Hashing

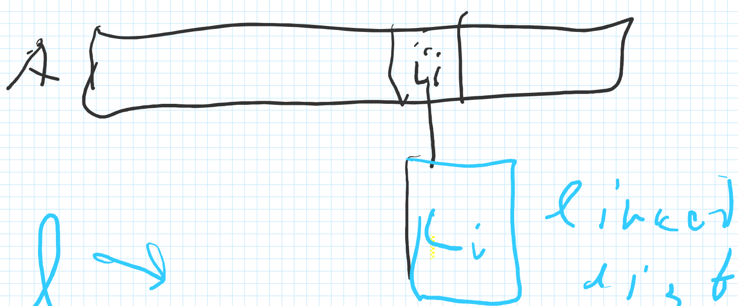
U: very large

$S \subseteq U$ maintain S

$f : U \to [0, m] = \{0, 1, 2, \dots, m\}$

hash func.

$A[0 \dots m] \leftarrow f(S)$



$U \to [0, \dots, m-1]$

u $\frac{u}{m}$

$U \to [0, \dots, m-1]$ $u >> m, n$ $|S| = n$

$u = |U|$ m^u huge

$\log_2 m^u = O(u \log m)$

Pick a small set of functions that is "good".

that is "good".

<u>Def</u> Given two elements $x, y \in U$
a family of function H is 2-Universal
if $\boxed{f(x) = f(y)}$

randomly pick $f \in H$ s.t.

$$P[f(x) = f(y)] \leq \frac{1}{m}$$

For any x, y.

Want H to be small.

$$Z_p = \{0, 1, \dots, p-1\}$$

p is a prime.

$[1, \dots n] \sim c\frac{n}{\log n} \leftarrow$ number of primes in this range

$1 \dots \sqrt{p} \quad p$

$Z_p \qquad x \bmod p =$ remainder of dividing x by p

<u>Lemma</u>

$x \in \mathbb{Z}_p \quad x \neq 0 \Rightarrow \exists y$ unique
s.t. $x \cdot y \equiv 1 \mod p$
$xy \equiv_p 1$

$a \in \mathbb{Z}_p \setminus \{0\} \quad b \in \mathbb{Z}_p$

size

$H = \{ax + b \mod p \mid a, b\} \quad p(p-1)$

$h(x) = (ax+b) \mod p$

Claim $ax + b$ is a permutation

$h(x) = ax + b$

$h(\mathbb{Z}_p) = \mathbb{Z}_p$

Proof $r, s \in \mathbb{Z}_p \quad h(r) = h(s)$

$\Leftrightarrow ar + b = as + b \mod p$

$a(r-s) = 0 \mod p$

$a^{-1} \quad r - s = 0$

$\Rightarrow r = s$. $\blacksquare$

Lemma

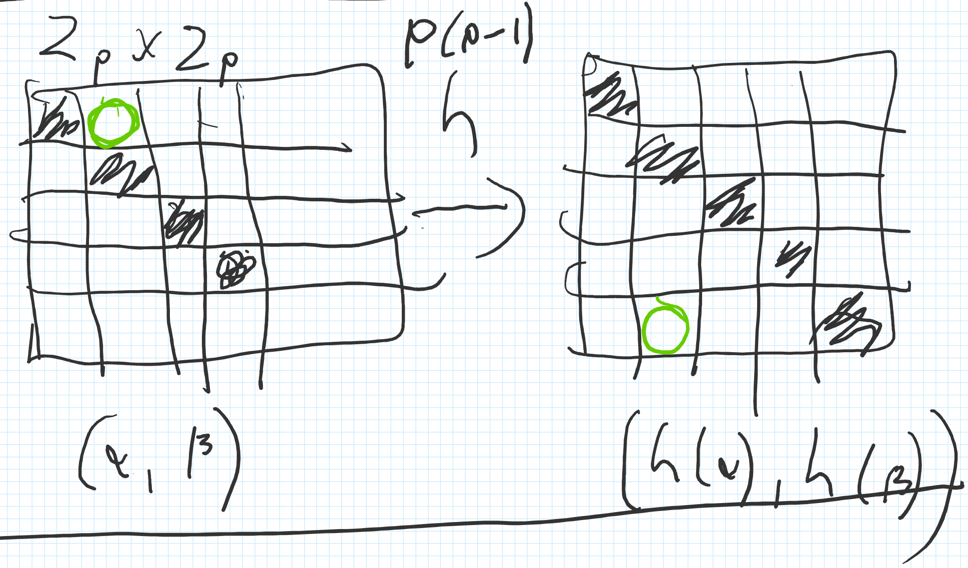
Fix a, b and consider x, y that are the solution to

