

Gestion de la Mobilité des Utilisateurs dans des Réseaux sans-fil de type "Cell-Free" avec Vieillessement des Canaux

Ala Eddine NOUALI¹ Mattia MERLUZZI¹ Jean-Baptiste DORE¹ Jean-Paul JAMONT²

¹CEA-Leti, 17 Avenue des Martyrs, 38054 Grenoble, , France

²LCIS, 50 Rue Barthélémy de Laffemas, 26000 Valence , France

Résumé – Cet article étudie l’association dynamique d’équipement utilisateur (UE) dans des réseaux scalables "cell-free", en considérant le vieillissement du canal et le délai du handover. Pour résoudre ce problème, nous proposons un algorithme basé sur l’apprentissage par renforcement profond (DRL). L’agent DRL ajuste dynamiquement le groupe de points d’accès (APs) pour chaque UE, en tenant compte des ressources radios limitées et les exigences de qualité de service (QoS) des UEs. Il est montré que l’agent DRL fournit une QoS favorable avec des connexions stables nécessitant un nombre limité de handovers au cours du temps.

Abstract – This paper studies the dynamic user equipment (UE) association in scalable cell-free networks, considering channel aging and handover delay. To solve this problem, we propose an algorithm based on deep reinforcement learning (DRL). The DRL agent adjusts dynamically the cluster of access points (APs) for each UE, taking into account the limited radio resources and quality of service (QoS) requirements. It is shown that the DRL agent provides a favorable QoS through stable connections with a limited number of handovers over time.

1 Introduction

Pour garantir la scalabilité des réseaux "cell-free", l’approche du regroupement centré sur l’équipement utilisateur (User-Centric Clustering, UCC) a été proposée, où chaque équipement utilisateur (UE) est servi par un groupe de points d’accès (APs) préférés [1]. Cependant, dans un réseau dynamique, la mobilité de l’UE cause des fluctuations rapides des canaux qui vieillissent au cours du temps [2]. Ainsi, les variations rapides des canaux nécessitent des handovers plus fréquents [3] pour ajuster les groupes d’APs servant chaque UE afin de maintenir une qualité de service (QoS) satisfaisante.

Dans un réseau cell-free mobile, plusieurs approches de handovers ont été proposées. Les auteurs de [4] proposent d’ajuster le cluster d’APs d’un UE uniquement lorsque la différence de performance entre l’état précédent et l’état actuel dépasse un seuil prédéfini. L’approche de [5] intègre en outre une marge d’hystérésis, garantissant que les performances de l’UE après les handovers dépassent celles actuelles avec au minimum cette marge. [6] donne la priorité aux ajustements de connexions pour les UEs subissant la dégradation des performances la plus critique, réduisant ainsi considérablement les handovers inutiles dans le réseau.

Néanmoins, l’approche UCC fait face plusieurs défis, principalement en termes de capacités de calcul des APs et des UEs [7]. De plus, chaque UE a des exigences spécifiques en fonction de ses besoins. Par conséquent, l’allocation des ressources doit être dynamiquement adaptée pour répondre à ces demandes variables et maintenir une QoS équitable.

Contrairement aux travaux précédents, nous proposons de former les clusters et de contrôler dynamiquement les handovers en fonction des variations des canaux et des demandes des UEs. Pour résoudre notre problème, nous proposons une approche basée sur l’apprentissage par renforcement profond (DRL) [8]. Grâce à une interaction continue avec l’environnement, notre agent DRL réussit à apprendre une stratégie de regroupement, déterminant quand les handovers doivent se produire et quel AP doit abandonner quel UE il sert pour maintenir une QoS satisfaisante par UE. Par conséquent, le DRL automatisera le réseau, permettant d’optimiser l’allocation de ressource avec une latence réduite et de maintenir une répartition équitable des ressources entre les UEs.

