IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 16, 2019, accepted January 28, 2019, date of current version February 27, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2898205

Intelligent User-Centric Network Selection: A Model-Driven Reinforcement Learning Framework

XINWEI WANG[1], JIANDONG LI[1], (Senior Member, IEEE), LINGXIA WANG[1], CHUNGANG YANG[1], (Member, IEEE), AND ZHU HAN[2,3], (Fellow, IEEE)

[1]State Key Laboratory of ISN, Xidian University, Xi'an 710071, China
[2]Department of Electrical and Computer Engineering, University of Houston, Houston, TX 77004, USA
[3]Department of Computer Science and Engineering, Kyung Hee University, Seoul 446-701, South Korea

Corresponding author: Chungang Yang (chgyang2010@163.com)

This work was supported in part by the National Science Foundation of China under Grant 61871454, in part by the Natural Science Basic Research Plan in Shaanxi Province of China under Grant 2017JZ021, in part by the Special Financial Grant from the China Postdoctoral Science Foundation under Grant 2016T90894, in part by the Special Financial Grant from the Shaanxi Postdoctoral Science Foundation under Grant 154066, in part by the National Science Foundation of China under Grant 61671062, under Grant ISN02080001 and Grant ISN90106180001, in part by the 111 Project under Grant B08038, in part by the Shaanxi Province Science and Technology Research and Development Program under Grant 2011KJXX-40, in part by the US MURI AFOSR MURI under Grant 18RT0073, and in part by the NSF Grant CNS-1717454, Grant CNS-1731424, Grant CNS-1702850, and Grant CNS-1646607.

ABSTRACT Ultra-dense heterogeneous networks, as a novel network architecture in the fifth-generation mobile communication system (5G), promise ubiquitous connectivity and smooth experience, which take advantage of multiple radio access technologies (RATs), such as WiFi, UMTS, LTE, and WiMAX. However, the dense environment of multi-RATs challenges the network selection because of the more frequent and complex decision process along with increased complexity. Introducing artificial intelligence to ultra-dense heterogeneous networks can improve the way we address network selection today, and can execute efficient and intelligent network selection. Whereas, there still exist difficulties to be noted. On one hand, the contradiction between real-time communications and time-consuming learning is exacerbated, which can result in slow convergence. On the other hand, the black-box learning mode suffers from oscillation due to the diversity of multi-RATs, which can result in arbitrary convergence. In this paper, we propose a model-driven framework with a joint off-line and on-line way, which is able to achieve fast and optimal network selection through an alliance of machine learning and game theory. Further, we implement a distributed algorithm at the user side based on the proposed framework, which can reduce the number of frequent switching, increase the possibility of gainful switching, and provide the individual service. The simulation results confirm the performance of the algorithm in accelerating convergence rate, boosting user experience, and improving resource utilization.

INDEX TERMS Game theory, heterogeneous networks, machine learning, model-driven, network selection.

I. INTRODUCTION

Although the fifth-generation mobile communication system (5G) standardization is ongoing, the ultra-dense heterogeneous characteristic has been prominent to satisfy the explosive traffic and promise ubiquitous connectivity and smooth experience in the fifth-generation mobile communication system [1]. Ultra-dense heterogeneous networks are composed of a large number of small base stations with different radio access technologies (RATs), which takes advantage of multiple RATs, such as UMTS, LTE, WiFi, and WiMAX, and brings multiplexing gain by parallel transmission across multi-RATs [2]. Ultra-dense heterogeneous networks make the base stations closer to the terminals for an improvement in network capacity, which also facilitates access among different RATs [3], [4]. Meanwhile, according to the statistics, there will be 11.6 billion intelligent terminals by 2021, which are equipped with multiple interfaces, such as WiFi, 3G, 4G,

The associate editor coordinating the review of this manuscript and approving it for publication was Tomasz Trzcinski.

and blue-tooth [5]. With multiple interfaces, the terminals can access to different networks with different RATs.

Recently, to give full play to the advantage of ultra-dense heterogeneous networks, network selection is of particular importance and has greatly attracted attention in the academia and industry community [6]–[10]. Network selection can be performed by taking either a network-centric or user-centric approach [11]. In the network-centric approach [12]–[16], there is a central controller that chooses networks for the users to achieve the global optimum. In the user-centric approach [17]–[21], each user chooses the network by itself. The former can get better performance with global information, while the latter can provide better individual service with less overhead.

However, typical challenges for network selection caused by ultra-dense heterogeneous networks need to be noted and addressed, such as:

- **complex decision:** Due to the dense deployment of a large number of base stations with different RATs, there exists more complex network selection process for the terminals, which may result in high decision delay. Therefore, more complex network selection process is a main challenge.
- **frequent switching:** It is easy for a terminal to be closed to several base stations in the ultra-dense heterogeneous networks. Because of the time-varying channel, the received signals from different base stations by the terminal may alternatively exceed. Conventional network selection approaches based on maximum received signal strength may cause serious frequent switching among the adjacent base stations, which results in high switching delay and high signal overhead. Therefore, frequent switching is a severe challenge.
- **useless switching:** The performance of the base station to be switched is not expectable, and thus the terminal may switch frequently if it switches to a dissatisfied base station. For example, a terminal has to switch again if it switches to a heavy-load base station and suffers from worse user experience. The useless switching may result in a waste of resources and poor user experience. Therefore, the unexpected switching performance is a significant challenge.
- **individual service:** Due to the heterogeneity and diversity of users' demands, it is not satisfied for the terminals to be served with the same mode. Heterogeneous and individual service for specific users is an inevitable challenge.

These challenges make the traditional network selection approaches no longer valid and call for trustworthy ones to be developed [3]. Therefore, how to develop efficient network selection approach, especially in the ultra-dense heterogeneous networks, is an essential and urgent issue.

With the development of artificial intelligence, intelligence has been a key feature of 5G [22]. Introducing artificial intelligence to ultra-dense heterogeneous networks can improve the way we address network selection today, and can execute efficient and intelligent network selection. At the same time, different from the existing access network architecture, 5G will adopt user-centric access network architecture, which will endow terminals with stronger capacity to satisfy the diverse communication demands. Terminals with intelligence have been a pioneer for evolving artificial intelligence, and also a commercial opportunity in the industrial community. On the one hand, it has become a major trend to transfer intelligent processing capability from the cloud to the terminals which means user-centric, due to the drive of privacy, latency, and reliability. On the other hand, the heterogeneous demands of different terminals also prompt the trend from network-centric to user-centric. Thus, we focus on user-centric network selection combined with artificial intelligence in this paper.

By introducing intelligence to the terminals, typical challenges for network selection in the ultra-dense heterogeneous networks can be solved.

- By taking full use of prior information, it is realizable for the terminals to learn from the complex environment, which is able to simplify and accelerate decision process. For example, the terminal which is usually located in a fixed region can make a fast decision based on the experience it has learned in advance.
- By intelligence, the terminals can use time-varying air interface received matrix indicators (RMIs) to predict a time-varying-tolerant indicator for determining the base station with good service. In this way, the terminals can relieve the impact of time-varying channel and access more stable base station based on the predicted indicator, which can significantly reduce frequent switching. For example, the terminal can predict the link quality using multiple RMIs such as reference signal received power (RSRP), reference signal received quality (RSRQ), and received signal strength indicator (RSSI), which is more credible to be referred to perform the network selection.
- In order to make a gainful switching, the terminals can predict or expect the switching performance before executing a strategy, which is able to avoid useless switching. For example, the terminals can predict if the network to be switched is heavy-loaded, and further decide whether to connect to or not.
- Due to introducing the intelligent capability to each terminal, the network selection is implemented by each terminal based on its own preference. For example, different users can produce customized network selection strategies based on their own behavior patterns. Therefore, heterogeneous and individual service for specific users can be achieved.

Machine learning is an important part of artificial intelligence to accomplish intelligent tasks [23]–[29]. The main categories for machine learning algorithms are supervised, unsupervised, and reinforcement learning [23]. Supervised learning, such as neural networks and decision trees, requires a supervisor, what is the expected output for each input,

to guide the agent. Unsupervised learning, such as K-means, has no need for a supervisor or expected output. Reinforcement learning, such as Q-learning and actor-critic, applies a reward mechanism to reflect the interaction with the environment. The system using reinforcement learning can update itself continuously, while the systems using supervised and unsupervised learning, in general, are static.

In this paper, we focus on intelligent user-centric network selection in the ultra-dense heterogeneous networks. Although some works have been done, introducing intelligence to the terminal for network selection in the ultra-dense heterogeneous networks is still at the beginning level. We summarize the main contributions of this paper as follows.

- We propose a model-driven learning framework consisting of feature-learning, game-modeling, and strategy-learning, which jointly exploits the diversity of both game theory and machine learning. The proposed model-driven framework adopts a joint off-line and on-line way, which can overcome the challenge between real-timing communication and time-consuming learning. Based on the framework, intelligent, fast, and efficient network selection can be fully implemented by the terminal itself.
- We develop a distributed algorithm at the terminal side based on the proposed framework, in which feature-learning and game-modeling are used to assist strategy-learning so as to improve user experience and accelerate convergence rate. More specifically, terminals utilize feature-learning to mine the complex correlation between multiple indicators and link quality. In this way, the shortage of only using a single and highly stochastic indicator in the current approaches can be effectively overcome, which can reduce frequent switching. In strategy-learning, terminals can make a decision considering the link quality and the load, which can avoid terminals to access heavy-loaded base stations. Before actually performing a strategy, terminals can previously estimate the performance to avoid ineffective switching. In addition, game theory is used to achieve fast and optimal convergence.
- We evaluate the performance of the proposed algorithm in a system-level platform constructed with python and simpy. We verify that the convergence rate can be accelerated by introducing game theory, and the frequent switching can be reduced by introducing feature-learning. Also the user experience can be enhanced with the combination of feature-learning, game-modeling, and strategy-learning. Simulation results confirm that the algorithm can achieve the user average delay reduction of 1.7 ms and the resource utilization ratio improvement of 9% than Q-learning.

The paper is organized as follows. We survey the works related to intelligent user-centric network selection in Section II. Section III introduces the system model and presents a model-driven framework. Section IV firstly models the network selection problem as a non-cooperative game and prove the convergence, and then proposes an intelligent algorithm for network selection to overcome the shortage of game solution. We present the evaluation in Section V. Followed by the conclusions in Section VI.

II. RELATED WORK

In this section, we survey and discuss the works that related to intelligent user-centric network selection. We survey the works based on multiple RMIs, data-driven, and model-driven, and discuss how our proposed approach advances the state of the art.

A. RELATED WORK BASED ON MULTI-RMIs

Conventional network selection based on single-RMI in the industry community may result in frequent switching, which causes high switching delay and poor user experience. In recent researches, the network selection considers more than one RMI in order to well reflect the communication scenario. Moon *et al.* [30] study network selection with the consideration of load in cellular networks using population game. Except for the traditional SINR RMI, the proposed algorithm considers the load status broadcasted by BS and the transmitting power. Based on the practical physical layer data rate and the weight of users, the QoE of users in dense networks has been considered in [31]. The authors model the network selection as transfer-matching game to optimal the users' QoE. More notably, this paper offers a modeling approach for users' QoE. Mar *et al.* [32] study the antenna selection problem in MIMO system. To fully reflect the character of channel, the paper take CQI, mobility, SINR and received antennas into consideration. This paper adopts an adaptive fuzzy neural network method to deduce the rule of antenna selection, which enhance the intelligence and adaptation of selection. To overcome the shortage of single-RMI, Lin *et al.* [33] study network selection with one user and two based stations and are devoted to solving the frequent switching problem based on multi-RMIs. The contribution of the paper is that the authors propose a prior-free algorithm which utilizes link quality to control the handoff.

