



HAL
open science

Stratégie d'apprentissage pour le multicast LTE avec communications D2D

Filippo Rebecchi, Lorenzo Valerio, Raffaele Bruno, Vania Conan, Marcelo Dias de Amorim, Andrea Passarella

► **To cite this version:**

Filippo Rebecchi, Lorenzo Valerio, Raffaele Bruno, Vania Conan, Marcelo Dias de Amorim, et al.. Stratégie d'apprentissage pour le multicast LTE avec communications D2D. Rencontres francophones sur la conception de protocoles, l'évaluation de performance et l'expérimentation des réseaux de communication (CoRes 2016), May 2016, Bayonne, France. hal-01305986

HAL Id: hal-01305986

<https://hal.science/hal-01305986v1>

Submitted on 22 Apr 2016

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.

Stratégie d'apprentissage pour le multicast LTE avec communications D2D

Filippo Rebecchi^{1 †}, Lorenzo Valerio², Raffaele Bruno², Vania Conan¹, Marcelo Dias de Amorim³ et Andrea Passarella²

¹Thales Communications & Security, France ²IIT-CNR, Italie ³CNRS/LIP6, France

Le multicast LTE représente, en principe, un moyen très efficace de distribuer du contenu à un groupe d'utilisateurs. Sur une cellule, son efficacité est dictée par les utilisateurs dont le canal est le moins bon. Afin de limiter leur impact, nous proposons d'exploiter des communications opportunistes entre les terminaux (D2D) en complément du multicast. Cependant, un tel mécanisme doit s'adapter à l'aléa introduit par les communications opportunistes pour sélectionner le meilleur compromis entre transmissions multicast et D2D. Dans ce but, nous évaluons l'apport d'une stratégie d'apprentissage automatique, dite du "bandit manchot multi-bras". Nos évaluations montrent que cette stratégie permet de configurer la transmission multicast d'une manière plus efficace, contribuant ainsi à l'économie des ressources radio.

Mots-clés : D2D, multicast, apprentissage.

1 Introduction

En exploitant la nature *diffusante* du canal sans fil, le multicast LTE utilise une seule liaison radio partagée entre plusieurs équipements utilisateurs (UE) au sein d'une même cellule. Les UE d'un même groupe reçoivent la même transmission permettant, en principe, une utilisation très efficace des ressources radio (RR) [LG12]. Cependant, en dépit de ses caractéristiques attrayantes, le multicast LTE présente des problèmes intrinsèques et irrésolus qui limitent son exploitation, notamment l'adaptation à l'utilisateur avec le pire canal et l'absence de fiabilité [BLHR06]. En misant sur des communications dispositif-à-dispositif (D2D), il est possible d'obtenir des gains de performance en matière de RR consommées à la station de base cellulaire (eNB) [RDdAC⁺15]. Au lieu d'atteindre tous les UE intéressés avec une seule transmission multicast, nous proposons de couvrir seulement un sous-ensemble d'entre eux (ceux qui présentent la meilleure qualité de canal), tout en exploitant des communications D2D pour atteindre les autres. Ceci permet une transmission multicast avec une modulation plus efficace. La conception d'une telle stratégie de distribution fait face à plusieurs défis spécifiques aux domaines des réseaux sans fil :

- Le bénéfice des transmissions D2D dépend de la mobilité des UE. Les communications D2D se produisent à des instants aléatoires, et ne peuvent donner qu'une assurance partielle de réception.
- Pour garantir la qualité de service requise, tout en offrant des économies de RR à l'eNB, il est essentiel de répartir, de façon efficace, les UE entre réception en multicast et en D2D.

L'incertitude qui régit les contacts entre les UE nous conduit à utiliser une approche d'apprentissage appelée *bandit manchot multi-bras* pour décider, à chaque paquet, la fraction d'UE à atteindre en multicast [SB98]. Chaque décision se traduit par une certaine utilisation de RR, ce qui génère une récompense. Cette récompense est ensuite utilisée pour guider (de façon probabiliste) les choix futurs. Le détail des résultats peut être trouvé dans la référence suivante [RVB⁺15].

2 Stratégie de distribution

Nous voulons transmettre des données multicast avec un délai maximal de réception D , et ceci au moindre coût pour l'infrastructure cellulaire (en nombre de RR utilisées). À cet égard, la Fig. 1 montre les trois étapes

[†]email: filippo.rebecchi@thalesgroup.com

de la stratégie de diffusion proposée. Dans un premier temps, l'infrastructure effectue une transmission en multicast qui, en général, ne couvre pas tous les UE. La modulation de transmission, décidée par le mécanisme d'apprentissage, a tendance à être un bon compromis en matière de couverture et consommation de RR. La diffusion du contenu se poursuit ensuite avec des transmissions D2D en direction de ces UE qui n'ont pas reçu le contenu à l'étape précédente. Lorsque le délai atteint sa valeur maximale, l'eNB envoie en unicast les données manquantes. Dans ce schéma, le coût total de diffusion, en matière de RR, provient à la fois de la transmission initiale (multicast) et des retransmissions finales (unicast).