2 Modèle du Système

2.1 Canal

Nous considérons un réseau "cell-free Multiple-Input Multiple-Output (CF-MIMO)" comportant M APs connectés à une unité centrale de traitement (CPU), et K UEs mobiles uniformément distribués dans le réseau. Soit $\mathcal{M} = \{1, \dots, M\}$ et $\mathcal{K} = \{1, \dots, K\}$ l’ensemble d’APs et d’UEs, respectivement. Chaque AP est équipé de N antennes, tandis que chaque UE dispose d’une seule antenne. Le canal $\mathbf{h}_{k,m} \in \mathbb{C}^{N \times 1}$ entre un UE k mobile et un AP m est modélisé par [9]

$$\mathbf{h}_{k,m}[n] = \sqrt{\frac{\kappa_{k,m}\beta_{k,m}}{\kappa_{k,m} + 1}} \tilde{\mathbf{h}}_{k,m} e^{j\phi_{k,m}^n} + \sqrt{\frac{\beta_{k,m}}{\kappa_{k,m} + 1}} \mathbf{R}_{k,m}^{1/2} \mathbf{f}_{k,m}[n] \triangleq \tilde{\mathbf{h}}_{k,m} e^{j\phi_{k,m}^n} + \tilde{\mathbf{R}}_{k,m}^{1/2} \mathbf{f}_{k,m}[n]. \quad (1)$$

où $\phi_{k,m}^n \sim \mathcal{U}_{[-\pi, \pi]}$ est le n -ième déphasage dans la composante du trajet direct (LoS) $\tilde{\mathbf{h}}_{k,m}$ et $\mathbf{f}_{k,m} \sim \mathcal{N}_{\mathbb{C}}(0, 1)$ est l’évanouissement à petite échelle de la composante multi-trajets (NLoS). Les termes $\kappa_{k,m}$, $\beta_{k,m}$ et $\tilde{\mathbf{R}}_{k,m}$ désignent le facteur de Rice, l’atténuation à grande échelle et la matrice de corrélation spatiale, respectivement. Nous supposons que ces paramètres restent fixés durant plusieurs blocs de ressources [9]. En raison de la mobilité de l’UE, le canal vieillit au fil du temps. Nous proposons de suivre cette modélisation [2]

$$\mathbf{h}_{k,m}[n] = \rho_k[\lambda - n] \mathbf{h}_{k,m}[\lambda] + \bar{\rho}_k[\lambda - n] \left(\tilde{\mathbf{h}}_{k,m} e^{j\phi_{k,m}^n} + \mathbf{g}_{k,m}[n] \right), \quad (2)$$

où $\bar{\rho}_k[\lambda - n] = \sqrt{1 - \rho_k^2[\lambda - n]}$. Le terme $\rho_k[\lambda - n]$ désigne le coefficient de corrélation temporelle du canal entre les instants λ et n . Nous supposons qu’il suit le modèle de Jake d’où $\rho_k[n] = J_0(2\pi f_{d,k} T_s (\lambda - n))$ [10]. Ici, $J_0(\cdot)$ est la fonction de Bessel d’ordre zéro de première espèce, T_s est le temps de cohérence du canal et $f_{d,k} = v_k f / c$ est le décalage Doppler tels que v_k , f et c désignent la vitesse de l’UE k , la fréquence porteuse et la vitesse de la lumière, respectivement. $\mathbf{g}_{k,m}[n] \sim \mathcal{N}_{\mathbb{C}}(0, \tilde{\mathbf{R}}_{k,m})$ représente la composante d’innovation.

Nous définissons la matrice de regroupement à l’instant t comme $\mathbf{C}(t) = \{c_{k,m}(t), \forall (k, m) \in \mathcal{K} \times \mathcal{M}\}$, représentant

les associations entre les K UEs et les M APs dans un bloc de cohérence de canal. Le terme $c_{k,m}$ dans \mathbf{C} indique si l'UE k est connecté à l'AP m , auquel cas $c_{k,m} = 1$; sinon $c_{k,m} = 0$.

