Cpts 515. 10/21/2020.

Today: Universal hash functions.

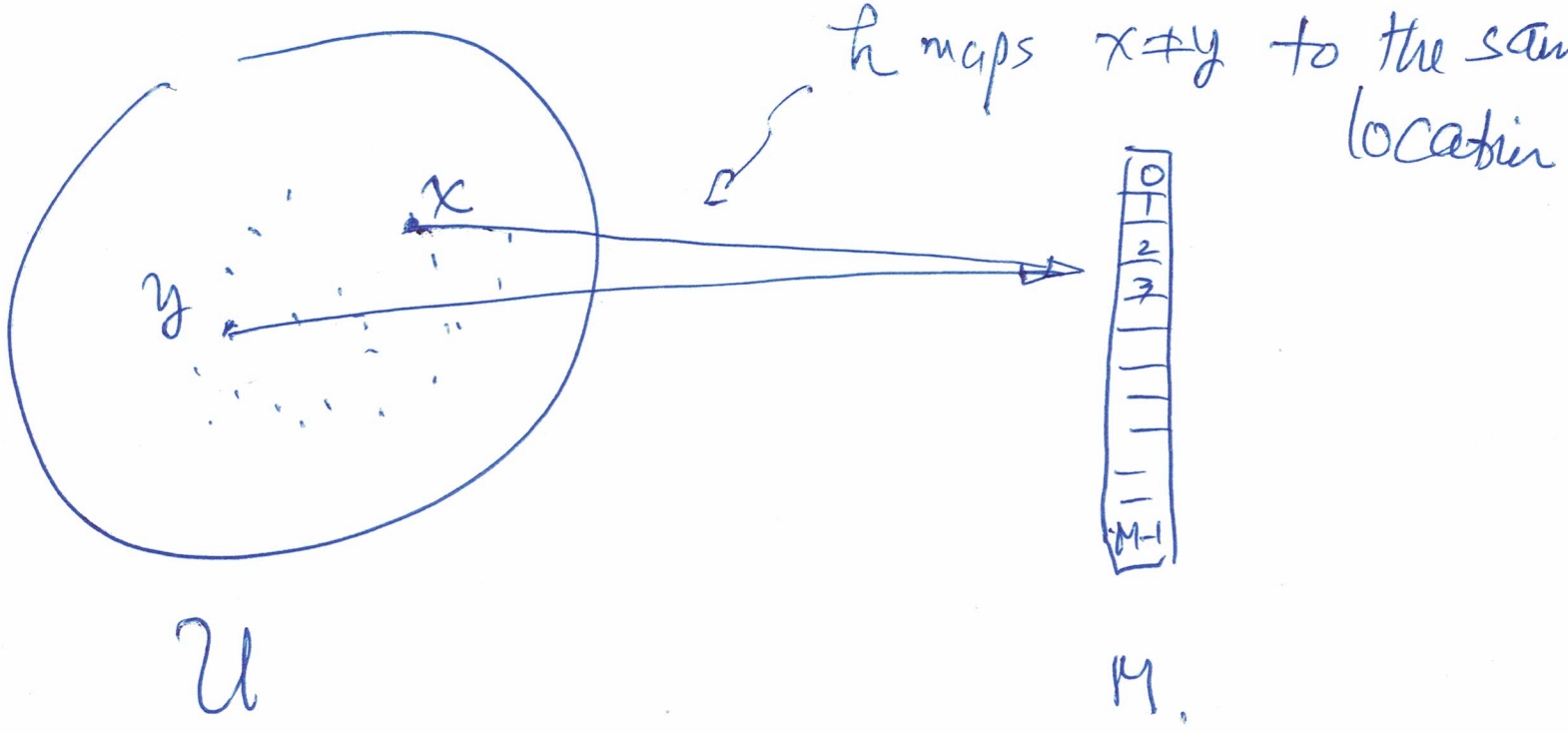
Key intuition: h is chosen randomly. However, once it's chosen, h is fixed (h itself is NOT random.)

Let $\mathcal{H}$ be a finite set of hash functions from U to $[M]$. $\mathcal{H}$ is universal if $\forall x \neq y \in U$,

$$\text{Prob}(h(x) = h(y)) \leq \frac{1}{M}.$$

collision.

where h is a random var on H.



---

h maps x to $h(x)$.

h maps y to $h(y)$.

How many choices of $h(y)$ when h is random?

M choices. ~~Among one of~~ all These choices,

One will make $h(x) = h(y)$.

φ

$\frac{1}{M}$.

Once h is chosen, h is fixed!
h is not a random function. h is randomly chosen function.

Thm. Let $\mathcal{H}$ be universal. For each $S \subseteq U$ with $|S| = N$,

$$\mathrm{Exp}(c(x, h, S)) \leq \frac{N}{M},$$

where $x \in S$, h is random var on $\mathcal{H}$.

$$c(x, h, S) = |\{y : h(x) = h(y), x \neq y \in S\}|.$$

S: 128 N

E:

Proof. Define $\delta_{xy} = \begin{cases} 1 & \text{if } h(x)=h(y) \\ 0 & \text{o.w.} \end{cases}$

collision count

For a given x,

$\sum_{x \neq y \in S} \delta_{xy}$ is the size of collision set; $C(x, h, S)$.

