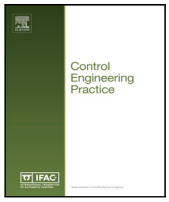
Control Engineering Practice 136 (2023) 105523

Contents lists available at ScienceDirect

Control Engineering Practice

journal homepage: www.elsevier.com/locate/conengprac



Hierarchical motion control strategies for handling interactions of automated vehicles

Balázs Németh *, Péter Gáspár

Institute for Computer Science and Control (SZTAKI), Eötvös Loránd Research Network (ELKH), Kende u. 13-17, H-1111 Budapest, Hungary

ARTICLE INFO

Keywords:
Automated vehicles
Vehicles interactions
Hierarchical control
Learning and control

ABSTRACT

The paper proposes motion control strategies for automated road vehicles to handle interactions among vehicles. The control strategies are built in a hierarchical structure, which contains a high-level learning-based control, a low-level robust control and an optimization-based supervisor. The purposes of the control strategies are to guarantee collision-free motion of the vehicles, and moreover to improve economy and traveling time performance levels. In the paper two configurations of the hierarchical structure are proposed, which are based on the centralized and independent design for automated vehicle motion. The paper presents the comparison of the two configurations through simulation examples. Moreover, the operation of the independent configuration on a small-scaled indoor test vehicle platform is also presented.

1. Introduction and motivation

Handling of automated vehicle interactions, e.g., vehicle following, intersection and roundabout scenarios, are challenging problems in the field of automated vehicle control design. One of the significant challenges is the handling of different types of objectives in the interactions among vehicles. The most important objective is the safe motion of the automated vehicles, e.g., guaranteeing collision-free motion and keeping speed limits. Due to the priority of safety in automated vehicles, these objectives impose hard constraints in the vehicle control strategy, i.e., these objectives must be guaranteed during the entire motion of the vehicles. Moreover, several non-safety performance requirements must be satisfied, e.g., trip time reduction, minimization of energy consumption and maintaining traveling comfort requirements. In the optimization of longitudinal vehicle motion, the non-safety performances form soft constraints, i.e., these objectives might be guaranteed. Moreover, the provided motion control strategy must be solved in real time, which poses the problem of forming low-complexity tasks on the multi-objective optimization.

Several papers have been published to provide some solutions to the previous problem, in the context of automated vehicles. Trajectory planning methods can provide efficient tools to handle interactions of automated vehicles (Katrakazas, Quddus, Chen, & Deka, 2015). An optimum-based method for trajectory generation of road vehicles has been proposed (Werling, Ziegler, Kammel, & Thrun, 2010), with results in the achievements of stopping, following, velocity keeping and merging functionalities. The Bézier curve optimization method has been proposed in Lattarulo and Pérez Rastelli (2021) and Moreau et al. (2019), while in Stahl, Wischnewski, Betz, and Lienkamp (2019) a planning method for high-speed vehicles has been provided, with which an action set of multiple paths has been provided. Focusing on the computation time of the trajectory, real-time solutions through dynamic path planning method (Hu, Chen, Tang, Cao, & He, 2018), or safe stop trajectory planning algorithm (Svensson et al., 2018) have also been proposed. Model Predictive Control solutions have been developed to follow the centerline of the lane, and similarly, the collision with obstacles has been avoided (Carvalho, Gao, Gray, Tseng, & Borrelli, 2013), or in Rosolia and Borrelli (2020) the predictive control solution has been extended with learning features to achieve minimum lap time vehicle motion. Using Pontryagin's minimum principle (Bichiou & Rakha, 2019), it is also possible to handle dynamic and static constraints in the same optimization process for achieving a minimum trip time motion. Reinforcement learning-based methods for trajectory design have been used in Chen et al. (2020), Isele, Rahimi, Cosgun, Subramanian, and Fujimura (2018), Wu, Jiang, and Zhang (2020) and Zhou, Yu, and Qu (2020). Its advantage is that some of these methods are model-free, which can provide solution to the problem of constraint formulation. The enhanced performance level is achieved through a control design method, which is based on a training process with high number of episodes. Although it can lead to promising results, the achieved neural-network-based agents may not guarantee avoidance of vehicles' collision.

In most of the papers, the control strategies have been developed on performing safe vehicle motion in intersections or in roundabouts. Its reason is that these scenarios have some specialties of their own. For example, in these scenarios increased number of vehicles are interacted. Moreover, various types of intersection and roundabouts are existing,

* Corresponding author.
E-mail addresses: balazs.nemeth@sztaki.hu (B. Németh), peter.gaspar@sztaki.hu (P. Gáspár).

https://doi.org/10.1016/j.conengprac.2023.105523
Received 29 June 2022; Received in revised form 20 March 2023; Accepted 4 April 2023
Available online 20 April 2023

i.e., with different number of lanes, with or without signalization, and varying route directions. Furthermore, in urban regions the individual intersections and roundabouts are connected to a network, which connections increase the complexity of developing optimal automated vehicle motion profile. Nevertheless, in these cases the motion of automated vehicles can have impact on the performance level of the traffic system, i.e., on traffic average speed or on density. Paper Yang, Almutairi, and Rakha (2021) has presented an eco-driving system, which is able to provide an energy-optimal motion profile for vehicles at multiple signalized intersections. The proposed system is scalable and can be applied in large networks. A solution to the problem of computational complexity has also been provided in Qian et al. (2019). Another specificity on urban regions is the presence of human-driven vehicles, and the automated vehicles must accommodate to their motions. An optimum in the coordination of automated vehicles might be achieved through a centralized control solution, but it requests data sharing between vehicles, which poses security and privacy challenges. Consequently, decentralized control structures should be designed, in which the autonomous vehicles are controlled individually, and their joint motion is coordinated. For example, in Xu, Xiao, Cassandras, Zhang, and Li (2022) a safe, energy- and time-optimal decentralized Connected and Automated Vehicle (CAV) control structure for handling vehicle motion in multi-lane unsignalized intersections has been proposed. In Chen et al. (2022) also a distributed control solution for CAVs using a graph-based modeling and optimality analysis has been presented.

Some relevant approaches to handle the problem of vehicle ordering in intersections and roundabouts are the game theory (Tian, Li, Kolmanovsky, Yildiz, & Girard, 2020), the queuing theory (Chai, Cai, ShangGuan, & Wang, 2017; Tachet et al., 2016), defining restricted zones (Riegger, Carlander, Lidander, Murgovski, & Sjöberg, 2016) or predictive optimization-based techniques (Morales Medina, Creemers, Lefeber, & van de Wouw, 2020; Németh & Gáspár, 2019). Nevertheless, the motion control for automated vehicles in urban areas can be extended to handle multiple intersections and vehicles. For example, decentralized control for CAVs at multiple adjacent intersections has been developed, see Chalaki and Malikopoulos (2022). A survey on intersection management for heterogeneous connected vehicles is found in Gholamhosseinian and Seitz (2022). Multi-agent approaches to control the vehicles in area of interconnected intersections is found in Dresner and Stone (2008) and Hausknecht, Au, and Stone (2011). The proposed method creates the possibility to reverse lanes under varying traffic conditions. A heuristic optimization method in multi-agent framework has been proposed in paper Zohdy and Rakha (2012). Papers Chen et al. (2020) and Masi, Xu, and Bonnifait (2018) have proposed virtual platooning approaches for autonomous vehicles to guarantee safe crossing. In Kumaravel, Malikopoulos, and Ayyagari (2022) a two-level framework has been presented for the coordination of CAV platoons, which cross unsignalized intersections. By the combination of human temporal behavior and tactical decision-making, Rodrigues, McGordon, Gest, and Marco (2018) has developed an adaptive tactical behavior planner to model human-like motion to be used for the control of autonomous vehicles in different types of roundabouts. Paper Debada, Makarem, and Gillet (2017) has designed a coordination strategy based on virtual vehicles approach that is used for mapping the states of CAVs. A risk model in the control strategy of CAVs for guaranteeing collision avoidance in a probabilistic sense at occluded intersection has been built, see Müller, Strohbeck, Herrmann, and Buchholz (2022).

In addition, several researchers have developed various learning-based methods for the control of automated vehicles to contribute to their safe motion in traffic. In Yan, Welschehold, Büscher, and Burgard (2022) a reinforcement-learning-based control solution has been proposed, which has yielded policy for a centralized controller to let CAVs at unsignalized intersections giving up their right of way. This control solution is able to result in optimized traffic flow. Self-learning integrated decision and control method to handle signalized intersections with mixed traffic flows can be found in Ren et al. (2022). A control algorithm tested in roundabout scenarios has been developed for the motion prediction of vehicles and it contains a dynamic Bayesian network with neural network models for prediction (Mehran & Nasser, 2021). Paper Tollner, Cao, and Zöldy (2019) has built up different neural network architectures for the decision of autonomous vehicles entering a roundabout and passing through it without collision, focusing on the minimization of errors. Papers Chen, Yuan, and Tomizuka (2019a, 2019b) have used learning methods for the control design of autonomous vehicles related to complex urban traffic situations. In these solutions an imitation control framework has been developed to learn driving strategy from offline collected data, and it is used for guaranteeing collision avoidance of autonomous vehicles in roundabouts. On the other hand, model-free deep reinforcement learning methods are built in the control framework of autonomous vehicles driving through roundabouts. Paper Chalaki et al. (2020) has used adversarial multi-agent reinforcement learning to control autonomous vehicles in roundabouts resulting in the reduction of traveling time.

The overview of vehicle control methods for handling vehicle interactions shows that in the recent years there has been increasing interest in the area with various approaches. The first conclusion is that controlling vehicles in roundabouts is a unique research field, but it is requested to find control algorithms with which further maneuvers, e.g., moving in intersections, vehicle following etc., can be carried out. Second, learning-based methods have increasing relevance in the motion control design for handling interactions. Nonetheless, control strategies for providing guarantees on safety performances must be formed in the context of vehicle-vehicle interactions, such as intersections or roundabouts. Third, CAV technologies in the field of roundabouts are promising. It requires an architecture for the connection of vehicles, e.g., cloud-based solutions, while the safe motion of the vehicles during the loss of the connection must also be guaranteed.

This paper provides a novel approach to handle the problems of guaranteed performance requirements for systems with learning-based agents. The motion optimization of automated vehicles through the separation of safety and non-safety performance requirements is carried out. The formulation of a quadratic optimization task is proposed, of which purpose is to guarantee safety performance requirements, i.e., avoidance of vehicle collision and keeping vehicle velocity limitation. For providing an efficient and fast online solution, the quadratic constraints of the optimization are approximated by linear constraints. The provided quadratic optimization task in each time step during the operation of the control is solved. Then, the improvement of the non-safety performances, e.g., reduction of the control energy, through a reinforcement-learning-based training process is achieved. The training process as an offline part of the control design is formed. Nevertheless, under the motion of the automated vehicle, the designed neural network works together with the quadratic optimization task.

In this paper a motion control strategy in a hierarchical structure for automated vehicles is proposed, with which vehicle-vehicle interactions can be handled. The contribution is a novel control strategy, which contains a robust controller and a supervisor on the vehicle level and a learning-based agent on the high level. The output of the hierarchical control is a longitudinal acceleration command of the automated vehicle, with which its safe motion, together with the reductions of traveling time a traction energy can be achieved. Moreover, in this paper centralized and independent configurations for the coordination of multiple automated vehicles have been formed, and the configurations through comparative simulation examples have been evaluated. It is concluded that using the independent configuration the related automated vehicle control problem with high efficiency can be solved and simultaneously, the computational complexity of the centralized configuration can be avoided. Some preliminary results on the centralized configuration can be found in Németh and Gáspár (2021a), and on the independent configuration in Németh and Gáspár (2021b). Nevertheless, a further new contribution of this paper is the

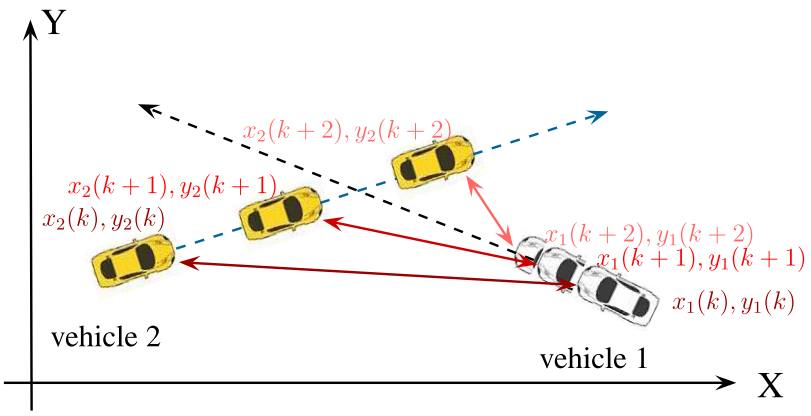


Fig. 1. Illustration of vehicle interaction scheme with Euler coordinates.

implementation of the hierarchical control method on a small-scale indoor test vehicle platform, and thus, the effectiveness of the method through Hardware-in-the-Loop (HiL) simulations is demonstrated.

The paper is organized as follows. The concept of hierarchical control with the proposed configurations is found in Section 2. The formulation of the vehicle interaction model is provided in Section 3. Using of reinforcement-learning-based design process for enhancing non-safety performance level, i.e., reduction of traveling time and control input is proposed in Section 4. The design process in two configurations, such as independent and centralized has been proposed. The comparison of their efficiency is illustrated via comparative simulation examples in Section 5. The implementation of independent configuration through HiL simulations is demonstrated in Section 6. Finally, the conclusions, the limitations and some future challenges of the method in Section 7 are discussed.

2. Concept of hierarchical motion control for automated vehicles

In this section the fundamentals of the proposed motion control are presented. First, the difficulty of vehicle interactions through an example is illustrated. Second, the concept of the design is presented for one automated vehicle, i.e., the interaction model and the control architecture. Third, the configurations for handling the control of multiple automated vehicles are presented.

