Workshop Prospectus
Consensus Safety Measurement Methodologies for ADS-Equipped Vehicles
June 25-26, 2019
Location: NIST Gaithersburg Campus

Overview

The automotive industry is planning to deploy levels 3 and above automated driving systems (ADS; SAE J3016) in the very near term with most major manufacturers (OEMs) targeting the early 2020s [1]. Community consensus around reliable, broadly-acceptable performance metrics for assessing ADS safety can facilitate the successful achievement of these deployments. This workshop will explore the need for a common method to evaluate safe performance operation of emerging ADS-equipped vehicle concepts and possible paths toward consensus

ADS-equipped vehicles are cyber-physical systems (CPS). CPS comprise interacting digital, analog, physical, and human components engineered for function through integrated physics and logic. These systems will provide the foundation of our critical infrastructure, form the basis of emerging and future smart services, and improve our quality of life in many areas [2]. In addition to ADS-equipped vehicles, CPS examples include smart grid, advanced manufacturing, and smart cities. Performance measurement and assurance in such systems will require new evaluation strategies and approaches that accommodate the complex, hybrid nature of these systems. This workshop will examine concepts for safety metrics, formal models, testing protocols, taxonomies, and other approaches for measuring the safety of ADS-equipped vehicles.

This workshop will explore the potential for a consensus-based, community-wide collaborative effort in advancing the safety of ADS-equipped vehicles by developing the appropriate metrics and methods. Visit the website https://www.nist.gov/news-events/events/2019/06/consensus-safety-measurement-methodologies-automated-vehicles-workshop for registration and further workshop information in the next few weeks.

Background

Industry-wide, globally-accepted methods for effective safety evaluation require vehicle-based measurements of ADS related behaviors. Based on an open, consensus-based standards that can facilitate successful ADS deployments that interoperate safely and securely in traffic, enhance mobility, save energy, improve environmental quality, and the quality of life. The emergence of such methods can be accelerated through a community-based approach and this workshop is intended to explore opportunities for such an approach.

NIST released in 2017 a Framework for Cyber-Physical Systems [2] that establishes the functional decomposition, the facets (groupings of activities) for and the concerns that drive the conceptualization, realization and assurance of CPS. This CPS methodology enables an approach to AV design, development and assurance that exposes the analysis of safety at every level of the overall system and relates safety concerns and actions to those associated with the high level concern of trustworthiness aspect of the CPS Framework (like security, privacy, resilience and reliability), as well as other aspects, including the Functional Aspect, Data Aspect and the Business Aspect. The benefit to AV Safety is an approach to AV Safety that integrates with subsystem/component decomposition and the broader concerns about the CPS.

Effective methods for AV safety measurement can build on existing safety measurement approaches but must overcome the challenges associated with applying them to automated vehicles. For example, rule-based and statistical methodologies are two fundamental approaches currently used to assess vehicle safety. Rule based assessment involves the definition of a comprehensive set of rules that precisely define what it means to function safely, and which vehicles can be empirically tested against. Statistical approaches track the performance of vehicles over millions of miles of real-world operation and calculate their probability of safe operation as an extrapolation of their observed frequency of safety violations. Both approaches are challenged by autonomous vehicles. These challenges include determining whether a comprehensive set of AV rules can be derived that is fully testable and whether statistical approaches can be designed to accurately assess the likelihood of correct and safe performance when vehicles are faced with all events, including rare, anomalous events that may not be well represented in historical performance databases.

Safety assessment is further challenged by individual technologies designed to enable vehicle autonomy. An example is deep neural nets, which are currently used to implement AV system perception and decision-making. Deep nets approximate non-linear relationships between sets of inputs and outputs, however, the statistical nature by which they do so is not understood in a manner that would enable verification or validation of safety. Corrections will typically be implemented by submitting them to new training data, but it is not clear how this data will impact overall performance (the new data can negatively impact performance against other previously validated safety requirements). Open questions are the degree to which complete retesting of compliance to all requirements, and/or the collection of extensive, new real-world operational and simulation data may be required for updates.

The challenges enumerated above and others underline that meeting the need for measuring AV safety "requires a shift from traditional paradigms of vehicle safety testing to a more comprehensive view of this matter" [3]. To date, academia, government and industry have ideated a variety of initial, candidate methodologies and frameworks. These include methods for implementing safety on the system, frameworks for modeling safe driving behavior and frameworks for measuring system safety. Examples of each are summarized below to frame some of the topics to be discussed at the proposed workshop. The inclusion of these examples should not be construed as endorsement of them nor as a complete characterization of the candidate methodologies. It is expected that detailed information and examples will be explored at the workshop.

