Self-Optimizing Data Offloading in Mobile Heterogeneous Radio-Optical Networks: A Deep Reinforcement Learning Approach

Sihua Shao, Mahmoud Nazzal, Abdallah Khreishah, and Moussa Ayyash

Abstract

In addition to the exploration of more spectrum at high-frequency bands, next-generation wireless networks will witness an intelligent convergence of radio frequency (RF) and non-RF links such as optical and visible light communication. Optical attocell (OAC) networks provide an additional layer to RF-based wireless networks with gigabit-per-second data transmission rate and centimeter-level location accuracy. However, the directionality, line-of-sight constraints, as well as strong sensitivity to the location and orientation of user terminals challenge the stringent requirements for throughput and latency. In this article, we consider mobile heterogeneous networks (HetNets) incorporating indoor OAC with femtocells and macrocells to provide a low-cost and energy-efficient solution. The HetNets solution satisfies diverse service requirements in terms of user-experienced data rate, mobility, latency, accuracy, and security in the Internet of Things. To support seamless connectivity and optimal resource allocation in the proposed HetNets with mobility awareness, handover in dynamic environments needs to be addressed efficiently. Incorporating rich environmental parameters into such a decision making problem facilitates the self-optimization process, but extensively expands the state space. To achieve a fast convergence speed, a deep reinforcement learning approach is proposed to optimize the handover parameters (e.g., time-to-trigger and hysteresis margin). This is a model-free and off-policy reinforcement setting that trains and employs a deep neural network to predict future rewards for successions of states and actions. Thus, the optimal parameters are obtained by selecting the best actions to take. Through numerical simulation and performance analysis, we discover the gain from enriching the state space and the adaptability of the system to dynamic environments.

Introduction

The majority of mobile traffic is usually indoors, especially in urban deployments, which is difficult to serve from outdoor base stations and is more of a challenge due to the use of ultra-high-frequency bands [1]. Ericsson recently reported that in a dense urban high-rise area, 37 percent of macro traffic was served to indoor mobile users during busy hours, indicating that in-building cell deployment could be increased to meet indoor mobile traffic demand. In exploring unused spectrum, optical attocell (OAC) is considered as a competitive non-radio frequency (RF) candidate for indoor wireless access [2] due to numerous advantages such as its dual-use nature, high energy efficiency, ubiquitous availability, and no interference to RF devices. Heterogeneous networks' (HetNets') [3] integration of non-RF OAC networks within RF-based femtocell and macrocell networks provides a low-cost and flexible solution to satisfy the specifications of user experienced data rate, mobility, and latency in the next-generation wireless network standards.

Vertical handover (VHO), referred to as automatic fallover from one technology to another in order to maintain communication, is different for the cellular-WiFi pair compared to that for the cellular-OAC pair. This is mainly caused by the directionality feature of OAC. Traditionally, two major handover parameters, time-to-trigger (TTT) and hysteresis margin (HM), are defined to decide whether and when the user device switches the connection from one cell to another. HM sets the threshold for the handover to start being considered and is used to avoid the ping-pong effect. TTT determines the observation duration for measurement after HM is met. Only if a certain event condition is satisfied for longer than the TTT is the handover triggered. The received signal strength (RSS) from an RF cell, either cellular or WiFi, has *a stable mean but a large deviation* for an indoor mobile user [4]. Accordingly, HM and TTT are designed to mitigate the false positive rate resulting from the deviation. In contrast, the RSS from an OAC has *a drastically changing mean but a stable deviation* during the walk in-and-out. Such unique features of OAC motivate a new holistic strategy to control the handover parameters.

In this article, we study a self-optimizing mobile HetNet, as shown in Fig. 1, and focus on the handover decision problem. The mobile user terminals (UTs) are located within the wireless coverage of an RF macrocell (e.g., cellular base station), an RF femtocell (e.g., WiFi access point), and optical attocells. The RF macrocell offers the widest coverage, but is constrained by the scarce licensed frequency bands. The RF femtocell offloads part of the data traffic from the RF macrocell; however, this

Digital Object Identifier: 10.1109/MNET.0072100606

Sihua Shao is with New Mexico Tech, USA; Mahmoud Nazzal and Abdallah Khreishah are with New Jersey Institute of Technology, USA; Moussa Ayyash is with Chicago State University, USA.

