# Association Intelligente et Équilibrage de Charge dans les Réseaux WiFi Denses

Mohamed Bellouch, Véronique Vèque, Lynda Zitoune, Iyad Lahsen-Cherif

# *Association Intelligente et Équilibrage de Charge dans les Réseaux WiFi Denses*

Mohamed Bellouch[1], Véronique Vèque[1], Lynda Zitoune[1] et Iyad Lahsen-Cherif[2]

[1] *Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des Signaux et Systèmes, bât. Bréguet, 3, rue Joliot Curie, 91190 Gif-sur-Yvette, France.*
*Email:* {*mohamed.bellouch, veronique.veque, lynda.zitoune*}*@l2s.centralesupelec.fr*
[2] *INPT, Département Systèmes de Communications, Rabat, Maroc.*
*Email:* *lahsencherif@inpt.ac.ma*

Le déploiement massif des réseaux WiFi, alimenté par les caractéristiques prometteuses des nouvelles générations telles que le 802.11be et au-delà, répond de mieux en mieux aux besoins de connectivité et de débit croissant des utilisateurs. Ces déploiements sont désormais confrontés à de nouveaux défis tels que la densité des points d'accès (APs) dans une zone restreinte et la nature dynamique des accès à ces APs. Par conséquent, le développement de techniques efficaces d'équilibrage de charge via une stratégie d'association optimale des stations (STAs) aux APs, garantissant une équité entre les utilisateurs tout en optimisant les performances du réseau, demeure un problème ouvert. Dans ce travail, nous présentons DQL-MultiMDP, un algorithme d'équilibrage de charge basé sur le Deep Q-Learning (DQL) permettant d'atteindre cet objectif en relèvant le défi de la complexité de l'environnement en alternant entre plusieurs Processus de Décision Markoviens (MDPs) en fonction de la variation dynamique du nombre d'APs et de STAs. Les résultats expérimentaux démontrent son efficacité en termes de qualité d'association résultante et d'optimisation du débit global du réseau, en comparaison avec l'approche traditionnelle basée sur la puissance du signal reçu.

**Mots-clefs :** WiFi, Load Balancing, Markov Decision Process, Deep Q-Learning

## 1 Introduction

The deployment of modern WiFi networks is characterized by its dense and dynamic nature; the user Stations (STAs) experience diverse Quality of Service (QoS) requirements, ranging from connected sensors to devices running intensive applications like Augmented/Virtual Reality (AR/VR), gaming, and 8K videos. Some STAs are attached to the users and exhibit complex mobility. The Access Points (APs) are close to each other, resulting in overlapping cells. They can be managed by energy efficiency mechanisms that turn them on/off depending on traffic and demands, similar to the STAs that connect and disconnect unpredictably. Therefore, the IEEE 802.11 standards aim to improve existing Wireless Local Area Networks (WLANs) to accommodate these circumstances. The IEEE 802.11be or WiFi 7, promises to deliver Extremely High Throughput (EHT) and reduce latency [1]. It comes with several novel enhancements, such as wider bandwidths, Multi-link Operation, Multi-AP Cooperation, and Seamless Roaming to ensure that STAs can seamlessly switch between different APs, thus providing solid support for load balancing techniques.

Traditionally, each STA is associated with the AP with the highest received signal strength indicator (RSSI), and this may result in overloading some APs, unfairness among users, and degradation of the overall network performance [2]. The early load balancing algorithms, e.g. [3, 4], often use simplistic metrics, such as RSSI measurements and the number of STAs connected to the

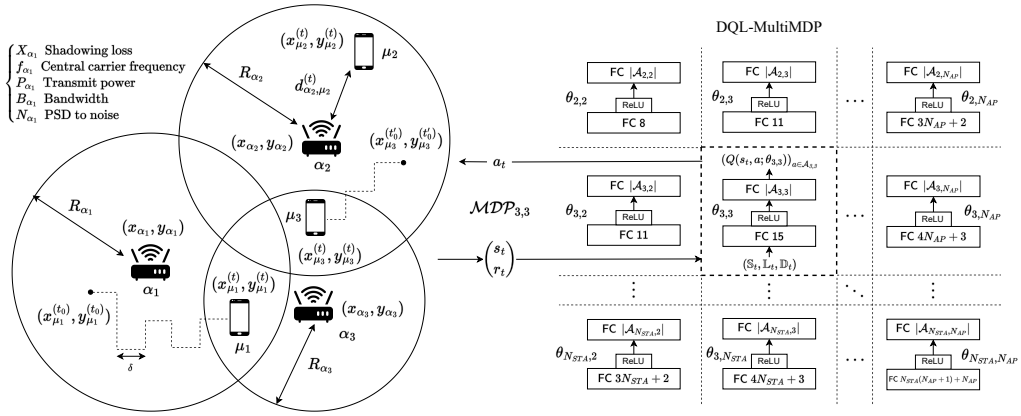


Fig. 1: Illustration of the used terminology

APs, which are not sufficient to establish the desired level of fairness while tackling the dynamic and dense nature of modern networks. Proposed Deep Reinforcement Learning (DRL)-based load balancing techniques, e.g. [5,6], successfully tackle the density and heterogenous requirements challenges. However, the design of these solutions and the experimental setup where their effectiveness is validated are often built upon settings where the number of nodes remains constant without considering their dynamic (on/off, mobility) behavior.

