Robust Game-Theoretic Decision-Making for Autonomous Vehicles

***Abstract*—**

I. Introduction

Despite many recent advances in connected and automated vehicles (CAVs) technology, the full automation systems that can provide the similar or better ability compared to the human drivers are still inherently flawed to be deployed in the market [1]. One of the most significant challenges is to plan the motion of the automated vehicles at the mixed-traffic conditions where the automated car coexists with the human-driven vehicles [2]. In particular, describing the human decision-making process is the most difficult since the humans do not always exhibit the optimized behavior due to limited rationality [3]. In order to ensure the safety, a conservative driving policy of the automated vehicles where the all possible driving situations are considered has been suggested [4], [5], but it may cause the disharmony with the human drivers. That is, too conservative behavior of the automated vehicles sometimes causes adverse effect on traffics, e.g., road congestion, car accidents, due to the uncertainties of the human drivers.

If the human-driven vehicle's response according to traffics, i.e., behavior of the surrounding vehicles, can be predicted, far less conservative motion planning of the automated vehicle can be designed [6]. As mentioned, however, modeling of the interaction between the vehicles may not very accurate. Moreover, since the communication time with the surrounding vehicles in reality is not sufficient to build the deterministic human driver models.

To overcome such shortcomings, we exploit the "level-k game theory" framework where the human thought processes in strategic games, i.e., driving at the mixed-traffics, are modeled in hierarchical structure [7]. The game theoretic approach has already proved its effectiveness when describing the interactions between the vehicles [8], [9], [10]. Although these approaches effectively describe rational decision-making for the human-driven vehicles, the model uncertainties that affect to the vehicle position are not considered. Therefore, it is assumed that the center of gravity of the vehicle follows the deterministic kinematic model with the constant safety constraint that prevents the physical safety violation between the vehicles, and depending on the size of the safety constraint, the level of conservatism in the autonomous vehicle's motion is determined.

In this paper, we adaptively adjust the safety constraint sizes of the interactive human-driven vehicles according to the confidence level to establish a balanced motion planning between the aggressive and conservative motion planning. The interactive vehicle's aggressiveness level is estimated and it is proportional to the size of safety constraint of the human-driven vehicle. Based on this, the autonomous vehicle's future trajectory is planned, which maximizes the cumulative reward. In [11], the caution level of the human-driven vehicle is used to describe the aggressiveness of the HDV, but only constant caution level is assumed, which eventually determine the level of conservatism of AV. To resolve this, we incorporate the level-k game theory with the caution level so that adaptive caution level is available, which enables adjustable motion planning of AV depending on the interactive vehicle's aggressiveness.

The contributions of this paper are summarized as follows. First and foremost, a balanced motion planning of AV is (Contributions is not stated yet)

The rest of the paper is organized as follows. In Section II, the vehicle dynamic model with the bounded uncertainty is introduced. We next present the game-theoretic decision making algorithm and motion planning procedure in Section III, the effectiveness of the proposed approach is described in Section IV, and the paper is concluded in Section V.

red: not super-clear, blue: added or modified.

A. Notation

II. Probelm statement

As illustrated in Fig. 1, depending on the motions of interactive vehicles (red), the autonomous vehicle can plan the different motions. If the interactive vehicles behave aggressively, the AV should choose the conservative motion planning, and vice versa. In addition, to avoid the too conservative or aggressive driving policy, the balanced motion planning is preferred in the actual application.

The problem we treat is to model the intent of the interactive vehicles using game theoretical approach, then AV exploits the estimated interactive vehicle's driving intention when making the decision in real-time. The confidence level of the estimated intention is represented by dashed-dotted boxes in Fig. 1 and their sizes are adjusted adaptively according to the confidence levels.

