IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON ARTIFICIAL INTELLIGENCE FOR PHYSICAL-LAYER WIRELESS COMMUNICATIONS

Received August 11, 2019, accepted August 20, 2019, date of publication August 26, 2019, date of current version December 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2937438

Joint Power Control and Channel Allocation for Interference Mitigation Based on Reinforcement Learning

GUOFENG ZHAO[1], YONG LI[1], CHUAN XU[1], (Member, IEEE), ZHENZHEN HAN[1], YUAN XING[1], AND SHUI YU[2], (Senior Member, IEEE)

[1]School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
[2]School of Software, University of Technology Sydney, Ultimo, NSW, Australia

Corresponding author: Chuan Xu (xuchuan@cqupt.edu.cn)

This work was supported in part by the Chongqing Technology Innovation and Application Demonstration Project under Grant cstc2018jszx-cyzdX0120, in part by the Innovation Funds of Graduate Ph.D. under Grant BYJS201803, and in part by the Chongqing Graduate Research and Innovation Project under Grant CYB18167 and Grant CYB19176.

ABSTRACT In dense Wireless Local Area Networks (WLANs), high-density Access Points (APs) bring severe interference that seriously affects the experience of users, resulting in lower throughput and poor connection quality. Due to the heavy computation workload raised by the sizable networking systems and the difficulty in estimating instantaneous Channel State Information (CSI), existing works are hard to solve interference problem. In this paper, we propose a Joint Power control and Channel allocation based on Reinforcement Learning (JPCRL) algorithm combining with statistical CSI to reduce interference adaptively. Firstly, we analyze the correlation between transmit power and channel, and formulate the interference optimization as a Mixed Integer Nonlinear Programming (MINLP) problem. Secondly, we use the statistical CSI method to take the power and channel state as the state and action space, the overall throughput increment as the reward function of Q-learning, and obtain the optimal joint optimization strategy through off-line training. Moreover, for the periodic reinforcement learning process leading to resource consumption, we design an event-driven mechanism of Q-learning, which triggers online learning to refresh the optimal policy by event-driven condition and the consumption of computing resources can be reduced. The evaluation results show that the proposed algorithm can effectively improve the throughput compared with the existing scheme.

INDEX TERMS Interference, throughput, reinforcement learning, channel allocation, power control.

I. INTRODUCTION

Recent years, Wireless Local Area Networks (WLANs) [1]–[3] have been widely deployed for its simplicity of deployment and low cost. According to the Cisco reports [4], the amount of mobile traffic offload will increase from 54 percent (13.4 exabytes/month) in 2017 to 59 percent (111.4 exabytes/month) by 2022. Moreover, the number of total Wi-Fi hotspots worldwide will grow four-fold from 2017 to 2022. To meet data traffic demand, lots of APs are densely deployed in different network scenarios, e.g. stadiums, shopping malls and conference venues. However, the dendely deployed APs results in severe interference, which causes frequent fluctuations of wireless communication links and degradation of quality of service (QoS) [5], [6].

The inter-AP interference will increase the packet drop rate, which limits seriously throughput performance of WLANs. The throughput optimization can be achieved by AP's Channel Allocation (CA) [7], [8] and transmission Power Control (PC) [9], [10] through interference mitigating. The implementation of interference mitigation requires CSI, which usually causes huge overhead, latency and power costs. The channel allocation schemes have been widely utilized in WLANs [7], [8], [11]. To optimize the system throughput, Li *et al.* [7] propose an interference-tolerant medium access method by utilizing Partially Overlapped Channels (POCs). In order to avoid the interference produced by adjacent APs and users, the APs can turn to idle channel to reduce co-channel interference [8], [11]. Moreover, power control

The associate editor coordinating the review of this article and approving it for publication was Guan Gui.

schemes have been introduced to adjust the coverage area and transmit signal strength to improve throughput [12], [13]. Existing power control schemes are often approach coverage problems by increasing power, which leads to detrimental results (i.e.,interference, delay). Lowering transmission power [10] has benefits in terms of both interference and energy consumption, but causes data rate decreasing.

Moreover, joint power control and channel allocation to further reduce interference has been studied by some researchers, due to the tight coupling of channel allocation and power control. For example, in work [14], a centralized network solution has been proposed to optimize overall wireless networks. Similarly, in work [15], authors optimize the channel allocation and power allocation components based on the characteristics of the desired video content and channel conditions to achieve a high visual experience quality for multiple users. Researches [16], [17] find the optimal channel assignment with fixed power allocation, and then select the optimal power allocation to maximize network utility with the fixed channel.

However, the current joint resource allocation algorithms only optimize one-dimensional, without joint optimization of power and channel in a single iteration. In addition, these schemes depend on various types of network information, such as the locations of users, the instantaneous channel state and interference parameters in time slot, which are hard to obtained instantaneously in dense WLANs. In dynamic, large and dense networks, eliminating interference may incur massive computational complexity and communication overhead, due to the changes of channel conditions, along with the system state of WLANs evolves over time. Therefore, in dense WALNs, how to optimize power and channel simultaneously to maximize throughput is still a problem needs to be solved.

Aiming to reduce the interference to improve the throughput, we propose an intelligent and efficient JPCRL algorithm in dense WLANs. We firstly formulate the interference problem as a mixed integer non-linear programming (MINLP) problem considering the transmission power and channel in each iteration. Based on the high efficiency of Reinforcement learning in analyzing the complex data and temporal correlation property, we introduce a Q-learning algorithm from classical RL to solve the objective problem. Further, for the throughput performance changes, we design an event-driven mechanism of Q-learning to determine whether to refresh the optimal strategy at a new round of training. To the best of our knowledge, it is the first time that implement RL in the area of joint power and channel allocation for dense WLANs.

The main contributions of this paper can be summarized as follows:

- To optimize the interference of dense WLANs, we not only consider the transmission power control of APs to achieve received signal strength requirements, but also avoid using the same spectrum in wireless channel allocation, and formulate the interference optimization as a mixed integer non-linear programming MINLP) problem.
- To reduce the computational complexity of joint optimization scheme, we introduce reinforcement learning to optimize dynamic power and channel allocation in dense WLANs, and obtain the optimal joint optimization strategy through off-line training, which effectively avoids solving complex repeated derivative calculations compared to traditional optimization methods.
- In order to reduce the complexity of online learning brought by network dynamics, we design an event-driven mechanism of Q-learning. When the offline strategy adjustment results the Q-value to exceed a specific threshold, a new optimal network configuration strategy will be determined by triggering a new round training.
- Simulation results demonstrate that our proposed scheme can perform efficaciously under heavy traffic flow as well. It performs better than traditional approaches in context of throughput, average interference and response time.

The structure of this paper is as follows. A review of related works is introduced in Section II. Section III describes the system model and the problem formulation. Section IV presents the optimal algorithm. Simulation results are presented in Section V. Finally, the conclusion of this work is made in Section VI.

II. RELATED WORKS

In order to improve network quality and system throughput, various schemes have been proposed such as PC [9], [10], [18], [19], CA [7], [20]–[24] and joint schemes [14], [16], [25]–[29].

For instance, some researchers focus on PC. In [9], authors study the problem of best probability distribution associated with power levels. Then, the optimal probability distribution problem is formulated as a mixed-strategy game, where each node strategically selects a probability distribution of transmission power levels to maximize throughput. A coordination Wi-Fi management platform has been designed in [18], which coordinates APs to reduce interference through power control, and the Nash bargaining-based power control model is formulated and solved in a distributed manner. Kim *et al.* [19] describe co-channel interference caused by 802.11 MAC ACK frames, and proposes a dynamic transmission power control algorithm for ACK frames to reduce interference. It is shown that the dynamic power control algorithm outperforms any fixed or predefined schemes.