By def., $Exp(\delta_{xy}) \leq \frac{1}{M}$.

↳ of universal $\mathcal{H}$.

So, $Exp\left(\sum_{x \neq y \in S} \delta_{xy}\right) \leq \frac{N-1}{M} \leq \frac{N}{M}$.

$\sum$ take this if $x \notin S$.

Example of universal hash functions.

(4-bit to 2-bit hash).

$$(a_1, a_2, a_3, a_4) \xrightarrow{h} (b_1, b_2).$$

$$\begin{pmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{pmatrix} \equiv_{\mod 2} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

2 by 4 matrix of random bits.

Suppose that the random matrix is

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

Then $h(1,0,0,1) =$

$$\begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} =_{mod_2} (2_{mod_2}, 0_{mod_2})$$

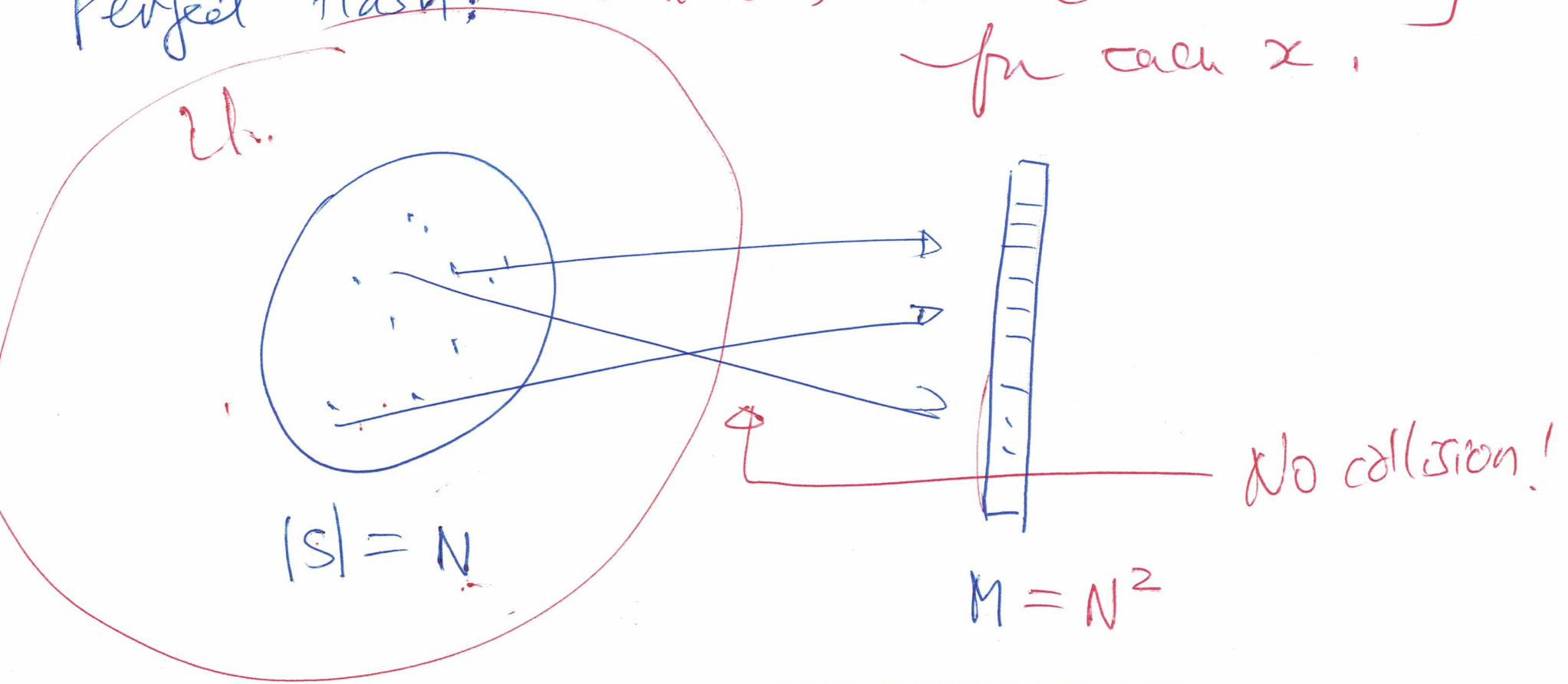
$$= (0,0).$$

One can prove:

$$\text{Prob}\,(h(x) = h(y)) = \frac{1}{4} \quad \text{for all } x \neq y \text{ in 4 bits.}$$

↑ taken over h.

Perfect Hash: collisions are constant many for each x.



Expectation of collision number for each x is upper bounded by $\frac{N}{M}$. ⇏ actual collision number is also $\leq \frac{N}{M}$.