2.1. Illustration of the vehicle interaction problem

The motivation of finding control design methods for handling interactions is illustrated as follows. Fig. 1 shows an example, in which an automated vehicle (Vehicle 1) and a non-automated disturbance vehicle (Vehicle 2) are in interaction. The motions of both vehicles on these short sections are considered to be straight. In the example it is also assumed that the motion of the disturbance vehicle is not influenced by the automated vehicle, but the motion of the automated vehicle must adapt to the motion of the disturbance vehicle. The goal of vehicle control is to find control inputs on horizon of N time intervals, i.e., longitudinal acceleration command $u(k)\dots u(k+N-1)$, with which the speed of Vehicle 1 $v_1(k+1)\dots v_1(k+N)$ during its journey is maximized, while a predefined safe distance s_{safe} from Vehicle 2 is kept. The condition for keeping s_{safe} is formed by the geometric inequality

$$\Big(x_1(k+t)-x_2(k+t)\Big)^2+\Big(y_1(k+t)-y_2(k+t)\Big)^2 \geq s_{safe}^2, \qquad \forall t\in[1;N], \quad (1)$$

where $x_1(k+t), y_1(k+t)$ are the predicted positions of Vehicle 1. The predictions of both vehicle motions on linear kinematic equations are based. For the prediction of $x_1(k+t), y_1(k+t)$, $t\in[1;N]$, the actual position $x_1(k), y_1(k)$, speed $v_1(k)$ and the control input sequence $u(k)\dots u(k+N-1)$ are used. The motion of Vehicle 2, i.e., positions $x_2(k+t), y_2(k+t)$, through its measured position $x_2(k), y_2(k)$, speed $v_2(k)$ and acceleration $a_2(k)$ values is predicted, together with the limitation of the predicted maximum speed $v_2(k+t)$ through v_{max}.

The goal of control for Vehicle 1 is to maximize its speed, and similarly, the avoidance of collision with Vehicle 2. The maximization requirement in the illustration example as an objective of the control for computing $u(k)\dots u(k+N)$ can be formulated, while the requirement of keeping safe distance forms constraints. The optimization problem is formulated as follows:

$$\max_{u(k)\dots u(k+N-1)} \sum_{i=1}^{N} v_1(k+i), \quad (2a)$$

subject to

$$X_1(k+t)=\mathrm{f}_1(x_1(k), y_1(k), v_1(k), u(k)\dots u(k+t-1)),$$

$$\forall t\in[1;N], \quad (2b)$$

$$X_2(k+t)=\mathrm{f}_2(x_2(k), y_2(k), v_2(k)), \forall t\in[1;N], \quad (2c)$$

$$u(k+i)\in\mathbf{U}, \forall i\in[0;N-1], \quad (2d)$$

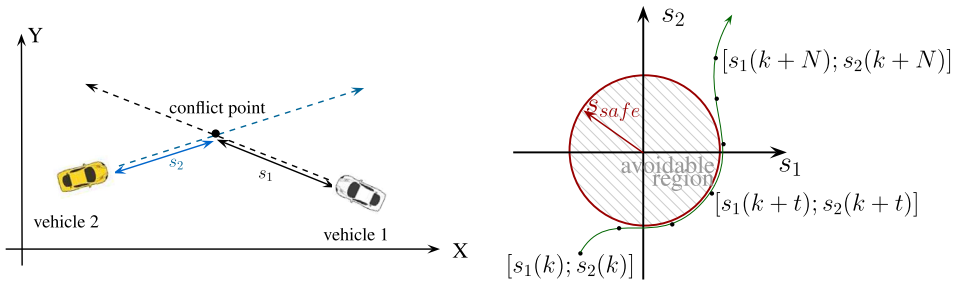
$$\Big(x_1(k+t)-x_2(k+t)\Big)^2+\Big(y_1(k+t)-y_2(k+t)\Big)^2 \geq s_{safe}^2,$$

$$\forall t\in[1;N], \quad (2e)$$

$$0\leq v_1(k+t)\leq v_{max}, \qquad \forall t\in[1;N], \quad (2f)$$

where (2a) is the objective function with the goal of speed maximization, and (2d) formulates the limitation on the control input, inequality constrains keeping of safe distance (2e). Inequality (2f) formulates the constraint on the speed of Vehicle 1, where v_{max} is determined by the path of the vehicle. $X_1(k+t)=\begin{bmatrix}x_1(k+t) & y_1(k+t) & v_1(k+t)\end{bmatrix}^T$, $X_2(k+t)=\begin{bmatrix}x_2(k+t) & y_2(k+t) & v_2(k+t)\end{bmatrix}^T$ incorporates in the position and speed predictions of the vehicles and the kinematic relationships are covered by $\mathrm{f}_1, \mathrm{f}_2$ in (2c). Remark that if Vehicle 2 is also an automated vehicle, the objective function of the maximization problem can also involve $v_2(k+i)$ terms, i.e., $\sum_{i=1}^{N}\Big(v_1(k+i)+v_2(k+i)\Big)$. Moreover, $u_1(k+i)$ and $u_2(k+i)$ control inputs of the vehicles are the variables of the optimization problem.

Formally, (2) is similar to a predictive control formulation. Thus, it might be possible to solve the problem (2) with the tools of the Model Predictive Control (MPC). In a vehicle control context, an enhanced solution is repetitive learning MPC (Rosolia & Borrelli, 2018), with which the optimization can be carried out directly, and terminal cost, terminal set through various test scenarios can be learned. In case of a vehicle interaction problem, it may lead to a high-performance control,



(a) Scheme of the interaction model (b) Illustration of the constraint

Fig. 2. Illustration of the proposed vehicle interaction model.

but the online computation cost during the operation of the vehicle can be high. Moreover, if the vehicle interaction scenario is complex (multiple intersections, roundabouts etc.), the model formulation in f_1, f_2 can be difficult. Therefore, in this paper an effective alternative solution to the optimization problem (2) is proposed.

2.2. Concept of the proposed motion profile design

The proposed motion profile design method is based on the idea that it is possible to form the interaction model of the vehicles without using $x_1(k+t), y_1(k+t), x_2(k+t), y_2(k+t), t \in [1; N]$ directly in the optimization problem. In the following, the reformulation of the task is shown.

First, it is assumed that it is possible to select critical points, where the routes of interacting vehicles cross each other, see Fig. 2(a). In the method the distances of the vehicles from this critical point is used, i.e., $s_1(k+t), s_2(k+t)$, instead of the vehicle coordinates $x_1(k+t), y_1(k+t), x_2(k+t), y_2(k+t)$. Its advantage is that a general formulation of the interaction model can be provided. Thus, the constraint on keeping s_{safe} is independent of the direction of the vehicle motions, because this information in the preliminary selection of conflict point is hidden.

Second, it is also assumed that the conflict point can be determined, and the distances between each vehicle and the conflict point can be measured, which leads to s_1 and s_2 distances in the presented example. Using of s_1, s_2 the problem on keeping s_{safe} is formed as follows, i.e., at the same time only one of the vehicles can be closer to the conflict point than as s_{safe}. This condition describes an avoidable region in the plane of $s_1 - s_2$, of which center is the conflict point. Geometrically, the condition on keeping s_{safe} entails that the trajectory of $s_1(k+t), s_2(k+t)$, $\forall t \in [1; N]$ must be out of a circle with s_{safe} radius, such as

$$\Big(s_1(k+t)\Big)^2 + \Big(s_2(k+t)\Big)^2 \geq s_{safe}^2, \quad \forall t \in [1; N]. \tag{3}$$

The condition is illustrated in Fig. 2(b). Although this concept is close to the approach of restricted zones (Riegger et al., 2016), it has the advantage of using only the terms of s_1, s_2 instead of the Euclidean coordinates x_1, y_1, x_2, y_2, which leads to constraint formulation with reduced complexity. In the concept of restricted zones, the areas of the zones can be defined geometrically, but in this concept the area in the plane of s_1, s_2 is depicted.

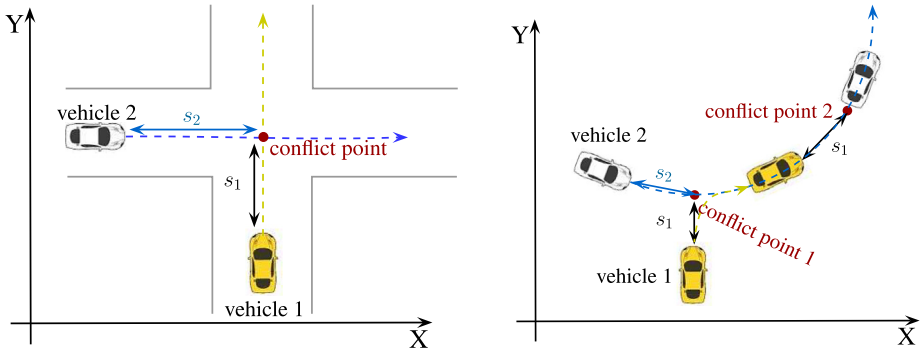
The selection of the conflict point for some relevant interactions, i.e., motion in intersection and roundabout, is illustrated in Fig. 3. In Fig. 3(a) the intersection scenario is shown, where the conflict point is defined by the crossing of the vehicle routes. Motion in roundabouts can be taken as joined vehicle motion in an intersection scenario and in a vehicle following scenario, i.e., safe motion requires the modification of the conflict point during the motion of the vehicle, see Fig. 3(b). Entering a roundabout can be handled as an intersection problem, and thus, the conflict point is defined by the crossing of the routes. In the next phase, the preceding Vehicle 2 by Vehicle 1 is followed and thus, the critical point is determined by the position of Vehicle 2, i.e., $s_2 \equiv 0$ and $s_1(k)$ is the distance between the vehicles.

Note that the limitation of the proposed constraint (3) is the assumption of point mass vehicle models. This simplification has two advantages, which motivate the application of this model. First, in the case of a point mass vehicle model the sizes of the vehicles have no impact on the constraint. As a result, the constraint is valid for all cases, independently of the angle between the vehicle routes. Thus, the optimization method (2) can be used for intersections, roundabouts and also for vehicle following. Second, the optimization method is also independent of the closest points of the vehicles from the conflict point. For example, in Fig. 2, if Vehicle 2 has priority over Vehicle 1, then the front left corner of Vehicle 1 and front right corner of Vehicle 2 are the closest points. Later, the front midline of Vehicle 1 and the right size of Vehicle 2 are the closest points. Finally, the front right corner of Vehicle 1 and the back right corner of Vehicle 2 are the closest to each other. Nevertheless, the real vehicle sizes must be considered to avoid collision, which can be carried out through the increasing of s_{safe}.

In the provided control solution, the elements of (2) to different control levels are separated. The maximization of the objective function through reinforcement learning on a high number of roundabout scenarios has been achieved. It has been carried out offline and during control operation only the resulted neural network has been used. The output of the RL-based controller $u_L(k)$ is considered to be a candidate control input of the system. Vehicle dynamics in the design of a robust control has been incorporated, which results in another candidate control input $u_K(k)$. In this way, the complex optimization problem (2) is converted to a simplified optimization, which has been built in a supervisor. The goal of the supervisor is to create $u(k)$ using $u_L(k), u_K(k)$, which $u(k)$ guarantees keeping safe distance.

In Fig. 4 the scheme of the vehicle control architecture is illustrated. The output of the automated vehicle control is $u(k)$, which in this paper is selected as longitudinal acceleration command $a_1(k)$. Depending on the complexity of the vehicle model, it is also possible to select $u(k)$ as traction force, see e.g., Németh, Gáspár, and Szabó (2021). The supervisor calculates the value of $u(k)$ using the expression $u(k) = u_K(k) + \Delta(k)$. In this form $u_K(k)$ represents the robust controller input, which has been implemented on the vehicle level. $\Delta(k)$ with its bounded domain $\mathbf{\Delta}$ is handled as an addition to $u_K(k)$. Moreover, the output of the RL-based controller is $u_L(k)$, which is an input of the supervisor. During the control operation $\Delta(k)$ is computed to minimize $(u(k) - u_L(k))^2$ difference, and simultaneously, to avoid the collision of the vehicle with another vehicles.

It is necessary to distinguish the roles of the RL-based and the robust controllers. A comprehensive introduction to the design can be found in Németh and Gáspár (2021). First, the role of the supervisor is to provide $u(k)$ control signal, with which the collision avoidance for all scenarios is guaranteed. In this design process two candidate control signals, such as $u_K(k)$ from the robust controller and $u_L(k)$ from the RL-based controller are used. The formulation of the supervisor in Section 3 can be found. Second, the robust controller is designed to provide $u_K(k)$, for which $\Delta(k)$ under all scenarios can be found. It requires the incorporation of the domain $\mathbf{\Delta}$ in the design of the robust controller,



(a) Motion in intersection (b) Motion in roundabout

Fig. 3. Illustration of conflict point selection.

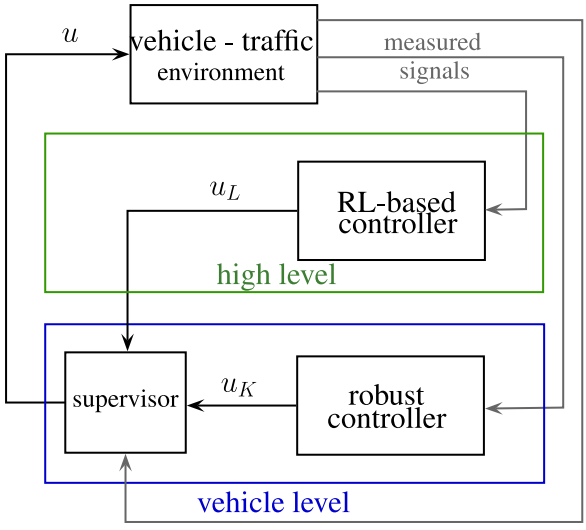


Fig. 4. Illustration of the control architecture.

which is interpreted as an input uncertainty domain in the context of the robust control, for details see Németh and Gáspár (2021b). Third, the role of the RL-based controller is to provide $u_L(k)$ candidate control input, with which non-safety performance requirements can be considered. Since the RL-based controller through a training process with a high number of varying episodes is required, $u_L(k)$ may be unable to guarantee collision avoidance for all scenarios in itself. The formulation of the RL-based controller in Section 4 can be found.

2.3. Multiple vehicle control configuration schemes

In the case of multiple vehicle scenarios, in which automated vehicles operate simultaneously, the proposed control scheme through various configurations can be extended. The possible configurations are influenced by the centralized or individual handling of automated vehicles. Two main configuration schemes for the coordination of automated vehicles in their interactions are illustrated in Fig. 5. In case of centralized method, the control inputs for n vehicles are computed with one optimization algorithm, while in case of independent method n optimization algorithms, i.e., one for each vehicle, are used.

Fig. 5(a) illustrates a centralized control scheme, which consists of one centralized multiple-input multiple-output learning-based agent on the high level for all vehicles and one centralized safe coordination level. The control inputs, i.e., longitudinal accelerations for the vehicles a_i, are transferred from the centralized safe coordination level. The advantage of this configuration is that motions of the vehicles in centralized layers are incorporated and thus, optimal control inputs on the global transportation system can be achieved. Nevertheless, a drawback of this configuration is that it requires centralized computation on the high level and on the coordination level, which can be difficult from an application point of view, e.g., all information must be simultaneously available, or data-privacy problems may also occur. This structure is recommended to be used especially for vehicle fleets, in which all of the vehicles are in the fleet of the same operator.

