iVRLS: In-coverage Vehicular Reinforcement Learning Scheduler

Taylan Şahin*, Mate Boban†, Ramin Khalili† and Adam Wolisz*
*Telecommunication Networks Group, Technische Universität Berlin, 10587 Berlin, Germany
†Huawei Munich Research Center, Huawei Technologies Duesseldorf GmbH, 80992 Munich, Germany
Emails: taylan.sahin.1@campus.tu-berlin.de, {mate.boban, ramin.khalili}@huawei.com, adam.wolisz@tu-berlin.de

***Abstract*—Cellular networks enable high reliability of vehicle-to-vehicle (V2V) communications thanks to centralized, efficient coordination of radio resources. Collision-free transmissions are possible, where base stations could allocate orthogonal resources to the vehicles. However, in case of limited resources in relation to the data traffic load, the resource allocation task becomes a challenge. Current solutions propose heuristic algorithms that focus on resource reuse, often based on the location of the vehicles. Such schedulers are mainly designed assuming ideal network coverage conditions and are prone to performance degradation in case of coverage loss. Further, they typically rely on frequent scheduling updates, which increases the dependency on coverage. In this paper, we propose a reinforcement learning-based approach to scheduling V2V communications. Our solution, called iVRLS, delivers higher reliability than an enhanced version of a state-of-the-art benchmark algorithm in case of intermittent coverage conditions, while requiring less frequent scheduling. Following this approach, we enable a unified scheduler deployment irrespective of coverage, which offers graceful performance behavior across varying coverage conditions, thus making iVRLS a robust alternative to existing schedulers.**

***Index Terms*—V2V, Reinforcement Learning, Scheduling, Radio Resource Allocation, Coverage**

I. Introduction

Radio resource allocation plays an important role in the performance of vehicle-to-vehicle (V2V) communications. Using distributed mechanisms, vehicles can select the resources themselves autonomously, such as by sensing other vehicles' transmissions. Yet, distributed methods are prone to collisions due to limited local view of the vehicles. Higher reliability of V2V communications could be achieved with the use of a centralized network entity, e.g., cellular base station, which coordinates the V2V transmission resources (cf. [1], [2], [3]). The centralized entity could allocate collision-free, e.g., time-orthogonal resources, to vehicles. However, given limited amount of resources in time and frequency, and demanding V2V traffic within a dynamic environment of vehicles, centralized resource allocation necessitates efficient algorithms to maintain V2V communication reliability.

Current solutions in the literature propose heuristics based on locations of vehicles to enable resource reuse. Nevertheless, the state-of-the-art algorithms typically assume ideal coverage conditions, without any loss of the control signaling between the base station and vehicles used to request and assign the resources. Furthermore, the algorithms rely on high-frequency, dynamic scheduling updates, which actually increases their dependency on control signaling reliability, and requires larger amount of resources for control signaling. Therefore, algorithms that operate at least equally well under realistic, intermittent coverage conditions would be definitely beneficial.

In this paper, we consider two ways to meet design of a such improved solution. On one hand, we propose a simple extension of existing centralized schedulers, where vehicles keep using their existing resources assigned by the network, extended for a (pre-)defined duration, e.g., until they connect back to the network. This would mitigate degradation of V2V communication quality due to otherwise deferred or delayed transmissions, or vehicles switching to autonomous resource allocation methods. On the other hand, we propose iVRLS (in-coverage vehicular reinforcement learning (RL) scheduler), a centralized RL-based solution similar to the VRLS presented in [4] as an efficient solution for the well-known out-of-coverage parts of the route (e.g., a tunnel). In comparison to VRLS that *pre-schedules* resources by means of one-time scheduling assignments before vehicles leave the coverage, iVRLS learns a scheduling policy by taking the advantage of assignments that are possible at all times and making use of instantaneous and exact knowledge of vehicular mobility. This enables better prediction of vehicle locations and resource utilization across varying conditions of coverage.

Our main contributions in this paper are as follows:

- We propose iVRLS, an RL-based centralized scheduler for V2V communications in coverage.
- We propose a simple yet effective enhancement to centralized scheduling of V2V communications to support intermittent coverage conditions, which could be applied to any state-of-the-art centralized algorithm.
- We evaluate the reliability performance of iVRLS under non-ideal coverage conditions within realistic vehicular network environments. In comparison to the enhanced version of a state-of-the-art centralized scheduling algorithm [5], iVRLS achieves better performance under lossy coverage conditions and relatively low frequency of scheduling updates.
- We demonstrate that iVRLS offers a unified solution for deployment irrespective of coverage, enabling simplified implementation and robustness to coverage variations in the network.

In the rest of the paper, Section II presents the related work; Section III provides the system model, defines the problem, and presents our enhancement to existing schedulers. Section IV describes iVRLS, and Section V presents the evaluation results. Finally, Section VI concludes the paper.

