



HAL
open science

Le caching proactif dans les réseaux cellulaires 5G

Kenza Hamidouche, Ejder Bastug, Mehdi Bennis, Merouane Debbah

► **To cite this version:**

Kenza Hamidouche, Ejder Bastug, Mehdi Bennis, Merouane Debbah. Le caching proactif dans les réseaux cellulaires 5G. La Revue de l'électricité et de l'électronique, 2014, 3 p. hal-01098856

HAL Id: hal-01098856

<https://hal.science/hal-01098856>

Submitted on 17 Jan 2015

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.

Le caching proactif dans les réseaux cellulaires 5G

Kenza Hamidouche[◇], Ejder Baştuğ[◇], Mehdi Bennis^{*} et Mérouane Debbah[◇],

[◇]Chaire Alcatel-Lucent - SUPÉLEC, Gif-sur-Yvette, France

^{*}Centre de Communications sans fils, Université de Oulu, Finlande

{kenza.hamidouche, ejder.bastug, merouane.debbah}@supelec.fr, bennis@ee.oulu.fi

Abstract—Cet article étudie l'un des outils clés pour aller au delà des réseaux sans fils 4G, en exploitant le déploiement des réseaux à petites cellules (RPC), et plus précisément à travers le *caching proactif*. Ce nouveau paradigme permet d'exploiter les données du big data dans le domaines des réseaux sans fils. En munissant les stations de bases de capacités de prédiction et en exploitant les développements récents dans le stockage, la sensibilité au contexte et les réseaux sociaux, la gestion du trafic peut être sensiblement améliorée. Ceci en servant les demandes prévisibles des utilisateurs de manière proactive, par la mise en cache des données dans les stations de base et les appareils des utilisateurs. Afin de montrer l'importance du caching proactif, nous étudions deux cas qui considèrent la structure spatiale et sociale du réseau, où le caching proactif joue un rôle crucial. Tout d'abord, nous proposons un mécanisme par lequel les fichiers sont mis dans le cache proactivement lorsque le trafic est faible en se basant sur la popularité des fichiers et les corrélations entre les utilisateurs et les fichiers. Ensuite, en s'appuyant sur les réseaux sociaux et les communications appareil-à-appareil (AàA), nous proposons une procédure qui exploite la structure sociale du réseau en prédisant l'ensemble des utilisateurs influents afin de mettre en cache (de manière proactive) les contenus stratégiques et de les diffuser sur les connexions sociales via les communications AàA. L'exploitation du paradigme de mise en cache proactif peut apporter des gains importants en termes de réduction de trafic sur les liens backhaul et de satisfaction des utilisateurs.

I. INTRODUCTION

La récente prolifération des smartphones a considérablement enrichi l'expérience des utilisateurs mobiles, conduisant à une vaste gamme de nouveaux services sans fil, tels que le streaming multimédia, les applications de navigation sur le Web et les réseaux sociaux. Ce phénomène a encore été alimenté par la diffusion de la vidéo mobile, qui représente actuellement près de 50 % du trafic de données mobiles, avec une augmentation qui sera 500 fois plus élevée au cours des 10 prochaines années [1]. Au même temps, les réseaux sociaux sont déjà le deuxième plus grand contributeur en volume de trafic avec une part de 15 % en moyenne [2]. Ce nouveau phénomène a poussé les opérateurs mobiles à revoir leurs réseaux actuels et à chercher des techniques plus avancées et sophistiquées pour augmenter la couverture, booster la capacité du réseau, et de manière rentable rapprocher les données des utilisateurs de ces derniers.

