Encoding and Monitoring Responsibility Sensitive Safety Rules for Automated Vehicles in Signal Temporal Logic

Mohammad Hekmatnejad, Shakiba Yaghoubi, Adel Dokhanchi, Heni Ben Amor, Aviral Shrivastava, Lina Karam, and Georgios Fainekos
{mhekmatn,syaghoub,adokhanc,hbenamor,ashriva6,karam,fainekos}@asu.edu
Arizona State University
Tempe, AZ, USA

ABSTRACT

As Automated Vehicles (AV) get ready to hit the public roads unsupervised, many practical questions still remain open. For example, there is no commonly acceptable formal definition of what safe driving is. A formal definition of safe driving can be utilized in developing the vehicle behaviors as well as in certification and legal cases. Toward that goal, the Responsibility-Sensitive Safety (RSS) model was developed as a first step toward formalizing safe driving behavior upon which the broader AV community can expand. In this paper, we demonstrate that the RSS model can be encoded in Signal Temporal Logic (STL). Moreover, using the S-TaLiRo tools, we present a case study of monitoring RSS requirements on selected traffic scenarios from CommonRoad. We conclude that monitoring RSS rules encoded in STL is efficient even in heavy traffic scenarios. One interesting observation is that for the selected traffic data, vehicle parameters and response times, the RSS model violations are not frequent.

KEYWORDS

Responsibility-Sensitive Safety, Signal-Temporal Logic, Monitoring, Robustness

ACM Reference Format:
Mohammad Hekmatnejad, Shakiba Yaghoubi, Adel Dokhanchi, Heni Ben Amor, Aviral Shrivastava, Lina Karam, and Georgios Fainekos. 2019. Encoding and Monitoring Responsibility Sensitive Safety Rules for Automated Vehicles in Signal Temporal Logic . In *17th ACM-IEEE International Conference on Formal Methods and Models for System Design (MEMOCODE '19), October 9–11, 2019, La Jolla, CA, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3359986.3361203

1 INTRODUCTION

Self-driving or Automated Vehicles (AV) promise to improve transportation efficiency and safety by eliminating human biases and driver errors [22]. Toward that goal, multiple technology and automotive companies have been working on different products to achieve full or partial autonomy. Even though autonomy has the potential to help save lives (directly and/or indirectly), software bugs and errors can negate the potential benefits of this technology. Bugs lurk in the software primarily due to complex interactions between sensors, vehicle dynamics, and software. In fact, the challenge is so critical that some of the recent AV incidents and accidents can be attributed in part to erroneous decisions by the software, e.g., [7, 16]. Similar challenges can be observed with Advanced Driver Assistance Systems (ADAS), e.g., arbitrary emergency braking system activation without an imminent collision [21]. On the other hand, it is well understood that not all accidents can be avoided in mixed-driver environments, e.g., a human driver losing control and colliding with an AV [2].

This raises the question under what assumptions can we establish AV or ADAS safety? Early on, Loos *et al.* [17] established assumptions under which an abstract model of an Adaptive Cruise Control system can be inductively proved safe using a theorem prover for hybrid dynamical systems. Such verified models can also be used for monitoring compliance at runtime [20]. More recently, Shalev-Shwartz *et al.* [27] provided a more extensive study on the formalization of rules for a safe-driver model referred to as Responsibility-Sensitive Safety (RSS) model. Namely, RSS attempts to formalize the interpretation of the "Duty of Care" from Tort law in different driving scenarios. The RSS model goes beyond the highway driving scenario introduced in [17] and includes urban and rural road driving scenarios as well as simplified lateral dynamics (not only longitudinal reasoning in each lane). The basic premise of [27] is that if all the road vehicles drive according to the RSS model, then all the vehicle interactions on the roads will be safe. In other words, it is possible to formalize what behaviors an AV should exhibit in order to never cause an accident. In addition, a formal model like RSS can help us assign blame when an accident happens, i.e., we can identify which vehicle violated the safety rules.

Works like [17, 27] establish provable safety under simplifying assumptions. However, as also highlighted in [24], there is a semantic gap between provable properties on simplified models and what can be proved about the real system or even about a high-fidelity model of the real system. For example, the simplest safety property which can be formulated is longitudinal safety between two cars on the highway. For this example, the assumption is that when the distance between the two cars becomes potentially unsafe, then the rear vehicle is allowed to accelerate up to a response time with an acceleration less than a maximum value and after that, it should slow down with at least a minimum braking deceleration that guarantees safety. If the assumption mentioned above is satisfied, then safety is guaranteed.

However, the question then becomes how do we establish that this assumption is satisfied on the real AV or a high-fidelity model

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
MEMOCODE '19, October 9–11, 2019, La Jolla, CA, USA
© 2019 Association for Computing Machinery.
ACM ISBN 978-1-4503-6997-8/19/10…$15.00
https://doi.org/10.1145/3359986.3361203

of the AV under a range of weather, road, and vehicle conditions? Toward that goal, in this paper, we formalize the RSS model assumptions [27] in Signal Temporal Logic (STL) [5, 18] as a way to encode the requirements for safe vehicle operation in logic. With the RSS assumptions encoded in assume-guarantee logical conditions in STL, a range of verification and testing tools could be employed to verify and validate AV compliance. For instance, [4], which encodes similar safety rules to RSS [27] in STL, recommends using existing STL based tools [5] for testing and monitoring. With an encoding of safety rules in STL, requirements-driven testing frameworks [6, 9, 23, 28] could be used to directly search for invalidating scenarios under probabilistic guarantees. Along the same lines, the authors in [24] categorize and review all the existing hybrid dynamical system tools which could be used to verify the RSS rules with more complex vehicle dynamics.

Contributions: In this paper, we demonstrate that the RSS model can be encoded in assume-guarantee STL requirements. Due to space limitations, we present in detail only two RSS scenarios, but the rest of the scenarios can be similarly encoded in STL. To motivate how the resulting STL requirements could be used in practice, we monitor multiple real driving data scenarios offline over some of the RSS rules written in STL [3]. Interestingly, it is observed that the RSS rules are not frequently violated by human drivers assuming fast reaction times. The code for the case study is distributed with S-TaLiRo [25].

2 PRELIMINARIES

In the following, $\mathbb{N}$ is the set of natural numbers and $\mathbb{R}$ the set of reals. For this work, we assume that there is a set of vehicles $V = \{e, 1, \ldots, m\}$, where e stands for the ego[1] vehicle and 1 to m are the vehicles that the ego vehicle needs to consider in its immediate environment. We view the behavior of each vehicle $v \in V$ as a discrete time signal (also referred to as *trace* in the following) $\sigma^{(v)} : N_v \to \mathcal{Y}$, where $N_v \subseteq \mathbb{N}$ is the domain of the signal and $(\mathcal{Y}, \mathbf{d})$ is a space equipped with a generalized quasi-metric $\mathbf{d} : \mathcal{Y} \times \mathcal{Y} \to \mathcal{V}$ [26]. In this context, $\mathcal{V}$ is a lattice which can be equipped with a complement (negation (-)) operator to define the set of truth values for STL (see [1] for a more detailed discussion). For example, if $\mathcal{V} = \overline{\mathbb{R}}_{\geq 0} \triangleq \mathbb{R}_{\geq 0} \cup \{+\infty\}$, then the truth values of STL would be over the set $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$.

We assume the same constant sampling rate Δt for all the signals $\sigma^{(v)}$ for each vehicle v as well as the same time domain $N_v = N$. That is, the value of the signal at sample i, i.e., $\sigma^{(v)}(i)$, is sampled at time $i\Delta t$. The results in this case study do not depend on the assumption of a constant sampling rate, but this assumption greatly simplifies the notation. Moreover, we drop the superscript (v) from $\sigma^{(v)}$ since the discussion applies equally to all the vehicles under consideration.

Typically, the space $\mathcal{Y}$ will depend on the fidelity of the vehicle model under consideration, e.g., the model of chassis, tires, suspension, powertrain, etc. Even though the RSS model for some scenarios considers rotational dynamics (besides translational), for the scenarios in this case study, we will only consider a 6-dimensional space: position on the plane, forward and lateral velocity, and forward and lateral acceleration, i.e., $\mathcal{Y} = \mathbb{R}^6$. In case there exists other variables characterizing the behavior of the vehicle, e.g., a Boolean signal indicating whether the vehicle is in danger or not, then these can also be included in $\mathcal{Y}$.

[1]The term *ego vehicle* refers to the vehicle which is under consideration or evaluation.

2.1 STL and STL Robustness

Originally, Signal Temporal Logic (STL) [18] was defined to express bounded time requirements over continuous-time (CT) signals. However, the extension over discrete-time (DT) signals, which we introduce here, is straightforward by using the standard Metric Temporal Logic (MTL) semantics as presented in [11].

Definition 2.1 (STL Syntax for DT signals). *Let x be a vector variable, i.e., $x = [x_1, \ldots, x_n]^T$, $p(x)$ be a function over the reals, and I be any non-empty interval of $\mathbb{R}_{\geq 0}$. The syntax for Signal Temporal Logic (STL) formulas is provided by the following grammar:*

$$\phi ::= \mathbf{T} \mid p(x) \geq 0 \mid \neg\phi \mid \phi \vee \phi \mid \bigcirc_I \phi \mid \phi \, \mathcal{U}_I \phi$$

where $\mathbf{T}$ is true, $\bigcirc_I$ is the next sample operator, and $\mathcal{U}_I$ is the until operator.