Réduire l'ensemble des UE atteints par le multicast réduit le coût d'envoi, car seuls les UE avec un bon canal sont pris en compte (donc moins de RR nécessaires pour la transmission). Toutefois, cela peut engendrer des coûts supplémentaires à cause des transmissions finales en unicast. Nous abordons donc le problème suivant : *comment sélectionner la configuration initiale des UE destinés à être atteints par le multicast tout en minimisant les ressources nécessaires pour la diffusion du contenu.*

Notre système utilise un seul paramètre I_0 qui représente la fraction d'UE qui reçoivent la transmission multicast.

Cela signifie que la transmission multicast, dans un premier temps, atteint seulement les meilleurs I_0 UE en matière de qualité de canal. Configurer de façon optimale I_0 n'est pas trivial. Alors que le coût de la transmission multicast est déterministe, le coût des transmissions finales en unicast est une variable stochastique.

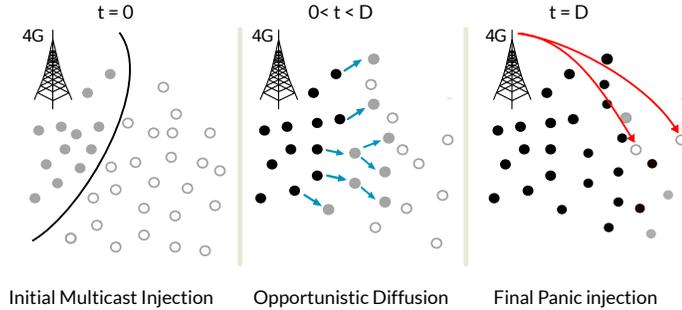


FIGURE 1: Stratégie de distribution : (1) Transmission en multicast à un sous-ensemble des UE; (2) diffusion opportuniste en D2D; (3) transmissions finales en unicast.

3 Algorithme d'apprentissage

Nous modélisons le problème décrit dans la section 2 à l'aide d'une approche d'apprentissage que nous résolvons avec la stratégie du *bandit manchot multi-bras*. Le système est capable d'apprendre la distribution des valeurs de I_0 de manière autonome et sans connaître au préalable la trame des contacts entre UE.

La formulation de l'algorithme peut être spécialisée comme suit. Tout d'abord, dans notre problème chaque bras du bandit correspond à une valeur de I_0 . Il en résulte que la distribution F_{d_i} représente la quantité de RR utilisées pendant le processus de diffusion lorsque I_0 est choisi. Plus précisément, $d_i = m_i + x_i$, où m_i est le nombre connu de RR utilisées pour une transmission multicast à la modulation nécessaire pour atteindre les I_0 meilleurs UE en matière de qualité de canal, et x_i est la variable aléatoire qui modélise le nombre total de RR utilisées pour les transmissions unicast à la fin de la période de validité du contenu (afin que tous les UE soient couverts). Chaque round de l'algorithme correspond à la diffusion d'un contenu qui est composé d'une multitude de paquets qui contribuent de façon indépendante à la mise à jour des distributions F_{d_i} . En supposant que $I_0 = i$ ait été utilisé pour la transmission du n -ième paquet, la récompense obtenue est donnée par $\mu_i(n) = \frac{1}{m_i + x_i(n)}$. Après dépassement du temps limite de chaque contenu, la récompense moyenne $\bar{\mu}_i(n)$ pour chaque seuil est mise à jour avec une moyenne mobile exponentielle de paramètre α . Si l'on définit $i^*(n) = \operatorname{argmax}_i \bar{\mu}_i(n-1)$, la politique la plus simple pour choisir la valeur du seuil de la (n) -ième transmission est de sélectionner $i^*(n)$ avec une probabilité $\pi_{i^*(n)} = (1 - \epsilon)$ (algorithme dit *ϵ -greedy*). Autrement, les récompenses peuvent être mises à jour avec la *méthode de la poursuite*, où les probabilités d'envoi sont renforcées comme suit :

$$\pi_{i^*(n)}(n) = \pi_{i^*(n)}(n-1) + \beta[\pi_{MAX} - \pi_{i^*(n)}(n-1)], \quad (1)$$

$$\pi_{i(n)}(n) = \pi_{i(n)}(n-1) + \beta[\pi_{MIN} - \pi_{i(n)}(n-1)], \forall i \neq i^*, \quad (2)$$

où π_{MAX} , π_{MIN} sont respectivement la limite supérieure et inférieure que la probabilité $\pi_{i(n)}(n)$ peut prendre. Cela garantit que la méthode de poursuite est en mesure de faire face à la possibilité de non-stationnarité