II. Related Work

Algorithms proposed in the literature for centralized scheduling of V2V communications are mainly based on spatial reuse of the resources, taking the vehicles' location into account [2], [5], [6], [7]. A resource allocation algorithm designed for superior reliability is proposed in [6], called "allocation with Maximum reuse Distance" (MD). MD uses a simple yet powerful heuristic, which allocates time resources to all vehicles in cyclic order following their positions. Thus, MD tries to maximize the average distance between the vehicles using the same resource, with the goal of minimizing the interference. MD is analytically shown to outperform other centralized scheduling algorithms in the literature from [2] and [7], as well as the random resource allocation. Yet, as also the authors indicate, MD is far from practical implementation in reality. The scheduling assignments are required to be sent for *all* vehicles in the environment simultaneously at a time, repeating with the V2V message generation rate, thus leading to impractical processing and signaling overhead in the network. Besides, MD considers only a single resource in the frequency dimension for the assignments (hence could only assign different time-resources).

A more practical version of MD is proposed in [5], with a similar name "Maximum Reuse Distance" (MRD). Sharing the same goal of maximizing the distance between reused resources with MD, MRD assigns a vehicle the last-occupied time resource, and within that resource, the last-occupied frequency resource, when vehicles ordered with respect to their distance to the requesting vehicle. In case unoccupied resources remain in the resource pool, then instead it assigns one of them randomly. MRD does not rely on heavy processing and signaling as MD, as it schedules vehicles asynchronously, in *multiples* of their message generation periodicity, by sending a single assignment to a single vehicle at a time. MRD is also shown to outperform the same benchmark algorithm in [2].

There are several other centralized resource allocation algorithms in the literature, however, considering impractical assumptions. An early work [8] within the framework of device-to-device (D2D) communications proposes an algorithm based on power control, using the channel state information (CSI) of all V2V links, which becomes infeasible in broadcast scenarios. Authors in [9] consider an algorithm based on clustering of vehicles and applying graph-based solutions, however, which is only applicable to road intersection scenarios. Other works by the same authors that propose clustering-based approaches [10] [11] are also challenging in terms of implementation due to their requirement of careful re-forming of clusters as vehicles move, which brings impractical complexity and processing overhead to the scheduler. Applicability of some other cluster-based solutions such as [12] are also limited to unicast or multicast V2V scenarios. There are also a large number of works assuming underlay conditions, i.e., V2V communications using the resources shared with cellular uplink and/or downlink communications, such as [13] and [14], which we do not see relevant in this study due to their different optimization tasks, such as maximizing the sum rate of cellular users.

RL within the context of resource allocation for vehicular networks has been applied so far mainly in the form of distributed algorithms or for the goal of communication mode selection between cellular and V2V links, as in [15] and [16], respectively. Based on our best knowledge, no RL-based algorithm is proposed for centralized in-coverage scheduling of V2V communications, up to now. In our previous work [4], we proposed a centralized RL-based scheduler, called VRLS, for resource allocation of out-of-coverage V2V communications. In this paper, we propose iVRLS, which follows a similar approach to the case of in-coverage scheduling, including the conditions of unstable connectivity. iVRLS learns a scheduling policy based solely on its experience with the vehicular network, without any pre-defined heuristic rule. In our evaluations, we select MRD algorithm in [5] to serve as a baseline, due to its realistic and practical assumptions, as well as its benchmarked performance.

III. System Model

A. Vehicular Network Environment

We consider a highway scenario where vehicles transmit broadcast periodic V2V traffic, as illustrated in Fig. 1. Vehicles transmit single-hop cooperative messages with periodicity T_m and size B_m to inform neighbors about their presence and status including current location, speed, etc. V2V transmissions use radio resources scheduled by the centralized network entity, a cellular base station (BS). Vehicles send scheduling requests (SRs) to the BS to request resources for their V2V transmissions, which contains their V2V traffic information such as T_m and B_m. In turn, BS informs the vehicles about the scheduling decision by sending scheduling assignments (SAs). In addition, BS can request vehicles to report their mobility information pertaining their location, speed, etc. to help making scheduling decisions.

Scheduling consists of assigning a single time-frequency resource r from a periodically-repeating resource pool with R resources, as shown in Fig. 1. Resource pool is organized in time slots and frequency subchannels, each being able to contain a single V2V message, given the modulation and coding scheme (MCS). Each vehicle sends SR with periodicity $T_{SR} = N_{SR}T_m$ (in multiples of T_m), starting from its first message generation after connecting to the network. Vehicle then keeps using the same scheduled resource for its V2V transmissions within that period, i.e., until the next SR/SA. Mobility information is requested by BS on demand; specifically, at every scheduling instance, which can be acquired via different positioning methods such as based on network-signaling or global positioning system [17]. SR, SA, and mobility information messages are transmitted in dedicated

