Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 5006770, 16 pages
https://doi.org/10.1155/2022/5006770

Research Article

State Aware-Based Prioritized Experience Replay for Handover Decision in 5G Ultradense Networks

Dong-Fang Wu,[1,2] **Chuanhe Huang**,[1,2] **Yabo Yin**,[1,2] **Shidong Huang**,[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 13 August 2021; Revised 2 December 2021; Accepted 9 April 2022; Published 5 May 2022

Academic Editor: Amr Tolba

The traditional handover decision methods depend on the handover threshold and measurement reports, which cannot efficiently resolve the frequent handover issue and ping-pong effect in 5G (5 generation) ultradense networks. To reduce the unnecessary handover and improve the QoS (quality of service), combine with the analysis of dwell time, we propose a state aware-based prioritized experience replay (SA-PER) handover decision method. First, the cell dwell time is computed by the geometrical analysis of real-time locations of mobile users in cellular networks. The constructed state aware sequence including SINR, load coefficient, and dwell time is normalized by max-min normalization method. Then, the handover decision problem in 5G ultradense networks is formalized as a discrete Markov decision process (MDP). The random sampling and small batch sampling affect the performance of deep reinforcement learning methods. We adopt the prioritized experience replay (PER) method to resolve the learning efficiency problems. The state space, action space, and reward functions are designed. The normalized state aware decision matrix inputs the DDQN (double deep Q-network) method. The competitive and collaborative relationships between vertical handover and horizontal handover in 5G ultradense networks are mainly discussed. And the high average network throughput and long average cell dwell time make sure of the communication quality for mobile users.

1. Introduction

The Internet of Things (IoT) and related technologies consist of the important parts of the new generation information technologies. The typical application scenarios of IoT include Internet of vehicles, intelligent transportation, smart factory, and smart home. The rapid development of communication, computation, and networking technologies has made more IoT devices connected. In the IoT, besides of the typical fixed equipment (e.g., sensors and cameras), it also includes huge amount of mobile user devices (e.g., cell phone, cars, and UAV). There is also high demand for mobile traffic and many time-sensitive typical applications (e.g., automatic drive and telemedicine). The high speed, low delay, and ubiquitous network characters of 5G networks support the Internet of everything, which is the critical guarantee for the high quality of communication services and big data business in IoT application scenarios.

The 5G low band, midband, and LTE (Long-Term Evolution) small cell techniques cannot meet the requirements of massive devices access, high data rate, and huge amount of mobile traffic in the next generation wireless networks [1]. Therefore, we adopt high frequency section and the ultradense deployment technique of 5G networks in our research. In ultradense networks (UDN), the 5G critical techniques consist of the millimeter wave technology [2]. By the ultradense deployment of small cells, the network throughput and number of access users in two-layer cellular

network architecture are improved [3–5]. And the QoS (quality of service) requirements of mobile users are also satisfied. However, the small coverage and network access limitations of small cells bring about the frequent handover and ping-pong effect which directly influence the quality and continuity of communication services in 5G ultradense networks [6–8]. The traditional handover decision methods depend on the handover threshold and measurement report, which cannot efficiently resolve the frequent handover and ping-pong effect.

To reduce the unnecessary handover and improve the QoS, from the point of state aware method, combine with the analysis of dwell time, the SA-PER handover decision method is proposed. The handover management process in wireless networks includes three steps: information collection, handover decision, and handover execution [9]. Most research works focus on the improvements of handover decision methods [10]. In the handover decision process, the optimal candidate cellular is determined by the multiple handover decision criteria and efficient handover decision strategies [11]. And the handover rate, ping-pong effect, radio link failure rate, throughput, and so on are selected as the evaluation criteria. In this paper, the dwell time and prioritized experience replay are selected as the new handover criteria and handover strategy, respectively.

As Figure 1 shows, the 5G ultradense networks consist of two-layer cellular architecture, included macro base station (MBS) and small base station (SBS) [9]. The communication services and data transmission of mobile users are realized with the connections of macro cell or small cell. Because of the ultradense deployment of small cells, the overlapped coverage of macro cell and small cell is obvious. The small coverage and access users' limitation of small cell lead to the frequent handover and ping-pong effect [10]. In our study, the complex handover decision problem includes vertical handover (MBS-SBS) and horizontal handover (MBS-MBS and SBS-SBS). How do ordinary mobile users choose between horizontal handover and vertical handover? How do we improve the performance and efficiency of deep reinforcement learning-based handover decision methods? The traditional weighted multiple handover decision method is easily affected by the training process of weighted coefficients, which unable to maintain stable performance. The handover threshold and priori knowledge cannot solve the ping-pong effect completely. Therefore, the cell dwell time is selected as the handover decision criteria and prefer to choose the cell which provides the long connection time not the cell which provides the optimal network services. We should be aware that if we select the cell obtained the optimal network service, the frequent changes of optimal cell lead to the frequent handover and degrade the QoS of mobile users [3]. To deal with the overestimates of DQN-based handover decision method, the DDQN is selected as the base method. To improve the learning efficiency, convergence rate, and handover performance, the prioritized experience replay mechanism is added into DDQN. Combining with the analysis of cell dwell time and PER method, a state aware-based prioritized experience replay handover decision method is proposed to deal with the frequent handover and communication interrupt problems in 5G ultradense networks.

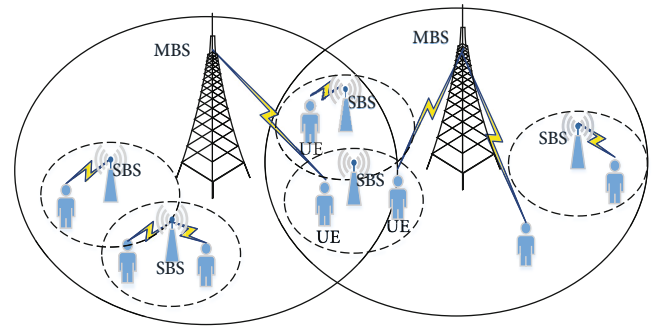

Figure 1: The scenario of horizontal handover and vertical handover for mobile users in 5G ultradense networks. The two-layer cellular architecture in 5G networks consists of MBS and SBS.

Our proposed method has good performance of handover and meets the demands of mobile communication service. In this research, our contributions are summarized as follows:

(1) The handover threshold and periodic measurement report cannot efficiently solve the frequent handover and ping-pong effect. And the ultradense deployment exacerbated the handover problems in 5G UDN. Aiming at the above handover problems in 5G UDN, we propose the SA-PER handover decision method to deal with the frequent handover and communication interrupt problems and reduce the ping-pong effect

(2) The dwell time of mobile users in cellular networks is analysed and calculated in detail. The proposed state aware method includes state aware sequence, max-min normalization, and normalized state decision matrix, which supports the preprocessing of data and assists the handover decision

(3) The handover decision problems of MBS-MBS, MBS-SBS, and SBS-SBS are carefully researched. Moreover, the competitive and collaborative relationships between vertical handover and horizontal handover in 5G UDN are concerned and analysed. Our analysis and discussion help mobile user better balance the choice between vertical handover and horizontal handover

The rest of this paper is organized as follows. The main research works of handover decision and existing challenges are introduced in Section 2. The system model is described in Section 3. The SA-PER handover decision method is proposed in Section 4. Simulation setups and experimental results are provided in Section 5. Finally, Section 6 concludes this paper. We summarize the definitions of the acronyms in this paper in Table 1.

2. Related Work

5G networks support the Internet of everything, which provides the ubiquitous communication services for the fixed IoT devices and mobile user devices. The mobility

Table 1: List of acronyms.

Symbol	Description
5G	5 generation
AHP	Analytic hierarchy process
A3C	Asynchronous advantage actor-critic
DDQN	Double deep Q-network
DNN	Deep neural networks
DQN	Deep Q-network
DRL	Deep reinforcement learning
ES	Evolution strategy
GRA	Grey relational analysis
HetNets	Heterogeneous networks
HOF	Handover failure rate
HOR	Handover rate
IoT	Internet of Things
KPIs	Key performance indicators
LTE	Long-Term Evolution
MBS	Macro base station
MDP	Markov decision process
PER	Prioritized experience replay
QoS	Quality of service
RL	Reinforcement learning
SAW	Simple additive weighting
SBS	Small base station
SDN	Software-defined network
SA-PER	State ware-based prioritized experience replay
TOPSIS	Technique for Order Preference by Similarity to Ideal Solution
UDN	Ultradense networks

management of the connected mobile devices is one critical challenge for the continuous communications and high quality of QoS. Therefore, many researchers focus on the handover problem of mobile devices. In high mobility scenario of IoT applications, such as UAV, the continuous communication connection and handover management are vital and nonignorable [12]. Sharma et al. [12] proposed a media independent handover-based fast handover security protocol in a heterogeneous IoT networks. The CoAP protocol is widely used in IoT networks. Chun and Park [13] proposed a CoAP-based mobility management protocol to realize the mobility management in IoT by the location management function. An SDN-based method realizes the mobility management in urban IoT heterogeneous networks [14]. Machine learning [15, 16] and reinforcement learning [17] have been widely applied to the research of handover management. As one new artificial intelligence method, DRL [18] is used in communications and networking to deal with many decision problems, e.g., handover decision. The high performance, online learning, and decision ability of DRL attracted much attention from the academia and industry.

