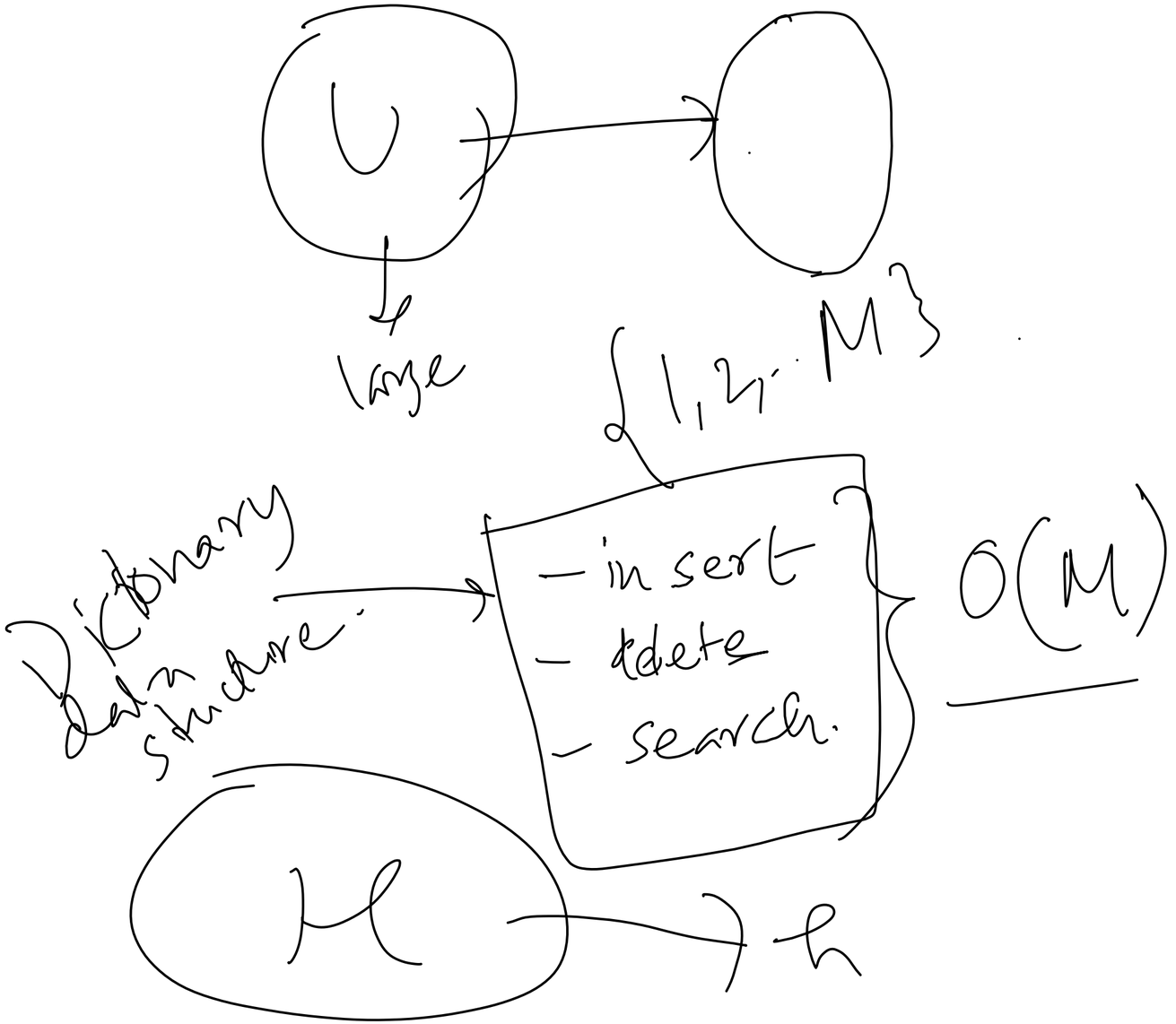
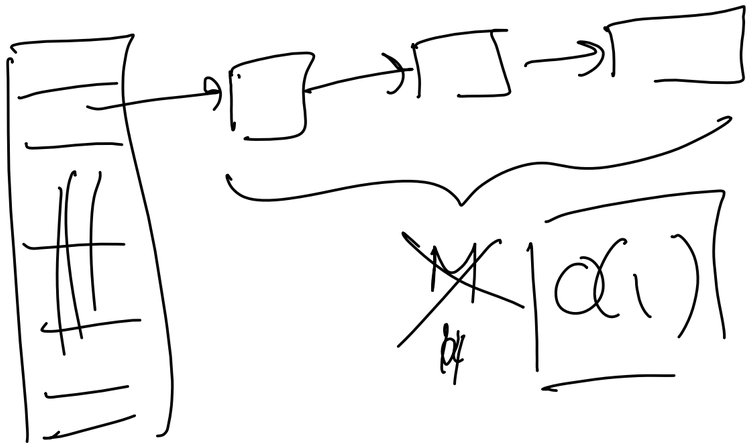
25.10.2024



$U/M \rightarrow$ collide.

