Learning to schedule joint radar-communication requests for optimal information freshness

Lee, Joash; Niyato, Dusit; Guan, Yong Liang; Kim, Dong In

2021

Lee, J., Niyato, D., Guan, Y. L. & Kim, D. I. (2021). Learning to schedule joint radar-communication requests for optimal information freshness. 2021 IEEE Intelligent Vehicles Symposium (IV), 8-15. https://dx.doi.org/10.1109/IV48863.2021.9575131

https://hdl.handle.net/10356/150718

https://doi.org/10.1109/IV48863.2021.9575131

Learning to Schedule Joint Radar-Communication Requests for Optimal Information Freshness

Joash Lee[1], Dusit Niyato[2], Yong Liang Guan[3], Dong In Kim[4]

***Abstract*— Radar detection and communication are two of several sub-tasks essential for the operation of next-generation autonomous vehicles (AVs). The former is required for sensing and perception, more frequently so under various unfavorable environmental conditions such as heavy precipitation; the latter is needed to transmit time-critical data. Forthcoming proliferation of faster 5G networks utilizing mmWave is likely to lead to interference with automotive radar sensors, which has led to a body of research on the development of Joint Radar Communication (JRC) systems and solutions. This paper considers the problem of time-sharing for JRC, with the additional simultaneous objective of minimizing the average age of information (AoI) transmitted by a JRC-equipped AV. We formulate the problem as a Markov Decision Process (MDP) where the JRC agent determines in a real-time manner when radar detection is necessary, and how to manage a multi-class data queue where each class represents different urgency levels of data packets. Simulations are run with a range of environmental parameters to mimic variations in real-world operation. The results show that deep reinforcement learning allows the agent to obtain good results with minimal a priori knowledge about the environment.**

I. Introduction

Self-driving vehicles of the future will have to perform a multitude of tasks to facilitate its end-goal of safe navigation, including sensing and perception, localization, mapping of obstacles and path finding. In a cooperative driving setting, large amounts of data will have to be communicated between vehicles (V2V) and between vehicles and infrastructure (V2X). With the commercialization of 5G communication technology and the associated transmission of data with higher frequencies, such as the mmWave band, interference with radar waves has become a possibility.

(a) (b)

Fig. 1: (a) Heavy precipitation causing low visibility for camera-based systems. (b) A busy junction.

The problem of jointly operating radar and communication functions within the same frequency range has been variously referred to as joint radar-communication (JRC) or communication and radar spectrum sharing (CRSS). Methods of realizing JRC systems have been variously reviewed by [1], [2]. These methods may be broadly categorized into three different approaches: using communications signals to perform object detection, modulating communication signals onto radar pulses, and time-division between radar operation and communication.

More traditional approaches to time-division use a fixed schedule to alternate between radar and communication [2], [3]. In contrast, a learning-based approach was proposed by [4] for use in automotive vehicles. The main advantage of such an approach is that the communication system is able to learn to respond to instantaneous changes in the environmental conditions that may require increased frequency of radar operation to support safe navigation.

Another challenge associated with the communication of data in self-driving vehicles is the high data rate and freshness of sensory information necessary for safe navigation, in combination with demand for low-latency Internet connection for in-vehicle entertainment systems. Self-driving vehicles are often equipped with a suite of sensors to reliably perceive its environment: cameras, LIDAR, radar, ultrasonic sensors, inertial measurement units and satellite-based positioning systems such as GPS. The exteroceptive sensors generate large amounts of data due to their multi-dimensional nature [5]. This may require transmission in a timely manner to surrounding vehicles or infrastructure in cooperative driving settings.

Traditional metrics on communication of data, such as throughput or delay, may be useful for in-vehicle enter-

[1]Joash Lee is with the Energy Research Institute @ NTU, Nanyang Technological University, Singapore. l.joash@ntu.edu.sg

[2]Dusit Niyato is with the School of Computer Science and Engineering, Nanyang Technological University, Singapore dniyato@ntu.edu.sg

[3]Yong Liang Guan is with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore eylguan@ntu.edu.sg

[4]Dong In Kim is with the School of Information and Communication Engineering, Sungkyunkwan University, South Korea. dikim@skku.ac.kr

This research is supported, in part, by Alibaba Innovative Research (AIR) Program, Alibaba-NTU Singapore Joint Research Institute, National Research Foundation, Singapore, under AI Singapore Programme (AISG-GC-2019-003), WASP/NTU grant M4082187 (4080), Singapore Ministry of Education Tier 1 (RG16/20), A*STAR under its RIE2020 Advanced Manufacturing and Engineering Industry Alignment Fund — Pre Positioning (A19D6a0053), and Ministry of Science and ICT, Korea, under the ICT Creative Consilience program (IITP-2020-0-01821) supervised by the IITP. Any opinions, findings, conclusions or recommendations in this material are those of the authors and do not reflect the views of the mentioned organizations.

tainment systems. However, they may be inadequate for evaluating the communication of sensory data in cooperative driving scenarios where timely responses to an ever-changing environment are required. The authors in [6] raised the example of how a data packet containing old information may be of little utility even if delivered with little delay. A more appropriate metric for the timeliness of communication in an automotive setting is the age-of-information (AoI), for which [7] was an early proponent. In the proposition by [7], "age" is the duration of time since the last received packet was generated.

In this paper, we consider a combined approach to jointly minimize AoI and maintain timely radar operation through time division for a single vehicle agent. Key considerations that motivate our approach to the problem formulation and our method of solving it are as listed below:

1) *Environment awareness* — The agent must be able to adjust its use of radar detection in response to situations where radar becomes more important.
2) *Urgency dependent* — Several classes of data may be transmitted by an autonomous vehicle. Our JRC system should be capable of prioritizing the communication of data of higher importance.
3) *Low prior knowledge* — The JRC system should require minimal prior knowledge on the vehicle's environment and the nature of the data it is to communicate.

