Un algorithme polynomial 2-approché pour le problème de dimensionnement des mémoires avec débit maximum intrinsèque

Olivier MARCHETTI & Alix MUNIER-KORDON
Université Pierre et Marie Curie
GdR ASR-MAO
Jeudi 3 Avril 2008 – ENS LYON

PLAN

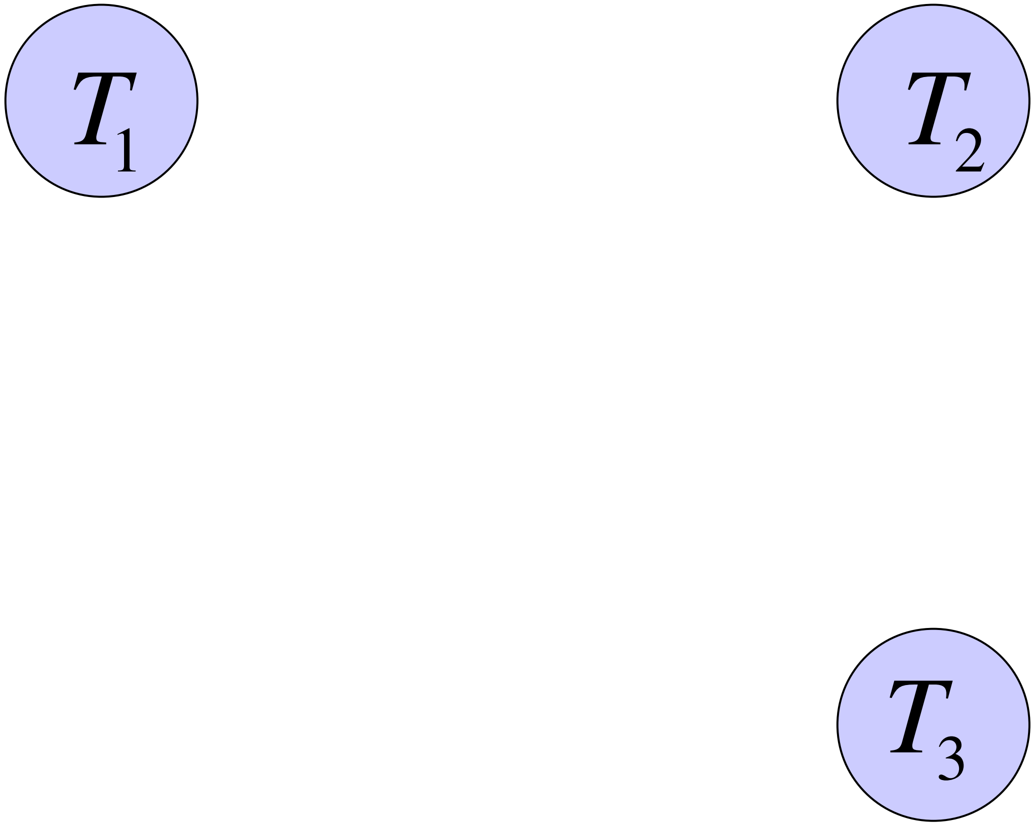
- Présentation & Contexte
- Notations
- Résultats de complexité
- Complexité de Débit Intrinsèque Maximum (DMI)
- Un algorithme 2-approché pour DMI

Présentation & Contexte

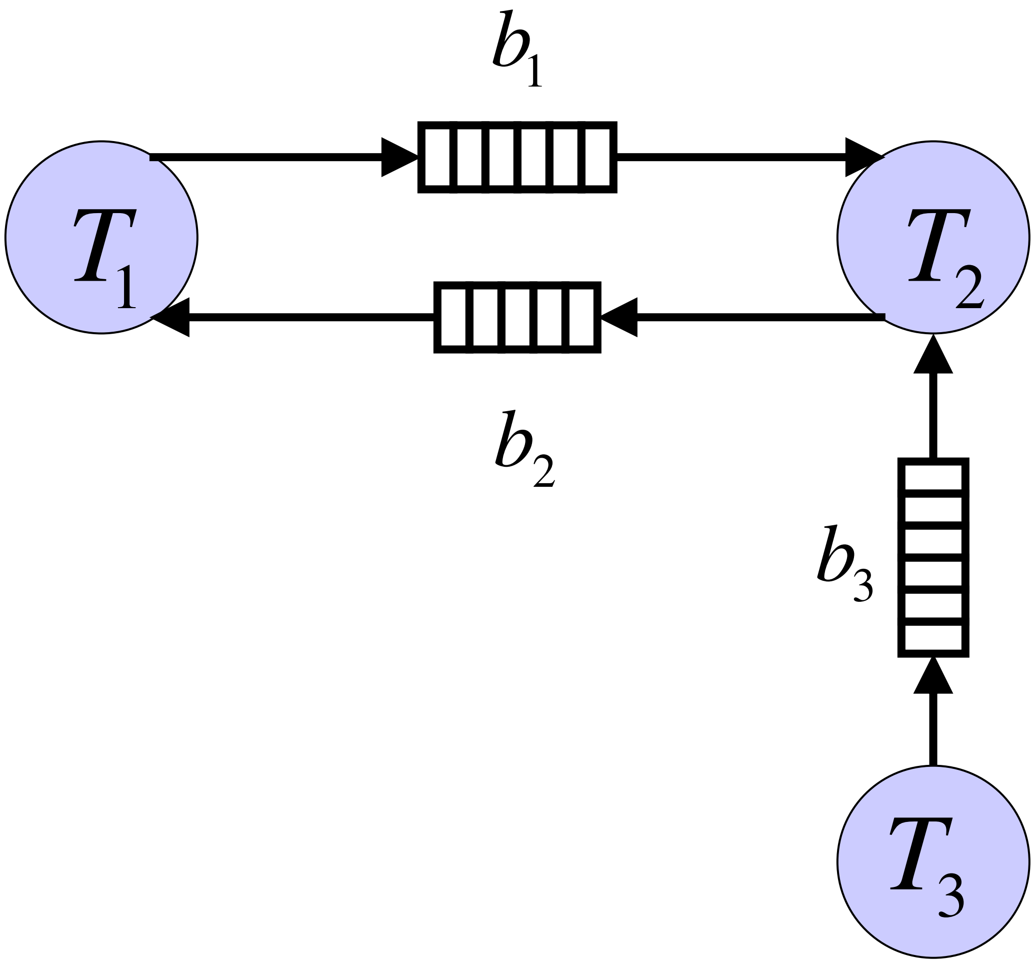
Présentation du problème - Introduction

- Tâches : programmes asynchrones.
- Communiquent avec buffers.
- Opérations séquentielles de lecture/écriture.
- Capacité d'un buffer : nombre de données pouvant y être stockées.
- Pour un buffer : données homogènes.

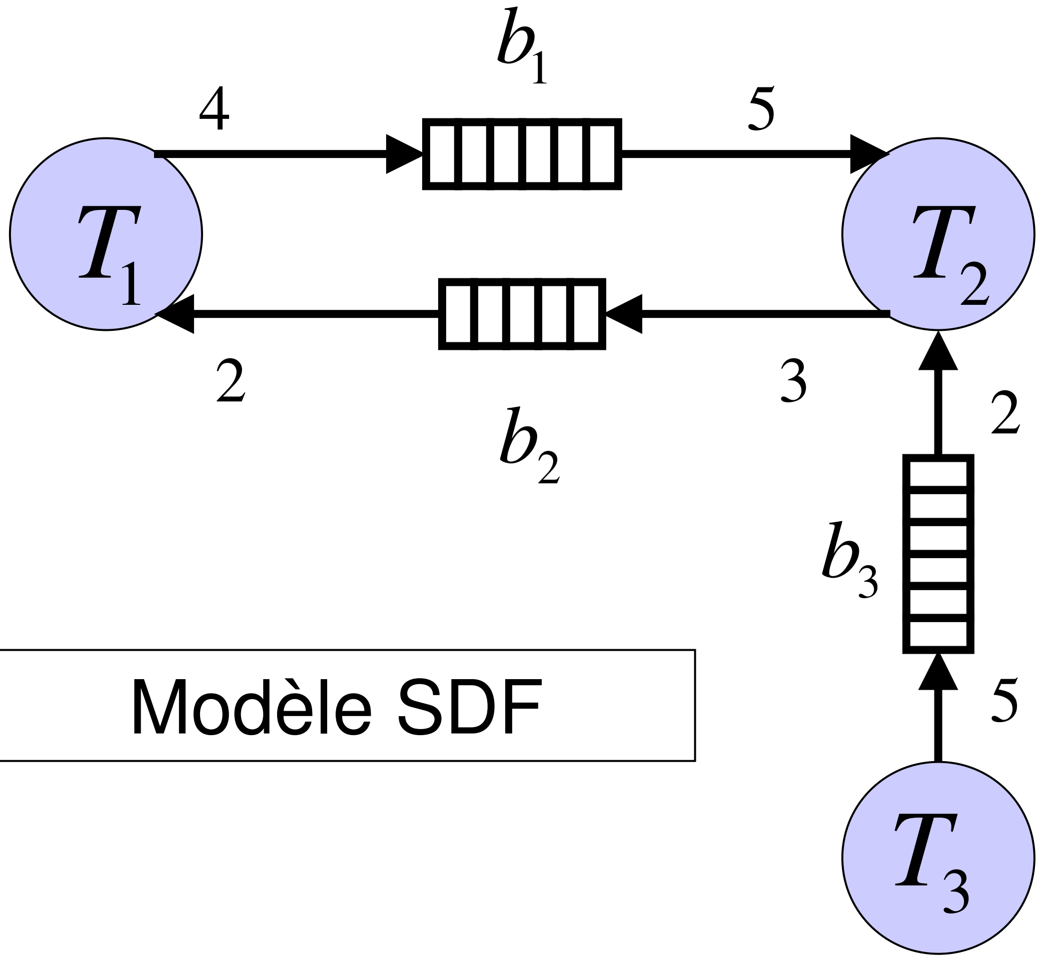
Présentation du problème - Introduction



Présentation du problème - Introduction

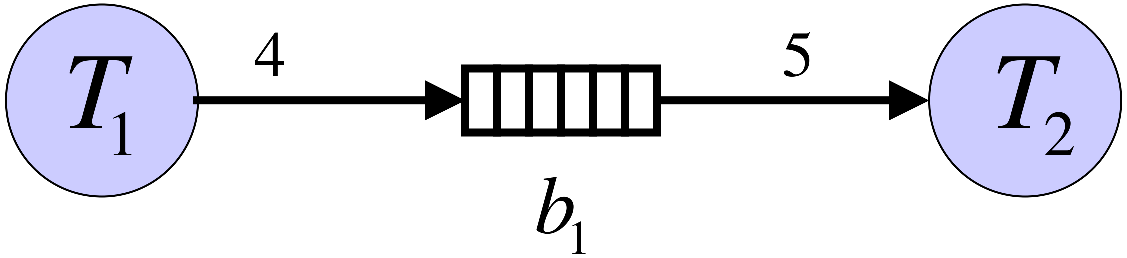


Présentation du problème - Introduction



Modèle SDF

Présentation du problème - Introduction



- La tâche T_1 peut écrire dans b_1.
- La tâche T_2 peut lire dans b_1.
- Les buffers sont gérés comme des FIFOs.

