

Cooperative perception in Vehicular Networks using Multi-Agent Reinforcement Learning

Mohamed K. Abdel-Aziz*, Sumudu Samarakoon*, Cristina Perfecto[†], and Mehdi Bennis*

*Centre for Wireless Communications, University of Oulu, Finland

[†]University of the Basque Country UPV/EHU, Spain

E-mails: {mohamed.abdelaziz, sumudu.samarakoon, mehdi.bennis}@oulu.fi, cristina.perfecto@ehu.eus

Abstract—Cooperative perception plays a vital role in extending a vehicle’s sensing range beyond its line-of-sight. However, exchanging raw sensory data under limited communication resources is infeasible. Towards enabling an efficient cooperative perception, vehicles need to address fundamental questions such as: what sensory data needs to be shared? at which resolution? In this view, this paper proposes a reinforcement learning (RL)-based content selection of cooperative perception messages by utilizing a quadtree-based point cloud compression mechanism. Furthermore, we investigate the role of federated RL to enhance the training process. Simulation results show the ability of the RL agents to efficiently learn the message content selection that maximizes the satisfaction of the vehicles in terms of the received sensory information. It is also shown that federated RL improves the training process, where better policies can be achieved within the same amount of time compared to the non-federated approach.

Index Terms—V2V, cooperative perception, federated reinforcement learning, quadtree.

I. INTRODUCTION

Nowadays, each vehicle is equipped with many sensors (e.g., RADAR, LiDAR, and cameras) whose quality varies widely. These sensors promote a wide range of applications that support and enhance the driving experience. However, the accuracy of these sensors is affected by weather conditions, blind spots, as well as sensors’ defects, all of which jeopardize these applications. In order to overcome this issue, vehicle-to-vehicle (V2V) communication, a promising facilitator for intelligent transportation systems, can be exploited. V2V communication is [1]. V2V can expedite the exchange of sensory information between vehicles to enhance the perception of the surrounding environment beyond their sensing range; such process is called *cooperative perception* [2]. The advantages of cooperative perception are validated in [3] demonstrating that it greatly improves the sensing performance. Motivated by its potential, several standardization bodies are currently focusing their efforts towards formally defining the cooperative perception message (CPM), its contents and generation rate [2], [4]. In addition, a growing body of literature has explored the use of cooperative perception in various ways [5]–[7]. In [5], the authors investigated which information

should be included within the CPMs to enhance a vehicle’s perception reliability. Cooperative perception from the sensor fusion point-of-view is tackled in [6] and a hybrid vehicular perception system that is able to fuse both local onboard sensor data as well as data received from a multi-access edge computing (MEC) server is proposed. Finally, the authors of [7] conducted a study focusing on raw-data level cooperative perception for enhancing the detection ability of self-driving systems; whereby sensory data collected from different positions and angles of connected vehicles are fused. Though interesting, neither of these works perform an in-depth analysis of the wireless connectivity impact.

Cooperative perception over wireless networks cannot rely on exchanging raw sensory data, due to the limited communication resources [2]. Therefore, this raw sensory data should be compressed efficiently to save both the storage and the available communication resources. One possible technique that could be useful for such spatial raw sensory data is called *region quadtree* [8]. Region quadtree is a tree data structure used to efficiently store data on a two-dimensional space. A quadtree recursively decomposes the two-dimensional space into four equal sub-regions (blocks) till all the locations within a block have the same state or till reaching a maximum predefined resolution (tree-depth). Tailoring the number and resolution of the transmitted quadtree blocks to bandwidth availability is a challenging problem.

The main contribution of this paper is to study the problem of selecting the contents of the exchanged CPMs between vehicular pairs. Specifically, for a given vehicular pair, the main objective of a vehicle is to learn which sensory information (quadtree blocks) should be transmitted and at which resolution level to maximize the satisfaction of its paired vehicle with the received sensory information. Solving such problem using conventional mathematical tools is complex and intractable. As a result, we resort to machine learning techniques, specifically deep reinforcement learning (DRL), which proved to be useful in such complex situations [9]. Moreover, we study the role of federated learning [10] within our RL scenario, and how it would affect the convergence speed of the learning algorithm. Simulation results show that the policies achieving higher vehicular satisfaction could be learned by the vehicles. It is also shown that federated RL improves the training process, where better policies can be achieved within the same duration compared to the non-

This work was supported in part by the Academy of Finland project CARMA, and 6Genesis Flagship (grant no. 318927), in part by the INFOTECH project NOOR, in part by the EU-CHISTERA projects LeadingEdge and CONNECT, and in part by the Kvantum Institute strategic project SAFARI.

