Correct-by-Construction Advanced Driver Assistance Systems based on a Cognitive Architecture

Francisco Eiras[1], Morteza Lahijanian[2], and Marta Kwiatkowska[3]

***Abstract*— Research into safety in autonomous and semi-autonomous vehicles has, so far, largely been focused on testing and validation through simulation. Due to the fact that failure of these autonomous systems is potentially life-endangering, formal methods arise as a complementary approach. This paper studies the application of formal methods to the verification of a human driver model built using the cognitive architecture ACT-R, and to the design of correct-by-construction Advanced Driver Assistance Systems (ADAS). The novelty lies in the integration of ACT-R in the formal analysis and an abstraction technique that enables finite representation of a large dimensional, continuous system in the form of a Markov process. The situation considered is a multi-lane highway driving scenario and the interactions that arise. The efficacy of the method is illustrated in two case studies with various driving conditions.**

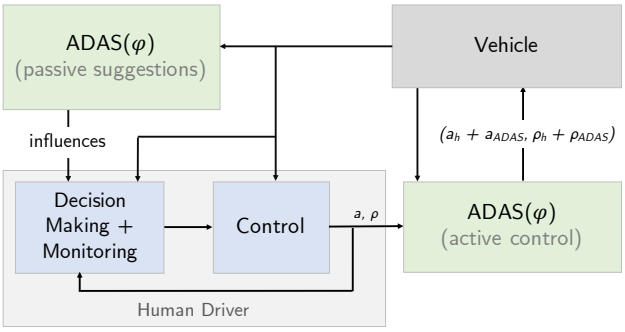


Fig. 1: Overview of the ADAS design with passive and active interventions for a specification φ.

I. Introduction

Humans do not have a good track record on the road. Road accidents kill 1.24 million people every year and over 90% of all crashes are mainly attributed to errors of human drivers [1]. While full self-driving technology is not yet available at scale, in an attempt to reduce these numbers, several car manufacturers have introduced semi-autonomous features in the form of Advanced Driver Assistance Systems (ADAS). Examples include Tesla's *Autopilot* and Ford's *Co-Pilot 360*. However, ensuring safety for semi-autonomous vehicles remains a major challenge with roots in the lack of coherent understanding of the human-ADAS interaction.

Existing methods to validate the safety of semi-autonomous systems rely on testing and simulation. Using real data to take statistically significant conclusions, however, is infeasible due to the time it takes to collect a sufficiently large amount of data [2]. Several approaches are based on modeling and simulating the semi-autonomous vehicle, as proposed in [3]–[6]. Despite this, it is imperative to recognize the shortcomings of simulation in safety evaluation of complex driver assistance systems which could have life-endangering impact [7], [8].

A promising direction is to employ formal verification techniques, which are based on rigorous mathematical reasoning, to obtain strong guarantees about the ADAS, as proposed by several recent works [9]–[11]. In [10], Nilsson *et al.* synthesized a provably-correct module for adaptive cruise control from specifications given in temporal logic. However, these studies ignore the human driver behavior variability presented by a driver model, which can lead to controllers that perform poorly in corner cases. On the other hand, [11] applies model checking techniques to the verification of data-driven models of human driver behavior, yet it does not explicitly model the human cognitive process nor does it leverage this analysis as a way to bootstrap safety in the form of an ADAS.

The overarching goal of this study is to provide safety guarantees in semi-autonomous vehicles through the integration of human cognition with formal methods. As a first step in this direction, this paper focuses on giving guarantees at the design level of the ADAS. Specifically, it employs the cognitive architecture known as Adaptive Control of Thought-Rational (ACT-R): a framework for specifying computational behavioral models of human cognitive performance, embodying both the abilities (e.g. memory storage and recall, perception or motor action) and constraints (e.g. memory decay and limited motor performance) of humans [12]–[17]. The work builds on the human driver model in a multi-lane highway driving scenario presented in [15]. It also expands upon [9] by applying verification techniques to an efficient abstraction of the model and extends it to allow the intervention of a provably-correct synthesized ADAS based on specifications given as temporal logic formulas.

The main contribution of this paper is threefold: first, it studies the verification of a human driver model built in a cognitive architecture through efficient model abstraction techniques. Second, it builds upon the model of human driving behavior as a way to bootstrap the desired properties

This work was partially supported by EPSRC Mobile Autonomy Program Grant EP/M019918/1. FiveAI provided a travel grant to support the presentation of this work.

[1]Francisco Eiras was a student at the Dept. of Computer Science, University of Oxford, UK, when this work was developed and is now with FiveAI francisco.eiras@five.ai

[2]Morteza Lahijanian is with the Dept. of the Ann and H.J. Smead Aerospace Engineering Sciences, University of Colorado Boulder morteza.lahijanian@colorado.edu

[3]Marta Kwiatkowska is with the Dept. of Computer Science, University of Oxford, UK marta.kwiatkowska@cs.ox.ac.uk

in the ADAS using formal methods. Third, it introduces a flexible framework in terms of specifications which allows for different guarantees to be obtained depending on the choices made by the ADAS designer. Other contributions of this work include case studies based on specific properties and an open source implementation of the framework. To the best of our knowledge, this is the first framework that brings formal reasoning to the design of semi-autonomous vehicle solutions by taking into account the cognitive process of the human.

II. Problem Formulation

We consider the driving scenario studied in [15], where a vehicle, called the *ego-vehicle*, is in an interaction with a lead vehicle in a multi-lane highway. We are interested in designing a correct-by-construction ADAS system for the ego-vehicle.

