Game theoretic decision making for autonomous vehicles' merge manoeuvre in high traffic scenarios

Mario Garzón, Anne Spalanzani

Game theoretic decision making for autonomous vehicles' merge manoeuvre in high traffic scenarios

Mario Garzón[1] and Anne Spalanzani[1]

***Abstract*— This paper presents a game theoretic decision making process for autonomous vehicles. Its goal is to provide a solution for a very challenging task: the merge manoeuvre in high traffic scenarios. Unlike previous approaches, the proposed solution does not rely on vehicle-to-vehicle communication or any specific coordination, moreover, it is capable of anticipating both the actions of other players and their reactions to the autonomous vehicle's movements.**

The game used is an iterative, multi-player level-k model, which uses cognitive hierarchy reasoning for decision making and has been proved to correctly model human decisions in uncertain situations. This model uses reinforcement learning to obtain a near-optimal policy, and since it is an iterative model, it is possible to define a goal state so that the policy tries to reach it.

To test the decision making process, a kinematic simulation was implemented. The resulting policy was compared with a rule-based approach. The experiments show that the decision making system is capable of correctly performing the merge manoeuvre, by taking actions that require reactions of the other players to be successfully completed.

I. INTRODUCTION

The past few years have seen an increasing interest on the research and development of autonomous cars. As new developments are made, more complex tasks and situations are being tackled, therefore requiring either more complex or more creative solutions. Among the many challenges that have arise for autonomous cars, the interaction with human drivers is one of the most complex and interesting. This is due to the fact that involving human drivers require to model their behaviour and anticipate their intentions, which may not be clear and therefore can not be predicted with certitude.

This paper proposes a solution to a very challenging scenario, the on-ramp merging of an autonomous car into a road with high traffic flow, namely during a traffic jam. Furthermore, the scenario does not assume any shared control or intercommunication with neighbourhood vehicles, therefore, it requires a non-coordinated bi-directional interaction between the autonomous cars and human drivers. A real world situation, such as the one proposed in this scenario is shown in Figure 1.

In order to solve a situation as the one proposed here, it is necessary to solve many challenges, they can be broadly divided in four areas: situation awareness of the autonomous vehicle or *ego-car*, prediction of actions or movements by other vehicles, decision making of the ego-car and execution of the manoeuvre. The work presented on this paper focuses on two of those challenges, the anticipation of actions from other vehicles and the decision making of the ego-car. Those two task are strongly dependent on each other, because the decision making should consider both the actions of the other vehicles as well as their reaction to the movements of the ego-car. Moreover, the solution proposed solves these two tasks simultaneously by involving the estimated actions and reactions of other vehicles in the decision making model.

Fig. 1: Example of the task and scenario: An autonomous vehicle (in the blue circle), needs to merge with the traffic.

Commonly used decision making process are not suitable in this case. The reason for this is that most techniques relay on overly defensive strategies, usually assuming constant velocity or simplistic models for other vehicles and human drivers, applying a "stay out of their way" strategy. Although those approaches may work in simpler scenarios, they are not suitable in the presence of high traffic, because there it is impossible to merge without affecting the cars surrounding our vehicle [1].

The approach used in this paper is based on Game Theory, which is a tool that provides a good framework to model and solve interactions between multiple agents [2], therefore allowing the vehicle to make a decision that can be seen as a risky manoeuvre if seen by traditional techniques, and at the same time anticipating both the movements of other cars as well as their reaction to the movements of the ego-car.

There are many different techniques within game theory, some of them have already been applied to autonomous cars [3]. Furthermore, usually the game theoretical decision making seeks equilibrium by assuming mutual rationality and mutual consistency (*i.e.* every player will always make the rational decision and all the players' beliefs about other players are consistent and true). This however is not necessary true, and therefore new game models have been proposed.

