Reaching Safe States in Autonomous Road Vehicles

Mario Gleirscher, Stefan Kugele

Reaching Safe States in Autonomous Road Vehicles

Mario Gleirscher* and Stefan Kugele†
Technische Universität München, Munich, Germany
Email: *mario.gleirscher@tum.de, †stefan.kugele@tum.de

***Abstract*—We outline core parameters to be taken into account when (i) defining safe states of autonomous passenger vehicles and (ii) deriving operational strategies of reaching them. We discuss the conception of fail-safe control strategies as well as the realization of such strategies in a control system architecture implementing them. This fast abstract outlines our research goals and our next steps.**

***Index Terms*—Fail-safe analysis, autonomous adaptive system.**

I. Background and Motivation

As an example of safety-critical autonomous systems, we consider *manned road vehicles in road traffic with an autopilot (AP) feature*, i.e., being able to automatically conduct a ride only given some valid target and minimizing human intermission. Let us consider the *system-level safety requirement*:

> The vehicle under consideration can always reach a safe state given a specific operational situation.

The Sections II to V develop on this statement. We furthermore assume the control system architecture consisting of a sensor subsystem, an actuator subsystem, and software-driven networked computing units, see Figure 1. A research system architecture for autonomous driving is also proposed in [1]. Based on this, Section VI discusses the relationship between control strategies and their realization. Section VII indicates how to choose from and switch between available strategies.

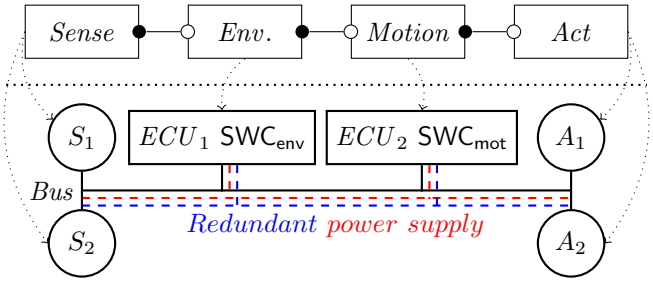

Figure 1: Control system architecture of a road vehicle.

II. Modeling the Road Vehicle Domain

Next, we introduce our vehicle domain modeling approach.

A. Modeling Abstract Vehicle Behavior

Given a transition system (Σ, Δ) with state space Σ and transition relation Δ, we model the system "driver-vehicle-road" as an abstract transition system for our strategy analysis (Figure 2). We consider subsets of Σ to model *operational situations* os and the abstract states σ_{haz}, σ_{saf}, and σ_{mis}. We consider subsets of Δ to model *system behaviors* within and across os as well as the transitions $comp$, $unsafe$, and $\neg comp$. Let $\mathcal{O}$ be set of all operational situations. Note, that our model allows to define transitions to be performed by any cooperation of driver, vehicle, and road.

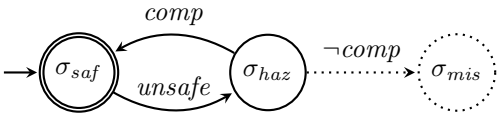

Figure 2: Abstract transition system for strategy analysis.

For a vehicle-side strategy in case of an internal error condition, the hazardous transition $unsafe$ can represent, e.g. a single component failure in the control system. The compensatory transition $comp$ represents any safety measure, e.g. degradation to a fail-operational slave component. Note, that we consider a successfully deployed airbag as part of $comp$ such that the mishap state is not reached in such an accident.

B. Modeling the State Space of the Vehicle Domain

For strategy analysis based on $\mathcal{O}$, we define the state space Σ using the following *parameter vector* $\vec{c}$:

- c_{driver} (driver cond.): physical presence, consciousness,
- c_{road} (road conditions): daylight, weather, traffic, road,
- c_{veh} (vehicle cond.): speed, location, fault conditions.

Then, the following functions of type $\mathcal{O} \to \mathbb{R}$ determine timing parameters important for our strategy analysis:

- t_{react} (driver reaction): est. time to take over control,
- t_{esc} (escalation time): est. time to hand over control,
- t_{loss} (control loss): estimated time till loss of control.

We do not regard all parameters and dependencies here.

