

Apprentissage machine pour l'optimisation énergétique des réseaux cellulaires hétérogènes sans-fil : une approche bandit à bras multiples

Navikkumar MODI¹, Philippe MARY², Christophe MOY¹

¹CentraleSupélec, IETR/SCEE, UMR CNRS 6164, 35510 Cesson-Sévigné, France

²INSA de Rennes, IETR, UMR CNRS 6164, F-35043 Rennes, France

{navikkumar.modi, christophe.Moy}@centralesupelec.fr, philippe.Mary@insa-rennes.fr

Résumé – Dans ce papier, nous proposons un algorithme d'apprentissage machine afin d'améliorer l'efficacité énergétique (EE) des réseaux cellulaires hétérogènes. Le but de la stratégie est d'apprendre quelle est la configuration optimale de stations de bases allumées et éteintes maximisant l'EE du réseau. Nous modélisons dans un premier temps les variations de charge du trafic par un processus de décision markovien (MDP), puis nous présentons un algorithme de la classe des UCB (Upper Confidence Bound) basé sur la théorie des bandits pour apprendre le meilleur déploiement. De plus, afin de palier à la perte de récompense due à la variation de la statistique de charge au cours du temps, nous intégrons la notion de transfert d'apprentissage (TL) se basant sur la solution obtenue lors de périodes d'apprentissage passées. Les simulations montrent que la solution proposée permet d'améliorer grandement l'EE du réseau par rapport aux techniques existantes.

Abstract – A machine learning algorithm is proposed in this paper to improve the energy efficiency (EE) of the heterogeneous cellular networks. The strategy aims at learning what is the optimal base stations (BS) configuration, i.e. ON or OFF, maximizing EE of the network. First, we formulate the traffic load variations as a Markov decision process (MDP), and present an algorithm from the upper confidence bound (UCB) class based on multi-armed bandit (MAB) theory to learn the best deployment of BS. Moreover, to cope with initial reward loss due to the variation of the load, a transfer learning (TL) framework for MAB, which benefits from the transferred knowledge observed in historical periods, is proposed. Extensive simulations demonstrate that the proposed solution can significantly increase EE of cellular networks.

1 Introduction

Les « data centers » et les réseaux cellulaires sont les sources principales de consommation énergétique de l'industrie de l'information et de la communication. Leur consommation énergétique représente 2% à 10% de la consommation globale [1]. Dans les réseaux cellulaires, la consommation énergétique des stations de bases (BSs) représente 60% à 80% de la consommation totale du réseau [1]. De plus la consommation statique d'une macro BS, la puissance utilisée pour alimenter et refroidir les équipements, peut monter jusqu'à 700 W [2]. D'un autre côté et à cause de la variation de la charge du trafic au cours de la journée, certaines BSs pourraient être éteintes et les utilisateurs ré-associés à d'autres BS afin de diminuer la consommation énergétique du réseau et donc augmenter son efficacité énergétique (EE).

De récents travaux se sont intéressés à l'allumage et extinction des BSs en fonction de la variation de la charge du réseau, e.g. [1, 3, 4]. La recherche de la configuration optimale des BS, i.e. allumées ou éteintes, maximisant l'EE du réseau est un problème NP-complet. Aussi certains auteurs, comme en [5], ont évalué l'impact d'une séquence de commutation déterministe des BS sur les probabilités de blocage et de coupure. Les auteurs de [1] ont modélisé le problème d'extinction des BSs comme un processus de décision markovien (MDP)

et ont proposé un algorithme d'apprentissage par renforcement qui s'adapte aux variations de la charge du trafic. De plus, les auteurs proposent l'algorithme TACT (Transfer Actor CriTic) qui utilise les connaissances acquises dans le passé à travers un transfert d'apprentissage (TL). Les travaux précédents requièrent des algorithmes complexes pour converger vers une solution.

