Journal of Engineering Science and Technology
Special Issue on ICIT2022 Vol. 19, No. 1 (2024) 121 - 130

DISTRIBUTED RESOURCE ALLOCATION IN V2V COMMUNICATION USING MULTI-AGENT DEEP REINFORCEMENT LEARNING

FARHAN AADIL[1], SABAHAT[2],
MUHAMMAD FAHAD KHAN[3], RAJERMANI THINAKARAN[4*]

[1,2,3]Department of Computer Science, COMSATS University Islamabad, Attock
[4]Faculty of Data Science and Information Technology, INTI International University, Malaysia
Corresponding author: rajermani.thina@newinti.edu.my

Abstract

Vehicular ad-hoc networks (VANET) enable vehicles to connect to the wireless network. In VANET, mobility is high, and the distribution of vehicles is uneven, leading to topology changes and disconnections of the network. In 5G, distributed resource allocation has an essential role in vehicle-to-vehicle (V2V) communication. Every node autonomously selects resources to disseminate Cooperative Awareness Messages (CAM). Due to limited resources, there is a challenge for distributed resource allocation in an urban scenario with traffic jams and providing a safer driving experience. In mode 4, nodes reserve the resources based on their local observations using semi-persistent scheduling (SPS). When two nodes, select the same resources, it will arise the resource contention problem. To overcome this problem, we proposed a distributed resource allocation for node communication based on multi-agent deep reinforcement learning. In this model, nodes are considered autonomous agent that makes their decisions based on local observation without any global information. It increases the packet reception ratio of a V2V communication. The experience of every agent is stored in the memory to be exploited for training. We train a model which is shared with other nodes and learn from experiences. Our results show that the performance of the network improves, aiming to achieve a higher Packet Reception Ratio (PRR), reduced collision, and enhanced network performance desired greatly.

Keywords: 3GPP, Cooperative awareness message, Deep reinforcement learning, Distributed resource allocation, Process innovation, Vehicular network.

1. Introduction

In this era of the world, we have reduced human effort by automating every task. V2V communications have become vital technology. It reduces road accidents and provides a safer driving experience. In our scenario, we are taking vehicles as a node. In a disaster scenario, packet delivery and active network communication are highly desirable. Reduce congestion and ensure utilization of resources of all nodes.

The revolution of automated nodes is about to begin, and a milestone is wireless communication between nodes [1]. There are two standards by international companies, IEEE 802.11p and short-range cellular-vehicle-to-anything (C-V2X), and both are under discussion. The former is only decentralized and based on sensing before transmitting, while later managed by an infrastructure based on orthogonal resources. Studies have been showing the advantages and disadvantages but still have doubts.

The standard Cooperative Awareness Messages (CAM) is one of the components that is defined by the European Telecommunication Standards Institute (ETSI) for disseminating information about vehicles [2]. In vehicular safety communication, each node broadcasts CAMs which contain the id, position, velocity, and direction of nodes with the emergency node and warnings of collision. OFDMA is an orthogonal frequency-division multiplexing digital modulation. Multiple access in OFDMA by assigning subcarriers to individual users. OFDMA is ideal bandwidth is low, reuse frequency, and increased efficiency.

For safety applications, 3GPP is working on 5G-V2X to support advanced driving applications. C-V2X supports both mode 3 and mode 4 communication. In mode 4, nodes reserve the resources based on their local observations using SPS. Each message needs a certain number of resources (time and frequency), depending on its size. High mobility of nodes leads to varying environments, so usage of conventional resource allocation methods which are mostly designed for static or low-mobility environment assumptions [3, 4].

To propose a technique that can reduce collision and enhance network performance mitigating the issue of resource allocation. We aimed to deal dynamic behaviour of nodes with unique resource allocation. But in reality, the situation is quite different. Nodes have scarce resources (e.g. bandwidth, processing power, frequency, and buffer) [5] and to save those resources [6, 7]. In case of dealing with some emergency situation, cooperation of nodes and congestion-free resources is highly desirable [8].

In V2V networks, mobility is high which causes rapid changes and in a short timescale, it’s not possible to assemble or keep track of full CSI. The traditional resource management approaches are hard to use. To deal with the node’s high mobility, a centralized allocation scheme was dependent on slowly varying the large-scale fading information [9]. Due to the interference, system performance degrades need to discover through approaches like multi-agent DRL for distributed resource assignment [10].