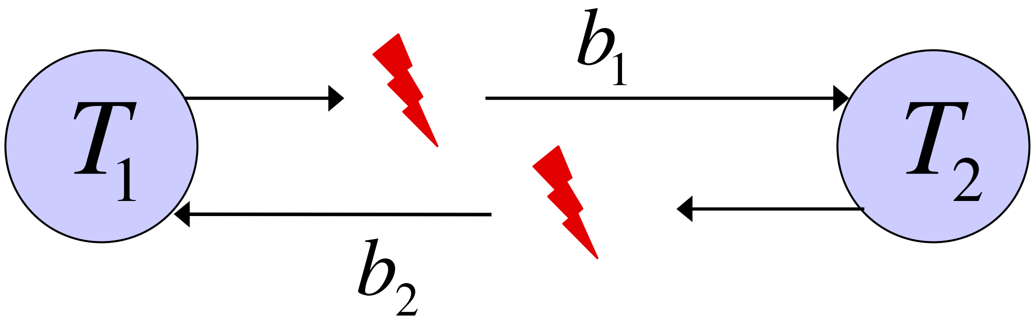
Présentation du problème - Introduction

- Une opération de lecture dans un buffer vide est impossible.
- Opération en attente.
- Tâche bloquée.

Présentation du problème - Introduction

- Une opération d'écriture dans un buffer plein est impossible.
- Opération en attente.
- Tâche bloquée.

Présentation du problème - Introduction



- Des blocages peuvent conduire à l'immobilisation complète du système.

Définition du problème

- Le problème du dimensionnement des mémoires consiste à définir pour chaque buffers une taille telle que :
  - garantie de l'absence d'interblocage (~vivacité),
  - optimiser la surface totale de ces buffers pour atteindre un débit donné (~bi-critère).

Présentation du problème - Applications Industrielles & Techniques

- Conception des DSP :
  - Applications Multimédias.
- Conception des SE temps réels.
- Modélisation de chaîne d'assemblage.
  - Limitation des stocks.

Applications Industrielles - Enjeux

- Comprendre et modéliser le fonctionnement de ces systèmes.
  - Vers des applications de plus en plus complexes.
  - Analyse de performances (débit).
- Actuellement : sur-dimensionnement.
  - Pb car coût très élevé du Silicium.
  - Pb car sur-coût en consommation.
  - Etape de vérification indispensable.

Notations

Choix de l'approche réseau de Petri

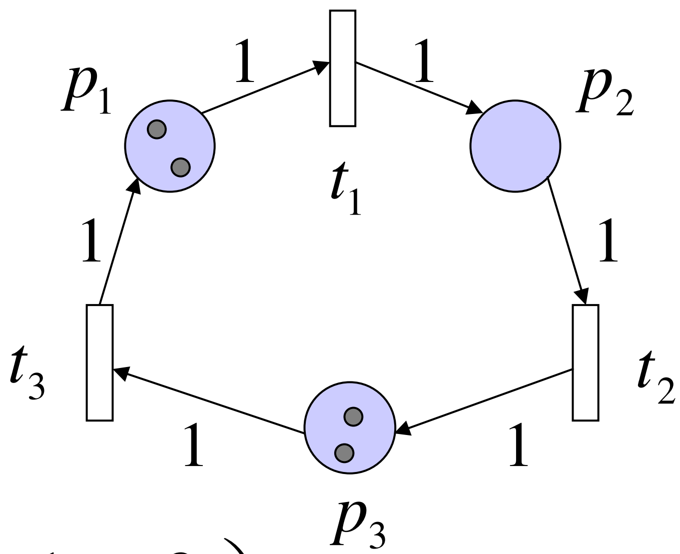
- GEGT équivalents aux SDF :
  - transition = tâche
  - place = buffer
  - jeton = donnée
- Formalisme plus éprouvé et plus standard.
- Variété des outils mis au point
  - Pb de décision,
  - Pb d'ordonnancement.

Notations - Représentation graphique

- Marquage initial :

$$M_0(p_1) = M_0(p_3) = 2$$

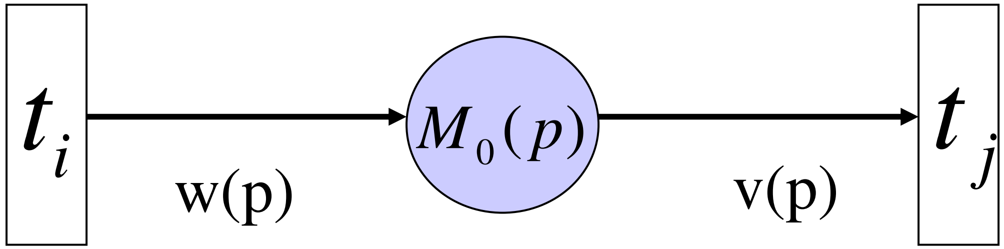
$$M_0(p_2) = 0$$



- Matrice d 'incidence :

$$\Gamma_G = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \end{pmatrix}$$

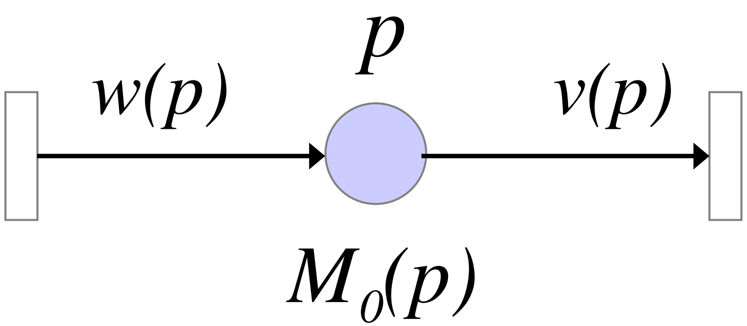
Graphe d'événements généralisé temporisé (GEGT)



- w(p) = nb de jetons émis lorsque la transition est franchie.
- v(p) = nb de jetons retirés pour franchir la transition.

Capacité d'une place

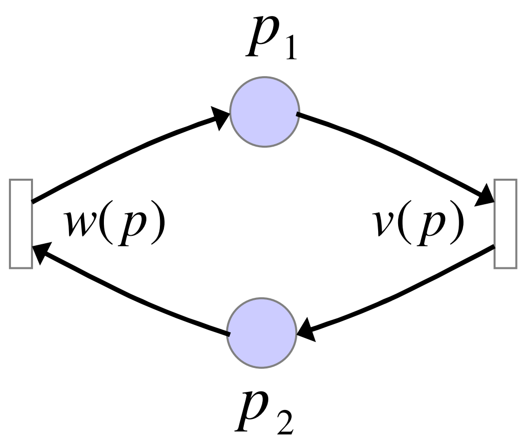
- Limitation d'une place :



$$M(p) \leq C(p)$$

Résultats sur la limitation de place

- Limitation d'une place :



$$M_0(p_1)+M_0(p_2)=C(p)$$

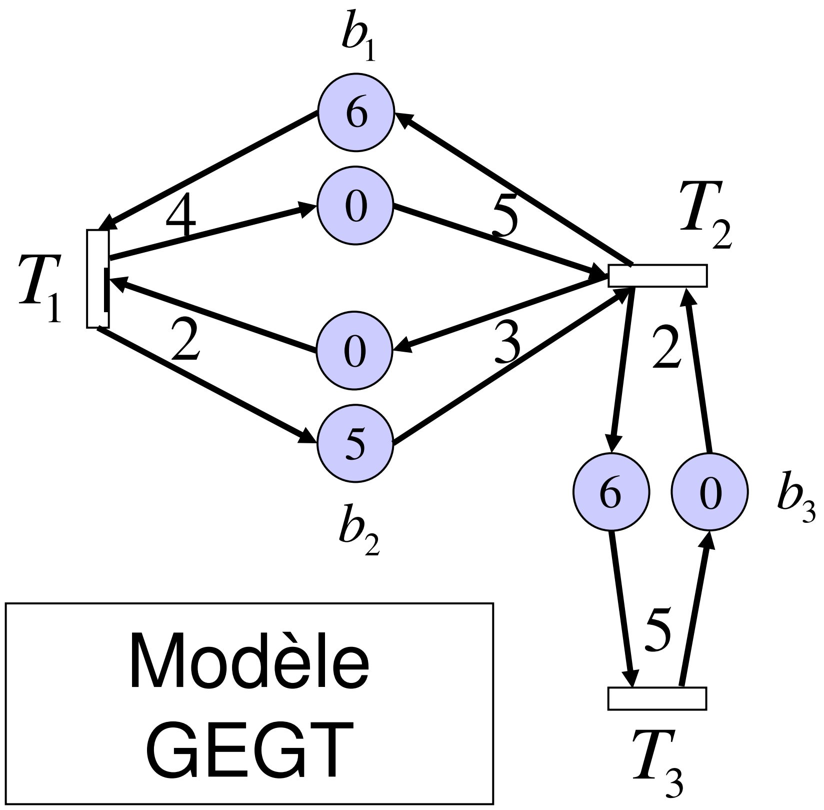
$$et \quad M_0(p_1)=M_0(p)$$

- La taille minimale T_{min} d'une place p vaut :

$$T_{min} = w(p)+v(p)\text{-}pgcd(w(p),v(p))$$

GEGT minimalement borné

- Un GEGT est minimalement borné si chaque couple de places associées à un buffer contient T_{min} jetons.



Propriétés associées à Γ_G

- Un P-semiflot est un vecteur Y tel que :

$$^{T}Y \cdot \Gamma_G = 0$$

  - utilisé dans les critères d'optimisation (combinaison linéaire des marquages constante).

- Un T-semiflot est un vecteur N tel que :

$$\Gamma_G \cdot N = 0$$

  - Le plus petit vecteur N non-nul désigne le nombre de répétition de chaque tâche qu'il faut faire pour atteindre de nouveau l'état initial.

Résultats de complexité

Problème de l'évaluation du débit

- Dans le cas des GEGT :
  - On ne connaît pas de certificat de vivacité de taille polynomiale.
  - Evaluer la vivacité (ou le débit) d'un GEGT se fait avec des algorithmes pseudo-polynomiaux.

- Pb : les valeurs numériques dans les SE sont très grandes.
  - Expansion (AMK) : explosion combinatoire.
  - Model Checking : idem

Restriction aux graphes d'événements temporisés (GET)

- Pour être dans NP, on travaille avec les GET.

- Soit $G = (P, T, l)$ un GET.
  - P l'ensemble de places.
    - une place modélise des contraintes.
  - T l'ensemble des transitions.
    - une transition modélise une tâche.

    $l : T \rightarrow R^{+*}$
  - Fonctions de marquage unitaire (i.e. w(p)=v(p)=1 pour toute place p).

