Reinforcement Learning Based Power Control for VANET Broadcast Against Jamming

Canhuang Dai*, Xingyu Xiao*, Liang Xiao*, Peng Cheng†

* Dept. of Communication Engineering, Xiamen University, China. Email: lxiao@xmu.edu.cn

† State Key Lab. of Industrial Control Technology, Zhejiang University, China. Email: pcheng@iipc.zju.edu.cn

***Abstract*—Broadcast of critical information such as emergency traffic messages in vehicular ad hoc networks (VANETs) has to address jamming with dynamic network topology. In this paper, we propose a deep reinforcement learning based cooperative power control scheme for VANET broadcast against reactive jammers who can observe the ongoing broadcast states. The neural episodic control based cooperative power control scheme uses the convolutional neural network and differentiate neural dictionary to accelerate the learning speed for the VANETs with dynamic topology. Simulation results have shown that the proposed scheme can effectively improve the packet delivery rate and reduce the energy consumption of the broadcast compared with other power control schemes.**

***Index Terms*—VANETs, jamming, deep reinforcement learning, cooperative power control**

I. Introduction

Vehicular ad hoc networks (VANETs) can provide the broadcast of critical information such as accident warning signals and routing maps among onboard units (OBUs) and roadside units (RSUs) using protocols such as IEEE 802.11p [1]. However, the broadcast process is vulnerable to jammers who aim to block the ongoing broadcast messages and cause denial of service (DoS) attacks [2]. By applying smart radios such as universal software radio peripherals, a smart jammer chooses jamming power and the target channel according to the ongoing broadcast and is less likely to be detected by traditional jamming detection schemes.

Traditional anti-jamming communication technique such as frequency hopping in wireless networks is not applicable to VANETs implementing IEEE 802.11p. More specifically, the VANET messages are sent on one of the seven channels, each with 10 MHz bandwidth, containing a control channel, four service channels and two reserved channels [1]. The number of the frequency channels is insufficient to resist jamming with frequency hopping schemes [3].

Power control can help resist jamming attacks in wireless networks. For instance, the power control schemes as proposed in [4]–[6] that depend on the radio channel condition and the jamming power can improve the quality of the received message against jamming in VANETs. The power control for relay node with Q-learning (PCRQ) in [7] applies Q-learning for each node to independently determine the transmit power in a cognitive radio network to address the jamming attack from a smart jammer, which is similar to the scenario of ours. In this paper, we investigate cooperative power control for message broadcast in VANETs against reactive jammers who observe the ongoing broadcast to decide the jamming power. Compared with the independent power control in which each RSU selects the transmit power according to its own strategy, the cooperative power control scheme exploits the mobility and the relative distance of the OBUs and the jammer to resist the jamming signals.

The transmit power of the broadcast signals will impact on the future policy of the reactive jammer, and on the other hand, the VANET dynamically adjusts the transmit power according to the historic jamming power. Thus, the broadcast power control process can be viewed as a finite Markov decision process (MDP). Therefore, reinforcement learning (RL) techniques such as Q-learning can be applied for a VANET to derive the optimal transmit power strategy with probability 1 after a long broadcast process and thus improve the anti-jamming broadcast efficiency via trial-and-errors without being aware of the current VANET model and the jamming model.

The RL-based VANET power control has to address the dynamic network topology with a large number of radio nodes and the radio propagation degradation such as fast fading and multi-path, which results in a high dimension state space and thus slow the learning speed. For example, Q-learning, a model free and a widely used algorithm suffers from the curse of dimensionality and even fails to achieve the optimal strategy in the decision process with a large number of feasible states [8]. Therefore, we propose a neural episodic control (NEC) based cooperative power control scheme for VANET broadcast, which uses the convolutional neural network (CNN) and differentiate neural dictionary (DND) to accelerate the learning process [9]. By applying the deep reinforcement learning technique, this scheme compresses the state space observed by the source node in the broadcast

This work was supported in part by the National Natural Science Foundation of China under Grant 61671396, 91638204 and 61761136012 and in part by the Open Research Project of the State Key Laboratory of Industrial Control Technology, Zhejiang University under Grant ICT180036.

and thus reduces the convergence time required to achieve the optimal policy. Simulation results show this scheme can significantly improve the broadcast efficiency, reduce the overall energy consumption and improve the packet delivery rate (PDR) in the VANET broadcast.

The main contributions of this work are summarized as follows:

- We propose a cooperative power control broadcast scheme to exploit the topology and mobility of the VANET to improve the anti-jamming broadcast performance and apply the NEC algorithm for the server to determine the transmit power of each RSU without being aware of the jamming model and the network model.
- Simulations are performed to evaluate the broadcast performance, showing that the proposed NEC-based power control scheme can improve the broadcast efficiency by improving PDR and reducing energy consumption.

The reminder of this paper is organized as follows. We first review the related work in Section II. We formulate the system model in Section III and propose the NEC-based cooperative power control scheme in Section IV. We present simulation results in Section V and conclude in Section VI.

