Optimization of Cell Individual Offset for Handover of Flying Base Station

Aida Madelkhanova and Zdenek Becvar
Dpt. of Telecommunication Engineering, Faculty of Electrical Engineering
Czech Technical University in Prague
Prague, Czech Republic
aida.madelkhanova@fel.cvut.cz, zdenek.becvar@fel.cvut.cz

2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring) | 978-1-7281-8964-2/20/$31.00 ©2021 IEEE | DOI: 10.1109/VTC2021-Spring51267.2021.9448970

***Abstract*—Flying base stations (FlyBSs) mounted on unmanned aerial vehicles (UAVs) are widely used in mobile networks to improve a coverage and/or quality of service for users. To ensure a seamless mobility of the FlyBSs among the static base stations (SBSs), an efficient handover mechanism is required. In this paper, we develop a novel handover mechanism determining the serving SBS for the FlyBS in order to increase the sum capacity of the users served by the FlyBS. We propose to dynamically optimize the handover by adjusting the cell individual offset of the SBS via Q-learning. The results show that the Q-learning converges promptly and the proposed approach increases the users capacity (by up to 18%) and their satisfaction with required minimum capacity (by up to 20%) comparing to state-of-the-art algorithms.**

***Index Terms*—Flying base station, handover, cell individual offset, reinforcement learning.**

I. Introduction

The FlyBSs, represented by the UAVs carrying a communication hardware, are seen as a suitable solution for the future mobile networks due to their flexibility in deployment and high mobility. The FlyBSs are deployed to the mobile networks to extend the network coverage or to boost service quality in a specific area [1]. As the FlyBSs can be deployed promptly, they are suitable for emergency situations or short-lasting events [2]. However, an integration of the FlyBSs to the mobile networks introduces new challenges.

In [3], the authors study the optimum altitude of the FlyBS by focusing on the reliability metrics in terms of power loss, outage probability, and bit error rate. An end-to-end capacity maximization via the optimal positioning of the FlyBS serving a single ground user is investigated in [4]. The outage probability of the half-duplex relaying in the network with the FlyBSs is minimized by optimizing the FlyBS trajectory and power allocation in [5]. In [6], the trajectory of the FlyBS and an allocation of the time for reception and forwarding of data are jointly optimized to maximize an efficiency in terms of an energy and a spectrum usage. In [7], the authors consider joint bandwidth and energy allocation for data gathering by the FlyBS to maximize the total data rate.

This work has been supported by Grant No. P102-18-27023S funded by Czech Science Foundation and by the grant of Czech Technical University in Prague No. SGS20/169/OHK3/3T/13..

Another challenge is related to ensuring a seamless mobility of the FlyBSs among the static base stations (SBSs). The FlyBS moves over time and their trajectory is arbitrary and hard to be predicted, as it is closely related to the movement of the user equipments (UEs) served by the FlyBS. The arbitrary trajectory together also with potentially high velocity of the FlyBS lead to rapid changes in the quality of channels between the FlyBS and the served UEs as well as between the FlyBS and the SBS that provides connectivity of the FlyBS to the network. These changes in the channel quality can result in performing handovers of the FlyBSs among the SBSs in an improper time and, consequently, to a handover failure and/or packet losses. Therefore, an efficient handover mechanism is required to provide a reliable communication of the UEs with the SBSs through the FlyBSs.

In conventional mobile networks, the handover of the UE between the SBSs is commonly initiated when a target SBS (i.e., the base station to which the handover should be performed) provides a channel of a higher quality than the current serving SBS. To avoid frequent handovers and/or handover failures, the decision on the handover initiation is tuned via control parameters, such as, a hysteresis, a time-to-trigger (TTT), or various offsets including cell individual offset (CIO) (please refer, for example, to [8] for more details). Thus, the handover usually takes place when the target SBS provides a channel quality that is higher than the channel quality of the serving SBS by the hysteresis and/or the offset(s) for a duration of the TTT.

In the conventional mobile networks (without FlyBSs), the optimization of the hysteresis, TTT, and the offsets is heavily addressed. For example, in [9], the authors propose an adaptation of the hysteresis according to relative qualities of the channels from the serving and neighboring SBSs. This approach reduces the number of handovers; however, it does not improve UEs' throughput. In [10] and [11], the authors propose to adapt the hysteresis via a fuzzy logic to minimize the number of performed handovers. However, the impact of the handover on the throughput of the UEs is not considered.