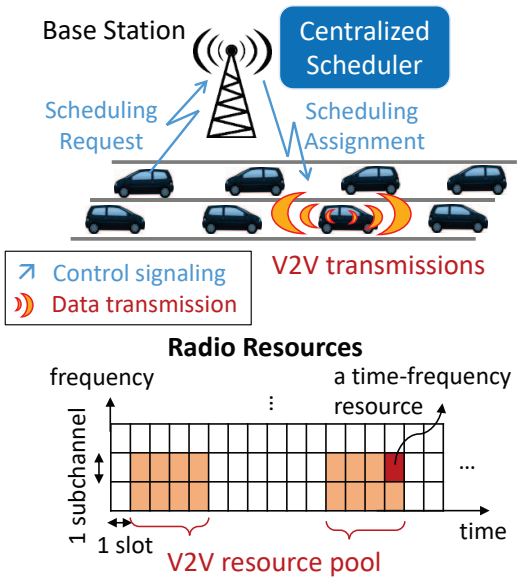


Fig. 1. System model with radio resources and the resource pool configured by the network for V2V communications.

control channels over uplink (UL) and downlink (DL), using resources different than the V2V resource pool.

Transmitted V2V messages are subject to path loss and fading effects of the wireless channel, besides the interference from any other transmission using the same resource. Thus, successful reception of a message depends on the signal-to-interference-plus-noise ratio (SINR) at the receiver. Vehicles are further assumed to be equipped with half-duplex (HD) radios for V2V communications, that is, they can either transmit or receive, but can not do both at a given time slot.

B. V2V Communication Reliability

We quantify the reliability of V2V transmissions with packet reception ratio (PRR), which is a metric specified by the 3GPP standard [18]. PRR of a single transmitted V2V message is calculated by X/Y, where Y is the number of vehicles located in the range (a, b) from the transmitter, and X is the number of vehicles with successful reception among Y. Average PRR is then calculated as $(X_1 + X_2 + X_3 + ... + X_n)/(Y_1 + Y_2 + Y_3 + ... + Y_n)$ with $a = i \times 20$ m, $b = (i + 1) \times 20$ m for $i = 0, 1, ...$ and n denoting the number of generated messages by all vehicles in the simulated duration [18].

C. Intermittent Coverage and Our Goal

Similar to V2V links, the link between the vehicles and the BS station is also subject to intermittent loss due to path loss and fading characteristics of the environment. Vehicles might lose connectivity to BS for a short duration, e.g., due to a physical blocking object or instantaneous fading in the wireless channel, resulting in the control channel errors. This leads to deferred V2V transmissions, degrading the reliability of V2V communications. If such conditions persist, e.g., a longer duration of coverage loss due to a road tunnel, vehicles resort to autonomous resource selection modes, such as the sensing-based semi-persistent scheduler Mode 4 defined by the 3GPP standard [19]. However, distributed methods suffer from limited view of the vehicles due to local and transient measurements, which degrades the quality of V2V communications [1], [2], [3].

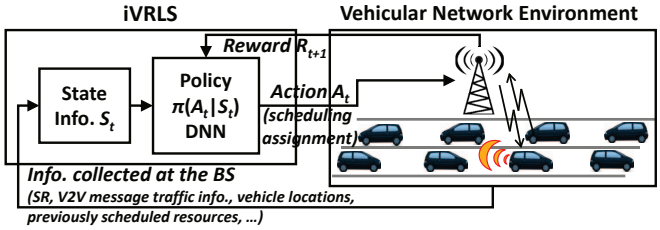


Fig. 2. In-coverage Vehicular Reinforcement Learning Scheduler (iVRLS).

Our objective is, therefore, to maximize reliability of the V2V transmissions by centralized scheduling of the resources under lossy conditions of network coverage. Achieving reliable V2V transmissions becomes a challenging task given limited amount of resources that needs to be managed for highly dynamic vehicular users. The centralized scheduler should minimize interfering transmissions, while taking the HD-limitation of the vehicle radios and the scheduling periodicity into account.

D. Proposed Enhancement for Centralized Scheduling to Support Intermittent Coverage

To account for intermittent network coverage loss, we first propose a simple enhancement to centralized V2V scheduling, which is in general applicable to any existing solution: upon experiencing the coverage loss, vehicles keep using their latest resource assignments for V2V communications until they connect back to the network and request a new resource. In our evaluations, we apply this enhancement to the MRD algorithm, so as to enable a fair comparison with iVRLS.

IV. iVRLS: In-coverage Vehicular Reinforcement Learning Scheduler

iVRLS targets the centralized scheduling problem using the framework of RL, as depicted in Fig. 2. To make scheduling decisions, iVRLS utilizes the information collected from the network, i.e., BSs, taking it in the form of a *state representation* as input. Given the current state of the network, iVRLS provides the scheduling assignments, defined as *actions*, based on its trained *policy*. Policy of iVRLS is modeled as a deep neural network (DNN), whose parameters are trained using state-of-the-art *actor-critic RL algorithm*. Training is conducted with a *reward* signal provided to iVRLS upon its each action, indicating how "good" the action was – in our case, the reliability of V2V transmissions. Our design of iVRLS is based on that of VRLS [4], with the main difference being the information used in the state representation of the environment, and the possibility of taking actions anytime and anywhere within the coverage of the base stations.

