An Anti-jamming Game in VANET Platoon with Reinforcement Learning

Yexian Fan[1], Xingyu Xiao[2], Wei Feng[1]
1. Dept. Computer,Ningde Normal University, Ningde , China ;
2. Dept. Communication Engineering,Xiamen University, Xiamen, China；

***Abstract*-- This paper proposes a game strategy of anti-jamming power control-based Dyna-Q reinforcement learning to counter the smart jamming attacks suffered by the periodic beacons in Vehicular ad-hoc networks(VANETs) platoon. Simulation results indicate that without knowing the radio channel model of the jammer, the proposed algorithm can not only realize the vehicle maximize efficiency and improve the signal-to-interference-plus-noise(SINR) of radio channel but also attain better convergence effects compared with a Q-learning based scheme.**

INTRODUCTION

As a general mode of vehicle-to-vehicle (V2V) communication in VANETs, platoon can not only improve the road capacity and save energy but also improve the efficiency of traffic management[1]. In the platoon, IEEE802.11p is adopted to periodically update the beacon information about vehicle position and velocity in the dedicated channel. Thus the beacon is vulnerable to suffer jamming Denial-of-Service(DOS) attacks[2]. A smart jamming strategy presented in recent years can adjust its attack strategy in real time by learning the launch behavior of legitimate users, which makes it more difficult for VANETs to resist jamming[3].

GAME MODEL

As shown in Fig.1, there are N smart vehicles and an intelligent jammer, and the smart vehicles are equipped with sensing devices, whereas the jammer can be a moving or stationary device. The N-smart vehicles are organized into a close-range platoon mode. To counter the smart jamming in VANETs platoon, the interaction between vehicles and jammer is formulated as the game model based on SINR and load overhead.

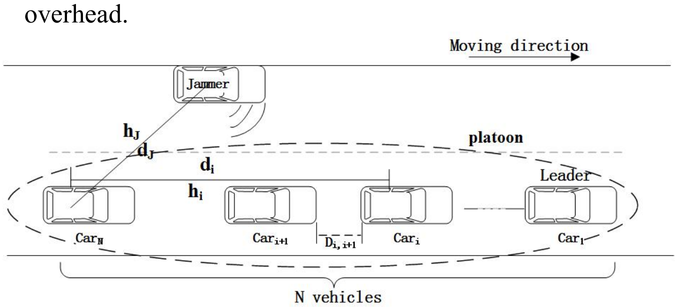

Fig.1 Network model

This research was supported by Scientific Research Fund of Fujian Provincial Education Department (JAT160539)

If CSMA/CA provides fair access to each vehicles, the time is divided into multiple slots and the vehicles are permitted to send beacons in slot *k*. In slot *k* vehicle *i* selectes the transmit power to obtain a higher SINR with minimum load overhead. The corresponding channel gain is h_i, and load overhead of unit power is C_i. The utility of the vehicle is defined as follows:

$$u_i(x_i, y) = \frac{h_i x_i}{\sigma + h_J y} - C_i x_i \tag{1}$$

The jamming power selected by jammer *J* in slot *k* is y , h_J denotes the corresponding channel gain, and C_J refers to the load overhead of unit power. The jammer aims to consume the energy of vehicle and block the ongoing transmission. The utility of it is defined as follows:

$$u_J(x_i, y) = -u_i - C_J y = -\frac{h_i x_i}{\sigma + h_J y} + C_i x_i - C_J y \tag{2}$$

THE ALGORITHMS

To obtain the optimal transmit power of dynamic jamming game in VANETs platoon, Dyna-Q learning [4] is introduced to control the power of VANETs and obtain the optimal utility by means of historic study.

