DQN for Multi-layer Game Based Mining Competition in VEC Network

Yijing Li, Xuefei Zhang, Huici Wu, Junjie Liu, Dian Tang, Xiaofeng Tao
National Engineering Lab for Mobile Network Technologies
Beijing University of Posts and Telecommunications, Beijing, China
Institue of Sensing Technology and Business, BUPT (Wuxi)
Email: {liyijing, zhangxuefei, dailywu, JunjieLiu, spotty, taoxf}@bupt.edu.cn

***Abstract*—Blockchain has been considered as a promising technology to improve the efficiency of data sharing in Vehicular Edge Computing (VEC) by data mining among RSUs and vehicles. In order to obtain more mining reward, RSUs and vehicles will compete to accomplish the data sharing mining task. However, it is easier for RSUs equipped with stronger computing capabilities to win the mining reward. In this condition, some vehicles provide their own computing resources to form a shared resource pool for the competition with RSUs. Meanwhile, the competition among the vehicles sharing the resource pool is non-negligible since some selfish vehicles divide the group reward but without providing resources. In this way, we provide a multi-layer game model that involving a deformed N Iterated Prisoners Dilemma (NIPD) among vehicles and a bargain game between resource pool and RSUs. Further, we use multi-agent Deep Q Network (DQN) to achieve the equilibrium between vehicles and RSUs in mining. Finally, numerical results show the optimal strategy can attain a stable data sharing mining system in VEC network.**

***Index Terms*—VEC, DQN, NIPD, blockchain, game theory**

I. Introduction

Proposed by the European Industry Standards Group (ISG) Telecommunications Standards Institute (ETSI) in 2014, mobile edge computing (MEC) is regarded as an effective way to drive the transition of mobile broadband networks to the programmable world and meet the expected latency, scalability and automation of 5G [1] [2]. As one of the most popular application, vehicular edge computing (VEC) network in which vehicles, on-vehicle users and road side units (RSU) are capable of data storage and processing [3] [4]. VEC has attracted a lot of interests in the 5G era since vehicular network is one of the important Internet of Things (IoT) application scenario [5]. However, the great amount of data has been a great challenge for storage and data reliability as well [6].

Blockchain with the advantages of decentralization, reliability, transparent and secure is considered as a promising approach in dealing with data sharing and data storage tasks challenge in VEC networks [7] [8] [9]. Recently, blockchain mining for data sharing applied in VEC is showed to be capable of improving the reliability and efficiency in data sharing as well as alleviating traffic problems to some extent [10] [11] [12]. In such a VEC mining process, the mining task is set as data sharing task among the RSU and vehicles. In order to gain the mining reward, the RSU and vehicles are willing to share the traffic data they generate and collect. The more data they share, the more mining reward they will receive.

However, it leads to a great issue when the RSU and vehicles accomplish such a mining task. Mobile vehicles and the RSU has the same purpose of gaining the mining reward for themselves, but vehicles belong to different individual owners and RSU has its manufacturer. Therefore they will form a competitive relationship in mining with the purpose of maximizing their own benefit. However, in this competition, RSU has more powerful electric capacity and computing resources than vehicles that the huge gap of computing resources results in disequilibrium in mining system [13]. In order to enhance the competing capability of vehicles and gain more rewards, vehicles tends to cooperate and share their resources as a resource pool.

However, the vehicles are selfish individuals that they may not contribute all their resource or even more worse, some of them may betray the others to gain reward with little effort. This is a block withholding attack in mining [14] that vehicles only contribute partial effort or even do nothing but share all rewards the whole group wins. Therefore, there are both competitive and cooperative relationships between vehicles in shared resource pool.

Prisoner's dilemma and bargain game in game theory are good methods to formulate the relationship problem among vehicles and between the RSU and vehicles respectively. Generally, prisoners dilemma indicates a problem between two parties, but the problem in this paper with shared resource pool is a N-party game, the shared resource pool in this paper contains N vehicles which is a more complicated NIPD.

Existing studies have provided many ideas in dealing with NIPD and bargain game. [15] proposed a leader follower based control solution by allocating more status and voting rights for contributed participants which will attract the imitation of others. [16] Changes the intensity of interaction based on individual behavior and study the repeated prisoners dilemma game in social networks. It also highlights the reputation of nodes for the purpose of good interaction and uses a lattice arrangement for propagation. Similarly, lattice arrangement is also applied in [17].