Let $\Sigma = T(c_{driver}) \times T(c_{road}) \times T(c_{veh})$ where T returns the type of $\vec{c}$'s components. Each $os \in \mathcal{O}$ is defined as a set $os = \{\sigma \in \Sigma \mid p_{os}(\sigma)\}$ using a predicate p_{os} over σ.

Based on this model, Section III sketches regions in Σ covered by σ_{saf} and σ_{haz}. Similarly, Sections IV and V sketch what happens during the transitions $comp$ and $unsafe$.

III. Safe, Hazardous, and Mishap States

The regions of the safe, hazardous, and mishap states vary with the given operational situation.

We define the *safe state* $\sigma_{saf} = \{\sigma \in \Sigma \mid \bigvee_{os \in \mathcal{O}} \mathsf{safe}(\sigma, os)\}$ where $\mathsf{safe}(\sigma, os)$ determines whether σ is in the safe region of os. We split σ_{saf} into a set partition $\{\sigma^j_{saf}\}_{j \in 0..n}$ with a predicate $\mathsf{safe}_j(\sigma, os)$, $n \in \mathbb{N}$, and $j = 0$ denotes the *criticality level* nearest to the hazardous state.

Next, we define the *hazardous state* σ_{haz} similarly to σ_{saf} and split σ_{haz} into a set partition $\{\sigma^j_{haz}\}_{j\in 0..m}$ with a predicate $\mathsf{hazardous}_j(\sigma, os)$, $m \in \mathbb{N}$, and $j = 0$ nearest to the safe state.

Finally, the *mishap state* σ_{mis} reached by $\neg comp$ encompasses states where unacceptable harm or damage already happened without appropriate compensation.

For *consistency*, we require $\sigma_{haz}, \sigma_{saf}, \sigma_{mis} \subseteq \bigcup_{os\in\mathcal{O}} os$, $\sigma_{saf} \cup \sigma_{haz} \cup \sigma_{mis} = \bigcup_{os\in\mathcal{O}} os$, and $\sigma_{saf} \cap \sigma_{haz} \cap \sigma_{mis} = \emptyset$. Given a set of *criticality levels* $\Sigma_{cl} = \{\sigma^j_{saf}\}_{j\in 0..n} \cup \{\sigma^j_{haz}\}_{j\in 0..m}$ and a metric $\mathsf{cl} : \Sigma_{cl} \to \mathbb{R}$, we also require

$$\forall u \in 0..n, v \in 0..m, \sim \in \{<, >, =\} : \mathsf{cl}(\sigma^u_{saf}) < \mathsf{cl}(\sigma^v_{haz}) \\ \wedge (\mathsf{cl}(\sigma^v_{saf}) \sim \mathsf{cl}(\sigma^u_{saf}) \vee \mathsf{cl}(\sigma^u_{haz}) \sim \mathsf{cl}(\sigma^v_{haz}) \Rightarrow u \sim v).$$

Let $n = m = 1$ and $os = \{\sigma \in \Sigma \mid c_{driver} = SA \wedge c_{road} = SDC \wedge c_{veh} = MR\}$ with $os \in \mathcal{O}$, SA="driver seated, awake," SDC="sunny day, dense traffic, crossroads," MR="medium speed, right lane" for all further examples.

IV. Reaching the Hazardous State ($\sigma_{saf} \to \sigma_{haz}$)

Table I outlines transitions from σ_{saf} to σ_{haz} according to os. Each $unsafe$ transition refers to a scenario described as a transition system, e.g. $unsafe_{10}$ represents the hazard scenario "sensor wears out." For hazard analysis, we ask questions such as "(To which extent) Is $unsafe_{10}$ containing $unsafe_{00}$?" We can use our model to answer such a question because $unsafe_{10}$ can be refined into partial transitions in Δ and intermediate states in Σ. Note, that an $unsafe$ transition can stem from fault conditions on all sides, driver, vehicle, and road.

	σ^0_{haz} (imprecise sensor data, ...)	σ^1_{haz} (component failure, ...)
σ^0_{saf} ($t_{loss} = short, \ldots$)	$unsafe_{00}$	$unsafe_{01}$
σ^1_{saf} ($t_{loss} = long, \ldots$)	$unsafe_{10}$	$unsafe_{11}$

Table I: Reaching σ_{haz} in os.