En conséquence, le rapport signal sur interférence plus bruit (SINR) perçu par l'UE k à l'instant t est défini par

$$\text{SINR}_k(t, \mathbf{C}) = \frac{\left| \sum_{m=1}^M \mathbf{h}_{k,m}^H(t) \mathbf{w}_{k,m}(t) c_{k,m}(t) \right|^2}{\sum_{\substack{j=1 \\ j \neq k}}^K \left| \sum_{m=1}^M \mathbf{h}_{k,m}^H(t) \mathbf{w}_{j,m}(t) c_{j,m}(t) \right|^2 + \sigma_n^2}, \quad (3)$$

où σ_n^2 représente la variance du bruit au niveau du récepteur et $\mathbf{w}_{k,m}(t) \in \mathcal{C}^{N \times 1}$ désigne le faisceau transmis par l'AP m vers l'UE k qu'il sert à l'instant t

$$\mathbf{w}_{k,m}(t) = \sqrt{P_{k,m}^{\text{tx}}(t)} \frac{\mathbf{v}_{k,m}(t)}{\sqrt{\mathbb{E}\{\|\mathbf{v}_{k,m}(t)\|^2\}}}. \quad (4)$$

Ici, $P_{k,m}^{\text{tx}}(t)$ désigne la puissance de transmission allouée par l'AP m vers l'UE k à l'instant t . Dans notre étude, nous obtenons $\mathbf{v}_{k,m}$ avec un précodage par forçage à zéro (ZF) appliqué par l'AP m vers les UEs qu'il sert. Enfin, le débit de données perçu par l'UE k à l'instant t est donné par

$$R_k(t, \mathbf{C}) = B \log_2(1 + \text{SINR}_k(t, \mathbf{C})). \quad (5)$$

2.2 Débit impacté par la mobilité

Chaque UE est mobile dans le réseau, et le changement de sa position impacte significativement la qualité de connexion avec son groupe d'APs. Donc, l'UE a besoin d'ajuster son cluster d'APs pour maintenir une QoS satisfaisante. Cependant, pendant l'association à des nouveaux APs, les connexions ne sont pas complètement établies, réduisant la période de transmission de données. Par conséquent, ce retard réduit le débit perçu par l'UE devenant [3]

$$R'_k(t, \mathbf{C}) = R_k(t, \mathbf{C})(1 - d_C H_k^{\text{AP}}), \quad (6)$$

où d_C est la durée du handover et H_k^{AP} représente le taux de handovers effectués. Comme le schéma du taux de handovers proposé dans [6], H_k^{AP} de l'UE k est déterminé par la moyenne de handovers effectués durant sa trajectoire de durée T .

2.3 Niveau de satisfaction par UE

Soit $R_k^{\text{req}}(t)$ le débit de données demandé par l'UE k et $\xi_k(t)$ son niveau de satisfaction en terme de QoS à l'instant t que nous définissons par

$$\xi_k(t, \mathbf{C}) = \min\left(1, \frac{R'_k(t, \mathbf{C})}{R_k^{\text{req}}(t)}\right). \quad (7)$$

En particulier, lorsque $\xi_k = 1$, on dit que l'UE k est complètement satisfait, signifie qu'il perçoit au minimum le débit de données qu'il a demandé.

2.4 Formulation du Problème

Nous nous intéressons au problème de UCC dynamique pour maintenir une QoS satisfaisante par UE. Nous formalisons le problème de la manière suivante :

$$\text{maximiser}_{\Psi(t)} \lim_{T \rightarrow +\infty} \frac{1}{T} \sum_{t=1}^T \min_{k \in \mathcal{K}} \xi_k(t, \mathbf{C}), \quad (\mathcal{P})$$

$$\text{s.t. } c_{k,m}(t) \in \{0, 1\}, \quad \forall k \in \mathcal{K}, m \in \mathcal{M}, \quad (\mathcal{C}_1)$$