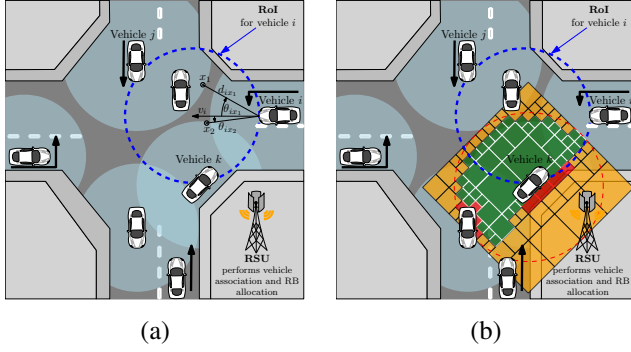


Figure 1. (a) Vehicles under the coverage of a single RSU, drive through a junction while dynamically exchanging sensory information. (b) Quadtree representation of a vehicle's sensing range, with a maximum resolution level $L = 5$. Green represents the unoccupied state s_- , red represents the occupied state s_+ and orange represents the unknown state s_0 .

federated approach.

The rest of this paper is organized as follows. Section II presents the system model. The network-wide problem is formulated in Section III, followed by our proposed federated RL solution within the cooperative sensing scenario, in Section IV. Finally, in Section V, simulation results are presented while conclusions are drawn in Section VI.

II. SYSTEM MODEL

A single RSU providing a coverage to a road junction, as shown in Fig. 1, is considered. Let $\mathcal{N} \triangleq \{1, 2, \dots, N\}$ denotes the set of vehicles served by the RSU. We denote the location of each vehicle $n \in \mathcal{N}$ at time slot t by $l_n(t)$. Without loss of generality, we assume that each vehicle is equipped with a single sensor having a fixed circular range of radius r . Each location sensed by vehicle n can have one of three states: occupied (s_+), unoccupied (s_-), or unknown (s_0). This latter unknown state corresponds to blind-spots, due to occlusion, or to points beyond the limits of the vehicle's sensing range. In this view, the probability of occupancy at location \mathbf{x} with respect to vehicle n is,

$$p_n(\mathbf{x}) = \begin{cases} \lambda_n & \text{if } s_n(\mathbf{x}) = s_+, \\ 1 - \lambda_n & \text{if } s_n(\mathbf{x}) = s_-, \\ 1/2 & \text{if } s_n(\mathbf{x}) = s_0, \end{cases} \quad (1)$$

where $s_n(\mathbf{x})$ is the state of location \mathbf{x} defined by vehicle n , and $\lambda_n \in (0.5, 1]$ corresponds to the reliability of its sensor. Let $q_n(\mathbf{x})$ denotes the worthiness (quality) of the sensed information at location \mathbf{x} that depends on the probability of occupancy $p_n(\mathbf{x})$, and the freshness of the sensed information, which can be quantified by the age of information (AoI) metric $\Delta_n(\mathbf{x})$ [11]. This worthiness is given by,

$$q_n(\mathbf{x}) = |2p_n(\mathbf{x}) - 1| \mu^{\Delta_n(\mathbf{x})}, \quad (2)$$

with a parameter $\mu \in (0, 1)$. Note that $q_n(\mathbf{x})$ decreases as its AoI increases (*outdated* information) or the probability of occupancy approaches 1/2 (*uncertain* information).

Moreover, vehicle n is interested in extending its sensing range by a duration of t_{int} seconds along its direction of

movement which is captured by a circular region of interest (RoI). The RoI of vehicle n has a diameter of $v_n t_{\text{int}}$, where v_n is the velocity of the vehicle. Within the RoI, the vehicle has higher interest in the locations closer to its current position as well as locations closer to its direction of movement over any other location. Therefore, the interest of vehicle n in a location \mathbf{x} is formally defined as follows

$$w_n(\mathbf{x}) = \begin{cases} \frac{v_n t_{\text{int}} \cos \theta - d}{v_n t_{\text{int}} \cos \theta} & d \leq v_n t_{\text{int}} \cos \theta \\ 0 & \text{o.w.,} \end{cases} \quad (3)$$

