Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 5871114, 11 pages
https://doi.org/10.1155/2021/5871114

Research Article

LB-DDQN for Handover Decision in Satellite-Terrestrial Integrated Networks

Dong-Fang Wu,[1,2] **Chuanhe Huang**,[1,2] **Yabo Yin**,[1,2] **Shidong Huang**,[1,2] **M. Wasim Abbas Ashraf**,[1,2] **Qianqian Guo**,[3] **and Lin Zhang**[4]

[1]*School of Computer Science, Wuhan University, Wuhan 430072, China*
[2]*Hubei Luojia Laboratory, Wuhan 430072, China*
[3]*School of Information Engineering, Zhengzhou Institute of Finance and Economics, Zhengzhou 450053, China*
[4]*Wuhan Maritime Communication Research Institute, Wuhan 430072, China*

Correspondence should be addressed to Chuanhe Huang; huangch@whu.edu.cn

Received 3 June 2021; Revised 22 October 2021; Accepted 15 November 2021; Published 8 December 2021

Academic Editor: Yanjie Fu

The frequent handover and handover failure problems obviously degrade the QoS of mobile users in the terrestrial segment (e.g., cellular networks) of satellite-terrestrial integrated networks (STINs). And the traditional handover decision methods rely on the historical data and produce the training cost. To solve these problems, the deep reinforcement learning- (DRL-) based handover decision methods are used in the handover management. In the existing DQN-based handover decision method, the overestimates of DQN method continue. Moreover, the current handover decision methods adopt the greedy strategy which lead to the load imbalance problem in base stations. Considering the handover decision and load imbalance problems, we proposed a load balancing-based double deep Q-network (LB-DDQN) method for handover decision. In the proposed load balancing strategy, we define a load coefficient to express the conditions of loading in each base station. The supplementary load balancing evaluation function evaluates the performance of this load balancing strategy. As the selected basic method, the DDQN method adopts the target Q-network and main Q-network to deal with the overestimate problem of the DQN method. Different from joint optimization, we input the load reward into the designed reward function. And the load coefficient becomes one handover decision factor. In our research, the handover decision and load imbalance problems are solved effectively and jointly. The experimental results show that the proposed LB-DDQN handover decision method obtains good performance in the handover decision. Moreover, the access of mobile users becomes more balancing and the throughput of network is also increased.

1. Introduction

As one important network of the future wireless networks, the STIN [1] offers mobile users communication services with wide coverage, high reliability, and low delay. This new integrated networks consist of terrestrial segment network and satellite segment network, which provides the future smart city with a ubiquitous network. Because of the increase of user's equipment and the mobility of users, the mobile users need to connect the optimal candidate cell to continue the communication. But the overlapped area, the access limitation of base station, and the random mobility of users make the handover more complex. How can mobile users decide the optimal candidate cell? Which decision factor is the key factor? Meanwhile, mobile users also need to deal with the handover decision problem in data transmission services. This research problem can directly influence the handover performance and Quality of Service (QoS) of mobile users in wireless networks. The handover management technology includes three steps: (1) information collection, (2) handover decision, and (3) handover execution [2]. Aiming at the frequent handover, handover failure, and load imbalance in the terrestrial segment of STIN, combining with the load balancing strategy, a load

balancing-based double deep Q-network (LB-DDQN) handover decision method is proposed.

The developments of handover decision algorithm attract more attention from academia and industry. The traditional handover decision methods include the simple additive weighting method, TOPSIS, decision function method, and Q-learning. These traditional methods focus on the measurement report, handover threshold and decision function, which relay on the priori knowledge and produce training cost. And the latest method is deep reinforcement learning (DRL) which combines the feature analysis ability of deep learning and decision-making ability of reinforcement learning. The DRL method effectively resolves the mobility of users and dynamics of networks. In the existing DRL methods, the value function-based DRL and the policy gradient-based DRL are the core of the fundamental approach and hot spots. The policy gradient-based DRL method is widely used in the Markov decision process with continuous action space. The DQN-based method is used in action space with discrete low dimension. And the researched handover decision problem in STIN is one typical deterministic discrete decision problem. And the DDQN method adopts the target Q-network and main Q-network to deal with the overestimate problem of DQN method. Therefore, we select the improved DDQN method to train the handover decision process.

Figure 1 shows that users' mobility patterns and distribution differences lead to the load imbalance in the overlapped area of MBS1 and MBS2. And the current handover decision methods adopt the greedy strategy which also lead to the load imbalance problem in base stations. In Figure 1, this arrow expresses the moving direction of LEO satellites along the periodic trajectory in the orbit. These vital decision factors affect the handover decision and lead to frequent handover and decreased network throughput. Recently, few studies pay attention to both handover decision and load balancing problems. Moreover, the unevenly distributed mobile users are forced to switch to the base stations with low load in the handover-based load balancing method. Unlike from this method, the SINR, delay, and load coefficient were selected as the handover decision factors in the LB-DDQN handover decision method. Combining with the load balancing strategy, the improved DDQN method constructed the Markov decision model to realize the handover decision. In our research work, the handover decision and load imbalance problems are solved effectively and jointly. The proposed method has good performance of handover and meets the demands of load balancing. In this research, our contributions are summarized as follows:

(1) We proposed the LB-DDQN handover decision method in STIN to deal with the frequent handover and load imbalance. The handover decision problem in the terrestrial segment of the STIN was resolved. Furthermore, the feasibility of the LB-DDQN handover method was proved by the experiments and analysis

(2) We constructed the load balancing strategy, including load coefficient, the reward of load, and load balancing evaluation function. The reward of load analyzed the load's influence on the handover in the LB-DDQN method. With the help of the load balancing evaluation function, the load condition of the whole network was intuitively evaluated

(3) We selected the load coefficient as the handover decision factor. In the training process of handover decision, the load balancing was also considered. The handover decision and load balancing were both realized. In our research work, the handover decision and load imbalance problems are solved effectively and jointly

The rest of this paper is organized as follows. The related works of handover decision are surveyed in Section 2. The system model is described in Section 3. The LB-DDQN handover decision method is proposed in Section 4. Simulation setups and results of experiments are provided in Section 5. Finally, Section 6 concludes this paper.