A. Vehicle Model

We consider the ego-vehicle kinematics are described by

$$\begin{aligned} \Delta x &= v\cos(\psi+\rho)\Delta t, & \Delta y &= v\sin(\psi+\rho)\Delta t, \\ \Delta v &= a\Delta t, & \Delta\psi &= \frac{2v}{l}\sin(\rho)\Delta t, \end{aligned} \tag{1}$$

where x and y are the coordinates of the vehicle's center of mass, v is the speed, and ψ is the heading angle of the vehicle. The control inputs are steering angle ρ and acceleration a. Finally, l is the length of the vehicle, and Δt is the time duration between two iterations of the model.

We assume that the motion of the lead vehicle is predictable. This simplifying assumption, even though not realistic in large scale, is reasonable for small road segments due to the predictability of highway driving and the possible improvements that can be introduced by using data [18].

B. Integrated Human Driver Model in ACT-R

The ego-vehicle is driven by a human, whose behavior is represented in ACT-R. ACT-R is a framework for specifying computational behavioral models of human cognitive performance [12]–[17]. It embodies two crucial cognitive aspects of humans: the abilities (e.g., memory storage, perception, and motor action) and the constraints (e.g. memory decay and limited motor performance). ACT-R can be generally described as two distinct layers: a perceptual-motor layer and a cognitive layer. The perceptual-motor layer corresponds to the interface of the cognition with the environment, being comprised of modules such as vision and motor actions. The cognitive layer is focused on memory, which can be divided into two different categories: declarative (consisting of factual knowledge and goals - e.g., *"The maximum driving speed in a typical US highway is 65 mph"* or *"Try to get to point B"*) and procedural (consisting of rules/procedures - e.g., *"If the lead vehicle is going slowly, attempt an overtake"*) [12].

Particularly, we focus on the model proposed by [15], which is an improved version of the model from [14] based on advances in ACT-R and real world data. It describes how a human controls a vehicle and performs an action (e.g. lane change), in the presence of other vehicles. The model, shown schematically in Fig. 2, consists of three distinct modules interacting in a sequential way:

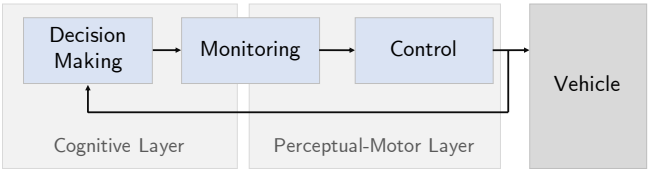


Fig. 2: Schematic overview of the interaction of the *control*, *monitoring* and *decision making* modules of the integrated human driver model in ACT-R.

- *control*, which manages both the lower level perception cues and the physical manipulation of the vehicle;
- *monitoring*, which maintains situational awareness through the awareness of the position of other vehicles around the ego-vehicle; and
- *decision making*, which uses the information gathered in the monitoring and control stage to determine the tactical decision to be taken (whether or not a lane change should happen).

A full description of the model and the governing equations for vehicle control generations are provided in Sec. IV.

C. ADAS Design

We consider an ADAS that corresponds, first and foremost, to determining the possible available interventions to the system at each point in time. The actions considered must be realistic in nature; otherwise, the obtained assistance system would prove to be incompetent in a real-world scenario. Fig. 1 summarizes the interventions we consider for the ADAS, which are divided into two types: *passive suggestions* and *active control*.

In passive suggestions, it is assumed that the assistance system cannot change the decision making directly (as it is a human cognitive process), but it can influence it to a certain degree through suggestions [19]. Hence, the control inputs to the vehicle are directly provided by the human (ACT-R), i.e., $a = a_h$ and $\rho = \rho_h$, where subscript h corresponds to the human, and the ADAS can only provide suggestions that can lead to safe and correct behaviors.

In active control, ADAS can have incremental control-based interventions at the level of acceleration and steering, i.e., $a = a_h + a_{\text{ADAS}}$ and $\rho = \rho_h + \rho_{\text{ADAS}}$, where the ADAS variables are constrained to ensure incremental interventions. The full details of the action availability, constraints, and intervention of the ADAS are presented in Sec. V.

D. Specification Language

To formally express the behavioral properties of interest for the semi-autonomous system, we use *Probabilistic Computation Tree Logic* (PCTL) [20]. A PCTL formula combines boolean and temporal operators with probabilistic reasoning, constituting a rich specification language.

Definition 1 (PCTL Syntax)**.** A *PCTL formula* Φ over a set of atomic propositions AP can be formed according to the following grammar:

$$\Phi := \texttt{true} \mid o \mid \Phi \wedge \Phi \mid \Phi \vee \Phi \mid \neg\Phi \mid \mathbb{P}_{\sim p}(\varphi)$$
$$\varphi := \bigcirc \Phi \mid \Phi_1 \mathcal{U}^{\leq k} \Phi_2 \mid \Phi_1 \mathcal{U} \Phi_2 \mid \Diamond \Phi$$

where $o \in AP$ is an atomic proposition, $\wedge$ ("and"), $\vee$ ("or"), and $\neg$ ("negation") are boolean operators, and $\bigcirc$ ("next"), $\mathcal{U}^{\leq k}$ ("bounded until") with $k \in \mathbb{N}$, $\mathcal{U}$ ("until"), and $\Diamond$ ("eventually") are temporal operators. $\mathbb{P}$ is the probabilistic operator, and $\sim p$ is a probability bound. The formulae Φ and φ are called state and path formalas, respectively.

