Design of driver-assist systems under probabilistic safety specifications near stop signs

Mojtaba Forghani, John M. McNew, Daniel Hoehener, and Domitilla Del Vecchio

***Abstract*—In this paper, we consider the problem of designing in-vehicle driver-assist systems that warn or override the driver to prevent collisions with a guaranteed probability. The probabilistic nature of the problem naturally arises from many sources of uncertainty, among which the behavior of the surrounding vehicles and the response of the driver to on-board warnings. We formulate this problem as a control problem for uncertain systems under probabilistic safety specifications and leverage the structure of the application domain to reach computationally efficient implementations. Simulations using a naturalistic data set show that the empirical probability of safety is always within 5% of the theoretical value in the case of direct driver override, validating our models and algorithm. In the case of on-board warnings, the empirical value is more conservative due primarily to driver's decelerating more strongly than requested. But in all cases, the empirical value is greater than or equal to the theoretical value, demonstrating a clear safety benefit.**

Note to Practitioners: ***Abstract*—Statistics show that a large percentage of vehicle crash fatalities and injuries happen in the proximity of intersections and stop signs. Many automotive companies have already released automated braking systems that warn drivers and reduce speed when approaching an obstacle. A major problem with the design of such driver-assist systems is to guarantee the absence of collisions even in the presence of uncertainty. In this work we present an approach using a probabilistic model for human driving behavior. The advantage of a probabilistic model is that it allows to distinguish between possible and probable scenarios. In particular, for any desired safety level P, our method guarantees safety as long as surrounding vehicles do not use behaviors from the $1-P$ probability tail of their behavior distribution. Leveraging the monotone structure of the system we obtain an efficient algorithm that can compute warnings and overrides online. Moreover, simulations on a naturalistic data set show that the resulting override is considerably less conservative than one obtained when driver behavior is modeled through bounded uncertainty. There are a number of simplifying assumption made in this work, which limit the application mainly to prevention of rear-end collisions. We plan to generalize the method in order to be able to cover more general collision scenarios.**

Primary and Secondary Keywords ***Index Terms*—Primary Topics: Probabilistic Safety, Collision avoidance, Human-Centered Automation**

I. Introduction

VEHICLE collisions in the proximity of intersections and stop signs continue to contribute substantially to the number of light vehicle crash fatalities and injuries worldwide. In the past several years, an average of 21% of the fatalities and about 50% of the most serious injuries in the United States have been attributed to intersections [30]. In response to these startling statistics, the problem of preventing or mitigating collisions near intersections (signaled or not) has become a priority for the Department of Transportation (DOT) [29]. A number of automotive companies have already released automated braking systems that, based on on-board sensors, warn the driver and reduce the vehicle's speed when the vehicle approaches an obstacle [3], [4], [31]. Using the ego-vehicle dynamics and relative distance and speed to the obstacle, currently available automated braking systems can compute the least conservative braking timing for collision avoidance in scenarios where the obstacle's speed remains constant. However, in a recent study [33] of the NASS-GES database, NHTSA showed that 56% of all rear-end crashes occur with a preceding vehicle that is decelerating. In particular, the highest percentage of these rear-end crashes (32%) occur with a preceding vehicle that is decelerating at a stop.

In this paper, we focus on scenarios where the preceding vehicle is highly likely to brake, such as in the proximity of an intersection. Consequently, an automated braking system should take into account the preceding vehicle's deceleration when deciding whether a safety intervention is necessary, a capability that no automated braking system currently has. In particular, when the ego-vehicle's maximum possible deceleration (e.g. a heavy vehicle) is less than that of the preceding vehicle, it is possible for the driver of the ego-vehicle to choose a following distance that is so close that the ego-vehicle cannot avoid collision unless it begins braking before the preceding vehicle begins braking. In such a case it may be preferable to begin automatic braking based on a prediction of the lead vehicle's deceleration rather than on the current relative acceleration of the two vehicles. However, this strategy has an increased possibility of being viewed by the driver as a false activation and so it is desirable to be able to establish formal guarantees on the necessity of automatic braking activation.

The need for extending the functionality of active safety systems to cases where the intentions of surrounding vehicles are not known *a priori* motivates the use of formal model-based approaches for verification and design [14]. These approaches include uncertainty in the model due to the behavior of other traffic participants and design least conservative interventions that guarantee safety (absence of collisions) despite this uncertainty. A possible approach models uncertainty as disturbances that take values in a bounded set. Then the maximum invariant safe set can be computed, which allows to synthesize a least restrictive control input, that is, the

Mojtaba Forghani, Daniel Hoehener and Domitilla Del Vecchio are with the Department of Mechanical Engineering, MIT, 77 Massachusetts Avenue, Cambridge, MA. mojtaba@mit.edu, hoehener@mit.edu, ddv@mit.edu

John M. McNew is with Toyota Technical Center, 1555 Woodridge Avenue, Ann Arbor, MI. john.mcnew@tema.toyota.com

least conservative intervention to prevent a collision (see for instance [23]–[25] and the references therein). While often only approximations of the maximum invariant safe set can be computed efficiently (see, for instance, [1], [19]), a number of ground transportation systems can be modeled by a special class of systems, called order preserving systems, which allow computationally efficient exact solutions, [5], [6], [8], [12], [26], [27]. Although this approach guarantees both safety and least restrictive control actions, assuring safety means that preventive interventions are designed to avoid collisions with surrounding vehicles that behave in the worst possible manner.

To overcome the resulting conservatism while still guaranteeing an acceptable safety level, a possible approach is to account for the fact that worst case behaviors occur with a very low probability, so that one can focus on preventing collisions for behaviors around the statistical mean. Guaranteeing safety in this framework means to design preventive interventions that result in avoiding collisions $100P\%$ of the time with P pre-fixed.

Stochastic (hybrid) dynamical systems provide a quite natural model for human driving behavior, as was shown for instance in [10], [11], [18], [28]. Methods to approximate the probability that executions of a stochastic hybrid system enter some set of undesirable states (bad set) were proposed in [2], [19], [20]. While such approaches can be used for danger assessment, they do not provide a safe control strategy. In [9], using a stochastic differential equation model, a stochastic optimal control approach was proposed to compute the maximal set of initial conditions from which the bad set can be avoided with a given probability P. Solving this optimal control problem, however, requires to solve a partial differential equation which may not be appropriate for real time implementation.

The methods discussed above guarantee the absence of collisions in $100P\%$ of all cases. This is in contrast with other approaches that use stochastic models but do not guarantee that the probability of a collision is less than or equal to a desired level P [17], [22]. In these approaches the probability that the driver's input leads to a collision is computed and safety interventions are then taken if this probability is above a certain threshold.

The main contribution of this paper is an on-line algorithm that computes a least restrictive control input that guarantees safety (absence of collisions) with probability P, i.e. it guarantees safety as long as surrounding vehicles do not use actions from the $1-P$ probability tail of their behavior distribution. The specific application scenario considered in this paper involves only vehicle's longitudinal dynamics, which are known to be input/output order preserving systems, that is, stronger braking leads to lower speed and higher acceleration leads to higher speed [13]. This feature allows us to transform the computationally difficult stochastic problem into a simple deterministic algebraic check that can be performed on-line using extremal inputs.

We apply the method for designing a driver-assist system that prevents rear-end collisions with a preceding vehicle at stop signs, intersections, or speed bumps. We consider two possible implementations. In the first one, the driver-assist system overrides the driver while in the second one the driver is only warned when deemed necessary to guarantee a probability of safety P. In either case, the model of the preceding vehicle is identified from data gathered from vehicles driving in the city of Ann Arbor (MI). A different data set is employed in simulations to emulate the preceding vehicle and validate our model and algorithm. For the second implementation, we issue a warning through a visual interface that displays the required braking next to the braking the driver is currently applying. Based on our observed data, we model the response of the driver to this warning as a simple time delay, whose probability distribution is constructed from data.

The paper is organized as follows. In Section II, we describe the application scenario and in Section III we introduce the corresponding mathematical model. The theoretical solution is presented in Section IV, followed by the solution algorithm and simulation results for the case with direct driver override in Sections V and VI, respectively. Finally, the case when warnings are issued is described in Section VII. We conclude with a discussion of some of the assumptions made and possible relaxations in future work.

II. Application Scenario

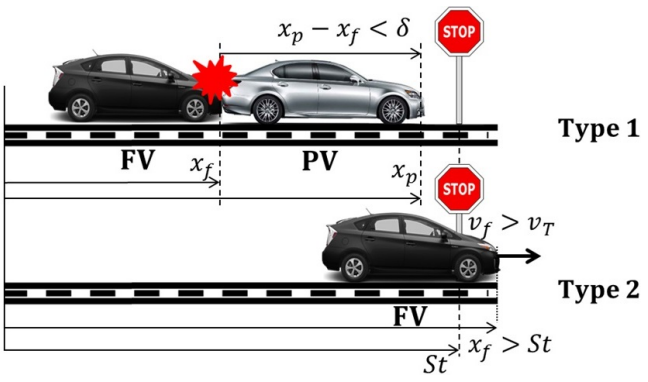

Fig. 1. The driver-assist system should prevent rear-end collisions with a preceding vehicle (PV) (collision of type 1). In addition, the ego-vehicle should not enter the study area (intersection or speed bump) at a high velocity (collision of type 2).

We consider the scenario of two consecutive vehicles approaching a study area, which can either be a stop sign, speed bump, or a non-signaled intersection. We assume that the following vehicle (FV) is equipped with a driver-assist system, while the preceding vehicle (PV) is fully human driven. The objective of the driver-assist system, which we seek to design, is to warn the driver and eventually apply automatic brake to mitigate two possible types of "collisions". In a type 1 collision, there is a rear-end collision between the two vehicles while in a type 2 collision, the FV crosses the study area with a velocity higher than a prefixed value. This value will be zero if the study area is a stop sign, while it will be strictly positive if the study area is only a speed bump or an intersection with the right of way. The scenario is depicted in Figure 1, in which, for simplicity, we have indicated the study area by a stop sign.

We denote longitudinal position and velocity of PV by x_p and v_p, respectively. Similarly, x_f and v_f are position and

velocity of FV, respectively. The longitudinal position of the study area according to a fixed reference frame is represented by St (Stop) and the maximum allowable velocity of FV at the study area is denoted by v_T. To keep the system away from type 1 collisions, we establish a minimum allowed distance between the two vehicles, which is denoted by $\delta > 0$. We consider a second order model to represent the longitudinal dynamics of both vehicles along their path. In particular, for FV we consider the model:

$$\ddot{x}_f = u - Dv_f^2 - a_r - a_s, \tag{1}$$

in which D captures the air drag, $a_r > 0$ models the deceleration due rolling resistance, a_s models the slope of the road, and u is the input that results from the brake or drive forces [21]. This input will be applied by the driver or by the driver-assist system in the case of an override. In particular, we will consider two possible implementations. In the first implementation, the driver-assist system will override the driver as soon as an imminent collision is detected. In the second implementation, the driver-assist system will only issue a warning and the control u will be the resulting input applied by the driver in response to that warning.

