An Oblivious Game-Theoretic Approach for Wireless Scheduling in V2V Communications

Xianfu Chen, Celimuge Wu, and Mehdi Bennis

***Abstract*—This paper addresses the problem of wireless resource scheduling in a vehicle-to-vehicle (V2V) communication network. The technical challenges lie in the fast changing network dynamics, namely, the channel quality and the data traffic variations. For a road segment covered by a road side unit (RSU), especially in a dense urban area, the vehicle density tends to be stable. The incoming service requests from the vehicle user equipment (VUE)-pairs compete with each other for the limited frequency resource in order to deliver data packets. Such competitions are regulated by the RSU via a sealed second-price auction at the beginning of scheduling slots. Each incumbent service request aims at maximizing the expected long-term payoff from bidding the frequency resource for packet transmissions. Markov perfect equilibrium (MPE) can be utilized to characterize the optimal competitive behaviors of the service requests. When the number of incumbent VUE-pairs becomes large, solving the MPE becomes infeasible. We adopt an oblivious equilibrium to approximate the MPE, which is theoretically proven to be error-bounded. The decision making process at each service request is hence transformed into a single-agent Markov decision process, for which we propose an on-line auction based learning scheme. Through simulation experiments, we show the potential performance gains from our proposed scheme, in terms of per-service request average utility.**

I. Introduction

The next generation vehicle-to-everything (V2X) technologies have been receiving significant attentions for enabling emerging mobile services and applications, such as traffic safety, congestion reporting and in-vehicle infotainment [1]. In particular, vehicle-to-vehicle (V2V) communication system, operating in an ad hoc model, is more flexible to render more attractive vehicular related applications. This type of vehicular applications have an ontological feature of requiring coordination between vehicles in close proximity. Without the infrastructure support, the high vehicle mobility and the dynamic network topology changes make the wireless resource scheduling in a V2V communication network extremely challenging. Therefore, it becomes very important to design an efficient scheme for a V2V system to improve the utilization of the limited wireless resources.

To address these technical obstacles, a number of recent works have focused on wireless resource allocation in V2V communications. In [2], Bai et al. proposed a low-complexity outage-optimal distributed channel allocation scheme for V2V communications based on maximum matching. In [3], Sun et al. investigated the radio resource management problem for device-to-device based V2V communication, for which a separate resource block and power allocation algorithm was proposed. Yao et al. proposed in [4] a loss differentiation rate adaptation scheme to meet the stringent delay and reliability requirements for V2V safety communications. In [5], Egea-Lopez et al. proposed a fair adaptive beaconing rate for inter-vehicular communications algorithm for the problem of beaconing rate control. Most of these works deal with the specific V2V communication scenarios limited by the vehicle density and fail to adequately address the network dynamics from the channel quality and the data traffic variations.

The framework of Markov decision process (MDP) has been applied to solve the resource scheduling problem in vehicular networks with time-varying nature. For example, Zheng et al. derived in [6] a stochastic learning algorithm for the delay-aware wireless resource scheduling in a software-defined vehicular network. However, the proposed linear decomposition technique ignores the coupling of the wireless resource scheduling among the participating agents. One of the main contributions in this paper is to address the problem of coupled wireless resource scheduling for a time-varying V2V communication network due to the limited frequency resource. Over the discrete scheduling slots, each vehicle user equipment (VUE)-pair competes with other VUE-pairs for the frequency resource to strike a balance between the transmit power consumption and the satisfaction of packet transmissions. The centralized frequency resource allocation for VUE-pairs arriving in the coverage of a road side unit (RSU) at each scheduling slot is regulated by sealed second-price auction [7][1]. We formulate the competitive resource scheduling problem as a stochastic game, for which the optimal solution can be characterized by the Markov perfect equilibrium (MPE). To handle the state-space explosion, we propose to approximate the MPE by an oblivious equilibrium (OE) [8], [9]. An online learning algorithm is further put forward to approach the OE solution. To the best of our knowledge, this is the first time to introduce the OE concept for wireless resource scheduling in V2V communications.

The rest of this paper is organized as follows. In following Section II, we introduce the considered system model and the

X. Chen is with the VTT Technical Research Centre of Finland Ltd, Oulu, Finland (email: xianfu.chen@vtt.fi). C. Wu is with the Graduate School of Informatics and Engineering, University of Electro-Communications, Tokyo, Japan (email: clmg@is.uec.ac.jp). M. Bennis is with the Centre for Wireless Communications, University of Oulu, Finland (email: bennis@ee.oulu.fi). This research was supported in part by AKA grants 310786 and 289611 and JSPS KAKENHI grant 16K00121.

[1] The dominant policy for a VUE-pair is to bid truthfully for the frequency resource.

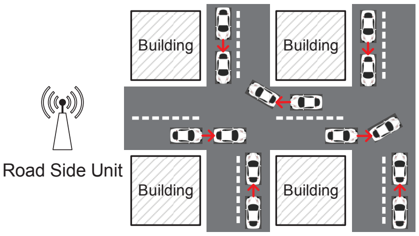

Fig. 1. Illustrative Manhattan grid vehicle-to-vehicle communication scenario.

related assumptions in this paper. In Section III, a MPE is defined to characterize the optimal solution for the problem of wireless resource scheduling in a Manhattan grid scenario, which is approximated by the OE. We propose an online learning algorithm to approach the OE solution and theoretically analyze the error between the MPE and the OE solutions in Section IV. In Section V, we evaluate the performance of our proposed algorithm through simulations. Finally, we draw the conclusions in Section VI.