The motivation is that single-RMI is inadequate to express the actual channel quality, and thus the decision based on single-RMI is not accurate. One reason is that there exits inherent measurement error for single-RMI. Another reason is that the actual channel quality or other high-level indicators depend on multi-RMIs [30], [31]. Due to the nonlinear and complex correlation between multi-RMIs and the link quality or specific indicators, artificial neural networks or its derivation is often used to learn the correlation in [32] and [33].

Corresponding to [30]–[33], in our paper, we focus on an ultra-dense heterogeneous network with multi-agent. Moreover, we take into account the influence of both the channel link and the switching performance on frequent switching, which is more efficient to avoid frequent switching in the ultra-dense heterogeneous networks.

B. RELATED WORK BASED ON DATA-DRIVEN

Data-driven intelligent network selection approaches have been developed in the academic community [34]–[36]. Data-driven means that the approaches are achieved by roughly putting machine learning algorithm and big data together.

El Helou *et al.* [34] study the network-assisted network selection problem between two orthogonal frequency division multiple access-based RATs in the heterogeneous cellular networks. Policy iteration algorithm is used to obtain optimal policies, which can jointly improve user experience and network performance. Further, a Q-learning-based algorithm is introduced to solve the information-incomplete problem. Perez *et al.* [35] study the network selection problem between a macro base station and dense WiFi access points. A Q-learning algorithm with improved reward mechanism, taking into account the load, the duration, and the signal-to-noise-ratio, is achieved to decide the optimal selection. Both papers use Q-learning to constantly generate network selection strategy, and the data collected in the procedure of communication is used to modify the Q-learning.

Different from [34] and [35], Perez *et al.* [36] propose a cognitive framework, in which the terminals firstly learn the varying environment for establishing states and then learn an optimal strategy. Based on the proposed framework, K-nearest neighbors and Q-learning are jointly used for network selection in heterogeneous networks. The K-nearest neighbors is used for mining the character of users and network's state. What's more, rather than pre-specified, the device itself can learn and label the state autonomously. Yang *et al.* [37] adopt multiple machine learning techniques jointly to enhance the prediction accuracy of network traffic. As a conclusion, except using for decision, many machine learning techniques are also used for prediction or regression in communication [36], [37].

The data-driven approaches face the problem that the absence of theory for making the performance expectable and the result explainable. The pure black-box framework of machine learning limits the long-term development of intelligence, which makes the intelligence difficult to be standardized and productization. The contradiction between real-time communications and time-consuming learning is also non-negligible [37]. Moreover, in fact, the approaches in [34]–[36] are not perfectly fitted to the actual multi-agent scenario because the design of state for each agent is independent of others' strategies. In addition, different RATs are assumed to adopt the same mechanism and provide co-localized service, which is inconsonant with the realistic system. Different from [34]–[37], we propose a model-driven learning framework based on game theory and machine learning, which can achieve fast convergence and make the result explainable by theoretical analysis.

C. RELATED WORK BASED ON MODEL-DRIVEN

Model-driven means that the machine learning algorithms are strengthened and guided by theoretical analysis. In general, model-driven approaches can be achieved by taking either a field-experience or theory-analysis way. In the field-experience way, the model is built based on the accumulated knowledge in the field. The theory-analysis way usually uses modeling tools, such as game theory, to describe the model. Game theory is a theoretical tool to model the complex problems in complex scenarios [38], [39]. Yang *et al.* [38] use coalition game to model the cooperative behaviors among small cells. More significantly, the paper proposes an incentive mechanism to form the cooperative coalitions. In [39], the hierarchical game is adopted to depict the behavior between the resource providers and resource requesters, and solve resource allocation problem in virtualized networks. Moreover, thanks to game theory, there exists an equilibrium to expect and guarantee network performance for those algorithms.

Several model-driven intelligent network selection works have been done using game theory and machine learning [40], [41]. Naghavi *et al.* [40] firstly model the network selection problem among the users as a non-cooperative game and obtain the network selection strategy that can result in the highest gain for each selection. Then, the convergence of the game is proved with the sufficient condition that the utility function is strictly decreasing. Next, a Q-learning algorithm is used to deal with the limited information. What is noteworthy is that the state and the reward of the proposed Q-learning algorithm is guided by the non-cooperative game to achieve the optimal strategy of network selection. Nguyen *et al.* [41] develop a network selection framework and a Hart's reinforcement learning algorithm to deal with slow convergence, high overheads, and undesirable equilibrium of the current network selection algorithm. Game theory is used to guide the action of Hart's reinforcement learning algorithm to obtain optimal strategy and achieve a fast convergence rate. The game theory in [40] and [41] are used for theory-analysis. As a presentative of field-experience, Morozs *et al.* [42] propose a reinforcement learning based dynamic spectrum access algorithm and offer Q-learning a heuristic guidance with prior knowledge. In both ways, the game theory shows a good property to guarantee and accelerate the convergence of learning.

The approach in [40] assumes that different RATs adopt the same mechanism, and all the users connected a same network is assumed to get the same rate. Corresponding to [40], in our paper, we describe the distinctions among different RATs and analyze the impact of different RATs' mechanisms on users' rate. Further, different from [41], we develop a reinforcement learning with feature-assisted and game-guided framework to achieve efficient, fast, and gainful network selection.

III. SYSTEM MODEL AND FRAMEWORK

In this section, we firstly introduce the system model in the ultra-dense heterogeneous networks. Then, we present a model-driven learning framework based on machine learning and game theory.

A. SYSTEM MODEL

We consider an ultra-dense heterogeneous network with multi-RATs, such as WiFi, UMTS, LTE, and WiMAX, as shown in Fig. 1. There are N serve nodes (SNs) and M terminals. In general, the base stations with different RATs serve as SNs. Let $M_j(t)$ denote the number of terminals connected to SN j at time t, and $\sum_{j=1}^{N} M_j(t) = M$. Each terminal accesses an appropriate SN for better service. The SNs that belong to different RATs are interference-free by using different bands, and the SNs that belong to the same RAT experience interference due to the frequency reuse. In today's intelligent terminals, each RAT has its own radio chip (transmit and receive chain), but only one RAT is used to route the traffic at any specific time. Thus, each terminal uses a single RAT at any given time, and each user is equipped with one terminal.

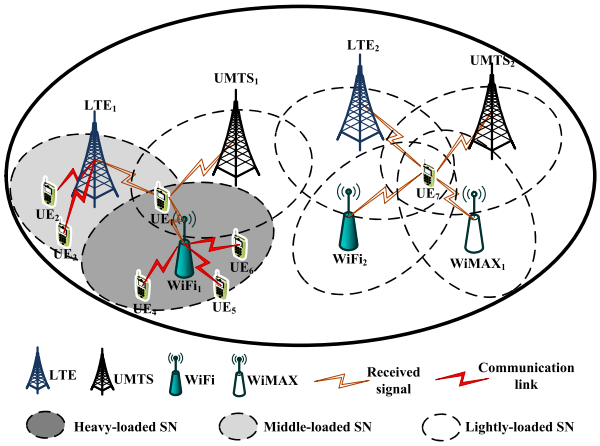


FIGURE 1. System model.

In Fig. 1, the communication link represents that the terminal is a subscriber of the associated SN. For example, UE_2 and UE_3 are the subscribers of LTE_1, and UE_4, UE_5, and UE_6 are the subscribers of $WiFi_1$. The received signal represents that the terminal is located in the server area of the SN. For example, UE_1 is located in the common server area of LTE_1, $UMTS_1$, and $WiFi_1$, and can receive signals from LTE_1, $UMTS_1$, and $WiFi_1$. UE_7 is located in the common server area of LTE_2, $UMTS_2$, $WiFi_2$, and $WiMAX_1$, and can receive signals from LTE_2, $UMTS_2$, $WiFi_2$, and $WiMAX_1$. The heavy-loaded SN, such as $WiFi_1$, represents the SN serves a lot of users. The middle-loaded SN, such as LTE_1, represents the SN serves a few users. The lightly-loaded SN, such as $UMTS_1$, represents the SN serves fewer users. In Fig. 1, UE_7 chooses an appropriate SN to connect to among LTE_2, $UMTS_2$, $WiFi_2$, and $WiMAX_1$. However, if the received signals are similar, UE_7 may frequently switch among neighboring SNs due to the time-varying channel. Moreover, UE_1 chooses an appropriate SN to connect to among LTE_1, $UMTS_1$, and $WiFi_1$. If UE_1 chooses $WiFi_1$ because of maximum received signal strength, the user experience may be worse than $UMTS_1$ due to the heavy-load of $WiFi_1$.

TABLE 1. Table of notations.

Symbol	Meaning
N	the number of SNs
M	the number of terminals
$\mathcal{M}$	the set of terminals
j	the index of the SN
i	the index of the terminal
$M_j(t)$	the number of terminals connected to SN j at time t
$U_{i,j}(t)$	the utility of terminal i connected to SN j at time t
$T_{i,j}(t)$	the throughput of terminal i connected to SN j at time t
$R_{i,j}(t)$	the instantaneous rate of terminal i connected to SN j at time t
L	the size of packet
ς	bits per resource block
z	the index of the packet
π_i	the strategy of terminal i
Π_i	the strategy set of terminal i
π	the strategy profile for all terminals
Π	the strategy profile set for all terminals
U_{i,π_i}	the utility of terminal i chosen strategy Π_i
π_{-i}	the strategy profile for all terminals except i
$U_{i,k}(t+1)$	the expected utility of terminal i on SN k at time $t+1$
$U_i^{j\to k}(t+1)$	the expected switching gain for terminal i from SN j to SN k at time $t+1$
μ	the threshold of the expected switching gain
B	the system bandwidth
$\gamma_{i,k}(t+1)$	the received SINR of terminal i on SN k at time $t+1$
ρ	the possibility for performing a switching
hv_i	the hysteresis value
X_i^z	the binary variable denoting the packet z of terminal i
$E(X_i^z)$	the expected value for approximating PSR
P	the resource block error probability
y	the number of resource block for transmitting the packet z
z^+	the number of the successful packets
z^-	the number of the failed packets
$s_i(t)$	the state of terminal i at time t
$a_i(t)$	the action of terminal i at time t
$R_i^{s,a}(t)$	the reward of terminal i at time t
$Q_i^{s,a}(t)$	the Q-value of terminal i at time t
$\boldsymbol{List_i}(t)$	the list vector of terminal i at time t
$\boldsymbol{FS_i^{List}}(t)$	the feature vector of terminal i at time t
$\boldsymbol{PSR_i^{List}}(t)$	the link quality vector of terminal i at time t
$\boldsymbol{Load_i^{List}}(t)$	the load vector of terminal i at time t
ε	the exploration rate
χ_i	the number of past consecutive concurrent switches observed by terminal i
α	the learning rate
γ_{th}	the discount factor

B. A MODEL-DRIVEN FRAMEWORK

As shown in Fig. 2, we present a model-driven learning framework with a joint off-line and on-line way, which jointly explores the diversity of both game theory and machine learning. Model-driven means applying theoretic analysis to black-box machine learning algorithm. As observed, the framework is segmented into three parts including feature-learning, game-modeling, and strategy-learning, in which feature-learning and strategy-learning are based on machine learning, and game-modeling is based on

game theory. In our framework, model-driven means applying game-modeling to strategy-learning. Feature-learning is an off-line way and strategy-learning is an on-line way. The model-driven framework with a joint off-line and on-line way has obvious advantages as follows.