The game theoretical model used in this paper, is based on the *cognitive hierarchy* model, proposed by *Camerer et al.* [4], which has been experimentally validated with

[1] The authors are with Univ. Grenoble Alpes, Inria, Grenoble INP, 38000 Grenoble, France - mario.garzon-oviedo@inria.fr, anne.spalanzani@inria.fr

cognition experiments [5]. Its main difference with most other game theory models, is that it assumes that some player's beliefs may be mistaken, and optimizes a policy based on that assumption, therefore yielding a higher level policy that is better than that of everyone else's. These games are also known as Level-k thinking, because starting from the assumption that every player will use a basic (level 0) policy, a player (level 1) will try to find a new optimal policy assuming every one else will use the previous one (level 0), and using the same logic, higher levels can be reached.

The main contribution of this paper, is the use of an iterative, multi-player level-k game model, in order to solve a very complex task for autonomous cars. This model, relies on the creation of an Iterative Semi Network-Form Game [6], which allows to obtain a near-optimal policy, taking into account the estimation of the possible actions of surrounding vehicles, as well as their reaction to the actions of the ego-car. Furthermore, the proposed technique does not require any vehicle-to-vehicle communication or coordinated behaviour. Also, the observation space and the set of actions proposed, can be applied to many other scenarios and other complex situations. In addition to this, the iterative model proposed, makes it possible to set a goal state and therefore model long term intentions of the ego-car and the additional vehicles.

The remainder of this paper is structured as follows Section II presents a brief summary of related works and states the main differences of the work presented here. Then, Section III describes the proposed methodology and their main components. Section IV details the implementation procedure and Section V describes the scenarios and experiments used to test the capabilities of the system and, finally, Section VII presents the concluding remarks.

II. RELATED WORK

Several different techniques to perform the merging manoeuvre have been reported in the past few years. One of the first works using game theory for this tasks was proposed by *Kita et at.* [2], it models a two-person game where players can either give-way or not and merge or pass. Also it obtains the reward function by finding correlation with real world observations. However, its drawback, apart from only considering two players, is that it assumes perfect information and knowledge of the other player.

Some of the works found on the literature are based on a shared control of all the vehicles in the scenario. Merging using on slot-based driving [7], or relying on vehicle-to-vehicle communication and strong cooperation, so as to yield space for the merge [8] have been proposed. Similarly, *Brechtel et al.* [9] uses continuous POMDP in order to generate safe, efficient, and goal-directed driving behaviours for a two-vehicles interaction, although only two possible actions (change lane or stay) were considered. A different work includes non-cooperative road users [10], however, those additional road users are assumed to follow a single trajectory that is not modified by the ego-car's actions.

Considering the possible reactions of other vehicles (uncoordinated approaches), a multi-policy decision-making process was proposed by *Cunningham et at.* [11]. It uses driving models to create a set of policies that evaluate the consequences of the ego-car actions. The effects that the actions of an autonomous car can take on human drivers are studied in the work of *Sadigh et al.* [12], this uses Inverse Reinforcement Learning to determine the cost functions for those possible actions. Another work uses a Receding Horizon Control and game theory to maximize mutual payoff of different players [3], However, all those works only consider one additional vehicle or player and two possible actions, rendering it invalid for high traffic scenarios.

Another approach generates many possible velocity profiles, and then it estimates an acceptable breaking or discomfort in other vehicles in reaction to each possible trajectory [1]. That estimation is based on the Intelligent Driver Model. A different approach, also based on a set of traffic rules and regulations, requires complex parameters and movements equations [13]. The main drawback of these approaches is that, since they are model based, they will produce mostly over safe manoeuvres, that will cause longer waits in order wait for a safe space for the merge manoeuvre.

Finally, the work proposed by *Li et al.* [14], which similarly to the work presented here, uses a level-k game theory approach to model and simulate traffic and it considers many players interacting and also many possible actions. However, that work is mainly focused on calibrating, testing and comparing decision and control systems for autonomous vehicles.