II. System Model

As illustrated in Fig. 1, this work considers a Manhattan grid V2V communication scenario, in which the VUEs compete for a common frequency resource. The whole system operates over discrete scheduling slots, each of which is of equal duration δ and is indexed by a positive integer $n \in \mathbb{N}_+$. When the service requests arrive within the coverage of the RSU, the frequency resource is allocated to the VUE transmitters. A service request is deemed as a request from a VUE-pair to the RSU for resource allocations throughout the service period. The centralized frequency resource allocation at the beginning of a scheduling slot is regulated by the RSU using a sealed second-price auction. The winner of the resource auction acquires the frequency during the slot for emptying the queued data packets. We assume that the service request arrivals from VUE-pairs constitute a sequence of independent and random variables. It has been established that for a well defined road segment, the vehicle density approaches to be steady [10]. The data queue of a service request may be terminated with a probability of $\gamma \in (0,1)$ after participating in the resource auction. In the following, we shall interchangeably use a service request and a VUE-pair.

Over the wireless scheduling horizon, each service request is assigned a unique positive integer index. At each scheduling slot n, the set of incumbent service requests is denoted by $\mathcal{K}(n)$ with $\lim_{n\to\infty} \mathsf{E}[|\mathcal{K}(n)|] = \kappa$ [10]. Let $q_k(n)$ and $a_k(n)$ be the queue length and the random new packet arrivals at slot n, respectively. The queue evolution for service request k can be expressed as

$$\begin{aligned} & q_k(n+1) = \\ & \begin{cases} \text{Null}, \ \text{if the queue is terminated at slot } n; \\ \min\left\{q_k(n) - \theta_k(n) d_k(n) + a_k(n), q^{(max)}\right\}, \\ \qquad \text{otherwise}, \end{cases} \end{aligned} \tag{1}$$

where $q^{(max)}$ is the maximum queue length that restricts $q_k(n) \in \mathcal{Q} = \{0, \cdots, q^{(max)}\}$, $d_k(n)$ is the scheduled number of packet departures during slot n, and $\theta_k(n)$ is an auction winner indicator that equals 1 if the service request wins the frequency resource and 0 otherwise. Let $b_k(n)$ be the bid submitted by each service request k at a scheduling slot n, then the winner is determined according to

$$\max_{\left\{(\theta_k(n)\in\{0,1\}:k\in\mathcal{K}(n)):\sum\limits_{k\in\mathcal{K}(n)}\theta_k(n)\le 1\right\}} \sum_{k\in\mathcal{K}(n)} \theta_k(n) b_k(n), \tag{2}$$

and the incurred payment for service request k at scheduling slot n is calculated as

$$\tau_k(n) = \theta_k(n) \max_{j\in\mathcal{K}(n)\setminus\{k\}} b_j(n). \tag{3}$$

To simplify the wireless communication model, perfect channel state information is assumed. Let $g_k(n)$ be the channel gain for a VUE-pair k over the frequency resource during a scheduling slot n, which is picked from a finite set $\mathcal{G}$. The transmit power consumption for delivering $\theta_k(n)d_k(n)$ error-free packets can be computed by

$$p_k(n) = \frac{w\sigma^2}{g_k(n)} \left(2^{\frac{\mu\theta_k(n)d_k(n)}{w\delta}} - 1\right), \tag{4}$$

where w is the bandwidth of the frequency resource, σ^2 is the power spectral density of background noise, and μ is the fixed size of a data packet.

III. Problem Formulation

At each scheduling slot n, the local state of a service request $k \in \mathcal{K}(n)$ is represented by $\mathbf{x}_k(n) = (g_k(n), q_k(n)) \in \mathcal{X} \triangleq \mathcal{G} \times \mathcal{Q}$, and $\mathbf{x}_{-k}(n) = (\mathbf{x}_j(n) : j \in \mathcal{K}(n)\setminus\{k\}) \in \mathcal{X}^{|\mathcal{K}(n)|-1}$ is defined to be the state of the competitors. A payoff function is associated with each service request k for the wireless resource scheduling at a slot n, which is chosen to be

$$\begin{aligned} & f_k(\mathbf{x}_k(n), \mathbf{x}_{-k}(n), \theta_k(n), d_k(n)) = \\ & u_k(\mathbf{x}_k(n), \mathbf{x}_{-k}(n), \theta_k(n), d_k(n)) - \tau_k(n), \end{aligned} \tag{5}$$

where the utility function

$$\begin{aligned} & u_k(\mathbf{x}_k(n), \mathbf{x}_{-k}(n), \theta_k(n), d_k(n)) = \\ & u_k^{(1)}(q_k(n)) + \alpha_k u_k^{(2)}(p_k(n)) + u_k^{(3)}(o_k(n)). \end{aligned} \tag{6}$$

In (6), α_k is a positive weight, and $u_k^{(1)}(\cdot)$, $u_k^{(2)}(\cdot)$ and $u_k^{(3)}(\cdot)$ are the positive monotonically decreasing functions measuring the satisfactions of the queue length $q_k(n)$, the transmit power consumption $p_k(n)$ and the packet overflows $o_k(n)$ which is given as follows

$$\begin{aligned} & o_k(n) = \\ & \max\left\{q_k(n) - \theta_k(n) d_k(n) + a_k(n) - q^{(max)}, 0\right\}. \end{aligned} \tag{7}$$

