

Deep Reinforcement Learning Based Resource Allocation for V2V Communications

Hao Ye , *Student Member, IEEE*, Geoffrey Ye Li, *Fellow, IEEE*, and Biing-Hwang Fred Juang

Abstract—In this paper, we develop a novel decentralized resource allocation mechanism for vehicle-to-vehicle (V2V) communications based on deep reinforcement learning, which can be applied to both unicast and broadcast scenarios. According to the decentralized resource allocation mechanism, an autonomous “agent,” a V2V link or a vehicle, makes its decisions to find the optimal sub-band and power level for transmission without requiring or having to wait for global information. Since the proposed method is decentralized, it incurs only limited transmission overhead. From the simulation results, each agent can effectively learn to satisfy the stringent latency constraints on V2V links while minimizing the interference to vehicle-to-infrastructure communications.

Index Terms—Deep Reinforcement Learning, V2V Communication, Resource Allocation.

I. INTRODUCTION

VEHICLE-TO-VEHICLE (V2V) communications [1]–[3] have been developed as a key technology in enhancing the transportation and road safety by supporting cooperation among vehicles in close proximity. Due to the safety imperative, the quality-of-service (QoS) requirements for V2V communications are very stringent with ultra low latency and high reliability [4]. Since proximity based device-to-device (D2D) communications provide direct local message dissemination with substantially decreased latency and energy consumption, the Third Generation Partnership (3GPP) supports V2V services based on D2D communications [5] to satisfy the QoS requirement of V2V applications.

In order to manage the mutual interference between the D2D links and the cellular links, effective resource allocation mechanisms are needed. In [6], a three-step approach has been proposed, where the transmission power is controlled and the spectrum is allocated to maximize the system throughput with constraints on minimum signal-to-interference-plus-noise ratio (SINR) for both the cellular and the D2D links. In V2V communication networks, new challenges are brought about by high mobility vehicles. As high mobility causes rapidly changing wireless channels, traditional methods on resource management

for D2D communications with a full channel state information (CSI) assumption can no longer be applied in the V2V networks.

Resource allocation schemes have been proposed to address the new challenges in D2D-based V2V communications. The majority of them are conducted in a centralized manner, where the resource management for V2V communications is performed in a central controller. In order to make better decisions, each vehicle has to report the local information, including local channel state and interference information, to the central controller. With the collected information from vehicles, the resource management is often formulated as optimization problems, where the constraints on the QoS requirement of V2V links are addressed in the optimization constraints. Nevertheless, the optimization problems are usually NP-hard, and the optimal solutions are often difficult to find. As alternative solutions, the problems are often divided into several steps so that local optimal and sub-optimal solutions can be found for each step. In [7], the reliability and latency requirements of V2V communications have been converted into optimization constraints, which are computable with only large-scale fading information and a heuristic approach has been proposed to solve the optimization problem. In [8], a resource allocation scheme has been developed only based on the slowly varying large-scale fading information of the channel and the sum V2I ergodic capacity is optimized with V2V reliability guaranteed.

Since the information of vehicles should be reported to the central controller for solving the resource allocation optimization problem, the transmission overhead is large and grows dramatically with the size of the network, which prevents these methods from scaling to large networks. Therefore, in this paper, we focus on decentralized resource allocation approaches, where there are no central controllers collecting the information of the network. In addition, the distributed resource management approaches will be more autonomous and robust, since they can still operate well when the supporting infrastructure is disrupted or become unavailable. Recently, some decentralized resource allocation mechanisms for V2V communications have been developed. A distributed approach has been proposed in [15] for spectrum allocation for V2V communications by utilizing the position information of each vehicle. The V2V links are first grouped into clusters according to the positions and load similarity. The resource blocks (RBs) are then assigned to each cluster and within each cluster, the assignments are improved by iteratively swapping the spectrum assignments of two V2V links. In [11], a distributed algorithm has been designed to

Manuscript received April 1, 2018; revised September 3, 2018 and December 1, 2018; accepted January 23, 2019. Date of publication February 4, 2019; date of current version April 16, 2019. This work was supported in part by a Research Gift from Intel Corporation and in part by the National Science Foundation under Grants 1815637 and 1731017. The review of this paper was coordinated by Editors of CVIS. (*Corresponding author: Hao Ye.*)

The authors are with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: yehao@gatech.edu; liye@ece.gatech.edu; juang@ece.gatech.edu).

Digital Object Identifier 10.1109/TVT.2019.2897134

optimize outage probabilities for V2V communications based on bipartite matching.

The above methods have been proposed for unicast communications in vehicular networks. However, in some applications in V2V communications, there is no specific destination for the exchanged messages. In fact, the region of interest for each message includes the surrounding vehicles, which are the targeted destinations. Instead of using a unicast scheme to share the safety information, it is more appropriate to employ a broadcasting scheme. However, blindly broadcasting messages can cause the broadcast storm problem, resulting in package collision. In order to alleviate the problem, broadcast protocols based on statistical or topological information have been investigated in [9], [10]. In [10], several forwarding node selection algorithms have been proposed based on the distance to the nearest sender, of which the p -persistence provides the best performance and is going to be used as a part of our evaluations.

In the previous works, the QoS of V2V links only includes the reliability of SINR and the latency constraints for V2V links has not been considered thoroughly since it is hard to formulate the latency constraints directly into the optimization problems. To address these problems, we use deep reinforcement learning to handle the resource allocation in unicast and the broadcasting vehicular communications. Recently, deep learning has made great stride in speech recognition [17], image recognition [16], and wireless communications [18]. With deep learning techniques, reinforcement learning has shown impressive improvement in many applications, such as playing videos games [14] and playing Go game [19]. It has also been applied in resource allocation in various areas. In [20], a deep reinforcement learning framework has been developed for scheduling to satisfy different resource requirements. In [21], a deep reinforcement learning based approach has been proposed for resource allocation in the cloud radio access network (RAN) to save power and meet the user demands. In [22], the resource allocation problem in vehicular clouds is solved by reinforcement learning, where the resources can be dynamically provisioned to maximize long-term rewards for the network and avoid myopic decision making. A new deep reinforcement learning based approach has been proposed in [23] to deal with the highly complex joint resource optimization problem in the virtualized vehicular networks.

We exploit deep reinforcement learning to find the mapping between the local observations, including local CSI and interference levels, and the resource allocation and scheduling solution in this paper. In the unicast scenario, each V2V link is considered as an agent and the spectrum and transmission power are selected based on the observations of instantaneous channel conditions and exchanged information shared from the neighbors at each time slot. Apart from the unicast case, deep reinforcement learning based resource allocation framework can also be extended to the broadcast scenario. In this scenario, each vehicle is considered as an agent and the spectrum and messages are selected according to the learned policy. In order to address the problems encountered in the V2V communication system, the observation, the action, and the reward functions are designed delicately for the unicast and the broadcast scenarios, respectively. In general, the agents will automatically balance between

minimizing the interference of the V2V links to the vehicle-to-infrastructure (V2I) links and meeting the requirements for the stringent latency constraints imposed on the V2V link.

Part of this work has been published in [12] and [13] for unicast and broadcast, respectively. In this article, we provide a unified framework for both the unicast and the broadcast cases. In addition, more experimental results and discussions are provided to better understand the deep reinforcement learning based resource management scheme. Our main contributions are listed in the followings.

- A decentralized resource allocation mechanism based on deep reinforcement learning has been proposed for the first time to address the latency constraints on the V2V links.
- To solve the V2V resource allocation problems, the action space, the state space, and the reward function have been delicately designed for the unicast and the broadcast scenarios.
- Based on our simulation results, deep reinforcement learning based resource allocation can effectively learn to share the channel with V2I and other V2V links under both scenarios.

