IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 18, 2018, accepted December 18, 2018, date of publication January 1, 2019, date of current version January 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2890210

A Distributed Multi-Agent RL-Based Autonomous Spectrum Allocation Scheme in D2D Enabled Multi-Tier HetNets

KAMRAN ZIA[1], NAUMAN JAVED[1], MUHAMMAD NADEEM SIAL[2], SOHAIL AHMED[3], ASAD AMIR PIRZADA[1], AND FARRUKH PERVEZ[1]

[1]Department of Avionics Engineering, National University of Sciences and Technology, Islamabad 44000, Pakistan
[2]Department of Informatics, King's College London, London WC2R 2LS, U.K.
[3]Department of Avionics Engineering, Air University, Islamabad 44000, Pakistan

Corresponding author: Kamran Zia (avionica70b@gmail.com)

ABSTRACT Multi-tier heterogeneous networks (HetNets) and device-to-device (D2D) communication are vastly considered in 5G networks. The interference mitigation and resource allocation in the D2D enabled multi-tier HetNets is a cumbersome and challenging task that cannot be solved by the conventional centralized resource allocation techniques proposed in the literature. In this paper, we propose a distributed multi-agent learning-based spectrum allocation scheme in which D2D users learn the wireless environment and select spectrum resources autonomously to maximize their throughput and spectral efficiency (SE) while causing minimum interference to the cellular users. We have employed the distributed learning in a stochastic geometry-based realistic multi-tier heterogeneous network to validate the performance of our scheme. The proposed scheme enables the D2D users to achieve higher throughput and SE, higher signal-to-interference-plus-noise ratio and low outage ratio for cellular users, and better computational time efficiency and performs well in the dense multi-tier HetNets without affecting network coverage compared with the distance based resource criterion and joint-resource allocation and link adaptation schemes.

INDEX TERMS D2D communication, multi-agent reinforcement learning, autonomous spectrum allocation, distributed reinforcement learning, heterogeneous networks, interference mitigation in D2D enabled HetNets.

I. INTRODUCTION

The increase in multimedia applications has given rise to the requirement of higher data rates in cellular networks. Moreover, applications such as Internet of Things (IoT), machine-to-machine communication, personalized TVs, video streaming, video conference calls and self-driven cars require network with high bandwidth, data rate and different latencies for their operation. The 5^{th} generation networks are envisioned to meet these requirements and provide higher data rates up to tens of Gbps to 1000-fold more devices [1]. Moreover, mobile phone usage has increased many fold over the past few years and is continuously rising. According to a Ericson Mobility Report 2017, mobile subscriptions have reached 7.79 billion in November 2017 and will rise to 9.12 billion by year 2023 [2]. In order to provide cellular services to such larger number of users, technologies like device to device communication and multi-tier heterogeneous networks (HetNets) are being developed and tested. HetNets comprises of micro, pico and femto base stations operating within the coverage of macro base stations (Figure 1). The radio access technology of each base station with asymmetrical transmit powers determine the cell sizes

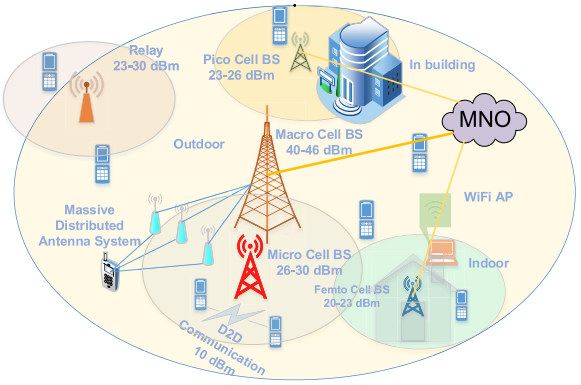

FIGURE 1. Infrastructure of heterogeneous network.

TABLE 1. List of abbreviations.

S No	Acronym	Definition
1	D2D	Device to Device Communication
2	SE	Spectral Efficiency
3	QoS	Quality of Service
4	DRC	Distance based Resource Criterion
5	RALA	Resource Allocation and Link Adaptation
6	HetNets	Heterogeneous Networks
7	FFR	Fractional Frequency Reuse
8	SIR	Signal to Interference Ratio
9	OPC	Opportunistic Power Control
10	eICIC	Enhanced Intercell Interference Coordination
11	RB	Resource Block
12	ACO	Ant Colony Optimization
13	KNN	K-Nearest Neighbors
14	SVM	Support Vector Machine
15	MIMO	Multiple Input Multiple Output
16	PCA	Principal Component Analysis
17	MDP	Markov Decision Process
18	MAB	Multi Arm Bandits
19	DUDe	Downlink-Uplink Decoupling
20	PPP	Poisson Point Process
21	OFDMA	Orthogonal Frequency Division Multiple Access
22	ASA	Autonomous Spectrum Allocation
23	UDN	Ultra Dense Network
24	SINR	Signal to Interference plus Noise Ratio

and add to the complex interference scenario in the uplink and downlink of the network [3].

Similarly, the device-to-device (D2D) communication is a promising technology that can provide higher data rates to the users in dense HetNets. D2D communication technology has become the focus of 5^{th} generation communication systems to support device-centric proximity based services, i.e., offloading calls to users located in close proximity, context aware applications, group multi-casting, multimedia data sharing, multi-player online gaming and public safety through emergency messaging, etc. The close proximity communication provides significant benefits in terms of throughput and hop gain thus lowering latency. The D2D communication can take place in either dedicated mode where base station assigns orthogonal frequency resources to D2D peers for communication or it can take place in shared mode in which it reuses frequency spectrum being used by conventional cellular users. Due to the increase in cellular user density, the networks are facing spectrum scarcity therefore performance benefits of D2D communication can be best achieved if it takes place in shared mode, also known as underlay mode. The D2D enabled heterogeneous networks can however, induce significant challenges in terms of co-tier and cross tier interference and resource allocation becomes quite challenging [4].

A. RELATED WORK

Different techniques have been employed in literature for resource allocation to D2D users operating in underlay mode. Author in [5] proposed distance based resource criterion (DRC) for sharing cellular resources with D2D users in single cell network. Although, this technique gives significant results in terms of throughput and outage of cellular users however it causes significant overhead to the base stations. Similar techniques are presented in [6] and [7] but they either require network assistance or cause significant overhead to the base stations. Ant colony optimization (ACO) based resource allocation is proposed in [8]. Graph coloring is used for mapping the interference among D2D users using Interference Level Indicator (ILI) term. Afterwards, ACO algorithm is employed to determine the optimum resource sharing among D2D and cellular users. Authors in [9] have proposed a network assisted interference mitigation scheme in which D2D users autonomously select RBs. The minimum transmission power level for D2D users is calculated based on the information received from base stations for all RBs and those RBs, which cause tolerable interference to macro and femto users, are selected. In order to reduce the overhead to the macro base stations, different machine learning based techniques are also proposed in the literature.

Machine learning has been considered as a powerful tool in solving different network problems in 5^{th} generation networks [10], [11]. Regression model, K-Nearest Neighbors (KNN) and Support Vector Machines (SVM) algorithms are employed in solving massive Multiple Input Multiple Output (MIMO) channel estimation/detection and user location/behavior classification problems. Similarly, Bayesian Learning is employed in spectrum sensing and detection, K-means clustering is employed in small cell and HetNets clustering and device to device user clustering problems, Principal Component Analysis (PCA) has been employed in anomaly, fault and intrusion detection problems and Markov Decision Process (MDP), Multi Armed Bandits (MAB) and Reinforcement learning have been employed in decision making under unknown channel conditions and spectrum sharing for D2D networks. The employment of machine learning in 5G is shown in Figure 2.