Problème Flot Ratio (Minty-1962)

- Le flot ratio $\rho(o(G))$ d'une orientation $o(G)$ vaut :

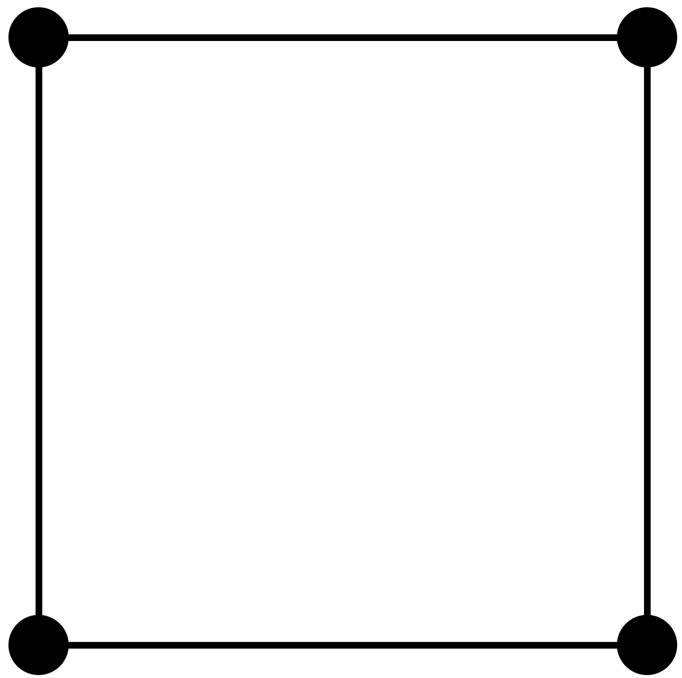
$$\rho(o(G)) = \max\left(\frac{n_c(o)}{m_c(o)}, \frac{m_c(o)}{n_c(o)}\right)$$

- <u>Flot Ratio :</u> Soient $G = (V, E)$ un graphe simple et K un entier. Existe-t-il une orientation des arêtes de E telle que le flot ratio soit inférieur à K ?

- <u>Théorème de Minty:</u> Soit $G = (V, E)$ un graphe simple. G est K-colorable si et seulement s'il existe une orientation des arêtes de flot ratio inférieur à $K-1$.

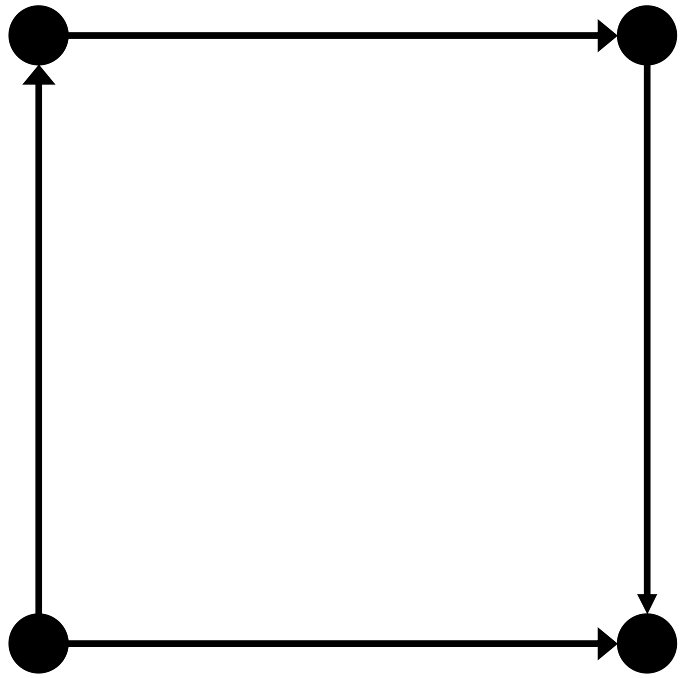
Problème Flot Ratio (Minty-1962)

- Exemple :



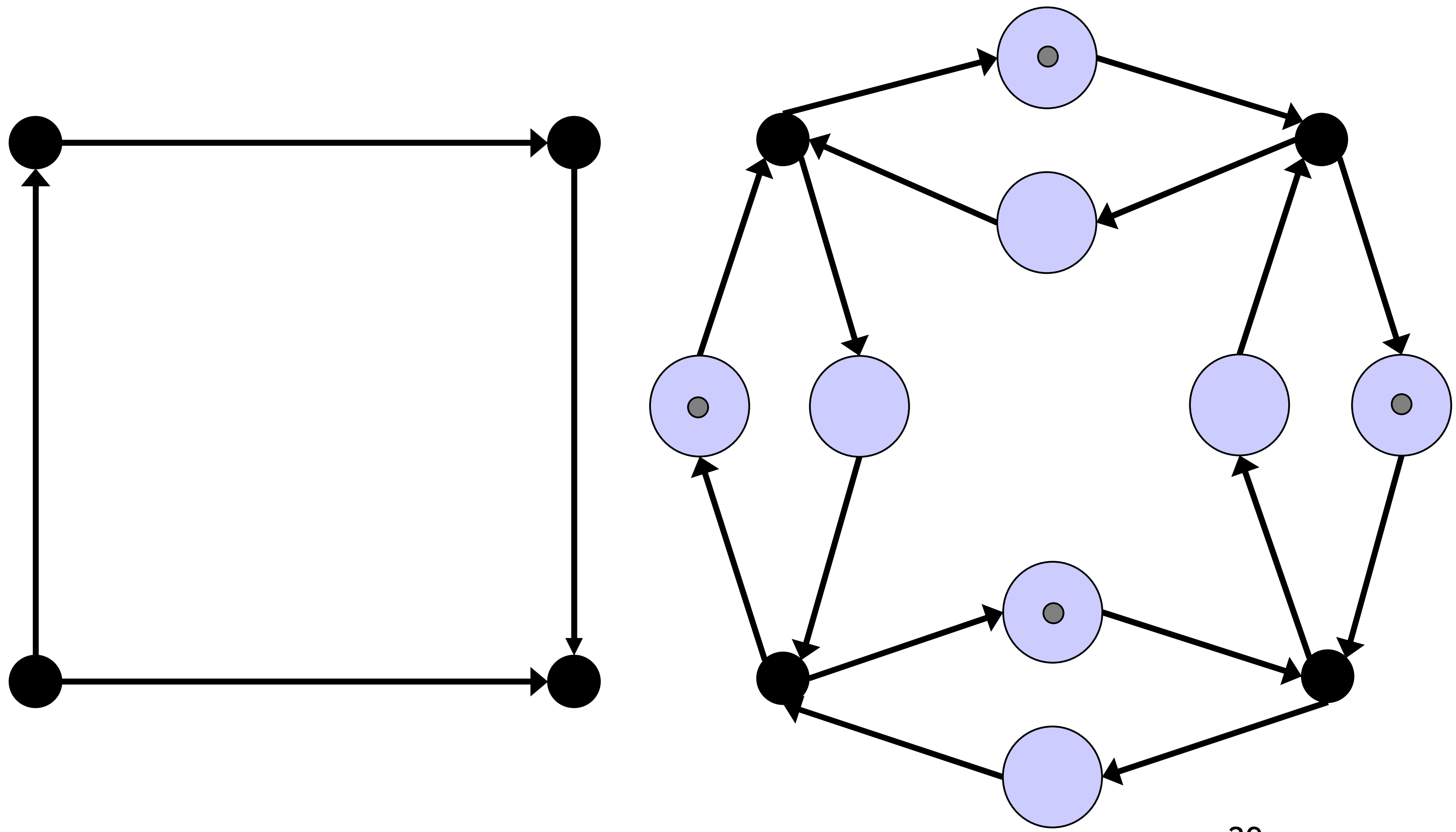
Problème Flot Ratio (Minty-1962)

- Exemple :



$$\rho(o(G)) = \max\left(\frac{3}{1}, \frac{1}{3}\right)$$

Lien avec les GET minimalement borné



Lien avec les GET minimalement borné

- Lemme (MMK-06) : Soit $o(G)$ une orientation de G et $M(G)$ le marquage minimalement borné associé à cette orientation. Alors :

$$\lambda(M(G)) = \frac{1}{\rho(o(G))+1}$$

- Conséquence : Calcul du Flot Ratio $\in P$.
  Flot Ratio $\in NP$.

Problème Flot Ratio (Minty-1962)

- Théorème 1 (MMK-06) :
  Flot ratio est $NP-Complet$.

  - Preuve : Flot ratio $\in NP$.
    D'après le lemme de Minty, on a
    K-coloration $\leq_P$ Flot Ratio .

Problème Débit Max - Borné Min

- Débit Max – Borné Min: Soient G=(T, P, l) un GET symétrique à tâche unitaire et un entier $K > 1$. Existe-t-il un marquage initial $M(G)$ tel que :

$$\sum_{p \in P} M(p) = \left|\frac{P}{2}\right| \quad \text{et} \quad \lambda(M(G)) \geq \frac{1}{K}$$

- Théorème 2 (MMK-06) :
Débit Max - Borné Min est $NP-Complet$.
  - Preuve : Débit Max - Borné Min $\in NP$.
D 'après le lemme (MMK-06),
Flot Ratio $\leq_P$ DM-BM
    - Remarque : Flot Ratio et DM-BM sont équivalents.

Problème Optimisation de marquage

- Problème classique en conception de chaîne d 'assemblage. Introduit par Laftit *et al.* en 1992.

- <u>Optimisation de marquage</u>: Soient G=(T, P, I) un GET, Y un P-semiflot, un entier K et Z$\in$ $[0,1]$. Existe-t-il un marquage initial $M(G)$ tel que :

$$^{T}Y \cdot M(G) \leq K \quad \text{et} \quad \lambda(M(G)) \geq Z$$

Problème Optimisation de marquage

- Théorème 3 (MMK-06) : Optimisation de marquage est $NP-Complet$.

  - Preuve: Optimisation de marquage $\in NP$. Soit I une instance de DM-MB. Alors le vecteur unitaire est un P-semiflot de Γ_G.

    DM-BM $\leq_P$ Optimisation de marquage

Complexité de Débit Maximum Intrinsèque

Importance du débit maximum intrinsèque

- Lorsque les buffers sont grands, c'est comme si il n'y avait plus de contraintes entre les différentes exécutions des tâches.
- Débit Maximum Intrinsèque d'un système.
  - Cadence maximale de l'application.
  - Un ou plusieurs composants limitent le débit des autres.
- Donne une borne supérieure du débit.

Problème Débit Maximum Intrinsèque

- Problème classique en conception de chaîne d 'assemblage. Introduit par Hillion *et al.* en 1987.

