Proactive Handover Type Prediction and Parameter Optimization Based on Machine Learning

Kai Sun, *Member, IEEE,* Qingfeng Han, Zongchang Yang, Wei Huang, *Member, IEEE,* Haijun Zhang, *Fellow, IEEE,* and Victor C. M. Leung, *Life Fellow, IEEE*

***Abstract*—With the explosive growth of smart devices and applications, the demand for mobile service with higher data rate and better quality of service is growing rapidly. Ultra-dense networks, capable of providing higher network throughput, remain one of the key technologies for next-generation mobile communications. However, the densification of network further reduces the coverage of base stations and the distance between each other, which in turn leads to unnecessary and frequent handovers (HOs), affects the stability and reliability of communication links. HO failures can even occur due to the improper HO control parameter (HCP) values. To this end, a HO type prediction and parameter optimization method based on machine learning is proposed. First, the HO is divided into four categories: successful handover (SHO), ping-pong handover (PPHO), too-late handover (TLHO), and too-early HO (TEHO). Second, we combine reinforcement learning with supervised learning and propose a novel adaptive HCP adjusting scheme. Specifically, deep Q-network dynamically selects HCP values through environmental information and supervised learning-based HO prediction results. Simulation results demonstrate that our proposed scheme achieves a prediction accuracy of 94.83%, while reducing the PPHO rate by 15%, the TEHO rate by 2%, and the TLHO rate by 3%.**

***Index Terms*—Handover type prediction, handover control optimization, machine learning, ultra-dense network.**

I. Introduction

In order to support the massive demands for services and address the ever-growing application scenarios, deploying dense network nodes to achieve higher capacity and better coverage is an effective solution [1]. Seamless handovers (HOs) can achieve the connection quality and stability of mobile user equipments (UEs) in the wireless network, thereby ensuring the reliable communication experience for users. However, the densification of networks poses challenges to traditional HO methods. The reduced sojourn time of UEs in small cells results in a significant increase in HOs, which not only increases the signaling cost of the system, but also affects the quality of service (QoS) of users, especially for the delay-sensitive users. Compared to macro cells, the signal strength fluctuations within overlapping areas of small cells are more pronounced. When UE traverses numerous irregularly overlapping areas, unnecessary ping-pong handover (PPHO) may occur, even radio link failure (RLF) and/or HO failure (HOF) may occur. The configuration of HO control parameters (HCPs) has a significant impact on the quality and stability of wireless network connections. How to set up HCPs reasonably has been one of the research focuses of wireless mobile communication network optimization [2].

A. Related Work

Generally, time-to-trigger (TTT), HO margin (HOM), and cell-individual offsets (CIO) are considered as the primary control parameters for managing the HO process [3]. Typically, these parameter values are established using the gold standards recommended by equipment vendors. However, these standards do not take into account differences in network deployment scheme, base station (BS) density, QoS requirements, and user mobility patterns, resulting in increased HOF rate [4]. Improper HCP values will result in unnecessary PPHO, too-late HO (TLHO) and too-early HO (TEHO), which in turn increase network signaling load, introduce HO delays, contribute to network congestion and connection interruption [5]. HCP optimization can effectively enhance success HO (SHO) rate and improve the HO performance of the UEs. The authors in [6] modeled the wireless network as Apollonian circles in geometry, proving the existence of optimal HO parameters that can minimize the PPHO rate and RLF rate. For frequent HOs in heterogeneous networks, a closed-form HO rate expression based on random geometry was provided in [7], and the simulation results demonstrated that HO rate has nothing to do with the mobility patterns and only dependents on the movement distance. The inter-frequency HO based on A5 trigger event was investigated in [8], and the impact of HCPs (i.e., TTT, HOM and CIO) on three key performance indicators was specifically analyzed. The authors in [9] summarized HO issues in both deterministic and random networks, with a focus on the performance of mobile-aware HOs based on trajectory and association methods. The study emphasizes the importance of BS selection based on link reliability to enhance user SHO rate. [10] investigated the HO management in cell-free networks and proposed an access point clustering and pilot allocation scheme to solve the link congestion problem

This work was supported in part by the National Natural Science Foundation of China under Grants 62371264, 62161035, Natural Science Foundation of Inner Mongolia Autonomous Region under Grants 2022MS06022, 2023MS06015. *(Corresponding author: Wei Huang.)*

Kai Sun, Qingfeng Han, Zongchang Yang, and Wei Huang are with the School of Electronic Information Engineering, the Inner Mongolia Key Laboratory of Intelligent Communication and Sensing and Signal Processing, Inner Mongolia University, Hohhot 020021, China. (e-mails: sunkai@imu.edu.cn; hqf6800@gmail.com; yzconepiece@gmail.com; huangwei@imu.edu.cn;).

Haijun Zhang is with Beijing Engineering and Technology Research Center for Convergence Networks and Ubiquitous Services, the University of Science and Technology Beijing, Beijing 100083, China (e-mail: haijunzhang@ieee.org).

Victor C. M. Leung is with the Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, BC, V6T 1Z4, Canada (e-mail: vleung@ieee.org).

by considering factors such as user association, power control, mobility patterns, and computational complexity. In summary, the optimization of HCPs is closely related to several factors, including the user's movement distance, the type of HO event, and link reliability. These studies emphasize that by properly configuring HCPs, it is possible to effectively reduce PPHO and RLF, thereby enhancing the SHO rate and overall network performance.

Machine learning (ML) has emerged as a primary tool for addressing HO management issues, which can effectively reduce the complexity of the problem while enhancing link performance. On the one hand, leveraging ML for extraction and learning of the temporal and spatial features in wireless networks to enable optimizing the HO process. These features include reference signal received power (RSRP), signal to interference plus noise ratio (SINR) / RSRQ, user mobility trajectories and traffic load, etc. [11] proposed a Q learning algorithm that adaptively adjusts TTT and HOM during the HO process according to the UE speed, thereby solving the problem of PPHO and RLF caused by unreasonable settings. In [12], a HCP optimization framework for the high-speed rail scenario was proposed to improve the reference signal received quality (RSRQ) of the UE by intelligently configuring TTT and HOM values. In [13], the Q-learning algorithm dynamically adjusts TTT and HOM according to HO types (such as SHO, TEHO, TLHO and PPHO) to maximize user experience at the cell edge. References [11], [12], [13] optimized HCPs using Q-networks, effectively improving the SHO rate. However, when learning the characteristics of wireless networks, Q-networks rely on robust network models, large-scale Q-tables, and sufficient training data, which presents challenges in algorithm implementation. To address this issue, reference [14] introduced deep Q-networks (DQN) based on Q-networks, effectively alleviating the aforementioned challenges by using neural networks to estimate the value function and advantage function. To reduce the frequent and unnecessary HOs in software-defined networking-enabled ultra-dense networks (UDNs), [15] proposed an adaptive HO control policy optimization method tailored to the dynamic characteristics of the network. Specifically, DQN is used to learn the wireless network status and user behavior, and dynamically adjusts the HO triggering points to ensure communication continuity and stability. In [16], the authors modeled the multiple-user HO and power control problem as a collaborative multi-agent reinforcement learning (RL) problem. In [17], the authors modeled the mobility load balancing problem as a Markov process and achieved the dynamic BS selection and resource allocation by adjusting CIO based on deep RL algorithms. A federated deep RL approach was proposed in [18] to solve the problem of load balancing and frequent HO. Although DQNs have achieved relatively mature applications in optimizing HCPs, they still have some limitations. The primary reason lies in the effectiveness of RL, which is influenced by various factors such as the complexity of the network environment, feature selection, reward function design, and network architecture. In complex network scenarios, issues such as convergence difficulties or low training efficiency may arise, thereby limiting the scalability of intelligent HCP optimization.

To overcome the challenges faced by DQN networks, precise classification and prediction of HO types can guide the optimization process of the RL algorithm in the early stages of training. This approach reduces the randomness of policy exploration, thereby enhancing convergence speed and improving model training efficiency. The authors in [19] classified HOF caused by RLF into three types: TEHO, TLHO, and wrong HO. Doing so can help establish a reasonable HCP range or speed up finding the HCP value. In addition, the classification and prediction capabilities of traditional ML rely on relatively small datasets and simpler model structures, enabling them to achieve high levels of classification and prediction accuracy [20], [21], [22]. In [20], the authors investigated the mobility of non-terrestrial networks, in which ground UEs make access decision by utilizing long short-term memory network based on historical trajectories of non-terrestrial BSs. In [21], the authors proposed a proactive HO management framework for millimeter networks, in which the future data rate fluctuations caused by moving obstacles are predicted and HO execution times are optimized based on the predicted values to ensure link stability. In [22], in order to solve the problem of frequent HO, the author combined recurrent neural network and dead reckoning to predict the next position of the vehicle, and proposed an active selection algorithm that selects target BS based on the predicted position. Although ML has become a crucial tool for studying HO management and optimization, the majority of current research focuses on the selection of target BS and resource allocation. There is relatively less attention given to wireless environmental learning for HO type classification and HCPs settings aiming at user link stability and service continuity. Therefore, we propose a proactive HO type and parameter optimization scheme. This scheme utilizes supervised learning to classify and predict HOs. Subsequently, it employs RL to optimize HCPs based on the prediction results and feature information. The introduction of prediction can correct the action output of RL to a certain extent, accelerate network convergence, and enhance training capabilities. This scheme uses the dual guarantees of HO prediction and RL to ensure that the HCPs of wireless network settings is accurate, thereby achieving excellent optimization results.

B. Main Contributions and Paper Organization

In this paper, we study the HO optimization scheme both consider the pre-classification of HOs and dynamic adjustment of HCPs based on ML. All the simulation results can be reproduced using the Python code and data files available at: https://github.com/HQF6800/HO. The main contributions of this paper can be summarized as follows.

