



# Specification-Guided Critical Scenario Identification for Automated Driving

Adam Molin<sup>1</sup>(✉), Edgar A. Aguilar<sup>2</sup>, Dejan Ničković<sup>2</sup>, Mengjia Zhu<sup>3</sup>,  
Alberto Bemporad<sup>3</sup>, and Hasan Esen<sup>1</sup>

<sup>1</sup> DENSO AUTOMOTIVE Deutschland GmbH, 85386 Eching, Germany  
{a.molin,h.esen}@eu.denso.com

<sup>2</sup> AIT Austrian Institute of Technology GmbH, 1210 Vienna, Austria  
{edgar.aguilar,dejan.nickovic}@ait.ac.at

<sup>3</sup> IMT School for Advanced Studies Lucca, 55100 Lucca, Italy  
{mengjia.zhu,alberto.bemporad}@imtlucca.it

**Abstract.** To test automated driving systems, we present a case study for finding critical scenarios in driving environments guided by formal specifications. To that aim, we devise a framework for critical scenario identification, which we base on open-source libraries that combine scenario specification, testing, formal methods, and optimization.

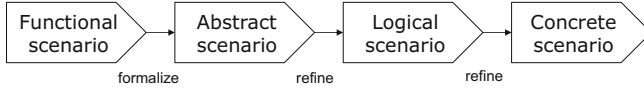
**Keywords:** Autonomous vehicles · Scenario based testing

## 1 Introduction

With the complexity of the automated driving (AD) system and its driving environment, verification and validation (V&V) is regarded as one of the major challenges of AD development [25]. *Scenario-based testing* (SBT) was introduced as an essential method for facilitating the overall safety assurance of ADs. In SBT, the expected behavior of an AD system is described by a representative set of scenarios that are relevant for its safe use. The SBT paradigm facilitates shifting the AD testing from the physical to the simulation environment. The use of virtual testing has manifold advantages – more specifically it allows to: (1) explore efficiently a large number of situations originating from the catalog of relevant scenarios, (2) reproduce environment conditions (fog, night, rain, etc.) that are hard to enforce in a physical environment, and (3) play dangerous scenarios without risk to humans, other vehicles or infrastructure.

Despite significant advances in research and standardization of SBT, there are still remaining open issues. One of them is to determine the critical scenarios among the virtually infinite number of scenarios with an abundance of influential factors ranging from weather or road conditions, to the behaviors of surrounding road users. A first attempt to keep the number of scenarios manageable is to

■ This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 956123.



**Fig. 1.** Scenario abstraction types according to [17].

restrict the operational design domain (ODD) of the AD system. According to [20], the ODD is defined as the operating conditions under which a given AD system is specifically designed to function. However, there are some factors, including the dynamic behavior of the road users, which cannot be controlled in the ODD. Thus, efficient methods are needed to identify the critical scenarios from the scenario space within the ODD. An extensive survey study on finding critical scenarios has been conducted in [25]. With regard to specification-guided critical scenario identification, our work is closely related to [8, 21, 23, 24].

In this paper, we present a specification-driven framework for critical scenario identification (CSI) entirely based on open-source software libraries and demonstrate its benefits with an automated emergency break case study. The proposed framework, based on the falsification testing paradigm [18], uses optimization-based methods for finding critical scenarios. We first describe the vanilla workflow and show how to tailor it with custom test generation and monitoring strategies. Hence, our aim is to share our experience in combining existing methods into a flexible and efficient SBT framework. To innovate the methodology for SBT within the framework, we investigate the separation between the AD system and the other road users, modeling their interplay with Assume/Guarantee (A/G) contracts. By using A/G contracts, we can improve the search for meaningful scenarios, assign responsibility for critical situations and distinguish between invalid behaviors originating from the AD system and from its environment. In that way, we can detect the violation of environment assumptions in the simulation execution, and discard the test run. By sharing our experience in SBT, we intend to nurture the innovation of prospective CSI methods that are based on specification-guided strategies.