The traditional handover decision methods in cellular networks include multi-attribute-based handover decision method [19], decision function-based handover decision method [15, 19], and context-aware-based handover decision method [20]. Bastidas-Puga et al. [19] proposed a predicted SINR-based handover decision method to deal with frequent handover and ping-pong effect. Singh and Singh [15] adopted the multiattribute decision method to obtain the weights of decision factors. By using the simple additive weighting (SAW), TOPSIS (Technique for Order Preference by Similarity to Ideal Solution), and grey relational analysis (GRA) methods, the candidate cells are decided. Hu et al. [20] proposed a velocity aware-based handover prediction method. The handover decision problem is formalized as the formal state-based shortest path problem in time expansion diagram. In [21], Goyal and Kaushal combined with the analytic hierarchy process method (AHP), TOPSIS, and reinforcement learning to optimize the selection of candidate cell. In addition, many researches adopt state aware in handover decision process, including context-aware [22, 23], mobility aware [6, 24], velocity aware [4, 20], and load aware [25]. The state aware method provides necessary data supports and decision basis for handover decision. In this paper, we adopt state aware method and cell dwell time to solve the performance fluctuation problem of traditional weighted multiple attribute handover decision methods.

There are many research works focus on the frequent handover, ping-pong effect, and handover failure problems in 5G

ultradense networks. Sun et al. [6] combined with the cell dwell time and movement state of users to match the candidate cells. By using movement aware handover decision method, the relations between dwell time and well connected cellular are balanced. In [26], by the assistance of unmanned aerial vehicles, the authors analysed the handover rate and dwell time of users in cellular networks. When the dwell time increases, the average handover numbers of users decrease, and the quality and continuity of communication services become better. Aiming at the frequent handover and increasing load of networks, Liu et al. [7] proposed a Q-learning-based handover decision method. The SDN (software-defined network) and 5G techniques were combined, and the entropy-based SAW handover decision method was proposed [8]. In recent researches, the base stations in cellular networks are selected as the edge computing node. Considering the migration of communication services, data services, and computing services, the researchers proposed a joint handover method and unloading decision method [27]. Huang et al. [16] firstly transformed the handover decision problem into the classification problem. Considering the changes of SINR parameter, the deep neural network (DNN) method realized the handover decision. Hasan et al. [28] classified the users into high speed users and ping-pong users. An elimination method of frequent handover was proposed. The energy cost issues of periodic measurements in 5G ultradense networks were also concerned [5].

The reinforcement learning-based handover decision method has good decision ability and handover performance, which is popular in handover decision researches in heterogeneous networks (HetNets) and UDN. Guidolin et al. [23] proposed an MDP-based handover decision method. By modelling the handover decision of mobile users, the optimal context handover decision standards were obtained. In [29], an MDP-based vertical handover method maximized the total expected rewards of handover. The AHP method computed the weight coefficients for the power, mobility, and energy cost decision factors. Yang et al. [30] and Sun et al. [31] adopted the multiarmed bandit handover decision method to produce handover decision strategies and reward. And the optimal candidate cell was determined. Tabrizi et al. [17] considered the state of networks and user devices and adopted Q-learning method to select candidate cells in handover decision process. The Q-learning-based handover decision method is widely used to solve the handover decision problems in terrestrial networks and satellite networks. The Q-learning-based handover decision method and relevant improved algorithms outperform the existing multiple attribute-based, decision function-based, and handover threshold-based methods. But, the Q-learning method needs to search the Q table for the optimal action in each iteration, which cost high searching time for the high dimensional state space. The Q-learning method is not suitable for the decision problem with high dimension state space. The DQN method replaces the Q table with DNN to describe the action value function, which is used to solve the decision problem with high dimension state space [32].

Google DeepMind team proposed the DRL method and obtained the superior performance in Atari 2600 games, which attracted more attentions from academia [33]. 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 [18]. In [34], the DQN-based handover decision method is used to deal with the frequent handover issue in UDN. The handover decision is formalized as a discrete Markov decision process. In [35], Sun et al. selected the evolution strategy (ES) to optimize the convergence speed and accuracy of backhaul network. And the DQN method was used in the vertical handover decision problem in HetNets. Wang et al. [36] creatively adapted the duelling network in reinforcement learning (RL). The proposed new network architecture represents two separate estimators, which express the state value function and the state-dependent action advantage function, respectively. The main benefit of this factoring is to generalize learning across actions without imposing any change to the underlying RL algorithm. To reduce the signalling overhead and solve the frequent handover, in [37], a double DRL method is proposed in 5G UDN, which reduces the handover numbers. By the trajectory-aware-based optimization method, the optimal candidate cell is determined with the trajectory of UE and topology of network. The connection time of UE-BS is increasing which reduces the handover overhead. Considering the handover decision problem in ultradense heterogeneous network, Song et al. [38] proposed a distributed DRL decision method. This proposed approach concerned the energy costs of transmission and handover load and minimized the total energy costs. In [39], the mobility patterns of users were classified, and the asynchronous multiagent DRL method was used in the handover decision process. In [40], the prior knowledge and supervised learning method are used to initialize the DNN, which offsets the bad effects of random exploration method. The frequent handover issue caused by deployment handover policy is solved by asynchronous advantage actor-critic-(A3C-) based handover method. In [41], the joint problem of handover and power allocation is formalized as the completely cooperated multiagent task, which is solved by the proposed proximal policy optimization-based multiagent reinforcement learning method. The global information is used in the training process of decentralized policy used in UE. In [32], Wu et al. proposed a load balancing-based double deep Q-network (LB-DDQN) method for handover decision. In the proposed load balancing strategy, a load coefficient is defined to express the conditions of loading in each base station. The supplementary load balancing evaluation function evaluates the performance of this load balancing strategy. The comparisons of different handover methods for cellular networks are shown in Table 2.

3. System Model

3.1. Network Model. In our research, the 5G UDN have tow-layer cellular architecture included M macro cells and N small cells. The deployment of heterogeneous cellular is shown as Figure 2. The communication services and data transmission of mobile users are provided by the connected macro cell or small cell. The state aware method periodically collects the data of network state, cellular state, and user

Table 2: Comparisons of different handover methods according to their characteristics.

Ref.	Problems and scenarios	Method	Contributions	Simulations	KPIs
[6]	Coordinated multipoint handover in 5G UDN	User-centric CoMP handover schemes	Characterize the movement trend through dwell time	Numerical simulation	HOR; Th
[7]	Handover triggering policy in 5G UDN	Clustering-based RL	Multiple decision criteria-based handover triggering mechanism	MATLAB; SD	HON; HOF; PPE; Th; latency
[12]	Handover failures and ping-pong effect in HetNet	HO triggers mechanism	Recursive least squares-based SINR prediction method	SD	HOF; PPE
[15]	Handover decision issue in LTE-A	AHP-TOPSIS; Q-learning	UE rank; the optimal triggering points of HO	MATLAB; SD	HOR; PPE
[17]	Handover in HetNets	Markov-based handover strategy	Context-aware handover policies	Monte Carlo simulations	Capacity
[22]	Handover in 5G	DNN	Reduce the handover problem to a classification problem	SD	RLF; PPE
[23]	Frequent handover in 5G UDN	Frequent handover mitigation algorithm	Dwell time estimation; user detection	NS3;	HON; Th
[27]	Handover decision in HetNets	Q-learning	Q-learning-based handover decision	SD	Cost; utility
[30]	Frequent handover in UDN	DQN	SDN-based UDN architecture; DQN-based handover decision	Mininet;	HOR; Th
[31]	VHO in HetNets	ES-DQN	Training the parameters of main Q-network with ES	Python; SD	HOF; Th; delay
[33]	Frequent handover in 5G	Double DRL	Trajectory-aware HO optimization approach	Wireless Insite software; SD	HON; Th
[34]	Handover decision in HetNets	Distributed DRL	MDP formulation; distributed DRL	SD	HON; energy cost
[36]	Handover in UDN	A3C	Mobility pattern-based user clustering; A3C-based HO policy	SD	HOR
[37]	Handover and power allocation in HetNets	Multiagent DRL	Proximal policy optimization; cooperative multiagent DRL	SD	HOR; Th
[43]	Load balancing and handover in 5G UDN	LB-DDQN	Load balancing strategy; load coefficient; load balancing evaluation function	Python; SD	HOR; Th
This paper	Handover decision in 5G UDN	SA-PER	State aware; analysis of dwell time; the relationships between VHO and HHO	Python; SD	RLF; HON; PPE; Th

VHO: vertical handover; HHO: horizontal handover; HON: handover number; HOR: handover rate; Th: throughput; HOF: handover failure rate; RLF: radio link failure; PPE: ping-pong rate; SD: simulated data.

state to support handover decision. The intelligent handover decision method is deployed in base stations, which collects the necessary data in real time and decides the optimal candidate cells.