The work presented in this paper is a significant improvement of the time-division based JRC method first proposed in [4]. We consider a single agent representing a JRC system on a self-driving vehicle. Our contributions are as follows:

1) The agent manages a multi-class data queue of finite length, where each class represents different urgency levels of data packets.
2) In addition to jointly coordinating radar and communication, the agent minimizes the AoI of its communication function.

We formulate our JRC-AoI problem as a Markov Decision Process (MDP), and investigate the use of Deep Q Networks (DQN) to jointly optimize the objectives of minimizing average AoI and managing JRC. Section II presents related work on automotive sensing and communication and their relation to traffic accident risk. Section III introduces our problem formulation, while Section IV provides a background on DQN. In Section V, we describe different experimental suites that mimic the variability in the environment that may be encountered by an automotive JRC systems. We conduct the experiments and discusses the results obtained. Code required to reproduce these experiments is available [1].

II. Background and Related Work

In this section, we review contemporary automotive sensing techniques and identify challenging scenarios for automotive perception. For each of these perceptually adverse scenarios, we establish a connection with prevailing literature on road safety risk. We also review related work on automotive communication and studies on AoI.

A. Automotive Sensing

In automotive vehicles, data from different sensors and sensor types is typically combined. This process, known as sensor fusion, results in more certainty than if sensor readings are used individually [8], [9]. Certainty in environmental perception can be gained through combining sensor types that have complementary strengths. For example, while there has been much success at object detection in the daytime with camera systems, the distance of the object of interest from the vehicle is measured more accurately using LIDAR or radar [10].

Another method of increasing the certainty of environmental perception is to increase the sampling rate of the available sensors. Methods to adjust the sampling rate of individual sensor nodes within a sensor network have been proposed by [11], [12] and reviewed in [13]. The key idea is that the sampling rate of a sensor node is increased when the sensor is observing an interesting event, and reduced otherwise. These adaptive sampling methods are motivated by constraints on energy use or overall system bandwidth, such as remotely-installed battery-powered monitoring [12] or wearable devices [11]. We note a similarity in JRC, in that there is a constraint on bandwidth availability. For JRC deployed in an automotive setting, events that are of interest to radar sensing are situations with higher risk, and situations where radar performs favorably in comparison to other available sensors.

B. Perception of Risky Conditions

Higher risk driving events include unfavorable road surface conditions, adverse weather, proximity of other vehicles, and excess vehicle speed. These factors have been shown by road safety studies [14]–[20] to be associated with higher risk of traffic accidents. For each of these environmental features that contribute to accident risk, we review how each feature can be measured and how they affect the risk level. We review literature in Section II-C on how this risk may be quantified.

Weather: Unfavorable weather conditions in an automotive setting include rain, snow or fog [17]. Classification of the weather condition can be conducted based on input from cameras [21] or LIDAR systems [22]. Our paper is concerned with quantitative evaluations of the intensity of weather conditions, which is a matter that was considered in [22]. A study on aggregated data from various regions of Europe showed that every additional 100 mm of rainfall led to an increase in the number of injury-causing traffic accidents by 0.2-0.3% [14]. In adverse weather and low light conditions, radar is known to be the most robust automotive sensing modality [10]. This is in contrast to camera systems, which experience a degradation in image quality such conditions; Figure 1a illustrates low visibility experienced by a camera under heavy precipitation.

[1] https://github.com/joleeson/JRC-AoI.git

Road surface: Unfavorable road surface conditions most associated with accidents were found by [16] to include irregular topographical characteristics. This may be quantified using standardized metrics such as the international roughness index (IRI). Measurement methods on non-specialized vehicles may utilize a combination of accelerometers and measurements from vehicle suspension components [23], [24]. These measurement modalities may be supplemented or substituted by a digital map database that contains information on road surface conditions [23]. In a road safety study [16], an increase in rut depth of 2.5 mm was found to increase the frequency of night-time accidents in Tennessee by 1.509 times. In higher risk situations that could cause loss of traction, increased radar operation would create a more accurate map of the vehicle's environment, which could become necessary for short-timescale corrective maneuvers.

Presence of moving objects: Successful driving is dependent on the sub-tasks of moving objective detection and identification, especially in urban settings. While identification of moving objects is typically best performed by camera systems, radar systems perform the best at detecting speed and distance of surrounding objects [10]. The volume of traffic flow, which is indicative of the number of vehicles in close proximity and interacting with one another, has been shown to have a positive relationship with the frequency of accidents on both motorways [18] and intersections [19]. On three-lane motorway segments in France, the average number of crashed vehicles doubled when the traffic flow rate increased from 1500 to 4000 vehicles per hour [18]. Figure 1b shows an example of a busy intersection with pedestrians and vehicle navigating in different directions.

Speed of ego vehicle: Many studies indicate that higher speeds or higher differences in speed relative to surrounding traffic are associated with increased frequency of road traffic accidents [15], [20]. A variation in speed by 1% was found to increase the frequency of accidents by 0.3% in [20]. A vehicle's speed can be measured by odometry sensors and corroborated with global navigation satellite systems. Increased frequency of radar operation would be useful in high speed scenarios because of its strengths in object speed detection.

As mentioned, the higher-risk conditions discussed in this paper coincide with situations where radar sensing is known to perform more favorably compared to alternative sensors such as LIDAR and camera systems [10]. However, there are few studies which consider the effect of such high-risk events in formulating strategies for automotive sensing. The typical approach is to sample data from each sensor at a constant pre-determined rate [5]. Solutions have been proposed in the form of sensor fusion methodologies [9], [25], although they do not assume any costs or constraints on sensor availability.

