Q-Learning Based Routing Algorithm for Urban Vehicular Networks

[1]Phouthone Vongpasith, [2]Soukpaseuth Banchong, [3]Vilaysack Keosouttha,
[4]Bouaketh Vannachit, [5]Lathsamy Chidtavong

Faculty of Natural Sciences
Department of Computer Science
National University of Laos
Vientiane Capital, Lao P.D.R P.O.BOX:7322

***Abstract*- A network with reliable and rapid communication is critical for urban Vehicular Ad Hoc Networks (VANETs). VANETs consisting of vehicular nodes moving on the roads with wireless communication. However, the highly dynamic topology of VANETs and limited communication of performance have brought great challenges to the routing design of VANETs. It is difficult for existing routing protocols for VANETs to adapt the high dynamics of VANETs. Moreover, few of existing routing protocols simultaneously meet the requirement of high packet delivery and low delay in Intelligent Transportation Systems (ITSs). This paper proposes a novel Q-learning based routing algorithm under next-hop selection optimization of geographic routing protocol. To adopt the decision making on the next-hop selection algorithms, an off-policy algorithm as it learns an optimal action maximize Q value based on a reward function that estimates a link reliability between nodes in the routing decision process to select the more reliable next-hop. The performance of the proposed routing approach is evaluated using comprehensive network simulation. Simulation results show that the proposed algorithm can provide higher packet arrival ratio, lower delay than existing routing protocols.**

***Keywords*-Vehicular Ad-hoc Networks; Q-learning Routing; Link Reliability; Geographic Routing**

I. INTRODUCTION

Vehicular Ad Hoc Networks (VANETs) consisting of a group of vehicular nodes that communicate over wireless links. Since the nodes are vehicular moving on the roads, the network topology may change rapidly and unpredictably over time. A major function in VANETs is the route discovery process, where a route from source node to destination node is discovered in order to transfer data packets. In VANETs, there are three classes of routing protocols: proactive, reactive, and geographic (Rahman.). The proactive protocols are table driven where each node maintains a route to every other node in the VANETs. Due to network topology and limited communication of performance this protocol model is less preferred in VANETs. Reactive routing strategy is popular in wireless ad hoc networks because of its less overhead and on-demand nature. Ad hoc on-demand distance vector (AODV) is the most popular examples of reactive routing protocols (Perkins). Greedy perimeter stateless routing (GPSR) (Karp.B.) is one of the most popular geographic routing protocols, which has also inspired varies extensions in (Li), (Jinqlao) and (Fan). The core idea of greedy routing is to forward the packet to the specific neighbor that is geographically closest to the destination.

In this work, the Q-learning technique is utilized to enable each node to learn how to select the optimal next-hop for data forwarding according to its trust Q value and reward function. Since Q-learning does not always need the detailed model description in computation, it is a widely used reinforcement learning method. It is the best method used to analyze autonomous agents that self-adapt to varying external environments (Anitha) (Arnau). Finally, a Q-learning routing algorithm based link reliability that calculates Q-value and link reliability vehicular nodes for next-hop selection is simulated in urban VANETs.

The rest of this paper is organized as follows. Section 2 gives an overview on the related work. Q-learning based routing protocol and the proposed algorithms are discussed in Section 3. Section 4 elaborates simulation environment and discusses the results on the efficiency of our algorithms. Finally, the last section concludes the paper and gives suggestion for further work in this area.

II. RELATED WORK

A wide range of solution approaches for specific routing protocols and different techniques of using reinforcement learning in data forwarding for vehicular networks has been proposed by literature (Jun), (Mujeeb), (Paúl), (Jae), and (Ehsan). Predictive Ad-hoc Routing fuelled by reinforcement learning and trajectory knowledge has been introduced in (Benjamin). (Jianmin) presented a novel Q-learning based multi-objective optimization routing protocol for FANETs to provide low-delay and low-energy service guarantees. Most of existing Q-learning based protocols use a fixed value for the Q-learning parameters. In contrast, Q-learning parameters can be adaptively adjusted in the proposed protocol to adapt to the high dynamics of FANETs. In the work presented by (Tauqeer), the decision making in next-hop selection is based on past and expected future routing decisions which is performed on the basis of reinforcement learning algorithms such as *Q*-Learning. (Rahul, Analysis of Reinforcement Based Adaptive Routing in MANET), (Ribal), (Zoubir), and (Rezoan) provide a comprehensive review of literature on the topic in RL-based routing protocols. A proper reward design and training mechanism and the multiple agents successfully learn demonstrated that with to cooperate in a distributed way to simultaneously improve the routing performance for VANETs (Le). To improve the performance of Q-routing the proposed analytical model considers the markov decision process and Q-learning the decision process of different task flows of VANETs (Elmustafa) and (Rahul, Analysis of Reinforcement Based Adaptive Routing in MANET). Aiming at the low efficiency