A. State Representation

State representation S_t, at the instance t a vehicle is requesting a resource, contains the information collected from the vehicular network environment in a compact and useful way. We design S_t to indicate expected "interference" level I_r on each resource r in the configured V2V resource pool, in case it was assigned to the requesting vehicle. We represent I_r by the inverse of the distance between the vehicles using the same resource and the requesting vehicle, over current and future instances of their message transmissions. Specifically, $S_t = [I_{r=1}, I_{r=2}, ..., I_{r=R}]$ is a vector with number of elements equal to the number of resources R in the resource pool, with each element $I_r = I_r^{u=1} + I_r^{u=2} + I_r^{u=3} + ...$ representing the sum of expected "interference" coming from the set of vehicles $\{u = 1, 2, 3, ...\}$ using the same resource r. In case no vehicle using a given resource, $I_r = 0$. Otherwise, I_r^u for a single vehicle u using the resource r is calculated as:

$$I_r^u = \frac{1}{N_{SR}}\left(\frac{1}{|\Delta x|} + \frac{1}{|\Delta x + \Delta v T_m|} + \frac{1}{|\Delta x + \Delta v 2T_m|} + ... + \frac{1}{|\Delta x + \Delta v (N_{SR}-1)T_m|}\right), \quad (1)$$

$$= \frac{1}{N_{SR}} \sum_{n=1}^{N_{SR}} \frac{1}{|\Delta x + \Delta v (n-1) T_m|}, \quad (2)$$

where $\Delta x = x_u - x_{req}$ and $\Delta v = v_u - v_{req}$ are the relative distance and the speed between the vehicle u and the requesting vehicle, respectively. Accordingly, the first term in the parentheses represents the "current interference", while every other term indicating the "expected interference" in the future, based on the changing positions of vehicles over time with message generation periodicity T_m, until the next scheduling event. Further, in case of a known out-of-coverage event by the network, such as due to a tunnel, the time to next scheduling event is calculated by dividing the length of the tunnel to the average speed of vehicles in the environment. An average over current and future instances is then taken to represent "overall" interference. In order to avoid singularity, we further modify the denominator in the sum as:

$$I_r^u = \frac{1}{N_{SR}} \sum_{n=1}^{N_{SR}} \frac{1}{\max(\Delta x_{min}, |\Delta x + \Delta v (n-1) T_m|)}, \quad (3)$$

where $\Delta x_{min} > 0$ is a fixed, minimum inter-vehicle distance. Overall, higher value of I_r indicates higher expected "interference" on a resource, in case assigned to the requesting vehicle.

B. Action Definition

Action A_t of iVRLS denotes assigning a single resource r to the vehicle requesting at instance t, from the resource pool configured for V2V communications. In case more than one vehicle request a resource at the same time, the actions are taken in a random order. iVRLS selects the resource to be assigned based on its policy $\pi(A_t|S_t) \rightarrow [0, 1]_R$, which is a mapping from state of the vehicular environment to a probability distribution over the set of possible actions, i.e., resources. The resource is then selected at random according to the distribution.

C. Reward

In order to train iVRLS, the reward R_{t+1} we provide upon its each action A_t is directly proportional to the reliability of V2V transmissions in the network. This is aligned with the aim of RL, which is to maximize the collected reward in the long run. Specifically, $R_{t+1} = -10 \times (1 - \text{PRR})$, i.e., a linear function of PRR, measured for the range of interest, e.g., a certain communication range required by a V2V use case. In case no transmission take place between the actions, e.g., when two vehicles request a resource at the same time, the reward of the previous action is provided.

D. Policy Deep Neural Network and the Training Algorithm

We represent the policy of iVRLS with a deep neural network (DNN), given the large number of possible state and action combinations [20]. The parameters of the policy DNN is trained using the state-of-the-art actor-critic RL algorithm. "Actor" refers to the policy taking actions, while "critic" is the *value function* that "judges" the actor's policy. In its simple terms, value function estimates the long-term collected rewards starting from the given state and following the policy onwards. We also approximate the value function with a DNN, given the large state space of our problem. Due to space considerations, we refer the reader to [4] for further details of the utilized DNNs and the training algorithm with their parameters.

