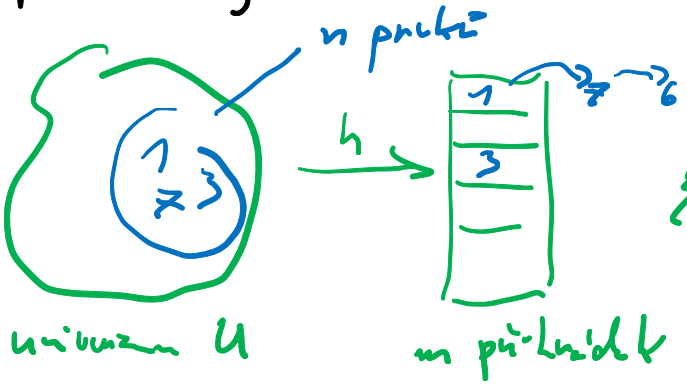
Rozptýlené (s rozptýlením)

DS pro množiny (slovník) s operacemi FIND, INSERT, DELETE – $O(1)$ očekávaný čas a $O(n)$ prostoru.



$h: U \to [m]$

$\{0,\dots,|U|-1\}$ $\{0,\dots,m-1\}$

$h(x) = h(y)$ kolize

$\alpha = n/m$ … faktor naplnění (hustota)

(očekávaná délka řetězku pokud h zcela náhodná)

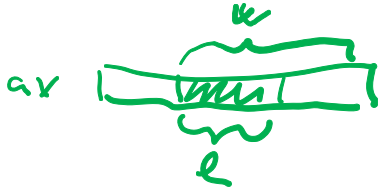
Př.:

- lineární kongruence

  $h(x) = ax \bmod m$ pro a, m nesoudělné

- vyšší bity součinu

  $h(x) = \lfloor (ax \bmod 2^w) / 2^{w-l} \rfloor$ pro a liché, $w \geq l \geq 1$



- skalární součin

  $h(\underbrace{x_1 \dots x_d}_{x \in \mathbb{Z}_m^d}) = \sum_i a_i x_i \bmod m$ pro $a \in \mathbb{Z}_m^d$, m prvočíslo

- polynom

  $h(x_1 \dots x_d) = \sum_i x_i a^i \bmod m$ pro $a \in \mathbb{Z}_m$, m prvočíslo

x náh. rovn. rozdělení $\Rightarrow$ $h(x)$ rovnoměrně rozdělení $\Rightarrow$ u/m očekávaná délka řetězu

Problém: pevná h. fce je snadno napadnutelná

$\Rightarrow$ zvol h náhodně z $\mathcal{H}$ (se silnějšími vlastnostmi)

Př. $\mathcal{H}$ = všechny fce $h: U \to [m]$ (zcela náh. fce)

$\Pr_{h \in \mathcal{H}}[h(x) = h(y)] = 1/m$ pro $x \neq y$

Otáz: potřebujeme $O(|U| \log m)$ bitů? (U bitů stačí příznakové pole)

Co je "dobrý" systém $\mathcal{H}$?

- parametrizovaný ~ $O(1)$ prostor pro h
- $O(1)$ čas pro výpočet ($O(\ell)$ pro řetězce délky ℓ)
- "chovat" stejně jako zcela náh. fce

Def: • Systém $\mathcal{H}$ fcí $h: U \to [m]$ je **c-univerzální** pro $c > 0$ pokud

$$P_{r_h}[h(x) = h(y)] \leq \frac{c}{m} \quad \text{pro lib. } x \neq y.$$

• $\mathcal{H}$ je **univerzální** pokud je c-univ. pro nějaké $c > 0$.

Věta: Nechť c-univ. systém $h: U \to [m]$, $x_1 \dots x_n, y$ navzájem různé. Pak

$$E_h[\underbrace{\#i \mid h(x_i) = h(y)}_{A}] \leq \frac{cn}{m}.$$

Dk: $A = \sum A_i$ kde $A_i = \begin{cases} 1 & \text{pokud } h(x_i) = h(y) \\ 0 & \text{jinak} \end{cases}$ (indikátory)

$$E[A_i] = P_r[A_i = 0] \cdot 0 + P_r[A_i = 1] \cdot 1 = P_r[A_i = 1] \leq \frac{c}{m}$$

$$E[A] = E[\sum A_i] = \sum_i E[A_i] \leq \frac{cn}{m} \qquad \square$$

↑ linearita střední hodnoty

Složitost hešování s řetězy

Je c_{univ}, $x_1 \dots x_n$ v tabulce, očekávaný čas:

FIND(x) ... $O(\frac{c_{univ} n}{m})$ neúspěšný, $O(\frac{c_{univ} n'}{m})$ úspěšný, n' = # prvků v tabulce po vložení x

INSERT(x) ... $O(\frac{c_{univ} n}{m})$ (bohužel nevíme, že tam není)

DELETE(x) ... $O(\frac{c_{univ} n'}{m})$

$\Rightarrow$ $O(1)$ očekávaný čas pokud $m = \Omega(n)$.