In this work, the atomic propositions represent boolean facts about the driving scenario. Through them, we can express properties of interest using PCTL, e.g., "The probability that eventually the distance to the nearest car becomes less than d_{safe} is less than 0.001" can be expressed as $\mathbb{P}_{<0.001}\big(\Diamond(\| \mathbf{x} - \mathbf{x}_{\text{near}} \| < d_{\text{safe}})\big)$.

E. Problem Statement

Given a vehicle, whose motion is described by (1), and a human driver represented by ACT-R, a set of initial conditions defined as a scenario $\mathcal{S}$, and a PCTL formula φ, we are interested in the following two problems:

Problem 1 (verification)**.** Compute the probability that the human-vehicle system satisfies φ in $\mathcal{S}$, i.e, $\mathbb{P}^{\mathcal{S}}(\varphi)$.

Problem 2 (synthesis)**.** Design an ADAS that optimizes the probability of satisfying φ by the human-vehicle-ADAS system in $\mathcal{S}$, i.e., $\mathbb{P}^{\mathcal{S}}_{\bowtie}(\varphi)$ with $\bowtie \ \in \{\max, \min\}$.

This is a flexible problem representation under which the specification φ comes from the designer of the ADAS. It should be noted that the two-vehicle scenario considered in this study is non-limiting as traffic in highways tends to be sparse, allowing to reason over each of the vehicles separately as in [15]. In addition, the proposed solution to Problems 1 and 2 is general and can be easily extended to more vehicles.

III. Preliminaries

In this study, we employ Markov models as the abstractions for the driving scenarios.

Definition 2 (Markov Chain (MC))**.** A MC is a tuple $\mathcal{M} = (S, \mathbf{P}, s_0, AP, \iota)$, where S is a finite set of states, $\mathbf{P} : S \times S \to [0, 1]$ is a transition probability function, $s_0 \in S$ is the initial state, AP is a set of atomic propositions, and $\iota : S \to 2^{AP}$ is a labelling function.

Definition 3 (Markov Decision Process (MDP))**.** An MDP is a tuple $\mathcal{M} = (S, Act, \mathbf{P}, s_0, AP, \iota)$, where S, s_0, AP, and ι are as in Definition 2, Act is a finite set of actions, and $\mathbf{P} : S \times Act \times S \to [0, 1]$ is a transition probability function. The set of actions available in state $s \in S$ is denoted by $Act(s)$.

Definition 4 (Path & Policy)**.** A *finite path* of an MDP is a finite sequence of states $s_0 s_1 \dots s_n$ such that the transition probability from s_i to s_{i+1} is non-zero under some action in $Act(s_i)$ for all $i \in \{0, \dots, n-1\}$. The set of all finite paths are denote by S^*. A *policy* for an MDP is a function $\pi : S^* \to Act$ that maps a finite path to an action such that $\pi(s_0 s_1 \dots s_n) \in Act(s_n)$. The set of all policies is denoted by Π.

IV. Abstraction and Verification of the Human-Vehicle System

To verify the human driver model under a specification φ, we first abstract it to a Markov Chain $\mathcal{M}_h$. We achieve this by discretizing the individual modules of the integrated human driver ACT-R model in [15] through the use of the vehicle model. We can then use off-the-shelf tools, e.g., PRISM [21], to perform the verification of the abstracted model. Below, for the purpose of clarify of presentation, we detail the abstraction procedure for a two-lane highway scenario, but we emphasize that the method extends trivially to n lanes.

A. Control Module

The control module of ACT-R is fully deterministic and can be divided into lateral (i.e. steering) and longitudinal (i.e. acceleration) control. The lateral control is determined by the existence of two artifacts that the driver obtains using low-level perception cues: the near and far points. In each ACT-R cycle, the model uses perception to determine the difference in visual angles $\Delta\theta_{\text{near}}$ and $\Delta\theta_{\text{far}}$ and the difference control law for the steering angle ρ_h is:

$$\Delta\rho_h = k_{\text{far}}\Delta\theta_{\text{far}} + k_{\text{near}}\Delta\theta_{\text{near}} + k_I \min(\theta_{\text{near}}, \theta_{\max})\Delta t, \quad (2)$$

where k_{far}, k_{near} and k_I are proportional control gains, and $\theta_{\max}$ is the maximum steering angle [15]. The process for the longitudinal control is similar. In each ACT-R cycle, the model starts by encoding the position of the lead vehicle and calculating the time headway to it, as well as the difference between this and the previous cycle, $\Delta t^{\text{hw}}_{\text{car}}$. The difference control law for the acceleration a_h can then be written as:

$$\Delta a_h = k_{\text{car}}\Delta t^{\text{hw}}_{\text{car}} + k_{\text{follow}}(t^{\text{hw}}_{\text{car}} - t^{\text{hw}}_{\text{follow}})\Delta t, \quad (3)$$

where k_{car} and k_{follow} are proportional gains of the control, and $t^{\text{hw}}_{\text{follow}}$ is the threshold time headway for following a vehicle [15]. To initiate a lane change, the driver begins following the near and far points of the destination lane instead of the current one [22].

The most direct approach to abstracting this module, widely seen in the literature for small scenarios (e.g. [18], [23]–[25]) is to represent the road as a grid with the position of the ego-vehicle being a cell in the grid. The error associated with this method of discretizing space can be reduced by decreasing the cell area, i.e., increase in resolution. However, this incurs in the problem of state explosion: as the resolution increases, the number of states in the system grows exponentially and the verification becomes intractable.