This paper presents DQL-MultiMDP, an algorithm for load balancing in dense and dynamic WiFi networks. Leveraging Deep Q-Learning (DQL), it takes complex state representations with advanced metrics to learn through autonomous decision-making an optimal STA-to-AP association policy that ensures long-term fairness among users, balances the load on the APs, and optimizes the overall network performance. Moreover, the approach alternates between the multiple Markov Decision Processes (MDPs) depending on the dynamic variation of the number of APs and STAs, thus providing a more fine-grained understanding of the environment's dynamics.

We proceed by introducing our system model for a dense WiFi network in Section 2, where the dynamics conform to the Markov property necessary in the RL formalism. In Section 3, we provide a brief overview of DQL-MultiMDP, and in Section 4, we discuss the results obtained to demonstrate its effectiveness. We conclude the paper in Section 5 with some insights and motivations for future work.

## 2 System model and assumptions

Let $t \in \mathbb{R}$ be a time variable. From the view of a single controller, the network at $t$ consists of a finite set of $N_{STA}^{(t)}$ active STAs $\mathcal{STA}^{(t)} = \{\mu_i\}_{1 \leq i \leq N_{STA}^{(t)}}$, and a finite set of $N_{AP}^{(t)}$ active APs $\mathcal{AP}^{(t)} = \{\alpha_j\}_{1 \leq j \leq N_{AP}^{(t)}}$. The association of the STAs at $t$ is modeled by a binary matrix $C^{(t)} = (c_{i,j}^{(t)})_{1 \leq i \leq N_{STA}^{(t)}, 1 \leq j \leq N_{AP}^{(t)}}$, such that, if $\mu_i$ is associated to $\alpha_j$ then $c_{i,j}^{(t)} = 1$, otherwise $c_{i,j}^{(t)} = 0$. To model the on/off switching behavior, we assign to any device $\beta$ (AP or STA) an independent Markov chain process $S_\beta = S_\beta^{(t_0)}, S_\beta^{(t_0+\eta)}, S_\beta^{(t_0+2\eta)}, \ldots$, where the state set is {'On', 'Off'}, and the index set is $\{t_0 + k\eta\}_{k \in \mathbb{N}}$, with $t_0$ being an initial time and $\eta > 0$ a time step. The APs are fixed in position, while the STAs move according to a Markov chain lattice random walk on $\mathbb{R}^2$ with a step size $\delta > 0$. Fig. 1 illustrates the used terminology with the considered radio channel characteristics (listed for $\alpha_1$). We denote by $SINR_{\mu,\alpha}^{(t)}$ the value of the Signal-to-Interference-and-Noise Ratio from the STA $\mu \in \mathcal{STA}^{(t)}$ to the AP $\alpha \in \mathcal{AP}^{(t)}$ at $t$. The achievable data rate of $\mu$ from $\alpha$ at $t$, denoted by $\rho_{\mu,\alpha}^{(t)}$, can be obtained from the Shannon-Hartley formula: $\rho_{\mu,\alpha}^{(t)} = B_\alpha \log_2(1 + SINR_{\mu,\alpha}^{(t)})$. We exclusively focus on the downlink transmission from the APs to the STAs, and we denote the required data rate by $\mu$ at $t$ as $d_\mu^{(t)}$.

