



HAL
open science

Test framework and key challenges for virtual verification of automated vehicles: the VeriCAV project

Thomas Levermore, Alan Peters

► To cite this version:

Thomas Levermore, Alan Peters. Test framework and key challenges for virtual verification of automated vehicles: the VeriCAV project. 39th International Conference on Computer Safety, reliability and Security (SAFECOMP), Position Paper, Sep 2020, Lisbon, Portugal. hal-02931723

HAL Id: hal-02931723

<https://hal.science/hal-02931723>

Submitted on 7 Sep 2020

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.

Test framework and key challenges for virtual verification of automated vehicles: the VeriCAV project

1st Thomas Levermore

*New Mobility Technologies Directorate
Connected Places Catapult
Milton Keynes, UK
thomas.levermore@cp.catapult.org.uk
<https://orcid.org/0000-0003-4029-9337>*

2nd Alan Peters

*New Mobility Technologies Directorate
Connected Places Catapult
Milton Keynes, UK
alan.peters@cp.catapult.org.uk*

Abstract—Virtual simulation is considered a crucial step in assuring the safety of highly automated vehicles. Significant challenges however remain in producing valid evidence of real-world system performance from virtual simulation and doing so in a robust, efficient manner. A system architecture for virtual verification of highly automated vehicles is implemented in the VeriCAV project and described here, with a focus on the interfaces between system elements. The framework includes elements for test generation and assessment, virtual simulation, realistic virtual actors as well as the automated driving system being tested.

Index Terms—Scenario-based testing, virtual verification, automated vehicles

I. INTRODUCTION

A highly automated vehicle (HAV) is a safety-critical system - whose failure would result in injury or fatality to occupants or those in the environment of the vehicle. Ensuring HAV safety is challenging due to the complexity of the systems coupled with the almost limitless complexity and variability of the road operating environment. New methods for producing safety cases are required and this paper presents a test framework to support the development and implementation of methods to build safety case evidence with virtual simulation.

There is wide industry acceptance that HAVs will need to be verified and validated using a combination of virtual simulation, closed-road testing and public road trials [1], [2]. Regulatory forums such as the United Nations Economic Commission for Europe (UNECE) also propose this approach, described by the UNECE working group on Validation Method for Automated Driving (VMAD) as the “multi pillar approach” [3].

The physical testing element is expensive and time-consuming, particularly when considering the extent of testing required to demonstrate sufficient safety. Maximising the use of virtual simulation, where suitable, can reduce the need

for physical testing. Virtual simulation is already used extensively during early development and testing of the Automated Driving Systems (ADSs) that control HAVs due to the cost of prototype systems and the consequences of failure in a physical test. The need for virtual testing to form a larger proportion of the system development lifecycle is increased as ADS software development is becoming increasingly iterative. Changes to software occur regularly with practices such as Continuous Integration [4] resulting in more frequent repeat testing. Virtual testing can identify design flaws earlier in the development cycle than physical testing.

Virtual simulation in this context is commonly proposed in conjunction with scenario-based testing [5], [6]. A scenario, as defined in [7] can be described by road information, stationary objects, movable objects and their movements and environment conditions. This can be extended to include vehicle state such as load, connectivity related conditions and the wider scenario environment. This scenario-based testing allows challenging, rare, recently observed and/or standardised sets of scenarios to be the focus of ADS testing while avoiding extended periods of routine, unchallenging driving that exist in on-road testing. The extensive use of machine learning algorithms in ADSs also means that comprehensive scenario-based testing remains a viable option where existing formal verification methods are not suitable [8].

It has been suggested that the amount of on-road testing required to demonstrate the safety of a HAV is in the order of hundreds of millions of miles [9], if not more. Such an onerous approach is clearly not feasible on the road, however a combined approach including virtual simulation makes extensive testing possible by exploiting parallelism [2]. Even with the benefits of virtual simulation, such testing requires significant effort and resource. Efficient, targeted scenario generation is necessary to enable a viable testing programme [10].

The Verification of Connected and Autonomous Vehicles (VeriCAV) project addresses the practical challenges that remain for virtual testing and these are discussed in the following section. The project outputs are expected to support

the combined testing approach ultimately leading to verified and validated HAVs. VeriCAV¹ is a £3m, 2-year collaborative research and development project consisting of four organisations: Connected Places Catapult, HORIBA MIRA (as industry lead), University of Leeds, and Aimsun.