Reinforcement learning has been greatly employed in providing solution to the resource allocation problems in 5G networks. A Q-learning based resource allocation is proposed in [12] in a single tier network. Resources are shared among D2D and cellular users using Q-learning based strategy to maximize the network throughput. Authors in [13] proposed an expected Q-learning based resource allocation for LTE-U networks with downlink-uplink decoupling (DUDe) technique. The spectrum allocation problem is formulated as game theoretic model which incorporates load balancing, spectrum allocation and user association. The resources are allocated by base stations in autonomous manner with only limited information about the network. Authors in [14] have proposed a Q-learning based resource allocation scheme in two tier network. The D2D users and femto cell user determine the resource blocks in cooperative manner for meeting their QoS requirements. The devices learn the optimal strategy by sharing learnt information, to select RBs by taking actions (selecting RB) and measuring the feedback of actions in the form of rewards (QoS requirements). In this way, the users selects RB in autonomous manner and adapts to the changing network conditions as well. However, the network model considered has single macro base station and only

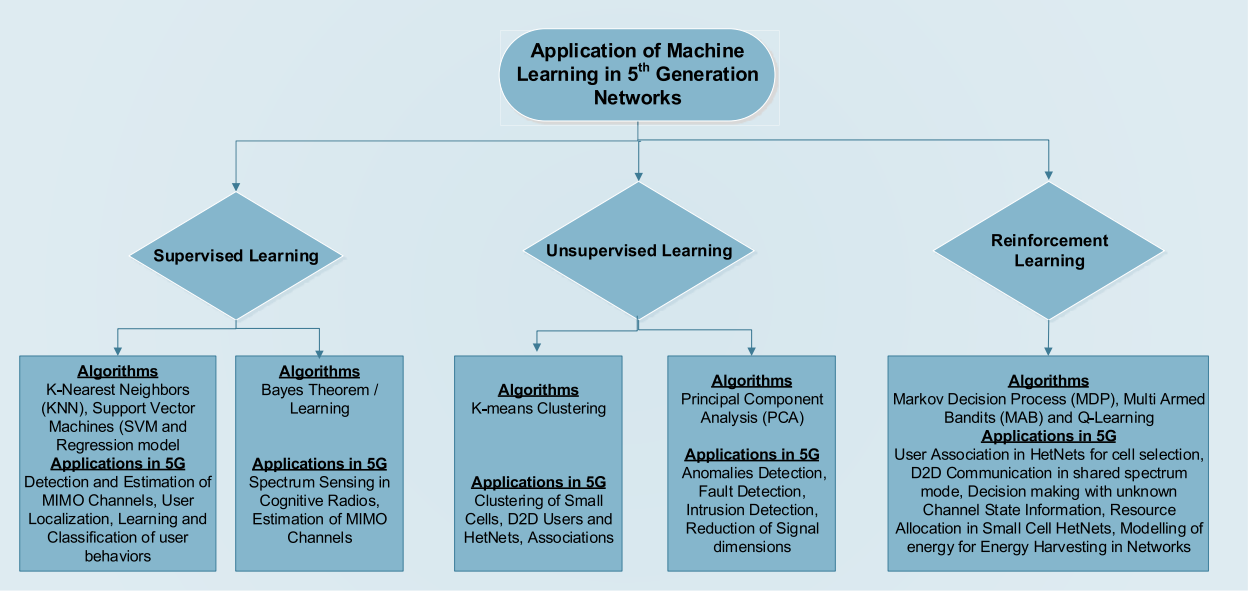

FIGURE 2. Machine learning in 5G.

one user associated to each femto base station, which is not practical. Moreover, sharing of information uses additional network resources, which is quite expensive. Authors in [15] proposed a Genetic Algorithm based joint resource allocation and user matching scheme to maximize the network throughput. The frequency resources are firstly allocated to the cellular users and then shared by the D2D users in a coordinated manner. Afterwards, optimal user matching is performed using Genetic Algorithm to maximize the system throughput.

B. MOTIVATION

The task of interference mitigation can be solved through power control techniques, fractional frequency reuse approach and resource allocation strategies. The Fractional Frequency Reuse (FFR) scheme [16] divides the region based on frequency selection as inner region and outer region and allocate resources to D2D users based on their location in certain region. Different power control schemes for single tier network have been proposed to mitigate interference among D2D and cellular users including Target Signal to Interference Ratio (SIR) tracking power control [17] and Opportunistic Power Control (OPC) [18]. Although, these power control techniques are distributive in nature however, they do not address the problem of interference in multi-tier heterogeneous network.

In this paper, we propose a distributive Reinforcement Learning (RL) based resource allocation scheme in multi-tier heterogeneous network to mitigate interference between D2D users and cellular users. In our previous work [19], we proposed a learning based autonomous spectrum allocation scheme for two-tier HetNets and shown the performance enhancement of D2D users in terms of throughput and spectral efficiency while meeting QoS requirements of conventional cellular users. In this paper, we have extended our work to multi-tier HetNets and analyzed the effects of increase in base station density and network tier over the performance of D2D users. Moreover, we have computed fairness of spectrum allocation among D2D users, additional memory requirements of the proposed algorithm and its computational time efficiency compared to other techniques.

Our work focuses on user-centric decision making for efficient spectrum selection by D2D users in a realistic and more practical stochastic geometry based multi-tier heterogeneous network. Enhanced inter cell interference coordination (eICIC) scheme is used to control the interference between multiple tiers of the network however D2D users employ online learning to mitigate interference. The D2D users employ an optimal learning strategy to determine suitable resource blocks (RBs) to meet their throughput requirements without the prior model of the environment. The prior model of the environment cannot be achieved due to unplanned deployment of femto base stations in the network and due to changing network conditions. The employed learning scheme is distributive in nature therefore; it reduces the complexity of system implementation. Moreover, it waves off the excessive overhead to the base stations, which is present in case of centralized scheduling of resources. The contribution of our paper is as follows:

- We propose an efficient scheme for resource sharing by D2D users in non co-operative manner using online learning in two-tier heterogeneous network. In our scheme, D2D users maximize their throughput through autonomous spectrum selection while causing minimal interference to the cellular users.

- A user-centric distributive decision making is proposed to shed off the processing load of base stations for resource allocation decisions.
- Our scheme maintains the QoS requirements of conventional cellular users in macro tier and secondary tier by meeting their SINR thresholds and outage ratios.
- It maintains the coverage probability of network in resource sharing mode by causing minimum degradation to the coverage of the network.
- Our scheme adapts to changing network conditions in fast and convenient manner to improve the throughput and spectral efficiency of D2D users.
- A stochastic geometry based realistic and practical network model with multiple macro and femto base stations is used to validate the effectiveness of our proposed scheme.
- The performance of proposed scheme is compared with other schemes to show its optimality in terms of throughput, spectral efficiency, coverage of the network, mean SINR of cellular users and their outage ratios.
- The computational optimality of our scheme is presented by showing the computational time comparison of our scheme with other schemes.
- Additional memory requirements of our scheme are presented and analysis is performed.
- The effect of increase in Base Stations Density on the performance of D2D users for proposed and comparison schemes is determined.
- The analysis of D2D performance with the increase in network tier is presented for proposed and comparison schemes.

C. REINFORCEMENT Q-LEARNING

The employed learning algorithm utilizes reinforcement Q-learning algorithm to determine the optimal policy π^* for taking decisions without the availability of perfect model of the environment. The cellular network environment keeps on changing with time and therefore it cannot be modeled. In order to determine the optimal policy for such a system, Q-learning is a useful tool for decision-making. Q-learning measures the quality of its action through feedback in the form of rewards and improves them without taking into account the factors affecting performance of the network like conditions of channel and user mobilities etc. Q-learning comprises of four parameters namely action a, state s, state transition probability $P_{s,s'}$ and reward $R_{s,a}$. The Q-learning algorithm runs on an agent that learns the environment through interaction with the environment and taking feedback of its actions. The state of an agent is its internal present condition or it may be external to agent such as usage of a particular spectrum resource. The reward is the measure of the feedback of the action that an agent takes in particular state. Consequently, the agent learns the environment and gains experience. Q-learning algorithm works as follows: The agent takes an action a_t in state s_t at some time instant t. As a result of the action a_t, the agent transitions to a new state according to $P_{s,s'}$ and feedback the measured reward $R(s_t,a_t)$ to the agent. The process is repeated for the next state and on in order to find the optimal Q-value for each pair of states and actions. The purpose of finding optimal Q-value is to determine the optimal strategy π_s^*. The optimal Q-value is determined as follows:

