Agreements of an Automated Driving System

Martin Skoglund
Department of Electronics
RISE Research Institutes of Sweden
Borås, Sweden
martin.skoglund@ri.se

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

Behrooz Sangchoolie
Department of Electronics
RISE Research Institutes of Sweden
Borås, Sweden
behrooz.sangchoolie@ri.se

***Abstract*—When introducing automated driving systems (ADS), it is imperative that there exist mutual agreements between the ADS and stakeholders – such as the ADS equipped vehicle user, other road users, and society at large – on how the ADS should behave. Lacking such agreements, the ADS may antagonize stakeholders and, even worse, pose severe safety risks. The ADS needs a complete and unambiguous set of machine-interpretable properties describing these interactions, while the human stakeholders need to understand and accept how the ADS is designed to behave. We propose to make these considerations explicit in the form of agreements. The completeness problem is tackled by cataloguing and categorizing all agreements that need to be considered during the lifetime of an ADS in a systematic way.**

Keywords— automated driving system, agreements, system safety, ethically aligned design

I. Introduction

Showing that an automated road vehicle is safe and can be trusted is a complex problem [1]. The scope of this work is to identify and address a small part of this problem - how to identify all agreements which must be explicitly conveyed and affirmed by the involved stakeholders and the ADS. An agreement that is extensively debated, and of great importance, is that there should be no disagreement in responsibility of control between a manual driver and the ADS feature, i.e. no mode confusion [2]. This type of agreement is a subclass of the agreements discussed in this work, which proposes a way to catalogue and categorize all agreements to consider during the life-time of an ADS.

The ADS needs a complete and unambiguous set of machine-interpretable properties describing these interactions, while the stakeholders need to understand and accept how the ADS is designed to behave. It is important to know how these agreements need to be addressed in order to avoid unreasonable risk of harm to persons, the environment or the economy. This is akin to making an ethical aligned design [3] of the ADS, where one concern is that the operation must be transparent to a wide range of stakeholders.

For users, transparency is important because it builds trust in the system, hence providing a simple way for the user to understand what the system is doing and why. For validation and certification of an ADS, on the other hand, the transparency is important because it exposes the system's processes for scrutiny. The formulation of agreements could help to ensure these qualities. The categorization enables the agreement-sets to be analyzed separately, in a methodical manner. Note that future work will be focused on how to represent handshake points using interval algebra, and how to analyze quantification and distribution of common knowledge before entering these agreements.

The research has been supported by VINNOVA in the ESPLANADE project (ref 2016-04268).

II. Elicit Agreements and Impose Structure

A. Elicit Agreements

We use the term agreement *(i)* to ensure and transfer the knowledge of how a specific automation works, linked with the user and the exact features installed; and at the same time *(ii)* to capture corresponding issues of how the ADS expects to be interacted with by the user. For an agreement to be valid, no change can occur that affect the conditions, of either party, under which the agreement was entered. The relationship is bi-directional, this is, a property that can be generalized to all arrangements needs to be identified. All contracts are of type agreement, but all agreements are not contracts; agreements that bind us legally are called contracts and are a subset of agreements. At this stage, the more general term of agreements is the most relevant, as repercussions of a broken agreement/contract are not yet analysed.

Since there is a need to make all considerations during the lifetime of an ADS explicit, the first step is to appropriately enumerate all agreements that can be found. The complexity of the obtained map of the complete agreement landscape is significant. In fact, the result is all combinations over the stakeholders that interacts with the ADS, all characteristics of these interactions, all types or classes of ADS as well as the life-time of ADS. Depending on the granularity of the design needs for each of these parameters, even if kept conservative, the enumeration of agreements would still result in hundreds of agreements. Hence, there is a need to divide the outlined problem. The solution we propose does not infringe on the completeness of the agreement set, rather visualize it in a more readable way thus enabling the agreement-sets to be analysed separately.

B. Impose Structure

In this section, we define the characteristics of the categorizations that was found when analysing the complete set of agreements identified during elicitation. The results can be divided into (1) stakeholders, (2) time scale, and (3) attributes, forming 27 categories to classify agreements. This is illustrated in Fig. 1, which also gives a few concrete examples of agreements. The categorization enables the agreement-sets to be analysed separately. Additional aspects applicable to all sets will encompass ADS type and the ADS lifetime. Below, we discuss each of the characteristics:

1) **Stakeholders.** Our approach to categorize the stakeholders is to divide them into three groups. The definitions of *Users* and *Passenger* was taken from SAE J3016 [4], and is appended with maintenance personnel and others that interact directly with the ADS. *Vehicle proximity* is defined as the close vicinity of the ADS equipped vehicle.

It contains other road users and V2X communication. The category *Society* contains stakeholder with more general concerns like insurance companies, national transport agencies, regulatory or certification entities etc

2) **Time Scale**. Here, we refer to time scale as a property contained within the agreements, not the time-scale of the agreements themselves. The *Operational* domain contains agreements regarding situations with one decision point and one action on short time-frame, e.g. correction to keep vehicle in lane. *Tactical* is on a medium time-frame containing more than two decision points and actions, that could be represented by a decision tree, e.g. an overtake manoeuvre. *Strategic* time scale contains agreements that concerns a long time-frame e.g. trip [4] or the operational lifetime of the vehicle.