For the considered dense urban area, there are an asymptotically large number of service requests, which are assumed to play a symmetric control policy $\mathbf{\Theta} = (\Theta^{(f)}, \Theta^{(p)})$ consisting of the frequency resource auction policy $\Theta^{(f)}$ and the packet scheduling policy $\Theta^{(p)}$. Note that $\Theta^{(p)}$ is local network state dependent. With $\boldsymbol{\Theta}$, after observing the global system state $(\mathbf{x}_k(n), \mathbf{x}_{-k}(n))$ at the beginning of each scheduling slot n, each VUE-pair $k \in \mathcal{K}(n)$ announces a bid $b_k(n)$ to the RSU for frequency allocation and then proceeds to schedule the packet transmissions based on the auction results, i.e., $\boldsymbol{\Theta}(\mathbf{x}_k(n), \mathbf{x}_{-k}(n)) = (b_k(n), d_k(n))$. Accordingly, we define the state value function $V_k(\mathbf{x}_k, \mathbf{x}_{-k}|\boldsymbol{\Theta}', \boldsymbol{\Theta})$ for service request k in state $\mathbf{x}_k(n) = \mathbf{x}_k$ when its competitors are in state $\mathbf{x}_{-k}(n) = \mathbf{x}_{-k}$ at current slot n, given that the competitors follow a common policy $\boldsymbol{\Theta}$ while the service request k follows $\boldsymbol{\Theta}'$. In specific,

$$V_k(\mathbf{x}_k, \mathbf{x}_{-k}|\boldsymbol{\Theta}', \boldsymbol{\Theta}) = \mathsf{E}_{(\boldsymbol{\Theta}', \boldsymbol{\Theta})}\left[\sum_{t=n}^{N_k} f_k(\mathbf{x}_k(t), \mathbf{x}_{-k}(t), \theta_k(t), d_k(t))|\mathbf{x}_k, \mathbf{x}_{-k}\right], \tag{8}$$

where N_k is the length of the service period after which the data queue is ineffective, and is hence a geometric random variable. The expected long-term payoff expressed by (8) can then be equivalently transformed into

$$V_k(\mathbf{x}_k, \mathbf{x}_{-k}|\boldsymbol{\Theta}', \boldsymbol{\Theta}) = \mathsf{E}_{(\boldsymbol{\Theta}', \boldsymbol{\Theta})}\left[\sum_{t=n}^{\infty} (\gamma)^t f_k(\mathbf{x}_k(t), \mathbf{x}_{-k}(t), \theta_k(t), d_k(t))|\mathbf{x}_k, \mathbf{x}_{-k}\right]. \tag{9}$$

Each VUE-pair k aims to find an optimal control policy that maximizes $V_k(\mathbf{x}_k, \mathbf{x}_{-k}|\boldsymbol{\Theta}', \boldsymbol{\Theta})$, $\forall(\mathbf{x}_k, \mathbf{x}_{-k})$.

Definition 1: A solution to the competitive wireless scheduling stochastic game can be characterized by the MPE $\boldsymbol{\Theta}^*$, which satisfies the condition: $\forall k \in \mathcal{K}(n)$,

$$V_k(\mathbf{x}_k, \mathbf{x}_{-k}) = \max_{\boldsymbol{\Theta}} V_k(\mathbf{x}_k, \mathbf{x}_{-k}|\boldsymbol{\Theta}, \boldsymbol{\Theta}^*), \forall(\mathbf{x}_k, \mathbf{x}_{-k}), \tag{10}$$

where $V_k(\mathbf{x}_k, \mathbf{x}_{-k}) = V_k(\mathbf{x}_k, \mathbf{x}_{-k}|\boldsymbol{\Theta}^*, \boldsymbol{\Theta}^*)$.

In general, dynamic programming [11] can be adopted to solve the MPE control policy for the VUE-pairs. However, the challenges for the method lie in the computational complexity and memory, which grow exponentially as the number of service request arrivals increases. The difficulty motivates an alternative solution, namely, the OE. The basic idea is that when there are a large number of service requests, the impacts from competitions among the VUE-pairs on the wireless scheduling can be averaged out such that the state of the competitors remains nearly unchanged across the scheduling horizon. With a OE policy, each VUE is thus able to make the near-optimal frequency auction and packet scheduling decisions based only on the local state information. Let $\hat{\boldsymbol{\Theta}} = (\hat{\Theta}^{(f)}, \Theta^{(p)})$ denote a symmetric oblivious control policy. The corresponding oblivious state value function for a service request $k \in \mathcal{K}(n)$ can be defined as

$$\hat{V}_k\left(\mathbf{x}_k|\hat{\boldsymbol{\Theta}}', \hat{\boldsymbol{\Theta}}\right) = \mathsf{E}_{(\hat{\boldsymbol{\Theta}}', \hat{\boldsymbol{\Theta}})}\left[\sum_{t=n}^{\infty} (\gamma)^t f_k(\mathbf{x}_k(t), \theta_k(t), d_k(t))\,|\mathbf{x}_k\right], \tag{11}$$

where the per-slot payoffs achieved from deploying $\hat{\boldsymbol{\Theta}}'$ when the competitors apply $\hat{\boldsymbol{\Theta}}$ then depend on the local states.