2. Related Work

The existing handover decision methods in the network can be divided into four categories [3]: the decision function method, the multiattribute decision method, the context-aware decision method, and the artificial intelligence decision method. In [4], a multiattribute decision method computed the weights of decision factors. The simple additive weighting method (SAW), the technique for order of preference by similarity to ideal solution (TOPSIS), and the grey relational analysis method (GRA) were adopted to select the optimal candidate cell base station. In [5], the analytic hierarchy process method (AHP) was used to obtain the weights of decision factors, and the orders of candidate base station were computed by the SAW method. In [6], the AHP method obtained the weights of decision factors, and the GRA method ranked the candidate cell base stations. In [7], combining with the property normalization and weight calculation, the improved multiattribute decision method ranked the candidate cell base stations. In [8], the designed decision maker performed the network selection and handover decision. In [9], a received signal strength indicator- (RSSI-) based fuzzy logic method executed fast handover and seamless handover. Considering the frequent handover and ping-pong effect, a predicted signal to interference plus noise ratio- (SINR-) based handover decision method was proposed [10]. In [11], a speed-aware-based handover decision method was proposed to deal with the influence of the handover process on the network throughput in the two layers' cell network. This handover decision method failed to choose the best candidate base station but the base station which kept the maximum service cycle. This selection makes sure of the cooperation between base stations and the elimination of interference. In [12], the improved fuzzy TOPSIS method reduced the scope of the candidate base stations, which increased the throughput of the network. In [13], a handover decision method using fuzzy logic and combinatorial fusion was proposed. According to the RSSI, data rate, and network delay, the handover was predicted. Combining with the TOPSIS, the proposed handover decision method

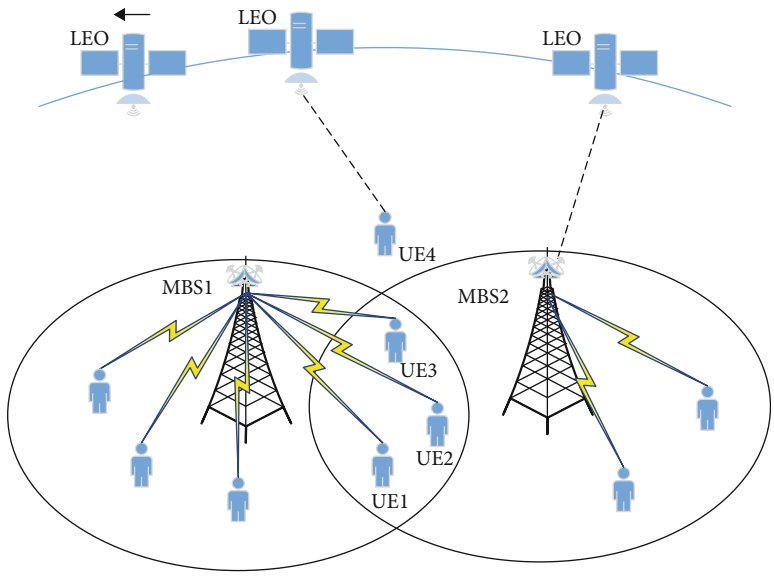

Figure 1: The scenario of handover decision and load imbalance in STIN.

assisted the mobile users to selecting the proper candidate base station [14]. In [15], combining with the improved competitive auction technique, the quality of uplink and downlink, and load factor, and the fast game-based handover decision method was proposed. The stochastic geometry analysis method was proposed [16]. By this method, the influence of different network topologies on handover decisions was analyzed. Moreover, the approximated handover number was estimated. In [17], an analysis framework of handover based on stochastic geometry was proposed to analyze the number of base stations, triggering time, and mobility patterns of users. The authors proposed a handover decision method based on fuzzy logic for saving the energy of mobile devices in an integrated LTE and Wi-Fi network [18]. And the traditional handover decision methods rely on the historical data and produce the training cost. The decision function method, the multiattribute decision method, and the context-aware decision method depend on the collected information about networks and users. This information plays an essential role in the handover decision. However, the collection of these vital information takes too much time. The delayed information and the dependence of prior knowledge lead to the frequent handover and load imbalance in the STIN.

The test of the reinforcement learning method on satellites designed by NASA proved that the artificial intelligence decision method had a good performance, and the deployment of this method was feasible [19]. In [20], the Markov handover decision model was constructed, and the hybrid vertical handover decision method was proposed. In [21], a Markov decision process- (MDP-) based handover decision method was proposed to optimize the QoS of network communication. In [22], considering the channel quality and QoS of communication, a reinforcement learning-based handover decision method was proposed from the point of user number. Google DeepMind team proposed the deep reinforcement learning method (DRL) which was evaluated in the Atari 2600, and this method achieved the excellent performance [23]. This new artificial intelligence method was used in communications and networking to deal with dynamic network access, data rate control, wireless caching, data offloading, and resource management [24]. In [25], a multiagent DRL method was proposed to resolve the distributed handover management problem. In this method, considering the cost factor, the user was modelled as an agent and the handover decision was optimized. In [26], the mobility patterns of users were classified, and the asynchronous multiagent DRL method was used in the handover decision process. In [27], the convergence speed and accuracy of the Q-network were optimized by the evolution strategy. The reinforcement learning-based handover decision method had good decision-making ability and handover performance. However, the state space, action space, and reward function of the different network scenarios need to be adjusted, which leads to the performance fluctuation. Moreover, the reinforcement learning-based handover decision method need to search the Q-table efficiently. This kind of method is suitable for the discrete state space problem. Replacing the Q-table by the neural network, the DRL-based handover decision method is good at dealing with the continuous state space problem. Therefore, our research adopts the improved DRL method to train the handover decision process.

3. System Model

3.1. Network Model and Problem Formulation. The terrestrial segment of the STIN makes up of M macro cells and N mobile users. In the satellite segment, there is always one LEO satellite that provides the satellite communication service. The mobile users select base station or LEO satellite to transmit data. Our research focused on the handover decision problem in the terrestrial segment of STIN. The network time T is divided into many time slots. In each time

slot t, every mobile user selects the optimal target base station from the M candidate base stations. When the mobile users move out of the range of network or the network time T is end, the state of the user update to tend, $T = \{t_0, t_1, \cdots, t_{\text{end}}\}$ includes many discrete handover decision scenarios. After the construction of the Markov handover decision model, the handover decision process is optimized by the neural network.