3.2. Channel Model. The channel model of MBS and SBS in 5G UDN describes the characteristics of wireless channel [7]. The path loss of wireless link connected cell i and user j defined as follows:

$$\mathrm{PL}_{ij} = \begin{cases} 32.4 + 20\lg(f) + 30\lg(d_{ij}) + \chi, & \text{it is macro cell} \\ 32.4 + 20\lg(f) + 31.9\lg(d_{ij}) + \chi, & \text{it is small cell} \end{cases},$$
$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}, \tag{1}$$

where the path loss parameter named PL_{ij}, f is carrier frequency, and d_{ij} is the straight line distance between cell i and user j. The coordinates (x_i, y_i) and (x_j, y_j) express the real positon of cell i and user j, respectively. χ is the interference and noise modelled by Gaussian random and Rayleigh random variables. The parameter SINR is defined as follows:

$$\mathrm{SINR} = 10 \bullet \log\left(\frac{P_s}{P_I + P_N}\right), \tag{2}$$

where P_S, P_I, and P_N are the effective power, interference signal power, and noise power, respectively. The network throughput of the occupied subchannel Th is defined as follows:

$$\mathrm{Th} = W * \log_2\left(1 + \frac{P_s}{P_I + P_N}\right), \tag{3}$$

where W is bandwidth of subchannel.

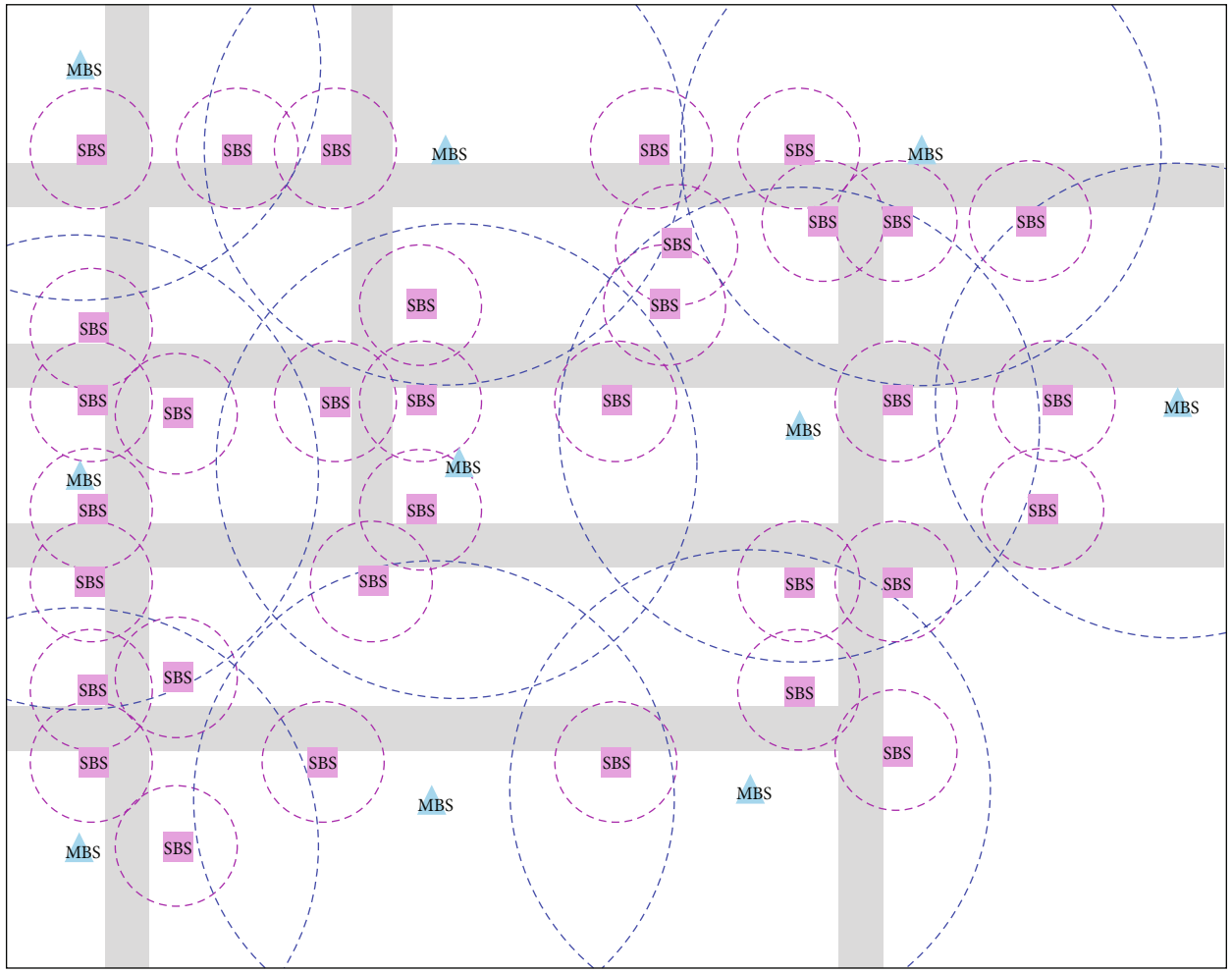

Figure 2: The deployment scenario of MBS and SBS in 5G UDN. The overlapped coverage of MBS and SBS is obvious. The coverage area of MBS is bigger than SBS.

3.3. Movement Model of Users. Figure 3 shows that the simulated scenario of smart city has multiple crossing roads, and many users move randomly. The MBS and SBS deploy in the both sides of roads, which provide wireless network access services, communication services, and data transmission with the covered users. In this city, there are N mobile users which appear randomly in different initial points and move at a constant speed along one road. The users' speed includes low speed, intermediate speed, and high speed which express the walk, bicycle riding, and drive scenes, respectively. Moreover, the users' number also has several values expressed the different user scenarios.

3.4. Problem Formulation and Algorithm Elements. In this paper, the handover decision problem in 5G UDN is formalized as a discrete Markov decision process, expressed with <S, A, and R>. And the parameters S and A are the state space and action space. The reward function is $r : S \times A -> R$. In the time slot t, s_t, s_{t+1}, a_t, and r_t are the network state, agent action, and immediate reward in handover decision process, respectively. The optimal candidate cells provide mobile users with better communication services. The research object of handover decision in this paper maximizes the long-term cumulative rewards. The discounted reward G_t in the interactions between agent and environment is defined as follows:

$$G_t = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}, \tag{4}$$

where R_t is the immediate reward in time slot t. The parameter γ is the discount coefficient of future reward. The action value function $Q(s_t, a_t)$ in the optimal Bellman operator is defined as follows:

$$Q^*(s_t, a_t) = E\left[R_{t+1} + \gamma \bullet \max_{a_{t+1}} Q^*(s_{t+1}, a_{t+1})\right], \tag{5}$$

where s_{t+1} is the network state in time slot $t + 1$. The maximum of $Q(s_{t+1}, a_{t+1})$ function is searched. The state space, action space, and reward function are defined as below, respectively.