For this case study, we will be using a quantitative interpretation of the STL semantics (see [5] for an overview). In order to define quantitative semantics with a topological interpretation over arbitrary predicates $p(x) \geq 0$, we will need to use a *generalized quasi-metric* $\mathbf{d}$ [26] (referred to simply as a metric for brevity in the following) to define a signed distance function:

Definition 2.2 (Signed Distance). *Let $x \in X$ be a point, $S \subseteq X$ be a set and $\mathbf{d}$ be a metric. Then, we define the Signed Distance from x to S to be*

$$\mathbf{Dist}_{\mathbf{d}}(x, S) := \begin{cases} -\inf\{\mathbf{d}(x, x') \mid x' \in S\} & \text{if } x \notin S \\ \inf\{\mathbf{d}(x, x') \mid x' \notin S\} & \text{if } x \in S \end{cases}$$

Intuitively, the distance function returns positive values when x is in the set S and negative values when x is outside the set S. The metric $\mathbf{d}$ must be at least a generalized quasi-metric as described in [1] which also includes the case where $\mathbf{d}$ is a metric as it was introduced in [11]. We should point out that we use the extended definition of supremum ($\sqcup$) and infimum ($\sqcap$). That is to say, the supremum of the empty set is defined to be the bottom element of the domain, while the infimum of the empty set is defined to be the top element of the domain. For example, when $\mathcal{V} = \overline{\mathbb{R}}_{\geq 0}$, then $\inf \emptyset := +\infty$.

We review STL semantics that map a formula φ and a trace σ to a value drawn from a lattice $\overline{\mathcal{V}} \triangleq \mathcal{V} \cup \{-v \mid v \in \mathcal{V}\}$. In this work, even though we need an arbitrary lattice of truth values $\mathcal{V}$ to treat Boolean signals, in most of the examples, we assume that $\mathcal{V} = \overline{\mathbb{R}}_{\geq 0}$ (with the usual negation (-) over the reals). We denote the robust valuation of the formula φ over the trace σ at sample i by $[\![\varphi]\!]_{\mathbf{d}}(\sigma, i)$. The semantics for a predicate $p(x) \geq 0$ evaluated at time i over trace σ is defined as the distance between $\sigma(i)$ and the set $[\![p(x) \geq 0]\!] \triangleq \{x \mid p(x) \geq 0\}$. Intuitively, this distance represents how robustly the point $\sigma(i)$ lies within (or is outside) the set $[\![p(x) \geq 0]\!]$. If this distance is zero, then the smallest perturbation of the point $\sigma(i)$ can affect the outcome of $\sigma(i) \in [\![p(x) \geq 0]\!]$.

Definition 2.3 (Discrete-Time Robust Semantics). *Consider an extended generalized quasi-metric space $(\mathcal{Y}, \mathbf{d})$. Let $\sigma : N \to \mathcal{Y}$*

be a trace, then the robust semantics of an STL formula φ with respect to σ at time sample i is defined as:

$$\begin{aligned}
[\![\mathbf{T}]\!]_{\mathbf{d}}(\sigma, i) &:= \bigsqcup \mathcal{V} := \top \\
[\![p(x) \geq 0]\!]_{\mathbf{d}}(\sigma, i) &:= \mathbf{Dist}_{\mathbf{d}}(\sigma(i), [\![p(x) \geq 0]\!]) \\
[\![\neg\varphi_1]\!]_{\mathbf{d}}(\sigma, i) &:= -[\![\varphi_1]\!]_{\mathbf{d}}(\sigma, i) \\
[\![\varphi_1 \vee \varphi_2]\!]_{\mathbf{d}}(\sigma, i) &:= [\![\varphi_1]\!]_{\mathbf{d}}(\sigma, i) \sqcup [\![\varphi_2]\!]_{\mathbf{d}}(\sigma, i)
\end{aligned}$$

$$[\![\bigcirc_{\mathcal{I}} \varphi_1]\!]_{\mathbf{d}}(\sigma, i) :=$$

$$\begin{cases} [\![\varphi_1]\!]_{\mathbf{d}}(\sigma, i+1) & \text{if } i+1 \in N \text{ and } (i+1)\Delta t \in (i\Delta t + \mathcal{I}) \\ -\top \triangleq \bot & \text{otherwise (i.e., the bottom element)} \end{cases}$$

$$[\![\varphi_1 \,\mathcal{U}_{\mathcal{I}}\, \varphi_2]\!]_{\mathbf{d}}(\sigma, i) :=$$

$$\bigsqcup_{i' \in \{j \in N \mid j\Delta t \in (i\Delta t + \mathcal{I})\}} \left([\![\varphi_2]\!]_{\mathbf{d}}(\sigma, i') \sqcap \bigsqcap_{i \leq i'' < i'} [\![\varphi_1]\!]_{\mathbf{d}}(\sigma, i'') \right)$$

where $\top$ is the top element of the lattice, and $t + \mathcal{I} = \{t'' \mid \exists t' \in \mathcal{I} \,.\, t'' = t + t'\}$.

Intuitively, the requirement $\bigcirc_{\mathcal{I}} \varphi$ states that φ should be true at the next sample, which should occur some time in the physical time interval $\mathcal{I}$. For example, consider $\Delta t = 0.1$ and the formula $\psi = \bigcirc_{[0,0.1]} \mathbf{T}$, then ψ is true ($\top$) at sample i since $(i+1)\Delta t - i\Delta t = \Delta t \in [0, 0.1]$. However, for $\Delta t = 0.2$, ψ would evaluate to false ($\bot$). The operator $\varphi_1 \,\mathcal{U}_{\mathcal{I}}\, \varphi_2$ states that φ_2 should be satisfied at some time in the interval $\mathcal{I}$ and until then φ_1 should hold. The other common Boolean and temporal operators can be defined as syntactic abbreviations (see [11, 18]). For example, $\Diamond_{\mathcal{I}} \phi \equiv \mathbf{T}\, \mathcal{U}_{\mathcal{I}}\, \phi$ stands for eventually at some time in the time interval $\mathcal{I}$, ϕ should be true, and $\Box_{\mathcal{I}} \phi \equiv \neg\Diamond_{\mathcal{I}} \neg\phi$ stands for always during the interval $\mathcal{I}$, ϕ should be true. When $\mathcal{I} = [0, \infty)$, we will be dropping $\mathcal{I}$ from the notation, e.g., $\Box_{[0,\infty)}\phi \equiv \Box\phi$.

An important operator that we will be using in this work is the *release* operator $\varphi_1 \mathcal{R}_{\mathcal{I}} \varphi_2 \equiv \neg(\neg\varphi_1 \,\mathcal{U}_{\mathcal{I}}\, \neg\varphi_2)$, which states that φ_2 should always hold during the time interval $\mathcal{I}$ up to (but not including) the time when φ_1 becomes true. In fact, we will need a slightly modified version of the release operator $\overline{\mathcal{R}}$ which does not require φ_2 to happen at all if φ_1 has happened in the past:

$$[\![\varphi_1 \overline{\mathcal{R}}_{\mathcal{I}} \varphi_2]\!]_{\mathbf{d}}(\sigma, i) :=$$

$$\bigsqcap_{i' \in \{j \in N \mid j\Delta t \in (i\Delta t + \mathcal{I})\}} \left([\![\varphi_2]\!]_{\mathbf{d}}(\sigma, i') \sqcup \bigsqcup_{i \leq i'' \leq i'} [\![\varphi_1]\!]_{\mathbf{d}}(\sigma, i'') \right)$$

We refer to the above operator as non-strict release operator. In fact, any non-strict release formula such as $\varphi_1 \overline{\mathcal{R}}_{\mathcal{I}} \varphi_2$ can be rewritten as $\varphi_1 \mathcal{R}_{\mathcal{I}} (\varphi_1 \vee \varphi_2)$ using the release operator.

In the following, we let $(\sigma, i) \models \varphi$ denote the standard Boolean STL satisfiability. Note that Boolean satisfiability reduces to an application of Def. 2.3 wherein the metric $\mathbf{d}$ is the discrete metric. It is easy to show that if the signal satisfies the property, then its robustness is non-negative and, similarly, if the signal does not satisfy the property, then its robustness is non-positive [11]. The robustness $[\![\varphi]\!]_{\mathbf{d}}(\sigma, i)$ can be computed in polynomial time in the size of the formula and the time domain N of σ (see [5] for different computation algorithms).

2.2 Responsibility-Sensitive Safety (RSS)

RSS is a safety modeling paradigm for autonomous driving cars which is based on responsibilities. In structured roads, the premise of RSS is that although a self-driving car might be in a car accident, it never causes an accident. RSS formalizes the following four common-sense rules [27]:

- "Keep a safe distance from the car in front of you, so that if it brakes abruptly you will be able to stop in time"
- "Keep a safe distance from cars on your side, and when performing lateral manoeuvres and cutting into another carâĂŹs trajectory, you must leave the other car enough space to respond"
- "You should respect *right-of-way* rules, but *right-of-way* is given not taken"
- "Be cautious of occluded areas, for example, a little kid might be occluded behind a parked car"

In this case study, we demonstrate the formalization in temporal logic of only the first two rules for monitoring real traffic scenarios, which are taken from the CommonRoad library. We specifically used real traffic scenarios, because autonomous vehicle driving datasets do not usually have ground truth data for the non-autonomous vehicles.

