Learning Based Security for VANET with Blockchain

Canhuang Dai*, Xingyu Xiao*, Yuzhen Ding*, Liang Xiao*, Yuliang Tang†, Sheng Zhou‡
*Department of Communication Engineering, Xiamen University, Xiamen, China. Email: lxiao@xmu.edu.cn
†Department of Communication Engineering, Xiamen University, Xiamen, China. Email: tyl@xmu.edu.cn
‡ Department of Electronic Engineering, Tsinghua University, Beijing, China. Email: sheng.zhou@tsinghua.edu.cn

***Abstract*—The security issue is one of the greatest challenges in vehicular ad hoc networks (VANETs) attracting a great deal of attention. Malicious onboard units (OBUs) can attack other OBUs with various manners to obtain illegal gains, such as jamming, eavesdropping spoofing and so on. To reduce the potential attackers in the network, we propose an indirect reciprocity security framework with a scalar reputation assigned to each OBU to evaluate their dangerous level to the VANET. A blockchain technique that uses consensus mechanism and encryption algorithms to protect information from being tampered is applied for the transmitter to record the behaviors of other OBUs. We also propose a reinforcement learning based action selection strategy for an OBU in the VANET to choose a reliable relay OBU or determine whether to follow the request of a source OBU or not. A hotbooting technique is applied for the OBUs with prior knowledge to accelerate the learning speed. Simulation results show that the proposed action selection strategy can efficiently increase the packet delivery ratio, the reputation and the utility of the each OBU.**

***Index Terms*—VANET, security, blockchain, reinforcement learning, game theory.**

I. Introduction

VEHICULAR ad hoc network (VANET) as an emerging technology provides the efficient dissemination of information among onboard units (OBUs) and roadside units (RSUs) using protocols such as IEEE 802.11p [1]. The periodic broadcast of information in the VANET such as collision avoidance, road blocks and accident notification in rush hours, plays a critical role in prevention of accident, mitigation traffic jams and enhancement on-road experience [2]. Since OBUs are rational and thus naturally selfish, they make any efforts to maximize their benefit, including behaving unfaithfully. For example, greedy OBU that wants to avoid congestion might send the forged messages to convince its neighbors that there is congestion ahead, thus the OBUs in its neighborhood choose other routes and the greedy OBU can get an unimpeded road. Besides, malicious OBUs may attack the network if it can obtain an illegal gain exceeds the cost of performing such an attack. Common attacks include jamming, spoofing, eavesdropping, sybil attacks and some other potential attack manners [3].

The indirect reciprocity principle with a common social norm and evolution method is a promising tool for improving security, reducing potential attackers and guaranteeing efficient communication in wireless networks [4]. Each node under the indirect reciprocity principle in the networks takes action to another node according to its reputation which indicating its dangerous level. The main idea of the indirect reciprocity principle is "I help you and someone else helps me" [5]. Monitoring nodes can detect what each node did in the pass transmit process and upload this profile to the certificate authority, which then updates the reputation of each node in the network.

Attackers may tamper the reputation in the broadcast process and thus paralyze the indirect reciprocity mechanism. The blockchain technique can be applied to avoid such a situation, which distributed stores and broadcasts data across a peer-to-peer network instead of a third-party intermediary. Each node in a decentralized system has a copy of the blockchain that consisting of the blocks recording the complete information about the actions that every node takes in the past. Due to the consensus mechanism and the dedicated encryption methods, it is promising that the blocks cannot be faked and modified retroactively. With the distributed, transparent and secure nature, blockchain is potentially suitable for the records management, and has been widely used in edge network [6], smart grids [7], [8], access control systems [9] and cognitive radio networks [10].