3) **Attributes**. Here, we classify attributes into three categories; The first category is refered to as *Safe*, capturing agreements pertaining to traditional functional safety, i.e. freedom from unacceptable risk of physical injury or of damage to the health. The *legal* category contains agreements connected to things regulated by law, unauthorized or unsanctioned. And finally the *proper* category capture agreements to conform to accepted behaviours or customs.

ADS Types. The categories for the ADS types, that might impact the agreements, were collected from the ERTRAC roadmaps for automated driving [5], and are as follows: *automated passenger cars*, *freight vehicles* and *urban mobility vehicles*.

ADS Life-Time. There is a need to make the continuous space of the life-time of ADS discrete, where agreements are valid over segments. This is done by introducing the concept of **handshake points,** which constitutes the start point of the segments. The end of each segment is when a change has occurred that affect the conditions of either party under which the agreement was entered. Here, an agreement is a bi-directional understanding reached between two parties, that is invalid when conditions change. Based on the global data protection regulation (GDPR) [6], the act of handshake, i.e. consent or agreeing, can be defined as any freely given, specific, informed and unambiguous indication of the wishes by which stakeholder or ADS, by a statement or by a clear affirmative action, signifies agreement.

III. Conclusions and Future Work

We set out to *divide* and *conquer* the complexity of the problem of making all the considerations between an ADS and a stakeholder explicit in the form of agreements. This was done by enumerating all agreements that needs to be considered during the lifetime of an ADS. Then identifying and applying an appropriate categorization, allowing the agreement sets to be analysed separately. For instance, now one can pick any specific category or consideration of interest – corresponding to the ADS type and application that is being analysed – and exclude all others, while remaining confident in completeness of the underlying agreement set.

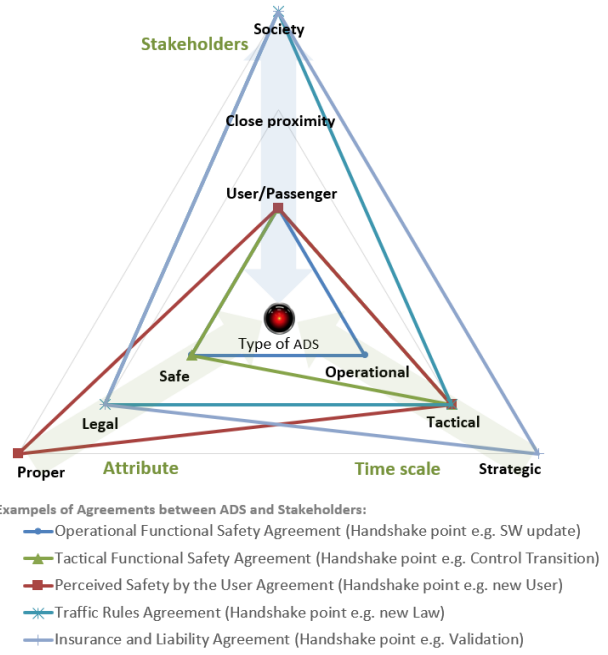

Fig. 1 Agreement characteristics.

What was achieved in this paper is the division of the problem, but the conquering part needs further work. Future work also includes investigation on how to handle cybersecurity and GDPR concerns, and to elaborate on the repercussions of a broken agreement. We also plan to use an interval algebra [7] to distribute handshake points more formally. Furthermore, there is an opportunity to analyse the same agreements from a different perspective, if the common knowledge is not equal [8], and there is a way to quantify how big this discrepancy is. The amount of discrepancy can provide us with a prioritized list of agreements that can benefit the most from a third-party intervention, like an independent assessment. This assessment can be part of a validation procedure performed by a certification agency to bridge the gap in common knowledge.

References

[1] P. Koopman and M. Wagner, 'Challenges in Autonomous Vehicle Testing and Validation', SAE International Journal of Transportation Safety, vol. 4, no. 1, pp. 15–24, Apr. 2016.

[2] R. Johansson, J. Nilsson, and A. Larsson, 'Safe Transitions Between a Driver and an Automated Driving System', International Journal On Advances in Systems and Measurements, vol. 10, no. 3 and 4, pp. 100–110, Dec. 2017.

[3] 'IEEE SA - 7001 - Transparency of Autonomous Systems'. [Online]. Available: http://standards.ieee.org/develop/project/7001.html. [Accessed: 25-July-2018].

[4] 'Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles J3016_201609'.

[5] ERTRAC, 'Automated Driving Roadmap', 2017.

[6] 'Data protection', European Commission - European Commission. [Online]. Available: https://ec.europa.eu/info/law/law-topic/data-protection_en. [Accessed: 25-July-2018].

[7] J. F. Allen, 'Maintaining knowledge about temporal intervals', vol. 26, no. 11, p. 12, 1983.

[8] Z. Hellman, 'Almost common priors', International Journal of Game Theory, vol. 42, no. 2, pp. 399–410, May 2013.