- We first model the HO types (such as SHO, PPHO, TEHO, and TLHO) based on standardized documents. We utilize simulations to replicate actual device HOs and collect HO data. Subsequently, we employ traditional supervised learning classifiers to predict these HO types, selecting the classifier with the highest prediction accuracy as the optimal one. By integrating classification and

prediction functionalities, our extended DQN network reduces the PPHO, TEHO, and TLHO rates to 0.2474, 0.0221, and 0.0632, respectively, achieving an approximately 1.04% reduction compared to the traditional DQN network. In terms of reward function convergence, our approach demonstrates a 7.41% improvement over the conventional DQN network.

- We propose a HCP optimization solution based on DQN. In detail, the UE's feature information is input to the HO type detection module. The HO type detection module outputs prediction results and feature information, which are then sent to the agent module. The agent module sends state information to the decision module. The decision module selects actions based on the received state information and returns them to the agent module. The agent module correct the actions based on the prediction results and then outputs them to the environment. The environment provides feedback rewards to the agent module, and finally, the agent module sends the rewards and actions to the decision module.
- To comprehensively evaluate the performance of the proposed scheme, three performance indicators are used: PPHO, TEHO, and TLHO rate. First, we compares these performance indicators of the DQN without supervised learning and the proposed extended DQN. Second, we compares four different HCPs optimization schemes. Simulation results show that, in terms of reducing the long-term cumulative number of HOs, the proposed solution outperforms the fixed HCP HO scheme, the non-ML HCP optimization scheme, the extreme HCP adjustment scheme based on supervised learning, and the Q-learning scheme.

The rest of the paper is organized as follows. The HO triggering condition, unnecessary HO types, and three performance metrics used in this paper are introduced in Section II. The details of HO classification prediction and HCP optimization scheme are presented in Section III. In Section IV, the proposed schemes are evaluated with plenty of simulation experiments. Finally, a conclusion is given in Section V.

II. System Model

In this paper, we consider a single-layer network architecture in the UDN scenario, which consists of N uniformly distributed small BSs (denoted as $\mathrm{BS}_i, i = 1, \cdots, N$). The downlink RSRP from BS_i to UE is expressed as

$$M_i = T_{\mathrm{x}} - PL\left(d_i\right), \tag{1}$$

where T_{x} is the transmitting power of BS_i, and $PL\left(d_i\right)$ is the pathloss [23]

$$PL\left(d_i\right) = 22.7 + 10\alpha \log_{10}\left(d_i\right) + 26\log_{10}\left(f\right) + X, \tag{2}$$

where α is the path loss exponent, d_i is the distance between the UE and BS_i which is calculated by the horizontal distance between them, the height of BS antenna and the height of UE, f is the carrier frequency in GHz, and X is the shadowing fading in dB. If BS_i is the serving BS of the UE, then the SINR of the UE can be expressed as

$$SINR_i = \frac{M_i}{\sum\limits_{i \neq j} M_j + N_0}, \tag{3}$$

where N_0 is the additive noise power.

A. HO Triggering Condition

The network-controlled UE-assisted HO mechanism is supported in 3GPP, in which HO can be generally divided into four stages: measurement report stage (T1), HO preparation stage (T2), HO execution stage (T3), and HO completion stage (T4) [24]. HO may be triggered when the RSRP received by the UE from a neighbor BS is higher than that of the serving BS. The simplified A3 event triggering condition can be expressed as

$$M_{\mathrm{T}} > M_{\mathrm{S}} + HOM, \tag{4}$$

where M_{T} and M_{S} represent the RSRP of target BS and serving BS, respectively. If M_{T} and M_{S} meet the condition in (4) continuously within TTT time period, the UE is triggered to switch to the target BS.

Fig. 1 illustrates the HO condition based on the simplified A3 event. In the figure, point A indicates that the UE satisfies the entry condition of the A3 event, Point B indicates that the UE has switched from the serving BS to the target BS, and the time interval between point A and point B is TTT. HOM is a parameter in dB, which is generally considered to the minimum received power or quality that triggers HO. TTT is the parameter in ms and is defined as the minimum time that must satisfy the HO triggering condition in (4).

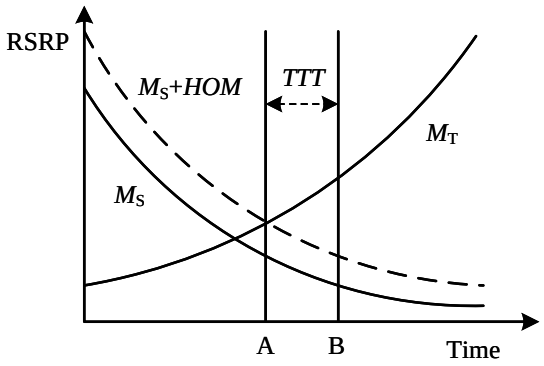


Fig. 1: An illustration of handover condition based on the simplified A3 event.

HOM and TTT are to ensure the stability of HO, but their values have a great impact on the HO performance. Higher HOM and TTT values can reduce the number of PPHOs, but also cause delays in HO decisions, resulting in TLHOs. On the contrary, lower values of HOM and TTT may reduce the delay of HO decisions, but may lead to TEHOs. Therefore, proper HOM and TTT settings are important to avoid unnecessary HOs.

B. Unnecessary HO Types

In this paper, we focus on unnecessary frequent HO and HOF due to RLF: PPHO, TEHO and TLHO [24]. In UDN, due to improper HCPs setting, one of the most common unnecessary HO is PPHO, as shown in Fig. 2(a). The definition of PPHO is that the UE switches back to the source cell within the preset dwell time after establishing connection to the target cell. Although PPHO achieves two SHOs, it will increase the signaling load and waste network resources, and will also reduce the UE's service quality. RLF may occur when the HCP settings are unreasonable, where the UE's SINR remains below a predefined threshold Q_{out}. RLF causes the UE to physically lose the radio connection, interrupting the UE's service and wasting network resources. Due to the reasons for RLF differ, the HOF is futher divided into TEHO and TLHO, as shown in Fig. 2(b) and Fig. 2(c). The definition of TEHO is that RLF occurs after the UE switches to the target cell, and then reconnects to the source cell. The definition of TLHO is that RLF occurs with the source cell during HO process, and then the UE reconnects to the target cell. Due to the need for additional retransmission and reconnection, it will result in greater signaling overhead and even worse call drop rates.

As shown in Fig. 3, the models of TEHO and TLHO are presented based on the standard document. TEHO and TLHO are divided according to the time when RLF occurs during the HO process. Here, State 1 refers to stage before entering the simplified A3 event. State 2 refers to the stage after the entry condition of simplified A3 event is satisfied and before the UE successfully receives the HO command. State 3 is the stage after the UE receives the HO command and before the UE successfully sends the HO completion. The stage after T4 is State 4. If the RLF occurs in State 2, it is TLHO, and if the RLF occurs after the HO is completed, it becomes TEHO. The definitions of PPHO, TEHO, and TLHO are provided below.

C. Performance Metrics

The RLF-related HOF rate is used as the first performance metric

$$R_{\text{HOF}} = \frac{N_{\text{HOF}}}{N_{\text{HO}}}, \tag{5}$$

where N_{HOF} is the total number of TEHOs and TLHOs, and N_{HO} is the number of HOs.

The PPHO rate is used as the second performance metric

$$R_{\text{PPHO}} = \frac{N_{\text{PPHO}}}{N_{\text{HO}}}, \tag{6}$$

where N_{PPHO} is the number of PPHOs.

One of the goals of this paper is to optimize HCP to improve the quality of HO decision by choosing appropriate HOM and TTT values, thereby increasing the number of SHOs. For this, we use SHO rate as the third performance metric

$$R_{\text{SHO}} = \frac{N_{\text{SHO}}}{N_{\text{HO}}}, \tag{7}$$

where N_{SHO} is number of the SHOs, and it meets $N_{\text{HO}} = N_{\text{SHO}} + N_{\text{HOF}} + N_{\text{PPHO}}$.

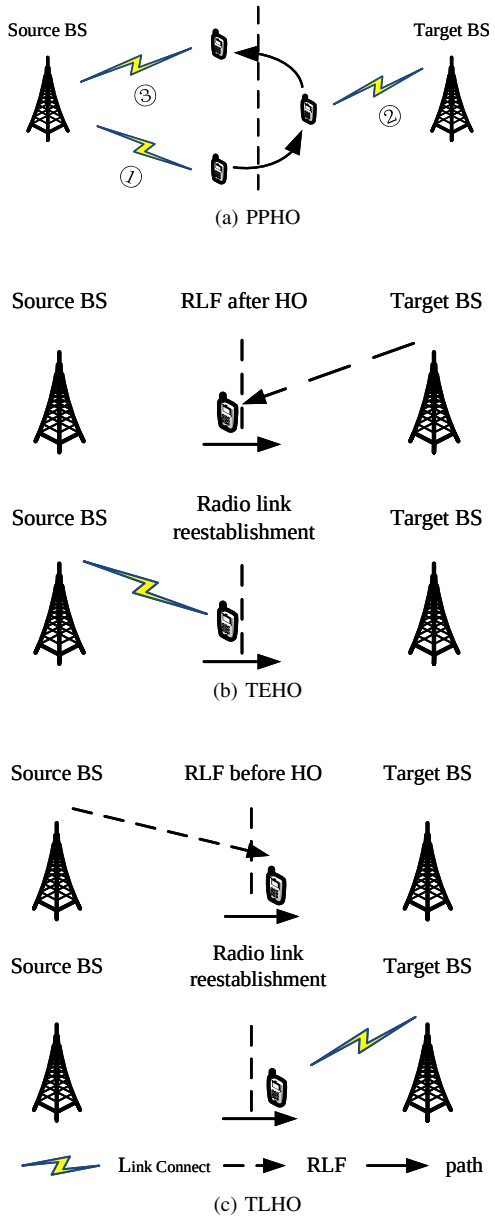


Fig. 2: Illustration of unnecessary HOs.

