TD 10

1 Tables de hachage

Un dictionnaire est un annuaire dont la fréquence de modifications est très faible devant la fréquence des consultations. Par exemple, les correcteurs orthographiques des traitements de texte ont recours à un dictionnaire.

Dans la suite, on s'intéresse a des dictionnaires implementés par des tables de hachage. Dans la mesure où les données du dictionnaire sont pratiquement inchangées durant une longue période, il est intéressant de choisir la fonction de hachage parmi un ensemble de fonctions de telle sorte qu'il n'y a pas de collisions.

Supposons par la suite que les identifiants sont n valeurs numérique compris entre 0 et $p-1$, avec p premier. Nous désirons stocker ce dictionnaire dans une table de hachage à m entrées, avec $m \leq p$.

La fonction de hachage que nous recherchons se trouve parmi l'ensemble $H = \{h_{a,b} \mid 0 < a < p \wedge 0 \leq b < p\}$ où

$$h_{a,b}(x) = (((ax+b) mod p) mod m) + 1$$

Exercice 1. *1. Soient i et i' deux identifiants différents et $h_{a,b} \in H$. Montrer que $ai + b \not\equiv ai' + b(modp)$.*

2. Soient i et i' deux identifiants différents. Montrer que

$$\forall 0 \leq u \neq v \leq p \exists! h_{a,b} \in H, ai + b \equiv u(modp) \wedge ai' + b \equiv v(modp)$$

3. Soit $q = \sum_{i<i'} |\{(a,b) \mid h_{a,b}(i) = h_{a,b}(i')\}|$ le nombre total des collisions. Montrer que $q \leq \frac{n(n-1)}{2} \frac{p(p-1)}{m}$.

4. Soit $m = n^2$. Montrer que $|H'| < \frac{1}{2}|H|$, où H' est le sous-ensemble de fonctions qui produisent au moins une collision.

5. Donner un algorithme probabiliste pour trouver une fonction de hachage sans collisions. Analyser l'algorithme.

Exercice 2. *La table de hachage dans l'exercice précèdent a n^2 entrées. Nous allons obtenir un meilleur résultat avec un* double *hachage.*

Le double hachage opère de la façon suivante:

- *Chaque entrée de la table de hachage* primaire *est une référence vers une table de hachage* secondaire.
- *Chaque table de hachage secondaire possède sa fonction de hachage propre. On note h_i la fonction de hachage associée à l'entrée i de la table primaire et m_i la taille de la table de l'entrée i.*
- *Lorsqu'un nouvel identifiant id est ajouté à la table on calcule $i = h(id)$ puis $j = h_i(id)$ afin de déterminer dans quelle entrée de la table de hachage référencée depuis l'entrée i, il doit être inséré. L'insertion se fait comme pour un table de hachage simple.*
- *La recherche et la suppression suivent un schéma similaire.*

1. *Soit $H'' \subseteq H$ le sous-ensemble des fonctions de hachage qui produisent au moins n collisions. Montrer que, lorsque $m = n$, $|H''| \leq \frac{1}{2}|H|$.*
2. *On choisit une fonction $f \in H''$ comme fonction de hachage primaire. Notons n_i le nombre d'identifiants id du dictionnaire t.q. $h(id) = i$. Par construction, $\sum_i n_i(n_i - 1)/2 < n$. Comment peut-on choisir les fonctions de hachage secondaires?*
3. *Quelle est la complexité en espace de ce double hachage?*

2 Test de primalité de Miller-Rabin

Exercice 3.

1. *(Fermat) Si p est premier et que $0 < a < p$ alors*

$$a^{p-1} \equiv 1 \mod p$$

2. *Si $x^2 \equiv 1 \mod p$ et que p est premier alors*

$$x \in \{1, -1\} \mod p$$

3. *Montrer que: si n est premier et $n - 1 = 2^s t$ ($s > 1$, t impair), alors:*

$$b^t \equiv 1 \mod n$$

ou

$$\exists 0 \leq r < s, b^{2^r t} \equiv -1 \mod n$$

$\forall b \in \{1, \ldots, n-1\}$

4. *Dans la suite, on va montrer que, si n est compose, alors*

$$|B| < \frac{1}{4}|\{1, \ldots, n-1\}|$$

ou $B = \{b \in \{1 \ldots n-1\} \mid$ condition (3) est satisfaite$\}$.

