

Ordonnancement contrôlé de migrations à chaud

*Vincent Kherbache, Fabien Hermenier,
Eric Madelaine*



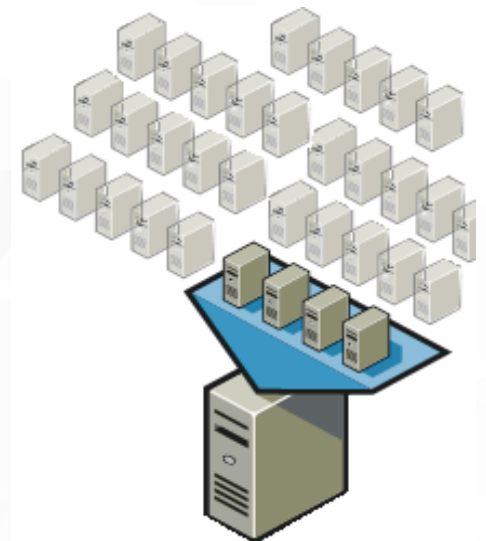
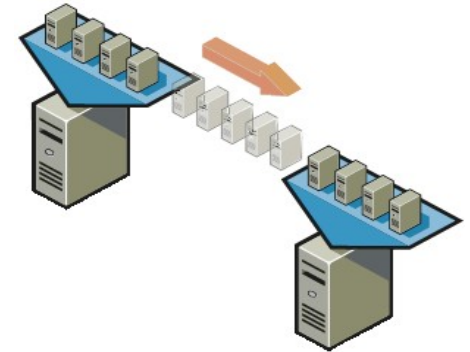
La migration à chaud

▼ Principe

- ▼ Déplacer une VM en cours de fonctionnement entre différents serveurs physique

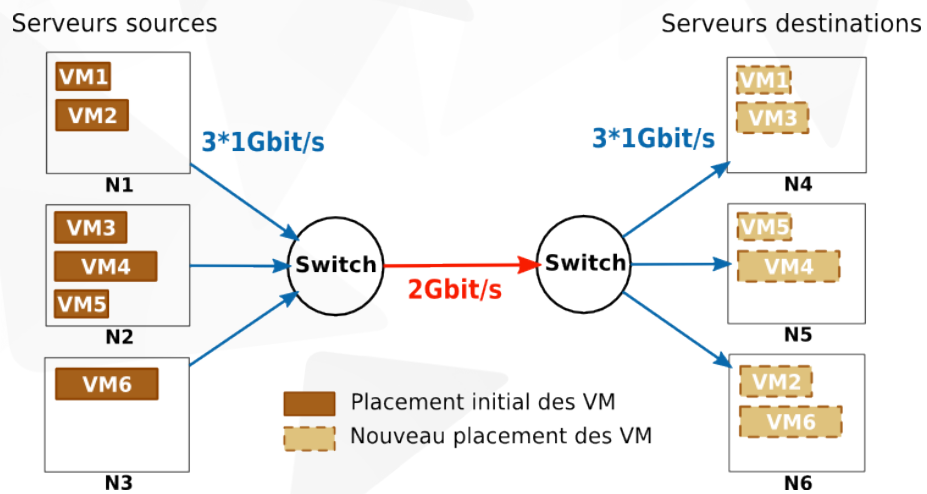
▼ Usages

- ▼ Répartition / gestion de charge
- ▼ Tâches de maintenance sur serveurs de production
- ▼ Réduction de la consommation énergétique



