26.10.2024

$2^k \le M < 2^{k+1}$

$M \leftarrow$ prime.

U : $(00 \longleftrightarrow) \; - \; - \; \underbrace{\overset{u \text{ bits}}{(\quad)}}_{k+1} \underbrace{(\quad)}_{k+1}$

$l = 4$: $M \to \{0, 1, \ldots, M-1\}$. $\longrightarrow c_1, c_2, c_3, c_4$.

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

c_1, c_4. $\downarrow$

x_1, x_2, x_3, x_4.

$|(\ )t| = M^4$.

$$\Pr_{h\in\mathcal{H}}\left[h(x_1,\dots,x_4) = h(y_1,\dots,y_4)\right].$$

$$\exists i:\ x_i \neq y_i \quad ;\ i=3.$$

$$= \Pr_{h\in\mathcal{H}}\left[\underline{c_1(x_1-y_1)+\cdots+c_4(x_4-y_4)} \equiv 0 \bmod M\right]$$

$$= \Pr_{h\in\mathcal{H}}\left[\underbrace{c_3}_{\{0,\dots,M-1\}}\underbrace{(x_3-y_3)}_{\neq 0} \equiv \underbrace{c_1(y_1-x_1)+c_2(y_2+x_2)+c_4(y_4-y_4)}_{d\,\in\,\{0,\dots,M-1\}} \bmod M\right]$$

$$\leq \frac{1}{M}.\ \square$$

c_3 is unique: $\boxed{d\cdot(x_3-y_3)^{-1} \bmod M}$

Fact: $a \in \{1, \ldots, p-1\}$ where p is prime.

$\exists$ a unique b such that

$$a \cdot \underline{b} \bmod p \equiv 1$$

b is called the inverse of a

Perfect Hash functions

Perfect $\equiv$ no collisions.

$U // M$

Assumption: The set we want to maintain is not changing.

$(x_1, \cdots, x_N)$ is known from before.

$$h : \underset{\text{large}}{U} \to \underset{\text{small.}}{[M]}$$

- no collision
- fast computation of h. $\to$ $\underline{O(1)}$

Claim: ~~The~~ Let $\mathcal{H}$ be a universal family that maps U to $[C \cdot N^2]$. Let $h \in \mathcal{H}$ be random.

Let $x_1, \ldots, x_N$ be a fixed set of items. ~~bucket size at every buck~~

The ~~prob~~ expected number of ~~collisions~~ collisions $\equiv$ bucket size max. $< \frac{1}{2}$.

proof: Indicator random variables

$C_{ij} = 1$ if $h(x_i) = h(x_j)$

$\phantom{C_{ij}} = 0$ o/w.

$i < j$

$M = C \cdot N^2 = N^2$

$C = 1$

$$\text{total \# of collisions} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} C_{ij}$$

$$\mathbb{E}\left[\qquad \right] = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \mathbb{E}[C_{ij}]$$

$$= \sum_{i=1}^{N} \sum_{j=i+1}^{N} \Pr_{h \in \mathcal{H}}[\underbrace{C_{ij} = 1}_{h(x_i) = h(x_j)}]$$

$$= \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{1}{N^2}$$

$$= \frac{N(N-1)}{2N^2} < \frac{1}{2}.$$

Corollary: $\exists$ some $h \in \mathcal{H}$ which has no collisions.

$$\underset{h \in \mathcal{H}}{\mathbb{E}}\left[\underset{\# \text{collision}}{\text{total}}\right] < \frac{1}{2} \Rightarrow$$

$$\sum_{x \in \Omega} pr(x \cap x)$$

What do we get?

$O(1)$ worst case time for insert, delete, search.

$\Rightarrow$ ~~search all~~

How to find such an $h \in \mathcal{H}$

Iterate over all h's of $\mathcal{H}$ until we find $h \in \mathcal{H}$ for which there are no collisions.