2.3 CommonRoad

CommonRoad is a composable framework for benchmarking motion planning on roads [3]. It provides an efficient format for storing all the necessary driving data in different driving scenarios. It is composed of the scenario data (road network, static and dynamic obstacles and the planning problem for the ego vehicle), ego vehicle model, and cost functions. CommonRoad scenarios are represented in XML files [15], and multiple real and handcrafted scenarios are available online on the CommonRoad website[2]. In this work, some of the real driving scenarios are monitored and tested against RSS specifications.

3 RSS FORMALIZATION PROBLEM

3.1 From Cartesian to Lane-Based Coordinate System

CommonRoad uses a global Cartesian coordinate system for localizing objects in scenarios, while safety requirements in RSS are defined in a lane-based coordinate system. In [27], the authors proposed a transformation from global positions in a Cartesian system to a lane-based coordinate system, which we use in this work (reviewed below).

In their model, the lane's center is represented as a smooth curve r formed as a concatenation of linear or arc-shaped pieces $r^{(1)}, ..., r^{(K)}$. The curve r can be represented as $r : [Y_{min}, Y_{max}] \rightarrow \mathbb{R}^2$ that maps a "longitudinal" position $Y \in [Y_{min}, Y_{max}] \subset \mathbb{R}$ into $(x, y) \in \mathbb{R}^2$, such that x is on the global $x - axis$, and y is on the global $y - axis$. Let $w : [Y_{min}, Y_{max}] \rightarrow \mathbb{R}_+$ be a continuous lane-width function that maps the longitudinal position Y into a positive width value and let $r^{\perp}(Y)$ denotes the normal unit-vector at the position Y. The subset of the points on the plane that fall into the

[2] https://commonroad.in.tum.de

lane is given by [27]:

$$D = \{r(Y) + \alpha w(Y) r^{\perp}(Y) \mid Y \in [Y_{min}, Y_{max}], \alpha \in [-\frac{1}{2}, +\frac{1}{2}]\}$$

The assumption in the above formula is that for any arc-shaped lane-piece $r^{(k)}$ with radius $\mathcal{R}_k$, the width of the lane along $r^{(k)}$ does not exceed $\mathcal{R}_k/2$. Note that what we refer to as a plane here is not a restriction, and all the definitions are extendable to surfaces. For points (x', y') belonging to the subset D, the transformation from the global coordinates (x', y') to the lane coordinates (Y', α') is $\gamma : D \to \mathbb{R}^2$ defined as [27]:

$$\gamma(x', y') = \{(Y', \alpha') \mid Y' \in [Y_{min}, Y_{max}], \alpha' \in [-\frac{1}{2}, +\frac{1}{2}],$$
$$(x', y') = r(Y') + \alpha' w(Y') r^{\perp}(Y')\}.$$

In this case, the term *lane-based coordinate system* is used to refer to $\gamma(D) = [Y_{min}, Y_{max}] \times [-1/2, +1/2]$ [27]. Let $\gamma(x', y')_y$, and $\gamma(x', y')_\alpha$ denote the first coordinate Y' and the second coordinate α'. Where it is needed, we denoted $\gamma(x', y')_x$ to refer to $\alpha' w(Y')$.

For two cars c_1 and c_2 that are on the same lane, it is desirable to define γ such that their logical ordering is preserved:

- If c_1 *is behind* c_2, then $\gamma(x_1, y_1)_y < \gamma(x_2, y_2)_y$, and
- If c_1 *is to the left of* c_2, then $\gamma(x_1, y_1)_\alpha < \gamma(x_2, y_2)_\alpha$.

If we use $w(Y')$-*units* to represent the width of curves in the above coordinate system, then the lateral movement on the curve falls in $\alpha \in [-1/2, +1/2]$. The lateral argument in the lane-based coordinates captures the notion of lateral maneuvers in widening/-narrowing lanes. In these lanes, if the car moves on one of the boundaries, it is not considered to have a lateral movement, even though it gets closer to/farther from the lane center. Longitudinal/lateral velocity/acceleration are the first/second derivatives of the longitudinal/lateral positions.

In CommonRoad the road network is composed of atomic, interconnected drivable road segments called lanelets. Lanelets are defined by their left and right bounds, which are represented by an array of points (a polyline)[3]. As a result, roads in CommonRoad consist of linear pieces $r^{(k)}$, and they are almost everywhere smooth. For points on the intersection of consecutive linear segments, circular approximations can be used to convert from the Cartesian coordinate to the lane-based coordinate system. The directed curve in the middle of the lane can be calculated using the lanelet's left and right boundaries.

3.2 Safe Longitudinal and Lateral Distances

Definition 3.1 (Safe Longitudinal Distance —same direction (Def. 1 in [27])). *A longitudinal distance between a car c_b that drives behind another car c_f, where both cars are driving at the same direction, is safe w.r.t. a response time ρ if for any braking of at most a^{lon}_{maxBr}, performed by c_f, if c_b will accelerate by at most a^{lon}_{maxAcc} during the response time, and from there on will brake by at least a^{lon}_{minBr} until a full stop then it will not collide with c_f.*

The below formula (Lemma 2 in [27]) calculates the safe distance between cars c_b, c_f w.r.t. the parameters in the above definition:

$$d_{min,lon} = max(d_{b,preBr} + d_{b,brake} - d_{f,brake}, 0),$$
$$d_{b,preBr} = v^{lon}_b \rho + \tfrac{1}{2} a^{lon}_{maxAcc} \rho^2,$$
$$d_{b,brake} = \frac{(v^{lon}_b + \rho a^{lon}_{maxAcc})^2}{2a^{lon}_{minBr}}, \quad d_{f,brake} = \frac{v^{lon^2}_f}{2a^{lon}_{maxBr}}$$

In the above equation, the maximum distance that two cars can traverse before they stop without a collision is $d_{b,preBr} + d_{b,brake}$ for the rear car and $d_{f,brake}$ for the front car. Therefore, the minimum needed distance for them not to collide is the total longitudinal movement of the rear car minus the longitudinal movement of the front car for the same period of time. Obviously, if the result became negative, then a collision happened.

Definition 3.2 (Safe Lateral Distance (Def. 6 in [27])). *The lateral distance between cars c_l, c_r driving with lateral velocities v_l, v_r is safe w.r.t. parameters $\rho, a^{lat}_{minBr}, a^{lat}_{maxAcc}$, and μ, if during the time interval $[0, \rho]$ the two cars will apply lateral acceleration of a^{lat}_{maxAcc} toward each other, and after that the two cars will apply lateral braking of a^{lat}_{minBr}, until they reach zero lateral velocity, then the final lateral distance between them will be at least μ.*

For calculation of the safe lateral distance we use the Lemma 4 in [27], by which for two cars c_l, c_r the minimum safe lateral distance between them where c_l is to the left of the c_r is

$$d_{min,lat} = \mu + max(d_{l,preBr} + d_{l,brake} - (d_{r,preBr} - d_{r,brake}), 0),$$
$$d_{l,preBr} = \frac{v^{lat}_l + v^{lat}_{l,\rho}}{2}\rho, \quad d_{l,brake} = \frac{v^{lat^2}_{l,\rho}}{2a^{lat}_{minBr}},$$
$$d_{r,preBr} = \frac{v^{lat}_r + v^{lat}_{r,\rho}}{2}\rho, \quad d_{r,brake} = \frac{v^{lat^2}_{r,\rho}}{2a^{lat}_{minBr}},$$
$$v^{lat}_{l,\rho} = v^{lat}_l + \rho a^{lat}_{maxAcc}, \quad v^{lat}_{r,\rho} = v^{lat}_r - \rho a^{lat}_{maxAcc}$$

In the above equation, the maximum distance that two cars can traverse before they stop without a collision is $d_{l,preBr} + d_{l,brake}$ for the left car, and $d_{r,preBr} - d_{r,brake}$ for the right car. Therefore, the minimum needed distance for them not to collide is the total movement of the left car (toward the positive latitude axis) minus the movement of the right car (toward the positive latitude axis) for the same period of time.

Definition 3.3 (Dangerous Longitudinal Situation and Dangerous Threshold time[3] (Definition 3 in [27])). *Time t is dangerous for cars c_b, c_f if the distance between them at time t is non-safe (according to Definition 1 or Definition 2 in [27]). Given a dangerous time t, its dangerous threshold time, denoted t^{lon}_b, is the earliest longitudinally dangerous time such that all the times in the interval $[t^{lon}_b, t]$ are dangerous. In particular, an accident can only happen at time t if it is dangerous, and in that case, we say that the dangerous threshold time of the accident is the dangerous threshold time of t.*

Next, we refer to what are "proper responses" to the dangerous situations.

