

Corrigé DM6

1. Par l'absurde, si on avait $r = s$ alors on aurait $0 = r - s = a(k - l) \pmod p$. L'élément a étant inversible modulo p , on en déduit que p divise $k - l$, autrement dit que $k = l$ dans $\mathbb{Z}/p\mathbb{Z}$ ce qui n'est pas par hypothèse.
2. Notons F cette fonction, D son ensemble de départ et A son ensemble d'arrivée. La question précédente montre qu'elle est bien définie puisqu'en effet son ensemble d'arrivée est constitué d'éléments différents de $\mathbb{Z}/p\mathbb{Z}$.

Comme p est premier, le cardinal de $(\mathbb{Z}/p\mathbb{Z})^*$ vaut $p - 1$ et on en déduit que $|D| = p(p - 1)$. D'autre part, pour chaque élément r de $\mathbb{Z}/p\mathbb{Z}$, il y a $p - 1$ éléments de $\mathbb{Z}/p\mathbb{Z}$ différents de r donc $|A| = p(p - 1)$.

En outre, F est surjective. En effet, soit $(r, s) \in A$. Observons

$$\begin{aligned} a &= (r - s)(k - l)^{-1} \pmod p \\ b &= (r - ak) \pmod p \end{aligned}$$

Comme $k - l \neq 0 \pmod p$, cet élément est inversible dans $\mathbb{Z}/p\mathbb{Z}$ donc a est bien défini. De plus, comme $r \neq s$, $a \neq 0 \pmod p$ et est par conséquent inversible dans $\mathbb{Z}/p\mathbb{Z}$. On a donc $(a, b) \in (\mathbb{Z}/p\mathbb{Z})^* \times \mathbb{Z}/p\mathbb{Z}$ et immédiatement $F(a, b) = (r, s)$. Donc (r, s) admet un antécédent par F dans D .

Les deux points précédents suffisent à conclure que F est bijective.

3. Il y a au plus $\lceil p/m \rceil$ éléments s égaux à r modulo m . Comme l'un d'entre eux est exactement r , le nombre recherché est inférieur à

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

La première inégalité vient du fait que pour tout $a, b \in \mathbb{N}^*$, $\lceil a/b \rceil \leq (a + b - 1)/b$. En effet dans ce contexte, par division euclidienne il existe $q \in \mathbb{N}$ et $r \in \llbracket 0, b - 1 \rrbracket$ tels que $a = bq + r$ et donc $\lceil a/b \rceil = a/b + (b - r)/b$. Si $r = 0$, l'inégalité est acquise et sinon $r \geq 1$ donc $(a + b - r)/b \leq (a + b - 1)/b$.

4. En reprenant les notations de la question 2, choisir h uniformément dans \mathcal{H}_{pm} revient à choisir (a, b) uniformément dans A . D'après la question 2, les probabilités que $(r, s) = F(a, b)$ vaille d sont les mêmes pour tout $d \in D$. Or, la probabilité de collision entre k et l est la exactement la probabilité d'avoir r et s congrus modulo m . On a donc

$$\mathbb{P}(h(k) = h(l)) = \frac{\text{nombre d'éléments } (r, s) \in D \text{ congrus modulus } m}{|D|} \leq \frac{p \frac{(p-1)}{m}}{p(p-1)} = \frac{1}{m}$$

5. On a $X = \sum_{k \neq l} \mathbb{1}_{\{h(k)=h(l)\}}$ donc $\mathbb{E}[X] = \sum_{k \neq l} \mathbb{P}(h(k) = h(l))$. Comme h est une fonction de hachage universelle, pour tout couple de clés différentes (k, l) , on a $\mathbb{P}(h(l) = h(k)) \leq 1/m$. Il ne reste donc plus qu'à compter le nombre de paires de clés : il y en a $\binom{n}{2}$. On en déduit que :

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

L'inégalité de Markov assure donc que $\mathbb{P}(X \geq 1) \leq \mathbb{E}[X]/1 < 1/2$ ce qui conclut.

6. Pour ce faire, il suffit de tirer uniformément h dans une classe de fonctions de hachage universelle jusqu'à ce qu'on en obtienne une pour laquelle il n'y a pas de collision dans notre table de taille n^2 . La question 5 montre que la probabilité de ne pas avoir trouvé de fonction h convenable après k essais indépendants est inférieure à $(1/2)^k$.

7. On a :

$$\begin{aligned}\mathbb{E} \left[\sum_{i=0}^{n-1} n_i^2 \right] &= \mathbb{E} \left[\sum_{i=0}^{n-1} \left(n_i + 2 \binom{n_i}{2} \right) \right] \text{ par l'indication} \\ &= \mathbb{E} \left[\sum_{i=0}^{n-1} n_i \right] + 2 \mathbb{E} \left[\sum_{i=0}^{n-1} \binom{n_i}{2} \right] \text{ par linéarité de l'espérance}\end{aligned}$$

Le premier terme de cette somme vaut n . En effet, peu importe la façon dont sont réparties les n clés au premier niveau de hachage, il y en a forcément n donc $\sum_{i=0}^{n-1} n_i = n$. D'autre part, $\sum_{i=0}^{n-1} \binom{n_i}{2}$ est exactement la variable aléatoire qui compte le nombre de collisions dans la table principale (celle calculée par h). D'après la question 5, on en déduit que :

$$\mathbb{E} \left[\sum_{i=0}^{n-1} \binom{n_i}{2} \right] \leq \frac{1}{m} \binom{n}{2} = \frac{1}{n} \frac{n(n-1)}{2} = \frac{n-1}{2}$$

puis que :

$$\mathbb{E} \left[\sum_{i=0}^{n-1} n_i^2 \right] \leq n + n - 1 < 2n$$

8. L'espace total occupé par les tables secondaires est précisément $\sum_{i=0}^{n-1} n_i^2$. La question précédente montre que la quantité moyenne de mémoire utilisée par toutes les tables secondaires est inférieure à $2n$. En appliquant l'inégalité de Markov, on a :

$$\mathbb{P} \left(\sum_{i=0}^{n-1} n_i^2 \geq 4n \right) \leq \frac{\mathbb{E} \left[\sum_{i=0}^{n-1} n_i^2 \right]}{4n} < \frac{1}{2}$$

9. Pour obtenir une table de hachage stockant les n clés de K avec garantie que la taille de la table est linéaire en n et est sans collisions on peut hacher sur deux niveaux selon la méthode décrite dans cette partie en choisissant les fonctions de hachage comme suit :

- Tant que $\sum_{i=0}^{n-1} n_i^2 \geq 4n$ avec n_i le nombre de clés hachées dans la case i par h , choisir uniformément une fonction h dans une classe de fonctions de hachage universelle comme \mathcal{H}_{pn} . D'après la question 8, la probabilité d'échouer à trouver une h convenable au bout de k essais indépendants est inférieure à $(1/2)^k$.
- Pour chaque $i \in \llbracket 0, n-1 \rrbracket$, déterminer une fonction de hachage h_i garantissant qu'il n'y a pas de collision dans la i -ème table secondaire de taille n_i^2 en suivant la méthode de la question 6.