TECHNICAL CHALLENGES FOR FULLY AUTOMATED DRIVING SYSTEMS

Steven E. Shladover
California PATH Program
University of California, Berkeley
1357 South 46th Street, Building 452
Richmond, CA 94804
+1-510-665-3514, steve@path.berkeley.edu

SUMMARY

Recent media coverage has created the impression that the technical challenges for fully automated driving systems are largely solved and that the primary impediments to widespread deployment of such systems are legal, institutional and customer acceptance issues. This paper explains the severity of the technical challenges that are not close to being solved, but that will need to be solved, before fully automated driving can be shown to be at least as safe as today's manual driving. The technical difficulties are illustrated by comparison with the automation of commercial aircraft operations. This supports the author's contention that the technical challenges are at least as difficult as the non-technical challenges to implementation of full automation.

Key words: automated driving, highway automation, driverless cars, autonomous vehicles, self-driving cars

INTRODUCTION – TRAFFIC SAFETY BASELINE

People have debated how safe automated vehicles will need to be in order to be socially acceptable, without much firm evidence to support their estimates (which have ranged from twice to ten times as safe as today's driving, all the way up to perfectly safe). It is not necessary to speculate about how large the improvement factor needs to be, but we can assume as a bare minimum that they will need to be no less safe than today's driving. Taking today's traffic safety statistics as a starting point, it is possible to quantify the minimum required intervals that will have to be achieved between fatal crashes and between injury crashes to demonstrate safety comparable to today's driving. Based on the latest U.S. traffic safety statistics published in the *2011 Traffic Safety Facts* (1), and assuming an average driving speed across all categories of roads of 30 mph, this leads to the following calculations:

$$\frac{100 \times 10^6 \text{ vehicle miles}}{1.10 \text{ fatalities}} \times \frac{32{,}367 \text{ fatalities}}{29{,}757 \text{ fatal crashes}} \times \frac{\text{hour}}{30 \text{ miles}} = 3.3 \text{ M hours/fatal crash}$$

$$\frac{100 \times 10^6 \text{ vehicle miles}}{75 \text{ injured persons}} \times \frac{2{,}217{,}000 \text{ injured persons}}{1{,}530{,}000 \text{ injury crashes}} \times \frac{\text{hour}}{30 \text{ miles}} = 64{,}400 \text{ hours/injury crash}$$

The sobering result is that fatal crashes occur about once per 3.3 million hours of vehicle travel and injury crashes occur about once per 64,000 hours of vehicle travel. How many

complicated, software-intensive electro-mechanical or electronic systems can claim comparable mean time between failure (MTBF) capabilities today? The comparable targets for other advanced industrialized countries are likely to be within a factor of two of these numbers (above or below).

These are very daunting targets to achieve for systems that are:

- Affordable to consumers (hence unlikely to include significant hardware redundancy)
- Operated and maintained by non-professional owners
- Electro-mechanical (mechatronic) systems that do not benefit from Moore's Law improvements in technology
- Software-intensive
- Required to operate in a harsh and unpredictable hazard environment.

WHAT DOES A FULLY-AUTOMATED DRIVING SYSTEM NEED TO DO?

It is important to identify what the responsibilities of an operational fully-automated driving system will have to be. If such a system is to be able to achieve Level 5 automation as defined in SAE J3016 (2), it must be able to take complete responsibility for driving under all conditions in which a typical skilled human driver can operate (the full range of traffic and weather conditions). If it is operating as an automated taxi to carry children, blind or otherwise disabled passengers, or as a driverless freight vehicle, or if it is serving as a repositioning system for a shared-vehicle system, there is no driver available to act as the safety backup so the system must safely handle virtually ALL hazards in the external driving environment and must have sufficient fault detection, identification and accommodation capabilities to recover from virtually ALL internal faults that it may encounter.

External Hazards

The external hazards include all other road users and objects on or adjacent to the roadway, which the system must be able to detect under all environmental conditions in which human drivers would be capable of driving. Static hazards such as roadside curbs, overhanging tree limbs, low clearance bridges and even road surface irregularities could potentially be documented in a detailed map database provided to the vehicles, so there is a reasonable chance of informing vehicles about them before they are encountered. Much more serious concerns revolve around dynamic hazards, which can appear with virtually no advance notice and must be detected by vehicle-based sensors.