$$Q^*(s,a) = E[R(s,a) + \gamma \sum_{s \in S} P_{s,s'}(a) \max_{b \in A} Q^*(s,b)] \quad (1)$$

where A and S represents the set of states and actions of the learning agent and γ represents the discount factor. The optimal policy is determined as follows:

$$\pi_s^* = arg \max_{a \in A} Q^*(s,a) \quad (2)$$

The employed learning algorithm, however determines the optimal Q-value in a recurring manner as follows:

$$Q(s,a) = (1-\alpha)Q(s,a) + \alpha[R(s,a) + \gamma \max_{b \in A} Q^*(s',b)] \quad (3)$$

where α is the learning rate of the agent. The appropriate actions are rewarded by increasing their Q-values while inappropriate actions are punished and their Q-values are decreased. The Q-values are stored in a 2-dimensional matrix consisting of state vs action entries. The convergence of this algorithm to an optimal Q-value has been proven in [20] provided the states and actions are visited infinitely often. Therefore, the Q-learning algorithm employs two stages namely exploration and exploitation. In exploration, the agent explores different actions chosen randomly to discover highly rewarded actions while in exploitation, the agent chooses best actions in a given state according to learnt policy π_s*. The level of exploration and exploitation is balanced through different techniques like ϵ-greedy action selection or softmax selection.

The organization of the paper is as follows. Section II describes the system model and assumptions and Section III presents problem formulation and section IV presents online learning parameters and the autonomous spectrum allocation scheme for D2D users and interference mitigation. Section V presents the evaluation results and finally section VI will conclude the paper.

II. SYSTEM MODEL

In this work, we have considered a network model consists of a multi-tier heterogeneous network with macro, micro, pico and femto base stations and D2D users who operate in underlay mode. The base stations are spatially distributed in R^2 according to a 2D Poisson Point Process (PPP) φ_n with intensity λ_n where $n = \{1, 2, \ldots, N\}$. The cellular users are spatially distributed with intensity λ_u in R^2 according to 2D-PPP φ_u. The base stations are employing OFDMA scheme for distributing resources to the cellular users associated with them. Due to OFDMA scheme, there will be only one interferer in the coverage of each cell. We assume that D2D users have already been discovered in the network. The D2D transmitters are uniformly distributed in the area

R^2 and D2D receivers are lying in the isotropic region of radius r_d around the D2D transmitters. The distance between a particular D2D transmitter and its intended receiver is assumed to be uniformly distributed in the isotropic region. This assumption is of practical interest and has been considered in many works [21]–[23]. Other distributions of the distance between D2D transmitter and receiver can easily be incorporated. A half-duplex D2D communication mode is considered employing Single Carrier Frequency Division Multiple Access (SC-FDMA) scheme. Since the devices are communicating directly in D2D communication therefore, energy efficiency is an important constraint to save devices power. Due to Low peak to average power ratio (PAPR) in SC-FDMA mode [24], D2D users save energy to increase battery timings. The network model is illustrated in Figure 3.

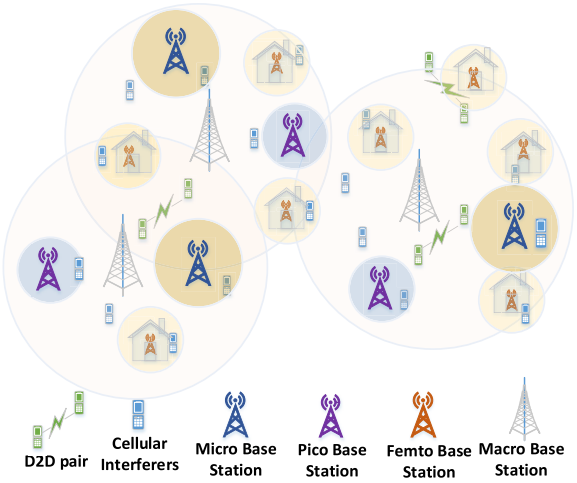

FIGURE 3. Illustration of network model.

The D2D users operate in underlay mode sharing same spectrum resources as used by the cellular users in coverage of macro and small cell base stations therefore they will cause interference to cellular users. The interference to the D2D users operating on a particular channel will come from one cellular user from each cell because of OFDMA scheme employed by them. Each D2D user from set of $\{1, 2, \ldots, D\}$ users will select one resource from set of $\{1, 2, \ldots, N\}$ spectrum resources available to them in an autonomous manner to maximize their throughput (TH_{D_n}) and keeping the interference to the cellular user under limits such that SINR of cellular user does not fall below a minimum required ζ_{min}. The objective of D2D users is to maximize their throughput while keeping the SINR of cellular user above required threshold.

We have considered a general path loss power model in which the signal power decays at the rate $r^{-\beta}$ with the propagation distance r. The path loss exponent is considered same for all the base stations and users in the network. Rayleigh fading has been considered to account for the channel fading, where its channel gain is assumed to be exponentially distributed with unit mean. Different notations used in our network model are listed in Table 2.

TABLE 2. Notations.

S No	Notations	Definitions
1	λ_n	Intensity Base Stations, 2D PPP
2	λ_u	Intensity of Cellular Users, 2D PPP
3	φ_n	Poisson Point Process for Base Stations
4	φ_u	Poisson Point Process for Cellular Users
5	β	Path Loss Exponent
6	P_m	Transmission Power of Macro BS
7	P_{mi}	Transmission Power of Micro BS
8	P_p	Transmission Power of Pico BS
9	P_f	Transmission Power of Femto BS
10	P_u	Transmission Power of Cellular Users
11	P_d	Transmission Power of D2D Transmitter
12	r_d	Distance between D2D Pair
13	TH_{D_n}	Throughput of D2D User
14	ζ_{min}	SINR Threshold for Cellular Users
15	θ	Intercell Interference Factor

The cellular users are associated with the base stations based on maximum received signal strength i-e the users will associate to the base station from which it receives the maximum power in downlink (DL). We have consider uplink (UL) spectrum sharing by D2D users with the cellular users. The small cell base stations also operate in underlay mode and utilize the same frequencies as used by the macro base stations. This will cause severe interference to the macro cell users thus causing Intercell interference. Several techniques have been introduced in LTE-A under the name enhanced Inter-cell Interference Coordination (eICIC) [25]. In this paper, we have assumed that eICIC have already been employed by the base stations to mitigate the inter-cell interference therefore our purpose is to mitigate interference caused by D2D users to the cellular users in shared mode. For inter-cell interference, we have assumed a inter-cell interference factor θ to denote the virtual reduction of interference among different cells and this factor θ is constant throughout the network. Therefore, the transmission power of the interferers will become:

$$P_I = P_u * \theta \tag{4}$$

We have employed Reinforcement Learning (RL) in a distributed manner where each D2D users learns it's environment to choose the best spectrum resource for communication to maximize its throughput while keeping interference to the cellular users under control. The learning is done in a recursive manner to determine a policy π^* to find the best available spectrum resource for communication.

III. PROBLEM FORMULATION

The resource allocation problem in our multi-tier network model with D2D users undelaying the uplink cellular spectrum has the objective of maximizing the D2D users throughput with minimal interference to the cellular users to keep their SINR above certain threshold (ζ_{min}). SINR of D2D receivers and SINR of cellular users are the main parameters that control the achievement of this objective. In order to calculate the SINR of D2D users, we need to find the sources

of interference to the D2D receiver D_r communicating with its transmitter using some resource block (RB) n from the set of $\{1, 2, \ldots, N\}$ resource blocks. As the base stations are employing OFDMA scheme, therefore, the D2D user will experience interference from one cellular user residing in each cell operating on the same RB n. The interference experienced by the D2D receiver D_r is given by:

$$I_{D_r} = \sum_{x_n \in X} P_I G(x_n, D_r) + \sum_{D_n \in D} P_d G(D_n, D_r) \tag{5}$$

where D_r is the underlay D2D receiver which is experiencing interference, x_n are the cellular users operating on RB n in each cell, P_I is the transmit power of cellular users and $G(x_n, D_r)$ is the interference gain of cellular user x_n towards D2D receiver D_r. Similarly, P_d is the transmit power of D2D transmitters, D_n are the D2D transmitters operating on RB n and $G(D_n, D_r)$ is the interference gain of D2D transmitters D_n towards D2D receiver D_r. The D2D transmitters are sharing the uplink resources of cellular users therefore; the interference experienced by D2D receivers will come from the cellular users, not the base stations. The interference gain of certain cellular user x_n or D2D transmitter D_n towards D_r over RB n is given by:

$$G(x_n, D_r) = F_{x_n, D_r} d_{x_n, D_r}^{-\beta} \tag{6}$$

where F_{x_n,D_r} is the channel fading component between x_n and D_r using RB n, d_{x_n,D_r} represents the distance between x_n and D_r and β represents the *Path Loss Exponent*. The SINR of the D2D receiver D_r over RB n is given by:

$$\zeta_{D_u} = \frac{P_d G(D_t, D_r)}{I_{D_r} + \sigma^2} \tag{7}$$

where ζ_{D_u} is the SINR of D2D receiver D_u, P_d is the transmit power of D2D transmitter D_t and $G(D_t, D_r)$ represents gain between D_t and D_r. σ^2 is the noise calculated by $\sigma^2 = N_o BW_{RB}$ where N_o is the thermal noise and BW_{RB} is the bandwidth of the resource block. In similar manner, the SINR of the conventional cellular user can also be calculated and is given by:

$$\zeta_{c_n} = \frac{P_u G(x_n, c_n)}{\sum_{D_n \in D} P_d G(D_n, c_n) + \sum_{x_n \in X} P_I G(x_n, c_n) + \sigma^2} \tag{8}$$

where ζ_{c_n} is the SINR of cellular user c_n operating on RB n and $G(D_n, c_n)$ and $G(x_n, c_n)$ are gains of D2D transmitters and cellular users in the direction of c_n respectively. $P_I = P_u\theta$ is the transmission power of the interferers incorporating the eICIC factor θ as described in section III. Note that the interference to the cellular user c_n is coming from other cellular users and D2D transmitters, operating on same RB n because of uplink (UL) resource sharing by D2D users. If the D2D users share downlink resources with the cellular users, then the interference to cellular user will come from base stations and D2D transmitters over particular RB. Our analysis and scheme is however, focused on UL resource sharing.

In order to calculate the throughput of D2D users and cellular users, SINR is required which can be calculated using 7 and 8 respectively. The throughput of a particular D2D user D_n over RB n can be calculated by Shannon formula as follows:

$$TH_{D_n} = BW_{RB} log_2(1 + \zeta_{D_n}) \tag{9}$$

The selection of a resource block n by a D2D user D_n in order to maximize its throughput is represented by a binary decision variable Z_{D_n} where

$$Z_{D_n} = \begin{cases} 1 & \text{if } D_n \text{ is transmitting over RB } n \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

The achievable throughput of the D2D user D_n can therefore, be written as:

$$TH_{D_n} = \sum_{n=1}^{N} Z_{D_n} BW_{RB} log_2(1 + \zeta_{D_n}) \tag{11}$$

In order to maintain the SINR of the cellular user above certain threshold (ζ_{min}), the interference caused by D2D users sharing uplink spectrum resources should be kept below the threshold of maximum tolerable interference by the cellular users I_{TH} as follows:

$$I_{D_n} = \sum_{D_n=1}^{D} Z_{D_n} P_d G(c_n, D_n) \leq I_{TH} \tag{12}$$

The objective of our spectrum allocation scheme is to maximize the throughput of the D2D users while maintaining the SINR of cellular users above ζ_{min}. Therefore, our optimization problem can be expressed as follows:

$$\begin{aligned} &\max_{n \in N} \sum_{n=1}^{N} Z_{D_n} BW_{RB} log_2(1 + \zeta_{D_n}) \\ &\text{subjected to } C.1 I_{D_n} \leq I_{TH} \quad \forall n \in N,\ D_n \in D \\ &\qquad C.2 Z_{D_n} \in \{0, 1\} \quad \forall n \in N,\ D_n \in D \\ &\qquad C.3 \zeta_{c_n} \geq \zeta_{min} \quad \forall n \in N,\ c_n \in C \end{aligned} \tag{13}$$

where ζ_{c_n} is presented in (8). The problem of allocation in (13) try to maximize the throughput of the D2D users while fulfilling the constraints defined in C.1, C.2 and C.3. The interference caused by the D2D users to the cellular users while reusing the cellular uplink resources is limited by a predefined tolerable interference threshold defined by constraint C.1. C.2 puts the restriction that each D2D user can use only one RB at a time and C.3 defines that the SINR of the cellular user remains above the predefined SINR threshold (ζ_{min}). Note that the C.3 is supported by the assumption that base stations can exchange their associated cellular users SINR with the D2D transmitters as they operate under their coverage [26].

IV. PROPOSED AUTONOMOUS SPECTRUM ALLOCATION SCHEME

We have considered a multi-agent reinforcement learning based system in which each D2D user acts as a learning agent. The agent collects the environment information from its own

TABLE 3. Basic working of autonomous spectrum allocation scheme.

Autonomous Spectrum Allocation Scheme
1. Initialize Q-Matrix Q for storing Q-values of the state action pairs
2. Observe the current state s_t of the environment at time t
3. Check ϵ for Exploration or Exploitation
4. **if** $Exploration$ **do**
5. Select action randomly and check the following
6. **if**
7. Throughput of the D2D user according to (11)
8. SINR of cellular users to guarantee its QoS according to (8)
9. C.1 and C.2 for fulfillment
10. **end** **if** $Exploitation$ **do**
11. Select action with the highest Q-value according to (17)
12. **end**
13. Evaluate Reward $R(s_t, a_t)$ for selected action a_t according to (15)
14. Update the Q-value in Q-Matrix against state-action according to (16)
15. Observe the new environment state s_{t+1} at time $(t+1)$
16. Repeat this process for next state s_{t+1}

and define it's state s_t at time t and performs an action a_t. Due to this action, the agent moves to new state and measures the reward $R(s_t, a_t)$ for the action. The agent calculates Q-value for the state s_t and action a_t using (3) and records it in its Q-table. The process is repeated with new state and after certain time, the agent populates it's Q-table with enough information to devise a strategy π^* to select RBs from set of $\{1, 2, \ldots N\}$ RBs in order to maximize its throughput while meeting all the constraints. Each agent (D2D user) has a role of learning agent, which tries to reach an optimum strategy for RB selection in different network states. The basic parameters of our learning system are as follows:

1) STATE

The learning agents rely on their locally observed environment conditions to define their state at a particular time slot t because of no cooperation among themselves. The state observed by agent i is defined as:

$$s_i^t = (i, n_i) \tag{14}$$

where n_i represents the RB selected by agent i.

2) ACTION

The agent has N number of resource blocks available to it for communication therefore the action of an agent is the selection of particular RB n i-e $(a_i) = (n_i)$

3) REWARD

The reward function is defined on the basis of throughput achieved by the agent for each state-action pair as $R_i(s_t, a_t)$ and is given by:

$$R_i(s_t, a_t) = \begin{cases} R_t(a_t) = TH_{D_n} & \text{if C.1 to C.3 satisfied} \\ -100 & \text{otherwise.} \end{cases} \tag{15}$$

Our learning system exploits the collected state-action Q values to determine the optimal strategy for spectrum allocation. The learning of the agents is divided into two stages namely exploration and exploitation. The agent continuously explores different actions in different environment states and updates it's Q-table in a recursive manner as follows:

$$Q_i(s_i, a_i) = (1-\alpha)Q_i(s_i, a_i) + \alpha[R_i(s_i, a_i) + \gamma \max_{l \in A_i} Q_i(s_i', l)] \tag{16}$$

where s_i is the current state and s_i' is the next state of the agent i. The learning agent tries to learn a strategy to maximize the future long term reward. The learning system relies on the SINR and data rates etc to take the actions in exploration phase. In order to balance the exploration and exploitation, ϵ-greedy method has been employed. The system starts with high exploration rate ϵ to find maximum actions with high Q-values to exploit. The actions that meet the constraints are rewarded while others are punished by giving negative reward. In this way, when the network conditions change, the agent automatically adapts to select a new action for meeting the constraints and maximizing it's throughput. The selection of actions in the exploitation phase are done as follows:

$$a_i = argmax(Q_i(s_i, a_i)) \tag{17}$$

In our scheme, each agent learns its own strategy to maximize its throughput. The pseudo code for the employed learning system for autonomous spectrum allocation by D2D users is given in Table 3.