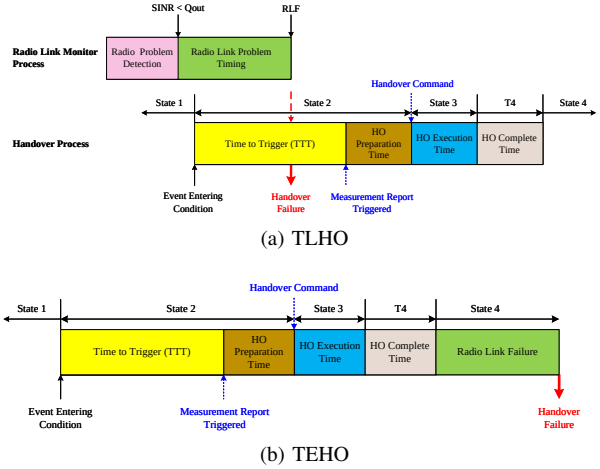


Fig. 3: The modeling of TLHO and TEHO due to RLF.

III. Proposed Algorithm

A DQN-based extension architecture under UDN is proposed. As shown in Fig. 4, the architecture mainly includes three modules: HO type detection module, agent module, and decision module.

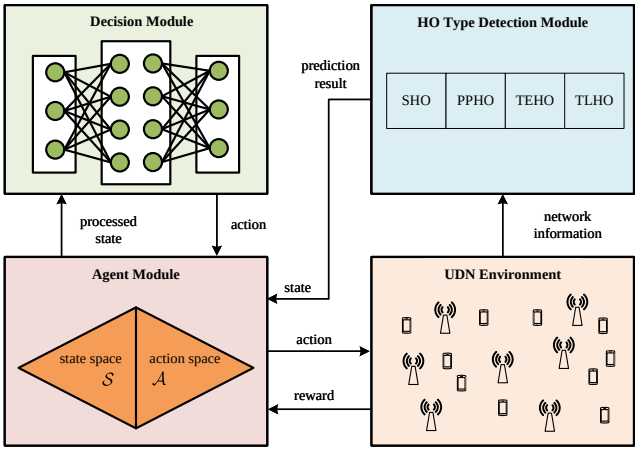


Fig. 4: DQN-based expansion architecture.

The HO type detection module classifies HO according to the input features, including SHO, PPHO, TEHO and TLHO. Its main functions are as follows:

- Obtain network information (e.g., RSRPs and SINRs of UEs, locations of UEs) from UDN environment.
- Judge whether the HO trigger condition is satisfied based on the obtained network information; if it is satisfied, the required feature information is sent to the trained HO type classifier.
- The classifier is used to predict which type of HO will occur in the UE, and the prediction result and state information are sent to the agent module.

The agent module is the main body that perceives the state and takes actions in UDN. Its main functions are as follows:

- Obtain state information and prediction result from the HO type detection module, and use the state information as the state.
- Process the state information and send the processed state information to the decision-making module.
- Correct the actions returned by the decision module based on the prediction results, apply the action to the network, and get the reward of the action from the UDN environment.

The decision module uses DNN as a function approximator to map states to Q-values. Its main functions are as follows:

- Read the state processed by the agent module and take it as input state.
- Choose an action based on the Q-value obtained from the input state mapping.
- Return the selected action to the agent module and get the reward after the agent module executes the action.

A. HO Type Prediction

In this paper, the HO type prediction problem is modeled as a multi-classification problem, and HO is divided into four categories, namely SHO, PPHO, TEHO and TLHO. Supervised learning algorithms create models by learning from labeled training data, which allow them to classify unlabeled data. The input features are composed of the measurement report data that triggers HO and the location information of the UE in the network: $F = [BS_S\ BS_T\ M_j\ M_i\ SINR_j\ SINR_i\ dis_j\ v\ dir]$, where BS_S and BS_T represent the current serving BS and target BS, respectively, M_j and M_i represent the RSRP between the UE and the target BS and the serving BS respectively. $SINR_j$ and $SINR_i$ represent the SINR between the UE and the target BS and serving BS respectively. At the same time, dis_j, v and dir represent the distance between the UE and the target BS, the moving speed and the moving direction of the UE when the UE entering the A3 event, respectively. Each set of inputs is labeled as one of SHO, TLHO, TEHO, and PPHO.

B. HCP Optimization Scheme

In the case where the HOM and TTT are fixed, the moving speed of UEs will affect the success of the HO. To better manage this effect and facilitate analysis, we divide movement speed into two classes: low speed and high speed. When the speed of the UE is lower than 40 km/h, it is called a low-speed UE, otherwise, when the speed is higher than 40 km/h, it is called a high-speed UE. For high-speed UEs, it is necessary to avoid the occurrence of TLHO and PPHO or reduce the frequency of their occurrence. For low-speed UEs, the main goal is to avoid the occurrence of or reduce the frequency of TEHO and PPHO. Although this paper discusses two speed cases, the proposed method can be extended to more fine-grained speed classification cases. For more detailed speed classification, we will discuss and analyze it in future work. For example, similar to [25], we can divide the speed into low speed, medium speed and high speed.

When the UE moves at a high speed, the UE will quickly leave the current serving BS and enter the coverage of other BSs, thereby triggering a HO event. If the TTT is set larger, the UE may experience RLF on the link with the serving BS before the TTT timer expires, resulting in TLHO. However, if the HOM is set within a reasonable range and the TTT is set relatively small, the UE can successfully complete HO, which can avoid RLF with the serving BS, that is, TLHO will not occur. Therefore, for high-speed UEs, we can optimize TTT to reduce the occurrence of TLHO.

When the UE moves at a low speed, if the HOM is set to be small, the HO event will be triggered relatively early. Although the UE can perform HO from the serving BS to the target BS, RLF may occur after completing the HO, resulting in TEHO. If the HOM is set larger, although the HO event will be triggered relatively late, due to the slow speed of the UE, the UE can also complete the HO within a reasonable TTT range, and the signal quality between the UE and the target BS is relatively stable. In this case, RLF will not occur, that is, TEHO will not occur. Therefore, for low-speed UEs, the HOM can be optimized to reduce the probability of TEHO.

In addition, regardless of whether the speed of the UE is high or low, if the HOM value is set to be small, the UE will

perform HO more frequently. If the moving direction of the UE changes after handing over to the target BS, the UE may switch from the target BS back to the source serving BS again, and PPHO occurs. Therefore, we can optimize the HOM for UEs where PPHO may occur.

C. MDP Modeling

In this paper, we model the UDN HO scenario as an environment in DQN and use MDP to model the DQN problem. Specifically, the modeling of MDP involves four aspects: state space, action space, reward function, and Q-value target.

- State space $\mathcal{S}$: The state of UE, $s \in \mathcal{S}$, is consistent with the input features of the classification, that is,

$$\mathcal{S} = \{BS_S, BS_T, M_j, SINR_j, M_i, SINR_i, dis_j, v, dir\}. \tag{8}$$

  The state information in (8) is crucial for determining the HO types. In real-world scenarios, the information related to distance, velocity and direction can be obtained through device sensors, while other state information can be directly obtained from BSs/UEs.
- Action space $\mathcal{A}$: The action space includes all actions that the agent can take, $a_1, a_2 \in \mathcal{A}$. Here, action $a_1 \in \{0, 2, 4, 6, 8, 10\}$, represents the value of HOM in dB, and $a_2 \in \{100, 300, 500, 700, 900\}$ represents the value of TTT in ms. When the HO prediction type is PPHO or TEHO (in the presence of low-speed UEs), the action a_1 is selected from the action space. When the HO prediction type is TLHO, and the speed of the UE is high, the action a_2 is selected.
- Reward function r: The reward function is the immediate reward obtained by the agent for taking action a_1 or a_2 in each state, and its role is to promote the algorithm to minimize PPHO and HOF. Therefore, the reward function is

$$r = -\omega_1 R_{\text{PPHO}} - \omega_2 R_{\text{TEHO}} - \omega_3 R_{\text{TLHO}}, \tag{9}$$

  where R_{TEHO} and R_{TLHO} are TEHO rate and TLHO rate, respectively. ω_1, ω_2 and ω_3 are the weights of PPHO, TEHO and TLHO, respectively. The weight setting of the reward function can reflect the priority of different tasks or goals. By increasing or decreasing the weight associated with a certain task, the agent can be instructed to pay more attention to or ignore specific tasks. In this way, the trade-offs between different tasks can be balanced in multi-task reinforcement learning to achieve better performance.
- Q-value $Q(s, a)$: Q-value represents the expected long-term cumulative reward obtained by taking action a in state s. It is calculated by combining the current immediate reward with the maximum expected reward in the next state.

$$Q(s, a) = r(s, a) + \gamma Q_{max}(s', a'), \tag{10}$$

  where $r(s, a)$ is the immediate reward for the current state-action pair (s, a). γ is the discount factor, which lies between 0 and 1. It is used to balance the relative importance of future rewards against immediate rewards. $Q_{max}(s', a')$ is the achievable maximum Q-value in the next state s'. It represents the expected cumulative reward that can be obtained by selecting the optimal action a' in the next state s'.

Based on the above analysis, we propose an HCP optimization strategy based on RL and supervised learning. The main steps are given in Algorithm 1.

