This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2018.2872013, IEEE Internet of Things Journal

Artificial Intelligence Inspired Transmission Scheduling in Cognitive Vehicular Communications and Networks

Ke Zhang, Supeng Leng, *Member, IEEE,* Xin Peng, Li Pan, Sabita Maharjan, *Member, IEEE,* and Yan Zhang, *Senior Member, IEEE*

***Abstract*—The Internet of things (IoT) platform has played a significant role in improving road transport safety and efficiency by ubiquitously connecting intelligent vehicles through wireless communications. Such an IoT paradigm however, brings in considerable strain on limited spectrum resources due to the need of continuous communication and monitoring. Cognitive radio (CR) is a potential approach to alleviate the spectrum scarcity problem through opportunistic exploitation of the underutilized spectrum. However, highly dynamic topology and time-varying spectrum states in CR-based vehicular networks introduce quite a few challenges to be addressed. Moreover, a variety of vehicular communication modes, such as vehicle-to-infrastructure (V2I) and vehicle-to-vehicle (V2V), as well as data QoS requirements pose critical issues on efficient transmission scheduling. Based on this motivation, in this paper, we adopt a deep Q-learning approach for designing an optimal data transmission scheduling scheme in cognitive vehicular networks to minimize transmission costs while also fully utilizing various communication modes and resources. Furthermore, we investigate the characteristics of communication modes and spectrum resources chosen by vehicles in different network states, and propose an efficient learning algorithm for obtaining the optimal scheduling strategies. Numerical results are presented to illustrate the performance of the proposed scheduling schemes.**

***Index Terms*—Cognitive radio, vehicular communication, Q-learning, transmission scheduling.**

I. Introduction

Advancements in information and communication technologies as well as proliferation of IoT devices have contributed much in expanding the reach of intelligent transportation system (ITS) in modern society. ITS is expected to play a crucial role in paving a path towards enabling the design, formation and proper function of coordinated transport networks, provide comfortable driving experience, improve traffic management and realize smart vehicular applications [1].

To realize the above ITS service provisioning, vehicles as well as traffic infrastructures and management systems need to be connected and informed through wireless vehicular communications. However, the exponential proliferation of radio-equipped vehicles and ubiquitous vehicular applications significantly increase transmission demands on wireless resources, and may lead to serious spectrum scarcity problem [2].

Cognitive Radio (CR) is a context-aware intelligent radio, which may relieve spectrum scarcity while improving spectrum efficiency through adaptively detecting and reusing underutilized portion of spectrum [3]. Enabled with the capabilities of exploiting spectrum resources, CR has been widely implemented in various wireless communication applications, such as public safety services and wireless sensor networks. In ITS, although the concept of cooperative wireless communications among mobile vehicles was proposed to make the driving experience safer and more comfortable, the existing allocated short range communication spectrum may not be enough to deliver data under strict delay constraints. To improve the performance of vehicular communications, additional available spectrum outside the dedicated channels should be efficiently detected and opportunistically utilizing in a CR framework.

Although employing CR technology for vehicular communications is a promising approach to enable vehicles to opportunistically access the underutilized spectrum, the inherent characteristics of vehicular networks, such as highly dynamic communication topology and complicated correlation between vehicular communication pairs, introduce critical challenges for efficient and reliable data transmission [4].

One of such challenges is time-varying data rate. In CR-based vehicular networks, highly dynamic topology caused by the variations in vehicle traffic distributions as well as the changes in available spectrum lead to intermittency, and make transmission scheduling in vehicular networks even more complex. In addition, vehicular networks always have heterogeneous transmission modes, such as vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) [5]. Working in V2V mode, intelligent vehicles can act as mobile transmission relays, which may help deliver data in a long distance. However, the correlation among various transmission modes that compete for limited spectrum resources always causes difficulties in efficiently utilizing available spectrum on data delivery. Jointly considering various QoS constraints of vehicular transmission and dynamic spectrum resource characteristics of the road sections, where the vehicles are being driven, make vehicular communication management further complicated. Nonetheless, recent advances in ITS technologies have greatly

K. Zhang and S. Leng are with the School of Information and Communication Engineering, University of Electronic Science and Technology of China (e-mail:{zhangke, spleng}@uestc.edu.cn).

X. Peng and L. Pan are with College of Information and Communication Engineering, Hunan Institute of Science and Technology, China (e-mail: xpeng@hnist.edu.cn, panli.hnist@gmail.com).

S. Maharjan is with Simula Metropolitan Center for Digital Engineering, and University of Oslo, Norway (e-mail: sabita@simula.no).

Y. Zhang is with Department of Informatics, University of Oslo, Norway (e-mail: yanzhang@ieee.org).

Corresponding author: Y. Zhang (e-mail: yanzhang@ieee.org).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2018.2872013, IEEE Internet of Things Journal

enhanced the caching capability of vehicles and have made cache-enabled vehicles a new promising approach to store and disseminate data [6]. However, utilizing onboard caching resources and high mobility characteristics of vehicles for efficient data delivery is still a critical challenge.

New approaches are required to address these challenges in order to provide reliable and efficient communications in CR-enabled vehicular networks. However, very few works have incorporated the mobility of vehicles and the states of spectrum resources into vehicular communication management, while the capability of mobile vehicle caching and multiple transmission modes also have not been fully exploited in CR scenarios.

To bridge this gap, in this paper, we focus on data transmission in cognitive vehicular networks. By integrating a wide range of communication resources and various types of transmission modes, we propose a learning-based optimal data scheduling scheme, which minimizes transmission costs while ensuring delay constraints. The main contributions of this paper are as follows:

- We formulate a Markov decision process model to analyze the transmission performance of CR-enabled vehicular networks, by jointly considering cognitive spectrum, states of vehicular caching, correlation between various transmission modes, mobility of vehicles as well as QoS requirements of data.
- We propose an optimal data transmission scheduling scheme based on a deep Q-learning approach to minimize costs with certain delay constraints by fully utilizing communication resources and vehicular transmission modes.
- We present an extensive analysis of the relation between transmission actions, delay constraints and incurred costs in various network states, and design an efficient learning algorithm for obtaining the optimal scheduling strategies.

The rest of this paper is organized as follows. Related works are reviewed in Section II. System model is presented in Section III. The transmission scheduling problem is formulated in Section IV. A deep Q-learning based scheduling scheme is introduced in Section V. Evaluation results are presented in Section VI and the paper is concluded in Section VII.

II. Related Works

CR has attracted significant attention over the last few years as a promising technology to address spectrum scarcity issues for emerging IoT applications. In [7], the authors focused on spectrum allocation in CR-based IoT, and proposed an allocation scheme with efficient spectrum utilization and high network throughput. In order to fully exploit the advantages of time efficiency, the authors in [8] investigated the information delivery dynamics of secondary users in cognitive sensor networks via epidemic models. In [9], the authors deployed CR technology in collecting the monitoring data of smart grid, and designed a traffic scheduling scheme based on binary exponential backoff algorithm. In [10], the authors studied security of communication between CR-enabled IoT devices, and proposed a probabilistic-based channel assignment mechanism to deal with jamming attacks. In [11], the authors investigated a wireless powered cognitive IoT network, and introduced an efficient information delivery scheme, which jointly optimizes energy harvesting and data transmission. However, the above works mainly focused on static communication topology with homogeneous IoT devices, where dynamic characteristics and various communication requirements of the devices have not been taken into consideration.

With the proliferation of smart vehicles, many works have studied CR-enabled vehicular networks. In [12], the authors investigated decision fusion techniques of spectrum sensing in cognitive vehicular communications. In [13], the authors focused on coexistence between CR-based vehicular and 802.22 networks, and optimized spectrum resource assignment as well as transmit power. The authors in [14] presented a semi-Markov decision policy based channel allocation scheme, which is priority- aware and improves overall system rewards. In [15], the authors designed energy efficient power allocation strategies for a cognitive vehicular network under primary user emulation attacks. The authors in [16] designed an adaptive double threshold spectrum sensing scheme to address spectrum scarcity in vehicular environments. However, only a few of these works have considered the influence of vehicle mobility and spectrum states of different regions on the design of data transmission strategies. In addition, joint optimization of spectrum access and vehicular transmission mode selection with QoS constraints has not been investigated in these studies.