$$\sum_{k \in \mathcal{K}} c_{k,m}(t) P_{k,m}^{\text{tx}}(t) \leq P_{\max}^{\text{tx}}, \quad \forall m \in \mathcal{M}, \quad (\mathcal{C}_2)$$

$$\sum_{k \in \mathcal{K}} c_{k,m}(t) \leq K_{\max}, \quad \forall m \in \mathcal{M}, \quad (\mathcal{C}_3)$$

$$\sum_{m \in \mathcal{M}} c_{k,m}(t) \leq M_{\max}, \quad \forall k \in \mathcal{K}, \quad (\mathcal{C}_4)$$

où $\Psi(t) = \{c_{k,m}(t), P_{k,m}^{\text{tx}}(t), \forall (k, m) \in \mathcal{K} \times \mathcal{M}\}$. Dans le problème (\mathcal{P}) , la contrainte (\mathcal{C}_1) indique que les variables d'association $c_{k,m}$ sont binaires. La contrainte (\mathcal{C}_2) garantit que la puissance d'émission d'un AP m ne dépasse pas sa puissance de transmission maximale, notée P_{\max}^{tx} . Enfin, la contrainte (\mathcal{C}_3) limite le nombre d'UEs servis par chaque AP m à au plus K_{\max} UEs et la contrainte (\mathcal{C}_4) assure que chaque UE k est connecté à au plus M_{\max} APs, en raison de leurs capacités de calcul limitées.

Le problème (\mathcal{P}) est un problème d'optimisation non-convexe et NP-hard, donc difficile à le résoudre avec les méthodes d'optimisation classiques. Pour réduire sa complexité, nous proposons une approche d'allocation de puissance distribuée qui respecte la contrainte (\mathcal{C}_2)

$$P_{k,m}^{\text{tx}}(t, \mathbf{C}) = \frac{c_{k,m}(t) R_k^{\text{req}}(t)}{\sum_{k' \in \mathcal{K}} c_{k',m}(t) R_{k'}^{\text{req}}(t)} P_{\max}^{\text{tx}}, \quad \forall k, m \in \mathcal{K} \times \mathcal{M}. \quad (8)$$

Cette stratégie permet d'améliorer la QoS des UEs ayant des demandes élevées tout en optimisant celle des UEs confrontés aux pires conditions de canal. De cette façon, le processus de regroupement détermine la stratégie d'allocation de puissance.

3 Regroupement dynamique avec DRL

Notre objectif est de déterminer à chaque instant t , la matrice de regroupement \mathbf{C} appartenant à un grand espace de $2^{K \times M}$ d'associations possibles. L'apprentissage par renforcement profond (DRL) [8] émerge comme un outil puissant permettant une prise de décision intelligente dans des environnements de grandes dimensions et dynamiques, où les trafics demandés et les conditions de canal fluctuent de manière aléatoire.

3.1 Composantes principales du DRL

- *État* : À l'instant t , l'état comprend les normes des canaux $\|\mathbf{h}_t\| = \{\|\mathbf{h}_{k,m}(t)\|\}_{(k,m) \in \mathcal{K} \times \mathcal{M}}$, les connexions établies à l'instant précédent \mathbf{C}_{t-1} , les débits demandés $\mathbf{R}_t^{\text{req}} = \{R_k^{\text{req}}(t)\}_{k=1}^K$ et les handovers effectués par chaque UE jusqu'à cet instant $\mathbf{N}_t^{\text{h}} = \{N_k^{\text{h}}(t)\}_{k=1}^K$

$$s_t = \{\|\mathbf{h}_t\|, \mathbf{C}_{t-1}, \mathbf{R}_t^{\text{req}}, \mathbf{N}_t^{\text{h}}\} \quad (9)$$

- *Action* : L'agent DRL applique l'action suivante

$$a_t = \{c_{1,1}(t), \dots, c_{K,M}(t)\} \quad (10)$$

L'action a_t n'est autre que la matrice de regroupement \mathbf{C} .

