Negotiation-Aware Reachability-Based Safety Verification for Autonomous Driving in Interactive Scenarios

Ran Tian, Anjian Li, Masayoshi Tomizuka, and Liting Sun

Abstract— **Safety assurance is a critical yet challenging aspect when developing self-driving technologies. Hamilton-Jacobi backward-reachability analysis is a formal verification tool for verifying the safety of dynamic systems in the presence of disturbances. However, the standard approach is too conservative to be applied to self-driving applications due to its worst-case assumption on humans' behaviors. In this work, we integrate a learning-based prediction algorithm and a game-theoretic negotiation model to online update the conservativeness of backward-reachability analysis. We evaluate our approach using real driving data. The results show that our approach can effectively reduce the conservativeness of the standard approach without sacrificing its safety verification ability.**

I. INTRODUCTION

Our society is rapidly advancing towards self-driving cars that interact with human-driven cars on public roads. Safety assurance is a critical yet challenging aspect when designing such autonomous systems. Hamilton-Jacobi (HJ) reachability analysis is a formal safety verification tool for verifying the safety of dynamic systems [1], [2]. Under HJ-reachability paradigm, both forward [3]–[5] and backward-reachability analysis [6]–[8] can be used to ensure safety in human-robot interaction applications. For instance, forward-reachability analysis can be used to compute the human's forward-reachable set (FRS) that should be avoided by the robot. Often, the FRS leads to overly conservative robot behaviors due to its open-loop spirit. Different from forward-reachability analysis, backward-reachability analysis formulates the human-robot interaction as a differential game and computes the full backward-reachable tube (BRT) that contains states that may result in the human-robot joint system being unsafe, assuming the human is adversarial and has full control bound[1], within some time horizon. When operating online, the BRT can either be used as a safety monitor that switches the robot's controller to the optimal safety controller when the human is about to breach the BRT [8] or infused into the robot's low-level controller design as constraints [6], [7]. Backward-reachability analysis is less conservative since it allows the robot to react to the human's adversarial actions. However, it is still impractical for self-driving applications due to its worst-case assumption on humans' behaviors (adversarial human with full control bound). Therefore, mitigation strategies are needed for generating a practical yet effective BRT.

Since the conservativeness of the full BRT is induced by assuming that the human-driven car utilizes its full control bound to drive the joint system to unsafe states, a natural solution is to directly limit the human-driven car's control bound. Such a technique has been exploited in both forward and backward-reachability analysis. For instance, [5] casts the human state prediction problem as a forward-reachability problem. When computing the human's FRS, [5] limits the human's control bound based on his/her likely actions. Despite being robust to mis-specified human models and priors, [5] separates the human from the robot and neglects the effects from interactions. In traffic settings, self-driving cars might get stuck with such an approach. The approach in [8] exploits a learning-based predictor to forecast the human-driven car's future trajectory and classifies the prediction into one of the pre-defined driving modes (e.g., turn left). Each driving mode is associated with a less-conservative human-driven car's control bound, and the corresponding BRT can be computed offline and queried online for safety verification. Such an approach works well if the learning-based predictor offers the human driver's future trajectories based on the self-driving car's future plan candidates, instead of only on map information and historical observations (environmental states). Without that, the estimated human-driven car's control bound can still be conservative in interactive scenarios since it doesn't account for the negotiation between the human-driven car and the self-driving car.

Our key insight is that: *in interactive scenarios, the distribution of human-driven car's actions is not only constrained by environmental states but also affected by the human driver's internal state that governs his/her behaviors in negotiations. Estimating the human driven-car's control bound online with both environmental states and the human driver's internal state accounted for can help to obtain a less-conservative while safety-preserving BRT.*

Overall, we make the following contributions towards practical safety verification for self-driving cars:

Developing a framework for constructing a practical yet effective BRT by leveraging learning-based prediction and game-theoretic reasoning. We propose to hierarchically estimate the human-driven car's control bound. We first utilize a learning-based method to estimate a control bound with the environmental information accounted for, and further modify the bound by reasoning about the human's willingness to cooperate in a game-theoretic fashion. Then the updated control bound is used to query an appropriate BRT online for safety verification.

Verifying the effectiveness of our approach with real driving data. We compare our approach with the standard

Ran Tian, Masayoshi Tomizuka, and Liting Sun are with University of California, Berkeley. ({rantian, tomizuka, litingsun}@berkeley.edu).