Dans ce papier, nous modélisons le problème de maximisation de l'EE par allumage et extinction des BS à l'aide de la théorie des bandits à bras multiples (MAB). Nous utilisons et adaptons un algorithme de type UCB que nous avons proposé en [6] dans le cadre de l'accès opportuniste au spectre (AOS) avec prise en compte de la qualité du canal, afin de trouver la configuration de déploiement qui maximise l'EE du réseau. Les simulations montrent que la solution proposée est meilleure que TACT et se rapprochent des performances d'une recherche exhaustive. La section 2 présente le modèle du système et le problème de maximisation de l'EE. La section 3 formule le problème avec la théorie MAB et présente le concept TL ainsi que l'algorithme TRQoS-UCB (Transfer Restless Quality of Service UCB) proposé pour résoudre le problème. Les résultats de simulations sont présentés en section 4 et nous concluons en section 5.

2 Modèle et hypothèses

2.1 Modèle de réseau

Soit un réseau cellulaire composé de macro et micro cellules dont l'ensemble des BSs $\mathcal{Y} = \{1, 2, \dots, Y\}$ est déployé dans \mathbb{R}^2 . On considère également qu'il existe un contrôleur centralisé gérant l'état des BSs, i.e. allumée, éteinte, et qui a la connaissance du trafic à l'instant t . A chaque itération n , une requête de transmission à la position x arrive suivant une loi de poisson avec un taux d'arrivée $\Lambda(x, n)$ par unité de temps. La taille des fichiers est distribuée suivant une loi exponentielle de paramètre $1/h(x, n)$. La charge de trafic instantanée est donc $L(x, n) = \Lambda(x, n)/h(x, n)$.

On considère le lien descendant et aucune interférence intra-cellulaire. Le rapport signal à bruit plus interférence (RSBI) en x à l'instant n s'écrit

$$\text{RSBI}_k(x, n) = \frac{g^k(x, n)P_k^{tx}}{\sum_{m \in \mathcal{Y}_n^{on} \setminus \{k\}} g^m(x, n)P_m^{tx} + \sigma^2} \quad (1)$$

où σ^2 est la puissance du bruit, $g^k(x, n)$ est le gain moyen du canal incluant la perte en espace libre, P_k^{tx} est la puissance de transmission de la k -ième BS. Le premier terme du dénominateur est la somme des puissances reçues à l'instant n en x des BSs interférentes. Enfin, $\mathcal{Y}_n^{on} \in \mathcal{Y}$ est l'ensemble des BSs actives à l'instant n . Soit B_a la bande passante du système, le taux de service de l'utilisateur en x de la part de la BS k à l'instant n est :

$$\Theta_k(x, n) = B_a \cdot \log_2(1 + \text{RSBI}_k(x, n)). \quad (2)$$

La charge de la BS k est la portion du temps requise pour assurer le trafic dans la cellule et a pour expression $\rho_k(n) = \sum_{x \in \mathcal{I}_k^n} (L(x, n)/\Theta_k(x, n))$ où $\mathcal{I}_k^n (k = 0, 1, 2, \dots)$ est la couverture de la BS k à l'instant n . Enfin, nous considérons un modèle linéaire de la puissance totale consommée par la BS k à l'instant n qui s'écrit

$$P_T^k(n) = a_k P_k^{tx}(n) + P_f^k \quad (3)$$

avec P_f^k la puissance statique consommée, indépendante de $P_k^{tx}(n)$, et a_k est un facteur d'échelle de puissance incluant les pertes de l'amplificateur.

2.2 Formulation du problème

La maximisation de l'EE dans un réseau cellulaire par allumage et extinction des BSs peut-être formulé comme suit

$$\mathcal{Y}_n^{on*} = \arg \max_{\mathcal{Y}_n^{on}} \left[\sum_{k \in \mathcal{Y}_n^{on}} \frac{\sum_{x \in \mathcal{I}_k^n} \Theta_k(x, n)}{P_T^k(n)} \right] \quad (4)$$

$$s.t. \quad 0 \leq \rho_k(n) \leq \rho_{th}, \forall k \in \mathcal{Y}_n^{on} \quad (4a)$$

$$\Theta_k(x, n) \geq \Theta^{\min}, \forall x \in \mathcal{I}_k^n, \forall k \in \mathcal{Y}_n^{on} \quad (4b)$$

où $\rho_{th} \leq 1$ en (4a) borne la charge afin de maintenir un compromis entre la stabilité du système et l'EE [1, 7]. La contrainte en (4b) assure une QoS minimale pour chaque utilisateur.