Different form PC approaches, some researchers mainly focus on the allocation of spectrum resources. In [21], authors propose an adaptive and distributed algorithm based on game-theoretic to select the channel width of APs. In [23], authors develop a joint optimization problem of channel selection and frame scheduling to maximize the summation throughput in LTE/WLAN. Since the high complexity of the formulated problem, a low-complexity heuristic algorithm has been proposed to select appropriate channels. Kala *et al.* [24] propose a channel allocation performance

prediction algorithm with a special emphasis on designing channel allocation schemes which alleviate the impact of interference on Wireless Mesh Network performance. However, techniques for optimizing the performance by considering joint channel allocation and power control are not investigated in these wireless networks.

Network optimization integrating CA and PC was studied in [26]–[29]. Ali *et al.* [26] consider the joint optimization of remote-radio-heads (RRH) association, sub-channel assignment, and power allocation for network sum-rate maximization in single-carrier frequency division multiple access based multi-tier cloud-radio access networks. The author in [27] are concerned with joint sub-channel and power allocation in a heterogeneous wireless network to optimize network performance. Kang *et al.* [28] characterize the final performance tradeoff between information decoding and energy harvesting, and an optimal power adaptation scheme for a nonlinear energy harvesting receiver operated only in the energy harvesting mode is proposed. Then, using this scheme derived the jointly optimal solution for the mode switching and power adaptation. However, the resource allocation algorithm is concerned only one-dimensional optimization in each step, and has the remarkably heavy computation workload.

With the rapid development of artificial intelligence (AI) [30]–[36], the use of artificial intelligence to optimize the network has become a trend with the advantages in processing large and complex data. Therefore, AI framework is a better choice and has been successfully used in wireless networks recently [37]–[40]. Xiao *et al.* [37] propose a RL-based power control scheme for downlink transmission to flexibly control their interference strategy. In [38], a RL solution is presented to adapt communization parameters of devices to the environment for maximizing energy efficiency and data transmissions. In [39], the author propose a handoff management scheme based on deep RL in WLANs, which can effectively improve the data rate. In [40], authors present a review of machine learning schemes in wireless sensor networks that used to increase resource utilization and prolong the lifespan of the network.

On the basis of the related works, it is noted that the most existing optimization methods do not address high-volume network data and high-quality communication between users and APs in dense WLANs. Therefore, we try to apply an RL-based joint power control and channel allocation approach to address aforementioned shortcomings.

III. SYSTEM MODEL

This section describes the correlation between the transmission power and channel. We formulate the interference and throughput problem from the power and channel parameters. Finally, we formulate the target problem through joint power control, channel allocation and other limitation factors.

A. NETWORK MODEL

In this subsection, we consider a centralized dense WLAN system as illustrated in Figure 1, which consists of a

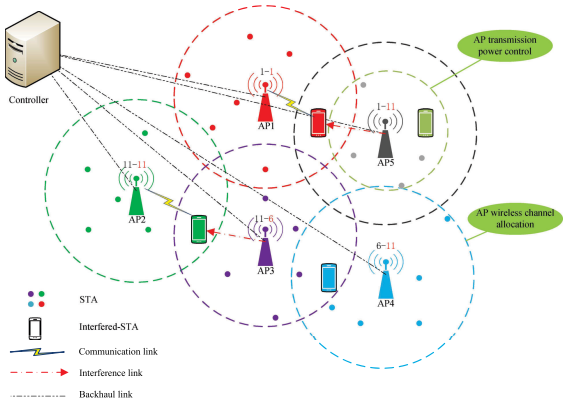

FIGURE 1. The simplified scenario for joint power control and channel control.

centralized controller, N APs and M users. Each user is associated with a surrounding AP to achieve high-speed and reliable Internet connectivity. The bands used by users refer to the 2.4 GHz band (2400−2483.5MHz). The transmission rate between a user and its associated AP depends on the distance between the user and the AP and the interference signals received from the environment. The set of APs is denoted as $\mathcal{N} = \{1, 2, 3, \ldots, N\}$, where N represents the total number of APs in the system. We also denote the set of users as $\mathcal{M} = \{1, 2, 3, \ldots, M\}$, where M represents the total number of users in the system.

In dense WLANs, we assume that one user is only connected to one AP, and any AP can over multiple users. We denote the minimum requirement of the active data rate for user j as l_j. Notice that for each AP, adjusting its power not only changes its transmission rate of own users, but also causes the interference variation to other APs. In this paper,the reasonable and effective optimization scheme can be achieved based on the information of APs collected by centralized controller. In the dual coordination optimization mechanism of interference, which can be used to finish the action of AP transmit power adjustment and AP channel switching to reduce interference and improve throughput. Moreover, the channel coefficients only include the path loss.

To given an example, let's consider Figure 1. Each AP has different coverage. As shown, using three non-overlapping channels of the 2.4 GHz band, the number above the AP icon indicates the allocated channel. There is interference between AP1 and AP5, AP2 and AP3, but one simple solution for allocating channels to APs would be to allocate channels 1,11,6,1 and 11(red numbers) to AP1, AP2, AP3, AP4 and AP5, respectively. In this case, the neighboring APs occupy different channels, so there is no interference between the APs. However, in real network, the system is more complicated, devising an interference management approach that considers all of these points is not straightforward, due to the following reasons:

- 1. Although increasing the transmission power of an AP can achieve acceptable communication quality, it is

TABLE 1. Symbol and definition.

Symbol	Definition
$\mathcal{M}$	Set of users; $\mathcal{M} = \{u_1, u_2, u_3, ..., u_M\}$
N	Set of deployed APs; $\mathcal{N} = \{AP_1, AP_2, AP_3, ..., AP_N\}$
$SINR$	Signal−to−noise−plus−interference ratio
$SINR_{min}$	SINR threshold value to avoid weak channels
p_i	The transmission power of AP_i
F	Set of the total available channels
f_i	Channel occupied by AP_i
d_{ij}	The Euclidean distance between u_j and AP_i
$\Delta(f_i, f_x)$	Interference effect of the channel allocated to AP_i on the channel allocated to AP_x
C_{ij}	The reachable downlink transmission rate of u_j to AP_i
g_{ij}	Channel gain between AP_i and u_j
γ	Path loss constant
B	Physical channel bandwidth
ϑ_{ij}	The association relationship between u_j and AP_i
I_j^{-i}	Sum interference signal from surrounding APs
I_j^x	The interference signal from the AP_x
e_{ix}	The adjacency relationship between AP_i and AP_x
L_{max}	Maximum AP load
U_a	The set of accosiated users
N_0	Gaussian noise

contradictory with reducing the transmission power will reduce the overlap between the interfering APs. In fact, increasing the transmission power results in an increase in the transmission range, which may subsequently lead to an increase in the overlap of the interfering APs.

- 2. To switch channels dynamically according to the status of network and avoiding channel overlapping as much as possible, which can greatly reduce the co-channel interference among APs. In addition, in order to avoid overlapping channel interference and guarantee the quality of communication, each AP can occupy one channel at a time. However, the number of interfering APs is reduced by applying orthogonal channels (eg, 1, 6, and 11), which results in less diversity of the available channels and subsequently increases of competing co-channel APs.

For ease of reference, the symbols and notations used in this paper are summarized in Table 1.

B. INTERFERENCE MODEL

In this paper, the transmit power of all APs can be reconfigured by the controller. In addition, AP's coverage is a circle centered on the AP, and its radius is positively correlated with transmit power. When a user is in the overlapping coverage range, the user will receive interference signals from other APs that have the same channel as the associated AP, and thus the performance of WLAN system would be severely degraded. We consider physical interference model [41], which computes all the links in wireless system that interferes user nodes. The physical interference model also overcomes the shortcomings of the protocol interference problem without considering the cumulative effect of interference, and can describe the interference in the real environment more accurately. Therefore, the interference accumulation effect between APs can be described by calculating the Signal-to-Interference-Plus-Noise Ratio (SINR) of the user link in current interference environment [42]. More intuitively, APs with large interference range have a low probability of utilizing same channels because a large number of users will be interfered by the AP. On the contrary, APs with small interference ranges are more likely to utilize same channels due to their limited interference to the WLANs.