- *Récompense* : Elle combine le minimum de QoS perçue avec le nombre de handovers effectués $N_{\text{total}}^{\text{h}}$ comme suit

$$r_t = \min_{k \in \mathcal{K}} \xi_k(t, \mathbf{C}) - \frac{N_{\text{total}}^{\text{h}}(t)}{N_{\max}^{\text{h}}}, \quad (11)$$

avec $N_{\max} = \min(K_{\max}M, M_{\max}K)$ désigne la borne supérieure de handovers possibles. Ici, notre objectif est de maximiser le minimum de QoS perçue tout en minimisant le nombre de handovers effectués à l'instant t .

3.2 UCC dynamique basé sur DRL

Il a été démontré que l'optimisation de politique proximale (PPO) surpasse d'autres algorithmes DRL tels que le q-learning profond (DQN), avec une complexité réduite [11]. En particulier, l'espace d'action de DQN croît exponentiellement à $2^{K \times M}$, ce qui le rend exhaustif en termes de calcul. En revanche, PPO utilise un espace d'action multi-discret, où chaque AP sélectionne indépendamment les UEs à servir, ce qui réduit considérablement la complexité faisant du PPO un choix plus pratique pour notre problème. L'algorithme 1 détaille notre approche de regroupement dynamique basé sur PPO. PPO comprend deux composantes clés, toutes les deux implémentées avec un réseau de neurones : l'*Acteur*, paramétré par ϕ , représente la politique π_ϕ responsable de la sélection des actions, et le *Critique*, avec le paramètre θ , estime la fonction de valeur V_θ pour guider les mises à jour de l'acteur. La fonction de perte totale de l'acteur est donnée par

$$\mathcal{L}(\phi) = \mathcal{L}_{\text{actor}}(\phi) - \beta \mathcal{H}(\pi_\phi), \quad (12)$$

où $\mathcal{L}_{\text{actor}}(\phi)$ est la fonction de perte de l'acteur calculée comme suit

$$\mathcal{L}_{\text{actor}}(\phi) = - \min \left(r^{\text{PPO}}(s_t, a_t) \hat{A}_t, \text{clip}(r^{\text{PPO}}(s_t, a_t), 1 - \epsilon, 1 + \epsilon) \hat{A}_t \right), \quad (13)$$

où

$$\hat{A}_t = \sum_{n=0}^{T-1} (\gamma \lambda_{\text{adv}})^n \delta_{t+n}. \quad (14)$$

$$\delta_t = r_t + \gamma V(s_{t+1}; \theta) - V(s_t; \theta) \quad (15)$$

et

$$r^{\text{PPO}}(s_t, a_t) = \frac{\pi(a_t | s_t; \phi)}{\pi_{\text{old}}(a_t | s_t)}. \quad (16)$$

Ici, \hat{A}_t désigne l'estimateur d'avantage généralisé (GAE) avec le paramètre λ_{adv} , δ_t est l'erreur de différence temporelle (TD), γ est le facteur d'actualisation, et r^{PPO} est le rapport de probabilité entre les nouvelles et les anciennes politiques. La fonction de clipping paramétrée avec ϵ permet d'éviter les mises à jour excessives et d'améliorer la stabilité.

Dans l'équation (12), $\mathcal{H}(\pi_\phi)$ désigne l'entropie dans un état donné et β est le coefficient d'entropie. Cette régularisation de l'entropie décourage la convergence prématurée vers une politique sous-optimale, encourageant l'exploration pour découvrir des meilleures stratégies de regroupement.