However, most of the researches consider NIPD problem by

978-1-7281-3316-4/19/$31.00 ©2019 IEEE

literately choose two random players of N and repeat two-party prisoners dilemma for times which is a suboptimal solution. And most of the works mentioned above assumed that players always interact with all neighbors with sufficient interaction during the evolution phase network [15]- [19]. Motivated by the observations, we provide a multi-layer game based model including NIPD and bargain game in VEC. Different from other researches on mining dilemma, we fully consider the characteristics in VEC that the energy of vehicles are limited, and form an NIPD among vehicles with energy constraints. And multi-agent DQN is applied to find an optimal strategy in NIPD and bargain game from the perspective of N players at the same time rather than traditionally repeating two-party prisoner's dilemma. This multi-agent DQN algorithm can achieve the equilibrium for both NIPD and the bargain game together.

The remainder of this paper is organized as follows. In section II, we explain the system model under our scenario in detail. The comprehensive problem formulation description is given in section III. We discuss the numerical results in Section IV and conclude concludes this paper Section V.

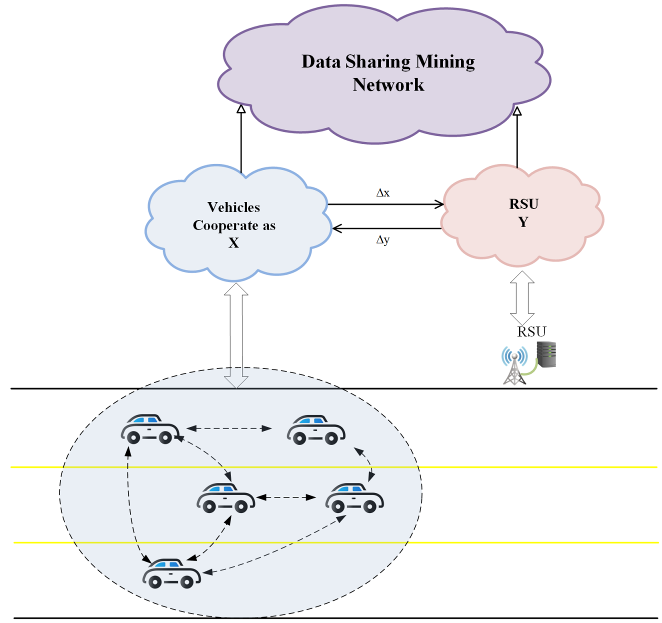

Fig. 1. Structure of the two-layer VEC mining network.

II. SYSTEM MODEL

As is seen in Fig.1, this section introduces the detailed data sharing mining task model in VEC network, in which there are a powerful RSU Y and N vehicles cooperate as a set X. Define m_x as the computing resource of each vehicles in X and m_y as the computing resource of RSU Y. X and Y compete in the mining network for the mining reward with their computing resources. Since vehicles in X have probability to betray the others or silence, we assume X has more computing resource than Y (1).

$$m_y < N \cdot m_x \tag{1}$$

There are both mining competition and cheat process between X and Y that X and Y both have the choice of $attack$ and $silence$ where $attack$ means one sacrifice part of resource from mining and use it to cheat another community for additional income and $silence$ means no cheating but only do its own mining computing. $\triangle x$ and $\triangle y$ denote the proportion of resources that X and Y sacrificed and used for cheating. M_X and M_Y denote the computing resources they contribute for mining and k is the number of vehicles in X choose $Cooperate$ (2) and (3). Constraint (4) limits the cheat resource cannot overcome the whole resources.

$$M_X = k \cdot m_x - \triangle x \tag{2}$$

$$M_Y = m_y - \triangle y \tag{3}$$

$$s.t. \quad \triangle x < k \cdot m_x, \triangle y < m_y \tag{4}$$

Assume X and Y have the same transition utility in mining resources that the mining reward is partitioned with the proportion they contribute mining resource and the total mining reward is 1. The mining income for X and Y can be expressed as R_X and R_Y in (5) and (6). And an utility function I_x and I_y is form to describe the total resource revenue conversion efficiency (7) and (8).