[3] Also called Blame Time in earlier versions of [27]

Definition 3.4 (Proper Response to Dangerous Longitudinal Situations (Definition 4 in [27]) —same direction). *Let t be a dangerous time for cars c_b, c_f and let t_b^{lon} be the corresponding dangerous threshold time. The proper behavior of the two cars (and say that c_b is the rear car) is to comply with the following constraints on the longitudinal speed:*

1. *c_b acceleration must be at most a_{maxAcc} during the interval $[t_b^{lon}, t_b^{lon}+\rho)$ and at most $-a_{minBr}$ from time $t_b^{lon}+\rho$ until reaching a safe longitudinal situation. After that, any non-positive acceleration is allowed.*
2. *c_f acceleration must be at least $-a_{maxBr}$ until reaching a safe longitudinal situation. After that, any non-negative acceleration is allowed.*

Lemma 3.1 (STL Translation of Def. 3.4 – Longitudinal Safety Specification). *An STL formula which formalizes Def. 3.4 of proper response to a dangerous longitudinal situation at dangerous threshold time t_b^{lon} is formulated as:*

$$\varphi_{resp}^{lon} \equiv \square\left((S_{b,f}^{lon} \wedge \bigcirc\neg S_{b,f}^{lon}) \rightarrow \bigcirc P^{lon}\right)$$

$$P^{lon} \equiv \big(S_{b,f}^{lon}\overline{\mathcal{R}}_{[0,\rho)}(A_{b,maxAcc}^{lon} \wedge A_{f,maxBr}^{lon}) \wedge S_{b,f}^{lon}\overline{\mathcal{R}}_{[\rho,+\infty)}(A_{b,minBr}^{lon} \wedge A_{f,maxBr}^{lon})\big)$$

where the definitions of the predicates are:

$$S_{b,f}^{lon} \equiv \gamma(x_f,y_f)_y - \gamma(x_b,y_b)_y - d_{min,lon} > 0,$$
$$A_{b,maxAcc}^{lon} \equiv a_b^{lon} \leq a_{maxAcc}^{lon},$$
$$A_{b,minBr}^{lon} \equiv a_b^{lon} \leq -a_{minBr}^{lon},$$
$$A_{f,maxBr}^{lon} \equiv a_f^{lon} \geq -a_{maxBr}^{lon},$$

and (x_b,y_b), and (x_f,y_f) are the Cartesian coordinates of c_b and c_f, respectively.

The above formula is an assume-guarantee requirement in which the antecedent of the implication ($\rightarrow$) becomes true if a longitudinally safe situation changes to the longitudinally dangerous situation. The consequent states what needs to be done immediately after the unsafe situation until the hazardous situation is resolved. The antecedent is a conjunction of two formulas representing a moment (dangerous threshold time) when the longitudinal distance between two cars is safe, but immediately after that, it becomes unsafe. The consequent starts with a next operator for which two formulas must hold. The first formula, $S_{b,f}^{lon}\overline{\mathcal{R}}_{[0,\rho)}(A_{b,maxAcc}^{lon} \wedge A_{f,maxBr}^{lon})$, is a release requirement in $[0,\rho)$. It necessities that the maximum allowed acceleration for the rear vehicle and the maximum allowed deceleration for the front vehicle should be observed up to time ρ or until the distance between the vehicles is safe again. The other formula, $S_{b,f}^{lon}\overline{\mathcal{R}}_{[\rho,+\infty)}(A_{b,minBr}^{lon} \wedge A_{f,maxBr}^{lon})$, is also a release requirement. It necessities that the minimum allowed deceleration for both vehicles should be observed after time ρ up to the time that the distance between the vehicles is safe again.

Definition 3.5 (Dangerous Lateral Situation and Danger Threshold time (Definition 7 in [27])). *Time t is laterally dangerous for cars c_l, c_r if the lateral distance between them at time t is non-safe (according to Definition 6 in [27]). Given a laterally dangerous time t, its Lateral Danger Threshold time, denoted t_b^{lat}, is the earliest laterally dangerous time such that all the times in the interval $[t_b^{lat}, t]$ are laterally dangerous. In particular, an accident can only happen at time t if it is laterally dangerous, and in that case, we say that the laterally threshold time of the accident is the laterally Danger Threshold time of t.*

Definition 3.6 (Lateral Proper Response (Definition 8 in [27])). *Let t be a laterally dangerous time for cars c_l, c_r, let t_b^{lat} be the corresponding laterally Danger Threshold time, and w.l.o.g. assume that at that time c_l was to the left of c_r. The laterally proper response of the two cars is to comply with the following constraints on the lateral speed:*

1. *If $t \in [t_b^{lat}, t_b^{lat}+\rho)$ then both cars can do any lateral action as long as their lateral acceleration, a, satisfies $|a| \leq a_{maxAcc}^{lat}$.*
2. *Else, if $t \geq t_b^{lat}+\rho$:*
   - *Before reaching μ-lateral-velocity of 0 (see Definition 5 in [27]), c_l must apply lateral acceleration of at most $-a_{minBr}^{lat}$ and c_r must apply lateral acceleration of at least a_{minBr}^{lat},*
   - *After reaching μ-lateral-velocity of 0, c_l can have any non-positive μ-lateral-velocity and c_r can have any non-negative μ-lateral-velocity.*

Lemma 3.2 (STL Translation of Definition 8 in [27] – Lateral Safety Specification). *An STL formula which formalizes Def. 3.6 of proper response to a dangerous lateral situation at dangerous threshold time t_b^{lat} is:*

$$\varphi_{resp}^{lat} \equiv \square\left((S_{l,r}^{lat} \wedge \bigcirc\neg S_{l,r}^{lat}) \rightarrow \bigcirc P^{lat}\right)$$

where P^{lat} is defined as

$$P^{lat} \equiv \left(P_{0,\rho}^{lat} \wedge P_{\rho,\infty}^{lat,1} \wedge P_{\rho,\infty}^{lat,2}\right)$$

and the subformulas $P_{0,\rho}^{lat}$, $P_{\rho,\infty}^{lat,1}$ and $P_{\rho,\infty}^{lat,2}$ are defined as

$$P_{0,\rho}^{lat} \equiv S_{l,r}^{lat}\overline{\mathcal{R}}_{[0,\rho)}(A_{l,maxAcc}^{lat} \wedge A_{r,maxAcc}^{lat})$$
$$P_{\rho,\infty}^{lat,1} \equiv \big(S_{l,r}^{lat} \vee V_{l,stop}^{lat}\big)\overline{\mathcal{R}}_{[\rho,+\infty)}A_{l,minBr}^{lat} \wedge \big(S_{l,r}^{lat} \vee V_{r,stop}^{lat}\big)\overline{\mathcal{R}}_{[\rho,+\infty)}A_{r,minBr}^{lat}$$
$$P_{\rho,\infty}^{lat,2} \equiv S_{l,r}^{lat}\overline{\mathcal{R}}_{[\rho,+\infty)}\big(V_{l,stop}^{lat} \rightarrow \bigcirc\square(V_{l,neg}^{lat})\big) \wedge S_{l,r}^{lat}\overline{\mathcal{R}}_{[\rho,+\infty)}\big(V_{r,stop}^{lat} \rightarrow \bigcirc\square(V_{r,pos}^{lat})\big)$$

and the definitions of the predicates are:

$$S_{l,r}^{lat} \equiv \gamma(x_r,y_r)_\alpha - \gamma(x_l,y_l)_\alpha - d_{min,lat} > 0,$$
$$A_{l,maxAcc}^{lat} \equiv |a_l^{lat}| \leq a_{maxAcc}^{lat},\ A_{r,maxAcc}^{lat} \equiv |a_r^{lat}| \leq a_{maxAcc}^{lat},$$
$$A_{l,minBr}^{lat} \equiv a_l^{lat} \leq -a_{minBr}^{lat},\ A_{r,minBr}^{lat} \equiv a_r^{lat} \geq a_{minBr}^{lat},$$
$$V_{l,stop}^{lat} \equiv v_l^{\mu-lat} = 0,\ V_{r,stop}^{lat} \equiv v_r^{\mu-lat} = 0,$$
$$V_{l,neg}^{lat} \equiv v_l^{\mu-lat} \leq 0,\ V_{r,pos}^{lat} \equiv v_r^{\mu-lat} \geq 0,$$

and (x_l,y_l) and (x_r,y_r) are the Cartesian coordinates of c_l and c_r, respectively.

The above formula is an assume-guarantee requirement, in which the antecedent of the implication ($\rightarrow$) becomes true if a laterally safe situation changes to the laterally dangerous situation. The consequent is what should be done immediately until the hazardous situation becomes safe again. The antecedent is a conjunction of two formulas representing a moment (dangerous threshold time) that the lateral distance between two cars is safe, but immediately after that, it becomes unsafe. The consequent starts with a next operator for which three formulas must hold. The first formula $P^{lat}_{0,\rho}$ is a release requirement, which states that in $[0, \rho)$ the maximum allowed acceleration should be respected by both cars immediately up to the time ρ or the time that the distance between the vehicles is safe again. The second formula $P^{lat,1}_{\rho,\infty}$ is a conjunction of two release requirement, which states that from time ρ, the minimum allowed deceleration for both vehicles should be observed up to the time that the distance between the vehicles is safe again or vehicles μ-lateral-velocity becomes zero. We separated release formulas for each vehicle because we want both cars to apply their brakes during a dangerous situation even if one car reached zero μ-lateral-velocity. Also, any vehicle that reaches zero μ-lateral-velocity has to satisfy one of the release formulas in the third formula. In the third formula $P^{lat,2}_{\rho,\infty}$, each of the release formulas has an implication sub-formula. The implications require that, at the next time, the left vehicle attains a non-positive μ-lateral-velocity and the right vehicle attains a non-negative velocity until they reach a safe lateral distance.

