# UNE CLASSE UNIVERSELLE DE FONCTIONS DE HACHAGE

Soient

- $p$ un nombre premier assez grand tq toute clé $k$ est dans $\{0, \dots, p-1\}$ et $p > m$.
- $\mathbb{Z}_p = \{0, 1, \dots, p-1\}$
- $\mathbb{Z}_p^* = \mathbb{Z}_p \setminus \{0\}$

Pour $a \in \mathbb{Z}_p^*$ et $b \in \mathbb{Z}_p$, soit

$h_{a,b} : \mathbb{Z}_p \to \mathbb{Z}_m$ avec

$$h_{a,b}(k) := (ak + b \pmod p) \pmod m$$

Cela nous permet de définir la classe suivante de fonctions de hachage

$$\mathcal{H}_{p,m} := \{ h_{a,b} \mid a \in \mathbb{Z}_p^*, b \in \mathbb{Z}_p \}$$

Au total $\mathcal{H}_{p,m}$ contient $p(p-1)$ fonctions de hachage

**Théorème :** La classe $\mathcal{H}_{p,m}$ de fonctions de hachage est universelle.

**Démo :** Soient $k, \ell \in \mathbb{Z}_p$ des clés distinctes tels que $k \neq \ell$.
Pour $h_{a,b} \in \mathcal{H}_{p,m}$ soit

$$r = ak + b \pmod p$$
$$s = a\ell + b \pmod p$$

On a $r - s = a(k - \ell) \pmod p$ donc $k \neq \ell$ et le fait que $p$ est un nombre premier impliquent que $r \neq s$.

Chacun des $p(p-1)$ choix de $a \in \mathbb{Z}_p^*$ et $b \in \mathbb{Z}_p$ donne une paire résultante $(r,s)$ différente avec $r \neq s$ :
Soient $(a_1, b_1), (a_2, b_2) \in \mathbb{Z}_p^* \times \mathbb{Z}_p$ distinctes et supposons que

$$a_1 k + b_1 = a_2 k + b_2 \pmod p \text{ et } a_1 \ell + b_1 = a_2 \ell + b_2 \pmod p$$

Si $a_1 = a_2$ alors $b_1 = b_2$, contradiction. Si $a_1 \neq a_2$ alors $a_1(k-\ell) = a_2(k-\ell)$ ce qui implique $k = \ell$, contradiction

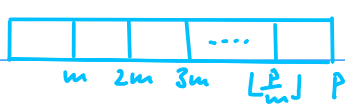
Comme il n'y a que $p(p-1)$ possibles avec $r \neq s$ il existe donc une bijection entre paires $(a,b)$ avec $a \neq 0$ et $(r,s)$ avec $r \neq s$.
Donc, pour toute paire donnée de clés $k \neq \ell$, si l'on choisit $(a,b)$ aléatoirement de manière uniforme dans $\mathbb{Z}_p^* \times \mathbb{Z}_p$ alors la paire résultante $(r,s)$ a la même probabilité d'être l'une quelconque de paires de valeurs distinctes mod $p$.

On a donc $\Pr[k \text{ et } \ell \text{ entrent en collision}] = \Pr[r \equiv s \pmod m]$ quand $r$ et $s$ sont choisis aléatoirement comme valeurs distinctes mod $p$.

Pour $r \in \mathbb{Z}_p$, le nombre de valeurs $s \in \mathbb{Z}_p$ tq $r \neq s$ et

$r - s \equiv 0 \pmod m$

$r \equiv s \pmod m$ est au plus $\lceil p/m \rceil - 1$ :




il y a au plus une valeur $s$ : $r - s \equiv 0 \pmod p$ dans chaque cellule sauf la première

Ensuite $\lceil p/m \rceil - 1 \le (p+m-1)/m - 1$
$= (p-1)/m$

Pour $r \in \mathbb{Z}_p$ on a $\Pr[\, r \equiv s \pmod m\,] \le \underbrace{((p-1)/m)}_{\#\text{ de collisions possibles}} / \overbrace{(p-1)}^{p-1 \text{ valeurs de } s \text{ tq } s \neq r} = \frac{1}{m}$

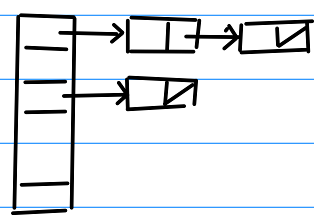
En conclusion, pour tout $k, \ell \in \mathbb{Z}_p$ distincts on a
$\Pr[h_{a,b}(k) = h_{a,b}(\ell)] \le \frac{1}{m}$

$\square$

## HACHAGE PARFAIT

Idée : Pour un ensemble de clés statique on utilise une stratégie de hachage à deux niveaux (hachage universel à chaque niveau) :

<u>$1^{\text{er}}$ étape</u> :
hachage universel avec listes chaînées



<u>$2^{\text{ème}}$ étape</u> :
transformer pour $0 \le j \le m-1$ la liste chaînée de l'alvéole $j$ dans un table de hachage secondaire $S_j$

En choisissant avec soin les fonctions de hachage pour $S_j$ on peut garantir qu'il n'y aura pas de collisions au niveau secondaire. Pour éviter des collisions à ce niveau-là, la taille $m_j$ de $S_j$ doit être le carré du nombre $n_j$ de clés hachées vers l'alvéole $j$ au premier niveau.

**Théorème** : Si l'on stocke $n$ clés dans une table de hachage de taille $m = n^2$ via une fonction de hachage $h$ choisie aléatoirement dans une famille universelle $\mathcal{H}$, alors la probabilité d'avoir une collision est $\le \frac{1}{2}$.

**Démo :** Pour $n$ clés distinctes il y a $\binom{n}{2}$ paires de clés susceptibles d'entrer en collision, chacune avec proba $\leq \frac{1}{m}$.
Quand $m = n^2$ le nombre attendu de collisions est

$X = \#$ de collisions

$$E[X] \leq \binom{n}{2} \cdot \frac{1}{m} = \frac{n^2 - n}{2} \cdot \frac{1}{n^2} < \frac{1}{2} \qquad (*)$$

Soit $I_{X \geq 1} = \begin{cases} 1 & \text{si } X \geq 1 \\ 0 & \text{sinon} \end{cases}$

$I_{X \geq 1}$ variable indicatrice

$I_{X \geq 1} \leq X$

$(*)$

On a $\Pr[I_{X \geq 1} = 1] = E[I_{X \geq 1}] \leq E[X] < \frac{1}{2}$

$\square$