where d is the euclidean distance between the location \mathbf{x} and the vehicle's position $l_n(t)$, and θ is the angle between the vehicle's direction of motion and location \mathbf{x} , as illustrated on Fig. 1(a). To capture the need of gathering new information, the interest $w_n(\mathbf{x})$ of vehicle n needs to be weighted based on the lack of worthy information, i.e., $1 - q_n(\mathbf{x})$. Hence, the modified interest of vehicle n in location \mathbf{x} is given by,

$$i_n(\mathbf{x}) = w_n(\mathbf{x})[1 - q_n(\mathbf{x})]. \quad (4)$$

Furthermore, a time-slotted communication with transmission slots of duration τ is considered. We assume that vehicles are already associated into pairs and can communicate simultaneously with each other, i.e. each vehicle is equipped with two radios, one for transmitting and other is for receiving. Additionally, a set $\mathcal{K} \triangleq \{1, 2, \dots, K\}$ of orthogonal resource blocks (RBs), with bandwidth ω per RB, is shared among the vehicles, where each radio is allocated with only one RB.

Let $h_{nn'}^k(t)$ be the instantaneous channel gain, including path loss and channel fading, from vehicle n to vehicle n' over RB k in slot t . We consider the 5.9 GHz carrier frequency and adopt the realistic V2V channel model of [12] in which, depending on the location of the vehicles, the channel model is categorized into: line-of-sight, weak-line-of-sight, and non-line-of-sight. Thus, the data rate from vehicle n to vehicle n' on time slot t (in packets per slot) is expressed as

$$R_{nn'}(t) = \frac{\tau}{M} \sum_{k \in \mathcal{K}} \omega \log_2 \left(1 + \frac{Ph_{nn'}^k(t)}{N_0 \omega + I_{nn'}^k(t)} \right), \quad (5)$$

where M is the packet length in bits, P is the transmission power per RB, and N_0 is the power spectral density of the additive white Gaussian noise. Here, $I_{nn'}^k(t) = \sum_{i,j \in \mathcal{N}/n,n'} Ph_{in'}^k(t)$ indicates the received aggregate interference at the receiver n' over RB k from other vehicles transmitting over the same RB k .

Finally, in order to save the available communication resources, we assume that the region quadtree compression technique is utilized by each vehicle [8]. Within this technique, each vehicle converts its sensing range into a squared-block of side-length $2r$. This block is divided recursively into 4 blocks until reaching a maximum resolution level L or until the state of every location \mathbf{x} within a block is the same. Without loss of generality, we assume that each block can be represented using M bits. Fig. 1(b) shows the quadtree representation of the sensing range of a vehicle when $L = 5$.

The state of a block b within the quadtree of vehicle n is said to be occupied if the state of any location \mathbf{x} within the

block is occupied, while the state of a block is said to be unoccupied if every location within the block is unoccupied. Otherwise, the block would have an unknown state. Let $\mathcal{B}_n(t)$ represents the set of quadtree blocks available for transmission by vehicle n at time slot t . For simplicity and without loss of generality, we assume that $\mathcal{B}_n(t)$ only contains blocks available from its own sensing range. Due to the quadtree compression, the cardinality of $\mathcal{B}_n(t)$ is upper bounded by: $|\mathcal{B}_n(t)| \leq \sum_{l=0}^{L-1} 4^l = \frac{1-4^L}{1-4}$.

III. PROBLEM FORMULATION

Each vehicular pair exchanges sensory information in the form of quadtree blocks with the objective of maximizing the satisfaction of each vehicle. The satisfaction of vehicle n with the sensory information received from its associated vehicle n' at time slot t can be defined as follows:

$$f_{nn'}(t) = \sum_{b \in \mathcal{B}_{n'}(t)} \sigma_{n'}^b(t) \left(\frac{\sum_{\mathbf{x} \in b} i_n(\mathbf{x})}{\text{Ar}(b)} q_{n'}(b) \right), \quad (6)$$

where $\sigma_{n'}^b(t) = 1$ if vehicle n' transmitted block b at time slot t , and $\sigma_{n'}^b(t) = 0$ otherwise, and $\text{Ar}(b)$ is the area covered by block b . Note that, vehicle n is more satisfied when receiving quadtree blocks with a resolution proportional to the weights of its RoI as per (4), which is captured by $\frac{\sum_{\mathbf{x} \in b} i_n(\mathbf{x})}{\text{Ar}(b)}$. Furthermore, vehicle n is more satisfied when receiving quadtree blocks having more worthy sensory information, which is captured by $q_{n'}(b)$. In this view, our cooperative perception network-wide problem can be formulated as follows:

$$\begin{aligned} \max_{\sigma(t)} \quad & \sum_{n, n' \in \mathcal{N}} f_{nn'}(t) \\ \text{s.t.} \quad & \sum_{b \in \mathcal{B}_n(t)} \sigma_n^b(t) \leq \sum_{n' \in \mathcal{N}} R_{nn'}(t), \quad \forall n \in \mathcal{N}, \forall t \end{aligned} \quad (7a)$$

$$\sigma_n^b(t) = \{0, 1\} \quad \forall t, \forall n \in \mathcal{N}, \quad (7b)$$

where the objective is to select the contents of the transmitted CPMs, $\sigma(t)$, in order to maximize the sum of the satisfaction of the associated vehicular pairs. Note that (7a) defines the maximum number of transmitted quadtree blocks of each vehicle by its Shannon data rate. Finding the optimal solution of (7) is complex and intractable, due to the need to frequently exchange fast-varying information between vehicles. Hence, we leverage machine learning techniques which have proved to be useful when dealing with such complex situations, specifically deep reinforcement learning (DRL) [9].

IV. RL IN COOPERATIVE SENSING

For a given vehicular association nn' and RB allocation, each vehicle n acts as an RL-agent whose goal is to learn which quadtree blocks to be transmitted to its associated vehicle n' in order to maximize its satisfaction. Accordingly, the *global state* of the RL environment is defined as $\langle \mathcal{B}_n(t), \mathcal{I}_{n'}(t), v_n, v_{n'}, l_n(t), l_{n'}(t) \rangle$, where $\mathcal{I}_{n'}(t)$ is the set of RoI weights of vehicle n' , as per (4), at time slot t . However, this global state cannot be observed by vehicle n , instead, the *local observation* of vehicle n is

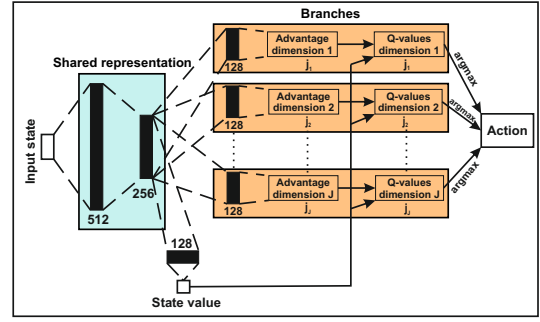


Figure 2. The BDQ neural network architecture utilized for our RL problem.

$\langle \mathcal{B}_n(t), v_n, v_{n'}, l_n(t), l_{n'}(t) \rangle$. At every time slot t and by utilizing this local observation, vehicle n would take an action $\sigma_n(t)$, selecting which quadtree blocks to be transmitted to its associated vehicle n' , and receives a feedback (*reward*) from vehicle n' equal to $f_{n'n}(t)$. In a nutshell, the elements of the RL problem at each vehicle n can be described as follows:

- Global state: $\langle \mathcal{B}_n(t), \mathcal{I}_{n'}(t), v_n, v_{n'}, l_n(t), l_{n'}(t) \rangle$
- Local observation: $\langle \mathcal{B}_n(t), v_n, v_{n'}, l_n(t), l_{n'}(t) \rangle$
- Action: $\sigma_n(t)$
- Reward: $f_{n'n}(t)$

To solve this RL problem, the well-known deep Q-Network (DQN) algorithm [9] can be utilized. Within DQN, the Q-value for each possible action should be estimated before deciding which action to take. As a result, its application to high dimensional action spaces is still arduous. It should be noted that our RL problem suffers from the high dimensionality of action spaces. Specifically, each vehicle needs to select $\sigma_n(t)$ whose dimension is $|\mathcal{B}_n|_{\max} \times 1$, yielding a total number of possible actions equal to $2^{|\mathcal{B}_n|_{\max}}$. These huge number of actions can seriously affect the learning behavior of DQN.

Recently, the authors in [13] have introduced a novel agent called Branching Dueling Q-Network (BDQ) leading to a novel neural network architecture that allows to distribute the representation of the action dimensions across individual network branches while maintaining a shared module that encodes a latent representation of the input state and helps to coordinate the branches. This neural network architecture exhibits a linear growth of the network outputs with increasing action space as opposed to the combinatorial growth experienced in traditional DQN network architectures. Fig. 2 demonstrates this neural network architecture, that we adopt in this work. As a result, for the message content selection $\sigma_n(t)$, the Q-value is estimated for $2 \times |\mathcal{B}_n|_{\max}$ actions only instead of $2^{|\mathcal{B}_n|_{\max}}$ actions.