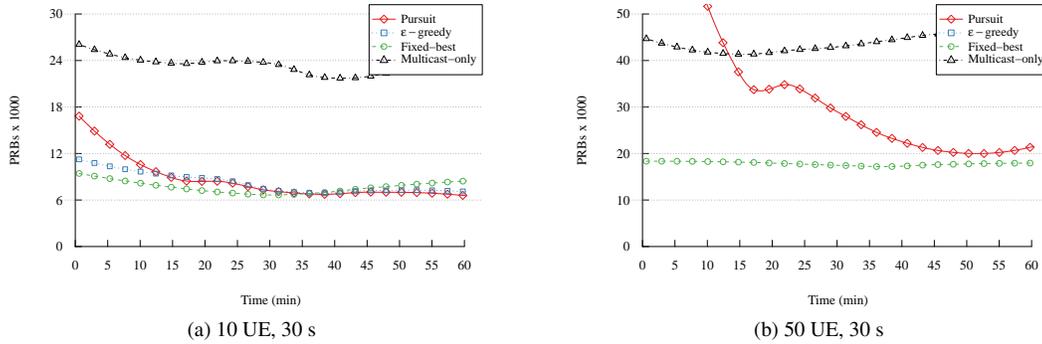


FIGURE 2: Utilisation de RR pour les stratégies *Multicast-only* (noir), ϵ -greedy (bleu), *Fixed-best* (vert), et *Poursuite* (rouge). En Fig. 2 (b), la stratégie ϵ -greedy n'est pas représentée car elle ne converge pas.

du problème que nous étudions, à savoir le fait que la distribution des récompenses puisse changer au fil du temps en raison de la mobilité sous-jacente.

4 Évaluation

Nous considérons la distribution de contenus dans un scénario piéton comme dans un centre commercial ou une zone d'intérêt touristique. Nous simulons des flux UDP, avec des paquets de 2048 octets et une taille du contenu total de 8 MO. La mobilité synthétique des UE suit le modèle Random-Waypoint sur une zone de $200 \times 200 \text{ m}^2$. Les UE se déplacent dans cet espace avec une vitesse se situant entre 1 et 2,5 m/s. Le réseau est composé d'un eNB placé au centre de la zone d'intérêt, un serveur distant qui fournit les contenus, et un nombre variable d'UE. Les principaux indices de performance que nous considérons dans l'évaluation sont (i) le nombre de RR utilisées par l'eNB pour transmettre le contenu et (ii) le pourcentage de RR économisées par rapport au cas avec uniquement le multicast. Dans nos évaluations, nous comparons les quatre stratégies suivantes :

- **Multicast-only** : les UE n'ont que le multicast cellulaire pour recevoir les données. Il s'agit de la stratégie de base.
- **Fixed-best** : l'allocation de la taille de multicast est statique (I_0 reste fixe pendant toute la durée de la simulation). Nous avons effectué des simulations pour trouver expérimentalement la valeur de I_0 qui minimise l'utilisation globale de RR. Cette stratégie *a posteriori* est irréalisable dans la pratique. Toutefois, elle est utilisée comme référence pour les stratégies d'apprentissage.
- **ϵ -greedy** : ce algorithme simple sélectionne avec une probabilité de $1 - \epsilon$ la valeur de I_0 avec la meilleure récompense. Dans notre réalisation, nous utilisons $\epsilon = 0.05$ et $\alpha = 0.5$.
- **Poursuite** : les probabilités sont mises à jour de manière à renforcer au maximum le meilleur I_0 estimé, tout en discriminant les autres valeurs. La probabilité de sélection minimale π_{MIN} permet de conserver une exploration constante. Dans la simulation, nous avons fixé $\beta = 0.3$, $\pi_{MIN} = 0.01$ et $\pi_{MAX} = 0.95$.

Les simulations effectuées montrent que les techniques d'apprentissage permettent d'économiser jusqu'à 88% de RR pour un scénario avec délai de livraison fixé à 90 s par rapport à la stratégie *Multicast-only*. Ces résultats confirment qu'une synergie dans l'utilisation des transmissions multicast et D2D permet de gérer plus efficacement les ressources à l'eNB. De plus, ces stratégies peuvent trouver de manière autonome le meilleur compromis entre multicast et D2D en un délai raisonnable – toujours inférieur à une heure dans nos simulations – sans explorer exhaustivement tout l'espace des paramètres. La Fig. 2 illustre le processus d'apprentissage pour le délai le plus serré que l'on a considéré (30 s). Les stratégies *Poursuite* et ϵ -greedy ont besoin de temps pour apprendre la distribution la plus appropriée pour I_0 . Après stabilisation, leur performance est souvent à la hauteur ou même meilleure que la stratégie représentée par *Fixed-best*, où la valeur de I_0 est pré-calculée. Dans cette dernière stratégie, la performance reste stable sur toute la période de diffusion, mais ce chiffre est le résultat d'un processus étendu d'essai-erreur. Un avantage des

