



HAL
open science

Adaptation des Réseaux d'Accès Radio Virtualisés à la Mobilité des Utilisateurs

Diala Naboulsi, Assia Mermouri, Razvan Stanica, Hervé Rivano, Marco Fiore

► **To cite this version:**

Diala Naboulsi, Assia Mermouri, Razvan Stanica, Hervé Rivano, Marco Fiore. Adaptation des Réseaux d'Accès Radio Virtualisés à la Mobilité des Utilisateurs. ALGOTEL 2018 - 20èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications, May 2018, Roscoff, France. pp.1-4. hal-01781620

HAL Id: hal-01781620

<https://inria.hal.science/hal-01781620v1>

Submitted on 30 Apr 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Adaptation des Réseaux d'Accès Radio Virtualisés à la Mobilité des Utilisateurs

D. Naboulsi¹ A. Mermouri² R. Stanica² H. Rivano² M. Fiore^{2,3}

¹Concordia University, Montreal, Canada, ²Université de Lyon, INSA Lyon, Inria, CITI, F-69621, Villeurbanne, France
³CNR – IEIT, Corso Duca degli Abruzzi 24, 10129 Torino, Italy

Le développement des techniques de virtualisation change l'architecture des réseaux mobiles. L'allocation de ressources et le traitement du signal peuvent devenir des fonctions logicielles hébergées dans des centres de calculs. La centralisation du calcul des ressources et l'association dynamique des unités de calculs de la bande de base (BBUs pour Base Band Units) aux antennes distantes (RRHs pour Remote Radio Heads) donnent plus de flexibilité aux opérateurs mobiles et permettent des réductions importantes des coûts de fonctionnement. La plupart des travaux sur ces architectures dites CRAN (pour Cloud Radio Access Networks) se placent du point de vue de l'opérateur et étudient des indicateurs de performance réseau. L'impact de ce nouveau paradigme sur l'expérience utilisateur a par contre été négligé. Dans cet article nous nous plaçons de ce point de vue. Nous montrons que l'affectation dynamique du calcul des ressources permise par CRAN génère une nouvelle classe de handover des terminaux mobiles qui peut affecter la qualité de service qui leur est fournie. Nous proposons donc un algorithme qui prend en compte le nombre de handovers pour optimiser l'affectation des BBUs aux RRHs sur un graphe dynamique représentant le système en version oracle et online. Nous montrons que ces deux versions obtiennent des résultats similaires avec des gains de 20% sur deux grands jeux de données réelles sans pour autant induire une sur-charge importante des centres de calcul par rapport à la littérature[†].

1 Introduction

La croissance du trafic Internet mobile [2] pousse les opérateurs à densifier toujours plus leur réseau d'accès radio (RAN). Déployer une station de base (BS) signifie installer une antenne (RRH) sur un point haut et la connecter à une unité de calculs de la bande de base (BBU) qui nécessite, entre autre, un système de climatisation coûteux et limitant les sites potentiels de déploiement.

En s'appuyant sur les méthodes de virtualisation des réseaux, le concept de Cloud-RAN (CRAN) a émergé [3]. La RRH est déployée sur site mais connectée via une fibre optique à un centre de calcul où le traitement du signal, l'allocation de ressource ou la gestion de la mobilité des utilisateurs deviennent des fonctions logicielles exécutées sur des ressources mutualisées. Cela permet un découplage des RRHs et des BBUs. En particulier une BBU peut désormais gérer plusieurs RRHs et l'affectation des RRHs aux BBUs peut se faire de manière dynamique pour s'adapter à la charge du réseau [4].

Les gains énergétiques et économiques [5] ou en performances réseau [6] de CRAN ont été étudiés, mais peu de travaux ont cherché à comprendre l'impact de cette mutualisation de coût sur les utilisateurs. Dans cet article, nous nous intéressons donc aux handovers qui sont particulièrement transformés dans cette nouvelle architecture: si certains handovers classiques sont absorbés lorsqu'ils ont lieu entre deux RRHs qui sont affectées à la même BBU, un nouveau type de handover, dit de reconfiguration des BBU, apparaît lorsqu'un utilisateur potentiellement statique doit changer de BBU. Nous modélisons l'évolution d'une architecture CRAN par un graphe dynamique prenant en compte ces deux notions de handover. L'affectation des RRHs aux BBUs devient un problème de clustering des sommets de ce graphe sous contrainte de capacité qui peut se résoudre par une modification de la méthode de Louvain [8] qui admet une version online. Nous évaluons cette méthode sur des données réelles pour montrer une réduction de plus de 20% du nombre de handovers sans pour autant induire une sur-charge importante des centres de calcul.