In order to understand the future motion of the human-driven PV based on the current state, we seek a model that captures typical driving behavior in the proximity of study areas such as speed bumps, non-signaled intersections and stop signs. Therefore, we assume a second order linear model in the form:

$$\ddot{x}_p = ax_p + bv_p + \mathbf{d}, \tag{2}$$

in which parameters a and b will be identified from data and $\mathbf{d}$ is a random variable whose probability distribution will also be identified from data (see Section VI). In this model, the probability distribution of $\mathbf{d}$ aims at capturing the variability among ways in which different drivers approach the study areas and also the variability among ways the same driver approaches these study areas. The full deceleration profile depends also on the distance to the study area and the current vehicle velocity. This is motivated by human factors studies that show that deceleration profiles depend on these variables, [15], [16]. In these studies, it was shown that the probability distribution of the average vehicle's deceleration when approaching the study area well captures the variability among drivers. This motivates the choice of $\mathbf{d}$ as random variable to capture the essentials of the variability among drivers.

Previous works have considered similar applications in which, in contrast to the present paper, $\mathbf{d}$ was modeled as a bounded disturbance whose bounds were set to capture all possible observed driver braking behaviors [26].

III. Mathematical Model and Problem Formulation

In this section, we introduce the mathematical model that we use as abstraction of the application scenario described in the previous section and then provide the formal problem formulation. We conclude the section by illustrating the important concepts with the application example.

A. Preliminaries and mathematical model

For any set $A \subset \mathbb{R}^n$, $\mathrm{cl}(A)$ denotes the closure of A, A^c its complement and ∂A the boundary of A. Furthermore, $S(A)$ is the set of piecewise continuous signals with images in A. We represent a *continuous system* by the collection $\Sigma = (X, U, \Delta, O, f, h)$, where $X \subset \mathbb{R}^n$ represents the state space, $U \subset \mathbb{R}^m$ is the set of input values, $\Delta \subset \mathbb{R}$ is the set of disturbance inputs, $O \subset X$ is the output space, $f\colon X \times U \times \Delta \to X$ is a vector field representing the system dynamics and finally, $h\colon X \to O$ is the output map.

The corresponding *flow* $\phi\colon [0, \tau_{end}] \times X \times S(U) \times \Delta \to X$ for some $\tau_{end} > 0$ is the map satisfying for all initial conditions $x \in X$, $\mathbf{u} \in S(U)$ and $d \in \Delta$, $\phi(0, x, \mathbf{u}, d) = x$ and $\dot{\phi}(t, x, \mathbf{u}, d) = f(\phi(t, x, \mathbf{u}, d), \mathbf{u}(t), d)$ for all $t \in [0, \tau_{end}]$. We assume that the flow exists and is continuous with respect to all its arguments. Notice also that here we have constant disturbance inputs.

The system that we consider is the parallel composition of two systems, formally introduced as follows.

Definition 1: For continuous systems $\Sigma^1 = (X^1, U^1, \Delta^1, O^1, f^1, h^1)$ and $\Sigma^2 = (X^2, U^2, \Delta^2, O^2, f^2, h^2)$ we define the *parallel composition* $\Sigma := \Sigma^1 \parallel \Sigma^2 := (X, U, \Delta, O, f, h)$, in which $X := X^1 \times X^2$, $U := U^1 \times U^2$, $\Delta := \Delta^1 \times \Delta^2$, $O := O^1 \times O^2$, $f := (f^1, f^2)^T$ and $h := (h^1, h^2)^T$.

In this paper, we consider systems whose flow is a monotone map, as this enables an efficient solution to the problem described in Section II. More precisely, we have the following.

Definition 2: The pair $(P, \preceq)$ consisting of a set P and a binary relation "$\preceq$" is a *partially ordered set* if for all $p_1, p_2, p_3 \in P$ the following relations hold:

i) $p_1 \preceq p_1$;
ii) $p_1 \preceq p_2$ and $p_2 \preceq p_1$ implies $p_1 = p_2$;
iii) $p_1 \preceq p_2$ and $p_2 \preceq p_3$ implies $p_1 \preceq p_3$.

Example 1: Defining for all $w, z \in \mathbb{R}^n$, $w \leq z$ if and only if $w_i \leq z_i$ for all $i \in \{1, 2, ..., n\}$, where v_i denotes the ith component of a vector v, we have that $(\mathbb{R}^n, \leq)$ is a partially ordered set. Similarly, for any set $A \subset \mathbb{R}^n$, we have that $S(A)$ together with the binary relation, $\mathbf{w} \leq \mathbf{z}$ if $\mathbf{w}(t) \leq \mathbf{z}(t)$ for all $t \in [0, \tau_{end}]$, forms a partially ordered set.

Definition 3: Let $(P, \preceq_P)$ and $(Q, \preceq_Q)$ be partially ordered sets. The map $f\colon P \to Q$ is *order preserving* (strict order preserving) provided that for $x, y \in P$ such that $x \preceq_P y$ ($x \prec_P y$), we have $f(x) \preceq_Q f(y)$ ($f(x) \prec_Q f(y)$), where $p \prec q$ is an abbreviation for $p \preceq q$ and $p \neq q$.

Definition 4: A continuous system $\Sigma = (X, U, \Delta, O, f, h)$ is called *input/output order preserving* (strict input/output order preserving) with respect to the control (disturbance), if the map $S(U) \ni \mathbf{u} \mapsto h(\phi(t, x, \mathbf{u}, d))$ ($\Delta \ni d \mapsto h(\phi(t, x, \mathbf{u}, d))$), is order preserving (strict order preserving) for any fixed t, x and d ($\mathbf{u}$).

This completes the preliminary definitions and we can introduce the class of systems which we will use in the rest of the paper.

Definition 5: Let $\Sigma^1 = (X^1, U, \emptyset, O^1, f^1, h^1)$ and $\Sigma^2 = (X^2, \emptyset, \Delta, O^2, f^2, h^2)$ be continuous systems. The parallel composition $\Sigma^* := \Sigma^1 \parallel \Sigma^2$ is an *order preserving system with*

stochastic disturbance (OPSD) if the following conditions are satisfied:

i) Σ^1 is input/output order preserving with respect to the control;
ii) Σ^2 is input/output order preserving with respect to the disturbance;
iii) There exists $u_m \in U$ such that $u_m \leq u$ for all $u \in U$;
iv) The disturbance input $\mathbf{d}$ of the system is a Δ-valued random variable with unimodal, invertible, cumulative distribution function $\Phi\colon \mathbb{R} \to [0,1]$.

B. Problem formulation

Assume we are given an OPSD Σ^* as defined in the previous section and a subset $B \subset X$, called *bad set*, of its state space. The objective is to keep the system outside of the bad set with a given probability P, that is, to keep a *safety level* P. The set of initial conditions such that there exists an open loop control for which this is possible is called the maximal open loop safe set.

Definition 6: Let Σ^* be an OPSD, $B \subset X$ be the bad set and $P \in (0,1)$ be given. The *maximal open loop safe set* is the set

$$\mathcal{W}(P) := \{x \in X \mid \exists \mathbf{u} \in S(U) \text{ s.t. } \Pr(\phi(t,x,\mathbf{u},\mathbf{d}) \notin B,\ \forall t \in [0,\tau_{end}]) \geq P\},$$

where ϕ denotes the flow of Σ^*.

Before giving a formal problem statement, let us introduce the assumption that we make on the shape of the bad set B.

Assumption 1: The bad set satisfies $B = B_1 \cup B_2$, with

$$B_1 := \bigcup_{i=1}^{r} \left\{x \in X \;\middle|\; Z_i^1 h^1(x^1) - Z_i^2 h^2(x^2) > H_i\right\},$$
$$B_2 := \bigcup_{j=1}^{s} \left\{x \in X \;\middle|\; G^j(x^1) > g^j\right\},$$

where Z^1 and Z^2 are $r \times \dim(O^1)$ and $r \times \dim(O^2)$ matrices with non negative coefficients, respectively, H is a r-dimensional vector, $G^j\colon X^1 \to \mathbb{R}^q$ and $g^j \in \mathbb{R}^q$. Moreover, for fixed $t \in [0,\tau_{end}]$, $x^1 \in X^1$, and $j \in \{1,...,s\}$, $G^j(\phi^1(t,x^1,\cdot))\colon S(U) \to \mathbb{R}^q$ is order preserving. The ith row of a matrix A is denoted by A_i and X, X^i, h^i are as in the definition of an OPSD for $i = \{1,2\}$. Finally, ϕ^1 is the flow corresponding to the continuous system Σ^1.

The goal is to construct a *safety supervisor*, that is, a controller that can enforce control inputs when necessary in order to achieve a desired safety level P. Moreover, this supervisor should act as late as possible, i.e. it should be a *least restrictive* controller. Formally:

Problem 1: Let Σ^* be an OPSD, $B \subset X$ satisfy Assumption 1 and $P \in (0,1)$ be given. Find a least restrictive feedback control map $\pi\colon X \to U$ such that

$$\Pr(\phi^\pi(t,x,\mathbf{d}) \notin B,\ \forall t \in [0,\tau_{end}]) \geq P, \qquad \forall x \in \mathcal{W}(P),$$

where ϕ^π denotes the flow of the closed-loop system corresponding to Σ^* and π.

In summary, we study systems of two agents, a controlled and an uncontrolled stochastic one. The controlled agent will have to prevent entering a bad set of states B_2 while also keeping a sufficient separation from the uncontrolled agent (set B_1).