The handover decision process of mobile users is modelled as the discrete Markov decision process, expressed by $<S, A, \pi, r, \gamma>$, where S is the discrete state space of the network, which is composed of network parameters. The parameter A is the action space that is composed of the candidate base stations set. The parameter $\pi : S->A$ shows that the action A can be determined by the state S. The function of reward $r : S \times A->R$ computes the positive or negative rewards from the network environment. The discount coefficient γ describes the value of future reward. Figure 2 shows that the agent obtains the recent state s_t in each time slot t. The action a_t is determined by the strategy π. After receiving the parameter of action a_t, the network environment returns a reward r_t which shows whether the action is proper. Then, the state of network environment is update to s_{t+1}.

The proposed LB-DDQN handover decision method is aimed at maximizing the total reward of the handover decision. In the interaction of agent and environment, the discounted total reward G_t is defined as

$$G_t = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1} = R_{t+1} + \gamma \bullet R_{t+2} + \gamma^2 \bullet R_{t+3} + \bullet\bullet\bullet, \quad (1)$$

where R is the immediate reward of handover. The discount coefficient of future reward is named γ. According to Equation (1), the Bellman operator updates the action-value function $Q(s_t, a_t)$.

$$Q(s_t, a_t) = E[R_{t+1} + \gamma \bullet Q(s_{t+1}, a_{t+1})], \quad (2)$$

where s_t and a_t are state and action, respectively, in time slot t. According to Equation (2), the optimal Bellman operator is defined as

$$Q^*(s_t, a_t) = E\left[R_{t+1} + \gamma \bullet \max_{a_{t+1}} Q^*(s_{t+1}, a_{t+1})\right]. \quad (3)$$

Equation (3) describes that the maximum value of the action-value function $Q(s_{t+1}, a_{t+1})$ is computed. In the LB-DDQN handover decision method, the neural network is used to estimate the action-value function.

3.2. State Space and Action Space. In each time slot t, the size of candidate base station set for mobile users is M. We select the SINR, delay, and load coefficient as the decision factors. As for the candidate base station i, the state space is expressed as $c_i(t) = \{\text{SINR}_i(t), D_i(t), L_i(t)\}$. The total network state space is defined as $c(t) = (c_0(t), c_1(t), c_2(t), \cdots, c_{M-1}(t))$. The public interface X2 shares the load information of each base station. Moreover, the SINR and the delay information are obtained by the regular measurement reports. These selected decision factors assist in the handover decision.

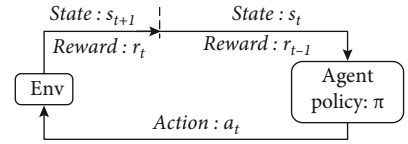

Figure 2: The framework of the reinforcement learning.

The action of handover for the mobile user is expressed by parameter a. And the action space is made up of all the indexes of the candidates base station, expressed as $A = \{0, 1, 2, \cdots, M-1\}$. According to the ε-greedy strategy, in each time slot t, the mobile user selects one proper base station to connect. When the action parameter $a = 0$, the base station whose index is 0 is selected.

3.3. Reward Function. Considering the selected network parameters in state space, the reward function is defined as

$$R_t = \sum_x w_x r_{x_i(t)}, \quad (4)$$

where the normalized reward of the parameter x for the candidate base station i in time slot t was expressed by $r_{xi(t)}$. The variable w_x is the weight coefficient of the parameter x ($0 \le w_x \le 1$). These weights are computed by the AHP method [6]. The load reward is computed from the load coefficient. The SINR can be obtained from the measurement report of base station. As the positive parameters such as SINR and load, the reward function is defined as

$$r_{x_i(t)} = \begin{cases} 1, & x_i(t) \ge X_{\max}, \\ \dfrac{x_i(t) - X_{\min}}{X_{\max} - X_{\min}}, & X_{\min} < x_i(t) < X_{\max}, \\ 0, & x_i(t) \le X_{\min}, \end{cases} \quad (5)$$

where the variables $X_{\max}$ and $X_{\min}$ are the maximum value and minimum value of network parameter x. The network parameter for candidate base station i in time slot t is expressed as $x_i(t)$ which is evaluated between [0,1]. As for the negative parameter, such as delay, the reward function is defined as

$$r_{x_i(t)} = \begin{cases} 1, & x_i(t) \le X_{\min}, \\ \dfrac{X_{\max} - x_i(t)}{X_{\max} - X_{\min}}, & X_{\min} < x_i(t) < X_{\max}, \\ 0, & x_i(t) \ge X_{\max}. \end{cases} \quad (6)$$

4. LB-DDQN-Based Handover Decision Method

4.1. Traditional Handover Decision Methods. The traditional handover decision methods include SAW [4], TOPSIS [17], and Q-learning [28]. Compared with the DRL handover decision method, these traditional methods depend on the

measurement report and prior knowledge. The traditional handover decision methods are unsuitable for the dynamically changing network environment. The load balancing problem is also not fully considered. The SAW method computes the order of the candidate base station. The sum of normalized parameters is defined as

$$\mathrm{Add}_i = \sum_{j=1}^{n} V_{ij} * \beta_j, \tag{7}$$

where the variable Add_i express the sum of normalized network parameter in the candidate base station i. The variable β_j is the weight of the parameter j. The variable V_{ij} is the normalized value of parameter j in the candidate base station i, where $i = 0, 1, 2, \cdots, M-1$ and $j = 1, 2, 3$. The TOPSIS [17] selects the optimal candidate base station. The Euclidean distance is defined as

$$\begin{aligned} U_i^* &= \sqrt{\sum_{j=1}^{n} \left(V_j^* - V_{ij}\right)^2}, \\ U_i' &= \sqrt{\sum_{j=1}^{n} \left(V_j' - V_{ij}\right)^2}, \end{aligned} \tag{8}$$

where U_i^* and U_i' are the Euclidean distances between V_{ij} and the optimal solution set $\{V_1^*, V_2^*, \cdots, V_n^*\}$ and the worst solution set $\{V_1', V_2', \cdots, V_n'\}$, respectively. We use the variable V_j^* and V_j' to express the maximum value and the minimum value of parameter j, respectively. And the closeness efficient C_i^* is defined as

$$C_i^* = \frac{U_i'}{U_i' + U_i^*}, \tag{9}$$

when the variable C_i^* close to 1 and the U_i^* is small which means the Euclidean distance between the candidate solution and optimal solution is small. The update of the action-value function in the Q-learning method is defined as

$$Q(s_t, a_t) = Q(s_t, a_t) + \eta \bullet \left(R_{t+1} + \gamma \bullet \max_a Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)\right). \tag{10}$$