The action selection and process of the nodes in a VANET that applies the indirect reciprocity principle can be view as a Markov decision process (MDP) due to the long-term evolution essence. More specifically, the action selection of a node affects not only its current utility but also its future expected utility. For instance, a node attacks the others may obtain a considerable reward immediately but suffer a communication failure in the future as no one helps it. Therefore, each node can apply reinforcement learning (RL) techniques such as Q-learning to achieve the optimal strategy via trials-and-errors. In this scheme, the actions are chose according to the current state that consists of the reputation of the VANET, the location of each OBU and the role the OBU

This paper was in part supported by National Natural Science Foundation of China under Grant 91638204 and 61671396 and in part by the open research fund of National Mobile Communications Research Laboratory, Southeast University (No. 2018D08).

plays in a transmission process based on a quality function or Q function for each state-action pair. Updated according to the iterative Bellman equation, the Q function provides the expected discounted long-term reward of the VANET. An OBU leaves a VANET may re-enter another one after some time. Therefore, a hotbooting technique is provided for improving the adaptation capability of the OBUs with prior knowledge, which utilizes the experience to initialize the Q function at the beginning.

The contributions of this paper can be summarized as follows:

- We propose an indirect reciprocity based security framework to guide the behavior of the OBUs in the VANET and reduce the potential attackers and apply the blockchain technique to protect the reputation from being tampered.
- We propose a Q-learning based action selection strategy in the indirect reciprocity game to achieve the optimal action without the knowledge of the channel model.
- A hotbooting technique is applied for the OBUs that re-enter a VANET to accelerate the learning speed.

The rest of this paper is organized as follows: We first review the related works in Section II. We present the system model in Section III and then design an indirect reciprocity framework with blockchain in Section IV. We propose a Q-learning based action selection strategy for OBUs in VANET and further apply the hotbooting technique to accelerate the learning speed in Section V. The simulation results are provided in Section VI. Finally, we make a conclusion in Section VII.

II. Related works

Numerous works have been done to deal with the security issues in the VANETs [11]. For example, a classification of current existing security issues and the possible cryptographic solutions are proposed in [3]. The hideaway strategy as proposed in [12] keeps the legitimate OBU silent if a jamming attack is detected based on the packet transmission ratio to improve the resistance against jamming. The location verification system as proposed in [13] takes the claimed location and raw observations across the receiving BS antennas to authenticate a location spoofing attack. The medium access control based method as proposed in [14] provides a real-time attack detection for safety application with low probability of false alarms.

The indirect reciprocity principle has been applied for the wireless network to facilitate cooperation between nodes and improve communication efficiency. The security system as proposed in [15] applies the indirect reciprocity to combat attacks in wireless networks and motivates the nodes to help each other. The indirect reciprocity based data sharing incentive scheme as proposed in [16] analyzes the uniqueness and stationary of the reputation distribution of the optimal action, showing that it will lead to a secure network. The indirect reciprocity based packet forwarding framework is developed in [17] to enforce cooperation of nodes in both structured and unstructured mobile ad hoc networks.

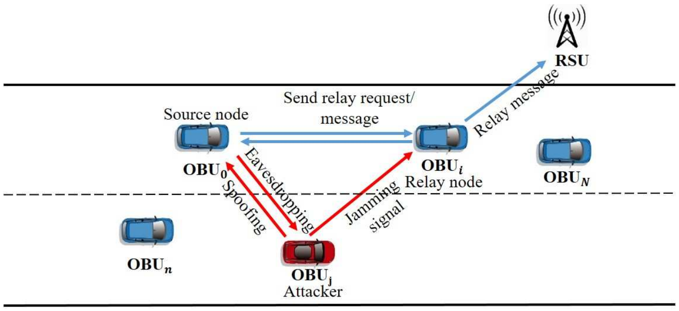


Fig. 1: Illustration of a transmission process in VANET.

Reinforcement learning techniques have been widely applied to improve security in wireless networks For example, the channel selection based on Q-learning can improve the resistance against jamming [18], and the channel accessing with Q-jamming [19] can resist smart jamming in hostile environment. A two-dimension anti-jamming communication system as proposed in [20] applies both frequency hopping and user mobility with deep Q-learning to increase the signal-to-interference-plus-noise ratio of legitimate nodes and accelerate the learning speed to obtain the optimal anti-jamming strategy. The physical layer authentication method with reinforcement learning as proposed in [21] exploits the radio channel information to detected spoofing attacks in wireless networks.