## 2 Specification-Driven Scenario-Based Testing

### 2.1 Traffic Scenario Description

In the operational domain in which the ADS will be deployed, it is exposed to a potentially infinite number of traffic scenarios. As a consequence, it is impractical to conduct testing - even in simulation - directly on these traffic scenarios. A first step towards a successful application of scenario-based testing to assure the correct behavior of an ADS within its ODD is the abstraction of traffic scenarios. While the argumentation for quality assurance is done on a higher level of abstraction, the creation of evidence is performed on simulating a variety of concrete traffic scenarios derived from the abstract ones. The PEGASUS project “for the establishment of generally accepted quality criteria, tools and methods as

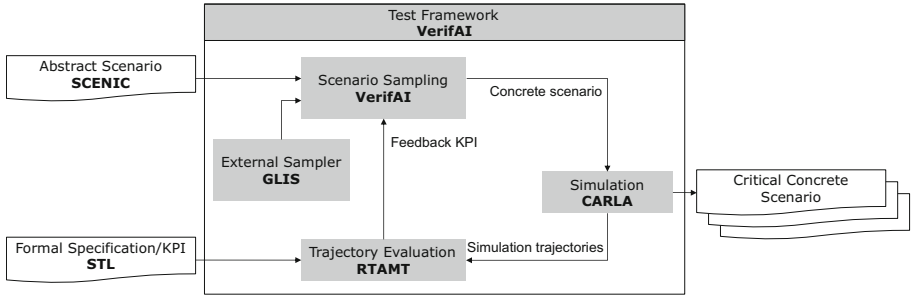
**Table 1.** Supported types and properties of scenario description formats

	OSC1.2	OSC2.0	SCENIC
Scenario types			
Functional	×	×	×
Abstract	×	✓	✓
Logical	✓	✓	✓
Concrete	✓	✓	✓
Properties			
Syntax format	XML <sup>a</sup>	DSL <sup>b</sup> , pythonic	DSL, pythonic
Language paradigm	imperative	Mostly declarative	Declarative/imperative
Map-agnostic scenario definition	×	✓	✓

<sup>a</sup> Extensible Markup Language

<sup>b</sup> Domain-specific language

well as scenarios and situations for the release of highly-automated driving functions”, introduced three abstraction types: functional, logical, concrete scenarios [16]. In this paper, we use an extended classification proposed in [17], see Fig. 1. Functional scenarios are defined as behavior-based, non-formal descriptions of traffic scenarios in natural language. Abstract scenarios are a formalization of functional scenarios using a declarative way to describe the scenario. Logical scenarios are defined as a parameterized set of traffic scenarios, while concrete scenarios are instances of a logical scenario with fixed parameters. They have a fixed scenery and road user behavior, that is based on the ego-vehicle movement. Abstract, logical, and concrete scenarios are machine-readable, and various realizations of traffic scenario description formats exist for simulation. In the following, we give a comparison between three non-proprietary, and openly available scenario description formats: OpenSCENARIO<sup>®</sup>1.2 (OSC1.2) [3], OpenSCENARIO<sup>®</sup>2.0 (OSC2.0) [4], and Scenic [12], see Table 1. With regard to the overall traffic scenario, their focus is on the initial placement and the dynamic behavior of the actors. The description of the scenery, such as the map, is defined outside these formats. OSC1.2 is mainly used for describing concrete traffic scenarios that can be directly run by the simulator. The actors’ placement and behavior are defined in an imperative fashion using pairs of actions and triggers that evoke these actions. OSC2.0’s and Scenic’s main intent is to define abstract scenarios, which can be concretized by a dedicated scenario generation engine. OSC2.0’s description is mostly declarative by constraining the road users’ behavior. The probabilistic programming language Scenic is declarative in the initial actor placement with a rich instruction set for relationships between entities, and uses an imperative description for behaviors. All three languages support parameterization of scenario parameters to describe logical scenarios. A distinctive feature of OSC2.0 and Scenic compared to OSC1.2 is that the location of the scenario does not need to be specified within the scenario definition. Instead, the scenario generation engine will find a suitable segment on the road map, on which the scenario can be executed with all actors in the simulator.



**Fig. 2.** Critical scenario identification framework with tool architecture.

Based on the scenario format, a database of abstract/logical scenarios needs to be created that covers all the relevant features in the considered ODD of the AD function. In this paper, we selected Scenic as our scenario format, due to both its flexibility in expressing abstract scenarios and the availability of an open-source testing framework [9] that is provided for Scenic.