Another configuration is shown in Fig. 5(b), in which each vehicle connects to its own high level control, and each vehicle has its own safe coordination level. As a result, it requires the sharing of limited information exchange between the vehicles, i.e., their positions and longitudinal speed, which facilitates the practical implementation of the control strategy. The advantage of this structure is that the number of vehicles is not limited by the computation capacity, because the computation tasks of each vehicle is independent. Due to the independent computation the coordination of the vehicles, e.g., minimum energy consumption on the global level, is not guaranteed.

In this paper the control design methods for both configuration schemes are presented. The effectiveness of the configurations through comparative scenarios from control engineering viewpoint is discussed. In spite of the results, both schemes can have their own application fields, and thus, the decision on the configuration requests careful considerations.

3. Model formulation for handling vehicle interactions

The formulation of the model for achieving collision avoidance in the vehicle interactions requires the involving of longitudinal vehicle models and the forming of constraints on the motion. In this paper the motions of the vehicles by linear kinematic relationships are described:

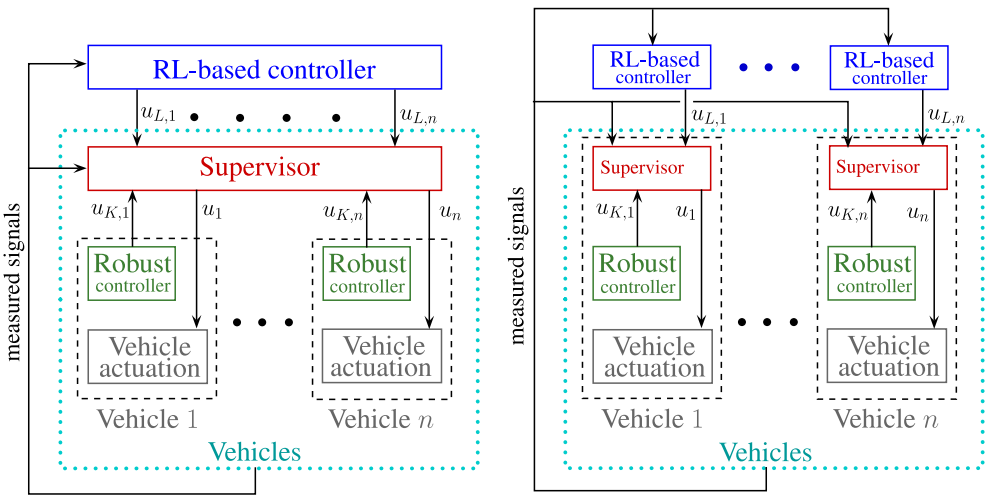
$$v_i(k+1) = v_i(k) + T a_i(k), \tag{4a}$$

$$s_i(k+1) = s_i(k) + T v_i(k) + \frac{T^2}{2} a_i(k), \tag{4b}$$

where T is time step of the discrete motion model. Index $_i$ is related to each vehicle and the total vehicle number is n. The longitudinal velocity and the displacement of vehicle i are $v_i(k)$ and $s_i(k)$. The control input of the vehicle is the longitudinal acceleration, such as $a_i(k) = u_i(k)$. In case of this model, s_i from the center of intersection is measured, which means that $s_i = 0$ represents the position of vehicle i in the center. Consequently, $s_i < 0$ means that vehicle i approaches to the intersection, and $s_i > 0$ represents that vehicle i moves toward.

The control input contains two terms (see Section 1):

$$a_i(k) = a_{K,i}(k) + \Delta_i(k), \tag{5}$$



(a) Centralized configuration (b) Configuration with independent vehicles

Fig. 5. Configuration schemes in the hierarchy structure.

where $a_{K,i} = u_{K,i}$ is the output of the robust controller of vehicle i, and the supervisor computes the additional term $\Delta_i(k)$. The computation of $\Delta_i(k)$ is based on the following principles.

- The objective of the computation is to provide $u_i(k) = a_i(k)$, which is as close as possible to $u_{L,i}(k) = a_{L,i}(k)$, i.e., their difference must be reduced. It reason is to approximate the level of non-safety performances, which is achieved through the RL-based controller.
- Since in the intersection the routes of some vehicles may be crossed, $a_i(k)$ must be selected for vehicle i to keep safe distance s_{safe} from the further vehicles. It forms a constraint for the selection process of Δ_i.
- The resulted $a_i(k)$ must guarantee keeping velocity limits, i.e., the maximum velocity is constrained by velocity limit v_{max} or the route curvature, which forms another constraint.

3.1. Constraint formulation

In Section 2 the concept of constraint formulation on keeping safe distance s_{safe} has already been introduced. It has been shown that quadratic constraint formulation for each interacting vehicle pair is able to guarantee collision avoidance. Although quadratic constraint provides an accurate description, it can be disadvantageous in real-time optimization problems due to increasing computational effort. It is motivated to find an approximation of the quadratic constraints, i.e., using linear constraints.

The idea behind the approximation in Fig. 6 is shown. The avoidable regions, i.e., half-planes are defined by tangent lines, which connect the actual state $\begin{bmatrix} s_i(k) & s_j(k) \end{bmatrix}^T$ and the circle. The tangent points on the circle in k is noted by $\begin{bmatrix} s_{T1,i}(k) & s_{T1,j}(k) \end{bmatrix}$, $\begin{bmatrix} s_{T2,i}(k) & s_{T2,j}(k) \end{bmatrix}$. The condition on the next point of the trajectory $\begin{bmatrix} s_i(k+1) & s_j(k+1) \end{bmatrix}^T$ is to be out of the half-plane, which forms two linear inequality constraints:

$$\begin{bmatrix} s_{T1,i}(k) \\ s_{T1,j}(k) \end{bmatrix}^T \begin{bmatrix} s_i(k) \\ s_j(k) \end{bmatrix} \leq \begin{bmatrix} s_{T1,i}(k) \\ s_{T1,j}(k) \end{bmatrix}^T \begin{bmatrix} s_i(k+1) \\ s_j(k+1) \end{bmatrix}, \tag{6a}$$

or

$$\begin{bmatrix} s_{T2,i}(k) \\ s_{T2,j}(k) \end{bmatrix}^T \begin{bmatrix} s_i(k) \\ s_j(k) \end{bmatrix} \geq \begin{bmatrix} s_{T2,i}(k) \\ s_{T2,j}(k) \end{bmatrix}^T \begin{bmatrix} s_i(k+1) \\ s_j(k+1) \end{bmatrix}. \tag{6b}$$

Note that this formulation leads to an outer approximation of the avoidable region. Although linear approximation leads to a simplified formulation, the outer approximation can result in increased conservativeness, i.e., there can be states in the avoidable region, which

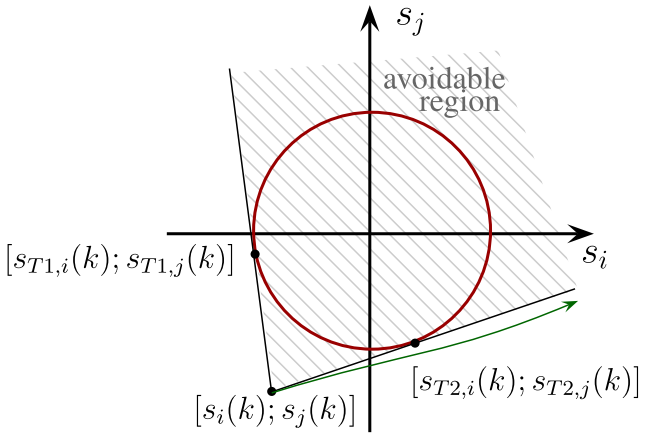


Fig. 6. Illustration on the approximation of the constraint.

are out of the circle. Moreover, Fig. 6 shows that the non-avoidable, i.e., reachable region is non-convex. The non-convex property of the reachable region results in the formulation of disjunctive inequalities (8).

In the next step, the longitudinal displacement at $k+1$ is expressed based on the motion Eq. (4) and using the expression on control input terms (5) leads to:

$$s_i(k+1) = s_i(k) + Tv_i(k) + \frac{T^2}{2}a_{K,i}(k) + \frac{T^2}{2}\Delta_i(k), \tag{7a}$$

$$s_j(k+1) = s_j(k) + Tv_j(k) + \frac{T^2}{2}a_{K,j}(k) + \frac{T^2}{2}\Delta_j(k), \tag{7b}$$

which transform the linear inequalities (6) to

$$\begin{bmatrix} s_{T1,i}(k) \\ s_{T1,j}(k) \end{bmatrix}^T \begin{bmatrix} -Tv_i(k) - \frac{T^2}{2}a_{K,i}(k) \\ -Tv_j(k) - \frac{T^2}{2}a_{K,j}(k) \end{bmatrix} \leq \frac{T^2}{2} \begin{bmatrix} s_{T1,i}(k) \\ s_{T1,j}(k) \end{bmatrix}^T \begin{bmatrix} \Delta_i(k) \\ \Delta_j(k) \end{bmatrix}, \tag{8a}$$

or

$$\begin{bmatrix} s_{T2,i}(k) \\ s_{T2,j}(k) \end{bmatrix}^T \begin{bmatrix} -Tv_i(k) - \frac{T^2}{2}a_{K,i}(k) \\ -Tv_j(k) - \frac{T^2}{2}a_{K,j}(k) \end{bmatrix} \geq \frac{T^2}{2} \begin{bmatrix} s_{T2,i}(k) \\ s_{T2,j}(k) \end{bmatrix}^T \begin{bmatrix} \Delta_i(k) \\ \Delta_j(k) \end{bmatrix}. \tag{8b}$$

A further constraint on the computation of $\Delta_i(k)$ is related to vehicle velocities. Constraint on $v_i(k+1)$ is formed through motion Eq. (4),

together with (5). The constraint on velocity limitation is formed as

$$0 \le v_i(k) + T a_{K,i}(k) + T\Delta_i(k), \qquad \forall i \in n, \tag{9a}$$

$$v_{max} \ge v_i(k) + T a_{K,i}(k) + T\Delta_i(k), \qquad \forall i \in n, \tag{9b}$$

which leads to

$$\begin{bmatrix} -1 \\ 1 \end{bmatrix} \Delta_i(k) \le \begin{bmatrix} \frac{v_i(k)}{T} + a_{K,i}(k) \\ \frac{v_{max} - v_i(k)}{T} - a_{K,i}(k) \end{bmatrix}, \quad \forall i \in n. \tag{10}$$

The last constraint in the computation of $\Delta(k)$ is resulted by the limitation on the achievable longitudinal acceleration, such as the bounds $a_{min,i}, a_{max,i}$, which are related to the physical limits on braking and driving. Since $a_i(k)$ depends on $a_{K,i}(k)$ and $\Delta_i(k)$, the constraints are

$$a_{min,i} - a_{K,i}(k) \le \Delta_i(k), \qquad \forall i \in n, \tag{11a}$$

$$a_{max,i} - a_{K,i}(k) \ge \Delta_i(k), \qquad \forall i \in n, \tag{11b}$$

which leads to the constraint

$$\begin{bmatrix} -1 \\ 1 \end{bmatrix} \Delta_i(k) \le \begin{bmatrix} a_{K,i}(k) - a_{min,i} \\ a_{max,i} - a_{K,i}(k) \end{bmatrix}, \qquad \forall i \in n. \tag{12}$$

Centralized configuration

In case of centralized configuration, one optimization problem for the computation of all $\Delta_i(k)$, $i \in n$ is formed as follows:

$$\min_{\Delta_1(k) \dots \Delta_n(k)} \sum_{i=1}^{n} \left(a_i(k) - a_{L,i}(k) \right)^2 \tag{13a}$$

subject to

$$s_i(k+1)^2 + s_j(k+1)^2 \ge s_{safe}^2, \qquad \forall i,j \in n, \tag{13b}$$

$$0 \le v_i(k+1) \le v_{max}, \qquad \forall i \in n, \tag{13c}$$

$$a_{min,i} - a_{K,i}(k) \le \Delta_i(k), \qquad \forall i \in n, \tag{13d}$$

$$a_{max,i} - a_{K,i}(k) \ge \Delta_i(k), \qquad \forall i \in n, \tag{13e}$$

$$\Delta_i \in \mathbf{\Delta}_i, \qquad \forall i \in n, \tag{13f}$$

where i, j represent the pairs of interacting vehicles. $\mathbf{\Delta}_i$ represents the domain on Δ_i.

The objective (13a) can be reformulated through (5), which leads to

$$\sum_{i=1}^{n} \left(a_i(k) - a_{L,i}(k) \right)^2 = \sum_{i=1}^{n} \left(a_{K,i}(k) + \Delta_i(k) - a_{L,i}(k) \right)^2 = \Delta(k)^T I_{n\times n} \Delta(k) + 2 f^T \Delta(k), \tag{14}$$

where $\Delta(k) = \left[\Delta_1(k) \dots \Delta_n(k) \right]^T$ and $I_{n\times n}$ is an identity matrix. The vector of f is formed as $f = \left[a_{K,1} - a_{l,1} \dots a_{K,n} - a_{l,n} \right]$. Thus, the optimization (13) for computing $\Delta_1(k) \dots \Delta_n(k)$ through relations (14), (8), (10) and (12) is transformed to

$$\min_{\Delta(k)} \ \Delta(k)^T I_{n\times n} \Delta(k) + 2 f^T \Delta(k) \tag{15a}$$

subject to

$$\begin{bmatrix} -1 \\ 1 \end{bmatrix} \Delta_i(k) \le \begin{bmatrix} \frac{v_i(k)}{T} + a_{K,i}(k) \\ \frac{v_{max} - v_i(k)}{T} - a_{K,i}(k) \end{bmatrix}, \quad \forall i \in n, \tag{15b}$$

and

$$\begin{bmatrix} -1 \\ 1 \end{bmatrix} \Delta_i(k) \le \begin{bmatrix} a_{K,i}(k) - a_{min,i} \\ a_{max,i} - a_{K,i}(k) \end{bmatrix}, \quad \forall i \in n, \tag{15c}$$

and

$$\Delta_i \in \mathbf{\Delta}_i, \qquad \forall i \in n, \tag{15d}$$

and

$$\begin{bmatrix} s_{T1,i}(k) \\ s_{T1,j}(k) \end{bmatrix}^T \begin{bmatrix} -T v_i(k) - \frac{T^2}{2} a_{K,i}(k) \\ -T v_j(k) - \frac{T^2}{2} a_{K,j}(k) \end{bmatrix} \le \frac{T^2}{2} \begin{bmatrix} s_{T1,i}(k) \\ s_{T1,j}(k) \end{bmatrix}^T \begin{bmatrix} \Delta_i(k) \\ \Delta_j(k) \end{bmatrix}, \quad \forall i,j \in n, \tag{15e}$$

or

$$\begin{bmatrix} s_{T2,i}(k) \\ s_{T2,j}(k) \end{bmatrix}^T \begin{bmatrix} -T v_i(k) - \frac{T^2}{2} a_{K,i}(k) \\ -T v_j(k) - \frac{T^2}{2} a_{K,j}(k) \end{bmatrix} \ge \frac{T^2}{2} \begin{bmatrix} s_{T2,i}(k) \\ s_{T2,j}(k) \end{bmatrix}^T \begin{bmatrix} \Delta_i(k) \\ \Delta_j(k) \end{bmatrix}, \quad \forall i,j \in n. \tag{15f}$$

The quadratic optimization in (15), due to the non-convex reachable regions, incorporates in disjunctive inequalities. Consequently, the presented optimization problem is a mixed-integer optimization problem (Belotti, Liberti, Lodi, Nannicini, & Tramontani, 2011). Nevertheless, it can be transformed to a set of quadratic optimization tasks, see at independent configuration below.