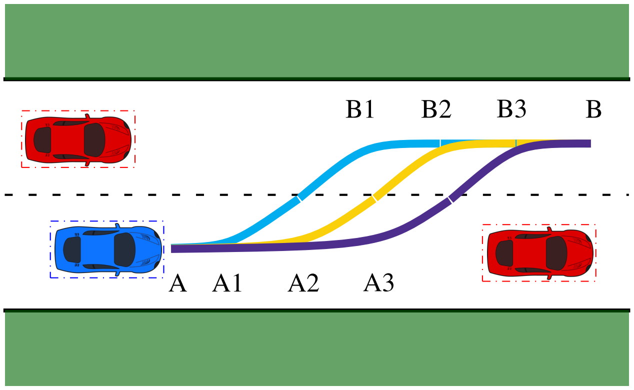

Fig. 1. Possible lane changing sequences; "aggressive motion planning"= {(A,A1), (A1,B1), (B1,B)}, "balanced motion planning"=$\{(A, A2), (A2, B2), (B2, B)\}$ and "conservative motion planning"=$\{(A, A3), (A3, B3), (B3, B)\}$ that can be chosen by the autonomous vehicle (blue) based on the motion of other non-autonomous vehicles (red).

III. Vehicle dynamic model

The vehicle dynamics are represented by the following discrete model [12]

$$x(t+1) = x(t) + v(t)\cos(\psi(t) + \beta(t))\Delta t + w_x(t) \tag{1a}$$

$$y(t+1) = y(t) + v(t)\sin(\psi(t) + \beta(t))\Delta t + w_y(t) \tag{1b}$$

$$\psi(t+1) = \psi(t) + \frac{v(t)}{l_r}\sin(\beta(t))\Delta t \tag{1c}$$

$$v(t+1) = v(t) + a(t)\Delta t \tag{1d}$$

$$\beta(t) = \arctan\left(\frac{l_r}{l_r + l_f}\tan(\delta_f(t))\right), \tag{1e}$$

where t denotes the discrete time instant; the pair $(x(t), y(t))$ represent the global position of the center of mass of the vehicle; the vehicle's speed is denoted by $v(t)$; $\beta(t)$ is the angle of $v(t)$ with respect to the longitudinal axis of the vehicle; $\psi(t)$ denotes the vehicle's yaw angle (the angle between the vehicle's heading direction and the global x-direction); $a(t)$ denotes the vehicle's acceleration at time t; Δt denotes the time step size; $\delta_f(t)$ represents the front steering angle; and l_f and l_r are the distance of the center of the mass of the vehicle to the front and rear axles, respectively; $w_x(t)$ and $w_x(t)$ denote the uncertainty in the position of the center of mass, respectively. It is assumed the uncertainties originate from a closed and compact disturbance set, $\mathcal{W} := \left\{w = (w_x, w_y) \,|\, \zeta w \le \theta,\, \zeta \in \mathbb{R}^{a\times 2},\, \theta \in \mathbb{R}^b\right\}$, with $b \in 2\mathbb{Z}^+$. The disturbance set is assumed to contain the origin. Furthermore, it is assumed that the rear wheels cannot be steered. Therefore, the control input to the model (1), represented by $\gamma(t) = (a(t), \delta_f(t))$, is the acceleration and front steering angle pair.

IV. Game-theoretic decision making

At each time instant, each vehicle selects an input pair from the finite action set, $\mathbf{\Gamma} = \{(0, 0), (0, \delta_{f,\text{nom}}), (0, -\delta_{f,\text{nom}}), (a_{\text{nom}}, 0), (-a_{\text{nom}}, 0), (a_{\max}, 0), (-a_{\max}, 0), (a_{\text{nom}}, \delta_{f,\max}), (a_{\text{nom}}, -\delta_{f,\max})\}$, where a_{nom}, $\delta_{f,\text{nom}}$ and $a_{\max}$, $\delta_{f,\max}$ are the nominal and maximum acceleration, front steer angle, respectively. The inputs pairs in $\mathbf{\Gamma}$ correspond to the actions, {"maintain", "turn slightly left", "turn slightly right", "accelerate", "decelerate", "maximum acceleration", "maximum deceleration", "turn left and accelerate", " turn right and accelerate" }, respectively. The input pair to be applied at every time step is decided based on optimizing a reward function.

