

Optimisation de la performance et de l'énergie dans un cloud.

Responsables : Hind Castel (Professeur à Télécom SudParis), Emmanuel Hyon (MCF Université Paris Ouest, Nanterre/LIP6).

Lieu du stage : Université Paris Saclay/Télécom SudParis (à Evry),

Contacts : H. Castel, hind.castel@telecom-sudparis.eu , E. Hyon emmanuel.hyon@lip6.fr .

Niveau : stage de Master 2

Durée : 6 mois

Compétences requises : connaissances en probabilités indispensables, et l'une des connaissances suivantes est requise (les compétences manquantes seront acquises au cours du stage):

- Recherche Opérationnelle et Optimisation ;
- Chaîne de Markov et Files d'attente ;
- Processus de Décision Markovien et méthodes stochastiques pour l'optimisation.

Financement : Oui suivant convention en vigueur.

Ce stage est financé par le projet BOBIBEE (Bornes et Bid Prices dans le But d'Economiser Energie) du programme PGMO de l'université Paris Saclay.

Suite : Une poursuite en thèse sur le même sujet est envisagée.

Description du sujet :

Ces dernières années, l'augmentation à la fois du nombre et de la taille des centres de données (data-centers) qui gèrent un ensemble croissant de données et de services (services de cloud), pose un problème énergétique crucial. En effet, la consommation électrique de ces centres sert à la fois à approvisionner les serveurs en énergie et simultanément à approvisionner l'infrastructure. Cette consommation peut être réduite car aujourd'hui, pour assurer la qualité de service des applications hébergées, les centres de données sont souvent généralement sur-dimensionnés. En vue d'améliorer le taux d'utilisation des serveurs, les centres de données ont déployé des systèmes de virtualisation dynamiques qui apportent la possibilité d'adapter les ressources (le nombre de machines virtuelles activées sur le serveur) à la charge. En effet, la consommation électrique d'un serveur n'est pas proportionnelle à son activité. Cette virtualisation offre la possibilité d'une réduction de l'énergie consommée, tout en maintenant un niveau de qualité de service pour les utilisateurs [Guohui2010] mais, cela nécessite de trouver les politiques d'adaptation des ressources appropriées [Benoit2017]. Ces politiques ont d'autant plus d'intérêt que la capacité d'économiser de l'énergie sera un point clef de l'efficacité d'un centre de données dans les années à venir.

Le but de ce stage est de trouver un mode efficace de gestion dynamique des ressources afin de réduire la consommation énergétique tout en gardant un niveau de qualité de service donné. Nous nous intéresserons plus particulièrement aux politiques d'activation et de désactivation des machines virtuelles. Nous chercherons à optimiser un coût moyen (coût par unité de temps) global et nous considérerons un problème d'optimisation stochastique, car, pour représenter les clouds, la plupart des travaux du domaine considèrent un système de file d'attente (avec des serveurs dans différents états [Kandi2017]). Dans ce cadre théorique on peut soit rechercher la politique optimale, soit s'intéresser à des restrictions, a priori sous optimales, mais extrêmement simples à mettre en pratique : les politiques dans lesquelles l'activation et la désactivation des machines virtuelles dépend d'un seuil ou de plusieurs seuils (politiques à hysteresis) [Lui1999].

Travail à réaliser

Le travail à réaliser se focalisera sur la recherche des politiques d'activation et de désactivation et sera abordé selon plusieurs axes. D'une part, par la recherche des politiques optimales avec des méthodes d'optimisation stochastique classiques et les Processus de Décisions Markovien. D'autre part, on se concentrera surtout sur les politiques à hysteresis et le calcul de la valeur des seuils. On cherchera à adapter des techniques d'agrégation ou de borne sur les coûts pour pouvoir appliquer

des techniques d'optimisation combinatoire classiques (« Branch and Bound » par exemple). Enfin, on cherchera des approximations en faisant appel à des formalismes issus du domaine du « revenue management » [Talluri 98] qui cherchent à déterminer l'opportunité de vendre un produit (place d'avion, nuit d'hôtel) pour approcher les seuils au moyen de la programmation linéaire.

Références

[Guohi2010] “*The impact of virtualization on network performance of amazon ec2 data center*”, W. Guohui and Ng, Eugene, IEEE INFOCOM, 2010.

[Benoit2017] “*Reducing the energy consumption of large scale computing systems through combined shutdown policies with multiple constraints*”, A. Benoit, L. Lefèvre, A.-C. Orgerie, and Issam Rais, International Journal of High Performance Computing Applications, 2017.

[Kandi2017] “*Analysis of performance and energy consumption in the cloud*”, M. Kandi, F. Ait Salaht, H. Castel and E. Hyon, EPEW 2017.

[Lui 1999] “*Stochastic complement analysis of multi-server threshold queues with hysteresis*”, J. C.S. Lui and L. Golubchik, Performance Evaluation, 1999.

[Talluri 98] “*An Analysis of Bid-Price Controls for Network Revenue Management*”, K. Talluri j G. van Ryzin, Management Science, 1998.