# 3 DQL-MultiMDP

A pair $(N_{STA}, N_{AP})$ is initially defined and called the *controller's capacity*. If $N_{STA}^{(t)} > N_{STA}$ or $N_{AP}^{(t)} > N_{AP}$, the controller employs a traditional association strategy; if $N_{AP}^{(t)} = 1$, it associates all the active STAs to the single AP; if $N_{STA}^{(t)} = 1$, it associates the single STA to the AP with the highest RSSI; otherwise, when $2 \leq N_{STA}^{(t)} \leq N_{STA}$ and $2 \leq N_{AP}^{(t)} \leq N_{AP}$, the association problem will be seen as a MDP $\mathcal{MDP}_{N_{STA}^{(t)},N_{AP}^{(t)}} = (\mathcal{S}_{N_{STA}^{(t)},N_{AP}^{(t)}}, \mathcal{A}_{N_{STA}^{(t)},N_{AP}^{(t)}}, \mathcal{P}_{N_{STA}^{(t)},N_{AP}^{(t)}}, r_t, \gamma)$, where the elements are the state space, the action space, the transition probabilities function, the reward function, and a discount factor $\gamma \in (0, 1)$, respectively. When a change in the number of devices occurs after a certain $\tau > 0$ (such as an AP being turned off or a new STA joining the network), and generally results in $N_{STA}^{(t+\tau)} \neq N_{STA}^{(t)}$ or $N_{AP}^{(t+\tau)} \neq N_{AP}^{(t)}$, the controller switches to $\mathcal{MDP}_{N_{STA}^{(t+\tau)},N_{AP}^{(t+\tau)}}$ which models the new context. The observed state $s_t$ at $t$ is the tuple $(\mathbb{S}_t, \mathbb{L}_t, \mathbb{D}_t)$ where $\mathbb{S}_t$ is the SINRs matrix, $\mathbb{L}_t$ is the APs load matrix, and $\mathbb{D}_t$ is the required data rates matrix, all given in Eq. 1. $s_t$ can practically be represented by a single vector of size $N_{STA}^{(t)}N_{AP}^{(t)} + N_{AP}^{(t)} + N_{STA}^{(t)}$.

$$\mathbb{S}_t = \left(SINR_{\mu_i,\alpha_j}^{(t)}\right)_{\substack{1\leq i\leq N_{STA}^{(t)} \\ 1\leq j\leq N_{AP}^{(t)}}}, \mathbb{L}_t = \left(\sum_{i=1}^{N_{STA}^{(t)}} c_{i,j}^{(t)}\right)_{1\leq j\leq N_{AP}^{(t)}}, \text{and } \mathbb{D}_t = \left(d_{\mu_i}^{(t)}\right)_{1\leq i\leq N_{STA}^{(t)}}. \quad (1)$$

The applied action at $t$ is a binary matrix $a_t = (a_{i,j}^{(t)})_{1\leq i\leq N_{STA}^{(t)},1\leq j\leq N_{AP}^{(t)}}$ that represents the controller's association of the STAs to the APs; if the controller associates the $i$-th STA to the $j$-th AP, then $a_{i,j}^{(t)} = 1$, otherwise $a_{i,j}^{(t)} = 0$. Let $2 \leq n \leq N_{STA}$ and $2 \leq m \leq N_{AP}$; as associating a single STA to multiple APs requires further analysis of the Multi-AP Cooperation feature [1], basic DQL-MultiMDP associates each STA to one and only one AP. Thus, $a_t$ becomes a matrix where the elements in each line are all zeros except one that is 1. Therefore, we have $|\mathcal{A}_{n,m}| = m^n$.

By letting $[\mathbb{L}_t]_j$ denote the $j$-th element in $\mathbb{L}_t$, the received reward at $t$ is expressed as

$$r_t = \sum_{i=1}^{N_{STA}^{(t)}} \sum_{j=1}^{N_{AP}^{(t)}} c_{i,j}^{(t)} \frac{\rho_{\mu_i,\alpha_j}^{(t)}}{d_{\mu_i}^{(t)} [\mathbb{L}_t]_j}, \quad (2)$$

so that maximizing the sum of $\gamma$-discounted future rewards implies finding the optimal association strategy that maximizes the ratio between the Shannon capacities and the required data rates while minimizing the load on the APs.

As illustrated in Fig. 1, the controller stores a Q-Network for all the possible $(N_{STA}-1)(N_{AP}-1)$ MDPs. Each Q-Network, with parameters $\theta_{n,m}$, consists of Fully Connected (FC) layers with ReLu activation functions. It outputs $(Q(s_t, a; \theta_{n,m}))_{a\in\mathcal{A}_{n,m}}$ the approximated Q-values of all the available actions in $\mathcal{A}_{n,m}$ given the current state $s_t$ taken as input.

# 4 Experimental results

We used the `ns3-gym` toolkit to interface between DQL-MultiMDP and an IEEE 802.11ax network simulated on `ns3` and whose topology is illustrated in Fig. 2. The simulation environment adheres to the dynamics presented in Section 2. The APs have identical characteristics, with coverage areas satisfying $R_{\alpha_1} = R_{\alpha_2} = R_{\alpha_3} > 3D$. The station requirements randomly change over time.