A reactive load balancing algorithm based on reinforcement learning is developed in [12]. The reactive algorithm is based on an adaptation CIO. The CIO is a cell-specific handover control parameter enabling to control cell association and to

978-1-7281-8964-2/21/$31.00 ©2021 IEEE

regulate cell coverage. The CIO is usually given positive or negative values; the CIO of the overloaded cell is reduced and that of the most eligible neighboring cell is increased. The authors suggest to adjust the CIO of serving and neighboring base stations by a specific step value so that the offset for the serving and neighboring base stations is of the same value, but opposite sign. The algorithm assumes that reinforcement learning states are defined based on a known distribution of the cell-edge users. However, such information is typically unknown to a network operator in practical scenarios. In [13], the authors present a deep reinforcement learning framework that adjusts the CIO of the base stations to balance the traffic among the base stations and to optimize the throughput of the network. This approach leads to an enhancement in the total throughput compared with the system with no CIO employment; however, the algorithm assumes only stationary UEs. In [14], a handover based on a reinforcement learning is proposed to maximize the received signal quality at the UEs represented by the UAVs while minimizing the number of handovers of these UEs. Despite the encouraging results, the framework adopted in [14] assumes the scenario with predefined and a priory known trajectory of the UAV. This assumption is, however, not valid in the scenario with the FlyBSs serving moving UEs as the trajectory of the FlyBSs is unknown. Handover management for the UAV acting as the UE via a dynamic adjustment of the SBSs' antenna tilt angles is outlined in [15]. The authors demonstrate that an intelligent antenna tilting reduces the handover rate for a simple mobility scenario with the UAV traveling along a linear trajectory.

None of the prior works [9] – [15] study the problem of the handover optimization for the FlyBSs serving mobile users, i.e., with unpredictable trajectory.

In this paper, we address the problem of the handover of the FlyBSs among the SBSs in a realistic and practical scenario with overloaded SBSs that are not able to serve all UEs. Thus, some UEs, which cannot be served by the overloaded SBSs, are served by the FlyBS. The FlyBS provides on-demand coverage in a relatively short-term peak traffic periods to the UEs that cannot be served by the SBS. The FlyBS follows the UEs movement and, hence, performs handover among the SBSs in order to provide a sufficient quality of the communication to the served UEs. We propose a framework for a setting of CIO for the individual SBSs to improve the handover decisions for the FlyBSs and, consequently, to improve the communication quality of the UEs not served by the SBSs. Mobility of FlyBS with unpredictable flying trajectory based on the UEs' movement has not been considered in most recent research contributions, which mainly focus on mobility of FlyBS with prior known trajectory or mobility of UEs while FlyBS remains static. Due to the dynamic properties of the network environment, such a complex problem requires the solution which is adaptive to the changes in the environment. Thus we suggest to solve the problem via reinforcement learning. Particularly, we employ the Q-learning to learn a proper setting of the CIO values for individual SBS according to the load of these SBSs. The CIO adjustment decisions are dynamically optimized using Q-learning to provide an efficient mobility support in the sky.The proposed solution leads to an increase in the UEs' capacity and to an increase in the ratio of the UEs satisfied with their experienced communication capacity. To our knowledge, it is a first attempt to optimize handover of the moving FlyBS by adjusting handover offsets and using reinforcement learning-based approach.

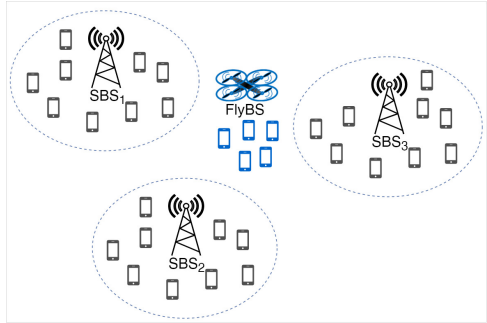

Fig. 1. System model with multiple SBSs serving UEs and one FlyBS serving the UEs that cannot be served by the SBSs, as the SBSs are overloaded.

The rest of this paper is organized as follows. Section II presents the system model and defines the problem addressed in this paper. Then, in Section III, we present our proposed Q-learning-based determination of the offsets of the SBSs for optimization of the FlyBS's handover via Q-learning. The simulation results and their discussion are provided in Section IV. Section V concludes the paper.

II. System model

In this section, we first outline the model of the system considered in this paper. Then, we formulate the targeted problem.

