Minimal Risk Condition for Safety Assurance of Automated Driving Systems

Magnus Gyllenhammar
Zenseact; KTH Royal Institute of Technology
Göteborg, Sweden
magnus.gyllenhammar@zenseact.com

Mattias Brännström
Zenseact
Göteborg, Sweden

Rolf Johansson
Astus AB
Göteborg, Sweden
rolf@astus.se

Fredrik Sandblom
Volvo Autonomous Solutions
Göteborg, Sweden
fredrik.sandblom@volvo.com

Stig Ursing
Semcon Sweden AB
Göteborg, Sweden
stig.ursing@semcon.com

Fredrik Warg
RISE Research Institutes of Sweden
Borås, Sweden
fredrik.warg@ri.se

***Abstract*—We have yet to see wide deployment of automated driving systems (ADSs) on public roads. One of the reasons is the challenge of ensuring the systems' safety. The operational design domain (ODD) can be used to confine the scope of the ADS and subsequently also its safety case. For this to be valid the ADS needs to have strategies to remain in the ODD throughout its operations. In this paper we discuss the role of the minimal risk condition (MRC) as a means to ensure this. Further, we elaborate on the need for hierarchies of MRCs to cope with diverse system degradations during operations.**

***Index Terms*—Automated driving systems, Safety, Minimal risk condition, Degraded operations, Safe state**

I. Introduction

One of the promises of increased automation of the transportation system is improving its safety. Paradoxically, this is also one of the main challenges when developing automated driving systems (ADSs). Similar to traditional automotive systems an ADS needs to be implemented to a high integrity to avoid accidents and fatalities when released in large volumes. Direct validation is both infeasible, due to the large amount of driving needed to prove sufficient integrity of the ADS [1], and potentially unsafe, as we have seen in the recent crashes with (partially) automated vehicles (Table 1.2 [2]). Since this brute force approach is not feasible, there is a need for alternatives. Raising the question; how can the safety of the ADS be ensured before release on public roads? This challenge is equivalent to showing that the residual risk after all design, implementation and verification efforts is sufficiently low. The operational design domain (ODD) has been proposed as a useful tool to achieve this goal [3].

In this paper, we explore what is required for the ADS to remain inside the ODD. The role of the minimal risk condition (MRC) is elaborated. In the light of the three decision levels, strategic, tactical and operational [4], we further discuss how a loss of capabilities can be handled with tactical decisions; i.e. through appropriately restricting the ADS's actions to the new capabilities or by re-planning the route to avoid unsuitable external operating conditions (OCs). If none of these are possible, the third option is to transition into MRC.

Related works are discussed in Sec. II and the role of the MRC is discussed in Sec. III. Handling capability reductions of the ADS in relation to the term restricted operational domain (ROD) is discussed in Sec. IV. Finally, conclusions and potential future works are given in Sec. V.

The research has been supported by the Strategic vehicle research and innovation (FFI) programme in Sweden via the SALIENCE4CAV project (ref 2020-02946) and the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

A. Preliminaries

In this paper, we focus on level 4 automated driving systems (ADSs), as defined in SAE J3016 [5]. A level 4 ADS is responsible for carrying out the entirety of the Dynamic Driving Task (DDT) within its specified ODD, including an ability to reach a minimal risk condition (MRC) if the mission cannot be completed. It may have a fallback-ready user ready to assist in such situations, but must also be able to resolve them on its own.

1) The two driving states of an ADS: According to SAE J3016, the DDT includes all the operational and tactical decisions required for operating the ADS in traffic to fulfil a strategic objective, but excluding the strategic functions such as destination and waypoint selection. The DDT is performed until either of the following four events happen:

(i) The user-defined mission given to the ADS is completed.
(ii) The driver[1] requests a takeover - the ADS tries to put itself in a state where such a handover can be conducted safely - and the handover is successfully completed.
(iii) A DDT performance-related system failure occurs - which is followed by the execution of a DDT Fallback (DDT-FB) to achieve MRC, or alternatively to a handover can be made.

[1]i.e. a conventional (in-vehicle) driver or a remote operator.