The rest of the paper is organized as follows. In Section II, the system model for unicast communications is introduced. In Section III, the reinforcement learning based resource allocation framework for unicast V2V communications is presented in detail. In Section IV, the framework is extended to the broadcast scenario, where the message receivers are all vehicles within a fixed area. In Section V, the simulation results are presented and the conclusions are drawn in Section VI.

II. SYSTEM MODEL FOR UNICAST COMMUNICATION

In this section, the system model and resource management problem of unicast communications are presented. As shown in Fig. 1, the vehicular network includes M cellular users (CUEs) denoted by $\mathcal{M} = \{1, 2, \dots, M\}$ and K pairs of V2V users (VUEs) denoted by $\mathcal{K} = \{1, 2, \dots, K\}$. The CUEs demand V2I links to support high capacity communication with the base station (BS) while the VUEs need V2V links to share information for traffic safety management. To improve the spectrum utilization efficiency, we assume that the orthogonally allocated uplink spectrum for V2I links is shared by the V2V links since the interference at the BS is more controllable and the uplink resources are less intensively used.

The SINR of the m^{th} CUE can be expressed as

$$\gamma^c[m] = \frac{P_m^c h_m}{\sigma^2 + \sum_{k \in \mathcal{K}} \rho_k[m] P_k^v \tilde{h}_k}, \quad (1)$$

where P_m^c and P_k^v denote the transmission powers of the m^{th} CUE and the k^{th} VUE, respectively, σ^2 is the noise power, h_m is the power gain of the channel corresponding to the m^{th} CUE, \tilde{h}_k is the interference power gain of the k^{th} VUE, and $\rho_k[m]$ is the spectrum allocation indicator with $\rho_k[m] = 1$ if the k^{th} VUE reuses the spectrum of the m^{th} CUE and $\rho_k[m] = 0$ otherwise. Hence the capacity of the m^{th} CUE is

$$C^c[m] = W \cdot \log(1 + \gamma^c[m]), \quad (2)$$

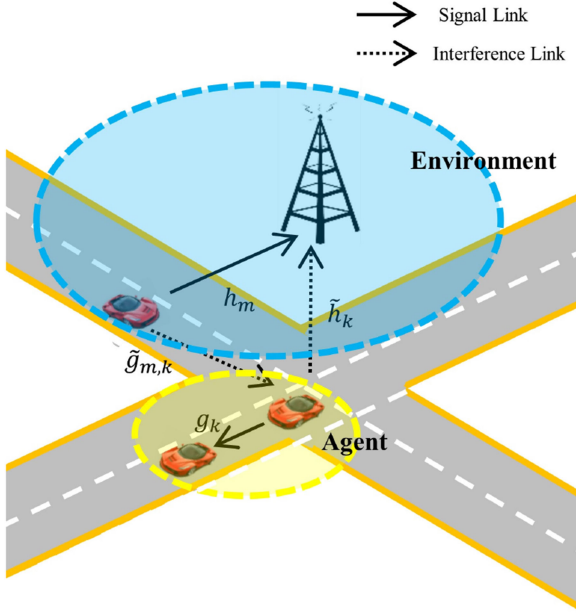


Fig. 1. An illustrative structure of unicast vehicular communication networks.

where W is the bandwidth.

Similarly, the SINR of the k^{th} VUE can be expressed as

$$\gamma^v[k] = \frac{P_k^v \cdot g_k}{\sigma^2 + G_c + G_d}, \quad (3)$$

where

$$G_c = \sum_{m \in \mathcal{M}} \rho_k[m] P_m^c \tilde{g}_{m,k}, \quad (4)$$

is the interference power of the V2I link sharing the same RB and

$$G_d = \sum_{m \in \mathcal{M}} \sum_{k' \in \mathcal{K}, k' \neq k} \rho_k[m] \rho_{k'}[m] P_{k'}^v \tilde{g}_{k',k}^v, \quad (5)$$

is the overall interference power from all V2V links sharing the same RB, g_k is the power gain of the k^{th} VUE, $\tilde{g}_{k,m}$ is the interference power gain of the m^{th} CUE, and $\tilde{g}_{k',k}^v$ is the interference power gain of the k'^{th} VUE. The capacity of the k^{th} VUE can be expressed as

$$C^v[k] = W \cdot \log(1 + \gamma^v[k]). \quad (6)$$

Due to the essential role of V2V communications in vehicle security protection, there are stringent latency and reliability requirements for V2V links while the data rate is not of great importance. The latency and reliability requirements for V2V links are converted into the outage probabilities [7], [8] in system design and considerations. With deep reinforcement learning, these constraints are formulated as the reward function directly, in which a negative reward is given when the constraints are violated. In contrast to V2V communications of safety information, the latency requirement on the conventional cellular traffic is less stringent and traditional resource allocation practices based on maximizing the throughput under certain fairness consideration remain appropriate. Therefore, the V2I sum rate

will remain a factor in the reward function for maximization in our method, as we can see in Section III.

Since the BS has no information on the V2V links, the resource allocation procedures of the V2I network should be independent of the resource management of the V2V links. Given resource allocation of V2I links, the objective of the proposed resource management scheme is to ensure satisfying the latency constraints for V2V links while minimizing the interference of the V2V links to the V2I links. In the decentralized resource management scenario, the V2V links will select the RB and transmission power based on the local observations.

The first type of observation that relates to resource allocation is the channel and the interference information. We assume the number of sub-channels is equal to the number of V2I links, M . The instantaneous channel information of the corresponding V2V link, $G_t[m]$, for $m \in \mathcal{M}$, is the power gain of sub-channel m . The channel information of the V2I link, $H_t[m]$, for $m \in \mathcal{M}$, characterizes the power gain of each sub-channel from the transmitter to the BS. The received interference signal strength in the previous time slot, $I_{t-1}[m]$, for $m \in \mathcal{M}$, represents the received interference strength in each sub-channel. Local observations also include information shared by the neighbors, such as the channel indices selected by the neighbors in the previous time slot, $N_{t-1}[m]$, for $m \in \mathcal{M}$, where each item represents the number of times the sub-channel has been used by the neighbors. In addition, information about the condition of the transmitted messages should also be involved, such as the remaining load for transmission, L_t , i.e., the proportion of bits remaining to transmit, and the time left before violating the latency constraint, U_t . This set of local information, including $H_t[m]$, $I_{t-1}[m]$, $N_{t-1}[m]$, for $m \in \mathcal{M}$, L_t , and U_t , will be used in the broadcast scenario as we will see in Section III. The observations are closely related to the optimal selection of power and the spectrum for transmission. Nevertheless, the relationship between the observations and the optimal resource allocation solution is often implicit and hard to establish analytically for network optimization. Deep reinforcement learning can find the relationship and accomplish optimization.

III. DEEP REINFORCEMENT LEARNING FOR UNICAST RESOURCE ALLOCATION

In this section, the deep reinforcement learning based resource management for unicast V2V communications is introduced. The formulations of key parts in the reinforcement learning are shown and the deep Q-network based proposed solution is presented in detail.

A. Reinforcement Learning

As shown in Fig. 2, the framework of reinforcement learning consists of agents and environment interacting with each other. In this scenario, each V2V link is considered as an agent and everything beyond the particular V2V link is regarded as the environment, which presents a collective rather than atomized condition related to the resource allocation. Since the behavior of other V2V links cannot be controlled in the decentralized setting, the action of each agent (individual V2V links) is thus

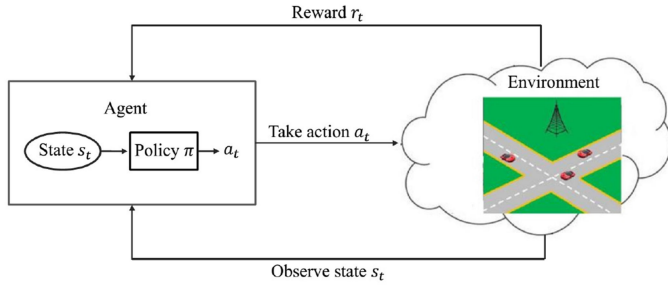


Fig. 2. Deep reinforcement learning for V2V communications.

based on the collective manifested environmental conditions such as spectrum, transmission power, etc.