$$\begin{cases} h(x) = s \\ h(y) = t \end{cases} \qquad s, t \in \mathbb{Z}_p \quad s \neq t$$

$$\begin{cases} ax + b = s \\ ay + b = t \end{cases}$$

$$\begin{cases} \alpha x + b = s \\ \alpha y + b = t \end{cases}$$

Unique solution.

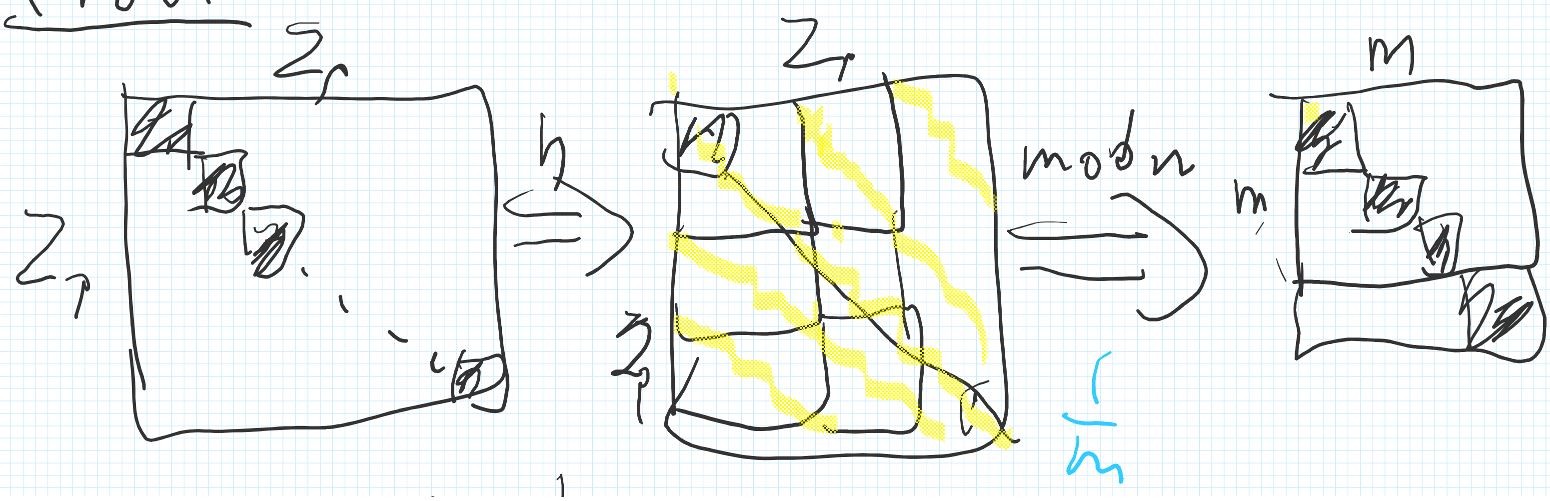


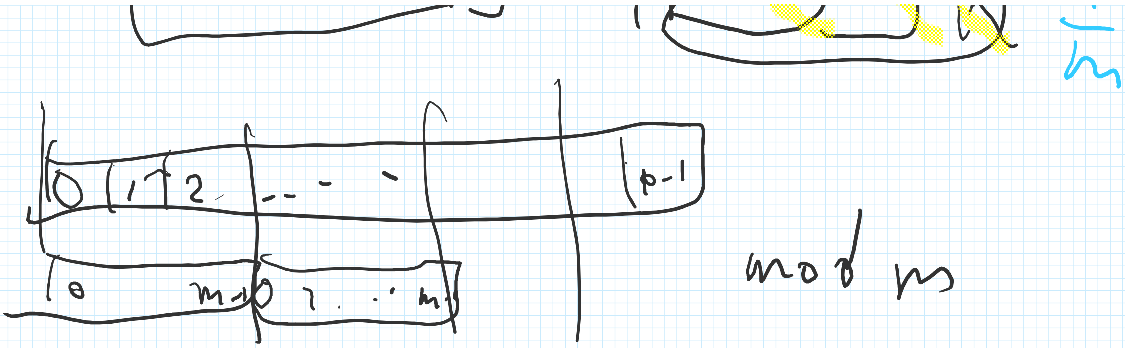
$$H = \{((ax+b) \bmod p) \bmod m\}$$

$$U = \mathbb{Z}_p \to \mathbb{Z}_m$$

Claim H is 2-universal.