Definition 2: A OE consists of a symmetric control policy $\hat{\boldsymbol{\Theta}}^*$ such that $\forall k \in \mathcal{K}(n)$,

$$\hat{V}_k(\mathbf{x}_k) = \max_{\hat{\boldsymbol{\Theta}}} \hat{V}_k\left(\mathbf{x}_k|\hat{\boldsymbol{\Theta}}, \hat{\boldsymbol{\Theta}}^*\right), \forall \mathbf{x}_k, \tag{12}$$

where $\hat{V}_k(\mathbf{x}_k) = \hat{V}_k(\mathbf{x}_k|\hat{\boldsymbol{\Theta}}^*, \hat{\boldsymbol{\Theta}}^*)$.

Corollary: The existence of a OE for the non-cooperative wireless scheduling problem considered in this paper is straightforward from the discussions in [8].

IV. Solving the OE Policy

In this section, we shall first propose an algorithm to solve the OE control policy and then quantify the error between the MPE and the OE solutions.

A. Proposed Algorithm

The problem in (12) is a typical infinite-horizon discounted MDP, for which we obtain the Bellman's optimality equation as (13), $\forall k \in \mathcal{K}(n)$, where $\hat{\Theta}^{(f),*}$ is the optimal oblivious frequency resource auction policy, $\hat{\boldsymbol{\Theta}}^{(f),*}(\mathbf{x}_{-k})$ denotes the auction bids from the competitors, and $\mathbf{x}_k'$ is the subsequent local state. Theorem 1 below gives the optimal auction bid b_k submitted by an incumbent service request k to the RSU at a current scheduling slot.

Theorem 1: With $\hat{\boldsymbol{\Theta}}^*$, the optimal auction bid from a VUE-pair $k \in \mathcal{K}(n)$ at a current scheduling slot n is given by (14), where $\mathbf{x}_k'$ and $\mathbf{x}_k''$ are the local states at the following scheduling slot under the different frequency resource auction and packet scheduling decisions.

Proof: Since the OE policy $\hat{\boldsymbol{\Theta}}^*$ is composed of the optimal oblivious frequency resource auction policy $\hat{\Theta}^{(f),*}$ and the optimal packet scheduling policy $\Theta^{(p),*}$, the Bellman's equation in (13) for a VUE-pair $k \in \mathcal{K}(n)$ can be rewritten as (15), where $\mathbf{x}_k''$ is the consequent local state if VUE-pair k fails the frequency resource auction. From the winner determination (2) and the payment calculation (3) rules, the optimal auction policy for VUE-pair k is to bid truthfully across the scheduling slots with the bid of the form in (14). □

However, one challenge of calculating an optimal bid at a current scheduling slot remains: the packet arrival distribution and the statistics of channel gain state variations may not be known a priori in practice. We thus define a post-decision state of each VUE-pair based on the observation that the packet arrivals are independent of the frequency resource auction and the packet scheduling decision makings. At current slot n, the post-decision state of a VUE-pair $k \in \mathcal{K}(n)$ is defined as $\tilde{\mathbf{x}}_k = (\tilde{g}_k, \tilde{q}_k) \in \mathcal{X}$, where $\tilde{g}_k = g_k$ and $\tilde{q}_k = q_k - \theta_k(\hat{\Theta}^{(f),*}(\mathbf{x}_k), \hat{\boldsymbol{\Theta}}^{(f),*}(\mathbf{x}_{-k}))\Theta^{(p),*}(\mathbf{x}_k)$. The introduction of a post-decision state enables factoring the utility function in (6) into two parts, which correspond to $u_k^{(1)}(\cdot) + \alpha_k u_k^{(2)}(\cdot)$ and

$$\hat{V}_k(\mathbf{x}_k) = \max_{\hat{\mathbf{\Theta}}(\mathbf{x}_k)} \tag{13}$$

$$\left\{ \gamma f_k\Big(\mathbf{x}_k, \theta_k\Big(\hat{\Theta}^{(f)}(\mathbf{x}_k), \hat{\mathbf{\Theta}}^{(f),*}(\mathbf{x}_{-k})\Big), \Theta^{(p)}(\mathbf{x}_k)\Big) + \gamma \sum_{\mathbf{x}'_k \in \mathcal{X}} \mathsf{Pr}\Big\{\mathbf{x}'_k | \mathbf{x}_k, \theta_k\Big(\hat{\Theta}^{(f)}(\mathbf{x}_k), \hat{\mathbf{\Theta}}^{(f),*}(\mathbf{x}_{-k})\Big), \Theta^{(p)}(\mathbf{x}_k)\Big\} \hat{V}_k(\mathbf{x}'_k) \right\}$$