II. KEY SIMULATION CHALLENGES

There is a wide range of safety aspects that will need to be addressed with the introduction of HAVs, including new passive safety for new vehicle types and safety assessment in the field. Here the focus is on the safety of the ADS decision making. Standards are being developed to address this area including ISO/PAS 21448 Safety of the intended functionality (SOTIF) [11] and the IEEE project developing P2846 - Formal Model for Safety Considerations in Automated Vehicle Decision Making [12]. Scenario based testing supports building a HAV safety case and is the focus of ongoing ISO standardisation efforts as ISO/WD 34502 Road vehicles — Engineering framework and process of scenario-based safety evaluation [13]. While these standardisation efforts are welcomed, creating a fully evidenced safety case for a HAV is still an unsolved challenge - virtual simulation is part of the answer.

The challenges for scenario-based virtual testing addressed in the VeriCAV project are:

- 1) How can scenarios be described in a way that:
 - a) captures the complexity of diverse scenario sets,
 - b) allows automated scenario creation,
 - c) and can be unambiguously interpreted by simulation tools? [14]
- 2) How can the onerous activity of manual scenario generation be automated and made more efficient?
- 3) How can validity and fidelity of the virtual simulation be ensured?
- 4) How can the use of the virtual simulation system be maximised across a range of different ADSs?
- 5) How can virtual simulation coverage be ensured in an efficient manner?
- 6) How can test success be assessed in virtual simulations?

Each of these challenges are addressed in the functional sub-systems of the VeriCAV system architecture outlined in the following section.

III. VERICAV TEST FRAMEWORK

The VeriCAV system is made up of several complex sub-systems developed by the consortium partners outlined in Figure 1. The system architecture includes core components:

- a Test Generator, to produce scenarios to test,
- a Simulation Master, to interface a simulation tool to the rest of the system,
- a Test Oracle to analyse test results and direct the test generator.

Additionally, the VeriCAV system architecture includes components that can be exchanged:

- a Simulation Tool, in which the virtual scenario unfolds,
- a Smart Actor Controller, to provide realistic actor behaviour
- the System under Test, the ADS.

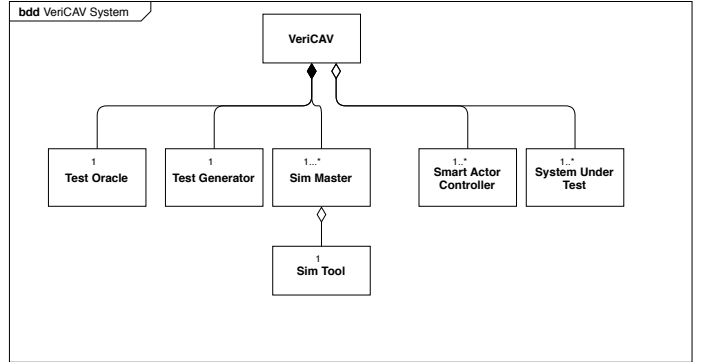


Fig. 1. Main components of the VeriCAV system expressed in SysML. The components that compose the core of the VeriCAV framework are shown on the left, with black diamond connector. The aggregated components (those with white diamonds) are modular components for which the framework is designed to be agnostic to their selection (i.e multiple simulation tools).

A. Test Generator

To address Challenge 2, an automated Test Generator is developed. The automated generation of test scenarios involves programmatically creating and combining the elements that make up a suitable scenario. The complexity in this process comes from the size of the scenario parameter space in terms of both variance within individual parameters and the number of parameters, or dimensions, of the parameter space – “the curse of dimensionality” [15]. The optimal selection of scenario parameters will be informed by the results of previous tests, provided by the Test Oracle and this optimisation is a rich area of research. While the VeriCAV consortium partners are exploring this area in the project, having the remaining elements of the test framework in place is critical to enabling future work on search space optimisation to move from the theoretical to the practical.

The output of the test generator is used by the Simulation Tool to execute the requested scenario-based test. To address Challenge 1 and maximise re-usability of interfaces, the scenario is described using the OpenSCENARIO [16] scenario description language (SDL). The OpenSCENARIO standardisation process undertaken by the Association for Standardization of Automation and Measuring Systems (ASAM), which Connected Place Catapult are involved in, is expected to result in wide use of the format.