The main differences of the our proposal with respect to the related works is as follows. Firstly, it can consider the reactions of multiple players to the ego-car actions, furthermore, it differs from the work of *Li et al.* [14] in that it can have different models or policies for each additional player, and it can also model end-of-roads or other variation of the road. Another important difference, is that it uses a time-iterative approach, that allows to define a goal state, and therefore it can be used to model many different scenarios, such as vehicles entering or leaving a highway. In addition to this, the assumptions of other players actions do not need to be perfect, and therefore it is possible to perform interactions with human drivers or other unconnected autonomous vehicles.

III. MERGE SCENARIO MODEL

The objective of the work presented here is to develop a decision making approach capable of solving a very complex task for autonomous vehicles, the on-ramp merge during a traffic jam. In order to solve this, it was necessary to define a world model, that allows to simulate the manoeuvre and the movements of all the vehicles. A model of the scenario, with its main components, is presented on Figure 2, which corresponds to a two lane road, that reduces to a single lane after a given distance.

A. World Model

In order to perform the simulated merge, a model of the world, which defines the movements of the vehicles,

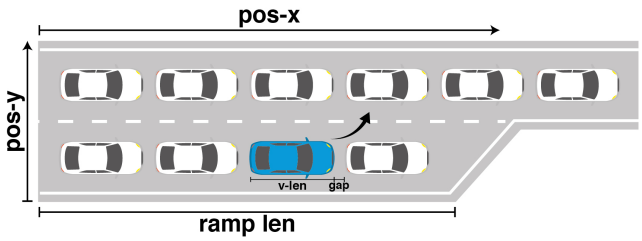


Fig. 2: Modelling of the world. The ego-car is shown in blue, and the variables considered in the world model are marked.

possible collisions and road restrictions has been defined. This world model uses discrete-time kinematic equations to define the movements of the vehicles, and rule-based checks for collision as well as to ensure the limits of the road. For each vehicle, the following parameters and measurements are defined (See Fig 2):

- v_len Length of the vehicle.
- pos_x, pos_y Position (forward and lateral).
- $accel\_x$ Linear acceleration.
- vel_x, vel_y Linear, (forward and lateral) velocity.

Having this, the equations of motion are defined according to Equation 1

$$\begin{aligned} pos\_x(t+1) &= pos\_x(t) + vel\_x(t) \cdot t\_step \\ vel\_x(t+1) &= vel\_x(t) + accel\_x(t) \cdot t\_step \\ pos\_y(t+1) &= pos\_y(t) + vel\_y(t) \cdot t\_step \end{aligned} \tag{1}$$

Some additional restrictions are also included on the world model, namely: The forward velocity is limited by a maximum speed parameter. The lateral position pos_y is only controlled by the change lane actions, thus, the lateral position corresponds to the lane the vehicle is on, and the lateral velocity vel_y will be either -1, 0 or 1 depending on the action taken. All the boundaries of the road, including the end of the merge ramp are ensured, as well as maximum acceleration and deceleration values. Furthermore, although the scope of this work does not include the use of a complex traffic simulator, the models of the world and vehicles, as well as the information available, have been selected so as to facilitate its future integration with a high traffic simulator previously developed [15].

B. Observation Space

As in any real world driving scenario, the vehicles involved are not capable of having complete information about every other vehicle. Moreover, when involving many other vehicles in the decision making, it will be very complex to process such amount of information. The solution proposed here relies on a different observation space, which, as any human driver will do, only considers the surroundings of the vehicle. Moreover, by having a broad set of discrete states, the observations can handle uncertainties that may be found in the measurements. The objective of these observations is to be able to provide a sufficient amount of information, without requiring complete or exact measurements.

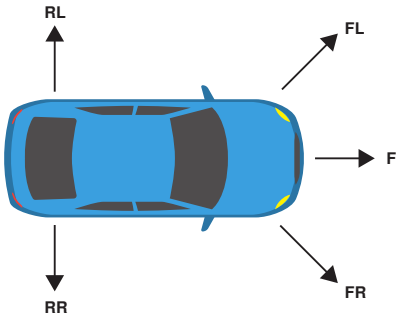


Fig. 3: Observation space. The possible states for each measurement are: forbidden, free, vehicle static, vehicle moving.

