Never feed it after midnight[1] – Testing un-intended consequences in simulation

Nicholas Moylan
XPI Simulation
Crawley, UK
Nick.Moylan@xpisimulation.com

ABSTRACT

Developing and deploying complex systems can be highly challenging, particularly with the issue of un-intended or emergent behaviors. Existing test methods are well adapted to low level test, demonstrating the performance of components in their un-integrated state, or high level demonstrations of integrated systems, but which lack the depth of test to find and fix un-intended emergent behaviors, particularly where those may be introduced by influences outside the system of interest. This is particularly important where the safe operation of the system matters to those whose mission matters.

One such complex system is Autonomous Vehicles (AV). This paper will seek to discuss the issues of identifying the impact of emergent behaviors, and approaches to verifying safe performance of complex systems in the real world, by testing in the synthetic environment. This paper will build upon the findings of the OmniCAV programme in verifying the safe behavior of an autonomous vehicle and draw parallels with other complex systems used in defense.

OmniCAV is a consortium project that is partly funded by the UK Government. It aims to deliver a highly realistic simulation environment for AV stack verification that considers all road users and road types. The programme has developed a unique synthetic environment integrating a high fidelity survey of urban and rural Oxfordshire, with other simulated road users allowing all conditions to be simulated and evaluated. This is being delivered in combination with real-world testing to support the assurance of the safety case for deploying a self-driving car. As well as XPI's work on OmniCAV, potential approaches for integrating multi-part simulators for generating coherent and rich digital twins associated with such autonomous systems testing will be addressed.

In concluding, some of the key technical challenges that remain in this domain will be identified, including interpretation of results and assuring completeness of testing.

ABOUT THE AUTHORS

Nicholas Moylan is a Product Specialist with XPI Simulation, leading the technical strategy and product line management of XPI's autonomous vehicle and driving simulation business. Nicholas has over 10 years' experience in airborne and maritime ISR (Intelligence, Surveillance and Reconnaissance) systems. Recently he has led research into intelligence and mission autonomy to improve survivability and situational awareness with Thales UK. Nicholas has also served in the British Army Reserves since 2015.

[1] In the 1984 film "Gremlins", the seemingly innocuous act of feeding small furry creatures called Mogwai after midnight had the un-intended consequence of turning otherwise cute creatures into murderous monsters. https://www.imdb.com/title/tt0087363/

Never feed it after midnight – Testing un-intended consequences in simulation

Nicholas Moylan
XPI Simulation
Crawley, UK
Nick.Moylan@xpisimulation.com

INTRODUCTION

The traditional approach to system testing and assuring safety have used a range of validation and verification approaches, generally, testing components and sub-systems against known interactions and well-controlled abstractions of the real world according to their specifications, for example using MIL-STD-461 for EMC Testing. However, this approach does little to re-create interactions with complex operating environments.

Systems deployed into complex environments have instead historically relied upon humans-in-the-loop as an effective feedback loop, controlling the system in a reasonable and for the most part safe way. With increasing complexity of systems, humans are relying more upon automation to both hide the complexity and take over some of the identification and management of system behavior. A critical example of this is Connected and Autonomous Vehicles (CAV) which are capable of operation without a human in the loop and hence responsibility for functional safety has shifted from the operator to the system.

The ICAO Accident Prevention Manual (ICAO, 1984) suggests that over time, there has been a trend for human error to be the primary cause of accidents versus machine causes (Figure 1). This leads to the intuitive conclusion that safety can be improved by reducing the level of human factors, by increasing the level of automation in the system. The World Health Organisation (WHO) 2018 Report on Road Safety indicates that there are 1.35 Million deaths on the roads each year (WHO, 2018), 90% of which is attributed to human error (Singh, 2015), it is clear that Automated Driving Systems (ADS) have the potential to dramatically reduce the level of fatal accidents on the roads.

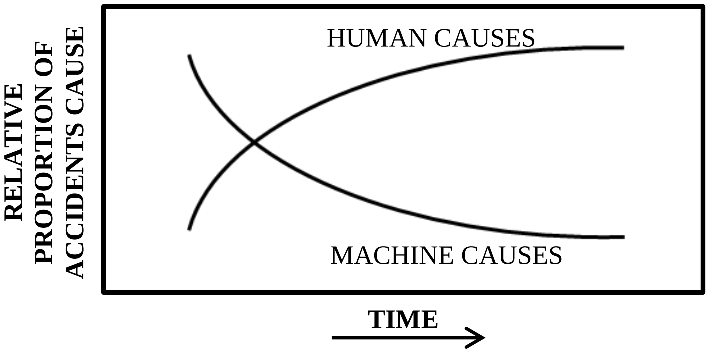


Figure 1. Changes in accident causal factors over time (ICAO, 1984)

With CAVs operating within a system of systems (SoS) environment, the impact and incidence of emergence is of concern for system designers and those responsible for assuring system safety. In particular, the possibility that the accident rate from machine causes increases beyond those previously seen from human causes.

Emergent behavior is functionality or characteristics of a system that does not depend on its individual parts, but on their relationships to one another within that system and other systems. Thus emergent behavior cannot be predicted by examination of a system's individual parts in the traditional way and usually only manifests once the system is in its operating state. A complex system such as a CAV operating on the roads with other road users is likely to exhibit strong emergent behavior.

CAVs will not be able to rely as heavily upon human monitoring and control of the system and will therefore require innovative new ways of robustly assuring safe operation of the system. One of these ways is by improving the quality of testing, but they cannot be fully tested using physical methods with any time or cost efficiency or with due regard to public road safety. Moving the testing from the physical world to a virtual one requires the interactions between the system of interest and the other systems in the operating environment, to create the conditions for emergent behavior, to be modelled and tested.

To meet these challenges, the UK Government funded the OmniCAV research project with the aim to help test these complex systems, and is focused on addressing key gaps in the current state-of-the-art to deliver a simulation platform with digital twin technology capable of testing the wide range of scenarios and interactions required for identifying un-intended behaviors.