Being a powerful tool in process control and resource management, learning has been applied in a wide range of areas. In [17], the authors designed an integrated resource management scheme for connected vehicles using a deep reinforcement learning approach. Recently, various learning techniques have been applied in the study of CR networks. In [18], the authors compared the performance of different machine learning approaches in terms of spectrum classification accuracy and computational time. The authors in [19] proposed a stochastic learning based spectrum access scheme that maximizes the throughput of CR networks. In [20], the authors incorporated reinforcement learning technology with Bayesian approach in cognitive channel sensing and selection, and designed a two-stage spectrum access scheme. In [21], by using reinforcement learning mechanism, the authors improved routing scalability and stability in the context of cognitive radio. The authors in [22] introduced a Q-learning based transmission scheme for CR-based networks, which improves system throughput through scheduling of cognitive nodes. However, none of the aforementioned works have incorporated learning techniques into designing data transmission schemes of cognitive vehicular networks. Different from these studies, in this paper, we jointly take the spectrum characteristics of different regions and the mobility of vehicles into account, and propose an efficient data scheduling scheme with optimal transmission mode selection and spectrum utilization integrating with a deep Q-learning approach.

III. System Model

Fig. 1 shows the architecture of a CR-enabled vehicular network in a unidirectional road scenario. There are M road-side units (RSUs) located along the road providing vehicular

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2018.2872013, IEEE Internet of Things Journal

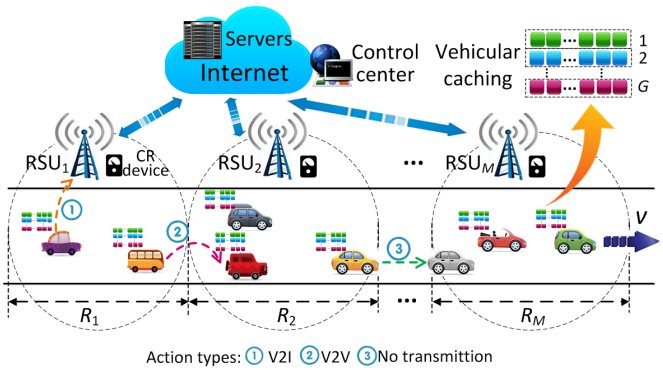


Fig. 1. Data transmission and caching in a CR-enabled vehicular network.

TABLE I
MAIN VARIABLES

Variables	Description
M	Number of RSUs
K_I	Number of licensed channels for V2I mode
K_V	Number of licensed channels for V2V mode
K_C	Number of CR channels
ρ	Traffic density
τ	Length of a time frame
N	Number of data types
b_j	Amount of type-j data
T_j	Transmission delay constraint of type-j data
η_j	Probability of type-j data
G	Number of caching queues in a vehicle
$x^l_{v,e}$,$y^l_{v,e}$,$z^l_{v,e}$	Transmission modes for vehicle v running on road segment e at time frame l
c_v,c_c	Costs for using a licensed channel and a CR channel, respectively
w	Timeout penalty
q^l_v	Caching state of vehicle v at frame l
a^l_v	Action of vehicle v at frame l

communication services to the vehicles on the road. The diameter of regions covered by these RSUs are $\{R_1, R_2, ..., R_M\}$, respectively. We consider that the licensed channels assigned for vehicular communication can be divided into two parts, namely K_I channels for V2I transmission mode and K_V channels for V2V mode. All the licensed channels are orthogonal, and the bandwidth of each channel is B. Besides the licensed channels, the V2I communication can opportunistically utilize CR spectrum resources. Each RSU is equipped with a CR device, which is able to detect available channels within the RSU's covering area. The number of CR channels is K_C. Birth-death process is a special case of continuous-time Markov process where the state transitions are of only two types: "births", which increase the state variable by one and "deaths", which decrease the state by one. For a given CR channel, it has two states. A busy state represents the period used by primary users and an idle state represents the unused period. These two states alternate with each other. Thus the state transition of the CR channel can be taken as a birth-death process. Considering that the primary users in different regions may have different working patterns on the CR channels, we model the activity of primary users in region m as a two-state birth-death process, and the on-time and off-time are exponentially distributed with rate λ^{on}_m and λ^{off}_m, respectively, $m \in \mathcal{M} = \{1, 2, ..., M\}$ [23].

Vehicles arrive at the starting point of the road following Poisson distribution with average density ρ, and move along the road at a constant speed V. It is noteworthy that the system model can be extended to a scenario, where the vehicles run at various speed. We divide the vehicles into different categories according to their speed. Each category of vehicles can be modeled as an independent Poisson arrival process with an dedicated arrival rate. Merging these Poisson processes can form a new Poisson process, whose arrival rate is sum of the rates of the separate process. The road is divided into E segments and the position of a vehicle is defined as the index of the road segment that the vehicle is located in. Each vehicle has a vehicular communication interface, which enables data transmission between vehicles and the RSUs. In an ITS system, various types of data are generated from onboard traffic monitoring and entertainment applications, and transmitted to the servers located in the Internet.

We focus on the management of uplink data transmission in the CR-enabled vehicular network. Specifically, as the RSUs connect to the servers via broadband wirelines whose transmission cost and delay can be ignored, we mainly investigate data transmission between vehicles and RSUs. In the vehicular network, there is a control center, which gathers traffic and network states from vehicles while scheduling vehicular transmission through a dedicated control channel. The transmission scheduling operates in a discrete time model with fixed length time frames. The length of a frame is denoted as τ. We consider τ is short and vehicular communication topology is constant during a time frame. At the end of each frame, a vehicle may generate some data that needs to be transmitted to the servers. We classify the generated data into N types. For each type of data, it is described in two terms as $\{b_j, T_j\}$, where b_j is the amount of data, T_j is the transmission delay constraint, and $j \in \mathcal{N} = \{1, 2, ..., N\}$. It is noteworthy that the unit of delay constraint is time frame. In other words, T_j indicates the time duration of τT_j. The probability that a vehicle generates type-j data in a time frame is η_j, where $\sum_{j=1}^{N} \eta_j \leq 1$.

The generated data can be transmitted from vehicles to the RSUs in V2I and V2V modes. When adopting direct V2I mode, a vehicle transmits data to the RSU whose communication range covers it. Moreover, considering that vehicles can communicate with each other through V2V connections, data may be delivered to remote RSUs through joint multi-hop V2V and V2I transmission. In addition, besides V2V transmission, cache-enabled vehicles can also bring cached data to remote road segments along with them, and then deliver the data to the RSUs in V2I mode.

The data caching in a vehicle is modeled as a multiple-queue system, which is shown in Fig. 2. Each queue consists of the data with identical remaining transmission time under its delay constraint. According to the types of the generated data, there are G caching queues in a vehicle indexed as $\{1, 2, ..., G\}$, respectively. Here $G = \max\{T_j, j \in \mathcal{N}\}$. For caching queue t of a given vehicle v, where $1 \leq t < G$, its input data can be categorized into three sources. The first one is the data generated by vehicle v itself, whose total delay

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2018.2872013, IEEE Internet of Things Journal

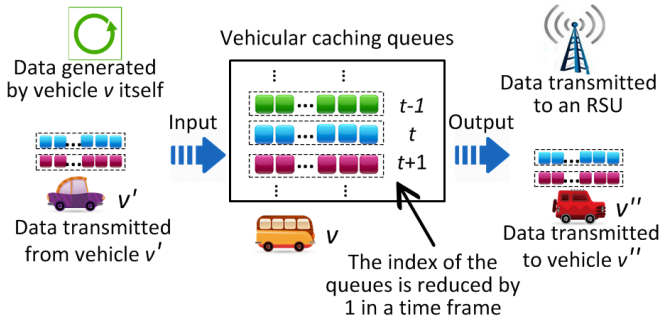


Fig. 2. Queue model of vehicular caching.

constraint is t. The second one is the data with remaining time $t+1$ transmitted from another vehicle, such as v' through V2V communication in the last time frame. The last one is the data cached in queue $t+1$ of vehicle v in the last time frame, which has not been transmitted. As time passes, the remaining time for transmission within the constraint decreases. Thus, the data needs to be moved to queue t. The output of a queue can be also divided into three types. Data in a queue of vehicle v can be transmitted to an RSU through V2I communication or to vehicle v'' in V2V mode. Besides these two types, data in queue t may be moved to queue $t-1$ of the same vehicle as time passes. For queue G, its input data only comes from the newly generated data of type G. To be able to transmit data under the specified delay constraints, the data cached in the queues with smaller index has higher priority to be transmitted by each vehicle.

