Универсальное семейство хеш-ф-ий

$\mathcal{H}$ – мн-во хеш ф-ий $h: K \to \{0 \dots (m-1)\}$

$\mathcal{H}$ – универсальное, если

$\forall k_1, k_2, k_1 \neq k_2 \quad P_{h \in \mathcal{H}}\{h(k_1) = h(k_2)\} \leq \frac{1}{m}$

Утв: Кол-во операций на безуспешный поиск в среднем $\Theta(1+\alpha)$

X_{ke} – случ. вел. $h_i(k) = h_i(\ell)$

$E_{h_i}[X_{ke}] \leq \frac{1}{m}$, $k \neq \ell$ $\quad E[X_{\ell\ell}] = 1$

$E[\text{длина цепочки}] = \sum_{k \neq \ell} E[X_{ke}] \leq \frac{1}{m} \cdot (n-1) < \alpha$.

Утв: Для успешного поиска то же самое.

$E[\text{длина цепочки}] = \sum_{k} E[X_{ke}] = 1 + \frac{n-1}{m} < 1 + \alpha$

*Построение универсального сем-ва хеш ф-ий

$\exists\, p > m$, p – простое, то семейство

$\{h_{ab}(k) = ((ak + b) \bmod p) \bmod m\}$ – универсальное

$a \in [1, p-1]$

$b \in [0, p-1]$

$ak_1 + b \bmod p = t_1$

$ak_2 + b \bmod p = t_2$

$k_1 \neq k_2 \bmod p \Rightarrow t_1 \neq t_2 \bmod p$

$(a, b) \longleftrightarrow (t_1, t_2)$

// ДЗ / Упр

$\#\{(a,b)\} = (p-1)\cdot p$

$\#\{(t_1,t_2)\} = (p-1)\,p$

(a,b) – распред. равномерно на $\{1,..p-1\}\times\{0,p-1\}$

$\Rightarrow (t_1,t_2)$ – распред. равномерно на различных парах из $\{0,..,p-1\}^2$

$$P_2[t_1 = t_2 \bmod m] \le \frac{1}{m}$$

$\exists\, t_1$ – фикс.

t_1

$$\#\{t_2, t_1 \ne t_2 \bmod m\} = \left\lceil \frac{p}{m} \right\rceil - 1 \le \frac{p+m-1}{m} - 1 = \frac{p-1}{m}$$

$$P_2[t_1 = t_2 \bmod m] \le \frac{1}{p}\cdot\frac{p-1}{m} < \frac{1}{m}$$

t_1, t_2

$t_1 \ne t_2$

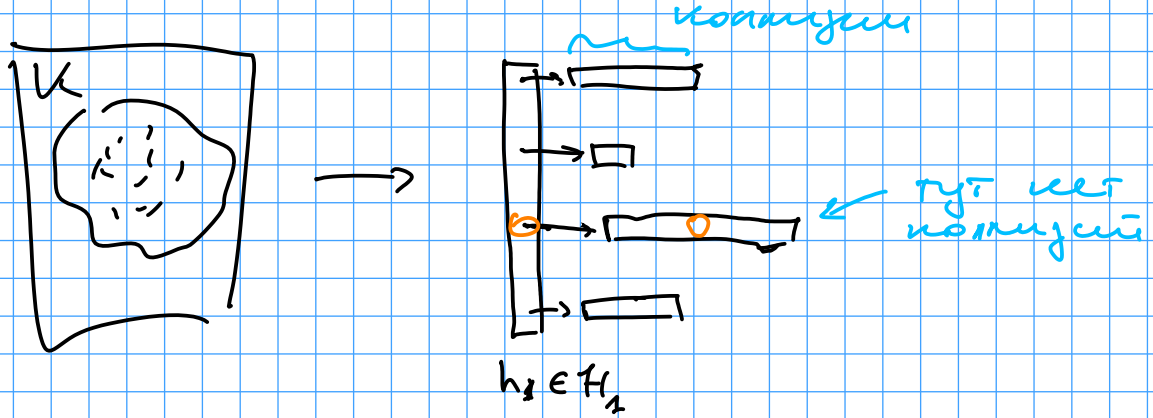
◁

Совершенное хеширование

n – # ключей

Набор ключей не изменяется.

$\exists\, H_1$ – универсальное семейство хеш ф-ий $K \to \{0,..n-1\}$



n_i – # эл-ов в ячейке i

$\forall$ ячейки i выберем сем-во $H_2^i: K \to \{0,..n_i-1\}$

$m = n^2$

Утв: Если $m = n^2 \Rightarrow$ с вер-тью $\geq \frac{1}{2}$ нет коллизий (для универс. сем-ва хеш-функций).

▷ X – случ. вел-на, равная кол-ву коллизий

$$E[X] \leq \frac{1}{m} C_n^2 = \frac{1}{m} \cdot \frac{n(n-1)}{2} = \frac{1}{n^2} \frac{n \cdot (n-1)}{2} < \frac{1}{2}$$

$$P[X \geq \lambda] \leq E[X] / \lambda$$

$$E[X] = \sum x_i P[X = x_i] = \sum_{x_i \geq \lambda} x_i P[X = x_i] + \sum_{x_i < \lambda} x_i P[X = x_i]$$

$$\geq \lambda \cdot \sum_{x_i \geq \lambda} P[X = x_i] = \lambda \cdot P[X \geq \lambda]$$

$\Rightarrow \lambda = 1 \Rightarrow P[X \geq 1] < \frac{1}{2}$ ◁

Сколько потребуется памяти?

$$\sum n_i^2 \leq \left(\sum n_i\right)^2 = n^2 \qquad n_i^2 = n_i + 2 \frac{n_i(n_i-1)}{2}$$

$$E\left[\underbrace{\sum n_i^2}_{\text{память}}\right] = E\left[\sum n_i\right] + E\left[2 \sum C_{n_i}^2\right] = \qquad 2 C_{n_i}^2$$

$$= n + 2 \cdot \underbrace{E\left[\sum C_{n_i}^2\right]}_{\#\text{ коллизий}} \leq n + \underbrace{\frac{2}{m} \cdot C_n^2}_{m = n} =$$

$$= n + \frac{2}{n} \cdot \frac{n(n-1)}{2} < 2n$$ ◁

$$P[X > 4n] \leq \frac{E[X]}{4n} < \frac{1}{2}$$