Review session next Tuesday: 217 Noyes Lab

Conflict exam: Wed 3-5pm

Register by Friday

Hashing : $O(1)$-time membership queries

Universe $\mathcal{U}$ of items → table of size m

hash function $h : \{0, 1, \ldots, u-1\} \to \{0, 1, \ldots, m-1\}$

store x at $T[h(x)]$

There will be collisions $x \neq y$ but $h(x) = h(y)$

~~"The items you store are random."~~

The hash function must be random.

Simplest : h is ideal random

Reality: h is only somewhat random

Uniform : For all $x \in \mathcal{U}$

For all $i \in [m] = \{0, 1, \ldots, m-1\}$

$$\Pr_h [h(x) = i] = 1/m$$

For any $a \in [m]$, let $\mathrm{const}_a(x) = a$ for all x

$\mathcal{H} = \{\mathrm{const}_a \mid a \in [m]\}$ is uniform

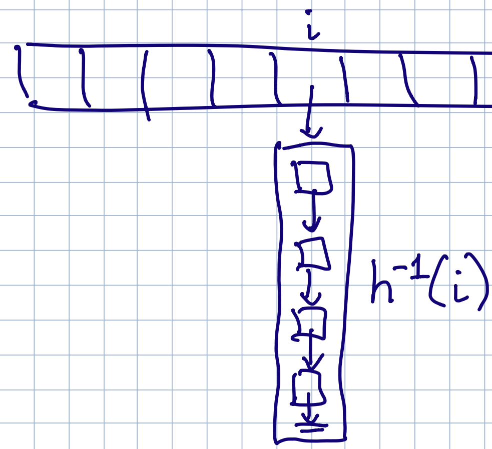
Near-Universal: For all $x \neq y \in \mathcal{U}$

$$\Pr_h [h(x) = h(y)] \leq \cancel{1/m}\ 2/m$$

Strongly universal:

$$\Pr_h\left[h(x)=i \wedge h(y)=j\right] \le 1/m^2$$

Chaining



$$E[\ell(x)] = \text{length of } x\text{'s chain} = \left|h^{-1}(h(x))\right|$$

$$= \sum_{y \in T} \Pr\left[h(x)=h(y)\right] \le \frac{n}{m} \text{ "load factor"}$$

assuming universal hashing

Universal hashing?

• Multiplicative:

prime $> |\mathcal{U}|$

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

$1 \le a \le p-1$ $\quad$ $0 \le b \le p-1$

$a \in [p]^+$ $\quad$ $b \in [p]$

a and b — salt of hashfunction

$$\mathcal{H} = \{ h_{a,b} \mid a \in [p]^+ \text{ and } b \in [p] \}$$ $p(p-1)$ functions

prime $p \Rightarrow$ For any $a \in [p^+]$ there is a unique $z \in [p]^+$ s.t. $a \cdot z \bmod p = 1$

Claim: $\mathcal{H}$ is universal.

Proof: Fix $x \neq y$. $\Rightarrow$ $ax+b \neq ay+b \bmod p$ $\qquad$ p prime, $p \geq |\mathcal{U}|$

Fix $r \neq s$

$$\boxed{\begin{aligned} ax+b &\equiv r \mod p \\ ay+b &\equiv s \mod p \end{aligned}}$$

has a unique solution (a, b), $a \in [p]^+$, $b \in [p]$

$$P_{a,b}\left[(ax+b \bmod p) = r \text{ and } (ay+b \bmod p) = s\right] = \frac{1}{p(p-1)}$$

$\Downarrow$

$p=7$

a	1	2	3	4	5	6
z	1	4	5	2	3	6

$a \cdot z \bmod p = 1$

$$\boxed{\begin{array}{l} ax+b \equiv r \mod p \\ ay+b \equiv s \mod p \end{array}}$$

$$\Rightarrow a = \frac{r-s}{x-y} \qquad b = \frac{sx-ry}{x-y}$$

$$\Pr_{a,b}[h_{ab}(x) = h_{ab}(y)] = \frac{N}{p(p-1)} \le \frac{1}{m} \quad \square$$

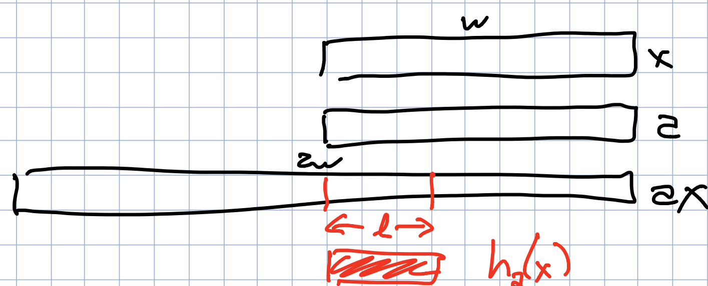
where $N = \#(r,s)$ s.t. $r \bmod m = s \bmod m$

$$N \le p \cdot \left\lfloor \frac{p}{m} \right\rfloor \le \frac{p(p-1)}{m}$$

$\mathcal{U} = [2^w]$ $\quad m = 2^\ell \quad$ odd $a \in \mathcal{U}$

$$h_a(x) = \left\lfloor \frac{(a \cdot x) \bmod 2^w}{2^{w-\ell}} \right\rfloor$$

```
#define HASH(a,x) ((a)*(x) >> (w-ℓ))
```



$\mathcal{H} = \{h_a \mid a \in \mathcal{U}$ and a odd$\}$ is near-universal

1997

For any set of items

Choose $h \in$ universal Family

$$E[\max_x T(x)] = O(1)$$

Even if $m = n$

Even if ideal random h

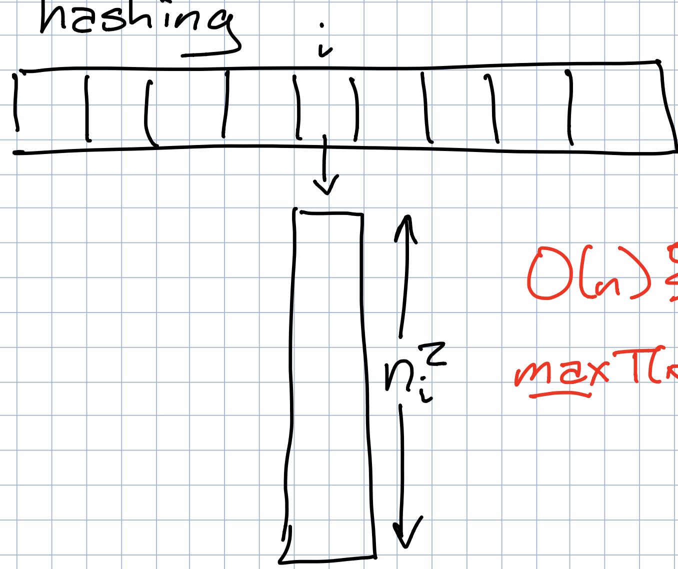
$$\max \ell(x) = \Theta\left(\frac{\log n}{\log\log n}\right) \text{ whp}$$

If $m = n^2$

Universal h

$$\Pr[\underline{\text{no}} \text{ collisions}] \geq 1/2$$

"Perfect" hashing



$n_i = |h^{-1}(i)|$

$O(n)$ expected space

$\underline{\max} T(x) = O(1)$

$$E\left[\sum_i n_i^2\right] = \sum_i E[n_i^2] \leq \cdots \leq 2n - 1$$