Ordonnancement de migrations

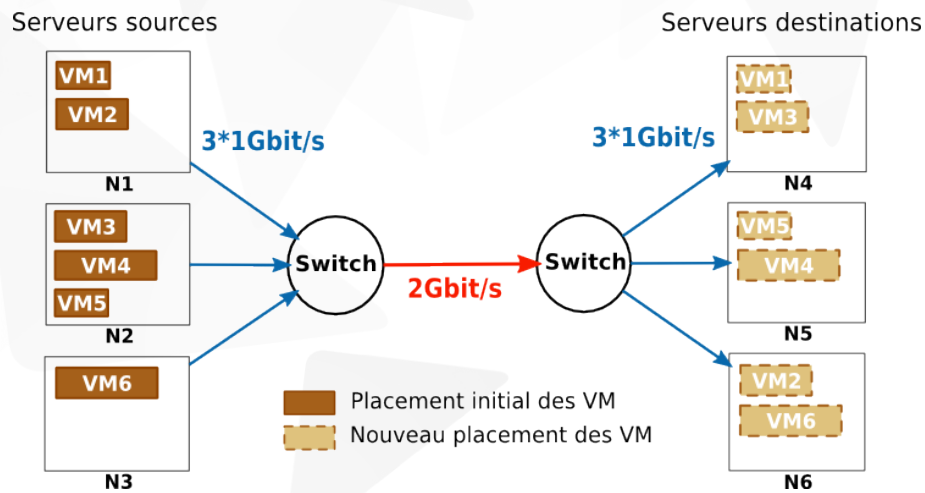
On veut migrer le plus rapidement possible



- Allouer la bande passante maximale pour chaque migration
- Paralléliser sans saturer le réseau

Ordonnancement de migrations

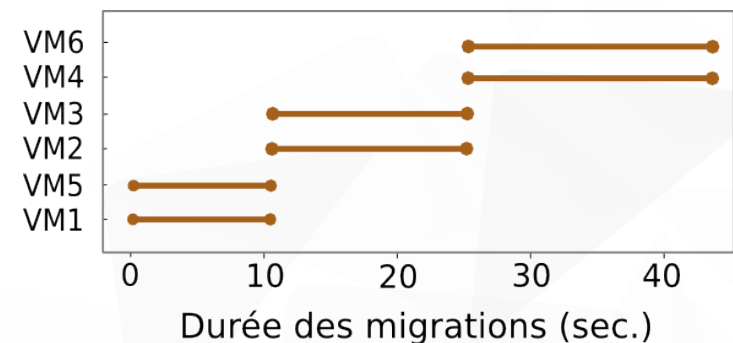
On veut migrer le plus rapidement possible



- ▶ Allouer la bande passante maximale pour chaque migration
- ▶ Paralléliser sans saturer le réseau

Déterminer pour chaque migration

- ▶ La bande passante à allouer
- ▶ Sa durée théorique
- ▶ Le moment où la lancer




- ▶ Parallélisme dépendant de la topologie
- ▶ Groupement par durées

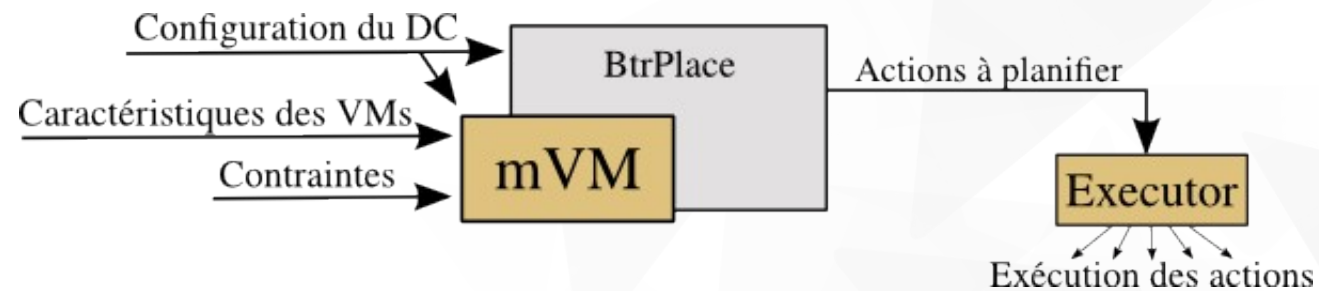
État de l'art

- ▼ Solutions proposées : [Entropy, BtrPlace, Memory Buddies, CloudSim, ..]
 - ▼ Réseaux non-bloquant
 - ▼ Workload ignorées
 - ▼ Parallélisation abusive ou inadaptée
- ▼ Conséquences :
 - ▼ Sous-estimation des durées
 - ▼ Migrations inutilement longues
 - ▼ Réduction des performances des VM