In dense WLANs, a user receives not only the signal from the associated AP, but also the interference signal from other APs and the noise from the environment. The reachable downlink transmission rate of a link can be characterized according to the SINR of the current network status. In this paper, we assume that the channel coefficient only include the path loss. Therefore, when u_j is associated with AP_i, the SINR of the link on current network is expressed as

$$SINR_j^i = \frac{p_i g_{ij}}{N_0 + I_j^{-i}}, \tag{1}$$

We suppose that the path loss depends only on the Euclidean distance between AP_i and u_j. Therefore, the path loss is given by $g_{ij} = d_{ij}^{-\gamma}$. Let d_{ij} represents the Euclidean distance between AP_i and u_j. γ represents the path loss constent, and the value is usually settled as 2-5 [43].

According to the measurement results, we recently reported that [44] the interference in WLANs is jointly determined by the following two factors: (1) the channel separation, and (2) the received signal strength indicator (RSSI). The *SINR* can be rewritten as

$$SINR_j^i = D_{RSSI} + 10\lg d_{ij}^{-\gamma} - 10\lg\left(\sum_{x=1, x\neq i}^{N} \Delta(f_i, f_x) d_{ij}^{-\gamma}\right) - 10\lg\left(1 + \frac{N_0}{I_j^{-i}}\right), \tag{2}$$

where N_0 is the power of the additive white Gaussian noise from the environment, $p_i g_{ij}$ represents the signal received by u_j, p_i represents the transmit power of AP_i, and g_{ij} represents the free space path loss factor from AP_i to u_j, I_j^{-i} is the cumulative interference power it receives from other APs in its range, D_{RSSI} represents the difference of RSSI received by the user from its associated AP and interferences, which can be expressed as $D_{RSSI} = p_i - \sum_{x\in N, x\neq i} p_x$, $\Delta(f_i, f_x)$ represents the channel allocation. To ensure successful reception, only when *SINR* is greater than a predefined threshold value $SINR_{min}$.

In order to describe the relationship between u_j and AP_i, we define the association factor as

$$\vartheta_{ij} = \begin{cases} 1, & u_j \text{ is associated with } AP_i, \\ 0, & otherwise. \end{cases} \tag{3}$$

The distance from u_j to AP_i is less than the coverage radius of AP_i, they are considered to be associated possibly. When the u_j is associated with the AP_i, and the value of ϑ_{ij} is 1, otherwise 0.

When u_j is associated with AP_i without interference from other APs, the received signal is only affected by the noise from the environment and the distance attenuation between the user and the AP. Therefore, according to the Shannon-Hartley theorem [45], the reachable downlink transmission rate of u_j can be obtained from AP_i is $C_{max,ij} = \vartheta_{ij} B \log_2(1 + SINR)$. p_i is the transmission power of the AP_i, and B represents the physical channel bandwidth. In order to receive a correct signal and ensure the quality of service (QoS), the user has to attain a transmission rate that is greater than the minimum transmission rate (C_{min}) requirement.

We also define I_j^{-i} as the inter-APs interference signal that u_j receives from surrounding APs except for the currently associated AP_i. Furthermore, I_j^x indicates the same channel interference that received by u_j from the AP_x which is adjacent to AP_i.

$$I_j^x = e_{ix} \Delta (f_i, f_x) p_x g_{xj}, \tag{4}$$

When there is an overlapping coverage area between AP_i and AP_x, we express the Euclidean distance between AP_i and AP_x as $d(AP_i, AP_x) = \sqrt{\left(x_a^i - x_a^x\right)^2 + \left(y_a^i - y_a^x\right)^2}$, and x_a^x, y_a^x represents the abscissa and ordinate of the AP_x, respectively. The adjacency relationship between AP_i and AP_x can be given by

$$e_{ix} = \begin{cases} 1, & d(AP_i, AP_x) < R_i + R_x, \\ 0, & otherwise, \end{cases} \tag{5}$$

where R_i, R_x is the effective coverage radius of AP_i and AP_j, respectively. If AP_i and AP_j are considered to be adjacent, and the value of e_{ix} is 1, otherwise 0.

$\Delta (f_i, f_x)$ represents the channel relationship between AP_i and AP_x. If f_i is equal to f_x, which represents AP_x and AP_i occupy the same channel, the value of $\Delta (f_i, f_x)$ is 1, otherwise 0.

$$\Delta (f_i, f_x) = \begin{cases} 1, & f_i = f_x, \\ 0, & otherwise, \end{cases} \quad f_i, f_x \in F. \tag{6}$$

It can be seen that the total interference received by u_j from all surrounding APs in the high-dense WIFI system is expressed as

$$I_j^{-i} = \sum_{x=1, x\neq i}^{n} I_j^x = \sum_{x=1, x\neq i}^{n} \left\{ e_{ix} \Delta (f_i, f_x) p_x g_{xj} \right\}. \tag{7}$$

Therefore, when u_j is associated with AP_i and receives interference from other APs, the link capacity between u_j and AP_i can be expressed as

$$C_{ij} = \vartheta_{ij} B \log_2 \left(1 + \frac{p_i g_{ij}}{N_0 + \sum_{x=1}^{n} e_{ix} \Delta (f_i, f_x) p_x g_{xj}} \right), \tag{8}$$

where B represents the radio channel bandwidth. 1

At the same time, the system throughput at t time $C_{total,t}$ can be expressed as

$$\begin{aligned} C_{total,t} &= \sum_{i=1}^{M} \sum_{j=1}^{N} \vartheta_{ij} B \log_2 \left(1 + SINR_j^i \right) \\ &= \sum_{i=1}^{M} \sum_{j=1}^{N} \vartheta_{ij} B \log_2 \left(1 + \frac{p_i g_{ij}}{N_0 + \sum_{x\neq i} \Delta (f_i, f_x) e_{ix} p_x g_{ix}} \right). \end{aligned} \tag{9}$$

C. PROBLEM FORMULATION

With the interference model, the joint transmit power control and channel allocation problem for throughput optimization under QoS consideration is formulated as follows.

$$\begin{aligned} \max \ & \sum_{i=1}^{M} \sum_{j=1}^{N} \vartheta_{ij} B \log_2 \left(1 + \frac{p_i g_{ij}}{N_0 + \sum_{x\neq i} \Delta (f_i, f_x) e_{ix} p_x g_{ix}} \right) \\ s.t. \ & C1: \vartheta_{ij}, \ e_{ix}, \ \Delta (f_i, f_x) = \{0, 1\}, \quad \forall i, x \in M, \ \forall j \in N, \\ & C2: SINR_{ij} \geq SINR_{\min}, \\ & C3: 0 \leq \sum_{j\in U_a} l_j \leq L_{\max}, \quad \forall j \in N, \\ & C4: \sum_{i=1}^{M} \vartheta_{ij} \leq 1, \quad \forall j \in N, \ \forall i \in M, \\ & C5: c_{ij} = \begin{cases} 1, & d_{ij} < R_i \\ 0, & otherwise, \end{cases} \\ & C6: \vartheta_{ij} \leq c_{ij}, \quad \forall j \in N, \ \forall i \in M. \end{aligned} \tag{10}$$

The objective (8) is introduced to measure the sum of the throughput of WIFI system. Solving the problem means that the corresponding algorithm should find the optimal power control vector P^* and channel switch vector f^*. The feasible domain of p_i, f_i, e_{ik}. $\Delta (f_i, f_x)$ and ϑ_{ij} are channel assignment and user assignment variables to be determined at t th time slot, respectively. The contraint e_{ik} ensures that only the user within the coverage of the AP can be associated with it. The constraint $C2$ shows the $SINR$ condition that each communication link to ensure successful communication, where $SINR_{min}$ is the minimum requirement. The constraint $C3$ shows that the total load of each AP is within its transmission capacity L_{max}, which has been measured in [46]. The constraint $C4 - C6$ shows that each user can only be associated with one AP, which covers the user.

The joint power control and channel allocation for maximum throughput is a Mixed Integer Nonlinear Programming problem that has been proven to be an nondeterministic polynomial time hard (NP-hard) problem [47], which cannot be solved directly by traditional optimization method. In order to achieve the goal of maximizing the throughput of the system, we present a joint power control and channel allocation algorithm based on RL in the next section.

