CSL 356 Lecture 25 Sept 27

Hashing contd.

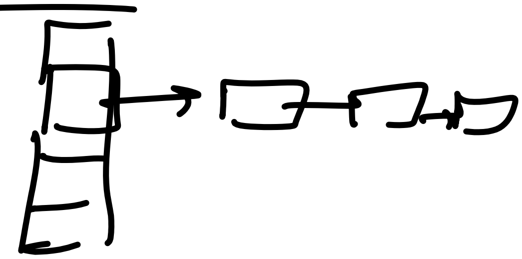
$$h_{a,b}(x) : x \to \left((ax + b) \bmod N \right) \bmod m$$

$a, b, x \in \mathcal{U}$ N is a prime

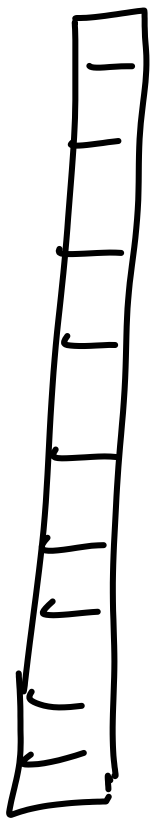
$$|\mathcal{U}| = N$$

m : size of table

Chaining method



Perfect hash function



$$S \to T$$

we must ensure that no more than one element is mapped to any location

$$|T| > |S|$$

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

Using universal hash function, the probability that $h(x) = h(y)$ for a randomly chosen $h \in \mathcal{H}$ is $\frac{c}{m}$

If X is a 0,1 random variable and $\text{prob}(X=1) = p$ then $E[X] = p$

The total expected # collisions in a set S, where $|S| = n$

$$\underset{h \in \mathcal{H}}{E}\left[\sum_{\substack{x,y \in S \\ x \neq y}} \delta_h(x,y)\right] = \sum_{x,y} E[\delta_h(x,y)]$$

all pairs

$$\leq \binom{n}{2} \cdot \frac{c}{m}$$

$f = \binom{n}{2}\frac{c}{m}$. Then by Markov's inequality Prob that the no. of collisions exceed $2 \cdot f \leq \frac{1}{2}$

For $c = 2$, $\boxed{m \geq 4n^2}$, the value of $2f$ is less than $\frac{1}{2}$

i.e. no collisions

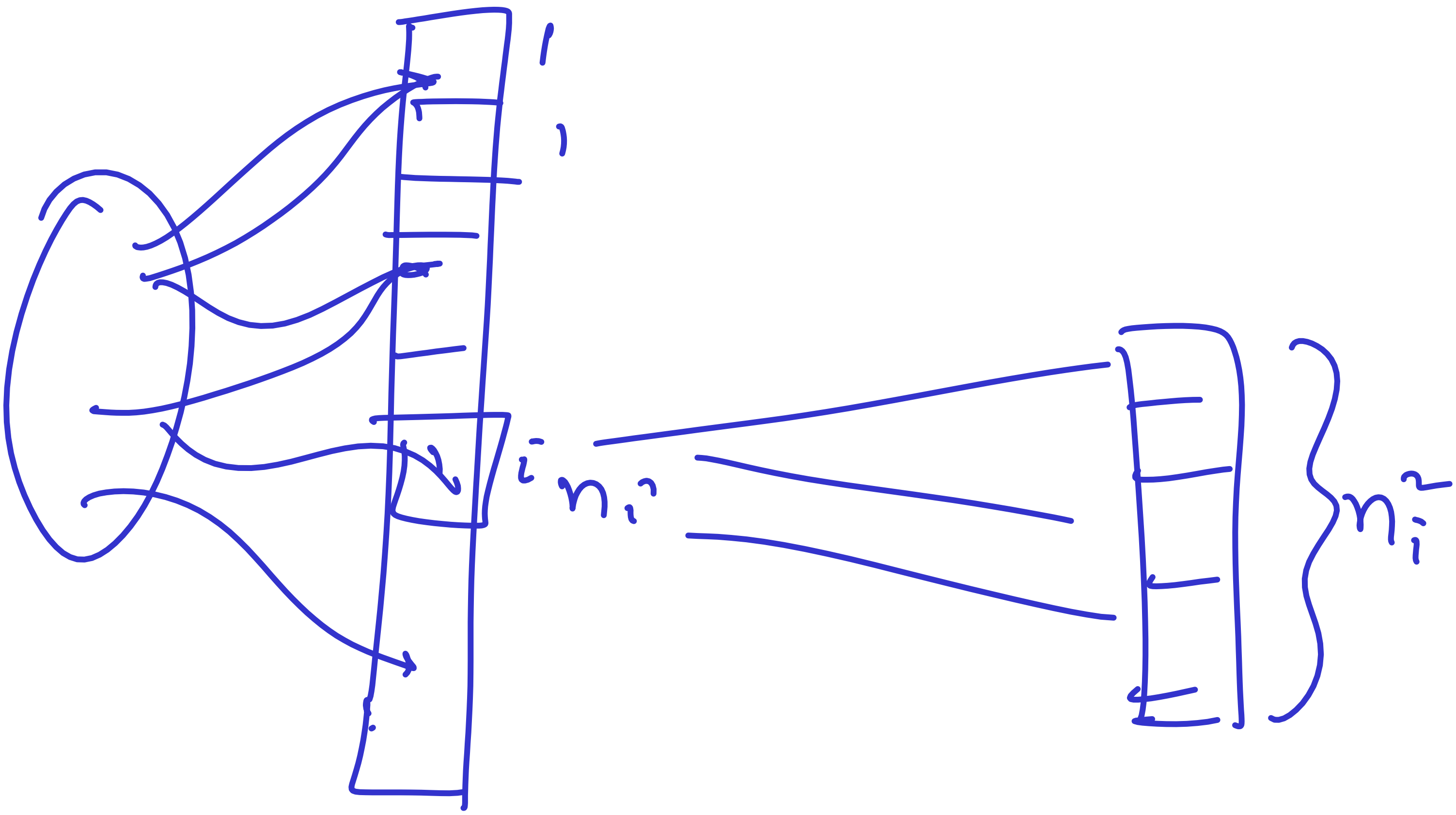
With prob $\frac{1}{2}$, there are no collisions if table size is about $\Omega(n^2)$

We use a two level strategy

→ First we hash the elements using a function from H.

(There could be collisions, suppose there are n_i elements hashed to location i, $n_i \geq 0$

→ Next level, for elements in location i, use the previous observation, i.e. hash these elements, S_i using $4n_i^2$ locations



$$E\left[\sum_i n_i^2\right] = O(n)$$

see proof in notes