IV. Data Transmission Scheduling: Analysis and Problem Formulation

In this section, we first investigate the performance of the CR-based vehicular networks with various transmission modes. Then we formulate an optimal data transmission scheduling problem that takes into account both delay constraints and transmission costs.

A. Analysis of CR-based Vehicular Communications

In vehicular networks, V2I communication is scheduled by the control center. When a vehicle needs to transmit data to an RSU, it sends a request to the control center through the dedicated control channel at the beginning of a time frame. Upon the request, the control center randomly chooses an available channel from the corresponding channel set, and allocates it to the vehicle. For simplicity, we consider that each vehicle at most gets one channel in one time frame.

Due to non-overlapping coverage of the RSUs, there is no interference between V2I transmissions in different coverage areas. Furthermore, as the channels are orthogonal, the V2I transmission performance is mainly affected by the distance between communication pairs and the availability of the spectrum resources. When vehicle v transmits data to RSU m on a licensed V2I channel, the transmission rate is given as

$$r_{v,m}^{L} = B\log(1+\frac{P_I}{L_0 d_{v,m}^{\alpha} P_w}), \tag{1}$$

where $d_{v,m}$ is the distance between vehicle v and RSU m, L_0 is the path loss at a reference unit distance, and α is path loss exponent. P_I and P_w are transmission power of a vehicle working in V2I mode and the power of additive white Gaussian noise, respectively. It is noteworthy that the transmission rate $r_{v,m}^{L}$ can be extended to a scenario with overlapping RSU coverage. In this scenario, besides path loss and white Gaussian noise, the interference between V2I communication pairs working on the same channels of different RSU coverage also needs to be incorporated in data transmission rates.

To exploit the underutilized CR spectrum resources, at the beginning of each time frame, the control center may allocate the detected available channels to the vehicles for V2I transmission. However, unlike assigned licensed channels to the vehicles, which is always available during a time frame for V2I transmission, a CR channel may be occupied by a primary user at any time. When this is the case, the vehicle should release the channel to the primary user. The average data transmission rate from vehicle v to RSU m through a CR channel is calculated as

$$r_{v,m}^{C} = \int_0^{\tau} \frac{\varpi B_c \lambda_m^{on} r_{v,m}^{L} Pr\{T_m^{off} > \varpi\}}{\tau B(\lambda_m^{on} + \lambda_m^{off})} d\varpi, \tag{2}$$

where $Pr\{T_m^{off} > \varpi\}$ represents the probability that no primary users arrive on the CR channel during time $\{0, \varpi\}$, given by $Pr\{T_m^{off} > \varpi\} = \exp(-\lambda_m^{off}\varpi)$. B_c is the bandwidth of a CR channel.

Vehicles may also deliver data in V2V mode. A vehicle can transmit data to another vehicle only under the condition that the two vehicles are located within the transmission range of each other, and the sending vehicle chooses V2V transmission while the receiving one does not take any transmission action in the same time frame. When there are multiple possible vehicular communication pairs, the pairs are formed from highest transmission rate to the lowest one. A control center in the vehicular network gathers the states of vehicles, and schedules vehicular transmission through a dedicated control channel. To make full use of the resources of channel set K_V, several V2V communication pairs can take place concurrently on the same channel using space division multiplexing [24]. In such a case, the data rate for V2V transmission from vehicle v to v' can be written as

$$r_{v,v'} = B\log(1+\frac{P_V / L_0 d_{v,v'}^{\alpha}}{P_w + \sum_{z\in\mathcal{Z}} P_V / L_0 d_{z,v'}^{\alpha}}), \tag{3}$$

where P_V is the transmission power of a vehicle in V2V mode. $d_{v,v'}$ is the distance between vehicles v and v'. $\mathcal{Z}$ denotes the set of other vehicles that communicate in the same channel within the interference range.

B. Problem Formulation

In a given time frame, each vehicle can transmit data through vehicular communication or keep the data in its cache. Let $x_{v,e}^{l} = 1$ indicate that vehicle v running on road segment e at time frame l chooses to transmit data through a CR channel in V2I mode. Similarly, we use $y_{v,e}^{l} = 1$ and $z_{v,e}^{l} = 1$ to indicate the transmission action taken by the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2018.2872013, IEEE Internet of Things Journal

vehicle through a licensed channel in V2I and V2V modes, respectively. Otherwise, these indicators are set to 0. The case $x^l_{v,e} = y^l_{v,e} = z^l_{v,e} = 0$ indicates that vehicle v does not transmit and stores the data in its cache instead.

In our proposed CR-enabled vehicular network framework, a proper transmission scheduling should consider maintaining required QoS, specifically the delay constraints. On the other hand, due to the scarcity of spectrum resources, efficient spectrum utilization should also be taken into consideration. To promote the vehicles to transmit data under delay constraints, we introduce a penalty mechanism. The transmission that does not meet the delay constraints brings a penalty to the system. A penalty amount of w is incurred to a vehicle for not meeting the latency constraint of a unit data. In addition, the costs for using a licensed channel and a CR channel in a time frame are c_v and c_c, respectively. Here the cost means the payment that a vehicle needs to pay to transmission service providers. Unlike the communication on CR channels that may be interrupted by the arrival of primary users, data transmission on licensed channels always has constant bandwidth during a time frame. As reliable and fast transmission service always has a higher benefit for a vehicle, we consider $c_v > c_c$. The objective of transmission scheduling is to minimize the penalty and the costs together.

Recall that the data cached in a vehicle is stored in multiple queues according to the remaining number of time frames for transmission under delay constraints. For the data cached in queue 1, which has only one time frame remaining for transmission, if the transmission fails, the deadline can not be met for the data. This transmission should be in V2I mode, since V2V mode cannot execute data delivery within one frame. Let sets $\mathcal{V}^l_{m,L}$ and $\mathcal{V}^l_{m,C}$ denote the vehicles in region m at time frame l, which chooses V2I transmission through a licensed channel and a CR channel, respectively. The vehicles in these sets may compete for limited spectrum resources, and their transmission performance may also be degraded by the intermittence of CR channels. The amount of data not meeting the delay constraints in time frame l can be calculated as

$$\begin{aligned} h^l = \sum_{m\in\mathcal{M}}\sum_{e\in\mathcal{E}_m}\sum_{v\in\mathcal{V}^l_e}(y^l_{v,e}\max\{0, q^l_{v,1} - \tfrac{r^L_{v,m}\tau K_I}{|\mathcal{V}^l_{m,L}|}\} \\ + x^l_{v,e}\max\{0, q^l_{v,1} - \tfrac{r^C_{v,m}\tau K_C}{|\mathcal{V}^l_{m,C}|}\} + (1-y^l_{v,e})(1-x^l_{v,e})q^l_{v,1}) \end{aligned}, \tag{4}$$

where $\mathcal{E}_m$ is the set of road segments within region m and $\mathcal{V}^l_e$ denotes the set of vehicles located in segment e at time frame l. $q^l_{v,t}$ is the length of data cached in queue t of vehicle v in frame l. $|\mathcal{V}^l_{m,L}|$ is the number of vehicles that belong to set $\mathcal{V}^l_{m,L}$, i.e., $|\mathcal{V}^l_{m,L}| = \sum_{e\in\mathcal{E}_m}\sum_{v\in\mathcal{V}^l_e} y^l_{v,e}$. Similarly, we have $|\mathcal{V}^l_{m,C}| = \sum_{e\in\mathcal{E}_m}\sum_{v\in\mathcal{V}^l_e} z^l_{v,e}$.

