (x,y) \in S \times S,\ x \neq y \text{ with } h_{a,b}(x) = h_{a,b}(y)\right]$$

$$\leq O\left(n^2 \cdot \frac{1}{m}\right)$$

$$= O\left(\frac{1}{c}\right) \qquad \text{set } m = cn^2$$

(Recall Markov's ineq: if $X \geq 0$,
$$\Pr[X \geq t] \leq \frac{E[X]}{t}.)$$

$$\Rightarrow \Pr[\#\text{colliding pairs} \geq 1] \leq \frac{O\left(\frac{1}{c}\right)}{1}$$

Set c suff. large $\Rightarrow$ this prob. $< \frac{1}{2}$

(rerun until success)

$\Rightarrow$ $\boxed{O(1)}$ worst-case query time

$\boxed{O(n^2)}$ space

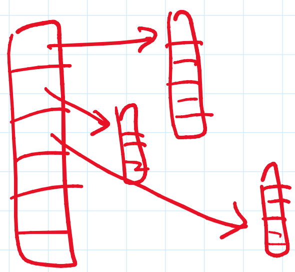
$\boxed{O(n)}$ expected preproc time ...

$\boxed{O(n^2)}$ space

$\boxed{O(n)}$ expected preproc time ...

Part II reduce space by bootstrapping!

Store each bucket $A[i]$ in a perfect hash table



2-level structure

query $\boxed{O(1)}$ $\boxed{\text{worst case}}$

$$\text{space} = O\left(m + \sum_{i=0}^{m-1} |A[i]|^2\right)$$

$$= O(m + \#\text{ colliding pairs})$$

$$\text{expected space} = O\left(m + n^2 \cdot \frac{1}{m}\right)$$

choose $m = n$ $\qquad = \boxed{O(n)}$ ← worst case by repeating until success

expected preproc time $\boxed{O(n)}$