4.2. Load Balancing Strategy. One base station services the fixed number of mobile users at the same time. When the number of connected users exceeds the upper limit of load or the mobile user cannot connect to the target base station, the handover request is failed. The proposed LB-DDQN handover decision method realizes the handover decision and load balancing simultaneously by the load balancing strategy. The limited resource block and unevenly distributed users fail the handover request. By the LB-DDQN method, the number of handover failure is effectively decreased. The load coefficient, load reward, and load balancing evaluation function are presented in the designed load balancing strategy. In time slot t, the load coefficient of base station i expressed by $L_{i,t}$ $(0 \le L_{i,t} \le 1)$ which is defined as

$$L_{i,t} = \frac{\mathrm{UEnum}_{i,t}}{\mathrm{Tnum}_i}. \tag{11}$$

The variable $L_{i,t}$ expresses the number of serving users. Assume that one mobile user connects up to one base station and occupies one resource block. In time slot t, the variable $\mathrm{UEnum}_{i,t}$ is the number of serving users. The variable $\mathrm{Tnu\,m}_i$ expresses the total number of resource blocks. When the load coefficient of the base station increases, the number of serving user increases. At this time, the load reward of the base station decreases, and the probability of handover selection decreases. On the contrary, when the load coefficient of the base station becomes small, the available source blocks and load reward become large. This base station is more possibly selected as the optimal candidate base station. In time slot t, the load reward of the base station i is defined as

$$\mathrm{HO\_reward}_{i,t} = \frac{1}{L_{i,t} + 1}, \tag{12}$$

where the load reward of the base station i in time slot t is expressed by $\mathrm{HO\_reward}_{i,t}$. Its value range is [0.5,1]. In time slot t, the load balancing evaluation function is defined as

$$\mathrm{LB}_t = \sqrt{\sum_{i \in M} \left(L_{i,t} - \frac{\sum_{i \in M} L_{i,t}}{M}\right)^2}, \tag{13}$$

where LB_t is the value of load balancing function in time slot t. When the value of this variable is bigger, the distribution of mobile users is more imbalanced. When the variable LB_t close to 0, the load of the base stations is balanced. The operation $\sum_{i \in M} L_{i,t}/M$ obtains the average load coefficient of all base stations.

4.3. Implementation of LB-DDQN. The DRL handover decision method adopts the neural network to estimate the optimal value of the action-value function. In the training process of handover decision, the normalized parameters of state space are regarded as the input of the neural network, and the optimal value of the action-value function is output.

$$Q(s_t, a_t; \theta) \approx Q^*(s_t, a_t). \tag{14}$$

In [23], the update of action-value function in the DQN method is defined as

$$Q_m(s_t, a_t) = Q_m(s_t, a_t) + \eta * \left(R_{t+1} + \gamma \bullet \max_a Q_t(s_{t+1}, \mathrm{a}) - Q_m(s_t, a_t)\right), \tag{15}$$

where Q_m is the action-value function of main Q-network and Q_t is the action-value function of target Q-network. We proposed the improved LB-DDQN handover decision method. When the maximum value of Q_m is obtained,

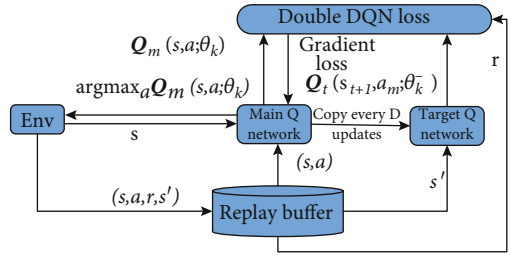

Figure 3: The framework of DDQN method.

the handover action a_m corresponding to the optimal Q_m is determined. The update of action-value function is defined as

$$a_m = \arg \max_a Q_m(s_{t+1}, a),$$
$$Q_m(s_t, a_t) = Q_m(s_t, a_t) + \eta * (R_{t+1} + \gamma \bullet Q_t(s_{t+1}, a_m) - Q_m(s_t, a_t)). \tag{16}$$

The loss function of DDQN method is the difference value between the target value y and the estimated action-value function $Q_m(s_t, a_t, \theta_k)$. The loss function is defined as

$$y = \begin{cases} R_{t+1}, & \text{if } s_{t+1} \text{ is end,} \\ R_{t+1} + \gamma \bullet Q_t(s_{t+1}, a_m; \theta_k^-), & \text{others,} \end{cases} \tag{17}$$
$$L_k(\theta_k) = E\left[(y - Q_m(s_t, a_t; \theta_k))^2\right].$$

In the training process of handover decision, the loss function returns the gradient loss to update the parameters of main Q-network at each iteration. With the updates of the parameters, the loss value of loss function decreases. And the performance of the handover becomes better. The loss function of the DDQN method is optimized by the stochastic gradient descent method. The gradient of loss function is defined as

$$\nabla_{\theta_k} L(\theta_k) = E\left[(y - Q_m(s_t, a_t; \theta_k)) \nabla_{\theta_k} \bullet Q_m(s_t, a_t; \theta_k)\right]. \tag{18}$$

As Figure 3 shows, the DDQN handover decision method adopts the main Q-network Q_m and the target Q-network Q_t. These two neural networks are initialized with the same network parameters. In the main Q-network Q_m, the network parameters are updated at each iteration. The main Q-network Q_m is used to estimate the value of action-value function. After every D steps, the network parameters of target Q-network are updated by the parameters of main Q-network. $\theta_k^- = \theta_k$. The target Q-network Q_t is used to compute the expected value of action-value function.

In the training process of handover decision, the experimental data is saved in the replay buffer. By the experience replay method and the small batch sampling method, the randomly sampled data is used as input data to train the parameters of Q-network. By the ε-greedy strategy, the exploration and the exploitation operations of the optimal handover action are realized. The detailed steps of the LB-DDQN handover decision method are described as Algorithm 1.

5. Simulation Results and Discussions

5.1. Simulation Environment. This research makes sure of the handover performance and load balancing requests simultaneously. A PC carries out the experiments with 3.2 GHz quad-core i5-1570 and 16GB of RAM. The OS is win 10, 64 bits, and the simulation platform is python3. Figure 4 shows that it is 2290 meters long and 1800 meters wide in the virtual town scenario. In this network area, there is one LEO satellite that provides 24-hour communication services. It includes 31 base stations whose communication range is 500 meters. Assume that the base station bandwidth is 10 MHz and the upper limits of connected users for the base station are 50. One user only occupies up to one resource block, and the bandwidth of each subchannel is 180 kHz. The starting point of the mobile user is randomly selected from 11 crossings. The speed of the mobile user is randomly selected from 5 km/h, 25 km/h, 50 km/h, 70 km/h, and 120 km/h. The mobile user is moving at a constant speed in straight lines. The number of mobile users is 50, 100, 200, and 200, respectively.