C. Characterization of Accident Risk

Studies on road safety commonly quantify the risk of a road accident in terms of the number of accidents over a given period of time [16]. Previous studies have used statistical regression techniques to evaluate how much each environmental condition contributes to the overall risk of traffic accidents for different sets of data from different geographical regions. Negative binomial regression is commonly used for such modeling [14], [16], [19], [20]. This is a generalization of Poisson regression which relaxes the assumption that the variance must be equal to the mean. The predicted mean number of traffic accidents μ within a given road segment for a particular time interval is given by:

$$\mu = \exp\left(\mathbf{e}\boldsymbol{\beta}\right) \tag{1}$$

where $\mathbf{e}$ is the vector of environmental features as introduced above, and $\boldsymbol{\beta}$ is the coefficient vector to be estimated. We adopt a similar mathematical relationship in our problem formulation in Section III-A. The learned relationship between environmental condition and risk level could be provided to individual vehicles through access to a digital map database.

D. Communication in Automotive Settings

Dissemination of information in V2V or V2X networks is typically handled by the IEEE Wireless Access in Vehicular Environments (WAVE) family [26]. However, existing standards and the prevailing body of literature on vehicular communications do not account for AoI. We use a definition of AoI consistent with earlier works [6], [27]:

Definition 1: The *age of information* $A^{<m>}$ for data class m at a receiver is the length of time from the generation to receipt of the most recently generated data packet of class m.

The initial study on AoI by [7] considers an M/M/1 model where data packets are generated at a constant rate and enter into a First-in-First-Out (FIFO) queue at the medium access control (MAC) layer. AoI was minimized by adapting the rate of packet generation at the source nodes. AoI has also been investigated for a system with a multi-class queue in [27]. The effect of interference and channel quality was not considered in these earlier works [6], [27]. We note that a limitation of the optimization-based approaches utilized by the above-mentioned studies is that they minimize AoI for the average system state by adjusting average rate parameters, and fail to account for instantaneous differences in the system. In contrast, we model a vehicular communication problem as a Markov Decision Process (MDP). A scheduling algorithm is then solved by using deep reinforcement learning.

III. Problem Formulation

In our study, the concept of AoI is extended by taking into account the urgency of each data class as a linear weight applied to its age. Given our focus on the transmitting agent, we consider "age" as the delay from packet generation to the instant of its transmission.

We formulate an online JRC scheduling problem as a MDP where the agent of interest is the JRC system of a single vehicle. A schematic of how the JRC scheduler interacts with the vehicle's sensory and perception systems is shown in Figure 2. At each discrete time step, the agent observes the state of its data queue along with information on the

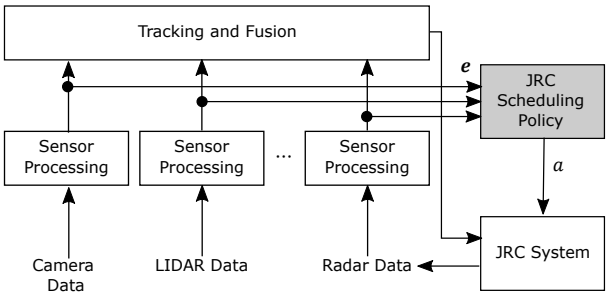

Fig. 2: A schematic showing how a Joint-Radar Communication system and its scheduling policy would interact with the sensor fusion system in an automotive application. The scheduling policy receives environmental features $\mathbf{e}$ and decides on scheduling actions a.

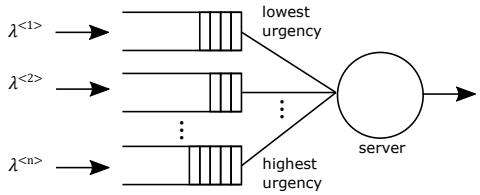

Fig. 3: A schematic of the multi-class data packet queue. Data packets with class n arrive into the queue.

surrounding environmental conditions. The environmental features considered in this paper are introduced in Section II, and have been shown in road safety studies [14]–[17] to be associated with the risk of road traffic accidents. From an information theoretic standpoint, a higher sampling rate of sensory input would be required to provide more certainty about safety-critical events [13]. We propose a reward function to encourage such behavior. Details on the state space, state transition, action space and reward function are further described in this section.

A. State Space and System Model

The state consists of features $\mathbf{e}$ representing the agent's environmental conditions, the condition of the communication channel c, the age $\boldsymbol{\alpha}$ and class $\boldsymbol{u}$ of the data packets that are queued, and the AoI for all classes of data $\mathbf{A}$:

$$\mathbf{s} = [\mathbf{e}, c, \boldsymbol{\alpha}, \mathbf{u}, \mathbf{A}]. \tag{2}$$

The environmental condition vector contains four environmental features, such that $\mathbf{e} = [\rho, w, m, v]$. These features represent the condition of the road, weather, the presence of moving objects nearby, and the speed of the ego vehicle respectively; they are identified in Section II to be associated with the risk level. Each feature is ranked on a discrete scale such that $\{\rho, w, m, v\} \in \{0, 1, \ldots, E\}$; a value of zero indicates the safest possible condition, while higher values indicate decreasing favorability in terms of safety. For example, $\rho = 0$ would indicate a dry well-maintained road, while higher values would indicate decreasing frictional coefficients. Higher values of w, m and v would indicate inclement weather and poor visibility, the presence of a moving object in the vicinity of the ego vehicle, and high speed of the ego vehicle respectively. For ease of presentation, we consider $E = 1$.