In many applications, deep learning techniques are also applied in the resource allocation area. Each node is act as agent and spectrum and select messages according to learn policy. We address the problem of V2V communication, action and designed the function of reward for the broadcasting scenarios. In order to

resolve the congestion problem and reduce delay, we use distributed resource allocation which is dependent on Multi-Agent Deep Reinforcement Learning.

2. Literature review

Distributed resource allocation in V2X communication has been standardized by two key radio access technologies like Cellular-V2X and Dedicated Short Range Communications (DSRC). This technology relies on 802.11p standard for medium access control (MAC) layers which deploy Carrier Sense Multiple Access [11-13]. Our analysis showed that near nodes are likely to choose the same resources [14].

In this paper, performs well as compared to traditional proportional fair scheduling, and reduced cost computational but not be able to adjust the time-varying traffic demands promptly [15-17]. Resource management is performed in a centralized manner and each node reports the local channel information to the central controller [6]. V2V communication plays an important role in intelligent transportation and road safety [18, 19]. The author uses DRL to deal with resource allocation. The proposed method is decentralized and not required global information in the Dense Urban scenario [20, 21].

To address the issue of congestion control, Mao et al. [22] define a range of message transfers which is the responsibility of the node. It is convenient to accept messages, forwarded and delivered them to the concerning destination. The proposed algorithm prohibits the node's selfish behaviour, decreases congestion, and ensures sharing of the resource by all nodes. Xu et al. [23] address the problem of the resource allocation between the nodes. Nodes communicate inside delimited out of coverage area (DOCA) and infrastructure connected with a centralized scheduler. This method is not for complex environments. A centralized scheduler is not capable enough to provide assignments immediately to the nodes that are not in the coverage.

As of the study Salahuddin et al. [24], it reduced the interference of signal by enabling zone formation. MADRL has been used in network problems. When number of nodes increases, number of nodes links increases, and not scaling computationally [25]. Ashraf et al. [20] and Cecchini et al. [26] found that deep reinforcement learning framework satisfied different resource requirements. As described from the following studies [15, 27], deep learning technique that solved the problem of resource allocation, and dynamically maximizes the rewards.

Distributed resource allocation can also be modelled in multi-agent. Node chooses the frequency and time resources for the transmission. To offer distributed solutions, multi-agent reinforcement learning is applied to various problems i.e. the decisions based on local information [28]. Sensing-Based SPS performance analysis [29]. The authors analyses Mode 4 resource pool configuration and the performance of SPS.

3. Proposed Methodology

3.1. Simulation environment

In this work, we used python programming language for the development and to test the algorithms as fast and as efficiently as possible. Python programming is faster to implement and test new ideas than the C++ language.

3.2. System overview

In literature, many schemes have been proposed for simulating and encouraging cooperation among nodes. Moreover, our work is intended to ensure the utilization of resources of all nodes in a network. Every node required resources to disseminate cooperative awareness messages. Each message needs a certain number of resources (time and frequency), depending on its size. Conventional resource allocation methods are mostly designed for static or low mobility. Multi-Agent DRL introduced resource allocation in V2V communication. The proposed method is based on a controlled entity which is act as a base station and does not require global information. Orthogonal frequency division multiple access (OFDMA) provides a higher frequency band and low transmission power.

Resource pool consists of resource blocks. When a User Equipment (UE) transmits a packet and decides when to transmit. When UEs select the same resource block and UEs will interfere with each other, and a collision happens. The proposed method is based on a controlled entity which is act as a base station and not required global information. MAC layer performed scheduling of packets. When node's MAC layer receives CAM messages from upper layer, it requests a resource for packet transmission from the MARL agent. The communication between the agent and node is realized through the python interface. After training i.e. decentralized execution, the same trained model is stored in the MAC layer of each node. Each node cooperates with the other nodes to maximize its objective. At each time-step action i.e. resource blocks of the nodes are determined by the MARL agent. Following are the parameters in Table 1.

Table 1. Parameters.

