473 | 10/14/21

L15: Hashing

(K, V): pairs

$K \in U$ - universe

insert$((u, v))$

find(u)

delete(u) : delete the pair associated with u

$O(1)$

$S \subseteq U$ set of elements we want to store

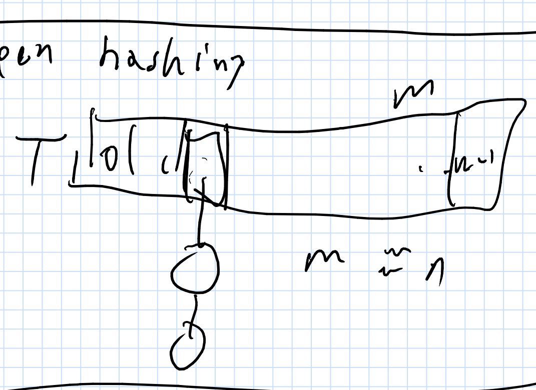
$n = |S|$.

$n << |U|$ open hashing

$h: U \to [m]$

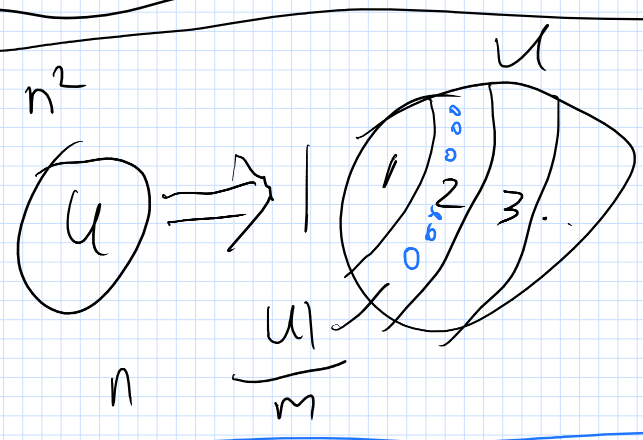
$T[h(e)] \leftarrow e$

h: hash function



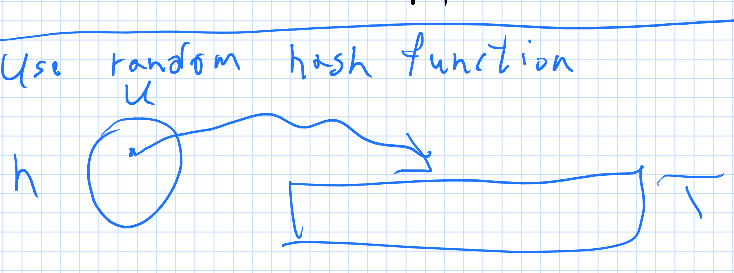
$m \approx n$

$|U| > n^2$



n $\frac{|U|}{m}$

Use random hash function



$\{0, \dots, m-1\}^U \ni f$

$| \quad " \quad | = m^{|U|}$

$\log_2 m^{|U|} = |U| \log_2 m$

---

p: prime number

$p > |U|$ large prime number

$\mathbb{Z}_p = \{0, 1, 2, \dots, p-1\}$

$(\mathbb{Z}_p, +, *)$

$a, b \in \mathbb{Z}_p$ $a \cdot b \pmod p \in \mathbb{Z}_p$

$a + b \pmod p$

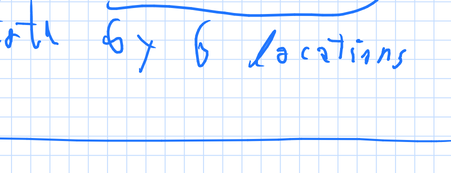
$\mathbb{Z}_p$ it is a finite field.