III. System Model

A. Network Model

To illustrate the system model clearly, we use a simple example network which is consisting of a RSU and N OBUs on the highway environment as shown in Fig. 1. OBU is the microwave device equipped on the vehicle for communication with other vehicles or the RSU which uploads messages such as the traffic accident alert or application data for the OBUs to the server. The OBUs transmit signals based the dedicated short range communications protocol that can offer a reliable communication within tens of meters and applies the carrier sense multiple access technique to avoid collisions. Let $\mathcal{U} = \{\mathrm{U}_n\}_{0 \leq n \leq N}$ be the node set of the VANET, in which each OBU is denoted by U_n with an unique ID $1 \leq n \leq N$ and the RSU is accordingly denoted by U_0.

Communication requests arrive at each individual OBU following a Poisson process with an identical parameter ρ. At the k-th transmission process, let $\mathrm{U}_t^{(k)}$, $1 \leq t \leq N$ be the source OBU that needs to upload messages. The superscript k is omitted if no confusion occurs. U_t first attempts to transmit the messages to the RSU directly. If communication between the source node and the RSU is unavailable due to the distance, the source node U_t can request another OBU, denoted to relay the messages for it. An indicator $\xi_n \in \{0, 1, 2\}$ is used to represent what role U_n plays in the network, with $\xi_n = 0$ means U_n is the source node, $\xi_n = 1$ means U_n is selected as a relay node and $\xi_n = 2$, otherwise.

Assuming that each OBU is self-interested, the selected relay OBU can determine to follow or refuse the relay request from U_t. For simplicity, we assume that the messages can be delivered to the RSU in one-hop if no attack happens, but it can be furthur extended to multi-hop scenerios by iteratively considering the selected relay OBU as a source OBU that continues selecting the successive relay OBU.

B. Attack Model

The other OBUs in the VANET beyond the source node may attack the legal communication via various manners to gain illegal rewards. In this paper, we consider that a malicious OBU is capable to launch three kinds of attack, including eavesdropping, jamming and spoofing. Specifically, an OBU can steal the privacy of U_t by eavesdropping the signal or blocks the legal communication by transmitting noisy signals at a jamming power denoted by P_J to the receiver, i.e., the selected relay OBU or the RSU. The non-relay OBUs can spoofing the source node by pretending that it is the relay node. In another case, the OBU that is selected as a relay OBU may be exactly the attacker that launches the spoofing attack by cheating the source node with faked responses. The actions of the OBUs, including the legal and illegal ones, are labelled with an action ID x, $1 \leq x \leq 5$, and evaluated with different dangerous level l, $0 \leq l \leq 4$ which are summarized in Table I. The action with higher dangerous level is more harmful to the network. By taking an action with ID x_n, node U_n, $1 \leq n \neq t \leq N$, obtains a corresponding instant payoff $\varpi(x_n)$ which is the difference between the cost of the action and the gain.

Extensive work on detecting the attack manners and the attackers has been done in the literature [22]. In this paper, we mainly focus on restricting the attackers and encouraging each OBU to behave itself with a reputation mechanism that is specified in the next section.

Action ID	Manner	Dangerous Level
1	Jamming	4
2	Spoofing	3
3	Eavesdropping	2
4	Disobey the request	1
5	Follow the request	0

TABLE I: Dangerous levels of different actions.

C. Channel Model

OBUs can communicate with its neighbors by periodical beacon messages and obtain channel state information such as the channel gain between any two node in the VANET, denoted by $h_{m,n}$, $0 \leq m \neq n \leq N$, which varies over time. In this paper, we refer to the two-ray path loss model proposed in [23] that is widely-used at the highway scenario which involves the wavelength λ, the distance $d_{m,n}$ and the heights c_m and c_n of U_m and U_n, respectively, and accordingly evaluate the channel gain as following:

$$h_{m,n} = \begin{cases} \lambda/4\pi d_{m,n}, & d_{m,n} \leq 4\pi c_m c_n/\lambda, \\ c_m c_n/d_{m,n}^2, & d_{m,n} > 4\pi c_m c_n/\lambda. \end{cases} \tag{1}$$