In this work, we take a different approach and focus on reducing the dimensionality of the problem into a less

error-prone space. We project the human-vehicle system state $(x, y, v, \psi, a, \rho, t) \in \mathbb{R}^7$ to $\mathbf{x} = (x, v, \lambda, a, t) \in \mathbb{R}^4 \times \{0, 1\}$, where x is bounded to a finite length of the road given by the scenario $\mathcal{S}$, and $\lambda \in \{0, 1\}$ represents the index of the lane (left or right). A time discretization is induced by Δt for all the continuous variables. Note that t is included in $\mathbf{x}$ to enable the tracking of the state of the other vehicle, whose motion is assumed to be known (see Sec. II-A). We further reduce the representation by compressing the lane change maneuver into a single transition, as described below.

The evolution of the compressed model is as follows. When the vehicle is following its current lane, λ remains the same, and x and v are given by (1) (y is the center of lane λ and $\psi = 0$) with the control input $\Delta\rho_h$ being zero and Δa_h given by (3). When a lane change is decided, the controls and state of the vehicle are given by (1)-(3). We declare the maneuver is complete when the vehicle has merged to the center of the final lane, updating λ. During the maneuver, we monitor the change in the truth values of the atomic propositions in addition to possible collisions. Then, we discard the maneuver trajectory and record only the two states, at which the lane-change maneuver starts and ends, and label the latter state with the propositions of the maneuver. These values can be pre-computed, stored in lookup tables, and used for deterministic transitions between the control and the next ACT-R step, producing significantly smaller models.

B. Decision Making and Monitoring

The decision making process to move from the right to the left lane consists of localizing the lead vehicle in the right lane and deciding whether or not to change lanes based on the time headway, $t_{\text{car}}^{\text{hw}}$. The lower this time headway, the more likely a driver is to perform the manoeuvre [26]. Let d to be the distance between the two vehicles. We represent the probability of the driver performing a lane change to the left lane with an exponentially decreasing function (as in [27], [28]):

$$P_{l_c}(t_{\text{car}}^{\text{hw}} \mid \lambda = 0) = e^{-\alpha t_{\text{car}}^{\text{hw}}}, \tag{4}$$

where α is a parameter of the decision making.

A similar approach can be applied for a driver in the left lane overtaking a vehicle behind it in the right lane, except in this case the opposite effect occurs in the decision making. In such a case, the probability of changing lane can be modelled as a normalized logarithmic function over the distance between the vehicles:

$$P_{l_c}(d, v \mid \lambda = 1) = \frac{\log(\beta d + 1)}{\log(\beta d_{\max} + 1)}, \tag{5}$$

where $d_{\max}$ is the maximum length considered in the scenario, and β is a parameter of the decision making. It should be noted that the values of α and β could be estimated from real data for a population of drivers [27], [28].

So far, this version of decision making is not influenced by the monitoring module at all, and it relies on the measurements of the values of $t_{\text{car}}^{\text{hw}}$ and d by the human. It is unrealistic to assume that the human's measurements are perfect. In order to reflect uncertainty in these values, stochastic noise is added to the measurement of d (as this is what human drivers have to instinctively measure through perception). The noise is considered to be normally distributed $w \sim \mathcal{N}(0, \sigma)$. For an integral resolution parameter, $\delta \in [0, d]$, and L as the number of discrete steps for x, we can define:

$$P'_{l_c}(d, v) = \sum_{i=-L}^{L} P_{l_c}(d + i, v) \int_{d+i-\delta/2}^{d+i+\delta/2} w(z)dz. \tag{6}$$

Similarly to the control module, the values of P'_{l_c} can be pre-computed and stored in a table to be used in stochastic transitions to the ACT-R control step of the following cycle.

C. Markov Chain Abstraction

We now define a finite MC $\mathcal{M}_h = (S, \mathbf{P}, s_0, AP, \iota)$ that unifies both modules using the discretization described above and a variable $\mu \in \{1, 2\}$, where $\mu = 1$ corresponds to the control step and $\mu = 2$ to the decision making stage.

We define a state $s \in S$ of $\mathcal{M}_h$ to be a tuple $s = (\mu, x, \lambda, a, v, t)$. For a given scenario $\mathcal{S} = (\lambda_0, x_0, v_0, \mathbf{x}^{ov})$, where $\mathbf{x}^{ov}$ is the state of the other vehicle, the state space S is automatically generated. The transition probabilities for all $s, s' \in S$ are given by:

$$\mathbf{P}(s, s') = \begin{cases} 1 & \text{if } \mu_s = 1 \wedge s' = \textsc{Control}(s), \\ \text{DMM}(s, s') & \text{if } \mu_s = 2, \\ 0 & \text{otherwise,} \end{cases}$$

where CONTROL is the lookup table for the control step described in Sec IV-A and DMM is the probability table for the decision making and monitoring stage described in Sec IV-B. The set AP and labeling function ι are naturally mapped according to the tuple elements of each state s. It is important to note that the generated model is symbolic in nature, adding to the flexibility of the framework. Furthermore, it is worth noting that $\mathcal{M}_h$ captures in a one-to-one mapping all the possible outcomes of the continuous integrated driver model, under the assumptions of the distributions given by (4), (5) and (6).

D. Verification of the Human-Vehicle System

Given the model $\mathcal{M}_h$, we are interested in computing the probability of satisfying a property φ for a given scenario $\mathcal{S}$, i.e, Problem 1. This probability is defined as:

$$\mathbb{P}^{\mathcal{S}}(\varphi) = \Pr(s_0 \models \varphi), \tag{7}$$

that is, the probability of φ holding in $\mathcal{M}_h$ from an initial state s_0. This problem has been extensively studied in the literature [20], and many linear programming based solutions for it exist using off-the-shelf tools, e.g. PRISM [21], hence solving Problem 1.

