IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received August 11, 2021, accepted August 26, 2021, date of publication August 31, 2021, date of current version September 10, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3109422

Simultaneous Data Rate and Transmission Power Adaptation in V2V Communications: A Deep Reinforcement Learning Approach

JUAN AZNAR-POVEDA, ANTONIO-JAVIER GARCIA-SANCHEZ, ESTEBAN EGEA-LOPEZ, AND JOAN GARCÍA-HARO, (Member, IEEE)

Department of Information and Communications Technologies, Universidad Politécnica de Cartagena, 30202 Cartagena, Spain

Corresponding author: Juan Aznar-Poveda (juan.aznar@upct.es)

This work was supported in part by the AEI/FEDER/UE [Agencia Estatal de Investigación (AEI), Fondo Europeo de Desarrollo Regional (FEDER), and Unión Europea (UE)] under Grant PID2020-116329GB-C22 [ARISE2: Future IoT Networks and Nano-networks (FINe)] and Grant PID2020-112675RB-C41 (ONOFRE-3), in part by the Fundación Séneca, Región de Murcia, under Grant 20889/PI/18 (ATENTO), and in part by the LIFE project (Fondo SUPERA COVID-19 through the Agencia Estatal Consejo Superior de Investigaciones Científicas CSIC, Universidades Españolas, and Banco Santander). The work of Juan Aznar-Poveda was supported by the Spanish Ministerio de Educación, Cultura y Deporte (MECD) through the Formación de Personal Investigador (FPI) Predoctoral Scholarship under Grant BES-2017-081061.

ABSTRACT In Vehicle-to-Vehicle (V2V) communications, channel load is key to ensuring the appropriate operation of safety applications as well as driver-assistance systems. As the number of vehicles increases, so do their communication messages. Therefore, channel congestion may arise, negatively impacting channel performance. Through suitable adjustment of the data rate, this problem would be mitigated. However, this usually involves using different modulation schemes, which can jeopardize the robustness of the solution due to unfavorable channel conditions. To date, little effort has been made to adjust the data rate, alone or together with other parameters, and its effects on the aforementioned sensitive safety applications remain to be investigated. In this paper, we employ an analytical model which balances the data rate and transmission power in a non-cooperative scheme. In particular, we train a Deep Neural Network (DNN) to precisely optimize both parameters for each vehicle without using additional information from neighbors, and without requiring any additional infrastructure to be deployed on the road. The results obtained reveal that our approach, called NNDP, not only alleviates congestion, leaving a certain fraction of the channel available for emergency-related messages, but also provides enough transmission power to fulfill the application layer requirements at a given coverage distance. Finally, NNDP is thoroughly tested and evaluated in three realistic scenarios and under different channel conditions, demonstrating its robustness and excellent performance in comparison with other solutions found in the scientific literature.

INDEX TERMS Vehicular ad-hoc networks, connected vehicles, Vehicle-to-Vehicle (V2V) communications, congestion control, power control, data rate control, deep reinforcement learning.

I. INTRODUCTION

Future Intelligent Transportation Systems (ITS) aim to reduce both the number and severity of accidents using connected vehicles. In ITS, Vehicle-to-Vehicle (V2V) communications [1], [2] periodically exchange broadcast single-hop messages, called beacons, to announce information which enables the tracking and prediction of neighboring vehicle behavior [3]. The goal is to achieve context awareness by means of cooperation among vehicles [3]. As the number of vehicles on the road increases, this context awareness is crucial, laying the foundations for many safety applications that reduce the risks of collision [4]–[6]. An overloaded channel compromises this feature because high packet losses may occur; affecting both periodical beaconing and event-related messages triggered in emergency cases [7]. Context-awareness can help leave a certain fraction of the channel capacity available to be used to deliver relevant messages, guaranteeing, a priori, the safety of drivers, passengers, and pedestrians.

Channel congestion may be controlled by different transmission parameters. The most common approach adjusts

The associate editor coordinating the review of this manuscript and approving it for publication was Moayad Aloqaily.

the beaconing rate [8]–[10]. Another approach fine-tunes transmission power, thus regulating the number of messages received by vehicles [11]–[13]. Finally, the data rate is also a parameter used to relieve channel congestion. However, very few proposals for data rate adaptation have been discussed in the scientific literature [14].

Adjusting a single parameter often entails worse outcomes in reducing channel load in highly congested scenarios. For instance, if the transmission power is decreased too much, messages would reach only a few neighboring vehicles; those located very near, leading to poorer cooperation among vehicles. In contrast, fine-tuning a combination of two or more parameters may yield better performance results since no drastic changes that would be detrimental to cooperative awareness occur. With this line of action, the most common approaches jointly consider transmission power and beaconing rate [15]–[17]. An optimal joint allocation of both parameters would be the best solution; however, the associated optimization problems are not always convex [18]. This issue may result in mixed-integer problems (MIPS), escalating computing complexity. To deal with this problem, some proposals in the literature suggested applying artificial intelligence (AI) techniques to reach an optimal joint allocation of beaconing rate and transmission power [19]. Although this set of parameters works significantly well, it is more consistent to combine data rate and transmission power since they are intrinsically interrelated. That is, high data rates reduce channel load but use more complex modulations, which are less robust against unfavorable channel conditions, and their efficiency depends on the transmission power radiated. Therefore, high transmission powers should be employed to balance the weakness of fading and attenuation at longer distances. Joint data rate and transmission power solutions are rarely found and simply treated in the scientific literature, which means they have limited ranges of data rates and transmission powers [20], [21].

In this paper, we apply the Deep Reinforcement Learning (DRL) framework to alleviate channel congestion through optimizing data rate and transmission power simultaneously. Given the nature of this problem, in which no a priori information or data about the (road) environment is available, we formulate it as a Markov Decision Process (MDP) and solve it using Deep Reinforcmeent Learning (DRL) algorithms. Previous proposals applying DRL to this problem are focused on infrastructure networks and disregard data rate control [22]–[24]. Our solution is addressed to an ITS G5 infrastructure-less (ad hoc) network; that is, a distributed environment where cooperation among vehicles naturally leads to Multi-Agent Reinforcement Learning algorithms (MARL) [25], [26]. However, the solution given in these papers are difficult to train and are not yet mature enough, especially regarding future real implementations. In our case, we train a single agent, whose resulting policy could be easily stored on-board any vehicle belonging to the network. Moreover, this shared policy obtains suitable results without the need to tackle the complexities associated with MARL approaches. In fact, in our previous paper [19], we demonstrated that this single agent, when appropriately trained, controls beaconing rate and transmission power by using a tabulated Q-learning method. Unlike [19], in this work we employ DRL, which is appropriate when the state space is large and continuous, as occurs in our case (road environment). The outcoming actions (data rate and transmission power) of the trained Deep Neural Network (DNN) are then applied by vehicles in a non-cooperative fashion, without the need to request additional information from neighboring vehicles.

This proposed mechanism, denoted Neural Network for Data rate and transmission Power (NNDP), controls overall channel congestion while assuring a certain transmission range with the most robust data rate possible. In short, we verify that training a single agent using our DRL approach is an appropriate solution to jointly adapt data rate and transmission power and thus adjust congestion levels in an effective way. The main contributions of this research work are summarized as follows:

- The policy is implemented through a DNN solution, which accepts continuous values as input. All the values stated in the standard are taken into consideration. This endows the algorithm with greater flexibility and accuracy than previous approaches [20], [21].
- The proposed method keeps the channel load below a certain threshold to avoid congestion, which notably reduces packet loss. At the same time, channel underutilization is avoided.
- Transmission power is adjusted to the necessary level to guarantee a given packet delivery ratio at a certain distance.
- Low data rates with more robust modulations schemes are rewarded, whenever possible, if the channel load allows their usage.
- The model obtained can be applied in a fully distributed fashion, without the need for a centralized network infrastructure.
- Finally, no information from neighboring vehicles is required to carry out actions, so any exchange with the application layer is disregarded for an appropriate resource allocation operation.

The remainder of the paper is organized as follows. First, Section II discusses the related work and congestion control from the viewpoint of the existing trade-off between transmission power and data rate. Then, in Section III, we detail the model proposed and the solving method used. Section IV conducts the performance evaluation, discussing different simulation scenarios and metrics, and compares the achieved results with other proposals of interest. Finally, Section V summarizes the main conclusions.

II. RELATED WORK

Vehicle communications are specified by the European Telecommunications Standards Institute (ETSI). In particular, ETSI defines the ITS-G5 radio channel comprising a 10 MHz control channel at the 5.9 GHz band of the IEEE 802.11p standard [27]. Transmissions over these networks are broadcast and employ Carrier-Sense Multiple Access with Collision Avoidance (CSMA/CA) as a medium access control (MAC) protocol. The ETSI Cooperative Awareness Service (CAS) also features periodic beaconing over one-hop communications as the basis of cooperative awareness. Such periodic messages or beacons, formally called Cooperative Awareness Messages (CAM) in Europe or Basic Safety Messages (BSM) in the US, are responsible for disseminating status and environmental information to vehicles on the control channel (G5CC in Europe and Channel 172 in the US, respectively). However, the aggregated load generated by dispatching beacons may cause packet loss, thereby negatively impacting safety applications. In addition, the Decentralized Environmental Notification (DEN) service, which is in charge of notifying about risk-related road events [7], needs some channel availability to guarantee the appropriate delivery of event-related messages in emergency cases, called Decentralized Environmental Notification Messages (DENM). To this end, the Cross-Layer Decentralized Congestion Control (DCC) Management Entity [28] is aimed at preventing ITS-G5 radio channel overload by adjusting different transmission parameters.

Regarding single-parameter control, beaconing or message rate is the most frequently employed congestion control parameter, and different authors have implemented algorithms which relate beaconing rate to the measured Channel Busy Rate (CBR) [8], [9], vehicle dynamics [3], [29], [30], or context information [10], [31]–[33], among others. However, in some cases, the only way to alleviate congestion is to decrease the beaconing rate excessively, which may degrade the necessary context awareness capabilities and hence vehicle safety [34].