and slow convergence of Q-learning, heuristic function and evaluation function are introduced to accelerate the update of Q-value of current optimal action, reduce unnecessary exploration, accelerate the convergence speed of Q-learning process and improve learning efficiency (Yang). To meet these features, many research works are focused on Q-learning routing.

III. Q-LEARNING BASED ROUTING ALGORITHM

In Figure 1, a current node interacts with its neighbouring nodes to make routing decisions for data forwarding. In such a case, the agent is a current node, the environment is router's neighbour in VANETs and actions are selections of next neighbor nodes to forward data packets. Q-learning algorithms are based on reward functions. The role of the reward, which is returned by the environment to the agent, is to provide feedback to the learning algorithm about the effect of the recent taken action. Whereas a reward function indicates what is good (or bad) in an immediate sense, a value function indicates what is good (or bad) in long-term (Benjamin) and (Jianmin).

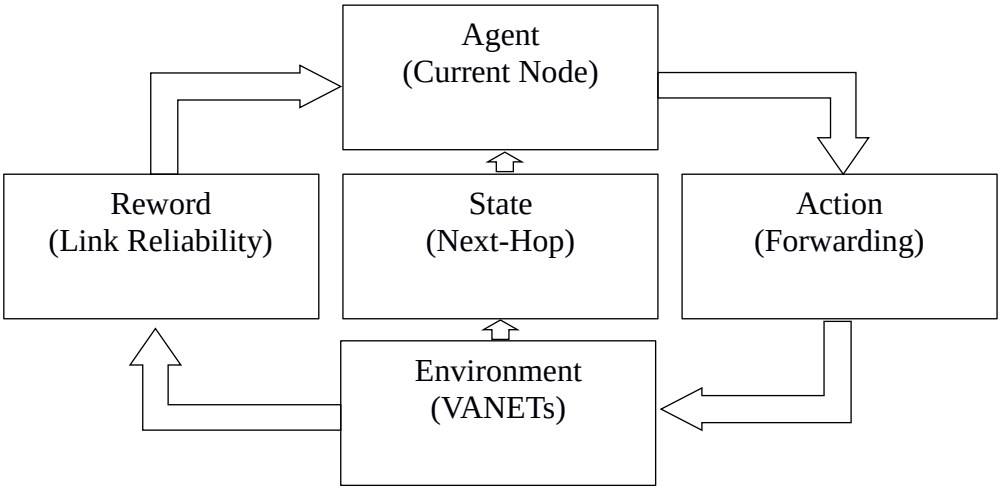

Figure 1: Q-Learning based on routing with a link reliability

A reward function calculated based on link reliability value that is a function of time, it considers the position, direction and speed from the HELLO packets for estimating the lifetime of communication between any two nodes, it defined as following.

$$R_{N_iD} = \frac{\sqrt{(a^2+c^2)R^2-(ad-bc)^2}-(ab+cd)}{(a^2+c^2)} \quad (1)$$

$$a = v_{N_i} cos\omega_{N_i} + v_D cos\omega_D$$
$$b = x_{N_i} + x_D$$
$$c = v_{N_i} sin\omega_{N_i} + v_D sin\omega_D$$
$$b = y_{N_i} + y_D$$

Where $R_{N_iD}$is a reward function calculated based on link reliability from $N_i$to D; $\omega_{N_i}$and $\omega_D$are the directions of $N_i$and D respectively; $v_{N_i}$and $v_D$are the speeds of $N_i$and D respectively; and$(x_{N_i}, y_{N_i})$and $(x_D, y_D)$are the positions of $N_i$and D respectively.