As in Fig. 2, at each time t , the V2V link, as the agent, observes a state, \mathbf{s}_t , from the state space, \mathcal{S} , and accordingly takes an action, a_t , from the action space, \mathcal{A} , selecting sub-band and transmission power based on the policy, π . The decision policy, π , can be determined by the state-action function, also called Q-function, $Q(\mathbf{s}_t, a_t)$, which can be approximated by deep learning. Based on the actions taken by the agents, the environment transits to a new state, \mathbf{s}_{t+1} , and each agent receives a reward, r_t , from the environment. In our case, the reward is determined by the capacities of the V2I and V2V links and the latency constraints of the corresponding V2V link.

As we have discussed in Section II, the state of the environment observable to by each V2V link consists of several parts: the instantaneous channel information of the corresponding V2V link, $\mathbf{G}_t = (G_t[1], \dots, G_t[M])$, the previous interference power to the link, $\mathbf{I}_{t-1} = (I_{t-1}[1], \dots, I_{t-1}[M])$, the channel information of the V2I link, e.g., from the V2V transmitter to the BS, $\mathbf{H}_t = (H_t[1], \dots, H_t[M])$, the selected of sub-channel of neighbors in the previous time slot, $\mathbf{N}_{t-1} = (N_{t-1}[1], \dots, N_{t-1}[M])$, the remaining load of the VUE to transmit, L_t , and the remaining time to meet the latency constraints U_t . Even though the state consists heterogeneous data, it will be shown later that with deep neural networks, useful information can be extracted from the heterogeneous data for learning the optimal policy. In summary, the state can be expressed as

$$\mathbf{s}_t = \{\mathbf{I}_{t-1}, \mathbf{H}_t, \mathbf{N}_{t-1}, \mathbf{G}_t, U_t, L_t\}.$$

At each time, the agent takes an action $a_t \in \mathcal{A}$, which consists of selecting a sub-channel and a power level for transmission, based on the current state, $\mathbf{s}_t \in \mathcal{S}$, by following a policy π . The transmission power is discretized into three levels, thus the dimension of the action space is $3 \times N_{RB}$ when there are N_{RB} resource blocks in all.

The objective of V2V resource allocation is as follows. An agent (i.e. a V2V link) selects the frequency band and transmission power level that incur only small interference to all V2I links as well as other V2V links while preserving enough resources to meet the requirement of the latency constraints. Therefore, the reward function that guides the learning should be consistent with the objective. In our framework, the reward function consists of three parts, namely, the capacity of the V2I links, the capacity of the V2V links, and the latency condition. The sum capacities of the V2I and the V2V links are used to

measure the interference to the V2I and other V2V links, respectively. The latency condition is represented as a penalty. In particular, the reward function is expressed as,

$$r_t = \lambda_c \sum_{m \in \mathcal{M}} C^c[m] + \lambda_d \sum_{k \in \mathcal{K}} C^v[k] - \lambda_p (T_0 - U_t), \quad (7)$$

where T_0 is the maximum tolerable latency and λ_c , λ_d , and λ_p are weights of the three parts. The quantity $(T_0 - U_t)$ is the time used for transmission; the penalty increases as the time used for transmission grows. In order to obtain good performance in the long-term, both the immediate rewards and the future rewards should be taken into consideration. Therefore, the main objective of reinforcement learning is to find a policy to maximize the expected cumulative discounted rewards,

$$R_t = \mathbb{E} \left[\sum_{n=0}^{\infty} \beta^n r_{t+n} \right], \quad (8)$$

where $\beta \in [0, 1]$ is the discount factor.

The state transition and reward are stochastic and modelled as a Markov decision process (MDP), where the state transition probabilities and rewards depend only on the state of the environment and the action taken by the agent. The transition from \mathbf{s}_t to \mathbf{s}_{t+1} with reward r_t when action a_t is taken can be characterized by the conditional transition probability, $p(\mathbf{s}_{t+1}, r_t | \mathbf{s}_t, a_t)$. It should be noted that the agent can control its own actions and has no prior knowledge on the transition probability matrix $\mathbf{P} = \{p(\mathbf{s}_{t+1}, r_t | \mathbf{s}_t, a_t)\}$, which is only determined by the environment. In our problem, the transition on the channels, the interference, and the remaining messages to transmit are generated by the simulator of the wireless environment.

B. Deep Q-Learning

The agent takes actions based on a policy, π , which is a mapping from the state space, \mathcal{S} , to the action space, \mathcal{A} , expressed as $\pi : \mathbf{s}_t \in \mathcal{S} \rightarrow a_t \in \mathcal{A}$. As indicated before, the action space spans over two dimensions, the power level and the spectrum subband, and an action, $a_t \in \mathcal{A}$, corresponds to a selection of the power level and the spectrum for V2V links.

The Q-learning algorithms can be used to get an optimal policy to maximize the long-term expected accumulated discounted rewards, G_t [24]. The Q-value for a given state-action pair (\mathbf{s}_t, a_t) , $Q(\mathbf{s}_t, a_t)$, of policy π is defined as the expected accumulated discounted rewards when taking an action $a_t \in \mathcal{A}$ and following policy π thereafter. Hence the Q-value can be used to measure the *quality* of certain action in a given state. Once Q-values, $Q(\mathbf{s}_t, a_t)$, are given, an improved policy can be easily constructed by taking the action, given by

$$a_t = \arg \max_{a \in \mathcal{A}} Q(\mathbf{s}_t, a). \quad (9)$$

That is, the action to be taken is the one that maximizes the Q-value.

The optimal policy with Q-values Q^* can be found without any knowledge of the system dynamics based on the following

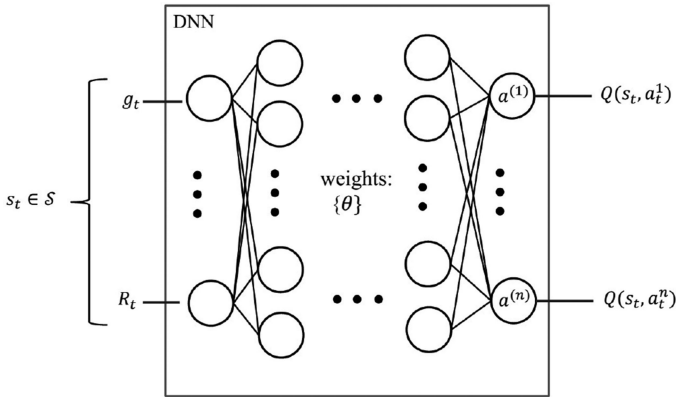


Fig. 3. Structure of deep Q-networks.

update equation,

$$Q_{new}(s_t, a_t) = Q_{old}(s_t, a_t) + \alpha[r_{t+1} + \beta \max_{s \in \mathcal{S}} Q_{old}(s, a_t) - Q_{old}(s_t, a_t)], \quad (10)$$

where α denotes the learning rate [25]. The second term on the right of the equation is the temporal difference error for updating the Q-value, which will be zero for the optimal Q-value Q^* .

It has been shown in [25] that in the MDP, the Q-values will converge with probability 1 to the optimal Q^* if each action in the action space is executed under each state for an infinite number of times on an infinite run and the learning rate α decays appropriately. The optimal policy, π^* , can be found once the optimal Q-value, Q^* , is determined.