- The model-driven framework with a joint off-line and on-line way is able to overcome the challenge in the current technology that the framework only depending on massive data may not be standardized and productized. The reason is that the data in the wireless communications is with small sample properties, and the black-box framework using machine learning makes it difficult to expect the performance and explain the result, which can be dealt with theoretical guidance.
- The model-driven framework with a joint off-line and on-line way is able to overcome the challenge in the current technology that the framework faces the contradiction between real-time communication and time-consuming learning. The reason is that the joint off-line and on-line way can reduce the decision delay, and the model-driven way can guarantee and speed up the convergence.
- The model-driven framework with a joint off-line and on-line way can be fully deployed and implemented at the terminal side, which can realize intelligent, fast, and efficient network selection by the terminal itself.

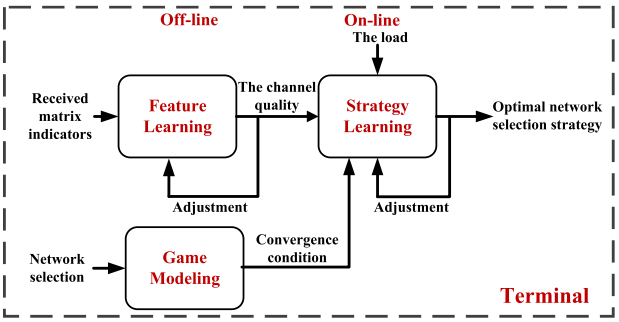


FIGURE 2. A model-driven framework.

In Fig. 2, feature-learning is implemented by supervised learning algorithms, such as random forest and neural networks. We use feature-learning to learn the link quality to deal with the frequent switching, such as UE_7 in Fig. 1. The motivation is that the link quality depends on multi-RMIs and cannot be evaluated well with single-RMI by a given equation for all cases. Moreover, the correlation between multi-RMIs and the link quality is complex and nonlinear. The input is multi-RMIs, such as RSRP, RSRQ, RSSI, signal to interference plus noise ratio (SINR), bit error rate (BER), and round-trip-time (RTT), and the output is the link quality. The terminal can adjust or modify the feature-learner when the environment changes or with a certain period. Game-modeling is used for modeling and analysing the network selection problem in a complex scenario based on game theory. We use game-modeling to model the multi-agent network selection in the ultra-dense heterogeneous networks and obtain essential convergence conditions for reaching an equilibrium. By theoretical guidance, the shortages that slow and arbitrary convergence can be solved, and fast and optimal convergence can be achieved. Strategy-learning is implemented by reinforcement learning algorithms, such as Q-learning and actor critic. We use strategy-learning to determine the optimal network selection strategy. Not only the link quality but also the SN' load is considered to make a decision, which can avoid accessing to heavy-loaded SN and resulting in poor user experience, such as UE_1 in Fig. 1. The terminal can continuously update strategy-learner.

IV. INTELLIGENT NETWORK SECTION

In this section, we firstly describe the network selection problem as a game problem, and then theoretically analyze the conditions to achieve the equilibrium. Through the analysis, we obtain a necessary convergence condition for network selection in the ultra-dense heterogeneous networks so as to guarantee the existence of equilibrium. However, if a user intently gets the best strategy using the game solution, it should obtain the information of all the users, which would lead to heavy signaling overhead. Thus, to overcome the shortage of game solution, we further propose an intelligent solution for network selection, which is guided by the equilibrium condition obtained from theoretical analysis to guarantee the convergence. That is, we use game solution to guide machine leaning solution and use machine learning solution to improve game solution.

A. NETWORK SELECTION GAME

We model the network selection problem in the ultra-dense heterogeneous networks as a non-cooperative network selection game. The players are terminals, and the strategies are the choice of SNs. At time t, for $\forall i \in \mathcal{M}$, the utility function of terminal i connected to SN j is

$$U_{i,j}(t) = T_{i,j}(t), \tag{1}$$

where $T_{i,j}(t)$ denotes the throughput of terminal i connected to SN j at time t. The throughput obtained by terminal i on SN j depends on the terminal's connected to network, the terminal-specific metric (e.g., instantaneous rate), and the other terminals sharing the same SN.

Depending on the different medium access control (MAC) protocols used by different RATs, the throughput that a user received may be different [43]. We divide the throughput models into two classes based on the MAC protocols.

1) WiFi-MODE

In this mode, the terminals who access the same network can receive the same throughput. An referable example of such MAC protocol is the distributed coordination function implemented in 802.11, in which a WiFi SN offers fair throughput to the subscribed terminals [44], [45]. In the downlink, the throughput of the subscribed terminals lies on the queuing technique used on the SN. Round robin is a

common technique to be used [11]. Therefore, the obtained throughput $T_{i,j}(t)$ of terminal i connected to SN j at time t is denoted as

$$T_{i,j}(t) = \frac{L}{\sum_{i'=1}^{M_j(t)} \frac{L}{R_{i',j}(t)}} = \frac{1}{\sum_{i'=1}^{M_j(t)} \frac{1}{R_{i',j}(t)}}, \tag{2}$$

where L is the size of packet. i' denotes any terminal. $R_{i,j}(t)$ is the instantaneous rate of terminal i connected to SN j at time t. The rate $R_{i,j}(t)$ lies on the channel condition and the selected modulation and coding scheme at time t. $M_j(t)$ is the number of terminals connected to SN j at time t. The mode is suitable for modeling the RATs with fair throughput, such as WiFi.

2) CELLULAR-MODE

In this mode, the throughput obtained by the terminals may be different and depends on the number of terminals accessed to the same SN. The examples of such MAC protocol are orthogonal frequency division multiple access and proportional fair scheduling [11], [46]. $T_{i,j}(t)$ is denoted as

$$T_{i,j}(t) = \frac{R_{i,j}(t)}{M_j(t)}. \tag{3}$$

The mode is applied to model time/frequency/time-frequency fair RATs, such as 3G, 4G, and WiMAX.

If the terminal accesses the WiFi network, the obtained utility is based on (2). Otherwise the utility is based on (3) when accessing the 3G/4G/WiMAX network.

We denote the strategy of terminal i as π_i, and the strategy set as Π_i. The strategy profile for all terminals is $\pi = (\pi_1, \pi_2, \ldots, \pi_i, \ldots, \pi_M)$ and the strategy profile set is Π.

Definition 1: We define the obtained utility as U_{i,π_i} if terminal i chooses strategy Π_i. Given the strategies of other terminals, each terminal chooses the best strategy. If the strategy profile π^* is a Nash equilibrium, we have

$$U_{i,(\pi_i^*,\pi_{-i}^*)} \geq U_{i,(\pi_i,\pi_{-i}^*)}, \quad \forall i \in M,\ \pi_i \in \Pi_i, \tag{4}$$

where π_{-i} is the strategy profile for all terminals except terminal i.

At any specific time, the rational or best response strategy of the terminal is to switch the connection to a SN that results in a higher utility. In order for terminal i to make a switching at time $t+1$ from SN j to SN k, we definite the expected switching gain as $U_i^{j\to k}(t+1)$. Then, we have

$$U_i^{j\to k}(t+1) = \frac{U_{i,k}(t+1)}{U_{i,j}(t)}, \tag{5}$$

where $U_{i,j}(t)$ is the obtained throughput of terminal i connected to SN k at time t. $U_{i,k}(t+1)$ is the expected throughput of terminal i switching to SN k at time $t+1$. $U_i^{j\to k}(t+1)$ denotes whether the switching can bring higher throughput or not. The terminal would perform a switching when the throughput after switching is higher than the one before switching. Thus, the expected gain $U_i^{j\to k}(t+1)$ should satisfy

$$U_i^{j\to k}(t+1) \geq \mu, \tag{6}$$

where μ is a threshold and $\mu \geq 1$.

The expect throughput $U_{i,k}(t+1)$ is roughly computed based on (2) or (3). Based on [47], we estimate the instantaneous rate of terminal i connected to SN k at time $t+1$ by

$$R_{i,k}(t+1) = B\log_2\left(1+\gamma_{i,k}(t+1)\right), \tag{7}$$

where B is the system bandwidth, $\gamma_{i,k}(t+1)$ is the received SINR of terminal i on SN k at time $t+1$. The same way is applied to estimate the instantaneous rate of other terminals. The 802.11u and access network discovery and selection function (ANDSF) are used to obtain the information on the number of terminals on other RATs and their instantaneous rates [48]. The WiFi Alliance proposes the Hotspot 2.0 standard based on 802.11u, which provides a mechanism to inform the terminal of information without requiring the terminal to contact the SN. The ANDSF is deployed in the core network and communicates with the terminals through specific interface. Thus, each terminal can evaluate its expected throughput if it determines to make a switching.

To make the switching more efficient, the expected throughput should be closed to the available throughput. However, if multiple terminals switch to the same SN at the same time, the expected throughput and the available throughput may be greatly different, which may cause the situation that the received signal is strong while the user experience is poor. Thus, we consider to minimize the number of concurrent switches to a SN. We consider the terminal switch with probability ρ, which depends on the congestion in the network and acts similarly to the 802.11 contention window mechanism. Similar to the binary exponential backoff in the 802.11 distributed coordination function, when terminal i observes that concurrent switches to a SN happens, it sets its probability as

$$\rho = \rho^{\chi_i}, \tag{8}$$

where χ_i is the number of past consecutive concurrent switches observed by terminal i.

B. CONVERGENCE TO GAME EQUILIBRIUM

In this section, we first analyze the convergence of the network selection game in which all SNs belong to WiFi-mode. Next, we consider the convergence of the network selection game in which all SNs belong to cellular-mode. Then, we discuss the convergence of the network selection game in which all SNs are with a mixture of the WiFi-mode and the cellular-mode.

1) CONVERGENCE WITH SINGLE-MODE

Theorem 1: The non-cooperative network selection game based on WiFi-mode can converge to a Nash equilibrium.

Proof: The proof is based on [49]. For simplicity, we use U_i as the utility of terminal i. We sort the utilities of all terminals as

$$U_1 \leq U_2 \leq \ldots \leq U_i \leq \ldots \leq U_M. \tag{9}$$

The users connected to the same SN can obtain the same throughput, while the users connected to different SNs may receive different throughput. We define a function G on the sorted utility as

$$G = U_1 \times S^{M-1} + U_2 \times S^{M-2} + \ldots + U_i \times S^{M-i} + \ldots + U_M, \tag{10}$$

where $S \to \infty$, and satisfy $S \gg U_i, \forall i \in M$. Assuming that terminal i makes a switching from SN j to SN k, the throughput of all the users in both SN j and SN k will vary. The throughput of all the users in SN j increase due to the leaving of terminal i, while the ones in SN k decrease due to the accessing of terminal i. However, the obtained throughput for terminal i on SN k should be higher than the one on SN j; otherwise the switch will not occur.