As discussed in Section II-C, the relationship between environmental conditions and the risk level is typically modeled using negative binomial regression. We interpret this quantity as related to the number of high-risk events for which radar operation is necessary. By making the assumption that the time step is sufficiently small, the arrival of a high-risk event X in each time step for an individual vehicle can be modelled by a Bernoulli distribution with probability parameter $k \exp(\mathbf{e}\boldsymbol{\beta})$, where k is a constant [28].

The channel state is an indicator of quality of the communication channel, where $c = 0$ indicates a good channel, and higher values indicate lower transmission rates.

Additional state features that we consider in this study relate to descriptors of data packets that arrive into the agent's queue, which has a maximum length of L. For each data packet l in the queue, its age is represented by α_l, such that the age vector is $\boldsymbol{\alpha} \in \mathbb{R}^L$. Note that this concept is related but different from the AoI of each data class, which we discuss further later in this section.

Definition 2: The age of each data packet is measured by the number of time steps that it remains in the queue before it is successfully transmitted by the agent.

The corresponding urgency class of each data packet is $u_l \in \{1, 2, \ldots, M\}$, such that $\mathbf{u} \in \mathbb{R}^L$. A value of $u_l = 0$ indicates that position l in the data queue is empty. If no data packets enter or leave the queue at time step t, the age transition function can be described as:

$$\alpha_l(t+1) = \alpha_l(t) + \mathbf{1}_{\mathbb{R}^+}(u_l(t)), \tag{3}$$

where $\mathbf{1}_{\mathbb{R}^+}$ is an indicator function for positive real numbers.

A schematic of the data packet queuing system is shown in Figure 3. New data packets enter the queue according to a Poisson distribution with means $(\lambda^{<1>}, \lambda^{<2>}, \ldots, \lambda^{<M>})$, where $\lambda^{<m>}$ is the parameter for the arrival of tasks with urgency level m. Let the random variables for the number of newly generated data packets be represented by $Y^{<m>}$. Each newly generated data packet is inserted into the first non-zero index of the urgency and action state features $\boldsymbol{\alpha}$ and $\mathbf{u}$. Data packets that exceed a threshold age of α_{max} are considered to be expired, and removed from the age and urgency state vectors.

The vector $\mathbf{A} \in \mathbb{R}^M$ maintains the AoI for each urgency class. Each element $A^{<m>}$ represents the number of time steps since the generation of the last data packet of urgency class m that was received by its intended receiver. The evolution of $A^{<m>}$ can be described as:

$$A^{<m>}(t+1) = \begin{cases} \min\limits_{l} \left(\alpha_l(t) \right. \\ \qquad \left. \mathbf{1}_{\{m\}}(u_l(t))\right) + 1 & \text{if } a(t) = a^{<m>} \\ A^{<m>}(t) + 1 & \text{otherwise,} \end{cases} \tag{4}$$

where $A^{<m>}(t+1)$ is the AoI of data class m at time step $(t+1)$, $\mathbf{1}_{\{m\}}(\cdot)$ is the indicator function for numbers equal to m, a_t is the action taken by the agent at time step t. The evolution of AoI with time for a particular class is shown in Figure 4. This model adopts the assumption that the time

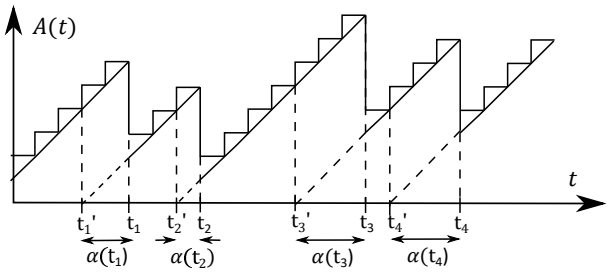

Fig. 4: The evolution of the Age of Information A of a specific data class with time step t. The time t_1' indicates when the first packet arrives into the queue. At time step t_1, a communication action is chosen, thus transmitting the packet after it has spent $\alpha(t_1)$ time steps in the queue.

taken from the moment of transmission to receipt of the data is constant and equal to zero. The actions available to the agent are described in the next sub-section.

B. Action

At each time step, the agent may choose to operate in radar mode a^r, or execute tasks with a chosen urgency level. The action set is thus: $\mathcal{A} = \{a^r, a^{<1>}, a^{<2>}, \ldots, a^{<M>}\}$, where $a^{<u=m>}$ indicates the choice of communicating data with urgency class m. A better channel condition c allows the agent to successfully transmit more data packets from the chosen class. The transmitted data packets are removed from the queue in the next time step. We define $T \in \mathbb{R}^L$ to be a vector that consists of binary indicators of whether each data packet l in the queue is transmitted during a given time step t. For simplicity, we consider a deterministic service time that is equal to the duration of one time step.

C. Reward

The reward received by the agent at each time step t is defined as a weighted sum that encourages the agent to minimize the age of the queued data packets while also carry out radar detection when necessary:

$$r(t) = w_{age} r_{age}(t) + w_{overflow} r_{overflow}(t) + w_{rad} r_{rad}(t), \tag{5}$$

where $w_{age}, w_{overflow}$ and w_{rad} are weights. r_{age} encourages the agent to minimize AoI, $r_{overflow}$ encourages the agent to transmit data packets before the queue overflows with new arrivals, while r_{radar} encourages the agent to perform a radar scan when there are unfavorable environmental conditions.

We formulate r_{age} to encourage minimization of the sum of urgency weighted ages for each data class.

$$r_{age}(t) = -\sum_{m=1}^{M} \big(m \times A^{<m>}(t+1)\big), \tag{6}$$

where m is the urgency class, and $A_{t+1}^{<m>}$ is the AoI of data class m at the next time step.

