IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received November 26, 2021, accepted January 24, 2022, date of publication February 4, 2022, date of current version February 17, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3149161

Reinforcement Learning-Based Control of Signalized Intersections Having Platoons

ANAS BERBAR[1], ADEL GASTLI[1], (Senior Member, IEEE), NADER MESKIN[1], (Senior Member, IEEE), MOHAMMED A. AL-HITMI[1], (Member, IEEE), JAWHAR GHOMMAM[2], MOSTEFA MESBAH[2], (Member, IEEE), AND FAÏÇAL MNIF[2], (Senior Member, IEEE)

[1]Electrical Engineering Department, College of Engineering, Qatar University, Doha, Qatar
[2]Department of Electrical & Computer Engineering, College of Engineering, Sultan Qaboos University, Muscat 123, Oman

Corresponding author: Adel Gastli (adel.gastli@qu.edu.qa)

This work was supported by the International Research Collaboration Co-Fund through Qatar University under Grant IRCC2020-15.

ABSTRACT Smart transportation cities are based on intelligent systems and data sharing, whereas human drivers generally have limited capabilities and imperfect traffic observations. The perception of Connected and Autonomous Vehicle (CAV) utilizes data sharing through Vehicle-To-Vehicle (V2V) and Vehicle-To-Infrastructure (V2I) communications to improve driving behaviors and reduce traffic delays and fuel consumption. This paper proposes a Double Agent (DA) intelligent traffic signal module based on the Reinforcement Learning (RL) method, where the first agent, the Velocity Agent (VA) aims to minimize the fuel consumption by controlling the speed of platoons and single CAVs crossing a signalized intersection, while the second agent, the Signal Agent (SA) proceeds to efficiently reduce traffic delays through signal sequencing and phasing. Several simulation studies have been conducted for a signalized intersection with different traffic flows and the performance of the single-agent with only VA, DA with both VA and SA, and Intelligent Driver Model (IDM) are compared. It is shown that the proposed DA solution improves the average delay by 47.3% and the fuel efficiency by 13.6% compared to the Intelligent Driver Model (IDM).

INDEX TERMS Traffic intersection, traffic signal control, platoon control, reinforcement learning, artificial intelligence.

I. INTRODUCTION

Human-driven vehicles are exposed to experience sudden traffic changes on many occasions resulting in consecutive vehicle stops named as ''Traffic Oscillation''. These oscillations include negative impacts such as increasing safety risks and maximizing fuel consumption [1]. The lack of data sharing among the drivers is one of the reasons for the triggering of these traffic oscillations. As shown in [2], vehicles on congested highways are forced to repeatedly deaccelerate and accelerate. Moreover, signalized intersections are another reason for traffic oscillations as they organize the traffic flow by alternating between green and red phases which results in a stop chain of vehicles in red phases [3]. Variable Speed Limit (VSL) is one solution that regulates the moving speed using real-time traffic data. VSL is implemented in signalized intersections [4], though its performance is dependent on the compliance of drivers and the variance in vehicle dynamics [5].

The emergence of smart cities has urged the need to implement smarter transportation systems that depend on Connected and Autonomous Vehicles (CAVs). Controlling CAVs to travel through signalized intersections has been a focus of research as a way of improving transportation safety and efficiency by employing Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I) communications [6], [7]. Introducing platoons of CAVs is more promising in the future as a group of CAVs is moving in the same direction with a single CAV in the front as the ''leader'', and CAVs succeeding as ''Followers'' maintaining a gap distance to the preceding vehicle while following the leader. CAVs are able of lane changing using different controllers such as longitudinal control and lateral control in [8] and PID controllers in [9]. That is required to merge into platoons, as in [10], in which a non-platoon CAV is merged into a cooperative adaptive cruise-controlled platoon. Also, unmerging is possible as shown

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

in [11], which includes merging and unmerging scenarios using distributed and consensus approaches. Platooning is intended to improve traffic management, shorten travel times, and enlarge traffic capacity [12].

The first main controller in intersections for the CAVs is responsible for speed reference. In unsignalized intersections, different controllers for CAVs are proposed, aiming to reach optimization through scheduling the crossing orders of all CAVs. Autonomous intersection management is proposed in [13] that splits the intersection into resources and ascribes them to CAVs in a First-In-First-Out approach. The system was later modified in [14] to account for all vehicle agents' dynamic information instead of simply applying FIFO. In [15], a decentralized energy-optimal control framework is proposed for CAVs, and the approach is extended in [16] to include turns and account for the joint energy-time optimal solution. Multiple intersection scenarios are simulated in [17], [18] using the optimal control approach.

In signalized intersections, trajectory optimization in [19] provides a smooth path for CAVs to cross the signals without stopping at red signals in static environments. Also, optimal eco-driving control is presented in [20] which employs a data-driven approach to account for the uncertainty in signalized time phasing based on dynamic programming for optimization. Lastly, RL-based velocity agents have been developed to locally control CAVs for avoiding obstacles [21] and risky behaviors [22]. In [23], a deep deterministic policy gradient RL-based approach is used to control CAVs behavior, which shows major improvements for various signalizing scenarios. It is also shown in [24], [25] that RL-based schemes can be utilized to improve traffic performance.

The second main controller is the traffic signal phasing and sequencing controller. The main aim of a traffic signal controller is to reduce the average delay of vehicles crossing an intersection, and consequently, increase the traffic throughput. While traditional fixed signalized intersections have poor performance, especially in asymmetric traffic flows, different methods have been developed to smartly control the traffic signals to reduce human involvement, reduce traffic delays and congestion, and most importantly, to keep pace with the development of smart cities in terms of communication. Signal controllers based on fuzzy logic with neural networks are presented in [26], model predictive controller in [27], and Reinforcement Learning (RL) controllers, which is considered a cheaper smart solution in India [28]. The large figure of states in such RL systems as signalized intersections motivated the researchers to find generalization techniques as linear function approximation in [29], state complexity reduction using self-organizing maps in [30], and deep learning in [31], [32].

This paper addresses the problem of combining two smart systems working simultaneously together in signalized intersections in smart cities, including platoons and single CAVs, as shown in Fig. 1, using the RL approach. Toward this goal, this study combines smart signal adaptation with CAVs speed controller to provide a full system and observe the results for a Double Agent (DA) setup working simultaneously. The DA setup decomposes into two agents. The first agent is the Velocity Agent (VA), which is responsible for providing the speed reference for CAVs and platoons inside the intersection. The second agent is the Signal Agent (SA), which is accountable for providing the change in traffic signal sequencing and time phasing for each traffic phase. In this paper, Reinforcement Learning (RL) is used as the main control scheme for both managing an intersection's traffic signals and specifying an optimum speed reference for each individual and platoon of CAVs crossing the intersection.

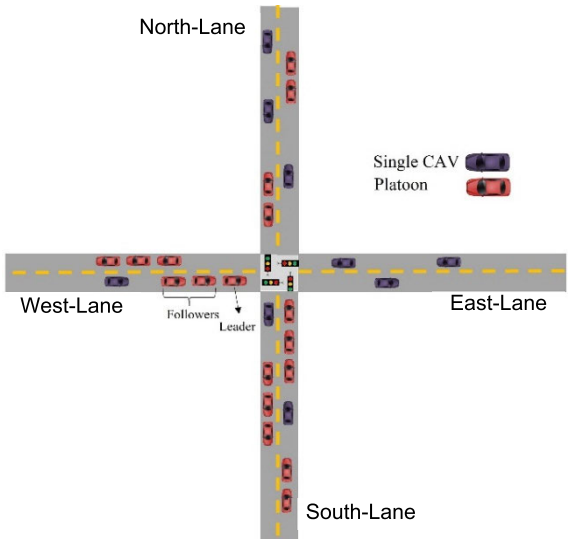


FIGURE 1. Traffic intersection with a mix of individual and platoon CAVs

The main contribution of the paper is that the proposed RL method combines signal sequencing to minimize all vehicle delays and speed trajectory referencing to minimize fuel consumption, and it is shown that in comparison with the benchmark of the Intelligent Driver Model (IDM), the proposed solution has significant improvement in both decreasing the vehicle delays and fuel consumption. To the best of the authors' knowledge, the simultaneous design of both CAVs speed control and traffic signal phasing control has not been considered in the literature, and previous works only investigated each problem separately.

The rest of this paper is organized as follows. In Section II the problem formulation is introduced, and Section III presents the proposed double agent RL-based methodology. Section IV provides the training procedures for both velocity and signal agents. Section V demonstrates the performance of the DA solution, and Section VI shows the results of SA and DA while comparing them with the IDM benchmark in various scenarios. Finally, the conclusion and future research directions are provided in Section VII.