4 Expérimentation et Évaluation

Dans nos simulations, nous considérons un réseau CF-MIMO composé d'un CPU, $M = 4$ APs uniformément répartis dans une zone de $300 \times 300 \text{ m}^2$, où chaque AP est équipé de $N = 16$ antennes et $K = 10$ UEs mobiles, équipé chacun d'une seule antenne. Pendant $T = 250 \times T_s$, chaque UE choisit une direction aléatoire et se déplace à une vitesse maximale de 3 m/s. De plus, le débit de données demandé par chaque UE suit une distribution de Poisson avec une intensité tirée aléatoirement de l'ensemble $\mathcal{R}^{\text{req}} = \{300, 500, 700\}$ Mb/s à chaque déploiement aléatoire, et reste fixe durant la période T . Les paramètres $\kappa_{k,m}$ et $\beta_{k,m}$ sont modélisés comme dans [9]

Algorithme 1 : UCC dynamique basé sur PPO

```

1 Initialisation : Initialiser l'acteur ( $\phi$ ) et le critique ( $\theta$ )
2 pour  $e = 1, 2, \dots, N_{\text{episodes}}$  faire
3   Collecter des expériences avec  $\pi_{\phi_e}$ 
4   pour  $t = 1, 2, \dots, T$  faire
5     Observer  $s_t$  et appliquer  $a_t \sim \pi(\cdot | s_t; \phi_e)$ 
6     Calculer  $w_{k,m}$  et  $P_{k,m}^{\text{tx}}$  (8),  $\forall k \in \mathcal{K}, m \in \mathcal{M}$ 
7     Calculer le nombre de handovers par UE
8     Observer l'état suivant  $s_{t+1}$ 
9   fin
10  Calculer  $R'_k(t)$  (6),  $\forall k \in \mathcal{K}$  et  $r_t$  (11)
11  Calculer  $\hat{A}_t$  en utilisant  $V(\cdot; \theta_e)$ 
12  Calculer  $G_t = \hat{A}_t + V(s_t; \theta_e)$ 
13  Stocker l'épisode  $e$  dans la mémoire  $m_{\text{episodes}}$ 
14  si  $|m_{\text{episodes}}| = l$  alors
15    pour  $ep = 1, 2, \dots, N_{\text{epochs}}$  faire
16      Tirer aléatoirement  $B_s$  épisodes
17      Mettre à jour  $\phi$  en minimisant  $\mathcal{L}(\phi)$ 
18      
$$\phi_{e+1} = \arg \min_{\phi} \frac{1}{B_s T} \sum_{b=1}^{B_s T} \mathcal{L}_b(\phi) \quad (17)$$

19      Mettre à jour  $\theta$  en minimisant  $\mathcal{L}(\theta)$ 
20      
$$\theta_{e+1} = \arg \min_{\theta} \frac{1}{B_s T} \sum_{b=1}^{B_s T} (G_b - V(s_b; \theta))^2 \quad (18)$$

21    fin
22  fin

```

et la matrice de covariance $\tilde{\mathbf{R}}_{k,m}$ est calculée comme dans le papier [12, Sec. 2.6]. Nous supposons que $\kappa_{k,m}$, $\beta_{k,m}$, $\bar{\mathbf{h}}_{k,m}$ et $\tilde{\mathbf{R}}_{k,m}$ restent fixes durant $50 \times T_s$. Les autres paramètres sont fixés comme suit : $f = 3.5$ GHz, $B = 100$ MHz, $T_s = 2$ ms, $K_{\max} = 8$, $M_{\max} = 3$, $P_{\max}^{\text{tx}} = 250$ mW, $\sigma_n^2 = -174$ dBm/Hz, $d_C = 0.1$ s, $\gamma = 0.99$, $\lambda_{\text{adv}} = 0.95$, $\epsilon = 0.2$, $\beta = 0.01$, $l = 25$, $B_s = 8$, $N_{\text{epochs}} = 10$ et $N_{\text{episodes}} = 10^5$. Nous déployons 2 couches de réseau de neurones au niveau de l'acteur et du critique dont chacune est composée de 64 neurones avec tanh comme fonction d'activation, et nous effectuons la descente du gradient avec un taux d'apprentissage de 3×10^{-4} .