C. Illustration on the application example

Consider the scenario of Section II, the system model is given by $\Sigma_{app} := \Sigma^1 || \Sigma^2$, where Σ^1 and Σ^2 are modeling FV and PV, respectively. Hence, $X^1 = X^2 = \mathbb{R} \times \mathbb{R}_+$ where $x^1 = (x_f, v_f)^T$, $x^2 = (x_p, v_p)^T$. Notice that we assume that the speed is always non-negative, meaning that vehicles cannot move backwards. The output maps are $h^1(x^1) = x_f$ and $h^2(x^2) = x_p$, i.e. the vehicle positions. The control input values are $U = [u_m, u_M] \subset \mathbb{R}$ for some constants $u_m < u_M$ and the set of disturbance inputs $\Delta = \mathbb{R}$. The vector fields $f^1\colon X^1 \times U \to X^1$ and $f^2\colon X^2 \times \Delta \to X^2$ are defined in accordance to (1) and (2):

$$f^1(x^1,u) = \begin{cases} (v_f,\ u - Dv_f^2 - a_r - a_s)^T & \text{if } v_f > 0, \\ 0 & \text{if } v_f \leq 0, \end{cases} \tag{3a}$$

$$f^2(x^2,d) = \begin{cases} (v_p,\ ax_p + bv_p + d)^T & \text{if } v_p > 0, \\ 0 & \text{if } v_p \leq 0. \end{cases} \tag{3b}$$

Finally, τ_{end} represents the time when FV leaves the study area. It can be shown that each of Σ_1 and Σ_2 is input/output order preserving [13]. This is qualitatively illustrated in Figure 2. Finally, the bad set models type 1 and type 2 collisions (Figure 1) through sets B_1 and B_2 respectively, where $Z^1 = Z^2 = 1$, $H = -\delta$, $G^1(x^1) = x^1$, $g^1 = (St, v_T)^T$ and $s = 1$, with δ, St and v_T as in Section II. Furthermore, since the flows of x_f and v_f are order preserving with respect to $\mathbf{u}$, $\mathbf{u} \mapsto G^1(\phi^1(t,x^1,\mathbf{u}))$ is order preserving.

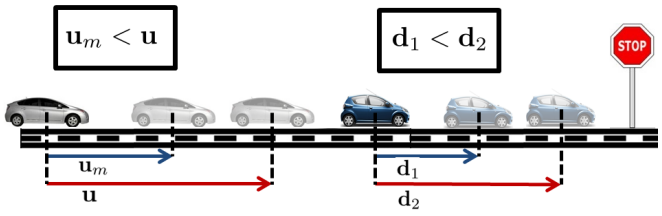

Fig. 2. The figure shows the resulting FV (light colored) positions when two different control inputs are applied over the same amount of time. By the order preserving property the position corresponding to the larger input is closer to the study area. For the PV (dark colored) an analogous property holds.

IV. Problem Solution

Throughout this section $P \in (0,1)$ and $\Sigma^* = \Sigma^1 \parallel \Sigma^2$, with $\Sigma^1 = (X^1, U, \emptyset, f^1, h^1)$ and $\Sigma^2 = (X^2, \emptyset, \Delta, f^2, h^2)$, denotes an OPSD with corresponding flow $\phi = (\phi^1, \phi^2)$.

We solve Problem 1 in three main steps. First, we will show (Theorem 1 below) that, thanks to the system being input/output order preserving with respect to the control, the maximal open loop safe set can be represented by

$$\mathcal{W}(P) = \{x \in X \mid \Pr(\phi(t,x,\mathbf{u}_m,\mathbf{d}) \notin B,\ \forall t \in [0,\tau_{end}]) \geq P\},$$

where $\mathbf{u}_m(t) = u_m$ for all $t \in [0, \tau_{end}]$. The interpretation of this result is that if the control $\mathbf{u}_m$ cannot prevent $100P\%$ of the collisions, then no other control can. This fact dramatically simplifies the problem of computing $\mathcal{W}(P)$ since it is not required to search for the best control input as this is always given by $\mathbf{u}_m$. Furthermore, since the system is input/output order preserving with respect to the disturbance, we can provide a simple deterministic check that allows to determine whether the system state is in $\mathcal{W}(P)$. Then we prove (Theorem 2 below) that we can guarantee a safety level P if we apply any arbitrary control $\mathbf{u}$ as long as the current system state is in the interior of $\mathcal{W}(P)$ and the control $\mathbf{u}_m \in S(U)$ otherwise.

Before presenting the solution to Problem 1, we define the (control dependent) P-safety capture set, representing the set of all states for which the probability of avoiding the bad set B is less than P for a fixed control input signal.

Definition 7: Let $P \in (0,1)$. The *P-safety capture set* for a given control input signal $\mathbf{u} \in S(U)$, OPSD Σ^* with flow ϕ and bad set B is defined as

$$C_{\mathbf{u}}(P) := \{x \in X \mid \Pr\left(\phi(t, x, \mathbf{u}, \mathbf{d}) \notin B,\ \forall t \in [0, \tau_{end}]\right) < P\}.$$

As above, let us define

$$\mathbf{u}_m \colon [0, \tau_{end}] \to U, \quad \mathbf{u}_m(t) = u_m\ \forall t \in [0, \tau_{end}]. \tag{4}$$

By the properties of the OPSD Σ^*, it is clear that $\mathbf{u}_m \leq \mathbf{u}$ for all $\mathbf{u} \in S(U)$, where "$\leq$" denotes the partial order on signals, see Example 1.

In the following we state the theoretical results that allow to solve Problem 1, the proofs of these results are provided in the Appendix.

Theorem 1: For an OPSD Σ^*, $P \in (0,1)$ and a bad set B satisfying Assumptions 1,

$$x \in \mathcal{W}(P) \quad \iff \quad x \notin C_{\mathbf{u}_m}(P).$$

By Theorem 1 it is clear that from some initial condition $x \in X$ we can avoid the set B with probability larger than or equal to P if the flow corresponding to the input $\mathbf{u}_m$ avoids B with probability at least P. The remaining question is when to apply this control in order to construct a least restrictive safety supervisor. It seems natural to wait until the state reaches the boundary of the safe set $\mathcal{W}(P)$. The following Theorem confirms this intuition.

Theorem 2 (Solution to Problem 1): Let Σ^* be an OPSD, $P \in (0,1)$ and $B \subset X$ a set satisfying Assumption 1. Define the feedback map $\pi \colon X \to U$ by

$$\pi(x) = \begin{cases} u_m & \text{if } x \in \mathrm{cl}(C_{\mathbf{u}_m}(P)), \\ U & \text{otherwise,} \end{cases} \tag{5}$$

where u_m is as in Definition 5 and $\mathbf{u}_m$ as in (4). Then for all $x \in \mathcal{W}(P)$ we have that $\Pr(\phi^\pi(t, x, \mathbf{d}) \notin B,\ \forall t \in [0, \tau_{end}]) \geq P$, where ϕ^π denotes the flow of the closed-loop system corresponding to Σ^* and π.

The basic idea of the feedback π is illustrated in Figure 3. Implementing this feedback controller requires an efficient way to check whether the state is in $C_{\mathbf{u}_m}(P)$. That is, the next step is to remove the need to integrate the dynamics of all possible disturbances to check whether a given initial condition $x \in X$ is in $\mathrm{cl}(C_{\mathbf{u}_m}(P))$. This problem is addressed by the following Proposition. For notational simplicity, let us define for all $t \in [0, \tau_{end}]$, $x \in X$, $\mathbf{u} \in S(U)$ and $d \in \Delta$:

$$F^{t,x,\mathbf{u}}(d) := Z^1 h^1(\phi^1(t, x^1, \mathbf{u})) - Z^2 h^2(\phi^2(t, x^2, d)), \tag{6}$$

with $F_i^{t,x,\mathbf{u}}$ denoting the ith component of $F^{t,x,\mathbf{u}}$.

Proposition 1: Let $P \in (0,1)$, B satisfy Assumption 1 and $\mathbf{u} \in S(U)$ be a control input signal. Define $\bar{d} := \Phi^{-1}(1-P)$, where Φ is the cumulative distribution function of the random variable $\mathbf{d}$, see Definition 5. Then we have that $C_{\mathbf{u}}(P) = \mathbf{S}_1^{\mathbf{u}} \cup \mathbf{S}_2^{\mathbf{u}}$ where

$$\mathbf{S}_1^{\mathbf{u}} := \left\{x \in X \mid \exists\, t \in [0, \tau_{end}], \exists\, i \in \{1, ..., r\}, \text{ such that } H_i < F_i^{t,x,\mathbf{u}}(\bar{d})\right\},$$
$$\mathbf{S}_2^{\mathbf{u}} := \left\{x \in X \mid \exists\, t \in [0, \tau_{end}], \exists\, j \in \{1, ..., s\}, \text{ such that } G^j(\phi^1(t, x^1, \mathbf{u})) > g^j\right\}.$$

To summarize, by Theorem 2 we can guarantee a prescribed safety level P by only overriding the driver input when the system state is on the boundary of the safe set $\mathcal{W}(P)$. In general, computation of the set $\mathcal{W}(P)$ is very demanding. However, exploiting that the system is input/output order preserving with respect to both control and disturbance we can derive via Theorem 1 and Proposition 1 a simple way to check whether the system state $x \in X$ is in $\mathcal{W}(P)$. Indeed, these results imply that it is sufficient to check whether the flow starting at x with control input $\mathbf{u}_m$ and deterministic disturbance input $\bar{d}$, where $\bar{d}$ is as in Proposition 1, enters the bad set B. We discuss this approach in detail in the next section.

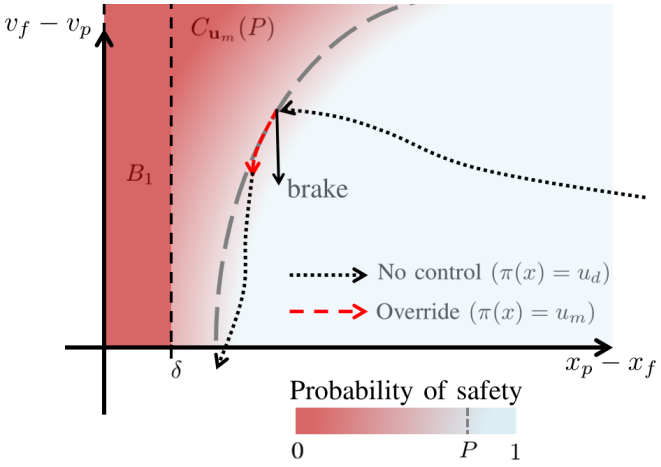

Fig. 3. The figure illustrates a slice of the four dimensional capture set of the application example in the case when type 2 collisions are ignored for simplicity. The axes show the relative position and speed of PV and FV, x denotes the state and u_d is the driver's input. The feedback π of Theorem 2 overrides the driver input at the boundary of the capture set. The resulting reduction in FV's speed allows to maintain the desired safety level.