II. Related work

Power control has been widely investigated in VANET. For instance, an anti-jamming reinforcement system as proposed in [4] tunes the parameters of rate adaptation and power to improve the network throughput in the presence of jammers in VANET. The transmit power adaption mechanism as developed in [5] dynamically adapts each vehicle's transmit power according to the fast changing conditions, network load and link quantities of upper-layeres based on the estimation at the physical layer and the feedback from an adaptive beaconing system to improve the communication efficiency. The interference resistance scheme as proposed in [6] leverages recurring interferences from an attacker by randomly selecting the transmit power according to a given probability distribution to improve the awareness quality and reduce the channel congestion. The dynamic transmission power and contention window size adaption scheme as proposed in [10] exploits the priorities of different messages and different traffic densities and uses a joint approach to adapt transmit power and quality-of-service parameters to improve the overall throughput and reduce the broadcast delay in the VANET. A UAV-aided VANET transmission scheme as proposed in [11] employs unmanned aerial vehicles to relay messages of an OBU to RSUs to avoid jamming attacks and decrease the bit-error-rate.

Reinforcement learning techniques have been widely applied to study anti-jamming communications in common wireless networks. For example, a reinforcement learning based power allocation scheme as presented in [12] employs the hotbooting dyna Q-learning algorithm for power allocation to address the smart jamming attack without being aware of the jamming model and radio channel model in a non-orthogonal multiple access communication system. The learning based channel access strategy as proposed in [13] uses Q-learning for channel selection in a competing mobile network game to improve the anti-jamming capability of users in present of a jammer. The two-dimensional anti-jamming communication scheme as proposed in [3] exploits the spread spectrum and user mobility and applies the deep Q-network for a secondary user to decide the direction to escape from a heavy jammed area or choose a frequency hopping pattern to defeat a smart jammer. The minimax-Q learning based channel selection strategy in [14] observes the spectrum availability, channel quality and the attackers' actions and adapts its strategy on staying or switching between control and data channels to achieve higher throughput in a time-varying spectrum environment.

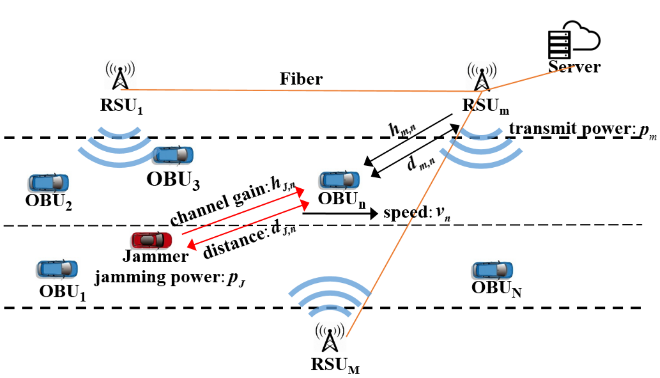


Fig. 1. Broadcast of a message to N OBUs via M RSUs against a jammer in a VANET with dynamically changing topology.

III. System model

We consider a VANET with M RSUs and N OBUs on the highway environment where OBUs and RSUs communicate in open and straight space, as shown in Fig. 1. A two-ray path loss model proposed in [15] is used to model the transmission loss in the highway scenario, which involves the wavelength λ, the distance d and the height of transmitter and receiver, h_t and h_r, and accordingly evaluates the path loss by

$$H(d)(dB) = \begin{cases} 20\log(4\pi d/\lambda), & d \leq 4\pi h_t h_r/\lambda, \\ 20\log\left(d^2/(h_t h_r)\right), & d > 4\pi h_t h_r/\lambda. \end{cases} \quad (1)$$

Note that the performance of the proposed cooperative power control scheme does not rely on the model of the path loss. In the VANET, RSU$_m$ downloads the messages with different emergency degrees from the server and broadcasts them to the neighboring OBUs on the road with the transmit power denoted by p_m, $1 \leq m \leq M$. The channel gain from RSU$_m$ to OBU$_n$ is denoted by $h_{m,n}$, $1 \leq m \leq M$, $1 \leq n \leq N$. Equipped with corresponding sensors, RSU$_m$ can measure the velocity and the distance of OBU$_n$, denoted by v_n and $d_{m,n}$, respectively.