## 2.2 Critical Scenario Identification

This section introduces the test framework to find critical concrete scenario instances within a specified abstract scenario efficiently and in a flexible manner. The framework depicted in Fig. 2 indicating the overall workflow is based on open-source software components highlighted in bold. It assumes two inputs, the abstract scenario given in the Scenic format, and a formal specification of the AD system defined in signal temporal logic (STL), that we use as a test oracle. The technical details on the formal specification are introduced in Sect. 2.3.

*Workflow.* The test execution framework is based on Berkeley’s VerifAI [9]. By applying a sampling strategy, VerifAI generates concrete scenarios from the Scenic scenario that are executed in the CARLA simulator [7]. To evaluate the resulting trajectories, we integrated RTAMT - an STL monitoring library [19] - into the VerifAI-based testing framework. RTAMT provides the automated generation of robustness monitors from STL specifications and therefore facilitates checking simulation traces against the formal specification. The robustness measure is then fed back as a criticality indicator to the scenario sampler that determines new test parameters that constitute the next concrete scenario to be simulated. Depending on the sampling strategy, the scenario search can be of explorative or exploitative nature. Instead of using the sampling strategies provided by VerifAI, we integrated an external sampling strategy, that is based on the global optimizer GLIS [5]. The details about GLIS are outlined in Sect. 2.4.

## 2.3 Formal Specifications

Concrete scenarios are typically evaluated against requirements. These requirements can cover various aspects, including safety, legal, comfort and ethical con-

siderations. In order to avoid ambiguities and facilitate their evaluation, there is a need to formulate requirements using a formal specification language. In this paper, we adopt signal temporal logic (STL) [15] as our specification formalism. There are several motivations to choose STL for requirement formalization: (1) an existing body of work already captures AD system requirements using STL, (2) STL admits quantitative semantics that can be used to guide the search for critical scenarios, and (3) there are runtime verification tools that enable evaluation of STL properties. The syntax of STL is given by the grammar

$$\varphi ::= \top \mid f(R) > 0 \mid \neg\varphi \mid \varphi_1 \vee \varphi_2 \mid \varphi_1 \mathcal{U}_I \varphi_2 \mid \varphi_1 \mathcal{S}_I \varphi_2,$$

where  $f(R)$  are terms in  $\Theta$  and  $I$  are real intervals with bounds in  $\mathbb{Q}_{>0} \cup \{\infty\}$ . As customary we use  $\diamond_I \varphi \equiv \top \mathcal{U}_I \varphi$  for *eventually*,  $\square_I \varphi \equiv \neg \diamond_I \neg \varphi$  for *always*,  $\diamond_I \varphi \equiv \top \mathcal{S}_I \varphi$  for *once* and  $\square_I \varphi \equiv \neg \diamond_I \neg \varphi$  for *historically*. The timing interval  $I$  may be omitted when  $I = [0, \infty)$  or  $I = (0, \infty)$ . STL can be naturally equipped with *quantitative semantics* based on the infinity norm [6] that measure how far is the observed behavior from satisfying or violating a requirement.

The evaluation of an AD system cannot be performed in isolation from its environment. For instance, an AD system cannot guarantee safety requirements, such as RSS, in presence of other road users that do not behave in a reasonable manner. The relation of the AD system and the environment under which it operates can be formalized in terms of a *contract*  $C = (\varphi, \psi)$ , a pair of properties where  $\varphi$  represents the assumptions on the environment and  $\psi$  guarantees of the system under these assumptions. This classical interpretation of  $C$  is given by the temporal logic formula

$$\square \varphi \rightarrow \square \psi.$$

According to the above formula, any violation of the assumption by the environment results in the (vacuous) satisfaction of the contract, even if the system also violates its guarantee. However, this definition neglects that these two violations may not be causally related – the violation of  $\psi$  by the system at time  $t$  before the violation of  $\varphi$  by the environment at time  $t' > t$  still results in the satisfaction of the contract. To address this situation, we propose a more refined notion of a contract that takes the intended temporal causality between the environment and the AD system into account. We denote our refined contract by  $\hat{C}$  and capture its meaning using the formula:

$$\square((\square_{[0,T]} \varphi) \rightarrow \psi)$$

where  $T$  specifies the maximum duration within which we consider the violation of  $\varphi$  to be causally related to the violation of  $\psi$ .