Independent configuration

In the case of the configuration with independent automated vehicles, the following optimization problem is formed for all automated vehicles:

$$\min_{\Delta(k)} \ \Delta(k)^2 + 2 f^T \Delta(k) \tag{16a}$$

subject to

$$\begin{bmatrix} -1 \\ 1 \end{bmatrix} \Delta(k) \le \begin{bmatrix} \frac{v_1(k)}{T} + a_K(k) \\ \frac{v_{max} - v_1(k)}{T} - a_K(k) \end{bmatrix} \tag{16b}$$

and

$$\begin{bmatrix} -1 \\ 1 \end{bmatrix} \Delta(k) \le \begin{bmatrix} a_K(k) - a_{min} \\ a_{max} - a_K(k) \end{bmatrix} \tag{16c}$$

and

$$\begin{bmatrix} s_{T1,1}(k) \\ s_{T1,j}(k) \end{bmatrix}^T \begin{bmatrix} -T v_1(k) - \frac{T^2}{2} a_K(k) \\ -T v_j(k) \end{bmatrix} \le \frac{T^2}{2} \begin{bmatrix} s_{T1,1}(k) \\ 0 \end{bmatrix}^T \Delta(k), \quad \forall j \in n_s \tag{16d}$$

or

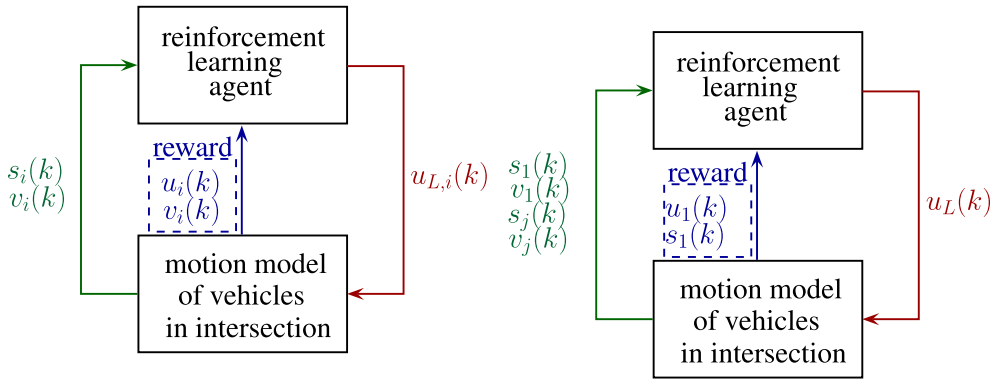
$$\begin{bmatrix} s_{T2,1}(k) \\ s_{T2,j}(k) \end{bmatrix}^T \begin{bmatrix} -T v_1(k) - \frac{T^2}{2} a_K(k) \\ -T v_j(k) \end{bmatrix} \ge \frac{T^2}{2} \begin{bmatrix} s_{T2,1}(k) \\ 0 \end{bmatrix}^T \Delta(k), \quad \forall j \in n_s, \tag{16e}$$

where n_s is the number of considered surrounding vehicles in the interaction scenario. The formed optimization problem (16) differs from the optimization problem of the centralized configuration (15). The main difference is formed in the objective, because it is formulated to only one vehicle, i.e., each vehicle has its own objective. Moreover, the constraints have the same mathematical structure, but their number is reduced, because the constraints are formulated not for n number of vehicles, but for n_s number of surrounding vehicles (Németh & Gáspár, 2021b). Consequently, the solution of (16) requires reduced computation effort, compared to (15).

Nonetheless, the mixed-integer optimization task (16) can be transformed to a set of quadratic optimization tasks. Its reason is that in the context of supervisor, the optimization problem has always the same structure, i.e., n_s leads to 2^{n_s} number of distinct quadratic optimization tasks, which tasks differ only in one of their constraints. Each of the $\Delta^{*m}(k), m \in [1; 2^{n_s}]$ solutions of the quadratic optimization tasks result in costs $\Delta^{*m}(k)^2 + 2 f^T \Delta^{*m}(k), m \in [1; 2^{n_s}]$. The solution of the mixed-integer optimization problem is $\Delta^{*m}(k)$, which results in the lowest cost from the set. Consequently, the online solution of the mixed-integer optimization task can be replaced with a set of fast quadratic optimization tasks and a selection process after that.

4. Design of motion profile using reinforcement learning

The previous parts of the paper have provided the formulation of the supervisor, which is responsible to guarantee safe performance requirements. In this section the design of the RL-based controller for each



(a) Centralized configuration (b) Independent configuration

Fig. 7. Control structure in the learning process for different configurations.

configuration is presented, with which the non-safety performance level of the vehicle control is improved. First, the learning-based process is formed to the centralized configuration and second, it is reformulated to the independent configuration.

The aim of the method is to design a neural-network-based agent, which provides candidate control signals $a_{L,i}(k)$ for improving non-safety performance level, such as minimization of $a_i(k)$ and of traveling time. The agent through the goal-directed reinforcement learning method is trained. Through the learning process the agent is trained to perform a task by interacting with an dynamic environment. This environment can be unknown, e.g., in case of the related control problem the number of vehicles, their distances etc. can vary. This learning method enables the agent to make decisions to maximize the expected cumulative reward for the task, where the reward function must be predefined. A survey on reinforcement learning and on its application to autonomous vehicles can be found in Kiran et al. (2022).

The scheme of the training process related to the centralized configuration is shown in Fig. 7(a). The model for learning incorporates in the supervisor within (15), the robust controller (see Németh & Gáspár, 2021a, 2021b for its design process) and the vehicle motion equations. This model is able to guarantee collision avoidance in each time step, independently from the values of $a_{L,i}(k)$ on each vehicle. Thus, due to the supervisor, the training process do not provide $a_{L,i}$ signals, which signals result in collision. During the learning process the safe vehicle motion, together with the enhancement of the non-safety performance level, is simultaneously guaranteed. The $r(k)$ reward for the learning process is determined by the values of $a_i(k)$ and $v_i(k)$, such as

$$r(k) = -Q_1 \sum_{i=1}^{n} a_i^2(k) + Q_2 \sum_{i=1}^{n} v_i(k), \tag{17}$$

in which Q_1, Q_2 are design weights with positive values. The purpose of Q_1 is to highlight the reduction of control input, such as the reduction of traction energy. The reason of negative sign is that through the reduction of $a_i^2(k)$ the value of $r(k)$ must be increased. The goal of Q_2 is to achieve increased velocity profile for the vehicle, i.e., to facilitate the reduction of traveling time. Through the selection of Q_1, Q_2 a balance between traction energy and traveling time reductions can be achieved. Selection of high value for Q_1 and low value for Q_2 can lead to $a_i(k) = 0$, i.e., to unacceptable slow motion for the vehicles. Similarly, high value for Q_2 and low value for Q_1 can lead to $a_i(k) = a_{max,i}$, i.e., to unacceptable high traction energy consumption. The appropriate selection of Q_1, Q_2 requires experimental tuning through the control designer. The observation for the agent contains the positions of the vehicles $s_i(k)$ and their velocities $v_i(k)$.

The training in the reinforcement learning through performing episodes is carried out. The goal of the training is to maximize the reward (17), which maximization through the evaluation of the simulation results in each episode is achieved. In the training method of this paper, the model-free, online and off-policy deep deterministic policy gradient (DDPG) has been applied, see Lillicrap et al. (2016). The resulted DDPG agent in the structure of an actor–critic relationship, with approximators is developed. During the training an optimal policy is calculated, which is able to maximize the reward on a long-term, where finding an optimal policy is equal to the learning of Bellman equation. The actor and critic approximators have observations, which are noted by S. The goal of the approximator on actor $\mu(S)$ is to find an action A, i.e., $a_{L,i}(k)$, with which the long-term reward is maximized. The purpose of critic $Q(S, A)$ element is to approximate the expected value of the long-term reward.

The training results in a neural network, whose output signals are the candidate control inputs, such as $a_{L,i}(k)$. The values of these signals in each k step are the inputs of the supervisor, see (15), and thus, the non-safety performance level is enhanced.

In the independent configuration (see Fig. 7(b)), the reward function $r(k)$ slightly differs from (17). For each vehicle it also contains $a_1(k), v_1(k)$, such as:

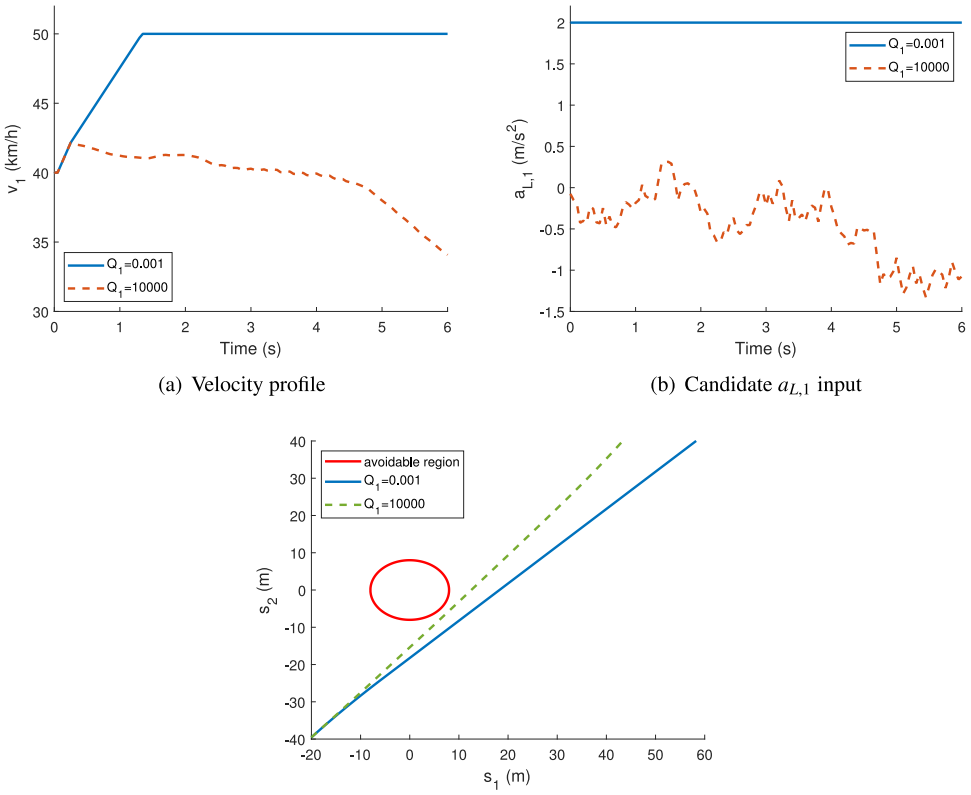
$$r(k) = -Q_1 a_1^2(k) + Q_2 v_1(k). \tag{18}$$

The reward contains the control input $a(k)$ and the longitudinal velocity of the automated vehicle $v_1(k)$. The reason of using these signals is the same as detailed in reward of the centralized configuration, i.e., to provide balance between control input reduction and velocity maximization.

Illustration on the impact of design weights

The impact of design weights on the performances of the control system through simulation examples is illustrated. In these example the interactions of two vehicles in an intersection are examined, where one of the vehicle is automated and the another is a non-automated with constant velocity.

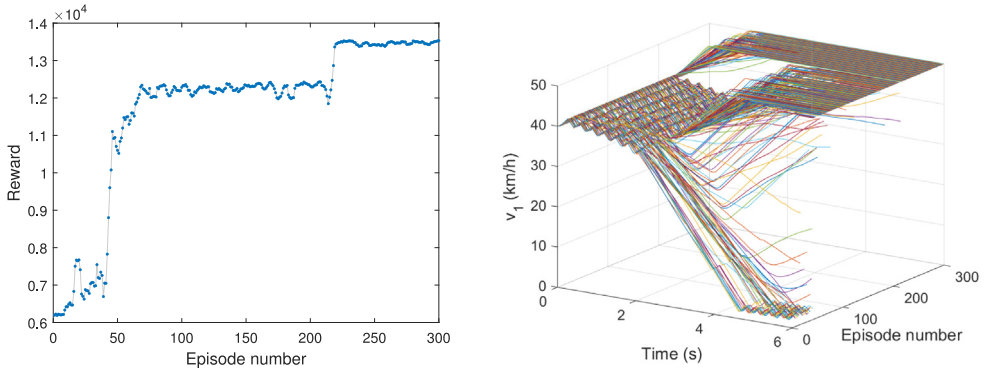
In the first example initial conditions $s_1(0) = -20$ m, $v_1(0) = 40$ km/h for the automated vehicle and $s_2(0) = -40$ m, $v_2(0) = 50$ km/h for the non-automated vehicle are selected. Thus, in this example the automated vehicle is able to leave the intersection first easily, i.e., it is possible to reduce its velocity without colliding to the non-automated vehicle. In this example the results with two scenarios with differently trained agents can be found, i.e., agents with $Q_1 = 0.001, Q_2 = 1$ and $Q_1 = 10\,000, Q_2 = 1$ selections. In case of $Q_1 = 0.001$ selection the high velocity motion of the automated vehicle is facilitated, while in case of $Q_1 = 10\,000$ the reduction of control intervention is highlighted. Fig. 8(a),(b) show the differences in the characteristics of v_1 and a_1 signals, which results fit to prior expectations. Nevertheless, both $s_1 - s_2$ trajectories are out of the avoidable region, i.e., the selection of Q_1, Q_2 does not influence the safe motion of the automated vehicle, see Fig. 8(c).



(a) Velocity profile (b) Candidate $a_{L,1}$ input

(c) Vehicle positions and avoidable region

Fig. 8. Comparison on the impact of different agents in Example 1.