This paper will explore the challenge of autonomous systems testing in synthetic environments using a digital twin to find and fix un-intended emergent behaviors through two lenses. The first will focus on activity carried out by XPI Simulation and the OmniCAV consortium on the OmniCAV project, which considers the automotive domain. The second will consider the defense domain in general, with a maritime autonomous system (MAS) test facility providing a particular use case. While highlighting considerations for the functional safety assurance, the role of regulators and standards and how these can enable assurance of increasing system autonomy. Outstanding technical challenges and areas for future research will also be elaborated upon.

SAFETY PROBLEM SPACE

Complex Systems And Historic Approach To Safety

Traditional system safety approaches have typically used risk based mitigations to certify that the system will not enter an unsafe state, through normal use, failure or abuse. This certification is typically supported by a range of standards based design and testing, which provides a level of assurance that the risk to life and property are above a minimum acceptable safety threshold and are As Low As Reasonably Practicable (ALARP).

For road vehicles, ISO 26262 is the major standard for safety related systems with electrical or electronic systems, based around a risk based system of assuring safety in the event of unit faults or system malfunction (ISO, 2018). As the approach is system centric, it does not take into account any significant interactions between the system of interest and the environment it operates within, therefore, it is unlikely to identify any safety issues associated with emergent behaviors of the vehicle under test.

A similar approach is taken by the aerospace and defense industry, for example the UK Design and Airworthiness Requirements for Military Aircraft, Def Stan 00-970 uses a series of standard tests to verify aircraft safety within the required operating envelope to support a safety case approach to airworthiness certification (MOD, 2019).

System Of Systems And The Need For A New Approach

As systems have become increasingly complex, automated and more integrated with other systems outside the design boundary, the likelihood and impact of emergent behavior has increased, and hence the coverage and effectiveness of isolated system testing to identify all the major risks to system safety has been reduced.

With the development of increasingly complex Advanced Driver Assistance Systems (ADAS), the focus is shifted to Safety of the Intended Function (SOTIF), which is concerned with the functionality in the absence of a fault, or as a result of emergent behaviors. This is in contrast with the traditional approach to system safety, which is concerned with managing the risk due to a system failure. ISO 21488 provides guidance to assuring the SOTIF where proper situational awareness is critical to safety, and where that situational awareness is derived from complex sensors and processing algorithms; especially emergency intervention systems (e.g. emergency braking systems) and ADAS (ISO, 2019).

The RAND Corporation has published research stating that, in order to demonstrate the same level of safety as human drivers – i.e. a fatality rate of 1.09 deaths per 100 million miles – that the vehicles would have to be driven 8.8 billion miles (assuming a 95% confidence level and a precision of 20%) (Kalra and Paddock, 2016). This seemingly unassailable mileage is compounded by the reality that in order to identify the emergent behaviors which might affect system safety, the testing must be representative of the operating environment. Billions of miles carried out exclusively on race tracks or straight highways will do nothing to demonstrate system safety during city driving.

To provide a significantly more robust assurance of complex system safety, the nature of testing has evolved to focus more upon system interactions in a SoS, and also upon a scenario based understanding of the system performance (NASA, 2010) in order to:

- Identify the most critical scenarios that can lead to the undesired consequences.
- Identify the risk drivers that cause these scenarios to be critical.
- Ensure that the controls are directed toward the risk drivers.

EMERGENT BEHAVIOR

Emergent behavior is functionality or characteristics of a system that does not depend on its individual parts, but on their relationships to one another while in a functional state, it does not occur as a result of any component failures. Emergent behavior is not always undesirable, in fact, good system engineering maximizes opportunity by understanding and exploiting emergence in engineered systems to create the required system level characteristics from synergistic interactions between the components, not just from the components themselves (Sillitto, 2010).

(Page, 2009) describes three types of emergence, simple, weak and strong:

- **Simple emergence** is generated by the combination of element properties and relationships and occurs in non-complex systems, for example, controlled flight cannot be attributed solely to the wings of an aircraft, it is as a result of the interaction between the wings, the engine and control system. Page suggests that simple emergence is the only type of emergence that can be predicted.
- **Weak emergence** is expected emergence which is desired (or at least allowed for) in the system structure. However, as a product of a complex system, it cannot be predicted just from knowledge of the characteristics of the individual system components. Taking the example of traffic jams: even if no vehicles have failed and no impediment to traffic flow exists, the overall system can still exhibit chaotic behavior.
- **Strong emergence** is unexpected emergence; behaviors are not observed until the system is simulated or tested or, more alarmingly, until the system encounters in operation a situation that was not anticipated during design and development. This is particularly likely where strong system interactions occur across a system boundary as shown in Figure 2.

A complex system interacts both with its own component parts, to produce simple and weak emergence, but also with the environment and actors it operates within as a SoS to produce strong emergence. Therefore, even for well understood systems, interactions outside the design system boundary are the significant source of emergence (Figure 2). It is therefore these interactions with the wider SoS environment which must be modelled accurately in order to understand strong emergent behaviors likely to occur.