$$R_X = \frac{M_X}{M_X + M_Y} \tag{5}$$

$$R_Y = \frac{M_Y}{M_X + M_Y} \tag{6}$$

$$I_x = \frac{R_X + \triangle x \cdot I_y}{M_X} \tag{7}$$

$$I_y = \frac{R_Y + \triangle y \cdot I_x}{M_Y} \tag{8}$$

And for vehicles in X, they have both cooperative and competition relationships as introduced in section I. We form an electric constraint NIPD problem with a Triad model as

$$Q_X = \{P, A, f_i(a_i|k)|i \in P, a_i \in A\} \tag{9}$$

where $P = 1, 2, ..., N$ is the number of vehicles in resource pool and $A \in \{C, D, S\}$ represents for the action set. An vehicle can choose $Cooperate$ to contribute all its resource for group interest, $Detect$ to contribute none of its effort but pretend as a "good" member for its own profit. Different from traditional NIPD, here a new action $Silence$ is added because vehicles suffer from electric limitation while this is not a problem before because players in normal NIPD do not has such limitation. $f_i(C|k)$, $f_i(D|K)$ and $f_i(S|k)$ indicate the income function of one vehicle under the situation that k vehicles among other N-1 vehicles in X choose $Cooperate$, $Detect$ and $Silence$ respectively. $f_i(C|k)$ and $f_i(D|K)$ will always increase as k increases from 1 to N-1 which means the more vehicles choose to cooperate, the more resource the pool will have and therefore the resource pool will gain more mining reward to allocation within vehicles. A simple income function setting is given in Table I where b is the coefficient defect reward. .

$$f_i(D|k) > f_i(C|k) \quad k \neq N-1 \tag{10}$$

$$f_i(C|N-1) > f_i(D|0) \tag{11}$$

$$f_i(S|k) \equiv 0 \tag{12}$$

Constraints are familiar with traditional prisoners dilemma (10)-(12). (10) represents that income for $Defect$ is always bigger than $Cooperate$ unless all members choose $Cooperate$. This is because vehicles choose $Cooperate$ and $Defect$ both are qualified to divide the group mining reward but $Defect$ costs no computing resource. (11) means that while all players choose $Cooperate$, they earn more than the situation of all $Defect$. It is easy to understand that if all vehicles in X choose $Defect$, X will not have any ability to compete with the RSU and will gain nothing. And vehicles choose $Silence$ will always gain nothing (12) which is a motivation for vehicles to take action, otherwise they will always go to the trend of all silence. A simple case of income that satisfy the constraints is illustrated in table 1

TABLE I
REWARD OF DIFFERENT ACTION AND COOPERATOR NUMBERS

Reward	**k cooperators in N-1 players**			
Action	***0***	***1***	**......**	***N-1***
0	***0***	***2***	**......**	***2(N-1)***
1	***b***	***2+b***	**......**	***2(N-1)+b***
2	***0***	***0***	**......**	***0***

III. Problem Fomulations

In this section, we describe the objective functions for each vehicles, the set X and Y, and introduce the multi-agent DQN algorithm to solve the problems.

In the bargain process of our proposed multi-layer game based model, the computing resources M_X contributed by X for mining is influenced by the number k of vehicles who choose $Cooperate$. Therefore, we analyze the NIPD game among vehicles first.

In the NIPD game, objective functions $P1$ for each vehicle i can be illustrated as finding the best action strategy $\Pi(a_i)^*$ under a certain number of k to maximize its income which can indicated as:

$$P1 : \Pi(a_i)^* = \max_{\Pi(a_i)} \{f_i(a_i|k)|i \in P, a_i \in A\} \tag{13}$$

$$s.t.(10) - (12)$$

And for the bargain game, problem comes as X and Y decide how many resources for mining and how many for cheating to maximum their resource transition utility I_x and I_y. With the basic expression of I_x and I_y in (7) and (8), we simplify the utility function as (14) and (15).

$$I_x = \frac{R_X \cdot M_Y + \triangle x \cdot R_Y}{M_X \cdot M_Y - \triangle x \cdot \triangle y} \tag{14}$$

$$I_y = \frac{R_Y \cdot M_X + \triangle y \cdot R_X}{M_X \cdot M_Y - \triangle x \cdot \triangle y} \tag{15}$$