A. System model

We consider a cellular mobile network consisting of a set of M SBSs $\boldsymbol{M}=\{m_1, m_2, ..., m_M\}$, a set of N UEs $\boldsymbol{N}=\{n_1, n_2, ..., n_N\}$ and one FlyBS, as shown in Figure 1. All UEs in the system require a certain communication capacity c_{req}. For a clarity of the following explanations, we assume the same c_{req} for all UEs. However, our proposed solution is suitable for any, even diverse c_{req} for individual UEs. Out of N UEs, N_u UEs cannot receive c_{req} from the SBSs (e.g., due to a high load of the SBSs). These N_u UEs are denoted as uncovered UEs and these are connected to the network via the FlyBS, which relays the communication from the adjacent SBS and the uncovered UEs.

The position of the n-th UE changes over time and is defined by Cartesian coordinates $\{x_n(t), y_n(t), z_n\}$. The FlyBS follows the uncovered UEs and the coordinates of the FlyBS at time t are denoted as $\{x_f(t), y_f(t), z_f\}$. Like in [16], [17], the position of the FlyBS is defined as a center of gravity of all UEs associated to the FlyBS. Note that the proposed solution does not depend on the position of the FlyBS and can be applied on any other approach for the determination of

the FlyBS position. The location of the SBS does not change over time and is denoted by $\{x_m, y_m, z_m\}$.

In our model, we consider the downlink communication from the SBS to the UEs either directly or via the FlyBS. The signal to interference plus noise ratio (SINR) at the n-th UE served by the FlyBS is:

$$\gamma_{f,n} = \frac{P_f h_{f,n}}{\sum_{m \in \boldsymbol{M}} P_m h_{m,n} + \sigma^2} \tag{1}$$

where P_f is the transmission power of the FlyBS, P_m is the transmission power of the SBS, $h_{f,n}$ stands for the channel gain between the FlyBS and the n-th UE, the term $\sum_{m \in \boldsymbol{M}} P_f h_{m,n}$ represents the co-channel interference from other SBSs, $h_{m,n}$ stands for the channel gain between the m-th SBS and the n-th UE, and σ^2 is the power of additive white Gaussian noise (AWGN) at the receiver.

Similarly, the SINR at the FlyBS receiving data from the m-th serving SBS is expressed as:

$$\gamma_{m,f} = \frac{P_m h_{m,f}}{\sum_{l \in \boldsymbol{M}, l \neq m} P_m h_{l,f} + \sigma^2} \tag{2}$$

where $h_{m,f}$ is the channel gain between the m-th serving SBS and the FlyBS, $\sum_{l \in \boldsymbol{M}, l \neq m} P_m h_{l,f}$ represents the interference from other SBSs, and $h_{l,f}$ is the channel gain between the l-th interfering SBS and the FlyBS.

We adopt a decode-and-forward (DF) relaying for the communication of the UEs via the FlyBS. The relay channel capacity of the DF system for the n-th user is defined as $C_n = \frac{B_n}{2} min\{log_2(1+\gamma_{m,f}), log_2(1+\gamma_{f,n})\}$ [18], where B_n denotes the bandwidth of the n-th UE's channel.

The m-th SBS serves a set of the UEs that generate a certain load ρ_m to this SBS. The load is defined as the ratio of the utilized bandwidth of the m-th SBS to serve the associated UEs versus the total amount of bandwidth available for the given m-th SBS, i.e.,:

$$\rho_m = \frac{\sum_{n \in \boldsymbol{N}} \beta_{m,n} B_n}{B_m} \tag{3}$$

where the binary parameter $\beta_{m,n} \in \{0, 1\}$ indicates if the n-th UE is associated to the m-th SBS (for $\beta_{m,n}$ = 1), or not (for $\beta_{m,n}$ = 0), B_m is the total bandwidth available for the m-th SBS.

As the FlyBS follows the moving users, handovers of the FlyBS among the SBSs are performed during the flight. Throughout the flight, the FlyBS measures the channel quality from all neighboring SBSs and periodically sends the measurement reports to the serving SBS in a similar way as the common UEs report their channel quality in the common mobile network. Based on the measurement results, the serving SBS decides to transfer the FlyBS to one of the neighboring SBSs if a higher signal quality can be reached. We consider the handover mechanism based on the A3 event [8] that involves the hysteresis, TTT, and CIO with the channel quality represented by the received signal strength (RSS) between the m-th SBS and the FlyBS expressed as $RSS_{m,f} = P_m h_{m,f}$. The handover from the serving SBS of the FlyBS to the neighboring SBS is triggered according to the A3 event if:

$$RSS_{j,f} + CIO_j - Hys > RSS_{m,f} + CIO_m \tag{4}$$

where $RSS_{j,f}$ denotes the RSS between the neighboring SBS and the FlyBS. CIO_j and CIO_m correspond to the CIOs of the neighboring and serving SBSs, respectively. Hys is the hysteresis value parameter in dB. The A3 event is triggered if and only if the condition (4) holds for a period of time that exceeds the TTT.