A. Action choice

The decision making process of the vehicle in choosing the optimal input pair follows a receding horizon strategy. A sequence of actions, $\boldsymbol{\gamma_t} = \{\gamma_t, \gamma_{t+1}, \cdots, \gamma_{t+N-1}\}$, is chosen that maximizes a cumulative reward given by

$$\mathcal{R}(\boldsymbol{\gamma_t}) = \sum_{j=0}^{N-1} \lambda^{j-1} R_{t+j}(\gamma_{t+j}), \tag{2}$$

where $R_{t+j}(\gamma_{t+j})$ is the stage reward at a prediction step j determined at time step t for an input, $\gamma_{t+j} \in \mathbf{\Gamma}$; $\lambda \in [0, 1]$ is the discount factor. By the receding horizon strategy, the input applied to (1), $\gamma(t)$, is the first element of $\boldsymbol{\gamma_t}^* = \left\{\gamma_t^*, \gamma_{t+1}^*, \cdots, \gamma_{t+N-1}^*\right\}$ is applied at each time instant t, i.e., $\gamma(t) = \gamma_t^*$. The stage reward at a prediction step j, $R_{t+j}(\gamma_{t+j})$, is defined as

$$R_{t+j}(\gamma_{t+j}) = R_{t+j}(\gamma_{t+j}|s_{t+j}) = \boldsymbol{\alpha}^T \boldsymbol{\phi_{t+j}} \tag{3}$$

where s_{t+j} is the traffic state at prediction step j; $\boldsymbol{\phi_{t+j}} = \{\phi_{1,t+j}, \phi_{2,t+j}, \cdots, \phi_{m,t+j}\}$ is the feature vector at step j and the weights for these features are in $\boldsymbol{\alpha} = \{\alpha_1, \alpha_2, \cdots, \alpha_m\}$, in which $\alpha_i > 0, \forall i \in \mathbb{Z}_{[0:m]}$. For the lane changing scenario in Fig. 1, the features considered are described below.

Rectangular outer approximation of the geometric contour of each vehicle is considered as shown by the dash-dotted boxes in Fig. 1. This outer approximation is referred as the collision avoidance zone (c-zone). The features, $\phi_{1,t}$, $\phi_{2,t}$ and $\phi_{3,t}$, are indicator functions based on the c-zone of the vehicles that respectively characterize:

- Collision status - The intersection of the c-zone of the ego vehicle with that of any other vehicle indicates a collision or a danger of collision. If an overlap is detected then $\phi_{1,t}$ is assigned a value -1; and 0, otherwise.
- On-road status - The intersection of the c-zone of the ego vehicle with that of green regions shown in Fig. 1 indicates that the ego vehicle is outside the road boundaries. The feature $\phi_{2,t} = -1$ if an overlap is detected; $\phi_{2,t} = 0$, otherwise.
- Safe zone violation status - A safe zone (s-zone) of a vehicle is a rectangular area that subsumes the c-zone of the vehicles with a safety margin. The safety margin is chosen based on the minimum distance to be maintained from the surrounding vehicles. If an overlap of the s-zone

of the ego vehicle with that of another vehicle is detected then $\phi_{3,\,t}$ is assigned a value -1; and 0, otherwise.

The other features considered in this work characterize:

- Distance to objective - In order to encourage the ego vehicle to change lane and reach a reference point in the new lane, $\left(x^{\text{ref}},\, y^{\text{ref}}\right)$, the feature $\phi_{4,\,t}$ is defined as

$$\phi_{4,\,t} = -\left(\left|x_t - x^{\text{ref}}\right| + \left|y_t - y^{\text{ref}}\right|\right). \tag{4}$$

- Distance to lane center - The feature, $\phi_{5,\,t}$, defined as

$$\phi_{5,\,t} = -\left|y_t - y^{lc}\right|, \tag{5}$$

  where y^{lc} is the y-coordinate of the center of the current lane that is included to encourage the ego vehicle to be at the middle of the current lane.
- Velocity error - The deviation of the velocity of the ego vehicle from a reference velocity, v^{ref}, is described by the feature $\phi_{6,\,t}$ as

$$\phi_{6,\,t} = -\left|v_t - v^{\text{ref}}\right|, \tag{6}$$

  where the reference velocity is typically chosen as the legislated speed limit.

B. Level-k framework

In a multi-agent traffic scenario, the interactive nature of the decision making process is captured by the features defined in the previous subsection.

features of the reward function take into account the states of the other agents in the

Based on the definitions of the features, the rewards of a vehicle not only depend on its own states and actions, but also depend on the states and actions of its opponent vehicle . Such an interdependence reflects the interactive nature of vehicle decision making in a multi- vehicle traffic scenario.

To compute the cumulative reward in (2), the state of the traffic, s_{t+j}, at prediction steps $j = 0, 1, \cdots, N-1$, are predicted using level-k game theoretic model.

have to be known to compute the .