$\forall a, b \in \mathbb{Z}_p$ $a \neq b$ $ab \equiv 0 \pmod p$

$a \in \mathbb{Z}_p$ $a \neq 0$ $a^{-1} \in \mathbb{Z}_p$ $a a^{-1} = 1 \bmod p$

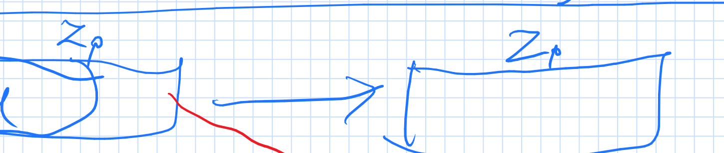
$\mathbb{Z}_p$ $a \in \mathbb{Z}_p$ $a \neq 0$

$f_a(x) = ax \pmod p$



$f_a(x) \neq f_b(x)$ $\forall a \neq b$

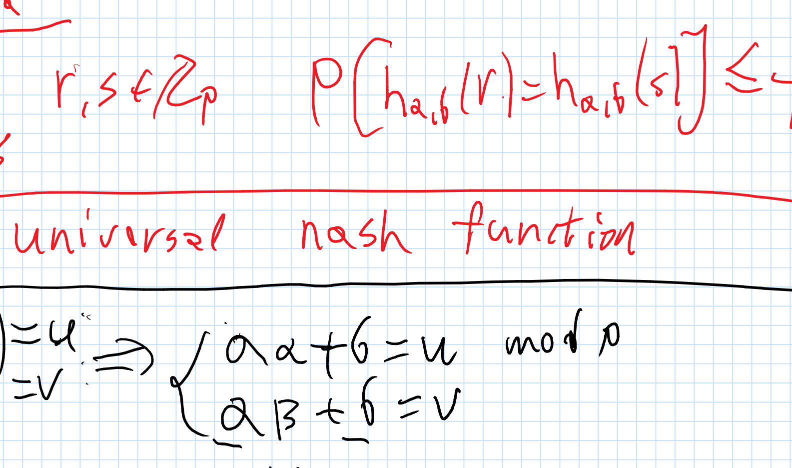
$b \in \mathbb{Z}_p$ $g_b(x) = x + b \pmod p$



rotate by b locations

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

$H = \{h_{a,b} \mid a, b \in \mathbb{Z}_p, a \neq 0\}$



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

Lemma

For $r, s \in \mathbb{Z}_p$, $r \neq s$ $P[h_{a,b}(r) = h_{a,b}(s)] \leq \frac{1}{m}$

2-universal hash function

$h(\alpha) = u$, $h(\beta) = v$ $\Rightarrow \begin{cases} a\alpha + b = u \bmod p \\ a\beta + b = v \end{cases}$

$\Downarrow$

Compute a and b

Lemma

$S \subseteq U$ n elements stored in a hash table of size m. The expected time for ins/del/search for specific element s

$O(n/m)$

Proof

$s_1, s_2, \dots, s_n$

$X_i = 1 \iff s$ and s_i collide

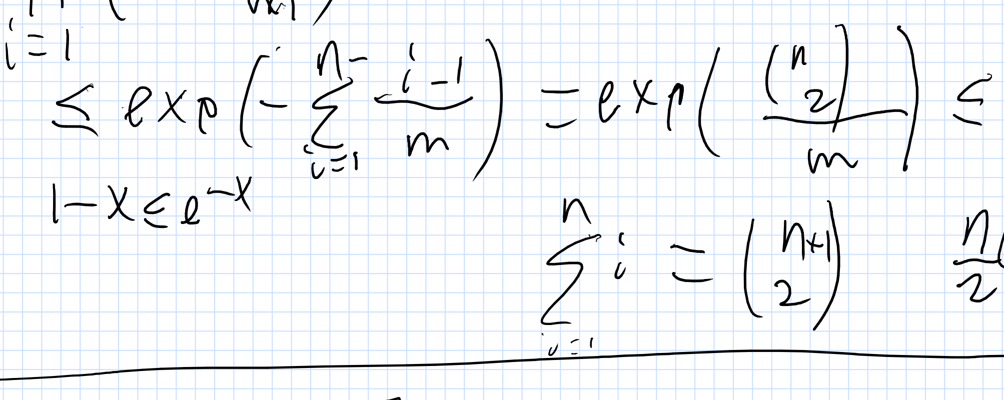
$E[X_i] = P[h(s) = h(s_i)] \leq \frac{1}{m}$