- <u>Débit Maximum Intrinsèque :</u> Soient G=(T,P,l) un GET, un entier K et $l^* = \max_{t_i \in T} (l(t_i))$.
  Existe-t-il un marquage initial $M(G)$ tel que :

$$M(G) \leq K \quad \text{et} \quad \lambda(M(G)) \geq \frac{1}{l^*}$$

Problème Débit Maximum Intrinsèque

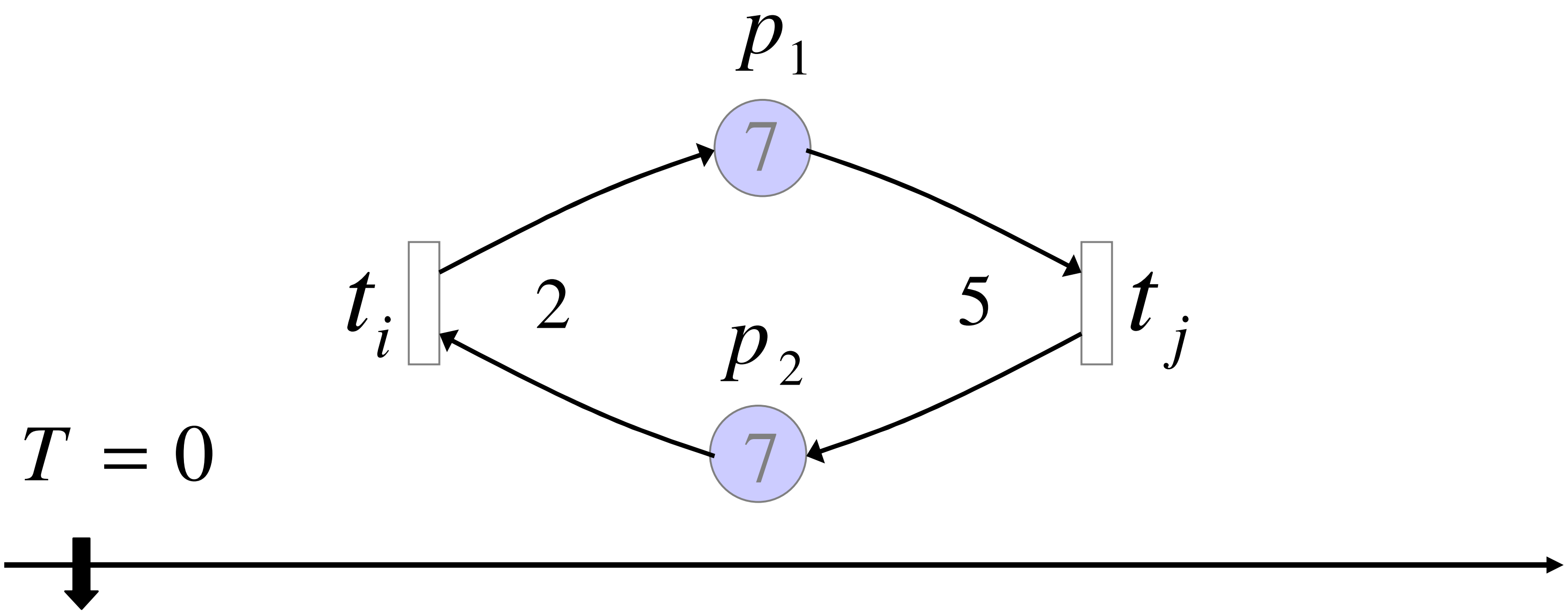
- Théorème 4 (MMK-06) : Débit Maximum Intrinsèque est $NP-Complet$.
  - Preuve : Débit Maximum intrinsèque $\in NP$.

    Soit I une instance de DM-BM. Ajout d 'une transition de durée K reliée à G par un couple de places. Le débit M.I. vaut $1/K$.

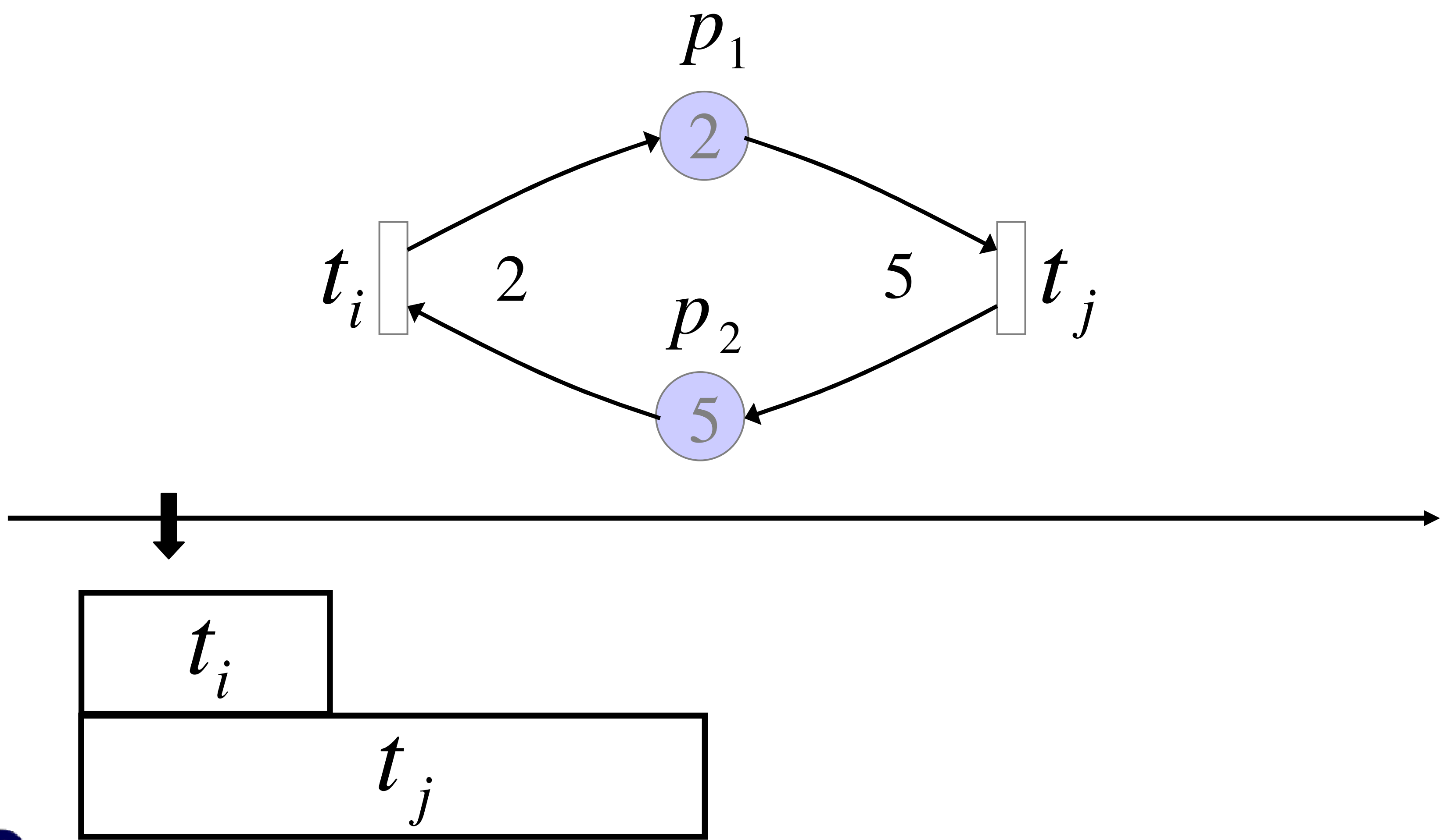
    DM-BM $\leq_P$ Débit Maximum Intrinsèque