The objective function of X and Y is to find a best strategy of $\{\triangle x, \triangle y\}$ can be indicated as:

$$P2 : \underset{\triangle x, \triangle y}{maximize}\, I_x(\triangle x, \triangle y) \tag{16}$$

$$P3 : \underset{\triangle x, \triangle y}{maximize}\, I_y(\triangle x, \triangle y) \tag{17}$$

$$s.t.(1) - (4)$$

The optimal strategy of $\{\triangle x, \triangle y\}$ exists with partial derivative constraints(18)-(19).

$$\frac{\partial^2 I_x}{\partial x^2} > 0 \quad \& \quad \frac{\partial^2 I_y}{\partial y^2} > 0 \tag{18}$$

$$\frac{\partial I_x}{\partial x} = 0 \quad \& \quad \frac{\partial I_y}{\partial y} = 0 \tag{19}$$

In order to find a global optimal solution, we combine the objective function $P1 - P3$ and apply a multi-agent DQN method for both bargain game and NIPD game to achieve the equilibrium. The final global objective function is to find a set combining action strategies of N vehicles and strategy for $\{\triangle x, \triangle y\}$ that maximize the income of each individuals as well as X and Y when reach an equilibrium:

$$P4 : \Pi^* = \max_{\Pi} V((a_1), ..., (a_N), \triangle x, \triangle y) \tag{20}$$

$$\Pi = \langle \Pi(a_1), ..., \Pi(a_N), \triangle x, \triangle y \rangle \tag{21}$$

$$s.t.(1) - (20)$$

The procedure of our algorithm can be seen in Fig.2 and Fig.3. First, we run a multi-agent DQN for NIPD game where vehicles in X are N players. Each of the players receives its observation OB_i^{t-1} from the $ENVIRONMENT_{t-1}$ which is constituted by the action of all players in $t-2$ slot. Then the player choose one random action $Action_i^{t-1}$ and gain a corresponding reward as a value function Q_i^{t-1}. These N actions in $t-1$ slot form the state of next slot $ENVIRONMENT_t$. Each player record the set $\{ENV_{t-1}, Action_i^{t-1}, Q_i^{t-1}, ENV_t\}$ in its own memory and repeat the learning process, until reach a optimum Q_value under strategy $\langle \Pi(a_1), ..., \Pi(a_N) \rangle$. Then the system will return $\langle \Pi(a_1), ..., \Pi(a_N) \rangle$ as the observation of X to the multi-agent DQN for bargain game which has the similar learning process. And we will repeat the whole two-layer learning process until achieve the objective function $P4$. Detailed algorithm flow is given in Algorithm 1. Different from normal DQN without interactions where only one agent is trained, the multi-agent DQN method applied in this paper can adapt the complicated NIPD environment where a agent gains its observations with the interactions of other $N-1$ agents. Research [20] indicates that multi-agent DQN can achieve better performance than normal DQN which can also be illustrated by the simulation in section IV.

Algorithm 1 Deep Q-learning with Experience Replay in Multi layer Game Theory

Input: number of players N, replay memory size, learning rate, reward decay, egreedy parameter
Output: Π^*, Q^*

