Hash tables

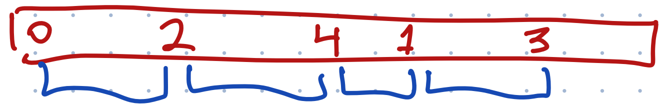
Universe $\mathcal{U} = \{0, 1, \ldots, 2^w - 1\} = [2^w]$ — w-bit words

Table $T[0, \ldots, m-1]$ — $m = 2^\ell$ — ℓ-bit labels

Hash function $h : \mathcal{U} \to [m]$

$\cancel{h(x) = \lfloor \phi x \rfloor \bmod m}$

NO



AT&T routers

Family $\mathcal{H}$ of hash functions ← fixed in code

When we init a hash table, choose $h \in \mathcal{H}$ at random

Use h for the lifetime of the table.

$h(s, x)$:
- x — object (different)
- s — salt (fixed randomly)

Assumptions about hash functions (families)

~~Uniform~~ $\Pr_{h \in \mathcal{H}}[h(x) = i] = \frac{1}{m}$ for all $x \in \mathcal{U}$, for all $i \in [m]$

$h_0(x) = 0 \quad h_1(x) = 1 \quad h_2(x) = 2 \quad \ldots$

$\mathcal{H} = \{h_i \mid i \in [m]\}$

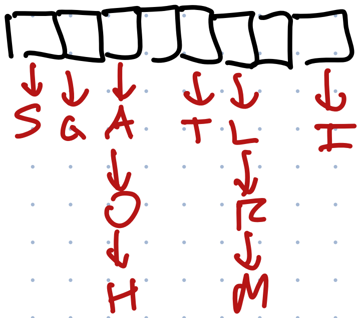
Universal: $\Pr_{h \in \mathcal{H}}[h(x) = h(y)] \leq \frac{1}{m}$ for all $x \neq y$

⇑

2-Uniform: $\Pr[h(x) = i \text{ and } h(y) = j] = \frac{1}{m^2}$ for all $x \neq y$, all i, all j

Ideal: All hash values are uniform and totally independent

= k-uniform for all k



Chained hashing

$T[i]$ = linked list of items with hash value i.

Expected Time to search for x = $O(E[\text{length}(T[h(x)])] + 1)$

Hash table stores $y_1 \dots y_n \neq x$

$$E[\text{len}(T[h(x)])] = \sum_{i=1}^{n} \Pr[h(x) = h(y_i)] \leq \frac{n}{m}$$

universal!

$E[\text{Time for unsuccesful search}] = O(\frac{n}{m} + 1)$

load factor

If load factor $= \Theta(1)$ h is universal

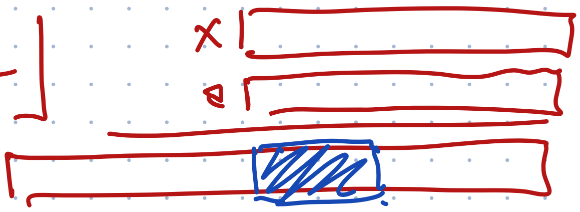
$E[\text{time}] = O(1)$!   How?

- $h(x) = ((ax+b) \bmod p) \bmod m$
  - salt
  - prime# $> m$
  - $0 \leq a \leq p-1$ $0 \leq b \leq p-1$

weakly universal

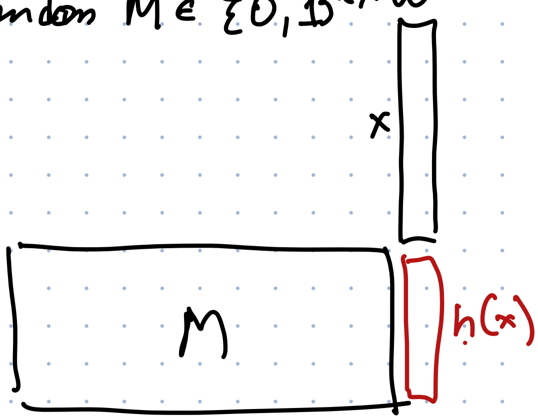
$\Pr[h(x) = h(y)] \leq \frac{2}{m}$

- $h(x) = \left\lfloor \frac{ax \bmod 2^w}{w^{w-\ell}} \right\rfloor$



- Random matrix $h : [2^w] \to [2^\ell] = \{0,1\}^w \to \{0,1\}^\ell$

Random $M \in \{0,1\}^{\ell \times w}$ $h_M(x) = Mx \bmod 2$



$$x = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

$$M \begin{bmatrix} 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$

$$\begin{matrix} 0 \\ 0 \\ 1 \\ 1 \end{matrix} \oplus \begin{matrix} 1 \\ 1 \\ 0 \\ 0 \end{matrix} \oplus \begin{matrix} 1 \\ 0 \\ 1 \\ 1 \end{matrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} h(x)$$

Fix $x \neq y$

WLOG $x_j = 0$ $\quad y_j = 1$

$$x = \begin{bmatrix} 0 \\ 1 \\ 1 \\ \boxed{0} \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \leftarrow j$$

$$M \begin{bmatrix} 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 & 1 & 0 \end{bmatrix} \quad \boxed{0}$$

Fix every element of M except column j

$\Rightarrow h(x)$ is fixed!

$h(y)$ does depend on jth column

2^ℓ possibilities for column

each yields a diff $h(y)$

$$\Pr[\underbrace{h(y)}_{\text{uniform}} = \underbrace{h(x)}_{\text{Fixed}}] = \frac{1}{m}$$

<u>2-uniform:</u> $\quad h(x) = \underset{\substack{\uparrow \\ \text{random matrix}}}{M} x \oplus \underset{\substack{\uparrow \\ \text{random vector}}}{B} \pmod 2$