Algorithme polynomial 2-approché pour DMI

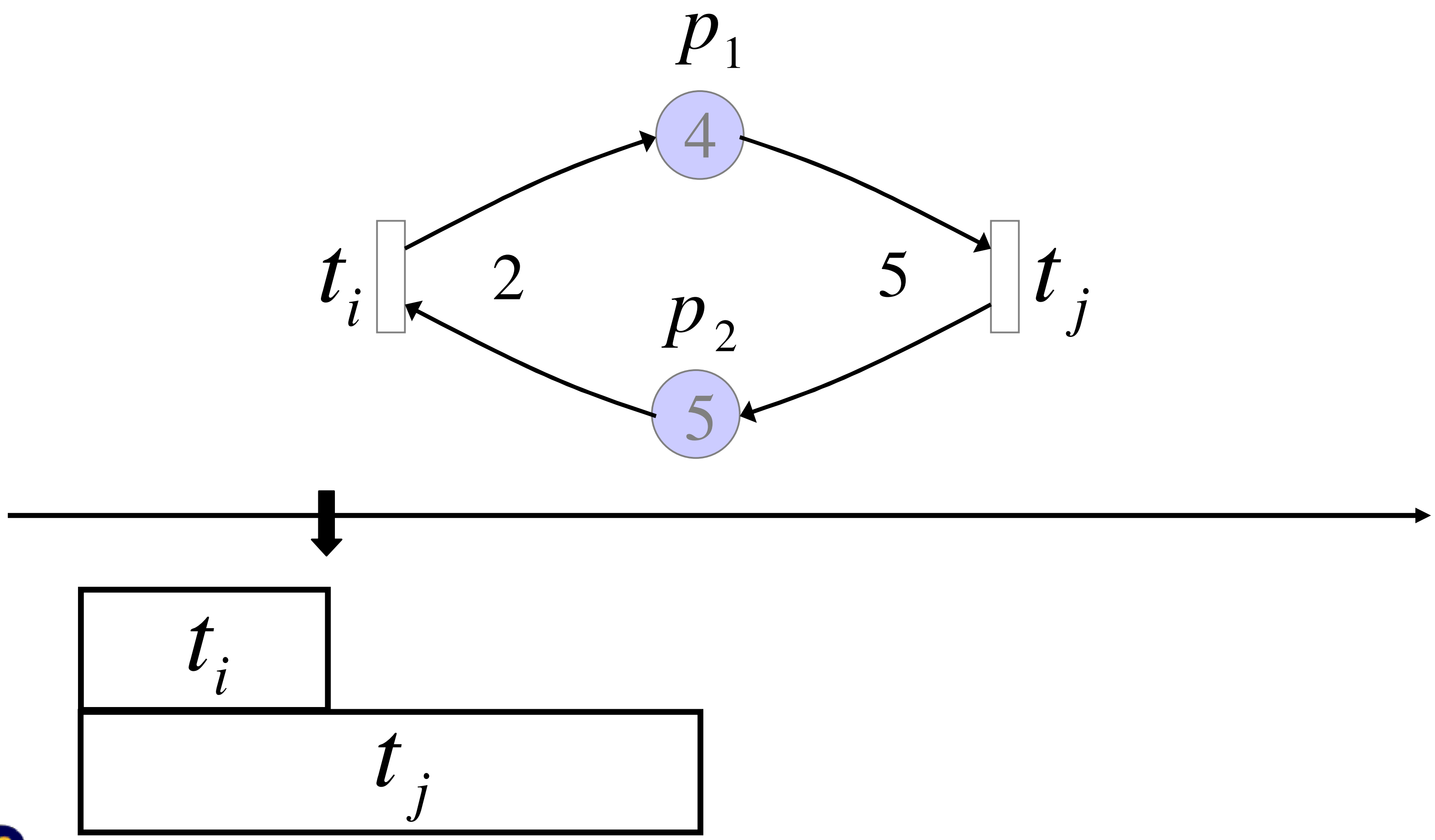
Exemple : cas à deux tâches



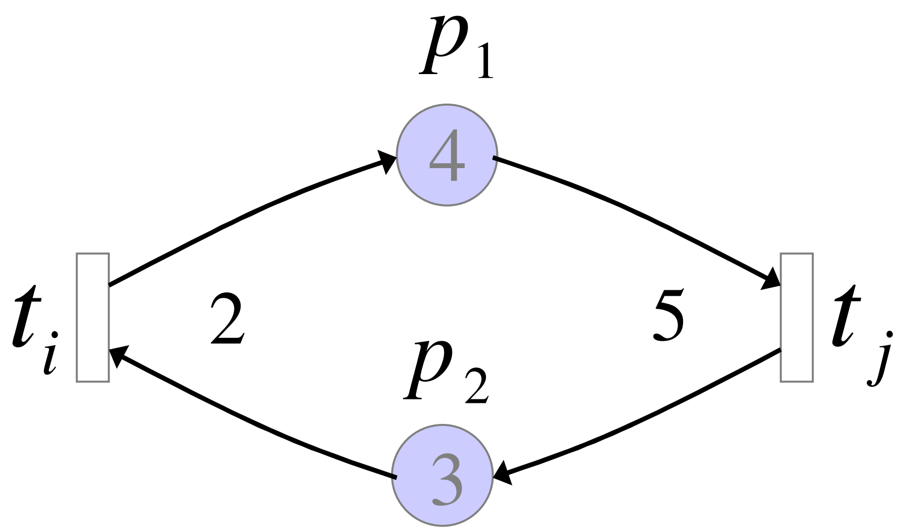
Exemple : cas à deux tâches

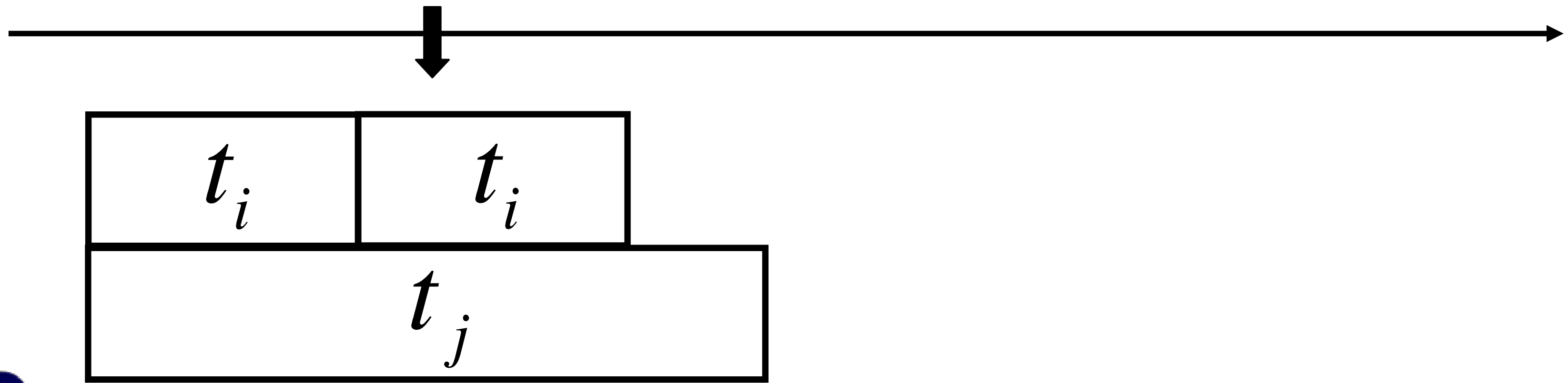


Exemple : cas à deux tâches

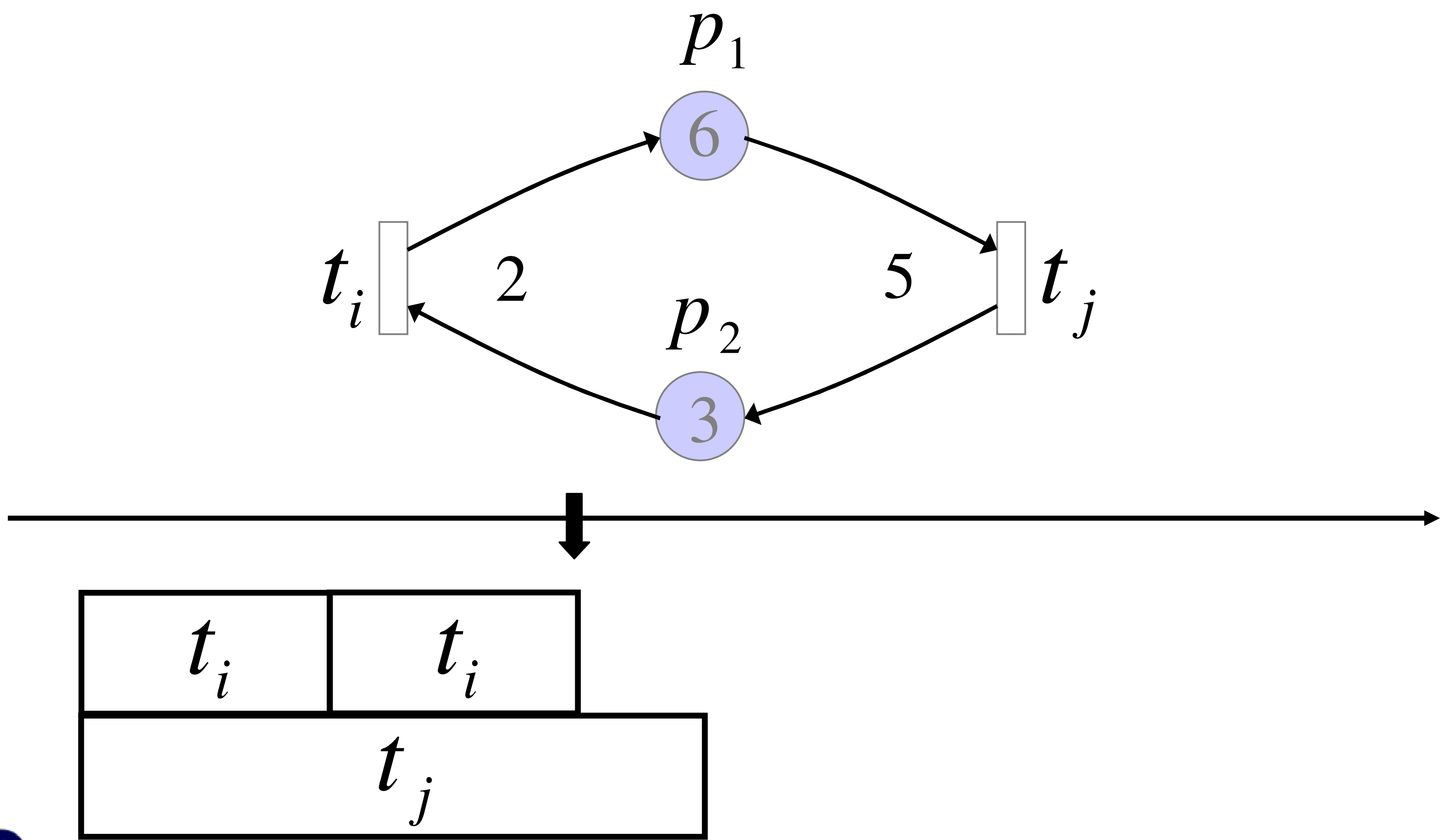


Exemple : cas à deux tâches

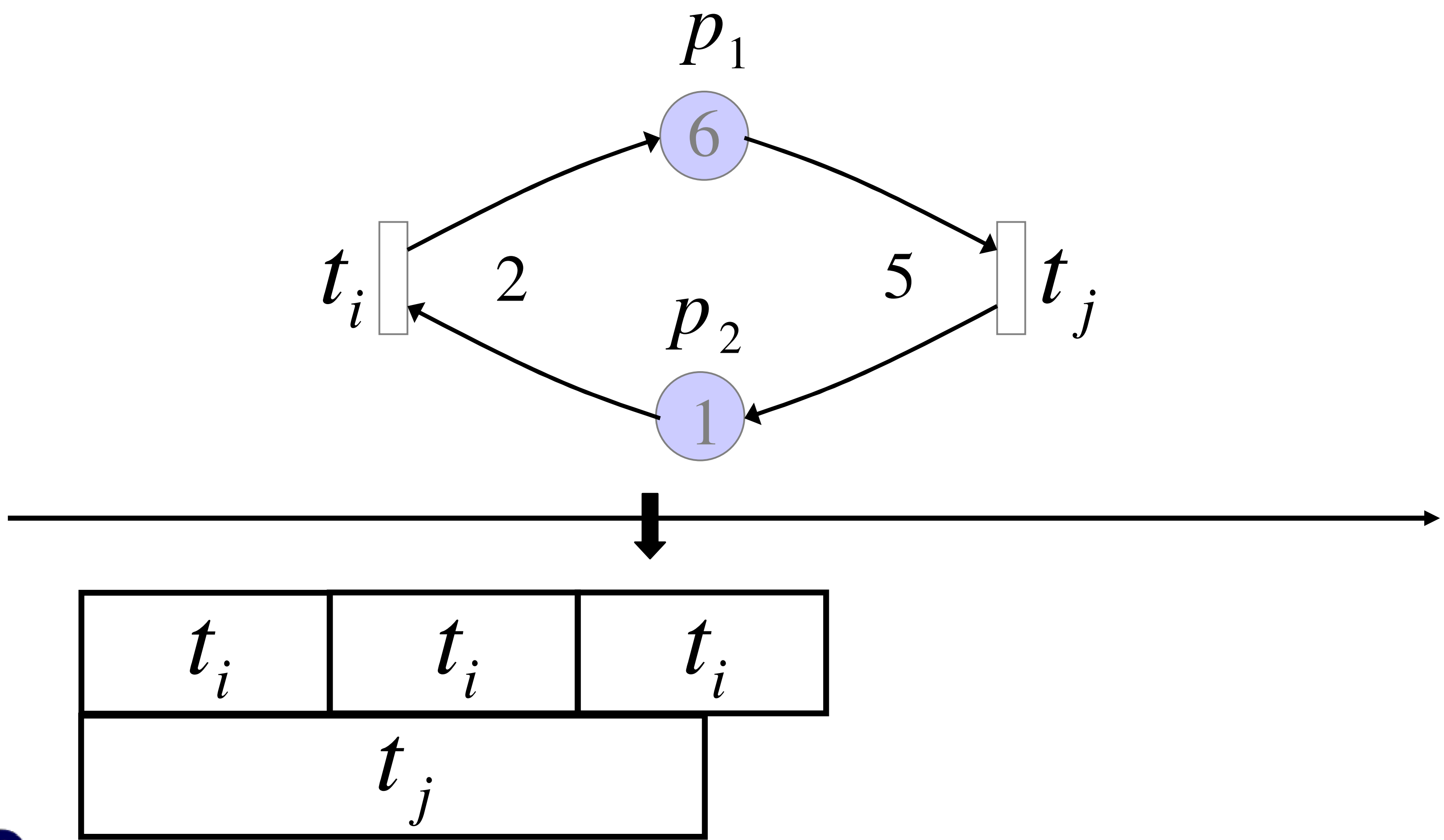




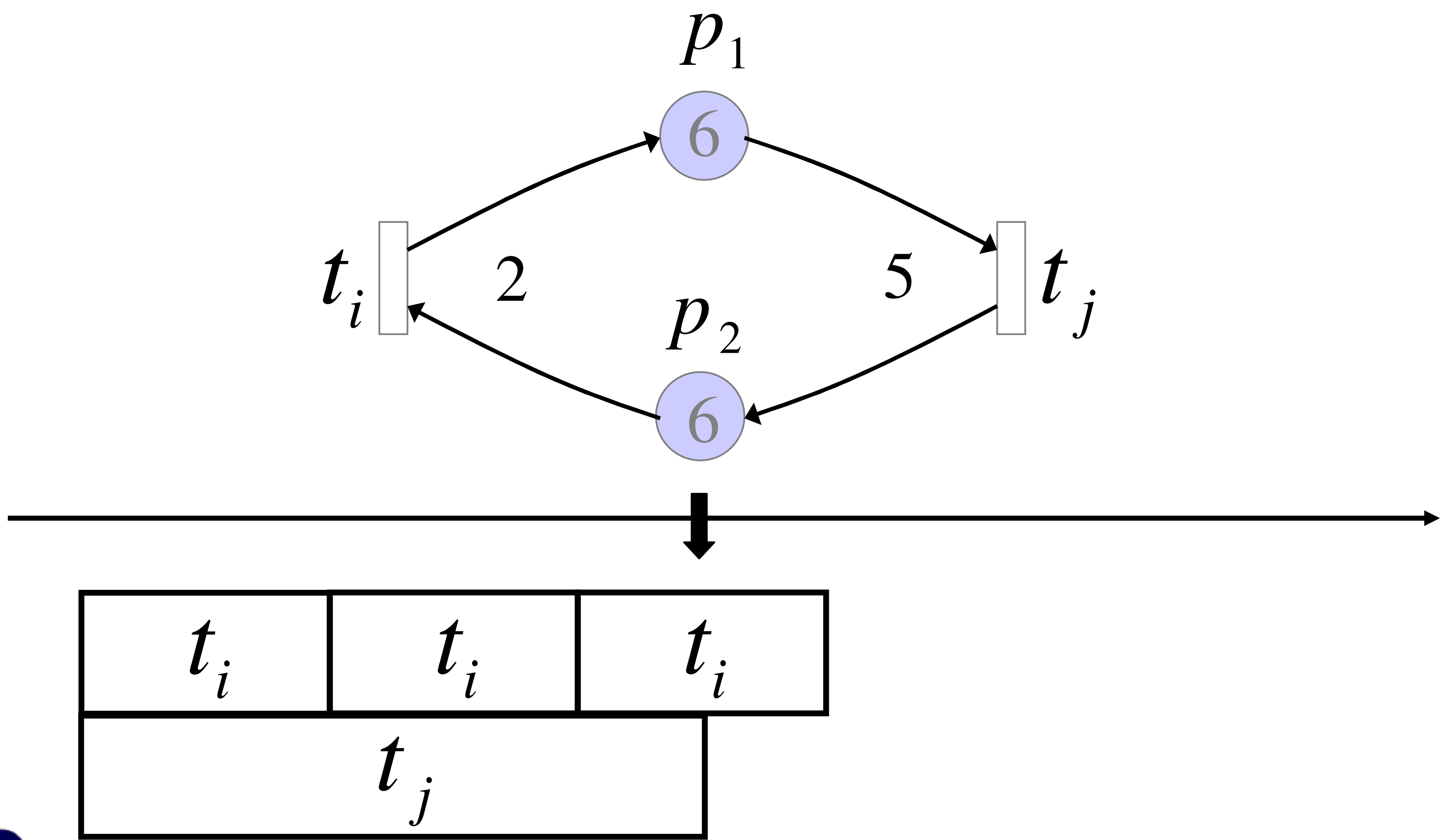
Exemple : cas à deux tâches



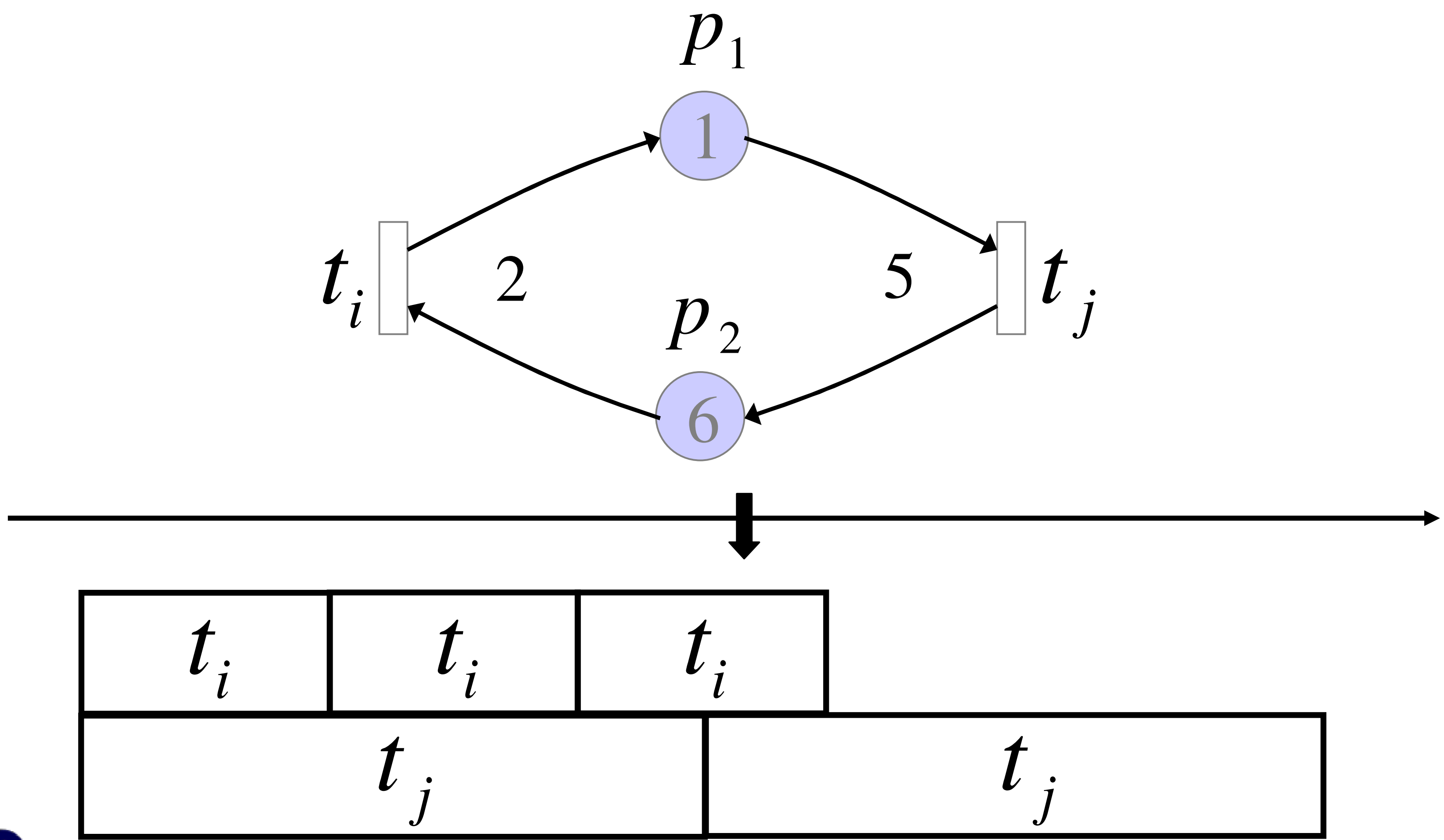
Exemple : cas à deux tâches



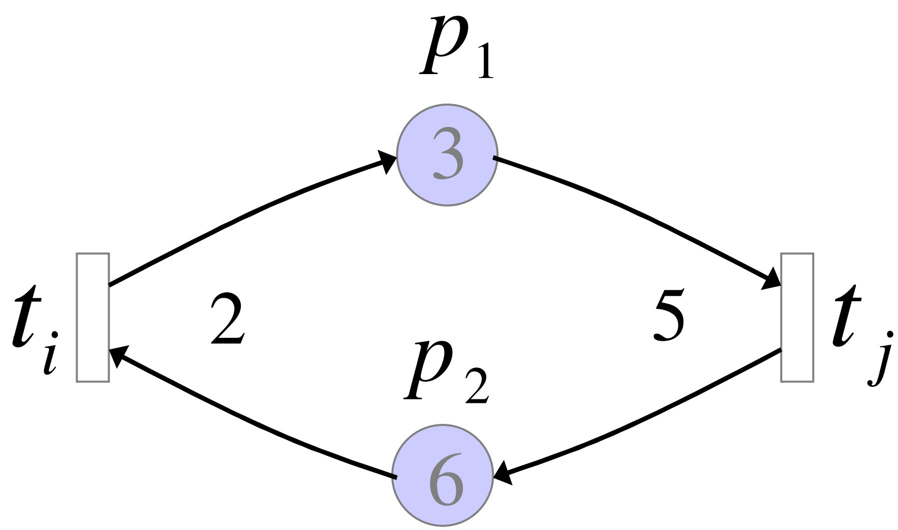