Solution

mVM : Un ordonnanceur de migrations

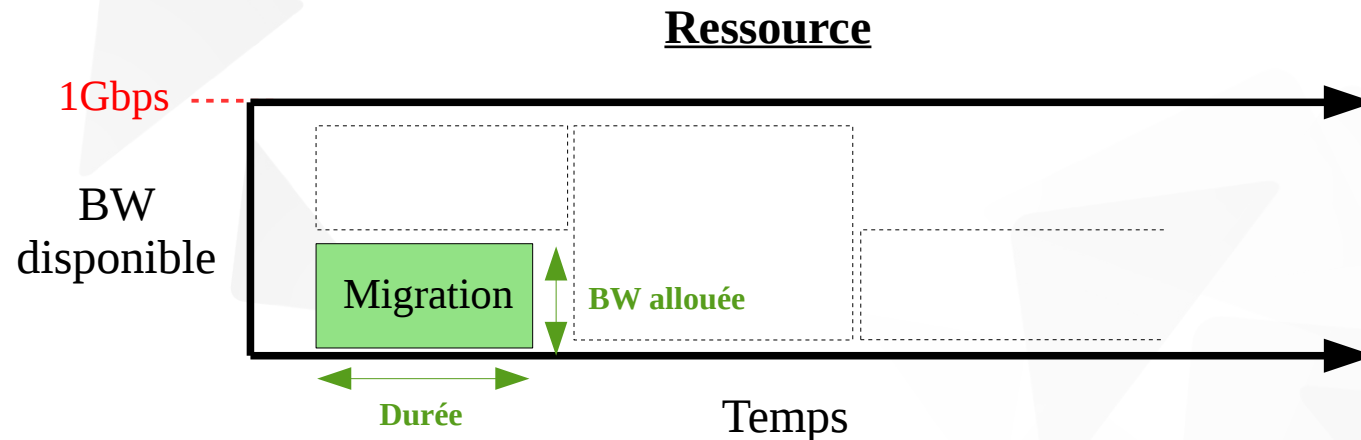
- ▼ Se substitue à l'ordonnanceur de  BtrPlace
- ▼ Propose un nouveau modèle d'ordonnancement
 - ▼ Modèle réseau
 - ▼ Modèle de migration
- ▼ ~ 1600 lignes de code



Modélisation réseau

Implémentation via des contraintes « cumulative » :

- ▼ **Placer des tâches à hauteurs et durées variables sur des ressources limitées.**
- ▼ 1 tâche => 1 migration ; ressources => éléments réseaux
- ▼ 2 ressources par lien : bande passante montante et descendante (full-duplex)
- ▼ 1 ressource par switch bloquant : fond de panier



- ▼ Permet d'établir le lien entre durée de migration et bande passante à allouer

Contraintes annexes

Ajout de **contraintes** permettant de **contrôler l'ordonnancement**

- ▼ **Contraintes temporelles :**

- ▼ `sync (vm[1-4]);`

- ▼ `seq (vm[5,8]);`

- ▼ `before (vm-1,vm-7);`

- ▼ **Contrainte énergétique :** (Implémentation du **modèle énergétique** dérivée de [*Liu et al., Cluster'13*])

