Deep Deterministic Policy Gradient for Handovers in Mobile Networks with Transparent UAV Relays

Ramsha Narmeen, Zdenek Becvar, Pavel Mach
Faculty of Electrical Engineering, *Czech Technical University in Prague*, Prague, Czech Republic
{narmeram, zdenek.becvar, machp2}@fel.cvut.cz

***Abstract*—In this paper, we introduce a novel framework jointly managing handovers of user equipments (UEs) and Unmanned Aerial Vehicles (UAVs) serving the UEs. The goal is to maximize the sum capacity of the UEs while considering a cost related to the handovers. To this end, we introduce a novel approach based on deep deterministic policy gradient (DDPG) adjusting the Cell Individual Offset (CIO) for handovers of the UEs among the UAVs and ground base stations (GBSs) as well as handovers of the UAVs among the GBSs. The UAVs playing the role of relays often face challenges related to the implementation cost and energy limitations. To address these challenges, the UAVs should operate in a transparent relaying mode. In such mode, unfortunately, the channels between the UEs and the UAVs are unknown as the transparent relays lack any communication control-related functionalities. Therefore, we adopt a deep neural network (DNN) to predict the channel qualities among the UEs and the UAVs for the handover purposes. We demonstrate that the proposal significantly increases the sum capacity of the UEs by dozens of percent and even reduces the number of handovers compared to state-of-the-art works. At the same time, the proposed DDPG-based CIO setting reduces a gap in the sum capacity between the predicted and the optimal (but practically not feasible) case with perfectly known channels among UEs and UAVs. Hence, the proposal is suitable for practical scenarios with not perfectly accurate channel quality information.**

***Index Terms*—Handover, machine learning, transparent relays, unmanned aerial vehicles, users, cell individual offset.**

I. Introduction

Unmanned aerial vehicles (UAVs) equipped with wireless communication hardware acting as flying base or relay stations, are considered to be a promising solution for future mobile networks due to a high adaptability and flexibility. The UAVs have a potential to extend the network coverage and enhance quality of service (QoS) in highly loaded areas [1]. However, integration of the UAVs into the mobile networks introduces challenges, such as optimizing the placement of the UAVs, the trajectory and power allocation of the UAVs, or handovers of user equipments (UEs) among the UAVs and ground base stations (GBSs) as well as handovers of the UAVs among the GBSs [2]. In this paper, we focus on the problem of handovers of both the UEs and the UAVs.

In traditional mobile networks, the handover of UEs among GBSs is typically triggered when a target GBS offers the channel with a superior quality compared to the channel from a serving GBS [3]. To mitigate frequent handovers, the decision to initiate the handover is fine-tuned using control parameters like hysteresis, time-to-trigger (TTT), or cell individual offset (CIO) [4]. The works targeting only GBSs, however, neglect aspects of dynamicity related to the UAVs' movement. The UAVs change their positions over time following an arbitrary movement of the served UEs [5]. Simultaneous movement of both UAVs and UEs may introduce rapid changes in the quality of all channels resulting in frequent and unpredictable handovers of the UEs among the UAVs and the GBSs as well as handovers of the UAVs among the GBSs [6]. Such frequent and unpredictable handovers can lead to handover failures, packet losses, and overloading of certain GBSs [7].

There are few studies targeting the optimization of the handover of the UAVs acting as the UEs, i.e., the UAVs are not serving any other UEs. In these works, the authors minimize the number of handovers via a dynamic adjustment of GBSs' antenna tilt [3], by alternation of the UAV handover parameters [8], or by reinforcement learning [9]. Despite promising results, [3], [8], [9] assume scenarios with predefined and a priory known UAV trajectories. However, this assumption does not hold in the scenarios with the UAV serving the UEs, where the UAV trajectories are unknown and depend on the movement of UEs. The same limitation applies also for traditional mobile relays deployed on trains or public transportation vehicles, as addressed, e.g., in [10]. Besides, the traditional mobile relays are (almost) static from the respective of the served UEs. Such assumptions do not hold for the UAVs serving the UEs.