V. Computation of Capture Set and Control Map

Based on Theorem 2, if we can calculate the capture set $C_{\mathbf{u}_m}(P)$, then the feedback map π defined by (5) guarantees

a safety level P. Using Proposition 1, we have that

$$C_{\mathbf{u}_m}(P) = \mathbf{S}_1^{\mathbf{u}_m} \cup \mathbf{S}_2^{\mathbf{u}_m},$$

and more precisely that the state $x \in X$ is in the capture set if at least one of the following two conditions is satisfied:

$$\begin{aligned} &1) \min_{\substack{t\in[0,\tau_{end}]\\ i\in\{1,\dots,r\}}} (H_i - F_i^{t,x,\mathbf{u}_m}(\bar{d})) < 0; \\ &2) \min_{\substack{t\in[0,\tau_{end}]\\ j\in\{1,\dots,s\}}} g^j - G^j(\phi^1(t,x^1,\mathbf{u}_m) < 0. \end{aligned} \tag{7}$$

These can be easily checked by integrating the system dynamics with the specified control and disturbance input, $\mathbf{u}_m$ and $\bar{d}$ respectively, over the time horizon $[0, \tau_{end}]$.

As mentioned in Section III, the feedback map π acts as a *safety supervisor*. That is, it overrides the driver when the system's state risks entering the capture set. The block diagram corresponding to this safety supervisor is depicted in Figure 4. The current system state $x \in X$ and the desired input u_d are given as an input to the supervisor. Using the conditions of (7), the supervisor verifies if the current state is in the closure of the capture set. If yes, then the desired control u_d is overridden by u_m. Otherwise u_d is applied.

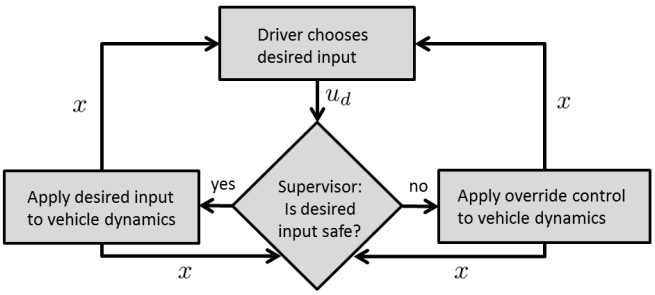

Fig. 4. Block diagram of the safety supervisor. In the figure x denotes the system state and u_d the desired input of the driver.

For the simulations, this safety supervisor was implemented in discrete time. Pseudo-code is outlined in Algorithm 1. The state of the discretized system at step k is denoted by $x[k] = (x^1[k], x^2[k])^T$. Future states are computed using Euler forward approximation with a fixed step size Δt. We also define the function $\mathbf{F}_i[k]$ as follows

$$\mathbf{F}_i[k] := Z_i^1 h^1(x^1[k]) - Z_i^2 h^2(x^2[k]). \tag{8}$$

All other notations are similar to the continuous-time model. Note that $x^1[k]$ and $x^2[k]$ in (8) depend on the initial condition and the inputs.

VI. Validation on Naturalistic Data Sets

In this section, we describe how the data of the preceding vehicle was collected and used to identify its parameters. We then provide results from simulations on two different data sets.

Algorithm 1 Safety supervisor

Require: $x[0] = (x^1[0], x^2[0])^T$: current state
Require: $u[0]$: desired input
  $k \leftarrow 1$;
  $\bar{d} \leftarrow \Phi^{-1}(1-P)$;
  $x[1] \leftarrow x[0] + \Delta t(f(x[0], u[0], \bar{d}))$; {One step lookahead}
  {Check whether $x[1]$ is safe, override if not}
  while $k\Delta t \leq \tau_{end}$ **do**
    for $i = 1$ to $i = r$ **do**
      if $H_i - \mathbf{F}_i[k] \leq 0$ **then**
        $u[0] \leftarrow u_m$; BREAK;
      end if
    end for
    for $j = 1$ to $j = s$ **do**
      if $g^j - G^j(x^1[k]) \leq 0$ **then**
        $u[0] \leftarrow u_m$; BREAK;
      end if
    end for
    $x[k+1] \leftarrow x[k] + \Delta t(f(x[k], u_m, \bar{d}))$;
    $k \leftarrow k + 1$;
  end while
  return $u[0]$;

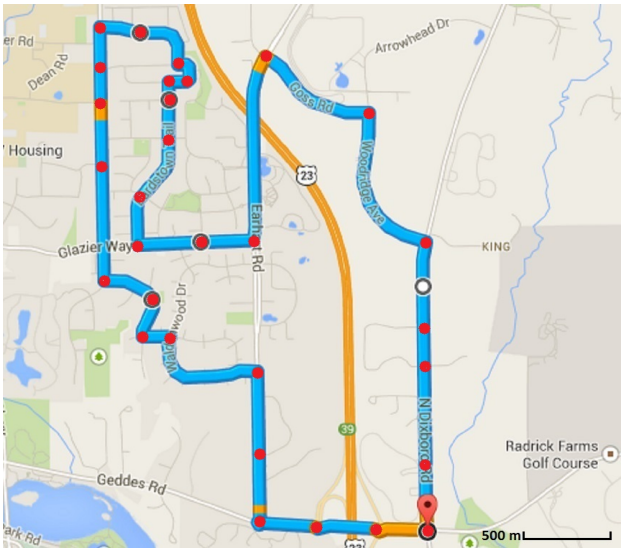

Fig. 5. The path that is used for PV data acquisition. The dots indicate study areas such as stop signs, speed bumps, roundabouts, etc.

A. Data collection and model identification

Figure 5 shows the path that was used to collect the data for both identifying the parameters of PV model and validating the safety guarantees of Algorithm 1. This path is located in Ann Arbor, Michigan, it is 11 km long, and consists of 30 study areas. The study area is a part of the road at which the driver frequently reduces his/her speed such as intersections, roundabouts, speed bumps or stop signs. For identifying the parameters of the PV model, we collected a total of 125 approaches to study areas, all from the same middle-aged, male driver. The data collected contains both speed and acceleration measurements from on-board sensors and position measurements obtained from GPS (Figure 6). We

call this data set the *test-vehicle data set*.

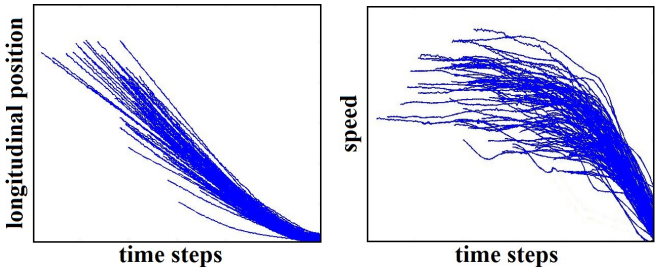

Fig. 6. Trajectories of position and speed of 125 profiles of the test vehicle at study areas [units and numbers are removed as they are proprietary information].

Based on this data we assumed a normal distribution for the disturbance input, i.e. $\mathbf{d} \sim \mathcal{N}(\mu, \sigma^2)$. We have used the least squares method to calculate the parameters a, b, μ and σ. In particular, using (3) and (3b) for $v_p[k] \geq 0$ we have

$$\begin{bmatrix} x_p[k+1] \\ v_p[k+1] \end{bmatrix} = \begin{bmatrix} x_p[k] + \Delta t v_p[k] \\ v_p[k] + \Delta t(a x_p[k] + b v_p[k] + \mu) \end{bmatrix}. \quad (9)$$

By replacing $x_p[k]$ in the second equation of (9) with $x_p[k-1] + \Delta t v_p[k-1]$, we obtain $v_p[k+1] = a(\Delta t x_p[k-1] + \Delta t^2 v_p[k-1]) + (1 + b\Delta t) v_p[k] + \mu \Delta t$. We define the new parameters $a' = a$, $b' = 1 + b\Delta t$ and $\mu' = \mu$. Minimizing the squared error for speed leads to the following optimization problem:

$$\min_X \|CX - D\|^2, \text{ with } X = (a', b', \mu')^T, \quad (10)$$

where $C = [C_{i,j}]_{\substack{i=1,\ldots,N_d \\ j=1,\ldots,3}}$, with N_d denoting the number of data points and $C_{1,1} = \Delta t x_p[1]$, $C_{1,2} = v_p[1]$, $C_{1,3} = \Delta t$, and for $k \geq 2$, $C_{k,1} = \Delta t x_p[k-1] + \Delta t^2 v_p[k-1]$, $C_{k,2} = v_p[k]$, $C_{k,3} = \Delta t$. Similarly, $D \in \mathbb{R}^{N_d}$ with components $D_k = v_p[k+1]$. Also, the variance can be calculated using

$$\sigma^2 = \frac{\sum_{i=1}^{N_d} |a x_p[i] + b v_p[i] + \mu - a_p[i]|^2}{N_d},$$

where N_d is the number of all data points, and $x_p[i]$, $v_p[i]$ and $a_p[i]$ are the measured position, velocity and acceleration of the ith data point.

B. Simulation results on a naturalistic data set

Algorithm 1 provides a safety supervisor that, given the model for the preceding vehicle, guarantees that independent from what the driver of the following vehicle does, at most $100(1-P)\%$ of all scenarios result in a crash. In order to verify this property we tested the algorithm on two different data sets. For the first test, we used the test-vehicle data set described in the previous section. In order to perform model identification and simulations on the same data set we used the cross validation method. For the second test, we collected through radar measurements the position and speed of vehicles that we encountered during a test drive in Ann Arbor on the same path of Figure 5. While this data was recorded, drivers were not aware of the ongoing experiment. This led to a data set consisting of 41 approaches to study areas of 41 different vehicles and drivers. This data set is called the *radar data set*.

In order to validate Algorithm 1, we determined the empirical safety level as follows. For each trial, we randomly and uniformly selected the initial speed of FV, in a range of 5 to 20 meters per second. Similarly we selected uniformly an initial relative position to PV in a range from δ to 50 meters. The FV would then move with a constant acceleration, chosen also according to a uniform distribution in [0,3] meters per square second. For the preceding vehicle, we chose for each trial randomly, according to a uniform distribution, either a trajectory from the test-vehicle data set (first test) or from the radar data set (second test). After performing a large number of trials $T = 5000$ we computed the *empirical safety level* $(1 - N/T)$, where N is the number of collisions encountered during T trials. Then we compared the obtained value with P. The logic diagram of our tests is shown in Figure 7.