Definition 3.7 (Dangerous Situation and Dangerous Threshold time (Definition 9 in [27])). *Time t is dangerous for cars c_1, c_2 if it is both longitudinally and laterally dangerous (according to Definition 3 and Definition 7 in [27]). Given a dangerous time t, its dangerous threshold time, denoted t_b, is $max\{t^{lon}_b, t^{lat}_b\}$ where t^{lon}_b, and t^{lat}_b are the longitudinal and lateral dangerous threshold times, respectively. In particular, an accident can only happen at time t if it is dangerous, and in that case the dangerous threshold time of the accident is the dangerous threshold time of t.*

Definition 3.8 (Basic Proper Response to Dangerous Situations (Definition 10 in [27])). *Let t be a dangerous time for cars c_1, c_2 and let t_b, t^{lon}_b, t^{lat}_b be the corresponding dangerous threshold time, longitudinal dangerous threshold time, and lateral dangerous threshold time, respectively. The basic proper response of the two cars is to comply with the following constraints on the lateral/longitudinal speed:*

- *If $t_b = t^{lon}_b$, then the longitudinal speed is constrained according to Def. 3.4.*
- *If $t_b = t^{lat}_b$, then the lateral speed is constrained according to Def. 3.6.*

Lemma 3.3 (STL Translation of Def. 3.8 for Monitoring—Basic Proper Response Specification). *Longitudinal and lateral safety requirement for an ego vehicle is formulated as*

$$\varphi^{lat,lon}_{resp} \equiv \varphi^{lon} \wedge \varphi^{lat} \wedge \varphi^{lat,lon}$$

$$\varphi^{lon} \equiv \square\Big(\big(\neg S^{lat}_{l,r} \wedge S^{lon}_{b,f} \wedge \bigcirc(\neg S^{lat}_{l,r} \wedge \neg S^{lon}_{b,f})\big) \rightarrow \bigcirc P^{lon}_{lat}\Big)$$

$$\varphi^{lat} \equiv \square\Big(\big(\neg S^{lon}_{b,f} \wedge S^{lat}_{l,r} \wedge \bigcirc(\neg S^{lat}_{l,r} \wedge \neg S^{lon}_{b,f})\big) \rightarrow \bigcirc P^{lat}_{lon}\Big)$$

$$\varphi^{lat,lon} \equiv$$

$$\square\Big(\big(S^{lat}_{l,r} \wedge S^{lon}_{b,f} \wedge \bigcirc(\neg S^{lat}_{l,r} \wedge \neg S^{lon}_{b,f})\big) \rightarrow \bigcirc(P^{lat}_{lon} \vee P^{lon}_{lat})\Big)$$

where P^{lat}_{lon} and P^{lon}_{lat} are modified versions of P^{lat} and P^{lon} where the propositions $S^{lat}_{l,r}$ and $S^{lon}_{b,f}$ are replaced with the formula $(S^{lat}_{l,r} \vee S^{lon}_{b,f})$.

Some important remarks are in order for the noticeable differences between Def. 3.8 and Lemma 3.3.

Remark 3.1. *At a first reading, Lemma 3.3 may appear to have more conditions than the original definition Def. 3.8. Lemma 3.3 determinizes the conditions and respective proper responses of Def. 3.8. Under a non-deterministic model of computation, if both conditions in Def. 3.8 became true at the same time, then the consequents would be checked for satisfaction non-deterministically and no additional case is required. Under a deterministic model of computation, if both conditions in Def. 3.8 became true, then we need to explicitly check for satisfaction of the disjunction of the consequents.*

It is important to note that in continuous time, the event $t_b = t^{lon}_b = t^{lat}_b$ is a measure zero event, so in practice such a case would never be observed in real driving scenarios. However, any implementation that monitors the requirement $\varphi^{lat,lon}_{resp}$ would use a digital clock (sampler), and, hence, the satisfaction of both conditions in Def. 3.8 is a possible event. In fact, in our case study, we observed many cases where both conditions were activated.

Remark 3.2. *Irrespective of whether a deterministic or a non-deterministic model of computation is used, Def. 3.8 states that the proper response for either violation is defined by the proper response of Def. 3.4 (Lemma 3.1) or of Def. 3.6 (Lemma 3.2). However, Def. 3.4 (and, hence, Lemma 3.1) considers only longitudinal safety and does not explicitly consider the possibility that the vehicle might later become laterally safe. Similarly, Def. 3.6 (Lemma 3.2) only considers lateral safety. The updated formulas P^{lat}_{lon} and P^{lon}_{lat} in φ^{lat}, φ^{lon} and $\varphi^{lat,lon}$ (as opposed to the original formulas P^{lat} and P^{lon} in φ^{lat}_{resp} and φ^{lon}_{resp}, respectively) are needed, because, now, we consider both longitudinal and lateral safety at the same time. For example, if the rear vehicle becomes longitudinally unsafe after being laterally unsafe, then in the future it may become safe again by either achieving lateral or longitudinal safety. The updated formulas P^{lat}_{lon} and P^{lon}_{lat} capture this possibility by releasing the safety requirements when $(S^{lat}_{l,r} \vee S^{lon}_{b,f})$ becomes true.*

Remark 3.3. *Definition 3.8 and its STL translation in Lemma 3.3 do not cover the cases in which both vehicles start from a longitudinally and laterally dangerous situation. The below is a revision of the Lemma 3.3 with a new conjunction formula $\varphi^{\neg lat,\neg lon}$ that*

completes the translation.

$$\varphi_{resp}^{lat,lon} \equiv \varphi^{lon} \wedge \varphi^{lat} \wedge \varphi^{lat,lon} \wedge \varphi^{\neg lat,\neg lon}$$
$$\varphi^{\neg lat,\neg lon} \equiv (\neg S_{l,r}^{lat} \wedge \neg S_{r,f}^{lon}) \rightarrow \bigcirc(P_{lon}^{lat} \vee P_{lat}^{lon})$$

In Section 5, we will use the above specification rule, i.e., Lemma 3.3 and Remark 3.3, for monitoring real scenario traffic with respect to the proper response requirement.

The STL formula P_{lon}^{lat} is the conjunction of $P_{0,\rho}^{lat}$, $P_{\rho,\infty}^{lat,1}$, and $P_{\rho,\infty}^{lat,2}$ formulas, in each, the main operators are non-strict release operators. In order for them to be satisfiable, the right-hand side subformula must be satisfiable all the time, or up to the time that the left-hand side becomes satisfiable. The left-hand side of all the formulas is a disjunction of lateral and longitudinal safety requirements, except for $P_{\rho,\infty}^{lat,1}$. In this formula, the disjunctive formula on the left-hand side of the first and second release operators have other propositions $V_{l,stop}^{lat}$ and $V_{r,stop}^{lat}$, respectively. In the case that an ego car does not perform any lateral maneuver or does the longitudinal maneuver, then the μ-lateral-velocity is zero which then stops the monitoring process. In other words, there are no more requirements concerning the cause of the RSS violation.

Remark 3.4. *RSS rules do not define control logic, but rather requirements for safety. As long as the RSS requirements are not violated, then any controller could be designed.*

RSS rules are not explicitly designed as controller rules. There is no emergency action as a command, such as braking, accelerating, or steering. All the predicates in the rules are referring to sensor measurements, not activation commands. Therefore, these rules are used for monitoring the closed loop – with controller – behavior of traffic participants rather than giving them commands or controlling them. It is important to note that still one can use the monitor while there is a controller in place. For example, the robustness output values of the monitor can be fed as inputs to the learning component of a driving controller. The trivial goal is to penalize the controller appropriately where the robustness values are negative and vice versa for the positive values.

4 RSS ROBUSTNESS THROUGH STL

In this section, we highlight the connection between the notion of STL robustness in monitoring a vehicle's behavior with respect to the surrounding traffic and the RSS model. The RSS rules require pairwise monitoring of the interactions between the ego vehicle and the vehicles in its surrounding environment. Given a pair of vehicles behaviors $\sigma^{(e)}$ and $\sigma^{(v)}$, then the STL formulas introduced in Sec. 3 can be evaluated for satisfaction over the combined vector $[(\sigma^{(e)})^T\ (\sigma^{(v)})^T]^T$. Since the STL formulas are interpreted using the robust semantics and there is a direct mapping between STL formulas and RSS rules, the robustness computed for the STL formulas implies a robustness interpretation for the RSS rules as well. Namely, a positive value for the STL formula implies satisfaction of the RSS rules, while a negative value implies violation of the RSS rules. The greater the value of the robustness, the less likely is to violate the RSS requirements and have an accident. Moreover, the robustness value for an STL specification translates to physical robustness values in terms of accelerations, velocities, or distances of the two vehicles depending on the specification. In other words, after evaluating the specification robustness, we can know exactly why a specification was satisfied or violated, e.g., an appropriate braking force was not applied. Besides, we can infer when a vehicle behavior came close to a violation, e.g., the distance between two vehicles was almost zero.

5 S-TALIRO APPLICATION ON REAL FREEWAY TRAFFIC DATA

By formalizing the RSS requirements as STL formulas, we can monitor the behavior of self-driving cars and measure their compliance against the RSS specifications. We used the semantics in Def. 2.3 to compute the robustness of some real traffic scenarios in CommonRoad. Specifically, we only considered highway traffic scenarios which by design do not include intersections and wherein all the self-driving cars are moving in the same direction (from west to east). Note that our approach is not restricted to the presented scenarios, and it applies to more general traffic scenarios that are discussed in [27]. We used DP-TaLiRo [12, 25] as our offline monitoring tool for computing the robustness of the traffic scenarios.