The simplest hazards to detect are the other vehicles with which the automated vehicle is sharing the road, since they are large and well-defined targets, and they normally move in relatively predictable ways. However, in some cases other vehicles can be more challenging to detect and interpret:

- Vehicles entering the roadway from blind driveways
- Vehicles violating traffic laws
- Vehicles moving erratically after having been engaged in crashes with other vehicles

- Law enforcement vehicles flashing their lights or using their sirens to alert drivers to respond in specific ways to their approach.

Non-vehicular objects sharing the roadway infrastructure pose more serious challenges. These include:

- Pedestrians (including small children moving erratically)
- Bicyclists
- Law enforcement officers (or even private citizens in emergency conditions) directing traffic
- Animals (domestic pets in urban areas, large wildlife in rural areas)
- People opening the doors of adjacent parked cars
- Debris from other vehicles that have crashed
- Unsecured loads falling off trucks
- Sand, gravel or rocks from landslides
- Any objects on the road surface that could cause damage or disrupt the trajectory of the vehicle if they are hit by the vehicle
- "Negative obstacles" – potholes in the road surface that are extremely hard to detect but can damage the vehicle and disrupt its trajectory.

An automated vehicle's threat detection systems face a classic signal detection dilemma – they must detect all genuine threats (to save the vehicle from crashing) and reject virtually all non-threatening targets (to avoid spurious braking or crash avoidance maneuvers). It can be remarkably difficult to discriminate between the genuine and spurious threats under a full range of driving conditions, particularly when the genuine threats are hard to see but the spurious threats are highly visible. For example, a deep pothole or a brick in the vehicle's tire track is very hard to detect while approaching at highway speed, but it can cause serious damage and cause the vehicle to veer off course into the path of another vehicle if the vehicle hits it, so it is essential that it be detected. In contrast, a metallized mylar balloon drifting in front of the vehicle is a conspicuous target for radar, lidar or camera-based systems, so it can be detected easily, but the threat detection system must be intelligent enough to recognize that it is benign so that the vehicle is not commanded to make an emergency stop or lane change to avoid hitting it. These are merely anecdotal examples from among the vast array of threats and potential threats that automated driving systems will encounter every day. They must be able to correctly discriminate between the hazardous and benign targets virtually every time, even for targets that they have never previously encountered or been trained about.

Environmental Conditions

The threat detection and discrimination functions need to be performed successfully under all environmental conditions in which a vehicle user expects the automated driving system to be available. For Level 5 automation systems, this means all conditions under which human drivers would expect to be able to drive. The system therefore has to be able to handle a full range of operational, lighting and weather conditions, including:

- Electromagnetic pulse disturbances interfering with operation of vehicle systems (e.g., lightning discharge)
- Precipitation (rain, snow, mist, sleet, hail, fog of different types…)
- Other atmospheric obscurants (smoke, dust, …)

- Night conditions with no illumination other than the vehicle headlights
- Low sun angle glare
- Glare off snowy and icy surfaces
- Reduced road surface friction based on rain, oil, snow or ice on the surface
- High winds
- Road surface markings and signs obscured by snow or ice
- Road surface markings obscured by reflections off wet surfaces at night
- Signs obscured by foliage or displaced by vehicle crashes.

In some of these conditions, the information normally conveyed by the traffic control devices simply becomes unavailable to the vehicle, as it is unavailable to the drivers using the same roads. Drivers improvise and make assumptions about the missing information, and if they make incorrect assumptions they need to learn and adapt. Programming automated driving systems to make similar accommodations is not a simple matter.

Internal Faults

Since no system designed and built by humans can be expected to be perfect, the automated driving systems will create their own hazards as a consequence of their internal faults. This means that even if the systems are able to avoid many (or even most) of the crashes that are currently caused by human errors, they will introduce new categories of crashes caused by their internal faults. The large challenge of the system designer is to ensure that these faults are sufficiently rare that the overall safety of the system is improved. This is where the notion of "functional safety" enters system design. The most widely accepted approach to ensuring functional safety of automotive systems is the international standard ISO 26262 (3).