The jammers can deteriorate the communication quality between any two nodes in the VANET. Let $\mathcal{U}_j \subseteq \mathcal{U}$ be the set of the jammers. The signal to interference plus noise ration (SINR) of the signal transmit from U_m to U_n is given by $\varrho_{m,n} = P_T h_{m,n} / \left(1 + \sum_{\mathrm{U}_j \in \mathcal{U}_j} P_J h_{j,n} + \sigma_n\right)$, where σ_n is the noise which follows a Gaussian distribution, i.e, $\sigma_n \sim \mathcal{N}(0, \delta_n)$. We assume that U_m can establish a stable communication link with another node U_n if the channel gain between them is larger than a given threshold τ, i.e., $\varrho_{m,n} > \tau$.

For ease of reference, the common used symbols and notations are summarized in Table II

N	The number of OBUs in the VANET
$\mathcal{U}$	The set of all nodes in the VANET
U_n	A node (an OBU or a RSU) with ID n
U_t	The source node
ξ_n	Relay indicator of node n
y_n	Scalar reputation value of node n
Y	Maximal reputation value
$d_{m,n}$	Distance between node U_m and U_n
x	Action ID
l	Dangerous level of different attack manner
$\varpi(x)$	Attack reward corresponding to action x
λ	Wavelength of the transmit signal
$P_{T/J}$	Transmit/Jaming power
$h_{m,n}$	Channel gain between node U_m and U_n
σ_n	Noise received by U_n
$\varrho_{m,n}$	SINR of the signal from node U_m to U_n
c_n	Height of node U_n
$Z^{(k)}$	Block of the k-th transmit process
$D^{(k)}$	Root of the Merkel tree
$T^{(k)}$	Time stamp
$\mathbf{s}^{(k)}$	System state at the k-th transmission process
α	Learning rate
γ	Discount factor

TABLE II: Summary of symbols and notations

IV. Indirect Reciprocity Based Security Framework with Blockchain

A. Reputation Mechanism

We design a security framework that applies the reputation mechanism based on the indirect reciprocity principle and the blockchain technique to reduce the potential attackers in a VANET. The reputation of U_n is denoted by y_n with $0 \leq y_n \leq Y$, where Y is the highest scalar reputation. The OBU with higher reputation is more likely to be selected as a relay OBU and its relay request is more likely to be accepted by the selected relay OBU. On the contrary, the OBU with lower reputation is viewed as more dangerous. The RSU maintains a highest reputation $y_0 = Y$ throughout as it never influences the network service. The reputation of U_n is updated based on a designed social norm $\mathbf{\Omega} = [\omega(x, y_t)]$, in which $\omega(x, y_t)$ is the instant scalar reputation assigned to U_n if it takes action x to the source node U_t with a reputation y_t.

Intuitively, the social norm should obeys the following principles: The nodes that helps a source node with high reputation or disobey the request of a source node with low reputation should be assigned with a high reputation. On the other hand, the nodes that attacks source node should be punished and assigned with a low reputation according to the dangerous level of the action, regardless of the reputation of source node. Therefore, we build the social norm as

$$\boldsymbol{\Omega} = \begin{bmatrix} 1 & 1 & \dots & 1 & 1 \\ 2 & 2 & \dots & 2 & 2 \\ 3 & 3 & \dots & 3 & 3 \\ Y & Y-1 & \dots & 2 & 1 \\ 1 & 2 & \dots & Y-1 & Y \end{bmatrix}. \tag{2}$$

Based on the social norm, the reputation of U_n at the k-th transmission process is updated by $y_n^{(k)} = (1-\eta)y_n^{(k-1)} + \eta\omega(x_n^{(k)}, y_t^{(k)})$. A forget factor η is used to weight how much the one-time action affects the reputation of a node.