3.4.1. State Space. In 5G UDN, the network state is obtained by state aware method. The state aware sequence consists of

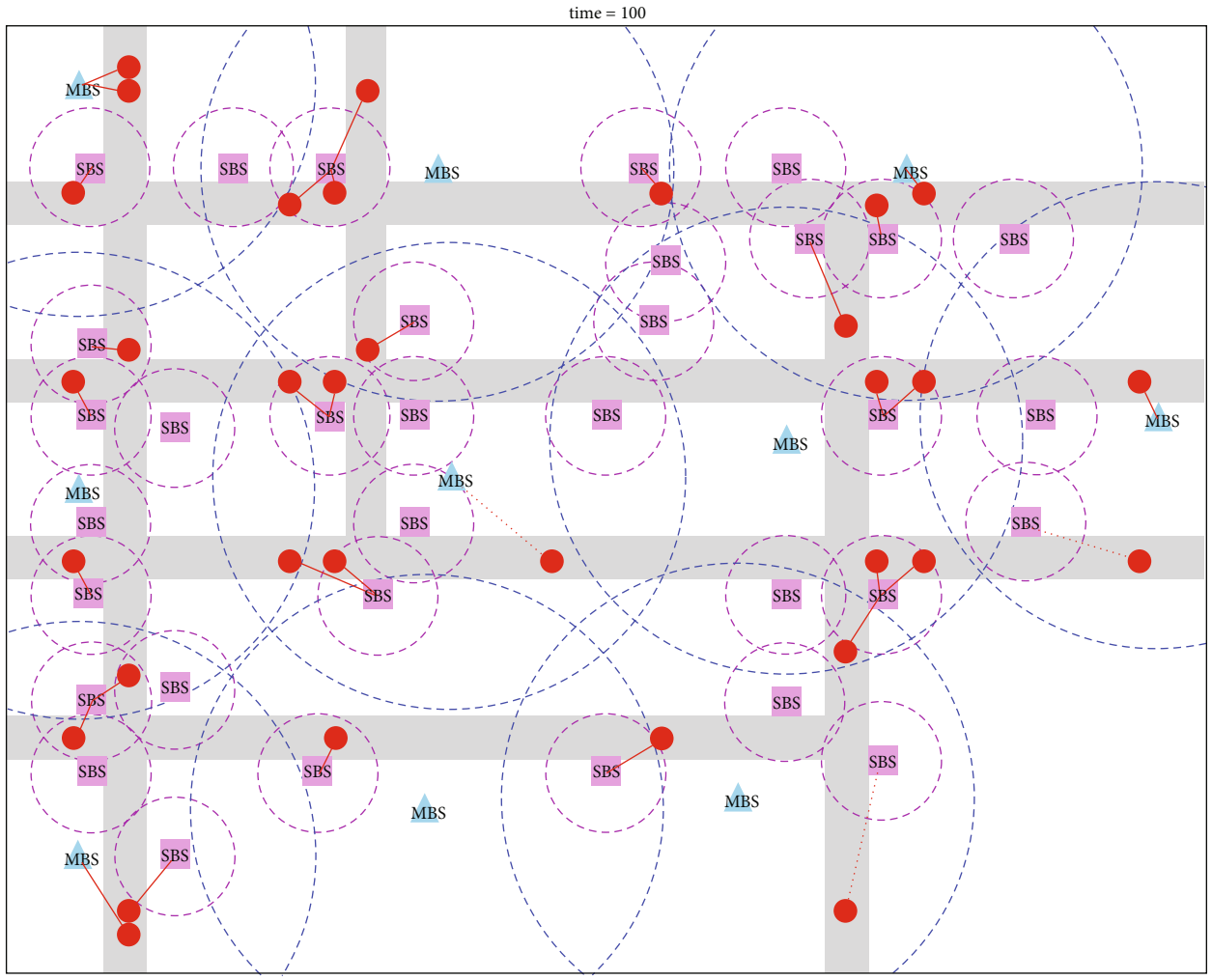

Figure 3: The handover scenario of mobile users in 5G UDN in 100 seconds. According to the SA-PER handover decision method, the mobile users select the candidate cell to connect.

SINR, dwell time *D*time, and load coefficient Load, expressed as $s_t = <\text{SINR}, D\text{time}, \text{Load}>$. Each time slot t, the state information of mobile users is updated in real time. The load coefficient is computed by Equation (14), and the load message is sharing by the public service interface $X2$ in base station. The dwell time *D*time is obtained by Equation (11) which is defined in Section 3.1.

3.4.2. Action Space. In network time slot t, the user selects a_t as the candidate cell to handover. The candidate cell index set in UDN is expressed with $A = \{0, 1, 2, \cdots, 42, 43\}$. The index 0 to 9 is macro cell, and others are small cell. Each time slot t, mobile users make a handover decision. If the handover is needed, the optimal candidate cell is determined.

3.4.3. Reward Function. The value of reward function is the immediate reward of action a_t. The reward function consisted of three decision factors is defined as follows:

$$R_t = \sum w_k \bullet c'_{t,i,k}, \tag{6}$$

where the parameter R_t is the immediate rewards in time slot t. The parameter w_k is the weight of network state factors which is produced by the AHP method, $k = 3$. The network state factors are the decision factors included SINR, *D*time, and Load. The parameter $c'_{t,i,k}$ is the normalized value of network state k in cell i in time slot t. The adopted normalization operation is the max-min normalization which is described in [29].

4. The State Aware-Based Prioritized Experience Replay Handover Decision Method

4.1. Analysis of Dwell Time in Cellular. According to the coverage area of heterogeneous cells, coordinates, and speed of mobile users, the dwell time in cell is computed [6]. Because the dwell time *D*time of mobile user is also a decision factor. The optimal candidate cell provided maximum dwell time is determined. In SA-PER handover decision method, a small amount of network performance is sacrificed. It is assumed that the mobile users move along the x-axis or y-axis in

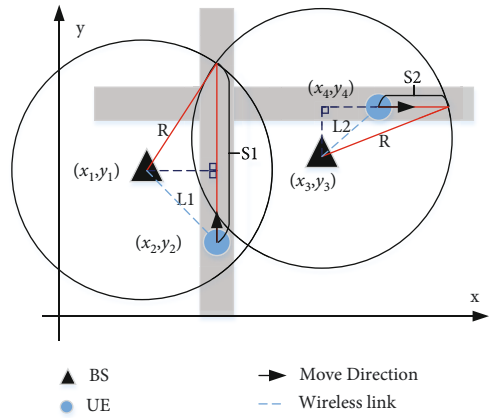

Figure 4: The dwell time for mobile users is analysed in 5G UDN. In the rectangular coordinate system, using the coordinates of mobile users and base station in cellular, the specific movement direction and dwell time are computed.

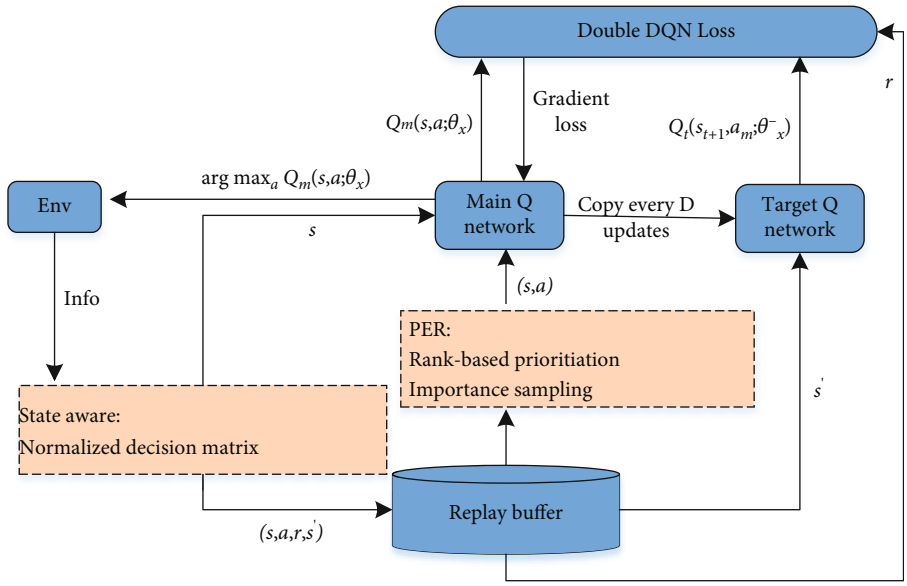

Figure 5: The framework of the proposed SA-PER handover decision method. The state aware method assists the handover decision, and the prioritized experience replay method improves the learning efficiency and accuracy.

the rectangular coordinate system as shown in Figure 4. The coordinates and speed of mobile users are collected in real time. Based on these collected state aware data, the dwell time of users in cell is computed. When users moving in the positive direction of the x-axis, the dis is

$$\text{dis} = \begin{cases} |y_1 - y_2| + \sqrt{R^2 - (x_2 - x_1)^2}, & y_1 \geq y_2 \\ \sqrt{R^2 - (x_2 - x_1)^2} - |y_2 - y_1|, & y_1 < y_2 \end{cases}, \tag{7}$$

where the parameter R is the communication radius of cell. The coordinates (x_1, y_1) and (x_3, y_3) are the locations of base station of cell. The coordinates (x_2, y_2) and (x_4, y_4) are the locations of mobile users. When users moving in the negative direction of the x-axis, the dis is

$$\text{dis} = \begin{cases} \sqrt{R^2 - (x_2 - x_1)^2} - |y_1 - y_2|, & y_1 \geq y_2 \\ \sqrt{R^2 - (x_2 - x_1)^2} + |y_2 - y_1|, & y_1 < y_2 \end{cases}. \tag{8}$$