V. Logical Compensation ($\sigma_{haz} \to \sigma_{saf}$)

Table II outlines transitions from σ_{haz} to σ_{saf} according to os. Each $comp$ transition refers to a strategy. The transition $comp$ can describe *behavioral safety tactics* [2] such as fail-operational, fail-silent, active (e.g. ABS, braking assistant) or passive (e.g. airbag) measures, or limp-home. When planning a compensation strategy, we ask questions such as "How can we adapt the system to reach the best σ^n_{saf} from the worst σ^m_{haz}?" We are particularly interested in $\mathsf{max}\{(\sigma^u_{haz}, \sigma^v_{saf}) \in \{\sigma^j_{haz}\}_{j\in 0..m} \times \{\sigma^j_{saf}\}_{j\in 0..n} \mid \mathsf{cl}(\sigma^u_{haz}) - \mathsf{cl}(\sigma^v_{saf})\}$.

In our example, $comp_{10}$ represents the fail-safe scenario "Switch back to basic driving control system." If σ^0_{saf} is accompanied with *dynamics* (i.e., physical movement) then we might choose a *fail-operational strategy* for $comp_{10}$ because the functions influencing the control of these dynamics need to be kept alive at least in a degraded version to quickly regain a *stable safe state*. Depending on t_{loss}, a strategy $comp_{11}$ can be difficult to plan or even impossible to realize.

	σ^0_{saf} (halted on lane, ...)	σ^1_{saf} (spreaded to safe side, ...)
σ^0_{haz} ($t_{loss} = long, \ldots$)	$comp_{00}$	$comp_{01}$
σ^1_{haz} ($t_{loss} = short, \ldots$)	$comp_{10}$	$comp_{11}$

Table II: Reaching σ_{saf} in os.

VI. Technical Compensation ($\sigma_{haz} \to \sigma_{saf}$)

Here, we consider details about the causal chains of $unsafe$ transitions and we can refine our compensation strategies $comp$. The constraint $t_{loss} \geq t_{esc} + t_{react} \wedge t_{esc} \geq t_{react}$ can act as an acceptance criterion for a viable $comp$ transition.

Based on our architecture in Figure 1, we can (1) refine the criticality levels and (2) introduce refined $comp$ transitions. At this level we model *dependability tactics* [3], [4] such as degradation or fail-over, MooN-D, or shutdown & repair.

The transition system of Figure 2 then has to be refined for the whole system, e.g. our vehicle architecture helps deriving failure modes for σ^u_{haz} and degraded modes for σ^v_{saf}.

VII. Combining Compensation Strategies

Having elaborated compensation strategies for relevant operational situations (Section V), and having refined them (Section VI), we can evaluate them according to several criteria:

- Which of the strategies are, e.g. time and energy efficient?
- Which require minimal or maximal human intermission?
- Which switches between operational situations happen?
- Which relationships among strategies, operational situations, and Σ_{cl} states help minimizing implementations?
- Which probabilities do $unsafe$ and $comp$ have?

Tables I and II lay a *basis for the evaluation, choice, and combination of strategies* according to $\vec{c}$, $\mathcal{O}$ (Section II-B), and the σ_{saf} to be currently reached.

VIII. Conclusion and Outlook

We sketched a multi-criteria, multi-stage approach to reaching a context-dependent safe state during an operational highly automated system, where critical parts of the compensation strategies have to be performed by cooperation of the autonomous system, its operators, and its environment.

Next, we want to formalize control system design patterns for safe autonomous systems, and perform probabilistic model checking of these strategies. We plan to evaluate our results in a project together with the automotive industry.

References

[1] J. Wei, J. M. Snider, J. Kim, J. M. Dolan, R. Rajkumar, and B. Litkouhi, "Towards a viable autonomous driving research platform," in *Intelligent Vehicles Symposium (IV), 2013 IEEE*, June 2013, pp. 763–770.

[2] M. Gleirscher, "Behavioral safety of technical systems," Dissertation, Technische Universität München, Dec 2014.

[3] C. Preschern, N. Kajtazovic, C. Kreiner *et al.*, "Catalog of safety tactics in the light of the IEC 61508 safety lifecycle," in *Proc VikingPLoP*, 2013, pp. 79–95.

[4] W. Wu and T. Kelly, "Safety tactics for software architecture design," in *Computer Software and Applications Conference (COMPSAC). Proc. 28th Ann. Int.*, vol. 1, Sep. 2004, pp. 368–75.