B. Problem formulation

In this paper, we focus on the handover of the moving FlyBS among the SBSs. At each decision time t, we determine which SBS should serve the FlyBS based on the available network resources. The objective is to optimize the decision on the handover by adjusting CIO of the SBSs $\boldsymbol{CIO}^*$ so that the total capacity of the UEs served by the FlyBS is maximized. Thus, our objective is defined as:

$$\boldsymbol{CIO}^* = \underset{\boldsymbol{CIO}^* \in O}{\operatorname{argmax}} \sum_{n \in \boldsymbol{N}_u} C_n \tag{5}$$

$$\text{subject to} \sum_{m \in \boldsymbol{M}} \beta_{m,n} = 1, \ \forall n \in \boldsymbol{N}, \tag{5a}$$

$$\sum_{m \in \boldsymbol{M}} \beta_{m,f} = 1, \tag{5b}$$

$$\sum_{n \in \boldsymbol{N}} \beta_{m,n} B_n \leqslant B_m, \ \forall m \in \boldsymbol{M}. \tag{5c}$$

where O=$\langle CIO_{min}, CIO_{max} \rangle$ defines the bound over the CIO values and CIO_{min} and CIO_{max} are the minimum and maximum possible CIO values in the system, respectively. The binary parameter $\beta_{m,f} \in \{0, 1\}$ indicates if the FlyBS is associated to the m-th SBS (for $\beta_{m,f}$ = 1), or not (for $\beta_{m,f}$ = 0). The constraint (5a) ensures that each UE is associated to just one SBS and the constraint (5b) ensures that the FlyBS is associated to just one SBS. Furthermore, the constraint (5c) guarantees that the SBSs do not allocate more bandwidth than available.

III. Proposed CIO adjustment based on Q-learning

An efficient solution to the optimization problem in (5) requires a prior knowledge of the network environment, e.g., the communication channel capacity due to unpredictable flying trajectory of FlyBS (dependent on the UEs' movement). However, such knowledge is not always available, and a significant amount of information should be exchanged to obtain at least a part of the required knowledge. Thus, a model-free deployment approach is required to solve this problem. We suggest to solve the problem via the reinforcement learning. The reinforcement learning-based method learns directly from observed experiences without a model, where a model represents the network environment's dynamics. In this section, we address the handover optimization problem with the reinforcement learning and we introduce a novel CIO adjustment scheme. We propose the Q-learning-based algorithm to obtain the optimal

CIO adjustment policy for the serving as well as target SBSs. In our proposal, each SBS considers only its individual traffic load for the CIO adjustment, and no information from the neighboring SBSs is required in order to avoid an additional overhead.

A. Preliminaries on Reinforcement Learning

The reinforcement learning is formulated as the triple (S, A, r), where S and A denote the sets of all possible states and actions, respectively, and r is the reward function [19]. The reinforcement learning agent aims at selecting a sequence of the optimal actions under different system states. Assuming π is a policy of choosing the actions, the action-value function $Q^{\pi}(s, a)$ for every state-action pair indicates how good is the action performed in that state. We use a well-known model-free reinforcement learning algorithm known as Q-learning. Thus, the action-value function $Q(s, a)$ is iteratively updated and preserved in a lookup table of the Q-values corresponding to each state-action pair. If the agent performs the action in the state s_t at the time t, it receives an immediate reward r_t and the system transits to the state s_{t+1}. The Q-values are updated based on the interaction with the environment according to:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[\, r_t + \lambda max Q(s_{t+1}, a) - Q(s_t, a_t)]\,, \tag{6}$$

where $\alpha \in (0, 1]$ is the learning rate that balances new information against previous knowledge, $\lambda \in (0, 1)$ is the discount factor that balances between the immediate and future rewards, and r_t is the reward received when moving from the state s_t to the state s_{t+1}.

Q-learning training process consists of multiple iterations, each with a number of steps μ. We adopt the ϵ-greedy policy, where the agent tries to obtain the highest reward at each training step, while also checks for other actions, which can improve the estimated future reward. The primary concept of ϵ-greedy policy is to pick a random number from [0; 1] and compare it with ϵ. If the chosen number is smaller than ϵ, the agent explores the state-action space by taking a random action; otherwise, the agent exploits the learned information and picks an optimal action with the highest Q-value. The learning starts with ϵ = 1, then, ϵ is reduced to ϵ = 0 by multiplying a decay factor η = 0.99 at each learning step [15] to end up with the optimal policy.

