Универсальное хеширование

≡ Универсальное семейство хеш-функций $H = \{h \mid h: K \to \{0, \dots, m-1\}\}$:

$$\forall\, k_1 \neq k_2 \qquad \Pr_{h \leftarrow H}[h(k_1) = h(k_2)] \leq \frac{1}{m}$$

Утв: Сложность неуспешного поиска для хеш-таблицы с УСХФ и разрешением коллизий методом цепочек $O(1+\alpha)$.

▷ X_{kl} – случайная величина: $h(k) = h(l)$

$$E[X_{kl}] = \frac{1}{m} \qquad E[X_{kk}] = 1$$

$$E[\#\text{операций}] = 1 + E\left[\sum_{\substack{l \text{ уже в} \\ \text{таблице} \\ k \neq l \\ \text{в таблице}}} X_{kl}\right] = 1 + \sum E[X_{kl}] =$$

$$= 1 + n \cdot \frac{1}{m} = 1 + \alpha \qquad \triangle$$

Утв: Сложность успешного поиска в тех же предположениях $O(1+\alpha)$.

▷ $E[\#\text{операций}] = 1 + E\left[\sum_{\substack{k \neq l \\ \text{в таблице}}} X_{kl}\right] = 1 + \frac{n-1}{m} + 1 <$

$$< 2 + \alpha = O(1+\alpha) \qquad \triangle$$

Утв: $\exists$ p – простое число больше m, тогда

$$H = \{h_{ab}(k) = (ak + b \bmod p) \bmod m \mid a \in \{1, \dots, p-1\},\ b \in \{0, \dots, p-1\}\} - \text{УСХФ}.$$

▷ Пусть все числа из $\{0, \dots, p-1\}$

$\forall k_1, k_2 : k_1 \neq k_2$

$$\begin{cases} a k_1 + b = t_1 \\ a k_2 + b = t_2 \end{cases} \qquad k_1 \neq k_2 \Rightarrow t_1 \neq t_2$$

$$a(k_1 - k_2) = t_1 - t_2 \quad \Rightarrow \quad a = \frac{t_1 - t_2}{k_1 - k_2}$$

$$a \neq 0 \Rightarrow t_1 \neq t_2$$

По t_1, t_2 можно восст. a и b

$$b = -\frac{t_1 - t_2}{k_1 - k_2} k_1 + t_1$$

$\Rightarrow$ Существует биекция $(a, b) \longleftrightarrow (t_1, t_2)$

Т.е. $\forall$ $\{1, \dots, p-1\} \times \{0, \dots, p-1\} \longleftrightarrow \{0, \dots, p-1\}^2 \setminus \{(x, x)\}$

Т.е. $\underset{h \leftarrow H}{P}[h(k_1) = t_1, h(k_2) = t_2] = \frac{1}{p(p-1)} \quad \forall t_1 \neq t_2$

Давайте зафиксируем t_1

Сколько $\exists t_2 : t_1 = t_2 \bmod m$?