0890-8044/22/$25.00 © 2022 IEEE

compromises the quality of experience (QoE) significantly when multiple UTs attempt to share the same spectrum resource. The OAC provides high-speed wireless connectivity to local users, while the signal quality highly depends on the location and orientation of the UTs. The parameters (e.g, TTT and HM) controlling the handover among different cells are updated regularly in a cloud server. The cloud server coordinates with the RF macrocell base station to periodically monitor the signal-to-interference-plus-noise ratio (SINR) of each mobile user and utilize the SINR feedback to optimize the handover parameters in a user-centric manner. To optimize the data offloading and resource allocation, the utilization of RF cell and OAC, the behavior and mobility pattern of user terminals, OAC deployment, and interference, the user preferences should be jointly analyzed, which requires massive iterations to reach a global optimum using traditional methods [5]. Moreover, the OACs are expected to satisfy the illumination requirements, the dynamics and uncertainty of which further complicate the problem.

Deep learning utilizes a neural network to handle complex and high-dimensional raw input data, and thereby can efficiently address the dynamics in the considered HetNets and avoid time-aggressive iterations of traditional mathematical methods. Reinforcement learning typically models the problem as a Markov decision process (MDP), where an agent at every time step is in a state, takes an action, receives a reward, and transitions to the next state according to environmental dynamics. The periodic updating process of handover parameters can be modeled as an MDP. Therefore, in this article, we utilize deep reinforcement learning (DRL) to optimize the handover decision making problem in a time-step-based manner. In particular, the online DRL is performed by directly measuring the QoE metrics (data rate, user speed, etc.) in each time step. In addition, offline DRL based on a fixed simulating dataset generated from the SINR feedback will be leveraged to speed up the training process of a specific environment. For both online and offline DRL, the neural network is adopted to predict a function value that estimates the future returns of taking action a from state s. Through the proposed DRL approach, we further discuss the existing challenges and future directions for self-optimizing mobile HetNets. Thus, the contributions of this article can be summarized as follows:

- We discuss the limitations and developments of the emerging technologies that enable the self-optimization of mobile HetNets incorporating RF and non-RF links. Specifically, an overview of key technologies including mobile-assisted handover, user profiling, and machine learning is presented.
- We analyze the handover decision making problem for the considered mobile HetNets. Then we propose a DRL-based data offloading approach, which can improve the QoE considering the mobility dynamics and different communication conditions.
- We verify through extensive simulations the fast convergence rate and the adaptability to different environments of the proposed self-optimizing system.

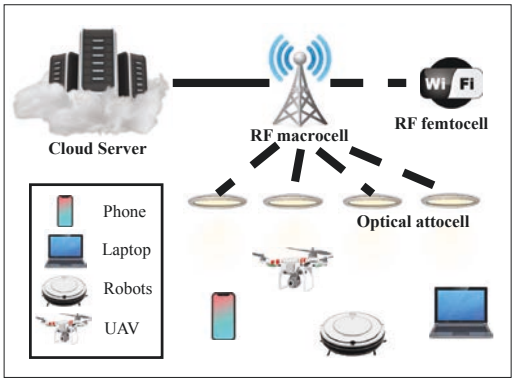

FIGURE 1. The considered self-optimizing mobile heterogeneous network.

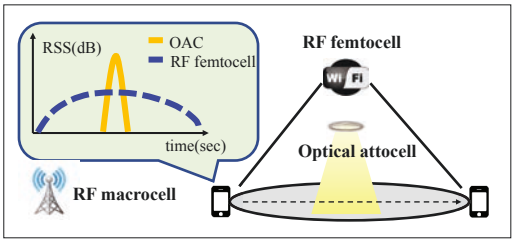

FIGURE 2. Difference in received signal strength when crossing a femtocell and an optical attocell.

Limitations of Current VHO Approaches

Cellular-WiFi Handover Is Different from Cellular-OAC Handover

The RSS from an RF cell does not increase or decrease sharply when the user crosses the cell. This is highly in contrast to the case in OAC. As shown in Fig. 2, the mobile user is walking in and out of an RF femtocell with an OAC located in the center. The RSS from the RF cell is maintained above a usable level for a relatively long period, while the non-zero RSS from the OAC may only last for several seconds. The difference indicates that if the cellular-WiFi handover strategy is directly applied to the cellular-OAC scenario, the UT will make frequent handover attempts to OAC; however, most of the attempts will end up failing since the mobile user walks out of the OAC before the handover is completed. The potential problems require us to rethink how to control the handover parameters such that the mobile user can take advantage of the ubiquitously deployed OACs and in the meantime keep the QoE above a satisfactory level. The diversity of QoE performance under different user mobility is validated through an extensive case study in [3].