## 2.4 Sampling Strategy

Different sampling strategies may be used to identify the parameters of the next concrete scenario to simulate. These strategies can be broadly divided into naïve (passive) and guided search (active) sampling strategies [25]. The naïve

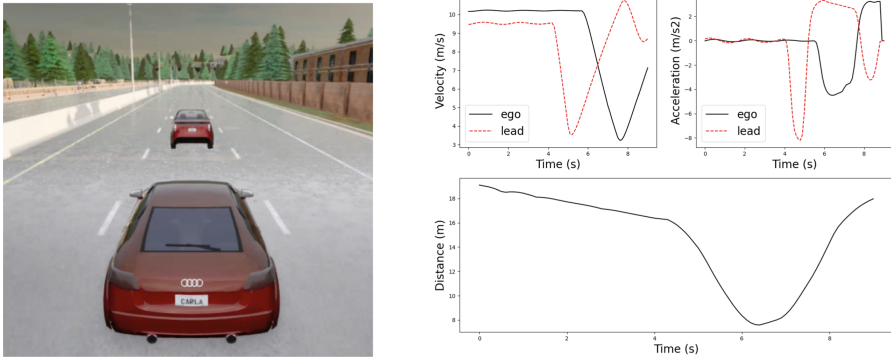
search strategies, such as random sampling, involve the independent selection of test parameters. In contrast, the guided search, such as optimization [10, 11], make the selection based on a specific selection criterion and the information of existing samples. Naïve search sampling strategies are useful if the simulation is computationally cheap to run since parallelization of the procedure is possible due to the independence among testing samples. On the other hand, when the test case simulation is computationally expensive to run and/or when the test cases interested (critical test cases in this case) are in a small region of the search domain, the guided search sampling strategies can be more sample efficient.

For the current study, guided-search sampling strategies such as surrogate-based black-box optimization methods are appropriate to efficiently identify relevant critical concrete scenarios for the AD system. It is because a closed-form expression of the KPI in terms of the test parameters is often unavailable. Specifically, we use the global optimization algorithm GLIS (Global optimization via Inverse distance weighting and Surrogate radial basis functions) [5] as the active guided-search sampler to identify the next test parameters of a concrete scenario for testing. The procedure of GLIS includes an initial sampling stage and an active learning stage. In the initial sampling stage,  $N_{\text{initial}}$  different test parameters are randomly selected within the search domain, and the corresponding concrete scenarios are simulated. The resulting quantitative evaluation of each test parameter from RTAMT monitors is fed back to GLIS (c.f. Fig. 2). A surrogate radial basis interpolation function (RBF) representing the correlation between the test parameters and the KPI is fitted to the initial samples. In the active learning phase, at each iteration, we identify a new test parameter, simulate the corresponding concrete scenario, and refit the surrogate function by including the newly identified test parameter and its KPI. The new test parameter is obtained by optimizing an acquisition function, which trades off the exploitation of the fitted RBF surrogate and exploration of an inverse distance weighting (IDW) function. IDW is a distance-based exploration function that promotes visiting points far away from the existing samples, which helps prevent the solver from being trapped in the local optima. GLIS terminates when the maximum allowed iteration is reached, or another user-defined criterion is met.

GLIS is chosen for this study, as it easily incorporates constraints and has a low computing cost [5]. If the computing cost is reasonable, GLIS may be replaced by other surrogate-based active samplers, such as Bayesian optimization.

### 3 Automatic Emergency Braking Case Study

To illustrate the methodology, we focus on testing a simple Automatic Emergency Braking (AEB) functionality using a highway scenario.



**Fig. 3.** (left) Snapshot of CARLA simulator running the AEB function test on a highway. (right) Example of telemetric data collected from all actors.

*Scenario Description.* The functional scenario is an ego vehicle following a leading vehicle on a highway, when suddenly the leading vehicle brakes abruptly. The ego vehicle is equipped with a simplistic distance-based AEB function which is activated when the ego is less than `safeDist` meters from the leading vehicle. Figure 3 shows a snapshot of the scenario running in CARLA v9.10.

