Universální hašovaní

23. března 2004

m je velikost hašovací tabulky
klíče k vyhovují nerovnosti $0 \leq k < K$ pro nějaké K
Zvolme prvočíslo p tak, aby bylo $K \leq p$
Pro $0 < a < p$ a $0 \leq b < p$ a $0 \leq k < p$ definujme

$$\lambda_{a,b}(k) = (ak + b) \bmod p \qquad h_{a,b}(k) = \lambda_{a,b}(k) \bmod m.$$

Dále definujme

$$\mathcal{L} = \{\lambda_{a,b} \mid 0 < a < p,\ 0 \leq b < p\}, \qquad \mathcal{H} = \{h_{a,b} \mid 0 < a < p,\ 0 \leq b < p\}$$

Pro libovolné $0 \leq i, j < p$, $i \neq j$ platí: jestliže $\lambda', \lambda'' \in \mathcal{L}$ a $\lambda' \neq \lambda''$, pak $(\lambda'(i), \lambda'(j)) \neq (\lambda''(i), \lambda''(j))$, neboli současně $\lambda'(i) \neq \lambda''(i)$ a $\lambda'(j) \neq \lambda''(j)$.
Důkaz: Předpokládejme, že $\lambda' = \lambda_{a',b'}$, $\lambda'' = \lambda_{a'',b''}$ $(\lambda'(i), \lambda'(j)) = (\lambda''(i), \lambda''(j))$.
Ppak

$$a'i + b' \equiv a''i + b'' \bmod p \quad \text{a} \quad a'j + b' \equiv a''j + b'' \bmod p,$$

$$(a' - a'')i \equiv (b'' - b') \bmod p \quad \text{a} \quad (a' - a'')j \equiv (b'' - b') \bmod p,$$

tedy

$$(a' - a'')(i - j) \equiv 0 \bmod p$$

a jelikož $i - j \neq 0 \bmod p$, musí být $a' - a'' \equiv 0 \bmod p$, protože p je prvočíslo, tedy $a' = a''$ a z toho také $b' = b''$.

Nechť $K_0, \ldots, k_{\ell-1}$ jsou navzájem různé klíče. Pro $h \in \mathcal{H}$ označme jako $\mathcal{C}(h)$ počet párů $0 \leq i, j < \ell$, $i \neq j$ takových že $h(k_i) = h(k_j)$, tedy počet konfliktů. Pak platí

$$\frac{1}{|\mathcal{H}|} \sum_{h \in \mathcal{H}} \mathcal{C}(h) \leq \frac{1}{m} \ell(\ell - 1).$$

Důkaz: Položme $c(h, i, j)$ rovno 1 pokud $h(k_i) = h(k_j)$ a rovno 0 jinak. Máme tedy odhadnout

$$\frac{1}{|\mathcal{H}|}\sum_{h\in\mathcal{H}}\sum_{j=0}^{\ell-1}\sum_{i=0;i\neq j}^{\ell-1}c(h,i,j)=\frac{1}{p(p-1)}\sum_{j=0}^{\ell-1}\sum_{i=0;i\neq j}^{\ell-1}\sum_{h\in\mathcal{H}}c(h,i,j).$$

Pro libovolné i,j, $i\neq j$ λ-obrazy dvojice (k_i,k_j) pro $\lambda\in\mathcal{L}$ pokrývají všechny uspořádané dvojice prvků y $\{0,\ldots,p-1\}$.

Kolik je dvojic $0\leq s_1,s_2<p$, $s_1\neq s_2$ takových že $s_1\equiv s_2 \bmod m$? Nechť s_1 je pevné, pak čísel $0\leq s<p$ takových že $s\neq s_1$ a $s\equiv s_1 \bmod p$ je

$$\lceil p/m\rceil-1\leq(p+m-1)/m-1=(p-1)/m,$$

tedy

$$\sum_{h\in\mathcal{H}}c(h,i,j)\leq\frac{p(p-1)}{m}$$

pro libovolné i a j a číslo, které máme odhadnout je nejvýše

$$\frac{1}{p(p-1)}\sum_{j=0}^{\ell-1}\sum_{i=0;i\neq j}^{\ell-1}\frac{p(p-1)}{m}=\frac{1}{m}\sum_{j=0}^{\ell-1}\sum_{i=0;i\neq j}^{\ell-1}1=\frac{1}{m}\ell(\ell-1).$$

Nakonec ještě spočteme pru uvedené klíče hodnotu, která je podstatná pro t.yv. perfektní párování:

$$\frac{1}{|\mathcal{H}|}\sum_{h\in\mathcal{H}}\sum_{i=0}^{m-1}|h^{-1}(i)|.$$

Tato hodnota je rovna

$$\frac{1}{|\mathcal{H}|}\sum_{h\in\mathcal{H}}\sum_{i=0}^{m-1}\big(|h^{-1}(i)|-|h^{-1}(i)|(|h^{-1}(i)|-1)\big)=$$

$$=\ell+\frac{1}{|\mathcal{H}|}\sum_{h\in\mathcal{H}}\mathcal{C}(h)=\ell+\frac{\ell(\ell-1)}{m}.$$