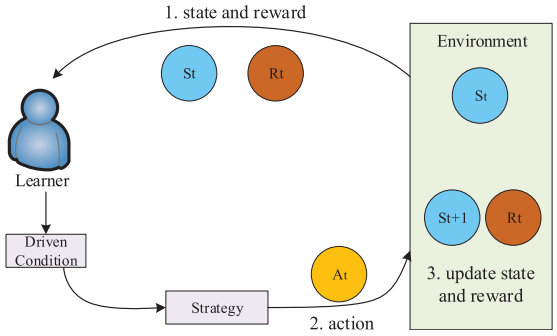

FIGURE 2. Intelligent Q-learning-based resource allocation: system environment and its elements.

IV. THE JOINT OPTIMAL ALGORITHM BASED ON REINFORCEMENT LEARNING

This section aims to discuss the most popular model-free RL algorithm used in this work, Q-learning, where an agent learns to take actions that would yield the most cumulative reward by interacting with the stochastic wireless network environment. Firstly, we establish Q-learning model based on state parameters in dense WLANs. Then, according to the Q-learning model, the main goal of the Q-learning algorithm is to learn an optimal strategy that exploits the total anticipated reward, which is given by Bellman's equation. Finally, the JPCRL algorithm depicts the specific steps performed to optimize the objective function.

A. REINFORCEMENT LEARNING

Reinforcement learning, such as Q-learning and actor-critic, applies a reward mechanism to reflect the interaction with the environment [48]. As presented in Figure 2, the learner-environment inter-action process is summarized. In such a case, the Q-learning model is likewise comprised of a learner, of a set of system states, S, and a set of actions, A, for every state. By performing an action in a particular state, the learner collects a reward r with the objective of maximizing its accumulated reward, i.e., maximizing throughput through joint power control and channel allocation in this paper.

In order to obtain the optimal policy, we must identify the action, state and reward functions in our Q-learning model, which will be described in the next following subsections.

System states (S): System state is an abstraction of the WLANs environment, and the learner makes action decisions based on the states of the WALNs. The key to affect the state of the network environment is the channel and transmit power of APs. The QoS of users is restricted by network environment. In addition, the current user information is also fed back to the network controller, so the network controller can further adjust according to user information to improve various network performance. Therefore, the system state $\boldsymbol{S}$ is defined as a countable non-empty set as

$$\boldsymbol{S} = S(\boldsymbol{u}, \boldsymbol{p}, \boldsymbol{f}) = \{S_0, S_1, \ldots, S_t, \ldots S_T\}, \tag{11}$$

where $\boldsymbol{u} = \{u_1, u_2, \ldots, u_j, \ldots, u_m\}(j = 1, 2, \ldots, m)$ represents the set of user imformation, and $u_j = \{(x_u^j, y_u^j), l_j\}$, where (x_u^j, y_u^j) represents the location information and l_j the transmission rate requirements of u_j, respectively. $\boldsymbol{p} = \{p_1, p_2, \ldots, p_i, \ldots, p_n\}$ and $\boldsymbol{f} = \{f_1, f_2, \ldots, f_i, \ldots, f_n\}$ represent the set of transmit power and the set of channel of the current AP, respectively. S_t is the system state at time t, T is the termination time, n is the number of APs, m is the number of users.

Action space (A): The learner takes an action by observing the state of the network, causing the network to operate in a new state to change the current state of the network. The action in the context means the transmission power control and the channel switch of the APs based on the state of network. Thus, the set of all actions is expressed as

$$\boldsymbol{A} = A(\boldsymbol{p}, \boldsymbol{f}) = \{A_1, A_2, \ldots, A_t, \ldots, A_T\}, \tag{12}$$

where, A_t represents the action taken by the learner at time slot t, $\boldsymbol{p} = \{p_1, p_2, \ldots, p_t, \ldots, p_n\}$ and $\boldsymbol{f} = \{f_1, f_2, \ldots, f_t, \ldots, f_n\}$ represent the transmit power and the channel of the action at time slot t, respectively.

Reward function (r): In this WLAN system, the learner tries to maximize the accumulated rewards by taken a set of actions, which directly affects the performance improvement of the system. In the optimization problem C_t, the goal is to maximize the system throughput. Thus, we define the immediate reward as the amount of change of the current system throughput and previous. The immediate reward is positive when the throughput increases, otherwise, negative. The benefit of the action is defined as the immediate reward r_t, which is a reward associated with the $(t-1)-t$th state transition. The immediate reward is denoted as

$$r_t = C_{total,t+1}(S, A) - C_{total,t}(S, A), \tag{13}$$

where the value of r_t is initialized to 0, the system throughput at t time is $C_{total,t}$.

When an action is taken, the learner will receive rewards or penalities and the status of the WLAN system will change. Thus, when the action taken by the learner increases the value of objective function, the learner receives a positive reward, conversely, a penalty i.e., a negative reward, which will reduce the cumulative rewards. The process of interaction between learner and the environment can be described as a strategy track

$$\tau = S_0, A_1, S_1, r_1, A_2, \ldots, A_t, S_t, r_t, A_{t+1}, S_{t+1}, \ldots. \tag{14}$$

B. SYSTEM UTILITY FUNCTION

Whereas the reward indicates what is good in an immediate sense, a utility function specifies long-term benefits. Roughly speaking, the utility of a state is the total amount of reward the learner can expect to accumulate over the future, starting from that state. Due to the number of users and the transmission rate requirements of users change over time, it is difficult to certain the system throughput

$C_t = \sum_{i=1}^{M}\sum_{j=1}^{N} \vartheta_{ij} B \log_2\left(1 + SINR_j^i\right)$ and the reward r. Consequently, it is reasonable to select the network that provides the best average utility. Since the user has no prior knowledge of the average performance of the available network, the learner must learn the optimal strategy from interaction with the environment. Mathematically, this learning problem can be formed to select an optimal strategy π^* that maximizes the accumulated average reward. According to the optimal strategy, the system takeing a series of action $\{A_1, A_2, A_3, \ldots, A_t\}$, the expected total return is maximized. The accumulated rewards R from the time t state is expressed as,

$$R_t\left(S, A, A'\right) = \sum_{k=0}^{T} r_{t+k}. \tag{15}$$

There are two value function to represent the feedbacks from each decision in the RL problem, namely state value function $V^{\pi}(S, A)$ and action-state value function $Q^{\pi}(S, A)$.

The expected state value is expressed as

$$V^{\pi} = E^{\pi}\left\{\sum_{k=0}^{T} \beta^k R_{t+k+1} | S_t = S\right\}, \tag{16}$$

where $E\{*\}$ denotes the mathematical expectation. Therefore, the maxsize sate value is

$$V^* = V^{\pi^*} = \max_{a \in A}[R_{t+1} + \beta V^*], \tag{17}$$

where, $\beta \in [0, 1]$ denotes the reward discount factor that reflects the importance of immediate reward and accumulated reward. When $\beta = 0$, the learner ignores future rewards, $\beta = 1$ represents the future rewards as important as the rewards in the current state.

According to the previous formula, based on the state S at the time t, the expectation of the future return can be obtained after the action being selected, which means the state-action value function

$$\begin{aligned} Q^{\pi}(S, A) &= E^{\pi}[R_t | S_t = S, A_t = A] \\ &= E^{\pi}[R_{t+1} + \beta Q(S', A') | S, A], \end{aligned} \tag{18}$$

State-value function iterative update formula

$$\begin{aligned} Q_{t+1}(S, A) = Q_t(S, A) + \alpha[R^A_{S \to S'} \\ + \beta \max Q_t(S', A') - Q_t(S, A)], \end{aligned} \tag{19}$$

According to increment $\alpha[R^A_{S \to S'} + \beta \max Q_t(S', A') - Q_t(S, A)]$ update state-action value function, α is the learning rate.