In the resource allocation scenario, once the optimal policy is found through training, it can be employed to select spectrum band and transmission power level for V2V links to maximize overall capacity and ensure the latency constraints of V2V links.

The classic Q-learning method can be used to find the optimal policy when the state-action space is small, where a look-up table can be maintained for updating the Q-value of each item in the state-action space. However, the classic Q-learning cannot be applied if the state-action space becomes huge, just as in the resource management for the V2V communications. The reason is that a large number of states will be visited infrequently and corresponding Q-value will be updated rarely, leading to a much longer time for the Q-function to converge. To remedy this problem, deep Q-network improves the Q-learning by combining the deep neural networks (DNN) with Q-learning. As shown in Fig. 3, the Q-function is approximated by a DNN with weights $\{\theta\}$ as a Q-network [24]. Once $\{\theta\}$ is determined, Q-values, $Q(s_t, a_t)$, will be the outputs of the DNN in Fig. 3. DNN can address sophisticated mappings between the channel information and the desired output based on a large amount of training data, which will be used to determine Q-values.

The Q-network updates its weights, θ , at each iteration to minimize the following loss function derived from the same Q-network with old weights on a data set D ,

$$Loss(\theta) = \sum_{(s_t, a_t) \in D} (y - Q(s_t, a_t, \theta))^2, \quad (11)$$

where

$$y = r_t + \max_{a \in \mathcal{A}} Q_{old}(s_t, a, \theta), \quad (12)$$

where r_t is the corresponding reward.

C. Training and Testing Algorithms

Like most machine learning algorithms, there are two stages in our proposed method, i.e., the training and the test stage. Both the training and test data are generated from the interactions of an environment simulator and the agents. Each training sample used for optimizing the deep Q-network includes s_t , s_{t+1} , a_t , and r_t . The environment simulator includes VUEs and CUEs and their channels, where positions of the vehicles are generated randomly and the CSI of V2V and V2I links is generated according to the positions of the vehicles. With the selected spectrum and power of V2V links, the simulator can provide s_{t+1} and r_t to the agents. In the training stage, the deep Q-learning with experience replay is employed [24], where the training data is generated and stored in a storage named *memory*. As shown in Algorithm 1, in each iteration, a mini-batch data is sampled from the *memory* and is utilized to renew the weights of the deep Q-network. In this way, the temporal correlation of generated data can be suppressed. The policy used in each V2V link for selecting spectrum and power is random at the beginning and gradually improved with the updated Q-networks. As shown in Algorithm 2, in the test stage, the actions in V2V links are chosen with the maximum Q-value given by the trained Q-networks, based on which the evaluation is obtained.

As the action is selected independently based on the local information, the agent will have no knowledge of actions selected by other V2V links if the actions are updated simultaneously. As a consequence, the states observed by each V2V link cannot fully characterize the environment. In order to mitigate this issue, the agents are set to update their actions asynchronously, where only one or a small subset of V2V links will update their actions at each time slot. In this way, the environmental changes caused by actions from other agents will be observable. In fact, the co-ordinations are introduced by allowing the agents sharing their actions with their neighbors and taking turns to make decisions [28]. Co-ordinations can bring improvement by selecting the actions jointly instead of independently. In the distributed resource allocation problem, the V2V links may run into collisions if different agents make their decisions independently and happen to select the same RB. However, when the agent has information about the neighbors' actions in its observation, it can automatically find out that using the same RB tends to result in low rewards, therefore avoid this selection in order to get more rewards.

IV. RESOURCE ALLOCATION FOR BROADCAST SYSTEM

In this section, the resource allocation scheme based on deep reinforcement learning is extended to the broadcast V2V communication scenario. We first introduce the broadcast system model. Then the key elements in reinforcement learning

Algorithm 1: Training Stage Procedure of Unicast.

-
- 1: **procedure** TRAIN
 - 2: **Input:** Q-network structure, environment simulator
 - 3: **Output:** Q-network
 - 4: **start:**
Randomly initialize the policy π .
Initialize the model.
Start environment simulator and generate vehicles, VUEs, and CUEs.
 - 5: **loop:**
Iteratively select the V2V link in the system.
For each V2V link, select the spectrum and power for transmission based on policy π .
Environment simulator generates states and rewards based on the action of agents.
Collect and save the data item {state, reward, action, post-state} into memory.
Sample a mini-batch of data from the memory.
Train the deep Q-network using the mini-batch data.
Update the policy π : chose the action with maximum Q-value.
 - 6: **end loop**
 - 7: **return:** Return the deep Q-network
-

Algorithm 2: Test Stage Procedure of Unicast.

-
- 1: **procedure** TESTING
 - 2: **Input:** Q-network, environment simulator
 - 3: **Output:** Evaluation results
 - 4: **start:**
Load the Q-network model.
Start environment simulator and generate vehicles, VUEs, and CUEs.
 - 5: **loop:**
Iteratively select a V2V link in the system.
For each V2V link, select the action by choosing the action with the largest Q-value.
Update the environment simulator based on the actions selected.
Update the evaluation results, i.e., the average of V2I capacity and the probability of successful VUEs.
 - 6: **end loop**
 - 7: **return:** Evaluation results
-

framework are formulated for the broadcast system and algorithms to train the deep Q-networks are shown.

A. Broadcast System

Fig. 4 shows a broadcast V2V communication system, where the vehicle network consists $\mathcal{M}_B = \{1, 2, \dots, M_B\}$ CUEs demanding V2I links. At the same time, $\mathcal{K}_B = \{1, 2, \dots, K_B\}$ VUEs are broadcasting the messages, where each message has one transmitter and a group of receivers in the surrounding area. Similar to the unicast case, the uplink spectrum for the V2I

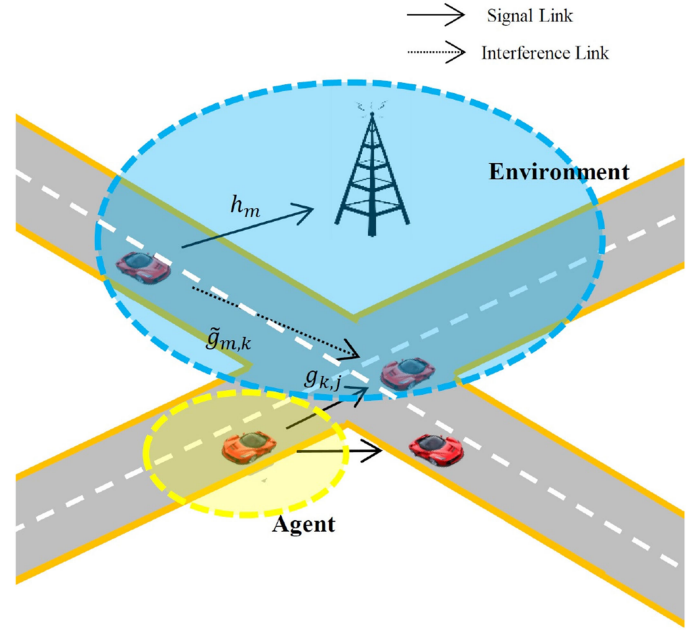


Fig. 4. An illustrative structure of broadcast vehicular communication networks.

links is reused by the V2V links as uplink resources are less intensively used and interference at the BS is more controllable.

In order to improve the reliability of broadcast, each vehicle will rebroadcast the messages that have been received. However, as mentioned in Section I, the broadcast storm problem occurs when the density of the vehicles is large and there are excessive redundant rebroadcast messages existing in the networks. To address this problem, vehicles need to select a proper subset of the received messages to rebroadcast so that more receivers can get the messages within the latency constraint while bringing little redundant broadcast to the networks.

