Proceedings Paper:
Birch, John, Blackburn, David, Botham, John et al. (5 more authors) (2020) A Structured Argument for Assuring Safety of the Intended Functionality (SOTIF). In: International Workshop on Artificial Intelligence Safety Engineering. Lecture Notes in Computer Science . Springer , pp. 408-414.

https://doi.org/10.1007/978-3-030-55583-2_31

A Structured Argument for Assuring Safety of the Intended Functionality (SOTIF)

John Birch[1], David Blackburn[2], John Botham[3], Ibrahim Habli[4], David Higham[5], Helen Monkhouse[1], Gareth Price[6], Norina Ratiu[7], Roger Rivett

[1] HORIBA MIRA Ltd
[2] Bentley Motors Ltd
[3] Ricardo UK Ltd
[4] University of York
[5] Imagination Technologies
[6] McLaren Applied
[7] Aston Martin Lagonda

Abstract. Current safety standards for automated driving recommend the development of a safety case. This case aims to justify and critically evaluate, by means of an explicit argument and evidence, how the safety claims concerning the intended functionality of an automated driving feature are supported. However, little guidance exists on how such an argument could be developed. In this paper, the MISRA consortium proposes a state machine on which an argument concerning the safety of the intended functionality could be structured. By systematically covering the activation status of the automated driving feature within and outside the operational design domain, this state machine helps in exploring the conditions, and asserting the corresponding safety claims, under which hazardous events could be caused by the intended functionality. MISRA uses a Traffic Jam Drive feature to illustrate the application of this approach.

Keywords: Safety Assurance, Safety Case, SOTIF, ODD, Automated Driving.

1 Problem

1.1 Safety Assurance of Automated Driving

Automated Driving (AD) promises to revolutionize the future of road transportation. However, the challenge of assuring its safety is significant and is subject to ongoing discussion and research. There are a variety of emerging standards such as ISO/PAS 21448 [1], UL 4600 [2] and ISO/TR 4804 [3] that relate to the safety of AD. These standards leave freedom for developers to reason about the safety of their systems by calling for the achievement of high-level goals or objectives, rather than conformance to prescriptive requirements, and by avoiding a declaration of what level of residual risk is reasonable or otherwise.

It is therefore not considered appropriate, nor feasible, to attempt to generate a compliance argument of the form "The Automated Driving System (ADS) is safe because its development complies with the requirements of standard X". Instead there is a professional responsibility placed on engineers to creatively justify, based on clear and rigorous evidence, why they believe their ADS is free from unreasonable risk. It is proposed that this justification should be communicated in the form of a safety argument, as part of a safety case [4], that will feature claims, assumptions and evidence related to a variety of standards, as acknowledged in [2]. This will help to ensure greater transparency in the development of ADS by enabling safety assessors and other stakeholders to critically evaluate the basis on which the system might be deployed.

1.2 Role of the Operational Design Domain

It is often the case that the Intended Functionality (IF), [1], of the ADS can only be achieved for a restricted set of vehicle, and external environmental, conditions referred to as the Operational Design Domain (ODD), [5], and defined as the "Operating conditions under which a given driving automation system or feature thereof is specifically designed to function (…)" [6]. This limitation may arise from known performance limitations or specification inefficiencies. To justify that the ADS is free from unreasonable risk it is necessary to reason about its IF when the vehicle is within the ODD, but also when the vehicle is transitioning into and out of the ODD.

The aim of this paper is to propose an approach to assuring ADS safety, initially aligned to ISO/PAS 21448, which is based on the central role played by the ODD and its transitions. It is illustrated with some example safety assurance considerations for a generic "Traffic Jam Drive" (TJD) feature.

2 Proposed Approach

2.1 ODD Transitions in an Example TJD Drive Cycle

Consider a typical drive cycle in which the generic TJD feature described in [7] may be used:

- The driver starts their journey by initializing the vehicle outside of the ODD before driving it into the ODD (e.g. onto a highway in clear weather with a lead vehicle etc.);
- TJD availability is indicated to the driver and the driver chooses to enable the feature, handing responsibility for the Dynamic Driving Task (DDT) [6] to the TJD feature;
- The TJD feature continues to control the DDT until either:
    - The driver chooses to deactivate the feature and resume control, or
    - The TJD hands control back to the driver without driver request;
- The driver leaves the highway (exiting the ODD), completes their journey and parks and secures the vehicle.

If the TJD feature were to be activated before entry to the ODD, or if the vehicle were to leave the ODD with the TJD still in control, the TJD feature would be responsible for controlling the DDT under conditions for which it was not designed. However, unless the driver is ready to resume control it may be unsafe for the TJD feature to relinquish DDT responsibility on exiting the ODD.