Quasi-Static Network Selection

One of the state-of-the-art VHO approaches is the quasi-static network model-based access point (AP)-user association [6], where the channel characteristics are assumed to be fixed in each coherent and equal-length time slot. According to the channel quality feedback at time t, a centralized coordinator (CC) computes the best strategy based on a proposed algorithm. However, the AP-user association strategy starts having an effect at time $t + \tau$. The assumption working behind the quasi-static network model is that the UTs barely change their

Although aggressively offloading data traffic from RF cells to OACs mitigates the congestion in RF-based wireless networks, the consequent user-experienced data rate and latency may be unacceptable even for a UT crossing the small-coverage OACs at a regular walking speed.

location or orientation within time τ, which keeps the timeliness of the association strategy for time t. According to the experimental results in [4, Fig. 4a], if we measure the signal-to-noise ratio (SNR) variation in a walking in-and-out scenario of an OAC, at the steepest region, it takes only 10 ms for the SNR to change by 10 dB. As a result, it is highly possible that the strategy for time t will be outdated at time $t + \tau$ in the mobile scenarios. Such outdated handover decisions will lead to either severe data rate dropping or intermittent disconnection.

Immediate Handover from RF Cells to OAC

Another very recent VHO approach is to control the dwell time (i.e., the amount of time by which handover is delayed after event condition is met) from OAC to RF cells [7]. This approach assumes the UT immediately hands over from RF cells to OAC whenever it is in OAC coverage. Optimistically connecting to the OAC and only optimizing the dwell time from OAC to RF cells overlook the non-negligible time cost of the handover process, which may end up with frequent disconnection. Although aggressively offloading data traffic from RF cells to OACs mitigates the congestion in RF-based wireless networks, the consequent user-experienced data rate and latency may be unacceptable even for a UT crossing the small-coverage OACs at a regular walking speed.

Key Technologies toward Self-Optimization

To enable self-optimization of mobile HetNets, we consider optimizing the handover strategies within each OAC in a centralized way while leaving the AP-user association to predefined handover triggering conditions. The optimal handover parameters are customized based on the classified type of user mobility and behavior toward self-optimization.

Mobile Assisted Handover

Handover can be classified based on the handover techniques used. Broadly, they can be classified into three types: network controlled handover (NCHO), mobile controlled handover (MCHO), and mobile assisted handover (MAHO). In NCHO, the network makes a handover decision based on the measurements of UTs at several APs. Information about the channel quality for all the UTs is available at a single point in the network that facilitates resource allocation. In MCHO, each UT takes complete control of the handover decision process. By measuring the RSS from surrounding APs, the UT initiates the handover when the RSS of the serving AP is worse than that of the target AP by a certain threshold. The fully decentralized handover control overlooks the network conditions and will severely degrade QoE with a large number of Internet of Things (IoT) devices. In MAHO, instead of the network making the measurements, the mobile UT collects the measurements, usually in the form of SINR, RSS, bit error rate, and so on, and sends them to the network to make a handover decision [8]. MAHO allows the considered mobile HetNets to self-optimize the handover strategies based on the best knowledge of the QoE of each UT.

User Profiling

An indoor mobile user exhibits similar mobility patterns, behaviors, and activities in daily work and life [3], especially in public places such as office buildings, hospitals, enterprises, schools, and so on. Grouping the user mobility information into profiles is of interest to the network. For instance, the users have a greater tendency to consume a video service when they are static, and the network would like to offload their traffic to OACs as much as possible. In contrast, when the users are in high mobility, they rather prefer video streaming, and the network may want to lower their chance to be switched to OACs. Such classifications help the correct estimation of user speed, providing important information about the consumption of network resources by the user and his/her mobility. The cognition of speed will then help the network operator to perform online network resource and handover optimization according to speed profiles [9].

Machine Learning

Machine learning (ML) models have been applied to a wide spectrum of applications and achieved state-of-the-art performance. In a problem such as optimizing handover control, an ML approach is a promising technique as it can learn to optimize handover based on observations and information from the environment. Along this line, RL and particularly its DRL breed seems to be the best fit among ML models. This can be seen in several aspects. One aspect is alleviating the need for explicit system modeling. The importance of this alleviation is expected to be more strongly pronounced in future heterogeneous wireless networks with increased system complexity. Moreover, DRL allows for accommodating additional user and system information to further guide the optimization process. This is due to its ability to handle high-dimensional state spaces. This promises to achieve fast and efficient performance and better exploitation of the available information about the system and the environment. With the anticipated surge of user types and mobility patterns in future networks, adaptivity and self-adjustment are key requirements in any handover mechanism.