V. SYNTHESIS FRAMEWORK FOR THE ADAS

In this section, we focus on the design of the ADAS and its representation as an MDP.

A. Passive Suggestions

At the decision making level, the human driver model in [15] has two options: it can either change lane or continue in the current lane. These options can be influenced using suggestions (e.g. through visual or auditive cues) [19]. If a driver can be influenced to make a conscious decision to decelerate (e.g. through the suggestions of the ADAS), then there is an argument for including this action in the decision making. Thus, we consider a 3-option ADAS with the following set of action suggestions:

$$Act = \{\mathrm{a_{cl}},\ \mathrm{a_{con}},\ \mathrm{a_{dec}}\},$$

where $\mathrm{a_{cl}}$, $\mathrm{a_{con}}$, and $\mathrm{a_{dec}}$ represent "*change lane*", "*continue driving in this lane*", and "*decelerate*" respectively. We assume that the human applies a constant deceleration value a_d when the deceleration decision is made, i.e., $a_h = a_d$. Then, we can abstract this human-vehicle-ADAS system as an MDP in a similar fashion to the MC abstraction above. Note that the MDP includes additional states that correspond to the decision (action) of deceleration. These states can be computed using the same procedure in Sec. IV-A and the use of $a_h = a_d$. The set of actions of the MDP is Act.

The transition probabilities of the MDP depend on how compliant the drivers are with the suggestions. For the case that they are fully compliant, the decision making at each step can be replaced by all the possible actions in Act, obtaining an MDP with three deterministic transitions at this level. However, full compliancy at all times is not realistic by any means. To capture all possibilities, we define $\gamma \in [0, 1]$ to be the responsiveness level of a driver to the suggestions given by the ADAS, where values 0 and 1 correspond to fully adamant and fully compliant driver, respectively. Building on the framework in [15] and Sec. IV-B, let p be the probability that a driver decides to change lane at state s, i.e., probability of deciding to continue in the current lane is $1-p$. Furthermore, denote by s'_i, the successor of state s if the vehicle performs action i. Then, the transition probabilities of the MDP from the states $s \in S$ that correspond to ACT-R decision making, i.e., $\mu_s = 2$, are given by:

$$\mathbf{P}(s, \mathrm{a}, s') = \begin{cases} \gamma + (1-\gamma)p & \text{if } \mathrm{a} = \mathrm{a_{lc}} \wedge s' = s'_{\mathrm{lc}}, \\ (1-\gamma)(1-p) & \text{if } \mathrm{a} = \mathrm{a_{lc}} \wedge s' = s'_{\mathrm{con}}, \\ \gamma & \text{if } \mathrm{a} = \mathrm{a_{dec}} \wedge s' = s'_{\mathrm{dec}}, \\ (1-\gamma)p & \text{if } \mathrm{a} = \mathrm{a_{dec}} \wedge s' = s'_{\mathrm{lc}}, \\ (1-\gamma)(1-p) & \text{if } \mathrm{a} = \mathrm{a_{dec}} \wedge s' = s'_{\mathrm{con}}, \\ \gamma + (1-\gamma)(1-p) & \text{if } \mathrm{a} = \mathrm{a_{con}} \wedge s' = s'_{\mathrm{con}}, \\ (1-\gamma)p & \text{if } \mathrm{a} = \mathrm{a_{con}} \wedge s' = s'_{\mathrm{lc}}, \\ 0 & \text{otherwise} \end{cases}$$

Note that, since $\gamma, p \in [0, 1]$, the transitions are guaranteed to sum up to one under each action.

B. Active Control

Active Acceleration Control: Active acceleration control by the ADAS is an incremental addition to the acceleration values applied by the human in the control module, i.e., $a = a_h + a_{\mathrm{ADAS}}$. Let the acceleration of the vehicle bounded by $a \in [a^{\min}, a^{\max}]$. In this module, a value $a_{\mathrm{ADAS}} \in \{a^{\min}_{\mathrm{ADAS}}, \ldots, a^{\max}_{\mathrm{ADAS}}\}$ is considered such that:

$$a^{\min}_{\mathrm{ADAS}} > a^{\min} \quad \text{and} \quad a^{\max}_{\mathrm{ADAS}} < a^{\max}. \tag{8}$$

Hence, the final acceleration applied to the vehicle becomes

$$a = \max(\min(a_h + a_{\mathrm{ADAS}}, a^{\max}), a^{\min}). \tag{9}$$

The restriction to the values of a_{ADAS} presented in (8) allows the system to be incremental instead of enforcing the specific values chosen by the ADAS, i.e., it is corrective instead of assertive.

Active Steering Control: The human driver model in ACT-R uses the control law in (2) for the steering angle ρ for given $k_{\mathrm{far}}, k_{\mathrm{near}}$ and k_I [15]. We design the active steering control of the ADAS (ρ_{ADAS}) to be given by the same control law with different sets of gains. Thus, actions in this part of the assistance system at the model level correspond to different sets of $\mathcal{C}_i = (k^{\mathrm{ADAS}}_{\mathrm{far}}, k^{\mathrm{ADAS}}_{\mathrm{near}}, k^{\mathrm{ADAS}}_I)_i$ available to the ADAS. The resulting steering angle applied to vehicle $\rho = \rho_h + \rho_{\mathrm{ADAS}}$ essentially becomes the control law in (2), where the gains are the sum of gains for the human and ADAS, i.e., incremental (corrective) control, as exemplified in Fig. 3.