When users moving in the positive direction of the y -axis, the dis is

$$\text{dis} = \begin{cases} |x_3 - x_4| + \sqrt{R^2 - (y_4 - y_3)^2}, & x_3 \geq x_4 \\ \sqrt{R^2 - (y_4 - y_3)^2} - |x_4 - x_3|, & x_3 < x_4 \end{cases}. \tag{9}$$

When users moving in the negative direction of the y

Input: Iteration number *NUM_EPISODES*, step number *MAX_STEPS*, node number *node_num*, measurement information *SINR*, 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. $\theta_x^- = \theta_x$.
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. (6), the immediate reward r_t is computed.
6: According to Eq. (11), the dwell time is computed. According to Eq. (14), the load coefficient *Load* is obtained. By the state aware method, the network state s_t in time slot t is constructed. According to Eq. (16, 17), the state decision matrix M_s is normalized.
7: By the ε-greedy method, the action a_t corresponding to state s_t is determined and the handover decision matrix $\boldsymbol{A}$ is updated.
8: The next state s_{t+1} is produced and the transition (s_t, a_t, r_t, s_{t+1}) is stored in buffer B.
9: In PER method, according to Eq. (18, 19), the priority and probability of sample are computed. According to Eq. (20), the weight of importance sampling method is computed. The sampling data is the input of main-Q network, and the action-value function $Q_m(s_t,a_t)$ is computed.
10: According to Eq. (22), 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 $Q_t(s_{t+1}, a_m)$ is computed.
11: Adopt the stochastic gradient descent method, according to Eq. (24), the parameters θ_x 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. $\theta_x^- = \theta_x$.
14: **end for**
15: **end for**
16: Return the handover decision matrix $\boldsymbol{A}$.

Algorithm 1: SA-PER handover decision algorithm.

Table 3: Simulation parameters of the network.

Parameters	Macro cell		Small cell
Total number of cell	10		34
Cell radius	500 m		50 m
Carrier frequency	2 GHz		28 GHz
System bandwidth	20 MHz		500 MHz
RB's bandwidth	180 kHz		1.75 MHz
Number of RBs	100		275
Thermal noise		-174 dBm/Hz	
Shadowing	7.8 dB		8.2 dB
Antenna gain	15 dBi		5 dBi
Cell transmit power	46 dBm		35 dBm
Path loss model	$32.4 + 20 \lg (f) + 30 \lg (d) + \chi$		$32.4 + 20 \lg (f) + 31.9\lg(d) + \chi$
Number of users		50, 100, 200, 300	
Speed of UE (km/h)		5, 25, 50, 70, 120	
Duration of simulation		600 seconds	
Sampling interval		0.1 second	

-axis, the dis is

$$\text{dis} = \begin{cases} \sqrt{R^2 - (y_4 - y_3)^2} - |x_3 - x_4|, & x_3 \geq x_4 \\ \sqrt{R^2 - (y_4 - y_3)^2} + |x_4 - x_3|, & x_3 < x_4 \end{cases}. \quad (10)$$

The dwell time $D\text{time}_{i,j}$ is computed by:

$$D\text{time}_{i,j} = \frac{\text{dis}_i}{v_j}, \quad (11)$$

where the parameter dis_i is the movement distance of user in

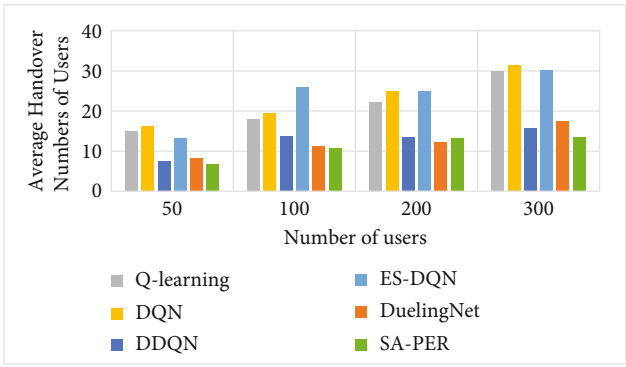

Figure 6: The average handover numbers of users with different methods. And the number of users affects the handover performance.

cell i. And v_j is the speed of user j. The average dwell time of mobile users Mtime is defined as follows:

$$M\text{time} = \sum_{j=0}^{N} \left(\frac{\sum_{i=0}^{M+S} D\text{time}_{i,j}}{HO\_\text{num}_j * BS\_\text{num}_j} \right) / N, \tag{12}$$

where HO_numl_j and BS_num_j are the total handover numbers and total connected cell numbers of user j, respectively. And M, S, and N are the total number of macro cells, small cells, and users, respectively.

4.2. State Aware Decision Matrix. In the state aware decision matrix, the state aware sequence is a vital input, which includes SINR, Dtime, and Load. SINR is the signal to interference plus noise ratio, which expresses the signal quality of BS. Dtime is the dwell time of UE in cellular, which expresses the connection time of UE-BS. Load is the load coefficient, which expresses the load condition of BS. In handover measurement procedure [42], when the neighbor cell's signal becomes stronger than serving cell's signal, the measurement is trigged. The serving cell sends the measurement control message to UE. In the measurement period, the UE measure the signal quality of cells in neighbor cell list (NCL). The SINR expresses the signal quality of cells, which is collected. Dtime is computed in Section 4.1, which needs the real-time position and velocity of UE. The real-time position and velocity of UE are the application layer information and collected by data collection coordinated function which is mentioned in 3GPP TR 23.700-91V17.0.0. And the public interface $X2$ shares the load information of each base station. By using state aware method, the state data of network, cell, and user is collected. Therefore, the network state aware sequence is defined as follows:

$$s_t = <\text{SINR}, D\text{time}, \text{Load}>, \tag{13}$$

where the parameter Dtime is obtained by Equation (11). The parameter Load is the load coefficient of cell.

$$\text{Load}_{i,t} = \frac{UE\text{num}_{i,t}}{T\text{num}_i}, \tag{14}$$

where $T\text{num}_i$ is the total number of subchannel in cell i. The parameter $UE\text{num}_{i,t}$ is the number of connected users in cell i in time slot t. The state decision matrix is defined as follows:

$$M_s = \begin{bmatrix} \text{SINR}_1 & D\text{time}_1 & \text{Load}_1 \\ \text{SINR}_2 & D\text{time}_2 & \text{Load}_2 \\ \vdots & \vdots & \vdots \\ \text{SINR}_L & D\text{time}_L & \text{Load}_L \end{bmatrix}, \tag{15}$$

where the parameter $L = M + S$ is the total number of cells. The parameter M_s contains the SINR, Dtime, and Load state data of every cells. The max-min normalization operation of state decision matrix is defined as follows:

$$c'_{t,i,k} = \begin{cases} \dfrac{c_{t,i,k} - \min(c_{t,k})}{\max(c_{t,k}) - \min(c_{t,k})}, & c_i \in \text{SINR or } D\text{time} \\ \dfrac{\max(c_{t,k}) - c_{t,i,k}}{\max(c_{t,k}) - \min(c_{t,k})}, & c_i \in \text{Load} \end{cases}. \tag{16}$$

The normalized state decision matrix is

$$M'_s = \begin{bmatrix} \text{SINR}'_1 & D\text{time}'_1 & \text{Load}'_1 \\ \text{SINR}'_2 & D\text{time}'_2 & \text{Load}'_2 \\ \vdots & \vdots & \vdots \\ \text{SINR}'_L & D\text{time}'_L & \text{Load}'_L \end{bmatrix}. \tag{17}$$

4.3. The Prioritized Experience Replay Based on DDQN Method. By the state aware method and normalization operation, the normalized state decision matrix is obtained which assists the handover decision. Combining with state aware method, the proposed SA-PER handover decision method adopts rank-based prioritization and importance sampling, which make sure of the learning efficiency and convergence of algorithm. The rank-based prioritization method computes the priority p_x of sample x.

$$p_x = \frac{1}{\text{rank}(x)}, \tag{18}$$

where the function rank (x) produces the order of sample x in experience buffer. The order of sample x is determined by its own absolute value of TD error. The probability of sample x is $P(x)$.