System Model

To optimize data offloading in the considered mobile HetNets in a self-organized manner with context awareness, we propose a system framework (Fig. 1) that is running a self-optimizing algorithm in a CC that may reside in a cloud server to control the handover parameters of the RF femtocells and OACs located under the coverage of the RF macrocell. We consider a control and user plane framework similar to the dual connectivity configuration proposed in [10]. However, we aim to self-optimize the handover parameters instead of heuristically adjusting them.

Typically, to design a handover policy, TTT and HM are set to appropriate values to lower the drop rate and the ping-pong rate. Nonetheless, since the line-of-sight signal is dominant in OAC, the value of SINR either increases or decreases monotonically when the UT crosses the OAC

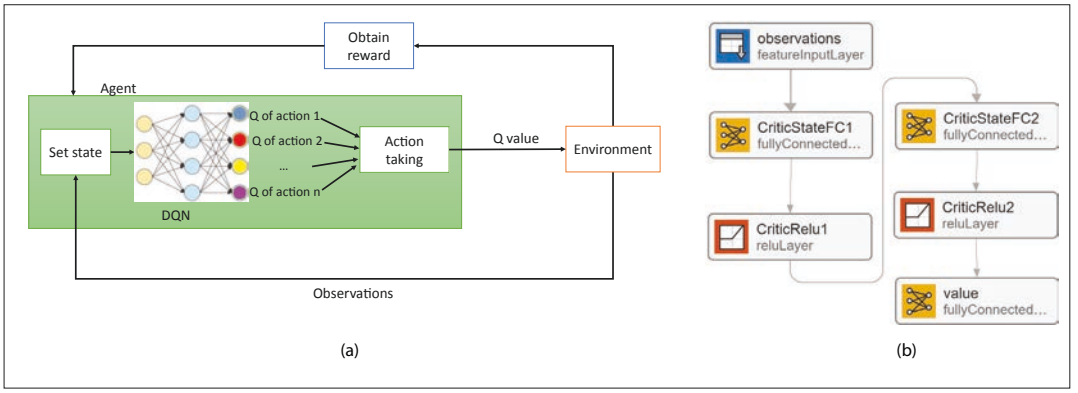

FIGURE 3. a) The proposed system's block diagram; b) the DNN's layer diagram. Q denotes the eventual reward value.

edge. Accordingly, the value of HM, which is leveraged to mitigate the ping-pong effect caused by the RSS fluctuation at the cell edge, is set to zero for the VHO from RF cell to OAC. Pertaining to handover control, the UT sends SINR reports of the potential serving cells periodically to the CC through RF macrocell uplink control signaling. Once the SINR reports meet a handover event condition, the TTT starts counting. If the event condition is satisfied until TTT counts to zero, the CC initiates the handover process by sending a Radio Resource Control (RRC) message [10] (i.e., shown in green) to the UT. By manipulating the value of TTT, the network adjusts the handover sensitivity to dynamic user mobility patterns.

The training process of TTT values is performed in a time-step-based manner. A certain period of time, which can typically be set as a day due to the statistically repetitive user translational movement patterns, is divided into multiple equal-length time steps, and the value of TTT is fixed in each time step duration. In order to enable refined optimization of the handover parameters and avoid sudden large changes, the value of TTT can only be increased or decreased by a single level in each time step. The training process has two forms: the online one directly measures the QoE metrics during the operation of the system, tunes the handover parameters, and updates the policy accordingly; while the offline one runs in the background by building a simulation environment based on the SINR feedback collected during the system operation. The offline training significantly shortens the convergence time, which could be a challenge due to the lack of training samples if only the online form is adopted.

Problem Formulation

This work considers the problem of optimizing the TTT values for the handover between RF cells and OACs. The objective is to maximize the average user throughput over APs. The optimization is an adjustment of the TTT values based on their current values and the status of the environment characterized by certain observations. Such observations may include, for example, average user speed, the number of users, and their RSSI values. The next system state depends on the current state and the action taken to change such a state. In other words, the system has a *Markov* property. Therefore, one may cast the optimization problem as a Markov decision process (MDP). To this end, deep Q-learning (DQN) [11] is a reasonable solution framework that promises to accommodate high-dimensional state spaces.