Another widely used parameter in congestion control is transmission power. Reducing transmission power means decreasing coverage distance and thus, the number of vehicles that receive the broadcast messages, so overall congestion is also alleviated. Several works propose controlling transmission power as a function of different metrics. Authors in [35] employed channel state information (CSI) to maximize energy efficiency in wireless cellular networks. The work in [11] exploited vehicle speed as a parameter to allocate transmission power. In particular, this approach extended the transmission range in the case of high speeds to raise awareness in neighboring vehicles with less time-to-collision. Vehicle density is also employed in [12] to decide whether to increase or decrease transmission power. Likewise, authors in [39] included an SNIR estimation in their study. Conversely, some proposals directly allocate transmission power as a function of the channel load [13], [40]. In [41], a parameter denoted as vehicle position prediction error determined the increase/decrease in transmission power. In general, if transmission power takes inadequately low values while attempting to mitigate congestion, the number of receivers would drop too much and overall awareness would be jeopardized. On top of this, abrupt transmission power variations also cause variations in the resource allocation mechanisms, as is dealt with in [13].

More advanced proposals combine two or more parameters simultaneously to take advantage of the benefits of each them as much as possible. In this case, an algorithm for joint optimal allocation of several parameters could be a good solution, but we find an important drawback in many cases: the optimization problem is usually non-convex. Even though there are solutions involving two or more parameters that clearly improve the usefulness and flexibility of congestion control [15], [18], [37], [38], [42]–[44], there is no silver bullet to resolve congestion control from a multi-parameter perspective. Given the complexity of the optimization problem, different advanced solutions have emerged. Similarly to [14] but including beaconing rate and transmission power as control parameters in the mathematical problem, an algorithm based also on game theory was proposed in [16]. Decision-making theory has also been an important tool to achieve optimal congestion control and endow a certain amount of intelligence to vehicles. In particular, the Markov Decision Process (MDP) framework is one of the decision-making techniques that provide the basis of reinforcement learning (RL) [45] commonly employed to solve complex problems. Congestion control is proposed by varying transmission power using both Q-Learning, in the particular case of LTE-V2V communications [22], and a MARL approach for overall wireless communications [25], [26], [36]. Regarding solutions where more than one parameter is optimized, authors in [23] derived the best selection of the frequency sub-band together with transmission power through a deep decision-making approach. In the case of C-V2V networks, a reinforcement learning framework offers a smart solution for balanced power control and rate adaptation [24]. Finally, in the context of the IEEE 802.11p standard, and consequently, in a distributed fashion, discrete Q-learning has also been proposed in [19] to optimize both beaconing rate and transmission power allocation.

Despite the fact that the IEEE 802.11p standard [27] defines up to 9 different data rates from 3 to 27 Mbps, a data rate of 6 Mbps is usually recommended. Moreover, authors in [46] provided a method to identify the optimum data rate according to different scenarios, assuming and fixing the 6 Mbps data rate. Higher data rates entail shorter packet durations, reducing the channel load, but these data rates employ high-order modulation schemes and coding rates. This means less robustness against adverse channel conditions over distance. To mitigate this effect, higher transmission power is required to guarantee an adequate Packet Delivery Ratio (PDR) at a given target distance. On the other hand, low data rates reduce the required transmission power levels to provide reliable communications at a certain target distance but increasing the transmission time and therefore, decreasing the throughput. The trade-off between data rate and transmission power in terms of transmission range and

TABLE 1. Comparison of our congestion control proposal (NNDP) and other related works.

	[8], [9]	[10], [31]	[12], [35]	[25], [45]	[14]	[16]	[40], [43]	[19], [23]	[20], [21]	**NNDP**
Controlled parameters	BR	BR	TP	TP	DR	BR, TP	BR, TP	BR, TP	DR, TP	**DR, TP**
Centralized/Distributed	D	D	D	C	D	D	D	D	D	**D**
Type of algorithm	O	O	O	RL	G	G	O	RL	O	**RL**
Non-cooperative	Yes	No	Yes	No	Yes	Yes	Yes	Yes	Yes	**Yes**

(BR): Beaconing Rate; (TP): Transmission Power; (DR): Data Rate; (O): Optimization; (G): Game theory; (RL): Reinforcement Learning.

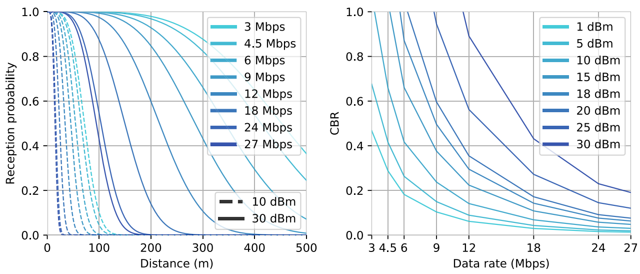


FIGURE 1. Reception probability variation over distance (left) and channel busy ratio (right) for different data rates and transmission powers using a Nakagami-m fading model.

reception probability is depicted in Figure 1. In [46], transmission power was adjusted to obtain the same PDR as the reference value obtained for 6 Mbps. However, this was discussed in [47], whose authors claimed that it is not clear whether the selected transmission power levels in [46] guarantee the communication range required by vehicular applications. Using both simulations and field experiments, authors in [47] demonstrated that 6 Mbps is not always the optimum data rate. As a consequence, there is limited research work which tackles data rate variations. We can highlight the study in [14], which is a non-cooperative approach based on game theory to successfully maintain congestion below a certain level.

As can be observed in Table 1, most of the aforementioned proposals integrate the beaconing rate and transmission power parameters to control channel congestion in their formulations. In light of the existing trade-off between data rate and transmission power, as explained above and in Figure 1, data rate variations could be compensated by simultaneously fine-tuning transmission power. Therefore, this double-parameter perspective is much more physically consistent due to the channel condition dependence of the data rate: high data rates are more affected by fading and attenuation, and thus, the effective transmission range is reduced, although it can be adjusted by increasing transmission power. Moreover, data rate and transmission power parameters can be directly controlled by the DCC Management Entity as defined by the standard [3]. In contrast, other parameters (e.g. receiver sensitivity) are more dependent on the particular hardware of each vehicle and may affect the MAC operation. We consider this issue out of the scope of this work. To the best of the authors' knowledge, there are only two works aimed at combining data rate and transmission power in vehicular ad-hoc networks [20], [21]. The first work proposed a look-up table to optimally select pairs of transmission power and data rates in terms of the PDR and end-to-end delay. However, available pairs of data rate and transmission power parameters are very limited, which leads to undesired behavior whenever the environment is slightly modified. Moreover, the validation results of this work are scarce and therefore weak. The second work, called CACC [20], analyzed the Received Signal Strength (RSS) of the received packets to determine whether their losses were due to weak signal or collisions and, based on this, decided to decrease or increase the transmission power or data rate. Despite obtaining fairly good results in terms of the PDR, the channel is underused or overused depending on the scenario, and only a few data rates among all the available range are analyzed for simpler scenarios. Keeping these weaknesses in mind, a more sophisticated scheme would be necessary to consider the full range of each parameter and select them according to different goals. Consequently, to contribute to filling this research gap, we propose a deep reinforcement learning approach, called NNDP, to (i) prevent congestion, leaving some of the channel capacity to deliver event-related messages available. Also, (ii) transmission power is intended to preserve adequate performance of safety applications at a certain distance, while (iii) the most robust data rate is set whenever possible.

III. CONGESTION CONTROL USING DRL

Congestion control is developed to guarantee an appropriate channel load, usually measured by the CBR metric, around a certain target value denoted Maximum Beaconing Load (MBL). According to several works [16], [48], [49], its optimal value is around 0.6 and 0.7. Higher channel loads may increase packet loss and hinder proper safety application operations. In this paper, we propose to control congestion by jointly adjusting both the data rate and transmission power. However, this is not trivial, and a subtle trade-off between both parameters is required to satisfy application layer requirements. In the case of transmission power, values that are too high increase congestion while values that are too low endanger vehicles' awareness.In terms of data rates, high rates alleviate congestion due to shorter packet transmission intervals. Nonetheless, high order modulations are required and robustness against fading and attenuation is lessened as distance increases. To this aim and as already mentioned in Section I, we first model the problem through a Markov Decision Process (MDP) framework. In general, MDP addresses congestion control in ad-hoc vehicular communications as an optimization procedure over discrete actions taken by the vehicles themselves in a distributed fashion. However, when the state space is large or continuous, novel approaches employ approximation methods, as in our proposal. Unlike in our previous work, where we used beaconing rate and transmission power [19] and the MDP was solved using tabulated policies, in this work we apply Deep Reinforcement Learning (DRL) to find the optimal transmission parameters more accurately. Within the DRL framework, we train DNN models using a simplified environment programmed in Python. Once the training is completed, we check whether the trained DNN (agent or model) successfully alleviates the channel congestion through individual actions of the vehicles in realistic scenarios using a discrete event simulator for networks.

A. REINFORCEMENT LEARNING FRAMEWORK

MDPs provide a mathematical framework to derive optimal sequences of actions, so they are commonly applied to formulate optimization problems. This is especially useful in those challenging environments where outcomes may be partially random or difficult to predict, as happens in vehicular settings. Formally, MDPs consist of the following elements:

- The *agent* is the learner entity that continuously seeks optimal behavior. In our case, the *agent* is every single vehicle on the road, whose goal is to reduce overall channel congestion in a distributed manner, jointly employing transmission power and data rate parameters.
- The environmental situation, along with the properties of the agent is called *state*. Usually, the state is defined as a vector $s \in \mathcal{S}$ that embraces both the outer and inner properties of the agent, with $\mathcal{S}$ being the set of possible states.
- The agent is able to perform an *action* $a \in \mathcal{A}(s)$. This action belongs to the available set of actions for each state. In our case, actions are a tuple consisting of the transmission power and data rate to be set in forthcoming transmissions.
- Every time the agent takes an action, the environment is modified, presenting a new situation to be explored. In this change of state from s to s', the agent obtains a *reward* r, considered as the feedback from the environment. It can be modeled as a function of the state s and the action taken a, i.e., $r(s, a) = f(s, a) \in \mathbb{R}$.