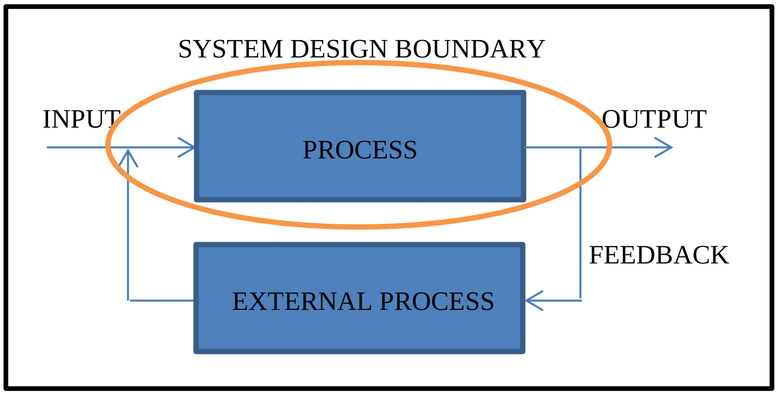


Figure 2. Simple System of Systems Representation

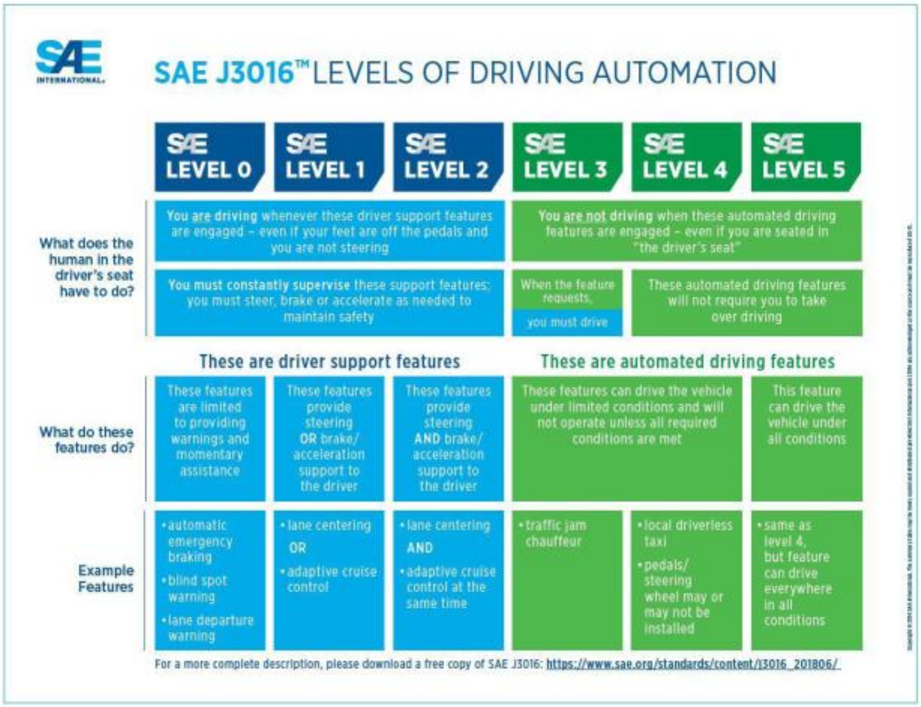
For human in the loop systems, the human operator acts in the control loop, managing the effects of strong emergence, potentially mitigating the risks to life. However, for autonomous systems, where the human is no longer in the loop, the likelihood of undesirable behavior going unchecked is commensurately higher.

Illustrating the potential impact of emergence in safety related system: human drivers as a population, are likely to behave differently to one another, when subjected to the same conditions reducing the overall effect of the emergent condition. Autonomous systems on the other hand will behave the same across a portion of the population. Therefore, if the behavior leads to an accident the entire population of autonomous systems are perceptible to any undesirable emergent behaviors.

As a means of predicting the likelihood or impact of emergent behavior in complex systems, (Mogul, 2006) proposes widening the system of interest beyond the system design boundary in order to capture the effects of SoS interactions. An approach supported by Cummins, is through use of scenario based testing in a modelling and simulation environment with one simulator representing the System of Interest, and one simulator representing the wider environment and agents, operating together to provide a SoS, in order to allow the functionality to be observed and measured (Cummings, 2015).

Limiting Emergent Behaviors

The final approach to limiting the likelihood and effect of emergent behavior in complex systems, is by limiting the Operating Design Domain (ODD) that they operate within. This approach is being used during CAV development, the SAE J3016^{TM} level 0-4 autonomy limit the conditions under which the CAV may be operated (SAE, 2019), see Figure 3. This approach significantly reduces the complexity of the SoS outside the system design boundary, though if level 5 autonomy is to be achieved a CAV could be reasonably expected to safely drive from London to Singapore, interacting with other systems which it may never have encountered previously.



SAE J3016™ LEVELS OF DRIVING AUTOMATION

	SAE LEVEL 0	SAE LEVEL 1	SAE LEVEL 2	SAE LEVEL 3	SAE LEVEL 4	SAE LEVEL 5
What does the human in the driver's seat have to do?	**You are driving** whenever these driver support features are engaged – even if your feet are off the pedals and you are not steering			**You are not driving** when these automated driving features are engaged – even if you are seated in "the driver's seat"		
	You must constantly supervise these support features; you must steer, brake or accelerate as needed to maintain safety			When the feature requests, you must drive	These automated driving features will not require you to take over driving	
	These are driver support features			**These are automated driving features**		
What do these features do?	These features are limited to providing warnings and momentary assistance	These features provide steering **OR** brake/ acceleration support to the driver	These features provide steering **AND** brake/ acceleration support to the driver	These features can drive the vehicle under limited conditions and will not operate unless all required conditions are met		This feature can drive the vehicle under all conditions
Example Features	• automatic emergency braking • blind spot warning • lane departure warning	• lane centering **OR** • adaptive cruise control	• lane centering **AND** • adaptive cruise control at the same time	• traffic jam chauffeur	• local driverless taxi • pedals/ steering wheel may or may not be installed	• same as level 4, but feature can drive everywhere in all conditions

For a more complete description, please download a free copy of SAE J3016: https://www.sae.org/standards/content/J3016_201806/

Figure 3. Levels of Driving Automation (SAE, 2019)

Understanding the ODD will guide the identification of system requirements for safe operation of the CAV. It also is key to understanding the test scenarios to be used during safety evaluation. For example, for a low-speed shuttle application in a city center, the ODD definition and testing scenarios would include attributes like type of urban roadways, speed range, environmental conditions, types of other road users and so on (Khastgir, 2020).

OMNICAV

OmniCAV[2] is a UK government funded cross-industry collaboration to create a virtual simulation and real-world testing and certification environment to accelerate the global development of CAVs. The "digital twin" represents a 32 km road-loop in Oxfordshire, which includes urban, peri-urban, and rural roads and a varied range of road layouts.