The abstract scenario, depicted in Listing 1.1, is formulated using Scenic<sup>1</sup>. The scenario first specifies the sampler and the map used to generate concrete simulations (lines 1 and 2). Then, it defines parameter variables that we partition into: (1) the *constant* variables (lines 3–4) that do not change across concrete scenarios and (2) the *optimization* variables (lines 6–7) that are fed to an external (VerifAI) sampler in order to find critical scenarios in a controlled fashion. There are also what we call *implicit* variables that are not explicitly part of the Scenic abstract scenario but still need to have a concrete value in the simulator. For example, the weather conditions, the exact starting position and orientation of each vehicle, the vehicle model, etc. In this case study, there are more than 25 implicit parameters. The scenario also defines the behavior of the ego (lines 9–12) and of the lead vehicle (lines 14–18). Both the ego and the lead vehicle follow the lane with some target speed as their default behavior. However, the lead vehicle abruptly breaks at regular intervals, while the ego breaks when it approaches any object at some minimum distance. The two vehicles are spawned at some uniformly chosen part of the map (line 20) that is sufficiently far away from an intersection (line 25). The lead car is initialized at some pre-defined distance in front of the ego vehicle (lines 22–23).

---

```

1 param verifaiSamplerType = `glis' # specify sampler
2 param carla_map = `Town04' # specify map to use
3 initDist = 30 # constant
4 leadSpeed = 10
5

```

<sup>1</sup> The shown scenario is simplified to facilitate presentation.

```

6  safeDist = VerifaiRange(25,45) # optimization variable
7  egoSpeed = VerifaiRange(9,11) # optimization variable
8
9  behavior AEB_Behavior: # define ego behavior
10  try:
11    FollowLaneBehavior(egoSpeed)
12    interrupt when withinDistanceToObjsInLane(self, safeDist):
13      take SetBrakeAction(1), SetThrottleAction(0)
14
15  behavior Brake_Behavior: # define behavior of lead car
16  try:
17    FollowLaneBehavior(leadSpeed)
18    interrupt when simulation().currentTime > delay:
19      take SetBrakeAction(1), SetThrottleAction(0)
20
21  spawnPt = Uniform(*HighwayRoads) # Highway part of map
22
23  ego = Car at spawnPt, with AEB_Behavior} # spawn ego
24  leadCar = Car at spawnPt + initDist, with Brake_Behavior # lead
25
26  require (distance from leadCar to intersection > 50)
27  # extra requirements for rejection sampling

```

---

**Listing 1.1.** AEB highway scenario in Scenic

*Formalized Requirements.* We illustrate the formalization of the requirements with the contract  $C = (\varphi, \psi)$ , which captures the assumption  $\varphi$  about the maximum allowed deceleration of the lead vehicle and the guarantee  $\psi$  as the Responsibility-Sensitive Safety (RSS) property of the ego vehicle. The assumption  $\varphi$  originates from the IEEE Standard 2846-2022 [1], that describes the minimal set of assumptions on the road users for safety-related models of AD. From the assumptions described in the standard, we focus on the maximum deceleration specification

$$\varphi = \beta \leq \beta_{max}.$$

The Responsibility-Sensitive Safety (RSS) rule specifies, under minimal assumptions, what longitudinal and lateral distances the ego vehicle must keep from other road users to ensure no collisions [22]. The RSS rules were formalized into temporal logic by [2, 14]. We adopt the STL specification from [2] for an ego vehicle (*back*) to keep a safe longitudinal distance to another vehicle (*front*):

$$\begin{aligned}
& \square (v_{front} \geq 0 \wedge v_{back} \geq 0) \\
& \square (a_{front} \in [a_{max-Br}, a_{max-Acc}] \wedge a_{back} \in [a_{max-Br}, a_{max-Acc}]) \\
& \square (d(front, back) < d_{safe} \rightarrow a_{back} \in [a_{max-Br}, a_{min-Br}])
\end{aligned}$$

where  $a, v$  are correspondingly acceleration and velocity. Similarly  $a_{max-Acc}$ ,  $a_{max-Br}$ ,  $a_{min-Br}$  are assumed maximum acceleration, maximum braking, and minimum braking acceleration. Finally,  $d_{safe}$  is determined dynamically depending on the velocities of both vehicles, and the reaction time  $\tau$  of the ego vehicle:

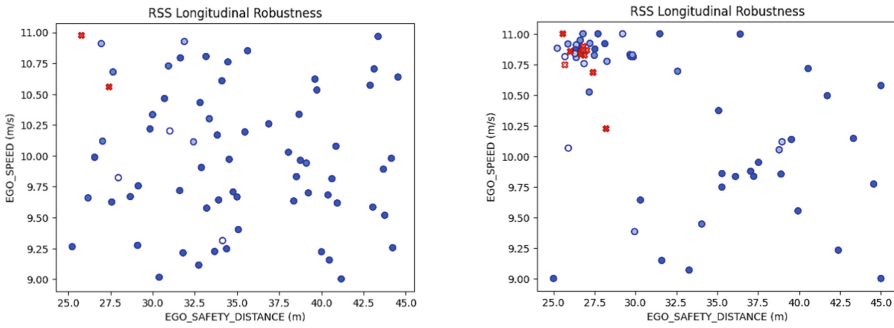


$$d_{\text{safe}} = \left( v_{\text{back}}\tau + \frac{a_{\text{max-Acc}}\tau^2}{2} + \frac{(v_{\text{back}} + a_{\text{max-Acc}}\tau)^2}{2a_{\text{min-Br}}} - \frac{v_{\text{front}}^2}{2a_{\text{max-Br}}} \right).$$

The safety distance is calculated in order to ensure that a collision is avoided as long as the ego vehicle is sufficiently far away from the leading vehicle. If it is momentarily closer than  $d_{\text{safe}}$  then a collision will still be avoided if the ego is reacting appropriately (by braking with at least  $a_{\text{min-Br}}$ ).

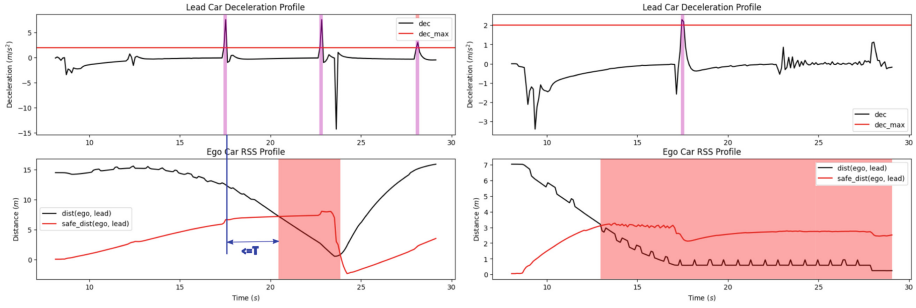
### 3.1 Simulation Results

In this section we present our evaluation outcomes. Figure 4 shows the results from simulating the abstract scenario 70 times using the described tool chain. Each point in the scatter plots represents a simulated concrete scenario, where the RSS longitudinal distance was monitored. If the ego vehicle managed to react adequately by braking in time, then this is represented as a blue circle, otherwise (if the specification was violated) it is represented by a red cross (the intensity of the color represents the robustness degree).



**Fig. 4.** Comparison between Halton sampling (left) and GLIS sampling [5] (right) for 70 concrete scenarios. The GLIS parameters are:  $\alpha = 1$ ,  $\delta = 0.5$ ,  $\epsilon_{\text{SVD}} = 0.01$ , and an inverse-quadratic basis function with  $\epsilon = 0.2$  was used.

Furthermore, we compare two different sampling strategies to find critical scenarios. In this case, we compare a passive sampling strategy (i.e. agnostic to feedback) which is based on Halton sequences [13], to an active strategy based on the GLIS optimization sampling. As expected, sampling scenarios with GLIS leads to the discovery of more critical scenarios (11 compared to 2 with Halton), and suggests variable regions which should be further investigated. In our example, the optimizer clearly was trying to exploit around the region of higher `egoSpeed`, and lower `safeDist` (as expected). In practice, both strategies are used to obtain a clear picture of the performance of the ADAS functionality.