DQN is a model-free and off-policy DRL method. Similar to other RL settings, DQN trains an agent based on states and rewards. Specifically, a DQN agent is a value-based RL agent that trains a deep neural network (DNN) referred to as the critic network. However, this critic network is trained to give an estimate of the long-run reward for each state-action pair. It is the deep characteristic of this network that makes it possible to include high-dimensional state spaces.

Similar to other RL approaches, the DQN setting works in two phases. First is a training phase where the DQN is trained over a set of episodes. The observed reward is used to update DNN using Bellman's equation. During the training phase, the weights of DNN are adjusted such that for a given action-state input, the Q value is predicted. The next phase is an inference phase representing the runtime operation. DNN can still be updated through the experience it gets during the inference phase. The exploration-exploitation trade-off of DQN is controlled by the ε-greedy approach, which is used as a probability threshold to either selecting an action at random or selecting it such that it maximizes the state-action value function.

Figure 3a shows a block diagram of the proposed DQN system. The DQN agent located at the CC interacts with the environment by adjusting the handover parameters (i.e., taking an action). The average throughput of RF and OAC users is estimated using the SINR feedback and regarded as the reward. After that, the information is used to update the critic DNN network.

The State-Space

Each state is a tuple composed of the TTT values, the user's speed, and the number of users. This information is readily available at the CC at no additional cost. One can think of incorporating other information/observations about the users or the environment once possible. This is expected to lead to more fine-grained state definitions and eventually better performance.

The Action Space

Actions are taken to adjust the handover parameters. This work assumes a set of predefined TTT values and each action corresponds to either increasing or decreasing one or two of the TTT

Parameter	Value
OAC coverage radius	1.5 m
TTT range	[0, 5.12] s
Number of time steps/episode	48
Initial ε	0.3
Minimum ε	0.001
ε-decay	0.001
Number of episodes	100
Initial state	[0, 5.12] s
Number of OAC APs	4
Number of RF APs	1
Experience buffer length	10000 samples
Mini-batch size	200

TABLE 1. Values of the main simulation parameters.

parameters by one step.[1] Hence these values will not exhibit abrupt changes. Thus, the action space is a discrete space and the role of the agent is to choose the best action. Still, one may additionally adopt a policy-based RL strategy whereby actions can be learned from the RL framework, rather than chosen from a predefined set.

The Reward Function

The reward function adopted in this work is the average throughput per user, across RF and OAC APs, as suggested in [3]. Technically, the throughput is obtained as a function of the SINR which is regularly being monitored at the CC.

DNN Configurations

A layer diagram of the adopted critic DNN structure is depicted in Fig. 3b. The critic DNN receives the state tuple as an input and outputs the expected cumulative long-term reward (Q-value) when the corresponding discrete action is taken. Therefore, the critic DNN used is composed of a feature input layer having the same size as the state dimension. This is followed by hidden layers: two stages of fully connected layers followed by rectified linear unit (ReLu) activation functions. Eventually, it is terminated by a fully connected output layer having the same dimension of the action space.[2] This approach of having all Q-values calculated with one pass through the network avoids having to run the network individually for every action and helps to increase speed significantly. Note that the DNN architecture is a generic one, and many others can be used.

Optimizers used in the training of DL models require independent and identically distributed data for their convergence. However, this is not the case with RL as states and their rewards and actions are correlated across time. Thus, naively training the DNN with sampled data results in oscillations and possible divergence in its training. To this end, a remedy to this problem is the advent of *experience replay*. In this setting, one continuously samples observations and rewards from the environment and stores them in a so-called *experience buffer*. Then the DNN is updated by a mini-batch of randomly selected data points from the experience buffer. In addition to breaking harmful correlations, experience replay allows learning more information from individual tuples multiple times, recalls rare occurrences, and in general makes better use of experiences.

Numerical Results and Performance Analysis

The performance of the proposed DQN-based handover control algorithm is characterized by its convergence rate and converged value. Simulations are conducted in MATLAB and Simulink 2021. Simulink is used to simulate the mobile user environment. Table 1 lists key simulation parameters and their chosen values. More specifically, we use a system composed of one RF AP[3] and four OAC APs. Each OAC AP has a coverage radius of 2 m. All OAC APs are uniformly disposed in the coverage area of the RF AP. In each time step, the average user speed and number of users are Gaussian functions of time, where the user speed is low and the number of users is high at midday. There is a total of 16 candidate values in [0, 5.12] s for each TTT value set according to [12]. This simulation considers the number of time steps per episode to be 48, and the duration of each time step is half an hour. The simulation results indicate the time cost of the offline training parallel to the online training during the system operation.