5.2. Simulation Parameters. The handover rate, handover failure rate, and throughput of the network are used to evaluate the handover performance. The simulation parameters are illustrated in Table 1.

The handover rate and handover failure rate are defined as

$$\text{HOR} = \frac{N_{\text{HO}}}{N_{\text{total}}},$$
$$\text{HOF} = \frac{N_{\text{re}} - N_{\text{HO}}}{N_{\text{total}}}, \tag{19}$$

where the variable N_{HO} is the number of successful handover and the variable N_{total} is the total number of handover decision. The N_{re} expresses the number of handover requests. The range of HOR and HOF is [0, 1]. The network parameter SINR is defined as

$$\text{SINR} = 10 \bullet \log\left(\frac{P_S}{P_I + P_N}\right), \tag{20}$$

where the variable P_S is the effective power and the variable P_I is the interference power. And the variable P_N is the noise power. The throughput of network Th is defined as

$$\text{Th} = W * \log_2\left(1 + \frac{P_S}{P_I + P_N}\right) \tag{21}$$

where the variable W is the bandwidth of the subchannel.

5.3. Simulation Results

5.3.1. Average Handover Number of User. As Figure 5 shows, the handover numbers of the LB-DDQN method are

Input: Iteration number NUM_EPISODES, step number MAX_STEPS, node number node_num, measurement information SINR and delay, length of update step D.
Output: Handover decision matrix $\boldsymbol{A}$.
1: Initialize action-value function Q, replay buffer B and handover decision matrix $\boldsymbol{A}$. The initialized parameters of the main Q-network and target Q-network are consistent. $\theta_k^- = \theta_k$.
2: **for** i=1, NUM_EPISODES **do**
3: **for** j=1, MAX_STEPS **do**
4: **for** k=1, node_num **do**
5: According to Eq. (4, 5, 6), the immediate reward $r_{j,k}$ is computed.
6: According to Eq. (11, 12, 13), the load coefficient $L_{i,t}$ and load reward HO_reward are obtained. Construct the sequences of state $s_{j,k}$ include: SINR, delay and HO_reward.
7: By the ε-greedy strategy, the handover action $a_{j,k}$ corresponding to the state $s_{j,k}$ is determined. And the handover decision matrix $\boldsymbol{A}$ is updated.
8: Construct the next state $s'_{j,k}$, the experience data $(s_{j,k}, a_{j,k}, r_{j,k}, s'_{j,k})$ is saved in the replay buffer B.
9: Using experience replay and small batch sampling method, the randomly sampled data from the replay buffer B is produced and input the main Q-network Q_m. Then the action-value function $Q_m(s,a)$ is obtained.
10: According to the Eq. (16, 17), the action a_m corresponding to the maximum value of Q_m is obtained and input the target Q-network Q_t. And the action-value $Q_t(s'_{j,k}, a_m)$ is computed.
11: Adopt the stochastic gradient descent method, according to Eq. (18), the parameters θ_k of main Q-network are updated.
12: **end for**
13: Every D steps, the parameters of target Q-network are updated by the parameters of main Q-network. $\theta_k^- = \theta_k$.
14: **end for**
15: **end for**
16: Return the handover decision matrix $\boldsymbol{A}$.

Algorithm 1: LB-DDQN handover decision algorithm.

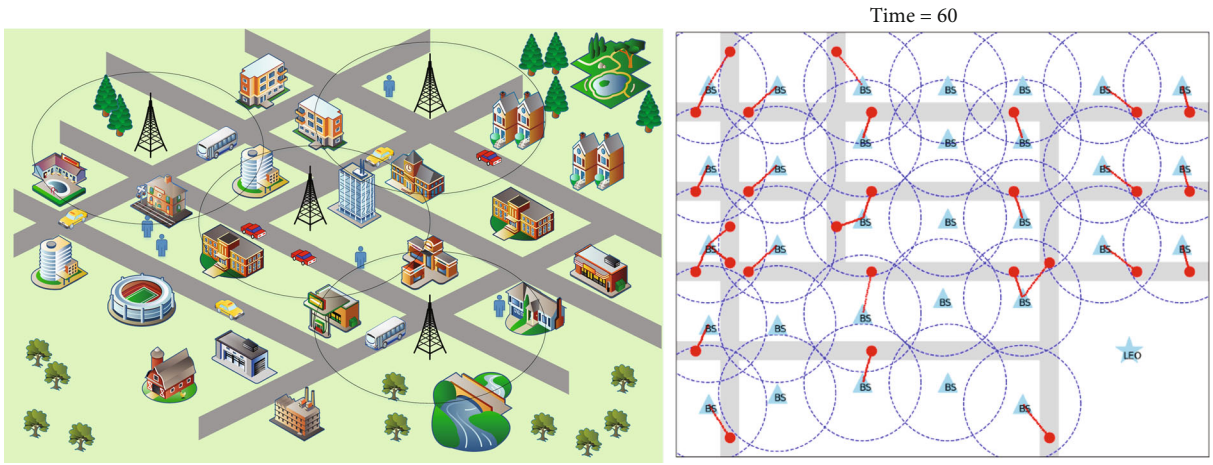

Figure 4: The scenario of the virtual town and the connection of mobile users at 60 second.

Table 1: Simulation parameters of the network.