II. PROBLEM FORMULATION

A signalized intersection is a solution to prevent potential crashes and organize the traffic flow. Traditional fixed signalizing has poor performance when running a high unsymmetrical traffic flow, resulting in huge delays. Also, introducing CAVs and platoons into smart cities requires speed referencing to cross the signalized intersection non-stop to minimize fuel consumption. These two problems can be solved by adopting a smart signalized intersection utilizing V2I and V2V communications to ensure the speed trajectory is provided for the CAVs and platoons alongside with the signal sequencing and time phasing of the traffic signals. In this work, the Mcity environment presented in [33] and [34] is considered an intelligent transportation city formulated for CAVs using different control methodologies and communication analogies such as V2V and V2I. The model has been modified to account for platoons and will be reviewed in this section. Table 1 shows the list of notations that are used in this paper for reference.

A. ENVIRONMENT MODEL

The intersection consists of four road segments: West (W), South (N), East (E), and North (N). Each individual segment is represented by a single lane. The traffic signal-enabled directions associated with this design are W-E and S-N. The intersection has a Control Zone (CZ), which is the whole area covered by the smart signal controller that is referred to as the ''Signal-Coordinator'', and the CZ lane length is R. The central square area of the intersection is called the Merging Zone (MZ), with width S, which of possible lateral platoon-CAV collisions, and its traffic flow is controlled by the traffic signals. The entry position of the i-th CAV at the Entry Point (EP) is denoted as $x_i = 0$. The overall model is presented in Fig.2(a). Let $M(t) \in \mathbb{N}$ be the overall number of CAVs entered the CZ following a first-in-first-out queue system. Before a platoon enters the intersection, the agent employs I2V communication with the platoon leader and assigns each platoon a unique ID which is an integer value $i = M(t) + 1$, and $M(t)$ is updated by adding N_i which is the number of CAVs in the i-th platoon as $M(t) = M(t) + N_i$. Continuously, integer i_p will be used to represent the platoon preceding the platoon i in the same lane as shown in Fig. 2(b). The system deployed V2V and V2I communication as shown in Fig. 2(b) where the signal coordinator continuously communicates with the platoon leader using V2I to send the vehicle information and receive speed reference. However, V2V is used continuously when two platoons exist in the same lane to prevent potential accidents.

B. PLATOON MODEL

One of the important aspects in controlling CAVs' speed is the safety distance between the vehicles. In general, the safety distance depends on the speed of the succeeding vehicle and is expressed in terms of time. Precisely, the 2-second rule applies to vehicles traveling at a speed below 12.5 m/s, whereas the 3-second rule applies to all vehicles with no speed limit [35].

TABLE 1. List of Notations.

Parameter	Notation
Entry Point	EP
CZ Lane Length	R
MZ width	S
Platoon ID	i
Vehicle Position	x
Total number of vehicles in CZ	$M(t)$
Number of CAVs in a platoon	N
Preceding vehicle of vehicle i	i_p
The gap between two CAVs	G
Vehicle Length	V_L
Minimum Gap Distance	S_0
IDM Desired Gap	S_i^*
IDM safe Headway Time	T
Acceleration	u
Velocity	v
Maximum Acceleration	$u_{\max}$
Minimum Acceleration	$u_{\min}$
Fuel Consumption from Velocity	f_v^i
Fuel Consumption from Acceleration	f_a^i
Instantaneous Fuel Consumption	f_i
Fuel Model 'q'	q
Fuel Model 'r'	r
Reference Speed	v_r
Maximum Road Speed	v_f
Velocity Agent Time Step	t_v
Velocity Agent State Vector	X_v^i
Vehicle (i) lane's signal status	$X_{v,1}^i$
Golden Binary State	$X_{v,2}^i$
Time left to switch signal status	$t_{left}^{lane_i}$
Velocity Agent Action Vector	A_v
Velocity Agent Reward	R_s^i
Signal Agent Action Vector	A_s
Signal Agent Time Step	t_s
Entry Velocity	v_e
Velocity Agent Reward Exponent Factor	m
Velocity Agent Discount Rate	γ_v
Signal Agent Discount Rate	γ_s
Velocity Agent Step Size	α_v
Signal Agent Step Size	α_s
Velocity Agent Initial Epsilon Value	ε_v
Velocity Agent Step 1 training	n_{v1}
Velocity Agent Step 1 decaying training	n_{v2}
Velocity Agent Step 2 training	n_{v3}
Velocity Agent Step 3 training	n_{v4}
Minimum Platoon Vehicles	$N_{\min}$
Maximum Platoon Vehicles	N_{max}
Training Session Steps	n_{sa}

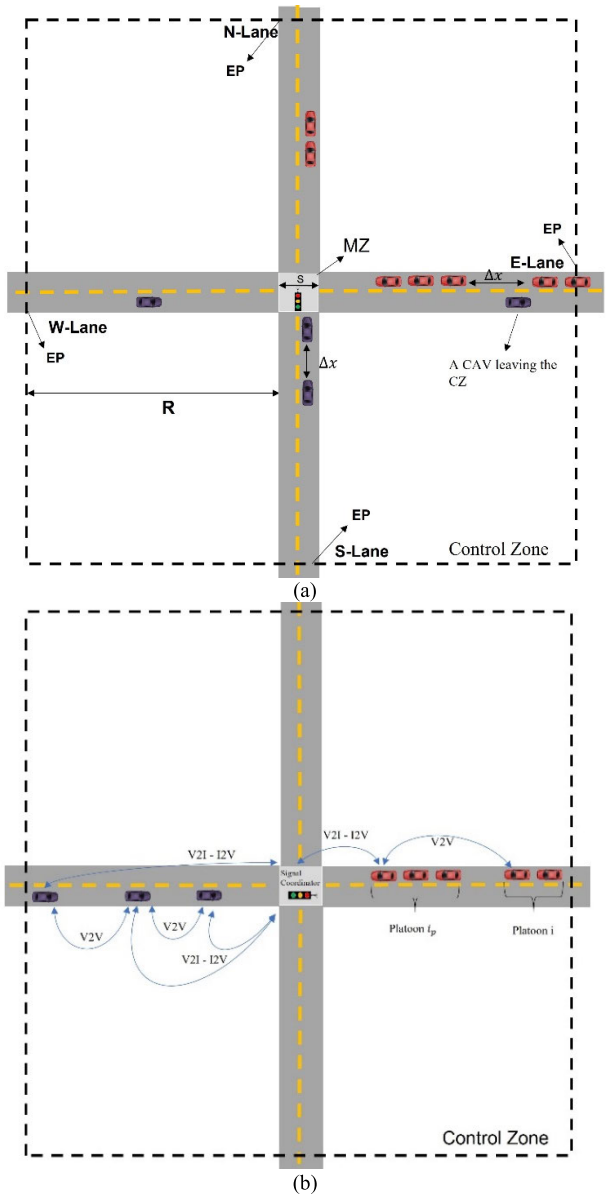


FIGURE 2. Model Review. (a) Control Zone consists of 4-Directions (W/E, S/N, E/W, N/S), single lane for each direction. (b) V2V and V2I Communication, where two leaders in the same lane utilize V2V, and the other CAVs focus on V2I-I2V.

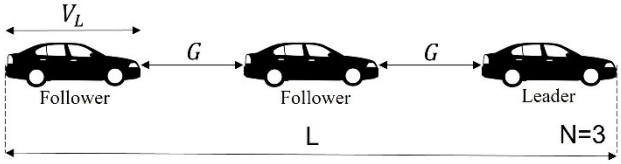


FIGURE 3. Platoon Model

The dynamics of a platoon of CAVs is assumed to be optimal, hence a platoon is represented by a length attribute, where each vehicle's average length is denoted as V_L, and the gap between two vehicles is G, and N is the number of CAVs in the platoon (see Fig. 3).

Accordingly, the length L_i of a platoon i can be continuously calculated as

$$L_i = V_L N_i + (N_i - 1)G_i \tag{1}$$

where $G_i = 2v_i + S_0$ is a variable gap that depends on the velocity of the platoon v_i and calculated maintaining the 2-second rule with a minimum gap distance S_0. Based on (1), each platoon can be represented by a long vehicle, where the followers in each platoon are following the leader using the 2-second rule.

C. INTELLIGENT DRIVER MODEL

The Intelligent Driver Model (IDM) will be used for comparison purposes and is illustrated by Treiber in [36]. IDM is used to model a human driving behavior deploying

$$S_i^* = S_0 + v_i T + \frac{v_i\left(v_i - v_{i_p}\right)}{2\sqrt{u_{\min} u_{\max}}} \tag{2}$$

$$u_i = u_{\max}[1 - \left(\frac{v_i}{v_f}\right)^4 - \left(\frac{S_i^*}{\Delta x_i - V_L}\right)^2] \tag{3}$$

where the instantaneous velocity and position of the ego vehicle is v_i and x_i, respectively, v_f is the maximum road velocity, S_0 is the least distance gap between vehicles, S_i^* is the desired gap between the ego vehicle and the preceding vehicle, and T is the safe headway time which depends on the reaction time of the driver. Consequently, $u_{\max}$ and $u_{\min}$ are the highest and lowest acceleration of the vehicle. The distance between the ego vehicle and the preceding vehicle is represented as

$$\Delta x_i = x_{i_p} - x_i \tag{4}$$

The calculated u_i is the acceleration that is extracted using S_i^*. It should be noted that the same notations can be referred to the platoon leader.