Name	Value
Access scheme	OFDMA
Carrier frequency	5.9 GHz
System bandwidth	10 MHz
Subcarrier spacing	30 kHz
PRBs	24
Transmit power	23.0 dBm
TFC index	4 (QPSK)
CAM message size	300 bytes
Periodicity	100 ms
Time-step	250000
Experience replay	1024
Batch size	512
Learning rate	0.0001
Hidden layers	256 neurons
Optimizer	ADAM
Activation function	ReLU
Discount factor	0.7
ε –decay	0.999

3.3. Deep Q network

Deep Q Network (DQN) is a model-free RL algorithm. It is a modern deep learning technique. DQN algorithm uses Q-learning, so that it can learn the best action in a given state and a deep neural network to estimate the Q value function. The state

and observation is the input of the neural network, and the number of actions shows that the agent can take.

In our system, each node acts as an agent. An agent observes a state s_t, from state space S, and the agent takes action accordingly which is based on policy π. An agent takes an action a_t, moves toward the next state $s_{r+1,}$ and received reward r_t. The goal of the agent is to maximize the accumulative reward, $G_t = r_t + \gamma G_{t+1.}$ γ is a discount factor $\gamma \in [0,1]$, which influences the future rewards. We use policy π to define the behaviour of an agent. The policy π, the expected total reward from the state s and action a, can be calculated from the action-value function. The action-value function can be calculated by using the Bellman equation where

$$Q(s, a) = Q(s, a) + \alpha\,[r + \gamma \max Q(s',a') - Q(s, a)] \quad (1)$$

Deep neural networks (DNN) are used when state and action space is large. We called this technique Deep Q Network (DQN), and Q-function is Q(s, a; θ) where θ represent the trainable weight of network. The least-square loss is defined as,

$$L(\theta) = [(r + \gamma \max Q(s' , a' ; \theta)) - Q(s, a; \theta)]^2 \quad (2)$$

Each node observes the position of other nodes on the road so that resources are allocated for CAM transmission. Each agent evolves its hidden state although they have the same DQN. We train only one model and shared it with other nodes and learn from experiences. DQN shows better results as compared to DDQN.

We deploy additional enhancements to stabilize the learning and improve the policy. DQN resolves the problem of convergence value function estimation faster training. DQN doesn't use ground-truth data to train the Q-value estimator. Thus, we adopt DQN as compared to DDQN because it drastically slows down learning and increases complexity. It performs poorly on problems with large action space.

3.4. Multi-agent deep reinforcement learning

Due to its dynamic nature, we formulate a multi-agent DRL, Fig. 1, in which more than one node is involved. It allows agents to make decisions. Agents behave distinctly due to different observations. In the training part, each node observes the position of other nodes, and determine the reward. After training, agents exploit the trained policy to make its decision. It maximizes the Packet Reception Ratio (PRR) and minimizes delay. In the beginning, the learning rate is 0.0001 and also utilizes the e-greedy policy and adaptive moment estimation method (Adam) for training.

Fig. 1. Deep Q Network.

DRL is used to deal with the high dynamicity of nodes. The multi-agent DRL is used to select resources for transmission of data/packet; this scheme enhances communication reliability and mitigates the network load. If different nodes independently make decisions and select the same resource block then V2V links may lead to collisions.

When an agent has the information about neighbours, easily finds out that using the same resource block leads to low rewards, hence avoiding this to increase the rewards. By using this scheme, to select the packet to transmit; our scheme deals with the network load and enhance the communication. This research addresses the resource allocation problem in V2V with higher PRR and minimum delay. So our scheme is novel in the sense that it encourages packet shifting rather than dropping and ensures awarding credits to every node who participated in process of packet delivery. As nodes increase, the node's links also increase and are not computationally scalable. The decisions of each node as a single decision can overcome this scaling problem. Describe the deployment of nodes in the environment and every node autonomously selects resources to disseminate Cooperative Awareness Messages (CAM). Every node shares its position information with each other. Nodes periodically broadcast CAM. Each message needs a certain number of resources (time and frequency), depending on its size.

Each node observes the position of other nodes on the road; it can allocate a resource for CAM transmission. We consider a model if two or more nodes exploit the same resource block within the range of the receivers; the receivers decode the packets of a closer node. Nodes act as an agent and observe state s_t, from state space S, and the agent takes an action a_t accordingly based on policy π. Resources are allocated with the help of multi-agent deep reinforcement learning. An agent takes an action, moves toward the next state $s_{t+1,}$ and gets a reward r_t. Each agent evolves its hidden state although they have the same DQN. To predict the mobility pattern of the nodes, we use the LSTM networks as a first layer and combine the observation over time.