Anjian Li is with Simon Fraser University, Burnaby, BC, Canada, V5A 1S6 ({anjianl}@sfu.ca).

[1] An agent has full control bound if it can maneuver freely and is only constrained by the maximal bound of controls that its actuator can provide.

full BRT using real traffic data. We show that our approach is significantly less conservative while maintaining safety verification ability.

II. PRELIMINARIES

A. Hamilton-Jacobi reachability analysis

In this section, we briefly introduce the HJ backward-reachability analysis. More details can be found in [1], [2].
Backward-reachable tube. We consider the evolution of a dynamical system governed by the differential equation: $\dot{s}(t) = f\big(s(t), u(t), d(t)\big)$, where $s \in \mathbb{R}^{n_s}$ denotes the state of the system, $u \in \mathbb{R}^{n_u}$ denotes the bounded control, and $d \in \mathbb{R}^{n_d}$ denotes the bounded disturbance. We assume that we have access to a target set $\mathbb{G}$ that contains the unsafe states of the dynamical system. Then we define the backward-reachable tube (BRT) of the dynamical system ($\mathcal{O}_\tau$) with respect of $\mathbb{G}$ as a set of initial states from which the disturbance can drive the dynamical system to $\mathbb{G}$ within a time horizon $|\tau|$ despite the optimal control efforts, namely,

$$\mathcal{O}_\tau = \{s_0 \in \mathbb{R}^{n_s} : \exists d(\cdot), \forall u(\cdot),$$
$$\exists t \in [\tau, 0], s(t|f, s_0, d(\cdot), u(\cdot)) \in \mathbb{G}\}, \quad (1)$$

where τ is a negative number since the system dynamics are propagated backwards in time.
Level-set method. The BRT $\mathcal{O}_\tau$ can be characterised as a zero sub-level set of a value function $V(s, \tau)$:

$$\mathcal{O}_\tau = \{s \in \mathbb{R}^{n_s} : V(s, \tau) < 0\}, \quad (2)$$

where the value function $V(s, t)$ is the viscosity solution of the following Hamilton-Jacobi Isaacs partial differential equation [9]:

$$\min\{\frac{\partial V(s, t)}{\partial t} + \min_{d \in \mathcal{D}} \max_{u \in \mathcal{U}} \nabla V(s, t)^\top f(s, u, d),$$
$$V(s, 0) - V(s, t)\} = 0, t \in [\tau, 0], \quad (3)$$

and $V(s, t)$ can be computed via dynamic programming [10] with terminal condition $V(s, 0) = l(s)$, where $l(s)$ is defined as $l(s) \leq 0 \iff s \in \mathbb{G}$.

B. Backward-reachability analysis in self-driving settings

System dynamics. We consider pairwise interactions between the self-driving car and each human-driven car in the environment (i.e., we verify the self-driving car's safety with respect to each human-driven car separately). Following [8], we consider a dynamical system that encodes the relative dynamics between the self-driving car and the human-driven car in a pairwise interaction. Specifically, the dynamics of the human-driven car is represented using an extended uni-cycle model and the dynamics of the self-driving car is modeled using a high-fidelity bicycle model. The state of the relative dynamic system is defined as $s = \big(x_{rel}, y_{rel}, \psi_{rel}, v_r, v_h\big)$, where x_{rel} (y_{rel}) is the x(y)-coordinate of the human-driven car in a coordinate frame centered at the geometric center of the self-driving car with x-axis aligned with the heading of the self-driving car, ψ_{rel} denotes the relative heading between the two cars, and v_r (v_h) denotes the speed of the self-driving car (human-driven car). We let $u_r(t) = (a_r(t), \delta_f(t))$ denote the control input of the self-driving car

with a_r denoting the acceleration and δ_f denoting the front wheel rotation, and let $u_h(t) = (a_h(t), \omega_h(t))$ denote the control input of the human-driven car with a_h denoting the acceleration and ω_h denoting angular speed. The evolution of the relative system is then governed by the following differential equation $\dot{s}(t) = f\big(s(t), u_r(t), u_h(t)\big)$:

$$\dot{x}_{rel} = \frac{v_r y_{rel}}{l_r} \sin(\beta_r) + v_h \cos(\psi_{rel}) - v_r \cos(\beta_r),$$
$$\dot{y}_{rel} = -\frac{v_r x_{rel}}{l_r} \sin(\beta_r) + v_h \sin(\psi_{rel}) - v_r \sin(\beta_r),$$
$$\dot{\psi}_{rel} = \omega_h - \frac{v_r}{l_r} \sin(\beta_r),$$
$$\dot{v}_r = a_r,$$
$$\dot{v}_h = a_h, \quad (4)$$