(a) Maximization of reward function (b) Improvement of velocity profile

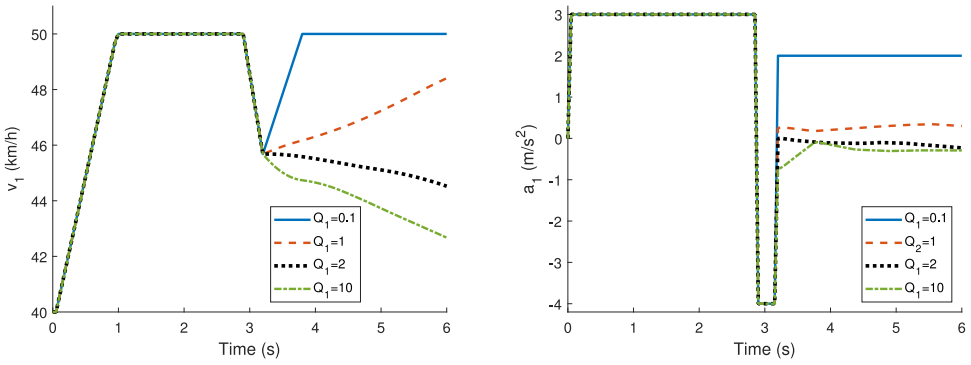
Fig. 9. Illustration on the learning process at $Q_1 = 0.001$ in Example 1.

The illustration on some results of the learning process at $Q_1 = 0.001$ can be found in Fig. 9. The results of the maximization process on the cumulative reward can be seen in Fig. 9(a). The improvement of v_1 characteristics during the learning process is illustrated in Fig. 9(b). It can be seen that the increasing of cumulative reward and the improvement of v_1 are in strong relationship, which is achieved through low Q_1 selection.

Fig. 10 illustrates the v_1, a_1 signals of Example 2. In these scenarios the initial conditions are modified to $s_1(0) = -40$ m, $v_1(0) = 40$ km/h and $s_2(0) = -40$ m, $v_2(0) = 40$ km/h, i.e., there is a conflict situation. Since the motion of the non-automated vehicle is not influenced, i.e., it accelerates to 50 km/h, the avoidance of the collision through the braking of the automated vehicle can be guaranteed, see Fig. 10(b) around 3 s. In these scenarios the impacts of various Q_1 selections on the signals are shown. Fig. 10(a) shows that the velocity profile after the braking significantly differs, depending on Q_1. Similar to the consequences of Example 1, low Q_1 results in high velocity and high Q_1 leads to reduced a_1. Moreover, the value of velocity reduction through Q_1 can be effectively set, see e.g., the results of $Q_1 = 1, Q_1 = 2$ selections between the another two Q_1 values. Thus, the selection of Q_1, Q_2 values for achieving the required performances through simulations can be carried out.

5. Comparison on the effectiveness of the configurations

In this section the effectiveness of the proposed centralized and independent configurations through simulation examples is illustrated. The example contains three automated vehicles, see Fig. 11. The goal of this section is to compare the effectiveness of the two configurations through simulation examples. Two scenarios at each of the



(a) Velocity profile (b) Candidate acceleration

Fig. 10. Comparison of v_1, a_1 signals in Example 2.

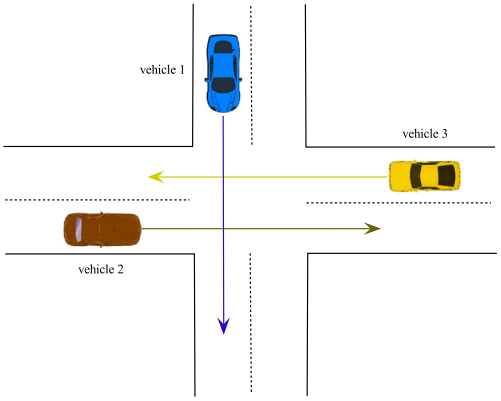


Fig. 11. Illustration on the comparative example.

configurations are examined. The initial positions of the vehicles are the same in both scenarios, such as $s_1(0) = -12$ m, $s_2(0) = -5$ m, $s_3(0) = -18$ m. The safety distance is selected to $s_{safe} = 8$ m, the input constraints are $a_{min,i} = -4$ m/s^2 and $a_{max,i} = 3$ m/s^2 for all vehicles. Sampling time is selected to $T = 0.05$ s. In the first scenario the initial velocities of the vehicles are $v_1(0) = 5$ m/s, $v_2(0) = 4$ m/s, $v_3(0) = 4$ m/s, which are modified to $v_1(0) = 5$ m/s, $v_2(0) = 4$ m/s, $v_3(0) = 6$ m/s in the second scenario.

5.1. Centralized configuration

In the examples of centralized configuration, the outputs of the RL-based controller are $a_{L,1}(k)$, $a_{L,2}(k)$, $a_{L,3}(k)$ and its observation contains the signals $s_1(k)$, $s_2(k)$, $s_3(k)$, $v_1(k)$, $v_2(k)$, $v_3(k)$, see Fig. 7(a). During the training process of the RL-based agent, the initial values of the vehicle states $(s_i(0), v_i(0))$ for the intersection scenarios in each of the episodes are generated randomly: $s_i(0)$ can vary between -10 m and -20 m, and $v_i(0)$ is between 0 km/h and 50 km/h. The actor network has 6 neurons in the input layer, 3 fully connected layers with 48 neurons and ReLu functions in each layer and 3 neurons with hyperbolic tangent functions in the output layer. The critic network has the same structure, but it also contains the actions as inputs. The sampling time in each episode is selected to $T = 0.05$ s and 500 episodes are carried out. The terms in the reward function are considered with the same design parameters, such as $Q_1 = Q_2 = 0.1$. The achieved value of the reward at the end of the training process is above 400.

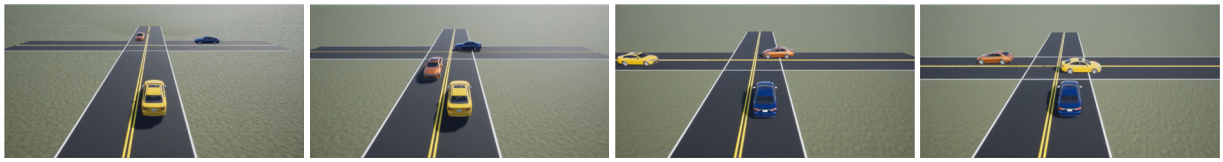
The scenarios with different initial values result in different control input and ordering in the intersection. The motions of the vehicles for each scenario are illustrated in Fig. 12. In both scenarios $Vehicle$ 2 reaches the intersection, but the ordering of the further vehicles is not the same in the two scenarios. In the first Scenario $Vehicle$ 1 is the second (see Fig. 12(b)), while in the second Scenario $Vehicle$ 3 is the second in the ordering (see Fig. 12(d)).

The keeping of s_{safe} is illustrated in Fig. 13 for all constraints and scenarios. Fig. 13(a) shows that the trajectories of the vehicles are close to the border of the avoidable region, but in case of the constraint on s_1 and s_3 increased distance between the vehicles are achieved, see Fig. 13(b). But, in the Scenario 2 both trajectories are close to the borders, see Fig. 13(c)–(d). These differences in the scenarios can also be seen in Fig. 12(b) and (d). In Scenario 1 $Vehicle$ 3 is far from $Vehicle$ 1 in the conflict situation (Fig. 12(b)), while in Scenario 2 the vehicles are close to each other (Fig. 12(b)). Moreover, the differences in the directions of the trajectories show the different ordering of $Vehicle$ 2 and $Vehicle$ 3. In Scenario 2 s_3 is significantly increased, which means that $Vehicle$ 3 reached the intersection earlier as $Vehicle$ 1.

The control input signals a_i and the output of the agents $a_{L,i}$ are shown in Fig. 14. It can be seen that all a_i signal keep the input constraints, which underlines the effectiveness of the robust longitudinal controller. Moreover, it can be seen that a_i and $a_{L,i}$ have the same values in most of the simulations. The signals of $a_{L,i}$ are generally overridden at the beginning of the simulations. For example, in Scenario 2 a_3 is slightly increased and a_2 is reduced to generate braking, with which the avoidance of the collision between $Vehicle$ 2 and $Vehicle$ 3 is achieved. Since $a_{L,i}$ signals are rarely overridden, the improved economy performance of the learning agent is preserved.

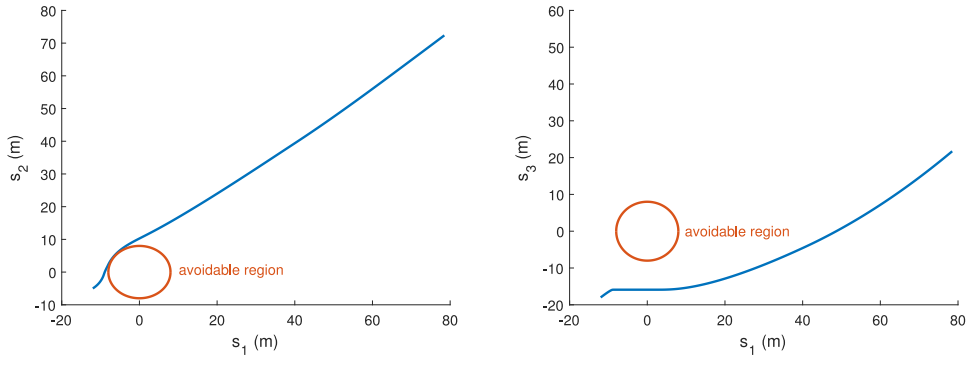
5.2. Independent configuration

In the rest of this section the effectiveness of the hierarchical control with independently controlled automated vehicles is presented, i.e., the

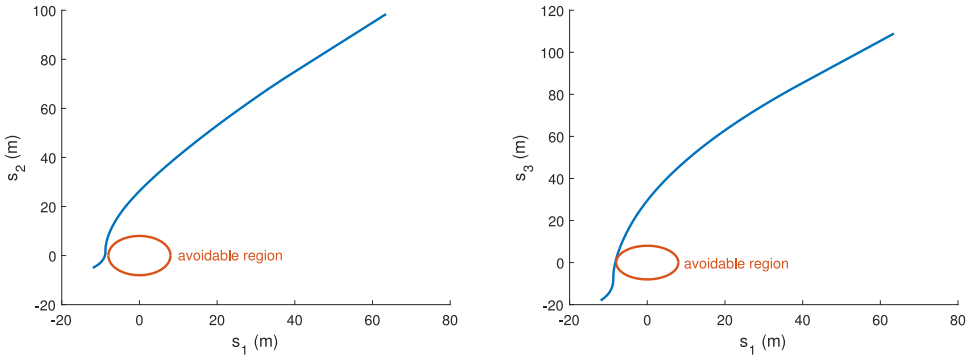


(a) Initial positions (Scenario 1) (b) Conflict situation (Scenario 1) (c) Initial positions (Scenario 2) (d) Conflict situation (Scenario 2)

Fig. 12. Illustration of the intersection scenarios (from the viewpoint of the vehicle with last ordering).

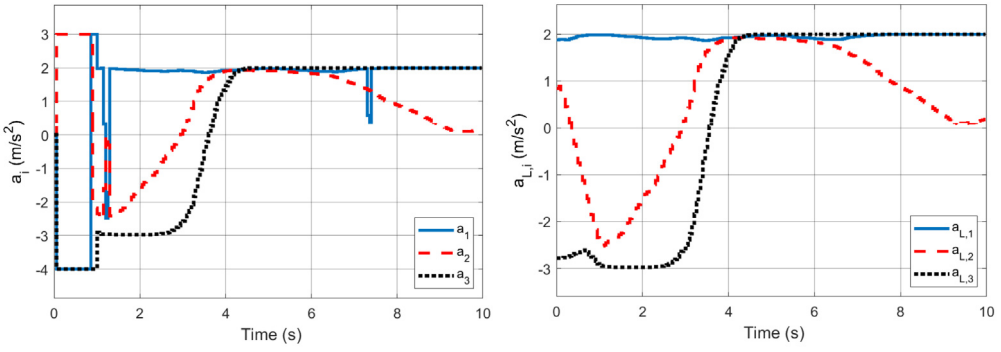


(a) $s_1 - s_2$ (Scenario 1) (b) $s_1 - s_3$ (Scenario 1)

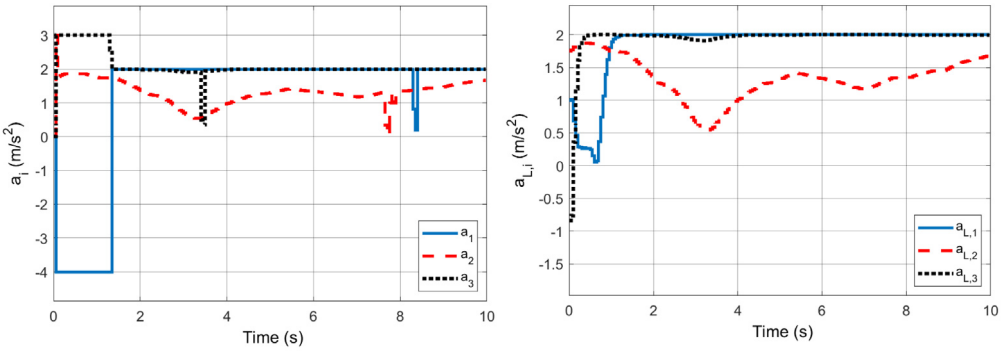


(c) $s_1 - s_2$ (Scenario 2) (d) $s_1 - s_3$ (Scenario 2)

Fig. 13. Illustration of the positions of the vehicles with the avoidable regions.

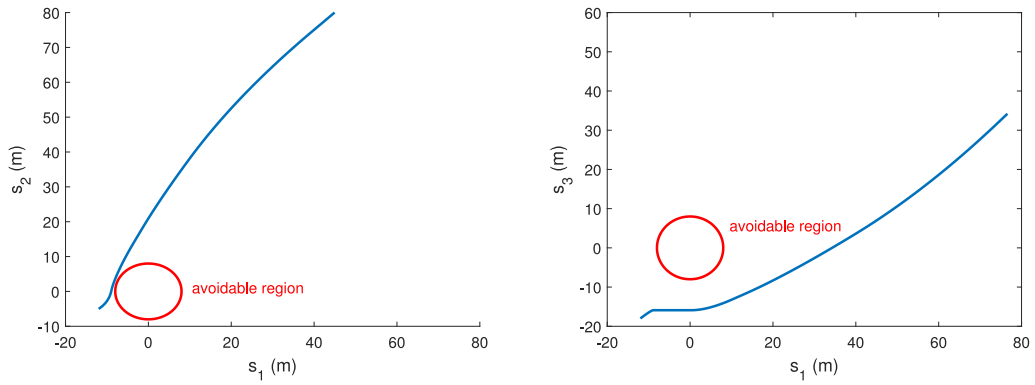


(a) Control inputs (Scenario 1) (b) Agent outputs (Scenario 1)



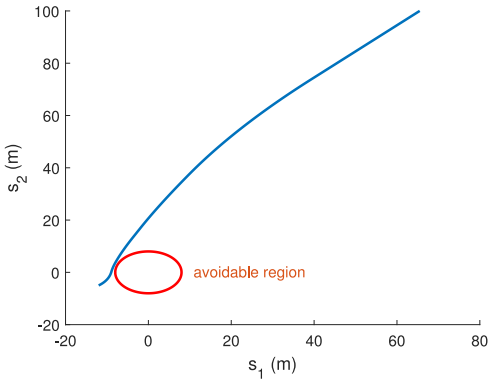
(c) Control inputs (Scenario 2) (d) Agent outputs (Scenario 2)

Fig. 14. Control inputs on the vehicles.