The objective of functional safety in ISO 26262 is the elimination of unacceptable or unreasonable risks to individuals caused by potential malfunctions in the electrical or electronic systems on a vehicle. Functional safety is different from active safety or passive safety. Active safety is concerned with systems aimed at avoiding crashes, while passive safety addresses mechanisms that minimize the severity of a crash if the crash cannot be avoided. The active or passive safety systems themselves need to be functionally safe because otherwise these systems could also cause injury. Functional safety primarily focuses on the risks arising from random hardware faults, as well as systematic faults in system design, in hardware or software development, or in production, throughout the entire service period of the vehicle.

The major categories of internal faults are expected to include:

- Mechanical and electrical component failures
- Computer hardware glitches
- Sensor condition or calibration faults
- Software coding bugs
- System specification errors
- System design errors.

Within each of these categories, many diverse faults should be expected. A careful and systematic design process can reduce the frequency and severity of these faults, but they cannot be eliminated entirely based on current system engineering processes. Designs based

on hardware or functional redundancy and voting systems can significantly reduce the consequences of the first four categories of internal faults by providing built-in backup capabilities, but this has serious cost consequences. If a vehicle's automated driving system needs to be equipped with triple redundant components plus the means for disengaging a faulty component, it could more than triple the cost of that automated driving system. A robust software development process would require separate teams to develop alternative code for each safety critical function and to have both sets of code running continuously and comparing their outputs for cross-validation. This would more than double the software development level of effort, which is already likely to be the dominant component of the system development cost.

The system design and specification errors are more difficult to handle because (by definition) they are not recognizable at the time the system is under development, so they remain hidden until they produce faulty behavior in practice. New methods of system design are needed to identify these problems efficiently early in the design process. Formal mathematical methods for verifying system specifications are so complicated that they have only been applied to very simple scenarios, and in those cases they have still required major investments of effort.

An example of the diversity of internal faults that can be expected in an automated driving system is shown in Table 1, which has been modified from the version published in (4).

Table 1 - Fault Conditions for Vehicle Automation Systems

Subsystems	**Fault Types**
Sensors	No output
	Output pegged at an extreme value
	Random noise
	Bias
	Drift
	Sensitivity change
	Interference
	Calibration error
Actuators	No response
	Response pegged at an extreme value
	Random noise
	Bias
	Intermittent operation
	Deadband
	Hysteresis
Control computer	Loss of power
	Software crash
	Operating system deadlock
	Overloaded processor or memory
	Software bugs
	Controller design errors

In-vehicle networks	Loss of signal
	Interference
	Overload
	Intermittent connections
	Software bugs
Vehicle faults	Tire burst
	Engine failure
	Brake failure
	Electrical failure
Wireless communications	Loss of signal
	Interference
	Intermittent drop-outs
	Overloaded channels
	Jamming
	Noise
	Software bugs
	Spoofing
	Lack of acknowledgment

TECHNOLOGICAL LIMITATIONS

Current technology cannot support a vehicle's ability to drive itself under all the environmental conditions previously identified and in the face of all the potential faults described above, based on a variety of limitations:

- Sensor signal processing ability to correctly detect all true hazards in the driving environment (zero false negatives needed to avoid crashes)
- Sensor signal processing ability to correctly reject nearly all false positive hazard detections (near zero false positives needed to be acceptable to users, because an incorrect emergency response action will be frightening and could cause a secondary crash)

There is an inherent conflict between minimizing false negatives and false positives within the same sensor signal processing application, which is one of the fundamental tenets of signal detection theory. If the sensitivity of detection is set high, so that the system flags nearly all threats even if their probability of being seriously hazardous is limited, it is likely to generate a high rate of false positives unless the sensor's ability to discriminate between true and false positive threats is exceptionally high. That is typically only possible with very expensive sensors such as multi-spectral sensors that use a variety of independent physical principles to identify the nature of the threats. The same challenge applies in reverse when the sensitivity of detection is set low to minimize false positives; in that case it becomes very likely that some true positives (true threats) will be missed, and the vehicle will fail to respond to some emergencies.