For further explaining it, we provide an illustrative example. We consider a network with 2 SNs and 5 terminals, in which terminal 1, terminal 2, and terminal 3 are the subscribers of SN j, and terminal 4 and terminal 5 are the subscribers of SN k. We sort the utilities as $U_1 = U_2 = U_3 < U_4 = U_5$, and thus we have

$$G = U_j \times \left(S^4 + S^3 + S^2\right) + U_k \times \left(S^1 + 1\right). \tag{11}$$

If terminal 3 makes a switching from SN j to SN k, the conditions $U_1 = U_2 < U_3{}' = U_4{}' = U_5{}'$, $U_j{}' > U_j$, $U_k{}' < U_k$, $U_k{}' > U_j$, and $G' = U_j{}' \times \left(S^4 + S^3\right) + U_k{}' \times \left(S^2 + S^1 + 1\right)$ should be satisfied. Thus, we have

$$G' - G > 0. \tag{12}$$

Therefore, G is strictly increasing. Since the number of terminals and SNs are limited, G does not indefinitely increase and would be steady finally, which means all switchings would terminate at some points. Since the users can no longer increase its utility by unilaterally changing its strategy, the terminate is a Nash Equilibrium. ■

Theorem 2: The non-cooperative network selection game based on cellular-mode can converge to a Nash equilibrium.

Proof: The proof is based on contradiction and is followed the one in [11]. We assume that a loop exists in the network selection process, which means that the initial state is equal to the end state. Between any two consecutive states, the throughput inequality denotes that the next throughput is larger than the previous throughput. Due to the characteristic of loop, whenever a user leaves a network, he will come back to this network later. If we collect all the throughput inequalities within the cycle and multiply them together, all terms will cancel each other. Then, we have $1 > 1$. Because a loop state sequence can not exist, the game would terminate at an equilibrium and no terminal can obtain higher utility by unilaterally changing its strategy. If the terminal could, the state would alter and it would not be an equilibrium. Therefore, it is a contradiction and there exists a Nash equilibrium. ■

2) CONVERGENCE WITH MIXED-MODE

Unlike the single-mode network, there exists infinite oscillation in the mixed-mode network [18]. Based on [18], the hysteresis mechanism is introduced to solve the infinite oscillation phenomenon and guarantee the convergence.

Definition 2 (Hysteresis Mechanism): Suppose that terminal i makes a switching from a mode of SNs to another mode of SNs. In order for terminal i to switch back to a SN in the previous mode, the expected throughput should be greater than the corresponding hysteresis value. The hysteresis value hv_i of terminal i in a given mode depends on its last achieved throughput in that mode prior to switching to a different mode of SNs.

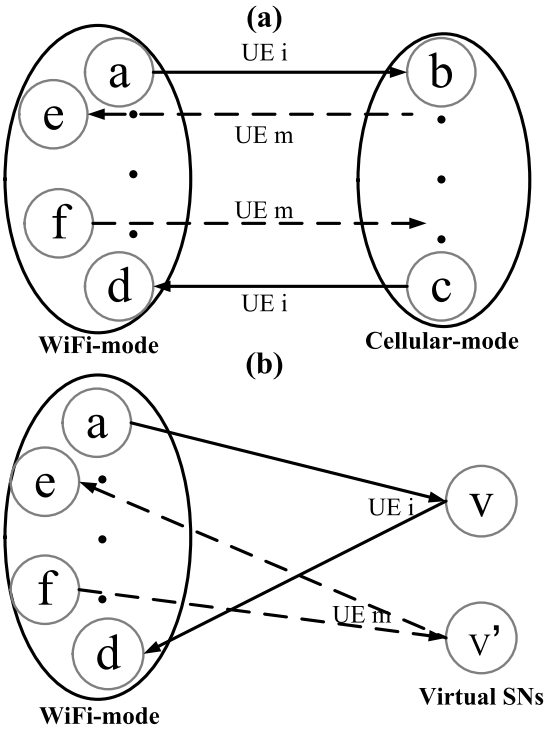


FIGURE 3. Network selection in the mixed-mode networks (a) and establishing the virtual SNs (b).

We classify all SNs into two modes based on the throughput modes, as shown in Fig. 3 (a). Terminal i switches from the WiFi-mode to the cellular-mode, after performing a series of switches within the cellular-mode, if terminal i wants to switch from the cellular-mode to the WiFi-mode, the following condition should be satisfied to perform switching.

$$U_{i,d} > U_{i,c} \& U_{i,d} > U_{i,a}, \tag{13}$$

where $U_{i,d}$ denotes the throughput of terminal i connected to SN d. The same meanings to $U_{i,c}$ and $U_{i,a}$. Here $U_{i,a}$ is the hysteresis value hv_i of terminal i in the Wifi-mode, which is defined in Definition 2.

Theorem 3: Based on the hysteresis mechanism described in Definition 2, a non-cooperative network selection game with a mixture of the WiFi-mode and the cellular-mode can converge to a Nash equilibrium.

Proof: The proof is based on contradiction [18]. The set of SNs and connected terminals are defined as the system state of the network. Assuming that there is an infinite loop in the system state evolution, which means that the initial state is equal to the end state at intervals. Assuming that the loop starts when a terminal switches from a mode to another mode. UE i and UE m in Fig. 3 (a) are such terminals that switch from one mode to another mode, and returns back to the previous mode after a while (to form a loop). Consider the second iteration of this loop. Due to the repetition, the terminals have historical knowledge in both classes.

We first assume that there is one virtual SN for each departure and return of any terminal in the WiFi-mode. The virtual SN serves only one specific terminal, and provides an average throughput which is equal to the average value of before leaving the WiFi-mode and immediately after returning to the WiFi-mode. For example, in Fig. 3 (b), the virtual SN v serves the terminal i and provides the throughput $U_{i,v} = \frac{U_{i,a}+U_{i,d}}{2}$. Based on the hysteresis mechanism, we have

$$U_{i,d} > U_{i,v} > U_{i,a}, \tag{14}$$

where terminal i connected SN v obtains gain after leaving SN a, and also obtains gain after leaving SN v.

We then consider terminal m switching from the cellular-mode to the WiFi-mode. Due to the loop, terminal m switches back to the cellular-mode after a period of time. Since we consider the second iteration of the loop, terminal m has historical knowledge of the WiFi-mode. Thus, we assume terminal m switches to the WiFi-mode from a virtual SN v', which provides the throughput $U_{i,v'} = \frac{U_{i,e}+U_{i,f}}{2}$. Based on the hysteresis mechanism, we have

$$U_{i,f} > U_{i,v'} > U_{i,e}, \tag{15}$$

where terminal m connected SN v' obtains gain after leaving SN e, and also obtains gain after leaving SN v'.

Each virtual SN only accommodates one terminal, and thus it can belong to either the WiFi-mode or the cellular-mode. We can assume all virtual SNs belong to the WiFi-mode. Consider the WiFi-mode and all virtual SNs, we can conclude that there is a loop in the WiFi-mode. However, there is no loop in the single-mode network, which has been proved in Section IV-B.1. Thus, it is a contradiction. The non-cooperative network selection game with a mixture of the WiFi-mode and the cellular-mode can converge to a Nash equilibrium. ■

C. INTELLIGENT RFEQG ALGORITHM

As mentioned before, we model the multi-agent network selection as a non-cooperative game, as presented in Section IV-A, and then analyze the Nash equilibrium in Section IV-B. However, it should obtain the information of all the users, which would lead to heavy signaling overhead. In this subsection, we use the convergence conditions got from game theory to strengthen traditional Q-learning and tackle the network selection problem. Meantime, the Q-learning overcomes the shortage of game theory. Based on the proposed framework in Section III, we firstly elaborate feature-learning and strategy-learning, and then propose RFEQG to provide good service with less signaling overhead.

1) FEATURE-LEARNING

Network selection in the industrial community is mainly based on max-RSRP, which causes inherent inaccuracy [33]. If the received RSRP of a terminal from different SNs are closed and may alternatively exceed due to the time-varying channel, the terminal may frequently switch among different SNs and result in high handover delay and a waste of network resources. The reason is that the actual link quality depends on multi-RMIs besides RSRP. However, the associated function is nonlinear and complex. Thus, we utilize machine learning to mine the nonlinear correlation between the link quality and multi-RMIs, such as RSRP, RSRQ, RSSI, SINR, and BER.

We use packet success ratio (PSR) as the estimation of the link quality. We define PSR as the probability of successfully transmitted packets to the total transmitted packets. The size of each packet is L, and each packet is transmitted on multiple resource blocks. Only when all resource blocks used to transmit one packet are successful, the packet is transmitted successfully. The error-transmitting probability of resource block is denoted as p. Let $X_i^z(t) = 1$ denotes the packet z of terminal i is successfully transmitted at time t, or failed if $X_i^z(t) = 0$. Due to the transmission of resource blocks obeying independent identical distribution, we can approximate PSR with the expected value $E\left(X_i^z\right)$ based on the weak large number law [33], which can be formulated as

$$E\left(X_i^z\right) = (1-p)^y, \tag{16}$$

where y is the number of resource blocks for transmitting the packet z and p is the error-transmitting probability of resource block. y depends on L and the bits per resource block, which is given by

$$y = \frac{L}{\varsigma}, \tag{17}$$

where ς denotes bits per resource block, which depends on the selected modulation and coding scheme.

To measure PSR in a realistic system, we construct a practical platform with python. In this platform, we count that the number of the successful packets of user i every 10 ms is z^+, and the number of the failed packets is z^-. Thus, $E(X_i)$ can be computed based on the weak large number law by

$$E(X_i) = \frac{z^+}{z^+ + z^-}. \tag{18}$$

Random forest (RF) [50] is chosen to learn the complex correlation between multi-RMIs and PSR because of its fast

decision rate, less computational resource, and good adaptability for small-sample learning. RF belongs to a bagging algorithm of ensemble learning [51], which can be used to deal with classify or regression problems. RF is an ensemble technique that trains several classifiers by random-back-sampling data on the original data set. Then, it uses the set of trained classifiers to classify the new samples, and obtains the final result from all classifiers using the majority votes or the mean. In this way, it is significantly enhanced in accuracy and generalization for the trained model comparing with one single decision-making model.

Classification and regression tree (CART) [52] is chosen as decision tree to generate several classifies in RF. The process of constructing a RF is roughly as follows: 1) Generate the training data set by multiple sampling from the original data set with a random-back-sampling method. 2) Train decision tree model using CART. 3) Split decision tree using the information gain ratio or Gini index. 4) Form RF using the multiple decision trees. The accuracy of RF mainly depends on its parameters, including the number of decision trees, the maximum number of features, the minimum number of samples, the ratio of train set to total data set, and the size of total data set. Different settings have different influences on the accuracy of RF. The RF algorithm for feature-learning is detailedly showed in Algorithm 1.