Algorithm 1 TTT and HOM Optimization Algorithm

Input:
$\mathcal{S} = \{BS_S, BS_T, M_j, SINR_j, M_i, SINR_i, dis_j, v, dir\}$;
Output:
SHO, PPHO, TEHO, TLHO and corresponding HCP paremeters;
Initiate action space $\mathcal{A} = \{a_1, a_2\}$;
Initiate reward function r, ω_1, ω_2, ω_3;
Initiate TTT and HOM values;
$n = 1$;
steps = 30000;
while $n <$ steps **do**
  Update the state information for all i and j in (8);
  if satisy (4) **then**
    The features are sent to HO type detection module and predict HO types;
    switch (HO type)
    case SHO**:**
      The agent module does not change TTT and HOM values;
    case PPHO**:**
      Enter the state into DQN and select an appropriate HOM value based on maximizing r in the action space $\{a_1\}$;
    case TLHO**:**
      if The UE is a high-speed UE **then**
        Enter the state into DQN and select an appropriate TTT value based on maximizing r in the action space $\{a_2\}$;
      else
        The agent will not change TTT and HOM values;
      end if
    case TEHO**:**
      if The UE is a low-speed UE **then**
        Enter the state into DQN and select an appropriate HOM value based on maximizing r in the action space $\{a_1\}$;
      else
        The agent will not change TTT and HOM values;
      end if
    end switch
    UE performs HO;
    Update the R_{PPHO}, R_{TEHO} and R_{TLHO} in (9);
  end if
end while

IV. Simulation Results and Analysis

In IV-A, the impacts of different parameters on HO performance are explored. First, the HO process of the UE under a specific trajectory and the evolution of RSRP and SINR during its movement are presented. Secondly, the influence of HOM, TTT and speed on HO performance is investigated. In IV-B, the effect of various classifiers on HO classification is verified. In IV-C, we verify the feasibility and effectiveness of the proposed algorithm. We compared the network performance under different HCP algorithms to verify the feasibility and effectiveness of the proposed algorithm. The basic simulation scenario is shown in Fig. 5. The supervised learning classifier used is XGBoost and the RL is DQN. The simulation parameters are shown in Table I. The DQN and XGBoost hyperparameters are set as shown in Table II.

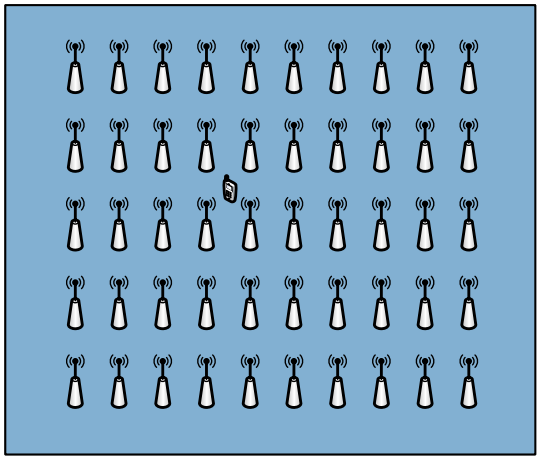


Fig. 5: Simulation scenario.

TABLE I: Simulation Parameters

Parameters	Values
Transmit power, T_{x}	30 dBm
Carrier frequency, f	3.5 GHz
Path loss exponent, α	3.67
Shadow fading, X	$N\,(0,4)$ dB
Height of BS's antenna	10 m
Height of UE	1.5 m
Noise power spectral density	-174 dBm/Hz
SINR threshold, Q_{out}	-12
UE speed, v	10, 20, 36, 40, 60, 72, 80, 100, 120 km/h
TTT	100, 300, 500, 700, 900 ms
HOM	0, 2, 4, 6, 8, 10 dB
Method of UE moving	straight line, random move
# of BSs	49, 50, 100
Weight of PPHO, ω_1	1
Weight of TEHO, ω_2	1
Weight of TLHO, ω_2	1

A. HO Performance Evaluation Under Different Parameters

Fig. 6 shows the evolution of RSRP and SINR when the UE moves along a straight line. In the figure, the abscissa and the ordinate on the left represent the coordinates of UE and BS respectively. The ordinate on the right represents the UE's RSRP, SINR or SINR threshold. In order to show the relationship between them more clearly, we share a common axis, although their units are different. The red triangle indicates the location of the BS, and the blue circle indicates its coverage. There are 49 BSs in the simulation environment, but for a clearer presentation, we only show the BSs that have been connected to the UE. Numbers in red brackets indicate the index of the BS. The green solid line indicates the movement trajectory of the UE, the green rightward arrow indicates the position of the UE when it initiates HO, and the green upward arrow indicates the position of the UE when it completes HO. The numbers above and below the UE movement trajectory indicate the index of the source BS and target BS respectively. The red dotted line represents the SINR threshold, the orange line represents the SINR of the UE, and the dark blue line represents the RSRP between the UE and the serving BS.

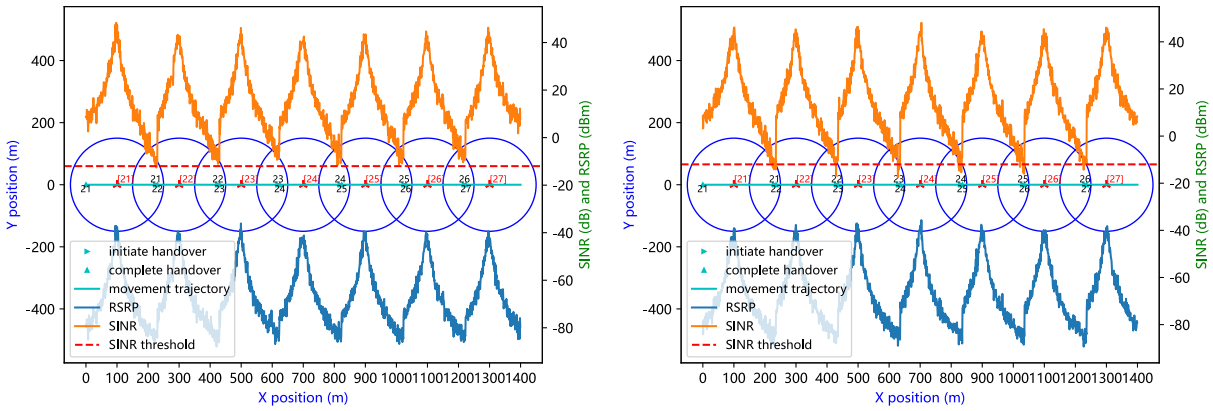
TABLE II: Network Hyperparameters

DQN	Values	XGBoost	Values
Learning rate	0.01	Learning rate	0.01
Replay buffer size	500	# of base estimators tree	500
Discount factor, γ	0.5	# of leaf nodes, γ	0.3
Greedy strategy, ϵ	0.9	Decision tree max depth	25

In Fig. 6(a), the speed of the UE is 10 m/s, and the HOM value is 6 dB. The UE has initiated a HO request before reaching the edge of the BS, and no RLF occurs. Therefore, TLHO does not occur during the UE's mobility. In Fig. 6(b), the velocity of UE is 10 m/s, and the HOM value is 10 dB. Since the selected HOM is too large, the UE does not initiate a HO request until it is at the edge of the BS. During the HO process, the UE undergoes RLF with the source BS, and then establishes a connection with the target BS, resulting in TLHOs. In Fig. 6(c), the HOM value is 6 dB, but the speed of UE is 20 m/s. Due to the higher speed, RLF occurs between the UE and the source BS during the HO process, and then a connection is established with the target BS, and TLHOs also occur. In Fig. 6(d), the HOM value and the speed of the UE are both high, so RLFs have occurred to the UE during all HO processes, and all of them belong to TLHOs.

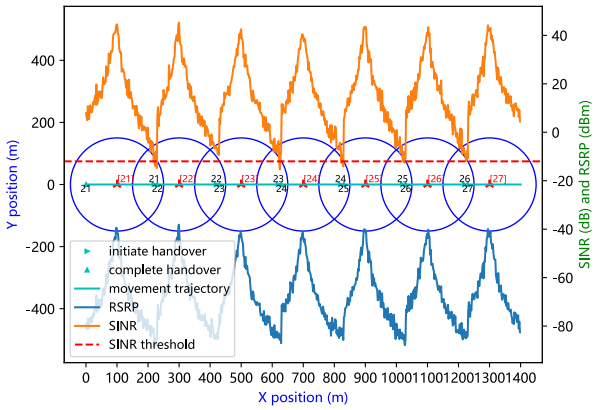
Further, we explore the effects of different HCPs and different speeds on the HO performance. In the simulation, the UE's mobility model is described as follows: UE starts to move from the randomly generated initial position in the rectangular simulation area. After that, every 6 seconds, the UE randomly selects one of eight directions to continue moving: east, south, west, north, northeast, southeast, southwest, and northwest. When the UE reaches the boundary of the rectangular area, it will reselect a moving direction and continue moving. In order to go through as many locations as possible, the UE moves 30,000 steps each time and repeats 200 times.