MDP-solving algorithms employ what is called *policy*, denoted as π, a mapping between states and actions; that is, $\pi : \mathcal{S} \to \mathcal{A}$. The main objective is to reach the optimal policy π^*, which maximizes the accumulated sum of rewards over the entire lifespan of the agent during training. This decision policy can be determined by the state-action function, also called the Q-function, $Q(s, a)$, which can be approximated using Deep Neural Networks (DNN). In general, Markovian systems operate from discrete spaces so the agent and environment interact with each other in a sequence of discrete-time steps. However, as occurs in our particular case, more complex problems comprising continuous variables could require some approximations to be solved. This will be detailed in the following subsection while particularizing the constituent elements of the proposed MDP model.

B. DATA RATE AND TRANSMISSION POWER

Roads are fairly complex environments which are influenced by many factors, not only the physical parameters of the road and vehicles (e.g., speed, position, acceleration, etc.) themselves but also numerous human factors. In this way, traffic conditions are quite unpredictable due to unforeseen events. The associated number of neighboring vehicles and their beaconing loads may increase both channel congestion and packet collisions, therefore drastically reducing packet reception probability. Furthermore, there is an additional randomness due to the channel fading and attenuation produced by the surroundings of the road. For instance, rural areas generally represent more favorable channel conditions, while urban areas cause higher fading (e.g., multipath effects caused by objects and buildings) and increase the number of weak signals in the environment. We assume a well-accepted Nakagami-m [50] fading and path loss propagation model in order to realistically characterize a wide range of channel conditions. From this model, we compute the average carrier sense range, r_{CS} (m) as a function of the transmission power. Basically, carrier sense range is defined as the average distance from the transmitter where the power is sensed by the receiver over its sensitivity (S), as suggested in [13]:

$$r_{CS} = \frac{\Gamma(m + \frac{1}{\beta})}{\Gamma(m)(SA\frac{m}{p})^{\frac{1}{\beta}}} \tag{1}$$

where $\Gamma(x)$ is the *gamma function*, p the transmission power, β the *path loss exponent*, A is defined by the expression $(\frac{4\pi}{\lambda})^2$ (λ is the wavelength of the carrier), and S is receiver sensitivity. Finally, m is the so called *shape* parameter, which indicates the severity of the fading conditions. For instance, $m = 1$ means severe fading, while $m = 5$ denotes the most favorable fading. As previously shown in Figure 1 in which the reception probability was computed from a Nakagami-m model, the carrier sense range depends on transmission power. This can be observed in the carrier sense range expression (1) as well. Therefore, as transmission power increases, a larger number of packets received from neighboring vehicles that are located at greater distances could be successfully decoded. That is, there is information available from a greater number of vehicles thus enriching context awareness. However, this increase in power also implies increasing the channel load. In contrast, if transmission power is excessively reduced, vehicles would receive packets only from closer neighbors. Therefore, there is a trade-off for achieving a certain channel load level without jeopardizing context aware vehicle information. To set appropriate transmission power while controlling congestion, a second parameter is usually considered. The most common approach consists of varying the beaconing rate by fixing the data rate by default to 6 Mbps. However, there is no reason not to propose

TABLE 2. IEEE 802.11p Data rates.

Data rate (Mbps)	Modulation	CR	Data bits per symbol	Coded bits per symbol	S_r [47] (dBm)
3	BPSK	½	24	48	-85
4.5	BPSK	¾	36	48	-84
6	QPSK	½	48	96	-82
9	QPSK	¾	72	96	-80
12	16-QAM	½	96	192	-77
18	16-QAM	¾	144	192	-73
24	64-QAM	⅔	192	288	-69
27	64-QAM	¾	216	288	-68

controlling congestion by dynamically adjusting data rates while fixing the beaconing rate to the maximum allowed (i.e., 10 Hz). Indeed, the IEEE 802.11p standard [27] defines up to 9 different data rates, from 3 to 27 Mbps. Note that, as shown in Table 2, higher data rates imply higher-order modulation schemes.

On the one hand, high data rates are more beneficial in terms of network throughput since packet transmissions are shorter, but they are also more prone to packet error due to interference and noise. Therefore, the higher the data rate, the higher the Signal-to-Interference-plus-Noise Ratio (SINR) threshold required for successful packet reception and the shorter the effective transmission range. Table 2 can be used to illustrate the trade-off (related to different data rates) between generated channel load and transmission power requirements over distance. Note that the receiver sensitivities stated in the standard [47], [51], denoted by S_r, are the minimum required values to keep the Packet Error Ratio (PER) below 10%; which means that, in the absence of interference/noise, at least 90% of the packets with that power level will be successfully received. Under these circumstances, the selection of the appropriate data rate should be based on its capacity to reduce the channel load while simultaneously guaranteeing the application requirements using the most suitable transmission power [47]. In short, we mainly combine both transmission power and data rate to make sure that channel loads are kept below the required MBL. Once congestion is alleviated, we determine the transmission power to ensure that, at a certain target distance, the received power is above the S_r required by a given data rate. In the following section, the safety distance, transmission power, and data rate for the measured CBR are related to each other.

1) AGENTS, ACTIONS, AND STATES

Agents, which are represented by every single vehicle on the road, continuously sense their environment to adequately adjust both transmission power and data rate. As previously stated, they are mainly intended to reduce overall channel congestion in a distributed manner by making use of their own metrics and without relying on any centralized infrastructure. To this end, each vehicle first computes the channel capacity (C, messages per second) that would be available according to the selected data rate, as illustrated in Equation (2).

$$C = \left(C_d \left\lceil \frac{b_{st} + M}{C_d} \right\rceil + t_{ps} \right)^{-1} \tag{2}$$

The data field of the Medium Access Control (MAC) frame/packet layer [27], also called MAC Protocol Data Unit (MPDU), is comprised of the packet length M in bits (536 B), plus 22 bits of service and tail (b_{st}), and additional padding destined to reach multiple coded bits (C_d) per Orthogonal Frequency Division Multiplexing (OFDM) symbol. This padding is represented in Equation (2) by the ceiling function and, according to Table 2, each data rate entails a different number of coded bits per OFDM symbol. Before transmission, the Physical layer (PHY) also includes a preamble and a signal field (t_{ps}, in seconds), which are transmitted applying the most robust data rate (3 Mbps), which translates into 40 μs. The whole packet structure is summarized in Table 3. Once the data rate has been selected and the channel capacity is calculated, each vehicle estimates the CBR that would be measured if all vehicles employ the same operating parameters. To this end, we use the average carrier sense area ($2 \times r_{CS}$), the vehicle density detected in the neighborhood (ρ), and the average beaconing rate (b), which is set to 10 Hz for every vehicle.

$$CBR = \frac{2 r_{CS} \rho b}{C} \tag{3}$$

The set of Equations (1), (2), and (3) allows vehicles to carry out congestion control. We also consider *actions* consisting of 2-tuples of transmission power (p) and data rate (d), $a = \langle a_p, a_d \rangle$. As stated in the standard [27], [28], transmission power may take both discrete and continuous values ranging from 1 to 30 dBm, whereas the data rate is constrained to some discrete values, as shown in Table 2. Notice that Equation (3) is only an estimation to express channel load as a function of the transmission parameters of every single vehicle. A more realistic calculation would include information from neighboring vehicles, which would turn the problem into a multi-agent approach. This type of approach is very complex to address, train, and deploy. Instead, we train a single agent to recognize and act against different levels of congestion. Agents define *states* to model their situation and their environment, so both data rate and transmission power should be relevant parts of these states. In addition, channel congestion has been included in the state by using the estimated vehicle density (ρ) within the neighborhood of each vehicle. The *states* are then defined as a 3-tuple containing the currently used transmission power, data rate, and estimated vehicle density, $s = \langle p, d, \rho \rangle$. When a vehicle executes an action $a = \langle a_p, a_d \rangle$, the environment response leads the vehicle to a new state s', as follows. The transmission power and data rate are applied as the action values to the state. For instance, if the current state transmits at 15 dBm and 6 Mbps and $a = \langle -4.8, 12 \rangle$, the new state will reduce the transmission power to 10.2 dBm and increase the data rate to 18 Mbps. Since each vehicle applies the same

TABLE 3. Packet structure for MAC and PHY layers.

MAC			Service	PSDU	Tail	Pad bits
PHY	Preamble	Signal	Data			

trained policy, the channel load measured by the vehicles will be also changed to the corresponding value, given by Equation (3). Therefore, the transition to a new state $s' = \langle p + a_p, d + a_d, \rho\rangle = \langle p', d', \rho\rangle$ is calculated depending on action $a = \langle a_p, a_d\rangle$. These state transitions describe the behavior of the vehicles, which is governed according to the rules imposed by the reward function.