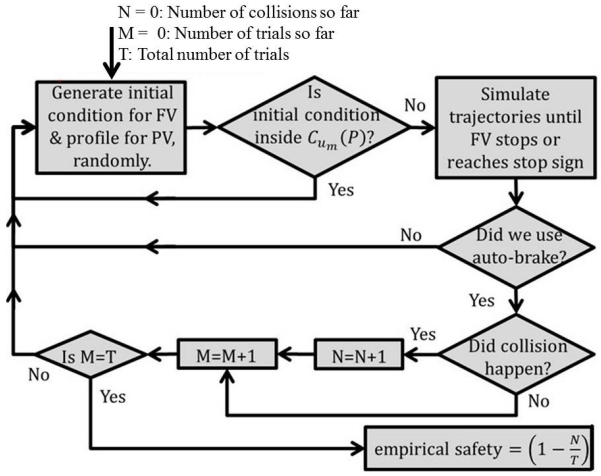

Fig. 7. Test to evaluate the empirical safety of the system.

As stated above, for the first test we use the k-fold cross validation method, see for instance [7], with $k = 5$. Thus, we partition the test-vehicle data set into 5 groups (each group with 25 trajectories), identify the parameters by solving the minimization problem (10) using the data of 4 groups and run the tests outlined in Figure 7 on the 5th group, called *test group*. This way, we compute for each test group the empirical safety level. The results are shown in Table I and indicate that the algorithm leads to an empirical safety level $1 - N/T$ close to P.

TABLE I
EMPIRICAL SAFETY LEVEL ON THE TEST-VEHICLE DATA USING 5-FOLD CROSS VALIDATION.

Test group	Empirical safety level for $P = 0.7$	$P = 0.8$	$P = 0.9$
Group 1	0.720	0.853	0.944
Group 2	0.677	0.821	0.934
Group 3	0.702	0.834	0.932
Group 4	0.695	0.812	0.912
Group 5	0.669	0.791	0.916
Average	**0.693**	**0.822**	**0.928**

For the second test, we trained the PV model using all 125 trajectories of the test-vehicle data set and for the simulations we randomly and uniformly selected trajectories from the naturalistic radar data set to perform the test outlined in Figure 7. Table II shows that the empirical safety level is still very close to the desired one, which also indicates that the identified model of the PV has good generalization ability.

TABLE II
EMPIRICAL SAFETY LEVEL ON THE NATURALISTIC RADAR DATA SET

Safety level (P)	Empirical safety level
0.7	0.681
0.8	0.832
0.9	0.929

Finally we compared our safety supervisor based on a stochastic disturbance model with the same supervisor based on a deterministic model as was proposed in [12] and [26]. In the deterministic model, the set of disturbance inputs is $\widetilde{\Delta} = [d_m, d_M]$, where the lower and upper bounds d_m and d_M correspond to the smallest and largest value of the disturbance within the 125 trajectories of the test-vehicle data set. Figure 8 shows a comparison of the timing when the supervisor based on the deterministic model first applies control input and when the supervisors based on the stochastic disturbance model with $P = 0.8$ and $P = 0.98$ do. The simulation was performed for a PV trajectory randomly chosen from the test-vehicle data set and the same FV model as for the other simulations was used. It is clear that using the deterministic model the safety supervisor overrides the driver sooner than with the stochastic one, and that for $P = 0.98$ overrides occur sooner than for $P = 0.8$.

Notice that in Figure 8(b), (d) and (f), we have used a counter that keeps the control input $\mathbf{u}_m$ on for at least $1s$ whenever the system exits $C_{\mathbf{u}_m}(P)$. As we can see from the plots, the number of switches between driver and supervisor input has been reduced significantly in Figure 8(f) compared to Figure 8(e). One can show that theoretically this modification of the control by introducing hysteresis does not affect the safety guarantees.

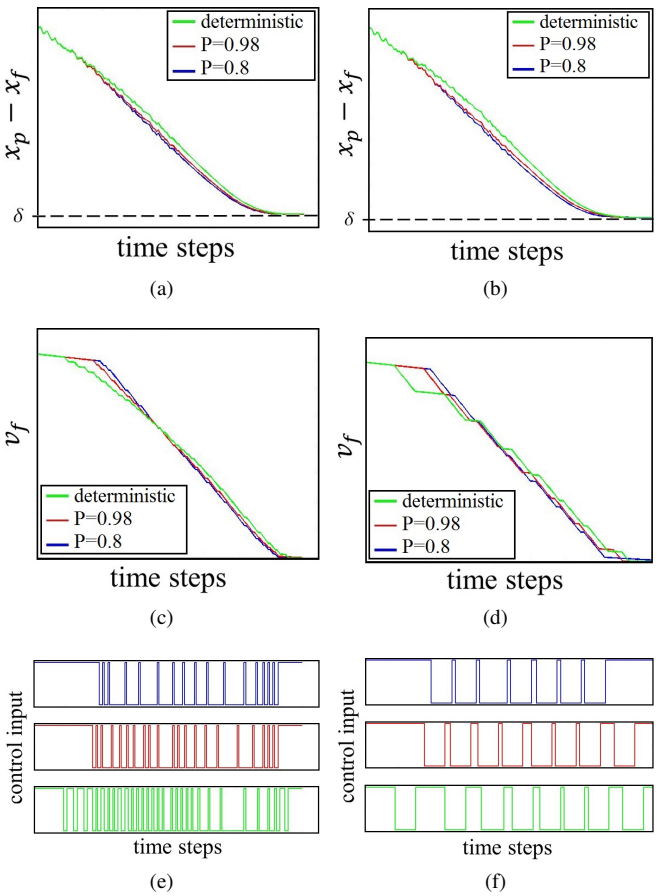

Fig. 8. Simulation results for safety $P \in \{0.8, 0.98, 1\}$ with the same PV profile and FV initial state. (a),(c) and (e) show from top to bottom the relative position between PV and FV, the speed of FV and the control input. (b),(d) and (f) show the same graphs for a modified controller with the hysteresis of $1s$ [units and numbers are removed as they are proprietary information].

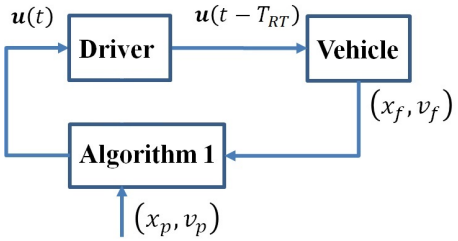

Fig. 9. Based on position and velocity of both vehicles, Algorithm 1 determines the safe input $u(t)$ which is applied by the driver with a random delay of T_{RT}.

VII. EXTENSION TO THE CASE WITH HUMAN IN THE LOOP

In the previous sections, we considered the case where the driver-assist system overrides the driver with emergency brake when deemed necessary to ensure the desired probability of safety. In this section, we design a driver-assist system that only warns the driver with the emergency brake level required to keep the desired probability of safety. This allows the driver to apply brake himself/herself, which consequently leads to minimizing automatic brake interferences. We model the human driver as an actuator with time delay, called T_{RT}, which is the time the driver takes to respond to the warning. We extend Algorithm 1 to the case in which we have "actuator delay" T_{RT}, distributed according to a probability distribution that will be identified from experimental data (see Section VII-B for the details). The architecture of this driver-assist system is depicted in Figure 9.

A. Extension of Algorithm 1 to stochastic actuator delay

In the following, for simplicity, we consider only rear-end collisions, that is, we assume $B_2 = \emptyset$. The safety supervisor is still in the form of (5), however $\mathbf{u}_m$ is not applied directly, but displayed to the driver (Section VII-B). The capture set is computed taking into account that when a control input is issued, it takes T_{RT} time before it can be applied, see Figure 9. That is, we have a system in the form of equation (3a) but with a control delay: $\forall v_f > 0$

$$f^1(x^1, u(t - T_{RT})) = \left(v_f,\ u(t - T_{RT}) - Dv_f^2 - a_r - a_s\right)^T,$$

where $T_{RT} \sim F_{T_{RT}}(t_{RT})$ and $F_{T_{RT}} : \mathbb{R}_+ \to [0,1]$ denotes the cumulative distribution function of the reaction time. As a consequence, the capture set is given by

$$C_{\mathbf{u}_{T_{RT}}}(P) = \{x \in X \mid \Pr(\forall t \in [0, \tau_{end}], \\ \phi_1^2(t, x^2, \mathbf{d}) - \phi_1^1(t, x^1, \mathbf{u}_{T_{RT}}) \geq \delta) < P\},$$

where $\phi_1^2(t, x^2, \mathbf{d})$ and $\phi_1^1(t, x^1, \mathbf{u}_{T_{RT}})$ are $x_p(t)$ and $x_f(t)$ of Σ_{app}, respectively, with $\mathbf{u}_{T_{RT}}$ defined as

$$\mathbf{u}_{T_{RT}}(t) = \begin{cases} u_0 & \text{if } t \in (0, T_{RT}), \\ u_m & \text{if } t \in [T_{RT}, \tau_{end}], \end{cases} \tag{11}$$

where $u_0 \in U$ is the value of the driver input when the warning is issued. Here, we assume that the value of u_0 is constant during the reaction time. In the following proposition we compute an approximation for $C_{\mathbf{u}_{T_{RT}}}(P)$.

Proposition 2: Let $P \in (0,1)$ and $p^* \in (0,1)$ be given and t^*_{RT} be such that

$$F_{T_{RT}}(t^*_{RT}) = \Pr(T_{RT} \leq t^*_{RT}) = p^*. \tag{12}$$

Moreover, let T_{RT} and $\mathbf{d}$ be independent random variables. Then

$$C_{\mathbf{u}_{T_{RT}}}(P) \subset C_{\mathbf{u}_{t^*_{RT}}}\left(\frac{P}{p^*}\right).$$

The proof of this proposition is provided in the Appendix.

For a given safety level P and choice of p^*, we can calculate t^*_{RT} based on the distribution function $F_{T_{RT}}(t_{RT})$ and use the result of Theorem 2 with $C_{\mathbf{u}_m}(P)$ replaced with $C_{\mathbf{u}_{t^*_{RT}}}\left(\frac{P}{p^*}\right)$ to obtain a safety supervisor for the case with actuator delay. Notice that p^* is a parameter of this method and different choices will lead to different approximations of the capture set, see Tables III and IV below. Further investigations are needed to develop a procedure that allows to choose the parameter p^* optimally. The tradeoff between p^* and P/p^* is illustrated in Figure 10.