[†] Une version plus complète de cet article a été publiée dans la conférence Infocom 2018 [1].

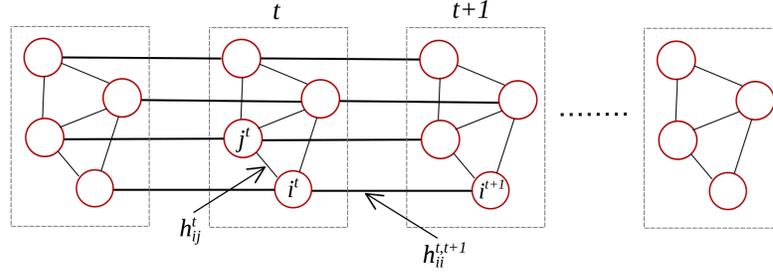


Fig. 1: Modélisation en graphe dynamique

2 Modélisation en clustering de graphe dynamique

Pour représenter la dynamique du trafic dans un réseau cellulaire, nous introduisons un graphe dynamique illustré en Figure 1. Le temps \mathcal{T} est discrétisé par intervalle ΔT_t , qui est le temps minimum entre deux reconfigurations de l'affectation RRH \leftrightarrow BBU. Le graphe comporte un sommet i^t pour chaque RRH i et intervalle de temps t . L'ensemble des sommets du graphe est donc $\mathcal{R} = \{i^t, i \text{ un RRH}, t \in \mathcal{T}\}$.

Il y a un arc e_{ij}^t de poids h_{ij}^t entre i^t et j^t dès qu'il y a h_{ij}^t utilisateurs actifs qui se "déplacent" du RRH i au RRH j pendant l'intervalle de temps t . Dans une architecture classique, cela représenterait autant de handovers dits de mobilités (et notés MHO).

On ajoute un arc $e_{ii}^{t,t+1}$ entre i^t et i^{t+1} de poids $h_{ii}^{t,t+1}$ égal au nombre d'utilisateurs associés au RRH i au passage de l'intervalle t à $t+1$. Si l'affectation de i à une BBU change à ce moment là, cela génère autant de handovers dits de reconfiguration (et notés RHO).

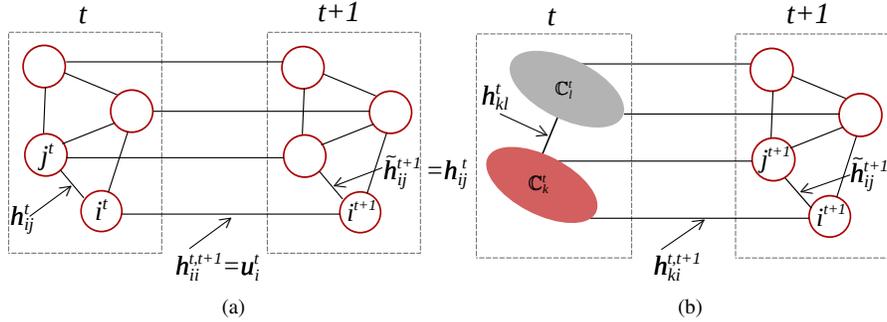
On considère enfin que chaque BBU $k \in \mathcal{C}$ a une capacité limitée c_k . La charge du RRH i à l'instant t est notée d_i^t . A chaque instant t , l'ensemble des RRH affectées à k forme un cluster noté \mathcal{C}_k^t et on a $\sum_{i \in \mathcal{C}_k^t} d_i^t \leq c_k; \forall k, \forall t$.

Le problème que nous traitons consiste alors à former des clusters de RRHs, chacun géré par une BBU, de manière à ce que chaque BBU ait la capacité de traiter le trafic de l'ensemble des RRHs de son cluster et que tous les RRHs soient dans un cluster. On considère que l'on a assez de BBUs pour couvrir l'intégralité du réseau (au pire, une par RRH). L'objectif est alors de minimiser le nombre de handovers subits par les utilisateurs, qu'ils soient MHO ou RHO. Considérant le graphe dynamique défini précédemment, si i^t et j^t n'appartiennent pas au même \mathcal{C}_k^t , il y aura h_{ij}^t MHO au temps t . De même, si $i^t \in \mathcal{C}_k^t$ et $i^{t+1} \notin \mathcal{C}_k^{t+1}$, il y aura $h_{ii}^{t,t+1}$ RHO. Il s'agit donc du problème d'optimisation suivant, où $\mathcal{C}_k = \bigcup_t \mathcal{C}_k^t$.

