Projet M1 OC 2025 - Video Streaming

Bornes Supérieures

EP 26/03/2025

Simplifications/aménagements de la modélisation MIP retenue :

Inputs :

- V : ensemble de vidéos
- D : Data Center
- E : ensemble de endpoints
- C : ensemble de cash servers
- R : ensemble de requêtes

- Attributs vidéos :
  - s_v : durée de la vidéo v∈ V (ms)
- Attributs endpoints :
  - l_e^D : latence de transfert d'une vidéo entre le data center D et le endpoint e∈ E (ms)
  - C_e : liste de cash centers connectés au endpoint e∈ E

Hypothèse : les listes de cash centers C_e incluent virtuellement le data center de numéro $D = |C| + 1$

- Attribut cash servers :
  - l_e^c : latence de transfert d'une vidéo entre le cash center c∈ C et le endpoint e∈ E (ms)
  - Q : capacité de stockage d'un cash center
- Attributs requêtes :
  - e_r : endpoint associé à la requête r∈ R
  - v_r : vidéo associée à la requête r∈ R
  - n_r : nombre de fois où la requête r∈ R est activée

Variables de décision :

- x_{vc} = 1 si la vidéo v∈ V est hébergée sur le cash center c∈ C, 0 sinon
- y_{rc} = 1 si la requête r∈ R est desservie par le cash center $c \in C_{e_r}$, 0 sinon.

Remarque : les variables u_r deviennent inutiles au vu de l'aménagement des ensemble C_e précisé ci-dessus.

Simplifications additionnelles :
- Les temps de transfert étant exprimés en micro secondes, on occulte la partie entière inférieure dans l'objectif
- Les contraintes [5] du modèle initial sont purement et simplement relaxées
- L'objectif de [VS] est ré-exprimé en minimisation pour se ramener au contexte du développement théorique donné dans le cours.

Modèle révisé :

$$[VS]: \min -K\sum_{r\in R} n_r \sum_{c\in C_{e_r}} (l_{e_r}^D - l_{e_r}^c)\, y_{rc}$$

$$Sc\begin{cases} [1]\ \forall c\in C, \sum_{v\in V} s_v x_{vc} \le Q \\ [2]\ \forall r\in R, \sum_{c\in C_{e_r}} y_{rc} = 1 \\ [3]\ \forall r\in R, \forall c\in C_{e_r}, y_{rc} \le x_{v_r c} \\ [4]\ \forall v\in V, \forall c\in C, x_{vc}\in\{0,1\} \\ [5]\ \forall r\in R, \forall c\in C_{e_r}, y_{rc}\in\{0,1\} \end{cases}$$

avec : $K = \dfrac{1000}{\sum_{r\in R} n_r}$

Borne Supérieures triviale UB_1 :

Comme suggéré dans l'heuristique duale en 2 phases proposée dans le cadre de l'étape précédente de ce projet, une solution (non réalisable, mais très optimiste) peut être générée en affectant chaque requête au cash center le plus favorable en termes de latence de transfert en occultant les contraintes de capacité portant sur ces cash centers. Notons c(r) le cash center desservant la requête $r\in R$. Rappelons que, par convention, c(r) = D signifie que la requête r est assurée directement à partir du data center D. Ce sera le cas notamment si $C_{e_r} = \varnothing$. On pose ainsi :

$$\forall r\in R, c(r) = \arg\min_{c\in C_{e_r}} l_{e_r}^c$$

Le saving de latence de transfert associé à la requête r est alors : $\pi_r = l_{e_r}^D - l_{e_r}^{c(r)}$

On aboutit alors à l'obtention d'une borne supérieure triviale : $UB_1 = K\sum_{r\in R} n_r \pi_r$

Borne Lagrangienne UB_2 :

Associons à chaque contrainte couplante $[3]\ \forall r\in R, \forall c\in C_{e_r}, y_{rc}\le x_{v_r c}$ un multiplicateur de Lagrange $\lambda_{rc}\ge 0$.

On obtient ainsi, en les dualisant, une décomposition lagrangienne dont la fonction de Lagrange s'écrit :

$$\Im(x,y,u,\lambda) = -K\sum_{r\in R} n_r \sum_{c\in C_{e_r}} (l_{e_r}^D - l_{e_r}^c)\, y_{rc} + \sum_{r\in R}\sum_{c\in C_{e_r}} \lambda_{rc}(y_{rc} - x_{v_r c})$$

$$\Im(x,y,u,\lambda) = \sum_{r\in R}\sum_{c\in C_{e_r}} -K\left[n_r(l_{e_r}^D - l_{e_r}^c) + \lambda_{rc}\right] y_{rc} - \sum_{r\in R}\sum_{c\in C_{e_r}} \lambda_{rc} x_{v_r c}$$

Pour une distribution λ fixée des multiplicateurs de Lagrange, on aboutit à deux sous-problèmes lagrangiens dissociés car indépendants :

Sous-problème Lagrangien 1 :

$$[SL_1(\lambda)]: \min \sum_{r\in R}\sum_{c\in C_{e_r}} \rho_{rc} y_{rc}$$

$$Sc\begin{cases} [2]\ \forall r\in R, \sum_{c\in C_{e_r}} y_{rc} = 1 \\ [5]\ \forall r\in R, \forall c\in C_{e_r}, y_{rc}\in\{0,1\} \end{cases}$$

avec : $\rho_{rc} = -K\left[n_r(l_{e_r}^D - l_{e_r}^c) + \lambda_{rc}\right]$

Ce sous-problème ne porte que sur les variables y_{rc}. Sa solution optimale, notée $\left[\overline{y}_{rc}(\lambda)\right]_{r\in R, c\in C}$, est obtenue par :

$$\forall r\in R, \overline{y}_{rc}(\lambda) = \begin{cases} 1\ pour\ \overline{c} = \arg\min_{c\in C_{e_r}} \rho_{rc} \\ 0 \text{ sinon} \end{cases}$$