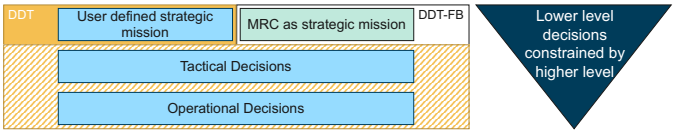

Fig. 1. The three types of driver efforts in a decision hierarchy. The higher levels limit what decisions are valid at the lower levels.

(iv) Nearing an ODD exit - DDT-FB is triggered to reach MRC, or the control is handed over to a driver before the exit occurs.

There are thus two distinct driving states: DDT and DDT-FB.

2) Three Levels of Driver Efforts in Context of ADS: To understand the different levels of decisions involved in operating an ADS, J3016 [5] introduces three levels of driver efforts, as suggested by Michon [4]. The three levels are: strategic, tactical and operational, and can be positioned in a decision hierarchy according to Fig. 1. On a strategic level, the ADS either operates to fulfil a user-defined mission or to achieve MRC. The tactical decisions involve taking safe actions to avoid other objects as well as to take appropriate actions to fulfil the strategic mission. Operational decisions, on the other hand, are the low-level vehicle control operations.

When the user-defined mission is applicable, the tactical decisions together with the operational ones are responsible for carrying out the DDT. In the case where the MRC is the strategic mission, the tactical and operational decisions will instead execute the DDT-FB.

II. Related Work

Reschka and Maurer [6] discuss what constitutes a safe state for an automated road vehicle and present four different conditions for such. The first three pertain to the DDT, performed either by (1) the human, by (2) a remote entity (e.g. through telecommunications) or by (3) the ADS itself. The fourth condition (4) is when the vehicle is standstill.We do not address the (1) and (2) but agree that the driver should be considered as one possible safe state. For (4) we note that the safety of a state is not only determined by the risk of the state itself (which is covered in [6]) but also by the propensity of the ADS to enter this state as well as the rate of recovery. Further, [6] suggest that the safety of condition (3) can be estimated as the difference between the available capabilities and the demanded capabilities. We suggest that this is (a) primarily covered by the ODD, when it comes to external conditions, (b) the prompt decision of entering an MRC if the capabilities can no longer be matched due to a system failure, and (c) the capability reporting together with appropriate tactical decisions for system failures which do not violate the demanded capabilities. Xue et al. [7] argue for a prolonged DDT-FB to reach a more safe MRC. We introduce a hierarchy of MRCs to be able to select the most appropriate (safest) one. Contrary to ISO/TR 4804, we consider an MRC as a final decision and further require the ADS to be at standstill, as does J3016 after the 2021 update [5]. There should be no recovery from MRC to the original user-defined goal (cf. Figure 8 of [8]). Reschka et al. [9] suggest that a set of performance criteria are matched with heuristic degradation actions to handle lost performance. This is an example of how to construct the tactical decisions to ensure safe actions considering the reported capabilities. However, we note that the approach of listing concrete heuristics for each different capability change is not scalable. For large discrete system degradations, Colwell et al. [10] coin the term restricted operational domain (ROD). We acknowledge the usefulness of the term, but also recognise the need for equivalent methods to handle more temporary and non-discrete degradations. Rather than a supervisory layer handling the different RODs, we believe that this comes as a natural consequence from having the tactical decisions receive capability reports from the different parts of the system and take appropriate actions based on this information. A method along these lines, for posing requirements on the perception sensors to report their capabilities, is suggested in [11].

III. Clarifying the Role of the MRC

Not all OCs of the ODD can be firmly ascertained throughout the operations of the ADS. To account for that, the ADS needs to bring itself to MRC before a violation of an OC. This is especially relevant when employing strategy III or IV of [3], using statically defined geographical and temporal conditions for where/when the OCs are valid (III), or run-time triggering conditions (IV). For instance, if an example ADS is travelling from A to B and is stuck in traffic, it might realise that it is unable to reach B before sunset. To avoid this the ADS will abandon its original strategic goal (B) and transition into MRC before the 200 lux ODD limit is violated. For the safety case to be valid the ADS should not be operational outside the ODD, thus the MRC itself cannot be a driving state of the system, else it would make no difference to transition into MRC to handle ODD exits. In the example, the illumination will be violated eventually, once the ADS is in MRC. The only way of arguing that the ADS is no longer driving, is by having the MRC as a state of standstill. This has also been recognised by the update of J3016 [5] in 2021 where it is defined as a "... *stopped condition* ...". In addition to that we would like to pose a yet improved definition:

Definition: Minimal risk condition (MRC) is a stable stopped condition at a position with an acceptable risk given the situation when the decision to enter MRC is taken. If an acceptable risk is not attainable, the position with the lowest risk should be selected. *The ADS is brought to this state by the user or the system itself, by performing the DDT-FB, when a given trip cannot or should not be completed. [5]*

The augmented definition highlights the need for arguing the safety for each MRC. This safety is made up of three components:

1) the frequency to enter the MRC,
2) the risk of the position selected, and
3) the rate of resolving the MRC, i.e. to bring the ADS out of MRC or have a driver take over the driving task.

Each OC strategy should be associated with a standstill position. This position could be all from at the curb of the road to neatly parked at the closest rest area. Depending on which, the frequency of entering that position can be higher or lower. As can the rate of recovery. Stopping on the side of the road on a highway is incurred with a significant risk and should be entered infrequently and/or rapidly be resolved, whereas parking at the rest area could be allowed more often and, as long as there is not a passenger in the vehicle, this state is safe for a very long time. Note that the acceptability of the risk of the state is related to the analysis pertaining to the three items listed above. Further, the lowest risk states selected, due to unattainability of an acceptable risk state, also needs to be included in the overall risk analysis of the system.

IV. Handling Performance Degradations

In addition to the ODD, the tactical decisions are limited by the current capabilities of the ADS. Let us assume that the ADS is configured with one perception block reporting to a decision-making block, which subsequently request actuation of paths from a vehicle platform. In such a case, the capabilities from the perception as well as the vehicle platform limit the tactical decisions to consider only manoeuvres that are both viable with respect to perception performance as well as attainable with the currently available (predicted) actuation capabilities. All possible capability reports should be understandable by the tactical and strategic decisions. The ADS needs to fulfil its DDT given these reports and if that is not the case, it should abandon the user-defined goal and go to an appropriate MRC. This is called a *DDT performance-related system failure* [5].

For the ADS to understand when a sub-optimal capability is sufficient, in terms of continued operations, this must have been part of the development and assessed and verified. Small temporary performance fluctuations would likely be included in the verification of the system, especially if the fluctuations are due to external factors. In that case, the range of such conditions should be encoded as OCs of the ODD. Such conditions could include occlusion of the vision sensors, due to precipitation, or increased braking distance due to a wet road surface. But what about the performance degradation in the system? What if we lose connection to one of the cameras? Or if the braking system is no longer redundant? Or if we have an occasional glitch in the communication with one of the sensors? The resulting capabilities of these ”new” subsystems need to be understood and assessed to allow continued operation of the ADS. Further, these capabilities need to be compared to what is required from the current DDT. Colwell et al. [10] use the term restricted operational domain (ROD) to describe where a permanently degraded subsystem is able to safely operate. However, we propose that it is the task of the tactical decisions to cope with any kind of capability degradation, not just large and discrete ones. Furthermore, the better the system is at self-diagnostics, the closer to a continuum is reported in the form of capability reports. Considering the RODs as a finite set of degraded states thus limits the usefulness of the concept. That said, the response from the tactical decisions is dependent on the amount of analysis and verification that has been conducted for each capability degradation. With increased development and verification efforts the availability of the system under diverse capability restrictions can be increased.

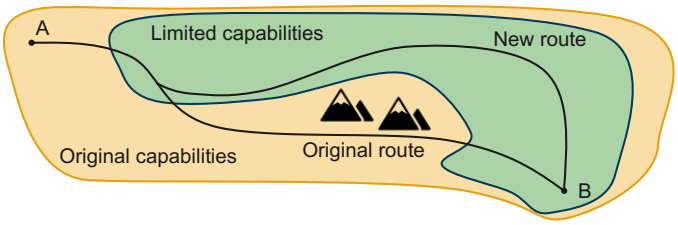

Fig. 2. Two routes between A to B are depicted. The original route across the mountains is no longer valid after a performance degradation, but the new route can be executed safely within the new capabilities (green area).