Vehicle *i* can iteratively update the vehicle's value function $Q_i(s_i, x_i)$ as the learning deepens. $V_i(s_i^k)$ means that the vehicle node *i* in a certain state s_i^k can obtain the maximum $Q_i(s_i, x_i)$ by selecting the effective transmit power. $V_i(s_i^k)$ and $Q_i(s_i, x_i)$ can be separately updated by using the following formulas:

$$Q(s_i^k, x_i^k) = (1-\alpha) Q_i(s_i^k, x_i^k) + \alpha(u_i(x_i^k, y^k) + \delta V_i(s_i^{k+1})) \tag{3}$$

$$V_i(s_i^k) = \max_{x_i \in X} Q_i(s_i^k, x_i) \tag{4}$$

It also should build the real experience record for each action - state pair and formulate the environment model. The experience in every slot *k* includes $< s_i^k, x_i^k, u_i^k, s_i^{k+1} >$, which refers to the utility s_i^k attained by vehicle *i* in the state of s_i^k by selecting transmit power x_i^k , and then vehicle *i* transfers to the next state s_i^{k+1} .With accumulation of interaction experience, the environmental model can approximate the state transfer probability function $\Pi(s_i^k, x_i^k, s_i^{k+1})$ and utility function $\tau_i(s_i^k, x_i^k)$ of the real environment. The above two functions can be updated as follows:

$$\Pi(s_i^k, x_i^k, s_i^{k+1}) = \frac{H_i'(s_i^k, x_i^k, s_i^{k+1})}{H_i(s_i^k, x_i^k)} \quad (5)$$

$$\tau_i(s_i^k, x_i^k) = \frac{1}{H_i(s_i^k, x_i^k)} u_i + \frac{H_i(s_i^k, x_i^k) - 1}{H_i(s_i^k, x_i^k)} \tau_i(s_i^k, x_i^k) \quad (6)$$

$H_i(s_i^k, x_i^k)$ stands for the real number of transmit power x_i^k in state s_i^k, and the real number of state s_i^{k+1} is expressed as $H_i'(s_i^k, x_i^k, s_i^{k+1})$. The utility function is attained by calculating the mean approximation utility of the same action - state pair.

With the help of the existing environment model, vehicle node i can generate E simulation experiences and update function Q. The state-action pair $(\hat{s}_i^e, x_i^e)$ is selected randomly and the next state $\hat{s}_i^{e+1}$ is attained according to Π_i. Function Q should be updated based on the following two formulas:

$$Q_i(\hat{s}_i^e, x_i^e) = (1-\alpha) Q_i(\hat{s}_i^e, x_i^e) + \alpha(\tau_i(\hat{s}_i^e, x_i^e) + \delta V_i(\hat{s}_i^{e+1})) \quad (7)$$

$$V_i(\hat{s}_i^e) = \max_{x_i \in X} Q_i(\hat{s}_i^e, x_i) \quad (8)$$

Algorithm 1.Power control strategy based on Dyna-Q.

Initialized $Q, V, \alpha, \delta, E, \Pi, \tau_i$;
for k=1, 2, 3,…
  Detect the current state $s_i^k = [h_i^k, y^{i-1}]$ and selects transmit power x_i^k with ε ;
  Obtain new state s_i^{k+1} and utility u_i^k ;
  Update $Q_i(s_i^k, x_i^k)$ and $V_i(s_i^k)$ via (3) and (4);
  Record the real experience $< s_i^k, x_i^k, u_i^k, s_i^{k+1} >$, then update the state transfer probability $\Pi(s_i^k, x_i^k, s_i^{k+1})$ and utility $\tau_i(s_i^k, x_i^k)$ via(5) and (6);
  for e=1 to E do
    Select state-action pair $(\hat{s}_i^e, x_i^e)$ randomly;
    Got next state $\hat{s}_i^{e+1}$ and utility u_i^k ;
    Update Q of $(\hat{s}_i^e, x_i^e)$ via (7) and (8);
  End for
End for