The definition of the observation space is shown in Figure 3. As can be seen there, each vehicle handles five different measurements, corresponding to the status of the front F, front-left FL, front-right FR, rear-left RL and rear-right RR. Observations of the back of the vehicle are not considered, because, as stated by *Li et al.* [14], they don't provide relevant enough information for the decision making process and including the will highly increment the size of the observation space.

There are four possible states for each of those measurements, they can be: *forbidden*, there is no road available; *free*, if there is no vehicle and the road is clear; *vehicle static*, if a vehicle is stopped or moving slower that the ego-car; and finally *vehicle moving*, meaning that a vehicle is present and moving at the same speed or faster than the ego-car.

C. Possible Actions

A set of five possible actions has been defined for the vehicles. As with the observation space, they are designed for the merge scenario, but they can also be used in many other situations. The actions are described next:

- *left:* Change lane to the left (if available).
- *right:* Change lane to the right (if available).
- *remain:* Maintain the same speed, direction and lane.
- *accel:* Increase speed at given rate. (up to max. velocity)
- *brake:* Decrease speed at given rate. (until velocity is zero)

D. Reward function

The final item in the formulation of the problem is the definition of a reward function, which defines the goals and/or behaviour of the vehicles. Since a time-iterative approach is used, it was possible to use a reward function where one or multiple states are defined as the goal, so as to obtain a policy that drives the ego-car towards this state. The reward function used is defined as per Equation 2.

$$Reward = \begin{cases} pos\_x, & \text{if } pos\_x > ramp\_len \\ & \text{and } pos\_y = goal\_lane \\ -350, & \text{if } constraint\_violation \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

For the merge scenario, the main goal will be to safely pass the bottleneck and get incorporated in the main lane

of the road, therefore, the goal state will be any position in the main (left) lane after the bottleneck. To ensure this goal, only three different values are used for the reward: If there is a constraint violation, such as a collision with other vehicles, or a movement towards a forbidden state (out of the road), a very low reward value is assigned. If the car is able to get to the goal lane, and pass the bottleneck, then its reward will be positive and it will increase as it moves forward so as to encourage faster movements. Finally, if the car has not merged, but is still on a valid state, a reward of zero is assigned.

IV. PROPOSED SOLUTION

The solution of the decision making problem presented here is based on the use of an Iterative Semi Network-Form Game. This type of game, which was first introduced by *Lee et al.* [6], combines Bayesian networks and Game Theory in order to generate a model where multiple human and autonomous components can interact. These form of games have been used to model cyber-physical security systems [16] and to predict pilot behaviours [17].

An iterative semi network-form game uses random variables and a probabilistic framework to represent the different components of the system. There are two types of nodes in this network, firstly the chance or Estimated Action nodes, which use a pre-defined fixed conditional probability distribution. The second type are the Decision Nodes, they represent the human players and their behaviour is learned or optimized. Each decision node requires a reward function (See Equation 2) and then game theory approach is applied in order to optimize their policy. An overview of the proposed iterative semi network-form game is presented in Figure 4.

A multi-player game, with up to 6 players, is proposed, however additional vehicles, following simple rules can be added to the simulation. For each of the players, a set of Estimated Action nodes is used to estimate the likely actions that the vehicles in the surroundings of the player may take at any given time. Each player then uses those possible actions to obtain an estimation of the new state, which will represent its near future status. Having this, it is possible for each player to tune its Decision Node's policy accordingly. The process to optimize that decision making policy is briefly described in Section IV-A.

A. Level-K Reinforcement Learning

In order to determine which is the best action to take in each case, a hierarchical or level-k reasoning process was used. The objective is to make meaningful predictions of the outcomes of the games assuming that the other players will use a basic (level-0) strategy, and then optimize the policy (i.e. conditional probability distribution) using that estimation, so as to obtain a higher level policy and so on, as aforementioned, this type of level-k solutions have been proved to produce results coherent with human behaviour.