The broadcast process can be interrupted by jamming attacks launched by malicious unit, which may hide in the VANET. In this work, we consider a reactive jammer that is

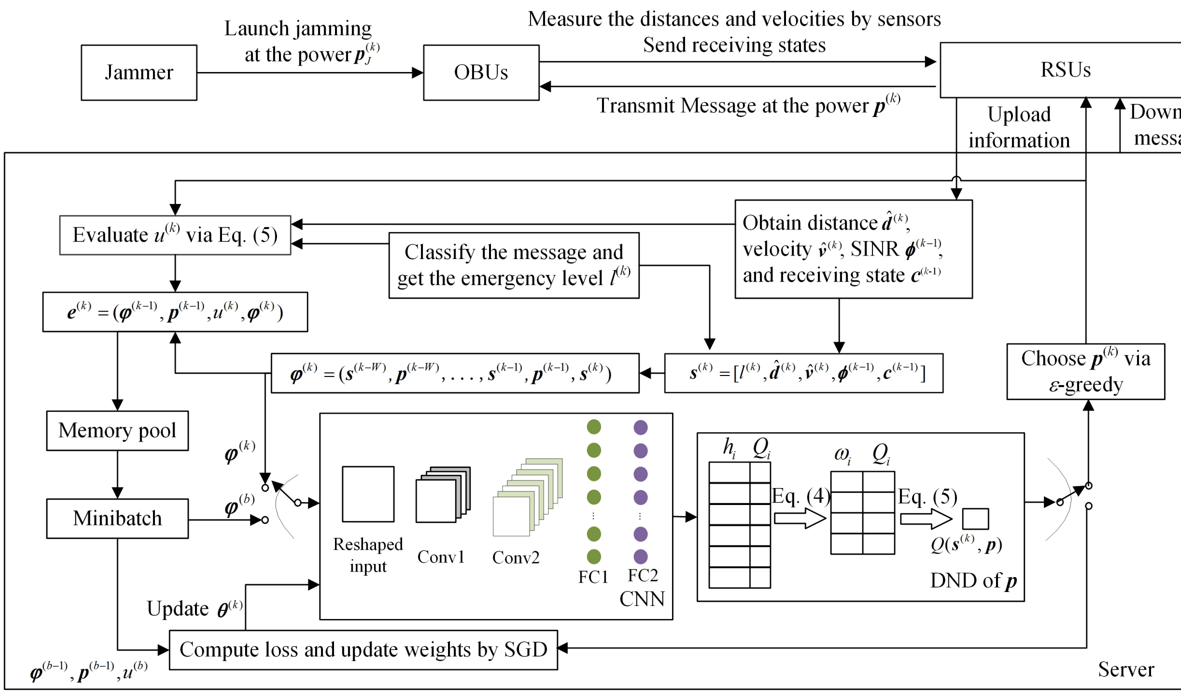


Fig. 2. Illustration of the NEC-based cooperative power control scheme in a VANET in the present of a jammer.

able to sense the ongoing transmission [16]. By applying universal software radio peripherals, the jammer can reactively adjust its jamming power denoted by p_J, according to the past transmit powers of the RSUs to avoid being detected by the intrusion detection strategies such as that in [17]. Similarly, the distance and the channel gain between the smart jammer and OBU$_n$ are denoted by $d_{J,n}$ and $h_{J,n}$, respectively.

The received signals at OBU$_n$ from other RSUs disturb the transmission from RSU$_m$ to OBU$_n$. Let σ_n be the environment noise and follow the normalized Gaussian distribution, i.e., $\sigma_n \sim \mathcal{N}(0, \delta_n)$. Therefore, the signal to interference plus noise ratio (SINR) of the received signal at OBU$_n$ from RSU$_m$ can be written as

$$\tau_{m,n} = \frac{p_m h_{m,n}}{\sum\limits_{1 \le i \ne m \le M} p_i h_{i,n} + p_J h_{J,n} + \sigma_n}. \tag{2}$$

OBU$_n$ can decode the message correctly if the SINR is higher than the threshold denoted by T_n, i.e., $\tau_{m,n} > T_n$. Let c_n represent the receiving state of OBU$_n$. More specifically, $c_n = 1$ indicates that the OBU$_n$ successfully decodes the broadcast message, and otherwise, $c_n = 0$. Each OBU informs the RSUs about its receiving state c_n and the SINR $\tau_{m,n}$ of last received signal via the periodic beacon frames [4].

The RSUs send the traffic information, the receiving state, the SINR of the signals received by the OBUs and the velocity and distance of each OBU to the server. If an OBU is close to the jammer but far away from the serving RSU, the RSU has to use a high transmit power to broadcast an urgent message. Nevertheless, if the message can tolerant more latency or the RSU cannot exceed the jamming signal, the servering RSU uses low transmit power until the OBU approaches another RSU. That is because the first RSU uses low transmit power to induce the jammer to reduce its jamming power and thus helps the other RSU to resist jamming signal for the approaching OBU. The emergency degree of a message is denoted by l, with $1 \le l \le L$, where a message with higher emergency degree such as an accident alert has to be broadcast to the OBUs more quickly.

IV. NEC-based Cooperative Power Control Scheme

The interactions between the RSUs and the jammer and OBUs make the broadcast process a finite MDP. Thus, a reinforcement learning algorithm can be used to derive an optimal cooperative transmit power selection policy without being aware of the jamming model and channel model. Note that the dynamic topology and complex channel model are hard to estimate accurately in VANETs, which results in an enormous state space, making the normal RL algorithms fall into the "curse of dimensionality" [8]. To address these problems, we apply the NEC algorithm, a deep reinforcement learning algorithm, for the server to accelerate the learning speed in this section.