Based on the above analysis, the proposed optimal data transmission scheduling problem, which intends to ensure that the delay constraints are met, while making full use of various spectrum resources, is formulated as follows:

$$\begin{aligned} \min_{\{x,y,z\}} Loss = \sum_{l=1}^{\infty}\{wh^l + \sum_{m\in\mathcal{M}}(c_v\sum_{e\in\mathcal{E}_m}\sum_{v\in\mathcal{V}^l_e}(y^l_{v,e}+z^l_{v,e}) \\ + c_c\sum_{e\in\mathcal{E}_m}\sum_{v\in\mathcal{V}^l_e}x^l_{v,e})\} \\ \text{s.t.} \quad \text{C1}: \quad x^l_{v,e}=\{0,1\}, \quad y^l_{v,e}=\{0,1\}, \quad z^l_{v,e}=\{0,1\} \\ \text{C2}: \quad x^l_{v,e}y^l_{v,e} = x^l_{v,e}z^l_{v,e} = y^l_{v,e}z^l_{v,e} = 0 \end{aligned}. \tag{5}$$

In (5), constraint C1 indicates that a vehicle can either take one transmission mode or not in time frame l. Constraint C2 shows that no vehicle can choose more than one transmission mode in the same time frame.

V. Learning Based Optimal Transmission Scheduling

In this section, we model the data transmission scheduling problem as a Markov decision process, and design a deep Q-learning based approach to derive optimal scheduling strategies.

A. Scheduling as a Markov Decision Process

In the proposed scheduling problem, the value of $Loss$ mainly depends on vehicular caching states and transmission mode selection. Moreover, vehicular caching operation is modeled as a queuing system and processes with state transitions. The caching state of next time frame is only related to the current state and the transmission modes, which are chosen by the vehicles according to the states as well as characteristics of the vehicular network in the current frame. Therefore, we can formulate transmission scheduling problem (5) as a Markov decision process (MDP).

The state of the MDP at time frame l can be defined as $S^l = \{Q^l_1, Q^l_2, ..., Q^l_{|\mathcal{V}|}, \Theta^l\}$, where $\mathcal{V}$ is the set of vehicles in the system. Q^l_v is the caching state of vehicle v at frame l, and it can be shown as $\{q^l_{v,1}, q^l_{v,2}, ..., q^l_{v,G}\}$, $v \in \mathcal{V}$. Θ^l is the set of vehicle positions in time frame l, which are valued by the index of the road segments where the vehicles are located in. The action taken by the vehicles at frame l is given as A^l. Specifically, for a vehicle located in road segment e, A^l can be written as $A^l_e = \{a^l_{1,e}, a^l_{2,e}, ..., a^l_{|\mathcal{V}|,e}\}$. $a^l_{v,e}$ is the action of vehicle v, which consists of the possible transmission modes and can be further expressed as $a^l_{v,e} = \{x^l_{v,e}, y^l_{v,e}, z^l_{v,e}\}$.

The MDP state transition between two consecutive time frames is a transition combination of both vehicle positions and caching states. The updated position Θ^{l+1} can be obtained based on the original position Θ^l and running speed. Now we focus on the caching state transitions. For a given vehicle v, its caching state transition can be further divided into transitions of multiple queue states, i.e., $Q^{l+1}_v = \{q^{l+1}_{v,1}, q^{l+1}_{v,2}, ..., q^{l+1}_{v,G}\}$. The amount of data cached in a queue between time frames may be affected by the newly generated data, outgoing or incoming data through vehicular communications, and the data from the queue with higher index. To facilitate the analysis of various factors that affect the transition of caching queue states, we introduce several variables. Let $t^l_{v,\min}$ be the smallest index of the queue with nonempty queuing data of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2018.2872013, IEEE Internet of Things Journal

vehicle v at time frame l. Then we define the amount of data that is delivered from v to RSU m in V2I mode and the data transmitted from vehicles v to v' in V2V mode at frame l as

$$D_{v,m}^{l,\mathrm{V2I}} = \frac{x_{v,e}^l r_{v,m}^C \tau K_C}{\left|V_{m,C}^l\right|} + \frac{y_{v,e}^l r_{v,m}^L \tau K_I}{\left|V_{m,L}^l\right|}, \tag{6}$$

and

$$D_{v,v'}^{l,\mathrm{V2V}} = z_{v,e}^l r_{v,v'} \tau / K_V, \tag{7}$$

respectively. The amount of data transmitted from queue $t+1$ of vehicle u to queue t of vehicle v in V2V mode can be expressed as

$$\begin{aligned} D_{u,v,t}^{l,\mathrm{V2V}} = (1-x_{v,e}^l)(1-y_{v,e}^l)(1-z_{v,e}^l) r_{u,v}\tau \\ \cdot \mathbf{1}(t_{u,\min}^l == t+1)/K_V \end{aligned}, \tag{8}$$

where $\mathbf{1}(\chi)$ is an indicator function which equals 1 if χ is true and 0 otherwise.

Then, given caching state Q_v^l, the states of queues that form Q_v^{l+1} can be shown in two parts, namely $q_{v,t_{v,\min}^l}^{l+1}$ and $q_{v,t}^{l+1}$ where $t \neq t_{v,\min}^l$, as follows

$$q_{v,t_{v,\min}^l}^{l+1} = \begin{cases} \max\{0, q_{v,2}^l + \eta_1 b_1 - D_{v,m}^{l,\mathrm{V2I}}\}, & t_{v,\min}^l = 1 \\ \max\{0, q_{v,t_{v,\min}^l+1}^l + \eta_t b_{t_{v,\min}^l} - D_{v,m}^{l,\mathrm{V2I}} \\ -D_{v,v'}^{l,\mathrm{V2V}} + D_{u,v,t_{v,\min}^l}^{l,\mathrm{V2V}}\}, & 1 < t_{v,\min}^l < G \\ \max\{0, \eta_G b_G - D_{v,m}^{l,V2I} - D_{v,v'}^{l,V2V}\}, & t_{v,\min}^l = G \end{cases} \tag{9}$$

$$q_{v,t}^{l+1} = \begin{cases} q_{v,t+1}^l + \eta_t b_t, & t \neq t_{v,\min}^l, 1 \leqslant t < G \\ \eta_t b_t, & t \neq t_{v,\min}^l, t = G \end{cases} \tag{10}$$

At time frame l, the sum of transmission penalty and cost from action A^l taken by the vehicles on state S^l, is calculated as

$$\begin{aligned} Loss^l = wh^l + \sum_{m\in\mathcal{M}} (c_v \sum_{e\in\mathcal{E}_m} \sum_{v\in\mathcal{V}_e^l} (y_{v,e}^l + z_{v,e}^l) \\ + c_c \sum_{e\in\mathcal{E}_m} \sum_{v\in\mathcal{V}_e^l} x_{v,e}^l) \end{aligned} \tag{11}$$

The goal of the MDP is to derive an optimal transmission scheduling policy that minimizes the cumulative value of $Loss^l$ over time frames. The optimal policy, which consists of data transmission actions for various vehicles at different time frames, can be written as

$$\pi^* = \arg\min_{\pi} \mathrm{E}(\sum_{l=1}^{\infty} \xi^l Loss^l), \tag{12}$$

where $0 < \xi < 1$ is a discount coefficient that indicates the effect of future reward on the current actions.

B. Deep Q-Learning Based Scheduling Schemes

As a problem with a large scale of state space, the formulated MDP is hard to solve directly [25]. To obtain optimal transmission scheduling strategy π^* in (12), we adopt a learning approach. Q-learning is an attractive approach to guide data transmission scheduling, as it learns from online information instead of pre-prepared training dataset. Furthermore, in the process of Q-learning, agents continue to take actions while getting quantitative reward. The action chosen in the next step is a function of currently learned values, which is similar to the operation of an MDP. Thus, solving MDP problem (12) can be formulated as a Q-learning process.