A decision-making strategy is a collection of specific actions taken when a state is given, i.e., $\pi = \pi(A|S)$ for all state-action pairs. The optimal strategy π^* is to maximize the accumulated reward of all states. Hence, The optimal state-action value of state S is defined as

$$A^* = \underset{A}{arg\max}\, Q^*(S, A) = \max_{A} Q(S, A). \tag{20}$$

To maximize the system long-term utility, the learner uses a state-action to guide its decision making. The strategy of accumulating the maximum reward is the optimal strategy. Therefore, we can obtain the optimal scheme to allocate power and channel according to the maximum state-action value function.

C. EVENT-DRIVEN CONDITION

In dense WLANs, a stochastic environment, significant changes in the operation of a system are the result of random event occurrences, so that, perceiving such events and reacting to them is crucial. In the interaction between the learner and the environment, in each learning step, the learner first observes the environment to collect channel information, and then formulates a strategy and learns. The whole learning process is periodic. When the learning environment is relatively stable, periodic collection of information and strategic search will inevitably consume unnecessary resources. In order to reduce the consumption of computing resources caused by the strategy periodically search for a large number of users, we introduce the event-driven mechanism into the joint power control and channel allocation algorithm based on Q-learning.

We define δ as the threshold of change tolerance. When degree of network disturbance reaches a certain threshold, the controller performs a new round of training based on the current network satus data. The degree of data change of network status is the same as the change between AP's current sense data and previous sence data, and it is a relative value. We represent event-driven conditions based on the degree of change in state value Q. The event-driven condition is designed as

$$\frac{Q_t\left(S_t, A_t\right) - Q_{t-1}\left(S_{t-1}, A_{t-1}\right)}{Q_t\left(S_t, A_t\right) - \left[\left(\sum_{k=1}^{T} r_{t+k}\left(S_{t+k}, A_{t+k}\right)/T\right) - r\left(S_t, A\right)\right]} > \delta. \tag{21}$$

where, δ is the threshold of event-driven condition. If the topology of APs had to change, when the condition value is greater than the threshold, the learner updates the strategy and action through a new round training, otherwise, performs the last action. If the topology of APs has to change, these strategies would become invalid, which would lead to a new run of the RL algorithm. The event-driven mechanism solves the performance degradation caused by network disturbances. In addition, learners don't need to perform trial and error and iteration in each learning step, which reduces the amount of computation and network resource waste caused by the periodic learning process.

D. THE JPCRL ALGORITHM

The details of JPCRL algorithm based on Q-learning method are given in Algorithm 1 and 2. The JPCRL algorithm includes two phases, one is training phase and the other is inference phase. In Algorithm 1 (training phase), the parameters related to WLAN system and Q-learning are

Algorithm 1 The JPCRL Algorithm - Training Phase

Input:
Q-table, $\alpha \in [0, 1]$, $\beta \in [0, 1]$, T, L_{max}, $\boldsymbol{S} = S(\boldsymbol{u}, \boldsymbol{p}, \boldsymbol{f})$, $\boldsymbol{A} = A(\boldsymbol{p}, \boldsymbol{f})$,$\alpha$, β.

Output:
The optimal strategy π^*, f^* and P^*.

1: **for** time-slot $t = 1, 2, \ldots, T$ **do**
2: Select a initial state S_0 randomly;
3: **while** $S_t != S_{goal}$ **do**
4: Select an action A_t based on greedy strategy, and obtain immediate reward r_t and next state S_{t+1};
5: Update the Q table according to $Q_t(S, A) \leftarrow Q_t(S, A)+\alpha[r^A_{S\rightarrow S'} + \beta \max Q_t(S', A') - Q_t(S, A)]$;
6: Select a α randomly to explore or utilize with greedy probability ε;
7: **if** explore **then**
8: Find the optimal action based on the Equation $A^* = \underset{A}{arg\max}\, Q^*(S, A)$;
9: Adjust AP according to the optimal action;
10: **else**
11: Adjust the AP based on the actions that have been obtained;
12: **end if**
13: **end while**
14: **end for**

Algorithm 2 The JPCRL Algorithm - Inference Phase

Input:
Q-table, $\alpha \in [0, 1]$, $\beta \in [0, 1]$, T, L_{max}, δ, State information S_t.

Output:
Optimal power strategy and channel handoff decision P^* and f^*.

1: Read the model saved in the training phase;
2: Read the state S_0 and preprocess it;
3: **for** time-slot $t = 1$ to T **do**
4: Input the state S_t to the evaluation network and output the Q table of all actions;
5: **if** $\frac{Q_t(S_t,A_t)-Q_{t-1}(S_{t-1},A_{t-1})}{Q_t(S_t,A_t)-\left[\left(\sum_{j=1}^{T} r_{t+j}(S_{t+j},A_{t+j})/T\right)-r(S_t,A)\right]} > \delta$ **then**
6: Repeat **Training Phase**;
7: **end if**
8: Select $A^* = \underset{A}{arg\max}\, Q^*(S, A)$;
9: Obtain reward $r^{A^*}_{S_t\rightarrow S_{t+1}}$ and the next state S_{t+1}.
10: **end for**

initialized first, as shown in step input. The training iteration period T is defined, working as the condition out of the training process and obtain the maximum Q value. The learner reads the state information S_0 and selects an action randomly to obtain immediate reward and update the Q values, as shown from step 1 to 4. The process of exploitation and exploration is given in steps 5 to 11, which guarantees that the final policy is a global optimum and not a local one. The optimal allocation policy can be obtained through massive training iterations.

In the Algorithm 2 (inference phase), the learner reads the initial state S_0 of the network, as shown in step 1. When the new state is fed into the optimal policy, the corresponding predicted output can be obtained immediately, because the computations in RL only contain several simple operations. Then, the learner selects actions based on the predicted out put. When the network changes greatly with abnormal behavior, we introduce event-driven strategy to determine the condition value for repeat training as shown step 5 to 10.

V. EXPERIMENT AND EVALUATION

In this section, we conduct simulations to evaluate the performance of WLAN under different settings. The proposed JPCRL algorithm is compared to three other approaches, CA scheme without PC [24], PC scheme without CA [18], and traditional JPC scheme [15]. The PC scheme without CA means that only the power of all APs is adjusted. The CA scheme without PC means that only channels are allocated to all APs. The traditional JPC scheme refers to optimizing joint power and channel in single iteration, only one-dimensional optimization is concerned in each step.

TABLE 2. Experiment parameters.

Parameter	Value	Commons
γ	2	Attenuation coefficient
P	[1mW/0dBm, 1W/30dBm]	Limit of AP transmit power
$L_{\max}$	70Mbps	Limit of AP load
f_c	2.4GHz	Carrier spectrum
N_0	$10^{-13}W/-100dBm$	Thermal noise power
B	20MHz	Bandwidth of channel
ε	0.4	greey rate
α	0.005	Learning rate
β	0.98	Reward discount
T	4000	The training iteration period

A. SIMULATION SETTING

We simulate a dense WLAN scenario where 15 APs are evenly deployed and different users densities are deployed randomly in an area of 100m*100m. In the simulations, we assume that the default value of AP's transmit power is 30dBm and can be adjusted, according to the coverage requirements of APs and the throughput demands of users. For the test data of power and channel information and related history data is measured according to our previous research [44]. The two learning stages are simulated in MATLAB simulation platforms.

Specifically, experiment parameters are shown in Table 2.

B. PERFORMANCE COMPARISONS

In this subsection, we present some evaluation results and provide a brief discussion. In the evaluation, we illustrate the impact caused by the value of learning rate α on learning efficiency.

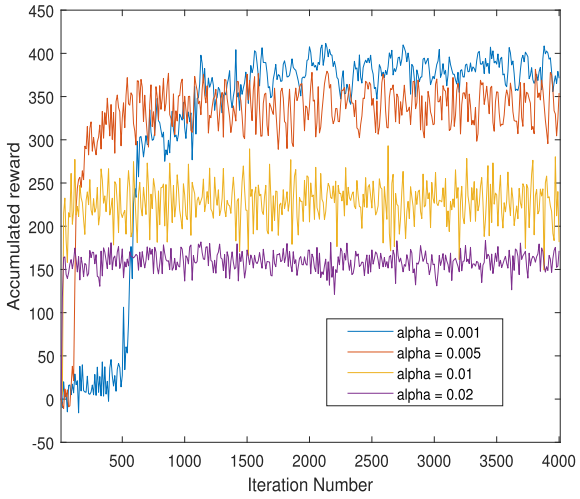

FIGURE 3. The accumulated reward with the diferent learning rate α.

