Devoir maison numéro 1
Correction

Exercice 1 Recherche de collisions

La fonction de hachage h devrait donner pour un fichier $\mathbf{f}$ un nombre aléatoire $h(\mathbf{f})$ entre 0 et $2^{128}-1$. Le problème de collisions est dans ce cas exactement le problème de l'anniversaire. La probabilité qu'il n'y ait pas de collision parmi k fichiers générés est (on note $N = 2^{128}$)

$$\bar{p}(k) = \frac{N(N-1)\dots(N-k+1)}{N^k} \tag{1}$$

$$= \prod_{i=0}^{k-1}(1-\frac{i}{N}) \tag{2}$$

$$\leqslant \exp\left(-\frac{k(k-1)}{2N}\right) \tag{3}$$

Une condition suffisante pour que $\bar{p}(k) \leqslant \frac{1}{2}$ est

$$k^2 - k \geqslant 2N\ln 2 \tag{4}$$

i.e.

$$k \geqslant \frac{1+\sqrt{1+8N\ln 2}}{2} \sim \sqrt{2\ln 2}\times 2^{64} \tag{5}$$

Exercice 2 Randomized Quicksort

1. Si $\frac{1}{3}|Y| \leqslant \mathrm{rg}(p) \leqslant \frac{2}{3}|Y|$ alors le choix de p est chanceux, donc la probabilité qu'un choix de pivot soit chanceux est $\geqslant \frac{1}{3}$.
2. On utilise Chernoff II avec $X_i \sim \mathcal{B}(p_i)$, $X_N = \sum_i X_i$. Par la question 1, $p_i \geqslant \frac{1}{3}$ donc $\mathbb{E}[X_N] \geqslant \frac{N}{3}$. Aussi, $\mathbf{P}\left(X_N \leqslant \frac{N}{3}(1-\epsilon)\right) \leqslant \mathbf{P}(X_N \leqslant \mathbb{E}[X_N](1-\epsilon))$.

$$\mathbf{P}\left(X_N \leqslant (1-\epsilon)\mathbb{E}[X_N]\right) \leqslant \exp\left(-\frac{\epsilon^2}{2}\mathbb{E}[X_N]\right) \leqslant \exp\left(-\epsilon^2\frac{N}{6}\right) \tag{6}$$

   Pour retrouver $\mathbf{P}(X_N \leqslant \frac{N}{4})$, on pose $\epsilon = \frac{1}{4}$ et on trouve $c = \frac{1}{6\times 16} = \frac{1}{96}$.
3. Soit $x \in X$. Soit Y_k^x l'ensemble contenant x à l'étape k. $Y_0^x = Y$.

$$|Y_k^x| \leqslant \frac{2}{3}|Y_{k-1}^x| \tag{7}$$

   Donc,

$$|Y_k^x| \leqslant \left(\frac{2}{3}\right)^k n \tag{8}$$

   Mais $|Y_k^x| \geqslant 1$ donc $\left(\frac{2}{3}\right)^k n \geqslant 1$:

$$k \leqslant \frac{\ln n}{\ln\frac{3}{2}} \tag{9}$$

   $C = \frac{1}{\ln\frac{3}{2}} \leqslant 3$
4. Soit $x \in X$. On applique la question 2 avec $N = kC\ln n$ avec k une constante $\geqslant 4$. Soit O_x la variable aléatoire qui note le nombre de comparaisons nécessaires pour trier x. A chaque choix de pivot, on fait n comparaisons pour placer le pivot. Le nombre de comparaisons totales pour trier X est borné par $n\max_x O_x$. Or, si $X_N > C\ln n$ alors x est trié en au plus N pivots (question 3). Donc, $\forall x$, $\mathbf{P}(O_x \leqslant N) \geqslant \mathbf{P}(X_N > C\ln n)$, et on obtient

$$\mathbf{P}(O_x > N) = 1 - \mathbf{P}(O_x \leqslant N) \leqslant \mathbf{P}(X_N \leqslant C\ln n) \tag{10}$$