B. Blockchain Based Security Framework

In the blockchain based security framework, the source OBU creates a new block $Z^{(k)}$ that records the actions of its neighboring OBUs including the selected relay OBU at the end of the k-th transmission process. The block header consists of hash value of the previous block $Z^{(k-1)}$, the time stamp $T^{(k)}$, a nonce $N^{(k)}$ and the root of a Merkel tree. The leaves of the Merkel tree are the hash values of the previous blocks. By repeatedly calculating the hash values of the child nodes to obtain the values of the parent nodes, we can construct a binary Merkel tree with a root denoted by $D^{(k)}$ which is unique for a given blockchain according to the property of the hash functions such as the SHA-256 algorithm. The block body compresses the verified data in the form of a hashed Merkle tree with its root included in the header.

Once a new block is created, the source OBU broadcasts it to the other OBUs so that each OBU is aware of what others did in the last transmission process and locally updates the reputation of each OBU according to the common social norm. The information related to this transmission process included in the block can be verified by all OBUs locally. If verified by a majority of OBUs in the VANET, the block will be linked to the previous blocks and thus forming a chain. The block that contains the message inconsistent with the reality cannot obtain the certification from the OBU and is abandoned by the system. Based on the blockchain based security system, the OBU has the access permission to inquire the previous transmission related message of all OBUs that cannot be faked and modified.

V. Q Learning Based Action Selection Strategy

The indirect reciprocity process in the VANET can be viewed as an MDP. More specifically, a node that helps the source node suffers a cost due to the energy consumption. But the other nodes is more willing to help it if it needs a relay, as its reputation increases. Thus each node can benefit from its current meritorious behavior in the future. Besides, the relay selection of the source node also affects the reputations of the other nodes in the network. Consequently, we can apply the RL techniques to derive the optimal action selection for each OBU in the VANET [24].

In the k-th transmission process, U_n is aware of the scalar reputation vector $\mathbf{y}^{(k)} = [y_n^{(k)}]_{1\leq n\leq N}$ which is distributed stored at every OBU in a blockchain system. Besides, through the ranging radars or the global position system, U_n is able to obtain the locations and velocities of other nodes denoted by $\mathbf{p}^{(k)} = [p_n^{(k)}]_{1\leq n\leq N}$ and $\mathbf{v}^{(k)} = [v_n^{(k)}]_{1\leq n\leq N}$, respectively. Including its role, the reputations and the positions and velocities of all OBUs in the network, U_n formulates a environment state denoted by $\mathbf{s}_n^{(k)} = [\xi_n^{(k)}, \mathbf{y}^{(k)}, \mathbf{p}^{(k)}, \mathbf{v}^{(k)}]$.

Let $Q_n\left(\varphi_n^{(k)}, x_n\right)$ denote the estimated Q-value, i.e., the long-term accumulate discount utility of U_n taking action x_n. A ε-greedy policy is applied for action selection to balance the exploitation and exploration based on the Q-value in the learning process. Specifically, the probability of selecting action x_n is given by

$$\Pr(x_n = x'_n) = \begin{cases} 1-\varepsilon, x'_n = \arg\max_{x'_n \in \mathcal{A}_{\xi_n}} Q_n\big(\mathbf{s}^{(k)}, x'_n\big) \\ \frac{\varepsilon}{|\mathcal{A}_{\xi_n}|-1}, \text{o.w.} \end{cases} \tag{3}$$

where $\mathcal{A}_{\xi_n}$ is the action set of U_n corresponding to the role indicator ξ_n. Specifically, the action set of a source node is the node set expect itself, i.e., $\mathcal{A}_0 = \mathcal{U}\backslash\{\mathrm{U}_n\}$. Besides, according to action ID in Table I, the action set of the relay and non-relay node are $\mathcal{A}_1 = \{2,3,4,5\}$ and $\mathcal{A}_2 = \{1,2,3,4,5\}$, respectively. The source node detects the behaviors of other nodes in the VANET, formulates a new block $Z^{(k)}$ and broadcasts it to the other nodes in the VANET, so that every node updates the reputation of the nodes in the VANET locally according to the common social norm, including the reputation of itself. The utility of each node in current transmission process is consisting of its reputation gain and the payoff of its action. Let μ denote the relative weight of the reputation gain, we formulate the utility by $u_n^{(k)} = \mu y_n^{(k)} + \varpi(x_n^{(k)})$.