$\downarrow$

all M items come from here.

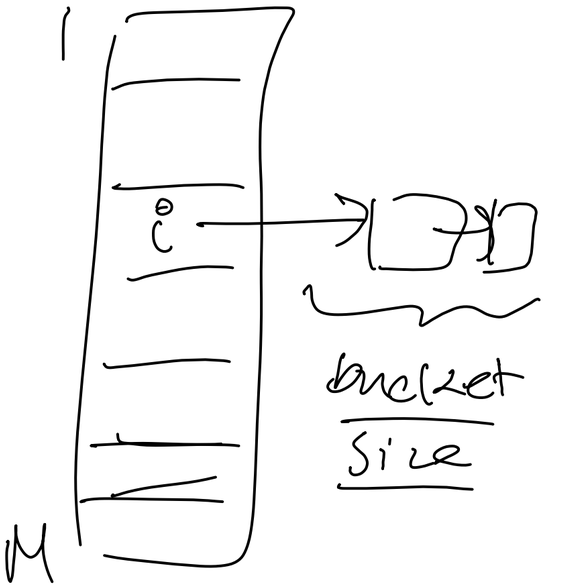


Claim: Let $\mathcal{H}$ be a universal family that maps from $U \to [M]$.

Let $x_1, \ldots, x_N$ be any set of N items.

If we choose an $h \in \mathcal{H}$ randomly; then the bucket size for any $i \in [M]$ in expectation is at most N/M

$$\mathbb{E}\left[\#\text{ collisions at any particular } i\right] \leq N/M.$$



proof: Fix i arbitrarily.

$$C_{ij} \leftarrow 1 \quad \text{if } h(x_i) = h(x_j)$$
$$\leftarrow 0 \quad \text{o/w.}$$

$$j = [M] \setminus \{i\}.$$

$$\text{Bucket size } i = \sum_{\substack{j=1 \\ j \neq i}}^{M} C_{ij}$$

$$\mathbb{E}[\text{bucket size at } i] = \mathbb{E}\left[\sum_{j \neq i} C_{ij}\right]$$

$$= \sum_{j \neq i} \mathbb{E}[C_{ij}]$$

$$= \sum_{j \neq i} \Pr[C_{ij} = 1]$$

$$= \sum_{j \neq i} \underbrace{\Pr_{h \in \mathcal{H}}[h(x_i) = h(x_j)]}$$

$$\leq \sum_{j \neq i} 1/M \qquad \text{[since } \mathcal{H} \text{ is universal]}$$

$$= \frac{N-1}{M} < \frac{N}{M}. \qquad \square$$

Corollary: Let $O_1, \ldots O_i, \ldots, O_N$ be a sequence of N insert/delete/search operation where each O_i uses item x_i.

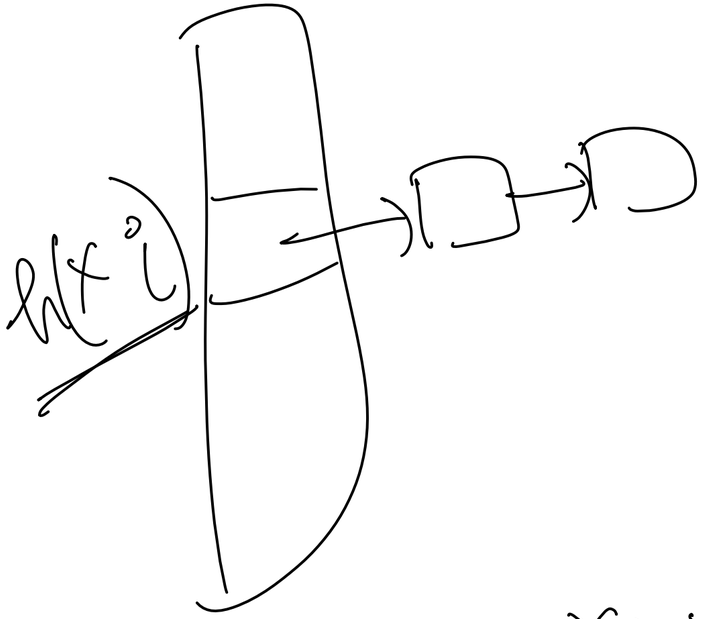
Then the expected ~~cos~~ total cost of all these operations is $O(N + N^2/M)$.

proof: Let t_i be the time taken for O_i.

