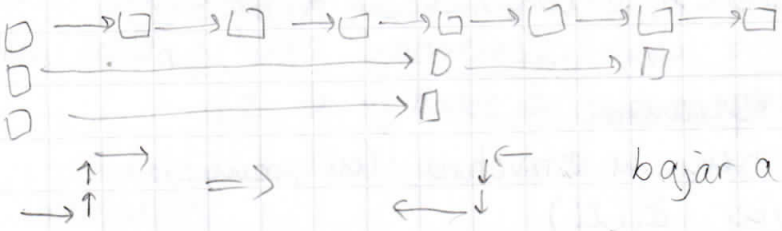
$$= \log n + \sum_{i=0}^{\infty} \frac{1}{2^i} = \log n + 2$$

★ Para ver el costo de la búsqueda, conviene ver el camino al revés



bajar a penas puedo!

Partimos en $L_1$ y nos movemos a la izq.

→ A penas encuentro un nodo que permite bajar, lo sigo
¿cuándo pasa esto? Cuando la moneda correspondiente salió cara.

La idea del camino inverso es llegar al último piso
⇒ "en cuántas monedas saco h caras?"

⇒ $\mathbb{E}[S] = \mathbb{E}[h + \sum_{r=0}^{\infty} S_r] = \mathbb{E}(h) + \sum_{r=0}^{\infty} \mathbb{E}[S_r]$ ← cuántos monedas antes de la 1ª cara

$S_r$ es tal que $\mathbb{E}(S_r) \leq 1$ y $S_r \leq \frac{n}{2^r}$ (largo esperado de la lista r.

$$\Rightarrow \mathbb{E}(S) \leq \sum_{i=0}^{\lfloor \log n \rfloor} 1 + \sum_{i=\lfloor \log n \rfloor + 1}^{\infty} \frac{n}{2^{i-1}} + \mathbb{E}(h)$$

$$= \mathbb{E}(h) + \log n + 3$$

$$= 2 \log n + 5$$

**P2** $A \; B \; C \quad AB \overset{?}{=} C$

$\Rightarrow$ Naive: mult. y comparar $\rightarrow \Theta(n^3)$

- Elegir al azar un vector $r$
- Calcular $A \cdot (B \cdot r)$ y comparar

$$\left.\begin{array}{l} \text{Si } A(Br) = Cr \Rightarrow \text{respondo sí} \\ \qquad\quad \sim \qquad\quad \text{no} \end{array}\right\} \; \Theta(n^2)$$

- Si $AB = C \Rightarrow A(Br) = Cr \quad \forall r$, $r$ es "testigo"

El algoritmo NO se va a equivocar cuando dice "NO"
$\Rightarrow$ el error es one-sided

El problema es tratar de estimar la "densidad" de buenos testigos
($\vec{r} = \vec{0}$ es el mal testigo)

elegimos el mal testigo

Mostraremos que si $AB \neq C$, $\mathbb{P}[(AB)r = Cr] \leq \frac{1}{2}$

Hay que especificar la elección de $r$:
→ se tiene un conjunto $S$, con $|S| \geq 2$
→ las componentes $r_i$ se eligen de $S$, con prob. i.i.d uniforme

idéntico dist. indep.

Sea $D = AB - C$, si $D \neq 0$, s.p.g. $d_{11} \neq 0$ (si no, permuto columnas y filas)

$$\text{Si } Dr = 0 \Rightarrow (Dr) = \sum_{i=1}^{n} d_{1i} \cdot r_i = 0$$

$$\Rightarrow r_1 = \frac{-1}{d_{11}} \left( \sum_{i=2}^{n} d_{1i} r_i \right)$$

- Para cada valor de $r_2 \dots r_n$ hay sólo 1 valor posible para $r_1$ (si fijo $r_2 \dots r_n$, hay un solo valor para $r_1$ para que sea mal testigo)
- Como $r_1$ se elige de forma uniforme, dado indep.

$$\mathbb{P}[r \text{ sea mal testigo}] \doteq \mathbb{P}[r_1 = r_1^* \mid r_2 \dots r_n = \dots] = \frac{1}{|S|} \leq \frac{1}{2}$$

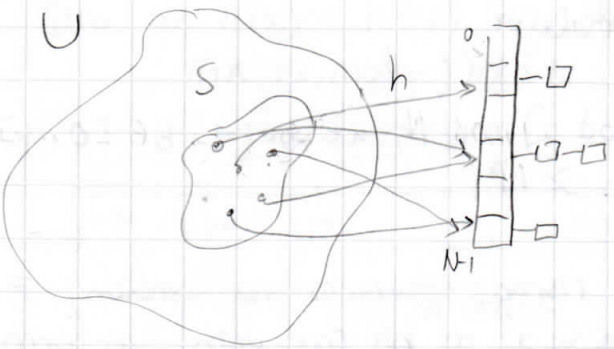
Si repito $t$ veces, la prob. de equivocarse es $\frac{1}{2^t}$

si alguna vez dice no $\Rightarrow$ no

$\sim$ $\Rightarrow$ sí

# Hashing Universal y Perfecto

2015年11月24日 (火)




S conjunto de elementos
N tamaño de la tabla
$\frac{|S|}{N}$

**Def:** Una familia $H$ de funciones de hashing es 2-universal si
$\forall\, x \neq y,\ \Pr_{h \in H}(h(x) = h(y)) \leq \frac{1}{N}$

"para todo par de eltos"

3-universal es más fuerte, sería para todo triple.