Parameters	Values
Total number of BS	31
Bandwidth	10 MHz
Thermal noise	-174 dBm/Hz
Shadowing	8 dB
Cell transmit power	46 dBm
Path loss model	128.1 + 37.6 ∗ lg (d)
Number of UE	50, 100, 200, 300
UE speeds (km/h)	5, 25, 50, 70, 120
Duration of simulation	600 seconds

compared. The speeds of the user are set to 5 km/h, 25 km/h, 50 km/h, 70 km/h, and 120 km/h, respectively. The amounts of mobile users are set to 50, 100, 200, and 300, respectively. This figure assists in analyzing the influence of user speed and amount of users on handover performance. As we can see, when the speed of mobile user increases, the handover number of users gradually decreases. This is because in the network time T, when the user speed increases, the user is earlier entering the final state. And the decrease of the effective sampling points in the simulation leads to the decreased of handover decision number and handover times. In the virtual town scenario of the network, when the user speed is 120 km/h, and the length of the road is 1.8 km, the number of effective sampling points is 545. When the user speed

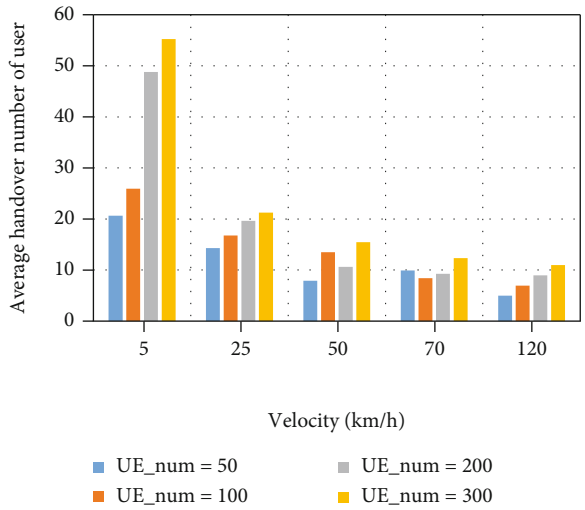

Figure 5: The handover number of LB-DDQN method with different user speeds.

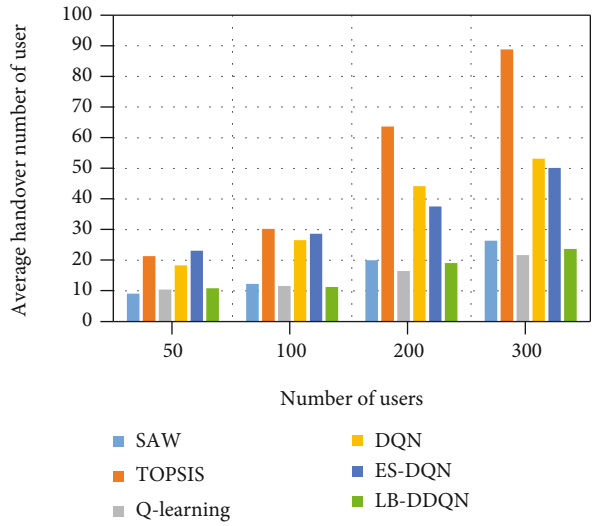

Figure 6: The handover performance of different handover decision methods.

changes to 5 km/h, the number of effective sampling points is 6000. Moreover, when the user speed is fixed, the increase of user number results in the increased of handover number. It is observed that the increase of mobile users leads to the increase of distribution difference, handover number, and interference signal. Meanwhile, handover management becomes more complex, because the SINR factor is also one of the handover decision factors, which leads to the increase of handover number.

As Figure 6 shows, the SAW [4], TOPSIS [17], Q-learning [28], DQN [27], and ES-DQN [27] handover decision methods are compared with the proposed LB-DDQN handover decision method in a different number of users. This figure shows the performance difference of these handover decision methods. As the increase of the number of users, the average handover number of user also increases. As for the traditional handover methods, the Q-learning-based handover method has the optimal handover performance. As for the artificial intelligence-based handover decision method, considering the load balancing factor as the decision factor, the proposed LB-DDQN method optimizes the handover decision process, and the average handover number of user decreases.

As we can see from Table 2, when the number of user is 50, the optimal handover decision method is the SAW method whose average handover number of a user is 9.1. The average handover number of the LB-DDQN method is 10.9. When the number of users is 100, the optimal handover decision method is the LB-DDQN method and the corresponding average handover number is 11.32. When the number of users is 200, the optimal handover decision method is Q-learning, and the average handover number is 16.53. And the average handover number of the LB-DDQN method is 19.07. When the number of user is 300, the optimal handover decision method is Q-learning, and the average handover decision method is 21.67. The average handover number of the LB-DDQN method is 23.74. The proposed LB-DDQN method has good handover, which is the same as the Q-learning. Moreover, the LB-DDQN method makes sure of the performance of handover and the continuity of data services.

Table 2: Average handover number of users for different handover decision methods.

UE_ Num	SAW	TOPSIS	Q-learning	DQN	ES-DQN	LB-DDQN
50	9.1	21.34	10.44	18.32	23.12	10.9
100	12.3	30.26	11.62	26.57	28.68	11.32
200	20.05	63.64	16.53	44.22	37.55	19.07
300	26.4	88.82	21.67	53.21	50.12	23.74

5.3.2. Handover Rate and Failure Rate. As Figure 7 shows, the handover rate and failure rate of different handover decision methods are compared. The number of user is 100. According to Equation (19), the handover rate and failure rate are computed. As for the handover rate, the optimal handover decision method is Q-learning. The handover rate of Q-learning is 0.0019. And the handover rate of LB-DDQN is 0.0021. As for the failure rate, the optimal method is the ES-DDQN method. Its failure rate is 0.0067. And the LB-DDQN is 0.007 which is better than Q-learning method. We find that the proposed LB-DDQN method has the good performance of handover rate and failure rate. Considering the load factor, the access of mobile users is more balanced, and the continuity of data services is enhanced. Compared with the DQN and ES-DQN methods, the LB-DDQN method decreases the handover rate, which makes sure of the request of handover. At the same time, by the load balancing strategy, the failure rate also decreases.

5.3.3. Throughput of Network. As Figure 8 shows, the throughput of the network for different handover decision methods are described. The number of users is 100. The optimal method is the proposed LB-DDQN handover

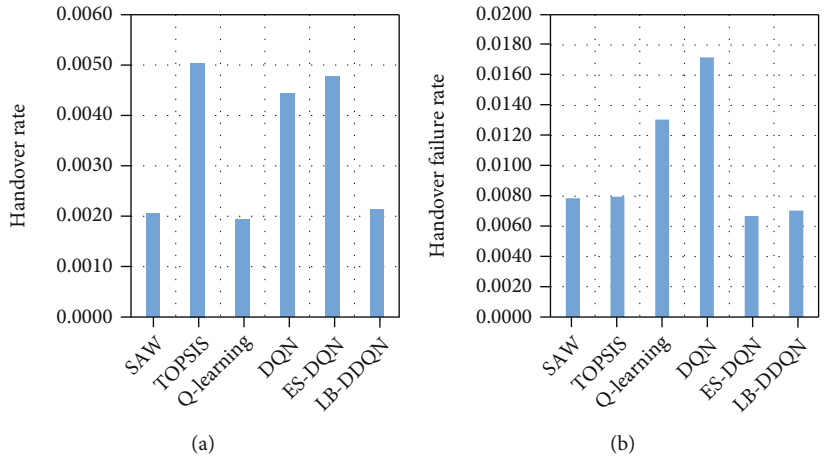

Figure 7: The handover rate (left) and failure rate (right) of different handover decision methods.