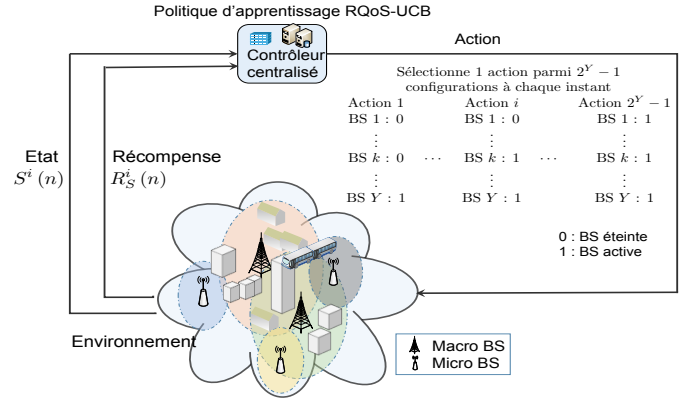


FIGURE 1 – Apprentissage par renforcement pour MAB.

3 MAB pour réseaux cellulaires « verts »

3.1 Model avec des bandits à bras multiples

On définit un n -uplet $\mathcal{M} = \langle \mathcal{S}, \mathcal{K}, \mathcal{P}, R \rangle$, où \mathcal{S} , \mathcal{K} , \mathcal{P} et R désignent l'espace des états, l'espace des actions, la matrice de transition et une fonction de récompense respectivement. La figure 1 illustre la modélisation de la commutation des BSs comme un problème de bandit à bras multiples. A chaque itération n , le contrôleur central choisit une action $A(n) = [A_1(n), \dots, A_Y(n)]^T$ avec $A_k(n) = 1$ si la BS k est allumée et 0 sinon. Cette action particulière conduit à un état de charge du réseau, $S^i(n) \in \{0, 1\}$ où $i \in \mathcal{K}$ représente l'index de l'action choisie avec $|\mathcal{K}| = 2^Y - 1$ et $S^i(n) = 1$ si toutes les contraintes en (4a) et (4b) sont satisfaites et 0 sinon. De plus, le choix de l'action i conduit à une récompense qui est l'EE de la configuration, i.e. $R_S^i(n) = \text{EE}(n)$. L'état et la récompense sont renvoyés dans le contrôleur afin que celui-ci prenne une décision quant à la prochaine action à tenir. Finalement, $\mathcal{P}^i = \{P_{k,l}^i, k, l \in \mathcal{S}, i \in \mathcal{K}\}$ représente la probabilité de transition du processus markovien ainsi défini.

3.2 Politique d'apprentissage

La politique RQoS-UCB (Restless QoS-UCB) proposée dans [6] mais dans un contexte AOS, est susceptible d'apprendre efficacement la meilleure configuration des BSs pour la maximisation de l'EE. La statistique de charge du réseau change au cours du temps dans la journée, mais également d'un jour sur l'autre. Par exemple, à une heure donnée de la journée, le taux d'arrivées des utilisateurs dans le réseaux peut varier d'un jour sur l'autre. Le transfert d'apprentissage consiste à utiliser les connaissances acquises durant les jours précédents afin de permettre à l'algorithme de repartir d'une connaissance sur la charge du réseau et donc d'avoir un *a priori* sur les actions intéressantes et celles qui le sont moins. C'est ce qu'illustre la figure 2 où pour chaque heure de la journée, la configuration apprise le jour précédent est appliquée au jour courant.

L'apprentissage s'effectue selon un bloc b décomposé en 3 sous-blocs, i.e. SB1, SB2 et SB3 [6]. Cette division a pour

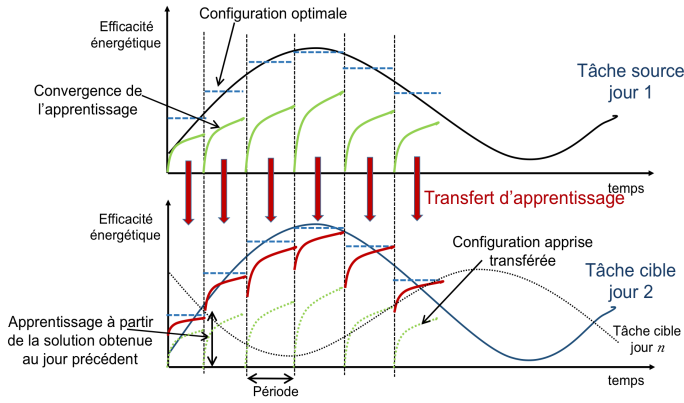


FIGURE 2 – Principe du transfert d'apprentissage.

