Modèle bi-niveau pour la tarification de ressources de cloud computing

Arnaud Laurent[1], Luce Brotcorne[2], Bernard Fortz[3]

[1] IMT Atlantique, LS2N
arnaud.laurent@imt-atlantique.fr
[2] Inria Lille
luce.brotcorne@inria.fr
[3] Université libre de Bruxelles
bernard.fortz@ulb.be

Mots-clés : *bi-niveau, cloud computing, tarification*

1 Introduction

Le terme de cloud computing utilisé pour la première fois dans les années 1990 représente l'abstraction des ressources de calculs et de stockage via le recours à des serveurs distants. Le cloud computing a pris de l'ampleur dans le domaine informatique ces dernières années et est utilisé dans de nombreux domaines. Nous considérons dans cette étude le cas où des ressources dans le cloud sont louées à des utilisateurs. Deux types de facturation sont généralement utilisés :

— la tarification "Pay As You Go" (PAYG) où l'utilisateur ne paye que les périodes durant lesquelles il consomme une ressource.
— le système d'abonnement, où l'utilisateur paye un forfait ne dépendant pas de son utilisation sur une période plus longue.

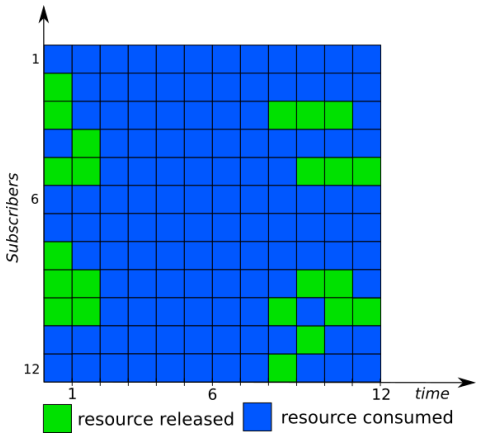
Ces deux tarifications qui cohabitent possèdent certaines limites :

— Les ressources sont souvent sous-utilisées sur certaines périodes lorsque qu'elles sont louées avec le système d'abonnement.
— Le Cloud Service Provider (CSP) n'a pas de levier d'action afin de guider la consommation de ses utilisateurs. Les tarifs étant fixes, le CSP ne peut par exemple pas encourager ses utilisateurs à utiliser leurs ressources durant des périodes où la consommation énergétique est moindre (par exemple, la nuit).

À partir de ces observations, nous proposons un nouveau modèle de tarification permettant aux abonnés de relâcher leurs ressources en échange d'une compensation financière, afin qu'elles soient revendues à des utilisateurs PAYG.

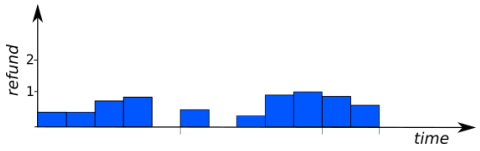
2 Problème

Afin de déterminer les prix de ventes et de compensation pour les ressources pour chaque période, nous proposons un modèle d'optimisation bi-niveau. L'optimisation bi-niveau est un sous-domaine de l'optimisation mathématiques où une partie des variables du modèle principal (meneur) correspondent a une solution optimale d'autres problèmes d'optimisation (suiveurs). Dans le modèle que nous proposons, le problème du meneur correspond à la tarification des ressources pour chaque période de temps ainsi que la définition des compensations financières allouées sur chaque période pour relâcher une ressource. L'objectif du meneur est de maximiser ses gains tout en assurant une bonne qualité de service pour tous les utilisateurs. Pour les problèmes des suiveurs, on distingue deux types d'utilisateurs :

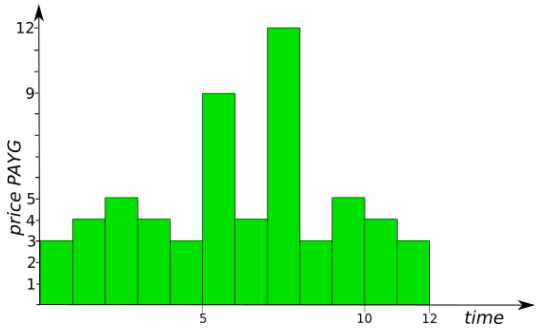


(a) Ressources relâchées par les abonnés

(b) Diagramme de Gantt de la consommation de ressources par les utilisateurs PAYG



(c) Prix de la compensations pour les abonnés en fonction du temps



(d) Prix de vente aux utilisateurs PAYG en fonction du temps

FIG. 1 – Exemple de solution optimale pour une instance avec 12 utilisateurs PAYG et 12 abonnés sur 12 périodes

— Les utilisateurs PAYG, qui ont pour objectif de consommer une ressource pendant plusieurs périodes consécutives, dans une fenêtre de temps donnée, au prix le plus bas.
— Les abonnés, ayant pour objectif de récupérer un maximum de compensations en relâchant leurs ressources sur des périodes peu contraignantes. Nous considérons un coût pour chaque abonné lié a un relâchement de ressources, en fonction de la volonté de l'abonné de changer sa consommation et de son niveau de sollicitation de la ressource sur la période.

Un exemple de solution optimale pour une petite instance est donné dans la Figure 1. On constate dans cette solution que les ressources relâchées par les abonnés (Figure 1a) sont entièrement consommées par les utilisateurs PAYG (Figure 1b).

3 Résolution

Afin de déterminer la tarification optimale pour le CSP, nous modélisons chacun des trois problèmes en variables entières et réelles pour le meneur et en variables réelles pour les suiveurs. Nous remplaçons ensuite les modèles des suiveurs par leur conditions d'optimalité grâce au théorème de dualité forte. On peut alors intégrer les modèles des suiveurs au modèle du meneur. Nous obtenons alors une formulation linéaire mixte après une linéarisation des produits des variables du meneur et des suiveurs.

Nous avons testé la résolution de ce modèle sur un ensemble d'instances générées à partir de données réelles. Une analyse de sensibilité sur plusieurs paramètres, notamment la capacité initiale en ressources de calculs du CSP est effectuée, et une approche de résolution par horizon roulants est aussi étudiée.