As shown in Fig. 2, the NEC algorithm consists of a CNN and a DND [9]. At each time slot k, the RSUs measure the distances and velocities of the OBUs, denoted by $\hat{\boldsymbol{v}}^{(k)} = [\hat{v}_n^{(k)}]_{1 \le n \le N}$ and $\hat{\boldsymbol{d}}^{(k)} = [\hat{d}_{m,n}^{(k)}]_{1 \le n \le N, 1 \le m \le M}$, respectively. Besides, the RSUs can obtain the receiving states $\boldsymbol{c}^{(k-1)} = [c_n^{(k-1)}]_{1 \le n \le N}$ and SINRs $\boldsymbol{\phi}^{(k-1)} = [\tau_{m,n}^{(k-1)}]_{1 \le n \le N, 1 \le m \le M}$ of the OBUs from the feedback frames. All information is uploaded to the server and thus the server obtains the

TABLE I
SUMMARY OF SYMBOLS AND NOTATIONS

Symbol	Meaning
N	Number of OBUs in the VANET
M	Number of RSUs in the VANET
L	Number of emergency levels of the messages
$p_{m/J}$	Transmit power of RSU$_m$/jammer
$h_{m/J,n}$	Channel gain from RSU$_m$/jammer to OBU$_n$
σ_n	Received noise power at OBU$_n$
$\tau_{m,n}$	SINR of the signal from RSU$_m$ to OBU$_n$
u	Utility of the server
ξ	Power coefficient
v_n	Velocity of OBU$_n$
c_n	Receiving state of OBU$_n$
$d_{m,n}$	Distance between RSU$_m$ and OBU$_n$
$\boldsymbol{s}$	System state
$\mathcal{P}$	Set of available power selections
$\boldsymbol{\varphi}$	Input of the CNN
W	Number of state-action pairs in $\boldsymbol{\varphi}$
γ	Discount factor
h	The key value in the DND
$\boldsymbol{\theta}$	Weights of the CNN
$f_{1/2}$	Number of filters in the first/second Conv layer
$y_{1/2}$	Size of filters in the first/second Conv layer
z	Number of units in the first FC layer

Algorithm 1 NEC-based cooperative power control scheme

Initialize $\boldsymbol{\theta}^0$, $\boldsymbol{\varphi}^0, \boldsymbol{c}^0, \boldsymbol{\phi}^0$ and $\mathcal{E} = \emptyset$
for $k = 1, 2, \cdots$ **do**
  Measure the velocities $\hat{\boldsymbol{v}}^{(k)}$ and distances $\hat{\boldsymbol{d}}^{(k)}$ via sensors
  Estimate the emergency level $l^{(k)}$
  $\boldsymbol{s}^{(k)} = [l^{(k)}, \hat{\boldsymbol{v}}^{(k)}, \hat{\boldsymbol{d}}^{(k)}, \boldsymbol{\phi}^{(k-1)}, \boldsymbol{c}^{(k-1)}]$
  if $k \leq W$ **then**
    Select $\boldsymbol{p}^{(k)} \in \mathcal{P}$ at random
  else
    $\boldsymbol{\varphi}^{(k)} = \left(\boldsymbol{s}^{(k-W)}, \boldsymbol{p}^{(k-W)}, \cdots, \boldsymbol{p}^{(k-1)}, \boldsymbol{s}^{(k)}\right)$
    $\boldsymbol{e}^{(k)} = [\boldsymbol{\varphi}^{(k-1)}, \boldsymbol{p}^{(k-1)}, u^{(k-1)}, \boldsymbol{\varphi}^{(k)}]$
    Append $\boldsymbol{e}^{(k)}$ to $\mathcal{E}$
    Input $\boldsymbol{\varphi}^{(k)}$ to the CNN and get the output, $\boldsymbol{h}$
    Generate $\omega_i^{(k)}$ via (4)
    Calculate the Q values for $\boldsymbol{p} \in \mathcal{P}$ via (5)
    Append $\boldsymbol{h}$ and Q to K_p and V_p, respectively
    Select $\boldsymbol{p}^{(k)}$ via ε-greedy strategy
  end if
  Broadcast message with $\boldsymbol{p}^{(k)}$
  Obtain the receiving state $\boldsymbol{c}^{(k)}$ and the SINR $\boldsymbol{\phi}^{(k)}$ from the beacon frames
  Evaluate $u^{(k)}$ via Eq. (6)
  Update $\theta^{(k)}$ by minibatch gradient descent.
end for

environment state $\boldsymbol{s}^{(k)} = [l^{(k)}, \hat{\boldsymbol{v}}^{(k)}, \hat{\boldsymbol{d}}^{(k)}, \boldsymbol{\phi}^{(k-1)}, \boldsymbol{c}^{(k-1)}]$. The server selects the transmit powers $\boldsymbol{p}^{(k)} = [p_m^{(k)}]_{1 \leq m \leq M}$, $0 \leq p_m^{(k)} \leq P_{max}$, based on current policy and state $\boldsymbol{s}^{(k)}$. Let $\mathcal{P}$ denote the set of available power selections, i.e., $\boldsymbol{p}^{(k)} \in \mathcal{P}$.