2) REWARD FUNCTION

Every time the agent (or the vehicle) performs an action and changes from state s to state s', a reward $r(s, a) \in \mathbb{R}$ is received. Maximizing accumulated rewards over time allows agents to learn the most suitable actions and, as a consequence, obtain an optimal policy. As mentioned above, the desired behavior is to maintain the channel load around a certain MBL, whose ratio over the channel capacity is typically between 0.6 and 0.7. Higher channel loads may increase packet loss and jeopardize the delivery of event-driven messages if an emergency arises. Conversely, lower channel loads decrease awareness of the surroundings and may cause channel to be underutilized. In order to achieve the desired behavior, we include the following function in our characterization:

$$g(x) = -sgn(x - x_T)x \tag{4}$$

where sgn is the signum function shifted by some target value x_T (in our case $x = CBR$ and $x_T = MBL$). As can be observed, a positive reward increase is obtained as long as the CBR approaches the target value (MBL). However, if the CBR exceeds that target value, an increasing negative reward is achieved. These penalties (negative rewards) intensify learning speed [45]. In this way, reaching the MBL $= 0.6$ not only allows us to reduce congestion and leave a certain fraction of the channel free to guarantee the delivery of emergency-related messages but also prevents channel underutilization. To move the agent toward this optimal behavior, we add $+10$ to the reward whenever the CBR reaches the MBL within a ± 0.025 error interval and -0.1 otherwise.

In addition to CBR control, some restrictions should be included to prevent the model from reaching undesired combinations of transmission parameters. For instance, the agent could learn to set the most robust data rate (longer transmission times) at the expense of reducing transmission power and thereby reaching a fewer number of neighbors. Despite achieving adequate channel load levels, overall awareness on the road would be seriously impacted; that is, transmitted messages would only reach the closest neighboring vehicles. To overcome this problem, we include a second term in the reward function aimed at satisfying reliability and awareness at a given distance. As already discussed in [47], higher data rates reduce congestion in an effective manner but entail less robustness against fading. This reduces the effective transmission range, requiring an increase in transmission power to obtain the same PDR at a certain distance. The sensitivities (S_r) specified in Table 2, also called *reliability sensitivities*, depend on the selected data rate and are used to improve the performance of the application layer, guaranteeing that at least 90% of the packets received are successfully decoded. Using a one-slope path loss model and the aforementioned sensitivities, $l = Ad_s^{\beta}$, we can shape the reward function to provide an acceptable PDR for safety applications, at least, up to a certain distance, called the safety distance (d_s). Therefore, the higher the received power over sensitivity, the higher the reward obtained, as indicated by the following equation:

$$r = -|(S_r + l) - p| \tag{5}$$

Note that this expression is aligned with the fact that from a logarithmic scale perspective, transmission power (p) minus path loss (l) results in power received ($p - l$) at a certain safety distance, which, in turn, should be greater or equal to sensitivity. It is true that lower data rates entail lower sensitivities, and the effective transmission range can be much higher than that for higher data rates (more vulnerable to channel conditions). This aspect is already included in expression (5). However, we also encourage low data rate usage whenever possible by adding a third term, so, the higher the data rate, the more negative the reward. In this way, excessive variations among higher data rates are most likely avoided. The total reward function is therefore aimed at controlling channel loads (see Equation (4)) while guaranteeing the proper operation of safety applications (Equation (5)) by intelligently exploiting the trade-off between transmission power and data rate, as shown in Equation (6):

$$r = \omega_c g(CBR) - \omega_p|(S_r + l) - p| - \omega_d(d)^{\omega_e} \tag{6}$$

Each term of the reward function is normalized and weighted using an iterative process to the following values: $\omega_c = 2$, $\omega_p = 0.25$, $\omega_d = 0.1$, and $\omega_e = 0.8$. As can be observed, channel load control assumes greater importance, while those terms that control single parameters play a minor role. For instance, excessive values of ω_c with respect to ω_p and ω_d result in satisfying the CBR limit, but some transmission power and data rate combinations may be undesired (e.g., too low transmission powers). In contrast, lower values of ω_c could violate the desired MBL objective, which means that congestion is no longer being controlled. Concerning the exponent of the data rate term, named ω_e, it governs how negative the rewards are as long as data rates increase. A 0.8 value is set to obtain a similar range for the rest of the terms. In essence, a balance among weights is required to satisfy the different constraints appropriately within the bounds of the parameters stated in the standard.

3) DERIVATION OF π^*

Once the proposed MDP model has been formulated, the next step is to derive the optimal policy (π^*), which determines the

TABLE 4. Environment and learning parameters and their values.

Parameter	Value
Channel frequency	5.9 GHz
Channel model fading	Nakagami-m
Path loss exponent (β)	2.5
Shape parameter (m)	2
Sensing power threshold (S)	-92 dBm
Safety distance (d_s)	100 m
Message size (M)	536 B
Learning rate (α)	0.0001
Discount factor (γ)	0.9
Batch size	128
Beaconing rate (b)	10 Hz
Min, Max vehicle density (ρ_{min}, ρ_{max})	0.0001, 0.8
Min, Max transmission power (p_{min}, p_{max})	1 dBm, 30 dBm
Min, Max data rate (d_{min}, d_{max})	3 dBm, 27 Mbps

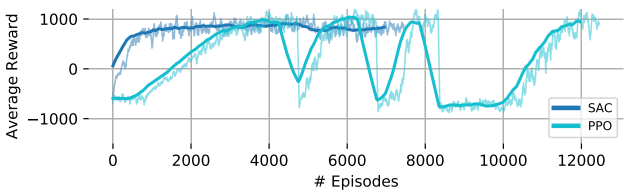


FIGURE 2. The average accumulated reward for PPO and SAC algorithms.

best action for every single state. Traditional MDP-solving algorithms, such as Q-learning [19], [52], use tabular methods which map $\pi : \mathcal{S} \rightarrow \mathcal{A}$ merely by employing a table. Despite achieving a convergent solution, and, a priori, good results, tabular methods are not appropriate to map every single state onto a suitable action, in particular when the state space is too large or continuous, as occurs in our case. Instead, we make use of Deep Neural Networks (DNN) to model π. Therefore, the policy is represented not as a table but as a parameterized functional form with a vector of weights, that is $\pi := f(\theta)$. By adjusting these weights θ, a wide range of functions can be implemented by the DNN. In our case, the DNN learns the best transmission parameters based on the road environment and vehicle situation.

There are many DRL algorithms based on DNNs [53]–[58] but not all of them accept the same type of states and actions. Recall that in our case, transmission power and vehicle density are continuous parameters. Concerning data rate, we consider it to be continuous, to later take the closest discrete value that satisfies the requirements stated in the standard. In this way, we resort to algorithms that feed on continuous actions, such as [53], [55]–[58], highlighting [57] and [58] for their good performance. The first one to consider is the so-called Proximal Policy Optimization (PPO) algorithm [57], which inputs multiple epochs of stochastic gradient ascent to perform each policy update. PPO exhibits the stability and reliability of trust-region methods (TRPO) but it is much simpler to implement. The second algorithm that also presents good results is the Soft Actor-Critic [58], whose main feature is entropy regularization. With SAC, the policy maximizes a trade-off between the expected return and entropy, a measure of randomness in the policy, which ensures greater robustness and stability. In our work, both algorithms were implemented in Python through RL-dedicated libraries [59], [60]. Basically, they iteratively calculate the maximum expected future rewards for each action at each state according to different policies. In particular, we selected a policy that implements the actor-critic method based on a multilayer perceptron (2 layers of 64 nodes). A hyperbolic tangent activation function is employed for PPO by default, whereas a Rectified Linear Unit (ReLU) is used for SAC. The initial weights of the DNN models for both PPO and SAC agents were randomly initialized. In practical terms, as shown in Figure 2, PPO results in much faster training than SAC, but eventually, the rewards decrease, which means that the algorithm forgets the good behavior learned. To avoid this situation, we automatically save the best model every few episodes. Conversely, SAC offers more stable rewards.

It is also important to highlight that the training process is performed by a single vehicle that monitors different levels of congestion, represented by the density of the vehicles sensed (ρ). Then, the trained model is loaded onto every single vehicle in the network to be evaluated (this process will be further explained in the following section). The rationale behind this is that channel loads are similar among neighboring vehicles so all of them will have the same requirements and thus, similar transmission parameters. This is just an assumption that enables channel loads to be estimated by taking the information from the vehicles (Equation (3)) into account. This estimation will be fairly close to the real load. Overall congestion is properly controlled in a distributed fashion as will also be shown in the next section. Note that as each vehicle applies the same policy with a similar channel load among neighbors, our proposal successfully converges to the same congestion level per vehicle. Finally, the environment and learning parameters used for the training of the PPO and SAC agents have been summarized in Table 4. In the next section, the trained DNN models are fed into realistic computer simulation software [61] to evaluate the performance of the algorithms in terms of channel congestion. The channel load estimate stated in this section and given by the expression (3) will also be thoroughly tested for different scenarios to prove the validity and robustness of the proposed algorithm.

IV. PERFORMANCE EVALUATION

In this section, we evaluate the performance of our proposed congestion control mechanism (NNDP), using OMNeT++ 5.3 [61] and including the INET 3.5 library [62]. This library implements the IEEE 802.11p standard along with realistic radio propagation and interference models. This simulation software as well as the RL libraries must be the cornerstones of the learning process. Once the learning process is finished and the weights of the DNN model have been thoroughly optimized, the vehicles will apply the resulting actions to alleviate any channel congestion episodes. In a real implementation, this would be achieved by installing the trained DNN model in the vehicle controller. The entire DNN model could be

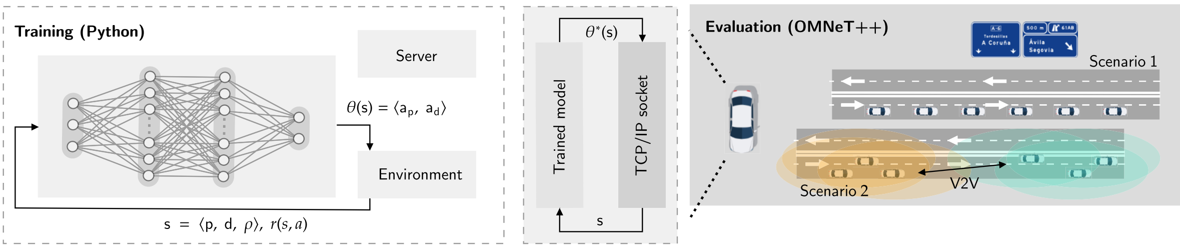


FIGURE 3. **Diagram of the training process of the DNN in a simplified environment developed in Python (left), and the subsequent evaluation of the trained model to control congestion in realistic vehicular networks (right). Different road scenarios have been simulated in OMNeT++, in which each vehicle individually sends its state over the socket and receives the optimal actions of transmission power and data rate from the DNN model (previously trained in Python).**

directly exported for this purpose but numerous compatibility problems may arise between OMNeT++ and RL libraries written in Python. To resolve this issue, one option is to use tables to store both actions and states, which would evaluate the DNNs, but at the expense of losing accuracy and reducing the potential advantages of DNNs against tabular methods like Q-learning. As a simple solution to evaluate the trained DNN model, we create a TCP/IP socket connection between Python, in which the model is contained, and OMNeT++, in which realistic road scenarios are simulated. The training process of the DNN model in Python using different DRL algorithms (PPO and SAC) and the subsequent evaluation of the resulting trained DNN models are illustrated in Figure 3.