Initialize replay memory M to capacity $MemorySize$
Initialize action-value function $Q_1, Q_2, ..., Q_N, Q_x, Q_y$ with random weights
for episode=1,2...E **do**
  Initialize N agents and sequence action $\Pi(a_1), ..., \Pi(a_N)$, initial state $STATE_1$ and preprocessed sequenced $\psi_1 = \psi[STATE_1]$
  for t=1,2...T **do**
    Initialize sequence action $\Pi_{a_i} \in (C, D, S)$, initial state $state_{i,1} = C_1, D_1, S_1$ prepocessed sequenced $\phi_1 = \phi(state_{i,1})$
    for players=1,2...N **do**
      Base on $state_{i,j-1}$, with probability ϵ select an action a_i as $\Pi(a_i)$
      Otherwise select $\Pi(a_i) = \max\limits_{\Pi(a_i)} Q^*(\phi(state_{i,1}), a_i; \theta)$
      Execute a_i in emulator and observe reward r_i^j and image $state_{i,j+1}$
      Store transition observation, action and reward in memory M
      Sample random minibatch of transition from memory M
      Set $V(a_i)_j = \begin{cases} V(a_i)_j, & for\,terminal\,\phi_{j+1} \\ V(a_i)_j + \gamma \max\limits_{\Pi(a_i)} Q(\phi_{j+1}, a_i; \theta) & \\ \quad for\,non-terminal\,\phi_{j+1} & \end{cases}$
      perform a gradient descent step on $(V(a_i)_j - Q(\phi_{j+1}, a_i; \theta))^2$
    end for
    Base on $\langle \Pi^*(a_1), ..., \Pi^*(a_N) \rangle$ initialize sequence action $\Pi_1 = \langle \Pi^*(a_1), ..., \Pi^*(a_N), x, y \rangle$, $x \in (0, 1)$ and initial state $STATE_1$ Base on $state_{j-1}$, with probability ϵ select an action (x_j, y_j) as Π_j
    Otherwise select $\Pi_j = \max\limits_{\Pi_j} Q^*(\psi(STATE_j), (x_j, y_j)); \theta)$
    Execute (x_j, y_j) in emulator and observe reward R_j^t and image $STATE_{j+1}$
    Store transition observation, action and reward in memory M
    Sample random minibatch of transition from memory M
    Set $V(\chi\prime)_j = \begin{cases} V(\chi\prime)_j, & for\,terminal\,\psi_{j+1} \\ V(\chi\prime)_j + \gamma \max\limits_{\Pi\prime} Q(\psi_{j+1}, (x_j, y_j); \theta) & \\ \quad for\,non-terminal\,\phi_{j+1} & \end{cases}$
    perform a gradient descent step on $(V(\chi\prime)_j - Q(\psi_{j+1}, (x_j, y_j); \theta))^2$
  end for
end for
return Π^*, Q^*

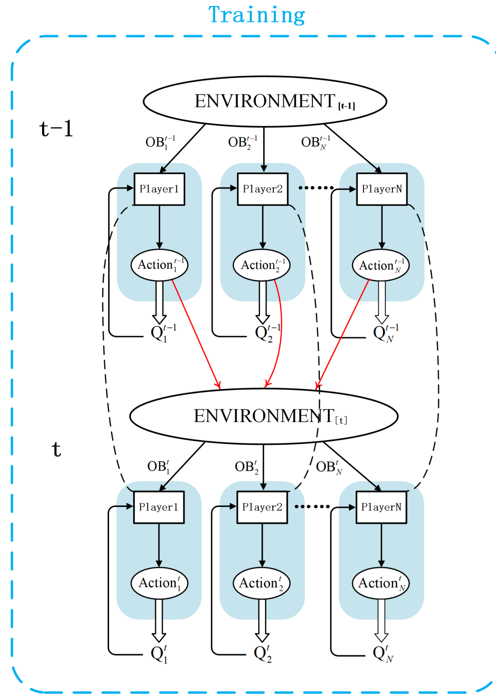

Fig. 2. Multi-agent DQN for NIPD.

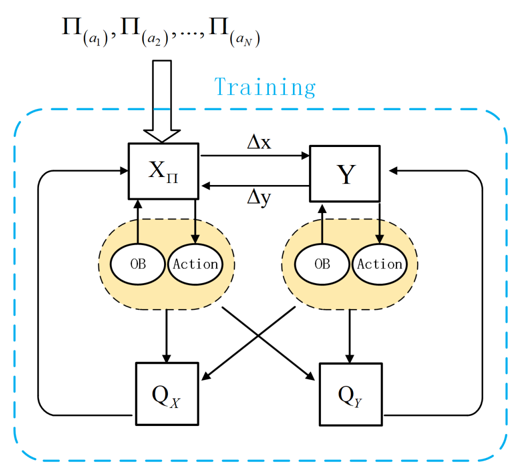

Fig. 3. Multi-agent DQN for bargain game.

IV. Numerical Results

To evaluate the performance of our proposed method, a python based simulation is developed. The simulation is set up to have 12 players in X. And for DQN process, parameters are set as learning rate=0.01, reward decay=0.9, e-greedy=0.9, replace-target-iter=200, and memory size=2000, epochs$\in (20000, 40000)$ according to different demands. Although we apply a global optimal solution in the multi-layer DQN for both games at the same time, we take them into two part to analyze for clearer observation in this section.