Equations (2) and (3) can be implemented in signalized intersections through 4 cases as follows: Case 1) No preceding vehicle and light is green in which we put $S_i^* = 0$. Case 2) No preceding vehicle but the light is red in which we substitute $v_{i_p} = 0$ in (2) and $V_L = 0$ in (3). Case 3) when a preceding vehicle exists, while the light is green, or the vehicle has already passed the signals. This case uses (2) and (3) as it is. Lastly, Case 4) when the preceding vehicle exists, and the light is red. In this case, it combines Case 2 and Case 3, both calculations presented in those two cases must be evaluated and the reference acceleration result will be the maximum acceleration result among the two evaluations.

D. FUEL CONSUMPTION MODEL

The fuel consumption model is calculated as [37], [38]:

$$\begin{aligned} f_v^i &= q_0 + q_1 v_i + q_2 v_i^2 + q_3 v_i^3 \\ f_a^i &= u_i\left(r_0 + r_1 v_i^2 + r_2 v_i^2\right) \\ f_i &= N_i(f_v^i + f_a^i) \end{aligned} \tag{5}$$

where coefficient vectors $q = [q_0, q_1, q_2]$ and $r = [r_0, r_1, r_2]$ are constants retrieved through experiment in [39], f_i is

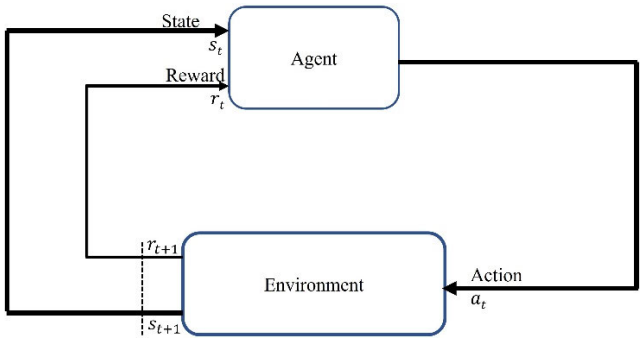


FIGURE 4. Reinforcement Learning Theory

the instantaneous fuel consumed which is the sum of f_v^i (fuel consumed from velocity) and f_a^i (fuel consumed from acceleration) multiplied by number of vehicles inside the platoon. For negative acceleration, then the acceleration value u_i is set to zero.

III. METHODOLOGY

The proposed traffic signal scheme employs a reinforcement learning approach to construct a DA system. The DA subsists into VA and SA. The SA controls the traffic signal phasing, and the VA sends a speed reference to the platoon leader. This section contains 4-sub sections as I-RL Background, II-Velocity Agent, III-Signal Agent, and IV-Override System.

A. REINFORCEMENT LEARNING BACKGROUND

Reinforcement learning (RL) is the process of learning machine learning models to make a set of decisions in a specific environment. The RL agent learns through trial and error to attain a goal and find the optimal sequence of decisions. The basic theory of reinforcement learning is shown in Fig. 4, where at each time step t, the agent gets an observation s_t from the environment's state space $\mathbf{S}$. Consequently, the agent will determine the next action a_t from the action space $\mathbf{A}$ based on the state s_t and apply it to the environment $\mathbf{E}$. The environment then responds to this action, resulting in a transition to a new state $s_{t+1} \in \mathbf{S}$, for which the agent receives a reward r_{t+1}.

1) Q-LEARNING

QL is an algorithm that is widely used in reinforcement learning for finding the optimal policy π^* through maintaining a Q-Matrix consisting of state-action pairs denoted as $Q(s, a)$ which is a matrix that contains the value action of a given action in each state. The action value is an estimation of future rewards that will be collected if this particular action is taken. The $Q(s_t, a_t)$ is estimated from multiple updates performed at time step $t+1$ after receiving r_{t+1} from performing action a_t in state s_t according to

$$\begin{aligned} &Q(s_t, a_t) \\ &= Q(s_t, a_t) + \alpha \\ &\quad \times [r_{t+1} + \gamma \max(Q(s_{t+1}, a_{t+1}))) - Q(s_t, a_t)] \end{aligned} \quad (6)$$

where α is the step size and γ is the discount factor of the future rewards. The actions are chosen using different policies such as ε-greedy where at each state there is a probability ε to perform an action that does not have the highest Q-value. However, a probability of $1-\varepsilon$ for the agent to be greedy and choose the highest Q-value action.

B. VELOCITY AGENT

The VA is the main agent in the system which has all the information of the vehicles in the system. As shown in Fig. 5, once a platoon is at the EP, the leader receives its unique identification number i, and sends the CAVs count inside the platoon N_i, and its lane to the VA. The velocity agent then continuously sends a reference speed v_r according to its current speed and position using RL with a time step t_v. Next, a state, action, and reward definitions of the VA are elaborated in this section. As a result of the constant flow of information inside the VA, it also computes the signal state which will be illustrated in the second part of this section.

1) VELOCITY AGENT STATE DEFINITION

The VA receives information from all platoons inside the CZ. The state definition of the VA is composed of a 4-element vector as:

$$X_v^i = [X_{v,1}^i, X_{v,2}^i, t_{left}^{lane_i}, x_i] \quad (7)$$

where $X_{v,1}^i$ refers to the current platoon lane signal light status whether it is green or not, $X_{v,2}^i$ represents the golden binary state that evaluates the possibility of the platoon's ability to pass as follows:

$$X_{v,2}^i = \begin{cases} 1 & t_i < t_{left}^{lane_i} \text{and light is green} \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

where

$$t_i = \frac{L_i + (R - x_i) - V_L}{v_i} \quad (9)$$

is the time required for the last vehicle inside the platoon 'i' to arrive at the MZ, $t_{left}^{lane_i}$ represents the time left for switching the platoon's lane $lane_i \in \{\mathrm{W-E}, \mathrm{S-N}\}$ signal status, and x_i is the current position of the platoon leader.

2) VELOCITY AGENT ACTION-REWARD DEFINITIONS

The VA action definition A_v is simply the linear distribution between 0 and the road speed v_f i.e. $A_v = [1, 2, \ldots, (v_f + 1)]$ and mapped to velocity reference $v_r = a_v - 1$ where $a_v \in A_v$ is the chosen action. The reward system is a normalized weighted sum of different rewards based on the platoon's velocity as $r_{v,1}$ and its weight as w_1, reaching the golden state as $r_{v,2}$ with weight w_2, and whether the platoon crossed a green or red signal as $r_{v,3}$ and $r_{v,4}$ with weights w_3 and w_4, respectively. The velocity reward is calculated as

$$r_{v,1} = -1 + 2\left(\frac{v}{v_f}\right)^m \quad (10)$$

where m is the reward exponent factor that can control the exponential level of reward for different velocities. The

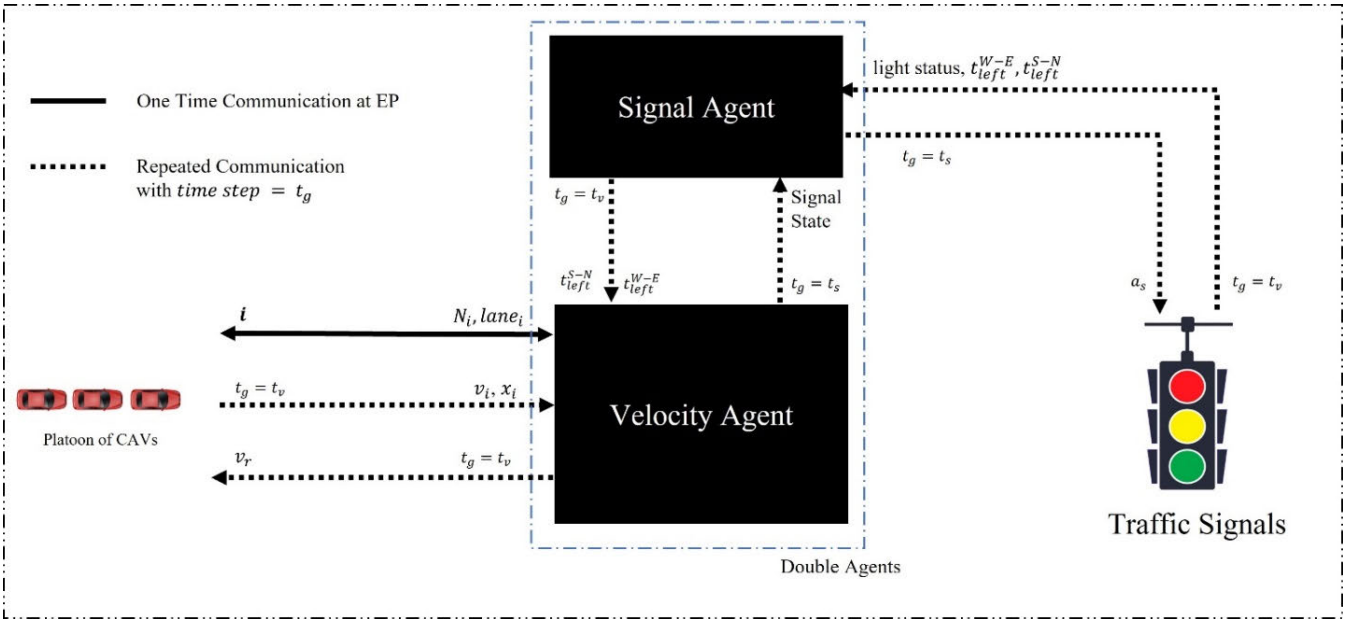


FIGURE 5. Double Agent Communication

TABLE 2. Velocity agent Reward Definition.