B. Q-learning for CIO adjustment

The objective defined in (5) is interpreted as the problem, where the agent maximizes its final cumulative rewards by interacting with an unknown environment over time. For our problem, each SBS can be considered as an agent; however, this can lead to an additional overhead due to information exchange directly among all SBSs. The FlyBS can also act as the agent for our purposes. Nevertheless, the FlyBS is usually constrained with a limited energy and any additional energy consumption is not welcome. Thus, we assume that the network is equipped with a central agent, e.g., in an edge server or a (software) entity in the operator's core network, that can monitor the load of the SBSs and implement Q-learning. Note that the proposed solution does not depend on the agent representation.

Algorithm 1 Q-learning for FlyBS handover optimization

1: **Input:** number of SBSs, SBS load, action set (possible CIO values)
2: Initialize $Q(s, a)$ and $S(1)$
3: **for** each learning step t **do**
4: observe current state $S(t)$ of the SBSs
5: choose action using ϵ-greedy policy
6: execute action $A(t)$ and update CIO values for the SBSs
7: calculate $r(t)$ using (7)
8: calculate the action, which maximizes Q-value in the next state and update Q-table using (6)
9: $S(t) \leftarrow S(t+1)$
10: **end for**
11: **Output:** Q-table

Now, let's define the state and action sets, and the reward function for our targeted problem. The detailed learning context is represented as follows. To formalize the Q-learning problem, the state of the network is represented by the load of the SBSs. Thus, the set of states $S(t)$ is defined as $S(t) = [\,\rho_1(t), \rho_2(t), \ldots, \rho_m(t)]$, where $\rho_m(t) \in [0, 1]$ and corresponds to the load of the m-th SBS at the time t. The action is understood as a selection of the CIO for each SBS. Thus, the central agent determines the CIOs for the SBSs via a selection of suitable actions $A(t) = [CIO_1(t), CIO_2(t), \ldots, CIO_m(t)]$, where the $CIO_m(t)$ corresponds to the CIO of the m-th SBS at the time t. Considering the problem formulation targeting the maximization of the total capacity of the UEs served by the FlyBS (see (5)), the reward function at the time t, $r(t) \in \langle 0, 1\rangle$, is defined:

$$r(t) = \frac{\sum_{n \in \boldsymbol{N}_u} C_n(t)}{N_u c_{req}}, \tag{7}$$

where $C_n(t)$ is the channel capacity of the n-th UE served by the FlyBS at the time t. The pseudo-code of the proposed Q-learning process is presented in Algorithm 1. In step 2, the Q-table is initiated with random values from interval (0; 1). The Q-value iterations for each training step are performed in steps 2-9. Based on the current state $S(t)$, ϵ-greedy policy is performed in step 5 to choose either the random or optimal action. In step 6, the chosen action is executed. The reward is calculated in step 7. Finally, in step 8, values for selecting different actions are stored in the Q-table, where the highest value represents the optimal choice.

IV. Performance Evaluation

In this section, we evaluate the performance of our proposed Q-learning-based handover optimization via simulations in MATLAB. Note that the proposed solution does not depend on the transmission power of base stations and the size of cells. Thus, we consider the mobile network containing three

TABLE I
SIMULATION PARAMETERS

Parameter	Value
Simulation area	1000m × 1000m
Carrier frequency	2 GHz
Tx power of SBS/FlyBS	23/23 dBm
Bandwidth of SBS	100 MHz
SBS/FlyBS/UE height	20/80/1.5 m
Number of UEs	125
Number of UEs served by FlyBS	25
Hysteresis margin	3 dB
Time step	1 s

SBSs, with transmission power of 23 dBm, and 125 UEs deployed in a square area of 1000 m × 1000 m. The SBSs are placed randomly with a minimum inter-site distance of 500 m. The UEs are also deployed randomly following a uniform distribution. The UEs served by FlyBS move in a crowd along the same direction (following the same crowd movement trajectory), but each UE can move arbitrary along the crowd trajectory.

The path-loss between the FlyBS and the UEs is modeled as the air-to-ground (A2G) communication according to [20], with suburban environment parameters ("suburban" channel model, i.e., a = 4.88, b = 0.43, ηLoS = 0.1 and ηNLoS = 21, see [21] for more details). A signal propagation for the SBSs is modeled according to [22] with the path loss model 128.1+37.6$log_{10}d$, where d (in km) is a distance between the UE and the SBS. The spectral density of noise is set to -174 dBm/Hz.