At each transmission process, each node can update the Q-value based on the utility according to the Bellman Equation [24] given by

$$\begin{aligned} Q_n\left(\mathbf{s}^{(k)}, x_n^{(k)}\right) &\leftarrow (1-\alpha)Q_n\left(\mathbf{s}^{(k)}, x_n^{(k)}\right) \\ &+ \alpha\left(u_n^{(k)} + \gamma \max_{x_n\in\mathcal{A}_{\xi_n}} Q_n\left(\mathbf{s}^{(k+1)}, x_n\right)\right), \end{aligned} \tag{4}$$

where the learning rate $\alpha \in (0,1]$ weights the current experience and the discount factor $\gamma \in [0,1]$ represents the weight of the future utility. The OBUs perform more greedy for current utility with lower γ. The Q-values are initialized with random values or zeros generally. For the re-enter OBUs that has ever connected to the a VANET under the indirect reciprocity principle, a hotbooting technique can be applied to utilized the prior knowledge to initialize the Q-values to accelerate the learning speed. With the experience including the required environment state, action and utility in the past time, the OBUs re-enter the VANET can first iteratively updated the initial Q-value via Eq. (4).

VI. Simulation Results

Simulations have been implemented to evaluate the performance of the proposed Q learning and hotbooting Q based action selection strategies for a VANET with 8 OBUs and 1 RSU. Each OBU moved with the same direction at a velocity between 60 and 90 km/h, which randomly changed over time. In the simulations, the communication frequency is 5.9 GHz according to the protocol IEEE 802.11p. We set $\lambda = 0.05$ m according to the communication frequency and $h_t = h_r = 1.2$ m for the channel model. The initial reputation of each OBU is randomly selected in $\{0, 1\}$ and maximal reputation is $Y = 4$. Unless specified otherwise, the transmit power, relay power and the jamming power is 200 mW, 200 mW and 100 mW, respectively. Besides, we set $\alpha = 0.7$, $\gamma = 0.5$ and $\varepsilon = 0.1$ in the learning process.

The simulation results show that the proposed hotbooting Q based action selection strategy outperforms the Q-learning based one with higher PDR, reputation and utility as shown in Fig. 2. For instance, as shown in Fig. 2(a), the Q-learning based scheme increases the PDR from 0.45% to 0.71% after transmitting 300 packets, and the hotbooting Q based scheme further increases it by 21.1%. At the same time, the Q-learning based scheme increases the utility from 20.96 to 121.81, and the hotbooting Q based strategy increases the utility of the OBUs by 25.8%, as shown in Fig. 2(c).Besides, as shown in Fig. 2(b), the Q-learning based action selection strategy increases the average reputation of the OBUs from 0.53 to 2.05 and the proposed hotbooting Q based strategy further increases this by 48.8% at the 1500-th time slot.