Une approche prometteuse pour répondre à ces demandes de trafic sans précédent se fait par le déploiement de réseaux à petites cellules [3]. Ces réseaux représentent un nouveau paradigme réseau basé sur l'idée de déploiement de petites stations de base à courte portée, de faible puissance, et à faible coût, superposées sur le réseau macro-cellulaire. Jusqu'à

présent, la grande majorité des travaux de recherche se focalisent sur les challenges liés à l'auto-organisation, la coordination des interférences inter-cellulaires (ICIC), la réduction de la charge du trafic, l'efficacité énergétique, etc (voir [3] et les références qui y sont citées). Ces études ont été réalisées dans le cadre du paradigme actuel des réseaux *réactifs*, dans lequel les demandes et les flux de trafic des utilisateurs doivent être servis en urgence à leur arrivée ou abondonnés provoquant des interruptions de service. Pour cela, le modèle de RPC actuel est loin de servir la demande de trafic élevée dont le déploiement à large échelle repose sur l'acquisition de sites coûteux, en plus du coût d'installation et de backhaul. Ces inconvénients sont en passe de devenir de plus en plus aigus, en raison de la flambée du nombre de périphériques connectés et l'avènement des réseaux ultra-denses, qui continueront à pousser les infrastructures des réseaux cellulaires actuels. Ces observations clés prescrivent un *nouveau* paradigme réseau qui va au-delà des déploiements hétérogènes actuels de petites cellules, mettant à profit les derniers développements dans le stockage, la sensibilité au contexte, et les réseaux sociaux.

Le paradigme réseau proposé est essentiellement proactif dans le sens où les nœuds du réseau (c-à-d, les stations de base, ordinateurs et smartphones) exploitent les informations sur les utilisateurs, anticipent les requêtes des utilisateurs et tirent profit de leurs capacités prédictives afin de réaliser d'importantes économies en ressources tout en garantissant les exigences en qualité de service (QoS) et réduisant les dépenses en coût/énergie [4]. Dotés de ces capacités prédictives, les utilisateurs sont ordonnancés de manière plus efficace et les ressources sont pré-alloués de façon plus intelligente. Ceci en servant de manière proactive, lorsque le trafic est faible dans le réseau (par exemple, la nuit), les demandes prédites pour les heures de pointe. En exploitant intelligemment les modèles statistiques du trafic et les informations sur les utilisateurs (c-à-d, les distributions de la popularité des fichiers, l'emplacement, et la vitesse de mobilité), le paradigme proposé permet de mieux prédire quand est ce que les données des utilisateurs seront demandées avec les ressources nécessaires pour les servir, et à quels emplacements dans le réseau les données doivent être mises en cache.

Une autre tendance d'actualité est les *réseaux sociaux* en ligne (e.g., Facebook) qui ont joué récemment un rôle déterminant dans la diffusion des données des utilisateurs [2]. En exploitant l'interdépendance entre les utilisateurs à travers les relations et les liens sociaux, les réseaux futurs peuvent tirer profit des modèles de corrélation dans les réseaux de liens sociaux et des données géographiques pour une meilleure prédiction et déduction du comportement des utilisateurs. La Figure. 1 montre une abstraction de la couche réseau technologique/spatiale recouverte d'une couche représentant le réseau social. Comme la diffusion du contenu

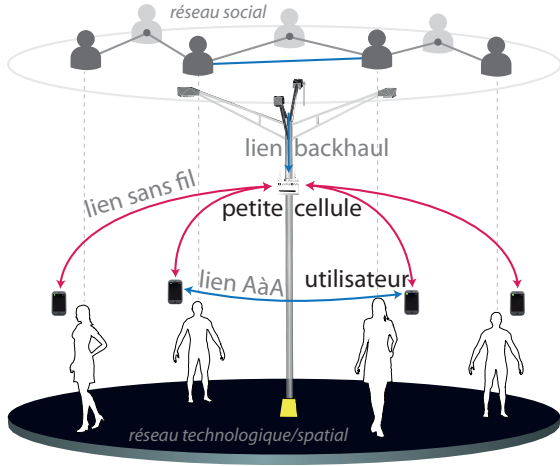


Figure 1: Une illustration d'une superposition d'un réseau social interconnecté et d'un réseau technologique/spatial.

des nœuds dans la couche réseau social est gérée en temps réel par l'intermédiaire des nœuds de la couche réseau technologique/spatiale, l'analyse des interactions entre ces deux couches pourrait engendrer des gains supplémentaires dans les réseaux du futur.