Fig.4 indicates the performance of $Loss$ and Q_value in terms of the decision epochs, where $Loss$ is the loss function value of Q network training and Q_value describes the reward feedback value of its action in one state. This shows that the more training epochs in DQN, the better performance of DQN will have. And the randomness of curve in Fig.4 as well as the figures below is caused by the fluctuation in learning states and somewhat the probability of $Silence$.

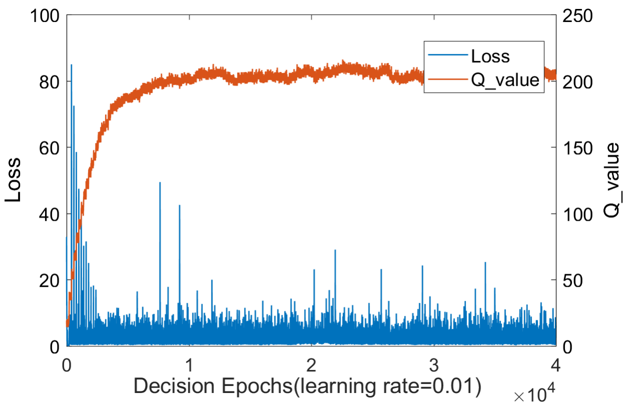

Fig. 4. Q value and loss in DQN process.

As is shown in Fig.5 cooperator numbers have different trends under different reward function coefficient b. In Fig.5, when the decision epoch is low (from 0-5000), players are willing to cooperate with each other for group profit while as decision epoch rises from 5000 to 7500, they find the fact that $Defect$ may get more individual income and therefore turn to the decision $Defect$. And in the learning process, DQN model with higher defect reward will find the fact earlier than those model with lower defect reward. This is the reason why cooperator numbers decay at different decision epochs. And with decision epoch continues rising from 7500 to 20000, players under cases "$b = 1$" and "$b = 3$" turn to choose $Cooperate$ again. This is because they weigh the pros and cons that with close reward of $Cooperate$ and $Defect$, $Cooperate$ can avoid the risk of all $Defect$ which leads to no gains for vehicles. On the contrary, under bigger gap of reward that "$b = 5$", players are more willing to take the risk mentioned above for higher benefit.

Fig.6 indicates the performance of $sumreward$ in terms of the decision epochs under different defect reward where sum reward is the total income of N players in NIPD. The curve of sum reward first increase from epoch 0-7500 and turns to a stable state where DQN model with high defect reward will reach the stable state with higher sum reward. The conclusion is easy to understand when combining Fig.5 and Fig.6, with higher gap between reward of $Cooperate$ and $Defect$, players will choose $Defect$ in high probability and the higher $Defect$ reward will obviously cause the higher sum reward.

And in Fig.7, a simulation of comparation of multi-agent DQN and one agent DQN is carried under the same reward coefficient "$b = 1$". The multi-agent algorithm is better than one agent DQN as shown in Fig.7 that under multi-agent DQN, players are more willing to $Cooperate$. This is because multi-agent DQN will do its training with comprehensive state of all

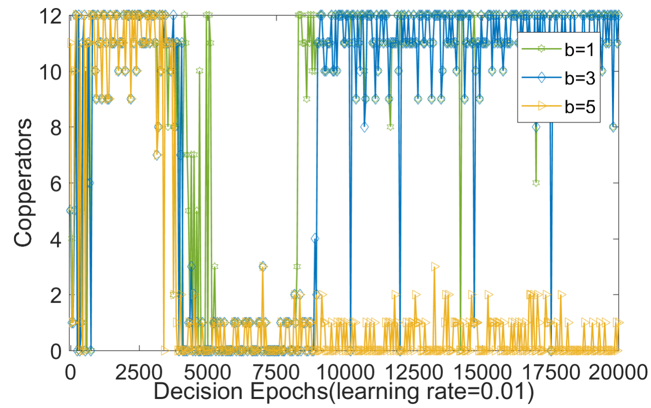

Fig. 5. Cooperator numbers in NIPD game with different a value for reward function.