We construct the input of the CNN as a sequence $\boldsymbol{\varphi}^{(k)}$ consisting of the current state and the previous W state-action pairs, i.e., $\boldsymbol{\varphi}^{(k)} = (\boldsymbol{s}^{(k-W)}, \boldsymbol{p}^{(k-W)}, \cdots, \boldsymbol{p}^{(k-1)}, \boldsymbol{s}^{(k)})$. The CNN is designed as a structure with two convolutional (Conv) layers and two fully connected (FC) layers. The first Conv layer contains f_1 filters with a size of $y_1 \times y_1$ and the second Conv layer contains f_2 filters with a size of $y_2 \times y_2$. The first FC layer maps the output of the second Conv layer to z rectified linear units and the second FC layer further maps them to $|\mathcal{P}|$ units. The weights of the CNN at time slot k are denoted by $\boldsymbol{\theta}^{(k)}$.

A DND is built for each given action $\boldsymbol{p}$, which is a memory module consisting of a dictionary structure (K_p, V_p). The output of the CNN, based on current sequence $\boldsymbol{\varphi}^{(k)}$, works as a key $\boldsymbol{h}$ and a Gaussian kernel function is applied to measure the distance between $\boldsymbol{h}$ and the key values in K_p, which is given by

$$k(\boldsymbol{h}, \boldsymbol{h}_i) = e^{-||\boldsymbol{h}-\boldsymbol{h}_i||_2^2/2}, \tag{3}$$

where $\boldsymbol{h}_i$ is the i-th key in K_p. Therefore, a weight for the i-th value in V_p, denoted by ω_i, is given by

$$\omega_i = k(\boldsymbol{h}, \boldsymbol{h}_i) / \sum_j k(\boldsymbol{h}, \boldsymbol{h}_j), \tag{4}$$

Note that the weights here are different from those in the CNN. Based on ω_i, we calculate the estimated Q value of each action on current state by

$$Q(\boldsymbol{s}^{(k)}, \boldsymbol{p}) = \sum_i \omega_i v_i, \tag{5}$$

where v_i is the i-th value in V_p. After the DND is queried, the new key $\boldsymbol{h}$ and Q value of each action is then appended to the end of K_p and V_p, respectively. Specially, if a key already exists in K_p, then the corresponding value is updated.

This power control algorithm uses the ε-greedy strategy to balance the exploitation and exploration in the learning process, the server selects the transmit power with the largest Q value with probability $1 - \varepsilon$ and another action with probability $\varepsilon/(|\mathcal{P}| - 1)$.

Each RSU transmits with the power $p_m^{(k)}$, receives the feedback information from the OBUs to extract the receiving states and SINR. The RSU also measures the distances and velocities of the OBUs in the area and uploads such information to the server. The server evaluates the utility $u^{(k)}$ based on the number of OBUs that successfully decode the message, the emergency level and the energy consumption by

$$u^{(k)} = \frac{\sum_{n=1}^{N} \left(c_n^{(k)} - c_n^{(k-1)}\right) l^{(k)}}{N - \sum_{n=1}^{N} c_n^{(k-1)}} - \xi \sum_{m=1}^{M} p_m^{(k)}, \tag{6}$$

where $\xi > 0$ is the power coefficient.

The server records the experience $\boldsymbol{e}^{(k)}$ given by $\boldsymbol{e}^{(k)} = [\boldsymbol{\varphi}^{(k)}, \boldsymbol{p}^{(k)}, u^{(k)}, \boldsymbol{\varphi}^{(k+1)}]$ in the memory pool $\mathcal{E} =$

$\{\boldsymbol{e}^{(1)}, \boldsymbol{e}^{(2)}, \cdots, \boldsymbol{e}^{(k)}\}$. The experience replay technique as presented in [18] is used to update the CNN weights based on the minibatch stochastic gradient descent and this process repeats B times, as shown in Algorithm 1.

V. Simulation results

Simulations were implemented to evaluate the performance of the NEC-based power control scheme for VANETs with a initial topology with 8 OBUs and 3 RSUs as shown in Fig. 3. The initial velocity vector of the OBUs was $\boldsymbol{v}^{(1)} = [90, 75, 75, 75, 70, 85, 85, 65, 80]$ km/h. Each OBU moved with the same direction at a velocity between 60 and 90 km/h, which randomly changed over time. In the simulations, we set the communication frequency as 5.9 GHz according to the protocol IEEE 802.11p and quantify the transmit power into 5 levels with the maximum transmit power $P_{max} = 500$mW. The jammer launched attacks according to the strategy proposed in [19]. We set $\lambda = 0.05$m according to the communication frequency and $h_t = h_r = 1.2$m for the path loss function.