[2] https://omnicav.com – OmniCAV stands for Connected and Autonomous Vehicles for All.

The OmniCAV programme builds upon previous work, including Certification of Automated Vehicles in Synthetic Environments (CAVinSE), which was reported on in the 2019 I/ITSEC papers (Coley, Fulker and McConnachie, 2019). The OmniCAV programme differs slightly from CAVinSE, in that CAVinSE focused upon certification and standards, OmniCAV focusses upon a real-world validated simulation-based framework for certifying autonomous vehicles.

OmniCAV addresses the need for a holistic simulation system through delivering a Proof-of Concept (PoC) of an integrated virtual and physical testing regime, covering rural and urban roads, and considering needs of key end-users. By 2021 the project will deliver a PoC testing regime, together with a viable virtual and physical platform for CAV testing validated against end-user need based on the following four stages (Figure 4):

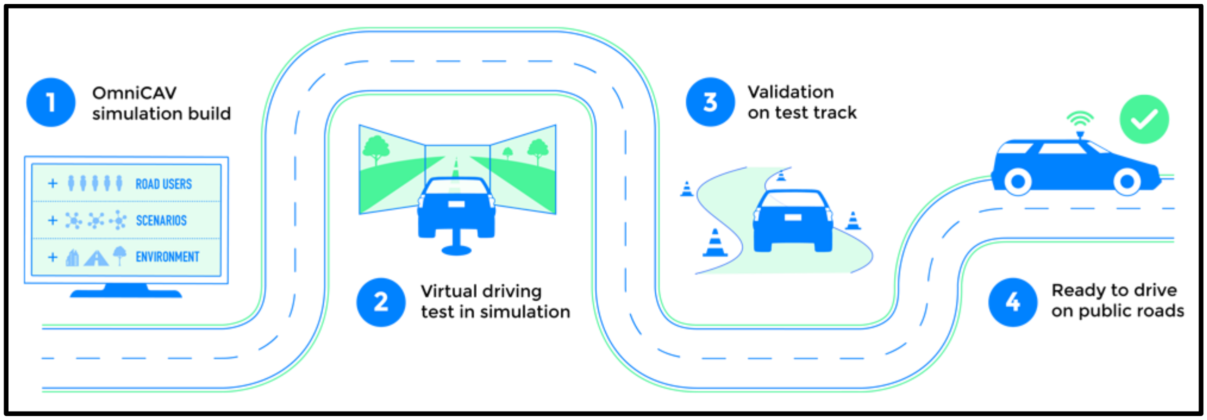

1 OmniCAV simulation build
+ ROAD USERS
+ SCENARIOS
+ ENVIRONMENT
2 Virtual driving test in simulation
3 Validation on test track
4 Ready to drive on public roads

Figure 4. OmniCAV Project Test Methodology

- Stage 1: Data collection - Collecting environment and agent data to produce a high fidelity digital twin of a combined rural/urban road loop in Oxfordshire in the U.K. (No simulation today covers rural as well as urban roads, despite these representing 93% of the U.K. and posing specific challenges (e.g. narrow lanes, lack of data, inadequate GPS coverage)). Extent and features of the test bed are shown in Figure 6.
- Stage 2: Scenario generation - Combining three hazard identification methodologies to develop a scenario library, scenario definition language and evaluation plan: addresses state-of-the-art gap on scenario generation.
- Stage 3: Novel Comprehensive simulation platform - Integrating these into a simulation platform, encompassing a traffic simulator to control traffic, a driving simulator for close-in interactions with other road users and multiple accurately modelled sensors: address state-of-the-art gap on modelling and simulation levels.
- Stage 4: Testing strategy - Validating simulated tests against results in a physical environment and completion of a proof-of-concept testing regime involving simulation, testing ground and on-road: addresses the objective of building confidence in CAV decision making and accelerating CAV development.

Building A Digital Twin Of Oxfordshire

The digital twin used in the OmniCAV programme required an accurate working model of the Oxfordshire test bed shown in Figure 6, using data from several sources. Firstly, an accurate 3D model of the route was built by Ordnance Survey, the UK national geospatial mapping agency, using a range of data sources, including LIDAR scans (Figure 5). Dynamic elements to the model such as traffic lights incorporated data provided by Oxfordshire Council and GPS signals accurately modelled for noise and signal degradation, to enable the model to mirror the form and function of the real world.