- Limited ability of system designers to anticipate every hazard scenario that needs to be accommodated safely

The driving environment is exceedingly complicated and diverse, and crashes typically occur when rare combinations of hazards are encountered. Designers of automated driving systems will need to expend large efforts in identifying the hazard scenarios that their systems could encounter. Each designer can hypothesize thousands of potential hazards and can combine them in many ways to represent the driving environment in even greater richness of detail. These can serve as the basis for designing the hazard avoidance strategies or maneuvers that they will instruct the vehicle control systems to follow. However, regardless of their best efforts, vehicles operating on public roads will inevitably encounter new hazard scenarios that their designers could not anticipate and therefore did not instruct them about. This leads to the need for systems with the ability to learn so that they can determine how to respond to the "new" scenarios. Learning systems such as this are non-deterministic, which makes it impossible to prove their safety or to know with certainty how they will respond to the new scenarios.

- Ethical challenges to the development of software systems that must make life-or-death value decisions in responding to emergency scenarios

Human drivers are sometimes confronted with life-or-death value decisions that they need to make instantaneously in cases of imminent crashes. They have to live with the consequences of those decisions and if they are determined to have made a "wrong" decision in a judicial proceeding they must bear the financial responsibility for that decision, normally through an insurance settlement. Our society does not have any mechanism for incorporating these types of ethical dilemmas in software design and development for automated driving systems, yet those systems will inevitably encounter them. For example, even with "perfect" sensor information, an automated vehicle will at some time be forced to choose between hitting a motorcyclist, leading to the death of that person, and hitting a large truck, leading to serious injury to the occupants of its own vehicle. How does anybody formulate the software that makes that decision, and who determines that that software is suitable for use on a public road? How does one account for the different uncertainties regarding the probability of death of the motorcyclist or the probability of serious injury to the automated vehicle's occupants? These questions are essentially unanswerable, yet until they are answered it is hard to conceive of how a vehicle automation system can be introduced into general public use.

- Lack of efficient formal methods to verify the safety of complicated software logic under all possible combinations of input conditions

Formal methods have been applied to verification of some simple automated vehicle maneuver protocols (5), however that exercise showed that even for a severely simplified circumstance with a very limited number of possible outcomes the process was still extremely complicated. Considering the complexity of the scenarios that an automated vehicle will encounter in everyday driving, the complexity of the verification process will be daunting indeed. Methods have not yet been developed that can verify safety for systems of this complexity, which means that the technical approaches for developing the software logic will have to remain ad-hoc and heuristic. This, in turn, means that errors in logic will be unavoidable.

- Infeasibility of testing under all possible input conditions to prove that they have been managed safely

An automated driving system is a complicated real-time control system, which has to select and execute actions based on the input conditions it detects. The diversity of those input conditions in the general road driving environment is huge – consider the diversity of the target objects to be identified, their locations relative to the automated vehicle and their motion trajectories relative to the automated vehicle and to each other, combined with diversity of weather and lighting conditions. The hazardous conditions that are most likely to cause crashes will be the "edge cases" that involve unusual combinations of input conditions that are rarely encountered.

The system designers can try to anticipate these edge cases and create test scenarios to represent all the hazards, but the number of combinations of targets and their locations, speeds and directions of travel, multiplied by the combinations of weather and lighting conditions, are such that it will be impossible to cover all the possibilities in test cases. Simply enumerating the possible cases is a large enough problem, but executing all the tests would be impossibly expensive and time consuming. This means that it is not possible to empirically demonstrate the ability of the automation system to handle all the hazards it will encounter. That leads to the open question about how the system developers as well as the purchasers, insurers and safety regulators will be able to ascertain the adequacy of the safety of an automated driving system's design.

- Immaturity of machine learning technology and its non-deterministic behavior in general, which means that its safety cannot be ascertained.

One of the possible technical approaches for designing a system to operate in an extremely complex and unpredictable environment is machine learning, by which an artificial intelligence system continuously improves its performance by aggregating past experience, in much the way that people learn. This has the theoretical advantage that it does not require explicit programming to handle each new condition, but its performance is heuristic rather than deterministic, making it impossible to ensure that it will respond "correctly" or "safely" to every condition that it encounters. Indeed, there is no certain way to quantify how well it will perform, which leaves the problem of ensuring system safety unsolved.

ANALOGIES TO AIR TRAFFIC CONTROL

Road vehicle automation is orders of magnitude more difficult to achieve safely than autopilot operation of commercial aircraft or the automated landing systems that are in current use. These challenges can be estimated based on a variety of measures of difficulty tabulated below, and for each such challenge the number in the right column of Table 2 represents the number of orders of magnitude of increased difficulty for road vehicle automation compared to commercial aircraft automation.