L'objectif principal de cet article est d'explorer les fondements des petites cellules activées en mode d'accès radio prédictif/proactif, et de faire de grandes avancées sur ce nouveau paradigme réseau. Dans la section suivante, nous détaillons les avantages et les perspectives de la mise en place des réseaux proactifs via deux études de cas, en s'appuyant sur le déploiements des réseaux à petites cellules, la notion d'apprentissage automatique et les réseaux sociaux.

II. CAS D'ÉTUDE I : LES RÉSEAUX À PETITES CELLULES PROACTIFS

Dans cette section, nous étudions le problème de réduction de la charge sur les liens backhaul dans les RPC, où le caching proactif joue un rôle crucial. En effet, le backhauling a une extrême importance pour le déploiement du RPC. Dans le modèle réseau considéré, les petites stations de base (PSB) sont déployées munis d'unités de stockage de grande capacité, mais les liens backhaul sont de capacité limitée. Nous nous basons sur [4], dans lequel une procédure de caching proactif est proposée pour stocker des fichiers en fonction de leur popularité, jusqu'à ce que la capacité de stockage soit atteinte. Dans ce modèle, les PSB connaissent la matrice de popularité $\mathbf{P}_{N \times F}$ où chaque ligne représente les utilisateurs et les colonnes, les fichiers préférés ou les notes données à ces fichiers. Néanmoins, dans la pratique, la matrice de popularité est grande, creuse et partiellement inconnue. Par conséquent, inspiré du *paradigme Netflix* et en utilisant des outils d'apprentissage automatique supervisé et spécifiquement le filtrage collaboratif (FC), nous proposons une procédure de caching proactif distribué qui exploite les corrélations utilisateurs-fichiers pour déduire la probabilité que le $u^{\text{ème}}$ utilisateur demande le $i^{\text{ème}}$ fichier.

La procédure de caching proposée est composée d'une partie formation et d'une partie placement. Dans la partie formation, l'objectif est d'estimer la matrice de popularité \mathbf{P}

(soit $\hat{\mathbf{P}}_{N \times F}$), où chaque PSB construit un modèle basé sur les informations disponibles sur les préférences/notes des utilisateurs. Cela se fait en résolvant le problème de minimisation des moindres carrés suivant :

$$\min_{\{b_u, b_i\}} \sum_{u,i} (r_{ui} - \hat{r}_{ui})^2 + \lambda \left(\sum_u b_u^2 + \sum_i b_i^2 \right) \quad (1)$$

où la somme est sur les paires utilisateur/fichier (u, i) dans l'ensemble de formation, où l'utilisateur u a noté le fichier i (r_{ui}), et la minimisation est sur les $N + F$ paramètres, où N est le nombre d'utilisateurs et F le nombre de fichiers dans l'ensemble de formation. En outre, $\hat{r}_{ui} = \bar{r} + b_u + b_i$ est le prédicteur de référence dans lequel b_i représente la qualité de chaque fichier i par rapport à la moyenne \bar{r} , et b_u représente la qualité de chaque utilisateur u par rapport à \bar{r} . Enfin, le poids λ est choisi pour équilibrer entre la régularisation et les données de formation appropriées. En particulier, la décomposition en valeurs singulières (SVD) régularisée peut être utilisée pour sa précision numérique (voir [5] pour d'autres méthodes de FC et leur comparaison). La SVD régularisée basée sur le FC construit $\hat{\mathbf{P}}$, comme une version de faible rang de \mathbf{P} . Dû à la faible densité de l'ensemble de formation, la décomposition se fait par une descente de gradient en exploitant la propriété des moindres carrés de la SVD. Après l'estimation de la matrice de popularité des fichiers $\hat{\mathbf{P}}$, la décision de mise en cache proactive peut être faite dans la phase placement des fichiers par le stockage des fichiers les plus populaires en utilisant un algorithme glouton (comme dans [4]) jusqu'à ce qu'il ne reste plus d'espace de stockage libre.

III. CAS D'ÉTUDE II : LE CACHING SOCIAL VIA LES COMMUNICATIONS AÀA

Dans cette section, nous montrons l'efficacité du caching proactif en s'appuyant sur les réseaux sociaux et les communications AàA. Plus précisément, nous considérons le déploiement d'un réseau où les utilisateurs accèdent à certains fichiers à partir d'une bibliothèque de fichiers F . Chaque utilisateur peut stocker des fichiers sur son appareil dans la limite de sa capacité de stockage. Comme le montre la Figure. 1, le réseau considéré peut être vu comme une superposition d'un réseau à petites cellules et d'un réseau social.