B. Simulation Tool

Many commercial and open-source simulation tools are available for virtual ADS testing [17]. The VeriCAV project is using the Prescan commercial simulation tool from Siemens [18] to test the framework concept. The open-source simulation tool CARLA [19] will also be used to assess the feasibility of interchanging the simulation tool in the test framework. The

¹<https://vericav-project.co.uk/>

project requirements for the simulation tool have been defined by industrial partners and allow Challenge 3 to be addressed objectively. These requirements consider visual and sensorial realism of the environment, physics models for the HAV and compatibility with commonly used standards. The ongoing and rapid development of simulation tools has been closely followed and the ability to upgrade the Simulation Tool to take advantage of this is key to the success of the test framework.

The interface between the Simulation Tool(s) and the ADS being tested is crucial in addressing Challenge 4. Publicly available ADSs such as Baidu Apollo [20] and Autoware [21] provide examples to test ADS interfacing in the absence of public information on commercial ADSs. Baidu Apollo is used in the VeriCAV test framework and involvement in interface standardisation working groups such as ASAM Open Simulation Interface (OSI) [22] will inform the interface development and ensure the framework is appropriate for state-of-the-art ADSs.

C. Smart Actors

The increase in realism of virtual simulation tools in recent years, with realistic *looking* models, has not been reflected in the *behaviour* of the actors occupying the virtual scenes. This is key in addressing Challenge 3. Actors such as pedestrians, cyclists or other vehicles have historically been controlled by relatively simple single-aspect models such as vehicle following or by pre-defined trajectories. The behaviour of an ADS being tested should not be known a-priori and this presents a challenge when the ADS interacts with other actors. An actor should react realistically to the range of possible behaviours the ADS may exhibit, something not achievable with basic actor models. A ‘Smart Actor’ can be any actor in a scenario that demonstrates behaviour which is more complex and nuanced than simple trajectory following or random movement [5]. A goal of VeriCAV is to have smart actors which demonstrate human-like behaviour but can also be constrained to a degree such that the test scenario maintains its integrity as expected by the Test Generator and Oracle.

Testing ADS decision making and path planning is the focus for the VeriCAV project, so the Smart Actor development is focused primarily on the behavioural realism rather than visual. This development includes data-driven models such as those that use machine-learning trained by real-world driving data [23] as well as top-down cognitive [24] and game theoretical models [25], [26]. Consortium partner, University of Leeds, are building on their experience in this field and the VeriCAV project will investigate assessment methods to validate the Smart Actor models such as [27]. To balance testing efficiency, realism, and validity of a given scenario, not all actors need to be equally complex - actors that are not scenario-critical or background traffic can be implemented with less processing overhead. A critical design decision being investigated is whether Smart Actor models are to be deterministic or probabilistic. A benefit of a deterministic model is that the same scenario can be repeated, and the Smart Actor behaviour will be the same, which is useful for

testing. However, a probabilistic model allows the variance in real human behaviour to be represented. Smart Actors that can utilise either approach are likely needed to fulfil the requirements of scenario-based testing.

D. Test Oracle

Automating test generation enables many complex scenarios to be tested in quick succession. The VeriCAV test framework advances this with smart actors which increase the complexity of scenarios. This makes assessing test success a challenge which consortium partner HORIBA MIRA are tackling with the development of a Test Oracle.

The automated analysis of test results is achieved by implementing a range of objective metrics by which test scenario performance is judged. These objective metrics are fed by data provided by the Simulation Master at the conclusion of each test scenario. Initially this work will focus on implementing existing metrics in literature such as the Responsibility-Sensitivity Safety (RSS) model [28] in an automated manner and subsequently developing further metrics focused on different aspects of ADS performance [29], [30]. This will also expose some of the practical challenges of extracting the required information from a virtual simulation and its automated processing in the Test Oracle. The Test Oracle is also used to inform subsequent test generation. While OpenSCENARIO is being used for the project SDL, it currently lacks a method for describing the criteria for success in a given scenario. Formal SDLs that do include criteria for verifying success have been presented [14] with conversion to OpenSCENARIO for executing and the Test Oracle development will be informed by the state-of-the-art. The definition of success criteria, the format in which the criteria are described and the method by which success is assessed are all areas of research by the VeriCAV consortium with the test framework used to provide practical, experimental results and inform further research.