If the ROD of a capability report is not understood the only option is to proceed to MRC. If such an analysis has been done however, the tactical decisions are left with two options:

(1) Keep the strategic goal and potentially change the route to avoid OCs that are not viable for the current ROD, or
(2) abandon the current strategic goal and go into MRC.

Within option (1) the task of remaining within the available capabilities (e.g. not exceeding a maximum speed of 80 km/h) is the task of the tactical decisions. Fig. 2 shows such an example. An example ADS experiences a reduced braking capability. It is still able to brake if driven at speeds below 80 km/h, but it is unable to handle steep slopes. Since these limitations are known, the tactical decisions decide to change the original route, over the mountains, to instead drive around and thus allowing the ADS to fulfil the original strategic goal. The change of the routes requires three things, a knowledge of the capabilities, an understanding of what is required to fulfil the strategic goal (cf. ODD exit strategy II [3]), and knowledge of the demanded capabilities of alternative routes. Within option (2) a detailed understanding of the capabilities of the sub-system makes it possible to select a more favourable MRC position. Instead of directly stopping in lane one can return to a dispatch centre or park at a nearby rest area.

A. Involving the User

Many of the ADS’s failures might be handled by a user taking back control. For systems with fall-back ready users, analysing each and every ROD is probably not a worthwhile effort. For dedicated vehicles, without drivers, this analysis will likely need to be expanded as it will be both risky and costly to allow the ADS to stop upon every small system failure. The value of understanding the RODs of the system will then outweigh the cost of doing the analysis.

If the user-defined goal is kept even after a take-back request it is fair to assume that some users might abuse this and deliberately refrain from taking back control to see if the ADS eventually reaches the user-defined goal anyways. If an ADS relies on a successful handover in e.g. 99% of the cases

when the ADS is approaching a road works, such abuse would result in our frequency of entering the MRC would be violated. Because of this, the decision of going into MRC must be a final one. If the user wishes to go back to its defined goal, then it has to initiate a negotiation for a new trip. Once the ADS is in MRC it is not in a driving state, but it is still operational and e.g. able to initiate a new trip.

B. A Hierarchy of MRCs

Let us explore the notion of having different MRCs depending on the situation the ADS finds itself in. It is evident that each strategy to ascertain the OCs could have different MRCs depending on which OC(s) it is tasked with ensuring and also related to the time until the OC(s) is violated. Consider an OC pertaining to sufficient lighting (e.g. illumination > 1000lux $\in$ ODD). It is possible to know well in advance when the sun will set and if the user-defined mission cannot be completed before that time an appropriate MRC position, e.g. stopping at the rest area closest to the destination, can be selected. However, while performing the DDT-FB to reach this parking lot there might be a queue which inhibits the ADS to reach this state before the OC is violated. This warrants a more prompt transition into MRC and it might be necessary for the ADS to eventually stop on the curb of the road to avoid an ODD exit. But there could also be another OC in risk of being violated when the ADS is in the process of achieving the MRC of the first. Rather than a queue, we might suddenly get a weather forecast saying that it will start hailing, which (let us assume) is not a valid OC. Also in this case the ADS needs to make a quick transition into an other MRC to accommodate this new information and avoid driving in hail. Similarly, an MRC should be defined for each system failure impacting the operating capabilities of the ADS resulting in an inability to fulfil the user-defined mission. Depending on the failure, the MRC might be more or less restrictive. Both the MRCs to avoid ODD exits as well as the MRCs to handle system failures can be put in a hierarchy related to the required final position and the allocated time to reach this state. A state diagram of the ADS with this hierarchy of MRCs is depicted in Fig. 3. Just as for abandoning the user-defined goal the choice of entering another MRC should be a definite one. This is why the MRCs can be ordered in a hierarchy as depicted.

Emergency manoeuvres, e.g. autonomous emergency braking (AEB), do not necessarily result in the abandonment of the strategic goal. Consider a successful AEB intervention to avoid collision with an obstacle. If the road is cleared the ADS might continue to fulfil its strategic goal. Emergency manoeuvres add a reactive component to the otherwise hierarchical decision making of the system. It is not the reaction itself but rather the outcomes that result in a change of the strategic goal.