The interference to the V2I links comes from the background noise and the signals from the VUEs that share the same sub-band. Thus the capacity of V2I link can be expressed as

$$\gamma^c[m] = \frac{P^c h_m}{\sigma^2 + \sum_{k \in \mathcal{K}_B} \rho_k[m] P^v \tilde{h}_k}, \quad (13)$$

where P^c and P^v are the transmission powers of the CUE and the VUE, respectively, σ^2 is the noise power, h_m is the power gain of the channel corresponding to the m^{th} CUE, \tilde{h}_k is the interference power gain of the k^{th} VUE, and $\rho_k[m]$ is the spectrum allocation indicator with $\rho_k[m] = 1$ if the k^{th} VUE reuses the spectrum of the m^{th} CUE and $\rho_k[m] = 0$ otherwise. Hence the capacity of the m^{th} CUE can be expressed as

$$C^c[m] = W \cdot \log(1 + \gamma^c[m]), \quad (14)$$

where W is the bandwidth.

Similarly, for the j^{th} receiver of the k^{th} VUE, the SINR is

$$\gamma_{k,j}^v = \frac{P^c g_{k,j}}{\sigma^2 + G_c + G_d}, \quad (15)$$

with

$$G_c = \sum_{m \in \mathcal{M}_B} \rho_{m,k} P^c \tilde{g}_{m,k}, \quad (16)$$

and

$$G_d = \sum_{m \in \mathcal{M}_B} \sum_{k' \in \mathcal{K}_B, k' \neq k} \rho_k[m] \rho_{k'}[m] P^v \tilde{g}_{k',k,j}^v, \quad (17)$$

where $g_{k,j}$ is the power gain of the j^{th} receiver of the k^{th} VUE, $\tilde{g}_{k,m}$ is the interference power gain of the m^{th} CUE, and $\tilde{g}_{k',k,j}^v$ is the interference power gain of the k'^{th} VUE. The capacity for the j^{th} receiver of the k^{th} VUE can be expressed as

$$C^v[k,j] = W \cdot \log(1 + \gamma^v[k,j]). \quad (18)$$

In the decentralized settings, each vehicle will determine which messages to broadcast and select which sub-channel to make better use of the spectrum. These decisions are based on some local observations and should be independent of the V2I links. Therefore, after resource allocation procedure of the V2I communications, the main goal of the proposed autonomous scheme is to ensure that the latency constraints for the VUEs can be met while the interference of the VUEs to the CUEs should be minimized. The spectrum selection and message scheduling should be managed according to the local observations.

In addition to the local information used in the unicast case, some information is useful and unique in the broadcast scenario, including the number of times that the message has been received by the vehicle, O_t , and the minimum distance to the vehicles that have broadcast the message, D_t . O_t can be obtained by maintaining a counter for each message received by vehicle, where the counter will increase one when the message has been received again. If the message has been received from different vehicles, D_t is the minimum distance to the message senders. In general, the probability of rebroadcasting a message decreases if the message has been heard many times by the vehicle or the vehicle is near to another vehicle that had broadcast the message before.

B. Reinforcement Learning for Broadcast

Under the broadcast scenario, each vehicle is considered as an agent in our system. At each time t , the vehicle observes a state, \mathbf{s}_t , and accordingly takes an action, a_t , selecting sub-band and messages based on the policy, π . Following the action, the state of the environment transfers to a new state \mathbf{s}_{t+1} and the agent receives a reward, r_t , determined by the capacities of the V2I and V2V links and the latency constraints of the corresponding V2V message.

Similar to the unicast case, the state observed by each vehicle for each received message consists of several parts: the instantaneous channel interference power to the link on each subband, $\mathbf{I}_{t-1} = (I_{t-1}[1], \dots, I_{t-1}[M])$, the channel information of the V2I link on each subband, e.g., from the V2V transmitter to the BS, $\mathbf{H}_t = (H_t[1], \dots, H_t[M])$, the selection of sub-channel of neighbors in the previous time slot, $\mathbf{N}_{t-1} = (N_{t-1}[1], \dots, N_{t-1}[M])$, and the remaining time to meet the latency constraints, U_t . Different from the unicast case, we have to

include the number of times that the message have been received by the vehicle, O_t and the minimum distance to the vehicles that have broadcast the message, D_t , in the state representation. In summary, the state can be expressed as

$$\mathbf{s}_t = \{\mathbf{I}_{t-1}, \mathbf{H}_t, \mathbf{N}_{t-1}, U_t, O_t, D_t\}.$$

At each time t , the agent takes an action at $a_t \in \mathcal{A}$, which includes determining the messages for broadcasting and the sub-channel for transmission. For each message, the dimension of action space is the $N_{RB} + 1$, where N_{RB} is the number of resource blocks. If the agent takes an action from the first N_{RB} actions, the message will be broadcast immediately in the corresponding sub-channel. Otherwise, the message will not be broadcast at this time if the agent takes the last action.

Similar to the unicast scenario, the objective of selecting channel and message for transmission is to minimize the interference to the V2I links with the latency constraints for VUEs guaranteed. In order to reach this objective, the frequency band and messages selected by each vehicle should have small interference to all V2I links as well as other VUEs. It also needs to meet the requirement of latency constraints. Therefore, similar to the unicast scenario, the reward function consists of three parts, the capacity of V2I links, the capacity of V2V links, and the latency condition. To suppress the redundant rebroadcasting, only the capacities of receivers that have not received the message are taken into consideration. Therefore, no capacity of V2V links is added, if the message to rebroadcast has already been received by all targeted receivers. The latency condition is represented as a penalty if the message has not been received by all the targeted receivers, which increases linearly as the remaining time U_t decreases. Therefore, the reward function can be expressed as,

$$r_t = \lambda_c \sum_{m \in \mathcal{M}} C^c[m] + \lambda_d \sum_{k \in \mathcal{K}, j \notin E\{k\}} C^v[k,j] - \lambda_p (T_0 - U_t), \quad (19)$$

where $E\{k\}$ represents the targeted receivers that have received the transmitted message.

In order to get the optimal policy, deep Q-network is trained to approximate the Q-function. The training and test algorithm for broadcasting are very similar to the unicast algorithms, as shown in Algorithm 3 and Algorithm 4.

V. EXPERIMENTS

In this section, we present simulation results to demonstrate the performance of the proposed method for unicast and broadcast vehicular communications.

A. Unicast

We consider a single cell system with the carrier frequency of 2 GHz. We follow the simulation setup for the Manhattan case detailed in 3GPP TR 36.885 [5], where there are 9 blocks in all and with both line-of-sight (LOS) and non-line-of-sight (NLOS) channels. The vehicles are dropped in the lane randomly according to the spatial Poisson process and each plans to communicate with the three nearby vehicles. Hence, the

Algorithm 3: Training Stage Procedure of Broadcast.

-
- 1: **procedure** TRAINING
 - 2: **Input:** Q-network structure, environment simulator
 - 3: **Output:** Q-network
 - 4: **start:** Random initialize the policy π .
Initialize the model.
Start environment simulator and generate vehicles, VUEs, and CUEs.
 - 5: **loop:** Iteratively select a vehicle in the system.
For each vehicle, select the messages and spectrum for transmission based on policy π
Environment simulator generates states and rewards based on the action of agents.
Collect and save the data item {state, reward, action, post-state} into memory.
Sample a mini-batch of data from the memory.
Train the deep Q-network using the mini-batch data.
Update the policy π : chose the action with maximum Q-value.
 - 6: **end loop**
 - 7: **return:** Return the deep Q-network
-

Algorithm 4: Test Stage Procedure of Broadcast.