where l_f (l_r) denotes the front (rear) axle length of the self-driving car and β_r is computed via $\beta_r = \tan^{-1}(\frac{l_r}{l_r + l_f} \tan(\delta_f))$.
BRT as a safety monitor. We treat the human-driven car's control as disturbance of the relative dynamical system (4), then we can compute the BRT that contains initial relative states that could potentially lead to collisions. Such a BRT can be used as a safety monitor that switches the self-driving car's low-level controller to the optimal safety controller when the current relative state is within the BRT [8] or can be infused into the self-driving car's low-level controller design as constraints [6], [7].
Disturbance bound. It can be observed from (4) and (3) that the conservativeness of the BRT highly depends on the human-driven car's control bound $\mathcal{D}$. The BRT computed with the full human-driven car's control bound is overly conservative in general.

III. NEGOTIATION-AWARE BRT

In this section, we propose a negotiation-aware BRT computation approach that estimates the human-driven car's control bound based not only on environmental information but also on the potential negotiation with the self-driving car. We propose to estimate such a bound hierarchically: 1) first utilize a learning-based method to estimate a control bound with the environmental information accounted for, and further modify the bound by reasoning about the internal state of the human driver in a game-theoretic fashion.

A. Estimate disturbance bound via learning-based predictor

As in [8], one way to obtain a practical estimate on $\mathcal{D}$ is to leverage a learning-based prediction algorithm. Typically, a learning-based prediction algorithm will generate the human-driven car's most-likely future trajectory and the statistical prediction error given the map information and the joint history states [11]. Hence, in the first layer, we utilize the most-likely trajectory and the prediction error to generate an initial control bound estimation. More specifically, we construct the Frenet frame taking the most-likely path as the reference path. In the Frenet frame, we can obtain an acceleration bound $\mathcal{D}_a = [a_{lower}, a_{upper}]$ from the most-likely trajectory and an angular speed bound $\mathcal{D}_\omega = [0, 0]$.

The statistical prediction error[2] can further provide additional safety margins to improve the predicted control bound's robustness in the presence of long-tailing problems in learning-based approaches.

B. Negotiation modeling via game-theoretic reasoning

One limitation of the first-layer's estimation approach is that the obtained control bound does not account for negotiations between agents (e.g., a human will slow down if a self-driving car gradually merges in front of him/her). Such a limitation can also lead to a conservative BRT when active maneuvers are needed in interactive scenarios. To tackle this problem, in the second layer, we exploit a game-theoretic approach to explicitly reason about negotiations and correct the predicted human-driven car's acceleration bound. We emphasize that in the second layer, the game-theoretic approach is exploited for its ability to model intense interactions between agents, mimic humans' decision-making behaviors, and reveal their internal states.

Assumptions and simplifications. We assume that the human reacts to the self-driving car via acceleration in the Frenet frame described in Section III-A. In other words, when modeling negotiations, we assume that the human-drive car will operate along its predicted most-likely path (reference path in the Frenet frame). Such an assumption is reasonable in light of two observations: 1) learning-based predictors predict a decent potential path of a human-driven car in structured roads, as they explicitly encode features of road environments during training; 2) a human driver tends to follow a reference path and react to other drivers mainly by adjusting his/her acceleration. Since the current motion plan of the self-driving car is available in the context of safety verification, we further assume that the self-driving car operates along its current planned path when modeling negotiations (such an assumption limits the self-driving car's mobility and tends to make the human's acceleration bound prediction more conservative, but it provides a computational gain by making the negotiation modeling online usable).

Negotiation as a Stackelberg game. With the above assumptions, we model the negotiation between the human-driven car and the self-driving car as a Stackelberg game [12] since it explicitly considers one player's advantages over the other player, which can be used to model different roles in a traffic negotiation. The dynamics of the game is governed by $s_{t+1} = g(s_t, a_t^r, a_t^h)$, where the subscript denotes a discrete time step, a denotes acceleration, and the dynamics function g governs the state evolution along agents' respective paths. We assume that both the human-driven car and the self-driving car have a finite set of acceleration controllers in their negotiation. A controller $\pi \in \Pi$ defines an agent's controls for a planning horizon with length T. At a time step t, an agent's controller π maps a time increment τ ($\tau \in [0, T]$) to the acceleration command at $t + \tau$.