V. PERFORMANCE EVALUATION

We have evaluated our Autonomous Spectrum Allocation (ASA) Scheme for D2D users underlaying UL cellular resources through extensive simulations in MATLAB.

The simulations were run in a multi-tier stochastic geometry based heterogeneous network with multiple macro, micro, pico and femto base stations, deployed randomly using PPP and are employing OFDMA scheme. Multiple D2D users are uniformly distributed in the network and reuse the UL cellular resources. A set of 15 resource blocks are distributed to the base stations as well as D2D users whereas each resource block has 12 sub-carriers having bandwidth of 180 KHz each. Therefore, each RB has a bandwidth of $12 \times 180 = 2160$ KHz available. The D2D users autonomously selects resource blocks from these 15 available RBs to maximize their throughput while maintaining QoS requirements of cellular users. We simulated the proposed scheme using Monte Carlo technique comprising of 10000 simulation runs. In order to capture the realistic and dynamically changing network state, the channel fading effects are generated on random basis in each simulation run. The cellular users association with base stations is determined in each simulation run to determine their associated cell IDs which are used to determine their SINR in the uplink direction. The simulation and ASA scheme parameters are chosen as per references and are listed in Table 4.

TABLE 4. Network model and simulation parameters.

Simulation Parameters Used	Value
Bandwidth of System, BW	30 MHz
Resource Blocks, N	15
Macro BS Tx Power, P_m	46 dBm
Micro BS Tx Power, P_{mi}	26 dBm
Pico BS Tx Power, P_p	23 dBm
Femto BS Tx Power, P_f	20 dBm
D2D Tx Power, P_d	10 dBm
Cellular User Tx Power, P_u	20 dBm
Macro BS Intensity, λ_m	$2\text{x}10^{-6}$
Micro BS Intensity, λ_{mi}	$6\text{x}10^{-6}$
Femto BS Intensity, λ_p	$8\text{x}10^{-6}$
Femto BS Intensity, λ_f	$1\text{x}10^{-5}$
Cellular User Intensity, λ_f	$6\text{x}10^{-4}$
No of D2D pairs, D	Variable
eICIC Factor, θ	-20 dB
Path Loss Exponent, β	4
Thermal Noise Density	-174 dBm/Hz
ASA Scheme Parameters	**Value**
Learning rate, α	0.2
Discount factor, γ	0.8
Exploration rate, ϵ	Dynamic

In order to analyze the impact of learning capabilities of the D2D users on their performance metrics (Throughput and Spectral Efficiency), we have compared the evaluation results of our scheme with Distance based Resource Criterion (DRC) scheme [5] and Joint-RALA scheme [27] which are separately implemented in MATLAB under identical circumstances in terms of users and network. The DRC scheme utilizes the distance between D2D users and cellular users to determine a RB allocation to the D2D users. The processing is centrally done by base stations and the GPS locations of all cellular users and D2D users are required by the base stations. Although, DRC gives significant performance in terms of throughput however, practical implementation of this technique is not possible due very large overhead of calculating distances in dense 5G HetNets. Moreover, a lot of cooperation will be required to employ this technique in multi-tier heterogeneous network. Joint-RALA scheme proposes a combined algorithm for resource allocation and link adaptation and it also supports functionality of carrier aggregation in 5G network.

A. PERFORMANCE EVALUATION OF D2D USERS

The performance of our scheme is compared with others in terms of aggregate throughput for D2D users (Figure 5), Spectral efficiency in bits/Hz (Figure 6) and CDF of average throughput of D2D users (Figure 7). Moreover, the fairness of our scheme is determined using Jane's Fairness Index calculated by $f(x_1, x_2, \ldots) = \frac{(\sum_{i=1}^{J} x_i)^2}{J \sum_{i=1}^{J} x_i^2}$ and is shown in Figure 4. Our scheme achieves good level of fairness among multiple D2D users by enabling all users to achieve same average throughput. This is because of the learning capability of each D2D users in the network who are populating their own Q-tables for intelligent selection of RBs. The fairness plot is clearly indicating the fair distribution of network resources among D2D users which in one of the major requirement in resource allocation problems.

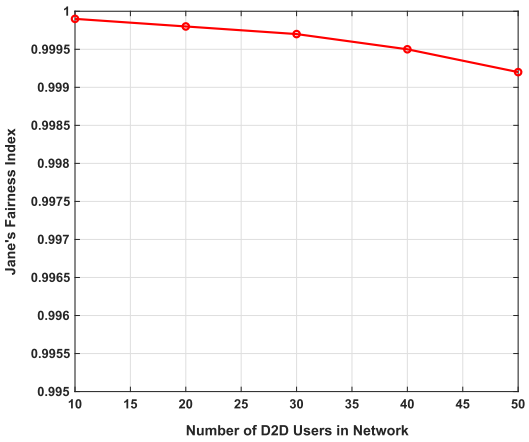

FIGURE 4. Jane's fairness index for ASA scheme.

Figure 5 and 6 show the plots of aggregate throughput of D2D users against number of D2D users in the network and spectral efficiency for different schemes. A higher aggregate throughput and spectrum efficiency is achieved by our scheme compared to other techniques. This is because our scheme devises the spectrum allocation problem with the objective of throughput maximization and makes use of distributed Q-learning as a tool to achieve this objective. As the network conditions change at every moment, the decision making needs to adapt to these changing network conditions. Our system undergoes plenty of interactions with the environment, therefore it is able to adapt to changing conditions and achieves higher throughput and spectral efficiency.

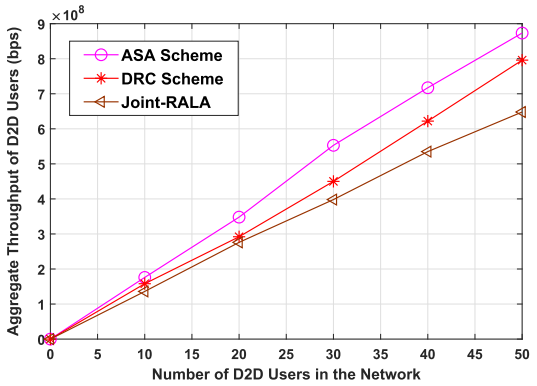

FIGURE 5. Aggregate throughput of D2D Users against number of D2D users in the network.

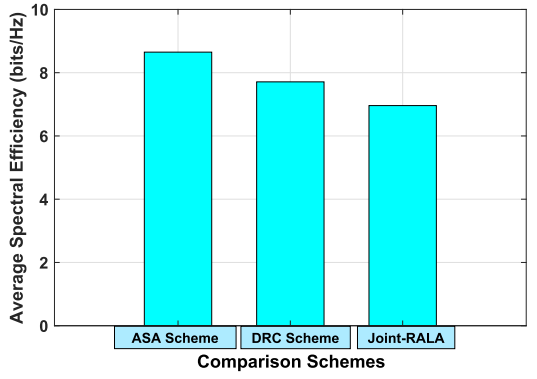

FIGURE 6. Spectral efficiency comparison of different schemes.

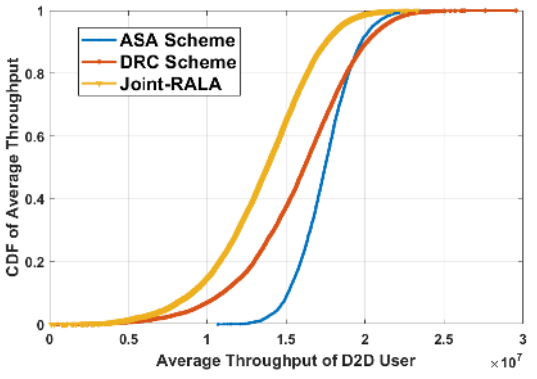

FIGURE 7. Comparison of CDF of average throughput of D2D user in network.