We consider 40 random deployments (realizations) and a duration of each is 200.000 seconds of real-time. Note that the positions of the UEs, corresponding trajectory of the FlyBS, and the positions of the SBSs change at each realization. The results of these realizations are then averaged out. Like in [12], the CIO of each SBS is selected from the range of -6 to 6 dB. Table I summarizes the major parameters used in our simulations.

For the Q-learning training purpose, different settings of α and λ have been tested and we have observed that α = 0.8 and λ = 0.6 are the most suitable for the proposed algorithm.

The performance of the proposed Q-learning algorithm is compared with three commonly exploited baseline approaches: i) handover without CIO, i.e., with CIO set to 0 dB for all SBSs (denoted as *No CIO* in figures); ii) the algorithm in [23], denoted as *Fixed step CIO*, which adjusts the CIO by adding or subtracting a fixed step when the difference in the load between neighboring base stations exceeds a threshold; and iii) the CIO adjustment algorithm from [24], denoted as *Adaptive CIO*, which sets CIO according to the predefined relation between the value of CIO and the average load of the SBSs. We also show performance for the case without FlyBS integration (denoted as *No FlyBS* in figures), i.e., all UEs are served only by SBSs.

We consider two performance indicators for the evaluation: i) the sum capacity of the UEs served by the FlyBS defined as $\sum_{n \in \boldsymbol{N}_u} C_n(t)$; ii) the UEs' satisfaction ratio, i.e., the ratio of the UEs for which $C_n \geq c_{req}, n \in \boldsymbol{N}_u$.

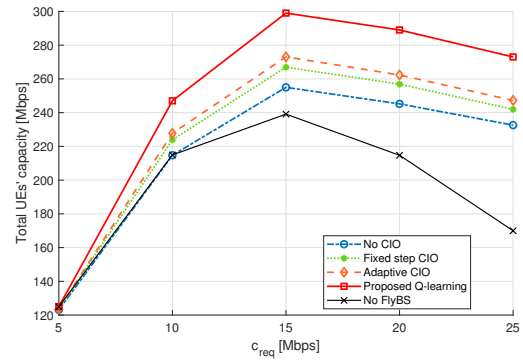

Fig. 2. Total capacity of UEs served by FlyBS.

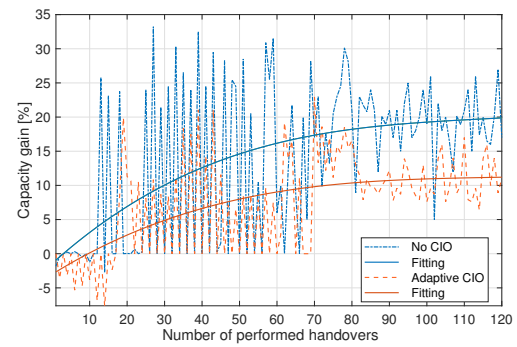

Fig. 3. Capacity gain UEs' served by FlyBS vs number of handovers performed by FlyBS for c_{req}=25 Mbps.

Figure 2 shows the total capacity of the UEs served by the FlyBS for various c_{req}. The capacity is raising with c_{req} up to c_{req} = 15 Mbps. Then, for c_{req} higher than 15 Mbps the total capacity starts decreasing. The decrease in the total capacity of the UEs with increasing of c_{req} is because the overloaded SBSs cannot provide all UEs with the resources required to meet c_{req}. The proposed algorithm increases the capacity by up to 18%, 11%, and 10% comparing to the No CIO, the Fixed step CIO and the Adaptive CIO algorithms, respectively. Note that the increase in the capacity by the proposed algorithm is more notable for the cases, when the system suffers from resource shortage. This improvement is because the proposed CIO adjustment algorithm considers the SBS' load and learns the most suitable CIO values for each state of the SBSs' loads. The significant decrease in the total capacity of UEs for the No FlyBS case is because, not all UEs can be served by SBSs.

