UNIVERSAL HASHING $h_{ab}(x) = ((ax+b) \% p) \% m \in \mathcal{H}$

- $\forall_{x \neq y} \in U \subseteq \mathbb{Z}_p \quad \Pr_{h \in \mathcal{H}}(h(x) = h(y)) \leq \frac{1}{m}$ collisions are uniformly distributed

- 2-way independence

$$\forall_{x,y \in U} \; \forall_{z,v \in [m]} \quad \Pr_{h_1, h_2 \in \mathcal{H}}(h_1(x) = z \wedge h_2(y) = v) = \Pr_{h_1 \in \mathcal{H}}(h_1(x) = z) \times \Pr_{h_2 \in \mathcal{H}}(h_2(y) = v)$$

HANDS-ON

Dictionary Hash Tables:

Cuckoo Hashing - the power of choice of 2 hash functions

Dictionary problem:

- Set S
  - membership: $x \stackrel{?}{\in} S$ — $O(1)$
  - insertion: $S \leftarrow S \cup \{x\}$ — $O(1)$
  - deletion: $S \leftarrow S \setminus \{x\}$ — $O(1)$

worst-case

DoS attack on hashing inside routers

$h(x) = x \% p$

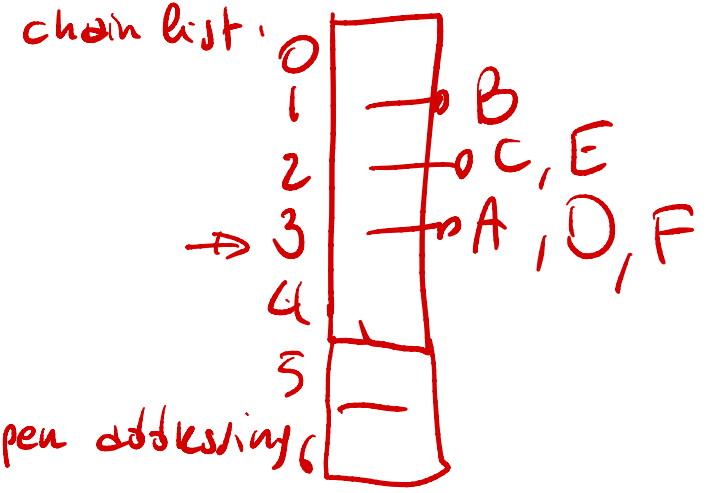
$O(1)$ ~~average~~ worst case

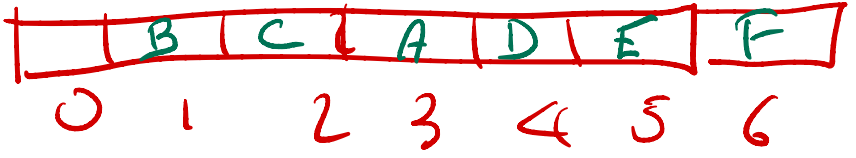
Big issue in hashing: handling collisions

$m = 7$

$x \in A = [\ A\ B\ C\ D\ E\ F\ ]$

$h(x) = 3\ 1\ 2\ 3\ 2\ 3$





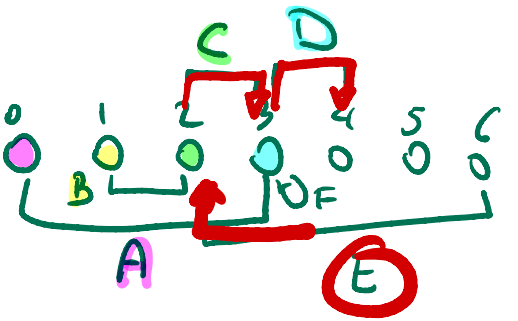
$x \in A = [A\ B\ C\ D\ E\ F]$

$h_1(x) = 3\ 1\ 2\ 3\ 2\ 3$ -

$h_2(x) = 0\ 2\ 3\ 4\ 6\ 3$

$m = 7$

HW
DEL



C

A	B	~~E~~	~~A~~ ~~D~~	D		
0	1	2	3	4	5	6

Insertion: path from node $h_1(x)$ to some node
cost = length of path

Code : $T[h_1(x)]$, $T[h_2(x)]$ see the pseudo code

$T[h_1(x)] = None$

cost : $O(1)$ time (worst-case!) membership, deletion

$O(1)$ expected amortized time for insertion

Insertion of $x \in S$, three cases for the table T:

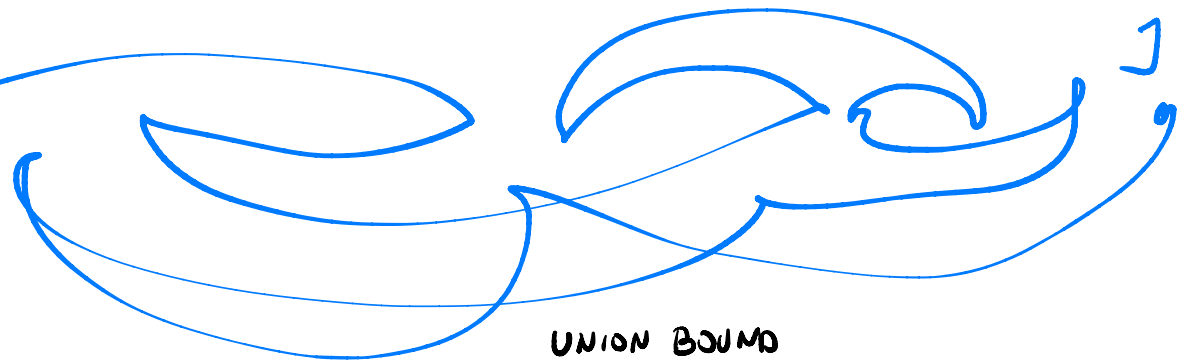
(1) T has the corresponding slot for x that is empty
$O(1)$ time

(2) slot $T[h_1(x)]$ is taken
path in G from node $h_1(x)$ to some node $j \in [m]$
ℓ = # traversed edges in G
cost is $O(1+\ell)$ time
(edges can be traversed more than once)
(multiple edges possibly)

(3) like (2) but the path is a cycle
$\Rightarrow$ REHASHING

2

$m > 2cn$, $c > 2$ hp



UNION BOUND

$Pr(A \cup B) \leq Pr(A) + Pr(B)$

$Pr(\text{path length} = \ell)$

$Pr(\exists \text{ edge } ij) = Pr(\exists x \in S : (h_1(x)=i \wedge h_2(x)=j) \text{ or } (h_1(x)=j \wedge h_2(x)=i))$

$\sum_{x \in S} Pr(h_1(x)=i \text{ OR } h_1(x)=j \wedge h_2(x)=i \text{ OR } h_1(x)=j \wedge h_2(x)=i) \leq 2 \sum_{x \in S} Pr(h_1(x)=i \wedge h_2(x)=j) \overset{\text{2-way ind}}{=}$

$2 \sum_{x \in S} (Pr(h_1(x)=i) \times Pr(h_2(x)=j)) \leq 2 \sum_{x \in S} \frac{1}{m^2} = \frac{2|S|}{m^2} = \frac{2n}{m^2} < \frac{1}{cm}$

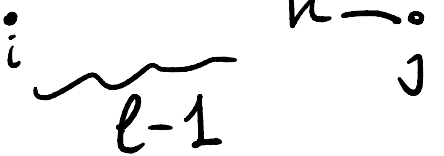
$m > 2cn \iff \frac{m}{c} > 2n$

$\frac{2n}{m^2} < \frac{m}{c} \cdot \frac{1}{m^2} = \frac{1}{cm}$

$i \overset{\ell}{\rightsquigarrow} j \equiv$ there exists a path from i to j of length $\ell > 1$ $\quad \leq \frac{1}{cm}$