Table 2 – Comparison of difficulty of road vehicle automation with commercial aircraft automation *by orders of magnitude*

Measure of difficulty	**Factor**
Number of potentially threatening targets a vehicle must track – this may only be one or two for a commercial aircraft, but could easily be a dozen for a road vehicle	1
Number of vehicles that a regional traffic management system needs to monitor at one time – This is likely to be in the range of hundreds for commercial aircraft at a major hub, but millions for road vehicles in a major metropolitan area	4
Accuracy of (lateral and longitudinal) range measurements needed relative to each target while cruising – For commercial aircraft, this would be of the order of 100 m in the lateral direction and 1000 m longitudinally. Road vehicles need to know their range to other vehicles laterally to an accuracy of 0.1 m and longitudinally 1 m, considering how much more closely they operate to other vehicles	3
Accuracy of range rate (speed difference) measurements needed relative to each target while cruising – For commercial aircraft, this should be known to within an accuracy of 10 m/s, but for road vehicles an accuracy in the range of 1 m/s is needed.	1
Amount of time available to respond to an emergency condition while cruising – With commercial aircraft cruising at altitude, a response within tens of seconds is likely to be acceptable because it takes much longer than that for the state of the aircraft to become critical. Road vehicles are in such close proximity to each other and to static obstacles that responses are needed within at most 0.1 second.	2
Acceptable cost for equipping each vehicle with automation capabilities (determining the extent to which redundancy can be employed) – Commercial aircraft flight control systems are economically viable at a unit cost of millions of dollars, but road vehicle automation systems will have to be sold in the range of thousands of dollars per unit to be viable in the market	3
Annual production volume of automation systems, providing opportunities for unit cost reductions – Commercial aircraft flight automation systems are produced in volumes of several hundred per year, while road vehicle automation systems could eventually be produced in volumes of several million per year.	-4
Sum total of orders of magnitude	**10**

The sum of the factors in the right column of Table 2 is 10. Remember that these numbers are the *exponents* on the factors of ten that apply here, so that 1 represents a factor of ten, 2 represents a factor of 100, etc. This means not that road vehicle automation is ten times as difficult as air traffic control automation, but rather that it is ten to the power 10 times as difficult – that is, ten billion times as difficult.

CONCLUSIONS

The feasibility of deploying a fully automated driving system that is at least as safe as today's driving is not determined by its ability to operate under normal conditions, but rather by its ability to handle the extreme "corner conditions" that are only encountered rarely, but are likely to become the causes of crashes when they do occur. Extensive research and

development work, including fundamental breakthroughs will be needed before these systems can become as safe as today's drivers. Those breakthroughs are needed in:

- Affordable sensing and sensor signal processing technology that can discriminate genuine hazards with a high enough degree of confidence that it can support the need for zero false negatives with near-zero false positives;
- Methods of specifying all of the automated driving system decision making processes (logic and algorithms) that can be systematically verified for completeness, correctness and safety;
- Methods of verifying the safety of software that has to be used to make safety-of-life decisions under unpredictably diverse operating conditions;
- Methods of rapidly detecting, identifying and accommodating failures in a complicated mechatronic system without significantly increasing its cost;
- Methods of proving the safety, durability and availability of a complicated mechatronic system that must operate under a very wide range of conditions, without being able to do exhaustive testing.

REFERENCES

1. National Highway Traffic Safety Administration (NHTSA), National Center for Statistics and Analysis, *Traffic Safety Facts Annual Report 2011*
2. *Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems*, Surface Vehicle Information Report J3016, SAE International, 2014.
3. *Road Vehicles – Functional Safety, International Standard ISO 26262*, International Organization for Standardization, 2009.
4. S.E. Shladover, "Cooperative (Rather than Autonomous)Vehicle-Highway Automation Systems", *IEEE Intelligent Transport Systems Magazine, Vol. 1*, No 1, 2009, pp
5. Hsu, A., Eskafi, F., Sachs, S. and Varaiya, P., "Protocol Design for an Automated Highway System", *Discrete Event Dynamic Systems, Vol 2*, No. 1, 1993, pp. 205-219.