En exploitant l'interaction entre le réseau social et le réseau technologique, chaque PSB traque et apprend sur l'ensemble des utilisateurs les plus *influents* en utilisant le graphe social, et détermine les probabilités d'influence en se basant sur l'historique des actions des utilisateurs rencontrés ainsi que leurs demandes de fichiers. Notamment, quand un utilisateur donné demande un fichier particulier, la PSB détermine si l'un des utilisateurs influents a le fichier demandé. Si oui, il redirige la demande vers l'utilisateur influent pour le servir à travers une communication AàA. Sinon, si le fichier n'est pas stocké par l'utilisateur influent, la PSB transmet le fichier directement à l'utilisateur à partir du cœur du réseau.

A. Les utilisateurs influents et les communautés

Afin de déterminer l'ensemble des utilisateurs influents, nous exploitons les relations et les liens sociaux entre les utilisateurs en se basant sur la métrique de *centralité* [6].

La métrique de centralité mesure l'influence sociale d'un nœud selon la manière dont il connecte le réseau, où le nœud ayant la centralité la plus élevée est considéré comme étant le plus important (i.e., influent) pour sa communauté. Typiquement, quatre mesures de centralité peuvent être considérées : (1) *le degré*, pour représenter le nombre de liens qui connectent un nœud avec d'autres nœuds voisins; (2) *la proximité*, pour représenter la distance entre un nœud et les autres nœuds à proximité. Par ailleurs, la mesure de proximité est la mesure clé pour déterminer les utilisateurs les plus influents; (3) *la relation d'intermédiaire*, qui représente la mesure dans laquelle un nœud se trouve sur les chemins les plus courts reliant à d'autres nœuds; (4) *le vecteur propre*, estime l'influence des nœuds dans le réseau en utilisant le vecteur propre correspondant à la plus grande valeur propre de la matrice d'adjacence du réseau. Ainsi, après avoir obtenu les K -utilisateurs les plus influents du réseau social, une méthode de classification (e.g., K -moyennes [7]) peut être appliquée pour former la communauté.

B. Diffusion de contenu

Après avoir défini les utilisateurs influents et leurs communautés, l'étape suivante consiste à déterminer le processus de diffusion de contenu à l'intérieur de chaque communauté. Pour ce faire, nous modélisons la diffusion de contenu par le processus de restaurant chinois (PRC) [8]. La motivation principale derrière l'utilisation de ce processus est de modéliser la procédure de partitionnement fichier-utilisateur qui constitue essentiellement une information préalable pour affecter les fichiers aux utilisateurs. Le PRC est basé sur une métaphore dans laquelle les objets sont des clients dans un restaurant, et les classes sont les tables auxquelles ils s'assoient. En particulier, dans un restaurant avec un grand nombre de tables, chacune avec un nombre infini de places, les clients entrent dans le restaurant les uns après l'autre, et chacun choisit une table au hasard. Dans le PRC avec le paramètre β , chaque client choisit une table occupée avec une probabilité proportionnelle au nombre d'occupants, et choisit la prochaine table vacante avec une probabilité proportionnelle à β . Plus précisément, le premier client choisit la première table avec une probabilité $\frac{\beta}{\beta} = 1$. Le deuxième client choisit la première table avec une probabilité $\frac{1}{1+\beta}$, et la seconde table avec une probabilité $\frac{\beta}{1+\beta}$. Après que le deuxième client ait choisi la seconde table, le troisième client choisit la première table avec une probabilité $\frac{1}{2+\beta}$, la deuxième table avec une même probabilité $\frac{1}{2+\beta}$ et la troisième table avec une probabilité $\frac{\beta}{2+\beta}$. Ce processus continue jusqu'à ce que tous les clients aient des places, en définissant la distribution sur l'affectation des personnes aux tables. Par conséquent, les décisions des clients suivants sont influencées par le feedback des clients précédents, selon lequel les clients changent leurs avis sur les fichiers et les probabilités avec lesquelles ils décident de consulter ces fichiers.