Position	Reward	Weight	Reward Range
$x_i < R/2$	Delay: $r_{v,1}^i$	1	[-1,1]
	Golden State: $r_{v,2}^i$	0	-1 or 1
	Cross Green Signal: $r_{v,3}^i$	0	1
	Cross Red Signal: $r_{v,4}^i$	0	-10
$R \geq x_i \geq R/2$	Delay: $r_{v,1}^i$	0.25	[-1,1]
	Golden State: $r_{v,2}^i$	0.75	-1 or 1
	Cross Green Signal: $r_{v,3}^i$	0	1
	Cross Red Signal: $r_{v,4}^i$	0	-10
$x_i > R$	Delay: $r_{v,1}^i$	0	[-1,1]
	Golden State: $r_{v,2}^i$	0	-1 or 1
	Cross Green Signal: $r_{v,3}^i$	1	1
	Cross Red Signal: $r_{v,4}^i$	1	-10

weights as shown in Table 2 are modified according to the position of the platoon leader in the intersection. It should be noted that in the first half distance of the lane, the reward is mainly biased to give all the weight to the velocity since the golden state is either achievable with the maximum speed, or unachievable in the case of the green signal phase and the long distance to the MZ.

However, after exceeding the halfway through, the golden state has more weight as it is the main goal of the agent to find the golden state with the minimum delay, in other words with the highest speed possible. Moreover, a reward is added for crossing a green signal, and punishment for passing a red signal, and each has weight 1 since the platoon can either pass a green or a red signal. Finally, the reward function can be expressed as $R_s^i = \sum_{j=1}^{4} r_{v,j} w_j$.

C. SIGNAL AGENT

In this paper, an RL agent is also used to control the traffic signals sequencing. Since this system is built for platoons of CAVs in the intersection, so the signal timing must be known prior to the platoon leader to determine if the platoon can pass within the green phase or not, and this information is utilized by the VA to find the optimum reference speed in case of not being able to pass. This leads to the action definition A_s as a fixed-sequencing variable timing of {10s, 15s, 20s, 25s}. So, the traffic signals will be alternating between S-N and W-E according to the action $a_s \in A_s$. The SA state vector is composed of five elements. The first element is the current signal phase whether S-N or W-E is green. The remaining elements represent the number of vehicles in the CZ that have not reached the MZ yet. We split each lane into two independent areas to count the vehicles inside each area, as the first area is 35% of R around the MZ, and the second area is the remaining area which mostly contains the vehicles that entered the MZ recently. We chose this proportion as the VA is expected to reach its minimum speed in the first area. Therefore, the second and third elements of the state vector is the number of vehicles in W/E first area and second area respectively. Consequently, the fourth and fifth elements are the number of vehicles in S/N first area and second area respectively. Lastly, the reward was the negative sum of the delay of all platoons which did not cross the MZ yet.

The time step of actions taken is t_s which is a variable that equals the action a_s at the previous time step (t-1).

The delay for the i-th platoon is calculated by evaluating the expected arrival time as

$$t_a^i = \frac{R}{v_e^i} \tag{11}$$

at the EP where v_e^i is the entry velocity of the platoon i, then compare it with the new expected arrival time

$$t_n^i = \frac{R - x_i}{v_i} \tag{12}$$

at each time step t_s. The delay is then multiplied by N_i to have a reward function as $R_s = -\sum_{i=j}^{z} N_i(t_a^i - t_n^i)$ where j and z are the earliest and latest platoon ID in the CZ which has not entered the MZ yet.

D. OVERRIDE SYSTEM

An override system for the VA is implemented for three main reasons as 1) to maintain a safe distance toward the preceding vehicle, 2) to assist the VA with finding the golden state faster, and) 3) to assure no red signal passing. The override system is built based on the condition that if the golden state is not achieved after passing 75% of the control zone width R, then keep reducing the speed until the golden state is achieved. Furthermore, to maintain a safe gap distance, a continuous test for the two-seconds rule

$$\Delta x = x_{i_p} - (x_i + L_i) \leq 2v_i + S_0 \tag{13}$$

must always apply, and once the test fails, the platoon leader sets its own speed to follow the preceding vehicle speed using V2V communication. Lastly, the system should ignore any accelerating speed action from the VA if the platoon will not be able to pass a green signal:

$$\text{if } \left\{ \left(\frac{v_i}{-u_{\min}} \geq t_{left}^{lane_i} \right) \text{ and } X_{v_2}^i = 0 \right\} \rightarrow v_r^i = 0 \tag{14}$$

To summarize, the override system takes control from the VA at any time whenever the following occurs:

1. Two Consecutive CAVs break the 2-seconds rule in Δx.
2. A red signal crossing might occur.
3. The golden state is not achieved after crossing $3/4$ R distance from the EP.

The override system takes control often in the range of (0-70%) depending on the lane traffic and the episode scenario (traffic signals phasing).

The full communications between the proposed DA intersection controller and the platoons-traffic signals are shown in Fig. 5. Note that light status is the current signal phase status and v_r is the velocity reference sent to the platoon leader from the VA set of actions.

IV. TRAINING SETTINGS

In this paper, we use Q-Learning to train the DA as illustrated in reinforcement learning background with Q-Learning review.

A. VELOCITY AGENT TRAINING

The VA is trained firstly through the following environment specifications:

1- The step size is α_v and the discount factor is γ_v.
2- Signals phasing are random actions with an equal probability from A_s.
3- There is a single platoon per lane so it's accident-free.
4- The entry velocity v_e is constant and is equal to v_f.
5- Number of vehicles in a platoon N_i is random between $[N_{\min}, N_{\max}]$ with equal probability.
6- The actions are chosen using ε-greedy policy

The training is divided into four phases and done as follows:

Step 1: The ε value is set to ε_s for n_{v1} episodes (1st phase) and then started decaying with a learning rate l_v for n_{v2} episodes until ε hits approximately zero (2nd phase).
Step 2: The override system is activated for n_{v3} episodes (3rd phase).
Step 3: The override system is deactivated for n_{v4} episodes and the system is trained in the final stage with $\varepsilon = 0$ (4th phase).

The episode starts by entering the platoon from the EP and ends by entering the MZ. Each training step is indeed essential as the first step is to ensure the agent should have enough initial exploration for all the actions in each state, the second step is to assist the agent to exploit more potential optimal actions and put the agent on the right path. Lastly, the third step is to make sure after the override system that any bad greedy actions should be penalized and its Q-value reduced. The values of n_{v1}, n_{v2} and n_{v3} are chosen through trial and error. The training results are shown in Fig. 6 where the average reward per step is calculated by dividing the total reward of the episode by the number of steps in the episode. The average reward per step is calculated since episodes are random for each vehicle due to the randomness of the signal phases which also resulted in big fluctuation in the training. Initially, with a high value of epsilon, the agent is crossing red signals in some episodes which leads to a significant drop in the average reward to an average reward of -0.2846. After Step 3, the drops are eliminated, and the reward averaged at 0.7021 and there is no red signal crossing detected.

B. SIGNAL AGENT TRAINING

As the VA is optimized and tested, it is required to optimize the SA to handle different traffic flows from all directions. The SA training is done through repeated training sessions where a training session is a session of constant traffic flow for n_{sa} steps.

The following SA training assumptions are followed

1) The SA step size is α_s and its discount factor is γ_s.
2) The entry velocity v_e is constant and is equal to v_f.
3) Traffic flows from each direction are constant for each training session.

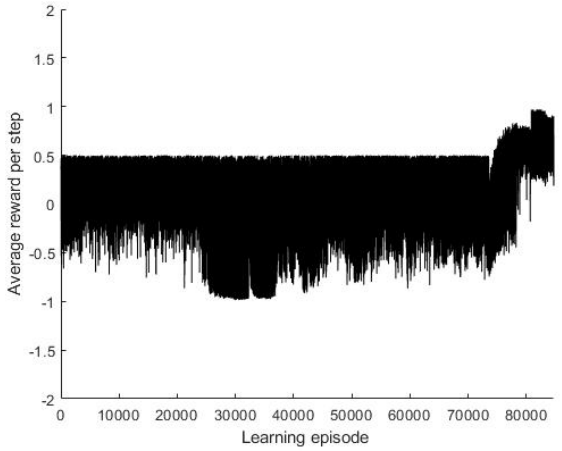


FIGURE 6. Velocity agent Training.