We selected a scenario from *CommonRoad (2017a)* in which there are six lanes (one of which was empty, so we omitted it) beginning from west and stretching to east. There are 43 vehicles that all move in the same direction as the lanes. All the vehicles except two of them, start and end at the same lane during the monitoring. The traffic scenario and the trajectories of the vehicles are shown in Figure 1 (some trajectories are omitted to reduce traffic congestion). Vehicles that are identified as *obs-775* and *obs-756* changed their lane some time later in their trajectories (e.g. see the trajectory of *obs-775* in Fig. 2). We chose vehicles *obs-775* and *obs-779* along with their trajectories for the case-study presented in this section. In the remainder of this section, *obs-775* and *obs-779* are referred to as the front car and the rear car, respectively. We only focus here on monitoring longitudinal safety aspects of the trajectories according to Def. 3.4 as the lateral safety aspects can be presented similarly.

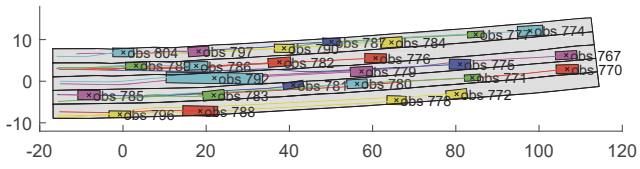

Figure 1: A traffic scenario taken from CommonRoad.

In the following discussion, the time units are in seconds, and the spatial coordinates are in meters. For the rear car, the trajectory start time is $t_s = 1.3$ and end time is $t_f = 8$, and for the front car they are $t_s = 0$ and $t_f = 8$. Each car's trajectory is sampled by delta-time $\Delta t = 0.1$. The start position of the rear car is $(x_b = -16.33, y_b = -0.59)$ and for the front car, it is $(x_f = -5.59, y_f = 2.71)$. The rear car started from lane 3, and the front car started from lane 2. In this longitudinal safety monitoring, we observed the world from the ego car's point of view that is the rear car. For the ego car, at each time, we monitor a car that is moving in the ego car lane if it is longitudinally the nearest car in front of the ego car. In our example, in the beginning, car *obs-775* is driving in lane 2, which

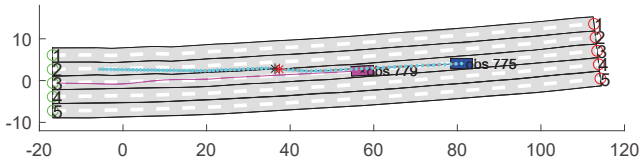

Figure 2: Two trajectories from which one vehicle (dashed-line and colored cyan) changed its lane from 2 to 3 and crossed the border lines (marked the exit and entry of the border crossing by asterisks). The trajectory of *obs-779* is the one corresponding to the ego vehicle in the case-study, and the trajectory of *obs-775* in magenta is for the front car in the case-study. The lanes are tagged with numbers from 1 to 5, and their center line is represented as white-dashed lines.

Table 1: The parameter values in the case-study and experiments.

parameter	value	parameter	value
ρ	0.5 sec	μ	0.4 m
δ_t	0.1 sec	a^{lon}_{minBr}	4 m/s^2
a^{lon}_{maxAcc}	5.5 m/s^2	a^{lat}_{minBr}	3 m/s^2
a^{lon}_{maxBr}	10 m/s^2	a^{lat}_{maxAcc}	3 m/s^2

excludes it from being monitored for evaluating the longitudinal safety of ego car *obs-779*. The same applies if we consider *obs-775* as the ego car. However, after 4.5 seconds of driving along its trajectory, the front car *obs-775* changes its lane to 3 in front of the car *obs-779*. At this time, the position of the rear car *obs-779* is ($x_b = 16.86, y_b = 0.25$), and the position of the front car *obs-775* is ($x_f = 36.36, y_f = 2.82$). We use $\gamma(x', y')_x$ to refer to the lateral distance of position (x', y') from the center-line of its lane. In this work, the above lateral distance definition and $\gamma(x', y')_\alpha$ are interchangeable as long as the width of the lanes are equal for all the trajectories in a scenario. At time $t = 4.5$, the longitudinal distance of the rear car from the start of lane 3 is 34.72 meters, and for the front car, it is 54.37 meters. The new longitudinal distances qualify the rear car (shorter longitudinal distance) as the ego car; therefore, in this instance, if the antecedent of φ^{lon} in Lemma 3.3 is satisfiable, then the proper response P^{lon}_{lat} is monitored. We calculated the speed and acceleration of cars with a sampling interval of $\Delta t = 0.1$ second.

5.1 Analysis of Unsafe Trajectories

The STL formula φ^{lon} in Lemma 3.3 is used for monitoring the longitudinal safety in this case-study. For this work, we computed the safe lateral distances for the monitored cars based on Def. 3.2 (Def. 6 in [27]). For longitudinal safety monitoring, we built multiple signals over time to represent safe longitudinal distances, safe lateral distances, rear car accelerations, and front car accelerations. These constitute the information we needed for monitoring longitudinal safety of both cars in our case-study based on Def. 3.7. For computing the minimum safe lateral/longitudinal distances, we used the parameters in Table 1.

We refer to two cars' lateral/longitudinal distances subtracted from their required safe lateral/longitudinal distances as lateral/-longitudinal robustness safety. As illustrated in Fig. 3, the lateral robustness safety between the two cars are negative through the whole monitoring duration in our example. Similarly, their longitudinal robustness safety are negative at all times except for two short intervals $[5.4, 5.4]$ and $[5.7, 5.8]$ in which they are positive. This caused the antecedent of the implication for the longitudinal safety requirement φ^{lon} in Lemma 3.3 to become active two times as it can be observed in the top diagram in Fig. 4. Next, the signals need to be monitored for the consequent of the implication, which is $\bigcirc P^{lon}_{lat}$ in Lemma 3.3, but in this example, we used $\bigcirc P^{lon}$ in Lemma 3.1 as all the lateral robustness safety values are negative in this case study.

Unsafe Behavior: The second and third diagrams in Fig. 4 represent the evaluation of the two release subformulas in P^{lon}. The first subformula shown in the second subplot of Fig. 4 was true for-all-times, but the second one toggled between true and false three times as represented in the third subplot of Fig. 4. At the first dangerous time $t = 5.4$, both subformulas are evaluated to true until $t = 5.7$ for which the longitudinal distance became safe; therefore, the situation became safe. However, for the second dangerous time $t = 5.8$, the second release and both longitudinal and lateral robustness safeties are false during the interval $[5.9, 7.5)$; therefore, the whole formula is false as can be observed in the last diagram in Fig. 4. The acceleration chart in Fig. 3 represented the cars' accelerations during the time, in which the ego car never applied the minimum required brake deceleration 4 m/s^2 from $t = 6.3$ to the end.

5.2 Analysis of Robustness

Besides computing the magnitude of the robustness of the undergoing monitoring example, we are interested in determining which predicate's violation dominated the magnitude of the robustness. Please note that here robustness values are concrete and we consider their unit (m, m/s, or m/s^2) abstract for this case-study when we compare the robustness values of predicates in each proposition. However, our monitoring tool can report the predicate whose robustness magnitude calculates the final robustness, and therefore, we can deduce its unit. The above implicitly implies that some measuring units could be more influential in determining the magnitude of the robustness.

Remark 5.1. *Under the assumption that the monitored trajectory signals are sampled at the same frequency, all the used predicates in the STL specification in Lemma 3.3 can be transformed into distance-based inequalities. Therefore, the semantics of the STL robustness computes a uniform distance-based magnitude.*

Case-study's Robustness: Assume that σ represents our input signals for this case-study as introduced in Section 2.1. The robustness of the signals against the longitudinal safety specification φ^{lon} is denoted as $[\![\varphi^{lon}]\!]_{\mathbf{d}}(\sigma, 0) = -7.00$. The negative sign of the result states that $(\sigma, 0) \not\models \varphi^{lon}$. Using DP-TaLiRo for the above example, it returns a piece of auxiliary information in addition to the magnitude of the robustness. Two data items from the auxiliary information are the predicate and the time that resulted in the

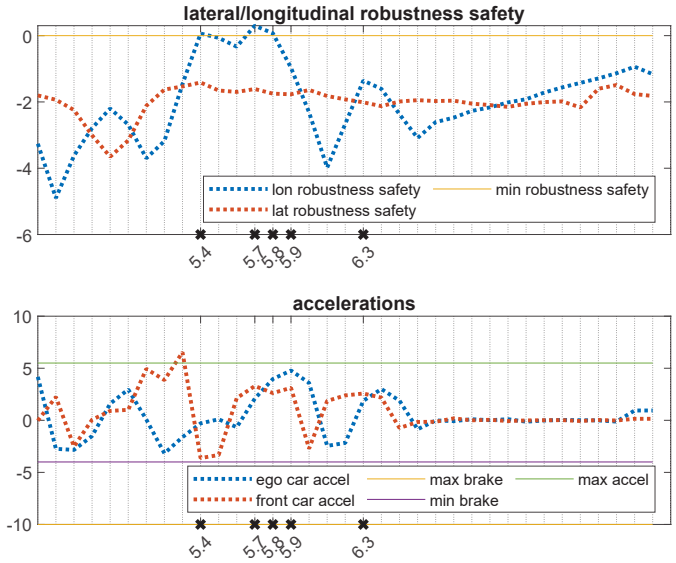

Figure 3: Top: the lateral and longitudinal robustness safeties are illustrated and compared with minimum robustness safety zero. Bottom: longitudinal accelerations of the ego car and the front car are depicted and compared with the minimum and maximum required brake accelerations, and the maximum safe accelerations.