Let $Q_\pi(S^l, A^l) = \mathrm{E}[(\sum_{l=1}^{\infty} \xi^l Loss^l) \,|(S^l, A^l)]$ denote the average system loss from taken action A^l at state S^l applying the transmission scheduling strategy π. Q functions $Q_\pi(S^l, A^l)$ and $Q_\pi(S^{l+1}, A^{l+1})$ are related as

$$Q_\pi(S^l, A^l) = \mathrm{E}_{S^{l+1}}\left[Loss^l + \xi Q_\pi(S^{l+1}, A^{l+1}) \,|(S^l, A^l)\right]. \tag{13}$$

Given action A^l performed in state S^l, the expected minimum system loss will be

$$Q^*(S^l, A^l) = \mathrm{E}_{S^{l+1}}\left[Loss^l + \xi \min_{A^{l+1}} Q^*(S^{l+1}, A^{l+1}) \,|(S^l, A^l)\right]. \tag{14}$$

To obtain $Q^*(S^l, A^l)$ and corresponding optimal transmission actions, we deploy an iterative approach. The updated value of $Q(S^l, A^l)$ in each iteration can be written as

$$\begin{aligned} Q(S^l, A^l) \leftarrow (1-\varphi) Q(S^l, A^l) + \varphi[Loss^l \\ + \xi \min_{A^{l+1}} Q^*(S^{l+1}, A^{l+1})], \end{aligned} \tag{15}$$

where $0 < \varphi < 1$ is the learning rate.

Although applying Q-learning technique can obtain the optimal scheduling strategies, this learning approach uses a Q-table to store learned state-action combinations and corresponding Q-values. The size of the table is equal to the dimension of the states multiplied by the dimension of the actions. Due to the constraint size of computer cache, it is a critical challenge to store a Q-table especially with high numbers of states and actions. To compensate the limitation of Q-learning, we further incorporate deep learning technology with Q-learning approach, and proposed a deep Q-learning based vehicular data transmission scheduling scheme [26].

In deep Q-learning, an efficient mapping construction between states, actions and awards plays a crucial role in learning optimal scheduling strategies from high-dimensional information. As neural networks are suitable for capturing the complex relationship between a large amount of data, we turn the representation of our Q-function into a function approximator formed by a four-layered neural network, with two hidden layers besides input layer and output layer. The inputs of the neural network are the system states and actions, and the outputs are the corresponding Q-function values, which is shown as $Q(S^l, A^l) \approx Q'(S^l, A^l; \theta)$ [27]. Here θ denotes the parameters of the neural network. Based on $Q'(S^l, A^l; \theta)$, the optimal action in state S^l is the one that results in the minimized Q-function value, which can be expressed as

$$A_{\mathrm{opt}}^l = \arg\min_{A^l} Q'(S^l, A^l; \theta) \tag{16}$$

To ensure the function approximation ability of the neural network, $Q'(S^l, A^l; \theta)$ should be trained to converge to the real value of $Q(S^l, A^l)$ over iterations. We define the difference between the value of these two Q-functions at frame l as

$$\Delta(\theta^l) = \mathrm{E}\left[\frac{1}{2}(Q_{\mathrm{tar}}^l - Q'(S^l, A^l; \theta^l))^2\right], \tag{17}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2018.2872013, IEEE Internet of Things Journal

where θ^l is the parameters of the formed neural network at frame l. $Q^l_{\rm tar}$ is the updated optimal value of $Q(S^l, A^l)$ in time frame l during the deep learning process, and can be written as

$$Q^l_{\rm tar} = Loss^l + \xi Q(S^l, \arg\min_{A^{l+1}} Q'(S^{l+1}, A^{l+1}; \theta^l)). \quad (18)$$

In each iteration, we deploy a gradient descent approach to modify θ. The gradient derived by differentiating $\Delta(\theta^l)$ will be

$$\nabla_{\theta^l}\Delta(\theta^l) = \mathrm{E}[\partial Q'(S^l, A^l; \theta^l)/\partial\theta^l (Q'(S^l, A^l; \theta^l) - Q^l_{\rm tar})]. \quad (19)$$

Then θ^l is updated according to

$$\theta^l \leftarrow \theta^l - \varsigma \nabla_{\theta^l}\Delta(\theta^l), \quad (20)$$

where ς is a step size coefficient.

In order to improve the learning efficiency while preventing local minimum Q-values, experience replay technique is utilized for parameter training. In the learning process, the experience in each iteration, including the actions, state transitions and corresponding Q-values, is stored in a replay memory [28]. When training the neural network, a batch of experience randomly drawn from the replay memory is used as samples instead of the most recently learned system information. This sampling approach breaks the similarity of subsequent training samples, which may lead to local optimization results. Moreover, during the learning process, we adopt ε-greedy policy to balance exploration and exploitation of action selection, where a random action is chosen with probability ε, otherwise a greedy action with the minimum Q-value is taken.

C. Efficient Deep Q-learning Algorithm for Optimal Scheduling

Although deep Q-learning scheme is a promising approach to find optimal data transmission strategies for the vehicles through iterative learning, large scale of network states as well as complex strategies with various communication modes and different types of spectrum resources may make the learning process converge slow. To further improve the learning efficiency, we give some criteria for choosing appropriate actions based on analysis of the transmission performance in different cases.

First, we focus on the vehicles whose $t^l_{v,\min} = 1$, i.e., these vehicles have data cached in queue 1. Considering the delay constraint, the data in queue 1 needs to be transmitted to RSUs within one time frame. Thus, the vehicles should communicate in V2I mode either through licensed channels or CR channels. The channel selection criterion is presented in Theorem 1.

Theorem 1. *Let $\Gamma^l_e = r^L_{v,m}\tau K_I/|V_{m,L}|$ and $\Psi^l_e = r^C_{v,m}\tau K_C/|V_{m,C}|$. For vehicle v that is located in road segment e at time frame l and having $t^l_{v,\min} = 1$, it chooses action $y^l_{v,e} = 1$ only under the condition that $\Gamma^l_e + (c_v - c_c)/w < \Psi^l_e < q^l_{v,1}$ or $\Gamma^l_e < q^l_{v,1} - (c_v - c_c)/w$ while $\Psi > q^l_{v,1}$.*

Proof. According to (11), in order to deliver the data cached in queue 1 within the delay constraint while minimizing the cost in time frame l, vehicle v only chooses a licensed channel for V2I transmission under the condition that $\max\{0, q^l_{v,1} - \Gamma^l_e\}w + c_v < \max\{0, q^l_{v,1} - \Psi^l_e\}w + c_c$. We consider two cases as follows. Case 1: $\Gamma^l_e \geqslant \Psi^l_e$. If $\Gamma^l_e \leqslant q^l_{v,1}$ and $\Psi^l_e \leqslant q^l_{v,1}$, the condition can be changed to $\Psi^l_e - \Gamma^l_e > (c_v - c_c)/w$, which contradicts with $\Gamma^l_e \geqslant \Psi^l_e$ due to $c_v > c_c$. In addition, if $\Gamma^l_e > q^l_{v,1}$ and $\Psi^l_e > q^l_{v,1}$, $(c_v - c_c)/w < 0$ that is contrary to the given condition $c_v > c_c$. Moreover, if $\Gamma^l_e > q^l_{v,1}$ and $\Psi^l_e \leqslant q^l_{v,1}$, we can get $\Psi > (c_v - c_c)/w + q^l_{v,1}$, which contradicts $\Psi \leqslant q^l_{v,1}$. Case 2: $\Gamma^l_e < \Psi^l_e$. In this case, if $\Gamma^l_e > q^l_{v,1}$ and $\Psi^l_e \leqslant q^l_{v,1}$, the condition turns to be $(c_v - c_c)/w < 0$. However, given $\Gamma^l_e \leqslant q^l_{v,1}$ and $\Psi^l_e \leqslant q^l_{v,1}$, the condition equivalents to

$$\Gamma^l_e + (c_v - c_c)/w < \Psi^l_e. \quad (21)$$

Moreover, if $\Gamma^l_e \leqslant q^l_{v,1}$ and $\Psi^l_e > q^l_{v,1}$, we have

$$\Gamma^l_e < q^l_{v,1} - (c_v - c_c)/w. \quad (22)$$

Combining (6(a)), (6(b)) and the given conditions of case 2 proves Theorem 1. □

Next, we investigate data transmission of the vehicles that have empty queue 1 but have data cached in queue 2. These vehicles can adopt various ways to deliver these data to the RSUs, e.g., two consecutive V2I mode transmissions, joint V2V and V2I delivery, and taking cached data to and sending at the following arrived road segment. To facilitate action selection in the learning process, we define the following criteria.