$$P(x) = \frac{p_x}{\sum_x p_x}. \tag{19}$$

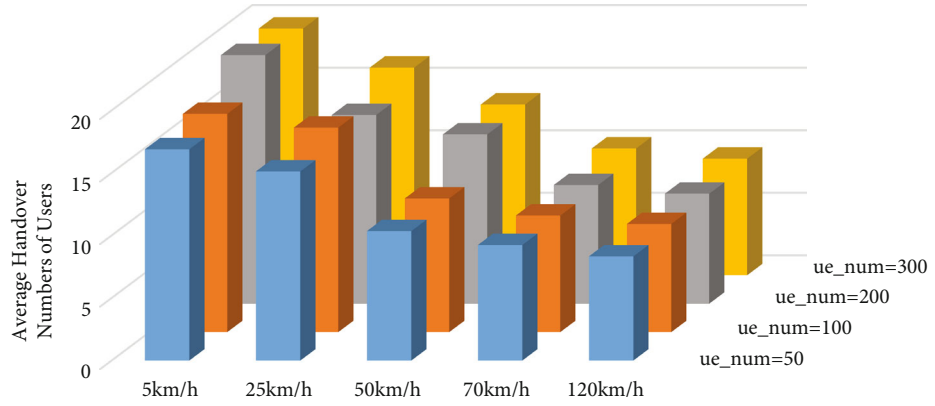

Figure 7: The average handover numbers of SA-PER method with different speeds and numbers of users. These two factors affect the handover performance.

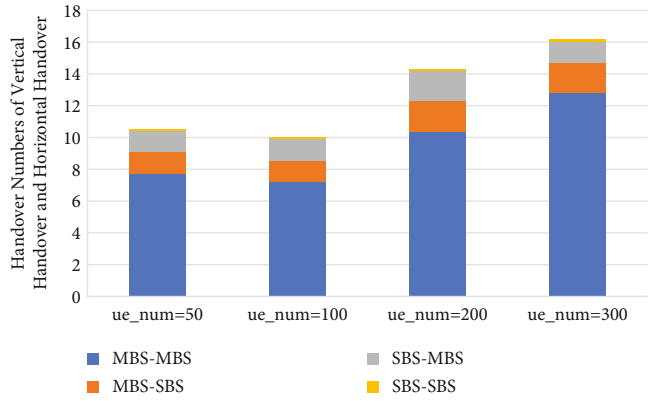

Figure 8: The average handover numbers of vertical handover and horizontal handover with different users' number. The number of horizontal handover is bigger.

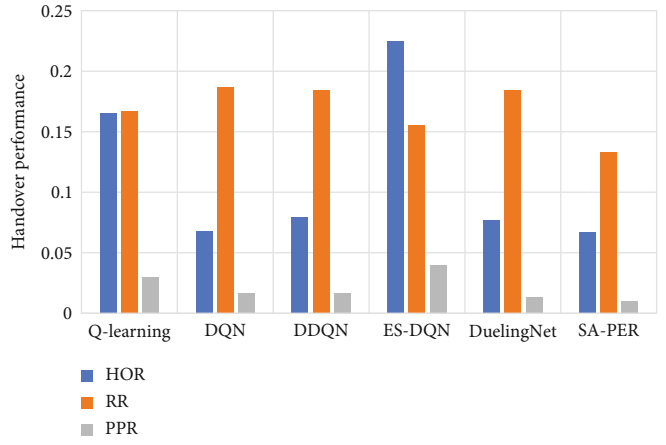

Figure 9: The handover rate, radio link failure rate, and ping-pong rate of different handover decision methods with the *ue*_num = 100.

The $P(x)$ is a ratio. For the stable distribution of sampling data, the weight coefficient of importance sampling is defined as

$$\omega_x = (C \bullet P(x))^{-\beta}, \tag{20}$$

where the parameter C is the total number of samples in buffer. The parameter $\beta = 0.4$ is a hyperparameter obtained from experiments. In the training process of handover decision, the normalized state decision matrix is the input of the Q-network, and the optimal value of the action-value function is output.

$$Q(s_t, a_t; \theta) \approx Q^*(s_t, a_t). \tag{21}$$

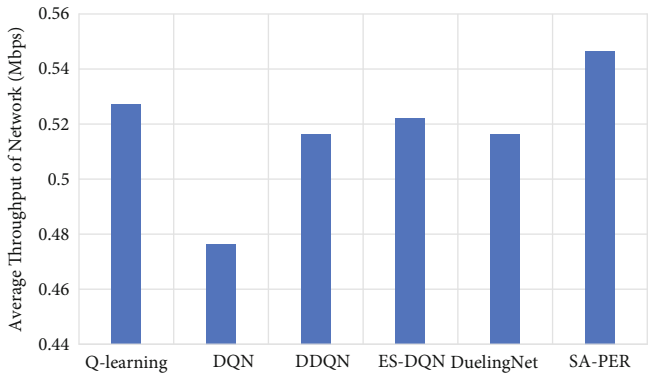

FIGURE 10: The average throughput of network for different handover decision methods with the *ue*_num = 100.

When the maximum value of Q_m is obtained, the corresponding handover action a_m is determined. The update of action-value function in DDQN method is defined as

$$a_m = \arg \max_a Q_m(s_{t+1}, a),$$
$$Q_m(s_t, a_t) = Q_m(s_t, a_t) + \eta * (R_{t+1} + \gamma \bullet Q_t(s_{t+1}, a_m) - Q_m(s_t, a_t)). \tag{22}$$

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

$$y = \begin{cases} R_{t+1}, & \text{if } s_{t+1} \text{ is end} \\ R_{t+1} + \gamma \bullet Q_t(s_{t+1}, a_m; \theta_x^-), & \text{others} \end{cases}, \tag{23}$$
$$L_x(\theta_x) = (y - Q_m(s_t, a_t; \theta_x))^2.$$

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 parameters, the value of loss function decreases. And the performance of handover becomes better. The loss function of DDQN method is optimized by the stochastic gradient descent method. The gradient of loss function is defined as

$$\nabla_{\theta_x} L(\theta_x) = \omega_x L_x(\theta_x) \nabla_{\theta_x} Q_m(s_t, a_t; \theta_x). \tag{24}$$

In Figure 5, the framework of the state aware-based prioritized experience replay method is illustrated clearly. In network environment, the necessary information and data collected by UE periodically input the state aware method. The obtained state decision matrix is normalized. Then, the obtained current state aware sequence $s = \{\text{SINR}, D \text{ time}, \text{Load}\}$, action a, reward r, and next state s' are stored in the replay buffer. The state aware method also sends the normalized state s to the main Q-network for the optimal action a which is determined and send to the network environment. The replay buffer provides transition (s, a), next state s', and reward r to the prioritized experience replay, target Q-network, and loss function, respectively. The prioritized experience replay includes the rank-based prioritization and importance sampling methods. The important samples usually have the big absolute value of TD error. These important samples came from the replay buffer are input the main Q-network. Different from the traditional DDQN method, the random sampling mechanism or mini-batch sampling method is improved by prioritized experience replay method. The basic DDQN method still includes the main Q-network and target Q-network which are used to determine the optimal action a_m and evaluate the Q value of a_m, respectively. Every D episodes, the network coefficients of target Q-network are updated by main Q-network. The main Q-network sends the $Q(s, a)$ to the loss function and get the corresponding gradient loss. At the same time, the target Q-network shares the $Q(s', a_m)$ with the loss function. By the state aware method and analysis of dwell time, the performance fluctuation of weighted multiattribute decision method is improved. The adopted prioritized experience replay method improves the performance of handover, the learning efficiency, and convergence speed.

5. Experimental Results and Discussions

5.1. Simulation Environment Setups. The targets of this research are to solve the frequent handover and communication interrupt. A PC carries out the simulation experiments with 3.2 GHz quad-core i5-1570 and 16 GB of RAM. The OS is win 10, 64 bits, and the simulation platform is Python 3. The simulated scenario of virtual city is shown as Figure 3. The width and length of simulated area in city are 2.5 kilometres and 2 kilometres. This scenario includes 7 roads, and the buildings, hills, rivers, and so on are unmarked. It contains 10 macro cells and 34 small cells. These base stations are deployed along the roads to cover as much area as possible. Note that the overlapping coverage is also evident. The movement model of UE is described as Section 3.3. The starting point of mobile user is randomly selected from 11 initial points. The speed of 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 300, respectively. The simulation environment of wireless heterogeneous cellular networks is realized by Python. In this simulation, the system bandwidth of macro cell and small cell is set to 20 MHz and 500 MHz, respectively. The wireless channels of macro cell and micro cell are modelled reference the TR 38.901 V16.1.0. The standard deviations of shadow fading are 7.8 dB and 8.2 dB, respectively. For the handover settings, TTT and A3 offset are set as 450 ms and 3 dB. If the SINR is below -3 dB for 500 ms, then the radio link is considered to have failed. The communication radius of macro cell and small cell is 500 meters and 50 meters, respectively. And the upper limits of connected users are 100 and 275, respectively. One user only occupies up to one resource block, and the bandwidth of subchannel in macro cell and small cell is 180 kHz and 1.75 MHz, respectively [43].