Proof





$$\boxed{\Pr[h(x)=h(y)] \le \frac{1}{m}}$$

$(ax+b, ay+b) \bmod p$

Lemma

Let $n = |S|$ the expected number of collision,

$n < m \qquad n << p$

$m = cn$

$$E[\#] = \binom{n}{2}\frac{1}{m} \le \frac{n}{2c}$$

Proof

$S = \{s_1, s_2, \ldots, s_n\} \subseteq \mathbb{Z}_p$

$\binom{n}{2}$ pairs

$Z_{i,i'} = 1 \iff s_i$ and $s_{i'}$ collide

$$P[Z_{i,i'} = 1] = \underset{h \in H}{P}[h(s_i) = h(s_{i'})] \leq \frac{1}{m}$$

$$E[\#] = E\left[\sum_{i<i'} Z_{i,i'}\right] = \sum_{i<i'} E[Z_{i,i'}]$$

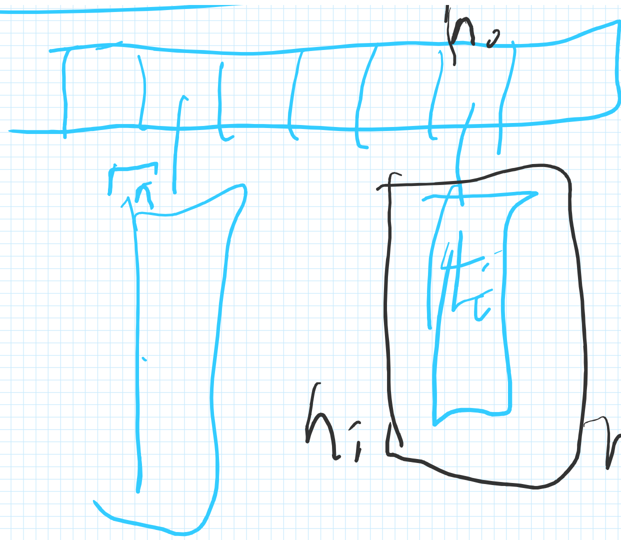
$$= \binom{n}{2} \frac{1}{m}$$

$m = 4n \implies \frac{n}{8}$ collisions

$$P[\# \geq n] \leq \frac{1}{8}$$

n_0

a, b



h_i hash table

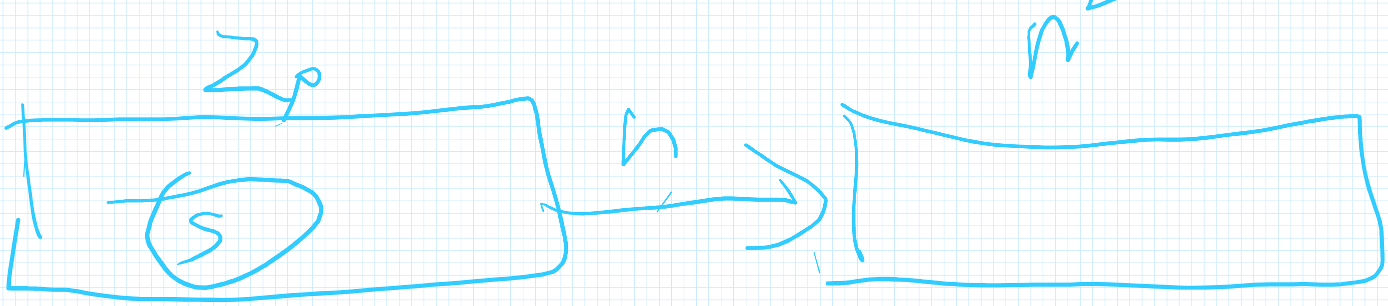
a, b

$\sum t_i^2 \le 6n$

t_i

t_i^2

$m = n^2$



$E[\#] = \frac{1}{2} \quad \le \frac{1}{2}$