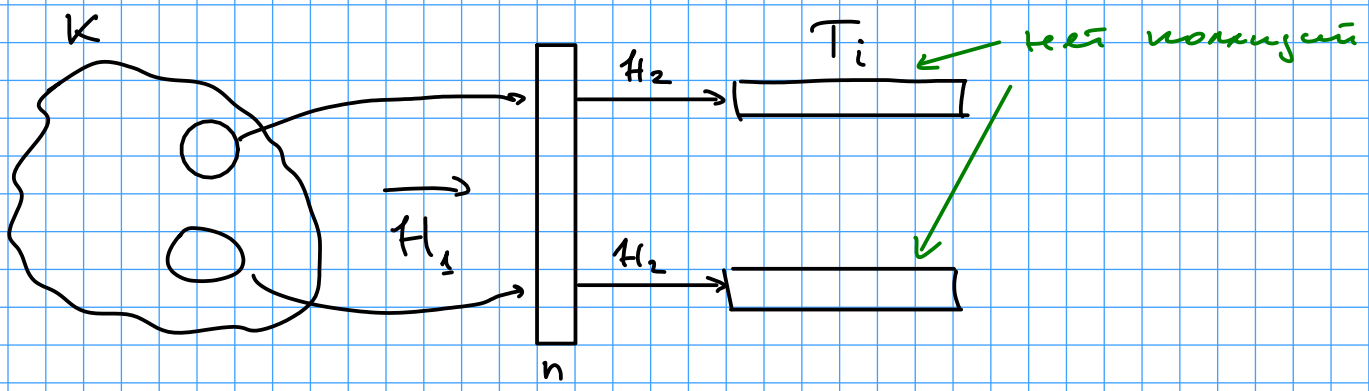
$$\#[t_2 : t_1 = t_2 \bmod m] \leq \left[\frac{p}{m}\right] - 1 \leq \frac{p + m - 1}{m} - 1 = \frac{p-1}{m}$$

$$\underset{h \leftarrow H}{P}[t_1 = t_2 \bmod m] \leq \frac{\#[t_2 : t_1 = t_2 \bmod m]}{\# t_2} =$$

$$= \frac{(p-1)/m}{p-1} = \frac{1}{m} \quad \triangle$$

Совершенное хеширование

Статическая хеш-таблица



Утв: $\exists\, m = n^2$ $P[\#\text{коллизий} \geq 1] \leq \frac{1}{2}$

(для УСХФ)

▷ $E[\#\text{коллизий}] = \sum_{k,\ell} E[X_{k\ell}] = \frac{1}{m} \cdot C_n^2 = \frac{1}{m}\frac{n\cdot(n-1)}{2} =$

кол-во пар

$= \frac{1}{n^2} \cdot \frac{n^2-n}{2} < \frac{1}{2}$

Нер-во Маркова: $P[X \geq \alpha] \leq \frac{E[X]}{\alpha}$

$P[\#\text{коллизий} \geq 1] \leq \frac{E[\#\text{коллизий}]}{1} = \frac{1}{2}$ ◁

$\exists\, n_i$ - это кол-во эл-ов в T_i

Проблема: $\sum n_i^2 = O(n^2) \Rightarrow$? много памяти ?

Утв: $P[\sum n_i^2 > 4n] \leq \frac{1}{2}$

▷ Хитрость: $n_i^2 = n_i + 2\frac{n_i\cdot(n_i-1)}{2}$

$E[\sum n_i^2] = E[\sum n_i + 2\cdot\sum \frac{n_i\cdot(n_i-1)}{2}] =$

$= n + 2\cdot\sum E[\frac{n_i(n_i-1)}{2}] = n + 2\cdot\underbrace{\sum E[C_{n_i}^2]}_{\text{общее \# коллизий}} =$

$= n + 2\cdot\frac{1}{m}\cdot C_n^2 = n + 2\cdot\frac{1}{n}\cdot\frac{n\cdot(n-1)}{2} = 2n-1 < 2n$

$$P\left[\sum n_i^2 \geq 4n\right] \leq \frac{E\left[\sum n_i^2\right]}{4n} = \frac{2n}{4n} = \frac{1}{2} \quad \triangleleft$$

Упр: Неравенство Маркова

$$P[X \geq \alpha] \leq \frac{E[X]}{\alpha}$$

$$E[X] = \sum_i p(X=i) \cdot i =$$

$$= \sum_{i<\alpha} p(X=i) \cdot i + \sum_{i \geq \alpha} p(X=i) \cdot i \geq$$

$$\geq \sum_{i \geq \alpha} p(X=i) \cdot i \geq \alpha \cdot \sum_{i \geq \alpha} p(X=i) =$$

$$= \alpha \cdot p(X \geq \alpha) \Rightarrow P(X \geq \alpha) \leq \frac{E[X]}{\alpha}$$