The CDF plots of average throughput for different schemes in figure 7 also indicates the achievement of higher average throughput by our scheme. The steepness in CDF plot for our scheme shows that D2D users are getting throughput higher than 10 to 15 Mbps all the time while throughput of less than 10 Mbps are achieved by other schemes in certain cases.

B. PERFORMANCE EVALUATION OF CELLULAR USERS

The effectiveness of Resource Allocation scheme cannot be validated if the performance metrics of Conventional Cellular Users are not met. The Key Performance Indicators (KPI) of

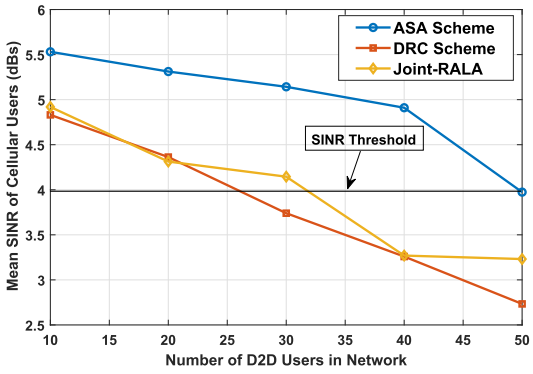

FIGURE 8. Mean SINR of conventional cellular users in network against number of D2D users in the network.

cellular users in the network include their SINR and outage ratio. Therefore, the mean SINR of conventional cellular users in the network and outage ratios for different number of D2D users in network is plotted in Figure 8 and 9 respectively. It is clear from Figure 8 that the mean SINR of cellular user in our scheme is higher compared to other schemes. This is because our scheme allocates spectrum while causing minimum interference to cellular users and keeps the SINR of cellular user above threshold ζ_{min}. The control over the interference to the cellular user is maintained through the assignment of the rewards to the learning agents (D2D Users). The RBs selected by D2D users which affects the SINR of cellular users below ζ_{min} are assigned negative rewards to avoid the selection of such RBs. DRC and Joint-RALA schemes do not ensure the ζ_{min} of cellular users therefore achieves lower mean SINR in the network. Due to the control over the interference caused to cellular user in our scheme, it achieves better outage ratio compared to other schemes as shown in Figure 9.

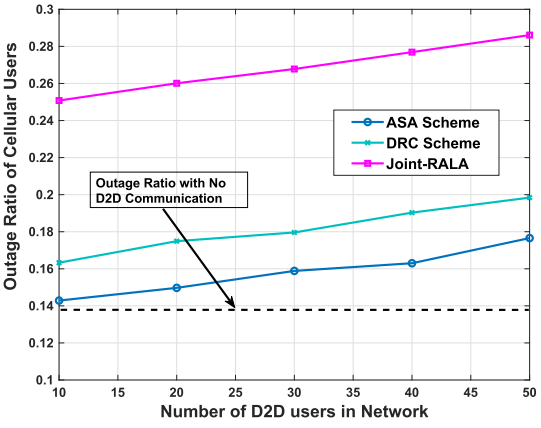

FIGURE 9. Outage ratio of conventional cellular users against number of D2D users in the network.

C. COVERAGE ANALYSIS OF NETWORK UNDERLAYING D2D COMMUNICATION

According to Andrews et al. [28], the coverage probability is simply $Pr[SINR > \zeta_{min}]$, where ζ_{min} is the minimum SINR threshold required by the cellular user. The SINR based

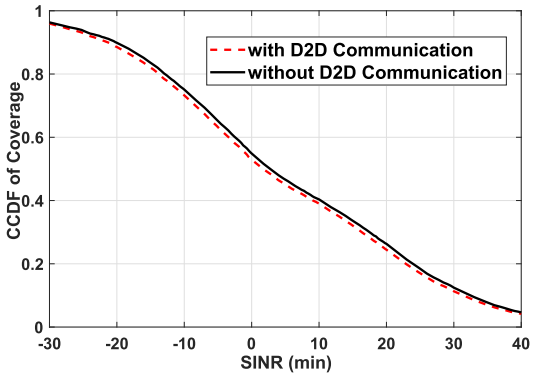

FIGURE 10. CCDF of coverage of network with 50 D2D users-ASA scheme.

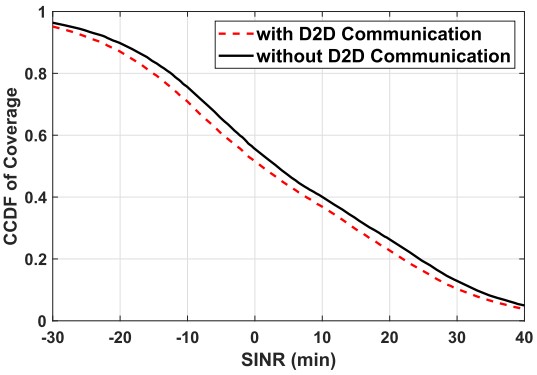

FIGURE 11. CCDF of coverage of network with 50 D2D users-DRC scheme.

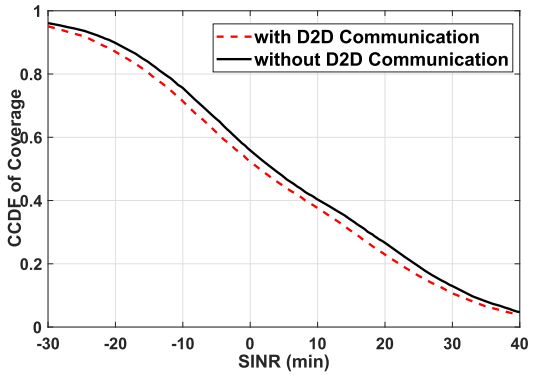

FIGURE 12. CCDF of coverage of network with 50 D2D users-joint RALA scheme.

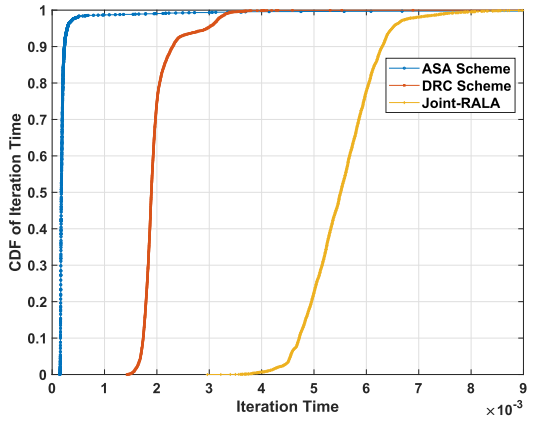

FIGURE 13. Computation time comparison of different schemes for resource allocation (50 D2D users in network).

coverage probability calculation is a good way to measure the performance of the network at the physical layer. The SINR based coverage analysis has been considered by many authors in their analysis [29], [30]. We have also performed SINR based coverage analysis of our network and in order to visualize the coverage, the cumulative distribution function of the coverage of our network are plotted for different schemes. The plots are taken with no D2D user in the network and with 50 D2D users in the network sharing the uplink resources. The objective of this analysis is to find the effect of spectrum sharing by D2D users over the coverage of the network. It can be seen in Figure 10, 11 and 12 that our scheme causes minimum reduction in the coverage of the network by underlaying 50 D2D users in the network. This is because of the interference control mechanism employed in our scheme to meet QoS requirements of the cellular users.

D. COMPUTATIONAL TIME ANALYSIS

Spectrum scarcity has tunneled the research in the field of D2D enabled HetNets to spectrum sharing based solutions. In order to allocate resources to the D2D users sharing resources with the cellular users, the decision time is of considerable importance. A longer decision time for spectrum selection will ultimately lead to lower throughput by the D2D users due to increased latency. Our scheme takes the decision of RB selection in considerable less time compared to other techniques and outperforms them in terms of throughput and spectral efficiency; please refer to Figure 13. This is because, the resource selection decisions have been distributed and taken by the D2D users themselves unlike the decisions taken by base station centrally in other schemes. The centralized resource allocation schemes [5], [27] are computationally more extensive and therefore takes more time in determining optimal resource for D2D users. This not only uses more computational power but also causes significant overhead to the base stations. Contrary to this, our distributed scheme is computationally more efficient both in terms of computation power and computation time.