2.2 Presence of the Vehicle in the ODD and Activation Status of the Intended Functionality

The two key parameters identified in the above drive cycle, whose combination is critical for considering safe control of the DDT, are:

1. The presence of the vehicle in the ODD, or otherwise;
2. The activation status of the (TJD) feature.

MISRA expresses the combination of these parameters in the form of a state termed the "ODD-Activation State" which can take one of four values:

- State 1 - IF is active whilst the vehicle is within the ODD
- State 2 - IF is active whilst the vehicle is outside of the ODD
- State 3 - IF is inactive whilst the vehicle is within the ODD
- State 4 - IF is inactive whilst the vehicle is outside of the ODD

These states and the possible transitions between them are depicted as a state machine in Figure 1. It might be argued that transitions could occur directly between State 1 and State 4 and also between State 2 and State 3. This would require the IF activation status to change at exactly the same time as the vehicle presence in the ODD changes. In practice this is very unlikely to occur, although it is recognized that the time spent in some of the states could be very short.

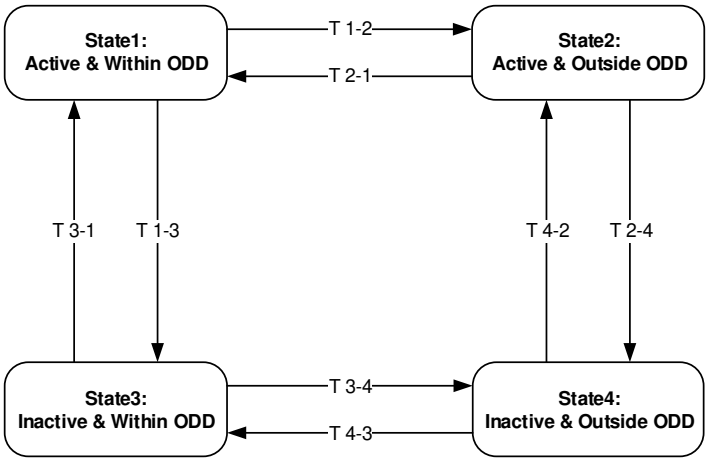

Fig. 1. ODD-Activation States and transitions

2.3 Example TJD Safety Claims

By explicitly defining the states and transitions in Figure 1 the corresponding safety implications and possible safety claims can be systematically identified. Let us illustrate this by returning to the TJD drive cycle example. Using the ODD-Activation state machine (Figure 1), Table 1 expands the steps previously outlined with some example informal claims that one may wish to make about the corresponding ADS behaviour.

Table 1. Example TJD ODD-Activation States and transitions and corresponding safety claims

ODD-Activation State or Transition	TJD Drive Cycle Step	Example Informal Safety Claims
State 4	Driver initialises vehicle outside of the ODD as the vehicle has not yet entered a highway with a lead vehicle, even though visibility is good.	The TJD feature will detect when the vehicle is outside of the ODD. Activation of the TJD is prevented until the vehicle enters the ODD.
T 4-3	Vehicle enters the highway behind a lead vehicle. Visibility remains good and so the vehicle has entered the ODD. The driver is still in control of the DDT.	-
T 3-1	TJD availability is indicated to the driver and the driver chooses to activate the feature, handing across control of the DDT.	The handover of DDT control to the TJD is as anticipated by the driver - it is intuitive and predictable and does not occur unless it is requested by the driver who is ready for it.
State 1	The TJD feature continues to control the DDT until...	The TJD controls the DDT within the ODD in a safe manner (e.g. successfully performing Object and Event Detection and Response (OEDR) [6] by keeping the vehicle in lane and at a safe distance to the lead vehicle, avoiding obstacle collision etc.)
T 1-3	...the driver chooses to deactivate the feature, taking back control of the DDT ...	The hand-back of control to the driver by the TJD is as anticipated by the driver - it is intuitive and predictable and does not occur until the driver is ready for it.
T 1-2 *State 2* *T 2-4*	...or the TJD hands DDT control back to the driver because, for example, visibility suddenly drops due to a change in the weather. Note: this would ultimately cause entry into State 4, via State 2.	The TJD will never cause exit from the ODD, e.g. by causing the vehicle to leave the highway. The TJD will detect the vehicle leaving the ODD (e.g. due to a sudden change of weather conditions, outside of its control) in a timely manner If the vehicle leaves the ODD whilst the TJD is in control of the DDT the TJD feature will take an appropriate and timely safe action, such as handing back control of the DDT to an alert driver or reaching a Minimal Risk Condition (MRC) [6]. The TJD feature will not regularly have to hand-back control of the DDT to the driver because of the inability of the feature to cope with commonly occurring, predictable, conditions (such as a change in weather conditions).
State 4	The driver completes the drive cycle, bringing the vehicle to rest and powering it down	-

2.4 MISRA SOTIF Argument Structure

From the example claims in Table 1, and by considering the states and transitions in Figure 1, we can extract some general high-level claims that one may wish to make about any ADS. These have been collated in a single argument structure expressed in Goal Structuring Notation (GSN) [8], the top level of which is shown in Figure 2.