Generally, the UAVs relay user data between the GBS and the UEs in a non-transparent mode. In such mode, the UAV relays carry out all the communication control and management functions, similar to the traditional GBSs. However, such comprehensive management leads to a high complexity, weight, and energy consumption [11], making the non-transparent relaying impractical for the energy-constrained UAV relays. Therefore, the UAV relays should operate in a transparent mode making the relays less complex, resulting in lighter, more cost-effective, and energy-efficient solution compared to the non-transparent relays [2]. In case of the transparent UAV relays, the GBS retains control of the communication management. Unfortunately, this also means that the channel quality between the UE and the UAV is unknown, since the transparent relays do not transmit their own reference signals and only forward the data symbols [12]. This is a serious obstacle in a deployment of the transparent relays in practice [13]. Still, the quality of the access channel between

This work was supported by the project No. 23-05646S funded by the Czech Science Foundation.

the UE and the UAV can be predicted from the quality of channels from the UE and the UAV to few surrounding GBSs using a deep neural networks (DNN) [14], making the use of transparent relay UAVs feasible [15].

In this paper, we adopt a scenario with predicted access channel quality between the UEs and the UAVs and we focus on the optimization of the CIO for the handover decision of all moving devices, i.e., UEs and UAVs. The key challenge is to develop a solution that is resistant to the access channel quality prediction error. Traditional Q-learning or actor-critic-based deep reinforcement learning (DRL), as adopted in [5], [9], does not cope well with the channel prediction errors. The lack of an experience replay buffer and an effective loss function do not allow the conventional DRL ability to learn from past experiences, making it inefficient for handover management with potentially inaccurate channel quality. To overcome this limitation, we propose a novel method based on deep deterministic policy gradient (DDPG) that intelligently predicts the CIO while minimizing the impact of prediction errors in channel quality. The DDPG is adopted due to: *i)* the dynamic and unpredictable behavior of the UEs and, consequently, also of the UAVs serving these UEs, and *ii)* an indirect and unknown relation between the current decisions (setting of CIO) and their future effects on the network.

The main contributions of this paper are summarized as follows:

- We introduce a novel framework optimizing CIO of GBSs and UAVs to increase the sum capacity of the UEs while minimizing the number of handovers. We adopt deep reinforcement learning based on DDPG to predict the optimized CIO settings. The DDPG incorporates a replay buffer mechanism storing experiences for efficient learning and enhances efficiency by allowing the system to learn from the past experience making it suitable for practical scenario with potentially inaccurate channel quality information.
- We show that the proposed DDPG-based CIO setting outperforms traditional DRL-based handover decision and, in addition, the proposal is also resilient to the potential channel quality errors, since the DDPG adaptively adjusts its policy learning based on a feedback from the environment, allowing it to adapt and optimize the actions despite uncertainties in the channel quality prediction.

The rest of this paper is organized as follows. Section II introduces the system model. Then, in Section III, the targeted problem is formulated. Section IV elaborates on our proposed solution based on DDPG to determine the CIOs of the GBS and UAVs. Simulation results and discussions are presented in Section V. Finally, Section VI concludes this paper.

II. System Model

In this section, we first provide details of the network, communication, and handover models adopted in this paper. Then, we describe the DNN employed to predict the channel qualities from the transparent UAVs to the UEs.

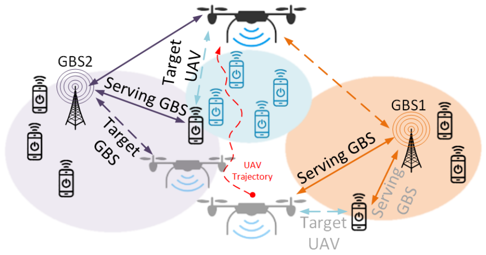

Fig. 1. System model with UEs handovering among all BSs and UAVs serving the UEs and performing handovers among GBSs.

A. Network model

We consider N UEs deployed in an area covered by G traditional GBSs and by F UAVs acting as the energy-efficient flying transparent relays (see Fig. 1). The coverage in the area is provided in total by $K = G + F$ base stations (BSs), encompassing GBSs and UAVs.

The position of the n-th UE is arbitrarily changed over time and the UAVs move in accordance with the movement of the connected UEs. Note that a mobility models of the UEs and the UAVs have no impact on the solution proposed in this paper. Hence, we do not specify these models in this section.