In Fig. 7, the average HO times and HO rates in the case of fixed UE's speed and TTT ($v = 20$ m/s, $TTT = 500$ ms) but different HOM values are presented. As shown in Fig. 7(a), the average number of HOs decreases with the increase of HOM value. This is because a higher HOM value makes it harder to meet the HO trigger condition, resulting in fewer HOs. Fig. 7(b) shows various HO rates under different HOM

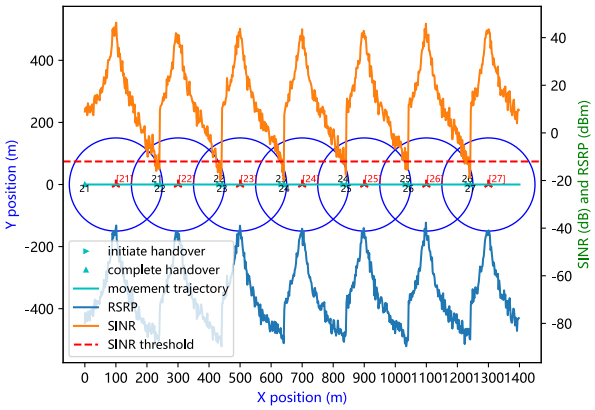


(a) $HOM = 6$ dB, $v = 10$ m/s

(b) $HOM = 10$ dB, $v = 10$ m/s



(c) $HOM = 6$ dB, $v = 20$ m/s



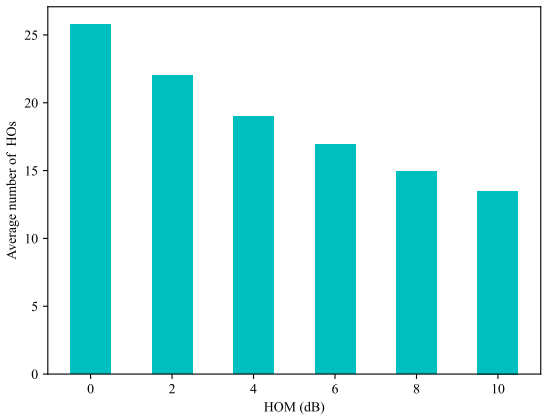
(d) $HOM = 10$ dB, $v = 20$ m/s

Fig. 6: Evolution of RSRP and SINR ($TTT = 500$ ms).

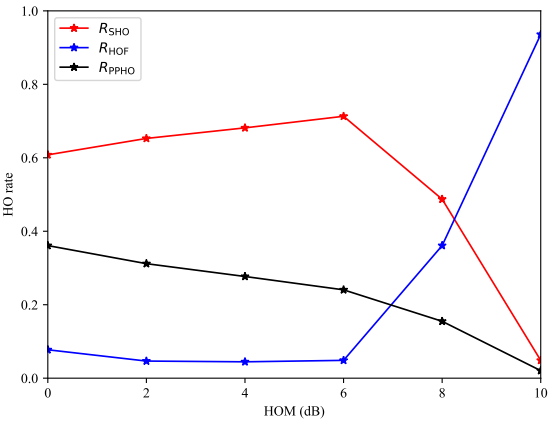
values. It can be found from the figure that with the increase of HOM, the PPHO rate decreases, the HOF rate increases, and the SHO rate first increases and then decreases. Although a high HOM value can reduce the number of PPHOs, it will also increase the number of HOFs. A high HOM will prevent UE from switching back to the source BS, thus reducing the number of PPHOs or PPHO rate. Too high HOM will lead to too late switching to the target BS, increasing the probability of RLF, that is, increasing the HOF rate. At this time, the HOF rate increases sharply, and the SHO rate decreases sharply. For example, the HOF rate rises sharply when the HOM is 8 dB and 10 dB.

In Fig. 8, the average HO times and HO rates in the case of fixed UE's speed and HOM ($v = 20$ m/s, $HOM = 2$ dB) but different TTT values are presented. As shown in Fig. 8(a), as the TTT value increases, the average number of HOs decreases gradually. Like the HOM, the larger the TTT value, the longer the UE needs to connect to the source BS, and the harder it is to meet the HO triggering condition, resulting in fewer HOs. From Fig. 8(b), it can be found that with the increase of TTT, the PPHO rate decreases, the HOF rate increases, and the SHO rate decreases slightly. A high TTT can also increase the difficulty for the UE to handover back to the source BS, thereby reducing the PPHO rate. If the TTT value is set too large, the UE may fail to trigger the HO in time, resulting in a significant drop in signal quality, resulting in more HOFs or an increase in HOF rate. Since the number of PPHOs is lower than that of HOFs, the SHO rate drops slightly. Since the decrease in PPHO times is less than the increase in HOF times, the SHO rate decreased slightly.

In Fig. 9, the average HO times and HO rates under the condition of fixed HOM and TTT values ($HOM = 2$ dB, $TTT = 500$ ms) but different speeds are given. It can be seen from Fig. 9(a) that as the UE speed increases, the average number of HOs also increases. This is because when the UE moves at a higher speed, the UE enters the coverage area of different BSs more frequently, that is, HOs to different BSs more frequently to maintain the communication connection. It can be seen from Fig. 9(b) that as the speed increases, the HOF rate increases, the PPHO rate decreases, and the SHO remains relatively stable. This is due to that the UE will leave the coverage of the source BS faster and enter the coverage

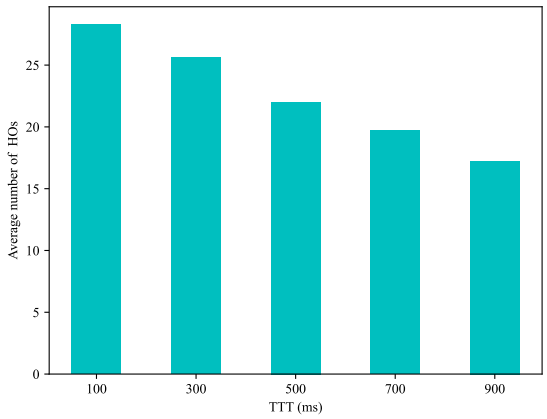


(a) Average number of HOs.

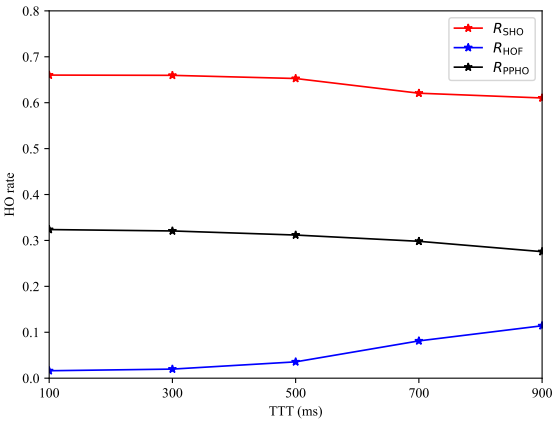


(b) HO rate of various types.

Fig. 7: HO performances ($v = 20$ m/s, $TTT = 500$ ms).



(a) Average number of HOs.



(b) HO rate of various types.

Fig. 8: HO performances ($v = 20$ m/s, $HOM = 2$ dB).

of the target BS, and it is more difficult to switch back to the source BS within a predetermined time, so the number of PPHOs or PPHO rate is reduced. At the same time, for a given HOM and TTT, it is more likely for the UE to fail to trigger the HO to the target BS in time, and the signal quality is severely degraded, resulting in a TLHO. This increases the HOF rate.

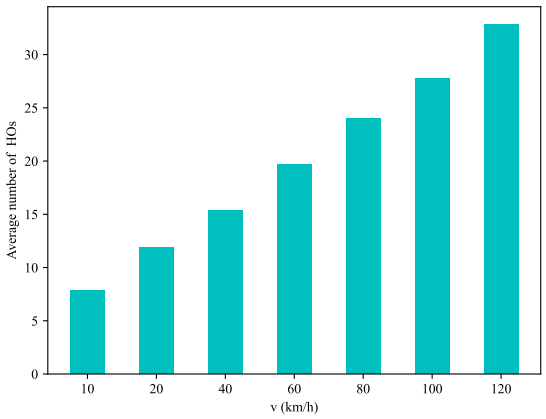
B. HO Prediction Evaluation of Different Classifiers

We compare and verify the effectiveness of various classifiers for HO classification. Since the UE has much more SHO than PPHO, TLHO and TEHO during the moving process, the collected HO data are unbalanced, that is, the number of the four HO types differs greatly. If a classification model is trained on an unbalanced dataset, this will lead to poor predictive performance of the model, especially for classes with fewer samples. Therefore, we use the SMOTE algorithm to perform data balance processing on the original data set. After that, we test the accuracy of seven classifiers including KNN, LR, DT, RF, MLP, AdaBoost and XGBoost on the balanced data set, and the results are shown in Table III. It can be found that most classifiers have higher classification accuracy for SHO and PPHO, but lower classification accuracy for TEHO and TLHO. The reason for the difference in the prediction accuracy of the four HO types is that during the data balance process of the data set, the newly generated data is generated based on the original data set, so the data quality is lower than the original data. Additionally, the acquisition of state information can also affect the HO rate. The distance, speed, and direction information of the user may introduce some errors. Through simulation verification, it was found that the lack of user speed, movement direction, and distance information in the network state negatively affects network performance, resulting in an inappropriate HO rate of 0.65. In contrast, when there is an error in the user position information, the inappropriate HO rate only increases by 0.025, which has a relatively small impact on network performance.

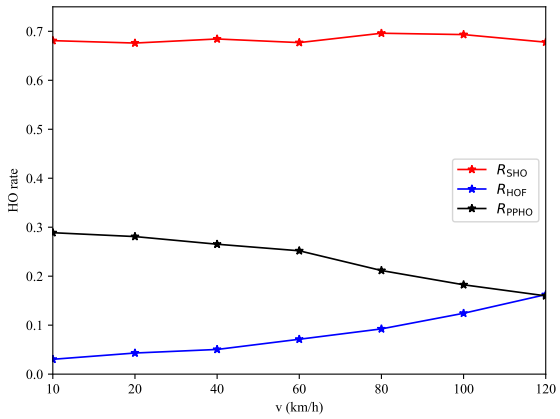
Fig. 10 shows the receiver operating characteristic (ROC) curves of XGBoost and MLP classifiers for four different HO types. In the figure, class0, class1, class2 and class3 represent SHO, PPHO, TLHO and TEHO respectively. The ROC curve is generated by computing the true positive rate (TPR) and

TABLE III: Classification Accuracy of Different Classifiers

HO types \ Classifiers	KNN	LR	DT	RF	MLP	Adaboost	XGboost
SHO	0.66	0.63	0.74	0.77	0.84	0.66	0.95
PPHO	0.72	0.66	0.76	0.83	0.97	0.67	0.94
TLHO	0.59	0.68	0.62	0.61	0.67	0.68	0.75
TEHO	0.59	0.88	0.57	0.57	0.60	0.91	0.62

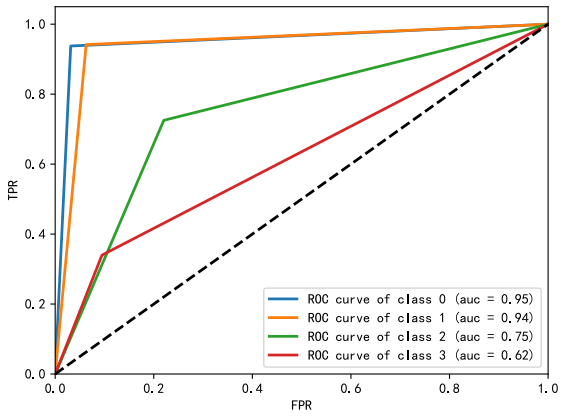


(a) Average number of HOs.