**Fig. 5.** Evaluation with A/G contracts.

In Fig. 5, we illustrate the discrepancy between the classical and the refined interpretation of A/G contracts. The figure depicts two simulations showing the deceleration  $\beta$  of the lead vehicle and the maximum allowed deceleration threshold  $\beta_{max} = 2 \text{ m/s}^2$  (top) and the distance between the ego and the lead vehicle, as well as the safe distance between them (bottom). We see that in the two simulations both the assumption  $\varphi$  and the guarantee  $\psi$  are violated (purple and red stripes, respectively). In the first simulation (left), there is a clear causality between the abrupt breaking of the lead vehicle and the longitudinal RSS violation – it follows that the contract is satisfied under both the classical and the refined interpretation. In the second simulation (right), the violation of the longitudinal RSS requirement happens before the lead vehicle breaks. Intuitively, we expect the contract to be violated since the behavior of the lead vehicle did not cause this critical scenario. However, under the classical contract interpretation, the contract is satisfied because the lead vehicle does violate the assumption at a later stage. On the other hand, the refined contract rightly indicates the contract falsification.

### 3.2 Lessons Learned

In this section, we share our experience about the scenario-based testing framework and collected during the case study evaluation.

*Passive vs. Active Sampling.* Both passive and active sampling have their merits in testing AD systems. Passive sampling methods such as Halton provide a coverage of the parameter space, facilitate detecting interesting patterns, if any, and help identifying parameter regions that are interesting to further explore. In contrast, active sampling methods such as GLIS can accelerate the detection of critical scenarios.

*Level of Scenario Abstraction.* Balance between keeping a scenario abstract, and letting the tools sample different variables, and having consistent concrete scenarios. If too many variables are left unspecified, drawing meaningful conclusions

from the experiments is difficult, but if too many parameters are specified, there is a risk of missing out on potential critical scenarios that are relevant (and it also needs more development time).

*Optimization with Implicit Variables.* It is interesting to note that from the point of view of the optimizer, the robustness function of a concrete scenario is non-deterministic. That is, there are many different concrete scenarios that result from having the same `egoSpeed` and `safeDist` which result in different robustness values. This is mostly due to different implicit parameters impacting the robustness, which the optimizer does not directly see (e.g. road geometry).

## References

1. IEEE standard for assumptions in safety-related models for automated driving systems. IEEE Std 2846-2022, pp. 1–59 (2022). <https://doi.org/10.1109/IEEEESTD.2022.9761121>
2. Aréchiga, N.: Specifying safety of autonomous vehicles in signal temporal logic. In: 2019 IEEE Intelligent Vehicles Symposium (IV), pp. 58–63 (2019). <https://doi.org/10.1109/IVS.2019.8813875>
3. Association for Standardization of Automation and Measuring Systems: ASAM OpenSCENARIO V1.2.0. Standard, Munich, Germany (2022). <https://www.asam.net/standards/detail/openscenario/>
4. Association for Standardization of Automation and Measuring Systems: ASAM OpenSCENARIO V2.0.0. Standard, Munich, Germany (2022). <https://www.asam.net/standards/detail/openscenario/v200/>
5. Bemporad, A.: Global optimization via inverse distance weighting and radial basis functions. *Comput. Optim. Appl.* **77**, 571–595 (2020). <http://cse.lab.imtlucca.it/~bemporad/glis>
6. Donzé, A., Maler, O.: Robust satisfaction of temporal logic over real-valued signals. In: Chatterjee, K., Henzinger, T.A. (eds.) FORMATS 2010. LNCS, vol. 6246, pp. 92–106. Springer, Heidelberg (2010). [https://doi.org/10.1007/978-3-642-15297-9\\_9](https://doi.org/10.1007/978-3-642-15297-9_9)
7. Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., Koltun, V.: CARLA: an open urban driving simulator. In: Proceedings of the 1st Annual Conference on Robot Learning, pp. 1–16 (2017)
8. Dreossi, T., Donzé, A., Seshia, S.A.: Compositional falsification of cyber-physical systems with machine learning components. *J. Autom. Reason.* **63**(4), 1031–1053 (2019). <https://doi.org/10.1007/s10817-018-09509-5>
9. Dreossi, T., et al.: VerifAI: a toolkit for the formal design and analysis of artificial intelligence-based systems. In: 31st International Conference on Computer Aided Verification (CAV), July 2019
10. Feng, S., Feng, Y., Sun, H., Bao, S., Zhang, Y., Liu, H.X.: Testing scenario library generation for connected and automated vehicles, part II: case studies. *IEEE Trans. Intell. Transp. Syst.* **22**(9), 5635–5647 (2020)
11. Feng, S., Feng, Y., Yu, C., Zhang, Y., Liu, H.X.: Testing scenario library generation for connected and automated vehicles, part I: methodology. *IEEE Trans. Intell. Transp. Syst.* **22**(3), 1573–1582 (2020)
12. Fremont, D.J., Dreossi, T., Ghosh, S., Yue, X., Sangiovanni-Vincentelli, A.L., Seshia, S.A.: Scenic: a language for scenario specification and scene generation. In: Proceedings of the 40th ACM SIGPLAN Conference on Programming Language Design and Implementation, pp. 63–78 (2019)