Construction time: $O(|\mathcal{H}|)$.

Markov's inequality:

Let X be a non-negative random variable.

$$\Pr[X \geq \overset{=1}{a}] \leq \frac{\mathbb{E}[X] = 1/2}{a = 1}$$

for any $a > 0$.

Example: Throwing a random Ludo die. (fair)..

$$\underbrace{\Pr[X > 7]}_{0} < \frac{3.5}{7} = \frac{1}{2}$$

Proof: Suppose not

$$\Pr[X \geq a] > \frac{E[X]}{a}.$$

$$E[X] = \sum_{x \in \Omega} \Pr[X = x] \cdot x$$

$$= \underbrace{\sum_{x \geq a} \Pr[X=x] \cdot x}_{} + \underbrace{\sum_{x < a} \Pr[X=x] \cdot x}_{}$$

$$\geq a \sum_{x \geq a} \Pr[X=x] \quad + \quad \geq 0$$

$$\geq a \cdot \underbrace{\Pr[X \geq a]}_{> \frac{E[X]}{a}}$$

$$> E[X].$$

<u>contradiction!</u>

$\Rightarrow$ If $E[\ ]$ is small, typically the r.v. takes small values.

$$\boxed{\Pr_{h \in \mathcal{H}}[\text{total \# collision} \geq 1] < \frac{1}{2}}$$

$$\boxed{\Pr_{h \in \mathcal{H}}[\underset{=0}{\text{total \# collision}} < 1] \geq \frac{1}{2}}$$