For the vehicular agents training purposes, DQN [9] is selected as the algorithmic basis¹. The detailed training algorithm is shown in Algorithm 1. Note that the per agent loss function used for training is as follows [13]:

$$L(\phi) = \mathbb{E}_{(s,a,r,s') \sim U(D)} \left[\frac{1}{J} \sum_i (y_i - Q_i(s, a_i))^2 \right],$$

¹DQN is selected for its simplicity, powerfulness, and off-policy algorithm.

Algorithm 1 Training a BDQ agent for cooperative sensing

- 1: **Initialize** the replay memory of each agent to capacity C .
- 2: **Initialize** each agent's neural network with random weights ϕ .
- 3: **Initialize** each agent's target neural network with weights $\phi^- = \phi$.
- 4: **foreach** slot t **do**
- 5: Each vehicle n computes its local observation $\langle \mathcal{B}_n(t), v_n, v_{n'}, l_n(t), l_{n'}(t) \rangle$.
- 6: With probability ϵ , each vehicle's agent selects random sensory blocks to be transmitted to its associated vehicle, otherwise it selects the sensory blocks with maximum Q-value.
- 7: The selected sensory blocks are transmitted over the allocated RB to the associated vehicle which only receives a random subset of these blocks depending on the data rate $R_{nn'}(t)$ as per (5).
- 8: Each vehicle n calculates its own satisfaction $f_{nn'}(t)$ with the received blocks and feeds it back as a reward to its associated vehicle.
- 9: Each vehicle n receives the reward, observes the next state and stores this experience (s_t, a_t, r_t, s_{t+1}) in its replay memory.
- 10: **if** vehicle n collected a sufficient amount of experiences **do**
- 11: Vehicle n samples uniformly a random mini-batch of experiences from its replay memory.
- 12: A gradient decent step is performed on $L(\phi)$ w.r.t. ϕ .
- 13: **end if**
- 14: **end for**

where ϕ are the weights of the corresponding Q-network. For an action dimension $i \in \{1, \dots, J\}$ with $|\mathcal{A}_i| = j_i$ discrete sub-actions, the individual branch's Q-value at state $s \in \mathcal{S}$ and sub-action $a_i \in \mathcal{A}_i$ is expressed in terms of the common state value $V(s)$ and the corresponding state-dependent sub-action advantage $A_i(s, a'_i)$ by $Q_i(s, a_i) = V(s) + (A_i(s, a_i) - \frac{1}{j_i} \sum_{a'_i \in \mathcal{A}_i} A_i(s, a'_i))$. Furthermore, $y_i = r + \gamma \frac{1}{j} \sum_i Q_i^-(s', \arg \max_{a'_i \in \mathcal{A}_i} Q_i(s', a'_i))$ is the temporal difference targets, with a discount factor of γ , and Q_i^- denoting branch i of the target neural network².

A. Federated RL

Note that, each vehicle n only uses its own experience to train its BDQ agent. Therefore, in order to have a resilient agent that performs well in different situations, the training process should run for a sufficient amount of time for each vehicle to have a diversified experience. Alternatively, during the training process, the vehicles could share their trained models with each other periodically to enhance the training process and obtain a better trained agent in less amount of time. In order to do so, we investigate the role of federated RL [14]. FL is a machine learning framework where different clients (vehicles) collaboratively train a model under the orchestration of a central entity (RSU), while keeping the training data (experience) decentralized and private [10]. Instead of applying FL within a supervised learning task, we apply FL within our cooperative perception vehicular RL problem. Specifically, after every X slots, each vehicle under the coverage of the RSU shares its latest model parameters (neural network weights) with the RSU, the RSU then averages all the received model parameters and broadcasts the outcome back to the vehicles.

²For more details on the choice of the loss function and its components, please refer to [13].