For completeness, the argument structure incorporates reference to the consideration of post-release SOTIF issues. Whilst this is an important topic it is not one considered to be central to the ideas presented in this paper and is thus not explored further.

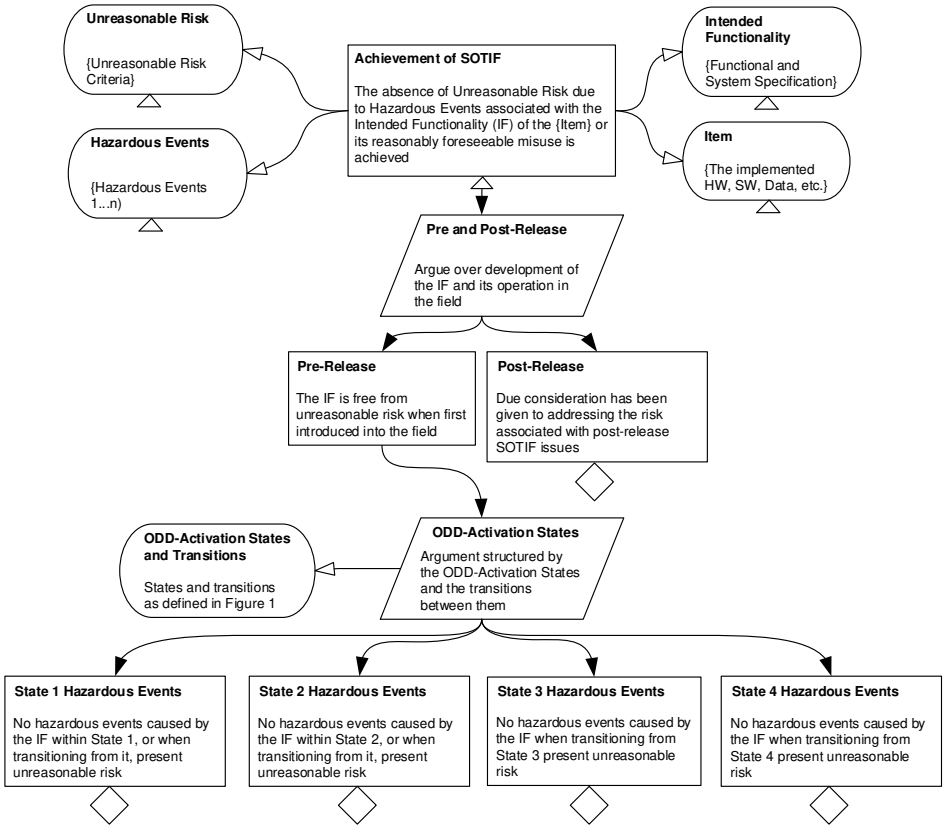

Fig. 2. Top-level SOTIF safety argument

3 Discussion and Further Work

The four-state model and corresponding safety argument represents MISRA's initial insight into an approach that highlights the central role played by the ODD in assuring ADS safety. The argument in Figure 2 represents an initial structure for a series of subsequent claims and items of evidence that will relate to a variety of topics in [1]. It

is anticipated that these claims will be categorized according to the following MISRA argument themes related to those introduced in [9]:

- The rationale for the SOTIF requirements used to specify the IF;
- The satisfaction of these requirements by the implemented IF;
- The means used to perform the various SOTIF-related activities;
- The development environment in which they have been performed.

Work is ongoing to further develop the argument structure and to recommend supporting claims. This will include broadening the argument scope (beyond [1]) to incorporate causes of hazardous events relating to malfunctions (functional safety) and vulnerabilities (cybersecurity). It is anticipated that this work will form a basis for a subsequent MISRA publication that follows on from [9].

References

1. ISO / PAS 21448:2019 Road Vehicles – Safety of the intended functionality.
2. UL 4600 UL Standard for Safety for Evaluation of Autonomous Products. First Edition, Apr 2020.
3. ISO / CD TR 4804 Road Vehicles – Safety and security for automated driving systems – Design, verification and validation methods.
4. R Hawkins, T Kelly, J Knight, P Graydon. A New Approach to Creating Clear Safety Arguments
5. M Gyllenhammar et al, Jan 2020. Towards an Operational Design Domain That Supports the Safety Argumentation of an Automated Driving System
6. SAE J3016, Jun 2018. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles
7. NHTSA DOT HS 812 623, Sep 2018. A Framework for Automated Driving System Testable Cases and Scenarios
8. Goal Structuring Notation Community Standard Version 2, Jan 2018.
9. MISRA Guidelines for Automotive Safety Arguments, ISBN 978-1-906400-24-8, Sep 2019.