$$\begin{aligned} b_k = {} & u_k\Big(\mathbf{x}_k, 1, \Theta^{(p),*}(\mathbf{x}_k)\Big) + \sum_{\mathbf{x}'_k \in \mathcal{X}} \mathsf{Pr}\Big\{\mathbf{x}'_k | \mathbf{x}_k, 1, \Theta^{(p),*}(\mathbf{x}_k)\Big\} \hat{V}_k(\mathbf{x}'_k) \\ & - \left(u_k\Big(\mathbf{x}_k, 0, \Theta^{(p),*}(\mathbf{x}_k)\Big) + \sum_{\mathbf{x}''_k \in \mathcal{X}} \mathsf{Pr}\Big\{\mathbf{x}''_k | \mathbf{x}_k, 0, \Theta^{(p),*}(\mathbf{x}_k)\Big\} \hat{V}_k(\mathbf{x}''_k) \right) \end{aligned} \tag{14}$$

$$\begin{aligned} \hat{\Theta}^{(f),*}(\mathbf{x}_k) = \underset{\hat{\Theta}^{(f)}(\mathbf{x}_k)}{\arg\max} \Bigg\{ & \Bigg(\gamma f_k\Big(\mathbf{x}_k, \theta_k\Big(\hat{\Theta}^{(f)}(\mathbf{x}_k), \hat{\mathbf{\Theta}}^{(f),*}(\mathbf{x}_{-k})\Big), \Theta^{(p),*}(\mathbf{x}_k)\Big) + \\ & \gamma \sum_{\mathbf{x}'_k \in \mathcal{X}} \mathsf{Pr}\Big\{\mathbf{x}'_k | \mathbf{x}_k, \theta_k\Big(\hat{\Theta}^{(f)}(\mathbf{x}_k), \hat{\mathbf{\Theta}}^{(f),*}(\mathbf{x}_{-k})\Big), \Theta^{(p),*}(\mathbf{x}_k)\Big\} \hat{V}_k(\mathbf{x}'_k) \Bigg) - \\ & \left(\gamma f_k\big(\mathbf{x}_k, 0, \Theta^{(p),*}(\mathbf{x}_k)\big) + \gamma \sum_{\mathbf{x}''_k \in \mathcal{X}} \mathsf{Pr}\Big\{\mathbf{x}''_k | \mathbf{x}_k, 0, \Theta^{(p),*}(\mathbf{x}_k)\Big\} \hat{V}_k(\mathbf{x}''_k) \right) \Bigg\} \end{aligned} \tag{15}$$

$u_k^{(3)}(\cdot)$, respectively. The probability of state transition from $\mathbf{x}_k$ to $\mathbf{x}'_k$ can be then expressed as

$$\begin{aligned} & \mathsf{Pr}\Big\{\mathbf{x}'_k | \mathbf{x}_k, \theta_k\Big(\hat{\Theta}^{(f),*}(\mathbf{x}_k), \hat{\mathbf{\Theta}}^{(f),*}(\mathbf{x}_{-k})\Big), \Theta^{(p),*}(\mathbf{x}_k)\Big\} \\ & = \mathsf{Pr}\{\mathbf{x}'_k | \tilde{\mathbf{x}}_k\} \\ & \times \mathsf{Pr}\Big\{\tilde{\mathbf{x}}_k | \mathbf{x}_k, \theta_k\Big(\hat{\Theta}^{(f),*}(\mathbf{x}_k), \hat{\mathbf{\Theta}}^{(f),*}(\mathbf{x}_{-k})\Big), \Theta^{(p),*}(\mathbf{x}_k)\Big\} \\ & = \mathsf{Pr}\{g'_k\} \, \mathsf{Pr}\{a_k\}. \end{aligned} \tag{16}$$

Let the oblivious post-decision state value function $\tilde{V}_k(\tilde{\mathbf{x}}_k)$ be

$$\tilde{V}_k(\tilde{\mathbf{x}}_k) = \gamma u_k^{(3)}(o_k) + \gamma \sum_{\mathbf{x}'_k \in \mathcal{X}} \mathsf{Pr}\{\mathbf{x}'_k | \tilde{\mathbf{x}}_k\} \, \hat{V}_k(\mathbf{x}'_k). \tag{17}$$

By substituting (17) into (14), we have

$$\begin{aligned} b_k = {} & u_k^{(1)}(q_k) + \alpha_k u_k^{(2)}\Big(p_k\left(1, \Theta^{(p),*}(\mathbf{x}_k)\right)\Big) + \frac{1}{\gamma}\tilde{V}_k(\tilde{\mathbf{x}}_k) \\ & - \left(u_k^{(1)}(q_k) + \alpha_k u_k^{(2)}(0) + \frac{1}{\gamma}\tilde{V}_k(\mathbf{x}_k)\right). \end{aligned} \tag{18}$$

The number of packet arrivals in the end of a scheduling slot is unavailable beforehand and so is the number of packet overflows at the slot. In this case, instead of directly computing the oblivious post-decision state value function as in (17), we propose an online learning algorithm to approach the oblivious post-decision state value function by exploring the conventional reinforcement learning techniques. Based on the observations of the local state $\mathbf{x}_k(n)$, the frequency resource allocation result $\theta_k(n)$ by the RSU, the number of packet departures $\theta_k(n)d_k(n)$, the number of new packet arrivals $a_k(n)$, the number of packet overflows $\max\{q_k(n) - \theta_k(n)d_k(n) + a_k(n) - q^{(max)}, 0\}$, the payment $\tau_k(n)$ at current slot n, and the resulting local state $\mathbf{x}_k(n+1)$ at the next scheduling slot $n+1$, each VUE-pair $k \in \mathcal{K}(n)$ updates the oblivious post-decision state value function on the fly according to

$$\begin{aligned} \tilde{V}_k^{n+1}(\tilde{\mathbf{x}}_k(n)) = {} & (1 - \zeta(n))\tilde{V}_k^n(\tilde{\mathbf{x}}_k(n)) \\ & + \zeta(n)\Big(\gamma u_k^{(3)}(o_k(n)) + \gamma \hat{V}_k^n(\mathbf{x}_k(n+1))\Big). \end{aligned} \tag{19}$$