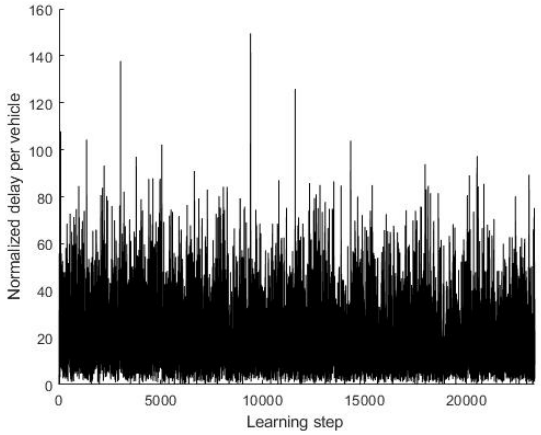


FIGURE 7. Signal Agent Case-1 Training.

4) The normalized reward is shown in Fig. 7 which is the negative sum of delay divided by the number of vehicles that got delayed for each step.

The same training steps are repeated for different traffic flows. In Fig. 7 which shows the training for Case-1 traffic flow in Table 3, the bad actions lead to huge delays and caused massive spikes, while by the end of the training these spikes are eliminated.

The signal agent is trained for urban intersections with a maximum traffic flow of 700veh/h for each direction. However, there are still small spikes since the normalized reward is only considering the vehicles that are being delayed not all the vehicles in the CZ. The values of S_0, A and T are obtained from [36] and shown in Table 3 with other used values for training both velocity and signal agents.

V. SYSTEM PERFORMANCE

We use the modified Mcity as the MATLAB/SIMULINK environment for measuring the performance. The model is supposed to work under various traffic demands from all directions. The N_i is generated between $[N_{\min} N_{\max}]$ with an equal probability. In this section, we will look deeply into one scenario which is Case I from Table 4.

The simulation trajectories corresponding to the position of CAVs-Platoons are plotted with respect to time for the double agent in W, S, E, and N lanes in Figs. 8, 9, 10, and 11, respectively. The platoons have learned to reduce their speed in advance before reaching the MZ (at 400m) to avoid traffic oscillations. Also, it can be noticed that the traffic signals phasing is changing depending on the traffic demand by

TABLE 3. Parameter Simulated Values.

Notation	Value
$u_{\max}$	2 m/s^2
$u_{\min}$	-5m/s^2
S_0	2m
v_f	13m/s
V_L	5m
T	1.6s
m	1.5
q	[0.1569 0.0245 -0.0007415]
r	[0.07224 0.09681 0.001075]
t_v	1s
t_s	Previous a_s
γ_v	0.8
γ_s	0.8
α_v	0.1
α_s	0.1
ε_v	0.9
n_{v1}	73560 episodes
n_{v2}	7287 episodes
n_{v3}	2402 episodes
n_{v4}	1471 episodes
$N_{\min}$	1 CAV
N_{max}	3 CAVs
n_{sa}	5000 steps

TABLE 4. Traffic Flow Cases.

Case	Entry Traffic Flow (Veh/h)			
	West	South	East	North
Case I	661	304	681	326
Case II	661	597	678	635
Case III	227	230	260	248
Case IV	661	452	230	59

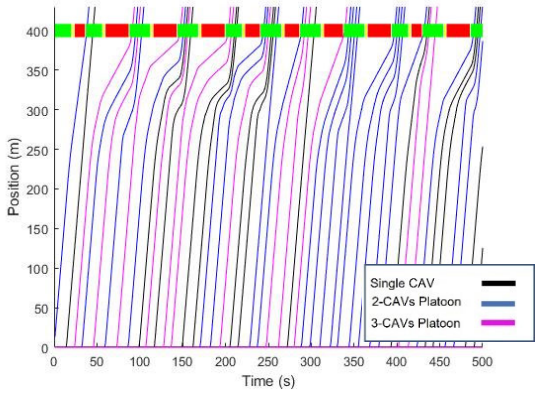


FIGURE 8. West lane trajectories of CAVs/Platoons corresponding to Case I using double-agent approach.

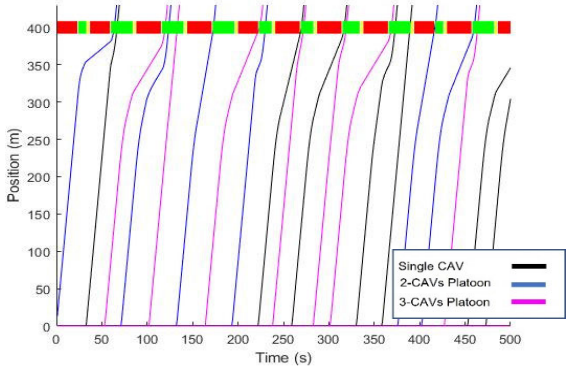


FIGURE 9. South lane trajectories of CAVs/Platoons corresponding to Case I using double-agent approach.

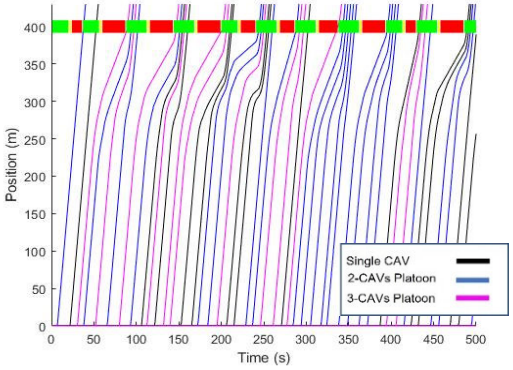


FIGURE 10. East lane trajectories of CAVs/Platoons corresponding to Case I using double-agent approach.

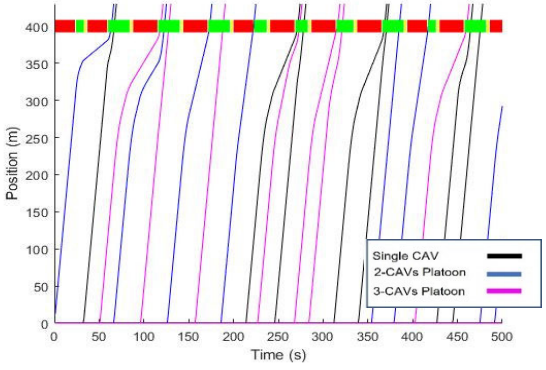


FIGURE 11. North lane trajectories of CAVs/Platoons corresponding to Case I using double-agent approach.

having the signal agent as it provides more green phases to the W-E than the S-N lane as it has higher traffic. The average velocity of all the vehicles in the four lanes that have not crossed the MZ yet is shown in Fig. 12. The minimum gap for each platoon is calculated with respect to the preceding platoon is continuously recorded and updated to ensure its an accident-free and follow S_0 constraint as shown in Fig. 13.

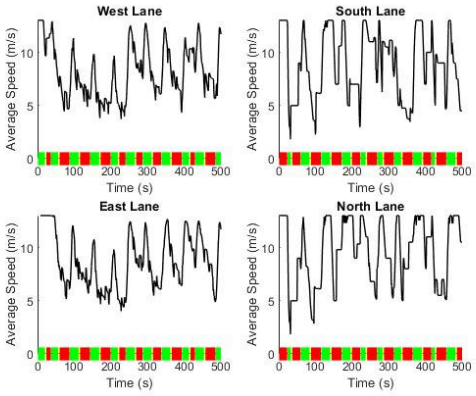


FIGURE 12. Average Speed of CAVs/Platoons corresponding to Case I using double-agent approach.

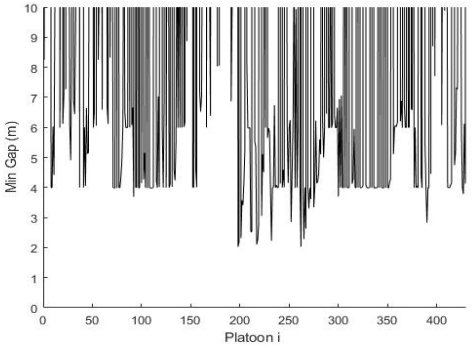


FIGURE 13. Minimum gap between CAVs/Platoon corresponding to Case I using double-agent approach.

One limitation of the DA is that the SA will only look to minimize the total delay in all lanes, giving longer signal phases to the higher traffic flow lanes can trigger a drop-down in the average speed of the low traffic flow lanes in which SA gives a very small phase, this can be seen in Fig. 12 North/South lanes at (Time=25s), where the platoon was forced to minimize its speed due to the low phase signal, and since it was the only platoon in the lane at that time (Fig. 9 and Fig. 11), the average speed was equal to the platoon's reduced speed which was significantly dropped. However, in most scenarios, this should not trigger a traffic oscillation because it happens with low traffic flow lanes, except if there is a succeeding vehicle that will be forced to stop and reach a minimum gap of S_0 as shown in several platoons in Fig. 13.

VI. SYSTEM EVALUATION COMPARISON WITH BENCHMARK

In this section, there are three different systems to evaluate the performance of the developed systems as:

1- The main benchmark that is the IDM with a fixed signalizing.
2- The single-agent that is the developed VA with a fixed signalizing.

TABLE 5. IDMs with Fixed Signalizing.

