# Hachage parfait

Antoine Dequay

21 septembre 2022

## Notes

— Prof : Loïc Hélouët.

— Leçon : 901, 921, 932.

— Référence :

  — Cormen.

On se donne un ensemble de clés fixé et on cherche à construire un hachage parfait, c'est à dire un hachage tel qu'une recherche se fasse en $O(1)$ en accès mémoire.

Pour cela, on va utiliser une table de hachage à 2 niveaux, avec un hachage universel à chaque niveau :

— Le premier niveau est un hachage par chaînage : on hache $n$ clés dans $m$ cases grâce à une fonction de hachage $h$ à choisir dans une famille de fonctions de hachage universelles.

— Au lieu de créer des listes chaînées des éléments hachés, on utilise une seconde table de hachage pour chaque case $j$, associée à une fonction de hachage $h_j$ et à l'ensemble $S_j$.

On cherche à choisir convenablement $h_j$ pour ne pas avoir de collisions au second niveau. Pour cela, on supposera que la taille $m_j$ de la table $S_j$ est $n_j^2$, où $n_j = \#S_j$.

En notant $\mathcal{H}_{p,m} = \left\{h_{a,b} : k \longmapsto ((ak+b) \mod p) \mod m,\ a \in \mathbb{Z}_p^* \text{ et } b \in \mathbb{Z}_p\right\}$ pour $p$ un nombre premier plus grand que les valeurs des clés, on peut prouver que $\mathcal{H}_{p,m}$ est un ensemble de tables universelles. On choisit $h \in \mathcal{H}_{p,m}$ et $h_j \in \mathcal{H}_{p,m_j}$.

Montrons qu'il n'y a pas de collision au second niveau :

**Théorème 1** Supposons qu'on stocke $n$ clés dans une table de hachage de taille $m = n^2$ par une fonction de hachage $h$ choisie aléatoirement dans une classe de fonctions de hachage universelles. Alors la probabilité qu'il y ait une ou plusieurs collisions est de moins de $\frac{1}{2}$.

*Preuve.* On a $\binom{n}{2}$ paires de clés qui peuvent entrer en collision, et chaque paire entre en collision avec une probabilité $\frac{1}{m}$.

Si on note $X$ la variable aléatoire comptant le nombre de collisions, on a :

$$\mathbb{E}(X) = \binom{n}{2}\frac{1}{n^2} = \frac{n(n-1)}{2n^2} < \frac{1}{2}.$$

Par inégalité de Markov, il vient bien :

$$\mathbb{P}(\{X \geq 1\}) \leq \mathbb{E}(X) < \frac{1}{2}.$$

□

**Théorème 2** Supposons qu'on stocke $n$ clés dans une table de hachage de taille $m = n$ par une fonction de hachage choisie aléatoirement dans une classe de fonctions de hachage universelles. Alors :

$$\mathbb{E}\left(\sum_{j=0}^{m-1} N_j^2\right) < 2n,$$

où $N_j$ est la variable aléatoire qui compte le nombre de clés hachées en case $j$.

*Preuve.* Avec l'identité $a^2 = a + 2\binom{a}{2}$ :

$$\begin{aligned}\mathbb{E}\left(\sum_{j=0}^{m-1} N_j^2\right) &= \mathbb{E}\left(\sum_{j=0}^{m-1} N_j + 2\binom{N_j}{2}\right) = \mathbb{E}\left(\sum_{j=0}^{m-1} N_j\right) + 2\mathbb{E}\left(\sum_{j=0}^{m-1}\binom{N_j}{2}\right) \\ &= n + 2\mathbb{E}\left(\sum_{j=0}^{m-1}\binom{N_j}{2}\right)\end{aligned}$$

La quantité $\mathbb{E}\left(\sum_{j=0}^{m-1}\binom{N_j}{2}\right)$ représente le nombre total de paires de clés qui entrent en collision. Par propriété du hachage universel, on a :

$$\mathbb{E}\left(\sum_{j=0}^{m-1}\binom{N_j}{2}\right) \leq \frac{1}{m}\binom{n}{2} = \frac{n-1}{2},$$

d'où :

$$\mathbb{E}\left(\sum_{j=0}^{m-1} N_j^2\right) \leq n + 2\frac{n-1}{2} = 2n - 1 < 2n.$$

□

**Corollaire 3** Supposons qu'on stocke $n$ clés dans une table de hachage de taille $m = n$ par un fonction de hachage choisie aléatoirement dans une classe de fonction de hachage universelles et qu'on choisit les tailles des tables de hachage secondaires comme $m_j = n_j^2$. Alors l'espace mémoire utilisé pour toute les tables secondaires dans le hachage parfait est en moyenne inférieur à $2n$.

*Preuve.* Le résultat est immédiat car $m_j = n_j^2$.

□

**Corollaire 4** Supposons qu'on stocke $n$ clés dans une table de hachage de taille $m = n$ par une fonction de hachage choisie aléatoirement dans une classe de fonction de hachage universelles et qu'on choisit les tailles des tables de hachage secondaires comme $m_j = n_j^2$. Alors la probabilité que l'espace mémoire utilisé pour les tables secondaires soit supérieur ou égale à $4n$ est de moins de $\frac{1}{2}$.

*Preuve.* Le corollaire précédent permet de montrer, via l'inégalité de Markov :

$$\mathbb{P}\left(\left\{\sum_{j=0}^{m-1} m_j \geq 4n\right\}\right) \leq \frac{\mathbb{E}\left(\sum_{j=0}^{m-1} m_j\right)}{4n} < \frac{2n}{4n} = \frac{1}{2}.$$

□

Ainsi, en testant des fonctions de hachage aléatoirement dans une famille universelle, on en trouvera "rapidement" une qui utilise un espace de stockage "raisonnable".