, where

$$\boldsymbol{\gamma_t}^* = \arg\max_{\boldsymbol{\gamma_t}\in\boldsymbol{\Gamma}} \mathcal{R}\left(\boldsymbol{\gamma_t}\right), \tag{7}$$

We use a game theoretic model for this prediction.

Numerous experimental results from psychology, cognitive science, and economics have suggested a hierarchical structure in human reasoning in games, see [13], [14], [15], [16]. The study of this reasoning hierarchy and its applications in game theoretic settings are addressed by the "level-k game theory." In [5], [6], [17], the level-k game theory is exploited to model vehicle interactions in highway traffic. The model has been compared to human traffic data in [6]. In this paper, we also exploit level-k game theory to model vehicle interactions, in particular, at intersections. Recently, level-k modeling of human agents was also considered in aerospace and energy applications [18], [19], [20], where human-to- human and human-to-automation interactions play a central role.

The model is premised on the idea that strategic agents (drivers/vehicles at an intersection, in our setting) have different reasoning levels. In particular, the level k indicates an agent's reasoning depth. The reasoning hierarchy starts from level-0. A level-0 agent makes instinctive decisions to pursue its goal without considering the interactions between itself and the others. On the contrary, a level-1 agent takes into account such interactions in its decision making process, in particular, by assuming that all the other agents in the game are level-0. Specifically, a level-1 agent assumes that all the other agents are level-0 so they make instinctive decisions; the level-1 agent predicts their actions as well as the evolutions of the game resulting from their actions based on this assumption; the level-1 agent then makes its own decision as the best response to such evolutions to pursue its own goal. Similarly, a level-k agent assumes that all the other agents are level-(k-1), makes predictions based on this assumption, and responds accordingly.

$$u_l^K = \arg\max_{u_l\in\mathcal{U}} \sum_{j=0}^{N-1} \gamma^j R\left(x_l\left(j+1\right), x_{i\neq l}\left(j+1\right), u_l\left(j\right), u_{i\neq l}^{K-1}\left(j\right)\right)$$

C. Driver model identification

$$\begin{aligned}
\tilde{K}\left(t\right) &= \arg\min_{K\in\{0,\,1\}} \left\|u_i^{\text{actual}}\left(t\right) - u_i^K\left(t\right)\right\| \\
P_{K_i=k}\left(t\right) &= P_{K_i=k}\left(t-1\right) + I_{\tilde{K}_i(t)=k}\Delta P \\
P_{K_i=k}\left(t\right) &= \frac{P_{K_i=k}\left(t\right)}{\sum_{k'=0}^{1} P_{K_i=k'}\left(t\right)}
\end{aligned}$$

Modified control policy taking model identification into account

$$u_l^D = \arg\max_{u_l\in\mathcal{U}} \sum_{k=0}^{1} P_{K_i=k}\left(t\right) \left[\sum_{j=0}^{N-1} \gamma^j R\left(x_l\left(j+1\right), x_{i\neq l}\left(j+1\right), u_l\left(j\right),\right.\right.$$

D. Robust approach

$$\mathcal{W}_i' = P_{K_i=0}\cdot\mathcal{W}_i$$

$$\tilde{u}_l^D = \arg\max_{u_l\in\mathcal{U}} \min_{w_x,\,w_y\in\mathcal{W}_i'} \left[\sum_{k=0}^{1} P_{K_i=k}\left(t\right) \left[\sum_{j=0}^{N-1} \gamma^j R\left(x_l\left(j+1\right), x_{i\neq l}\left(j\right.\right.\right.\right.$$

V. Simulation results

The effectiveness of the proposed approach is verified through the case studies. Starting from the same initial conditions, the AV (blue car) switches to the left lane and the HVs (red cars) keep the lane as illustrated in Fig. 2. We specify the level-1 decision making strategy to all HVs. That is, cautious drivers are assumed, but unknown to AV.