but de gérer la caractéristique « restless » de la modélisation MAB. La récompense obtenue dans l'état $S^{i,h}$ avec la configuration $i \in \mathcal{K}$ du réseau pendant la période précédente H_2^i est $R_S^{i,h}(H_2^i)$, où H_2^i est le nombre d'itérations pendant lequel l'action i a été sélectionnée dans le SB2 dans la tâche source. A la fin d'un bloc b , TRQoS-UCB retourne l'index $i \in \mathcal{K}$ de l'action à appliquer lors du prochain bloc $b + 1$, qui maximise $B^{i,h}(n_2, T_2^i) \forall i \in \mathcal{K}$ tel que :

$$B^{i,h}(n_2, T_2^i) = \bar{S}^{i,h}(T_2^i) - Q^{i,h}(n_2, T_2^i) + C^{i,h}(n_2, T_2^i). \quad (5)$$

Les termes $\bar{S}^{i,h}$, $Q^{i,h}$ et $C^{i,h}$ sont définis en (6) dans lesquels $M^{i,h}(n_2, T_2^i) = G_{\max}^S - G_S^{i,h}(T_2^i)$, $G_{\max}^S = \max_{i \in \mathcal{K}} G_S^i(T_2^i)$, $G_S^{i,h}(T_2^i) = 1/T_2^i \sum_{k=1}^{T_2^i} R_S^i(k) + 1/H_2^i \sum_{k=1}^{H_2^i} R_S^{i,h}(k)$.

La figure 3 illustre l'algorithme. Un état ζ^i récurrent est défini qui permet d'entrer dans SB2, phase pendant laquelle les termes en (6) sont mis à jours. Pendant SB1, on se contente d'observer les états $S^i(n)$, et la récompense associé $R_S^i(n)$.

4 Résultats de simulations

On considère un réseau cellulaire hétérogène avec 5 BSs macros et 5 micros déployées aléatoirement sur une surface de $5 \times 5 \text{ km}^2$. La table 1 rassemble les paramètres utilisés pour les simulations. On étudie l'efficacité de la solution proposée lorsque la charge de trafic fluctue périodiquement. Les simulations sont faites en moyennant sur 100 réalisations de réseaux, chacune consistant en 3000 itérations sur une période d'une heure. Notre solution sans et avec TL, i.e. RQoS-UCB et TRQoS-UCB respectivement, est comparée avec les algorithmes ACT and TACT (ce dernier étant l'algorithme "Actor CriTic" (ACT) avec un transfert d'apprentissage) en [1], un algorithme glouton décentralisé [7] et une recherche exhaustive.

La figure 4 montre tout d'abord que l'EE varie de manière périodique et est maximale lors des périodes de faibles charges, i.e. de 1h à 8h du matin, et est minimale lors des pics de trafic, de 12h à 22h. On remarque que lors des pics de trafic tous les algorithmes ont les mêmes performances et sont confondus avec l'EE obtenue avec une recherche exhaustive. Seul l'algorithme

TABLE 1 – Paramètres de simulation

Paramètres	Valeurs
Surface	5km × 5km
Puissance de transmission	Macro : 20W, Micro : 1W
Puissance opérationnelle	Macro : 865W, Micro : 38W
Bande passante	1.25MHz
Modèle de propagation	COST 231 modifié
Taux d'arrivée $\Lambda(x, n)$	$5 \cdot 10^{-5} \cdot [\cos(2\pi(n+10)/24)] + 10^{-4}$
Taille fichier $1/h(x)$	100Kbyte
Seuil de charge ρ_{th}	0.6
Débit bit min Θ_{\min}	122kbps
Paramètres d'exploration	$\alpha = 0.25$ and $\beta = 0.32$