Algorithm 1 RF Algorithm for Feature-Learning

Input: i, $\boldsymbol{List_i}$, $\boldsymbol{FS_i^{List}}$
Output: $\boldsymbol{PSR_i^{List}}$
1: Initialization.
2: **while True**
3: Obtain the trained RF learning machine f_i
4: Compute $\boldsymbol{PSR_i^{List}}$ based on (21) using f_i constructed by CARTs
5: **end while**

When the terminals have a request for transmitting traffic, the terminals firstly detect and get the service list vector which is composed of the adjacent and serviceable SNs. The list vector can be gotten by ANDSF which can inform the terminals of the information about LTE/WiFi/WiMAX by the way of pull or push. The list vector $\boldsymbol{List_i}(t)$ of terminal i at time t can be denoted as

$$\boldsymbol{List_i}(t) = [1, 2, \ldots, j, \ldots]. \tag{19}$$

Supposing that D-dimensional RMIs are considered to estimate PSR, and thus the feature vector $\boldsymbol{FS_i^{List}}(t)$ of terminal i at time t can be denoted as

$$\boldsymbol{FS_i^{List}}(t) = \{\boldsymbol{x_1}(t), \boldsymbol{x_2}(t), \ldots, \boldsymbol{x_d}(t), \ldots, \boldsymbol{x_D}(t)\}, \tag{20}$$

where $\boldsymbol{x_d}(t)$ is the dth feature vector at time t in the feature space, and $\boldsymbol{x_d}(t) = \left(\ldots, x_d^k(t), \ldots x_d^j(t), \ldots\right)^T$. For example, $x_d^k(t)$ and $x_d^j(t)$ can represent the received RSRP from SN k and SN j by terminal i at time t, respectively. The terminals obtain the RMIs by pilot signal measurement for 3G/4G SNs and carrier sense for WiFi SNs. Therefore, we construct input-output as

$$\boldsymbol{PSR_i^{List}}(t) = f_i\left(\boldsymbol{FS_i^{List}}(t)\right), \tag{21}$$

where $\boldsymbol{PSR_i^{List}}(t)$ is the link quality vector for the SNs in the $\boldsymbol{List_i}(t)$ estimated by terminal i at time t. f_i is the RF-learner trained and used by terminal i.

2) STRATEGY-LEARNING

At time t, each terminal generates the optimal network selection based on strategy-learning. Q-learning [53] is chosen to implement strategy-learning due to its low complexity and efficiency. Q-learning belongs to reinforcement learning, which enables the agent to decide the optimal action from its own experience. However, traditional Q-learning is not suitable for the network selection problem we address in the ultra-dense heterogeneous networks. On the one hand, traditional Q-learning may be slow and arbitrary convergence, and even may be not workable in the multi-agent scenario. This is because it is fit for the situation that the states of agents are independent of each other, or the situation that there is only one agent. While the states of agents we consider in our paper are dependent of each other. Generally, game theory and sharing Q-table are main ways to deal with the multi-agent scenario. On the other hand, there exists infinite oscillation in the heterogeneous mixed-mode networks [18], which also makes traditional Q-learning misconvergence. We have discussed and analyzed in Section IV-A and Section IV-B that game theory can be used to avoid infinite oscillation and enable the mixed-mode network selection game converge. Based on the hysteresis mechanism and selfish behavior deduced from the game theory, we develop an enhanced Q-learning with game theory (EQG) algorithm as shown in Algorithm 2.

In Algorithm 2, the agent is terminal i, $\forall i \in \mathcal{M}$. Each terminal decides the network selection based on not only the link quality $\boldsymbol{PSR_i^{List}}$ but also the SN' load $\boldsymbol{Load_i^{List}}$. The state of terminal i at time t can be denoted as

$$s_i(t) = \left(\boldsymbol{List}(t), \boldsymbol{PSR_i^{List}}(t), \boldsymbol{Load_i^{List}}(t)\right), \tag{22}$$

where $\boldsymbol{Load_i^{List}}(t) = [n_1, n_2, \ldots, n_j, \ldots]$ is the load vector of the SNs in $\boldsymbol{List_i}(t)$. The load can be gotten by terminal i based on ANDSF and Hotspot 2.0 in 802.11u standard.

The action of terminal i at time t is to choose an appropriate SN from the service list. Thus, the action can be showed as

$$a_i(t) = \boldsymbol{List}(t), \tag{23}$$

where $a_i(t) = j$ denotes that terminal i chooses SN j from $\boldsymbol{List}(t)$ to connect to at time t.

The reward of terminal i at time t is the achieved throughput after performing $a_i(t)$, and

$$Re_i^{s,a}(t) = U_i(t), \tag{24}$$

where $U_i(t)$ denotes the achieved throughput by terminal i after performing $a_i(t)$ at time t. If terminal i connects to SN j

Algorithm 2 EQG Algorithm for Strategy-Learning

Input: $i, \boldsymbol{List_i}, \boldsymbol{RSRP_i^{List}}, \boldsymbol{Load_i^{List}}$
Output: a_i

```
Initialization: Q_i, hv_i, s_i
while True
  if rand () > ε, ε is the exploration rate
    Select a_i randomly.
  else
    Select a_i = arg max_{a_i} Q_i (s_i, a_i).
  end if
  if U_{i,a_i} / U_{i,j} > μ, j is the last action
    if class (a_i) = class (j)
      if rand () < ρ^{χ_i}
        a_i = a_i
        if Concurrency happens
          χ_i = χ_i + 1
        else
          χ_i = 0
      else
        a_i = j
    else
      if U_{i,a_i} > hv_i [class (a_i)]
        if rand () < ρ^{χ_i}
          hv_i [class (j)] ← U_{i,j}
          a_i = a_i
          if Concurrency happens
            χ_i = χ_i + 1
          else
            χ_i = 0
        else
          a_i = j
  else
    Q_i (s_i, a_i) = 0
    a_i = j
  end if
  Execute a_i, obtain R_i^{s,a}, observe s_i'
  Calculate Q_i^{s,a}
  Calculate Q_i^{s,a}
  s_i ← s_i'
end while
```

at time t, the achieved throughput $Re_i^{s,a}(t) = U_{i,j}(t)$, where $U_{i,j}(t)$ are based on (2) with the WiFi network and (3) with the 3G/4G/WiMAX network.

The state-action value function is computed based on

$$Q_i^{s,a}(t) = (1-\alpha)\, Q_i^{s,a}(t) + \alpha \left[Re_i^{s,a,s'}(t) + \gamma \max_{a_i(t+1)} Q_i^{s',a'}(t+1) \right], \quad (25)$$

where s' is the next state after performing a, and a' denotes all actions when staying state s'. α is the learning rate, and γ is the discount factor.

Firstly, terminal i initializes its Q-table, hysteresis vector, and state. The hysteresis vector of terminal i consists of the hysteresis value of terminal i in the WiFi-mode and the cellular-mode, respectively. Then, from Line 3 to Line 7, the terminal choose the action according to ε-greedy algorithm. Next, the terminal makes a estimation for the selected action. Due to the selfish user, in Line 8 and Line 29, the user estimates whether the switching can result in a higher utility or not. In Line 9, Line 19, and Line 21, hysteresis mechanism stated in Definition 2 is adopted. From Line 10 to Line 17 and Line 20 to Line 28, probabilistic switching is considered to avoid concurrency. Terminal i firstly judge whether the selected SN belongs to the same mode with the currently accessing SN. If it is, terminal i observes whether concurrency happens and determine the final action. If it is not, terminal i first judge whether the hysteresis mechanism is satisfied, and then observes whether concurrency happens and determine the final action. In Line 30, Q-value is set to 0 to penalize the action with bad utility, which can speed up the convergence rate.

3) RFEQG ALGORITHM

We propose a distributed RFEQG algorithm combining feature-learning, game theory, and strategy-learning, as shown in Algorithm 3, which is performed by each terminal. Firstly, by accurately estimating the link quality with Algorithm 1, the frequent switching and the switching delay can be reduced. Thus, the user experience can be improved. Next, based on Algorithm 2, not only the link quality but also the load is considered to make a decision. Moreover, the terminals perform switching that results in a higher utility by expecting the utility before executing the strategy. The exponential back-off mechanism is used to lessen the concurrent switching, which avoids network congestion and improves service quality. In addition, we also taking the hysteresis mechanism inferred from game theory into account to guarantee the optimal convergence and accelerate the convergence rate. Based on Algorithm 3, user experience can be significantly enhanced.

Algorithm 3 RFEQG Algorithm for User-Centric Network Selection

Input: $i, \boldsymbol{List_i}, \boldsymbol{FS_i^{List}}, \boldsymbol{Load_i^{List}}$
Output: a_i

```
Initialization
while True
  Perform Algorithm 1
  Perform Algorithm 2
end while
```

The proposed RFEQG algorithm consists of feature learning and strategy learning, and thus we analyze the complexity of the two parts, respectively. In fact, it is relatively hard to theoretically analyze the complexity of RF and Q-learning duo to the uncontrollable convergence. Therefore, we adopt the way of qualitative analysis to clarify the complexity. The complexity of feature-learning mainly appears on the off-line training of RF, which may need not to put too much

concentration because it can be achieved using server. As for the on-line operation in algorithm 1, it is just a series of if-else judgements which possesses low complexity. In strategy learning, we use game theory to speed up the learning and convergence rate of strategy. In algorithm 2, the related operations with game theory are line 8, line 9, line 18, line 19, and line 21. We can see there only exist some simple if-else judgements, and the terminal only should store two scalar values when introducing hysteresis mechanism. Besides game theory, we also use Q-learning to implement strategy-learning due to its low complexity and high efficiency. We simulate the convergence rate of strategy learning in Fig. 7 and Fig. 9 in Section V, in which we can observe that our proposed strategy learning achieves lower complexity than pure Q-learning. Therefore, on the whole, the proposed RFEQG is with acceptable complexity.

V. SIMULATION RESULTS AND DISCUSSIONS

In this section, we evaluate and demonstrate the effectiveness of the proposed algorithm. We consider an ultra-dense heterogeneous network with two kinds of different RATs (LTE and WiFi) in a square area of 80*80 meters. There are 4 SNs (2 LTE SNs and 2 WiFi SNs) and 15 terminals in the square area. Each user is equipped with a terminal.The users are distributed randomly within the square area. In another words, average three to four users are covered with one BS. To completely and particularly show the independent intelligence and behavior of every terminal in the simulation, we focus on a part of the whole ultra-dense network. From another perspective, the LTE SNs' path loss model is simulated as COST 231 Walfish-Ikegami model, which reflect the non-light-of-sight propagation of the SNs, and the transmitter of SNs is 30 dBm low power transmitter. The path loss model is denoted as $PL\,(dB) = -35.4 + 26\log_{10}(d) + 20\log_{10}(f_c)$. The unit of d is meter and f_c is the system frequency. The simulation parameters are detailedly described as Table II. The basic configurations are based on 3GPP TR36.814.

We use python and simpy to construct a simplified system-level simulation platform in order to evaluate and verify the performance of the intelligent network selection algorithm. In our platform, we use simpy to implement discrete-event simulation and construct multi-agent asynchronous networks. The traffic arrival time interval of terminals obeys exponential distribution. Every 10 ms, the terminals with traffic to be transmitted trigger network selection and choose the best SN to access. Then, the SN performs power allocation among all resource blocks. Next, the SN determines modulation scheme, code rate, etc.. We design two schedulers based on (2) and (3), i.e., LTE scheduler and WiFi scheduler, to allocate resource block for terminals every 1 ms. Finally, we evaluate the transmission performance with some metrics such as the resource block error probability. We use the proposed RFEQG algorithm to accomplish the network selection of the platform.

TABLE 2. Table of simulation parameters.

