Safe System Behavior: From Requirements to Concrete Specifications.

A discussion paper describing an approach, submitted by the experts from Germany.

1. Background

The current draft of Document No 5 (*GRVA-FRAV-02-05-rev.2*) contains a set of miscellaneous requirements and specifications for the safe system behavior, however the process of the derivation of specifications from requirements is not transparent.

An example for this is the definition of top-level requirements for safe system behavior. While the revised framework document on automated/autonomous vehicles states the following as a key issue and working principle for further development (ECE/TRANS/WP.29/2019/34/Rev.2, paragraph 9b):

> "When in the automated mode, the automated/autonomous vehicle should be free of unreasonable safety risks to the driver and other road users and ensure compliance with road traffic regulations.",

there is also a safety vision included in paragraph 7 that goes beyond the principle as stated above:

> "This level of safety implies that an automated/autonomous vehicle shall not cause any non-tolerable risk [introduce unreasonable risks], meaning that automated/autonomous vehicle systems, while in automated mode, shall not cause any traffic accidents [incidents/events] resulting in [destruction of property,] injury or death that were reasonably foreseeable and preventable."

It is not transparent how the fact that an automated vehicle is *expected to not cause a number of accidents* (namely those that are not reasonably foreseeable and preventable) has been derived from the requirement that an automated vehicle shall be *free of unreasonable safety risks*.

2. Approach

2.1 Basics

Given that the framework document asks the subsidiary bodies of WP.29 to further develop requirements based on the key issues and common working principles, we propose to define functional requirements for automated vehicles in a transparent top-down-approach, starting with the most top-level requirements that are available and going down to more detailed specifications until a sufficient level of detail has been achieved. This sufficient level of detail has been reached when the resulting requirements fulfil the following properties: unitary (the requirement addresses only one thing), consistent, unambiguous, verifiable[1].

Ideally the resulting requirements will allow a formalized way of verifying the vehicle behavior with figures and dimensions, like requirements in a mathematical formulation. This result would then allow a pass-fail-assessment of any given set of vehicle behavior (typically described as vehicle motion variables over time and in relation to the surroundings and surrounding traffic), regardless of whether this behavior has been measured on the test track or in real traffic or has been calculated using

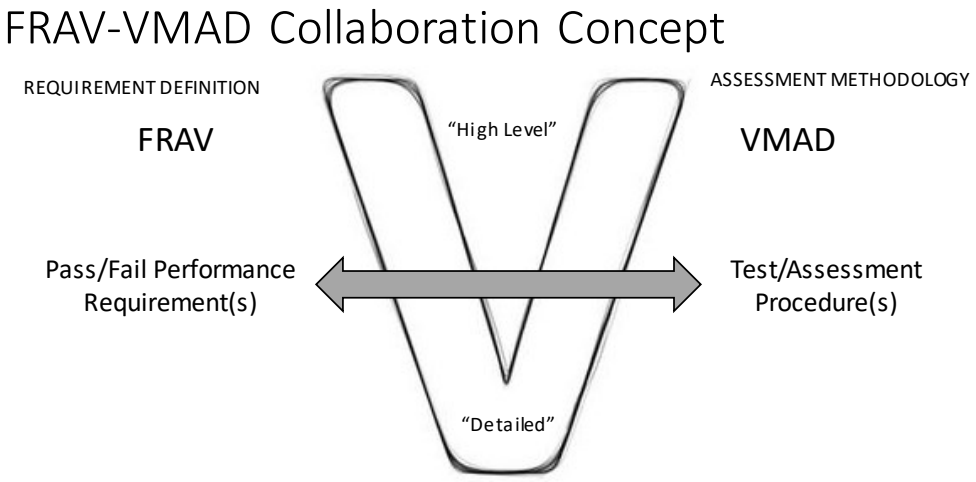
[1] [1]. For more details refer e.g. to https://en.wikipedia.org/wiki/Requirement

simulation methods. Criteria for the pass-fail-assessment might e.g. be speed reduction, collision avoidance, time headway etc.

This approach would also allow a clear separation of the work between the FRAV and VMAD working groups:

> FRAV is to deliver requirements for automated vehicles that allow a pass-fail-assessment of any vehicle behavior, regardless how this vehicle behavior has been generated. FRAV does not look into test procedures, data bases etc., FRAV does not deliver any new validation methods.
>
> VMAD is to deliver methods to generate vehicle behavior data, especially with new assessment and test methods. The pass-fail-assessment will be carried out with a set of requirements defined within the FRAV group. VMAD in this context should not deliver any additional requirements (like classifying situations as preventable or not preventable).

The roles of the two working groups would then be the left and the right part of the V model for product development, as it was proposed in FRAV-01-15-rev.1 (see figure below).



2.2. Top-level requirements and first-level specifications

As a first step, the contents of the framework document shall serve as a starting point:

> "When in the automated mode, the automated/autonomous vehicle should be free of unreasonable safety risks to the driver and other road users and ensure compliance with road traffic regulations." (Paragraph 9b).

This allows the formulation of the following top-level requirements:

> The automated vehicle follows road traffic regulations.
>
> The automated vehicle is free of unreasonable safety risks to the driver.
>
> The automated vehicle is free of unreasonable safety risks to other road users.

According to the V-model for product development, only the top-level requirements are called requirements. Everything derived from these requirements is referred to as specifications. Yet, also specifications (of all levels!) need to fulfil the characteristics for good requirements.

The framework document itself derives the following specifications from the requirements:

> "This level of safety implies that an automated/autonomous vehicle shall not cause any non-tolerable risk [introduce unreasonable risks], meaning that automated/autonomous vehicle systems, while in automated mode, shall not cause any traffic accidents [incidents/events] resulting in [destruction of property,] injury or death that were reasonably foreseeable and preventable." (Paragraph 7).