During training, the agent randomly selects a mini-batch of 200 data points from an experience buffer of 104 training samples. We set the initial value of ε to 0.3, the minimum value of ε to 0.01, and the ε decay rate to 0.001. The DNN architecture has an input layer of 9 neurons, followed by two fully connected hidden layers of 12 and 48 neurons, respectively. The model is terminated by an output layer of 9 neurons.

The learning reward convergence of the proposed algorithm is evaluated in two simulations. In the first simulation, we evaluate the convergence performance with different state-space dimensions. In the second simulation, we evaluate the adaptation capability to different operational environments and conditions. The results are averaged over 10 trials to have statistical significance.

The results of the first simulation are presented in Fig. 4a, which shows the convergence for three state-space settings. First is a two-dimensional (2D) setting where a state is composed of the two TTT values only. Second is a three-dimensional (3D) setting where a state is a concatenation of the TTT values, and one observation that is the average user speed. Similarly, the third setting is four-dimensional (4D), where each state is a concatenation of the two TTT values, the average user speed, and the user bandwidth.

Several observations can be made in view of Fig. 4a. First, it is clearly evident that DQN can learn in a small number of episodes. This agrees with the intuition that the use of DNN is expected to enhance the speed of convergence. More specifically, the 2D case converges in around 20 iterations. Next is the 3D case in around 35 iterations, and then the 4D case in around the same number of iterations. The results are intuitively sound in the sense that higher-dimension state space corresponds to higher computational burdens despite promising better optimization. Second, the added benefit of increasing the dimension of state space is seen in converging to higher terminal values

[1] We assume actions of incrementing or decrementing the values of TTT by one step.

[2] Because the action space is discrete, the output corresponds to this space.

[3] Incorporating more RF APs does not incur substantial changes to the proposed approach since the RF AP with the strongest SINR for each OAC is mostly determined. The evaluation of the impact of "more than one RF AP" on the convergence rate is considered as our future work.

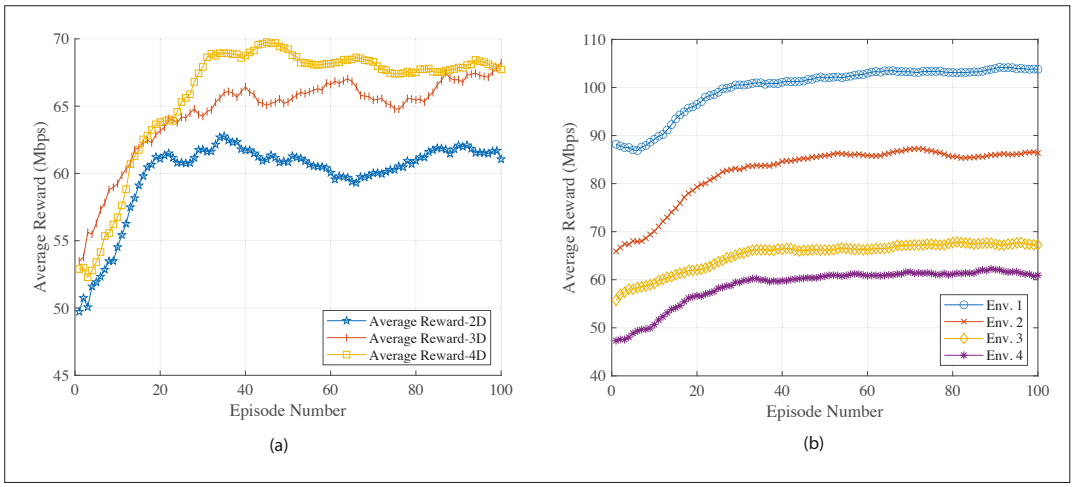

FIGURE 4. Average reward vs. episode for 10 trails with a) different state space dimensions; b) different operation environments. For a), the 95 percent confidence intervals are [61.19, 64.34] Mb/s, [66.63, 69.83] Mb/s, and [68.05, 71.19] Mb/s, respectively for the 2D, 3D and 4D settings. For b), the 95 percent confidence intervals of the Env. 1 through Env. 4 are [103.44, 104.97] Mb/s, [85.77, 88.54] Mb/s, [66.00, 69.70] Mb/s, and [61.25, 63.29] Mb/s, respectively.

albeit after a few more iterations. This observation is consistently correct as one moves from the 2D to 4D settings. In conclusion, one can clearly see the added benefit of using more fine-grained states. Essentially, each dimension corresponds to a certain measurement or knowledge about the system. To this end, the DQN framework allows one to afford high-dimensional state spaces efficiently and seamlessly.