1. NHTSA's "A Novel Method to Evaluate the Safety of Highly Automated Vehicles" [3]

This paper introduces "the concept of an Instantaneous Safety Metric. The fundamental constructs of its implementation, both geometric region of interest and classification of vehicle interactions are discussed".

This methodology aims to determine the action to be taken by a subject vehicle by considering the possible actions of the surrounding vehicles and the impact of those actions on the subject vehicle's choice of action. This method considers the ADS' operational design domain (ODD), whether there is a choice of action for the vehicles that 'continues to provide' an escape path or a path that avoids an eventual collision and assesses the severity of the maneuver required to follow that path.

All of the above steps are taken at one point in time and the positions/headings and impact and subject vehicle choice of actions are determined for discrete points in time in the future. The space occupied by a vehicle at a given point in time, as a result of any choice of action, is referred to as the vehicle's profile. The 'possible space' of a vehicle at a future point in time is the union of all profiles of that vehicle at that

point in time over all possible choices of action of the vehicle. The intersection of all profiles at a given point in time in the future is the vehicles unavoidable space at that time.

Given a point in time in the future and two vehicles A and B, if there is no intersection between the spaces of A and B at time T this is termed an 'impossible interaction'. If the intersection is non-empty, then this is termed a 'possible interaction'. A 'critical interaction' occurs when there exists a set of profiles for traffic vehicle(s) that intersects all profiles for the subject vehicle at a given point in time and it is called an 'immanent interaction' if all possible actions by traffic vehicles result in contact with the subject vehicle at some point in the future.

There are multiple opportunities for future work associated with this approach to assessing the safety of ADS decision-making, including:

> Assessing the feasibility of the required calculations. For example, the real action sets of the vehicles involved require one to consider an infinite number of possible inputs, whether it be for lateral or longitudinal motion (i.e., braking, steering, and acceleration). Extending those considered herein, full acceleration and full deceleration as well as full right and full left lateral motion, would be useful in assessing the 'worst case' scenarios associated with the actions.

One potential approach to achieving feasibility could be:

- Bounding the search for critical or imminent interactions,
- Computing the escape paths that would not require severe driving maneuvers; and
- Identifying feasible action sets that demonstrate corresponded to the worst cases.

Progress on relevant research questions could enable a demonstration of 'feasibility and correctness' of the approach.

The authors state in their concluding remarks: "The proposed method of safety analysis provides a quantitative window into the predominantly qualitative world of subtle and nuanced traffic interactions. The authors believe that this method and approach provide a sure path toward developing a tool which can be used to quantify ADS performance in simulation, test-track, and on-road evaluations."

2. Intel/Mobileye's "On a Formal Model for AV Safety" [4]

Intel/Mobileye have proposed having an open-source and technology-neutral model for the safety of ADS decision making process. A vehicle would be deemed to be *safe relative to the model*, if its driving behaviors conformed to the model. In their paper [4], the authors propose a driving philosophy that states that ADSs have a *responsibility* to drive safely. As a duty to the ADS-equipped vehicle occupants and other vulnerable road users in the environments in which the ADS is operating. Moreover, it assert the industry has a *responsibility* to deliver a verifiably safe product.

The proposed approach is classified as flexible. It suggests that an ADS must be *sensitive* to what it means to drive safely. Different geographies around the world may have slightly different definitions of what it means to drive safely and may have correspondingly different risk tolerances and assumed behavior for the human drivers or other agents in that region. Thus, any formal model for safety must be *sensitive* to geographical differences to ensure that ADS-equipped vehicles only make safe driving decisions as agreed to and defined by the people and governments in the different geographies in which AV's will operate.

Responsibility Sensitive Safety (RSS) is an open and transparent technology-neutral, formally-verifiable model for safe decision making of an ADS that features specific parameters that can be adjusted and tuned as desired by the people and governments in which ADS-equipped vehicles will be deployed.

RSS is based on human centered concepts of what it means to drive safely. Things like, always maintain a safe following distance, be cautious in areas where other vehicles or pedestrians may be included, and right of way is given, not taken.

These are common sense human principles for safe driving may not be found in any legal 'rules of the road' but instead are a set of safety principles that are widely understood and accepted by humans as the safe way to drive. It is suggested in the paper that humans that strictly follow these principles are among the safest drivers on the road today. Conversely, humans that follow a lead vehicle too closely, are careless in areas where others may be present, and aggressively taking right-of-way from others are considered as not driving safely, and as often the cause of crashes.