Then total time $T = \sum_{i=1}^{N} t_i$

$$\mathbb{E}[T] = \mathbb{E}\left[\sum_{i=1}^{N} t_i\right]$$

$$= \sum_{i=1}^{N} \underbrace{\mathbb{E}[t_i]}$$



$$= \sum_{i=1}^{N} \mathbb{E}\left[(\text{bucket size at } i) + c\right]$$

$$\leq O(N) + \sum_{i=1}^{N} \frac{i}{M}$$

$x_1, x_2, \ldots, x_{i-1}$

$$= O(N) + O(N^2/M) \quad \square$$

What is the best choice of M?

$$O\left(N + N^2/M\right) \qquad \boxed{M = \Theta(N)}$$

$$= O(N) \text{ for } N \text{ operations}$$

$$\boxed{O(1) \text{ per operation in expectation.}}$$

Assumption: Computing the hash function takes constant!

Can we get universal family? Yes.

$\boxed{\text{Construction 1}}$

$h : \underbrace{U}_{} \to [M]$

10 character
IIITK ID

Number

W.l.o.g. we'll assume U is an integer which a power of 2.

$$2^k \leq N < 2^{k+1}$$

$|U| = 2^u \equiv U$ can be encoded by u bits.

$M = 2^m \equiv \underline{\underline{M}}$ can be encoded by m bits.

$$h : \underbrace{2^u}_{\text{large}} \to \underbrace{2^m}_{\text{small}}.$$

$\mathcal{H}$ is defined as follows.

Fix a matrix A of dimension: $m \times u$

$$m \left\{ \begin{array}{c} \\ \\ \end{array} \right. \underbrace{\begin{bmatrix} & 0/1 & \end{bmatrix}}_{u} \longrightarrow h_A.$$

2^{mu} many matrices.

$h_A(\underset{\sim}{x})$:

$\in \{0,1\}^n$

n bit strings.

$$\begin{bmatrix} & & \\ & A_{m\times n} & \\ & & \end{bmatrix} \begin{bmatrix} \\ \\ \\ \end{bmatrix} = \begin{bmatrix} \\ \\ \end{bmatrix}_{m\times 1}$$

$x_{n\times 1}$

$\in \{0,1\}^m$.

multiplications and additions are modulo 2.

$1, 0$: $1+1=0$.

for each fixed A, h_A defines a mapping from $\{0,1\}^n$ to $\{0,1\}^m$.

$\mathcal{H}$ consists of h_A s for all possible A s

$$|\mathcal{H}| = 2^{mu}.$$

<u>Example</u>:

$U = \{00, 01, 10, 11\}$. $u = 2$.

Range $= \{0, 1\}$. $m = 1$

dimension $= 1 \times 2$. $\underbrace{[0\ 0], [0\ 1], [1\ 0], [1\ 1]}_{4 \text{ choices of } A.}$

$h_{00}, h_{01}, \underline{h_{10}}, h_{11}$

	00	01	10	11
h_{00}	0	0	0	0
h_{01}	0	1	0	1
h_{10}	0	0	1	1
h_{11}	0	1	1	0

$$\begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$= \boxed{0}$$

$$(1 + 0) \bmod 2$$

$$= 1$$

Verify This is universal

Claim: The above hash family is universal.

proof: Fix x, y arbitrarily.

$$\Pr_{h \in \mathcal{H}} \left[\underline{h(x)} = \underline{h(y)} \right].$$

$$= \Pr_{h \in \mathcal{H}} \left[\underline{Ax} \equiv \underline{Ay} \mod 2 \right]$$

$$\begin{pmatrix} \\ \\ \end{pmatrix}_{m \times 1} \qquad \begin{pmatrix} \\ \\ \end{pmatrix}_{m \times 1}$$

Since $x \neq y$. therefore define $z = (x - y)$

$$= \Pr_{h_A \in \mathcal{H}} \left[\underline{\underline{A}} z \equiv \underline{\underline{0}} \bmod 2 \right] . \quad z = \underline{\underline{x-y}} .$$