In the NEC algorithm, we formulated the input sequence φ with 6 state-action pairs. The server selected the action with the largest Q value with the probability 0.9, i.e., $\varepsilon = 0.1$. We constructed the CNN with $f_1 = 20$, $f_2 = 40$, $y_1 = 8$, $y_2 = 4$ and $z = 1024$. The input sequence φ was reshaped into a matrix at the size of 19×19. The weights of the CNN are updated via minibatch gradient descend with 32 pieces of experience sampled from the memory pool at every time slot. We compared the performance of the proposed NEC-based cooperative power control scheme with the PCRQ algorithm in [7] in which each RSU selected the transmit power independently.

The simulation results show that the proposed NEC-based cooperative power control scheme achieves lower energy consumption, higher PDR and higher utility compared with the independent power control scheme PCRQ. For instance, as shown in Fig. 4(a), the PCRQ algorithm decreases the energy consumption from 907 mW to 803 mW at the 1000-th time slot, while the NEC-based cooperative power control scheme further decreases it by 13.1%. At the same time, as shown in

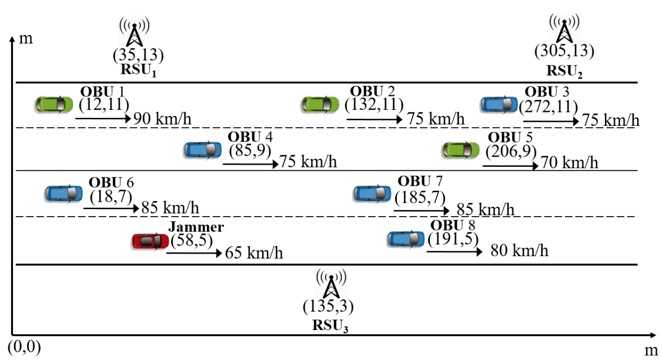
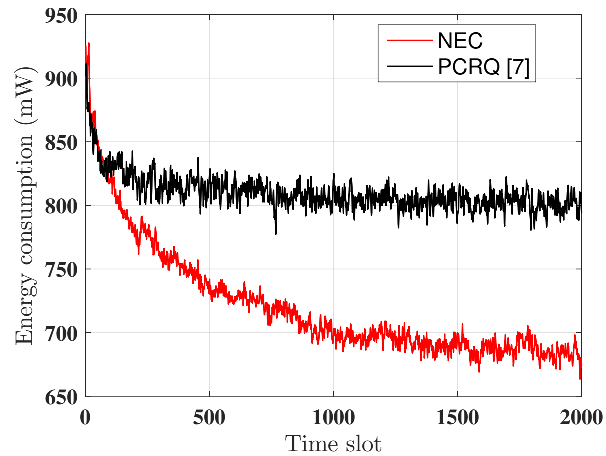
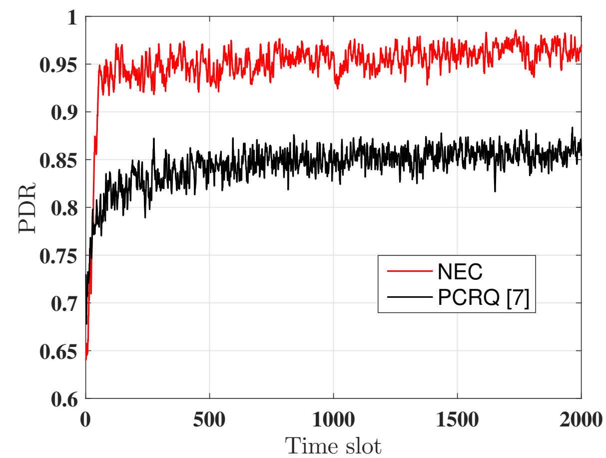


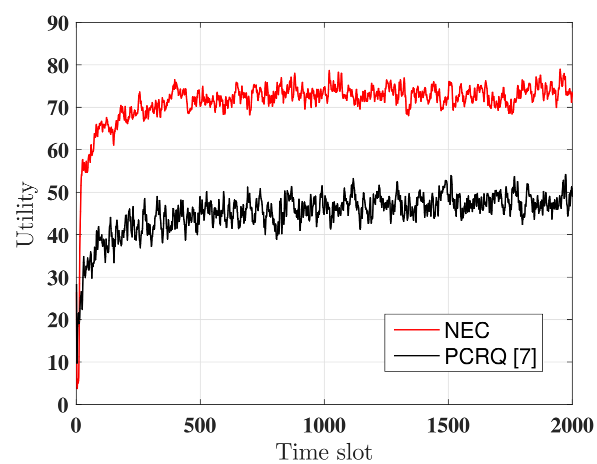
Fig. 3. Initial topology of the VANET in the simulations with 3 RSUs and 8 OBUs.