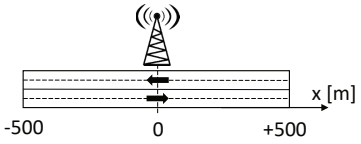
V. Evaluation

A. Simulation Setup

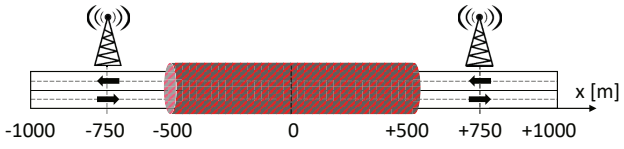
We evaluate the performance of iVRLS in two scenarios A and B that represent various network coverage conditions, as illustrated in Fig. 3. Scenario A contains a single BS that serves a large area, where edge-sections of the road are prone to higher loss of coverage due to increased channel path loss on UL/DL. In Scenario B, two BSs serve smaller areas, with higher frequency of scheduling, thus providing a better coverage accompanied by increased control on V2V transmissions. Yet, there is a "coverage gap" (a tunnel) between the two BSs, where it is not possible to signal for any scheduling at all. Specifically, in Scenario A, the highway is of 1000 m length, having the BS deployed at its center with a 45 m longitudinal offset. Vehicles send scheduling requests every $T_{SR} = 10$ s, i.e., $N_{SR} = 50$ messages. V2V transmit powers are assumed to be -5 dBm, which enables multiple collision domains across the highway. In Scenario B, the highway is 2000 m with two BSs deployed at positions 750 m from its center with a 45 m longitudinal offset. Vehicles are assumed to send more frequent scheduling requests, at every $T_{SR} = 1$ s ($N_{SR} = 5$). The middle 1000 m section of the highway is assumed to be a tunnel-zone without any BS coverage, where vehicles are not able to send/receive any SR/SA. V2V transmit powers are set to 23 dBm, which is the maximum value allowed by 3GPP [21].

TABLE I
SIMULATION PARAMETERS

Mobility	1000 m (Scenario A) and 2000 m (Scenario B) highway with 2 lanes/direction of 4 m width; vehicle lengths of 5 m; Poisson arrival per direction with $\sim$Exp(0.4) [18] with speeds $\sim\mathcal{N}(75, 45)$ km/h (Scenario A) and $\sim\mathcal{N}(120, 36)$ km/h (Scenario B).
V2V message traffic	$T_m = 200$ ms; $B_m = 190$ Bytes.
V2V resources	Carrier frequency = 5.9 GHz; Bandwidth = 10 MHz (50 RBs) with 32 RBs active; MCS index = 9; 1 subchannel = 16 RBs, 1 slot = 1 ms; Resource pool of 2 subchannels by 10 slots, periodically repeating with 80 ms.
V2V channel model	3GPP Channel Model [18] with Path loss: LOS model in WINNER+B1; path loss at 3 m is used for distances < 3 m; Shadowing fading: log-normal distr. with 3 dB std. dev. and 25 m decorr. distance. V2V Tx power = -5 dBm (Scenario A) and 23 dBm (Scenario B); Thermal noise level = -174 dBm/Hz; Antennae: 1 Tx and 2 Rx omni-directional with 1.5 m height, 3 dBi gain, and 9 dB noise figure.
UL/DL channel model	COST-Hata Channel Model [22] at 2.1 GHz with 10 dB shadowing fading, 10 MHz (50 RBs) bandwidth; BS DL Tx power = 30 dBm, vehicle UL Tx power = 23 dBm BS antenna: isotropic with 30 m height and 5 dB noise figure.



(a) Scenario A.



(b) Scenario B.

Fig. 3. Evaluation scenarios.

Mobility of vehicles is simulated using the realistic traffic simulator SUMO [23]. Vehicles travel according to the utilized car-following [24] and lane-changing [25] models, which result in a realistic, stochastic driving behavior. In both scenarios, the highway has 2 lanes per direction, where vehicles arrive at the both ends with probability 0.4/s (leading to an average inter-vehicle gap of 2.5 s [18]), with speeds distributed normally $\sim\mathcal{N}(75, 45)$ km/h and $\sim\mathcal{N}(120, 36)$ km/h in scenarios A and B, respectively. Slower speeds with higher deviation in Scenario A result in a denser and more variant road traffic than in Scenario B. Yet, in both scenarios, the realistic mobility models lead to a highly dynamic traffic with density varying over time and space.

Communications is simulated using a full-stack LTE V2V system-level simulator in ns-3 [26], which is developed by ourselves by extending the openly available LTE [27] and D2D [28] modules. Vehicles generate V2V messages with $T_m = 200$ ms and $B_m = 190$ Bytes, that are typical of CAM transmissions [29]. Configured V2V resource pool is assumed to consist of 2 subchannels and 10 time slots, each able to carry a single V2V message combined with the control information and protocol overhead. V2V and UL/DL channels are simulated using realistic path loss and fading with parameters indicated in Table I, where further simulation details are also provided.

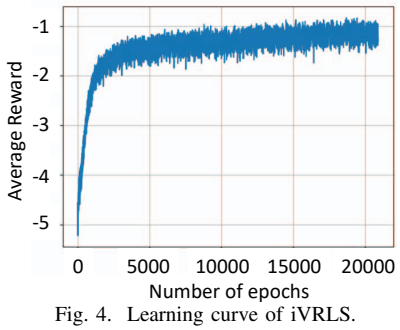


Fig. 4. Learning curve of iVRLS.