Figure 3 depicts the gain achieved by the proposal in the total capacity of UEs served by FlyBS with respect to the No CIO and to the Adaptive CIO algorithms. The figure illustrates the learning progress of the proposal after individual learning events, i.e., after each handover performed by the FlyBS. At the beginning of the learning (up to roughly 17 handovers) the gain becomes negligible or even slightly negative in some steps compared to the No CIO (up to -1%) and to the Adaptive CIO

(up to -7%). This slightly negative gain is a result of the initial "random" learning when (almost) no information that would guide the selection of the CIO is available. However, after this short initial learning phase, i.e., after about first 17 handovers, the gain becomes always non-negative and increases with new performed handovers. The capacity gain achieved by the proposed Q-learning based algorithm converges approximately after 60 handovers. Moreover, after about 70 handovers, the gain stabilizes and becomes notably positive ranging typically from about 10% to 30% and from about 8% to 20% with respect to the No CIO and Adaptive CIO algorithms, respectively. The figure also illustrates fitting function for the gain with respect to the No CIO and Adaptive CIO algorithms. The fitting function demonstrates that the proposed algorithm outperforms the No CIO and the Adaptive CIO in the sum capacity of UEs by 19% and 11%, respectively, after 90 handovers performed by the FlyBSs. Requiring only tens of handovers to learn the suitable values of CIO is sufficiently fast to deploy the proposed algorithm in real networks.

In Figure 4, we show the satisfaction of the UEs served by FlyBS with the received capacity, i.e., the ratio of the UEs that receive at least c_{req}. For all compared algorithms, the UEs' satisfaction level is decreasing with an increasing c_{req}. The reason for the decrease in the satisfaction is the fact that the total capacity required by all UEs in the network increases with c_{req}, while the amount of bandwidth available to the SBSs and the FlyBS remains the same. Due to the resources shortage, the overloaded SBSs are not able to assigning the required bandwidth to all UEs and less UEs achieves c_{req}. For very low requires capacity (c_{req} = 5 Mbps), all UEs are satisfied disregarding the CIO setting, as there are enough resources in the system to satisfy all UEs. However, as c_{req} increases, the satisfaction starts decreasing. The decrease in the satisfaction is more notable for all competitive algorithms than for the proposed algorithm. The proposed algorithm leads to a an improvement in the satisfaction of about 11%, 8%, and 7% percent-points with respect to the No CIO, Fixed step CIO, and Adaptive CIO algorithms, respectively. This corresponds to an increase in the UEs' satisfaction ratio by 20%, 16%, and 14% comparing to the satisfaction ratio achieved by the No CIO, Fixed step CIO, and Adaptive CIO algorithms, respectively. The improvement in the UEs' satisfaction is achieved by the adjustment of CIO dynamically according to the load of SBSs so that the FlyBS is associated to the SBS that offers required communication capacity to the FlyBS for a longer time.

V. Conclusion

In this paper, we have proposed a novel algorithm based on Q-learning managing handover of the FlyBS among the SBSs to maximize the total capacity of the UEs served by the FlyBS. The proposed algorithm adjusts CIO values according to the load of SBSs. The states of the Q-learning agent are described in terms of the load of the SBSs and the reward function is defined in terms of the capacity of UEs served by the FlyBS. The results show an enhancement in the UEs' capacity by up to 18% and by 20% in the level of the UEs' satisfaction with respect to state-of-the-art solutions. We also demonstrate that the Q-learning process converges quickly and only tens of handovers are required to reach a significant gain.

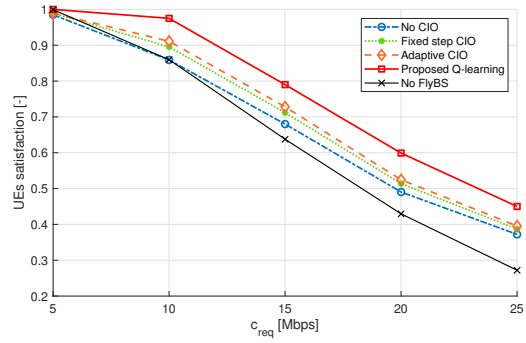

Fig. 4. Satisfaction of UEs served by FlyBS vs c_{req}.

Future extensions should target extension towards multiple FlyBSs. In addition, the work should be also extended towards joint handover decision and the FlyBS positioning.

References

[1] Y. Zeng, R. Zhang and T. J. Lim, "Wireless communications with unmanned aerial vehicles: opportunities and challenges," in *IEEE Communications Magazine*, vol. 54, no. 5, pp. 36-42, May 2016.

[2] B. Li, Z. Fei and Y. Zhang, "UAV communications for 5G and beyond: Recent advances and future trends," in *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2241-2263, April 2019.

[3] J. Chen, U. Mitra and D. Gesbert, "Optimal UAV relay placement for single user capacity maximization over terrain with obstacles," *2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Cannes, France, 2019.

[4] Y. Chen, W. Feng and G. Zheng, "Optimum placement of UAV as relays," *IEEE Communications Letters*, vol. 22, no. 2, pp. 248-251, Feb. 2018.