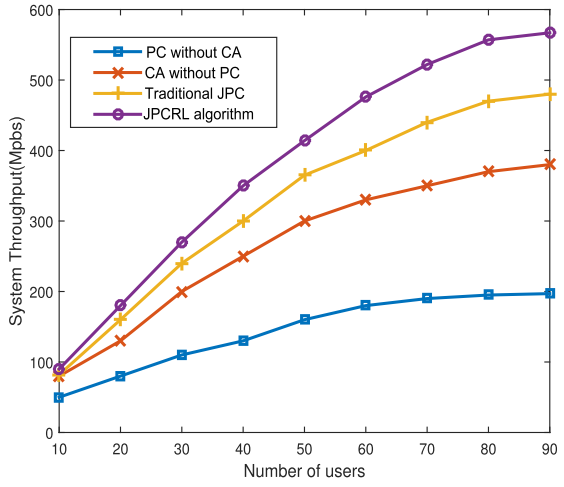

FIGURE 4. System throughput compared with different number of users.

Figure 3 shows the convergence situations of accumulated reward under different learning rate α in the iteration process of our algorithm. The accumulated reward is obtained by calculating the amount of change that the system throughput in every iteration. The figure shows that as the iteration increases, the accumulated reward gradually converges to the optimal value. When the learning rate is very small ($\alpha = 0.001$), the RL learner has to take nearly 1000 time steps to converge. When increasing the learning rate to 0.005, the optimal policy can be learned, that is about 400 time steps until convergence. From Figure 3, we can observe that the learning rate has some effects on the accumulated reward of the proposed scheme. Choosing a learning rate that is too small will result in a slower convergence rate, but the accumulated reward will be higher. Conversely, if the learning rate is too large, the convergence rate will be faster, but the accumulated reward is lower. Therefore, learning rate should be selected properly, neither too large nor too small.

As is shown in Figure 4, the system throughput of the four scenarios increases as the number of users increases. As the

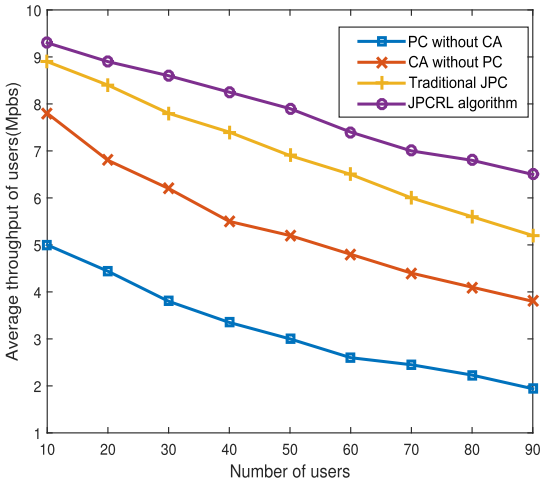

FIGURE 5. Average throughput with different number of users.

number of users increases, the throughput increases with an increase rate, and after the number reaches 60, the throughput growth rate decreases. The proposed algorithm in this paper improves the throughput by 1.7 times compared to the PC scheme without CA. The throughput is increased by 50% relative to the CA scheme without PC. Compared with the traditional JPC algorithm, the JPCRL algorithm is also improved by 16%. This is because as the number of users increases, the competition between users and users increases, and the co-channel interference between the AP and another AP increases, and the relationship between the users and the AP is more complicated. The JPCRL algorithm obtains an optimization strategy through continuous training. In the experiment, the algorithm maximizes throughput to meet user throughput requirements by allocating channels and adjusting power at each iteration.

In Figure 5, we compare the average throughput of users under different number of users. The results show that as the number of users increases, the average throughput of users decreases. We can observe that the average throughput of the JPCRL algorithm is better than the other three schemes. There are two reasons for the decrease of the average throughput of users: one of them is that as the number of users increases, the density of users in the scenario increases, and the competition between users increases. The other is that the APs have a large transmission power, resulting in an increase in the same frequency interference between the APs. The event-driven reinforcement learning algorithm can more effectively update the optimal strategy, resulting in lower throughput performance.

Figure 6 shows that the aggregarte throughput of APs under different number of users. We increas the number of users from 10 to 90, the average throughput of APs will decrease. From Figure 6, with the number of users increases, the aggregate throughput of APs is gradually decreased for all schemes. This is because increasing the user density also increases the number of clients present in the overlapping coverage of multiple APs, thereby increasing interference

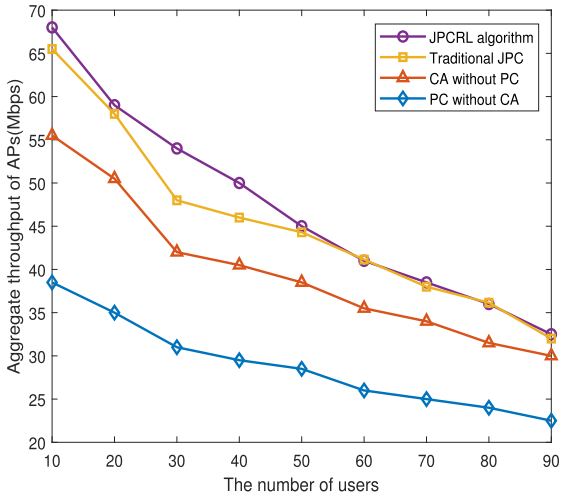

FIGURE 6. Aggregate throughput of APs compared with three schemes.

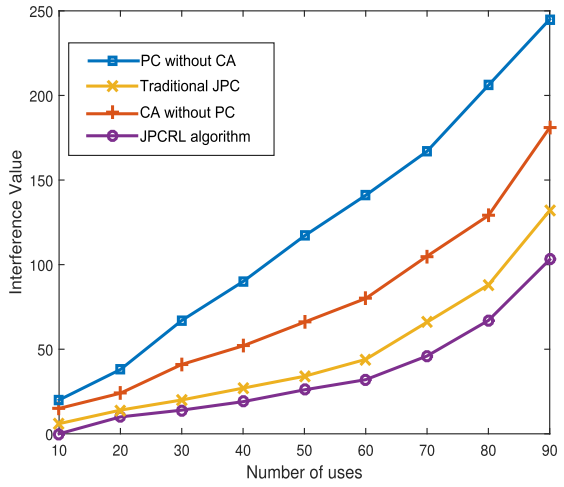

FIGURE 7. Interference compared with three schemes.

between multiple APs, which results in a decrease in total throughput. When the number of users is smaller than 50, the JPCRL algorithm outperforms the other three. When the number of users is greater than 50, the JPCRL algorithm has almost the same performance as the traditional JPC scheme. This is because the adjustment of the algorithm is weakened as the number of users increases. In addition, the AP single-scheme adjustment effect is limited, and the joint optimization algorithm can significantly improve performance.

Figure 7 shows that as the number of users increases, co-channel interference between APs increases. It can be seen from the results that the JPCRL algorithm is obviously superior to the other three algorithms. As can be seen from this figure, the total amount of interference in the system is increasing with the number of users increases. The reason is that more users increases the more links, all links aim to share the fixed radio resource, consequently leading to the growing amount of generated interference. The interference of our proposed algorithm is reduced by 80% compared with the PC scheme without CA, which 30% lower than that

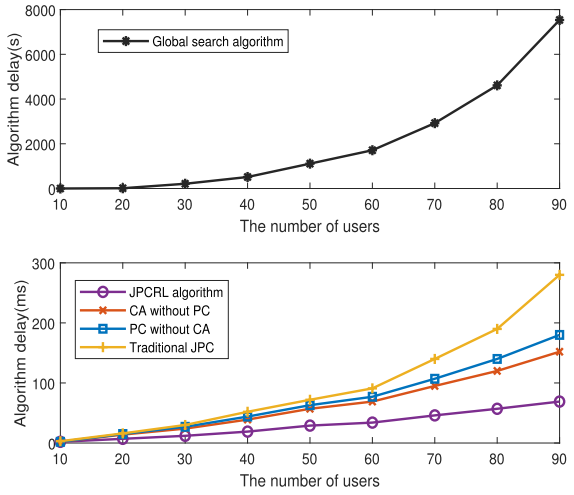

FIGURE 8. Algorithm delay compared with three schemes.

of the traditional JPC scheme, and 65% lower than that of the CA scheme without PC. Therefore, the JPCRL algorithm simultaneously adjusts the channel and power algorithms at each step, which can further reduce interference.