en [7] échoue à maintenir une bonne EE sur cette période. Ces performances similaires durant les gros pics de trafic sont dues au fait que l'ensemble des BSs est sollicité pour satisfaire l'importante charge réseau et donc l'apprentissage est moins nécessaire. A contrario, lors des périodes de plus faibles charges, i.e. la nuit, le nombre de BSs nécessaires pour satisfaire les utilisateurs est moins important et l'apprentissage du meilleur déploiement fait sens. On constate que TRQoS-UCB surpasse son homologue TACT, i.e. $4.3 \cdot 10^6$ bits/Joule vs $3 \cdot 10^6$ bits/Joule respectivement sur la plage horaire 2h-6h. De plus, on constate l'intérêt du TL puisque TRQoS-UCB surpasse d'environ 23% l'EE de RQoS-UCB. Enfin, on remarque que TRQoS-UCB est à moins de 5% de la solution optimale sur cette plage horaire.

La figure 5 décrit l'évolution de l'énergie économisée en pourcentage par rapport à la situation où toutes les BSs seraient allumées sur le temps d'une journée. On remarque que TRQoS-UCB permet d'économiser plus de 70% d'énergie sur la plage horaire 2h-6h, ce qui est mieux que RQoS-UCB, pour lequel 60% d'économie d'énergie est réalisé sur la même plage horaire. De même, TACT et ACT économisent à peu près la même énergie, i.e. 60% et 58% respectivement, sur la période de faible charge. De plus, TRQoS-UCB est à moins de 5% des performances de la solution optimale. On remarque que sur la période de forte charge, 12h-22h, tous les algorithmes, sauf l'algorithme glouton décentralisé, ne permettent pas une économie d'énergie significative par rapport au déploiement total. Le comportement de l'algorithme glouton ici éclaire ses performances sur la figure 4. C'est la solution qui permet la plus grande économie d'énergie, i.e. elle atteint quasiment la solution optimale sur la plage horaire 2h-8h, et économise près de 20% d'énergie sur 12h-22h. En effet, cette solution conduit à éteindre plus de BS au détriment de l'efficacité spectrale du réseau, contrairement aux autres solutions, et donc conduit à une EE moins bonne également.

Enfin, lorsque le nombre de BSs croît, l'apprentissage se fait sur un ensemble dont la cardinalité augmente exponentiellement. Par conséquent, l'algorithme TRQoS-UCB mettra plus de temps pour trouver la configuration optimale. Cependant la complexité algorithmique ne dépend pas du nombre de BSs

$$\bar{S}^{i,h}(T_2^i) = \frac{\sum_{t=1}^{T_2^i} S^i(t) + \sum_{t=1}^{H_2^i} S^{i,h}(t)}{T_2^i + H_2^i}, Q^{i,h}(n_2, T_2^i) = \frac{\beta M^{i,h}(n_2, T_2^i) \log(n_2 + H_2^i)}{T_2^i + H_2^i}, C^{i,h}(n_2, T_2^i) = \sqrt{\frac{\alpha \log(n_2 + H_2^i)}{T_2^i + H_2^i}} \quad (6)$$

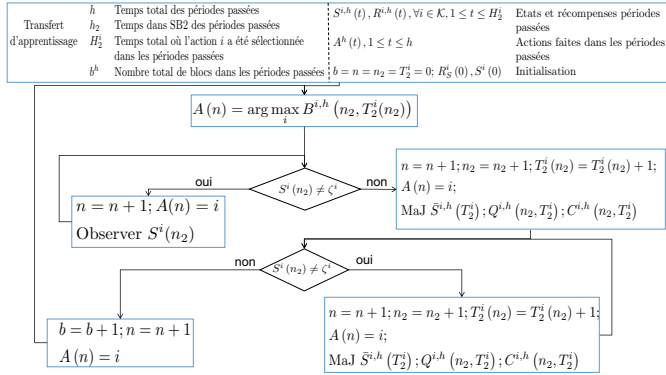


FIGURE 3 – Schéma bloc de la politique TRQoS-UCB

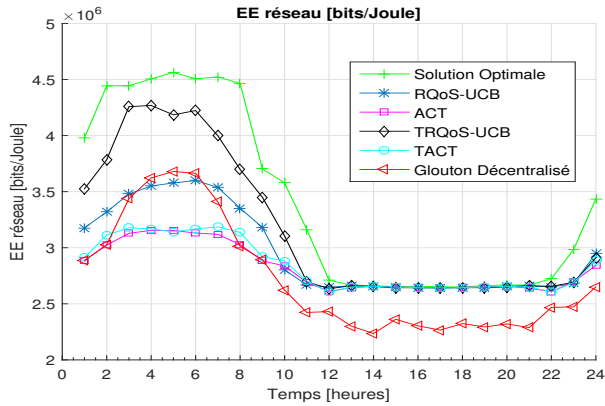


FIGURE 4 – EE du réseau en fonction du moment de la journée comme cela peut se voir en (5).