**Def:** Variable indicadora

$$C_{xy} = \begin{cases} 1 & \text{si } h(x) = h(y) \\ 0 & \text{o} \end{cases}$$

$$C_{xS} = |\{ y \in S,\ C_{xy} = 1 \}|$$

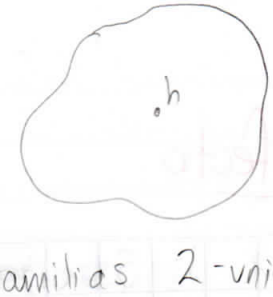
**OBS:** $C_{xS}$ es el costo de buscar la clave $x$, o insertarla

$$C_{xS} = \sum_{y \in S} C_{xy} = 1 + \sum_{\substack{y \in S \\ y \neq x}} C_{xy}$$

$$\mathbb{E}(C_{xS}) = 1 + \sum_{\substack{y \in S \\ y \neq x}} \mathbb{E}(C_{xy})$$

$$= 1 + \sum_{\substack{y \in S \\ y \neq x}} \Pr(h(x) = h(y)) \overset{H \in S \text{ 2-universal}}{<} 1 + \frac{|S|}{N}$$

∴ Usando $N = \Theta(|S|)$ celdas, el costo esperado es $\Theta(1)$




La prob de que h aleatoria, para 2 (x e y) de mi conjunto fijo de claves, cualesquiera, h se porta bien con alta prob.

Veamos algunas familias 2-universal

$$H_p = \{ h_{ab}(x) = ((ax + b) \bmod p) \bmod N,\ a \in [1 \ldots p-1],\ b \in [0 \ldots p] \}$$

donde p es un primo $> N$.

para cada a y b en sus resp. rangos tenemos una función $\neq$.
Al crear la estructura, elegimos a y b con prob. uniforme.
(el mod N es para que quepa en la tabla)
Debo hacerlo aleatorio para poder hablar de probabilidades. Dentro de H puede haber una h muy mala, pero con cierta probabilidad
Lo bacán de 2-universal es que hay alta densidad de funciones que se portan bien, que tienen pocas colisiones entre x e y.

**Teorema:** $H_p$ es 2-universal.

Dem: Fijemos $r \neq s$ y calculemos, para $x \neq y$

$$Pr(ax+b = r \bmod p \ \wedge\ ay+b = s \bmod p)$$

Si eso ocurre

$$\begin{array}{r} ax+b = r \bmod p \\ ay+b = s \\ \hline a(x-y) = r-s \bmod p \end{array}$$

$$a = \frac{(r-s)}{(x-y)} \bmod p$$

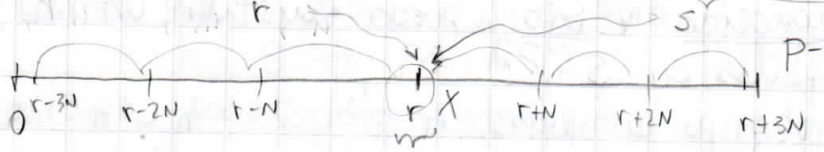
y además $b = r - ax \quad (= s - ay)$

∴ hay exactamente un valor de a y un valor de b que cumplen eso

$$Pr = \frac{1}{p(p-1)}$$

Para que $h(x) = h(y)$ necesitamos que

$$((ax+b) \bmod p) \bmod N = ((ay+b) \bmod p) \bmod N$$




este r no se cuenta pues no se puede dar $r = s$ (si no $a = 0$)

Hay a lo más $\lceil p/N \rceil - 1$ valores de $s$ que colisionan con ese valor de $r$, $r \neq s$

→ tenemos $p$ opciones para $r$

→ para cada $r$ tenemos $\lceil p/N \rceil - 1$ opciones para $s$.

→ la prob. de que se dé cada par $(r,s)$ específica es $\frac{1}{p(p-1)}$

Entonces, $\Pr(h(x) = h(y)) \leq \frac{1}{p(p-1)} \cdot \underbrace{p(\lceil p/N \rceil - 1)}_{< \frac{p(p-1)}{N}} < \frac{1}{N}$

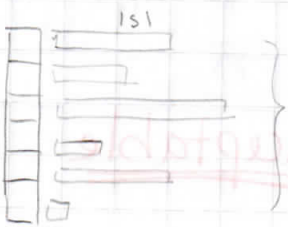
- $\lceil p/N \rceil - 1 = \left\lfloor \frac{p+N-1}{N} \right\rfloor - 1$

$\leq \frac{p+N-1}{N} - 1$

$= \frac{p-1}{N}$

* hay que escoger algún primo más grande que $N$.

Si $U$ es más grande que $|s| \cdot N$ siempre hay un $h$ que hace caer muchos en una misma celda.




siempre hay una celda con $|s|$ o más elementos colisionados