Theorem. Let $\mathcal{H}$ be universal. $|S| = N$ and $M = N^2$. Then for each S $(\subseteq U)$,

$$\text{Prob}\left(\sum_{x \neq y \in S} \delta_{xy} = 0 \right) \geq \frac{1}{2}$$

(the Prob is taken on h randomly chosen from $\mathcal{H}$).

How to generate a collision-free hash:

① Setup: Given S, $\mathcal{H}$.

Goal: Pick a $h \in \mathcal{H}$ so that S is collision free wrt the h; i.e.,

$$\sum_{x \neq y \in S} \delta_{xy} = 0.$$

S is collision-free wrt h.

Alg of generating the h:

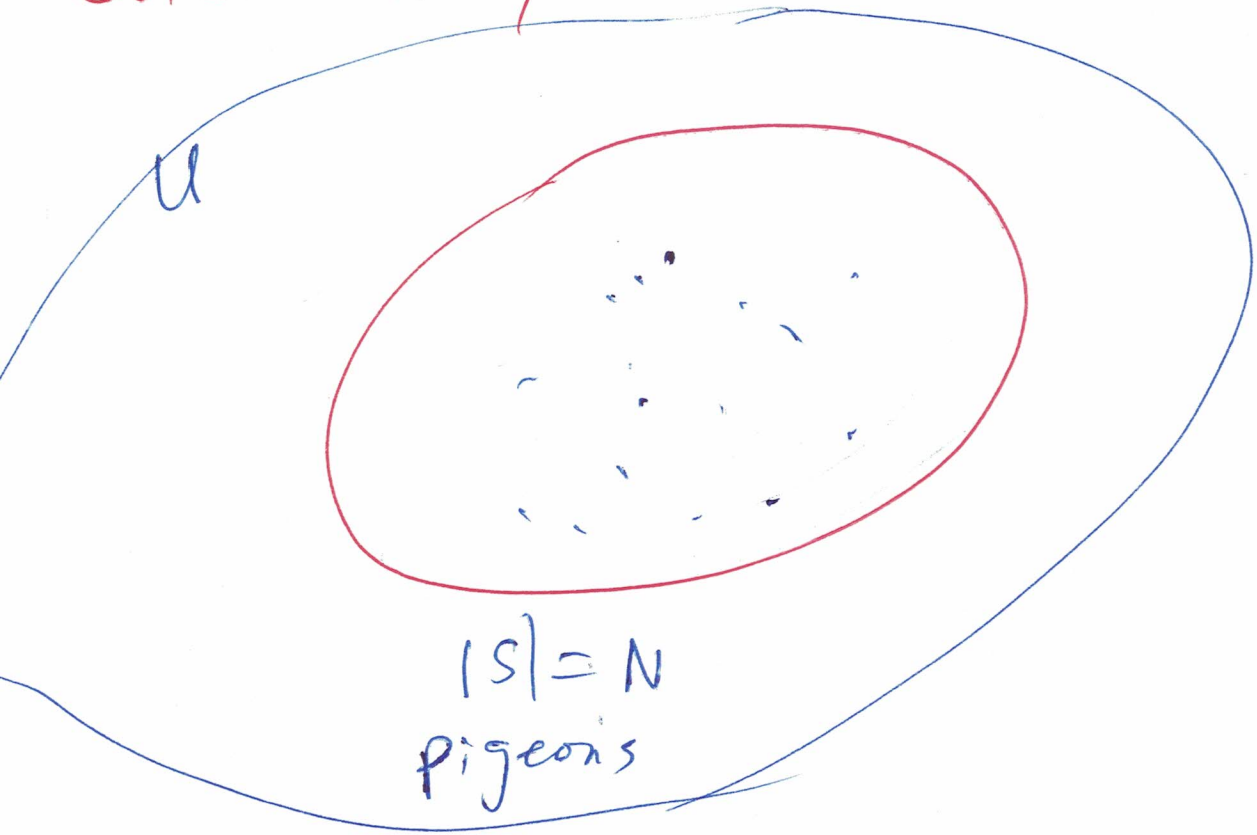
(1). Pick a random $h \in \mathcal{H}$.

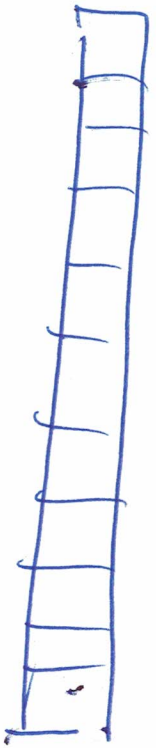
(2). Checks $\sum_{x \neq y \in S} \delta_{xy} = 0$? ($\Leftrightarrow$ Is there any collision in S wrt h)

if yes (no collision), ret h;

o.w. go back to (1).

One tricky case.





M = O(N) holes.

Confusion:

U: all 32-bit strings. // Universe size $\approx$ 4.7 billion.

For each $S \subseteq U$, with $|S| = 16$, (you pick arb. 16 32-bit strings), I wait hash those 32-bit strs into.

16 drawers.