13. Halton, J.H., Smith, G.B.: Algorithm 247: radical-inverse quasi-random point sequence. *Commun. ACM* **7**(12), 701–702 (1964). <https://doi.org/10.1145/355588.365104>
14. Hekmatnejad, M., et al.: Encoding and monitoring responsibility sensitive safety rules for automated vehicles in signal temporal logic. In: *Proceedings of the 17th ACM-IEEE International Conference on Formal Methods and Models for System Design. MEMOCODE 2019* (2019). <https://doi.org/10.1145/3359986.3361203>
15. Maler, O., Nickovic, D.: Monitoring temporal properties of continuous signals. In: *Formal Techniques, Modelling and Analysis of Timed and Fault-Tolerant Systems, Joint International Conferences on Formal Modelling and Analysis of Timed Systems, FORMATS 2004 and Formal Techniques in Real-Time and Fault-Tolerant Systems, FTRTFT 2004, Grenoble, France, 22–24 September 2004, Proceedings*, pp. 152–166 (2004)
16. Menzel, T., Bagschik, G., Maurer, M.: Scenarios for development, test and validation of automated vehicles. In: *2018 IEEE Intelligent Vehicles Symposium (IV)*, pp. 1821–1827. IEEE (2018)
17. Neurohr, C., Westhofen, L., Butz, M., Bollmann, M.H., Eberle, U., Galbas, R.: Criticality analysis for the verification and validation of automated vehicles. *IEEE Access* **9**, 18016–18041 (2021)
18. Nghiem, T., Sankaranarayanan, S., Fainekos, G., Ivancic, F., Gupta, A., Pappas, G.J.: Monte-carlo techniques for falsification of temporal properties of non-linear hybrid systems. In: *Proceedings of the 13th ACM International Conference on Hybrid Systems: Computation and Control, HSCC 2010, Stockholm, Sweden, 12–15 April 2010*, pp. 211–220 (2010)
19. Ničković, D., Yamaguchi, T.: RTAMT: online robustness monitors from STL. In: Hung, D.V., Sokolsky, O. (eds.) *ATVA 2020. LNCS*, vol. 12302, pp. 564–571. Springer, Cham (2020). [https://doi.org/10.1007/978-3-030-59152-6\\_34](https://doi.org/10.1007/978-3-030-59152-6_34)
20. On-Road Automated Driving (ORAD) committee: J3016 Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. Technical report (2021). <https://www.sae.org/standards/content/j3016-202104/>
21. Qin, X., Aréchiga, N., Best, A., Deshmukh, J.: Automatic testing with reusable adversarial agents. *arXiv preprint arXiv:1910.13645* (2019)
22. Shalev-Shwartz, S., Shammah, S., Shashua, A.: On a formal model of safe and scalable self-driving cars, August 2017. *arXiv e-prints*. [arXiv:1708.06374](https://arxiv.org/abs/1708.06374)
23. Tuncali, C.E., Fainekos, G., Prokhorov, D., Ito, H., Kapinski, J.: Requirements-driven test generation for autonomous vehicles with machine learning components. *IEEE Trans. Intell. Veh.* **5**(2), 265–280 (2019)
24. Tuncali, C.E., Pavlic, T.P., Fainekos, G.: Utilizing s-taliro as an automatic test generation framework for autonomous vehicles. In: *19th IEEE International Conference on Intelligent Transportation Systems, ITSC 2016, Rio de Janeiro, Brazil, 1–4 November 2016*, pp. 1470–1475 (2016)
25. Zhang, X., et al.: Finding critical scenarios for automated driving systems: a systematic mapping study. *IEEE Trans. Softw. Eng.* (2022)