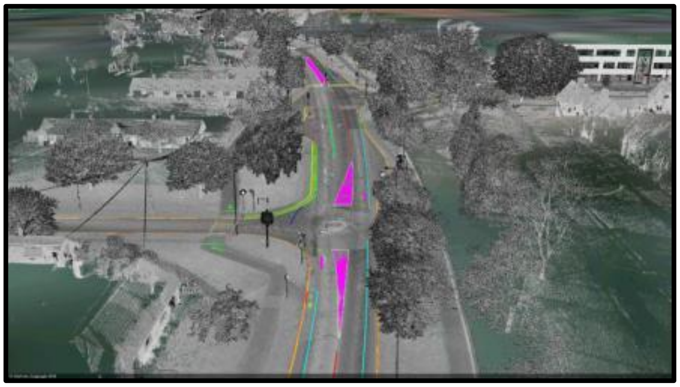


Figure 5. Extracting information from a point cloud

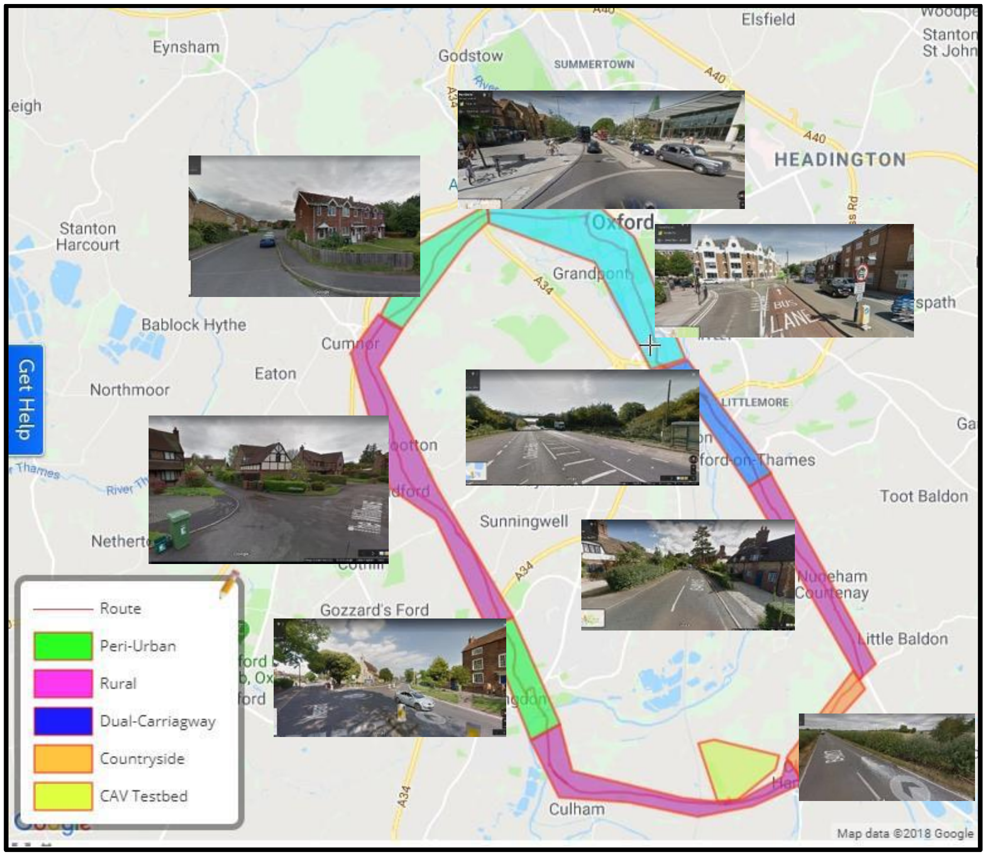


Figure 6. Test Bed Location

The system under test, the CAV is modelled using data provided from the vehicle manufacturer, Arrival and incorporating as much of the real system software and hardware in the loop as possible to minimize behavioral artifacts induced during testing as a result of simulation only configurations.

During testing, the digital twin is populated with other systems, which will interact with the CAV, including environmental and weather effects and other road users. This SoS model sets the conditions for testing the interactions between the CAV and systems outside the design boundary, which are required for strong emergent behaviors to be tested.

Simulator Architecture

In order to support a PoC as well as position for future (post project) development, it is critical to define an architecture that is modular and employs, where available, standard interfaces between core components. Given the complexity, immaturity and in some cases, disparity of the current CAV eco-system between nations and amongst manufacturers, this is not straightforward. This is further compounded by having no current common interface across sensors, requiring development to be based around specific supplier products and future modifications for every new sensor to be simulated. The OmniCAV high-level architecture is shown in Figure 7.

As can be seen in Figure 7, the CAV node is separated from its sensors and actuators, as these are physical devices which cannot operate within a virtual environment. Instead, the sensors and actuators are modelled within the simulator engine and validated separately to ensure they match the characteristics of the real subsystems as closely as possible. Where possible, realistic CAV to simulator interfaces (such as GMSL and CAN) are used to stimulate the heart of the CAV to test its detection and decision making algorithms under normal and edge case conditions without further change within the CAV system design boundary.