Figure 8 shows the comparison of execution time of different algorithms. Although the global search algorithm can get the optimal resource allocation scheme, it can be seen that the algorithm consumes a lot of time, and the time consumed shows an exponential rising trend. The global search algorithm is very inefficient and cannot be applied to practical problem solving. In addition, the speed of all algorithm solutions decreases as the number of network nodes increases. However, in the reasoning phase, the time delay value of the algorithm is reduced to a tolerable range and is superior to the other three algorithms. Please note that we are more concerned with the performance in the real application, the inference phase. It is worthwhile to sacrifice a small amount of complexity in exchange for a significant increase in throughput. Therefore, the JPCRL algorithm compromises complexity and gain and is practically feasible.

VI. CONCLUSION

This work has sudied the resource assignation in dense WLANs and improves the throughput. We proposed a more practical and suitable algorithm for WLANs, which joints power control and channel allocation based on RL to improve throughput. In the JPCRL algorithm, the channel parameters and power levels are obtained through actual measurements and an optimal resource allocation strategy that maximizes long-term system benefits is calculated. In the absence of any disturbances or minor disturbances applied to the learning system, the system can operate under optimal conditions according to the optimal strategy. When there is a large disturbance in the system, we introduce an event-driven strategy to trigger the learning process and re-acquire the optimal strategy. It is shown that our proposed JPCRL algorithm achieves significant improvements in terms of reducing the overall interference in the network and increasing the throughput.

The scheme could provide helpful guidance for dense APs deployment and network-intensive applications in future.

REFERENCES

[1] B. Dappuri and T. G. Venkatesh, "Design and performance analysis of multichannel MAC protocol for cognitive WLAN," *IEEE Trans. Veh. Technol.*, vol. 67, no. 6, pp. 5317–5330, Jun. 2018.

[2] C. Xu, J. Wang, Z. Zhu, and D. Niyato, "Energy-efficient WLANS with resource and re-association scheduling optimization," *IEEE Trans. Netw. Service Manage.*, vol. 16, no. 2, pp. 563–577, Apr. 2019. doi: 10.1109/TNSM.2019.2910203.

[3] M. Derakhshani, X. Wang, D. Tweed, T. Le-Ngoc, and A. Leon-Garcia, "AP-STA association control for throughput maximization in virtualized WiFi networks," *IEEE Access*, vol. 6, pp. 45034–45050, 2018.

[4] Cisco. (Feb. 2019). *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update*, 2017–2022. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-738429.html.

[5] F. Gringoli, R. Klose, M. Hollick, and N. Ali, "Making Wi-Fi fit for the tactile Internet: Low-latency Wi-Fi flooding using concurrent transmissions," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC)*, May 2018, pp. 1–6.

[6] J. Chen, B. Liu, H. Zhou, Q. Yu, G. Lin, and X. Shen, "QoS-driven efficient client association in high-density software-defined WLAN," *IEEE Trans. Veh. Technol.*, vol. 66, no. 8, pp. 7372–7383, Aug. 2017.

[7] J. Li, T. Y. Cheng, X. Jia, and L. M. Ni, "Throughput optimization in WLAN/Cellular integrated network using partially overlapped channels," *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 157–169, Jan. 2018.

[8] Y. Zhang, C. Jiang, J. Wang, Z. Han, J. Yuan, and J. Cao, "Green Wi-Fi management: Implementation on partially overlapped channels," *IEEE Trans. Green Commun. Netw.*, vol. 2, no. 2, pp. 346–359, Jun. 2018.

[9] M. Zou, S. Chan, H. L. Vu, and L. Ping, "Throughput improvement of 802.11 networks via randomization of transmission power levels," *IEEE Trans. Veh. Technol.*, vol. 65, no. 4, pp. 2703–2714, Apr. 2016.

[10] K.-S. Shin and O. Jo, "Joint scheduling and power allocation using non-orthogonal multiple access in directional beam-based WLAN systems," *IEEE Wireless Commun. Lett.*, vol. 6, no. 4, pp. 482–485, Aug. 2017.

[11] Y. Su, Y. Wang, Y. Zhang, Y. Liu, and J. Yuan, "Partially overlapped channel interference measurement implementation and analysis," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Apr. 2016, pp. 760–765.

[12] Q. Fan, H. Lu, P. Hong, and Z. Zhu, "Throughput–power tradeoff association for user equipment in WLAN/cellular integrated networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3462–3474, Apr. 2017.

[13] J. Yao, W. Lou, C. Yang, and K. Wu, "Efficient interference-aware power control for wireless networks," *Comput. Netw.*, vol. 136, pp. 68–79, May 2018.

[14] H. Y. Lee, W. J. Lee, M. Shin, and M. Y. Chung, "Channel allocation and transmission power management scheme in software defined network-based WLAN environments," in *Proc. IEEE Int. Conf. Inf. Netw. (ICOIN)*, Jan. 2017, pp. 138–142.

[15] C. He, Y. Hu, Y. Chen, X. Fan, H. Li, and B. Zeng, "MUcast: Linear uncoded multiuser video streaming with channel assignment and power allocation optimization," *IEEE Trans. Circuits Syst. Video Technol.*, to be published. doi: 10.1109/TCSVT.2019.2897649.

[16] K. Lee, Y. Kim, S. Kim, J. Shin, S. Shin, and S. Chong, "Just-in-time WLANs: On-demand interference-managed WLAN infrastructures," in *Proc. IEEE 35th Annu. Int. Conf. Comput. Commun.*, Apr. 2016, pp. 1–9.

[17] F. Bouhafs, M. Seyedebrahimi, A. Raschella, M. Mackay, and Q. Shi, "Per-flow radio resource management to mitigate interference in dense IEEE 802.11 wireless LANs," *IEEE Trans. Mobile Comput.*, to be published. doi: 10.1109/TMC.2019.2903465.

[18] Y. Zhang, C. Jiang, Z. Han, S. Yu, and J. Yuan, "Interference-aware coordinated power allocation in autonomous Wi-Fi environment," *IEEE Access*, vol. 4, pp. 3489–3500, 2016.

[19] S. Kim, J. Yi, Y. Son, S. Yoo, and S. Choi, "Quiet ACK: ACK transmit power control in IEEE 802.11 WLANs," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, May 2017, pp. 1–9.

[20] S. Byeon, H. Kwon, Y. Son, C. Yang, and S. Choi, "RECONN: Receiver-driven operating channel width adaptation in IEEE 802.11ac WLANs," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Apr. 2018, pp. 1655–1663.

[21] T. Song, T.-Y. Kim, W. Kim, and S. Pack, "Adaptive and distributed radio resource allocation in densely deployed wireless lans: A game-theoretic approach," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 4466–4475, May 2018.

[22] A. Baid and D. Raychaudhuri, "Understanding channel selection dynamics in dense Wi-Fi networks," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 110–117, Jan. 2015.

[23] H. Ko, J. Lee, and S. Pack, "Joint optimization of channel selection and frame scheduling for coexistence of LTE and WLAN," *IEEE Trans. Veh. Technol.*, vol. 67, no. 7, pp. 6481–6491, Jul. 2018.

[24] S. M. Kala, V. Sathya, M. P. K. Reddy, B. Lala, and B. R. Tamma, "A socio-inspired CALM approach to channel assignment performance prediction and WMN capacity estimation," *J. Netw. Comput. Appl.*, vol. 125, no. 1, pp. 42–66, 2019.