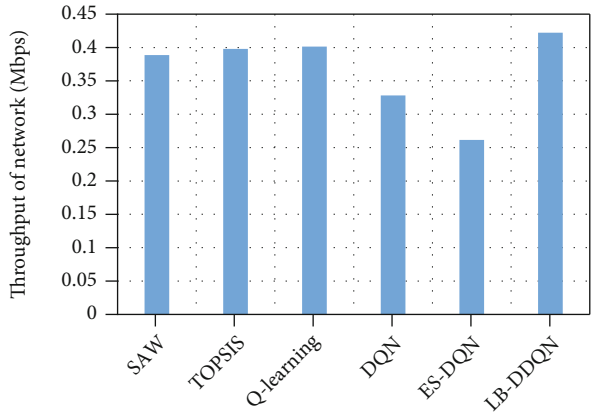

Figure 8: The throughput of network for mobile users.

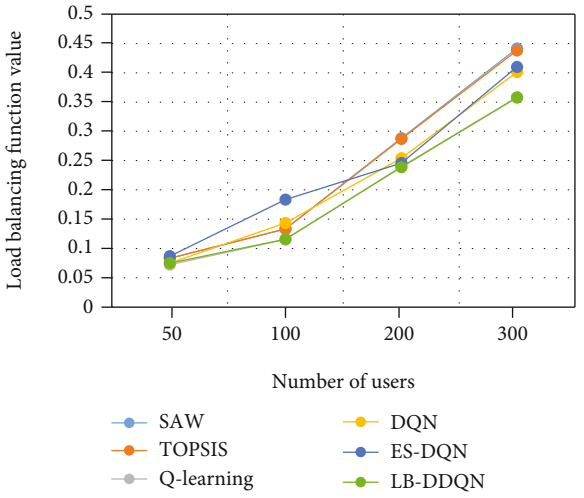

Figure 9: The values of load balancing function for different handover decision methods.

decision method whose throughput of the network is 0.4221 Mbps. The network throughput of the -learning method is 0.4012 Mbps. We find that by combining the load factor, the network throughput of the LB-DDQN method is higher than those of the others. Our load balancing strategy eliminates the effects of frequent handover, handover failure, and load imbalance.

5.3.4. Load Balancing Function Value. As Figure 9 shows, the evaluation of load balancing for these methods is described. The number of mobile users is 50, 100, 200, and 300, respectively. The load is smaller, and the distribution of mobile user is more balanced. As the number of users increases, the value of load balancing function increases. Moreover, the optimal method is the LB-DDQN handover decision method. This is because our method combines the load balancing strategy, and the load coefficient is also the decision factor.

As Table 3 shows, the detailed value of load balancing function is described. The number of users is 50, 100, 200, and 300, respectively. When the number of users is 50 and 100, respectively, the optimal method is the Q-learning method. The values of load balancing function are 0.0724 and 0.1154, respectively. When the number of users is 200 and 300, respectively, the optimal method is the LB-DDQN handover method. The values of load balancing function are 0.2388 and 0.3577, respectively. As the number of users increases, the difference in distribution of mobile users is more complex. The performance of the proposed load balancing strategy for the mobile user is good and proper.

Table 3: The values of load balancing function for different handover decision methods.

UE_ Num	SAW	TOPSIS	Q-learning	DQN	ES-DQN	LB-DDQN
50	0.0834	0.0832	0.0724	0.075	0.0868	0.075
100	0.1336	0.1333	0.1154	0.1436	0.1836	0.1158
200	0.2882	0.287	0.2396	0.2541	0.2456	0.2388
300	0.4416	0.438	0.3583	0.4012	0.4101	0.3577

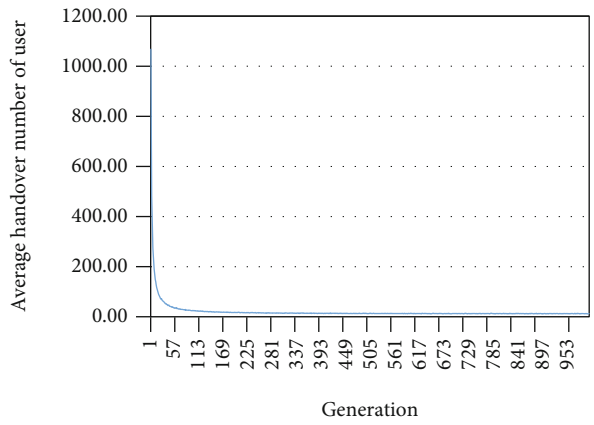

Figure 10: The convergence of LB-DDQN handover decision method.

5.3.5. The Convergence of the LB-DDQN Method. As Figure 10 shows, with the increases of generation number, the average handover number of users is convergent quickly. The number of mobile users is 100. When the number of generations is 127, the average handover number of users is 19.91. When the number of generations is 1000, the average handover number of a user is 11.32. Because of the random initialization for the Q-network, the initial value of average handover number for mobile users is very high. By the multiple iterations, experience replay, and small batch sampling methods, the weights of the main Q-network and the target Q-network are matured. The results of our method are also convergent rapidly.

6. Conclusions

The LB-DDQN handover decision method and the load balancing strategy are proposed in this paper. And the validation of our method is realized in the virtual simulation of the STIN. The designed load balancing strategy combines the load coefficient and load reward to assist the training of handover decision. The frequent handover and handover failure are optimized by the LB-DDQN handover decision method and load balancing strategy. The distributions of mobile users with different numbers are more balancing, and the number of handover failures is decreased. Furthermore, the LB-DDQN method adapts to the different conditions of user speeds, movement routes, and user numbers. Its adaptability and performance of handover decisions are good and low cost.

Data Availability

The data used to support the findings of this study are available from Dong-Fang Wu (at wudongfang@whu.edu.cn).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 61772385).

References

[1] J. Liu, Y. Shi, Z. M. Fadlullah, and N. Kato, "Space-air-ground integrated network: a survey," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 2714–2741, 2018.