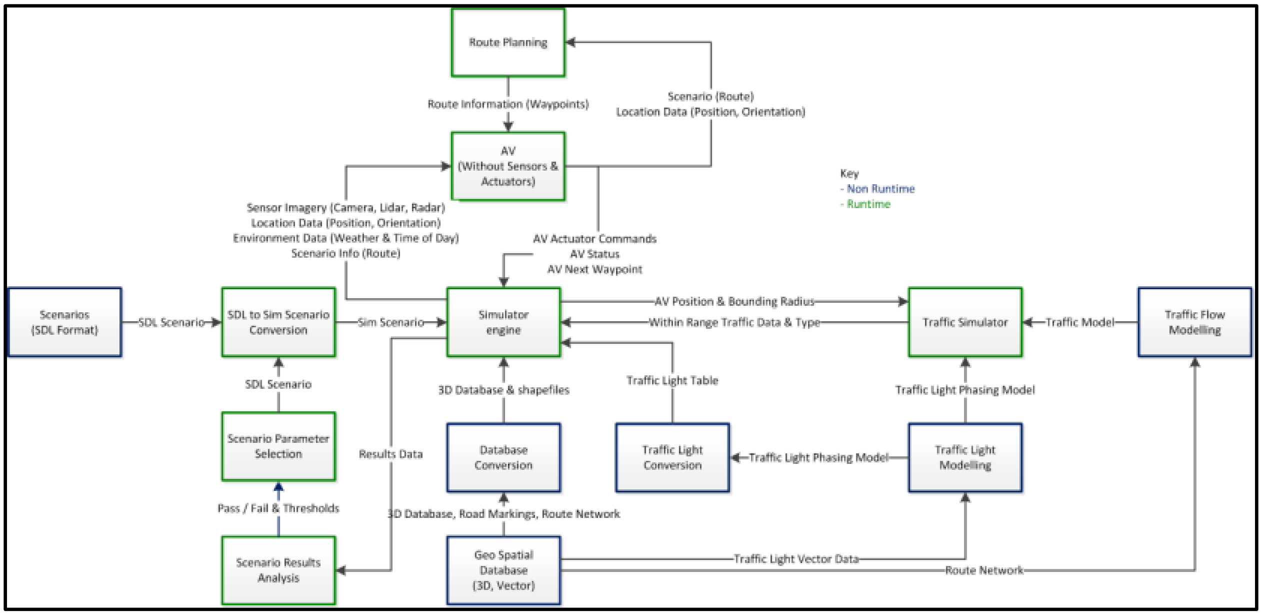

Route Planning
Route Information (Waypoints)
Scenario (Route)
Location Data (Position, Orientation)
AV
(Without Sensors & Actuators)
Key
- Non Runtime
- Runtime
Sensor Imagery (Camera, Lidar, Radar)
Location Data (Position, Orientation)
Environment Data (Weather & Time of Day)
Scenario Info (Route)
AV Actuator Commands
AV Status
AV Next Waypoint
Scenarios
(SDL Format)
SDL Scenario
SDL to Sim Scenario Conversion
Sim Scenario
Simulator engine
AV Position & Bounding Radius
Within Range Traffic Data & Type
Traffic Simulator
Traffic Model
Traffic Flow Modelling
SDL Scenario
3D Database & shapefiles
Traffic Light Table
Traffic Light Phasing Model
Scenario Parameter Selection
Results Data
Database Conversion
Traffic Light Conversion
Traffic Light Phasing Model
Traffic Light Modelling
Pass / Fail & Thresholds
3D Database, Road Markings, Route Network
Scenario Results Analysis
Geo Spatial Database
(3D, Vector)
Traffic Light Vector Data
Route Network

Figure 7. OmniCAV Simulation Schematic Diagram

Getting The Testing Right

As suggested previously, to prove CAVs are at least as safe as human-driven vehicles, requires over 8.8 billion road test miles (Karla and Paddock, 2016), however, not all road miles are critical to safety. To this end, Hazard Based Testing (HBT) has been proposed which focusses on identifying "how a system fails" as compared to "how a system works" (Khastgir et al, 2018). Based on the HBT concept, OmniCAV uses a novel hybrid approach to create test scenarios focusing on identifying emergent behaviors and failures, with three sources of scenario data:

- OmniCAV uses accident database analysis to identify accident hotspots and parameters which contribute to causation of accidents (with varying levels of severity). Stats19 and RAIDS accident databases, which are both UK based databases are used for this purpose.
- Anonymized vehicle insurance claim records collected by one of the partners of OmniCAV (Admiral), is used to identify trends in the situations that lead to insurance claims.
- A Systems Theoretic Process Analysis (STPA) (Leveson, 2012), which is a systems based method safety analysis method to identify failures was used as a foundation and its extension (Khastgir et al, 2020), was used to create test scenarios. STPA was done for both the Autonomous Control System and the Brake-by-Wire actuation system. Using the three methods of scenario generation, a scenario library is being created which is aligned with U.K.'s National CAV Test scenario Database, part of the Midlands Future Mobility[3]. STPA of higher abstraction of the ACS revealed 1190 requirements and over 3000 scenarios. STPA of BBW revealed 3736 requirements and over 5000 scenarios.

As the output of the scenario is complex, a simple pass/fail approach is not suitable. Following each test, the behavior of the CAV is confirmed to have fallen within an acceptable range, according to a set of pre-defined rules on CAV safe behavior, including "did the behavior of the CAV cause an avoidable accident". If a CAV fails to meet these rules, the results are always of interest in identifying system safety failures. For those cases where the CAV meets all the rules, the data is checked to confirm that the test scenario was valid and correctly executed and so can be added to the body of evidence of the safety case.

Research is actively on-going to define a more nuanced definition of acceptably safe CAV behavior, with the MUSICC programme proposing a 5 part criteria (Myers and Saigol, 2019) allowing safe interactions (human/CAV or CAV/CAV), fault/error tolerance and legal and considerate driving to be assessed.

[33] https://midlandsfuturemobility.co.uk

Validated Testing

OmniCAV project has two types of Proof-of-Concept (PoC) testing. First, is the testing of the ADS and second is the testing of the OmniCAV simulation platform to evaluate its ability to represent the real-world environment. While the former requires us to test first in the simulation environment and then in the real-world environment, the later calls for a reverse approach. For testing the simulation environment, we first gather data from driving the ADS in the real-world and use the gathered data to replay the scenarios in simulation to compare the results.

Once the simulator is validated and representative of the physical world, the OmniCAV project will demonstrate a PoC of how the Autonomous Driving System could be validated using both physical and virtual testing. The project will be using the Arrival Autonomous Driving System (ADS) fitted to the Arrival Van, a light commercial vehicle designed from the ground up to be capable of being driven autonomously.

Prior to real-world deployment the Arrival ADS will pass through a series of gateway tests, starting in simulation and only progressing into the physical (controlled) and physical (public road) tests once its safety has been proven at the previous gateway. The PoC will use the intensive validation capability of the simulator in scenarios specifically generated based on the vehicle architecture using the STPA of the Arrival ADS done as part of the project, accident data analysis and insurance claims data analysis.

Key Outputs And Findings

This simulation architecture is designed to be used to validate the synthetic environment against the real world to show how viable it is as a digital twin. This will demonstrate that simulation could be used as a key tool for providing CAV test evidence to support a CAV safety case. It is not expected that simulation, by itself, will provide all the evidence for CAV accreditation, but instead will be a key aspect of a mixed synthetic/on road test programme. Simulation will give highly repeatable, controlled, testing, at potentially faster than real time and critically, will be able to put the CAV in situations that would be dangerous or difficult to construct in the real world.

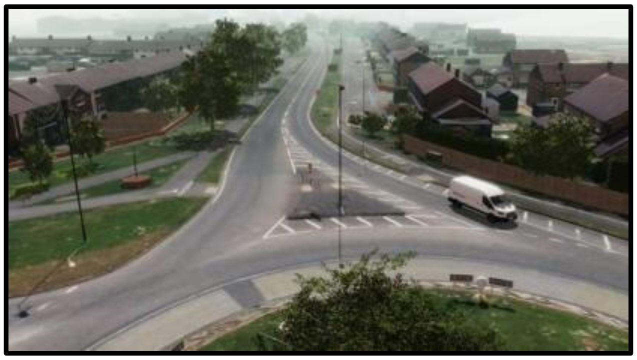


Figure 8. Example of OmniCAV Simulation

Further, through closed loop testing of the CAV and understanding of the CAV internal architecture, it is possible to identify potential stress test cases for the CAV. It is one of the aims of this project to be able to demonstrate how scenario parameter selection can be automated to determine the worst case combination for the CAV and hence critical scenarios under which strong emergent behaviors are likely to occur (Brackenstone et al, 2020).