Finding the optimal equilibrium for multiple players, in an iterative game using methods such as Nash equilibrium or quantal response equilibrium, may not be feasible because

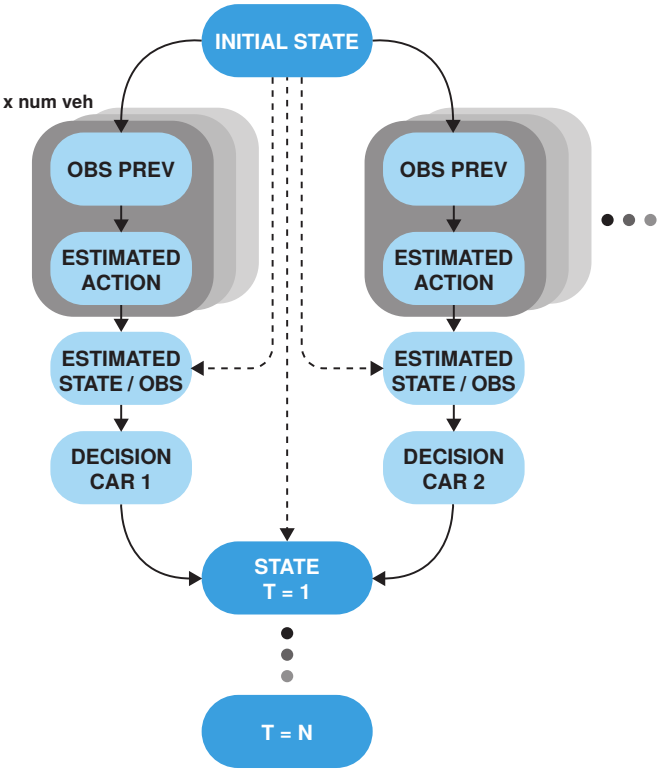


Fig. 4: One time-step and two player example of the proposed semi network-form game. Each car generates an estimation of the movements of the other players and makes the decision accordingly.

those methods become extremely complicated when a high number of player and possible actions is involved. Therefore, it is necessary to use a different approach. In this case, the solution is based on reinforcement learning.

The system does not consider every possible combination of actions, instead it assumes that the players define a single policy for all the network, and then execute this policy over all iterations of the network, thus making the computational complexity independent of the number of time-steps. Also, by defining a single policy, it is not necessary to sample for different actions of the players, but rather it is possible to sample for different policies.

Once a policy has been defined, its utility is obtained by adding the rewards of every time step. Using this utility value, an optimization problem can be defined by trying to find the policy that leads to the highest expected infinite sum of discounted rewards. This optimization problem is solved using a Monte Carlo reinforcement learning approach, which searches for a near-optimal policy and can be seeded with lower level policies to improve its results.

V. EXPERIMENTS DESCRIPTION

This section describes the test performed in order to validate the decision making process and to evaluate the performance of the system itself. Firstly, the scenario, its initialization and the overall procedure of the simulation is explained. Then, the basic behaviour, used by level 0 players

is explained, and finally a comparison of the performance of different level players is carried out.

A. Scenario set-up

The scenario used for the experiments is based on the world model presented on Figure 2. It is composed of a two-lane road, each lane has a width of $3.2\,\mathrm{m}$. The left lane is continuous and unlimited in distance, whereas the right lane ends after a distance of $25\,\mathrm{m}$. The vehicles are assumed to drive on the centre of the lane, and they only move laterally to perform a lane change manoeuvre.

Since the scenario modelled is a traffic jam, the maximum velocity of the cars is limited to $5\,\mathrm{m\,s^{-1}}$. The acceleration of the vehicles is $2.5\,\mathrm{m\,s^{-2}}$ and the breaking action decelerates the vehicle at $5\,\mathrm{m\,s^{-2}}$.

The initialization of the vehicles is as follows: the ego-car and one more vehicle are placed on the merging ramp and the rest of the vehicles on the main lane. The vehicles located on the same lane are separated by a distance of $2.5\,\mathrm{m}$, and all vehicles are assigned an initial speed of $2.5\,\mathrm{m\,s^{-1}}$.