-
- 1: **procedure** TESTING
 - 2: **Input:** Q-network, environment simulator
 - 3: **Output:** Evaluation results
 - 4: **start:** Load the Q-network model.
Start environment simulator and generate vehicles, VUEs, and CUEs.
 - 5: **loop:** Iteratively select a vehicle in the system.
For each vehicle, select the messages and spectrum by choosing the action with the largest Q-value.
Update the environment simulator based on the actions selected.
Update the evaluation results, i.e., the average of V2I capacity and the probability of successful VUEs.
 - 6: **end loop**
 - 7: **return:** Evaluation results
-

number of V2V links, K , is three times of the number of vehicles. Our deep Q-network is a five-layer fully connected neural network with three hidden layers. The numbers of neurons in the three hidden layers are 500, 250, and 120, respectively. The activation function of Relu is used, defined as

$$f_r(x) = \max(0, x). \quad (20)$$

The learning rate is 0.01 at the beginning and decreases exponentially. We also utilize ϵ -greedy policy to balance the exploration and exploitation [24] and adaptive moment estimation method (Adam) for training [26]. The detail parameters can be found in Table I.

TABLE I
SIMULATION PARAMETERS

Parameter	Value
Carrier frequency	2 GHz
Bandwidth per channel	1.5 MHz
BS antenna height	25 m
BS antenna gain	8 dBi
BS receiver noise figure	5 dB
Vehicle antenna height	1.5 m
Vehicle antenna gain	3 dBi
Vehicle receiver noise figure	9 dB
Vehicle speed	36 km/h
Neighbor distance threshold	150 m
Number of lanes	3 in each direction (12 in total)
Latency constraints for V2V links T_0	100 ms
V2V transmit power level list in unicast	[23, 10, 5] dBm
Noise power σ^2	-114 dBm
$[\lambda_c, \lambda_d, \lambda_p]$	[0.1, 0.9, 1]
V2V transmit power in broadcast	23 dBm
SINR threshold in broadcast	1 dB

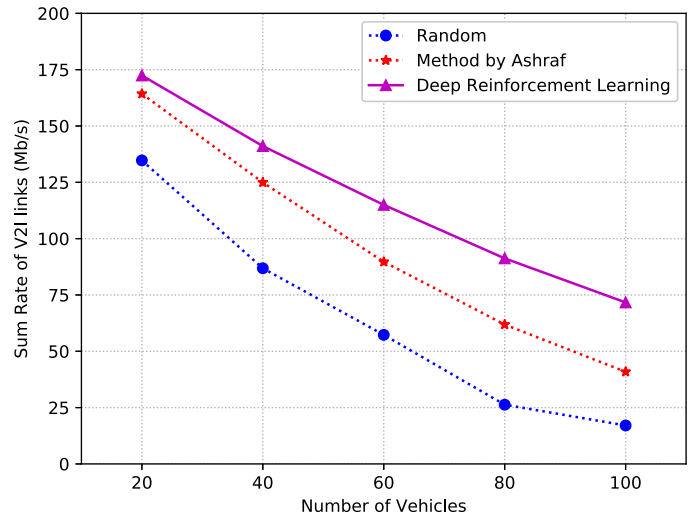


Fig. 5. Mean rate versus the number of vehicles.

The proposed method is compared with other two methods. The first is a random resource allocation method. At each time, the agent randomly chooses a sub-band for transmission. The other method is developed in [15], where vehicles are first grouped by the similarities and then the sub-bands are allocated and adjusted iteratively to the V2V links in each group.

1) *V2I Capacity:* Fig. 5 shows the summation of V2I rate versus the number of vehicles. From the figure, with the increase of the vehicles, the number of V2V link increases as a result, the interference to the V2I link grows, therefore the V2I capacity drops. the proposed method has much better performance to mitigate the interference of V2V links to the V2I communications.

2) *V2V Latency:* Fig. 6 shows the probability that the V2V links satisfy the latency constraint versus the number of vehicles. Similarly, with the increase vehicles, the V2V links in all increase, as a result it is more difficult to ensure every vehicle satisfy the latency constraints. From the figure, the proposed method has a much larger probability for the V2V links to satisfy the latency constraint since it can dynamically adjust the

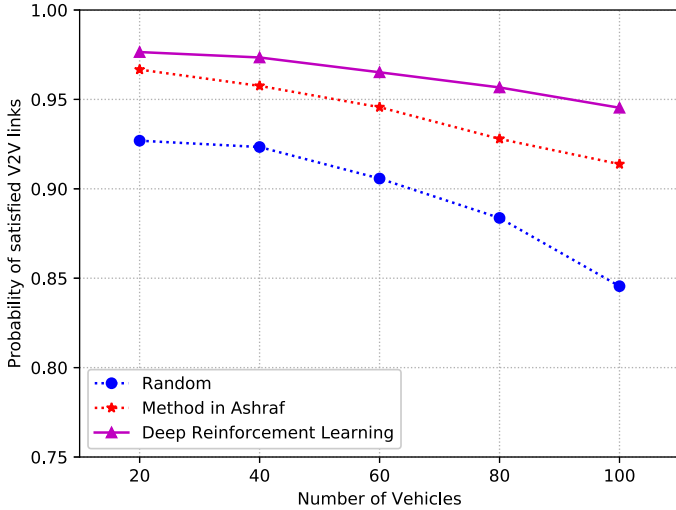


Fig. 6. Probability of satisfied V2V links versus the number vehicles.

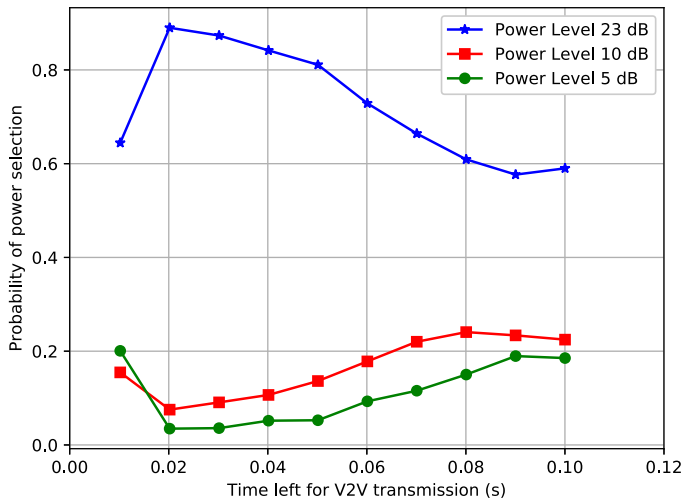


Fig. 7. The probability of power level selection with remaining time for transmission.

power and sub-band for transmission so that the links likely violating the latency constraint have more resources.

In order to find the reason for the superior performance of the learning based policy, we examine the power selection behaviors at different time during the transmission. Fig. 7 shows the probability for the agent to choose power levels with different time left for transmission. In general, the probability for the agent to choose the maximum power is low when there is abundant time for transmission, while the agent will select the maximum power with a high probability to ensure satisfying the V2V latency constraint when only a small amount of time left. However, when only 10 ms left, the probability for choosing the maximum power level suddenly drops to about 0.6 because the agent learns that even with the maximum power the latency constraints will be violated with high probability and switching to a lower power will get more reward by reducing interference to the V2I and other V2V links. Therefore, we can infer that the improvement of deep reinforcement learning based approach

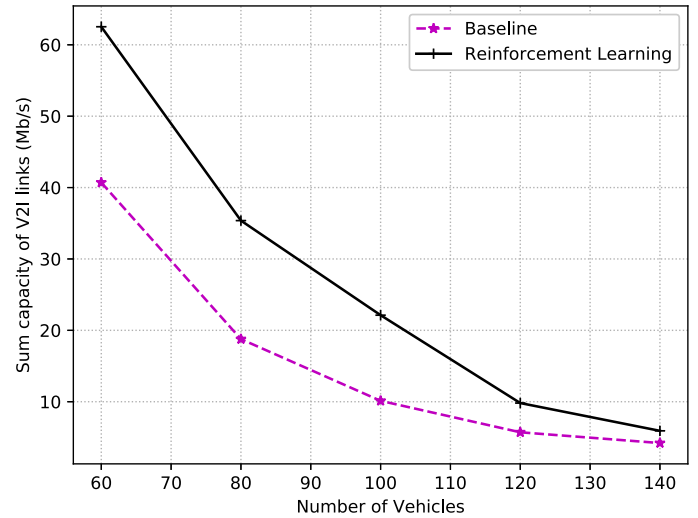


Fig. 8. Sum capacity versus the number of vehicles.

comes from learning the implicit relationship between the state and the reward function, which the benchmark methods fail to characterize.