In Q-Learning, an action is executed base on the reword received from the environment (Arnau). In the proposed work, current node is formulated as an agent (Figure 3). It calculates the Q value based on the reward which calculates from the link reliability value of each it's neighbouring nodes and the destination node. Generally the Q value (Q-Learning score) is defined as (Anitha).

$$Q_{N_i}(s_{t+1}, a_{t+1}) \leftarrow (1-\alpha)Q_C(s_t, a_t) + \alpha\left[R_{N_iD}(s_{t+1}, a_{t+1}) + \gamma min Q_{N_i}(s_{t+1}, a_{t+1}) - Q_C(s_t, a_t)\right] \quad (2)$$

where $Q_C(s_t, a_t)$ represents the actual estimate Q value of current node C for the state-action pair(s_t, a_t); $Q_{N_i}(s_{t+1}, a_{t+1})$ is the new estimation Q value of neighbor node N_i for the future state-action pair(s_{t+1}, a_{t+1});$R_{N_iD}(s_{t+1}, a_{t+1})$ define the reward value obtained from link reliability between $N_i$and$D$; α represents step size parameter ($0 < \alpha \leq 1$); γ is the discount factor ($0 < \gamma \leq 1$) and t is the current step number. Therefore, the estimated Q value as stated above is impacted by the reward value. The proposed algorithm here in is an off-policy algorithm as it learns an optimal action max $Q_{N_i}(s_{t+1}, a_{t+1})$, where the agent selects the current optimal action based on a reward function $R_{N_iD}$and selects a random action with optimal action max$Q_{N_i}(s_{t+1}, a_{t+1})$. The Q-learning based routing is presented in Algorithm 1.

Algorithm 1 Q-learning Based Routing

1: Initialize Q value of current node C$Q_C(s_t, a_t)$for the state-action pair(s_t, a_t)
2: Initialize next-hop = -1;
3: Define value of ∝ and γ;
4: whileN_i do // i=1,2,...
4: Calculate link reliability reward R_{N_iD}using equation (1);

5: Update$Q_{N_i}(s_{t+1}, a_{t+1})$by equation (2) based on R_{N_iD}
6: if($Q_C(s_t, a_t)$<$Q_{N_i}(s_{t+1}, a_{t+1})$) then
7: $Q_C(s_t, a_t) \leftarrow Q_{N_i}(s_{t+1}, a_{t+1})$
8: next-hop ← 1
9: end if
10: end while
11: return next-hop

IV. SIMULATION RESULTS

We measured the packet delivery ratio and average end-to-end delay that directly reflects the performance of routing protocols. To study the performance of the proposed Q-learning routing and corresponding compared protocols (Benjamin), (Perkins), and (Rahman.) under impact of number and speed of nodes.

Figure 2 shows the packet delivery ratio of varying number of nodes in the urban VANETs. When the number of nodes is small, most of the optimal next-hop nodes can be selected using the proposed algorithm. Thus, forwarding data packet is almost 90% to 99% under the proposed condition which is always 2% to 5% more than the existing routing protocols. Whenever there is a change in route due to link failure, the intermediate nodes should share this information. Rarely, intermediate nodes can also misbehave, and thus the data packet delivery decreases with the increase in the number of nodes. The packet delivery ratio decreases more in existing GPSR as the number of nodes increases.