The handover rate (HOR), radio link failure rate (RLF rate), and ping-pong rate (PPR) are selected as evaluation

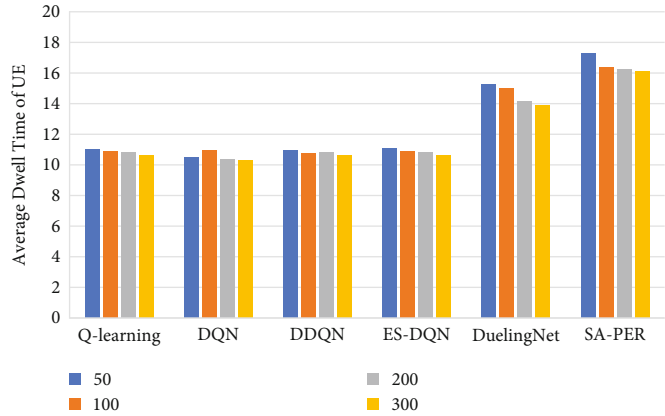

Figure 11: The average dwell time of users for different handover decision methods with different numbers of users. The SA-PER method has an excellent performance.

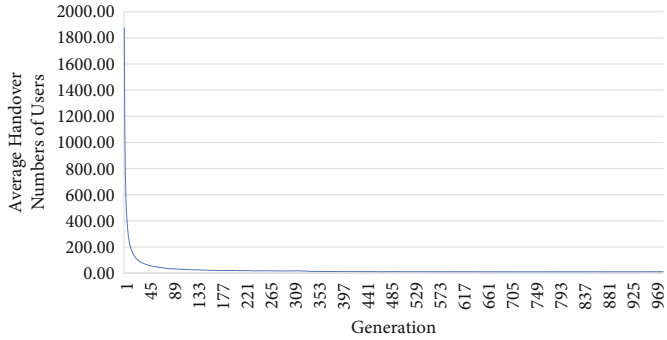

Figure 12: The convergence condition of SA-PER method when the number of user is 100.

criterions.

$$\begin{aligned} \mathrm{HOR} &= \frac{N_{HO}}{N_{\mathrm{total}}}, \\ \mathrm{RR} &= \frac{N_{\mathrm{RLF}}}{N_{\mathrm{total}}}, \\ \mathrm{PPR} &= \frac{N_{PP}}{N_{\mathrm{total}}}, \end{aligned} \tag{25}$$

where N_{HO} is the number of successive handover, N_{RLF} is the number of RLF, N_{pp} is the number of ping-pong, and N_{total} is the number of handover requests. The value of HOR, RR, and PPR is between [0, 1]. According to Reference [7, 44], the parameters of 5G UDN are determined. To compare the proposed method, several previous popular handover decision methods are considered: Q-learning [29], DQN [34], DDQN [45], ES-DQN [35], and DuelingNet [36] handover decision methods.

Reference to [39, 41], the simulation parameters of the network are show as Table 3

5.2. Analysis and Discussion of Experimental Results

5.2.1. Average Handover Numbers of UE. Figure 6 shows the average handover numbers of different handover decision methods while the numbers of users are 50, 100, 200, and 300, respectively. When the number of users increases, the handover numbers increase. And the proposed SA-PER handover decision method has the excellent performance, and the performance of DuelingNet method is much closed. When a number of users are 50, 100, 200, and 300, the average handover numbers of SA-PER are 6.82, 10.76, 13.12, and 13.36, respectively.

In the proposed SA-PER method, the state aware method makes full use of the state aware data and provides the decision basis for the handover decision. Moreover, the PER method improves the sampling method, and the learning efficiency and accuracy of DRL algorithm are optimized. In the DDQN method, the main Q-network trains the network coefficients, and the target Q-network updates Q-network. The learning performance of DDQN method is better than the traditional DQN method. Based on DDQN, the DuelingNet method updates the network structure and improves the learning ability. According to the comparative analysis, we found that the proposed SA-PER handover decision method solved the frequent handover problem. And the average handover numbers decreased obviously, which meets the communication demands of mobile users.

Figure 7 shows the average handover numbers of SA-PER method with different speeds and numbers of users. When the number of user is fixed, the increase of user speed leads to the decrease of handover numbers. This is because that when the user speed is bigger, the number of sampling

is smaller, and the number of handover request is smaller. When the user speed is fixed, the increase of users' number leads to the increase of average handover number, because the load coefficient is one handover decision factor. In the process of users' movement, the mobile users prefer to connect the candidate cell which has a low load coefficient.

Figure 8 shows the vertical handover (MBS-SBS) and horizontal handover (MBS-MBS and SBS-SBS) performance of SA-PER method with different numbers of users. With the increase of users' number, the total handover numbers are increased. Because the increase of users' number affects the load of cell directly, in the SA-PER method, the number of vertical handover is smaller than horizontal handover. This is because that in the ultradense deployment of small cells, the overlapped coverage between macro cell and small cell is obvious. In the handover decision process, the macro cell is mostly selected as the candidate cell. This is because that the dwell time is also one decision factor. When the dwell time is longer, the handover number is smaller. The total handover numbers of vertical handover change a little. When the coverage of cellular network is poor, the mobile user only connects MBS or SBS. The collaborative relationship between horizontal handover and vertical handover is dominated. When the coverage of cellular network is good, the candidate cellular set is big. The competitive relationship between horizontal handover and vertical handover is dominated. When the speed of UE increases, the UE selects the macro cell to handover, which has the long dwell time. Our research analyses the relations between vertical handover and horizontal handover, which provides good preparations for the real deployment and increases the successive handover rate.

5.2.2. Handover Rate, Radio Link Failure Rate, and Ping-Pong Rate. Figure 9 shows the average value of the handover rate, radio link failure rate, and ping-pong rate of different handover decision methods with the $ue\_num = 100$.

When the values of HOR, RR, and PPR are smaller, the performance of handover decision method is better. Because of the random motion of UE, the N_{total} is different for the different handover decision methods. The HOF, RR, and PPR of the proposed method are 0.066, 0.133, and 0.009, respectively. The SA-PER outperforms other selected methods. By the analysis of dwell time and PER, the average handover number is minimum. The evolution strategy of ES-DQN method initializes the deep neural network and produces some unnecessary handovers. The number of ping-pong effect is less than the total number of handover, which explains the smaller value of PPR than HOR. The increase of handover requests leads to the increase of radio link failure. Therefore, the RR of DQN, DDQN, and DuelingNet increase a little.

5.2.3. The Throughput of Networks. Figure 10 shows the average throughput of network for different handover decision methods while the number of user is 100. In comparison, the proposed SA-PER handover decision method has a higher throughput 0.5465 Mbps. The performance of network throughput for Q-learning method is in the second place. Because the Q-learning method usually used in the discrete problems not the continuity problems, the state aware and PER method optimize the data collection and batch sampling. Therefore, the proposed method meets the demands of communication services for the mobile users.

5.2.4. Average Dwell Time of User. The average dwell time of different handover decision methods with different numbers of users is shown in Figure 11. When the number of users increases, the average dwell time decreases. And the SA-PER method has a longer dwell time than others. Because the state aware and PER method improve the learning efficiency and accuracy, according to Equation (12), when the total dwell time is fixed, the decrease of handover number and connected cell number leads to the increase of dwell time. The proposed SA-PER method has the longest dwell time which means the lower handover numbers. And this proposed method meets the demand of communication continuity for mobile users.

5.2.5. The Convergence of SA-PER Method. Figure 12 shows the convergence condition of SA-PER method when the number of user is 100. The average handover numbers correspond to each generation. In the proposed SA-PER method, the coefficients of Q-network have the random initial parameters, which leads to a high handover number. With the training process, the handover performance of our method becomes stable, and the handover number becomes small. When the number of generation is 100, the convergence of our method is obvious, and the handover number is 30.54. When the number of generation increases to 1000, the minimum handover number is 8.88. The proposed method has a good handover performance and improves the efficiency of handover management.