B. Broadcast

The simulation environment is same as the one used in unicast, except that each vehicle communicates with all other vehicles within certain distance. The message of the k^{th} VUE is considered to be successfully received by the j^{th} receiver if the SINR of the receiver, $\gamma_{k,j}^v$, is above the SINR threshold. If all the targeted receivers of the message have successfully received the message, this V2V transmission is considered as successful.

The deep reinforcement learning based method jointly optimizes the scheduling and channel selection while historical works usually treat the two problems separately. In the channel selection part, the proposed method is compared with the channel selection method based on [15]. In the scheduling part, the proposed method is compared with the p -persistence protocol for broadcasting, where the probability of broadcasting is determined by the distance to the nearest sender [10].

1) *V2I Capacity*: Fig. 8 demonstrates the summation of V2I rate versus the number of vehicles. From the figure, the proposed method has better performance to mitigate the interference of the V2V links to the V2I communications.

2) *V2V Latency*: Fig. 9 shows the probability that the VUEs satisfy the latency constraint versus the number of vehicles. From the figure, the proposed method has a larger probability for VUEs to satisfy the latency constraint since it can effectively select the messages and sub-band for transmission.

C. Discussion

In summary, in both unicast and broadcast scenarios, deep reinforcement learning can find the relationship between local observations and the solution maximizing the accumulated reward function in a trial-and-error manner while this relationship is hard for the conventional approaches to characterize. For

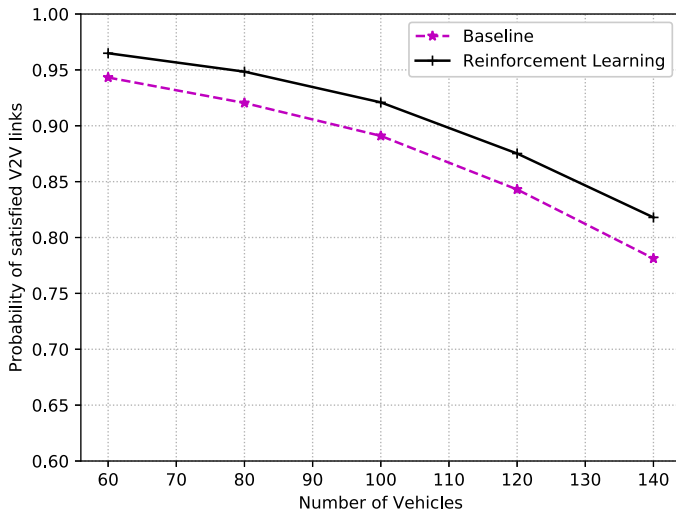


Fig. 9. Probability of satisfied VUEs versus the number vehicles.

instance, as shown in Fig. 7, the deep reinforcement learning can automatically figure out how to use different power levels according to the remaining time. But the relationship between the remaining time and the reward function cannot be formulated explicitly. Therefore, deep reinforcement learning allocates the resources wisely according to the local observations and has significant gains in the V2V successful rate and the V2I capacity comparing to the conventional methods.

A major concern on the deep learning based methods is the computation complexity. In our implementation, the time for each selection is 2.4×10^{-4} second, using GPU 1080 Ti. The speed is acceptable since the constraint on the vehicles is not very stringent. In addition, there are several prior works that reduce the computation complexity of the deep neural networks, such as binarizing the weights of the network [27]. How to exploit these new technologies to reduce the complexity of the proposed approach will be a future direction.

VI. CONCLUSION

In this paper, a decentralized resource allocation mechanism has been proposed for the V2V communications based on deep reinforcement learning. It can be applied to both unicast and broadcast scenarios. Since the proposed methods are decentralized, the global information is not required for each agent to make its decisions, the transmission overhead is small. From the simulation results, each agent can learn how to satisfy the V2V constraints while minimizing the interference to V2I communications.

ACKNOWLEDGMENT

The authors would like to thank Dr. M. Wu, Dr. S. C. Jha, Dr. K. Sivanesan, Dr. L. Lu, and Dr. J.-B. Kim from Intel Corporation for their insightful comments, which have substantially improved the quality of this paper.

REFERENCES

- [1] L. Liang, H. Peng, G. Y. Li, and X. Shen, "Vehicular communications: A physical layer perspective," *IEEE Trans. Veh. Technol.*, vol. 66, no. 12, pp. 10647–10659, Dec. 2017.
- [2] H. Peng, L. Liang, X. Shen, and G. Y. Li, "Vehicular communications: A network layer perspective," *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 1064–1078, Feb. 2019.
- [3] H. Ye, L. Liang, G. Y. Li, J. Kim, L. Lu, and M. Wu, "Machine learning for vehicular networks," *IEEE Veh. Technol. Mag.*, vol. 13, no. 2, pp. 94–101, Jun. 2018.
- [4] H. Seo, K. D. Lee, S. Yasukawa, Y. Peng, and P. Sartori, "LTE evolution for vehicle-to-everything services," *IEEE Commun. Mag.*, vol. 54, no. 6, pp. 22–28, Jun. 2016.
- [5] *3rd Generation Partnership Project: Technical Specification Group Radio Access Network: Study LTE-Based V2X Services: (Release 14)*, Standard 3GPP TR, Jun. 2016.
- [6] D. Feng, L. Lu, Y. Yuan-Wu, G. Li, S. Li, and G. Feng, "Device to-device communications in cellular networks," *IEEE Commun. Mag.*, vol. 52, no. 4, pp. 49–55, Apr. 2014.
- [7] W. Sun, E. G. Ström, F. Brännström, K. C. Sou, and Y. Sui, "Radio resource management for D2D-based V2V communication," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6636–6650, Aug. 2016.
- [8] L. Liang, G. Y. Li, and W. Xu, "Resource allocation for D2D-enabled vehicular communications," *IEEE Trans. Commun.*, vol. 65, no. 7, pp. 3186–3197, Jul. 2017.
- [9] S.-Y. Ni, Y.-C. Tseng, Y.-S. Chen, and J.-P. Sheu, "The broadcast storm problem in a mobile ad hoc network," in *Proc. 5th Annu. ACM/IEEE Int. Conf. Mobile Comput. Netw.*, Aug. 1999, pp. 151–162.
- [10] O. Tonguz, N. Wisitpongphan, J. Parikh, F. Bai, P. Mudalige, and V. Sadekar, "On the broadcast storm problem in ad hoc wireless networks," in *Proc. 3rd Int. Conf. Broadband Commun. Netw. Syst.*, Oct. 2006, pp. 1–11.
- [11] B. Bai, W. Chen, K. B. Letaief, and Z. Cao, "Low complexity outage optimal distributed channel allocation for vehicle-to-vehicle communications," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 1, pp. 161–172, Jan. 2011.
- [12] H. Ye, G. Y. Li, and B.-H. F. Juang, "Deep reinforcement learning for resource allocation in V2V communications," in *Proc. IEEE Int. Conf. Commun.*, May 2018, pp. 1–6.
- [13] H. Ye and G. Y. Li, "Deep reinforcement learning based distributed resource allocation for V2V broadcasting," in *Proc. IEEE Int. Wireless Commun. Mobile Comput. Conf.*, Jun. 2018, pp. 440–445.
- [14] V. Mnih *et al.*, "Playing Atari with deep reinforcement learning," 2013, arXiv:1312.5602.
- [15] M. I. Ashraf, M. Bennis, C. Perfecto, and W. Saad, "Dynamic proximity-aware resource allocation in vehicle-to-vehicle (V2V) communications," in *Proc. IEEE Globecom Workshops*, Dec. 2016, pp. 1–6.
- [16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Conf. Neural Inf. Process. Syst.*, Dec. 2012, pp. 1097–1105.
- [17] C. Weng, D. Yu, S. Watanabe, and B. H. F. Juang, "Recurrent deep neural networks for robust speech recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2014, pp. 5532–5536.
- [18] H. Ye, G. Y. Li, and B.-H. F. Juang, "Power of deep learning for channel estimation and signal detection in OFDM systems" *IEEE Wireless Commun. Lett.*, vol. 7, no. 1, pp. 114–117, Feb. 2018.
- [19] D. Silver *et al.*, "Mastering the game of go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484–489, Jan. 2016.
- [20] H. Mao, M. Alizadeh, I. Menache, and S. Kandula, "Resource management with deep reinforcement learning," in *Proc. 15th ACM Workshop Hot Topics Netw.* Nov. 2016, pp. 50–56.
- [21] Z. Xu, Y. Wang, J. Tang, J. Wang, and M. C. Gursoy, "A deep reinforcement learning based framework for power-efficient resource allocation in cloud RANs," in *Proc. IEEE Int. Conf. Commun.*, May 2017, pp. 1–6.
- [22] M. A. Salahuddin, A. Al-Fuqaha, and M. Guizani, "Reinforcement learning for resource provisioning in the vehicular cloud," *IEEE Wireless Commun.*, vol. 23, no. 4, pp. 128–135, Jun. 2016.
- [23] Y. He, N. Zhao, and H. Yin, "Integrated networking, caching and computing for connected vehicles: A deep reinforcement learning approach," *IEEE Trans. Veh. Technol.*, vol. 67, no. 1, pp. 44–55, Jan. 2018.
- [24] V. Mnih *et al.*, "Human-level control through deep reinforcement learning," *Nature* vol. 518, no. 7540, pp. 529–533, Feb. 2015.
- [25] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, Cambridge, MA, USA: MIT Press, 1998.
- [26] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, arXiv:1412.6980.