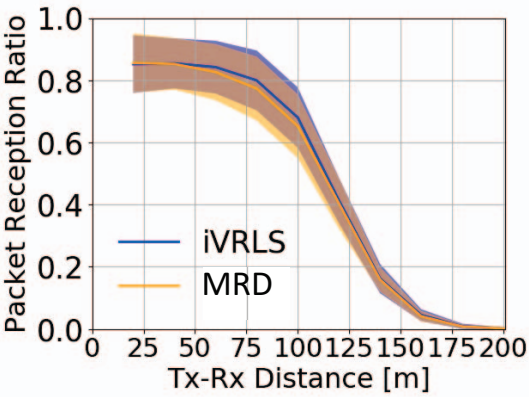
B. Training of iVRLS

We train iVRLS before its deployment, on a setting with simpler mobility and communication model as compared to evaluation scenarios, which enables faster training. In the training environment, 50 vehicles are initially placed uniformly at random on a 1000-m-long two-way highway without any tunnel, and travel with constant speeds randomly selected from normal distribution $\sim\mathcal{N}(75, 25)$ km/h. Upon their exit, they return back to the highway from the opposite direction after a time offset $\sim$Exp(0.4). V2V communications in the *training environment* is abstracted with the protocol model [30] (note that in the evaluations, the realistic propagation model is used), using fixed V2V transmission ranges of 120 m. In this model, unsuccessful receptions are assumed to result from any transmission whose range intersect with another one using the same resource, at the receiver, besides the errors due to HD radio. Vehicle-to-BS links in the training are assumed to be error-free. Considering the V2V transmission range, reward R_{t+1} is calculated using PRR measured at $0 - 100$ m Tx-Rx distance. $\Delta x_{min} = 3$ m is utilized, in line with the V2V path loss model used in our evaluations. We provide the training curve of iVRLS in Fig. 4. It shows the average reward collected by iVRLS over the training epochs (each epoch is a sequence of 60 state-action-reward tuples). iVRLS converged to a stable performance level at around 20000 epochs in the training environment.

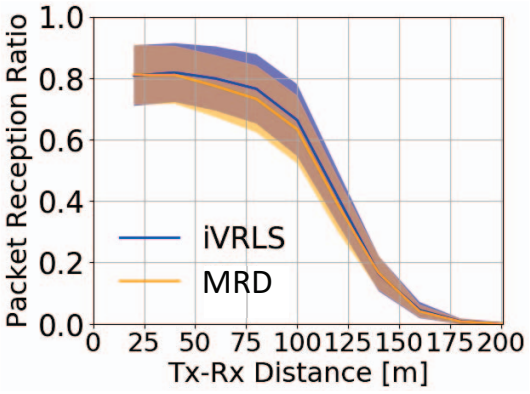
C. iVRLS Reliability

We evaluate the reliability performance of iVRLS with respect to PRR of V2V messages as a function of the transmitter-receiver (Tx-Rx) distance among the vehicles in the evaluation scenarios A and B described in Section V-A. In the results, we provide the mean (solid curve) and the standard deviation (shade) of PRR measured every 10 s, during a total simulation time of 1000 s.

Fig. 5 shows the results of iVRLS in comparison to MRD in Scenario A. In Fig. 5(a), we report the PRR from the central part of the road, measured at locations $[-250, +250]$ m, while Fig. 5(b) presents the PRR at the remaining edges of the road

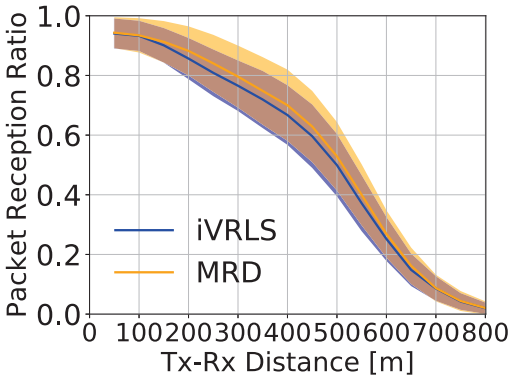


(a) Central section.

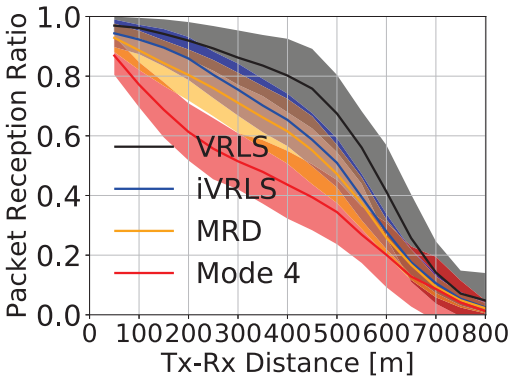


(b) Edge sections.

Fig. 5. PRR in Scenario A, measured as a function of Tx-Rx distance, at central and edge sections of the road.



(a) In-coverage.



(b) Tunnel.

Fig. 6. PRR in Scenario B, measured as a function of Tx-Rx distance, at in-coverage and tunnel sections of the road.

(measured at locations $[-500, -250]$ and $[+250, +500]$ m). It could be observed from Fig. 5 that, at all sections of the road, iVRLS performs marginally better than MRD, with a gain more pronounced at larger Tx-Rx distances. In case of edge sections, vehicles suffer from larger path loss to the BS, resulting in increased loss of SR/SAs, which degrades the performance of both algorithms. iVRLS shows a larger gain over MRD at the road edges, as compared to the central section of the road.