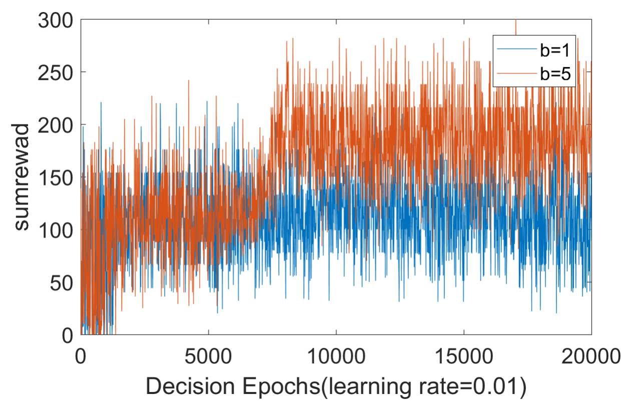

Fig. 6. Sum reward in NIPD game with different a value for reward function.

agents in network and get feedback as own observation rather than only observe from the perspective of one agent.

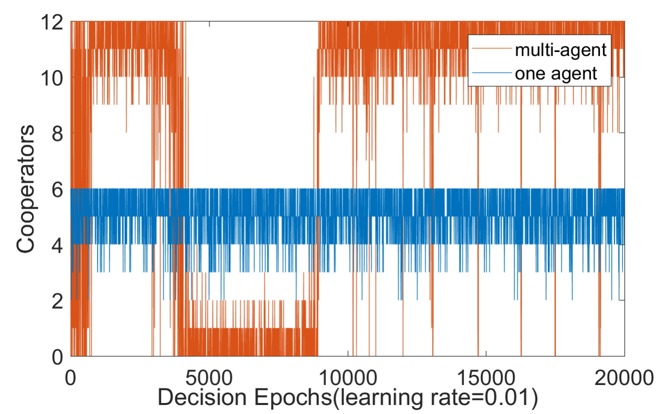

Fig. 7. Cooperator numbers in NIPD game under different agent.

After receiving the learning result of N agents for N players, the NIPD system reaches the stable state with a certain cooperator number. We regard the number as an input of X in the bargain game to analyze the utility discrepancy of I_x and I_y. As the simulation result illustrates in Fig.8.

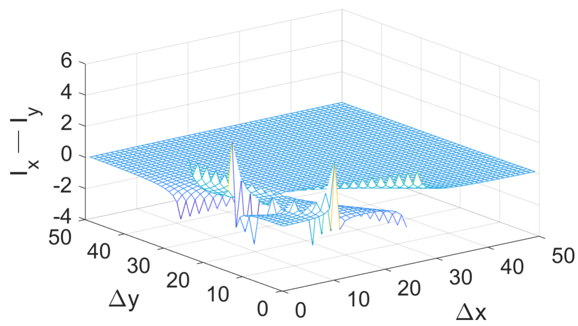

Fig. 8. Utility discrepancy of I_x and I_y in bargain game.

Axis x and y are the percentage of resource that X and Y sacrificed for cheating, axis z is the utility discrepancy of I_x and I_y. Positive peaks and negative peaks correspond to the circumstances where X and Y achieve their best utility functions respectively.

V. Conclusions

This paper introduces a multi-layer game based data sharing mining competition for encouraging data sharing in VEC network. RSU and vehicles are motivated to accomplish data sharing mining task for gaining mining reward. With comprehension of the computing resource gap of RSU and vehicles, NIPD for both competition and cooperation relationship among vehicles and bargain game between RSU and vehicles competition are considered. On this basis, we apply a multi-agent DQN for the multi-layer game.

Numerical results demonstrate that under this model, the multi-agent DQN method can obtain a global optimal solution for both NIPD game and the bargain game while achieving the equilibrium. Simulations prove the high learning effect and show the great advantages of multi-agent DQN when comparing with traditional learning model.

Acknowledgement

This work was supported in part by the National Science and Technology Major Project under Grant 2018ZX03001009, in part by the Fundamental Research Funds for the Central Universities, in part by the National Nature Science Foundation of China under Grant 61701037 and 61325006, and in part by the 111 Project of China B16006.

References

[1] Mobile-edge computingIntroductory technical white paper, White Paper, ETSI, Sophia Antipolis, France, Sep. 2014. [Online]

[2] Q. Cui et al., "Preserving Reliability of Heterogeneous Ultra-Dense Distributed Networks in Unlicensed Spectrum," in IEEE Communications Magazine, vol. 56, no. 6, pp. 72-78, June 2018.