After opening the aforementioned socket connection between Python and OMNeT++, which saves us from exporting the whole trained model to the OMNeT++ simulator, each vehicle initialize its transmission parameters (23 dBm and 6 Mbps), and our proposed congestion alleviation mechanism (NNDP) starts to run. As can be observed in Algorithm 1, it first reads the current transmission power and data rate and calculates the vehicle density of the environment. To this end, each vehicle uses the average carrier sense range along with the number of neighboring vehicles detected. Note that the vehicle density is only an estimation that represents the channel load measured in the environment at a given time. Once the vehicles are aware of their state, they send these 3-tuples (p, d, and ρ) to the Python server. Before giving actions back to OMNeT++, the server evaluates the model as many times as there are available actions (per state) to avoid overlooking possible inaccuracies in the training process and to guarantee that proper transmission parameters are reached for every state. From the simulator's viewpoint, the server immediately responds with the action recommended for that state in a single execution time of the algorithm, so our solution is also useful in highly variable scenarios. Finally, each vehicle adjusts its transmission parameters according to the received action.

NNDP allocates data rate and transmission power in a distributed and non-cooperative manner, without relying on any base station or infrastructure. Therefore, we compare it with a similar existing congestion control mechanism called Channel-Aware Congestion Control (CACC) [20]. Basically, CACC adjusts transmission power and data rate according to the cause of packet loss. This is discerned by the Packet Delivery Ratio (PDR) and Packet Collision Rate (PCR) metrics, which, in turn, are based on a given RSS threshold (ξ). Therefore, CACC is able to achieve the optimal MBL = 0.6 but only when the RSS threshold is properly set. For the sake of clarification, we will show how setting different values for the RSS threshold (ξ = −85.72 and 92.26 dBm) may result in different CBR levels. In general, the comparison among the different approaches is conducted for the following metrics:

- Channel Busy Ratio (CBR) is defined as the fraction of time (typically 1 second) in which the channel is busy either due to transmissions or receptions. The CBR indicates the best channel utilization so higher CBR values are closely related to a greater number of packet losses. In these cases, situation awareness is damaged, and the adequate operation of safety applications may be hindered.
- Neighboring vehicles (N). Together with the CBR, the number of neighboring vehicles is essential to provide insight into how information is distributed on the road.
- Packet Delivery Ratio (PDR) is usually defined as the ratio of successfully received packets by all the receivers with respect to the total number of packets

Algorithm 1 NNDP Evaluation (OMNeT++)

$s \leftarrow \langle p_0, d_0, \rho_0 \rangle$
loop over time t
  for all $v \in \mathcal{V}$ **do**
    Calculate r_{CS} according to Eq. (1)
    $\rho \leftarrow \frac{n}{2r_{CS}}$
    $s \leftarrow \langle p, d, \rho \rangle$
    $a \leftarrow \theta(s) = \langle a_p, a_d \rangle$
    $p \leftarrow p + a_p$
    $d \leftarrow d + a_d$
  end for
end loop

TABLE 5. OMNeT++ simulation settings.

Parameter	Value
Channel frequency	5.9 GHz
Channel model fading	Nakagami-m
Path loss exponent (β)	2.5
Shape parameter (m)	2
Sensing power threshold (S)	-92 dBm
SNIR threshold	4 dB
Background noise	-110 dBm
Message size (M)	536 B
Beaconing rate (b)	10 Hz
Min, Max transmission power (p_{min}, p_{max})	1 dBm, 30 dBm
Min, Max data rate (d_{min}, d_{max})	3 dBm, 27 Mbps

transmitted [21], [50], [63], [64]. The PDR is said to be an estimate of situation awareness, intrinsically related to radio channel propagation and medium access control packet losses. Therefore, the highest possible PDR is desirable. Instead, authors of CACC [20] established their own interpretation of the PDR as the relation between the number of decoded packets (N_s) and the sum of decoded packets and packets lost due to weak signal reception (N_w). From our point of view, this definition differs notably from the original definition of the delivery ratio [21], [50], [63], [64] since the authors of CACC did not consider collisions in the total number of packets lost. Despite using the PDR proposed in [20] to implement CACC reliably, we compute the PDR in a traditional way. In our case, the PDR is a transmitter-centric approach, defined as the ratio between the number of packets transmitted that are successfully received at a certain distance and the total number of packets transmitted. Note that this PDR is a function of the distance from which packets are successfully received. More concretely, the PDR is calculated at 50 m steps. This provides more accurate information in terms of transmission power changes and their effects on coverage range, which is of major interest for the problem addressed here.

- Total number of decoded packets (N_s). The total number of beacons successfully received in the entire network under the same scenario also provides additional information about the proper operation of the different algorithms.

The simulations are conducted using a fixed beaconing rate of 10 Hz and a beacon size of 536 bytes. The resulting PHY packet duration and channel capacity will depend on the data rate [27]. For instance, 6 Mbps means a packet duration of 760 μs and a total channel capacity of $C = 1315.78$ beacons per second. All the simulation parameters are specified in Table 5. The different scenarios tested are described below.

A. UNIFORMLY SPACED VEHICLES

To validate our proposed congestion control mechanism, we compare the trained agents (PPO and SAC) in our NNDP solution versus CACC. To this end, we first deploy a simple scenario consisting of a row of evenly spaced vehicles in

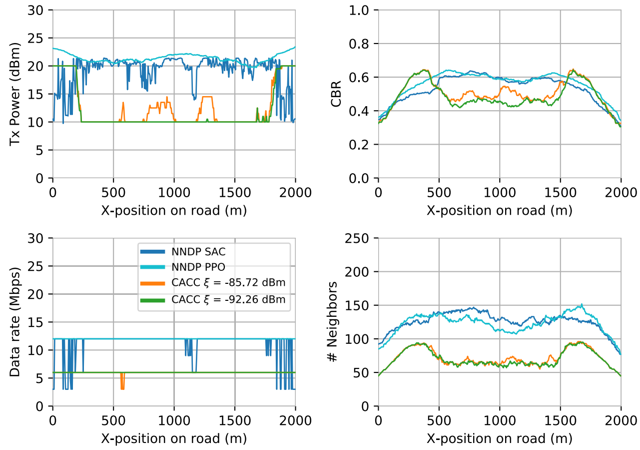


FIGURE 4. Comparison of NNDP and CACC in a congested scenario based on a single row of evenly spaced vehicles.

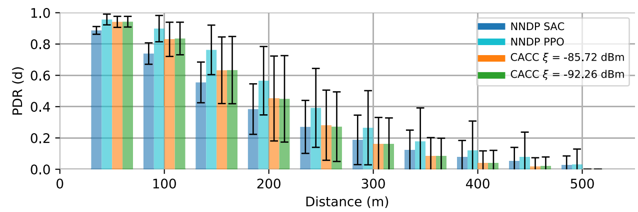
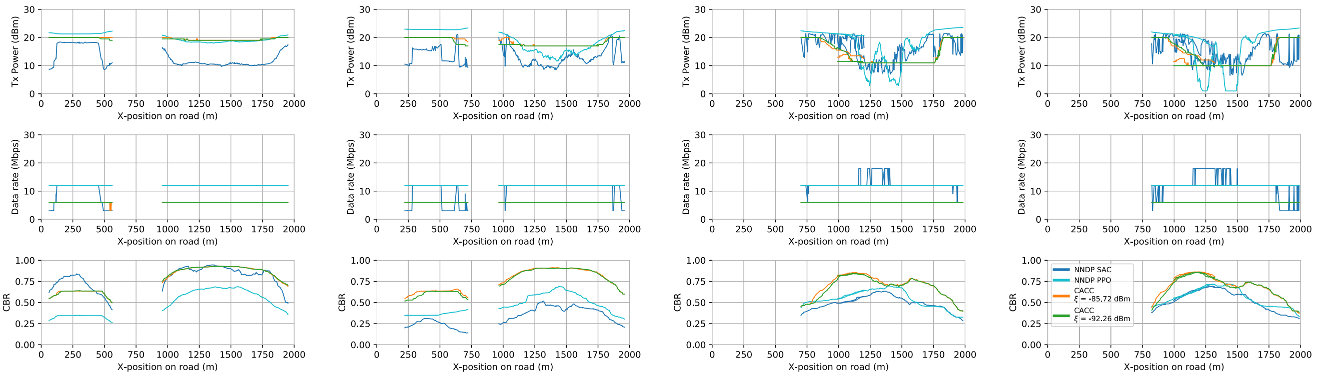


FIGURE 5. The PDR versus distance for a single row of evenly spaced vehicles.