A more formal version of these specifications is the following:

> The automated vehicle shall not cause any traffic accidents that are reasonably foreseeable and preventable resulting in destruction of property.
>
> The automated vehicle shall not cause any traffic accidents that are reasonably foreseeable and preventable resulting in injury.
>
> The automated vehicle shall not cause any traffic accidents that are reasonably foreseeable and preventable resulting in death.

These further specifications are not verifiable yet. It is considered to be the task of FRAV to operationalize the words "unreasonable risk" and "reasonably foreseeable and preventable".

One major issue is that these specifications are covering just accidents caused by the automated vehicle, but they are not addressing the performance of the automated vehicle to avoid or mitigate accidents caused by other traffic participants. In addition, it is questionable why the automated vehicle should be allowed to cause unpreventable or unforeseeable accidents. A way out would be to remove the "preventable/foreseeable" from these specifications (relevant only for accidents *caused* by the ADS) and add a set of specifications addressing accidents *NOT caused* by the ADS.

> Case a) Specifications for accidents caused by the ADS:
>
> The automated vehicle shall not cause any traffic accidents resulting in destruction of property.
>
> The automated vehicle shall not cause any traffic accidents resulting in injury.
>
> The automated vehicle shall not cause any traffic accidents resulting in death.
>
> =
>
> The automated vehicle shall not cause any traffic accidents.
>
> Case b) Specifications for accidents caused by other traffic participants:
>
> The automated vehicle shall avoid any traffic accidents that are reasonably foreseeable and preventable resulting in destruction of property.
>
> If this is not possible, the automated vehicle shall avoid any traffic accidents that are reasonably foreseeable and preventable resulting in injury.
>
> If this is not possible, the automated vehicle shall avoid any traffic accidents that are reasonably foreseeable and preventable resulting in death.

These sets of requirements now cover both cases (accidents caused and not caused by the ADS) and therefore are a more comprehensive set of requirements. However, they are not yet verifiable, since no definition of foreseeable, preventable, and possibly reasonably is provided, and the definitions influence the resulting performance specifications to a large extend. A proposal for a usable definition could look like this:

> "preventable" means accidents that are preventable using state-of-the-art perception and actuation technologies within the limits given by physical laws. Any sensor that is available in a production vehicle at the time of writing shall be part of the sensor portfolio; sensor fusion technologies shall be used to combine different sensor systems. (this means: use all available sensors and combine them).
>
> "reasonably foreseeable" means all accidents that would not occur if the system had reacted with the minimum technically possible reaction/latency time of the sensor/actuator system.

As in ALKS it could be the task of FRAV to determine which system and physical properties (like e.g. built-up-times or achievable decelerations) can be regarded as state-of-the-art or technically possible and which not. It could be the task of FRAV to further specify the physically possible system behavior by deriving concrete expected performances for specific accident-prone maneuvers, as next steps. This approach allows to replace the terms *preventable and reasonably foreseeable* by *physically avoidable*.

It could then be subsequent task of FRAV to specify system performance for the rest of situations in which the accident is not avoidable.

3. First proposal for concrete Top-Level Specifications

The following set of verifiable specifications and definitions could serve as a proposal for the future work in FRAV:

> (1) The automated vehicle shall not cause any traffic accidents.
>
> (2a) The automated vehicle shall follow road traffic regulations.
>
> (2b) The automated vehicle shall limit the resulting horizontal acceleration to values below [3] m/s² ([2,4]m/s² in case of standing passengers). *[3 m/s² can be considered to be comfortable for passengers, thus it will not cause any injuries for unprepared vehicle occupants. The value 2.4 m/s² has been found to be appropriate for standing passengers]*.
>
> (3) If (2a) or (2b) are not possible due to irregular actions of other traffic participants, the automated vehicle shall carry out an emergency maneuver according to specifications 4-9.
>
> (4) The automated vehicle shall avoid any traffic accidents that are physically avoidable (according to specifications 7-9) resulting in destruction of property.
>
> (5) If this is not possible, the automated vehicle shall avoid any traffic accidents that are physically avoidable (according to specifications 7-9) resulting in injury [of first vulnerable road users, second vehicle occupants]. It shall mitigate the destruction of property.
>
> (6) If this is not possible, the automated vehicle shall avoid any traffic accidents that are physically avoidable (according to specifications 7-9) resulting in death [of first vulnerable road users, second vehicle occupants]. It shall mitigate injuries.
>
> (7) The reaction of the ADS in terms of giving a braking or steering demand shall start latest if the movement of the other traffic participant or object was visible for [0.5 s]
>
> (8) During the emergency maneuver the longitudinal and lateral accelerations are allowed to exceed [3] m/s² ([2,4] m/s² in case of standing passengers), however it shall under all

circumstances achieve at least these values. If necessary in the emergency maneuver the vehicle shall decelerate up to its full braking performance and/or perform an evasive maneuver when appropriate. Full braking performance shall be reached within [0.5 s].

(9) In case the collision cannot be avoided the ADS shall not deactivate or unreasonably switch the control strategy in order to mitigate the consequences of the collision as far as possible.

These specifications do completely describe the required behavior of an automated vehicle on top level, but they could probably require more specifications for specific maneuvers.

It could be the task of FRAV to further operationalize the top-level specifications by second-, third- etc. level specifications (define time headways, TTCs etc.), e.g. in order to fix how the automated vehicle shall follow road traffic regulations and in order to determine the borderline for “physically avoidable”.