Compte tenu de cela, la diffusion de contenu dans le réseau social est analogue à la sélection de la table dans un PRC. En effet, si on considère le réseau comme étant un restaurant chinois, où le contenu correspond au très grand nombre de fichiers, et les utilisateurs aux clients, nous pouvons modéliser

le processus de diffusion de contenu en ligne par un PRC. C'est au sein de chaque communauté que les utilisateurs demandent séquentiellement de télécharger les données qu'ils recherchent, et quand un utilisateur télécharge le contenu, cet accès est enregistré dans l'historique de l'utilisateur. À son tour, cette action influence la probabilité que ce contenu soit demandé par d'autres utilisateurs au sein de la même communauté sociale.

On définit le nombre d'utilisateurs par N et le nombre total de données par F . Compte tenu de l'important volume des données disponibles, nous supposons que $F = F_0 + F_h$, où F_h représente l'ensemble des données ayant déjà été consultées et F_0 l'ensemble des données sans historique. On considère également $\mathbf{Z}_{N \times F}$ comme étant une matrice binaire aléatoire indiquant quelle donnée a été choisie par chaque utilisateur, où $z_{n,f} = 1$ si l'utilisateur n sélectionne la donnée f et 0 sinon. On peut montrer que [8] :

$$P(\mathbf{Z}) = \frac{\beta^{F_h} \Gamma(\beta)}{\Gamma(\beta + N)} \prod_{f=1}^{F_h} (m_f - 1)! \quad (2)$$

où $\Gamma(\cdot)$ représente la fonction Gamma, m_f est le nombre d'utilisateurs actuellement affectés à la donnée f (ou l'historique de nombre de consultations) et F_h est l'ensemble des données ayant déjà été consultées avec $m_f > 0$. Par conséquent, pour une $P(\mathbf{Z})$ donnée, les fichiers les plus populaires dans chaque communauté peuvent être stockés en se basant sur des algorithmes glouton dans le cache des K -utilisateurs les plus influents.

IV. CONCLUSION

Nous avons présenté les réseaux réactifs actuels et proposé un nouveau paradigme de réseau proactif où le caching joue un rôle crucial. En exploitant les capacités prédictives des réseaux 5G, ainsi que les notions de sensibilité au contexte et les réseaux sociaux, la surcharge du réseau en demandes de données peut être sensiblement réduite en servant de manière proactive les demandes prévisibles des utilisateurs, grâce au caching stratégique des données dans les stations de base et les appareils des utilisateurs. Ce réseau prédictif, avec des capacités de stockage suffisantes sur les bordures du réseau, permet d'aider les opérateurs mobiles à apprivoiser le tsunami de données, qui continuera à mener à rude épreuve les réseaux actuels.

REFERENCES

- [1] Cisco, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2013-2018," *Livre blanc*, [En ligne] <http://goo.gl/l77HAJ>, 2014.
- [2] Ericsson, "5G radio access - research and vision," *Livre blanc*, [En ligne] <http://goo.gl/Huf0b6>, 2012.
- [3] J. G. Andrews, "Seven ways that HetNets are a cellular paradigm shift," *IEEE Communications Magazine*, vol. 51, no. 3, pp. 136-144, 2013.
- [4] E. Baştuğ, J.-L. Guénégo, and M. Debbah, "Proactive small cell networks," in *20th International Conference on Telecommunications (ICT)*, Casablanca, Maroc, 2013.
- [5] J. Lee, M. Sun, and G. Lebanon, "A Comparative Study of Collaborative Filtering Algorithms," [En ligne] *arXiv: 1205.3193*, 2012.
- [6] M. Newman, *Networks: an introduction*. Oxford University Press, 2009.
- [7] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651 - 666, 2010.
- [8] T. L. Griffiths and Z. Ghahramani, "The Indian Buffet Process: An Introduction and Review," *J. Mach. Learn. Res.*, vol. 12, pp. 1185-1224, Jul. 2011.