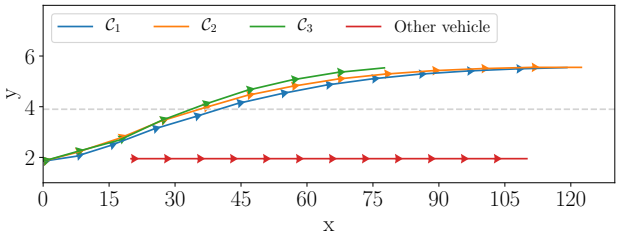


Fig. 3: Example of the simulation of a lane change (x and y in meters) for 3 different sets of parameters $\mathcal{C}_i = (k^{\mathrm{ADAS}}_{\mathrm{far}}, k^{\mathrm{ADAS}}_{\mathrm{near}}, k^{\mathrm{ADAS}}_I)$ for the steering control law, with $\mathcal{C}_1 = (15, 3, 5)$, $\mathcal{C}_2 = (17, 3, 6)$ and $\mathcal{C}_3 = (14.5, 3, 7)$.

We augment the MDP obtained in Sec. V-A by adding the deterministic actions described to the control stage of the model. That is the transition probabilities of the MDP for the states $s \in S$ that correspond to ACT-R control, i.e., $\mu_s = 1$, are given by:

$$\mathbf{P}(s, \mathrm{a}, s') = \begin{cases} 1 & \text{if } s' = \textsc{ActControl}(s, \mathrm{a}), \\ 0 & \text{otherwise}, \end{cases}$$

where ACTCONTROL is the state-action lookup table for the active control step described above.

C. Policy Synthesis

Recall that we are interested in designing an ADAS that optimizes the probability of satisfying a given property φ for a scenario $\mathcal{S}$, i.e, Problem 2. The finite MDP constructed by adding the actions at the decision making and control levels, $\mathcal{M}_{\mathrm{ADAS}}$, represents all the possible choices of the ADAS at every Δt step of the driving scenario $\mathcal{S}$. Therefore, the optimal ADAS problem is reduced to finding an optimal policy over $\mathcal{M}_{\mathrm{ADAS}}$.

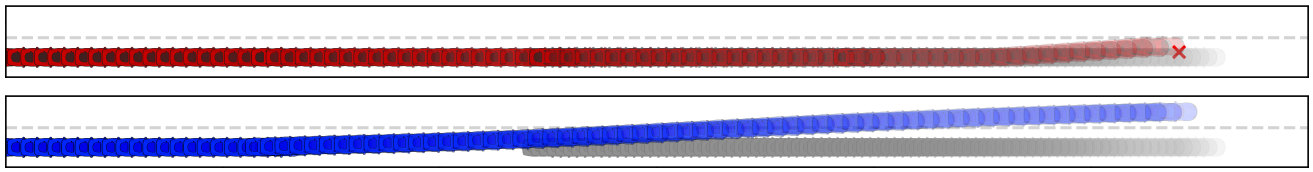
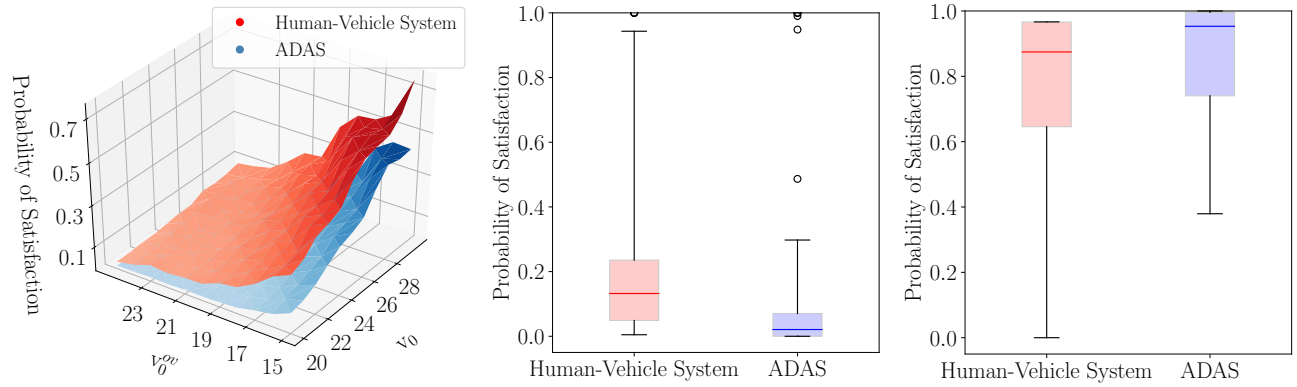


Fig. 4: Example of a run for φ_1 with $\mathcal{S} = (0, 0m, 25\frac{m}{s}, 50m, 15\frac{m}{s})$. Top in red: human-vehicle system (no ADAS). Bottom in blue: human-vehicle system with ADAS. Gray: the other vehicle. For readability purposes, the opacity of the cars decreases with time. The red 'x' marks a collision between the vehicles.



(a) φ_1, $\bowtie = \min$ (b) φ_1, $\bowtie = \min$ (c) φ_2, $\bowtie = \max$

Fig. 5: Analysis of the probability of satisfaction of φ_1 and φ_2 in various conditions. (a) varying scenarios $\mathcal{S} = (0, 0m, v_0, 50m, v_0^{ov})$ with $v_0 \in \{20, ..., 30\}\frac{m}{s}$ and $v_0^{ov} \in \{15, ..., 20\}\frac{m}{s}$; (b) a randomly sampled population of 100 different scenarios S; and (c) a randomly sampled population of 100 different scenarios S and $T = 21s$.