- ▼ `powerBudget (500 Watts, [22:00-06:30]);`

2 objectifs

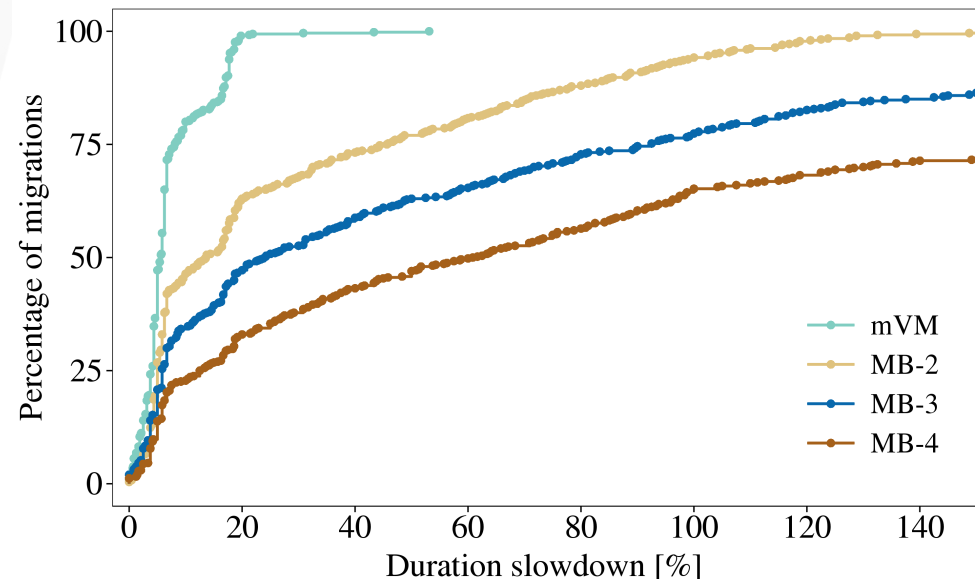
- ▼ Minimiser la somme des temps de fin de migration :
 - ▼ Migrer chaque VM le plus rapidement possible
 - ▼ Assurer un faible temps de complétion
- ▼ Minimiser la consommation énergétique :
 - ▼ S'adapter à l'utilisation d'énergie renouvelable
 - ▼ Variabilité du coût de l'énergie

Évaluation : 50 scénarios aléatoires

Configuration :

- ▼ 4 serveurs
- ▼ 10 VMs
- ▼ Placements aléatoires

Durée des migrations VS ordonnancement séquentiel



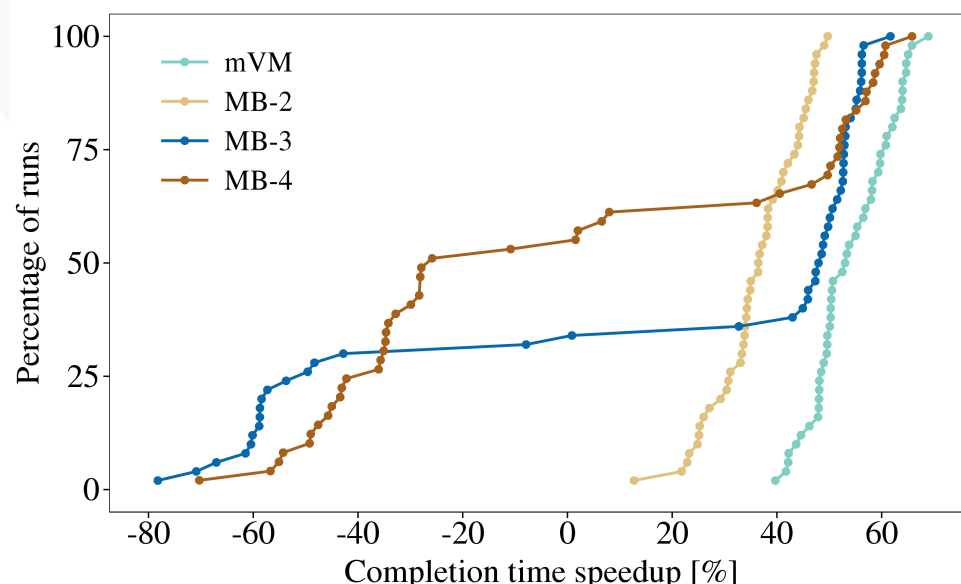
| Scheduler | mVM | MB-2 | MB-3 | MB-4 |
|-------------------------------|--------|---------|---------|---------|
| Tps de migration moyen (sec.) | 45,55 | 57,22 | 113,2 | 168,6 |
| Slowdown moyen (%) | 7,35 % | 29,69 % | 141,3 % | 259,2 % |

Évaluation : 50 scénarios aléatoires

Configuration :