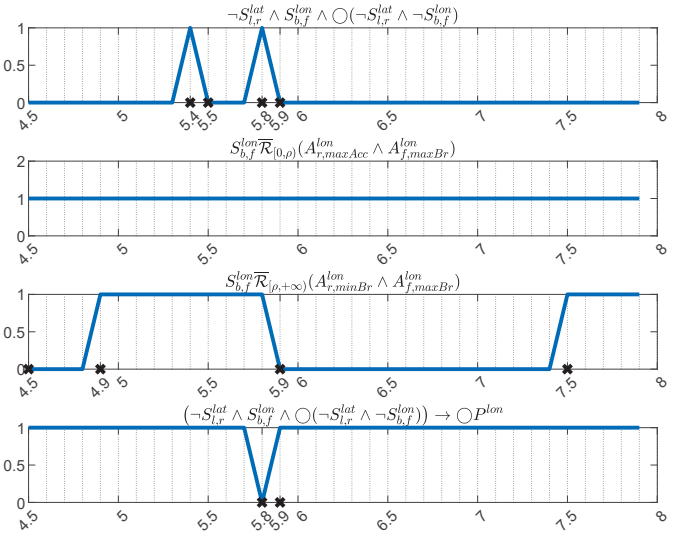

Figure 4: A signal as in Figure 3 is checked against the longitudinal safe response formula φ^{lon} in Lemma 3.3 (without the global operator). Each segment of the above figure represents the truth value for a sub-formula in φ^{lon}. The topmost chart detected two dangerous threshold times: first one responses safely, and the second fails to response safely.

computed robustness. In our case, the predicate is $A^{lon}_{r,minBr}$ and the time is $t = 6.4$ from the beginning of the monitoring.

Table 2: Top-left, bottom-left, top-right: the first columns in these tables represent the predicates in the specification formula φ^{lon}, φ^{lat}, and $\varphi^{lat,lon}$, respectively. The other two columns represent the number of violations that each predicate caused in each subformula. Subformula φ^{lon} is defined in Lemma 3.1, and subformula φ^{lat} is defined in Lemma 3.2. Subformulas φ^{lon}_{lat}, φ^{lat}_{lon}, and $\varphi^{lat,lon}$ are defined in Lemma 3.3. Formula $\overline{\varphi}^{lat,lon}$ is the same as $\varphi^{lat,lon}$ except that we replaced P^{lat}_{lon} and P^{lon}_{lat} in it with P^{lat} and P^{lon}, respectively. Bottom-right: the first row represents the total number of violations that happened in all the scenarios; The second row represents the violation percentages.

predicates	φ^{lon}	φ^{lon}_{lat}
$S^{lon}_{b,f}$	2	2
$S^{lat}_{l,r}$	1	0
$A^{lon}_{b,maxAcc}$	18	18
$A^{lon}_{b,minBr}$	190	184
$A^{lon}_{f,maxBr}$	9	9
predicates	φ^{lat}	φ^{lat}_{lon}
$S^{lon}_{b,f}$	0	0
$S^{lat}_{l,r}$	9	8
$A^{lat}_{l,maxAcc}$	188	186
$A^{lat}_{l,minBr}$	0	0
$A^{lat}_{r,maxAcc}$	256	256
$A^{lat}_{r,minBr}$	0	0
$V^{lat}_{l,stop}$	39	36
$V^{lat}_{r,stop}$	0	0
$V^{lat}_{l,neg}$	0	0
$V^{lat}_{r,pos}$	0	0

predicates	$\overline{\varphi}^{lat,lon}$	$\varphi^{lat,lon}$
$S^{lon}_{b,f}$	0	0
$S^{lat}_{l,r}$	0	0
$A^{lat}_{l,maxAcc}$	0	0
$A^{lat}_{l,minBr}$	0	0
$A^{lat}_{r,maxAcc}$	5	3
$A^{lat}_{r,minBr}$	0	0
$V^{lat}_{l,stop}$	0	0
$V^{lat}_{r,stop}$	0	0
$V^{lat}_{l,neg}$	0	0
$V^{lat}_{r,pos}$	0	0
$A^{lon}_{b,maxAcc}$	0	0
$A^{lon}_{b,minBr}$	4	0
$A^{lon}_{f,maxBr}$	1	1
Execution Statistics (Lemma 3.3)		
# of violation	722	703
violation %	5.9%	5.74%

6 DISCUSSION AND FUTURE WORK

We ran a set of experiments on 29 real traffic scenarios taken from CommonRoad (2017a) (including all the highway scenarios), and the results are presented in Tables 2-3. As indicated in the first row of Table 4, the performance of the monitoring algorithm of S-TaLiRo for thousands of monitoring cases is still feasible. Some statistics about the experiments are summarized in Table 4. In these experiments, we used a Windows 10 machine with Intel Core i7 CPU 8550U @ 1.8GHZ, 16GB RAM, Matlab R2018b, and DP-TaLiRo.

In the following, we are interested only in the cases in which the safety rules became violated. Therefore, we do not distinguish between cases where the antecedent of the rule was never activated, and cases which the antecedent was activated, but then the consequent became satisfied. However, we remark that such finer categorization is possible using the interfaces proposed in [13]. In our experiments (Tables 2-3), we used two versions of the RSS specifications in Def. 3.8. First, in relation to Remark 3.2, we modified the formulas in Lemma 3.3 by replacing P^{lat}_{lon} and P^{lon}_{lat} with P^{lat} and P^{lon}, respectively. This modification changed the consequents of the formulas to be the same as what is stated in Lemma 3.1 and

Table 3: Top: simillar to Table 2, formula $\overline{\varphi}^{\neg lat,\neg lon}$ is the same as $\varphi^{\neg lat,\neg lon}$ except that we replaced P^{lat}_{lon} and P^{lon}_{lat} with P^{lat} and P^{lon}, respectively. Bottom: the first row represents the total number of violations that happened in all the scenarios; the second row represents the violation percentages.

predicates	$\overline{\varphi}^{\neg lat,\neg lon}$	$\varphi^{\neg lat,\neg lon}$
$S^{lon}_{b,f}$	0	0
$S^{lat}_{l,r}$	0	0
$A^{lat}_{l,maxAcc}$	172	166
$A^{lat}_{l,minBr}$	0	0
$A^{lat}_{r,maxAcc}$	177	161
$A^{lat}_{r,minBr}$	0	0
$V^{lat}_{l,stop}$	420	350
$V^{lat}_{r,stop}$	0	1
$V^{lat}_{l,neg}$	0	0
$V^{lat}_{r,pos}$	0	0
$A^{lon}_{b,maxAcc}$	6	7
$A^{lon}_{b,minBr}$	5	3
$A^{lon}_{f,maxBr}$	0	1
Execution Statistics (Remark 3.3)		
# of violation	780	689
violation %	6.37%	5.63%

Table 4: Some statistics on our experiments.

item	value
average runtime per monitor execution	21 *ms*
average number of cars in each scenario	48
average number of surrounding cars to be monitored	8.8
average length of trajectories per car	6.8 *s*

Lemma 3.2. These consequents of the rules are actively monitored when the distance of the ego car from a secondary car becomes unsafe both laterally and longitudinally. On the other hand, part of the conditions to dismiss a dangerous longitudinal/lateral situation is if the longitudinal/lateral distance becomes safe again. As it is stated in Remark 3.2, the above response is too conservative in real scenarios for which if either of the distances became safe before an accident happens, then the situation is not dangerous anymore. Second, we ran our experiments with the formulas in Lemma 3.3, and the results are illustrated in the third columns of the Tables 2-3. The new result confirms that by using the disjunction of lateral and longitudinal safe distances to stop monitoring a dangerous situation, the number of violations reduces in general (due to the relaxed conditions).

Without considering Remark 3.3, about 5.9% of monitored trajectories showed unsafe behaviors in the evaluated scenarios. As it is shown in Table 3, by considering the rules in Remark 3.3, the number of violations increased by a factor of 2.1, and the percentage of unsafe behaviors increased to 12.3% (calculated by adding 6.37% as the newly violation percentage to the former 5.9% violation percentage in Table 2). Among the conditions that contributed to an unsafe behavior, the minimum longitudinal brake and maximum lateral acceleration were the most frequently violated ones (i.e., see Table 2). Note that the selected parameters (see Table 1) for the RSS model directly affect the results. There are various types of cars that are allowed to drive on motorways, and the CommonRoad dataset did not include information on the types of the vehicles. Also, the dataset did not include information on the road conditions that can be used to determine the proper values for the RSS parameters. As a result, we chose parameters in Table 1 for the RSS rules that are reasonable for average passenger vehicles. For the sake of completeness, in the following, we discuss the results of a sensitivity analysis on the RSS parameters [10, 30]. We believe that our results highlight the role of regulatory bodies in determining a meaningful set of parameters for the RSS models.