B. Simulation work-flow

Once the vehicles are assigned their initial position and speed, the main loop starts. The simulation is executed for $10\,\mathrm{s}$, and as aforementioned, the movements have discrete time steps of $1\,\mathrm{s}$. At each time step, each vehicle obtains an estimation of the actions of the vehicles in its surroundings. Then, using those estimated actions, their observations are updated, and the best action is selected accordingly. Each vehicle performs only one action per time step. Finally all selected actions are applied simultaneously and the state of each car is updated according to Equation 1. After the new state has been obtained, the reward function (See Section III-D) is computed, and the process continues with a new time-step.

C. Level 0 Behaviour

A set of basic rules was defined as the Level 0 policy. These rules generate a behaviour that can be used as reference for training as well as a point of comparison with the higher level policies found by the reinforcement learning algorithm.

The main idea of that behaviour is that all vehicles will continue on their lane until it is not possible to move forward (*i.e.* when they reach the end of the ramp), and then they will wait to have a clear space, change lane and continue their movement. This Level 0 policy is achieved by following the next rules:

$$brake = \begin{cases} \text{if } vel\_x > 0 \text{ and } F = \textit{forbidden} \\ \text{if } vel\_x > 0 \text{ and } F = \textit{vehicle static} \\ \text{if } \textit{any collision} \end{cases}$$

$$accel = \begin{cases} \text{if } vel\_x = 0 \text{ and } F = \textit{free} \end{cases}$$

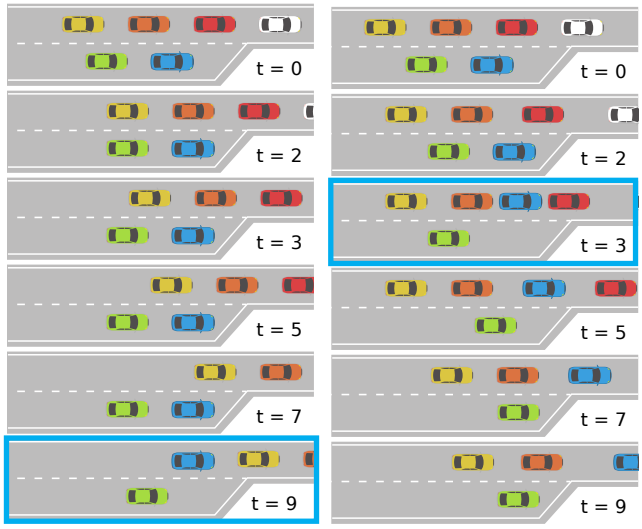
$$left = \begin{cases} \text{if } vel\_x = 0,\ RL = \textit{free} \text{ and } F = \textit{forbidden} \end{cases}$$

$$remain = \begin{cases} \textit{elsewhere} \end{cases}$$

Using this basic policy, all vehicles are capable of merging if enough time is given. For the case of the ego-vehicle, in the $10\,\mathrm{s}$ of the simulation, it will be able to place immediately after the bottleneck. Moreover, additional vehicles, not participating on the game, are controlled using this policy.

VI. RESULTS AND EVALUATION

In order to have a qualitative evaluation of the decision making process, an example of the trajectories obtained using the proposed decision making system, as well as the reference basic behaviour, are presented in Figure 5, the main difference, as shown in the image, is that using the rule based approach the ego-car will always wait for a safe space in order to merge, in this case at time $t = 9\,\mathrm{s}$, whereas using the game theoretical approach, the ego-car can anticipate that the orange vehicle will break if it merges, and as shown, it correctly manages to do so at a much earlier time, at $t = 3\,\mathrm{s}$.