$$\Pr[\exists\, i \overset{\ell}{\rightsquigarrow} j] =$$

$$\Pr[\exists k : i \overset{\ell-1}{\rightsquigarrow} k \wedge \text{edge } kj] \leq$$



$$\sum_k \Pr[\underbrace{i \overset{\ell-1}{\rightsquigarrow} k}_{A} \wedge \underbrace{\text{edge } kj}_{B}] =$$

$$\Pr[B \mid A] = \frac{\Pr[A \wedge B]}{\Pr[A]}$$

$$\sum_k \Pr[i \overset{\ell-1}{\rightsquigarrow} k] \times \Pr[\underbrace{\text{edge } kj}_{B} \mid \underbrace{i \overset{\ell-1}{\rightsquigarrow} k}_{A}]$$

$\sum_k$: $\leq m$

$\Pr[i \overset{\ell-1}{\rightsquigarrow} k]$ (A): $\leq \frac{1}{c^{\ell-1} m}$ by induct

$\Pr[\text{edge } kj \mid i \overset{\ell-1}{\rightsquigarrow} k]$ (B | A): $\leq \frac{1}{cm}$

like $(\exists \text{edge } i)$: $\sum_{x \in S}$ $\quad \sum_{x \in S - i \rightsquigarrow k}$

$$\leq \cancel{m} \times \frac{1}{c^{\ell-1} m} \times \frac{1}{c \cancel{m}} = \frac{1}{c^{\ell} m}$$

$$\sum_{\ell} \ell \cdot \frac{1}{c^{\ell} m} = \text{expected path length}$$

① empty slot → $O(1)$

② path from $i = h_1(x)$ $\underbrace{\text{run}}_{\ell}$ cost $= O(1+\ell)$ with pr $\leq \frac{1}{c^{\ell} m}$

expected $O\left(1 + \sum_{\ell \geq 1} \ell \cdot \frac{1}{c^{\ell} m}\right) = O(1)$ (no cycle)

D. Knuth, Concrete Mathematics, p. 33

$$\sum_{\ell \geq 1} \ell \left(\frac{1}{c}\right)^{\ell} < \frac{c}{(c-1)^2}, \quad c > 2$$

③ we cycle when inserting ⇒ we need REHASHING

choose $h_1, h_2 \in \mathcal{H}$ again
reinsert all the keys from scratch
$O(n)$ time if it succeeds

Note: during REHASHING maybe we need REHASHING again

$$\Pr(\text{REHASHING}) \le \Pr(\exists \text{ cycle in } G) = \sum_{i=0}^{m-1} \Pr[\exists \text{ path from } i \text{ to } j=i]$$

$$= \sum_{i=0}^{m-1} \sum_{\ell \ge 1} \Pr(\exists \text{ } i \xrightarrow{\ell} j=i) \overset{\text{UNION BOUND}}{\le} \sum_{i=0}^{m-1} \sum_{\ell \ge 1} \frac{1}{c^\ell m} = \frac{1}{m} \sum_{i=0}^{m-1} \sum_{\ell \ge 1} \frac{1}{c^\ell}$$

$\Pr(\exists \text{ } i \xrightarrow{\ell} j=i) \le \frac{1}{c^\ell m}$

$\sum_{\ell \ge 1} \frac{1}{c^\ell} < \frac{1}{c-1}$

$$< \frac{1}{m} \sum_{i=0}^{m-1} \frac{1}{c-1} = \frac{1}{c-1} = p < 1$$

$c > 2$

- REHASHING occurs with probability nearly $p = \frac{1}{c-1} < 1$
- REHASHING takes $O(n)$ if it succeeds

(a bit sloppy)

Q: How many REHASHINGs ?

# REHASHINGs	joint probability
1	p
2	p^2
3	p^3
⋮	⋮
t	p^t

Expected number of REHASHINGs:

$$\sum_{t \geq 1} t\, p^t = O(1) \quad \text{as } p < 1$$

Note: no infinite loop of REHASHINGs as $\lim_{t \to \infty} p^t = 0$