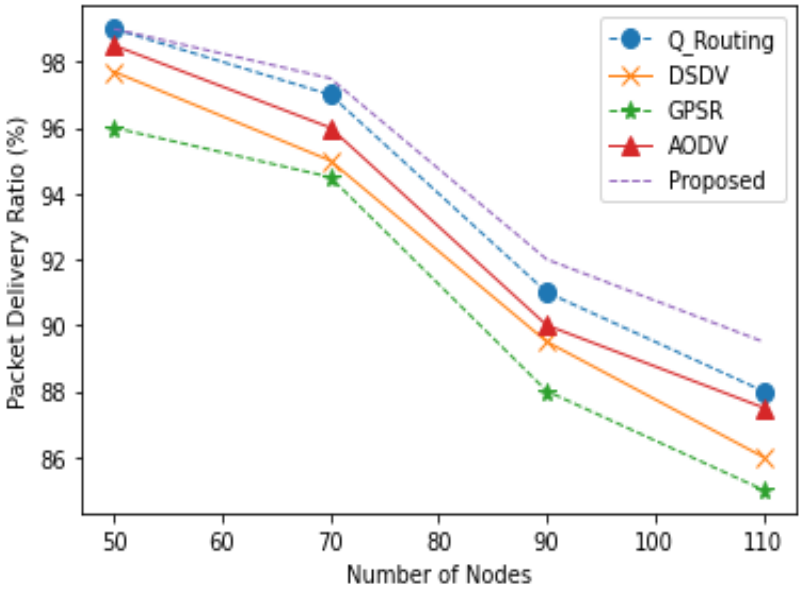

Figure 2: Impact of number of nodes on packet delivery ratio

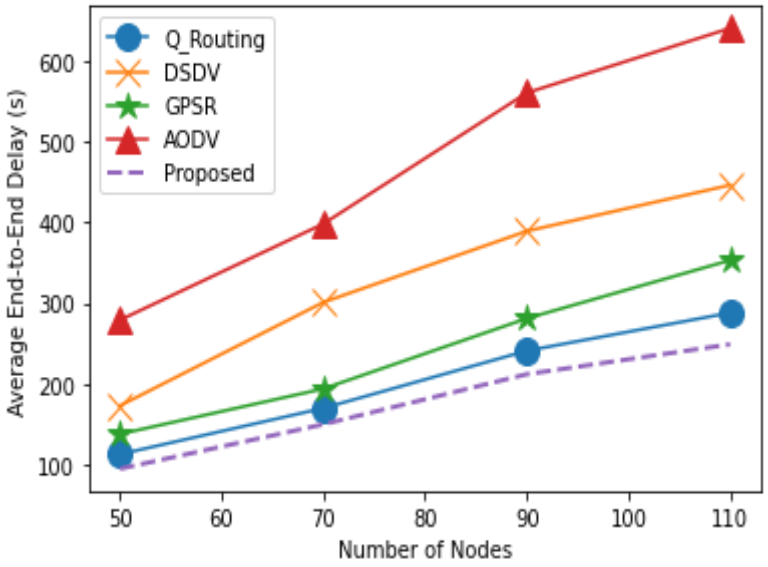

Figure 3: Impact of number of nodes on average end-to-end delay

In Figure 3, we compared the average end-to-end delay of routing protocols. Average end-to-end delay is less, when there is less number of nodes in the network; as the number of nodes increases in a network, the proposed can explore more optimal next-hop

nodes and can avoid local maximum problem, results in less end-to-end delay compared to the other routing protocols. Experimental results suggest that when number of nodes is 110, end-to-end delay is high in AODV. It is the critical value of nodes in the network that produces bad output in terms of packet delivery ratio and average end-to-end delay.

In the second simulation setup, we increased the maximum speed of nodes in the network from 10 to 25 m/s. In Figure 4, the results are plotted between maximum speed of nodes and packet delivery ratio. The experimental results of the proposed show that, as the maximum speed of nodes increases to 25 m/s, the packet delivery ratio increases to 88%. With Q_Routing, the packet delivery ratio is 78%, and then it increases with the increase in maximum speed. When the maximum speed increases to 25 m/s the packet delivery ratio gradually increases to 85%. Based on the observations, the use of larger maximum speed can increase the performance of ad hoc networks under the routing protocols.