All the devices are initially active except $\alpha_2$ and $\mu_1$. The controller then interacts with $\mathcal{MDP}_{8,2}$ and updates $\theta_{8,2}$. As shown in Figure 3, the episodes' sum of rewards evolves until $\alpha_2$ is turned on nearly after the 50th episode, so the environment transits to $\mathcal{MDP}_{8,3}$, resulting in a drop in the plot. $\mathcal{MDP}_{8,2}$ is observed again around the 120-th episode, and the evolution starts where it left off because $\theta_{8,2}$ was already trained. After the 160-th episode, the algorithm reaches the optimal policy in all the visited MDPs, outperforming the single MDP approach that models the

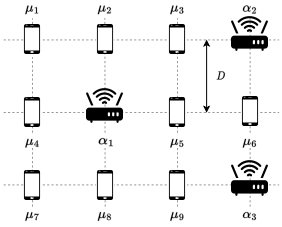


Fig. 2: Test setting

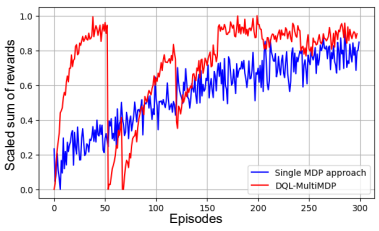


Fig. 3: MDP decomposition

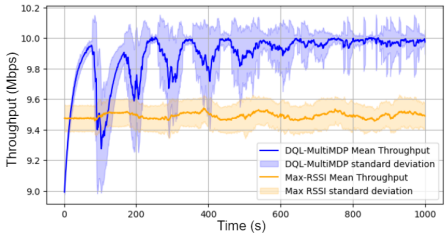


Fig. 4: Network performance

environment as only one MDP and uses only one Q-Network. Tab. 1 demonstrates the intelligent association obtained by the learned policy, which is more flexible than the Max-RSSI association strategy that associates each STA to the closest AP, causing a high load on $\alpha_1$. ($\mu_6$ is turned off).

Tab. 1: Association results

| **STA** | $\mu_1$ | $\mu_2$ | $\mu_3$ | $\mu_4$ | $\mu_5$ | $\mu_7$ | $\mu_8$ | $\mu_9$ |
|---|---|---|---|---|---|---|---|---|
| **STA's required rate (Mbps)** | 1 | 10 | 1 | 10 | 10 | 10 | 10 | 5 |
| **Max-RSSI association strategy** | $\alpha_1$ | $\alpha_1$ | $\alpha_2$ | $\alpha_1$ | $\alpha_1$ | $\alpha_1$ | $\alpha_1$ | $\alpha_3$ |
| **DQL-MultiMDP** | $\alpha_2$ | $\alpha_2$ | $\alpha_3$ | $\alpha_1$ | $\alpha_2$ | $\alpha_1$ | $\alpha_3$ | $\alpha_3$ |

Recent experiments and investigations revealed that the main factor behind imposing the controller capacity is the action space's growth. The controller capacity can be set for such worst-case scenarios at $(N_{STA}, N_{AP}) = (10, 4)$. However, DQL-MultiMDP can be adapted to control a network with a large number of devices by partitioning it into clusters. When tested in a network of 50 STAs and 16 APs divided into 5 clusters, Fig. 4 shows how the Max-RSSI strategy is outperformed after reaching the optimal policy in all the visited MDPs in about 400 s on average.

## 5 Conclusion

This paper presented DQL-MultiMDP, a DQL-based load-balancing algorithm that leverages a MDP decomposition approach to learn the optimal STA-to-AP association strategy in dense and dynamic WiFi networks. Experimental results validate the algorithm's effectiveness in both ensuring fairness among users and optimizing the network's performance; encouraging further improvements to meet WiFi's new features, such as Muli-link Aggregation and Multi-AP Cooperation.

## References

[1] E. Khorov, I. Levitsky, and I. F. Akyildiz, "Current status and directions of IEEE 802.11 be, the future wi-fi 7," *IEEE access*, vol. 8, pp. 88664–88688, 2020.

[2] I. Papanikos and M. Logothetis, "A study on dynamic load balance for IEEE 802.11 b wireless LAN," in *Proc. COMCON*, vol. 2001, 2001.

[3] S.-T. Sheu, C.-C. Wu, *et al.*, "Dynamic load balance algorithm (DLBA) for IEEE 802.11 wireless LAN," *Journal of Applied Science and Engineering*, vol. 2, no. 1, pp. 45–52, 1999.

[4] H. Gong and J. Kim, "Dynamic load balancing through association control of mobile users in WiFi networks," *IEEE Transactions on Consumer Electronics*, vol. 54, no. 2, pp. 342–348, 2008.

[5] M. A. Kafi, A. Mouradian, and V. Vèque, "On-line client association scheme based on reinforcement learning for wlan networks," in *2019 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–7.

[6] R. Ahmad, M. D. Soltani, M. Safari, and A. Srivastava, "Reinforcement learning-based near-optimal load balancing for heterogeneous LiFi WiFi network," *IEEE Systems Journal*, vol. 16, no. 2, pp. 3084–3095, 2021.