Table 5: Sensitivity analysis of the RSS parameters. Units for accelerations, time and distances are m/s^2, *seconds*, and *meters*, respectively. The value of μ is 0.4 as in Table 1.

parameter	values								
$a^{lon}_{max,accel}$	2.75			5.5			8.25		
$a^{lat}_{max,accel}$	1.5			3			4.5		
$a^{lon}_{max,brake}$	5			10			15		
$a^{lon}_{min,brake}$	6			4			2		
$a^{lat}_{min,brake}$	4.5			3			1.5		
ρ	0.3	0.5	2	0.3	0.5	2	0.3	0.5	2
# of violations	**20**	**33**	**435**	**90**	**204**	**606**	**261**	**583**	**902**

6.1 Sensitivity Analysis on RSS Parameters

We chose a busy scenario from our monitoring experiments to do sensitivity analysis on the RSS parameters. There are 9 different parameter configurations based on two parameter categories: acceleration/deceleration and response time. In order to create the different configurations for our analysis, we used three multiplicative factors which are applied to the base parameters in Table 1. The lists of multiplicative factors are (0.5, 1, 1.5) for acceleration and maximum-deceleration, (1.5, 1, 0.5) for minimum-deceleration, and (0.6, 1, 4) for response time. For example for the response time, in Table 5, we have $0.6\times0.5 = 0.3$, $1\times0.5 = 0.5$, and $4\times0.5 = 2$. The results are summarized in Table 5. Based on Table 5, a pattern was observed that by increasing the acceleration/deceleration bounds, the number of violations increases. Another interesting observation is that for slower response time, the number of violations increases drastically. Most of the new violations are the result of violating $\varphi^{\neg lat,\neg lon}$. The slower the response time, the longer the required safe longitudinal/lateral distances (i.e., see Def. 3.1 and Def. 3.2). The longer safe distances cause both lateral and longitudinal robustness safety to become negative in most of the cases and exclude the other longitudinal and lateral safety rules from being triggered. In the case of longer safe distances, the predicates that caused the highest number of violations are $A^{lat}_{l,maxAcc}$, $A^{lat}_{r,maxAcc}$, and $V^{lat}_{l,stop}$.

7 CONCLUSIONS

In this paper, we present a translation of the Responsibility-Sensitive Safety (RSS) [27] rules into Signal Temporal Logic (STL) [5] formulas. The encoded formulas could be used for Automated Driving System (ADS) model verification and/or automated test case generation for discovering control software bugs. In fact, the requirements as presented in this work can be used directly for testing ADS using our Sim-ATAV framework [28, 29]. We view this as a major motivation for formalizing the RSS model in STL. Now, it is straightforward to test the control and perception system stack against the RSS model.

In this paper, however, we provide an alternative – but equally important – application. We utilized the STL formulas to monitor off-line naturalistic driving data provided with CommonRoad [15]. We demonstrated that the computation is efficient, and, most importantly, that the RSS rules are satisfied in the majority of the actual vehicle trajectories (assuming fast reaction times by the drivers). Finally, we remark that we are currently working toward on-line (runtime) robustness monitoring [8] with the goal of deploying the monitoring system on FPGAs similar to [14, 19].

ACKNOWLEDGMENTS

This work was supported in part by NSF award 1350420 and by a gift from Intel Corporation.

REFERENCES

[1] Houssam Abbas, Georgios E. Fainekos, Sriram Sankaranarayanan, Franjo Ivancic, and Aarti Gupta. 2013. Probabilistic Temporal Logic Falsification of Cyber-Physical Systems. *ACM Transactions on Embedded Computing Systems* 12, s2 (May 2013).

[2] abc15.com staff. 2018. Self-driving car crash in Arizona: Red light runner hits Waymo van. *ABC15 Arizona* (5 2018).

[3] Matthias Althoff, Markus Koschi, and Stefanie Manzinger. 2017. CommonRoad: Composable benchmarks for motion planning on roads. In *2017 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 719–726.

[4] Nikos Arechiga. 2019. Specifying Safety of Autonomous Vehicles in Signal Temporal Logic. In *IEEE Intelligent Vehicles Symposium (IV)*. 58–63.

[5] Ezio Bartocci, Jyotirmoy Deshmukh, Alexandre Donzé, Georgios Fainekos, Oded Maler, Dejan Nickovic, and Sriram Sankaranarayanan. 2018. Specification-based Monitoring of Cyber-Physical Systems: A Survey on Theory, Tools and Applications. In *Lectures on Runtime Verification*. LNCS, Vol. 10457. Springer, 128–168.

[6] Glen Chou, Yunus Emre Sahin, Liren Yang, Kwesi J. Rutledge, Petter Nilsson, and Necmiye Ozay. 2018. Using Control Synthesis to Generate Corner Cases: A Case Study on Autonomous Driving. *IEEE Trans. on CAD of Integrated Circuits and Systems* 37, 11 (2018), 2906–2917.

[7] Alex Davies. 2016. Google's Self-Driving Car Caused Its First Crash. *Wired* (2 2016).

[8] Adel Dokhanchi, Bardh Hoxha, and Georgios Fainekos. 2014. On-Line Monitoring for Temporal Logic Robustness. In *Runtime Verification (LNCS)*, Vol. 8734. Springer, 231–246.

[9] Tommaso Dreossi, Alexandre Donze, and Sanjit A. Seshia. 2018. Compositional Falsification of Cyber-Physical Systems with Machine Learning Components. *arXiv:1703.00978v3* (2018).

[10] Laura Eboli, Gabriella Mazzulla, and Giuseppe Pungillo. 2016. Combining speed and acceleration to define car users safe or unsafe driving behaviour. *Transportation research part C: emerging technologies* 68 (2016), 113–125.

[11] Georgios E. Fainekos and George J. Pappas. 2006. Robustness of Temporal Logic Specifications. In *Formal Approaches to Testing and Runtime Verification (LNCS)*, Vol. 4262. Springer, 178–192.

[12] Georgios E Fainekos, Sriram Sankaranarayanan, Koichi Ueda, and Hakan Yazarel. 2012. Verification of automotive control applications using s-taliro. In *2012 American Control Conference (ACC)*. IEEE, 3567–3572.

[13] Thomas Ferrère, Dejan Nickovic, Alexandre Donzé, Hisahiro Ito, and James Kapinski. 2019. Interface-aware signal temporal logic. In *Proceedings of the 22nd ACM International Conference on Hybrid Systems: Computation and Control*. ACM, 57–66.

[14] Stefan Jaksic, Ezio Bartocci, Radu Grosu, Reinhard Kloibhofer, Thang Nguyen, and Dejan Nickovic. 2015. From signal temporal logic to FPGA monitors. In *13 ACM/IEEE International Conference on Formal Methods and Models for Codesign (MEMOCODE)*. 218–227.

[15] Markus Koschi, Stefanie Manzinger, and Matthias Althoff. [n. d.]. CommonRoad: Documentation of the XML Format. ([n. d.]).

[16] Timothy B. Lee. 2018. Report: Software bug led to death in UberâĂŹs self-driving crash. *Ars Technica* (5 2018).

[17] Sarah M. Loos, Andre Platzer, and Ligia Nistor. 2011. Adaptive Cruise Control: Hybrid, Distributed, and Now Formally Verified. In *Formal Methods (LNCS)*, Vol. 6664. Springer, 42–56.

[18] Oded Maler and Dejan Nickovic. 2004. Monitoring Temporal Properties of Continuous Signals. In *Proceedings of FORMATS-FTRTFT (LNCS)*, Vol. 3253. 152–166.

[19] Mohammadreza Mehrabian, Mohammad Khayatian, Aviral Shrivastava, John C Eidson, Patricia Derler, Hugo A Andrade, Ya-Shian Li-Baboud, Edward Griffor, Marc Weiss, and Kevin Stanton. 2017. Timestamp Temporal Logic (TTL) for Testing the Timing of Cyber-Physical Systems. *ACM Transactions on Embedded Computing Systems (TECS)* 16, 5s (2017).

[20] Stefan Mitsch and André Platzer. 2016. ModelPlex: verified runtime validation of verified cyber-physical system models. *Formal Methods in System Design* 49, 1-2 (2016), 33–74.

[21] Tanya Mohn. 2019. 2017-2018 Nissan Rogue Automatic Emergency Braking Presents 'Unreasonable Risk,' Safety Group Says. *Forbes* (3 2019).

[22] National Highway Traffic Safety Administration (NHTSA). [n. d.]. Automated Vehicles for Safety. https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety

[23] Matthew O'Kelly, Houssam Abbas, and Rahul Mangharam. 2017. Computer-Aided Design for Safe Autonomous Vehicles. In *Resilience Week*.

[24] Nima Roohi, Ramneet Kaur, James Weimer, Oleg Sokolsky, and Insup Lee. 2018. Self-driving vehicle verification towards a benchmark. *arXiv preprint arXiv:1806.08810* (2018).

[25] S-TaLiRo Tools. [n. d.]. https://sites.google.com/a/asu.edu/s-taliro/.

[26] Anthony Karel Seda and Pascal Hitzler. 2008. Generalized Distance Functions in the Theory of Computation. *Comput. J.* 53, 4 (2008), bxm108443–464.

[27] Shai Shalev-Shwartz, Shaked Shammah, and Amnon Shashua. 2018. On a formal model of safe and scalable self-driving cars. *arXiv:1708.06374v6* (2018).

[28] Cumhur Erkan Tuncali, Georgios Fainekos, Hisahiro Ito, and James Kapinski. 2018. Simulation-based Adversarial Test Generation for Autonomous Vehicles with Machine Learning Components. In *IEEE Intelligent Vehicles Symposium (IV)*.

[29] Cumhur Erkan Tuncali, Georgios Fainekos, Danil Prokhorov, Hisahiro Ito, and James Kapinski. 2019. Requirements-driven Test Generation for Autonomous Vehicles with Machine Learning Components. *arXiv 1908.01094* (2019).

[30] Jin Xu, Kui Yang, YiMing Shao, and GongYuan Lu. 2015. An experimental study on lateral acceleration of cars in different environments in Sichuan, Southwest China. *Discrete Dynamics in nature and Society* 2015 (2015).