The penalty for queue overflows is proportional to the number of newly-generated data packets that the queue cannot accommodate, after accounting for data transmission T and data expiration at the current time step:

$$\begin{aligned} r_{overflow}(t) = \min \Big(0, -\Big(\sum_{l=1}^{L} [\mathbf{1}_{\mathbb{R}^+}(u_l(t)) \\ - \mathbf{1}_{\mathbb{R}^+}(\alpha_l(t) - \alpha_{max}) - T_l(t)] \\ + \sum_{m=1}^{M} Y^{<m>}(t) - N \Big)\Big). \end{aligned} \tag{7}$$

In our simulated environment, the activation of radar detection mode is deemed necessary whenever an unexpected high-risk event occurs. The term r_{radar} is proportional to the number of unfavorable environmental features in $\mathbf{e}$ whenever a high-risk event occurs (i.e. $X_t = 1$):

$$r_{rad}(t) = -\big(\rho(t) + w(t) + m(t) + v(t)\big) \times X(t). \tag{8}$$

IV. Deep Q-Learning

This section provides a brief background on the DQN algorithm that we use to solve our JRC-AoI problem. DQN is a reinforcement learning algorithm that combines Q-learning algorithm with the use of deep neural networks. This method was chosen for its abilities in handling MDPs with high complexity or high-dimensional state spaces, such as the Atari game environments for which the algorithm received notable recognition [29], and learning optimal actions with minimal a priori knowledge of system parameters.

At the beginning of each time step t, the state s_t as observed by the agent is input into the Q network, parameterized by θ_t at time step t, which then predicts the optimal values Q^* of taking each possible action a_t. Based on these predictions, the agent chooses an action based on the ϵ-greedy policy, which leads to receipt of reward r_t. The agent's experience (s_t, a_t, r_t, s_{t+1}) is stored in a dataset D_t.

At regular intervals, the agent performs Q-learning through experience replay by sampling a minibatch (s_i, a_i, r_i, s_{i+1}) randomly from its memory D_t, and updates its Q network using the following equation:

$$\theta_{t+1} = \theta_t - \beta \nabla_\theta Q_{\theta_t}(s_i, a_i)\big(Q_{\theta_t}(s_i, a_i) - y_i\big), \tag{9}$$

where β is the learning rate, $\nabla_\theta Q_{\theta_t}(s_i, a_i)$ is the gradient of the Q-network at the point (s_i, a_i) with respect to its parameters θ_t, and y_i is the target Q-value based on a one-step Bellman backup on the sampled experience:

$$y_i = r_i + \gamma \max_a Q_{\theta_t'}(s_{i+1}, a), \tag{10}$$

where γ is the discount rate, and θ_t' are the parameters of the target Q-network that are updated at fixed intervals with the parameters θ_t from the online Q-network. The steps mentioned above are repeated across many time steps and training iterations. In our study, we utilize the double Q-learning extension [30] and a dueling network architecture [31] for improved learning of the Q values. We term this as Dueling DDQN and also refer to it as DQN for simplicity.

V. EXPERIMENTS

In this section, we first evaluate the effectiveness of Dueling DDQN in scheduling just-in-time radar operation while minimizing the AoI of transmitted data. Crucially, we then simulate conditions that our JRC system might encounter in service by considering a scenario where the ego vehicle encounters different environmental conditions across road segments. In each of these experiments, the performance of a JRC scheduler trained using Dueling DDQN is compared with the non-learning Round Robin algorithm and a one-step planner with prior knowledge. Performance is evaluated in terms of the total reward received in each episode, as well as the AoI for data from each urgency class. We also consider the more traditional metric of throughput.

A. Performance

The performance of the DQN agent is compared against the three benchmark algorithms introduced below:

Q-learning: This more classical form of algorithm uses the Bellman update equation introduced in (9) to update a Q-table containing the agent's evaluation of each state-action pair.

Round Robin: This algorithm alternates between radar operation a^r and a communication action; the communication actions cycle between urgency levels, such that the action sequence is $\{a(t=1), a(t=2), \ldots\} = \{a^r, a^{<1>}, a^r, a^{<2>}, \ldots\}$.

One-step Planner: This one-step planner has perfect knowledge of the environment's state transition model and reward function, which it exploits to greedily choose the action with the highest expected instantaneous reward at each time step.

To compare the above-mentioned algorithms of interest, experiments are run in an environment with a greatly reduced state space by restricting the length of the data queue to $N = 3$. This allows the Q-tables to be stored in the memory and storage facilities available on a consumer-level desktop computer. To discourage the agent from letting data packets in the queue become stale, the threshold age is also reduced to $\alpha_{max} = 2$.

We set each episode to comprise of 400 time steps, and conduct training across 2500 episodes. For simplicity, we set the data packet arrival parameters to be equal $\lambda^{<m>} = \lambda$ and set the environmental condition indicators to be discrete values $\{\rho, w, m, v\} \in \{0, 1\}$. For each set of experimental parameters, experiments with different random seeds are performed. Performance variations across random seeds are represented on the graphically plotted results by shaded areas.

The training progress in terms of total reward achieved is shown in Figure 5 for the environment with reduced state space. The DQN agent outperforms both the tabular Q-learning and the non-learning Round Robin agents. The superior performance of the DQN-trained agent in terms of both learning speed and final performance indicates that the use of neural networks allows the agent to generalize across

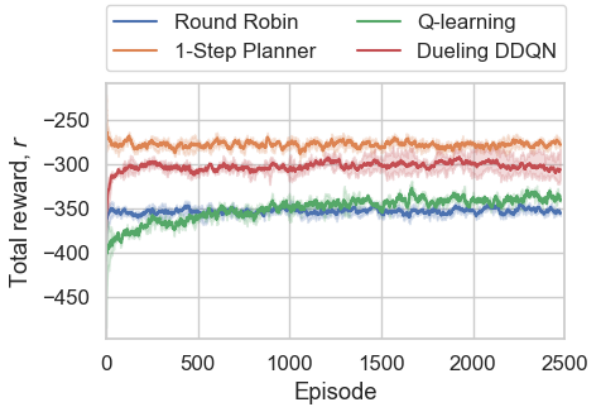

Fig. 5: Total reward r obtained for each episode during the training processes of the DQN and Q-learning agents, compared with the testing process for the Round Robin algorithm and one-step planner.