$$\min \left(\sum_{i^t \in \mathcal{C}_k, j^t \in \mathcal{C}_l} h_{ij}^t + \sum_{i^t \in \mathcal{C}_k, j^t \in \mathcal{C}_l} h_{ii}^{t,t+1} \right) \text{ s.t. } \sum_{i^t \in \mathcal{C}_k} d_i^t \leq c_k; \forall k, \forall t \quad (1)$$

En faisant abstraction de la contrainte de capacité, le problème se réduit à une détection de communauté classique fondée sur la maximisation d'une métrique de modularité Q , problème NP-difficile que l'on peut approcher avec la méthode de Louvain [8]. Cette méthode peut se modifier pour prendre en compte la contrainte de capacité, dans une version avec oracle connaissant tout le graphe, et une version "online". Les développements algorithmiques (ainsi qu'une version plus précise de la modélisation et de l'évaluation des algorithmes) sont présentés dans la version longue de cet article [1]. L'idée principale de la modification de la méthode de Louvain consiste simplement à rajouter au test de pertinence de l'ajout d'un sommet du graphe à une communauté, la vérification que la contrainte de capacité reste satisfaite.

Le principe de la version online de l'algorithme est illustré dans les Figures 2(a) et 2(b). L'idée principale consiste à traiter l'intervalle t en se fondant sur une estimation du temps $t+1$ puis d'agréger les noeuds d'un même cluster pour reconstruire un graphe qui permettra de traiter l'intervalle suivant lorsque l'information exacte sera disponible. Dans les évaluations que nous avons menées, nous avons opté pour une estimation conservatrice : $\widetilde{h_{ij}^{t,t+1}} = h_{ij}^t$.


Fig. 2: Représentation (a) et construction (b) du graphe dans l'algorithme online.

Mobilité	Strategie			
	Oracle	Frequency	Capacity	Online
Low (Abidjan)	0.77	4.79	4.4	0.77
High (Abidjan)	0.76	1.35	0.95	0.71
Low (Dakar)	0.85	3.06	2.74	0.78
High (Dakar)	0.79	1.19	0.85	0.79

Tab. 1: Gain en nombre de handovers par rapport à une approche sans CRAN.

3 Jeux de données et évaluation

Nos algorithmes sont évalués sur deux ensembles de données réelles, fournies par Orange dans le cadre du challenge "Data for Development" (D4D) [9]. Ces jeux de données ne contiennent pas toutes les informations nécessaires à la constitution du graphe. Ils sont donc complétés par des modèles stochastiques. Les jeux de données indiquent les charges horaires des BS et la durée moyenne des appels. Les appels arrivent alors selon un processus de Poisson d'intensité la charge. Leurs durées suivent une distribution Log-normale, centrée sur la moyenne, de variance arbitrairement fixée à 1s. La mobilité des utilisateurs, et donc les handovers MHO, est aussi estimée par une probabilité de se déplacer et une destination choisie uniformément dans les cellules voisines. Deux scénarios sont testés, un de faible mobilité (5% des appels subissent un MHO) et un de forte mobilité (50%). Nos algorithmes sont confrontés à deux stratégies, une dite "frequency-oriented" cherchant à optimiser la réutilisation spatiale des fréquences [7], et une autre dite "capacity-oriented" cherchant à minimiser le nombre de BBUs utilisées.

Dans un premier temps, la capacité d'une BBU est fixée à la demande maximum d'une station de base GSM haute capacité avec 18 fréquences porteuses. On évalue alors le nombre de BBUs utilisées par les différentes méthodes ainsi que le gain en handovers par rapport à une architecture sans CRAN. Les résultats sont présentés en Table 1 et Figure 3 (avec en plus les ressources GSM utilisées par le trafic).

Premier élément donné par la Table 1, nos méthodes obtiennent des gains inférieurs à 1. CRAN est donc à même de faire diminuer le nombre de handovers par rapport à une architecture sans CRAN et donc améliorer l'expérience utilisateur. Puisque nos méthodes sont optimisées pour la minimisation des handovers, il est heureux, mais normal, qu'elles donnent de meilleurs gains que les méthodes orientées capacité et fréquence. Par contre, ces méthodes génèrent un nombre de handovers parfois très supérieur à ce qu'obtient une architecture sans CRAN. Ces objectifs d'optimisation peuvent donc amener à une dégradation du service rendu aux utilisateurs, potentiellement en compromis avec les gains économiques de l'opérateur. L'autre élément est que l'Oracle (non utilisable en pratique) et la version online donnent des résultats très similaires sur notre métrique.