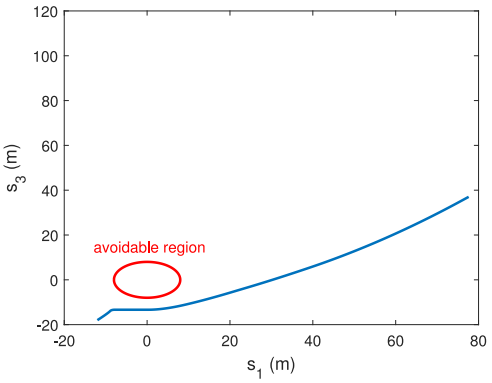


(a) $s_1 - s_2$ (Scenario 1)

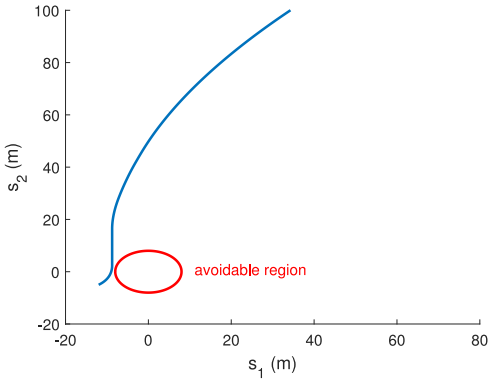
(b) $s_1 - s_3$ (Scenario 1)



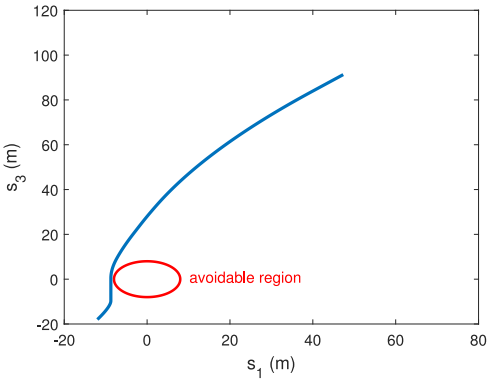
(c) $s_1 - s_2$ (Scenario 2)



(d) $s_1 - s_3$ (Scenario 2)



(e) $s_1 - s_2$ (Scenario 3)



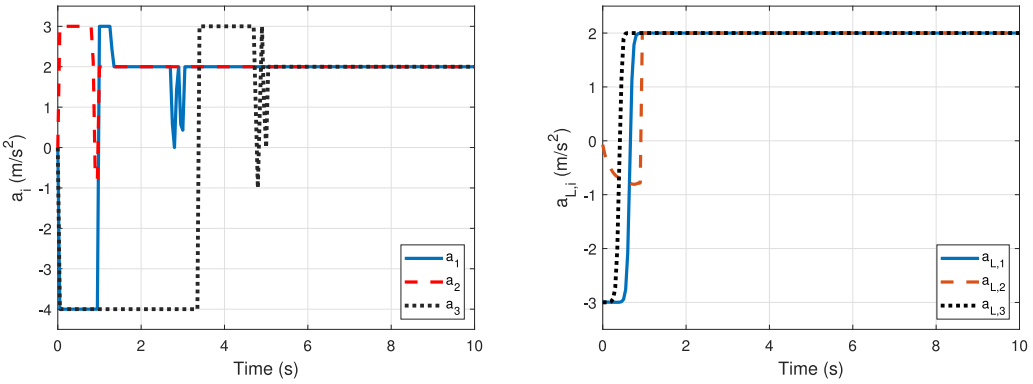
(f) $s_1 - s_3$ (Scenario 3)

Fig. 15. Illustration of the positions of the vehicles with the avoidable regions.

control systems of the vehicles using (16) and (18) are designed. Fig. 15 shows the trajectories of $s_1 - s_2$ and $s_1 - s_3$ for all scenarios. The first contribution of the trajectories is that the collision of the vehicles in case of all scenarios is guaranteed, see Fig. 15(a)–(d). Nevertheless, there are some differences in the scenarios, depending on the configurations. The second contribution is that the vehicles are further from each other, i.e., in case of independent configuration the $s_1 - s_i$ trajectories are slightly further from the avoidable region, see e.g., Figs. 13(a) and 15(a). It results in that $Vehicle\ 2$, and consequently, $Vehicle 3$ spend slightly more time in the intersection. For example, at the end of the 10 s simulation time s_1 is smaller in case of the independent configuration. The difference through the comparison of a_i and $a_{L,i}$ signals can be illustrated, see Figs. 14(a),(b) and 16(a),(b). The tendency of a_i signals is the same in both configurations, but in case of the centralized configurations the signals are smoother, whose reason is in the difference of $a_{L,i}$ signals.

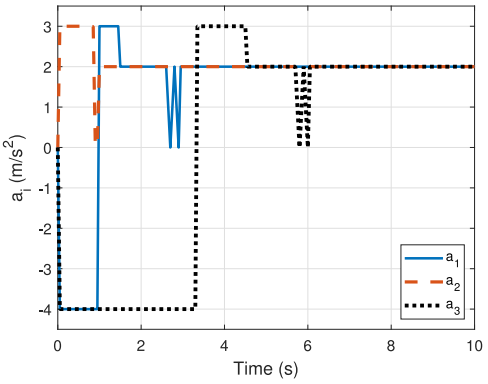
The third contribution of the comparison is that the ordering of the vehicles in case of Scenario 2 is the same as in Scenario 1. Thus, the increase of $v_3(0)$ does not result in priority for $Vehicle\ 3$ against $Vehicle\ 1$, while in the centralized cases the slight initial speed variation has been used for increasing the speed of $Vehicle\ 3$. The initial speed $v_3(0)$ must be increased to 9 m/s to modify vehicle ordering, which is known as Scenario 3, see Fig. 15(e),(f). Fig. 16(e) shows that a_3 in Scenario 3 is increased earlier as in Scenario 2. It results in that the last vehicle in Scenario 3, i.e., $Vehicle\ 1$ is further from the intersection at the end of the simulation ($s_1(10) = 47$ m), as the last vehicle in Scenario 2, i.e,., $Vehicle\ 3$, such as $s_3(10) = 39$ m.

Finally, the comparison of the simulations illustrate that the proposed hierarchical control guarantees safety performances (1) in both configurations. Evaluating non-safety performances, i.e., which are formed in the rewards (17),(18), the hierarchical control with

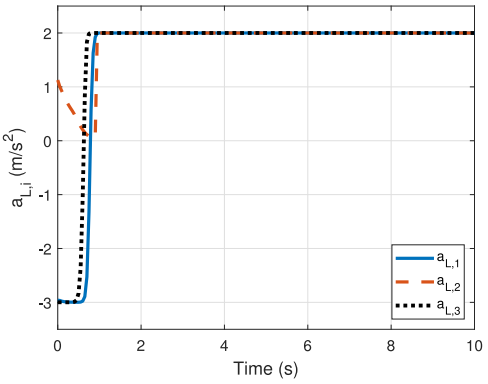


(a) Control inputs (Scenario 1)

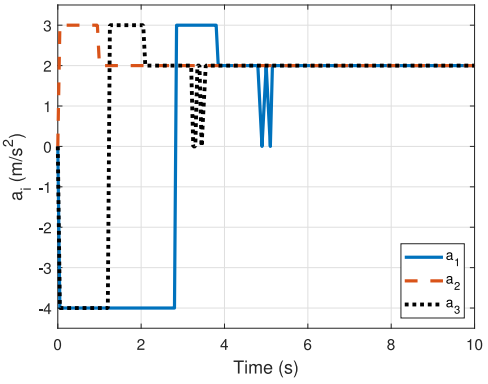
(b) Agent outputs (Scenario 1)



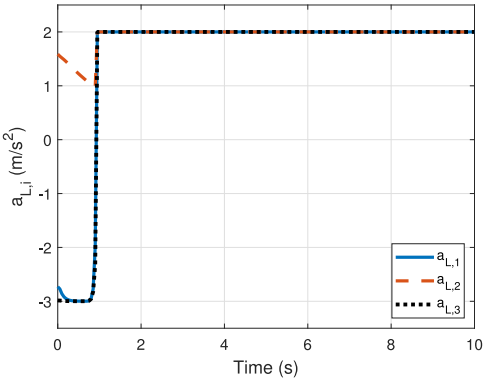
(c) Control inputs (Scenario 2)



(d) Agent outputs (Scenario 2)



(e) Control inputs (Scenario 3)



(f) Agent outputs (Scenario 3)

Fig. 16. Control inputs on the vehicles.

centralized configuration is more efficient than with independent configuration. Nevertheless, the centralized configuration results in only a slight improvement of the non-safety performance level. In spite of this improvement, the advantage of the control with independent configuration is that the complexity of the RL-based controller, i.e., the neural network, and the optimization task in the supervisor is lower.

This conclusion is confirmed by the simulation result below. Table 1 summarizes the results of simulations with different n_s number of vehicles in an intersection. The simulations have been performed at independent and centralized configurations and the computation times T_{comp} of the supervisory algorithm, related to a given input set, have been measured. For simulation purposes MATLAB 2020a on PC with Intel Core i7 10th Gen processor has been used. It can be seen that the increase of n_s has significant impact on T_{comp}, and that is more relevant on the centralized configuration. The increase in T_{comp} is resulted that the number of cases in the mixed-integer optimization is 2^n. Nevertheless, in case of centralized configuration further increasing is resulted that the number of optimization variables is $n_s + 1$, while in case of independent configuration it is 1 constantly. Consequently, high n_s at centralized configuration requires high T_{comp}, and thus, the algorithm in real-time may not be performed. But, in case of the independent configuration the considered number of vehicles during the computation is fixed. In the example, $n_s = 3$ is selected, and thus, $T_{comp} < T$, which is a criterion against the implementation.

6. Implementation of the motion control algorithm

The goal of this section is to demonstrate the effectiveness of the proposed control algorithm through its implementation on a HiL simulator environment.

Table 1
Computation time of the supervisory algorithm.

n_s	1	2	3	4	5	6	7	8	9	10
T_{comp} (s) for ind.	0.012	0.022	0.044	0.057	0.080	0.106	0.226	0.426	0.866	1.904
T_{comp} (s) for centr.	0.019	0.028	0.057	0.065	0.091	0.116	0.269	0.529	1.001	2.189

n_s	11	12	13	14	15
T_{comp} (s) for ind.	3.097	6.610	16.292	34.886	66.817
T_{comp} (s) for centr.	4.015	9.429	21.355	43.820	138.362

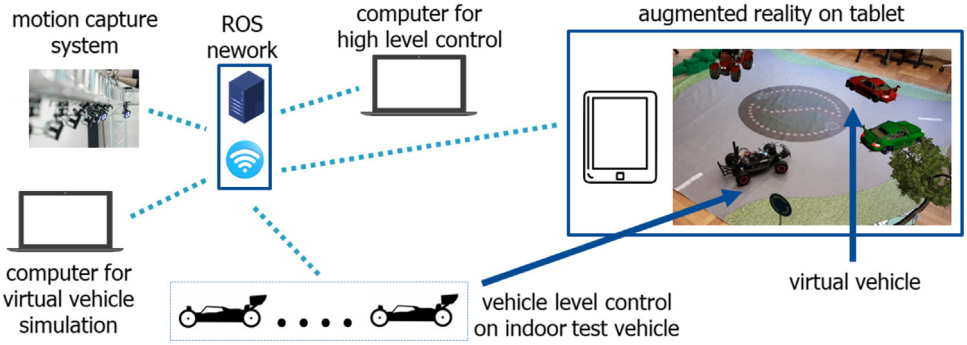


Fig. 17. Scheme of the HiL architecture with multiple vehicles.

The scheme of the HiL architecture is illustrated in Fig. 17. The communication between the elements through their connection to Robot Operating System (ROS) is realized. All of the elements in the HiL architecture, i.e., the test vehicles, the motion capture system, the computer for high level control, the computer for virtual vehicle simulation and the tablet for visualization as nodes into the ROS network are connected.

In the given setup the test vehicles are F1TENTH type of 1/10 sized wheeled RC vehicles. The vehicle contains a Jetson Xavier NX computer and various sensors, i.e., camera, LiDAR, IMU etc., but for achieving high precision positioning of the vehicles (± 0.2 mm), the OptiTrack motion capture system is used. The lateral motion of the vehicle is influenced by front wheel steering, whose value through a PID controller based on the actual lateral error is computed. Measuring the longitudinal speed of the vehicles is also the role of the OptiTrack motion capture system. The system contains 6 cameras, the measurements through reflective passive markers on the vehicles are realized. The measured position and speed signals through Wi-Fi are transmitted to the ROS server, and thus, these signals for all equipment with ROS nodes are available. The motions of the virtual vehicles on a computer using Matlab based on the vehicle model (4) are simulated. The advantage of virtual vehicles is that a high number of vehicles without their expensive physical realization in training and evaluation process can be incorporated. The augmented reality, with which the virtual vehicles can be visualized, is implemented on a tablet based on Android. For visualization purposes, an application based on Vuforia engine in Unity is developed. The operation requires a marker (e.g., on the floor) for the positioning of the tablet. The position, orientation information on the virtual vehicles to the tablet via the ROS network are transmitted. Moreover, the RL-based agent of the control loop on a computer is implemented, which is also connected to the ROS network. The RL-based agent through ROS gets information on the position and speed values of all test vehicles and virtual vehicles, which are used for the computation of the candidate control input $a_{L,i}(k)$.

In the rest of this section the effectiveness of the hierarchical motion control algorithm with independent configuration is demonstrated, i.e., two scenarios demonstrate the efficient motion of physical and virtual automated vehicles in a roundabout with and without high level connection.

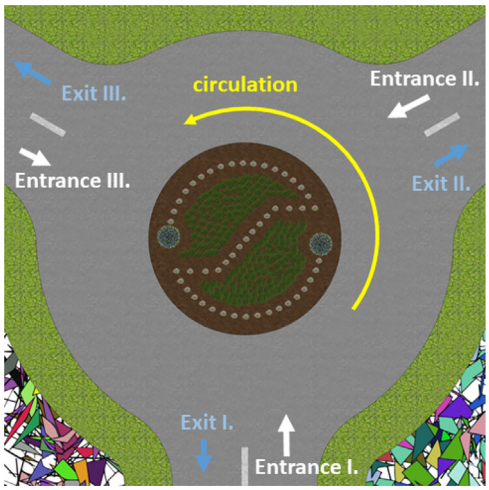


Fig. 18. Illustration of the roundabout example.