OMNeT++. This scenario seeks a situation in which the channel loads measured by the different neighboring vehicles are similar. In particular, we employ a single row of 400 static vehicles, uniformly distributed along 2000 m. The outcomes of this scenario, after an exhaustive simulation during intervals of 25 s, are shown in Figure 4. As can be observed, the trained PPO and SAC agents lead the vehicles to the desired behavior previously described; that is, the CBR is properly limited to 0.6 by adjusting both transmission power and data rate. Although all the algorithms provide a similar response in terms of channel load, CACC leaves the channel underused with a CBR $= 0.4$, clearly below the MBL. This may be a consequence of its narrower range of available parameters (transmission power is subject to discrete steps of 0.5 dBm and only 3 and 6 Mbps data rates are available). Moving back to NNDP, on top of CBR control, the data rate is intended to be robust against channel fading so NNDP agents attempt to always set the lowest rate while the channel load is successfully limited. Such an effect is illustrated by NNDP SAC at the end of the road, where there is less congestion, and therefore, lower data rates are used. Instead, NNDP PPO chooses not to vary the data rate and to increase transmission power. Note that despite having been trained according to the same rules, each algorithm learns in a different way, which results in different behavior for the vehicles. Higher data rates are allocated by NNDP with regards to CACC while matching the PDR levels. This means that more information has been shared among vehicles so better context awareness



(a) t_{sim} = 1 s **(b)** t_{sim} = 5 s **(c)** t_{sim} = 15 s **(d)** t_{sim} = 20 s

FIGURE 6. Evaluation of different algorithms jointly controlling transmission power and data rates in a realistic traffic jam scenario comprised of two approaching clusters of vehicles. The ongoing progress is described for several simulation times (i.e. 1, 5, 15, and 20 s).

is achieved. Regarding transmission power, it is high enough to satisfy application layer requirements at a certain safety distance ($d_s = 100$ m). More specifically, we seek such received power 100 m away from the transmitter, where the PDR metric is equal to or greater than 0.8, as shown in Figure 5. It should also be pointed out that, on average, NNDP variants reach a similar PDR value to the CACC algorithm, which does not employ any CBR target. This contributes to supporting the idea that 0.6 is a suitable target fraction of channel utilization.

B. TWO RANDOMLY DISTRIBUTED MOVING CLUSTERS

The robustness of NNDP is thoroughly tested in a realistic scenario in which the assumption related to channel load is not satisfied. In this situation, vehicles are not evenly spaced so channel load similarity between close vehicles does not hold. Unlike the first scenario, we employ two different clusters of vehicles bounded within a 500 m and 1000 m long road section, respectively, and located 450 m away from each other. The vehicles are also randomly located in a row in a Poisson distribution of average density $\rho = 0.2$ and 0.4 vehicles per meter, respectively. This results in the first cluster (A) being comprised of approximately 100 vehicles located from 0 to 500 m, an empty road section from 500 to 950 m, and the second cluster (B) composed of about 400 vehicles distributed along the next 1000 m (950 to 1950 m). A realistic traffic jam scenario is represented, in which all the vehicles are heading in the same direction. The vehicles located in the front of cluster A are approaching, supposing free flow, the rear of cluster B. For this purpose, the speed of cluster A is 40 mps, which is considerably higher than the maximum permissible speed of 34 mps, whereas vehicles in cluster B are moving very slowly (2 mps).

This dynamic scenario certainly requires an adaptation of the transmission parameters throughout the entire simulation time to alleviate congestion. For instance, cluster A is lightly congested at the beginning, and this congestion increases as it approaches the second cluster (B). We simulate this scenario for 25 seconds until both clusters come together, increasing vehicle density and provoking channel congestion. Under this premise, all the compared algorithms attempt to reduce channel congestion, mainly by decreasing transmission power, although they show slightly different behavior. As illustrated in Figure 6, both NNDP PPO and NNDP SAC alleviate channel congestion properly by maintaining the CBR around 0.6-0.7. Conversely, CACC exceeds this desired CBR range during the entire simulation time, which would jeopardize the delivery of event-related messages broadcast in emergency cases. Meanwhile, the data rate is set at a constant 6 Mbps by the CACC algorithm. In contrast, NNDP agents better exploit data rate usage, which, acting together with the transmission power, notably reduces channel congestion. However, the NNDP SAC approach attempts to lower the data rate to provide transmissions with more robust modulations. Since the main priority of NNDP is to reduce congestion, this is only possible when the channel is not congested. In fact, when the two clusters join and congestion drastically increases, NNDP SAC increases the data rate to avoid reducing transmission power too much and to maintain PDR levels. As shown in the previous scenario, NNDP PPO is an algorithm that tries to not vary the data rate in a similar way as CACC. The only difference is that, in NNDP PPO the MBL is satisfied through sharp decreases in transmission power, as shown at $t_{sim} =$ 15 and 20 s. As regards the PDR, the bar plot of Figure 7 reveals similar performance to the CACC algorithm. The PDR has been averaged for the entire simulation time and for all the vehicles. This is largely due to the fact that the scenario is now moving, and a more global and robust perspective is required. The standard deviation is included for 10 different distances from 50 to 500 m. In essence, the results obtained illustrate that our proposal attains a similar PDR to CACC. However, NNDP clearly improves it at long distances both for NNDP PPO and NNDP SAC. This means that transmitted beacons reach the farthest neighbors with higher probability, which makes the vehicles aware of risks earlier (e.g., jams).

C. ROBUSTNESS AGAINST ATTENUATION

Despite being trained for certain channel conditions, as stated in Section III, the goal is to demonstrate that NNDP works well even when these conditions vary. To do this, different path loss exponents are tested to verify the robustness of NNDP beyond the training conditions. As described in Algorithm 1, channel load is represented by vehicle density, which is derived from the number of neighboring vehicles divided by twice the average carrier sense range. The carrier sense range depends greatly on channel conditions so it should be updated over time to provide the most accurate estimation. The shape parameter m and the path loss exponent β characterize the severity of fading and attenuation, respectively, whereas the sensitivity of the receiver and the frequency of the carrier remain constant. The shape parameter barely varies vehicle density since the gamma functions in both the numerator and denominator are compensated and, in the remaining terms, the influence of m is almost negligible with regards to changes of the exponent $1/\beta$, as shown in equation 1. Indeed, this is why the path loss exponent β takes a more important in vehicle density estimation than m. This can also be inferred from the results obtained in the simulation of the CACC algorithm [20]. Keeping this premise in mind, we evaluate the previous moving scenario IV-B for different path loss exponent values to demonstrate that the proposed NNDP works properly.

The results achieved are illustrated in Figure 8, employing bar plots and averaging throughout the entire simulation time. Firstly, the carrier sense range is remarkably high when the value of β is set to 2.25, which is considered close to free space attenuation. In this scenario, vehicles receive messages from more vehicles separated by large distances so the channel load increases rapidly. Under these circumstances, all the compared approaches reduce transmission power. Particularly,as congestion increases, NNDP SAC raises the data rate to transmit faster and thus reduces the beaconing load. Conversely, NNDP PPO and CACC keep constant data rates of 12 and 6 Mbps, respectively. The CBR is properly adjusted to the MBL by both NNDP algorithms and, as occured in the previous scenarios, the CACC solution results in a much more congested channel. This could threaten the delivery of event-related messages triggered in emergency situations. Concerning the PDR at 50 m, similar values are obtained by each one of the algorithms analyzed. The rest of the distances, which are not shown in Figure 8, are aligned with the results previously provided for $\beta = 2.5$ in Figure 7. Moreover, as β increases ($\beta = 2.75$), channel attenuation is higher, which (i) reduces the average carrier sense range and, in turn, (ii) senses a fewer number of neighboring vehicles. In this context, the CACC algorithm remains indifferent in terms of data rate, while NNDP and, in particular, the SAC agent, decides to reduce the data rate, resulting in greater robustness over attenuation. This is immediately reflected by reaching a higher PDR. Given less congestion due to higher attenuation, transmission power is slightly increased, which brings the CBR to suitable values. In short, the DNN trained (using both PPO and SAC algorithms) with $\beta = 2.5$ operates appropriately, even when channel conditions vary (i.e. using $\beta = 2.25$, 2.75). NNDP behaves similarly to CACC, which does not depend on channel conditions (β). In fact, it is worth noting that our proposed mechanism not only alleviates congestion but also supports the transmission of information much faster while reaching a similar PDR and greater throughput (total number of decoded packets) than CACC. In other words, NNDP obtains a similar PDR to CACC but with greater throughput and employing higher data rates, which results in improved channel availability for DENM messages (lower CBR).

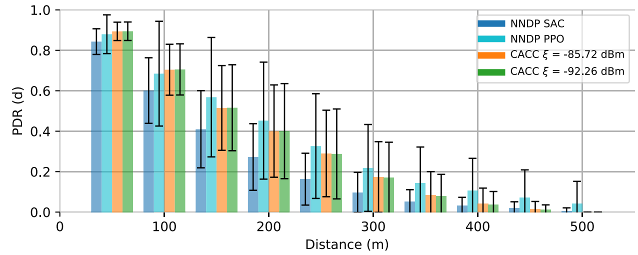


FIGURE 7. The PDR versus distance for two approaching clusters.

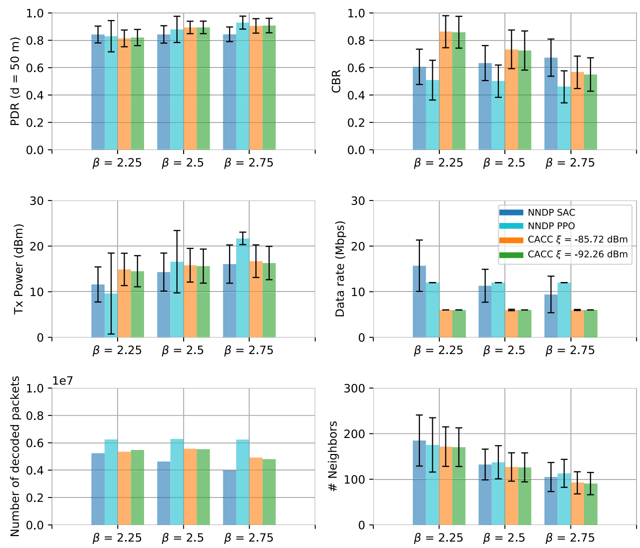


FIGURE 8. Comparison of NNDP and CACC for different path loss exponents and for two approaching clusters.