Parameter	Value
RATs	LTE, WiFi
The number of SNs N	4
The number of terminals M	15
The arrival interval of service	Exponential distribution with $\lambda = 0.1$
System bandwidth B	20 MHz
System frequency f_c	2.6 GHz
Noise density	-176 dBm/Hz
The maximal power of each SN	30 dBm
Path loss model	COST 231 Walfish-Ikegami model
Fast fading	Rayleigh fading
Exploring rate ε	0.9
Switching gain μ	1
Switching probability ρ	0.9
Learning rate α	0.9
Discount factor γ_{th}	0.9

A. CONVERGENCE

Fig. 4 shows the convergence of the proposed RFEQG algorithm. We arbitrarily select 4 terminals from 15 terminals as an illustration, including terminals 2, 3, 10, and 11. The simulation is 5000 ms. The X-axis is simulation times, and the Y-axis is network selection strategy. SN 0 and SN 2 are LTE SNs, while SN 1 and SN 3 are WiFi SNs. We can observe that the number of connecting to 4 SNs is almost the same when the algorithm does not converge. The number of connecting to non-optimal SNs is in close proximity to 0 when the algorithm gradually converges. Meanwhile, the terminal can steadily connect to the optimal SN. For example, the optimal SN for terminal 2 is SN 3. Before 700ms, the algorithm gradually converges, and the number of accessing 4 SNs for terminal 2 is almost the same. After 700ms, the algorithm gradually converges, the number of accessing SN 0, SN 1, and SN 2 for terminal 2 drastically decreases and approached 0, and the number of accessing SN 3 for terminal 2 distinctly increases and is stable for a long time. We can see that terminal 2 occasionally accesses other SNs besides SN 3. This is because our algorithm considers exploration and exploitation mechanism to avoid getting into the local optimal solution. If other SNs still provide worse service, terminal 2 can quickly converge to the optimal SN again. The analysis and conclusion apply to any other terminals.

B. PERFORMANCE EVALUATION

The accuracy of the RF algorithm which is used to evaluate the link quality PSR is presented in Fig. 5. In the simulation, we consider RSRP, SINR, BER, and the type of SNs as the RMIs to evaluate PSR. The accuracy using RF to evaluate PSR mainly depends on the number of decision trees, the maximum number of features, the minimum number of samples, the ratio of train set to total data set, and the size of total data set. We initially set the number of decision trees as 100, the maximum number of features as 2, the minimum number of samples as 50, and the ratio of train set to total data set as 0.8. The size of total data set is 20, 000 in Fig. 5 (a), (b), (c), and (d). Fig. 5 (a) shows that different numbers of decision trees have different influence on the accuracy.

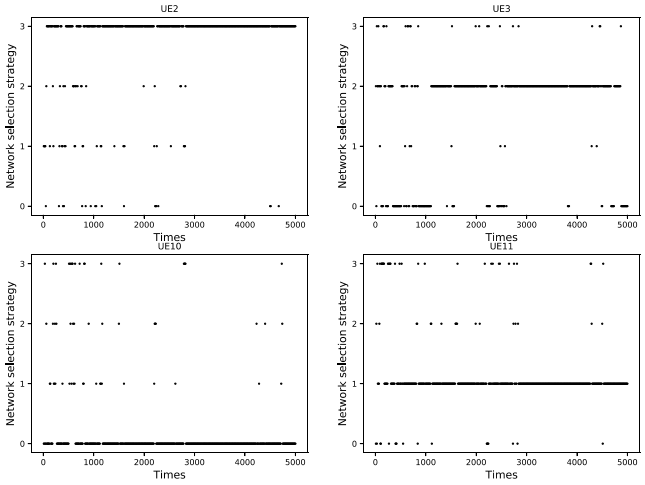


FIGURE 4. The convergence of the proposed RFEQG algorithm.

(a)

RF parameters	Value	Accuracy
The maximum number of features	2	
The minimum number of samples	50	
The ratio of train set to total data set	0.8	
The number of decision trees	20	0.808706
	70	0.808863
	120	0.809546
	200	0.809651
	220	0.809336

(b)

RF parameters	Value	Accuracy
The number of decision trees	200	
The minimum number of samples	50	
The ratio of train set to total data set	0.8	
The maximum number of features	1	0.809283
	2	0.809231
	3	0.809966
	4	0.806868

(c)

RF parameters	Value	Accuracy
The number of decision trees	200	
The maximum number of features	3	
The ratio of train set to total data set	0.8	
The minimum number of samples	30	0.806238
	40	0.809073
	50	0.809388
	60	0.808601

(d)

RF parameters	Value	Accuracy
The number of decision trees	200	
The maximum number of features	3	
The minimum number of samples	50	
The ratio of train set to total data set	0.2	0.785386
	0.6	0.807545
	0.7	0.809423
	0.8	0.809966
	0.9	0.807603

(e)

RF parameters	Value
The number of decision trees	200
The maximum number of features	3
The minimum number of samples	20
The ratio of train set to total data set	0.8

(f)

The size of total data set	Accuracy
20000	0.809966
50000	0.841332
200000	0.877944

FIGURE 5. The accuracy of RF algorithm versus different RF parameters.

We can see that the highest accuracy is 0.809651 when the number of decision tree is 200. Fig. 5 (b) shows that the highest accuracy is 0.809966 when the maximum number of features is 3. Fig. 5 (c) shows that the highest accuracy is 0.809388 when the minimum number of samples is 50. Fig. 5 (d) shows that the highest accuracy is 0.809966 when the ratio of train set to total data set is 0.8. Thus, the RF parameters are set as Fig. 5 (e) to achieve the highest accuracy. Based on the setting in Fig. 5 (e), Fig. 5 (f) simulates the correlation of the accuracy and the size of total data set. We can observe the highest accuracy is achieved when the size of total data set is 200, 000. Thus, the RF we use in the following is based on the parameters in Fig. 5 (e) and the size of total data set is 200, 000.

We confirm the effectiveness of feature-learning with the RF algorithm in Fig. 6. We select 2 terminals from 15 terminals as an illustration, in which terminal 12 frequently switches among 4 SNs, especially between SN 0 and SN 3, and terminal 1 is normal without frequent switching. The simulation is 60, 000 ms. The X-axis is simulation times, and the Y-axis is network selection strategy. UE12 (with EQG) and UE1 (with EQG) indicate that the EQG algorithm is used, while UE 12 (with RFEQG) and UE 1 (with RFEQG) indicate that the RFEQG algorithm is used. Compared with UE 12 (with EQG), we can observe that the frequent switching is obviously reduced for UE 12 (with RFEQG). This is because by accurately estimating PSR, SN 0 can provide better service compared with SN 3. In addition, we can see that the number of accessing SN 1 and SN 2 is also obviously reduced, and terminal 12 can keep connection to SN 0 for a long time. Thus, by considering feature-learning with RF, the frequent switching for terminals can be drastically avoided. We can also find that feature-learning with RF can reduce the number of accessing non-optimal SNs and accelerate the convergence rate by comparing UE 1 (with EQG) with UE 1 (with RFEQG). It can be concluded that feature-leaning with RF has the ability of reducing frequent switching, maintaining steady access to the optimal SN, and accelerating the convergence rate. Therefore, the aims can be achieved with feature-learning, including cutting down the switching delay, improving the quality of service, and lessening the resource waste.

Further, we represent the average delay of the user in Fig. 7. The average delay means the statistic average of times and users. The X-axis is simulation times, and the Y-axis is average delay of the user. The simulation is 60, 000 ms, and we make a statistic at intervals of 6, 000 ms. We compare the performance of RFEQG algorithm with Q algorithm, enhanced Q-learning without game (EQWG) algorithm, EQG algorithm. The Q algorithm represents traditional Q-learning. The EQWG algorithm takes into account the switching effect and the concurrent switching, but without the hysteresis mechanism obtained from game-modeling. The EQG algorithm involves the switching effect, the concurrent switching, and the hysteresis mechanism. The RFEQG algorithm involves the feature-learning, the switching effect, the concurrent switching, and the hysteresis mechanism. The delay consists of the queue delay and the transmission delay. It is easy to understand that the delay decreases with the growth of simulation times, which is because terminals can gradually receive better service with the convergence of the algorithm. We can observe that the EQWG algorithm has an prominent advantage over the traditional Q algorithm, which is because the switching effect and the concurrent switching are considered to guarantee the switching gain. The delay reduction of almost 1.5 ms is achieved when the simulation time is 30, 000 ms. It can also be found that lower delay and faster convergence are achieved by comparing EQWG and Q algorithms with EQG algorithm. Before 60, 000 ms, EQWG and Q algorithms do not converge, while EQG algorithm converges before 35, 000 ms. The reason is that the hysteresis mechanism obtained by game-modeling can guarantee and speed up the convergence rate. Moreover, the superiority of the RFEQG algorithm is more obvious than all other algorithms whether it is in the convergence rate or in the low-delay

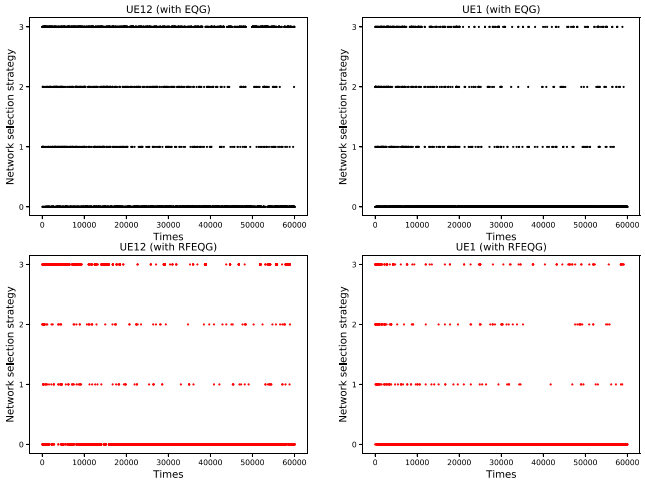


FIGURE 6. The effectiveness of feature-learning with RF algorithm.

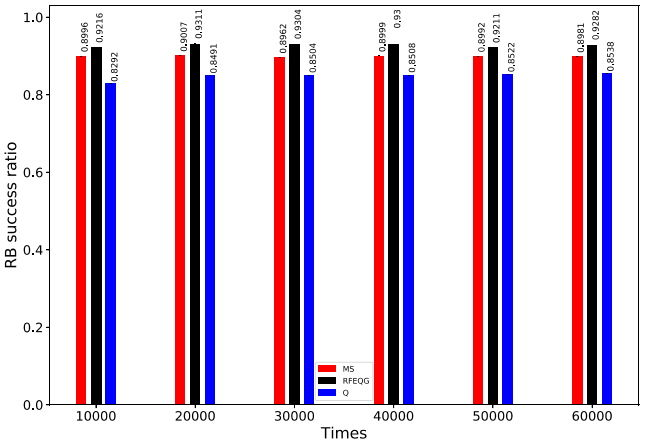


FIGURE 8. Comparison of the RB success ratio for different algorithms after convergence.