B. Communication model

We consider downlink communication, where signals from the GBSs reach the UEs either directly from the GBSs or via the UAVs. The Signal-to-Interference-plus-Noise Ratio (SINR) $\gamma_{i,j}$ experienced by the j-th receiver directly served by the i-th transmitter, is defined as:

$$\gamma_{i,j} = \frac{p_i h_{i,j}}{\sigma^2 + \sum_{\forall_{i'} \in K/i} p_{i'} h_{i',j}} \tag{1}$$

where p_i is the transmission power of the i-th transmitter, $h_{i,j}$ is the channel quality between the i-th transmitter and the j-th receiver, σ^2 is spectral noise density, $\sum_{\forall_{i'} \in K/i} p_{i'} h_{i',j}$ is the interference from other transmitters, using the same band, $p_{i'}$ is the transmission power of the i'-th transmitter, and $h_{i',j}$ is the quality of the channel from the i'-th transmitter to the j-th receiver. The transmitter is represented either by the g-th GBS or by the f-th UAV (i.e., $i \in \{g, f\}$) and the receiver is either the f-th UAV or the n-th UE (i.e., $j \in \{f, n\}$).

The communication capacity for the n-th UE directly served by the g-th GBS is defined as:

$$c_{g,n} = B_n \log_2(1 + \gamma_{g,n}) \tag{2}$$

where B_n is the bandwidth requested by the n-th UE to meet c_{req} and is defined as $B_n = \frac{c_{\text{req}}}{\log_2(1+\gamma_{g,n})}$. The bandwidth allocation is independent of the handover decision. Therefore, we assume that bandwidth is assigned to the UEs to fulfill the minimum required capacity c_{req} considering the UEs' SINR. The bandwidth is assigned in descending order, prioritizing the UEs with the highest SINR first, since these UEs require the lowest bandwidth. This process is iterated for subsequent UEs until the available bandwidth is sufficient to meet the requirements of additional UEs [5].

The communication capacity in the case of relaying from the g-th GBS via the f-th UAV to the n-th UE with each hop assigned with a half of the resources, as in [16], is defined as:

$$c_{g,f,n} = \frac{1}{2} B_n \min\left\{\log_2(1+\gamma_{g,f}), \log_2(1+\gamma_{f,n})\right\} \quad (3)$$

Then, in general, the communication capacity of the n-th UE communicating with the GBS directly or via the UAV is:

$$c_n = \begin{cases} c_{g,n}, & \text{if the UE is directly served by GBS} \\ c_{g,f,n}, & \text{if the UE is served via UAV} \end{cases} \quad (4)$$

The k-th BS serves a group of receivers imposing the total load ρ_k^{UE}. The load is characterized as ratio of bandwidth B_n allocated to n-th UE served by k-th BS to the total bandwidth B. Then, the total load of k-th BS is the sum of loads imposed by all UEs (ρ_k^{UE}) and UAVs (ρ_k^{UAV}) and is calculated as:

$$\rho_k = \rho_k^{\text{UE}} + \rho_k^{\text{UAV}} = \frac{\sum_{\forall_{n \in N}} \beta_{k,n}^{\text{UE}} B_n}{B} + \frac{\sum_{\forall_f \in F} \beta_{g,f}^{\text{UAV}} B_n}{B} \quad (5)$$

where $\beta_{k,n}^{\text{UE}} \in \{0,1\}$ indicates if the n-th UE is attached to the k-th BS ($\beta_{k,n}^{\text{UE}} = 1$) or not ($\beta_{k,n}^{\text{UE}} = 0$) and $\beta_{g,f}^{\text{UAV}} \in \{0,1\}$ indicates if the f-th UAV is attached to the g-th GBS ($\beta_{g,f}^{\text{UAV}} = 1$) or not ($\beta_{g,f}^{\text{UAV}} = 0$).

C. Handover decision

The handover of the UEs between the serving BS and the target BS is initiated based on the commonly adopted A3 event defined by 3GPP [17]. Therefore, the UEs perform handover from the serving BS to the target BS when the following inequality holds true for a period of Time-To-Trigger (TTT):

$$p_t h_{t,n} + CIO_t - \Delta > p_s h_{s,n} + CIO_s \quad (6)$$

where the indexes s and t indicate the serving and target BSs, respectively, $CIO_{s/t}$ is the CIO of the serving/target BS, and Δ is the handover hysteresis.

The UAVs also perform handovers among the GBSs. Similar to operation of common UEs in mobile networks, UAV measures the channel quality from neighboring GBSs. The channel quality measurement report is transmitted to serving GBS periodically in the same way as common UEs report their channel quality. Based on the measurement results, UAV initiates handover to a neighboring GBS if the condition specified in (6) is met for a period of TTT.

