Jeśli $a_0, \dots, a_{n-1}, b_0, \dots, b_{n-1} \in \mathbb{N}$. Obliczanie możemy wykonywać w $\mathbb{Z}_p$, $p \in \mathbb{P}$, gdzie p jest odpowiednio duże i w $\mathbb{Z}_p$ jest n-ty pierwotny pierwiastek z jedności, który ma własność:

$$(*) \quad \forall_{\substack{k \geq 0 \\ n \nmid k}} \sum_{j=0}^{n-1} (\omega_n^k)^j = 0$$

FAKT: Niech n, w - potęgi liczby 2 ($\neq 1$)

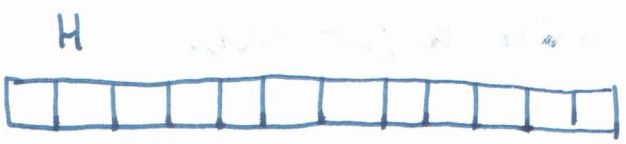
$$m = \omega^{n/2} + 1$$

Wówczas n i w są odwracalne w pierścieniu reszt mod m oraz ω jest n-tym pierwiastkiem pierwotnym z jedności spełniającym $(*)$.

Haszowanie

U - universum kluczy, $S \subseteq U$

operacje pożądane:
- insert
- delete
- search



$H[i]$ - true iff $i \in S$

nierealne, gdy $|U|$ - wielkie

$m = O(|S|)$
$\uparrow$
n

Chcemy mieć funkcję $h: U \to \{0, \dots, m-1\}$
↖ f-ga haszująca

Kolizja - sytuacja, w której $\forall_{\substack{k,l \in S \\ k \neq l}} h(k) = h(l)$

Jak radzić sobie z kolizjami:

Met. pamiętanie klucza

- nawlekanie ($H[i]$ - lista kluczy z S, dla których h przyjmuje wartości i)
- adresowanie otwarte

Chcielibyśmy, by h rozmieszczała klucze równomiernie.

Co by było, gdyby h - rozmieszczała losowo? - analogia ze schematami urnowymi?

Jaka byłaby oczekiwana długość listy (np. listy $H[0]$)? Odp: $O(n/m)$

$X_i = \begin{cases} 1, & \text{jeśli klucz wpada do listy } 0\text{-ej} \\ 0, & \text{w.p.p.} \end{cases}$

$X_1, X_2, \ldots, X_n$

$X = \sum_{i=1}^{n} X_i$ - długość listy $H[0]$, $E[X] = \sum_{i=1}^{n} E[X_i] = \frac{n}{m}$

$\frac{1}{m}$

Zatem jeśli $n \leq m$ to jest $O(1)$.

Jaka jest oczekiwana wartość długości najdłuższej listy?

Gdy $n = m$, to $\frac{\log n}{\log(\log(n))}$

Używamy f-cji pseudolosowych:

- deterministyczne
- „zachowują się" podobnie do losowych
- szybko (h w $O(1)$) wyliczalne

Przykłady tablic f-cji:

- $h(k) = k \bmod m$

  OSTROŻNOŚĆ W WYBORZE m-a ; zalecane m - liczba pierwsza oddalona od potęgi 2

- $h(k) = \lfloor m(k \cdot A - \lfloor k \cdot A \rfloor) \rfloor$, gdzie A - ustalona liczba z $(0,1)$

  dobre A: $A = \frac{\sqrt{5} - 1}{2}$, tu m może być potęgą 2-ki

Adresowanie otwarte:

- klucze pamiętamy w tablicy H
- k chcemy zapamiętać w $H[h(k)]$, a jeśli jest konflikt to stosujemy jakąś strategię znajdowania wolnej lokalizacji

  klasyczne strategie:

  - metoda liniowa

    $h(k,i) = (h'(k) + i) \bmod m$, i ozn. nr próby

    Najpierw k próbujemy wstawić w $h(k,0)$, potem $h(k,1) \ldots h(k,$

  - metoda kwadratowa

    $h(k,i) = (h'(k) + c_1 i + c_2 i^2) \bmod m$

    c_1, c_2 - pewne stałe dobrane tak, by zachodził warunek (dost): $h(k,0), h(k,1), \ldots, h(k,m-1)$ - permutacja liczb $0, \ldots, m-1$

  - podwójne haszowanie

    $h(k,i) = (h_1(k) + i\,h_2(k)) \bmod m$, h_1, h_2 - f-cje haszujące

    $\forall k$ $h_2(k)$ - względnie pierwsze z m

met. liniowa:

m permutacji wg których są próbowane lokalizacje

met. kwadratowa:

m permutacji

met haszowania podwójnego:

m^2 permutacji

Jak dobrać m?

Gdy $n \geq 2m$, to robimy przehaszowanie do dwa razy większej tablicy

Adresowanie otwarte c.d.