VII. Conclusion

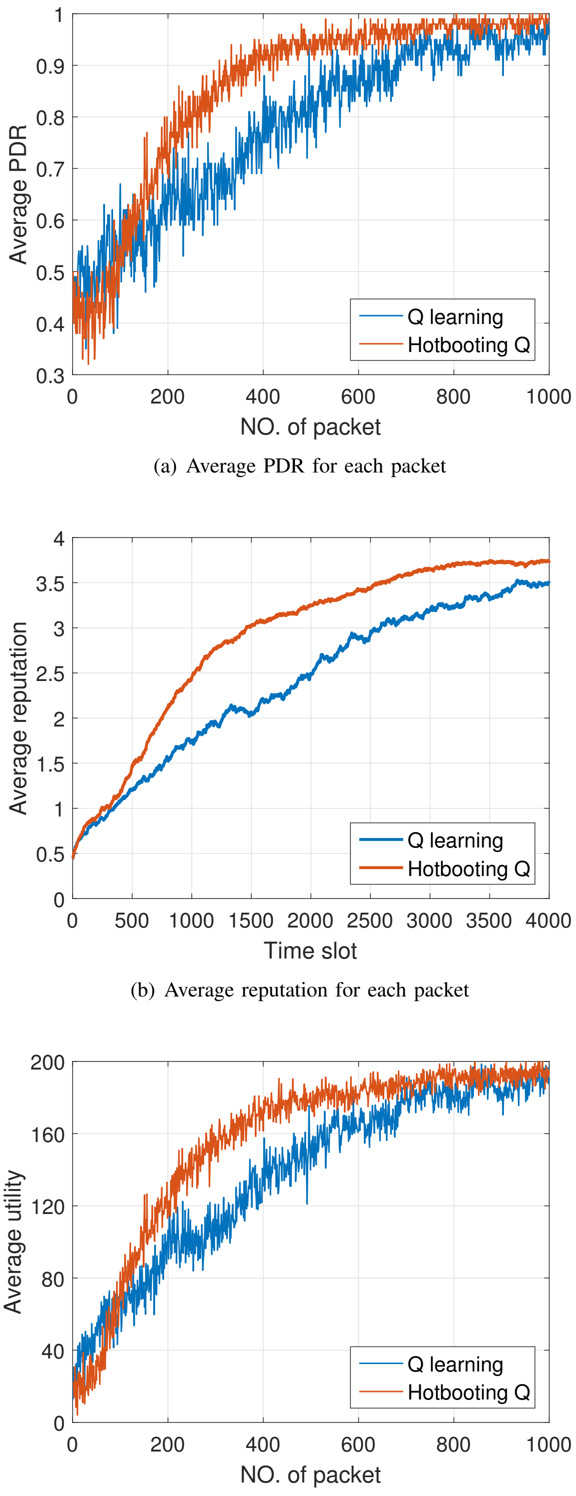
In this paper, we have proposed a indirect reciprocity based security framework to encourage OBUs in the VANET to help each other and reduce the potential attackers. A scalar reputation is assigned to each OBU which is updated according to a designed social norm. A blockchain technique is used for information recording to prevent the reputation vector of the VANET from being tampered by the malicious OBUs. A Q learning based action selection strategy has been proposed to choose the optimal actions for the OBUs without being aware of the channel model. A hotbooting technique that utilizes the prior experience to initialize the Q-values is used to accelerate the learning rate. As shown in the simulation results, the proposed hotboting Q based action selection strategy can improve the average reputation by 3.52 after 3000 transmissions and improve the packet delivery rate and utility by 55% and 151.64, respectively, after transmitting 600 packets.

References

[1] S. Sultan, M. Doori, A. Bayatti, *et al.*, "A comprehensive survey on vehicular ad hoc network," *J. Netw. Comput. Appl.*, vol. 37, pp. 380–392, Jan. 2014.

[2] H. Hartenstein and L. Laberteaux, "A tutorial survey on vehicular ad hoc networks," *IEEE Commun. Mag.*, vol. 46, no. 6, Jun. 2008.

[3] M. N. Mejri, J. Ben-Othman, and M. Hamdi, "Survey on VANET security challenges and possible cryptographic solutions," *Vehicular Comm.*, vol. 1, no. 2, pp. 53–66, Apr. 2014.



(a) Average PDR for each packet

(b) Average reputation for each packet

(c) Utility

Fig. 2: Performance of the RL based action selection strategies for indirect reciprocity among OBUs in VANETs with 8 OBUs.

[4] B. Niu, H. V. Zhao, and H. Jiang, "A cooperation stimulation strategy in wireless multicast networks," *IEEE Trans. Signal Proc.*, vol. 59, no. 5, pp. 2355–2369, Feb. 2011.

[5] N. Martin and K. Sigmund, "Evolution of indirect reciprocity," *Nature*, vol. 437, pp. 1291–1298, Oct. 2005.

[6] D. Xu, L. Xiao, L. Sun, *et al.*, "Game theoretic study on blockchain based secure edge networks," in *Proc. IEEE/CIC Int. Conf. Comm.*, Qingdao, China, Sep. 2017.