D. Prediction of channel qualities between UAVs and UEs

In this paper, we assume the UAVs as transparent relays, presenting a challenge due to their inability to directly acquire the channel quality between UEs and UAVs for handover purposes [11], [13]. To address the issue of unknown channel quality, we adopt DNN to predict the channel quality between UEs and UAVs solely from information available in the network, as suggested in [15]. The DNN-based UAV to UEs channel quality prediction leverages the known quality of channels between UEs and serving and few neighboring GBSs and between UAV and few neighboring GBSs. By utilizing the known channel quality from UEs and UAV to the serving and neighboring GBSs, DNN predicts the quality of direct channel between UE and UAV [15].

The architecture of DNN predicting the UAV to UEs channel quality consists of an input layer, H hidden layers, and an output layer. Initially, the channel qualities of UEs and UAVs to the serving and few neighboring GBSs are inserted into the DNN's input layer. Subsequently, the input channel qualities undergo processing through H hidden layers that are fully connected and are followed by a sigmoid activation function. The output layer is activated with function allowing to predict continuous access channel quality between UAV and UE.

III. Problem Formulation

In this paper, we optimize the handover of the UEs among the UAVs and GBSs jointly with the handover of the UAVs among GBSs. The primary goal is to adjust the CIOs of all BSs $\text{CIO}^* = \text{CIO}_1, ..., \text{CIO}_K$ to maximize the sum capacity of the UEs. Setting a low CIO for an overloaded BS while assigning a higher CIO to neighboring BSs enables to redistribute the UEs from the overloaded BS to adjacent BSs, see (6). Thus, incorporating CIO into the handover decision allows to improve the communication capacity of the UEs by steering handovers towards underutilized BSs that can provide more radio resources. Solely optimizing the sum capacity may result in an excessive number of handovers leading to an increased signaling overhead and energy consumption, which is undesirable for the UAVs. Therefore, we also consider a cost of performed handovers μ represented in practice, e.g., by signaling, handover interruption, or extra energy consumption [5]. Then, we define the targeted problem as:

$$\begin{aligned} \text{CIO}^* = \underset{\text{CIO} \in O}{\arg\max} & \sum\nolimits_{n \in N} c_n - \mu \\ a) \quad & c_n > c_{\text{req}}, \forall n \\ b) \quad & \sum\nolimits_{k=1}^{K} \beta_{k,n}^{\text{UE}} = 1, \forall_n \\ c) \quad & \sum\nolimits_{g=1}^{G} \beta_{g,f}^{\text{UAV}} = 1, \forall_f \end{aligned} \quad (7)$$

where $O = \langle CIO_{min}, CIO_{max} \rangle$. The constraint (7a) ensures that each UE receives the minimum required capacity, the constraint (7b) ensures that each UE is associated with just one BS (either GBS or UAV), and the constraint (7c) limits each UAV to be associated to just one GBS.

The major challenge in solving the problem outlined in (7) is a high randomness driven by arbitrary and hard-to-predict mobility patterns of both the UEs and the UAVs. Addressing this challenge typically necessitates the application of nonlinear optimization techniques. However, such techniques rely on having a precise knowledge of the network state, including the locations of the UEs and the UAVs, which may not always be available and accurate. Furthermore, even having perfect information of all relevant parameters, solving this optimization problem is NP-hard due to its formulation as a non-convex function. In addition, any future impact of the current decision (CIO setting) is unpredictable due to the unknown future movement of UEs and UAVs. Hence, we adopt deep reinforcement learning based on DDPG to adjust the CIO for the handover of UEs among all BSs and the CIO for the handover of UAVs among GBSs, described in the next section.

IV. Proposed CIO Adjustment using DDPG

In this section, we first provide a brief overview of the background in deep reinforcement learning relevant to our specific problem. After that, we present details of the proposed DDPG-based approach for the CIO adjustment in the networks with transparent UAVs.

A. Converting CIO adjustment problem to MDP framework

To apply the actor-critic-based DDPG algorithm for the dynamic CIO setting in the frame of the handover, we interpret the optimization problem as the MDP. The MDP is defined by a 4-tuple (S, A, P, R), where S and A represent finite sets of states and actions, respectively, P denotes the probability of transition from the state s to the state s' based on the taken action a, and R presents the immediate reward obtained by taking the action a. Individual components of the MDP in relation to our problem are elaborated as follows.

State: The state comprises load of BSs performing handover. The implied load is proportional to required resources to meet c_{req} for UEs/UAVs performing the handover. The state space $S(t)$ at time t is defined as $S(t) = [\rho_1(t), \dots, \rho_k(t)]$.

Action: The action $A(t)$ is defined as a selection of the CIO for all K BS at the time t, i.e., the action space is defined as $A(t) = [CIO_1(t), \dots, CIO_K(t)]$.