RELATED DEFENSE APPLICATIONS

Similar complex systems are used in Defense and Aerospace particularly where those systems have strong interactions with wider SoS in their operating environment.

Plymouth Sound Case Study

Maritime Autonomous Vehicles (MAV), to a greater extent than CAVs operate in a highly complex environment, with a wide range of other entities in the operating domain. One area of interest is Plymouth Sound in the UK. Plymouth is home to the largest naval base in Western Europe, servicing surface and sub-surface vessels, is used by recreational and commercial vessels and has recently started being used as a testbed location for MAVs. The hydrography is also complex, as the Plymouth Sound connects 3 waterways, 2 freshwater inlets and one to the English Channel, see Figure 9.

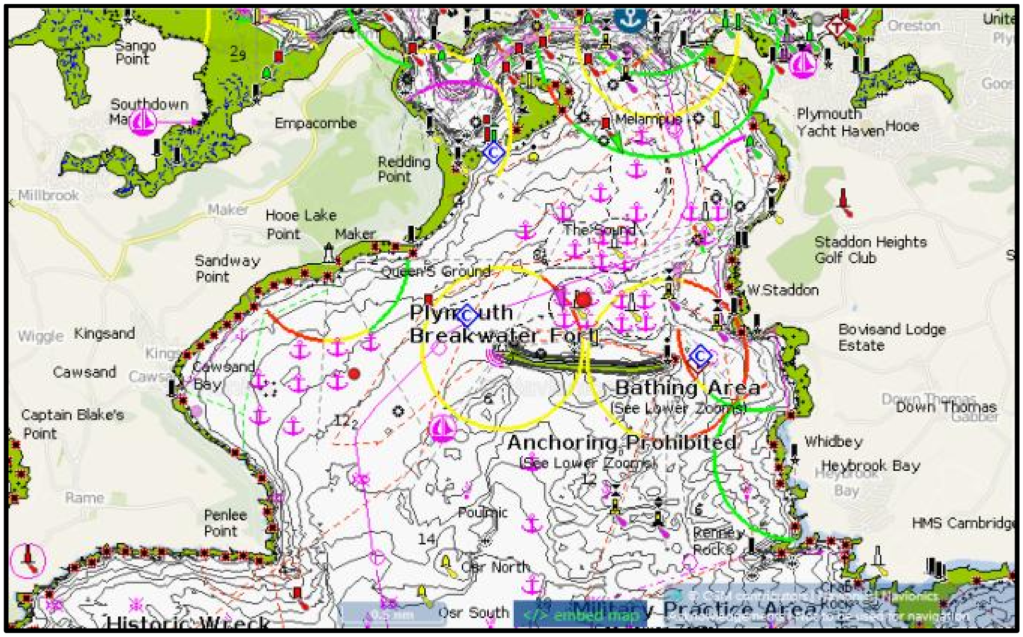


Figure 9. Complexity of Maritime Operating Environment – Plymouth Sound

Thales UK is developing a digital twin of Plymouth Sound, to enable MAV testing and vessel movement monitoring by incorporating real time data from bathymetric and meteorological sensors, vessel tracking and accurate hydrological scans of the Sound along with live data collected on and by the MAV itself. This digital twin will go further than the approach used in OmniCAV, by not only testing MAVs prior to deployment on the water, but in real time during operation, to validate decision making by the systems and ensure un-intended behaviors do not result in hazardous functionality and also to develop a better understanding of the wider effects of interactions of SoS.

ROLE OF STANDARDS AND REGULATION

Highly autonomous, complex systems are a rapidly developing landscape, and one in which both standards bodies and regulatory authorities have a central role to play. For rich digital twins to be a successful tool, open standards are required, particularly in the following areas:

- Architectures – enabling integration of a different functionality into the digital twin, such as the traffic model into the OmniCAV simulation. Open architectures are required to provide a framework which does not produce false results in testing for emergent behaviors through interactions within the simulator or between the system under test and the test environment.
- Datasets and models – such as mapping data, scenarios, sensor and environmental models, have a significant role in assuring the completeness and quality of testing, for example, the UK National CAV Test Database (WMG, 2020).
- Live Data – to allow real time update of digital twins and further understand the effect of interactions between the SoS. An excellent example of this nature of data being made open and public is Transport for London's (TfL) OpenData[4].
- Design and Test methodologies – such as the BSI PAS 1883 (BSI, 2020) for specifying and constraining the ODD.

The role of the regulator is different from the standards bodies, in that they may take a more directive approach, requiring CAV designers to adopt certain standards and methodologies in such as the CAVPASS process published by the UK Centre for Connected and Autonomous Vehicles (CCAV) (CCAV, 2019) and defining what is an acceptable level of risk for systems, based upon based safe performance criteria, such as that proposed by MUSICC (Myers and Saigol, 2019).

[4] https://tfl.gov.uk/info-for/open-data-users/our-open-data

Finally, the study commissioned by UK Department for Transport into CAV public acceptability showed participants indicated they wanted to know that CAVs would bring safety benefits and wanted to see Government regulations on safety standards and data security (McCool, 2019). Furthermore, the need to develop and communicate a set of transparent protocols, rules and regulations, driven by an independent and impartial organization (McCool, 2019), was seen as essential to obtaining public trust and acceptance of CAV.

CONCLUSIONS AND FURTHER RESEARCH

This paper set out to show how emergent behaviors present in complex autonomous systems, which have an impact upon system safety can be identified early in the development lifecycle using a digital twin. It provided an example of building a digital twin test facility to conduct scenario based testing on a self-driving car to provide assurance of safe operation within its ODD. It also showed how a similar approach could be applied to a range of other domains including aerospace and defense, where increasing levels of autonomy and system complexity are changing the requirements for determining whether a system is safe for use.