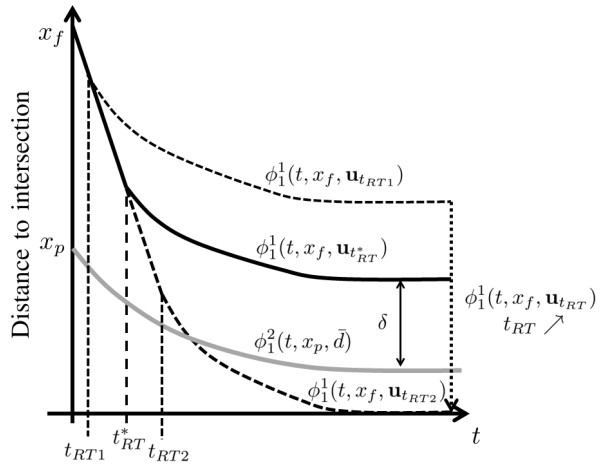

Fig. 10. The figure shows the position trajectories of both vehicles. The dashed lines correspond to position trajectories of FV for increasing values of t_{RT}. If $\Pr(\mathbf{d} \leq \bar{d}) = 1 - P^*$ and $F(t^*_{RT}) = p^*$then it is clear that increasing p^* requires increasing t^*_{RT} which means that the driver has to be warned earlier.

B. Identification of driver's reaction time distribution

In order to determine the distribution function $F_{T_{RT}}$ from data, a set of experiments on the test path depicted in Figure 5 was performed, in which we showed a visual warning to the driver, and measured the time the driver took to press the brake after the warning was issued. These experiments will be referred to as *reaction time experiments*. In these experiments, we have used Algorithm 1, with $P/p^* = 0.9$ and control input in the form of (11) with $t^*_{RT} = 1.5s$ preset reaction time [32]. All data was collected from the same middle-aged, male driver. Figure 11 shows the response of this driver (the u term of acceleration in (3a)) from onset of the warning, and the frequency of his response time to the warning. We collected a total of 104 of these braking trajectories of the test vehicle starting at the time the warning was issued. Figure 12 shows the schematic of the HMI (Human Machine Interface) that has been used to show the warning to the driver. This HMI allows the driver to compare his/her acceleration input with the required acceleration u_m in order to help him/her adjust the acceleration of the vehicle.

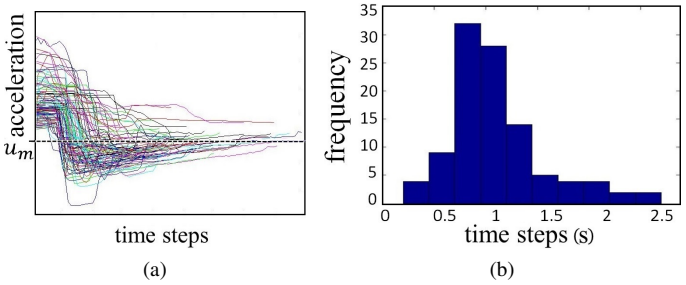

Fig. 11. (a) shows the braking profiles collected during the reaction time experiments. (b) shows the empirical distribution of the response time to the warning as observed during the reaction time experiments.

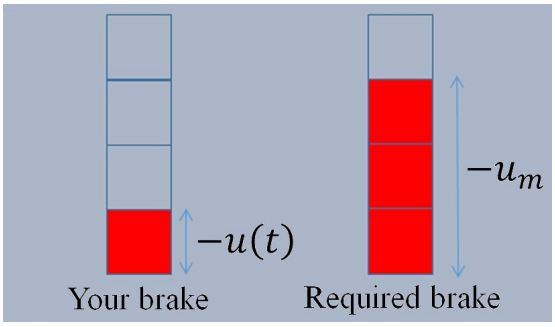

Fig. 12. The HMI used for warning.

C. Validation on naturalistic data set

Similar to Section VI-B, where we performed simulations using two different data sets to validate the safety of the supervisor based on direct overrides, we computed the empirical safety level (see Figure 7), with $\mathbf{u}_m$ replaced by $\mathbf{u}_{t^*_{RT}}$ and t^*_{RT} corresponding to different choices of p^*.

We used half of the data collected during the reaction time experiments for model identification, that is, to determine

the empirical distribution of T_{RT}. The remaining half of the braking profiles, not used for model identification, was used during the simulations instead of the fixed override input $\mathbf{u}_m$. Thus at each trial we randomly selected a braking profile among the profiles not used for model identification[1]. As for the direct override case, we performed two different tests corresponding to different data sets for the preceding vehicle. In the first test, we used 62 trajectories of the test-vehicle data set for model identification and the rest of the data for the simulations. The empirical safety levels for different values of P and p^* are shown in Table III.

TABLE III
EMPIRICAL SAFETY LEVEL ON THE TEST-VEHICLE DATA SET.

	Empirical safety level for		
Delay p^*	$P = 0.7$	$P = 0.8$	$P = 0.9$
1.0	0.923	0.948	0.974
0.94	0.831	0.862	0.945
0.91	0.796	0.872	0.940
0.87	0.770	0.851	-
0.83	0.737	0.848	-

For the second test, we used the entire test-vehicle data set for model identification and during simulations the trajectories from the radar data set were used for PV. The results of this second experiment are shown in Table IV.

TABLE IV
EMPIRICAL SAFETY LEVEL ON THE RADAR DATA SET.

	Empirical safety level for		
Delay p^*	$P = 0.7$	$P = 0.8$	$P = 0.9$
1.0	0.948	0.956	0.977
0.94	0.812	0.891	0.927
0.91	0.793	0.860	0.941
0.87	0.752	0.852	-
0.83	0.731	0.823	-

We conclude this section with a discussion of the driver's input model, equation (11). Since the rear-end collision situation corresponds to $x_p - x_f < \delta$, we compare the FV's positions corresponding to real driver's inputs measured during the reaction time experiments (Figure 11(a)), denoted by x_{driver}, with the FV's positions corresponding to a model input of the form (11), with T_{RT} replaced by $t^*_{RT} = 1.5s$, denoted by x_{model}. This comparison shows that for 98.1% of the trajectories, $x_{driver}(t) < x_{model}(t)$ for all $t \in [t^*_{RT}, t_l]$, where t_l corresponds to the length of the observed trajectory. Thus for the reaction time t^*_{RT}, which satisfies $\Pr(T_{RT} \leq t^*_{RT}) = 0.87$, the human driver managed in 98.1% of the cases to provide at least as much deceleration as the model x_{model} predicted. This is primarily due to the fact that the driver of the FV would in most cases reduce the acceleration in the vicinity of the intersection already before actively pushing the brakes. This explains the conservatism that we see as the model assumes a constant driver input during the reaction time.

[1]Since during the simulations we may require a profile of FV which is longer than trajectories of Figure 11(a), for the rest of such simulations we have used $\mathbf{u} = \mathbf{u}_m$.

VIII. CONCLUSION

In this paper, we have proposed a model-based approach to design driver-assist systems with a guaranteed probability of safety P. In particular, we have focused on a case study where a vehicle approaching a stop sign, speed bump or intersection, has to prevent a collision with a preceding vehicle. We modeled driver behavior (both for the preceding and following vehicles) through probabilistic models, in which the probability distributions of unknown parameters were identified from data. Our solution approach leveraged the monotone structure of the dynamics to provide an efficient algorithm for the real-time implementation of the driver-assist system. Simulations on a naturalistic data set demonstrate that the algorithm can indeed guarantee the desired safety level while being substantially less conservative than the deterministic counterpart.

There are many assumptions and simplifications made in this work that should be relaxed in future work. On the algorithmic side, we seek to extend the efficient computation of the probabilistic capture set to more general forms of the bad set B and system's dynamics. In these algorithms, we would also like to extend the control input model to actuators with stochastic delays and actuation uncertainty, to capture the possible difficulty a driver has to track the required control input and more accurately model physical limitations. More generally, it will be interesting to extend these approaches to cases where the driver model (both for the vehicle under study and for the surrounding vehicles) is identified and adjusted online to provide better adaptation to different drivers, different paths, and different vehicle parameters. Finally, human drivers are often modeled as multimodal systems [18] to take into account fundamentally different dynamical behaviors such as for instance braking and accelerating. An extension of the approach to such a hybrid setting should also be investigated.

APPENDIX

In the following we provide the proofs of the main results.

Proof of Theorem 1: ($\Leftarrow$) If $x \notin C_{\mathbf{u}_m}(P)$, then $\Pr(\phi(t, x, \mathbf{u}_m, \mathbf{d}) \notin B,\ \forall t \in [0, \tau_{end}]) \geq P$, which implies $x \in \mathcal{W}(P)$.

($\Rightarrow$) We prove the contrapositive, i.e. if $x \in C_{\mathbf{u}_m}(P)$ then $x \notin \mathcal{W}(P)$. From Proposition 1 we can distinguish two cases:

Case (1): If $x \in \mathbf{S}_2^{\mathbf{u}_m}$ then there exist a time $t \in [0, \tau_{end}]$, and $j \in \{1, ..., s\}$ such that $G^j(\phi^1(t, x^1, \mathbf{u}_m)) > g^j$, which from the order preserving property of Assumption 1 implies that for all $\mathbf{u} \in S(U)$, $G^j(\phi^1(t, x^1, \mathbf{u})) > g^j$. Hence, $\phi^1(t, x^1, \mathbf{u}, d) \in B_2$ and we conclude $x \notin \mathcal{W}(P)$.