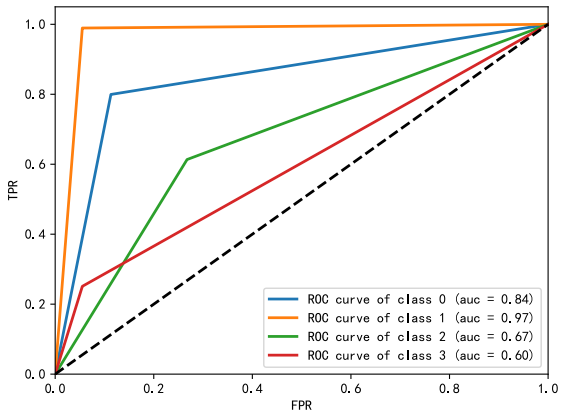


(b) HO rate of various types.

Fig. 9: HO performances ($HOM = 2$ dB, $TTT = 500$ ms).



(a) XGBoost.



(b) MLP.

Fig. 10: ROC curves of two classifiers.

false positive rate (FPR). The location and area of these curves can help us evaluate the performance of the classifiers. It can be seen from the figure that the ROC curves of XGBoost and MLP are above the black dotted line, that is, the area under ROC curve is greater than 0.5, which indicates that both classifiers have good classification prediction capabilities, and their classification prediction accuracy is better than that of random classification.

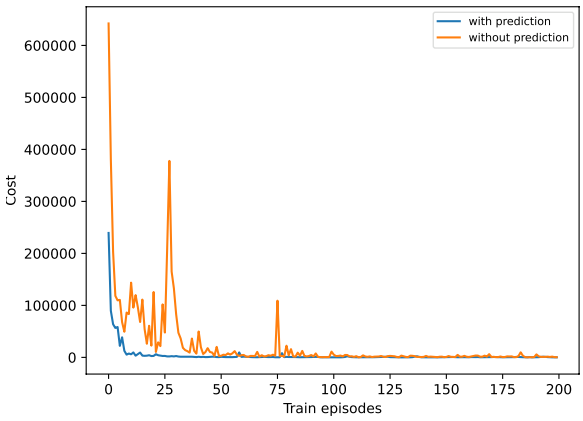
C. Performance Verification of the Proposed Algorithm

We compare the performance of the DQN without supervised learning predictions to that of the DQN with supervised learning. We set the UE speed to 20 m/s and the BS number to 50. In Fig. 11(a), the loss function of the DQN is presented, the DQN with predictions converges faster in terms of loss compared to the DQN without predictions. In our proposed extended DQN, we introduced the HO type detection process. This helps the agent output the correct actions, which improves the network's performance to some extent and accelerates the convergence of the loss function. The DQN training process without supervised learning causes 92% TEHO and does not suppress the occurrence of RLF. The DQN network training with supervised learning can reduce PPHO to around 24%, TEHO to 3%, and TLHO to 7%. This indicates that the

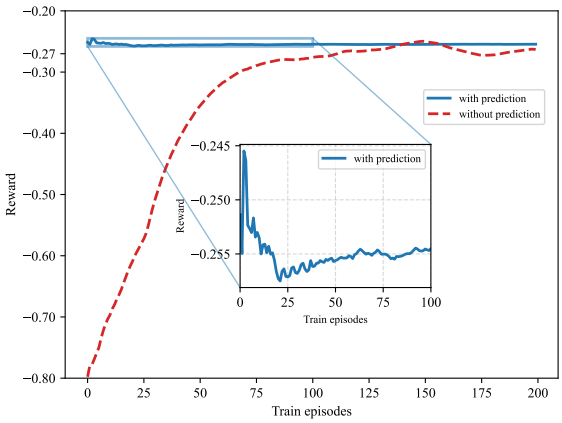
addition of supervised learning enables our network to support both offline and online learning. As shown in Fig. 11(b), the blue solid line represents the reward function of DQN with the prediction module, while the red dashed line represents the reward function of DQN without the prediction module. It can be observed that the reward function of the blue solid line is higher than that of the red dashed line and converges more rapidly. Specifically, the reward function of the blue solid line initially increases sharply to -0.246, then decreases to -0.257, and eventually converges to -0.255. The initial upward fluctuation can be attributed to the prediction module, which reduces the action space of the agent in the early training stages by predicting HO types. This reduction allows the agent to select several higher-reward actions within a smaller action space. However, as training progresses, the reward begins to decline because the agent needs to explore more extensively and try different actions and strategies, which may lead to a higher number of incorrect action choices and, consequently, a decrease in reward. As exploration gradually decreases, the agent begins to adjust its strategy based on long-term returns, and the reward function stabilizes and converges. It is important to note that the value function used in this study emphasizes long-term cumulative rewards, and therefore, the converged reward value may be lower than the short-term optimal solution. In contrast, the DQN network represented by the red dashed line, which does not incorporate the prediction module, lacks corrective actions for its choices. In the early stages of training, the reward function is relatively low, but as training progresses, the reward gradually increases, ultimately converging to approximately -0.27. This process reflects that, without the prediction module, the agent gradually gains higher rewards by continuously adjusting its strategy. However, the convergence speed is slower, and the final reward is slightly lower than that of the DQN network with the prediction module.

In this paper, variations in the weighting factors affect the occurrence of different types of HO events and performance metrics. In order to verify the feasibility of the proposed scheme, we discuss the effects of different weighting parameters on PPHO rate, TLHO rate and TEHO rate. As shown in Table IV, our proposed extended DQN can reduce the PPHO rate, TEHO rate and TLHO rate by about 25%, 3% and 7%, respectively. When fixing ω_2 and ω_3 to 1 and increasing ω_1 from 1 to 10, the PPHO rate of the reference UE decreases by 1.62%, but the TEHO and TLHO increase by 2.49% and 17.05%, respectively. When changing the weighting factors, the variation in the TEHO and TLHO rates are smaller than the variation in the PPHO rate. For example, when fixing ω_1 and ω_3 to 1 and increasing ω_2 from 1 to 10, the PPHO rate of the reference UE increases by 14.79% while the TEHO rate decreases by 6.23%. Therefore, by adjusting the weights of the reward function, the inappropriate HO rate experienced by users can be changed, allowing for optimization of specific types of inappropriate HOs.

We investigated the impact of classifier prediction accuracy on the network performance of the proposed scheme. In particular, the impact of the seven classifiers mentioned in Section IV-B is compared. The changes in the loss function are



(a) DQN loss function.



(b) DQN reward function.

Fig. 11: DQN objective function.

presented in Fig. 12 and the corresponding performance metrics of the extended DQN with different prediction accuracies are presented in Table V. It can be found that classifiers with higher prediction accuracy can enhance the performance of the DQN, while those with lower prediction accuracy degrade the performance of the DQN. For example, when the XGBoost classifier with higher accuracy is combined with DQN, the loss function converges faster, and the PPHO rate, TEHO rate and TLHO rate are reduced to 26.8%, 6.4% and 2.6%, respectively. The lower prediction accuracy of the logistic regression classifier, combined with DQN, makes it difficult for the network to converge, and the TLHO rate is as high as 84.1%.

We compare the proposed algorithm with the other four algorithms in terms of PPHO rate, TEHO rate and TLHO rate performance metrics. The four algorithms used for comparison are traditional HCP algorithm (THA), non-ML, supervised-learning-based extremely HCP algorithm (SLEHA) and Q-learning-based HCP algorithm (QHA).

- THA: Similar to [24], this method neither performs classification prediction nor HCP optimization for HOs.

TABLE IV: Effect of Weight Parameters on HO Rates

HO rate \ weights	$\omega_1=1$ $\omega_2=1$ $\omega_3=1$	$\omega_1=0.1$ $\omega_2=1$ $\omega_3=1$	$\omega_1=10$ $\omega_2=1$ $\omega_3=1$	$\omega_1=1$ $\omega_2=0.1$ $\omega_3=1$	$\omega_1=1$ $\omega_2=10$ $\omega_3=1$	$\omega_1=1$ $\omega_2=1$ $\omega_3=0.1$	$\omega_1=1$ $\omega_2=1$ $\omega_3=10$
R_{PPHO}	0.244	0.269	0.231	0.243	0.254	0.251	0.252
R_{TEHO}	0.025	0.023	0.026	0.027	0.020	0.034	0.022
R_{TLHO}	0.068	0.065	0.071	0.066	0.070	0.076	0.062

TABLE V: Performance Metrics of Different Classifiers

HO rate \ Classifiers	KNN	LR	DT	RF	MLP	Adaboost	XGboost
R_{PPHO}	0.411	0.188	0.345	0.311	0.305	0.340	0.269
R_{TLHO}	0.010	0.031	0.010	0.030	0.030	0.012	0.034
R_{TEHO}	0.057	0.610	0.030	0.003	0.011	0.189	0.017

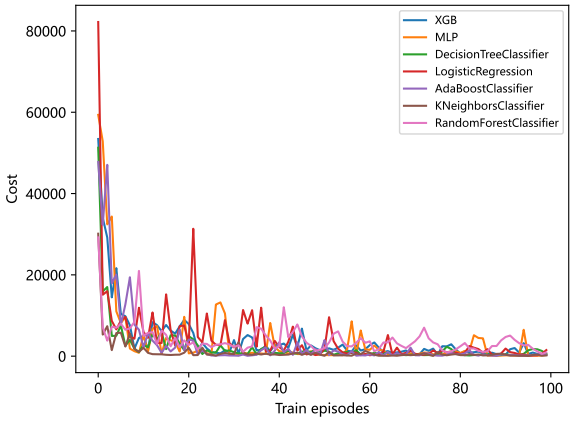


Fig. 12: Loss function of different classifiers.