6. Conclusions

In this research, the proposed SA-PER handover decision method reduced the frequent handover and ping-pong effect in 5G ultradense networks. The quality and continuity of communication services are upgraded and improved. The state aware method and the analysis of cell dwell time reduced the frequent handover and ping-pong effect. The prioritized experience replay method improved the learning efficiency and convergence rate of DDQN-based handover decision method. The analysis of competitive and collaborative relationships between different handovers helps the network operators balance the resource efficiency and QoS. In addition, by means of the decision ability of DDQN method, the online learning of handover decision is more adapted to the dynamics of networks and mobility of users.

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] M. I. Rochman, V. Sathya, N. Nunez et al., "A comparison study of cellular deployments in Chicago and Miami using apps on smartphones," in *Proceedings of the 15th ACM Workshop on Wireless Network Testbeds, Experimental evaluation & CHaracterization*, pp. 61–68, New Orleans, LA, USA, 2022.

[2] S. Khosravi, H. S. Ghadikolaei, and M. Petrova, "Learning-based load balancing handover in mobile millimeter wave networks," in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, pp. 1–7, Taipei, Taiwan, 2020.

[3] V. Sathya, "Evolution of small cell from 4G to 6G: past, present, and future," https://arxiv.org/abs/2101.10451.

[4] 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.

[5] Y. Z. H. Wang, X. Yang, and C. Wei, "METRE measurement task recommendation for energy-efficient handover in dense networks," in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, pp. 1–6, Taipei, Taiwan, 2020.

[6] L. W. W. Sun, J. Liu, N. Kato, and Y. Zhang, "Movement aware CoMP handover in heterogeneous ultra-dense networks," *IEEE Transactions on Communications*, vol. 69, no. 1, pp. 340–352, 2021.

[7] Q. Liu, C. F. Kwong, S. Wei, L. Li, and S. Zhang, "Intelligent handover triggering mechanism in 5G ultra-dense networks via clustering-based reinforcement learning," *Mobile Networks and Applications*, vol. 26, pp. 27–39, 2021.

[8] M. Cicioğlu, "Multi-criteria handover management using entropy-based SAW method for SDN-based 5G small cells," *Wireless Networks*, vol. 27, no. 4, pp. 2947–2959, 2021.

[9] 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.

[10] D. Xenakis, N. Passas, L. Merakos, and C. Verikoukis, "Mobility management for femtocells in LTE-advanced: key aspects and survey of handover decision algorithms," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 1, pp. 64–91, 2014.

[11] 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.

[12] V. Sharma, J. Guan, J. Kim et al., "MIH-SPFP: MIH-based secure cross-layer handover protocol for Fast Proxy Mobile IPv6-IoT networks," *Journal of Network and Computer Applications*, vol. 125, pp. 67–81, 2019.

[13] S.-M. Chun and J.-T. Park, "Mobile CoAP for IoT mobility management," in *2015 12th Annual IEEE Consumer Communications and Networking Conference (CCNC)*, pp. 283–289, Las Vegas, NV, USA, 2015.

[14] D. Wu, D. I. Arkhipov, E. Asmare, Z. Qin, and J. A. McCann, "UbiFlow: Mobility Management in Urban-Scale Software Defined IoT," in *2015 IEEE Conference on Computer Communications (INFOCOM)*, pp. 208–216, Hong Kong, China, 2015.

[15] 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.

[16] Z. H. Huang, Y. L. Hsu, P. -K. Chang, and M. -J. Tsai, "Efficient handover algorithm in 5G networks using deep learning," in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, pp. 1–6, Taipei, Taiwan, 2020.

[17] H. Tabrizi, G. Farhadi, and J. Cioffi, "Dynamic handoff decision in heterogeneous wireless systems_ Q-learning approach," *2012 IEEE International Conference on Communications (ICC)*, 2012, pp. 3217–3222, Ottawa, ON, 2012.

[18] 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.

[19] 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.

[20] X. Hu, H. Song, S. Liu, and W. Wang, "Velocity-aware handover prediction in LEO satellite communication networks," *International Journal of Satellite Communications and Networking*, vol. 36, no. 6, pp. 451–459, 2018.

[21] 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.

[22] A. Stamou, N. Dimitriou, K. Kontovasilis, and S. Papavassiliou, "Context-aware handover management for HetNets: performance evaluation models and comparative assessment of alternative context acquisition strategies," *Computer Networks*, vol. 176, article 107272, 2020.

[23] F. Guidolin, I. Pappalardo, A. Zanella, and M. Zorzi, "Context-aware handover policies in HetNets," *IEEE Transactions on Wireless Communications*, vol. 15, no. 3, pp. 1895–1906, 2016.

[24] J. Liu, X. Tao, and J. Lu, "Mobility-aware centralized reinforcement learning for dynamic resource allocation in HetNets," in *2019 IEEE global communications conference (GLOBECOM)*, pp. 1–6, Waikoloa, HI, USA, 2019.

[25] S. He, T. Wang, and S. Wang, "Load-aware satellite handover strategy based on multi-agent reinforcement learning," in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, pp. 1–6, Taipei, Taiwan, 2020.

[26] M. Salehi and E. Hossain, "Handover rate and sojourn time analysis in mobile drone-assisted cellular networks," *IEEE Wireless Communications Letters*, vol. 10, no. 2, pp. 392–395, 2021.

[27] W. Nasrin and J. Xie, "A joint handoff and offloading decision algorithm for mobile edge computing," in *2019 IEEE Global Communications Conference (GLOBECOM),*, pp. 1–6, Waikoloa, HI, USA, 2019.

[28] M. M. Hasan, S. Kwon, and S. Oh, "Frequent-handover mitigation in ultra-dense heterogeneous networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 1, pp. 1035–1040, 2019.

[29] J. Chen, X. Ge, and Q. Ni, "Coverage and handoff analysis of 5G fractal small cell networks," *IEEE Transactions on Wireless Communications*, vol. 18, no. 2, pp. 1263–1276, 2019.

[30] B. Yang, X. Wang, and Z. Qian, "A multi-armed bandit model-based vertical handoff algorithm for heterogeneous wireless networks," *IEEE Communications Letters*, vol. 22, no. 10, pp. 2116–2119, 2018.

[31] L. Sun, J. Hou, and T. Shu, "Optimal handover policy for mmWave cellular networks a multi-armed bandit approach," in *2019 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, Waikoloa, HI, USA, 2019.

[32] D.-F. Wu, C. Huang, Y. Yin et al., "LB-DDQN for handover decision in satellite-terrestrial integrated networks," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 5871114, 2021.

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

[34] M. Wu, W. Huang, K. Sun, and H. Zhang, "A DQN-based handover management for SDN-enabled ultra-dense networks," in *2020 IEEE 92nd Vehicular Technology Conference (VTC2020-Fall),*, pp. 1–6, Victoria, BC, Canada, 2020.

[35] 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.

[36] Z. Wang, T. Schaul, M. Hessel, H. Hasselt, M. Lanctot, and N. Freitas, "Dueling network architectures for deep reinforcement learning," in *Proceedings of The 33rd International Conference on Machine Learning, Proceedings of Machine Learning Research*, pp. 1995–2003, New York, USA, 2016.

[37] M. S. Mollel, A. I. Abubakar, M. Ozturk et al., "Intelligent handover decision scheme using double deep reinforcement learning," *Physical Communication*, vol. 42, article 101133, 2020.

[38] Y. Song, S. H. Lim, and S. W. Jeon, "Distributed online handover decisions for energy efficiency in dense HetNets," in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, pp. 1–6, Taipei, Taiwan, 2020.

[39] 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.

[40] L. L. Z. Wang, Y. Xu, H. Tian, and S. Cui, "Handover optimization via asynchronous multi-user deep reinforcement learning," in *2018 IEEE International Conference on Communications (ICC)*, pp. 1–6, Kansas City, MO, 2018.

[41] D. Guo, L. Tang, X. Zhang, and Y.-C. Liang, "Joint optimization of handover control and power allocation based on multi-agent deep reinforcement learning," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 11, pp. 13124–13138, 2020.

[42] V. M. Nguyen, C. S. Chen, and L. Thomas, "A unified stochastic model of handover measurement in mobile networks," *IEEE/ACM Transactions on Networking*, vol. 22, no. 5, pp. 1559–1576, 2014.

[43] M. T. Nguyen, S. Kwon, and H. Kim, "Mobility robustness optimization for handover failure reduction in LTE small-cell networks," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 5, pp. 4672–4676, 2018.

[44] 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.

[45] J. Cai, C. Wang, M. Lei, and M. J. Zhao, "An intelligent routing algorithm based on prioritized replay double DQN for MANET," in *2020 IEEE 92nd Vehicular Technology Conference (VTC2020-Fall)*, pp. 1–5, Victoria, BC, Canada.