Case (2): $x \in \mathbf{S}_1^{\mathbf{u}_m} \cap (\mathbf{S}_2^{\mathbf{u}_m})^c$. For a control input signal $\mathbf{u} \in S(U)$ we define

$$\Omega_{\mathbf{u}} := \left\{ d \in \Delta \;\middle|\; \exists\ t \in [0, \tau_{end}], \exists i \in \{1, ..., r\} \text{ such that } F_i^{t,x,\mathbf{u}}(d) > H_i \right\}.$$

From the order preserving properties of Σ^* and Assumption 1 it follows that $\mathbf{u} \mapsto F_i^{t,x,\mathbf{u}}(d)$ is an increasing function for all t, x, d. This implies that $\Omega_{\mathbf{u}_m} \subset \Omega_{\mathbf{u}}$ for all $\mathbf{u} \in S(U)$. Finally, using that $x \in \mathbf{S}_1^{\mathbf{u}_m}$ we have for all $\mathbf{u} \in S(U)$,

$$1 - P < \Pr(\Omega_{\mathbf{u}_m}) \leq \Pr(\Omega_{\mathbf{u}}),$$

which shows that $x \notin \mathcal{W}(P)$. ∎

Proof of Theorem 2: Let $\bar{d} \in \Delta$ be as in Proposition 1. Then it is clear that $\Pr(\mathbf{d} \geq \bar{d}) = P$. Consequently, it suffices to show that

$$\phi^{\pi}(t, x, d) \notin B, \quad \forall t \in [0, \tau_{end}], \quad \forall d \geq \bar{d}.$$

Assume to the contrary that there exist $d \geq \bar{d}$ and $t^* \in [0, \tau_{end}]$ such that

$$\phi^{\pi}(t^*, x, d) \in B. \tag{13}$$

As $B \subset C_{\mathbf{u}_m}(P)$, we have also that $\phi^{\pi}(t^*, x, d) \in C_{\mathbf{u}_m}(P)$. Thus defining $\bar{t} = \sup\{t \in [0, t^*] \mid \phi^{\pi}(t, x, d) \in C_{\mathbf{u}_m}(P)^c\}$, it follows from the continuity of the flow with respect to time that $\bar{x} := \phi^{\pi}(\bar{t}, x, d) \in \partial C_{\mathbf{u}_m}(P)$. Moreover, by the very definition of $\bar{t}$ and the feedback map π, for all $t \in [\bar{t}, t^*]$

$$\phi^{\pi}(t, x, d) = \phi(t - \bar{t}, \bar{x}, \mathbf{u}_m, d). \tag{14}$$

Next notice that $C_{\mathbf{u}_m}(P)$ is open (see Corollary 1 below), which in turn by Proposition 1 implies that for all $j \in \{1, \ldots, s\}$ and $t \in [0, \tau_{end}]$, there exists $i \in \{1, \ldots, q\}$ such that

$$F^{t,\bar{x},\mathbf{u}_m}(\bar{d}) \leq H \quad \text{and} \quad G_i^j(\phi^1(t - \bar{t}, \bar{x}^1, \mathbf{u}_m)) \leq g_i^j. \tag{15}$$

Finally, by the order preserving property of $d \to \phi^2(t, x, d)$, this implies also that

$$F^{t,\bar{x},\mathbf{u}_m}(d) \leq H, \quad \forall t \in [0, \tau_{end}]. \tag{16}$$

However, (14)-(16) assure that $\phi^{\pi}(t, x, d) \notin B$ for all $t \in [\bar{t}, t^*]$ contradicting (13). ∎

Remark 1: Notice that with similar arguments one can prove that if system Σ^2 is strictly input/output order preserving with respect to disturbance input then the probability of avoiding B with the feedback map π is exactly P.

Proof of Proposition 1: Step 1: We show that

$$C_{\mathbf{u}}(P) = \tilde{\mathbf{S}}_1^{\mathbf{u}} \cup \mathbf{S}_2^{\mathbf{u}},$$

where

$$\tilde{\mathbf{S}}_1^{\mathbf{u}} := \left\{x \in X \;\middle|\; \Pr(F^{t,x,u}(\mathbf{d}) \leq H \; \forall t \in [0, \tau_{end}]) < P\right\}.$$

Based on Assumption 1 the bad set can be written as $B = B_1 \cup B_2$. The P-safety capture set for input signal $\mathbf{u}$ for this bad set is given by

$$C_{\mathbf{u}}(P) = \{x \in X \mid \Pr\left(\phi(t, x, \mathbf{u}, \mathbf{d}) \notin B_1 \wedge \phi(t, x, \mathbf{u}, \mathbf{d}) \notin B_2, \; \forall t \in [0, \tau_{end}]\right) < P\}.$$

As a direct consequence of the definition of $\mathbf{S}_2^{\mathbf{u}}$, $\mathbf{S}_2^{\mathbf{u}} \subset C_{\mathbf{u}}(P)$. Therefore,

$$C_{\mathbf{u}}(P) \cap \mathbf{S}_2^{\mathbf{u}} = \mathbf{S}_2^{\mathbf{u}} = (\mathbf{S}_2^{\mathbf{u}} \cup \tilde{\mathbf{S}}_1^{\mathbf{u}}) \cap \mathbf{S}_2^{\mathbf{u}}. \tag{17}$$

From relationship (17), it suffices to prove[2]

$$C_{\mathbf{u}}(P) \cap (\mathbf{S}_2^{\mathbf{u}})^c = (\mathbf{S}_2^{\mathbf{u}} \cup \tilde{\mathbf{S}}_1^{\mathbf{u}}) \cap (\mathbf{S}_2^{\mathbf{u}})^c = \tilde{\mathbf{S}}_1^{\mathbf{u}} \cap (\mathbf{S}_2^{\mathbf{u}})^c. \tag{18}$$

[2]Since for sets A,B and C, if $A \cap B = C \cap B$ and $A \cap B^c = C \cap B^c$, then $A = A \cap (B \cup B^c) = (A \cap B) \cup (A \cap B^c) = (C \cap B) \cup (C \cap B^c) = C \cap (B \cup B^c) = C$.

From the definition of $C_{\mathbf{u}}(P)$, we have

$$C_{\mathbf{u}}(P) \cap (\mathbf{S}_2^{\mathbf{u}})^c = \{x \in (\mathbf{S}_2^{\mathbf{u}})^c \mid \Pr(\phi(t, x, \mathbf{u}, \mathbf{d}) \notin B_1 \wedge \phi(t, x, \mathbf{u}, \mathbf{d}) \notin B_2, \forall t \in [0, \tau_{end}]) < P\}. \tag{19}$$

Also, from the definition of $\mathbf{S}_2^{\mathbf{u}}$, it follows that if $x \in (\mathbf{S}_2^{\mathbf{u}})^c$, then for all $t \in [0, \tau_{end}]$, $\phi(t, x, \mathbf{u}, \mathbf{d}) \notin B_2$. Therefore, we can write (19) in the form of

$$C_{\mathbf{u}}(P) \cap (\mathbf{S}_2^{\mathbf{u}})^c = \{x \in (\mathbf{S}_2^{\mathbf{u}})^c \mid \Pr(\phi(t, x, \mathbf{u}, \mathbf{d}) \notin B_1, \forall t \in \mathbb{R}_+) < P\},$$

which based on the definition of $\tilde{\mathbf{S}}_1^{\mathbf{u}}$ and Assumption 1, can be further simplified as

$$C_{\mathbf{u}}(P) \cap (\mathbf{S}_2^{\mathbf{u}})^c = \{x \in (\mathbf{S}_2^{\mathbf{u}})^c \mid x \in \tilde{\mathbf{S}}_1^{\mathbf{u}}\} = (\mathbf{S}_2^{\mathbf{u}})^c \cap \tilde{\mathbf{S}}_1^{\mathbf{u}},$$

which proves (18).

Step 2: We have to show that $\tilde{\mathbf{S}}_1^{\mathbf{u}} = \mathbf{S}_1^{\mathbf{u}_m}$

($\subset$) We show instead $(\mathbf{S}_1^{\mathbf{u}_m})^c \subset (\tilde{\mathbf{S}}_1^{\mathbf{u}})^c$. Let $x \in (\mathbf{S}_1^{\mathbf{u}_m})^c$, then for all $t \in [0, \tau_{end}]$, $H \geq F^{t,x,\mathbf{u}_m}(\bar{d})$. Since the function $h^2(\phi^2(t, x^2, d))$ is order preserving with respect to d by the definition of an OPSD, then based on Assumption 1, $Z^2 h^2(\phi^2(t, x^2, d))$ is also order preserving with respect to d. Since on the other hand $h^1(\phi^1(t, x^1, \mathbf{u}))$ is not a function of d, then $F^{t,x,\mathbf{u}}(d)$ is a decreasing function of d. Hence, we have

$$\forall t \in [0, \tau_{end}], \forall d \geq \bar{d} : H \geq F^{t,x,\mathbf{u}_m}(d). \tag{20}$$

From the definition of $\bar{d}$ and (20), it follows that

$$\Pr(\forall t \in [0, \tau_{end}], H \geq F^{t,x,\mathbf{u}_m}(d)) \geq \Pr(d \geq \bar{d}) = P,$$

which implies that $x \in (\tilde{\mathbf{S}}_1^{\mathbf{u}})^c$.

($\supset$) Let $x \in \mathbf{S}_1^{\mathbf{u}_m}$, then there exists a $t \in [0, \tau_{end}]$, $i \in \{1, ..., r\}$ and $\epsilon > 0$ such that

$$F_i^{t,x,\mathbf{u}_m}(\bar{d}) = H_i + \epsilon.$$

Because of the continuity of $F_i^{t,x,\mathbf{u}_m}(\cdot)$, it follows that for some $\delta > 0$,

$$F_i^{t,x,\mathbf{u}_m}(\bar{d} + \delta) \geq H_i + \frac{\epsilon}{2} > H_i. \tag{21}$$

Moreover, because of the decreasing property of $F_i^{t,x,\mathbf{u}_m}(\cdot)$, (21) implies that

$$\forall d \leq \bar{d} + \delta : F_i^{t,x,\mathbf{u}_m}(d) > H_i,$$

and because the cumulative distribution function Φ is increasing, we have

$$\Pr(\exists t \in [0, \tau_{end}], \exists i \in \{1, ..., r\}, F_i^{t,x,\mathbf{u}_m}(d) > H_i) \geq \Pr(d \leq \bar{d} + \delta) = \Phi(\bar{d} + \delta) > 1 - P. \tag{22}$$

Finally, noticing that (22) is equivalent to

$$\Pr(\forall t \in [0, \tau_{end}], F^{t,x,\mathbf{u}_m}(d) \leq H) < P,$$

completes the proof. ∎

A direct and useful consequence of Proposition 1 is the following:

Corollary 1: The set $C_{\mathbf{u}_m}(P)$ is open.