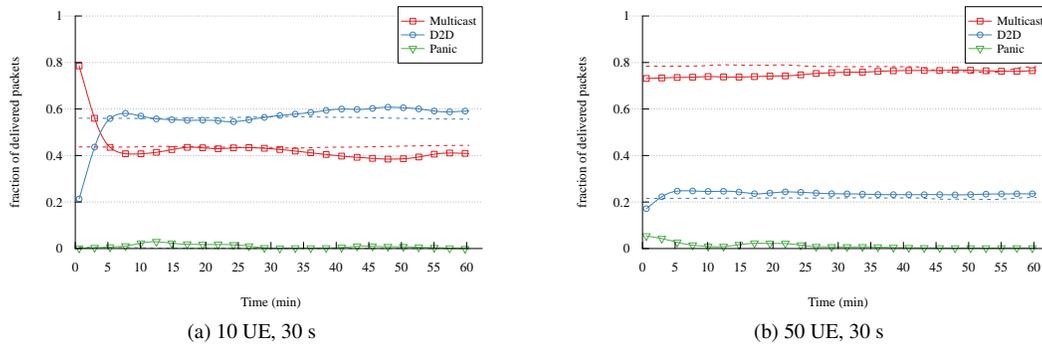


FIGURE 3: Méthode de réception dans *Poursuite*. Les lignes pointillées se réfèrent à la stratégie *Fixed-best*.

techniques d'apprentissage est que, contrairement à l'allocation statique de *Fixed-best*, elles continuent à explorer l'espace des solutions, étant capables de faire face à la non-stationnarité du processus de contact qui régit la diffusion D2D. Ceci explique pourquoi, à certains moments, ces stratégies peuvent avoir des meilleures performances que *Fixed-best*. En revanche, nous nous rendons compte que la méthode ϵ -greedy, en raison de sa simplicité, ne correspond pas bien aux scénarios qui présentent une variabilité importante de la diffusion D2D (par exemple, en Fig. 2(b)). Dans ces cas, la méthode de *poursuite* est mieux adaptée. Néanmoins, dans les scénarios où la variabilité du processus est faible – comme dans le cas où le délai de livraison est long, ou le nombre d'UE faible – l'approche ϵ -greedy permet des temps de convergence plus rapides (Fig. 2(a)). Enfin, la Fig. 3 montre l'apport des différentes méthodes de réception comme pourcentage des paquets distribués dans l'une des trois phases.

5 Conclusion

Nous avons proposé une stratégie de dissémination de contenus qui mise sur l'utilisation conjointe du multicast cellulaire et du D2D. Même si le multicast est une option avantageuse pour distribuer des données dans un réseau LTE, la performance globale est déterminée par l'utilisateur avec la qualité de canal la moins bonne. Nous avons donc proposé une solution basée sur des techniques d'apprentissage pour lutter contre les inefficacités du multicast et distribuer une partie du trafic en utilisant des communications D2D. Les résultats de simulation montrent que la stratégie proposée permet de configurer la transmission multicast d'une manière plus efficace, contribuant ainsi à l'économie des ressources et à l'augmentation du débit global de la cellule.

Références

- [BLHR06] R. Bhatia, L. Li, L. Haiyun, and R. Ramjee. ICAM : integrated cellular and ad hoc multicast. *IEEE Trans.Mobile Comput.*, 5(8) :1004–1015, Aug 2006.
- [LG12] D. Lecompte and F. Gabin. Evolved multimedia broadcast/multicast service (eMBMS) in LTE-advanced : overview and rel-11 enhancements. *IEEE Commun. Mag.*, 50(11) :68–74, November 2012.
- [RDdAC⁺15] F. Rebecchi, M. Dias de Amorim, V. Conan, A. Passarella, R. Bruno, and M. Conti. Data offloading techniques in cellular networks : A survey. *IEEE Commun. Surveys Tuts.*, 17(2) :580–603, Secondquarter 2015.
- [RVB⁺15] F. Rebecchi, L. Valerio, R. Bruno, V. Conan, M. Dias de Amorim, and A. Passarella. A joint multicast/{D2D} learning-based approach to {LTE} traffic offloading. *Computer Communications*, 72 :26–37, December 2015.
- [SB98] Richard S Sutton and Andrew G Barto. *Introduction to reinforcement learning*. MIT Press, 1998.