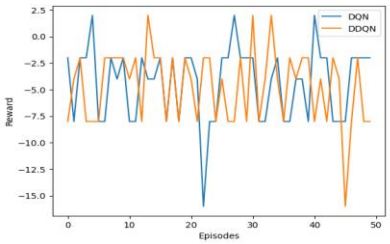
The performance evaluated based on Packet Reception Rate (PRR), reward, and reduced collision. It also reduces the error rate and lowers queue delay. Our performance analysis is based on a highway scenario involving faster dissemination of messages among the connected vehicles and efficient resource utilization. The experience of every agent is stored in the memory to be exploited for training. In proposed scheme, every node in the network will receive some positive reward for successful transmission and a negative reward for congestion.

4. Results and Discussion

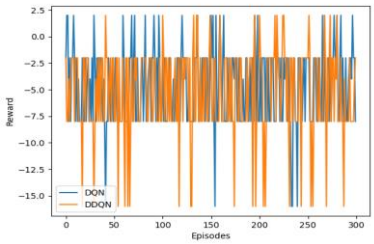
We consider a wireless vehicular network. The network consists of a set of nodes N = {1,2,3,…,N}, and moving along a road. Their distance to every node is changing dynamically. Each node autonomously select a resource from resource block R = {1,2,3,…,R}. First, we evaluated the performance of proposed approach for faster analysis. We considered a model, if two or more nodes, select the same resources, it arises the resource contention problem. A time-slotted system is considered, in which all nodes are scheduled simultaneously at the time t for the allocation of resources in R. We maintain the high reliability of periodic broadcast messages. Each node present in the environment strives to increase the number of nodes that decodes its packet.

Table 1 shows the parameters. We evaluate the performance of the proposed work as shown in Table 1. We used an ADAM optimizer for training. It works efficiently when we are working with a large problem including a large amount of data. ReLU is a non-linear function and outputs the input directly if it is positive, otherwise the output is zero. The learning rate is 0.0001, the discount factor is 0.7 and ε-decay is 0.999. We trained the model with ε-greedy policy 25000 time-steps.

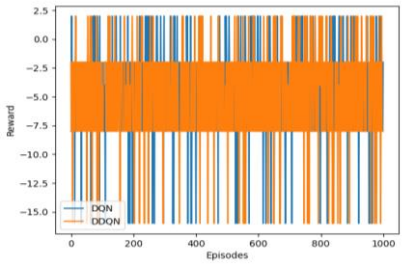
The proposed method works well when a large number of nodes are present in the environment. DQN resolves the problem of convergence value function estimation faster training; don't use ground-truth data to train the Q-value estimator. While DDQN drastically slows down learning and increases sample complexity. It performs poorly on problems with large action space. Figures 2(a)-(c) show that the proposed approach has good results as compared to DDQN. It also performed well with a large number of nodes.



(a) Scenario.



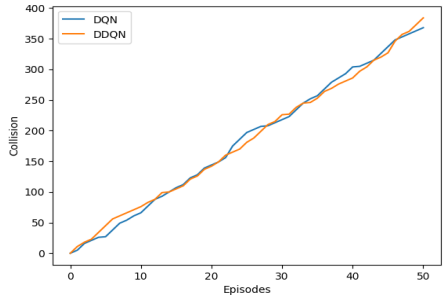
(b) Scenario.



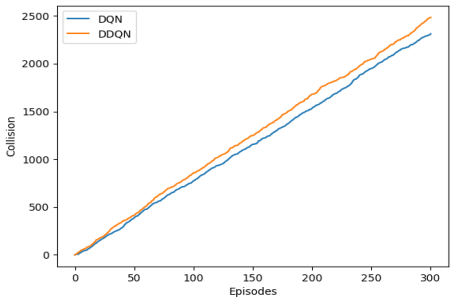
(c) Scenario.

Fig. 2. Performance of congested scenarios.