Nous comparons notre algorithme avec une approche de jeu de correspondance appelée *acceptation différée* (DA) [13] et un algorithme de regroupement proposé dans [14]. Dans un jeu DA, à chaque itération, les UEs demandent de s'associer aux APs en fonction de leurs préférences. Chaque AP, à son tour, ne conserve que K_{\max} UEs préférés dans sa liste d'attente, reportant la décision finale d'association à la dernière itération. Dans ce travail, nous choisissons la norme du canal comme métrique de préférence. Les auteurs de [14] proposent que chaque UE se connecte initialement au meilleur AP non saturé en terme de norme du canal. Ensuite, les UEs peuvent étendre leurs clusters en se connectant aux APs non saturés. Nous modifions cet algorithme, noté **MD**, afin de prendre en compte les limitations sur le nombre d'associations M_{\max} que chaque UE peut établir. Nous proposons de comparer notre **PPO** avec **DA** et **MD**, où nous appliquons chaque algorithme à chaque instant t pour déterminer les clusters. Étant donné la complexité de l'apprentissage qui croît exponentiellement avec la taille du réseau et la durée de simulation, nous avons pré-entraîné notre

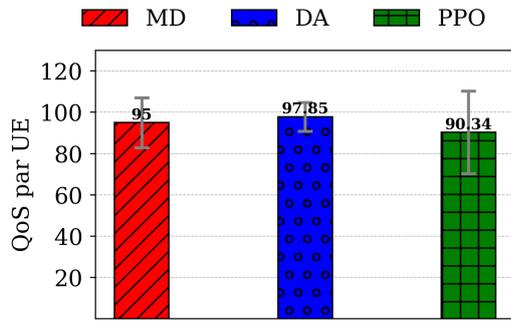


FIGURE 1 : Niveau de satisfaction par UE.

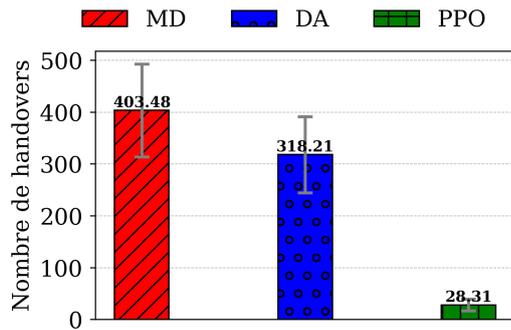


FIGURE 2 : Nombre de handovers durant une période T .

acteur-critique avec 1000 épisodes heuristiques générés par l'algorithme **DA**, pour bien initialiser les réseaux de neuronaux. Ensuite, la stratégie de regroupement dynamique a été affinée par un apprentissage sur 10^5 épisodes. Les résultats présentés sont issus de 1000 réalisations indépendantes du réseau.

La figure 1 montre que notre approche basée sur **PPO** garantit un niveau de QoS satisfaisant, en assurant en moyenne au moins 90% du débit requis pour chaque UE. En comparaison, notre méthode atteint un niveau de QoS similaire à celui obtenu par les approches **MD** et **DA**, avec une supériorité de **DA** sur **MD**, grâce à la formation de clusters favorisant des liens de meilleure qualité. Toutefois, la figure 2 met en évidence un nombre excessif de handovers induit par les approches **MD** et **DA**, soulignant que notre méthode offre un meilleur compromis entre QoS par UE et stabilité des associations. En particulier, pendant la phase d'apprentissage, notre DRL apprend une stratégie efficace de regroupement via l'exploration. Compte tenu de l'état actuel de l'environnement, l'agent DRL peut ajuster les clusters, contrairement aux algorithmes **MD** et **DA**, qui s'appuient uniquement sur les réalisations instantanées des canaux pour former les clusters et ignorent les connexions établies précédemment. Il en résulte que notre approche permet une réduction significative de la signalisation du lien fronthaul, où les connexions initialement établies étant suffisamment robustes nécessitant quelques handovers pendant une période de durée T comme le montre la figure 2. Ainsi, notre stratégie permet à un groupe d'APs de maintenir une association stable avec les mêmes UEs limitant la signalisation excessive et la latence induites par les reconfigurations fréquentes des groupes d'APs.