- ▼ 4 serveurs
- ▼ 10 VMs
- ▼ Placements aléatoires

Temps de complétions VS ordonnancement séquentiel

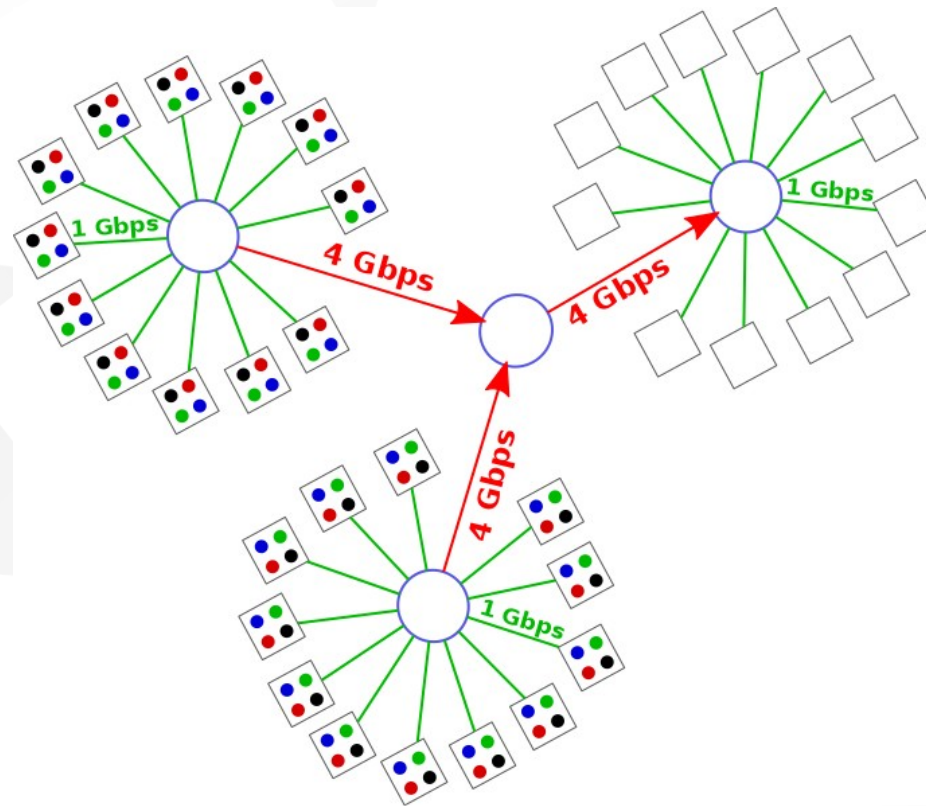


| Scheduler | mVM | MB-2 | MB-3 | MB-4 |
|--------------------------------|---------|---------|---------|---------|
| Tps de completion moyen (sec.) | 212,8 | 295,9 | 394,6 | 479,4 |
| Speedup moyen (%) | 54,18 % | 36,42 % | 15,94 % | -2,64 % |

