

## DM6 : Hachage parfait

L'objectif de ce sujet est d'étudier une structure de données utilisant l'aléatoire. Le contexte est le suivant : on considère un univers  $U$  fini et inclus dans  $\mathbb{N}$  dans lesquelles les clés d'une table de hachage prennent leurs valeurs. L'ensemble des clés est noté  $K \subset U$  : il est connu en amont et est de taille  $n$ . On souhaite construire une table de hachage dont les clés seront celles de  $K$  et garantissant que le nombre d'accès mémoire lors d'une recherche dans cette table se fait en temps constant : on parle alors de hachage parfait.

### Partie 1 Classe de fonctions de hachage universelle

Soit  $E$  un ensemble fini de fonctions de hachage de  $U$  dans  $\llbracket 0, m-1 \rrbracket$  pour  $m \in \mathbb{N}^*$ . On dit que  $E$  est une *classe de fonctions de hachage universelle* si, lorsqu'on tire  $h$  uniformément aléatoirement dans  $E$ , la probabilité que deux clés différentes entrent en collision via  $h$  est inférieure à  $1/m$  ; autrement dit :

$$\text{Pour toutes clés } l, k \text{ telles que } l \neq k, \mathbb{P}(h(l) = h(k)) \leq \frac{1}{m}$$

Soit  $p$  un nombre premier strictement plus grand que  $|U|$  et  $m$  un entier. Pour tout  $a \in (\mathbb{Z}/p\mathbb{Z})^*$  et tout  $b \in \mathbb{Z}/p\mathbb{Z}$ , on introduit la fonction  $h_{ab}$  suivante (où  $U$  est considéré en tant que sous ensemble de  $\mathbb{Z}/p\mathbb{Z}$ ) :

$$h_{ab} : \begin{cases} U \longrightarrow \llbracket 0, m-1 \rrbracket \\ k \longmapsto ((ak + b) \bmod p) \bmod m \end{cases}$$

On note  $\mathcal{H}_{pm} = \{h_{ab} \mid a \in (\mathbb{Z}/p\mathbb{Z})^*, b \in \mathbb{Z}/p\mathbb{Z}\}$ .

L'objectif de cette partie est de montrer que  $\mathcal{H}_{pm}$  est une classe de fonctions de hachage universelle. Dans les questions suivantes, on a fixé un entier premier  $p$  plus grand que  $|U|$  et un entier  $m$  ainsi que deux clés  $k, l \in \mathbb{Z}/p\mathbb{Z}$  différentes.

1. Si  $a \in (\mathbb{Z}/p\mathbb{Z})^*$  et  $b \in \mathbb{Z}/p\mathbb{Z}$ , montrer que  $r = (ak + b) \bmod p$  et  $s = (al + b) \bmod p$  sont différents.

2. Montrer que la fonction suivante est une bijection bien définie :

$$\begin{cases} \{(a, b) \mid a \in (\mathbb{Z}/p\mathbb{Z})^*, b \in \mathbb{Z}/p\mathbb{Z}\} & \longrightarrow \{(r, s) \in (\mathbb{Z}/p\mathbb{Z})^2 \mid r \neq s\} \\ (a, b) & \longmapsto ((ak + b) \bmod p, (al + b) \bmod p) \end{cases}$$

3. Montrer qu'étant donné  $r \in \mathbb{Z}/p\mathbb{Z}$ , le nombre d'éléments  $s \in \mathbb{Z}/p\mathbb{Z}$  tels que  $r \neq s$  et  $r \equiv s \pmod{m}$  est inférieur à  $(p-1)/m$ .

4. Dédire des questions précédentes que la classe de fonctions  $\mathcal{H}_{pm}$  est universelle.

### Partie 2 Hachage sur un niveau

Dans cette partie, on suppose que les  $n$  clés sont hachées dans une table de taille  $m = n^2$  à l'aide d'une fonction de hachage  $h$  choisie uniformément aléatoirement dans une classe de fonctions de hachage universelle à valeurs dans  $\llbracket 0, m-1 \rrbracket$  (par exemple,  $\mathcal{H}_{pm}$  pour un certain  $p \in \mathbb{P}$ ). On note  $X$  la variable aléatoire comptant le nombre de collisions dans cette table (c'est-à-dire le nombre de paires de clés  $\{k, l\}$  tels que  $h(k) = h(l)$ ).

5. Majorer l'espérance de  $X$  et en déduire que la probabilité d'avoir au moins une collision dans cette table est strictement inférieure à  $1/2$ .

6. A l'aide du résultat précédent, expliquer comment obtenir une table de taille quadratique en le nombre de clés dans laquelle il est garanti qu'il n'y ait pas de collision.

### Partie 3 Hachage sur deux niveaux

La table construite à la question 6 garantit qu'il n'y a pas de collision mais il est regrettable que sa taille soit quadratique en le nombre de clés. On cherche dans cette partie à construire une table sans collisions mais de taille linéaire en  $n$  plutôt que quadratique.

Pour ce faire, on commence par hacher les  $n$  clés dans une table principale de taille  $m = n$  via une fonction  $h$ . Puis, dans chaque case de cette table, on construit une table de hachage secondaire de taille  $m_i = n_i^2$  via une fonction de hachage  $h_i$  où  $n_i$  est le nombre de clés qui sont hachées dans la case  $i$  par  $h$  au premier niveau.

7. Montrer que, si on hache  $n$  clés dans une table de taille  $m = n$  via  $h$  choisie uniformément aléatoirement dans une classe de fonctions de hachage universelle, alors  $\mathbb{E} \left[ \sum_{i=0}^{n-1} n_i^2 \right] < 2n$  où  $n_i$  est la variable aléatoire comptant le nombre de clés hachées dans la case  $i$  par  $h$ .

*Indication : pour tout  $k \in \mathbb{N}^*$ ,  $k^2 = k + 2\binom{k}{2}$ .*

8. Montrer qu'avec la stratégie consistant à hacher sur deux niveaux, la probabilité que l'espace total occupé par l'ensemble des tables secondaires dépasse  $4n$  est strictement inférieur à  $1/2$ .
9. En déduire comment construire une table de hachage sans collision de taille linéaire en le nombre de clés à hacher.