Co bohužel dopředu nevíme n?

if $\alpha = n/m > 1$ then $m' = 2m$ + přehešování

if $\alpha < 1/4$ then $m' = m/2$ + přehešování (pokud potřeba)

$\Rightarrow$ $\alpha \in [1/4, 1]$ (flexibilní pole)

$\Rightarrow$ amortizované $O(1)$ očekávaný čas

Konstrukce univ. systémů

- $h_t(x) = x \cdot t$ pro $x, t \in \mathbb{Z}_m^d$, m prvočíslo
  ($x = (x_1 \dots x_d)$, $t = (t_1 \dots t_d)$)

Věta: $\mathcal{H} = \{ h_t \mid t \in \mathbb{Z}_m^d \}$ je 1-univerzální pro lib. prvočíslo m, $d \geq 1$.

Důk: $x \neq y$ DÚNO $x_d \neq y_d$

bilinearita

$$P_t[h(x) = h(y)] = P_t[x \cdot t = y \cdot t] = P_t[(x-y)t = 0] =$$

$$= P_t\left[\sum_i (x_i - y_i) t_i = 0\right] = P_t\left[(x_d - y_d) t_d = \underbrace{-\sum_{i=1}^{d-1} (x_i - y_i) t_i}_{*}\right] = 1/m$$

($x_d - y_d \neq 0$)

volba h: postupně $t_1 \dots t_{d-1}, t_d \in \mathbb{Z}_m$

a) $* = 0$ $\quad P_t[t_d = 0] = 1/m$

b) $* \neq 0$ $\quad P_t[t_d = */(x_d - y_d)] = 1/m$ $\quad \square$

$\Rightarrow$ $\mathcal{H} = \{ h_a \mid a \in \mathbb{Z} \}$ je 1-univ. pro lib. prvočíslo m

$h_a(x) = ax \bmod m$

Co když m není prvočíslo?

$h_{a,b}(x) = ((ax+b) \bmod p) \bmod m$ $\quad$ p prvočíslo, $p \geq m$, $a,b \in \mathbb{Z}_p$ $\quad$ $2_{m \geq}$

Věta: Systém $\mathcal{H} = \{h_{a,b} \mid a,b \in \mathbb{Z}_p\}$ je 2-univ. pro lib. prvočíslo $p \geq m$.

Důk: $x \neq y$ pevné. Pro $a,b \in \mathbb{Z}_p$

$r = (ax+b) \bmod p$

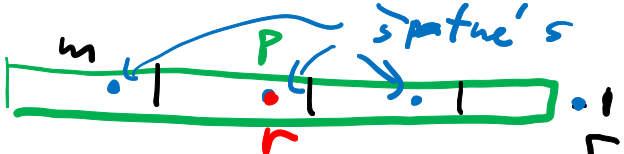
$s = (ay+b) \bmod p$

$(a,b) \mapsto (r,s)$ bijektivní

$\Rightarrow (r,s)$ rovnoměrně rozdělení

kolize $h_{a,b}(x) = h_{a,b}(y) \iff r \equiv s \bmod m$

(r,s) ... "špatná" dvojice



pro dané r je $\leq \lceil p/m \rceil$ špatných s

$$\Pr[h_{a,b}(x) = h_{a,b}(y)] \leq \frac{p\lceil p/m \rceil}{p^2} \leq \frac{p+m-1}{p \cdot m} \overset{m \leq p}{\leq} \frac{2p-1}{pm} \leq \frac{2}{m}$$

$\lceil p/m \rceil \leq \frac{p+m-1}{m}$

↳ pozn.: $(1+\varepsilon)$-univ. pro dost. velké p.

Věta: Systém $\mathcal{H} = \{ h_{a,b} \mid a,b \in \mathbb{Z}_p, a \neq 0 \}$ je 1-univ. pro lib. prvočíslo $p \geq m$.

Důk: podobně $(a,b) \mapsto (r,s)$ bij. $a \neq 0$, $r \neq s$

(r,s) je špatné: $r \equiv s \bmod m$, $r \neq s$

pro dané r je $\leq \lceil p/m \rceil - 1$ špatných s

$$\Pr[h_{a,b}(x) = h_{a,b}(y)] \leq \frac{p(\lceil p/m \rceil - 1)}{p(p-1)} \leq \frac{(p-1)/m}{p-1} = \frac{1}{m} \quad \square$$

Př: $\mathcal{H} = \{h_0, h_1\}$ $h_i(x_0 x_1) = x_i$

x	00	01	10	11
$h_0(x)$	0	0	1	1
$h_1(x)$	0	1	0	1

$$\Pr[h(x) = h(y)] = \begin{cases} 0 & x = \bar{y} \\ \frac{1}{2} & x \neq \bar{y} \end{cases} \quad (x \neq y)$$

$\Rightarrow$ $\mathcal{H}$ je 1-univ.

Alg: $\Pr[h(00) = 0] = 1$ $\quad \#(h(00)) = 0$

$\Pr[h(01) = 0 \wedge h(10) = 1] = 1$ $\quad \#(h(01)) = 0$, $\#(h(10)) = 1$

$\#(h(y)) = \begin{cases} 1 & \forall y \neq 01 \\ 0 & \end{cases}$ $\quad a,b \in \{0,1\}$

Def: Systém $\mathcal{H}$ funkcí $h: U \to [m]$ je **(k, c)-nezávislý** pro m; $k \geq 1$, $c > 0$

$$\Pr_{h \in \mathcal{H}}\left[h(x_1) = a_1 \wedge \dots \wedge h(x_k) = a_k\right] \leq \frac{c}{m^k}$$

pro lib. $x_1 \dots x_k$ různé
lib $a_1 \dots a_k$ (ne nutně různé)