Curiosités

Olivier

Août 2025

Table des matières

1 bonjour

Définition 1.1. *On pose $\mathcal{A}$ un ensemble. $\mathcal{A}$ sera appelé* ***alphabet*** *et les éléments de $\mathcal{A}$ seront appelés* ***caractères*** *ou* ***lettres****.*

On veut envoyer une suite finie de liens de vidéos YouTube. Pour cela, il suffit de n'envoyer que les IDs de chaque vidéo puisqu'on peut retrouver chaque lien à partir de l'ID correspondant.
Définition 1.2. *Un* ***ID*** *est une suite de 11 caractères.*

Objectif : Envoyer le moins de caractères possible.

On suppose que les IDs sont composées de caractères aléatoires de $\mathcal{A}$ choisis uniformément.

Soit $N \in \mathbb{N}^*$ le nombre de vidéos que l'on souhaite envoyer. Soit $i_1, \ldots, i_N \in \mathcal{A}^{11}$ les IDs des vidéos que l'on souhaite envoyer.

1.1 Coller tous les caractères

Une première approche est de simplement concaténer les IDs les uns à la suite des autres. On aura alors $11N$ caractères au total

1.2 Encodage et décodage en binaire

On observe que $\#\mathcal{A} = 64 = 2^6$ est une puissance de 2. Alors on peut encoder sans "prendre plus de place" chaque caractère de chaque ID en binaire sur 6 bits.

On définit la fonction $\text{bin}_6 : \mathcal{A} \to \{0,1\}^6$ qui convertit un caractère en binaire en utilisant une bijection $\mathcal{A} \to [\![0, 63]\!]$ (par la définition de la finitude de $\mathcal{A}$)

On obtient une nouvelle chaîne de caractères :

$$B_{i_1,\ldots,i_N} := \text{bin}_6(i_1) \ldots \text{bin}_6(i_N)$$

où $\text{bin}_6 : \mathcal{A}^{11} \to \{0,1\}^6$ est défini **abusivement** par :

$$\text{bin}_6 : \begin{cases} \mathcal{A}^{11} \to (\{0,1\}^6)^{11} \\ c = (c_1, \ldots, c_{11}) \mapsto (\text{bin}_6(c_1) \ldots \text{bin}_6(c_{11})) \end{cases}$$

Et alors, $B_{i_1,\ldots,i_N}$ a une longueur de $6 \times 11 \times N = 66N$.

Puis on peut faire des "paquets" de 8 bits sur $B_{i_1,\ldots,i_N}$. Et ré-encoder chaque "paquet" en un de 256 différents choisis.

Puisqu'il est probable (si N n'est pas un multiple de 4) que $66N$ ne soit pas un multiple de 8, on complète le dernier "paquet" par des 0 pour arriver à une longueur multiple de 8.

Concrètement, on va ajouter $\text{offset} := \begin{cases} (66N \mod 8) + 1 & \text{si } 8 \nmid 66N \\ 0 \quad \text{sinon} \end{cases}$ zéros.

et on va écrire offset, puis les caractères encodés depuis 8 bits pour avoir une nouvelle chaîne de caractères.

Plaçons-nous dans le cas où $8 \mid 66N$. Alors la chaîne de caractères finale commencera par un 0 et aura $\frac{66N}{8} = \frac{33N}{4} \in \mathbb{N}$ paquets. Donc $1 + \frac{33N}{4}$ caractères au total.

Et on a le rapport : (on omet le caractère de décalage)

$$\frac{\{\text{longueur\_finale}\}}{\{\text{longueur\_initiale}\}} = \frac{\frac{33N}{4}}{11N} = \frac{3}{4}$$

Donc asymptotiquement ($N \to +\infty$) le ratio tend vers $\frac{3}{4}$.

On remarque que le ratio ne dépend pas des bases de décodage/ré-encodage choisies.

Plus généralement, en notant $k_1, k_2 \in \mathbb{N}^*$ le nombre de bits de décodage et de ré-encodage respectivement, le ratio asymptotique est $\frac{k_1}{k_2}$.