Herein, $\zeta(n) \in [0, 1)$ is the learning rate satisfying $\sum_{n=1}^{\infty} \zeta(n) = \infty$ and $\sum_{n=1}^{\infty} (\zeta(n))^2 < \infty$ to ensure the convergence property, the packet departures are determined by (20), and the subsequent local state $\mathbf{x}_k(n+1)$ can be evaluated as in (21), where the average payment $\bar{\tau}_k^n(\mathbf{x}_k)$, $\forall \mathbf{x}_k \in \mathcal{X}$, across the scheduling horizon is updated according to

$$\bar{\tau}_k^{n+1}(\mathbf{x}_k) = \begin{cases} \frac{Y_k^n(\mathbf{x}_k)\bar{\tau}_k^n(\mathbf{x}_k) + \tau_k(n)}{Y_k^n(\mathbf{x}_k) + 1}, & \text{if } \mathbf{x}_k = \mathbf{x}_k(n); \\ \bar{\tau}_k^n(\mathbf{x}_k), & \text{otherwise,} \end{cases} \tag{22}$$

with $Y_k^n(\mathbf{x}_k)$ being the number of times the service request in state $\mathbf{x}_k$ until scheduling slot $n+1$.

B. Solution Error Analysis

In this subsection, we aim to quantitatively analyze the error between a MPE and a OE solutions to the stochastic wireless resource scheduling game. We first define an asymptotically Markovian property for a OE policy as follows.

$$d_k(n) = \arg\max_{d} \left\{ \gamma \left(u_k^{(1)}(q_k(n)) + \alpha_k u_k^{(2)} \left(p_k(1, d)\right) - \bar{\tau}_k^n(\mathbf{x}_k(n)) \right) + \tilde{V}_k^n(\tilde{\mathbf{x}}_k(n)) \right\} \tag{20}$$

$$\begin{aligned} \hat{V}_k^n(\mathbf{x}_k(n+1)) = & \max_{\hat{\mathbf{\Theta}}(\mathbf{x}_k(n+1))} \Big\{ \gamma \Big(u_k^{(1)}(q_k(n+1)) \\ & + \alpha_k u_k^{(2)} \Big(p_k \left(\theta_k \left(\hat{\Theta}^{(f)}(\mathbf{x}_k(n+1)), \hat{\mathbf{\Theta}}^{(f),*}(\mathbf{x}_{-k}(n+1)) \right), \Theta^{(p)}(\mathbf{x}_k(n+1)) \right) \Big) - \bar{\tau}_k^n(\mathbf{x}_k(n+1)) \Big) + \tilde{V}_k^n(\tilde{\mathbf{x}}_k(n+1)) \Big\} \end{aligned} \tag{21}$$

Definition 3: A OE control policy $\hat{\mathbf{\Theta}}^*$ is said to possess the asymptotically Markovian property if

$$\lim_{\kappa \to \infty} \mathsf{E}\left[V_k\left(\mathbf{x}_k, \mathbf{x}_{-k} | \mathbf{\Theta}^*, \hat{\mathbf{\Theta}}^*\right) - V_k\left(\mathbf{x}_k, \mathbf{x}_{-k} | \hat{\mathbf{\Theta}}^*, \hat{\mathbf{\Theta}}^*\right) \right] = 0, \tag{23}$$

which means that the performance gain achieved by a VUE-pair $k \in \mathcal{K}(n)$ from following the MPE policy $\mathbf{\Theta}^*$ instead of the OE policy approaches zero when the average number of VUE-pairs goes to infinity.

Theorem 2: The asymptotically Markovian property holds for a OE policy $\hat{\mathbf{\Theta}}^*$.

Proof: For any given control policy $\mathbf{\Theta}$, let us define $\forall k \in \mathcal{K}(n)$,

$$\begin{aligned} & \hat{V}_k\left(\mathbf{x}_k | \mathbf{\Theta}, \hat{\mathbf{\Theta}}^*\right) = \\ & \mathsf{E}_{(\mathbf{\Theta}, \hat{\mathbf{\Theta}}^*)}\left[\sum_{t=n}^{\infty} (\gamma)^t f_k(\mathbf{x}_k(t), \theta_k(t), d_k(t)) \, | \mathbf{x}_k \right]. \end{aligned} \tag{24}$$

Note that using [11, Theorem 5.5.1], we have

$$\begin{aligned} \hat{V}_k\left(\mathbf{x}_k | \hat{\mathbf{\Theta}}, \hat{\mathbf{\Theta}}^*\right) & \leq \max_{\mathbf{\Theta}} \hat{V}_k\left(\mathbf{x}_k | \mathbf{\Theta}, \hat{\mathbf{\Theta}}^*\right) \\ & = \bar{V}_k\left(\mathbf{x}_k | \mathbf{\Theta}^*, \hat{\mathbf{\Theta}}^*\right). \end{aligned} \tag{25}$$