Exemple : cas à deux tâches

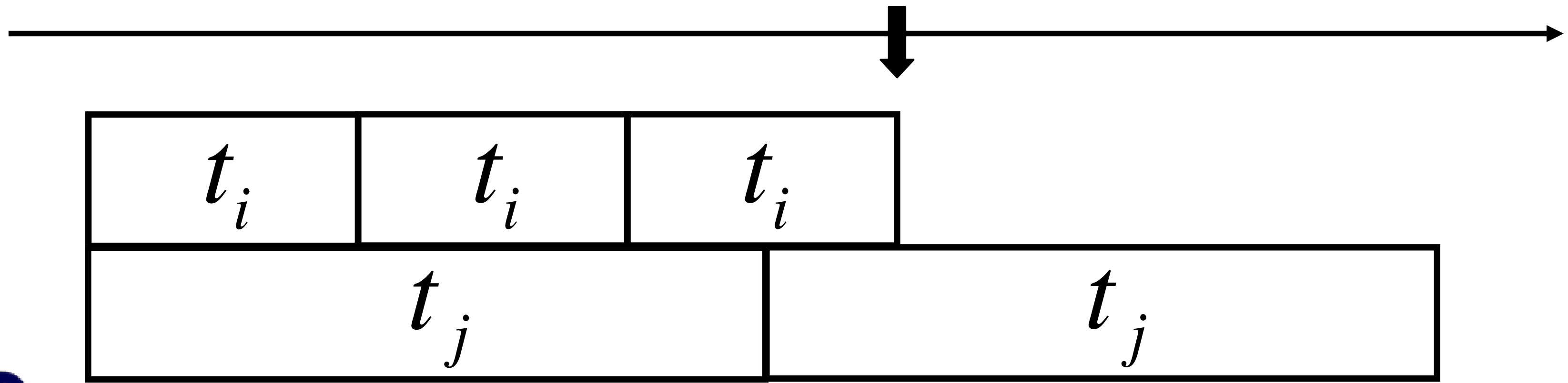


Exemple : cas à deux tâches

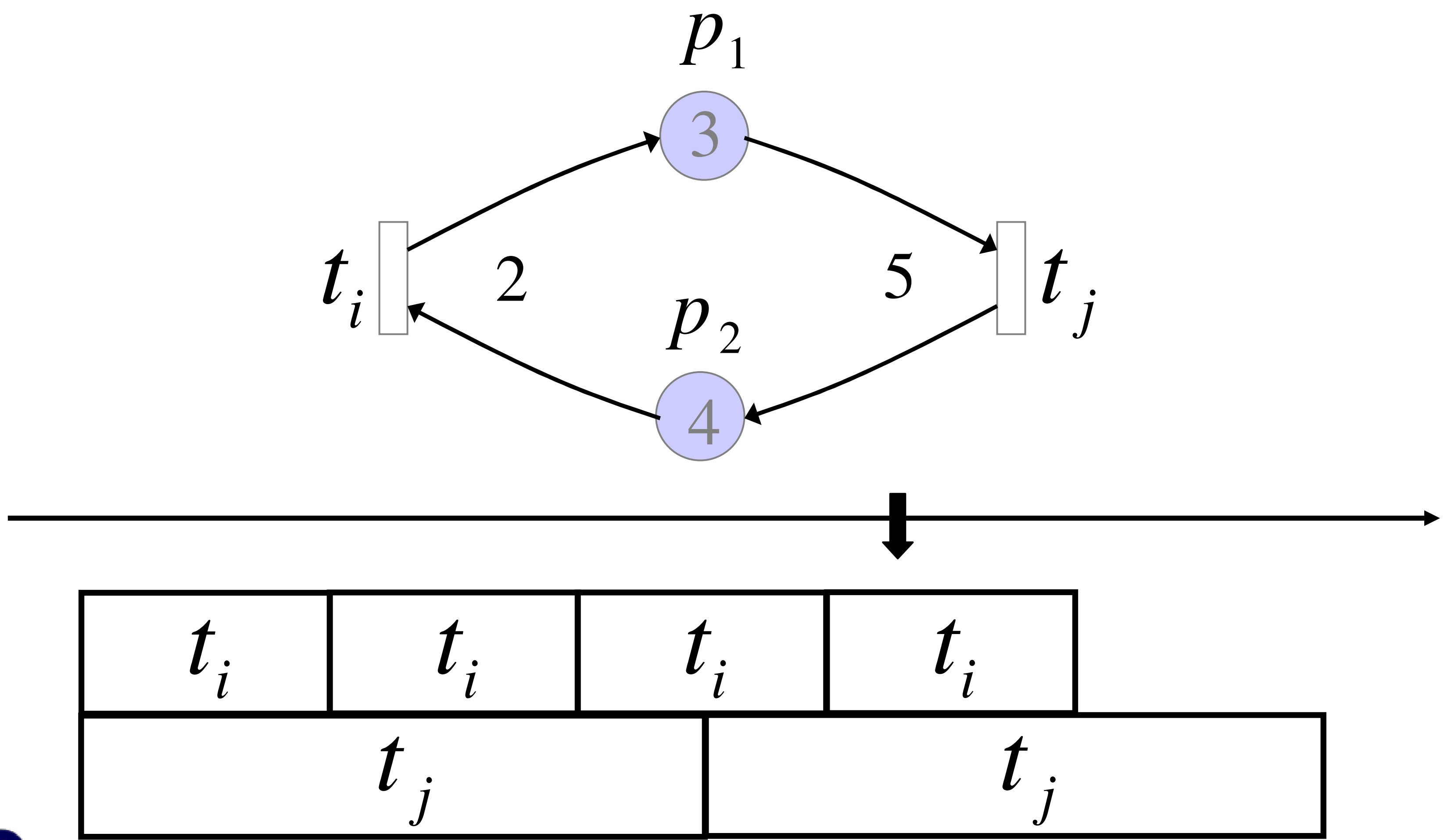


Exemple : cas à deux tâches

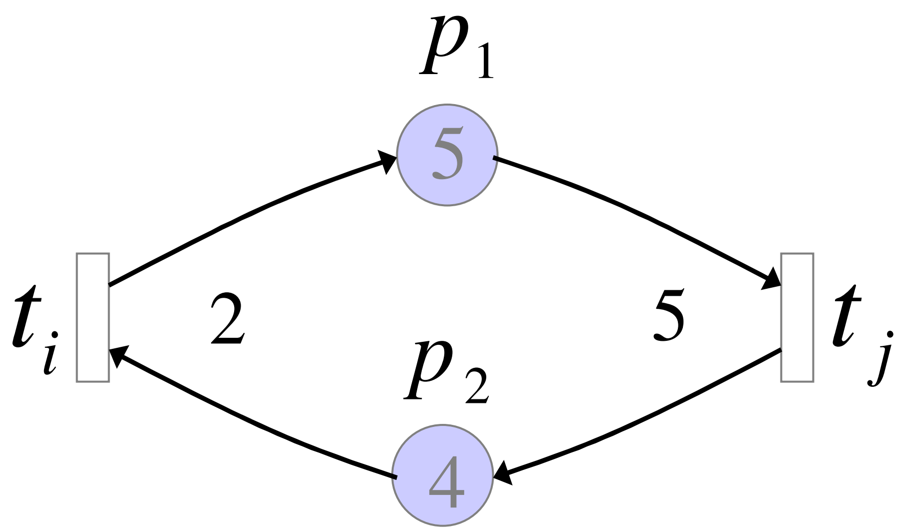


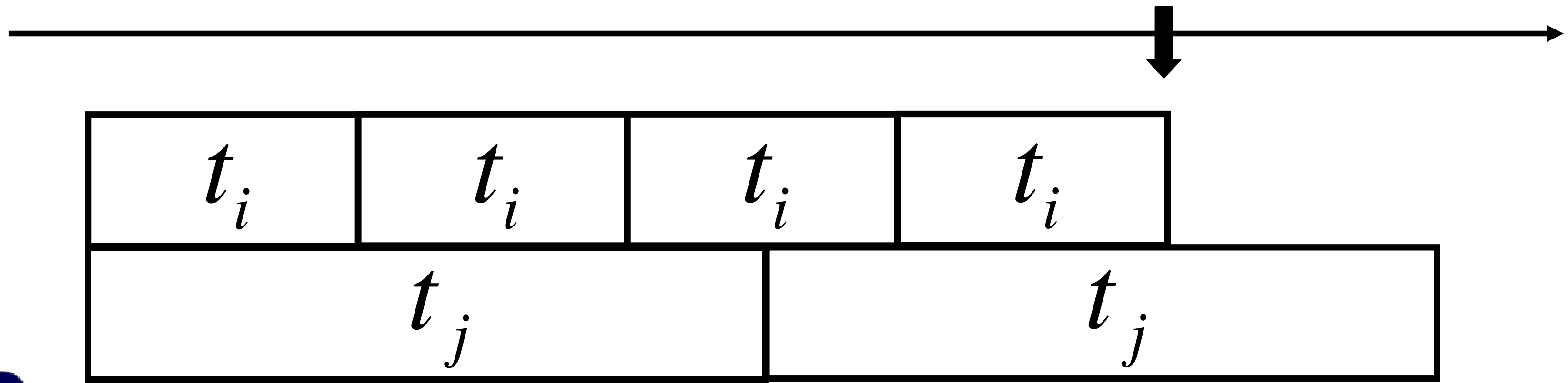


Exemple : cas à deux tâches

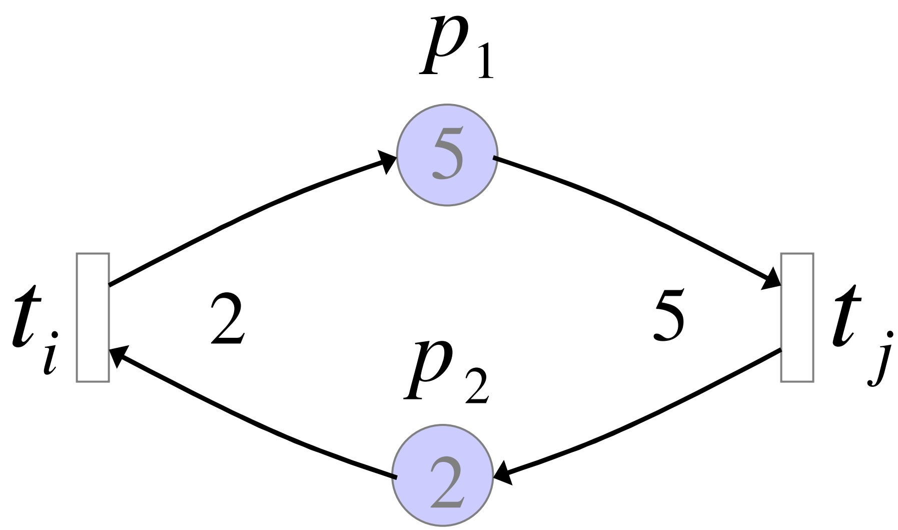


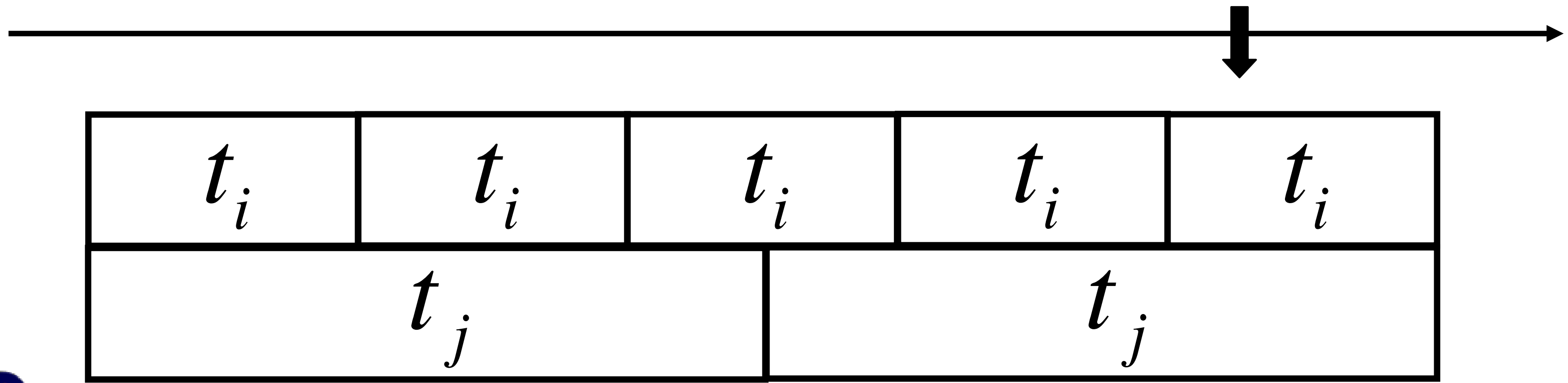
Exemple : cas à deux tâches



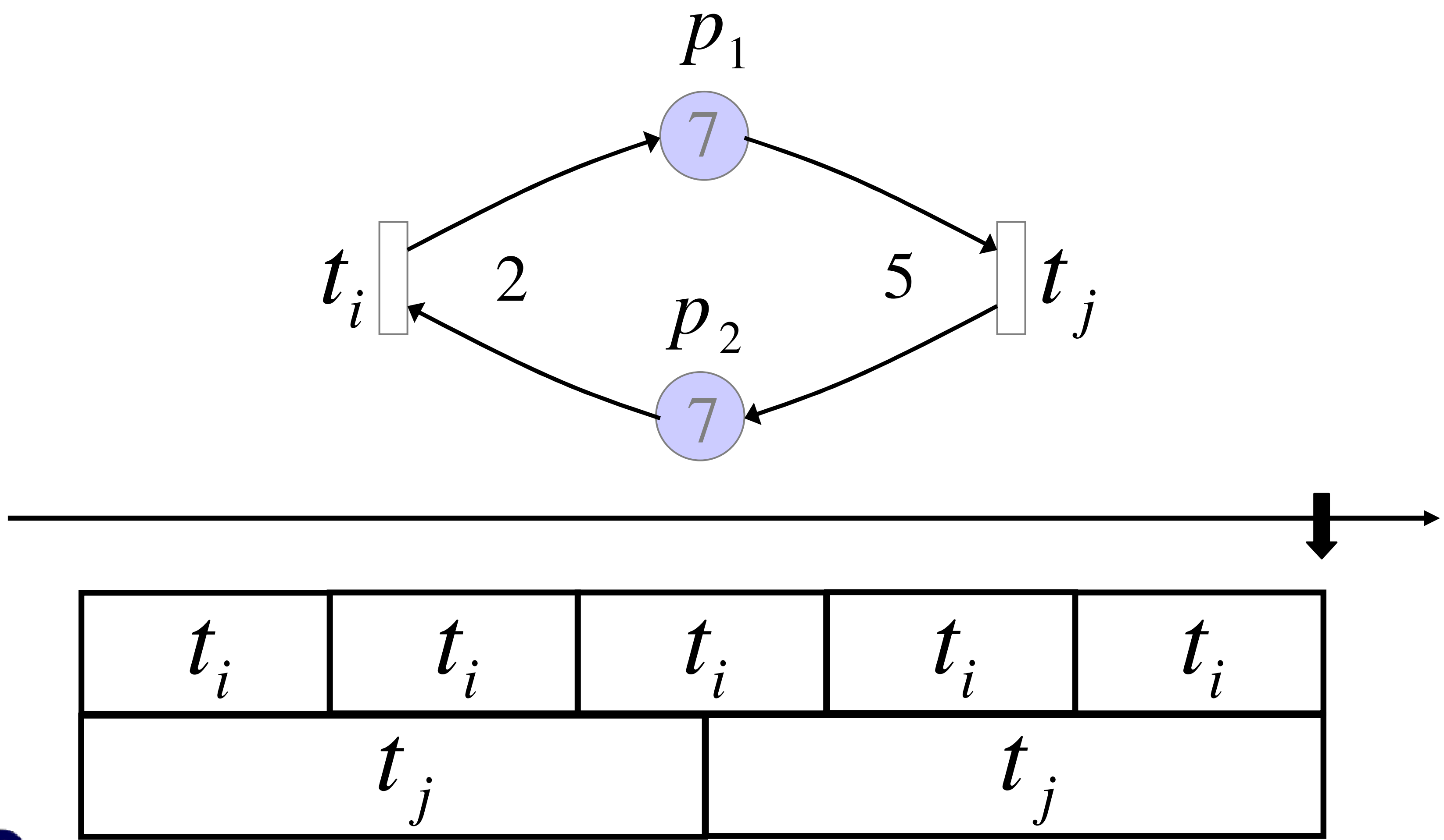


Exemple : cas à deux tâches

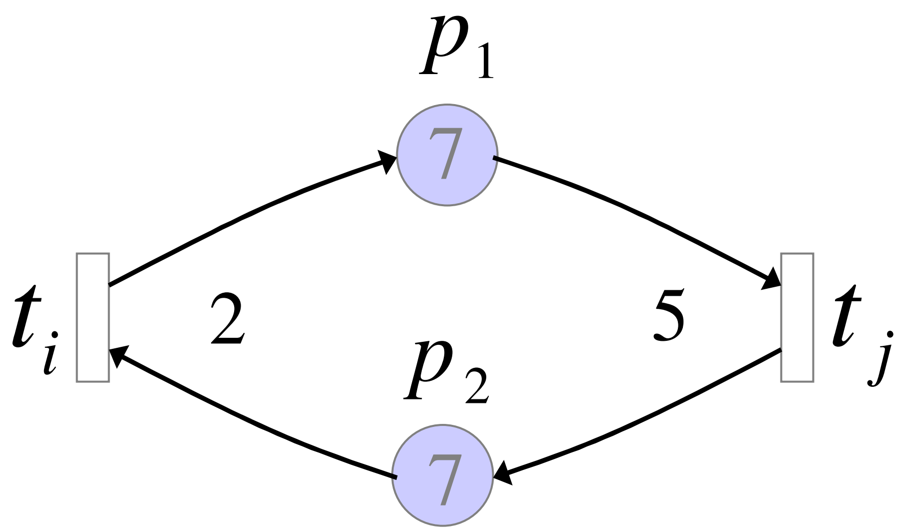


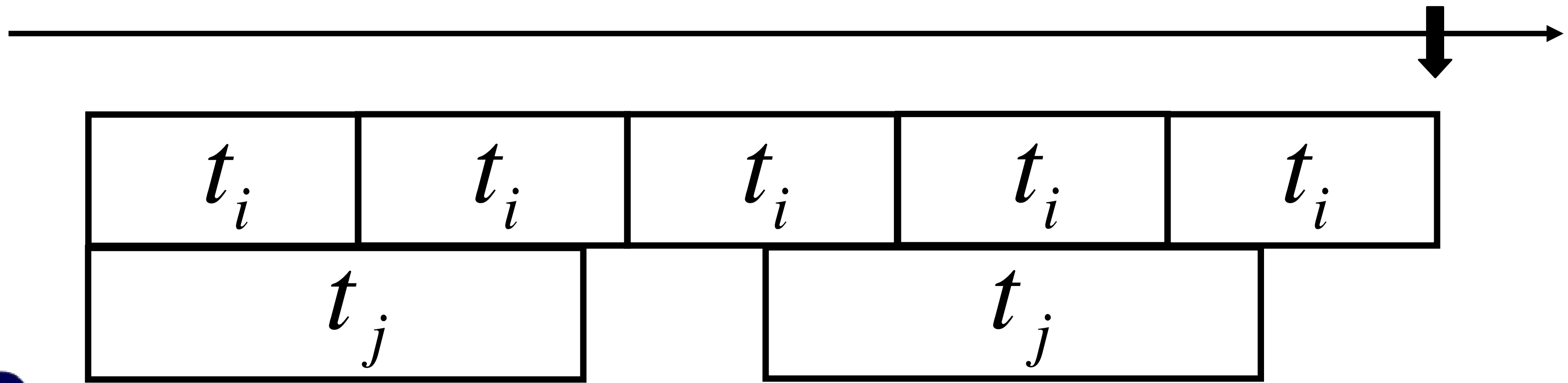