(a) Ruled based (Level 0) (b) Game Theoretic (Level 2)

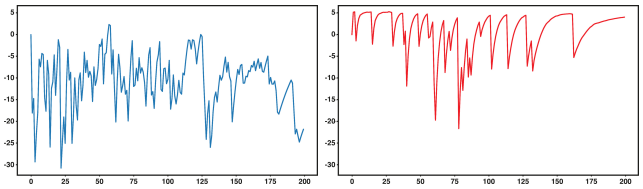
Fig. 5: Sequence of movements using rule-based and game theoretic approaches. The ego-car is shown in blue and the instant of the merge manoeuvre is highlighted.

A. Resulting policy evaluation

Since obtaining a Nash equilibria for the resulting policy (*i.e.* probability distribution), is not possible in this case. It is necessary to use a different technique to provide a formal evaluation of the quality of the obtained policies.

This evaluation is obtained by using a Monte Carlo Importance Sampling estimation of the Predictive Game Theory (PGT) model [18]. This sampling provides a quantification of the level of rationality inherent in a player's behaviour [19]. It uses as reference a welfare function, which in this case is the same as the reward function defined in Section III-D, where negative values represent collisions or undesired states, whereas, positive values show that the car was able to pass the bottleneck, moreover, larger positive rewards mean that the given car reached a further position, and since the

simulation time is fixed, it also means that it was capable of crossing faster.



(a) Level 1 (b) Level 2

Fig. 6: Level 1 vs level 2 average welfare, from Importance Sampling estimation.

The resulting values of the importance sampling are shown in Figure 6. For both, level 1 and level 2 players, the expected welfare can vary highly, this is because using the game theoretical approach, there may be risky situations, and thus it may lead to lower rewards. As can be seen in Fig. 6a, for the Level 1 player, most of the expected rewards are negative, meaning that there is a considerable probability of collisions. However, when training a higher level policy (Fig 6b), it can be seen that, although there may be still some collisions, most of the rewards is positive, and is overall higher than the reward obtained by the level 1 policy, thus proving the efficiency of the decision making to find a solution to this scenario.

VII. CONCLUSIONS

This paper presents a possible solution for a very complex task of autonomous vehicles: the merge manoeuvre in very high traffic scenarios. The solution does not require any specific cooperation or vehicle-to-vehicle communication, and it is capable of anticipate the reactions of other vehicles to specific actions of the ego-car.

The proposal uses an iterative, multi-player level-k game, and finds a near-optimal solution using Monte Carlo reinforcement learning techniques. Moreover, the set of observation, possible actions and reward proposed, can be applied to different situations, and it can yield goal-oriented behaviours.

A kinematic simulation were used in order to test the decision making algorithm and a pre-defined rule-based policy was used for comparison. The results show the capacity of the proposed methodology to solve the task by taking actions that may be consider unsafe unless they take into account the reactions of the other players.

ACKNOWLEDGMENT

This work was funded under project CAMPUS (Connected Automated Mobilty Platform for Urban Sustainability) sponsored by Programme d'Investissements d'Avenir (PIA) of french Agence de l'Environnement et de la Maîtrise de l'Énergie (ADEME).

REFERENCES

[1] N. Evestedt, E. Ward, J. Folkesson, and D. Axehill, "Interaction aware trajectory planning for merge scenarios in congested traffic situations," in *2016 IEEE International Conference on Intelligent Transportation Systems (ITSC)*, Nov 2016, pp. 465–472.

[2] H. Kita, "A merging–giveway interaction model of cars in a merging section: a game theoretic analysis," *Transportation Research Part A: Policy and Practice*, vol. 33, no. 3, pp. 305 – 312, 1999. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0965856498000391

[3] F. Meng, J. Su, C. Liu, and W. Chen, "Dynamic decision making in lane change: Game theory with receding horizon," in *2016 UKACC 11th International Conference on Control (CONTROL)*, Aug 2016, pp. 1–6.