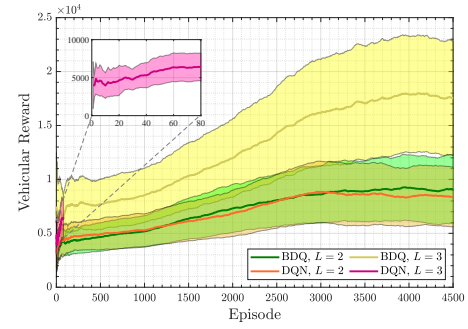


Figure 3. Learning curves for the vehicular RL environment using BDQ vs DQN agents, for different L . The solid lines represent the average over all the vehicles, where the learning curve of each vehicle is smoothed by the moving average over a window size of 1000 episodes, while the shaded areas show the 90% confidence interval over the vehicles.

Table I
SIMULATION PARAMETERS.

Parameter	Value	Parameter	Value
ω	180 KHz	N_0	-174 dBm/Hz
τ	2 ms	P	10 dBm
M	100 Byte	t_{int}	2 sec
λ_n	1	L	5
X	5 slots	r	20

V. SIMULATION RESULTS AND ANALYSIS

In this section, simulations are conducted based on practical traffic data to demonstrate the effectiveness of the proposed approach. A traffic light regulated junction scenario is considered. The scenario contains vehicles of different dimensions to mimic assorted cars, buses and trucks. The vehicles' mobility traces are generated using Simulation of Urban MObility (SUMO) application [15]. Unless stated otherwise, the simulation parameters are listed in Table I.

Moreover, the hyperparameters used for training the vehicular agents are discussed next. Common to all agents, training always started after the first 1000 steps, after which one step of training is run at every time step. Adam optimizer was used with a learning rate of 10^{-4} . Training was done with a mini-batch size of 64 and a discount factor $\gamma = 0.99$. The target network is updated every 1000 time steps. A rectified non-linearity (ReLU) is used for all hidden layers and a linear activation is used on the output layers, for all the neural networks. Each neural network had two hidden layers with 512 and 256 units in the shared network module and one hidden layer per branch with 128 units. Finally, a buffer size of 10^6 is set to the replay memory of each agent.

First of all, we verify that the BDQ agent can deal with the huge action space problem without any performance degradation compared to the classical DQN agent. For this purpose, we vary the size of the action space of the RL problem by varying the maximum quadtree resolution L . Note that, when $L = 2$, the maximum number of blocks available is $\frac{1-4^L}{1-4} = 5$, resulting in a total number of actions of $2^5 = 32$, and when $L = 3$, the maximum number of blocks available is 21, leading to a total number of $2^{21} \approx 2 \times 10^6$ actions. Fig. 3 shows the learning curve of both BDQ and DQN agents, for

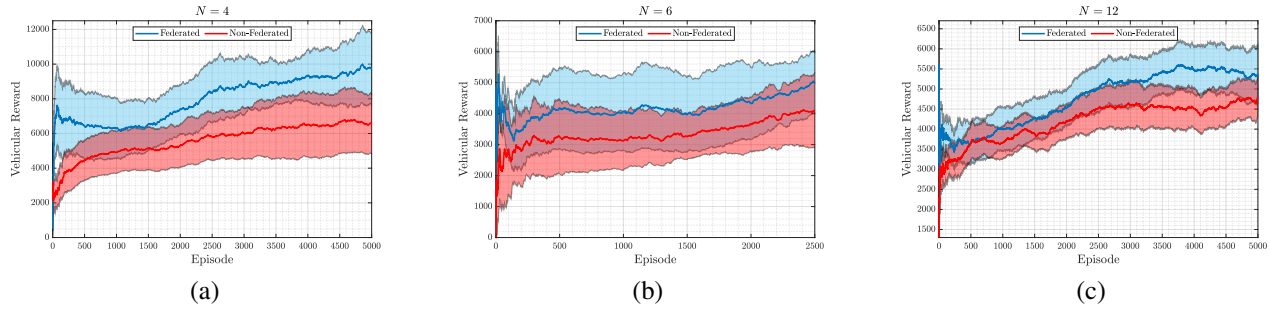


Figure 4. Learning curves for the federated vs non-federated scenarios of vehicular cooperative perception environment. The solid lines represent the average over all the vehicles, where the learning curve of each vehicle is smoothed by the moving average over a window size of 1000 episodes, while the shaded areas show the 90% confidence interval over the vehicles.

each case of L . When $L = 2$ (small action space), the learning curve of both BDQ and DQN agents are comparable and they learn with the same rate. However, when L increases to 3 (huge action space), the training process of the DQN agent could not be completed because it is computationally expensive. This is due to the huge number of actions that need to be explicitly represented by the DQN network and hence, the large number of network parameters that need to be trained every iteration. On the other hand, the BDQ agent shows more robustness against the huge action space, which demonstrates the suitability of BDQ agents to overcome the frequent scalability problems faced by other forms of RL.