[2] G. Gódor, Z. Jakó, Á. Knapp, and S. Imre, "A survey of handover management in LTE-based multi-tier femtocell networks: requirements, challenges and solutions," *Computer Networks*, vol. 76, pp. 17–41, 2015.

[3] A. Stamou, N. Dimitriou, K. Kontovasilis, and S. Papavassiliou, "Autonomic handover management for heterogeneous networks in a future internet context: a survey," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 4, pp. 3274–3297, 2019.

[4] N. P. Singh and B. Singh, "Vertical handoff decision in 4G wireless networks using multi attribute decision making approach," *Wireless Networks*, vol. 20, no. 5, pp. 1203–1211, 2014.

[5] S. Bhosale and R. Daruwala, "Multi-criteria vertical handoff decision algorithm using hierarchy modeling and additive weighting in an integrated WiFi/WiMAX/UMTS environment– a case study," *KSII Transactions on Internet and Information Systems*, vol. 8, no. 1, pp. 35–57, 2014.

[6] M. Alhabo, L. Zhang, and N. Nawaz, "GRA-based handover for dense small cells heterogeneous networks," *IET Communications*, vol. 13, no. 13, pp. 1928–1935, 2019.

[7] B. R. Chandavarkar and R. M. R. Guddeti, "Simplified and improved multiple attributes alternate ranking method for vertical handover decision in heterogeneous wireless networks," *Computer Communications*, vol. 83, pp. 81–97, 2016.

[8] I. Bisio and A. Sciarrone, "Fast multiattribute network selection technique for vertical handover in heterogeneous emergency communication systems," *Wireless Communications and Mobile Computing*, vol. 2019, 17 pages, 2019.

[9] A. Çalhan and M. Cicioğlu, "Handover scheme for 5G small cell networks with non-orthogonal multiple access," *Computer Networks*, vol. 183, article 107601, 2020.

[10] E. R. Bastidas-Puga, Á. G. Andrade, G. Galaviz, and D. H. Covarrubias, "Handover based on a predictive approach of signal-to-interference-plus-noise ratio for heterogeneous cellular networks," *IET Communications*, vol. 13, no. 6, pp. 672–678, 2019.

[11] R. Arshad, H. ElSawy, S. Sorour, T. Y. Al-Naffouri, and M.-S. Alouini, "Velocity-aware handover management in two-Tier cellular networks," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1851–1867, 2017.

[12] Y. S. Hussein, B. M. Ali, M. F. A. Rasid, A. Sali, and A. M. Mansoor, "A novel cell-selection optimization handover for long-term evolution (LTE) macrocellusing fuzzy TOPSIS," *Computer Communications*, vol. 73, pp. 22–33, 2016.

[13] I. Kustiawan, C.-Y. Liu, and D. F. Hsu, "Vertical handoff decision using fuzzification and combinatorial fusion," *IEEE Communications Letters*, vol. 21, no. 9, pp. 2089–2092, 2017.

[14] X. Duan, J. Wei, D. Tian et al., "Adaptive handover decision inspired by biological mechanism in vehicle ad-hoc networks," *Computers, Materials & Continua*, vol. 61, no. 3, pp. 1117–1128, 2019.

[15] C.-C. Tseng, H.-C. Wang, K.-C. Ting, C.-C. Wang, and F.-C. Kuo, "Fast game-based handoff mechanism with load balancing for LTE/LTE-A heterogeneous networks," *Journal of Network and Computer Applications*, vol. 85, pp. 106–115, 2017.

[16] T. M. Duong and S. Kwon, "Vertical handover analysis for randomly deployed small cells in heterogeneous networks," *IEEE Transactions on Wireless Communications*, vol. 19, no. 4, pp. 2282–2292, 2020.

[17] X. Xu, Z. Sun, X. Dai, T. Svensson, and X. Tao, "Modeling and analyzing the cross-tier handover in heterogeneous networks," *IEEE Transactions on Wireless Communications*, vol. 16, no. 12, pp. 7859–7869, 2017.

[18] T. Coqueiro, J. Jailton, T. Carvalho, and R. Francês, "A fuzzy logic system for vertical handover and maximizing battery lifetime in heterogeneous wireless multimedia networks," *Wireless Communications and Mobile Computing*, vol. 2019, Article ID 1213724, 13 pages, 2019.

[19] P. V. R. Ferreira, R. Paffenroth, A. M. Wyglinski et al., "Reinforcement learning for satellite communications: from LEO to deep space operations," *IEEE Communications Magazine*, vol. 57, no. 5, pp. 70–75, 2019.

[20] A. M. Vegni and E. Natalizio, "A hybrid (N/M)CHO soft/hard vertical handover technique for heterogeneous wireless networks," *Ad Hoc Networks*, vol. 14, pp. 51–70, 2014.

[21] S. Zang, W. Bao, P. L. Yeoh, B. Vucetic, and Y. Li, "Managing vertical handovers in millimeter wave heterogeneous networks," *IEEE Transactions on Communications*, vol. 67, no. 2, pp. 1629–1644, 2019.

[22] Y. Sun, G. Feng, S. Qin, Y.-C. Liang, and T.-S. P. Yum, "The SMART handoff policy for millimeter wave heterogeneous cellular networks," *IEEE Transactions on Mobile Computing*, vol. 17, no. 6, pp. 1456–1468, 2018.

[23] V. Mnih, K. Kavukcuoglu, D. Silver et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[24] N. C. Luong, D. T. Hoang, S. Gong et al., "Applications of deep reinforcement learning in communications and networking: a survey," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 4, pp. 3133–3174, 2019.

[25] A. D. D. M. Sana, E. C. Strinati, and A. Clemente, "Multi-agent deep reinforcement learning for distributed handover management in dense MmWave networks," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8976–8980, Barcelona, Spain, 2020.

[26] Z. Wang, L. Li, Y. Xu, H. Tian, and S. Cui, "Handover control in wireless systems via asynchronous multiuser deep reinforcement learning," *IEEE Internet of Things Journal*, vol. 5, no. 6, pp. 4296–4307, 2018.

[27] J. Sun, Z. Qian, X. Wang, and X. Wang, "ES-DQN-based vertical handoff algorithm for heterogeneous wireless networks," *IEEE Wireless Communications Letters*, vol. 9, no. 8, pp. 1327–1330, 2020.

[28] T. Goyal and S. Kaushal, "Handover optimization scheme for LTE-advance networks based on AHP-TOPSIS and Q-learning," *Computer Communications*, vol. 133, pp. 67–76, 2019.