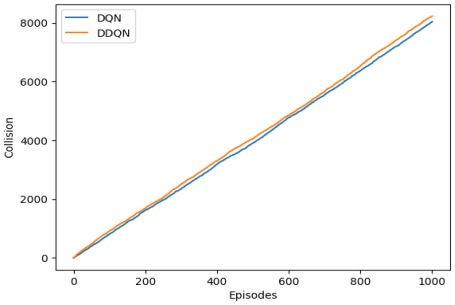
As seen in Figs. 3(a)-(c), the system achieves higher PRR values, reduce collision as compared to DDQN. It can be observed that DQN is quicker than DDQN as it uses fewer numbers of parameters. The performance of DQN algorithm is also affected by changing the values of parameters such as replay memory size and batch size selected for experience replay.



(a) Scenario.



(b) Scenario.



(c) Scenario.

Fig. 3. Performance analysis.

5. Conclusion

Due to limited resources, there is a challenge for distributed resource allocation. The neighbouring nodes can benefit from those messages to reduce road accidents. Congested resource allocation needs to be developed for each node, to provide a unique resource. To overcome this problem, we proposed a distributed resource allocation for vehicle-to-vehicle communication based on multi-agent deep reinforcement learning. Our proposed method DQN is compared with DDQN, and our approach shows better results in congested scenario. We adopt DQN as compared to DDQN because it drastically slow down learning and increases complexity. By using proposed approach, the performance of the network improved, achieving a higher Packet Reception Ratio (PRR), reduced collision and enhanced network performance desired greatly.

References

1 Bazzi, A.; Cecchini, G.; Menarini, M.; Masini, B.M.; and Zanella, A. (2019). Survey and perspectives of vehicular Wi-Fi versus sidelink cellular-V2X in the 5G era. *Future Internet*, 11(6), 122.

2 Ben Hamida, E.; Noura, H.; and Znaidi, W. (2015) Security of cooperative intelligent transport systems: Standards, threats analysis and cryptographic countermeasures. *Electronics*, 4(3), 380-423.

3 Rizwan, M.; Aadil, F.; Durrani, M.Y.; and Thinakaran, R. (2023). Online signature verification for forgery detection. *International Journal of Advanced Computer Science and Applications*, 14(3), 478-484.

4 Liang, L.; Ye, H.; and Li, G.Y. (2018). Towards intelligent vehicular networks: A machine learning framework. *IEEE Internet of Things Journal*, 6(1), 124-135.

5 Ansa, G.; Cruickshank, H.; Sun, Z.; and Alshamrani, M. (2017). A security scheme to mitigate denial of service attacks in delay tolerant networks. *Journal of Computer Sciences and Applications*, 5(2), 50-63.

6 Jiang, Q.; Men, C.; and Tian, Z. (2016). A credit-based congestion-aware incentive scheme for DTNs. *Information*, 7(4), 71.

7 Chakrabarti, C.; Banerjee, A.; and Roy, S. (2014). An observer-based distributed scheme for selfish-node detection in a post-disaster communication environment using delay tolerant network. *Proceedings of the* 2014 *Applications and Innovations in Mobile Computing* (*AIMoC*), Kolkata, India, 151-156.

8 Silva, A.P.; Burleigh, S., Hirata, C.M.; and Obraczka, K. (2015). A survey on congestion control for delay and disruption tolerant networks. *Ad Hoc Networks*, 25 (Part B), 480-494.

9 Liang, L.; Li, G.Y.; and Xu, W. (2017). Resource allocation for D2D-enabled vehicular communications. *IEEE Transactions on Communications*, 65(7), 3186-3197.

10 Sun, W.; Ström, E.G.; Brännström, F.; Sou, K.C.; and Sui, Y. (2015). Radio resource management for D2D-based V2V communication. *IEEE Transactions on Vehicular Technology*, 65(8), 6636-6650.

11 Hassan, M.I.; Vu, H.L.; and Sakurai, T. (2011). Performance analysis of the IEEE 802.11 MAC protocol for DSRC safety applications. *IEEE Transactions on Vehicular Technology*, 60(8), 3882-3896.

12 Nomor Research (2018). Comparison of v2x based on 802.11p, lte and 5g. *Technical report*, Nomor Research GmbH, Germany, White Paper.

13 5G Automotive Association (2018). *V2X Technology Benchmark Testing*. 5GAA.