Reward: The reward function reflects targeted problem formulated in (7), i.e., to maximize sum capacity of UEs served by BS while avoiding handovers failures taking into account the penalty associated to the cost of handovers μ. Hence, the reward function is defined as $r(t) = \sum_{n \in N} c_n - \mu$.

B. Proposed DDPG-based CIO adjustment

The proposed concept of DDPG for CIO setting in the environment with DNN-based prediction of channel quality between the UEs and transparent UAVs is shown in Fig. 2. The UEs to UAV channels are predicted using DNN based on [15] and are fed to DDPG. The DDPG is built on the actor-critic framework, where the actor generates actions, and the critic directs the actor to adjust the actions towards a higher reward. On top of the critic and actor DNNs, DDPG includes also a loss function, and a replay memory buffer helping to cope with uncertainty in the accuracy of the DNN-based channel quality prediction. The replay memory stores the experience tuples with the present state $S(t)$, the chosen action $A(t)$, the immediate reward $r(t)$, and the next state. The replay memory plays a crucial role in mitigating the problem of temporal correlation in the experience tuples and improving the overall stability and efficiency of the highly dynamic handover environment. By randomly selecting experiences from the memory, the agent can break the sequential nature of the experiences, learn from a broader range of situations, and ultimately improve the CIO setting of BSs. The replay buffer allows the system to learn from past experience to mitigate the propagation of the channel quality prediction error from the DNN to the CIO setting.

The critic DNN in DDPG evaluates the actions, i.e., the CIO setting, by estimating the total reward resulting from the taken actions. More specifically, the critic DNN is trained to estimate the cumulative reward $R(s(t), a(t))$ representing the total expected sum of the rewards that the agent can accumulate by taking the action a in the state s. The critic DNN estimates the Q-value using the following steps: 1) the critic takes the value of the state and the action (i.e., the load of the BSs and the CIO setting), 2) the critic DNN performs a forward pass through the DNN using a function approximation, and 3) the estimated Q-value is extracted from the output layer of the critic DNN.

The actor DNN measures a target Q-value or a target cumulative reward, defined recursively using Bellman equation:

$$R(s(t), a(t)) = r(t) + \epsilon \max R(s(t+1), a(t+1)) \qquad (8)$$

where, ϵ is the discount factor, $s(t+1)$ is the next state for the next action $a(t+1)$ and $\max R(s(t), a(t))$ represents the maximum expected cumulative reward in the next state.

The training of the actor and critic DNNs aims at minimizing the difference between the predicted and target Q-values by mean squared error loss function L, defined as:

$$L = \frac{1}{D} \sum_{l} [R(s_l(t), a_l(t)) - R'(s_l(t), a_l(t))]^2 \qquad (9)$$

where D is the number of samples in the training batch and $\sum_l$ represents the summation over all transitions, where the transition is an experience tuple consisting of state, action, reward, and next state. The critic DNN is updated by minimizing the loss function (9) with respect to the parameter θ^Q, which is updated until convergence is met. The actor DNN is updated by computing the gradient of the expected return J with respect to the actor parameters $\theta^\delta, \nabla_{(\theta^\delta)} J \approx 1/D \sum_l \nabla_a R(s, a|\theta^Q)|_{(s=s_l, a=\delta(s_l))} \nabla_{(\theta^\delta)} \delta(s|\theta^\delta)|_{(s_l)}$, where $\nabla_a R(s, a|\theta^Q)$ computes the gradient of the critic's policy with respect to the action a, and $\nabla_{(\theta^\delta)}\ \delta(s|\theta^\delta)$ evaluates the actor's policy with respect to the critic's policy θ^δ for each state s. The expected return is updated with respect to the actor parameter θ^δ using the gradient $\nabla_{(\theta^\delta)} J$, which guides the learning process of the actor DNN to improve the policy for reinforcement learning-based CIO adjustment.

The proposed DDPG for CIO setting has a low computational complexity with the number of mathematical operations being approximately 2444 for $G = 4$ and $F = 4$ [15]. Therefore, the computational demands are considered negligible and do not hinder real-time processing at BSs with the computing power currently available in the mobile networks.

V. Performance Evaluation

In this section, we outline simulation models and setup. Then, we discuss related state-of-the-art works considered for comparison. Last, we present and discuss simulation results.