Par conséquent, notre approche permet non seulement d'automatiser le réseau, réduisant ainsi la latence induite par la signalisation requise pour établir les connexions entre les différents APs et UEs, mais également de limiter considérablement le nombre de handovers, faisant de notre solution une alternative plus réaliste que les approches **MD** et **DA**, souvent pénalisées par une complexité opérationnelle excessive.

5 Conclusion

Dans cet article, nous avons étudié le problème d'association dynamique des UEs dans les réseaux CF-MIMO pour maximiser la QoS, en considérant le vieillissement des canaux et le délai des handovers. Pour ce faire, nous avons proposé un algorithme basé sur le DRL. Les simulations montrent que notre solution fournit une QoS satisfaisante tout en réduisant le nombre de handovers. En tenant compte à la fois des demandes hétérogènes des UEs et des connexions établies, l'agent DRL apprend à ajuster dynamiquement les clusters d'APs associés à chaque UE, tout en minimisant le nombre de handovers. Par conséquent, l'agent parvient à apprendre une stratégie de regroupement d'APs centré sur l'UE, où les connexions sont suffisamment robustes dans le temps pour ne requérir que peu de handovers sur une période de mobilité donnée.

Les travaux futurs exploreront une implémentation multi-agents pour améliorer la résilience et la scalabilité du réseau.

Références

- [1] Stefano BUZZI *et al.* : Cell-free massive MIMO : User-centric approach. *IEEE Wireless Commun. Letters*, 2017.
- [2] Jiakang ZHENG *et al.* : Impact of channel aging on cell-free massive MIMO over spatially correlated channels. *IEEE Transactions on Wireless Communications*, 2021.
- [3] Yunlu XIAO *et al.* : Mobility performance analysis of scalable cell-free massive MIMO. *In ICC 2022-IEEE International Conference on Communications*. IEEE, 2022.
- [4] Robbert BEERTEN *et al.* : Cell-free massive MIMO in the O-RAN architecture : Cluster and handover strategies. *In IEEE Global Communications Conference*, 2023.
- [5] Mahmoud ZAHER *et al.* : Soft handover procedures in mmWave cell-free massive MIMO networks. *IEEE Transactions on Wireless Communications*, 2023.
- [6] Yunlu XIAO et Ljiljana SIMIĆ : A Novel Socially-Differentiated Handover Scheme for UE-Centric Cell-Free Massive MIMO. *In 2024 IEEE Wireless Communications and Networking Conference (WCNC)*, 2024.
- [7] Jianfeng SHI *et al.* : Power-efficient transmission for user-centric networks with limited fronthaul capacity and computation resource. *IEEE Trans. on Commun.*, 2020.
- [8] Kai ARULKUMARAN *et al.* : Deep reinforcement learning : A brief survey. *IEEE Signal Process. Mag.*, 2017.
- [9] Özgecan ÖZDOĞAN *et al.* : Performance of cell-free massive MIMO with Rician fading and phase shifts. *IEEE Transactions on Wireless Communications*, 2019.
- [10] William C. JAKES et Donald C. COX : *Microwave Mobile Communications*. Wiley-IEEE Press, 1994.
- [11] John SCHULMAN, Filip WOLSKI, Prafulla DHARIWAL, Alec RADFORD et Oleg KLIMOV : Proximal policy optimization algorithms. *arXiv :1707.06347*, 2017.
- [12] Emil BJÖRNSSON *et al.* : Massive MIMO networks : Spectral, energy, and hardware efficiency. *Foundations and Trends® in Signal Processing*, 2017.
- [13] Alvin E ROTH : Deferred acceptance algorithms : History, theory, practice, and open questions. *international Journal of game Theory*, 36:537–569, 2008.
- [14] Marx FREITAS *et al.* : Matched-decision AP selection for user-centric cell-free massive MIMO networks. *IEEE Transactions on Vehicular Technology*, 2023.