In a negotiation, the human-driven car can either be a *follower* who is willing to cooperate or be a *leader* who aims to dominate the interaction. Specifically, a follower maximizes

[2]Such statistical results can be obtained on the validation sets of the learning algorithms.

his/her own reward function while accommodating the self-driving car. Assuming that humans are noisy-rational, then a follower selects controller in response to the self-driving car's current motion plan according to the following quantal response model:

$$\mathbb{P}(\pi_h|s_t, \pi_r^g) \propto \exp(\beta Q_t^h(s_t, \pi_h, \pi_r^g)), \quad (5)$$

$$Q_t^h(s_t, \pi_h, \pi_r^g) = \sum_{n=0}^{T/\Delta t} R_h(s_{t+n}, \pi_h(n\Delta t), \pi_r^g(n\Delta t)), \quad (6)$$

where π_g^r denotes the self-driving car's current planned controls along its desired path, Δt is the time integral between successive time steps, and R_h is human's reward function learned via Inverse Reinforcement Learning [13]. In contrast to a follower, a leader makes decisions assuming the self-driving car is a follower who optimally responds to his/her controller:

$$\mathbb{P}(\pi_h|s_t) \propto \exp(\beta Q_h^t(s_t, \pi_h, \arg\max_{\pi_r \in \Pi} Q_r^t(s_t, \pi_h, \pi_r)), \quad (7)$$

where the expression of Q_r^t is defined analogously to (6) with reward function R_r.

Leading-following role inference. Since the leading-following role of the human-driven car in a negotiation is unknown, we treat it as an internal state that needs to be inferred online. Normally the self-driving car's sensing module runs at a much higher frequency than the safety verification module. We let ξ_t denote the collection of controls of the human-driven car observed by the self-driving car between step $t-1$ and step t. We let $\theta \in \Theta = \{l, f\}$ denote the human-driven car's leading-following role and define $b_t(\theta) = \mathbb{P}(\theta^* = \theta)$ as the belief distribution over Θ at step t. Then the self-driving car can update the human-driven car's role via Bayesian inference:

$$b_t(\theta) \propto \mathbb{Z}(\xi_t|\theta)b_{t-1}(\theta), \quad (8)$$

where $\mathbb{Z}(\xi_t|\theta)$ is an observation model that specifies the probability of observing ξ_t if the human-driven car's role were θ. We define a distance measure d that evaluates the accumulative difference (Euclidean distance) between the observed human-driven car's controls (ξ_t) and controls obtained from a controller $\pi \in \Pi$ evaluated at the times steps that the observations were made, and use d to find the controller that best matches the observed controls. Then the observation model is defined as:

$$\mathbb{Z}(\xi_t|\theta) = \begin{cases} \mathbb{P}\big(\arg\min_{\pi_h \in \Pi} d(\pi_h, \xi_t)|s_{t-1}, \pi_r^g\big), \text{if } \theta = f, \\ \mathbb{P}\big(\arg\min_{\pi_h \in \Pi} d(\pi_h, \xi_t)|s_{t-1}\big), \text{if } \theta = l, \end{cases} \quad (9)$$

where the probabilities are computed using (5) and (7). With (8) and (9), we can "correct" the human's predicted acceleration bound from the learning-based predictor by explicitly modeling and inferring the human's willingness to be negotiated.

C. Online update the backward-reachable tube.

Since Π is finite, we have a finite set of potential acceleration bounds ($\bar{\mathcal{D}}_a$) of the human-driven car associated with the controllers. For each acceleration bound $\mathcal{D}_a^i$, we offline compute a BRT, $\mathbb{O}_\tau^i$, using the control bound $(\mathcal{D}_a^i, \mathcal{D}_\omega)$, where $\mathcal{D}_\omega$ denotes the human-driven car's angular speed

bound obtained from the learning-based predictor with safety margin described in Section III-A. When operating online, at each step, we first update $b_t(\theta)$ based on the current observation, then run the following algorithm to augment BRTs based on the distribution over the human-driven car's future control bounds with respect to $b_t(\theta)$. The notation Δ denotes a desired confidence bound (we aim to augment BRTs such that their marginal probability is larger than Δ). When $\Delta = 1$, the resulting BRT is equivalent to the one that considers the whole acceleration bound (most conservative). Note that the obtained BRT is in the Frenet frame described in Section III-A.