Proof: For all $t \in [0, \tau_{end}]$ and $i \in \{1, ..., r\}$ we define

$$S_{t,i} := \left\{x \in X \;\middle|\; H_i < F_i^{t,x,\mathbf{u}_m}(\bar{d})\right\}.$$

Then by continuity of $F_i^{t,x,\mathbf{u}_m}(\bar{d})$ with respect to x, $S_{t,i}$ is an open set. From Proposition 1, we have

$$\mathbf{S}_1^{\mathbf{u}_m} = \bigcup_{\substack{t\in[0,\tau_{end}]\\ i\in\{1,\ldots,r\}}} S_{t,i}.$$

Therefore $\mathbf{S}_1^{\mathbf{u}_m}$ is also an open set. Since $\mathbf{S}_2^{\mathbf{u}_m}$ is open, $C_{\mathbf{u}_m}(P) = \mathbf{S}_1^{\mathbf{u}_m} \cup \mathbf{S}_2^{\mathbf{u}_m}$ is open as well. ∎

Proof of Proposition 2: If $x \in C_{\mathbf{u}_{T_{RT}}}(P)$, then from the definition of the capture set, we obtain

$$\Pr(\forall t \in [0,\tau_{end}], \phi_1^2(t,x^2,\mathbf{d}) - \phi_1^1(t,x^1,\mathbf{u}_{T_{RT}}) \geq \delta) < P. \tag{23}$$

By conditioning on the two events $T_{RT} < t_{RT}^*$ and $T_{RT} \geq t_{RT}^*$, we can write relationship (23) in the form of

$$\begin{aligned}\Pr(\forall t, \phi_1^2(t,x^2,\mathbf{d}) - \phi_1^1(t,x^1,\mathbf{u}_{T_{RT}}) \geq \delta \mid T_{RT} < t_{RT}^*)\cdot \\ \Pr(T_{RT} < t_{RT}^*) + \Pr(\forall t, \phi_1^2(t,x^2,\mathbf{d}) - \\ \phi_1^1(t,x^1,\mathbf{u}_{T_{RT}}) \geq \delta \mid T_{RT} \geq t_{RT}^*)\Pr(T_{RT} \geq t_{RT}^*) < P.\end{aligned} \tag{24}$$

Using (12) and the fact that the second term in (24) is non-negative, we derive that

$$\begin{aligned}\Pr(\forall t \in [0,\tau_{end}], \phi_1^2(t,x^2,\mathbf{d}) - \phi_1^1(t,x^1,\mathbf{u}_{T_{RT}}) \geq \delta \mid \\ T_{RT} < t_{RT}^*) < \frac{P}{p^*}.\end{aligned} \tag{25}$$

Since $u_0 \geq \mathbf{u}_m$, for all times $t_{RT} < t_{RT}^*$ we have $\mathbf{u}_{t_{RT}}(t) \leq \mathbf{u}_{t_{RT}^*}(t)$. This holds since for $t \in [0, t_{RT}) \cup [t_{RT}^*, \tau_{end}]$ we have $\mathbf{u}_{t_{RT}}(t) = \mathbf{u}_{t_{RT}^*}(t)$, and for $t \in [t_{RT}, t_{RT}^*)$ we have $\mathbf{u}_{t_{RT}}(t) - \mathbf{u}_{t_{RT}^*}(t) = \mathbf{u}_m - u_0 \leq 0$. Therefore, by the order preserving property with respect to the input, for all $t \in [0,\tau_{end}]$, $\phi_1^2(t,x^2,\mathbf{d}) - \phi_1^1(t,x^1,\mathbf{u}_{t_{RT}}) \geq \phi_1^2(t,x^2,\mathbf{d}) - \phi_1^1(t,x^1,\mathbf{u}_{t_{RT}^*})$. Since this relationship is valid for all $t_{RT} < t_{RT}^*$ and T_{RT} is independent of $\mathbf{d}$, we have

$$\begin{aligned}\Pr(\phi_1^2(t,x^2,\mathbf{d}) - \phi_1^1(t,x^1,\mathbf{u}_{T_{RT}}) \geq \delta\ \forall t \mid T_{RT} < t_{RT}^*) \\ \geq \Pr(\phi_1^2(t,x^2,\mathbf{d}) - \phi_1^1(t,x^1,\mathbf{u}_{t_{RT}^*} \geq \delta\ \forall t).\end{aligned} \tag{26}$$

Relationships (25) and (26) together imply that $x \in C_{\mathbf{u}_{t_{RT}^*}}(\frac{P}{p^*})$. ∎

REFERENCES

[1] M. Althoff, J. M. Dolan, "Online verification of automated road vehicles using reachability analysis", *IEEE Trans. Robotics*, vol. 30, pp. 903-918, 2014.

[2] M. Bujorianu, "Extended stochastic hybrid systems and their reachability problem", in *Lecture notes in computer science: Hybrid systems: computation and control*, R. Alur, G. Pappas (Eds.), vol. 2993, pp. 234249, 2004.

[3] E. Coelingh, L. Jokobsson, H. Lind, M. Lindman, "Collision warning with auto brake - A real-life perspective", Nat. Highway Traffic Safety Admin., April, 2007.

[4] E. Coelingh, A. Eidehall, M. Bengtsson, "Collision warning with full auto brake and pedestrian detection - a practical example of automatic emergency braking", in *Proc. of IEEE Conference on Intelligent Transportation Systems*, 2010, pp. 155-160.

[5] A. Colombo, D. Del Vecchio, "Efficient algorithms for collision avoidance at intersections", in *Proc. of the ACM International Conference on Hybrid Systems: Computation and Control*, 2012, pp. 145-154.

[6] V. Desaraju, H. C. Ro, M. Yang, E. Tay, S. Roth, D. Del Vecchio, "Partial order techniques for vehicle collision avoidance: Application to an autonomous roundabout test-bed", in *Proc. IEEE Conference on Robotics and Automation*, 2009, pp. 82-87.

[7] P. A. Devijver, J. Kittler, *Pattern Recognition: A Statistical Approach*, Prentice-Hall, 1982.

[8] J. Duperret, M. Hafner, D. Del Vecchio, "Formal design of a provably safe robotic roundabout system", in *Proc. IEEE/RSJ Conference on Intelligent Robots and Systems*, 2010, pp. 2006-2011.

[9] P. M. Esfahani, D. Chatterjee, J. Lygeros, "On a problem of stochastic reach-avoid set characterization", in *Proc. IEEE Conference on Decision and Control*, 2011, pp. 7069-7074.

[10] V. Gadepally, A. Kurt, A. Krishnamurthy, Ü. Özgüner, "Driver/Vehicle state estimation and detection", in *Proc. of IEEE Conference on Intelligent Transportation Systems*, 2011, pp. 582-587.

[11] D. Hoehener, P. A. Green, D. Del Vecchio, "Stochastic hybrid models for predicting the behavior of drivers facing the yellow-light-dilemma", in *Proc. of American Control Conference*, 2015, pp. 3348-3354.

[12] M. Hafner, D. Cunningham, L. Caminiti, and D. Del Vecchio, "Cooperative collision avoidance at intersections: Algorithms and experiments", *IEEE Trans. Intelligent Transportation Systems*, vol. 14, pp. 1162-1175, 2013.

[13] M. Hafner, D. Del Vecchio, "Computational tools for the safety control of a class of piecewise continuous systems with imperfect information on a partial order", *SIAM J. Control Optim.*, vol. 49, pp. 24632493.

[14] P. Kafka, "The automotive standard ISO 26262, the innovative driver for enhanced safety assessment & technology for motor cars", *Procedia Engineering*, vol. 45, pp. 2-10, 2012.

[15] R. Keifer, D. LeBlanc, M. Palmer, J. Salinger, R. Deering, M. Shulman, "Development and validation of functional definitions and evaluation procedures for collision warning/avoidance systems", *Final Report DOT HS 808 964*, Washington, DC: U.S. Department of Transportation, 1999.

[16] S. E. Lee, S. B. Brown, M. A. Perez, Z. R. Doerzaph, V. L. Neale, "Normal and hard braking behavior at stop signs and traffic signals", in *Proc. of Human Factors and Ergonometrics Society Annual Meeting*, 2005, pp. 1897-1901.

[17] J. C. McCall, M. M. Trivedi, "Driver behavior and situation aware brake assistance for intelligent vehicles", *Proceeding of IEEE*, vol. 95, pp. 374-387, 2007.

[18] A. Pentland, A. Liu, "Modeling and prediction of human behavior", *Neural Computation*, vol. 11, pp. 229242, 1999.

[19] S. Prajna, A. Jadbabaie, G. J. Pappas, "A framework for worst-case and stochastic safety verification using barrier certificates", *IEEE Trans. on Automatic Control*, vol. 52, pp. 1415-1428, 2007.

[20] M. Prandini and J. Hu, "Application of reachability analysis for stochastic hybrid systems to aircraft conflict prediction", in *Proc. IEEE Conference on Decision and Control*, 2008, pp. 4036-4041.

[21] R. Rajamani, *Vehicle Dynamics and Control*, Springer, 2012.

[22] V. A. Shia, Y. Gao, R. Vasudevan, K. Driggs Campbell, Th. Lin, F. Borrelli, R. Bajcsy, "Semiautonomous vehicular control using driver modeling", *IEEE Trans. Intelligent Transportation Systems*, vol. 15, pp. 2696-2709, 2014.

[23] C. Tomlin, J. Lygeros, S. S. Sastry, "Controller synthesis for hybrid systems: the Hamilton-Jacobi approach,", in *AAAI Spring Symposia*, 1999, pp. 192-197.

[24] C. Tomlin, J. Lygeros, S. S. Sastry, "A game theoretic approach to controller design for hybrid systems", *Proceedings of IEEE*, vol. 88, pp. 949-970, 2000.

[25] C. Tomlin, I. Mitchell, A. M. Bayen, M. Oishi, "Computational techniques for the verification of hybrid systems", *Proceeding of IEEE*, vol. 91, pp. 986-1001, 2003.

[26] R. Verma, D. Del Vecchio, "Semiautonomous multivehicle safety: A hybrid control approach", *IEEE Robotics and Automation Magazine*, vol. 18, pp. 44-54, 2011.

[27] R. Verma, D. Del Vecchio, "Safety control of hidden mode hybrid systems", *IEEE Trans. on Automatic Control*, vol. 57, pp. 62-77, 2012.

[28] C. L. Walton, D. Del Vecchio, R. M. Murray, "Identification of decision rules in a human controlled system: vehicles at a traffic intersection", in *Proc. of International Conference of Robotics and Automation*, 2004, pp. 1173 - 1178.

[29] "2012 motor vehicle crashes: Overview. U.S. Department of Transportation National Highway Safety Administration", http://www-nrd.nhtsa.dot.gov/Pubs/811856.pdf

[30] "Intersection Safety", http://safety.fhwa.dot.gov/intersection/

[31] "Quick work: Better autobrake helps more models earn top ratings for front crash prevention", Insurance Institute for Highway Safety, Status Report, vol. 49, May 2014, http://www.iihs.org/iihs/news/desktopnews/quick-work-better-autobrake-helps-more-models-earn-top-ratings-for-front-crash-prevention.

[32] "Driver Distraction, Warning Algorithm Parameters, and Driver Response to Imminent Rear-end Collisions in a High-Fidelity Driving Simulator", National Highway Traffic Safety Administration, Technical Report, March 2002.

[33] "Automatic Emergency Braking System (AEB)", National Highway Traffic Safety Administration, Research Report, August 2014, http://www.automotivesafetycouncil.org/files/NHTSA%20AEB%20Report.pdf.