$E[\sum X_i] = \sum_{i=1}^{n} E[X_i] = \frac{n}{m}$ ▨

Birthday pardox

balls and bins

$m = \frac{1}{365}$



1 first ball

$(1 - \frac{1}{m})$ second ball

probability of no collosion

$(1 - \frac{i-1}{m})$ i th ball not to collide | no earlier ball collision

$\prod_{i=1}^{n} (1 - \frac{i-1}{m})$

$\leq \exp(-\sum_{i=1}^{n} \frac{i-1}{m}) = \exp(\frac{\binom{n}{2}}{m}) \leq \exp(-\frac{n^2}{2m}) \leq \frac{1}{2}$

$1 - x \leq e^{-x}$

$\sum_{i=1}^{n} i = \binom{n+1}{2}$ $\frac{n(n+1)}{2}$ $n = \sqrt{2m}$

$\sqrt{700} \approx 30$

$E[\#\text{ of collusions}] = \binom{n}{2} E[X_{\text{specific pair collide}}]$

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

---

n balls m entries

$E[\#\text{ of collusions}] = \frac{\binom{n}{2}}{m} \leq \frac{1}{2}$

increase m

n elements $m \approx 2n^2$

close hashing

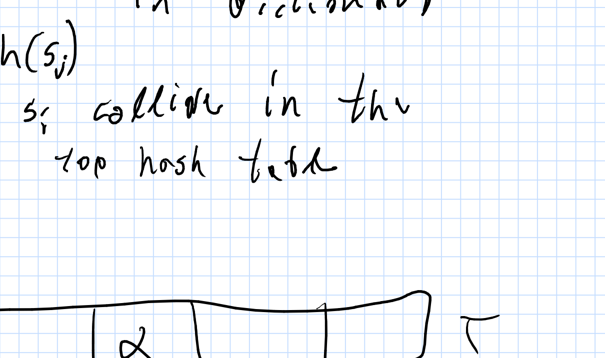
$m = 2n^2$

$P[\#\text{ of collusions} \geq 1] = P[Z \geq 2E[Z]]$

$\leq \frac{1}{2}$.

---

top level open hashing $m = 2n$



$O(\ell_i^2)$

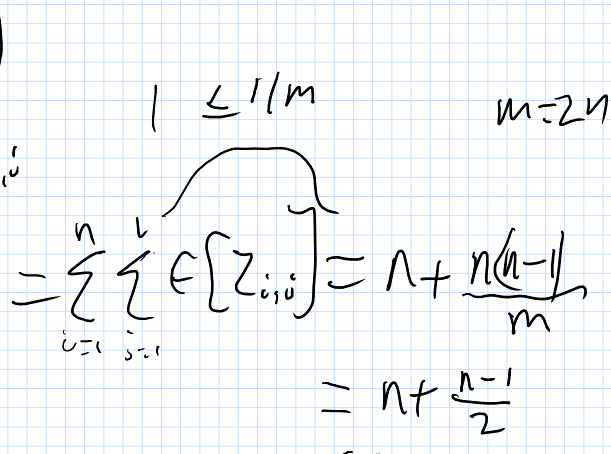
Perfect hashing.

$s_1, s_2, \dots, s_n \equiv n$ elements that we are storing in dictionary

$Z_{i,j} = 1 \iff h(s_i) = h(s_j)$
s_i and s_j collide in the top hash table

$i = j$ $Z_{i,j} = 1$

$Z_{i,j} = Z_{j,i}$



indices

$(i, j) \in \{\ell_1, \ell_2, \dots \ell_k\} \times \{\ell_1, \ell_2, \dots, \ell_k\}$

$Z_{i,j} = 1$

$\sum_{\alpha=1}^{|T|} \left|\text{size of } \alpha\text{th bucket}\right|^2$

$= \sum_{i=1}^{n} \sum_{j=1}^{n} Z_{i,j}$

1 $\leq 1/m$ $m = 2n$

$E[\sum_{i=1}^{n} \sum_{j=1}^{n} Z_{i,j}] = \sum_{i=1}^{n} \sum_{j=1}^{n} E[Z_{i,j}] = n + \frac{n(n-1)}{m}$

$= n + \frac{n-1}{2}$

$\leq 2n$