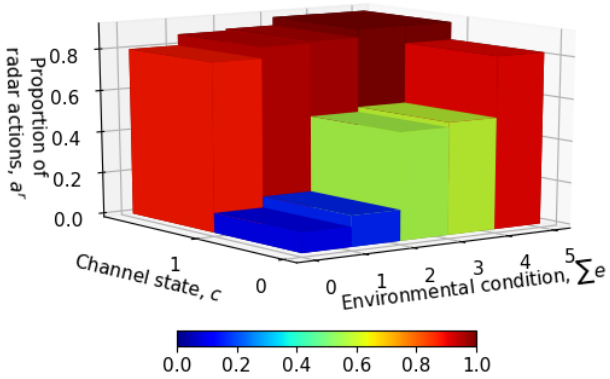

Fig. 6: The proportion of radar operation actions a^r taken by a trained DQN agent across a range of state features e and c.

the state space of the JRC environment. An analysis of the Q-learning algorithm at test time shows that most states visited by the agent are estimated by the Q-table to have their default value of zero, indicating that these states were previously unexplored. This shows that even with the greatly reduced state-space, the representation of the Q-function as a table is inadequate and does not allow for sample-efficient learning.

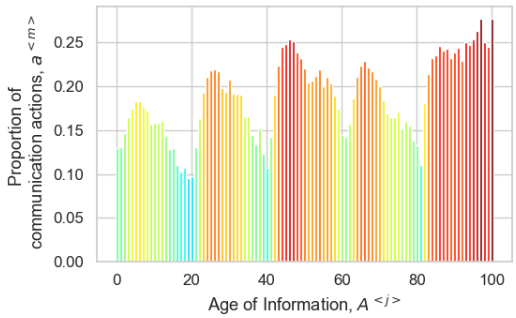

Fig. 7: The proportion of communication actions taken by a trained DQN agent across a range of class ages A.

We produce heat maps of a policy learned by a DQN agent

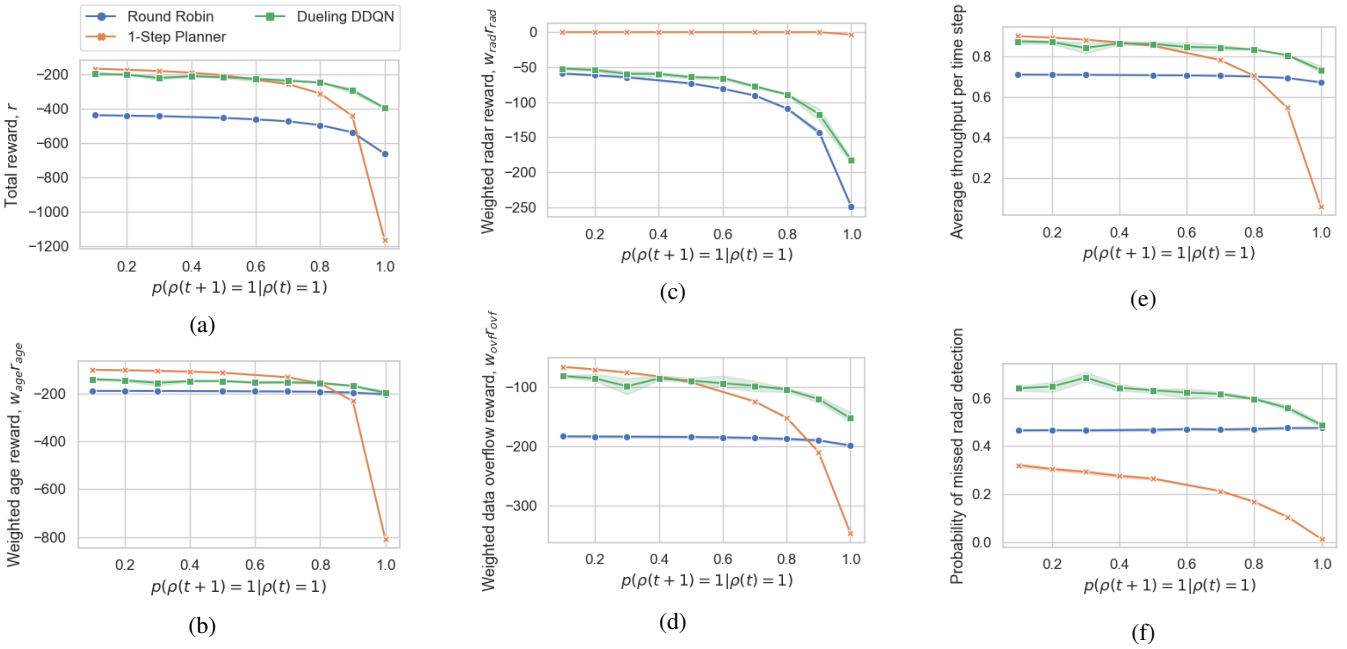

Fig. 8: A comparison the effect of road condition transition probability $p(\rho_{t+1} = 1|\rho_t = 1)$ on (a) the sum of total reward r per episode, (b) total urgency-weighted age reward r_{age} per episode, (c) total radar reward r_{rad} per episode, (d) total data overflow reward $r_{overflow}$ per episode, (e) throughput and (f) probability of missed radar detection.

in Figures 6 and 7. Figure 6 shows the sampled probability that a DQN-trained agent selects the radar operation action a^r given state features $\mathbf{e}$. The results show that the agent learned to choose radar detection more often when the environmental conditions were poorer (indicated by higher values of $\sum \mathbf{e}$), and at almost every event where poor channel state ($c = 1$) decreased data throughput. On the other hand, Figure 7 shows the sampled probability that a DQN-trained agent chooses to transmit data for a given class m for different levels of AoI $\Delta^{<m>}$. The results show that a data class with a higher AoI is generally more likely to have packets transmitted.