5 Conclusion

Nous nous sommes intéressés au problème de l'extinction des BS afin de maximiser l'EE d'un réseau cellulaire hétérogène. Afin d'améliorer la convergence de l'algorithme UCB utilisant l'EE comme QoS, nous avons proposé d'utiliser les connaissances acquises lors des jours précédents par un mécanisme de transfert. Les simulations ont montré que notre solution permettait de se rapprocher à moins de 5% de la solution optimale sur la plage horaire de faible trafic. De plus, notre solution permet d'économiser 70% d'énergie sur cette plage, qui est la plage intéressante pour l'apprentissage d'un déploiement.

Remerciements

Ce travail a bénéficié d'une aide de l'Etat attribuée au labex CominLabs et gérée par l'Agence Nationale de la Recherche au titre du

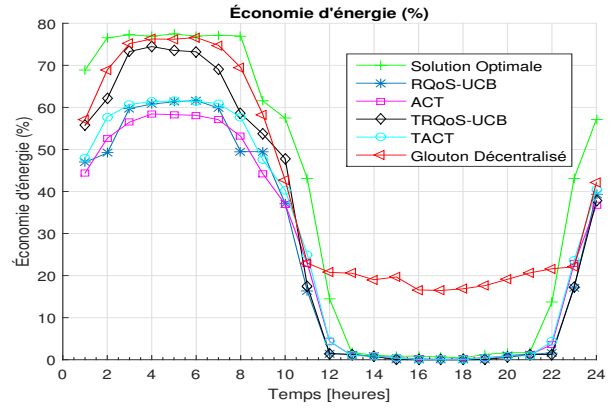


FIGURE 5 – Economie d'énergie moyenne sur une journée par rapport à un déploiement total.

programme « Investissements d'avenir » portant la référence ANR-10-LABX-07-01. Les auteurs remercient également la région Bretagne pour son support sur ces travaux.

Références

- [1] R. Li *et al.*, “TACT : A Transfer Actor-Critic Learning Framework for Energy Saving in Cellular Radio Access Networks,” *IEEE Trans. Wireless Commun.*, vol. 13, no. 4, pp. 2000–2011, April 2014.
- [2] G. Auer *et al.*, “Framework for energy efficiency analysis of wireless networks,” in *Int. Conf. Wireless Comm., Veh. Tech., Inf. Theory, Aero. Elec. Syst. Tech. (Wireless VITAE)*, 2011, pp. 1–5.
- [3] L. Suárez *et al.*, “A multi-criteria BS switching-off algorithm for 5G heterogeneous cellular networks with hybrid energy sources,” *Trans. Emerging Telecom. Tech.*, vol. 27, no. 7, pp. 923–938, 7 2016.
- [4] Z. Li, D. Grace, and P. Mitchell, “Traffic-Aware Cell Management for Green Ultradense Small-Cell Networks,” *IEEE Trans. Veh. Technol.*, vol. 66, no. 3, pp. 2600–2614, March 2017.
- [5] F. Han, Z. Safar, and K. J. R. Liu, “Energy-Efficient Base-Station Cooperative Operation with Guaranteed QoS,” *IEEE Trans. Commun.*, vol. 61, no. 8, pp. 3505–3517, August 2013.
- [6] N. Modi, P. Mary, and C. Moy, “QoS driven Channel Selection Algorithm for Cognitive Radio Network : Multi-User Multi-armed Bandit Approach,” *IEEE Trans. Cognitive Commun. Netw.*, 2017.
- [7] E. Oh, K. Son, and B. Krishnamachari, “Dynamic Base Station Switching-On/Off Strategies for Green Cellular Networks,” *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 2126–2136, May 2013.