When $\kappa \to 0$, [9, Theorem 5.1] implies

$$\begin{aligned} & V_k\left(\mathbf{x}_k, \mathbf{x}_{-k} | \mathbf{\Theta}^*, \hat{\mathbf{\Theta}}^*\right) - V_k\left(\mathbf{x}_k, \mathbf{x}_{-k} | \hat{\mathbf{\Theta}}^*, \hat{\mathbf{\Theta}}^*\right) \\ & = V_k\left(\mathbf{x}_k, \mathbf{x}_{-k} | \mathbf{\Theta}^*, \hat{\mathbf{\Theta}}^*\right) - \bar{V}_k\left(\mathbf{x}_k | \mathbf{\Theta}^*, \hat{\mathbf{\Theta}}^*\right) \\ & + \bar{V}_k\left(\mathbf{x}_k | \mathbf{\Theta}^*, \hat{\mathbf{\Theta}}^*\right) - V_k\left(\mathbf{x}_k, \mathbf{x}_{-k} | \hat{\mathbf{\Theta}}^*, \hat{\mathbf{\Theta}}^*\right) \\ & \leq V_k\left(\mathbf{x}_k, \mathbf{x}_{-k} | \mathbf{\Theta}^*, \hat{\mathbf{\Theta}}^*\right) - \hat{V}_k\left(\mathbf{x}_k | \hat{\mathbf{\Theta}}, \hat{\mathbf{\Theta}}^*\right) \\ & + \bar{V}_k\left(\mathbf{x}_k | \mathbf{\Theta}^*, \hat{\mathbf{\Theta}}^*\right) - V_k\left(\mathbf{x}_k, \mathbf{x}_{-k} | \hat{\mathbf{\Theta}}^*, \hat{\mathbf{\Theta}}^*\right) \to 0, \end{aligned} \tag{26}$$

which concludes the proof. □

V. Simulation Results

We perform simulations to evaluate our proposed auction based learning scheme for wireless resource scheduling in a V2V communication network. For comparison purpose, three baselines are simulated as well, namely,

1) Channel-aware control policy – A VUE-pair evaluates the need of occupying the frequency resource for packet transmissions based on the channel state at each scheduling slot and does not take into account the queue state.

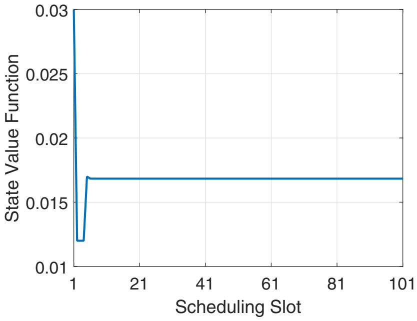

Fig. 2. Illustration of the convergence property of our proposed scheme.

2) Queue-aware control policy – A VUE-pair announces at each scheduling slot the preference of obtaining the frequency resource to maximize the expected long-term number of packets to be transmitted [12].
3) Random control policy – This policy randomly generates the bid of having the frequency resource for a VUE-pair at each scheduling slot, which means that the random policy does not consider any dynamics in the system.

For all simulations, we consider a single RSU scenario. The bandwidth of the frequency resource is 5 MHz. The maximal queue length is set to be $q^{(max)} = 10$, and each packet is of 5000 bits. Packets arrive according to a Poisson process with average rate of λ. Furthermore, all incumbent VUE-pairs are assumed to experience independent and identically distributed channel gains from a common set $\mathcal{G} = \{-18.82, -13.79, -11.23, -9.37, -7.8, -6.3, -4.68, -2.08\}$ (dB), the mean of which is given by $\bar{g}$. $u_k^{(1)}(q_k(n))$, $u_k^{(2)}(p_k(n))$ and $u_k^{(3)}(o_k(n))$ of a service request $k \in \mathcal{K}(n)$ are chosen to be

$$\begin{aligned} & u_k^{(1)}(q_k(n)) = e^{-q_k(n)}, u_k^{(2)}(p_k(n)) = e^{-p_k(n)}, \\ & u_k^{(3)}(o_k(n)) = e^{-o_k(n)}. \end{aligned} \tag{27}$$

Other parameters are set to be: $\alpha_k = 10^{-3}$ and $\gamma = 0.9$.

We first validate the convergence property of our proposed learning based wireless scheduling scheme. In the simulation, the average number of service requests in the steady V2V network is set to be $\kappa = 20$. We fix the mean of channel states and the average packet arrival rate to be $\bar{g} = -2$ and $\lambda = 0.3$. Without loss of the generality, we plot the simulated variations in post-decision state value function, $\tilde{V}_k((-7.8, 6))$, for a k, versus the slots in Fig. 2, which tells that the proposed scheme converges reasonably fast.

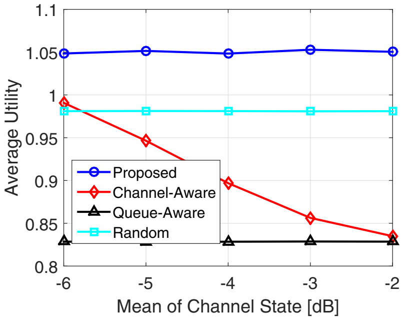

Fig. 3. Average utility per service request across the learning procedure versus means of channel states, where $\kappa = 20$ and $\lambda = 0.3$.