Sous-problème Lagrangien 2 :

$$\left[SL_2(\lambda)\right] : \max \sum_{r\in R} \sum_{c\in C_{e_r}} \lambda_{rc} x_{v_r c}$$

$$Sc \begin{cases} [1] \ \forall c \in C, \sum_{v\in V} s_v x_{vc} \leq Q \\ [4] \ \forall v \in V, \forall c \in C, x_{vc} \in \{0,1\} \end{cases}$$

Ce sous-problème ne porte que sur les variables x_{vc}. Sa solution optimale, notée $\left[\overline{x}_{v_r c}(\lambda)\right]_{r\in R, c\in C_{e_r}}$ est obtenue en résolvant les $|C|$ problèmes de knapsack (indépendants) par Programmation Dynamique :

$$\left[KSP(c)\right] : \max \sum_{r\in R} \lambda_{rc} x_{v_r c}$$

$$Sc \begin{cases} [1'] \ \sum_{r\in R} s_{v_r} x_{v_r c} \leq Q \\ [4'] \ \forall r \in R, x_{v_r c} \in \{0,1\} \end{cases}$$

La valeur de la fonction duale en λ est alors donnée par :

$$L(\lambda) = Opt[SL_1(\lambda)] + Opt[SL_2(\lambda)] = K \sum_{r\in R} \sum_{c\in C_{e_r}} \rho_{rc} \overline{y}_{rc}(\lambda) + \sum_{c\in C} \sum_{r\in R} \lambda_{rc} \overline{x}_{v_r c}(\lambda)$$

L'objectif est alors de déterminer la distribution de multiplicateurs de Lagrange minimisant $L(\lambda)$, ce qui revient à résoudre le problème dual : $L(\lambda^*) = \min_{\lambda\geq 0} L(\lambda)$. Une procédure de sous-gradient est tout à fait adaptée à ce but. A titre indicatif, le sketch de l'algorithme s'y rapportant est rappelé ci-après :

RELAX-LAGR[]

Iter = 0 *// Compteur d'itérations à 0 (méthode de sous-gradient) /*

Init(α,ρ) *// Initialisation des coefficients de relaxation (α =0,9,ρ =2) p.e.)*

$\lambda = 0$ // Initialisation des multiplicateurs de Lagrange/

$L^* = +\infty$ // Initialisation de la borne lagrangienne

Tant que iter < itmax faire

Résolution du sous-problème Lagrangien $\left[SL_1(\lambda)\right]$ *:* $\rightarrow \left[\overline{y}_{rc}(\lambda)\right]_{r\in R, c\in C}$

$\rho_{rc} = -K\left[n_r(l^D_{e_r} - l^c_{e_r}) + \lambda_r\right]$ */ Calcul des coûts régularisés*

$$\left[SL_1(\gamma)\right] : \min \sum_{r\in R}\sum_{c\in C_{e_r}} \rho_{rc}\, y_{rc}$$

$$Sc\begin{cases}[2]\ \forall r\in R, \sum_{c\in C_{e_r}} y_{rc} = 1 \\ [5]\ \forall r\in R, \forall c\in C_{e_r}, y_{rc}\in\{0,1\}\end{cases} \qquad \rightarrow \forall r\in R, \overline{y}_{rc}(\lambda) = \begin{cases}1\ pour\ \overline{c} = \arg\min_{c\in C_{e_r}} \rho_{rc} \\ 0 \text{ sinon}\end{cases}$$

Résolution du sous-problème Lagrangien $\left[SL_2(\lambda)\right]$ *:* $\rightarrow \left[\overline{x}_{v_r c}(\lambda)\right]_{r\in R, c\in C_{e_r}}$

Résoudre chaque problème de knapsack $\left[KSP(c)\right]_{c\in C}$ indépendamment :

$$\left[KSP(c)\right] : \max \sum_{r\in R} \lambda_r x_{v_r c}$$

$$Sc\begin{cases}[1']\ \sum_{r\in R} s_{v_r} x_{v_r c} \le Q \\ [4']\ \forall r\in R, x_{v_r c}\in\{0,1\}\end{cases}$$

Calcul de la fonction duale :

$$L(\lambda) = \sum_{r\in R}\sum_{c\in C_{e_r}} \rho_{rc}\overline{y}_{rc}(\lambda) + \sum_{c\in C}\sum_{r\in R} \gamma_r \overline{x}_{v_r c}(\lambda)$$

Sauvegarde de l'optimum dual courant :

Si $L(\lambda) < L^*$ Alors

$L^* = L(\lambda)$, $\lambda^* = \lambda$

FinSi

Calcul des composantes du sous-gradient de L au point γ :

$\forall r\in R, \forall c\in C_{e_r},\ \gamma_{rc} = y_{rc} - x_{v_r c}$

Mise à jour des multiplicateurs de Lagrange :

$$\forall \mathrm{r}\in \mathrm{R}, \forall \mathrm{c}\in \mathrm{C}_{\mathrm{e_r}}, \lambda_{\mathrm{rc}} = \lambda_{\mathrm{rc}} + \rho \frac{[\mathrm{LB} - \mathrm{L}(\lambda)]}{\sum_{\mathrm{r}\in\mathrm{R}}\sum_{\mathrm{c}\in\mathrm{C_{e_r}}} \gamma^2_{\mathrm{rc}}} \gamma_{\mathrm{rc}},$$

Iter=iter+1 *// Incrémentation du compteur d'itérations /*

Révision de ρ : *$\rho = \alpha\rho$ // Révision coefficient de relaxation /*

Finttq