6.1. Demonstration on roundabout scenario

The scheme of the roundabout of the demonstration example in Fig. 18 is illustrated. The roundabout has anticlockwise circulation and three entrance/exit connections. This type of roundabout in the learning process with different entrance/exit variations of the vehicles has been used. During the simulations it has been assumed that the motions of the non-automated surrounding vehicles are not influenced by the automated vehicles, i.e., the automated vehicle must adapt to the other vehicles in the actual roundabout scenario. With regards to the roundabout scenario, the results of four simulation cases are shown. In simulation Case 1 and Case 2 the high level control is connected to the vehicles, while in Case 3 and Case 4 the connection has been lost, i.e., $a_{L,i} \equiv 0$. Moreover, in Case 1 and Case 3 the initial position of a virtual vehicle with automated control differs from its position in Case 2 and Case 4.

In the roundabout scenario three automated vehicles in independent configuration are involved, i.e., two F1TENTH test vehicles and a virtual automated vehicle (red), and two further non-automated virtual

(a) Red virtual vehicle has priority against test vehicle I (Case 1)

(b) Test vehicle I enters, keeping s_{safe} (Case 1)

(c) Test vehicle I leaves roundabout (Case 1)

(d) Test vehicle II follows burgundy virtual vehicle (Case 1)

(e) Test vehicle I decides to enter before red virtual vehicle (Case 2)

(f) Keeping s_{safe} (Case 2)

Fig. 19. Illustration on the roundabout scenario. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

vehicles (green, burgundy) are found, see the snapshots of Case 1 in Fig. 19. Fig. 19(a) shows the interaction of two automated vehicles, i.e., virtual vehicle and test vehicle I. Since the virtual vehicle starts its motion close to Entrance I, test vehicle I has to stop to keep s_{safe}. Nevertheless, when the virtual vehicle leaves Entrance I, the test vehicle enters into the roundabout, see Fig. 19(b). During its motion the distances of s_{safe} from both vehicles, i.e., the preceding and the follower test vehicle II are kept. In the next phase of the HiL simulation, virtual vehicle leaves the roundabout on Exit II and the motion of test vehicle I must be adapted to the preceding green virtual vehicle. Fig. 19(c) illustrates that test vehicle I also leaves the roundabout on Exit III and then, test vehicle II must adapt to the entering burgundy virtual vehicle, which has a slow motion. Finally, the virtual vehicle leaves the roundabout on Exit I and then the test vehicle II also leaves the roundabout on Exit I, while s_{safe} has been kept, see Fig. 19(d).

In Fig. 19(e) a snapshot on another scenario is illustrated, in which the initial position of red virtual vehicle is modified (Case 2), i.e., the red vehicle is further from Entrance I. The goal of this snapshot is to show that the automated test vehicle I is able to make an alternative decision, i.e., to enter prior to the red vehicle, if s $_{safe}$ can be kept. This difference can be seen through the comparison of Fig. 19(a) and (e).

The results on the roundabout scenario at Case 1 are shown in Fig. 20. The acceleration command of test vehicle I, that is a_1, is illustrated in Fig. 20(a). It can be seen that a_1 is close to $a_{L,1}$, which means that the high level control provides efficient candidate control inputs. Between 1 s and 2.5 s, the value of a_1 is reduced to -1 m/s^2 to guarantee priority for the red virtual vehicle, see also Fig. 19(a). Its impact is also shown in Fig. 20(d), see v_1 speed signal. Due to the priority, red virtual vehicle is able to move with its maximum speed 0.6 m/s almost during the entire route, see Fig. 20(c). Nevertheless, the test vehicle I and the test vehicle II must adapt to the surrounding vehicles, which adaptations result in varying acceleration profiles, see Fig. 20(a)–(b). The distance between the vehicles, i.e., s_i values for Case 1 are shown in Fig. 22(a). For example, between 4 s and 5 s the keeping of s_1 above s_{safe} (see Fig. 22(a)) requires the reduction of a_1 (see Fig. 20(a)), which also leads to decreased v_1 (see Fig. 22(d)). Since in this phase of the HiL simulation test vehicle I moves already in the roundabout, test vehicle II must also adapt to this situation, see its acceleration profile between 4 s and 5 s (Fig. 20(b)), its varying speed profile (Fig. 20(d)) and the keeping of s_2 above s_{safe} in (Fig. 22(a)). The similar effect between 13 s and 14.5 s is shown, when test vehicle II has to adapt to the motion of burgundy non-automated vehicle, see Fig. 20(b).

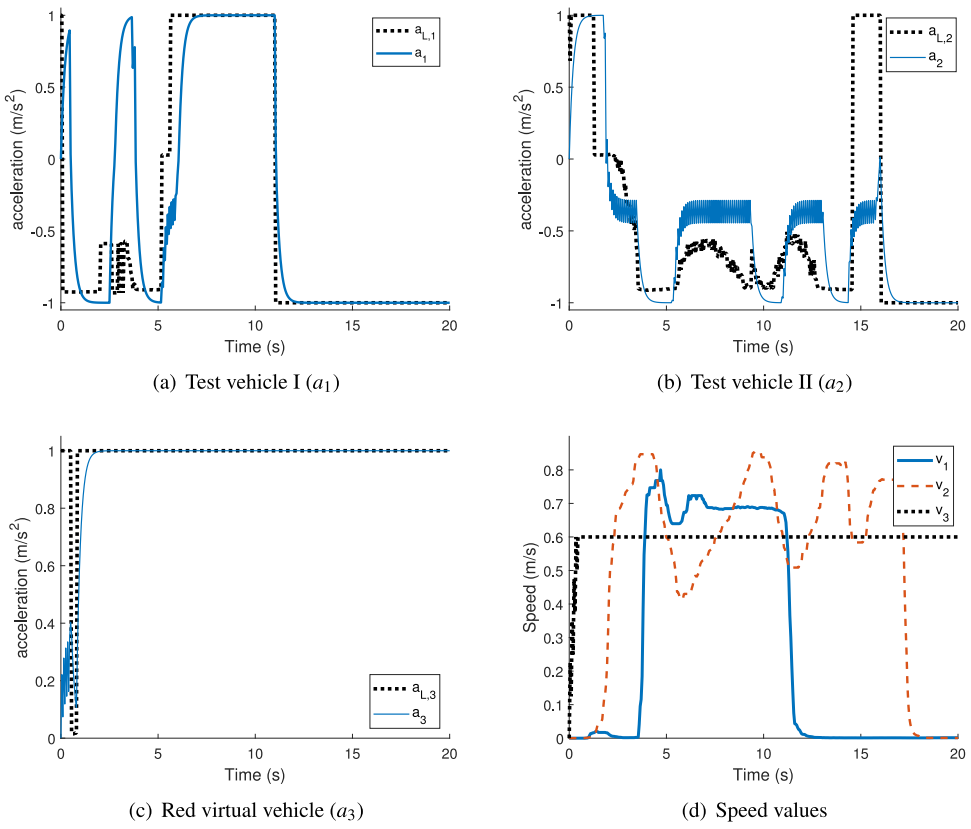
The speed profile for each automated vehicle in Case 2 and Case 3 for comparison purposes in Fig. 21 can be found. The difference of Case 1 and Case 2 is only the initial position of the red virtual vehicle, see also Fig. 19(a),(e). In Case 2 test vehicle I enters the roundabout before red virtual vehicle arrives at Entrance I, and thus, v_1 increases sharply at 1 s, while v_3 is reduced at 1.5 s. In Case 2 the speed profile of all vehicles differ from Case 1, compare Figs. 20(d) and 21(a). It reason is that after the entrance of test vehicle I, the ordering of the four vehicles in the roundabout is different. I.e., at Case 1: green virtual vehicle, red virtual vehicle, test vehicles I, test vehicle II, at Case 2: green virtual vehicle, test vehicle I, red virtual vehicle, test vehicle II, compare Fig. 19(e),(f). Moreover, in Fig. 21(b) the speed profiles of the automated vehicles at Case 3 are illustrated. It is shown that the loss of the high level control connection has only slight impact on the vehicle speeds, which is beneficial for safety purposes, see Fig. 21(b). Nevertheless, through the loss of the connection, the speed value and its variation on test vehicle I and on test vehicle II are increased. It means that the economy performance, i.e., minimization of a_i in the reward (see (18)) has not been achieved.

Finally, it is demonstrated that the variation of the initial speed or the loss of the high level control connection are not able to result in the violation of safety performance. In Fig. 22 the s_i values in all cases are illustrated. It can be seen that at all cases, all of the s_i values are above s_{safe}. It means that the proposed hierarchical control can provide guarantees on collision avoidance.

7. Conclusions

The paper has proposed the design of a hierarchical control strategy for guaranteeing the safe motion of automated vehicles under the scenarios of their interactions. The hierarchical structure has been formed for two configurations, i.e., centralized and independent configurations. Each configuration has a learning-based control on the high level, a robust control and a supervisor on the vehicle level. The advantages of the configurations through a comparative simulation scenario have been presented.

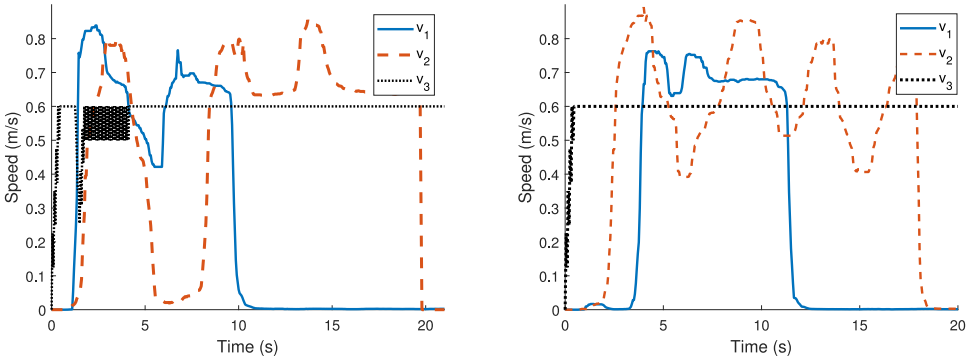
The hierarchical control strategy in independent configuration of the automated vehicles has been implemented. It has been demonstrated through HiL simulations under intersection and roundabout



(a) Test vehicle I (a_1) (b) Test vehicle II (a_2)

(c) Red virtual vehicle (a_3) (d) Speed values

Fig. 20. Results on roundabout scenario (Case 1). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



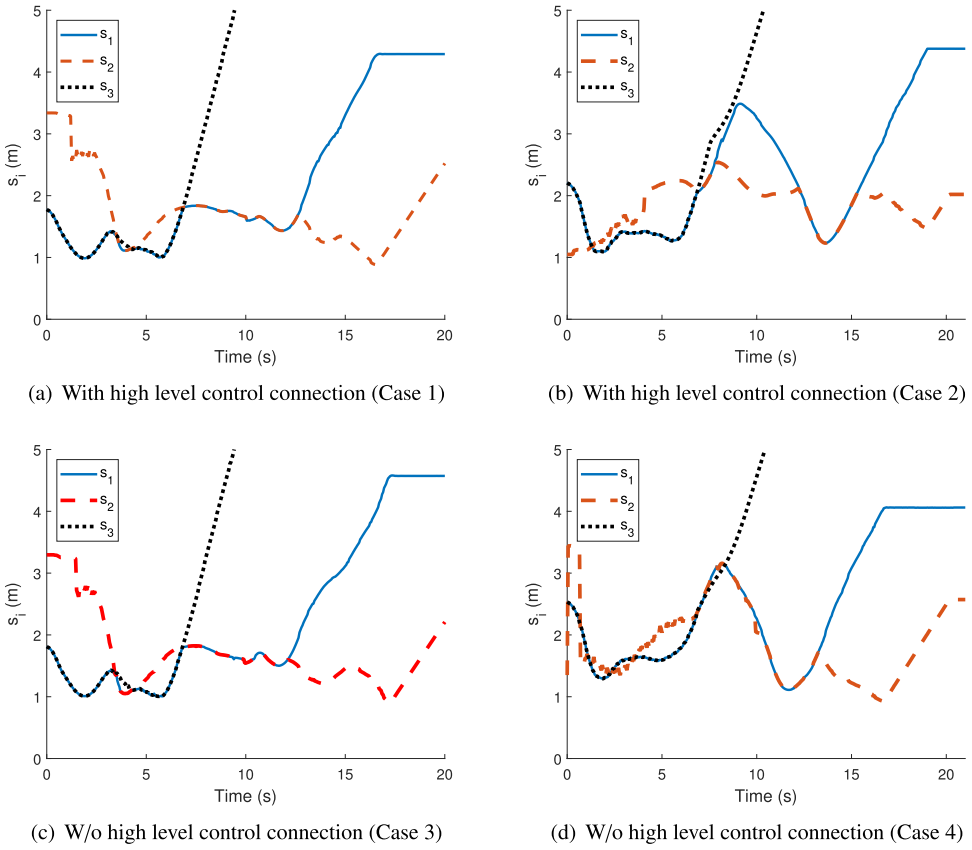
(a) Speed values (Case 2) (b) Speed values (Case 3)

Fig. 21. Speed profiles at different cases.

environments, that the safe motion of the vehicles in all of the scenarios can be guaranteed. Moreover, through the learning-based control the economy performance of the automated vehicles can be improved. The demonstrations have also shown that the independently-controlled automated vehicles are able to adapt to the motions of each other and the those of non-automated vehicles.

In spite of the safe and effective operation of the hierarchical control strategy, some limitations on the method can be found. The applicability of the method can be increased with the improvement of the kinematic vehicle model in the supervisor to dynamic vehicle model. This improvement can result in the application of the method for handling vehicle interactions in high speed roundabouts. Another limitation of the method is the point mass model of the vehicle. Although this limitation through increased safety distance between the vehicles can be compensated, but through consideration of real vehicle sizes the distance between the vehicles can be reduced.