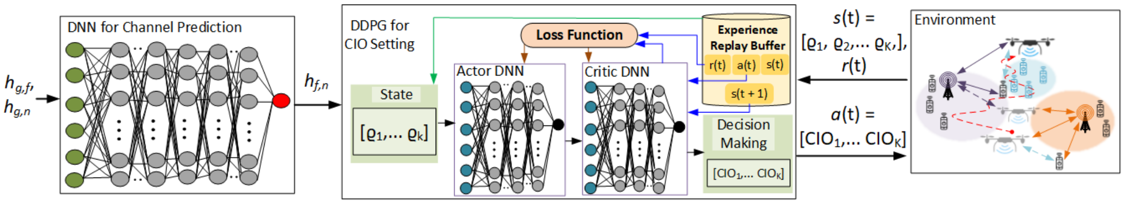

Fig. 2. Proposed DDPG for CIO setting suitable for the environment with transparent UAV relays exploiting the UAV to UE channel prediction via DNN.

A. Simulation models and settings

Simulations are conducted in a suburban setting with an area of 1000 × 1000m where four traditional GBSs are deployed randomly with a minimum inter-site distance of 500 m. Additionally, up to six UAVs are added to the simulation area. In order to show the effects of the handover setting, we utilize a comprehensive model incorporating dynamic movement of both UEs and UAVs, as presented in [5], which allows for a detailed examination of the complex interactions and transitions in the network. There are in total 150 UEs, each moving with a random speed ranging from 1 to 3 m/s, with all UEs consistently active. Among these UEs, 60 UEs are uniformly distributed around GBSs in a circular area with a radius of 150 m, moving arbitrarily within this region. Another 30 UEs are evenly dispersed across the entire simulation area, following a random waypoint mobility model [18]. The remaining 60 UEs adhere to a cluster movement model [19], where UEs are distributed evenly across six clusters, each cluster containing a varying number of UEs. UEs within a cluster are restricted to a circular region with an 80-meter radius, and all UEs follow the movement of the cluster center. The cluster movement within the simulation area aligns also with a random waypoint mobility model [18]. The movement of UAVs corresponds to the center of gravity of the UEs attached to this UAV [20]. The code of implemented proposal is available at GitLab [1].

B. Competitive state-of-the-art works

We compare the proposed DDPG-based CIO setting (denoted as *Proposed DDPG CIO*) to the following competitive state-of-the-art algorithms:

- Actor-critic DRL for CIO setting (*AC-DRL CIO*): The actor critic-based DRL used for the CIO setting according to the load of GBSs and handover cost, as presented in [5] to maximize the sum capacity is the closest and most recent state-of-the-art work.
- *Adaptive CIO*: The algorithm, described in [21], sets CIO according to predefined load thresholds of GBSs, aiming to minimize the number of handovers.

For all algorithms, we evaluate performance for the *predicted* and *actual* channel qualities labeled in figures as *Pred. chan* and *Act. chan.*, respectively. The prediction of the UEs to UAV channel quality is done using DNN proposed in [15]. In case of the actual channel qualities, the DNN shown in Fig. 2 does not apply, as we assume the access channels are known even though it is practically infeasible for the transparent UAV relays. Still, considering also the theoretical case with actual channels in evaluations allows to demonstrate the effectiveness of the proposed DDPG-based CIO setting for handling the error imposed by the DNN channel quality prediction.

C. Discussion of results

Fig. 3 investigate the sum capacity over varying minimum required capacity by UEs c_{req}. The sum capacity increases with c_{req}, as the resources are used in a more efficient way (resources are allocated first to the UEs with a good channel) for all algorithms. The improvement in sum capacity by the proposed DDPG is up to 12.2% and 36.1% compared to AC-DRL and adaptive CIO, respectively, for the predicted channel quality case. The gain results from the capabilities of DDPG to effectively deal with prediction errors in DNN-based channel quality prediction. If the actual channels between UEs and UAVs would be theoretically known, the proposed DDPG for the CIO setting improves sum capacity by up to 5 % compared to the case with predicted channels. However, for AC-DRL and Adaptive CIO, the sum capacity is degraded notably by about 6.6% and 9%, respectively. Hence, the DDPG suppresses negative impact of DNN-based channel quality prediction for transparent relays by more than 24% and 44% compared to AC-DRL CIO and Adaptive CIO, respectively.