Algorithm 1: Update BRT

---

1 For each controller $\pi_h^i \in \Pi$, compute its expected probability $\mathbb{P}(\pi_h^i)$ with respect to $b_t(\theta)$;
2 Sort Π with respect to $\mathbb{P}(\pi_h^i)$ in descending order;
3 Initialize $\mathbb{O}_\tau^* \leftarrow \emptyset$, $P = 0$;
4 **for** $i = 1, 2, ..., |\Pi|$ **do**
5     $P = P + \mathbb{P}(\pi_h^i)$; $\mathbb{O}_\tau^* \leftarrow \mathbb{O}_\tau^* \cup \mathbb{O}_\tau^i$;
6   **if** $P >= \Delta$ **then**
7   break;
8   **end if**
9 **end for**
10 Return $\mathbb{O}_\tau^*$;

IV. EXPERIMENT AND RESULT

A. Experiment design

We extracted 15 pairs of two-car interaction trajectories at an uncontrolled roundabout from [14]. For each interaction log, we use a BRT to monitor the safety performance of one selected car (treated as the self-driving car) when it interacts with the other car (treated as the human-driven car). Both cars are set to execute their original controls from the log. When the human-driven car breaches the self-driving car's BRT, the interaction is identified as unsafe, and a safety controller overrides the self-driving car's controls.

Manipulated variables. The manipulated variable is the choice of different algorithms for generating a BRT for safety verification. In addition to our approach that online updates the BRT using the proposed hierarchical method, we also use the full BRT (the one that assumes the worst outcomes) in standard backward-reachability analysis as a baseline.

Dependent measures. Since all the extracted interactions are safe, we quantify the conservativeness and effectiveness of a BRT using two measures: relative speed and relative distance between two cars when the human-driven car breaches the self-driving car's BRT.

B. Implementation details.

We model the acceleration controller in the negotiation modeling (Section III-B) as a second-order polynomial and sample 200 acceleration controllers to construct Π (Fig. 1). When evaluating (5) and (7), we use multi-thread computing to evaluate each controller in parallel so that the negotiation modeling is real-time feasible. The planning horizon T is $2s$, and the confidence bound in Algorithm 1 is $\Delta = 0.9$. We initialize the distribution over the human-driven car's leading-following role as a uniform distribution over Θ.
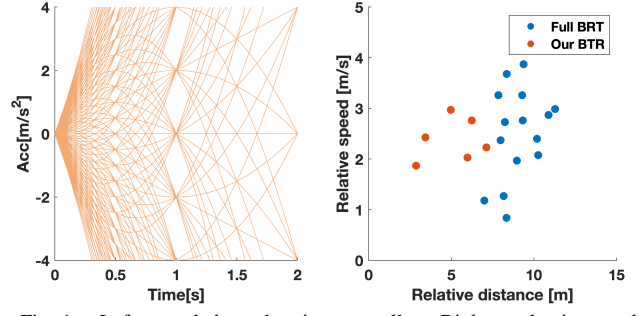

Fig. 1. Left: sampled acceleration controllers. Right: evaluation result.

C. Result

In the left plot of Fig. 1, we show the relative speed/distance between two cars when the human-driven car breaches the self-driving car's BRT. It can be observed that, even though all 15 interactions are safe, the human-driven car breaches the full BRT in all the interactions in our experiment, which shows the conservativeness of the full BRT. With our approach, only 6 interactions are identified as unsafe interactions by the BRT and the human-driven car breaches the BRT with an average time-to-collision $TTC = 2.4s$, which shows that our approach to dynamically updating the BTR reduces conservativeness of the full BRT while maintaining the safety verification ability of backward-reachability analysis. In Fig. 2, we show an instance of our experiment in a strong negotiation. In (a-1)-(a-2), the full BRT identifies the interaction as unsafe and the safety controller stops the self-driving car, although the human-driven car is trying to yield and stops at $6m$ away from the self-driving car. In contrast, our approach identifies that the human-driven car is willing to cooperate and quickly adjusts the BRT to ensure a safe and effective interaction.
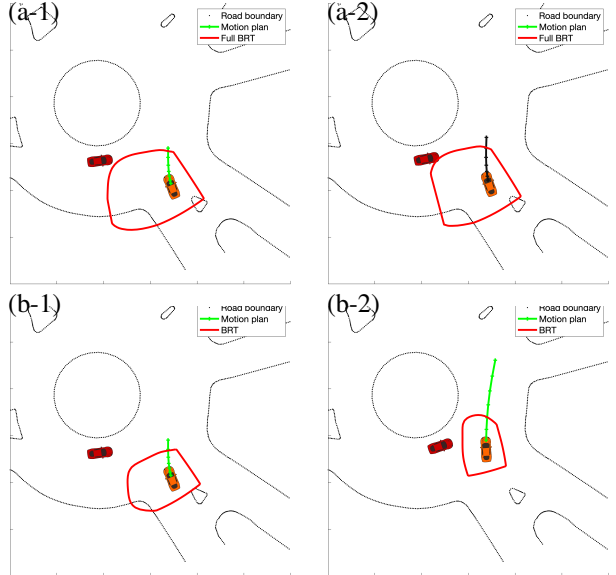