[5] S. Zeng, H. Zhang, K. Bian and L. Song, "UAV relaying: Power allocation and trajectory optimization using decode-and-forward protocol," *IEEE International Conference on Communications (ICC 2018) workshops*, Kansas City, MO, 2018.

[6] J. Zhang, Y. Zeng and R. Zhang, "Spectrum and energy efficiency maximization in UAV-enabled mobile relaying," *2017 IEEE International Conference on Communications (ICC)*, Paris, 2017, pp. 1-6.

[7] H. Zanjie, N. Hiroki, K. Nei, O. Fumie, M. Ryu and Z. Baohua, "Resource allocation for data gathering in UAV-aided wireless sensor networks," *2014 4th IEEE International Conference on Network Infrastructure and Digital Content*, Beijing, 2014, pp. 11-16.

[8] 3GPP, "E-UTRA radio resource control (RRC) protocol specification (Release 8)," 3GPP, Tech. Rep. 36.331 V8.16.0, Dec. 2011.

[9] Z. Becvar, P. Mach, "Adaptive hysteresis margin for handover in femtocell networks", *International Conference on Wireless and Mobile Communications (ICWMC 2010)*, pp. 256-261, 2010.

[10] K. da Costa Silva, Z. Becvar, E. Cardoso, C. R. Francês, "Self-tuning handover algorithm based on fuzzy logic in mobile networks with dense small cells", *IEEE Wireless Communications and Networking Conference (IEEE WCNC 2018)*, pp. 1-6, 2018.

[11] K. da Costa Silva, Z. Becvar, C. R. Francês, "Adaptive Hysteresis Margin Based on Fuzzy Logic for Handover in Mobile Networks with Dense Small Cells", IEEE Access, vol. 6, pp. 17178-17189, 2018.

[12] S. S. Mwanje and A. Mitschele-Thiel, "A Q-Learning strategy for LTE mobility Load Balancing," *IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, London, 2013.

[13] K. Attiah et al., "Load balancing in cellular networks: A reinforcement learning approach," *IEEE Consumer Communications & Networking Conference (CCNC)*, Las Vegas, NV, USA, 2020.

[14] Y. Chen, X. Lin, T. Khan and M. Mozaffari, "Efficient drone mobility support using reinforcement learning," *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, Seoul, Korea (South), 2020, pp. 1-6.

[15] M. M. U. Chowdhury, W. Saad and I. Güvenç, "Mobility Management for Cellular-Connected UAVs: A Learning-Based Approach," *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*, Dublin, Ireland, 2020, pp. 1-6.

[16] O. Esrafilian, R. Gangula and D. Gesbert, "UAV-relay placement with unknown user locations and channel parameters," *2018 52nd Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, 2018, pp. 1075-1079, doi: 10.1109/ACSSC.2018.8645508.

[17] Z. Becvar, M. Vondra, P. Mach, J. Plachy and D. Gesbert, "Performance of mobile networks with UAVs: Can flying base stations substitute ultra-dense small cells?," *European Wireless 2017; 23th European Wireless Conference*, Dresden, Germany, 2017, pp. 1-7.

[18] J. N. Laneman, D. N. C. Tse and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062-3080, Dec. 2004.

[19] R. Sutton "Reinforcement Learning: An Introduction" MIT Press, 1998.

[20] A. Al-Hourani, S. Kandeepan and S. Lardner, "Optimal LAP Altitude for Maximum Coverage," *IEEE Wireless Communications Letters*, vol. 3, no. 6, pp. 569-572, Dec. 2014.

[21] R. I. Bor-Yaliniz, A. El-Keyi and H. Yanikomeroglu, "Efficient 3-D placement of an aerial base station in next generation cellular networks," *2016 IEEE International Conference on Communications (ICC)*, Kuala Lumpur, 2016, pp. 1-5.

[22] 3GPP TR 36.814, "Evolved Universal Terrestrial Radio Access (E-UTRA); Further Advancements for E-UTRA Physical Layer Aspects (Release9)," 2010.

[23] R. Kwan, R. Arnott, R. Paterson, R. Trivisonno and M. Kubota, "On mobility load balancing for LTE systems," *2010 IEEE 72nd Vehicular Technology Conference - Fall*, Ottawa, ON, 2010, pp. 1-5.

[24] S. Su, T. Chih and S. Wu, "A novel handover process for mobility load balancing in LTE heterogeneous networks," *2019 IEEE International Conference on Industrial Cyber Physical Systems (ICPS)*, Taipei, Taiwan, 2019, pp. 41-46.