The five panels on the left in Fig. 2(a) shows the behavior of each vehicle at 1 second intervals when the AV changes the lane aggressively. The uncertainty size from the perspective of AV is illustrated by red dashed line. In this case study, there is almost no safety margin as the AV takes an aggressive lane change strategy. As a result, the left steering input is initiated as soon as the simulation begins and lane change occurs between 60m and 70m. Looking more closely, due to the close distance between AV, the HV slightly moves to left at time $t = 2s$ and $3s$. It is possible to avoid collision because HV is cautious driver, but collision may happen if there is aggressive driver. This aggressive strategy should not be adopted because a certain degree of conservatism is preferred rather than car crashes.

We next illustrate the results when adopting the conservative lane change strategy as shown in Fig. 2(c). Unlike the previous case, the large uncertainty

VI. Conclusions

Acknowledgement

We would like to express out appreciation to Nan Li who provided expertise that greatly assisted the research and it should be noted that the code publicly available (https://github.com/gokulsivasankar/RobustDecisionMaking) is based on his previous study [9].

References

[1] R. Okuda, Y. Kajiwara, and K. Terashima, "A survey of technical trend of adas and autonomous driving," in *Technical Papers of 2014 International Symposium on VLSI Design, Automation and Test*, pp. 1–4, IEEE, 2014.

[2] D. A. Lazar, R. Pedarsani, K. Chandrasekher, and D. Sadigh, "Maximizing road capacity using cars that influence people," in *2018 IEEE Conference on Decision and Control (CDC)*, pp. 1801–1808, IEEE, 2018.

[3] T. L. Griffiths, F. Lieder, and N. D. Goodman, "Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic," *Topics in cognitive science*, vol. 7, no. 2, pp. 217–229, 2015.

[4] L. Claussmann, A. Carvalho, and G. Schildbach, "A path planner for autonomous driving on highways using a human mimicry approach with binary decision diagrams," in *2015 European Control Conference (ECC)*, pp. 2976–2982, IEEE, 2015.

[5] S. Brechtel, T. Gindele, and R. Dillmann, "Probabilistic decision-making under uncertainty for autonomous driving using continuous pomdps," in *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pp. 392–399, IEEE, 2014.

[6] D. Sadigh, S. Sastry, S. A. Seshia, and A. D. Dragan, "Planning for autonomous cars that leverage effects on human actions.," in *Robotics: Science and Systems*, vol. 2, Ann Arbor, MI, USA, 2016.

[7] D. O. Stahl, "Evolution of smartn players," *Games and Economic Behavior*, vol. 5, no. 4, pp. 604–617, 1993.

[8] 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, 2017.

[9] N. Li, I. Kolmanovsky, A. Girard, and Y. Yildiz, "Game theoretic modeling of vehicle interactions at unsignalized intersections and application to autonomous vehicle control," in *2018 Annual American Control Conference (ACC)*, pp. 3215–3220, IEEE, 2018.

[10] R. Tian, S. Li, N. Li, I. Kolmanovsky, A. Girard, and Y. Yildiz, "Adaptive game-theoretic decision making for autonomous vehicle control at roundabouts," in *2018 IEEE Conference on Decision and Control (CDC)*, pp. 321–326, IEEE, 2018.

[11] G. I. Jin, S. Bastian, M. M. Richard, and A. Matthias, "Risk-aware motion planning for automated vehicle among human-driven cars," in *2019 American Control Conference (ACC)*, pp. –, IEEE, 2019.

[12] J. Kong, M. Pfeiffer, G. Schildbach, and F. Borrelli, "Kinematic and dynamic vehicle models for autonomous driving control design," in *2015 IEEE Intelligent Vehicles Symposium (IV)*, pp. 1094–1099, IEEE, 2015.