14 Tanenbaum, A.S.; and Wetherall. D.J. (1996). *Computer networks*. (5th ed.). Pearson Education.

15 Gupta, J.K.; Egorov, M.; and Kochenderfer, M. (2017). Cooperative multi-agent control using deep reinforcement learning. *Proceedings of the Autonomous Agents and Multiagent Systems* (*AAMAS* 2017), São Paulo, Brazil, 66-83.

16 Ye, H.; and Li, G.Y. (2018). Deep reinforcement learning based distributed resource allocation for V2V broadcasting. *Proceedings of the* 2018 14*th International Wireless Communications & Mobile Computing Conference* (*IWCMC*), Limassol, Cyprus, 440-445.

17 Shi, K.; and Gu, X. (2021). Performance of V2V communication distributed resource allocation scheme in dense urban scenario. *Proceedings of the* 2021 *IEEE Wireless Communications and Networking Conference* (*WCNC*), Nanjing, China, 1-6.

18 Jiang, Q.; Men, J.; Tian, Z.; and Meijuan, J. (2016). A congestion-aware node cooperation mechanism based on double auction for opportunistic networks. *International Journal of Future Generation Communication and Networking*, 9(10), 105-122.

19 Wang, C.; Gong, Z.; Wu, C.; Zhao, B.; and Zhang, Z. (2012). CRASP: congestion control routing algorithm against selfish behavior based on pigeonhole principle in DTN. *Proceedings of the* 2012 *IEEE* 9*th International Conference on Mobile Ad-Hoc and Sensor Systems* (*MASS* 2012), Las Vegas, NV, USA, 1-6.

20 Ashraf, M.I.; Bennis, M.; Perfecto, C.; and Saad, W. (2016). Dynamic proximity-aware resource allocation in vehicle-to-vehicle (V2V) communications. *Proceedings of the* 2016 *IEEE Globecom Workshops* (*GC Wkshps*), Washington, DC, USA, 1-6.

21 Nasir, Y.S.; and Guo, D. (2019). Multi-agent deep reinforcement learning for dynamic power allocation in wireless networks. *IEEE Journal on Selected Areas in Communications*, 37(10), 2239-2250.

22 Mao, H.; Alizadeh, M.; Menache, I.; and Kandula, S. (2016). Resource management with deep reinforcement learning. *Proceedings of the* 15*th ACM Workshop on Hot Topics in Networks*, Atlanta, Georgia, USA, 50-56.

23 Xu, Z.; Wang, Y.; Tang, J.; Wang, J.; and Gursoy, M.C. (2017). A deep reinforcement learning based framework for power-efficient resource allocation in cloud RANs. *Proceedings of the* 2017 *IEEE International Conference on Communications* (*ICC*), Paris, France, 1-6.

24 Salahuddin, M.A.; Al-Fuqaha, A.; and Guizani, M. (2016). Reinforcement learning for resource provisioning in the vehicular cloud. *IEEE Wireless Communications*, 23(4), 128-35.

25 Ye, H.; Li, G.Y.; and Juang, B.H. (2019). Deep reinforcement learning based resource allocation for V2V communications. *IEEE Transactions on Vehicular Technology*, 68(4), 3163-73.

26 Cecchini, G.; Bazzi, A.; Masini, B.M.; and Zanella, A. (2017). Performance comparison between IEEE 802.11p and LTE-V2V in-coverage and out-of-coverage for cooperative awareness. *Proceedings of the* 2017 *IEEE Vehicular Networking Conference* (*VNC*), Turin, Italy, 109-114.

27 Hüttenrauch, M.; Šošić, A.; and Neumann, G. (2017). Guided deep reinforcement learning for swarm systems. arXiv preprint arXiv:1709.06011.

28 Sukthankar, G.; and Rodriguez-Aguilar, J.A.(2017). Autonomous agents and multiagent systems. *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems* (*AAMAS* 2017), São Paulo, Brazil.

29 Nabil, A.; Kaur, K.; Dietrich, C.; and Marojevic, V. (2018). Performance analysis of sensing-based semi-persistent scheduling in C-V2X networks. *Proceedings of the* 2018 *IEEE* 88*th vehicular technology conference* (*VTC-Fall*), Chicago, IL, USA, 1-5.