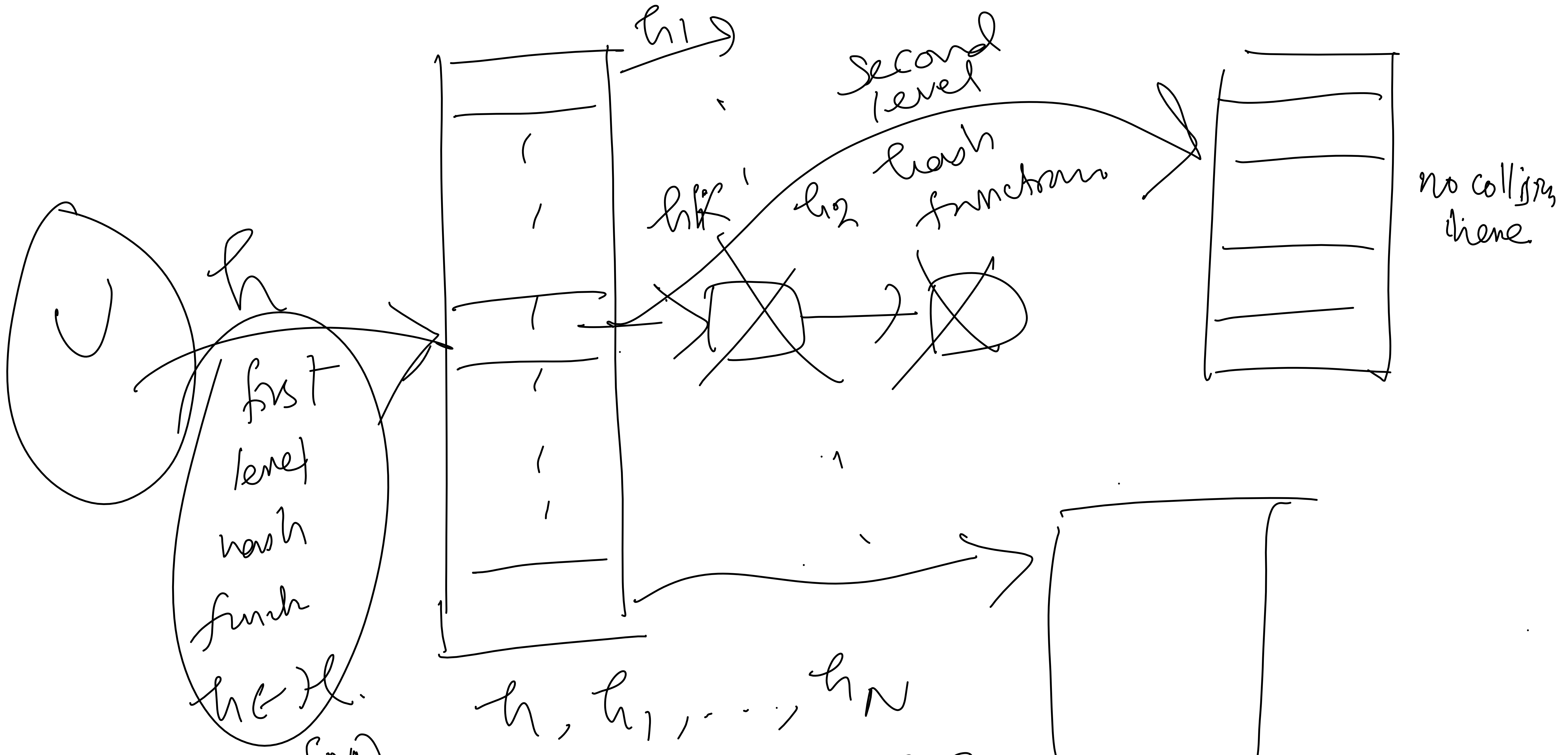
repeat: $\left(\frac{1}{2}\right)^{t}$

$(2$ tries in expectation.

$O(\ln 1/\delta)$ repetitions the random $h \in \mathcal{H}$ will have 0 collisions w.p. $\geq 1-\delta$ [for any δ]

Issue: Space usage is $O(N^2)$ to keep N items.

$\boxed{\text{two level hash functions}}$



$U \to [N]$

$h_k : U \to [L_k^2]$

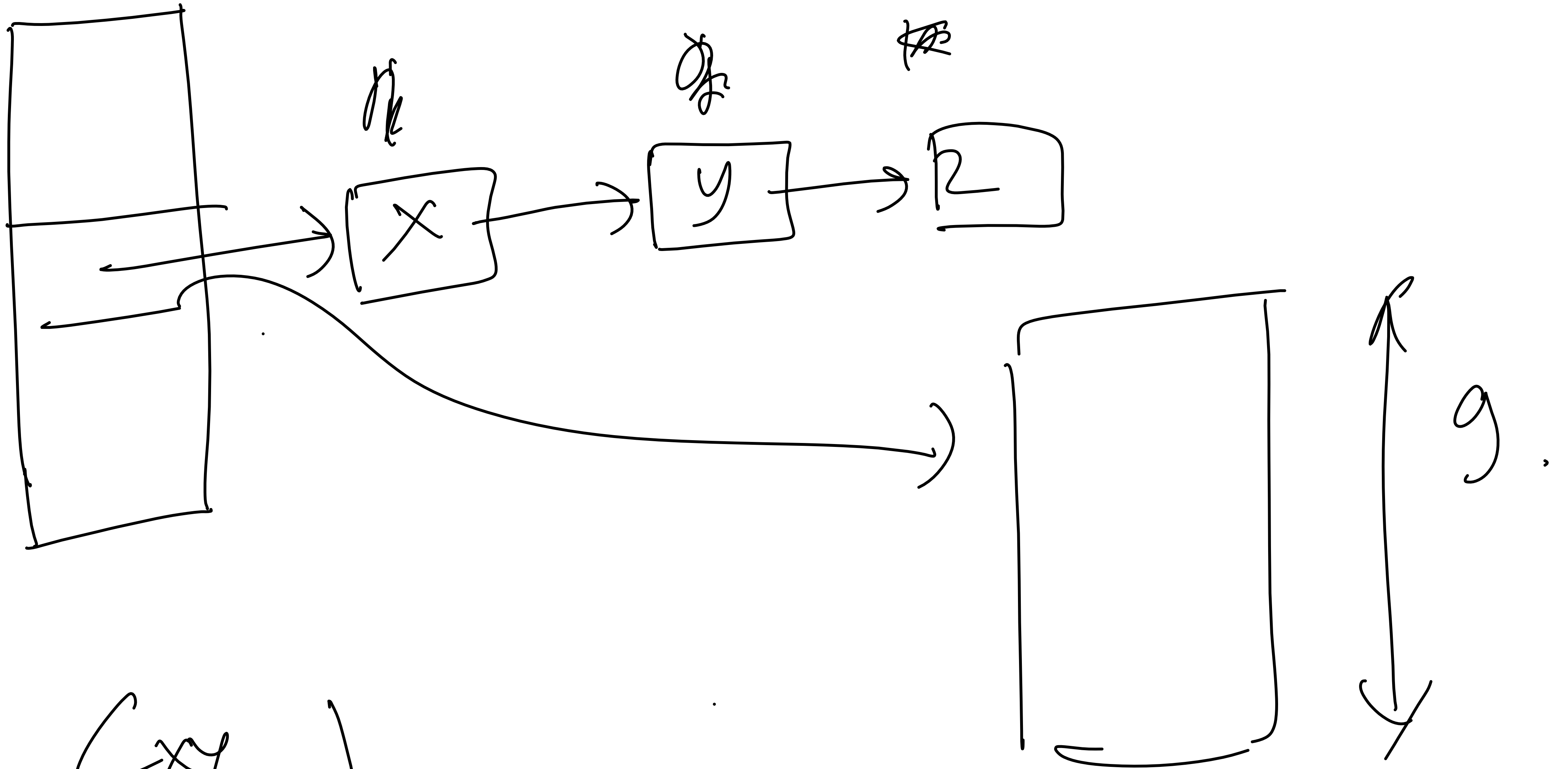
L_k = bucket size at k.