V. Conclusions and Future Work

In this paper we elaborate on how the MRC should be used to avoid ODD exists and to cope with system failures of an ADS and gives an elaborated definition to the term. Further, the MRC should be the result from a definite decision

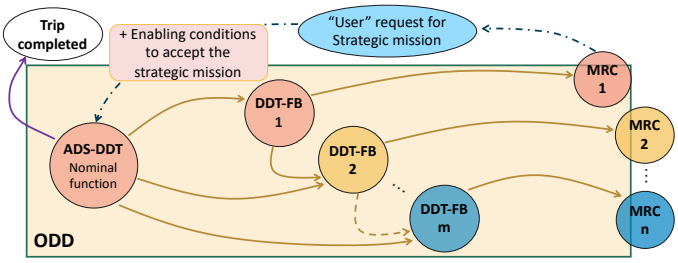

Fig. 3. Illustrates that the ADS can be associated with a set of MRCs. Before the ADS has fully reached the intended MRC the strategic goal can change to an MRC with higher requirements, which is indicated as the transition from DDT-FB 1 to DDT-FB 2 etc. Once the MRC is reached the ADS will remain there until a new trip is requested.

to abandon the previous strategic mission. Capabilities of the ADS should be reported to, and eventual reductions should be handled by, the tactical decisions. This view of handling degraded performance is a more refined way of incorporating the ROD, as proposed in [10].

As future work we suggest to investigate the implications of capability reporting on functional architecture.

References

[1] R. A. Young, "Automated driving system safety: Miles for 95% confidence in "vision zero"," *SAE International Journal of Advances and Current Practices in Mobility*, vol. 2, no. 2020-01-1205, pp. 3454–3480, 2020.

[2] M. Carre, "Autonomic Framework For Safety Management In The Autonomous Vehicle," Theses, Université de Pau et des Pays de l'Adour, Dec. 2019. [Online]. Available: https://hal-univ-pau.archives-ouvertes.fr/tel-02455266

[3] M. Gyllenhammar, R. Johansson, F. Warg, D. Chen, H.-M. Heyn, M. Sanfridson, J. Söderberg, A. Thorsén, and S. Ursing, "Towards an operational design domain that supports the safety argumentation of an automated driving system," in *Proceedings of the 10th European Congress on Embedded Real Time Systems (ERTS)*, Toulouse, France, Jan. 2020.

[4] J. A. Michon, "A critical view of driver behavior models: what do we know, what should we do?" in *Human behavior and traffic safety*. Springer, 1985, pp. 485–524.

[5] SAE, "SAE J3016:202104 - SURFACE VEHICLE RECOMMENDED PRACTICE - Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles," 2020.

[6] A. Reschka and M. Maurer, "Conditions for a safe state of automated road vehicles," *it - Information Technology*, vol. 57, Jan 2015.

[7] W. Xue, B. Yang, T. Kaizuka, and K. Nakano, "A fallback approach for an automated vehicle encountering sensor failure in monitoring environment," in *2018 IEEE Intelligent Vehicles Symposium (IV)*, June 2018, pp. 1807–1812.

[8] ISO, "ISO/TR 4804:2020 Road vehicles — Safety and cybersecurity for automated driving systems — Design, verification and validation," 2020.

[9] A. Reschka, J. R. Böhmer, T. Nothdurft, P. Hecker, B. Lichte, and M. Maurer, "A surveillance and safety system based on performance criteria and functional degradation for an autonomous vehicle," in *2012 15th International IEEE Conference on Intelligent Transportation Systems*, Sep. 2012, pp. 237–242.

[10] I. Colwell, B. Phan, S. Saleem, R. Salay, and K. Czarnecki, "An automated vehicle safety concept based on runtime restriction of the operational design domain," in *proceedings of 2018 IEEE Intelligent Vehicles Symposium (IV)*, Changshu, China, Jun. 2018.

[11] R. Johansson, S. Alissa, S. Bengtsson, C. Bergenhem, O. Bridal, A. Cassel, D.-J. Chen, M. Gassilewski, J. Nilsson, A. Sandberg, S. Ursing, F. Warg, and A. Werneman, "A strategy for assessing safe use of sensors in autonomous road vehicles," in *Computer Safety, Reliability, and Security*, S. Tonetta, E. Schoitsch, and F. Bitsch, Eds. Cham: Springer International Publishing, 2017, pp. 149–161.