V. CONCLUSION

Vehicular communications support the transmission of real-time periodic messages (beacons), which allow vehicles to be aware of their changing environment. Most of the safety applications which are conceived to guarantee driver and passenger protection are based on the information exchanged by beacons. However, an increase in beaconing loads may result in higher packet loss and compromise the appropriate functioning of these applications. Therefore, the design of effective congestion control mechanisms, while maintaining a certain fraction of the channel free, is essential for the successful delivery of messages, especially those triggered under emergency incidences. In this paper, we propose an innovative congestion control mechanism that simultaneously tunes transmission power and data rate parameters. Since the associated optimization problem is not convex, ordinary optimization methods are usually inapplicable. Instead, we employ different Deep Reinforcement Learning algorithms.

The proposed mechanism, called NNDP, alleviates congestion in a non-cooperative way, without requiring any additional information from neighbors or centralized infrastructure. Simulation results reveal that NNDP (i) successfully keeps channel loads at the desired levels, leaving channel capacity free enough for successful DENM reception. Once congestion is alleviated, NNDP is intended to (ii) prevent transmission power from reducing too much, guaranteeing a given packet delivery ratio at a certain distance, and (iii) setting the most robust data rate against fading and attenuation whenever possible. Despite being a non-cooperative scheme, all vehicles are geared toward the same goal, which successfully alleviates congestion while reaching higher throughput (number of decoded packets) and a similar PDR to other related mechanisms. The proposed solution operates reasonably well, even in conditions that differ notably from those used in the training environment. Our future work will focus on the study of its cost-effective implementation and improved capabilities to allow the algorithms to learn while driving. Other transmission parameters dependent on the particular hardware of vehicles and their effect on the MAC layer will be also studied.

REFERENCES

[1] L. Liang, H. Peng, G. Y. Li, and X. Shen, "Vehicular communications: A physical layer perspective," *IEEE Trans. Veh. Technol.*, vol. 66, no. 12, pp. 10647–10659, Dec. 2017.

[2] H. Peng, L. Liang, X. Shen, and G. Y. Li, "Vehicular communications: A network layer perspective," *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 1064–1078, Feb. 2018.

[3] *Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Part 2: Specification of Cooperative Awareness Basic Service*, document ETSI 302 637-2 v1.4.1, Aug. 2019.

[4] *Intelligent Transport Systems (ITS); V2X Applications; Part 1: Road Hazard Signalling (RHS) Application Requirements Specification*, document ETSI TS 101 539-1 v1.1.1, Aug. 2013.

[5] *Intelligent Transport Systems (ITS); V2X Applications; Part 2: Intersection Collision Risk Warning (ICRW) Application Requirements Specification*, document ETSI TS 101 539-2 v1.1.1, Jun. 2018.

[6] *Intelligent Transport Systems (ITS); V2X Applications; Part 3: Longitudinal Collision Risk Warning (LCRW) Application Requirements Specification*, document ETSI TS 101 539-3 v1.1.1, Nov. 2013.

[7] *Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Part 3: Specifications of Decentralized Environmental Notification Basic Service*, document ETSI TS 102 637-3 v1.1.1, Intelligent Transport Systems (ITS), 2010.

[8] G. Bansal, J. B. Kenney, and C. E. Rohrs, "LIMERIC: A linear adaptive message rate algorithm for DSRC congestion control," *IEEE Trans. Veh. Technol.*, vol. 62, no. 9, pp. 4182–4197, Nov. 2013.

[9] T. Tielert, D. Jiang, Q. Chen, L. Delgrossi, and H. Hartenstein, "Design methodology and evaluation of rate adaptation based congestion control for vehicle safety communications," in *Proc. IEEE Veh. Netw. Conf. (VNC)*, Nov. 2011, pp. 116–123.

[10] J. Aznar-Poveda, E. Egea-Lopez, A.-J. Garcia-Sanchez, and P. Pavon-Mariño, "Time-to-collision-based awareness and congestion control for vehicular communications," *IEEE Access*, vol. 7, pp. 154192–154208, 2019.

[11] M. Joseph, X. Liu, and A. Jaekel, "An adaptive power level control algorithm for DSRC congestion control," in *Proc. 8th ACM Symp. Design Anal. Intell. Veh. Netw. Appl. (DIVANet)*, 2018, pp. 57–62.

[12] O. M. Akinlade, "Adaptive transmission power with vehicle density for congestion control," M.S. thesis, Univ. Windsor, Windsor, ON, Canada, 2018. [Online]. Available: https://scholar.uwindsor.ca/etd/7420etd/7420

[13] E. Egea-Lopez, "Fair distributed congestion control with transmit power for vehicular networks," in *Proc. IEEE 17th Int. Symp. World Wireless, Mobile Multimedia Netw. (WoWMoM)*, Jun. 2016, pp. 1–6.

[14] H. Amer, H. Al-Kashoash, M. J. Khami, M. Mayfield, and L. Mihaylova, "Non-cooperative game based congestion control for data rate optimization in vehicular ad hoc networks," *Ad Hoc Netw.*, vol. 107, Oct. 2020, Art. no. 102181.

[15] SAE International, "On-board system requirements for V2V safety communications," SAE Tech. Rep. J2945/1, 2019.

[16] F. Goudarzi, H. Asgari, and H. S. Al-Raweshidy, "Fair and stable joint beacon frequency and power control for connected vehicles," *Wireless Netw.*, vol. 25, pp. 1–12, Nov. 2019.

[17] G. S. Rawat and K. Singh, "Joint beacon frequency and beacon transmission power adaptation for internet of vehicles," *Trans. Emerg. Telecommun. Technol.*, p. e4124, 2020, doi: 10.1002/ett.4124.

[18] E. Egea-Lopez and P. Pavon-Mariño, "Fair congestion control in vehicular networks with beaconing rate adaptation at multiple transmit powers," *IEEE Trans. Veh. Technol.*, vol. 65, no. 6, pp. 3888–3903, Jun. 2016.

[19] J. Aznar-Poveda, A.-J. Garcia-Sanchez, E. Egea-Lopez, and J. Garcia-Haro, "MDPRP: A Q-Learning approach for the joint control of beaconing rate and transmission power in VANETs," *IEEE Access*, vol. 9, pp. 10166–10178, 2021.

[20] B.-M. Cho, M.-S. Jang, and K.-J. Park, "Channel-aware congestion control in vehicular cyber-physical systems," *IEEE Access*, vol. 8, pp. 73193–73203, 2020.

[21] L. J. Wei and J. M.-Y. Lim, "Identifying transmission opportunity through transmission power and bit rate for improved VANET efficiency," *Mobile Netw. Appl.*, vol. 24, no. 5, pp. 1630–1638, Oct. 2019.

[22] S. Sharma and B. Singh, "Context aware autonomous resource selection and Q-learning based power control strategy for enhanced cooperative awareness in LTE-V2V communication," *Wireless Netw.*, vol. 26, pp. 1–16, Aug. 2020.

[23] H. Ye, G. Y. Li, and B.-H. F. Juang, "Deep reinforcement learning based resource allocation for V2V communications," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 3163–3173, Apr. 2019.

[24] E. Ghadimi, F. D. Calabrese, G. Peters, and P. Soldati, "A reinforcement learning approach to power control and rate adaptation in cellular networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–7.

[25] K. K. Nguyen, T. Q. Duong, N. A. Vien, N.-A. Le-Khac, and N. M. Nguyen, "Non-cooperative energy efficient power allocation game in D2D communication: A multi-agent deep reinforcement learning approach," *IEEE Access*, vol. 7, pp. 100480–100490, 2019.

[26] Z. Li and C. Guo, "Multi-agent deep reinforcement learning based spectrum allocation for D2D underlay communications," *IEEE Trans. Veh. Technol.*, vol. 69, no. 2, pp. 1828–1840, Feb. 2019.

[27] *Intelligent Transport Systems (ITS); Access Layer Specification for Intelligent Transport Systems Operating in the 5 GHz Frequency Band*, document ETSI EN 302 663 v1.3.0, May 2019.

[28] *Intelligent Transport Systems (ITS); Cross Layer DCC Management Entity for Operation in the Its G5A and Its G5B Medium*, document ETSI TS 103 175 v1.1.1, Jun. 2015.

[29] T. Lorenzen and H. Tchouankem, "Evaluation of an awareness control algorithm for VANETs based on ETSI EN 302 637-2 V1.3.2," in *Proc. IEEE Int. Conf. Commun. Workshop (ICCW)*, Jun. 2015, pp. 2458–2464.

[30] N. Lyamin, A. Vinel, M. Jonsson, and B. Bellalta, "Cooperative awareness in VANETs: On ETSI EN 302 637-2 performance," *IEEE Trans. Veh. Technol.*, vol. 67, no. 1, pp. 17–28, Jan. 2018.

[31] H.-H. Nguyen and H.-Y. Jeong, "Mobility-adaptive beacon broadcast for vehicular cooperative safety-critical applications," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 6, pp. 1996–2010, Jun. 2018.

[32] F. Lyu, M. Li, and X. Shen, "Safety-aware and distributed beacon congestion control," in *Vehicular Networking for Road Safety*. Cham, Switzerland: Springer, 2020, pp. 129–157.

[33] W. Li, W. Song, Q. Lu, and C. Yue, "Reliable congestion control mechanism for safety applications in urban VANETs," *Ad Hoc Netw.*, vol. 98, Mar. 2020, Art. no. 102033.