B. Effect of varying environmental conditions

While navigating a given route, a vehicle is likely to transit between varying road conditions. In our model, a deterioration in road condition from good to poor corresponds with an increase in the environmental variable ρ from $\rho = 0$ to $\rho = 1$. The study [16] showed that the frequency of accidents scales with $\exp(0.5\text{IRI})$. Consequently, if the IRI transitions from a good condition of 50 ($\rho = 0$) to a fair value of 100 ($\rho = 1$), the crash frequency would increase by 1.65 times.

Consider a scenario where the vehicle takes a route with poorer road conditions, such that the state transition model dictates that the probability of remaining in a poor road $p(\rho_{t+1} = 1|\rho_t = 1)$ increases. The overall risk of a high-risk event would increase, causing a corresponding increase in the need for radar operation. We investigate the effect of increasing $p(\rho_{t+1} = 1|\rho_t = 1)$ from 0 to 1.0 on system performance in this sub-section.

Figure 8 shows the average total episode reward attained by agents trained across a range of values for $p(\rho_{t+1} = 1|\rho_t = 1)$. This is compared against the value of its components $w_{age}r_{age}$ and $w_{rad}r_{rad}$. We also plot supporting performance metrics, such as Figure 8e, which shows the average number of data packets sent per step (throughput), and Figure 8f, which shows the probability that the agent wrongly chooses a communication action when there is a high-risk event.

A higher rate of occurrence of high-risk events X (due to a higher value of $p(\rho_{t+1} = 1|\rho_t = 1)$) increases the number of opportunities to accrue penalties r_{rad}, as shown in Figure 8c. The DQN agent and one-step planner respond to the higher probabilities of high-risk events by more frequently choosing radar detection a^r, as reflected by the declining probability of missed detection in Figure 8f. While Figure 8f shows that the DQN agent misses a higher number of radar events than the Round Robin agent, Figure 8c shows that the DQN agent achieves a higher radar detection reward r_{rad}. This indicates that the DQN agent learned to identify the more safety-critical moments as demarcated by higher values of the environmental features $\mathbf{e}$. The overall performance of the DQN agent, as measured by the total reward per episode (see Figure 8a) is comparable to that of the one-step planner for $p(\rho_{t+1} = 1|\rho_t = 1) = [0.1, 0.6]$. As $p(\rho_{t+1} = 1|\rho_t = 1)$ increases, the performance of the DQN agent exceeds that of the planner. This may be attributed to more information in the reward signal that helps the DQN function approximator to learn the reward function more accurately. Figure 8d indicates that at higher values of $p(\rho_{t+1} = 1|\rho_t = 1)$, the DQN agent is better able to balance the conflicting requirements of increased radar operation and transmitting a constant stream of data without queue overflows. Overall, the results show that the combination of the problem formulation

and the characteristics of the DQN algorithm allow the agent to learn an effective solution specific to the prevailing environmental conditions with minimal knowledge of the environmental parameters.

VI. Conclusion

We considered the problem of joint radar communication (JRC) for autonomous vehicles where the objective is to jointly optimize radar operation and age of information (AoI) of a multi-class data queue by finding a real-time time division policy. By reviewing contemporary literature in automotive sensing and perception in combination with modern understanding of road traffic safety, we established how a JRC scheduling policy module can be integrated into an automotive perception system, and how judicious scheduling of radar sensing can be used to support the objective of road safety. We framed the problem as a Markov Decision Process (MDP), and solved it using an extension of the Deep Q Networks (DQN) algorithm. Experimental results show that the DQN method outperforms the standard tabular Q-learning algorithm and the more traditional round robin queuing algorithm. The DQN method also performs well compared to an exhaustive one-step planner, even with minimal information about the environment. Future work may consider methods to implicitly learn to adapt to continually varying environmental parameters and consider the effect of a multi-agent environment.

References

[1] F. Liu, C. Masouros, A. Petropulu, H. Griffiths, and L. Hanzo, "Joint radar and communication design: Applications, state-of-the-art, and the road ahead," *IEEE Transactions on Communications*, pp. 1–1, 2020.

[2] N. C. Luong, X. Lu, D. T. Hoang, D. Niyato, and D. I. Kim, "Radio resource management in joint radar and communication: A comprehensive survey," *arXiv preprint arXiv:2007.13146*, 2020.

[3] R. Hult, F. E. Sancar, M. Jalalmaab, A. Vijayan, A. Severinson, M. D. Vaio, P. Falcone, B. Fidan, and S. Santini, "Design and experimental validation of a cooperative driving control architecture for the grand cooperative driving challenge 2016," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 4, pp. 1290–1301, 2018.

[4] Q. H. Nguyen, T. H. Dinh, C. L. Nguyen, and D. Niyato, "irdrc: An intelligent real-time dual-functional radar-communication system for automotive vehicles," *IEEE Wireless Communications Letters*, 2020.

[5] J. Choi, V. Va, N. Gonzalez-Prelcic, R. Daniels, C. R. Bhat, and R. W. Heath, "Millimeter-wave vehicular communication to support massive automotive sensing," *IEEE Communications Magazine*, vol. 54, no. 12, pp. 160–167, 2016.

