Reaching Safe States in Autonomous Road Vehicles
Mario Gleirscher, Stefan Kugele

To cite this version:
Mario Gleirscher, Stefan Kugele. Reaching Safe States in Autonomous Road Vehicles. Fast abstracts at International Conference on Computer Safety, Reliability, and Security (SAFECOMP), 2016, Trondheim, Norway. hal-01370229

HAL Id: hal-01370229
https://laas.hal.science/hal-01370229
Submitted on 22 Sep 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
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 intervention. 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.

II. MODELING THE ROAD VEHICLE DOMAIN

Next, we introduce our vehicle domain modeling approach.

A. Modeling Abstract Vehicle Behavior

Given a transition system \((\Sigma, \Delta)\) with state space \(\Sigma\) and transition relation \(\Delta\), we model the system “driver-vehicle-road” as an abstract transition system for our strategy analysis (Figure 2). We consider subsets of \(\Sigma\) to model operational situations \(os\) and the abstract states \(\sigma_{saf}, \sigma_{haz},\) and \(\sigma_{mis}\). We consider subsets of \(\Delta\) to model system behaviors within and across \(os\) as well as the transitions \(comp, unsafe,\) and \(¬comp\).

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 \(\Sigma\) 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 \(\mathcal{O} \rightarrow \mathbb{R}\) 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 \(c\)’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 \(\sigma\).

Based on this model, Section III sketches regions in \(\Sigma\) covered by \(\sigma_{saf}\) and \(\sigma_{haz}\). 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 \(\sigma_{saf} = \{\sigma \in \Sigma \mid \forall os \in \mathcal{O} safe(\sigma, os)\}\) where \(safe(\sigma, os)\) determines whether \(\sigma\) is in the safe region of \(os\). We split \(\sigma_{saf}\) into a set partition \(\{\sigma_{saf}^j\}_{j \in \mathbb{N}}\) with a predicate \(safe_j(\sigma, os), n \in \mathbb{N}\), and \(j = 0\) denotes the criticality level nearest to the hazardous state.
Next, we define the hazardous state $\sigma_{\text{haz}}$ similarly to $\sigma_{\text{saf}}$ and split $\sigma_{\text{haz}}$ into a set partition $\{\sigma_{\text{haz}}^j\}_{j \in 0..m}$ with a predicate hazardous $\{\sigma_{\text{o}}\}$, $m \in \mathbb{N}$, and $j = 0$ nearest to the safe state.

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

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

$$\forall u \in 0..n, v \in 0..m, \in \{<,>,=\} : c(\sigma_{\text{saf}}^u) < c(\sigma_{\text{haz}}^u) \land (c(\sigma_{\text{saf}}^v) \sim c(\sigma_{\text{saf}}^u) \lor c(\sigma_{\text{haz}}^v) \sim c(\sigma_{\text{haz}}^u)) \Rightarrow u \sim v).$$

Let $n = m = 1$ and $\sigma_{\text{o}} = \{\sigma \in \Sigma | c_{\text{driver}} = \text{SA} \lor c_{\text{road}} = \text{SDC} \land c_{\text{veh}} = \text{MR}\}$ with $\sigma_{\text{o}} \in \mathcal{O}$, $\text{SA} = \text{“driver seated, awake,” SDC} = \text{“sunny day, dense traffic, crossroads,” MR} = \text{“medium speed, right lane”}$ for all further examples.

**IV. REACHING THE HAZARDOUS STATE ($\sigma_{\text{saf}} \rightarrow \sigma_{\text{haz}}$)**

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

<table>
<thead>
<tr>
<th>$\sigma_{\text{haz}}^0$ (imprecise sensor data, . . .)</th>
<th>$\sigma_{\text{haz}}^1$ (component failure, . . .)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_{\text{saf}}^{00}$ (t$_{\text{loss}}$ = short, . . .)</td>
<td>unsafe$_{00}$</td>
</tr>
<tr>
<td>$\sigma_{\text{saf}}^{10}$ (t$_{\text{loss}}$ = long, . . .)</td>
<td>unsafe$_{10}$</td>
</tr>
</tbody>
</table>

Table I: Reaching $\sigma_{\text{haz}}$ in $\sigma_{\text{o}}$.

**V. LOGICAL COMPENSATION ($\sigma_{\text{haz}} \rightarrow \sigma_{\text{saf}}$)**

Table II outlines transitions from $\sigma_{\text{haz}}$ to $\sigma_{\text{saf}}$ according to $\sigma_{\text{o}}$. 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 $\sigma_{\text{saf}}$ from the worst $\sigma_{\text{haz}}$?” We are particularly interested in $\max\{\sigma_{\text{saf}}^{\text{haz}}, \sigma_{\text{saf}}^{\text{comp}}\} \in \{\sigma_{\text{haz}}^j\}_{j \in 0..m} \times \{\sigma_{\text{saf}}^j\}_{j \in 0..n} \mid c(\sigma_{\text{saf}}^j) - c(\sigma_{\text{saf}}^{\text{haz}})\}$. In our example, $\text{comp}_{10}$ represents the fail-safe scenario “Switch back to basic driving control system.” If $\sigma_{\text{saf}}^j$ is accompanied with dynamics (i.e., physical movement) then we might choose a fail-operational strategy for $\text{comp}_{11}$ 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 $t_{\text{loss}}$, a strategy $\text{comp}_{11}$ can be difficult to plan or even impossible to realize.

**VI. TECHNICAL COMPENSATION ($\sigma_{\text{haz}} \rightarrow \sigma_{\text{saf}}$)**

Here, we consider details about the causal chains of unsafe transitions and we can refine our compensation strategies comp. The constraint $t_{\text{loss}} \geq t_{\text{esc}} \lor t_{\text{read}} \geq t_{\text{react}}$ 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 $\sigma_{\text{saf}}$ and degraded modes for $\sigma_{\text{saf}}$. The failure modes for $\sigma_{\text{saf}}^0$ and the whole system, e.g. our vehicle architecture helps deriving failure modes for $\sigma_{\text{saf}}^0$ and degraded modes for $\sigma_{\text{saf}}$.

**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 intervention?
- Which switches between operational situations happen?
- Which relationships among strategies, operational situations, and $\Sigma_{\text{cl}}$ 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 $\mathcal{C}$, $\mathcal{O}$ (Section II-B), and the $\sigma_{\text{saf}}$ 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**