IDM Fixed Signalizing	Fixed Time	Avg Delay (s)	Avg Fuel Consumption (mL)
Case I	15	160.1	42.4
	20	124.8	40.48
	25	96.3	38.2
Case II	15	109.7	38.9
	20	89.8	33.3
	25	67.9	31.4
Case III	10	6.6	21.5
	15	7.85	22
	20	9.64	22.9
Case IV	10	91.0	36.77
	15	81.1	32.33
	20	64	28
	25	52	27.2

3- DA setup using the combination of the velocity and signal agents.

To choose a suitable fixed signalizing period to compare with our system, we analyzed the different options available in A_s with IDMs for the four cases mentioned below where each case with a fixed time has been run for 1hour of simulation and its results are shown in Table 5. The best performing time was 25s in Case I, Case II, and Case IV. However, 10s was sufficient for Case III. In order to compare with a constant fixed signalizing instead of alternating between the best timing since alternating would be considered as a smart system itself, hence we chose the 20-second phasing as the average best action.

The traffic signals have 3 seconds yellow light between switching. There are four different traffic flow scenarios are simulated as Case I: Unsymmetrical considerably high traffic flow, Case II: Symmetrical considerably high traffic flow, Case III: Unsymmetrical low traffic flow, and Case IV: Extremely unsymmetrical traffic flow as shown in Table 4.

The comparison among different solutions is mainly focused on measuring the average delay, and average fuel consumption. The simulation is conducted for each scenario for 1 hour and the corresponding results are presented in Table 6. The delay is calculated by subtracting the estimated arrival time $t_{a,i}$ at the entry point from the actual arrival time to MZ, and the fuel consumption is accumulated as briefly illustrated in the previous sections until passing the MZ. The

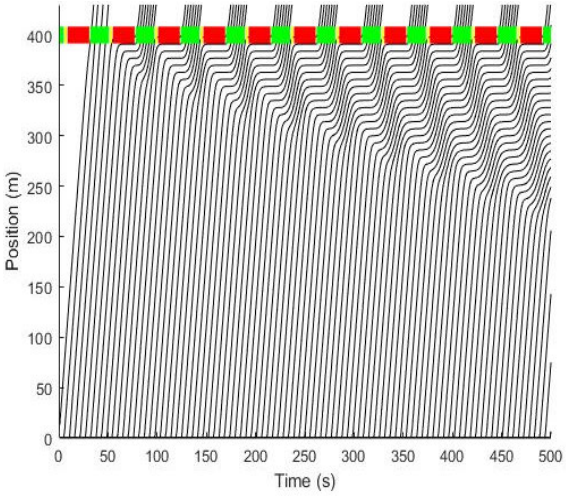


FIGURE 14. West lane trajectories of CAVs corresponding Case II traffic scenario using IDM approach.

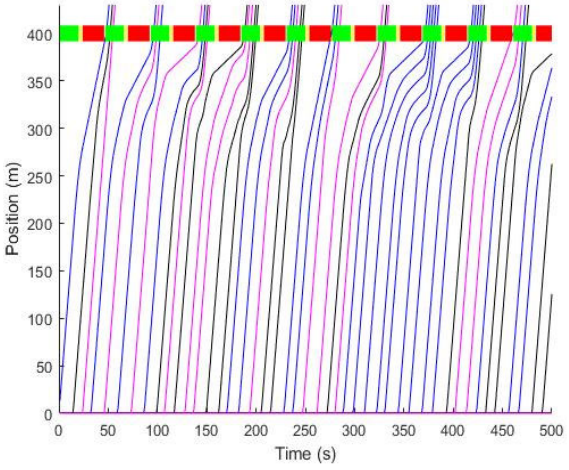


FIGURE 15. West lane trajectories of CAVs corresponding Case II traffic scenario using single-agent approach.

improvement is measured compared to the IDM system as the benchmark.

Table 6 shows that VA and DA perform better than the IDM approach in three out of four cases. In case III, the traffic flow is low and symmetric as well which is the perfect scenario for IDMs with the fixed signalizing. However, it is not a practical scenario as most urban intersections have unsymmetrical traffic flows. In the remaining cases, the single/double-agent outperform the IDM approach with a significant delay/fuel efficiency improvement as a result of two main reasons as I) The platoons are more efficient in maximizing the throughput of the green signals where more CAVs are passing the green signal together since they are arriving together as a platoon and II) The velocity agent eliminates the traffic oscillations caused generally in human-driven vehicles which causes huge delays and extra fuel consumption. Fig. 14 shows the traffic oscillations caused by the IDM approach. Fig. 15 and

TABLE 6. The Comparison Results of Approaches.

Entry Traffic	Approach	Avg Delay (s)	Avg Fuel Consumption (mL)	Avg W-E Green Time (s)	Avg S-N Green Time (s)	Delay Improvement %	Fuel Consumption Improvement %
Case I	Double RL	29.0	26.7	18.3	12.46	76.7%	34%
	Fixed Signals with Velocity agent	41.4	30.8	20	20	66.8%	23.9%
	Fixed Signals with IDM	124.8	40.48	20	20	Benchmark	Benchmark
Case II	Double RL	33.6	28.95	19.5	18.6	62.5%	13%
	Fixed Signals with Velocity agent	39.5	31.18	20	20	56%	6.33%
	Fixed Signals with IDM	89.8	33.3	20	20	Benchmark	Benchmark
Case III	Double RL	11.59	18.59	16.9	15.5	-20.2%	-10%
	Fixed Signals with Velocity agent	11.1	18.3	20	20	-15.15%	-8.7%
	Fixed Signals with IDM	9.64	16.84	20	20	Benchmark	Benchmark
Case IV	Double RL	19.17	23.1	19.7	16.6	70.0%	17.5%
	Fixed Signals with Velocity agent	21.1	24.14	20	20	67%	13.8%
	Fixed Signals with IDM	64	28	20	20	Benchmark	Benchmark

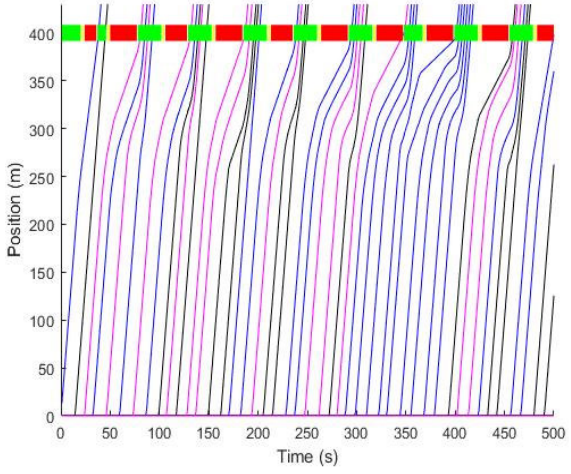


FIGURE 16. West lane trajectories of CAVs corresponding Case II traffic scenario using double-agent approach.

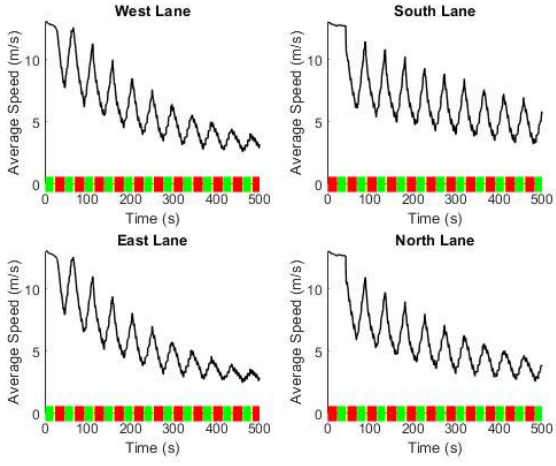


FIGURE 17. Average Speed of vehicles corresponding to Case II using IDM approach.

Fig. 16 are showing the single agent and DA setup where the traffic oscillations are eliminated, and in DA the signal phase timings improved (all figures present Case II W-Lane). The average speed of IDM, Single-Agent and DA is plotted in Fig. 17, Fig. 18 and Fig. 19 respectively. We can see that the average speed of the single agent (Fig. 18) and DA (Fig. 19) does not have the limitation mentioned earlier of low phases, which eliminated those sharp drops in the average speed. This is mainly due to having high symmetrical traffic flow, wherein the single-agent signal phases are always 20-seconds, and in the DA, the signal phases reach 25 seconds in many states.

The final results of single-agent average delay improvement is 43.7%, and fuel consumption average improvement is 8.8% based on the four cases compared to the benchmark. Consequently, implementing the DA system leads to more efficient results especially in the unsymmetrical traffic flow cases which is the most practical scenario. The DA delay

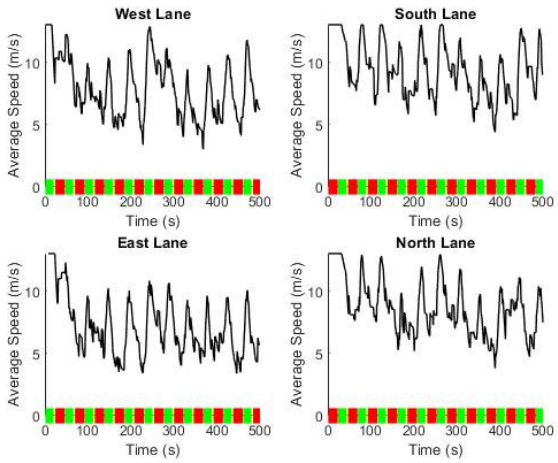


FIGURE 18. Average Speed of CAVs/Platoons corresponding to Case II using single-agent approach.