[6] R. Talak, S. Karaman, and E. Modiano, "Optimizing information freshness in wireless networks under general interference constraints," *IEEE/ACM Transactions on Networking*, vol. 28, no. 1, pp. 15–28, 2020.

[7] S. Kaul, M. Gruteser, V. Rai, and J. Kenney, "Minimizing age of information in vehicular networks," in *2011 8th IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks*, June 2011, pp. 350–358.

[8] F. Kunz, D. Nuss, J. Wiest, H. Deusch, S. Reuter, F. Gritschneder, A. Scheel, M. Stübler, M. Bach, P. Hatzelmann, C. Wild, and K. Dietmayer, "Autonomous driving at ulm university: A modular, robust, and sensor-independent fusion approach," in *2015 IEEE Intelligent Vehicles Symposium (IV)*, 2015, pp. 666–673.

[9] D. Nienhuser, T. Gumpp, and J. M. Zollner, "A situation context aware dempster-shafer fusion of digital maps and a road sign recognition system," in *2009 IEEE IV*, 2009, pp. 1401–1406.

[10] J. Steinbaeck, C. Steger, G. Holweg, and N. Druml, "Next generation radar sensors in automotive sensor fusion systems," in *2017 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, 2017, pp. 1–6.

[11] C. Habib, A. Makhoul, R. Darazi, and R. Couturier, "Real-time sampling rate adaptation based on continuous risk level evaluation in wireless body sensor networks," in *2017 IEEE 13th International Conference on WiMob*, 2017, pp. 1–8.

[12] A. Pal and K. Kant, "On the feasibility of distributed sampling rate adaptation in heterogeneous and collaborative wireless sensor networks," in *2016 25th ICCCN*, 2016, pp. 1–9.

[13] D. Giouroukis, A. Dadiani, J. Traub, S. Zeuch, and V. Markl, "A survey of adaptive sampling and filtering algorithms for the internet of things," in *14th ACM International Conference on Distributed and Event-based Systems*, 2020, p. 27–38.

[14] R. Bergel-Hayat, M. Debbarh, C. Antoniou, and G. Yannis, "Explaining the road accident risk: Weather effects," *Accident Analysis & Prevention*, vol. 60, pp. 456–465, 2013.

[15] M. A. Abdel-Aty and A. E. Radwan, "Modeling traffic accident occurrence and involvement," *Accident Analysis & Prevention*, vol. 32, no. 5, pp. 633–642, 2000.

[16] C. Y. Chan, B. Huang, X. Yan, and S. Richards, "Investigating effects of asphalt pavement conditions on traffic accidents in tennessee based on the pavement management system (pms)," *Journal of Advanced Transportation*, vol. 44, no. 3, pp. 150–161, 2010.

[17] J. B. Edwards, "The relationship between road accident severity and recorded weather," *Journal of Safety Research*, vol. 29, no. 4, pp. 249–262, 1998.

[18] J.-L. Martin, "Relationship between crash rate and hourly traffic flow on interurban motorways," *Accident Analysis & Prevention*, vol. 34, no. 5, pp. 619–629, 2002.

[19] H. C. Chin and M. Quddus, "Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections," *Accident Analysis & Prevention*, vol. 35, no. 2, pp. 253–259, 2003.

[20] M. Quddus, "Exploring the relationship between average speed, speed variation, and accident rates using spatial statistical models and gis," *Journal of Transportation Safety & Security*, vol. 5, no. 1, pp. 27–45, 2013.

[21] H. Kurihata, T. Takahashi, I. Ide, Y. Mekada, H. Murase, Y. Tamatsu, and T. Miyahara, "Rainy weather recognition from in-vehicle camera images for driver assistance," in *2005 IEEE IV*, 2005, pp. 205–210.

[22] R. Heinzler, P. Schindler, J. Seekircher, W. Ritter, and W. Stork, "Weather influence and classification with automotive lidar sensors," in *2019 IEEE IV*, 2019, pp. 1527–1534.

[23] N. Abulizi, A. Kawamura, K. Tomiyama, and S. Fujita, "Measuring and evaluating of road roughness conditions with a compact road profiler and arcgis," *Journal of Traffic and Transportation Engineering (English Edition)*, vol. 3, no. 5, pp. 398–411, 2016.

[24] I. Fialho and G. J. Balas, "Road adaptive active suspension design using linear parameter-varying gain-scheduling," *IEEE Transactions on Control Systems Technology*, vol. 10, no. 1, pp. 43–54, 2002.

[25] O. Mees, A. Eitel, and W. Burgard, "Choosing smartly: Adaptive multimodal fusion for object detection in changing environments," in *2016 IEEE/RSJ International Conference on IROS*, 2016, pp. 151–156.

[26] D. Jia, K. Lu, J. Wang, X. Zhang, and X. Shen, "A survey on platoon-based vehicular cyber-physical systems," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 263–284, 2016.

[27] L. Huang and E. Modiano, "Optimizing age-of-information in a multi-class queueing system," in *2015 IEEE ISIT*, June 2015, pp. 1681–1685.

[28] D. P. Bertsekas, *Introduction to probability*, 2nd ed. Belmont, Mass: Athena Scientific, 2008, ch. 6.2, p. 311.

[29] V. Mnih, K. Kavukcuoglu, D. Silver, A. Rusu, J. Veness, M. Bellemare, A. Graves, M. Riedmiller, A. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[30] H. van Hasselt, "Double q-learning," *NeurIPS*, no. 23, pp. 2613–2621, 2010.

[31] Z. Wang, T. Schaul, M. Hessel, H. Hasselt, M. Lanctot, and N. Freitas, "Dueling network architectures for deep reinforcement learning," in *Proceedings of The 33rd ICML*, vol. 48, 2016, pp. 1995–2003.