E. MEMORY REQUIREMENTS

Look up tables in the form of Q-tables are an essential component of the Reinforcement Q-learning algorithm in order to determine the optimal strategy through mapping of best state-action pairs for every state. These Q-tables help D2D users in effective decision making for resource selection. Since memory is a critical resource, it becomes necessary to determine the memory requirements of the learning D2D users. Considering all the values in Q-tables of type double that occupies 8 bytes in memory, the memory space required for our scheme is 1800 bytes $=$ 1.757 KB

(15 States, 15 Actions). As each D2D user is separately storing its Q-table in its memory therefore, each user will be utilizing 1.757 KB of memory space. On the contrary in case of centralized resource allocation, the base stations will be computing the distance and measuring link gains for all the admitted D2D users in the network therefore, memory requirements for storing and processing this information will be considerably large compared to our scheme. The increase in number of D2D or cellular users will not increase memory requirements of our algorithm whereas, the memory requirements of centralized resource allocation schemes increase with the number of admitted users in the network.

F. EFFECT OF BASE STATIONS DENSITY ON D2D PERFORMANCE

The 5^{th} generation networks are envisioned to provide network coverage to a larger number of users. This will require Ultra Dense Networks (UDN) with a larger number of base stations to extend network coverage. Therefore, it is necessary to analyze the effect of increase in base station density on the performance of D2D users in terms of aggregate throughput for the proposed scheme. Figure 14 show the effect of increase in macro base station density on the aggregate throughput of 50 D2D users in our network. The increase in macro base stations density will increase the small cell base stations density as well due to linear relationship ($\lambda_{mi} = 3 * \lambda_m$, $\lambda_p = 4 * \lambda_m$ and $\lambda_f = 5 * \lambda_m$) between them. The figure indicates a decrease in aggregate throughput with the increase in base stations density. This is because the increase in base stations causes increase in number of cellular users operating on a particular resource block due to OFDMA assumption. This increase in number of cellular users in the network increases the interference over available resource blocks and therefore reduce the aggregate throughput of D2D users. The D2D users are exploiting spatial reuse of spectrum resources in the network through Reinforcement Learning therefore, increase in cellular users will limit this spatial reuse and hence the aggregate throughput reduces. However, our scheme is still performing better in UDN and enabling D2D users to achieve higher throughput compared to other schemes.

G. EFFECT OF NETWORK TIER ON D2D PERFORMANCE

Similar to the effect of increase in base station density over D2D users performance, we have analyzed the performance of our scheme with the increase in network tier in our network model. In order to find effect of tier on D2D users performance, we have considered two, three and four tier network. In three tier network, we have considered macro, micro and pico base stations. The relationship between macro and micro base stations is kept $\lambda_{mi} = 3 * \lambda_m$ while relationship between macro and pico base stations is kept $\lambda_p = 4 * \lambda_m$. The transmit power of micro base station is $P_{mi} = 26\ dBm$ while that of pico base station is $P_p = 23 dBm$. Similarly, in four tier network we have considered femto base stations as well in addition to macro, micro and pico base stations. The transmit

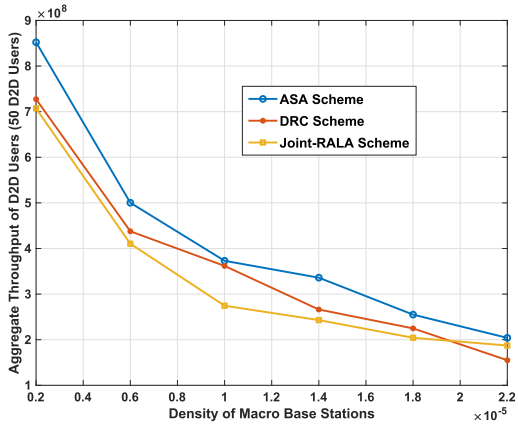

FIGURE 14. Effect of increase in base stations density on aggregate throughput of D2D users in network.

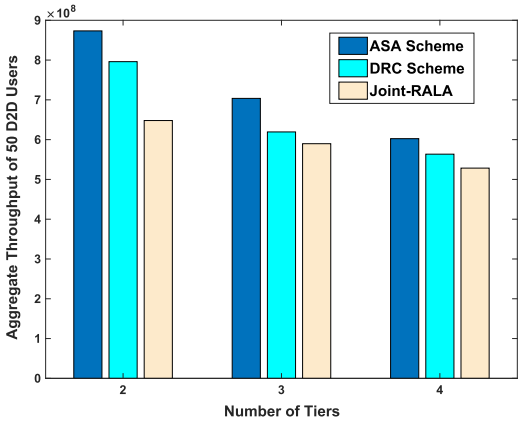

FIGURE 15. Effect of network tier on aggregate throughput of D2D users in the network.

power of femto base stations in kept 20 dBm in four tier network. Figure 15 shows the effect of tier on performance of D2D users.

The increase in network tier reduces the aggregate throughput of D2D users. Due to the OFDMA assumption in our network model, there is at least one user in each cell operating on some particular resource block therefore, total number of users operating on a particular resource block equals the number of cells or base stations. As the network tier increases, the number of base stations and the cellular interferers over a particular resource block will increase causing more interference to D2D users using same resource blocks. This increased interference limits the spatial reuse of resource blocks by D2D users thus reducing their aggregate throughput. The comparison plot of effect of tier on D2D users performance in Figure 15 clearly indicates that our scheme outperforms other schemes in multitier HetNets as well.

VI. CONCLUSION

In this paper, we have proposed an autonomous spectrum allocation scheme with distributed Q-learning based algorithm to allocate spectrum to the underlaying D2D users and control interference in D2D enabled multi-tier HetNets. Our scheme is distributed and will run on devices to enable them to select spectrum autonomously thus shedding off the

processing load of the base stations. The objective of our proposed scheme is to maximize the throughput of D2D users while maintaining the QoS requirements of the cellular users and cellular outage ratio. The learning methodology employed makes use of interactions with the environment to adapt to changing network conditions in fast and convenient manner to improve the throughput and spectral efficiency. The performance of proposed scheme is compared with other schemes to show its optimality in terms of Throughput, Spectral Efficiency, Mean SINR of cellular users, Outage ratios Coverage of the network and Computation time. Analysis have shown that the proposed technique outperforms other techniques in both ultra-dense networks and multi-tier HetNets.

REFERENCES

[1] R. El Hattachi and J. Erfanian, "5G white paper," Next Gener. Mobile Netw. Alliance, White Paper, 2015. [Online]. Available: https://www.ngmn.org/fileadmin/ngmn/content/images/news/ngmn_news/NGMN_5G_White_Paper_V1_0.pdf

[2] N. Heuveldop *et al.*, "Ericsson mobility report," Ericsson AB, Technol. Emerg. Business, Stockholm, Sweden, Tech. Rep. EAB-17, 2017, vol. 5964.

[3] T. O. Olwal, K. Djouani, and A. M. Kurien, "A survey of resource management toward 5G radio access networks," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1656–1686, 3rd Quart., 2016.

[4] W. H. Chin, Z. Fan, and R. Haines, "Emerging technologies and research challenges for 5G wireless networks," *IEEE Wireless Commun.*, vol. 21, no. 2, pp. 106–112, Apr. 2014.

[5] H. Wang and X. Chu, "Distance-constrained resource-sharing criteria for device-to-device communications underlaying cellular networks," *Electron. Lett.*, vol. 48, no. 9, pp. 528–530, Apr. 2012.

[6] T. Peng, Q. Lu, H. Wang, S. Xu, and W. Wang, "Interference avoidance mechanisms in the hybrid cellular and device-to-device systems," in *Proc. IEEE 20th Int. Symp. Pers., Indoor Mobile Radio Commun.*, Sep. 2009, pp. 617–621.

[7] M. Zulhasnine, C. Huang, and A. Srinivasan, "Efficient resource allocation for device-to-device communication underlaying LTE network," in *Proc. IEEE 6th Int. Conf. Wireless Mobile Comput., Netw. Commun. (WiMob)*, Oct. 2010, pp. 368–375.