Évaluation : établir un seuil de puissance

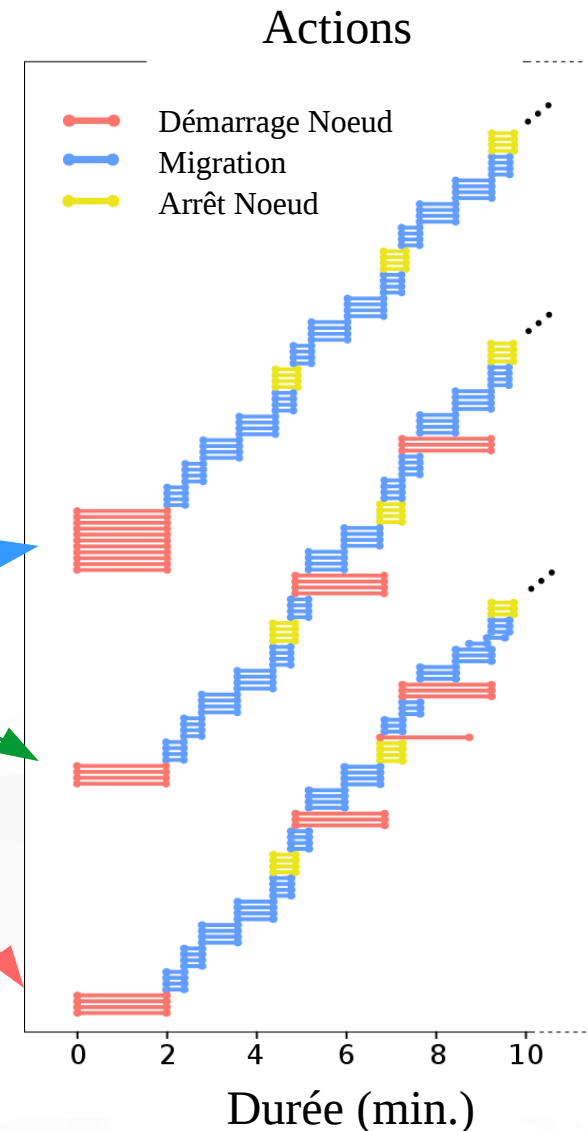
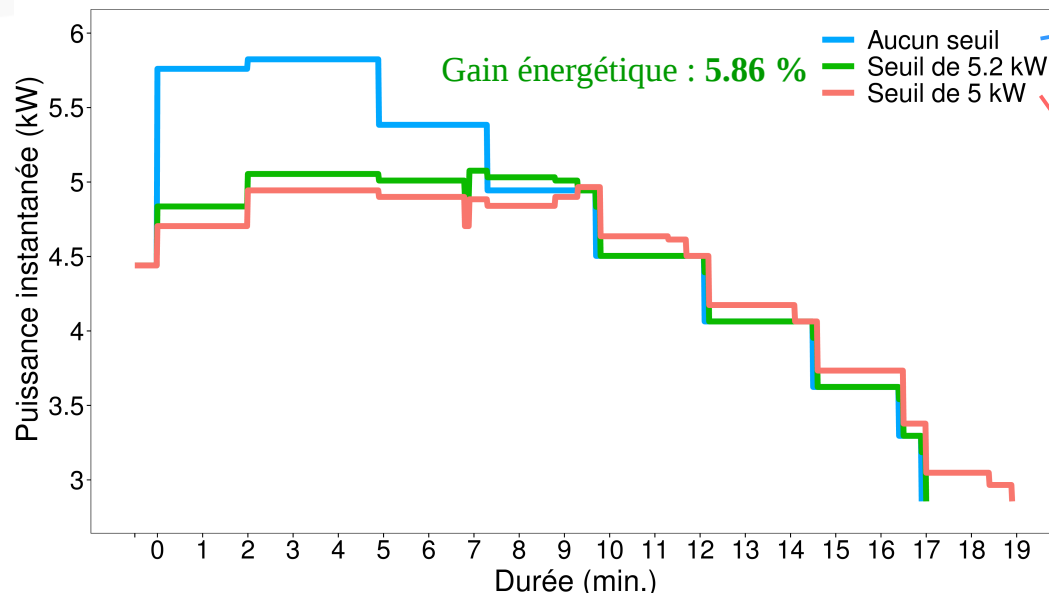
Scénario de décommissionnement :

- ▼ 3 * 12 serveurs (2 racks vers 1)
- ▼ 4 VM par serveur



Évaluation : établir un seuil de puissance

- ▼ Migrations 4 par 4, parallélisme optimal
 - ▼ Report des actions de boot
- ▼ 2 min. supplémentaires pour un seuil à 5kW



Ordonnancement de migrations

- ▼ mVM considère la charge mémoire et le réseau
 - ▼ Ordonnanceur de migrations précis (> 90 %)
 - ▼ Migrations 20.4 % plus rapides que Memory Buddies
- ▼ Contrôle de l'ordonnancement via des contraintes haut niveaux
 - ▼ Synchronisation, séquentialisation / parallélisation
 - ▼ Gestion énergétique
 - ▼ contraintes de « power capping »
 - ▼ 20 % d'énergie sauvegardée lors d'un scénario de décommissioning

Travaux futurs

- ▼ Intégration de la **problématique de placement**
=> Décisions de placement tenant compte de l'ordonnancement
- ▼ **Downtime** contrôlable => **variable** du modèle
- ▼ Papier accepté à UCC'15
=> Code open-source

Ordonnancement contrôlé de migrations à chaud

*Vincent Kherbache, Fabien Hermenier,
Eric Madelaine*