Automating the analysis of single test scenarios is critical, however it is also critical to assess how each test scenario result contributes to the overall assessment of the ADS performance, a concept referred to as coverage and highlighted in Challenge 5. Running many test scenarios that only vary a small amount doesn’t provide useful evidence of ADS performance away from the narrow subset of scenarios tested but as noted previously the parameter space is simply too large to test every possible combination, therefore a reliable assessment of coverage is critical. Using existing literature in the area of coverage based testing [31], [32] and capitalising on novel methods from other fields, the VeriCAV project aims contribute to this important field.

IV. SUMMARY

In order to progress virtual verification as a pillar of a structured ADS testing programme, a framework of robust, modular interfaces needs to be in place. The framework presented here enables this within an end-to-end system representative of industry ADS testing, while also enabling solutions to the particular practical challenges of virtual verification to be

explored. The significant areas of research are discussed and the VeriCAV approach to modular interfacing, test generation and assessment and smart actors has been outlined. The pragmatic approach taken in the project acknowledges that significant developments in Simulation Tools and ADSs will occur within and beyond the lifetime of the project. However, by developing the framework and sub-systems in a modular manner, future developments in virtual verification methods can be incorporated into the test framework. By balancing the short term objectives of demonstrating an end-to-end test system with the longer term objectives of creating a flexible, modular framework, the VeriCAV project aims to advance virtual verification and support the deployment of HAVs. The authors believe that scenario based testing in virtual simulation will play a vital role in ensuring HAV safety. Engagement with the wider safety community to ensure the test framework outlined here aligns with research, development and demonstration of verification techniques is a key activity of the VeriCAV project.

ACKNOWLEDGMENT

This work is undertaken as part of the VeriCAV project which is part-funded Innovate UK through the Centre for Connected and Autonomous Vehicles (CCAV). The authors wish to acknowledge the valued technical input and collaborative approach of VeriCAV project partners HORIBA MIRA, the University of Leeds and Aimsun.