Il suit

$$\begin{aligned}
\mathbf{P}(\max_x O_x > kC\ln n) &\leqslant n\mathbf{P}(O_x > kC\ln n) && \text{(Borne de l'union)} && (11)\\
&\leqslant n\mathbf{P}(X_N \leqslant C\ln n) && \text{Equation 10} && (12)\\
&= n\mathbf{P}\left(X_N \leqslant \frac{N}{k}\right) && && (13)\\
&\leqslant n\mathbf{P}\left(X_N \leqslant \frac{N}{4}\right) && \frac{N}{k} \leqslant \frac{N}{4} && (14)\\
&\leqslant n\exp(-kcC\ln n) && && (15)
\end{aligned}$$

Pour $k = 100$, $100cC = \frac{100}{96\ln 3/2} > 1$ Donc le nombre total d'opérations est inférieur à $100Cn\ln n$, i.e. est en $\mathcal{O}(n\ln n)$ avec grande probabilité.

Exercice 3 Routage de paquets par un algorithme probabiliste

1. On considère n pair. Pour tout $x \in \{0,1\}^n$, on écrit $x = (x^1, x^2)$ avec $x^1, x^2 \in \{0,1\}^{n/2}$. On considère une permutation $\pi : 0,1^n \to \{0,1\}^n$ qui vérifie $\pi((x^1,0)) = (0,x^1)$. Ces $2^{n/2}$ chemins passent tous par le sommet $(0,0)$. En particulier, pour tous les x^1 ayant 1 comme premier bit, on a $2^{(n/2)-1}$ chemins et chacun passe par l'arrête qui relie 0_n à $(0_{n/2}, 1, 0_{(n/2)-1})$. L'arrête ne pouvant en laissant passer qu'un seul à chaque étape, il faut au moins $2^{(n/2)-1}$ à l'algorithme pour finir.
2. Soit e l'arrête orienté qui relie $(e_0,\dots,e_{i-1},0,e_{i+1},\dots,e_n)$ à $(e_0,\dots,e_{i-1},1,e_{i+1},\dots,e_n)$. L'itinéraire naïf de y à $\tau(y)$ passe par e si et seulement si

$$\begin{aligned}
e_1,\dots,e_{i-1} = \tau(y)_1,\dots,\tau(y)_{i-1} && (16)\\
y_{i+1},\dots,y_n = e_{i+1},\dots,e_n && (17)\\
y_i = 0, \tau(y)_i = 1 && (18)
\end{aligned}$$

On note Y_e l'ensemble des éléments $y \in V \setminus x$ vérifiant $y_{i+1},\dots,y_n = e_{i+1},\dots,e_n$ et $y_i = 0$. On a $|Y_e| = 2^{i-1}$. De plus, pour $y \in Y_e$, $\mathbf{P}(y \in H_e) = \mathbf{P}(\tau(y)_i = 1, \tau(y)_1 = e_1,\dots,\tau(x)_{i-1} = e_{i-1}) = \frac{1}{2^i}$.

$$\mathbf{E}\left[|H_e|\right] = \mathbf{E}\left[\sum_{y\in V} \mathbb{1}(x \in H_e)\right] = \sum_{y\in V}\mathbf{E}\left[\mathbb{1}(x\in H_e)\right] \leqslant \sum_{y\in Y_e}\mathbf{P}(x\in H_e) = 2^{i-1}\frac{1}{2^i} = \frac{1}{2}$$

3. Pour tout $x \in V$, on écrit P_x l'ensemble des arrêtes de l'itinéraire naïf de x à $\tau(x)$. On a $|K_x| \leqslant \sum_{e\in P(x)}|He|$. donc $\mathbf{E}[|K_x|] \leqslant \mathbf{E}[\sum_{e\in P(x)}|He|] = \sum_{e\in P(x)}\mathbf{E}[|He|] \leqslant n$ car un itinéraire naïf est au maximum de longueur n. De plus, en écrivant $I_{x,y}$ l'indicatrice de l'évènement "P_x s'intersecte avec P_y", les $I_{x,y}$ sont des variable de Bernouilli indépendantes et $K_x = \sum_{y\in V\setminus x} I_{x,y}$. On peut donc appliquer l'inégalité de Chernoff II avec la majoration (cf. partie bonus).