- [27] M. Rastegari, V. Ordonez, J. Redmon, and A. Farhadi, “Xnor-net: Imagenet classification using binary convolutional neural networks,” in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2016, pp. 525–542.
- [28] C. Zhang and V. R. Lesser, “Coordinated multi-agent reinforcement learning in networked distributed POMDPs,” in *Proc. AAAI Conf. Artif. Intell.*, Aug. 2011, pp. 764–770.



Hao Ye (S'18) received the B.E. degree in information engineering from Southeast University, Nanjing, China, in 2012 and the M. E. degree in electrical engineering from Tsinghua University, Beijing, China, in 2015. He is currently working toward the Ph.D. degree in electrical engineering with the Georgia Institute of Technology, Atlanta, GA, USA. His research interests include machine learning, signal detection, vehicular communications, and radio resource management.



Geoffrey Ye Li (S'93–M'95–SM'97–F'06) received the B.S.E. and M.S.E. degrees from the Nanjing Institute of Technology, Nanjing, China, in 1983 and 1986, respectively, and the Ph.D. degree from Auburn University, Auburn, AL, USA, in 1994. He was a Teaching Assistant and then a Lecturer with Southeast University, Nanjing, China, from 1986 to 1991, a Research and Teaching Assistant with Auburn University, Auburn, AL, USA, from 1991 to 1994, and a Postdoctoral Research Associate with the University of Maryland at College Park, College Park, MD, USA

from 1994 to 1996. He was with AT&T Labs—Research, Red Bank, NJ, USA, as a Senior and then a Principal Technical Staff Member from 1996 to 2000. Since 2000, he has been with the School of Electrical and Computer Engineering with the Georgia Institute of Technology as an Associate Professor and then a Full Professor. His general research interests include statistical signal processing and machine learning for wireless communications. In these areas, he has authored more than 250 journal papers in addition to more than 40 granted patents and many conference papers. His publications have been cited more than 34,000 times and he has been recognized as the World's Most Influential Scientific Mind, also known as a Highly-Cited Researcher, by Thomson Reuters almost every year. He was awarded IEEE Fellow for his contributions to signal processing for wireless communications in 2005. He won 2010 IEEE ComSoc Stephen O. Rice Prize Paper Award, 2013 IEEE VTS James Evans Avant Garde Award, 2014 IEEE VTS Jack Neubauer Memorial Award, 2017 IEEE ComSoc Award for Advances in Communication, and 2017 IEEE SPS Donald G. Fink Overview Paper Award. He was the recipient of the 2015 Distinguished Faculty Achievement Award from the School of Electrical and Computer Engineering, Georgia Tech. He has been involved in editorial activities for more than 20 technical journals for the IEEE, including founding Editor-in-Chief of the IEEE 5G Tech Focus. He has organized and chaired many international conferences, including Technical Program Vice-Chair of the IEEE International Conference on Communications 2003, Technical Program Co-Chair of the IEEE International Workshop on Signal Processing Advances in Wireless Communications 2011, General Chair of the IEEE Global Conference on Signal and Information Processing 2014, Technical Program Co-Chair of the IEEE Vehicular Technology Conference (VTC) 2016 (Spring), and General Co-Chair of the IEEE VTC 2019 (Fall).



Bing-Hwang Fred Juang received the Ph.D. degree from the University of California, Santa Barbara, CA, USA. He is the Motorola Foundation Chair Professor and a Georgia Research Alliance Eminent Scholar with the Georgia Institute of Technology, Atlanta, GA, USA. He is also enlisted as Honorary Chair Professor with several renowned universities. He had conducted research work with Speech Communications Research Laboratory and Signal Technology, Inc. in the late 1970s on a number of government-sponsored research projects and with Bell Labs during the 80s and 90s until he joined Georgia Tech in 2002. His notable accomplishments include development of vector quantization for voice applications, voice coders at extremely low bit rates (800 bps and 300 bps), robust vocoders for satellite communications, fundamental algorithms in signal modeling for automatic speech recognition, mixture hidden Markov models, discriminative methods in pattern recognition and machine learning, stereo—and multi-ponic teleconferencing, and a number of voice-enabled interactive communication services. He was the Director of Acoustics and Speech Research at Bell Labs (1996–2001). He has authored extensively, including the book-Fundamentals of Speech Recognition, coauthored with L.R. Rabiner, and holds nearly two dozen patents. He was the recipient of the Technical Achievement Award from the IEEE Signal Processing Society in 1998 for contributions to the field of speech processing and communications and the Third Millennium Medal from the IEEE in 2000. He was also the recipient of two Best Senior Paper awards, in 1993 and 1994 respectively, and a Best Paper Award in 1994, from the IEEE Signal Processing Society. He served as the Editor-in-Chief of the IEEE Transactions on Speech and Audio Processing from 1996 to 2002. He was elected the IEEE fellow (1991), a Bell Labs fellow (1999), a member of the US National Academy of Engineering (2004), and an Academician of the Academia Sinica (2006). He was named recipient of the IEEE Field Award in Audio, Speech and Acoustics, the J.L. Flanagan Medal, and a Charter Fellow of the National Academy of Inventors, in 2014.