This paper defines the following concepts relative to the RSS model:

- **Driving Safely**: A vehicle is driving safely if its decisions conform with the RSS model.
- **Driving Unsafely**: A vehicle is driving unsafely if its decisions violate the RSS model.
- **Dangerous Situation**: A state of the vehicle such that there is the possibility of an accident and is typically the result of some unexpected behavior from other road users.
- **Safe State**: A Safe State is a measure of the current state of the AV with respect to all other relevant agents around the vehicle such that no Dangerous Situation exists. An ADS in a Safe State by definition can neither be causing nor involved in a Dangerous Situation.
- **Danger Threshold**: The Danger Threshold is the moment in time at which a vehicle transitions from a Safe State into a Dangerous Situation. In other words, the danger threshold is the moment in time at which a vehicle began driving unsafely by initiating a Dangerous Situation.
- **Proper Response**: A Proper Response is what the AV should do to get out of a Dangerous Situation and return to a Safe State.

The purpose of publishing the model openly, in a series of academic papers, is not to claim that the model is perfect. Instead, the goal is to help start an open/honest conversation about safety verification of the ADS decision making across the community so that we can arrive together at a common solution for verifiably safe vehicles. Research is needed to compare the model against new scenarios, to determine whether RSS's "Safe State" can be realized under all ODDs and whether a complete RSS rule set can be specified or how this rule set can be fully implemented and validated in the real world .

The RSS model emphasizes safe driving principles. Additional topics for future research along these lines would include:

- What should be our method for developing a set of safe driving principles?
- How are we to validate these principles?
- How are we to test whether an ADS adheres to the principles?

3. Rand Corporation's "Measuring Automated Vehicle Safety, Forging a Framework" [5]

The Rand Corporation defines a framework for measuring ADS safety. It is stated that "This report is intended to foster broader understanding and discussion of ADS safety that can aid the public and policymakers in the debate over a new product category expected to reshape economic, social, and community activity." Summarizing all the Rand team's findings and their proposed framework is beyond

the scope of this white paper. Readers can access the full report at https://www.rand.org/content/dam/rand/pubs/research_reports/RR2600/RR2662/RAND_RR2662.pdf

Instead, it is more instructive to summarize the pertinent recommendations made in this report to the automated vehicle community. In addition to the proposed framework for technology- and vendor-neutral safety measurement, other recommendations are:

1. Develop a formal protocol for the demonstration of AVs (in simulators, on private tracks and on public roads with safety drivers) to facilitate comparisons across companies and evidence of safety for the public and policymakers.
2. "During development, demonstration, and early deployment, when sufficient exposure has not been accumulated to allow for statistically valid comparison of rates, outcomes (e.g., crashes or an absence of crashes) should be evaluated as case studies."
3. Enable data-sharing (both between companies and with government) to improve safety across the industry and to enable real-time sharing (e.g., around roadworks). Regarding outcome measures, a protocol for reporting to government entities could be codified in terms of measures, context, format, frequency, data security, governance, and other factors.
4. "A formal taxonomy around ODD is needed. Such a taxonomy should specify how ODDs convey where, when, and under what circumstances the AV can operate."
5. Given the challenges of measuring safety where the system changes, constantly and at irregular intervals, more research is needed.

4. NHTSA's "A Framework for Automated Driving System Testable Cases and Scenarios" [6]

The following summary is extracted from the NHTSA report. "The goal of this research was to develop an example of a preliminary test framework for ADS that are in development and may come to market in the near to mid future. The following steps were conducted to support the development of the sample test framework:

1. Identify concept ADS
2. Identify attributes that define the operational design domain (ODD)
3. Identify object and event detection and response (OEDR) capabilities
4. Identify and assess failure modes and failure mitigation strategies

Technologies of interest in this work included light-duty automated driving functions that fell within Level 3 (L3) to Level 5 (L5) of the SAE levels of driving automation (SAE International, 2018). The functions were identified based on prototype vehicles and conceptual systems. A literature review which included popular media, press releases, technical journals, and conference proceedings was performed. This review identified potential concept ADS being developed or proposed by original equipment manufacturers (OEMs), suppliers, technology companies, and other organizations. The identified ADS were categorized into a set of generic names. The terminology was modified to ADS features (as opposed to functions) to more closely align with the standardization community's language. Twenty-four conceptual features were identified, and although a thorough search was conducted, the list is not exhaustive.