$$\mathbf{P}(|K_x| \geqslant (1+4)n) \leqslant \exp\left(\frac{-4^2}{2+4}n\right) \leqslant \exp(-2n) \leqslant 2^{-2n} \quad (19)$$

4. $\mathbf{P}(\exists x \in V : |K_x| \geqslant 5n) \leqslant \sum_{x\in V}\mathbf{P}(|K_x| \geqslant 5n) \leqslant 2^n 2^{-2n} = 2^{-n}$. Ainsi, avec grande probabilité $(1-2^n)$, et en utilisant le lemme admis de l'énoncé, il faut moins de $5n$ étapes pour envoyer tous les paquets de x à $\tau(x)$. De même, toujours avec grande probabilité, il faut moins de $5n$ étapes pour envoyer tous les paquets de $\tau(x)$ à $\pi(x)$. Ainsi, avec grande probabilité, il faut moins de $10n$ étapes pour envoyer tous les paquets de x à $\pi(x)$.
5. On considère $n = 4p$ un multiple de 4. On considère X l'ensemble des sommets de "poids" p (le poids est le nombre de 1). On a $|X| = \binom{4p}{p}$. On considère une permutation qui associe chaque élément de X à un autre, de support disjoint (le support est l'ensemble des indices où les bits valent 1). Pour un élément de X, la probabilité que de passer par le sommet $v = (0,\dots,0)$ est de $q = \frac{1}{\binom{2p}{p}}$. Le nombre d'élément passant par v suit donc une loi $\mathcal{B}(|X|,q)$, et son espérance est donc $\mu_n = \frac{\binom{4p}{p}}{\binom{2p}{p}} \sim a\left(\frac{256}{108}\right)^p > \exp(p)$ pour n assez grand (a est une constante). Donc $\mu_n \geqslant \exp((1/4)n)$ pour n assez grand. Donc en particulier, en appliquant Chernoff II, $P(I_n < \mu_n/2) \leqslant \exp(-\mu_n/12)$. Donc avec grande probabilité, il y a plus de $\exp(n/4)/2$ paquets passant par v. Le sommet n'ayant que n arrête, il faudra au moins $\exp(cn)/2n = \Omega(2^{c'n})$ étapes où $c = \frac{1}{4}$ et $c' < c$.

BONUS : Chernoff II avec majoration de l'espérance

Supposons $\mu = \mathbb{E}[X] \leqslant m$. Montrons que $\mathbf{P}(X \geqslant (1+\varepsilon)m) \leqslant \exp\left(-\frac{\varepsilon^2}{2+\varepsilon}m\right)$.

Pour $t > 0$ à déterminer, on a

$$\mathbf{P}(X \geqslant (1+\varepsilon)m) = \mathbf{P}(e^{tX} \geqslant e^{t(1+\varepsilon)m}) \leqslant e^{-t(1+\varepsilon)m} M_X(t).$$

On a $M_{X_i}(t) = (1-p_i) + p_i e^t = 1 + p_i(e^t - 1) \leqslant \exp(p_i(e^t-1))$, et donc par indépendance

$$M_X(t) = \prod M_{X_i}(t) \leqslant \exp(\mu(e^t - 1)) \leqslant \exp(m(e^t-1)).$$

On choisit maintenant la valeur $t_1 = \ln(1+\varepsilon)$ pour obtenir

$$\mathbf{P}(X \geqslant (1+\varepsilon)) \leqslant \exp(m(e^{t_1}-1) - (1+\varepsilon)t_1 m) = \left(\frac{e^\varepsilon}{(1+\varepsilon)^{1+\varepsilon}}\right)^m.$$

Le reste de la preuve est identique au cours.