CSL 356 Lecture 24 Sept 26

Universal Hash functions

Given a Universe $\mathcal{U}: 0, 1, 2, \ldots N-1$

a subset $S \subset \mathcal{U}$ $|S| = n$

a table T : $0, 1, 2, \cdot\, m-1$

we want to find a mapping

$h : \mathcal{U} \to T$ such that

S "behaves well"

Collision : $h(x) = h(y)$ for $x, y \in \mathcal{U}$, $x \neq y$

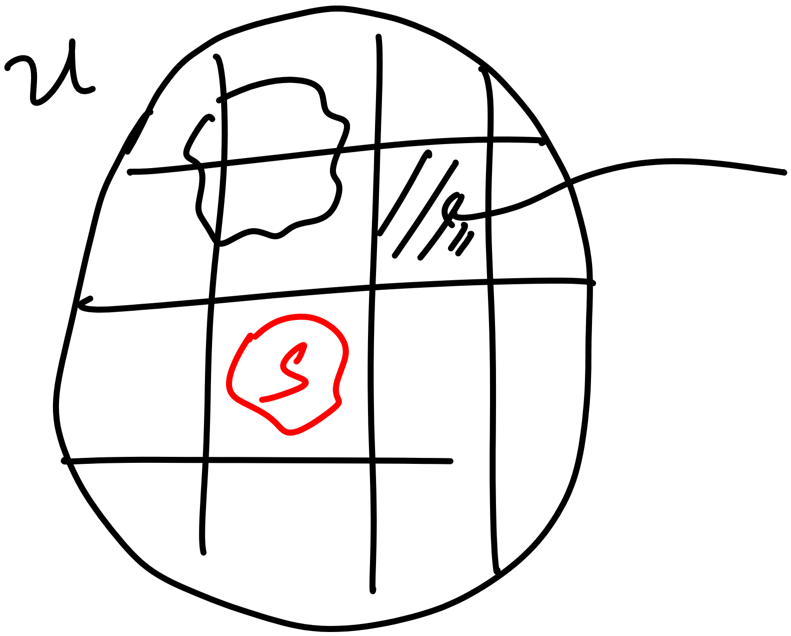
$$\delta_h(x, y) = \begin{cases} 1 & \text{if } h(x) = h(y) \\ 0 & \text{otherwise} \end{cases}$$

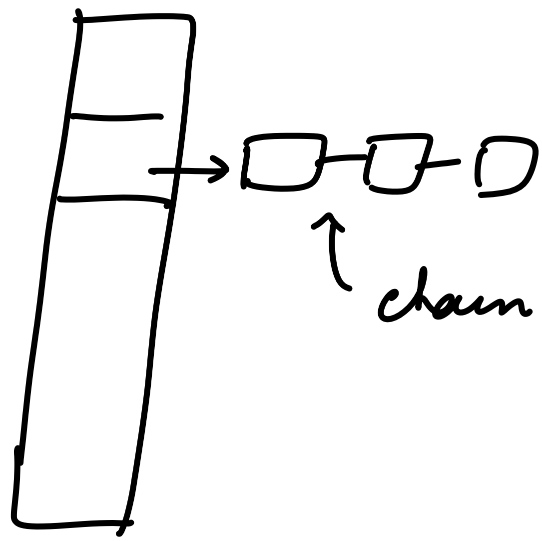
$$\delta_h(x, S) = \sum_{y \in S} \delta_h(x, y)$$

For any $x \in S$, $\delta_h(x, S)$ is "minimal"

$|\mathcal{U}| \gg |T|$ and therefore

collisions are inevitable





Hash table is a dictionary that supports Search, Insert, Delete

$$O_1(x_1)\ O_2(x_2)\ O_3(x_3)\ \ldots\ O_n(x_n)$$

$x_i \in \mathcal{U} \quad O_i \in$ Search, Insert, Delete

Under the assumption that elements are chosen uniformly at random from $\mathcal{U}$ and the hash function is "balanced" we can get good average performance.

<u>Balanced</u>: $|h^{-1}(i)| = |h^{-1}(j)|$
$\forall\, i, j \in T$

$$E\left[\delta_h(x, S)\right] = \left[\text{Prob}\,(h(x) = h(y))\right] \times |S|$$

$$\frac{1}{m} \cdot n = \frac{n}{m}$$

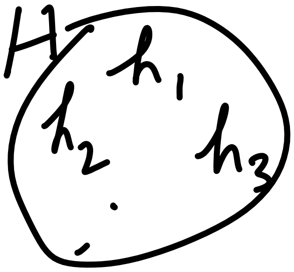
If n is $\leq c \cdot m$, this is $O(1)$

i.e. each chain has expected constant size, so total expected # operations is $O(n)$ for a sequence of n ops

A mod function, i.e. mod m is "balanced"

Lesson: We want to expand the scope of the hash function

We choose a set of hash functions H



We want to show that for an arbitrary subset S, there are many "good" hash functions in H

Universal hash family:

A collection of hash functions H is called c-universal if

$\forall x, y \in \mathcal{U}$

$$\left| \{ h \mid h(x) = h(y),\ h \in H \} \right| \leq \frac{c \cdot |H|}{m}$$

$|T|$

Claim $\underset{\text{choice of } h}{E}[\delta_h(x,S)] = O\left(\frac{n}{m}\right)$

$$\sum_{h\in H}\sum_{y\in S}\delta_h(x,y) = \sum_{y\in S}\sum_{h\in H}\delta_h(x,y)$$

$$\leq \sum_{y\in S} c\cdot\frac{|H|}{m}$$ (from defn of universal)

$$\leq n\cdot c\frac{|H|}{m}$$

So $\underset{h\in H}{E}[\delta_h(x,S)] = \frac{1}{\cancel{|H|}}\cdot n\cdot\frac{c\cdot\cancel{|H|}}{m} = \frac{c\cdot n}{m}$

Existence of Universal hash family

$$H = h_{a,b}(x) : x \to ((ax+b) \bmod N) \bmod m$$

$x,\ a, b \in \mathcal{U}$

table size

claim $\forall\ x, y \in \mathcal{U},\ |\{h(x) = h(y)\}| \leq c \frac{|H|}{m}$

$$|H| = |\mathcal{U}| \times |\mathcal{U}| - (0, \mathcal{U})$$

N is prime $\quad N^2 - N \sim N^2$

For how many choices of a, b

$h(x) = h(y)$

$$\Rightarrow ax + b = (q + rm) \bmod N$$

$$ay + b = (q + sm) \bmod N$$

There is a unique soln for each choice of $q \in \{0, 1 \ldots m-1\}$

and $r, s \in 0, m, 2m, \ldots \frac{N-1}{m}$

Total solns : $\cancel{m} \times \frac{N-1}{\cancel{m}} \times \frac{N-1}{m} \sim \frac{(N-1)^2}{m}$