[3] X. Hou, Y. Li, M. Chen, D. Wu, D. Jin and S. Chen, "Vehicular Fog Computing: A Viewpoint of Vehicles as the Infrastructures," in IEEE Transactions on Vehicular Technology, vol. 65, no. 6, pp. 3860-3873, June 2016.

[4] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari and M. Ayyash, "Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications," in IEEE Communications Surveys & Tutorials, vol. 17, no. 4, pp. 2347-2376, Fourthquarter 2015.

[5] W. Xu et al. "Internet of Vehicles in big data era" IEEE/CAA J. Automatica Sinica vol. 5 no. 1 pp. 19-35 Jan. 2018.

[6] A. R. Biswas and R. Giaffreda, "IoT and cloud convergence: Opportunities and challenges," 2014 IEEE World Forum on Internet of Things (WF-IoT), Seoul, 2014, pp. 375-376.

[7] L. Yue H. Junqin Q. Shengzhi W. Ruijin "Big data model of security sharing based on blockchain" Proc. 3rd Int. Conf. Big Data Comput. Commun. (BIGCOM) pp. 117-121 2017.

[8] Z. Xiong Y. Zhang D. Niyato P. Wang Z. Han "When mobile blockchain meets edge computing" IEEE Commun. Mag. vol. 56 no. 8 pp. 33-39 Aug. 2018.

[9] Y. Zhang J. Wen "The IoT electric business model: Using blockchain technology for the Internet of Things" Peer-to-Peer Netw. Appl. vol. 10 no. 4 pp. 983-994 2017.

[10] J. Kang et al., "Blockchain for Secure and Efficient Data Sharing in Vehicular Edge Computing and Networks," in IEEE Internet of Things Journal, vol. 6, no. 3, pp. 4660-4670, June 2019.

[11] Q. Ren, K. L. Man, M. Li and B. Gao, "Using Blockchain to Enhance and Optimize IoT-based Intelligent Traffic System," 2019 International Conference on Platform Technology and Service (PlatCon), Jeju, Korea (South), 2019, pp. 1-4.

[12] Y. Yang, L. Chou, C. Tseng, F. Tseng and C. Liu, "Blockchain-Based Traffic Event Validation and Trust Verification for VANETs," in IEEE Access, vol. 7, pp. 30868-30877, 2019.

[13] I. Eyal, "The Miner's Dilemma," 2015 IEEE Symposium on Security and Privacy, San Jose, CA, 2015, pp. 89-103.

[14] M. Rosenfeld, Analysis of Bitcoin pooled mining reward systems, arXiv preprint arXiv:1112.4980, 2011.

[15] L. Guo, Z. Liu and Z. Chen, "A leader-based cooperation-prompt protocol for the prisoner's dilemma game in multi-agent systems," 2017 36th Chinese Control Conference (CCC), Dalian, 2017, pp. 11233-11237.

[16] J. Li, C. Zhang, Q. Sun, Z. Chen and J. Zhang, "Changing the Intensity of Interaction Based on Individual Behavior in the Iterated Prisoners Dilemma Game," in IEEE Transactions on Evolutionary Computation, vol. 21, no. 4, pp. 506-517, Aug. 2017.

[17] X. Wang, C. Luo, S. Ding and J. Wang, "Imitating Contributed Players Promotes Cooperation in the Prisoners Dilemma Game," in IEEE Access, vol. 6, pp. 53265-53271, 2018.

[18] X. Xu Z. Rong C. K. Tse "Bounded rationality optimizes the performance of networked systems in Prisoners dilemma game" Proc. IEEE Int. Symp. Circuits Syst. (ISCAS) pp. 1-5 May 2018.

[19] Y. Xie, S. Chang M. Yan Z. Zhang X. Wang "Environmental influences on cooperation in social dilemmas on networks" Phys. A Stat. Mech. Appl. vol. 492 pp. 2027-2033 Feb. 2018.

[20] R. Lowe, Y. Wu et al. Multi-agent actor-critic for mixed cooperative-competitive environment Advances in Neural Information Processing Systems. pp. 63796390, 2017.