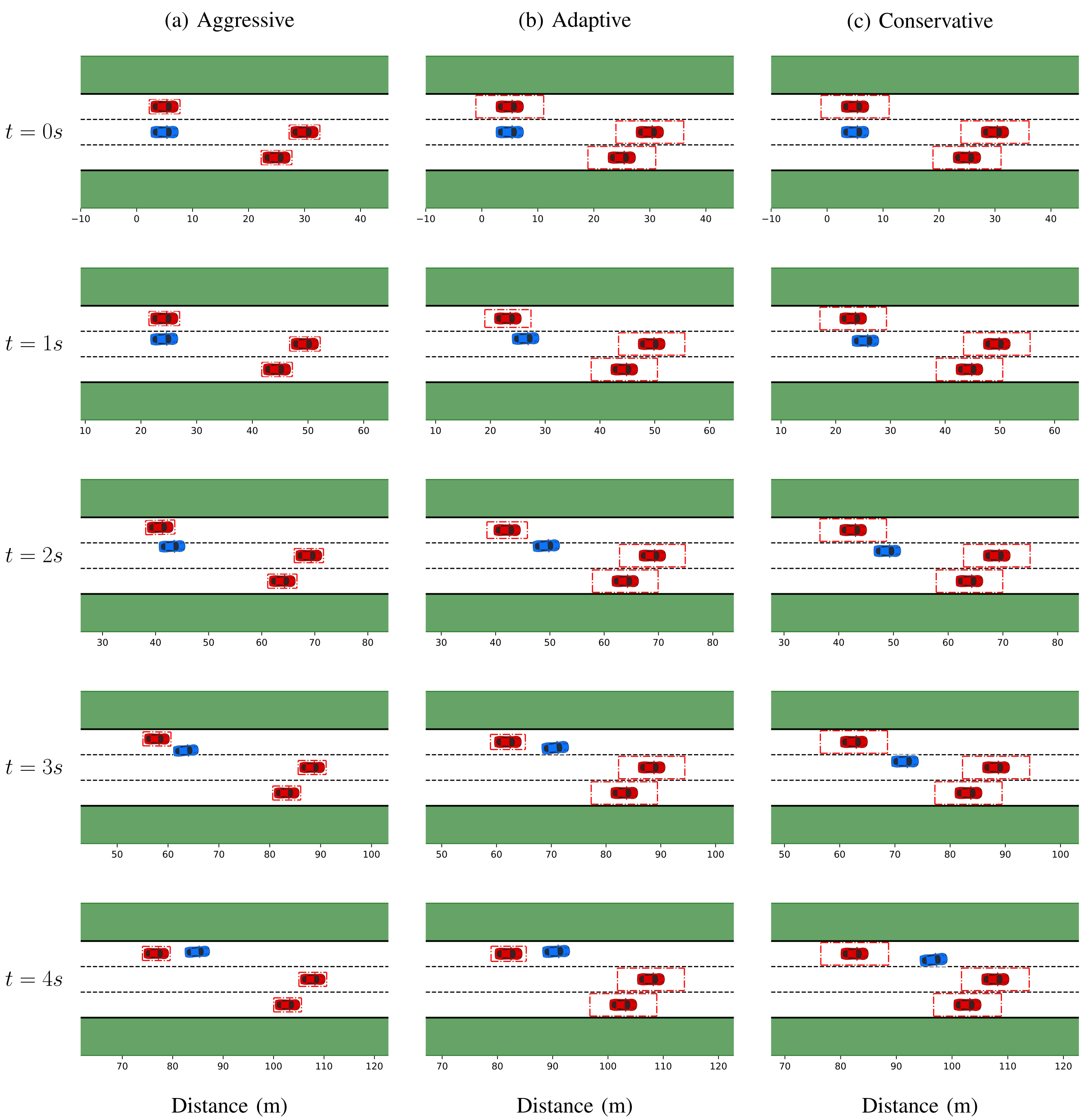
(a) Aggressive
(b) Adaptive
(c) Conservative
$t = 0s$
$t = 1s$
$t = 2s$
$t = 3s$
$t = 4s$
Distance (m)
Distance (m)
Distance (m)

Fig. 2. .

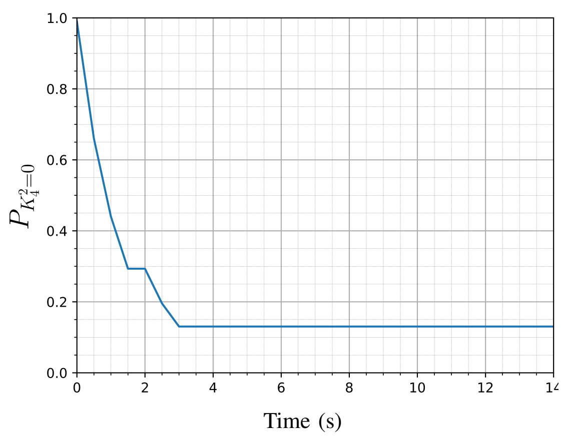

Fig. 3. Probability that vehicle 4 can be modeled as level-0 from the perspective of the ego vehicle.