Theorem 2. *For vehicle v that is located in road segment e at time frame l and will arrive at segment e' at frame $l+1$, if $t^l_{v,\min} = 2$, vehicle v takes action $x^l_{v,e} = y^l_{v,e} = z^l_{v,e} = 0$ under the condition that $q^l_{v,2} + \eta_1 b_1 + D^{l,V2V}_{u,v,1} < r^C_{v,m'}\tau$.*

Proof. Considering the transmission costs, when vehicle v chooses to deliver data through any mode in time frame l, it needs to pay at least c_c. However, if the vehicle takes no transmission at l and carries the cached data to the following road segment, no transmission cost occurs. During its running in frame l, vehicle v may generate $\eta_1 b_1$ data by itself and receive $D^{l,V2V}_{u,v,1}$ data from another vehicle. If vehicle v can transmit the cached data as well as the newly generated and received data through a CR channel in frame $l+1$, i.e., $q^l_{v,2} + \eta_1 b_1 + D^{l,V2V}_{u,v,1} < r^C_{v,m'}\tau$, the vehicle pays the same transmission cost as c_c but delivers more data to an RSU than it does in frame l. Under this condition, the vehicle prefers no transmission in frame l. □

Theorem 3. *Let $r_{v,m} = \max\{r^L_{v,m}, r^C_{v,m}\}$. For vehicle v that is located at road segment e in time frame l and that will arrive at e' in frame $l+1$, the vehicle can deliver data to vehicle v' located at e'' through a V2V mode. Vehicle v should not take action $z^l_{v,e} = 1$ at frame l under one of the following conditions, $r_{v,m} > r_{v,v'}$ or $r_{v',m''} + (c_v - c_c)/w < r_{v,m} + r_{v,m'} < q^l_{v,2} + \eta_1 b_1$ or $r_{v',m''} < q^l_{v,2} + \eta_1 b_1 - (c_v - c_c)/w$ while $r_{v,m} + r_{v,m'} > q^l_{v,2} + \eta_1 b_1$.*

Proof. When vehicle v transmits in V2V mode, it should pay transmission cost c_v, which should not be less than the cost

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2018.2872013, IEEE Internet of Things Journal

of V2I transmission either through a licensed channel or a CR channel. In addition, as V2V transmission is always associated with V2I transmission to execute complete data delivery, the transmission step with the slowest rate may be the bottleneck. Thus, if V2V transmission rate $r_{v,v'}$ is less than the maximum V2I rate $r_{v,m}$, vehicle v chooses to transmit data in direct V2I mode instead of joint V2V and V2I transmission. Furthermore, vehicle v should not choose V2V mode in time frame l if $\max\{0, q_{v,2}^l - (r_{v,m} + r_{v,m'})\tau\}w + c_v < \max\{0, q_{v,2}^l - r_{v',m''}\tau\}w + c_c$. The discussion and simplification of the condition is similar as the proof of Theorem 1. □

Based on the above action selection criteria, we propose an efficient deep Q-learning algorithm for deriving the optimal transmission scheduling strategies, which is shown in Algorithm 1.

Algorithm 1 Deep Q-learning algorithm for optimal scheduling

Initialization:

Initialize Q-network with weights θ, action-value function Q, and experience replay memory.

For a give steady vehicular traffic flow **Do**
  Observe the initial state S^0;
  For time frames $l = 0, ..., L_{\max}$ **Do**
    Based on the criteria in Theorems 1, 2 and 3, select a random action A^l with probability ε, otherwise choose action $A^l = \arg\max_A Q(S^l, A; \theta)$;
    Execute action A^l, derive the next state S^{l+1} and obtain utility $Loss^l$ according to (9), (10) and (11);
    Store the experience $(S^l, A^l, Loss^l, S^{l+1})$ into the experience replay memory;
    Get a batch of samples from the replay memory, and calculate difference function $\Delta(\theta^l)$ according to (17);
    Calculate the gradient of $\Delta(\theta^l)$ with respect to θ^l according to (19);
    Update θ^l according to (20);
  End For
End For

VI. Numerical Results

In this section, we evaluate the performance of the proposed optimal transmission scheduling scheme. We consider a scenario where 3 RSUs are randomly located in a 1000-meter unidirectional road. The generated data of the vehicles running on the road is classified into 5 types. The amount of each type of data is randomly distributed in the interval (10, 30), and the transmission delay constraints for these data types are 1 to 5, respectively. The speed of the running vehicles is 100 km/h. Transmission costs c_v and c_c are 1 and 0.5, respectively. Transmission power P_I and P_V are set as 33 dbm and 30 dbm, respectively [29].

Fig. 3 shows the average transmission loss of the system in a time frame with different scheduling schemes. It is clear that our proposed deep-Q learning scheme has the lowest loss compared to the results of the other two schemes, especially

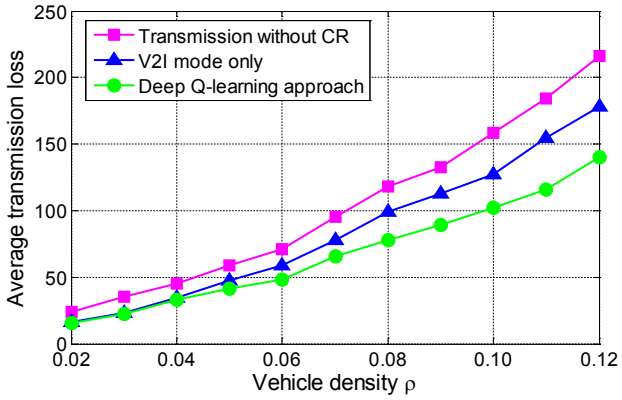


Fig. 3. Average transmission loss with different schemes.

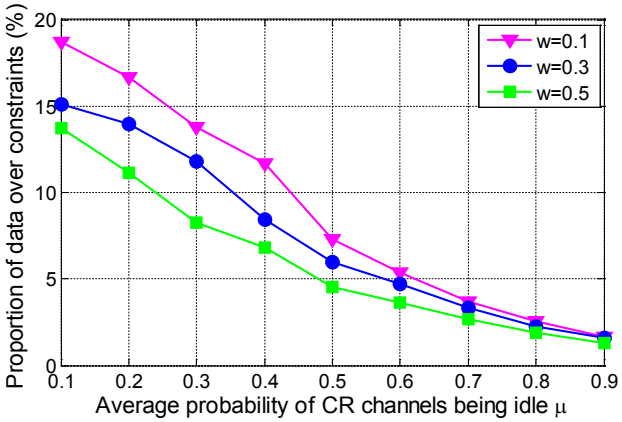


Fig. 4. Average proportion of data not meeting delay constraints with various CR channel availability.

with high vehicle density. The reason is that our scheme incorporates various transmission modes while fully exploiting CR spectrum resources that do not only belong to the local road region but also parts of the remote regions through joint V2V and V2I transmissions. In contrast to the learning scheme, when vehicles only take V2I mode, all available spectrum for transmission are limited to the resources belonging to a road region that the vehicles are currently located in. For a road region with poor CR channel availability but high vehicle density, although the vehicles in this region can utilize the local spectrum efficiently, constrained by the total amount of local available resources, high proportion of data transmission not meeting delay constrains occurs leading to high loss. When the vehicles adopt the transmission scheme without CR, they suffer the highest loss among the three schemes. As CR spectrum helps transmit data with low costs, not utilizing the CR resources increases the possibility of transmission overtime and greatly increases the system loss.