Donner un algorithme probabiliste pour tester si un nombre est premier en utilisant (3). Analyser l'algorithme.

5. *Si $d = gcd(k, m)$, alors il y a d elements dans le group $\{g, g^2, \ldots, g^m = 1\}$ qui satisfont*
$$x^k = 1$$

6. *Si p est un nombre premier et que $p = 2^{s'}t'$ (t' impair), alors le nombre de $x \in \{1, \ldots, p-1\}$ qui satisfont*
$$x^{2^s t} \equiv -1 \mod p$$
(ou t est impair) est:
   - 0, *si* $r \geq s'$
   - $2^r gcd(t, t')$, *si* $r < s'$

7. *Soit n un entier compose impair. Notons avec f le nombre de b qui satisfont la condition (3). Alors $f < \frac{1}{4}$.*
   - *Supposer que $p^2|n$ pour un p premier. Utiliser (5) pour conclure.*
   - *Supposer que $n = pq$, pour p et q nombres premiers. (on suppose par la suite $p - 1 = 2^{s'}t'$ (t' impair), $p - 1 = 2^{s''}t''$ (t'' impair), $s' \leq s''$)*
     - *Montrer que $x \equiv 1 \mod pq$ ssi $x \equiv 1 \mod p$ et $x \equiv 1 \mod q$*
     - *Montrer que*
$$f \leq \frac{t't'' + t't'' + 4t't'' + 4^2t't'' + \ldots + 4^{s'-1}t't''}{n-1}$$
     - *Montrer que*
$$f \leq \frac{t't'' + t't'' + 4t't'' + 4^2t't'' + \ldots + 4^{s'-1}t't''}{2^{s'+s''}t't''} = 2^{-s'-s''}(1+\frac{4^{s'}-1}{4-1})$$
     - *Conclure pour $s' < s''$.*
     - *Si $s' = s''$, alors $gcd(t, t')gcd(t, t'') < \frac{1}{3}t't''$.*
     - *Conclure pour $s' = s''$.*
   - *Supposer que $n = p_1 p_2 \ldots p_k$ pour $k \geq 3$. Conclure.*

3 NP-completude

Exercice 4. *NP-completude du probleme de PL en nombres entiers On considere le probleme de PL en nombres entiers $Ax = b$, $x \geq 0$ ou A est une matrice $m \times n$ d'entiers et b est un m-vecteur d'entiers et la solution x doit etre un vecteur d'entiers. On fixe $a = \max_{i,j}(|a_{i,j}|, |b_i|)$.*

- *Montrer que le probleme est NP-difficile: transformer une instance de 3-SAT en une instance du probleme en remplacant chaque formule atomique A_i par deux variables entieres x_i et y_i liees par $x_i + y_i = 1$.*

- *On suppose que le probleme a une solution. Montrer alors qu'il existe une solution x telle que $\max_i x_i \leq 2n^2(ma)^{2m+1}$. Pour cela, on considere une solution avec $\max_i x_i$ minimal.*
  - *On pose $M = (ma)^m$. Conclure dans le cas $\max_i x_i \leq M$.*
  - *On suppose que les k premieres composantes de x sont plus grandes que M. Soient $v_1, \ldots, v_k$ les k premieres colonnes de A. Conclure dans le cas ou il existe des α_i entiers $\leq M$ tels que $\Sigma_i \alpha_i v_i = 0$.*
  - *Sinon, on peut extraire une matrice $k \times k$ inversible V des v_i. En resolvant le systeme ${}^tVu = 1$, on peut obtenir un vecteur ligne h de taille m. En utilisant les formules de Cramer, borner h et conclure.*
- *Montrer que le probleme est NP-complet.*