(a) Energy consumption

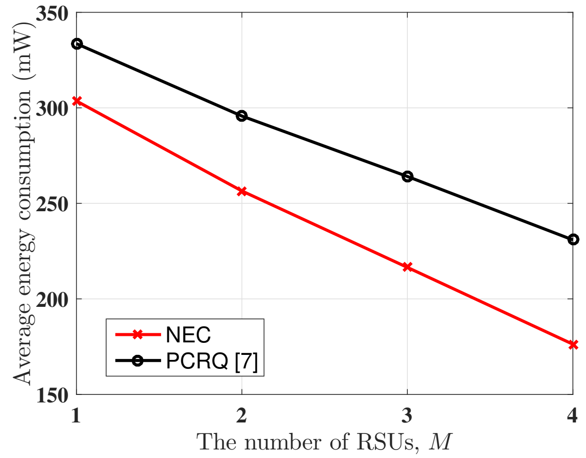


(b) Packet delivery rate

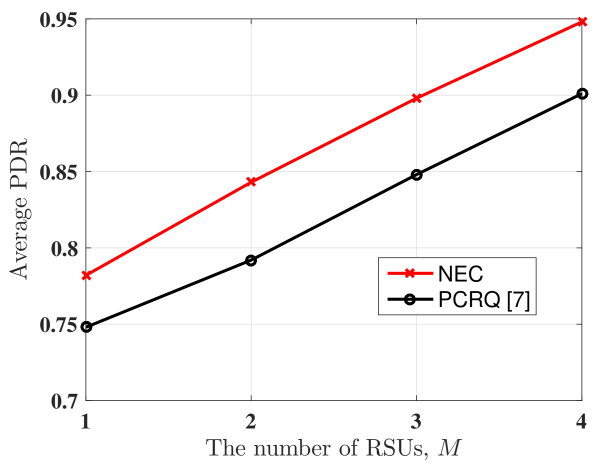


(c) Utility of the server

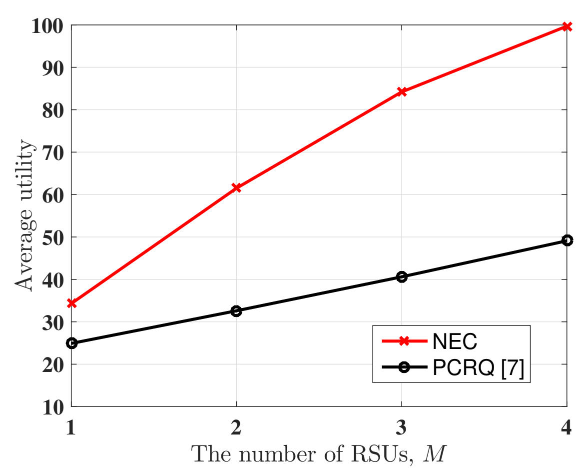
Fig. 4. Performance of the NEC-based cooperative power control algorithm for anti-jamming broadcast in a VANET with 3 RSUs, 8 OBUs and the maximum transmit power $P_{max} = 500$mW. The CNN inputs 6 state-action pairs and updates the weights with 32 pieces of experience at each time slot.



(a) Average energy consumption



(b) Average PDR



(c) Average Utility of the server

Fig. 5. Performance of the NEC-based cooperative power control algorithm for anti-jamming broadcast in a VANET versus the number of RSUs over 1000 time slots with 8 OBUs and the maximum transmit power $P_{max} =$ 500mW. The CNN inputs 6 state-action pairs and updates the weights with 32 pieces of experience at each time slot.

Fig. 4(b), the PCRQ algorithm increases the packets delivery rate from 65.4% to 85.2% and the NEC-based cooperative scheme further improves it by 12.0%. Besides, the PCRQ algorithm increases the utility from 9.3 to 47.8 and the NEC-based cooperative power control scheme further increases it by 54.2%, as shown in Fig. 4(c).

Fig. 5 shows that the average broadcast performance over 1000 time slots improves with the number of the RSUs in a specific area. For instance, if the number of RSUs increases from 1 to 4, the average energy consumption of the VANET with NEC-based cooperative power control scheme decreases by 41.8% and the packet delivery rate and utility increase by 23.3% and 1.9 times, respectively. If the number of RSUs is 4, the NEC-based cooperative power control scheme has 24.2% lower energy consumption, 5.3% higher PDR and 92.7% higher utility compared with the PCRQ algorithm.

VI. Conclusion

In this paper, we have proposed a NEC-based cooperative power control scheme for anti-jamming broadcast in VANET against smart jamming that enables a server to derive the optimal transmit power for each RSU without being aware of the VANET model and the jamming model. By applying the CNN and DND structure to abstractly map the VANET environment states to actions, this scheme can accelerate the learning speed and improve the broadcast performance. For instance, the energy consumption to broadcast messages to 8 OBUs in a VANET with 3 RSUs is reduced by 13.1% and the packet delivery rate and utility are increased by 12.0% and 54.2%, respectively, at the 1000-th time slot, compared with the PCRQ algorithm in which each RSU determines its transmit power independently.