[34] B. Shabir, M. A. Khan, A. U. Rahman, A. W. Malik, and A. Wahid, "Congestion avoidance in vehicular networks: A contemporary survey," *IEEE Access*, vol. 7, pp. 173196–173215, 2019.
[35] R. Aslani and M. Rasti, "A distributed power control algorithm for energy efficiency maximization in wireless cellular networks," *IEEE Wireless Commun. Lett.*, vol. 9, no. 11, pp. 1975–1979, Nov. 2020.
[36] Y. S. Nasir and D. Guo, "Multi-agent deep reinforcement learning for dynamic power allocation in wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 10, pp. 2239–2250, Oct. 2019.
[37] B. Aygun, M. Boban, and A. M. Wyglinski, "ECPR: Environment-and context-aware combined power and rate distributed congestion control for vehicular communications," *Comput. Commun.*, vol. 93, pp. 3–16, Nov. 2016.
[38] Y. Yoon and H. Kim, "Balancing power and rate control for improved congestion control in cellular V2X communication environments," *IEEE Access*, vol. 8, pp. 105071–105081, 2020.
[39] A. Haider and S.-H. Hwang, "Adaptive transmit power control algorithm for sensing-based semi-persistent scheduling in C-V2X mode 4 communication," *Electronincs*, vol. 8, no. 8, p. 846, 2019.
[40] H. Chang, Y. E. Song, H. Kim, and H. Jung, "Distributed transmission power control for communication congestion control and awareness enhancement in VANETs," *PLoS ONE*, vol. 13, no. 9, Sep. 2018, Art. no. e0203261.
[41] M. Wang *et al.*, "Research on adaptive beacon message transmission power in VANETs," *J. Ambient Intell. Humanized Comput.*, 2020, doi: 10.1007/s12652-020-02575-x.
[42] A. Rostami, H. Krishnan, and M. Gruteser, "V2V safety communication scalability based on the SAE J2945/1 standard," in *Proc. Workshop ITS Amer.*, 2018, pp. 1–10.
[43] M. Sepulcre, J. Gozalvez, and M. C. Lucas-Estañ, "Power and packet rate control for vehicular networks in multi-application scenarios," *IEEE Trans. Veh. Technol.*, vol. 68, no. 9, pp. 9029–9037, Sep. 2019.
[44] S. Bolufé, S. Montejo-Sánchez, C. A. Azurdia-Meza, S. Céspedes, R. D. Souza, and E. M. G. Fernandez, "Dynamic control of beacon transmission rate and power with position error constraint in cooperative vehicular networks," in *Proc. 33rd Annu. ACM Symp. Appl. Comput.*, Apr. 2018, pp. 2084–2091.
[45] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 2018.
[46] D. Jiang, Q. Chen, and L. Delgrossi, "Optimal data rate selection for vehicle safety communications," in *Proc. 5th ACM Int. Workshop Veh. Inter-NETworking (VANET)*, 2008, pp. 30–38.
[47] M. Sepulcre, J. Gozalvez, and B. Coll-Perales, "Why 6 Mbps is not (always) the optimum data rate for beaconing in vehicular networks," *IEEE Trans. Mobile Comput.*, vol. 16, no. 12, pp. 3568–3579, Apr. 2017.
[48] C.-L. Huang, Y. Fallah, R. Sengupta, and H. Krishnan, "Adaptive intervehicle communication control for cooperative safety systems," *IEEE Netw.*, vol. 24, no. 1, pp. 6–13, Jan. 2010.
[49] S. Subramanian, M. Werner, S. Liu, J. Jose, R. Lupoaie, and X. Wu, "Congestion control for vehicular safety: Synchronous and asynchronous mac algorithms," in *Proc. 9th ACM Int. Workshop Veh. Inter-Netw., Syst., Appl.*, 2012, pp. 63–72.
[50] X. Ma, J. Zhang, and T. Wu, "Reliability analysis of one-hop safety-critical broadcast services in VANETs," *IEEE Trans. Veh. Technol.*, vol. 60, no. 8, pp. 3933–3946, Oct. 2011.
[51] *IEEE Standard for Information Technology-Telecommunications and Information Exchange Between Systems-Local and Metropolitan Area Networks-Specific Requirements Part 11: Wireless Lan Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, IEEE Standard 802.11, 2007.
[52] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "A brief survey of deep reinforcement learning," 2017, *arXiv:1708.05866*. [Online]. Available: http://arxiv.org/abs/1708.05866
[53] J. X Wang, Z. Kurth-Nelson, D. Tirumala, H. Soyer, J. Z Leibo, R. Munos, C. Blundell, D. Kumaran, and M. Botvinick, "Learning to reinforcement learn," 2016, *arXiv:1611.05763*. [Online]. Available: http://arxiv.org/abs/1611.05763
[54] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, and S. Petersen, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.
[55] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," 2015, *arXiv:1509.02971*. [Online]. Available: http://arxiv.org/abs/1509.02971
[56] S. Fujimoto, H. Hoof, and D. Meger, "Addressing function approximation error in actor-critic methods," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 1587–1596.
[57] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," 2017, *arXiv:1707.06347*. [Online]. Available: http://arxiv.org/abs/1707.06347
[58] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 1861–1870.
[59] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, and R. Kern, "Array programming with numpy," *Nature*, vol. 585, no. 7825, pp. 357–362, 2020.
[60] A. Raffin, A. Hill, M. Ernestus, A. Gleave, A. Kanervisto, and N. Dormann. (2019). *Stable Baselines3*. [Online]. Available: https://github.com/DLR-RM/stable-baselines3
[61] A. Varga, "OMNet++," in *Modeling and Tools for Network Simulation*. Berlin, Germany: Springer, 2010, pp. 35–59.
[62] *INET Framework Inet Framework*. Accessed: Jul. 3, 2019. [Online]. Available: https://inet.omnetpp.org
[63] S. Malik and P. K. Sahu, "A comparative study on routing protocols for VANETs," *Heliyon*, vol. 5, no. 8, Aug. 2019, Art. no. e02340.
[64] B. Choudhury, V. K. Shah, A. Dayal, and J. H. Reed, "Joint age of information and self risk assessment for safer 802.11p based V2V networks," 2020, *arXiv:2012.04774*. [Online]. Available: http://arxiv.org/abs/2012.04774

JUAN AZNAR-POVEDA received the bachelor's and master's degrees in telecommunications engineering from the Universidad Politecnica de Cartagena, Spain, in 2016 and 2018, respectively, where he is currently pursuing the Ph.D. degree with the Information Technologies and Communications Department. His research interests include vehicular networks, deep reinforcement learning, machine learning, optimization algorithms, electronics, and electrochemical sensing. His bachelor's final project was awarded 2nd place in the Liberalization Telecommunications Award, given by the National Association of Technical Telecommunication Engineers, in 2016, and his master's final project was awarded the Extraordinary Master Award, by the Universidad Politecnica de Cartagena.

ANTONIO-JAVIER GARCIA-SANCHEZ received the bachelor's degree in industrial electronic engineering and the master's degree in industrial engineering from the Technical University of Cartagena (UPCT), Spain, in 1997 and 2000, respectively, and the Ph.D. degree in telecommunication engineering from the Information and Communication Technologies Department (DTIC), UPCT, in 2005. Since 2002, he has been with DTIC, UPCT. He is currently a Full Professor at UPCT, Telematic Engineering area, and the Head of DTIC. He is the coauthor of more than 100 conferences and journal papers, 50 of them indexed in the Journal Citation Report (JCR). His research interests include the areas of wireless sensor networks (WSNs) and the Internet of Things, streaming services, intelligence artificial applied to communication networks, smart data processing, and nanocommunications. He has belonged to more than 35 conferences/workshops as a TPC Staff and to eight as a member of the organizing committee. He has led or leads many regional/national projects and private contracts remarking those related to Proof of Concept whose goal is to place on the consumer market devices, procedures, and tools. He has been responsible for a European H2020 project. In addition, he currently leads a project related to the COVID-19 area funded by CSIC and Banco Santander. Finally, in the field of technology transfer, he is a Co-Inventor of 14 patents/intellectual property, having been recognized by the Ministry of Science and Innovation a six-year transfer. Since July 2020, he has been a Co-Founder and a CTO of the spinoff QARTECH Innovations, dedicated to endow Industria 4.0 with high-performance technology. He is a Reviewer of notable communication journals listed in the ISI-JCR, highlighting IEEE INTERNET OF THINGS journal, IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, and IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT, COMPUTER NETWORKS.

ESTEBAN EGEA-LOPEZ received the bachelor's degree in telecommunications engineering from the Universidad Politecnica de Valencia (UPV), Spain, in 2000, the master's degree in electronics from the University of Gavle, Sweden, in 2001, and the Ph.D. degree in telecommunications from the Universidad Politecnica de Cartagena (UPCT), in 2006. He is currently an Associated Professor with the Department of Information Technologies and Communications, UPCT. His research interests include vehicular networks and MAC protocols.

JOAN GARCÍA-HARO (Member, IEEE) received the M.S. and Ph.D. degrees in telecommunication engineering from the Universitat Politecnica de Catalunya, Spain, in 1989 and 1995, respectively. He has been a Visiting Scholar with Queens University at Kingston, Canada, from 1991 to 1992, and Cornell University, Ithaca, USA, from 2010 to 2011. He is currently a Professor with the Universidad Politecnica de Cartagena. He is the author or coauthor of more than 80 journal articles mainly in the fields of switching, wireless networking and performance evaluation. He also received an Honorable Mention for the IEEE Communications Society Best Tutorial Paper Award, in 1995. He served as the Editor-in-Chief for the IEEE GLOBAL COMMUNICATIONS NEWSLETTER, included in the *IEEE Communications Magazine*, from April 2002 to December 2004. He has been a technical editor of the same magazine, from March 2001 to December 2011.