Fig. 4 illustrates average proportion of data that does not meet the delay constraints with various CR channel availability. We define the average probability of the CR channels being idle in road region m as $\mu_m = \lambda_m^{on}/(\lambda_m^{on} + \lambda_m^{off})$. In this simulation scenario, all the road regions have identical CR channel characteristic, i.e., $\mu_1 = \mu_2 = \mu_3 = \mu$. As μ increases, the proportion of data not meeting the delay constraints is

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2018.2872013, IEEE Internet of Things Journal

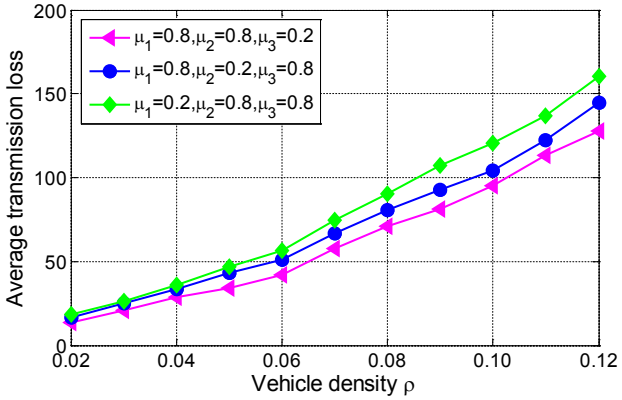
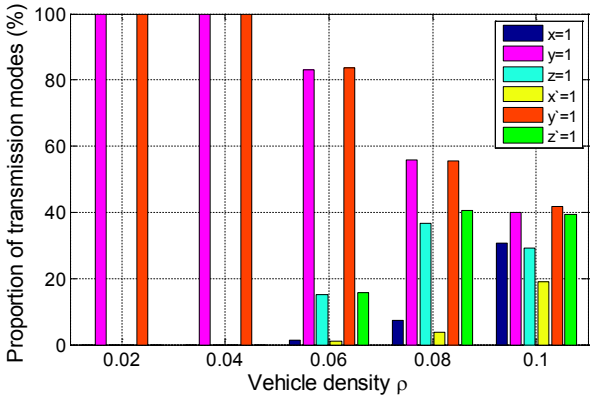
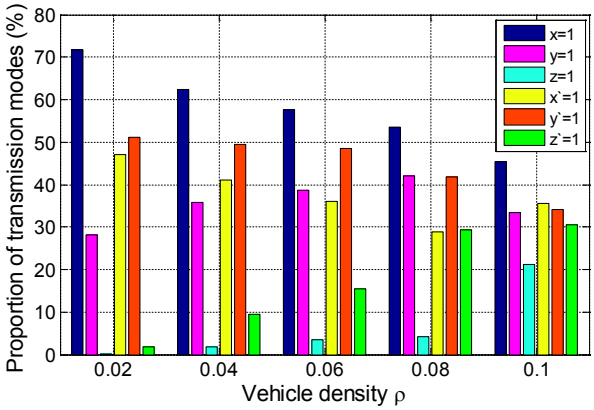


Fig. 5. Average transmission loss with different CR channel availability in various road regions.



(a) Penalty w=1



(b) Penalty w=0.3

Fig. 6. Average proportion of selected transmission modes with different CR channel availability in various road regions.

reduced. Available CR channels alleviate the spectrum scarcity of the vehicular network, and help transmit data with fast rates and low time delay. Thus, μ with higher value indicates that more CR spectrum resources can be utilized by vehicular data transmission, and less number of data transmissions exceed delay constraints. Furthermore, from this figure we can find that the value of penalty w greatly affects the data proportion, especially when μ is small. Higher penalty means more costs need to be paid for data transmission exceeding delay constraints, and it has a stronger promotion effect for the vehicles to choose transmission actions that have higher possibilities to ensure the delay constraints. In contrast, when w is small, to reduce the total loss value, vehicles are inclined to exploit the low-cost CR spectrum resources, although data transmission may be interrupted by the arrival of primary users.

Fig. 5 presents the effects of the CR characteristics in different road regions on the average system loss in a time frame. Compared to the case when the regions with high possibility of idle CR channels are located near the starting point of the road, the case when these regions are located at the farther end of the road has higher average loss. The reason is that in the former case, the available CR channels belonging to the starting regions can be utilized directly for data transmission by the vehicles without or with few V2V relays. However, in the latter case, the starting region has poor CR spectrum resources. Spectrum scarcity occurs in this region especially with high vehicle density. Consequently, to exploit more available channels, parts of the vehicles may deliver data to the remote regions with rich CR spectrum resources through multi-hop V2V transmission. High cost is brought by V2V data delivery, and thus the average loss is increased.

Fig. 6 shows the comparison of selected transmission modes of the vehicles located in the first road region with different CR channel availability in various regions. Here sets $\{x, y, z\}$ and $\{x', y', z'\}$ indicate the selected transmission actions with CR channel availability $\mu_1 = 0.8, \mu_2 = 0.2, \mu_3 = 0.8$ and $\mu_1 = 0.2, \mu_2 = 0.8, \mu_3 = 0.8$, respectively. Fig. 6(a) presents the performance of the case when the vehicles transmit urgent data with strict delay constraints. In this case, to promote vehicular transmission under the delay constraints, the value of penalty w is high. In order to avoid being subject to timeout penalty, the vehicles preferentially select licensed channels to transmit data. CR V2I transmission only helps alleviate congested licensed channels when vehicle density is high. In contrast, for transmitting data with relatively loose delay constraints, such as entertainment and map update data, timeout penalty w has low value and transmission cost turns to be a primary consideration. We show the proportion of selected transmission modes of this case in Fig. 6(b). With the increase of vehicle density from 0.02 to 0.08, the proportion of the vehicles that choose V2I mode with CR channels decreases. Although vehicles prefer to use CR spectrum with low cost, when there is a large amount of vehicles running on the road, the vehicles should turn to utilize licensed channels for alleviating spectrum scarcity. However, there are significant differences in the proportion of V2I transmission through license channels, between the scenarios with different CR channel availability. The value of y raises up while y' hardly changes as ρ increases from 0.02 to 0.06. The reason is that when the second region has poor CR available spectrum resources, in order to deal with a large number of vehicle transmission requirements, the vehicles in the first region have to take action $y = 1$ for delivering data. However, when the available CR resources in the second region are more than that in the first region, with the increase of vehicle density, a large

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2018.2872013, IEEE Internet of Things Journal

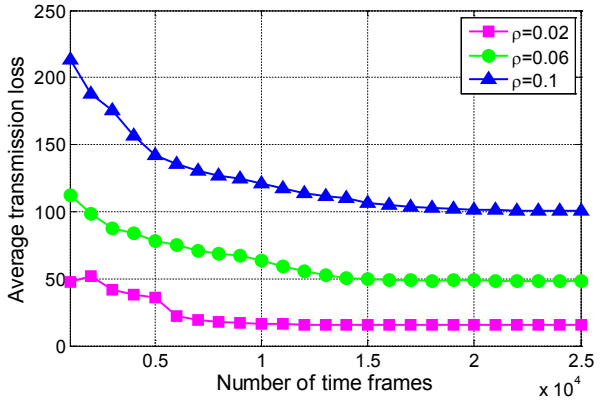


Fig. 7. The convergence of the proposed deep Q-learning based transmission scheme.

part of newly added data can be transmitted from the first region to the second one by V2V transmission, and then to be delivered to an RSU through low-cost CR channels. Thus, the proportion of action $y' = 1$ changes slightly as ρ increases.

Fig. 7 presents the convergence of our proposed deep Q-learning based transmission scheme. The learning process takes about 13000 to 20000 time frames to obtain the optimal transmission strategies with different vehicle density ρ. For these scenarios, the process of the proposed deep Q-learning scheme took about 25 minutes on average to execute on a computer with an i7 CPU and 8 GB RAM.

In this section, numerical results show that our proposed deep Q-learning based data transmission scheme outperforms conventional scheduling strategies in terms of reduced average transmission loss especially with high vehicle density. In addition, we investigate the issues that may affect the performance of our scheme. We find that the increase of timeout penalty incentivizes the vehicles to transmit data in licensed channels instead of CR channels for keeping delay constraints, especially in the scenario with poor CR channel availability. Furthermore, the effects of CR characteristics in different road regions on the average transmission loss and transmission modes selection are also demonstrated.