[4] C. F. Camerer, T.-H. Ho, and J.-K. Chong, "A Cognitive Hierarchy Model of Games*," *The Quarterly Journal of Economics*, vol. 119, no. 3, pp. 861–898, 08 2004. [Online]. Available: https://doi.org/10.1162/0033553041502225

[5] M. A. Costa-Gomes and V. P. Crawford, "Cognition and behavior in two-person guessing games: An experimental study," *American Economic Review*, vol. 96, no. 5, pp. 1737–1768, December 2006. [Online]. Available: http://www.aeaweb.org/articles?id=10.1257/aer.96.5.1737

[6] R. Lee, D. H. Wolpert, J. W. Bono, S. Backhaus, R. Bent, and B. Tracey, "Counter-factual reinforcement learning: How to model decision-makers that anticipate the future," *CoRR*, vol. abs/1207.0852, 2012. [Online]. Available: http://arxiv.org/abs/1207.0852

[7] D. Marinescu, J. Čurn, M. Bouroche, and V. Cahill, "On-ramp traffic merging using cooperative intelligent vehicles: A slot-based approach," in *2012 IEEE International Conference on Intelligent Transportation Systems (ITSC)*, Sep. 2012, pp. 900–906.

[8] V. Milanes, J. Godoy, J. Villagra, and J. Perez, "Automated on-ramp merging system for congested traffic situations," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 2, pp. 500–508, June 2011.

[9] S. Brechtel, T. Gindele, and R. Dillmann, "Probabilistic decision-making under uncertainty for autonomous driving using continuous pomdps," in *2014 IEEE International Conference on Intelligent Transportation Systems (ITSC)*, Oct 2014, pp. 392–399.

[10] M. Düring, K. Franke, R. Balaghiasefi, M. Gonter, M. Belkner, and K. Lemmer, "Adaptive cooperative maneuver planning algorithm for conflict resolution in diverse traffic situations," in *2014 International Conference on Connected Vehicles and Expo (ICCVE)*, Nov 2014, pp. 242–249.

[11] A. G. Cunningham, E. Galceran, R. M. Eustice, and E. Olson, "Mpdm: Multipolicy decision-making in dynamic, uncertain environments for autonomous driving," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, May 2015, pp. 1670–1677.

[12] D. Sadigh, S. S. Sastry, S. A. Seshia, and A. D. Dragan, "Planning for autonomous cars that leverage effects on human actions," in *Robotics: Science and Systems*, 2016.

[13] Y. Rasekhipour, A. Khajepour, S. Chen, and B. Litkouhi, "A potential field-based model predictive path-planning controller for autonomous road vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 5, pp. 1255–1267, May 2017.

[14] N. Li, D. W. Oyler, M. Zhang, Y. Yildiz, I. Kolmanovsky, and A. R. Girard, "Game theoretic modeling of driver and vehicle interactions for verification and validation of autonomous vehicle control systems," *IEEE Transactions on Control Systems Technology*, vol. 26, no. 5, pp. 1782–1797, Sep. 2018.

[15] M. Garzón and A. Spalanzani, "An hybrid simulation tool for autonomous cars in very high traffic scenarios," in *2018 International Conference on Control, Automation, Robotics and Vision (ICARCV)*, Nov 2018, pp. 803–808.

[16] S. Backhaus, R. Bent, J. Bono, R. Lee, B. Tracey, D. Wolpert, D. Xie, and Y. Yildiz, "Cyber-physical security: A game theory model of humans interacting over control systems," *IEEE Transactions on Smart Grid*, vol. 4, no. 4, pp. 2320–2327, Dec 2013.

[17] Y. Yildiz, A. Agogino, and G. Brat, "Predicting pilot behavior in medium-scale scenarios using game theory and reinforcement learning," *Journal of Guidance, Control, and Dynamics*, vol. 37, no. 4, pp. 1335–1343, 2019/04/12 2014. [Online]. Available: https://doi.org/10.2514/1.G000176

[18] D. H. Wolpert and J. W. Bono, "Distribution-valued solution concepts," *Review of Behavioral Economics*, vol. 1, no. 4, pp. 381–443, 2014. [Online]. Available: http://dx.doi.org/10.1561/105.00000015

[19] D. H. Wolpert, "A predictive theory of games," *arXiv preprint nlin/0512015*, 2005.