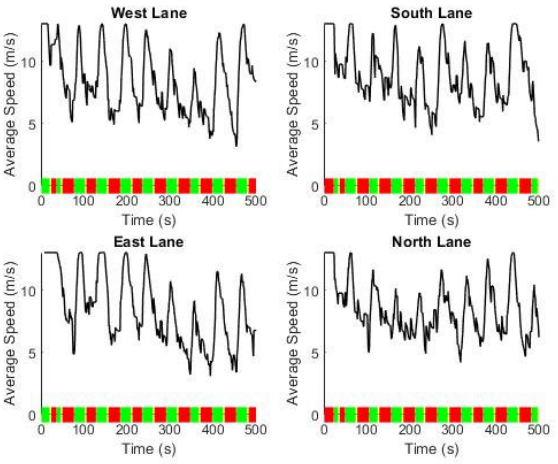


FIGURE 19. Average Speed of CAVs/Platoons corresponding to Case II using double-agent approach.

average improvement is 47.3%, and fuel consumption was 13.6% average more efficient than the benchmark.

VII. CONCLUSION

In this study, we proposed a double QL agents RL based to control the platoons of CAVs into signalized intersections in the promised smart cities. The training of the double agent is performed in a decentralized manner where the velocity agent is trained and executed to train the SA. There are two main improvements in the presented system:

1) The first improvement is reducing the average delay of CAVs passing urban intersections with an average improvement of 47.3%.
2) The second improvement is the fuel efficiency of an average of 13.6%, which is a critical part to consider in the long term.

Two main points are drawn from the results, the first point is that introducing platoons in higher traffic flows can effectively reduce fuel consumption and average delays. The second point is that the SA can adapt to symmetrical/unsymmetrical traffic flows which is highly needed. In this proposed design, the override system has played a major role in assisting the VA to avoid potential accidents and find the golden state. Our aim in the future is to add neural networks to both agents in order to remove the override system in the VA to make it a fully smart system and reach a better optimal policy in the SA. Also, we are looking to simulate traffic conflicts for accident detection.

ACKNOWLEDGMENT

The findings achieved herein are solely the responsibility of the authors. The open-access publication of this article was funded by Qatar National Library.

REFERENCES

[1] D. Chen, J. Laval, Z. Zheng, and S. Ahn, "A behavioral car-following model that captures traffic oscillations," *Transp. Res. B, Methodol.*, vol. 46, no. 6, pp. 744–761, Jul. 2012.
[2] X. Li, J. Cui, S. An, and M. Parsafard, "Stop-and-go traffic analysis: Theoretical properties, environmental impacts and oscillation mitigation," *Transp. Res. B, Methodol.*, vol. 70, pp. 319–339, Dec. 2014.
[3] F. Zhou, X. Li, and J. Ma, "Parsimonious shooting heuristic for trajectory design of connected automated traffic Part I: Theoretical analysis with generalized time geography," *Transp. Res. B, Methodol.*, vol. 95, pp. 394–420, Jan. 2017.
[4] M. Sanchez, J.-C. Cano, and D. Kim, "Predicting traffic lights to improve urban traffic fuel consumption," in *Proc. 6th Int. Conf. ITS Telecommun.*, Jun. 2006, pp. 331–336.
[5] C. Fuhs and P. Brinckerhoff, "Synthesis of active traffic management experiences in Europe and the United States," Federal Highway Admin., Washington, DC, USA, Tech. Rep. FHWA-HOP-10-031, 2010.
[6] S. I. Guler, M. Menendez, and L. Meier, "Using connected vehicle technology to improve the efficiency of intersections," *Transp. Res. C, Emerg. Technol.*, vol. 46, pp. 121–131, Sep. 2014.
[7] L. Li, D. Wen, and D. Y. Yao, "A survey of traffic control with vehicular communications," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 1, pp. 425–432, Feb. 2014.
[8] R. Zhou, Z. Li, and Z. Yang, "Optimal lane change analysis for vehicle platooning based on lateral and longitudinal control," in *Proc. 39th Chin. Control Conf. (CCC)*, Jul. 2020, pp. 5534–5539.
[9] A. Mehmood, M. Liaquat, A. I. Bhatti, and E. Rasool, "Trajectory planning and control for lane-change of autonomous vehicle," in *Proc. 5th Int. Conf. Control, Automat. Robot. (ICCAR)*, Apr. 2019, pp. 331–335.
[10] J. Hu, H. Wang, X. Li, and X. Li, "Modelling merging behaviour joining a cooperative adaptive cruise control platoon," *IET Intell. Transp. Syst.*, vol. 14, no. 7, pp. 693–701, Jul. 2020.
[11] S. Santini, A. Salvi, A. S. Valente, A. Pescapè, M. Segata, and R. L. Cigno, "Platooning maneuvers in vehicular networks: A distributed and consensus-based approach," *IEEE Trans. Intell. Vehicles*, vol. 4, no. 1, pp. 59–72, Mar. 2019.
[12] S. Badnava, N. Meskin, A. Gastli, M. A. Al-Hitmi, J. Ghommam, M. Mesbah, and F. Mnif, "Platoon transitional maneuver control system: A review," *IEEE Access*, vol. 9, pp. 88327–88347, 2021.
[13] K. Dresner and P. Stone, "A multiagent approach to autonomous intersection management," *J. Artif. Intell. Res.*, vol. 31, pp. 591–656, Mar. 2008.
[14] Q. Jin, G. Wu, K. Boriboonsomsin, and M. Barth, "Multi-agent intersection management for connected vehicles using an optimal scheduling approach," in *Proc. Int. Conf. Connected Vehicles Expo (ICCVE)*, Dec. 2012, pp. 185–190.
[15] A. A. Malikopoulos, C. G. Cassandras, and Y. Zhang, "A decentralized energy-optimal control framework for connected automated vehicles at signal-free intersections," *Automatica*, vol. 93, pp. 244–256, Jul. 2018.
[16] Y. Zhang and C. G. Cassandras, "Decentralized optimal control of connected automated vehicles at signal-free intersections including comfort-constrained turns and safety guarantees," *Automatica*, vol. 109, Nov. 2019, Art. no. 108563.

[17] Y. J. Zhang, A. A. Malikopoulos, and C. G. Cassandras, "Optimal control and coordination of connected and automated vehicles at urban traffic intersections," in *Proc. Amer. Control Conf. (ACC)*, Jul. 2016, pp. 6227–6232.

[18] B. Chalaki and A. A. Malikopoulos, "Optimal control of connected and automated vehicles at multiple adjacent intersections," *IEEE Trans. Control Syst. Technol.*, early access, May 31, 2021, doi: 10.1109/TCST.2021.3082306.

[19] J. Ma, X. Li, F. Zhou, J. Hu, and B. B. Park, "Parsimonious shooting heuristic for trajectory design of connected automated traffic Part II: Computational issues and optimization," *Transp. Res. B, Methodol.*, vol. 95, pp. 421–441, Jan. 2017.

[20] C. Sun, J. Guanetti, F. Borrelli, and S. J. Moura, "Optimal eco-driving control of connected and autonomous vehicles through signalized intersections," *IEEE Internet Things J.*, vol. 7, no. 5, pp. 3759–3773, May 2020.

[21] H.-T. Tseng, C.-C. Hsieh, W.-T. Lin, and J.-T. Lin, "Deep reinforcement learning for collision avoidance of autonomous vehicle," in *Proc. IEEE Int. Conf. Consum. Electron.-Taiwan (ICCE-Taiwan)*, Sep. 2020, pp. 1–2.

[22] S. Mo, X. Pei, and C. Wu, "Safe reinforcement learning for autonomous vehicle using Monte Carlo tree search," *IEEE Trans. Intell. Transp. Syst.*, early access, Mar. 11, 2021, doi: 10.1109/TITS.2021.3061627.

[23] M. Zhou, Y. Yu, and X. Qu, "Development of an efficient driving strategy for connected and automated vehicles at signalized intersections: A reinforcement learning approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 1, pp. 433–443, Jan. 2020.

[24] C. Wu, A. M. Bayen, and A. Mehta, "Stabilizing traffic with autonomous vehicles," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, May 2018, pp. 6012–6018.

[25] Z. Cao, H. Guo, J. Zhang, F. Oliehoek, and U. Fastenrath, "Maximizing the probability of arriving on time: A practical Q-learning method," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 1–7.

[26] D. Jia and Z. Chen, "Traffic signal control optimization based on fuzzy neural network," in *Proc. Int. Conf. Meas., Inf. Control*, vol. 2, 2012, pp. 1015–1018.