To quantitatively study the execution time of the proposed algorithm, we average execution time over 10 trials for 100 episodes. The tic-toc function in MATLAB 2021-a is run on a Dell OPTIPLEX workstation with Intel Core i7 (8-core and 3.81 GHz) and 8 GB RAM. The 2D, 3D, and 4D cases take 159.39, 161.58, and 161.83 s to finish, respectively.

The second simulation considers different environmental settings, including mobility pattern and number of OAC APs. In this work, we adopt two mobility patterns: a straight-line pattern and a zigzag pattern where the user performs an orthogonal turn after a certain travel distance.

In particular, we compare the convergence performance as exposed to the following four settings:

- Env. 1: Two OAC APs, straight-line pattern
- Env. 2: Three OAC APs, straight-line pattern
- Env. 3: Three OAC APs, zigzag pattern
- Env. 4: Four OAC APs, zigzag pattern.

In view of Fig. 4b, the average reward is higher for simpler mobility patterns. We also observe that adding more OAC APs does not necessarily enhance the reward. The reason is that the increased complexity of handover decision making may reduce the average throughput experienced by each user. Overall, the simulation results validate that the proposed DQN algorithm can converge fast and adapt to different environments. This is a key requirement for future communication networks where systems are expected to be strongly dynamic and changing.

Finally, we compare our proposed scheme to two state-of-the-art approaches, namely quasi-static (QS) [6] and immediate handover (IH) [7] schemes, under Env. 2 and Env. 3 settings. We compare the three schemes in terms of average Th and average number of handover operations (NHO) per user. As shown in Fig. 5, our proposed scheme outperforms QS and IH due to the reduction of unnecessary handover operations.

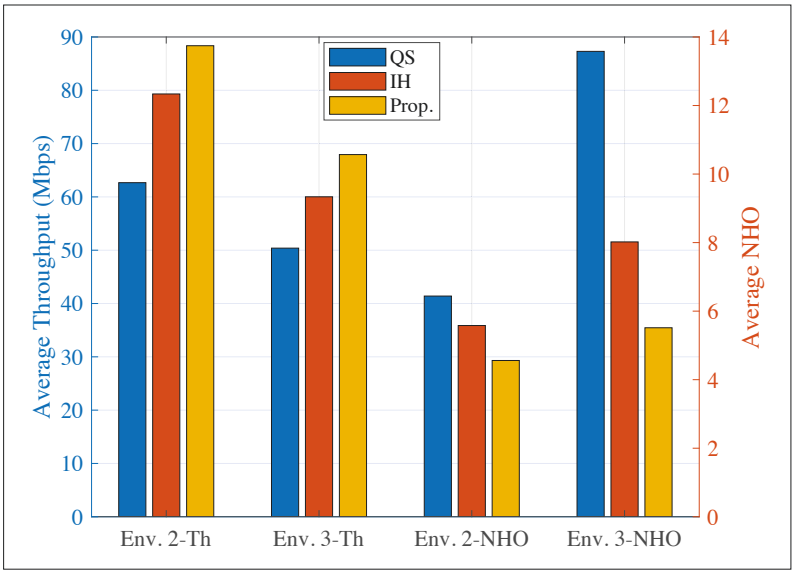

FIGURE 5. Performance comparison of the proposed approach vs. quasi-static (QS) and immediate handover (IH) schemes, compared in terms of average throughput (Th) on the left vertical axis, and average number of handovers (NHO) per user on the vertical right axis, under Env. 2 and Env. 3 settings.

Conclusion and Future Work

In this article, we study the next generation mobile wireless HetNet and consider optimizing the data offloading policy. We introduce the difference between inter-RF cells handover and the handover between RF cells and OACs. We highlight the limitations of the current VHO approaches and the developments of emerging technologies toward self-optimization. Then we discuss how the technologies are integrated into the system model. We formulate a handover parameter optimization problem and propose a DRL-based offloading strategy. The simulation results validate that the system adapts to dynamic environments with a fast convergence rate, which is

The simulation results validate that the system adapts to dynamic environments with a fast convergence rate, which is an essential feature of network self-optimization.

an essential feature of network self-optimization. As our future work, we will incorporate more realistic mobility models in the simulation, such as detailed models tailored for specific scenarios and real trace models. We will also study the integration of multiple access technologies in the reward function design [13], channel aggregation and load balancing among different links [14], and secure data transmission using the narrow-interception-range OAC links [15].