In Fig. 6, we provide the PRR results from Scenario B. Fig. 6(a) reports the results from the two ends of the road under (partial) coverage, measured at locations $[-1000, -250]$ and $[+250, +1000]$ m, while Fig. 6(b) presents the PRR from the very central part of the tunnel, measured at $[-250, +250]$ m. From Fig. 6(a), it is observable that MRD provides marginally larger reliability of V2V transmissions under coverage. At close Tx-Rx ranges, both iVRLS and MRD yield around 95% PRR in average, with achievable values up to 100%. Compared to Scenario A, less dense vehicles in the environment create less load on the resources, and scheduling is done more frequently, which results in overall higher reliability of V2V transmissions. Also, higher transmit power increases the rate of successful transmissions at larger Tx-Rx distances, up to 800 m. Such close-to-ideal conditions benefit the MRD algorithm, for which it is designed. On the other hand, V2V transmissions taking place within the tunnel, as observed from Fig. 6(b), suffer from degraded reliability as compared to the transmissions within coverage. Under such conditions, iVRLS is able to yield higher performance than MRD at almost all Tx-Rx distances. We here also note that the policy learned by iVRLS is well applicable to different V2V transmission ranges resulting from different V2V transmit powers in the evaluation environments.

In Fig. 6(b), we further evaluate the performance of distributed sensing-based semi-persistent scheduler Mode 4 from the 3GPP standard [19], as well as our centralized solution VRLS [4] designed for out-of-coverage communications, for reference purposes (cf. [4] for their configurations). Without our proposed method that extends the use of the scheduled resources under intermittent coverage conditions, one has to combine in-coverage schedulers with such schedulers for out-of-coverage V2V transmissions. Fig. 6(b) shows that, with our proposed method, centralized schedulers perform better than the distributed scheduler Mode 4, while iVRLS provides better performance than MRD. On the other hand, VRLS, designed solely for out-of-coverage coverage operation, outperforms both of the extended schedulers, as well as Mode 4. Accordingly, one could consider combining in-coverage solutions with VRLS for even better OOC performance, yet at the expense of increased cost and complexity.

D. Challenges for Deploying iVRLS

The performance gain of iVRLS comes with the cost of training and higher operational complexity, as compared to the heuristic-based approaches, such as MRD. As any other RL solution, iVRLS requires a training, which could be conducted off-line, i.e., before its deployment. On the other hand, during its operation, iVRLS needs to compute the state representation and process it with its policy DNN, which incurs higher processing cost than the operations required by heuristic algorithms. Despite these challenges, iVRLS would be preferable for deployment to serve areas with rather non-ideal conditions such as intermittent network coverage given its advantageous performance, where area-specific training and operation could be performed.

VI. Conclusions

We proposed iVRLS, an RL-based centralized scheduler for in-coverage V2V communications, specifically targeting imperfect network coverage conditions. To support such conditions, state-of-the-art centralized schedulers could be extended by a simple method we propose, where vehicles continue keeping their resources for their V2V transmissions during the periods of coverage loss. Nevertheless, iVRLS performs better than the enhanced version of a state-of-the-art heuristic algorithm under intermittent coverage conditions. In particular, iVRLS performs better in case of high traffic load and less frequent scheduling of V2V messages, as well as within road tunnels without any network coverage. By delivering higher V2V communication reliability as the network conditions move away from the ideal, iVRLS offers a robust alternative to existing schedulers across varying conditions of coverage.

References

[1] V. Vukadinovic, K. Bakowski, P. Marsch, I. D. Garcia, H. Xu, M. Sybis, P. Sroka, K. Wesolowski, D. Lister, and I. Thibault, "3GPP C-V2X and IEEE 802.11 p for vehicle-to-vehicle communications in highway platooning scenarios," *Ad Hoc Networks*, vol. 74, pp. 17–29, 2018.

[2] G. Cecchini, A. Bazzi, B. M. Masini, and A. Zanella, "Performance comparison between IEEE 802.11p and LTE-V2V in-coverage and out-of-coverage for cooperative awareness," in *2017 IEEE Vehicular Networking Conference (VNC)*, 2017, pp. 109–114.

[3] R. Molina-Masegosa, J. Gozalvez, and M. Sepulcre, "Comparison of IEEE 802.11p and LTE-V2X: An evaluation with periodic and aperiodic messages of constant and variable size," *IEEE Access*, vol. 8, pp. 121 526–121 548, 2020.

[4] T. Sahin, R. Khalili, M. Boban, and A. Wolisz, "VRLS: A unified reinforcement learning scheduler for vehicle-to-vehicle communications," in *2019 IEEE 2nd Connected and Automated Vehicles Symposium (CAVS)*. IEEE, 2019, pp. 1–7.

[5] G. Cecchini, A. Bazzi, M. Menarini, B. M. Masini, and A. Zanella, "Maximum reuse distance scheduling for cellular-v2x sidelink mode 3," in *2018 IEEE Globecom Workshops (GC Wkshps)*, 2018, pp. 1–6.