In Fig. 4, we illustrates the sum capacity for varying number of UAVs. The sum capacity increases with the number of UAVs as more UAVs improve SINR in the network despite increased interference among the UAVs. The increase in sum capacity by the proposed DDPG is up to 10.1% and 20.2% compared to AC-DRL and Adaptive CIO, respectively, for the predicted channel qualities. Compared to the theoretical case with all access channels perfectly known, the channel quality prediction leads to a decrease in sum capacity of the proposed CIO setting by up to 4.6%, while AC-DRL and Adaptive CIO leads to a drop of 7.2% and 9.7%, respectively. This shows that DDPG suppresses negative impact of the DNN-based channel quality prediction by roughly 36% and 53% compared to AC-DRL CIO and Adaptive CIO, respectively.

In Fig. 5, we show sum of the number of handovers of UEs and UAVs. The number of handovers increases with number of UAVs as more UAVs perform handovers and also UEs have more opportunities for handover as well. The proposed DDPG-based CIO setting for predicted channel qualities reduces number of handovers by up to 13.0% and 35.2% compared to AC-DRL and Adaptive CIO, respectively. The channel quality prediction results in additional handovers for all algorithms

[1] Code of the proposal in Matlab: https://gitlab.fel.cvut.cz/mobile-and-wireless/codes/publications/deep-reinforcement-learning-for-handovers-in-mobile-networks-with-transparent-uav-relays

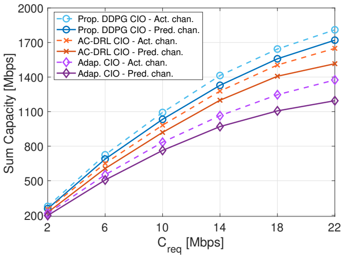

Fig. 3. Sum capacity for varying values of c_{req}, Number of UAVs $F = 4$.

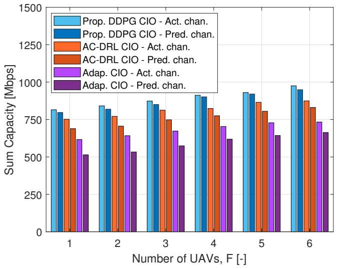

Fig. 4. Sum capacity for a varying number of UAVs, $c_{\text{req}} = 8$ Mbps.

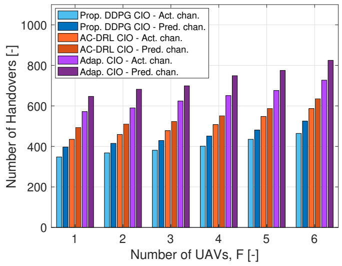

Fig. 5. Total number of handovers by UEs and UAVs, $c_{\text{req}} = 8$ Mbps.

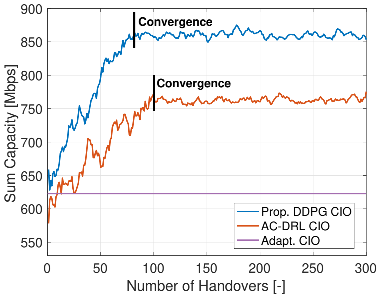

Fig. 6. Convergence of sum capacity as a function of the number of handovers performed in the network, $c_{\text{req}} = 8$ Mbps (using predicted channels).

compared to theoretical case with known channels, since prediction error incurs additional (wrong) handovers. While the number of handovers for AC-DRL CIO and Adaptive CIO is increased by 13.3% and 15.6%, respectively compared to the actual channel knowledge, the proposed DDPG CIO leads only up to 9.8% increase in number of handovers. Thus, the DDPG mitigates negative impact of the channel quality prediction error on the number of handovers by up to 26% and by 37% compared to AC-DRL and Adaptive CIO, respectively.

The convergence of the proposal, depicted as the sum capacity over the number of performed handovers required to train DDPG is shown in Fig. 6. The proposed DDPG CIO approach convergences about 20% faster compared to the AC-RL CIO. Besides, the sum capacity reached by the proposal is always notably higher than that of competitive state-of-the-art AC-DRL as well as Adaptive CIO algorithms.

VI. Conclusion

In this paper, we have addressed the challenges related to handover in networks with the transparent UAV relays employed due to cost and energy limitations. By proposing a novel framework for setting the CIO of all BSs, we have developed a DDPG-based solution maximizing the sum capacity of the UEs while reducing the number of handovers. We have also demonstrated, that potential accuracy in the channel quality is suppressed significantly by the proposed DDPG-based CIO setting compared to state-of-the-art works.

References

[1] M. Dai *et al.*, "Unmanned-Aerial-Vehicle-Assisted Wireless Networks: Advancements, Challenges, and Solutions," *IEEE Internet Things J.*, vol. 10, no. 5, pp. 4117-4147, March 2023.