[25] S. Alam, N. Aqdas, I. M. Qureshi, S. A. Ghauri, and M. Sarfraz, "Joint power and channel allocation scheme for IEEE 802.11 af based smart grid communication network," *Future Gener. Comput. Syst.*, vol. 95, pp. 694–712, Jun. 2019.

[26] S. Ali, A. Ahmad, R. Iqbal, S. Saleem, and T. Umer, "Joint RRH-association, sub-channel assignment and power allocation in multi-tier 5G C-RANs," *IEEE Access*, vol. 6, pp. 34393–34402, 2018.

[27] B. Khamidehi, A. Rahmati, and M. Sabbaghian, "Joint sub-channel assignment and power allocation in heterogeneous networks: An efficient optimization method," *IEEE Commun. Lett.*, vol. 20, no. 12, pp. 2490–2493, Dec. 2016.

[28] J.-M. Kang, I.-M. Kim, and D. I. Kim, "Joint optimal mode switching and power adaptation for nonlinear energy harvesting SWIPT system over fading channel," *IEEE Trans. Commun.*, vol. 66, no. 4, pp. 1817–1832, Apr. 2018.

[29] M. Seyedebrahimi, F. Bouhafs, A. Raschella, M. Mackay, and Q. Shi, "Fine-grained radio resource management to control interference in dense Wi-Fi networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2017, pp. 1–6.

[30] Y. Wei, F. R. Yu, M. Song, and Z. Han, "User scheduling and resource allocation in HetNets with hybrid energy supply: An actor-critic reinforcement learning approach," *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 680–692, Jan. 2018.

[31] C. Jiang, H. Zhang, Y. Ren, Z. Han, K.-C. Chen, and L. Hanzo, "Machine learning paradigms for next-generation wireless networks," *IEEE Wireless Commun.*, vol. 24, no. 2, pp. 98–105, Apr. 2017.

[32] C. Zhang, P. Patras, and H. Haddadi, "Deep learning in mobile and wireless networking: A survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2224–2287, 3rd Quart., 2019. doi: 10.1109/COMST.2019.2904897.

[33] M. Liu, T. Song, and G. Gui, "Deep cognitive perspective: Resource allocation for NOMA-based heterogeneous IoT with imperfect SIC," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2885–2894, Apr. 2019.

[34] J. Wang, Y. Ding, S. Bian, Y. Peng, M. Liu, and G. Gui, "UL-CSI data driven deep learning for predicting DL-CSI in cellular FDD systems," *IEEE Access*, vol. 7, pp. 96105–96112, 2019.

[35] Y. Wang, M. Liu, J. Yang, and G. Gui, "Data-driven deep learning for automatic modulation recognition in cognitive radios," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 4074–4077, Apr. 2019.

[36] M. Liu, J. Yang, T. Song, J. Hu, and G. Gui, "Deep learning-inspired message passing algorithm for efficient resource allocation in cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 641–653, Jan. 2018.

[37] L. Xiao, Y. Li, C. Dai, H. Dai, and H. V. Poor, "Reinforcement learning-based NOMA power allocation in the presence of smart jamming," *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 3377–3389, Apr. 2018.

[38] A. Azari and C. Cavdar, "Self-organized low-power IoT networks: A distributed learning approach," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2018, pp. 1–7.

[39] Z. Han, T. Lei, Z. Lu, X. Wen, W. Zheng, and L. Guo, "Artificial intelligence-based handoff management for dense WLANs: A deep reinforcement learning approach," *IEEE Access*, vol. 7, pp. 31688–31701, 2019.

[40] M. A. Alsheikh, S. Lin, D. Niyato, and H. P. Tan, "Machine learning in wireless sensor networks: Algorithms, strategies, and applications," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 4, pp. 1996–2018, 4th Quart., 2014.

[41] Y. Zhou, X.-Y. Li, M. Liu, Z. Li, S. Tang, X. Mao, and Q. Huang, "Distributed link scheduling for throughput maximization under physical interference model," in *Proc. IEEE INFOCOM*, May 2012, pp. 2691–2695.
[42] Y. Zhou, X.-Y. Li, M. Liu, X. Mao, S. Tang, and Z. Li, "Throughput optimizing localized link scheduling for multihop wireless networks under physical interference model," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 10, pp. 2708–2720, Oct. 2014.
[43] H. Zhai and Y. Fang, "Physical carrier sensing and spatial reuse in multirate and multihop wireless ad hoc networks," in *Proc. IEEE 25th Int. Conf. Comput. Commun. (INFOCOM)*, Apr. 2006, pp. 1–12.
[44] C. Xu, Z. Han, Q. Wang, G. Zhao, and S. Yu, "Modelling the impact of interference on the energy efficiency of wlans," *Concurrency Comput., Pract. Exper.*, vol. 31, no. 17, p. e5217, Sep. 2019. doi: 10.1002/cpe.5217.
[45] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Jul./Oct. 1948.
[46] C. Xu, Z. Han, G. Zhao, and S. Yu, "A sleeping and offloading optimization scheme for energy-efficient WLANs," *IEEE Commun. Lett.*, vol. 21, no. 4, pp. 877–880, Apr. 2017.
[47] R. Cohen, L. Katzir, and D. Raz, "An efficient approximation for the generalized assignment problem," *Inf. Process. Lett.*, vol. 100, no. 4, pp. 162–166, 2006.
[48] R. S. Sutton and A. G. Barto, *Introduction to Reinforcement Learning*, vol. 135. Cambridge, MA, USA: MIT Press, 19980.

GUOFENG ZHAO received the M.E. degree from Northwestern Polytechnic University, in 1996, and the Ph.D. degree from Chongqing University, in 2003.

He is currently a Professor with the School of Communication Engineering and the Director of the Research Center of Future Internet, Chongqing University of Posts and Telecommunications. He has undertaken more than 20 projects or programs, including the National Basic Research Program of China and the Natural Science Foundation of China. He has published more than 100 articles and holds six patents. His research interests include the future Internet, the mobile Internet, network management, and network security.

YONG LI is currently a Research Staff Member with the Future Networks Research Institute, Chongqing University of Posts and Telecommunications, China. His research interests include wireless network resource management, machine learning, and software-defined networking.

CHUAN XU received the B.E. and M.E. degrees in communication engineering from the Chongqing University of Posts and Telecommunications, Chongqing, China, in 2003 and 2006, respectively, and the Ph.D. degree in control theory and engineering from Chongqing University, China, in 2012.

He is currently a Professor with the School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications. His research interests include the areas of wireless communication, software-defined networking, and the Internet of Things.

ZHENZHEN HAN is currently pursuing the Ph.D. degree in information and communication engineering with the Future Networks Research Institute, Chongqing University of Posts and Telecommunications, China. Her research interests include wireless network resource management, network measurement, and software-defined networking.

YUAN XING is currently pursuing the Ph.D. degree in information and communication engineering with the Future Networks Research Institute, Chongqing University of Posts and Telecommunications, China. Her research interests include the Industrial Internet of Things and time-sensitive networking.

SHUI YU (SM'12) is currently a Professor with the School of Software, University of Technology Sydney, Australia. His research interests include security and privacy, networking, big data, and mathematical modeling. He has published two monographs and edited two books, more than 200 technical articles, including top journals and top conferences, such as IEEE TPDS, TC, TIFS, TMC, TKDE, TETC, ToN, and INFOCOM. He initiated the research field of networking for big data, in 2013. His h-index is 35. He actively serves his research communities in various roles.

Dr. Yu is also a member of AAAS and ACM and a Distinguished Lecturer of the IEEE Communication Society. He is currently serving on a number of prestigious editorial boards, including the IEEE COMMUNICATIONS SURVEYS AND TUTORIALS (Area Editor) and *IEEE Communications Magazine* (Series Editor). He has served many international conferences as a member of the organizing committee, such as the Publication Chair of the IEEE Globecom 2015 and IEEE INFOCOM 2016 and 2017, and the General Chair of the ACSW 2017.