The policy that maximally satisfies φ is defined as:

$$\pi^* \in \arg\sup_{\pi \in \Pi} \mathbb{P}^{\pi}(\varphi) \tag{10}$$

and, respectively, $\arg\inf$ for minimally. The computation algorithms for such policies are well-studied in the formal synthesis literature, and there exist many off-the-shelf tools, e.g., PRISM [21], that solve this optimization problem efficiently. In addition to the optimal policy π^*, these tools compute the probability of satisfying φ under π^*, denoted by $\mathbb{P}^{\pi^*}(\varphi)$.

VI. Experimental Results

The proposed framework is implemented as an open source tool in Python using PRISM[1]. To illustrate its efficacy, we performed a series of case studies using various scenarios and specifications. Due to space constraints, we can show only two of them here. We refer the reader to [29] for the full report on all the case studies.

We considered a two-lane highway scenario with both the ego-vehicle and lead vehicle driving on the right lane on a road segment that is 500 meters long ($x_{\max} = 500m$). The lead vehicle is assumed to be moving at a constant speed with $\mathbf{x}^{ov}(0) = (x_0^{ov}, v_0^{ov})$. The analysis below is performed based on the ACT-R parameters given in [15].

Case Study 1: We are interested in minimizing ($\bowtie = \min$) the safety property of crashing, i.e,

$$\varphi_1 = \Diamond \, \text{CRASH},$$

for initial conditions given by $\mathcal{S} = (\lambda_0, x_0, v_0, x_0^{ov}, v_0^{ov})$. For $\mathcal{S} = (0, 0m, 25\frac{m}{s}, 50m, 15\frac{m}{s})$ with no ADAS intervention, the verification framework generates $\mathbb{P}^S(\varphi_1) = 0.489$. In this situation, the ego-vehicle is travelling at a high speed when compared to the lead vehicle, leaving the human with little room for mistakes. The constraints imposed by the human cognitive modeling, such as memory decay, distraction and limited motor performance, inevitably lead to a high probability of crashing. By adding the ADAS to the ego-vehicle, however, this probability is reduced by more than half to $\mathbb{P}^{S,\pi^*}_{\min}(\varphi_1) = 0.242$, showing the effectiveness of the ADAS. Fig. 4 shows an example run for the human driver model (top), which results in a crash, and ADAS system (bottom), which avoids a crash by suggesting and actively contributing to the lane changing action early on.

Fig. 5a presents the variation of the probability of satisfaction of the safety specification with the change of v_0 and v_0^{ov}. As it can be observed, the introduction of the ADAS reduces the probability of crashing significantly in all the cases. Fig. 5b shows boxplots for the same safety property in a randomly generated sample of 100 different scenarios, obtained by uniformly sampling over bounded intervals for each of the variables. In this case, it is also observed a decrease in the probability of satisfaction of φ_1, with the first, second and third quartiles in Fig. 5b being lower for the system with the ADAS than those for the human driver alone.

[1]Github repository: https://github.com/fgirbal/cbc_adas

Case Study 2: We are interested in maximizing ($\bowtie = \max$) the liveness property of completing the road segment in under T seconds, i.e,

$$\varphi_2 = (\neg \text{CRASH})\, \mathcal{U}\, \big((x = x_{\max}) \wedge (t \leq T)\big),$$

for various sets of initial conditions given by $\mathcal{S}$.

Fig. 5c shows boxplots of the probability of the liveness property for $T = 21s$ in a randomly generated sample of 100 different scenarios $\mathcal{S}$, obtained using the method previously described. In this situation, an increase in the probability of satisfaction of φ_2 is observed, with the first, second and third quartiles in Fig. 5c being higher for the system with the ADAS than those for the human driver alone. This again illustrates the efficacy of the ADAS in terms of the satisfaction of the liveness property, i.e., the ADAS makes the system reach the end of the road safely and faster as required by φ_2.

VII. Final Remarks

In this work, we proposed a framework for providing guarantees in (i) analyses of semi-autonomous driving scenarios and (ii) designing ADAS through the means of formal methods and modeling of the driver's cognitive process. We achieved this by employing ACT-R to represent the human and a novel abstraction method that enables the representation of the infinite, continuous system of human-vehicle by a finite Markov model. In the future, a data driven approach should be followed to validate the obtained results and evaluate the assumptions made about the drivers. A similar perspective can be taken for the design of the specifications, which could be learned in a closed loop fashion to minimize the difference between the full system with the ADAS and expert drivers. This is only possible due to the flexibility of the specifications allowed in the framework. Once the models are accurate according to the real world data, it is possible to deploy the obtained solutions.

References

[1] S. Singh, "Critical reasons for crashes investigated in the national motor vehicle crash causation survey," Tech. Rep., 2015.

[2] N. Kalra and S. M. Paddock, "Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability?" *Transportation Research Part A: Policy and Practice*, vol. 94, pp. 182–193, 2016.

[3] D. Sportillo, A. Paljic, M. Boukhris, P. Fuchs, L. Ojeda, and V. Roussarie, "An immersive virtual reality system for semi-autonomous driving simulation: a comparison between realistic and 6-dof controller-based interaction," in *Proceedings of the 9th International Conference on Computer and Automation Engineering*. ACM, 2017, pp. 6–10.