24.05.2018

FAKT Przy założeniu (dper)* i $\alpha = \frac{n}{m} < 1$ (współczynnik zapełnienia tablicy) oczekiwana liczba prób wykonanych przy wyszukiwaniu klucza zakończonego fiaskiem jest $< \frac{1}{1-\alpha}$

* (dper) - każda permutacja liczb $0, \ldots, m-1$ jest jednakowo ppb jako ciąg $h(k,0), h(k,1), \ldots, h(k,m-1)$

Niech p_i - ppb wykonania i nieudanych prób

Szukamy $(x) = \sum i p_i$

Niech q_i - ppb wykonania co najmniej i nieudanych prób

$$p_i = q_i - q_{i+1}$$

$$(x) = \sum_i^{\infty} i(q_i - q_{i+1}) = \sum_i^{\infty} q_i$$

$$q_1 = \frac{n}{m}, \quad q_2 = \frac{n}{m}\left(\frac{n-1}{m-1}\right) \leq \left(\frac{n}{m}\right)^2, \quad \ldots, \quad q_i \leq \left(\frac{n}{m}\right)^i = \alpha^i$$

$$(x) = \sum_i^{\infty} q_i \leq \sum_i^{\infty} \left(\frac{n}{m}\right)^i = \sum_i^{\infty} \alpha^i = \frac{1}{1-\alpha}$$

FAKT

Dla wyszukiwania zakończonego powodzeniem mamy $\leq \frac{1}{\alpha} \ln\left(\frac{1}{1-\alpha}\right) + \frac{1}{\alpha}$

Def

Niech H będzie rodziną f-cji haszujących z U w $\{0, \dots, m-1\}$. Rodzinę H nazywamy uniwersalną jeśli:

$$\forall_{\substack{x,y \in U \\ x \neq y}} \quad |\{h \in H : h(x) = h(y)\}| \leq \frac{|H|}{m}$$ (innymi słowy jest mała)

Przykłady rodzin uniwersalnych:

- Niech m, t.że $m \in \mathbb{P}$ oraz $m^{r+1} > |U|$

  $0 \leq a < m^{r+1}$ przedstawiamy w systemie m-arnym $(a_0, a_1, \dots, a_r)$

  definiujemy: (to też jest w sys m-arnym)

  $$h_a(x) = \left(\sum_{i=0}^{r} a_i x_i\right) \bmod m$$

  (d-d, że to jest r. uniwersalna pomijamy)

- Niech $p \in \mathbb{P}$, $p > |U|$, m - wielkość tablicy haszującej

  $$\forall_{\substack{a \in \mathbb{Z}_p^* \\ b \in \mathbb{Z}_p}} \quad \text{Niech } h_{a,b}(x) = ((ax+b) \bmod p) \bmod m$$

FAKT

Zbiór f-ji $H_{pm} = \{h_{ab} : a \in \mathbb{Z}_p^*, b \in \mathbb{Z}_p\}$ jest rodziną uniwersalną

D-d

$h_{a,b}(x) = ((ax+b) \bmod p) \bmod m)$, niech $h'_{ab} = (ax+b) \bmod p$

Niech k, l dowolne różne klucze $s = h'(k)$, $t = h'(l)$

Spostrzeżenie:

- $s \neq t$ $\quad s = (ak+b) \bmod p$, $t = (al+b) \bmod p$

  gdyby $s = t$, to $0 = t - s = a(k-l) \bmod p$, ale $p \nmid a$, bo $a \in \{1, \dots, p-1\}$

  oraz $p \nmid (k-l)$, bo $k \neq l$ i $p > |U|$

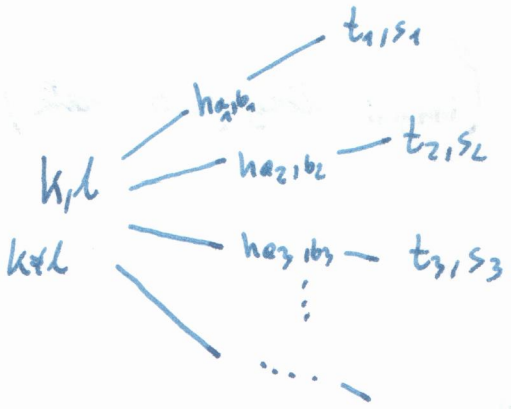
- Każda z funkcji $h'_{a,b}$ przekształca parę k,l na inną parę t,s

$$t = (ak+b) \bmod p \qquad s = (al+b) \bmod p$$

$$t - s = a(k-l) \bmod p \Rightarrow a = \left((k-l)^{-1} \bmod p\right)(t-s) \bmod p$$

$$b = (t - ak) \bmod p$$

-

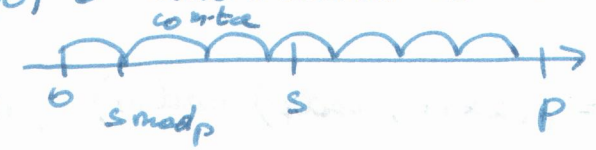


Różnych funkcji $h_{a,b}$ jest $(p-1)p$. (tyle możliwości na a, tyle na b) Różnych możliwych par jest $p(p-1)$, bo $t \neq s$ i $t,s \in \mathbb{Z}_p$

Istnieje więc bijekcja między funkcjami $h'_{a,b}$ a parami t,s (t. że $t \neq s$)

- Zatem ppb kolizji kluczy k,l przy haszowaniu jest równe ppb-stwu wylosowania pary s,t takiej, że $t \equiv s \bmod m$ spośród wszystkich par t. że $t \neq s$

Dla dowolnego (ale ustalonego) s liczb t takich że $t \equiv s \bmod m$ i $t \neq s$ jest równa $\lceil \frac{p}{m} \rceil - 1$



, a

$$\lceil \frac{p}{m} \rceil - 1 \leqslant \frac{p-1+m}{m} - 1 = \frac{p-1}{m}$$

- Ponieważ różnych t dla danego s jest $p-1$, więc ppb, że s i t kolidują jest $< \frac{1}{m}$