The ODD describes the specific operating domains in which the ADS is designed to function. The ODD will likely vary for each ADS feature on a vehicle and specifies the condition in which that feature is intended and able to operate with respect to roadway types, speed range, lighting conditions, weather conditions, and other operational constraints. The ODD is specified by the technology developer, and the ADS should be able to identify whether it is operating within or outside of that ODD. This taxonomy is hierarchical and includes the following top-level categories: 1) Physical Infrastructure, 2) Operational Constraints, 3) Objects, 4) Connectivity, 5) Environmental Conditions, and 6) Zones. OEDR refers to the

subtasks of the DDT that include monitoring the driving environment (detecting, recognizing, and classifying objects and events and preparing to respond as needed) and executing an appropriate response to such objects and events (i.e., as needed to complete the DDT and/or DDT fallback; (SAE International, 2018).

Potential failures could have significant impacts, ultimately resulting in collisions that could damage the vehicle or harm its occupants or other roadway users. Failure mitigation strategies, including both fail-operational (FO) and fail-safe (FS) techniques, were then identified and analyzed. FS techniques are used when the ADS cannot continue to function, and may include options such as the following: 1) Transitioning control to fallback-ready user, 2) Safely stopping in lane, 3) Safely moving out of travel lane/park. FO techniques can be used to allow the ADS to function at a reduced capacity, potentially for a brief period of time or with reduced capabilities, and may include options such as adaptive compensation – weighting data from a complementary component or subsystem more heavily (e.g., weighting camera data more heavily if lidar fails). The appropriate failure mitigation strategy is highly dependent on the nature of the failure and the initial conditions under which the failure occurs."

Workshop Objectives

Other concepts for consideration could include geofencing for ADS-equipped vehicle operational design domain, examples from other sectors such as Federal Aviation Regulations Part 135 for commercial aircraft, and others that workshop participants may put forward.

The proposed workshop objectives are as follows:

- Develop ideal criteria for what any approach to automated vehicle decision-making safety must satisfy.
- Review existing or proposed methods to establish safety requirements and safety metrics.
- Identify gaps and key challenges, and explore opportunities for progress, including alternative methods that should be considered.
- Explore consensus around priorities for enabling progress on the gaps and key challenges that are identified.
- Publish a review of the identified ideal criteria, gaps and challenges, possible methods to address these gaps and articulate a preferred path to developing ADS safety requirements and testing.
- Explore opportunities for the workshop attendees to continue to collaborate, on the identified gaps, to drive towards consensus safety metrics across the community**.**

Please join us June 25th and 26th at NIST in Gaithersburg, Maryland to contribute to this community-wide effort to advance performance metrics to ensure ADS-equipped vehicle safety.

References

[1] "The Self-Driving Car Timeline – Predictions from the Top 11 Global Automakers", July 19, 2018, Jon Walker, from https://emerj.com/ai-adoption-timelines/self-driving-car-timeline-themselves-top-11-automakers/

[2] NIST Framework for Cyber-Physical Systems, Special Publications: SP 1500-201 (https://www.nist.gov/publications/framework-cyber-physical-systems-volume-1-overview), SP 1500-202 (https://www.nist.gov/publications/framework-cyber-physical-systems-volume-2-working-group-reports) and SP 1500-203 (https://www.nist.gov/publications/framework-cyber-physical-systems-volume-3-timing-annex)

[3] NHTSA's "A Novel Method to Evaluate the Safety of Highly Automated Vehicles" Joshua L. Every Frank Barickman, John Martin Sughosh, Rao Scott Schnelle, Bowen Weng, Paper Number 17-0076
[4] On a Formal Model of Safe and Scalable Self-driving Cars", last revised 27 Oct 2018, Shai Shalev-Shwartz, Shaked Shammah, Amnon Shashua., Cornell University Library, arXiv.org > cs > arXiv:1708.06374
[5] Measuring Automated Vehicle Safety, Forging a Framework, Laura Fradde-Blanar, Marjory S. Blumenthal, James M. Anderson, Nidhi Kalra, Rand Corporation, 2018, ISBN: 978-1-9774-0164-9
[6] Thorn, E., Kimmel, S., and Chaka, M. (2018, September). A framework for automated driving system testable cases and scenarios (Report No. DOT HS 812 623). Washington, DC: National Highway Traffic Safety Administration.