La Figure 3 donne une vision plus opérateur du résultat, le nombre de BBUs pouvant être relié à la charge du centre de calcul. Évidemment, la méthode orientée capacité utilise un nombre réduit de BBUs comparée à nos méthodes. L'écart reste cependant faible, le compromis n'est donc pas à l'avantage de cette méthode. La méthode orientée fréquence utilise, elle, un nombre de BBUs très supérieur. Par ailleurs, la version Oracle de l'algorithme suit mieux l'évolution de la charge du réseau que la version online qui, notamment lorsque la mobilité est faible, tarde à répercuter une baisse du trafic. Les solutions calculées par les deux méthodes sont donc assez différentes malgré leurs performances similaires en termes de handovers.

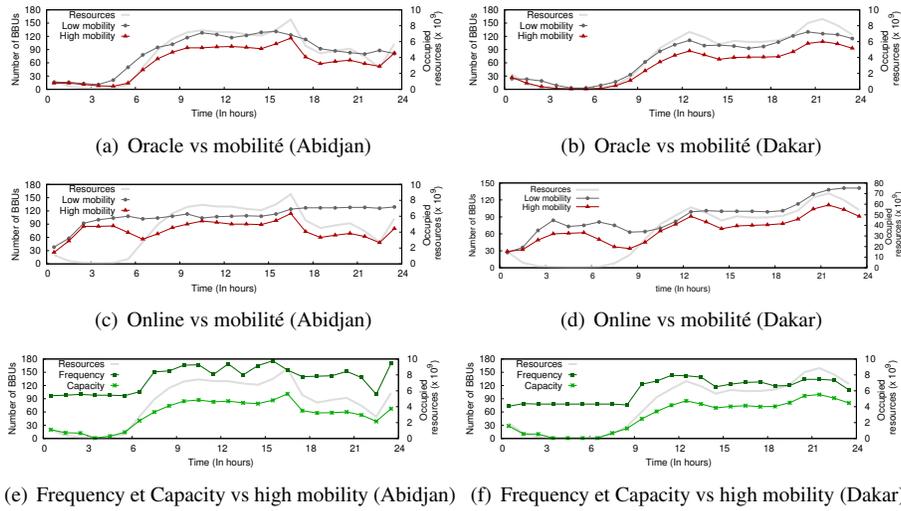


Fig. 3: Nombre de BBUs affectées par les différentes méthodes de calcul

4 Conclusion

Ce papier étudie l’affectation de BBUs à des RRHs dans une architecture Cloud-RAN. Ces architectures font émerger un nouveau type de handover. Nous proposons un modèle fondé sur des graphes dynamiques pour exprimer l’évolution du réseau et la mobilité des utilisateurs dans lequel notre problème se résout par un clustering des sommets sous contrainte de capacité. Nous proposons un algorithme de type oracle et un online dont nous montrons qu’ils ont des performances similaires avec un gain en nombre de handovers de l’ordre de 20% par rapport à une architecture sans CRAN. Dans la version complète de l’article [1], l’étude de performances est plus poussée et montre que l’on peut obtenir des gains de 30% en s’autorisant une faible probabilité de pertes d’appels. La comparaison avec les propositions de la littérature optimisant d’autres métriques montre que nos méthodes n’induisent pas une sur-charge importante des centres de calcul, la charge restant de toute façon très inférieure à celle exigée par une architecture classique.

References

- [1] D. Naboulsi *et al.*, “On User Mobility in Dynamic Cloud Radio Access Networks”, *Proc. IEEE Infocom* 2018.
- [2] Cisco, “Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2015–2020”, White Paper, 2016.
- [3] Z. Zhu *et al.*, “Virtual Base Station Pool: Towards a Wireless Network Cloud for Radio Access Networks”, *Proc. ACM CF* 2011.
- [4] C. Liu *et al.*, “The Case for Re-Configurable Backhaul in Cloud-RAN based Small Cell Networks”, *Proc. IEEE Infocom* 2013.
- [5] S. Bhaumik *et al.*, “CloudIQ: A Framework for Processing Base Stations in a Data Center”, *Proc. ACM MobiCom* 2012.
- [6] J. Tang *et al.*, “System Cost Minimization in Cloud RAN with Limited Fronthaul Capacity”, *IEEE Transactions on Wireless Communications*, 16(5):3371–3384, May 2017.
- [7] K. Wang *et al.*, “Graph-based dynamic frequency reuse in Cloud-RAN”, *Proc. IEEE WCNC* 2014.
- [8] V. D. Blondel *et al.*, “Fast Unfolding of Communities in Large Networks”, *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10): P10008, Oct. 2008.
- [9] V. Blondel *et al.*, “Data for Development: The D4D Challenge on Mobile Phone Data”, *arXiv preprint arXiv*, 1210.0137 (2012).