Exemple : cas à deux tâches

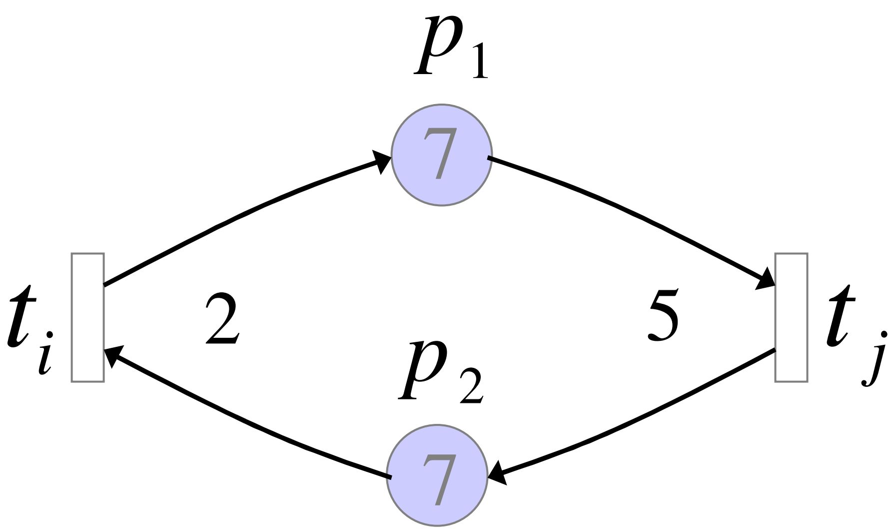


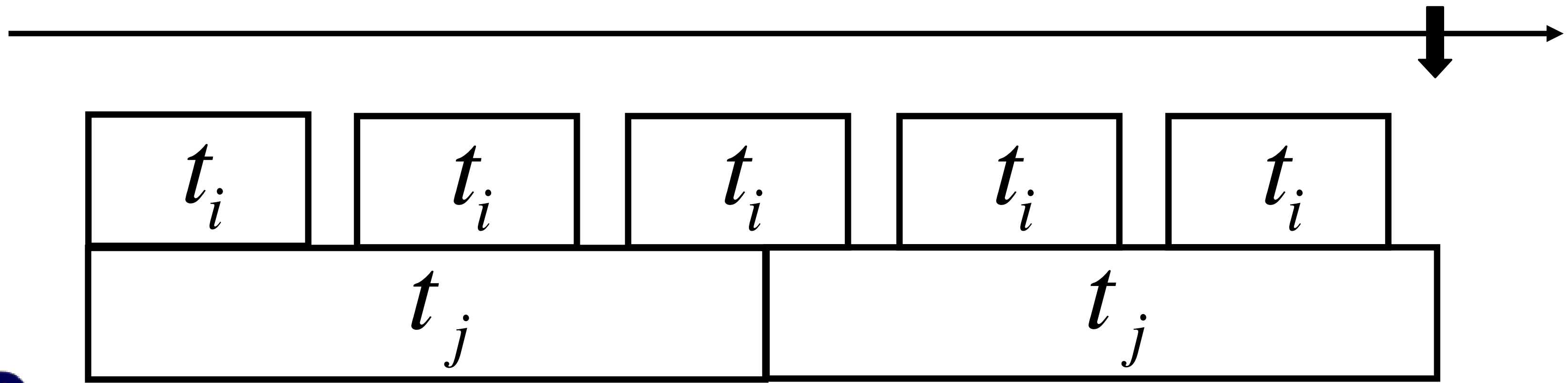
Exemple : cas non-critique





Exemple : cas non-critique





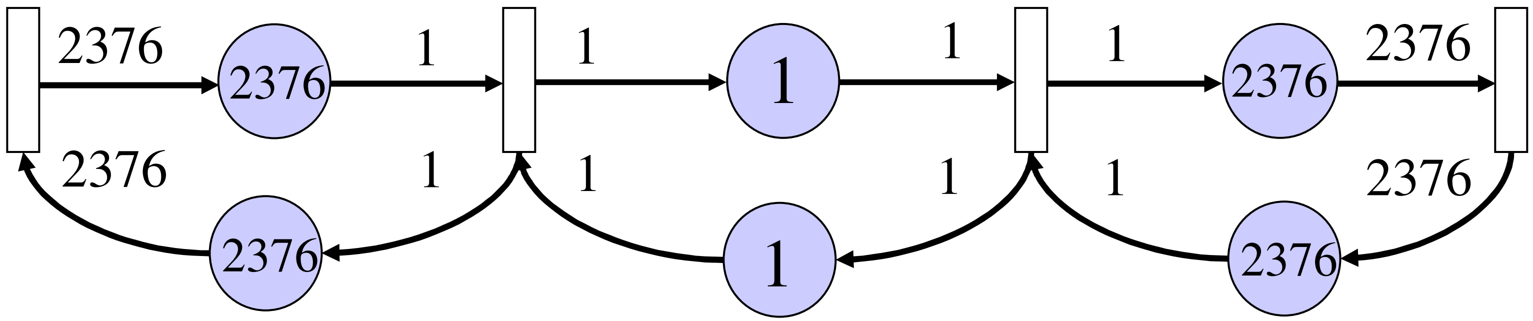
Problème de marquage avec contrainte sur le débit

- Théorème 5 (MMK-06) : En mettant sur chaque place T_{min} jetons, on obtient une solution 2-approchée pour le problème du débit maximum intrinsèque.

- L'algorithme est exact si pour toute transition :

$$l(t_i) \cdot N_i = \rho, \quad \forall\, t_i \in T$$

Exemple : H.263 video codec



Analyser le débit avec les méthodes existantes

expansion $\approx$ simulation

⟹ Algorithme en $\Theta(m)$

Conclusion & Perspectives

Conclusion & perspectives

- Lien fort entre les problèmes de coloration et ces problèmes d 'ordonnancement cyclique.
- 3 Problèmes bi-critère d 'ordonnancement cyclique sont $NP - Complets$.
- GEGT symétrique : Elaborer une heuristique pour le problème du débit maximum intrinsèque.
- Etude des ordonnancements 1-périodiques (algorithme polynomiaux pour le débit et la vivacité)

Questions