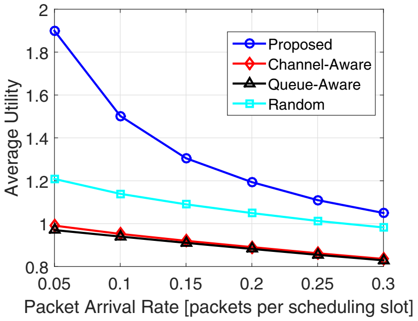

Fig. 4. Average utility per service request across the learning procedure versus average packet arrival rates, where $\kappa = 20$ and $\bar{g} = -2$.

We then simulate the average utility achieved by the service requests during the competition for the frequency resource under various settings of the mean channel state and the average packet arrival rate. The results are depicted in Figs. 3 and 4. From both plots, our proposed scheme outperforms the baselines. As the channel quality becomes better, the average utility performance from the channel-aware scheme becomes worse and the other three schemes realize nearly stable performance. This observation can be explained that due to a choice of small value of α_k, $u_k^{(1)}(\cdot)+u_k^{(3)}(\cdot)$ dominates the value of utility function. The performance from all schemes decreases when the average packet arrival rate increase. The reason is obvious. More packet arrivals lead to more queued and dropped packets, and hence smaller average utility.

Finally, the average utility performance from all schemes under different network size is exhibited in Fig. 5. Similar observations as in Fig. 4 can be made from Fig. 5. More incumbent service requests mean higher competition for the frequency resource, which indicates worse utility performance for the wireless resource scheduling schemes.

VI. Conclusions

In this paper, we study the problem of non-cooperative wireless resource scheduling in a V2V communication network. Under the assumptions of time-varying channel qualities and data arrivals, the problem is formulated as a stochastic game. The interactions among the incumbent service requests are controlled by the RSU through a sealed second-price auction. The technical challenge lies in the network state-space explosion due to the large number of service requests in an urban Manhattan area. Therefore, we propose to adopt a OE to approximate the MPE solution. Without a priori statistics knowledge of channel quality and packet arrival variations, we derive an online learning algorithm to approach the OE control policy. The performance gap between the MPE and the OE solutions is theoretically analyzed. From simulations, significant performance gain from our proposed scheme is showcased in terms of average utility performance.

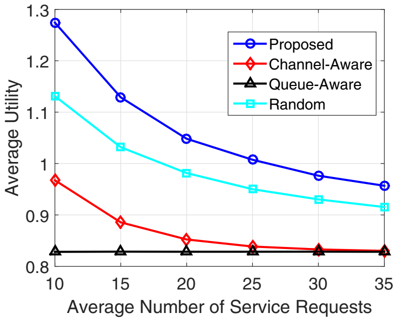

Fig. 5. Average utility per service request across the learning procedure versus numbers of service requests, where $\bar{g} = -3$ and $\lambda = 0.3$.

References

[1] P. Rost, A. Banchs, I. Berberana, M. Breitbach, M. Doll, H. Droste, C. Mannweiler, M. A. Puente, K. Samdanis, and B. Sayadi, "Mobile network architecture evolution toward 5G," *IEEE Wireless Commun.*, vol. 54, no. 5, pp. 84–91, May 2016.

[2] 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.

[3] 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.

[4] Y. Yao, X. Chen, L. Rao, X. Liu, and X. Zhou, "LORA: Loss differentiation rate adaptation scheme for vehicle-to-vehicle safety communications," *IEEE Trans. Veh. Technol.*, vol. 66, no. 3, pp. 2499–2512, Mar. 2017.

[5] E. Egea-Lopez and P. Pavon-Mariño, "Distributed and fair beaconing rate adaptation for congestion control in vehicle network," *IEEE Trans. Mobile Comput.*, vol. 15, no. 12, pp. 3028–3041, Dec. 2016.

[6] Q. Zheng, K. Zheng, H. Zhang, and V. C. M. Leung, "Delay-optimal virtualized radio resource scheduling in software-defined vehicular networks via stochastic learning," *IEEE Trans. Veh. Technol.*, vol. 65, no. 10, pp. 7857–7867, Oct. 2016.

[7] W. Vickrey, "Counterspeculation, auctions and competitive sealed tenders," *J. Finance*, vol. 16, no. 1, pp. 8–37, Mar. 1961.

[8] S. Adlakha, R. Johari, G. Y. Weintraub, and A. Goldsmith, "On oblivious equilibrium in large population stochastic games," in *Proc. IEEE CDC*, Atlanta, GA, Dec. 2010.

[9] G. Y. Weintraub, L. Benkard, and B. Van Roy, "Oblivious equilibrium: A mean field approximation for large-scale dynamic games," in *Proc. NIPS*, Vancouver, Canada, Dec. 2005.

[10] Y. Zhuang, J. Pan, V. Viswanathan, and L. Cai, "On the uplink MAC performance of a drive-thru internet," *IEEE Trans. Veh. Technol.*, vol. 61, no. 4, pp. 1925–1935, May 2012.

[11] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. New York, NY: John Wiley & Sons, 1994.

[12] F. Fu and M. van der Schaar, "Learning to compete for resources in wireless stochastic games," *IEEE Trans. Veh. Technol.*, vol. 58, no. 4, pp. 1904–1919, May 2009.