Finally, in Fig. 4, we compare the training process evolution of the federated and non-federated scenarios for different values of N . It is observed that for the same training period, if compared to the non-federated scenario, the federated scenario achieves better rewards, and, hence, better policies over all the vehicles. This result corroborates that FL is instrumental in enhancing and boosting the RL training process.

VI. CONCLUSION

In this paper, we have studied the problem of selecting the contents of the vehicular CPMs in order to maximize the vehicles' satisfaction in terms of the received sensory information while considering the impact of the wireless communication. To solve this problem, we have resorted to the DRL techniques to model our cooperative perception RL problem. In order to overcome the huge action space inherent to the formulation of the RL problem, we applied the dueling and branching concepts. Moreover, we have studied the role of federated RL in enhancing the RL training process. Simulation results show that policies achieving higher vehicular satisfaction could be learned by the vehicles, leading to a higher vehicular satisfaction. It is also shown that federated RL achieves better policies within the same amount of time compared to the non-federated approach.

REFERENCES

- [1] J. Park, S. Samarakoon, H. Shiri, M. K. Abdel-Aziz, T. Nishio, A. Elgabli, and M. Bennis, "Extreme URLLC: Vision, Challenges, and Key Enablers," 2020.
- [2] ETSI TR 103 562 V2.1.1, "Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Analysis of the Collective Perception Service (CPS); Release 2," Dec. 2019.
- [3] Y. Wang, G. de Veciana, T. Shimizu, and H. Lu, "Performance and scaling of collaborative sensing and networking for automated driving applications," in *2018 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE, 2018, pp. 1–6.
- [4] 3GPP TR 22.886 V16.2.0, "3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Study on enhancement of 3GPP Support for 5G V2X Services (Release 16)," Dec. 2018.
- [5] G. Thandavarayan, M. Sepulcre, and J. Gozalvez, "Generation of cooperative perception messages for connected and automated vehicles," *arXiv preprint arXiv:1908.11151*, 2019.
- [6] M. Gabb, H. Digel, T. Müller, and R.-W. Henn, "Infrastructure-supported Perception and Track-level Fusion using Edge Computing," in *2019 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2019, pp. 1739–1745.
- [7] Q. Chen, S. Tang, Q. Yang, and S. Fu, "Cooper: Cooperative perception for connected autonomous vehicles based on 3d point clouds," in *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2019, pp. 514–524.
- [8] H. Samet, "The Quadtree and Related Hierarchical Data Structures," *ACM Comput. Surv.*, vol. 16, no. 2, pp. 187–260, Jun. 1984. [Online]. Available: <https://doi.org/10.1145/356924.356930>
- [9] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, Feb 2015.
- [10] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," ser. Proceedings of Machine Learning Research, A. Singh and J. Zhu, Eds., vol. 54. Fort Lauderdale, FL, USA: PMLR, 20–22 Apr 2017, pp. 1273–1282. [Online]. Available: <http://proceedings.mlr.press/v54/mcmahan17a.html>
- [11] S. Kaul, M. Gruteser, V. Rai, and J. Kenney, "Minimizing age of information in vehicular networks," in *2011 8th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks*, 2011, pp. 350–358.
- [12] T. Mangel, O. Klemp, and H. Hartenstein, "A validated 5.9 ghz non-line-of-sight path-loss and fading model for inter-vehicle communication," in *2011 11th International Conference on ITS Telecommunications*, 2011, pp. 75–80.
- [13] A. Tavakoli, F. Pardo, and P. Kormushev, "Action Branching Architectures for Deep Reinforcement Learning," *CoRR*, vol. abs/1711.08946, 2017. [Online]. Available: <http://arxiv.org/abs/1711.08946>
- [14] C. Nadiger, A. Kumar, and S. Abdelhak, "Federated Reinforcement Learning for Fast Personalization," in *2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*, 2019, pp. 123–127.
- [15] P. A. Lopez and M. Behrisch and L. Bieker-Walz and J. Erdmann and Y. Flötteröd and R. Hilbrich and L. Lücken and J. Rummel and P. Wagner and E. Wiessner, "Microscopic Traffic Simulation using SUMO," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 2018, pp. 2575–2582.