SIMULATION RESULTS

To verify the effectiveness of the proposed anti-jamming game strategy, the paper also designs an anti-jamming strategy based on Q-learning, which mainly evaluates the changes in vehicle nodes utility and SINR of channels. If $c_i = 0.01$, $c_J = 0.01$ and $\sigma = 0.5 \times 10^{-9}$, the jammer selects its transmit power p_J according to the transmit power P_i of vehicle in the previous slot and the transmit power can be quantized as $p_J = p_i = \{1, 3, \cdots, 9\}$, $i = 1, 2, 3$. According to Fig.2, the convergence of the Dyna-Q strategy is superior to that of the Q-learning strategy. For example, the utility of a vehicle employing the Dyna-Q strategy in 200 slots is 0.1486, whereas the utility of the Q-learning strategy is 0.1316. During the dynamic game process, the vehicle nodes adjust the transmit power to obtain the optimal utility and reduce the jamming of the jamming node via continuous learning. As shown in Fig.3, with increasing simulation slots, SINR of vehicle is improved continually and converges to the stable value; thus, the efficiency of the vehicle node to send information is improved. Meanwhile, the convergence of the SINR under the Dyna-Q strategy is superior to that of the Q-learning strategy.

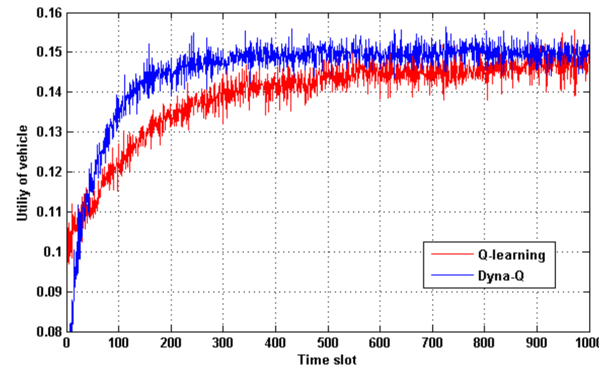

Fig. 2 Utility of vehicle

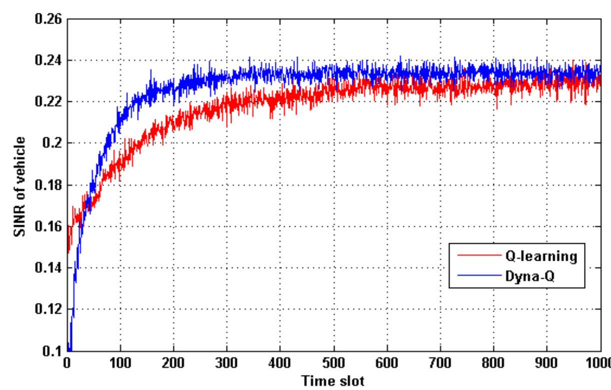

Fig. 3 SINR of vehicle

CONCLUSION

In this paper, a power control game strategy of VANETs platoon is proposed to counter the smart jamming. With the help of the Dyna-Q learning method, the power control strategy is equipped with better anti-jamming and convergence performance. The simulation results indicate that the algorithm can not only obtain the optimum utility but also improve the SINR of the channel effectively.

RERERRENCES

[1] Maiti Santa, Winter Stephan, Kulik Lars. A conceptualization of vehicle platoons and platoon operations. Transportation Research Part C,2017,:1-19

[2] Zeadally S, Hunt R, Chen Y S. Vehicular ad hoc networks (Vanets): status, results, and challenges. Telecommunication Systems, 2012, 50(4):217-241.

[3] L.Xiao, C. Xie, M. Min, and W. Zhuang, "User-centric view of unmanned aerial vehicle transmission against smart attacks," IEEE Transactions on Vehicular Technology, to appear. DOI: 10.1109/TVT.2017.2785414.

[4] Sutton R S. Dyna, an integrated architecture for learning, planning, and reacting[J]. ACM SIGART Bulletin, 1991, 2(4): 160-163.