Specifically, HOM and TTT are set as fixed values during the whole simulation process, such as $HOM = 3$ dB, $TTT = 500$ ms.

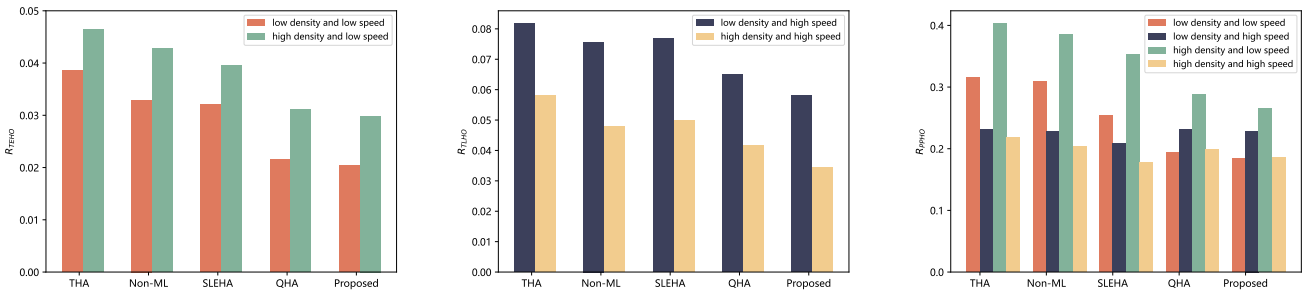
- Non-ML: Similar to [11], this scheme adjusts the HCPs by increasing or decreasing the fixed values. When PPHO and TEHO occurred, fixed values were added to the original TTT and HOM values. And when TLHO occurs, the fixed values are subtracted from the original TTT and HOM values.
- SLEHA: Similar to [26], supervised learning is used to transform the HO problem into a classification problem. Although this method classifies HO, it does not perform adaptive HCP optimization, that is, HOM is set to 0 dB or 10 dB, and TTT is set to 100 ms or 900 ms according to the HO classification results. The initial settings of HOM and TTT of this optimization algorithm are the same as those of THA.
- QHA: Similar to [27], Q-learning is used to learn policies by mapping states and actions to a Q- value, with the ultimate goal of finding policies that maximize cumulative rewards. The actions and rewards of this algorithm are the same as those of the algorithm proposed in this paper.

It should be noted that the initial HCP of all the above methods is the same, $HOM = 3$ dB, $TTT = 500$ ms. For SLEHA, QHA and the proposed scheme, when the predicted result is SHO, the initial HCP is used to perform HO, and when the prediction result is other types, each scheme performs HO according to its adjusted HCP. For the purpose of demonstrating the impact of HCP adjustment and optimization, we compare TEHO rate, TLHO rate and PPHO rate under different BS density and velocity combinations: low density BS (50 BSs) and high density (100 BSs), low speed (5m/s) and high speed (20m/s), as shown in Fig. 13.

In Fig. 13(a), compared with THA, Non-ML, SLEHA and QHA, in the low-density BS and low-speed UE environment, the proposed algorithm reduces the TEHO rate by 45.66%, 38.39%, 35.79% and 6.03% respectively; while in high-density BS and low-speed UE deployment environment, the proposed algorithm reduces the TEHO rate by 37.62%, 30.19%, 26.83% and 6.31% respectively. Furthermore, the greater the deployment density of BSs, the higher the probability of TEHO. This is because, as the BS density increases, the distance between BSs decreases, and the signals in the overlapping area are more unstable and the interference is stronger, all of which lead to an increase in the number of RLFs and thus an increase in TEHO.

In Fig. 13(b), for high-speed UEs, the proposed method outperforms the other three methods in improving TLHO performance regardless of high or low BS density. For example, in a low-density BS and high-speed UE deployment environment, compared with THA, Non-ML, SLEHA and QHA, this algorithm reduces the TLHO rate by 29.07%, 23.28%, 24.25% and 10.55% respectively. The higher the deployment density of BSs, the UE receives stronger signals from more neighboring target BSs. Therefore, HO can be triggered in a timely manner, and the probability of TLHO occurrence is lower.

Fig. 13(c) illustrates the comparison of PPHO rates across four combinations of BS density and UE motion speed: low-density low-speed, low-density high-speed, high-density low-speed and high-density high-speed environment. In the low-speed UE environment, high density of BSs will produce more PPHOs compared to low density of BSs. Since in high-density BS environment, the user crosses more overlapping area of BSs with unstable signals. In high-speed UE environment, high



(a) TEHO. (b) TLHO. (c) PPHO.

Fig. 13: HO performances of four schemes.

density of BSs instead produces fewer PPHOs compared to low density BSs. Therefore, with the BS density increases, there are more candidate BSs with HO options for UE in the overlap area, which reduce the likelihood for UE to randomly walk to reconnect to the serving BS. Without loss of generality, high-speed UE produces fewer PPHOs compared to low-speed UE since the user will quickly cross the base station overlap area.

Nevertheless, QHA and the proposed algorithm yield different outcomes. In environments involving both low-density and high-speed as well as high-density and high-speed, the performance of QHA and the proposed algorithm falls short compared to SLEHA. In SLEHA, when the classifier classifies an upcoming HO type as PPHO based on the features of the input network, it assumes that the value of the HOM setting at this point will cause the UE to experience PPHOs. The classifier adjusts the value of the HOM straight up to the highest to avoid the occurrence of PPHOs. In the QHA and Proposed algorithm, when the UE is identified to be on the verge of experiencing PPHOs, both algorithmic decisions select HOM actions that is not adjusted to the highest. Instead, a lower HOM is selected. However, compared to the highest HOM value, there is an increase in the occurrence of PPHOs. Therefore, in the low-density and high-speed conditions, the QHA and Proposed algorithm only reduce the PPHO rate by 0.5% and 1.3% relative to the PPHO rate of the traditional HO algorithm. In low-density and low-speed as well as low-density and high-speed environments, both the QHA and Proposed algorithms exhibit a higher occurrence of PPHOs at higher speeds compared to lower speeds. The reason is that, in the HO problem, a state is influenced of multiple factors, such as signal strength, interference, user velocity, and so on. These factors vary in a discrete space, DQN and Q-learning algorithms may yield different results when learning in such a complex state space.

According to the above analysis, the HCP optimization methods based on HO classification are superior to THA. Even under extreme HCP tuning, the HO performance can be improved to a certain extent. This shows the value of HO classification in HO performance optimization. Moreover, the proactive adjustment of HCP using ML can further improve the HO performance. Both QHA and our proposed algorithm can make adaptive adjustments for HCP, so they have better performance than SLEHA. Finally, the advantage of our algorithm over QHA is that traditional Q-learning uses tables to store Q-values, which becomes intractable for continuous state spaces or large discrete state spaces; DQN approximates the Q-value function through neural networks, which can better deal with continuous state space; and DQN introduces an experience playback mechanism, stores previous experience samples in the playback buffer, and randomly samples them for network training, which can solve the problem of sample correlation and improve the efficiency and stability of training.

V. Conclusion

In this paper, the main focus is on addressing unnecessary HOs and RLFs in UDNs, aiming to enhance QoS of users. This paper proposes a scheme that proactively predicts HO types and dynamically adjusts HCPs, and performs a performance analysis on this scheme. We first classify unnecessary HOs into PPHO, TEHO and TLHO, and utilize seven classifiers commonly used in ML for HO prediction. The simulation results indicate that the XGBoost classifier achieves the best prediction performance with an accuracy of 94.83%. Subsequently, the proposed scheme applies a DQN to dynamically adjust TTT and HOM based on different predicted HO type. Comparative simulations with SLEHA, QHA and the proposed algorithm show that the proposed algorithm performs better, reducing TEHO and TLHO rates to 2.03% and 3.43%, respectively. Therefore, the solution of pre-classifying HOs and adjusting HCP according to HO types can effectively reduce the occurrence of unnecessary HOs.

Future work can focus on conducting specific analysis for the optimal HCPs and considering more practical scenarios, such as heterogeneous networks and variable-speed reference UEs. In addition, future research can also explore the utilization of multi-agent systems to reduce unnecessary HOs for multiple mobile UEs, and use average UE throughput as a performance metric for discussion.

References

[1] Chen, T. Zhao, H. -H. Chen and W. Meng, "Network densification and path-loss models versus UDN performance—A unified approach," *IEEE Trans. Wireless Commun.*, vol. 20, no. 7, pp. 4058-4071, Jul. 2021.