[27] H. Nakanishi and T. Namerikawa, "Optimal traffic signal control for alleviation of congestion based on traffic density prediction by model predictive control," in *Proc. 55th Annu. Conf. Soc. Instrum. Control Eng. Jpn. (SICE)*, Sep. 2016, pp. 1273–1278.

[28] N. Bhave, A. Dhagavkar, K. Dhande, M. Bana, and J. Joshi, "Smart signal—Adaptive traffic signal control using reinforcement learning and object detection," in *Proc. 3rd Int. Conf. I-SMAC (IoT Social, Mobile, Analytics Cloud) (I-SMAC)*, Dec. 2019, pp. 624–628.

[29] L. N. Alegre, T. Ziemke, and A. L. C. Bazzan, "Using reinforcement learning to control traffic signals in a real-world scenario: An approach based on linear function approximation," *IEEE Trans. Intell. Transp. Syst.*, early access, Jun. 29, 2021, doi: 10.1109/TITS.2021.3091014.

[30] M. Miletić, K. Kušić, M. Gregurić, and E. Ivanjko, "State complexity reduction in reinforcement learning based adaptive traffic signal control," in *Proc. Int. Symp. ELMAR*, Sep. 2020, pp. 61–66.

[31] A. Jaleel, M. A. Hassan, T. Mahmood, M. U. Ghani, and A. U. Rehman, "Reducing congestion in an intelligent traffic system with collaborative and adaptive signaling on the edge," *IEEE Access*, vol. 8, pp. 205396–205410, 2020.

[32] K. Yang, I. Tan, and M. Menendez, "A reinforcement learning based traffic signal control algorithm in a connected vehicle environment," in *Proc. 17th Swiss Transp. Res. Conf. (STRC)*, May 2017, pp. 1–5, doi: 10.3929/ethz-b-000130809.

[33] Y. Zhang, C. G. Cassandras, W. Li, and P. J. Mosterman, "A discrete-event and hybrid simulation framework based on SimEvents for intelligent transportation system analysis," *IFAC-PapersOnLine*, vol. 51, no. 7, pp. 323–328, 2018.

[34] Y. Zhang, C. G. Cassandras, W. Li, and P. J. Mosterman, "A simevents model for hybrid traffic simulation," in *Proc. Winter Simulation Conf. (WSC)*, Dec. 2017, pp. 1455–1466.

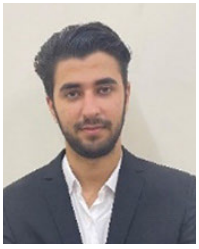
[35] T. Hailemariam Yimer, C. Wen, X. Yu, and C. Jiang, "A study of the minimum safe distance between human driven and driverless cars using safe distance model," 2020, *arXiv:2006.07022*.

[36] M. Treiber, A. Hennecke, and D. Helbing, "Congested traffic states in empirical observations and microscopic simulations," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 62, no. 2, p. 1805, Aug. 2000.

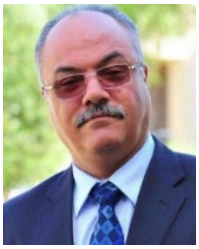
[37] J. Rios-Torres and A. A. Malikopoulos, "Automated and cooperative vehicle merging at highway on-ramps," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 4, pp. 780–789, Apr. 2017.

[38] M. A. S. Kamal, M. Mukai, J. Murata, and T. Kawabe, "Model predictive control of vehicles on urban roads for improved fuel economy," *IEEE Trans. Control Syst. Technol.*, vol. 21, no. 3, pp. 831–841, Apr. 2012.

[39] W. Zhao, D. Ngoduy, S. Shepherd, R. Liu, and M. Papageorgiou, "A platoon based cooperative eco-driving model for mixed automated and human-driven vehicles at a signalised intersection," *Transp. Res. C, Emerg. Technol.*, vol. 95, pp. 802–821, Oct. 2018.



ANAS BERBAR received the B.Sc. degree in electrical engineering from Qatar University, in 2020. His undergraduate studies mainly focused on power electronics and electric vehicles charging stations. Since 2020, he has been working as a Research Assistant in the field of machine learning and intelligent transportation systems.

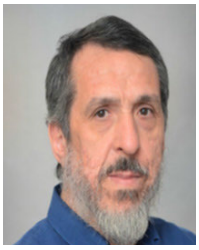


ADEL GASTLI (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from the National School of Engineers of Tunis, Tunisia, in 1985, and the M.Sc. and Ph.D. degrees in electrical & computer engineering from the Nagoya Institute of Technology, Japan, in March 1990 and March 1993, respectively. From September 1985 to September 1987, he worked with the National Institute for Standards and Intellectual Property, Tunisia. He worked with Mitsubishi Electric Corporation, Japan, from April 1993 to July 1995. He joined the Electrical and Computer Engineering Department, Sultan Qaboos University, Oman, in August 1995. He worked as the Head of the Department, from September 2001 to August 2003, and from September 2007 to August 2009. He was appointed as the Director of Sultan Qaboos University, Quality Assurance Office, from February 2010 to January 2013. In February 2013, he joined the Electrical Engineering Department, Qatar University, as a Professor and the Kahramaa-Siemens Chair of Energy Efficiency. From August 2013 to September 2015, he was appointed the College of Engineering Associate Dean for Academic Affairs. His current research interests include energy efficiency, renewable energy, electric vehicles, and smart grid.

NADER MESKIN (Senior Member, IEEE) received the B.Sc. degree from the Sharif University of Technology, Tehran, Iran, in 1998, the M.Sc. degree from the University of Tehran, Tehran, in 2001, and the Ph.D. degree in electrical and computer engineering from Concordia University, Montreal, QC, Canada, in 2008. He was a Postdoctoral Fellow at Texas A&M University at Qatar, Doha, Qatar, from January 2010 to December 2010. He is currently an Associate Professor at Qatar University, Doha, and an Adjunct Associate Professor at Concordia University. He has published more than 230 refereed journals and conference papers. His research interests include FDI, multiagent systems, active control for clinical pharmacology, cyber-security of industrial control systems, and linear parameter varying systems.

MOHAMMED A. AL-HITMI (Member, IEEE) received the B.Sc. degree in electrical engineering from Qatar University, Doha, Qatar, in 1992, and the M.S. and Ph.D. degrees in control engineering from the University of Sheffield, in 1994 and 2002, respectively. He is currently an Associate Professor in electrical engineering with Qatar University, where he is also working as the Head of the Department of Electrical Engineering. He has conducted many research projects funded by national and industrial funding agencies. He has authored more than 50 research articles in top peer-reviewed journals and conferences. He is involved in several administrative committees in leadership roles with Qatar University. He is also serving as a reviewer for many top journals. His research interests include control systems theory, neural networks, fuzzy control, and electric drive systems.

JAWHAR GHOMMAM received the B.Sc. degree in computer and control engineering from the National Institute and Applied Sciences and Technology (INSAT), Tunis, in 2003, the D.E.A. (M.Sc.) degree from the Laboratoire d'Informatique, Robotique et Micro-électronique (LIRMM), University of Montpelier, France, in 2004, and the joint Ph.D. degree in control engineering from the National Engineering School of Sfax and the University of Orleans, in 2008. From 2008 to 2017, he was with the National Institute of Applied Sciences and Technology, where he held a tenured Associate Professor with the Department of Physics and Instrumentation. In January 2018, he joined the Department of Electrical and Computer Engineering, Sultan Qaboos University, Oman. He is a member of the Control and Energy Management Laboratory and also an Associate Researcher with the GREPCI-Lab, Ecole de Technologie Superieure, Montreal, QC, Canada. His research interests include fundamental motion control concepts for nonholonomic/underactuated vehicle systems, nonlinear and adaptive control, intelligent and autonomous control of networked unmanned systems, team cooperation, consensus achievement, and sensor networks. He serves as a regular referee and an associate editor for many international journals in the field of control and robotics.



MOSTEFA MESBAH (Member, IEEE) received the Ph.D. degree in the area of learning control systems from the Department of Electrical and Computer Engineering, University of Colorado Boulder, Colorado, USA. He is currently an Associate Professor with the Department of Electrical and Computer Engineering, Sultan Qaboos University (SQU), Oman. Before joining SQU, he holds teaching and research positions at several universities, namely the University of Colorado Boulder (UCB), USA, the Queensland University of Technology (QUT), The University of Queensland (UQ), and the University of Western Australia (UWA), Australia. He has published more than 150 publications in international journals and conferences and supervised more than ten Ph.D. students. He was a Lead Guest Editor of a EURASIP special issue on advances in non-stationary electrophysiological signal analysis and processing. His research interests include intelligent control systems and signal processing & their applications.

FAÏÇAL MNIF (Senior Member, IEEE) received the Ph.D. degree from the Ecole Polytechnique de Montréal (University of Montreal), in 1996. He is currently an Associate Professor in control engineering with the Department of Electrical and Computer Engineering, Sultan Qaboos University. His research interest includes control theory and applications and robotics.