References

[1] Y. Yao, L. Rao, X. Liu, and X. Zhou, "Delay analysis and study of IEEE 802.11 p based DSRC safety communication in a highway environment," in *Proc. IEEE Int. Conf. Computer Commun.*, pp. 1591–1599, Turin, Italy, Apr. 2013.

[2] M. S. Al-Kahtani, "Survey on security attacks in vehicular ad hoc networks (VANETs)," in *Proc. Int. Conf. Signal Processing and Commun. Systems*, pp. 1–9, Gold Coast, Australia, Dec. 2012.

[3] G. Han, L. Xiao, and H. V. Poor, "Two-dimensional anti-jamming communication based on deep reinforcement learning," in *IEEE Int. Conf. Acoustics, Speech and Signal Processing*, pp. 2087–2091, New Orleans, LA, Mar. 2017.

[4] K. Pelechrinis, I. Broustis, S. V. Krishnamurthy, and C. Gkantsidis, "A measurement-driven anti-jamming system for 802.11 networks," *IEEE/ACM Trans. Networking*, vol. 19, no. 4, pp. 1208–1222, Feb. 2011.

[5] M. Sander Frigau, "Cross-layer transmit power and beacon rate adaptation for VANETs," in *Proc. ACM Int. Symp. Design and Analysis of Intelligent Vehicular Networks and Applications*, pp. 129–136, Barcelona, Spain, Nov. 2013.

[6] B. Kloiber, J. Harri, and T. Strang, "Dice the TX power - improving awareness quality in VANETs by random transmit power selection," in *Proc. IEEE Vehicular Networking Conf.*, pp. 56–63, Seoul, South Korea, Nov. 2012.

[7] L. Xiao, Y. Li, J. Liu, *et al.*, "Power control with reinforcement learning in cooperative cognitive radio networks against jamming," *J. Supercomputing*, vol. 71, no. 9, pp. 3237–3257, Apr. 2015.

[8] V. Mnih, K. Kavukcuoglu, D. Silver, *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, Feb. 2015.

[9] A. Pritzel, B. Uria, and S. Srinivasan, "Neural episodic control," *arXiv preprint arXiv:1703.01988*, 2017.

[10] D. B. Rawat, D. C. Popescu, G. Yan, and S. Olariu, "Enhancing VANET performance by joint adaptation of transmission power and contention window size," *IEEE Trans. Parallel and Distributed Systems*, vol. 22, no. 9, pp. 1528–1535, Jan. 2011.

[11] L. Xiao, X. Lu, D. Xu, Y. Tang, L. Wang, and W. Zhuang, "UAV relay in VANETs against smart jamming with reinforcement learning," *IEEE Transactions on Vehicular Technology*, Jan. 2018.

[12] L. Xiao, Y. Li, C. Dai, H. Dai, and H. V. Poor, "Reinforcement learning-based NOMA power allocation in the presence of smart jamming," *IEEE Trans. Vehicular Technology*, vol. 67, no. 6, pp. 3377–3389, Apr. 2018.

[13] Y. Gwon, S. Dastangoo, C. Fossa, and H. Kung, "Competing mobile network game: Embracing antijamming and jamming strategies with reinforcement learning," in *Proc. IEEE Conf. Commun. Network Security*, pp. 28–36, Paris, France, July. 2013.

[14] B. Wang, Y. Wu, K. R. Liu, and T. C. Clancy, "An anti-jamming stochastic game for cognitive radio networks," *IEEE J. Sel. Areas in Commun.*, vol. 29, no. 4, pp. 877–889, Mar. 2011.

[15] C. Sommer and F. Dressler, "Using the right two-ray model a measurement based evaluation of PHY models in VANETs," in *Proc. ACM Int. Mobile Ad Hoc Networking and Computing*, pp. 1–3, Nevada, LV, Sep. 2011.

[16] X. Tang, P. Ren, Y. Wang, Q. Du, and L. Sun, "Securing wireless transmission against reactive jamming: A stackelberg game framework," in *Proc. IEEE Global Commun. Conf.*, pp. 1–6, San Diego, CA, Dec. 2015.

[17] H. Nguyen-Minh, A. Benslimane, and A. Rachedi, "Jamming detection on 802.11 p under multi-channel operation in vehicular networks," in *Proc. Int. Conf. Wireless and Mobile Computing, Networking and Commun.*, pp. 764–770, Shanghai, China, Sep. 2015.

[18] M. Li, T. Zhang, Y. Chen, and A. J. Smola, "Efficient mini-batch training for stochastic optimization," in *Proc. Int. Conf. Knowl. Disc. Data Mining*, pp. 661–670, New York, NY, Aug. 2014.

[19] Z. Yang, P. Cheng, and J. Chen, "Learning-based jamming attack against low-duty-cycle networks," *IEEE Trans. Dependable and Secure Computing*, vol. 14, pp. 650–663, Nov. 2015.