Fig. 2. (a-1)-(a-2) show two sequential steps in an experiment with the full BRT. (b-1)-(b-2) show steps in a similar simulation with the BRT generated by our approach. The BRTs showed in the cartesian frame above are projected from the Frenet frame.

References

[1] M. Chen and C. J. Tomlin, "Hamilton–jacobi reachability: Some recent theoretical advances and applications in unmanned airspace management," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 1, pp. 333–358, 2018.

[2] S. Bansal, M. Chen, S. Herbert, and C. J. Tomlin, "Hamilton-jacobi reachability: A brief overview and recent advances," in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*. IEEE, 2017, pp. 2242–2253.

[3] M. Althoff and J. M. Dolan, "Online verification of automated road vehicles using reachability analysis," *IEEE Transactions on Robotics*, vol. 30, no. 4, pp. 903–918, 2014.

[4] S. B. Liu, H. Roehm, C. Heinzemann, I. Lütkebohle, J. Oehlerking, and M. Althoff, "Provably safe motion of mobile robots in human environments," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 1351–1357.

[5] S. Bansal, A. Bajcsy, E. Ratner, A. D. Dragan, and C. J. Tomlin, "A hamilton-jacobi reachability-based framework for predicting and analyzing human motion for safe planning," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 7149–7155.

[6] K. Leung, E. Schmerling, M. Zhang, M. Chen, J. Talbot, J. C. Gerdes, and M. Pavone, "On infusing reachability-based safety assurance within planning frameworks for human–robot vehicle interactions," *The International Journal of Robotics Research*, vol. 39, no. 10-11, pp. 1326–1345, 2020.

[7] X. Wang, K. Leung, and M. Pavone, "Infusing reachability-based safety into planning and control for multi-agent interactions," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 6252–6259.

[8] A. Li, L. Sun, W. Zhan, M. Tomizuka, and M. Chen, "Prediction-based reachability for collision avoidance in autonomous driving," *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 2021.

[9] L. C. Evans and P. E. Souganidis, "Differential games and representation formulas for solutions of hamilton-jacobi-isaacs equations," *Indiana University mathematics journal*, vol. 33, no. 5, pp. 773–797, 1984.

[10] I. M. Mitchell, A. M. Bayen, and C. J. Tomlin, "A time-dependent hamilton-jacobi formulation of reachable sets for continuous dynamic games," *IEEE Transactions on automatic control*, vol. 50, no. 7, pp. 947–957, 2005.

[11] S. Mozaffari, O. Y. Al-Jarrah, M. Dianati, P. Jennings, and A. Mouzakitis, "Deep learning-based vehicle behavior prediction for autonomous driving applications: A review," *IEEE Transactions on Intelligent Transportation Systems*, 2020.

[12] T. Başar and G. J. Olsder, *Dynamic noncooperative game theory*. SIAM, 1998.

[13] B. D. Ziebart, A. L. Maas, J. A. Bagnell, and A. K. Dey, "Maximum entropy inverse reinforcement learning." in *Aaai*, vol. 8. Chicago, IL, USA, 2008, pp. 1433–1438.

[14] W. Zhan, L. Sun, D. Wang, H. Shi, A. Clausse, M. Naumann, J. Kümmerle, H. Königshof, C. Stiller, A. de La Fortelle, and M. Tomizuka, "INTERACTION Dataset: An INTERnational, Adversarial and Cooperative moTION Dataset in Interactive Driving Scenarios with Semantic Maps," *arXiv:1910.03088 [cs, eess]*, Sept. 2019.