VII. Conclusion

In this paper, we have investigated data transmission in a CR-enabled vehicular network. We formulate an MDP model to analyze the performance of CR-based vehicular communications, where the characteristics of the CR channels in different road regions, the mobility of vehicles as well as the QoS requirements of data transmission are taken into account. To minimize transmission costs while ensuring data delay constraints, we obtain optimal transmission scheduling strategies in an efficient deep-Q learning approach, which fully exploits various spectrum resources and the benefits from proper transmission mode selection. Analytical results illustrate that the proposed scheme for vehicular data transmission efficiently reduces transmission costs and helps data delivery under delay constraints.

Reliable and delay-constrained data delivery plays an important role in the implementation of ITS. However, how to effectively utilize communication, caching and computing resources as well as various vehicular transmission modes for data delivery in the context of CR network is still a fundamental but unexplored question. In addition, traffic safety related information always requires real-time transmission. The way to cater for the emergency data delivery through incorporating road traffic state prediction and CR-enabled resource management requires future study.

Acknowledgment

Work in this article was supported by fundamental research funds for the central universities, China, under Grant No. 2672018ZYGX2018J001, the National Natural Science Foundation of China under Grants No. 61374189 and 61772195, the joint fund of the Ministry of Education of China and China Mobile under Grant No. MCM 20160304, and the Natural Science Foundation of Hunan Province under Grants No. 2018JJ2156.

References

[1] L. Cai, J. Pan, L. Zhao and X. Shen, "Networked electric vehicles for green intelligent transportation," *IEEE Communications Standards Magazine*, vol. 1, no. 2, pp. 77-83, July, 2017.

[2] S. Shah, E. Ahmed, M. Imran and S. Zeadally, "5G for vehicular communications," *IEEE Communications Magazine*, vol. 56, no. 1, pp. 111-117, Jan. 2018.

[3] J. Ren, Y. Zhang, R. Deng, N. Zhang, D. Zhang and X. Shen, "Joint channel access and sampling rate control in energy harvesting cognitive radio sensor networks," *IEEE Trans. Emerging Topics in Computing*, accepted.

[4] Z. Zhou, H. Yu, C. Xu, Y. Zhang, S. Mumtaz and J. Rodriguez, "Dependable content distribution in D2D-based cooperative vehicular networks: A big data-integrated coalition game approach," *IEEE Trans. Intelligent Transportation Systems*, vol. 19. no. 3, pp. 953-964, Jan. 2018.

[5] G. Qiao, S. Leng, K. Zhang and Y. He, "Collaborative task offloading in vehicular edge multi-access networks," *IEEE Communications Magazine*, vol. 56, no. 8, pp. 48-54, Aug. 2018.

[6] C. Wu, T. Yoshinaga, Y. Ji, T. Murase and Y. Zhang, "A reinforcement learning-based data storage scheme for vehicular Ad Hoc networks," *IEEE Trans. Vehicular Technology*, vol. 66, no. 7, pp. 6336-6348, July, 2017.

[7] R. Han, Y. Gao, C. Wu and D. Lu, "An effective multi-objective optimization algorithm for spectrum allocations in the cognitive-radio-based Internet of things," *IEEE Access*, vol. 6, pp. 12858-12867, Jan. 2018.

[8] P. Chen, S. Cheng and H. Hsu, "Analysis of information delivery dynamics in cognitive sensor networks using epidemic models," *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 2333-2342, Sep. 2017.

[9] T. Jiang, H. Wang, M. Daneshmand and D. Wu, "Cognitive radio-based smart grid traffic scheduling with binary exponential backoff," *IEEE Internet of Things Journal*, vol. 4, no. 6, pp. 2038-2046, Dec. 2017.

[10] H. A. B. Salameh, S. Almajali, M. Ayyash and H. Elgala, "Spectrum assignment in cognitive radio networks for internet-of-things delay-sensitive applications under jamming attacks," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 1903-1913, June 2018.

[11] B. Lyu , H. Guo, Z. Yang and G. Gui, "Throughput maximization for hybrid backscatter assisted cognitive wireless powered radio networks," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 2015-2024, June 2018.

[12] A. Paul, A. Daniel, A. Ahmad and S. Rho, "Cooperative cognitive intelligence for Internet of vehicles," *IEEE Systems Journal*, vol. 11, no. 3, pp. 1249-1258, Sept. 2017.

[13] Y. Han, E. Ekici, H. Kremo and O. Altintas, "Resource allocation algorithms supporting coexistence of cognitive vehicular and IEEE 802.22 networks," *IEEE Trans. Wireless Communications*, vol. 16, no. 2, pp. 1066-1079, Feb. 2017.

[14] M. Li, L. Zhao and H. Liang, "An SMDP-based prioritized channel allocation scheme in cognitive enabled vehicular Ad Hoc networks," *IEEE Trans. Vehicular Technology*, vol. 66, no. 9, pp. 7925-7933, Sept. 2017.

[15] H. Ghafoor and I. Koo, "CR-SDVN: A cognitive routing protocol for software-defined vehicular networks," *IEEE Sensors Journal*, vol. 18, no. 4, pp. 1761-1772, Feb. 2018.

[16] E. Hill and H. Sun, "Double threshold spectrum sensing methods in spectrum-scarce vehicular communications," *IEEE Trans. Industrial Informatics*, vol. 14, no. 9, pp. 4072-4080, Mar. 2018.

[17] Y. He, N. Zhao and H. Yin, "Integrated networking, caching, and computing for connected vehicles: A deep reinforcement learning approach," *IEEE Trans. Vehicular Technology*, Vol. 67, No. 1, pp. 44-55, Jan. 2018.

[18] F. Azmat, Y. Chen and N. Stocks, "Analysis of spectrum occupancy using machine learning algorithms," *IEEE Trans. Vehicular Technology*, vol. 65, no. 9, pp. 6853-6860, Sept. 2016.

[19] H. Cao and J. Cai, "Distributed opportunistic spectrum access in an unknown and dynamic environment: A stochastic learning approach," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 5, pp. 4454-4465, May, 2018.

[20] V. Raj, I. Dias, T. Tholeti and S. Kalyani, "Spectrum access in cognitive radio using a two-stage reinforcement learning approach," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 1, pp. 20-34, Feb. 2018.

[21] Y. Saleem, K. Yau, H. Mohamad, N. Ramli, M. Rehmani and Q. Ni, " Clustering and reinforcement-learning-based routing for cognitive radio networks," *IEEE Wireless Communications*, vol. 24, no. 4, pp. 146-151, Aug. 2017.

[22] J. Zhu, Y. Song, D. Jiang and H. Song, "A new deep-Q-learning-based transmission scheduling mechanism for the cognitive Internet of things," *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 2375-2385, Aug. 2018.

[23] R. Yu, W. Zhong, S. Xie, Y. Zhang and Y. Zhang, "QoS differential scheduling in cognitive-radio-based smart grid networks: An adaptive dynamic programming approach," *IEEE Trans. Neural Networks and Learning Systems*, vol. 27, no. 2, pp. 435-443, Feb. 2016.

[24] X. Guan, Y. Huang, M. Chen, H. Wu, T. Ohtsuki and Y. Zhang, "Exploiting interference for capacity improvement in software-defined vehicular networks," *IEEE Access*, vol. 5, pp. 10662-10673, June, 2017.

[25] O. Maillard, T. Mann and S. Mannor, "How hard is my MDP? The distribution-norm to the rescue," *Advances in Neural Information Processing Systems*, pp. 1835-1843, 2014.

[26] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra and M. Riedmiller, "Playing atari with deep reinforcement learning," *arXiv: Learning*, 2013.

[27] H. Hasselt, A. Guez and D. Silver, "Deep reinforcement learning with double Q-learning," in *Proc. National Conference on Artificial Intelligence*, pp. 2094-2100, 2016.

[28] Z. Ren, D. Dong, H. Li and C. Chen, "Self-paced prioritized curriculum learning with coverage penalty in deep reinforcement learning," *IEEE Trans. Neural Networks and Learning Systems*, vol. 29, no. 6, pp. 2216-2226, Feb. 2018.

[29] A. Bazzi, B. M. Masini, A. Zanella and I. Thibault, "On the performance of IEEE 802.11p and LTE-V2V for the cooperative awareness of connected vehicles," *IEEE Trans. Vehicular Technology*, vol. 66, no. 11, pp. 10419-10432, Sept. 2017.