[4] S. Baltodano, S. Sibi, N. Martelaro, N. Gowda, and W. Ju, "The rrads platform: a real road autonomous driving simulator," in *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, 2015, pp. 281–288.

[5] M. Zhou, X. Qu, and S. Jin, "On the impact of cooperative autonomous vehicles in improving freeway merging: a modified intelligent driver model-based approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 6, pp. 1422–1428, 2017.

[6] A. Gruber, M. Gadringer, H. Schreiber, D. Amschl, W. Bösch, S. Metzner, and H. Pflügl, "Highly scalable radar target simulator for autonomous driving test beds," in *Radar Conference (EURAD), 2017 European*. IEEE, 2017, pp. 147–150.

[7] P. Koopman and M. Wagner, "Challenges in autonomous vehicle testing and validation," *SAE International Journal of Transportation Safety*, vol. 4, no. 1, pp. 15–24, 2016.

[8] J. Somers, "The coming software apocalypse," http://www.theatlantic.com/technology/archive/2017/09/saving-the-world-from-code/540393/, Sep 2017 (accessed August 18, 2018).

[9] T. Chen, M. Kwiatkowska, A. Simaitis, and C. Wiltsche, "Synthesis for multi-objective stochastic games: An application to autonomous urban driving," in *International Conference on Quantitative Evaluation of Systems*. Springer, 2013, pp. 322–337.

[10] P. Nilsson, O. Hussien, Y. Chen, A. Balkan, M. Rungger, A. Ames, J. Grizzle, N. Ozay, H. Peng, and P. Tabuada, "Preliminary results on correct-by-construction control software synthesis for adaptive cruise control," in *Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on*. IEEE, 2014, pp. 816–823.

[11] D. Sadigh, K. Driggs-Campbell, A. Puggelli, W. Li, V. Shia, R. Bajcsy, A. Sangiovanni-Vincentelli, S. S. Sastry, and S. Seshia, "Data-driven probabilistic modeling and verification of human driver behavior," in *2014 AAAI Spring Symposium Series*, 2014.

[12] J. R. Anderson, *The Architecture of Cognition*. Psychology Press, 2013.

[13] J. R. Anderson, M. Matessa, and C. Lebiere, "ACT-R: A theory of higher level cognition and its relation to visual attention," *Human-Computer Interaction*, vol. 12, no. 4, pp. 439–462, 1997.

[14] D. Salvucci, E. Boer, and A. Liu, "Toward an integrated model of driver behavior in cognitive architecture," *Transportation Research Record: Journal of the Transportation Research Board*, no. 1779, pp. 9–16, 2001.

[15] D. D. Salvucci, "Modeling driver behavior in a cognitive architecture," *Human factors*, vol. 48, no. 2, pp. 362–380, 2006.

[16] N. A. Taatgen, C. Lebiere, and J. R. Anderson, "Modeling paradigms in ACT-R," *Cognition and multi-agent interaction: From cognitive modeling to social simulation*, pp. 29–52, 2006.

[17] T. Balke and N. Gilbert, "How do agents make decisions? A survey," *Journal of Artificial Societies and Social Simulation*, vol. 17, no. 4, p. 13, 2014.

[18] D. S. González, J. S. Dibangoye, and C. Laugier, "High-speed highway scene prediction based on driver models learned from demonstrations," in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2016, pp. 149–155.

[19] R. Fuller, "Towards a general theory of driver behaviour," *Accident analysis & prevention*, vol. 37, no. 3, pp. 461–472, 2005.

[20] C. Baier and J.-P. Katoen, *Principles of model checking*. MIT press, 2008.

[21] M. Kwiatkowska, G. Norman, and D. Parker, "PRISM 4.0: Verification of probabilistic real-time systems," in *International conference on computer aided verification*. Springer, 2011, pp. 585–591.

[22] D. D. Salvucci and A. Liu, "The time course of a lane change: Driver control and eye-movement behavior," *Transportation research part F: traffic psychology and behaviour*, vol. 5, no. 2, pp. 123–132, 2002.

[23] B. D. Ziebart, N. Ratliff, G. Gallagher, C. Mertz, K. Peterson, J. A. Bagnell, M. Hebert, A. K. Dey, and S. Srinivasa, "Planning-based prediction for pedestrians," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2009, pp. 3931–3936.

[24] C. Bererton, G. J. Gordon, and S. Thrun, "Auction mechanism design for multi-robot coordination," in *Advances in Neural Information Processing Systems*, 2004, pp. 879–886.

[25] M. Lahijanian, J. Wasniewski, S. B. Andersson, and C. Belta, "Motion planning and control from temporal logic specifications with probabilistic satisfaction guarantees," in *2010 IEEE International Conference on Robotics and Automation*. IEEE, 2010, pp. 3227–3232.

[26] H. C. Manual, "Highway capacity manual," *Washington, DC*, vol. 11, 2000.

[27] D. Kong and X. Guo, "Analysis of vehicle headway distribution on multi-lane freeway considering car–truck interaction," *Advances in Mechanical Engineering*, vol. 8, no. 4, p. 1687814016646673, 2016.

[28] A. Maurya, S. Dey, and S. Das, "Speed and time headway distribution under mixed traffic condition," *Journal of the Eastern Asia Society for Transportation Studies*, vol. 11, pp. 1774–1792, 2015.

[29] F. Eiras, "To err is human: Designing correct-by-construction driver assistance systems using cognitive modelling," Master's Thesis, University of Oxford, 2018.