REFERENCES

- [1] L. Fraade-Blanar, M. Blumenthal, J. Anderson, and N. Kalra, *Measuring Automated Vehicle Safety: Forging a Framework*. RAND Corporation, 2018. [Online]. Available: https://www.rand.org/content/dam/rand/pubs/research_reports/RR2600/RR2662/RAND_RR2662.pdf
- [2] Waymo LLC, "On the road to fully self-driving," Tech. Rep., 2017.
- [3] United Nations Economic Commission for Europe World Forum for Harmonization of Vehicle Regulations, "New Assessment/Test Method for Automated Driving (NATM)," 2020.
- [4] Uber Advanced Technologies Group, "A principled approach to safety," Tech. Rep., 2018. [Online]. Available: <https://uber.app.box.com/v/UberATGSafetyReport>
- [5] Z. Saigol and A. Peters, "Verifying automated driving systems in simulation: framework and challenges," in *25th ITS World Congress*, 2018. [Online]. Available: <http://www.zeynsaigol.com/ITS2018VerifyingADSinSimulation.pdf>
- [6] R. Hillman, "Test methods for interrogating autonomous vehicle behaviour—findings from the humandrive project," 2019. [Online]. Available: <https://humandrive.co.uk/wp-content/uploads/2020/03/HumanDrive-Test-Methods-for-Interrogating-Autonomous-Vehicle-Behaviour.pdf>
- [7] T. Menzel, G. Bagschik, L. Isensee, A. Schomburg, and M. Maurer, "From functional to logical scenarios: Detailing a keyword-based scenario description for execution in a simulation environment," 2019. [Online]. Available: <http://arxiv.org/abs/1905.03989v1>
- [8] J. Norden, M. O'Kelly, and A. Sinha, "Efficient black-box assessment of autonomous vehicle safety," 2019. [Online]. Available: <https://arxiv.org/abs/1912.03618v1>
- [9] N. Kalra and S. M. Paddock, "Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability?" *Transportation Research Part A: Policy and Practice*, vol. 94, pp. 182–193, 2016.
- [10] C. Robert, J. Guiochet, and H. Waeselynck, "Testing a non-deterministic robot in simulation - How many repeated runs ?" in *The fourth IEEE International Conference on Robotic Computing (IRC 2020)*, Taichung, Taiwan, Mar. 2020, p. 8p. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-02444350>
- [11] International Organization for Standardization, "ISO/PAS 21448:2019 Road vehicles — Safety of the intended functionality," 2019.
- [12] Institute of Electrical and Electronics Engineers Standards Association, "P2846 - formal model for safety considerations in automated vehicle decision making," 2019. [Online]. Available: <https://standards.ieee.org/project/2846.html>
- [13] International Organization for Standardization, "ISO/WD 34502 Road vehicles — Engineering framework and process of scenario-based safety evaluation." [Online]. Available: <https://www.iso.org/standard/78951.html>
- [14] W. Damm, E. Möhlmann, and A. Rakow, "Traffic sequence charts for the ENABLE-s3 test architecture," in *Validation and Verification of Automated Systems*. Springer International Publishing, 2019, pp. 45–60.
- [15] R. Bellman, *Dynamic Programming*. Princeton University Press, 1957.
- [16] Association for Standardization of Automation and Measuring Systems, "OpenSCENARIO," 2020. [Online]. Available: <https://www.asam.net/standards/detail/openscenario/>
- [17] Y. Kang, H. Yin, and C. Berger, "Test your self-driving algorithm: An overview of publicly available driving datasets and virtual testing environments," *IEEE Transactions on Intelligent Vehicles*, vol. 4, no. 2, pp. 171–185, 2019.
- [18] Siemens, "Simcenter prescan," 2020. [Online]. Available: <https://www.plm.automation.siemens.com/global/en/products/simcenter/prescan.html>
- [19] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: an open urban driving simulator," 2017. [Online]. Available: <http://arxiv.org/abs/1711.03938v1>
- [20] Baidu, "Baidu apollo open platform," 2019. [Online]. Available: <https://apollo.auto/>
- [21] S. Kato, E. Takeuchi, Y. Ishiguro, Y. Ninomiya, K. Takeda, and T. Hamada, "An open approach to autonomous vehicles," *IEEE Micro*, vol. 35, no. 6, pp. 60–68, 2015.
- [22] Association for Standardization of Automation and Measuring Systems, "Open simulation interface," 2020. [Online]. Available: <https://www.asam.net/standards/detail/osi/>
- [23] F. Behbahani, K. Shiarlis, X. Chen, V. Kurin, S. Kasewa, C. Stirbu, J. Gomes, S. Paul, F. A. Oliehoek, J. Messias, and S. Whiteson, "Learning from demonstration in the wild," in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 775–781. [Online]. Available: <https://www.fransoliehoek.net/docs/Behbahani19ICRA.pdf>
- [24] C. C. Macadam, "Understanding and modeling the human driver," *Vehicle System Dynamics*, vol. 40, no. 1-3, pp. 101–134, 2003.
- [25] F. Camara, R. Romano, G. Markkula, R. Madigan, N. Merat, and C. Fox, "Empirical game theory of pedestrian interaction for autonomous vehicles," in *Measuring Behavior 2018: 11th International Conference on Methods and Techniques in Behavioral Research*, R. Grant, T. Allen, A. Spink, and M. Sullivan, Eds. Manchester Metropolitan University, 2018, pp. 238–244. [Online]. Available: <http://eprints.whiterose.ac.uk/129303/>
- [26] K. Kang and H. A. Rakha, "A repeated game freeway lane changing model," *Sensors*, vol. 20, no. 6, p. 1554, 2020.
- [27] J. Bernhard, K. Esterle, P. Hart, and T. Kessler, "Bark: Open behavior benchmarking in multi-agent environments." [Online]. Available: <http://arxiv.org/abs/2003.02604v1>
- [28] S. Shalev-Shwartz, S. Shammah, and A. Shashua, "On a formal model of safe and scalable self-driving cars," 2017. [Online]. Available: <https://arxiv.org/abs/1708.06374v6>
- [29] R. Myers and Z. Saigol, "Pass-fail criteria for scenario-based testing of automated driving systems," 2020. [Online]. Available: <http://arxiv.org/abs/2005.09417v1>
- [30] B. Klamann, M. Lippert, C. Amersbach, and H. Winner, "Defining pass-fail-criteria for particular tests of automated driving functions," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*. IEEE, 2019.
- [31] Z. Tahir and R. Alexander, "Coverage based testing for v&v and safety assurance of self-driving autonomous vehicle : A systematic literature review," in *The Second IEEE International Conference On Artificial Intelligence Testing*, 1 2020. [Online]. Available: http://eprints.whiterose.ac.uk/155680/16/camera_ready.pdf
- [32] R. B. Abdessalem, A. Panichella, S. Nejati, L. C. Briand, and T. Stifter, "Testing autonomous cars for feature interaction failures using many-objective search," in *Proceedings of the 33rd ACM/IEEE International Conference on Automated Software Engineering - ASE 2018*. ACM Press, 2018.