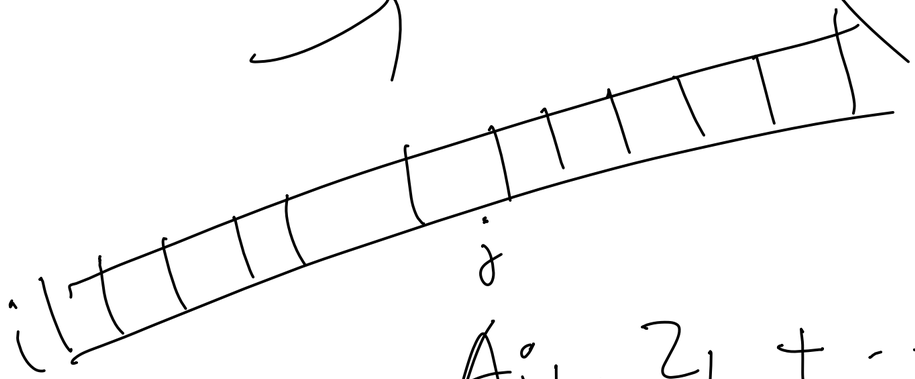
$$\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}_{m \times 1}$$

$$= \Pr_{A \in \{0,1\}^{m \times n}} \left[\underline{\underline{A z}} \equiv 0 \bmod 2 \right] .$$

$$\begin{array}{c} m \\ \text{rows.} \end{array} \begin{pmatrix} \text{———} \\ \text{———} \\ \text{———} \\ \text{———} \end{pmatrix} \begin{pmatrix} A_1 \\ \vdots \\ A_i \\ \vdots \\ A_m \end{pmatrix} \underset{m \times 1}{\begin{pmatrix} z \end{pmatrix}} = \begin{pmatrix} A_1 z \\ \vdots \\ A_i z \\ \vdots \\ A_m z \end{pmatrix}$$

$$A_i \in \{0,1\}^n .$$

$$= \prod_{i=1}^{m} \underbrace{\Pr_{A_i \in \{0,1\}^n} \left(A_i \cdot z \equiv 0 \mod 2 \right)}_{\text{focus on any particular } i}$$



$$A_{i1} \cdot z_1 + \cdots + A_{in} \cdot z_n \equiv 0 \mod 2.$$

$x \neq y$ $\exists$ some j such that $z_j \neq 0$.

$$\cancel{A_{ij} z_j} = \underbrace{A_{i1} z_1 + \cdots + A_{i(j-1)} z_{j-1} + A_{i(j+1)} z_{j+1} + \cdots + A_{in} z_n}_{\text{very end.}} \mod 2$$

Therefore this equation holds with prob $= \frac{1}{2}$.

$$= \frac{1}{2^m} = \frac{1}{M}. \qquad \square$$

Therefore this family is universal.

Construction 2.

$U \to$ u bit strings

$$\underbrace{(0\ 1\ \cdot)(\ \cdot\ )(\ 1)(0\ 1)}_{u = 32}$$

$a_1 \cdot a_2 \cdot a_3 \cdot a_4$

$\in [0, 255]$ 257 prime.

$U \to \{0, 1, \ldots, K-1\}^{\ell}$

$K = 256$

$\ell = 4$.

$M \leftarrow$ prime number which is $\geq K$.

$\ell = 9$ here

Fact: For every $x > 1$ $\exists$ a prime between x & $2x$.

$\mathcal{H}$ is defined as follows.

fix 4 number $c_1, c_2, c_3, c_4 \in [0, M-1]$

$$h(x) = (c_1 x_1 + c_2 x_2 + c_3 x_3 + c_4 x_4) \bmod M$$

x_1, x_2, x_3, x_4

each $0 \le x_i \le K-1$ $= \{0, \ldots, M-1\}$

$\mathcal{H}$: class of all h as c_1, c_2, c_3, c_4 varies. $(M)^4$

Claim: Above $\mathcal{H}$ is a universal family.

Proof: lets prove for $l = 4$.

Fix $x \neq y$ arbitrarily.

$$\Pr_{h \in \mathcal{H}} \left[h(\underset{x_1 \ldots x_4}{x}) = h(\underset{y_1 \ldots y_4}{y}) \right]$$

$$= \Pr_{h \in \mathcal{H}} \left[\begin{aligned} & c_1 x_1 + c_2 x_2 + c_3 x_3 + c_4 x_4 \\ & \equiv c_1 y_1 + c_2 y_2 + c_3 y_3 + c_4 y_4 \mod M \end{aligned} \right]$$

$$= \Pr_{(c_1, c_2, c_3, c_4)} \left[c_1(x_1 - y_1) + c_2(x_2 - y_2) + c_3(x_3 - y_3) + c_4(x_3 - y_4) \equiv 0 \bmod M \right]$$

Let $x_3 \neq y_3$.

$$= \Pr_{(c_1 \ldots c_4)} \left[c_3(x_3 - y_3) \equiv c_1(x_1 - y_1) + c_2(x_2 - y_2) + c_4(y \right]$$