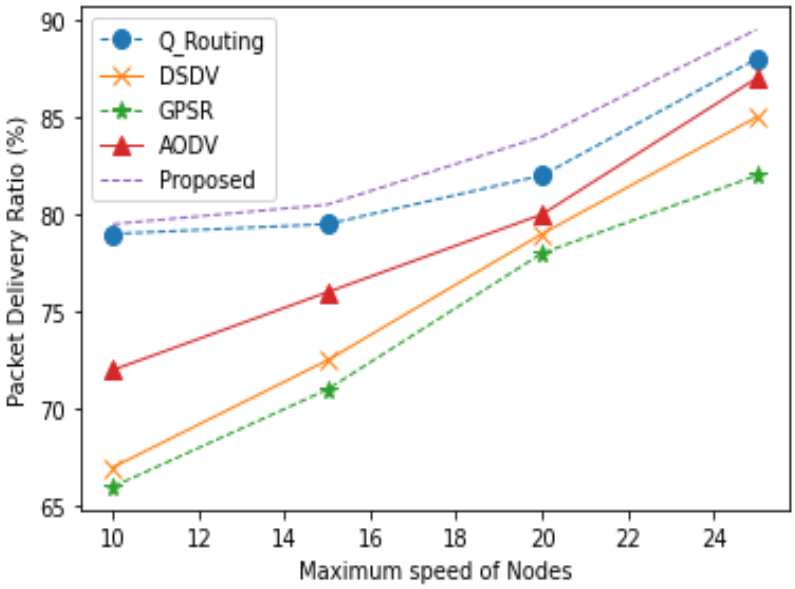

Figure 4: Impact of maximum speed of nodes on packet delivery ratio

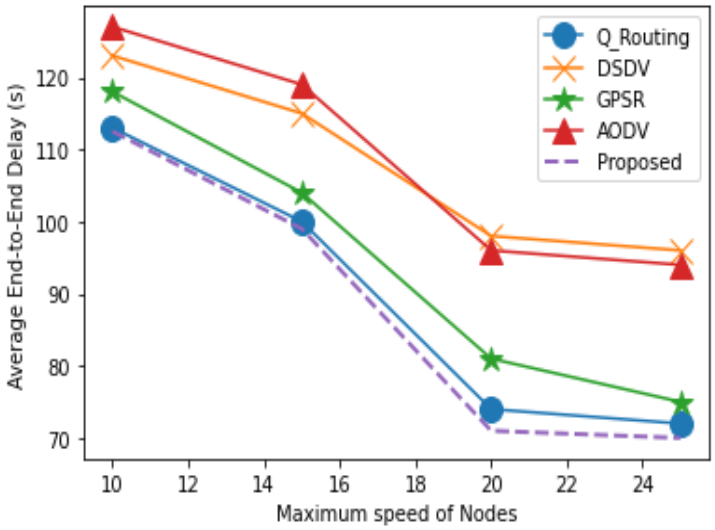

Figure 5: Impact of maximum speed of nodes on average end-to-end delay

The graph in Figure 5 illustrates the relationship between maximum speed of nodes and average end-to-end delay. In both of the algorithms of Q-Routing and the proposed the delay decreases steeply as the maximum speed of nodes reaches 20 m/s. When the maximum speed of nodes increases to 20 m/s, delay is 81 ms, 75 ms, and 72 ms in GPSR, Q-Routing and the proposed, respectively. As the maximum speed of nodes increases to 25 m/s, the delay is 98 ms and 99 ms under AODV and DSDV respectively which is greater than GPSR, Q-Routing and the proposed implementation.

V. CONCLUSION

We have proposed a Q-learning routing algorithms with a reward function that calculated based on link reliability. The geographic routing is enhanced by reinforcement learning an optimal next-hop forwarders using Q-learning routing algorithms. We analyzed the performance of the proposed for urban VANET scenarios with various number and speed of nodes. The large number of nodes is the critical value of nodes in the network that produces bad output in terms of packet delivery ratio and average end-to-end delay. However, the use of larger maximum speed can increase the performance of the routing protocols under urban VANET scenarios.

The foremost advantage is the performance of the proposed compares AODV and DSDV. The routing problem in VANETs requires the optimization of many conflicting objectives. This work can be further extended by applying exportation of Q-learning approach and exploitation in all stages of reward function to improve the routing performance.

REFERENCES:

1. Anitha, V. K., & Akilandeswari J. *Self-Adaptive Trust Based ABR Protocol for MANETs Using Q-Learning. The Scientific World Journal.* . 2014. <http://dx.doi.org/10.1155/2014/452362>.
2. Arnau, R. S., & Fatemeh A. *Fully-echoed Q-routing with Simulated Annealing Inference for Flying Adhoc Networks. IEEE Transactions on Network Science and Engineering, 99(1), doi:10.1109/TNSE.2021.3085514.* 2021.
3. Benjamin, S., & Cedrik S. *PARRoT: Predictive Ad-hoc Routing Fueled by Reinforcement Learning and Trajectory Knowledge. IEEE 93rd Vehicular Technology Conference (VTC-Spring), journals/corr/abs-2012-05490.* 2020. < https://arxiv.org/abs/2012.05490>.
4. Ehsan, M., & Amir N. "An Enhanced Dynamic Source Routing Algorithm for the Mobile Ad-Hoc Network using Reinforcement learning under the COVID-19 Conditions." *Journal of Computer Science, 16 (10), 1477-1490, DOI: 10.3844/jcssp.2020.1477.1490.* (2020).
5. Elmustafa, S. A., & Mohammad K. H. *Machine Learning Technologies for Secure Vehicular Communication in Internet of Vehicles: Recent Advances and Applications. Security and Communication Networks,* . 2021. <https://doi.org/10.1155/2021/8868355.>.
6. Fan, Y., & Jaogao W. *A Double Q-Learning Routing in Delay Tolerant Networks. International Conference on Communications (ICC), 1-6. doi:10.1109/ICC.2019.8761526.* 2019.
7. Jae, H. C., & Hayoun L. *Dynamic Topology Model of Q-Learning LEACH Using Disposable Sensors in Autonomous Things Environment. Appl. Sci. 10(24), doi:10.3390/app10249037.* 2020.
8. Jianmin, L., & Qi W. *QMR:Q-learning based Multi-objective optimization Routing protocol for Flying Ad Hoc Networks. Computer Communications, Elsevier, 150, 304-316. doi: 0.1016/j.comcom.2019.11.01. Hal-02970649*. 2020.
9. Jinqlao, W., & Min F. H. *RSU-Assisted Traffic-Aware Routing Based on Reinforcement Learning for Urban VANETts. IEEE Access, 8. 5733-5748. doi:10.1109/ACCESS.2020.2963850*. 2020.
10. Jun, Y., & Hesheng Z. *Learning-based Routing Approach for Direct Interactions between Wireless Sensor Network and Moving Vehicles. Proceedings of the 16th International IEEE Annual Conference on Intelligent Transportation Systems (ITSC 2013), 1971-1976, doi:10.1109/ITSC.2013.6*. 2013.
11. Karp.B., & Kung H.T. *GPSR: Greedy perimeter stateless routing for*. 2000.
12. Le, L., & Hao Y. *Spectrum Sharing in Vehicular Networks Based on Multi-Agent Reinforcement Learning. Journal Selection Areas in Communications, 37(10), 2282-2292.* 2019.
13. Li, R,.& Li F. *QGrid: Q-learning based routing protocol for vehicular ad hoc networks. IEEE 33rd International Performance Computing and Communications Conference (IPCCC),1-8. doi:10.1109/PCCC.2014.7017079.* 2014.
14. Mujeeb, U. R., & Dost M. K. *A Novel Density-based Technique for Outlier Detection of High Dimensional Data Utilizing Full Feature Space. Information Technology and Control, 50(1), 138-152.* 2021.
15. Paúl, V., & Jack B. "A Framework for Trustworthy Exchanges of Information in VANETs Based on Blockchain and a Virtualization Layer." *Journals Applied Sciences, 10(21), Doi: 10, 7930; doi:10.3390/app10217930.* (2020).
16. Perkins, C., & Belding-Royer C. *RFC3561: Ad Hoc On-Demand Distance Vector (AODV) Routing. RFC Editor, USA, @rfc{10.17487/RFC3561.* 2003.
17. Rahman., & Zukarnain Z. " Performance comparison of AODV, DSDV and I-DSDV routing protocols in mobile ad hoc networks.European." *journal of scientific research,31, 566-576.* (2009).
18. Rahul, D., & Patil B. D. "Analysis of Reinforcement Based Adaptive Routing in MANET." *Indonesian Journal of Electrical Engineering and Computer Science, 2(3), 684 ~ 694.* (2016).
19. —. *Enhanced Confidence Based Q Routing for an Ad Hoc Network American Journal of Educational Science, 1(3), 60-68.* 2015. <http://www.aiscience.org/journal/ajes.>.
20. Rezoan, A. N.,& Sangman M. *Reinforcement Learning-Based Routing Protocols for Vehicular Ad Hoc Networks: A Comparative Survey. in IEEE Access, 9, 27552-27587, doi: 10.1109/ACCESS.2021.3058388.* 2021.
21. Ribal, A., & Chadi A. *Deep Reinforcement Learning-based Scheduling for Roadside Communication Networks. International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), 1-8, doi: 10.23919/WIOPT.2017.7959912.* 2017.
22. Tauqeer, S., & Malik M. H. *Reinforcement Learning-Based Routing Protocol to Minimize Channel Switching and Interference for Cognitive Radio Networks. Hindawi Complexity,*. 2020. < https://doi.org/10.1155/2020/8257168>.
23. Yang, X., & Zhang M. *V2V Routing in VANET Based on Heuristic Q-Learning. International Journal of Computers Communications & Control,15(5),*. 2020. <https://doi.org/10.15837/ijccc.2020.5.3928.>.
24. Zoubir, M. *Reinforcement Learning Based Routing in Networks: Review and Classification of Approaches. Translations and content mining are permitted for academic research only, 2169-3536.* 2019.