Identifying emergent behaviors and their impacts requires a good understanding of the system of interest and the systems it interacts with in its ODD. The RAND paper (Kalra and Paddock, 2016) suggested proving a CAV is safe would require billions of miles of road testing, but this paper has shown how this can be reduced by identifying appropriate critical test scenarios in order to test the corner cases under which emergent behavior or system failures are most likely to occur.

Key Technical Challenges And Limitations

As the OmniCAV programme is only a PoC, there are recognized limitations with its specific implementation, however the limitations identified here are against a hypothetical complete test facility:

- The external sensors and interfaces of a system under test will typically require a separate model when integrated within a digital twin test environment, breaking the design system boundary. These models will need to be validated separately in order to ensure that they exhibit the same behavior and uncertainties as the real sensor or interface.
- Un-identified interactions and systems within the operating design domain will reduce the effectiveness of the digital twin to identify emergent behaviors, conversely, including unnecessary systems which do not interact with the system under test, will reduce the value and performance of the simulation.
- In-ability of digital twins to exactly model continuous systems, such as environmental weather effects. The quality of the digital twin is directly proportional to the quality of the data it is produced from. Ensuring data which has come from disparate sources at different times with different levels of quality is managed appropriately is a key technical challenge of future digital twins.
- Even a single scenario is capable of having an unlimited number of permutations if controls on critical scenario selection are not appropriately applied. As CAVs near Level 5 autonomy (SAE, 2019) and an unlimited ODD, the range of scenarios can also increase exponentially, with a proportionate increase in time and cost of testing.

Next Steps And Further Research

With a global market for CAV technology testing up to £30bn per annum by 2035 (Brackenstone et al, 2020), the outputs of PoCs such as OmniCAV will require maturing through increasingly high TRL (Technology Readiness Levels) in the coming years to enable the market. Expansion of the scope, quality and quantity of testing (particularly through adding richness to the digital twin and expanding the test scenario database) will be a key enabler for a CAV market which offers significant long term benefits.

ACKNOWLEDGEMENTS

The work presented here is supported by the OmniCAV project, part-funded by the Centre for Connected and Autonomous Vehicles, delivered in partnership with Innovate UK. (Grant No. 104529). The authors will also like to thank the project partner organisations for providing the required infrastructure support for this work.

REFERENCES

Brackenstone, M et al. (2020). *OmniCAV: A Simulation and Modelling System that enables "CAVs for All"*, 23rd IEEE International Conference on Intelligent Transportation Systems: IEEE

British Standards Institute (2020). *BSI/PAS 1883:2020 Operational Design Domain (ODD) taxonomy for an automated driving system (ADS) – Specification*, BSI

Centre for Connected and Autonomous Vehicles (CCAV) (2019). *Code of Practice: Automated vehicle trialling*, UK Department for Transport

Coley, T, Fulker, D and McConnachie, R (2019). *Simulation-Based Autonomous Systems Testing – From Automotive to Defence*, I/ITSEC 2019

Cummings, MA (2015). *Identifying And Quantifying Emergent Behavior Through System Of Systems Modeling And Simulation*, Naval Postgraduate School Monterey, CA

International Civil Aviation Organization (1984). *Accident prevention manual*. Montreal, Canada: ICAO

ISO (2018). *ISO 26262:2018 Road Vehicles – Functional Safety*, ISO

ISO (2019). *ISO/PAS 21488:2019 Road vehicles — Safety of the intended functionality*, ISO

Kalra, N. and Paddock, S. (2016). *Driving To Safety: How Many Miles Would it Take to Demonstrate Autonomous Vehicle Reliability.* RAND Corporation

Khastgir, S (2020). *Enabling safe CAV deployment*, BSI

Khastgir, S, Brewerton, S, Thomas, J and Jennings, P (2020). *Systems Approach to Creating Test Scenarios for Automated Driving Systems*, Reliability Engineering and System Safety, vol. (in review).

Khastgir, S, Birrell, S, Dhadyalla, G and Jennings P (2018). *The Science of Testing: An Automotive Perspective,* SAE Technical Paper: 2018-01-1070.

Leveson, N. G. (2012). *Engineering a Safer World: Systems Thinking Applied to Safety*, MIT Press.

McCool, S (2019). *CAV Public Acceptability Dialogue: Engagement Report*, UK Department for Transport

Mogul, J (2006). *Emergent (Mis)behavior vs. Complex Software Systems*, HP Laboratories: Palo Alto

Myers, R and Saigol, Z (2020). *Pass-Fail Criteria for Scenario-Based Testing of Automated Driving Systems*, Connected Places Catapult, UK

NASA (2010). *NASA/SP-210-580 NASA System Safety Handbook, Volume 1: System Safety Framework and Concepts for Implementation*, NASA

Page, S.E. (2009). *Understanding Complexity*. The Great Courses. Chantilly, VA, USA: The Teaching Company.

SAE (2019). *SAE J3016^{TM} Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*, SAE

Sillitto, H.G. (2010). *Design principles for ultra-large-scale systems*, Proceedings of the 20th Annual International Council on Systems Engineering (INCOSE) International Symposium, July 2010, Chicago, IL, USA

Singh, S (2015). *Critical reasons for crashes investigated in the National Motor Vehicle Crash Causation Survey*, Washington, DC

UK Ministry of Defence (2019). *DEF STAN 00-970 Design and Airworthiness Requirements for Service Aircraft*, UK MOD

Midlands Future Mobility, "Midlands Future Mobility," 2020. [Online]. Available: https://midlandsfuturemobility.co.uk/. [Accessed: 27-Mar-2020].

Warwick Manufacturing Group (WMG) (2020). *Autonomous Vehicle safety standards to be set by Warwick academic*, Retrieved May 16, 2020, from https://warwick.ac.uk/newsandevents/pressreleases/autonomous_vehicle_safety

World Health Organisation (WHO) (2018). 2018 Report on Road Safety, World Health Organisation