- first level: possible collisions
- second level: no collisions; (use previous construction.)

Claim: The space usag is $\underline{\underline{O(N)}}$ in expectations.

proof: $\underline{O(N)} \leftarrow$ first level

Space usage at second level: $\boxed{\sum_{k=1}^{N} \underline{L_k^2} = \sum_{i=1}^{N}\sum_{j=1}^{N} \underline{\underline{C_{ij}}}}$



C_{xy}

C_{xx}

$C_{\cancel{x}xz}$

$g \to 1$

$C_{ij} = 1$ if $h_k(x_i) = h_k(x_j)$

$$\mathbb{E}\left[\sum_{k=1}^{N} l_k^2\right] = \sum_{i=1}^{N}\sum_{j=1}^{N} \mathbb{E}[c_{ij}]$$

$$= O(N) + \sum_{i=1}^{N}\sum_{\substack{j=1 \\ j\neq i}}^{N} \underbrace{\mathbb{E}[c_{ij}]}$$

$$= O(N) + \sum_{i=1}^{N}\sum_{\substack{j=1 \\ j\neq i}}^{N} \frac{1}{N}.$$

$$\Pr\left(h(x_i) = h(x_j)\right]$$

$$\leq \frac{1}{M}$$

$$= \underbrace{O(N)}$$

$$\leq CN.$$

$$\Pr\left[\sum_{k=1}^{N} \cancel{} L_k^2 > 2 \cdot CN\right] \le \frac{1}{2}.$$

Repeat until $\left(\sum_{k=1}^{N} L_k^2 \le CN\right)$ for the first time.

$O(\log 1/\delta)$ repetitions ensure the above happens with prob $\geq (1-\delta)$.

- worst case space usage is $O(N)$
- worst time complexity is $O(1)$ per operation.

H: $2^u \to 2^m$

$$\frac{2^{um} \text{ many}}{M^4}$$