Acknowledgment

This work was supported by NSF grants OIA-1757207 and CNS-1647170.

References

[1] Ericsson, "Planning In-Building Coverage for 5G: From Rules of Thumb to Statistics and AI"; https://www.ericsson.com/en/mobility-report/articles/indoor-outdoor.
[2] W. Saad *et al.*, "A Vision of 6G Wireless Systems: Applications, Trends, Technologies, and Open Research Problems," *IEEE Network*, vol. 34, no. 3, May/June 2019, pp. 134–42.
[3] S. Shao *et al.*, "Optimizing Handover Parameters by Q-Learning for Heterogeneous Radio-Optical Networks," *IEEE Photonics J.*, vol. 12, no. 1, 2020.
[4] J. Zhang *et al.*, "Dancing with Light: Predictive In-Frame Rate Selection for Visible Light Networks," *Proc. IEEE INFOCOM*, 2015, pp. 2434–42.
[5] S. Shao *et al.*, "Joint Link Scheduling and Brightness Control for Greening VLC-Based Indoor Access Networks," *J. Optical Commun. Networking*, vol. 8, no. 3, 2016, pp.148–61.
[6] T. M. Duong and S. Kwon, "Vertical Handover Analysis for Randomly Deployed Small Cells in Heterogeneous Networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, 2020, pp. 2282–92.
[7] A. Zeshan and T. Baykas, "Location Aware Vertical Handover in a VLC/WLAN Hybrid Network," *IEEE Access*, vol.9, 2021, pp. 129,810–19.
[8] L. Jiao *et al.*, "Enabling Efficient Blockage-Aware Handover in RIS-Assisted mmWave Cellular Networks," *IEEE Trans. Wireless Commun.*, vol. 21, no. 4, 2022, pp. 2243–57.
[9] I. Saffar *et al.*, "Deep Learning Based Speed Profiling for Mobile Users in 5G Cellular Networks," *Proc. IEEE GLOBECOM*, 2019, pp. 1–7.
[10] M. Polese *et al.*, "Improved Handover Through Dual Connectivity in 5G mmWave Mobile Networks," *IEEE JSAC*, vol. 35, no. 9, 2017, pp. 2069–84.
[11] V. Mnih *et al.*, "Human-Level Control Through Deep Reinforcement Learning," *Nature*, vol. 518, no. 7540, 2015, pp. 529–33.
[12] T. ETSI, 136 331 v13. 0.0 (Jan. 2016) LTE, "Evolved Universal Terrestrial Radio Access (E-UTRA)," Jan. 2016, pp. 2016–670.
[13] A. Farhadi Zavleh and H. Bakhshi, "Resource Allocation in Sparse Code Multiple Access-Based Systems for Cloud-Radio Access Network in 5G Networks," *Trans. Emerging Telecommun. Technologies*, vol. 32, no. 1, 2021, p. e4153.
[14] Y. Li *et al.*, "Deep Reinforcement Learning for Dynamic Spectrum Sensing and Aggregation in Multi-Channel Wireless Networks," *IEEE Trans. Cognitive Commun. Networking*, vol. 6, no. 2, 2020, pp. 464–75.
[15] G. Pan *et al.*, "Secure Cooperative Hybrid VLC-RF Systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 11, 2020, pp. 7097–7107.

Biographies

Sihua Shao [M'18] (sihua.shao@nmt.edu) received his Ph.D. degree and the Hashimoto Prize for best doctoral dissertation from the New Jersey Institute of Technology in 2018. Currently, he is an assistant professor with the Department of Electrical Engineering at New Mexico Tech. His research interests include wireless communication and wireless networks.

Mahmoud Nazzal [S'14] (mn69@njit.edu) received his Ph.D. degree from Eastern Mediterranean University in 2015. He is currently a Ph.D. student with the Department of Electrical and Computer Engineering at New Jersey Institute of Technology. His research interests include sparse coding, machine learning, and signal processing for wireless communication.

Abdallah Khreishah [S'07, M'11, SM'18] (abdallah@njit.edu) received his Ph.D. degree from Purdue University in 2010. He is currently an associate professor with the Department of Electrical and Computer Engineering at New Jersey Institute of Technology. His research spans wireless networks, visible light communication, congestion control, edge computing, and network security.

Moussa Ayyash [SM'12] (msma@ieee.org) is currently a professor of computing at Chicago State University. His current research interests span digital and data communication, wireless networking, visible light communication, network security, and machine learning. He is a recipient of the 2018 Best Survey Paper Award from IEEE Communications Society.