[8] E. Liotou, D. Tsolkas, N. Passas, and L. Merakos, "Ant colony optimization for resource sharing among D2D communications," in *Proc. IEEE 19th Int. Workshop Comput. Aided Modeling Design Commun. Links Netw. (CAMAD)*, Dec. 2014, pp. 360–364.

[9] A.-H. Tsai, L.-C. Wang, J.-H. Huang, and T.-M. Lin, "Intelligent resource management for device-to-device (D2D) communications in heterogeneous networks," in *Proc. 15th Int. Symp. Wireless Pers. Multimedia Commun. (WPMC)*, Sep. 2012, pp. 75–79.

[10] 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.

[11] 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.

[12] Y. Luo, Z. Shi, X. Zhou, Q. Liu, and Q. Yi, "Dynamic resource allocations based on Q-learning for D2D communication in cellular networks," in *Proc. 11th Int. Comput. Conf. Wavelet Act. Media Technol. Inf. Process. (ICCWAMTIP)*, Dec. 2014, pp. 385–388.

[13] Y. Hu, R. MacKenzie, and M. Hao, "Expected Q-learning for self-organizing resource allocation in LTE-U with downlink-uplink decoupling," in *Proc. 23th Eur. Wireless Conf. Eur. Wireless*, May 2017, pp. 1–6.

[14] I. AlQerm and B. Shihada, "A cooperative online learning scheme for resource allocation in 5G systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1–7.

[15] C. Yang, X. Xu, J. Han, W. ur Rehman, and X. Tao, "GA based optimal resource allocation and user matching in device to device underlaying network," in *Proc. IEEE Wireless Commun. Netw. Conf. Workshops (WCNCW)*, Apr. 2014, pp. 242–247.

[16] H. S. Chae, J. Gu, B.-G. Choi, and M. Y. Chung, "Radio resource allocation scheme for device-to-device communication in cellular networks using fractional frequency reuse," in *Proc. 17th Asia–Pacific Conf. Commun. (APCC)*, Oct. 2011, pp. 58–62.

[17] G. J. Foschini and Z. Miljanic, "A simple distributed autonomous power control algorithm and its convergence," *IEEE Trans. Veh. Technol.*, vol. 42, no. 4, pp. 641–646, Nov. 1993.

[18] K.-K. Leung and C. W. Sung, "An opportunistic power control algorithm for cellular network," *IEEE/ACM Trans. Netw.*, vol. 14, no. 3, pp. 470–478, Jun. 2006.

[19] K. Zia, N. Javed, M. N. Sial, S. Ahmed, and F. Pervez, "Multi-agent RL based user-centric spectrum allocation scheme in D2D enabled hetnets," in *Proc. IEEE 23rd Int. Workshop Comput. Aided Modeling Design Commun. Links Netw. (CAMAD)*, Barcelona, Spain, Sep. 2018, pp. 1–6.

[20] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Mach. Learn.*, vol. 8, nos. 3–4, pp. 279–292, 1992.

[21] C. Ma, W. Wu, Y. Cui, and X. Wang, "On the performance of successive interference cancellation in D2D-enabled cellular networks," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Apr./May 2015, pp. 37–45.

[22] X. Lin and J. G. Andrews, "Optimal spectrum partition and mode selection in device-to-device overlaid cellular networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2013, pp. 1837–1842.

[23] X. Lin, R. Ratasuk, A. Ghosh, and J. G. Andrews, "Modeling, analysis, and optimization of multicast device-to-device transmissions," *IEEE Trans. Wireless Commun.*, vol. 13, no. 8, pp. 4346–4359, Apr. 2014.

[24] H. G. Myung, J. Lim, and D. J. Goodman, "Peak-to-average power ratio of single carrier FDMA signals with pulse shaping," in *Proc. IEEE 17th Int. Symp. Pers., Indoor Mobile Radio Commun.*, Sep. 2006, pp. 1–5.

[25] D. Lopez-Perez, I. Guvenc, G. de la Roche, M. Kountouris, T. Q. S. Quek, and J. Zhang, "Enhanced intercell interference coordination challenges in heterogeneous networks," *IEEE Wireless Commun.*, vol. 18, no. 3, pp. 22–30, Jun. 2011.

[26] I. Alqerm and B. Shihada, "Energy-efficient power allocation in multitier 5G networks using enhanced online learning," *IEEE Trans. Veh. Technol.*, vol. 66, no. 12, pp. 11086–11097, Dec. 2017.

[27] S. Rostami, K. Arshad, and P. Rapajic, "A joint resource allocation and link adaptation algorithm with carrier aggregation for 5G LTE-Advanced network," in *Proc. 22nd Int. Conf. Telecommun. (ICT)*, Apr. 2015, pp. 102–106.

[28] J. G. Andrews, R. K. Ganti, M. Haenggi, N. Jindal, and S. Weber, "A primer on spatial modeling and analysis in wireless networks," *IEEE Commun. Mag.*, vol. 48, no. 11, pp. 156–163, Nov. 2010.

[29] M. N. Sial and J. Ahmed, "Analysis of K-tier 5G heterogeneous cellular network with dual-connectivity and uplink–downlink decoupled access," *Telecommun. Syst.*, vol. 67, no. 4, pp. 669–685, 2018.

[30] K. Smiljkovikj, P. Popovski, and L. Gavrilovska, "Analysis of the decoupled access for downlink and uplink in wireless heterogeneous networks," *IEEE Wireless Commun. Lett.*, vol. 4, no. 2, pp. 173–176, Apr. 2015.

KAMRAN ZIA received the bachelor's degree in avionics engineering from the College of Aeronautical Engineering (CAE), National University of Sciences and Technology, Islamabad, Pakistan, in 2011, where he is currently pursuing the M.S. degree in avionics engineering. He received the Best Avionics System Design while pursuing the bachelor's degree from the Department of Avionics Engineering, CAE. His research interests include digital signal processing, wireless communication, interference mitigation and resource allocation in mobile networks, software-defined radios, and radar DSP.

NAUMAN JAVED received the bachelor's degree from the College of Aeronautical Engineering (CAE), National University of Sciences and Technology (NUST), Islamabad, and the Ph.D. degree in electrical and computer engineering from the University of Massachusetts Amhesrt. He is currently an Associate Professor with CAE, NUST, where he is also the Head of the Communications Group. His research interests include brain computing, neuromorphic computing, computing architectures and processor design, and artificial intelligence and machine learning.

MUHAMMAD NADEEM SIAL received the B.E. degree (Hons.) in avionics engineering from the College of Aeronautical Engineering, National University of Science and Technology, Pakistan, in 1997, the M.S. degree in information security from Sichuan University, Chengdu, China, in 2007, and the Ph.D. degree in wireless communication networks from the Department of Electrical Engineering, COMSATS University Islamabad, Pakistan. He was a Research Fellow with King's College London, where he was involved in V2X communication. His research interests include 5G heterogeneous cellular networks, including the analysis of decoupled cell association, dual connectivity, device-to-device communication, downlink-uplink resources allocation optimization, and interference management of HetNets.

SOHAIL AHMED received the bachelor's degree in avionics engineering from the College of Aeronautical Engineering, Risalpur, Pakistan, in 1992, the M.S. degree in telecomm engineering from the National University of Sciences and Technology, Islamabad, Pakistan, in 2003, and the Ph.D. degree in wireless communication from the University of Southampton, Southampton, U.K., in 2007. He is currently an Assistant Professor with Air University, PAC Campus, Islamabad. His research interests include frequency hopping systems, software defined radios, radar detection, and cognitive radars.

ASAD AMIR PIRZADA received the master's degrees in computer science and information security and the Ph.D. degree in trust and security issues in ad-hoc wireless networks from the University of Western Australia. He is currently the Commandant of the College of Aeronautical Engineering, National University of Sciences and Technology, Pakistan. His research interests include data communications and network security.

FARRUKH PERVEZ received the bachelor's degree in avionics engineering from the National University of Sciences and Technology (NUST), Islamabad, Pakistan, in 2009, and the M.S. degree in electrical engineering from NUST, in 2018. He is currently an Assistant Professor with NUST. His research interests include wireless technologies for emergency response, 5G backhaul challenges, and user cell association schemes for HetNets.