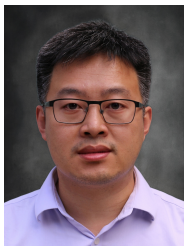
[2] S. Jaffry, R. Hussain, X. Gui and S. F. Hasan, "A comprehensive survey on moving networks," *IEEE Commun. Surv. Tutor.*, vol. 23, no. 1, pp. 110-136, Firstquarter 2021.
[3] C. Lee, H. Cho, S. Song and J. -M. Chung, "Prediction-based conditional handover for 5G mm-Wave networks: A deep-learning approach," *IEEE Veh. Technol. Mag.*, vol. 15, no. 1, pp. 54-62, Mar. 2020.
[4] M. Manalastas, M. U. B. Farooq, S. M. A. Zaidi, A. Abu-Dayya and A. Imran, "A data-driven framework for inter-frequency handover failure prediction and mitigation," *IEEE Trans. Veh. Technol.*, vol. 71, no. 6, pp. 6158-6172, Jun. 2022.
[5] S. M. A. Zaidi, M. Manalastas, M. U. B. Farooq, H. Qureshi, A. Abu-Dayya and A. Imran, "A data driven framework for QoE-aware intelligent EN-DC activation," *IEEE Trans. Veh. Technol.*, vol. 72, no. 2, pp. 2381-2394, Feb. 2023.
[6] M. -T. Nguyen and S. Kwon, "Geometry-based analysis of optimal handover parameters for self-organizing networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp. 2670-2683, Apr. 2020.
[7] T. M. Duong and S. Kwon, "A framework of handover analysis for randomly deployed heterogeneous networks," *Comput. Netw.*, vol. 217, no. 109351, 2022.
[8] M. Umar Bin Farooq *et al.*, "A data-driven self-optimization solution for inter-frequency mobility parameters in emerging networks," *IEEE Trans. Cogn. Commun. Netw.*, vol. 8, no. 2, pp. 570-583, Jun. 2022.
[9] H. Tabassum, M. Salehi and E. Hossain, "Fundamentals of mobility-aware performance characterization of cellular networks: A tutorial," *IEEE Commun. Surv. Tutor.*, vol. 21, no. 3, pp. 2288-2308, thirdquarter 2019.
[10] M. Zaher, E. Björnson and M. Petrova, "Soft handover procedures in mmWave cell-free massive MIMO networks," *IEEE Trans. Wireless Commun.*, doi: 10.1109/TWC.2023.3330199.
[11] R. Karmakar, G. Kaddoum and S. Chattopadhyay, "Mobility management in 5G and beyond: A novel smart handover with adaptive time-to-trigger and hysteresis margin," *IEEE Trans. Mob. Compt.*, vol. 22, no. 10, pp. 5995-6010, 1 Oct. 2023.
[12] C. Wu, X. Cai, J. Sheng, Z. Tang, B. Ai and Y. Wang, "Parameter adaptation and situation awareness of LTE-R handover for high-speed railway communication," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 3, pp. 1767-1781, Mar. 2022.
[13] M. L. Marí-Altozano, S. S. Mwanje, S. L. Ramírez, M. Toril, H. Sanneck and C. Gijón, "A service-centric Q-learning Algorithm for Mobility Robustness Optimization in LTE," *IEEE Trans. Netw. Service Manag.*, vol. 18, no. 3, pp. 3541-3555, Sept. 2021.
[14] N. Zhao, Y. C. Liang, D. Niyato, Y. Pei, M. Wu and Y. Jiang, "Deep reinforcement learning for user association and resource allocation in heterogeneous cellular networks," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5141-5152, Nov. 2019.
[15] W. Huang, M. Wu, Z. Yang, K. Sun, H. Zhang, and A. Nallanathan, "Self-adapting handover parameters optimization for SDN-enabled UDN," *IEEE Trans. Wireless Commun.*," vol. 21, no. 8, pp. 6434-6447, Aug. 2022.
[16] 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 Trans. Veh. Technol.*, vol. 69, no. 11, pp. 13124-13138, Nov. 2020.
[17] G. Alsuhli *et al.*, "Mobility load management in cellular networks: A deep reinforcement learning approach," *IEEE Trans. Mob. Compt.*, vol. 22, no. 3, pp. 1581-1598, 1 Mar. 2023.
[18] Y. Cao, S. -Y. Lien, Y. -C. Liang, K. -C. Chen and X. Shen, "User access control in open radio access networks: A federated deep reinforcement learning approach," *IEEE Trans. Wireless Commun.*, vol. 21, no. 6, pp. 3721-3736, Jun. 2022.
[19] M. T. Nguyen, S. Kwon and H. Kim, "Mobility robustness optimization for handover failure reduction in LTE small-cell networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 4672-4676, May 2018.
[20] Y. Cao, S. -Y. Lien and Y. -C. Liang, "Deep reinforcement learning for multi-user access control in non-terrestrial networks," *IEEE Trans. Commun.*, vol. 69, no. 3, pp. 1605-1619, Mar. 2021.
[21] Y. Koda, K. Nakashima, K. Yamamoto, T. Nishio and M. Morikura, "Handover management for mmWave networks with proactive performance prediction using camera images and deep reinforcement learning," *IEEE Trans. Cogn. Commun. and Netw.*, vol. 6, no. 2, pp. 802-816, Jun. 2020.
[22] Q. Liu, G. Chuai, J. Wang and J. Pan, "Proactive mobility management with trajectory prediction based on virtual cells in ultra-dense networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 8, pp. 8832-8842, Aug. 2020.
[23] 3GPP TR 36.814: Further advancements for E-UTRA physical layer aspects, 3rd Generation Partnership Project, 2017.
[24] 3GPP TS 36.300: Evolved Universal Terrestrial Radio Access, 3rd Generation Partnership Project, 2019.
[25] K. Sun, J. Yu, W. Huang, H. Zhang and V. C. M. Leung, "A multi-attribute handover algorithm for QoS enhancement in ultra dense network," *IEEE Trans. Veh. Technol.*, vol. 70, no. 5, pp. 4557-4568, May 2021.
[26] Z. -H. Huang, Y. -L. Hsu, P. -K. Chang and M. -J. Tsai, "Efficient handover algorithm in 5G networks using deep learning," *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, Taipei, Taiwan, 2020, pp. 1-6.
[27] S. Wang and L. Zhang, "Q-learning based handover algorithm for high-speed rail wireless communications," *2023 IEEE Wireless Communications and Networking Conference (WCNC)*, Glasgow, United Kingdom, 2023, pp. 1-6.

Kai Sun (Member, IEEE) received the B.S. degree and M.S. degree in Communication Engineering and Signal and Information Processing from Inner Mongolia University (IMU), Hohhot, China, in 1999 and 2002, respectively. He received the Ph. D. degree in Circuits and Systems from Beijing University of Posts and Telecommunications (BUPT), Beijing, China in 2009. Now he is a Professor in IMU. His research interests include distributed and cooperative communications, radio network planning and optimization, radio resource allocation and scheduling.

Qingfeng Han is currently working toward the M.S. degree with the School of Electronic Information Engineering, Inner Mongolia University (IMU), Hohhot, China. His research interests include communications theory, radio resource allocation, reinforcement learning.

Zongchang Yang received the B.S. degree in Information Engineering from Shijiazhuang Tiedao University (STDU), Shijiazhuang, China, in 2019, and received the M. S. degree in Information and Communication Engineering from Inner Mongolia University, Hohhot, China, in 2023. His research interest includes mobility management and radio resource allocation in wireless networks.

Wei Huang (Member, IEEE) received the B.S. degree and M.S. degree in Communication Engineering and Signal and Information Processing from Inner Mongolia University (IMU), Hohhot, China, in 1999 and 2004, respectively. She is currently an Associated Professor in IMU. Her research interest includes radio resource allocation and scheduling, green communication, heterogeneous networks, deep learning, VLSI design, and hardware acceleration.



Haijun Zhang (Fellow, IEEE) was a Post-Doctoral Research Fellow with the Department of Electrical and Computer Engineering, The University of British Columbia((UBC), Canada. He is currently a Full Professor with University of Science and Technology Beijing, China. He is a IEEE ComSoc Distinguished Lecturer. He received IEEE CSIM Technical Committee Best Journal Paper Award in 2018, IEEE ComSoc Young Author Best Paper Award in 2017, and IEEE ComSoc Asia-Pacific Best Young Researcher Award in 2019. He serves/served as Track Co-Chair of VTC Fall 2022 and WCNC 2020/2021, Symposium Chair of Globecom 2019, TPC Co-Chair of INFOCOM 2018 Workshop on Integrating Edge Computing, Caching, and Offloading in Next Generation Networks, and General Co-Chair of GameNets 2016. He serves as an Editor of IEEE Transactions on Information Forensics and Security, IEEE Transactions on Communications, and IEEE Transactions on Wireless Communications.

Victor C. M. Leung (Life Fellow, IEEE) is a Distinguished Professor of Computer Science and Software Engineering at Shenzhen University, China. He is also an Emeritus Professor of Electrical and Computer Engineering and Director of the Laboratory for Wireless Networks and Mobile Systems at the University of British Columbia (UBC), Canada. His research is in the broad areas of wireless networks and mobile systems, and he has published widely in these areas. His published works have together attracted more than 60,000 citations. He is named in the current Clarivate Analytics list of "Highly Cited Researchers". Dr. Leung is serving on the editorial boards of the IEEE Transactions on Green Communications and Networking, IEEE Transactions on Computational Social Systems, and several other journals. He received the 1977 APEBC Gold Medal, 1977-1981 NSERC Postgraduate Scholarships, IEEE Vancouver Section Centennial Award, 2011 UBC Killam Research Prize, 2017 Canadian Award for Telecommunications Research, 2018 IEEE TCGCC Distinguished Technical Achievement Recognition Award, and 2018 ACM MSWiM Reginald Fessenden Award. He co-authored papers that were selected for the 2017 IEEE ComSoc Fred W. Ellersick Prize, 2017 IEEE Systems Journal Best Paper Award, 2018 IEEE CSIM Best Journal Paper Award, and 2019 IEEE TCGCC Best Journal Paper Award. He is a Life Fellow of IEEE, and a Fellow of the Royal Society of Canada (Academy of Science), Canadian Academy of Engineering, and Engineering Institute of Canada.