[7] J. Kang, R. Yu, X. Huang, *et al.*, "Enabling localized peer-to-peer electricity trading among plug-in hybrid electric vehicles using consortium blockchains," *IEEE Trans. Ind. Inform.*, vol. 13, no. 6, pp. 3154–3164, Dec. 2017.

[8] N. Z. Aitzhan and D. Svetinovic, "Security and privacy in decentralized energy trading through multi-signatures, blockchain and anonymous messaging streams," *IEEE Trans. Dependable Security Comput.*, Oct. 2016.

[9] G. Zyskind, O. Nathan, *et al.*, "Decentralizing privacy: Using blockchain to protect personal data," in *Security and Privacy Workshops*, pp. 180–184, San Jose, CA, Jul. 2015.

[10] K. Kotobi and S. G. Bilén, "Blockchain-enabled spectrum access in cognitive radio networks," in *Wireless Telecommun. Symp.*, pp. 1–6, Chicago, IL, Apr. 2017.

[11] L. Xiao, W. Zhuang, S. Zhou, and C. Chen, "Learning-based vanet communication and security techniques," Springer, 2019, ISBN: 978-3-030-01731-6.

[12] I. K. Azogu, M. T. Ferreira, J. A. Larcom, and H. Liu, "A new anti-jamming strategy for VANET metrics-directed security defense," in *IEEE Global Commun. Conf.*, pp. 1344–1349, Atlanta, GA, Dec. 2013.

[13] S. Yan, R. A. Malaney, I. Nevat, and G. W. Peters, "Location verification systems for VANETs in rician fading channels," *IEEE Trans. Vehicular Technology*, vol. 65, no. 7, pp. 5652–5664, Jul. 2016.

[14] A. Benslimane and H. Nguyen-Minh, "Jamming attack model and detection method for beacons under multichannel operation in vehicular networks," *IEEE Trans. Vehicular Technology*, vol. 66, no. 7, pp. 6475–6488, Jul. 2017.

[15] L. Xiao, Y. Chen, W. S. Lin, *et al.*, "Indirect reciprocity security game for large-scale wireless networks," *IEEE Trans. Inf. Forensic Secur.*, vol. 7, no. 4, pp. 1368–1380, Aug. 2012.

[16] B. Zhang, Y. Chen, J.-L. Yu, B. Chen, and Z. Han, "Indirect-reciprocity data fusion game and application to cooperative spectrum sensing," *IEEE Trans. Wireless Commun.*, vol. 16, no. 10, pp. 6571–6585, Oct. 2017.

[17] C. Tang, A. Li, and X. Li, "When reputation enforces evolutionary cooperation in unreliable manets," *IEEE Trans. Cybernetics*, vol. 45, no. 10, pp. 2190–2201, Oct. 2015.

[18] W. G. Conley and A. J. Miller, "Cognitive jamming game for dynamically countering ad hoc cognitive radio networks," in *Military Commun. Conf.*, pp. 1176–1182, San Diego, CA, Nov. 2013.

[19] Y. Gwon, S. Dastangoo, C. Fossa, and H. Kung, "Competing mobile network game: Embracing antijamming and jamming strategies with reinforcement learning," in *IEEE Conf. Commun. Netw. Security*, pp. 28–36, National Harbor, MD, Oct. 2013.

[20] 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 Proc.*, pp. 2087–2091, New Orleans, LA, Mar. 2017.

[21] L. Xiao, Y. Li, G. Han, G. Liu, and W. Zhuang, "Phy-layer spoofing detection with reinforcement learning in wireless networks," *IEEE Trans. Vehicular Technology*, vol. 65, no. 12, pp. 10037–10047, Dec. 2016.

[22] Y. Zou, J. Zhu, X. Wang, *et al.*, "A survey on wireless security: Technical challenges, recent advances, and future trends," *Proc. IEEE*, vol. 104, no. 9.

[23] 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 Netw. and Computing*, pp. 1–3, Nevada, LV, Sep. 2011.

[24] R. S. Sutton, A. G. Barto, *et al.*, *Reinforcement learning: An introduction*. MIT press, 1998.