Moreover, a further theoretical challenge of the research is to provide a control strategy, which is able integrate the benefits of centralized and independent configurations. For example, in the proposed roundabout scenario the modification of the initial position of a vehicle has led to different vehicle ordering in the roundabout. The different ordering has also led to different speed profiles. Nevertheless, it can be a further challenge to select the speed profile of the individual automated vehicles to improve the economy performances of all vehicles. It requests a control strategy, in which the centralized control can be decentralized to achieve independent configuration, but the performance level of the centralized configuration can be preserved, see e.g., the concept of global/local control (Nguyen, Hara, & Tsumura, 2017; Wang, Fujimoto, & Hara, 2017), or distributed reinforcement learning (Jiang, Huang, Jafari, & Jalayer, 2022; Li, Wu, & He, 2022; Muzahid et al., 2023; Wang, Zhang, Xu, Li, & Ran, 2019). Using these methods, due to the distributed functionality in the individual



(a) With high level control connection (Case 1)

(b) With high level control connection (Case 2)

(c) W/o high level control control connection (Case 3)

(d) W/o high level control control connection (Case 4)

Fig. 22. Distances between the vehicles in the roundabout scenario.

configuration, further improvement on the reductions of traveling time and traction energy can be achieved. These approaches can lead to the design of automated vehicle control in an individual configuration, and the global performances on the local level can be considered.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The research was supported by the European Union within the framework of the National Laboratory for Autonomous Systems (RRF-2.3.1-21-2022-00002). The paper was partially funded by the National Research, Development and Innovation Office (NKFIH) under OTKA Grant Agreement No. K 135512.

References

Belotti, P., Liberti, L., Lodi, A., Nannicini, G., & Tramontani, A. (2011). Disjunctive inequalities: Applications and extensions. In *Wiley encyclopedia of operations research and management science*. American Cancer Society.

Bichiou, Y., & Rakha, H. A. (2019). Developing an optimal intersection control system for automated connected vehicles. *IEEE Transactions on Intelligent Transportation Systems*, *20*(5), 1908–1916.

Carvalho, A., Gao, Y., Gray, A., Tseng, H. E., & Borrelli, F. (2013). Predictive control of an autonomous ground vehicle using an iterative linearization approach. In *16th international IEEE conference on intelligent transportation systems (ITSC 2013)* (pp. 2335–2340).

Chai, L., Cai, B., ShangGuan, W., & Wang, J. (2017). Connected and autonomous vehicles coordinating method at intersection utilizing preassigned slots. In *2017 IEEE 20th international conference on intelligent transportation systems (ITSC)* (pp. 1–6).

Chalaki, B., Beaver, L. E., Remer, B., Jang, K., Vinitsky, E., Bayen, A. M., et al. (2020). Zero-shot autonomous vehicle policy transfer: From simulation to real-world via adversarial learning. In *2020 IEEE 16th international conference on control automation (ICCA)* (pp. 35–40). http://dx.doi.org/10.1109/ICCA51439.2020.9264552.

Chalaki, B., & Malikopoulos, A. A. (2022). Optimal control of connected and automated vehicles at multiple adjacent intersections. *IEEE Transactions on Control Systems Technology*, *30*(3), 972–984.

Chen, C., Xu, Q., Cai, M., Wang, J., Wang, J., & Li, K. (2022). Conflict-free cooperation method for connected and automated vehicles at unsignalized intersections: Graph-based modeling and optimality analysis. *IEEE Transactions on Intelligent Transportation Systems*, *23*(11), 21897–21914.

Chen, X., Xu, B., Qin, X., Bian, Y., Hu, M., & Sun, N. (2020). Non-signalized intersection network management with connected and automated vehicles. *IEEE Access*, *8*, 122065–122077.

Chen, X., Xu, B., Qin, X., Bian, Y., Hu, M., & Sun, N. (2020). Non-signalized intersection network management with connected and automated vehicles. *IEEE Access*, *8*, 122065–122077.

Chen, J., Yuan, B., & Tomizuka, M. (2019a). Deep imitation learning for autonomous driving in generic urban scenarios with enhanced safety. In *2019 IEEE/RSJ international conference on intelligent robots and systems (IROS)* (pp. 2884–2890). http://dx.doi.org/10.1109/IROS40897.2019.8968225.

Chen, J., Yuan, B., & Tomizuka, M. (2019b). Model-free deep reinforcement learning for urban autonomous driving. In *2019 IEEE intelligent transportation systems conference (ITSC)* (pp. 2765–2771). http://dx.doi.org/10.1109/ITSC.2019.8917306.

Debada, E., Makarem, L., & Gillet, D. (2017). A virtual vehicle based coordination framework for autonomous vehicles in heterogeneous scenarios. In *2017 IEEE international conference on vehicular electronics and safety (ICVES)* (pp. 51–56). http://dx.doi.org/10.1109/ICVES.2017.7991900.

Dresner, K., & Stone, P. (2008). A multiagent approach to autonomous intersection management. *Journal of Artificial Intelligence Research*, *31*(1), 591–656.

Gholamhosseinian, A., & Seitz, J. (2022). A comprehensive survey on cooperative intersection management for heterogeneous connected vehicles. *IEEE Access*, *10*, 7937–7972.

Hausknecht, M., Au, T. C., & Stone, P. (2011). Autonomous intersection management: Multi-intersection optimization. In *2011 IEEE/RSJ international conference on intelligent robots and systems* (pp. 4581–4586).

Hu, X., Chen, L., Tang, B., Cao, D., & He, H. (2018). Dynamic path planning for autonomous driving on various roads with avoidance of static and moving obstacles. *Mechanical Systems and Signal Processing*, *100*, 482–500.

Isele, D., Rahimi, R., Cosgun, A., Subramanian, K., & Fujimura, K. (2018). Navigating occluded intersections with autonomous vehicles using deep reinforcement learning. In *2018 IEEE international conference on robotics and automation (ICRA)* (pp. 2034–2039).

Jiang, S., Huang, Y., Jafari, M., & Jalayer, M. (2022). A distributed multi-agent reinforcement learning with graph decomposition approach for large-scale adaptive traffic signal control. *IEEE Transactions on Intelligent Transportation Systems*, *23*(9), 14689–14701.

Katrakazas, C., Quddus, M., Chen, W.-H., & Deka, L. (2015). Real-time motion planning methods for autonomous on-road driving: State-of-the-art and future research directions. *Transportation Research Part C (Emerging Technologies)*, *60*, 416–442.

Kiran, B. R., Sobh, I., Talpaert, V., Mannion, P., Sallab, A. A. A., Yogamani, S., et al. (2022). Deep reinforcement learning for autonomous driving: A survey. *IEEE Transactions on Intelligent Transportation Systems*, *23*(6), 4909–4926.

Kumaravel, S. D., Malikopoulos, A. A., & Ayyagari, R. (2022). Optimal coordination of platoons of connected and automated vehicles at signal-free intersections. *IEEE Transactions on Intelligent Vehicles*, *7*(2), 186–197.

Lattarulo, R., & Pérez Rastelli, J. (2021). A hybrid planning approach based on MPC and parametric curves for overtaking maneuvers. *Sensors*, *21*(2).

Li, G., Wu, J., & He, Y. (2022). HARL: A novel hierachical adversary reinforcement learning for automoumous intersection management. URL https://arxiv.org/abs/2205.02428.

Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., et al. (2016). Continuous control with deep reinforcement learning. In *International conference on learning representations*.

Masi, S., Xu, P., & Bonnifait, P. (2018). Adapting the virtual platooning concept to roundabout crossing. In *2018 IEEE intelligent vehicles symposium (IV)* (pp. 1366–1372). http://dx.doi.org/10.1109/IVS.2018.8500611.

Mehran, Z. A., & Nasser, L. A. (2021). On-line situational awareness for autonomous driving at roundabouts using artificial intelligence. *Journal of Machine Intelligence and Data Science*, *2*, 17–24. http://dx.doi.org/10.11159/jmids.2021.003.

Morales Medina, A. I., Creemers, F., Lefeber, E., & van de Wouw, N. (2020). Optimal access management for cooperative intersection control. *IEEE Transactions on Intelligent Transportation Systems*, *21*(5), 2114–2127.

Moreau, J., Melchior, P., Victor, S., Cassany, L., Moze, M., Aioun, F., et al. (2019). Reactive path planning in intersection for autonomous vehicle. *IFAC-PapersOnLine*, *52*(5), 109–114, 9th IFAC Symposium on Advances in Automotive Control AAC 2019.

Müller, J., Strohbeck, J., Herrmann, M., & Buchholz, M. (2022). Motion planning for connected automated vehicles at occluded intersections with infrastructure sensors. *IEEE Transactions on Intelligent Transportation Systems*, *23*(10), 17479–17490.

Muzahid, A. J. M., Kamarulzaman, S. F., Rahman, M. A., Murad, S. A., Kamal, M. A. S., & Alenezi, A. H. (2023). Multiple vehicle cooperation and collision avoidance in automated vehicles: survey and an AI-enabled conceptual framework. *Scientific Reports*, *603*(13).

Németh, B., & Gáspár, P. (2019). Coordination of automated and human-driven vehicles in intersection scenarios. In *2019 American control conference (ACC)* (pp. 5278–5283).

Németh, B., & Gáspár, P. (2021). Guaranteed performances for learning-based control systems using robust control theory. In A. Koubaa, & A. T. Azar (Eds.), *Deep learning for unmanned systems* (pp. 109–142). Cham: Springer International Publishing.

Németh, B., & Gáspár, P. (2021a). Design of learning-based control with guarantees for autonomous vehicles in intersections. *IFAC-PapersOnLine*, *54*(2), 210–215, 16th IFAC Symposium on Control in Transportation Systems CTS 2021.

Németh, B., & Gáspár, P. (2021b). The design of performance guaranteed autonomous vehicle control for optimal motion in unsignalized intersections. *Applied Sciences*, *11*(8).

Németh, B., Gáspár, P., & Szabó, Z. (2021). Guaranteed performances for a learning-based eco-cruise control using robust LPV method. *IFAC-PapersOnLine*, *54*(8), 83–88, 4th IFAC Workshop on Linear Parameter Varying Systems LPVS 2021.

Nguyen, B. M., Hara, S., & Tsumura, K. (2017). Hierarchically decentralized control for in-wheel-motored electric vehicles with global and local objectives. In *2017 11th Asian control conference (ASCC)* (pp. 1176–1181).

Qian, B., Zhou, H., Lyu, F., Li, J., Ma, T., & Hou, F. (2019). Toward collision-free and efficient coordination for automated vehicles at unsignalized intersection. *IEEE Internet of Things Journal*, *6*(6), 10408–10420.

Ren, Y., Jiang, J., Zhan, G., Li, S. E., Chen, C., Li, K., et al. (2022). Self-learned intelligence for integrated decision and control of automated vehicles at signalized intersections. *IEEE Transactions on Intelligent Transportation Systems*, *23*(12), 24145–24156.

Riegger, L., Carlander, M., Lidander, N., Murgovski, N., & Sjöberg, J. (2016). Centralized MPC for autonomous intersection crossing. In *2016 IEEE 19th international conference on intelligent transportation systems (ITSC)* (pp. 1372–1377).

Rodrigues, M., McGordon, A., Gest, G., & Marco, J. (2018). Autonomous navigation in interaction-based environments - A case of non-signalized roundabouts. *IEEE Transactions on Intelligent Vehicles*, *3*(4), 425–438. http://dx.doi.org/10.1109/TIV.2018.2873916.

Rosolia, U., & Borrelli, F. (2018). Learning model predictive control for iterative tasks. A data-driven control framework. *IEEE Transactions on Automatic Control*, *63*(7), 1883–1896.

Rosolia, U., & Borrelli, F. (2020). Learning how to autonomously race a car: A predictive control approach. *IEEE Transactions on Control Systems Technology*, *28*(6), 2713–2719.

Stahl, T., Wischnewski, A., Betz, J., & Lienkamp, M. (2019). Multilayer graph-based trajectory planning for race vehicles in dynamic scenarios. In *2019 IEEE intelligent transportation systems conference (ITSC)*. IEEE.

Svensson, L., Masson, L., Mohan, N., Ward, E., Brenden, A. P., Feng, L., et al. (2018). Safe stop trajectory planning for highly automated vehicles: An optimal control problem formulation. In *2018 IEEE intelligent vehicles symposium (IV)* (pp. 517–522).

Tachet, R., Santi, P., Sobolevsky, S., Ignacio Reyes-Castro, L., Frazzoli, E., Helbing, D., et al. (2016). Revisiting street intersections using slot-based systems. In *PLoS One*. *11*, Article e0149607.

Tian, R., Li, N., Kolmanovsky, I., Yildiz, Y., & Girard, A. R. (2020). Game-theoretic modeling of traffic in unsignalized intersection network for autonomous vehicle control verification and validation. *IEEE Transactions on Intelligent Transportation Systems*, 1–16.

Tollner, D., Cao, H., & Zöldy, M. (2019). Artificial intellgence based decision making of autonomous vehicles before entering roundabout. In *2019 IEEE 19th international symposium on computational intelligence and informatics and 7th IEEE international conference on recent achievements in mechatronics, automation, computer sciences and robotics (CINTI-MACRo)* (pp. 000181–000186). http://dx.doi.org/10.1109/CINTI-MACRo49179.2019.9105322.

Wang, Y., Fujimoto, H., & Hara, S. (2017). Driving force distribution and control for EV with four in-wheel motors: A case study of acceleration on split-friction surfaces. *IEEE Transactions on Industrial Electronics*, *64*(4), 3380–3388.

Wang, C., Zhang, J., Xu, L., Li, L., & Ran, B. (2019). A new solution for freeway congestion: Cooperative speed limit control using distributed reinforcement learning. *IEEE Access*, *7*, 41947–41957.

Werling, M., Ziegler, J., Kammel, S., & Thrun, S. (2010). Optimal trajectory generation for dynamic street scenarios in a Frenet Frame. In *2010 IEEE international conference on robotics and automation* (pp. 987–993). http://dx.doi.org/10.1109/ROBOT.2010.5509799.

Wu, T., Jiang, M., & Zhang, L. (2020). Cooperative multiagent deep deterministic policy gradient ($CoMADDPG$) for intelligent connected transportation with unsignalized intersection. *Mathematical Problems in Engineering*.

Xu, H., Xiao, W., Cassandras, C. G., Zhang, Y., & Li, L. (2022). A general framework for decentralized safe optimal control of connected and automated vehicles in multi-lane signal-free intersections. *IEEE Transactions on Intelligent Transportation Systems*, *23*(10), 17382–17396.

Yan, S., Welschehold, T., Büscher, D., & Burgard, W. (2022). Courteous behavior of automated vehicles at unsignalized intersections via reinforcement learning. *IEEE Robotics and Automation Letters*, *7*(1), 191–198.

Yang, H., Almutairi, F., & Rakha, H. (2021). Eco-driving at signalized intersections: A multiple signal optimization approach. *IEEE Transactions on Intelligent Transportation Systems*, *22*(5), 2943–2955.

Zhou, M., Yu, Y., & Qu, X. (2020). Development of an efficient driving strategy for connected and automated vehicles at signalized intersections: A reinforcement learning approach. *IEEE Transactions on Intelligent Transportation Systems*, *21*(1), 433–443.

Zohdy, I., & Rakha, H. (2012). Optimizing driverless vehicles at intersections. In *19th ITS world congress*. Vienna, Austria.