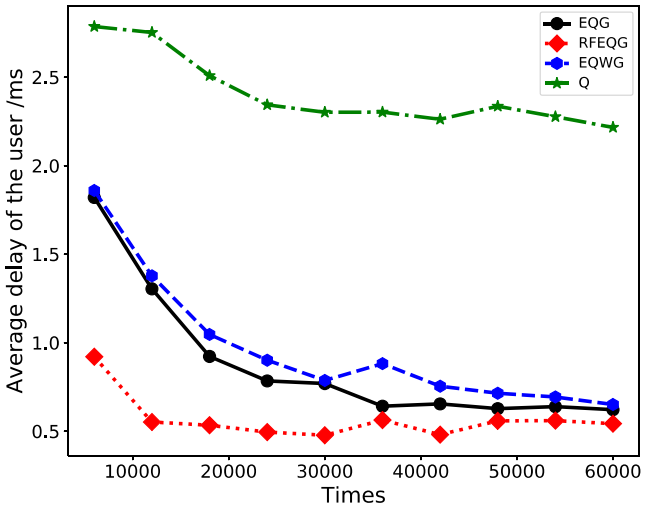


FIGURE 7. Comparison of the user average delay for different algorithms.

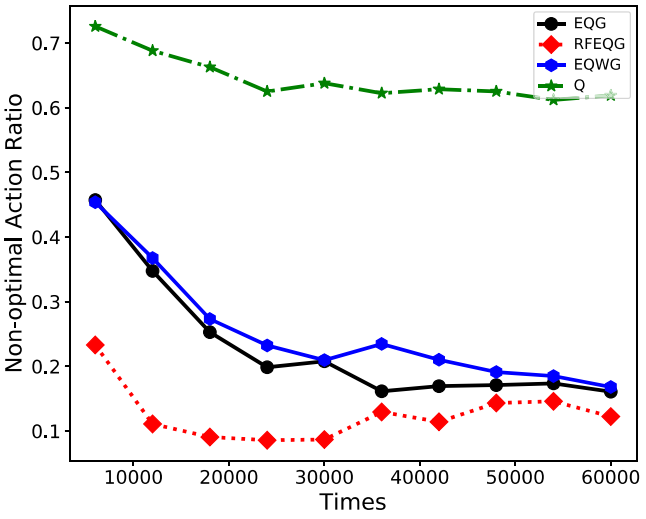


FIGURE 9. Comparison of the non-optimal selection rate for different algorithms.

performance. RFEQG algorithm reaches the convergence before 12, 000 ms. The minimum delay is about 0.6 ms for each terminal with RFEQG algorithm, which means that each terminal can obtain good service experience.

As shown in Fig. 8, the RB success ratio using different algorithms after convergence are compared. The X-axis is simulation times, and the Y-axis is RB success ratio. The simulation is 60, 000 ms, and we make a statistic at intervals of 10, 000 ms. The RFEQG algorithm and the Q algorithm are the same as Fig. 7. The max-RSRP (MS) algorithm represents the terminals connect to the network that can provide the maximum RSRP. We can see that the RB success ratio with the MS algorithm is always higher than the one with the Q algorithm, and the promotion of the RB success ratio is about 5% when the simulation times is 30, 000. While the RB success ratio with the RFEQG algorithm is always higher than the one with the MS algorithm, and the promotion of the RB success ratio is about 4% when the simulation times is 30, 000. It can be calculated that the RB success ratio with the RFEQG algorithm is 9% higher than the one with the Q algorithm when the simulation times is 30, 000. Therefore, our proposed RFEQG algorithm, which jointly considers feature-learning, strategy-learning, and game-modeling, has an obvious advantage than the traditional Q algorithm and the classical MS algorithm. By intelligently selecting the network by the terminal itself with the RFEQG algorithm, the network resource and the user experience can be improved.

The total non-optimal selection ratio for terminals is illustrated in Fig. 9. The non-optimal selection ratio means that the ratio of the terminals connecting to the SNs expect the optimal SN. The X-axis is simulation times, and the Y-axis is non-optimal selection ratio. The simulation is 60000 ms, and we make a statistic at intervals of 6000 ms. The involved algorithms are the same as Fig. 7. It is easy to understand that the total non-optimal selection ratio decreases with the growth of simulation times, which is because terminals can access to the optimal SN when the algorithm converges. We can also

observe that the EQWG algorithm has an prominent advantage over the traditional Q algorithm, which is because the switching effect and the concurrent switching are considered to guarantee the switching gain. When the simulation time is 30000 ms, the EQWG algorithm achieves the ratio reduction of almost 42% than the Q algorithm. Further, It can be found that lower ratio and faster convergence are achieved by comparing EQWG algorithm with EQG algorithm. The reason is that game-modeling can guide the terminals to make the best decision, implementing the fast convergence. Moreover, the superiority of the RFEQG algorithm is more obvious than all other algorithms whether it is in the convergence rate or in the performance. The supreme ratio reduction of almost 55% is reached with RFEQG algorithm. We can observe there exists fluctuation in RFEQG algorithm, which is due to the influence of time-varying channel and load.

VI. CONCLUSIONS AND FUTURE WORK

We have studied the intelligent user-centric network selection problem in the ultra-dense heterogeneous networks. We proposed a model-driven learning framework, which combines machine learning and game theory, to achieve fast and optimal network selection. Further, We implemented a fully distributed intelligent network selection algorithm at the user side based on the proposed framework. We introduced feature-learning to mine and learn the nonlinear and complex correlation between multi-RMIs and the link quality. By this way, we could reduce frequent switching and switching delay. Not only the link quality but also the load was considered to select the strategy, which could avoid accessing heavy-loaded SN. We also considered the switching effect and the concurrency, which could reduce unnecessary signaling overhead and guarantee switching effect. Game theory was used to avoid the infinite oscillation of the network selection in the mixed-model networks and guarantee the convergence. Simulation results confirmed the effectiveness of the proposed algorithm in reducing frequent switching, reducing average delay, enhancing user experience, and increasing resource utilization. Moreover, game theory was demonstrated to have a crucial impact on guaranteeing the convergence.

There also exist some other issues about network selection to be studied for improving our work. In UDNs, user mobility has a drastic impact on frequent hand-off and severe signaling overhead. In future work, we can take the regular mobility into consideration in the network selection. We can predict the moving path and regard the location as the input of strategy learning to reduce frequent hand-off.

REFERENCES

[1] Y. Benchaabene, N. Boujnah, and F. Zarai, "5G cellular: Survey on some challenging techniques," in *Proc. 18th Int. Conf. Parallel Distrib. Comput., Appl. Technol. (PDCAT)*, Taipei, Taiwan, Dec. 2017, pp. 348–353.

[2] S. H. Chae, J.-P. Hong, and W. Choi, "Optimal access in OFDMA multi-RAT cellular networks: Can a single RAT be better?" *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, pp. 4778–4789, Jul. 2016.

[3] D. Liu *et al.*, "User association in 5G networks: A survey and an outlook," *IEEE Trans. Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 1018–1044, 2nd Quart., 2016.

[4] J. Ding, R. Xu, Y. Li, P. Hui, and D. Jin, "Measurement-driven modeling for connection density and traffic distribution in large-scale urban mobile networks," *IEEE Trans. Mobile Comput.*, vol. 17, no. 5, pp. 1105–1118, May 2018.

[5] *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021 White Paper*. Accessed: Mar. 2017. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html

[6] J. Montoya, A. Sethi, and N. G. Gómez, "A load-based and fair radio access network selection strategy with traffic offloading in heterogeneous networks," in *Proc. 7th Int. Conf. Comput. Commun. Control (ICCCC)*, Oradea, Romania, May 2018, pp. 193–202.

[7] G. Araniti, P. Scopelliti, G.-M. Muntean, and A. Iera, "A hybrid unicast-multicast network selection for video deliveries in dense heterogeneous network environments," *IEEE Trans. Broadcast.*, to be published.

[8] S. Khan, M. I. Ahmad, and F. Hussain, "Exponential utility function based criteria for network selection in heterogeneous wireless networks," *Electron. Lett.*, vol. 54, no. 8, pp. 529–531, Apr. 2018.

[9] M. Lahby and A. Sekkaki, "A graph theory based network selection algorithm in heterogeneous wireless networks," in *Proc. Int. Conf. New Technol., Mobility Secur. (NTMS)*, Paris, France, Feb. 2018, pp. 1–4.

[10] N. Zarin and A. Agarwal, "A hybrid network selection scheme for heterogeneous wireless access network," in *Proc. IEEE 28th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Montreal, QC, Canada, Oct. 2017, pp. 1–6.

[11] E. Monsef, A. Keshavarz-Haddad, E. Aryafar, J. Saniie, and M. Chiang, "Convergence properties of general network selection games," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Hong Kong, Apr./May 2015, pp. 1445–1453.

[12] S. Singh and J. G. Andrews, "Joint resource partitioning and offloading in heterogeneous cellular networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 888–901, Feb. 2014.

[13] C. Liu, M. Li, S. V. Hanly, and P. Whiting, "Joint downlink user association and interference management in two-tier HetNets with dynamic resource partitioning," *IEEE Trans. Veh. Technol.*, vol. 66, no. 2, pp. 1365–1378, Feb. 2017.

[14] H. M. ElBadawy, "Optimal RAT selection algorithm through common radio resource management in heterogeneous wireless networks," in *Proc. 28th Nat. Radio Sci. Conf. (NRSC)*, Cairo, Egypt, Apr. 2011, pp. 1–9.

[15] H. Ding, H. Zhang, J. Tian, S. Xu, and D. Yuan, "Energy efficient user association and power control for dense heterogeneous networks," in *Proc. Int. Conf. Comput., Netw. Commun. (ICNC)*, Maui, HI, USA, Mar. 2018, pp. 741–746.

[16] V. Sagar, R. Chandramouli, and K. P. Subbalakshmi, "Software defined access for HetNets," *IEEE Commun. Mag.*, vol. 54, no. 1, pp. 84–89, Jan. 2016.

[17] E. Aryafar, A. Keshavarz-Haddad, M. Wang, and M. Chiang, "RAT selection games in HetNets," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Turin, Italy, Apr. 2013, pp. 998–1006.

[18] A. Keshavarz-Haddad, E. Aryafar, M. Wang, and M. Chiang, "HetNets selection by clients: Convergence, efficiency, and practicality," *IEEE/ACM Trans. Netw.*, vol. 25, no. 1, pp. 406–419, Feb. 2017.

[19] A. Awad, A. Mohamed, and C.-F. Chiasserini, "User-centric network selection in multi-RAT systems," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Doha, Qatar, Apr. 2016, pp. 97–102.

[20] H. AlNashwan and A. Agarwal, "User-centric network selection in wireless heterogeneous networks," in *Proc. IEEE 30th Can. Conf. Elect. Comput. Eng. (CCECE)*, Windsor, ON, Canada, Apr./May 2017, pp. 1–6.

[21] M. Zhang, X. Yang, T. Xu, and M.-T. Zhou, "Congestion-aware user-centric cooperative base station selection in ultra-dense networks," in *Proc. Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Nanjing, China, Oct. 2017, pp. 1–6.

[22] X. X. Wang Li and V. C. M. Leung, "Artificial intelligence-based techniques for emerging heterogeneous network: State of the arts, opportunities, and challenges," *IEEE Access*, vol. 3, pp. 1379–1391, 2015.

[23] R. Li *et al.*, "Intelligent 5G: When cellular networks meet artificial intelligence," *IEEE Wireless Commun.*, vol. 24, no. 5, pp. 175–183, Oct. 2017.

[24] J. Bendriss, I. G. Ben Yahia, P. Chemouil, and D. Zeghlache, "AI for SLA management in programmable networks," in *Proc. Design Reliable Commun. Netw. (DRCN)*, Munich, Germany, Mar. 2017, pp. 1–8.