[6] A. Bazzi, A. Zanella, G. Cecchini, and B. M. Masini, "Analytical investigation of two benchmark resource allocation algorithms for lte-v2v," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 6, pp. 5904–5916, 2019.

[7] L. Hu, J. Eichinger, M. Dillinger, M. Botsov, and D. Gozalvez, "Unified device-to-device communications for low-latency and high reliable vehicle-to-x services," in *2016 IEEE 83rd Vehicular Technology Conference (VTC Spring)*, 2016, pp. 1–7.

[8] S. Zhang, Y. Hou, X. Xu, and X. Tao, "Resource allocation in d2d-based v2v communication for maximizing the number of concurrent transmissions," in *2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2016, pp. 1–6.

[9] L. F. Abanto-Leon, A. Koppelaar, and S. H. de Groot, "Parallel and successive resource allocation for v2v communications in overlapping clusters," in *2017 IEEE Vehicular Networking Conference (VNC)*, 2017, pp. 223–230.

[10] ——, "Subchannel allocation for vehicle-to-vehicle broadcast communications in mode-3," in *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, 2018, pp. 1–6.

[11] L. F. Abanto-Leon, A. Koppelaar, and S. Heemstra de Groot, "Network-assisted resource allocation with quality and conflict constraints for v2v communications," in *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*, 2018, pp. 1–5.

[12] M. I. Ashraf, M. Bennis, C. Perfecto, and W. Saad, "Dynamic proximity-aware resource allocation in vehicle-to-vehicle (v2v) communications," in *2016 IEEE Globecom Workshops (GC Wkshps)*, 2016, pp. 1–6.

[13] L. Liang, G. Y. Li, and W. Xu, "Resource allocation for d2d-enabled vehicular communications," *IEEE Transactions on Communications*, vol. 65, no. 7, pp. 3186–3197, 2017.

[14] W. Sun, D. Yuan, E. G. Ström, and F. Brännström, "Cluster-based radio resource management for d2d-supported safety-critical v2x communications," *IEEE Transactions on Wireless Communications*, vol. 15, no. 4, pp. 2756–2769, 2016.

[15] H. Ye, G. Y. Li, and B. F. Juang, "Deep reinforcement learning based resource allocation for v2v communications," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 4, pp. 3163–3173, 2019.

[16] H. Yang, X. Xie, and M. Kadoch, "Intelligent resource management based on reinforcement learning for ultra-reliable and low-latency iov communication networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 5, pp. 4157–4169, 2019.

[17] J. A. del Peral-Rosado, R. Raulefs, J. A. López-Salcedo, and G. Seco-Granados, "Survey of cellular mobile radio localization methods: From 1g to 5g," *IEEE Communications Surveys Tutorials*, vol. 20, no. 2, pp. 1124–1148, 2018.

[18] 3GPP TR 36.885 V14.0.0, *Study on LTE-based V2X services (Release 14)*, 3GPP Std., June 2016.

[19] 3GPP TR 37.985 V16.0.0, *Overall description of radio access network (RAN) aspects for vehicle-to-everything (V2X) based on LTE and NR (Release 16)*, 3GPP Std., June 2019.

[20] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[21] 3GPP TS 36.101 V16.4.0, *User Equipment (UE) radio transmission and reception (Release 16)*, 3GPP Std., Dec. 2019.

[22] C. Action, "231,"digital mobile radio towards future generation systems," final report," tech. rep., European Communities, EUR 18957, Tech. Rep., 1999.

[23] P. A. Lopez, M. Behrisch, L. Bieker-Walz, J. Erdmann, Y.-P. Flötteröd, R. Hilbrich, L. Lücken, J. Rummel, P. Wagner, and E. WieBner, "Microscopic traffic simulation using SUMO," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018, pp. 2575–2582.

[24] S. Krauß, "Microscopic modeling of traffic flow: Investigation of collision free vehicle dynamics," Ph.D. dissertation, 1998.

[25] J. Erdmann, "Lane-changing model in sumo," *Proceedings of the SUMO2014 modeling mobility with open data*, vol. 24, pp. 77–88, 2014.

[26] "The ns-3 network simulator." [Online]. Available: www.nsnam.org

[27] G. Piro, N. Baldo, and M. Miozzo, "An LTE module for the ns-3 network simulator," in *Proceedings of the 4th International ICST Conference on Simulation Tools and Techniques*. ICST, 2011, pp. 415–422.

[28] R. Rouil, F. J. Cintrón, A. Ben Mosbah, and S. Gamboa, "Implementation and Validation of an LTE D2D Model for ns-3," in *Proceedings of the Workshop on ns-3*. ACM, 2017, pp. 55–62.

[29] ETSI TC ITS, *Intelligent Transport Systems; Vehicular Communications; Basic Set of Applications; Part 2: Specification of Cooperative Awareness Basic Service*, Std. ETSI EN Std 302 637-2 V.1.3.1, 2014.

[30] P. Gupta and P. Kumar, "The Capacity of Wireless Networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388–404, Mar 2000.