[2] M. Najla, Z. Becvar, P. Mach and D. Gesbert, "Positioning and Association Rules for Transparent Flying Relay Stations," *IEEE Wireless Commun. Lett.*, vol. 10, no. 6, pp. 1276-1280, June 2021.

[3] M. M. U. Chowdhury *et al.*, "Mobility Management for Cellular-Connected UAVs: A Learning-Based Approach," *IEEE Int. Conf. on Commun. Workshops*, Dublin, Ireland, 2020, pp. 1-6.

[4] 3GPP, "Handover Procedures," 3GPP, TS 23.009 V16.0.0, Jul. 2020.

[5] A. Madelkhanova *et al.*, "Optimization of Cell Individual Offset for Handover of Flying Base Stations and Users," *IEEE Trans. on Wireless Commun.*, vol. 22, no. 5, pp. 3180-3193, May 2023.

[6] I. A. Meer *et al.*, "Mobility Management for Cellular-Connected UAVs: Model-Based Versus Learning-Based Approaches for Service Availability," *IEEE Trans. on Net. and Service Manag.*, vol. 21, no. 2, pp. 2125-2139, April 2024.

[7] M. Huang and J. Chen, "Proactive Mobility Load Balancing through Interior-point Policy Optimization for Open Radio Access Networks," *IEEE Trans. on Mobile Computing*, 2024.

[8] W. Dong *et al.*, "An Enhanced Handover Scheme for Cellular-Connected UAVs," *IEEE ICCC*, 2020.

[9] Y. Jang *et al.*, "UAVs Handover Decision using Deep Reinforcement Learning," *16th Int. Conf. on Ubiquitous Information Manag. and Commun.*, Seoul, Korea, Republic of, 2022, pp. 1-4.

[10] M. -S. Pan *et al.*, "An Enhanced Handover Scheme for Mobile Relays in LTE-A High-Speed Rail Networks," *IEEE Trans. Veh. Technol.*, vol. 64, no. 2, pp. 743-756, Feb. 2015.

[11] 3GPP, "Understanding on Type 1 and Type 2 Relay," TSG RAN WG1 Meeting # 57bis, R1-092370, Huawei, LA, CA, USA, June 2009.

[12] R. N. Braithwaite, "Improving Data Throughput for Cell-Edge Users in a LTE Network Using Up-Link HARQ Relays," *IEEE Veh. Tech. Conf.*, San Francisco, CA, USA, 2011, pp. 1-5.

[13] 3GPP, "Type 2 Relay Summary", RAN1 Meeting #60, R1-100951, ALU, ALU Shanghai Bell, CHTTL, San Francisco, USA, Feb. 2010.

[14] Z. Becvar, D. Gesbert, P. Mach and M. Najla, "Machine Learning-based Channel Quality Prediction in 6G Mobile Networks", *IEEE Commun. Mag.*, volume 61, no. 7, 2023.

[15] M. Najla *et al.*, "Predicting Device-to-Device Channels From Cellular Channel Measurements: A Learning Approach," *IEEE Trans. on Wireless Commun.*, vol. 19, no. 11, pp. 7124-7138, Nov. 2020.

[16] K. Ma *et al.*, "Reliability-Constrained Throughput Optimization of Industrial Wireless Sensor Networks With Energy Harvesting Relay," *IEEE Internet Things J.*, vol. 8, no. 17, pp. 13343-13354, Sept., 2021.

[17] 3GPP, "E-UTRA radio resource control (RRC) protocol specification (Release 8)," 3GPP, Tech. Rep. 36.331 V16.3.0, Jan. 2021.

[18] C. Bettstetter *et al.*, "The node distribution of the random waypoint mobility model for wireless ad hoc networks," *IEEE Trans. on Mobile Comput.*, vol. 2, no. 3, pp. 257-269, July-Sept. 2003.

[19] X. Hong *et al.*, "A group mobility model for ad hoc wireless networks", *MSWiM*, 1999.

[20] R. Amer, W. Saad and N. Marchetti, "Mobility in the Sky: Performance and Mobility Analysis for Cellular-Connected UAVs," *IEEE Trans. on Commun.*, vol. 68, no. 5, pp. 3229-3246, May 2020.

[21] S. -L. Su *et al.*, "A Novel Handover Process for Mobility Load Balancing in LTE Heterogeneous Networks," *IEEE ICPS*, 2019.