[25] C. Sieber, A. Obermair, and W. Kellerer, "Online learning and adaptation of network hypervisor performance models," in *Proc. IFIP/IEEE Symp. Integr. Netw. Service Manage. (IM)*, Lisbon, Portugal, May 2017, pp. 1204–1212.

[26] J. Ahmed, A. Johnsson, F. Moradi, R. Pasquini, C. Flinta, and R. Stadler, "Online approach to performance fault localization for cloud and data-center services," in *Proc. IFIP/IEEE Symp. Integr. Netw. Service Manage. (IM)*, Lisbon, Portugal, May 2017, pp. 873–874.

[27] K. Ma, H. Fu, T. Liu, Z. Wang, and D. Tao, "Deep blur mapping: Exploiting high-level semantics by deep neural networks," *IEEE Trans. Image Process.*, vol. 27, no. 10, pp. 5155–5166, Oct. 2018.

[28] W. Hu, Y. Jin, Y. Wen, Z. Wang, and L. Sun, "Toward Wi-Fi AP-assisted content prefetching for an on-demand TV series: A learning-based approach," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 7, pp. 1665–1676, Jul. 2018.

[29] Y. Zhang, M. Pezeshki, P. Brakel, and S. Zhang, "Towards end-to-end speech recognition with deep convolutional neural networks," in *Proc. INTERSPEECH*, Sep. 2016, pp. 410–414.

[30] S. Moon, H. Kim, and Y. Yi, "BRUTE: Energy-efficient user association in cellular networks from population game perspective," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 663–675, Jan. 2016.

[31] H. Shao, H. Zhao, Y. Sun, J. Zhang, and Y. Xu, "QoE-aware downlink user-cell association in small cell networks: A transfer-matching game theoretic solution with peer effects," *IEEE Access*, vol. 4, pp. 10029–10041, Nov. 2016.

[32] J. Mar, M. B. Basnet, and G.-Y. Liu, "An intelligent transmit power control and receive antenna selection scheme for uplink MIMO-transceiver in high mobility environments," in *Proc. IEEE Int. Symp. Broadband Multimedia Syst. Broadcast. (BMSB)*, Nara, Japan, Jun. 2016, pp. 1–4.

[33] T. Lin, C. Wang, and P.-C. Lin, "A neural network based context-aware handoff algorithm for multimedia computing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 2, no. 5, Mar. 2005, pp. 1129–1132.

[34] M. E. Helou, M. Ibrahim, S. Lahoud, K. Khawam, D. Mezher, and B. Cousin, "A network-assisted approach for RAT selection in heterogeneous cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 6, pp. 1055–1067, Jun. 2015.

[35] E. Fakhfakh and S. Hamouda, "Optimised Q-learning for WiFi offloading in dense cellular networks," *IET Commun.*, vol. 11, no. 15, pp. 2380–2385, Nov. 2017.

[36] J. S. Perez, S. K. Jayaweera, and S. Lane, "Machine learning aided cognitive RAT selection for 5G heterogeneous networks," in *Proc. IEEE Int. Black Sea Conf. Commun. Netw. (BlackSeaCom)*, Istanbul, Turkey, Jun. 2017, pp. 1–5.

[37] Y. Yang, Y. Chen, C. Li, X. Gui, and L. Li, "Network traffic prediction based on LSSVM optimized by PSO," in *Proc. IEEE 11th Int. Conf. Ubiquitous Intell. Comput. IEEE 11th Int. Conf. Autonomic Trusted Comput. and IEEE 14th Int. Conf. Scalable Comput. Commun. Associated Workshops*, Bali, Indonesia, Dec. 2014, pp. 829–834.

[38] C. Yang *et al.*, "DISCO: Interference-aware distributed cooperation with incentive mechanism for 5G heterogeneous ultra-dense networks," *IEEE Commun. Mag.*, vol. 56, no. 7, pp. 198–204, Jul. 2018.

[39] L. Wang, C. Yang, X. Wang, F. R. Yu, and V. C. M. Leung, "User oriented resource management with virtualization: A hierarchical game approach," *IEEE Access*, vol. 6, pp. 37070–37083, 2018.

[40] P. Naghavi, S. H. Rastegar, V. Shah-Mansouri, and H. Kebriaei, "Learning RAT selection game in 5G heterogeneous networks," *IEEE Wireless Commun. Lett.*, vol. 5, no. 1, pp. 52–55, Feb. 2016.

[41] D. D. Nguyen, H. X. Nguyen, and L. B. White, "Reinforcement learning with network-assisted feedback for heterogeneous RAT selection," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 6062–6076, Sep. 2017.

[42] N. Morozs, T. Clarke, and D. Grace, "Distributed heuristically accelerated Q-learning for robust cognitive spectrum management in LTE cellular systems," *IEEE Trans. Mobile Comput.*, vol. 15, no. 4, pp. 817–825, Apr. 2016.

[43] X. Cao, Z. Song, B. Yang, and Z. Han, "Full-duplex MAC protocol for Wi-Fi/LTE-U coexistence networks," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Kansas City, MO, USA, May 2018, pp. 1–6.

[44] M. Heusse, F. Rousseau, G. Berger-Sabbatel, and A. Duda, "Performance anomaly of 802.11b," in *Proc. IEEE Int. Conf. Comput. Commun. (INFOCOM)*, San Francisco, CA, USA, Mar./Apr. 2003, pp. 836–843.

[45] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 3, pp. 535–547, Mar. 2000.

[46] E. Liu, Q. Zhang, and K. K. Leung, "Asymptotic analysis of proportionally fair scheduling in Rayleigh fading," *IEEE Trans. Wireless Commun.*, vol. 10, no. 6, pp. 1764–1775, Jun. 2011.

[47] P. Kyritsi, R. A. Valenzuela, and D. C. Cox, "Effect of the channel estimation on the accuracy of the capacity estimation," in *Proc. IEEE VTS 53rd Veh. Technol. Conf. (VTC)*, Rhodes, Greece, May 2001, pp. 293–297.

[48] M. A. Khan, S. Leng, W. Xiang, and K. Yang, "Architecture of heterogeneous wireless access networks: A short survey," in *Proc. IEEE Region 10 Conf.*, Macao, China, Nov. 2015, pp. 1–6.

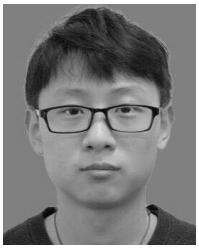
[49] E. Even-Dar, A. Kesselman, and Y. Mansour, "Convergence time to nash equilibrium in load balancing," *ACM Trans. Algorithms*, vol. 3, no. 3, p. 32, Aug. 2007.

[50] R. Gomes, M. Ahsan, and A. Denton, "Random forest classifier in SDN framework for user-based indoor localization," in *Proc. IEEE Int. Conf. Electro/Inf. Technol. (EIT)*, Rochester, MI, USA, May 2018, pp. 0537–0542.

[51] J. A. Benediktsson, J. R. Sveinsson, and P. H. Swain, "Hybrid consensus theoretic classification," in *Proc. Int. Geosci. Remote Sens. Symp. (IGRSS)*, Lincoln, NE, USA, May 1996, pp. 1848–1850.

[52] C. Yin, J. Xiang, H. Zhang, and J. Wang, "A new classificaiton method for short text based on SLAS and CART," in *Proc. Int. Conf. Comput. Intell. Theory, Syst. Appl. (CCITSA)*, Yilan, Taiwan, Dec. 2015, pp. 133–135.

[53] D. Pandey and P. Pandey, "Approximate Q-learning: An introduction," in *Proc. Int. Conf. Mach. Learn. Comput. (ICMLC)*, Bangalore, India, Feb. 2010, pp. 317–320.



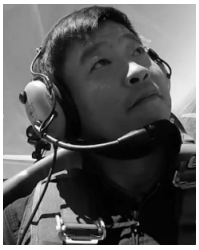
XINWEI WANG received the bachelor's degree in communication engineering from Xidian University, in 2012, where he is currently pursuing the master's degree. He is currently with the GUIDE Research Team, under the supervision of Dr. C. Yang. His research interest includes intelligent network association for cellular networks.

JIANDONG LI (SM'05) received the bachelor's, master's, and Ph.D. degrees in communications and electronic system from Xidian University, in 1982, 1985, and 1991, respectively, where he has been with Xidian University since 1985, was an Associate Professor from 1990 to 1994, has been a Professor since 1994, has been a Ph.D. Student Supervisor since 1995, has been the Dean of School of Telecommunication Engineering since 1997, and also serves as the Executive Vice Dean of the Graduate School.

Dr. Li is a Senior Member of the China Institute of Electronics and the Fellow of the China Institute of Communication. He was a member of PCN Specialist Group for China 863 Communication High Technology Program from 1993 to 1994 and from 1999 to 2000. He is also the member of Communication Specialist Group for The Ministry of Information Industry. His current research interest and projects are funded by the 863 High Tech Project, NSFC, National Science Fund for Distinguished Young Scholars, TRAPOYT, MOE, and MOI.

LINGXIA WANG received the bachelor's degree in communication engineering, in 2012. She is currently pursuing the master's degree with Xidian University. She is also with the GUIDE research team, under the supervision of Dr. C. Yang. Her research interest includes intelligent resource management for cellular networks.

CHUNGANG YANG (S'09–M'12) received the bachelor's and Ph.D. degree from Xidian University, Xi'an, China, in 2006 and 2011, respectively. From 2010 to 2011, he was a Visiting Scholar with the Department of Electrical and Computer Engineering, Michigan Technological University. From 2015 to 2016, he was a Visiting Scholar with the Department of Electrical and Computer Engineering, University of Houston. He is currently an Associate Professor with Xidian University, where he leads the research team of GUIDE, Game, Utility, Intelligent computing Design for Emerging communications. He has edited two books: *Game Theory Framework Applied to Wireless Communication Networks* (IGI Global, 2016) and *Interference Mitigation and Energy Management in 5G Heterogeneous Cellular Networks* (IGI Global, 2017). His research interests include resource and interference management, network optimization, and mechanism design for cognitive radio networks, heterogeneous cellular networks, and game theory for wireless communication and computing networks.



ZHU HAN (S'01–M'04–SM'09–F'14) received the B.S. degree in electronic engineering from Tsinghua University, in 1997, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Maryland at College Park, College Park, in 1999 and 2003, respectively.

From 2000 to 2002, he was an R&D Engineer of JDSU, Germantown, MD, USA. From 2003 to 2006, he was a Research Associate with the University of Maryland. From 2006 to 2008, he was an Assistant Professor with Boise State University, ID, USA. He is currently a Professor with the Electrical and Computer Engineering Department and the Computer Science Department, University of Houston, TX, USA. His research interests include wireless resource allocation and management, wireless communications and networking, game theory, big data analysis, security, and smart grid. He was a recipient of an NSF Career Award, in 2010, the Fred W. Ellersick Prize of the IEEE Communication Society, in 2011, the EURASIP Best Paper Award for the *Journal on Advances in Signal Processing*, in 2015, the IEEE Leonard G. Abraham Prize in communications systems (Best Paper Award in IEEE JSAC), in 2016, and several best paper awards in IEEE conferences. He is currently an IEEE Communications Society Distinguished Lecturer.